Neural network approach for estimating biophysical attributes during vegetative stages of potential canopies of maize in southeastern of Buenos Aires, Argentina

Andrea Inés Irigoyen[1], Carolina Maune[2], Lucas Bonelli[1,3]

[1] Unidad Integrada Balcarce: Facultad de Ciencias Agrarias UNMdP-INTA Balcarce, Ruta Nac. 226, km 73.5 (7620) Balcarce, Argentina

irigoyen.andrea@inta.gob.ar

[2] Dow Agrosciences, Ruta 8 km 264 (2720) Colón, Argentina

[3] Conicet

Abstract. Leaf Area Index (LAI) is a key input for many crop models. The LAI patterns measured *in situ* are time consuming and labor intensive and could be substituted by intelligent techniques of approximation as artificial neural networks (ANNs). The objective of this study was to evaluate the possibility to estimate the evolution of LAI and height of maize canopies in southeastern of Buenos Aires Province, Argentine using neural network models. A field experiment under non limiting condition was carried out to generate a range of environmental conditions (four planting dates and three hybrids with contrasting maturity). Periodical measurements of LAI on tagged plants were used to develop, evaluate and test the neural networks to approximate variation of leaf area index and height at plot scale. Data from canopy structure properties as leaf area, height and leaf area density profile were obtained by non-destructive methods. Planting date (PD), relative maturity of the hybrid (MR) and thermal time from emergence (TTE) were the inputs to the ANNs. A decomposition method based on Garson's algorithm was applied to quantify the relative importance for each input variable. The method provides a better description of the knowledge learned by the networks during the training process. Sensitivity analysis was performed to identify relevant variables and quantify the risk of a given combination of variables. The RM showed a major contribution in ANNs to estimate LAI and HLL. Both trained ANNs were most sensitive to TTE than the remaining inputs.

Key words: Leaf Area Index, Plant Height, Planting Date, Maturity Class

1 Introduction

Leaf Area Index (LAI) is a key input for many crop models. Direct LAI measurements are time consuming and labor intensive. Indirect methods help to estimate it, but all methods exhibit its limitations [1]. Another attribute of interest in biophysical modeling is the height of canopy, because the distribution vertical of canopy is determinant on how interacts with the environment. Rates of evapotranspiration can be estimated based on resistances of canopy [2] [3]. In maize, leaf area and vertical leaf area distribution affects radiation interception and, then, dry matter accumulation and grain yield [4].

Evolution of LAI can be described according a few variables of environment and management. In the absence of water stress, LAI is controlled by the temperature-dependent processes (leaf initiation and expansion). The rates of temperature-dependent development can be approximated by thermal functions [5] [6].

Crop development by simulation model also requires specific genetic parameters that describe genotype adaptation to environment [7] [8]. These approaches provide an estimation of response on biomass or final yield, but they are not suitable for use in modeling crop structure [7] [9] [10]. Some changes introduced more recently in DSSAT [11] have enhanced the predicted LAI. The knowledge of how vary the leaf area in different hybrids as a relative expression to maximum achievable can be of interest to assess the potential of hybrids. Some traits of canopies are maintained over different environments, when are showed in a normalized expression as relative profiles of LAI [4], [12]. This relation is mainly useful for developing new models.

The artificial neural networks (ANNs) are mathematical models similar to the biological structures, with computational capacity to solve problems of approximation, classification and optimization [13]. Simple ANNs with a vector of inputs, one or more hidden layers and following a feed-forward procedure can approximate any mathematical function. A sole hidden layer is sufficient to approximate a continuous function [14].

The use of ANN on crop modeling is entirely justified, due the non linearity in most of relationships between crop attributes (functional and physic) and environmental variables. This works specifically focuses on the estimation of canopy development which is a key determinant of the radiation interception by the crop and, then, of the energy and water exchanges.

As thermal time can be considered a good estimator of crop development, when it is combined with some indicator of management practices can give an approximation of attributes of canopy. Then, ANN models to estimate other biophysical processes linked to dynamics of leaf area (i.e., evapotranspiration or soil water) have showed satisfactory results, when thermal time was regarded as input [15].

ANNs are cited in literature as models capable to approximate attributes at leaf, individual and regional scales, but no references were found at plot scale. [16] [17] [18].

The objective was to develop ANN models to approximate leaf area index and height of plants during vegetative stages in maize canopies growing on non limiting conditions in southeastern of Buenos Aires, as affected by planting date and maturity class.

2 Material and methods

2.1 Site and experimental condition

Data used to develop ANN models were acquired from a field experiment, conducted at Balcarce Unidad Integrada (INTA-UNMdP), Balcarce, Argentina (37º 45' S; 58º 18' W) during 2011-2012 growing season, under no tillage practices. Three hybrids of different relative maturity (RM) class (DK692, RM119; Illinois 1550, RM102; Pioneer 39B77, RM89) were sowed in four dates (04-October, 14-November, 15-December and 04-January). The duration of the period exhibiting foliar expansion (from emergence to tasseling) ranged from 708 to 1038 °C day^{-1}. Non limiting conditions for the crop were performed with agronomic management (density, fertilization, weed control) and control of the soil water availability. Detailed information can be found in [12], [19].

2.2 Field data collections

At the three leaf stage (V3), five successive plants were tagged in the central row of each treatment. Leaves were tinned to identify the leaf position over the measurement period. Lamina length (L) and maximum lamina width (W) of each leaf were weekly registered and used to calculated leaf area (LA), as originally proposed by Montgomery (1911) , cited by [20], and without major modifications in actual genotypes:

$$LA = L \times W \times 0.75 \qquad \textbf{(1)}$$

The measurements started when plants had three leaves fully expanded (V3). The final value in each determination coincided with LAI at tasseling (LAIVT) of each plant and was an indicator of maximum LAI. Total plant leaf area was calculated as the sum of the areas of all individual leaves per plant, and then was expressed per unit of land occupied by the plants (LAI). The media density was 9.8 plants m^{-2}. The final number of leaves ranged from 17 to 22.

Crop development was recorded according [21]. Other biometric attributes as height of insertion of the ligule of last expanded leaf (HLL), total height, total number of leaves and number of fully expanded leaves were recorded. Canopy height was represented in this work by HLL, due this attribute was best correlated with LAI [12] than total height. Average values of the five plants per plot were used to represent the attributes of the canopy.

2. 3 Development of ANN models

The variables used as inputs to approximate LAI and HLL target values were: a) planting date, PD (expressed as number day of the year calendar), b) relative maturity class, RM (as indicator of duration of crop growing season) [22] and c) thermal time from emergence, TTE (accumulated residues from daily mean air temperatures minus 8°C) [7]. The selection of inputs was based on: a) the readily availability of management data, b) the relation of maturity class or planting dates and final number of leaves in the plant [23], [24], [25] and c) the simplicity on estimation of crop development and growth by thermal time [5], [7], [26]. Limitations of the method are given by [27]. In particular for maize, different relations have been recently discussed [28].

Activations functions for the hidden units are needed to introduce non-linear components. In this study, two types of transformed of 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, in general, allows a network to map a nonlinear process.

Training of a ANN with the above topology was achieved by adjusting the weights of the neurons through an iterative algorithm that minimizes the error between the predicted outputs and the actual data. The training was carried out under the Broyden-Fletcher-Goldfarb-Shanno (BFGS) or Quasi-Newton, a powerful second order algorithm, using the measured values of IAF and HLL as target outputs in the ANNS. Each ANNs architecture was trained under automatic early stopping criterion associated to cross validation method. For this reason, the data set was split into training, validation and test groups to apply cross-validation. Training, testing, and validation sets were representative of the same population. Total data ($n = 503$ for LLH and $n = 310$ for LAI) were split in three groups in the proportion 60:20:20 (training, test and validation groups, respectively). In order to evaluate the hypotheses of equality of frequencies distribution between values of training set with test and validation ones, respectively the non parametrical Kolmogorov-Smirnov test was applied ($p < 0.05$).

The selection of ANNs architectures was based on the application of a selected algorithm integrated on the IPS (Intelligent Problem Solver) of the Neural Network module of Statistica Software [29]. 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 training. The model with the lowest cross validation error was chosen and then, the ANN with best performance for each combination was retained and evaluated.

Based on the cross validation, once the best architecture for each ANN model was identified the selected models were subjected to further evaluation for their effectiveness in estimating the target values. The description of each model was according: a) inputs b) sequence n-m-x, where n is input number, m is number of hidden layers and x is number of outputs and c) activation function.

Table 1. Descriptive of data sets used for development and evaluation of ANN models to approximate leaf area index (LAI) and height (HLL) of maize canopies.

		Inputs			Targets	
Set	Descriptive	RM	PD	TTE (∘ day^{-1})	LAI (m^2 m^{-2})	HLL (m)
Train	Minimum	89	4	96.2	0.07	0.03
	Maximum	119	349	948.4	6.67	2.28
	Mean				2.84	0.79
Validation	Minimum	89	4	96.2	0.08	0.04
	Maximum	119	349	1038.9	6.97	2.43
	Mean				2.85	0.80
Test	Minimum	89	4	96.2	0.07	0.04
	Maximum	119	349	948.4	6.97	2.33
	Mean				3.44	0.67

RM: relative maturity of maize hybrid, PD: date of planting following day of calendar.

2.4 Evaluation of ANN models

The model performance was carried out (from the test set) by correlation measures (regression analysis) and three difference based errors: root-mean-square error (RMSE), mean bias error (MBE), mean absolute error (MAE). The Student test was used to statistically evaluate the value of either the intercept (H0: a=0) or slope of the straight line (H0: b=1) at the 5% probability level. More information about the complementary of errors is given in [30], [31], [32], [33]. All indicators were computed for LAI and HLL ANN models.

2.5 Extraction of knowledge from ANN models

Once the ANNs were trained on a specific network topology, then the modeling of attributes process using ANN involved the extracting knowledge from each network. The embedded knowledge is in the form of connection weights. Garson's method [34] was performed from adjusted synaptic weights of each ANNs. The contribution of each input neuron to the output (c_{ijo}) was computed via each hidden neuron as the product of the input-hidden connection (w_{ij}) and the hidden-output connection (w_{jo}):

$$\mathrm{c_{ijo}} = w_{ij} \times w_{jo} \quad \textbf{(2)}$$

The relative contribution of each input *k* to hidden neuron *j* can be expressed as:

$$r_{ijo} = \frac{|\mathrm{c_{ijo}}|}{\sum_{K=1}^{m} |\mathrm{c_{kjo}}|} \quad \textbf{(3)}$$

The total contribution of input *i* is:

$$S_i = \sum_{j=1}^{n} r_{ijo} \quad \textbf{(4)}$$

Finally, the relative contribution of each input is:

$$RI \quad = S_i / \sum_{k=1}^{m} Sk \quad \textbf{(5)}$$

The Sensitivity Analysis (SA) possibilities understand the influence of each input to the model [35]. The SA was performed following procedures from Neural Network module from STATSOFT program [29] test how the ANN responses and, hence, the error would increase or decrease of each of its inputs variable were to undergo a change.

3 Results and discussion

The ANN structures showed in Table 2 exhibited the smallest errors during the cross validation. LAI values estimated by the ANN were close to field measurements (Fig. 1). The MAE represents 6% of mean observed values. Errors are smaller than reported from literature [10] for models of wide diffusion for cool environments. HLL values were also well performed by the ANN (Fig. 2). In this case, the MAE represents 7% of mean observed values (Table 3). A strong correlation between LAI and HLL has been discussed for each maturity hybrid, whichever the planting date under non limiting conditions [12], therefore similar results with models based in the same inputs were expected.

Table 2. Characteristics of ANNs trained to approximate leaf area index (LAI) and height (HLL) of potential maize canopies.

Attributes	Inputs	ANN Structure	Activation in hidden layer	Number of free parameters
LAI	RM PD TTE	MLP 3-5-1	Logistic	26
LLH	RM PD TTE	MLP 3-6-1	Hyperbolic tangent	31

MLP: multilayer perceptron

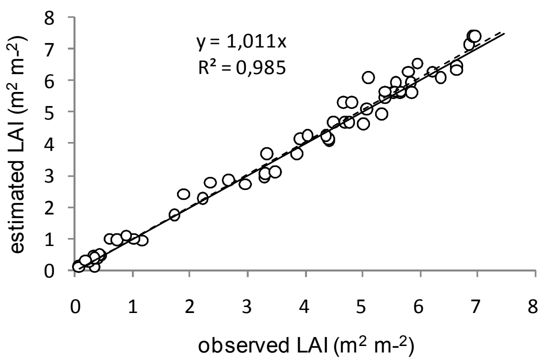

Fig.1. Scatterplots of observed and estimated values of leaf area index (LAI) in test set (n=62). The dashed line is the regression line and the entire line represents the 1:1 relation for observed values.

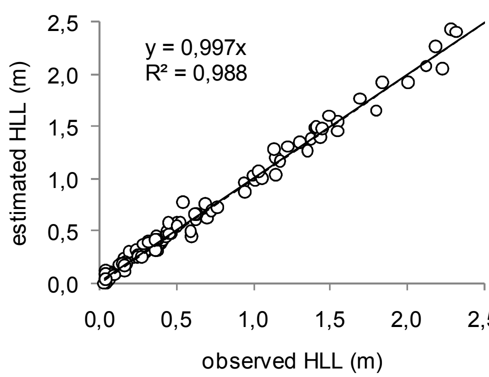

Fig. 2. Scatterplots of observed and estimated values of height of canopy represented by the height of insertion of last ligulated leaf (HLL) in test set (n=100). The dashed line is the regression line and the entire line represents the 1:1 relation for observed values.

Table 3. Errors of estimations of attributes of maize canopies by the proposed ANNs computed on the test set.

ANN for Attributes	RMSE	MBE	MAE
LAI	0.2858	0.0379	0.2163
LLH	0.0655	-0.0002	0.0499

(Error units: m^2 m^{-2} for LAI errors and m for HLL errors)

The analysis of synaptic weights in ANN trained to approximate LAI showed the highest relative contribution relative of maturity class (RM) respect the remaining variables in the trained ANN (Table 4). Also the contribution of RM to the ANN trained to estimate HLL was the highest of the inputs.

On the other hand, sensitivity to each input was significantly different, with TTE as major indicator of development, and similar for both LAI and HLL models (Table 5). It is reasonably given the nature of a model representing evolution of attributes over time.

Table 4. Relative contribution of inputs (RI) to neural network models to approximate leaf area index (LAI) and height of insertion of last liguled leaf (LLH).

	RI to trained ANN	
Inputs	LAI	HLL
RM	0.67	0.49
PD	0.22	0.23
TTE	0.11	0.28

Table 5. Relative sensitivity of ANN models to approximate leaf area index (LAI) and height of insertion of last liguled leaf (LLH) to inputs.

	Sensitivity (%) to inputs of trained ANN	
Inputs	LAI	HLL
RM	0.02	0.03
PD	0.26	0.29
TTE	0.72	0.68

The ANN model can be applied for a range of management practices, due the information about hybrids and planting dates is input as numerical variables. The proposed simplification lies in determining the appearance and the expansion of leaves as only a function of air temperature.

4 Conclusions and future work

In the present study, two attributes of maize canopies growing under potential conditions were approximated by artificial neural networks (ANNs) over a range of environmental conditions generated by planting date and length of crop season. In general, the ANN model using readily available variables provided accurate and reliable leaf area index (LAI) and height (HLL) predictions, whichever planting date or maturity class.

We conclude that the developed ANN models has a great ability to learn and build up a neural system for crop growth prediction, and the results provide a useful guidance for yield estimation or crop water use. Furthermore, a model capable of generalization as ANN represents a support from ground measurements to generate LAI products at the regional scale.

The models developed in this study represent the canopy expansion under non limiting conditions for the crop. Improvements in the development of ANN models is in progress to predict the growth and duration of active canopy (green leaf area) of maize under different crop management practices (plant density and soil water availability) in the southeastern of Buenos Aires region. Future work will concentrates on the evolution of the green canopy over the full growing season.

Acknowledgements

The authors are grateful to Universidad Nacional de Mar del Plata and Instituto Nacional de Tecnología Agropecuaria (INTA) for financial support. Data used in this work were obtained as part of the graduation work of C. Maune. We thank to G. Eriksen for his valuable help on data collection.

References

1. Weiss, M., Baret, F., Smith, G.J., Jonckheere, I., Coppin, P.: Review of methods for in situ leaf area index (LAI) determination: Part II. Estimation of LAI, errors and sampling. Agricultural and Forest Meteorology 121 (2004) 37–53
2. Monteith, J.L.: Evaporation and environment. In: "The state and movement of water in living organisms". XIXth Symposium Society for Experimental Biology, Swansea, Cambridge University Press. (1965) 205–234
3. Shuttleworth, W., Wallace, J.: Evaporation from sparse crops: an energy combination theory. Quarterly Journal of the Royal Meteorological Society 111 (1985) 1143–1162
4. Valentinuz, O.R.; Tollenaar, M.: Effect of genotype, nitrogen, plant density and row spacing on the area-per-leaf profile in maize. Agron. J. 98 (2006) 94–99
5. Ritchie, J. T.; NeSmith, D. S.: Temperature and crop development. In: Modeling plant and soil systems. Agronomy Monograph, 31, American Society of Agronomy, Madison, Wisc. USA (1991): 5-29
6. Stewart, D. W; Dwyer, L. M.: Appearance time, expansion rate and expansion duration for leaves of field-grown maize (Zea mays L.). Can J. Plant Sci. 74 (1994): 31–36
7. Birch, C.J.; Vos, J.; Van Der Putten, P.E.L.: Plant development and leaf area production in contrasting cultivars of maize grown in a cool temperate environment in the field. European Journal of Agronomy. 19 (2003) 173–188
8. Lizazo, J.T.; Batchelor, W.D.; Wesgate, M. E.: A leaf area model to simulate cultivar-specific expansion and senescence of maize leaves. Field Crops Research 80 (2003) 1–17
9. Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor,W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J.,Ritchie, J.T.: The DSSAT cropping system model. Eur. J.Agron. 18 (2003) 235–265
10. Lopez-Cedrón, F.J., Boote, K.J., Ruíz-Nogueira, B. Sau, F.: Testing CERES-Maize versions to estimate maize production in a cool environment. Europ. J. Agronomy 23 (2005) 89–102
11. Lizaso, J. I., K. J. Boote, J. W. Jones, C. H. Porter, L. Echarte, M. E. Westgate, Sonohat, G.: CSM-IXIM: A New Maize Simulation Model for DSSAT Version 4.5 Agron. J. 103 (2011) 766–779
12. Maune, C.: Fechas de siembra y desarrollo del área foliar pre-antesis en híbridos de maíz de diferente madurez relativa. Tesis de graduación, Fac. Cs. Agrarias UNMdP, Balcarce, Argentina (2014) 46 p.
13. Haykin, S. Redes Neurais: princípios e prática. Trad. M. Engel. Porto Alegre. Bookman (2001)
14. Cybenko, G.: Approximations by superpositions of a sigmoidal function. Mathematical of Control, Signal and Systems 2 (1989) 303–324
15. Irigoyen, A. I., Della Maggiora, A.I.: Knowledge extraction from artificial neural networks: A case study on total soil water during maize crop season. In: De Campos, A., Gabriel, D. Luján, D.L. (ed.) Impacts of Agrosystem on the Environment, Unesco Chair of Eremology, Ghent University and International Centre of Eremology, Belgium. (2015) 29–34
16. Odabas, M.S. Ergun, E., Oner, F.: Artificial neural network approach for the prediction of the corn (Zea mays L.) leaf area. Bulgarian Journal of Agricultural Science 19 (2013) 766–769

17. Verger, A., Baret, F., Camacho, F.: Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: Evaluation over an agricultural area with CHRIS/PROBA observations. Remote Sens. Environ. 115 (2011) 415–426
18. Yin, G., Li, J. Liu, Q. Fan, W. Xu, B. Zeng, Y. Zhao, J.: Regional leaf area index retrieval based on remote Sensing: the role of radiative transfer model selection. Remote Sens. 7 (2015) 4604-4625
19. Bonelli, L.E.: Rendimiento potencial de maíz en Balcarce en función de la fecha de siembra y la duración del ciclo del hibrido. Tesis Magister Scientiae. Facultad de Ciencias Agrarias, Universidad Nacional de Mar del Plata. Balcarce, Buenos Aires, Argentina (2014) 86 p
20. Stewart, D.W., Dwyer, L.M.: Mathematical characterization of leaf shape and area of maize hybrids. Crop Science 39 (1999) 422–427
21. Ritchie, S. W., Hanway, J. H., Benson, G. O.: How a corn plant develops. Ames: Coop. Extension Service. Special Report. (1997) 21p
22. Dwyer, L.M., Stewart, D.W., Carrigan, L., M, B.L., Neave, P., Balchin, D.: Guidelines for comparisons among different maize maturity rating systems. Agron. J. 91 (1999) 946–949
23. Cirilo, A. G.: Desarrollo, crecimiento y partición de material seca en cultivos de maíz sembrados en diferentes fechas. Tesis Magister Scientiae, Fac. Cs. Agrarias. Universidad Nacional de Mar del Plata, Balcarce, Buenos Aires, Argentina (1994) 86 p
24. Andrade, F.H., Cirilo, A., Uhart, S., Otegui, M.: Ecofisiologia del cultivo de maíz. Editorial Médica Panamericana. 292 p (1996)
25. Capristo, P.R., Rizzalli, R.H., Andrade, F.H.: Ecophysiological yield components of maize hybrids with contrasting maturity. Agronomy Journal 99 (2007) 1111–1118
26. Ben Nouna, B., Katerji, N., Mastrorilli, M.: Using the CERES-Maize model in a semi-arid Mediterranean environment. New modelling of leaf area and water stress functions. Europ. J. Agronomy 19 (2003) 115–123
27. Bonhomme, R.: Bases and limits to using 'degree.day' units. European Journal of Agronomy 13 (2000) 1–10
28. Kumudini, S., Andrade, F.H., Boote, K.J., Brown, G.A., Dzotsi, K. A., Edmeades, G. O., Gocken, T., Goodwin, M., Halter, A. L., Hammer, G. L., Hatfiled, J. L., Jones, J.W., Kemanian, A. R., Kim, S.H., Kiniry, J., Lizazo, J.I., Nendel, C., Nielsen, R. L., Parent, B., Stockle, C.O., Tardieu, F., Thomison, P.R., Timlin, D.J., Vyn, T.J., Wallach, D., Yang, H.S., Tollennaar, M.: Predicting maize phenology: intercomparison of functions for developmental response to temperature. Agronomy Journal 106 (2014) 2087–2097
29. Statsoft. Statistica Neural Networks Module (computer program). Vers. 9. Statsoft, Inc. Tulsa, 2009. 1 CD-ROM (2009)
30. Willmott, C.J.: Some comments on the evaluation of model performance. Bulletin of American Meteorology Society (1982) 1309-1013
31. Willmott, C. and Matsuura, K.: Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in assessing average model performance, Clim. Res. 30 (2005) 79–82
32. Willmott, C. J., Matsuura, K., and Robeson, S. M.: Ambiguities inherent in sums-of-squares-based error statistics, Atmos. Environ. 43 (2009) 749–752
33. Chai, T.; Draxler, R.R.: Root mean square error (RMSE) or mean absolute error (MAE)? –Arguments against avoiding RMSE in the literature. Geosci. Model Dev., 7 (2014) 1247–1250
34. Garson, G.D.: Interpreting neural network connection weights. Artificial Intelligence Expert 6 (1991) 47–51
35. Saltelli, A.: Sensitivity analysis for importance assessment. Risk Analysis 22 (2002) 579–590