Sólo para uso académico
Prohibida su comercialización

Artículo provisto por Biblioteca Central FAUBA

Development of a Regional Soil Productivity Index Using an Artificial Neural Network Approach

Josefina Luisa De Paepe* and Roberto Alvarez

ABSTRACT

Soil productivity indices represent ratings of the potential plant biomass production of soils. Inductive approaches determine productivity based on inferred effects of soil properties on yield. Conversely, deductive approaches use yield information to estimate productivity. Our objective was to compare the performance of both types of productivity indices for assessing regional soil productivity for wheat (*Triticum aestivum* L.) yield in the Pampas. Soil data from soil surveys and interpolated climate information were utilized. Wheat yield data from a 40-yr period and representing ~45 Mha were used. Inductive productivity indices showed a low correlation with observed yield (**$R^2 < 0.45$, $P = 0.05$). The best performance of deductive empirical methods was attained using a blind guess option, but soils could only be rated when yield data were available. Yield models based on the neural network approach had good performance ($R^2 = 0.614$, root mean square error [RMSE] = 548 kg ha^{-1}) and was used for regional productivity index development. This index could be extrapolated to soils for which yield data are not available, and its validation with yield averages was optimal ($R^2 = 0.728$, $P = 0.05$). Regional high productivity was achieved for combinations of medium to high levels of soil organic C and soil available water storage capacity variables, which showed a positive interaction. This methodology for assessing soil productivity based on an empirical yield-based model may be applied in other regions of the world and for different crops.**

Soil productivity can be defined as the capacity of a soil to produce plant biomass or crop seeds (Yang et al., 2003). To understand causal relationships with crop yield, soil types have been grouped and classified for comparison of their productivity (Sarkar et al., 1966; Allgood and Gray, 1978). The current reduction of soil productivity by human actions or natural processes has increased the need to develop methods for its quantification (Kim et al., 2000). A soil productivity index represents the potential of a particular soil to generate harvest products in relation to the optimal yield that an ideal soil would have during its first year of cropping (Huddleston, 1984).

The inductive approach attempted to determine soil productivity based solely on the inferred effects of various soil properties on yield (Huddleston, 1984). Numerical ratings were developed without the use of yield data, but they can afterward be calibrated or evaluated with yield information for particular regions and crops. A multiplicative model was developed by the FAO (Riquier et al., 1970) that considered soil moisture, drainage, rooting depth, texture, base saturation and salinity, organic matter content, mineral exchange capacity, clay type, and mineral reserves. Although the methodology was information intensive, it was applied in some countries and for some crops, and good performance was obtained. More recently, Kiniry et al. (1983) and Pierce et al. (1983) independently developed a simpler quantitative assessment of soil productivity related to soil suitability for crop root growth in the Corn Belt based on the work of Neill (1979). The two indices can be combined and called the Kiniry–Pierce soil productivity index. The soil environment was described in terms of sufficiency for root growth as influenced by five variables: soil available water storage capacity (SAWSC), aeration, bulk density, pH, and electrical conductivity. According to this multiplicative approach, any variable could be limiting, and the method is flexible in allowing modifications of the sufficiency factors. Several modified versions exist in which model statements were altered or new ones included to account for local soil characteristics (Rijsberman and Wolman, 1985).

Deductive productivity indices are generated using information from empirical models or process-based models to estimate crop yield and are therefore directly validated against yield data (Huddleston, 1984). Empirical models attempt to determine functional relationships of soil, climate, and management factors with yield using either existing or specially designed agronomic experiments (Jame and Cutforth, 1996). When the necessary data are available, these models offer a reliable method for investigating crop response to environmental conditions and are relatively simple to develop. Examples of deductive soil productivity indices based on empirical approaches could be found for corn (*Zea mays* L.) and soybean [*Glycine max* (L.) Merr.] in Illinois (García-Paredes et al., 2000), for soybean in Mexico (Yang et al., 2003), and for sorghum [*Sorghum bicolor* (L.) Moench] in Australia (Potgieter et al., 2005). Process-based models are built using mathematical equations to quantitatively model the interactions between environmental and crop factors (Sinclair and Seligman, 1996). The main limiting factor for

Facultad de Agronomía, Univ. de Buenos Aires, CONICET, Av. San Martín 4453 (1417), Buenos Aires, Argentina. Received 7 Feb. 2013. *Corresponding author (depaepe@agro.uba.ar).

Published in Agron. J. 105:1803–1813 (2013)
doi:10.2134/agronj2013.0070

Copyright © 2013 by the American Society of Agronomy, 5585 Guilford Road, Madison, WI 53711. All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Abbreviations: PET, potential evapotranspiration; RMSE, root mean square error; SAWSC, soil available water storage capacity.

their application is the high demand for information for the parameterization–validation procedures that may not always be available, especially in developing countries. Uncertainty can be high because this information is often taken from previous research conducted under inadequate environmental conditions or from expert opinion.

The Argentine Pampas covers an area of approximately 60 Mha, and because of its extension and yield potential it is considered one of the main cropping regions of the world (Satorre and Slafer, 1999). Wheat production is widespread across the region, under both humid and semiarid climates. Local soil productivity estimations were not developed, and the FAO methodology was used with soil survey information for assessing productivity without validation with yield data (Instituto Nacional de Tecnología Agropecuaria, 2013). In the Pampas, artificial neural network (ANN) models had a better performance for in-season prediction of wheat yield than common regression models (Alvarez, 2009). Artificial neural networks are empirical modeling techniques that are much simpler than process-based models and that have a high predictive quality (Özesmi et al., 2006). This empirical method is based on the neuronal structure and processing of the brain, which is capable of learning relationships from information (Jorgensen and Bendoricchio, 2001). The advantage over other modeling methods is that this method does not assume a prior structure of the data and is well suited for fitting nonlinear relationships (Batchelor et al., 2002).

At the regional scale of the Pampas, few empirical models have assessed the relation between soil properties and wheat yield. Lower wheat yields were detected in areas with drainage problems (Verón et al., 2004), and SAWSC in the upper 100 cm of the soil profile showed the strongest correlation with yield (Alvarez, 2009). Soil organic C explained more yield variability under semiarid conditions (Díaz-Zorita et al., 2002). Nevertheless, there are no studies that have analyzed the interaction among soil variables and the interaction of these soil properties with regional climate characteristics.

Our objective was to compare the performance of inductive productivity indices, the FAO index and the Kiniry–Pierce index, with the deductive approach for assessing regional soil productivity for wheat in the Pampas. The hypothesis of this study was that deductive productivity indices are more accurate than inductive indices because they are directly validated with yield data. Different empirical modeling techniques were compared for yield estimation for the deductive index development, ranging from simple blind guess methods to ANNs.

MATERIALS AND METHODS

Study Area

The Argentine Pampas (located between 28 and 40°S and between 68 and 57°W) is a vast plain with relief that is flat or slightly rolling. The natural vegetation consists of grasslands on which graminaceous plant species dominate. The mean annual temperature ranges from 14°C in the south to 23°C in the north, and the mean annual rainfall varies from 200 to 1200 mm from west to east. Rainfed crops are widespread in the humid and semiarid portions of the region (with annual rainfall >500 mm), on well-drained soils, mainly Mollisols, formed on loess-like materials (Hall et al., 1992; Alvarez and Lavado, 1998). At present, approximately 60% of the area is under agriculture, and the main crops are soybean, wheat, and corn. Because Pampean soils are very fertile and crops were commonly rotated with livestock production on leguminous pastures, fertilization has become a common practice only since approximately 1995 (Ministerio de Agricultura, Ganadería y Pesca, 2013) but at low rates (FAO, 2004). Wheat has been the most common crop during the last decades, but currently it is being replaced by soybean in many areas. Its growing cycle starts in June or July, depending on the sowing date, and it is harvested in December. The fallow period usually runs from April to June.

For this study, we used information from more than 150 counties in five provinces (Buenos Aires, Córdoba, Entre Ríos, La Pampa, and Santa Fe) in which wheat is an important constituent of rotations. Forty of these counties were removed from the analysis because hydromorphic soils predominated and the cultivated area was <30% according to seeded area information at the county level (Ministerio de Agricultura, Ganadería y Pesca, 2013). To reduce statistical noise due to large differences in county areas (of up to 30-fold), information was aggregated into 41 geographic units with an average area of 1 ± 0.5 Mha each (Fig. 1). We used the isohyet of 800 mm that divides the region to define the humid Pampas to the east and the semiarid (subhumid) Pampas to the west. On average, these geographic units aggregated information from three counties. The spatial aggregation was performed taking into account previously defined relief, type of landscape, and soil classes of Pampean subregions (Alvarez and Lavado, 1998). County-level information was aggregated up to the geographic unit level by applying weighted averages for each county, corrected for their corresponding areas.

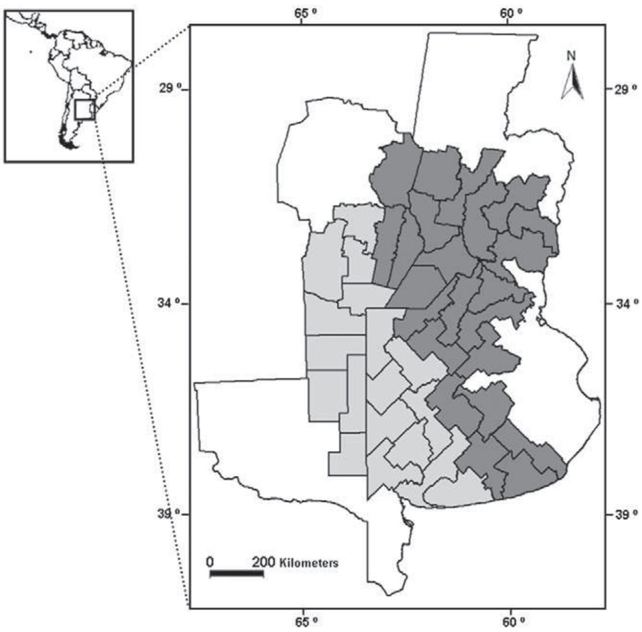

Fig. 1. Map of the Pampean Region of Argentina and its 41 geographical units. Darker geographical units located in the eastern portion correspond to the humid Pampas and lighter geographical units correspond to the semiarid Pampas in the western portion. The division is based on the annual rainfall isohyet of 800 mm.

Data Sources

Wheat yield data at the county scale were calculated from available records of annually harvested area and grain production from 1967 to 2006 (Ministerio de Agricultura, Ganadería y Pesca, 2013).

Climate records for the same period were obtained from the Servicio Meteorológico Nacional (www.smn.gov.ar) and the Instituto Nacional de Tecnología Agropecuaria (www.inta.gob.ar). Monthly temperature and rainfall information from approximately 80 meteorological stations, well distributed across the study area and the surroundings, were used. The number of meteorological stations varied from year to year. The average number of stations was 80 for the 40 yr of analysis. The year that had the lowest number of meteorological stations was 1967 with 74 stations, and in 2005 the number of stations was the greatest, with 86 stations. Averages at the county scale were interpolated using the inverse distance weighting method in the Spatial Analyst extension of ArcGIS 9.1 (ESRI). This method used information from meteorological stations in the county or very close to it for estimating an areal average (Zimmerman et al., 1999). We evaluated the method by performing point estimations for 30 stations spread across the Pampas. The software performed good estimations of temperature as indicated by the comparison between modeled and observed values ($R^2 > 0.90$, F test $P = 0.01$, ordinate and slope equal to 0 and 1, respectively, as indicated by a t-test with $P = 0.05$). Good estimations of monthly rainfall were achieved ($R^2 > 0.7$, F test $P = 0.01$, but intercept and slope were not equal to 0 and 1, respectively). Because the difference between estimated and observed values was small (5%), we still used the method for interpolation. The wheat growing cycle was divided into three periods for the analysis: fallow (April–June), vegetative growing phase (July–September), and flowering to maturing phase (October–November).

Potential evapotranspiration (PET) and k_c (a crop coefficient that accounts for differences in crop type, cultivar, and development stage that should be considered when assessing evapotranspiration) were calculated as previously described using estimated temperature and rainfall at the county level (Allen et al., 1998; Alvarez, 2009). Because no k_c values were available for the fallow period, we assumed the same value as for the sowing period (i.e., 0.5). The ratios between rainfall (R) and PET for the fallow and crop growing periods were calculated. The photothermal quotient was estimated for the critical period of 1 mo before crop anthesis, with variations related to the latitude of the geographic units. Incoming radiation was calculated by applying a locally developed method for atmospheric transmittance estimation at the county level (Alonso et al., 2002), and a base temperature of 4.5°C was used (Magrin et al., 1993).

Soil data from soil surveys of the provinces of Buenos Aires (Instituto Nacional de Tecnología Agropecuaria, 1989a), La Pampa (Instituto Nacional de Tecnología Agropecuaria, 1980), Córdoba (Instituto Nacional de Tecnología Agropecuaria, 2003), Santa Fe (Instituto Nacional de Tecnología Agropecuaria, 1981, 1989b), and Entre Ríos (Instituto Nacional de Tecnología Agropecuaria, 1984) were used. These surveys were performed, mainly, between 1960 and 1980. More than 1000 soil profile descriptions with their corresponding areas of influence were available. Soil variables taken into account were organic C (Nelson and Sommers, 1996), textural composition (Gee and Bauder, 1986), and pH (Thomas, 1996) from the soil surface to the 100-cm depth or the upper limit of the petrocalcic horizon. Because these variables were described at different depth intervals, depending on the genetic horizons, the variation in depth was modeled fitting different functions using Table Curve 2D (Systat Software) with good results ($R^2 > 0.90$). Using the fitted models for each soil variable and soil profile, the estimated values were assessed in layers of 25 cm up to the 1-m depth or the upper limit of the petrocalcic horizon. Because some surveys reported soil organic C and others reported organic matter, a correction factor of 1.72 was applied for transforming organic matter to organic C (Nelson and Sommers, 1996). Data at the profile level were transformed to the cartographic scale and afterward up to the county scale, taking into account the area of influence of each profile and cartographic unit (Alvarez and Lavado, 1998).

Soil organic C data published in these surveys were 40 yr old on average. A recent study published information at the county scale that resulted from a sampling performed during 2007 to 2008 (Berhongaray et al., 2013). We therefore used the soil survey data for the first years of the study and the recent data for the final years, linearly interpolating the C content for intermediate years. The SAWSC for the upper 1 m of the profile or the upper limit of the petrocalcic horizon was estimated using textural information and organic C content (Rawls et al., 1982). Bulk density was estimated by the Rawls (1983) method, expressed on a per hectare basis and corrected by a factor of 4% to adjust the demonstrated overestimation of this method for Pampean soil density (Berhongaray et al., 2013).

Inductive Soil Productivity Indices

The FAO methodology (Riquier et al., 1970) was used in the Pampas for rating soil productivity; this information is available in published soil surveys (Instituto Nacional de Tecnología Agropecuaria, 1981, 1984, 1989a, 1989b, 2003, 2013). Soil productivity was reported at the soil cartographic unit or county scale. We aggregated published data to our geographic unit scale taking into account the relative contribution of each value related to its area of influence. The range of the FAO index was from 0 to 100, but we adapted it to range from 0 to 1 to make comparison with other tested productivity indices possible.

Both versions of the index developed by Neill (1979) were combined into a single index, called the Kiniry–Pierce index, to take advantage of available information. Each of the root response functions described in this method represents the fractional sufficiency (0.0–1.0) for values of each soil variable. The product of all sufficiencies was considered as describing the fractional sufficiency of any soil layer for root growth. Electrical conductivity was not measured, and its sufficiency was assumed to be 1.0 according to Kiniry et al. (1983). Aeration and the sufficiency for aeration were also established at 1.0 because flooding areas were discarded from this analysis. All the other sufficiencies were calculated for layers of 25 cm and summed to a depth of 1 m or the upper limit of the petrocalcic horizon. Results at the county scale were aggregated up to the geographic unit level.

We estimated the profile root fractions that would exist under ideal soil conditions with plant-determined rooting depths using Horn's equation (Horn, 1971). We considered 100 cm as the optimal depth for wheat growth in the Pampas (Calviño and Sadras, 2002). The equation predicted profile fractional depletion with depth restrictions.

The sufficiency for SAWSC was estimated assuming that a value of 0.20 or larger indicated nonlimiting conditions (Kiniry et al., 1983).

The bulk density sufficiency function was estimated according to an adaptation proposed by Udawatta and Henderson (2003). Clay, silt, and sand contents were regressed, using the equations proposed in this work for estimations of growth-limiting, critical, and nonlimiting bulk densities for each soil layer. Values of 1 were assigned to bulk density values larger than the nonlimiting density, with a critical sufficiency of 0.83. A value of 0 was assigned to bulk densities greater than the growth-limiting threshold. Linear interpolation was applied for estimating intermediate values.

Sufficiency for pH per soil layer was calculated according to the equations of Pierce et al. (1983).

Relative sufficiencies of weather were calculated by a response surface model at the geographic unit scale using yearly climate information, with detrended yield as the dependent variable, according to Kiniry et al. (1983) (see below for a description of yield detrending). Independent variables included in the model were R/PET ratio for fallow, vegetative growth, and flowering periods and the photothermal quotient. Training and validation data sets (75 and 25%, respectively) were used for this weather sufficiency. Models were fitted with the training set and validated against the validation set. Values of all sufficiencies were multiplied for each soil layer, and layer values were summed. The highest index value estimated by these models received a value of 1, and the results were normalized with respect to this value.

A modification of this index was tested by adding a sufficiency factor for organic matter. This sufficiency was included as an additive factor to the estimated inductive productivity index because its effect was not proportional to yield. Not every increase in soil organic matter content results in increasing yield (Pierce et al., 1983; Wilson et al., 1991). We followed the method proposed by Wilson et al. (1991) for the inclusion of an organic matter sufficiency factor. The slope between average organic matter and yield was 0.23, indicating a 23% increase in yield for each 1% increase in organic matter; however, Wilson et al. (1991) showed that using a value near 0.25 as an additive factor to the Kiniry–Pierce index caused the level of organic matter to be too influential in the model. Therefore, we chose a slope of 0.15 as reasonable because other values for this sufficiency were tested without affecting the final result.

Deductive Soil Productivity Indices

Four yield-modeling techniques were tested: blind guess, polynomial regression, regression trees, and ANNs. The final data set generated had 1640 data (n = 41 geographic units × 40 yr). The blind guess method corresponds to yield estimations for each geographic unit based on information on yield averages (Alvarez, 2009). The estimated yield value does not account for any climate or soil information. The average yield for the first 30 yr of this study and the average yield of 1, 2, 3, 5, 7, or 10 yr chosen at random were used to estimate the yield by this method. The estimated yield resulting from these combinations of years was validated against the yield average of the last decade of the analyzed period for each geographic unit. The R^2 of these linear regressions was used as a statistical measure of the performance of the tested blind guess method. The resulting crop yield had a highly positive trend with time, largely resulting from improvements in technology, mainly the adoption of modern cultivars and the increased use of fertilizers; the yield increase with time was detrended using the yield equivalent of the year 2006 (Lobell and Field, 2007).

Response surface models have been frequently used in the evaluation of agronomic experiments, with expected positive linear effects and negative quadratic effects (Colwell, 1994). Before applying polynomial regression, the normality of the variables was tested by the Shapiro–Wilk method. As evidence of normality was not found, the data were transformed by applying logarithms, exponentials, arcsines, powers, and finally the Box–Cox method (Peltier et al., 1998). Because normality was not attained but the variables were close to a normal distribution, we were restrictive with the size of the hypothesis test (P = 0.01) because of asymptotic arguments (Amemiya, 1985). Linear, quadratic, and interaction terms were incorporated in the models only if they were significant at P = 0.01 and the whole model at P = 0.01 (F test). A forward stepwise method was used for predictor selection until the maximum R^2 was attained. Yield was modeled using as predictors time (year), rainfall, temperature, potential crop evapotranspiration, the R/PET ratio, photothermal quotient, soil depth, clay, sand, and silt contents, soil organic C, and SAWSC. Climate variables were tested separately for fallow, vegetative, and reproductive crop stages and for combinations of periods. The variance inflation factor was used to check the autocolinearity between independent variables (Neter et al., 1990). To assess the generalization ability of models, these were fitted using 75% of the data, randomly selected (training set), and validated against the independent remaining 25% (validation set). A hierarchical approach was also used to combine independent variables for calculation of other variables with the purpose of including the effects of the variables in the first level and allowing the simplification of the selected model (Schaap and Bouten, 1996). The regression of estimated vs. observed yield, slopes, and intercepts were compared using the t-test in the IRENE software (Fila et al., 2003).

The regression tree approach represents a nonparametric statistical method used to explain the variation of a single numerical response variable by one or more explanatory variables (Digby and Kempton, 1994). Normality of variables was not a requisite for this technique. A tree was constructed by repeatedly splitting the data by a simple rule for every single explanatory variable (Steinberg and Colla, 1995). At each split, the data were partitioned into two mutually exclusive groups, each of which was as homogeneous as possible and provided the best explanation of the response variable (McKenzie and Ryan, 1999). The splitting procedure was then applied to each new group separately. Splits minimized the sums of squares within groups. Complex structured trees, with too many ramifications, were discarded in our study to avoid overlearning (De'Ath and Frabricius, 2000), using the R^2 as the decision criterion for tree selection. The predictors tested to generate a regression tree model were the same as those used for the response surface regression fit. The whole data set was partitioned into 75 and 25% for training and validation, respectively, using the same sets as in the regression fitting. Regression trees were fitted using the Cubist 2.05 software (Rulequest).

Feed-forward ANNs were tested as modeling methods for yield prediction, fitting weights by the back-propagation algorithm (Kaul et al., 2005) The common network architecture of three layers was used: input, hidden, and output layers. Linear transfer functions were used from the input layer to the hidden layer and from the output layer to the network output, while a sigmoid function connected the hidden layer to the output layer (Lee et al., 2003). Simplification of the network architecture, scaling methods, learning rate, and epoch size were similar to those described by Alvarez (2009). We aimed for the maximum simplification of networks, reducing input variables and neurons in the hidden layer as much as possible without affecting the R^2. Sensitivity analysis was performed to weight the effect of different inputs on wheat yield (the output) by calculating a sensitivity ratio (Miao et al., 2006). Only predictors with a sensitivity ratio >1 were selected because lower values indicated no impact of the input variable on the network output. The same predictors used for the other modeling methods were tested as network inputs. To avoid overlearning, the data were partitioned into 50% for training, 25% for testing, and 25% for validation. Models were adjusted with the training set, and early stopping of weight fitting was achieved when the R^2 of the test set became lower than the R^2 of the training set (Kleinbaum and Kupper, 1979). The validation data set was the same as used for the regressions and regression trees adjustment. Neural networks were fitted using Statistica Neural Networks (version 2011, StatSoft).

Model performance was compared using the R^2 and RMSE (Kobayashi and Salam, 2000). Possible differences between R^2 values were tested by a specific test using Fisher′s Z transformation (Kleinbaum and Kupper, 1979). Yield-based productivity indices of the empirical models were calculated, assigning a value of 1 to the maximum modeled yield and relating all other values to this maximum. The selected empirical model for productivity index development was validated against the average observed yield per geographic unit of the independent data set. Information from field experiments located in both the humid and semiarid Pampas were used to validate the developed regional productivity at the field scale (Alvarez and Grigera, 2005; Bono et al., 2010).

RESULTS

A high variability was observed in the climate and soil variables and wheat yield (Table 1). Rainfall was the climate variable that showed the largest variability among geographic units. In the semiarid west, rainfall during the crop growing season averaged 264 mm, while in the eastern humid Pampas, it reached a mean of 352 mm. As the temperature increased from south to north, potential evapotranspiration increased as well, ranging from 230 to 345 mm. This temperature gradient determined a photothermal quotient gradient from south to north when related to incoming radiation. Soil properties also showed a wide range of variation. The SAWSC contrasted deeply between soils of the semiarid and humid Pampas, mainly because of textural differences. Rich organic C soils from the eastern geographic units had approximately threefold higher C levels than poor organic C soils from the west. In 12 geographic units, no petrocalcic layer was present within the upper 100 cm of the soil profile, but in some units of the southern portion of the Pampas, this was the main soil-related constraint for normal crop root development.

Regarding the temporal variability of soil variables for the 40 yr of this study, we assumed no significant changes in soil texture and soil depth between past and present. Accordingly, we used constant SAWSC values for each geographic unit. Conversely, soil organic C changes were observed when comparing available information of the first vs. the last decade. Some soils from the semiarid Pampas gained organic C, while some soils located in the humid Pampas lost organic C, according the a previous study by Berhongaray et al. (2013). These changes were accounted for when using linear interpolations for the intermediate decades.

Annual wheat yield varied considerably among geographic units and across the 40 growing seasons. Yield increased with time by 56% from the start of our time series to the final analyzed years, with an average yield gain for the whole region of 37 kg ha^{-1} yr^{-1}. In the humid eastern geographic units, yields were approximately 71% higher than in the semiarid western geographic units of the study area. Because of climatic variability, yield could vary up to sixfold in the same geographic unit.

Table 1. Variability of climate, soil variables, and wheat yield throughout the 41 geographical units in the Argentine Pampas from 1967 to 2006. Minimum and maximum values were calculated from the total data set for the 40 yr.

Variable†	Min.	Mean ± SD	Max.
Rainfall during fallow period, mm	21.3	161 ± 77.8	555
Rainfall during vegetative growth period, mm	11.6	113 ± 57.0	379
Rainfall during flowering period, mm	56.0	196 ± 71.2	486
Temperature during fallow period, °C	8.14	13.1 ± 1.55	17.3
Temperature during vegetative growth period, °C	7.7	11.3 ± 1.54	16.0
Temperature during flowering period, °C	12.3	17.9 ± 1.62	21.5
PET during fallow period, mm	73.9	130 ± 13.7	175
PET during vegetative growth period, mm	77.8	134 ± 14.4	191
PET during flowering period, mm	185	305 ± 29.3	412
R/PET ratio during fallow period	0.150	1.26 ± 0.677	4.340
R/PET ratio during vegetative growth period	0.071	0.887 ± 0.554	3.680
R/PET ratio during flowering period	0.136	0.653 ± 0.281	1.620
Photothermal quotient, MJ m^{-2}d^{-1} °C^{-1}‡	0.93	1.37 ± 0.22	2.05
Soil depth, cm§	73	92.5 ± 7.32	100
Clay, %§	10.2	26.5 ± 9.93	49.0
Sand, %§	6.26	42.3 ± 24.3	88.7
SAWSC, mm§	91.3	138 ± 33.1	192
Soil organic C, Mg ha^{-1}¶	31.0	64.5 ± 17.4	109
Wheat yield, kg ha^{-1}	448	1886 ± 695	4519

† PET, potential crop evapotranspiration; *R*, rainfall; SAWSC, soil available water storage capacity, 0–100-cm depth.

‡ During critical period of 1 mo before anthesis.

§ 0–100 cm or up to the upper layer of the petrocalcic horizon.

¶ 0–50-cm depth.

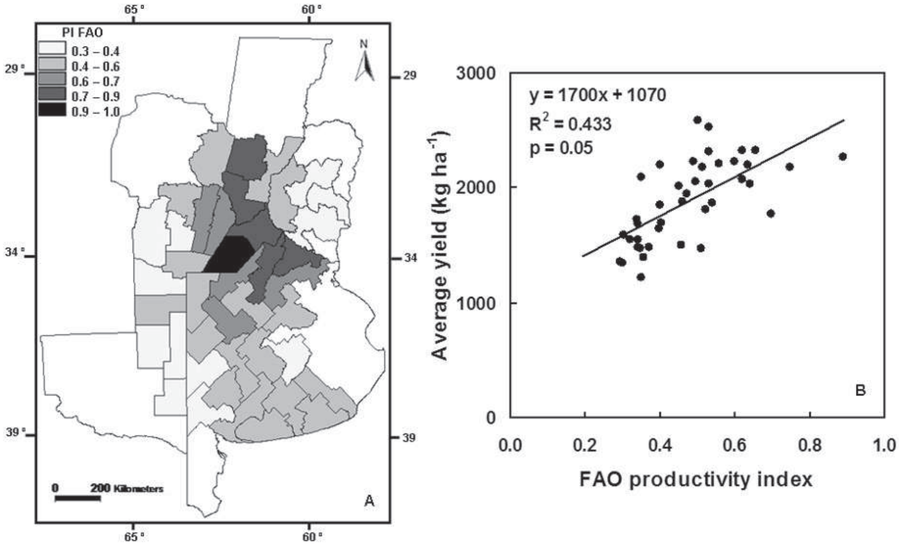

Fig. 2. (A) Spatial distribution of the inductive soil productivity index based on the FAO methodology for the geographical units of the Pampean Region (darker colors indicate higher productivity indices), and (B) regression between the FAO inductive soil productivity index and the average wheat yield during 1967 to 2006 for the geographical units. The original scale of the FAO index (0–100) was rescaled to 0 to 1 for comparison with other productivity indices.

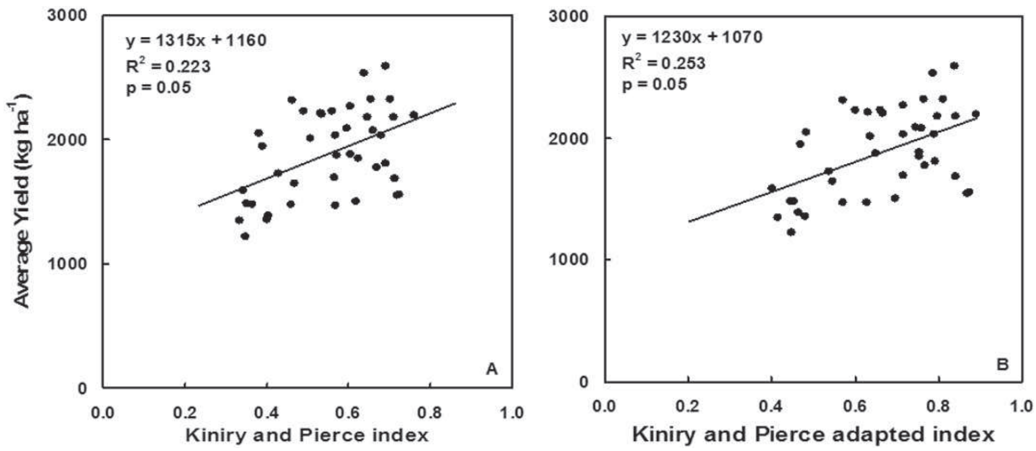

Fig. 3. (A) Regression of the inductive Kiniry-Pierce soil productivity index developed by Kiniry et al. (1983) and Pierce et al. (1983) and (B) the adapted inductive Kiniry–Pierce soil productivity index with the inclusion of a sufficiency for organic matter vs. the average wheat yield during 1967 to 2006 for the geographical units.

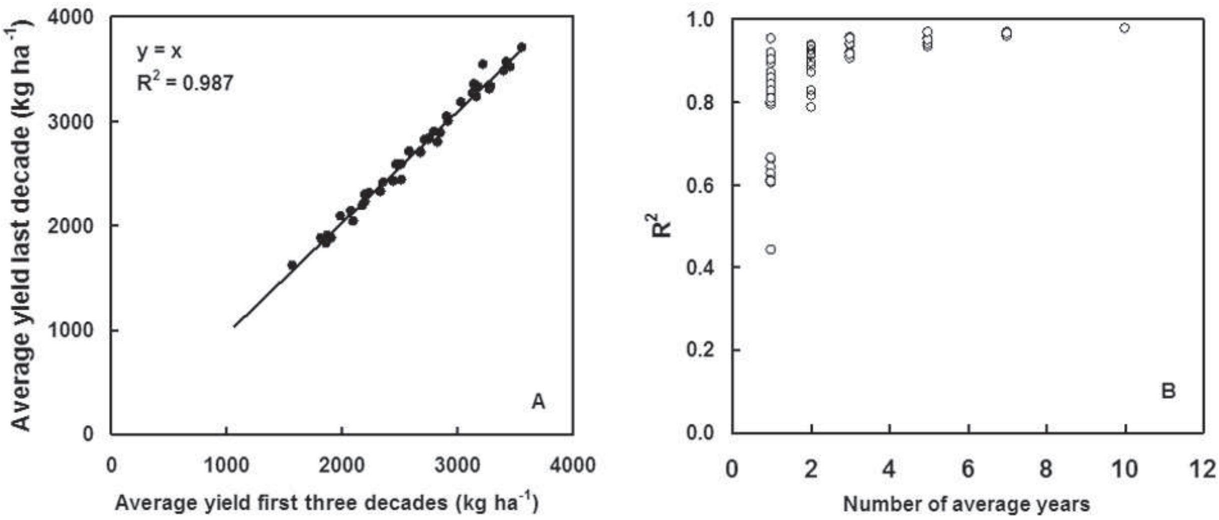

Fig. 4. (A) Regression of average yield of the first three decades (1967–1996) vs. average yield of the last decade (1997–2006) of analysis; and (B) the relationship between R^2, which represents the adjustment of the linear regression of average yield of the last decade (1997–2006) and averages generated from combinations of yield information of various numbers of years, vs. the number of years of yield data used (1, 2, 3, 5, 7, and 10 yr) to estimate the average yield of the last decade of analysis (1997–2006).

Climate and soil properties were poorly correlated ($R^2 < 0.15$), which discarded autocorrelation problems that could arise when modeling. Yield was moderately correlated with time ($R^2 = 0.37$), but poor relationships were also found with organic C, SAWSC, and rainfall during the fallow and crop growing cycle ($R^2 < 0.13$).

The spatial distribution of the FAO productivity index across the Pampas showed that values >0.7 were obtained in six geographic units of the humid Pampas (Fig. 2A). This inductive soil productivity approach estimated low rates for geographic units located in both the humid and semiarid Pampas. The FAO soil productivity index ranged from 0.33 up to 0.89 in Pampean soils, with a relatively low correlation with average observed yield data (Fig. 2B). Regarding the slope of this regression, an increase of 0.5 up to 0.7 in this index resulted in an average increase of 250 kg ha^{-1}. Similar results were obtained with the Kiniry–Pierce index, which ranged from 0.33 to 0.76, and its modification, which obtained slightly higher values (ranging 0.40–0.89). Both the original Kiniry–Pierce inductive method and its adaptation gave poor results when correlated with observed yield averages for the 41 geographic units (Fig. 3A and 3B). The dispersion of the linear correlation of both the original and the adapted Kiniry–Pierce indices with average yield was greater than that for the FAO productivity index. Regarding the slope of this index, an increase of 0.2 units (from 0.5 up to 0.7, for example) resulted in a yield increase of 200 kg ha^{-1}.

The blind guess method allowed a very good estimation of current wheat yield (average yield of the last decade) using past yield data. For example, using the average yield of the first three decades to estimate the yield of the last decade of our time series resulted in optimal results ($R^2 = 0.987$) (Fig. 4A). Also, when we used random averages of past yield data of 1, 2, 3, 5, 7, and 10 yr to estimate the average yield of the last 10 yr, at least 3 to 5 yr had to be averaged to obtain an acceptable performance (Fig. 4B).

All three empirical modeling methods tested in this study allowed fitting models with a good adjustment ($R^2 \geq 0.528$, RMSE $\leq$ 547 kg ha^{-1}) (Fig. 5). The generalization capacity of these models was good, as no significant differences were detected in R^2 between training and validation data sets. Slopes of observed vs. estimated yields were not different from 1, and ordinates were equal to 0 in all cases ($P = 0.05$). Regression trees and ANNs were more successful in predicting wheat yields, attaining higher R^2 values than the polynomial regression ($P = 0.05$). The generalization capacity of both regression trees and ANNs was not statistically different. The regression model included as predictors time, R/PET ratio during fallow and vegetative growth, the photothermal quotient, and SAWSC, with an $R^2 = 0.528$ and RMSE = 549 kg ha^{-1} (Fig. 5A). The model predicted that yield increased with time and with increasing R/PET ratio, SAWSC, and photothermal quotient. Visual inspection of the residuals showed that the frequency distribution was close to the normal distribution.

The regression tree model that best fitted yield data used the same predictors as polynomial regression but detected an additional effect of soil organic C ($R^2 = 0.577$, RMSE = 547 kg ha^{-1}) (Fig. 5B). The regression tree had seven rules, and the variable SAWSC defined the first split with a value of 92 mm. The model indicated that the highest yields were obtained after 1994 in soils with a SAWSC >117 mm and soil organic C contents >64 Mg ha^{-1}. The lowest yields were attained in soils with a SAWSC <92 mm. When each variable was analyzed for its contribution to the yield tree construction (attribute usage for splitting), we observed that time was used 100%, SAWSC 84%, organic C 79%, R/PET ratio for the fallow period 74%, R/PET ratio for the vegetative period 71%, and photothermal quotient 32% of the time.

The best ANN fitted had seven neurons in the hidden layer and included as inputs the same predictors as the regression tree but also R/PET ratio during the flowering to maturing period ($R^2 = 0.614$, RMSE = 548 kg ha^{-1}) (Fig. 5C). Time had a positive linear relation with yield. The R/PET ratio during the three periods, organic C, and SAWSC showed curvilinear relationships with yield. Because this empirical model based on an ANN approach achieved a slightly higher fit, although not significantly, than the regression tree method, it was chosen for the productivity index development. The maximum productivity index was attained for combinations of medium-to-high levels of both organic C and SAWSC.

Regarding the spatial distribution across the Pampas of this deductive productivity index based on the ANN approach, the highest productivities were observed in the humid eastern counties and the lowest in the counties that correspond to the western semiarid Pampas (Fig. 6A). Estimated productivity was >0.9 in approximately 17% of Pampean geographic units,

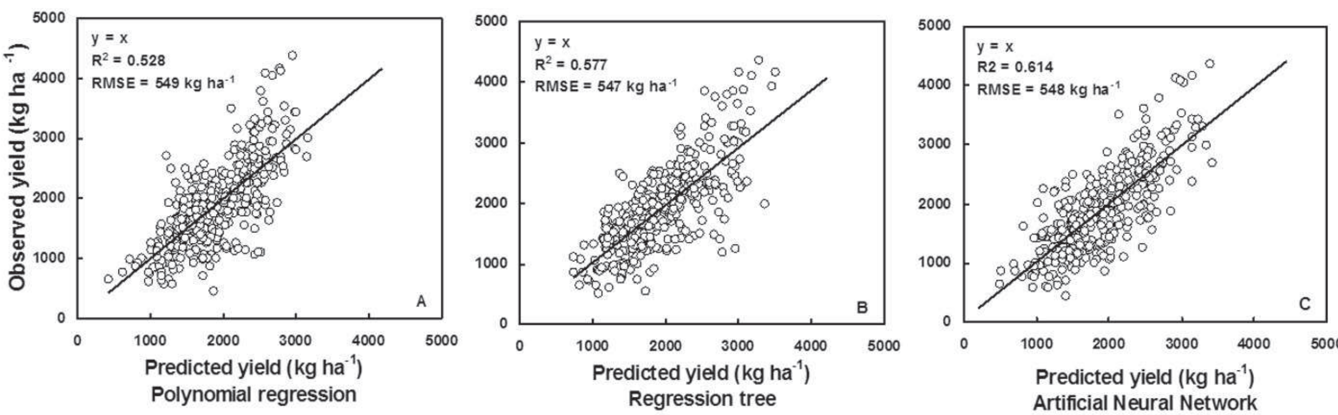

Fig. 5. Relationship between predicted and observed wheat yield values by (A) the polynomial regression method, (B) the regression tree method, and (C) the artificial neural network method. For the three empirical models, the model was validated against an independent validation set (25% of total data set).

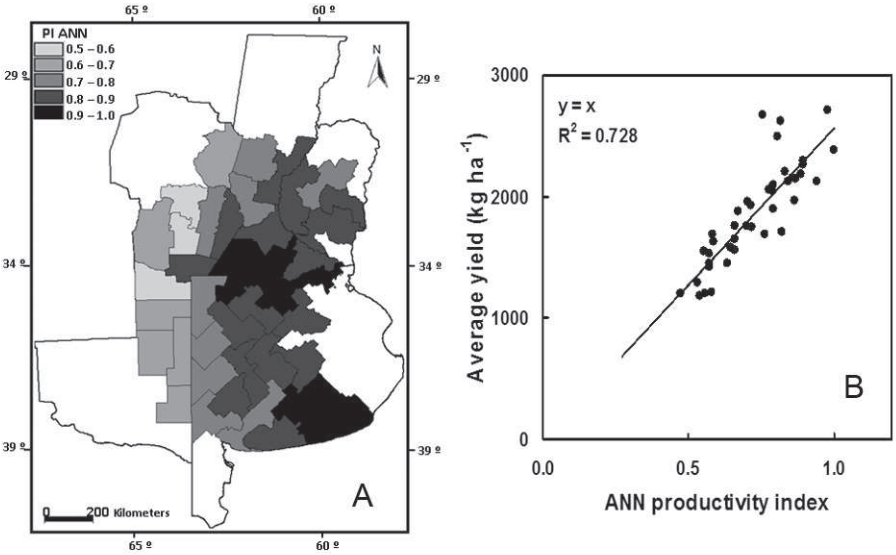

Fig. 6. (A) Spatial distribution of the deductive productivity index based on the artificial neural network (ANN) approach for the geographical units of the Pampean Region (darker colors indicate higher productivity indices), and (B) relationship between the predicted productivity index for wheat based on the ANN empirical approach and the average wheat yield during 1967 to 2006 for the geographical units.

another 60% had productivity indices between 0.7 and 0.9, and 23% of the Pampean region showed low productivity, <0.7. Average productivity indices for the 40 yr of analysis were calculated for each geographic unit and correlated with the average observed yield data from the independent data set with an adjustment of $R^2 = 0.728$ (Fig. 6B).

The neural network model was used for exploring the effects of environmental factors on wheat yield and predicting productivity for combinations of inputs. The model showed that SAWSC had a major impact on yield under low-rainfall scenarios (low average R/PET ratio), but this effect disappeared when rainfall was not limiting (Fig. 7A). It also showed that soil productivity depends on the positive interaction between SAWSC and soil organic C (Fig. 7B). The maximum productivity index was achieved in soils with 70 Mg ha^{-1} organic C and SAWSC of 140 mm up to a depth of 1 m as described by the model.

Finally, we tested whether the developed ANN productivity index based at the regional scale could be used with information from field experiments (Fig. 8). We observed no significant correlation between the estimated yield at the regional scale and the observed yield of field trials. The average yield achieved by the ANN model for these climate and soil inputs was 2000 kg ha^{-1}, with a minimum of 500 kg ha^{-1} and a maximum of 2900 kg ha^{-1}, while the average observed yield was 3400 kg ha^{-1}, ranging from 600 to 6200 kg ha^{-1}. The observed yield of the field trials almost doubled the yield estimated by the ANN model for those specific locations.

DISCUSSION

Our results showed that both inductive methods tested, the FAO productivity index commonly applied to the Pampas for soil evaluation and the Kiniry–Pierce index, were not suitable tools for regional Pampean soil ratings for wheat. The inclusion of a sufficiency factor for soil organic matter in the latter did not

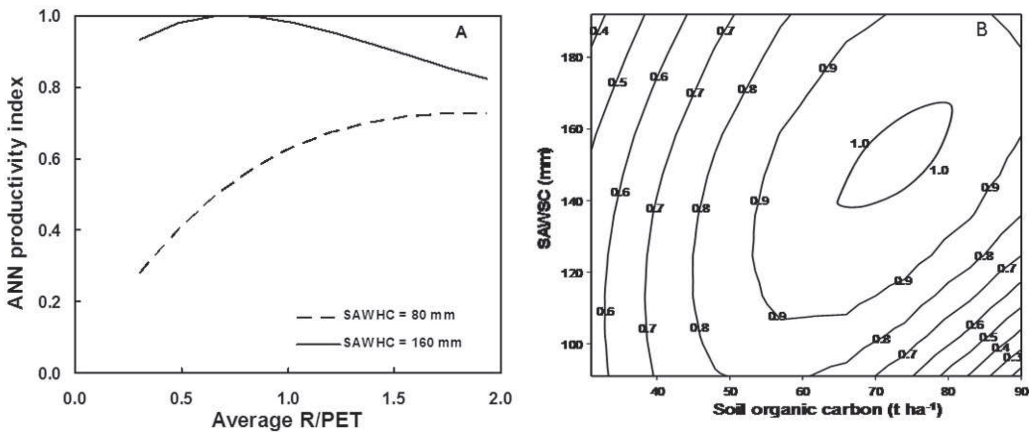

Fig. 7. (A) Productivity index based on the artificial neural network (ANN) approach for an average rainfall/potential evapotranspiration (R/PET) ratio during fallow, vegetative crop growth, and flowering for two soil available water storage capacity (SAWSC) scenarios: the soil scenario with 160-mm SAWSC up to 1-m depth represents a typical soil of the humid Pampas, while the soil scenario with 80-mm SAWSC represents a typical soil of the semiarid Pampas; and (B) interaction between soil organic C and SAWSC selected by the ANN model as soil inputs and the resulting wheat yield productivity index.

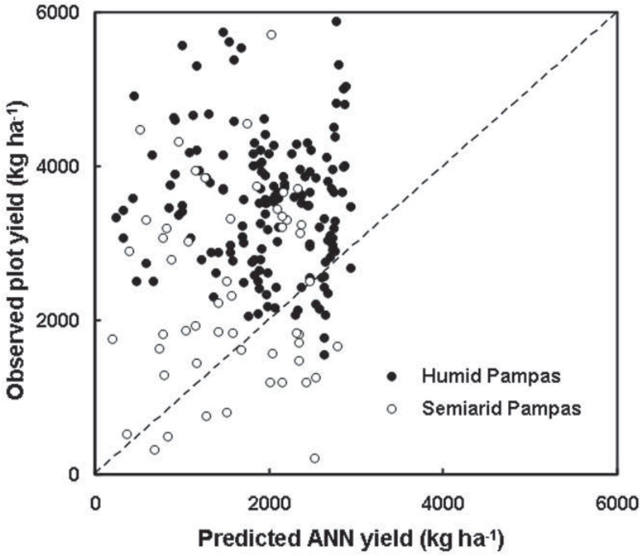

Fig. 8. Relationship between predicted wheat yield based on the artificial neural network (ANN) approach vs. the observed yield data from field experiments distributed along the Pampas and separated into humid and semiarid Pampas.

greatly improve the fit with yield values (Wilson et al., 1991). When correlating small-grain yield in Montana with this index, low relationships were also found ($R^2 < 0.35$) (Gerhart, 1989).

Yield data seem essential for adequate soil evaluation for wheat production in the Pampas. The simplest blind guess method was a good option when yield data were available for the soil to be rated. Long yield records were not needed; stable results were attained with averages of three to five growing seasons. In cases in which detailed climate and soil information is not available, this kind of yield-based model should be used.

Empirical modeling allowed good wheat yield prediction at the regional scale using easily obtainable information under a very large range of conditions. The fraction of yield variability not explained by the methods tested in our study (roughly 40%) may be accounted for by environmental factors not included in the ANN model. Agriculture has been a common regional practice for more than a century in the Pampas, but despite its economic and ecological importance, official agencies generate information with great uncertainties (Paruelo et al., 2004). Crop yield information was available at the county scale, while climate and soil data were available at smaller scales, which makes aggregation for modeling purposes difficult because information scales should be homogeneous.

In this study, we observed that when the modeling method was more sophisticated, more environmental effects (predictors in models) were detected and a higher proportion of yield variability could be explained. The selected neural network model allowed analysis of the interaction among predictors and the development of a soil productivity index. This index can be used for soil ratings for which wheat yield data are not available. A previous neural network model was developed in the Pampas for in-season wheat production forecasting that could not perform this job because it was fitted to a small data set ($n = 100$) that did not allow a deep analysis of interactions among soil and climate variables (Alvarez, 2009).

The effect of time on yield can be attributed mainly to two factors: wheat genetic improvement (Calderini and Slafer, 1998) and the adoption of fertilization (Ministerio de Agricultura, Ganadería y Pesca, 2013). In our data set, the fertilizer rate was closely correlated with time (data not presented). Consequently, we could not include this management variable as a predictor for modeling and instead used the variable time as a partial surrogate for this management improvement (Lobell et al., 2005).

Better fits were attained when the R/PET ratio was calculated separately during the fallow, vegetative, and crop flowering to maturity phases than if all three periods were combined into one or two periods (data not presented). This result may be attributed to the importance of soil water content at sowing on crop yield, which has already been quantified in on-farm experiments in the semiarid Pampas (Bono et al., 2010). The effect of soil water content during fallow on yield was included in the R/PET ratio for this period. The same occurred for the other crop cycle periods because the impact of water availability on growth depends on the crop growth stage (Brisson et al., 2001).

The SAWSC in Pampean soils depends mainly on soil texture and depth and only to a minimum extent (2%) on soil organic matter content (De Paepe and Alvarez, 2012). In the Pampas, as in other cultivated areas, a strong influence of SAWSC on crop yield has been observed. For example, in Australia, linear relationships have been observed between SAWSC and wheat yield up to a threshold of 65 mm (Lawes et al., 2009), yield regulation by SAWSC was described in France (Wassenaar et al., 1999), and the geographic patterns of SAWSC have been determined in the United States because of their importance on productivity (Kern, 1995). Determining SAWSC was necessary for the development of models that help understand the soil influence on yield under varying climate and soil conditions throughout the Pampas.

Organic matter has a positive influence on wheat yield; this might be a possible consequence of its function as a nutrient source. In the Pampas, it was demonstrated that at the time of wheat sowing, NO_3 levels were higher in C-rich soils (Alvarez et al., 2002) and as organic matter increased, the capacity of the soils to mineralize N during wheat growing cycle increased (Alvarez and Steinbach, 2011). As we observed at the regional scale for the Pampas, significant relationships of soil organic C and crop yield were established in other studies from other parts of the world (García-Paredes et al., 2000); however, this relation was not established in all regions (Jiang and Thelen, 2004), and this could be associated with soil C gradients and management conditions (low or high fertilizer consumption).

When analyzing the impact of different soil scenarios on wheat yield, as predicted by the ANN model output, SAWSC clearly determined higher productivity in areas with fine- to medium-textured soils. With this empirical model, we were able to analyze the interaction between soil variables for different Pampean subregions. When the humid and semiarid Pampas were analyzed separately (data not shown), the interaction between soil organic C and SAWSC remained positive; however, some differences were observed. In the semiarid area, both soil variables had a greater effect on wheat yield than in the more humid portion of the region. In the semiarid geographic units, when soil organic C increased from 40 to 70 Mg ha^{-1}, the productivity index increased from 0.3 to 0.9, whereas in the humid part this resulted in a smaller increase from 0.7 to 0.9 (data not shown). Similarly, when SAWSC increased from 100 to 140 mm in the semiarid Pampas, the index increased 0.5 units, while in the humid area this increase was only 0.2 units. The optimal SAWSC value in the semiarid portion appeared to be 170 mm up to 1-m depth, while in the humid portion this value appeared to be 130 mm.

Some unexpected results were observed in fine-textured soils, with some decreases in predicted productivity at high average R/PET ratios. This may be attributed to a confounding effect because in very humid areas (the eastern edge of the Pampas), in which fine-textured soils predominate, high rainfall scenarios determine disease problems in wheat (Annone, 2001) and possible temporarily flooding conditions. Confounding effects between environmental variables were one of the main problems of yield modeling and special care must be taken when interpreting model predictions (Bakker et al., 2005).

Upscaling the information on crop production at the site scale across greater areas is necessary to obtain estimates of crop production at aggregated regional levels (Olesen et al., 2000). Modeling yield estimations at the regional scale results in improved fits because outliers are averaged (Bakker et al., 2005). We tested the generalization ability of our data aggregation method by developing an ANN model for yield estimation using county information instead of geographic unit data; we attained similar results to those reported here (n = 4440, R^2 = 0.608, RMSE = 331 kg ha^{-1}). The climate and soil variables included in this county-scale model were the same as the model that upscaled information to the geographic unit level, with a similar impact on yield. Regarding regional aggregation, Easterling et al. (1998) noted that there seems to be a trade-off between crop estimations at the regional scale gained by aggregating environmental variables vs. loss of statistical properties of these variables. For our study, this implies that the regional aggregation of yield and environmental variables to the geographic unit level or county level preserved these variables sufficiently, leading to agreement between the estimated and observed yields. Nevertheless, this study showed that using data aggregated at the geographic unit level with an ANN approach allowed working not only with averages of soil information but also with soil information that was related to similar landscape properties.

CONCLUSION

The ANN approach for the development of a regional deductive productivity index allowed a better correlation with wheat yield in the Pampas than the two tested inductive approaches. These foreign inductive soil productivity indices should not be extrapolated to other regions or crops from which they were developed without yield validation. The ANN productivity index showed that regional wheat production was determined by the positive interaction of soil organic C and SAWSC. The regional productivity index model developed could not be validated with site-specific experiment information and was suited only to the regional scale. When only yield data are available, averages from 3 to 5 yr are adequate for regional soil ratings of wheat productivity. The principles of the ANN methodology used here as a tool for regional productivity estimation can be applied in other regions of the world and for different crops.

ACKNOWLEDGMENTS

This work was supported by grants from the University of Buenos Aires (W617 and G033), CONICET (PIP 02608) and FONCYT (PID-BID 37164- 49). We thank Prof. R. Ceulemans and his research group in Plant Ecology at the University of Antwerp for their helpful suggestions and critical reading of earlier versions of this manuscript.

REFERENCES

Allen, R.G., L.S. Pereira, D. Raes, and M. Smith. 1998. Crop evapotranspiration: Guidelines for computing crop requirements. Irrig. Drain. Pap. 56. FAO, Rome.

Allgood, F.P., and F. Gray. 1978. Utilization of soil characteristics in computing productivity ratings of Oklahoma soils. Soil Sci. 125:359–366. doi:10.1097/00010694-197806000-00004

Alonso, M., R. Rodriguez, S. Gomez, and R. Giagnoni. 2002. Un método para estimar la radiación global con la amplitud térmica y la precipitación diarias. Rev. Fac. Agron. Vet., Univ. Buenos Aires 22:51–56.

Alvarez, R. 2009. Predicting average regional yield and production of wheat in the Argentine Pampas by an artificial neural network approach. Eur. J. Agron. 30:70–77. doi:10.1016/j.eja.2008.07.005

Alvarez, R., C.R. Alvarez, and H. Steinbach. 2002. Association between soil organic matter and wheat yield in humid Pampa of Argentina. Commun. Soil Sci. Plant Anal. 33:749–757. doi:10.1081/CSS-120003063

Alvarez, R., and S. Grigera. 2005. Analysis of soil fertility and management effects on yields of wheat and corn in the Rolling Pampa of Argentina. J. Agron. Crop Sci. 191:321–329. doi:10.1111/j.1439-037X.2005.00143.x

Alvarez, R., and R.S. Lavado. 1998. Climate, organic matter and clay content relationships in the Pampa and Chaco soils, Argentina. Geoderma 83:127–141. doi:10.1016/S0016-7061(97)00141-9

Alvarez, R., and H.S. Steinbach. 2011. Modeling apparent nitrogen mineralization under field conditions using regressions and artificial neural networks. Agron. J. 103:1159–1168. doi:10.2134/agronj2010.0254

Amemiya, T. 1985. Advanced econometrics. Harvard Univ. Press, Cambridge, MA.

Annone, J.G. 2001. Criterios empleados para la toma de decisiones en el uso de fungicidas en trigo. Rev. Tecnol. Agropecu. 6:16–20.

Bakker, M.M., G. Govers, F. Ewert, M. Rounsevell, and R. Jones. 2005. Variability in regional wheat yields as a function of climate, soil and economic variables: Assessing the risk of confounding. Agric. Ecosyst. Environ. 110:195–209. doi:10.1016/j.agee.2005.04.016

Batchelor, W.D., B. Basso, and J.O. Paz. 2002. Examples of strategies to analyze spatial and temporal yield variability using crop models. Eur. J. Agron. 18:141–158. doi:10.1016/S1161-0301(02)00101-6

Berhongaray, G., R. Alvarez, J. De Paepe, C. Caride, and R. Cantet. 2013. Land use effects on soil carbon in the Argentine Pampas. Geoderma 192:97–110. doi:10.1016/j.geoderma.2012.07.016

Bono, A., J. De Paepe, and R. Alvarez. 2010. In-season wheat yield prediction in the semiarid Pampa of Argentina using artificial neural networks. In: A.J. Greco, editor, Progress in food science and technology. Vol. 1. Nova Sci. Publ., Hauppauge, NY. p. 133–150.

Brisson, N., E. Guevara, S. Meira, M. Maturano, and G. Coca. 2001. Response of five wheat cultivars to early drought in the Pampas. Agronomie 21:483–495. doi:10.1051/agro:2001139

Calderini, D.F., and G.A. Slafer. 1998. Changes in yield and yield stability in wheat during the 20th century. Field Crops Res. 57:335–347. doi:10.1016/S0378-4290(98)00080-X

Calviño, P., and V. Sadras. 2002. On-farm assessment of constraints to wheat yield in the south-eastern Pampas. Field Crops Res. 74:1–11. doi:10.1016/S0378-4290(01)00193-9

Colwell, J. 1994. Estimating fertilizer requirements. A quantitative approach. CAB Int., Wallingford, UK.

De'Ath, G., and K.E. Frabricius. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. Ecology 81:3178–3192. doi:10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2

De Paepe, J., and R. Alvarez. 2012. Estimacion de la capacidad de almacenamiento de agua util de suelos pampeanos con informacion textural. Paper presented at the XIX Congreso Latinoamericano de la Ciencia del Suelo y XXII Congreso Argentino de la Ciencia del Suelo, Mar del Plata, Argentina. 16–17 Apr. 2012.

Díaz-Zorita, M., G.A. Duarte, and J.H. Grove. 2002. A review of no-till systems and soil management for sustainable crop production in the subhumid and semiarid Pampas of Argentina. Soil Tillage Res. 65:1–18. doi:10.1016/S0167-1987(01)00274-4

Digby, P., and R. Kempton. 1994. Multivariate analysis of ecological communities. Chapman and Hall, London.

Easterling, W.E., A. Weiss, C.J. Hays, and L.O. Mearns. 1998. Spatial scales of climate information for simulating wheat and maize productivity: The case of the US Great Plains. Agric. For. Meteorol. 90:51–63. doi:10.1016/S0168-1923(97)00091-9

FAO. 2004. Fertilizer use by crop in Argentina. FAO, Rome.

Fila, G., G. Bellocchi, M. Acutis, and M. Donatelli. 2003. IRENE: A software to evaluate model performance. Eur. J. Agron. 18:369–372. doi:10.1016/S1161-0301(02)00129-6

García-Paredes, J.D., K.R. Olson, and J.M. Lang. 2000. Predicting corn and soybean productivity for Illinois soils. Agric. Syst. 64:151–170. doi:10.1016/S0308-521X(00)00020-2

Gee, G.W., and J.W. Bauder. 1986. Particle-size analysis. In: A. Klute, editor, Methods of soil analysis. Part 1. Physical and mineralogical methods. SSSA Book Ser. 5. SSSA, Madison WI. p. 383–412. doi:10.2136/sssabookser5.1.2ed.c15

Gerhart, K.E.S. 1989. Evaluating the performance of the soil productivity index (PI) model in Cascade County, Montana. M.S. thesis. Montana State Univ., Bozeman.

Hall, A.J., C. Rebella, C. Guersa, and J. Culot. 1992. Field-crop systems of the Pampas. In: C.J. Pearson, editor, Field crop ecosystems. Elsevier, Amsterdam.

Horn, F.W. 1971. The prediction of amounts and depth distribution of water in a well-drained soil. M.S. thesis. Univ. of Missouri, Columbia.

Huddleston, J.H. 1984. Development and use of soil productivity ratings in the United States. Geoderma 32:297–317. doi:10.1016/0016-7061(84)90009-0

Instituto Nacional de Tecnología Agropecuaria. 1980. Inventario integrado de los recursos naturales de la provincia de La Pampa. INTA, Buenos Aires.

Instituto Nacional de Tecnología Agropecuaria. 1981. Mapa de suelos de la provincia de Santa Fe, Parte I. NTA, Buenos Aires.

Instituto Nacional de Tecnología Agropecuaria. 1984. Suelos y erosión de la provincia de Entre Ríos. INTA-EEA, Entre Ríos, Paraná, Argentina.

Instituto Nacional de Tecnología Agropecuaria. 1989a. Mapa de suelos de la provincia de Buenos Aires. INTA, Buenos Aires.

Instituto Nacional de Tecnología Agropecuaria. 1989b. Mapa de suelos de la provincia de Santa Fe, Parte II. INTA, Buenos Aires.

Instituto Nacional de Tecnología Agropecuaria, 2003. Recursos naturales de la Provincia de Cordoba. INTA, Buenos Aires.

Instituto Nacional de Tecnología Agropecuaria. 2013.Suelos de la República Argentina. INTA, Buenos Aires. www.geointa.inta.gov.ar (accessed 2 Jan. 2013).

Jame, Y.W., and H.W. Cutforth. 1996. Crop growth models for decision support systems. Can. J. Plant Sci. 76:9–19. doi:10.4141/cjps96-003

Jiang, P., and K.D. Thelen. 2004. Effect of soil and topographic properties on crop yield in a north-central corn–soybean cropping system. Agron. J. 96:252–258. doi:10.2134/agronj2004.0252

Jorgensen, S., and G. Bendoricchio. 2001. Fundamentals of ecological modelling. 3rd ed. Elsevier, Amsterdam.

Kaul, M., R.L. Hill, and C. Walthall. 2005. Artificial neural networks for corn and soybean yield prediction. Agric. Syst. 85:1–18. doi:10.1016/j.agsy.2004.07.009

Kern, J.S. 1995. Geographic patterns of soil water-holding capacity in the contiguous United States. Soil Sci. Soc. Am. J. 59:1126–1133. doi:10.2136/sssaj1995.03615995005900040026x

Kim, K., B.L. Barham, and I. Coxhead. 2000. Recovering soil productivity attributes from experimental data: A statistical method and an application to soil productivity dynamics. Geoderma 96:239–259. doi:10.1016/S0016-7061(00)00020-3

Kiniry, L., C. Scrivener, and M. Keener. 1983. A soil productivity index based upon predicting water depletion on root growth. Res. Bull. 1051. Missouri Agric. Exp. Stn., Columbia.

Kleinbaum, D.G., and L.L. Kupper. 1979. Applied regression analysis and other multivariable methods. Duxbury Press, North Scituate, MA.

Kobayashi, K., and M.U. Salam. 2000. Comparing simulated and measured values using mean squared deviation and its components. Agron. J. 92:345–352.

Lawes, R.A., Y.M. Oliver, and M.J. Robertson. 2009. Integrating the effects of climate and plant available soil water holding capacity on wheat yield. Field Crops Res. 113:297–305. doi:10.1016/j.fcr.2009.06.008

Lee, J.H.W., Y. Huang, M. Dickman, and A.W. Jayawardena. 2003. Neural network modeling of coastal algal blooms. Ecol. Modell. 159:179–201. doi:10.1016/S0304-3800(02)00281-8

Lobell, D.B., and C. Field. 2007. Global scale climate–crop yield relationships and the impacts of recent warming. Environ. Res. Lett. 2:014002. doi:10.1088/1748-9326/2/1/014002

Lobell, D.B., I.J. Ortiz-Monasterio, G.P. Asner, R.L. Naylor, and W.P. Falcon. 2005. Combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape. Agron. J. 97:241–249.

Magrin, G., A.J. Hall, G. Baldi, and M. Grondona. 1993. Spatial and interannual variations in the photothermal quotient: Implications for the potential kernel number of wheat crops in Argentina. Agric. For. Meteorol. 67:29–41. doi:10.1016/0168-1923(93)90048-M

McKenzie, N.J., and P.J. Ryan. 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89:67–94. doi:10.1016/S0016-7061(98)00137-2

Miao, Y., D. Mulla, and P. Robert. 2006. Identifying important factors influencing corn yield and grain quality variability using artificial neural networks. Precis. Agric. 7:117–135. doi:10.1007/s11119-006-9004-y

Ministerio de Agricultura, Ganadería y Pesca. 2013. Producción agricola y forestal. Ministerio de Agricultura, Ganadería y Pesca, Buenos Aires. www.minagri.gob.ar (accessed 1 May 2013).

Neill, L. 1979. An evaluation of soil productivity based on root growth and water depletion. M.S. thesis. Univ. of Missouri, Columbia.

Nelson, D.W., and L.E. Sommers. 1996. Total carbon, organic carbon; and organic matter. In: D.L. Sparks, editor, Methods of soil analysis. Part 3. Chemical methods. SSSA Book Ser. 5. SSSA, Madison WI. p. 961–1010. doi:10.2136/sssabookser5.3.c34

Neter, J., W. Wasserman, and M. Kutner. 1990. Applied linear statistical models. 3rd rd. Irwin Publ., Burr Ridge, IL.

Olesen, J.E., P.K. Bøcher, and T. Jensen. 2000. Comparison of scales of climate and soil data for aggregating simulated yields of winter wheat in Denmark. Agric. Ecosyst. Environ. 82:213–228. doi:10.1016/S0167-8809(00)00227-9

Özesmi, S.L., C.O. Tan, and U. Özesmi. 2006. Methodological issues in building, training, and testing artificial neural networks in ecological applications. Ecol. Modell. 195:83–93. doi:10.1016/j.ecolmodel.2005.11.012

Paruelo, J., J.P. Guerschman, G. Baldi, and C. Di Bella. 2004. La estimacion de la superficie agricola: Antecedentes y una propuesta metodologica. (In Spanish, summary in English.) Interciencia 29:421–427.

Peltier, M.R., C.J. Wilcox, and D.C. Sharp. 1998. Application of the Box–Cox data transformation to animal science experiments. J. Anim. Sci. 76:847–849.

Pierce, F., W. Larson, R. Dowdy, and W. Graham. 1983. Productivity of soils: Assessing long-term changes due to erosion. J. Soil Water Conserv. 38:39–44.

Potgieter, A.B., G.L. Hammer, A. Doherty, and P. de Voil. 2005. A simple regional-scale model for forecasting sorghum yield across north-eastern Australia. Agric. For. Meteorol. 132:143–153. doi:10.1016/j.agrformet.2005.07.009

Rawls, W. 1983. Estimating soil bulk density from particle size analysis and organic matter content. Soil Sci. 135:123–125. doi:10.1097/00010694-198302000-00007

Rawls, W., D. Brakensiek, and K. Saxton. 1982. Estimation of soil water properties. Trans. ASAE 25:1316–1332.

Rijsberman, F.R., and M.G. Wolman. 1985. Effect of erosion on soil productivity: An international comparison. J. Soil Water Conserv. 40:349–354.

Riquier, J., D. Bramao, and J. Cornet. 1970. A new system of soil appraisal in terms of actual and potential productivity. Mimeo AGT: TESR/70/6. FAO, Rome.

Sarkar, P.K., O.W. Bidwell, and L.F. Marcus. 1966. Selection of characteristics for numerical classification of soils. Soil Sci. Soc. Am. J. 30:269–272. doi:10.2136/sssaj1966.03615995003000020035x

Satorre, E., and G. Slafer. 1999. Wheat production systems of the Pampas. In: E. Satorre and G. Slafer, editors, Wheat: Ecology and physiology of yield determination. Hayworth Press, Binghamton, NY. p. 333–348.

Schaap, M.G., and W. Bouten. 1996. Modeling water retention curves of sandy soils using neural networks. Water Resour. Res. 32:3033–3040. doi:10.1029/96WR02278

Sinclair, T.R., and N.G. Seligman. 1996. Crop modeling: From infancy to maturity. Agron. J. 88:698–704. doi:10.2134/agronj1996.00021962008800050004x

Steinberg, D., and D. Colla. 1995. CART: Tree-structured non-parametric data analysis. Salford Systems, San Diego.

Thomas, G.W. 1996. Soil pH and soil acidity. In: D.L. Sparks, editor, Methods of soil analysis. Part 3. Chemical methods. SSSA Book Ser. 5. SSSA, Madison WI. p. 475–490. doi:10.2136/sssabookser5.3.c16

Udawatta, R.P., and G.S. Henderson. 2003. Root distribution relationships to soil properties in Missouri oak stands: A productivity index approach. Soil Sci. Soc. Am. J. 67:1869–1878. doi:10.2136/sssaj2003.1869

Verón, S.R., J.M. Paruelo, and G.A. Slafer. 2004. Interannual variability of wheat yield in the Argentine Pampas during the 20th century. Agric. Ecosyst. Environ. 103:177–190. doi:10.1016/j.agee.2003.10.001

Wassenaar, T., P. Lagacherie, J.-P. Legros, and M.D.A. Rounsevell. 1999. Modeling wheat yield responses to soil and climate variability at the regional scale. Clim. Res. 11:209–220. doi:10.3354/cr011209

Wilson, J.P., S.P. Sandor, and G.A. Nielsen. 1991. Productivity index model modified to estimate variability of Montana small grain yields. Soil Sci. Soc. Am. J. 55:228–234. doi:10.2136/sssaj1991.03615995005500010039x

Yang, J., R. Hammer, A. Thompson, and R. Planchar. 2003. Predicting soybean yield in a dry and wet year using a soil productivity index. Plant Soil 250:175–182. doi:10.1023/A:1022801322245

Zimmerman, D., C. Pavlik, A. Ruggles, and M. Armstrong. 1999. An experimental comparison of ordinary and universal kriging and inverse distance weighting. Math. Geol. 31:375–390. doi:10.1023/A:1007586507433