IEEE-SEM, Volume 9, Issue 2, February-2021
ISSN 2320-9151

Predicting Maximum Daily Temeperatures to Aid Farming for the State of Maharashtra, India

1st Ajinkya Datalkar
Software Engineer/ Data Analyst
Agrolly LLC
Jersey City, NJ, USA
ajinkyadatalkar1@gmail.com

2nd Manoela Morais
Data Analyst
Agrolly LLC
Jersey City, NJ, USA
manoengquimica@gmail.com

3rd Te-Yi Tsai
Software Developer
Agrolly LLC
New York City, NY, USA
helentsaidev@gmail.com

***Abstract*—Agriculture industry is one of the industries that is highly affected by the changes in weather. Annual weather predictions can help farmers determine which crops to grow in advance. Farmers will be aware of disasters like draught, well in advance. Annual weather predictions can not only help in planning crops to be grown but can also help in managing the crop requirements once the seeds have been sown. Farmers can be prepared for any adversaries they may face in advance, thus helping in increasing the crop yield. Determining maximum daily temperatures will help farmers in understanding crop water requirements. These values are essential to determine the amount of water transpiration from the plant tissues and leaves. Maximum daily temperatures will also help in determining the water irrigation needs. Since crops are sensitive to the rising temperatures, here a model for predicting maximum daily temperatures for the state of Maharashtra, India is discussed. The methodologies adopted to forecast weather parameters were the simple and seasonal times series. Comparative study has been carried out by using error percentage, mean square error and root mean square error. We found that time-series models work well for predicting averages monthly temperatures, however for daily values adjusting for seasonality presented a better result.**

Keywords*— Weather, Forecast, Time-series, Ordinal, Recursion, Regression, Temperature, Rainfall, Precipitation, Machine learning, Artificial Intelligence, Predictions, Crops, Farming, Agriculture, Irrigation, Water Requirements, Livestock, R, R-studio

I. INTRODUCTION

Global climate change has had adverse effects on all walks of life, in the past decade these effects have become much more observable as we have started to notice the rise in temperatures, melting of ice and glaciers, extinction of plants and animal species, etc. As the world population rises, the need for developing methods that will help in effective farming also rises. *In a study done by Kibrom A. Abay and Nathaniel D. Jensen, it has been seen households exposed to more unpredictable weather are less likely to engage in livestock production for markets, rather they are more likely to engage in livestock production for precautionary savings and insurance. Furthermore, greater rainfall uncertainty influences livestock portfolio allocation toward those which can be easily liquidated while also discouraging investment in modern livestock inputs.*[1]

In several cases it has also been seen that the amount of fertilizers to be used while sowing the seeds also depends upon the weather. *For example in a study by Terrance Hurley, Jawoo Koo and Kindie Tesfaye it was seen that the amount of fertilizer used affected the crop yield and the amount of fertilizer to used depended upon the weather conditions. In their study they found that unpredictable weather makes maize farming inherently risky—how much will be produced is not known when seed is planted. This and other types of risk have a significant impact on farmer decisions such as the decision to adopt improved seed varieties or use fertilizer (for a review, see Hurley, 2010).*[2] In this study the authors Terrance Hurley, Jawoo Koo and Kindie Tesfaye have clearly shown how weather risks and some other kinds of risks affect the decisions of the farmers. Therefore, the success of agricultural yield productivity relies on good decisions, and a successful cropping season involves the ability to plan months ahead of starting planting. Crop productivity could be increased by avoiding losses due to the effect of low or high weather damage.

Temperatures are heavily affected by seasons, for example summer season will experience high temperatures, whereas the rainy season may experience moderate to low temperatures and winters will experience low temperatures. India in general has 6 seasons namely the Spring, Summer, Monsoon, Autumn, Pre-Winter and Winter. The State of Maharashtra however experiences only three seasons namely the Summer, Winter and the Monsoon. This shows that the state will experience high, medium and low temperatures depending upon the season. Therefore, states such as Maharashtra demands for a regional model to predict temperature forecast.

Time series (ts) forecasting can be used to predict future values based on past data. This modeling has attracted attention from researchers and communities over the few decades due to the usage in numerous practical fields such as finance, economy, engineering, weather, etc. The main goal of this research is to verify which times series model (non-seasonal and seasonal) present a lower RSME and MPE. This paper is organized as follows: description of the data, the methodology used. Analysis of using times series models to predict maximum daily temperature, Experimental results obtained, and finally we conclude this paper in last section.

II. DATA

The weather variable used in this study was the maximum temperature. The dataset consists in the extraction of the last 5 years of meteorological data between January 2015 and May 2020 from the Nasa website for the state of Maharashtra.

III. METHODOLOGY

For modeling we used the "ts" method and "forecast" method to generate forecasts using time series

IEEE-SEM, Volume 9, Issue 2, February-2021
ISSN 2320-9151

analysis. We used the “ts” method in R (programming language) to perform time series analysis on the data provided. Before generating weather predictions, the “forecast” library has to be imported, this library contains the “forecast” method. The “forecast” method is used to predict the future values based on the time series analysis done on past values. While performing time series analysis, it is always a better idea to provide data that covers wide range of events. This enables us to take all possibilities into consideration. Take an example where we want to predict annual weather forecast for a region that is hit by drought every three years, it is safe to assume that for more accurate weather predictions data covering at least two drought events is provided. Minimum temperature, Maximum Temperature and Precipitation are the three basic parameters that affect crop growth. *Two scientists in India, K. S. Kavi Kumar and Jyoti Parikh, assessed the effect of higher temperatures on wheat and rice yields. Basing their model on data from 10 sites, they concluded that in north India a 1-degree Celsius rise in mean temperature did not meaningfully reduce wheat yields, but a 2-degree rise lowered yields at almost all the sites. When they looked at temperature change alone, a 2-degree Celsius rise led to a decline in irrigated wheat yields ranging from 37 percent to 58 percent. When they combined the negative effects of higher temperature with the positive effects of CO2 fertilization, the decline in yields among the various sites ranged from 8 percent to 38 percent* [3].

Since these factors fluctuate a lot over a period of time, annual weather forecast predicted using past values is bound to be erroneous. Consider an example of city of Achalpur located in the state of Maharashtra. This city experiences very high temperatures in Summers and moderate temperatures in Winters and Monsoon Season.

IV. MODELING

A. Time Series Analysis To Predict Maximum Daily Temperatures

In order to perform time series analysis on the data, import weather data for the past few years. This data should include multiple instances of events draughts, floods and other natural disasters. The past data can be easily acquired from the internet or by using libraries in R like the nasapower library provided by NASA. Once the data has be imported, it is required to be cleaned, the purpose of this step is to interpolate null values and replace any outliers present in the data. Once the data is cleaned, a time series analysis can be performed on the data. The “ts” function can be used to create a time-series object used for forecasting. The syntax for using the “ts” function is as given below:

ts (data, start = c (start time), end =c (end time),

frequency=365)

Here data is the past weather data, start is the date or time from which time series analysis is to be performed and end is the date or time until which the analysis is to be performed. Frequency is set to 365 so that 365 observations are made while analyzing the data. Once the time-series object has been created, forecasting of the data can be done, with the help of the forecast function. The syntax for the forecast function is as given below:

forecast (time-series object, h=365)

Here the h is the number of periods for forecasting. Consider the example given below, in the graph in Fig (1) forecasting is done for the month of January 2021. Data has been forecasted for 31 days.

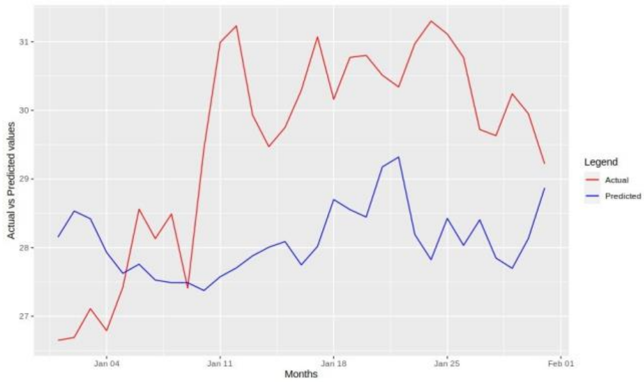

Fig 1. Maximum Temperature Prediction for January 2021 using Time-Series

B. Analyzing Seasonality And Trends of the Data

The basic ts function does not take into consideration the seasonality of the data. As a result, the deviation of the predicated value from the actual value is very high and this can be seen in the graph above. This deviation of the predicted values from the actual values is due to the fact that the seasonality of the data isn’t taken into consideration. Consider the graph given in Fig (2) for the same period with seasonality taken into consideration.

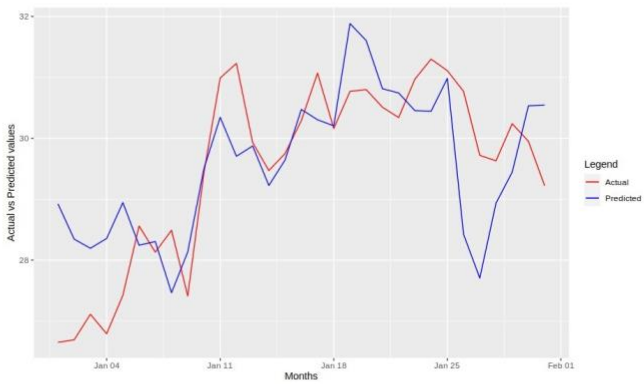

Fig 2. Maximum Temperature Prediction for January 2021 using Seasonal Time-Series

It can be seen in the above graph Fig (2) that the deviation between the predicted and the actual values is very negligible in comparison to the deviation between predicted and actual values in the graph in Fig (1). Since seasonality is taken into consideration in case of the second graph, the RMSE (Root Mean Square Error) of the prediction is low, resulting in better prediction. The closer the RMSE value to 0, the better the prediction.

A time-series data can exhibit a variety of patterns, the data

**For the seasonal time series graphs presented in this paper, the predicted values may be offset by a factor ranging between 0.9 to 1.6*

IEEE-SEM, Volume 9, Issue 2, February-2021
ISSN 2320-9151

can be seasonal, trend following or cyclic. We decompose the time-series data into these components. STL (Seasonal and Trend decomposition using Loess) decomposition is one such method that allows us to take seasonal windows into consideration. STLM is one such model under STL that allows to take into consideration the seasonal window. For the graph in Fig (2), STLM model was used.

The syntax for using STLM method and forecast method for STLM is shown in the images Fig (3) and Fig (4).

```
stlm(
  y,
  s.window = 13,
  robust = FALSE,
  method = c("ets", "arima"),
  modelfunction = NULL,
  model = NULL,
  etsmodel = "ZZN",
  lambda = NULL,
  biasadj = FALSE,
  xreg = NULL,
  allow.multiplicative.trend = FALSE,
  x = y,
  ...
)
```

Fig 3: STLM method syntax [5]

```
# S3 method for stlm
forecast(
  object,
  h = 2 * object$m,
  level = c(80, 95),
  fan = FALSE,
  lambda = object$lambda,
  biasadj = NULL,
  newxreg = NULL,
  allow.multiplicative.trend = FALSE,
  ...
)
```


Fig 4: Syntax for forecast method for STLM [5]

There are several other methods like SEATS and X11, that allow seasonal decomposition, however the advantage of using STL/ STLM over other methods is that it can handle any type of seasonality, not only monthly or quarterly data. The rate of change over time and the smoothness of the trend cycle can be controlled by the user.

The first attribute "y" of the STLM method has to be replaced with a ts object.

C. Seasonal window

The state of Maharashtra has a tropical climate, and the three distinct seasons are Summer (March - May), Monsoon (June-September) and Winter (October-February). Summers are short and extremely hot. Thus we see that the "s.window" attribute of the stlm method needs to be set 3, indicating 3 seasonal windows. The graph in Fig (2) was generated by setting in STLM method:

Attributes	***Values***
y	Ts object
s.window	3
robust	FALSE
method	c("ets")
etsmodel	"MAN"
biasadj	TRUE
xreg	NULL
allow.multiplicative.trend	FALSE
x	y

1. Attributes used in STLM analysis in Fig. 2

STLM method takes the time series object and applies STL decomposition to it. It seasonally models the and the data and passes it to the model function. STLM method returns the STL decomposition and the seasonally adjusted time series object.

STLM method take the ts object and passes it to the modelfunction. In the example in Fig (2), ETS (Error, Trend, Seasonal) model is used. ETS method has a seasonality limit of 24 and as a result in the given example, the seasonality is ignored. In the above example Fig (2), ETS method is used to smooth out any errors that may have occurred. ETS model takes the parameters A, M, N and Z. A = additive, M = multiplicative, N = none and Z = automatic. In the above case we have made use of MAN model in the ETS method.

The s.window value will usually change with the state as the number of seasons will change with each state. For example, most of the North Indian states have 6 seasons whereas, the South Indian state of Kerala has Monsoon twice a year and a total of 4 seasons including the 2 Monsoon seasons.

An improperly supplied s.window value will give a higher RMSE (Root Mean Square Error) and MSE (Mean Square Error). The value has to be set depending upon how the seasonality evolves. While predicting annual weather forecast for the state of Maharashtra, setting s.window value anything other than 3 results in a higher RMSE and MSE. Providing a value lesser than 3 will cause the data to be overfitted, whereas providing a value more than three will cause the data to be underfitted.

D. Forecasting STLM

In order to forecast the time-series data adjusted for seasonality, the STLM object is passed to the forecast method and the value of h is set to 31 (h = 31), the other parameters are passed as they appear in the syntax of the forecast method for STLM.

V. Results

A. Annual Forecasting: Time Series VS Seasonal Time Series

The main purpose of providing this forecast is to help farmers plan their crops in advance and be prepared for any calamity that may occur. Providing daily values in the annual forecast can help farmers in determining the sowing month, week and days. Therefore, it is also important to predict near accurate values.

For the purpose of predicting data for the month of January 2021, a time series analysis on the data from January-01-1981 to December-31-2020 was done. After forecasting the result is presented in the graph in shown in Fig (1). In order to determine the accuracy of both the models, namely Time Series as well as Seasonal Time Series, data for the past year

(2020), can be predicted and compared to the actual values. To perform time-series analysis and predict data for 2020, weather data until December-31-2019 is imported.

After the time-series analysis is complete, for the seasonal prediction, pass the ts object to the stlm function. Before forecasting make sure the frequency is set to 365 or 366 in case of the year 2020.

Given in the figures Fig (5) and Fig (6) the forecasts for the basic time series and seasonal time series:

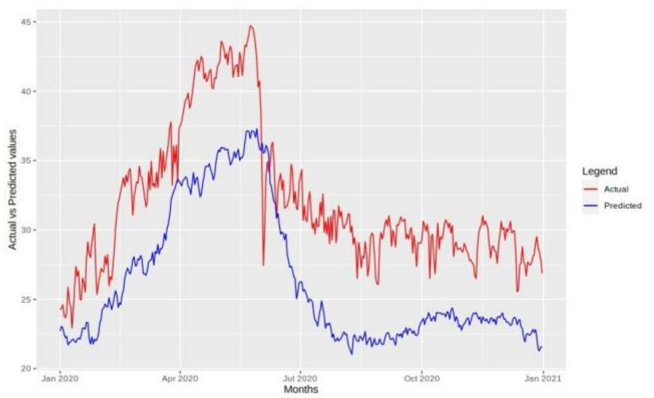

Fig 5. Predictions using Time-Series Analysis

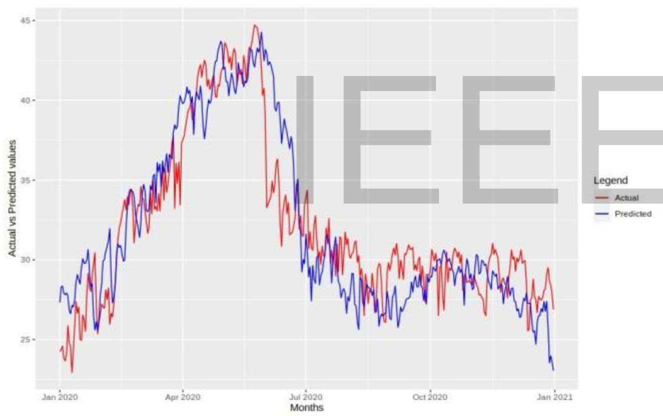

Fig 6. Predictions using Seasonal Time-Series Analysis

The accuracy of time series model vs seasonal time series model can be judged by looking at the two graphs. The predicted values in the time-series graph rarely intersect with the actual values, whereas it can be seen that in case of seasonal time-series graph, the predicted values and actual values merge and intersect on multiple occasions.

B. Error Margins For Seasonal Time Series

Here the errors only for seasonal time series model are taken into consideration as the errors are relatively small and the model is usable. For the purpose of predicting annual forecast using seasonal time series shown in the graph in Fig (6), we used the ANN model of the ETS method:

```
Model Information:
ETS(A,N,N)

Call:
 ets(y = x, model = etsmodel, allow.multiplicative.trend = allow.multiplicative.trend)

  Smoothing parameters:
    alpha = 0.8125

  Initial states:
    l = 32.107

  sigma:  0.1679

     AIC     AICc      BIC
85404.33 85404.33 85427.02

Error measures:
                        ME      RMSE        MAE          MPE       MAPE       MASE       ACF1
Training set -0.0001294536 0.1679061 0.03071545 -0.002858901 0.09689655 0.01258983 0.02134145
```

Fig 7. Model Information and Error Measurement for seasonal time series model, for 2020 maximum daily temperature predictions in Achalpur

Low value of RSME and MAE (0.1678061 and 0.03071545) were achieved by changing the ETS model from MAN to AAN and taking into consideration the data for the past 38 years. It was seen that the more data was crunched into the seasonal time series model, the more RMSE value dropped. The seasonal time series model predicted extremely accurate data for 49 days out of 366 days. The table given below shows error percentages and days:

Error %	*Days*
Less than 1% (difference < 0.32°C)	49 days
1% to 5% (0.32°C < diff < 1.6°C)	132 days
5% to 10% (1.6°C < diff < 3.2°C)	107 days
10% to 15% (3.2°C < diff <4.8°C)	56 days
15% to 20% (4.8°C < diff <6.4°C)	13 days
20% to 25% (6.4°C < diff < 8.00°C)	5 days
25% to 30.20% (6.40°C < diff < 10.0°C)	4 days

2. Days with errors for predicted maximum temperatures in Achalpur

From the above table it can be seen that the model produced very close to accurate date for 288 days i.e., the margin for error was less than 10%, for the remaining 78 days, it can be seen in the graph that city of Achalpur experienced a sudden drop in temperature on the 154th day of the year 2020. The temperature difference between actual and predicted values for that day was 9.76°C. There is an error in the values generated for the next 20 days after day 154. These types of errors are difficult to avoid. Thus, we can see that the model predicted acceptable results for 79% of the days and erroneous results for 21% of the days. If we do not take into consideration the days from 154 to 174, the model predicted acceptable results for 82% of the days.

**For the seasonal time series graphs presented in this paper, the predicted values may be offset by a factor ranging between 0.9 to 1.6*

C. *Testing Model on Other Cities in Maharashtra*

1) *For Mumbai:*

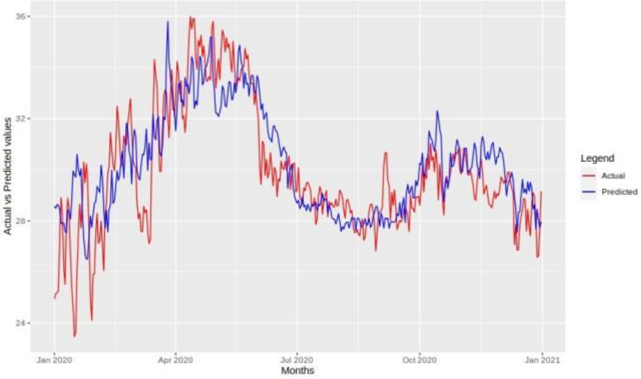

Fig 8. Actual vs Predicted Temperatures for 2020 in Mumbai

Error %	***Days***
Less than 1% (difference < 0.302°C)	63 days
1% to 5% (0.302°C < diff < 1.51°C)	202 days
5% to 10% (1.51°C < diff < 3.02°C)	82 days
10% to 15% (3.02°C < diff <4.51°C)	16 days
15% to 20% (4.51°C < diff <6.04°C)	1 day
20% to 25% (6.04°C < diff < 7.55°C)	2 days
25% to 30% (7.55°C < diff < 9.06°C)	0 days

[3.]Days with errors for predicted maximum temperatures in Mumbai

```
> summary(temperature_fcast_all_Max)

Forecast method: STL +  ETS(A,N,N)

Model Information:
ETS(A,N,N)

Call:
 ets(y = x, model = etsmodel, allow.multiplicative.trend = allow.multiplicative.trend)

  Smoothing parameters:
    alpha = 0.9939

  Initial states:
    l = 30.1531

  sigma:  0.1009

     AIC     AICc      BIC
70901.25 70901.25 70923.94

Error measures:
                        ME      RMSE        MAE         MPE       MAPE       MASE        ACF1
Training set -7.272426e-05 0.1009182 0.01751242 -0.00103074 0.05840943 0.01310367 0.001136407
```

Fig 9. Model Information and Error Measurement for seasonal time series model, for 2020 maximum daily temperature predictions in Mumbai

2) *For Nashik:*

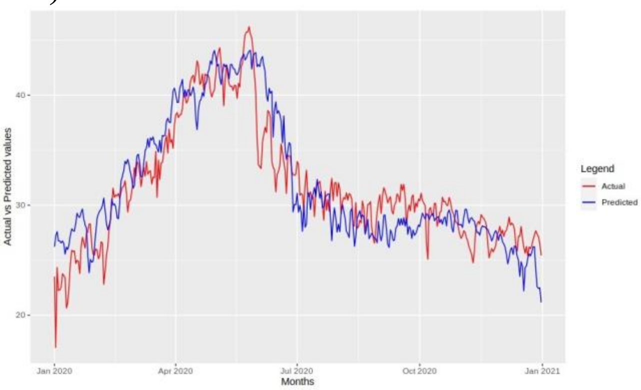

Fig 10. Actual vs Predicted Temperatures for 2020 in Nashik

Error %	***Days***
Less than 1% (difference < 0.3181°C)	27 days
1% to 5% (0.3181°C < diff < 1.59°C)	140 days
5% to 10% (1.59°C < diff < 3.18°C)	118 days
10% to 15% (3.18°C < diff <4.77°C)	57 days
15% to 20% (4.77°C < diff <6.362°C)	16 day
20% to 25% (6.362°C < diff < 7.95°C)	3 days
25% to 30% (7.95°C < diff < 9.543°C)	3 days

[4.]Days with errors for predicted maximum temperatures in Nashik

```
Forecast method: STL +  ETS(A,N,N)

Model Information:
ETS(A,N,N)

Call:
 ets(y = x, model = etsmodel, allow.multiplicative.trend = allow.multiplicative.trend)

  Smoothing parameters:
    alpha = 0.8363

  Initial states:
    l = 32.3477

  sigma:  0.1738

     AIC     AICc      BIC
86392.64 86392.64 86415.33

Error measures:
                        ME      RMSE        MAE          MPE       MAPE       MASE       ACF1
Training set -0.0001600321 0.1738333 0.03093455 -0.003155738 0.09733162 0.01260024 0.01808735
```

Fig 10. Model Information and Error Measurement for seasonal time series model, for 2020 maximum daily temperature predictions in Nashik

VI. Observations

In our study we also found that Time-Series model works well for predicting averages monthly temperatures, however if daily temperatures are to be taken into consideration, the seasonal factor has to be taken into consideration. If we take a look at the graphs in figures 5 and 6, we will realize that the even if we step up the predicted values in graph in figure 5 by adding 2 or 3 to all predicted values to reduce or adjust for error, we will still notice that the toughs and the crests are not fluctuating as much as they are fluctuating in the graph for predicted values in the figure 6. This shows us the importance of adjusting for seasonality.

VII. Distributing and Making Information Available to the Farmers

While all the research and development is being done, it is important that these findings reach the farmers in a way that it is easily understood by the farmers and put to use effectively.In a study by Heidi Kaila and Finn Tarp, they found that with the introduction of internet to agriculture, the agricultural output increased by 6.8%. It was also found in the study that households with younger household heads benefited the most from access to the internet [6].

In a country like India that has more than 50% of its population with ages less than 25 years and more than 65% of the population with ages less than 35 years [7]. Making information available through mobile devices with the help of applications can be beneficial.With the help of data available through weather predictions, alerts and notifications can be sent to the farmers informing them about any calamity and any precautions that are needed to be taken.

With the help of predicted weather data, predictions for crop requirements can be made. Given below is an example of an app that shows information to the farmers regarding crop requirements with the help of predicted data.

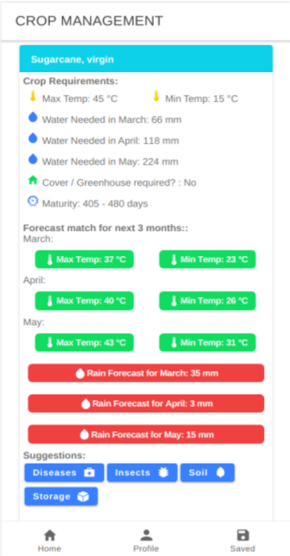

Fig 11. Agrolly app showing crop data in comparison with weather prediction.

The app shown in the figure above is developed by Agrolly LLC and is currently under testing in India, it calculates crop water requirements based on the methods provided by Food and Agriculture Organization and based on the weather predictions generated using time series model, gives farmers insights about various problems they might face in a way that is easily understood by farmers.

VIII. Conclusion

We found that the ETS method with model ANN presented a better result with the RSME and MAE lower than 0.2847 and 0.09337713. Therefore, we used the same models for predicting values for all the other cities. However, the ETS models are needed to be tested for regions to see if we could get results better than this.

In this study we only took into consideration, the maximum daily temperature values for the past years. These predictions can be improved by taking into consideration the various other factors like rainfall, humidity, minimum temperatures, etc. Furthermore, other models for predictions using time-series analysis needs to be studied before determining the best model for predicting maximum temperatures in the state of Maharashtra, India.

While development of these models is an ongoing and never-ending process, innovative ways that help spread of this information are also needed to be developed.

IX. References

[1] Abay KA, Jensen ND.Access to markets, weather risk, and livestock produc-tion decisions: Evidence from Ethiopia. AgriculturalEconomics. 2020;51:577–593.https://doi.org/10.1111/agec.12573

[2] Terrance Hurley, Jawoo Koo and Kindie Tesfaye. Weather risk: how does it change the yield benefits of nitrogen fertilizerand improved maize varieties in sub-Saharan Africa? AgriculturalEconomics. 49 (2018) 711–723. https://doi.org/10.1111/agec.12454

[3] Lester R. Brown, *Plan B 3.0: Mobilizing to Save Civilization*, Chapter 3: Rising temperatures and rising seas, The crop yield effect, Earth Policy Institute, Rutgers University, [E-book] Available: http://www.earth-policy.org/books/pb3/

[4] R-core R-core@R-project.org, *R Documentation*, Time-Series Objects, Available: https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/ts

[5] Rob Handyman, *R Documentation*, Forecasting Using Stl Objects, Available: https://www.rdocumentation.org/packages/forecast/versions/8.12/topics/forecast.stl

[6] Kaila H, Tarp F. Canthe Internet improve agricultural production? Evidencefrom Viet Nam. AgriculturaEconomics. 2019;50:675–691.https://doi.org/10.1111/agec.12517

[7] Demographics of India. https://en.wikipedia.org/wiki/Demographics_of_India

**For the seasonal time series graphs presented in this paper, the predicted values may be offset by a factor ranging between 0.9 to 1.6*