Australian Journal of Crop Science AJCS
AJCS 4(5):372-377 (2010) ISSN:1835-2707

Assessment of agricultural mechanization status of potato production by means of Artificial Neural Network model

Morteza Zangeneh, Mahmoud Omid* and Asadollah Akram

Department of Agricultural Machinery Engineering, Faculty of Agricultural Engineering and Technology, School of Agriculture & Natural Resources, University of Tehran, Karaj, Iran

***Corresponding author: omid@ut.ac.ir**

Abstract

An artificial neural network (ANN) model was developed to assess mechanization status of potato farms in Iran. Mechanization index (MI) and level of mechanization (LOM) were used to characterize farming system of potato production in the region. To develop ANN model, data were obtained from farmers, government officials as well as from relevant databases. A wide range of explanatory parameters of farming activities were examined. Finally, 19 explanatory parameters were used as input variables to predict MI and LOM. Based on performance measures, single hidden layers with 8 and 3 neurons in the hidden layer were finally selected as the best configuration for predicting MI and LOM, respectively. For the optimal ANN models, the values of the model's outputs correlated well with actual outputs, with coefficient of determination (R^2) of 0.98 and 0.99 for MI and LOM, respectively. Sensitivity analyses were also conducted to investigate the effects of various explanatory parameters on the outputs. Since the ANN model can predict the two mechanization indicators for a target farming system with high accuracy, it could be a good alternative to regression for assessing agricultural mechanization of regional farms with similar conditions.

Keywords*:* Potato, Farming system, Neural networks, Assessment, Mechanization index, Level of mechanization.

Introduction

Better knowledge of the past and present is a key component for the improvement of the planning process that will impact Iran's agricultural sector in the years to come. Findings from this and similar studies can be used to set new directions for the analysis of the technological status of Iran's agriculture. In this paper special attention is paid to mechanical inputs. In order to maximize the efficiency of introducing agricultural technology to farms in a target region, the farming system of the region should be first characterized, especially to identify possible resource constraints and to capture the diversity of farming systems (Sims, 1987, Collado and Calderón, 2000, and Oida, 2000) Monitoring the mechanization status of target region, in combination with other agronomic indicators such as productivity potential, would lead to a better assessment of the sustainability of the farming system (Garcia et al., 2005). Singh, 2002 stated that growth of crop production depends on the three sources: arable land expansion, increase in cropping intensity and yield growth. Okurut and Odogola, 1999 reported that besides land, farm power is the second most important input to agricultural production. Warkentin, 1991 stated that water use efficiency is highly concern of crop management. Barton, 1999 reported that farm power determines the scale and intensity of farm operation. Sarker, 1999 reported that adoption of power tillers for tilling has brought some significant changes on overall production and sustainability of small farm systems. Borlaug and Dowswell, 1993 stated that crop production environment including the generation and transfer of appropriate technology must be improved to increase the fertilizer use efficiency to meet the challenge of feeding increased population. Mechanization is a concept and cannot be measured directly. Appropriate indicators must be selected to determine levels of mechanization. An indicator of mechanization is a variable that allows describing and monitoring the processes, states and tendencies of systems at the farm, regional, national or worldwide levels. Evaluation of the performance of agricultural production system has been made with a set of evaluating indicators/parameters. The parameters include machinery energy ratio (MER), mechanization index (MI), level of mechanization (LOM), productivity level of consumed power (PLCP), energy input to produce per unit energy output (EIO), energy input to generate per unit GDP output, energetic efficiency of the system (EE), etc. A part of the present study was the formulation of an index to measure the mechanization status achieved at the farm level. To assess mechanization status of a farming system some of these indicators may be used to characterize the production of various agro-products in a region. Here, MI and LOM are chosen to investigate farming system of potato production in the Hamadan province of Iran.

Table 1. Socio-economic structure of potato farms (per farm)

Item	Value
Population (person)	5.2
Age of farmer (year)	42.56
Total area (ha)	101.71
Potato area (ha)	23.06
Number of potato plots	1.86
Number of crop planted	4.12
Maximum yield (tonnes ha^{-1})	62.95
Tractor ownership (number)	2.10

The two indicators (MI and LOM) were defined because they would allow us to identify which farming systems in the region would benefit most from mechanization and to estimate the intensity of mechanization as part of an agricultural modernization program. LOM is based on the premise that a mechanized farmer is the one that finds a way to utilize amounts of mechanical energy that are higher than the typical values using locally available technology. The second index (MI) elaborated here is an expression of the deviation of the actual amount of motorized farm work from the normal values at the regional level. This index is based on the premise that a mechanized farmer is the one that finds a way to utilize amounts of mechanical energy that are higher than the typical values using locally available technology. It is implied that these technologies are mechanized agricultural practices that have been successfully incorporated into the farming systems. The Artificial Neural Network (ANN) gives estimations of the mechanization indicators using limited data available from the target region, without the need to calculate them directly, which would require more data. The model is based on statistical analyses of actual data, and enables us to distinguish between necessary and unnecessary items of raw data. A fundamental hypothesis of this study is being feasible to train an ANN model to establish a non-explicit function, which corresponds to the ANN network itself, between a selected set of simple inputs, such as farm size, and number of tractors owned, and two mechanization indicators as the outputs. Recently, Zangeneh et al., 2010 have investigated energy use pattern in potato production in Hamadan province of Iran. The population investigated was divided into two strata based on tractor and farm machinery ownership and level of farming technology. Group I farmers were owners of agricultural machinery and practiced high level of farming technology, whereas Group II were non-owners of machinery and exercised low level of farming technology. Total energy consumption of Group I and group II was 157.151 and 153.071 GJ ha^{-1}, respectively which we studied Group I in present paper. The amount of nonrenewable energy (NRE) in both groups was rather high. They concluded a reduction in the total NRE ratio, specifically in chemical fertilizer usage would have positive effects on the sustainability of potato production as well as other positive environmental effects. In this sequel, we develop ANN models to predict mechanization indicators for potato production based on energy and power consumption. The potential practical application of this work consists on mapping the proposed mechanization indicators for a much wider area without direct calculations. Further analysis based on the interrelation between the produced data with complementary parameters already available in local databases, would contribute to assess the mechanization status in the region.

Material and methods

Data source and processing

The study was carried out on potato farms in Hamadan province, Iran. Data were collected from the farmers by using a face to face questionnaire method. The additional materials used in this study were collected from the previous studies and publications by some institutions like FAO. The sample size was determined using Cochran technique as 68 farms (Snedecor and Cochran, 1989). The original data set was consisted of 300 explanatory parameters for each farm covering all characteristics of farming system of potato in the region. To assess the technological status and the agricultural production strategies, the farming system was analyzed according to its energy input-output flow and consumption of power for various farming operations.

Socio-economic structures of farms

Socio-economic structure of studied farms is shown in Table 1. According to the results of this table, maximum yield was 62.95 tonnes ha^{-1}. Average farm size was 23.06 ha, and potato production occupied 22.67% of total farm lands. The other vegetables grown besides potato were wheat, alfalfa, corn and barley.

Input and output parameters

Based on data availability and how representative they were of all the data, a set of input items, including 94 explanatory parameters, were chosen as the first candidate set of the input items for all outputs. Underlying distributions were uniform, representativeness was checked. Using a regression method (forward method) different collection of input items selected for two mechanization indicators as outputs. Because the items have different scales, the data were normalized by converting them using natural logarithm to maintain the neural network sensitivity as per Drummond et al., 2003, Abdullakasim et al., 2005, Zhang et al., 1998. Table 2 shows the selected parameters fed into the ANN model during the training process. These items represent key factors of the farming system and were identified as factors in the mechanization status. They produced superior performance during the teaching process.

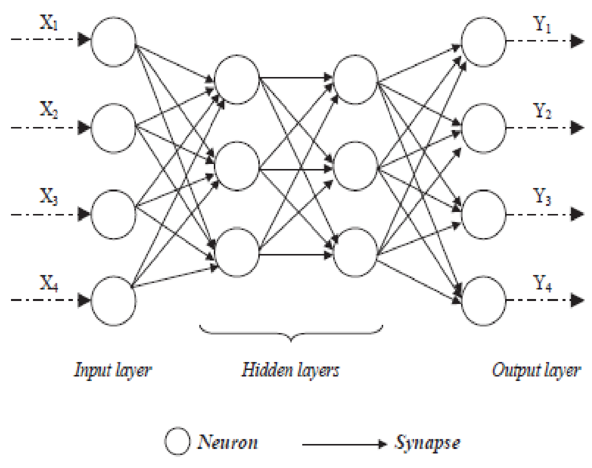

Fig 1. The structure of multilayer feed forward neural network

Table 2. Selected input items for characterizing LOM and MI of farming system of potato

Input items	Outputs	
	LOM	MI
1	Area under potato cultivation (ha)	Experience of potato production (year)
2	Number of potato plots	Total land size (ha)
3	Total power of tractors (hp)	Area under potato cultivation (ha)
4	Total number of tractors	Number of land preparation operations
5	Number of land preparation operations	Required time for hand collecting and bagging potatoes
6	Required time for hand collecting and bagging potatoes per hectare	Cost of fixed labor per year ($)
7	Cost of fixed labor per year ($)	Number of non-subsidized potassium fertilizer (50 kg bag)
8	Number of non-subsidized nitrogen fertilizer (50 kg bag)	Amount of insecticides (L ha^{-1})
9	Amount of insecticides (L ha^{-1})	Number of irrigation
10	Number of irrigation	Total hours of farm machinery work (h ha^{-1})
11	Sum of fixed costs ($ ha^{-1})	Cost of seed ($ ha^{-1})
12	-	Equal energy of seed (MJ ha^{-1})

Definitions of mechanization indices

Based on the general concept of mechanization (Sims, 1987) and the structure of ANN models, two mechanization indicators, namely, Mechanization Index (MI) and Level of Mechanization (LOM), were chosen to characterize farming system of potato in the target region. These indicators are defined mathematically as equations (1) and (2). The MI elaborated here is an expression of the deviation of the actual amount of motorized farm work from the normal values at the regional level.

$$MI = \frac{1}{n}\sum_{i=1}^{n}\frac{M_{e(i)}}{M_{av}}.\frac{L_i}{TL_i} \qquad (1)$$

where: MI = Mechanization Index for the production unit \`a\`, $M_{e(i)}$ = Overall input energy due to machinery in the production unit \`a\`, M_{av} = Regional-average energy due to machinery, L_i = Land area cultivated in the production unit \`a\`, TL_i = Total farm land ownership of production unit \`a\`, n = Number of farms . The MI index, proposed by Andrade and Jenkins, 2003 is an indication of the amount of machinery a given farmer uses for farm work compared with the average in the region. The second term in Eq. (1) includes a ratio between the land area cultivated with potato crop and the total land ownership. This term was introduced because it reflects the importance of land demand for cultivation. The LOM index is based on the premise that a mechanized farmer is the one that finds a way to utilize amounts of mechanical energy that are higher than the typical values using locally available technology.

$$LOM = \sum_{i=1}^{n}\frac{P_i \times \eta}{L_i} \qquad (2)$$

where: LOM = level of mechanization, P_i= power of tractors , η = correction factor for utilized power (0.75). Field capacity was multiplied by rated power so the quantification of energy expenditure was made in work units (kWh). The regional normal was obtained after compiling a full dataset of all respondents and then it was defined the mode for the number of passes for each operation as well as the mode in tractor size and field capacity. It is implied that these technologies are mechan-

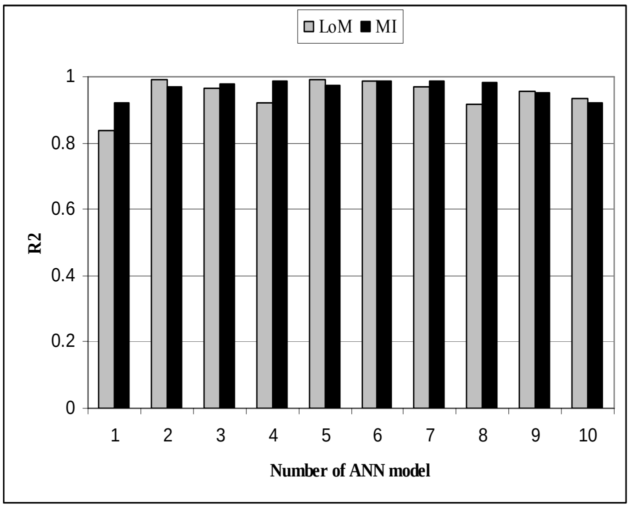

Fig 2. Correlation between the ANN model's outputs and calculated outputs

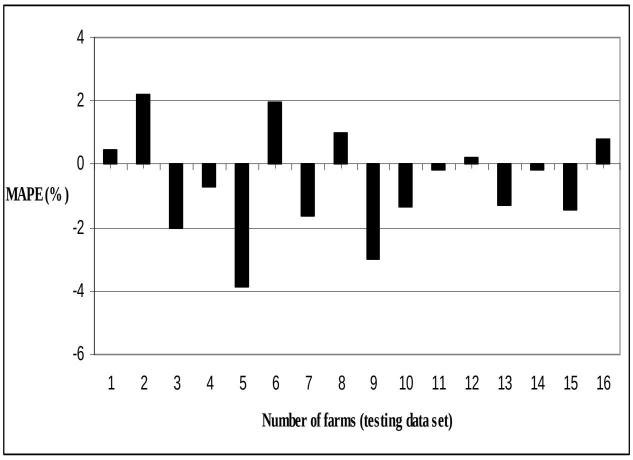

Fig 3. MAPE in the MI estimation over the test set

Table 3. Alternative configuration of ANN models for LOM and MI

Outputs	# ANN model	# Hidden layer	# Neurons	MSE	MAE	MAPE	R^2
LOM	1	1	2	0.0792	0.1467	8.4137	0.839
	2	**1**	**3**	**0.0234**	**0.0692**	**1.8151**	**0.992**
	3	1	4	0.0202	0.0735	6.0146	0.965
	4	1	5	0.0256	0.1300	6.6020	0.923
	5	1	6	0.0240	0.0772	2.1966	0.990
	6	1	7	0.0040	0.0536	2.7926	0.985
	7	1	10	0.0068	0.0705	3.2190	0.969
	8	1	12	0.0357	0.1558	9.8229	0.915
	9	2	4-4	0.0250	0.0917	6.9694	0.954
	10	2	4-8	0.0317	0.0994	6.9999	0.936
MI	1	1	2	0.0655	0.1808	3.0602	0.921
	2	1	3	0.0246	0.1291	2.5296	0.970
	3	1	4	0.0953	0.2366	3.7092	0.978
	4	1	5	0.0168	0.1122	2.2225	0.988
	5	1	6	0.0532	0.1625	2.8564	0.974
	6	**1**	**8**	**0.0093**	**0.0733**	**1.3859**	**0.989**
	7	1	10	0.0166	0.0821	1.4996	0.988
	8	1	12	0.0190	0.1167	2.3176	0.984
	9	2	4-5	0.0450	0.1749	3.2799	0.951
	10	2	8-8	0.1095	0.1863	3.0602	0.921

ized agricultural practices that have been successfully incorporated into the farming systems. During the interview, data was recorded on all the mechanized operations performed by farmers in the sample providing an estimation of the field capacity (hours of work per unit land). The ANN models were trained to output these indicators from the data of the 19 input parameters, included in Table 2.

Multilayer Feed forward ANN

There are multitudes of ANN structures and different classification frameworks. For examples, ANN could be classified according to the learning method or to the organization of the neurons (Chester, 1993). The one that have been used in this work is called Multi Layer Perception (MLP), in which neurons are organized in several layers: the first is the input layer (fed by a pattern of data), while the last is the output layer (which provides the answer to the presented pattern). Between input and output layers there could be several other hidden layers (see Fig. 1). The number of hidden layers has an important role in determining the generalization ability of the MLP. MLP represents a tool, which is able to identify the relationships between different data sets, although the form of these relationships is not defined exactly. For this reason they are called ''universal approximation or regression tools'' (Hornik et al., 1989). The ANN model was calibrated using the Neural Solutions 5.0 software package. During the calibration process, 80 architecture combinations were trained. Variations of the back propagation learning algorithm were applied. As presented by Zhang et al., 1998, the square error of the estimates between the observed and actual output is fed-back through the network causing changes of the weights, with the purpose of preventing that the same error will happen again. Batch back propagation provided smooth curves, with results generally better than those of the other training back propagation methods. At this stage, results from cross-validation analysis in relation to network size and number of training cycles were analyzed to select the best combination to keep the model simple. The data sets of the 68 farm patterns were divided randomly into three subsets, containing 41 patterns for training, 16 patterns for MI and 15 patterns for LOM testing, and 10 patterns for the validation phase. The number of patterns in the training subset was set to about 60% of the total data, as per Zhang et al., 1998. Extraction of the training subset was repeated several times at random to check the quality of the trained networks generated, as indicated by high R^2 and low MSE, MAE and MAPE values and a wide range of outputs. The validity of the model was checked by comparing its output values with those calculated using equations 1 and 2, calculating and comparing mean squared error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and determination coefficient (R^2). The values of the coefficients of ANN models that have been assigned in order to minimize the MAPE are defined in Eq. (3):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\left|\frac{y_i - \hat{y}_i}{y_i}\right| 100\right) \qquad (3)$$

where $\hat{y}$ and y are predicted and actual value, and n is the total number of predictions.

Results

Number of hidden units

To determine the optimal architecture of the ANN, the architectures of networks with hidden units ranging from one to 20 were trained, tested and validated. The validation subset contained 10 patterns, which were not used in the training and testing phases. This subset was used to test the correlation between the values of the outputs given by the ANN model and those calculated from Eqs. (1) and (2). Network architectures with hidden units ranging from one to 20 were simultaneously trained and tested with the respective subsets. The accuracy of the ANN model was validated using the testing data set. Figure 2 shows the correlation between the model's output and actual (calculated) output. In general, networks containing two to eight hidden units showed better performance. Various numbers of hidden layers were tested, but single hidden layer networks showed better results. Furthermore, ANN architecture with

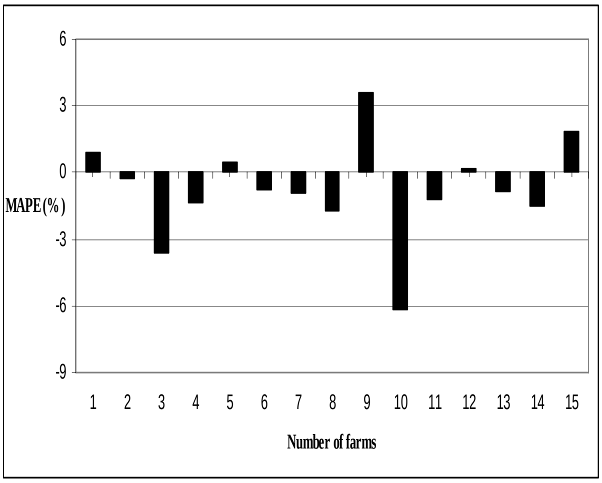

Fig 4. MAPE in the LOM estimation over the test set

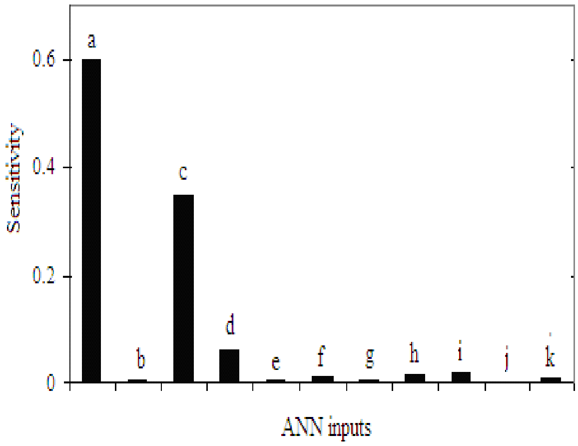

Fig 5. Sensitivity analysis of ANN inputs on LOM. (**a**) area under potato cultivation, (**b**) number of potato plots, (**c**) total power of tractors, (**d**) total number of tractors, (**e**) number of land preparation operations, (f) required time for hand collecting and bagging potatoes, (**g**) cost of fixed labor, (**h**) number of non-subsidized nitrogen fertilizer, (**i**) amount of insecticides, (**j**) number of irrigation, and (**k**) sum of fixed costs.

fewer hidden layers can avoid over fitting problems observed during the trial-and-error procedure. Therefore, ANN with single hidden layer was selected for further investigation. Based on performance measures, single hidden layers with 3 and 8 neurons in the hidden layer were finally selected as optimum configuration for predicting LOM and MI, respectively. The related characteristics of ANN architectures' are given in Table 3.

The testing data set was also tested for the MAPE, MAE and MSE. Data patterns that generated great errors and recognized to be outlier for all of the indicators were rejected from the original data set in order to determine a boundary which represents the applicable range of the ANN model. Rejecting such data patterns enhances the network forecasting capability. The remaining, representative values of the items redefine our target farming system. The rejected data sets were regarded as those that could be analyzed under different farming systems. Figures 3 and 4 show MAPE of the two output indicators for each validation pattern, obtained by comparing the outputs of the best ANN model of every desired output and the outputs calculated using Eqs. (1) and (2) (actual model).

Discussion

Sensitivity analysis

In order to assess the predictive ability and validity of the developed ANN model, two sensitivity analyses were conducted using the best single output network in the case of each output (LOM or MI). In each case, the robustness and sensitivity of the model were determined by examining and comparing the outputs produced during the validation stage with the calculated values. The ANN model was trained by removing one explanatory parameter at a time while not changing any of the other items for every pattern. Results of sensitivity analyses for LOM and MI are shown in Figs. 5 and 6, respectively. According to the obtained results share of each input item of developed ANN model on desired output can be seen clearly. Based on results of sensitivity analyses for each output, three explanatory parameters that have most effect on related outputs are selected and argued.

Analysis of selected parameters

The most meaningful explanatory parameters for the ANN models were area under potato cultivation (ha), total power of tractors (hp) and total number of tractors (for LOM prediction), and total land size (ha), area under potato cultivation (ha) and total hours of farm machinery work (h ha^{-1}) (for MI prediction). Area under cultivation has an important effect on the number and level of farm machinery used and handled. Facilities on farms increased by amplifying owned land size, in this way developing farm size provide additional potential of apply more farm machinery and level of agricultural mechanization (LOM) consequently, and the MI also has this trend. Generally land size has a positive effect on agricultural mechanization that government endeavors to en bloc small farms to import appropriate technology in farms reducing production risk and increase benefit. Total mechanical power implemented in farm can change production status by reducing timeliness losses and increasing quality of farming operations. Agricultural pundits define agricultural mechanization as tractorization, this indicates importance of tractor and mechanical power in agricultural production specially in some crops in different sections of production process such as corn, sugar beet, cane, potato and etc. Because of financial limitations different farms don't have all of the required farm machinery for their farming operations. Hours of farm machinery work on different farms varies with different level of mechanization, thus renting and hiring tractors and other machineries for different operations are required and this phenomenon affect mechanization status in farms.

Conclusions

The developed ANN model predicted well the two indicators, Mechanization Index (MI) and Level of Mechanization (LOM), for the potato farms in the study area in Iran. The correlation

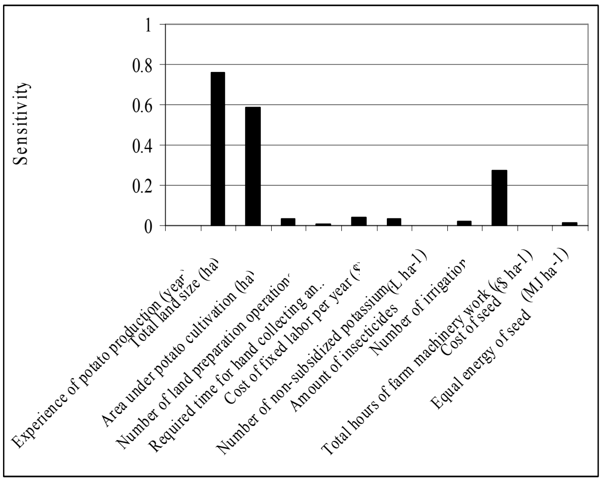

Fig 6. Sensitivity analysis of input items on MI

between the model's outputs and the calculated values of the indicators was quite strong according to the results after the validation phase (10 cases), with R^2=0.989 and 0.992 for MI and LOM, respectively. Furthermore, the developed mechanization indicators provided sufficient information to identify the target farming system as well as to assess their mechanization status. The models are based on a single hidden layer ANN. During the simulation process, the model was sensitive enough while predicting information which agrees well with the observed performance of the target farming system. Therefore, each of the selected input variables contributed the improveement in the performance of the ANN. The wide range of the actual output values for the MI and LOM (0.078 to 3.83 and 0.483 to 0.726 respectively) in the studied farming system suggests that ANN models may be applied to other regions in the country with conditions similar to those in this study. We recommend that the ANN model is tested using specific inputs from different farming systems in other regions of the country; especially where the tractor type described in this study is not the main power source. Further practical application of this work consists on generating a map of mechanization indicators for a much wider area. Analyzing the interrelation between this baseline data, in conjunction with available farm monitor reports could allow between others: resolving indications of average effectiveness of energy conversion, to identify priority areas to replace obsolete agricultural machinery, as well as, to asses the suitability of introducing new tractor units in the region.

References

Abdullakasim W, Koike M, Takigawa T, Hasegawa H, Yoda A, Bahalayodhin B, Usaborisut P (2005) Neural network modeling and sensitivity analysis of the bulk density response in sandy loam under vehicle-induced stresses. Journal of JSAM 67(2): 77-85.

Andrade P, Jenkins BM, (2003) Identification of patterns of farm equipment utilization in two agricultural regions of central and northern Mexico. Agricultural Engineering International: the CIGR Journal of Scientific Research and Development. Invited Overview Paper 5: 276-288.

Barton D, (1999) Options for power tillers and draught animals on small farms. Proc. Workshop on Options for Farm Power Use in Primary Cultivation on Small Farms, Vol. 2. Mymensingh: Bangladesh Agricultural University.

Borlaug NE, Dowswell CR (1993) To nourish infertile soil that feed a fertile population that crowds a fragile world. Fertilizer News, 38: 11-20.

Chester J (1993) Neural Networks, A Tutorial. Prentice Hall Pub. New York.

Collado M, Calderón E (2000) Energía de utilizada para la producción agrícola en el estado de Guanajuato: Cuantificación y análisis. Memorias del III Congreso Latinoamericano de Ingeniería Agrícola. Guanajuato. México. (In Spanish).

Drummond ST, Sudduth KA, Joshi A, Birrel SJ, Kitchen NR (2003) Statistical and Neural Methods for site-specific yield prediction. Transactions of the ASAE. 46(1): 5-14.

García H, García OJ, Ramírez A (2005) INIFAP. Sistema de consulta del potencial
agroproductivo del estado de Guanajuato. SICOPOT-GTO. Guanajuato. México. (CD-ROM In Spanish).

Hornik K, Stinchcombe M, White H, (1989) Multilayer feedforward networks are universal approximators. Neural Networks 2: 359-366.

Oida A (2000) Agricultural machinery management using personal computer (Text & Practice), Agricultural Machinery Management Course organized by Japan International Cooperation Agency. Japan.

Okurut S, Odogola WR (1999) Farm power development in Uganda: Past, present and future trends. Proc.99 Intl. Conf. Agril. Engg., Beijing, China, 1: 232.

Sarker RI (1999) Farm power issues in Bangladesh: Options for primary cultivation on small farms. Proc. Workshop on Options for Farm Power Use in Primary Cultivation on Small Farms, Vol. 2 under the NRI Research Project A0784(X0395) Mymensingh: Bangladesh Agricultural University.

Sims B (1987) Mecanización para el pequeño agricultor. Secretaría de Agricultura y Recursos Hidráulicos. Instituto Nacional de Investigaciones Forestales y Agropecuarias. México, D.F.México. (In Spanish).

Singh RB (2002) The state of food and agriculture in Asia and Pacific. First version
published by IFA and FAO, Parries, France, April 2002. pp 1-21. www.agroindonesia.com/Publications/2002_Singh.pdf.

Snedecor GW, Cochran WG (1989) Statistical Methods. 8th edn. Iowa State Press, USA.

Warkentin BP (1991) Protection of groundwater quality through efficient irrigation. Proc. Intl. Seminar on Efficient Water Use in Mexico City, Oct. 21-25, pp: 1.

Zangeneh M, Omid M, Akram A (2010) A comparative study on energy use and cost analysis of potato production under different farming technologies in Hamadan province of Iran. Energy 35(7):2927-2933.

Zhang G, Patuwo BE, Hu MY (1998) Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting 14: 35-62.