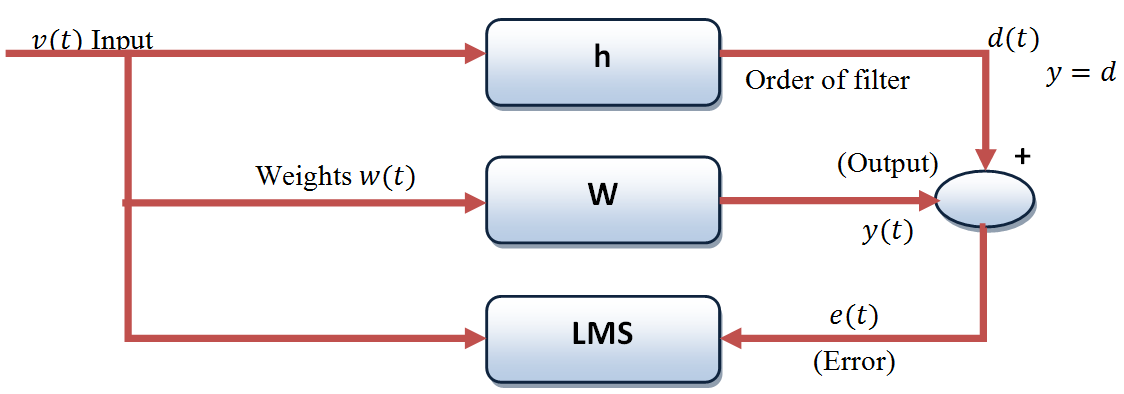
**Chapter 2**

**Short Term Forecasting of Wind Speed based on Statistical and Time Series Models**

*This chapter presents statistical and Time Series (TS) models for short-term forecasting of wind speed. Two types of statistical and four TS models are investigated with actual measured wind speed data of three sites. Least Mean Square (LMS), Holt Winters model ARMA, ARIMA, Transfer Function, GARCH and Wavelet-ARIMA models are developed. Analytical and iterative algorithms for optimally selecting ARIMA models are investigated. A case study is carried out to know effect of volume of data and horizon on wind speed forecasting accuracy.*

**2.1 Least Mean Square (LMS) Algorithm**

LMS method is introduced by Hoff and Widrow in 1959 [42]. Earlier to this a nonlinear adaptive filter is first introduced by Gabor with a Volterra series [43]. LMS includes a procedure that makes unchanged modifications to minimize mean square error. Association function and matrix inversions are not required in LMS algorithm.

****

**Fig. 2.1: Block diagram of an adaptive filter LMS**

Fig. 2.1 presents adaptive filter LMS model. For a data signal at t, denoted as and y(t) as output of a is a linear combination of prior N trials of data weighed by analogous weight vector . M is an integer parameter used by filters. Output is compared to input data stream. Filter attempts to adapt to and fed into adaptation algorithm, so filter weights are restructured. Vector i.e., weights are updated at every time step *t* in order to reduce mean square error. LMS algorithm is defined through three expressions given by:

|  |  |  |
| --- | --- | --- |
|  | Filter output:  Estimation error:  Weight adaptation: | (2.1a) |

Where,

is reference wind speed in m/s,

and denote M × 1 column vectors are given by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.1b) |
|  |  | (2.1c) |

Where,

is recent wind speed value at time t,

is forecasted wind speed at t+k time period,

m is adaptive rate of filter ,

N(k) is norm of input wind speed at time step k

**2.1.1 Wind Speed Forecast with LMS**

Fundamental forecast is delaying recent event value by one time occurrence and use it as reference signal. Filter then calculates an estimation of input signal at time occurrence n, as a linear combination of N prior readings. Subtracting prediction signal from desired signal gives value of error which is fed back to adapt filter weights.

Elements of LMS algorithm framed out by Hayes, Haykin (1996) [44] are given by:

|  |  |
| --- | --- |
| Weight vector: | (2.2a) |
| Signal input: | (2.2b) |
| Filter output: | (2.2c) |
| Cost function: | (2.2d) |

LMS elements in (2.1a-d) are utilized to minimize cost function. A simple algorithm presented as:

Where are defined as earlier and *h* is an adaptation rate. Training rule is given by:

|  |  |  |
| --- | --- | --- |
|  |  |  |

Where, denotes a learning rate.

Additionally, LMS algorithm is extended and Ordinary Least Square (OLS) regression is used based on weight vector learning process. A simple regression is given by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.3) |

Where,

is a dependent variable

is a independent variable

are estimated polynomials

is tap-weight vector after learning process.

Relation (2.3) is extended by including additional variables. Final forecast is expressed by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.4) |

Where,

is previous event of dependent variable.

LMS steps for a pth order is summarized by:

Computation: For n = 0, 1, 2...

|  |  |  |
| --- | --- | --- |
|  |  | (2.4a)  (2.4b)  (2.4c) |

Where,

p is filter order

is step size and

As LMS algorithm use approximate values of anticipations, weights may never reach an optimal in unconditional situation, but there is a possible mean convergence. Weights can alter by small amount, and changes possible for optimal weights. Though, if variance is large convergence in mean may become uncertain. Hence, a lower bound on µ is required. This is given by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.5) |

Where,

is an auto-correlation matrix, its Eigen values are non-negative.

If condition is not fulfilled, then algorithm becomes unbalanced. Convergence of algorithm is inversely proportionate to Eigen value and correlation matrix R. When Eigen values of R are extensive, convergence may be slow. If µ value is selected very small then algorithm converges very slowly. A large value of µ may lead to a quicker convergence but it may be less constant around minimum value. Maximum convergence speed is achieved by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.6) |

Where, is smallest Eigen value.

Given µ is less than optimum, speed of convergence is determined by **λ**min. Faster convergence is achieved when **λ**max is near to **λ**min andhighest attainable convergence depends on Eigen value of R.

**2.1.2 Results and Discussion**

Wind speed forecasting results for 24hr wind speed forecast using LMS for Bagalkot, Vijayapura and Bengaluru sites are presented in Fig. 2.2-3.4. Sub-figures (a) presents comparison of actual vs forecasted wind speed, (b) indicates correlation between actual and forecasted wind speed and (c) gives frequency distribution of forecast error. Regression plot shows R2=0.606 which is moderate. Error distribution shows that both negative and positive errors are equally distributed. It is observed that forecasted curves are nearly following pattern, but curves are not smooth enough. Table 2.1 presents comparison of forecasting MAPE and t-test results of LMS method. MAPE values range from 20.67%, 19.06% and 21.43% for three sites. Fig. 2.5-3.7 present results for 168hr prediction of wind speed from LMS method for three sites. In this case MAPE values range from 27.62%, 27.57% and 26.10% for three sites. Value of h=0 indicates that, hypothesis for error from normal distribution cannot be rejected at 95% confidence interval. Hence LMS model’s error failed in t-test. Forecasted wind speed curves present spikes causing increased uncertainty in forecasting.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2.1: Forecasting result and Test result for error and forecasted wind speed for 24hr and 168hr prediction of LMS method** | | | | | | | | | | | | | | |
|  | **24Hour Prediction** | | | | | | | **168 Hour Prediction** | | | | | | |
| **Site** | **MAPE in %** | **t-test** | | | | **Stats** | | **MAPE in %** | **t-test** | | | | **Stats** | |
| **h** | **P** | **ci** | | **T-stat** | **Sd** | **h** | **P** | **ci** | | **t-stat** | **Sd** |
| Bagalkot | 20.67 | 0 | 0.471 | -0.089 | 0.190 | 0.406 | 0.800 | 27.62 | 0 | 0.677 | -0.099 | 0.151 | 0.417 | 0.820 |
| Vijayapura | 19.06 | 0 | 0.492 | -0.107 | 0.201 | 0.413 | 0.901 | 27.57 | 0 | 0.628 | -0.101 | 0.167 | 0.486 | 0.878 |
| Bengaluru | 21.43 | 0 | 0.461 | -0.193 | 0.212 | 0.412 | 0.925 | 26.10 | 0 | 0.887 | -0.208 | 0.180 | 0.442 | 1.275 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.2: (a) Forecasted result for 24hr of LMS Algorithm,(b) regression plot actual/forecasted wind (c)speed distribution error distribution and for Bagalkot Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.3: (a) Forecasted result for 24hr of LMS Algorithm, regression plot actual/forecasted wind speed distribution error distribution and for Vijayapura Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.4: (a) Forecasted result for 24hr of LMS Algorithm,(b) regression plot actual/forecasted (c) wind speed distribution error distribution and for Bengaluru Site** |
|  | | |
| |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.5: (a)Forecasted result for 168hr of LMS Algorithm,(b) regression plot actual/forecasted wind (c)speed distribution error distribution and for Bagalkot Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.6: Forecasted result for 168hr of LMS Algorithm, regression plot actual/forecasted wind speed distribution error distribution and for Vijayapura Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.7: (a) Forecasted result for 168hr of LMS Algorithm,(b) regression plot actual/forecasted (c) wind speed distribution error distribution and for Bengaluru Site** |

**2.2 Exponential Smoothing Method**

LMS method resulted in spikes in forecasted values. To avoid peaks, effect of smoothing is investigated with General Exponential Smoothing and Holt’s Winter Exponential smoothing method.

**2.2.1 General Exponential Smoothing (GES)**

In this method, wind speed at time t, y(t) is modeled using a fitting function expressed as [18]:

|  |  |  |
| --- | --- | --- |
|  |  | (2.7) |

Where,

is fitting function vector for process

is coefficient vector

is a white noise

Estimates of coefficients are found using weighted or discounted least mean square error for recent N sampled intervals to minimize function given as :

|  |  |  |
| --- | --- | --- |
|  |  | (2.8) |

Minimization of (2.8) gives estimated vector of coefficients is written as:

|  |  |  |
| --- | --- | --- |
|  |  | (2.9) |

Where,

|  |  |  |
| --- | --- | --- |
|  |  | (2.10) |
|  |  | (2.11) |

Forecast of series at lead time using (2.9) –(2.11) is written as:

|  |  |  |
| --- | --- | --- |
|  |  | (2.12) |

Coefficient estimates and forecasts are updated respectively using (2.12) as:

|  |  |  |
| --- | --- | --- |
|  |  | (2.13) |
|  |  | (2.14) |

Where,

|  |  |  |
| --- | --- | --- |
|  |  | (2.15) |

L matrix is called transition matrix and is constructed on basis that model have a fitting function satisfying relationship.

**2.2.2 Holt-Winter’s Exponential Smoothing (HWES)**

GES method is extended to deal with both seasonal and trend variations because each season has different mean value. Holt-Winters method introduced by Chatfield et al.(1988), has two versions multiplicative and additive [45].

General forecast function for HWES method is given by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.16) |

Where,

is component of level

is slope component

is a seasonal component

*s* is a seasonal period.

Updating formulae for these components need a constant for smoothing constant. If are used as parameter for level, slope, constant, respectively, then updating expressions are written by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.17) |
|  |  | (2.18) |
|  |  | (2.19) |

It is possible to estimate smoothing constant by minimizing sum of squared errors. Once again, and all lie between zero and one. If additive version of HWES (2.16) is utilized, and then seasonal part is added into one step ahead forecast function, then (2.16) is written as:

|  |  |  |
| --- | --- | --- |
|  |  | (2.20) |

Seasonal and level components entail differences of ratios are given by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.21) |

Slope component remains unchanged. For initial values, it is sensible to set component equalling to average observation, i.e.:

|  |  |  |
| --- | --- | --- |
|  |  | (2.21a) |

Initial value for slope is taken from averaged difference between first and next year averages. That is:

|  |  |  |
| --- | --- | --- |
|  |  | (2.21b) |

Seasonal index initial value is calculated as:

|  |  |  |
| --- | --- | --- |
|  |  | (2.21c) |

To take care of seasonality, introduced a third relationship given by:

|  |  |  |
| --- | --- | --- |
|  | Forecast: | (2.22a) |

Where,

|  |  |  |
| --- | --- | --- |
|  | Trend:  Seasonal: | (2.22b) |

Forecast value is obtained from level, trend and seasonal components as in (2.22a-b). Flowchart for Holt Winter’s smoothing based forecasting of wind speed is presented in Fig. 2.8. Wind speed time series are used to initialize values of trend component and seasonal component. Values of α, β and are initialized to 0.0001.

Parameters are updated after reaching α, β, and = 1. These updating is done for all observations. Forecast MAPE is estimated for initial values. MAPE for all intervals are stored. Updating of parameters is stopped when a minimum MAPE reaches its minimum set value.

Yes

No

No

Yes

No

Yes

Input data

Estimate Initial value of trend

Input values of d and s

Compute initial values of seasonal component

Set values of α, β and ϒ equal to 0.0001

A

A

t=t+1

Update parameters

If t<Total No. of Observations ?

If MAPE minimum

Save MAPE, α, β,

α, β, and = 1

Print Y, F, MAPE, α, β, and

Stop

Increase

α, β, and by 0.0001

**Fig. 2.8: Holt Winters Smoothing Forecasting**

**3.2.3 Results for Wind Speed Prediction with Smoothing Method**

## Wind speed data over 10 years is taken as test data and exponential smoothing algorithm is applied to it. A model is developed using hourly wind speed of first nine years tenth year wind speed data is predicted. Fig. 2.9 to 3.14 depict forecasting results of wind speed for 24hr and 168 hr ahead for three sites suing HWES.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.9: Forecasted result for 24hr of HWES, regression plot actual vs forecasted wind speed error distribution and for Bagalkot Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.10: Forecasted result for 24hr of HWES, regression plot actual vs forecasted wind speed error distribution and for Vijayapura Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.11: Forecasted result for 24hr of HWES, regression plot actual vs forecasted wind speed error distribution for Bengaluru Site** |
|  | | |
| |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.12: Forecasted result for 168hr of HWES, regression plot actual vs forecasted wind speed distribution error distribution and for Bagalkot Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.13: Forecasted result for 168hr of HWES, regression plot actual vs forecasted wind speed distribution error distribution and for Vijayapura Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.14: Forecasted result for 168hr of HWES, regression plot actual vs forecasted wind speed distribution error distribution and for Bengaluru Site** |
|  | | |

**Table 2.2: Comparison MAPE for Holt Winters Exponential Smoothing forecast and test results**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Site** | **24Hour Prediction** | | | | | | | **168 Hour Prediction** | | | | | | |
| **MAPE  in %** | **t-test** | | | | **Stats** | | **MAPE  in %** | **t-test** | | | | **Stats** | |
| **h** | **P** | **ci** | | **T-stat** | **Sd** | **h** | **P** | **ci** | | **t-stat** | **Sd** |
| Bagalkot | 16.655 | 0 | 0.552 | -0.097 | 0.209 | 0.446 | 0.880 | 19.778 | 0 | 0.744 | -0.108 | 0.166 | 0.458 | 0.902 |
| Vijayapura | 16.368 | 0 | 0.657 | -0.117 | 0.221 | 0.454 | 0.991 | 19.801 | 0 | 0.690 | -0.111 | 0.183 | 0.534 | 0.965 |
| Bengaluru | 16.498 | 0 | 0.509 | -0.212 | 0.233 | 0.453 | 1.017 | 19.818 | 0 | 0.975 | -0.228 | 0.198 | 0.486 | 1.402 |

Table 2.2 compares forecasting MAPE and test results for 24hr and 168hr for three sites. It is observed that there is an improvement of 7% decrease in MAPE value as compared with LMS method with use of smoothing method. Test results indicate smoothing effect increases probability of model getting passed in t-test. It is observed that model predicts wind speed value one time step before target series is fed to system. It is observed that predicted values have a greater error margin at extremes of series. Main drawback of this model is that it is not capable of predicting wind speed values beyond extent of original target series. It is observed that accuracy is low at extremes making it hard to use it in turbulent conditions. MAPE of 16.655% for Bagalkot, 16.368% for Vijayapura and 16.498% for Bengaluru site are observed for 24hr prediction and 19.778% for Bagalkot, 19.801% for Vijayapur and 19.818% for Bengaluru site.

Fig. 2.9 to 3.14 presents 24hr and 168hr prediction for three sites. Performance of Holt Winters Exponential smoothing is improved for both 24hr and 168hr prediction as compared with LMS method described. Forecasted curves are smooth and spikes are avoided in Bagalkot, Vijayapura and Bengaluru sites. Further regression coefficient is 0.239 for Vijayapura site which very less as compared to 0.611 and 0.639 for Bagalkot and Bengaluru sites respectively.

**2.3 Time Series Analysis**

Wind speed Time Series is a stochastic non-stationary process for which properties of TS are described by Box EP & Jenkins [46]. Time Series is a set of sequential events measured with fixed interval of time.

**Stochastic Process** is a nature of statistical phenomenon of that evolves in time in laws of probability.

**Stationary Stochastic Process** is in a meticulous state of statistically equilibrium and said strictly stationary when joint probability distribution associated with events, at any set of intervals is same as that associated with events observed at times.

**Variance and Mean of a Stationary Process:** when, stationarity assumption implies that stochastic progression has a constant mean. Mean and variance of is estimated by sample mean and variance given by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.25) |
|  |  | (2.26) |

**Auto-covariance and Auto-correlation Coefficients:** stationarity assumption implies that combined distribution is identical for all times which are at constant interval apart. Auto covariance at is given by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.27) |

Similarly auto-correlation at is given by:

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  | (2.28) |

For stationary process, variance is same at time and. Thus auto-correlation at , that is correlation between and , is Which implies that .

**Positive Definiteness and Auto-covariance Matrix:**

Covariance vector associated with stationary process for events is made at successive intervals given by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.29) |

**Satisfactory Conditions by Auto-correlations:**

Positive definiteness of auto-correlation matrix implies that its determinant and all principal minors are greater than zero. In particular for,

|  |  |  |
| --- | --- | --- |
|  |  | (2.30) |

So that, and hence, Similarly for ,

|  |  |  |
| --- | --- | --- |
|  |  | (2.30a) |

Which implies that, ,

**2.3.1 Stationary Models: Mixed Autoregressive Moving Average ARMA**

A Stochastic model is particularly useful in representing practical occurring series. Current value of even is derived as a finite, linear aggregate of previous values of event and shock at. Let values of a event denoted at equally spaced times t, t-1, t-2, … by zt, zt-1, zt-2, ….. and zt, zt-1, zt-2,.. be deviation from µ; for example , . Then ARMA is written as [46]:

|  |  |  |
| --- | --- | --- |
|  |  | (2.31) |

And it is written as infinite AR process as

|  |  |  |
| --- | --- | --- |
|  |  | (2.32) |

If event are really MA (1), then a non-parsimonious depiction in terms of an AR model is obtained. However, AR (1) is not parsimoniously derived using a MA process. To obtain parameterization parsimoniously, it is necessary to include both AR and MA terms in model. Thus,

|  |  |  |
| --- | --- | --- |
|  |  | (2.33) |
| Or |  |  |
|  |  | (2.34) |

It is called as mixed ARMA process of order (p,q), which is abbreviated as ARMA(p,q).., (2.34) is written as:

|  |  |  |
| --- | --- | --- |
|  |  | (2.35) |

Since where is average of event in stationary, general ARMA(p,q) in (2.35) is expressed as:

|  |  |  |
| --- | --- | --- |
|  |  | (2.36) |

Where constant term in model is given by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.37) |

Behavior of AR, MA and ARMA models are described in Table 2.3.

**Table 2.3: Summary of properties of AR, MA and Mixed ARMA process [46]**

|  |  |  |  |
| --- | --- | --- | --- |
| **Condition** | **AR** | **MA** | **ARMA** |
| Model expressed as previous |  |  |  |
| Model expressed as previous a |  |  |  |
|  | Series is Finite | Series is Infinite | Series is Infinite |
|  | Series is Finite | Series is Finite | Series is Infinite |
| Stationary condition | roots lie outside unit circle | Constantly stationary | Roots lie outside unit circle |
| Invertibility Condition | Always invertible | Roots of lie outside unit circle | Roots lie outside unit circle |
| ACF | damped Infinite exponentials and /or damped sine waves)  Off Tails | Finite  Cuts off after lag q | Infinite exponentials and/or sine waves damped after q-p lags  Off Tails |
| PACF | Finite | Infinite damped exponentials and sine waves | Infinite damped exponentials and/or sine waves after first p-q lags |
| Cuts off after lag p | Tails off | Tails off |

**2.3.2 Non-stationary Model: Autoregressive Integrated Moving Average Model**

If roots of lie outside unit circle, process of ARMA is stationary conversely exhibit volatile non-stationary behavior if roots lie inside unit circle. If roots of lie on unit circle, resulting models are of great importance in expressing homogeneous TS of non-stationary characteristics. Let us consider model given by Box Jenkins [45]:

|  |  |  |
| --- | --- | --- |
|  |  | (2.38) |

Where is a non-stationary AR operator such that of roots of are unity and remainder lie in outside unit circle, then (2.38) is written as:

|  |  |  |
| --- | --- | --- |
|  |  | (2.39) |

Where is a stationary AR operator. Since, for d ≥ 1, where is difference operator, model is written as:

|  |  |  |
| --- | --- | --- |
|  |  | (2.40) |

Equivalently, it is defined by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.41) |

Thus assuming dth difference of series is derived by a stationary invertible ARMA process.

**Three Explicit Forms of ARIMA**

Observing ARIMA model characteristics, three explicit forms are presented as:

**Forecasts from Difference Equation:** Taking conditional expectations at time t in (2.31), obtained as

|  |  |
| --- | --- |
|  |  |

(2.42)

**Forecasts from Integrated Form:**

|  |  |
| --- | --- |
|  |  |

**Forecasts as a weighted average of previous observations and forecasts made at previous lead times from same origin**

Taking conditional expectations in (2.42) yields

|  |  |  |
| --- | --- | --- |
|  |  | (2.44) |

**Table 2.4: Behavior of ACF of dth difference of an ARIMA process of order (p,d,q) [45]**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameters** | **Order** | | | | |
| **(1,d,0)** | **(0,d,1)** | **(2,d,0)** | **(0,d,2)** | **(1,d,1)** |
| Behavior of | Decays exponentially | Only nonzero | Mixture of exponentials or damped sine wave | Only and nonzero | Reduce exponentially from first lag |
| Behavior of |  | Exponential dominates decay | Only and nonzero | Dominated by mixture of exponentials or damped sine wave | Exponential reduce from first lag. |
| Preliminary estimates from: |  |  |  |  |  |
| Admissible region |  |  |  |  |  |

Behaviour of ACF for dth difference is shown Table 2.4 indicates that ACF lags decays exponentially from first lag itself. After dth difference lags reduce exponentially from first lag. Dominance by mixture of exponentials and damped sine waves for.

**2.3.3 Criteria for Selection of Models**

In proposed method of optimizing model parameters, Akaike information criterion (AIC) and Final Prediction Error (FPE) are used as model selection criteria’s are used to find best model [100].

**Akaike Information Criteria (AIC):** It measures relative quality of statistical models for a given set of data and parameters. AIC deals with complexity and [goodness of fit](http://en.wikipedia.org/wiki/Goodness_of_fit" \o "Goodness of fit). For statistical model, value of AIC is given by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.45) |

Where, k is quantity of parameters, and L is value of likelihood maximized function. Model selected which has minimum AIC value. AIC rewards goodness of fit and includes a penalty with increasing function parameters. This discourages [over-fitting](http://en.wikipedia.org/wiki/Overfitting" \o "Overfitting) of models.

**Final Prediction Error:** It gives an evaluation of model quality in situation where models are tested with different pattern. Model with smallest FPE is given by [100]:

|  |  |  |
| --- | --- | --- |
|  |  | (2.46) |

Where, *d* is quantity of parameters, *V* is function of loss, and *N* is quantity of parameters. Function of loss *V* is given by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.47) |

Where,  expresses parameters estimated.

Analytical method of identifying ARMA model with help of ACF and PACF is shown in Fig. 2.15. Model orders are identified by plotting ACF and PACF. By manually observing ACF and PACF, it is found that there exists an AR of order 1. Since ACF has infinite lags which remains positive. But PACF drops after first lag. Since model is identified as AR(1). Although this is AR (1), now have to consider MA (1) also to consider white noise error in data which helps to improve forecasting accuracy. ARMA model is developed using armax command in matlab (model = armax (data, order). This model is used in predicting future values using predict command (YP=predict (model, data, horizon). predicted variable YP contains same size of data as original data. But first YP (1: horizon) data is set to zero, data after this period is predicted data. Predicted data YP (horizon: 2\*horizon). This predicted data is compared with actual value of future.

**2.3.4 Analytical and Iterative Method of Identifying Model**

Fig. 2.16 presents procedure for optimizing model parameters to find best model with lowest AIC & FPE values. Data is plotted to visualize outliers, missing, trend, seasonality and cyclic pattern in order to make data suitable for developing ARIMA model. Outliers and is replaced by suitable value by means of averaging neighboring data.

Yes

NO

Use p,q in armax(data, [p q])

Predict(model, data)

Error, MAPE

Plot Actual, Forecasted, Error

A

A

Plot ACF, PACF, and Spectrum to find stationarity

Identify AR MA orders

Is Stationary?

Data=diff(data)

Data

Remove Outliers, missing data

**Fig. 2.15:** **Flowchart for Analytical Method.**

Seasonality, cyclic pattern and stationarity is found by plotting ACF, PACF and Spectrum plots. Non-stationarity is converted to stationary by suitably differencing data.

**Algorithm for Selection of Model**

An algorithm is developed for optimally selecting models with lesser MAPE and t-test passed models are selected. Procedure to select models is as given as follows:

Step 1: Load hourly wind speed data and process, remove outliers, missing data.

Step 2: Specify maximum order value for pmax, dmax, qmax of ARIMA (p,d,q).

Step 3: Construct table maximum possible combination of orders using pmax, dmax, qmax.

Step 4: Select combination 1. If d > 0, then difference data d times.

Step 5: Construct model using orders selected in step 4.

Step 6: Calculate forecasting value using constructed model. Calculate forecasting error.

Step 8: Calculate Mean Absolute Percentage Error (MAPE)

Step 9: Store forecasted values, errors, MAPE for further use. Select next model.

Step 10: Repeat Step 4 to Step 9 until MAPE become optimum.

Step 11: If optimum value of MAPE not obtained from values of pmax, dmax, qmax selected in step 2, increase orders and again select a different set of orders and repeat steps from step 3 to step 10. Choose Model which gives optimum MAPE.

|  |
| --- |
| YES  YES  YES  YES  NO  NO  NO  NO  Load Model Data, test data  Select d=0  Select p=0  Select q=0  Select ARIMA(p,d,q)  Select q=q+1  Select p=p+1  Select d=d+1  Predict ARIMA(p,d,q)  AIC, BIC, MAPE  Select model for further forecast  d>=dmax  AIC,BIC,MAPE<min  q>=qmax  P>=pmax  (popt,dopt,qopt)  Select q=0  Select q=0 |
| **Fig. 2.16: Iterative Algorithm for Identification of Model by varying p,d,q orders** |

YES

NO

NO

YES

YES

YES

NO

Tuning of parameters

A

DIFF

Check goodness of fit

Check normally distribution

Durbin-watson test

Ljung box test

Engles ARC HT Test

If DW,LJ, Arch=1

Forecast

Monte Carlo simulation

Compare model output &

Measured output

Optimum model order

Forecast from Optimum model

NO

Data

ACF & PACF

PLOT ACF & PACF

Is ACF line decaying

Is Stationary?y

PDQ

AIC BIC for Seasonal, Non-Seasonal

Create SFM ARIMA Trade

Calculate Parameters for Non seasonal

Calculate Parameters seasonal

Is AIC, BIC<min (AIC,BIC)

Change Orders P,D,Q

A

**Fig. 2.17: Box Jenkins Methodology for TS Forecasting**

Flowchart for identification of models is shown in Fig. 2.17. Orders of ARIMA model, (p,d,q) are initialized to zero. A random value of AIC and FPE are set. Model developed based on initial (p,d,q) orders. Then AIC and FPE value is calculated. In iteration process if AIC and FPE is calculated and updated for each of model. Best model with pbest and qbest are used to develop model and used for forecasting future values. Forecasted values for required horizon are compared with actual value. Forecasted error is measured using Mean Absolute Percentage Error (MAPE).

**2.3.5 Results of Time Series Models**

Two methods are used to select models for wind speed forecasting. Analytically models are identifying by visualizing ACF and PACF plots. It is observed that mean is varying over all time period, spectrum plot shows higher frequency dominance, ACF plot shows a exponentially decaying cyclic lags and PACF plot indicates one significantly higher lag indicating AR order 1.

Fig. 2.18(a)-3.20(a) presents training data for three sites. ACF, PACF, spectrum plot for original wind speed and first differenced wind speed data are plotted. There is no accurate order found from original data ACF. Further it is found after differencing of original data in Fig. 2.18(b)-3.20(b). Mean of differenced data became constant, ACF has one significant lag and one significant PACF lag indicating ARIMA (1,1,1). Results of ARIMA (1,1,1) is presented in Fig. 2.21-3.23.

**Table 2.5: Comparison of MAPE values for 24hr and 168hr ahead forecast for Analytical method ARIMA (1,1,1)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Site** | **MAPE  in %** | **t-test** | | | | **Stats** | | **MAPE  in %** | **t-test** | | | | **Stats** | |
| **h** | **P** | **ci** | | **T-stat** | **Sd** | **h** | **P** | **ci** | | **t-stat** | **Sd** |
| Bagalkot | 27.496 | 0 | 0.276 | -0.049 | 0.105 | 0.223 | 0.440 | 25.027 | 0 | 0.372 | -0.054 | 0.083 | 0.229 | 0.451 |
| Vijayapura | 25.905 | 0 | 0.329 | -0.059 | 0.111 | 0.227 | 0.496 | 25.515 | 0 | 0.345 | -0.056 | 0.092 | 0.267 | 0.483 |
| Bengaluru | 29.772 | 0 | 0.255 | -0.106 | 0.117 | 0.227 | 0.509 | 31.269 | 0 | 0.488 | -0.114 | 0.099 | 0.243 | 0.701 |

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Bagalkot** | **Vijayapura** | **Bengaluru** |
|  |  |  |
| **(a)Original data** | **(a)Original data** | **(a)Original data** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| **(b)First Differenced data** | **(b)First Differenced data** | **(b)First Differenced data** |
| **Fig. 3.18: Training data of Bagalkot site for Analytical method.** | **Fig. 2.19: Training data of Vijayapura site for Analytical method.** | **Fig. 2.20: Training data of Bengaluru site for Analytical method.** |

**Iterative Method**

Analytical method has MAPE value between 25.905% for 24hrs and 31.269% for 168hrs forecasting. This MAPE value is higher as compared to other method. Since, in analytical method only one model is identified during identification process. Possibility of checking performance of other models is lost. Hence an Iterative method explained in Fig. 2.17 is developed for wind speed forecast in which number of alternatives ARIMA(p,d,q) models is tested to find an optimum model. Table 2.6 gives results of iterative ARIMA Model. Maximum order of 3 is initialized for ARIMA model i.e pmax, dmax, qmax = 3,3,3. Hence 27 nos of alternative models are obtained during iteration. Out of 27 models, model with least MAPE is selected as optimum and shown in Table 2.6. For all three sites have different set of orders. Observed that, orders are selected based purely on data pattern. Also forecasting error minimized as compared to analytical method of wind speed forecast. With this method an improvement of 3% is observed.

Forecasted results for analytical ARIMA model are shown in Fig. 2.21-3.24 for 24hr and Fig. 2.25-3.27 for 168hr prediction. Results reveal that forecasted curves are nearly following actual values and curves are smoother as compared to analytical method. Further highest regression found from iterative model. Distribution of error is uniform in iterative model. This concludes that, although analytical method is simple but failed to forecast accurately as compared to iterative identification method. Hence iterative approach of identification of model parameters is selected for further in combination process.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 2.6: Comparison of MAPE values for 24 hr and 168 hr ahead forecast for Iterative method** | | | | |
| **Site** | **Model identified** | **MAPE for 24 hr**  **forecast in %** | **Model Identified** | **MAPE for 168 hr**  **forecast in %** |
| Bagalkot | ARIMA(1,1,2) | 21.643 | ARIMA(3,1,2) | 23.741 |
| Vijayapura | ARIMA(2,1,2) | 22.709 | ARIMA(3,1,1) | 24.710 |
| Bengaluru | ARIMA(2,1,3) | 22.243 | ARIMA(3,1,3) | 23.076 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.21: (a) Actual Vs Forecasted wind speed for 24hr using Analytical method (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Bagalkot Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.22: (a) Actual Vs Forecasted wind speed for 24hr using Analytical method (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Vijayapura Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.23: (a) Actual Vs Forecasted wind speed for 24hr using Analytical method (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Bengaluru Site** |
|  | | |
| |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.24: (a) Actual Vs Forecasted wind speed for 168hr using Analytical method (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Bagalkot Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.25: (a) Actual Vs Forecasted wind speed for 168hr using Analytical method (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Vijayapura Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.26: (a) Actual Vs Forecasted wind speed for 168hr using Analytical method (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Bengaluru Site** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.27: (a) Actual Vs Forecasted wind speed for 24hr using Iterative method (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Bagalkot Site** | |  |  | | --- | --- | |  | | |  |  | |   **Fig. 2.28: (a) Actual Vs Forecasted wind speed for 24hr using Iterative method (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Vijayapura Site** | |  |  |  | | --- | --- | --- | |  | | | |  |  |   **Fig. 2.29: (a) Actual Vs Forecasted wind speed for 24hr using Iterative method (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Bengaluru Site** |
|  | | |
| |  |  | | --- | --- | |  | | |  |  |   **Fig. 3.30: (a) Actual Vs Forecasted wind speed for 168hr using Iterative method (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Bagalkot Site** | |  |  |  | | --- | --- | --- | |  | | | |  |  |   **Fig. 2.31: (a) Actual Vs Forecasted wind speed for 168hr using Iterative method (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Vijayapura Site** | |  |  |  | | --- | --- | --- | |  | | | |  |  |   **Fig. 2.32(a) Actual Vs Forecasted wind speed for 168hr using Iterative method (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Bengaluru Site** |

**2.4 Generalized Autoregressive Conditional Heteroskedasticity Modeling (GARCH)**

In practice, volatility of wind speed tends to change over time. Consequently, conventional TS models seem unattractive for variables like wind speed. Engle (1982) are first to develop AR Conditional heteroskedasticity (ARCH) model for forecasting conditional variance of TS. In an ARCH (p) model, conditional variance depends upon p lagged square errors [51].

**2.4.1 Mathematical Modeling of GARCH (1,1)**

In an ARCH (1) approach, next phase variance is only non independent on last phase squared remaining so a disaster that affected a huge residual would not have class of persistence that authentic disasters are inspected. Because of this reason it has leads to addition of ARCH approach to a GARCH, or indiscriminate ARCH approach are initially developed by Bollerslev (1986), which is identical to ARMA approach, in a GARCH(1,1) approach[52]. Variance of GARCH is given by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.48) |

Where is variance, are non-negative constraints, is previous non-conditional variance, is squared residuals in GARCH

Observing eq.(2.48) as an ARCH (1) approach is an AR(1) approach on residuals, and ARCH(1,1) approach is an ARMA(1,1) approach on squared residuals by making similar substitutions as before, variance at time is given by

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  | (2.49) |
|  |  | (2.50) |
|  |  | (2.51) |

variance of unconditional αt is

|  |  |  |
| --- | --- | --- |
|  |  | (2.52) |

“Var(at )is stationery task”

GARCH(1,1) is written as an ARCH():

|  |  |  |
| --- | --- | --- |
|  |  | (2.53) |

It is written as:

|  |  |  |
| --- | --- | --- |
|  |  | (2.54) |

Since non conditional returns of variance is .

GARCH (1, 1) expressed as:

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  | (2.55) |

Variance is biased combination of non conditional variance of returns, proceeding phase’s squared residuals and previous phase’s non unconditional variance with loads. Prediction for next phase variance is given by:

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  | (2.56) |

It is observed from (2.56) that, so as to predict horizon which goes to ∞.

Spectrum Analysis

Identification of Stationary

Model Identification

Model Estimation

Forecast and Evaluation

Heteroskedasticity Test

Evaluation of ARIMA/GARCH Model Performances

Examination of Forecast Accuracy Measures

Model Identification and Estimation

ARIMA

GARCH

**Fig. 3.33: Flowchart of GARCH analysis**

**2.4.2 Results of GARCH(1,1)**

ARIMA models used to develop GARCH model by deploying (2.56) for estimating future variance and hence future value of wind speed. Following results obtained from GARCH (1,1) model for three sites with forecasting horizon of 24hr and 168 ahead wind speed.

Results of GARCH model is shown in Table 2.7. There is an improvement of 2% to 3% in MAPE as compared to conventional ARIMA model. Test results shows significant improvement over ARIMA model. MAPE for Bagalkot is 19.618%, 17.788% for Vijayapur and 19.594% obtained for Bengaluru site. Forecasted wind speed values, error distribution and regression between actual vs forecasted values are represented in Fig. 2.34 to 3.39. Regression value range form 0.551 to 0.656 for 24hr and 0.458 to 0.631 for 168hr prediction.

**Table 2.7: Comparison of MAPE values for 24hr and 168hr ahead forecast for GARCH(1,1)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Site** | **MAPE  in %** | **t-test** | | | | **Stats** | | **MAPE  in %** | **t-test** | | | | **Stats** | |
| **h** | **P** | **ci** |  | **T-stat** | **Sd** | **h** | **P** | **ci** |  | **t-stat** | **Sd** |
| Bagalkot | 19.618 | 0 | 0.414 | -0.073 | 0.157 | 0.335 | 0.660 | 20.576 | 0 | 0.559 | -0.082 | 0.125 | 0.344 | 0.677 |
| Vijayapura | 17.788 | 0 | 0.493 | -0.088 | 0.166 | 0.341 | 0.743 | 19.634 | 0 | 0.518 | -0.083 | 0.138 | 0.401 | 0.724 |
| Bengaluru | 19.594 | 0 | 0.382 | -0.159 | 0.175 | 0.340 | 0.763 | 21.107 | 0 | 0.732 | -0.172 | 0.149 | 0.365 | 1.052 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.34: (a) Actual Vs Forecasted wind speed for 24hr using GARCH(1,1) model (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Bagalkot Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.35: (a) Actual Vs Forecasted wind speed for 24hr using GARCH(1,1) model (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Vijayapura Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.36: (a) Actual Vs Forecasted wind speed for 24hr using GARCH(1,1) model (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Bengaluru Site** |
| |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.37: (a) Actual Vs Forecasted wind speed for 168hr using GARCH(1,1) model (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Bagalkot Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.38: (a) Actual Vs Forecasted wind speed for 168hr using GARCH(1,1) model (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Vijayapura Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.39: (a) Actual Vs Forecasted wind speed for 168hr using GARCH(1,1) model (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Bengaluru Site** |

**2.5 Transfer Function ARIMA**

Although uni-variate ARIMA model can dynamically describe relationship of Time Series data, it cannot check casualty between independent variable and interpreted variable. Given an input series wind speed and output series temperature in raw data form, four main stages and various sub-stages in complete process of transfer function model building is described as:

**Stage 1:** Identification of model form

* 1. : Preparation of input and output series, Pre-whitening input and output series
  2. : Computing cross and autocorrelations for pre-whitened input and output series
  3. : Direct estimation of impulse weights
  4. : Specifying (r,s,b) for transfer function model relating input and output series.
  5. : Preliminary estimation of noise series (nt) and computation of acf and pacf for this series. Specifying (pn,qn) for ARIMA(pn,0,qn) model of noise series(nt)

**Stage 2:** Estimation of parameters of Transfer Function model

2-1: Preliminary estimates of parameters final estimates of parameters

**Stage 3:** Diagnostic testing of transfer function model

3-1: Computation of ACF for residuals of (r,s,b) model linking input and output series.

3-2: Computation of cross-correlation between residuals in 3-1 and pre-whitened noise series.

**Stage4:** Using Transfer function model for forecasting

Three case studies are taken up to find effect of volume of data, forecasting horizon and orography on forecasting accuracy using Transfer Function ARIMA model.

**2.5.1 Mathematical Modeling of Transfer Function ARIMA Model**

Transfer function model TRFU(r,s,b) is given by[46]:

|  |  |  |
| --- | --- | --- |
|  |  | (2.57) |

where, is a polynomial in B of order , is a polynomial in B of order , and disturbance process. Transfer function ARIMA model for present work is given by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.58) |

Where,

is wind speed time series as output

is temperature time series as input

Original time series are converted to pre-whiten series in order to make them suitable for transfer function model.

|  |  |  |
| --- | --- | --- |
|  |  | (2.59) |
|  |  | (2.60) |

**2.5.2 Estimation of Parameters for Transfer Function ARIMA Model**

Parameters for TRFU-ARIMA model are estimated by using hourly data and 10min data are calculated as shown in Table 2.8 to 3.11. While calculating parameters wind speed data for 2007 to 2017 of BEC Bagalkot, Karnataka, India is considered.

# Table 2.8: ARIMA Equation for 10min interval data

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data** | **Model**  **ARIMA**  **(p,d,q)** | **AIC** | **FPE** | **Phi** | **Theta** |
| 2007 | 2,1,2 | -0.7809 | 0.4620 | 1+0.05101q1-0.5471 q2 | 1-0.0422p1-0.6053p2 |
| 2008 | 1,1,1 | -0.7774 | 0.5036 | 1-0.680388q1 | 1-0.7441 p1 |
| 2009 | 2,1,2 | 0.6507 | 0.9250 | 1-0.2098q1-0.2651 q2 | 1-0.3021 p1-0.4565 p2 |
| 2010 | 1,1,2 | -0.7112 | 0.5021 | 1-0.8440 q1 | 1-1.0635 p1+0.6661 p2 |
| 2011 | 1,1,1 | -0.6042 | 0.3654 | 1-0.53001 q1 | 1-0.8852 p1 |
| 2012 | 1,1,1 | -0.7215 | 0.6122 | 1-0.5861 q1 | 1-0.8802 p1 |
| 2013 | 2,1,2 | -0.6912 | 0.4521 | 1+0.06321 q1-0.6512 q2 | 1-0.0533 p1-0.5042 p2 |
| 2014 | 2,1,2 | -0.7021 | 0.6625 | 1+0.06532 q1-0.5932 q2 | 1-0.0635 p1-0.6132 p2 |
| 2015 | 2,1,2 | 0.7109 | 0.8012 | 1-0.1999 q1-0.2531 q2 | 1-0.1287 p1-0.6662 p2 |
| 2016 | 1,1,2 | -0.7036 | 0.5054 | 1-0.9541 q1 | 1-1.1112 p1+0.3250 p2 |
| 2017 | 1,1,1 | -0.5236 | 0.4502 | 1-0.4518 q1 | 1-0.9565 p1 |

**Table 2.9: ARIMA parameters for Hourly data**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data** | **Model ARIMA**  **(p,d,q)** | **AIC** | **FPE** | **Phi** | **Theta** |
| 1 month | 1,1,1 | 0.00230 | 0.61698 | 1-0.77109 q1 | 1-0.8818 p1 |
| 6 months | 2,1,2 | 0.02530 | 1.97497 | 1-1.8106 q1+0.8674 q2 | 1-1.905 p1+0.9201 p2 |
| 2007 | 2,1,2 | 0.02658 | 1.90470 | 1-1.7895 q1+0.8354 q2 | 1-1.889 p1+0.9136 p2 |
| 2008 | 2,1,2 | 0.02666 | 1.23650 | 1-1.7779 q1+0.8374 q2 | 1-1.8915 p1+0.9043 p2 |
| 2009 | 2,1,1 | 0.02709 | 1.73110 | 1-0.9517 q1+0.1503 q2 | 1-0.9837 p1 |
| 2010 | 2,1,2 | 0.02769 | 1.36584 | 1-0.8210 q1-0.1591 q2 | 1-0.8210 p1-0.1591 p2 |
| 2011 | 2,1,2 | 0.02804 | 2.43119 | 1-0.5812 q1-0.2351 q2 | 1-0.8865 p1-0.2236 p2 |
| 2012 | 2,1,2 | 0.02947 | 1.36147 | 1-0.8965 q1-0.2265 q2 | 1-0.8615 p1-0.2312 p2 |
| 2013 | 2,1,2 | 0.02949 | 1.30701 | 1-0.5568 q1-0.1658 q2 | 1-0.8896 p1-0.2254 p2 |
| 2014 | 2,1,1 | 0.02951 | 1.96051 | 1-0.7812 q1-0.2384 q2 | 1-0.9024 p1 |
| 2015 | 2,1,2 | 0.02954 | 2.94077 | 1-0.9063 q1-0.2014 q2 | 1-0.8854 p1-0.2036 p2 |
| 2016 | 2,1,2 | 0.02960 | 2.82314 | 1-0.8362 q1-0.1674 q2 | 1-0.8825 p1-0.2152 p2 |
| 2017 | 2,1,2 | 0.02967 | 1.97497 | 1-0.8815 q1-0.1785 q2 | 1-0.9047 p1-0.1265 p2 |

**Table 2.10: ARIMA parameters for Multiyear hourly data**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data** | **Model**  **ARIMA**  **(p,d,q)** | **AIC** | **FPE** | **Phi** | **Theta** |
| 2007-08 | 2,1,2 | 0.04491 | 1.88560 | 1-1.7929 q1+0.8851 q2 | 1-1.8986 p1+0.9120 p2 |
| 2007-09 | 2,1,2 | 0.04381 | 1.36500 | 1-1.7885 q1+0.84558 q2 | 1-1.8991 p1+0.9120 p2 |
| 2007-10 | 2,1,2 | 0.04222 | 1.91100 | 1-1.7645 q1+0.8221 q2 | 1-1.8861 p1+0.8979 p2 |
| 2007-11 | 2,1,2 | 0.04012 | 1.50778 | 1-1.7956 q1+0.8014 q2 | 1-1.8851 p1+0.7991 p2 |
| 2007-12 | 2,1,2 | 0.03699 | 2.68385 | 1-1.689 q1+0.8809 q2 | 1-1.779 p1+0.8684 p2 |
| 2007-13 | 2,1,2 | 0.03695 | 1.50295 | 1-1.853 q1+0.8874 q2 | 1-1.983 p1+0.9047 p2 |
| 2007-14 | 2,1,2 | 0.0364 | 1.44284 | 1-1.8891 q1+0.93654 q2 | 1-1.8704 p1+0.9147 p2 |
| 2007-15 | 2,1,2 | 0.0354 | 2.16425 | 1-1.5846 q1+0.8863 q2 | 1-1.9987 p1+0.9047 p2 |
| 2007-16 | 2,1,2 | 0.0236 | 3.24638 | 1-1.689 q1+0.9054 q2 | 1-1.8561 p1+0.9153 p2 |
| 2007-17 | 2,1,2 | 0.0254 | 3.11653 | 1-1.658 q1+0.8855 q2 | 1-1.8847 p1+0.9684 p2 |

**Table 2.11: Transfer function ARIMA parameters**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **ACF**  **α** | **ACF**  **β** | **ACF**  **(diff2((xt)** | **ACF**  **(diff2((yt)** | **ACF**  **noise** | **Cross**  **Corr**  **α,β** | **PACF**  **noise** | **Spectrum**  **Noise** |
| 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | -0.004 | 1.000 | 0.000 |
| 0.232 | 0.048 | -0.371 | -0.525 | -0.518 | -0.074 | -0.518 | 0.000 |
| -0.346 | -0.379 | -0.161 | -0.027 | -0.039 | -0.0063 | -0.391 | 0.000 |
| -0.231 | -0.142 | 0.002 | 0.024 | 0.029 | -0.053 | -0.293 | 0.000 |
| -0.091 | -0.049 | 0.007 | -0.003 | -0.001 | 0.015 | -0.241 | 0.000 |
| -0.032 | -0.02 | -0.002 | 0.003 | 0.003 | 0.006 | -0.201 | 0.001 |
| 0.001 | -0.007 | 0.004 | -0.003 | -0.011 | 0.003 | -0.176 | 0.001 |
| 0.004 | 0.001 | 0.013 | 0.004 | -0.001 | 0.001 | -0.158 | 0.001 |
| -0.002 | 0.009 | -0.018 | -0.007 | -0.009 | -0.002 | -0.163 | 0.002 |
| 0.014 | 0.015 | 0.010 | 0.018 | 0.029 | -0.007 | -0.114 | 0.003 |
| 0.004 | 0.004 | 0.006 | -0.001 | -0.001 | 0.009 | -0.094 | 0.006 |
| 0.001 | 0.01 | 0.001 | 0.001 | -0.002 | 0.009 | -0.094 | 0.007 |
| -0.002 | 0.007 | -0.008 | 0.011 | 0.013 | -0.005 | -0.068 | 0.008 |
| 0.008 | -0.011 | 0.009 | -0.015 | -0.013 | -0.001 | -0.07 | 0.009 |
| 0.005 | -0.004 | 0.006 | 0.008 | 0.007 | 0.011 | -0.055 | 0.012 |

It shows that hourly wind speed data is more suitable for wind speed forecasting. AIC value decreases as volume of wind data increases. Expression for transfer function -ARIMA model is developed using parameters given in Table 2.11 is formulated by:

|  |  |
| --- | --- |
|  | (2.61) |

Where,

is wind speed

is temperature at time interval

is non seasonal differencing.

Procedure of building transfer function model involves steps of model identification, estimate, fitting check and forecasting is presented in Fig. 2.40.

|  |
| --- |
| Wind Speed and Temperature data  Sequence Plot and Examine trend for differencing  Assumption of stationarity Decide order of differencing  Model Identification Plot and examine ACF and PACF to specify model  Parameter Estimation Calculate parameter values and model fitting statistics  Forecast Perform Future Prediction  Is Model Adequate  ?  Is it most suitable model?  No  No  Yes  Yes |
| **Fig. 2.40: Flowchart for Transfer Function ARIMA model** |

|  |  |
| --- | --- |
| **(a) ACF** | **(b) PACF** |
|  |  |
| **Fig. 2.41(a): ACF and PACF plot for two years 2016-17** | |

ACF and PACF remain same for remaining years of data. ACF indicates order of MA model and PACF indicates order of AR model. ACF plot for original wind speed data is presented in Fig. 2.41. This requires differencing of data in order to make it suitable for development of model. A first order differencing is done.

**2.5.3 Case Study: Effect of Volume of Data and Horizon on Forecasting Accuracy**

Wind speed for a period of one month is used as test data for comparison of forecasted wind speed data. A sample forecasted wind speed data for one day ahead is shown in Fig. 2.42 using one year data.

Fig. 2.42: Comparison of Actual Vs Forecasted wind speed for one day ahead using one year data

Similarly results found for different forecast horizon for three sites are shown in Table 2.12. Graphical representation of same is given in Fig. 2.43. It is found that up to 20hr change in MAPE is very small. After 24hr MAPE starts increasing rapidly and at 60hr reaches near 45%.

**Table 2.12: Comparison of forecasting results for 3 sites for different forecasting horizon**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Site/ Prediction  Horizon** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **20** | **24** | **30** | **40** | **50** | **60** |
| **Bagalkot** | 8.50 | 8.9 | 9.11 | 9.23 | 9.56 | 9.88 | 10.1 | 10.66 | 11.01 | 11.87 | 14.6 | 17.66 | 18.11 | 28.22 | 34.28 | 41.25 |
| **Vijayapura** | 9.27 | 9.7 | 9.93 | 10.06 | 10.42 | 10.77 | 11.01 | 11.62 | 12.00 | 12.94 | 15.91 | 17.59 | 19.74 | 30.76 | 37.37 | 44.96 |
| **Bengaluru** | 9.10 | 9.19 | 9.68 | 9.73 | 9.88 | 9.96 | 10.19 | 10.85 | 11.39 | 11.95 | 15.3 | 17.20 | 18.98 | 29.34 | 35.58 | 41.25 |

**Fig. 2.43: Comparison of forecasting error for different horizon for three Sites**

It is found that as forecasting horizon increases as forecasting error increases. Among three sites Bengaluru site has highest forecasting accuracy for all prediction horizons due to consistency in data. For 1 to 20 hour prediction there is not much reduction in forecasting performance. After 20 hour forecasting error increases drastically due to short term prediction ability of ARIMA model. For longer horizons ARIMA model fails to predict well. A case study to know effect of orography on forecasting accuracy is carried out using 10min and hourly wind speed data of Bagalkot BEC Site.

**Fig. 2.44: Actual Vs Forecasted wind speed data for one year data**

**Fig. 2.45: Actual vs Forecasted wind speed for two years**

Actual Predicted Error

**Fig. 2.46: Actual vs Forecasted wind speed for three years data**

**Fig. 2.47: Actual vs Forecasted wind speed for four years data**

**Fig. 2.48: Actual vs Forecasted wind speed for five years data**

Fig. 2.44 to 48 present actual vs forecasted wind speed curves for 1 year to 5 years. It is found that forecasted curve follows actual wind speed only for few intervals. For remaining intervals difference is large enough to discard this method. This method fails to capture peak values.

**Table 2.13: Comparison of MAPE Value for different volume of data**

|  |  |  |  |
| --- | --- | --- | --- |
| **Years of**  **data used** | **MAPE**  **Bagalkot** | **MAPE**  **Vijayapura** | **MAPE**  **Bengaluru** |
| One year | 28.5457 | 29.3451 | 28.6547 |
| Two year | 24.6788 | 25.0265 | 24.9580 |
| Three year | 23.7311 | 24.2365 | 23.9587 |
| Four year | 23.5544 | 24.3108 | 23.9651 |
| Five year | 23.5287 | 24.3509 | 24.0698 |
| Six Year | 23.5229 | 24.2501 | 24.1253 |
| Seven Year | 23.5022 | 24.2207 | 24.1022 |
| Eight Year | 23.5005 | 24.1999 | 24.0356 |
| Nine Year | 23.4999 | 24.1967 | 24.0250 |
| Ten Year | 23.4550 | 24.1652 | 24.0010 |

**Fig. 2.49: Comparison of forecasting Error for different volume of data**

Table 2.23 and Fig. 2.37 give comparison of forecasting Error for different volume of data that after three years data, forecasting error not vary significantly. **It is found that, it is sufficient to develop forecasting model using three year wind speed data. Further forecasting MAPE increases rapidly after 24hr horizon.**

**2.6 Wavelet Analysis**

Wavelet basis function is used to represent many scale functions. This is generated by translated and scaled mother wavelet. Wavelet is defined as small wave. Wavelet transforms have an infinite possible set of source functions, whereas Fourier transform uses sine and cosine functions. Wavelet transforms do not have single set of source functions. Because of that wavelet analysis gives immediate admission to data that which is masked by new time frequency approaches. Very small source functions are utilized to separate signal breaks. Apart from that, when complete frequency analysis is needed, very large source functions are utilized.

Let consider signal Ψ (t) and Fourier transform of same as **.** Then components of decayed wave given by [54]:

|  |  |  |
| --- | --- | --- |
|  |  | (2.62) |

After dilation and conversion of Ψ (t), Ψa,b() is obtained as a.

|  |  |  |
| --- | --- | --- |
|  |  | (2.63) |
| Where, a is scale factor,b is conversion factor. | |  |

**Continuous Wavelet Transforms (CWT)**

Continuous wavelet transform deduces wavelet analysis is given by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.64) |

**Discrete Wavelet Transform (DWT)**

Discrete wavelet is implemented complement of CWT. DWT is better than CWT in decaying and reconstructing most of wind velocity conflict. It reduces processing time and gives enough and sufficient data. Detailed wavelet transform is given by [54]:

|  |  |  |
| --- | --- | --- |
|  |  | (2.65) |

Where,

|  |  |  |
| --- | --- | --- |
|  |  | (2.66) |

Parameters a, b, co sampled parallel to wavelets on “Dyadic” WT.

Discrete wavelets function is given by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.67) |

Where,

Square integrals of a signal f () is viewed as demerit of estimation. To reduce arithmetic process, Mallat Algorithm deals with signal [62]. Depend on this method pattern f() is decayed into detail TS at various resolutions d1,d2……..dj and a common level number of wavelet decay are given by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.68) |

But after wavelet decay, coarse approximation and points number are divided into half with real time series f(t), because of which coarse approximation and details are regenerated by (2.68) is rewritten as:

|  |  |  |
| --- | --- | --- |
|  |  | (2.69) |

Where H\*And G\* represents operators of H and G , Dj and Aj whose point number TS is equal to genuine TS which are regenerated series of dj and aj and they satisfy S=D1*+D2+D3…….+Dj+ Aj.*

Decompose selected wavelet function in to different 1-D wavelet decayed coefficient D1, D2, D3,D4 ,D5………………. DJ and AJ where J is a scaling function. These coefficients approximately meet original signal.

**Steps for ARIMA (P, D,Q) Approach in each Wavelet Scale Domain**

Step 1: Inspect D (1<j) and AJ may be stationery. If TS is non identical then follow difference approach else follow step 2

Step 2: Verify truncating and trailing characteristics of D (1<J) and AJ TS PACF and ACF respectively .then verify factors (p,q) approaches which are built in procedure 3

Step 3: Construct ARMA models of DJ and AJ respectively and then estimate factors of ARMA Approach.

Step 4: Testing is done to check faults in a ARIMA approach which are constructed in step3.

Step 5: Utilize verified ARIMA (p,q) approach to predict upcoming values of DJ and AJ TS.

**Model Prediction**

 and are decayed signal of DJ and AJ then forecasting speed of a original wind speed is defined as .

**2.6.1 ARIMA-Wavelet Model**

Steps followed are:

* Original wind speed recurred from wind form is decayed into different scales(resolution),d1,d2,d3………dJ with coarse estimation of a TS aJ after reconstruction of decayed data DJ and AJ are obtained.
* For a different value of DJ and AJ of a TS ARIMA model is developed to predict respectively.

Detailed flowchart for Wavelet Analysis based forecasting is given in Fig. 2.50.

Wind Speed Data

Discrete Wavelet Transform

a1

d1

d2

d3

d4

d5

Apply threshold on coefficients

D1

D2

D3

D4

D5

Inversed WT

De-noised Series

Fluctuation Series

Smoothing model

Smoothing

Forecasted detailed series

Forecasted approximate series

Final Wind Speed Forecast

**Fig. 2.50: Flow chart of Wavelet analysis**

**2.6.2 Results for Wavelet ARIMA Model**

Fig. 2.54(a to d) represents Wavelet Analysis of Wind Speed TS of Bagalkot site for one year. Following ARIMA models are identified from approximate and detailed coefficients obtained in Fig. 2.54(b) for level 5 haar wavelet families. Initially prediction is done for these individual coefficients using algorithm developed in section 3.5 and then forecasts are decomposed to a single forecasting function.



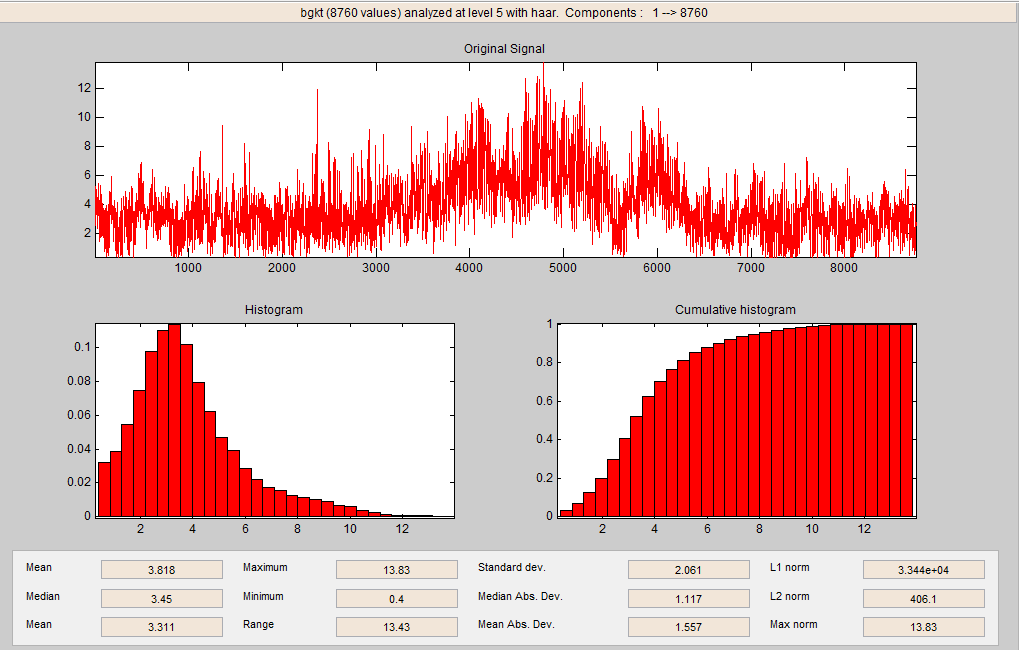
**Fig. 2.51 (a): Original and Compressed Wind Speed series of Bagalkot Site**

|  |
| --- |
|  |

**Fig. 2.51 (b): Approximations of wind speed series of data of Bagalkot Site**

|  |
| --- |
|  |
|  |
| **Fig. 2.51 (c): Debauchies 5 Level decomposition of Wind Speed** |
|  |

**Fig. 2.51 (d): Approximations of wind speed series of data of Bagalkot Site**



**Fig. 2.51 (e): Histogram and Cumulative distribution of wind speed**

**Table 2.14: ARIMA models obtained for approximate and detailed coefficients**

|  |  |  |  |
| --- | --- | --- | --- |
| **Approximate**  **coefficients** | **ARIMA**  **Model** | **Detailed**  **Coefficients** | **ARIAM Model** |
| A1 | ARIMA(2,1,1) | D1 | ARIMA(1,2,1) |
| A2 | ARIMA(2,1,2) | D2 | ARIMA(1,2,2) |
| A3 | ARIMA(2,2,1) | D3 | ARIMA(2,2,2) |
| A4 | ARIMA(2,2,1) | D4 | ARIMA(3,2,3) |
| A5 | ARIMA(2,2,2) | D5 | ARIMA(3,2,4) |

Table 2.16 presents wind speed forecasting results obtained for decomposed series after combining these approximate and detailed coefficients. Results revealed that, performance of Wavelet-ARIMA model found better than ARIMA model alone. Significant improvement in forecasting accuracy is achieved.

**Table 2.15: Comparison of MAPE values for 24hr and 168hr ahead forecast for Wavelet ARIMA model**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Site** | **24 hr Forecast** | | **168 hr Forecast** | |
| **MAPE**  **in %** | **R2** | **MAPE**  **in %** | **R2** |
| Bagalkot | 16.518 | 0.893 | 17.211 | 0.803 |
| Vijayapura | 15.104 | 0.298 | 16.887 | 0.267 |
| Bengaluru | 18.088 | 0.754 | 18.914 | 0.673 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.52: (a) Actual Vs Forecasted wind speed for 24hr using Wavelet ARIMA model (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Bagalkot Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.53: (a) Actual Vs Forecasted wind speed for 24hr using Wavelet ARIMA model (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Vijayapura Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.54: (a) Actual Vs Forecasted wind speed for 24hr using Wavelet ARIMA model (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Bengaluru Site** |
| |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.55: (a) Actual Vs Forecasted wind speed for 168hr using Wavelet ARIMA model (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Bagalkot Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.56: (a) Actual Vs Forecasted wind speed for 168hr using Wavelet ARIMA model (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Vijayapura Site** | |  |  | | --- | --- | |  | | |  |  |   **Fig. 2.57: (a) Actual Vs Forecasted wind speed for 24hr using Wavelet ARIMA model (b) Regression plot of Actual Vs Forecasted (c) Error Distribution for Bengaluru Site** |

**2.7 Comparison of Forecasted Results of Statistical Models**

In this section results of statistical models are presented graphically for both 24hr and 168hr wind speed forecasting for comparison of forecasted curves from different models.

**Fig. 2.58: Comparison of Statistical Model 24hr wind speed forecast results for Bagalkot Site**

**Fig. 2.59: Comparison of Statistical Model 24hr wind speed forecast results for Vijayapura Site**

**Fig. 2.60: Comparison of Statistical Model 24hr wind speed forecast results for Bengaluru Site**

**Fig. 2.61: Comparison of Statistical Model 168hr wind speed forecast results for Bagalkot Site**

**Fig. 2.62: Comparison of Statistical Model 168hr wind speed forecast results for Vijayapura Site**

**Fig. 2.63: Comparison of Statistical Model 168hr wind speed forecast results for Bengaluru Site**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2.16: Comparison of MAPE of statistical methods for 24hr and 168hr wind speed forecasting** | | | | | | | |
| **Sl.** | **Site** | **Bagalkot** | | **Vijayapura** | | **Bengaluru** | |
| **Method** | **24hr** | **168hr** | **24hr** | **168hr** | **24hr** | **168hr** |
| 1. | LMS | 27.671 | 27.622 | 29.060 | 27.574 | 27.430 | 26.105 |
| 2. | HW | 16.655 | 19.778 | 16.368 | 19.801 | 18.498 | 19.818 |
| 3. | Analytic | 27.496 | 25.027 | 25.905 | 25.515 | 29.772 | 31.269 |
| 4. | Iterative | 21.643 | 23.741 | 22.709 | 24.710 | 22.243 | 23.076 |
| 5. | GARCH | 19.618 | 20.576 | 17.788 | 19.634 | 19.594 | 21.107 |
| **6.** | **Wavelet-ARIMA** | **16.518** | **17.211** | **15.104** | **16.887** | **18.088** | **18.914** |
|  |  |  |  |  |  |  |  |

**Fig.3.64: Comparison of forecasting performance of statistical models**

In Table 2.17, Wavelet ARIMA model outperforms all models. Performance of combinational Wavelet ARIMA model with MAPE of 15.104% has outperformed remaining models. Analytic ARIMA (identification by ACF and PACF) model performed very bad since forecasting errors are very high and also not consistent.

Table 2.18 represents R2 values of statistical method. Wavelet-ARIMA with R2 value of 0.898 has outperformed remaining methods.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2.17: R2 Values for Statistical Methods** | | | | | | | | | |
| **Method** | **24Hour** | | | | | **168Hour** | | | |
| **Bagalkot** | | **Vijayapura** | **Bengaluru** | | **Bagalkot** | **Vijayapura** | | **Bengaluru** |
| LMS | 0.606 | | 0.537 | 0.471 | | 0.658 | 0.546 | | 0.462 |
| HWES | 0.766 | | 0.447 | 0.565 | | 0.611 | 0.236 | | 0.639 |
| Analytical ARIMA | 0.464 | | 0.596 | 0.596 | | 0.629 | 0.541 | | 0.547 |
| Iterative ARIMA | 0.796 | | 0.655 | 0.637 | | 0.546 | 0.554 | | 0.671 |
| GARCH | 0.556 | | 0.551 | 0.604 | | 0.631 | 0.458 | | 0.559 |
| **Wavelet-ARIMA** | **0.823** | | **0.828** | **0.814** | | **0.813** | **0.807** | | **0.803** |
| **Table 2.18: Regression Equations for Actual Vs Forecasted Wind Speed values for 24hour forecast** | | | | | | | | | |
| **Method** | | **24Hour** | | | | | | | |
| **Bagalkot** | | | **Vijayapura** | | | **Bengaluru** | |
| LMS | | -0.104x2+2.735x-6.394 | | | 0.011x2+0.626x+2.033 | | | -0.503x2+4.665x-6.504 | |
| HWES | | 0.964x+1.399 | | | 0.715x+2.973 | | | -0.415x2+3.744x-3.969 | |
| Analytical ARIMA | | 0.772x+3.264 | | | 0.016x2+1.090x-0.093 | | | -0.106x2+2.991x-0.099 | |
| Iterative ARIMA | | 1.011x+0.299 | | | 1.079x+0.179 | | | -0.128x2+3.062x-7.050 | |
| GARCH | | -0.015x2+1.075x-0.095 | | | 0.979x-0.165 | | | -0.148x2+2.013x-6.078 | |
| Wavelet-ARIMA | | 1.249x-0.922 | | | 0.825x+0.931 | | | 0.180x2-0.55x+3.141 | |
| **Method** | | **168Hour** | | | | | | | |
| **Bagalkot** | | | **Vijayapura** | | | **Bengaluru** | |
| LMS | | -0.025x2+1.425x-1.479 | | | -0.069x2+1.891x-2.717 | | | -0.076x2+1.573x-0.986 | |
| HWES | | 1.021x-0.226 | | | -0.022x2+0.645x-1.453 | | | 0.986x+1.142 | |
| Analytical ARIMA | | 1.194x-0.591 | | | 0.919x+0.359 | | | 0.030x2+0.516x+1.143 | |
| Iterative ARIMA | | 1.005x+0.381 | | | 0.899x+0.450 | | | -0.103x2+2.562x-5.045 | |
| GARCH | | 0.909x+0.345 | | | 0.856x+0.351 | | | -0.213x2+3.061x-4.036 | |
| Wavelet-ARIMA | | 1.150x-0.633 | | | 1.104x-0.550 | | | -0.026x2+0.741x+0.589 | |

**2.8 Statistical Tests of Forecasted Results**

Conducting multiple tests on forecasted and actual wind speed values helps in mitigating misclassification and misspecification of models. With regards to this paired t-test, F-test and regression tests are conducted for forecasted results. A test statistical hypothesis is a decision rule leading to either acceptance or rejection of hypothesis. In this thesis paired t-test, F-test and regression are conducted on model output forecasted wind speed data and error measured after prediction.

**Paired t-test**: Vartype is decided depending on whether the data has a equal mean or not. If vartype is equal, test computes a sample standard deviation using

|  |  |  |
| --- | --- | --- |
|  |  | (2.70) |

Where *Sx* and *Sy* are sample standard deviations of X and Y, respectively, and *N* and *M* are sample sizes of X and Y, respectively. It returns *P* value a probability, under null hypothesis, t-statistic is given by:

|  |  |  |
| --- | --- | --- |
|  |  | (2.71) |

Where *Sx* and *Sy* are sample standard deviations, and are sample means, and *N* and *M* are sample sizes. Test statistic, has Student's *t* distribution with *N* + *M*– 2 degrees of freedom under null hypothesis.

**F-Test for Two-Sample Assuming Equal Variances**

Probability of variances in X and Y are not significantly different determined by F-test. This function is to know two samples whether have different variances. Table 2.19 presents test results of Paired t-test and Two Sample F-Test for variances (Assuming equal variance for Actual Vs Forecasted Wind Speed for 6models. It is found that, Pearson correlation for Wavelet-ARIMA model is 0.896 for Bagalkot, 0.846 for Vijayapur and 0.818 for Bengaluru site. Two Sample F-Test for variances for actual vs forecasted values assuming equal variance are given in Table 2.20. It is observed that, multiple R2, R2 and adjusted R2 are higher for Wavelet-ARIMA model. Further standard error of ARIMA model for Bagalkot is 0.941, 0.925 for Vijayapur and 0.943 for Bengaluru site. The results reveal that Wavelet-ARIMA model outperformed LM, HWES, ARIMA and GARCH models. Regression plots for residual, fit and Normal probability are given in Fig. 2.65 to 3.70. Residual vs forecasted values is presented in sub-figure (a). Sub-fig (b) represent line fit plot for forecasted wind speed values. Normal probability plot for forecasted wind speed values is given in sub-fig. (c). Results reveal that residual, fit and probability of Wavelet-ARIMA model are uniform and consistent. Lower range of residuals for Wavelet-ARIMA model shows superiority of this model.

**Table 2.19: Paired t-test and Two Sample F-Test for variances assuming equal variance for Actual Vs Forecasted Wind Speed for 6models**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Site** | **Method** | **Pearson Correlation** | **t-stat** | **t-critical one tail** | **t-critical two tail** | **F Critical one-tail** | **P Critical one-tail** | **F critical** |
| Bagalkot | LMS | 0.810 | -52.602 | 1.339 | 1.294 | 0.604 | 0.001 | 0.629 |
| Holt Winter | 0.701 | -94.246 | 1.479 | 1.594 | 0.831 | 0.116 | 0.701 |
| Analytical | 0.794 | -46.518 | 1.312 | 1.243 | 0.440 | 0.008 | 0.617 |
| Iterative | 0.736 | -71.138 | 1.222 | 1.078 | 0.540 | 0.001 | 0.574 |
| GARCH | 0.754 | -77.506 | 1.248 | 1.124 | 0.751 | 0.033 | 0.586 |
| Wavelet ARIMA | **0.896** | -70.220 | **1.082** | **1.105** | 0.609 | 0.001 | 0.696 |
| Vijayapura | LMS | 0.585 | -54.200 | 0.967 | 0.675 | 0.436 | 0.001 | 0.454 |
| Holt Winter | 0.708 | -116.100 | 1.336 | 1.288 | 0.815 | 0.094 | 0.628 |
| Analytical | 0.736 | -72.540 | 1.217 | 1.068 | 0.640 | 0.002 | 0.572 |
| Iterative | 0.755 | -75.640 | 1.297 | 1.124 | 0.570 | 0.001 | 0.609 |
| GARCH | 0.713 | -77.510 | 1.266 | 1.074 | 0.751 | 0.033 | 0.595 |
| Wavelet ARIMA | **0.846** | -81.720 | **0.903** | **0.589** | 0.234 | 0.001 | 0.424 |
| Bengaluru | LMS | 0.676 | -110.50 | 1.448 | 1.901 | 0.472 | 0.001 | 0.680 |
| Holt Winter | 0.719 | -192.900 | 1.321 | 1.26 | 1.184 | 0.138 | 0.621 |
| Analytical | 0.736 | -72.540 | 1.217 | 1.068 | 0.634 | 0.002 | 0.572 |
| Iterative | 0.788 | -52.070 | 1.303 | 1.224 | 0.556 | 0.001 | 0.612 |
| GARCH | 0.787 | -70.020 | 1.249 | 1.314 | 0.622 | 0.001 | 0.587 |
| Wavelet ARIMA | **0.818** | 86.200 | **1.053** | **1.021** | 0.621 | 0.001 | 0.636 |

**Regression Test between Actual and Forecasted Wind Speed Values**

Two sample F-test for variances assuming equal variance for actual vs forecasted wind speed is obtained in Table 2.20. It is found that R2values for Wavelet-ARIMA model is in range of 0.818 to 0.896. Similarly Standard Error SE is between 0.925 to 0.941. Results reveal that Wavelet-ARIMA model has outperformed LMS, ARIMA and GARCH method.

**Table 2.20: Two Sample F-Test assuming equal variance for actual vs forecasted values**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Site** | **Method** | **Multiple**  **R2** | **R2** | **Adjusted**  **R2** | **SE** | **SS Regr** | **F-Value** | **t-stat** | **Lower**  **95%** | **Upper\_**  **95%** |
| Bagalkot | LMS | 0.610 | 0.656 | 0.653 | 1.973 | 599.230 | 317.790 | 9.255 | 2.063 | 3.180 |
| Holt Winter | 0.701 | 0.711 | 0.710 | 1.516 | 741.810 | 618.940 | 1.634 | 0.091 | 0.966 |
| Analytical | 0.694 | 0.630 | 0.627 | 1.824 | 575.000 | 283.990 | 10.890 | 2.471 | 3.560 |
| Iterative | 0.739 | 0.546 | 0.543 | 1.392 | 197.690 | 200.850 | 9.352 | 1.634 | 2.577 |
| GARCH | 0.754 | 0.569 | 0.567 | 0.962 | 203.102 | 219.320 | 8.669 | 1.483 | 3.166 |
| **Wavelet ARIMA** | **0.896** | **0.803** | **0.802** | **0.941** | 734.228 | 77.480 | 8.416 | 1.432 | 2.550 |
| Vijayapura | LMS | 0.585 | 0.342 | 0.338 | 1.910 | 104.937 | 104.940 | 14.389 | 3.770 | 2.097 |
| Holt Winter | 0.808 | 0.652 | 0.650 | 1.599 | 200.198 | 313.500 | 3.590 | 0.545 | 1.877 |
| Analytical | 0.736 | 0.541 | 0.538 | 1.693 | 193.130 | 195.850 | 9.380 | 1.648 | 2.523 |
| Iterative | 0.784 | 0.569 | 0.562 | 1.362 | 203.100 | 219.330 | 8.669 | 1.483 | 2.358 |
| GARCH | 0.766 | 0.544 | 0.456 | 0.965 | 222.500 | 204.600 | 8.225 | 1.550 | 2.114 |
| **Wavelet ARIMA** | **0.846** | **0.858** | **0.864** | **0.925** | 19.451 | 70.521 | 23.742 | 3.860 | 4.560 |
| Bengaluru | LMS | 0.676 | 0.656 | 0.653 | 1.974 | 46.179 | 140.250 | 12.622 | 1.583 | 2.170 |
| Holt Winter | 0.799 | 0.788 | 0.736 | 1.668 | 64.570 | 294.730 | 2.161 | 0.034 | 0.794 |
| Analytical | 0.736 | 0.741 | 0.738 | 1.793 | 193.130 | 195.850 | 9.380 | 1.648 | 1.885 |
| Iterative | 0.788 | 0.765 | 0.768 | 1.446 | 566.970 | 271.120 | 8.257 | 1.944 | 2.651 |
| GARCH | 0.755 | 0.766 | 0.760 | 0.996 | 504.880 | 208.550 | 7.225 | 1.785 | 1.956 |
| **Wavelet ARIMA** | **0.818** | **0.829** | **0.837** | **0.943** | 238.740 | 95.650 | 1.343 | 1.343 | 1.562 |

**Regression Test: Residual, Fit Plot and Normal Probability Plot**

**Bagalkot Site**

|  |  |  |  |
| --- | --- | --- | --- |
| LMS |  |  |  |
|  |  |  |  |
| Holt Winter |  |  |  |
|  |  |  |  |
| Analytical ARIMA |  |  |  |
| **Fig. 2.65: Regression test results of Bagalkot Site (a) Residual Plot, (b) Line Fit Plot (c) Normal Probability Plot for LM, HW and Analytical Models** | | | |
| Iterative ARIMA |  |  |  |
|  |  |  |  |
| GARCH |  |  |  |
|  |  |  |  |
| Wavelet ARIMA |  |  |  |
| **Fig. 2.66: Regression test results of Bagalkot Site (a) Residual Plot, (b) Line Fit Plot (c) Normal Probability Plot for Iteravive ARIMA, GARCH and Wavelet ARIMA Models** | | | |

**Vijayapura Site**

|  |  |  |  |
| --- | --- | --- | --- |
| LMS |  |  |  |
|  |  |  |  |
| Holt Winter |  |  |  |
|  |  |  |  |
| Analytical ARIMA |  |  |  |

**Fig. 2.67: Regression test results of Vijayapura Site (a) Residual Plot, (b) Line Fit Plot (c) Normal Probability Plot for LM, HW and Analytical Models**

|  |  |  |  |
| --- | --- | --- | --- |
| Iterative ARIMA |  |  |  |
|  |  |  |  |
| GARCH |  |  |  |
|  |  |  |  |
| Wavelet ARIMA |  |  |  |

**Fig. 2.68: Regression test results of Vijayapura Site (a) Residual Plot, (b) Line Fit Plot (c) Normal Probability Plot for Iteravive ARIMA, GARCH and Wavelet ARIMA Models**

**Bengaluru Site**

|  |  |  |  |
| --- | --- | --- | --- |
| LMS |  |  |  |
|  |  |  |  |
| Holt Winter |  |  |  |
|  |  |  |  |
| Analytical ARIMA |  |  |  |

**Fig. 2.69: Regression test results of Bengaluru Site (a) Residual Plot, (b) Line Fit Plot (c) Normal Probability Plot for LM, HW and Analytical Models**

|  |  |  |  |
| --- | --- | --- | --- |
| Iterative ARIMA |  |  |  |
|  |  |  |  |
| GARCH |  |  |  |
|  |  |  |  |
| Wavelet ARIMA |  |  |  |

**Fig. 2.70: Regression test results of Bengaluru Site (a) Residual Plot, (b) Line Fit Plot (c) Normal Probability Plot for Iteravive ARIMA, GARCH and Wavelet ARIMA Models**

**2.9 Salient Outcomes of Statistical Models**

Variant statistical and TS models are critically investigated and tested for three selected sites Bagalkot, Vijayapura and Bengaluru. Following are key features observed in this work:

* Two persistence models namely Least Mean Square Algorithm and Holt Winters Smoothing methods are investigated. Holt’s Winter smoothing method performed well as compared to LMS method.
* Two novel techniques of identifying ARIMA models are developed using analytical and iterative method. Iterative method of identifying models outperformed analytical method.
* Transfer function ARIMA, GARCH and Wavelets-ARIMA are developed with a case study to investigate effect of volume of historical wind speed data on model parameters and on forecasting accuracy.
* Forecasting accuracy decreases with increased horizon as well as decreased data volume and vice versa. It is observed that as stochasticity/ uncertainty of data increases forecasting accuracy decreases.
* Pearson correlation for Wavelet-ARIMA model is 0.896 for Bagalkot, 0.846 for Vijayapur and 0.818 for Bengaluru site.
* Standard error of ARIMA model for Bagalkot is 0.941, 0.925 for Vijayapur and 0.943 for Bengaluru site.
* The results reveal that Wavelet-ARIMA model outperformed LM, HWES, ARIMA and GARCH models. Regression plots for residual, fit and Normal probability are presented. Residual vs forecasted values indicate line fit plot for forecasted wind speed values. Normal probability plot for forecasted wind speed values obtained. Results reveal that residual, fit and probability of Wavelet-ARIMA model are uniform and consistent. Lower range of residuals for Wavelet-ARIMA model shows superiority of this model.
* The t-test, h-test and regression tests are conducted on forecasted wind speed data. Following inferences drawn out from tests.
  + Wavelet ARIMA model has highest Pearson correlation factor, lowest critical one tail and two tail values indicating superiority of this model compared to other five models.
  + Model with highest regression coefficient has highest accuracy.
  + Forecasting accuracy is directly a function of error distribution found from regression tests.
  + Significant improvement in wind speed prediction is found from Wavelet-ARIMA model.

Forecasted results proved that combination of Wavelet and ARIMA model resulted in smoother forecast curves with lesser MAPE value. Test results reveal that Wavelet-ARIMA model is consistent and can be used for various patterns.