



# **iJRASET**

International Journal For Research in  
Applied Science and Engineering Technology



---

# **INTERNATIONAL JOURNAL FOR RESEARCH**

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 10    Issue: VI    Month of publication: June 2022**

**DOI: <https://doi.org/10.22214/ijraset.2022.44649>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# Forecasting: The Operating Parameters of Industrial Equipment

Atul Abhishek

B.Tech, Petroleum Engineer, Hazaribagh, Jharkhand-825301

**Abstract:** There are wide variety of equipment used in the production and operation phase in fields like chemical, energy, manufacturing, power etc. All sectors want to reduce their CAPEX (capital expenditure) and OPEX (operational expenditure) to minimize the overall cost. Companies are aiming to decrease the downtime due to breakdown and untimely maintenance of the parts of these equipment and vessels, but still industries are oriented towards primitive preventive maintenance practises which is effective but still poses threat of failure. It is high time industries develop and adopt predictive maintenance approach using analytics and forecasting using their historical data.

We will be focusing on the forecasting option as it gives an idea of the possible future values based on the historical values. This technique is used by large retailers and warehouse companies to forecast sales and demand of their products, so that they can keep their inventory and assets updated accordingly. We will look at Quantitative forecasting method, which is driven by data, follows a pattern which repeats, has no bias, and captures complex patterns. The historical data on which the forecasting is to be done should be long, relevant, and accurate.

**Keywords:** Predictive Maintenance, Holt's Winter, ARMA, ARIMA, SARIMA, SARIMAX, Autocorrelation, Stationary

## I. INTRODUCTION

Forecasting is a traditional method mostly used to predict the demand, sales, growth of parameters or products. Although its not widely applied for forecasting the operational parameters of equipment. But as the values of parameters show trends, seasonality, cyclicity, a combination of these or all of these, which could make forecasting applicable in this case.

In future a dynamic system could also be made in which data would be collected in real time from SCADA or any other system installed by the company and based on the nature of data the most suitable model will be applied to forecast the parameters. It is a thing of future but would still need human supervision. The accuracy of the forecasted values and the real values would be checked with multiple test cases, and after that a suitable conclusion could be drawn whether to accept or reject this method. Just like business segment it might become an essential part of operations and engineering space.

## II. LITERATURE SURVEY

There are different types of components associated with time series-level, trend, seasonality, cyclicity, noise [1].

- 1) **Level:** The baseline of the series on which different components could be added
- 2) **Trend:** It gives an indication of increase or decrease of a series in long run

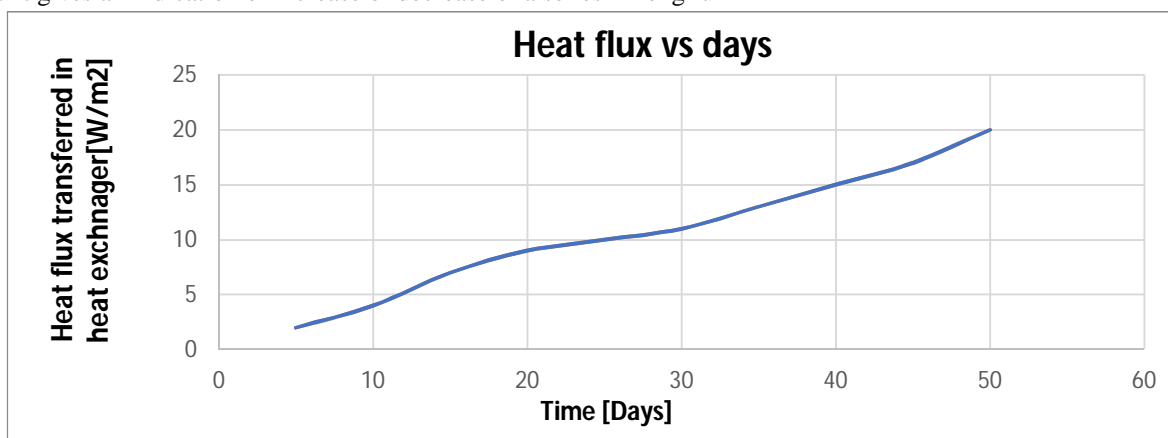


Fig. 1 Heat flux transferred in heat exchanger on daily basis

3) *Seasonality*: It is a pattern which repeats itself in a period

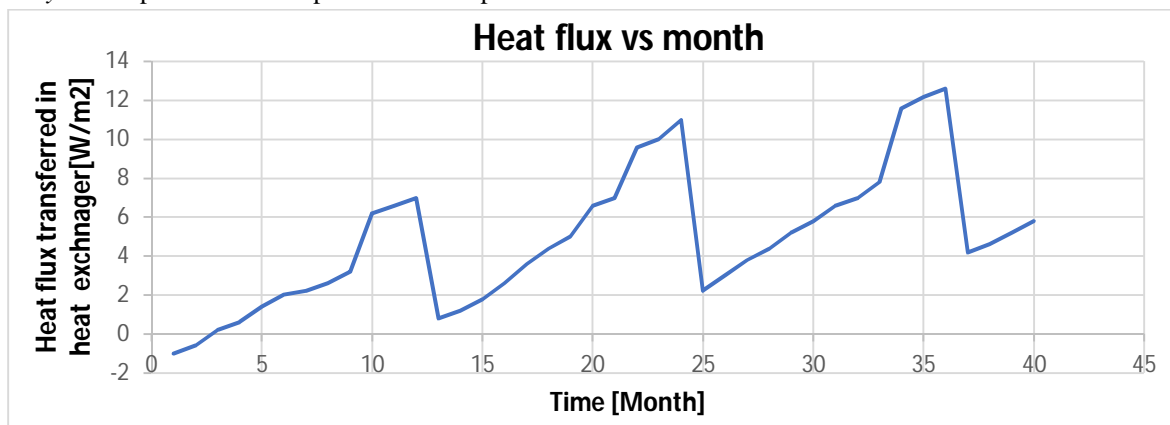


Fig. 2 Heat flux transferred in heat exchanger on monthly basis

4) *Cyclicality*: It is also a pattern but repeats periodically

5) *Noise*: It is random fluctuations in the time series

There are various methods used for forecasting like Naïve method, Simple Moving Average, Exponential Smoothing [2], Holt Winter's Smoothing, Auto Regressive model, and their family.

Naïve method is used when number of observations are 10 or below and has no noise.

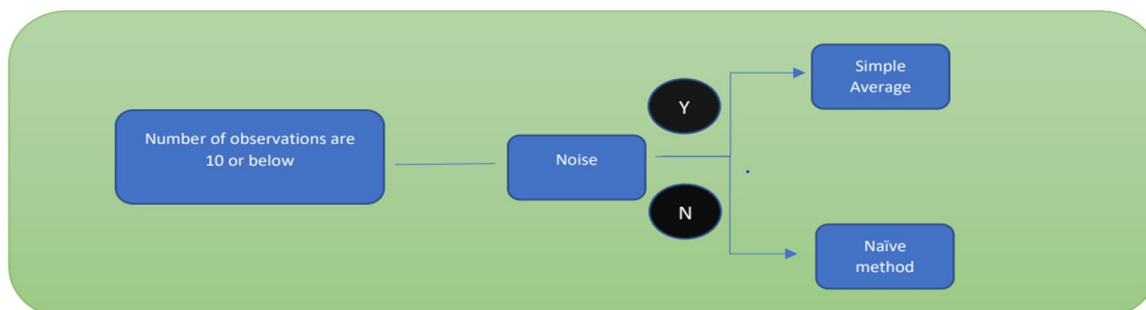


Fig. 3 Forecasting methods for observations less than 10

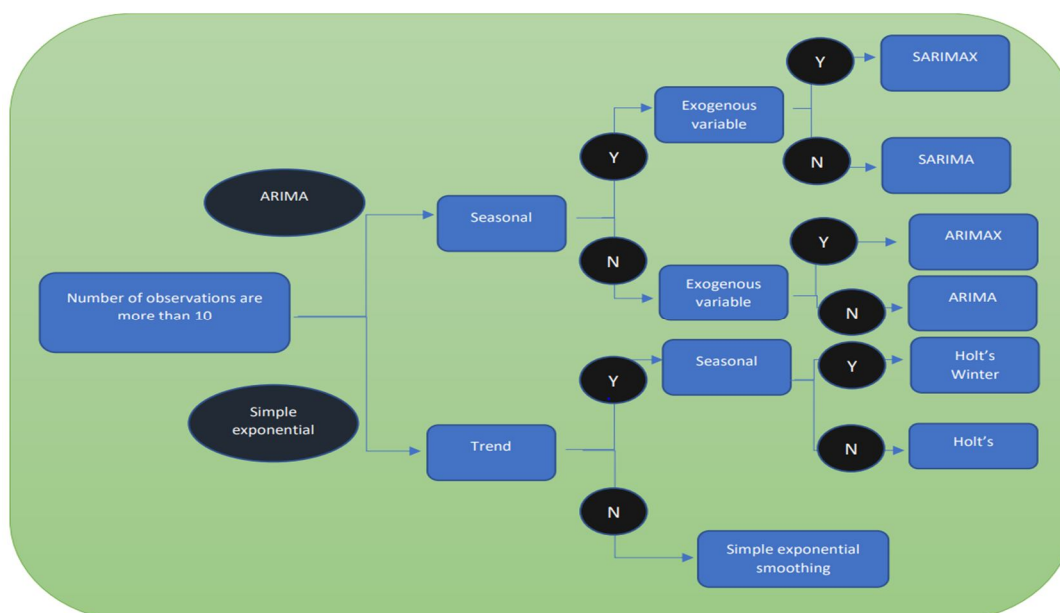


Fig. 4 Forecasting methods for observations more than 10

For forecasting GDP, population of a country for coming years, sales for next week based on previous month, shares value of a company for next couple of days, Naïve method and Simple Average are the best option as they work well when observations are less, and they give weightage to the last reading for forecasting values

We will see two most popular forecasting methods – Holt's Winter and Autoregressive family of methods (ARMA, ARIMA, SARIMA, SARIMAX).

- 6) *Holt's Winter Smoothing [3]*: The Holt-Winters' exponential smoothing technique forecasts based on the level, trend, and seasonality of a time series.
- 7) *Holt' Winter Additive Method*: This method is used when the difference between the crest and trough of seasonality component remains same even though the trend increases or decreases.
- 8) *Holt' Winter Multiplicative Method*: This method is used when the difference between the crest and trough of seasonality component changes with change in trend.

Equation with seasonal components, including level and trend i.e.

$$y_{t+1} = l_t + b_t + S_{t+1-m}$$

$l_t$  : level component

$b_t$  : trend component

The seasonal component is calculated using the following equation:

$$S_t = \gamma \cdot (y_t - l_{t-1} - b_{t-1}) + (1 - \gamma) \cdot s_{t-m}$$

$m$  is the number of times a season repeats during a period

The weight of  $\gamma$  is assigned to this component, whereas  $1 - \gamma$  is the weight assigned to the seasonal component of the recent observations.

The trend and the level equations respectively are as follow:

$$\hat{b}_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}$$

$$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1})$$

The weight of  $\beta$  is assigned to this component, whereas  $1 - \beta$  is the weight assigned to the trend component of the recent observations. The weight of  $\alpha$  is assigned to this component, whereas  $1 - \alpha$  is the weight assigned to the level component of the recent observations. This method is having one of the lowest **MAPE** and **RMSE** and considers level, trend and seasonality component which makes this method very handy and effective.

In the autoregression methods correlation is established between the past values of the time series, like linear regression. Only difference between AR and linear regression is that in addition to the past values of time series, we are also using the correlation of past values with the current values. Without getting much into the details we will see few important topics before building an AR model.

### III. PROCESS AND RESULTS

- Stationarity
- Autocorrelation

For a **stationary series**, statistical property such as **variance**, **mean** and **covariance** should be constant irrespective of the time of observation. Stationarity properties are easier to model and analyse.

There are series like **white noise** and **random walk** which are stationary and non-stationary respectively as white noises are uncorrelated observations with no trend, seasonal or cyclic component and random walks are time series where current observation is sum of previous observation and a random change.



1) *Stationarity Tests* : There are 2 types of tests to check stationarity

TABLE I [4]  
STATIONARITY TEST

	Augmented Dickey-Fuller (ADF) Test	Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test
Null Hypothesis (H0)	The series is not stationary	The series is stationary
Alternate Hypothesis (H1)	The series is stationary	The series is not stationary
P-value > 0.05	Not Stationary	Stationary
P-value < 0.05	Stationary	Not Stationary

ADF

P-value > 0.05: fail to reject null hypothesis, P-value <= 0.05: reject null hypothesis

KPSS

P-value <= 0.05: reject null hypothesis, **P-value > 0.05**: fail to reject null hypothesis

```
KPSS Statistic: 0.148564
Critical Values @ 0.05: 0.46
p-value: 0.100000
```

Fig. 5 Stationary test results from forecasting model

This is how the test results are displayed and as p-value is more than 0.05 for KPSS test, the time series would be stationary and we don't need to go for treatments like Box-Cox Transformation [5] and Differencing, which we will discuss now

In case of non-stationary series, we need to use some tools to convert non-stationary points to stationary points (Not getting into the mathematical part)

2) *Box-Cox Transformation*: It makes variance, and covariance constant

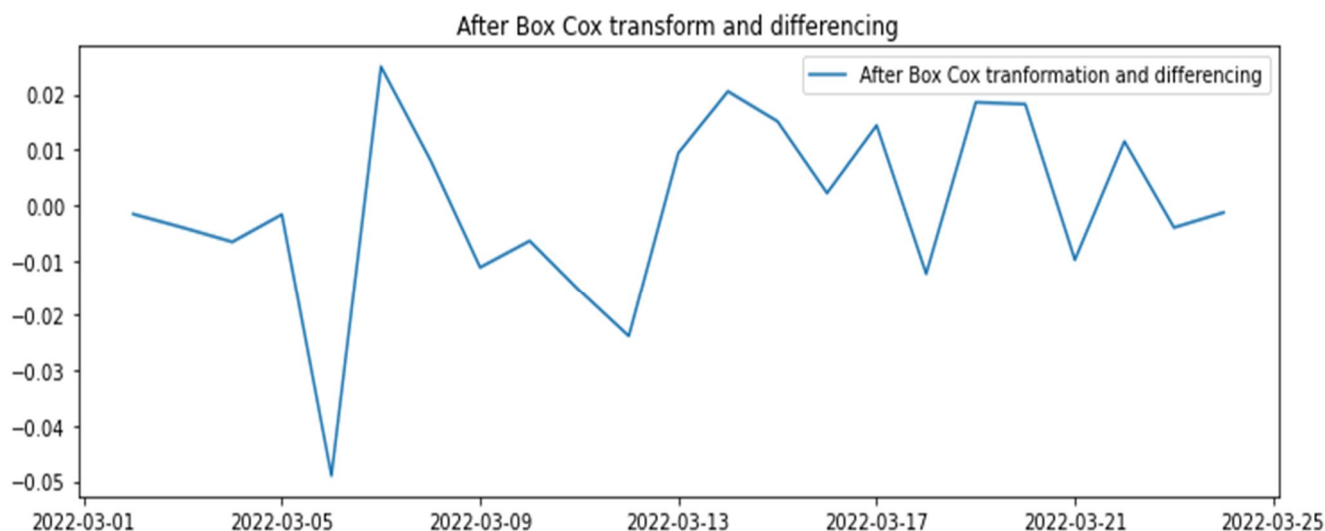


Fig. 6 Dataset after applying Box Cox transformation and differencing

- 3) *Differencing*: It makes the series linear by reducing the trend and seasonality. We get value of **d**, which is the number of times difference operation was done to make the series linear.

Autocorrelation helps in establishing relationship between the observed value with its lagged value.

- a) *ACF (Autocorrelation Function) [6]*: It captures both direct and indirect relationships between the variables.

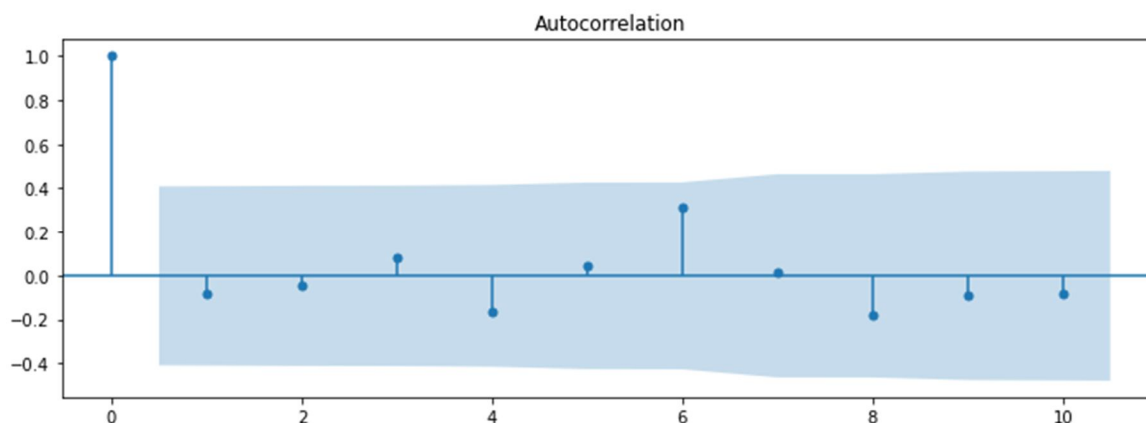


Fig. 7 ACF shows that correlation is relevant after 0 lag

- b) *PACF (Partial Autocorrelation Function [7])*: To capture the direct relationship Partial Autocorrelation is used

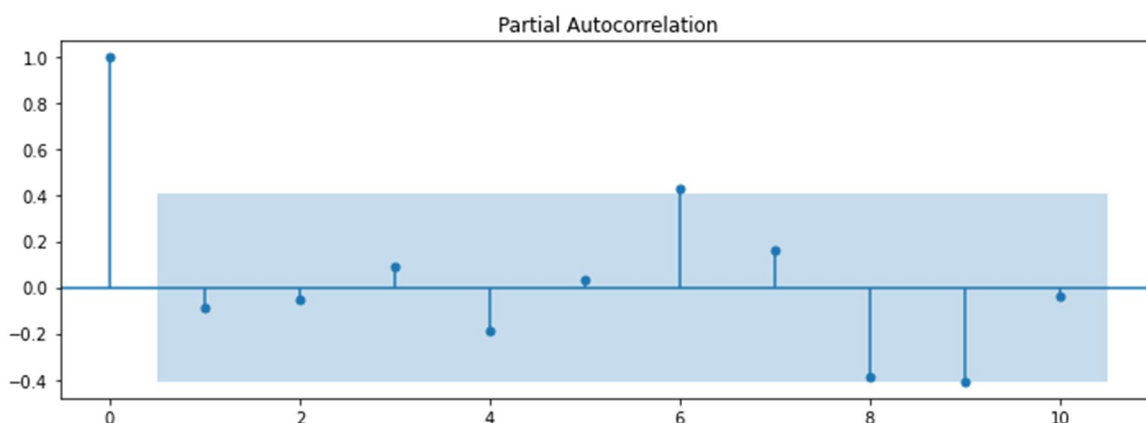


Fig. 8 PACF shows that correlation is relevant after 0 lag

- For both ACF and PACF we can see that the correlation between the observed and previous values is only relevant for lag 0. So, value of  $q=0$  and  $p=0$  in above case.
- If the line is above or below the shaded region (significant level), then only the correlation is relevant for that lag.
- We will get the value of  $p$  and  $q$  based on PACF and ACF lag values, respectively which would be used while modelling AR family of models.
- AR (Auto Regressive) ( $p$ ) [8] forecasts the value using linear regression of one or more previous values.
- MA (Moving Average) ( $q$ ) [9] forecasts the future value as a linear regression of past forecast errors.

- c) *ARMA (Auto Regressive Moving Average) [10]*:

- It captures only trend (non-seasonal component)
- It consists of both **AR(p)** and **MA(q)**
- Values are forecasted as a linear regression of past observations and past forecast errors.
- $p$  and  $q$  values can be found from PACF and ACF.

d) *ARIMA (Auto Regressive Integrated Moving Average) [11]:*

- It also captures only trend (non-seasonal component)
- New parameter 'I' stands for integrated
- The original series is differenced to make it stationary by removing trend
- Later trend is integrated to the original series
- Apart from p and q, d is new parameter which is degree of differencing to makes series linear
- 

e) *SARIMA (Seasonal Auto Regressive Integrated Moving Average):*

- Apart from trend (non-seasonal components) it captures seasonal components too
- p, q, d elements are for trend and P, Q, D for seasonality
- m is also a new element which is the number of time steps for a single seasonal period.

f) *SARIMAX (Seasonal Auto Regressive Integrated Moving Average)*

- Same as SARIMA but here an **exogenous variable** can be added which is an external variable and helps in modelling future values
- Exogenous variable increases the accuracy

E.g., of exogenous variable: If sales of a clothing brand for a year with trend and seasonality is given, but for some months sales was high and it was found that company carried out advertisement campaigns in those or nearby months. So, the data of months when advertisement campaign was carried out be used to forecast the future values

In the case of static equipment, we can use the data of months when any type of maintenance activity has been carried out and use it as an exogenous variable for forecasting the parameters for future.

#### IV. CONCLUSION

In current times a lot of equipment have their operating parameters saved in SCADA or another database. This data can be used for making the forecasting to prevent possible cases of failures. This could emerge as a pivotal method in reducing the OPEX by planning the time of maintenance without shutting down the operations. The coding has been done in Python using Machine Learning and could be automated to make a dynamic model which forecasts in real time. This would maintain integrity of the data as no changes could be made. Results could be displayed in form of dashboards so that it is understood by everyone. Forecasting sales, products, demand etc has been taking place from some time now but forecasting operating conditions of equipment is somewhat new and could be a bit tricky. Nevertheless, companies should try it once and see for themselves if it helped them in saving any revenue or not.

#### V. ACKNOWLEDGEMENT

Atul Abhishek is currently a part of the Energy team of a leading Management Consulting firm. Atul has previously worked in the Upstream and Midstream segment of Oil & Gas companies and has experience in field operations as well. He also has a diploma in Advance certification program in Data Science from IIIT, Bangalore, and has prepared forecasting models on Python using Machine Learning (ML). He has right blend of knowledge when it comes to equipment, business, and ML.

#### REFERENCES

- [1] <https://online.stat.psu.edu/stat510/lesson/1/1.1>
- [2] <https://online.stat.psu.edu/stat510/lesson/5/5.2>
- [3] Dasgupta SS, Mahanta P, Roy R, Subramanian G. Forecasting Industry Big Data with Holt Winter's Method from a Perspective of In-Memory Paradigm. On the Move to Meaningful Internet Systems: OTM 2014 Workshops, Lecture Notes in Computer Science, 2014;8842:80-85
- [4] [Time Series: Interpreting ACF and PACF | Kaggle](#)
- [5] Bergmeir, C., R. J. Hyndman, and J. M. Benitez (2016). Bagging Exponential Smoothing Methods using STL Decomposition and Box-Cox Transformation. International Journal of Forecasting 32, 303-312
- [6] <https://online.stat.psu.edu/stat510/lesson/1/1.2>
- [7] <https://online.stat.psu.edu/stat510/lesson/2/2.2>
- [8] <https://online.stat.psu.edu/stat510/lesson/1/1.2>
- [9] <https://online.stat.psu.edu/stat510/lesson/2/2.1>
- [10] Time Series: Interpreting ACF and PACF | Kaggle
- [11] <https://online.stat.psu.edu/stat510/lesson/3>



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)