

CHANGE POINT ANALYSIS: A STATISTICAL APPROACH TO DETECT POTENTIAL ABRUPT CHANGE

Siti Nur Afifah Mohd Arif^a, Mohamad Farhan Mohamad Mohsin^b, Azuraliza Abu Bakar^{a*}, Abdul Razak Hamdan^a, Sharifah Mastura Syed Abdullah^c

^aData Mining and Optimization Research Group, Centre for Artificial Intelligence Technology, Faculty of Science & Information Technology, Universiti Kebangsaan Malaysia, 43600 Selangor, Malaysia

^bSchool of Computing, Universiti Utara Malaysia, 06010, Sintok, Kedah, Malaysia

^cInstitute of Climate Change, Universiti Kebangsaan Malaysia, 43600 Selangor, Malaysia

Article history

Received

21 January 2017

Received in revised form

14 April 2017

Accepted

31 May 2017

*Corresponding author
azuraliza@ukm.edu.my

Abstract

Change-point analysis has proven to be an efficient tool in understanding the essential information contained in meteorological data, such as rainfall, ozone level, and carbon dioxide concentration. In this study, change-point analysis was used to discover potential significant changes in the annual means of total rainfall, temperature and relative humidity from 25 years of Malaysian climate data. Two methods, the CUSUM and bootstrap, were used in the analysis, where the CUSUM was used to analyze the data trends and patterns and bootstrapping was used to calculate the occurrence of change points based on the confidence level. The results of the analysis showed that potential abrupt shifts seem to have taken place in 1999, 2001 and 2002 with respect to the annual means for relative humidity, temperature and total rainfall, respectively. These identified change points will be further analyzed as potential candidates of abrupt change by extending the proposed method in a future study.

Keywords: Change-point analysis, abrupt change, CUSUM, bootstrap, climate

Abstrak

Analisis titik perubahan telah menunjukkan keupayaannya sebagai satu alat yang berkesan dalam memahami maklumat-maklumat penting yang terdapat pada data meteorologi seperti hujan, paras ozon, dan kepekatan karbon dioksida. Dalam kajian ini, analisis titik perubahan diimplementasikan untuk mengesan sebarang perubahan yang signifikan pada data iklim Malaysia sepanjang tempoh 25 tahun untuk purata tahunan jumlah hujan, suhu, dan kelembapan relatif. Dua teknik diaplikasikan dalam analisis ini iaitu CUSUM dan bootstrap, di mana teknik CUSUM digunakan untuk menganalisis tren dan corak data manakala teknik bootstrap adalah untuk mengira tahap keyakinan sesebuah titik perubahan. Hasil kajian mendapati bahawa perubahan mendadak berpotensi berlaku pada tahun 1999, 2001, dan 2002 untuk purata tahunan kelembapan relatif, suhu, dan jumlah hujan. Kesemua titik perubahan yang telah dikenalpasti kemudiannya dianalisis sebagai calon perubahan cuaca mendadak dengan mengembangkan lagi kaedah yang dicadangkan pada kajian akan datang.

Kata kunci: Analisis titik perubahan, perubahan mendadak, CUSUM, bootstrap, cuaca

© 2017 Penerbit UTM Press. All rights reserved

1.0 INTRODUCTION

An abrupt change is defined as a change that occurs suddenly or unfolds faster than expected, thereby forcing a reactive rather than proactive mode of behavior [1]. Such changes can spread systemically and can rapidly affect numerous interconnected areas of concern. The application of abrupt change detection in climate studies is important because it can provide an early warning of an impending disaster and the consequent damage that could occur. An abrupt climate change occurs when the climate system is forced to cross some threshold, thereby triggering a transition to a new state at a rate that is determined by the climate system itself and which is faster than the cause [2]. It has been argued that there is an urgent need for research on ways to detect and monitor abrupt change due to the current state of knowledge on potential abrupt changes to the ocean, atmosphere, ecosystems, and high latitude areas [3]. In the field of data mining, abrupt change can be seen as a sudden change in a data stream. A data stream is time series data that are collected on a consistent ongoing basis. Recently, research has been conducted on pattern discovery in data streams, which is a topic that is viewed as a big data problem. A data stream consists of a huge amount of data—sometimes many decades worth—that potentially contains important patterns that could provide evidence of abrupt changes. However, to date, many algorithms have failed to detect meaningful change because mostly they interpret the change as an outlier. If they were able to detect these meaningful changes perhaps they could be used to predict abrupt changes earlier than the currently available technology is able to do. On the other hand, statistical analysis has proved to be a useful technique that can be used to detect climatic patterns and changes in the past [4]. Change-point analysis is one of the statistical approaches that have been widely adopted by researchers to detect change points in a data stream and it has been shown to have the potential to detect abrupt climate change candidates [5].

Recently, the use of change-point methods to detect past abrupt shifts in time series data, especially climate data, has become widespread. A change-point is a point that shows where a shift in the data pattern occurs. It is also defined as a point at which the parameters (such as such as mean, variance, and trend) of an underlying distribution or the parameters of a model used to describe time series abruptly changes [6]. In previous studies, especially in earth sciences, change-point analysis has been applied to detect changes in temperature [7, 8], precipitation [9, 10], and the river water level [11], as well as to study past changes in the land uptake of carbon [12]. The method has also been used to study changes in the variance of oceanographic time series [13]. In this study, change-

point analysis was chosen as analytic tool to solve a climate-related problem. Specifically, 25 years of temperature, rainfall, and relative humidity information contained in a climate dataset for North Kedah, Malaysia was investigated. The main objective of this study was to discover any interesting trends and potential abrupt shifts in the dataset.

The remainder of this paper is organized as follows: Section 2.0 explains the change point analysis approach and the climate dataset used in this study. In Section 3.0, it presents the experimental results and discussion, while Section 4.0 provides the conclusions and future work on the present study.

1.1 Related Work

In recent years, the change-point analysis method has proven to be a useful analytic tool in analyzing time series datasets and identifying underlying trends. Most of the results presented in previous studies show that change-point analysis is capable of revealing the existence of hidden change points in time series or sequence datasets. It has been widely used in various fields, such as medical [14, 15], astronomy [16], and finance [17], as well as climate.

Change-point analysis has also been used in public health control. For example, [18] applied this method to active syndromic monitoring data to detect changes in the incidence of emergency department visits for influenza-like sickness during the H1N1 pandemic. The study used three change-point analysis techniques—the cumulative sum (CUSUM) chart, Bayesian, and structural change model—to detect the changes in daily time series data from the United States Distributed Project. Change-point analysis produced a good result because the method was able to identify change points and indicate the trend in the disease.

The effectiveness of change-point analysis in analyzing time series data has encouraged climate researchers to implement this method in the field of meteorology. The keen interest in this type of analysis can be seen in the increased number of change-point analysis techniques that have been used in this field, especially in research related to time series-based datasets such as those for temperature and rainfall. For example, [10] employed a Bayesian technique to analyze extreme changes in an annual rainfall dataset for Peninsular Malaysia. The study used a dataset covering 29 years from 1975 to 2004 collected from 50 meteorological stations. The changes in the average level of the extreme rainfall value were determined based on a single-shift model. Based on the analysis, the results showed that some of the stations observed significant changes, especially in the early 1990s. The study also revealed that more than 75% of the stations that experienced a significant change lay on the west coast of Peninsular Malaysia. However, the study also noted that the existence of anomalies in time series data may have affected the results because the analysis

relied on the average value, which is very sensitive to anomalies. Nonetheless, the research showed that change-point analysis is an effective tool that can be used to gain an understanding of the significant changes that occur in climate data.

Change-point analysis has also been used to investigate changes in the patterns of annual and seasonal weather in Iran [19]. Three statistical techniques were used in the study: Pettitt's test, the Mann-Kendall sequential test (MK SQ test) and the Mann-Kendall rank test (MK test). The study used a 57-year weather dataset (1950–2007) that consisted of annual and seasonal rainfall data, average temperature, maximum temperature and minimum temperature. From the experiment, it was found that there was a mutation or a positive change-point in 1990 for temperature and no significant turning point in annual rainfall. Nevertheless, the overall finding showed that change-point analysis is capable of analyzing climatic time series datasets.

In another work, [11] tested a change-point analysis method on hydrological time series data. Specifically, the study examined stream-typed data of river waters in Shunde, Guangdong, China for the period 1952–1997 by using two techniques: gray relational and translational relational. The results of the study were in line with intuitive observations that changes occurred in 1991, 1992 and 1993.

In addition, [7] discovered a change-point in a time series dataset of mountain temperatures. The study involved the investigation of an average daily temperature dataset covering 46 years from 1969 to 2007, which was compiled from four meteorological stations in the mountainous areas of the Romanian Carpathians, namely Ceahlau Toaca, Omu, Tarcu and Vladeasa. The study applied a two-phase regression model adapted from [20] and [21]. The results revealed a change-point in the mid-1980s. Similar to the above studies, [22] also utilized the capability of change-point analysis. In this instance, it was used to analyze changes in temperature information. For the analysis the study used a dataset covering 69 years from 1941 to 2010, which was taken from the Krishnanagar weather observatory in India. The study employed two change-point analysis techniques, CUSUM and bootstrap, to analyze change points in the yearly average temperature, maximum temperature and minimum temperature. This method revealed that several abrupt changes occurred in 1951, 1964, 1971, 1978, 1984, and 1991 for the yearly maximum and in 1971, 1978, 1986, 1991, 2000 for the yearly minimum temperature time series.

Some other studies based in the United States have also utilized change-point analysis. For instance, [12] employed a two-phase regression model technique to detect undocumented change points in temperature data taken from Chula Vista, California in the United States over a period of 78 years from 1919 to 1996. Through this technique, the study was able to demonstrate the existence of change points in 1966, 1982 and 1985. Change-point analysis was also used as a tool to detect abrupt change in carbon dioxide concentration in Mauna Loa, Hawaii, where [6] and [12] presented the results of a change-point analysis based on the informational approach. In light of the above, change-point analysis seems to have the ability to detect potential abrupt change with good detection results. In particular, change-point analysis that employs the CUSUM and bootstrap techniques, as implemented in [22], has shown good results that also seem to be more comprehensive than others. Inspired by the success reported in [22] and elsewhere, this study applied a similar method to investigate variation in the climate data for Malaysia. Table 1 provides an overview of the above-discussed previous works on change-point analysis conducted on time series datasets.

2.0 METHODOLOGY

In this study, two change-point analysis techniques are combined, namely CUSUM charts and bootstrap rank statistics, as suggested by [23]. Generally, a CUSUM chart indicates an out-of-control activity when a process intersects a boundary by an ascending or descending drift of the cumulative sum, while the bootstrap technique introduced by [24] is a common resampling technique for reckoning the distributions of statistics based on independent observations. Change-point analysis uses a recursive algorithm to identify multiple changes and an iterative algorithm decomposes the dataset into sub-datasets having different means. This method produces a series of estimated change points with a different confidence level and the optimal solution is determined by minimizing the false positive changes through a backward elimination procedure. The results are then analyzed to identify any potential abrupt change.

Table 1 Recent Change Point Analysis Studies

No.	Techniques	Researchers	Location	Domain	Data for >10 Years
1	Circular binary segmentati	Olshen et al. [14]	United States	Medical	No
2	CUSUM chart, bootstrap	Zhiheng Xu et al. [15]	United States	Medical	Yes
3		Seo-Won Chang et al. [16]	Seoul, Korea	Astronomy	No
4		Dipak Bisai et al. [22]	Krishnanagar, India	Climate (temperature)	Yes
5	Linear regression model	Zeileis A. et al. [17]	China	Finance	Yes
6	CUSUM chart, Bayesian, structural change model	Taha Kass-Hout et al. [18]	United States	Medical	No
7	Bayesian	Wan Zawiyah et al. [10]	Peninsular Malaysia	Climate (rainfall)	Yes
8	Pettitt's test, Mann-Kendall sequential test (MK SQ-test), Mann-Kendall rank test (MK-test)	Mohammad Zarenistanak et al. [19]	Iran	Climate (rainfall, temperature)	Yes
9	Gray relational	H. Wong et al. [11]	Guangdong, China	Climate (river)	Yes
10		Adlina-Eliza et al. [7]	Romanian Carpathians	Climate (temperature)	Yes
11	Two-phase regression model	Robert Lund et al. [20]	Chula Vista, California, USA	Climate (temperature)	Yes
12		Robert Lund et al. [12]	Mouna Loa, Hawaii	Climate (carbon dioxide)	Yes
13	Informational approach	Claudie Beaulieu et al. [6]	Mauna Loa, Hawaii	Climate (carbon dioxide)	Yes

2.1 Change-Point Analysis Algorithm

Let x_1, x_2, \dots, x_n symbolize n data points in time series and let S_0, S_1, \dots, S_n define the cumulative sum of the points. Change-point analysis is computed by applying three steps to the initial dataset $D_0 = \{X_1, \dots, X_n\}$ of size n ($n_0 = |D_0|$). The mean \bar{x} of x_1, x_2, \dots, x_n is formulated by

$$\bar{x} = \frac{x_1+x_2+\dots+x_n}{n} \tag{1}$$

The cumulative sum always starts at zero, 0. Therefore, let S_0 be equal to zero, $S_0 = 0$.

Then, S_i is calculated repeatedly as follows:

$$S_i = S_{i-1} + (x_i - \bar{x}), i = 1, 2, 3, \dots, n \tag{2}$$

Before bootstrap analysis is executed, an estimation of the magnitude of the changes is needed to create a boundary for the chart. It is computed as

$$S_{diff}^i = \max_{i=0, \dots, n} S_i - \min_{i=0, \dots, n} S_i = S_{max} - S_{min} \tag{3}$$

The bootstrap analysis is then executed N times on D_0 , after the magnitude of change has been calculated. A single bootstrap is executed as follows:

- i. A bootstrap dataset D_1 of size n from data points of time series in dataset D_0 is represented as x_j ($j = 1, 2, 3, \dots, n$). This dataset is generated by original n values which are randomly reordered, which is also known as sampling without replacement (SWOR).
- ii. The bootstrap CUSUM is computed by following a similar method based on the bootstrap sample and is defined as S_j .
- iii. The magnitude of change for the bootstrap CUSUM is calculated as follows

$$S_{diff}^j = \max_{j=0, \dots, n} S_j - \min_{j=0, \dots, n} S_j = S_{max} - S_{min} \tag{4}$$

- iv. Then, where the original magnitude of change is more than the magnitude of change of bootstrap CUSUM, $S_{diff}^i > S_{diff}^j$, the number of bootstraps is counted. Let N be the number of bootstrap sample executed and K be the number of

bootstraps for which $S_{diff}^i > S_{diff}^j$, where the confidence level that a change has occurred as a percentage is defined as

$$CL = \frac{\sum_{i=1}^N I(S_{diff}^i > S_{diff}^j)}{N} \times 100 = \frac{K}{N} \times 100 \quad (5)$$

The bootstrapping result is a distribution-free approach with only a single assumption, namely that of an independent error structure [18]. An independent error structure is when the data points are distributed as follows:

$$x_i = m_i + e_i \quad (6)$$

where m_i is the mean at time i . Normally, $m_i = m_{i-1}$ except for a small number of values of i that are called change points. Meanwhile, e_i is a random error correlated with the i -th value and the independent e_i is assumed to have a zero mean value, to be identically distributed and to be normally distributed.

An estimation of when the change occurred can be made when a change has been detected. The estimator, which is the CUSUM estimator, is computed as follows. Let m be such that

$$|S_m| = \max_{i=0, \dots, n} |S_i| \quad (7)$$

where S_m is the furthest point from the zero value in the CUSUM chart and the last point before the change occurred is estimated by point m while point $m+1$ estimates the first point after the change occurred. The mean square error (MSE) is used as the second estimator when the change happens:

Let $MSE(m)$ be defined as $MSE(m) = \min_{i=1, \dots, |D|} MSE(i)$ for a given sub-dataset D as follows:

$$MSE(i) = \sum_{j=1}^i (X_j - \bar{X}_1)^2 + \sum_{j=i+1}^{|D|} (X_j - \bar{X}_2)^2 \quad (8)$$

where $\bar{X}_1 = \frac{\sum_{j=1}^i X_j}{i}$ and $\bar{X}_2 = \frac{\sum_{j=i+1}^{|D|} X_j}{i}$

The MSE estimation is performed by dividing the data series into two parts: 1 as m and $m+1$ as n , before it is estimated how well the data in each part suits their equivalent mean. The value of m that minimizes $MSE(m)$ represents the best estimator of the last point before the change, whereas $m+1$ represents the first point after the change. The data are divided into sub-segments as soon as a change is detected. In a similar manner, the same method is applied to the data of each segment to get another change point that splits equivalent segments into sub-segments. Multiple changes can be detected by performing a repeated analysis to get other significant change points at consequent levels and the confidence limits and levels are calculated by the bootstrapping technique.

After the change points with their confidence levels are identified, a backward elimination procedure is performed to remove the insignificant points. When a point is eliminated, the surrounding change points and their significance levels are re-estimated to reduce the proportion of false detections. It should be noted that outliers are treated as isolated points. If the data contain outliers, this method analyzes the data by rank rather than by actual data values. Therefore, the change-point analysis only detects less continuous changes with little intrusion from outliers. Hence, significant change points can be discovered by applying this method to time series data on rainfall, temperature and relative humidity, which are the climate data of interest to this study.

2.2 The North Kedah Climate Dataset

This study tested the proposed change-point analysis approach on a Malaysian time series climate dataset, the North Kedah climate dataset, which was collected by a meteorological station at Sultan Abdul Halim Airport, Alor Setar, Kedah, in the northwest of Peninsular Malaysia (see Figure 1). The station is located at 6.1248° N, 100.3678° E and has an average elevation of 3.9 meters above mean sea level. The dataset contains about 24 years of climate information on North Kedah dating from 1985 to 2008. It consists of 19,872 records and three variables: hourly mean rainfall, hourly mean temperature and hourly relative humidity. According to the Malaysian Meteorology Department, this station is in an area that has a mean annual rain of about 1,990 mm and a mean annual relative humidity of 82% and where the temperatures are rather consistent throughout the year at around 27.9 degrees Celsius, with an average high temperature of around 32.0 degrees Celsius and an average low temperature of around 23.0 degrees Celsius.

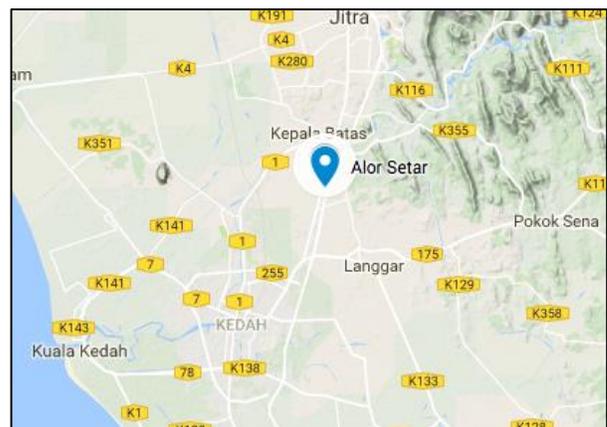


Figure 1 Meteorological station at Sultan Abdul Halim Airport, Alor Setar, Kedah, Malaysia

As Malaysia is mainly surrounded by ocean, both land and sea breezes have a huge impact on the wind flow pattern, especially on clear days. According to the Malaysian Meteorological Department, sea breezes with a wind speed of between 10 and 15 knots often occur and can travel several tens of kilometers inland on bright afternoons. The reverse process also occurs, where land breezes of weaker strength travel to coastal areas on clear nights. Generally, the monthly rainfall pattern shows two episodes of maximum rainfall separated by two episodes of minimum rainfall. The major maximum episodes of rainfall mostly take place from October to November while the minor maximum episodes generally appear from April to May. On the other hand, the major minimum rainfall occurs from January to February while the minor minimum takes place from June to July. In contrast, in the northwest of Peninsular Malaysia including Kedah, the major minimum rainfall occurs from June to July and the minor minimum in February. As for the major maximum and minor minimum, these remain the same. These observations gave a potential opportunity to research the local weather from North Kedah.

2.3 Research Methodology

In this study, the research methodology consisted of three main activities: data preprocessing, data mining by using the proposed approach and analysis of the results, as shown in Figure 2. In the data preprocessing step, all raw data were clean and well prepared before the data mining process. The preprocessing step in this work is conducted through applying basic data mining preprocessing step. In this study, the dataset was separated into two groups so that the data could be analyzed annually and hourly. After the dataset had been preprocessed, the change-point analysis approach described in Section 2.1 was applied to mine the data. The results were then analyzed to identify the occurrence of change points which are considered as potential abrupt change.

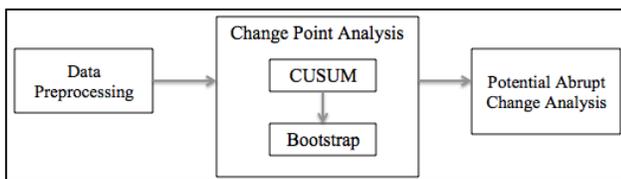
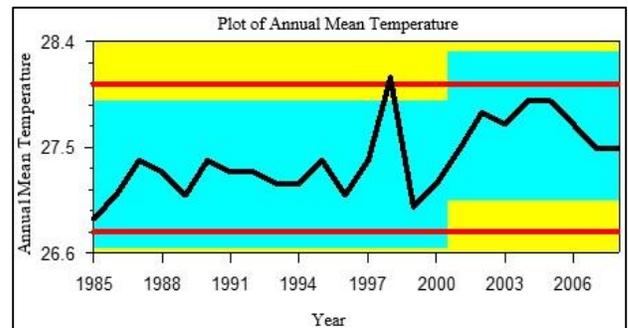


Figure 2 Research methodology

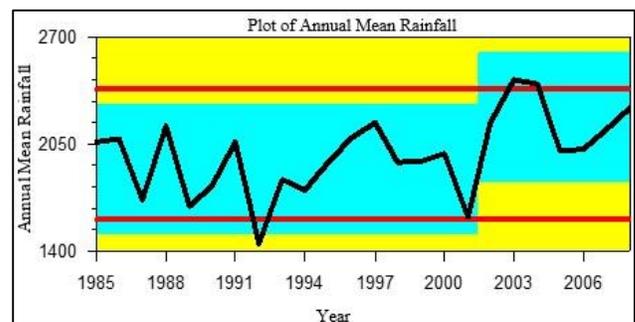
3.0 RESULTS AND DISCUSSION

This section presents the trends in climate change for North Kedah that were identified by using the proposed change-point analysis approach. From the experiment, change-point analysis was able to

detect several change points in the annual mean temperature, annual mean rainfall, and annual mean relative humidity of North Kedah. Previously, this method was used to investigate the hourly mean data for temperature, rainfall and relative humidity in North Kedah from 2000 to 2008, the results of which are shown in Appendices A, B and C, respectively. Figure 3(a)–Figure 3(c) shows the trends in the annual mean temperature, annual mean rainfall, and annual mean relative humidity of North Kedah from 1985 to 2008. In this figure, the blue region indicates the change points that have occurred throughout the years monitored while the red line represents the upper and lower limits. Based on the CUSUM, a point is considered to fall beyond the boundary if the value falls outside the limits and it is labeled as an outlier or known as an extreme value. From Figure 3, it is clear that some points can be labeled as outliers because the values of those points fall outside the boundary. In Figure 3(a), one point that is slightly outside the boundary from 1997 to 2000, while in Figure 3(b), points around 1992 and during the period 2002–2004 are outside the upper and lower boundary limits, respectively, whereas in Figure 3(c), all the points are inside the boundary.



(a)



(b)

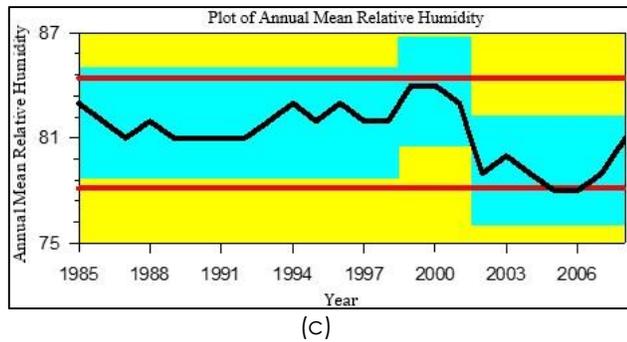


Figure 3 Annual trends for (a) mean temperature (b) mean rainfall (c) mean relative humidity

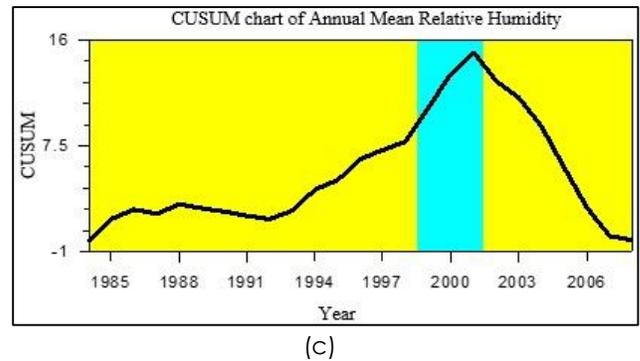
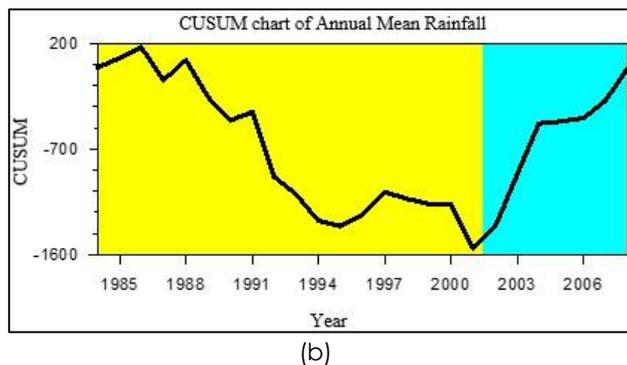
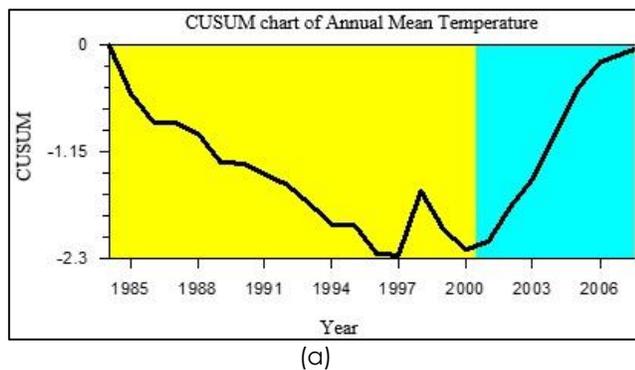


Figure 4 CUSUM chart for (a) annual mean temperature (b) annual mean rainfall (c) annual mean relative humidity

Figure 4(a)–(c) shows a graphical representation of the CUSUM analysis. The CUSUM chart was used to illustrate the cumulative sum of the data points, where the blue region represents the existence of a change point. A segment with an ascending trend (black upward moving line in blue region) in the CUSUM chart signifies a period of time where the value is above the total average and vice versa. From Figure 4(a)–(c), it can be seen that several significant changes occurred from 2000 to 2006 as the changes are in the blue region. For example, Figure 4(a) and (b) shows that there was an ascending trend after 2000. In contrast, Figure 4(c) shows that a descending trend was recorded after 2000. Hence, it is clear that changes have occurred in the climate dataset.



As explained in Section 2.1, CUSUM with bootstrapping identifies significant changes in a dataset based on the confidence level which can be classified into several levels. A level-1 change denotes the change that is detected first by the proposed change-point analysis approach and this change is highlighted by the blue region in the CUSUM chart. In the proposed approach, the detection process is repeated through several iterations to find other changes at other levels (level-2 to level-4). A level-1 change in an early iteration is the most noticeable change-point in the CUSUM chart and it is easier to detect by the proposed procedure. Therefore, the higher the level of change (e.g. level-2, level-3, level-4) the higher the possibility that a change has occurred. Table 2(a)–(c) shows the significant changes that were hidden inside the analyzed dataset, where the last column of the table represents the level of change.

For every subsequent segment of annual mean rainfall and annual mean relative humidity, change-point analysis assumes that the data are in the form of an independent error structure. No outliers were detected for either of the above characteristics (rainfall, humidity). However, the finding for annual mean temperature differed, where the data appeared to violate the assumption of independent errors. Since change-point analysis is a mean-shift model, dependent errors may be affected by the confidence level and confidence interval. The errors are positively correlated if one value from the data point is above average, and the next several values are also likely to be above average. Hence, the result of the analysis may falsely indicate that a change has occurred. However, this issue can be overcome by analyzing the change-point by ranking. In other words, the data are sorted into order and the values are replaced by the relative positions in the order, where a higher rank is associated with a larger value and vice versa. This makes the interpretation of the plots much simpler and more robust to outliers. The subsequent application of a bootstrapping technique confirmed that some change points had occurred in the data. Specifically, the results of the

bootstrapping analysis detected four change points in the data with a confidence level of more than 95%, as shown in Table 2(a)–(c).

Table 2 Significant Changes in (a) Annual Mean Temperature, (b) Annual Mean Rainfall and (c) Annual Mean Relative Humidity

Table of Significant Changes for Annual Mean Temperature

Year	Conf. Interval	Conf. Level	From	To	Level
2001	(1997, 2003)	97%	27.275	27.688	1

(a)

Table of Significant Changes for Annual Mean Rainfall

Year	Conf. Interval	Conf. Level	From	To	Level
2002	(1995, 2005)	97%	1903	2213.7	1

(b)

Table of Significant Changes for Annual Mean Relative Humidity

Year	Conf. Interval	Conf. Level	From	To	Level
1999	(1996, 1999)	98%	81.857	83.667	3
2002	(2002, 2002)	96%	83.667	79.143	1

(c)

From Table 2(a), it can be seen that the analysis discovered a significant change at level 1 for annual mean temperature with a confidence interval from 1997 to 2003 at a confidence level of 97%. Prior to the change in 2001, the annual mean temperature was 27.28 degrees Celsius before it increased slightly to 27.69 degrees Celsius. In Table 2(b), a significant level 1 change was detected for annual average rainfall in 2002 with a 97% confidence level. The analysis indicated that earlier, before the first change, the mean total rainfall was 1903 mm and after the first change it was 2213.7 mm.

From the experiment, two major changes were detected for mean average relative humidity. As shown in Table 2(c), the first change was estimated to have taken place around 1999. This period signifies the first year in which the change-point occurred before the second change, which was likely to have occurred around 2002. The confidence level associated with each change indicates the accuracy of the analysis in respect of the change that occurred. The first change has a higher confidence level (98%) than that of the second change (96%). The first change took place between 1996 and 1999 and the second change occurred

during 2002. The confidence interval for the first change (i.e. 1996–1999) is wider than that for the second change (i.e. during 2002). This indicates the time of the first change cannot be as precisely determined as the second change. As stated above, the level of change signifies the importance of the change. The level-1 change is the first change detected in the analysis, which is the most obviously apparent in the CUSUM chart in Figure 3(c), whereas the level-3 changes are detected on a third pass of the data. The magnitude of change (refer to formula (4)) for the second change is 4.524, whereas that for the first change is 1.81 (a difference of 2.714). Hence, the second change is more significant than the first change.

From the experiment, it is clear that a potential abrupt shift can be detected by change-point analysis. The idea of using change-point analysis to detect potential abrupt change is by analyzing the change points and the level of changes. In this study, only two levels were identified: level 1 and level 3. This might be because annual mean data were used. A level-3 change appears to denote a potential abrupt shift and as such it will be used in future work. Based on the results of this study, a potential abrupt shift in all three variables occurred during the period between 1995 and 2002, as summarized in Table 3. In terms of climatology, two extreme El-Niño events occurred between 1982 and 1983 and between 1997 and 1998. These events might have contributed to the changes in climate detected by the analysis. This is in line with [10], which discovered these change points during the El-Niño events by using Bayesian a change-point analysis method.

Table 3 Occurrence of potential abrupt shift

Variable	Year range
Temperature	1997 – 2003
Rainfall	1995 – 2005
Relative Humidity	1996 – 1999 2002 – 2002

The results of the study show that change-point analysis is a useful tool in detecting smaller continual changes. It can also handle all types of time series data including data from non-normal distributions, poorly behaved data and data with outliers. Moreover, this method can characterize a detected change pattern more accurately by providing correlated confidence levels and confidence intervals for the times of the changes. However, regardless of its various advantages, change-point analysis has a few inadequacies. The biggest shortcoming is that this method is unable to detect isolated abnormal points. Rather, when an outlier is found, it uses a rank technique on the data so that it can provide results that are robust to outliers. Another limitation is related to bootstrapping, where the technique is unable to create similar results each

time it executes because of random selection during bootstrap sampling. In order to generate results that are more precise, the number of bootstraps must be increased, which indirectly doubles the duration of the analysis. Nevertheless, change-point analysis is still one of the most widely used methods for detecting change points because of its flexibility and simplicity.

4.0 CONCLUSION

This paper presented the implementation of a change-point analysis method to detect change points in the North Kedah climate dataset. This study was motivated by the previous research in various fields that successfully discovered hidden trends through the utilization of change-point analysis. Two change-point techniques were used in this study: the CUSUM and bootstrapping. From the analysis, several major changes in mean annual temperature, mean annual total rainfall and mean annual relative humidity were found to have occurred between 1995 and 2002. These identified change points can be regarded as potential abrupt shifts. This finding can be considered proven because it is in line with previous research about general climatological conditions in Malaysia that saw the occurrence of extreme El-Niño events in 1997 and 1998. Moreover, the results of the study showed that change-point analysis can be used to detect past climate trends and variations. It is not suggested that this method can predict future shifts. However, in future work, the results of change-point analysis will be used to help to identify any potential abrupt changes and more improvements will be made to the approach in order to develop a robust framework for observing abrupt shifts and making more accurate predictions.

Acknowledgments

The authors gratefully acknowledge the Universiti Kebangsaan Malaysia (UKM), Institute of Climate Change UKM, as well as the Fundamental Research Grant Scheme for supporting this research project through grant FRGS/1/2014/ICT02/UKM/01/2.

References

- [1] Alley, R. B. et al. 2002. Abrupt Climate Change: Inevitable Surprises. US National Research Council Report. National Academy Press, Washington DC. 27.
- [2] Alley, R. B. 2002. Abrupt Climate Change: Inevitable Surprises. US National Research Council Report. National Academy Press. Washington DC. 14.
- [3] Elena A Khapalova, Venkata K. Jandhyala, & Stergios B. Fotopoulos. 2013. Change-point Analysis of Annual Mean Precipitation for Northern Tropical and Southern Latitudes of the Globe in the Past Century. *Journal of Environmental Statistics*. 4(3): 1-21.
- [4] S. Shanmugasundram. 2012. *Statistical Analysis to Detect Climate Change and Its Implications on Water Resources* (Doctoral dissertation). Victoria University, Australia.
- [5] Zhao, F. F., Xu, Z. X. & Zhang, L. 2012. Changes in Stream Flow Regime Following Vegetation Changes from Paired Catchments. *Hydrological Processes*. 26: 1561-1573.
- [6] Beaulieu C., J. Chen, & J. L. Sarmiento. 2012. Change-Point Analysis as a Tool to Detect Abrupt Climate Variations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 370 (1962): 1228-1249.
- [7] Adina-Eliza Croitoru, Dorin Drignei, Iulian-Horia Holobaca, & Carmen Sofia Dragota. 2012. Change-point Analysis for Serially Correlated Summit temperatures in the Romanian Carpathians. *Theor Appl Climatol*. 108: 9-18.
- [8] A. Shirvani. Change Point Detection of the Persian Gulf Sea Surface Temperature. 2015. *Theoretical and Applied Climatology*. 127(1): 123-127.
- [9] Elena A Khapalova, Venkata K. Jandhyala, and Stergios B. Fotopoulos. 2013. Change-point Analysis of Annual Mean Precipitation for Northern Tropical and Southern Latitudes of the Globe in the Past Century. *Journal of Environmental Statistics*. 4(3): 1-21.
- [10] Wan Zawiah Wan Zin, Abdul Aziz Jemain and Kamarulzaman Ibrahim. 2012. Bayesian Change Point Analysis of the Extreme Rainfall Events. *Journal of Mathematics and Statistics*. 8(1): 85-91.
- [11] H. Wong, B.Q. Hu, W.C. Ip, & J. Xia. 2006. Change-point Analysis of Hydrological Time Series Using Grey Relational Method. *Journal of Hydrology*. 324: 323-38.
- [12] Robert Lund & Jaxk Reeves. 2002. Detection of Undocumented Change Points: A Revision of the Two-phase Regression Model. *Journal of Climate*. 15: 2547-2554.
- [13] Killick, Rebecca, Idris A. Eckley, Kevin Ewans, & Philip Jonathan. 2010. Detection of Changes in Variance of Oceanographic Time-series using Change Point Analysis. *Ocean Engineering*. 37(13): 1120-1126.
- [14] Jie Chen & Arjun K. Gupta. 2012. *Parametric Statistical Change Point Analysis With Applications to Genetics, Medicine, and Finance*. 2: 212-215.
- [15] Zhiheng Xu, Taha Kass-Hout, Colin Anderson-Smits, and Gerry Gray. 2015. Signal Detection Using Change Point Analysis in Postmarket Surveillance. *Pharmacoepidemiology and Drug Safety*. 24: 663-668.
- [16] Seo-Won Chang & Yong-Ik Byun. 2012. Variability Detection by Change-point Analysis. *Statistical Challenges in Modern Astronomy*. Springer New York. 5: 491-493.
- [17] Zeileis A., Ajay Shah, and Ila Patnaik. 2010. Testing, Monitoring, and Dating Structural Changes in Exchange Rate Regimes. *Computational Statistics & Data Analysis*. 54(6): 1696-1706.
- [18] Taha A Kass-Hout, Zhiheng Xu, Paul McMurray, Soyoun Park, David L Buckeridge, John S. Brownstein, Lyn Finelli, & Samuel L Groseclose. 2012. Application of Change Point Analysis to Daily Influenza-Like Illness Emergency Department Visits. *Journal of the American Medical Informatics Association : JAMIA*. 19(6): 1075-1081.
- [19] Zarenistanak M., Dhorde A.G., & Kripalani R.H. 2014. Trend Analysis and Change Point Detection of Annual and Seasonal Precipitation and Temperature Series Over Southwest Iran. *Journal of Earth System Science*. 2: 281-295.
- [20] Lund, R. 2007. Change-point Detection in Periodic and Autocorrelated Time Series. *Journal of Climate*. 20: 5178-5190.
- [21] Toreti A, Desiato F, Fioravanti G, & Perconti W. 2010. Seasonal Temperatures Over Italy and Their Relationship with Low Frequency Atmospheric Circulation Patterns. *Journal of Climate Change*. 99: 211-227.
- [22] D. Bisai, S. Chatterjee, A. Khan, & N. K. Barman. 2014. Long Term Temperature Trend and Change Point : A Statistical

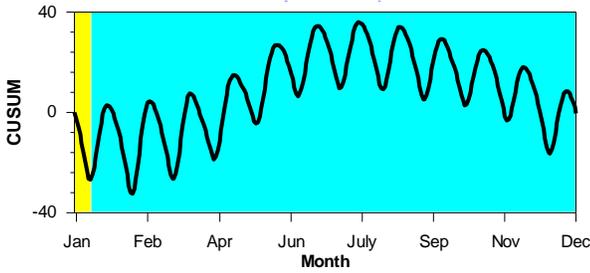
- approach. *Open Journal of Atmospheric and Climate Change*. 1(1): 32-42.
- [23] Taylor W. 2010. *Change-Point Analysis: A Powerful New Tool for Detecting Changes*. Taylor Enterprises.
- [24] B. Efron. 1979. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*. 7(1): 1-26.

Appendix

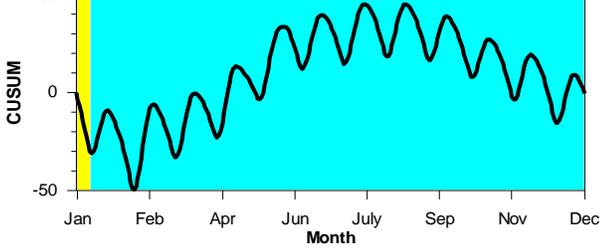
APPENDIX A

CUSUM Chart of Hourly Mean Temperature for 2000 – 2008

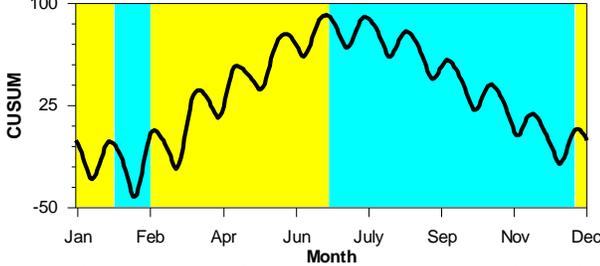
CUSUM Chart of Hourly Mean Temperature in 2000



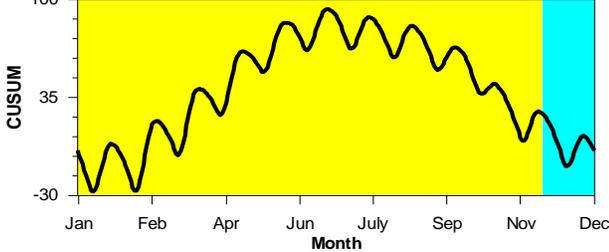
CUSUM Chart of Hourly Mean Temperature in 2001



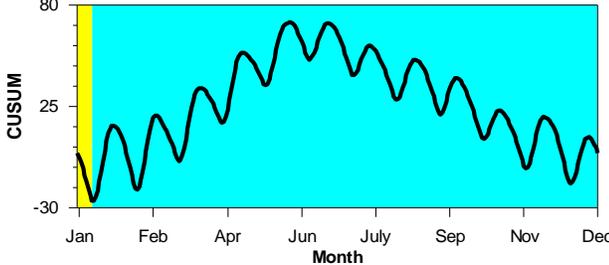
CUSUM Chart of Hourly Mean Temperature in 2002



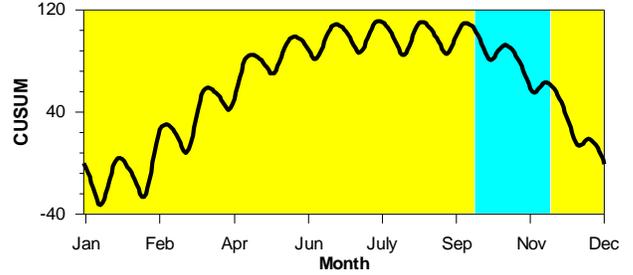
CUSUM Chart of Hourly Mean Temperature in 2003



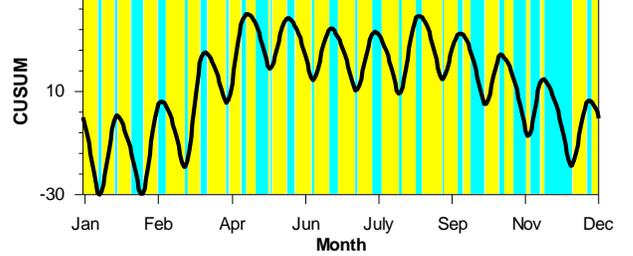
CUSUM Chart of Hourly Mean Temperature in 2004



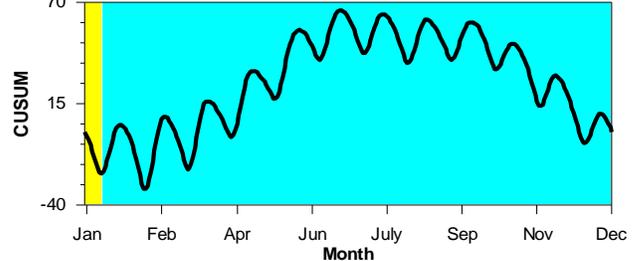
CUSUM Chart of Hourly Mean Temperature in 2005



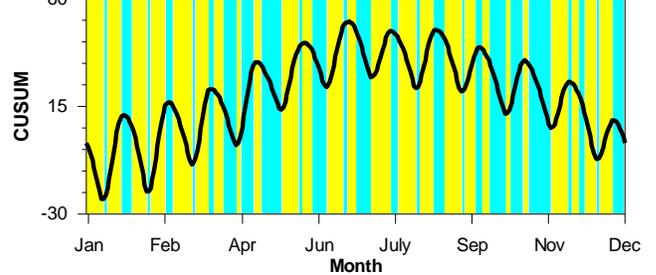
CUSUM Chart of Hourly Mean Temperature in 2006



CUSUM Chart of Hourly Mean Temperature in 2007



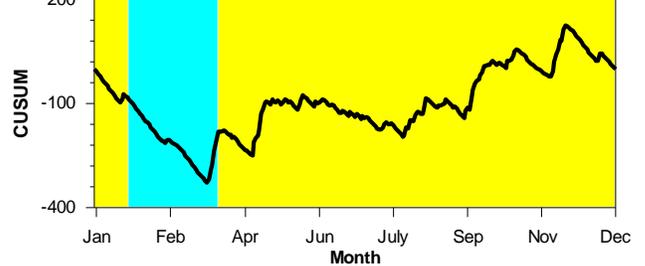
CUSUM Chart of Hourly Mean Temperature in 2008

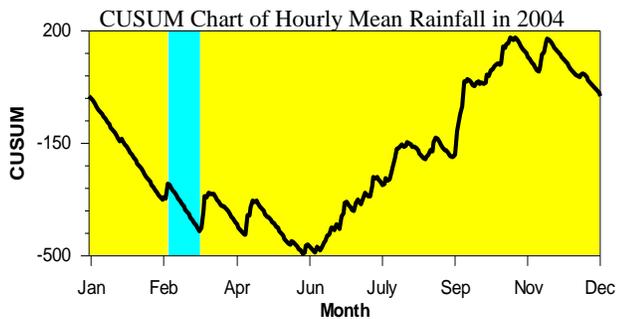
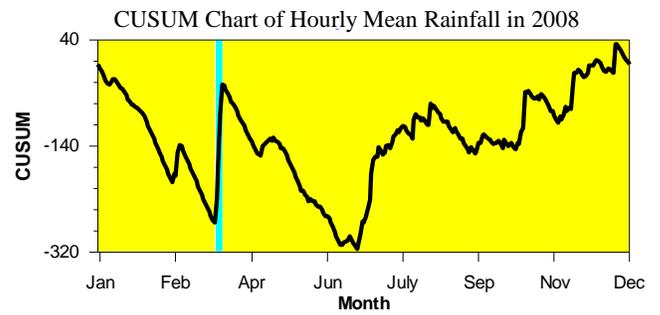
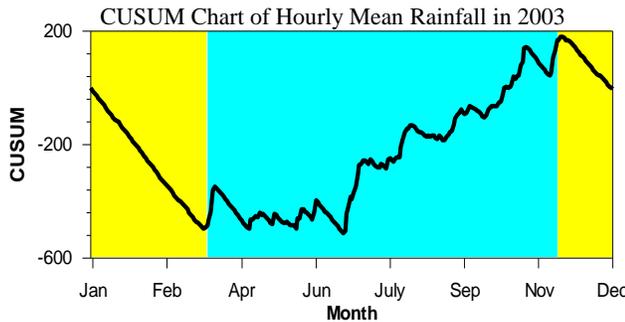
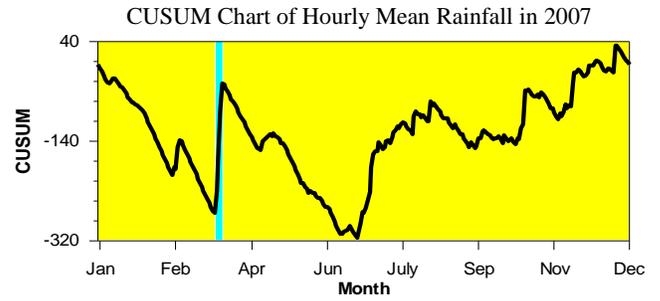
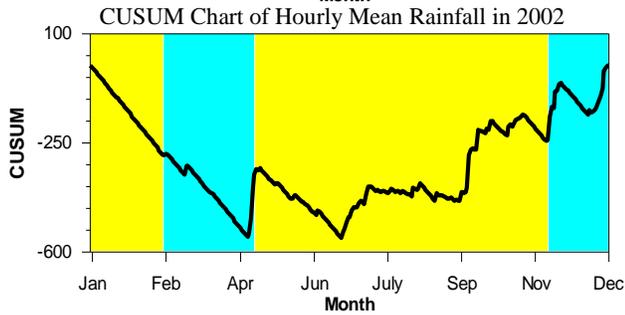
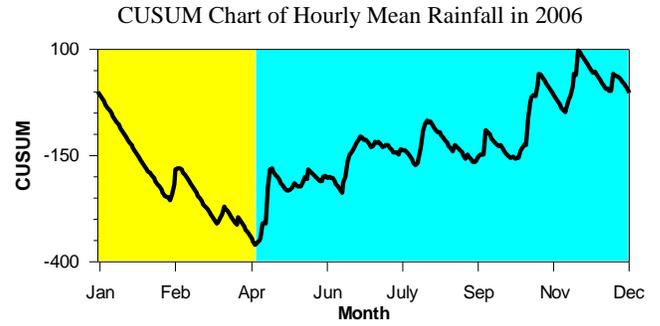
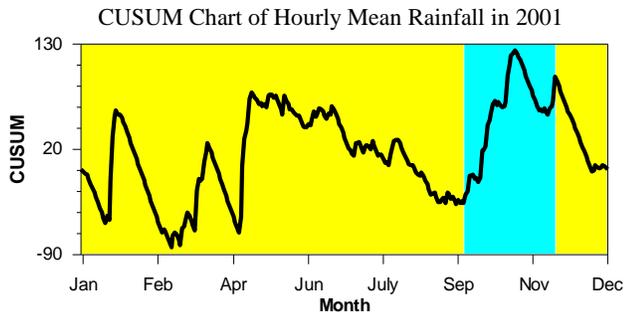


APPENDIX B

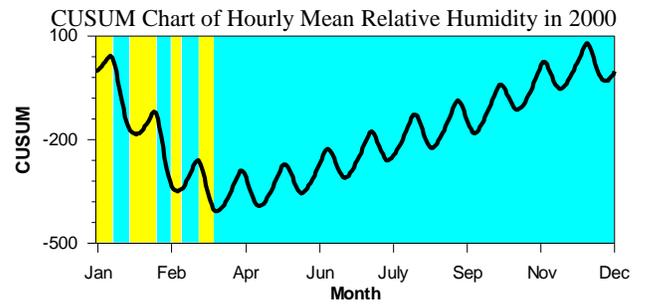
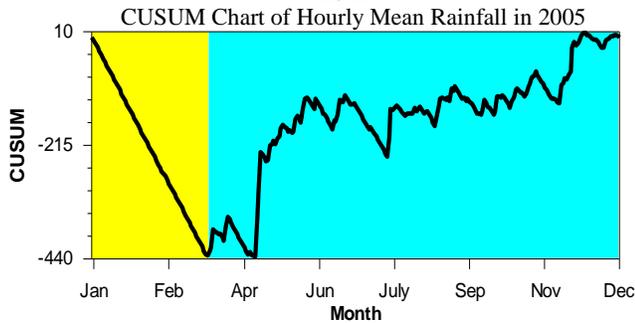
CUSUM Chart of Hourly Mean Rainfall for 2000 – 2008

CUSUM Chart of Hourly Mean Rainfall in 2000

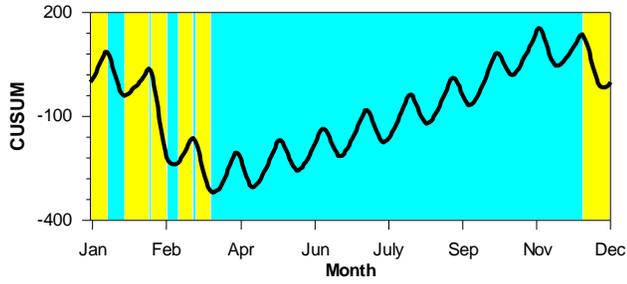




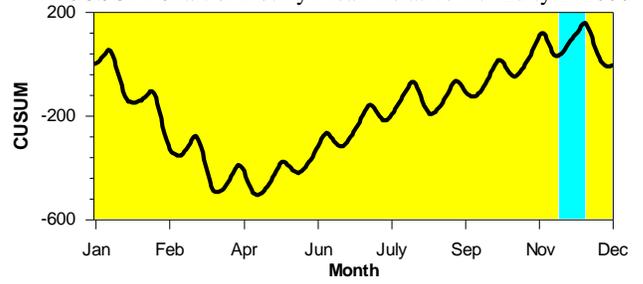
APPENDIX C
CUSUM Chart of Hourly Mean Relative Humidity
for Year 2000 – 2008



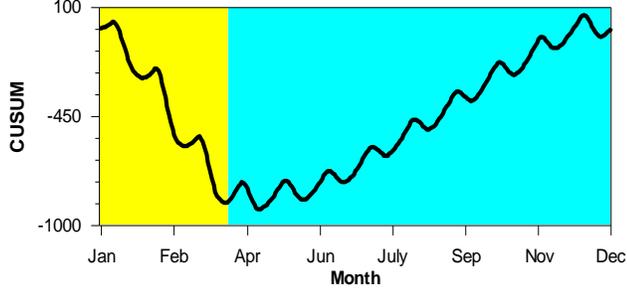
CUSUM Chart of Hourly Mean Relative Humidity in 2001



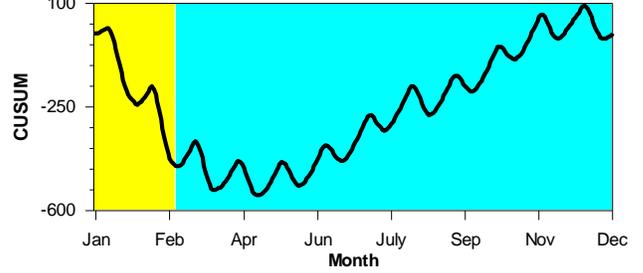
CUSUM Chart of Hourly Mean Relative Humidity in 2006



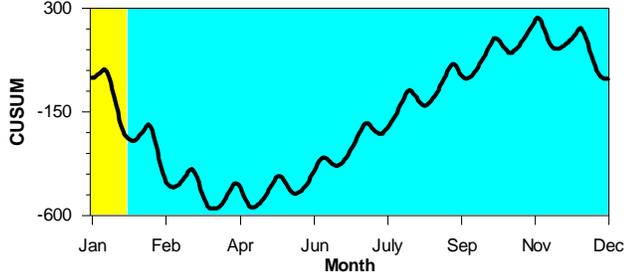
CUSUM Chart of Hourly Mean Relative Humidity in 2002



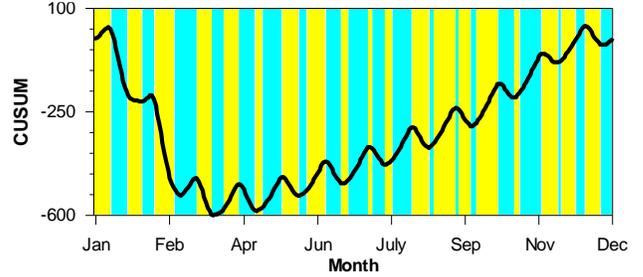
CUSUM Chart of Hourly Mean Relative Humidity in 2007



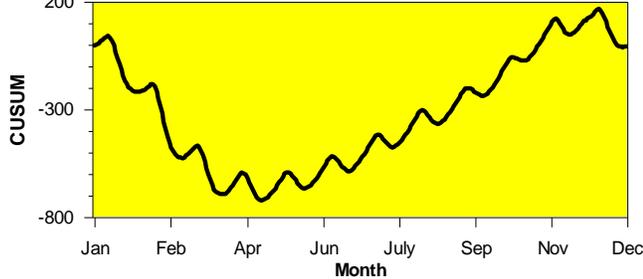
CUSUM Chart of Hourly Mean Relative Humidity in 2003



CUSUM Chart of Hourly Mean Relative Humidity in 2008



CUSUM Chart of Hourly Mean Relative Humidity in 2004



CUSUM Chart of Hourly Mean Relative Humidity in 2005

