**Scientific Papers Series Management, Economic Engineering in Agriculture and Rural Development**
**Vol. 15, Issue 2, 2015**
**PRINT ISSN 2284-7995, E-ISSN 2285-3952**

# EFFECT OF VARIOUS INPUTS ON PADDY PRODUCTION - A COMPARISON OF ARTIFICIAL NEURAL NETWORKS AND LINEAR REGRESSION ANALYSIS

**Mohammad Ali HORMOZI[1], Abbas ABDESHAHI[2], Mohammad Amin ASOODAR[1]**

[1]Ramin Agriculture and Natural Resources University of Khuzestan, Department of Agricultural Machinery and Mechanization, Mollasani, Iran, Phone/fax: +98.61.3652.3922,
Emails: hormozi@ramin.ac.ir, asoodar@scu.ac.ir
[2]Ramin Agriculture and Natural Resources University of Khuzestan, Department of Agricultural Economics, Mollasani, Iran, Phone/fax: +98.61.3652.2434, Email: ahmadreza1378@yahoo.com

***Corresponding author.*** hormozi@ramin.ac.ir

***Abstract***

*We analyzed the effect of chemical fertilizer, seed, biocide, farm machinery and labor hours on production of paddy (paddy rice) in the Khuzestan province in the South Western part of Iran. Here we test two methods (linear regression and neural network). We conclude that the results gotten by neural network with no hidden layer and linear regression are closed to each other. We insist that for a data set of this type the regression analysis yields more reliable results compared to a neural network. They suggest that machinery has a very clear positive effect on yield while fertilizer and labor doesn't affect on it. One can say that there is no necessity that increasing the amount of some "useful input" increase paddy production.*

***Key words:*** *neural networks, linear regression, paddy*

## INTRODUCTION

Rice is one of the most important cereal crops grown globally. Rice is important as a staple human food source in many areas of Iran, where the per capita consumption of rice is approximately 100 grams per day. In 2008, which was a drought year, statistics show that Iran had 527,000 hectares of paddy cultivation. The Khuzestan province is normally known as the fifth largest rice producer, with 15,000 hectares devoted to rice production; nonetheless, due to low productivity, the paddy production of Khuzestan was announced to be sixth highest in Iran in 2008 [5]. In a normal year, rice-cultivated areas may reach 59,000 hectares [4]; however, the nearly 1400-kg/ha difference among the regions of Khuzestan shows there are other variables that affect productivity [1]. Moreover, some reasons make it difficult to predict the effect of variables on productivity. For example, the situation of climate; the climate of Khuzestan is generally hot and occasionally humid, particularly in the south, while winters are much more pleasant and dry. Summertime temperatures routinely exceed 50 degrees Celsius and in the winter it can drop below freezing, with occasional snowfall, all the way south to Ahvaz. Khuzestan province is known to master the hottest temperatures on record for a populated city anywhere in the world. Many sandstorms and dust storms are frequent with the arid and desert-style terrains.

In this project, we will give the effect of chemical fertilizer, seed, biocide, farm machinery and labor hours on production of paddy (paddy rice) in the Khuzestan province by using two methods (linear regression and neural network).

## MATERIALS AND METHODS

The research survey was undertaken in the Khuzestan province in South Western Iran.

The data used in this study were collected during a survey covering the crop year of 2009 in two climatic regions: (I) mountainous North Eastern Khuzestan and (II) the plains (the rest of Khuzestan). The Baghmalek and

Izeh districts from region (I) and the Ahvaz, Shushtar, Dezful, Shush, Ramhormoz and Dashte-Azadegan districts from region (II) were chosen. The data set consists of 93 observations. It includes amongst six variables: The output of production as dependent variable is kilogram per hectare (kg/ha) of paddy, the inputs as quantitative variables are namely chemical fertilizer (kg/ha) that was applied in the farm, seed planted per hectare (kg/ha), biocides including insecticide, herbicide, and fungicide that were used by farmers (kg/ha), farm machinery that energy equivalent is used in this study (mega joule per hectare) and labor hours that were applied (h/ha) [3].

The main goal of this project is to detect the effect of chemical fertilizer, seed, biocides, farm machinery and labor hours on paddy production. To achieve this we use two methods.

The "Neural Network" as employed here is obviously not a biological concept. It is artificial. In the regime of neural networks, "Neural Networks" and "Artificial Neural Networks" are interchangeable. Neural networks are inspired by the functioning of the biological network of neurons in human brain. A neural network resembles the brain in two respects: First, knowledge is acquired by the network from its environment through a learning process; second, interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge. It is certain that a neural network model is composed of many neurons, which are basic unites that work like computer processors. The units are connected by communication paths (connections) with weights. The units operate only on the inputs that they receive via the connections, and then send the outputs to the next layer of units. As a neural network model, it should have some sort of "training" rule whereby the weights of connections are adjusted on the basis of training set data. Neural networks use a series of neurons in what is known as the hidden layer that apply nonlinear activation functions to approximate complex functions in the data [2]. In this study we consider a Neural Network with no hidden layer.

Linear regression is an approach to modelling the relationship between a scalar dependent variable y and one or more explanatory variables denoted X. The case of one explanatory variable is called simple regression. More than one explanatory variable is multiple regression. (This in turn should be distinguished from multivariate linear regression, where multiple correlated dependent variables are predicted, (citation needed) rather than a single scalar variable.)

Linear regression models are often fitted using the least squares approach, but they may also be fitted in other ways, such as by minimizing the "lack of fit" in some other norm (as with least absolute deviations regression), or by minimizing a penalized version of the least squares loss function as in ridge regression [2].

## RESULTS AND DISCUSSIONS

**Neural Network.** A summary of the main results is provided by Table 1-2 and Fig. 1-2. Where "t.y." stands for "to yield". The process needed 7098 steps until all absolute partial derivatives of the error function were smaller than 0.01 (the default threshold). The estimated weights range from -18.25 to 550.48. The output of the neural network, i.e. the fitted values o(x), is provided by Table 2.

Table 1. Coefficients for neutral network

| Intercept.t.y. | Fertilizer.t.y. | Biocide.t.y. | Seed.t.y. | Machinery.t.y. | Labor.t.y. |
|---|---|---|---|---|---|
| 550.48 | 1.26 | -18.25 | -4.51 | 3.21 | 0.54 |

Table 2. Output of the neural network

| Fertilizer | Biocide | Seed | Machinery | Labor | Output |
|---|---|---|---|---|---|
| 258 | 2 | 150 | 1067.45 | 306.5 | 3759.03 |
| 184 | 6.5 | 100 | 1046.05 | 284.1 | 3728.47 |
| 46 | 6 | 150 | 770.63 | 533 | 2588.21 |
| 115 | 1 | 100 | 697.63 | 490 | 2734.24 |
| 115 | 0 | 200 | 1050.95 | 320 | 3343.85 |
| 161 | 3.5 | 150 | 822.23 | 531 | 2943.52 |

Figure 1 suggests that machinery has positive effect to the yield status. The generalized

weights are given for all covariates within the same range (Fig. 2).

The distribution of the generalized weights suggests that the covariate fertilizer and labor have no effect on the yield status since all generalized weights are nearly zero.

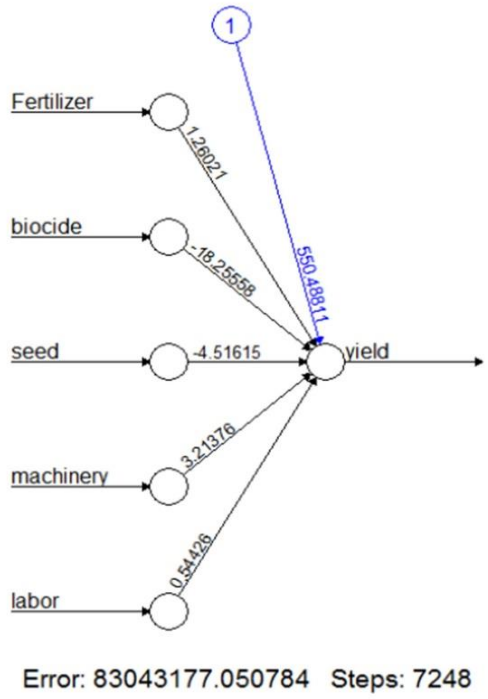


Fig. 1. Summary of the main results for neutral networks

**Linear Regression.** A summary of the main results is provided by Table 3-4 and Fig. 3.

In the output from this first analysis, the p-value for the full model is 0.002540815, small enough to suggest that at least one of the predictor variables is significant in the model. The adjusted R square is 0.14 which indicates that the predictor variables explain 14 percent of the variance in the response variable.

We believe that "Machinery" is the predictor variable which is useful for prediction, this result is same as what we get from neural network but **for a data set of this type the regression analysis yields more reliable results compared to a neural network.** Coefficients for linear regression are given in Table 4.

It suggests that machinery has positive effect to the yield while fertilizer and labor doesn't affect on it. This result is quite similar (with a little difference) to what we get from neural network (see Fig. 1).

The plot in the upper of Fig. 3 shows the residual errors plotted versus their fitted values.

The second plot (from up to down) is a standard Q-Q plot, which should suggest that the residual errors are normally distributed. The scale-location plot (third one) shows the square root of the standardized residuals (sort of a square root of relative error) as a function of the fitted values.

Finally, the plot in the lower shows each points leverage, which is a measure of its importance in determining the regression result.

For many observations, Cook's distance is less than 1 hence most of them may not be considered an outlier.

Table 3. Output of the linear regression

| Input | Estimate | Std. Error | *t* value | *Pr*(>\|t\|) |
|---|---|---|---|---|
| Intercept | 550.48 | 1077.09 | 0.51 | 0.61 |
| Fertilizer | 1.26 | 2.09 | 0.60 | 0.54 |
| Biocide | -18.25 | 81.72 | -0.22 | 0.82 |
| Seed | -4.51 | 3.82 | -1.17 | 0.24 |
| Machinery | 3.21 | 0.86 | 3.69 | 0.0004*** |
| Labor | 0.54 | 0.56 | 0.95 | 0.34 |

Table 4. Coefficients for linear regression

| Intercept | Fertilizer | Biocide | Seed | Machinery | Labor |
|---|---|---|---|---|---|
| 550.48 | 1.26 | -18.25 | -4.51 | 3.21 | 0.54 |

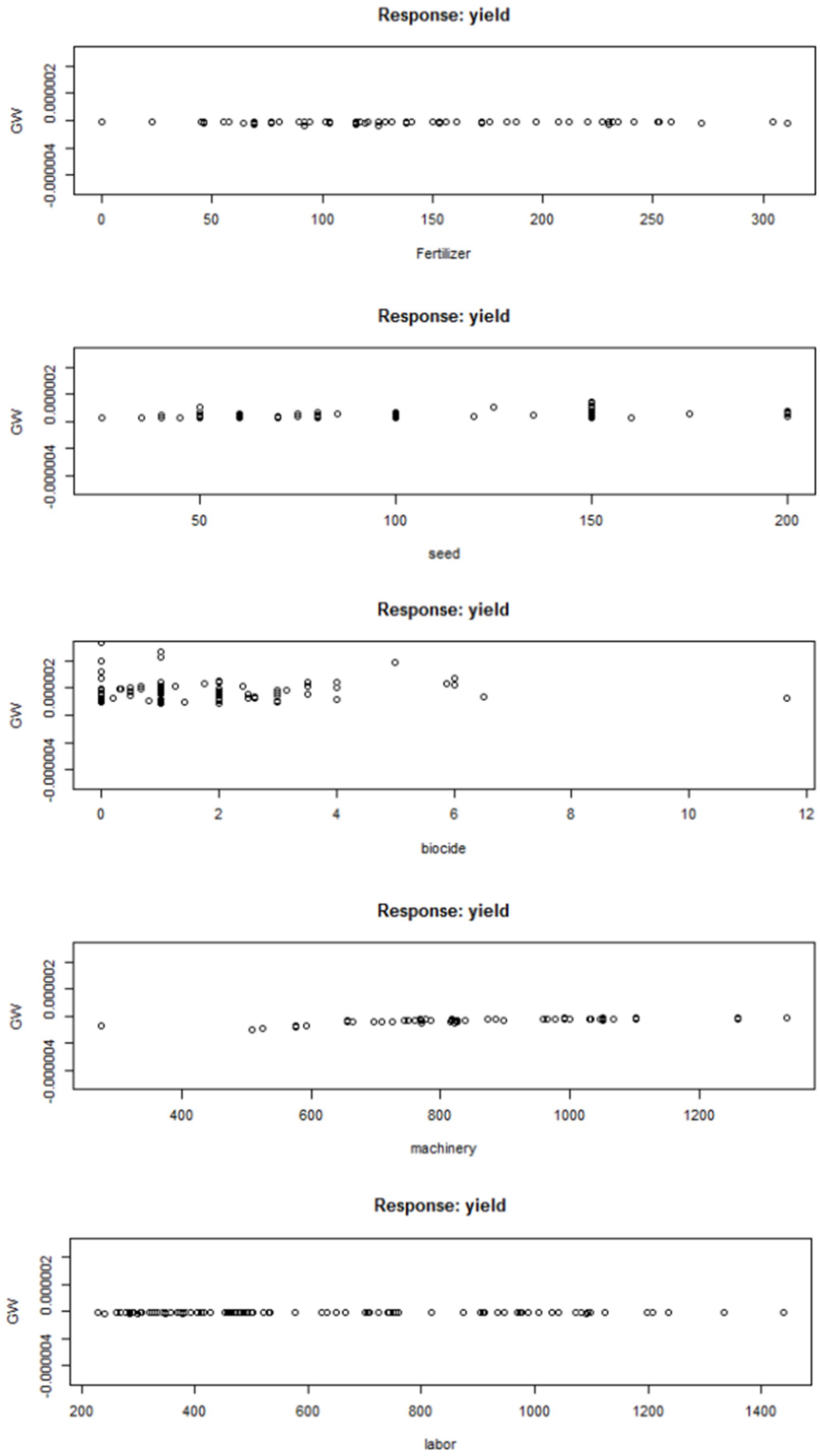


Fig. 2. Plots of generalized weights with respect to each covariate.

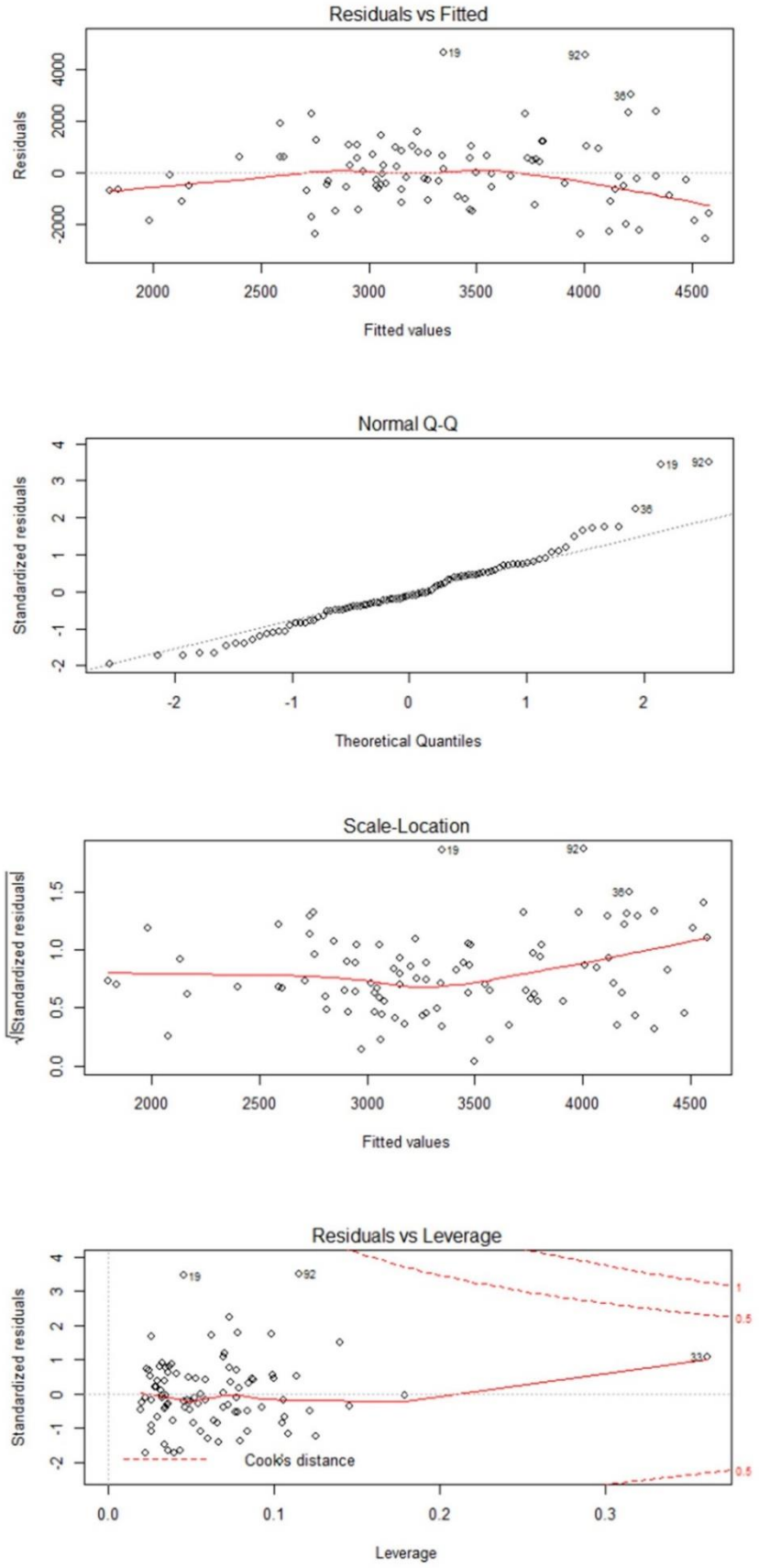


Fig. 3. Plots of linear regression

## CONCLUSIONS

We conclude that the results gotten by neural network with no hidden layer and linear regression are closed to each other. We insist that for a data set of this type the regression analysis yields more reliable results compared to a neural network. They suggest that machinery has a very clear positive effect on yield while fertilizer and labor does not affect on it. One can say that there is no necessity that increasing the amount of some "useful object" increase plant growth. Hence result makes sense.

As a future work someone can use qualitative variable e.g "transplanting in mountainous region", "transplanting in plain region", "wet seeding in poor drainage lands", "wet seeding in normal drainage lands" and "Dry seeding" or using other methods.

## REFERENCES

[1]AOKP, 2007, Status report of rice production in Khuzestan province. Agriculture Organization of Khuzestan Province. Ahvaz. 55 pp.

[2]Bishop, C.M., 2006, Pattern Recognition and Machine Learning, New York: Springer. 738 pp.

[3]Hormozi, M.A., Asoodar, M.A., Abdeshahi, A., 2012, Impact of mechanization on technical efficiency: A case study of rice farmers in Iran, Procedia Economics and Finance. 1:176–185.

[4]MJA, 2006, Agricultural statistics. Department of Statistics and Info-Technology. Ministry of Jihad-Agriculture. Tehran, Iran. Vol I.

[5] MJA, 2008, Agricultural statistics. Department of Statistics and Info-Technology. Ministry of Jihad-Agriculture. Tehran, Iran. Vol I.