Journal of Agrometeorology 22 (1) : 41-47 (March 2020)

Modeling yield and backscatter using satellite derived biophysical variables of rice crop based on Artificial Neural Networks

MAHESH PALAKURU[1], SIRISHA ADAMALA[2] and HARISH BABU BACHINA[3*]

[1]*School of Civil Engineering, Vellore Institute of Technology (VIT), Vellore, Tamil Nadu*

[2]*Natural Resources Management Division, Central Island Agricultural Research Institute (CIARI), Port Blair, Andaman & Nicobar Islands*

[3]*Department of Applied Engineering, Vignan's Foundation for Science, Technology and Research (VFSTR) University, Vadlamudi*

Corresponding author's email: bachina.harish@gmail.com

ABSTRACT

In this study, 'observed rice yield (ton acre^{-1})' and 'remotely sensed backscatter'are modelled using artificial neural network (ANN) and multiple linear regression (MLR) methods for East and West Godavari districts of Andhra Pradesh in India. The biophysical variables viz. backscatter (bs), normalized difference vegetation index (NDVI), Chlorophyll (chfl), fraction of absorbed photosynthetically active radiation (FAPAR), leaf area index (LAI), canopy water content (CWC), and fraction of vegetation cover (F_{cover}) were derived from Scatterometer Satellite-1 (SCATSAT-1), Moderate Imaging Spectrometer (MODIS) and Sentinel-2 satellite data.Inputs selected are bs, NDVI, chfl, FAPAR, LAI, CWC, and F_{cover} for rice yield model, whereas NDVI, chfl, FAPAR, LAI, CWC, and F_{cover} are inputs for backscatter models. The performance of ANN and MLR models was evaluated using three indices such as root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). The results concluded that the ANN models achieved R^2 of 0.908 and 0.884 which are 42.73% and 28.85% higher than that of the MLR method for rice yield and backscatter, respectively.

Keywords: Rice yield, backscatter, artificial neural network, model, MLR

Rice (*Oryzasativa*) is one of the world's major staple and higher demand foods, especially in India.The rice production as a whole is sensitive to weather fluctuations (Mallick *et al.,* 2007; Dari *et al.,* 2017), under threat due to decrease in its acreage and increase in occurrence of extreme events like floods, droughts, hailstormsetc (Bal and Minhas, 2017). The timely and accurate assessment of rice yield on larger spatial areas is crucial for assessing food supplies and productivity. Biophysical parameters: Chlorophyll (chfl), Leaf Area Index (LAI), Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), Fraction of Cover (F_{cover}), and Canopy Water Content (CWC) extracted from satellite data play an important role in mapping therice and other crops' yield (Bal *et al.,* 2013; Haldar *et al.,* 2014; He *et al.,* 2018). Sentinel-2 multi-spectral imagery (MSI) provided by the European space agency is used to extract biophysical parameters of the rice crop.

Scatterometer Satellite-1 (SCATSAT-1) is a miniature satellite launched by Indian Space Research Organisation (ISRO)provides Ku band scatterometer data with two polarisations. Scatterometer sensor works with 13.55 GHz frequency with 2 km spatial resolution and it provides continuous day and night data to the user community.Sigma horizontal (SH) and sigma vertical (SV) are the products of SCATSAT-1,where 'SH' is used to extract the backscatter (bs) values.

Several authors have used satellite data to extract biophysical parameters, microwave backscatter and normalized difference vegetation index (NDVI) for mapping rice yield.Oza*et al.* (2007) derived information on important phenological phases of rice crop using QuikSCAT Ku band scatterometer and Special Sensor Microwave/Imager (SSM/I) passive microwave radiometer data. Lopez-Sanchez *et al.* (2011) investigated the potential of polarimetric synthetic aperture radar (SAR) imagery with X-band for identifying the different phenological stages of rice fields. This approach exploits the known sensitivity of polarimetry to the structure or morphology of the observed scene to differentiate the physical changes and conditions followed by rice crops during its growth cycle. Kotani *et al.* (2017) used NDVI to

describe the temporal variation of the photosynthetic activity. Palakuru*et al.* (2019) and Palakuru and Yarrakula (2019 a, b) identified rice phenology stages and mapped of rice yield using SCATSAT-1, Moderate Imaging Spectrometer (MODIS) NDVI, and soil map active passive (SMAP) L-band data. Symmetric distribution and anomaly method were used to define phenology of the crop and multiple linear regression (MLR) model was used to map rice yield.

Measuring rice yield at every point location and modeling backscatter values using conventional methods is often tedious and time consuming. Artificial Neural Networks (ANNs) are suitable models to overcome the limitations of conventional methods (Adamala *et al.,* 2014). ANN models have found potential application in agricultural and hydrological modelling (Adamala *et al.,* 2015a,b). Zhang *et al.* (2018) mapped the rice phenology using ANN approach in China with Landsat 8 satellite imagery. They have used land surface temperature (LST) and NDVI as auxiliary input data. The above all studies confirms the potentiality of ANN models to estimate rice crop yield using satellite derived biophysical parameters as input.

The East and West Godavari district of Andhra Pradesh is dominated with rice cultivated area due to the rich source of Godavari river. The measurement of rice yield andextraction of backscatter data at large spatial area of Godavari basin is possible through remote sensing and geographical information system (GIS) integration. Further, their accurate estimation is achieved using ANN model. Therefore, in this study SCATSAT-1, MODIS and Sentinel-2 satellite derived biophysical variables are used as input in ANN models to predict the rice yield and backscatter for the Godavari basin.

MATERIALS AND METHODS

Study area and description

East and West Godavari districts of Andhra Pradesh are rich in crop production area and are blessed with good water resources. Godavari river is the main source of water for irrigation in the study area. The total geographic area of study area is around 20547 km^2. The study area located in between 80^{o}21'39" E to 82^{o}21'48" E longitude and 16^{o}25'50" N to 18^{o}90'0" N latitude at the seashore of Bay of Bengal. Agriculture is the main source of income to the people and principle crop grown in study area is rice crop. East and West Godavari are occupied with the highest position in rice production in Andhra Pradesh state.

The details of satellite data used in this study is shown in Table 1.SCATSAT-1 scatterometer data is in microwave wavelength region, which ranges from 0.7 cm to 100cm in the EMR spectrum. The 'bs'and NDVI values used in this study were extracted from SCATSAT-1and MODIS thematic map (TM) sensor data, respectively.Sentinel-2 sensor datawere used to extract biophysical parameters (chfl, LAI, FAPAR, F_{cover}, and CWC) of the rice crop.

SCATSAT-1 sensor provides data in non image farmat with backscatter values as digital number (DN), which have no units. Therefore, DN values were converted into decible (dB) to use for thisstudy. The DN to dB is converted as below:

$$dB = (DN \times scale) + offset \quad (1)$$

Artificial neural networks (ANNs)

ANNs are based on the highly interconnected structure of brain cells. This approach is fast and robust in noisy environments, flexible in the range of problems it can solve, and highly adaptive to newer environments. ANNs currently have numerous real-world applications, such as time series prediction, rule-based control and hydrological modeling. The multilayer feedforward neural network consists of a set of sensory units that constitute the input layer, one or more hidden layers of computation nodes and an output layer of computation nodes. The input signal propagates through the network in a forward direction, layer by layer. These neural networks are commonly referred to as multilayer perceptrons. A typical three-layer feedforward ANN is shown in Fig.1. The mathematical form of a three-layer feedforward ANN is given as:

$$O_k = g_2\left[\sum_j w_{kj} g_1\left(\sum_i w_{ji} I_i + w_{jo}\right) + w_{ko}\right] \quad (2)$$

where O_k= output at node k of the output layer, g_1 and g_2 = activation functions for the hidden and output layers respectively, I_i = input value to node i of the input layer, V_j = hidden value to node j of the hidden layer, O_k = output at node k of the output layer,w_{ji}= weight controls the strength of connection between the input node i and hidden node j, w_{kj}= weight controls the strength of connection between the hidden node j and output node k,I_o = input layer bias term (=1) with bias weights w_{jo} and V_o= output layer bias term (=1) with bias weights w_{ko}.

The Levenberg–Marquardt (LM) algorithm is a

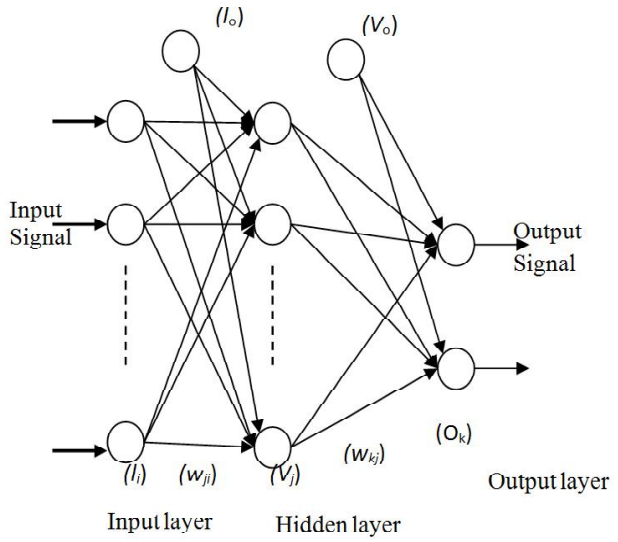

Fig. 1: A basic overview of artificial neural network topology

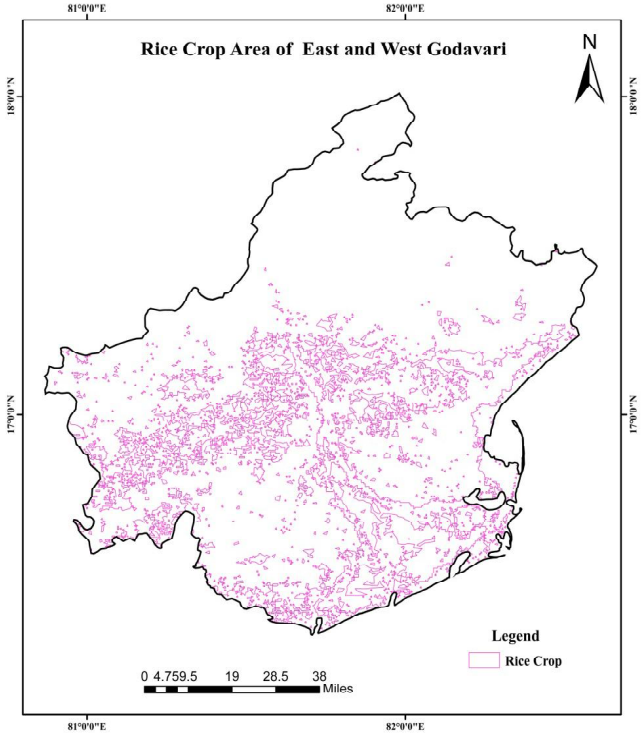

Fig. 2: Rice crop area map generated for Godavari command area

modification of the classic newton algorithm for finding an optimum solution to a minimization problem. It is designed to approach second-order training speed and accuracy without having to compute the Hessian matrix (H). Hessian matrix contains second derivative of network errors with respect to network weights and biases. Jacobeans matrix (J) contains first derivative of the network error matrix with respect to weights and biases. When the performance function has the form of a sum of squares (as is typical in training feed-forward networks), then the Hessian matrix can be approximated as $H = J^T J$ (3)

The gradient (g) can be computed as: $g = J^T e$ (4)

Where e = matrix of network errors. The 'J' can be computed through a standard back propagation technique that is much less complex than computing 'H'. Thus, in LM algorithm, H is approximated in terms of J following Newton-like update in following way: $w_{k+1} = w_k - [J^T J + \mu I]^{-1} J^T e$ (5) $J = \nabla E$ (6)

Where, w = weights of neural network; μ = learning parameter that controls the learning process. When μ is zero, this is just Newton's method, using the approximate H. When μ is large, this becomes gradient descent with a small step size. Newton's method is faster and more accurate near an error minimum, so the aim is to shift towards Newton's method as quickly as possible. Overall, second-order nonlinear optimization techniques are usually faster and more reliable. Therefore, in this study, LM algorithm was used for multi-layer perceptron training. To reduce the impact of the weight initialization on model results, each training process is repeated three times from different random starting points.

Multiple linear regression (MLR) model

MLR is a conventional approach in statistical modeling, which models the relationship between one dependent and number of independent variables. The general form of the MLR model is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_{n-1} x_{n-1} + \beta_n x_n \quad (7)$$

where y = dependent variable that is represented as a function of n independent variables (x); $^{2}_{0}$ = unknown intercept; $^{2}_{1}$, $^{2}_{2}$,..., $^{2}_{n-1}$, $^{2}_{n}$ = partial regression coefficients of the function.

Evaluation of models

To evaluate the performance of the developed ANN and MLR models, statistical analysis was done involving the root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(T_i - O_i)^2} \quad (8)$$

Where T_i and O_i = observed and estimated values at the i^{th} step, respectively; n = number of data points.

Table 1: Satellite data used to derive biophysical parameters

Satellite	Sensor	Row/Path/Orbit	Date of passing	Spatial resolution
SCATSAT-1	Scatterometer	-	2018-02-28	2 km
MODIS	TM	139/44	2018-02-28	500 m
Sentinel-2	MSI	R119	2018-02-25	20 m

Table 2: Performance of ANN and MLR models for rice yield and backscatter estimation

Parameter	Performance	Training		Testing	
		ANN	MLR	ANN	MLR
Yield (ton/acre)	RMSE	0.031	0.119	0.049	0.177
	MAE	0.025	0.096	0.047	0.146
	R^2	0.969	0.570	0.908	0.520
Backscatter	RMSE	0.013	0.936	0.170	0.878
	MAE	0.006	0.651	0.145	0.831
	R^2	0.993	0.511	0.884	0.629

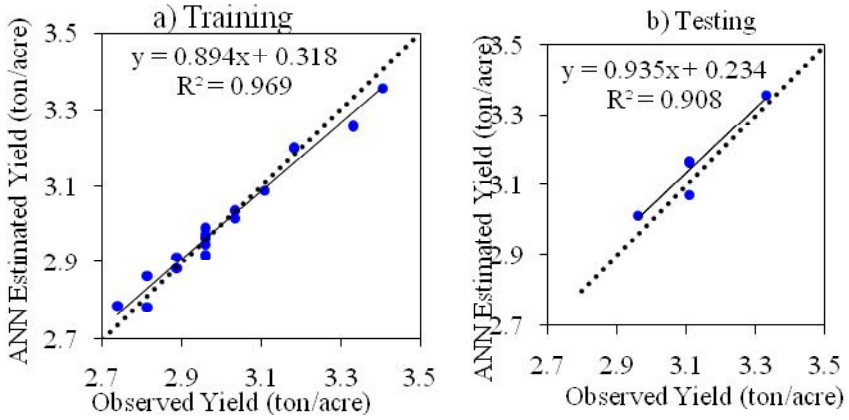

Fig. 3: Scatter plots of observed and ANN estimated yield during a) training and b) testing

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|T_i - O_i| \quad (9)$$

$$R^2 = \frac{\left[\sum_{i=1}^{n}\left(O_i - \overline{O}\right)\left(T_i - \overline{T}\right)\right]^2}{\sum_{i=1}^{n}\left(O_i - \overline{O}\right)^2 \sum_{i=1}^{n}\left(T_i - \overline{T}\right)^2} \quad (10)$$

Where $\overline{T}$ and $\overline{O}$= average of observed and estimatedvalues, respectively.

Model development

The ANN models in this study were based on the most popular form, i.e. a feed-forward network with one hidden layer. The number of input neurons were selected as the number of most significant inputs and the single output neuron was taken as a separate model was developed to predict rice yield and backscatter.The input layer for the ANN yield prediction model consists of seven nodes such as bs, NDVI, chfl, FAPAR, LAI, CWC, and F_{cover} and the output layer consists of one node such as rice yield (ton/acre). Similarly, the input layer for the ANN backscatter prediction model consists of six nodes such as NDVI, chfl, FAPAR, LAI, CWC, and F_{cover} and the output layer consists of one node such as backscatter.

The numbers of hidden neurons were selected based on Adamala *et al.* (2014) research work to calculate the hidden neuron number that should be good enough in capturing the general patterns. In network training, back-propagation is applied using the LM implementation. The logistic sigmoid transfer function is used in the hidden and

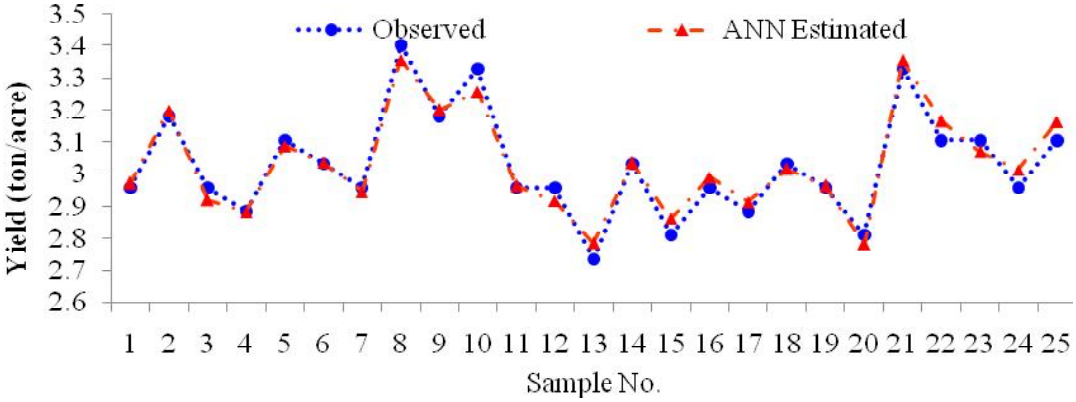

Fig. 4: Comparison of observed and ANN estimated rice yield

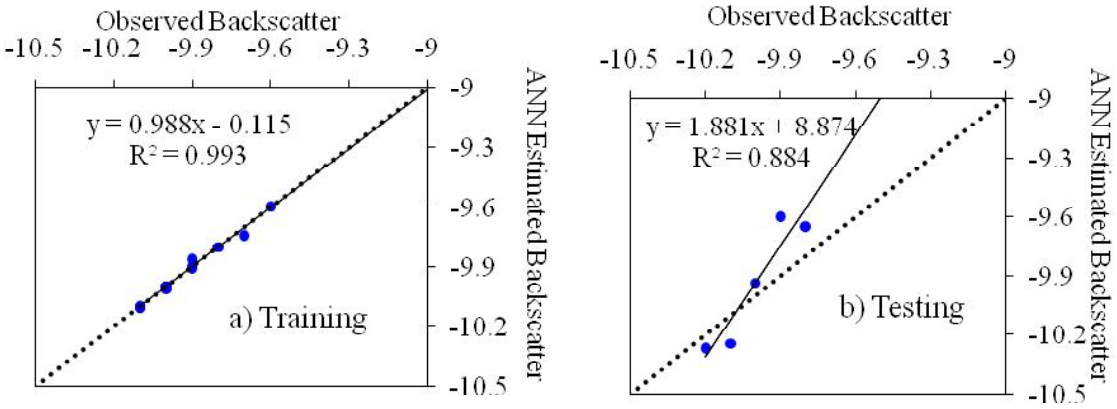

Fig. 5: Scatter plots of observed and ANN estimated backscatter a) training b) testing

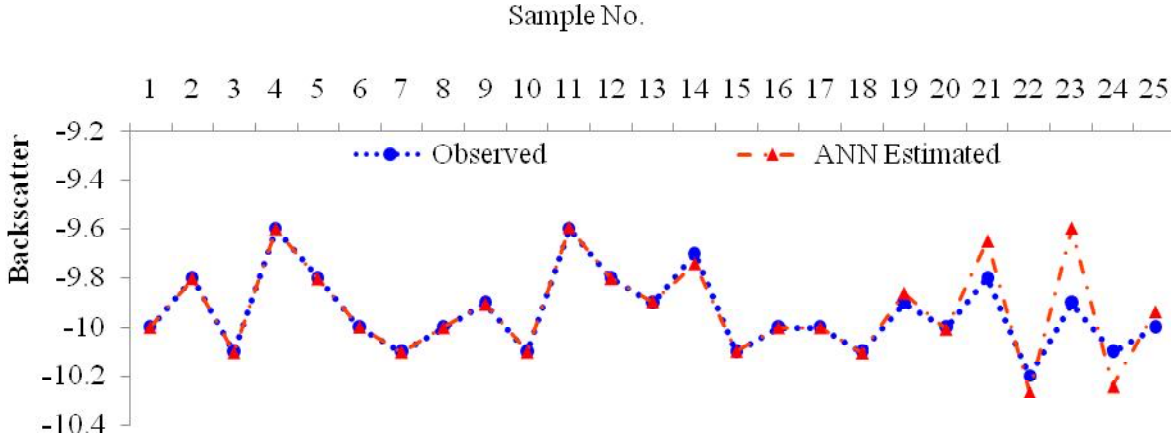

Fig. 6: Comparison of observed and ANN estimated backscatter

output layers. The learning function for weights and bias was based on gradient descent with momentum. Low rates of learning (0.2) and momentum (0.1) were applied in all the model development. The mean squared error was used as the performance function in training. All the observed rice yield and backscatter data were preprocessed and normalized in the range [0, 1] for ANN modeling. For developing the ANN models, the code was written using Matlab 7.0 programming language (The Mathworks, Inc., Natick, Mass.).

RESULTS AND DISCUSSION

Rice crop area maps

The crop acreage for Godavari command area was estimated from the satellite imageries and this was based on the unsupervised classification using Erdas imagine software. The downstream area of study area is dominated with more rice crop area as compared to upstream area (Fig. 2). The biophysical parameters (bs, NDVI, LAI, chfl, FAPAR,

LAI, CWC, and F_{cover}) of rice crop are extracted from the identified rice crop dominant areas.

ANN and MLR models

Table 2 shows the ANN and MLR models performance for estimating rice yield and backscatter. The estimates of RMSE (ton acre^{-1}), MAE (ton acre^{-1}) and R^2 for ANN model were good with values 0.031, 0.025 and 0.969 as compared to MLR model in yield prediction during training. Similarly, ANN models performed well with low RMSE and MAE and high R^2 values as compared to MLR during testing. ANN estimated backscatter models gave low RMSE of 0.049, MAE of 0.047 and high R^2 values of 0.908 as compared to MLR during training and similar was the performance during testing.

Fig.3 (a & b) show the scatter plots of observed and ANN estimated yield during training and testing, respectively. ANN estimated yield agreed well with the observed yield and are distributed evenly on both sides of the 1:1 line. The fit line equations in Fig.3 gave the values of a_0 and a_1 coefficients closer to one and zero, respectively. It is revealed that the spread of ANN estimated yield during testing around 1:1 line was less than that during training. The ANN estimated yield were well agreed with the observed values with high values of R^2 (0.908). The above results indicated that agreement of ANN model with the observed values was good.

Fig.4 show the time series plot of best performed ANN and observed rice yield. This figure illustrates the close relationship between observed and ANN estimated rice yield. The ANN estimated rice yield values were exactly superimposed with the observed values.

Fig. 5 (a &b) show the scatter plots of observed and ANN estimated backscatter during training and testing, respectively. ANN estimated backscatter values agreed well with the observed backscatter values and are distributed evenly on both sides of the 1:1 line. It is revealed that the spread of ANN estimated backscatter during testing around 1:1 line was less than that during training. This is due to the use of less data points for testing as compared to training. The ANN estimated backscatter values were well agreed with the observed values with high values of R^2 (= 0.884). Fig. 7 show the time series plot of best performed ANN and sensor derived backscatter. This figure illustrates the close relationship between observed and ANN estimated backscatter.

CONCLUSION

Satellite input data (bs, NDVI, chfl, FAPAR, LAI, CWC, and F_{cover}) from SCATSAT-1, MODIS and Sentinel-2 were used to estimate backscatter and rice crop yield in Godavari irrigation command area using ANN models. The feed-forward ANN models were trained number of times with LM training algorithm to optimize training parameters (learning and momentum rate, hidden neurons and layers, error, etc,). Both the ANN based backscatter and rice yield models were compared with the MLR models using evaluation indices such as RMSE, MAE and R^2 values. The results concluded that the ANN models achieved R^2 of 0.908 and 0.884, which are 42.73% and 28.85% higher than that of the MLR method for rice yield and backscatter, respectively. Similarly, ANN based rice yield and backscatter models gave better performance in terms of low RMSE and MAE values as compared to MLR models. This study demonstrating that the potential of remote sensing, GIS and ANN modeling in large scale rice mapping using moderate spatial resolution satellite data.

REFERENCES

Adamala, S., Raghuwanshi, N.S., Mishra, A. and Tiwari, M.K. (2014a). Evapotranspiration modeling using second-order neural networks. *J.Hydrol. Eng.*, 19(6):1131-1140.

Adamala, S., Raghuwanshi, N.S. and Mishra, A. (2015a). Generalized quadratic synaptic neural networks for ET_o modeling. *Environ.Process.*,2(2): 309-329.

Adamala, S., Raghuwanshi, N.S., Mishra, A. and Tiwari, M.K. (2015b). Closure to evapotranspiration modeling using second-order neural networks. *J.Hydrol. Eng.*,20(9): 07015015.

Bal, S.K., Choudhury, B.U., Sood, A., Saha, S., Mukherjee, J., Singh, H., and Kaur, P. (2013). Relationship between leaf area index of wheat crop and different spectral indices in Punjab. *J. Agrometeorol.,* 15(2): 98-102.

Dari, B., Sihi, D., Bal, S.K. and Kunwar, S. (2017). Performance of direct-seeded rice under various dates of sowing and irrigation regimes in semi-arid region of India. *Paddy Water Environ.,* 15: 395-401. doi - 10.1007/s1003-016-0557-8

Haldar, D., Patnaik, C. and Chakraborty, M.(2014). Jute crop discrimination and biophysical parameter monitoring using multi-parametric SAR data in West Bengal, India. *Open Access Libr. J.*, 1(e817): 1–11.

He, Z., Li, S., Wang, Y., Dai, L. and Lin, S.(2018). Monitoring rice phenology based on backscattering characteristics of

multi-temporal RADARSAT-2 datasets. *Remote Sens.*, 10(2):340, DOI: 10.3390/rs10020340.

Kotani, A., Hiyama, T., Ohta, T., Hanamura, M., Kambatuku, J.R., Awala, S.K. and Iijima, M. (2017). Impact of rice cultivation on evapotranspiration in small seasonal wetlands of north-central Namibia. *Hydrol. Res. Lett.*, 11(2): 134-140.

Lopez-Sanchez, J.M., Cloude, S.R. and Ballester-Berman, J.D. (2011). Rice phenology monitoring by means of SAR polarimetry at X-Band. *IEEE Trans. on Geosci. Remote Sens.*, 50(7): 2695-2709.

Mallick, K., Mukherjee, J., Bal, S.K., Bhalla, S.S and Hundal, S.S. (2007). Real time rice yield forecasting over central Punjab region using crop weather regression model. *J. Agrometeorol.,* 9(2): 158-166.

Oza, S.R., Panigrahy, S. and Parihar, J.S. (2008). Concurrent use of active and passive microwave remote sensing data for monitoring of rice crop. *Int. J. Appl. Earth Obs. Geoinform.*, 10(3): 296-304.

Palakuru, M., Yarrakula, K., Chaube, N.R., KhadarBabu, S.K. and Rao, Y.S. (2019). Identification of rice crop phenological parameters using dual polarized SCATSAT-1 (ISRO, India) scatterometer data. *Environ. Sci. Pollut. Res.*, 26(2): 1565-1575.

Palakuru, M. and Yarrakula, K. (2019a). Study on paddy phenomics ecosystem and yield estimation using space-borne multi sensor remote sensing data. *J. Agrometeorol.*, 21(2): 171-175.

Palakuru, M. and Yarrakula, K. (2019b). Study on paddy phenomics eco-system and yield estimation using multi-temporal remote sensing approach. *Indian J. Ecol.*, 46(2):293-297.

Zhang, M., Lin, H., Wang, G., Sun, H. and Fu, J. (2018). Mapping paddy rice using a convolutional neural network (CNN) with Landsat 8 datasets in the Dongtinglake area, China. *Remote Sens.*, 10(11):1840, DOI:10.3390/rs10111840.

Received : November 2019 ; Accepted : January 2020