SJSS
SPANISH JOURNAL OF SOIL SCIENCE

Application of artificial neural networks to estimate soil organic carbon in a high-organic-matter Mollisol

Aplicación de redes neuronales artificiales para estimar el carbono orgánico del suelo en un Mollisol con elevado contenido de materia orgánica

Aplicação de redes neuronais artificiais para estimar o carbono orgânico num Molisolo com elevado teor de matéria orgânica

AUTHORS

Moreno R.[1]

Irigoyen A. I.[1,2]

Monterubbianesi M. G.[1]

Studdert G. A.[@,1]
studdert.guillermo@inta.gob.ar

[@] Corresponding Author

[1] Facultad de Ciencias Agrarias, Universidad Nacional de Mar del Plata, Unidad Integrada Balcarce. Ruta Nac. 226 km 73,5 – CC 276. 7620 Balcarce, Prov. Buenos Aires, Argentina.

[2] Associate Researcher of Comisión de Investigaciones Científicas. Prov. de Buenos Aires, Argentina.

Received: 14.08.2017 | **Revised:** 25.10.2017 | **Accepted:** 25.10.2017

ABSTRACT

Soil organic carbon (SOC) has a key role in the global carbon (C) cycle. The complex relationships among the components of C cycle make the modelling of SOC variation difficult. Artificial neural networks (ANN) are models capable to determine interrelationships based on information. The objective was to develop and evaluate models based on the ANN technique to estimate the SOC in Mollisols of the Southeastern of Buenos Aires Province, Argentina (SEBA). Data from three long term experiments were used. Management and meteorological variables were selected as input. Management information included numerical variables (initial SOC (SOCI); number of years from the beginning of the experiment (Year), proportion of soybean in the crop sequence; (Prop soybean); crop yields (Yield), proportion of cropping in the crop rotation (Prop agri), and categorical variables (Crop, Tillage). In addition, two meteorological inputs (minimum (Tmin) and mean air temperature (Tmed)), were selected. The ANNs were adequate to estimate SOC in the upper 0.20 m of Mollisols of the SEBA. The model with the best performance included six management variables (SOCI, Year, Prop soybean, Tillage, Yield, Prop agri) and one meteorological variable (Tmin), all of them easily available and with low level of uncertainty. Soil organic C changes related to soil use in the SEBA could be satisfactorily estimated using an ANN developed with simple and easily available input variables. Artificial neural network technique appears as a valuable tool to develop robust models to help predicting SOC changes.

RESUMEN

El carbono orgánico del suelo (SOC) tiene un papel clave en el ciclo global del carbono. Las relaciones complejas entre los componentes del ciclo de C hacen difícil la modelización de la variación del SOC. Las Redes Neuronales Artificiales (ANN) son modelos capaces de determinar las interrelaciones existentes basadas en información disponible. El objetivo fue desarrollar y evaluar modelos basados en la técnica de ANN para estimar el SOC en Mollisoles del sudeste de la Provincia de Buenos Aires, Argentina (SEBA). Fueron empleados datos provenientes de tres experimentos de larga duración conducidos en el SEBA. Variables de manejo y meteorológicas fueron seleccionadas como entradas de las ANN. La información de manejo incluyó variables numéricas (SOC inicial (SOCI); número de años desde el inicio del experimento (Year), proporción de soja (Prop soybean), rendimiento de cultivos (Yield), proporción de la agricultura en la secuencia (Prop agri)) y variables categóricas (cultivo (Crop), sistema de labranza (Tillage)). Además, dos variables meteorológicas (temperatura mínima (Tmin) y temperatura promedio (Tmed))

DOI: 10.3232/SJSS.2017.V7.N3.03

fueron consideradas. Las ANN estimaron adecuadamente el SOC en los 0,20 m superiores de Mollisoles del SEBA. El modelo con mejor desempeño fue desarrollado a partir de una variable meteorológica (Tmin) y seis variables de manejo (SOCI, Year, Prop sowbean, Tillage, Yield, Prop agri), todas ellas fácilmente accesibles y con bajo nivel de incertidumbre.

RESUMO

O carbono orgânico do solo (SOC) tem um papel fundamental no ciclo global do carbono. As relações complexas entre os componentes do ciclo do C dificulta a modelação da variação do SOC. As redes neuronais artificiais (ANN) são modelos capazes de determinar as inter-relações existentes com base em informação disponível. O objetivo deste trabalho foi desenvolver e avaliar modelos baseados na técnica de ANN para estimar o SOC em Molisolos do sudeste da província de Buenos Aires, Argentina (SEBA). Foram utilizados dados de três ensaios de longa duração conduzidos em SEBA. Variáveis meteorológicas e de gestão foram selecionadas como dados de entrada das ANN. Informações de gestão incluíram variáveis numéricas (concentração inicial de SOC (SOCI); número de anos desde o início do ensaio (Year), a proporção de soja na sequência da colheita; (Prop soja), rendimento da colheita (Yield); proporção de cultivo na sequência da rotação da cultura (Prop agri)) e variáveis categóricas (cultivo (Crop), e sistema de lavoura (Tillage)). Além disso, consideraram-se duas variáveis meteorológicas (temperatura média do ar (Tmed) e temperatura mínima do ar (Tmin)). Os modelos baseados em ANN demonstraram ser adequados para estimar o SOC nas camadas superiores (0,20 m) dos Molisolos do SEBA. O modelo com melhor desempenho foi desenvolvido a partir de uma variável meteorológica (Tmin) e seis variáveis de gestão (SOCI, Year, Prop soja, Tillage, Yield, Prop agri), sendo todas as varáveis facilmente acessíveis e com baixo nível de incerteza. As alterações no SOC relacionadas com o uso do solo no SEBA poderiam ser satisfatoriamente estimadas usando uma ANN desenvolvida a partir de variáveis simples e facilmente disponíveis. A técnica de ANN parece ser uma ferramenta válida para desenvolver modelos robustos para ajudar a prever as alterações de SOC.

KEY WORDS
Modeling, cropping systems, tillage systems.

PALABRAS CLAVE
Modelación, sistemas de cultivo, sistemas de labranza.

PALAVRAS-CHAVE
Modelação, preparação do solo, sistemas de cultivo.

1. Introduction

Soil organic carbon (SOC) is both source and sink of atmospheric C dioxide and plays a key role in the global carbon (C) cycle. Besides, its content impacts on soil nutrient supply and on soil water storage capacity and, therefore, on crop yields. In addition, it is one of the most sensitive soil components to land use (Quiroga and Studdert 2015). However, the relationships among the components of C cycle and the factors that determine their fluxes, are very complex and, therefore, their study and prediction turn difficult (Parton et al. 1987; Smith et al. 1997).

Empirical and stochastic models have been developed to describe complex interactions (Parton et al. 1987; Hansen et al. 1991; Franko et al. 1995; Liang et al. 2008; Kemanian and Stôckle 2010). However, their results tend to be over-simplified since they cannot take into account all the critical factors and non-linear relationships that influence C dynamics. On the other hand, some models are complex and/or require very detailed information that is not usually available or is difficult to estimate (e.g. Century model) (Levine and Kimes 1998), that make them unfeasible for generalized use.

Some researchers appealed to the artificial neural networks (ANN) technique to overcome some limitations of other modeling techniques. Artificial neural networks allow describing complex interrelationships based on simple information available. The technique has been applied to estimate either properties or processes that define soil status variables, and among them, to characterize SOC dynamics in different environments (Levine and Kimes 1998; Ingleby and Crowe 2001; Somaratne et al. 2005). In Argentina, some estimation of SOC in soils of the Pampas and Chaco were satisfactorily performed (Álvarez 2008; Álvarez et al. 2009, 2011, 2012; De Paepe and Álvarez 2013).

Despite the high and stable SOC content of the soils of the Southeastern of Buenos Aires Province, Argentina (SEBA) soils, the progressive increase of cropping in the last decades, has led to a sharp SOC loss (Sainz Rozas et al. 2011; Reussi Calvo et al. 2014). The sustainable use of these soils requires the knowledge of the impact of management practices on SOC dynamics to be able to use soil preserving its health. Some simulation models have been locally calibrated and validated with acceptable results (Studdert et al. 2011; Moreno et al. 2016), but they were not developed for the SEBA conditions. On the other hand, some preliminary attempts were done to estimate and interpret the variation of SOC in soils of the SEBA under conventional tillage, using ANN with promising results (Moreno et al. 2014a, 2014b). We hypothesized that ANN models developed using available local information will satisfactorily estimate SOC changes in loam-high-organic-matter-content soils under different cropping systems. The objective of this work was to develop and evaluate ANN models to estimate SOC content changes in soils of the SEBA.

2. Materials and methods

2.1. Experimental site

Data from three long-term soil management experiments carried out in the experimental field of the *Unidad Integrada Balcarce*, Balcarce, Buenos Aires Province, Argentina (37° 45' S, 58° 18' W, 138 m over sea level) between 1976 and 2012 was used. The experiments were set on a soil complex of Typic Argiudoll (Soil Survey Staff 2014) (Mar del Plata series (INTA 1979)) and Petrocalcic Argiudoll (Soil Survey Staff 2014) (Balcarce series, with petrocalcic horizon below 0.7 m depth (INTA 1979)). Clay, silt, sand and soil organic matter concentrations of the soil complex surface layer (0-20 cm depth) are 232, 343, 425, and 63.0 g kg^{-1}, respectively, and the texture class is loam (INTA 1979). Cation exchange capacity, base saturation and pH are 24.0 cmol$_c$ kg^{-1}, 74.1% and 6.1, respectively. Bulk density varies between 1.1 and 1.25 Mg m^{-3}. The slope is less than 2% and, therefore, soil water erosion was considered negligible. Climate is mesothermal sub-humid to humid (according to Thornthwaite) or temperate-humid without a dry season (according to Köppen). The median annual rainfall is 939 mm yr^{-1} and annual mean daily temperature is 13.9 °C (Agri-Weather Station, Unidad Integrada Balcarce, located ~1000 m away from the experiments).

2.2. Experiment description

Information from three long term experiments carried out with a randomized complete block design and a split-plot treatment arrangement, was used:

1) "Continuous Cropping": carried out between 1984 and 1995 with 16 crop sequences including wheat (*Triticum aestivum* L.), soybean (*Glycine max* (L) Merr.), maize (*Zea mays* L.), and sunflower (*Helianthus annuus* L.) under conventional tillage (CT, moldboard plow, disk harrow, and field cultivator) and with and without N (WN and WON, respectively). This experiment is more thoroughly described in Studdert and Echeverría (2000).

2) "Crop-pasture Rotations": carried out between 1976 and 2006 with different combinations of periods under cropping (wheat, soybean, maize, sunflower, potato (*Solanum tuberosum* L.), and oat (*Avena sativa* L.) and vetch (*Vicia sativa* L) or red clover (*Trifolium pratensse* L.) as green manures) with and without N (WN and WON, respectively), and periods under

grass-based pastures. Between 1976 and 1993 tillage system was CT and between 1994 and 2006 some treatments were under CT and other under no-tillage (NT). More information about this experiment between 1976 and 1993 can be found in Studdert et al. (1997). The phase between 1994 and 2003 has been described in Eiza et al. (2005). Between 2004 and 2006, treatments and tillage systems were the same as described by Eiza et al. (2005).

3) “Tillage systems”: carried out from 1997 with the sequence maize, sunflower, wheat, under two tillage systems (CT and NT) and with and without N (WN and WON, respectively). More information about this experiment can be found in Diovisalvi et al. (2008).

Soil organic C concentration at 0-0.20 m depth in the fall of most of the years of each experiment (Moreno et al. 2016) had been determined through wet combustion with maintenance of the reaction temperature (120 °C) for 90 min (a variant of the Walkley-Black method, Schlichting et al. 1995). Concentration of SOC was converted into stock (Mg C ha^{-1}) using bulk density determined or estimated as described by Studdert et al. (2011). Furthermore, crop productivity data was available as grain yield at commercial humidity content (14.0% for wheat, 14.5% for maize, 13.5% for soybean and 11.0% for sunflower), as tuber yield for potato and dry matter of aboveground biomass for oat and vetch (Moreno et al. 2016). Yields for grass-based pastures, expressed as dry matter of aboveground biomass, were estimated according to Agnusdei et al. (2001).

2.3. ANN-based models

An ANN is a parallel processing structure constituted by units (neurons) organized in layers that emulate biological neurons (Haykin 2001). The ANN have the capacity of identifying complex relationships from input information (different input variables, x_1 … x_n, **Figure 1**) through the approximation of any mathematical function along a training procedure to yield a desired output. Besides, ANN are capable of storing knowledge about the relationships among input variables and about its proper functioning, that could be made available through different analysis techniques (Braga et al. 2007).

An ANN is characterized by its structure or architecture, the training algorithm and the activation functions (Braga et al. 2007) and it is imperative to define them to develop an ANN-based model. A schematic representation of an artificial neuron (basic unit in an ANN model) is shown in **Figure 1**.

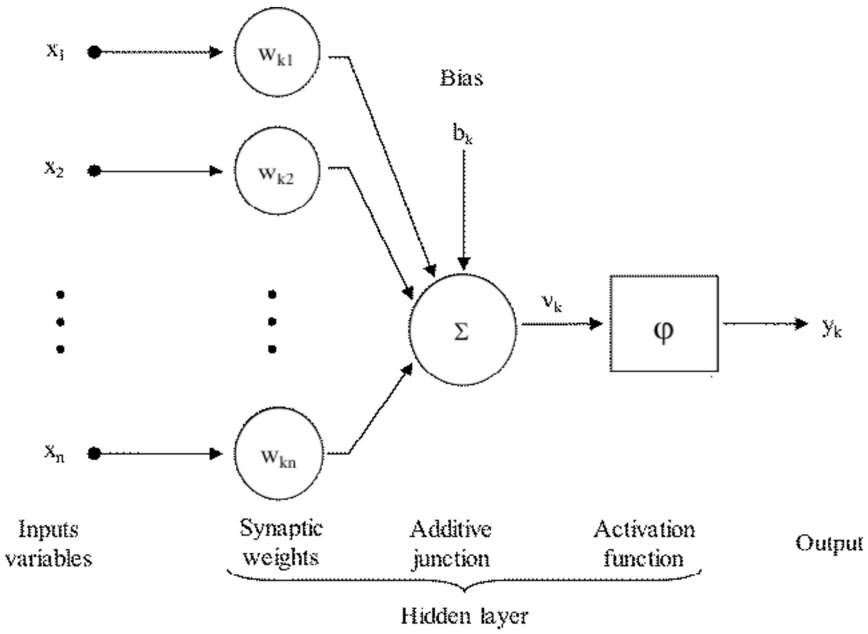

Figure 1. Scheme of an artificial neuron, adapted from Haykin (2001).

The multilayer perceptron network (MLP) is one of the most commonly used feed forward ANN type. A MLP network consists of one input layer, one or more hidden layers and one output layer. The strength of the connection between two neurons in adjacent layers is represented by what is known as a 'synaptic weight'. The additive junction (Σ) represents the addition of signals in the inputs layer weighted by their respective synaptic weights (w_k). Then, the activation function (φ) limits the amplitude of the output of the neuron. The bias b_k increases or decreases the input to activation function, assigning positive or negative values. According to the bias (positive or negative), the relationship between the induced field or activation potential (v_k) and the output (y_k) is transformed.

Mathematically, an ANN can be described by the equations:

Eq. 1

$$y_k = \varphi\left(\sum_{i=1}^{n} x_i w_{ki}\right) + b_k$$

Eq. 2

$$v_k = u_k + b_k$$

where y_k is the output neuron; φ is the activation function; x_i is the i-th input variable; w_{ki} is the synaptic weight of the k neuron for the i-th input variable, and b_k is the bias. The artificial neuron computes its output (y_k) according to the **Equation (1)**. In **Equation (2)** v_k indicates the weighted inputs (u_k) affected by the bias (b_k).

The size of the network is linked to the nature of the problem to be solved and the number of patterns or training pairs of inputs (x) - outputs (y) (Rogers and Dowla 1994). Then, the dimensionality of the models tends to be much higher in more complex problems (Maier and Dandy 2000). In addition, network architecture determines the number of connection weights (free parameters) and the way information flows through the network (Maier and Dandy 2000). The number of free parameters (N) is defined by:

Eq. 3

$$N = (n * m + m) + (m * x + x)$$

where n is the number of inputs, m is the number of hidden layers and x is the number of outputs.

2.4. Development of ANN-based models

Multilayer perceptron models with a unique hidden layer were developed to estimate SOC in the soil upper 0.20 m. It has been shown that only one hidden layer is required to approximate any continuous function (Cybenko 1989). In this study we developed MLP network models with one hidden layer and one output layer. Therefore, the size of each network was defined by the number of input variables and the number of neurons in the hidden layer.

To go through the mechanism of model development we pre-selected 16 input variables (three categorical variables and 13 quantitative variables) based on availability of information and potential relationships with SOC stock variation:

- Nitrogen fertilization (WN or WON) (categorical).

- Tillage: tillage system (NT or CT) (categorical).

- Crop: preceding crop to soil sampling for SOC content determination (categorical).

- Year: number of years since the beginning of the experiment up to soil sampling for SOC content determination for each treatment (quantitative).

- Yield: average grain yield of all the crops in the sequence (kg grain ha^{-1}) since the beginning of the experiment up to the year before soil sampling for SOC content determination (quantitative).

- C Input: average input of C by crop sequence (Mg ha^{-1}) since the beginning of the experiment up to the preceding crop to soil sampling for SOC content determination (quantitative). To calculate C input, wheat, soybean, sunflower and maize grain yields, potato tuber yield, and oat and vetch aboveground dry matter production were used. The calculation of residue input mass by wheat, soybean, sunflower, maize, and potato was done using the grain or tuber yield, and harvest indexes (HI) and the below- (root biomass + rhizodeposition)/aboveground biomass (RB/TAB) relationship used by Studdert et al. (2011).

For oat and vetch, RB/TAB was assumed the same as for wheat (Studdert et al. 2011). Pasture aboveground dry matter production was estimated as reported by Agnusdei et al. (2001) for similar pastures. Pasture RB/TAB was estimated according to Bélanger et al. (1992). Carbon content of plant tissues was assumed as 0.43 kg C kg^{-1} (Sánchez et al. 1996).

- Prop agri: proportion of cropping in the whole crop rotation (quantitative).
- Prop SC: proportion of summer crops in the whole crop rotation (quantitative).
- Prop SC agri: proportion of summer crops in the cropping phases (quantitative).
- Prop soybean: proportion of soybean in the cropping phases (quantitative).
- SOCI: SOC at the beginning of each experiment (Mg ha^{-1}).
- Pp: accumulated precipitation (mm) (quantitative).
- ET0: annual reference evapotranspiration estimated by Penman-Monteith model (mm) (quantitative).
- Tmin: mean annual minimum air temperature (°C) (quantitative).
- Tmax: mean annual maximum air temperature (°C) (quantitative).
- Tmed: mean annual mean air temperature (°C) (quantitative).

Values for each meteorological variable were the result of the summation (Pp) or average (Tmin, Tmax, Tmed) of data over the 12 months previous to each soil sampling for SOC content determination.

Total data was split into training, test and validation groups with the proportion 60:20:20. The training group (n = 1083) was used during model training. The validation group (n = 359) was used for cross-validation (Maier and Dandy 2000) and the test group (n = 359) was used to evaluate the final performance of each model (Haykin 2001). Data for each group was randomly selected and distribution of frequencies among groups were homogenous (Kruskal-Wallis test, $p > 0.05$) (data not shown). To define which of the 16 pre-selected variables would be used as better input variables we based on Spearman correlation analysis between observed SOC in the upper 0.20 m and each one of them (**Table 1**).

Table 1. Spearman correlation coefficients ($p < 0.05$) between soil organic carbon stock in the upper 0.20 m of soil and different quantitative management and meteorological variables

Variable*	Overall	Data group		
		Training	Validation	Test
Year	-0.58	-0.56	-0.62	-0.56
Yield (kg ha^{-1})	-0.22	-0.25	-0.19	-0.16
C input (Mg ha^{-1})	-0.12	-0.16	NS§	NS§
Prop agri	0.18	0.23	NS§	NS§
Prop SC	-0.08	NS§	NS§	-0.12
Prop SC agri	-0.20	-0.22	-0,16	-0.17
Prop soybean	-0.32	-0.34	-0.31	-0.29
SOCI (Mg ha^{-1})	0.42	0.39	0.46	0.45
Pp (mm)	NS§	NS§	NS§	NS§
ET0 (mm)	NS§	NS§	NS§	-0.10
Tmin (°C)	-0.27	-0.26	-0.31	-0.24
Tmed (°C)	-0.23	-0.22	-0.26	-0.22
Tmax (°C)	-0.05	NS§	NS§	NS§

* For variable description see text. § NS: not significant.

Initial SOC and Year were the variables that showed the highest correlation with SOC, both for all data and after splitting it into the different data groups, with coefficients close to 0.45 (SOCI) and -0.50 (Year). Likewise, correlation coefficients for Prop soybean, Yield and Prop agri, were high. On the other hand, among meteorological variables, only Tmin and Tmed showed significant correlation with SOC stock. Therefore, the interpretation of correlation coefficients led to the selection of five basic management variables as quantitative input variables (SOCI; Year, Prop soybean; Yield, Prop agri) and two meteorological variables (Tmin and Tmed). In addition, two categorical variables (Crop and Tillage) were selected due to their relationship with SOC stock (Moreno et al. 2014b). Even though this work was done for only one soil type and climatic condition, many of the selected variables resulted the same as those selected by other authors who developed ANN-based models for a broader range of environmental conditions of Argentina (Álvarez 2008; Álvarez et al. 2009, 2011, 2012; De Paepe and Álvarez 2013).

Artificial neural network-based models were performed including different combinations of management and meteorological variables, and trained to estimate SOC stock. Methods followed to arrange the inputs in each combination were based on *a priori* knowledge of the system being modelled and on correlations analysis.

Models defined were organized in three subsets as follows:

* **Subset 1:** ANN-based models with two management input variables resulting from the combinations of five input management variables, taken by two. Management variables used were the three quantitative variables with the highest correlation with SOC stock (SOCI, Year and Prop soybean, **Table 1**) and two categorical variables (Crop and Tillage). Seven out of ten possible models with only management variables (basic models) were chosen because of best performance. Additionally, other 21 models were defined including either each and both meteorological variables most correlated with SOC stock (Tmin and Tmed, **Table 1**). In summary, Subset 1 included 28 models.

* **Subset 2:** ANN-based models with three management input variables resulting from the combinations of the seven selected management input variables (SOCI; Year, Prop soybean; Yield, Prop agri, Crop, Tillage, **Table 1**) taken by three. It was imposed the restriction that all models performed always had to include the two management input variables showing the highest correlation with SOC (SOCI and Year, **Table 1**) and one of the other management variables (a total of five basic models). Fifteen additional models were defined including either each and both meteorological variables most correlated with SOC stock (Tmin and Tmed, **Table 1**). In summary, Subset 2 included 20 models.

* **Subset 3:** ANN-based models with more than three management input variables resulting from the combination of the three selected management input variables showing the highest correlation with SOC (SOCI, Year, and Prop soybean) with one (four-variable models), two (five-variable models) or three (six-variable models) of the other selected input variables. Most models in this subset were defined without meteorological, but some of them were also defined including either each and both meteorological variables most correlated with SOC stock (Tmin and Tmed, **Table 1**). Total of models defined in Subset 3 was nine.

To solve estimation problems, a supervised training has to be carried out for which input variables and target observed outputs are provided to the ANN. Training or learning of an ANN with a defined structure is achieved by adjusting the weights of the neurons through an iterative algorithm that minimizes the error between the predicted and the target outputs. This process is equivalent to parameter adjustment in conventional statistical model fitting. Bias values were initially set as 1 and the final value for each ANN was determined in the process. In this work, the selection of ANN architectures was based on the application of a selected algorithm integrated on the Intelligent Problem Solver (IPS) of the Neural Network module of Statistica Software (Statsoft 2009). The inputs and the outputs of data sets were automatically normalized to improve the performance of ANN models. The maximal number of neurons was fixed related to the number of examples trained. The Automated Network Search (ANS) of the software, was set to retain the five models with the lowest cross-validation error (over 200 ANN for each combination of input variables it was asked to train) and then, the ANN with the best performance for each combination was chosen and evaluated. Two types of transformed sigmoid activation functions (i.e. logistic and hyperbolic tangent) were applied in the hidden layer and linear ones in the output layer. The sigmoid response allows a network to map a non-linear process and is recommended to avoid saturation and convergence in approximation problems.

2.5. Evaluation of ANN model performance

The performance of the ANN was evaluated on test data group using several standard statistical performance evaluation criteria based on the difference between observed and simulated SOC stock values. Those statistical indicators were: mean of the differences between observed and simulated values (bias error, BE, Mg C ha^{-1}), mean of those differences relative to the observed values (bias relative error, BRE, %), and root mean square error (RMSE, expressed as stock, Mg C ha^{-1}) (Fox 1981). Performed ANN models were sort (increasing order) through each of the mentioned error types and ranked from the lowest to the highest and assigned a ranking number according to each of all three sorts. A final hierarchical overall ranking of performance was calculated as the average of the three ranking numbers achieved by each ANN-based model for all three sorts. This procedure enabled the determination of the ANN-based model with the best and that with the worst performance. Model performance was also evaluated through simple regression analyses between observed and simulated SOC stock values. The joint hypothesis of equality of intercept and slope of each simple linear regression to 0 and 1, respectively, was evaluated through F tests. All statistical analyses were performed with the R statistical package (R Core Team 2015).

3. Results and discussion

3.1. Description of models

A total 57 ANN-based models were developed (28 in Subset 1 (**Table 2**), 20 in Subset 2 (**Table 3**), and nine in Subset 3 (**Table 4**) to estimate SOC stock including between two and eight input variables and a maximum of ten neurons in the hidden layer. Most of the models had adequate structure, without problems during training, given the large number of training data (n = 1083). The models with Crop as input variable resulted in a higher number of free parameters, since this categorical variable presented 10 input options (i.e. ten different crops). Artificial neural networks with large structure (i.e. high number of input variables and/or of neurons in the hidden layer) could present problems of sub-training. However, Rogers and Dowla (1994) indicated that if the number of weights (or free parameters) does not exceed the number of examples for training, such training problems would not be expected to occur.

3.2. Model performance

According to the evaluation on test data group, linear regression analyses between observed and simulated SOC stock values were all significant ($p < 0.05$) (**Tables 5, 6, 7**). The joint hypothesis of equality of intercept and slope of each simple linear regression to 0 and 1, respectively, was not rejected ($p > 0.05$) in any case (**Tables 5, 6, 7**). However, R^2 ranged only between 0.1 and 0.6 (**Tables 5, 6, 7**). Other authors reported higher R^2 values when estimating SOC concentrations with ANN for several soil types of Argentina (Álvarez et al. 2011, 2012; Berhongaray et al. 2013). The low R^2 obtained in this work could be associated to the large variability in observed SOC stocks among experiment replications. Studdert et al. (1997) reported significant differences ($p < 0.01$) for observed SOC stocks among blocks in the "Crop-pasture rotations" experiment, with 50% of standard deviations ranging between 1.8 Mg C ha^{-1} and 5.2 Mg C ha^{-1}, and an average standard deviation of 3.4 Mg C ha^{-1}. Likewise, Studdert and Echeverría (2000) also reported significant

Table 2. Input variables, structure and activation functions of artificial neural networks (ANN) of multiperceptron type (MLP) trained to estimate soil organic carbon stock in the soil upper 0.20 m for Subset 1

Number of ANN	Input variables*	Structure[§]	Activation function [❐]	N[†]
1	**Year-SOCI**	**MLP 2-10-1**	**Logistic**	**41**
2	Year-SOCI-Tmin	MLP 3-10-1	Hyp Tang	51
3	Year-SOCI-Tmed	MLP 3-4-1	Logistic	21
4	Year-SOCI-Tmin-Tmed	MLP 4-8-1	Hyp Tang	49
5	**Year-Prop soybean**	**MLP 2-8-1**	**Hyp Tang**	**33**
6	Year-Prop soybean-Tmin	MLP 3-7-1	Hyp Tang	36
7	Year-Prop soybean-Tmed	MLP 3-7-1	Hyp Tang	36
8	Year-Prop soybean-Tmin-Tmed	MLP 4-4-1	Hyp Tang	25
9	**Year-Tillage**	**MLP 3-6-1**	**Logistic**	**31**
10	Year-Tillage-Tmin	MLP 4-10-1	Hyp Tang	61
11	Year-Tillage-Tmed	MLP 4-7-1	Hyp Tang	43
12	Year-Tillage-Tmin-Tmed	MLP 5-10-1	Hyp Tang	71
13	**SOCI-Prop soybean**	**MLP 2-10-1**	**Hyp Tang**	**41**
14	SOCI-Prop soybean-Tmin	MLP 3-9-1	Logistic	46
15	SOCI-Prop soybean-Tmed	MLP 3-8-1	Hyp Tang	41
16	SOCI-Prop soybean-Tmin-Tmed	MLP 4-8-1	Hyp Tang.	49
17	**SOCI-Tillage**	**MLP 3-10-1**	**Logistic**	**51**
18	SOCI-Tillage-Tmin	MLP 4-7-1	Hyp Tang	43
19	SOCI-Tillage-Tmed	MLP 4-6-1	Hyp Tang	37
20	SOCI-Tillage-Tmin-Tmed	MLP 5-9-1	Logistic	64
21	**Prop soybean-Tillage**	**MLP 3-5-1**	**Hyp Tang**	**26**
22	Prop soybean-Tillage-Tmin	MLP 4-6-1	Hyp Tang	37
23	Prop soybean-Tillage-Tmed	MLP 4-8-1	Hyp Tang	49
24	Prop soybean-Tillage-Tmin-Tmed	MLP 5-6-1	Hyp Tang	43
25	**Year-Crop**	**MLP 11-8-1**	**Logístic**	**105**
26	Year-Crop-Tmin	MLP 12-8-1	Tang Hip.	113
27	Year-Crop-Tmed	MLP 12-5-1	Tang Hip.	71
28	Year-Crop-Tmin-Tmed	MLP 13-8-1	Tang Hip.	121

* See input variable description in text; [§] Structure: MLP n-m-x: n: number of input variables, m: number neurons in the hidden layer, x: number of output layers; [❐] Activation function in the hidden layer: Hyp tang: hyberbolic tangent; [†] N: number of free parameters.

differences ($p < 0.01$) for SOC stocks among blocks of the experiment "Continuous cropping", with an average standard deviation of 3.6 Mg C ha^{-1} (50% of standard deviations ranging between 2.0 and 5.4 Mg C ha^{-1}). This fact, together with the relative low observed SOC stock change (i.e. decrease) over the 26 years evaluated, may have contributed to the low R^2 values obtained.

Table 3. Input variables, structure and activation functions of artificial neural networks (ANN) of multiperceptron type (MLP) trained to estimate soil organic carbon stock in the soil upper 0.20 m for Subset 2

Number of ANN	Input variables*	Structure§	Activation function ❐	N†
29	**Year-SOCI-Prop soybean**	**MLP 3-8-1**	**Hyp Tang**	**41**
30	Year-SOCI-Prop soybean-Tmin	MLP 4-8-1	Hyp Tang	49
31	Year-SOCI-Prop soybean-Tmed	MLP 4-10-1	Hyp Tang	61
32	Year-SOCI-Prop soybean-Tmin-Tmed	MLP 5-9-1	Logistic	64
33	**Year-SOCI-Tillage**	**MLP 4-5-1**	**Hyp Tang**	**31**
34	Year-SOCI-Tillage-Tmin	MLP 5-10-1	Hyp Tang	71
35	Year-SOCI-Tillage-Tmed	MLP 5-9-1	Hyp Tang	64
36	Year-SOCI-Tillage-Tmin-Tmed	MLP 6-8-1	Logistic	65
37	**Year-SOCI-Yield**	**MLP 3-5-1**	**Hyp Tang**	**26**
38	Year-SOCI-Yield-Tmin	MLP 4-5-1	Hyp Tang	31
39	Year-SOCI-Yield-Tmed	MLP 4-8-1	Hyp Tang	49
40	Year-SOCI-Yield-Tmin-Tmed	MLP 5-9-1	Hyp Tang	64
41	**Year-SOCI-Prop Agri**	**MLP 3-3-1**	**Logistic**	**16**
42	Year-SOCI-Prop agri-Tmin	MLP 4-10-1	Logistic	61
43	Year-SOCI-Prop agri-Tmed	MLP 4-6-1	Hyp Tang	37
44	Year-SOCI-Prop agri-Tmin-Tmed	MLP 5-10-1	Hyp Tang	71
45	**Year-SOCI-Crop**	**MLP 12-5-1**	**Hyp Tang**	**71**
46	Year-SOCI-Crop-Tmin	MLP 13-7-1	Hyp Tang	106
47	Year-SOCI-Crop-Tmed	MLP 13-10-1	Hyp Tang	151
48	Year-SOCI-Crop-Tmin-Tmed	MLP 14-8-1	Hyp Tang	129

* See input variable description in text; § Structure: MLP n-m-x: n: number of input variables, m: number of neurons in the hidden layer, x: number of output layers; ❐ Activation function in the hidden layer: Hyp tang: hyperbolic tangent; † N: number of free parameters.

Table 4. Input variables, structure and activation functions of artificial neural networks (ANN) of multiperceptron type (MLP) trained to estimate soil organic carbon stock in the soil upper 0.20 m for Subset 3

Number of ANN	Input variables*	Structure§	Activation function ◻	N†
49	Year-SOCI-Prop soybean-Tillage	MLP 5-8-1	Hyp Tang	57
50	Year-SOCI-Prop soybean-Yield	MLP 4-4-1	Hyp Tang	25
51	Year-SOCI-Prop soybean-Prop agri	MLP 4-8-1	Logistic	49
52	Year-SOCI-Prop soybean-Tillage-Yield	MLP 6-4-1	Logistic	33
53	Year-SOCI-Prop soybean-Tillage-Prop agri	MLP 6-5-1	Hyp Tang	41
54	Year-SOCI-Prop soybean-Tillage-Yield-Prop agri	MLP 7-8-1	Hyp Tang	73
55	Year-SOCI-Prop soybean-Tillage-Yield-Prop agri-T min	MLP 8-8-1	Hyp Tang	81
56	Year-SOCI-Prop soybean-Tillage-Yield-Prop agri-T med	MLP 8-6-1	Logistic	61
57	Year-SOCI-Prop soybean-Tillage-Yield-Prop agri-T min-Tmed	MLP 9-9-1	Logistic	100

* See input variable description in text; § Structure: MLP n-m-x: n: number of input variables, m: number of neurons in the hidden layer, x: number of output layers; ◻ Activation function in the hidden layer: Hyp tang: hyberbolic tangent; † N: number of free parameters.

Table 5. Statistical indicators obtained on test data group from simple linear regressions between observed and simulated with artificial neural network (ANN) models of soil organic carbon stock at 0.20 m for Subset 1 (**Table 2**)

Number of ANN	Input variables*	Statistical indicators§			
		a	b	R^2	p value a = 0 and b = 1
1	Year-SOCI	0.78	0.98	0.47	0.88
2	Year-SOCI-Tmin	-0.60	1.00	0.52	0.97
3	Year-SOCI-Tmed	0.32	0.99	0.52	0.95
4	**Year-SOCI-Tmin-Tmed**	**0.22**	**0.99**	**0.56**	**0.92**
5	Year-Prop soybean	-1.03	1.01	0.36	0.76
6	Year-Prop soybean-Tmin	2.21	0.97	0.41	0.89
7	Year-Prop soybean-Tmed	3.84	0.95	0.39	0.74
8	**Year-Prop soybean-Tmin-Tmed**	**0.53**	**0.99**	**0.44**	**0.96**
9	Year-Tillage	2.25	0.97	0.32	0.93
10	Year-Tillage-Tmin	2.68	0.96	0.43	0.77
11	Year-Tillage-Tmed	-1.60	1.01	0.42	0.92
12	**Year-Tillage-Tmin-Tmed**	**1.57**	**0.97**	**0.48**	**0.90**

13	SOCI-Prop soybean	0.55	0.99	0.28	0.91
14	SOCI-Prop soybean-Tmin	0.98	0.98	0.47	0.44
15	SOCI-Prop soybean-Tmed	1.56	0.97	0.35	0.93
16	**SOCI-Prop soybean-Tmin-Tmed**	**0.64**	**0.98**	**0.52**	**0.45**
17	SOCI-Tillage	-8.40	1.09	0.23	0.31
18	SOCI-Tillage-Tmin	-2.40	1.02	0.46	0.67
19	SOCI-Tillage-Tmed	-6.20	1.07	0.40	0.40
20	SOCI-Tillage-Tmin-Tmed	-4.74	1.05	0.56	0.24
21	Prop soybean-Tillage	6.26	0.92	0.10	0.78
22	Prop soybean-Tillage-Tmin	5.25	0.93	0.39	0.45
23	Prop soybean-Tillage-Tmed	4.01	0.95	0.28	0.82
24	Prop soybean-Tillage-Tmin-Tmed	-4.07	1.05	0.46	0.65
25	Year-Crop	-0.51	1.01	0.35	0.86
26	Year-Crop-Tmin	5.54	0.93	0.45	0.41
27	Year-Crop-Tmed	2.16	0.97	0.46	0.72
28	Year-Crop-Tmin-Tmed	6.35	0.91	0.50	0.08

* See input variable description in text; § Statistical indicators: a: intercept, b: slope, R^2: coefficient of determination.

Table 6. Statistical indicators obtained on test data group from simple linear regressions between observed and simulated with artificial neural network (ANN) models of soil organic carbon stock at 0.20 m for Subset 2 (**Table 4**)

Number of ANN	Input variables*	Statistical indicators§			
		a	b	R^2	p value a = 0 and b = 1
29	**Year-SOCI-Prop soybean**	**1.55**	**0.97**	**0.49**	**0.88**
30	Year-SOCI-Prop soybean-Tmin	1.93	0.97	0.54	0.80
31	Year-SOCI-Prop soybean-Tmed	2.36	0.96	0.53	0.64
32	Year-SOCI-Prop soybean-Tmin-Tmed	-0.13	0.99	0.56	0.57
33	**Year-SOCI-Tillage**	**2.52**	**0.96**	**0.48**	**0.67**
34	Year-SOCI-Tillage-Tmin	3.67	0.95	0.53	0.45
35	Year-SOCI-Tillage-Tmed	1.68	0.97	0.56	0.77
36	Year-SOCI-Tillage-Tmin-Tmed	2.27	0.96	0.59	0.61
37	**Year-SOCI-Yield**	**-1.58**	**1.01**	**0.50**	**0.75**
38	Year-SOCI-Yield-Tmin	-0.30	1.00	0.54	0.67
39	Year-SOCI-Yield-Tmed	1.18	0.98	0.52	0.91
40	Year-SOCI-Yield-Tmin-Tmed	0.41	0.99	0.56	0.45

41	**Year-SOCI-Prop agri**	**-0.11**	**1.00**	**0.49**	**0.98**
42	Year-SOCI-Prop agri-Tmin	1.52	0.97	0.54	0.78
43	Year-SOCI-Prop agri-Tmed	1.89	0.97	0.54	0.86
44	Year-SOCI-Prop agri-Tmin-Tmed	3.88	0.95	0.57	0.53
45	**Year-SOCI-Crop**	**3.45**	**0.95**	**0.48**	**0.70**
46	Year-SOCI-Crop-Tmin	5.13	0.93	0.57	0.21
47	Year-SOCI-Crop-Tmed	3.99	0.94	0.58	0.23
48	Year-SOCI-Crop-Tmin-Tmed	5.18	0.93	0.60	0.11

* See input variable description in text; § Statistical indicators: a: intercept, b: slope, R^2: coefficient of determination.

Table 7. Statistical indicators obtained on test data group from simple linear regressions between observed and simulated with artificial neural network (ANN) models of soil organic carbon stock at 0.20 m for Subset 3 (**Table 4**).

Number of ANN	Input variables*	Statistical indicators§			
		a	b	R^2	p value a = 0 and b = 1
49	Year-SOCI-Prop soybean-Tillage	5.16	0.93	0.50	0.42
50	Year-SOCI-Prop soybean-Yield	-1.12	1.01	0.50	0.95
51	Year-SOCI-Prop soybean-Prop Agri	3.42	0.95	0.50	0.65
52	Year-SOCI-Prop soybean-Tillage-Yield	0.47	0.99	0.53	0.99
53	Year-SOCI-Prop soybean-Tillage-Prop agri	3.41	0.96	0.53	0.49
54	Year-SOCI-Prop soybean-Tillage-Yield-Prop agri	7.03	0.91	0.54	0.07
55	Year-SOCI-Prop soybean-Tillage-Yield-Prop agri-Tmin	6.58	0.91	0.58	0.14
56	Year-SOCI-Prop soybean-Tillage-Yield-Prop agri-Tmed	4.31	0.94	0.57	0.36
57	Year-SOCI-Prop soybean-Tillage-Yield prom-Prop agri-Tmin-Tmed	5.46	0.93	0.59	0.24

* See input variable description in text; § Statistical indicators: a: intercept, b: slope, R^2: coefficient of determination.

Root mean square error, BRE, and BE values obtained on test data group when contrasting observed vs. simulated SOC stocks are presented in **Figure 2**. Most of the 57 ANN-based models defined showed acceptable results (Smith et al. 1997). In general, RMSE ranged between 4.97 and 7.39 Mg C ha^{-1} and did not differ from those reported by Álvarez et al. (2009) and some models yielded better indicators than those reported by Álvarez et al. (2011). Bias relative errors ranged between 4.69 and 7.26% and BE ranged between -0.39 and 0.49 Mg C ha^{-1}. Other authors (Levine and Kimes 1998; Somaratne et al. 2005) reported even lower errors but they used both management variables and chemical properties as input variables.

Error variability among ANN-based models including only two management variables (Subset 1, **Table 2, Figure 2**) was high. On the other hand, error variability among ANN-based models of Subsets 2 (three management variables, **Table 3, Figure 2**) and 3 (more than three management variables, **Table 4, Figure 2**), were lower than those of Subset 1 and similar between them. In all cases, the inclusion of the selected meteorological input variables (i.e. Tmin and/or Tmed, **Tables 2, 3, 4**) improved model performance through reducing errors (**Figure 2**). Therefore, SOC stock could be satisfactorily estimated with ANN models including only three management input variables and selected meteorological input variables.

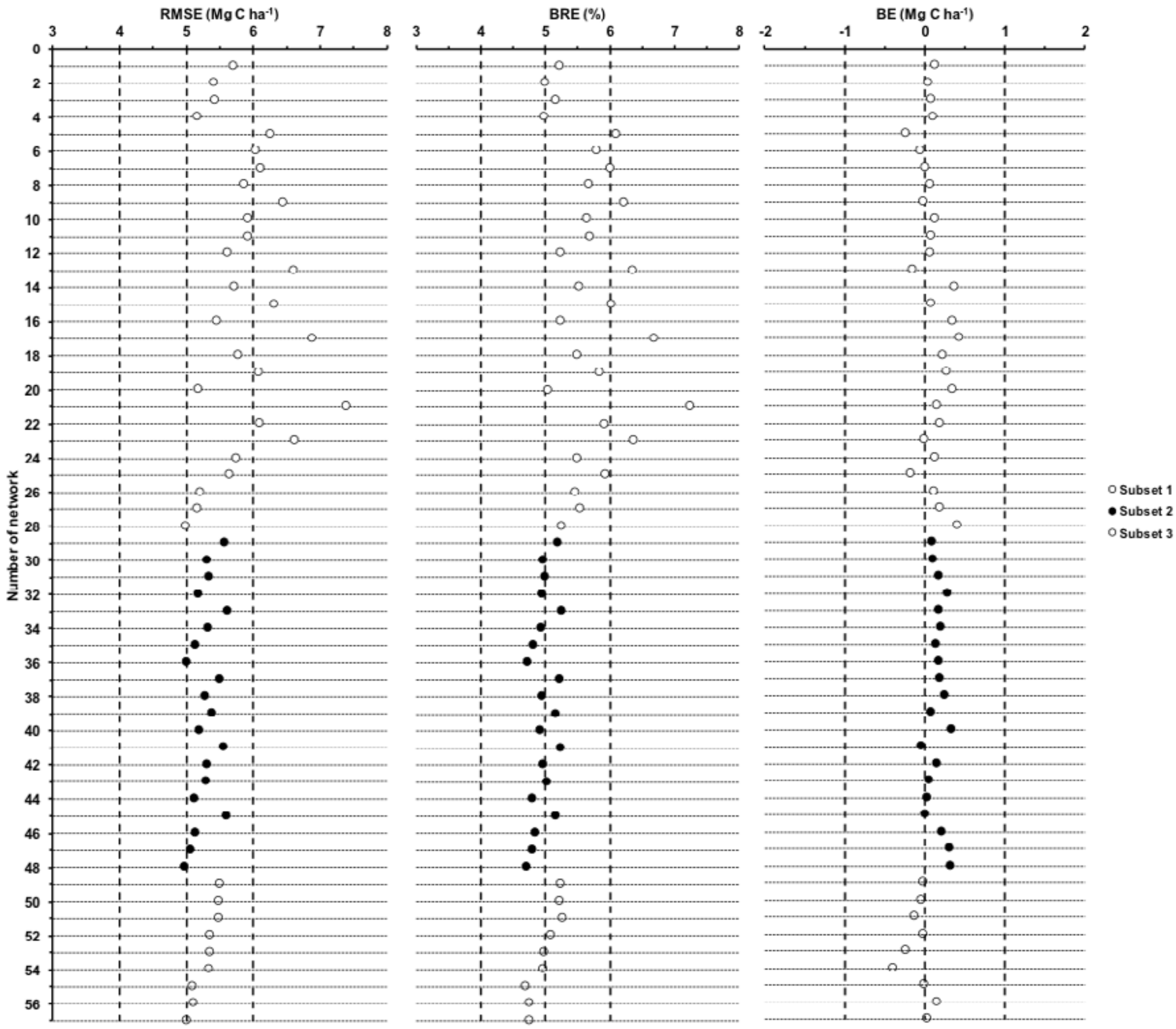

Figure 2. Statistical indicators (root mean square error (RMSE), bias error (BE); relative bias error (BRE)) for each artificial neuron network (ANN) based model trained (number of ANN-based models can be seen in **Tables 2, 3,** and **4** for model Subsets 1, 2, and 3, respectively).

3.3. Models with the best and the worst performances

Table 8 shows the 10 ANN models with the lowest (best models, first 10 hierarchical positions) and the highest (worst models, last ten hierarchical positions) average of individual positions of ranking through RMSE, BRE, and BE. Root mean square errors of the ten best models ranged between 4.97 and 5.36 Mg C ha^{-1}, BRE ranged between 4.70 and 5.09%, and BE ranged between -0.01 and 0.33 Mg C ha^{-1}. Only one of the ten best models belongs to Subset 1 (ANN 4, **Tables 2, 8**) and only one did not include meteorological input variables (ANN 52, **Tables 4, 8**). Five out of nine of the best models including meteorological variables, included both Tmin and Tmed (ANN 57, 44, 36, 4, and 48, **Tables 2, 3, 4, 8**). The best model (ANN 55, **Tables 4, 8**) also showed one of the highest R^2 of observed vs. simulated SOC stock linear regressions (0.58, **Table 7**). On the other hand, all the worst models integrated Subset 1. Their RMSE ranged between 5.66 and 7.40 Mg C ha^{-1}, BRE ranged between 5.50 and 7.26, and BE ranged between -0.23 and 0.44 Mg C ha^{-1}. Five out of all the worst models did not include meteorological variables and the rest, included only one (either Tmin or Tmed). The R^2 of linear regression of observed vs. simulated with the worst model (ANN 17, **Tables 2, 8**) SOC stocks, was very low (R^2 = 0.23, **Table 5**).

Table 8. Best and worst positions within the hierarchical ranking of the trained artificial neural network (ANN) models on the basis of the average of the ranking positions (increasing order) sorting by three statistical indicators. RMSE: root mean square error (Mg C ha^{-1}); BRE: bias relative error (%); BE: bias error (Mg C ha^{-1}). The ANN are described in **Tables 2, 3,** and **4**.

	Number of ANN	Subset	RMSE ranking	BRE ranking	BE ranking	Error ranking average	Final hierarchy ranking
Best models	**55**	**3**	**6**	**1**	**1**	**2.7**	**1**
	57	3	4	4	9	5.7	2
	44	2	8	6	8	7.3	3
	36	2	3	3	36	14.0	4
	56	3	7	5	33	15.0	5
	35	2	10	8	29	15.7	6
	4	1	11	18	22	17.0	7
	52	3	25	23	4	17.3	8
	43	2	18	21	14	17.7	9
	48	2	1	2	50	17.7	10
Worst models	15	1	52	51	17	40.0	48
	25	1	39	49	34	40.7	49
	18	1	43	39	43	41.7	50
	14	1	41	41	54	45.3	51
	22	1	49	48	39	45.3	52
	13	1	54	54	30	46.0	53
	19	1	48	47	47	47.3	54
	21	1	57	57	32	48.7	55
	5	1	51	52	44	49.0	56
	17	1	56	56	57	56.3	57

Studdert et al. (2011) and Moreno et al. (2016) reported that the performance of RothC (Jenkinson et al. 1987) and AMG (Andriulo et al. 1999) models, respectively, to simulate SOC stock showed some differences between nitrogen fertilization levels and/or tillage systems. Therefore, we also evaluated the best (**Figure 3**) and worst (**Figure 4**) ANN-based model performances through RMSE and BE discriminated by agronomic management (i.e. separately for each tillage system level (regardless nitrogen fertilization level) and for each nitrogen fertilization level (regardless tillage system level). According to RMSE, the best ANN estimated better (lower RMSE) SOC stock under NT and WN. However, dispersion of BE was a little higher and some ANN-based models showed no difference between levels of tillage system nor between nitrogen fertilization levels, but some others showed an inverse trend than that of RMSE. Anyway, the best ANN model (ANN 55, **Table 8**) did not show differences between the levels of both management practices, and, despite the differences, the RMSE were all within acceptable levels (Smith et al. 1997).

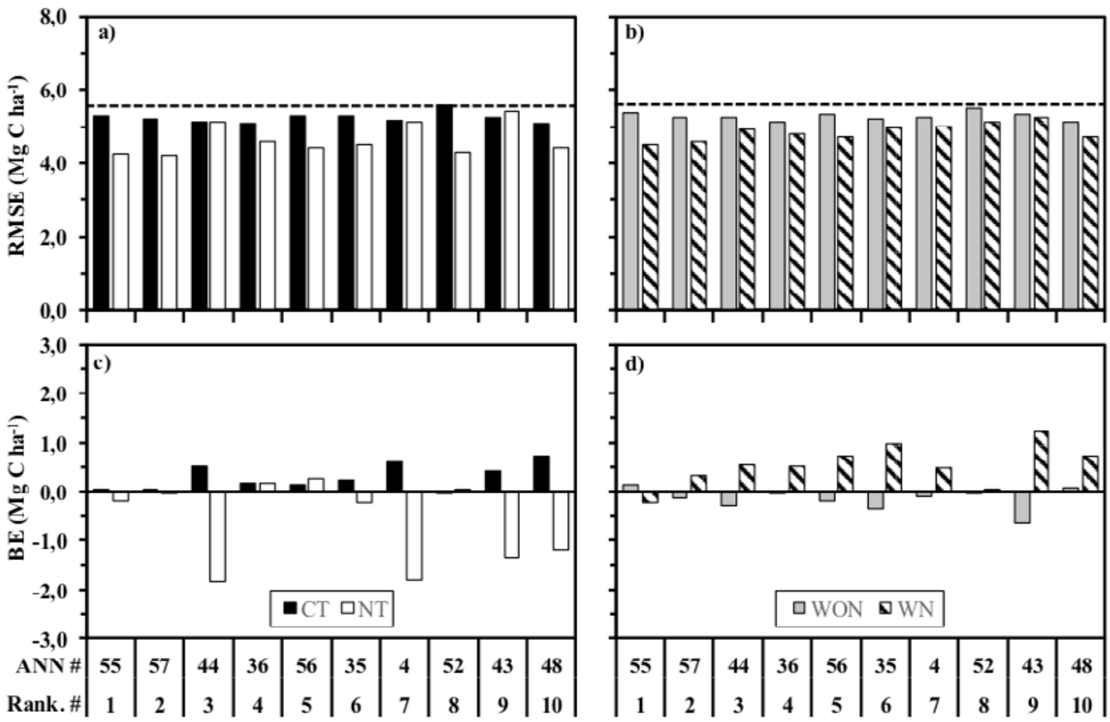

Figure 3. Statistical indicators of the ten trained artificial neural network (ANN)-based models with the best performance (**Table 8**) discriminated by management treatment: tillage system (a, c) and nitrogen fertilization (b, d). ANN #: ANN number (**Tables 2, 3, 4**); Rank. #: ranking position according to **Table 8**; RMSE (a, b): root mean square error; BE (c, d): bias error. WON: treatments without nitrogen fertilization; WN: treatments with nitrogen fertilization; CT: treatment under conventional tillage; NT: treatments under no tillage. Dashed line in plots a) and b) indicate the general mean of RMSE (for the 57 ANN trained).

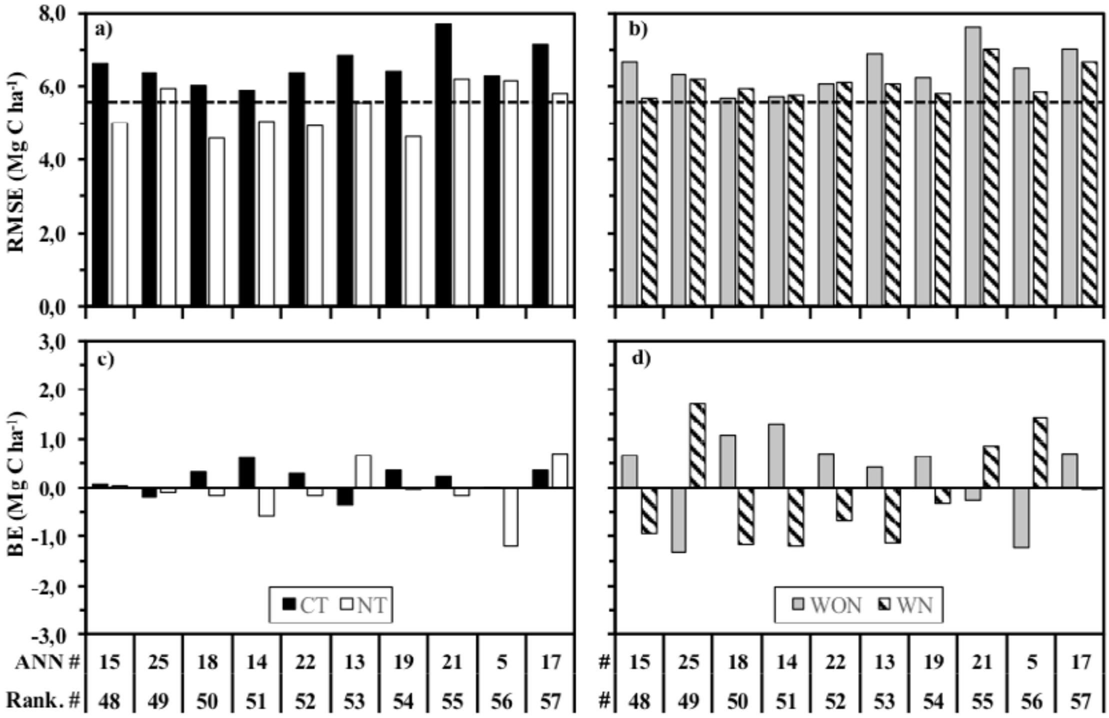

Figure 4. Statistical indicators of the ten trained artificial neural network (ANN)-based models with the worst performance (**Table 8**) discriminated by management treatment: tillage system (a, c) and nitrogen fertilization (b, d). ANN #: ANN number (**Tables 2, 3, 4**); Rank. #: ranking position according to **Table 8**; RMSE (a, b): root mean square error; BE (c, d): bias error. WON: treatments without nitrogen fertilization; WN: treatments with nitrogen fertilization; CT: treatment under conventional tillage; NT: treatments under no tillage. Dashed line in plots a) and b) indicate the general mean of RMSE (for the 57 ANN trained).

The ANN-based model with the best performance (ANN 55, **Table 4, Figure 3**) was developed based on all (five) management variables combined with Tmin (**Table 4**). On the other hand, the worst performance was achieved by the ANN-based model with only two management variables (SOCI and Tillage) (ANN 17, **Tables 4, 8, Figure 4**). The differences in statistical indicators between the best and the worst ANN-based models (**Figure 2**) were of 1.8 and 0.44 Mg C ha^{-1}, and 1.9 percent points in RMSE, BE and BRE, respectively. Taking into account the complexity of the processes and interactions involved in SOC formation and degradation in relation to soil use, those differences can be considered negligible (Smith et al. 1997). However, even though small, the improvement of an ANN-based model performance including six management variables and one meteorological one (ANN 55, **Table 4**), could be assumed as better representing the factors that define surface SOC dynamics in Mollisols of the SEBA. Besides, the input variables used by ANN 55 (**Table 4**) do not mean additional complication for potential users since they are easily available everywhere.

The distribution of simulated and observed SOC stock over time, is a visual tool that can help to interpret model performance. **Figure 5** shows the evolution of observed SOC stock values and those estimated with the best (ANN 55, **Tables 4, 8**) (**Figure 5a**) and the worst (ANN 17, **Tables 2, 8**) (**Figure 5b**) ANN-based models. **Figure 6** shows the evolution of both observed and estimated with the best and worst models SOC stock values discriminated by tillage system and nitrogen fertilization levels. Whichever the fertilization treatment, both ANN (the best (ANN 55) and the worst (ANN 17) showed better performance over time under NT than under CT (**Table 9**).

Soil organic C stocks estimated with ANN

55 (**Table 4**) showed the best match with observed values, especially up to 18 years since the beginning of the experiments. This model showed a better estimation to the observed changes and variability of SOC stock (**Figures 3, 5a, 6**). This may be attributed to the number of input variables involved, which made ANN 55 more representative of the variables influencing SOC dynamics. Anyway, input variables in ANN 55 are very few in relation to the high number of factors driving SOC variation.

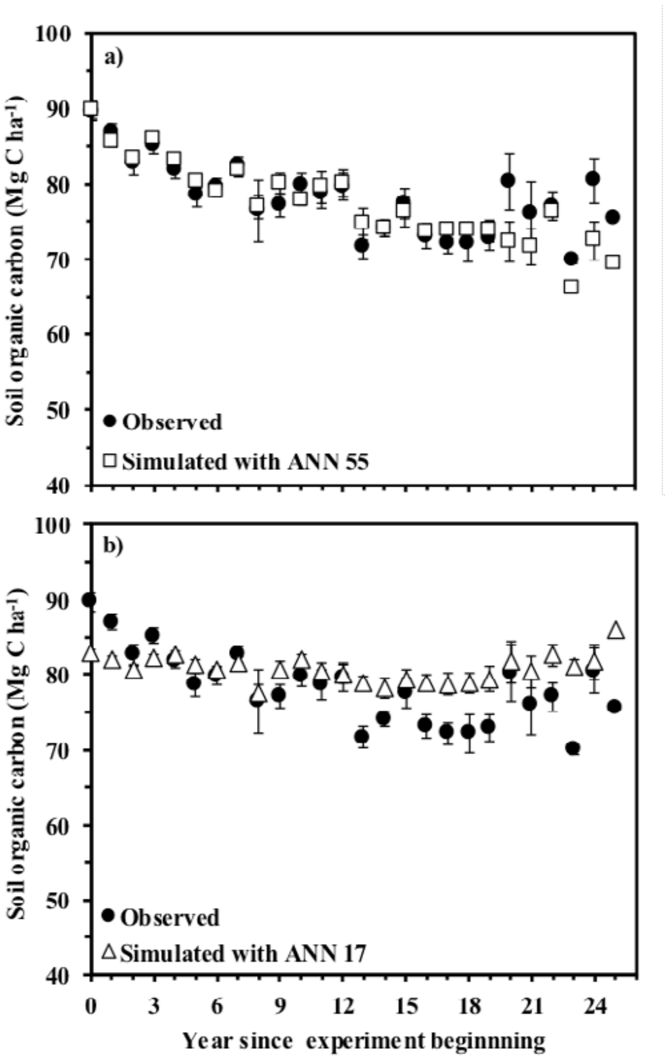

Figure 5. Evolution of observed and simulated soil organic carbon since the beginning of the experiments. a) simulation with the best trained artificial neural network (ANN) model (ANN 55, **Tables 4, 8**); b) simulation with the worst trained ANN model (ANN 17, **Tables 2, 8**). Vertical bars indicate standard error of the mean.

Other ANN models developed in Argentina to predict SOC variations based on some other input variables showed different statistical indicators than those achieved in this work. Álvarez (2008) used the average C input, silt plus clay content and air temperature as input variables and reported an RMSE of 4.7 Mg C ha^{-1} (similar to that achieved with ANN 55, **Figures 2, 3**). However, the R^2 reported by Álvarez (2008) (R^2 = 0.93) was much higher than that shown by ANN 55 (R^2 = 0.58, **Table 7**). Likewise, Álvarez et al. (2011) also developed ANN models based on crop type, average grain yield and precipitation to predict gains and losses of SOC under different cropping systems. They obtained better statistical indicators (R^2= 0.85 and RMSE = 0.63) than those obtained with our ANN 55, although lower than those reported by Alvarez (2008). On the other hand, the ANN with the worst performance (ANN 17, **Tables 2, 8**) did not match observed SOC changes over time (**Figures 5b, 6**).

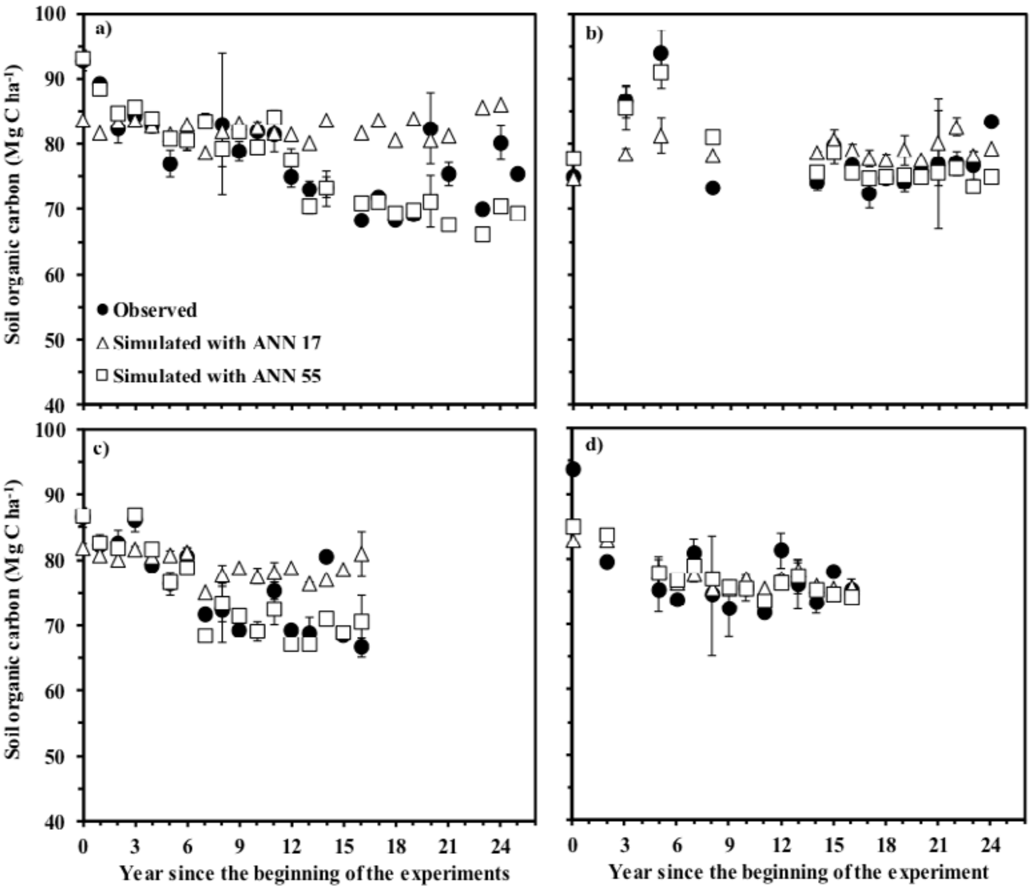

Figure 6. Evolution of observed and simulated soil organic carbon since the beginning of the experiments under different tillage and nitrogen fertilization treatments. Simulated values with the ANN showing the best (ANN 55, **Tables 4, 8**) and the worst (ANN 17, **Tables 2, 8**) performances. a) conventional tillage without nitrogen; b) no-tillage without nitrogen; c) conventional tillage with nitrogen; d) no-tillage with nitrogen.

Table 9. Statistical indicators of models ANN with the best (ANN 55, **Table 4**) and the worst (ANN 17, **Table 2**) performances. RMSE: root mean square error; BRE: bias relative error: BE: bias error; CT: conventional tillage; NT: no tillage; WN: with nitrogen, WON: without nitrogen

Number of ANN	Tillage system	Nitrogen fertilization	RMSE	BRE	BE
			Mg C ha^{-1}	%	Mg C ha^{-1}
55	CT	WN	4.09	4.47	-0.32
		WON	5.11	4.99	0,26
	NT	WN	3.90	4.37	0.01
		WON	3.66	4.15	-0.36
17	CT	WN	6.40	7.11	-0.09
		WON	6.49	6.83	0.62
	NT	WN	4.20	4.82	0.17
		WON	5.71	6.44	1.13

4. Conclusions

Artificial-neural-network-based models were adequate to estimate SOC in the upper 0.20 m of Mollisols of the SEBA. All ANN-based models trained could be used in the SEBA under different management situations. The model with the best performance (ANN 55) was developed including six management variables (SOCI, Year, Prop soybean, Tillage, Yield, Prop agri) and one meteorological variable (Tmin) as input variables, all of them easily available and with very low level of uncertainty. For our Mollisols, the composition of the ANN-based models with better performances (top average hierarchical ranking order) showed that management variables were predominant over the meteorological ones. The number of input variables used is yet recommendable and does not imply serious difficulties for users under environmental and management conditions of SEBA. However, future studies based on knowledge extraction from ANN should allow improving interpretations of these results and to support the use of the technique of ANN to develop models using simple and easily available local information.

5. Acknowledgements

The information shown in this work integrates the Master of Science Thesis (Facultad de Ciencias Agrarias, Universidad Nacional de Mar del Plata) of the first author. The first author was granted a Research Fellowship (Comisión de Investigaciones Científicas, Provincia de Buenos Aires) to go through her Graduate Program. This work was funded by the Agencia Nacional de Promoción Científica y Tecnológica (research project PICT 2012-1092), the Programa Cooperativo para el Desarrollo Tecnológico Agroalimentario y Agroindustrial del Cono Sur (research project FONTAGRO RG-T1654), the Universidad Nacional de Mar del Plata (research project AGR481/15) and the Instituto Nacional de Tecnología Agropecuaria (research project PNCYO 1127032).

REFERENCES

• Agnusdei MG, Colabelli MR, Fernández Grecco RC. 2001. Crecimiento estacional de forraje de pasturas y pastizales naturales del Sudeste Bonaerense. Boletín Técnico N° 152. Balcarce, Argentina: Estación Experimental INTA.

• Álvarez R. 2008. Comparación de técnicas de regresión, arboles de regresión y redes neuronales para estimar el contenido de carbono orgánico de los suelos. In: Proceedings of the 21th Congreso Argentino de la Ciencia del Suelo; 2008 May 13-16; Potrero de los Funes, San Luis, Argentina. In CD.

• Álvarez R, Berhongaray G, De Paepe J, Mendoza MR, Steinbach HS, Caride C, Cantet R. 2012. Productividad, fertilidad y secuestro de carbono en suelos pampeanos: efecto del uso agrícola. Anales Acad Nac Agron Vet. LXVI:381-426.

• Álvarez R, Steinbach HS, Bono A. 2011. An artificial neural network approach for predicting soil carbon budget in agroecosystems. Soil Sci Soc Am J. 75:965-975.

• Álvarez R, Steinbach HS, Bono A, Berhongaray G. 2009. Balance de carbono y nivel de materia orgánica en suelos agrícolas de la Región Pampeana. In: Proceedings of the Simposio Efectos de la Agricultura, la lechería y la ganadería en el recurso natural suelo: impactos y propuestas; 2009 August 20-21; Montevideo, Uruguay. In CD.

• Andriulo A, Mary B, Guerif J. 1999. Modelling soil carbon dynamics with various cropping sequences on the Rolling Pampas. Agronomie 19:365-377.

• Bélanger G, Gastal F, Warembourg FR. 1992. The effects of nitrogen fertilization and the growing season on carbon partitioning in a sward of tall fescue (Festuca arundinacea Schreb.). Ann Bot. 70:239-244.

• Berhongaray G, Álvarez R, De Paepe J, Caride C, Cantet R. 2013. Land use effects on soil carbon in the Argentine Pampas. Geoderma 192:97-110.

• Braga AP, Carvalho APL, Ludermir TB. 2007. Introdução. In: Braga AP, Carvalho API, Ludemir TB, editors. Redes Neurais Artificiais: Teoria e Aplicações. Rio de Janeiro, Brasil: LCT editora. p. 67-85.

• Cybenko G. 1989. Approximations by superpositions of a sigmoidal function. Math Control Signal 2:303-324.

• De Paepe JL, Álvarez R. 2013. Desarrollo de un índice de productividad de suelo regional a través de una red neuronal artificial. Informaciones Agronómicas de Hispanoamérica 12:23-26.

• Diovisalvi NV, Studdert GA, Domínguez GF, Eiza MJ. 2008. Fracciones de carbono y nitrógeno orgánicos y nitrógeno anaeróbico bajo agricultura continua con dos sistemas de labranza. Ciencia del Suelo 26:1-11.

• Eiza MJ, Fioriti N, Studdert GA, Echeverría HE. 2005. Fracciones de carbono orgánico en la capa arable: efecto de los sistemas de cultivo y de la fertilización nitrogenada. Ciencia del Suelo 23:59-68.

• Fox DG. 1981. Judging air quality model performance: a summary of the AMS workshop on dispersion model performance. Bull Am Meteorol Soc. 62:599-609.

• Franko U, Oeschligel B, Schenk S. 1995. Simulation of temperature, water, and nitrogen dynamics using the model CANDY. Ecol Model. 81:213-222.

• Hansen S, Jensen HE, Nielsen NE, Svendsen H. 1991. Simulation of nitrogen dynamics and biomass production in winter wheat using the Danish simulation model DAISY. Fert Res. 27:245-259.

• Haykin S. 2001. Redes Neurais: princípios e prática. Porto Alegre, Brasil: Bookman. 900 p.

• Ingleby HR, Crowe TG. 2001. Neural network models for predicting organic matter content in Saskatchewan soils. Can Biosyst Eng. 43:1-5.

• INTA. 1979. Carta de Suelos de la República Argentina. Buenos Aires (Argentina): Instituto Nacional de Tecnología Agropecuaria, Secretaría de Agricultura, Ganadería y Pesca.

• Jenkinson DS, Hart PBS, Rayner JH, Parry LC. 1987. Modelling the turnover of organic matter in long-term experiments. INTECOL Bull. 15:1-8.

• Kemanian AR, Stôckle CO. 2010. C-Farm: A simple model to evaluate the carbon balance of soil profiles. Europ J Agron. 32:22-29.

• Levine ER, Kimes DS. 1998. Predicting soil carbon in mollisols using neural networks. In: Lal R, Kimble JM, Follett RF, Stewart BA, editors. Soil Processes and the Carbon Cycle. Florida, USA: CRC Press LLC. p. 473-484.

• Liang, Y, Gollany H, Rickman R, Albrecht S, Follett R, Wilhelm W, Novak J, Douglas C. 2008. CQESTR simulation of management practice effects on long-term soil organic carbon. Soil Sci Soc Am J. 72:1486-1492.

• Maier HR, Dandy GC. 2000. Neural network for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ Model Softw. 15:101-124.

• Moreno R, Irigoyen AI, Studdert GA. 2014a. Aplicación de redes neuronales artificiales para estimar el carbono orgánico del suelo. In: Proceedings of the 24th Congreso Argentino de la Ciencia del Suelo and 2nd Reunión Nacional “Materia Orgánica y Sustancias Húmicas”; 2014 May 5-9; Bahía Blanca, Buenos Aires, Argentina. In CD.

• Moreno R, Irigoyen AI, Studdert GA. 2014b. Carbono orgánico del suelo: Extracción de conocimiento desde redes neuronales artificiales. In: Proceedings of the 24th Congreso Argentino de la Ciencia del Suelo and 2nd Reunión Nacional “Materia Orgánica y Sustancias Húmicas”; 2014 May 5-9; Bahía Blanca, Buenos Aires, Argentina. In CD.

• Moreno R, Studdert, GA, Monterubbianesi MG, Irigoyen AI. 2016. Soil organic carbon changes simulated with the AMG model in a high-organic-matter Mollisol. Spanish J Soil Sci 6:212-229.

• Parton W, Shimel D, Cole C, Ojima D. 1987. Analysis of factors controlling soil organic matter levels in Great Plains grasslands. Soil Sci Soc Am J. 51:1173-1179.

• Quiroga RA, Studdert GA. 2015. Manejo del suelo e intensificación agrícola: agua y materia orgánica, dos aspectos clave. In: Echeverría HE, García FO, editors. Fertilidad de suelos y fertilización de cultivos. 2nd ed. Buenos Aires (Argentina): Ediciones INTA. p. 73-100.

• R Core Team. 2015. R: A Language and Environment for Statistical Computing. Vienna (Austria): R Foundation for Statistical Computing.

• Reussi Calvo NI, Studdert GA, Calandroni MB, Diovisalvi NV, Cabria FN, Berardo A. 2014. Nitrógeno incubado en anaerobiosis y carbono orgánico en suelos agrícolas de Buenos Aires. Ciencia del Suelo 32:189-196.

• Rogers L, Dowla FU. 1994. Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling. Water Resour Res. 30:457-481.

• Sainz Rozas H, Echeverría H, Angelini H. 2011. Niveles de materia orgánica y pH en suelos agrícolas de la Región Pampeana y Extrapampeana Argentina. Informaciones Agronómicas de Hispanoamérica 2:6-12.

• Sánchez SR, Studdert GA, Echeverría HE. 1996. Descomposición de residuos de cosecha en un Argiudol típico. Ciencia del Suelo 14:63-68.

• Schlichting E, Blume HP, Stahr K. 1995. Bodenkundliches Praktikum. Hamburg, Berlin: Paul Parey. 209 p.

• Smith P, Smith J, Powlson D, McGill W, Arah J, Chertov O, Coleman K, Franko U, Frolking S, Jenkinson D. 1997. A comparison of the performance of nine soil organic matter models using datasets from seven long-term experiments. Geoderma 81:153-225.

• Soil Survey Staff. 2014. Keys to soil taxonomy. Washington (USA): USDA-Natural Resources Conservation Service.

• Somaratne G, 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.

• Statsoft 2009. Statistica Neural Module [CD-ROM] version 9.0. Tulsa, USA: Statsoft Inc.

• Studdert GA, Echeverría H. 2000. Crop rotations and nitrogen fertilization to manage soil organic carbon dynamics. Soil Sci Soc Am J. 64:1496-1503.

• Studdert GA, Echeverría H, Casanovas EM. 1997. Crop pasture rotation for sustaining the quality and productivity of a Typic Argiudoll. Soil Sci Soc Am J. 61:1466-1472.

• Studdert GA, Monterubbianesi MG, Domínguez GF. 2011. Use of RothC to simulate changes of organic carbon stock in the arable layer of a Mollisol of the Southeastern Pampas under continuous cropping. Soil Tillage Res. 117:191-200.