Statistical Methods for Analyzing Rainfall in Thailand and the Southern Oscillation Index

Mayuening Eso

A Thesis Submitted in Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Research Methodology

Prince of Songkla University

2016

Thesis Title Statistical Methods for Analyzing Rainfall in Thailand and the Southern Oscillation Index

Author Miss Mayuening Eso

Major Program Research Methodology

Major Advisor:

……………………………………
(Emeritus Prof. Dr. Don McNeil)

Co-advisors:

……………………………………
(Asst. Prof. Dr. Metta Kuning)

……………………………………
(Dr. Somporn Chuai-Aree)

Examining Committee:

…..………………………………Chairperson
(Asst. Prof. Dr. Nittaya McNeil)

……………………………………
(Emeritus Prof. Dr. Don McNeil)

……………………………………
(Asst. Prof. Dr. Metta Kuning)

……………………………………
(Dr. Somporn Chuai-Aree)

……………………………………
(Prof. Dr. Khairil Anwar Notodiputro)

The Graduate School, Prince of Songkla University, has approved this thesis as fulfillment of the requirements for the Doctor of Philosophy Degree in Research Methodology.

………..……………………………...
(Assoc. Prof. Dr. Teerapol Srichana)
Dean of Graduate School

This is to certify that the work here submitted is the result of the candidate's own investigations. Due acknowledgement has been made of any assistance received.

……………..……………………Signature
(Emeritus Prof. Dr. Don McNeil)
Major Advisor

………………………………Signature
(Miss Mayuening Eso)
Candidate

I hereby certify that this work has not been accepted in substance for any degree, and is not being currently submitted in candidature for any degree.

……………..……………………Signature

(Miss Mayuening Eso)

Candidate

ชื่อวิทยานิพนธ์ วิธีการทางสถิติสำหรับวิเคราะห์ปริมาณน้ำฝนของประเทศไทยและดัชนีความผันแปรของระบบอากาศบริเวณซีกโลกใต้

ผู้เขียน นางสาวมายือนิง อิสอ

สาขาวิชา วิธีวิทยาการวิจัย

ปีการศึกษา 2558

บทคัดย่อ

วิทยานิพนธ์ฉบับนี้นำเสนอการประยุกต์ใช้วิธีการทางสถิติกับข้อมูลปริมาณน้ำฝนรายวันจากสถานีตรวจอากาศ 114 สถานีในประเทศไทย ปี พ.ศ. 2544 ถึง พ.ศ. 2555 และข้อมูลดัชนีความผันแปรของระบบอากาศบริเวณซีกโลกใต้ (Southern Oscillation Index: SOI) ปี พ.ศ. 2419 ถึง พ.ศ. 2557

การศึกษาแรกมีวัตถุประสงค์เพื่อจัดกลุ่มสถานีตรวจอากาศ 114 สถานีในประเทศไทยด้วยวิธีการวิเคราะห์ปัจจัย (Factor analysis) และศึกษารูปแบบการกระจายของปริมาณน้ำฝนรายวันของสถานีเหล่านั้นด้วยตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงแบบแกมมา (Gamma-distributed generalized linear model) ผลการศึกษาพบว่า การวิเคราะห์ปัจจัยสามารถแบ่งสถานีตรวจอากาศ 114 สถานีในประเทศไทยเป็นเจ็ดกลุ่ม ได้แก่ ภาคเหนือตอนบน ภาคเหนือตอนล่าง ภาคตะวันออกเฉียงเหนือ ภาคกลาง ภาคใต้ตอนบน ภาคใต้ฝั่งตะวันตก และภาคใต้ฝั่งตะวันออก ซึ่งวิธีการวิเคราะห์ปัจจัยสามารถอธิบายความแปรปรวนของข้อมูล 58.9% ทั้งนี้เมื่อพิจารณารูปแบบการกระจายของปริมาณน้ำฝนแต่ละภาค และความสัมพันธ์ของปริมาณน้ำฝนระหว่างภาคต่าง ๆ สามารถจัดกลุ่มสถานีตรวจอากาศใหม่ลดลงเหลือสี่กลุ่ม ได้แก่ ภาคเหนือ (รวมสถานีที่อยู่ทางภาคเหนือและภาคตะวันออกเฉียงเหนือ) ภาคกลาง (รวมสถานีที่อยู่ทางภาคกลางและภาคตะวันออก) ภาคใต้ฝั่งตะวันออก และภาคใต้ฝั่งตะวันตก

ขั้นต่อไปวิเคราะห์ข้อมูลปริมาณน้ำฝนทั้งสี่ภาคด้วยตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงแบบแกมมาโดยเลือกเฉพาะช่วงฤดูฝน สำหรับฤดูฝนของภาคเหนือและภาคกลาง คือ ช่วงเดือนเมษายนถึงตุลาคม (ประมาณหกเดือน) ฤดูฝนของภาคใต้ฝั่งตะวันออก คือ ช่วงปลายเดือนกันยายนถึงต้นเดือนมกราคม (ประมาณสี่เดือน) และฤดูฝนของภาคใต้ฝั่งตะวันตก คือ ช่วงเดือนเมษายนถึงพฤศจิกายน (ประมาณแปดเดือน) ผลการศึกษาพบว่า ปริมาณน้ำฝนรายวันเฉลี่ยแต่ละปีของภาคเหนือ ภาคกลาง และภาคใต้ฝั่งตะวันตกมีความแปรปรวนค่อนข้างน้อย แต่ปี พ.ศ. 2554 ภาคเหนือมีปริมาณน้ำฝนรายวันเฉลี่ยสูงกว่าค่าเฉลี่ยรวมอย่างมีนัยสำคัญ ส่วนภาคใต้ฝั่งตะวันออกมี

ปริมาณน้ำฝนรายวันเฉลี่ยแต่ละปีไม่คงที่ ในปี พ.ศ. 2548 2551 2553 และ 2554 มีปริมาณน้ำฝนรายวันเฉลี่ยสูงกว่าค่าเฉลี่ยรวม

เนื่องด้วยการศึกษาในอดีตรายงานว่า ดัชนีความผันแปรของระบบอากาศบริเวณซีกโลกใต้มีความสัมพันธ์กับปริมาณน้ำฝนในประเทศต่าง ๆ เช่น อินโดนีเซีย ออสเตรเลีย บราซิล และจีน เป็นต้น ดังนั้นการศึกษาที่สองมีวัตถุประสงค์เพื่อหาความสัมพันธ์ระหว่างดัชนีความผันแปรของระบบอากาศบริเวณซีกโลกใต้กับปริมาณน้ำฝนรายวันจากทั้งสี่ภาคของประเทศไทยซึ่งเป็นผลลัพธ์จากการศึกษาแรก และเพื่อตรวจสอบคุณสมบัติ (Characteristics) ของข้อมูลดัชนีความผันแปรของระบบอากาศบริเวณซีกโลกใต้ช่วงปี พ.ศ. 2419 ถึง พ.ศ. 2557 ผลการศึกษาพบว่า ปริมาณน้ำฝนรายวันเฉลี่ยทั้งสี่ภาคของประเทศไทย ปี พ.ศ. 2544 ถึง พ.ศ. 2555 ไม่มีความสัมพันธ์กับดัชนีความผันแปรของระบบอากาศบริเวณซีกโลกใต้ และเมื่อพิจารณาค่าสะสมรายเดือนของดัชนีความผันแปรของระบบอากาศบริเวณซีกโลกใต้ พบว่า สามารถแบ่งข้อมูลออกเป็นสี่ช่วงเวลา ดังนี้ ช่วงแรก คือ ปี พ.ศ. 2419 ถึง พ.ศ. 2462 ค่าดัชนีสะสมรายเดือนไม่ปรากฏแนวโน้มเพิ่มขึ้นหรือลดลง ช่วงที่สอง คือ ปี พ.ศ. 2463 ถึง พ.ศ. 2518 ค่าดัชนีสะสมรายเดือนมีแนวโน้มเพิ่มขึ้น ช่วงที่สาม คือ ปี พ.ศ. 2519 ถึง พ.ศ. 2538 ค่าดัชนีสะสมรายเดือนมีแนวโน้มลดลง และช่วงที่สี่ คือ ปี พ.ศ. 2539 ถึง พ.ศ. 2557 ค่าดัชนีสะสมรายเดือนไม่ปรากฏแนวโน้มเพิ่มขึ้นหรือลดลง

ดังนั้นเพื่อยืนยันว่า ข้อมูลดัชนีความผันแปรของระบบอากาศบริเวณซีกโลกใต้มีแนวโน้มเพิ่มขึ้นหรือลดลงในช่วงเวลาต่าง ๆ เหล่านั้นจริงหรือไม่ วิธีการทางสถิติที่เหมาะสม ได้แก่ ตัวแบบการถดถอยเชิงเส้น (Linear regression model) และตัวแบบถดถอยบนตัวเอง (Autoregressive model) ถูกนำมาประยุกต์ใช้ในการวิเคราะห์ข้อมูล โดยตัวแปรตาม คือ ดัชนีความผันแปรของระบบอากาศบริเวณซีกโลกใต้ และตัวแปรอิสระ คือ Boxcar function ซึ่งเป็นฟังก์ชันที่ใช้ระบุแนวโน้มในช่วงเวลาต่าง ๆ ของข้อมูล ผลการศึกษาพบว่า ไม่พบแนวโน้มใด ๆ ในสี่ช่วงเวลาข้างต้น จึงสรุปได้ว่า ดัชนีความผันแปรของระบบอากาศบริเวณซีกโลกใต้เป็นค่าที่เกิดขึ้นอย่างสุ่ม ๆ ไม่สามารถคาดการณ์ได้ (Random noise process)

Thesis Title Statistical Methods for Analyzing Rainfall in Thailand and the Southern Oscillation Index

Author Miss Mayuening Eso

Major Program Research Methodology

Academic Year 2015

ABSTRACT

This dissertation applies statistical methods to two different datasets. The thesis comprises two studies. The objectives of the first study are to classify daily rainfall of 114 stations in the period 2001-2012 across Thailand as well as to investigate their patterns. Factor analysis and gamma-distributed generalized linear model (GLM) are used to fit the data and the following results were observed. Factor analysis has been able to divide the area of Thailand into seven regions with explanation of 58.9% of the total variance. Those regions are North, Lower north, North-east, Central, Upper south, South-west and South-east. After observing the correlation between the stations, seven regions can be finally reduced into four distinct regions namely: north (combining the stations from north and north-east regions), central (combining the stations from central and east), south-east and south-west. The rainfall in rainy season for each region is then chosen for further analysis. In north and central, the rainy season is in the period of April to October (six months). In south-east and south-west, the rainy season is in late September to early January (four months) and April to November (eight months), respectively. Moreover, the response variables of the study are 5-day averages rainfall in the period of rainy season of each region and the determinants are year and month. The gamma models are used to fit the data for those

regions. The deviance residual plots from the models provide a reasonable fit. The results suggest that daily rainfall has less variation in north, central and south-west for the 12 years, but the north receives significantly heavy rainfall in 2011. However, daily rainfall in south-east fluctuates throughout the 12 years. The unusual heavy rain occurs in 2005, 2008, 2010 and 2011.

Some previous studies reported that rainfall has a relationship with Southern Oscillation Index (SOI) in various countries such as Indonesia, Australia, Brazil and China, etc. Therefore, the second study examines the relationship between large-scale SOI data and rainfall in Thailand and explores the characteristics of the SOI between 1876 and 2014. The correlation coefficients suggest that there are no association between rainfall and the SOI for the four regions of Thailand. In addition, the cumulative monthly SOI has shown that the data can be categorized into four periods. The first period, from 1876 to 1919, shows no trend. An increasing trend is apparent in the second period from 1920 until 1975 while a decreasing trend is apparent in the third period, 1976 to 1995. In the last period, between 1996 and 2014, the SOI appears fairly stable. The SOI is further examined using linear regression and autoregressive (AR) models on boxcar function where the function model the trends of the SOI. Finally, we can conclude that the SOI is quite similar to a random noise process.

Acknowledgements

This thesis would not have been successful without the help of mentioned people here, I would like to grab this opportunity to give a big thanks for them.

First of all, I would like to say “thank you very much” to my supervisors Emeritus Prof. Dr. Don McNeil, Asst. Prof. Dr. Metta Kuning and Dr. Somporn Chuai-Aree for their invaluable assistance encouragement and helpful guidance during my study.

Secondly, if without financial support, this thesis cannot go through so far. I would also like to acknowledge for a huge scholarships from (1) grant funds from the program of Strategic Scholarships Fellowships Frontier Research Networks (Specific for Southern Region) for the Ph.D. Program Thai Doctoral degree from the Office of the Higher Education Commission, Thailand, (2) the Centre of Excellence in Mathematics, the Commission on Higher Education, Thailand, (3) the Graduate School, Prince of Songkla University, Thailand, (4) International Statistical Institute (ISI) South East Asia Regional Network and (5) the organization of Deep South Coordination Center (DSCC), Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Thailand.

Thirdly, I am grateful to Thai Meteorology Department (TMD), Thailand for providing the data. Moreover, a many thanks to Prof. Alfredo Huete from University of Technology Sydney, Australia for his acceptance as a visitor student in a university in Sydney and providing me with exposure in Remote Sensing field during 24 February-23 May 2015 and 25 July-22 October 2015. I am thankful to the staffs in National Health and Medical Research Council (NHMRC) Clinical Trials Center, University of Sydney, Australia for providing a chance to engage in activities organized by them.

Fourthly, I would like to acknowledge teachers and staffs of Research Methodology at Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus consist of Asst. Prof. Dr. Chamnein Choonpradub, Asst. Prof. Dr. Phattawan Tongkumchum, Asst. Prof. Dr. Apiradee Lim, Asst. Prof. Dr. Nittaya McNeil and Dr. Attachai Ueranantasun for their encouragement and suggestions. Many thanks to Dr. Hilary Green the lecturer from Department of Statistics, Faculty of Science and Engineering, Macquarie University, Australia for editing my English writing of second manuscript. I would also like to mention about my friends for best memories, their best laughs, their jokes, and their joint activities together that help me to be relax and happy during my study.

Finally, I would like to say a huge thanks to my family for their encouragement in all aspects, their praying, and their understanding throughout this study.

Mayuening Eso

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

In this chapter we describe the background, rationale, objective and literature review of the present study.

1.1 Background

According to Leewatchanaku (2000) and Wangwongwiroj (2008), and the reports from Thai Meteorological Department (TMD) (2012), Thailand is located in the tropical area between latitudes 5°37' N to 20°27' N and longitudes 97°22' E to 105°37' E, covering an area of about 513,115 square kilometers or around 200,000 square miles. Thailand is bordered by Myanmar and Laos on the north, by Laos, Cambodia and the Gulf of Thailand on the east, by Malaysia on the south and by Myanmar and the Andaman Sea on the west.

Climates of Thailand including rainfall are under the influence of southwest monsoon and northeast monsoon. The southwest monsoon brings a stream of warm moist air from the Indian Ocean in mid-May to mid-October towards Thailand, and causes abundant rain over the country, especially on the windward side of the mountains. The northeast monsoon brings considerable rain along the eastern coast of southeastern Thailand from mid-October until mid-February. The northeast monsoon also brings the cold and dry air from anticyclone in mainland China over major parts of Thailand, especially around the northern and northeastern parts which are higher latitude areas. These particular areas usually experience dry spell because of this monsoon.

In addition, Thailand gets rain from other winds such as trade winds starting from mid-February to mid-April, and cyclones that come from the Pacific Ocean, South China Sea, Bay of Bengal and Andaman Sea. Typically, the cyclones occur through Thailand on an average three or four times per year, normally from April until December as shown in Figure 1.1.

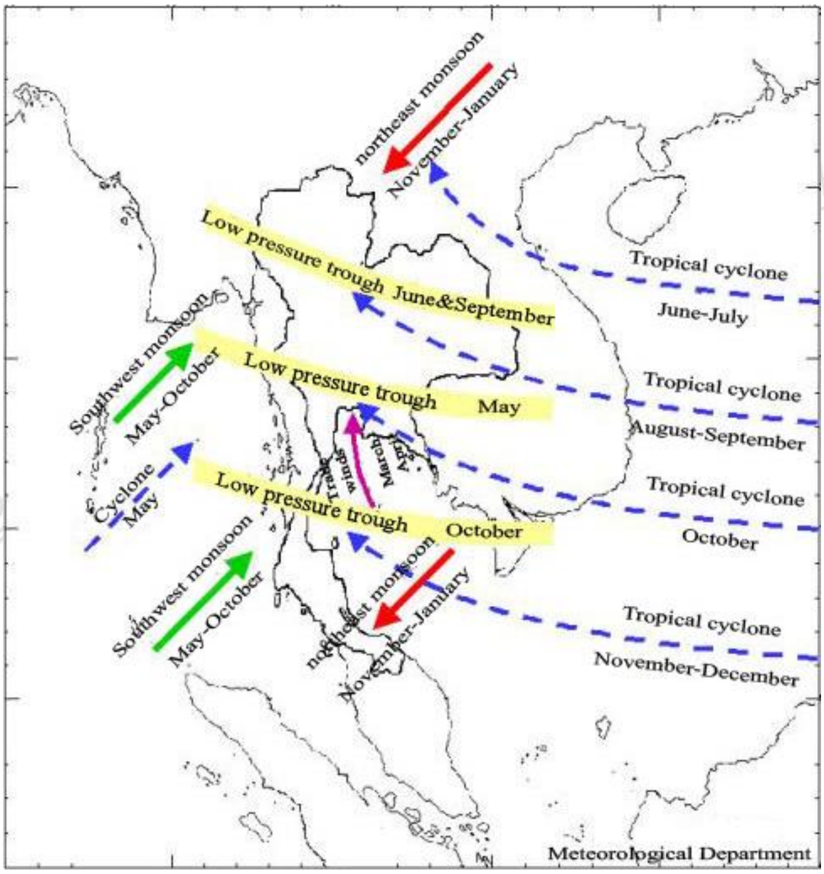

Figure 1.1: Direction of monsoon through Thailand (Source: This map from the book of *Hydrology*, page 61 (Wangwongwiroj, 2008))

The climate of Thailand can be divided into three seasons. The first season is rainy or southwest monsoon season starting from mid-May to mid-October. The wettest period of the year is from August to September in which some areas are probably flooded. Except southeastern Thailand, abundant rain remains until the end of the year while November is the wettest month. The winter or northeast monsoon season, the second season, starts from mid-October to mid-February. As a result, it is quite cold in upper part of Thailand in December and January. In contrast, there is a great amount of rainfall in southeastern Thailand, especially during October to November. The last season is summer or pre-monsoon starting from mid-February to mid-May. In this season, the weather becomes warmer, especially in the upper part of Thailand. April is the hottest month. The example of annual rainfall and the number of rainy days from the year 1971 to 2000 in each season are shown in Table 1.1.

Table 1.1: Seasonal rainfall in Thailand during 1971-2000

Regions	Average rainfall (mm)			Annual rainy days
	Winter	Summer	Rainy	
North	105.5	182.5	952.1	123
North-east	71.9	214.2	1,085.8	117
Central	124.4	187.1	903.3	113
East	187.9	250.9	1,417.6	131
South-east	759.3	249.6	707.3	148
South-west	445.9	383.7	1,895.7	176

Moreover, rainfall can be classified into five groups. It is called the trace day when the rainfall is less than 0.1 millimeters/day (mm/day), light rain when rainfall is between 0.1-10.0 mm/day, moderate rain when rainfall is between 10.1-35.0 mm/day, heavy rain when rainfall is between 35.1-90.0 mm/day and very heavy rain when rainfall is more than 90.1 mm/day.

However, total annual rainfall in most areas of the country is around 1,564.8 mm. The areas in upper part of Thailand usually have heavy rain in August or in September. Almost all of the areas, especially, in the west and east of the country get lots of rain because they are located in front of a mountain or on the windward side of the southwest monsoon, particularly KlongYai district, Trat province, in which the total annual rainfall is more than 4,000 mm. Whereas, areas with less rain are located on the leeward side of mountain, such as north-central, central and western areas of the northeast. For the southwest of Thailand, the areas receive heavy rain throughout the year in which the heaviest rainfall is in September. For the southeast, most areas are on the windward side of the northeast monsoon, and thus receive more rain than the southwest part. The heaviest rainfall in this part appears in November.

The rainfall in Thailand might be related to the phenomenon of El Niño Southern Oscillation (ENSO) as the study of Singhrattna et al. (2005) found that Southern Oscillation Index (SOI) which is one of the indicators of ENSO was a significant relationship with the summer rainfall during the post-1980 period (the years of the study were from 1951 to 2001). Some studies investigated the correlation between ENSO and rainfall in the countries which are in the tropical and sub-tropical zone such as Indonesia (Quinn et al., 1978), Australia (Chiew et al., 1998; Chambers, 2003), Brazil (Stoeckenius, 1981; Rao and Hada, 1990), China (Renhe et al., 1999), northeastern South America and southeastern Africa (Ropelewski and Halpert, 1987). In the next section, we further describe what the ENSO phenomenon is, and how it is correlated to El Niño, La Niña and SOI.

The SOI is a large-scale fluctuation of the Mean Sea Level Pressure (MSLP) in the tropical Pacific between Tahiti (12°27' S, 130°50' E), French Polynesia and Darwin (17°40' N, 149°25' E), Australia. This value gives an indication of the development and intensity of El Niño or La Niña events in the Pacific Ocean. Prolonged periods of negative SOI values below -8 called El Niño events where air pressure is below-normal at Tahiti and above-normal at Darwin. This situation refers to sea surface temperatures in the central and eastern tropical Pacific Ocean significantly warmer than normal, appearing off the coasts of Peru and Ecuador, with a decrease in the strength of the Pacific trade winds. The opposite condition occurs when there are prolonged periods with strong positive SOI values above +8 called La Niña events. This is associated with stronger Pacific trade winds and warmer sea temperatures to the north of Australia. The water in the central and eastern tropical Pacific Ocean become cooler during this time (see, Figure 1.2). These events occur in every three to eight years as part of a natural occurring cycle. Alternatively, the events may be called warm and cold ENSO events for El Niño and La Niña, respectively (Australian Government Bureau of Meteorology, 2005).

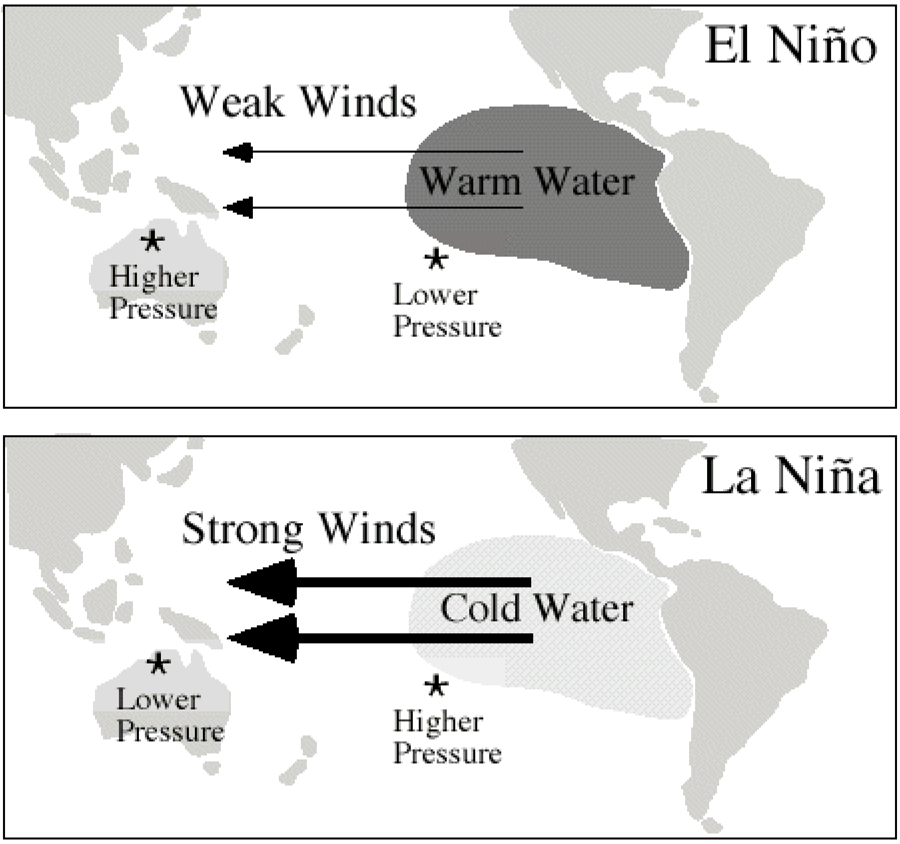

Figure 1.2: Situation of occurring ENSO, asterisks represent the locations of Tahiti, French Polynesia and Darwin, Australia (Source: This picture from the website "http://www.worldclimatereport.com/archive/previous_issues/vol1/v1n24/feature.htm")

1.2 Rational

Rainfall is a major source of the input in the water resource system. Rainfall is also an important factor in many fields, such as agriculture, industrial, livestock, woodland, and ecosystem, etc., but excessive rainfall may cause natural disasters. Due to continuous and excessive rainfall, Thailand has been badly experienced many times with landslides and floods, for instance, the severe flooding in Hat Yai 2010 and in Bangkok 2011. These contributed to costly damage of assets and loss of life. After the floodwater receded, millions of dollars were spent for recovery, reconstruction and rehabilitation of the overall infrastructure such as agriculture, farm, road, bridge and energy, etc. Similarly, the problem of both physical and mental health had to be suddenly mitigated for the people who were affected by flooding (Assanangkornchai et al., 2004; The World Bank, 2012).

Unexpected heavy rains have been observed in some years, and it may be caused by the phenomenon of the ENSO which already has been highlighted in the Introduction section. As the phenomenon can be indicated by SOI data, Figure 1.3 shows the cumulative monthly SOI from which Watts (2010) has deduced that “...from 1920 went into a long La Niña-dominate trend that ended with the Great Pacific Climate Shift of 1976” and “The subsequent El Niño-dominated trend from 1976 to 1995 was almost three times as fast as the rise”.

As of the impact of El Niño, prolonged dry periods and bush fires may occur in northern Australia, Indonesia, Philippines and South Africa. Likewise, heavy rainfall, sometimes related to extensive flooding, may occur over parts of Peru, Ecuador, and areas around the Gulf of Mexico. The impact of La Niña is a tendency to cause heavy rainfall and flooding in Australia, Indonesia, Philippines and southern Africa. It also causes of drought in East Africa and southern region of South America.

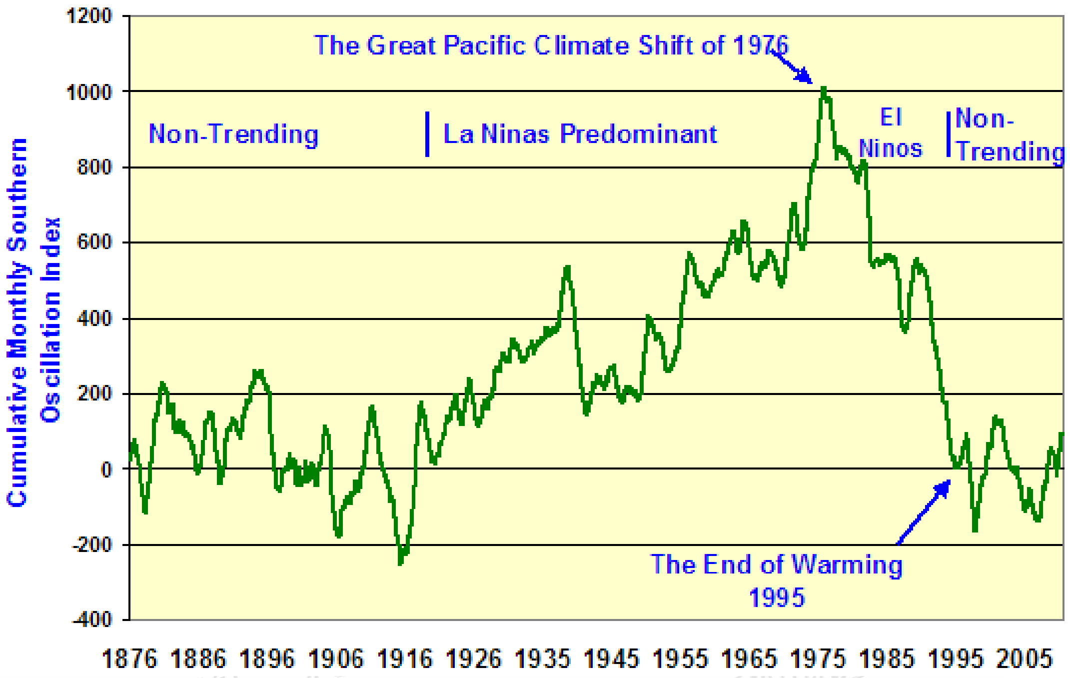

Figure 1.3: Cumulative monthly SOI from 1876-2010 (Source: This picture from the website “http://wattsupwiththat.com/2010/12/09/the-story-told-by-the-southern-oscillation-index”)

However, the rainfall in each region of Thailand fluctuates by season and topography as mentioned in the section of Introduction. It is important to understand the distribution of rainfall and its characteristics, and fit an appropriate statistical model to see the pattern of rainfall. Moreover, statistical analysis of rainfall would give important resources to manage and improve the existing water regimes which would be cost effective utilization of water resources. The information also will be a valuable resource to prevent floods and droughts, or develop better drainage system in the country.

In this study, we examine a distribution of daily rainfall during the period of 2001 to 2012 from various weather stations over Thailand using statistical methods. We also use the SOI data between 1876 and 2014 to explore characteristics of the data by using appropriate scientific methods. We further analyze the correlation between the daily rainfall and the SOI. Graphical methods are thereafter used to disseminate the results.

1.3 Objectives

1. To classify daily rainfall regions during 2001 to 2012 throughout Thailand.
2. To examine the patterns of daily rainfall during 2001 to 2012 throughout Thailand.
3. To examine the correlation between daily rainfall in Thailand and the Southern Oscillation Index during 2001 to 2012.
4. To investigate the characteristics of the Southern Oscillation Index during 1876 to 2014.

1.4 Literature review

1.4.1 Rainfall region classification

There are various studies on how to analyze the data with large variables for minimizing any loss of information and easy interpretation of the data. Principal component analysis (PCA), cluster analysis and factor analysis were used to classify the climatic variables. For example of the studies using PCA, Lee (2005) studied the annual rainfall distribution characteristic of Chia-Nan plain area in Southern Taiwan using annual rainfall of 178 stations from the period of 1978 to 2000. Two factors that were the upper and lower portion of this area were produced by PCA with explanation of 73% of the total variance. Comrie and Glenn (1998) determined rainfall regions of the United States-Mexico border region based on seasonality and variability of monthly rainfall at 309 stations from the period of 1961 to 1990. The study retained nine regions of those stations, explaining 65% of the variance in the data using PCA.

For the studies using factor analysis, Fotiadi et al. (1999) applied factor analysis to mean monthly precipitation from 54 stations from the period of at least 20 years in northwest Greece. The 12 calendar months were classified into two groups (cool and warm period) with distinctive spatial distributions of rainfall, explaining 88.5% of the total variance. Nikolakis (2008) used mean monthly precipitation totals from 63 stations: 46 stations over 89 years (1911-2000) and 17 stations over 59 years (1911-1970), in Cyprus. Factor analysis was used to classify the spatial distribution of precipitation. The 12 months of a year were classified in two groups, composing the wet and dry seasons, and corresponding to two statistically significant factors, which explained 92.6% of the total variance.

A combination of two methods is employed by some researchers. For example, Hussain and Lee (2009) used factor analysis and cluster analysis to classify regions in Pakistan with a technique involving 10-days rainfall parameter. The data comprised the daily rainfall at 32 stations, over 27 years for the period 1980 to 2006. The results from factor analysis consisted of three factors and the total variance for these factors were 94.6%. The first factor related with early autumn, winter and late spring. The second factor was summer monsoon, and the third factor showed deep relationship with winter rainfall. The result of cluster analysis was applied on those factors to divide the study area into six regions. Puvaneswaran (1990) worked on 28 climatic variables with 113 stations from 1931-1960 in Queensland, Australia. Factor analysis can divide the variables to three factors namely a humidity factor, a temperature factor and rainfall factor with 65% of total variance. Then, from those results, the stations could be divided into three major homogeneous climatic regions and 14 sub climatic regions by cluster analysis. The final example, the grouping of stations with similar characteristics and recognition of climatic regions using PCA and cluster analysis were applied to two sets of precipitation data in Switzerland, with one containing 47 stations during the year 1961 to 1980, and the other comprising 101 stations during the year 1981 to 1993. For the first data set, the PCA resulted in 5 factors accounting for 84% of the total variance. From the five factors, they re-classified again using cluster analysis that divided the study areas into seven zones. For the second data set, the PCA reduced the number of variables from 101 to 10, then re-grouping the 10 components by cluster analysis into 13 regions (Baeriswyl and Rebetez, 1997).

1.4.2 Rainfall pattern

Many studies have explored the situation of rainfall in several places. Some of them used non-parametric Mann-Kendall test, such as Taxak et al. (2014). The study was to assess the trends of annual and seasonal rainfall data. It was found that there was an increasing rainfall trend in Wainganga river basin located in Central India during the period 1901-1948, and there was a decreasing rainfall trend in the basin during the period 1949-2012. Likewise, Gandomkar et al. (2011) found that mean annual rainfall had a negative trend in Sefid-roud basin between 1956 and 2005. Also, in almost gauged stations in the Mediterranean areas, annual and seasonal rainfall appeared to be a predominantly negative trend between 1918 and 1999 (Longobardi, 2009). Keggenhoff (2014) had concluded that almost all extreme rainfall indices showed an increase trend, but the changes were not statistically significant. While, the total rainfall increased from 1971-2010 in Georgia.

Furthermore, there are some studies that used the difference methods which are explained as follows. Feng et al. (2007) reported that the decreasing trends in extreme rainfall events were mainly observed in northern China, and increasing trends were observed in the Yangtze River basin and in northwestern China in the period of 1951 to 2000. They employed annual maxima of daily, 2-day, 5-day and 10-day rainfall with the technique of generalized extreme value (GEV) distribution. In the southern South America, Castañeda and González (2008) used MOD-T principal component analysis, T-Student and Mann-Kendall test applied to monthly and annual rainfall in the period of 1950-1999. The results showed positive trends in north and south of Patagonia, while there was a decrease in the western and central zone of Argentinean Patagonia. Using Kendall-tau test, there was statistically significant increasing in intensity and frequency of extreme daily rainfall events at the stations located in the Typhoon belt of Philippines in the period of 1951-2010, compared to the mean 1961-1990 baseline values, and there was no station showing a decreasing trend (Cinco et al., 2014). For the last example, Doty (1982) applied a simple linear regression technique to annual rainfall. The result suggested that the rainfall in the windward side of the island of Hawaii during 1900 to 1977 showed a general downward trend and in the southeast side of the island a general upward trend.

It can be seen from previous studies that most of the studies used non-parametric methods to explore daily rainfall trends. However, in order to explore the distribution of rainfall, some other methods were used. For examples, Aksoy (2000) expressed that gamma distribution fitted very well to 30-year daily rainfall series in the Asian part of Istanbul. Ejieji (2004) showed that the results from gamma distribution gave better fit to observed data than exponential distribution for almost daily rainfall stations in Nigeria. The distribution of rainfall data from the Tropical Rainfall Measuring Mission (TRMM) satellite were examined to show that gamma fitted better than lognormal fitted in wet regions (Cho et al., 2004).

1.4.3 Rainfall and the SOI

Earlier work demonstrated that the ENSO was responsible for rainfall in various countries by using the SOI as one of the indicators to measure the ENSO. For example, Bhalme and Jadhav (1984) reported that the Indian monsoon rainfall in the period 1875 to 1980 was significantly correlated with the SOI using linear correlation coefficients. The all drought years in India were the simultaneous occurrence with low the SOI as well as all the flood years were showing when high the SOI happened. Hadiani et al. (2012) applied the Pearson correlation coefficient and analysis of nonlinear regression variance to evaluate the relation between SOI and rainfall. They found that the SOI from winter until the beginning of spring in the area of Pacific Ocean was effective on rainfalls from 1971 to 2005 of late spring and the beginning of winter in the north of Iran. In eastern Australia, the phase representing consistently positive SOI generally corresponded with above median rainfall amounts, while the phase representing consistently negative SOI corresponded with below median rainfall amounts (Stone, 1992). The study also used the Kruskal-Wallis test methods to explore the different rainfall probability distributions in each phase of the SOI. In Sri Lanka, the correlation coefficients between the rainfall (1881-1980) and SOI were positive and significant in the first intermonsoon season (March to April) and the southwest monsoon season (May to September). Cumulative rainfall of the peak monsoon months in July and August revealed a strong and positive correlation with the seasonal SOI. The rainfall in the second intermonsoon season (October to November) and the SOI had negative correlation as well as in the northeast monsoon season (December to February). However, they were weak correlation in the northeast monsoon season. Those correlations were identified by cross-correlation method (Suppiah, 1989). Tigona and de Freitas (2012) had showed that the La Niña conditions corresponded with high average rainfall (1971- 2009), and the El Niño conditions coincided with drier rainfall in Vanuatu region of the tropical South Pacific. The strength of relationships between the SOI and rainfall depended on the time periods used, and depended on the particular geographic region of Vanuatu. The study used linear regression and Pearson correlation to examine those correlations. For the last example, Xu, et al. (2004) showed that the El Niño events significantly reduced the precipitation in southern part of Australia. In contrast, the La Niña events increased the precipitation in that particular area. This was also true in Indonesia, Malaysia, Philippines, Thailand and northern part of Australia, although the tendency was not as strong as of southern part of Australia. However, the correlation of those events were not observed in China, Japan and Korea, as the study used of Pearson correlation coefficient and Mann-Kendell to test those correlations.

From the results of previous studies above, it should be noted that the positive SOI has a correlation to the high rainfall amount, and the negative SOI has a correlation to the low rainfall amount. Those correlations are observed in the areas around the western equatorial Pacific.

1.4.4 Rainfall in Thailand

As a part of rainfall investigation in Thailand, from the past there were the variety of data formats, regions and methods as the following examples. To see the distribution of rainfall, Phien et al. (1980) fitted the distribution of the monthly rainfall 20 years in northeast of Thailand for 56 stations using the method of leakage law with data included zero, and using lognormal and gamma distributions for the data without zero values. Those models fitted the data very well. Szyniszewska and Waylen (2012) analyzed daily rainfall in order to establish the relationship to monthly rainfall in central and northeastern Thailand. The study used the data from 17 stations with at least 30 years of rainfall recorded from 1951-2006 and 1970-2006. Markov process and transitions probability were used to find the probability of daily rainfall occurrence (rain or no rain) for each month. Gamma distribution was then used to calculate the amount of rainfall on wet days. It was concluded that the majority of months the mean and standard deviation of daily rainfall magnitudes increased steadily with increasing monthly rainfall. The gamma distribution described the magnitudes of daily rainfall well.

In order to forecast rainfall in Thailand, Singhrattna et al (2005) developed a statistical forecasting method for summer monsoon rainfall over Thailand using linear regression and a local polynomial-based nonparametric method. The study used the rainfall data of August to October from 1951 to 2001. The results showed that the predictor variables, such as sea-surface temperature (SST) and sea level pressure (SLP) in the Indo-Pacific region and also ENSO indices (i.e. SOI) had a significant relationship with the summer rainfall only during the post-1980 period. It was concluded that the two models could predict the rainfall at 1-3 month lead time. Hung et al. (2009) used an Artificial Neural Network (ANN) technique to improve rainfall forecast performance in Bangkok. The hourly data were used from 75 rain gauge stations from 1997 to 1999, and the data of the year 2003 were used as a testing set. The results showed that the ANN model could predict the rainfall in Bangkok with highly satisfactory from 1 to 3 ahead. Weesakul (2005) developed a mathematical model for forecasting annual rainfall in Thailand. ARMA and ARIMA models were used to fit the time series of annual rainfall during 1951 to 1990 of 3l rainfall stations distributed in all regions of Thailand. It was shown that ARIMA model was able to forecast annual rainfall for all regions of Thailand.

1.5 Conceptual framework

The conceptual framework of dissertation is shown in Figure 1.4.

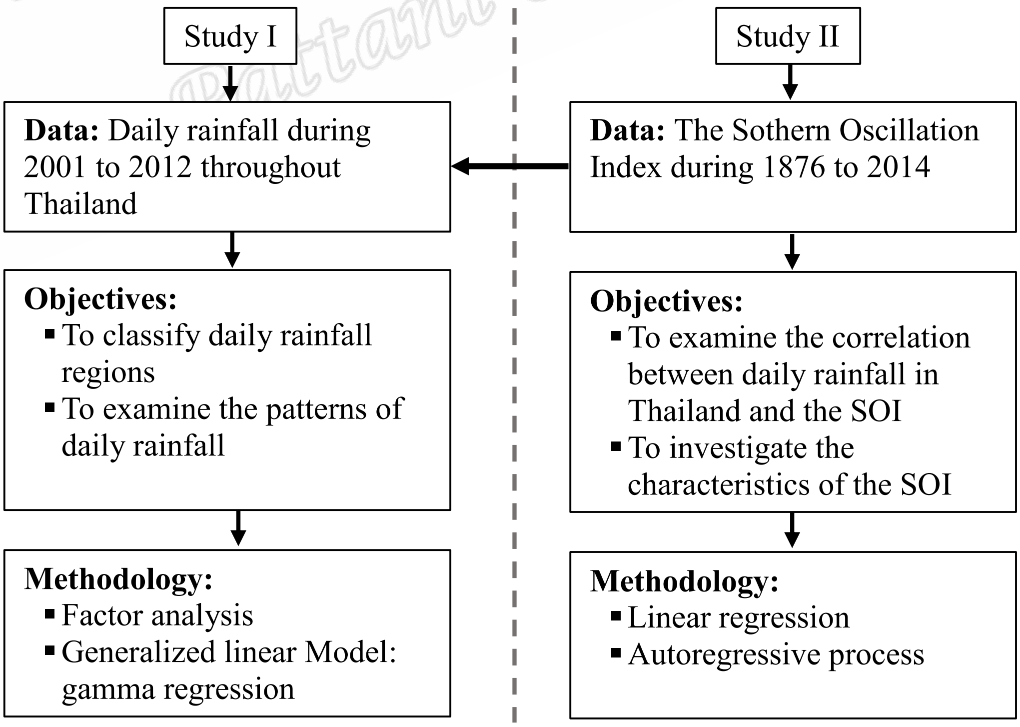

Figure 1.4: Conceptual framework of dissertation

The present study uses two different datasets that are daily rainfall during 2001 to 2012 throughout Thailand and the Sothern Oscillation Index during 1876 to 2014. The first study, the rainfall data are revealed with two objectives which are to classify and examine daily rainfall using factor analysis and generalize linear model with gamma regression. The second study, the SOI data also are presented with two objectives which are to examine the correlation between daily rainfall in Thailand and the SOI and to investigate the characteristics of the SOI using linear regression and autoregressive process.

In summary, from the previous studies, three methods were widely used to classify the climatic variables which were Principal Component Analysis (PCA), cluster analysis and factor analysis. In order to examine the pattern of rainfall, several methods were used by researchers, such as Mann-Kendall test, Kendall-tau test, simple linear regression and generalized extreme value (GEV) distribution. In addition, the majority of methods used to define the correlation between rainfall and the SOI was a correlation coefficient. In Thailand, there were a number of studies concerning rainfall in the country in several regions, different years, different formats of data set and using various methods such as Markov process, lognormal distribution, gamma distribution, linear regression, a local polynomial-based nonparametric method, Artificial Neural Network, ARMA model and ARIMA model.

In the present study, we have also used the existing statistical methods to fit the daily rainfall data across Thailand and the SOI. The data sources and methodologies are presented in the next chapter.

CHAPTER 2

Methodology

Two different datasets from two studies are contained in this chapter. The first study is concerned with daily rainfall data using the approach of factor analysis and gamma-distributed generalized linear model. The second study is concerned with Southern Oscillation Index (SOI) data using linear regression and autoregressive model.

2.1 Daily rainfall data

2.1.1 Data source and variables

The secondary data used in this study are provided by Thai Meteorology Department (TMD). Daily rainfall amounts are measured every three hours starting from 07.00 a.m. until 07.00 a.m. of the following day. The total rainfall in millimeters (mm) for one day (24 hours) throughout Thailand are available on the website of TMD for three previous months. The rainfall patterns across Thailand are presented by 122 weather stations. The number of weather stations in the North, North-east, East, Central, South-east and South-west regions are 29, 27, 15, 20, 23 and 8, respectively (the locations of the stations are shown in Appendix I).

The data are collected from 1st January, 2001 to 7th February, 2013. The report consisting of 54 stations with missing values is as shown in Table 2.1. A criterion is set up to overcome the stations with a lot of missing values. The stations which have more than 13% of missing values are removed from the study. The 13% is the appropriate percentage to be the cut point for handling the stations that contain a lot of missing value of this dataset. Such eight stations which belong to this criteria are Bang Khen, Burirum, Khanom, Krabi, Mae Jo, Nongbualumphu, Sukhothai and Suwanabhum Airport. The 32 stations are observed to have incomplete data for January 2001, thus this month is also excluded from the analysis. The rain falls on 29th February for each leap year, 2004, 2008 and 2012, are simply omitted. Therefore, data from 5th February, 2001 to 4th February, 2013 for 114 stations are used in the study.

Table 2.1: Stations with missing values

Stations	%	Stations	%
Bang.Khen	100.00	Kamphaeng.Saen.Agromet	0.70
Khanom	90.41	Loei.Agromet	0.70
Nongbualumphu	66.09	Nan.Agromet	0.70
Suwanabhum.Airport	54.47	Nong.Phlup.Agromet	0.70
Krabi	44.06	Nakhon.Phanom.Agromet	0.70
Mae.Jo	43.32	Nakhorn.Sri.Thammarat.Agromet	0.70
Burirum	40.60	Pak.Chong.Agromet	0.70
Sukhothai	40.60	Phriu.Agromet	0.70
Doi.Muser.Agromet.Stn	12.64	Pichit.Agromet	0.70
Lampang.Agromet	12.60	Ratchaburi	0.70
Si.Samrong.Agromet	12.37	Roi.Et.Agromet	0.70
Takua.Pa	10.54	Sawi.Agromet	0.70
Sattahip	8.37	Sakon.Nakhon.Agromet	0.70
Donmuang	3.19	Surat.Thani	0.70
Huai.Pong.Agromet	2.76	Surat.Thani.Agromet	0.70
Ko.Lanta	2.01	Surin.Agromet	0.70
Pathumthani.Agromet	1.67	Tak.Fa.Agromet	0.70
Phatthalung.Agromet	1.40	Tha.Phra.Agromet	0.70
Chon.Buri	1.27	Ubon.Ratchathani.Agromet	0.70
Chiang.Rai.Agromet	0.95	U.Thong.Agromet	0.70
Kabin.Buri	0.86	Yala.Agromet	0.70
Si.Sa.Ket.Agromet	0.84	Phra.Sang	0.52
Ayuttaya.Agromet	0.70	Pilot.Station	0.36
Bang.Na	0.70	Phatthaya	0.32
Chacherngsao.Agromet	0.70	Lam.Chabang	0.27
Chai.Nat	0.70	Aranyaprathet	0.05
Kho.Hong.Agromet	0.70	Bangkok.Metropolis	0.05

The spatial distribution of the 114 stations is shown in Figure 2.1. From this dataset, it is found that 11 provinces have no stations namely Amnat Charoen, Yasothon, Saraburi, Ang Thong, Sing Buri, Nonthaburi, Uthai Thani, Samut Sakhon, Samut Songkhram, Nakhon Nayok and Nong Bua Lam Phu. Provinces have different numbers of stations, for instance, there are 35 provinces with only one station, 19 provinces with two stations, five provinces with three stations, four provinces with four stations and two provinces with five stations.

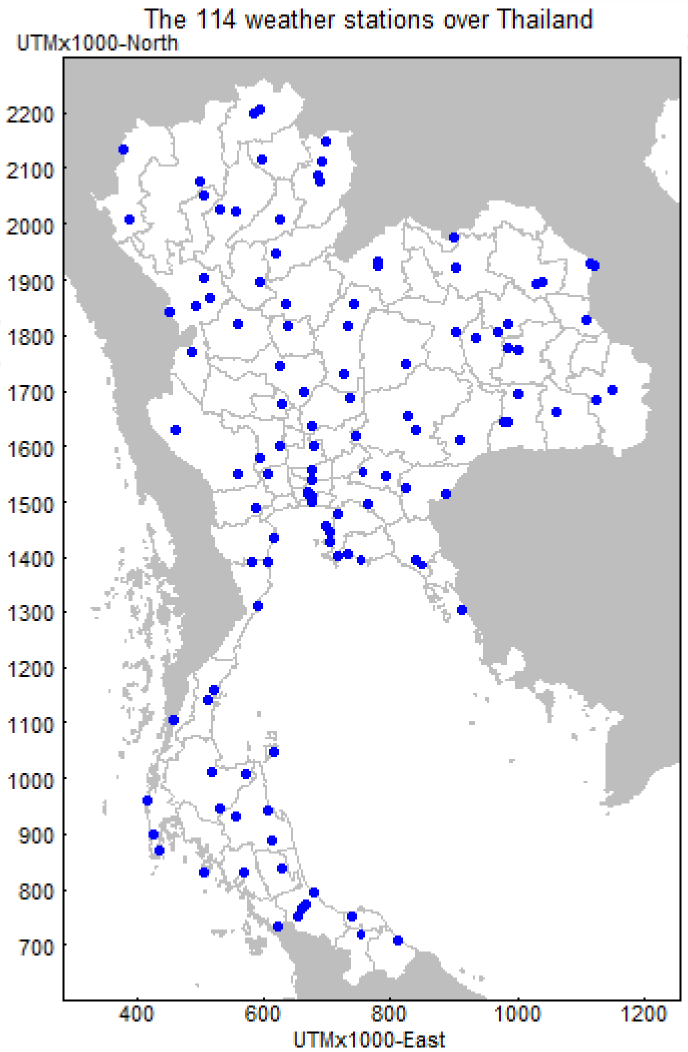

Figure 2.1: Weather stations across Thailand

The data structure consists of the variables year, month, day and total rain in 24 hours for 114 stations as shown in Table 2.2.

Table 2.2: Data structure of daily rainfall

no	year	month	day	Mae.Hong.Son	Mae.Sariang	…	Narathiwat
1	2001	1	1	0	0		0
2	2001	1	2	0	0		2.4
3	2001	1	3	0	0		3.1
⋮	⋮	⋮	⋮	⋮	⋮		⋮
11316	2001	12	31	0	0		0.4
11317	2002	1	1	0	0		0
11318	2002	1	2	0	0		0
11319	2002	1	3	0	0		1.1
⋮	⋮	⋮	⋮	⋮	⋮		⋮
22540	2002	12	31	0	0		12.1
⋮	⋮	⋮	⋮	⋮	⋮		⋮
1108872	2012	1	1	0	0		11
1108873	2012	1	2	0	0		1.2
1108874	2012	1	3	6.4	0		0
⋮	⋮	⋮	⋮	⋮	⋮		⋮
1120186	2012	12	31	0	0	…	218.8

However, rainfall is time series data. The rainfalls on consecutive days are serially correlated and the conventional statistical models assume independent errors. Thus, we effectively reduce these serial correlations by restructuring the data as 5-day averages. Then, the data are reduced to 73 periods in each year**.**

Variables

The outcome of the first study is 5-day averages rainfall of 114 stations $(y_1, y_2, \dots, y_{114})$ from the year 2001-2012 and the determinants are year and month.

2.1.2 Statistical analysis

To start with, preliminary analysis is demonstrated to reveal the rough results using the descriptive statistics including mean, median, maximum and frequency. After that, statistical models are used to handle the data. Because the dataset has more than one response variable, it can be summarized by data matrices *y* with *n* observations and *p* variables called multivariate analysis. One of the widely used methods for multivariate analysis is factor analysis dividing the variables into groups.

Factor analysis

Factor analysis is a statistical method to provide linear combinations of the variables, called factors (Johnson and Wichern, 2007; Rencher, 2002; Venables and Ripley, 2002). In the study, this method is used to allocate the 114 stations to be a smaller number of groups that reduce the spatial correlation and contain the bulk of the relevant information in the data. Basically, this method maximize correlations between variables within factors and minimize correlations between variables in different factors. The model can be written as:

$$y_{ij} = \mu_j + \sum_{k=1}^{p} \lambda_{jk} f_{ik} + z_{ij} \qquad (1)$$

where y_{ij} is 5-day averages rainfall on date *i* at stations *j*, μ_j is the average of rainfall at stations *j*, *p* is the number of factors, f_{ik} is the factor score for factor *k* on date *i*, λ_{jk} is the factor loading for factor *k* at station *j* and z_{ij} is residual errors. The covariance matrix with Spearman correlation which is applied to positive skewed distribution of data is used to fit the factor model. To estimate factor loading scores, maximum likelihood and promax are used in the present study for factor extraction and factor rotation, respectively.

After getting the number of clusters from factor analysis, factor scores for each cluster are calculated using average rainfall amount from their membered stations. Then, each cluster has been analyzed using generalized linear model with gamma distributions.

Generalized linear model: gamma regression

Generalized linear models (GLMs) are based on the exponential family of distributions to extend linear models for accommodating non-normal response distributions and transformations to linearity. The outcome variable and linear predictors are linked by canonical link function that is presented in Table 2.3 (Gill, 2001; McCullagh and Nelder, 1989; Venables and Ripley, 2002).

Table 2.3: Canonical link functions are widely used for each exponential family

Family	Canonical link	Name	Variance
Binomial	$\log(\mu/(1-\mu))$	Logit	$\mu(1-\mu)$
Gamma	$-1/\mu$	Inverse	μ^2
Gaussian (Normal)	μ	Identity	1
Inverse Gaussian	$-2/\mu^2$	1/mu^2	μ^3
Poisson	$\log\mu$	log	μ
Source: Venables and Ripley (2002), page 185			

Gamma distribution is particularly useful to model the outcome with nonnegative value. The gamma distribution is most commonly written as

$$f(y|\alpha,\beta)=\frac{1}{\Gamma(\alpha)}\beta^{\alpha}y^{\alpha-1}e^{-\beta y},\ y,\alpha,\beta>0 \qquad (2)$$

where y is a random sample from gamma distribution, α is shape parameter, β is inverse-scale parameter and $\Gamma(\alpha)$ is the gamma function.

In the study, rainfall measurements have positively skewed distributions, even when aggregated into 5-day averages, and thus violate the statistical assumption of normally distributed errors due to the GLMs are built on the framework of the classic linear model. To account for such skewness, generalized linear model with gamma distributions is fitted. The model with inverse canonical link function takes the form

$$1/\text{mean}[R] = \text{constant - factor (year) - factor (month)} \quad (3)$$

where R is average daily rainfall in 5-day periods during rainy seasons, year starts from 2001-2012 and month has the value from 1 to 12.

Graphical methods

Graphical methods are quite useful to present the results from statistical process more clearly. In this study, graph of correlogram of autocorrelation function (ACF) (Chatfield, 1996) is used to check the serial correlations of the data. Bar plot and histogram are applied to explore the distribution of data. In order to display the Spearman correlation of 114 stations, bubble matrix chart is plotted among the stations. For checking the goodness of fit of models, normal quantile plot of deviance residuals is used to evaluate models. The graph of 95% confidence intervals (95% CIs) can be used to show the pattern and extent of average daily rainfall. Furthermore, the confidence intervals are based on standard errors to present the differences between the average rainfall and its overall mean by method of sum contrasts (Tongkumchum and McNeil, 2009).

Diagram of data analysis

The first study can be summarized by the diagram below.

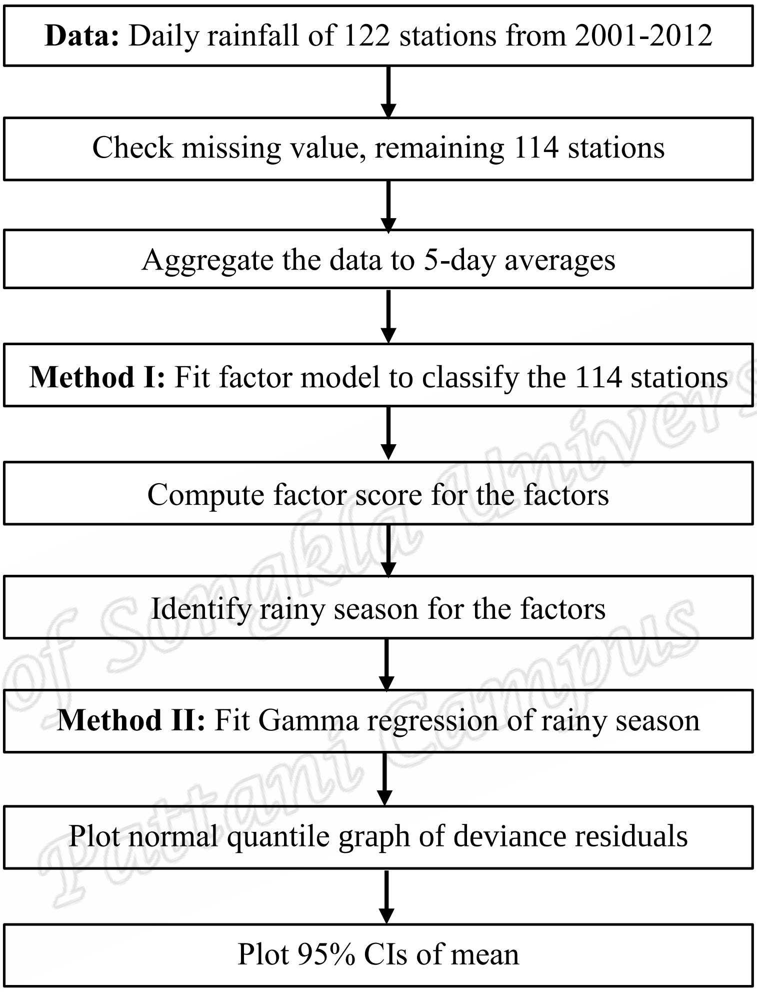

Figure 2.2: Diagram of rainfall analysis

Factor analysis fits the 5-day average rainfall to allocate the stations to be a smaller group. Then, factor scores of each group are computed by finding average daily rainfall from the stations located in the same group. Bar plots of the groups are demonstrated to define their rainy season. Next, Gamma regression is used to fit the data for the groups with the outcome is 5-day averages daily rainfall during rainy season and the determinants are year and month. Finally, the graphs of normal quantile of deviance residuals for evaluating the models and 95% CIs for presenting year and month effects are plotted.

2.2 The Southern Oscillation Index (SOI)

2.2.1 Data source and variables

The Australian Government Bureau of Meteorology (BOM), National Climate Centre, Climate Analysis Section provides the Monthly average SOI data from 1876 to 2014, totalizing 139 years with 1,664 observations. The data can be downloaded from "http://www.bom.gov.au/climate/current/soihtm1.shtml". Table 2.4 shows the data structure which contains the SOI value in particular months and years. The SOI values can be calculated as:

$$\mathrm{SOI} = 10 \times \frac{[\mathrm{P_{diff}} - \mathrm{P_{diffav}}]}{\mathrm{SD(P_{diff})}} \tag{4}$$

where P_{diff} is the average of Mean Sea Level Pressure (MSLP) at Tahiti for the particular month minus the average of MSLP at Darwin for the particular month. P_{diffav} is the long term average of P_{diff} for the particular month. $SD(P_{diff})$ is the long term standard deviation of P_{diff} for the particular month. The dataset that BOM uses based on means and standard deviations calculated over the period 1933 to 1992. The multiplication by 10 is a convention, and using this scale factor the SOI ranges from about -35 to +35.

Table 2.4: Data structure of the SOI

Year	Jan	Feb	Mar	Apr	⋯	Sep	Oct	Nov	Dec
1876	11.3	11	0.2	9.4	⋯	10.5	-8	-2.7	-3
1877	-9.7	-6.5	-4.7	-9.6	⋯	-17.2	-16	-12.6	-12.6
1878	-8.7	-21.1	-15.5	-8.8	⋯	17.7	10.9	15.1	17.9
⋮	⋮	⋮	⋮	⋮	⋯	⋮	⋮	⋮	⋮
2014	12.2	-1.3	-13.3	8.6	⋯	-7.5	-8	-10	-5.5

Variables

The outcome of the second study is monthly average SOI in the period of 1876 to 2014 and the determinant is boxcar function.

2.2.2 Statistical analysis

The second study attempts to describe the variation of the SOI using appropriate scientific approach which involves fitting an appropriate model to the data. The methods suggest that the models have result in a p-value. The p-value is the probability that a data confuguration is at least as unusual as that observed could have arisen purely by chance, that is assuming that the null hypothesis is true. By convention, p-values smaller than 0.05 provide sufficient evidence to reject the null hypothesis. Incidentally, the null hypothesis of the study is the fluctuations in the SOI are random. However, the methods that are recommended by the present study are linear regression then followed by autoregssive model. In addition, Analysis of Varince (ANOVA) (Schumacker, 2015) is used to confirm the p-value for testing the null hypothesis after applying the two methods.

Linear model

Firstly, a linear regression model that assumes independent errors taking the formulation below (Eqn 5) is used to fit the data (Faraway, 2005; Venables and Ripley, 2002)

$$y_t = b_0 + b_i x_{it} + z_t \,. \qquad (5)$$

In this model y_t represents the SOI for month *t* where *t* equals 1 for January 1876, while x_{it} represents boxcar function (see Weisstein, 2002; Weisstein, 2015) taking the values of 1 on interested intervals and 0 elsewhere (Figure 2.3). The function of boxcar fuction is

$$\prod_{a,b}(x) = H(x-a) - H(x-b)\,, \qquad (6)$$

where $\prod_{a,b}(x)$ is equal 1 for $a \le x \le b$ and 0 otherwise and *H*(*x*) is the Heaviside step function (see Weisstein, 2002; Weisstein, 2015).

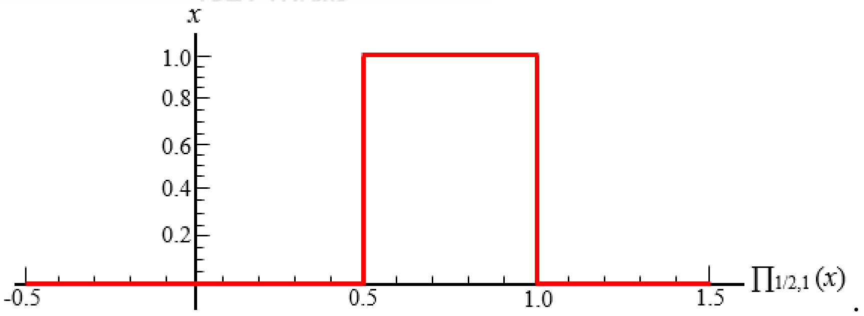

Figure 2.3: Boxcar function (Source: The original graph from the website "http://mathworld.wolfram.com/BoxcarFunction.html")

The b_0 and b_i are parameters for estimating mean and slope, respectively. The z_t constitutes successive values which may be a series of auto-correlated normally-distributed errors (noise). The boxcar function is treated like a dummy variable. A number of the dummy variable is equal to a number of level of covariate variable.

For each the dummy variable, the component of the variable takes the value 1 if the observation of the covariate variable at the levels has interested values and 0 otherwise (McCullagh, P. and Nelder, 1989). Similarly, the boxcar function is defined as 1 in the period of interested time and 0 elsewhere.

Autoregressive model

Secondly, for a time series, the residuals z_t are often called the noise, and the model which does not include the noise is called the signal. If the residuals arise from uncorrelated errors, the noise is called white noise, otherwise it is coloured noise. A commonly implemented model for the coloured noise follows an autoregressive (AR) process. The AR model takes the form (Eqn 7)

$$z_t = \sum_i a_i z_{t-s_i} + w_t \qquad (7)$$

where w_t is white noise. This means that each value of z_t is expressed as a linear combination of previous values at specified lags (s_1, s_2, s_3, etc.) plus an independent white noise series (Chatfield, 1996). Finally, the correlogram of autocorrelation function and partial autocorrelation function are plotted to check those white noise.

Autocorrelation function

Autocorrelation function (ACF) is a measured the correlation between observations in the same time series dataset at different points of time lag (*s*). The correlation coefficient (r_s) between observations with a distance *s* apart can be computed from

$$r_s = \frac{\sum_{t=1}^{n-s} (Y_t - \overline{Y})(Y_{t+s} - \overline{Y})}{\sum_{t=1}^{n} (Y_t - \overline{Y})^2}, s = 1, 2, 3, \ldots \quad (8)$$

where Y_t is the observed value at time *t*, *n* is a number of sample size, $\overline{Y}$ is the average of observation. The coeffient of r_s should be in the intervals of -1 to 1. When $|r_s|$ is close to zero, there is less correlation of the observations at different time *t*, in other word the observations in a dataset are random (Chatfield, 1996).

Partial autocorrelation function

Partial autocorrelation function (PACF) is commonly used for identifying the order of an autoregressive model. PACF is a measured the correlation between the observations with a distance *s* apart by controlling the other lags to be constant values. The correlation coefficient (r_{ss}) of PACF can be computed from

$$r_{ss} = \begin{cases} r_{ss} = r_1 & s = 1 \\ \dfrac{r_s - \sum_{j=1}^{s-1} r_{s-1,j} r_{s-j}}{1 - \sum_{j=1}^{s-1} r_{s-1,j} r_j} & s = 2, 3, 4, \ldots \end{cases} \quad (9)$$

where $r_{s,j} = r_{s-1,j} - r_{ss} \times r_{s-1,s-j}$, *s* is lagged terms.

In addition, a set of correlation coefficients r_s (or r_{ss}) can be exhibited by the correlogram graph (Chatfield, 1996) in order to show more clearly that those values are close to zero. The correlogram is plotted by r_s (or r_{ss}) against the lags *s*. The values are expected to present in the intervals of $\pm 2/\sqrt{n}$ with the statistical significance level of 0.05.

Diagram of data analysis

The second study can be summarized by the diagram below.

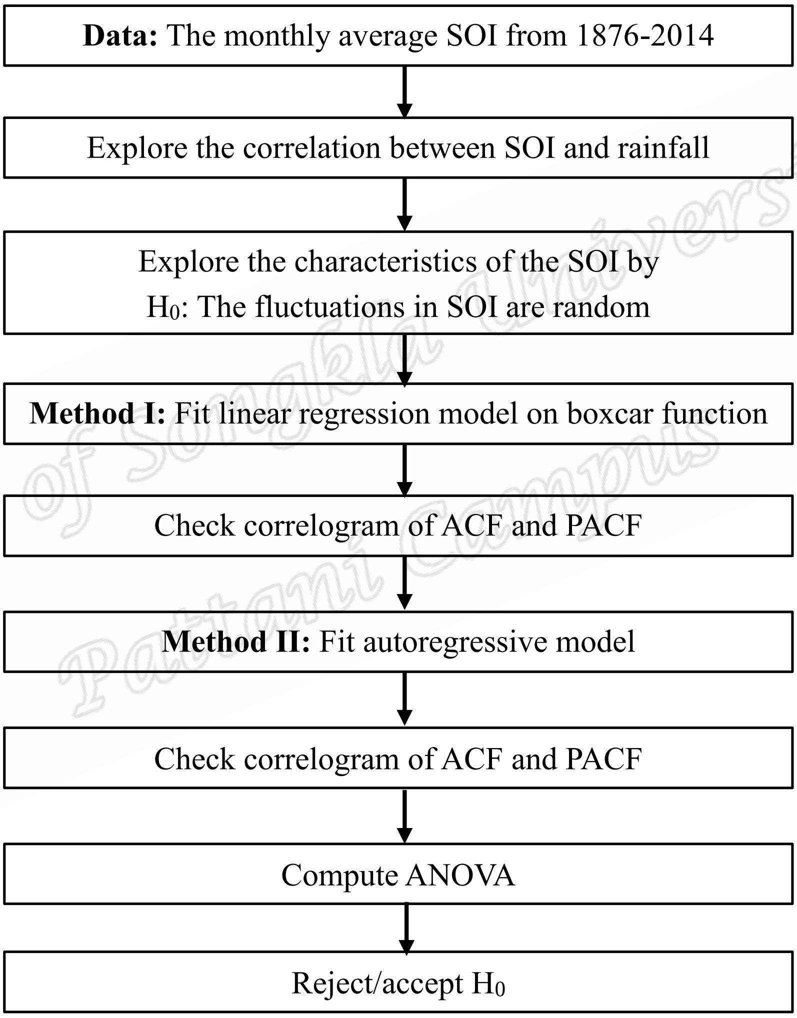

Figure 2.4: Diagram of the SOI analysis

After the correlation between rainfall and the SOI are examined, the linear regression model on boxcar functions has firstly been fitted to reveal the characterisitc of the SOI. Then, the correlogram of ACF and PACF of residuals z_t (Eqn 5) are plotted. The

ACF is used to check the serial correlation, and the PACF is used to indicate which lagged terms need to be included in the AR model to ensure that the w_t (Eqn 7) component represents white noise. Next, AR model with the same predictors as in the linear regression model is analyzed to get the p-values which of each lag are considered; these should be significant (<0.05), for all lagged terms. Again, the correlogram of ACF and PACF of the w_t (Eqn 7) are plotted for checking the reasonable fit of the model. Finally, the Analysis of Variance (ANOVA) procedure is used to test the predictor from the final model to confirm that at least one of the parameters from Eqn 5 is not equal to zero.

2.3 Statistical software

All analysis and graphs are produced by R software which is a freeware an open-source project. It can be easily downloaded from the internet. In general, R is a functional language and software environment for statistical computing and graphics. The URL (Uniform Resource Locator) of R is “http://www.r-project.org” (R Development Core Team, 2013).

In summary of this chapter, the first study, 5-day averages of rainfall are conducted to avoid the problem of serial correlation. Factor analysis is used to classify 114 stations to reduce the spatial correlation. Gamma-distributed generalized linear model is used to see the pattern and extent of daily rainfall across Thailand. In the second study, for the SOI data, linear regression model and autoregressive model are appropriate scientific approach to test the null hypothesis that the SOI is a random noise.

In the next chapter, the results of daily rainfall analysis have been shown.

CHAPTER 3

Analysis of Daily Rainfall

In this chapter, the results of the analysis of the daily rainfall during 2001 to 2012 in Thailand have been explained. Factor analysis is used to reduce the correlation between rainfall stations, while generalized linear model (GLM) by using gamma regression is used to examine the month and year effects. The manuscript of this study entitled "Analysis of daily rainfall during 2001-2012 in Thailand" is published in Songklanakarin Journal of Science and Technology, Vol. 37, No. 1 (Jan.-Feb., 2015), (see Appendix II).

3.1 Preliminary analysis

To start with, the number of rainy days in the period 2001 to 2012 for the regions of Thailand such as North, North-east, Central, East, South-east and South-west are explored and classified into five groups, based on the amount of rainfall, as shown in Table 3.1. In the North, the number of days with light rain is around 84 days/year, moderate rain is around 33 days/year and heavy rain is around 10 days/year. With more than 0.1 mm/day rainfall as the base, the average daily rainfall for the three seasons including summer, rainy and winter have been shown in Table 3.2. The average daily rainfall in the summer is 11.26 mm/day, rainy season is 11.07 mm/day and winter is 8.91 mm/day. There is not much difference between the average rainfall during summer and rainy seasons. In 12 years, the maximum daily rainfall is 263.70 mm/day at the station of Uttaradit occurring on 22nd May, 2006.

In the North-east, the number of days with light rain is around 75 days/year, moderate rain 31 days/year and heavy rain 12 days/year. The average daily rainfall is 13.36 mm/day in rainy season, 10.99 mm/day in the summer and 8.14 mm/day in the winter. The maximum rainfall is 405.79 mm/day and occurs at the station of Nong Khai on 30th July, 2011.

In the Central, light rain is around 79 days/year, moderate rain is 29 days/year and heavy rain is 10 days/year. There is not much difference for the average rainfall in all three seasons. It is around 11.47, 10.95 and 9.73 mm/day in the summer, rainy and winter, respectively. The maximum rainfall for the 12 years in this area is 216.80 mm/day at the station of Bangkok Metropolis on 27th April, 2009.

In the East, the number of days with light rain is around 83 days/year, moderate rain is around 37 days/year and heavy rain is around 15 days/year. The maximum number of heavy rain days is 45 days/year at Klong Yai, 32 days/year at Phriu Agromet and 29 days/year at Chanthaburi stations. In 12 years, the maximum rainfall at Klong Yai is around 445.30 mm/day on 2nd July, 2006. Klong Yai receives rainfall of more than 200 mm/day for 12 days which are one day in 2001, three days in 2006, two days in 2007, a day in 2008, three days in 2009 and a day in 2011. At Phriu Agromet, the maximum rainfall is 242.20 mm/day which occurs on 12nd July, 2012. At Chanthaburi, the maximum rainfall is 273.00 mm/day and occurs on 5th October, 2006. The average daily rainfall for all stations in this area is over 10.00 mm/days in the summer and rainy seasons.

In the South-east, the number of days is 102, 38 and 13 days/year for light, moderate and heavy rain, respectively. The rainy season of this area is during the winter season of the country. The average rainfall for almost all stations in the rainy season are more than 10.00 mm/day, except Phra Sang and Chawang which receive rain around 8.95 and 9.56 mm/day, respectively. It can be seen clearly that the average rainfall is quite high at the stations of Nakhon Si Thammarat, Nakhorn Sri Thammarat Agromet, Phatthalung Agromet, Songkhla and Narathiwat. In the summer, the rainfall for almost all stations are more than 10.00 mm/day, except the stations of Songkhla and Pattani Airport. The maximum rainfall in this area is 521.80 mm/day at the station of Songkhla and occurs on 23rd November, 2005.

In the South-west, the number of days with light rain is 106 days/year, moderate rain is 51 days/year and heavy rain is 23 days/year. The number of rainy days (185 days/year) is more than the number of trace days (180 days/year). Moreover, the number of moderate and heavy rainy days, ranged from 41-61 and 16-40 days/year, respectively, are quite high compared to other regions. The average rainfall is almost similar for the three seasons. Likewise, the stations of Ranong and Takua Pa exhibit high rainfall in rainy season compared to other stations.

Table 3.1: Number in average of rainy days of six regions in Thailand

Regions	Rainy day (days/year)				
	TD	LR	MR	HR	VHR
North	237 (182-261)	84 (63-114)	33 (21-45)	9 (5-15)	1 (1-1)
North-east	249 (226-262)	75 (66-88)	31 (24-43)	11 (6-20)	1 (1-2)
Central	248 (203-266)	79 (68-104)	29 (24-49)	9 (5-13)	1 (1-2)
East	230 (163-264)	83 (67-97)	37 (27-63)	13 (7-34)	2 (1-11)
South-east	213 (187-261)	102 (78-119)	38 (20-54)	11 (5-17)	2 (1-5)
South-west	180 (141-205)	106 (96-116)	51 (41-61)	20 (15-33)	3 (1-7)

Trace day (TD) is rain <0.1 mm/day, light rain (LR) is rain 0.1-10.00 mm/day, moderate rain (MR) is rain 10.1-35.0 mm/day, heavy rain (HR) is rain 35.1-90.0 mm/day and very heavy rain (VHR) is rain >90.0 mm/day

Table 3.2: Average, median and maximum daily rainfall 114 stations of three seasons

Stations	Summer (mid-Feb to mid-May) (mm/day)			Rainy season (mid-May to mid-Oct) (mm/day)			Winter (mid-Oct to mid-Feb) (mm/day)		
	Mean	Median	Max	Mean	Median	Max	Mean	Median	Max
North	***11.26***	***5.00***	***107.03***	***11.07***	***5.10***	***145.75***	***8.91***	***3.29***	***83.96***
Mae.Hong.Son	10.09	4.80	70.10	9.74	4.80	99.90	9.85	3.00	128.00
Mae.Sariang	10.48	5.50	71.40	8.77	4.90	135.50	8.77	3.70	70.70
Chiang.Rai	10.43	4.00	89.30	14.01	7.10	138.50	9.89	3.80	138.60
Chiang.Rai.Agromet	9.82	4.80	69.70	13.16	6.85	129.40	8.31	3.85	147.40
Phayao	11.32	5.00	89.10	10.13	4.60	154.30	9.85	3.70	105.50
Chiang.Mai	10.80	5.15	113.80	10.07	5.20	144.40	12.12	5.30	114.60
Lampang	11.65	5.10	191.20	9.28	3.60	135.40	7.72	2.40	77.30
Lampang.Agromet	12.89	6.85	91.90	10.84	4.90	104.30	10.35	4.70	114.40
Lamphun	9.76	4.00	82.90	10.33	4.40	156.00	8.61	3.70	74.20
Phrae	11.00	4.70	158.10	10.29	4.35	218.20	7.22	2.40	60.00
Nan	12.51	6.50	129.40	11.48	4.80	143.50	6.63	1.90	49.80
Nan.Agromet	12.23	5.65	92.50	11.64	5.00	159.40	7.45	2.55	47.50
Tha.Wang.Pha	10.05	4.35	104.20	11.98	5.10	141.80	7.28	2.25	49.20
Thung.Chang	12.47	7.70	100.70	14.17	6.90	259.00	10.43	4.00	71.90
Uttaradit	11.70	4.30	108.30	13.19	5.30	263.70	8.20	3.10	44.70
Si.Samrong.Agromet	14.09	5.30	116.40	11.58	5.30	139.30	9.60	2.70	102.10
Tak	12.56	5.10	99.80	10.02	4.00	123.40	9.44	3.20	123.50
Mae.Sot	10.77	4.50	135.10	12.12	6.60	121.90	8.39	2.90	73.10
Bhumibol.Dam	15.51	6.20	247.10	9.85	3.25	170.70	11.12	5.00	100.90
Doi.Muser.Agromet.Stn.	10.08	4.65	93.80	10.49	5.70	125.50	8.36	2.45	90.60
Umphang	11.09	5.70	93.00	9.28	5.30	124.70	6.63	2.80	50.00
Phitsanulok	10.61	4.00	85.30	12.11	5.60	124.10	8.92	2.40	93.80

Stations	Summer (mid-Feb to mid-May) (mm/day)			Rainy season (mid-May to mid-Oct) (mm/day)			Winter (mid-Oct to mid-Feb) (mm/day)		
	Mean	Median	Max	Mean	Median	Max	Mean	Median	Max
Phetchabun	10.01	4.20	85.60	9.92	4.60	143.10	7.21	3.40	60.10
Lom.Sak	8.05	3.80	62.80	9.40	4.20	115.60	6.67	3.30	34.30
Wichian.Buri	12.44	5.85	94.00	11.92	5.30	110.80	11.25	4.50	97.70
Kamphaeng.Phet	11.70	4.00	117.20	10.99	4.40	112.10	11.03	3.20	84.30
Pichit.Agromet	9.94	3.30	97.20	12.11	5.70	140.80	9.35	2.60	62.80
North-east	***10.99***	***4.42***	***116.34***	***13.36***	***6.13***	***174.09***	***8.14***	***2.82***	***72.40***
Nong.Khai	10.54	4.70	99.20	15.11	6.10	405.90	8.80	2.70	83.20
Loei	9.58	4.00	67.80	11.23	4.80	152.00	7.75	2.30	69.20
Loei.Agromet	10.87	4.40	114.40	10.86	4.35	128.30	7.63	1.90	86.10
Udon.Thani	8.24	2.50	94.80	13.05	5.45	192.60	8.19	2.00	64.40
Sakon.Nakhon	11.09	4.00	141.50	13.96	6.60	155.60	6.38	2.50	70.20
Sakon.Nakhon.Agromet	11.81	5.10	110.00	13.91	6.80	173.90	8.51	2.90	66.10
Nakhon.Phanom	11.22	4.05	105.00	19.76	10.60	203.00	4.44	1.90	43.80
Nakhon.Phanom.Agromet	11.04	4.90	114.30	17.91	9.70	243.60	5.78	2.10	54.40
Khon.Kaen	12.64	4.50	221.90	11.92	5.40	127.20	9.55	3.30	92.30
Tha.Phra.Agromet	11.43	4.65	164.20	11.79	5.40	141.40	10.57	3.80	90.60
Mukdahan	9.23	3.20	104.30	14.32	6.50	141.10	5.92	2.60	55.00
Kosum.Phisai	11.69	5.30	147.80	13.46	6.60	182.30	8.03	3.20	112.60
Kamalasai	10.72	4.40	109.70	14.10	6.20	188.10	8.28	3.10	62.40
Chaiyaphum	10.72	4.95	162.50	11.90	4.50	133.20	9.66	3.10	68.60
Roi.Et	12.07	5.00	96.50	13.42	6.50	198.60	9.97	3.90	68.40
Roi.Et.Agromet	11.09	3.60	103.50	13.84	6.45	128.10	12.46	4.20	73.10
Ubon.Ratchathani.Agromet	9.53	3.70	115.90	14.44	6.60	125.60	7.47	1.40	65.00
Ubon.Ratchathani	11.37	4.40	116.70	15.19	7.90	153.00	9.28	2.40	122.40
Si.Sa.Ket.Agromet	11.95	5.00	89.50	15.00	6.90	263.40	7.27	3.90	37.40

Stations	Summer (mid-Feb to mid-May) (mm/day)			Rainy season (mid-May to mid-Oct) (mm/day)			Winter (mid-Oct to mid-Feb) (mm/day)		
	Mean	Median	Max	Mean	Median	Max	Mean	Median	Max
Nakhon.Ratchasima	10.90	4.00	92.00	10.41	3.70	129.70	7.36	2.90	61.30
Pak.Chong.Agromet	10.49	4.85	81.20	9.67	4.40	145.90	7.01	2.70	60.10
Chok.Chai	11.16	5.00	100.30	9.85	4.20	114.00	7.48	2.75	52.60
Surin	12.99	5.55	144.40	13.30	6.00	153.00	7.71	2.40	91.40
Surin.Agromet	11.24	4.35	90.70	13.53	6.00	241.60	7.65	2.35	55.10
Tha.Tum	10.95	4.10	116.20	13.80	7.00	177.70	10.80	3.70	83.90
Nang.Rong	11.13	4.75	120.50	11.56	4.80	127.60	7.77	3.20	92.70
Central	***11.47***	***4.76***	***108.34***	***10.95***	***4.66***	***142.76***	***9.73***	***3.71***	***95.90***
Nakhon.Sawan	9.68	2.70	102.70	11.20	4.10	136.40	9.74	2.90	86.50
Tak.Fa.Agromet	11.44	5.10	85.30	11.43	4.50	149.30	8.68	3.15	79.70
Chai.Nat	10.50	5.05	75.00	10.23	4.20	96.30	9.03	2.65	87.90
Ayuttaya.Agromet	10.84	4.75	94.50	10.90	4.10	144.60	11.48	4.70	94.20
Pathumthani.Agromet	11.62	4.30	107.30	11.18	4.50	180.50	7.95	2.40	98.00
Ratchaburi	9.29	3.40	98.20	9.58	3.65	108.90	9.04	3.90	141.30
Suphan.Buri	10.03	3.15	103.40	8.88	3.40	190.40	9.20	3.20	84.70
U.Thong.Agromet	9.76	3.15	73.00	9.03	3.20	161.90	10.10	5.15	68.70
Lop.Buri	12.15	4.60	105.60	11.32	4.80	164.90	8.58	3.75	70.10
Bua.Chum	8.76	3.10	104.90	11.24	4.10	172.20	9.02	3.00	80.70
Pilot.Station	10.71	6.10	87.80	9.58	3.80	135.20	11.25	4.00	153.30
Kanchanaburi	13.82	6.10	100.50	8.97	3.40	119.00	10.22	4.00	124.70
Thong.Pha.Phum	11.35	5.60	84.00	11.68	6.60	109.80	7.91	3.85	103.50
Kamphaeng.Saen.Agromet	11.45	4.75	94.40	9.71	4.30	124.30	10.58	4.30	71.50
Bangkok.Metropolis	13.94	5.45	216.80	13.26	6.40	157.40	10.82	4.90	69.60
Klong.Toey	13.72	5.70	140.30	12.81	6.00	152.20	10.75	3.80	77.90
Bang.Na	13.79	7.15	185.90	13.01	6.30	148.40	10.34	3.70	110.60

Stations	Summer (mid-Feb to mid-May) (mm/day)			Rainy season (mid-May to mid-Oct) (mm/day)			Winter (mid-Oct to mid-Feb) (mm/day)		
	Mean	Median	Max	Mean	Median	Max	Mean	Median	Max
Donmuang	13.57	5.60	90.50	13.15	6.50	118.00	10.42	3.50	123.30
East	***12.37***	***5.49***	***109.58***	***13.92***	***6.70***	***176.52***	***9.63***	***4.33***	***88.75***
Chacherngsao.Agromet	12.34	6.75	101.60	11.51	5.90	130.50	7.04	3.60	88.90
Prachin.Buri	12.75	4.95	121.30	14.66	6.85	145.60	8.74	3.70	72.00
Kabin.Buri	11.65	4.70	94.50	11.83	6.30	104.00	7.80	2.65	99.70
Aranyaprathet	10.46	4.00	98.50	11.17	5.60	95.20	8.48	3.50	51.00
Sa.Kaew	10.64	5.60	95.40	11.79	6.15	127.50	6.26	2.70	96.50
Chon.Buri	11.03	5.30	84.20	10.84	4.50	150.00	10.24	6.00	107.20
Ko.Sichang	11.92	5.65	87.90	10.52	4.60	137.60	12.35	6.95	99.40
Phatthaya	10.71	5.00	66.80	10.69	4.40	189.40	10.08	5.50	63.40
Sattahip	14.43	5.20	156.20	10.29	4.50	160.30	10.83	5.40	59.70
Lam.Chabang	10.87	6.50	78.40	11.13	5.20	126.00	10.13	5.20	57.00
Rayong	12.98	5.60	128.40	11.68	4.70	137.10	9.41	4.10	78.80
Huai.Pong.Agromet	12.75	4.50	112.30	10.75	4.60	183.90	10.53	4.50	111.30
Chanthaburi	14.33	5.50	117.70	19.18	9.25	273.00	10.94	3.40	113.80
Phriu.Agromet	14.91	7.40	177.00	21.02	10.70	242.40	11.30	3.90	130.10
Khlong.Yai	13.84	5.75	123.50	31.72	17.30	445.30	10.36	3.80	102.40
South-east	***12.52***	***4.86***	***181.64***	***9.09***	***4.10***	***121.81***	***16.03***	***6.36***	***258.10***
Phetchaburi	11.91	3.90	161.10	8.29	3.00	148.10	12.50	5.40	118.30
Prachuap.Khiri.Khan	14.74	5.80	199.80	7.22	2.90	131.60	12.27	5.15	173.40
Hua.Hin	14.05	3.55	162.00	7.39	2.40	126.20	11.39	3.50	188.00
Nong.Phlup.Agromet	11.72	4.25	195.80	7.41	2.40	226.00	10.82	3.30	187.10
Chumphon	13.52	5.20	178.10	9.27	4.80	153.50	16.46	6.70	181.10
Sawi.Agromet	13.89	5.80	131.10	8.85	4.65	122.30	17.80	7.00	194.30

Stations	Summer (mid-Feb to mid-May) (mm/day)			Rainy season (mid-May to mid-Oct) (mm/day)			Winter (mid-Oct to mid-Feb) (mm/day)		
	Mean	Median	Max	Mean	Median	Max	Mean	Median	Max
Surat.Thani	12.23	4.10	241.50	8.30	3.50	124.90	12.69	4.80	239.50
Ko.Samui	13.77	4.70	414.70	8.82	3.95	114.60	18.92	6.80	363.90
Surat.Thani.Agromet	13.52	5.50	275.10	9.63	4.50	116.00	16.28	6.35	242.50
Phra.Sang	11.30	5.20	87.10	9.11	4.70	113.00	8.95	4.05	139.50
Nakhon.Si.Thammarat	15.78	6.00	290.40	9.27	4.20	111.10	23.99	9.80	405.00
Nakhorn.Sri.Thammarat.Agromet	14.54	5.40	215.00	9.72	4.40	135.80	21.35	8.20	365.40
Chawang	10.39	4.50	104.20	11.03	6.70	79.20	9.56	4.00	195.00
Phatthalung.Agromet	13.68	5.05	199.10	7.91	4.00	78.70	20.73	9.70	292.20
Kho.Hong.Agromet	11.36	4.60	106.80	9.80	4.60	152.60	17.50	7.40	278.00
Sa.Dao	11.45	6.00	77.50	9.53	4.50	110.60	14.03	6.80	174.70
Songkhla	9.48	4.60	87.50	9.11	3.90	94.80	21.43	8.55	521.80
Hat.Yai.Airport	10.81	4.80	89.90	8.45	3.40	106.20	14.76	6.60	208.00
Pattani.Airport	9.75	4.30	100.60	9.24	4.20	87.80	17.83	6.70	274.30
Yala.Agromet	11.61	4.10	101.20	12.00	4.80	122.00	16.90	5.85	279.10
Narathiwat	13.35	4.70	395.90	10.50	4.50	103.00	20.38	6.90	399.00
South-west	***13.09***	***5.91***	***134.63***	***17.49***	***9.00***	***187.19***	***10.42***	***4.44***	***116.54***
Ranong	16.43	6.80	182.60	25.43	15.00	208.00	9.28	3.55	95.50
Takua.Pa	14.66	7.70	145.00	22.81	11.50	207.40	11.62	5.45	81.60
Phuket	11.81	6.40	101.60	14.43	7.20	180.70	10.19	5.00	92.90
Phuket.Airport	12.63	4.75	185.40	16.95	7.95	211.90	9.98	4.50	107.30
Ko.Lanta	11.40	5.20	87.30	16.62	8.00	235.10	9.64	3.85	162.70
Trang.Airport	11.89	4.90	121.10	12.82	6.70	103.60	10.79	4.20	150.00
Satun	12.79	5.60	119.40	13.36	6.65	163.60	11.46	4.50	125.80

Due to the rainfall in closely located stations are spatially correlated which can be seen from the matrix plots as shown in Figure 3.1. The bubble chart presents the correlation of rainfall between 114 stations ordered by regions from North (NN), North-east (NE), Central (CC), East (EE), South-east (SE) and South-west (SW). The size of bubble determines the magnitude of Spearman correlation, while the black bubbles define positive correlation and the red bubbles define negative correlation. The stations in the North, North-east, Central and East have shown high correlation between each other, but that not for South-east and South-west, they fairly separate from others. The correlation coefficients vary from -0.11 to 0.91. In order to reduce the spatial correlation, factor analysis is used to allocate those stations to a smaller number of clusters with similar patterns. It is noted that because of the serial correlation, the 5-day averages of rainfall are used for the next step of the analysis. All missing value, Chapter 2, page 18, also are disappeared due to this aggregating data.

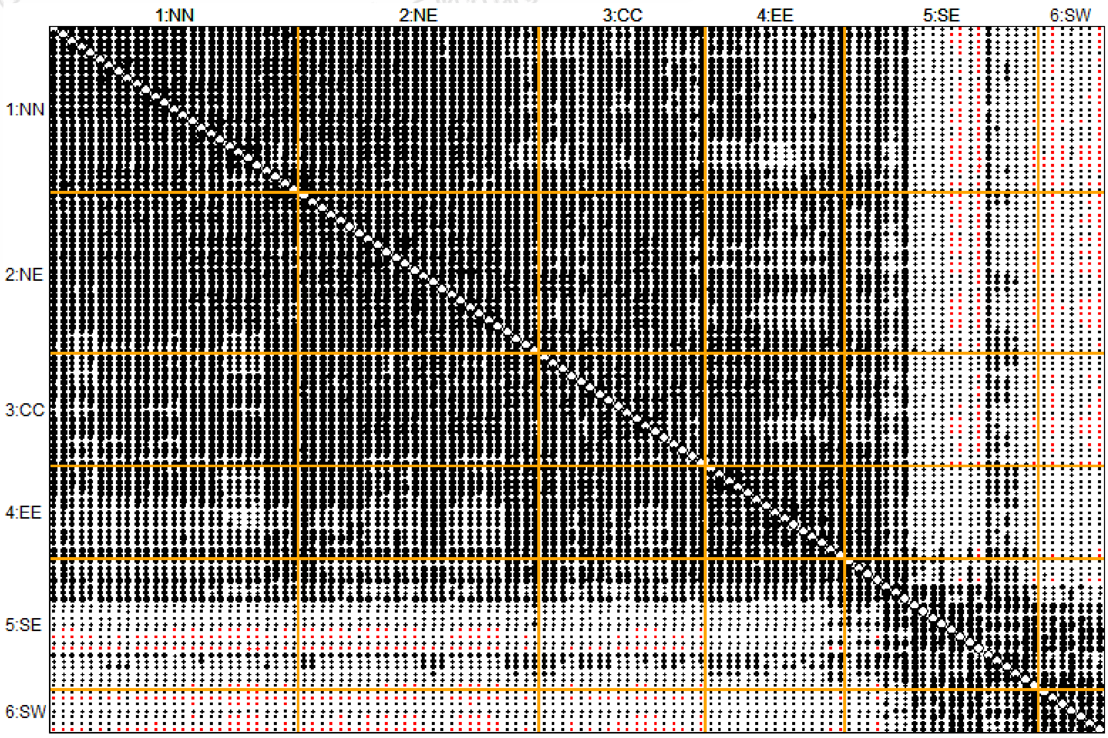

Figure 3.1: Spearman correlation between stations

3.2 Factor analysis

Table 3.3 shows loading scores from the factor analysis indicating the correlations between factors and stations. When the loading scores greater than 0.30, a station is included in a factor found that 86 of the 114 stations correlate with a single factor, and 28 stations correlate with two factors called mixed factor, which is the number in red. However, in the analysis, the stations which are the mixed factor are dominated in the factors which have the highest loading score. Therefore, this technique can reduce 114 stations to seven factors and explain 58.9% of the total variance.

Note that, the 58.9% of total variation is as high as possible. When eight factors are chosen, the total variation is 59.2% but the stations can be allocated only for factors 1-7, and factor 8 does not have any stations. Factor 8 has only factors mixed with factor 1. If nine factors are chosen, the total variation is 58.7% and factor 9 also adds no new information factor. We get only eight factors. The total variation increases to 76.3% when the number of factors is 25 but these remain uninformative. Therefore, the 58.9% of the variation accounted for by the factor analysis is actually as high as can be obtained for these data. As explained in chapter 11 of Venables and Repley (2002), unlike principal component analysis, a factor model cannot always be fitted to a given set of data, and in the present case the 40% of the variation remaining is largely noise.

Table 3.3: Loading scores of seven factors and mixed factor from factor analysis

Stations	F1	F2	F3	F4	F5	F6	F7	Uniq
Nakhon.Phanom.Agromet	1.13	-0.13			-0.24		0.19	0.19
Mukdahan	1.11				-0.16		0.13	0.20
Sakon.Nakhon	1.10				-0.24			0.19
Nakhon.Phanom	1.09	-0.12			-0.21		0.21	0.18
Sakon.Nakhon.Agromet	1.08				-0.24			0.24
Roi.Et	1.01							0.19
Kamalasai	1.01					-0.12		0.18
Roi.Et.Agromet	1.01							0.23
Kosum.Phisai	0.95							0.21
Tha.Phra.Agromet	0.91							0.23
Khon.Kaen	0.86	0.11			-0.12			0.23
Ubon.Ratchathani	0.84						0.16	0.23
Udon.Thani	0.84			0.22				0.23
Tha.Tum	0.80			-0.11				0.25
Surin.Agromet	0.78			-0.13				0.28
Nong.Khai	0.78			0.26				0.24
Surin	0.77			-0.12	0.11			0.28
Ubon.Ratchathani.Agromet	0.77				0.13		0.18	0.27
Si.Sa.Ket.Agromet	0.75				0.16		0.10	0.26
Chaiyaphum	0.57	0.18					-0.19	0.25
Loei	0.57	0.11		0.23			-0.19	0.24
Loei.Agromet	0.55	0.12		0.24			-0.19	0.23
Chok.Chai	0.49	0.42		-0.11			-0.12	0.32
Lom.Sak	0.49			0.17	0.28			0.24
Wichian.Buri	0.44	0.21			0.17		-0.16	0.27
Nakhon.Sawan	0.30	0.28			0.28			0.27
Nang.Rong	0.62	0.34		-0.17				0.30
Nakhon.Ratchasima	0.57	0.32					-0.18	0.28
Sa.Kaew	0.48	0.46					0.10	0.30
Bua.Chum	0.42	0.32			0.11		-0.13	0.29
Khlong.Yai	0.39	0.38	-0.25			0.20	0.26	0.28
Phetchabun	0.44			0.16	0.31			0.27
Pichit.Agromet	0.37	0.11		0.13	0.36			0.28
Huai.Pong.Agromet	-0.19	1.05			-0.19			0.30
Sattahip	-0.25	1.05			-0.13			0.29
Lam.Chabang		1.04			-0.18			0.26
Phatthaya	-0.14	1.04			-0.13			0.30
Rayong	-0.12	0.97			-0.16			0.33
Ko.Sichang		0.94			-0.10			0.28
Pilot.Station		0.91						0.31
Klong.Toey	0.12	0.89				-0.15		0.23
Chon.Buri		0.87						0.28
Bang.Na	0.11	0.87				-0.12		0.22

Stations	F1	F2	F3	F4	F5	F6	F7	Uniq
Bangkok.Metropolis	0.11	0.86				-0.13		0.24
Phetchaburi		0.78					0.23	0.31
Kamphaeng.Saen.Agromet		0.74			0.15			0.26
Ratchaburi		0.70			0.14		0.16	0.28
Donmuang	0.14	0.68						0.27
Hua.Hin		0.66					0.25	0.37
Kanchanaburi		0.66			0.16			0.34
Chacherngsao.Agromet	0.20	0.65				0.11		0.36
Pathumthani.Agromet	0.14	0.64			0.13			0.29
Nong.Phlup.Agromet		0.63					0.16	0.38
U.Thong.Agromet		0.62			0.23			0.26
Suphan.Buri	0.13	0.61			0.21			0.28
Phriu.Agromet	0.28	0.57	-0.16			0.16	0.15	0.25
Ayuttaya.Agromet	0.20	0.55			0.18			0.29
Lop.Buri	0.23	0.44			0.24			0.28
Pak.Chong.Agromet	0.27	0.39		-0.10	0.17	0.15	-0.17	0.35
Tak.Fa.Agromet	0.24	0.35			0.24		-0.10	0.27
Chanthaburi	0.36	0.56	-0.11				0.14	0.24
Aranyaprathet	0.36	0.47						0.38
Kabin.Buri	0.35	0.45						0.30
Prachin.Buri	0.37	0.43						0.26
Chai.Nat	0.21	0.41			0.32			0.27
Prachuap.Khiri.Khan		0.60			0.11		0.34	0.43
Kho.Hong.Agromet			0.86					0.32
Songkhla			0.83					0.37
Nakhorn.Sri.Thammarat.A	-0.11		0.81					0.34
Yala.Agromet	0.19		0.79					0.45
Nakhon.Si.Thammarat			0.79					0.37
Phatthalung.Agromet	-0.16		0.77					0.39
Hat.Yai.Airport			0.77					0.39
Pattani.Airport			0.77					0.41
Narathiwat			0.73			-0.15		0.54
Sa.Dao			0.62			0.25		0.49
Surat.Thani.Agromet			0.62			0.11	0.20	0.43
Surat.Thani			0.56			0.19	0.20	0.40
Ko.Samui	-0.11	0.19	0.55				0.23	0.52
Chiang.Rai.Agromet	0.19	0.12		0.70				0.21
Chiang.Rai	0.22	0.12		0.67				0.23
Phayao	0.11	0.13		0.67				0.23
Chiang.Mai	0.12			0.52	0.26			0.23
Mae.Hong.Son	0.28			0.44	0.19		0.17	0.28
Thung.Chang	0.32			0.70	-0.17			0.19
Tha.Wang.Pha	0.32			0.70	-0.12			0.16
Nan.Agromet	0.36			0.66				0.16
Nan	0.39			0.62	-0.11			0.18

Stations	F1	F2	F3	F4	F5	F6	F7	Uniq
Lampang.Agromet				0.52	0.37			0.20
Lamphun				0.51	0.30			0.24
Phrae	0.16			0.50	0.31			0.23
Lampang		0.12		0.50	0.30			0.22
Uttaradit	0.16	0.12		0.41	0.33			0.20
Tak	-0.10	0.11		0.18	0.74			0.20
Bhumibol.Dam	-0.15	0.15		0.23	0.69			0.26
Doi.Muser.Agromet.Stn.	0.14			0.22	0.65		0.19	0.15
Mae.Sot	0.29			0.24	0.55	-0.14	0.29	0.16
Si.Samrong.Agromet	0.17	0.11		0.25	0.45			0.26
Umphang	0.21		-0.10	0.15	0.45		0.14	0.28
Thong.Pha.Phum	0.22	0.18	-0.14	0.12	0.42		0.22	0.24
Kamphaeng.Phet	0.17	0.19		0.15	0.41			0.29
Mae.Sariang	0.21	0.11		0.29	0.38		0.20	0.24
Phitsanulok	0.31	0.15		0.15	0.37			0.27
Phuket.Airport	-0.18		0.11			0.84		0.28
Takua.Pa						0.82		0.27
Phuket			0.17			0.78		0.32
Ko.Lanta			0.18			0.72		0.32
Satun			0.27		-0.13	0.64		0.46
Trang.Airport		-0.15	0.33			0.65		0.35
Phra.Sang		-0.13	0.34			0.58	0.11	0.42
Chawang		-0.15	0.33			0.56	0.17	0.38
Ranong	0.28					0.49	0.37	0.21
Sawi.Agromet		0.25	0.25			0.13	0.40	0.52
Chumphon		0.25	0.27				0.39	0.50

F1=Factor 1, F1=Factor 2, F3=Factor 3, F4=Factor 4, F5=Factor 5, F6=Factor 6, F7=Factor 7, Uniq=Uniquenesses

The bubble plots in Figure 3.2 and 3.3 show the correlation between stations before and after fitting the factor analysis, respectively, ordered by the seven factors. It can be seen clearly in Figure 3.3 that the factor model can reduce the correlations of the closed stations. They are presented by the smaller bubbles that are the residuals after applying the factor model.

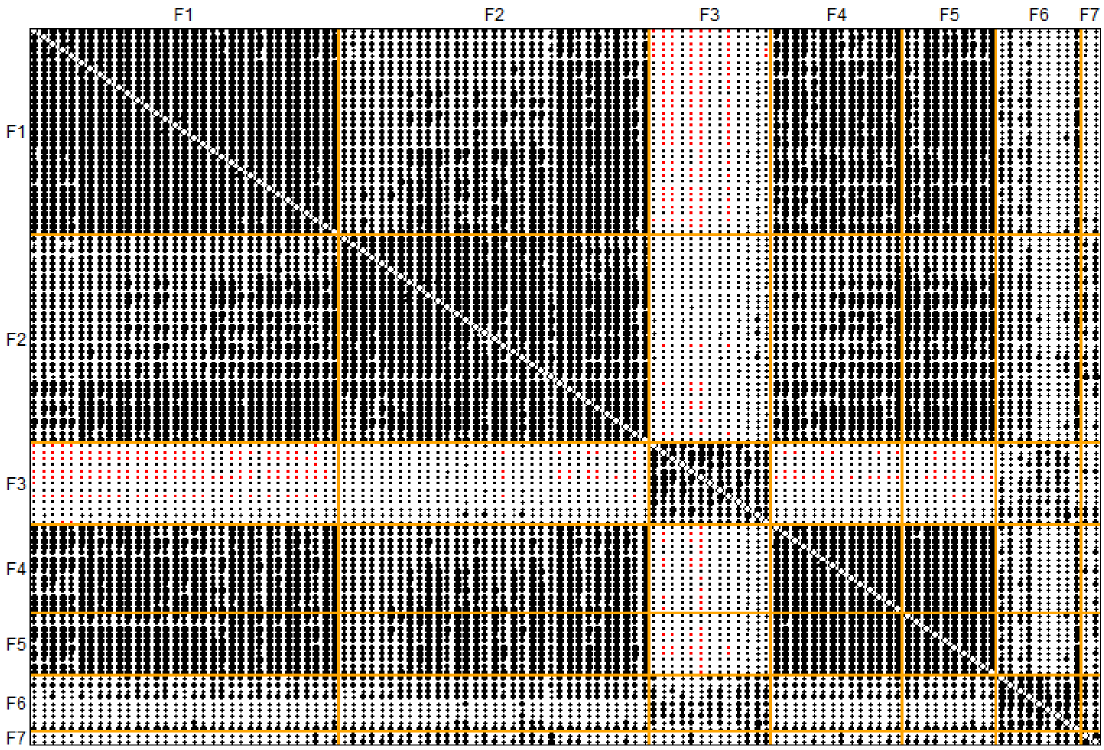

Figure 3.2: Correlation between stations before fitting factor analysis

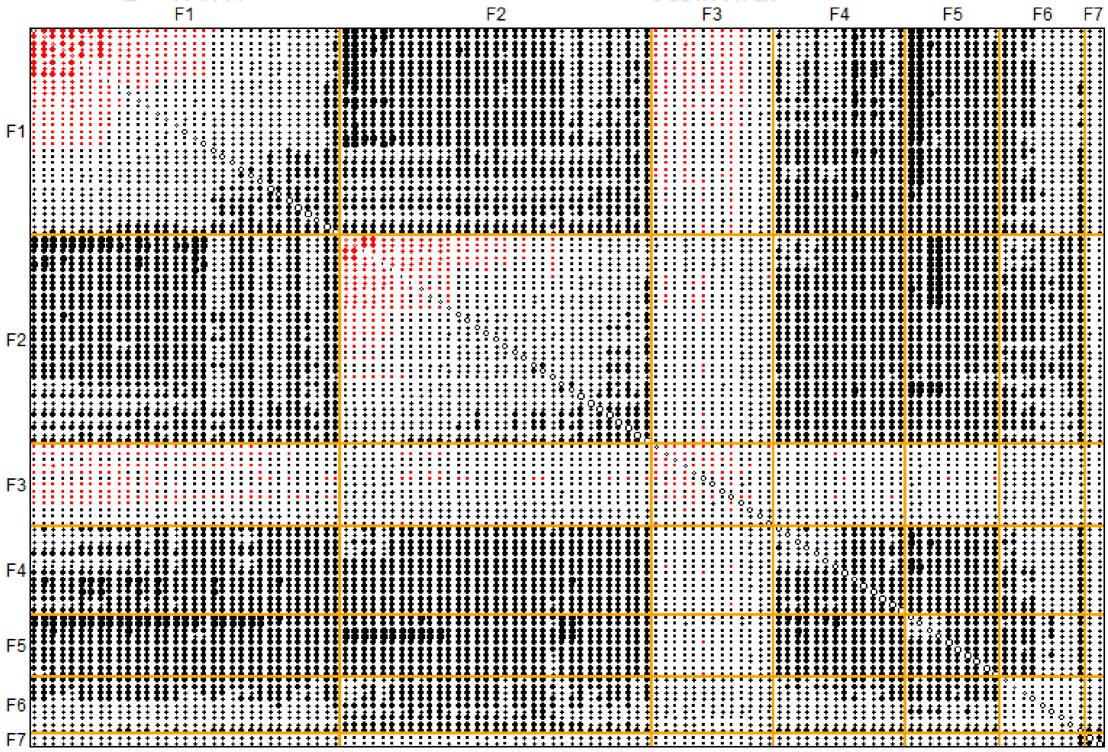

Figure 3.3: Correlation between stations after fitting factor analysis

Moreover, the seven factors have been demonstrated on the map of Thailand (Figure 3.4) which comprises of the all regions. It is noted that double circles on the map denote the mixed factor stations. The map shows that the first factor covers 32 stations in northeastern part of Thailand called north-east. The second factor includes the stations in central and eastern part of Thailand with 34 stations called central. The third factor consists of 13 stations from the area of the south-east called south-east. The fourth factor represents the upper northern part of Thailand with 14 stations called north. The fifth factor combines 10 stations from the lower north called lower north. The sixth factor comprises of nine stations from the south-west called south-west, and the last factor covers the area of the upper south including two stations called upper south. The following section throws light on the rainy season of the seven regions.

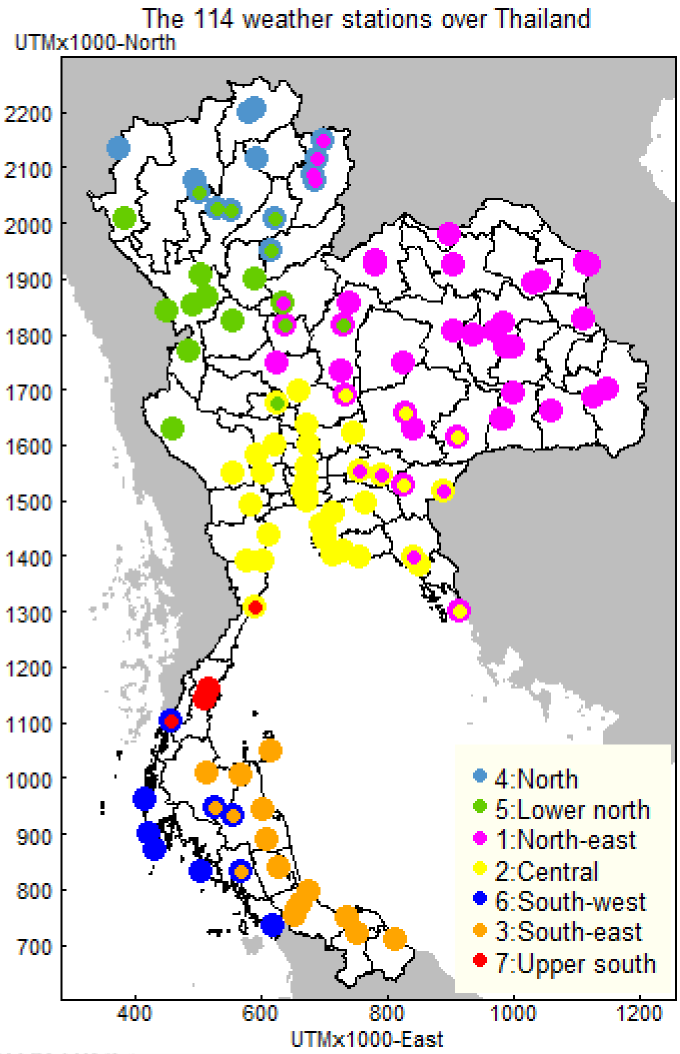

Figure 3.4: Identifying the stations of seven factors on map of Thailand, double circles indicate mixed factor stations

3.3 Rainy season

In order to identify the factor scores for seven regions, the average rainfall is computed from the stations located in the same region for each. Moreover, the overall mean of daily rainfall throughout Thailand is around 4.43 mm/day. Figure 3.5 shows the average daily rainfall of 73 days for 12 years. Successive days with average rainfall above the overall mean are defined as rainy season as shown in green bar. The graphs indicate that the rainy season of the north is in the period of April to

September around six months. The lower north, north-east and central have the rainy season around seven months starting from late April to early October.

In contrast to rainy seasons in the southern region of Thailand, the upper south has four rainy months during September to December. There are also around four months of rainy season for the south-east starting from late September to early January and the south-west has quite a longer rainy season around eight months starting from May to November.

In addition, the daily average rainfall during the rainy season is 11.04, 9.78, 8.19, 7.90, 7.33, 7.07 and 6.30 mm for the region of south-east, south-west, upper south, north-east, north, lower north and central, respectively.

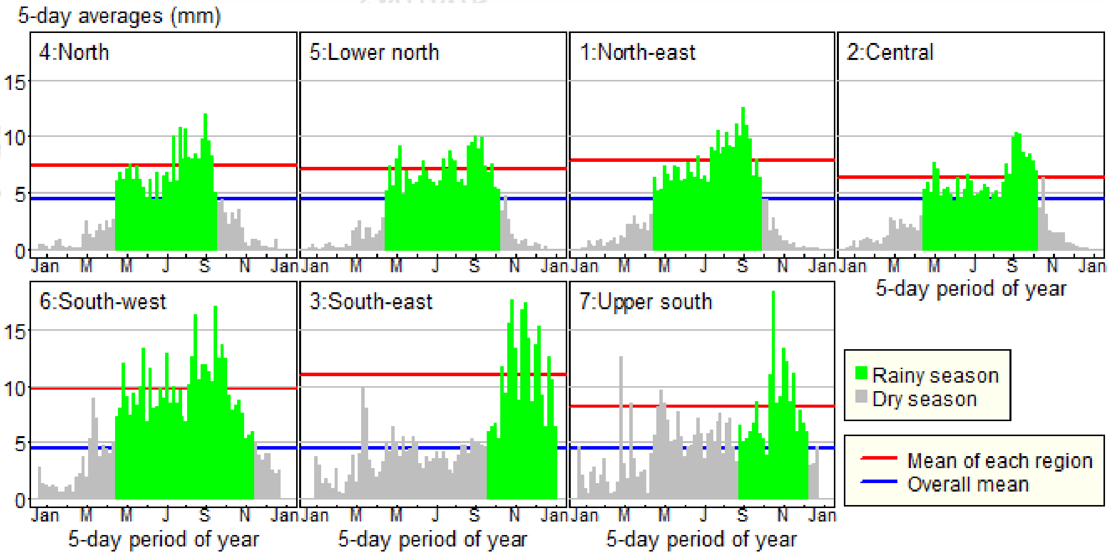

Figure 3.5: Rainy season of seven regions

Figure 3.6 shows the daily rainfall distribution for all regions. The results show that all histograms are skewed to the right even when data are transformed. Thereby, gamma regression model is applied to the period of rainy season of seven regions in next section. Since the outcome that contains the values of zero of gamma model may present some problems to the model, thus, the provided zeroes are replaced by a small amount (0.01 mm, say).

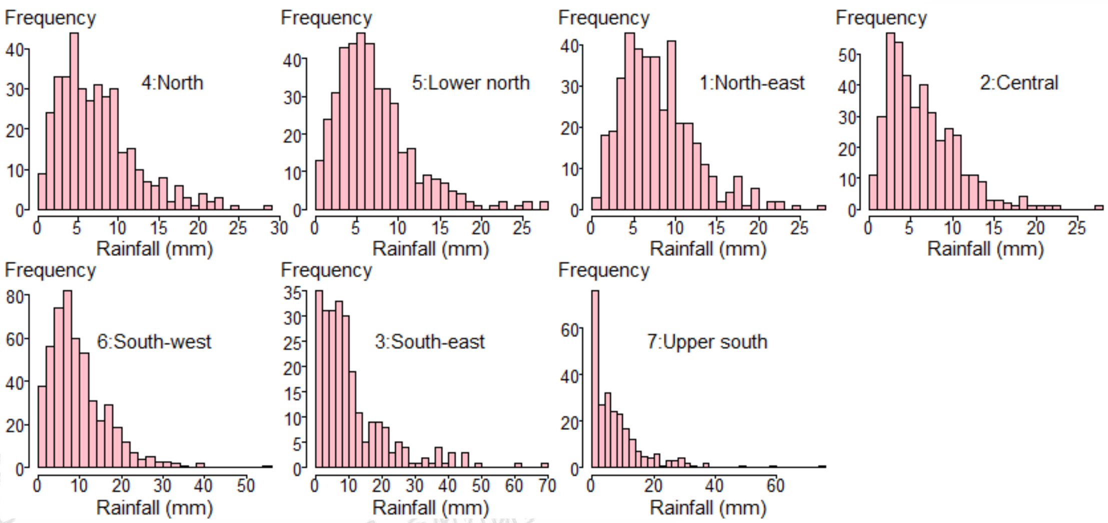

Figure 3.6: Histogram of seven regions have positive skew

3.4 Generalize linear model: gamma regression

Firstly, the normal quantile plots of the deviance residuals from gamma model are considered as shown in Figure 3.7. The graphs indicate that the models provide a reasonable fit for all regions, except the region of upper south where the lower amount of rainfall departs from the normal line (red line).

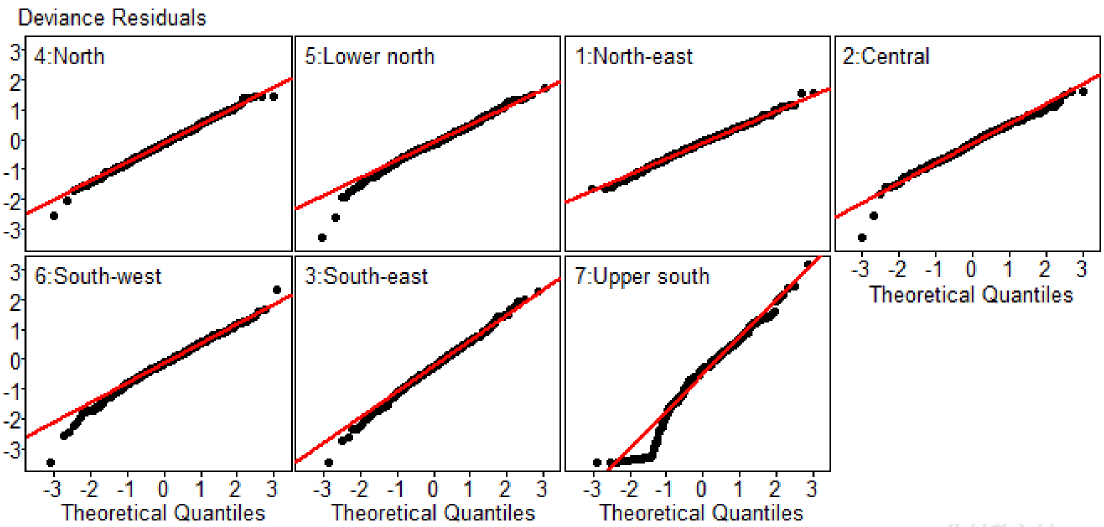

Figure 3.7: Normal quantile plots of deviance residuals from gamma regression model

The results in Figure 3.8 show similar variation of daily rainfall in the three northern regions. These are north, lower north and north-east. Each region has unusually heavy rainfall in their 2011 rainy seasons. When the correlation coefficients are checked between the regions, the results show that the coefficient of correlation between them is over 0.70 (Figure 3.9). Accuracy is improved by aggregating data from those regions. Even though, the upper south and south-east seem to have similar rainy season pattern with similar heavy rainfall of in 2005 and 2010, their coefficients of correlation are lower than 0.50 (Table 3.4). The upper south is then not included in the next step of the analysis.

Table 3.4: Correlation coefficients between regions

Regions	4:NN	5:LN	1:NE	2:CC	6:SW	3:SE	7:US
4:NN	1	0.71	0.70	0.53	0.35	-0.10	0.11
5:LN		1	0.71	0.70	0.41	-0.12	0.17
1:NE			1	0.65	0.45	-0.16	0.13
2:CC				1	0.51	-0.05	0.24
6:SW					1	0.26	0.35
3:SE						1	0.43
7:US							1

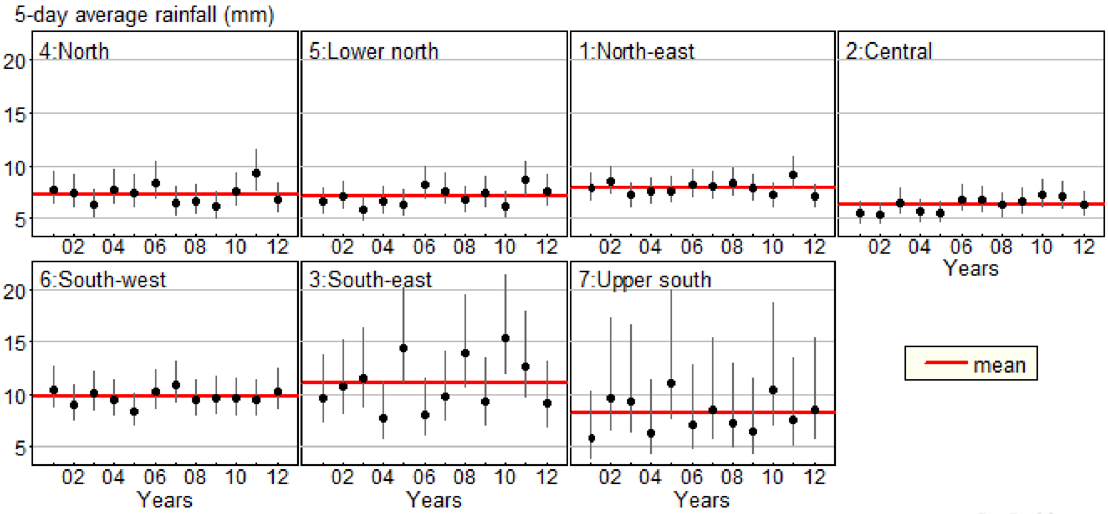

Figure 3.8: 95% confidence intervals of year effects

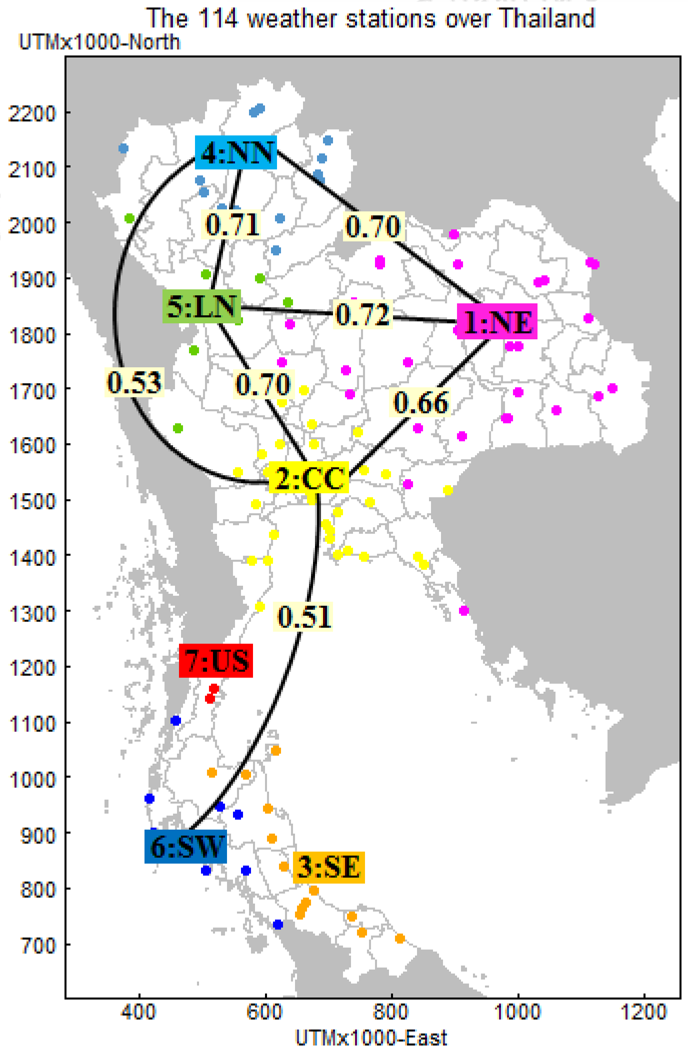

Figure 3.9: Correlation coefficients between regions

Now, the number of regions is reduced to four regions, these are north, central, south-east and south-west. Furthermore, graphs of the autocorrelation function of deviance residuals are plotted as shown in Figure 3.10 to ensure that the independence assumption has been handled properly. The graph of the regions show that the autocorrelations are below 0.20 and no statistically significant correlations are at lag 1 with the p-value of 0.066, 0.907, 0.205 and 0.176 for north, central, south-east and south-west, respectively.

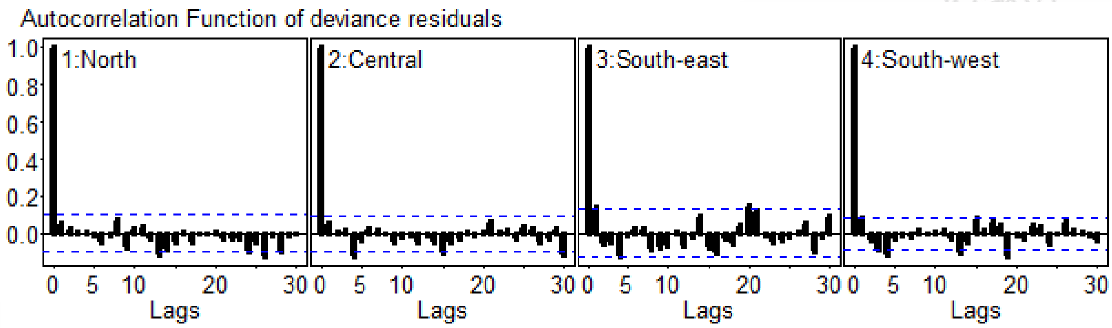

Figure 3.10: Autocorrelation function of deviance residuals of four regions

The deviance residuals from gamma model of each region in Figure 3.11 improve after combining the regions with the same pattern of rainfall.

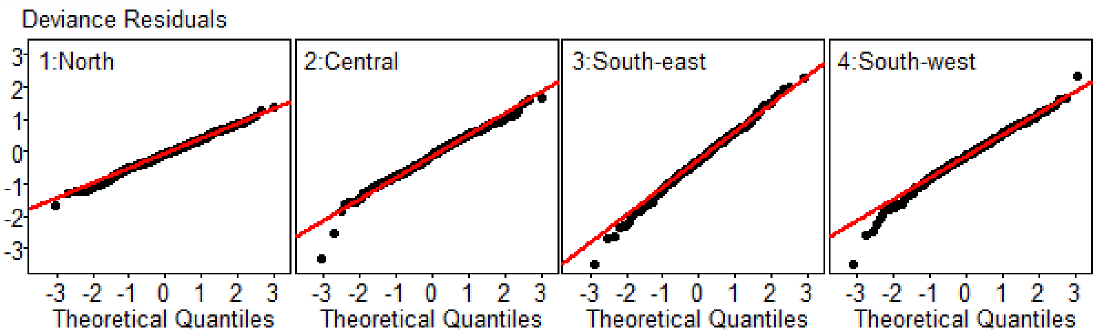

Figure 3.11: Normal quantile plots of deviance residuals for four regions

Figure 3.12 shows the year effects of daily rainfall. These results suggest unusual rainfall occurs in 2011 in the northern and north-east regions of Thailand, the average rainfall is around 9.13 mm/day with 7.98-10.65 mm/day of 95% CI. The lowest average rainfall occurs in 2003 around 6.68 mm/day with 5.84-7.81 mm/day of 95% CI. The overall mean rainfall in rainy season for these region is 7.59 mm/day. The patterns of the average rainfall are mostly under and on the line of its overall mean, except in 2006 and 2011. However, there is only one year, 2011, indicated to be statistically significant.

In the south-east, the graph displays the oscillations of the average. It can be seen clearly the average has increased from 2001 to 2003 where the average is on the line of overall mean, the average decline then rapidly to the bottom in 2004 where it is the lowest average with 7.77 mm/day and 5.89-11.40 mm/day of 95% CI. After that, it goes sharply up to 2005, 14.42 mm/day with 11.19-20.26 mm/day of 95% CI, and goes down again in 2006. Again, it rises until the year 2008 with the average 13.87 mm/day and 10.74-19.55 mm/day of 95% CI. While the average in 2009 decreases again before reaching its peak in 2010 with average 15.39 mm/day and 11.98-21.51 mm/day of 95% CI. Finally, the average decreases slowly till 2012. However, there are two years of 2005 and 2010 where the averages are statistically significantly different from overall mean.

In contrast, in the central and south-west, the daily rainfall has less variation. For central there are four years where the average is under the overall mean including 2001, 2002, 2004 and 2005, whereas all years in the region do not appear to be statistically significant. For south-west, the majority of years reveal the average under and on the line of overall mean except in 2001 and 2007 and they are also not statistically significant.

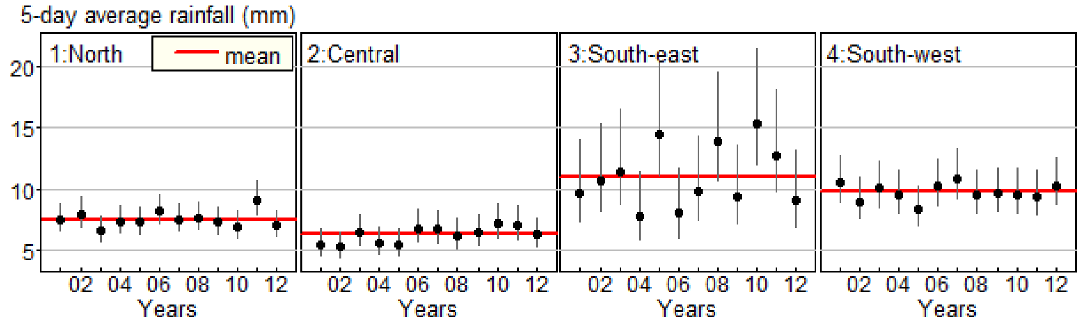

Figure 3.12: 95% confidence intervals of year effects of four regions

When considering the month effects, in the north, the heavy rainfalls are obviously in August where the average is 9.50 mm/day with 8.44-10.86 mm/day of 95% CI and September where the average is 9.57 mm/day with 8.47-11.01 mm/day of 95% CI.

In the central, June seems to have the lowest rainfall when compared to other months. It is significantly under the overall mean in the rainy season of this region, the average is around 5.19 mm/day with 4.57-6.01 mm/day 95% CI. The heavy rainfalls in September and October have the average of 8.89 mm/day with 7.65-10.61 mm/day 95% CI and 7.67 mm/day with 6.43-9.51 mm/day 95% CI, respectively.

In the south-east, the heaviest rainfall occurs in November with the average of 15.64 mm/day and 95% CI between 12.08-22.17 mm/day, likewise in December, the average seems to be above the overall mean. The last region is the south-west where the heaviest rainfall in September is around 13.02 mm/day and 95% CI between 11.07-15.08 mm/day. The average of the two months, August and October, still occur above the overall mean and that of other two months, June and July, appear on the line of the overall mean.

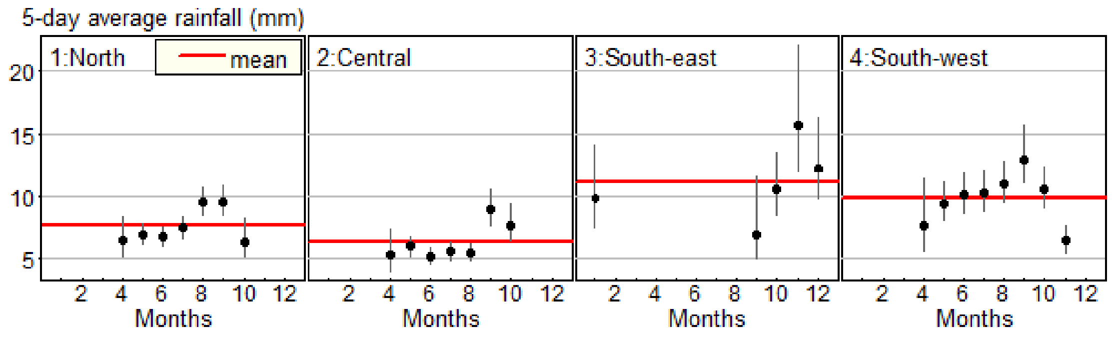

Figure 3.13: 95% confidence intervals of month effects of four regions

In summary, it might be huge task to analysis a large set of response such as the rainfall from the 114 stations used in this study. In order to make the analysis easier, meaningful and informative, factor analysis is applied to rainfall stations to allocate those stations with the same pattern. The stations can then be divided into seven geographical regions which are north, lower north, south-east, central, upper south, south-east and south-west. After investigating the pattern of rainfall and correlation coefficient in the rainy season to all the regions, they remain four regions which are north, central, south-east and south-west. Finally, the generalized linear model with gamma regression fit the data of four regions quit well as revealed by the deviance residuals.

In the next chapter, we examine the relationship between the Southern Oscillation Index (SOI) and rainfall. Results from some previous studies found that the SOI had correlation with the rainfall and other climatic data.

CHAPTER 4

Analysis of the Southern Oscillation Index

In this chapter, we have focused on describing the characteristic of the Southern Oscillation Index in the period of 1876 to 2014. Linear regression and autoregressive model are used to examine the trends of the data. The manuscript of this study entitled "The Southern Oscillation Index as a Random Walk" will be published in the Volume 13, Number 8 of the Walailak Journal of Science and Technology due in August 2016, (see Appendix III).

4.1 Preliminary analysis

The variations of the monthly SOI from 1876 to 2014 are shown in Figure 4.1. The blue lines present the value above zero, likewise, the red lines present the value below zero. The graph suggests that the data fluctuate between -42.60 and 34.8 and do not show any trend. However, there are quite uncommon values of the SOI which express less than -30 that occur in 1896, 1905, 1953, 1982 and 1983 with -42.20, -42.60, -31.90, -31.10 and -33.30, respectively. Three values of the SOI are more than +30 that occur in 1904, 1917 and 1973 with 31.70, 34.80 and 31.60, respectively. According to the Introduction from Chapter 1, the strong El Niño and strong La Niña might occur in those years.

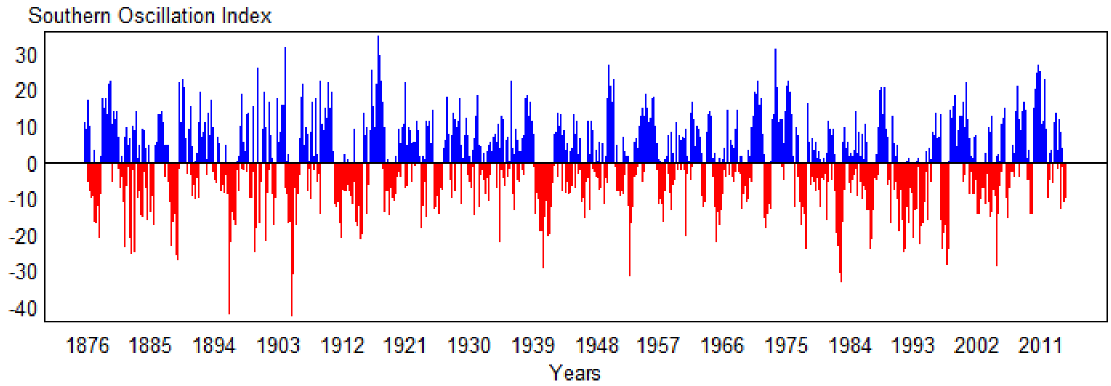

Figure 4.1: Variations of the SOI from 1876 to 2014

Due to the prolonged period of the SOI value less than -8, here define as at least successive four months as shown in Table 4.1, it is realized for occurrence El Niño event. When the SOI more than +8, it is indicated for occurrence La Niña event. With reference to these criteria, an analysis of the last 25 years, would indicate an appearance of the El Niño in August 1991 to April 1992 (9 months), October 1992 to October 1993 (13 months), March to December 1994 (10 months), March 1997 to April 1998 (14 months), July to October 2002 (4 months), December 2009 to March 2010 (4 months) and August to November 2014 (4 months). The values of the last three periods are showed nearly -8 (not strong values). Thus, the expectation of less rain for areas in the western equatorial Pacific should be exhibited in those periods.

Meanwhile, the La Niña might be detected in June 1998 to April 1999 (11 months), October 1999 to April 2000 except January 2000 (6 months), September 2000 to February 2001 (6 months), September 2000 to February 2001 (6 months), November 2007 to March 2008 (5 months), August 2008 to February 2009 (7 months), April 2010 to April 2011 except June 2010 (12 months), July 2011 to January 2012 except August 2011 (6 months) and March to July 2013 (5 months). Unusual heavy rain for areas in the western equatorial Pacific should also be exhibited in those periods.

Table 4.1: Values of the SOI from 1990 to 2014

Years	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec
1990	-1.1	-17	-8.5	-0.5	13.1	1	5.5	-5	-7.6	1.8	-5.3	-2.4
1991	5.1	0.6	-11	-13	-19	-5.5	-1.7	-7.6	-17	-13	-7.3	-17
1992	-25	-9.3	-24	-19	0.5	-13	-6.9	1.4	0.8	-17	-7.3	-5.5
1993	-8.2	-7.9	-8.5	-21	-8.2	-16	-11	-14	-7.6	-14	0.6	1.6
1994	-1.6	0.6	-11	-23	-13	-10	-18	-17	-17	-14	-7.3	-12
1995	-4	-2.7	3.5	-16	-9	-1.5	4.2	0.8	3.2	-1.3	1.3	-5.5
1996	8.4	1.1	6.2	7.8	1.3	13.9	6.8	4.6	6.9	4.2	-0.1	7.2
1997	4.1	13.3	-8.5	-16	-22	-24	-9.5	-20	-15	-18	-15	-9.1
1998	-24	-19	-29	-24	0.5	9.9	14.6	9.8	11.1	10.9	12.5	13.3
1999	15.6	8.6	8.9	18.5	1.3	1	4.8	2.1	-0.4	9.1	13.1	12.8
2000	5.1	12.9	9.4	16.8	3.6	-5.5	-3.7	5.3	9.9	9.7	22.4	7.7
2001	8.9	11.9	6.7	0.3	-9	1.8	-3	-8.9	1.4	-1.9	7.2	-9.1
2002	2.7	7.7	-5.2	-3.8	-15	-6.3	-7.6	-15	-7.6	-7.4	-6	-11
2003	-2	-7.4	-6.8	-5.5	-7.4	-12	2.9	-1.8	-2.2	-1.9	-3.4	9.8
2004	-12	8.6	0.2	-15	13.1	-14	-6.9	-7.6	-2.8	-3.7	-9.3	-8
2005	1.8	-29	0.2	-11	-15	2.6	0.9	-6.9	3.9	10.9	-2.7	0.6
2006	12.7	0.1	13.8	15.2	-9.8	-5.5	-8.9	-16	-5.1	-15	-1.4	-3
2007	-7.3	-2.7	-1.4	-3	-2.7	5	-4.3	2.7	1.5	5.4	9.8	14.4
2008	14.1	21.3	12.2	4.5	-4.3	5	2.2	9.1	14.1	13.4	17.1	13.3
2009	9.4	14.8	0.2	8.6	-5.1	-2.3	1.6	-5	3.9	-15	-6.7	-7
2010	-10	-15	-11	15.2	10	1.8	20.5	18.8	25	18.3	16.4	27.1
2011	19.9	22.3	21.4	25.1	2.1	0.2	10.7	2.1	11.7	7.3	13.8	23
2012	9.4	2.5	2.9	-7.1	-2.7	-10	-1.7	-5	2.7	2.4	3.9	-6
2013	-1.1	-3.6	11.1	0.3	8.4	13.9	8.1	-0.5	3.9	-1.9	9.2	0.6
2014	12.2	-1.3	-13	8.6	4.4	-1.5	-3	-11	-7.5	-8	-10	-5.5

4.2 Correlation between the SOI and rainfall

A graph of the SOI and rainfall from the four regions (results from Chapter 3), i.e. north, central, south-east and south-west, is plotted as shown in Figure 4.2. The graph explains that there are similar patterns of rainfall every year for north, central and south-west.

Interestingly, for south-east, the rainfall reaches its peak in November 2008 (29.87 mm/day) with the SOI of 17.1. The La Niña is observed during August 2008 to February 2009 where the value of the SOI varies from 9.1 to 17.1. Otherwise, the high rainfall in other months does not seem show. Furthermore, in December 2005, November 2009, November 2010 and March 2011 have also shown quite heavy rainfall. Likewise, the observed SOI in December 2005 and November 2009 are 0.6 and -6.7, respectively, and that during July 2010 to April 2011 ranges from 10.0 to 27.1. On the other hand, in February 2005, the SOI has strong negative (-29.0) and rainfall for all regions appear accidentally almost zero. Thus, we can see that there are not consistency patterns between the SOI and rainfall.

Therefore, which indicate that the association between monthly SOI and monthly rainfall of those regions becomes interesting. We found that the correlation coefficients are -0.13, -0.05, 0.18 and -0.05 for north, central, south-east and south-west, respectively. These numbers confirm that there is no relationship between the SOI and daily rainfall during 12 years throughout Thailand. However, more interesting characteristics of the SOI are examined in the next section.

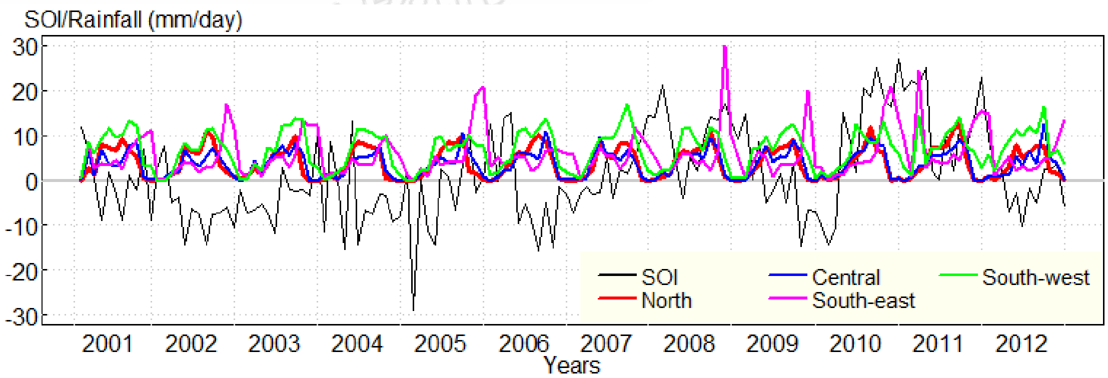

Figure 4.2: The SOI and daily rainfall in the period 2001 to 2012

4.3 Cumulative monthly SOI

Watts (2010) has pointed out the interesting issue of cumulative monthly SOI graph. From the graph, he mentioned that a long La Niña-dominate trend occurred from 1920 to 1976 and El Niño-dominated trend occurred from 1976 to 1995.

As we can see the raw data of the SOI from the graph in Figure 4.1, it does not show the trend because the data fluctuate. When the cumulative monthly SOI are plotted, we can see an apparent trend from 1876 to 2014 as shown in Figure 4.3. It can be seen clearly the patterns suggest four periods. The first period from 1876 to 1919 seems to have no trend. The second period shows an increasing trend between 1920 and 1975. The third period shows a decreasing trend between 1976 and 1995. The fourth period from 1996 to 2014, it appears again that there is no trend in the period.

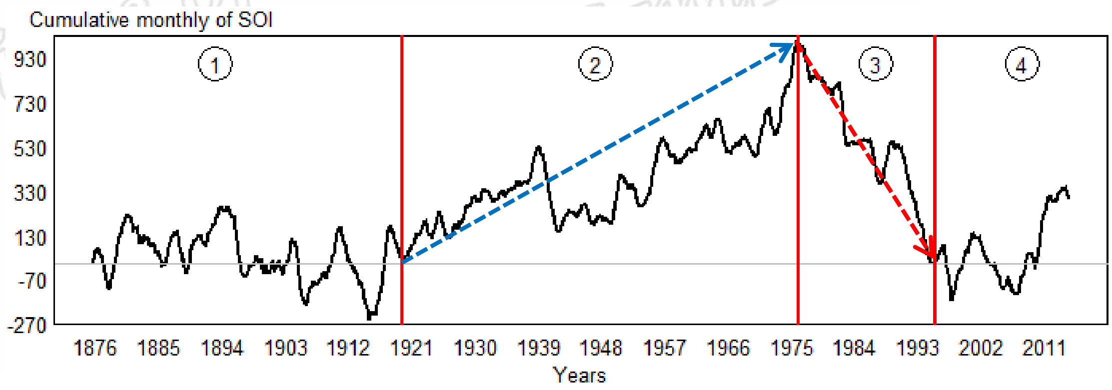

Figure 4.3: Cumulative monthly SOI from 1876 to 2014

In order to investigate those trends, the appropriate scientific methods are applied to three subsets of the data. The first series includes the data from the second and third periods, from 1920 to 1995, in which the increasing and decreasing trend appear. The second series includes the data from 1920 to 2014 which are three periods, the first period of the data is excluded from the analysis. The third series includes all data, from 1876 to 2014. Remember that the null hypothesis (H_0) of the study is the fluctuations in the SOI are random.

4.3.1 Analysis of two periods of data

Firstly, only the data from the second and third periods, between 1920 and 1995, are considered, while the first and last period are assumed to be constant (Figure 4.4). The two periods are analysed using linear model and autoregressive model on one boxcar function.

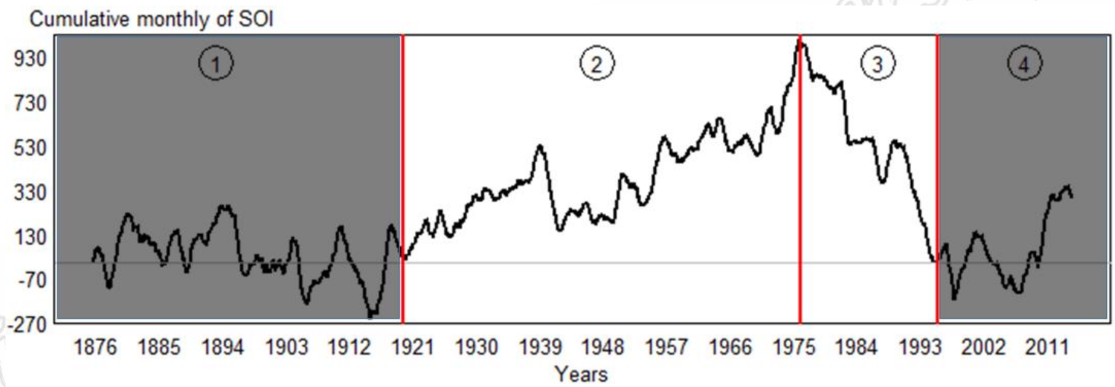

Figure 4.4: Two periods of the data from 1920 to 1995

The null hypothesis is H_0: $b_0 = b_1 = 0$, where b_0 and b_1 refer to the slopes of the second and third periods, respectively. The linear model from Eqn 5, Chapter 2, page 27 can be written as

$$y_t = b_0 + b_1 x_{1t} + z_t .$$

The boxcar function (x_1) takes value of 1 starting at the changing point during the year 1976 to 1995 and 0 elsewhere.

Results from linear model

Table 4.2 shows the coefficient of parameters b_0 and b_1 from the linear model. The model indicates that the estimates of the coefficients of b_0 and b_1 are statistically significant with p-values less than 0.050. In other words, there are significant the upward trend in the second period, (b_0 = 1.409, p-value < 0.001), and downward trend in the third periods, (b_1 = -5.417, p-value < 0.001).

Table 4.2: Results from linear model for two periods

Parameters	Coefficients	SE	t-value	P-value
b_0	1.409	0.363	3.888	<0.001
b_1	-5.417	0.707	-7.667	<0.001
The p-value of ANOVA test is <0.001				

The ACF graph of the residuals from the linear regression model shows that the data are highly serially correlated, referring to Chapter 2, page 28, the z_t still be coloured noise. Furthermore, the PACF graph suggests the correlation of terms at lags 1, 2 and 3 months (Figure 4.5).

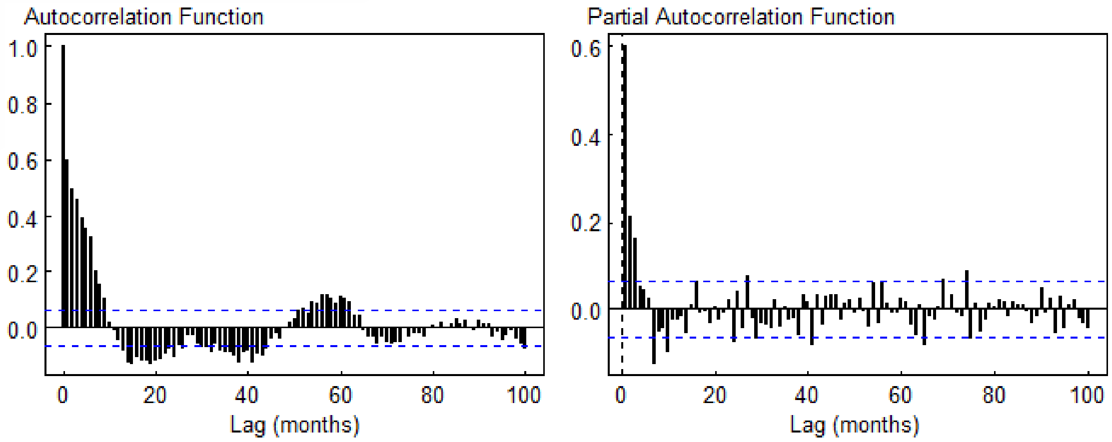

Figure 4.5: ACF (left panel) and PACF (right panel) graphs from linear model for two periods

Results from autoregressive model

Then, those lagged terms are included in AR model giving the results as shown in Table 4.3. As all AR terms at lags 1, 2 and 3 are statistically significant, (z_1 = 0.437, p-value < 0.001; z_2 = 0.139, p-value < 0.001; z_3 = 0.158, p-value < 0.001), we have derived an appropriate model.

Table 4.3: Results from AR(3) for two periods

Parameters	Coefficients	SE	P-value
z_1	0.437	0.033	<0.001
z_2	0.139	0.035	<0.001
z_3	0.158	0.033	<0.001
b_0	1.433	1.038	0.168
b_1	-5.437	1.961	0.006
The p-value of ANOVA test is 0.021			

The ACF graph from this model in the left panel of Figure 4.6 indicates that successive values w_t (Eqn 7, Chapter 2, page 28) are not correlated (white noise), most of the correlation coefficients are close to zero and most of them also are in the 95% confidence intervals. Also, the graph of the PACF in the right panel of Figure 4.6 shows that the model fits the data quite well. The p-value from the ANOVA test for parameters b_0 and b_1 is 0.021 (Table 4.3) indicating that there is appropriat evidence against the null hypothesis and that there is at least one parameter which the coefficient is not equal to zero. It is in accordance with the p-value of less than 0.050 for the third period, (b_1 = -5.437, p-value = 0.006). It can be interpreted that there is no significant trend in the second period and a significant decreasing trend in the third period.

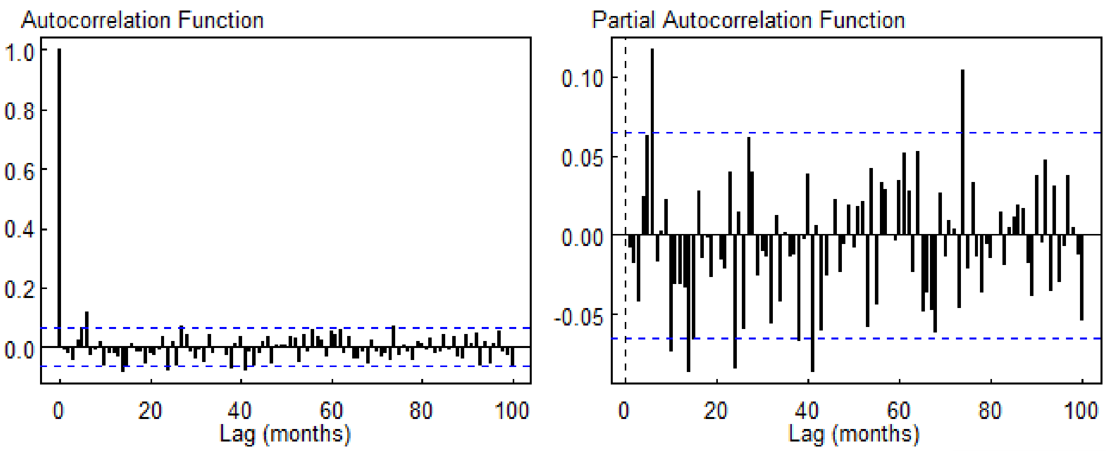

Figure 4.6: ACF (left panel) and PACF (right panel) graphs from AR(3) for two periods

4.3.2 Analysis of three periods of data

Next, the same methods as above are applied with the data from three periods during 1920 to 2014, which include the data from the second, third and fourth periods as shown in Figure 4.7. Likewise, the first period from 1876 to 1919 is presumed to have no trend. In this analysis, two boxcar functions are considered. The first boxcar function (x_1) starts at the change point, that is during the year 1976 to 1995, and during the year 1996 to 2014 for the second boxcar function (x_2).

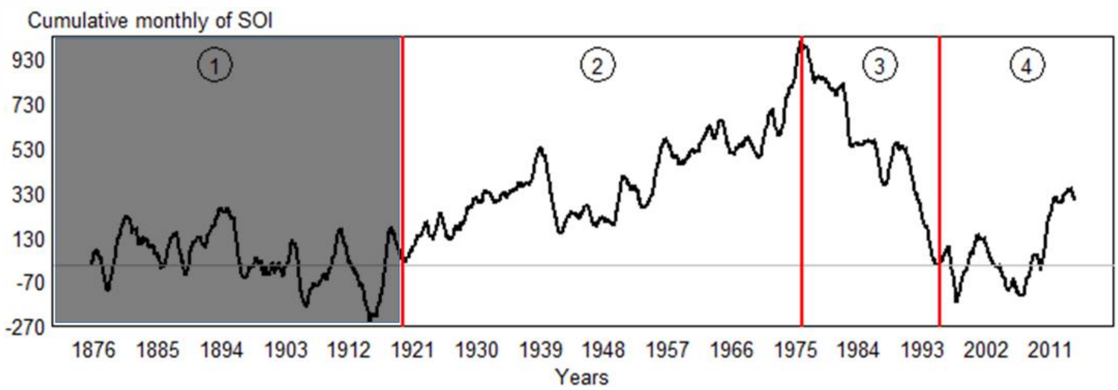

Figure 4.7: Three periods of the data from 1920 to 2014

The null hypothesis is H_0: $b_0 = b_1 = b_2 = 0$, where b_0, b_1 and b_2 refer to the slopes of the second, third and fourth periods, respectively. The linear model from Eqn 5, Chapter 2, page 27 takes the form

$$y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + z_t \ .$$

Results from linear model

Table 4.4 shows the estimated coefficients of the paramters b_0, b_1 and b_2 from the linear model. These results are consistent with those from the previous analysis, and show significant increasing and decreasing trends in the second and third periods, respectively, and no trend in the fourth period. After the ACF and PACF graphs presented in Figure 4.8, are examined, it is again apparent that three lagged terms should be in the AR model.

Table 4.4: Results from linear model for three periods

Parameters	Coefficients	SE	t-value	P-value
b_0	1.409	0.373	3.779	<0.001
b_1	-5.417	0.727	-7.452	<0.001
b_2	-0.242	0.736	-0.329	0.742
The p-value of ANOVA test is <0.001				

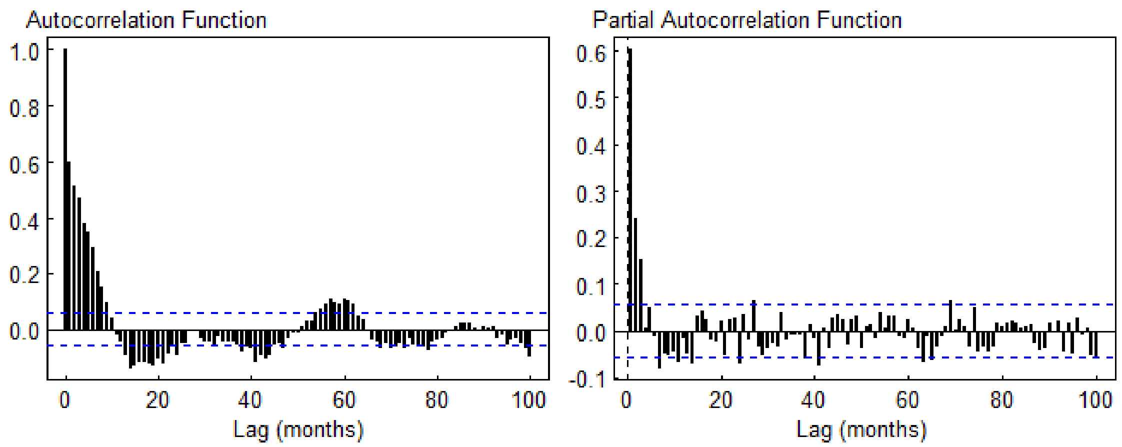

Figure 4.8: ACF (left panel) and PACF (right panel) graphs from linear model for three periods

Results from autoregressive model

Figure 4.9 shows that the model fits the data quite well with presenting uncorrelated w_t by ACF plot. The p-value from the ANOVA test of parameters b_0, b_1 and b_2 from the subsequent AR(3) model is 0.037 (Table 4.5) indicating that there is at least one parameter not equal to zero. The results of this model are shown in Table 4.5. It can be seen that there are insufficient evidence of the trends in the second, (b_0 = 1.442, p-value = 0.188) and the fourth periods, (b_2 = -0.159, p-value = 0.940). However, there has been a significant decreasing trend in the third period, (b_1 = -5.675, p-value = 0.005). Note that, from this dataset the p-value of ANOVA test is bigger than that of the previous dataset from 0.021 to 0.037.

Table 4.5: Results from AR(3) for three periods

Parameters	Coefficients	SE	P-value
z_1	0.420	0.029	<0.001
z_2	0.173	0.031	<0.001
z_3	0.151	0.029	<0.001
b_0	1.442	1.096	0.188
b_1	-5.675	2.039	0.005
b_2	-0.159	2.122	0.940
The p-value of ANOVA test is 0.037			

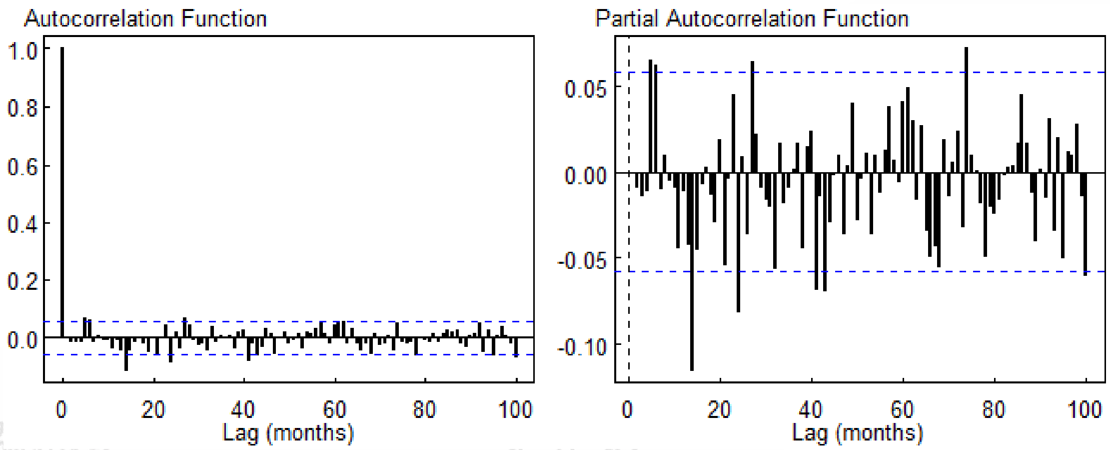

Figure 4.9: ACF (left panel) and PACF (right panel) graphs from AR(3) for three periods

4.3.3 Analysis of four periods of data

In this final analysis, the four periods from 1876 to 2014 (Figure 4.10) are analysed with three boxcar functions. The first boxcar function (x_1) takes value of 1 starting at the changing point during the year 1920 to 1975 and 0 elsewhere, the second boxcar function (x_2) takes value of 1 during the year 1976 to 1995 and 0 for others and the last boxcar function (x_3) takes value of 1 during the year 1996 to 2014 and 0 elsewhere.

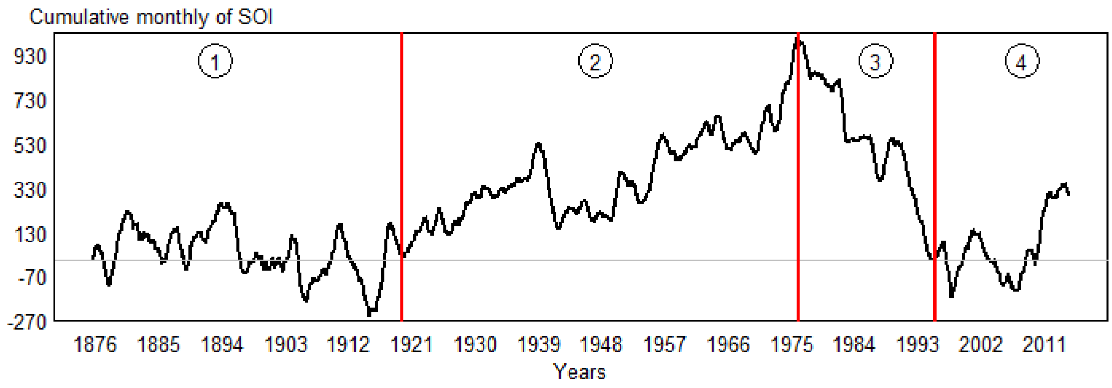

Figure 4.10: Four periods of the data from 1876 to 2014

The null hypothesis is H$_0$: $b_0 = b_1 = b_2 = b_3 = 0$, where b_0, b_1, b_2 and b_3 refer to the slopes of the first, second, third and fourth periods, respectively. The linear model from Eqn 5, Chapter 2, page 27 takes the form

$$y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + b_3 x_{3t} + z_t \, .$$

Results from linear model

Table 4.6 shows the coefficient of parameters b_0, b_1, b_2 and b_3 from the linear model. The model indicates that there is no significant trend in either the first (b_0 = 0.037, p-value = 0.935) or fourth (b_3 = 1.130, p-value = 0.165) periods, but there is a significant increasing trend in the second period, (b_1 = 1.373, p-value = 0.023) and decreasing trend in the third period, (b_2 = -4.045, p-value < 0.001).

Table 4.6: Results from linear model for four periods

Parameters	Coefficients	SE	t-value	P-value
b_0	0.037	0.450	0.081	0.935
b_1	1.373	0.601	2.284	0.023
b_2	-4.045	0.804	-5.028	<0.001
b_3	1.130	0.814	1.389	0.165
The p-value of ANOVA test is <0.001				

The ACF graph of the residuals from the linear regression presented in Figure 4.11 shows that the data are highly serially correlated, and the PACF graph also suggests that the first three lagged terms should be the order in AR model.

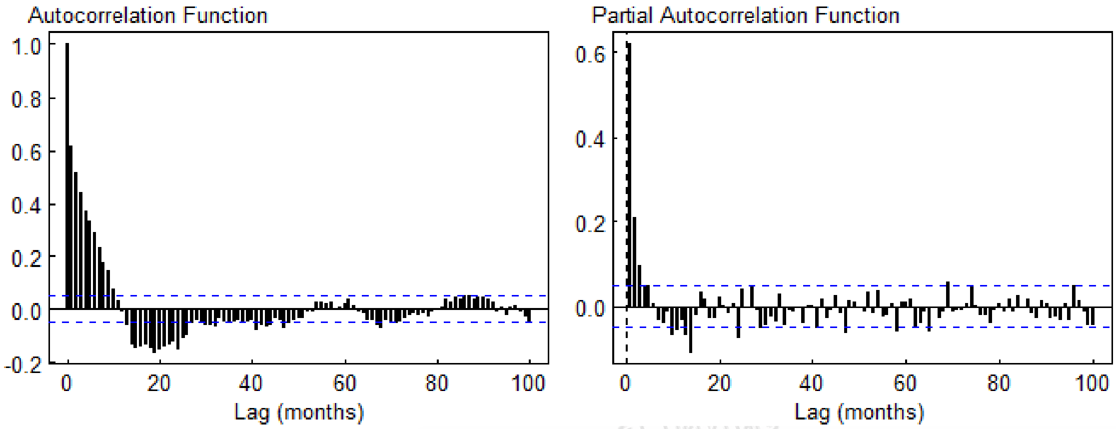

Figure 4.11: ACF (left panel) and PACF (right panel) graphs from linear model for four periods

Results from autoregressive model

Figure 4.12 shows the ACF plot with uncorrelated errors (w_t). The p-value from the ANOVA test for parameters b_0, b_1, b_2 and b_3 from AR model is 0.078 indicating that there is insufficient evidence against the null hypothesis and that the coefficient of each parameter may indeed equal zero. It can be interpreted that there is no trend for all periods, in which it is in accordance with the p-value of more than 0.050 for those four parameters as shown in Table 4.7. The p-value of b_2, third period, is nearly 0.050 which is 0.058. Moreover, from the last dataset, the p-value of ANOVA test is bigger than that the two previous datasets.

Table 4.7: Results from AR(3) for four periods

Parameters	Coefficients	SE	P-value
z_1	0.467	0.024	<0.001
z_2	0.168	0.027	<0.001
z_3	0.095	0.024	<0.001
b_0	-0.083	1.249	0.947
b_1	1.645	1.652	0.319
b_2	-4.142	2.185	0.058
b_3	1.367	2.234	0.541
The p-value of ANOVA test is 0.078			

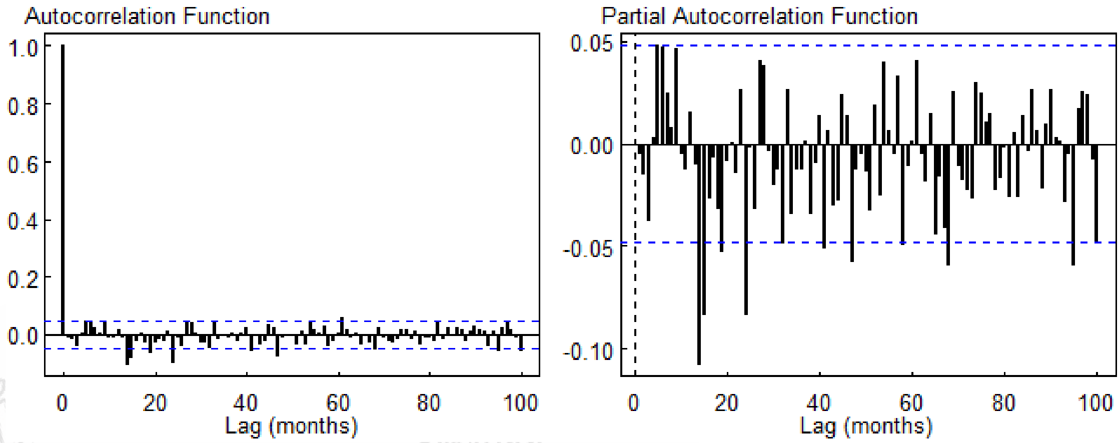

Figure 4.12: ACF (left panel) and PACF (right panel) graphs from AR(3) for four periods

In summary, the correlation coeffients suggest that there are no correlation between monthly SOI and daily rainfall of the regions of north, central, south-east and south-west. Even though, it seems there are heavy rainfall in various years when the SOI express strong positive value, and vice versa, there are a few years that have regular rainfall amount even though the SOI is quite strong negative/positive. To extract more characteristics of the SOI, the raw monthly SOI between 1876 and 2014 shows no an apparent trend, but the cumulative data suggest that the data should be divided into four periods including the first period from 1876 to 1919 with no trend, the second period from 1920 to 1975 with an increasing trend, the third period form 1976 to 1995 with a decreasing trend and the last period from 1996 to 2014 with again no trend. In order to investigate those trends, the linear regression and AR model have been fitted with three data subsets consist of (1) the dataset from 1920 to 1995, (2) the dataset from 1920 to 2014 and (3) the dataset from 1876 to 2014. For the linear regression model, the outcome, SOI, is regressed against boxcar function, where the function model the trends of the SOI. An autoregressive process is used to account for serial correlation in the residuals. Notice, the results from ANOVA test have shown that when the dataset is larger, the p-value also is going up to be more than 0.050. From the analysis, we can conclude that the SOI is quite similar to a random noise process.

In the next chapter, the discussions and conclusions of the two studies have been mentioned.

CHAPTER 5

Discussions and Conclusions

The final chapter presents discussions and conclusions for the thesis covering two studies "Analysis of Daily Rainfall during 2001-2012 in Thailand" and "The Southern Oscillation Index as a Random Walk". The chapter is classified into four sections which are (1) overall finding, (2) discussions, (3) conclusions and (4) limitations and recommendations.

5.1 Overall finding

Daily rainfall in the period of 2001-2012, containing in 114 stations over Thailand is finally divided into four regions based on their spatial correlation investigated by factor analysis technique. They are north; covering areas in north and north-east, central; covering areas in central and east, south-east; covering areas in south-east side and south-wet; covering areas in south-west side. The results from gamma-distributed GLM suggest that there are less variation in daily rainfall in north, central and south-west, but the average rainfall appears as heavy rain in 2011 in the north, while the situation of daily rainfall in south-east fluctuates all times. The unusual heavy rain occurs in 2005, 2008, 2010 and 2011.

When analyzing the correlation between rainfall and the SOI, we found that there is no association between them. Investigating the SOI data using linear regression and autoregressive model, the results suggest that even though the cumulative of SOI has shown an increasing trend between 1920 and 1975, and a decreasing trend between 1976 and 1995, but in the study we can conclude that the SOI data is a random noise.

The SOI may not be able to forecast other variables even itself. Carberry et al (2007) said that "Like familiar biological systems, the weather is not static. This means that the SOI is also continuously changing and fluctuating".

5.2 Discussions

In terms of statistical methods that have been used in this study, it is not frequent to find the study using factor analysis to fit climatic data as well as time series data. Focusing on rainfall data in particular, the previous studies fitted the factor analysis to categorize 12 months in a year by rainfall amount (Fotiadi et al., 1999; Hussain and Lee, 2009; Nikolakis, 2008). Meanwhile, they used principal component analysis (Baeriswyl and Rebetez, 1997; Comrie and Glenn, 1998; Lee, 2005) or cluster analysis (Baeriswyl and Rebetez, 1997; Hussain and Lee, 2009; Puvaneswaran, 1990) to classify geographical rainfall. The present study suggests that factor analysis can also be applied to geographical variable as Cheung et al. (2015) employed this technique to classify 144 stations of solar radiation absorption data over the years of 1990-2012 in Australia. Wanishsakpong and McNeil (2015) used this method to classify temperatures from 85 stations during 1970 to 2012 over Australia and Chooprateep and McNeil (2014) also used this method to classify surface temperature patterns during 1973 to 2008 in Southeast Asia. Moreover, many fields from the previous study had applied the factor analysis to their data, for instance, classifying species of bird assemblage (Rittiboon et al., 2012), grouping species of fish standing crop in the Na Thap River, Southern Thailand (Saheem et al., 2015) and clustering the districts in Nepal (Paudel et al., 2015) and provinces in Thailand (Paudel et al., 2015).

Similarly, the using generalized linear model with gamma regression is quite rare to find from the previous study undertaken, we find that the study from Chandler and Wheater (2002) who observed the daily rainfall in the south Galway region of western Ireland and they found that the model provided a powerful tool for interpreting historical rainfall records. Coe and Stern (1982) also showed the gamma-distributed GLM was fitted the data at the sites Kharja in Jordan, Lunuwila in Sri Lanka, Zinder in Niger and Shakawe in Botswana, the results suggested that the parameters estimating from the models had to be modified. Moreover, the benefit of GLMs is the method can easily produce the 95% confidence intervals of mean.

However, in the second study, we show the ability of the two methods to identify the trends of the cumulative SOI and have finally found the results interesting. They are the linear regression that is used widely for continuous outcome together with covariate variables to find the association between them and autoregressive model is also used commonly in time series analysis. The special situation of this study is we have applied both methods on covariate variable of boxcar function.

In terms of the results of the first study, it is evident that the rainy season of the region in the upper part of Thailand is in the period of the southwest monsoon and also for the south-east in the period of the northeast monsoon. For south-west region, the rainy season covers both the south-west and north-east monsoon. The results from gamma model suggest that a significantly high rainfall occurs in the north part of Thailand in 2011. Likewise, central region shows that rainfall in the rainy season for all years is not different from its overall mean, but in the same year the areas around Bangkok suffered severe flooding starting around the end of July. This was due to five categories of storms experienced by Thailand during June to October in this year.

These included 'Hai Ma', 'Nok Ten', 'Hai Tang', 'Nesat' and 'Nalgae'. They caused heavy rainfall in the upper part of Thailand, which then resulted in the worst flood affecting 684 districts of 65 provinces in and around Bangkok (IPSR, 2012). Even though, only one storm occurred in 1st November, 2010 hitting the South-east region particularly Songkhla province, it effected to wide flash flooding the areas in this region including Chumpon, Nakhon Si Thammarat, Surat Thani, Phatthalung, Songkhla, Pattani, Yala, Narathiwat, Trang, Krabi and Satun (Climatological Center Meteorological Development Bureau, 2011). Also the average rainfall in this year from the gamma model shows the statistical difference from the overall mean.

Considering the El Niño and La Niña events, there are several indices used to measure the occurrence of the events, one of those is monthly SOI data. The Australian Government Bureau of Meteorology (BOM), National Climate Centre, Climate Analysis Section also has emphasized this data (Australian Government Bureau of Meteorology, 2015). The La Niña is detected in the period of July 2010 to April 2011 and the rainfall in south-east has also been unusually high in November 2010 and March 2011. Sometimes, the opposite magnitude has appeared between the SOI and rainfall, such as in February 2008 where the SOI value is highly positive SOI around 21.3, but the rainfall is just 2.16 mm/day at that time. It seems that the La Niña has no effect on the levels of rainfall in Thailand. Also, the correlation coefficients confirm that there is no correlation between the SOI and four regions. Nevertheless, in the study of Singhrattna et al. (2005) found that the SOI is significantly related to the summer rainfall during 1980-2001 in Thailand. Tularam (2010) suggested that there was an apparent relationship between the SOI and rainfall, but the SOI might not be strong enough for longer terms of forecasting rainfall. Bhalme and Jadhav (1984) also revealed the limited skill in the prediction of monsoon rainfall of India from the SOI alone, they said that the atmosphere was extremely complex and it cannot be expected that the SOI could account for most of the monsoon variability. Likewise, Chiew et al. (1998) showed that rainfall and ENSO indicators which were the SOI and sea surface temperature (SST) were statistically significant in most parts of Australia but they were not sufficiently strong to consistently predict rainfall accurately. Moreover, some researches have suggested that the SOI might be used to predict rainfall in Fiji Islands (Dennett et al., 1978) and southern Ghana (Adiku and Stone, 1995). However, Thai Meteorology Department has mentioned that the effect of the La Niña on rainfall in Thailand depends on the strength of that event (Jutakorn, 2010).

The historical data of 12 years just show the pattern and magnitude of rainfall across Thailand. Sometimes we have to learn from the existing data to reveal the pattern and magnitude of rainfall in the previous year for helping the decision of planning and management the situations of the future. The 12 years data may not be enough to use for forecasting rainfall. Likewise, the rainfall depends on many factors such as seasonal wind, low pressure trough, tropical cyclone and the location of the areas, etc. For the next study, the data should be extended for more years and with the same methods used in the present study, we may see the patterns of rainfall more clearly.

For the second study, after the individual monthly SOI is examined in this study, we could say that accumulating time series data over time can give a misleading impression, suggesting that a purely random process appears to have an informative structure. As the results have shown, there is insufficient evidence to conclude that the fluctuations in the SOI are due to anything other than random noise. Likewise, the study of Wu and Huang (2004) used the SOI as noisy data to illustrate their developed method. On the other hand, Chu and Katz (1985) found that the monthly SOI of the current month depends on the SOI of one month and seven or nine months ago using the method of autoregressive moving average (ARMA).

Furthermore, given that the selected periods of possible changes are based on the data, the p-value is not strictly correct, since the null hypothesis is postulated from the same data used to test it. Even though, we could test the null hypotheses from different periods with this study to define the boxcar function on the data by other criteria such as using the period of every 10, 20 or 30 years, etc. to determine whether or not the variability in SOI is just a random process, an appropriate method needs to be based on sound statistical methodology. It is almost impossible to duplicate this method with the same data because the SOI signal is unique. We could use other methods, including bootstrapping, to test the null hypotheses. However, in this study, we emphasize that the cumulative data has appeared to show some trends, but that we do not have sound statistical evidence to verify this.

5.3 Conclusions

Typically, the rainy season of Thailand is in the period of mid-May to mid-October, except South-east where the rainy season is between late September and early January. In 12 years, the average daily rainfall in rainy season of six regions are 11.07, 13.36, 10.96, 13.92, 16.03 and 17.47 mm for North, North-east, Central, East, South-east and South-west, respectively. Also, the average daily rainfall in the summer seems to be similar with that in rainy season. They are 11.26, 10.99, 11.47, 12.37, 12.52 and 13.09 mm for North, North-east, Central, East, South-east and South-west, respectively. In particular, in South-west, the average daily rainfall in the winter also has the amount which is 10.42 mm close to both rainy and summer seasons.

However, in this study, we have reclassified the stations of those regions. Firstly, the data are aggregated to 5-day averages to remove the serial correlation because the conventional statistical models assume independent errors. After that we investigate correlation between stations and found the spatial correlation among close rainfall stations. Thus, the approach of factor analysis is applied to allocate 114 station across Thailand to be new group depending on the correlation between the stations. Thus, this method is able to separate the regions of Thailand with overall 114 weather stations into seven clusters, comprising of upper north, lower north, north-east, central, upper north, south-west and south-east with the total variance of 58.9%.

Then, the successive days with an average rainfall over 4.43 mm/day (overall mean) are specified as a rainy season for each region. Finally, after investigating the correlation coefficient between the regions, the seven regions are further reduced into four regions, namely north, central, south-east and south-west. The results suggest that the rainy season is in the period of April to October for north and central (6 months), late September to early January (4 months) for south-east and April to November for south-west (8 months).

Since, the average daily rainfall of four regions is skewed to the right, gamma-distributed generalized linear model is fitted to those data. The results show that four models are capable of fitting data reasonably well examining by normal quantile plot of deviance residuals. The models suggest that, in terms of month effects, in north, the average rainfall in August and September are statistically significantly different (above) from its overall mean in its rainy season. In central, the highest rainfall occurs in September followed by October. In south-east, the highest rainfall occurs in November and in September for south-west. Likewise, the year effects have been addressed that the amounts of rainfall in the year 2011 in north is above the overall mean. In south-east, the levels of rainfall in the year 2005 and 2010 are also above its overall mean. In the year 2008 and 2011 though the rainfall seems to be quite high, but it is not statistically significantly. In central and south-west, daily rainfall has less variation in 12 years. Furthermore, the average daily rainfall in rainy season of the regions are 7.59, 6.30, 11.04 and 9.78 mm for north, central, south-east and south-west, respectively.

Interestingly, it is found from several studies that the monthly SOI and rainfall data are related (e.g. Hadiani et al., 2012; Tigona and de Freitas, 2012). In this study, we found that there are no correlation between the SOI and daily rainfall of all four regions of Thailand. The correlation coefficients have been revealed as -0.13, -0.05, 0.18 and -0.05 for north, central, south-east and south-west, respectively.

Moreover, the previous studies found that there were the relationship between the SOI and temperature (e.g. Halpert and Ropelewski, 1992; Jones and Trewin, 2000). Otherwise, the study of Al-Zuhairi et al. (2013) found that there was no relationship detected between temperature and the SOI from 1900 to 2008 over Iraq. Therefore, the characteristics of the SOI are examined in the second study. Actually, the raw SOI data do not present any trend, but cumulative SOI data show apparent trends that can be further divided into four periods comprising the years (1) 1876-1919, (2) 1920-1975, (3) 1976-1995 and (4) 1996-2014. We have investigated the apparent increasing and decreasing trends in the second and third periods, respectively.

In order to examine those trends, the three datasets have been analysed using the linear regression and autoregressive model on the boxcar function. The results from this study show that the appropriate autoregressive model for each of the three datasets is AR(3). For the model which includes only the data from periods two and three, we found no significant trend in the period 1920 to 1975 but a significant decreasing trend in the period 1976 to 1995. However, the results from the analysis of the data including all four periods show no evidence of trend for any periods. The p-value from ANOVA test is going up to be more than 0.05 when the dataset is larger. Consequently, the evidence suggests that the fluctuations in the SOI seem to be similar with a random noise process.

5.4 Limitations and recommendations

The present study shows the powerful of existing statistical methods which can be applied to time series data. The daily rainfall analysis is based on modeling of month and year effects only. We suggest that, the models would be more valuable if some more variables, such as temperature, altitude of the stations, wind velocity and air pressure, etc. are accounted. The models may give more information and may be able to predict the daily rainfall. Also, the extended data may show some patterns of daily rainfall in Thailand with the SOI data. Moreover, the methods used in this study could be applied to other either data or places to compare those results with this study.

As the results in Table 3.1 (Chapter 3), the dry spell per year takes quite long time. It is an interesting topic to demonstrate for the next study. The data might also be modelled in two stages (Chandler et al. and Wheater, 2002). Firstly, the pattern of occurrence of wet and dry days are concerned using logistic regression and secondly only wet days are chosen to model using gamma regression for computing the amount of rainfall. Moreover, the data in format of weekly total, monthly total, yearly total and extreme rainfall also prove very interesting to be established through those models.

Lastly, mixed model is one of the interesting models that could be applied to explore the rainfall data for obtaining more accurate results. The mixed model comprises fixed and random factors such as in the present study season and region might be treated as fixed and random variables, respectively.

References

Adiku, S.G.K. and Stone, R.C. 1995. Using the Southern Oscillation Index for improving rainfall prediction and agricultural water management in Ghana. Agricultural Water Management, 29, 85-100.

Aksoy, H. 2000. Use of gamma distribution in hydrological analysis. Turkish Journal of Engineering and Environmental Sciences. 24, 419-428.

Al-Zuhairi, M.F., Kadhum, J.H. and Farhan, A.J. 2013. Time series analysis of Southern Oscillation Index and temperature over Iraq. Diyala journal for pure scinces. 9, 57-68.

Assanangkornchai, S., Tangboonngam, S. and Edwards, J.G. 2004. The flooding of Hat Yai: predictors of adverse emotional responses to a natural disaster. Stress and Health. 20, 81-89. doi: 10.1002/smi.999.

Australian Government Bureau of Meteorology. 2005. El Nino, La Nina and Australia's Climate. Available from: http://www.bom.gov.au/info/leaflets/nino-nina.pdf. [March 27, 2014].

Australian Government Bureau of Meteorology. 2015. Southern Oscillation Index. Available from: http://www.bom.gov.au/climate/glossary/soi.shtml. [April 30, 2015].

Baeriswyl, P.A. and Rebetez, M. 1997. Regionalization of Precipitation in Switzerland by Means of Principal Component Analysis. Theoretical and Applied Climatology. 58, 31-41.

Bhalme, H.N. and Jadhav, S.K. 1984. The southern oscillation and its relation to the monsoon rainfall. Journal of Climatology. 4, 509-520.

Carberry, P. George, D. and Buckley, D. 2007. The southern oscillation index and southern Australia, Primefacts 615, NSW department of primary industries, May 2007, Available online: http://www.dpi.nsw.gov.au/__data/assets/pdf_file/0007/138139/the-southern-oscillation-index-and-southern-australia.pdf. [April 30, 2015].

Castañeda, M. and González, M. 2008. Statistical analysis of the precipitation trends in the Patagonia region in southern South America. Atmósfera. 21, 303-317.

Chambers, L.E. 2003. South Australian rainfall variability and trends. Bureau of Meteorology Research Center Australia. BMRC research report no. 92.

Chandler, R.E. and Wheater, H.S. 2002. Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland. Water Resources Research. doi: 10.1029/2001WR000906.

Chatfield, C. 1996. The Analysis of Time Series An Introduction. Fifth ed. Chapman and Hall/CRC, London, U.K., pp.18-24, 30-40.

Cheung, K.K.W., Chooprateep, S. and Ma, J. 2015. Spatial and temporal patterns of solar absorption by clouds in Australia as revealed by exploratory factor analysis. Solar Energy. 111, 53-67.

Chiew, F.H.S., Piechota, T.C., Dracupm, J.A. and McMahon, T.A. 1998. E1 Nino/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting. Journal of Hydrology. 204, 138-149.

Cho, H., Bowman, K.P. and North, G.R. 2004. A Comparison of Gamma and Lognormal Distributions for Characterizing Satellite Rain Rates from the Tropical Rainfall Measuring Mission. Journal of Applied Meteorology. 43, 1586-1597.

Chooprateep, S. and McNeil, N. 2014. Temperature changes in Southeast Asia: 1973-2008. Chiang Mai University Journal of Natural Sciences. 13, 105-116. doi: 10.12982/cmujns.2014.0025.

Chu, P. and Katz, R.W. 1985. Modeling and forecasting the Southern Oscillation: A time-domain approach*. Monthly Weather Review. 113, 1876-1888.

Cinco, T.A., de Guzman, R.G., Hilario, F.D. and Wilson, D.M. 2014. Long-term trends and extremes in observed daily precipitation and near surface air temperature in the Philippines for the period 1951–2010. Atmospheric Research. doi: http://dx.doi.org/10.1016/j.atmosres.2014.03.025.

Climatological Center Meteorological Development Bureau, 2011, Weather of Thailand in 2010. Available online: http://www.joelandchoom.net/ tmdsummaries/T_weather2553.pdf [November 11, 2015].

Coe, R. and Stern, R.D. 1982. Fitting models to daily rainfall data. Journal of Applied Meteorology. 21, 1024-1031.

Comrie, A.C. and Glenn, E.C. 1998. Principal components-based regionalization of precipitation regimes across the southwest United States and northern Mexico, with an application to monsoon precipitation variability. Climate Research. 10, 201-215.

Dennett, M.D., Elston, J. and Prasad, P.C. 1978. Seasonal rainfall forecasting in Fiji and the Southern Oscillation. Agricultural Meteorology. 19, 11-22.

Doty, R.D. 1982. Annual Precipitation on the Island of Hawaii between 1890 and 1977. Pacific Science. 36, 421-425.

Ejieji, C.J. 2004. Evaluation of two models for the distribution of daily rainfall amounts in Nigeria. Journal of Agricultural Research and Development. 3, 61-74.

Faraway, J.J. 2005. Linear Models with R, Chapman and Hall/CRC, Florida, U.S.A., pp.11-14.

Feng, S., Nadarajah, S. and Hu, Q. 2007. Modeling annual extreme precipitation in China using the Generalized Extreme Value distribution. Journal of the Meteorological Society of Japan. 85, 599-613.

Fotiadi, A.K., Metaxas, D.A. and Bartzokas, A. 1999. A statistical study of precipitation in Northwest Greece. International Journal of Climatology. 19, 1221-1232.

Gandomkar, A., Gord faramarzi, T.S., Chafi, P.S. and Amani, A. 2011. Analysis of precipitation and temperature trends in Sefid-Roud basin. World Academy of Science, Engineering and Technology. 59, 974-978.

Gill, J. 2001. Generalized Linear Models: A Unified Approach, Sage Publications, Inc., Thousand Oaks, California, pp. 29-39.

Hadiani, M., Asl, S.J., Banafsheh, M.R., Dinpajouh, Y. and Yasari, E. 2012. Investigation the Southern Oscillation Index Effect on Dry/Wet Periods in North of Iran. International Journal of Agriculture and Crop Sciences. 4, 1291-1299.

Halpert, M.S. and Ropelewski, C.F. 1992. Surface temperature patterns associated with the Southern Oscillation. Journal of climate. 5, 577-593.

Hung, N.Q., Babel, M.S., Weesakul, S. and Tripathi, N.K. 2009. An artificial neural network model for rainfall forecasting in Bangkok Thailand. Hydrology and Earth System Sciences. 13, 1413-1425.

Hussain, M.S. and Lee, S. 2009. A classification of rainfall regions in Pakistan. Journal of the Korean Geographical Society. 44, 605-623.

Institute for Population and Social Research, Mahidol University (IPSR). 2012. Food Security the Illusion of Money VS the Reality of Food. Thai Health. 9, 34-40.

Johnson, R.A. and Wichern, D.W. 2007. Applied Multivariate Statistical Analysis, Pearson Prentice Hall, New Jersey, U.S.A., pp. 481-526.

Jones, D.A. and Trewin, B.C. 2000. On the relationships between the El Niño-Southern Oscillation and Australian land surface temperature. International journal of climatology. 20, 697-719.

Jutakorn, J. 2010. El Niño and La Niña phenomenon effected on southwest monsoon in Thailand. Meteorological department. Weather report no. 551.557-21-01-2010.

Kate-Eiam, S. 2003. Forecasting Technique. Thaksin University, Songkhla, Thailand, pp. 202-209.

Keggenhoff, I., Elizbarashvili, M., Amiri-Farahani, A. and King, L. 2014. Trends in daily temperature and precipitation extremes over Georgia, 1971–2010. Weather and Climate Extremes. 4, 75-85.

Lee, C.Y. 2005. Application of rainfall frequency analysis on studying rainfall distribution characteristics of Chia-Nan plain area in southern Taiwan, Environment and Bio Informatics. 2, 31-38.

Leewatchanaku, K. 2000. Hydrology, First ed. Siam Stationery Supplies Ltd., Part, Bangkok, pp. 3.10-3.23.

Longobardi, A. and Villani, P. 2009. Trend analysis of annual and seasonal rainfall time series in the Mediterranean area. International Journal of Climatology. doi: 10.1002/joc.2001.

McCullagh, P. and Nelder, J.A. 1989. Generalized Linear Models, Second ed., Chapman and Hall/CRC, Washington, D.C., U.S.A., pp. 54, 287-306.

Nikolakis, D. 2008. A statistical study of precipitation in Cyprus. Hellenic Journal of Geosciences. 43, 67-74.

Paudel, B.K., Lim, A. and Dureh, N. 2015. Population distribution pattern of 76 provinces in Thailand: Application of factor analysis. Asian Social Science. 11, 55-61.

Paudel, B.K., Lim, A. and Guragain, A.M. 2015. Analyzing age-sex structure patterns in Nepal using factor analysis. Songklanakarin Journal of Science and Technology. 37, 485-491.

Phien, H.N., Arbhabhirama, A. and Sunchindah, A. 1980. Distribution of monthly rainfall in northeast Thailand, Southeast Asian Studies. 18, 110-123.

Puvaneswaran, M. 1990. Climatic classification for Queensland using multivartate statistical techniques. International Journal of Climatology. 10, 591-608.

Quinn, W.H., Zopf, D.O., Short, K.S. and Kuo Yang, R.T.W. 1978. Historical trends and statistics of the southern oscillation, El Nino, and Indonesian droughts. Fishery Bulletin. 76, 663-678.

Rao, V.B. and Hada, K. 1990. Characteristics of rainfall over Brazil: Annual variations and connections with the Southern Oscillation. Theoretical and Applied Climatology. 42, 81-91.

Renhe, Z., Sumi, A. and Kimoto, M. 1999. A diagnostic study of the impact of El Nino on the precipitation in China. Advances in Atmospheric Sciences. 16, 229-241.

Rencher, A.C. 2002. Methods of Multivariate Analysis, John Wiley and Sons, Inc., New York, U.S.A., pp. 408-447.

Rittiboon, K., Tongkumchum, P. and Karntanut, W. 2012. Bird assemblages in the Thale Noi non-hunting area, southern Thailand. Journal of Sustainability Science and Management. 7, 42-48.

Ropelewski, C.F. and Halpert, M.S. 1987. Global and regional scale precipitation patterns associated with the E1 Nino/Southern Oscillation. Monthly Weather Review. 115, 1606-1626.

R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Austria: R Foundation for Statistical Computing. Available from: http://www.R-project.org. [April 18, 2014].

Saheem, N., Chesoh, S. and Lim, A. 2015. Factor analysis for clustering and estimating fish distribution pattern in a tropical estuary in southern Thailand. Kasetsart Journal: Natural Science. 49, 188-199.

Schumacker, R.E. 2015. Learning Statistics Using R. Sage Publications, Inc., Washington, D.C., U.S.A., pp. 270-281.

Singhrattna, N., Rajagopalan, B., Clark, M. and Kumar, K.K. 2005. Seasonal forecasting of Thailand summer monsoon rainfall. International Journal of Climatology. 25, 649-664.

Stoeckenius, T. 1981. Interannual variations of tropical precipitation patterns. Monthly Weather Review. 109, 1233-1247.

Stone, R. and Auliciems, A. 1992. SOI phase relationships with rainfall in eastern Australia. International Journal of Climatology. 12, 625-636.

Suppiah, R. 1989. Relationships between the Southern Oscillation and the rainfall of Sri Lanka. International Journal of Climatology. 9, 601-618.

Szyniszewska, A.M. and Waylen, P.R. 2012. Determining the daily rainfall characteristics from the monthly rainfall totals in central and northeastern Thailand. Applied Geography. 35, 377-393.

Taxak, A.K., Murumkar, A.R. and Arya, D.S. 2014. Long term spatial and temporal rainfall trends and homogeneity analysis in Wainganga basin, Central India. Weather and Climate Extremes. 4, 50-61.

Thai Meteorological Department. 2012. The climate of Thailand. Available online: www.tmd.go.th/en/archive/thailand_climate.pdf [August 18, 2015].

The World Bank. 2012. Thai Flood 2011 Rapid Assessment for Resilient Recovery and Reconstruction Planning. Available online: https://www.gfdrr.org/sites/gfdrr/files/publication/Thai_Flood_2011_2.pdf [August 18, 2015].

Tigona, R.S. and de Freitas, C.R. 2012. Relationship between the Southern Oscillation and rainfall in Vanuatu. Weather and Climate. 32, 43-50.

Tongkumchum, P. and McNeil, D. 2009. Confidence intervals using contrasts for regression model. Songklanakarin Journal of Science and technology. 31, 151-156.

Tularam, G.A. 2010. Relationship between El Niño Southern Oscillation Index and rainfall (Queensland, Australia). International Journal of Sustainable Development and Planning. 5, 378-391.

Venables, W.N. and Ripley, B.D. 2002. Modern Applied Statistics with S, Springer Science and Business Media, LLC, New York, U.S.A., pp. 139-145, 183-189, 321-325.

Wanishsakpong, W. and McNeil, N. 2015. Modelling of daily maximum temperatures over Australia from 1970 to 2012. Meteorological Applications. doi: 10.1002/met.1536.

Wangwongwiroj, N. 2008. Hydrology, First ed. King Mongkut's University of Technology Thonburi, Bangkok, Thailand, pp. 57-71.

Watts, A. 2010. The story told by the Southern Oscillation Index. Available: http://wattsupwiththat.com/2010/12/09/the-story-told-by-the-southern-oscillation-index/. [April 04, 2015].

Weesakul, U. and Lowanichchai, S. 2005. Rainfall Forecast for Agricultural Water Allocation Planning in Thailand. Thammasat International Journal of Science and Technology. 10, 18-27.

Weisstein, E.W. 2002. CRC Concise Encyclopedia of Mathematics. Second ed. Chapman and Hall/CRC, New York, U.S.A., pp. 276.

Weisstein, E.W. 2015. Boxcar function. MathWorld. Available: http://mathworld.wolfram.com/BoxcarFunction.html. [April 04, 2015].

Wu, Z. and Huang, N.E. 2004. A study of the characteristics of white noise using the empirical mode decomposition method. Proceedings of the Royal society of London A, 460, 1597–1611. doi: 10.1098/rspa.2003.1221.

Xu, Z.X., Takeuchi, K. and Ishidaira, H. 2004. Correlation between El Niño-Southern Oscillation (ENSO) and precipitation in South-east Asia and the Pacific region. Hydrological Processes. 18, 107-123.

Appendix I

The location of 123 stations

No	Station code	Station name	Latitude	Longitude	UTM (m)		Province	Province code	Region
					x	y			
1	300201	Mae Hong Son	19.18.0	97.50.0	377423.52	2134434.84	Mae Hong Son	58	North
2	300202	Mae Sariang	18.10.0	97.56.0	387181.14	2008952.50	Mae Hong Son	58	North
3	303201	Chiang Rai	19.57.41	99.52.53	592223.15	2207450.78	Chiang Rai	57	North
4	303301	Chiang Rai Agromet	19.52.15	99.46.58	581951.20	2197378.12	Chiang Rai	57	North
5	310201	Phayao	19.8.0	99.54.0	594652.28	2115824.27	Phayao	56	North
6	327301	Mae Jo	18.55.0	99.0.0	500000.00	2091606.84	Chiang Mai	50	North
7	327501	Chiang Mai	18.47.24	98.58.37	497570.38	2077591.84	Chiang Mai	50	North
8	328201	Lampang	18.17.0	99.31.0	554608.27	2021610.15	Lampang	52	North
9	328301	Lampang Agromet	18.19.0	99.17.0	529940.51	2025244.15	Lampang	52	North
10	329201	Lamphun	18.34.0	99.2.0	503517.31	2052881.51	Lamphun	51	North
11	330201	Phrae	18.10.0	100.10.0	623396.76	2009016.79	Phrae	54	North
12	331201	Nan	18.46.47	100.46.40	687379.75	2077390.56	Nan	55	North
13	331301	Nan Agromet	18.52.0	100.45.0	684356.34	2086985.13	Nan	55	North
14	331401	Tha Wang Pha	19.6.38	100.48.9	689612.03	2114036.95	Nan	55	North
15	331402	Thung Chang	19.24.43	100.53.7	697959.91	2147491.94	Nan	55	North
16	351201	Uttaradit	17.37.0	100.6.0	616703.79	1948114.94	Uttaradit	53	North
17	373201	Sukhothai	17.6.22	99.48.0	585108.83	1891468.42	Sukhothai	64	North
18	373301	Si Samrong Agromet	17.10.0	99.52.0	592171.80	1898198.45	Sukhothai	64	North
19	376201	Tak	16.52.42	99.8.36	515266.69	1866101.29	Tak	63	North
20	376202	Mae Sot	16.39.33	98.33.3	452103.26	1841904.80	Tak	63	North
21	376203	Bhumibol Dam	17.14.0	99.3.0	505315.53	1905368.52	Tak	63	North
22	376301	DoiMuser Agromet Stn.	16.45.0	98.56.0	492894.44	1851900.32	Tak	63	North
23	376401	Umphang	16.0.57	98.51.56	485616.70	1770691.43	Tak	63	North
24	378201	Phitsanulok	16.47.0	100.16.0	634991.29	1856017.46	Phitsanulok	65	North
25	379201	Phetchabun	16.26.0	101.9.0	729575.12	1818088.23	Phetchabun	67	North

No	Station code	Station name	Latitude	Longitude	UTM (m)		Province	Province code	Region
					x	y			
26	379401	LomSak	16.46.25	101.14.58	739775.45	1855869.89	Phetchabun	67	North
27	379402	Wichian Buri	15.39.25	101.6.30	725998.63	1732111.62	Phetchabun	67	North
28	380201	Kamphaeng Phet	16.29.0	99.32.0	556923.67	1822475.40	Kamphaeng Phet	62	North
29	386301	Pichit Agromet	16.26.17	100.17.33	637991.19	1817832.12	Pichit	66	North
30	352201	Nong Khai	17.52.0	102.43.0	258047.00	1976913.85	Nong Khai	43	North-east
31	353201	Loei	17.27.0	101.44.0	790332.52	1931415.21	Loei	42	North-east
32	353301	Loei Agromet	17.24.0	101.44.0	790411.68	1925878.53	Loei	42	North-east
33	354201	Udon Thani	17.23.0	102.48.0	266258.65	1923303.26	Udon Thani	41	North-east
34	356201	Sakon Nakhon	17.9.0	104.8.0	407819.97	1896354.49	Sakon Nakhon	47	North-east
35	356301	Sakon Nakhon Agromet	17.7.0	104.3.0	398937.85	1892707.97	Sakon Nakhon	47	North-east
36	357201	Nakhon Phanom	17.25.0	104.47.0	476988.82	1925662.75	Nakhon Phanom	48	North-east
37	357301	Nakhon Phanom Agromet	17.26.0	104.47.0	476990.91	1927506.59	Nakhon Phanom	48	North-east
38	360201	Nongbualumphu	17.10.53.79	102.26.18.59	227533.88	1901445.64	Nongbualumphu	39	North-east
39	381201	Khon Kaen	16.27.48	102.47.12	263695.74	1821481.81	Khon Kaen	40	North-east
40	381301	Tha Phra Agromet	16.20.0	102.49.0	266745.10	1807057.70	Khon Kaen	40	North-east
41	383201	Mukdahan	16.32.0	104.43.0	469767.34	1827952.48	Mukdahan	49	North-east
42	387401	Kosum Phisai	16.14.50	103.4.5	293521.08	1797256.26	Mahasarakam	44	North-east
43	388401	Kamalasai	16.19.57	103.35.18	349202.94	1806237.72	Kalasin	46	North-east
44	403201	Chaiyaphum	15.48.0	102.2.0	182156.06	1749054.65	Chaiyaphum	36	North-east
45	405201	Roi Et	16.3.0	103.41.0	359152.28	1774913.60	Roi Et	45	North-east
46	405301	Roi Et Agromet	16.4.0	103.37.0	352031.91	1776804.12	Roi Et	45	North-east
47	407301	Ubon Ratchathani Agromet	15.23.33	105.3.33	506349.11	1701739.86	Ubon Ratchathani	34	North-east
48	407501	Ubon Ratchathani	15.15.0	104.52.0	485682.45	1685981.79	Ubon Ratchathani	34	North-east
49	409301	Si Sa Ket Agromet	15.2.0	104.15.0	419379.90	1662149.68	Si Sa Ket	33	North-east
50	431201	Nakhon Ratchasima	14.57.46	102.4.36	185546.48	1656281.57	Nakhon Ratchasima	30	North-east
51	431301	Pak Chong Agromet	14.38.38	101.19.15	749975.08	1620219.57	Nakhon Ratchasima	30	North-east
52	431401	Chok Chai	14.43.8	102.10.7	195101.71	1629149.48	Nakhon Ratchasima	30	North-east

No	Station code	Station name	Latitude	Longitude	UTM (m)		Province	Province code	Region
					x	y			
53	432201	Surin	14.53.0	103.30.0	338635.59	1645964.83	Surin	32	North-east
54	432301	Surin Agromet	14.53.0	103.27.0	333255.64	1646001.62	Surin	32	North-east
55	432401	Tha Tum	15.19.0	103.41.0	358648.19	1693780.30	Surin	32	North-east
56	436201	Burirum	15.13.0	103.14.0	310236.63	1683058.57	Burirum	31	North-east
57	436401	Nang Rong	14.35.0	102.48.0	262980.60	1613387.80	Burirum	31	North-east
58	400201	Nakhon Sawan	15.48.0	100.10.0	624954.31	1747159.12	Nakhon Sawan	60	Center
59	400301	Tak Fa Agromet	15.21.0	100.30.0	661011.83	1697596.16	Nakhon Sawan	60	Center
60	402301	Chai Nat	15.9.0	100.11.0	627135.91	1675259.92	Chai Nat	18	Center
61	415301	Ayuttaya Agromet	14.31.0	100.43.0	684986.86	1605563.10	Ayuttaya	14	Center
62	419301	Pathumthani Agromet	14.6.0	100.37.0	674529.88	1559385.82	Pathumthani	13	Center
63	424301	Ratchaburi	13.29.23	99.47.32	585740.54	1491431.57	Ratchaburi	70	Center
64	425201	SuphanBuri	14.28.28	100.8.20	622740.10	1600503.48	Suphan Buri	72	Center
65	425301	U Thong Agromet	14.18.0	99.52.0	593472.47	1581079.85	Suphan Buri	72	Center
66	426201	Lop Buri	14.48.0	100.37.0	673984.44	1636832.43	Lop Buri	16	Center
67	426401	Bua Chum	15.15.50	101.11.30	735378.87	1688699.21	Lop Buri	16	Center
68	429201	Pilot Station	13.22.38	100.35.58	673200.46	1479411.34	Samutprakarn	11	Center
69	429601	Suwanabhum Airport	13.41.11	100.46.3	691156.81	1513740.71	Samutprakarn	11	Center
70	450201	Kanchanaburi	14.1.21	99.32.9	557860.17	1550280.26	Kanchanaburi	71	Center
71	450401	Thong Pha Phum	14.44.32	98.38.11	460862.42	1629846.55	Kanchanaburi	71	Center
72	451301	Kamphaeng Saen Agromet	14.1.0	99.58.0	604387.96	1549782.86	Nakhon Pathom	73	Center
73	455201	Bangkok Metropolis	13.43.35	100.33.36	668681.64	1518011.61	Bangkok	10	Center
74	455203	Klong Toey	13.42.25	100.34.5	669566.84	1515866.12	Bangkok	10	Center
75	455301	Bang Na	13.40.0	100.37.0	674854.55	1511444.76	Bangkok	10	Center
76	455302	Bang Khen	13.51.0	100.35.0	671115.14	1531703.23	Bangkok	10	Center
77	455601	Donmuang	13.55.9	100.36.18	673405.80	1539370.93	Bangkok	10	Center
78	423301	Chacherngsao Agromet	13.30.56	101.27.30	766097.23	1495484.96	Chacherngsao	24	East
79	430201	Prachin Buri	14.3.0	101.22.0	755587.37	1554538.10	Prachin Buri	25	East

No	Station code	Station name	Latitude	Longitude	UTM (m)		Province	Province code	Region
					x	y			
80	430401	Kabin Buri	13.59.0	101.42.26	792472.87	1547553.62	Prachin Buri	25	East
81	440201	Aranyaprathet	13.42.0	102.35.0	238618.01	1515854.31	Sa Kaew	27	East
82	440401	Sa Kaew	13.47.20	102.2.5	179359.43	1526357.56	Sa Kaew	27	East
83	459201	Chon Buri	13.22.0	100.59.0	714793.82	1478544.28	Chon Buri	20	East
84	459202	Ko Sichang	13.9.42	100.48.7	695307.11	1455713.42	Chon Buri	20	East
85	459203	Phatthaya	12.55.12	100.52.10	702822.45	1429028.87	Chon Buri	20	East
86	459204	Sattahip	12.41.0	100.59.0	715383.92	1402935.90	Chon Buri	20	East
87	459205	Lam Chabang	13.4.37	100.52.33	703387.91	1446398.25	Chon Buri	20	East
88	478201	Rayong	12.37.56	101.20.37	754578.04	1397604.24	Rayong	21	East
89	478301	Huai Pong Agromet	12.44.0	101.8.0	731634.73	1408597.18	Rayong	21	East
90	480201	Chanthaburi	12.37.0	102.6.48	186371.62	1396471.73	Chanthaburi	22	East
91	480301	Phriu Agromet	12.30.31	102.10.23	192736.64	1384438.62	Chanthaburi	22	East
92	501201	Khlong Yai	11.46.0	102.53.0	269336.88	1301621.46	Trat	23	East
93	465201	Phetchaburi	12.59.58	100.3.38	615014.29	1437313.75	Phetchaburi	76	South-east
94	500201	Prachuap Khiri Khan	11.50.0	99.50.0	590774.21	1308259.45	Prachuap Khiri Khan	77	South-east
95	500202	Hua Hin	12.35.10	99.57.45	604549.54	1391557.40	Prachuap Khiri Khan	77	South-east
96	500301	Nong Phlup Agromet	12.35.0	99.44.0	579656.22	1391169.92	Prachuap Khiri Khan	77	South-east
97	517201	Chumphon	10.29.0	99.11.0	520062.14	1158858.15	Chumphon	86	South-east
98	517301	Sawi Agromet	10.20.0	99.6.0	510948.20	1142269.08	Chumphon	86	South-east
99	551201	Surat Thani	9.8.8	99.9.7	516694.57	1009842.39	SuratThani	84	South-east
100	551202	Phunphin Airport	9.12.31	99.17.50	532655.65	1017929.42	SuratThani	84	South-east
101	551203	Ko Samui	9.28.0	100.3.0	615264.12	1046619.93	SuratThani	84	South-east
102	551301	Surat Thani Agromet	9.6.0	99.38.0	569594.32	1005968.76	SuratThani	84	South-east
103	551401	Phra Sang	8.34.0	99.16.0	529344.55	946954.78	SuratThani	84	South-east
104	552201	Nakhon Si Thammarat	8.32.16	99.56.50	604246.33	943878.74	Nakhon Si Thammarat	80	South-east
105	552202	Khanom	9.14.35	99.51.27	594190.85	1021837.12	Nakhon Si Thammarat	80	South-east
106	552301	Nakhorn Sri Thammarat Agromet	8.3.0	100.0.0	610191.13	889960.11	Nakhon Si Thammarat	80	South-east

No	Station code	Station name	Latitude	Longitude	UTM (m)		Province	Province code	Region
					x	y			
107	552401	Chawang	8.25.55	99.30.43	556355.61	932087.40	Nakhon Si Thammarat	80	South-east
108	560301	Phatthalung Agromet	7.35.0	100.10.0	628701.52	838408.10	Phatthalung	93	South-east
109	568301	Kho Hong Agromet	7.0.0	100.30.0	665695.03	774013.24	Songkhla	90	South-east
110	568401	Sa Dao	6.47.53	100.23.26	653668.13	751645.47	Songkhla	90	South-east
111	568501	Songkhla	7.12.14	100.36.17	677187.76	796599.39	Songkhla	90	South-east
112	568502	Hat Yai Airport	6.55.0	100.26.0	658357.21	764775.24	Songkhla	90	South-east
113	580201	Pattani Airport	6.47.0	101.9.0	737632.14	750324.03	Pattani	94	South-east
114	581301	Yala Agromet	6.31.0	101.17.0	752512.65	720889.84	Yala	95	South-east
115	583201	Narathiwat	6.25.0	101.49.0	811597.26	710120.65	Narathiwat	96	South-east
116	532201	Ranong	9.59.0	98.37.0	457985.95	1103594.13	Ranong	85	South-west
117	561201	Takua Pa	8.41.3	98.15.8	417736.47	960015.88	Phang-nga	82	South-west
118	564201	Phuket	7.53.0	98.24.0	433860.66	871447.71	Phuket	83	South-west
119	564202	Phuket Airport	8.8.42	98.18.52	424477.53	900391.90	Phuket	83	South-west
120	566201	Ko Lanta	7.32.0	99.3.0	505516.04	832708.02	Krabi	81	South-west
121	566202	Krabi	8.3.0	98.54.0	488981.42	889826.78	Krabi	81	South-west
122	567201	Trang Airport	7.31.0	99.37.0	568035.08	830913.11	Trang	92	South-west
123	570201	Satun	6.39.0	100.5.0	619749.21	735188.97	Satun	91	South-west

Appendix II:

Study I

Songklanakarin J. Sci. Technol.
37 (1), 81-88, Jan. - Feb. 2015

SJST SONGKLANAKARIN JOURNAL OF SCIENCE AND TECHNOLOGY
http://www.sjst.psu.ac.th

Original Article

Analysis of daily rainfall during 2001-2012 in Thailand

Mayuening Eso[1,2], Metta Kuning[1,2]*, and Somporn Chuai-Aree[1,2]

[1] *Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Mueang, Pattani, 94000 Thailand.*

[2] *Centre of Excellence in Mathematics, Commission on Higher Education, Ratchathewi, Bangkok, 10400 Thailand.*

Received: 21 June 2014; Accepted: 3 November 2014

Abstract

The presented study aims to classify precipitation regions, analyze trends, and fit an appropriate model for daily precipitation in Thailand. Factor analysis and generalized linear model (GLM) with gamma regression are performed on the historical records of daily rainfall amounts from 114 weather stations during 2001 to 2012. The study shows that the factor analysis divides the area of Thailand into seven regions with explanation of 58.9% of the total variance. The conducted gamma models reveal a good fit for the upper part, south-east, and south-west of the examined regions. The deviance residual plots from these models also provide a reasonable fit.

Keywords: daily rainfall, daily precipitation, factor analysis, gamma regression, GLM

1. Introduction

Rainfall in Thailand is under the influence of the monsoon seasonal winds i.e. southwest monsoon and northeast monsoon. The southwest monsoon starts in mid-May and continues to mid-October, and the northeast monsoon occurs from mid-October until mid-February. Moreover, the precipitation in Thailand is affected by other winds such as cyclone coming from Pacific Ocean, South China Sea, Bay of Bengal, and Andaman Sea. Typically, the cyclone comes through Thailand at an annual average of 3-4 times, between April and December for each year (Leewatchanaku, 2000; Wangwongwiroj, 2008).

The phenomenon of El Niño and La Niña also influence the amount of rainfall. These events occur every three to eight years as a part of a natural cycle. The term El Niño refers to the situation when sea surface temperatures in the east-central equatorial Pacific Ocean are significantly warmer than normal, and La Niña effects are opposite to those of El Niño and refer to a period when the eastern Pacific Ocean is much cooler than the normal sea surface temperature (Australian Government Bureau of Meteorology, 2005). The damage caused by El Niño is normally quite extensive. For example, it can involve drought and bush fires in Australia, Southeast Asia (especially in Indonesia), and southern Africa. High rainfalls are also observed over parts of Peru, Ecuador, and areas around the Gulf of Mexico from El Niño effect. The impact of La Niña on the world's weather is a tendency to heavy rainfall and flooding in Australia, Indonesia, Philippines, and southern Africa. It also causes a significant drought event occurring in east Africa and southern region of South America. From 2001 to 2012, El Niño occurred in May 2002 to March 2003 (11 months), June 2004 to February 2005 (9 months), August 2005 to January 2006 (6 months), and June 2009 to January 2010 (8 months). For the corresponding period, La Niña occurred in January to February 2001 (2 months) and in September 2007 to May 2008 (9 months) (Jutakorn, 2010), and also between late 2010 and early 2011 (Davey *et al.*, 2011). According to the Australian Government Bureau of Meteorology (2014), the La Niña during 2010-2011 occurred in October and December 2010, and in February and March 2011.

* Corresponding author.
Email address: metta.kuning@gmail.com

There are many advantages from water and rainfall, but excessive water supply might result in natural disasters such as landslides and floods. Thailand has been badly flooded many times due to heavy rainfall, which has contributed to costly damage of assets and agriculture, and significant loss of life. Therefore, it is important to find an appropriate statistical model for forecasting daily precipitation amounts.

In Thailand, precipitation trend analysis as part of climate change for the past several decades has been the topic of many studies. In 2005, Singhrattna *et al.* developed a statistical forecasting method for summer monsoon rainfall (August-October) over Thailand using traditional linear regression and a local polynomial-based nonparametric method in the west central region and in the Chao Phraya River basin. Hung *et al.* (2009) used an Artificial Neural Network (ANN) technique to improve rainfall forecast performance in Bangkok. Szyniszewska and Waylen (2012) studied daily rainfall characteristics from the monthly rainfall totals in central and northeastern Thailand (Lopburi, Chachoengsao, Buriram and Sisaket province). Phien *et al.* (1980) fitted the distribution in time of the monthly rainfall sequences at various stations in northeast Thailand. This study aims to figure out more precisely how daily rainfall varied over the period 2001-2012, to classify rainfall regions in Thailand and to develop an appropriate generalized linear model.

2. Data and Methodology

Daily precipitation for Thailand for the period 2001-2012 (12 years) was obtained from Thai Meteorology Department (TMD). The dataset contains 122 stations starting from January 1, 2001 to February 7, 2013. The criterion is set up to overcome the station with a large amount of missing values. The data in which the values are missing for more than 13% are removed. Eight stations which meet these criteria are Bang Khen, Burirum, Khanom, Krabi, Mae Jo, Nongbualumphu, Sukhothai, and Suwanabhum Airport. Thirty two stations are observed to have incomplete data for January 2001, and are also excluded from the analysis. The rain falling on February, 29 for each leap year, 2004, 2008, and 2012, are simply omitted. As a result, data from February 5, 2001 to February 4, 2013 for 114 stations are used for the study.

The spatial distribution of those 114 stations is shown in Figure 1. It is found that 11 provinces have no stations from this data set. They are Amnat Charoen, Yasothon, Saraburi, Ang Thong, Sing Buri, Nonthaburi, Uthai Thani, Samut Sakhon, Samut Songkhram, Nakhon Nayok, and Nong Bua Lam Phu. Provinces have different numbers of stations. For instance, there are 35 provinces with only one station, 19 provinces with two stations, five provinces with three stations, four provinces with four stations, and two provinces with five stations.

However, rainfalls on successive days are serially correlated. Since conventional statistical models assume independent errors, we effectively reduce these serial correlations by restructuring the data as 5-day averages, thus reducing the data to 73 periods in each year. The autocorrelation function (ACF) is applied for checking the independence assumption (Shumway and Stoffer, 2011; Venables and Ripley, 2002).

Rainfalls in closely clustered stations are also spatially correlated. In this study, factor analysis is used to allocate the 114 stations to a smaller number of clusters with similar patterns. The model for factor analysis can be written in matrix notation as

$$y-\mu=\Lambda f+\varepsilon \tag{1}$$

where

$$y=(y_1, y_2, ..., y_p)', \ \mu=(\mu_1, \mu_2, ..., \mu_p)',$$

$$f=(f_1, f_2, ..., f_m), \ \varepsilon=(\varepsilon_1, \varepsilon_2, ..., \varepsilon_p)', \text{ and}$$

$$\Lambda=\begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pm} \end{pmatrix}.$$

Likewise, $y_1, y_2, ..., y_p$ are 5-day average of rainfall, $\mu_1, \mu_2, ..., \mu_p$ are the mean rainfall amount of each station, $f_1, f_2, ..., f_m$ are common factors, and $\varepsilon_1, \varepsilon_2, ..., \varepsilon_p$ are unique factors (Johnson and Wichern, 2007; Rencher, 2002). In this study,

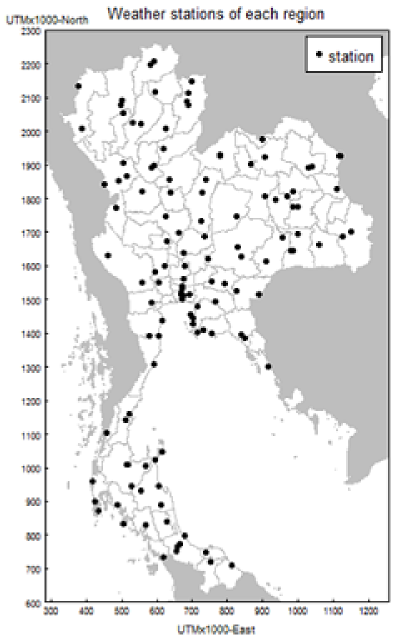

Figure 1. Geographical distribution of the 114 meteorological stations in Thailand

the covariance matrix is used to fit the factor model. Maximum likelihood and promax are used for factor extraction and factor rotation, respectively, to estimate factor loadings. Then, the average daily rainfall from the stations in each factor is computed.

Rainfall measurements have skewed distributions, even when only rainy days (R), with over 4.43 mm/day are selected, and therefore violating the statistical assumption of normally distributed errors. The generalized linear model (GLM) with gamma distributions is fitted for this skewed data. The function is most commonly written as

$$f\left(y|\alpha,\beta\right)=\frac{1}{\Gamma(\alpha)}\beta^{\alpha}y^{\alpha-1}e^{-\beta y}\ ,y,\alpha,\beta>0 \qquad (2)$$

where y is a random sample from the gamma distribution, α is an inverse-scale parameter, β is a shape parameter, and $\Gamma(\alpha)$ is a gamma function.

For GLM, the canonical link for the gamma family is $-1/\mu$. Mean for gamma distribution is μ and variance is μ^2 (Gill, 2001). The analyzed multiple regression model is shown as follows;

$$1/\text{mean}[R] = \text{constant - factor(year) - factor(month)} \qquad (3)$$

To check the goodness of fit, deviance residual is used to evaluate models. A 95% confidence interval plot can be used to show the pattern of rainfall amounts for each factor. These confidence intervals are based on standard errors to present the differences between the average rainfall and its overall mean by sum contrasts (Tongkumchum and McNeil, 2009; Venables and Ripley, 2002). The dependent variable of this research is the rainfall amount as 5-day average, and the determinants are months and years. R programming language is used for data analysis and graphical displays (R Development Core Team, 2013).

3. Results

3.1 Factor analysis

Factor analysis is used to produce the rainfall regions in Thailand. This technique is useful for a very large dataset and data reduction with minimum loss of information in order to have a good understanding and interpretation of the data results (Baeriswyl and Rebetez, 1997).

The loading values from factor analysis in Table 1 denote correlations between factors and stations. This technique reduces the number of meteorological stations from 114 to 7 regions and can be explain 58.9% of the total variance. When taking into account the loadings above 0.30, 87 of the 114 stations are correlated with a single factor, and 26 stations are correlated with two factors.

Figure 2 demonstrates the seven regions identified by factor analysis, which comprises of the following regions. The first region (F1) covers the area of the northeastern part of Thailand for 32 stations. The second region (F2) includes the central and eastern part of Thailand with 34 stations, and therefore is defined as central. The third region (F3) consists of 13 stations from the area of the south-east. The fourth region (F4) represents the upper northern part of Thailand with 14 stations. The fifth region (F5) combines 10 stations from the lower north. The sixth region (F6) comprises of nine stations from the south-west, and the last region (F7) covers the area of the upper south including two stations.

3.2 The pattern of daily rainfall

The annual daily average rainfall of each region during 2001-2012 is shown in Figure 3. The daily average rainfall of the upper north, lower north, northeastern, and central region range from 2.99 to 5.01 mm/day. The highest daily average rainfall is in the year 2011 for these regions. These are around 5.01, 4.96, 4.99 and 4.29 mm/day for each region, respectively. In 2009, the upper north has approximately 3.13 mm/day of rainfall on the average, the lowest amount compared to other regions. For the south-west, the daily average rainfall is more than 5.5 mm/day for all years; it reaches its peak of 7.99 mm/day in 2012. For the upper south, the rainfall fluctuates from 4.61 to 6.61 mm/day. There are obvious peaks in 2002 and 2011 with 6.30 and 6.61 mm/day, respectively. The rainfalls of the south-east varies from 4.37 to 8.64 mm/day. Unusual heavy rainfall is also observed in 2005, 2008, 2010, and 2011 with the average of 6.57, 6.86, 7.22, and 8.64 mm/day, respectively.

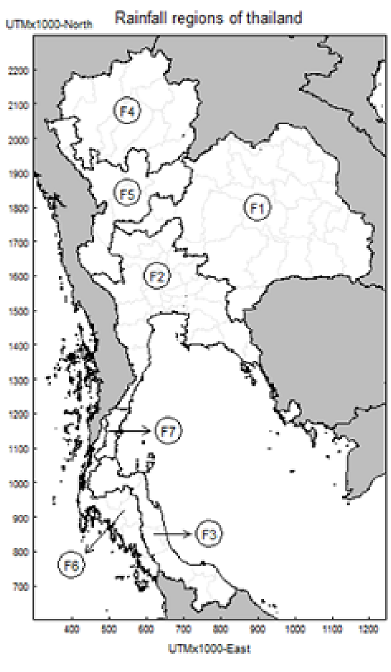

Figure 2. The rainfall regions from factor analysis over Thailand

Table 1. Loading scores from factor analysis

Station	F1	F2	F3	F4	F5	F6	F7	Uniq
NPnomA	**1.133**	-0.125			-0.237		0.191	0.189
MDahan	**1.105**				-0.155		0.133	0.202
SNakhon	**1.100**				-0.240			0.189
NPnom	**1.087**	-0.123			-0.206		0.209	0.175
SNakhonA	**1.082**				-0.238			0.238
RoiEt	**1.008**							0.194
KMlasai	**1.007**					-0.120		0.184
RoiEtA	**1.005**							0.232
KSPhiSai	**0.949**							0.209
TPhraA	**0.905**							0.234
KKaen	**0.859**	0.113			-0.120			0.234
UBRthani	**0.843**						0.164	0.227
Uthani	**0.835**			0.220				0.226
ThaTum	**0.799**			-0.110				0.249
SurinA	**0.779**			-0.127				0.278
Nkhai	**0.778**			0.255				0.244
Surin	**0.767**			-0.122	0.114			0.278
UBRthaniA	**0.765**				0.127		0.181	0.270
SSKetA	**0.754**				0.157		0.103	0.259
ChPhum	**0.566**	0.183					-0.188	0.253
Loei	**0.565**	0.107		0.228			-0.192	0.242
LoeiA	**0.546**	0.118		0.238			-0.191	0.234
LomSak	**0.487**			0.170	0.284			0.239
WBuri	**0.438**	0.214			0.174		-0.158	0.272
NRong	**0.618**	**0.342**		-0.172				0.298
NRsima	**0.569**	**0.317**					-0.176	0.281
ChChai	**0.492**	**0.424**		-0.105			-0.121	0.319
SKaew	**0.476**	**0.460**					0.100	0.300
BChum	**0.423**	**0.320**			0.111		-0.129	0.290
Pchabun	**0.441**			0.158	**0.314**			0.267
PichitA	**0.367**	0.110		0.129	**0.356**			0.280
Nsawan	**0.299**	0.277			0.281			0.268
KHongA			**0.856**					0.317
Skhla			**0.830**					0.372
NSTA	-0.112		**0.809**					0.337
YalaA	0.186		**0.791**					0.446
NST			**0.788**					0.367
PTlungA	-0.155		**0.773**					0.393
HYaiAir			**0.765**					0.392
PTNAir			**0.765**					0.411
NRTwat			**0.733**			-0.146		0.535
SDao			**0.618**			0.248		0.486
SRThaniA			**0.615**			0.107	0.200	0.425
SRThani			**0.561**			0.187	0.203	0.404
KSmui	-0.105	0.190	**0.551**				0.228	0.520
ChRaiA	0.187	0.118		**0.698**				0.207
ChRai	0.215	0.118		**0.671**				0.225
Pyao	0.113	0.126		**0.666**				0.228
ChMai	0.117			**0.522**	0.257			0.229
Lphun				**0.508**	0.296			0.243
Lpang		0.121		**0.502**	0.295			0.224
MaeHS	0.278			**0.442**	0.193		0.169	0.279
TChang	**0.318**			**0.698**	-0.172			0.192
ThaWP	**0.323**			**0.697**	-0.124			0.160
NanA	**0.362**			**0.660**				0.158
Nan	**0.394**			**0.621**	-0.110			0.176
LpangA				**0.518**	**0.367**			0.201
Phrae	0.161			**0.496**	**0.314**			0.225
Udit	0.160	0.122		**0.411**	**0.327**			0.202

Station	F1	F2	F3	F4	F5	F6	F7	Uniq
HPongA	-0.185	**1.052**			-0.193			0.302
SThip	-0.253	**1.050**			-0.127			0.289
Lbang		**1.042**			-0.183			0.260
PTya	-0.140	**1.035**			-0.134			0.295
RYong	-0.124	**0.969**			-0.158			0.333
KSchang		**0.941**			-0.102			0.283
Pilot		**0.905**						0.312
KToey	0.123	**0.885**				-0.148		0.229
ChBuri		**0.874**						0.276
BNa	0.110	**0.870**				-0.119		0.217
BkokM	0.108	**0.862**				-0.129		0.235
Pburi		**0.775**					0.230	0.311
KPSaenA		**0.739**			0.147			0.263
Ratburi		**0.702**			0.143		0.155	0.279
DMuang	0.138	**0.681**						0.271
HuaHin		**0.663**					0.253	0.369
Kanburi		**0.656**			0.157			0.337
ChChsaoA	0.197	**0.654**				0.113		0.356
PTthaniA	0.136	**0.637**			0.125			0.287
NPhlupA		**0.629**					0.158	0.384
UThongA		**0.616**			0.230			0.262
SuPBuri	0.129	**0.610**			0.213			0.276
PhriuA	0.281	**0.571**	-0.161			0.159	0.146	0.245
AtayaA	0.204	**0.552**			0.183			0.287
LBuri	0.233	**0.438**			0.240			0.278
PChongA	0.265	**0.393**		-0.103	0.167	0.154	-0.172	0.349
TakFaA	0.240	**0.345**			0.244		-0.103	0.268
ChTburi	**0.364**	**0.561**	-0.111				0.144	0.238
Aprathet	**0.356**	**0.469**						0.375
KBBurin	**0.346**	**0.450**						0.301
PrBuri	**0.369**	**0.433**						0.262
KLYai	**0.392**	**0.376**	-0.247			0.203	0.262	0.283
PRKhan		**0.595**			0.105		**0.337**	0.429
ChNat	0.210	**0.413**			**0.316**			0.268
Tak	-0.100	0.109		0.183	**0.738**			0.203
BmbolDam	-0.149	0.147		0.229	**0.685**			0.262
DoiMSA	0.135			0.222	**0.649**		0.190	0.149
MaeSot	0.285			0.243	**0.553**	-0.136	0.287	0.163
SSrongA	0.168	0.108		0.252	**0.451**			0.256
Umphang	0.211		-0.101	0.149	**0.449**		0.135	0.276
TPPhum	0.218	0.176	-0.140	0.120	**0.420**		0.224	0.239
KPPhet	0.171	0.188		0.146	**0.413**			0.288
MaeSr	0.207	0.107		0.292	**0.381**		0.200	0.237
Pnulok	**0.314**	0.151		0.154	**0.374**			0.268
PKetAir	-0.175		0.108			**0.839**		0.276
TKPa						**0.824**		0.265
Pket			0.169			**0.778**		0.322
KLanta			0.181			**0.722**		0.317
Satun			0.270		-0.130	**0.637**		0.464
TrangAir		-0.147	**0.334**			**0.648**		0.353
PSang		-0.128	**0.336**			**0.576**	0.106	0.418
Chawang		-0.154	**0.334**			**0.556**	0.168	0.377
RNong	0.276					**0.490**	**0.371**	0.206
SawiA		0.249	0.251			0.129	**0.396**	0.519
Chphon		0.252	0.274				**0.393**	0.501

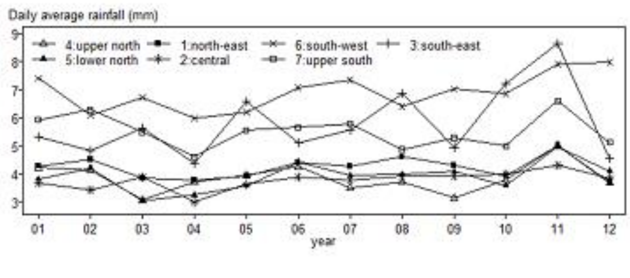

Figure 3. the average of daily rainfall

3.3 Rainy season

The rainy season is defined here as the successive days with an average of daily rainfall of at least 4.43 mm/day. The rainy season is showed by the dark grey color in Figure 4. In the upper part of Thailand, the rainy season for upper north is in the period between April and September, which lasts six months. The lower north region also has the rainy season around six months starting from May to early October. In the north-east, the rainy season occurs in late April to early October, and lasts roughly for seven months. In the central region, the rainy month starts one month after that of north-east from May to October and it lasts six months.

In contrast to rainy seasons in the south of Thailand, the south-west has a quite longer rainy season around nine months starting from late March to November. The upper north has four rainy months during September to December, and there are five months of a rainy season for the south-east starting from September to January.

The overall mean of daily average rainfall during the rainy season are 10.71, 9.16, 8.19, 7.90, 7.40, 7.24, and 6.25 mm/day for the region of south-east, south-west, upper south, north-east, upper north, lower north, and central, respectively.

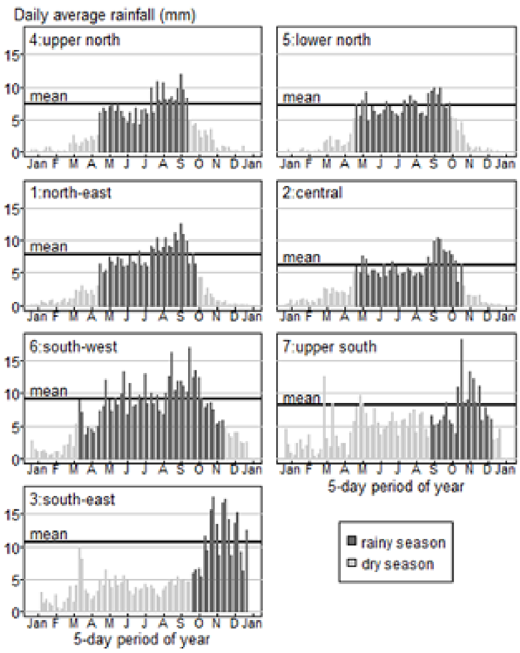

Figure 4. The rainy season of seven regions

3.4 Generalized linear model: gamma regression

Gamma regression is fitted to the data using *glm*() function in R. The four regions in the upper part of Thailand have correlation coefficients of over 0.60 (Figure 5), and thus these regions are grouped into one region. In the southern part, the upper south and south-east are also combined because there is a lower number of stations in the upper south region. Gamma regression is thus fitted into three regions; upper, south-west and south-east in the period of rainy seasons.

The autocorrelation function of deviance residuals from gamma model is analyzed to ensure the independence assumption has been handled properly. The ACF plots for each factor show that no statistically significant correlations are at lag 1, and all autocorrelation are below 0.2 as shown in Figure 6.

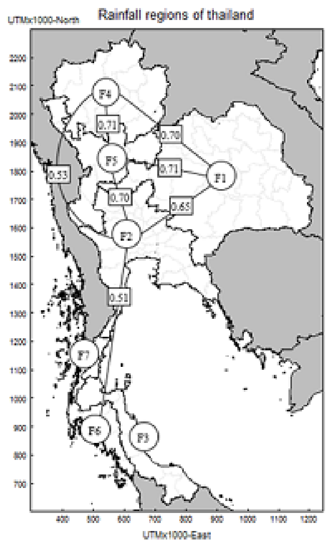

Figure 5. The correlation coefficients between regions

Figure 7 shows the confidence intervals from the fitted gamma model. For the region of upper part of Thailand, it is found that there are discrepancies for the month of June, August and September from the overall mean of 7.05 mm/day. For these months, the average rainfalls are 6.17, 7.98, and 9.34 mm/day, and 95% confidence intervals are 5.60-6.87, 7.22-8.92, and 8.36-10.59 mm/day, respectively. The daily rainfall in June is under the overall mean and different from other months. In the south-east, the month of November is evidently different from the overall mean of 10.25 mm/day. The average rainfall in this region is 14.89 mm/day and 95% confidence interval is 11.56-20.94 mm/day. In the south-west, there are six months that are statistically significantly different from its overall mean of 9.16 mm/day. It is shown that the averages of rainfall in both April and November are less than the overall mean by having 5.04 mm/day with 4.38-5.93 mm/day of 95% confidence interval, and 6.32 mm/day with 5.34-7.75 mm/day of 95% confidence interval, respectively. For the remaining four months, July, August, September, and

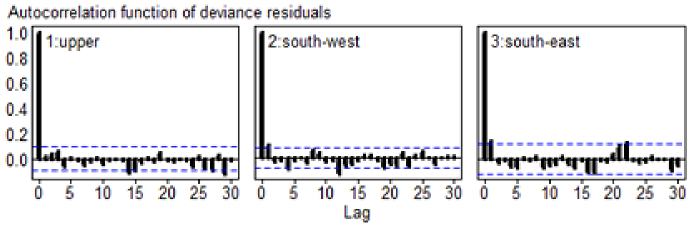

Figure 6. The autocorrelation function of residuals

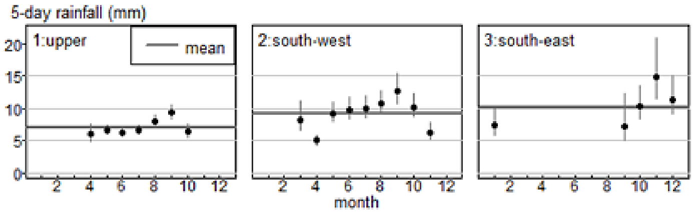

Figure 7. Confidence interval from gamma models of each month

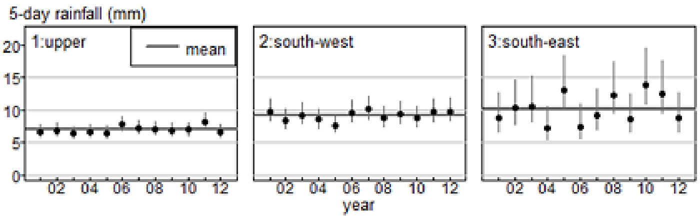

Figure 8. Confidence interval from gamma models of each year

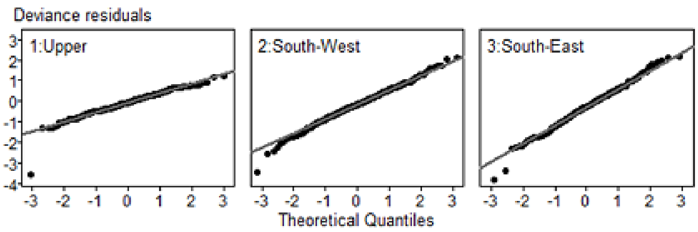

Figure 9. Deviance residuals of three regions from gamma model

October, the averages of rainfall are 10.00, 10.71, 12.73, and 10.27 mm/day with 95% confidence interval of 8.60-11.96, 9.26-12.70, 10.85-15.40, and 8.82-12.29 mm/day, respectively.

When considering year variables (Figure 8), the average of rainfall in the year 2011 is different from the overall mean for the region of the upper part of Thailand. The average of rainfall is 8.30 mm/day, and 95% confidence interval is 7.31-9.60 mm/day. In the south-east, the average rainfalls are different from their overall mean in 2005 and 2010, the average rainfall of 13.14 and 13.99 mm/day and 95% confidence interval of 10.21-18.44 and 10.90-19.51 mm/day, respectively. However, the rainfall in the south-west is not different from its overall mean in each year.

The deviance residual plots from the model are shown in Figure 9, and indicate that the model provides a reasonable fit for all regions.

4. Conclusions and Discussion

The approach of factor analysis is able to separate the regions of Thailand with overall 114 weather stations into seven clusters, comprising of upper north, lower north, north-east, central, upper north, south-west and south-east with the total variance of 58.9%. The successive days with an average rainfall over 4.43 mm/day are specified as a rainy season for each region. Finally, after investigating the correlation coefficient between regions, the seven regions are further reduced into three regions, namely the upper part of Thailand, south-west, and south-east.

The result shows that three models are capable of fitting data reasonably well. The average of rainfall in June, August, and September of the upper part region are statistically significantly different from its overall mean. For the south-west, there are six months that the amounts of rainfall are different from its overall mean starting from April to December excluding May and June, and only the one month of November for the south-east. The amounts of rainfall in the year 2001 of the upper part of Thailand are different from the overall mean and for the south-east, the levels of rainfall in the year 2005 and 2010 are also different from its overall mean.

It is evident that the rainy season of the region in the upper part of Thailand is in the period of the southwest monsoon and also for the south-east in the period of the northeast monsoon. When considering the phenomenon of the La Niña, it seems that the La Niña has no effect on the levels of rainfall in Thailand. The reason for this can be seen from the final model in which the upper part is statistically significantly different from the overall mean in September 2011 when there was no La Niña. Likewise, for the south-east, there was no La Niña event in November 2005 and November 2010.

In addition, the result of year variable from the gamma model for the central region is not different from other years, but in 2011 the area around Bangkok suffered severe flooding starting around the end of July. During June to October in 2011, Thailand experienced five categories of storms. These include 'Hai Ma', 'Nok Ten', 'Hai Tang', 'Nesat', and 'Nalgae'. They caused heavy rainfall in the upper part of Thailand, which then resulted in the worst flood affecting 684 districts of 65 provinces around and in Bangkok (IPSR, 2012).

Acknowledgments

We gratefully acknowledge the Thai Meteorological Department (TMD) for providing the daily rainfall data. We are also grateful to Emeritus Professor Don McNeil for his helpful suggestions. This research was supported by grant funds from the program of Strategic Scholarships Fellowships Frontier Research Networks (Specific for Southern Region) for the Ph.D. Program Thai Doctoral degree from the Office of the Higher Education Commission, Thailand, and also by the Centre of Excellence in Mathematics, the Commission on Higher Education, Thailand.

References

Australian Government Bureau of Meteorology. 2005. El Nino, La Nina and Australia's Climate. Available from: http://www.bom.gov.au/info/leaflets/nino-nina.pdf. [March 27, 2014].

Australian Government Bureau of Meteorology. 2014. Record-breaking La Niña events: An analysis of the La Niña life cycle and the impacts and significance of the 2010-11 and 2011-12 La Niña events in Australia. Available from: http://www.bom.gov.au/climate/enso/history/ln-2010-12/index.shtml. [April 3, 2014]

Baeriswyl, P.A. and Rebetez, M. 1997. Regionalization of precipitation in Switzerland by means of principal component analysis. Theoretical and Applied Climatology. 58, 31-41.

Davey, M., Huddleston, M. and Brookshaw, A. 2011. Global impact of El Niña and La Niña. Available from: http://www.changingclimate.org.uk/uploads/content_documents/361/Global_Impact_of_ElNino_and_LaNina_Implications_for_Financial_Markets.pdf. [April 1, 2014].

Gill, J. 2001. Generalized Linear Models: A Unified Approach, Sage Publications, Inc., Thousand Oaks, U.S.A., pp. 30-39.

Hung, N.Q., Babel, M.S., Weesakul, S. and Tripathi, N.K. 2009. An artificial neural network model for rainfall forecasting in Bangkok Thailand. Hydrology and Earth System Sciences. 13, 1413-1425.

Institute for Population and Social Research, Mahidol University (IPSR). 2012. Food Security the Illusion of Money VS the Reality of Food. Thai Health. 9, 34-40.

Johnson, R.A. and Wichern, D.W. 2007. Applied Multivariate Statistical Analysis, Pearson Prentice Hall, New Jersey, U.S.A., pp. 481-526.

Jutakorn, J. 2010. El Niño and La Niña phenomenon effected on southwest monsoon in Thailand. Meteorological department. Weather report no. 551.557-21-01-2010.

Leewatchanaku, K. 2000. Hydrolog, Siam Stationery Supplies Ltd. Part, Bangkok, Thailand, pp. 3.10-3.23.

Phien, H.N., Arbhabhirama, A. and Sunchindah, A. 1980. Distribution of monthly rainfall in northeast Thailand. Southeast Asian Studies. 18, 110-123.

R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Austria: R Foundation for Statistical Computing. Available from: http://www.R-project.org. [April 18, 2014].

Rencher, A.C. 2002. Methods of Multivariate Analysis, John Wiley and Sons, Inc., New York, U.S.A., pp. 408-447.

Shumway, R.H. and Stoffer, D.S. 2011. Time Series Analysis and Its Applications with R Examples, Springer Science and Business Media, LLC, New York, U.S.A., pp. 21.

Singhrattna, N., Rajagopalan, B., Clark, M. and Kumar, K.K. 2005. Seasonal forecasting of Thailand summer monsoon rainfall. International Journal of Climatology. 25, 649-664.

Szyniszewska, A.M. and Waylen, P.R. 2012. Determining the daily rainfall characteristics from the monthly rainfall totals in central and northeastern Thailand. Applied Geography. 35, 377-393.

Tongkumchum, P. and McNeil, D. 2009. Confidence intervals using contrasts for regression model. Songklanakarin Journal of Science and Technology. 31, 151-156.

Venables, W.N. and Ripley, B.D. 2002. Modern Applied Statistics with S, Springer Science and Business Media, LLC, New York, U.S.A., pp. 183-189, 389-391.

Wangwongwiroj, N. 2008. Hydrology, King Mongkut's University of Technology Thonburi, Bangkok, Thailand, pp. 57-71.

Appendix III:

Study II

WALAILAK JOURNAL

http://wjst.wu.ac.th Physical Sciences

The Southern Oscillation Index as a Random Walk

Mayuening ESO[1, 2], Metta KUNING[1, 2, *], Hilary GREEN[3], Attachai UERANANTASUN[1] and Somporn CHUAI-AREE[1, 2]

[1]*Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Mueang, Pattani 94000, Thailand.*
[2]*Centre of Excellence in Mathematics, CHE, Si Ayutthaya Rd., Bangkok 10400, Thailand.*
[3]*Department of Statistics, Faculty of Science and Engineering, Macquarie University, NSW, 2109, Australia.*

(*Corresponding author's e-mail: metta.kuning@gmail.com)

Received: xxx, Revised: xxx, Accepted: xxx

Running title

The SOI as a Random Walk Text

Abstract

The Southern Oscillation Index (SOI) has been used as a predictor of variables associated with climatic data, such as rainfall and temperature, and is related to the El Nino and La Nina phenomena, also called El Nino Sothern Oscillation (ENSO). The present study aims to describe the characteristics of the SOI between 1876 and 2014 using statistical methods. The graph of the cumulative monthly SOI in the period 1876-2014 shows that the data can be divided into four periods. The first period, from 1876 to 1919, shows no trend. An increasing trend is apparent in the second period from 1920 until 1975 while a decreasing trend is apparent in the third period, 1976 to 1995. In the last period, between 1996 and 2014, the SOI appears fairly stable. In order to investigate those trends, the linear regression and autoregressive (AR) model have been fitted. For the linear regression model, the outcome, SOI, is regressed against boxcar function, where the functions model the trends of the SOI. An autoregressive process is used to account for serial correlation in the residuals. The conclusion is that the SOI is quite similar to a random noise process.

Keywords: Autoregressive model, boxcar function, serial correlation, Southern Oscillation Index, white noise

Introduction

The southern oscillation (SO) is normally identified by the southern oscillation index (SOI) [1], that is, the large-scale fluctuations of the Mean Sea Level Pressure (MSLP) in the tropical Pacific between Tahiti and Darwin, Australia. The SOI is expressed as a number which ranges from about -30 to +30. Prolonged periods of negative SOI values are called *El Niño* events where air pressure is below-normal at Tahiti and above-normal at Darwin. The opposite condition occurs when there are prolonged periods with strong positive SOI values called *La Niña* events. Those events are also called El Nino Southern Oscillation (ENSO) events [2-4].

The SOI time series is useful for research into climatic data. There are several studies on the relationship between SOI and rainfall (e.g. Hadiani et al [5]; Tigona and de Freitas [6]) as well as SOI and temperature (e.g. Jones and Trewin [7]; Halpert and Ropelewski [8]). Some studies have indicated that positive values of SOI correlate with above average rainfall and negative values of SOI correlate with below average rainfall, for example, in Australia (Stone and Auliciems [9]) and Ghana (Adiku and Stone [10]), whereas the correlation between SOI and rainfall was shown to be weaker in Western Australia (Carberry et al [11]). In addition, Carbone et al [2] reported that the total ozone in southern Brazil correlated with SOI; such that a reduction in total ozone occurred during *El Niño* episodes (negative SOI) and an increase in ozone occurred during *La Niña* episodes (positive SOI) from 1997 to 2003. On the other hand, Al-Zuhairi et al [12] studied the correlation between temperature and SOI over Iraq where there was no relationship detected in the period from 1900 to 2008.

That "the weather is not static", Carberry et al [11], implies that the SOI is also continuously fluctuating. The SOI index is calculated using a moving 30-day average in order to try to eliminate some *Noise*, small random changes. A section of these data is illustrated in Figure 1 which shows the monthly average SOI from 1876 to 2014. The graph does not show the trend of SOI because the data fluctuate.

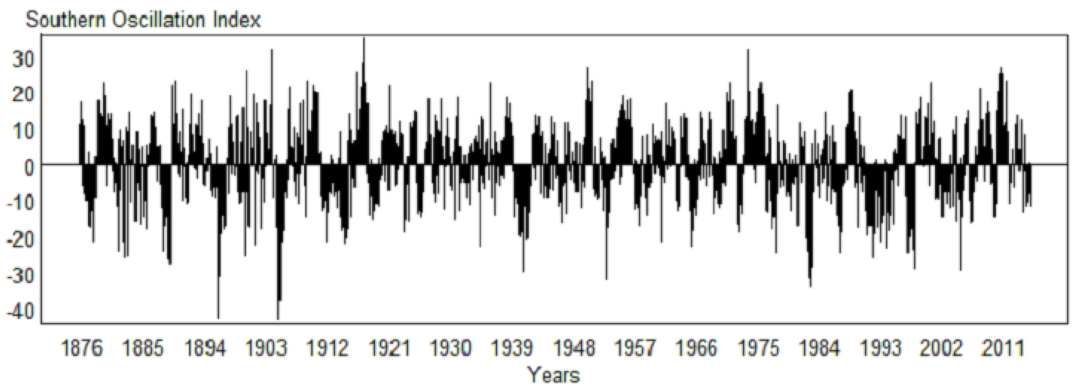

Figure 1 The variation of SOI from 1876 to 2014.

However, when we look at the graph of the cumulative SOI values in Figure 2, we can see an apparent trend from which Watts [13] deduced that "...from 1920 went into a long La Nina-dominate trend that ended with the Great Pacific Climate Shift of 1976" and "The subsequent El Nino-dominated trend from 1976 to 1995 was almost three times as fast as the rise". Thus, the focus of this paper is to evaluate the evidence for these claim using appropriate scientific methods.

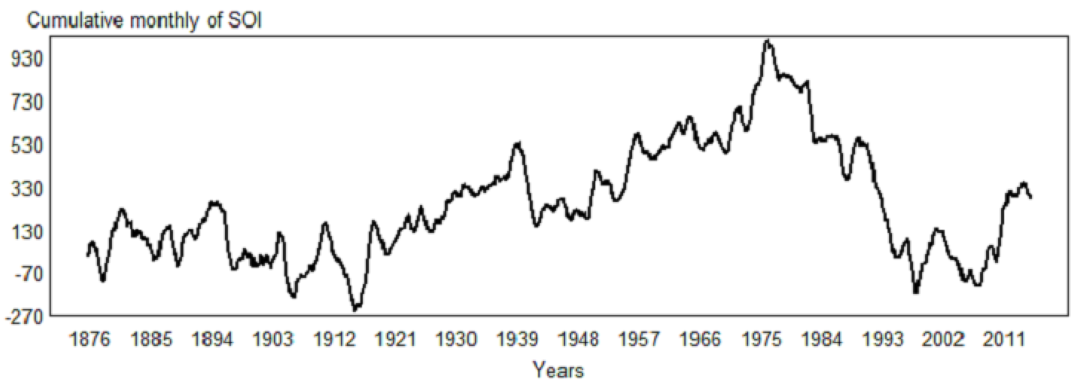

Figure 2 Cumulative monthly of SOI.

The data and methodology used in this study are described in the following section. Results and discussion are then presented, summarized in sections 4.

Materials and methods

Data

The data of the present study were downloaded from the Australian Government Bureau of Meteorology website, http://www.bom.gov.au/climate/current/soihtm1.shtml. The data are the average monthly SOI calculated over 139 years, so there are 1,664 observations from the year 1876 to 2014. There are several methods used to calculate the SOI; the method used by the Australian Bureau of Meteorology is the Troup SOI [14] (see, Australian Government Bureau of Meteorology [15]).

Methodology

The null hypothesis of this analysis is that the data are just noise, and in turn, the fluctuations in the SOI are random. In order to test this hypothesis, firstly, the scientific approach involves fitting an appropriate model to the data, and then applying an appropriate statistical test of the null hyphothesis, resulting in a p-value. The p-value is the probability that a data confuguration is at least as unusual as that observed could have arisen purely by chance, that is assuming that the null hypothesis is true. By convention, p-values smaller than 0.05 provide sufficient evidence to reject the null hypothesis.

The first model fitted to the data is a linear regression model (see, Gill [16]; Venables and Ripley [17]) taking the formulation below (Eqn 1)

$$y_t = b_0 + b_i x_{it} + z_t \, . \qquad \text{(Eqn 1)}$$

In this model y_t represent the SOI for month t where t equals 1 for January 1876, while x_{it} represent boxcar function (see, Weisstein [18]) taking the values of 1 on intervals starting with a changing point and 0 elsewhere. The z_t constitute successive values which may be a series of auto-correlated normally-distributed errors (noise).

For a time series [19], the residuals z_t are often called the noise, and the model which does not include the noise is called the signal. If the residuals arise from uncorrelated errors, the noise is called white noise, otherwise it is coloured. A commonly implemented model for coloured noise follows an autoregressive (AR) process. The AR model takes the form

$$z_t = \sum_i a_i z_{t-s_i} + w_t \qquad \text{(Eqn 2)}$$

where w_t is white noise. This means that each value of z_t is expressed as a linear combination of previous values at specified lags (s_1, s_2, s_3, etc.) plus an independent white noise series.

The linear regression model that assumes independent errors using the function *lm*() in R [20] is used to determine the number of parameters in the AR model. The validity of a model can be checked by using the *acf*() function in R to plot the auto-correlation function (ACF) of the z_t (Eqn 1). The *pacf*() function, the partial auto-correlation function (PACF) of the residuals, indicates which lagged terms need to be included in the AR model (Eqn 2) to ensure that the w_t component represents white noise. The function *arima*() is used to analyse an AR model with the same predictor as in the linear regression model (Eqn 1). From the AR model, the p-values of each lag are considered; these should be significant (<0.05), for all lagged terms. Finally, the Analysis of Variance (ANOVA) procedure is used to test the predictor from the final model to confirm that at least one of the parameters from Eqn 1 is not equal to zero.

Results and discussion

The cumulative data from 1876 to 2014 are shown in Figure 3 and the patterns suggest four periods. The first period from 1876 to 1919 seems to show no trend. An increasing trend is apparent between the years 1920 and 1975 followed by a decreasing trend between the years 1976 and 1995. Again, it appears that there is no trend in the period from 1996 to 2014.

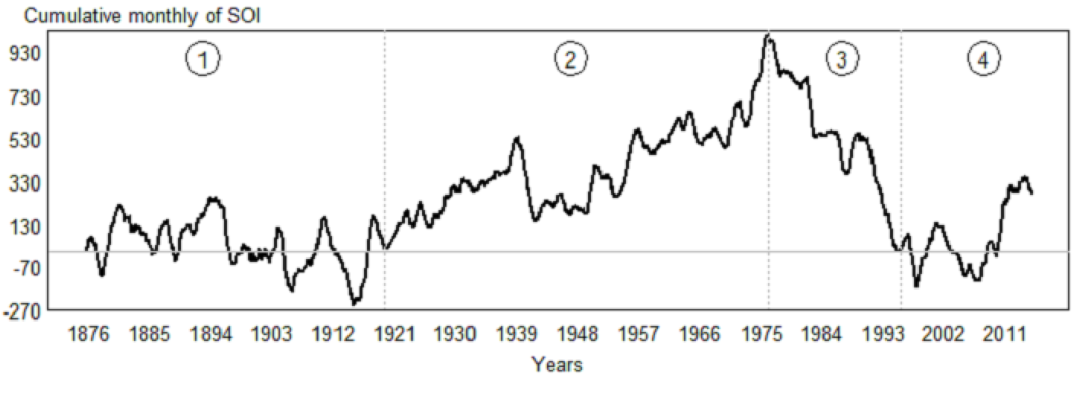

Figure 3 The four periods of the data.

The same statistical methods are applied to three subsets of the raw data. The first series includes all the data, from 1876 to 2014. The second series includes the data from 1920 to 2014, while the third series only includes the data from 1920 to 1995. The results from the analyses of the data from the three different periods are presented in the next section.

Analysis of four periods of data

Firstly the four periods are analysed using the linear model on three boxcar functions. The first boxcar function takes value of 1 starting at the changing point during the year 1920 to 1975 and 0 elsewhere (x_1), the second boxcar function takes value of 1 during the year 1976 to 1995 and 0 for others (x_2) and the last boxcar function takes value of 1 during the year 1996 to 2014 and 0 elsewhere (x_3).

Table 1 shows the coefficient of parameters b_0, b_1, b_2 and b_3 from the linear model when b_0, b_1, b_2 and b_3 refer to the slopes of the first, second, third and last period, respectively. The model indicates that the estimates of the coefficients of b_0 and b_3 are not statistically significant, but those of for b_1 and b_2 are significant with p-values less than 0.05. In other words, there is no significant trend in either the first or fourth period, but there is a significant increasing trend in the second period, (b_1 = 1.373, p-value =0.023) and decreasing trend in the third period, (b_2 = -4.045, p-value < 0.001).

Table 1 The coefficient of parameters from linear model for four periods.

Parameters	Coefficients	SE	t-value	P-value
b_0	0.037	0.450	0.081	0.935
b_1	1.373	0.601	2.284	0.023
b_2	-4.045	0.804	-5.028	<0.001
b_3	1.130	0.814	1.389	0.165
The p-value of ANOVA test is <0.001				

The ACF graph of the residuals from the linear regression model above shows that the data are highly serially correlated, and the PACF graph suggests the correlation of terms at lags 1, 2 and 3 months (see Figure 4).

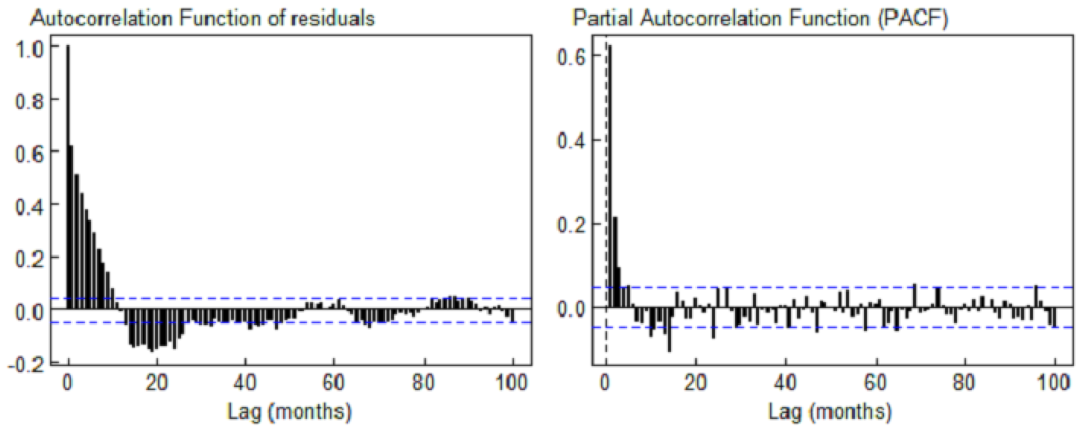

Figure 4 ACF graph (left panel) shows serial correlation and PACF graph (right panel) shows lagged terms of serial correlation from the linear model for four periods.

Then those lagged terms are included in the function *arima*() giving the results as shown in Table 2. As all AR terms at lags 1, 2 and 3, are statistically significant, we have derived an appropriate model.

Table 2 The result from AR(3) for four periods.

Parameters	Coefficients	SE	P-value
ar1	0.467	0.024	<0.001
ar2	0.168	0.027	<0.001
ar3	0.095	0.024	<0.001
b_0	-0.083	1.249	0.947
b_1	1.645	1.652	0.319
b_2	-4.142	2.185	0.058
b_3	1.367	2.234	0.541
	The p-value of ANOVA test is 0.078		

The ACF graph from this model in the right panel of Figure 5, indicates that successive values w_t are not correlated and the graph of the PACF in the right panel of Figure 5 shows that the model fits the data quite well. However, The p-value from the ANOVA test for parameters b_0, b_1, b_2 and b_3 is 0.078 indicating that there is insufficient evidence against the null hypothesis and that the coefficients of each parameter may indeed equal zero. It can be interpreted that there is no trend for all periods, in which it is in accordance with the p-value of more than 0.05 for those four parameters as shown in Table 2.

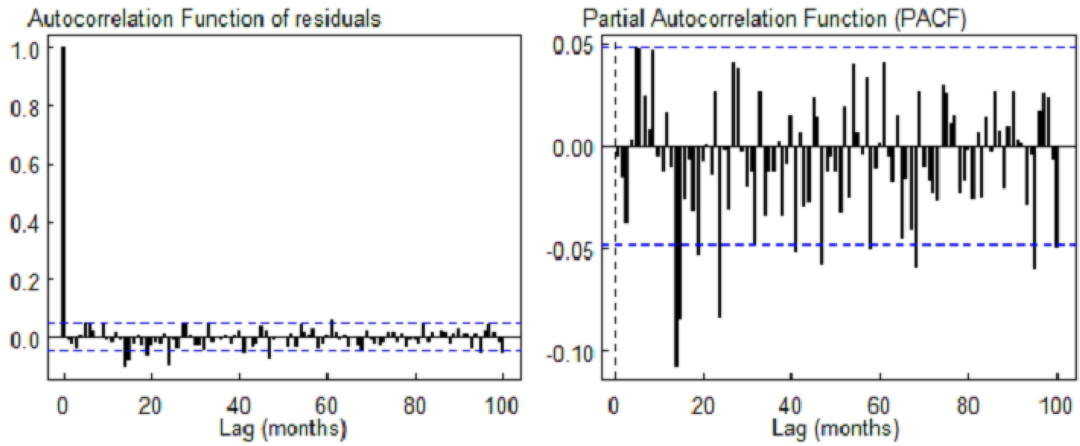

Figure 5 ACF graph in the left panel and PACF graph in the right panel from AR(3) for four periods.

Analysis of three periods of data

Next, we apply the same method as above, but we only include the data from three periods from 1920 to 2014, which includes the data from the second, third and fourth periods as shown in Figure 3. The first period is presumed to have no trend. In this analysis, two boxcar functions are considered, that is during the year 1976 to 1995 (x_1) and 1996 to 2015 (x_2). Table 3 shows the estimated coefficients of the paramters b_0, b_1 and b_2 from the linear model. These results are consistent with those we found from the previous analysis, and shows significant increasing and decreasing trends in the second and third periods respectively, and no trend in the fourth period. After the ACF and PACF graphs presented in Figure 6, are examined, it is again apparent that three lagged terms should be in the AR model.

Table 3 The coefficient of parameters from linear model for three periods.

Parameters	Coefficients	SE	t-value	P-value
b_0	1.409	0.373	3.779	<0.001
b_1	-5.417	0.727	-7.452	<0.001
b_2	-0.242	0.736	-0.329	0.742
			The p-value of ANOVA test is <0.001	

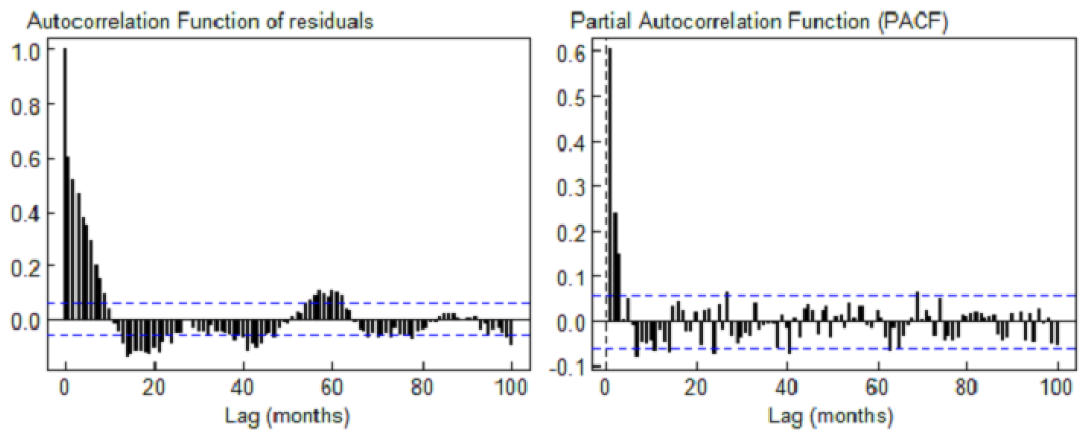

Figure 6 ACF graph (left panel) shows serial correlation and PACF graph (right panel) shows lagged terms of serial correlation from linear model for three periods.

The p-value from the ANOVA test of parameters b_1, b_2 and b_3 from the subsequent AR(3) model is 0.037 indicating that there is at least one parameter not equal to zero. The results of this model are shown in Table 4 and we see there is insufficient evidence of a trend in the period 1920 to 1975, but there has been a significant decreasing trend between 1976 and 1995 and there is no trend in the period 1996 to 2014.

Table 4 The coefficient of parameters from AR(3) for three periods.

Parameters	Coefficients	SE	P-value
ar1	0.420	0.029	<0.001
ar2	0.173	0.031	<0.001
ar3	0.151	0.029	<0.001
b_0	1.442	1.096	0.188
b_1	-5.675	2.039	0.005
b_2	-0.159	2.122	0.940
The p-value of ANOVA test is 0.037			

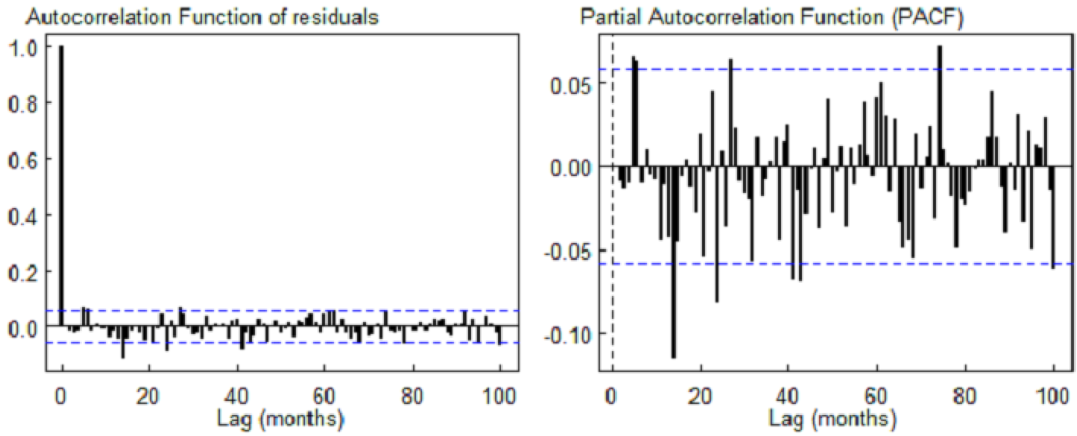

Figure 7 ACF graph in the left panel and PACF graph in the right panel from AR(3) for three periods**.**

Analysis of two periods of data

In this final analysis, only the data from the second and third periods, between 1920 and 1995, are considered, while The first and last period are assumed to be constant. One boxcar function starting at the change point that takes the value of 1 in the period 1976 to 1995 and 0 elsewhere (x_1) is the determinant in the linear model. Table 5 shows that there are significant the upward trend and downward trend in the second and third periods, respectively. The result from PACF graph presented in Figure 8 also suggests that three lagged terms should be the order in AR model.

Table 5 The coefficient of parameters from linear model for two periods.

Parameters	Coefficients	SE	t-value	P-value
b_0	1.409	0.363	3.888	<0.001
b_1	-5.417	0.707	-7.667	<0.001
		The p-value of ANOVA test is <0.001		

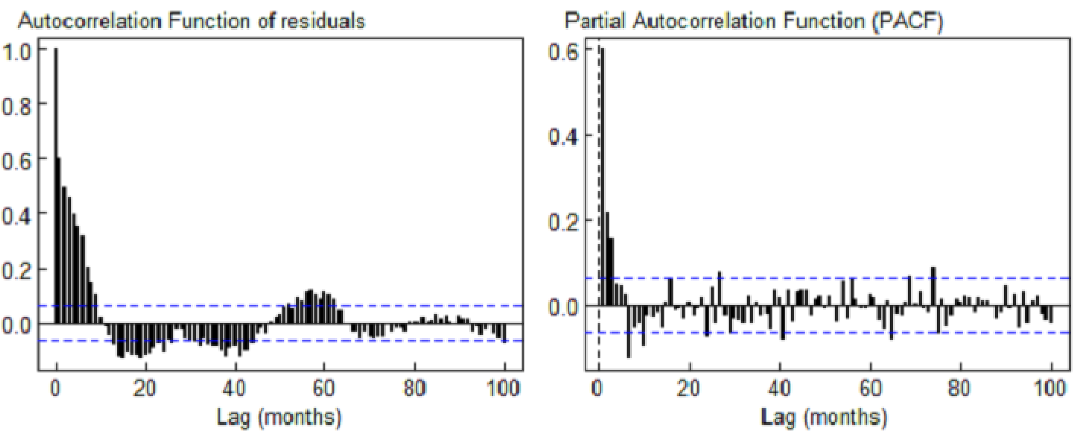

Figure 8 ACF graph (left panel) shows serial correlation and PACF graph (right panel) shows lagged terms of serial correlation from linear model for two periods**.**

The ANOVA test of the parameters b_1 and b_2 from AR(3) model is 0.021. Thus we can conclude there is no significant trend in the second period, a significant decreasing trend in the third period (see Table 6).

Table 6 The coefficient of parameters from AR(3) for two periods.

Parameters	Coefficients	SE	P-value
ar1	0.437	0.033	<0.001
ar2	0.139	0.035	<0.001
ar3	0.158	0.033	<0.001
b_0	1.433	1.038	0.168
b_1	-5.437	1.961	0.006
The p-value of ANOVA test is 0.021			

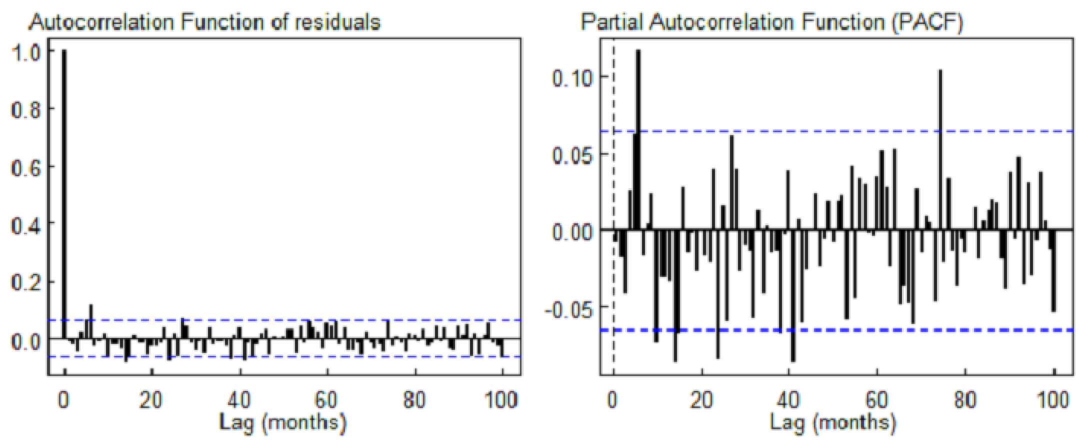

Figure 9 ACF graph in the left panel and PACF graph in the right panel from AR(3) for two periods**.**

Accumulating time series data over time can give a misleading impression, suggesting that a purely random process appears to have an informative structure. As the results have shown, there is insufficient evidence to conclude that the fluctuations in the SOI are due to anything other than random noise. Likewise, the study of Wu and Huang [21] used the SOI as noisy data to illustrate their developed method. However, some researches have suggested that the SOI might be used to predict rainfall in Fiji Islands. (Dennett et al [22]), southern Ghana (Adiku and Stone [10]) and some parts of Australia (Chiew et al [23]) and also predict itself (Chu and Katz [24]) indicating that some other correlation structure may be involved with other variables.

Given that the selected periods of possible changes are based on the data, the p-value is not strictly correct, since the null hypothesis is postulated from the same data used to test it. Even though, we could test the null hypotheses from different periods with this study to define the boxcar function on the data by other criteria such as using the period of every 10, 20 or 30 years, etc. to determine whether or not the variability in SOI is just a random process. An appropriate method needs to be based on sound statistical methodology; it is almost impossible to duplicate this method with the same data because the SOI signal is unique. We could use other methods, including bootstrapping, to test the null hypotheses. However, in this study, we emphasize that the cumulative data has appeared to show some trends, but that we do not have sound statistical evidence to verify this.

Conclusions

It seems reasonable that the cumulative monthly SOI data can be divided into four periods; comprising the years 1876-1919, 1920-1975, 1976-1995 and 1996-2014. We have investigated the apparent increasing and decreasing trends seen in the second and third periods respectively. To examine those trends, the three data sets have been analysed using the linear regression and autoregressive model on the boxcar function. The results from this study show that the appropriate autoregressive model for each of the three data sets is AR(3). For the model which included only the data from periods two and three, we found no significant trend in the period 1920 to 1975 but a significant decreasing trend in the period 1976 to 1995. However, the result from the analysis of the data including all four periods shows no evidence of trend for any periods. Consequently, the evidence suggests that the fluctuations in the SOI seem to be similar with a random noise process.

Acknowledgements

This research is supported by the Centre of Excellence in Mathematics, Commission on Higher Education, Thailand. We gratefully acknowledge Emeritus Professor Don McNeil and Assistant Professor Phattrawan Tongkumchum for their helpful suggestions.

References

[1] SG Philander. *Encyclopedia of Global Warming and Climate Change*. 2nd ed. SAGE publication, Inc, Thousand Oaks, 2012, p. 1453.

[2] S Carbone, RL Manfro, AMP Schuch, MB da Rosa, DK Pinheiro, MT Kayano, and NJ Schuch. Correlation between southern oscillation index and total ozone column using Brewer and Toms at southern Brazil (29.4°S, 53.8°W) from 1997 to 2003, EGU General Assembly 2005, Available at: http://mtc-m18.sid.inpe.br/col/sid.inpe.br/ePrint@1905/2005/09.26.17.44/doc/correlation %20between.pdf, accessed April 2015.

[3] YH Jin, A Kawamura, K Jinno and Y Iseri. On the long-term variability of Southern Oscillation Index. *In*: Proceeding of 2003 Korea Water Resources Association. 2003, p. 151-158.

[4] JI Salisbury and M Wimbush. Using modern time series analysis techniques to predict ENSO events from the SOI time series. *Nonlinear Proc. Geoph*. 2002; **9**, 341-345.

[5] M Hadiani, SJ Asl, MR Banafsheh, Y Dinpajouh and E Yasari. Investigation the Southern Oscillation Index Effect on Dry/Wet Periods in North of Iran. *Int. J. Agric. Crop Sci*. 2012; **4**, 1291-1299.

[6] RS Tigona and CR de Freitas. Relationship between the Southern Oscillation and rainfall in Vanuatu. *Weather Clim*. 2012; **32**, 43-50.

[7] DA Jones and BC Trewin. On the relationships between the el nino-southern oscillation and Australian land surface temperature. *Int. J. Climatol*. 2000; **20**, 697-719.

[8] MS Halpert and CF Ropelewski. Surface temperature patterns associated with the southern oscillation. *J. Clim*. 1992; **5**, 577-593.

[9] R Stone and A Auliciems. SOI phase relationships with rainfall in eastern Australia. *Int. J. Climatol*. 1992; **12**, 625-636.

[10] SGK Adiku and RC Stone. Using the Southern Oscillation Index for improving rainfall prediction and agricultural water management in Ghana. *Agr. Water Manage*. 1995; **29**, 85-100.

[11] P Carberry, D George and D Buckley. The southern oscillation index and southern Australia, NSW department of primary industries, Available at: http://www.dpi.nsw.gov.au/__data/assets/pdf_file/0007/138139/the-southern-oscillation-index-and-southern-australia.pdf, accessed April 2015.

[12] MF Al-Zuhairi, JH Kadhum and AJ Farhan. Time Series Analysis of Southern Oscillation Index and Temperature over Iraq. *Diyala J. Pure Sci*. 2013; **9**, 57-68.

[13] A Watts. The story told by the Southern Oscillation Index, Available at: http://wattsupwiththat.com/2010/12/09/the-story-told-by-the-southern-oscillation-index/, accessed April 2015.

[14] AJ Troup. The Southern Oscillation. *Q. J. R. Meteorol. Soc*. 1965; **91**, 490-506.

[15] Australian Government Bureau of Meteorology. Sothern Oscillation Index, Available at: http://www.bom.gov.au/climate/glossary/soi.shtml, accessed May 2015.

[16] J Gill. *Generalized Linear Models*: *A Unified Approach*. Sage Publications, Inc., Thousand Oaks, 2001, p. 6-7.

[17] WN Venables and BD Ripley. *Modern Applied Statistics with S*. 4th ed. Springer Science and Business Media, LLC, New York, 2002, p. 144-145.

[18] EW Weisstein. Boxcar function, Available at: http://mathworld.wolfram.com/BoxcarFunction.html, accessed April 2015.

[19] RH Shumway and DS Stoffer. *Time series analysis and its applications with R examples*. Springer Science and Business Media, LLC, New York, 2011, p. 11-17, 37.

[20] R Development Core Team. R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna. Austria: R Foundation for Statistical Computing, Available at: http://www.R-project.org, accessed March 2015.

[21] Z Wu and NE Huang. A study of the characteristics of white noise using the empirical mode decomposition method. *In*: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. 2004, p. 1597-1611.

[22] MD Dennett, J Elston and PC Prasad. Seasonal rainfall forecasting in Fiji and the Southern Oscillation. *Agr. Meteorol*. 1978; **19**, 11-22.

[23] FHS Chiew, TC Piechota, JA Dracup and TA McMahon. E1 Nino/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting. *J. Hydrol*. 1998; **204**, 138-149.

[24] P Chu and RW Katz. Modeling and forecasting the Southern Oscillation: A time-domain approach. *Mon. Weather Rev*. 1985; **113**, 1876-1888.

Appendix IV:

Proceeding

ISI-RSC 2014 ABSTRACT BOOK

Fitting Daily Rainfall Amount in Thailand Using Gamma Regression

Mayuening Eso*
Prince of Songkla University, Pattani, Thailand
mayening@hotmail.com

Metta Kuning
Prince of Songkla University, Pattani, Thailand
metta.kuning@gmail.com

The objective of this study is to develop an appropriate generalized linear model of daily rainfall during 2001-2012 in Thailand. In this study, serial correlations are removed by restructuring the data as 5-day means, because conventional statistical models assume independent errors. In multivariate analysis, factor analysis is used to reduce a large numbers of stations. It is found that seven regions with similar patterns of daily rainfall in Thailand are the upper north, lower north, north-east, central, south-west, upper south and south-east. Gamma regression is fitted in the period of rainy season for each region.

Key Words: Daily Rainfall in Thailand, Factor Analysis, Generalized Linear Model, GLM

Confidence Bands for Confidence Intervals from Data of Two Parameters Exponential Distribution under Complete Censoring (Study Case: Waiting Time of Earthquake Disasters in Indonesia at March 2013)

Puteri Pekerti Wulandari
Students of Statistics
Universitas Islam Indonesia, Yogyakarta, Indonesia
pekertiputeri@yahoo.com

Akhmad Fauzy
Lecturer of Statistics
Universitas Islam Indonesia, Yogyakarta, Indonesia
afauzy@yahoo.com

Earthquakes are one of the frequent disasters in Indonesia. Based on data from the National Disaster Mitigation Institution (BNPB) recorded in March 2013 has occurred 7 times earthquake. This research was aims to find confidence bands for confidence interval from data of exponential distribution two parameters under complete censoring.
The results of the data analysis showed that the earthquake in March 2013 has been distributed exponential with time interval earthquake, the fastest is 1 day and the longest was 7 days. The average waiting time for an earthquake parameter μ is 1 day and the parameter θ is 3 day. For the average waiting time of 1 day have a probability 86.7% and the average to waiting time of 7 days have a probability 14.3%. From the plot it can be seen that the resulting confidence band moves decreases exponentially.

Key Words: Exponential, Estimates, Earthquake, Intervals, The Complete Sensor

International Statistical Institute
Regional Statistics Conference 2014
Statistical Science for a Better Tomorrow
Kuala Lumpur, Malaysia | 16-19 November 2014

Certificate of Appreciation

This is to certify that

Mayuening Eso

Contributed to the International Statistical Institute
Regional Statistics Conference 2014

Marzunisham Omar
Chairman, Local Organising Committee

Organised by: International Statistical Institute South East Asia Regional Network (SEA NETWORK)
In collaboration with: BANK NEGARA MALAYSIA (CENTRAL BANK OF MALAYSIA), Jabatan Perangkaan MALAYSIA, MALAYSIA INSTITUTE OF STATISTICS

International Statistical Institute
Regional Statistics Conference 2014
Statistical Science for a Better Tomorrow
Kuala Lumpur, Malaysia | 16-19 November 2014

Certificate of Attendance

This is to certify that

Mayuening Eso

Attended the International Statistical Institute
Regional Statistics Conference 2014

Marzunisham Omar
Chairman, Local Organising Committee

Organised by: International Statistical Institute South East Asia Regional Network (SEA NETWORK)
In collaboration with: BANK NEGARA MALAYSIA (CENTRAL BANK OF MALAYSIA), Jabatan Perangkaan MALAYSIA, MALAYSIA INSTITUTE OF STATISTICS

Fitting Daily Rainfall Amount in Thailand Using Gamma Regression

Mayuening Eso*

Prince of Songkla University, Pattani, Thailand, mayening@hotmail.com

Metta Kuning

Prince of Songkla University, Pattani, Thailand, metta.kuning@gmail.com

The objective of this study is to develop an appropriate generalized linear model of daily rainfall during 2001-2012 in Thailand. In this study, serial correlations are removed by restructuring the data as 5-day means, because conventional statistical models assume independent errors. In multivariate analysis, factor analysis is used to reduce a large numbers of stations. It is found that seven regions with similar patterns of daily rainfall in Thailand are the upper north, lower north, north-east, central, south-west, upper south and south-east. Gamma regression is fitted in the period of rainy season for each region.

Key Words: daily rainfall in Thailand, factor analysis, generalized linear model, GLM

Introduction

Studying of the characteristics and patterns of rainfall can provide important information for water management to prevent floods and landslides especially in agricultural countries. Rainfall data have been analyzed by researchers in several methods and several countries. Aksoy (2000) used 2-parameter gamma distribution to fit the daily rainfall data in Istanbul, Turkey. Ejieji (2004) analyzed the distribution of daily rainfall amounts which were modeled to gamma and exponential distribution in vegetative zones, Nigeria. Cho et al. (2004) compared gamma and lognormal distributions of rainfall dataset provided by the Tropical Rainfall Measuring Mission

(TRMM) satellite. Duan et al. (1998) investigated daily rainfall data for U.S. Pacific Northwest using weibull and gamma distribution.

In Thailand, Szyniszewska and Waylen (2012) used markov chain and gamma distribution for finding the probability of rainfall occurrence and calculating amount of rainfall in wet days, respectively. The monthly rainfall totals were used to examine the daily rainfall characteristics in central and northeastern Thailand. Phien et al. (1980) analyzed the distribution of the monthly rainfall sequences at different stations in the northeast of Thailand. The leakage law was fitted very well for monthly rainfall sequences data. The lognormal and gamma distributions were also fitted well for the data without zero values.

It can be seen that gamma distribution method is commonly used for analyzing daily rainfall data series. Consequently, daily rainfall data in this study are analyzed to develop an appropriate generalized linear model with gamma regression. R program (2013) is used to be a tool for analyzing data and creating graphs.

Methodology

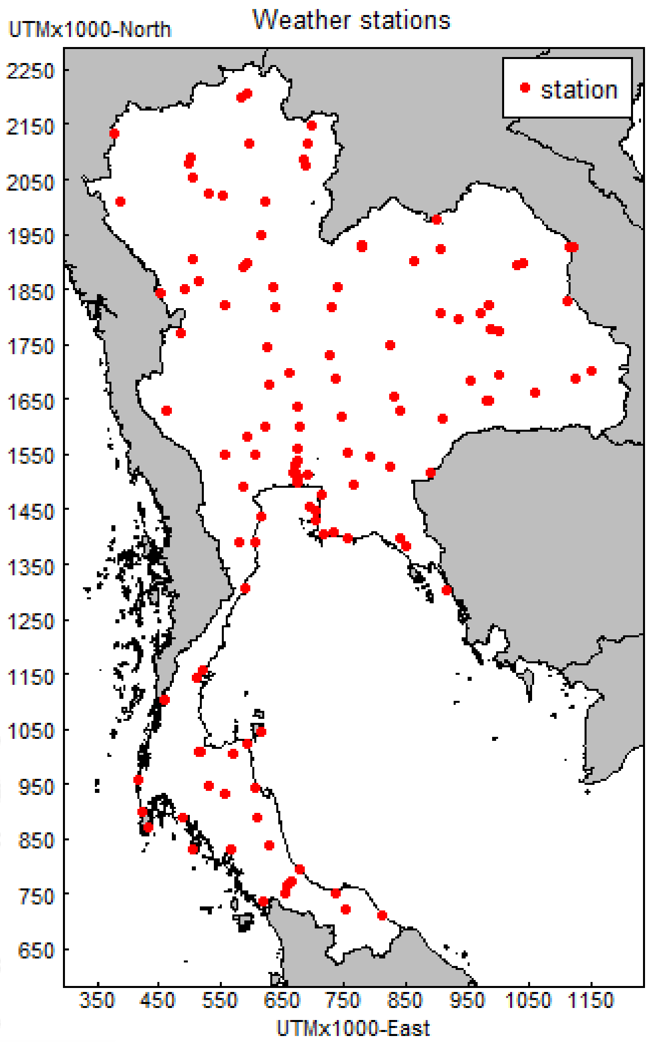

Figure 1 the weather stations of the study in Thailand

In the study, the daily rainfall data are provided by Thai Meteorological Department (TMD). The locations of the 114 weather stations in this study can be seen in Figure 1. The data are obtained from February 5, 2001 to February 4, 2013. The rain falling on February, 29 for every leap year are excluded. The daily data are manipulated to 5-day average for eliminating the problem of serial correlations because of the conventional statistical assumption. Therefore, the data are reduced to 73 periods in each year.

Factor analysis technique (Johnson and Wichern (2007); Rencher (2002); Venables and Ripley (2002)) is used to distinguish the 114 stations to be a smaller number of regions with promax rotation and Maximum likelihood extraction estimating factor loadings. The 5-day average in rainy season (*y*), which is a set of skewed data of each region, is fitted using the generalized linear model (GLM) with gamma distributions (Gill (2001); McCullagh and Nelder (1989); Venables and Ripley (2002)). The model is shown as follows:

$$1/\text{mean}(y) = \text{constant} - \text{factor}(year) - \text{factor}(month)$$

Sum contrasts (Venables and Ripley (2002); Tongkumchum and McNeil (2009)) is used to compare rainfall mean in each year and month with the overall rainfall mean.

Results

The factor analysis allows a reduction of the number of stations from 114 to 7 regions, which comprise of upper north (14 stations), lower north (10 stations), north-east (33 stations), central (33 stations), south-west (9 stations), upper south (2 stations) and south-east (13 stations). The aforementioned seven regions are displayed on the map in Figure 2.

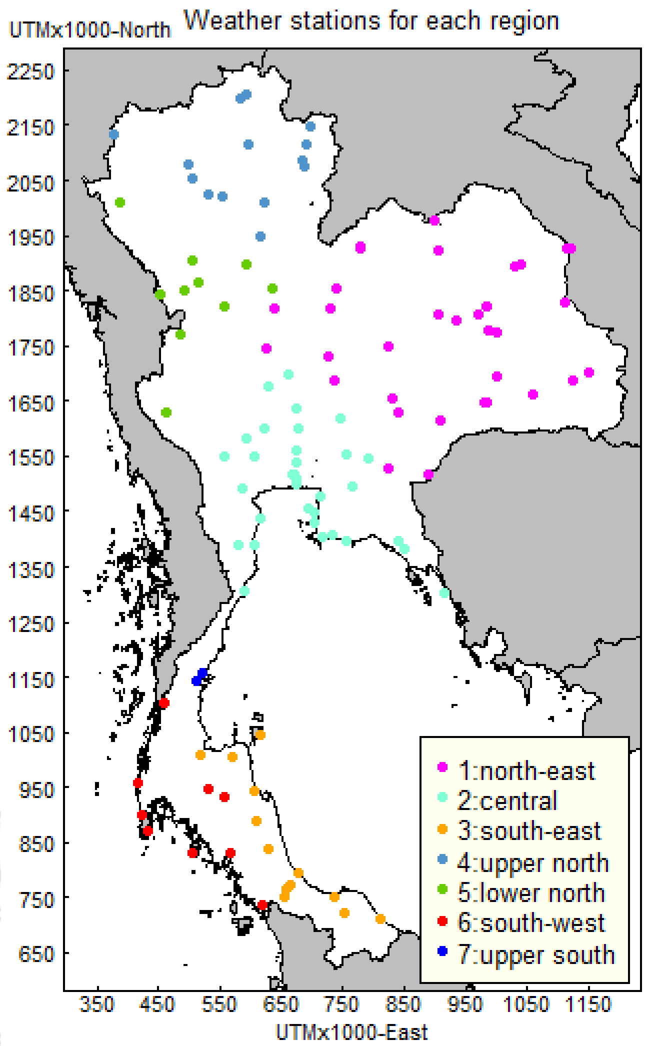

Figure 2 the monthly average of seven regions

The overall mean of the rainfall amount for all regions is 4.43 mm/day, and a series of successive days with an average of daily rainfall over than the overall mean is defined a rainy season. The overall average of rainfall in the rainy season for south-east, south-west, upper south, north-east, upper north, lower north and central are 10.71, 9.16, 8.19, 7.90, 7.40, 7.24, 6.25 mm/day, respectively.

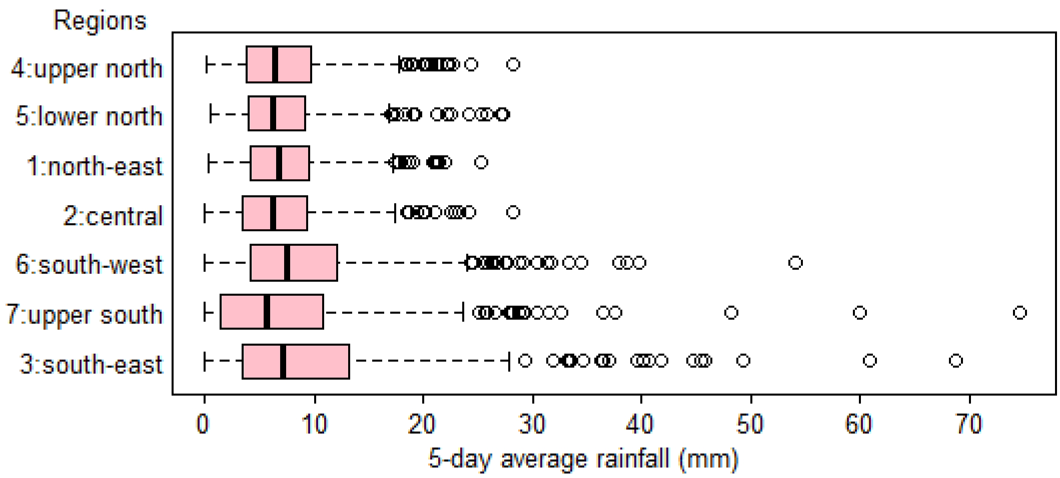

Figure 3 Boxplot of seven regions

The boxplot in Figure 3 shows the distribution of daily rainfall of all regions. It is evident that the daily rainfalls for all regions are skewed to the right.

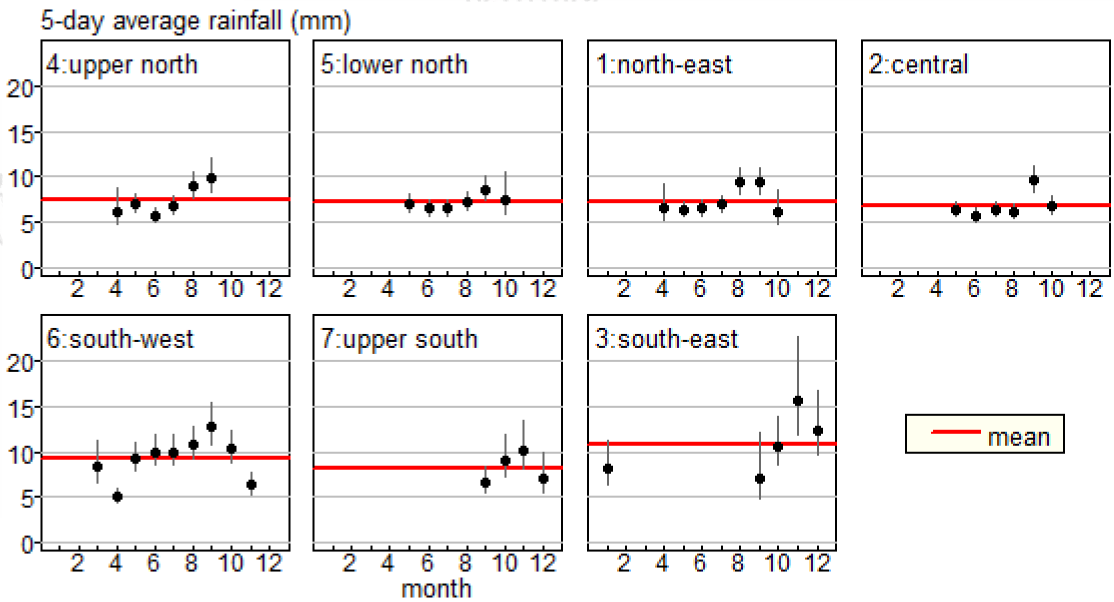

Figure 4 Confidence interval of daily rainfall in each month from gamma model

The result from the gamma model is shown in Figure 4 and 5. There are seven models corresponding to the seven regions. In the upper north of Thailand, the months of June, August, and September are statistically significantly different from its overall mean of the region, and the last two are over the mean. For the lower north, the average rainfall is over the overall mean in September. In the north-east, the model reveals that the average rainfalls in August and September are different from its overall mean as the results of both months are above the overall mean. In the central region, the months June and September are different from its overall mean by statistical significance. While the average rainfall in June is under the overall mean, that of September is clearly more than the overall mean in the region.

For the southern part of Thailand, the south-west region has six months that are different from its overall mean. The average of rainfall in April and November are less than its overall mean and others are over the overall mean. The upper south and south-east have only one month, November, different from their overall means.

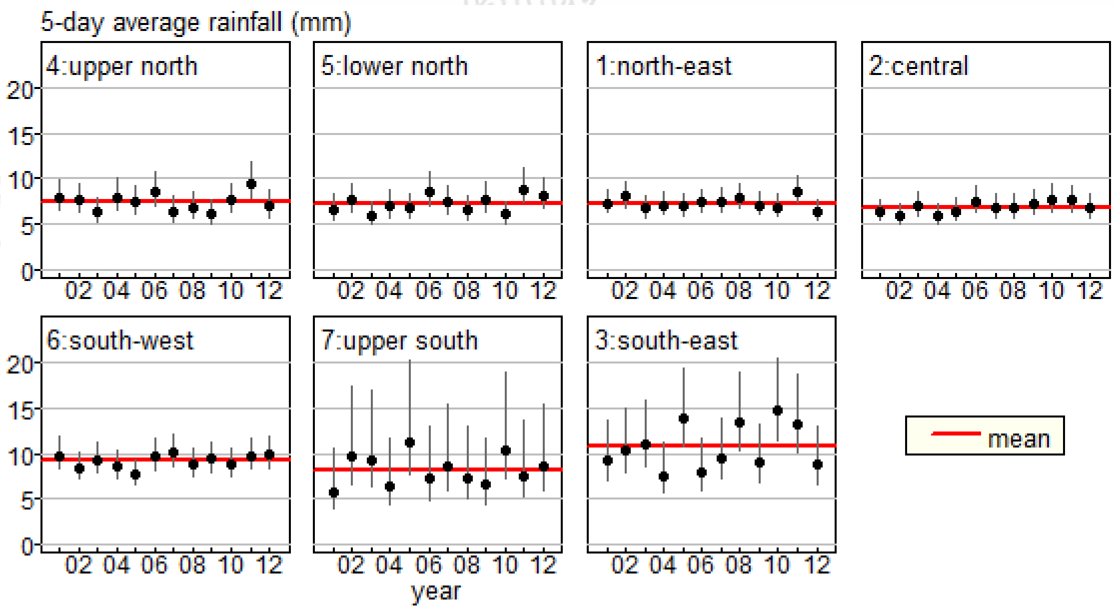

Figure 5 Confidence interval from gamma model of daily rainfall for each year

In terms of yearly results, the upper north, lower north and north-east have only one year, 2011, in which the average rainfall is different from their overall mean by statistical significance. The south-east has three years, 2005, 2008, and 2010, in which their average rainfalls differ from their overall means. However, there is not a single year that the daily rainfall is different from its overall mean in the central, the south-west, and the upper south. The analysis also reveals that the regions that show statistical significant daily average rainfall in any year have daily average rainfall above the overall mean.

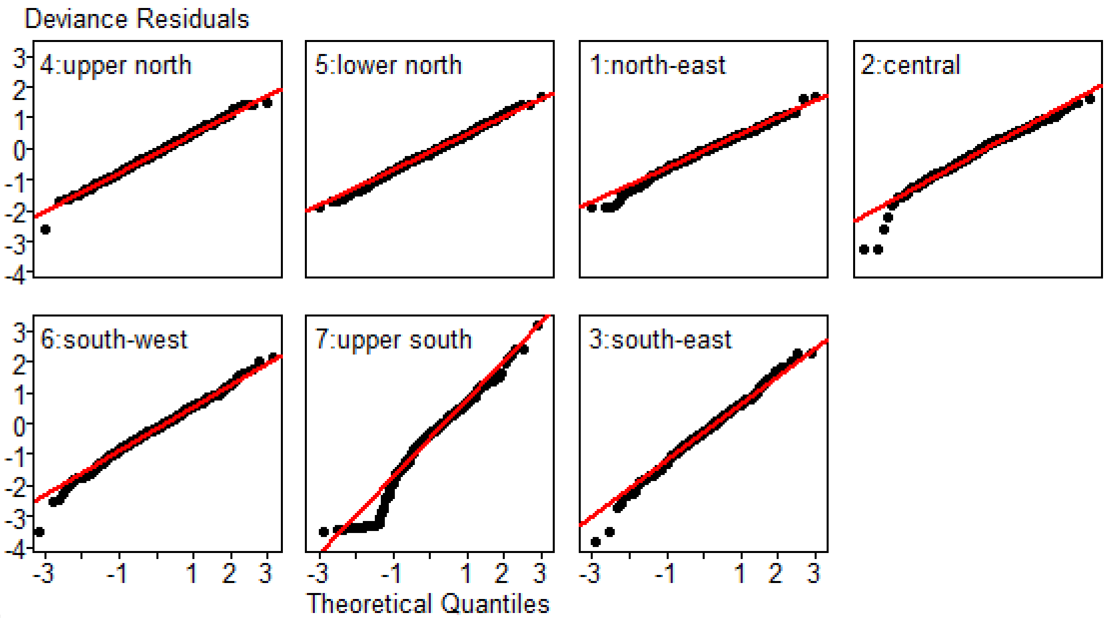

Figure 6 Deviance residuals from gamma model for seven regions

The deviance residuals plots from the model are shown in Figure 6, and it reveals that the model provides a reasonable fit for six regions except for the upper south.

Conclusions and discussions

The technique of factor analysis can divide the regions of Thailand to seven regions including the upper north, the lower north, the north-east, the central, the south-west, the upper south and the south-east. Because of right skewed data, the gamma regression is fitted for all regions. The deviances residuals of six regions from this model reveal moderately fit except the upper south, because there are only two stations in the region. In terms of the analysis tool, it can be seen that the result is similar to the studies from Phien et al. (1980), Duan et al (1998), Aksoy (2000), Cho et al. (2004), Ejieji (2004), and Szyniszewska and Waylen (2012) which have indicated that gamma distribution can be utilized for long sequences daily rainfall data on wet days.

Moreover, this study shows that the months with high rainfall during a rainy season are September for five regions, including the upper north, lower north, north-east, central, and south-west, and November for other two regions which are the south-east and upper south. The results from this study can be beneficial for the water resources management in planning the periods to release water from a reservoir for preventing floods, and to reserve water to avoid an extensive drought problem for each region.

Acknowledgement

This research is supported by grant funds from the program Strategic Scholarships Fellowships Frontier Research Networks (Specific for Southern region) for the Ph.D. Program Thai Doctoral degree from Office of the Higher Education Commission, Thailand and also by the Centre of Excellence in Mathematics, the Commission on Higher Education, Thailand. We gratefully acknowledge Emeritus Professor Don McNeil for his assistance and suggestions. We are also grateful to the Thai Meteorological Department (TMD) for providing the daily rainfall data.

References

Aksoy, H. (2000) “Use of Gamma Distribution in Hydrological Analysis”, *Turkish Journal of Engineering and Environmental Sciences*, 24, 419-428.

Cho, H., Bowman, K. P. and North, G. R. (2004) “A comparison of gamma and lognormal distributions for characterizing satellite rain rates from the tropical rainfall measuring mission”, *Journal of Applied Meteorology*, 43, 1586-1597.

Duan, J., Belker, J. and Grant, G. E. (1998) "Evaluation of probability density functions in precipitation models for the Pacific Northwest", *Journal of the American Water Resources Association*, 34, 617-627.

Ejieji, C. J. (2004) "Evaluation of two models for the distribution of daily rainfall amounts in Nigeria", *Journal of Agricultural Research and Development*, 3, 61-74.

Gill, J. (2001) *Generalized Linear Models: A Unified Approach*, Sage Publications, Inc., Thousand Oaks, U.S.A.

Johnson, R. A. and Wichern, D. W. (2007) *Applied Multivariate Statistical Analysis*, Pearson Prentice Hall, New Jersey, U.S.A.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, Chapman and Hall, London, England.

Phien, H. N., Arbhabhirama, A. and Sunchindah, A. (1980) "Distribution of monthly rainfall in northeast Thailand", *Southeast Asian Studies*, 18, 110-123.

R Development Core Team. (2013) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna. Austria: R Foundation for Statistical Computing. Available: http://www.R-project.org. [April 18, 2014].

Rencher, A. C. (2002) *Methods of Multivariate Analysis*, John Wiley and Sons, Inc., New York, U.S.A.

Szyniszewska, A. M. and Waylen, P. R. (2012) "Determining the daily rainfall characteristics from the monthly rainfall totals in central and northeastern Thailand", *Applied Geography*, 35, 377-393.

Tongkumchum, P. and McNeil, D. (2009) “Confidence intervals using contrasts for regression model”, *Songklanakarin Journal of Science and technology*, 31, 151-156.

Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S*, Springer Science and Business Media, LLC, New York, U.S.A.

Appendix V:

Research Collaborations

UTS:C3

UNIVERSITY OF TECHNOLOGY SYDNEY

Plant Functional Biology and Climate Change Cluster
City Campus
PO Box 123 Broadway
NSW 2007 Australia
www.c3.uts.edu.au
UTS CRICOS PROVIDER CODE 00099F

January 20, 2015

Attn of: Dr Yupadee Chaisuksan
Vice President for Academic and International Affairs,
Pattani Campus Prince of Songkla University

Dear Dr. Chaisuksant,

I would like to welcome and formally invite students, Mayuening Eso and Potjamas Chuangchang, to visit Sydney, Australia and participate in collaboration and research activities in my remote sensing laboratory within the Plant Functional Biology and Climate Change Cluster at the University of Technology Sydney.

A 3-month visit would be very helpful with a suggested timeframe from 22 February to 22 May 2015.
Dr. Don McNeil, their official advisor, will assist with this collaborative visit and help arrange accommodation and travel.

I very much look forward to their visit and for working together in joint research activities in remote sensing and Thailand ecosystem and biotic resources.
Please let me know if you would like further details regarding this proposed visit or anything else,

sincerely,

Alfredo Huete, Prof.

Plant Functional Biology and Climate Change Cluster
School of Environmental Sciences
University of Technology, Sydney
15 Broadway Road, Ultimo, NSW 2007
Australia

UNIVERSITY OF TECHNOLOGY SYDNEY

Plant Functional Biology and Climate Change Cluster
City Campus
PO Box 123 Broadway
NSW 2007 Australia
www.c3.uts.edu.au
UTS CRICOS PROVIDER CODE 00099F

May 4, 2015

Attn of: Dr Yupadee Chaisuksan
Vice President for Academic and International Affairs,
Pattani Campus Prince of Songkla University

Dear Dr. Chaisuksant,

I would like to welcome and formally invite student, Ms. Mayuening Eso as a visiting student scholar in my Remote Sensing research program at the University of Technology, Sydney (UTS), in Australia. Ms. Eso will participate in collaboration and research activities in my laboratory on topics of time series analyses, land surface dynamics, and climate change.

A 3-month visit would be very helpful with a suggested timeframe from 4 July to 3 October 2015. This will provide us an opportunity to generate a joint research publication. Dr. Don McNeil will assist with this collaborative visit and help arrange accommodation and travel.

I very much look forward to this opportunity for scholarly exchange. Please let me know if you would like further details regarding this proposed visit or anything else,

sincerely,

Alfredo Huete, Prof.

Plant Functional Biology and Climate Change Cluster
School of Environmental Sciences
University of Technology, Sydney
15 Broadway Road, Ultimo, NSW 2007
Australia

Vitae

Name: Miss Mayuening Eso

Student ID: 5420330012

Educational Attainment:

Degree	Name of Institution	Year of Graduation
Bachelor of Science (Applied Mathematics)	Prince of Songkla University	2007
Master of Science (Research Methodology)	Prince of Songkla University	2010

Scholarship Awards during Enrolment:

1. Grant funds from the program of Strategic Scholarships Fellowships Frontier Research Networks (Specific for Southern Region) for the Ph.D. Program Thai Doctoral degree from the Office of the Higher Education Commission, Thailand.

2. Research scholarship from (i) the Centre of Excellence in Mathematics, the Commission on Higher Education, Thailand, (ii) the Graduate School, Prince of Songkla University, Thailand, and (iii) the organization of Deep South Coordination Center (DSCC), Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Thailand.

3. Scholarship for proceeding from International Statistical Institute (ISI) South East Asia Regional Network.

4. Scholarship for visiting University of Malaya, Malaysia from the Faculty of Science and Technology, Prince of Songkla University, Thailand.

Work-Position and Address:

Position: Programmer

Address: Deep South Coordination Center, 3th floor 51B Building, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus.

List of Publications and Proceedings

Publications:

Eso, M., Kuning, M. and Chuai-Aree, S. 2015. Analysis of daily rainfall during 2001-2012 in Thailand. Songklanakarin Journal of Science and Technology. 37, 81-88.

Eso, M., Kuning, M., Green, H., Ueranantasun, A. and Chuai-Aree, S. 2016. The Southern Oscillation Index as a Random Walk. Walailak Journal of Science and Technology (in press).

Proceedings:

Eso, M. and Kuning, M. 2014. Fitting Daily Rainfall Amount in Thailand Using Gamma Regression. International Statistical Institute Regional Statistics Conference 2014: Statistical Science for a Better Tomorrow. 16-19 November 2014, Sasana Kijang, Kuala Lumpur, Malaysia.