Europ. J. Agronomy 30 (2009) 70–77

Contents lists available at ScienceDirect

European Journal of Agronomy

journal homepage: www.elsevier.com/locate/eja

Predicting average regional yield and production of wheat in the Argentine Pampas by an artificial neural network approach

R. Alvarez

Facultad de Agronomía-Universidad de Buenos Aires, Avenue San Martín 4453, (1417) Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 27 November 2007
Received in revised form 20 June 2008
Accepted 28 July 2008

Keywords:
Wheat
Yield estimation
Argentine Pampas

ABSTRACT

A regional analysis of the effects of soil and climate factors on wheat yield was performed in the Argentine Pampas in order to obtain models suitable for yield estimation and regional grain production prediction. Soil data from soil surveys and climate data from meteorological records were employed. Grain production information from statistics at county level was integrated at a geomorphological level. The Pampas was divided into 10 geographical units and data from 10 growing season were used (1995–2004). Surface regression and artificial neural networks (ANN) methodologies were tested for analyzing the data. Wheat yield was correlated to soil available water holding capacity (SAWHC) in the upper 100 cm of the profiles (r^2 = 0.39) and soil organic carbon (SOC) content (r^2 = 0.26). The climate factor with stronger effect on yield was the rainfall/crop potential evapotranspiration ratio (R/CPET) during the fallow and vegetative crop growing cycle periods summed (r^2 = 0.31). The photothermal quotient (PQ) during the pre-anthesis period had also a significant effect on yield (r^2 = 0.05). A surface regression response model was developed that account for 64% of spatial and interannual yield variance, but this model could not perform a better yield prediction than the blind guess technique. An ANN was fitted to the data that accounted for 76% of yield variability. Comparing predicted versus observed yield a lower RMSE (P = 0.05) was obtained using the ANN than using the regression or the blind guess methods. Regional production estimations performed by the ANN showed a good agreement with observed data with a RMSE equivalent to 7% of the whole surveyed area production. As variables used for the ANN development may be available around 40–60 days before wheat harvest, the methodology may be used for wheat production forecasting in the Pampas.

1. Introduction

The Argentine Pampas is a vast plain of around 50 Mha (Alvarez and Lavado, 1998) where nearly 50% of the area is devoted to agriculture, being wheat one of the main crops (Hall et al., 1992). The region is considered as one of the most suitable areas for grain crops production in the world (Satorre and Slafer, 1999).

The effects of soil properties and climate on wheat yield have been assessed in some areas of the Pampas at the farmers fields scale. In the southern portion (Southern Pampa), a humid sub region with soils of high organic matter content, water deficit during 30 days before to 10 days after the flowering period, and temperature during grain filling, accounted for over half of the variance in crop yield (Calviño and Sadras, 2002). In this Pampean subregion wheat yields are also higher in deep soils (100–120 cm free rotting depth) than in shallow ones (Sadras and Calviño, 2001). In the east portion of the Pampas (Inland Pampa and West Pampa), with semiarid climate and medium to low organic matter soils, wheat yield is correlated to soil organic matter (SOM) following a linear-plateau tendency (r^2 = 0.48) with a critical level at 72 t SOM ha^{-1} in the upper 20 cm of the profile (Díaz-Zorita et al., 1999). In the northern portion of the Pampas (Rolling Pampa), a humid sub region with soils of medium organic matter content and deep profiles, rainfall and nutrients availability accounted for 50–70% or yield variability (Alvarez and Grigera, 2005; Sain and Jauregui, 1993).

At the scale of the whole Pampean Region there are no studies relating both soil properties and climate to wheat yield, because available soil data came form surveys at the series level and no integration has been performed to county or geomorphological levels, which allow relating those data to existing statistical yield information. It had only been detected that wheat yield is lower in areas with drainage problems (Veron et al., 2004). Conversely, climate effects on wheat had been assessed in the past. Using results from field experiments widespread along the Pampas, under water and nutrients non-limiting scenarios, the photothermal quotient (PQ = ratio between incident radiation to temperature during the critical period of one month prior to anthesis) accounted for nearly 50% of interannual wheat yield variability (Magrin et al.,

E-mail address: ralvarez@agro.uba.ar.

1161-0301/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.eja.2008.07.005

1993). Combining county statistical yield data and sparse information on harvest index, a regional analysis of climate factors effects on the above ground net primary productivity of the crop showed that rainfall and temperature accounted for 63% of the variance (Veron et al., 2002). Also using this county information a model that explains 34% of wheat yield variability was developed with PQ and rainfall as independent variables (Veron et al., 2004).

As there is no models for yield prediction developed in the Pampas, wheat production is forecasting before harvest by estimation of surface seeded to the crop, using reports from local informers, and average yield of different areas. In some cases, in-season yield estimated in the field is used for adjusting average predictions.

Artificial neural networks (ANN) had become a popular technique in biological sciences due to their predictive quality and because there are simplest than process based models (Joergensen and Bendoricchio, 2001; Özesmi et al., 2006). They are adaptive analytical methodologies based on neuronal structures and processing of the brain capable of learning relationships in patters of information (Joergensen and Bendoricchio, 2001). ANN had the advantage over other empirical modeling techniques that do not assume an a priory structure for the data, are well suited for fitting non-linear relationships and complex interactions, and can expose hidden relationships among input variables (Batchelor et al., 1997). As other empirical models they cannot extrapolate outside the range of data inputs.

Typically an ANN is structured in three neuronal layers: an input layer in which numbers of neurons correspond to the number of input variables, a hidden layer with a complexity determined empirically during ANN development, and an output layer with a neuron for each output variable (Fig. 1). Information flows from the input layer to the output layer through the hidden layer and the learning process consists in adjusting the weights associated to the transfer functions between neurons comparing ANN outputs with observed data by an iterative procedure (Joergensen and Bendoricchio, 2001). This learning process is performed usually by the back propagation algorithm that fits the weights from the output layer through the input layer (Kaul et al., 2005). The most common transfer function used between the hidden layer and the output layer is the sigmoid, and the lineal function is generally used to pass information from the input layer to the hidden layer (Kaul et al., 2005). Agronomic examples of ANN uses are as variable as environmental correlation (Park and Vlek, 2002), perdition of soil organic carbon content (Somaratgne et al., 2005), generate fertilizer recommendations (Broner and Comstock, 1997), estimation of soil hydraulic properties (Nemes et al., 2003), prediction of crop development (Elizondo et al., 1994), epidemic severity evaluation (Batchelor et al., 1997), and yield prediction (Kaul et al., 2005).

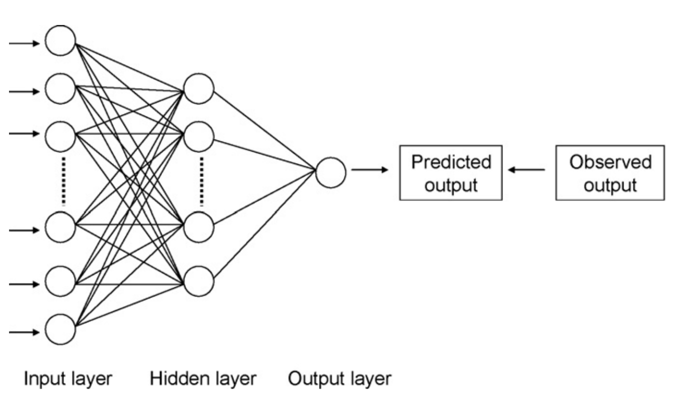

Fig. 1. Representation of a feed-forward artificial neuronal network showing layers and connections.

My objective was to analyze at the regional scale the effects of soil properties and climate on wheat yield in the Pampas in order to generate models suitable for accounting spatial and interannual yield variability. These models may be used for understanding which are the main factors controlling crop yield, for the development of productivity indexes, and for grain production forecasting. Artificial neural networks were used as tools for yield predictions.

2. Materials and methods

The Pampas plain runs from 28 to 40°S in Argentina. The relief is flat or slightly rolling and its natural vegetation consist of grasslands in which graminaceous vegetation species are dominant. Mean annual rainfall ranged from 200 mm in the west to 1200 mm in the east and mean annual temperature ranged from 14 °C in the south to 23 °C in the north. Agriculture is performed in the semiarid and humid portions of the region on well drained soils, mainly Mollisols formed on loess like materials, and areas with hydromorphic soils are devoted to pastures (Hall et al., 1992). Wheat (*Triticum aestivum*), corn (*Zea mays*) and soybean (*Glicine max*) are the main crops, being wheat widespread all over the region. Around 6 Mha are sown to wheat annually (SAGYP, 2004). The fallow period falls usually between April and June and the crop growing cycle from July to end of November, with some variability between Pampean subregions.

Soil data were obtained from soil surveys of the provinces of La Pampa (INTA, 1980), Buenos Aires (INTA, 1989) and Santa Fe (INTA, 1981, 1983). In these surveys, typical profiles and the area they occupied were reported. The surface of the surveyed area was divided into 10 geographic units (Fig. 2) according to geomorphological and soil classification considerations previously defined (INTA, 1980, 1989) and taking into account that rainfall and temperature throughout each unit were homogenously. On the basis of the soil profiles characteristics described in soil surveys and their corresponding area, the weighed average values of soil organic carbon (SOC), clay, silt, and sand were calculated as previously described (Alvarez and Lavado, 1998) for different soil layers: 0–20, 20–50 and 50–100 cm. Each value obtained was the mean SOC, clay, silt,

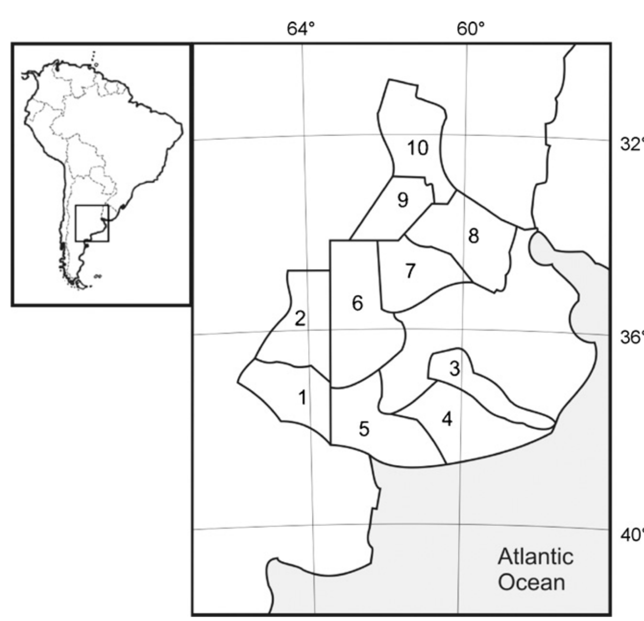

Fig. 2. Map of the Pampean Region and the geographic units studied.

or sand content of the geographic unit for each depth, and includes the mean of all the soils within the unit weighed by the area that occupied. Soil bulk density was estimated from texture and organic matter contents (Rawls, 1983) and SOC contents to 50 cm depth were expressed on a surface basis. Soil available water holding capacity (SAWHC) was estimated to 100 cm depth by the method of Rawls et al. (1982). In the south and west portions of the Pampas many soils presented a strong petrocalcic layer within the upper 100 cm of the profile that completely restricts roots grow. This was taken into account in the integration of data when averaging SAWHC of the geographic units considering for these soils, calculating SAWHC only within the free rotting depth. Total surveyed area integrated in this study rounded 26 Mha and include around 60% of surface destined to wheat in the Pampas. Yield information from other argentine provinces was not used because of the lack of adequate soil surveys. Neither data from the Flat Pampa, a subregion of the Buenos Aires Province characterized by very low slopes and hidromorphic soils, was integrated in this study because agriculture is restricted to small areas on summit positions and averaging soil properties to the geomorphic level would give biased results for cropped soils.

Climatic records were obtained from unpublished data of the Servicio Meteorológico Nacional available upon request. Monthly precipitation and mean monthly temperature were calculated from 1995 to 2004. Potential evapotranspiration was estimated by a modification of the Penman formula (Linacre, 1977). Locally adjusted k_c coefficients (Doorenbos and Pruitt, 1977; Totis and Perez, 1994) were applied to estimate wheat potential evapotranspiration. For the fallow period as no k_c coefficients were available it was assumed a value equivalent to that corresponding to the sowing period (0.5). The ratio rainfall/crop potential evapotranspiration (R/CPET) was calculated for the fallow period and for different periods during the crop growing cycle. For estimation of incoming solar radiation, a locally developed modification of the Hunt et al. (1998) method was employed for estimation of atmosphere transmittance which allows a closer agreement between esteemed radiation versus radiometric measurements in the Pampean Region (Alonso et al., 2002). Solar radiation at the top of the atmosphere was calculated using algorithms included in RadEst 3.00 (Donatelli et al., 2003). The PQ was calculated for the critical period of one month before anthesis using esteemed incoming radiation and mean daily temperature above a base temperature of 4.5 °C (Magrin et al., 1993). Anthesis dates varied with latitude in the Pampas from 30 September in the north to 10 November in the south. Anthesis dates at different latitudes were taken from experiments published in Magrin et al. (1993) and for intermediate latitudes estimations were performed using unpublished data (F. Menéndez, personal communication).

Yield data were calculated from unpublished records of the Secretaría de Agricultura, Ganadería y Pesca of Argentina of annual harvested surface and grain production at county level for the period 1995–2004. Yearly data of seeded surface with wheat were also available. Integration of data for the geographical units was performed as the weighed averages of the county yield averages affected by the corresponding surfaces.

A blind guess methodology was tested for prediction of yield of the geographical units previously defined. Average yield for each unit was calculated for the period 1995–2004 and annual observed yield correlated with these averages taken as predictions. Regression techniques were also tested for yield forecasting. In a fist step, relationships between yield and soil or climate variables were tested with linear and quadratic simple regression using the r^2 as a decision criterion. In a second step, a polynomial surface response model was developed of the form:

$$\text{Yield} = a_0 + a_1 v_1 - a_2 v_1^2 + a_3 v_2 - a_4 v_2^2 + a_5 v_1 v_2 + \cdots + a_{n-2} v_x - a_{n-1} v_x^2 + a_n v_x v_{x-1}$$

where a_0–a_n are the regression coefficients and v_1–v_x are the independent variables.

The model incorporates linear and quadratic terms for assessing linear and curvilinear effects of independent variables on the dependent variable and interaction terms between independent variables. It has been extensively used in agronomic experiment evaluation with positive expected linear effects and negative quadratic effects (Colwell, 1994). A combination of forward, backward and stepwise regression adjustments were used in order to obtain the simplest model with the highest r^2. Terms were maintained in the final model only when they were significant at $P = 0.05$ and the whole regression at $P = 0.01$ by the F test. Autocolinearity of independent variables were checked by means of the VIF value (Neter et al., 1990). A hierarchical approach was implemented combining variables for calculating others that include the effects of the variables in the fist level but allowed the simplification of models. (Schaap et al., 1998). Ten-fold cross validation was used for assessing the ability of the best regression model obtained of generalization to other possible data sets.

A feed-forward back propagation ANN was them tested for yield prediction. This kind of ANN, known as multilayer preceptors, had shown to be well suited for yield prediction at the plot scale and other agronomic uses when managing sets of data of similar size to that available in this study (Kaul et al., 2005; Starrett et al., 1997). Lineal transfer functions were used from the input layer to the hidden layer and from the output layer to the network output, meanwhile a sigmoid function (Lee et al., 2003) connected the hidden layer to the output layer. Input variables were scaled by the minimax procedure between 0 and 1 to make variation ranges uniform and data suitable for the sigmoid function (Park and Vlek, 2002). Network outputs were de-scaled to original units. The ANN was developed by a supervised learning procedure using the back propagation algorithm for weights fitting (Rogers and Dowla, 1994). A hierarchical approach was implemented for model simplification during the selection of input variables, preferring those that resulted from the integration of variables used in their construction, which effects resumed (Park and Vlek, 2002). The stepwise methodology was applied for inputs selection during ANN testing (Gevrey et al., 2003). The learning rate controls the size of weight change made by the back propagation algorithm during each iteration (Kaul et al., 2005). A larger learning rate may lead to faster convergence but may lead to a local minimum (Lee et al., 2003). Consequently, a low learning rate of 0.1 was used during ANN development. The epoch size represents the number of epochs (iterations) for which the algorithm will run. On each epoch, the entire training set is fed through the network, and used to adjust the network weights (Somaratgne et al., 2005). Around 50 epochs are adequate for convergence in some situations (Schaap and Bouten, 1996; Schaap et al., 1998). An epoch size of 100 was used here.

As the number of neuron in the hidden layer increase, the model fits better to the training data but the problem of possible overlearning (overfitting) increase too (Özesmi et al., 2006). Consequently, a balance between prediction ability of the ANN and complexity must be reach. Maximum initial number of neurons in the hidden layer was set by methods describe by Somaratgne et al. (2005) and neurons were deleted one at a time till model simplification reduced its ability to fit the data using the r^2 as decision criterion. Cross-validation is recommended to avoid overlearning (Özesmi et al., 2006), with early stopping of weights adjustment, when deviation from the verification set becomes higher than from the training

Table 1
Range of variability of soil and climate variables and wheat yield

	Clay (%)[a]	Silt (%)[a]	Sand (%)[a]	Organic carbon (%)[b]	Temperature (°C)[c]	Radiation (MJ m^{-2} d^{-1})[d]	Rainfall			Yield (kg ha^{-1})
							Fallow period (mm)	Vegetative period (mm)	Reproductive period (mm)	
Mean	23.8	34.6	41.6	2.07	13.5	19.0	184	117	211	2500
Minimum	9.70	21.1	5.90	1.17	10.3	6.70	64.0	20.0	30.0	947
Maximum	33.6	60.6	69.2	3.42	17.3	24.1	492	259	389	4130

[a] 0–100 cm depth or down to petrocalcic horizon.
[b] 0–50 cm depth.
[c] During crop growing cycle.
[d] During critical period of one month before anthesis.

Table 2
Correlation coefficients between independent variables

	Available water holding capacity	Organic carbon	Fototermal quotient	Rainfall fallow period	Rainfall vegetative period
Organic carbon	0.127				
Fototermal quotient	−0.234	0.289			
Rainfall fallow period	0.163	0.254	0.009		
Rainfall vegetative period	−0.104	0.536	0.002	0.200	
Rainfall reproductive period	0.112	0.194	−0.247	0.313	0.445

$R > 0.195$; 0.254; and 0.321; significant at $P = 0.05$; 0.01 and 0.001, respectively.

set (Park and Vlek, 2002). Data were randomly partitioned in 70% training: 30% verification and iteration stopped when the r^2 of the verification set tended to be lower when comparing to the r^2 of the training set. To test the generalization capacity of the models developed a modification of the procedure outlined by Schaap and Bouten (1996) was applied. Data were partitioned 10 times in 70:30 for training and verification, respectively, and best models generated with the first 70:30 partition run against the remain 70:30 data groups. Comparing r^2 between groups shown if model were able to predict yield independently of the partition of data and thus may generalize.

Wheat production estimations were performed using yield predictions generated by the three methodologies tested and an estimation of harvested surface. This later estimation was obtained regressing harvested surface with seeded surface. Slopes and intercepts of predicted versus observed yield and grain production regressions were compared by the *t* test using IRENE (Fila et al., 2003). Root mean square error (RMSE) (Kobayashi and Salam, 2000) was calculated for each estimation methodology and significant differences between RMSE tested by an *F* test (Xiong and Meullenet, 2006).

3. Results

A broad range of variability was observed in the soil and climate properties of the Pampas Region that leads to a 4-fold difference in wheat yield throughout regions and years (Table 1). Soil texture in the 0–50 cm layer varied from sandy loam to silty clay loam. Five geographical units had no impedance constraints within the upper 100 cm of the profile, meanwhile in the other five, average depths to petrocalcic horizon ranged from 77 to 88 cm. As the consequence of the combination of texture and free rooting depth, SAWHC ranged from 79 to 187 mm. Soil fertility, evaluated thought organic carbon content, was also very different between units (41–126 t C ha^{-1} in the 0–50 cm soil layer). Variability of climatic conditions was even greater, with a 4-fold difference of incoming solar radiation during the crop critical period. This produced, when related to temperature, a PQ range of 1.09–2.22 MJ m^{-2} d^{-1} $°C^{-1}$. Rainfall was the environmental variable with the greater variability. A 5-fold range of rainfall during the fallow and whole crop growing periods summed occurred, which produced that R/CPET ranged from 0.30 to 2.0.

Relationships between environmental variables were low. Soil organic carbon was not significantly correlated with SAWHC, meanwhile positive correlations were observed between SOC and rainfall (Table 2). Soil available water holding capacity was mainly determined by the clay + silt contents of the soils (SAWHC (mm) = 31 + 0.012clay + silt (t ha^{-1}), $r^2 = 0.98$, $P = 0.01$). Positive significant associations were also observed between rainfall during the fallow and crop growing periods, and a negative correlation existed between PQ and rainfall during the reproductive period.

Wheat yield was significantly correlated with some environmental variables, increasing along time with an average gain of 52 kg ha^{-1} y^{-1} in the study area (Table 3). A curvilinear relationship was observed between yield and SAWHC, reaching maximum yield values in soils that can store around 150 mm of available water in the upper 100 cm of the profile. Texture appeared as the main soil factor controlling wheat yield though its effect on soil water properties. Clay + silt mass in the top 100 cm depth accounted for 37% of yield variability ($P = 0.01$) using the quadratic model. Yield was also significantly correlated with soil organic carbon content, increasing from low carbon levels to around 90 t C ha^{-1} and thereof stabilizing. As the PQ increased, wheat yield increased too but with a low correlation coefficient. A rise in PQ of 1 MJ m^{-2} d^{-1} $°C^{-1}$ determined an average yield increase of 640 kg ha^{-1}. Rainfall during the fallow and the crop vegetative periods were significantly correlated with yield, but no significant association was detected between yield and rain-

Table 3
Significance of regression terms from regressions between wheat yield and some independent variables, and corresponding determination coefficients

Independent variable	Lineal term	Quadratic term	Determination coefficient (R^2)
Year	0.05	ns	0.05
Available water holding capacity	0.001	0.001	0.39
Organic carbon	0.01	0.05	0.26
Fototermal quotinet	0.05	ns	0.04
Rainfall fallow period	0.001	0.001	0.24
Rainfall vegetative period	0.001	ns	0.11
Rainfall reproductive period	ns	ns	ns

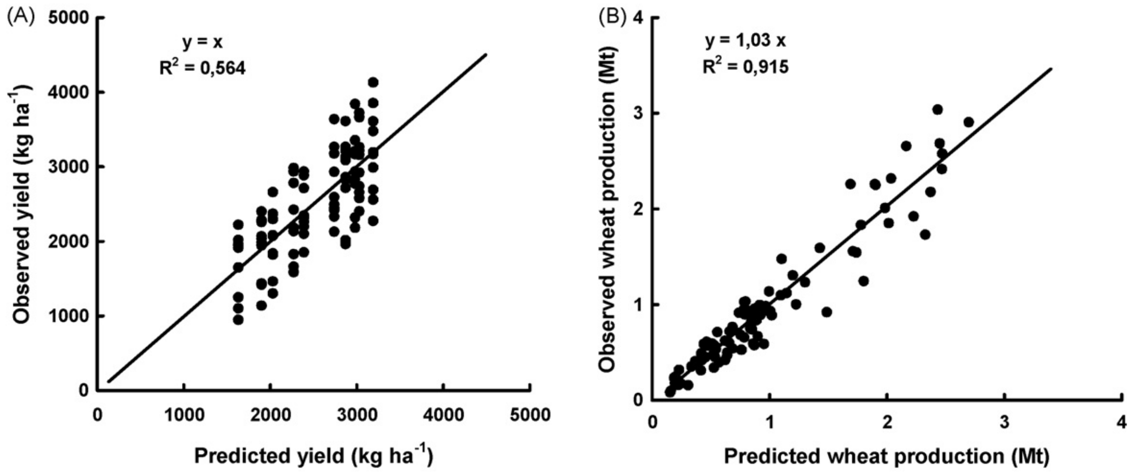

Fig. 3. Relationships between observed and predicted wheat yield (A) and Pampean grain production (B) using the blind guess estimation methodology.

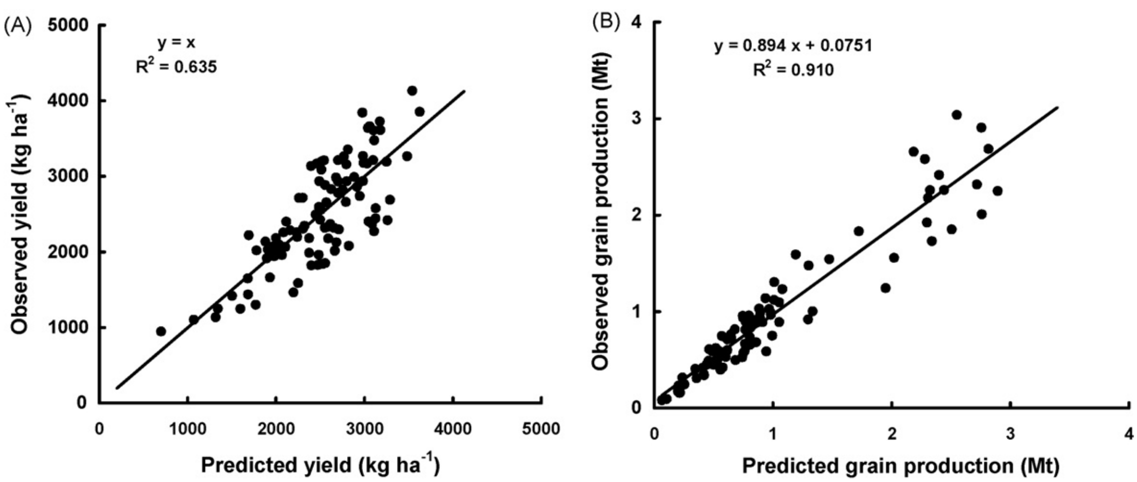

Fig. 4. Relationships between observed and predicted wheat yield (A) and Pampean grain production (B) using lineal surface regression as estimation methodology.

fall during the reproductive stage. When regressing yield against R/CPET of fallow and vegetative periods summed a quadratic model accounted for 31% of yield variability. This correlation coefficient could not be improved including the reproductive stage.

Estimation of wheat yield by the blind guess methodology accounted for 56% of spatial yield variability (Fig. 3A). By definition this methodology could not account for interannual variability. Regression between observed versus predicted yield had intercept non-different from 0 and slope equal to 1 ($P = 0.05$). Despite this medium determination coefficient for yield perdition, forecasting grain production by multiplying predicted yield and estimated harvested surface was very much better (Fig. 3B). This was the consequence of the deep impact of harvested surface on production and of the close agreement between seeded surface and harvested surface (harvested surface = 0.981 × seeded surface, $r^2 = 0.998$, $P = 0.001$). Harvested surface was in average 2% lower than seeded surface and could be estimated with precision some months before crop harvest. Grain production forecasting by the blind guess method was around 3% lower that the observed production across all regions and years. Integrating grain production for the whole surveyed area, RMSE was equivalent to 10.6% of annual average production.

A surface regression response model could be fitted to yield data that accounted for around 64% of the variance (Fig. 4A). The model included year of harvest, SAWHC, R/CPET and PQ as independent variables, and the regression of observed against estimated values showed intercept equal to 0 and slope of 1 ($P = 0.05$). Year of harvest and PQ showed positive effects on wheat yield, meanwhile SAWHC and R/CPET presented linear positive effects and curvilinear negative terms. The average determination coefficient of a 10-fold cross validation was 0.53 indicating that the generalization ability of the regression method was not high. Estimating grain production of the surveyed region as the product of yield predicted by the surface regression and estimated harvested surface gave similar results that the blind guess methodology, with a RMSE equivalent to 9.2% of whole area production (Fig. 4B). The regression methodology could not improved yield and grain production predictions when compared to the blind guess strategy (Table 4).

Table 4
Root mean square errors from different estimation methodologies of yield and production of wheat

Methodology	Yield (kg ha^{-1})	Production (Mt)
Blind guess	450 a	1020000 a
Regression	411 a	881000 a
Neural network	333 b	700000 b

Values followed by the same letters (a and b) in a column are not different at $P = 0.05$.

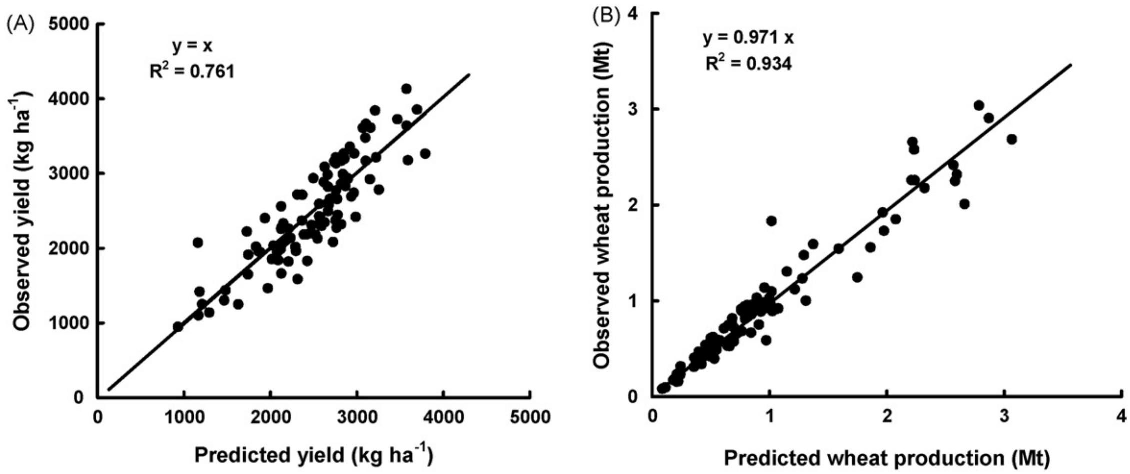

Fig. 5. Relationships between observed and predicted wheat yield (A) and Pampean grain production (B) using an artificial neural network as estimation methodology.

The ANN that best fitted to yield data used as network inputs year of harvest, SAWHC, SOC, R/CPET and PQ. It was structured with five neurons in the hidden layer and accounted for 76% of yield variance (Fig. 5A). Regression of observed versus predicted yield had intercept non-different from 0 and slope equal to 1 ($P = 0.05$). When data were partitioned 10 times 70:30 for training and verification, respectively, and the network run, all determinations coefficients from the validation data sets ranged from 0.76 to 0.80 showing a good generalization ability of the ANN method. Grain production could be well predicted by the ANN with an average sub estimation of 3% across regions and years (Fig. 5B). This methodology allowed better predictions of yield and grain production than the blind guess and the regression methods. The RMSE of the ANN approach was significant lower that those of the other strategies (Table 4).

4. Discussion

In the study area of this work annual yield gain averaged 4% per year during the 1995–2004 period. This increase may be attributed both to genetic improvement (Calderini et al., 1995) and better management practices (Satorre and Slafer, 1999).

Soil available water holding capacity had a strong impact on wheat yield in the Pampas when averaging data at a macro regional level. In other parts of the World studies performed at different scales showed significant effects on crops yield of SAWHC or some related soil properties. Analyzing sub-field yield variability, plant available water storage capacity of soils regulated wheat productivity in Southern Australia (Wong and Asseng, 2006), meanwhile soil texture was highly correlated to soybean yield in Mississippi soils (Cox et al., 2003). Collecting data from field experiments or production fields, significant relationships had been found between texture and cotton yield in Central Greece (Kalivas and Kollias, 2001), or free rooting depth and different crop yields in Spain (De la Rosa et al., 1981). Productivity index generated for specific soil types, using texture and rooting depth among other properties, may explain around 50% of corn and soybean yield variance in Illinois soils and are useful for average yield estimation at county level when climate inter annual variability is not taken into account (García-Paredes et al., 2000). Consequently, the determination of the capacity of soils to store available water appeared to be necessary when developing yield prediction models under a variety of climate, soil conditions and scales of analysis.

The relationship founded between SOC and wheat yield seems to be based on the impact of organic matter as a source of nutrients. On-farm local studies showed that nitrate nitrogen levels at wheat sowing are higher in organic matter rich soils (Alvarez et al., 2002), and mineralization during the crop growing cycle also increase in Pampean soils of high SOC content (González Montaner et al., 1997). In the semiarid portion of the Pampas soils present a wide spectrum of SOC contents, textures and free rotting depths. Field experiments performed in this area showed that SOC is correlated to wheat yield, independently of soil texture or depth (Bono and Alvarez, 2006). Conflicting results had been obtained in studies of the effects of SOC on crops yield worldwide. In some cases, significant relationships could be established between SOC and yield (Catching et al., 2002; García-Paredes et al., 2000), meanwhile in others not significant association was detected between both variables (Alvarez and Grigera, 2005; Jiang and Thelen, 2004). As a consequence, the inclusion of this soil property in models developed for predicting yield seems to be useful only in some situations; especially in cases where the range of the variable is very broad, whit some data in the very low SOC levels that restricts crop yield (Díaz-Zorita et al., 1999).

In the present study, a better adjustment was obtained when correlating wheat yield with the R/CPET during the fallow and vegetative growing period summed than when including the reproductive period too. This result may be attributed partially to the importance of soil water content at sowing on crop yield which has been yet quantified in the semiarid portion of the Pampas executing on-farm experiments (Bono and Alvarez, 2006). Soil water content at sowing is taken into account indirectly in this study including the fallow period in the water index R/CPET. Water deficits during the vegetative stage affect wheat yield in the Pampas (Brisson et al., 2001; González Montaner et al., 1997) but also around the critical flowering period (Calviño and Sadras, 2002). This later expected effect was not detected by the ANN model. The use of the water index R/CPET, which integrated variables related to water availability to the crop at different stages, allowed a better explanation of wheat yield variance that the use of the simple variables in the construction of the ANN model (results not presented). The hierarchical approach used here resulted in a simple model with good predictive capacity.

Using both soil productivity rates and climate variables for yield prediction, ANN had shown to be better tools than regression methods when analyzing corn and soybean yield data generated in field trials (Kaul et al., 2005). Integration of data at regional scales, as performed in this study, allows improving fits averaging outliers, with higher improvement as surface increase (Bakker et al., 2005). Confounding effects, generated by autocolineality between independent variables is a potential problem in this kind of studies, which may be overwhelming by experimentation, fixing all conditions except the one is tested (Bakker et al., 2005). Correlation between independent variables was generally low in the Pampas and only variables not significantly correlated were included in the ANN model, so confounding effects may be discarded. As all variables used in the ANN model are available 40–60 days before wheat harvest, in-season yield and production predictions are possible. Different methodologies for in-season prediction of crop yield had been tested in other agricultural regions as the use of the NDVI for wheat (Freeman et al., 2003) or the application of agro-climatic models for sorghum (Potgieter et al., 2005), but this techniques are not available in the Pampean Region at present.

The ANN approach allowed a better prediction of wheat yield and production than other methodologies when applied at a regional scale. These results may be considered as a first step in the developing of methods suitable for yield prediction for the whole Pampean Region and the methodology may be applied in other cropping areas of the World and for different crops.

Acknowledgement

This study was granted by the University of Buenos Aires, Proyect UBACyT G004.

References

Alonso, M.R., Rodriguez, R.O., Gomez, S.G., Giagnoni, R.E., 2002. Un método para estimar la radiación global con la amplitud térmica y la precipitación diarias. Rev. Fac. Agron. UBA 22, 51–56.
Alvarez, R., Grigera, S., 2005. Analysis of soil fertility and fertilizer effects on wheat and corn yield in the Rolling Pampa of Argentina. J. Agron. Crop Sci. 191, 321–329.
Alvarez, R., Lavado, R.S., 1998. Climate, organic matter and clay content relationships in the Pampa and Chaco soils, Argentina. Geoderma 83, 127–141.
Alvarez, R., Alvarez, C.R., Steinbach, H.S., Salas, J., Grigera, S., 2002. Materia orgánica y fertilidad de los suelos en la Pampa Ondulada. Informaciones Agronómicas (INPOFOS) 14, 11–14.
Bakker, M.M., Govers, G., Ewert, F., Rounsewell, M., Jones, R., 2005. Variability in regional wheat yield as a function of climate, soil and economic variables: assessing the risk of confounding. Agric. Ecosyst. Environ. 110, 195–209.
Batchelor, W.D., Yang, X.B., Tschanz, A.T., 1997. Development of a neural network for soybean rust epidemics. Trans. ASAE 40, 247–252.
Bono, A., Alvarez, R., 2006. Rendimiento de trigo en la Región Semiárida y Subhumeda Pampeana: un modelo predictivo de la respuesta a la fertilización nitrogenada. In: XX Congreso Argentino de la Ciencia del Suelo, Proceedings on CD, p. 5.
Brisson, N., Guevara, E., Meira, S., Maturano, M., Coca, G., 2001. Response of five wheat cultivars to early drought in the Pampas. Agronomie 21, 483–495.
Broner, I., Comstock, C.R., 1997. Combining expert systems and neural networks for learning site-specific conditions. Comp. Elec. Agric. 19, 37–53.
Calderini, D.F., Dreccer, M.F., Slafer, G.A., 1995. Genetic improvement in wheat yield and associated traits. A re-examination of previous results and the latest trends. Plant Breeding 114, 108–112.
Calviño, P., Sadras, V., 2002. On-farm assessment of constraints to wheat yield in the southeastern Pampas. Field Crop Res. 74, 1–11.
Catching, W.E., Hawkins, K., Sparrow, L.A., McCorkell, B.E., Rowley, W., 2002. Crop yields and soil properties on eroded slopes of red ferrosols in north-west Tasmania. Aust. J. Soil Res. 40, 625–642.
Colwell, J.D., 1994. Estimating Fertilizer Requirements. A Quantitative Approach. CAB International, UK, p. 259.
Cox, M.S., Gerard, P.D., Wardlaw, M.C., Abshire, M.J., 2003. Variability of selected soil properties and their relationship with soybean yield. Soil Sci. Soc. Am. J. 67, 1296–1302.
De la Rosa, D., Cardona, F., Almorza, J., 1981. Crop yield predictions based on properties of soil in Sevilla, Spain. Geoderma 25, 267–274.
Díaz-Zorita, M., Buschiazzo, D.E., Peinemann, N., 1999. Soil organic matter and wheat peoductivity in the Semiarid Argentine Pampas. Agron. J. 91, 276–279.
Donatelli, M., Bellocchi, G., Fontana, F., 2003. RasEst3.00: software to estimate daily radiation data from commonly meteorological variables. Eur. J. Agron. 18, 363–367.
Doorenbos, J., Pruitt, W.O., 1977. Crop water requirements. FAO. Irrigation and Drainage Paper No. 24. Rome, Italy, p. 193.
Elizondo, D.A., McClendon, R.W., Hoogenboom, G., 1994. Neural network models for predicting flowering and physiological maturity of soybean. Trans. ASAE 37, 981–988.
Fila, G., Bellocchi, G., Acutis, M., Donatelli, M., 2003. IRENE: a software to evaluate model performance. Eur. J. Agron. 18, 369–372.
Freeman, K.W., Raun, W.R., Jonson, G.V., Mullen, R.W., Stone, M.L., Solie, J.B., 2003. Late-season prediction of wheat yield and grain protein. Commun. Soil Sci. Plant Anal. 34, 1837–1852.
García-Paredes, J.D., Olson, K.R., Lang, J.M., 2000. Predicting corn and soybean productivity for Illinois soils. Agric. Syst. 64, 151–170.
Gevrey, M., Dimopoulos, I., Lek, S., 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. Mod. 160, 249–264.
González Montaner, J.H., Maddonni, G.A., DiNapoli, M.R., 1997. Modeling grain yield and grain yield response to nitrogen in spring wheat crops in the Argentinean Pampa. Field Crop Res. 51, 241–252.
Hall, A.J., Rebella, C.M., Ghersa, C.M., Culot, J.P., 1992. Field crop systems of the Pampas. In: Pearson, C.J. (Ed.), Field Crop Ecosystems of the World, vol. 18. Elsevier, Amsterdam, pp. 413–450.
Hunt, L.A., Kuchar, L., Swanton, C.J., 1998. Estimation of solar radiation for use in crop modeling. Agric. Forest Meteorol. 91, 293–300.
INTA, MEPLP, FALP, 1980. Inventario de los recursos naturales de la Provincia de la Pampa, p. 493.
INTA, MAGPSF, 1981. Mapa de suelos de la Provincia de Santa Fe. Parte I, p. 245.
INTA, MAGPSF, 1983. Mapa de suelos de la Provincia de Santa Fe. Parte II, p. 216.
INTA, 1989. Mapa de suelos de la Provincia de Buenos Aires, p. 525.
Jiang, P., Thelen, K.D., 2004. Effect of soil and topographic properties on crop yield in a north-central corn–soybean cropping system. Agron. J. 96, 252–258.
Joergensen, S.E., Bendoricchio, G., 2001. Fundamentals of Ecological Modelling, third ed. Elsevier, Oxford, UK, p. 530.
Kalivas, D.P., Kollias, V.J., 2001. Effects of soil, climate and cultivation techniques on cotton yield in Central Greece, using different statistical methods. Agronomie 21, 73–89.
Kaul, M., Hill, R.L., Walthall, C., 2005. Artificial neural networks for corn and soybean yield prediction. Agric. Syst. 85, 1–18.
Kobayashi, K., Salam, M.U., 2000. Comparing simulated and measured values using mean square deviation and its components. Agron. J. 92, 345–352.
Lee, J.H.W., Huang, Y., Dickman, M., Jayawardena, A.W., 2003. Neural network modeling of coastal algal blooms. Ecol. Mod. 159, 179–201.
Linacre, E.T., 1977. A simple formula for estimating evapotranspiration rates in various climates, using temperature data alone. Agric. Meteorol. 18, 409–424.
Magrin, G.O., Hall, A.J., Baldy, C., Grondona, M.O., 1993. Spatial and interannual variations in the phototermal quotient: implications for the potential kernel number of wheat crops in Argentina. Agric. Forest Meteorol. 67, 29–41.
Nemes, A., Schaap, M.G., Wösten, J.H.M., 2003. Functional evaluation of pedotransfer functions derived from different scales of data collection. Soil Sci. Soc. Am. J. 67, 1093–1102.
Neter, J., Wasserman, W., Kutner, M.H., 1990. In: Irwin Inc. (Ed.), Applied Linear Statistical Models, IL, USA, p. 1172.
Özesmi, S.L., Tan, C.O., Özesmi, U., 2006. Methodological issues in building, training, and testing artificial neural networks in ecological applications. Ecol. Mod. 195, 83–93.
Park, S.J., Vlek, P.L.G., 2002. Environmental correlation of three-dimensional soil spatial variability: a comparison of three adaptive techniques. Geoderma 109, 117–140.
Potgieter, A.B., Hammer, G.L., Doherty, A., Voil, P., 2005. A simple regional-scale model for forecasting sorghum yield across north-eastern Australia. Agric. Forest Meteorol. 132, 143–153.
Rawls, W.J., 1983. Estimating soil bulk density from particle size analysis and organic matter content. Soil Sci. 135, 123–125.
Rawls, W.J., Brakensiek, D.L., Saxton, K.E., 1982. Estimation of soil water properties. Trans. ASAE 25, 1316–1328.
Rogers, L.L., Dowla, F.U., 1994. Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling. Water Res. 30, 457–481.
Sadras, V.O., Calviño, P.O., 2001. Quantification of grain response to soil depth in soybean, maize, sunflower, and wheat. Agron. J. 93, 577–583.
SAGYP, 2004. Estadísticas de producción agrícola. http://siiap.sagyp.mecon.ar/.
Sain, G.E., Jauregui, M.A., 1993. Deriving fertilizer recommendation with a flexible functional form. Agron. J. 85, 934–937.
Satorre, E.H., Slafer, G.A., 1999. Wheat production systems of the Pampas. In: Satorre, E.M., Slafer, G.A. (Eds.), Wheat Ecology and Physiology of Yield Determination. The Haworth Press Inc., New York, pp. 333–348.
Schaap, M.G., Bouten, W., 1996. Modeling water retention curves of sandy soils using neural networks. Water Res. 32, 3033–3040.
Schaap, M.G., Leij, F.J., van Genuchten, M.Th., 1998. Neural networks analysis for hierarchical prediction of soil hydraulic properties. Soil Sci. Soc. Am. J. 62, 847–855.
Somaratgne, S., Seneviratne, G., Coomaraswamy, U., 2005. Prediction of soil organic carbon across different land-use patterns: a neural network approach. Soil Sci. Soc. Am. J. 69, 1580–1589.
Starrett, S.K., Starrett, S.K., Adams, G.L., 1997. Using artificial neural networks and regression to predict percentage of applied nitrogen leached under turfgrass. Commun. Soil Sci. Plant Anal. 28, 497–507.
Totis, L., Perez, O., 1994. Relaciones entre el consumo de agua máximo de la secuencia de cultivo trigo/soja y la evapotranspiración potencial para el cálculo de la dosis de riego. INTA Pergamino-Carpeta de Producción Vegetal 12, 1–4.

Veron, S.V., Paruelo, J.M., Sala, O.E., Lauenroth, W.K., 2002. Environmental controls of primary production in agricultural systems of the Argentine Pampas. Ecosystem 5, 625–635.
Veron, S.V., Paruelo, J.M., Slafer, G.A., 2004. Interannual variability of wheat yield in the Argentine Pampas during the 20th century. Agric. Ecosyst. Environ. 103, 177–190.
Wong, M.T.F., Asseng, S., 2006. Determining the causes of spatial and temporal variability of wheat yields at sub-field scale using a new method of upscalling a crop model. Plant Soil 283, 203–215.
Xiong, R., Meullenet, J.F., 2006. A PLS dummy variable approach to assess the impact of jar attributes on linking. Food Qual. Preferen. 17, 188–198.