**The rebuplıc of turkey**

**BAHCESEHIR unıversıty**

**graduate school of natural and applıed scıences**

**computer engıneerıng**

**TIME SERIES anomaly detectıon**

**Master Thesis**

**samet yazak**

**Supervisor: Asst. prof. cemal okan şakar**

**İSTANBUL, 2018**

**The republıc of turkey**

**BAHCESEHIR UNIVERSITY**

**GRADUATE SCHOOL of natural and applıed scıences**

**computer engıneerıng**

Name of the thesis: Time Series Anomaly Detection

Name/Last Name of the Student: Samet Yazak

Date of the Defense of Thesis:

The thesis has been approved by the Graduate School of Natural and Applied Sciences.

Signature

Asst. Prof. Yücel Batu SALMAN

Graduate School Director

I certify that this thesis meets all the requirements as a thesis for the degree of Master of Arts.

Signature

Asst. Prof. Tarkan AYDIN

Program Coordinator

This is to certify that we have read this thesis and we find it fully adequate in scope, quality and content, as a thesis for the degree of Master of Arts.

Examining Committee Members Signature\_\_\_\_

Thesis Supervisor -----------------------------------

Asst. Prof. Cemal Okan ŞAKAR

Thesis Co-supervisor ----------------------------------

Title Name and Surname

Member -----------------------------------

Title Name and Surname

Member -----------------------------------

Title Name and Surname

**ABSTRACT**

TIME Serıes anomaly detectıon

Samet Yazak

Computer Engineering

Thesis Supervisor: Asst. Prof. Cemal Okan Şakar

June 2018, 0

This study deals with anomaly detection for time series data using unsupervised methods. There are many robust techniques that only work well on specific domains. It is challenging to find a complete method for detecting various anomaly types in different datasets. For this reason, this study aims to find such a method (or ensemble of methods).

**Keywords**: Anomaly detection, time series data

**CONTENTS**

[**LIST OF TABLES** vii](#_Toc513324569)

[**LIST OF FIGURES** viii](#_Toc513324570)

[**ABBREVIATIONS** ix](#_Toc513324571)

[**SYMBOLS** x](#_Toc513324572)

[**1.** **INTRODUCTION** 1](#_Toc513324573)

[**1.1** **TIME SERIES PROPERTIES** 1](#_Toc513324574)

[**1.1.1** **Univariate vs Multivariate** 1](#_Toc513324575)

[**1.1.2** **Time Series Characteristics** 1](#_Toc513324576)

[**1.2** **ANOMALY TYPES** 2](#_Toc513324577)

[**1.2.2** **Point Anomaly** 2](#_Toc513324578)

[**1.2.2** **Contextual Anomaly** 3](#_Toc513324582)

[**1.2.3** **Collective Anomaly** 4](#_Toc513324583)

[**1.3** **ANOMALY DETECTION APPLICATIONS** 5](#_Toc513324584)

[**1.3.1** **Intrusion Detection** 5](#_Toc513324585)

[**1.3.2** **Fraud Detection** 5](#_Toc513324586)

[**1.3.3** **Medical Diagnosis** 5](#_Toc513324587)

[**1.3.4** **Flight Safety** 6](#_Toc513324588)

[**2** **LITERATURE** 7](#_Toc513324589)

[**3** **DATASETS AND METHODS** 8](#_Toc513324590)

[**3.1** **DATASETS** 8](#_Toc513324591)

[**3.1.1** **Data Preparation** 8](#_Toc513324592)

[**3.2** **METHODS** 8](#_Toc513324593)

[**3.2.1** **Moving Average** 9](#_Toc513324594)

[**3.2.2** **Auto-Regressive Integrated Moving Average (ARIMA)** 9](#_Toc513324595)

[**3.2.3** **Kalman Filter** 13](#_Toc513324596)

[**3.2.4** **Time Series Decomposition** 16](#_Toc513324597)

[**4** **RESULTS** 19](#_Toc513324598)

[**5** **DISCUSSION** 20](#_Toc513324599)

[**6** **CONCLUSION** 21](#_Toc513324600)

# **LIST OF TABLES**

# **LIST OF FIGURES**

**Figure 1.1: Seasonal time series** 1

**Figure 1.2: Trend in time series** 2

**Figure 1.3: Point Anomaly** 3

**Figure 1.4: Contextual Anomaly** 4

**Figure 1.5: Collective Anomaly** 5

**Figure 3.1: Original Data ADF Test** 11

**Figure 3.2: Differenced Data ADF Test** 11

**Figure 3.3: ACF Plot** 12

**Figure 3.4: PACF Plot** 12

**Figure 3.5: Reference values for uncertainty parameters** 14

**Figure 3.6: Lower covariance uncertainty** 14

**Figure 3.7: Higher environment uncertainty** 15

**Figure 3.8: Lower environment uncertainty** 15

**Figure 3.9: STL Output** 17

**Figure 3.10: Decompose Output** 17

**Figure 3.11: Moving Avarage vs Moving Median** 18

# **ABBREVIATIONS**

Sd: Standard Deviation

CI: Confidence Interval

# **SYMBOLS**

# **INTRODUCTION**

Anomaly is part of a data or event that does not conform to the excepted behavior. Ex; a web site that is normally receives 0 to 10 requests per hour, suddenly gets 1000 requests. This kind of behavior can also be called outlier, discord, exception. For the rest of this experiment, it will be called anomaly for consistency.

Anomaly detection is finding unexpected behavior of data or events (anomalies). It is studied in many years in diverse domains such as security, finance, health care, fraud etc. Comprehensive surveys of anomaly detection techniques are provided by Hodge and Austin (2004), Agmeyang et al (2006) and Chandola et al (2008). Also, many problem specific studies exist such as Aleskorov (1997) on credit card fraud, Axelsson (2000) on intrusion detection, Spence et al (2001) on MRI image anomaly detection, Markov and Singh (2003a, 2003b) on novelty detection, Fujumaki et al (2005) on air craft signal processing for anomaly detection.

Time series anomaly detection is different from traditional techniques because of nature and behavior of sequence data. Techniques for time series data focus on previously seen observations while traditional techniques focus on general behavior. These techniques can be classified into three categories as Kernel Based, Window Based and Markovian (Chandola et al 2009). Chandola et al (2009) provides a survey on time series anomaly detection with variety of techniques in each group.

Motivation of this experiment is to observe the strengths and drawbacks of present time series anomaly detection techniques on 3 different anomaly types and assemble different methods in one technique that will work well on all types on real time. Supervised techniques that based on sliding window algorithm which will be detailed in Section 3.2 will be investigated. Results of used techniques will be provided with recall, f1 score and performance measures.

Rest of this thesis is organized as follows. Current literature will be provided briefly in Section 2. Datasets, pre-processing tasks and techniques will be described in Section 3. Section 4 consists results of experiments. Section 5 discusses strengths and drawbacks of techniques, which technique is best for what type of anomaly and what can be done beyond this experiment to get better results. Section 6 includes closing statements.

## **TIME SERIES PROPERTIES**

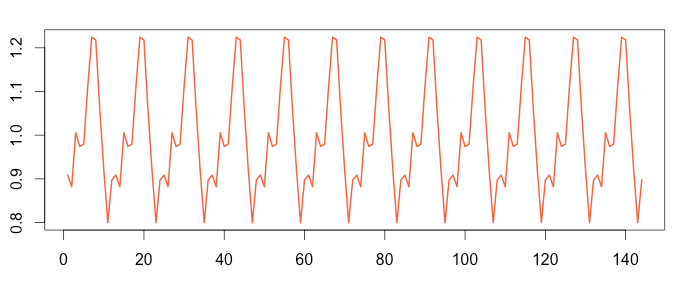
### **Univariate vs Multivariate**

Univariate data consists of single observation that is collected over equal time intervals while multivariate means data consists of many attributes. In this study, univariate time series were used.

### **Time Series Characteristics**

**Seasonality:** Fixed length periodic fluctuations in data. See Figure 1.1.

**Figure 1.1: Seasonal time series**



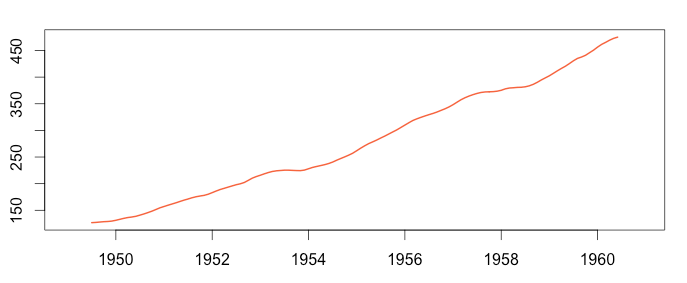
**Cyclic:** Periodic fluctuations with variable length.

**Trend:** Long-term increase/decrease in data.

**Stationary:** Mean and variance does not change over time.

**Residual:** Remainder of time series when season and trend components are removed.

**Figure 1.2: Trend in time series**



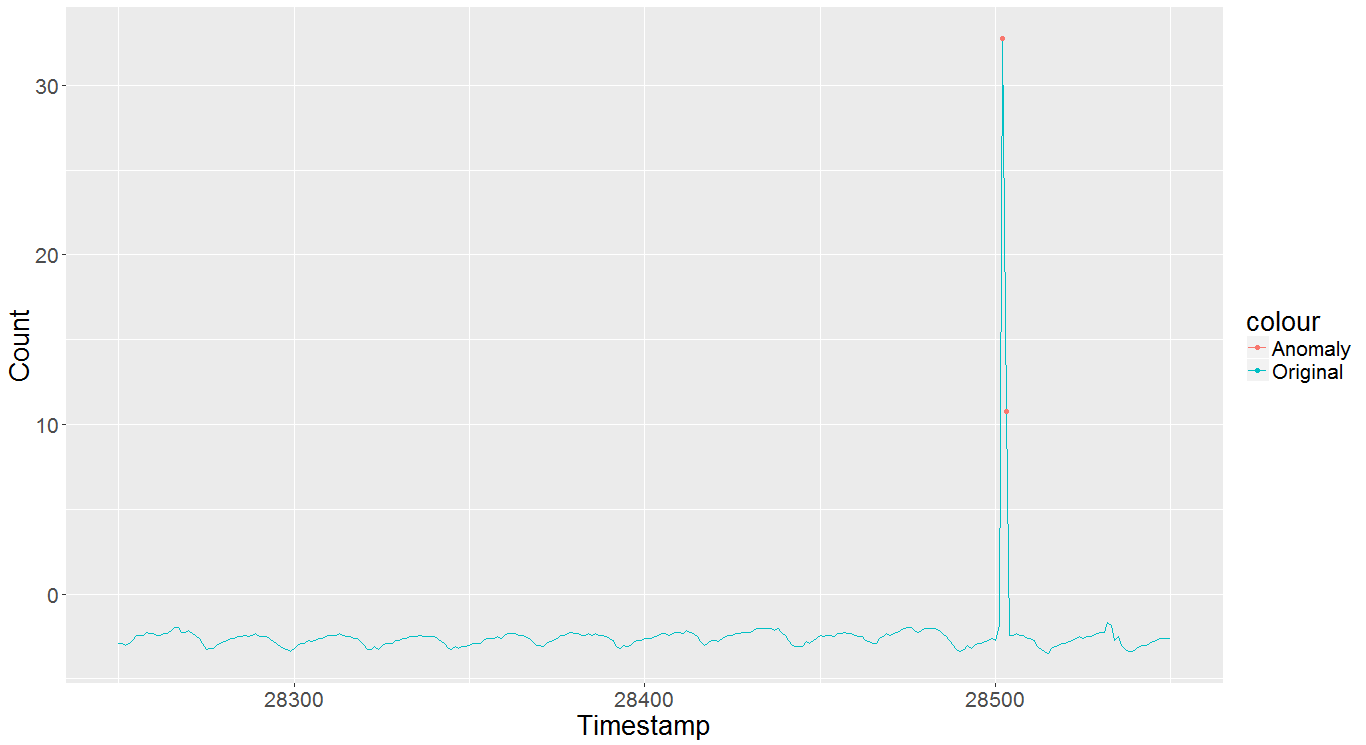
## **ANOMALY TYPES**

It is important to know what kind of anomaly/anomalies that problem includes before determining or proposing a method to sort it out. Because, each type requires different interests and algorithms to detect anomalies. In the literature, anomalies are categorized into three groups (Chandola et al 2009).

### **Point Anomaly**

Individual data instance that can be identified as anomalous is named point anomaly. Figure 1.1 shows an example of this type. If data is considered as web site traffic per hour, red points does not conform with the nature of previous data instances, so they can be identified as anomalous.

**Figure 1.3: Point Anomaly**

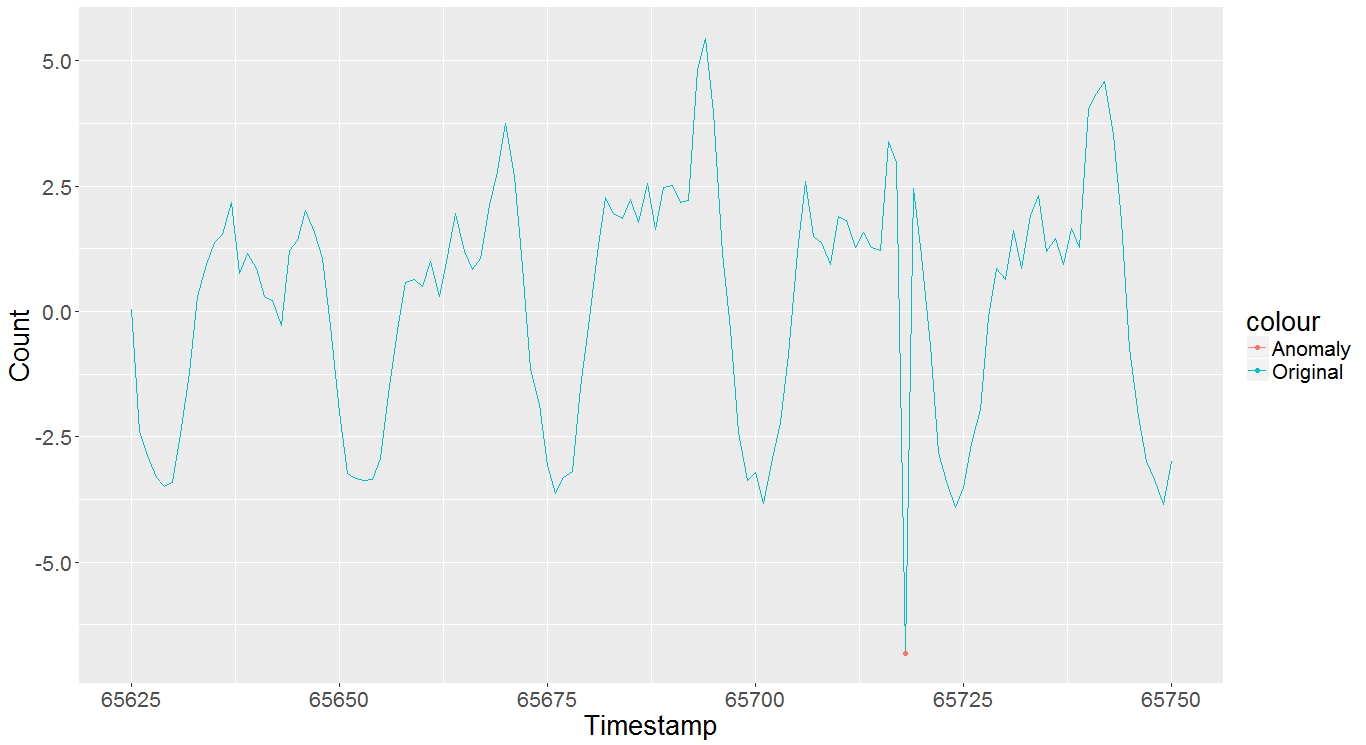




### **Contextual Anomaly**

If data instance is anomalous only in a certain context but not otherwise, then it is a contextual anomaly. Anomaly in Figure 1.2 is a contextual anomaly for breaking periodicity in current context. As in the previous example, if we consider data as web traffic per hour in a web site or purchase of certain product online, this type of anomaly detection can point to a short time of a system failure.

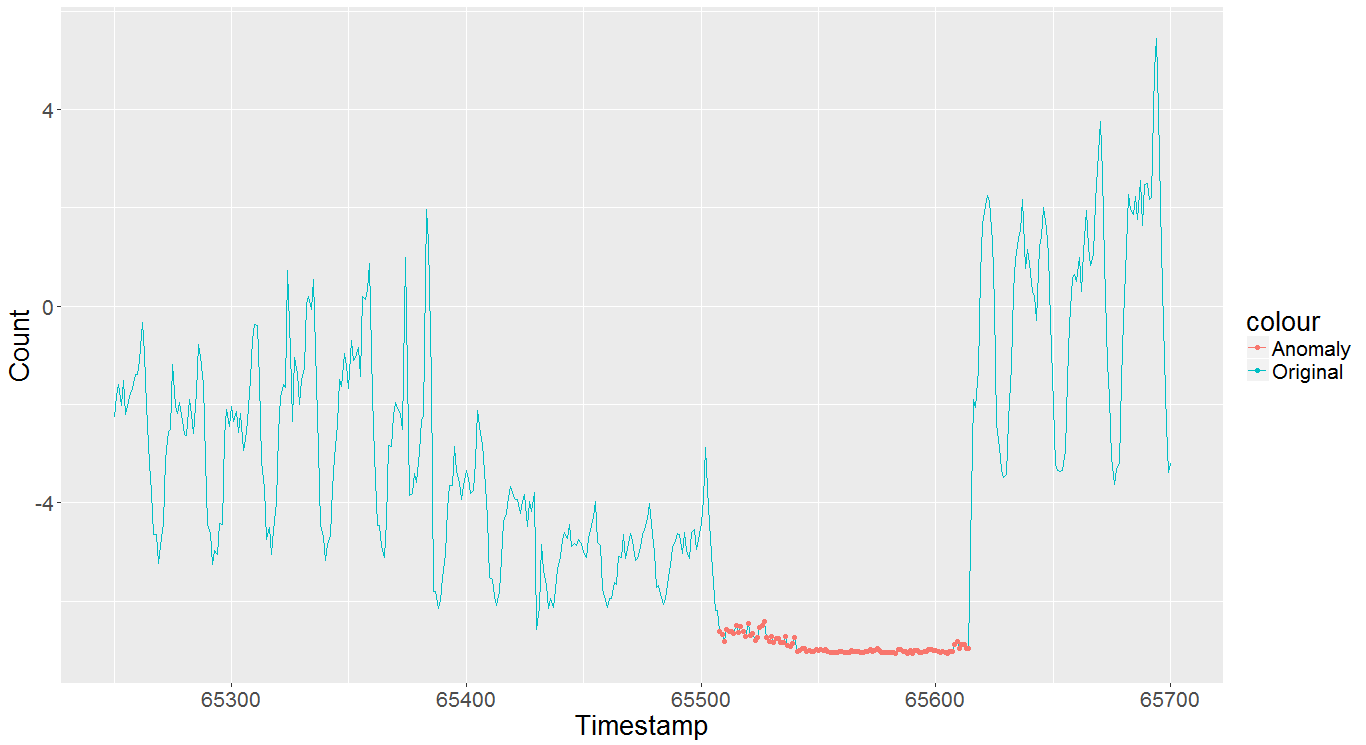
**Figure 1.4: Contextual Anomaly**



### **Collective Anomaly**

If a group of data instances does not conform with respect to the nature of complete data, but they cannot be declared as anomalous individually, then this group of instances are named as collective anomaly. Figure 1.3 shows an example of a collective anomaly. For a considerable amount of time, web site traffic is measured at lowest level (although few data instances can have low values) which can point a serious system failure. Techniques used for detecting collective anomalies are different from point and contextual anomalies and requires detailed examination.

**Figure 1.5: Collective Anomaly**



## **ANOMALY DETECTION APPLICATIONS**

### **Intrusion Detection**

Intrusion detection is a software tool to detect malicious activity or unauthorized access to a computer or network. Since anomaly means different pattern from normal behavior of data, intrusion detection software are good for anomaly detection techniques and they are widely studied (Denning 1987, Axelsson 2000).

### **Fraud Detection**

Fraud detection is a billion-dollar industry that aims to protect customers from unauthorized transactions. There are many applicable domains for fraud like credit cards, banks, stock markets. It generally requires detecting malicious activity in real time to prevent financial loss.

### **Medical Diagnosis**

Medical diagnosis is a very crucial domain and this type of anomaly model needs high accuracy. It typically uses patient history or recent measurements to find irregularities. Electrocardiogram (ECG) records are example of a medical time series data.

### **Flight Safety**

To monitor flight status, sequences of data are collected from aircrafts with various sensors. These sequential data is used to detect faults or unexpected behavior of signals.

# **LITERATURE**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Publication Name** | **Year** | **Writer(s)** | **Abstract** |
| **1** |  |  |  |  |
| **2** |  |  |  |  |
| **3** |  |  |  |  |
| **4** |  |  |  |  |
| **5** |  |  |  |  |
| **6** |  |  |  |  |
| **7** |  |  |  |  |

# **DATASETS AND METHODS**

## **3.1 DATASETS**

In this experiment, a benchmark dataset that is created for Time Series Anomaly Detection by Yahoo is used. Dataset is combination of four parts A1, A2, A3, A4 with each has different characteristics. Two of those sets are used in experiments; A1 and A2. A1 consists real world web request statistics of some Yahoo properties. A2 is a synthetic data which anomalies, trend and seasonality are placed randomly. Both has three attributes:

* Timestamp: hourly representation of time when data is measured
* Value: measurement of data
* Is\_anomaly: indicates whether current instance is anomalous or not, 0 for non-anomalous, 1 for anomalous.

### **3.1.1 Data Preparation**

A1 and A2 datasets are separated into multiple files (67 files for A1, 100 files for A2) and each file has different magnitude of measurement value for instances. So, to merge them into single file with compatible value attributes, a pre-processing method has to applied to standardize data (Nayyar et al 2016). These steps can be listed as:

* Scale each file around mean
* Calculate local minimum and maximum values for each file
* Calculate scale factor that will be applied all files with following formula:

avgMean = mean of local minimums

avgMax = mean of local maximums

scaleFactor = (avgMin + avgMax) / 2

* Multiply each value in all files with scale factor calculated in previous step.
* Merge all files into one.

## **3.2 METHODS**

In this chapter, methods that are applied to datasets during experiment will be explained in detail. Essential part of methods is based on same algorithm which is to calculate a confidence interval that next expected instance will most likely be in, using previous observed instances. These previous observed instances are formed using Sliding Window technique. Sliding Window handles fixed number of instances (window) at a time, the earliest instance is dropped out of window when a new instance is observed. Window size is hyper parameter of this method. Algorithm 3.1 explains calculating confidence interval.

**Algorithm 3.1:**

1. Set an appropriate window size for data, n.
2. Iterate fixed size window though whole date set.

window-1 = {1, 2, 3, 4, …, n - 1, n}

window-2 = {2, 3, 4, …, n, n + 1}

window-3 = {3, 4, …, n + 1, n + 2}

1. Calculate confidence interval (CI) for next instance. (Calculate lower and upper bounds)
2. If next observed instance is inside of CI, it is not anomalous, otherwise, mark it as anomaly.

Remainder of this chapter will provide methods to calculate CI for step 3.

### **3.2.1 Moving Average**

Moving Average is a method that only uses mean and standard deviation of window to calculate CI. It is based on the assumption that next instance will be most likely be similar with the window records. In this domain, it is mean and standard deviation that makes a record coherent with entire window. One can generate many useful formulas that can be effective for calculating CI, here is a sample to express idea better:

|  |  |
| --- | --- |
|  | (3.1) |

where mean is mean, sd is standard deviation of window records, n is window size and m is a hyper parameter to adjust confidence interval.

### **3.2.2 Auto-Regressive Integrated Moving Average (ARIMA)**

ARIMA is a popular and widely used model for time series forecasting and data analysis. Forecasting the next instance can result calculating better confidence intervals, which leads sliding window technique to apply better anomaly test.

ARIMA is a generalization of Auto-Regressive Moving Average (ARMA) model. Fitting an ARIMA model is sometimes called Box-Jenkins Model. It has three key parts, hence it is notated by 3 parameters as ARIMA (p, d, q). There are briefly listed here:

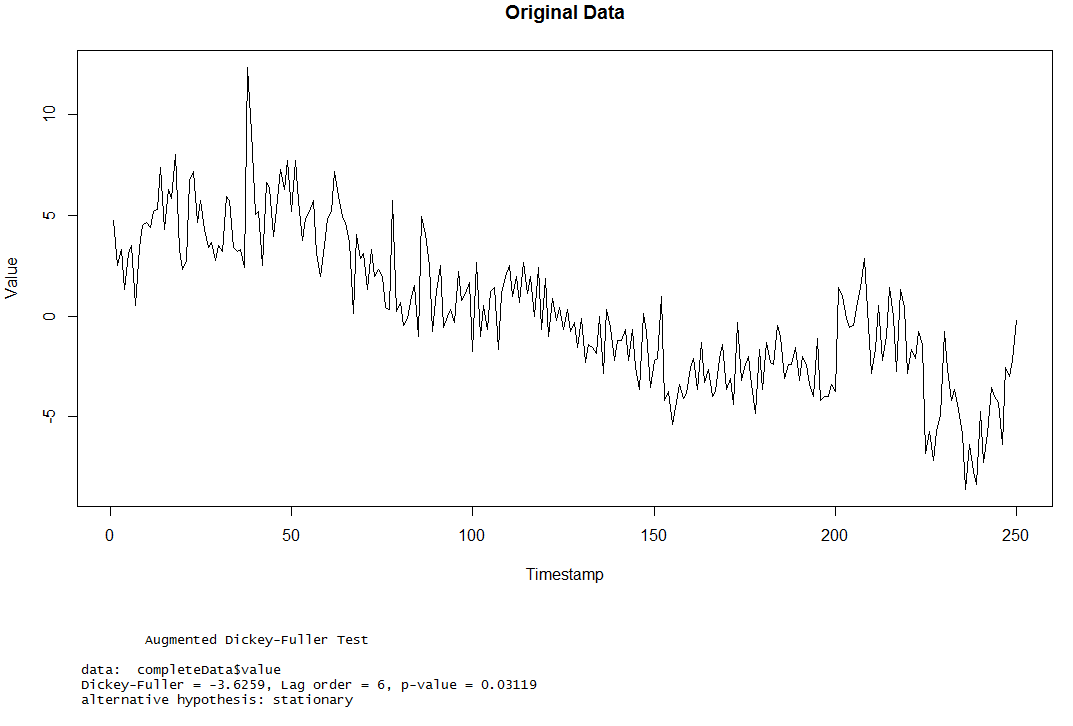
* AR(p): Auto-regression. Number of lags used in the model.
* I(d): Integrated. Degree of differencing between past and current values d times. It is used to make series stationary.
* MA(q): Moving average. MA refers to residual error of model and q is the number of instances in moving window.

In order to make ARIMA model forecast better results, following steps must be applied.

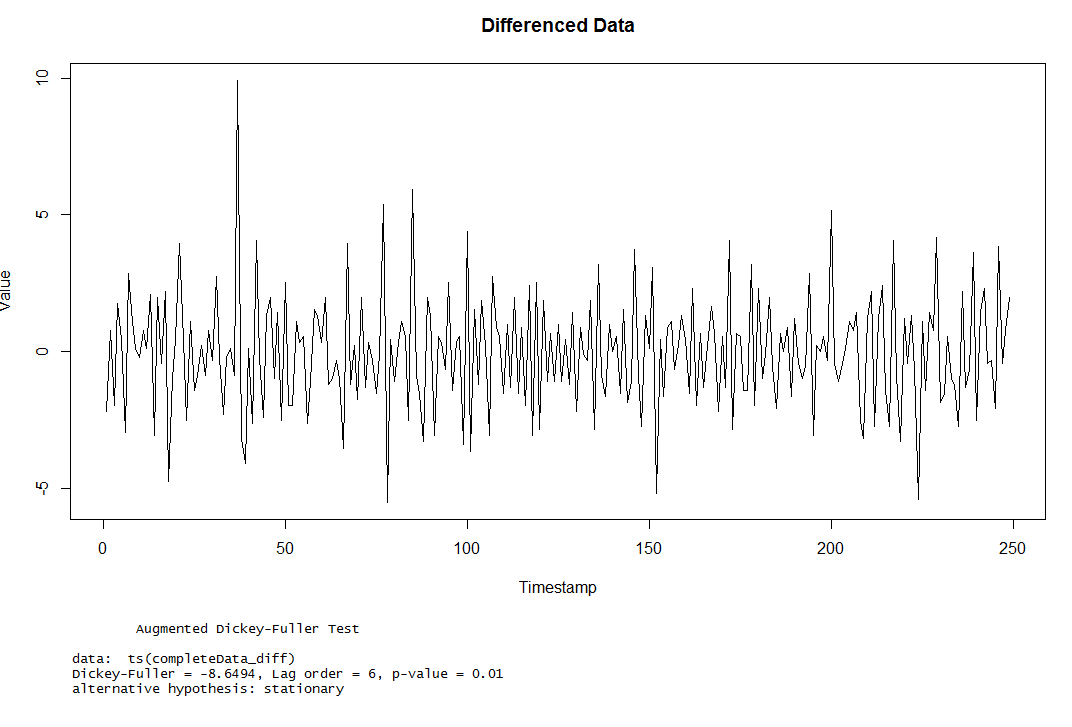
1. Visualize data: To model a successful model, trend, seasonality or random behavior must be determined. Because, ARIMA works best if data is stable and stationary.
2. Make data stationary: Augmented Dickey-Fuller (ADF) test is a popular stationary test for ARIMA models. As a result of the test, if data is non-stationary, differencing which is Integration part of model can be applied to data. Decomposition is also an effective method. It will be detailed in Section 3.2.4.

The lower p-value output is generated from ADF test; the better d parameter is gained for model. To determine d parameter, first, original data is passed to ADF test (Figure 3.1). Then, a differenced version of data with degree 1 is tested. Figure 3.2 shows an ADF test with differenced data and after this test d can be determined as 1.

**Figure 3.1: Original Data ADF Test**

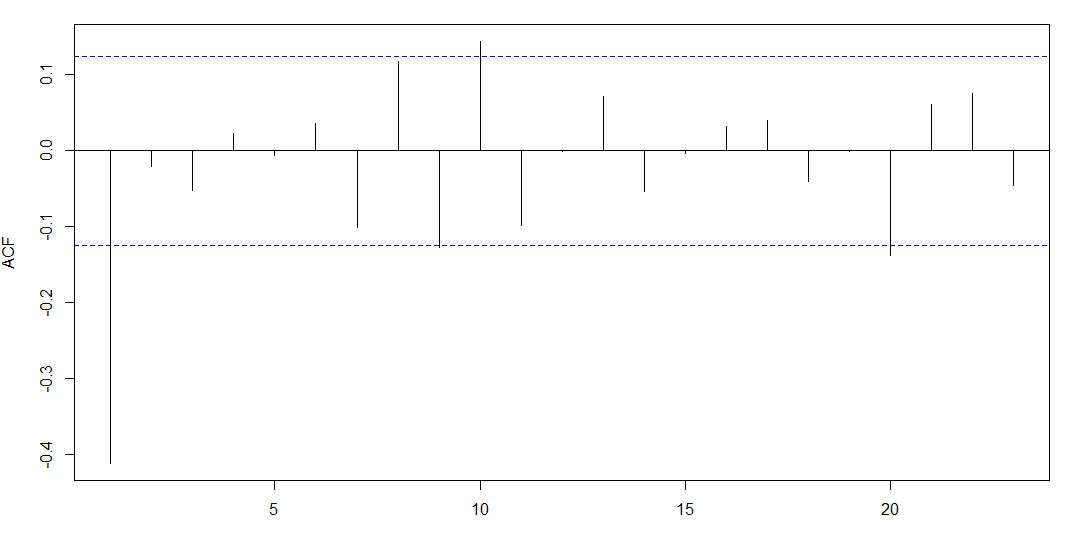


**Figure 3.2: Differenced Data ADF Test**

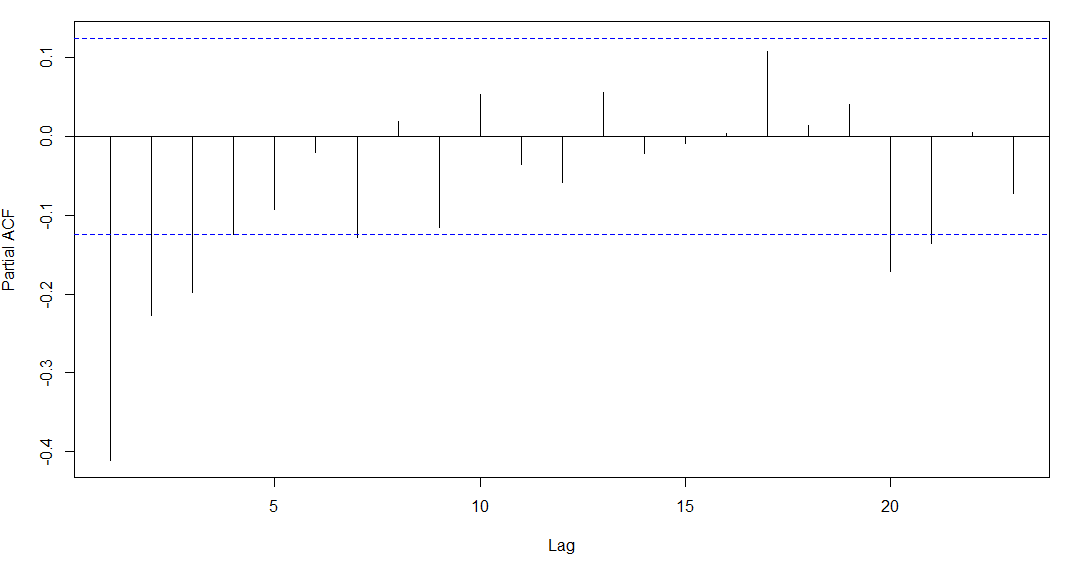


1. Find p and q: Other two essential parameters of model, p and q can be found using ACF (Figure 3.3) and PACF (Figure 3.4) plots. Figures show that both have a peak at Lag 1. This means p and q can be 1.

**Figure 3.3: ACF Plot**



**Figure 3.4: PACF Plot**



1. Build ARIMA model with p, d and q: Detail of building is beyond this experiment, hence a popular online R library “forecast” is used to build a model.

Main goal of building an ARIMA is to calculate a better CI. “Forecast” library also provides parametric CI results. Thus, for this model, probability of CI that next instance will most likely be in is a hyper-parameter. As it is known anomaly is a rare case, probability can be between 90% and 99%.

### **3.2.3 Kalman Filter**

Kalman Filter (Kalman 1960) is an algorithm that uses Gaussian Distribution and previously measured instances to produce estimation of unknown variables. Despite its almost 60-year history, it is still a powerful solution to tracking and data prediction tasks. Typical usage areas of Kalman Filter are global positioning system, guidance of aircrafts, data smoothing, robotic motion control, time series signal processing etc.

In this experiment, Kalman Filter was used to predict next time series instance value and then calculate CI based on it. ARIMA can fail producing good results with volatile time series. It is essential to make data stationary before building ARIMA model. Kalman Filter is an alternative solution to smooth data and make predictions about future. Algorithm 3.2 explains how Kalman Filter smooths time series data:

**Algorithm 3.2**

1. Let be previous observed value, be current value, be covariance uncertainty, be uncertainty of environment like noise, be error.
2. Initialize starting error with 1.
3. Calculate covariance;

|  |  |
| --- | --- |
|  | (3.2) |

1. Calculate gain;

|  |  |
| --- | --- |
|  | (3.3) |

1. Predict next value using previous value and gain;

|  |  |
| --- | --- |
|  | (3.4) |

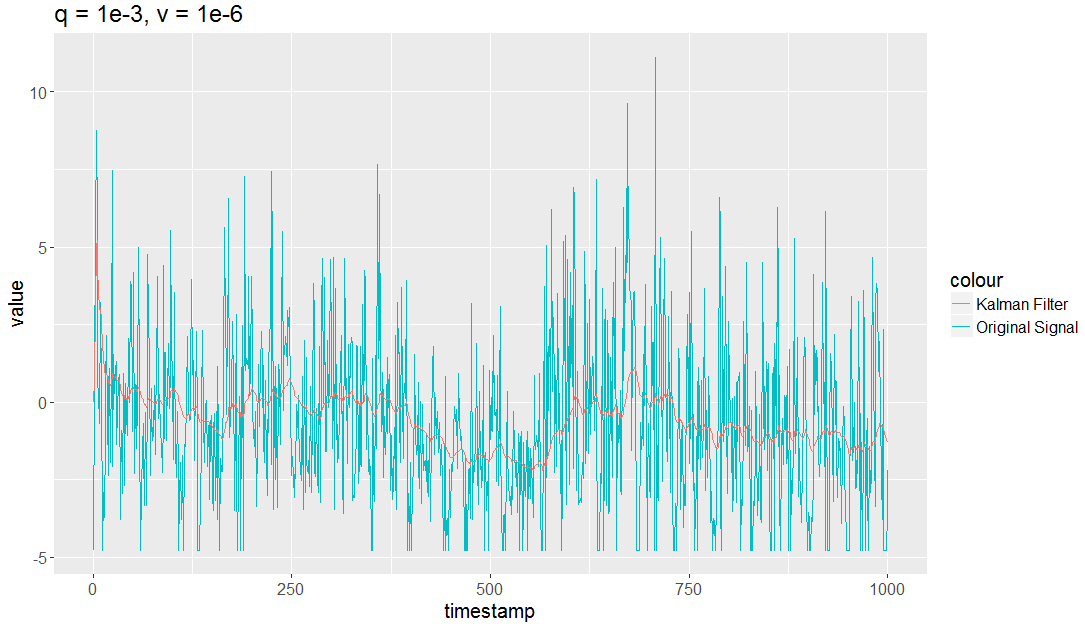
1. Calculate error which will be used in next iteration;

|  |  |
| --- | --- |
|  | (3.5) |

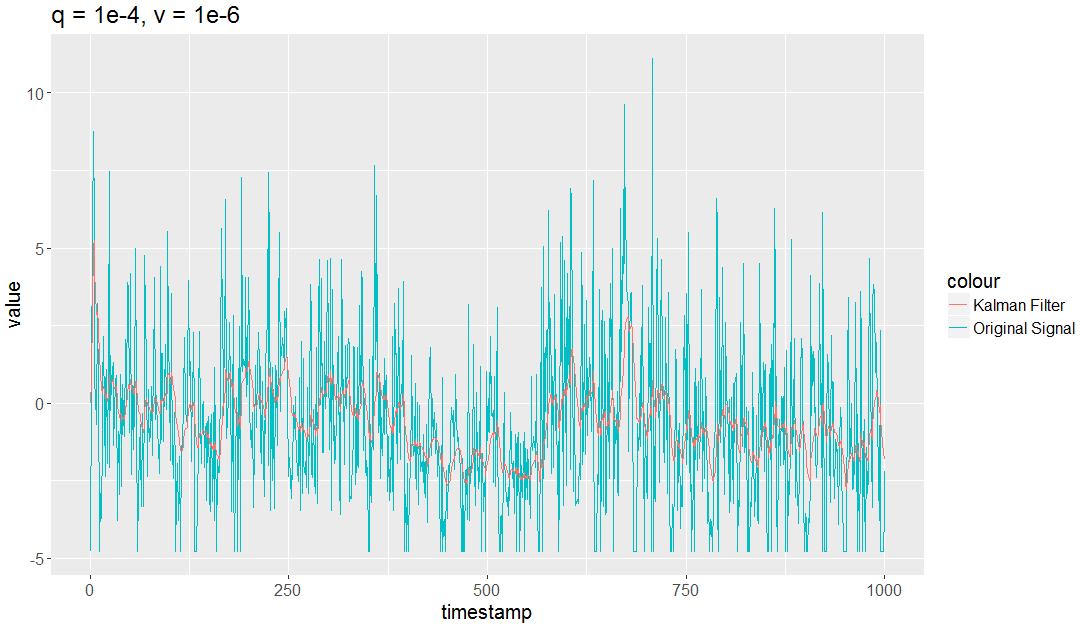
1. Iterate step 3 to 6 until all records are processed.

In algorithm 3.2, and are hyper-parameters. Following images demonstrate effects of different parameters on signal smoothing.

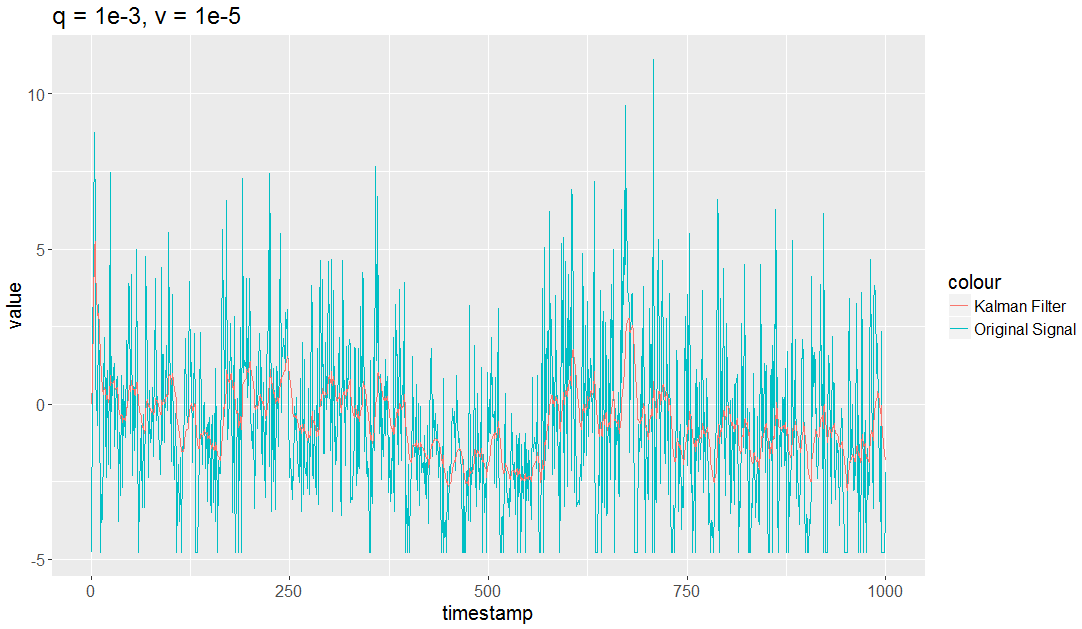
**Figure 3.5: Reference values for uncertainty parameters**



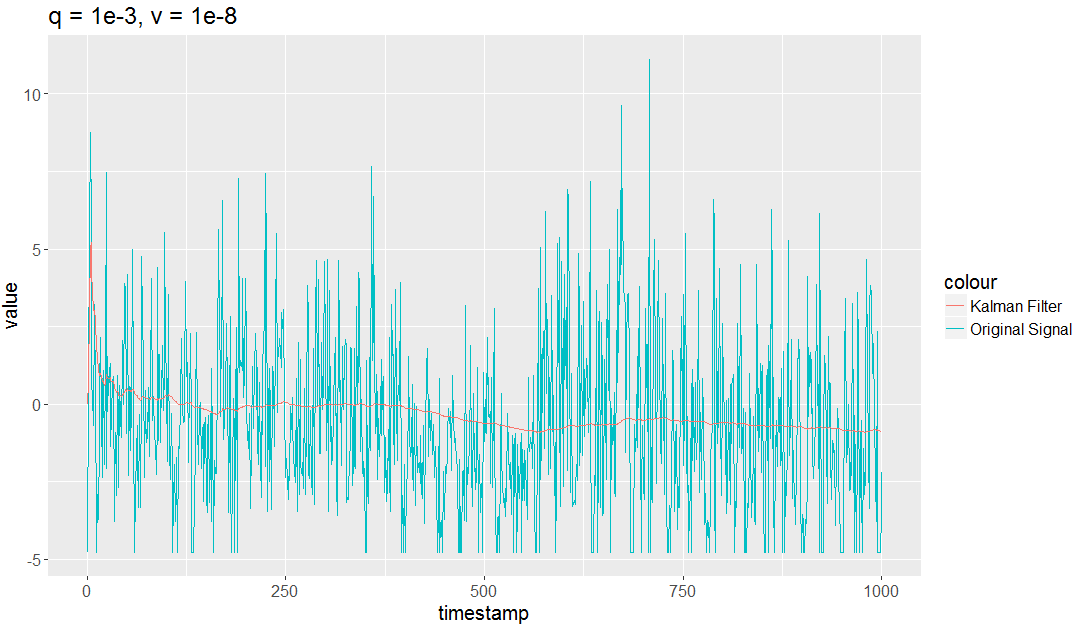
**Figure 3.6: Lower covariance uncertainty**



**Figure 3.7: Higher environment uncertainty**



**Figure 3.8: Lower environment uncertainty**



Difference of Figure 3.5 and Figure 3.6 shows that for a lower covariance uncertainty, more volatile signal is obtained. For environment uncertainty, Figure 3.7 and Figure 3.8 demonstrate that high value gets a volatile signal while low value gets a smother signal.

Once signal is smoothed, CI can be calculated using equation 3.1 and signal properties like mean and standard deviation.

### **3.2.4 Time Series Decomposition**

Time series decomposition is splitting time series data into three parts; seasonality, trend and residual. Residual part is used to detect anomalies.

Residual is remainder of data when seasonality and trend are removed. To remove them, first, one must find seasonal behavior of data and calculate trend. In this experiment 2 different seasonality determination method are used; Discrete Fourier Transform and correlation of time series windows.

**Discrete Fourier Transform:** DFT is one of the most powerful digital signal processing tools. It is basically converting finite sequence of data into same-length of Discrete-Time Fourier Transform (DTFT) sequences. It can provide multiple period lengths for seasonality of data. In this experiment, period which is the smallest amongst numbers that are greater than 24 is used for sliding windows. The reason of number 24 is that data is collected by hourly observation and 24 hours mean a day. A day of periodicity looks reasonable for data.

Drawback of DFT is it always gives a periodicity length result whether data has seasonality component or not. Threating data like it has seasonality results incorrect residual calculation, thus it results false anomaly detections.

**Correlation:** If time series data has seasonality, then successive periods of data should be highly correlated. Based on this definition, if a data window has seasonality, period length can be found using Algorithm 3.3.

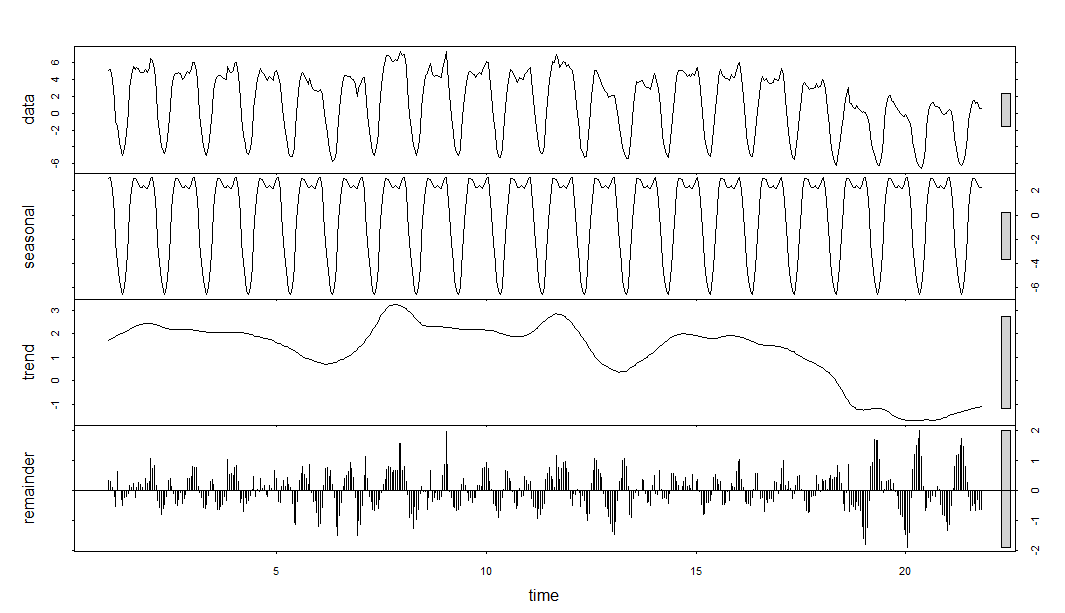
**Algorithm 3.3**

1. Set a possible number of period length (p) values.
2. **S**elect a new p value and split data into p parts.
3. Calculate correlation of all periods with each other.
4. If correlation value is greater than threshold, then p is seasonal length.
5. If correlation value is lower than threshold, repeat Step 2.

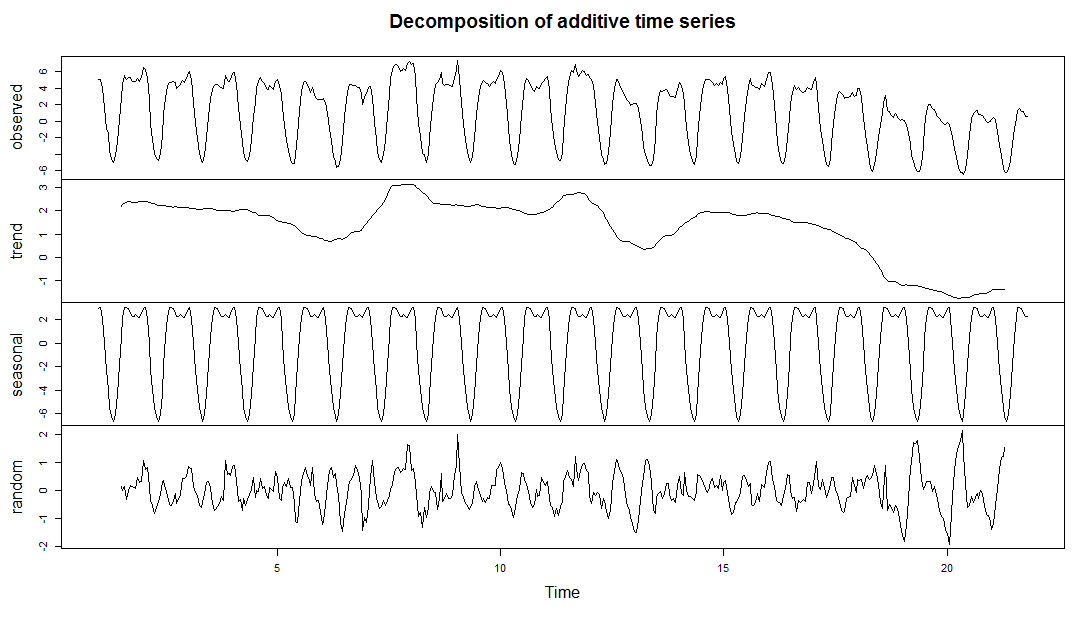
Drawback of Algorithm 3.3 is complexity of calculation which is O(n2) in worst case.

As an alternative of these two methods, built-in R methods stl() and decompose() can be used. But both methods require period length as input. Figure 3.9 and Figure 3.10 show sample outputs of stl() and decompose(). As It can be seen on images, outputs also consist trend and residual part which is important for anomaly detection procedures.

**Figure 3.9: STL Output**

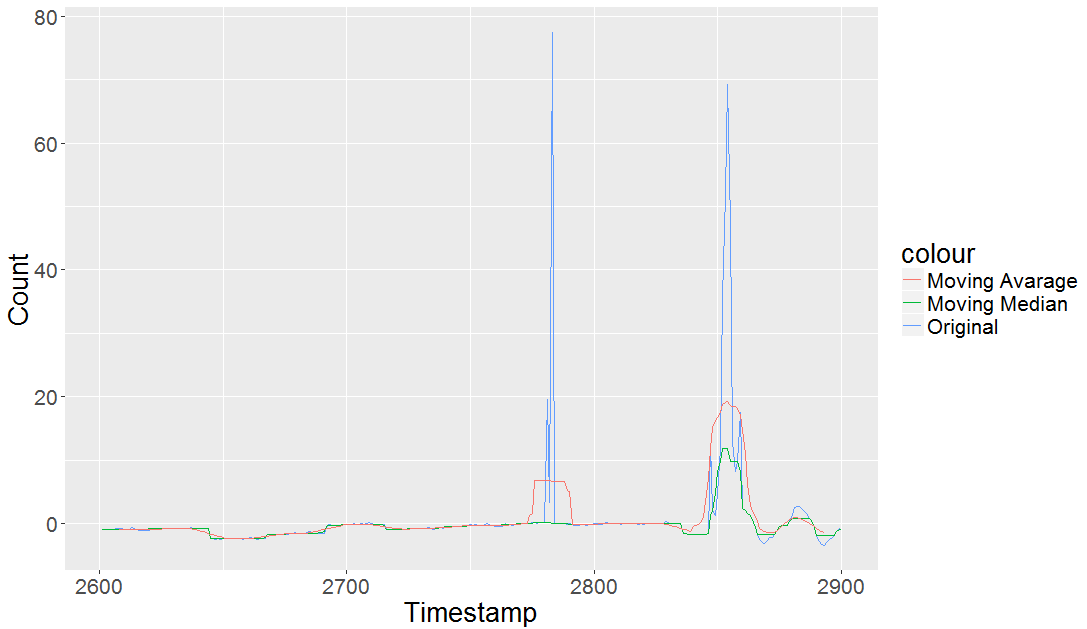


**Figure 3.10: Decompose Output**



Before starting anomaly detection, trend in data must also be calculated. Moving average and moving median both can be used to calculate trend. However, moving median is more robust to anomalies, thus, it fits better for anomaly detection procedures. Figure 3.11 shows the difference between two methods in case of anomaly exists in data. Moving average has higher spikes than moving median in case of anomalies.

**Figure 3.11: Moving Average vs Moving Median**



As it is mentioned before, residual is what is left when seasonality and trend have been removed. Confidence interval (CI) can be calculated based on mean and standard deviation of residual according to Equation 3.1.

# **RESULTS**

# **DISCUSSION**

# **CONCLUSION**

**REFERENCES**

***Books***

Kumar, V., Tan, P., N., Steinbach, M. 2014. *Introduction to data mining.*

Jain, A. K., Dubes. R. C. 1988. *Algorithms for Clustering Data*. Prentice-Hall, Inc.

***Term Publications***

Chandola, V., Banerjee, A., and Kumar, V. 2009. Anomaly detection: A survey. *ACM Comput. Surv.* 41, 3

Chandola, V., Mithal, V., Kumar, V. 2008. A comparative evaluation of anomaly detection techniques for sequence data. *In Proceedings of International Conference* *on Data Mining*.

Denning, D.E., 1987. An Intrusion Detection Model. *Software Engineering, IEEE Transactions,* 222-32

Axelsson, S., 2000. Intrusion Detection Systems: A Survey and Taxonomy

Nayyar, K., Vishwasrao, S., Chakravarty, S., Dabiri, S. 2016. Anomaly Detection for Univariate Time-Series Data. *ICML 2015*.

Kalman, R., E. 1960. A new approach to linear filtering and prediction problems. *J. Basic Eng.* 82, 1.

Ahmed, M., Mahmood, A. 2014. Network traffic analysis based on collective anomaly detection. *IEEE 9th conference on industrial electronic sand applications (ICIEA)*. 1141–46.

Agyemang, M., Barker, K., Alhajj, R. 2006. A comprehensive survey of numeric and symbolic outlier mining techniques. *Intell. Data Anal.* 521–538.

Wei, L., Keogh, W., Xi, X. 2006. Saxually explicit images: Finding unusual shapes. *ICDM ’06. Sixth International Conference*. 711–720.

Fox, A. J. 1972. Outliers in time series. *Journal of the Royal Statistical Society. Series B(Methodological).* 34(3):350–363

Ma, J., Perkins, S. 2003. Time-series novelty detection using one-class support vector machines. 1741–1745.

Hodge, V., Austin, J. 2004. A survey of outlier detection methodologies. *Artif. Intel. Rev*. 22, 2, 85–126.

Aleskerov, E., Freisleben, B., Rao, B. 1997. Cardwatch: A neural network based database mining system for credit card fraud detection. *In Proceedings of the IEEE Conference on Computational Intelligence for Financial Engineering*. 220–226.

Keogh, E., Lin, J., Lee, S.-H., Herle, H. V. 2006. Finding the most unusual time series subsequence: Algorithms and applications. *Knowl. Inform. Syst*. 11, 1, 1–27.

Keogh, E., Lonardi, S., Chiu, B. Y. 2002. Finding surprising patterns in a time series database in linear time and space. *In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM Press.* 550–556.

Janakiram, D., Reddy, V., Kumar, A. 2006. Outlier detection in wireless sensor networks using Bayesian belief networks. *In Proceedings of the 1st International Conference on Communication System Software and Middleware*. 1–6.

Kumar, V. 2005. Parallel and distributed computing for cybersecurity. *IEEE Distrib. Syst. Online.* 6, 10.

Spence, C., Parra, L., Sajda, P. 2001. Detection, synthesis and compression in mammographic image analysis with a hierarchical image probability model. *In Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis. IEEE Computer Society*.

Fujimaki, R., Yairi, T., Machida, K. 2005. An approach to spacecraft anomaly detection problem using kernel feature space. *In Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*. ACM Press, 401–410.

Teng, H., Chen, K., Lu, S. 1990. Adaptive real-time anomaly detection using inductively generated sequential patterns. *In Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy*. IEEE Computer Society Press, 278–284.

Rousseeuw, P. J. Leroy, A. M. 1987. Robust Regression and Outlier Detection.

Markou, M. Singh, S. 2003a. Novelty detection: A review-part 1: Statistical approaches. *Sig. Proc*. 83, 12, 2481–2497.

Markou, M. Singh, S. 2003b. Novelty detection: A review-part 2: Neural network based approaches. *Sig. Proc*. 83, 12, 2499–2521.

Edgeworth, F. Y. 1887. On discordant observations. *Philosoph. Mag*. 23, 5, 364–375.

Knorn, F., Leith, D. J. 2008. Adaptive kalman filtering for anomaly detection in software appliances. *IEEE Conference on Computer Communications Workshops*. 1–6.

***Other Publication***

Yahoo! Webscope dataset. <http://labs.yahoo.com/Academic_Relations> [accessed 6 November 2017]

Forecast R Package. <https://cran.r-project.org/web/packages/forecast/index.html> [accessed 12 December 2017]

Time Series Components. <https://www.otexts.org/fpp/6/1> [accessed 1 April 2018]

**APPENDICES**