AATCC Review

07 July 2024: Received
29 July 2024: Revised
03 September 2024: Accepted
05 October 2024: Available Online

https://aatcc.peerjournals.net/

Original Research Article

Open Access

Pre-harvest Forecasting of Paddy Yield Based on Weather Parameters Using Different Statistical Methods for Middle Gujarat

Patel N. D.[1] Shukla X. U.[2], Patel H. K[3]. and Parmar D. J.[4]

[1]Department of Basic Sciences & Humanities, BACA, AAU, Anand, India
[2]College of Agricultural Information Technology, AAU, Anand, India
[3]Regional Research Station, AAU, Anand, India
[4]Department of Agricultural Statistics, BACA, AAU, Anand, India

ABSTRACT

An attempt to forecast paddy yield based on weather parameters, at the pre-harvest stage has been documented in this paper for middle Gujarat. Different supervised machine learning models have been tested under the study namely Linear Regression, Polynomial Regression and Neural Network. Weather variables such as maximum temperature, minimum temperature, average relative humidity, sunshine hours and accumulated rainfall have been initially used on a fortnightly basis for the period from 1980–81 to 2018–19 to train these models. Subsequent feature selection suggests that the minimum temperature and relative humidity (morning & evening) play important roles as predictors over others. Further, 37th and 39th as harvested weeks were identified with the lowest error for Linear Regression and Neural Network models respectively.

Keywords: *Machine Learning, Neural Networks, Linear Regression, polynomial Regression, Paddy and yield*

INTRODUCTION

Since three billion people consume rice and it delivers 35 to 60% more calories to humans than any other crop, rice is one of the most significant food crops farmed worldwide. After the start of the green revolution, a vast rise in area coverage, output, and productivity was seen. The state's ongoing adoption of the rice-wheat cropping system has resulted in extensive usage of agrochemicals, which eventually lowers the yield from rice production.

To predict crop production, several academics have attempted to create pre-harvest forecast models based on meteorological data.Crop yield has been influenced by different weather parameters with different relationships. Therefore, production of maximum crop yield it is necessary to select weather factors with the highest correlation with yield while using crop yield modeling [2] [6]. Therefore, it will be crucial to establish a mechanism for predicting rice yield. A large portion of India's population depends on seasonal crops cultivated in rainy environments for its food supply.With the objective of Selection weather parameters for pre-harvest forecasting of paddy yield and comparison of the different statistical methods for forecasting of paddy yield, the entitled study on Pre-harvest forecasting of paddy yield based on weather parameters using different statistical methods for middle Gujarat was carried out.

SOURCE OF DATA

The actual paddy yield data and meteorological data from the year 1988 to 2018 was collected from the Directorate of Agriculture, Gandhinagar Gujarat and weather data like Maximum temperature, Minimum Temperature, Rainfall, Morning Relative Humidity, Evening Relative Humidity from the Department of Agricultural Meteorology, B.A. College of Agriculture, AAU, Anand, respectively.

MATERIALS AND METHOD

Here, three different models to predict the paddy yield Stepwise Regression, Polynomial Regression and Artificial Neural Network were fitted using the data series of paddy yield & weather parameters from 1988-2021. All three models have predicted the paddy yield of the year 2016 to 2021.

1. Polynomial Regression

Quadratic Regression Approach (Montgomery et al., 2003)

The fitted regression equation was as under

$$Y = a + bt + ct^2$$

The unknown parameter viz, a, b and c were estimated by using the 'Principle of least square' method.

2. Artificial Neural Network:

The multilayer perceptron (MLP) model belongs to feed forward neural Network. In terms of functionality. Additionally Multilayer perceptron have been proven to be able to approximate any continuous function by adjusting the number of nodes in the hidden layer [5], with numerous cases of successful applications [3]. Figure 1 illustrates the general structure of a three-layer MLP with one hidden layer of L nodes, a p-dimensional input vector X, and a q-dimensional output vector Y. The relationship between the input and output components for this MLP can be generally expressed as

$$y_k = \varphi\left(\sum_{j=1}^{L} b_{kj}\,\psi\left(\sum a_{ji}\,x_i\right)\right)$$

Corresponding Author: **Patel N. D.*

DOI: https://doi.org/10.21276/AATCCReview.2024.12.03.383

where φ and ψ are the transfer functions, $_{ji}a$ denotes the input to hidden layer weights at the hidden neuron, j and $_{kj}b$ is the hidden to output layer weights at the output unit k.

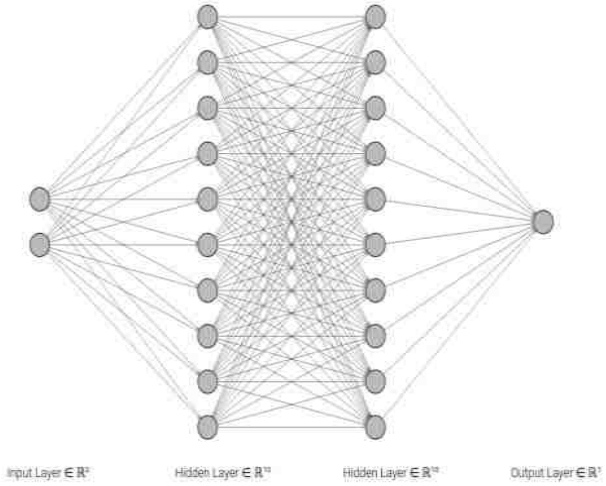

Fig.1a. General Structure for 13&15th week
(10 node x 2 hidden layers, 2 Input node)

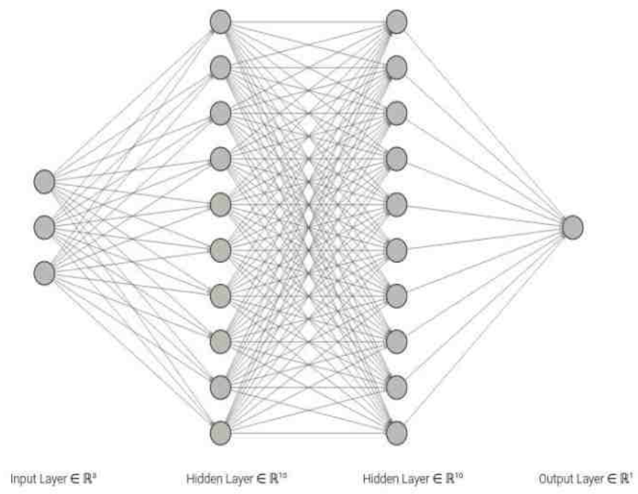

Fig.1b. General Structure for 17th week
(10 node x 2 hidden layers, 3 Input node)

RESULTS

For each weather variable, two weather indices were developed, one as simple accumulation of weather variables and the other one as the weighted accumulation of weekly weather variable, weights being correlation coefficients of weather variables in respective weeks with yield (adjusted for trend effect, if present). Similarly, for interaction of weather variables, indices were generated using weekly products of weather variables taking two at a time. A stepwise regression technique was used to select the important weather indices.

Table 1: Variables used in Model development and their description

Weather Parameter	Unweighted coefficients	Weighted coefficients
Tmax	Z10	Z11
Tmin	Z20	Z21
RH1	Z30	Z31
RH2	Z40	Z41
RF	Z50	Z51

The crop yield forecasting models were developed to forecast paddy yield for Anand district.Stepwise regression analysis, Polynomial Regression and Artificial Neural Networks were used with weather parameters like as Maximum Temperature (Tmax), Minimum Temperature (Tmin), Relative Humidity (RH1&RH2) and Rainfall(RF) for 35, 37and 39th weeks as input variables.The independent variables of weather parameters were developed by using unweighted and weighted with correlation as shown in Table 1.

Table 2: Developed Different Stepwise Regression Models for Three Different Weeks

<table>
<tr><th>Sr. No.</th><th>Districts Anand</th><th colspan="5">Regression Formula</th></tr>
<tr><td rowspan="5">1</td><td rowspan="5">23- 35th Week</td><td>Coefficients</td><td>Estimate</td><td>Std. Error</td><td>t value</td><td>Pr(>|t|)</td></tr>
<tr><td>Intercept</td><td>-4108.11</td><td>3042.81</td><td>-1.35</td><td>0.1895</td></tr>
<tr><td>TIME</td><td>29.06</td><td>7.48</td><td>3.88</td><td>0.0007***</td></tr>
<tr><td>Z41</td><td>27.34</td><td>7.61</td><td>3.59</td><td>0.0014**</td></tr>
<tr><td>Z20</td><td>15.49</td><td>9.00</td><td>1.72</td><td>0.09829</td></tr>
<tr><td rowspan="5">2</td><td rowspan="5">23- 37th Week</td><td>Coefficients</td><td>Estimate</td><td>Std. Error</td><td>t value</td><td>Pr(>|t|)</td></tr>
<tr><td>Intercept</td><td>-7823.45</td><td>2728.92</td><td>-2.86</td><td>0.00849 **</td></tr>
<tr><td>TIME</td><td>28.20</td><td>6.96</td><td>4.04</td><td>0.00046 ***</td></tr>
<tr><td>Z31</td><td>41.25</td><td>10.22</td><td>4.03</td><td>0.00048 ***</td></tr>
<tr><td>Z10</td><td>8.61</td><td>4.64</td><td>1.85</td><td>0.07612</td></tr>
<tr><td rowspan="6">3</td><td rowspan="6">23- 39th Week</td><td>Coefficients</td><td>Estimate</td><td>Std. Error</td><td>t value</td><td>Pr(>|t|)</td></tr>
<tr><td>Intercept</td><td>-4621.47</td><td>1780.09</td><td>-2.59</td><td>0.01615 *</td></tr>
<tr><td>TIME</td><td>26.47</td><td>8.06</td><td>3.28</td><td>0.00325 **</td></tr>
<tr><td>Z31</td><td>21.26</td><td>8.97</td><td>2.37</td><td>0.02652 *</td></tr>
<tr><td>Z11</td><td>67.41</td><td>33.73</td><td>1.99</td><td>0.05768</td></tr>
<tr><td>Z40</td><td>0.58</td><td>0.44</td><td>1.31</td><td>0.20286</td></tr>
</table>

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

As per Table 2, the models were developed using unweighted and weighted variables as independent variable and dependent variable yield (Y) of paddy. For 35th week the partial regression coefficient of time, Z41(RH2) and Z20(Tmin) showed a significant impact on the yield of paddy. In 37th week the regression coefficient Z31(RH1), time and Z10(Tmax) showed significant impact on yield. Also, in 39th week Z40(RH2), time, Z11(Tmax) and Z31(RH1) showed significant impact on yield.

Table 3: Actual & Predicted yield for the year 2016-18 of Three Different Weeks using stepwise Regression

Time Period	Year	Actual Yield (kg/ha)	Predicted Yield (kg/ha)	Absolute Difference (kg/ha)	Prediction Percentage Error (FPE)	Mean Percentage Prediction Error (MPPE)
23-35th Week	2016	2745.30	2415.06	330.24	12.03	16.37
	2017	2706.53	2267.65	438.88	16.21	
	2018	2792.83	2209.90	582.93	20.87	
23- 37th Week	2016	2745.30	2784.84	39.54	1.44	3.39
	2017	2706.53	2800.62	94.09	3.47	
	2018	2792.83	2645.65	147.18	5.26	
23- 39th Week	2016	2745.30	3023.99	278.69	10.15	9.35
	2017	2706.53	3074.91	368.38	13.61	
	2018	2792.83	2912.99	120.16	4.30	

Predicted yield by Stepwise Regression equations for the year 2016-18 of three different weeks were shown in Table 3. Here, the best yield prediction with a coefficient of determination (R^2) was 0.68 in the 37th week followed by 35th (0.64) and 39th (0.67) week.

As per Table 4, the models were developed using unweighted and weighted variables as independent variables and dependent variable yield (Y) of paddy. For 35th week the partial regression coefficient of Z41(RH2) and Z20(Tmin) showed a significant impact on the yield of paddy. In 37th week the regression coefficients Z31(RH1) and Z10(Tmax) showed significant impact on yield. Also, in 39th week Z40(RH2) and Z11(Tmax) showed a significant impact on yield.

Table 4: Developed Different Polynomial Regression Models for Three Different Weeks

Sr. No.	Districts Anand	Regression equation
1	23- 35th Week	$Y = -79079.4 + (400.39 * Z20) + (908.63 * Z41) + (-0.48 * Z20^2) +(-2.52 * Z20 * Z41) + (-0.76 * Z41^2)$
2	23- 37th Week	$Y = -17843 + (102.94 * Z31) + (26.90 * Z10) +(-0.1025 * Z31 * Z10)$
3	23- 39th Week	$Y = -4548.54 + (104.16 * Z11) + (-0.42 * Z40) +(0.038 * Z11 * Z40)$

Predicted yield by Polynomial Regression equations for the year 2016-18 of three different weeks were shown in Table 5. Here, best yield prediction with coefficient of determination (R^2) was 0.48 in the 35th week followed by 37th (0.47) and 39th (0.45) week.

Table 5: Actual & Predicted yield for the year 2016-18 of Three Different Weeks using Polynomial Regression

Time Period	Year	Actual Yield (kg/ha)	Predicted Yield (kg/ha)	Absolute Difference (kg/ha)	Prediction Percentage Error (FPE)	Mean Percentage Prediction Error (MPPE)
23-35th Week	2016	2745.30	2299.70	445.60	16.23	7.38
	2017	2706.53	2696.66	9.87	0.36	
	2018	2792.83	2637.18	155.65	5.57	
23- 37th Week	2016	2745.30	3484.23	738.93	26.91	24.76
	2017	2706.53	3496.88	790.35	29.20	
	2018	2792.83	3300.88	508.05	18.19	
23- 39th Week	2016	2745.30	3432.95	687.65	25.04	22.76
	2017	2706.53	3463.13	756.60	27.95	
	2018	2792.83	3219.85	427.02	15.28	

Predicted yield by Artificial Neural Network for the year 2016-18 of three different weeks were shown in Table 6. Here, best yield prediction with less mean percentage prediction error was 2.84 in the 37th week followed by 39th (2.93) and 35th (4.01) week.

Table 6: Actual & Predicted yield for the year 2016-18 of Three Different Weeks usingArtificial Neural Network

Time Period	Year	Actual Yield (kg/ha)	Predicted Yield (kg/ha)	Absolute Difference (kg/ha)	Prediction Percentage Error (FPE)	Mean Percentage Prediction Error (MPPE)
23-35th Week	2016	2745.30	2917.97	172.67	6.28	4.01
	2017	2706.53	2808.39	101.86	3.76	
	2018	2792.83	2736.43	56.4	2.01	
23- 37th Week	2016	2745.30	2793.62	48.32	1.76	2.84
	2017	2706.53	2875.93	169.41	6.25	
	2018	2792.83	2778.24	14.59	0.52	
23- 39th Week	2016	2745.30	2684.18	61.12	2.22	2.93
	2017	2706.53	2681.59	24.93	0.92	
	2018	2792.83	2634.29	158.54	5.67	

Table 7: Comparsion of Actual & Predicted yield for the year 2016-21 of 35th week by Stepwise Regression Analysis / Polynomial Regression / Artificial Neural Network

Time Period	Year		Polynomial Regression	Stepwise Regression Analysis	Neural Network
		Actual (kg/ha)	Predicted (kg/ha)	Predicted (kg/ha)	Predicted (kg/ha)
23-35th Week	2016	2745.30	2299.70	2415.06	2917.97
	2017	2706.53	2696.66	2267.65	2808.39
	2018	2792.83	2637.18	2209.90	2736.43
	2019	1835.49	2420.89	2347.15	2855.71
	2020	2422.87	2266.46	2219.78	2653.13
	2021	2532.22	2073.48	2451.09	2920.15

Table 8: Comparsion of Actual & Predicted yield for the year 2016-21 of 37th week by Stepwise Regression Analysis / Polynomial Regression / Artificial Neural Network

Time Period	Year		Polynomial Regression	Stepwise Regression Analysis	Neural Network
		Actual (kg/ha)	Predicted (kg/ha)	Predicted (kg/ha)	Predicted (kg/ha)
23-37th Week	2016	2745.30	3484.24	2784.99	2793.63
	2017	2706.53	3496.88	2800.76	2875.94
	2018	2792.83	3300.89	2645.79	2778.24
	2019	1835.49	3096.44	2478.03	2612.07
	2020	2422.87	3262.68	2607.97	2682.49
	2021	2532.22	2931.32	2350.83	2546.14

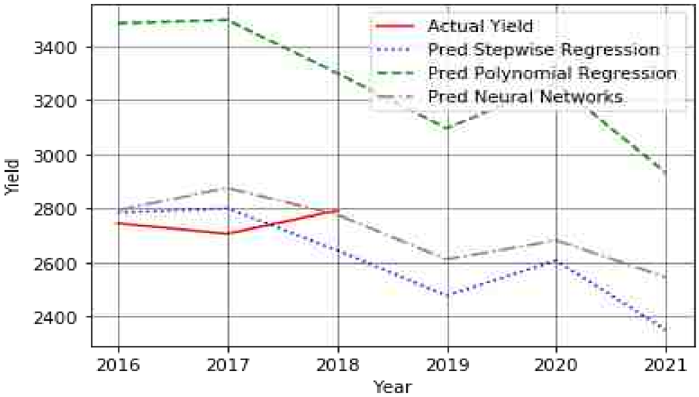

Fig.2. Comparison of Actual & Predicted yield for the year 2016-21 of 15th week by Stepwise Regression Analysis / Polynomial Regression / Artificial Neural Network

In Table 8 and Figure 2 we can easily seen that the forecasted paddy yield by using Artificial Neural Network is near by the actual yield.

Table 9: Comparsion of Actual & Predicted yield for the year 2016-21 of 39th week by Stepwise Regression Analysis / Polynomial Regression / Artificial Neural Network

Time Period	Year		Polynomial Regression	Stepwise Regression Analysis	Neural Network
		Actual yield (kg/ha)	Predicted (kg/ha)	Predicted (kg/ha)	Predicted (kg/ha)
23-39th Week	2016	2745.30	3432.95	3023.96	2684.18
	2017	2706.53	3463.14	3074.87	2681.60
	2018	2792.83	3219.85	2912.95	2634.29
	2019	1835.49	3574.82	2951.01	2622.64
	2020	2422.87	3167.97	2824.35	2571.21
	2021	2532.22	3396.81	2821.99	2555.40

Table 10: Comparsion of Mean Percentage Prediction Error (MPPE) and Root Mean Squre Error (RMSE)by Stepwise Regression Analysis / Polynomial Regression / Artificial Neural Network

Time Period	Polynomial Regression Analysis		Stepwise Regression		Neural Network	
	MPPE	RMSE	MPPE	RMSE	MPPE	RMSE
23-35th Week	7.38	272.56	16.37	462.32	4.01	120.23
23-37th Week	24.76	690.12	3.39	103.40	2.84	102.05
23-39th Week	22.76	639.70	9.35	275.52	2.93	103.28

The Mean Percentage Prediction Error for 37th week was 3.39 followed by 9.35 for 39th week and 16.37 for 35th week by Stepwise Regression Analysis. The Root Mean Square Error for 37th week was 103.40 followed by 275.52 for 39th week and 462.32 for 35th week respectively by Stepwise Regression Analysis.

The Mean Percentage Prediction Error for 35th week was 7.38 followed by 22.76 for 39th week and 16.37 for 37th week by Polynomial Regression. The Root Mean Square Error for 35th week was 272.56 followed by 639.70 for 39th week and 690.12 for 37th week respectively by Polynomial Regression.

The Mean Percentage Prediction Error for the 37th week was 2.84 followed by 2.93 for 39th week and 4.01 for 35th week by Neural Network. The Root Mean Square Error for 37th week was 102.05 followed by 103.28 for 39th week and 120.23 for 35th week respectively by Neural Network.

Thus, the predicted paddy yield computed by ANN have lowest MPPE & RMSE among other two models. Therefore, ANN is a preferable model to predict the yield.

CONCLUSION

It can be concluded that among all three methods like as Stepwise Regression Analysis / Polynomial Regression / Artificial Neural Network the weather parameters the minimum temperature, and relative humidity (morning & evening) play key role as predictor's during crop yield prediction for Anand district. Simultaneously Relative humidity (morning & evening) has a significant relation with paddy yield. As per the Stepwise Regression Analysis and Artificial Neural Network. the 37th week was forecasted as harvested week. Among 35, 37 and 39th number of weeks as forecasted error was minimum for it. Unavailability of actual paddy yield data validation of these models was done for three years 2015-16, 2016-17 and 2017-18 and obtained minimum Mean Percentage Prediction Error (MPPE) in 37th week *i.e.*2.84 and 3.39 of artificial neural Network and Stepwise Regression Analysis respectively. So, we suggested that among all three methods, Artificial Neural Network would be preferable to predict the paddy yield as we found the lowest Mean Percentage Prediction Error (MPPE) among other methods.

DATA AVAILABILITY

The original contributions presented in the study are included in articles/supplementary material. Further inquries can be directed to the corresponding authors.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interest personal relationship that could have appeared to influence the work reported in this paper.

ACKNOWLEDGMENT

The authors are thankful to the Director of Research and Dean PG studies, Anand Agricultural University, Principal and Dean, B.A.College of Agriculture, Anand Agricultural University for approval of work, technical guidance. Authors thanks to Head of the Department, Basic Science & Humanities Department, B.A.College of Agriculture, Anand Agricultural University, Anand.

References

1. Agrawal, R and Mehta, S.C., 2007. Weather based forecasting of crop yields, pest and diseases- IASRI. Models. J. Ind. Soc. Agril. Statist, 61(2): 255-263.

2. Baier, W. 1977. Crop weather models and their use in yield assessment. Tech. note. no. 151, WMO, Geneva, 48 pp.

3. Guo, W.W. (2010), Incorporating statistical and neural network approaches for student course satisfaction analysis and prediction," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3358–3365

4. Handrick, W.A. and Scholl, J.C. (1943). Technique in measuring joint relationship. The joint effects of temperature and precipitation on crop yield. N. Carolina Agric. Exp. Sta. Tech. Bull.

5. Hornik, K., Stinchcombe, M. and White, H. (1989), Multilayer feedforward Network are universal approximators," *Neural Network*, vol. 2, no. 5, pp. 359–366.

6. Koocheki, A., Kamali, G. H. and Banaian, M., 1993. Simulation of primary production. The center of agrobiological research and department of theoretical production ecology, Wageningen, Netherlands. Published by World Meteorological Organization. Geneva, July, (Translated in Persian, Tehran).

