

RESEARCH ARTICLE

Developing Prediction and Forecasting model for coconut under weather parameters and external factors in Coimbatore district

K. Madeshwaran¹, A. Eswari¹, M.R. Duraisamy², M. Praneetha³, B. Senthamizh Selvi⁴

¹Department of PS & IT, AEC & RI, TNAU, Coimbatore,

²Coconut Research Station, Aliyarnagar, Coimbatore,

³Department of Spices and Plantation Crops, HC & RI, TNAU, Coimbatore.

ABSTRACT

Coconut is the most important plantation crop, planted in nearly every country. Coimbatore is the major producer of coconut in Tamil Nadu, followed by Thanjavur and Kanyakumari, hence this research is focused on the Coimbatore district. West Coast Tall is the popular variety which gives better yield than other varieties. This study was carried out on West Coast Tall (WCT) variety. In this paper, Coconut yield prediction model is developed under weather and external factors such as Minimum temperature, Relative Humidity, Rainfall, Plant height, Stem girth, Female flowers in inflorescence, Leaf length and Copra content. Correlation and multiple regression analysis is carried out to obtain the final form of the yield prediction model. Further, we validate this model using field-level data from TNAU coconut research farm using MATLAB and python software. The actual yield in comparison with the predicted yield using software is found to be in satisfactory agreement. In addition, we developed forecasting model for coconut for forthcoming years using proposed model.

Received : 13 November 2023

Revised : 25 November 2023

Revised : 04 December 2023

Accepted: 14 December 2023

Keywords: Coconut; Regression analysis; Field level data; External factors; Simulation.

INTRODUCTION

Coconut tree is known as the 'Tree of Heaven' because every part of the coconut palm may be used in some fashion. Food, drink, shelter, and industrial supplies are all available. Coconut farming has been practiced in India for a long time, and it is an important element of the people's social, economic, and cultural activities. The country has made significant success in raising coconut yield and productivity. India ranks third in the production of coconut in the world. The production of coconut in the country stood at 21207 million nuts during 2020-2021 which is 34% of the global production (Source: Ministry of Agriculture and Farmers Welfare). In India, Tamil Nadu stands second in coconut production with 3751.26 thousand tonnes, contributing around 27.47% to the total production in India (Source: National Horticultural Board). Crop yield is a complicated variable that is influenced by a number of factors including genotype, environment, and their interactions. Accurate yield prediction necessitates a fundamental understanding of the functional link between yield and these interaction components, which

necessitates both large datasets and sophisticated algorithms to uncover. Weather is a possible yield defining factor in agriculture, influencing the growth and development of a crop both directly and indirectly, and affecting around 50% of crop yield. On an average, the coconut palm produces one bunch per month and the bunch requires 38 months to fully ripened (Ranasinghe *et al.*, 2015). Regression analysis between coconut productivity and the number of extreme events during the first four months after the flowering was performed, and the result shows coconut yield was affected by climatic conditions when the inflorescence opened (Pathmeswaran *et al.*, 2018). According to Ranasinghe *et al.*, (2015), yield of coconut is determined by early fruit setting, which might be hampered by adverse weather conditions. Shripad Vishweshwar *et al.*, (2019) employed regression model to predict nut yield using climatic variables as independent factors and nut yield per palm as dependent factors. Primordium initiation and ovarian development stages were more strongly and significantly influenced by weather parameters. To estimate the yield, basic mathematical



or statistical relationships based on agronomic and meteorological data were formed (Dadhwal & Ray, 2000), and also crop development models (Thorp, De Jonge, Kaleita, Batchelor, and Paz, 2008); and remote sensing techniques proved to be an excellent tool for predicting and forecasting of crop yield (Geipel *et al.*, 2014). Naresh Kumar *et al.*, (2007) investigated the combined effect of dry spells and rainfall on coconut production and determined that the presence of a dry spell in one year will reduce coconut yield in succeeding years. Bappa Das *et al.*, (2020) predicted coconut yield using six multivariate techniques and the study concluded that the ELNET method had been more accurate for predicting coconut yield. Benny Antony (2021) predicted coconut yield with respect to rainfall using different regression and machine learning techniques and it was found that random forest regression was more effective in prediction of coconut yield in Kerala. Recently Rohit *et al.* (2023) have analysed the characterization of mother palms and novel techniques to produce elite seedlings of coconut. Though it has previously been explored how to forecast coconut yield using basic and advanced regression models based on meteorological factors. Rainfall, temperature, fertilizers, pesticides, pH level, fruit set, nut weight and other genotypic characteristics influence crop yield. Based on our review, there is no evidence that weather parameters and external factors were combined for the prediction of coconut yield. So, this study was attempted to use both weather parameters and external factors to predict coconut yield more precisely.

MATERIAL AND METHODS

The daily weather parameters were collected in Agro-Climatic Research Centre, Tamil Nadu Agricultural University, Coimbatore, and the external factors were collected year-wise from Coconut Research Station, Aliyar, for the years (2010 - 2016). The daily weather data are taken on an average to form year wise weather data. The primary data for validation had been collected from Coconut Farm, Tamil Nadu Agricultural University for two years (2021 and 2022).

Correlation:

Correlation coefficients are used to quantify the strength of a linear association between two variables, x and y. A linear correlation coefficient greater than zero indicates a positive relationship. A number less than zero indicates a negative relationship. Finally, a value of 0 denotes that the variables x and y are unrelated.

$$\text{Correlation} = \rho = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$$

In this study, the dependent variable is nut yield and the independent variables are plant height, stem girth, female flowers in inflorescence, husk thickness, copra content, minimum temperature, rainfall, and relative humidity. Correlation is carried out between nut yield and all other factors to find the factors which are all highly responsible for multi-collinearity between predictor variables.

Multiple Linear regression:

Multiple regression, or MLR, is a statistical method for predicting the outcome of a response variable by integrating a number of explanatory factors. The linear relationship between explanatory (independent) and response (dependent) variables is attempted to be represented using multiple linear regression. Multiple regression is simply an extension of ordinary least-squares (OLS) regression since it includes more than one explanatory variable.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + e$$

where, for $i=n$ observations:

Y_i =dependent variable

X_i =explanatory variables

β_0 =y-intercept (constant term)

β_p =slope coefficients for each explanatory variable

e =the model's error term (also known as the residuals).

ARIMA:

ARIMA, or autoregressive integrated moving average, is a statistical analysis model that uses time series data to understand the data set better or anticipate future trends. The trend and seasonal components are the most common causes of non-stationarity in time series data. The differencing step is used to convert non-stationary data to stationary data. To eliminate the trend component in the data, one or more times of differencing steps can be used. Autoregressive statistical models anticipate future values based on previous values. The ARIMA methodology is a statistical method for analyzing and developing a forecasting model that best represents a time series by modelling the data correlations. ARIMA smooth time series data using lagged moving averages. The parameters of the ARIMA model is as follows:

P: The number of lag observations included in the model, also called the lag order.

D: The number of times that the raw observations are differenced, also called the degree of differencing.

Q: The size of the moving average window, also called the order of moving average.



Software Used for the Study:

The correlation and multiple linear regression analysis carried out using MATLAB. The simulation of observed and predicted yield had been done using R and SPSS software.

Analysis of Yield Prediction Model for various cases:

a. Coconut yield prediction using Weather parameters:

The multiple linear regression model for weather parameters alone is as follows:

$$Y = 43764.71 - 4899.5 * X_1 + 2016.9 * X_2 + 732.94 * X_3 + 1528.01 * X_4 - 56.02 * X_5 \text{-----(1)}$$

where,

X_1 - Maximum Temperature, X_2 - Minimum Temperature, X_3 - Relative Humidity, X_4 - Wind Speed, X_5 - Rainfall.

The above equation explains the linear relationship between yield and weather parameters. It implies that the Maximum Temperature and Rainfall were having negative impact on coconut yield whereas the other parameters such as Minimum Temperature, Relative humidity and Wind Speed were positively contributed to coconut yield.

b. Coconut yield prediction using External Factors:

The multiple linear regression model for external factors alone is as follows:

$$Y = 1430.793 - 38.91 * X_1 + 2.67 * X_2 + 14.15 * X_3 + 414.81 * X_4 + 7.44 * X_5 \text{-----(2)}$$

where, X_1 -Plant Height, X_2 -Stem girth, X_3 -Husk thickness, X_4 -Female flowers in inflorescence, X_5 - Copra content.

It can be seen that from the above equation Husk Thickness and Female flowers in inflorescence having higher impact on yield followed by Copra content and Leaf length. The yield is having negative impact on Stem girth that means if the stem girth increases, the yield will decrease.

c. Coconut yield prediction using weather parameters and external factors:

The yield prediction model for coconut is obtained by using weather parameters and external factors is as follows.

$$Y = 0 + 0 * X_1 + 19.606 * X_2 + 1.4537 * X_3 - 1.1593 * X_4 - 26.072 * X_5 + 4.8373 * X_6 - 10.928 * X_7 + 7.3649 * X_8 \text{----- (3)}$$

- X_1 - Minimum Temperature
- X_2 - Relative Humidity
- X_3 - Rainfall
- X_4 - Plant Height
- X_5 - Stem Girth
- X_6 - Leaf Length
- X_7 - Female flower in inflorescence
- X_8 - Copra content

From the above equation, it can be seen that the intercept is zero, which means the dependent variables were mostly correlated to the independent variable. It also seen that the Minimum Temperature (X_1) has no effect on nut yield and the variables Relative Humidity (X_2), Rainfall (X_3), Leaf length (X_6) and Copra content(X_8) were positively contributed to the coconut yield. The model also indicates that the variables Plant Height (X_4), Stem girth (X_5) and Female flower in inflorescence (X_7) were negatively contributed to coconut yield.

RESULTS AND DISCUSSION

Figure 1 explains that the variables were selected based on the correlation coefficient between yield and other variables. It is evident that the correlation between yield and leaf length is high (0.88) followed by Copra content (0.67). Like that, we had selected Minimum Temperature, Relative humidity and Rainfall as weather parameters and external factors such as Plant Height, Stem girth, Leaf length, Female flowers in inflorescence and Copra content for the same aspects.

Table 1-3 shows the observed yield and predicted yield for weather parameters alone and external factors alone and for both, respectively. It was observed that the residuals were high in using weather parameters & external factors alone than combined used prediction model.

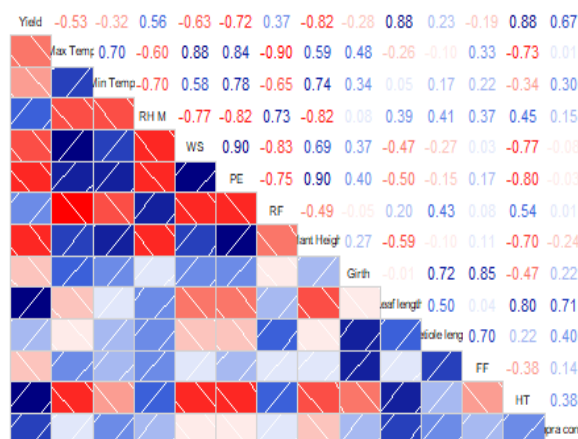


Fig 1. Correlogram showing correlation between yield and all factors



Comparison of Numerical and Analytical result:

Figure 2 and 3 explains the observed yield and predicted yield using weather parameters alone and

external factors alone. It had been observed that the actual and model yield follows similar trend.

Figure 4 shows the observed yield and predicted yield by using weather parameters and external factors. It had been observed that the predicted yield was almost similar to observed yield. The validation had been done for two years that also plotted in this figure. The MATLAB program is also given in Appendix – A.

Figure 5 depicts a box plot of observed and predicted yield. The boxplot implies that there is no outlier in the data. It is evident from the boxplot that the deviation is not high in the observed and predicted yield. As a result, the median of the created model is 1550.422 gms/plant, while the observed yield is 1550.5 gms/plant. As a result, it is suggested that the created model was found to be acceptable for predicting coconut yield in the research region.

Validation of the model using primary data:

The model created using weather parameters and external factors had been validated using primary data collected for two years (2021-2022) in Coconut Farm, Tamil Nadu Agricultural University, Coimbatore. The data was collected in 25 trees of West Coast Tall Variety and the data had taken on an average to form year-wise data. It was found that the validated yield and observed yield is almost nearer (Fig 4.). From Table 4, we found that the model built using weather parameters and external factors were having more R² value (0.99) and less RMSE (3.93) and MAPE (0.16%) values than separately built prediction models. Based on RMSE and MAPE values, it is inferred that the prediction model developed using weather parameters and external factors is more accurate than separately built models.

Forecasting Model:

The coconut yield is forecasted for the next ten years by using Autoregressive Moving Average (ARIMA) model. The ARIMA model for forecasting is selected based on Lower AIC value which is ARIMA (1, 0,0). The equation for ARIMA (1, 0,0) is given by

$$y_t = 1769.53 + 0.0635y_{t-1}$$

The code for forecasting model in python were given in Appendix – B.

where, y_t, y_{t-1} : yield of present year and previous year respectively, t – time in years.

Figure 6 shows the forecasted yield for ten years (2017-2026). It can be seen that the forecasted yield is increasing from year to year and having an upward trend.

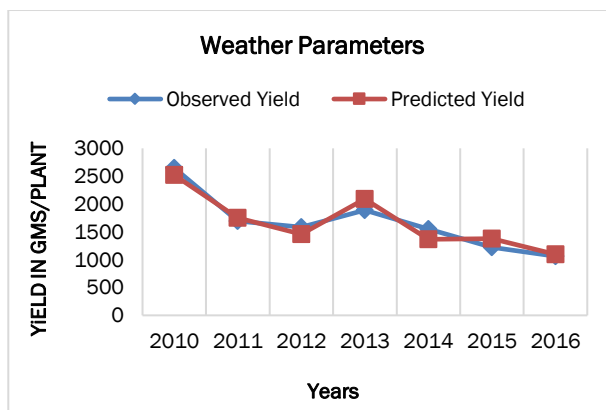


Fig 2. Line chart showing observed and predicted yield using weather parameters alone.

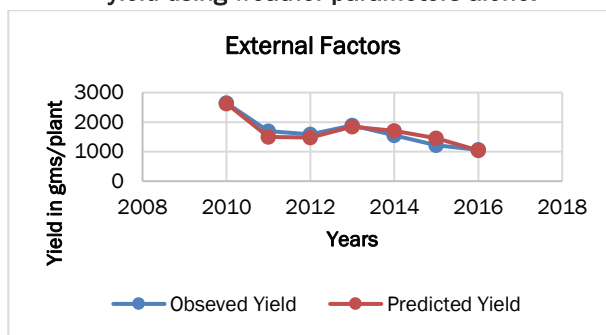


Fig 3. Line chart showing observed and predicted yield using external factors alone.

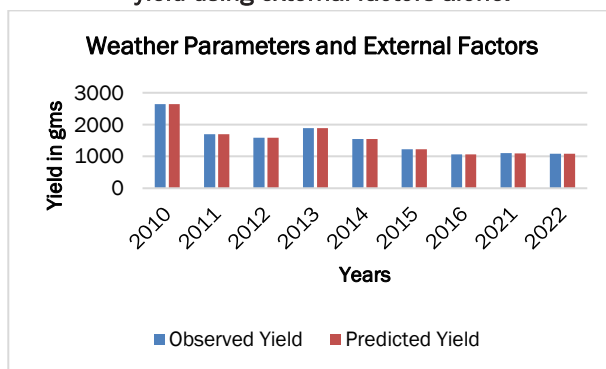


Fig 4. Bar chart showing observed and predicted yield using weather parameters external factors

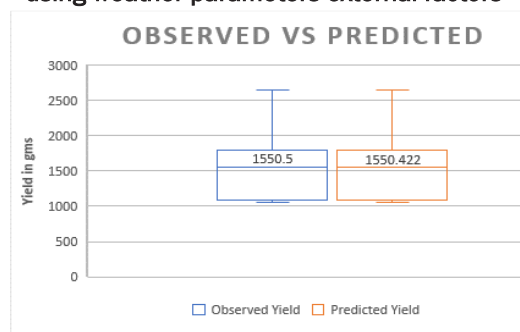


Fig 5. Comparison between observed yield and predicted yield (Eq. 3) using Boxplot.

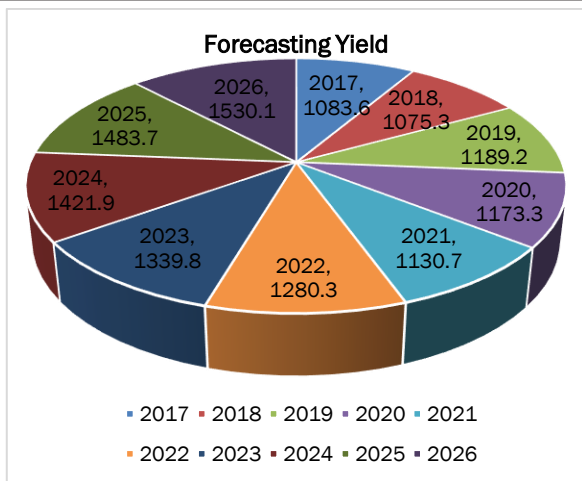


Fig 6. showing the forecasted yield for the years 2017 to 2026.

Table.1 showing observed and model predicted yield for weather parameters for the years (2010-2016).

| Year | Observed Yield(gms/plant) | Predicted Yield(gms/plant) |
|------|---------------------------|----------------------------|
| 2010 | 2647.5 | 2519.658 |
| 2011 | 1695 | 1748.705 |
| 2012 | 1587.5 | 1460.287 |
| 2013 | 1892 | 2092.918 |
| 2014 | 1550.5 | 1363.895 |
| 2015 | 1222.3 | 1375.925 |
| 2016 | 1062 | 1095.412 |

Table.2 showing observed and model predicted yield for external factors for the years (2010-2016).

| Year | Observed Yield (gms/plant) | Predicted Yield (gms/plant) |
|------|----------------------------|-----------------------------|
| 2010 | 2647.5 | 2630.196 |
| 2011 | 1695 | 1500.959 |
| 2012 | 1587.5 | 1472.158 |
| 2013 | 1892 | 1843.564 |
| 2014 | 1550.5 | 1709.54 |
| 2015 | 1222.3 | 1459.144 |
| 2016 | 1062 | 1041.239 |

Table 3. indicates observed and predicted yield for weather parameters and external factors for the years (2010-2016) and validated yield (2021-2022).

| Year | Observed Yield (gms/plant) | Predicted Yield (gms/plant) |
|------|----------------------------|-----------------------------|
| 2010 | 2647.5 | 2647.543 |
| 2011 | 1695 | 1694.975 |
| 2012 | 1587.5 | 1587.489 |
| 2013 | 1892 | 1891.979 |
| 2014 | 1550.5 | 1550.422 |
| 2015 | 1222.3 | 1222.211 |
| 2016 | 1062 | 1061.982 |
| 2021 | 1105.50 | 1094.747 |
| 2022 | 1085.00 | 1080.115 |

Table 5. shows the forecasted yield for 2017 to 2026.

| Year | Forecasted Yield (gms/plant) | Year | Forecasted Yield (gms/plant) |
|------|------------------------------|------|------------------------------|
| 2017 | 1083.6 | 2022 | 1280.3 |
| 2018 | 1075.3 | 2023 | 1339.8 |
| 2019 | 1189.2 | 2024 | 1421.9 |
| 2020 | 1173.3 | 2025 | 1483.7 |
| 2021 | 1130.7 | 2026 | 1530.1 |

CONCLUSION

The coconut yield prediction model was developed using weather and external factors. This proposed model is validated using field level data. Further, we developed forecasting model for coconut yield and forecasted for ten years from 2017 to 2026. The proposed model will be helpful for farmers by predicting the yield before the harvest period so that they can plan the post-harvest process earlier. The model can also be used for policy making for coconut growers.



Table 4. showing R square, RMSE and MAPE for Equation (1) and (2):

| | Model Equation | R ² | RMSE | MAPE |
|---|--|----------------|--------|-------|
| Using Weather parameters | $Y = 43764.71 - 4899.5 * MaT + 2016.9 * MiT + 732.94 * RH + 1528.01 * WS - 56.02 * RF$ | 0.91 | 139.03 | 7.76% |
| Using external factors | $Y = 1430.793 - 38.91 * X_1 + 2.67 * X_2 + 14.15 * X_3 + 414.81 * X_4 + 7.44 * X_5$ | 0.91 | 118.29 | 6.12% |
| Using weather parameters and external factors | $Y=0+0*X_1+19.606*X_2+1.4537*X_3-1.1593*X_4-26.072*X_5+4.8373*X_6-10.928*X_7+7.3649*X_8$ | 0.99 | 3.93 | 0.16% |

Acknowledgement:

The authors are very grateful to Department of Physical Sciences & Information Technology for providing facilities required for research.

Ethics statement:

No specific permits were required for the described field studies because no human or animal subjects were involved in this research.

Consent for publication:

All the authors agreed to publish the content.

Competing interests:

There was no conflict of interest in the publication of this content.

Data availability:

All the data of this manuscript are included in the MS. No separate external data source is required.

REFERENCES

V. Abeywardena, 1968. Forecasting coconut crops using rainfall data – A preliminary study, Ceylon Coconut Quarterly Review **19**, 161-176. LINK: <https://dl.nsf.gov.lk/handle/1/19883>

T.S.G Peiris, R.O. Thattil, R. Mahindapala. 1995. An analysis of the effect of climate and weather on coconut (*Cocosnucifera*). Journal of Experimental Agriculture **31(4)**: 451-460. DOI: <https://doi.org/10.1017/S0014479700026430>

T. S. G. Peiris and R. O. Thattil., 1998. The Study of Climate Effects on the Nut Yield of Coconut using Parsimonious Models, Journal of Experimental Agriculture, Volume **34**: 189-206. DOI: <https://doi.org/10.1017/S0014479798002051>

C. Palaniswami, R. Dhanapal, A.K. Upadhyay, C. Manojkumar and K. Samsudeen 2008. A fuzzy neural network for coconut yield prediction, Journal of Plantation Crops, **36 (1)**: 24-29. LINK: <https://www.researchgate.net/publication/257465383>

T. S. G. Peiris, J. W. Hansen and Lareef Zubai, 2008. Use of seasonal climate information to predict coconut production in Sri Lanka, International Journal of Climatology, *Int. J. Climatol.***28**: 103-110. DOI:[10.1002/joc.1517](https://doi.org/10.1002/joc.1517)

Ranasinghe, C.S., Silva, L.R.S., Premasiri, R.D.N., 2015. Major determinants of fruit set and yield fluctuation in coconut (*Cocos nucifera* L.). *J. Natl. Sci. Found. Sri Lanka* **43(3)**:253-264. DOI:[10.4038/jnsfr.v43i3.7955](https://doi.org/10.4038/jnsfr.v43i3.7955)

C. Pathmeswaran, E. Lokupitiya , K.P. Waidyarathne , R.S. Lokupitiya, 2018.Impact of extreme weather events on coconut productivity in three climatic zones of Sri Lanka, *European Journal of Agronomy*,**96**:47-53. DOI : <https://doi.org/10.1016/j.eja.2018.03.001>

Shripad Vishweshwar, Shankar Meti, B.V. Champa and M. Nagaraja, 2019. Climate Based Coconut Yield Analysis in Chanrayapatna Taluk of Hassan District of Karnataka, India, *International Journal of Current Microbiology and Applied Sciences* ISSN: 2319-7706 Volume **8(7)**. DOI:[10.20546/ijcmas.2019.807.357](https://doi.org/10.20546/ijcmas.2019.807.357)

Bappa Das, Bhakti Nair, Vadivel Arunachalam, K. Viswanatha Reddy, 2020. Comparative evaluation of linear and nonlinear weather-based models for coconut yield prediction in the west coast of India, International Journal of Biometeorology ISSN 0020-7128 *Int J Biometeorol.* DOI: [10.1007/s00484-020-01884-2](https://doi.org/10.1007/s00484-020-01884-2)

Benny Antony, 2021. Prediction of production of crops with respect to rainfall, Environmental Research, ISSN 0013-9521, Vol-202. DOI : [10.1016/j.envres.2021.111624](https://doi.org/10.1016/j.envres.2021.111624)

S. Rohith1 , S. Kavibalan1 , K. Thangaraj2 , J. Suresh4 , M. Ananthan3 and R. Renuka1*, Characterization of mother palms and novel techniques to produce elite seedlings of coconut var. Chowghat Orange Dwarf, Electronic Journal of Plant Breeding, Vol **14(3)** : 867 - 875, doi: 10.37992/2023.1403.112.



Appendix-A (MATLAB code for yield prediction model):

```
>> y = [18475 16950 15875 1892015505 12223
10620];
>> x1 = [22.32 21.48 22.04 22.25 22.46 22.54
22.77];
>> x2 = [87.33 88.25 84.5 84.25 83.42 84.67 84];
>> x3 = [63.52 91.96 31.14 44.63 61.10 57.43
22.65];
>> x4 = [693 749 840 920 1040 1080 1103];
>> x5 = [100 100.4 96 96 100 98 105];
>> x6 = [620 508 520 520 535 506 504];
>> x7 = [20 20 11.3 10.7 13 19 25];
>> x8 = [200 146.4 131.75 182.8 162 142
167.5];
>> X = [ones(size(x1)) x1 x2 x3 x4 x5 x6 x7 x8];
>> a = X\y
```

A

```
0
0
19.606
1.4537
-1.1593
-26.072
4.8373
-10.928
7.3649
```

Appendix – B Code for ARIMA in Python software:

```
1. pip install pmdarima
2. import pandas as pd
import numpy as np
3. df=pd.read_csv(r'C:\Users\mades\Desktop\d.csv', index_col='Year')
4. df
5. from statsmodels.tsa. stattools import adfuller
def ad_test(dataset):
dftest=adfuller(dataset, autolag='AIC')
print("1.ADF:",dftest[0])
print("2.P-value:",dftest[1])
print("3.Num of lags:",dftest[2])
```

```
print("4.Num of observations used for ADF Regression
and Critical Values Calculation:",dftest[3])
```

```
print("5.Critical Values:")
for key,val in dftest[4].items():
print("\t",key,".",val)
```

6. ad_test(df['Yield'])

7. from pmdarima import auto_arima

```
import warnings
```

```
warnings.filterwarnings("ignore")
```

8. stepwise_fit=auto_arima(df['Yield'],trace=True,

```
supress_warnings=True)
```

```
stepwise_fit.summary()
```

9. from statsmodels.tsa.arima_model import ARIMA

```
print(df.shape)
```

```
10. train=df.iloc[:-3]
```

```
test=df.iloc[-3:]
```

```
print(train.shape,test.shape)
```

11. model=ARIMA(train['Yield'],order=(1,0,0))

```
model=model.fit()
```

```
model.summary()
```

12. start=len(train)

```
end=len(train)+len(test)-1
```

13.

```
pred=model.predict(start=start,end=end,typ="levels")
```

```
pred.index=df.index[start:end+1]
```

```
print(pred)
```

14. model2=ARIMA(df['Yield'],order=(1,0,0))

```
model2=model2.fit()
```

```
model2.summary()
```

15. start=len(df)

```
end=len(df)+5
```

```
pred2=model2.predict(start=start,end=end,typ="level
s").rename('ARIMA Predictions')
```

```
index=[2022,2023,2024,2025,2026,2027]
```

```
pred2.index=index
```

```
print(pred2)
```