Neural Networks Sorghum Yield Forecasting Model Using Satellite-based Vegetation Indices in the Rain-fed Sectors in Sudan

Adil M. Alkhidir¹, Hanan H. A. Adlan², and Isam A. Basheir²

¹Department Civil Engineering, Faculty of Engineering, University of Khartoum, Khartoum, Sudan
²Department of Computer Sciences, Faculty of Mathematical Sciences, Khartoum, Sudan

Abstract: Neural network is universal function approximators. The most commonly used type of neural network is the Radial Basis Function (RBF). Artificial neural networks (ANNs) are made up of input layers, hidden layers and output layers. The RBF neural network has an input layer where the data samples are fed, typically after being normalized. A model has been developed based on a Radial Basis Function (RBF). Study area that has been selected is Gedaref state. The input data are visible Red-Green-Blue (RGB) images of agricultural areas used as predictors to predict the Sorghum yield. Vegetation Index green (VIg) in the Sorghum production is implemented to the Satellite images as input which is expressed as follows:

\[ \text{VIg} = \frac{\text{green} - \text{red}}{\text{green} + \text{red}} \]

Neural Network architectures were developed, investigated, and tested for forecasting Sorghum yield for the selected study area. A monthly imagery input data and Sorghum yield output data sets are used for training and validation of the Neural Network models in the selected study area. Correlation between the measured output data (actual) and the forecasted output data using the model have given high value of R. The model can be used with great efficiency to forecast Sorghum (output) in targeted agriculture area given RGB image of the study area.

Keywords: Neural Network; Vegetation Indices; Radial Basis Function (RBF); Forecasting and Sorghum yield

INTRODUCTION

A neural network is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:
1. Knowledge is acquired by the network from its environment through a learning process.
2. Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge. The modification of synaptic weights provides the traditional method for the design of neural networks. [1].

In the latest two decades scientists who are working in the fields of food security and climate change use ANN for crops forecasting. Because crop yield forecasting plays an important role in farming planning and management, domestic food supply, international food track, ecosystem sustainability [2]. It has been proven that a linear regression is insufficient to show the interactions of the factors and crop yield compared with ANN crop forecasting [3]. Sudhanshu et al apply ANN for agricultural crop yield prediction using vegetation indices. [4]. A computer model that forecasts growth and yield of rain-fed sorghum crop was developed by Adil et al. using ZIMSched (a spreadsheet based model). Climate and soil data is used as input to the model and the overall results of the model application were satisfactory. [5]. Water budget model is developed for forecasting Sorghum in the rain-fed sector in (Gedaref) using special purpose software packages for hydrometeorological analysis and crop water simulator (RAINBOW, CROPWAT and BUDGET). It was recommended that the BUDGET program is not a 100% indicator to specify the actual yield. This calls for revising BUDGET program by adding additional modules to consider other factors of production[6].

Cereal grains are the most important calorie source in Sudanese diet. For most Sudanese people the Sorghum is critical for national food security. Therefore developing Sorghum yield forecasting models will fill the gap of Sorghum forecasting research areas which is crucial to adapt the climate change consequences. Average of Vegetation Index green (VIg) for all pixels under Sorghum production triangle is used as an input to artificial neural network models. Radial Basis Functions Neural Network models are
developed and investigated to forecast Sorghum yield in the rain-fed sectors in Sudan (Gedaref as study area) using MATLAB (2013a).

**STUDY AREA**

The Gedaref state is located in eastern Sudan (figure 3.1), boarded by Kassala state to the north, Khartoum state to the northwest, Sinnar state to the south, Gezira state to the west and Eritrea to the east. The state covers a total area of 75,263 Km². It lies between latitude 12° 45' N and 14° 15' N and longitude 34° E and 37° E, its average altitude is 600 meters above the see level. Also, the region under consideration is about 490 km from the capital Khartoum and 770 km from Port Sudan city, the main sea port of Sudan. Thus the region’s geographical position is favorable to domestic and foreign trade. The black triangle (Fig 1.) boards the areas of Sorghum production.

Fig. 1: Gedaref state, Triangle of Sorghum production.

2. METHODOLOGY

Radial-Basis Function (RBF) network involves three layers with entirely different roles. The input layer is made up of source nodes (sensory units) that connect the network to its environment. The second layer, the only hidden layer in the network, applies a nonlinear transformation Gaussian Radial Basis Function (RBF) from the input space to the hidden space; in most applications the hidden space is of high dimensionality. The output layer is linear, supplying the response of the network to the activation pattern (signal) applied to the input layer.

![Artificial Neural Network neuron model](image)

Fig. 3: Artificial Neural Network neuron model

This neuron model is formed by the following components:

**Parameters:** parameters are usually known as synaptic weights.

**Transfer function:** The function whose argument is the combination between synaptic weights and input vector.

In this research the neuron model will be focused on as shown in Fig. 3.

**Architecture:** The architecture of a neural model gives information about how neurons are arranged. The number of neurons in both input and output, whereas the number of hidden layers [11].

**Radial Basis Function (RBF) Training:** RBF training is based on a general method for building linear models. The function that is used for training iteratively creates a radial basis network one neuron at a time. Neurons are added to the network until the sum-squared error falls beneath an error goal or a maximum number of neurons have been reached. The criterion is typically chosen to maximize the generalization capability of the network. [12], [13].

**Cost function:** The cost function define what is desired, given by the Mean Square Error (MSE) as.

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - y_{di})^2
\]  

Where \(y_i\) is the predicted value obtained from the neural network model, \(y_{di}\) is the actual value.

\[
MBE = \frac{1}{n} \sum_{i=1}^{n} (I_{pi} - I_i)
\]  

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (I_{pi} - I_i)^2}
\]

Where \(I_{pi}\) is the predicted value, \(I_i\) is the measured value, and \(n\) is the number of observations.

![Flowchart](image)

Fig. 4 The flowchart in Fig. 4 describes the steps of developing the neural network model.
DATA

Satellite images are collected from the website:
Images are in tiff format and exported into RGB bmp images. Images are resized into 512x512, image pixel is 250m. Each pixel represents about 14.9 Feddan. An algorithm for converting the image set into Vegetation Index green (VIG) indices using Matlab for months (June, July, August, September and December) for all years 2001 to 2013.[13] Vegetation Index green (VIG) is expressed as follows:

\[
VIG = \frac{\text{green} - \text{red}}{\text{green} + \text{red}}
\]  

(4)

Models are constructed using months of Jun, July, Aug, Sep and Oct RGB values as input to ANN model. Yield data (production and area) are provided by Ministry of Agriculture (Gedaref state).

MATLAB is used in the development of the model. Inputs and outputs are normalized using MATLAB function (prestd) which preprocesses the network training set by normalizing the inputs (p) and targets (t) so that they have means of zero and standard deviations of 1.

Input: Average of Vegetation Index green (VIG) in the Sorghum production triangle is implemented to the Satellite images as input.

Output: Sorghum yield (Sacks/Feddan). Models are differentiated by architectures by varying the number of inputs.

Fig. 5: Gerdaref – June.2001.

Fig. 6: Gedaref - July.2001

Fig. 7: Gedaref – Aug.2001

Fig. 8: Gedaref – Sep.2001.

Fig. 9: Gedaref – Oct.2001.

Fig. 10: VIG indices of July for the years 2001 to 2013
RESULTS AND DISCUSSION

The neural network models are trained with different size of training and validating sets. Selection the size of training and validating depends on the results without specific percentage. The models are differentiated by their architecture and number of input samples. Topology is obtained after training procedure by trial and error. Radial Basis Function network is used for developing Sorghum yield forecasting model with the following parameters values:

- SPREAD = 100.
- The transfer function in the hidden layer is Gaussian function with a linear activation function in the output layer.
- GOAL specified with 0.000001 except Model 9 (August and September model), GOAL specified with 0.0000001

Graphical representation of the five selected models out of nine models. Table 1 shows the model names description.

<table>
<thead>
<tr>
<th>Model name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1 - JAS</td>
<td>Model 1 - Jul, August and September indices</td>
</tr>
<tr>
<td>M2 - JASO10</td>
<td>Model 2 - Jul, August, September and 10th October indices</td>
</tr>
<tr>
<td>M3 - JASO13</td>
<td>Model 3 - Jul, August, September and 15th October indices</td>
</tr>
<tr>
<td>M4 - JASO</td>
<td>Model 4 - Jul, August, September and October indices</td>
</tr>
<tr>
<td>M5 - JAS</td>
<td>Model 5 - Jun, Jul, August and September</td>
</tr>
<tr>
<td>M6 - JASO15</td>
<td>Model 6 - Jun, Jul, August and 15th October indices</td>
</tr>
<tr>
<td>M7 - JAS</td>
<td>Model 7 - Jun, Jul, August and September</td>
</tr>
<tr>
<td>M8 - JJA</td>
<td>Model 8 - Jun, Jul and August</td>
</tr>
<tr>
<td>M9 - JAS</td>
<td>Model 9 - August and September</td>
</tr>
</tbody>
</table>

3.1 Model 2 result:
Input: Jul, Aug and Sep (JAS).
Output: Sorghum yield (Sacks/Feddan)
Size of training set: 9
Size of validating set: 4
Number of neurons in the hidden layer: 9
R = 0.77

MODEL 3 RESULT:
Output: Sorghum yield (Sacks/Feddan)
Size of training set: 10
Size of validating set: 3
Number of neurons in the hidden layer: 9
R = 0.87

MODEL 7 RESULT
Input: Jun, Jul, Aug, Sep, Oct31 (JJASO31).
Output: Sorghum yield (Sacks/Feddan)
Size of training set: 10
Size of validating set: 3
Number of neurons in the hidden layer: 9
R = 0.74
MODEL 8 RESULT

Input: Jun, Jul, Aug, Sep, Oct31 (JJAS).
Output: Sorghum yield (Sacks/Feddan)
Size of training set: 7
Size of validating set: 6
Number of neurons in the hidden layer: 7
R = 0.77

3.5 MODEL 9 RESULT:

Input: Aug and Sep (AS).
Output: Sorghum yield (Sacks/Feddan) - SY
Size of training set: 8
Size of validating set: 5
Number of neurons in the hidden layer: 8
R = 0.82

CONCLUSION

Neural network models of Sorghum production at rain-fed area (Gedatref) are developed using VIg image indices as input and Sorghum yield as output. The model architectures are varied based on the number of the inputs. RBF network with prestd function can be adopted for Sorghum yield forecasting accompanied with architectures that give high values of R. As the selected models using average of vegetation indices in the Sorghum production locations gives high values of R therefore we can continue adopting the average vegetation indices for Sorghum forecasting models. Vegetation indices and ANN can be considered as good base to develop software tool for crop forecasting models in various locations in Sudan, regional and global rain-fed sectors.

REFERENCES

[1] C. M. Bishop, Neural Networks for Pattern Recognition, Clarenndon Press.


