Predicting CO_2 Emissions from Farm Inputs in Wheat Production using Artificial Neural Networks and Linear Regression Models

"Case study in Canterbury, New Zealand"

Majeed Safa
Department of Land Management and Systems, Lincoln University, Christchurch, New Zealand

Peter Nuthall
Department of Land Management and Systems, Lincoln University, Christchurch, New Zealand

Mohammadali Nejat
Department of Agricultural of Sciences, Payame Noor University, Saveh, Iran

Bruce Greig
Department of Land Management and Systems, Lincoln University, Christchurch, New Zealand

***Abstract*—Two models have been developed for simulating CO_2 emissions from wheat farms: (1) an artificial neural network (ANN) model; and (2) a multiple linear regression model (MLR). Data were collected from 40 wheat farms in the Canterbury region of New Zealand. Investigation of more than 140 various factors enabled the selection of eight factors to be employed as the independent variables for final the ANN model. The results showed the final ANN developed can forecast CO_2 emissions from wheat production areas under different conditions (proportion of wheat cultivated land on the farm, numbers of irrigation applications and numbers of cows), the condition of machinery (tractor power index (hp/ha) and age of fertilizer spreader) and N, P and insecticide inputs on the farms with an accuracy of ±11% (± 113 kg CO_2/ha). The total CO_2 emissions from farm inputs were estimated as 1032 kg CO_2/ha for wheat production. On average, fertilizer use of 52% and fuel use of around 20% have the highest CO_2 emissions for wheat cultivation. The results confirmed the ANN model forecast CO_2 emissions much better than MLR model.**

Keywords*—Artificial neural networks; modelling; CO_2 emissions; wheat cultivation

I. Introduction

Around the world wheat is used as one of the main food sources to provide a large proportion of the calories and protein needed by human beings [1]. The world wheat production forecasted for 2020 varied depending on the prediction method used: 746 Mt [2], 840 Mt [3] and 1050 Mt [4]. To meet the target 2020 wheat production, the current average wheat yield of 2.7 t/ha needed to be increased by 40% [5]. The three options available to lift wheat production to meet the 2020 target include: expansion of cultivated land, intensification of cultivated land and increases in production per ha [6].

The use of plant genetics, new pest control methods, and more efficient fertilizers have increased farm production over the last 30 years [7]. At a global level, it would be too difficult to find additional areas for agriculture as most cultivable area is already under use. Intensification of the area currently cultivated involves adopting more rigorous farm operation systems and the application of more chemical inputs (pesticides, fertilizers and fuel). It was expected that the newly-developed seed varieties would have improved yields over the last few decades; but, in many areas, due to the use of traditional farming methods by farmers and other technical limitations, yields are still lower than the desired production [4, 8].

Overall, New Zealand agriculture is dominated by high farm inputs [9, 10]. Agricultural production is a victim of, and contributor to, global warming [10-13]. CO_2 contributes significantly to greenhouse gases (GHG) [14]. The link among production, energy consumption and CO_2 emissions in agricultural activities is well understood [10, 15-18]. CO_2 is emitted during different farming activities, such as land use changes, application of fertilizers and pesticides, the ignition of fossil fuels and plant waste, decay of organic matter and microorganisms in the soil [12, 19, 20]. GHGs could change the current environmental conditions that have uncontrolled impacts on agricultural production. To monitor CO_2 emission reduction targets, the effects of direct and indirect factors on CO_2 emissions should be investigated.

MLR models have been used widely in agricultural projects more than other prediction techniques [21, 22]. A simple model with a high r^2 can be developed through the use of sufficient numbers of samples and independent variables. Input variables are always maintained in the best model if the actual and predicted data are correlated with a p value of 0.05 [23]. In the first step, corroboration between CO_2 emission and each input variable is verified with simple MLR using r^2 as the decision criterion. A MLR model is then established to predict CO_2 emissions as:

$$Y=a_0+a_1V_1+a_2V_2+\cdots+a_nV_n+\epsilon \qquad [1]$$

where a0-an = coefficients of regression, V1-Vn = the input variables and ϵ = the error. The linear model represents the links between the independent (input) variables and the dependent (output) variable.

Artificial neural networks (ANN) have been used recently for investigating the connections between input and output parameters [24, 25]. Based on an analysis of the already-entered data, neural networks can find a link among the input and outputs, as well as the controlled and uncontrolled parameters [26]. To develop an effective ANN model, the number and accuracy of the data sampled are key issues, as ANNs require enough data to develop suitable connections, ANNs cannot develop the correct connections by themselves. ANN models are simple applications that can predict or classify different data to give with robust results. ANNs can estimate nonlinear input-output applications with high accuracy, so can play a vital role in simulating complex systems [27].

The feed-forward multi-layered perception (MLP) paradigm is the most common ANN structure used in modelling studies. The feed-forward MLP paradigm consists of independent variables, hidden layers and an output layer trained by the back propagation (BP) learning method. MLPs trained by BP are capable of modelling any function, so they are widely used for prediction models [28-30]. The neurons associated with the first hidden layer analysis, the weighted independent variables, use a transfer function to lead to the results. The most commonly used transfer functions include: logistic, linear, sine, Gaussian and hyperbolic-tangent. The results from the first hidden layer are then directed to the second hidden layer via weighted connections. Summation of the weighted inputs is processed by the neurons in the hidden layer using their transfer functions. The neuron outputs associated with the output layer are termed called the predicted output [22, 25].

The mean square error (MSE) between the predicted results and the measured data is minimized by adjusting the weights. The following relation is used for estimating the mean square error for a basic network having one output neuron:

$$MSE = \frac{1}{2N}\sum_{i}^{N}(t_i - z_i)^2$$

where z_i = predicted outputs associated with the ith training pattern, t_i = the actual outputs associated with the ith training pattern and N = the sample size of the training patterns [25]. Furthermore, the root mean square error (RMSE) is used to show the errors in the units of the actual and predicted data.

Models with a minimum of input variables are preferable for problem solving. Therefore, data reduction is useful if the number of input variables is high and the available sample size is limited [25].

II. METHOD

The experiments were conducted on irrigated and dry land arable farms totalling 35,300 ha in Canterbury, New Zealand.

Canterbury is the dominant wheat production region in the country and shares almost 90% of wheat cultivation farms and wheat yield in New Zealand [31]. CO_2 emissions from wheat farms were investigated by considering different energy sources such as: fertilizers, pesticides, electricity, fuel and machinery. The following relationship was used to calculate total CO_2 emissions (E).

$$E = \Sigma(A_i\, C_i) \qquad [2]$$

where Σ = summation, A_i = input factor and C_i = the CO_2 emission conversion coefficient for each factor.

Different conversion coefficients were used to convert farm inputs into CO_2 emissions. Selecting accurate conversion coefficients was the key point of this study. Apart from farm inputs, the impact of around 140 factors comprising both technical and social aspects such as the farmers' social status, the properties of tractors and equipment, farm conditions and yield, were investigated.

Except for fuel burning, where carbon dioxide was released directly, CO_2 was also released indirectly from farming activities. The use of most inputs associated with agriculture were converted into energy coefficients to obtain kg CO_2/MJ. Three different sources of data collection were included: a survey, a literature review and field measurements. This study was based on an analysis called the 'cradle-to-gate analysis', which meant that the transport and waste disposal components of the products' life cycles were not involved after they left the farm gate.

A limited number of independent variables were selected to ensure a practical model. The input variables were reduced by applying pre-processing based on correlation analysis, followed by principle component analysis (PCA). Analysis of various variables associated with the components of PCA led to the identification of a cumulative variance with eigenvalues greater than 1. Around 140 input variables were applied in the final ANN model. The analysis consisted of two steps. In the first step (pre-processing), input variables, which had no little correlation with each other but had a significant impact on CO_2 emissions, were selected. In the second step, 16 variables that demonstrated high links with CO_2 production were selected, and included: area of wheat cultivation (ha), percentage of wheat cultivation area on farms, number of cows, annual rainfall, age of farmers, educational background of farmers, irrigation frequency, capacity of tractors (hp), farm size (ha), inputs such as N, P, fungicides, and insecticides, age of fertilizer spreader, number of plough passes, number of sprayer passes, and age of sprayer. The PCA process was guided to select eight independent variables to be applied as independent variables in the ANN model. The eight independent variables selected included plough passage numbers, the proportion of wheat area on farms, irrigation frequency, number of cows, age of fertilizer spreader and farm inputs, nitrogen input (kg), insecticide input (kg), phosphate input (kg), age of sprayer and tractor power index (hp/ha). The selected eight independent variables had a threshold cumulative variance of around 72.3%.

Data were collected from 40 wheat farms. For training purposes 30 farms were selected randomly and the remaining 10 farms were used for model validation.

A limited number of hidden neurons were enough to describe simple nonlinear problems. In contrast, to solve the very nonlinear problems associated with large amounts of input variables large numbers of neurons were essential to predict an output variable with a low margin of error. Currently, neuron numbers were selected based mostly on trial and error rather than science [32]. For the purpose of this study, the different aspects of the ANN model were optimised by using a genetic algorithm-based optimisation to determine a satisfactory model structure. A number of trials led to the selection of a modular neural network with two hidden layers containing two sub-networks (Figure 1).

For function approximation, the optimised model was trained. In the training process, the weight change followed by subsequent batch processing was controlled by the learning rate. A training process with a higher learning rate would be quicker; but the weights may oscillate around the lowest level of error, but never reach it [25]. Subsequently, this study used a learning rate of 0.01 (low). The learning method adopted (Quick Prop) was very fast in reducing flaws when finding promising results. Quick Prop adjusted weights by indirectly using the second derivative of error. In each trial of Quick Prop, weights were revised using following relationship:

$$w_{m+1} = w_m + \Delta w_m$$

$$\Delta w_m = \frac{d_m}{d_{m-1} - d_m} \Delta w_{m-1}$$

$$d_m = \sum_{n-1}^{N} \left[\frac{\partial E}{\partial w_m} \right]_n$$

where Δwm = the existing weight, dm = the average derivative of the error for the current epoch (batch) m; and ∂E/∂wm = the current error gradient for a particular input vector [25].

This study examined different functions, that included the logistic sigmoid, hyperbolic tangent (tanh), sine, Gaussian and linear functions. To propose the final model, the hyperbolic tangent function was selected for the input layer and the first hidden layer; the logistic function was applied for the second hidden layer; and the linear function was selected for the output layer (Figure 3). These functions can be written as:

$$L(u) = [1 + e^{-u}]^{-1}$$

$$tanh(u) = \frac{1 + e^{-u}}{1 - e^{-u}}$$

where tanh (u) = the hyperbolic tangent function, L(u) = the logistic function, and u = weighted sum of inputs into a neuron [25].

III. Results and Discussion

A. CO_2 Emissions

The study revealed that an average of 1032 kg CO_2/ha, was released from each wheat cultivation farm. To achieve a wider perspective, farm inputs were divided into five categories: electricity, fertilizers, agrichemicals, machinery and fuel. As shown in Table 1, fertilizer (mostly nitrogen) was ranked highest on the farms studied, with 52% of total CO_2 emissions.

TABLE I. Total CO2 Production from Different Agricultural Inputs Applied on Wheat Cultivation Farm (Kg CO2/Ha)

	Fertilizer	*Agrichemicals*	*Electricity*	*Machinery*	*Fuel*	*Total*
Total	539 (52%)	55 (5%)	86 (8%)	149 (14%)	203 (20%)	1032

B. Model Development

MLR and ANN models were developed for forecasting CO_2 emissions from the wheat cultivation farms.

1) Multiple Linear Regression Model

MLR demonstrates the linear relationships between the input and output variables. In this study the MLR model was compared with an ANN for which data from 25% randomly-selected samples were used for model validation and the remaining 75% data were used for model training. The MLR model developed was used to estimate, the validation data. The MLR model was able to be fitted to the CO_2 emission data and accounted for around 35% and 70% of the variance in the validation and training data, respectively. Figures 1 and 2 demonstrate the relationships between the forecasted and measured CO_2 emissions, respectively, for the training and validation data. The MSE and RMSE estimated for validation data were 6977 and 84 kg CO_2/ha, respectively.

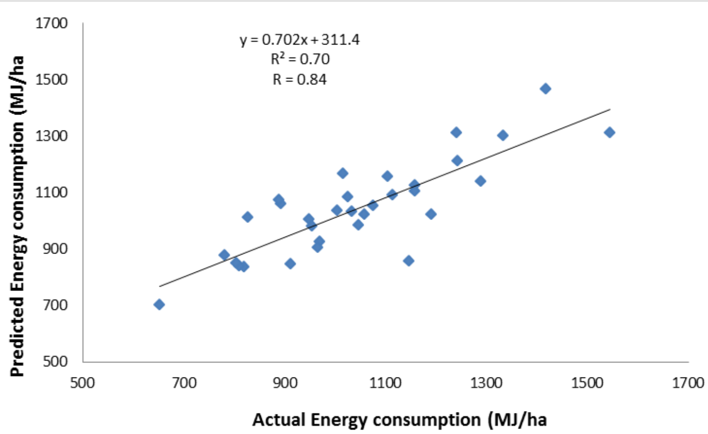

Fig. 1. Relationships between the field measurements and model-predicted CO2 emissions (training) based on the MLR model

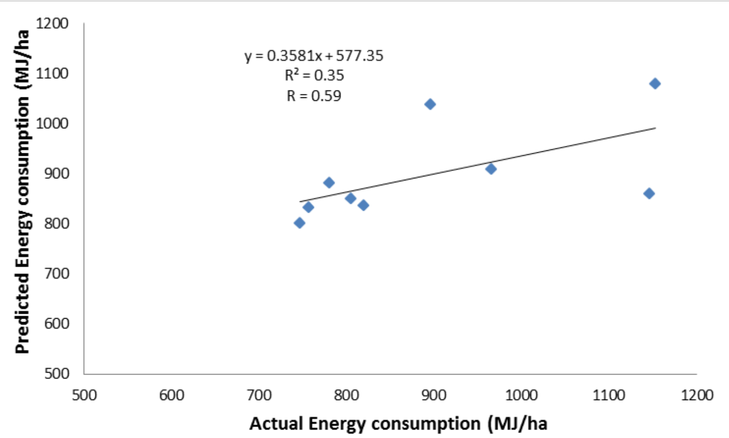

Fig. 2. Relationships between the field measurements and model-predicted CO_2 emissions (validation) based on the MLR model

2) *Artificial Neural Network Model*

After trialling different neuron activation functions, learning algorithms and network structures, a modular network with two hidden layers was developed (Figure 3). After the input layer, the modular network was divided into two parts. There were 20 neurons in the input layer with two and 12 neurons, respectively, for the first and second parts of the modular network. The final output (CO_2 emissions) was produced by combining the results in the output layer.

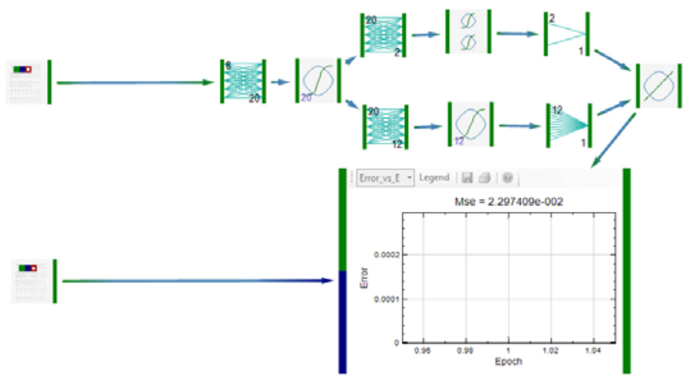

Fig. 3. Structure of the modular network and the number of neurons in each layer

Trials (1500) led to the production of the most satisfactory ANN model with a scaled MSE of 2.3×10-2 (with inputs and outputs ranged between -1 and +1). Compared to MLR, the ANN model predicted CO_2 emissions effectively and accounted for almost 90% of the variance (Figure 5) in the validation data. Figures 4 and 5 show the relationship between the actual and predicted data for the training and validation of the ANN model. The r^2 was estimated at 0.82 and 0.89, respectively, for training and validation of the ANN model.

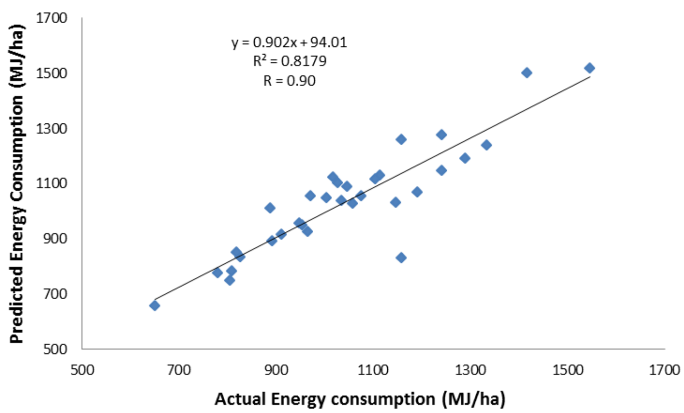

Fig. 4. Relationships between the field measurements and model-predicted CO_2 emissions (training) based on the ANN model

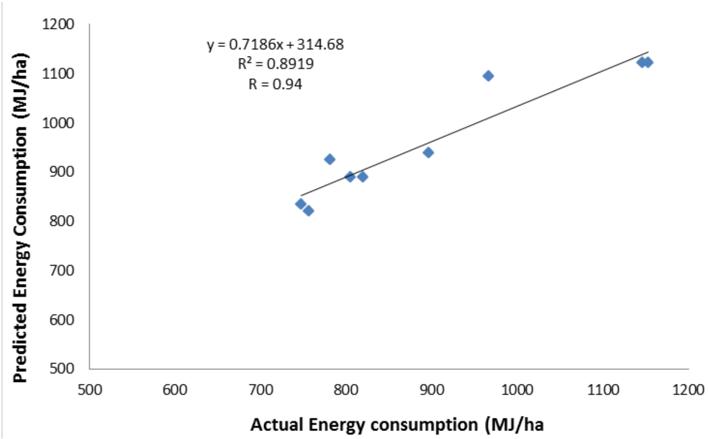

Fig. 5. Relationships between the field measurements and model-predicted CO_2 emissions (validation) based on the ANN model

Figures 6 and 7 illustrate the ANN predictions for the training and validation data, respectively. The four lines in each picture represent the desired output, network output, and the high and low boundaries of the confidence intervals. The region within which the correct answer was within the 95% confidence level, as indicated by the grey area. As Figure 7 shows, the final model can predict CO_2 emissions up to ±113 kg CO_2/ha within the 95% confidence level. The results indicated the chance that the predicted errors would be more than ±113 kg CO_2 /ha was only 5%.

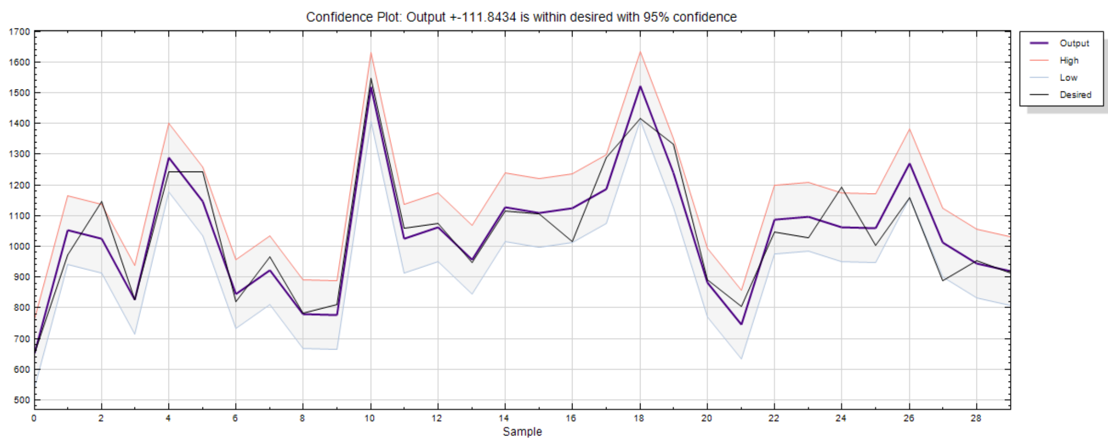

Fig. 6. Predicted, observed and 95% confidence interval for CO_2 emissions based on the ANN model (training data)

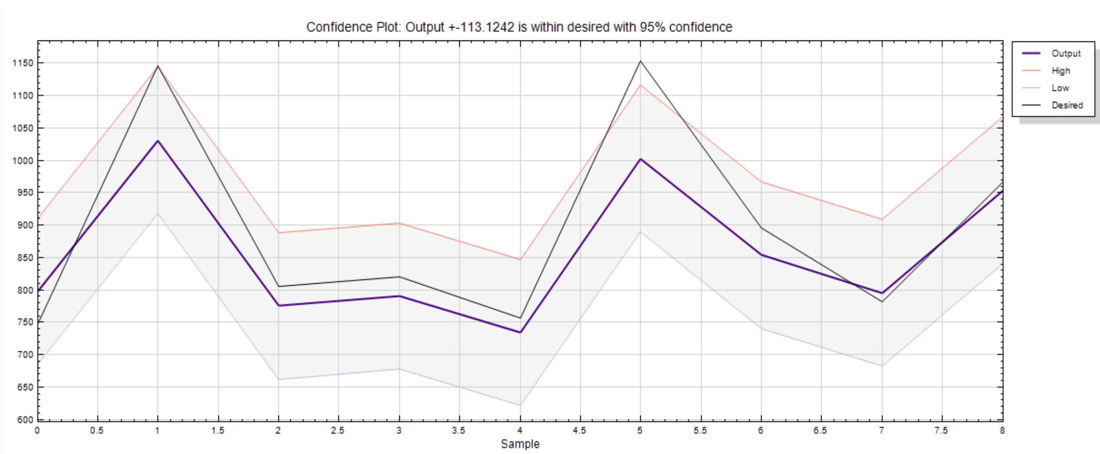

Fig. 7. Predicted, observed and 95% confidence interval for CO_2 emissions based on the ANN model (validation data)

For the training and validation data, the link between the measured and forecasted CO_2 emissions for the ANN model was much greater compared to that of the MLR model. Compared to MLR, the ANN model had a noticeably lower RMSE for the validation data (Table 2).

TABLE II. MSE AND RMSE OF THE MLR AND ANN MODELS

	Linear		*ANN*	
	Training	Validation	Training	Validation
MSE	5635	6977	3641	3307
RMSE	75	83	60	57

There were a number of uncontrolled parameters that could affect CO_2 emissions in agricultural farms, and they made the results of these experiments quite interesting. The proposed model can predict CO_2 emissions in wheat farms within tolerable margins of error. There were some fixed independent parameters in the proposed model that could not be changed: such as farm conditions. Some of the independent variables, such as the farmer's education, would also influence the CO_2 emission indirectly. This calls for future studies to investigate the detailed links between the input parameters and CO_2 emissions from agricultural farms.

The ANN model can predict CO_2 emissions from farm inputs. The model can help farmers identify the factors with more potential to minimise CO_2 emissions on their farms. In addition, scientists and decision makers can evaluate CO_2 emissions in Canterbury.

IV. CONCLUSIONS

In this study an ANN model was developed to forecast CO_2 emissions from farm inputs using direct and indirect factors to predict CO_2 emissions from wheat farms. The proposed ANN model could forecast CO_2 emissions from farm inputs depending on each farm's conditions, such as the proportion of wheat cultivation area on farms, frequency of irrigation and number of cows, the condition of machinery (tractor power index (hp/ha) and age of fertilizer spreader) and inputs on the farm (N, P and insecticide use) in Canterbury agricultural farms with a margin of error of ±11% (±113 kg CO_2/ha). As there were numbers of uncontrolled factors in agricultural production, the size of error was acceptable. In addition, the results showed that the ANN model using heterogeneous data can better forecast CO_2 emissions than the MLR model (Table 2). Using dissimilar inputs, such as farm conditions and social factors, would help the relevant agencies view the problem from various angles.

The finding from these experiments indicated the capability of an ANN model for forecasting CO_2 emissions from agricultural inputs by adopting indirect factors. This improved model can support decision makers by providing information on predicted CO_2 emissions from a wide range of farm products. Analysis of the results made it clear that it was not possible to change some input parameters in the short term. However, for the scientific community and decision makers, the model would provide useful information to judge the best directions for CO_2 emission reductions in the future.

Testing the results for at least five years with larger sample sizes would lead a more accurate model for forecasting the trend of CO_2 emissions in agricultural farms under various situations. The outcomes of this research can be recognised as a first effort to propose methods appropriate for estimating CO_2 emissions by considering geographical, social and technical parameters together. This proposed approach can be replicated to other farm production systems and cropping areas.

REFERENCES

[1] Breiman A, Graur D. Wheat Evaluation. Israel J Plant Sci. 1995; 43: 58-95.

[2] OECD-FAO. OECD-FAO Agricultural Outlook. Paris: ECD Publishing, OECD & FAO,; 2011.

[3] Kronstad WE. Agricultural development and wheat breeding in the 20th century., Developments in Plant Breeding vol. 6. Kluwer Academic Publishers, Dordrecht, The Netherlands, p. 1-10.

[4] Rosegrant MW, Agcaolli-Sombilla A, Perez N. Global Food Projections to 2020: implications for investment 2020 Vision Discussion papers. 1995;5.

[5] Kole C. Cereals and millets. Berlin ; New York: Springer, 2006.

[6] Vlek PLG, Rodríguez-Kuhl G, Sommer R. Energy Use and CO2 Production in Tropical Agriculture and Means and Strategies for Reduction or Mitigation Environment, Development and Sustainability. 2004;6.

[7] Pimentel D, Pimentel M. Food, energy, and society. 3rd ed. Boca Raton, FL: CRC Press, 2008.

[8] Ozkan B, Akcaoz H, Fert C. Energy input-output analysis in Turkish agriculture. Renewable Energy. 2004;29(1):39-51.

[9] Wells C. Total energy indicators of agricultural sustainability : dairy farming case study. Wellington [N.Z.]: Ministry of Agriculture and Forestry; 2001. p. vii, 81 p.

[10] Safa M, Samarasinghe S. CO2 emissions from farm inputs "Case study of wheat production in Canterbury, New Zealand". Environmental Pollution. 2012;171(0):126-32.

[11] Sauerbeck DR. CO2 emissions and C sequestration by agriculture – perspectives and limitations Nutrient Cycling in Agroecosystems. 2001;60:253-66.

[12] Hillier J, Walter C, Malin D, Garcia-Suarez T, Mila-i-Canals L, Smith P. A farm-focused calculator for emissions from crop and livestock production. Environmental Modelling & Software. 2011;26(9):1070-8.

[13] Rajaeifar MA, Ghobadian B, Safa M, Heidari MD. Energy life-cycle assessment and CO2 emissions analysis of soybean-based biodiesel: a case study. Journal of Cleaner Production. 2014;66(0):233-41.

[14] Janardhan V, Fesmire B. Energy explained. Lanham, Md.: Rowman & Littlefield, 2011.

[15] USDA. U.S. Agriculture and Forestry Greenhouse Gas Inventory: 1990-2005. Washington, D.C: United States Department of Agriculture (USDA); 2008.

[16] Stout BA. Handbook of energy for world agriculture. London; New York: Elsevier Science Pub. Co., 1990.

[17] Snyder CS, Bruulsema TW, Jensen TL, Fixen PE. Review of greenhouse gas emissions from crop production systems and fertilizer management effects. Agriculture, Ecosystems & Environment. 2009;133(3-4):247-66.

[18] Safa M, Samarasinghe S. Determination and modelling of energy consumption in wheat production using neural networks: "A case study in Canterbury province, New Zealand". Energy. 2011;36(8):5140-7.

[19] Smith KA, Conen F. Impacts of land management on fluxes of trace greenhouse gases. Soil Use and Management. 2004;20:255-63.

[20] Jeffrey S A. Effects of atmospheric CO2 concentration on wheat yield: review of results from experiments using various approaches to control CO2 concentration. Field Crops Research. 2001;73(1):1-34.

[21] Colwell JD. Estimating fertilizer requirements : a quantitative approach. Wallingford: CAB International, 1994.

[22] Safa M, Samarasinghe S, Nejat M. Prediction of Wheat Production Using Artificial Neural Networks and Investigating Indirect Factors Affecting It: Case Study in Canterbury Province, New Zealand. Journal of Agricultural Science and Technology. 2015.

[23] Alvarez R. Predicting average regional yield and production of wheat in the Argentine Pampas by an artificial neural network approach. European Journal of Agronomy. 2009;30(2):70-7.

[24] Parten C, Hartson C, Maren A. Handbook of neural computing applications: San Diego, CA (USA); Academic Press Inc, 1990.

[25] Samarasinghe S. Neural networks for applied sciences and engineering : from fundamentals to complex pattern recognition. Boca Raton, FL: Auerbach, 2007.

[26] Kalogirou SA. Applications of artificial neural-networks for energy systems. Applied Energy. 2000;67(1-2):17-35.

[27] Hagan M, Demuth H, Beale M. Neural network design: Boston, USA: PWS Publishing Company, 2002.

[28] Hornik K, Stinchocombe M, White H. Multilayer feedforward networks are universal approximators. Elsevier Science Ltd Oxford, UK, UK 1989;2(5).

[29] Heinzow T, Tol RSJ. Prediction of crop yields across four climate zones in Germany: an artificial neural network approach. Centre for Marine and Climate Research, Hamburg University, Hamburg. 2003.

[30] Jebaraj S, Iniyan S. A review of energy models. Renewable and Sustainable Energy Reviews. 2006;10(4):281-311.

[31] Statistics New Zealand. Agricultural Production Statistics (Final): June 2010. Wellington: Statistics New Zealand; 2010.

[32] Kermanshahi B, Iwamiya H. Up to year 2020 load forecasting using neural nets. International Journal of Electrical Power & Energy Systems. 2002;24(9):789-97