Spatio and Temporal Statistical Modeling of Earth Surface Temperatures

Wandee Wanishsakpong

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

Prince of Songkla University

2015

Thesis Title Spatio and Temporal Statistical Modeling of Earth Surface Temperatures

Author Miss Wandee Wanishsakpong

Major Program Research Methodology

Major Advisor	**Examining Committee:**
………………………………….	…………………………..Chairperson
(Emeritus Prof. Dr. Don McNeil)	(Dr. Somporn Chuai-Aree)
	....
Co-advisors	(Emeritus Prof. Dr. Don McNeil)
…………………………………	…………………………………
(Asst. Prof. Dr. Nittaya McNeil)	(Asst. Prof. Dr. Nittaya McNeil)
....	....
(Asst. Prof. Dr. Phattrawan Tongkumchum)	(Asst. Prof. Dr. Phattrawan Tongkumchum)
	…………………………………..
	(Assoc. Prof. Dr. Rohana Binti Jani)

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 Wandee Wanishsakpong)

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 Wandee Wanishsakpong)

Candidate

Thesis Title Spatio and Temporal Statistical Modeling of Earth Surface Temperatures

Author Miss Wandee Wanishsakpong

Major Program Research Methodology

Academic Year 2015

ABSTRACT

In this thesis, graphical and statistical methods were used to determine the trends and patterns of temperature change as well as to forecast of future temperatures of earth surface temperature change. The thesis comprises of two studies carried out for different issues, temperature change in Arctic and temperature change in Australia.

In the first study, monthly temperature anomalies variation from 1973 to 2008 above latitude 45 degrees north, covering the Arctic Ocean, northern areas of the Atlantic and Pacific Oceans, and the Asian and European Continents, were examined. Temperature data were obtained from the Climate Research Unit (CRU) of the United Kingdom. First, a linear regression model was used to investigate the trends and patterns of temperature change of 69 sub-regions. A second order autoregressive process was used to reduce auto correlation at lag 1 and 2 months. Factor analysis was then used to account for spatial correlation. Twelve large regions having similar temperature change patterns in each large region were identified. A 95% confidence interval (CI) of temperature change was estimated for each of the 12 large regions. Each large region was reclassified into three levels. High rates of temperature increase (0.20˚C - 0.32˚C) occurred in the North Pacific Ocean, Alaska and Eastern Siberia. Moderate temperature increases (0.13˚C - 0.19˚C)

occurred in northern Canada, Greenland, Iceland, Norway, Sweden and Finland. Northern Siberia and part of the North Atlantic had low level increases (0.09˚C - 0.129˚C) while northeast Canada and its surrounding seas did not show evidence of warming.

In the second study, daily maximum temperature data from 1970 to 2012 in Australia were described. The data were obtained from 85 weather stations randomly selected from a total of more than 700 stations of the Australia Bureau of Meteorology (BOM). This study is based on two data sets, temperature change in daily maximum temperatures over consecutive 5-day periods and monthly maximum temperatures. For the first data set, the variation of daily maximum temperatures over consecutive 5-day periods was examined. A linear regression model was initially used to model seasonally adjusted daily maximum temperatures. The data were fitted with a first order autoregressive process to reduce auto correlation at lag 1 month. Factor analysis was used to classify the temperatures from the 85 stations into seven factors corresponding to seven geographical regions. Average maximum annual temperature in these seven regions ranged from 23˚C to 36˚C. A sixth order polynomial regression model was fitted in these seven regions. Trends and patterns of temperatures were found to be similar in the central, eastern, southern and southeastern parts of the country. These trends and pattern show an increase in temperature after 1974 and decrease in temperature around 1984 and with another steady increase from 2000 to 2005 and decrease through 2012. In the second data set, maximum monthly temperatures were defined as the highest daily temperature in a particular month. Missing values in the data were estimated using a regression model accounting for information from the nearest stations as well as the time periods. Then factor analysis was utilized to reduce the dimensions of the data set. A limitation of this factor analysis is that some of the stations were not clearly separated. Cluster analysis was used to classify the factor loading produced four clusters of stations. A quartic trend model with 3^{nd} order time lag was fitted for each cluster and forecasting maximum temperatures over the short period. The results showed that the forecasted maximum monthly temperatures were decreasing during the period of 2013-2015. A 95% confidence interval (CI) of the maximum monthly temperature predictions ranged from about 27˚C - 44˚C, 26˚C - 40˚C, 31˚C - 42˚C and 26˚C - 42˚C in clusters 1, 2, 3 and 4, respectively.

ชื่อวิทยานิพนธ์ แบบจำลองทางสถิติของอุณหภูมิพื้นผิวโลกต่อพื้นที่และเวลา

ผู้เขียน นางสาววันดี วณิชย์ศักดิ์พงศ์

สาขาวิชา วิธีวิทยาการวิจัย

ปีการศึกษา 2558

บทคัดย่อ

วิทยานิพนธ์ฉบับนี้ใช้วิธีกราฟและวิธีการทางสถิติศึกษาแนวโน้ม รูปแบบ และพยากรณ์การเปลี่ยนแปลงของอุณหภูมิพื้นผิวโลก โดยประกอบด้วยสองกรณีศึกษาคือ การเปลี่ยนแปลงของอุณหภูมิพื้นผิวบริเวณอาร์กติก และการเปลี่ยนแปลงของอุณหภูมิพื้นผิวของประเทศออสเตรเลีย

การศึกษาแรก ศึกษาข้อมูลอุณหภูมิรายเดือนในบริเวณเหนือละติจูด 45 องศา เหนือ ในช่วงปี ค.ศ.1973 ถึง ค.ศ.2008 ครอบคลุม มหาสมุทรอาร์กติก พื้นที่ตอนเหนือของมหาสมุทรแอนแลนติกและแปซิฟิก ทวีปเอเชีย และทวีปยุโรป ข้อมูลได้มาจากการเก็บรวบรวมจากหน่วยวิจัย Climate Research Unit ประเทศ สหราชอาณาจักร การวิเคราะห์ข้อมูลเริ่มต้นด้วยการศึกษาการเปลี่ยนแปลงของอุณหภูมิใน 69 พื้นที่ โดยใช้ตัวแบบการถดถอยเชิงเส้น ขจัดสหสัมพันธ์ในตัวเองกับช่วงเวลาที่ศึกษา ด้วยขบวนการสหสัมพันธ์ในตัวเองอันดับสอง และใช้การวิเคราะห์ปัจจัยเพื่อจัดกลุ่มพื้นที่ที่มีลักษณะการเปลี่ยนแปลงของอุณหภูมิที่คล้ายคลึงกันเป็นกลุ่มเดียวกัน ผลการศึกษาพบว่า สามารถจัดกลุ่มพื้นที่ได้ 12 กลุ่มพื้นที่ และประมาณการเปลี่ยนแปลงของอุณหภูมิในแต่ละกลุ่มพื้นที่ด้วยช่วงความเชื่อมั่น 95% ของการเปลี่ยนแปลงของอุณหภูมิ และจัดกลุ่มการเปลี่ยนแปลงของอุณหภูมิเป็น 3 ระดับดังนี้ การเปลี่ยนแปลงของอุณหภูมิระดับสูงอยู่ในช่วง 0.2 องศาเซลเซียส ถึง 0.32 องศาเซลเซียสต่อทศวรรษ เกิดขึ้นในบริเวณเหนือมหาสมุทรแปซิฟิก อลาสก้า และทางตะวันออกของไซเบียเรีย การเปลี่ยนแปลงของอุณหภูมิระดับปานกลางอยู่ในช่วง 0.13 องศาเซลเซียส ถึง 0.19 องศาเซลเซียสต่อทศวรรษ เกิดขึ้นในบริเวณตอนเหนือของแคนาดา กรีนแลนด์ ไอซ์แลนด์ นอร์เวย์ สวีเดน และฟินแลนด์ สำหรับการเปลี่ยนแปลงของอุณหภูมิระดับต่ำอยู่ในช่วง 0.09 องศาเซลเซียส ถึง 0.129 องศาเซลเซียสต่อทศวรรษ เกิดขึ้นในบริเวณทางตอน

เหนือของไซเบียเรีย บางส่วนทางตอนเหนือของแอนแลนติก ขณะที่ทางตะวันออกเฉียงเหนือของแคนาดาและทะเลที่อยู่ล้อมรอบไม่พบการเปลี่ยนแปลงของอุณหภูมิ

การศึกษาที่สอง ศึกษาอุณหภูมิสูงสุดรายวันของประเทศออสเตรเลีย ในช่วง ค.ศ.1970 ถึง ค.ศ. 2012 ข้อมูลถูกสุ่มมา 85 สถานีจาก Australia Bureau of Meteorology (BOM) ซึ่งมีสถานีในการเก็บรวบรวมข้อมูลอุณหภูมิมากกว่า 700 สถานี นำข้อมูลมาจัดการเป็นสองชุด ข้อมูลชุดแรก ศึกษาความผันแปรของอุณหภูมิสูงสุดราย 5 วัน โดยเริ่มต้นใช้รูปแบบการถดถอยเชิงเส้นในการปรับอิทธิพลของฤดูกาล วิเคราะห์ขบวนการสหสัมพันธ์ในตัวเองอันดับหนึ่งเพื่อลดสหสัมพันธ์ในตัวเอง และใช้วิธีการวิเคราะห์ปัจจัย เพื่อจัดกลุ่มความสัมพันธ์ของอุณหภูมิจาก 85 สถานีไปเป็น 7 กลุ่มของสถานีซึ่งสอดคล้องกับลักษณะทางภูมิศาสตร์ของประเทศออสเตรเลีย อุณหภูมิเฉลี่ยรายปีทั้ง 7 กลุ่มอยู่ในช่วง 23 องศาเซลเซียส ถึง 36 องศาเซลเซียส รูปแบบสมการถดถอยโพลิโนเมียลดีกรีหกถูกกำหนดเพื่อศึกษาแนวโน้มและรูปแบบของอุณหภูมิใน 7 กลุ่มดังกล่าว ผลการศึกษาพบว่า แนวโน้มและรูปแบบการเปลี่ยนแปลงของอุณหภูมิคล้ายคลึงกันในตอนกลาง ตะวันออก ใต้ และบางส่วนของทางตะวันออกเฉียงใต้ รูปแบบและแนวโน้มแสดงการเพิ่มขึ้นของอุณหภูมิหลังปี ค.ศ.1974 และลดลงในปี ค.ศ.1984 และมีการเพิ่มขึ้นอีกครั้งจากปี ค.ศ.2000 ถึง ค.ศ.2005 หลังจากนั้นอุณหภูมิมีการลดลงถึงปี ค.ศ.2012 สำหรับข้อมูลชุดที่สองศึกษาข้อมูลอุณหภูมิสูงสุดรายเดือนโดยเลือกวันที่มีอุณหภูมิสูงสุดในเดือนนั้น ๆ ข้อมูลรายวันที่สูญหายถูกประมาณด้วยรูปแบบสมการถดถอยเชิงเส้นโดยใช้ข้อมูลสถานีและช่วงเวลาที่ใกล้เคียงกัน การวิเคราะห์ปัจจัยถูกใช้เพื่อช่วยลดมิติของข้อมูลชุดนี้ ข้อจำกัดของการวิเคราะห์ปัจจัยคือ มีบางสถานีไม่สามารถถูกจัดกลุ่มไปยังกลุ่มใดกลุ่มหนึ่งได้อย่างชัดเจน การวิเคราะห์กลุ่มถูกนำมาใช้ในการจัดกลุ่มของค่า factor loading ผลการวิเคราะห์กลุ่มพบว่า สามารถจัดกลุ่มสถานีได้เป็น 4 กลุ่ม แต่ละกลุ่มถูกกำหนดด้วยรูปแบบ quartic trend model with 3nd order time lag และพยากรณ์อุณหภูมิสูงสุดในช่วงเวลาสั้นๆ ผลการศึกษาพบว่า การพยากรณ์อุณหภูมิสูงสุดรายเดือนมีแนวโน้มลดลงในช่วงปี ค.ศ.2013 ถึง ค.ศ. 2015 ทั้ง 7 กลุ่ม ช่วงความเชื่อมั่น 95% ของการพยากรณ์อุณหภุมิสูงสุดรายเดือนอยู่ระหว่าง 27 องศา

เซลเซียส ถึง 44 องศาเซลเซียส ในกลุ่มที่ 1, 26 องศาเซลเซียส ถึง 40 องศาเซลเซียส ในกลุ่มที่ 2, 31 องศาเซลเซียส ถึง 42 องศาเซลเซียส ในกลุ่มที่ 3 และ 26 องศาเซลเซียส ถึง 42 องศาเซลเซียส ในกลุ่มที่ 4

Acknowledgements

I would like to acknowledge my sincere gratitude to my advisor, Professor Dr. Don McNeil, and my co-advisors, Asst. Prof. Dr. Nittaya McNeil and Asst. Prof. Dr. Phattrawan Tongkumchun, for their valuable assistance, encouragement and helpful guidance. Also grateful thanks are given to Asst Prof Dr. Chamnein Choonpradub, Asst. Prof. Dr. Apiradee Lim and their Ph.D. students for helpfulness and support during the past 3 years.

My special thanks go to Dr. Kehui Luo and Professor Dr. Khairil Anwar Notodiputro, who together have kindly advised my thesis and work. Also my thanks are extended to Professor Dr. Robert Orr for his assistance with my academic English writing and suggestions.

I am grateful to Dr. Kehui Luo and Professor Dr. Xia Yingcun for providing a comfortable and enjoyable place to visit in the Department of Statistics, Macquarie University, Australia in 2013 and the Department of Statistics and Applied Probability, National University of Singapore (NUS), Singapore in 2014. Also, I would further like to thank the Department of Statistics, Faculty of Science, Kasetsart University and the Graduate school, Prince of Songkla University for funding this study.

Finally, I wish to thank my parent, my sisters and brother and my colleagues for their support and encouragement.

Wandee Wanishsakpong

Contents

Contents (cont.)

Contents (cont.)

List of Tables

List of Figures

List of Figures (conts.)

List of Figures (conts.)

Page

List of Figures (conts.)

CHAPTER 1

Introduction

1.1 Background and Rationale

Climate change is one of the most important problems facing the world today. Attempts by governments to ameliorate its effect, such as attempting to reduce carbon dioxide emissions are hotly debated. Even scientists do not agree on the extent to which these emissions effect harmful global warming or on the scope, variation and magnitude of global warming itself. Earth surface temperatures have changed over the past 150 years with a slightly higher rate of warming in the 20^{th} century (Jones *et al*., 1999).

Since 1880, the average surface temperature of the earth has increased by about 0.8˚C (NASA, 2010). During the relatively short period of 1925-1944, warming increased by 0.37˚C (Jones *et al*., 1999), and since 1975, warming increased in the range of 0.15˚C - 0.2˚C (NASA, 2010). More recently, in 2010 alone, the annual average earth surface temperature increase was 0.53˚C ± 0.09˚C. This was slightly higher than the annual average temperature increase in 1998 and 2005 which was estimated to be 0.51˚C and 0.52˚C, respectively (WMO, 2011).

In addition, changes in sea surface temperatures increased over the 20^{th} century and continued to rise. For the period of 1910-2013, average sea surface temperatures increased by about 0.13˚F per decade (NOAA, 2014). The warming is greater over land than over the oceans, because water is slower to absorb and release heat (NASA, 2010).

Changes in earth surface and sea surface temperatures are associated with change in sea levels, destruction of ecosystems, shrinkage of mountain glaciers, reduction of ice cover (National Academies, 2008) and altered ocean circulation patterns (Houghton *et al.*, 2001). In addition, climatic variability may provide indicators for several seasonal-related weather events and may also be related to natural disasters, such as floods, droughts, tropical cyclones, rising sea levels and the El Niño Southern Oscillation (Hughes, 2003).

Temperature variability on the earth has been studied in different regions. In the Northern Hemisphere and the Southern Hemisphere based on data collected during 1961-1990, the annual average surface temperature were 14.08˚C and 13.48˚C (Jones *et al*., 1999). In the Arctic, the average temperature has increased by about 0.6˚C since the beginning of the 20th century with maximum temperatures increasing by approximately 1.2˚C since 1945 (Overpeck *et al*., 1997). Temperatures in South-East Asia have also increased, ranging from 0.091˚C to 0.240˚C per decade during the period of 1973-2008 (Chooprateep and McNeil, 2014).

In Australia, average temperatures increased by about 0.32˚C from 1981 to 2005 (Collins *et al*., 2000). Moreover, the mean surface temperatures over the period 1991, 1992, 2003 and 2006 in four Asian cities also increased; 1.8˚C in Bangkok, 2.2˚C in Osaka, 2.5˚C in Seoul and 3.1˚C in Tokyo (Taniguchi *et al*., 2007). In the Pacific Ocean, average temperatures increased by about 3.1˚C in 1976 when compared to average temperatures during the period of 1951-1975 (Hartmann and Wendler, 2005). Sea surface temperatures in the North Atlantic Ocean increased by approximately 0.13˚C per decade during the period 1973-1989, and at least 0.4˚C per decade during 1990-2008 (McNeil and Chooprateep, 2014).

Many scientists and researchers have studied patterns of earth surface temperature change in different regions of the world using various methodologies, including computer simulation models (Johannessen *et al*., 2003), empirical orthogonal functions (Semenov, 2007) and appropriate statistical techniques, such as multiple linear regression (Lean and Rind, 2009), multiple regression with non-Gaussian correlated errors and autoregressive moving average (Hughes *et al*., 2006) models, Pearson correlation analysis (Griffiths *et al*., 2005), linear and quadratic regression (Maiyza *et al*., 2009), principle components analysis (Jones, 1999), linear spline functions, multivariate linear regression models and factor analysis (Chooprateep and McNeil, 2014).

Studies of earth surface temperature changes and the interpretations of these trends are complicated by sensitivity to the time period and geographic region. In this study, earth surface temperature trends in two different geographic regions, the Arctic and Australia, are investigated.

The Arctic is a polar region located at the northern most part of the earth. The study area is above latitude 45 degree north consisting of a vast, ice-covered ocean and surrounded by treeless permafrost. Accordingly, the Arctic region is an important component of the earth's climate system. Arctic air temperatures have risen at almost twice the rate of the global average over the past few decades. The surface waters of the Arctic Ocean have been warming due to declining sea-ice cover, allowing the water to absorb more heat from the sun. Changes in the Arctic climate also affect the rest of the world through increased global warming and rising sea levels (ACIA, 2004).

Australia is surrounded by the Indian and Pacific Oceans. The island-continent is located between latitudes 9˚ and 44˚S and longitudes 112˚ and 154˚E. The largest part of Australia, its inner core, is desert or semi-arid. The south-east and south-west corners have a temperate climate. The northern part of the country has a tropical climate with tropical rainforests, grasslands and deserts. Extreme temperatures have increased consistent with the warming of the climate (IPCC, 2007), the variation of maximum temperature in part of Australia during the 20^{th} century (Nicholls *et al*., 2004) and also warming over recent decades (Murphy and Timbal, 2007).

The primary objective of this study is to examine the patterns of temperature change in the region centered around the middle part of the northern hemisphere of the earth's surface in the Arctic region, including both land and sea surface temperatures, and in the Australian region, covering only land surface temperatures, using a combination of statistical techniques including linear regression, auto-regressive model, time series filtering to reduce auto-correlations between successive monthly and daily values, factor analysis, cluster analysis, quartic trend model combined with the 3^{rd} order time lag and polynomial regression models.

1.2 Research Objective

This primary research objective will be attained through investigation of the trends, patterns and classification of temperature changes in the Arctic and Australia including, application of an appropriate statistical model to forecast temperature change.

1.3 Literature Review

Various studies have assessed temperature change over the Arctic region and in Australia. Temperature change in the Arctic has been a major topic of international review and indigenous observations during the last decade (Krupnik and Jolly, 2012) which parallels research on temperature change in Australia over recent decades (Murphy and Timbal, 2007). Since natural temperature change and pressure variations are complex, many different statistical methods have been used to explain the variability and trends of temperature change and to quantify the size and speed of the rate of change.

Temperature change in the Arctic region and statistical methods

The Arctic climate has been changing rapidly since 1980. The mean seasonal, annual daily maximum and minimum air temperatures and diurnal temperature range in the Arctic over the period 1951-1990 were studied by Przybylak (1996). The decrease in the mean minimum temperature was about twice as great as the rate of increase of the mean maximum temperature. The increased temperatures of the diurnal temperature range occurred in the summer rather than the winter. Linear trends of annual and seasonal maximum and minimum temperatures were calculated. These trends were estimated using the student's t-test. Standard deviations were then used to compare the behavior of the maximum and minimum temperatures. Wang *et al*. (2012) focused on the variability and trends of Arctic climate from satellite observations over the years 1982-2004. The annual mean surface temperature increased at a rate of 0.34˚C per decade and decreased significantly at an annual rate of -0.037˚C in winter. Major cooling occurred around the central and eastern Arctic Ocean. During the warmer seasons from spring to autumn, surface temperatures increased at an annual rate of 0.068˚C, 0.070˚C and 0.045˚C, respectively. The temperature trends were derived using least square regression over the period 1982-2004. The dependent variable was surface temperature, year was the independent variable, and the trend value was the slope of the linear regression line along with the standard deviation (SD) of the slope.

The variability and trends of temperature change are reflected in the Arctic Oscillation (AO) which implies a linkage between global climate change and Arctic climate change (Wang *et al*., 2012). Comprehensive information about temperature changes in and around the Arctic sub-regions were also studied. Anisimov *et al*. (2007) investigated the changes of air temperature in Russia over the period 1900-2004. The trend of annual average temperature increase was about 0.5˚C in the north of European Russia and, 1.4˚C - 1.6˚C in the south of Ural Siberia and the Far East. On average, for the whole of Russia, the trend was an increase of 1.1˚C. The coefficients of correlation between the regional average temperatures, linear regression and time series were analyzed for this study. In Greenland, the annual warming trend during 1919-1932 was 33% greater than the period of 1994-2007, and the recent warming was greatest in western Greenland during autumn and southern Greenland in winter. The periods of these apparent temperature trends were estimated using least square linear regression and time series analysis. Moreover, spatial interpolations of regional bias patterns were estimated by Kriging interpolation procedures (Box *et al*., 2009). In North Carolina, South Finland, and Southeast England, temperatures from 1971 to 1997 have increased by 1.0˚C and 2.1˚C, respectively. However, temperatures in Southeast England were unchanged (Donaldson *et al*., 2002).

In addition, the sea surface temperatures around the Arctic were studied. In the eastern Arctic Ocean, winter season temperatures have increased by approximately 1˚C per decade, whereas in the western Arctic Ocean, temperatures decreased by about 1˚C per decade. During the spring, the eastern Arctic Ocean showed significant warming. This warming was as much as 2˚C during the period of 1979-1997 (Rigor *et al.*, 1999). Hartmann and Wendler (2005) focused on Pacific climate shift in the climatology of Alaska. The pacific decadal oscillation index shifted in 1976. Mean annual and seasonal temperatures were studied and compared during the periods of 1951-1975 and 1977-2001. The temperature increased by 3.1˚C over the period 1977-2001, higher than for the period 1951-1975. The differences in the means were analyzed using a Student's t-test and determined the temperature trend by least square linear regression.

Sea surface temperatures in the North Atlantic Ocean over the period 1973-2008 were analyzed using linear spline functions, auto-regressive models to account for the auto-correlation, and multivariate linear regression models and factor analysis to account for the spatial correlation. From the result of factor analysis, three factors were identified. Temperature changes were found to increase by approximately 0.13˚C per decade during

1973-1989 in these three factors, but during 1990-2008, the temperatures increased by at least 0.4˚C per decade in two of the three identified factors (McNeil and Chooprateep, 2014).

Temperature changes in Australia and statistical methods

The average temperature in Australia has increased by approximately 0.8˚C since 1910 (Collins, 2000). From 1910 until 1990, trends in temperature change of the average maximum and minimum temperature were also studied by Torok and Nicholls (1996) who found that the occurrence of changes and trends were 0.54˚C per 100 years and 0.98˚C per 100 years respectively. The trends were investigated employing a linear regression model. The strongest trend was found in the middle of the century. Correlation analysis was used to assess the accuracy of the dataset between the all-Australia mean maximum and minimum temperature and the all-Australia annual rainfall average. The multiple correlations were 0.78. For the period of 1910 -2002, Karoly and Braganza (2004) investigated the variations of the Australian-average mean temperature and diurnal temperature range. Temperature variability was measured by the standard deviations of the temperature. Simple linear detrending was used to remove any possible anthropogenic signal in the observed indices. The research findings were that the observed temperature variations may still include variation of natural or anthropogenic external forcing in time scales less than a century.

Murphy and Timbal (2007) investigated the temperature trends during the period 1910-2006. They found that mean temperatures reached 0.09˚C per decade while temperatures have been rising by 0.19˚C per decade for 1970-2006 by using a linear regression model as well. Moreover, part of the period 1910 to 2008 also was studied. For example, most parts of Australia increased in temperature by 0.1˚C - 0.2˚C per decade since 1951. This warming has been greatest in Queensland and the southern half of Western Australia (Suppiah *et al*., 2001) and during 1957 to 1996, the trends in annual frequency of warming events have generally increased and the number of cool extremes has decreased. Time series analysis of annual index values was investigated for each temperature record. Trends in the series were calculated for each station using linear regression (Collins *et al*., 2000).

Many techniques were applied to study temperature change and prediction. For instance, Jones (1999) used principal component analysis to investigate the characteristics and mechanisms of Australian temperature variability for the period 1950 to 1994. Approximately 80 percent of the temporal variance was explained by five independent models. The analysis showed that there were warming trends across much of Australia, but also revealed that about 80 to 90 percent of the temporal variance was unrelated to this trend. Grainger and Frederiksen (2008) focused on the potential predictability of Australian surface maximum and minimum temperature using monthly data from December 1950 to November 2000. They found that monthly mean variance within a season can be modeled by a linear relationship, and inter-monthly correlations can be assumed to be stationary. Potential predictability was estimated by removing the intraseasonal variance from the total interannual variance. In addition, this study indicated that the intraseasonal variance of Australian surface maximum and minimum temperature estimated using the stationary variance assumption and the linear assumption showed qualitatively and quantitatively similar patterns of distribution. The results showed that the four seasons for northern Australia have high potential predictability.

1.4 Conceptual Framework

The conceptual framework for analyzing data in the Arctic and Australia is shown in Figure 1.1

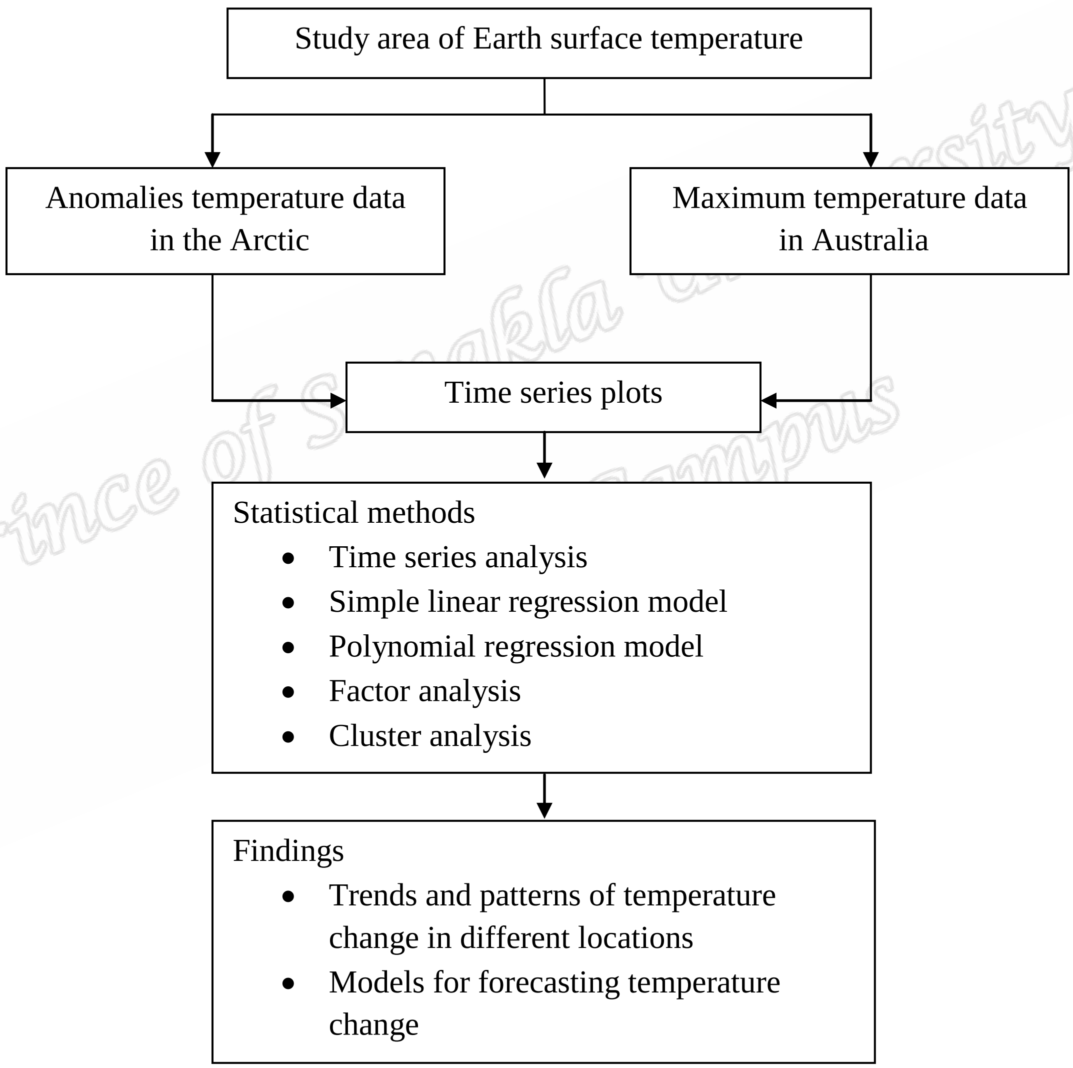


Figure 1.1 Conceptual Frameworks

From Figure 1.1, the first data set is monthly anomalies temperature data in the Arctic during 1973 - 2008 and the second data set is maximum temperature data in Australia during 1970 - 2012. Patterns in the data were derived by plotting time series. Similar main statistical methods were employed to analyze both data sets. Time series analysis was used to describe the correlation of time and variation of data. A simple linear regression model and a polynomial regression model were used to investigate the trends, patterns and rate of temperature change. Factor analysis and cluster analysis were used to classify similar patterns into groups. This method reduced the correlation between variables (study area). Trends and patterns of temperature change were estimated for the different locations as well as for assessing the ability of models to forecast temperature change for short time periods.

1.5 Organization of thesis

In this thesis, the five chapters are organized as follows:

Chapter1 provides an introduction to the background and rationale of temperature change of the Earth's surface temperatures, the objectives of the study, the literature review and the conceptual framework. Chapter 2 provides a description of all methodologies used for this study. Chapter 3 presents the analysis of anomalies temperature data in the Arctic. Chapter 4 presents the analysis of maximum temperature data in Australia. Chapter 5 concludes and discusses the research findings. Moreover, Chapter 5 gives some suggestions for further research.

CHAPTER 2

Methodology

This chapter consists of a discussion of data sources, data management, study diagrams and statistical methods. These methods are comprised of the seasonally adjusted method, linear regression analysis, times series analysis, auto correlation, autoregressive process and filtering, factor analysis, cluster analysis and polynomial regression modeling. All data analyses and graphical display were carried out using R program (R Development Core Team, 2009).

2.1 Data Source and Data Management

The data were obtained from two different data sources of temperature change in the Arctic region and Australia.

Arctic region

Monthly average temperature anomalies data were used to study temperature trends and patterns. The data were downloaded as the file HadCRUT3 from the Climate Research unit at University of East Anglia, UK (CRU, 2009), (http://www.cru.uea.ac.uk/). Since January 2006, HadCRUT3 has produced a new dataset version which combined land and marine anomalies from temperature anomalies for 5˚ by 5˚ latitude- longitude grid-boxes on the Earth's surface. The dataset was collected from 4,349 weather stations for the period of 1850-2008. The temperature anomalies data are raw monthly temperatures subtracted from average temperatures from 1961 to 1990. Each grid-box value is the mean of all available station anomaly values, except that station outliers in excess of five standard deviation are omitted (Brohan et al,, 2006). Data from observation points above latitude 45˚N from 1973 to 2008 was selected in this study. The research locations are defined by 648 grid-boxes of 5˚ by 5˚ dimension. These grid-boxes cover the Arctic Ocean, the northern areas of the Atlantic and Pacific Oceans, the North American, Asian and European Continents. The data included 432 monthly average temperature anomalies for 5˚ by 5˚ latitude-longitude grid-boxes. Data were missing in a number of those grid-boxes, particularly in the polar zones. Therefore, the 648 grid-boxes were combined into 69 sub-regions following a design in the shape of an igloo. This design visually illustrates correlation more effectively than from the perspective of pairs of adjoining sub-regions. The 69 sub-regions included 1 sub-region encompassing up to 5˚ below the North Pole (85˚N-90˚N), 8 in the 75˚N-85˚N band, 12 in the 65˚N-75˚N band, 24 in the 55˚N-65˚N band and another 24 in the 45˚N-55˚N band. However, a number of the sub-regions had still missing data, as shown in brackets in Figure 2.1.

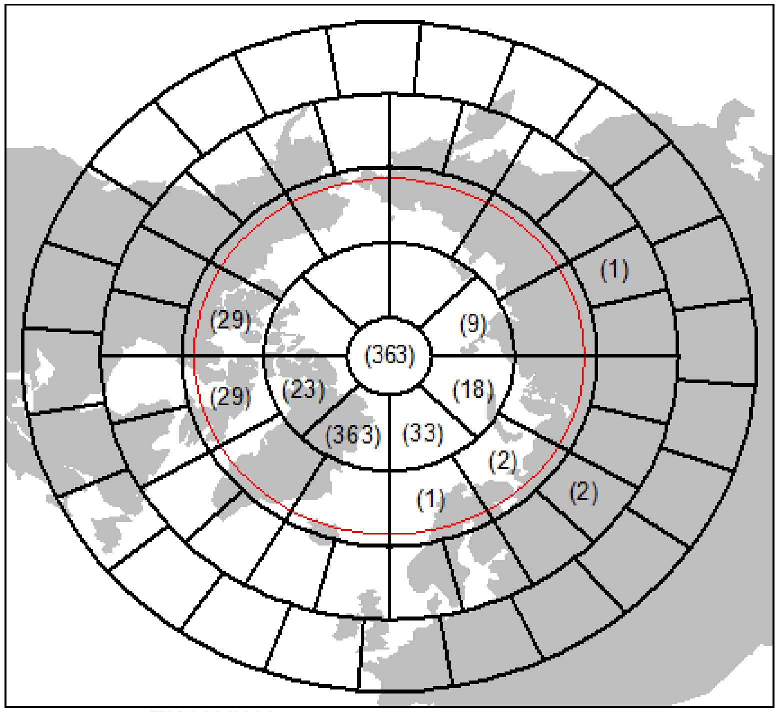


Figure 2.1: Map of the 69 sub-regions above latitude 45˚N. Numbers in brackets are the number of months with missing data for 1973-2008.

Australia

Daily maximum temperature data from 1970 to 2012 were downloaded from the website of the Australian Bureau of Meteorology (BOM, 2013), (http://www.bom.gov.au). The Australia BOM provided daily maximum temperature data for over 700 weather stations. The maximum temperature data recorded in some stations started in different time periods. Some stations also had incomplete records due to weather phenomena and changing reporting requirements (BOM, 2013). A sample of 85 stations was randomly selected. These stations were distributed across Australia. These selections of the stations were in line with Australia's Reference Climate Station Network (RCS) and the Australia Climate Observations Reference Network Surface Air Temperature (ARORN-SAT).

These stations had approximately 80% complete data from 1970 to 2012, as shown in Figure 2.2. In each station, daily maximum temperature data were collected. Observations on February 29 were omitted to have equal numbers of observations for each year, totaling 15,695 observations including missing data over the 43 year period. For this data set, the data were managed to form two new data sets. One data set, for the maximum temperatures over consecutive 5-day periods, was used for analysis to reduce the number of observations to 73 periods in each year, as well as reducing the correlation between successive observations. This data set provided 3,139 observations over the 43-year period and another one for the monthly maximum temperatures which were defined as the highest daily temperature in a particular month. Therefore, the second data set provided 516 monthly maximum temperature observations over the 43 year period.

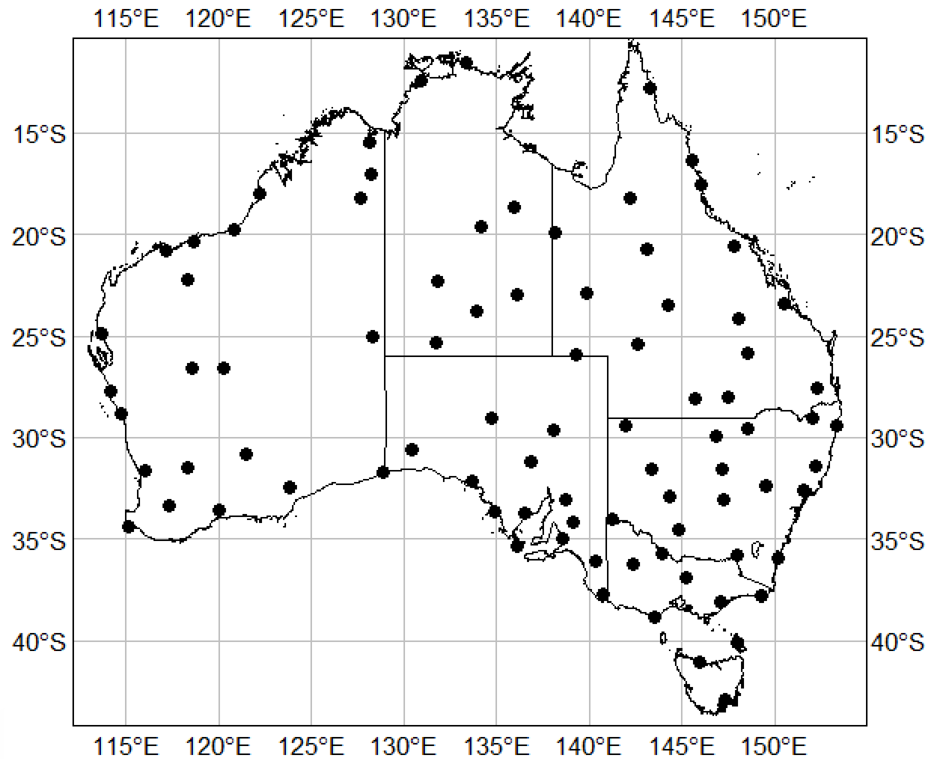


Figure 2.2: Station locations where daily maximum temperature data were collected. The state boundaries in Australia are also given.

2.2 variables

The dependent variables are the surface temperatures which includes both land and sea surface temperature and the independent variable are units of time (Monthly, daily and period of year).

2.3 Study diagrams

The study diagrams for the Arctic region and Australia are shown as follows:

Arctic region

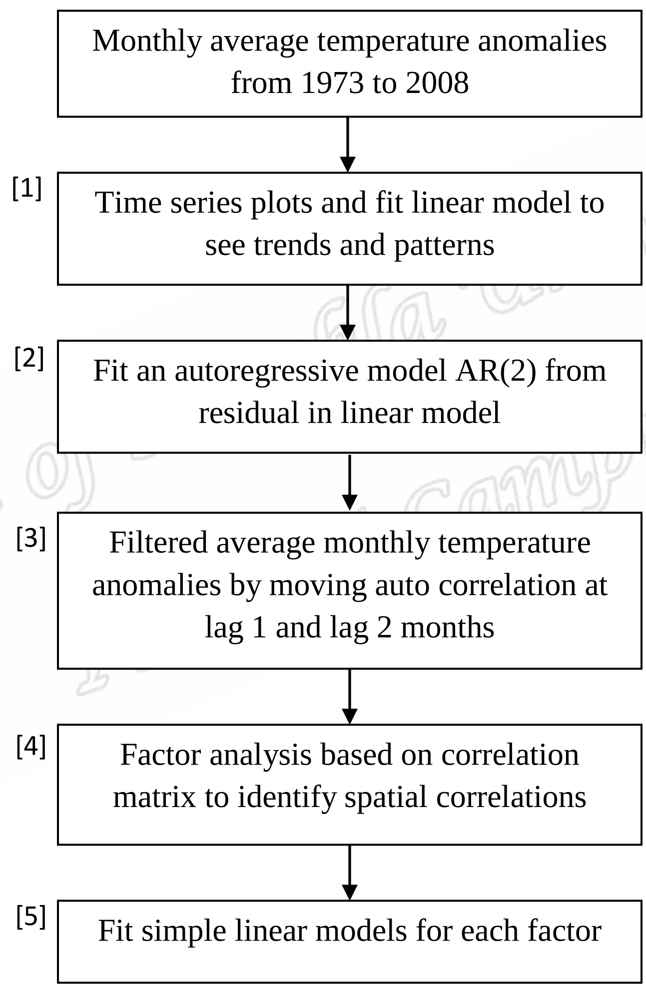


Figure 2.3: The diagram of data analysis in Arctic region

Figure 2.3 shows the steps of analysis for monthly average temperature anomalies data from 1973 to 2008. In the first step, [1] the data were plotted and fitted to a linear regression model to see the temperature trends and patterns. [2] Since the data had time correlates, an auto-regressive model (AR2) was fitted to reduce these correlations. [3] After that, the data were filtered by moving auto-correlation at lag 1 and lag 2 months. [4] The filtered data were used for factor analysis to identify spatial correlation. [5] Finally, a linear regression model was fitted for each factor. The coefficients of the linear regression model were used to explain the rate of temperature change with a confidence interval of the temperature changes in each factor.

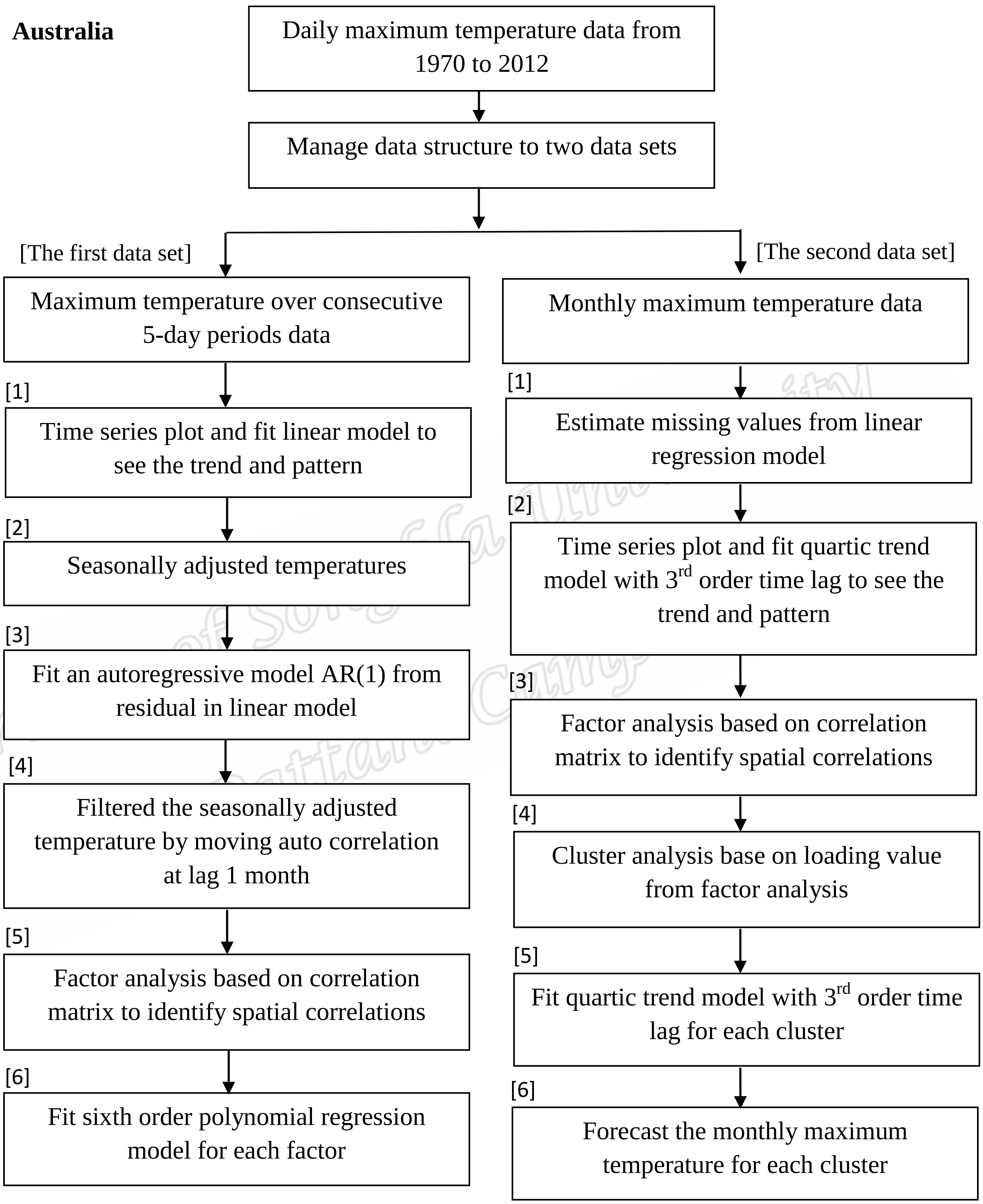


Figure 2.4: The diagram of data analysis in Australia

The daily maximum temperatures in Australia from 1970 to 2012 were managed to form two data sets as shown in diagram of figure 2.4. The first data set is the maximum temperature over consecutive 5-day periods data. [1] Start, the data were plotted and fitted by linear regression model to see the trends and patterns. [2] Since the data showed the patterns of seasonal variation, the data were adjusted seasonally by subtracting the average of the maximum temperature in each period and then adding them back to the overall mean for all 85 stations. [3] Next, an auto-regressive model AR(1) was fitted to reduce correlation between the data and time periods. [4] Then, the data were filtered by removing auto-correlation at lag 1 month. [5] After that, factor analysis was applied to the filtered temperatures to identify spatial correlation and [6] fitted to a sixth order polynomial regression model to account for trends in the data as displayed by a curvilinear pattern in each factor.

For the second data set, the monthly maximum temperature data were investigated for the presence of trends and patterns of temperature change. Since there were missing data in several stations, [1] a linear regression model was fitted to estimate these missing values by assuming that the maximum monthly temperature at a particular month and station could be predicted using the previous and next month temperatures as well as the first and second nearest station temperatures at the pre-defined month. After interpolating the missing data, [2] the data were plotted and fitted to the quartic trend model with 3rd order time lag to see trends and patterns and [3] to identify spatial correlations by factor analysis. From the results of factor analysis, there were several stations having similar factor loadings. [4] Cluster analysis was used to regroup the 85 stations based on their factor loadings. The distance or dissimilarity between groups of stations was defined using the completed linkage and based on the Euclidian distance. Since, the data showed the curve pattern in each cluster. [5] The quartic trend model with 3rd order time lag was fitted reasonably well [6] and this model was used to forecast for short time periods in each cluster.

2.4 Statistical methods

Seasonally adjusted method

The maximum temperatures over consecutive 5-days periods were seasonally adjusted to remove variation in temperature. The adjustment was carried out using the backward difference operator Δ as well as the backshift operator B (Montgomery *et al*., 2008)

If x_{kj} is the maximum temperature over consecutive 5-day periods in station k for period j and the backward difference is $\Delta^d = (1-B)^d$ whereas $B^d x_{kj} = x_{k(j-d)}$, then the lag-d seasonal operator is defined as $\Delta^d x_{kj} = (1-B)^d x_{kj} = x_{kj} - x_{k(j-d)}$

Hence, y_{kj} denotes the maximum temperature over consecutive 5-day periods in station k for period j. From the above equation by choosing d = 73 as follows,

$$y_{kj} = \Delta^{73} x_{kj} = (1-B)^{73} x_{kj} = x_{kj} - x_{k(j-73)} \quad (2.1)$$

Time series analysis

Auto correlation

A time series is a set of observations which each one is recorded at a specific time (Brockwell and Davis, 2002). The correlation between members of a series and numbers arranged in time is called auto correlation or serial correlation. An important guide to the persistence in a time series is provided by the series of quantities called sample auto correlation coefficients, which measure the correlation between observations separated by l time

The auto correlation coefficient at lag l, is given by

$$r_l = \frac{\sum_{t=1}^{n-l}(x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2} \quad (2.2)$$

where $\bar{x}$ is the mean of *n*-*l* observations.

The auto correlation coefficient at lag l is interpreted graphically by auto correlation function (ACF) plot, sometimes called a correlogram where r_l is plotted against the lag l (Cryer, 1986; Chatfield, 1996).

If a time series is completely random with a large *n*, auto correlation should be near zero for any and all time-lag separations. In fact, for independent and identically distributed (iid) noise with finite variances, auto correlation is approximately iid N(0,1/ *n*) for *n* large. Therefore, approximately 95% of auto correlation should fall within the bounds of

$\pm 1.96/\sqrt{n}$. The dotted lines on the auto correlation function graph are the bounds $\pm 1.96/\sqrt{n}$ (Brockwell and Davis, 2002) as shown in Figure 2.5.

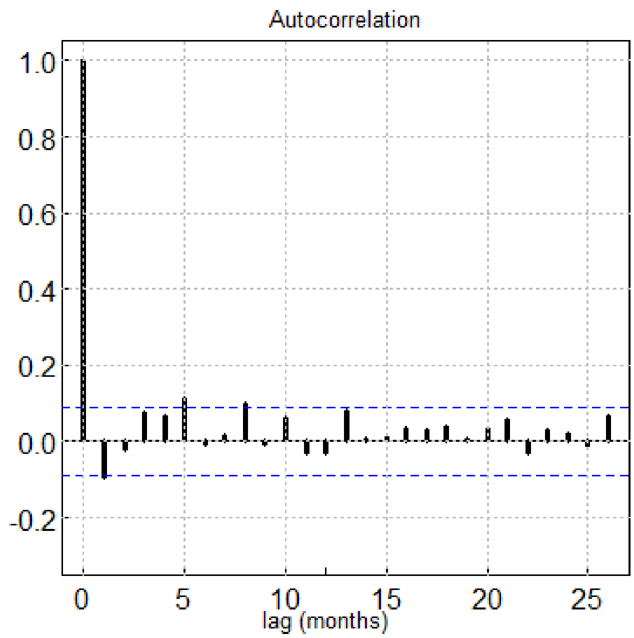


Figure 2.5: Example of auto correlation plot. The dotted lines indicate $\pm 1.96/\sqrt{n}$.

Autoregressive process

Autoregressive process AR is applied to account for auto correlation among residuals from the fitted linear model (Venables and Ripley, 2002). AR(p) is an autoregressive model of the pth order, and AR(p) satisfies the equation

$$e_t = \emptyset_1 e_{t-1} + \emptyset_2 e_{t-2} + \cdots + \emptyset_p e_{t-p} + a_t \tag{2.3}$$

where e_t is the difference between temperature (y_t) and the corresponding fitted value $(\hat{y}_t)$.Thus, $e_t = y_t - \hat{y}_t$ is the residual value at the time t, $t - i$ is the order of the past time for i = 1, 2, …, p, $\emptyset_1, \emptyset_2, \ldots, \emptyset_p$ are the parameters of model, and a_t is the value that is not explained by the past values. The series is assumed to be stationary, and a_t is independent of $e_{t-1}, e_{t-2}, \dots, e_{t-p}$ (Cryer, 1986).

Filtering

Filtered temperature data were obtained using the equation (Chatfield, 1996) as follows:

$$z_t = e_t - (\emptyset_1 e_{t-1} + \emptyset_2 e_{t-2} + \cdots + \emptyset_p e_{t-p}) \qquad (2.4)$$

where z_t is the filtered temperature data at the time t, e_t is the difference between temperature (y_t) and the corresponding fitted value $(\hat{y}_t)$. Thus $e_t = y_t - \hat{y}_t$ is the residual value at the time t, $t - i$ is the order of the past time for i = 1, 2, …, p, and $\emptyset_1, \emptyset_2, \dots, \emptyset_p$ are the average coefficients of the fitted autoregressive model AR(p).

Regression analysis

Linear regression is an approach used to model the relationship between a dependent variable and one or more explanatory variables. In this study, a simple linear regression model was fitted to investigate the temperature trends, and a multiple linear regression model was fitted to estimate the missing data for temperatures in Australia.

The simple linear regression model takes the following form:

$$Y = \beta_0 + \beta_1 t + \in \qquad (2.5)$$

where Y denotes the temperature, β_0 is intercept, β_1 corresponds to the temperature change, t is time and $\in$ is the error assumed to follow a normal distribution with constant variance.

The multiple linear regression model takes the following form:

$$Y_{tk} = \beta_0 + \beta_1 Y_{t,k-} + \beta_2 Y_{t,k+} + \beta_3 Y_{t-1,k} + \beta_4 Y_{t+1,k} + \epsilon_{tk} \qquad (2.6)$$

for t = 1, 2, … , 516 and k = 1, 2, …, 85, where Y_{tk} denotes the maximum temperature at month t of station k, β_j are regression coefficients, j= 0, 1, 2, 3, 4, $Y_{t,k-}$ is the maximum temperature at month t for the nearest station from station k, $Y_{t,k+}$ is the maximum temperature at month t for the second nearest station from station k, $Y_{t-1,k}$ is the maximum temperature at month $t-1$ for station k, $Y_{t+1,k}$ is the maximum temperature at month $t+1$ for station k and ϵ_{tk} is the error assumed to follow a normal distribution with constant variance at month t of station k (Kedem and Fokianos, 2002).

Polynomial regression model

Polynomial regression is a form of linear regression for quantitative independent variables and the dependent variable. They are among the most frequently used curvilinear response models (Neter *et al*., 1996). The general polynomial regression model, as an n^{th} degree polynomial takes the following form

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \cdots + \beta_n t^n + \in_t \qquad (2.7)$$

where Y_t is temperature at time t, β_0 represents the mean response of Y_t when t = 0, $\beta_1, \beta_{2,\ldots,}\beta_n$ are effect coefficients and ϵ_t is the error term.

A quartic trend model combined with the 3rd order time lag was fitted to account for the trend and to predict the maximum monthly temperature. The model takes the following form

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \delta_1 Y_{t-1} + \delta_2 Y_{t-2} + \delta_3 Y_{t-3} + \epsilon_t \quad (2.8)$$

where Y_t temperature at time t, and $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \delta_1, \delta_2, \delta_3$ are the parameters of the model, whereas ϵ_t is the error term.

Factor analysis

Factor analysis is another variable-directed technique concerned with explaining the covariance and/or correlation structure among the measured variables and determining the correlation between a large set of variables in terms of a small number of underlying factors (Johnson, 1998).

The factor analysis model with *m* factors, denoted by $f_1, f_2, \dots, f_m$, takes the form

$$y_t = \mu + \sum_{q=1}^{m} \lambda_q f_q \quad (2.9)$$

where y_t is the temperature for time t, μ is the average temperature, λ_q is the factor loadings on the q^{th} factor and f_q is the q^{th} common factor.

The factors were identified using factor loadings. The cut-off value for the factor loading was 0.3 as the criteria for classifying the region in each factor (Hair *et al*., 1998). The loadings (usually between -1 and 1) were controlled by rotating the factors. Most rotation procedures require many factors loadings as close as possible to zero and maximize as many of the others as possible. Moreover, it would be highly desirable that the factors are completely independent. However, this is not critical if response variables are not located heavily on more than one factor (Johnson, 1998).

The Varimax rotation method and Promax rotation method were used in this study. The Varimax rotation method is an orthogonal rotation algorithm with procedures that keep the factors uncorrelated whenever the method starts with a set of uncorrelated factors. The orthogonal rotations of factor loading matrices do not affect the communalities of the response variables and the specific variance of the variables. The Promax rotation method is an alternative non-orthogonal (Oblique) rotation method. In practice, Oblique rotations do not produce new factors that remain uncorrelated, which is a contradiction of the initial factor analysis assumption. That is, the oblique rotation allows the factors to correlate loadings to form a conceptually clearer picture (Johnson, 1998).

Cluster analysis

Cluster analysis is a technique that produces classification from data that are initially unclassified. Cluster analysis provides the ability to measure the similarity or dissimilarity between two individual observations and subsequently the similarity or dissimilarity between two clusters of observations (Johnson, 1998).

One simple measure of dissimilarity is standard Euclidean distance. This is the distance between two observations when plotting the two observations in the p-dimensional sample space and measuring the distance between them using a formula. The formula that calculates Euclidean distance takes the form of (Johnson, 1998).

$$d_{rs} = [(x_r - x_s)'(x_r - x_s)]^{1/2} \quad (2.10)$$

where x_r and x_s are two points of dissimilarity.

Classifying a set of objects mean establishing or constructing clusters or a hierarchy. There are two basic methods to distinguish whether the objects are either hierarchical or nonhierarchical.

Hierarchical clustering method was used in this study. The observed data points were grouped into clusters in a nested sequence of clustering methods in the complete linkage of two groups between individuals A and B. The complete linkage between A and B is the greatest distance between an element in A and an element in B (Husson, 2011). Moreover, a hierarchical tree diagram is a means of providing assistance in deciding when to stop the clustering process. It contains branches connecting data points and shows the order in which the points are assigned to clusters. The lengths of its branches are proportional to the distance between points and clusters when points and cluster are combined (Johnson, 1998).

CHAPTER 3

Analysis of temperature anomalies data in the Arctic

This chapter reports the analysis of monthly average temperature anomalies data in order to investigate the trends and patterns of temperature changes above latitude 45 degrees North from 1973 to 2008.

3.1 Summary of the temperature data

Temperature changes above latitude 45 degrees North were investigated from 1973 to 2008 since this period has the most complete data and also corresponds with global temperature anomalies which have increased during this period (CRU, 2009). The results also appeared in Wanishsakpong *et al*. (2014).

The study area consisted of 648 grid-boxes which were combined into 69 sub-regions covering the Arctic Ocean, northern areas of the Atlantic and Pacific Oceans and the North American, Asian and European Continents. The numbers on the map in Figure 3.1 indicate the locations of these sub-regions. In each sub-region, there were 432 observations with some of the sub-regions having missing values. Of the 69 sub-regions, 57 sub-regions had complete data, 10 sub-regions had less than 10% missing data and 2 sub-regions had more than 80% missing data.

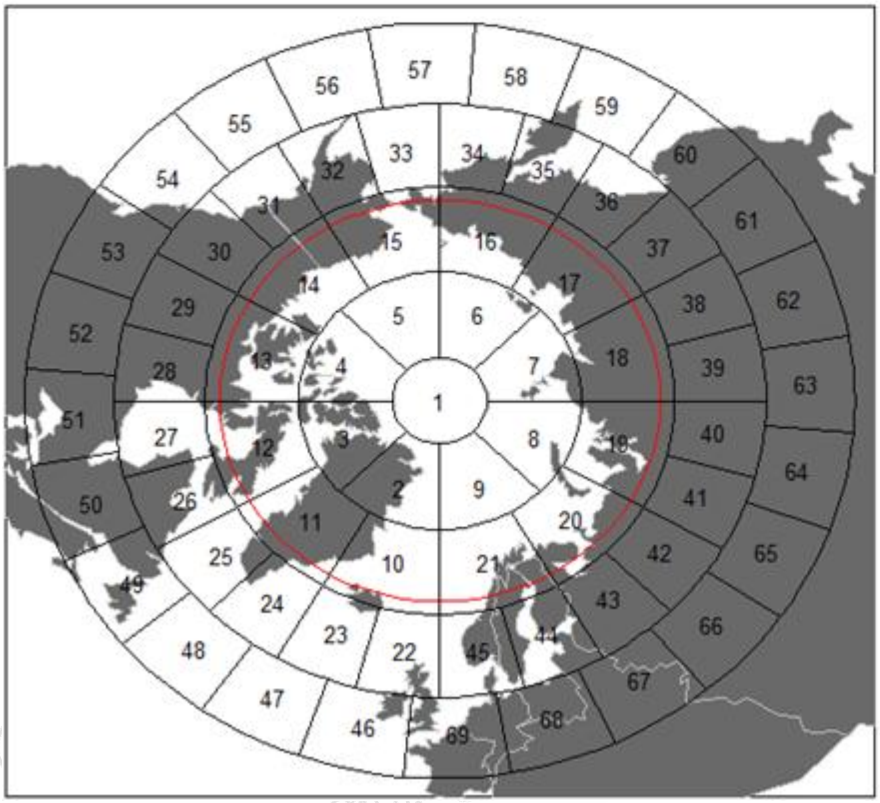


Figure 3.1: Map of the 69 sub-regions above latitude 45˚N

Summaries of monthly average temperature anomalies data in each of the 69 sub-regions are shown in Table 3.1.

Sub-region	Latitude and longitude	n	Mean	Min	Max
1	North pole	69	2.00	0.01	6.80
2	75-85˚N – 180-135˚E	69	1.90	0.07	6.90
3	75-85˚N – 135-90˚W	409	0.82	-6.00	9.40
4	75-85˚N – 90-45˚W	432	0.61	-9.70	8.90
5	75-85˚N – 45-0˚W	432	0.43	-6.70	8.00
6	75-85˚N – 0-45˚E	432	0.61	-6.12	9.40
7	75-85˚N – 45-90˚E	423	0.76	-9.40	13.90
8	75-85˚N – 90-135˚E	414	0.61	-9.70	12.40
9	75-85˚N – 135-180˚E	399	0.61	-6.40	6.80
10	65-75˚N – 180-150˚W	432	0.67	-7.52	12.14
11	65-75˚N – 150-120˚W	432	1.07	-6.98	12.67
12	65-75˚N – 120-90˚W	403	0.64	-7.02	8.84
13	65-75˚N – 90-60˚W	403	0.32	-8.04	8.62
14	65-75˚N – 60-30˚W	432	0.33	-9.05	5.23
15	65-75˚N – 30-0˚W	432	0.41	-1.58	3.67
16	65-75˚N – 0-30˚E	432	0.28	-1.32	2.51
17	65-75˚N – 30-60˚E	432	0.49	-3.07	3.30
18	65-75˚N – 60-90˚E	432	0.91	-7.80	10.30
19	65-75˚N – 90-120˚E	432	0.56	-11.00	12.28
20	65-75˚N – 120-150˚E	430	0.53	-5.20	8.87
21	65-75˚N – 150-180˚E	431	0.59	-6.33	8.20
22	55-65˚N – 180-165˚W	432	0.32	-3.13	3.17
23	55-65˚N – 165-150˚W	432	0.40	-5.18	6.07
24	55-65˚N – 150-135˚W	432	0.21	-4.23	5.18
25	55-65˚N – 135-120˚W	432	0.45	-9.09	9.61
26	55-65˚N – 120-105˚W	432	0.67	-11.30	10.98
27	55-65˚N – 105-90˚W	432	0.67	-8.17	8.00
28	55-65°N – 90-75˚W	432	0.68	-6.66	7.96
29	55-65˚N – 75-60˚W	432	0.25	-7.05	8.70
30	55-65˚N – 60-45˚W	432	0.009	-2.56	2.51
31	55-65˚N – 45-30˚W	432	0.20	-1.62	2.67
32	55-65˚N – 30-15˚W	432	0.09	-0.98	1.74
33	55-65˚N – 15-0˚W	432	0.11	-1.10	1.82
34	55-65˚N – 0-15˚E	432	0.29	-2.10	3.30
35	55-65˚N – 15-30˚E	432	0.29	-3.52	3.90

Table 3.1(a): Data summaries of monthly average temperature anomalies in each sub-region

Sub-region	Latitude and longitude	n	Mean	Min	Max
36	55-65˚N – 30-45˚E	432	0.57	-9.57	9.76
37	55-65˚N – 45-60˚E	432	0.59	-10.95	10.00
38	55-65˚N – 60-75˚E	431	0.62	-8.63	9.82
39	55-65˚N – 75-90˚E	432	0.71	-12.40	11.80
40	55-65˚N – 90-105˚E	432	0.71	-9.36	10.47
41	55-65˚N – 105-120˚E	432	0.64	-10.25	10.03
42	55-65˚N – 120-135˚E	430	0.70	-5.90	9.49
43	55-65˚N – 135-150˚E	432	0.43	-3.50	5.51
44	55-65˚N – 150-165˚E	432	0.23	-3.06	3.41
45	55-65˚N – 165-180˚E	432	0.11	-4.04	4.58
46	45-55˚N – 175-160˚W	432	0.04	-1.62	1.37
47	45-55˚N – 160-145˚W	432	0.11	-1.73	2.13
48	45-55˚N – 145-130˚W	432	0.21	-1.69	2.20
49	45-55˚N – 130-115˚W	432	0.17	-5.47	3.91
50	45-55˚N – 115-100˚W	432	0.41	-8.98	9.63
51	45-55˚N – 100-85˚W	432	0.39	-5.55	8.57
52	45-55˚N – 85-70˚W	432	0.09	-6.21	6.63
53	45-55˚N – 70-55˚W	432	0.20	-3.45	3.20
54	45-55˚N – 55-40˚W	432	0.03	-2.25	2.12
55	45-55˚N – 40-25˚W	432	0.21	-1.59	2.34
56	45-55˚N – 25-10˚W	432	0.24	-1.78	2.26
57	45-55˚N – 10-5˚E	432	0.35	-1.83	2.38
58	45-55˚N – 5-20˚W	432	0.51	-3.03	4.47
59	45-55˚N – 20-35˚E	432	0.35	-7.26	6.86
60	45-55˚N – 35-50˚E	432	0.32	-8.83	9.90
61	45-55˚N – 50-65˚E	432	0.44	-9.59	10.42
62	45-55˚N – 65-80˚E	432	0.44	-9.42	9.27
63	45-55˚N – 80-95˚E	432	0.63	-7.41	7.53
64	45-55˚N – 95-110˚W	432	0.68	-6.64	7.84
65	45-55˚N – 110-125˚E	432	0.70	-4.39	7.35
66	45-55˚N – 125-140˚E	432	0.45	-3.90	5.50
67	45-55˚N – 140-155˚E	432	0.21	-1.41	2.38
68	45-55˚N – 155-170˚E	432	0.12	-1.37	2.13
69	45-55˚N – 170-175˚E	432	0.09	-1.67	1.73

Table 3.1(b): Data summaries of monthly average temperature anomalies in each sub-region

Table 3.1 presents the numerical summaries. There were 432 observations in most of the sub-regions except for sub-regions 1, 2, 3, 8, 9, 12, 13, 20, 21, 38 and 42. The range of monthly average temperature anomalies varied from -12.4˚C in sub-region 39 to 13.9˚C in sub-region 7.

A time series plot of the average temperature anomalies for all 69 sub-regions is presented to show overall temperature patterns. The dots shown in Figure 3.2, suggest that there is a statistical relationship between temperature and time, and the tentative assumption of the straight line model (equation 2.5, page 25) appears to be reasonable. Therefore, a simple linear regression model (equation 2.5, page 25) was fitted with these data in order to investigate temperature changes in all 69 sub-regions from 1973 to 2008.

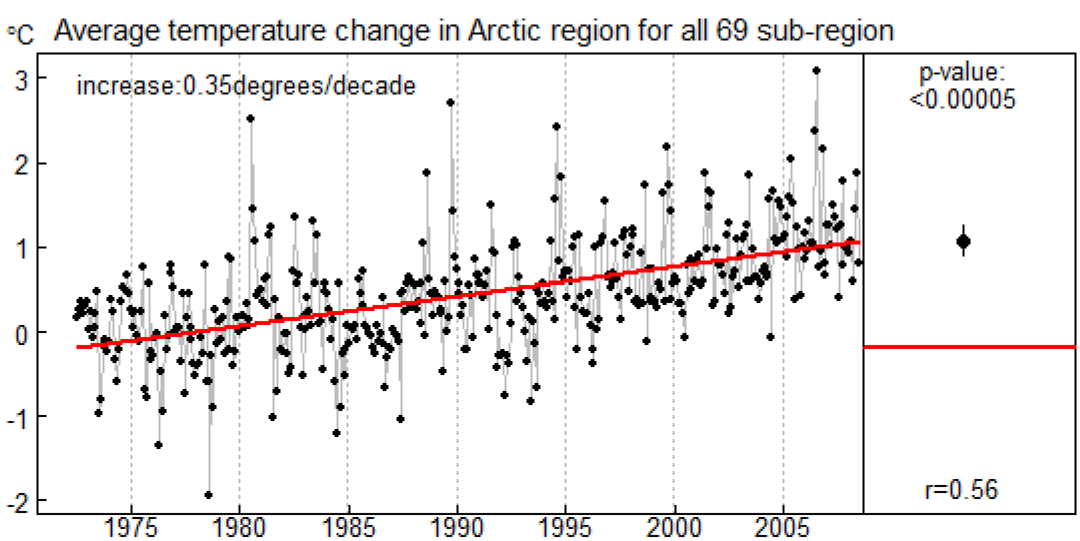


Figure 3.2: Average temperature anomalies from 1973 to 2008 (432 months). The correlation is denoted by r.

In Figure 3.2, temperatures show a significant rate of increased (slope) by 0.35˚C per decade over this period (P-value < 0.05). A 95% confidence interval (CI) on the slope is 0.30˚C - 0.39˚C. The correlation between temperature and time is r = 0.56. The horizontal line on the right panel shows the average temperature predicted by the model at the first time point (January 1973).

A separate linear regression model (equation 2.5, page 25) was fitted for each of the 67 sub-regions to explain temperature change per decade, except in sub-regions 1 (North Pole) and sub-region 2 (75-85˚N – 180-135˚E) that do not have enough data. The results of temperature change for each of the 67 sub-regions are shown in Appendix 1. Moreover, a 95% confidence interval (CI) of temperature changes per decade for each of the 67 sub-regions was estimated from the simple linear regression model (equation 2.5, page 25) to confirm the results shown in Figure 3.4.

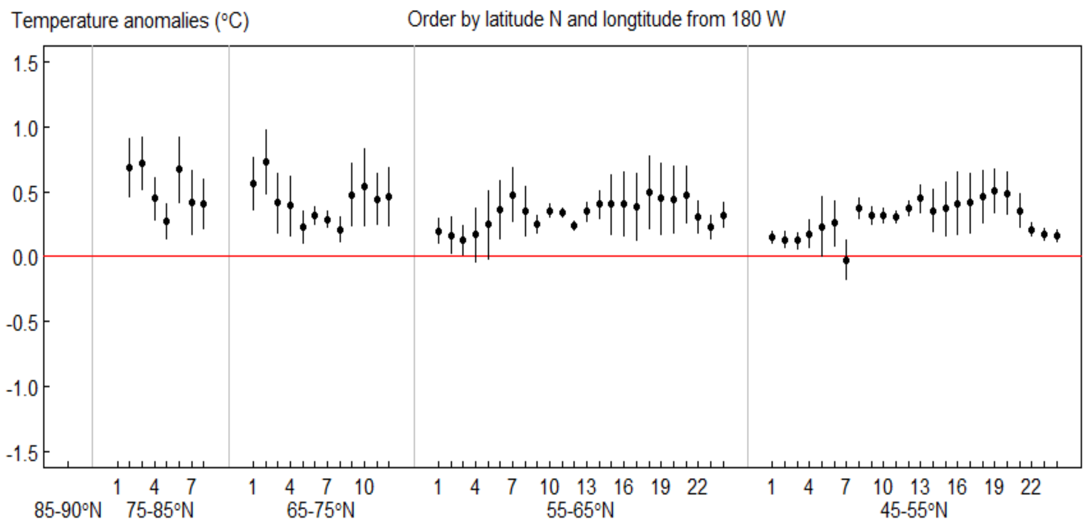


Figure 3.3: 95% confidence interval of monthly temperature anomalies changes ordered by latitude N and longitude from 180 W. The horizontal axis shows sub-regions in each group of latitudes.

Figure 3.3 shows 95% confidence intervals (CI) of temperature change per decade ordered by latitude N and longitude from 180 W. The 95% confidence intervals covering zero imply no significant temperature change. It is clear that there are 64 sub-regions in which temperatures have increased. There is no evidence of change in 3 sub-regions located at latitude 55˚N-65˚N longitude 135˚W-120˚W (p-value =0.116), latitude 55˚N - 65˚N longitude 120˚W-105˚W (p-value =0.065) and latitude 45˚N-55˚N longitude 85˚W-70˚W (p-value =0.764). The temperature changes in the sub-regions are displayed on the map in Figure 3.4

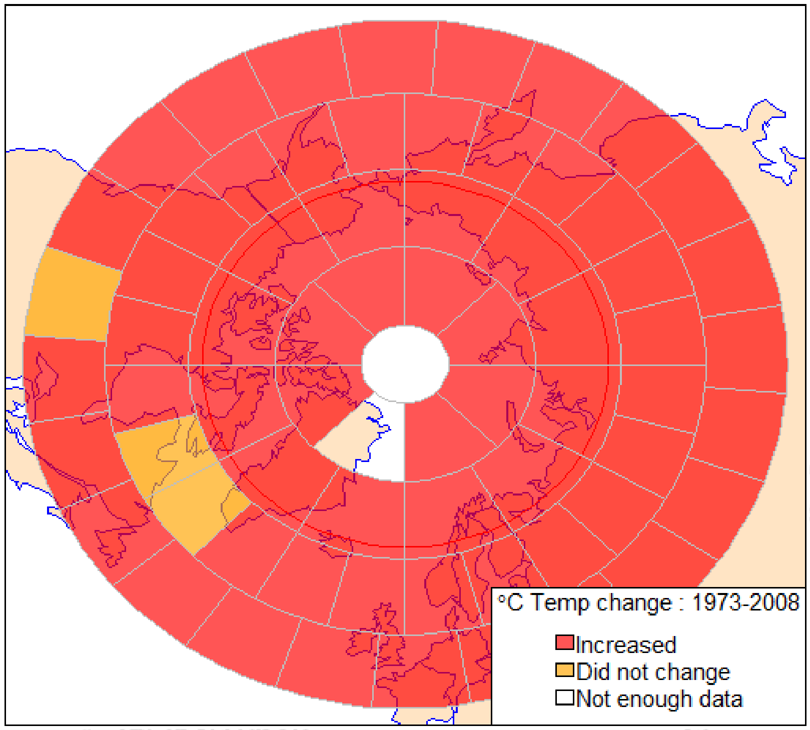


Figure 3.4: Map of temperature changes of the 69 sub-regions from 1973 to 2008.

Figure 3.4, red represents a significant increase (p-value < 0.05), orange no evidence of change and white insufficient data in the sub-regions which were omitted from further analysis.

3.2 Time series analysis with autoregressive process and filtering

A common assumption in time series data is that the data are stationary. A stationary process has an auto correlation structure that does not change over time. Auto correlation of the residuals from a fitted simple linear model (equation 2.5, page 25) shows the correlations separated by time in Figure 3.5.

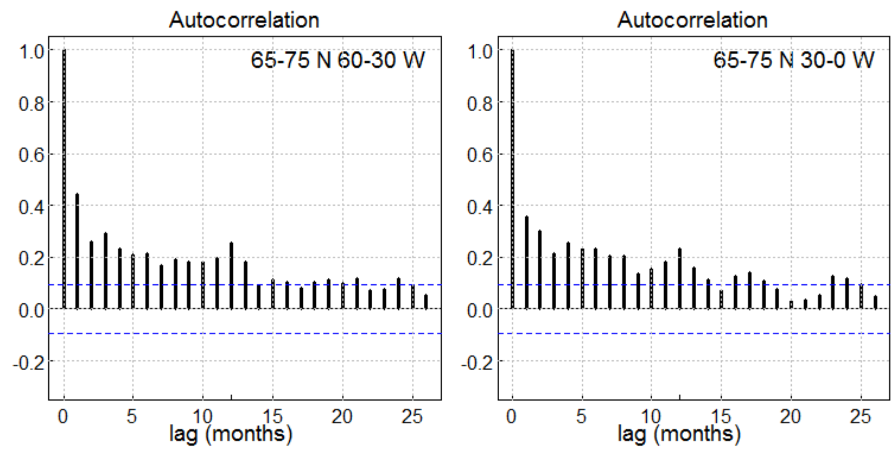


Figure 3.5: Auto correlation function plots of the residuals for two of the 67 sub-regions. The dotted line represents the 95% confidence interval of a zero correlation.

Figure 3.5 is a graph of the auto correlation function (ACF) of the residuals for two of the 67 sub-regions where temperatures have increased. This figure indicates that the residuals are not stationary, as evidenced by significant auto correlation with some lags. The graph of auto correlation (ACF) does not fall within the 95% confidence interval. Therefore, a second order autoregressive process was fitted by equation (2.3), page 24 to reduce these correlations. These residuals were assumed to be stationary. Coefficient $\emptyset_1$and $\emptyset_2$were obtained with two parameters of the data in each region, and the coefficients $\emptyset_1$and $\emptyset_2$ were then averaged across all the 67 sub-regions. The average coefficients are 0.333 and 0.04.

After removing the auto correlation structure using those average coefficients, the filtered residuals were obtained from equation (2.4), page 25. The auto correlation function (ACF) of the filtered residuals for the same two of the 67 sub-regions is shown in Figure 3.6.

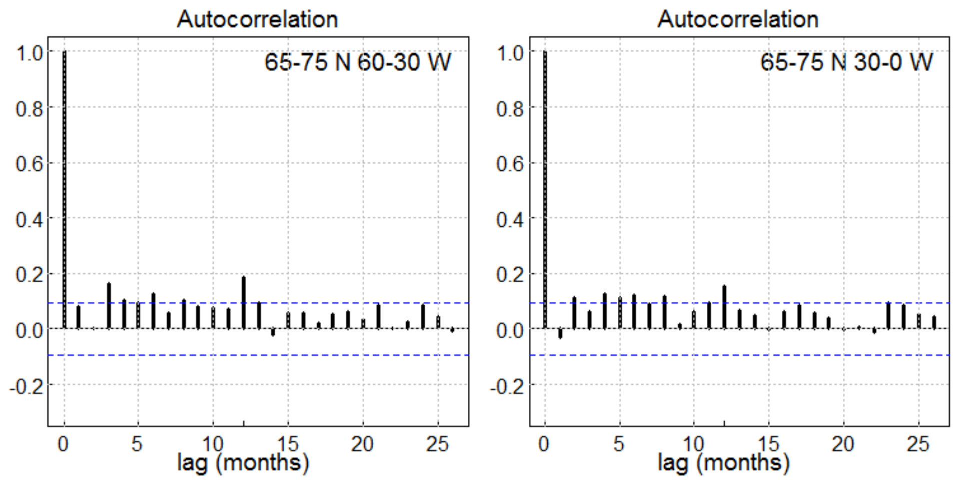


Figure 3.6: Auto correlation function plots of the filtered residuals for the two sub-regions illustrated in Figure 3.5. The dotted line represents the 95% confidence interval of a zero correlation.

Figure 3.6 reveals that the filtering residuals are not significantly auto correlated, because most of the correlations in each sub-region are within the 95% confidence interval.

3.3 Spatial correlation and factor analysis

Spatial correlation between adjacent sub-regions would need to be considered because there are still correlations remaining after filtering. The correlation coefficients among 67 sub-regions range from 0.01 to 0.83, with a median of 0.04.

Factor analysis was applied to identify correlations between the filtered average monthly temperature anomalies in the 67 sub-regions by using model (equation 2.9, page 27). The factor rotation used is the Promax method. The results of factor analysis are shown in 12 factors. Factors with similar patterns represent a large region that consists of two to nine adjoining sub-regions, as shown in Table 3.2 and also shown in map of Figure 3.7.

Sub regions	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Uniqueness
25	**0.772**		0.521				0.112
26	**0.869**	0.152	0.235				0.100
27	**0.623**	0.611		-0.104			0.170
49	**0.786**	-0.203					0.402
50	**0.939**		-0.134				0.134
51	**0.728**	0.186	-0.305				0.276
3		**0.404**	0.334		0.138		0.521
4	-0.101	**0.434**	0.292				0.557
12	0.183	**0.621**	0.258		0.108		0.350
13	-0.134	**0.701**	0.115				0.458
28		**0.912**	-0.123		-0.105		0.208
29	-0.124	**0.766**			-0.116		0.398
52	0.293	**0.468**	-0.313	-0.203			0.557
53		**0.445**	-0.168	0.159			0.713
10	-0.121		**0.722**				0.270
11	0.182	0.163	**0.606**	0.106			0.404
22	-0.129		**0.644**				0.446
23			**0.883**				0.245
24	0.548		**0.718**				0.161
14	-0.125		0.162	**0.351**			0.800
15		-0.159		**0.408**		-0.137	0.703
30		0.153		**0.455**			0.750
31				**0.748**			0.483
32				**0.917**			0.241
33		-0.117		**0.691**	-0.150		0.302
54		0.198		**0.390**	0.129		0.766
55		0.120		**0.400**			0.740
56				**0.487**	-0.104	0.167	0.686
17		-0.143			**0.543**	0.450	0.458
36		-0.100		-0.12	**0.860**		0.107
37					**0.829**	0.260	0.074
59				-0.127	**0.522**	-0.194	0.360
60					**0.640**	-0.142	0.222
18					0.135	**0.832**	0.275
38					0.400	**0.702**	0.078
39						**0.840**	0.073
40						**0.629**	0.085

Table 3.2(a): Results of the factor analysis. Bold values represent factor loadings greater than 0.3.

Sub regions	Factor 7	Factor 8	Factor 9	Factor 10	Factor 11	Factor 12	Uniqueness
61	**0.863**						0.131
62	**0.813**		0.205	0.117			0.110
63	**0.553**	0.129	0.408	0.185			0.176
16	-0.309	**0.447**					0.437
34		**0.857**	0.101				0.190
35	-0.145	**0.706**					0.175
57		**0.639**					0.487
58	0.148	**0.762**					0.442
64	0.195	0.126	**0.871**	0.159	-0.172		0.198
65			**0.783**		0.263		0.210
9	0.131	0.129	-0.177	**0.492**	0.157		0.592
21				**0.895**			0.230
43				**0.585**	0.391		0.455
44				**0.792**			0.470
45				**0.516**	-0.26	0.129	0.604
19			-0.158		**0.519**		0.388
20			-0.135	0.434	**0.655**		0.290
41			0.502	-0.218	**0.557**		0.100
42					**0.957**		0.081
66			0.353		**0.602**		0.380
46						**0.916**	0.212
47						**0.795**	0.345
48						**0.519**	0.521
68		-0.128		0.156		**0.424**	0.635
69				0.108		**0.675**	0.418
5	-0.283						**0.857**
6	-0.155						**0.919**
7	-0.105	0.169	-0.296	0.134			**0.701**
8		0.147	-0.235	0.281			**0.714**
67		-0.123	0.135	0.156	0.158	0.141	**0.788**

Table 3.2(b): Results of the factor analysis. Bold values represent factor loadings greater than 0.3.

Table 3.2, the results were considered for factor loadings greater than 0.3 for each factor. Ideally, each sub-region is correlated with only one factor. Sub-regions with high uniqueness above 0.7 are sub-region 5, 6, 7, 8 and 67. These sub-regions have variations in temperature that are not similar to the other sub-regions. Although sub-regions 14, 15, 30, 54 and 55 have high uniqueness, their factor loadings are more than 0.3. Therefore, these sub-regions could be grouped into a single factor. Overall, there are 62 sub-regions classified into 12 factors and 5 sub-regions with a high degree of uniqueness that are not combined with the other sub-regions.

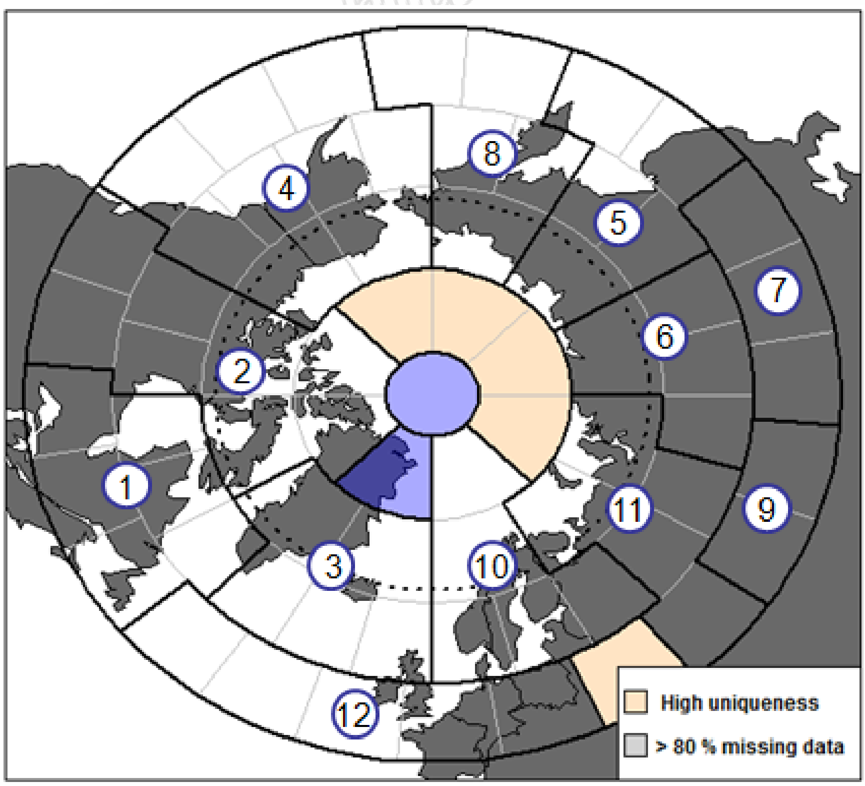


Figure 3.7: Twelve factors (code1-12) identified from factor analysis

In Figure 3.7, the map shows code from 1 to 12 factors, and 5 (in cream color) of 67 sub-regions included in the factor analysis with a high degree of uniqueness.

In addition, the factor model can account for the correlation between those sub-regions by indicating that the results of factor analysis can remove the correlation between sub-regions as shown in Figure 3.8.

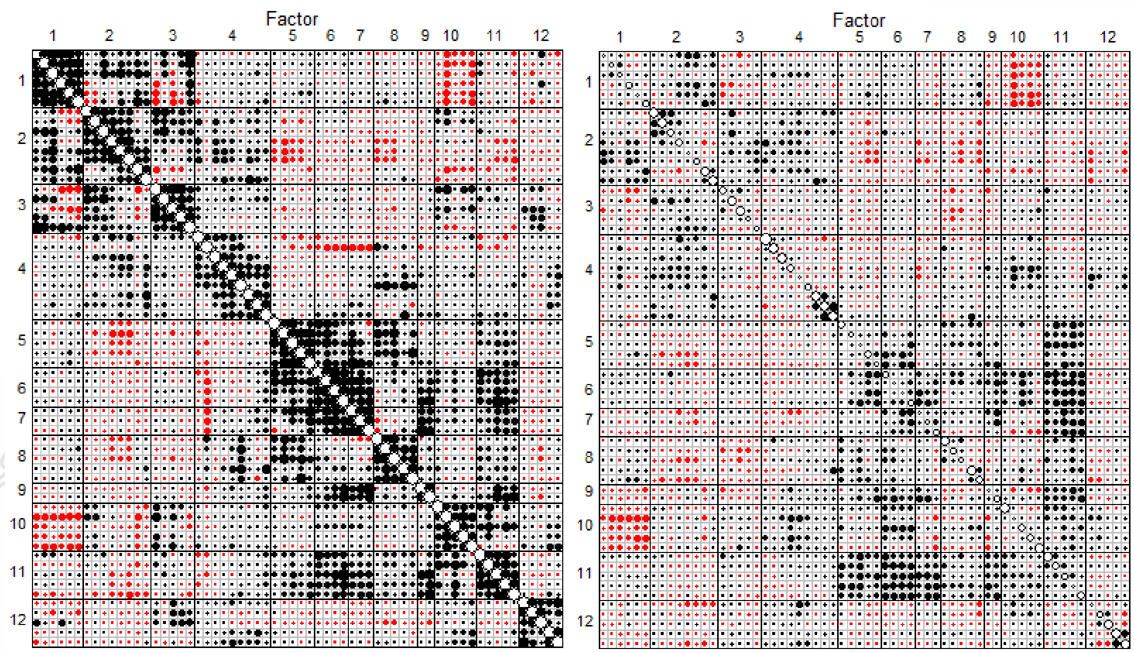


Figure 3.8: Bubble plots matrix of correlations between filtered monthly temperature anomalies in regions before (left panel) and after (right panel) fitting the factor model.

Figure 3.8 presents the correlation matrix of the filtered temperatures between pairs of the sub-regions within each of the 12 factors before doing factor analysis, as displayed in the left panel. The sub-regions within each factor are highly positively correlated, and there is little correlation between pairs of sub-regions from different factors. After doing factor analysis, the factor loadings were obtained. These factor loadings show correlation between sub-regions within a factor. The correlation matrix of the filtered temperatures was removed by the correlation matrix of the factor loading for considering the efficiency of the factor model. These residuals are shown in the right panel. Clearly, pairs of sub-regions even within the same factor show very little correlation, meaning the factor model can reduce the correlation between sub-regions reasonably well. The black dots represent a positive correlation, red dots represent a negative correlation, and the size of each dot shows proportionality to the absolute value of the correlation coefficient.

The filtered monthly temperature anomalies were fitted to the simple linear regression model (equation 2.5, page 25) to see the slope of temperature change. A 95% confidence interval (CI) of the temperature change was estimated for each of the 12 factors, as well as the five sub-regions with high uniqueness in the factor analysis as shown in Table 3.3.

Variables	95 % CI	Increase per decade
Factor 1	(-0.031,0.503)	no significant increase
Factor 2	(0.185,0.558)	0.371
Factor 3	(0.160,0.552)	0.356
Factor 4	(0.259,0.337)	0.298
Factor 5	(0.123,0.617)	0.370
Factor 6	(0.095,0.831)	0.463
Factor 7	(0.119,0.740)	0.430
Factor 8	(0.301,0.474)	0.388
Factor 9	(0.259,0.740)	0.499
Factor 10	(0.151,0.489)	0.320
Factor 11	(0.193,0.788)	0.465
Factor 12	(0.101,0.191)	0.146
Sub-region 5	(0.113,0.448)	0.281
Sub-region 6	(0.094,0.307)	0.186
Sub-region 7	(0.204,0.683)	0.443
Sub-region 8	(0.091,0.492)	0.267
Sub-region 67	(0.090,0.181)	0.136

Table 3.3: 95% confidence interval (CI) of temperature change per decade in each of 12 factors and five sub-regions.

Table 3.3, 11 of the 12 factors and all the 5 sub-regions showed a significant increase in temperature as the 95% CI does not cover zero, except in factor 1 in which there was no evidence of temperature change.

Each such region was reclassified into three levels according to the lower boundary of the 95% confidence interval for temperature change per decade. The three levels were grouped as follow between 0.090˚C and 0.129˚C represents small change, 0.130˚C -

0.199˚C moderate change and 0.200˚C – 0.320˚C large change. Yellow, orange and red colors were used to indicate the three different levels as shown in Figure 3.9.

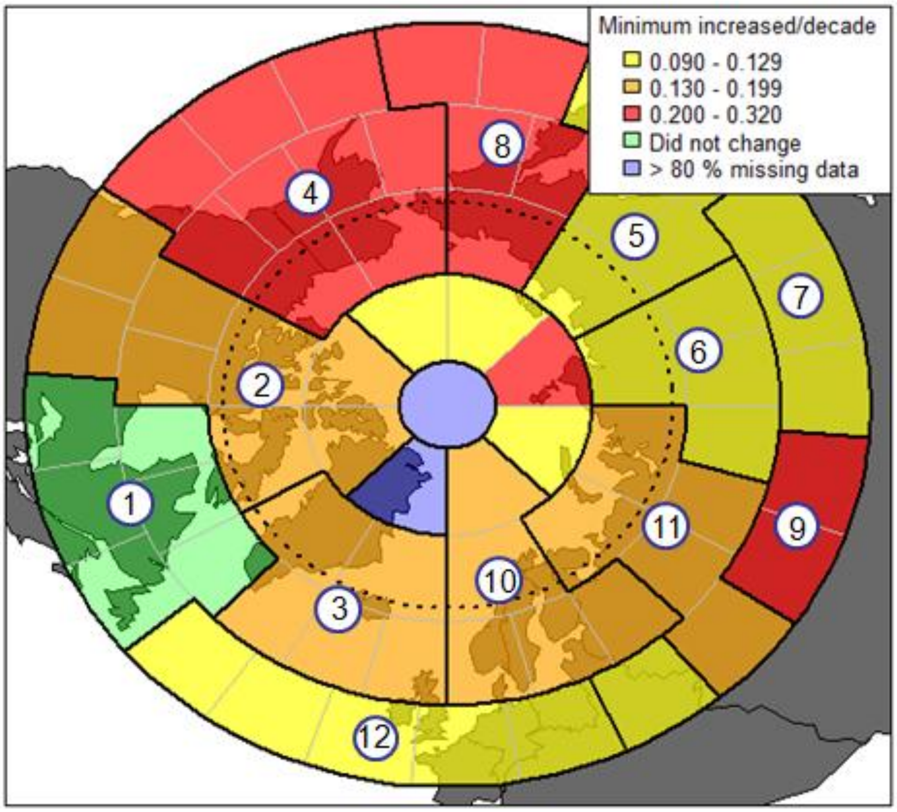


Figure 3.9: The minimum temperature increase per decade in regions.

Obviously in Figure 3.9, the northern Pacific Ocean, Alaska and eastern Siberia have incurred a high increase in temperature (factor 4, 8 and 9); northern Canada, Greenland, Iceland, Norway, Sweden and Finland have shown a moderate increase in temperature (factor 2, 3, 10 and 11); and northern Siberia and part of the north Atlantic have experienced only a slight increase in temperature (factor 5, 6, 7 and 12); whereas northeast Canada and its surrounding seas (factor1) did not experience significant temperature change. Four out of the five sub-regions in yellow experienced a small temperature increase per decade, and another sub-region covering Severanya Zemlya had a high degree of temperature increase.

3.4 Conclusions

This study provides information about temperature changes on 69 sub-regions of the earth surface above latitude 45 degrees north. Monthly average temperature anomalies were examined from 1973 to 2008. A linear regression model was used to investigate the trends and patterns of temperature changes. 64 sub-regions out of 69 experienced a significant increase in temperature, 2 sub-regions had insufficient data, and only 3 sub-regions remained unchanged. Auto correlations of temperatures over time were also checked in time series analysis. Autoregressive process and filtering were used to remove time trends and auto correlation. Factor analysis was then used to model filtered temperatures for identification and classification of regions with similar temperature changes. Twelve factors from factor analysis, each having similar increases per decade, were identified. Each factor had a different quantity of temperature increase. Areas experiencing large temperature increases per decade included the northern Pacific Ocean, Alaska and eastern Siberia which temperature increased from 0.200˚C to 0.320˚C.

CHAPTER 4

Analysis of maximum temperature data in Australia

This chapter contains analyses of maximum temperature data in Australia from two main data sets during the period 1970 to 2012. The first data set is comprised of the maximum temperatures recorded over consecutive 5-day periods. The second data set consists of monthly maximum temperatures.

4.1 Results from the first data set

4.1.1 Preliminary data analysis

Daily maximum temperature data totaling 15,695 observations including missing data over 43 years were recorded in each of the 85 Australian meteorological reporting stations. Observations on February 29 were omitted in order to have equal numbers of observations for each year. The first data set, maximum temperatures over consecutive 5-day periods, was investigated to reduce the correlation between successive data and also to reduce the number of observations to 73 periods in each year. Therefore, there are 3,139 observations over the 43-year period in each of the 85 stations. The results of analysis of maximum temperatures over consecutive 5-day periods also appeared in Wanishsakpong and McNeil (2015).

First, the data were plotted and fitted to a linear regression model (equation 2.5, page 25) to see the trends and patterns of 5-day maximum temperatures for all stations. Figure 4.1 shows the five selected stations because these stations have both increased and decreased.

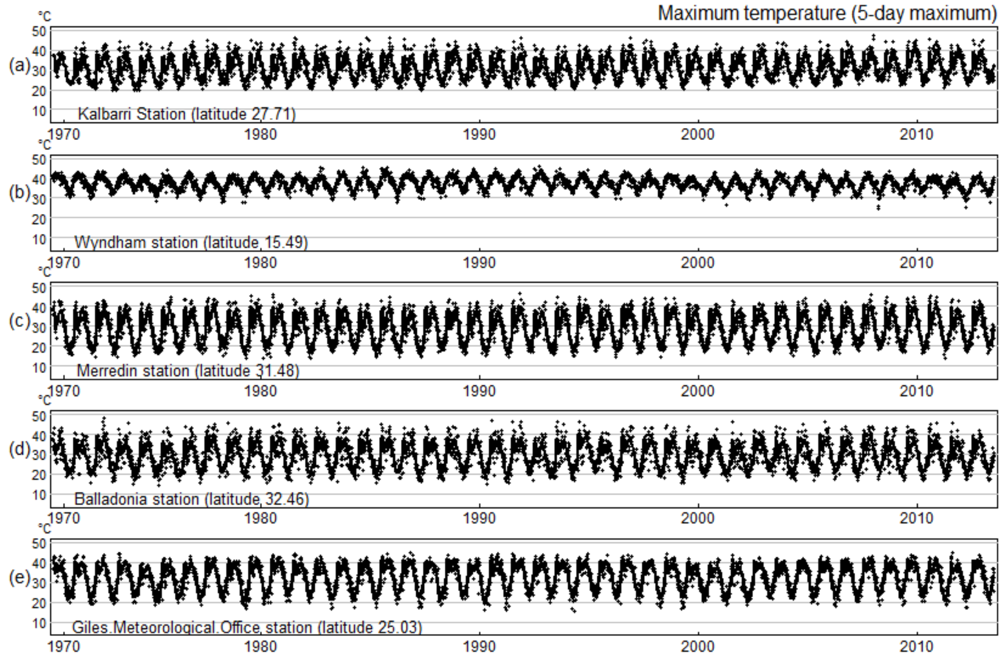


Figure 4.1: Maximum temperatures over consecutive 5-day periods for 5 selected stations (a) Kalbarri station (b) Wyndham station (c) Merredin station (d) Balladonia station and (e) Giles Meteorological station.

In Figure 4.1, the graph demonstrates the maximum temperatures over consecutive 5-day periods for 5 selected stations. The dots represent the data, and the curve represents their fitted values using model (equation 2.5, page 25). Overall, the data clearly show a noticeable seasonal periodicity pattern. Therefore, the data were seasonally adjusted by using equation 2.1, page 22 as shown in the graph for the five stations selected in Figure 4.2.

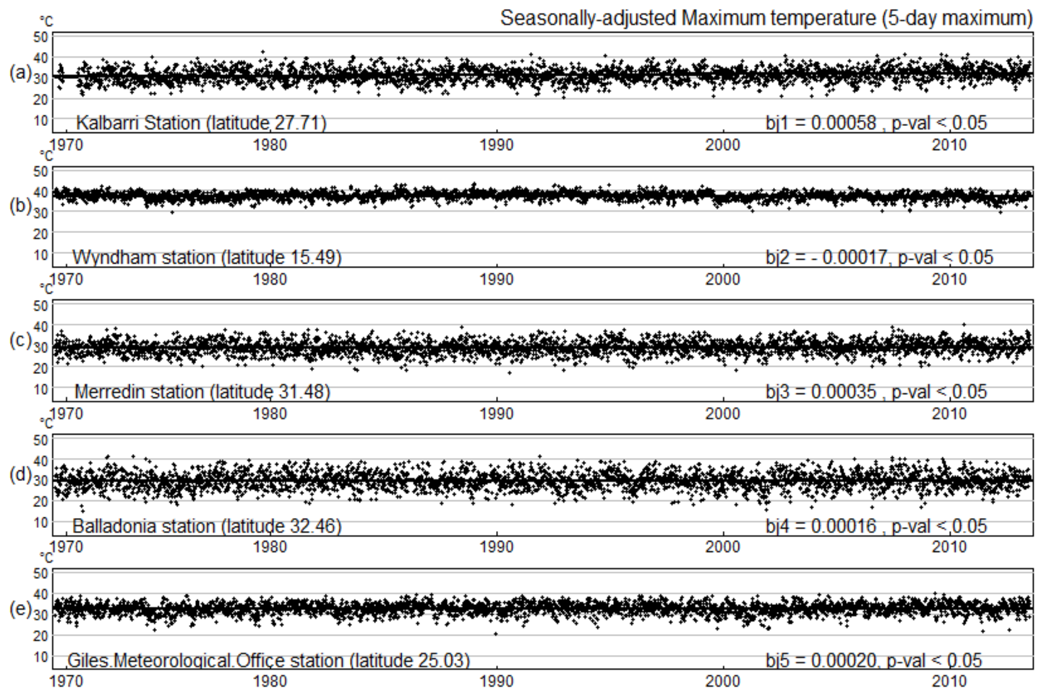


Figure 4.2: Seasonally adjusted maximum temperatures over consecutive 5-day periods of 5 selected stations (a) Kalbarri station (b) Wyndham station (c) Merredin station (d) Balladonia station and (e) Giles Meteorological station.

Figure 4.2 presents the seasonally adjusted maximum temperatures over consecutive 5-day periods for 5 selected stations. The dots represent the seasonally adjusted data and the line represents the temperature trend using model (equation 2.5, page 25). It is clear that the seasonal patterns were removed. Moreover, this seasonally adjusted method is applicable to the other stations. A linear regression model (equation 2.5, page 25) was fitted to the seasonally adjusted data to see the trend of temperature change in each station. For the 5 selected stations, the temperatures increased significantly by about 0.0058˚C , 0.00035˚C, 0.0016˚C and 0.00020˚C in the Kalbarri Station (a), Merredin Station (c), Balladonia Station (d) and Giles Meteorological Office Station (e) respectively, except for Wyndham Station (b) which shows a significant decrease in temperatures of 0.00017˚C. Temperature changes in the other stations are shown in Table 4.1.

Stations	b_{jk}	p-value	Stations	bjk	p-value	Stations	bjk	p-value
a8251	0.00058	< 0.0005	a1013	-0.00018	< 0.0005	a99005	0.00029	< 0.0005
a11003	0.00038	0.00028	a4035	0.00018	0.00017	a75031	0.00037	< 0.0005
a6011	0.00030	0.00001	a2012	-0.00001	0.82617	a90015	0.00012	0.14892
a48031	0.00022	0.00017	a62013	0.00044	< 0.0005	a15085	0.00013	0.00451
a10647	0.00034	< 0.0005	a37010	0.00009	0.06904	a33013	0.00039	< 0.0005
a44026	0.00032	< 0.0005	a44010	0.00016	0.00681	a38002.38	0.00040	< 0.0005
a15528	0.00054	< 0.0005	a15511	0.00043	< 0.0005	a85072	0.00034	0.00001
a49019	0.00060	< 0.0005	a43015	0.00030	< 0.0005	a69018	0.00015	0.02255
a10633	0.00016	0.02764	a18069	0.00046	< 0.0005	a32025	0.00028	< 0.0005
a5026	0.00016	0.00026	a7045	0.00026	< 0.0005	a4032	0.00018	0.00029
a18012	0.00031	0.00121	a17031	0.00014	0.02027	a50052	0.00040	< 0.0005
a46043	0.00016	0.01519	a94029	0.00026	0.00010	a51039	0.00023	0.00009
a28008	0.00015	< 0.0005	a4019	0.00010	0.06615	a23034	0.00027	0.00049
a88109	0.00038	< 0.0005	a9518	0.00022	0.00001	a80023	0.00048	< 0.0005
a2032	0.00006	0.14374	a13017	0.00020	0.00012	a26021	0.00031	0.00011
a48015	0.00015	0.01769	a78077	0.00032	0.00001	a56032	0.00005	0.29877
a15135	0.00011	0.01879	a35065	0.00011	0.02424	a24511	0.00048	< 0.0005
a3003	-0.00008	0.08989	a15602	0.00014	0.01823	a14401	0.00006	0.00346
a38024	0.00035	< 0.0005	a18115	0.00021	< 0.0005	a38003	0.00025	< 0.0005
a14015	0.00008	0.00013	a16001	0.00048	< 0.0005	a10092	0.00036	< 0.0005
a29012	-0.00005	0.11128	a40082	0.00053	< 0.0005	a13012	0.00031	< 0.0005
a36031	0.00016	0.00139	a18110.18	0.00036	0.00004	a84070	0.00039	< 0.0005
a30045	-0.00004	0.35443	a16007.16	0.00012	0.08696	a39083	0.00030	< 0.0005
a25507	0.00023	0.00252	a47016	0.00049	< 0.0005	a11017	0.00017	0.03703
a72043	0.00050	< 0.0005	a19062	0.00034	< 0.0005	a18014	0.00045	< 0.0005
a61250	0.00039	< 0.0005	a9053	0.00044	< 0.0005	a8051	0.00025	0.00069
a31037	0.00015	< 0.0005	a60085	0.00018	0.00173	a58012	0.00037	< 0.0005
a46037	0.00050	< 0.0005	a15590	0.00039	< 0.0005			
a12038	0.00032	< 0.0005	a91009	0.00010	0.00688			

Table 4.1: Regression coefficients of 85 stations

Table 4.1 presents the coefficients for year i and period j at station k (b_{jk}) $(i = 1,2,\ldots,43, j = 1,2,\ldots,3139 \; and \; k = 1,2,\ldots,85)$ which represent temperature change. Overall, temperatures increased in 74 stations, decreased in one station (a1013) in which the coefficient was significantly different from zero (p-value<0.05) and remained the same in 10 stations for which the coefficients were not statistically significant from zero.

After the linear regression model (equation 2.5, page 25) was fitted, the normality assumption requires that residuals be normally distributed. This assumption was verified and assessed by plotting residuals against normal quantiles as shown on Figure 4.3.

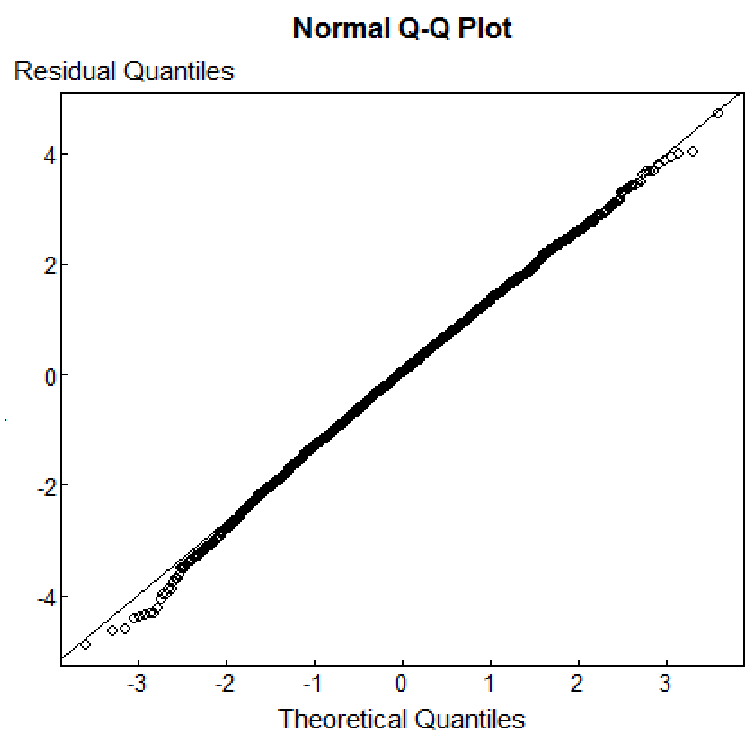


Figure 4.3: Residual quantile plot

Figure 4.3 shows that the residuals lie along a straight line. The straight line corresponds to an exact fit of the data to a normal distribution. Thus, the residuals are normally distributed. In addition, the assumption of independence should be considered because there are correlations between the temperature data and time. These correlations were removed by fitting the first autoregressive model (AR1) from equation 2.3 page 24 and then filtering the data by removing auto correlation at lag 1 month from equation 2.4, page 25. Coefficient $\emptyset_1$ was obtained with a parameter of the data in each station, and the coefficient $\emptyset_1$ was then averaged across all the 85 stations. The average coefficient is 0.2513. The auto correlation function (ACF) of the filtered data for the same five selected stations is shown in Figure 4.4.

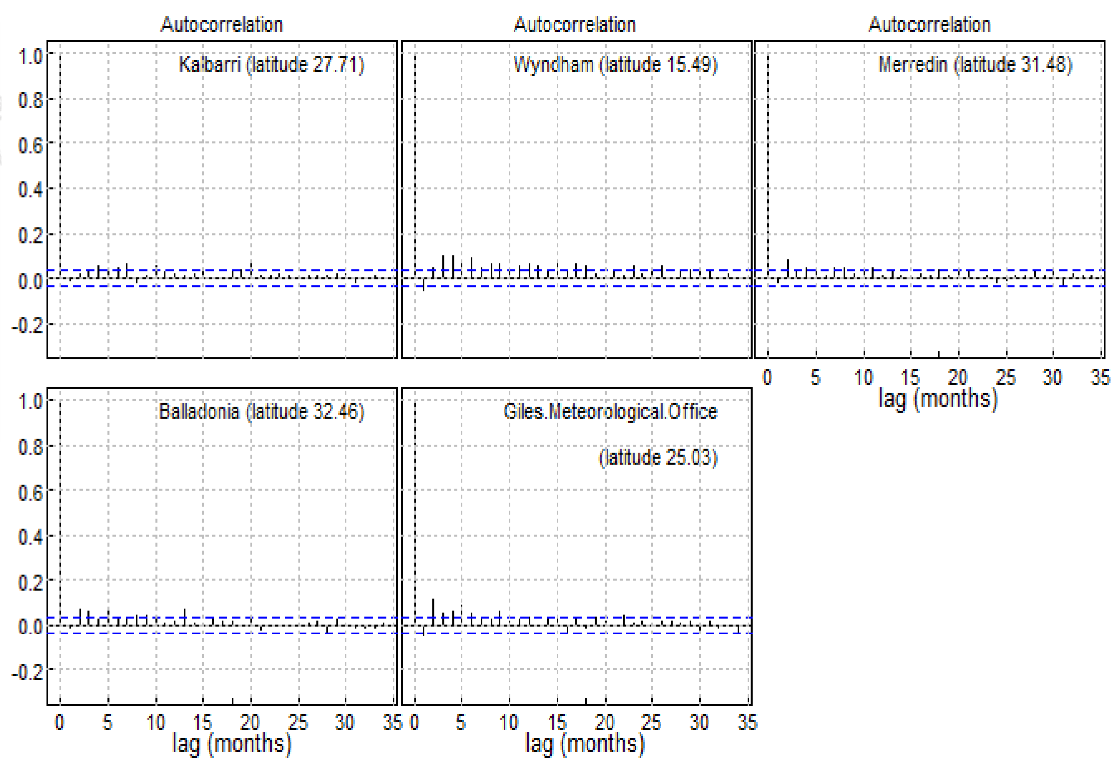


Figure 4.4: Auto correlation function plots of the filtered residuals for the five selected stations. The dotted line represents the 95% confidence interval of a zero correlation.

Figure 4.4 shows that most of correlations in the five stations are within the 95% confidence interval. Therefore, the assumption of independence is supported, and the filtered seasonally adjusted maximum temperatures were obtained for analysis in the following section.

4.1.2 Spatial correlation and factor analysis

Spatial correlation of the filtered seasonally adjusted maximum temperatures was investigated. After removing auto correlation in each station, there are still correlations among the 85 stations. Correlation coefficients among the 85 stations ranged from 0.00007 to 0.925, with a median of 0.167.

A factor model (equation 2.9, page 27) was used to reduce these spatial correlations by classifying the 85 stations into factors considering the value of the factor loadings greater than 0.3 for each factor as shown in Table 4.2. The rotation of factors used the Promax method.

Stations	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Uniqueness
a88109	**0.93**							0.23
a80023	**0.87**							0.17
a99005	**0.87**		0.11	-0.17				0.42
a85072	**0.87**			-0.14				0.36
a78077	**0.85**				0.12			0.13
a75031	**0.75**	0.17		0.15				0.16
a90015	**0.75**	-0.10	0.11	-0.16	0.24			0.38
a84070	**0.75**						-0.13	0.48
a94029	**0.73**	-0.12	0.11	-0.15	0.15			0.55
a72043	**0.73**	0.22						0.36
a47016	**0.63**			0.13	0.21			0.31
a49019	**0.55**	0.25		0.26				0.29
a91009	**0.52**							0.79
a46043	**0.43**	0.28		**0.43**				0.19
a19062	**0.47**			0.27	**0.41**			0.16
a50052	**0.49**	**0.46**	-0.13	0.17	-0.11			0.24
a24511	**0.56**			0.14	**0.51**			0.11
a25507	**0.70**	-0.10			**0.56**	-0.13		0.10
a26021	**0.70**	-0.16	0.12	-0.12	**0.55**	-0.15		0.19
a56032		**0.83**						0.35
a40082	-0.16	**0.81**		-0.18	0.11	0.10		0.39
a60085		**0.80**	-0.12	-0.14		0.16		0.39
a48031		**0.79**						0.28
a44010		**0.77**		0.18				0.22
a43015	-0.11	**0.75**	0.24		0.10			0.31
a61250	0.20	**0.71**	-0.16		-0.10	0.16	-0.11	0.39
a48015	0.16	**0.66**		0.25				0.28
a35065	-0.14	**0.65**	0.25		0.12			0.36
a62013	0.28	**0.63**	-0.15					0.27
a51039	0.29	**0.60**	-0.16	0.22				0.23
a39083	-0.21	**0.53**	0.12	-0.18				0.50
a58012	-0.18	**0.50**		-0.12		0.11		0.72
a44026		**0.69**		**0.33**		-0.11		0.21
a69018	**0.30**	**0.38**			-0.12	0.16	-0.13	0.71
a30045		0.27	**0.70**			-0.18		0.35
a29012			**0.69**					0.51
a15528			**0.52**	0.11		0.23		0.38
a31037			**0.40**	-0.12				0.75
a14401			**0.39**	-0.12		0.17		0.84
a14015	0.12		**0.36**			0.24		0.82
a32025			**0.36**	-0.14				0.79

Table 4.2(a): Factor loadings of the 85 stations from factor analysis by Promax method.

Stations	**Factor 1**	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**	**Factor 6**	**Factor 7**	**Uniqueness**
a15135			**0.86**	**0.45**				0.18
a15085	0.11		**0.85**	**0.35**				0.27
a37010			**0.82**	**0.39**			0.11	0.29
a2032	0.12	-0.13	**0.73**	**0.13**		**0.41**		0.32
a2012		-0.15	**0.69**	**0.21**		**0.47**	0.12	0.28
a1013		-0.15	**0.67**			**0.40**		0.43
a38003		0.22	**0.63**	**0.55**		-0.22		0.23
a36031		**0.49**	**0.47**	**0.17**		-0.20		0.26
a33013	-0.12	**0.39**	**0.47**	**-0.16**				0.52
a38002.38		0.24	0.28	**0.75**		-0.13		0.22
a15511	-0.14		0.16	**0.70**	0.14	0.29	-0.15	0.27
a17031	0.20	0.11		**0.60**	0.17			0.21
a15590			**0.42**	**0.76**		0.11		0.21
a15602			**0.54**	**0.75**				0.22
a46037	0.22	**0.35**		**0.59**				0.18
a16007.16				**0.56**	**0.34**	0.16		0.27
a13017	-0.18		0.13	**0.53**	0.18	**0.50**	-0.13	0.30
a38024		**0.46**	**0.41**	**0.48**		-0.19	0.11	0.22
a16001	0.26			**0.43**	**0.41**			0.18
a18012	0.12		-0.12		**0.76**	0.11		0.24
a18069	0.20				**0.75**			0.29
a18110.18			-0.12		**0.64**	0.25		0.37
a11003					**0.57**	0.22		0.54
a23034	**0.52**				**0.65**			0.12
a18014	**0.36**				**0.64**			0.24
a18115	**0.43**			-0.12	**0.61**			0.37
a5026						**0.74**		0.39
a4035						**0.68**		0.51
a4032			0.13			**0.62**		0.57
a13012	-0.12		-0.13		0.21	**0.61**	0.22	0.39
a4019			0.15			**0.57**		0.64
a3003			0.16	-0.11		**0.41**	0.14	0.80
a7045			-0.13		0.13	**0.60**	**0.35**	0.37
a10647			0.12				**0.86**	0.26
a10092						0.24	**0.80**	0.20
a9053					-0.11		**0.77**	0.42
a10633					0.11	0.16	**0.75**	0.35
a8051					-0.22		**0.56**	0.64
a9518		-0.12					**0.56**	0.68
a8251				-0.12	-0.22		**0.48**	0.71
a6011				-0.14	-0.18		**0.37**	0.80
a12038					0.28	**0.41**	**0.49**	0.35
a11017					**0.33**	0.29	**0.39**	0.52
A28008			0.27	-0.11			-0.12	**0.86**

Table 4.2(b): Factor loadings of the 85 stations from factor analysis by Promax method.

In Table 4.2, 85 stations are classified into 7 factors with one station (Lockhart River Airport; a28008) having a uniqueness value 0.86. Thus, there is no dominating factor for this station. The factor loadings are ordered from highest to lowest within each factor (excluding the mixed factor loading). The results of factor analysis are also presented in the map of Australia as shown in Figure 4.4.

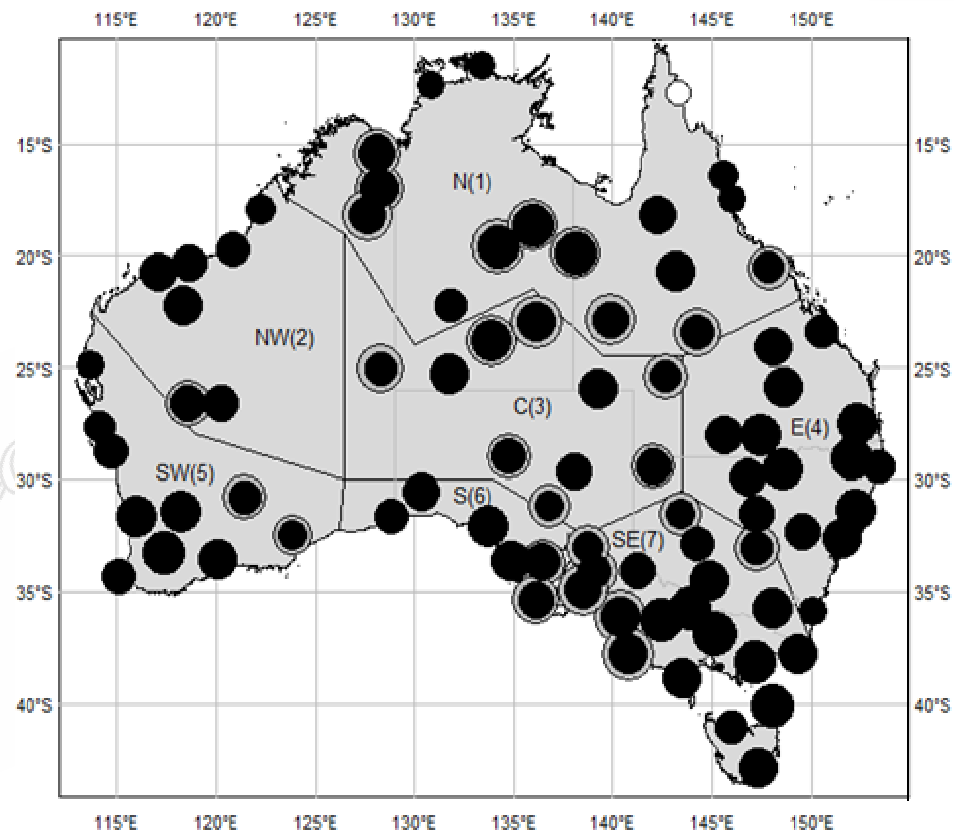


Figure 4.5: Factor analysis divides the stations into 7 geographical groups. Numbers in brackets are the number of factors. E = east, SE = southeast, S = south, SW = southwest, NW = northwest, N = north, C = central.

Figure 4.5 shows a map of Australia in seven regions as determined by the factor model. The circles represent the 85 stations. The size of each circle is proportionate to the factor loading for each station. The superimposed circles show thirty stations with mixed factor loadings, which were omitted in subsequent analyses. The hollow circle shows one station in northern Australia with a high degree of uniqueness.

4.1.3. Polynomial regression model

The means of maximum temperatures for consecutive 5-day periods over 43 years are presented to demonstrate the patterns in each factor as shown in Figure 4.6.

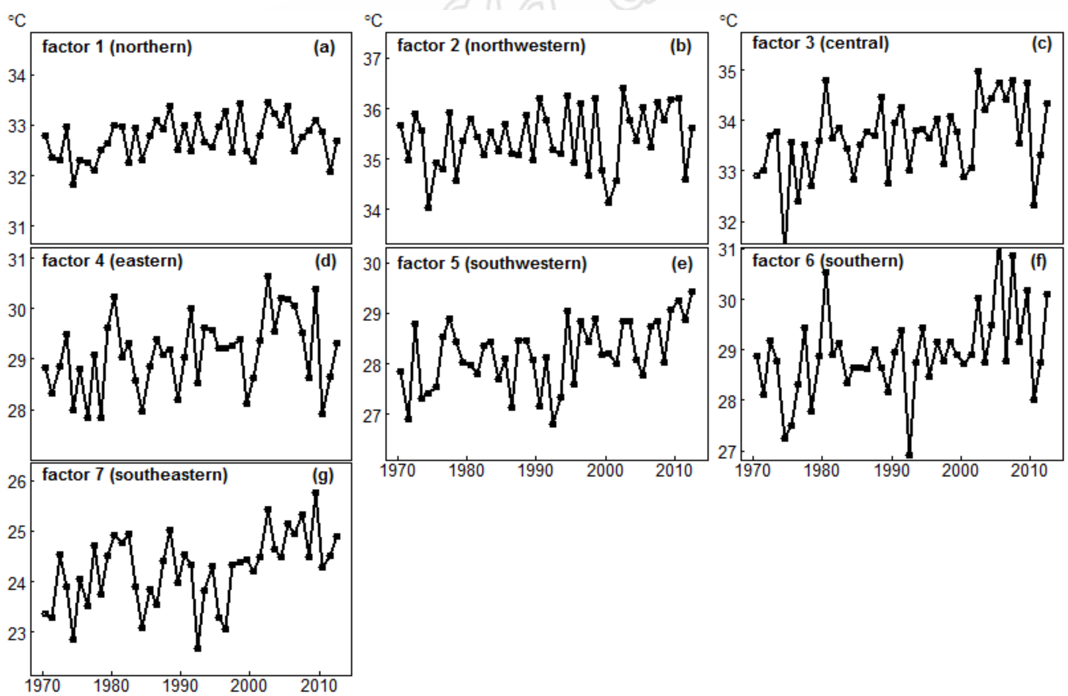


Figure 4.6: The mean of maximum temperatures for consecutive 5-day periods in each factor.

In Figure 4.6, the graph shows the pattern of temperature changes in the seven factors. The dots represent the means of maximum temperatures over consecutive 5-day periods based on the model (equation 2.5, Page 25). The data display a curvilinear pattern in each factor. Therefore, a polynomial regression model (equation 2.7, page 26) was applied with this pattern. The coefficients of the polynomial regression model increased significantly to the sixth order as shown in Table 4.3.

Factor1	**Coefficients**	**p-value**
b_{01}	32.99	< 0.0005
b_{11}	-0.005900	< 0.0005
b_{21}	0.000015	< 0.0005
b_{31}	-1.61E-08	0.00001
b_{41}	7.98E-12	0.00021
b_{51}	-1.89E-15	0.00150
b_{61}	1.71E-19	0.00657

Factor2	**Coefficients**	**p-value**
b_{02}	35.8	< 0.0005
b_{12}	-0.006000	0.03730
b_{22}	0.000012	< 0.0005
b_{32}	-1.04E-08	< 0.0005
b_{42}	3.63E-12	< 0.0005
b_{52}	-4.50E-16	< 0.0005
b_{62}	-1.14E-19	< 0.0005

Factor3	**Coefficients**	**p-value**
b_{03}	33.63	< 0.0005
b_{13}	-0.00780	0.00869
b_{23}	0.000024	0.00292
b_{33}	-2.87E-08	0.00394
b_{43}	1.54E-11	0.00743
b_{53}	-3.88E-15	0.01496
b_{63}	3.68E-19	0.02908

Factor4	**Coefficients**	**p-value**
b_{04}	28.97	< 0.0005
b_{14}	-0.004330	0.0490
b_{24}	0.000014	0.0202
b_{34}	-1.84E-08	0.0155
b_{44}	1.06E-11	0.0166
b_{54}	-2.80E-15	0.0219
b_{64}	2.76E-19	0.0326

Factor5	Coefficients	p-value
b_{05}	27.58	< 0.0005
b_{15}	0.004700	0.0052
b_{25}	-8.25E-06	0.0135
b_{35}	5.94E-09	0.0275
b_{45}	9.98E-12	0.0480
b_{55}	-3.10E-15	0.0312
b_{65}	3.53E-19	0.0205

Factor6	Coefficients	p-value
b_{06}	29.39	< 0.0005
b_{16}	-0.011960	0.01187
b_{26}	0.000037	0.00441
b_{36}	-4.56E-08	0.00383
b_{46}	2.57E-11	0.00491
b_{56}	-6.74E-15	0.00757
b_{66}	6.68E-19	0.01238

Factor7	Coefficients	p-value
b_{07}	23.55	< 0.0005
b_{17}	-0.001810	< 0.0005
b_{27}	0.000012	< 0.0005
b_{37}	-1.88E-08	0.04850
b_{47}	1.11E-11	0.04830
b_{57}	-2.88E-15	0.04020
b_{67}	2.73E-19	0.04420

Table 4.3: Sixth order polynomial regression coefficients of 7 factors.

Table 4.3 shows coefficients of the sixth order polynomial regression model in each of the 7 factors. All coefficients are significantly different from zero (p-value < 0.05). The data also were plotted with a sixth order polynomial regression model to see the trends and patterns of temperature change in each factor as shown in Figure 4.7.

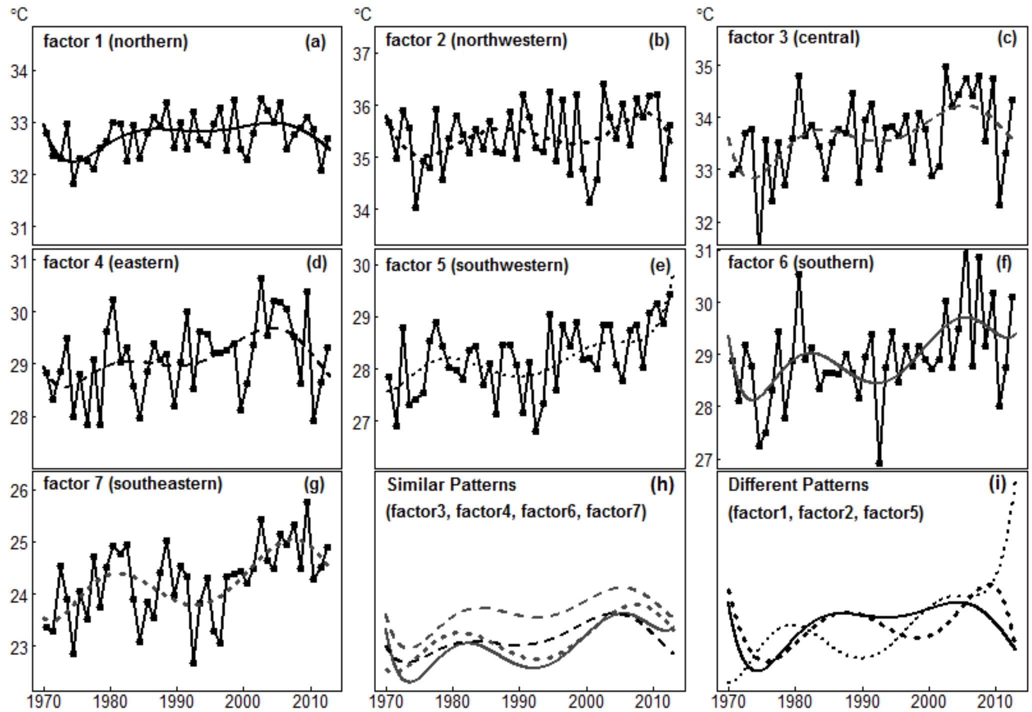


Figure 4.7: The patterns of temperature change in seven factors corresponding to seven geographical (a)-(g) regions. The bottom centre panel shows regions with similar patterns, while the bottom right panel shows regions with dissimilar patterns.

Figure 4.7(a) – 4.7(g) present the temperature trends in each of the 7 factors. The dots are means of maximum temperature over consecutive 5-day periods based on a linear regression model (equation 2.5, page 25) over 43 years. The average maximum temperatures in all factors ranged from 23^{o}C to 36^{o}C. Factor 2 represents the desert area of the northwest, the area with the highest average temperature of approximately 36^{o}C in 2002. The data were fitted to a polynomial regression model (equation 2.7, page 26) in each factor. Four geographic regions, namely the central (factor3), eastern (factor4), southern (factor6) and southeastern (factor7) regions were grouped with similar patterns as shown in figure 4.7(h). These patterns show an increase in temperatures after 1974 followed by a steady decrease in temperatures around 1984 and with another gradual increase from 2000 until 2005, and finally another decrease through 2012. Three geographic regions, the northern region (factor1), the northwestern region (factor2) and the southwestern region (factor5), had different temperature patterns as shown on figure 4.7(i).

4.1.4 Conclusions

Maximum temperatures over consecutive 5-day periods were adjusted to remove seasonal effects in the data. A linear regression model was initially used to model the seasonally adjusted temperatures in order to examine the trends and patterns of temperatures in the 85 Australian meteorological stations. Temperatures increased at 74 stations, decreased at 1 station, and there was no evidence of change at 10 stations. Factor analysis was used to classify the temperatures of the 85 stations into seven factors, each having different patterns of temperature changes. These seven factors corresponded to seven geographical regions, namely the northern, northwestern, central, eastern, southwestern, southern and southeastern regions. Four geographic regions (central, eastern, southern and southeastern) had similar temperature patterns, whereas three geographic regions (northern, northwestern and southwestern) had different temperature patterns.

4.2 Results of the second data set

4.2.1 Preliminary data analysis

The monthly maximum temperature data studied were the highest daily temperature recorded during the particular month. Therefore, 516 months over 43 years were included in each of the 85 meteorological reporting stations. The sample of 85 stations was randomly selected using the same criteria as in the first data set as set in section 4.1.1. The percentage of missing data from each of the 85 meteorological stations is shown in Figure 4.8. The results of analysis of monthly maximum temperatures also appeared in Wanishsakpong *et al*. (2015).

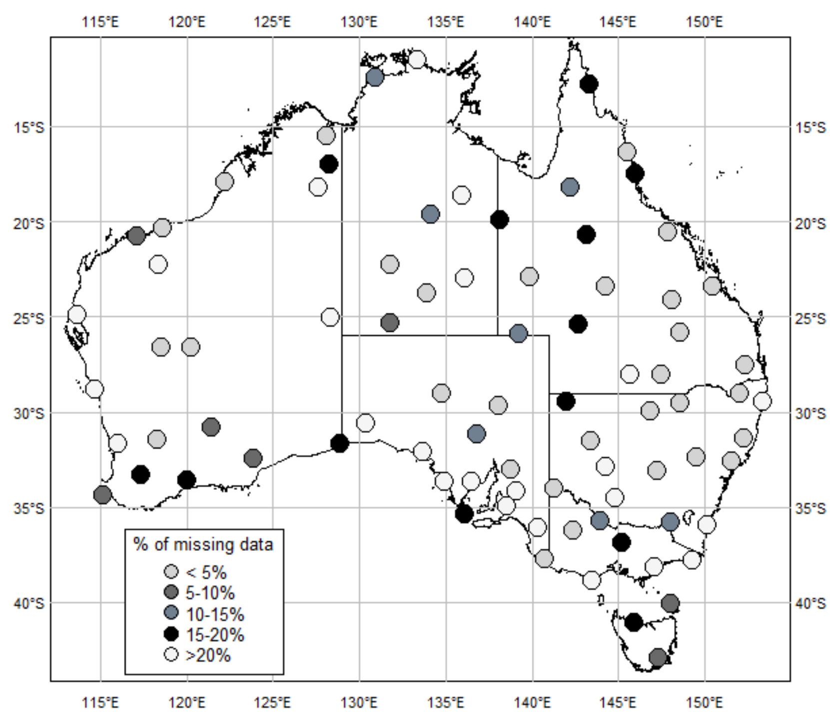


Figure 4.8: An area-based sample of 85 stations with indications of percentage missing data in each station. The state boundaries of Australia are also shown.

Figure 4.8 shows the percentage of missing data in each station. Most of the stations have missing data of less than 5%. These missing data were estimated by using multiple linear regression models (equation 2.6, page 26). Ideally, this model assumes that the maximum monthly temperature of a particular month and station can be predicted using the previous and subsequent monthly temperatures as well as the first and second nearest station temperatures at the predefined month.

The model (equation 2.6, page 26) relating temperatures at month t and temperatures at neighboring stations as well as temperatures at month$(t-1)$ and $(t+1)$ was fitted with a random sample of 10 stations as shown in Figure 4.9.

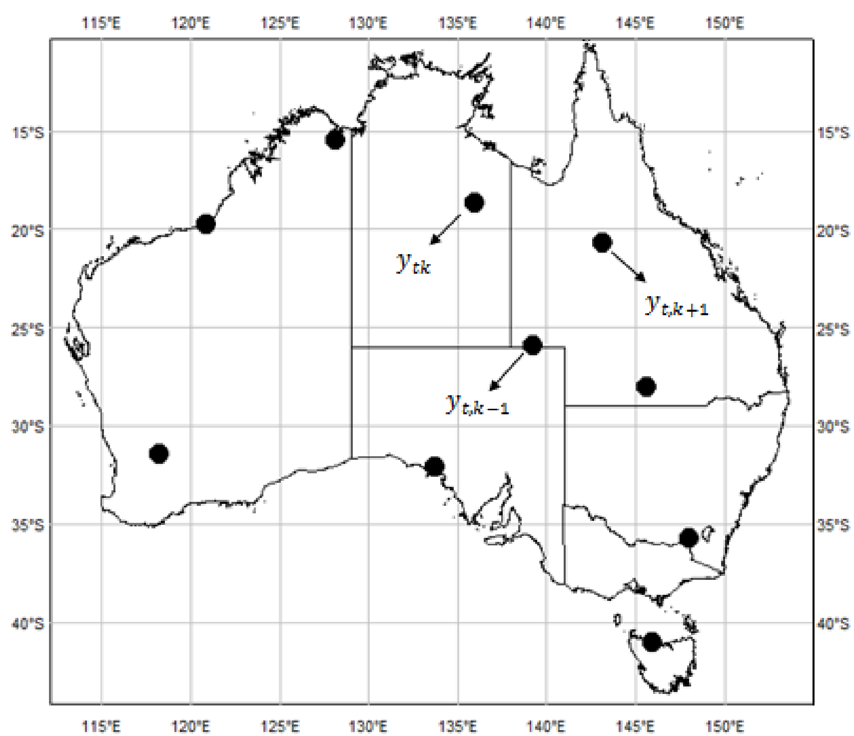


Figure 4.9: A random sample of 10 stations and example of configuration of Y_{tk}, $Y_{t,k-1}$ and $Y_{t,k+1}$.

The results of regression equation 2.3 are shown in Table 4.4.

Predictor	Coefficient	P-value	Adjusted R^2
Constant	-2.1186	< 0.05	0.942
Nearest 1 (y_{t,k^-})	0.2869	< 0.05	
Nearest 2 (y_{t,k^+})	0.2945	< 0.05	
Previous month $(y_{t-1,k})$	0.2464	< 0.05	
Next month $(y_{t+1,k})$	0.2407	< 0.05	

Table 4.4: The coefficients of regression model (equation 2.6)

Table 4.4 shows all coefficients of the regression model (equation 2.6, page 26) are statistically significant (P-value < 0.05), and the adjusted R-square is 94.2%. Thus, it would be confirmed that this model can estimate all missing data in the 85 stations.

After the missing values were interpolated, maximum monthly temperatures were then plotted to investigate the patterns in four selected stations out of 85 stations, one each in the northern, southern, eastern and western regions of Australia. The regional station locations reflected the four different temperature curve patterns during the 10- year periods. Therefore, a quartic trend model combined with the 3rd order time lag (equation 2.8, page 27) was applied with these stations as shown in Figure 4.10.

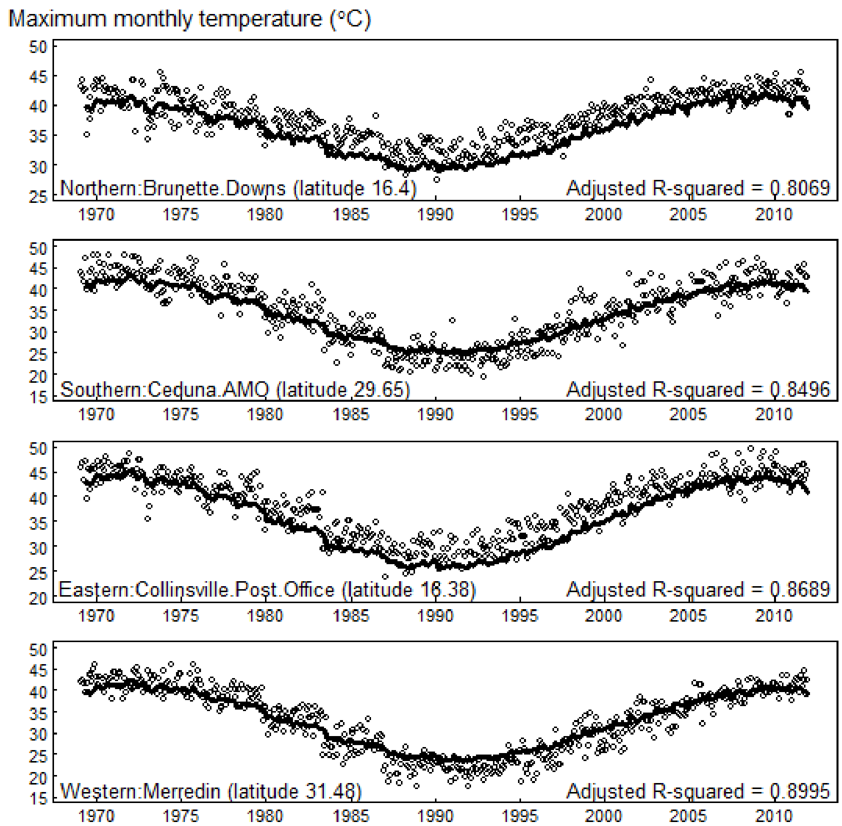


Figure 4.10: Observed maximum monthly temperatures with fitted values using the regression equation 2.8 for four stations.

Figure 4.10, the dots represent the maximum monthly temperature over 43 years in four selected stations. The data were fitted with the quartic trend model combined with the 3rd order time lag (equation 2.8, page 27). It is clear that this model can estimate the pattern and trend of the time series data reasonably well as shown by an adjusted R-squared of more than 80% for the four selected stations. Moreover, this model was also applied equally well to the other stations and was indicative of a polynomial pattern with a range between 20˚C and 50˚C in the 85 meteorological stations.

4.2.2 Spatial correlation, factor analysis and cluster analysis

The patterns and trends of temperatures for the 85 stations have similar polynomial patterns. Correlation coefficients among the 85 stations ranged from 0.06 to 0.95, with a median of 0.86. A factor analysis model (equation 2.9, page 27) was used to group these 85 stations into several factors to reduce these spatial correlations. The rotation of factors used was the Promax method. The results of factor analysis are shown in 3 factors but the correlations between factors were high, ranging from 0.88 to 0.95 with a median of 0.9415. Therefore, in this case the Varimax rotation method was used because this procedure keeps the factors uncorrelated. The resulting factor loadings are presented in Table 4.5.

Stations	Factor1	Factor2	Factor3
a9053	**0.88**	-0.35	-0.22
a8251	**0.87**	-0.33	-0.22
a8051	**0.86**	-0.37	-0.22
a10647	**0.85**	-0.39	-0.26
a26021	**0.83**	-0.34	-0.38
a10633	**0.83**	-0.40	-0.29
a6011	**0.83**	-0.30	-0.19
a91009	**0.82**	-0.26	-0.35
a10092	**0.82**	-0.43	-0.31
a9518	**0.81**	-0.18	0.02
a18115	**0.81**	-0.25	-0.38
a99005	**0.81**	-0.27	-0.40
a25507	**0.80**	-0.39	-0.42
a88109	**0.79**	-0.37	-0.43
a72043	**0.79**	-0.36	-0.46
a90015	**0.79**	-0.31	-0.41
a78077	**0.79**	-0.40	-0.44
a23034	**0.78**	-0.40	-0.44
a94029	**0.78**	-0.33	-0.40
a85072	**0.78**	-0.34	-0.46
a24511	**0.77**	-0.40	-0.46
a12038	**0.76**	-0.47	-0.36
a80023	**0.76**	-0.40	-0.47
a18014	**0.76**	-0.41	-0.46
a7045	**0.76**	-0.51	-0.35
a11017	**0.76**	-0.46	-0.35
a19062	**0.75**	-0.43	-0.47
a13012	**0.74**	-0.52	-0.36
a75031	**0.74**	-0.42	-0.50
a47016	**0.73**	-0.42	-0.48
a18069	**0.73**	-0.44	-0.43
a18012	**0.72**	-0.46	-0.45
a84070	**0.71**	-0.35	-0.49
a18110.1	**0.71**	-0.48	-0.42
a4035	**0.71**	-0.58	-0.31
a11003	**0.71**	-0.45	-0.38
a49019	**0.70**	-0.45	-0.51
a50052	**0.70**	-0.43	-0.55
a5026	**0.69**	**-0.59**	-0.35
a62013	**0.69**	-0.42	-0.56
a4032	**0.68**	**-0.59**	-0.31
a17031	**0.68**	-0.50	-0.51

Table 4.5(a): The first three factor loadings of 85 stations with cumulative variances of 90.14%

Stations	Factor1	Factor2	Factor3
a46043	**0.68**	-0.48	-0.53
a16001	**0.68**	-0.49	-0.51
a51039	**0.66**	-0.46	-0.56
a13017	**0.65**	**-0.60**	-0.41
a46037	**0.64**	-0.51	-0.55
a16007.1	**0.64**	-0.53	-0.50
a15511	**0.62**	**-0.60**	-0.45
a48015	**0.61**	-0.50	**-0.58**
a31037	**0.61**	**-0.59**	-0.34
a32025	**0.60**	-0.50	-0.26
a44026	**0.58**	**-0.55**	**-0.57**
a29012	0.28	**-0.85**	-0.32
a1013	0.36	**-0.84**	-0.27
a2032	0.39	**-0.82**	-0.32
a15085	0.42	**-0.78**	-0.36
a2012	0.45	**-0.78**	-0.35
a14015	-0.06	**-0.78**	0.00
a14401	0.37	**-0.76**	-0.21
a3003	0.32	**-0.76**	-0.09
a30045	0.41	**-0.75**	-0.42
a37010	0.44	**-0.75**	-0.40
a15135	0.49	**-0.74**	-0.39
a36031	0.43	**-0.71**	-0.49
a33013	0.46	**-0.70**	-0.42
a38003	0.48	**-0.69**	-0.48
a15602	0.51	**-0.67**	-0.48
a15528	0.56	**-0.66**	-0.43
a38024	0.51	**-0.64**	-0.53
a15590	0.55	**-0.64**	-0.48
a35065	0.49	**-0.64**	-0.51
a4019	0.59	**-0.64**	-0.29
a39083	0.48	**-0.64**	-0.46
a38002.3	0.54	**-0.62**	-0.53
a28008	0.54	**-0.60**	-0.29
a43015	**0.52**	**-0.57**	**-0.56**

Table 4.5(b): The first three factor loadings of 85 stations with cumulative variances of 90.14%

Stations	Factor1	Factor2	Factor3
a60085	0.51	-0.51	**-0.62**
a61250	0.59	-0.47	**-0.61**
a40082	0.48	-0.55	**-0.59**
a44010	0.54	-0.55	**-0.59**
a48031	**0.58**	-0.51	**-0.59**
a56032	**0.58**	-0.49	**-0.59**
a69018	0.49	-0.39	**-0.58**
a58012	0.26	-0.53	**-0.56**
Var	36.39	24.48	16.01
%Var	0.428	0.288	0.188

Table 4.5(c): The first three factor loadings of 85 stations with cumulative variances of 90.14%

Table 4.5 shows the results of factor analysis. Three factors were obtained with a cumulative variance of 90.4%. Although the Varimax rotation method produced less correlation between these factors than the Promax rotation method, there are still high correlations ranging from 0.80 to 0.94, with a median of 0.934. Moreover, there are several stations having similar absolute factor loadings, which in this situation mean that it is not easy to group the stations based on the results of this factor analysis. The stations with similar absolute factor loadings are a5026, a4023, a13017, a15511, a48015, a31037, a44026, a43015, a48031 and a56032. Therefore, cluster analysis was defined based on the similarity of the factor loadings.

The hierarchical clustering method and distance or dissimilarity between groups of stations was defined by using the complete linkage based on the Euclidian distance. The resulting cluster analysis produced grouped into four clusters. A description of the results of the cluster analysis follows in Table 4.6.

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster 1	39	0.6486	0.1183	0.2541
Cluster 2	14	0.2448	0.1165	0.3322
Cluster 3	12	0.1615	0.1083	0.2489
Cluster 4	20	0.3055	0.1121	0.2578

Table 4.6: Description of resulting clusters based on factor loadings

Table 4.6 shows the statistical information about the cluster solution. The table summarizes each cluster by the number of observations, the within cluster sum of squares, the average distance from the observation to the cluster centroid and the maximum distance of the observation to the cluster centroid. In general, a cluster with a small sum of squares is more compact than one with a large sum of squares. Centroid distance is a measure of the homogeneity of merged clusters, and its value should be small. Therefore, this indicates that the four clusters are reasonably sufficient for the partition. These four clusters are shown in the map of Australia in Figure 4.11.

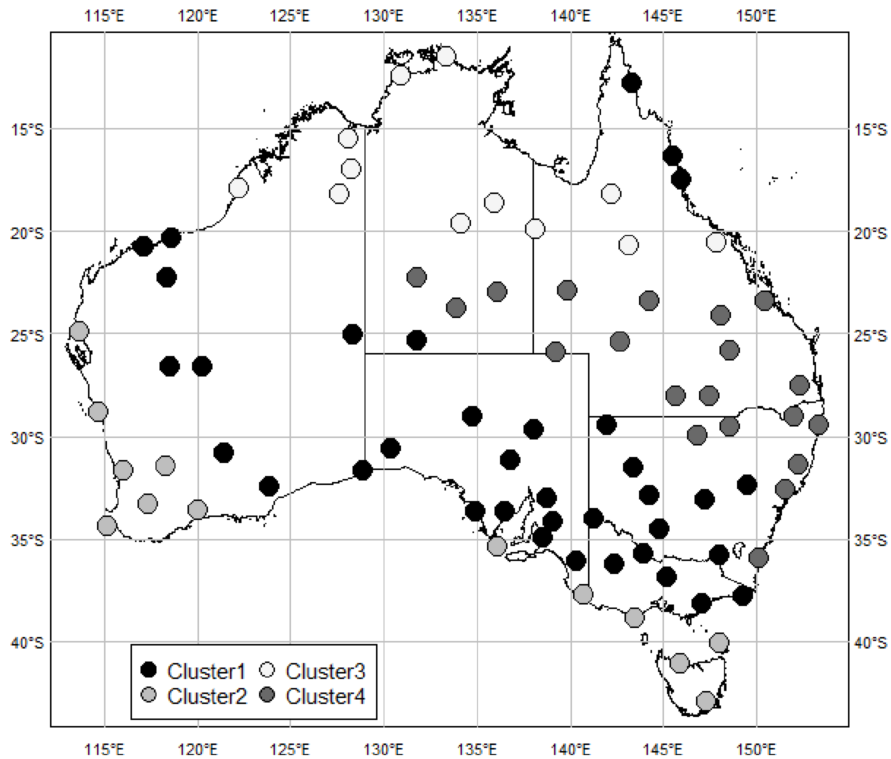


Figure 4.11: Cluster analysis divides the stations into 4 groups

Figure 4.11 shows the corresponding geographical regions for the four clusters are the north-west with some parts of the central and south regions, the south-eastern region including some parts of the northern region (Cluster1); the lower boundary of the south-western and, southern region including Tasmania (Cluster2); the north (Cluster3) and some parts of the central and eastern region (Cluster4).

In each cluster, the data were plotted to see the trends and patterns of temperature as shown in Figure 4.12. The data in each cluster show similar polynomial patterns. Therefore, the quartic trend model with 3rd order time lag (equation 2.8, page 27) was also fitted to each cluster.

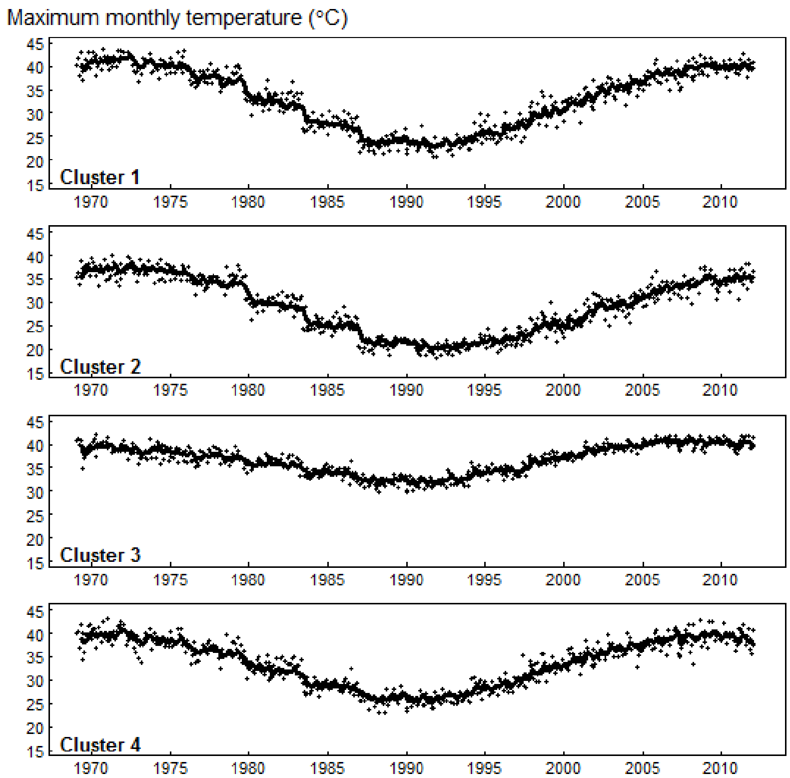


Figure 4.12: The fitted values of maximum monthly temperatures in each cluster based on model (equation 2.8).

In Figure 4.12, the dots represent the maximum monthly temperatures in each cluster. The temperatures decreased from 1970 to 1990 and then increased until year 2012. The temperatures ranged from 20˚C to 45˚C over the 43years. The curve line is the fitted values from the quartic trend model with 3rd order time lag (equation 2.8, page 27).

After fitting a quartic trend model with 3rd order time lag in each cluster, it is necessary to plot the residuals to check that the assumptions of a normal distribution and independence of this model have been satisfied. The residuals from the model were plotted with normal quantiles to confirm a normal distribution as shown in Figure 4.13

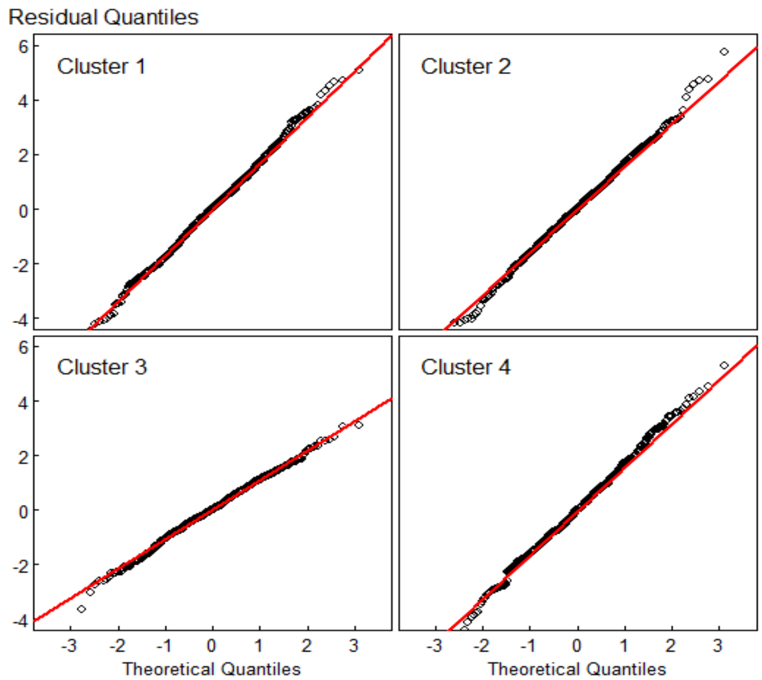


Figure 4.13: Residual Q-Q plots (quantiles plots) for each cluster

Figure 4.13, the linearity of the residual in each of the four clusters suggests that the residuals are normally distributed, because they lie along a straight line.

In addition, the residuals were also plotted for auto correlation to confirm the independence among residuals as shown in Figure 4.14.

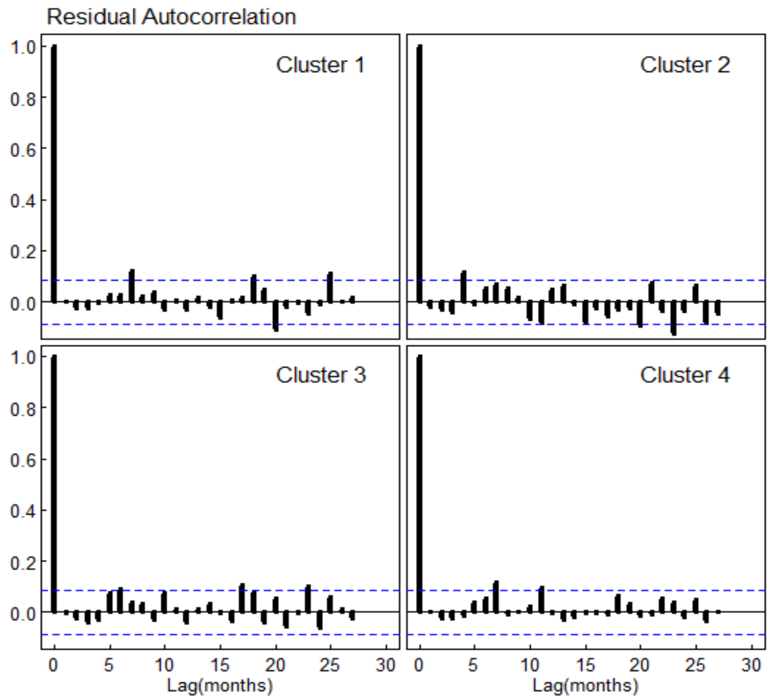


Figure 4.14: Residual auto-correlation function (ACF) plots for each cluster

Figure 4.14 shows that there are no auto-correlations, because the correlations in all clusters stay within the 95% confidence interval of zero correlation. That is, the residuals are independent.

Since the model was successfully examined for the adequacy of these assumptions, this model can be used to estimate the data patterns and trends as shown in the results of the regression coefficients for each cluster in Table 4.7.

	Coefficients			
	Cluster1	**Cluster2**	**Cluster3**	**Cluster4**
Intercept	17.9900	16.8900	15.5100	19.7000
t	0.0647	0.0663	0.0293	0.0498
t^2	-0.0010	-0.0010	-0.0004	-0.0009
t^3	3.36×10^{6}	3.04×10^{6}	1.51×10^{6}	2.98×10^{6}
t^4	-3.19×10^{9}	-2.78×10^{9}	-1.49×10^{9}	-2.91×10^{9}
y_{t-1}	0.1842	0.1545	0.1499	0.1631
y_{t-2}	0.1932	0.1850	0.2597	0.2208
y_{t-3}	0.1615	0.1777	0.1800	0.1000
Adjusted R^2	0.9287	0.9251	0.8694	0.8924

Table 4.7: Regression coefficients for each cluster

Table 4.7 shows the regression coefficients in each cluster. The coefficients indicate that the regression lines almost perfectly fit the data, because adjusted R^2 are close to 1. The adjusted R^2 are 0.9287, 0.9251, 0.8694 and 0.8924 in cluster 1, 2, 3 and 4, respectively.

Figure 4.12 indicates that the trends of temperature changes in each cluster occur over 10 year periods. This means that there will be 4 different forms of curve during 10 years. Therefore, the most accurate forecast that can be predicted by this model will be a short-term forecast. In this case, the forecast of monthly maximum temperatures is from the year 2013 up to 2015 or 36 months as shown in Figure 4.15.

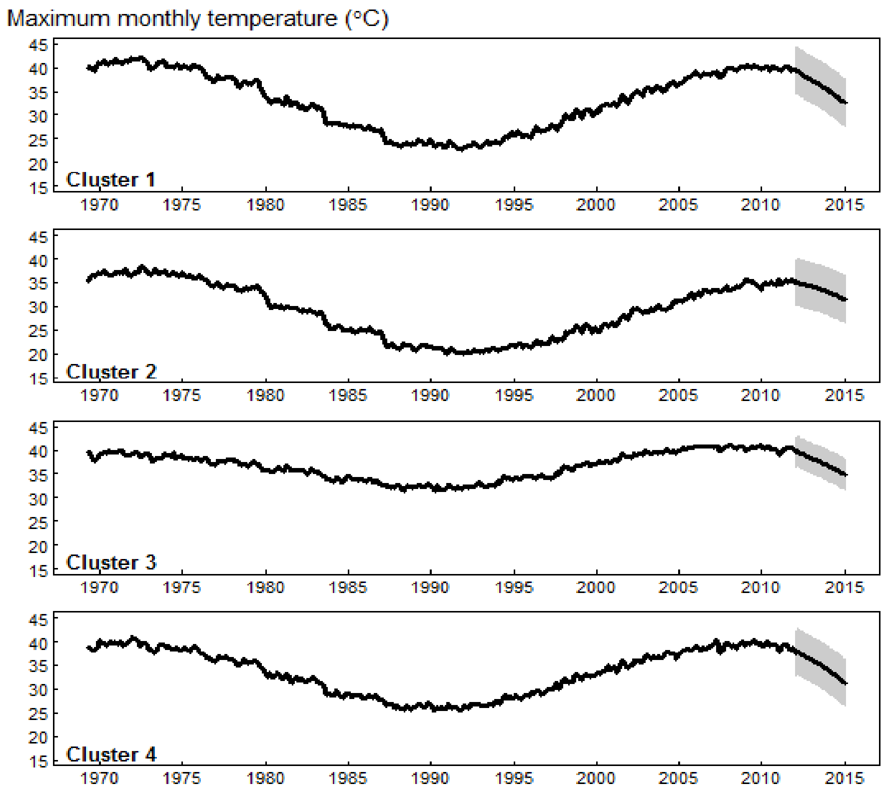


Figure 4.15: Forecasting the maximum monthly temperature from 2013 to 2015. The curved lines are trend estimations based on model (equation 2.10). The highlighted lines are the 95% confidence interval of prediction.

In Figure 4.15, the forecasted temperatures gradually decrease during the 2013-2015 period in each cluster as shown in the highlighted curved line. The highlight shows the 95% confidence interval (CI) of the maximum monthly temperatures predictions for the 3 years in each cluster. The 95% CI ranges from 27˚C - 44˚C, 26˚C - 40˚C, 31˚C - 42˚C and 26 ˚C - 42˚C in cluster 1, 2, 3 and 4, respectively. The confidence interval of cluster 3 is narrower than for the other clusters, and the temperature changes are higher than for other clusters (ranging from 32˚C - 43˚C) which occurred in the north. The highest predicted temperature of approximately 45˚C occurs in cluster 1 which is in a desert area.

Maximum monthly temperatures from January 2013 to September 2015 were compared with forecasted temperatures in order to check the accuracy of the model as shown in Figure 4.16.

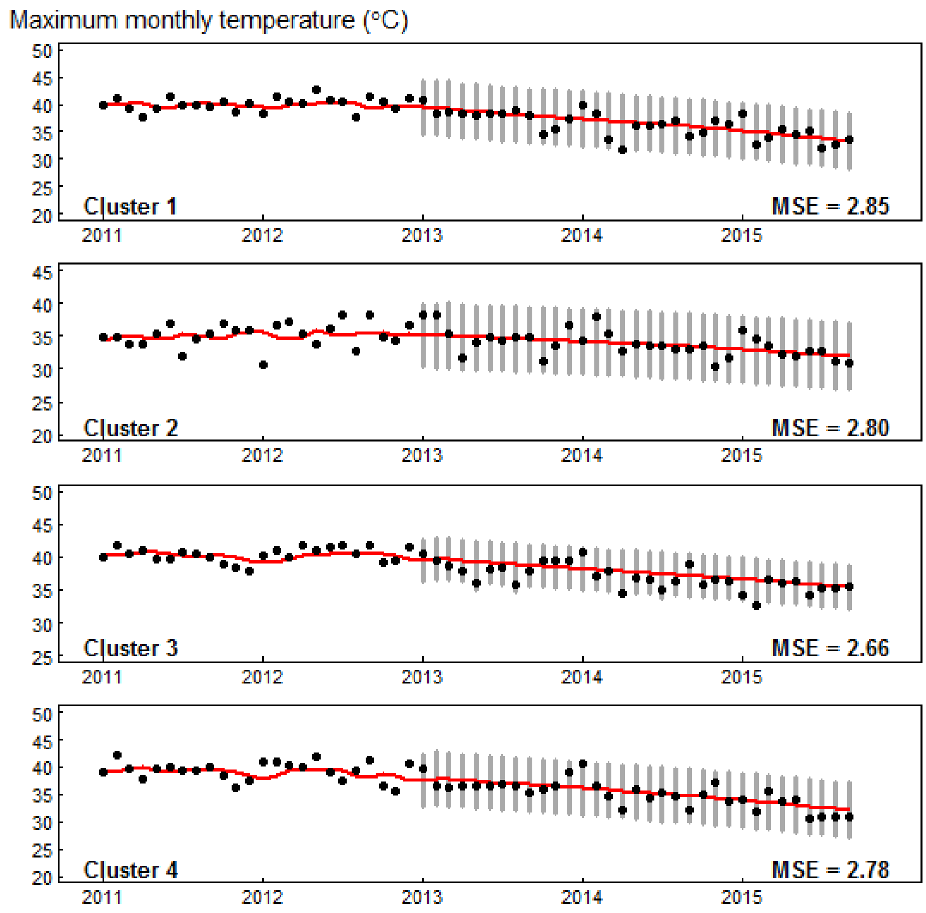


Figure 4.16: Forecasting the maximum monthly temperature from January 2013 to September 2015. The dots are maximum monthly temperature. The curved lines are trend estimations based on the model (equation 2.8). The highlighted lines are the 95% confidence interval of prediction.

In Figure 4.16, the dots represent the maximum monthly temperatures which are within a 95% confidence interval of the maximum monthly temperature predictions for 33 months in each cluster. Mean square errors of forecasting are 2.85, 2.80, 2.66 and 2.78 in cluster 1, 2, 3 and 4, respectively. Therefore, it is confirmed that this model can accurately forecast temperatures for short time periods.

4.2.3 Conclusions

The maximum monthly temperature data of 85 Australian meteorological stations from 1970 to 2012 were investigated. Missing values were estimated by a linear regression model to account for information from the nearest stations as well as time periods.

Factor analysis was applied to group the numbers of stations, but there were some stations having similar factor loadings. Therefore, the results of factor loading values were used as the input for cluster analysis. Cluster analysis grouped the number of station from 85 stations into 4 clusters corresponding to geographical regions. The first cluster consisted of 39 stations located in the northwest and some parts of the northern, central, southern and southeastern regions of Australia. The second cluster consisted of 14 stations distributed within the boundary of the southwestern and southern regions of Australia. The third cluster consisted of 12 stations located in northern Australia. The fourth cluster consisted of 20 stations distributed in parts of central and eastern of Australia. The highest temperatures that occurred were approximately 45˚C in cluster 1, which is in a desert area.

A quartic trend model with 3rd order time lag exhibited high goodness of fit the data in each cluster. In addition, this model can be used to forecast over relatively short time periods. The results of forecasting for the 3-year time period of 2013-2015 showed a maximum monthly 95% confidence interval of 27˚C - 44˚C, 26˚C - 40˚C, 31˚C - 42˚C and 26˚C - 42˚C in clusters 1, 2, 3 and 4, respectively.

CHAPTER 5

Conclusions and Discussions

This chapter consists of summaries of the main findings based on the application of statistical methods used to explain the trends and patterns of temperature change in the Arctic and Australia. Discussions of the findings, limitations and recommendations are also presented.

5.1 Conclusions

In the Arctic

The trends and patterns of monthly averaged temperature anomalies were investigated for the period of 1973 through 2008 (432 months) on grid-boxes of the earth's surface above 45° North, including both land and sea. This area has 648 grid-boxes which cover the Arctic Ocean, northern areas of the Atlantic and Pacific Oceans and the Asian and European Continents. The 648 grid-boxes were combined to form 69 sub-regions in order to better work through issues of missing data. Statistical techniques were applied to determine the trends, patterns and extent of temperature change in the large Arctic region.

Separate linear regression models were first fitted to the 69 sub-regions to determine the trends of temperature change. Of the 69 sub-regions, the temperatures increased in all but two sub-regions which had insufficient data for analysis and in three sub-regions which showed no evidence of temperature change. Auto regressive process and filtering were applied to reduce time correlations. Spatial correlations also were considered because similar temperature changes occurred in each sub-region. Classification of sub-regions with similar temperature changes was analyzed by factor analysis.

Twelve factors were identified from factor analysis. Each of these factors was classified into three levels based on the lower bound of the 95% CI of temperature change per decade. Three different temperature levels were indicated. Larger temperature increases occurred in the northern Pacific Ocean, Alaska and eastern Siberia (0.200^{o}C to 0.320^{o}C). Increases in the temperatures of northern Canada, Greenland, Iceland, Norway, Sweden and Finland were moderate (0.130^{o}C to 0.199^{o}C). Small increases occurred in northern Siberia and part of the northern Atlantic Ocean (0.090^{o}C to 0.129^{o}C). Almost all sub-regions experienced some level of temperature increase. An exception to the trend of increased temperatures occurred in northeast Canada and its surrounding seas which did not experience much change.

In Australia

Maximum temperatures over consecutive 5-day periods data and the maximum monthly temperatures data were studied from 1970 to 2012 for a random sample of 85 meteorological reporting stations. These stations were spread approximately evenly over Australia and had 80% complete data. Missing data were estimated using linear interpolation. This study was conducted using statistical methods including linear regression model, polynomial regression model, time series analysis, factor analysis, cluster analysis and forecasting to investigate the characteristics and classification of temperature variability.

Analysis of maximum temperatures over consecutive 5-day periods indicated that there were different patterns of temperature change in the seven factors corresponding to seven geographic regions in Australia. Average maximum temperatures ranged from 23^{o}C to 36^{o}C. The highest average temperature, approximately 36^{o}C, occurred in the desert area of the northwest in 2002. Similar temperature trends occurred in the central, eastern, southern and southeastern parts of Australia.

For the maximum monthly temperature data, 4 clusters were found to have different patterns of temperature change. The highest temperatures, approximately 45^{o}C, occurred in the desert area of the northwest and some parts of the central, southern and southeastern regions. In addition, a quartic trend model combined with a 3rd order time lag fitted the data well for each cluster. This model can be used to forecast maximum monthly temperature for the 3-year period of 2013-2015. The forecasted maximum monthly temperatures decreased during this period. The 95% CI of the maximum monthly temperatures ranged from about 27^{o}C-44^{o}C, 26^{o}C-40^{o}C, 31^{o}C-42^{o}C and 26^{o}C-42^{o}C for clusters 1, 2, 3 and 4, respectively.

5.2 Discussions

In this study, temperature change in the Arctic region of the Northern Hemisphere showed a temperature increase almost everywhere, which is consistent with the findings of Jones *et al*., 1999 that temperatures increased on average by 0.37^{o}C from 1925-1944 and 0.32^{o}C over 1978-1997. Temperature increase in the Arctic was investigated in the same region during the same period and in similar periods by Rigor *et al*., 1999 and Polyakov *et al*., 2002. They found that both land and sea parts of the northern Atlantic Arctic region experienced temperature increases. Later research by McNeil and Chooprateep (2014) reported similar findings of rising temperatures in their study of sea surface temperatures on the northern Atlantic Ocean. The rate of temperature increase is greater in the polar region than in lower latitudes, which is then manifested in the variation of surface air temperatures between the Arctic and northern Atlantic regions (Polyakov *et al*., 2002). Moreover, Delworth and Knutson, (2000) and Mann (2000) suggest that variability in the North Atlantic is influenced by slow change in oceanic thermohaline circulation.

In the analysis of temperature change in Australia, monthly maximum temperatures showed a polynomial pattern with a range between 20^{o}C and 50^{o}C. This result was similar to the general pattern of maximum temperatures over consecutive 5-day periods, but it differed in the range of maximum temperatures, which was between 23^{o}C and 36^{o}C. These differences have occurred because different times were used. In addition, the trends and patterns of temperature change of maximum temperatures over consecutive 5-day periods in the central, eastern, southern and southeastern regions were similar within regions when grouped into 4 clusters corresponding to these geographic regions. Temperature change in these clusters is influenced by reduced rainfall and duration of droughts since 1973, with increased temperatures during 2002 and 2003 in the southeastern region (Deo *et al*., 2009; Nicholls, 2004). The eastern region experienced increased temperatures by approximately 2^{o}C to2.5^{o}C during 1982-1983 and 2002-2003 in the summer periods with decreased rainfall in this period. Central New South Wales and northern Victoria experienced increased rainfall by about 15% and found of droughts about 3-5 days for the period 1951-2003 (Deo *et al*., 2009). Moreover, there was a decrease in the Pacific trade winds and a warming of the central and eastern tropical Pacific Ocean. (Nicholls *et al*., 2004; Alexander *et al*., 2007; Murphy and Timbal, 2007; Deo *et al*., 2009).

The methodological approach used in the present study can explain the variability of temperature on the Earth's surface by using simple linear regression modeling as the first step in identifying trends (Nicholls *et al*., 2004), but may not be the best way to estimate trends for shorter periods (Benestad, 2003). Temperature change showed periodic variations. A polynomial regression model was considered to account for the detailed description of the trend within a calibration interval (Benestad, 2003). However, this model is not suitable for forecasting long-term trends because of uncertainly with other factors in the future. Predicting regional climate change with uncertainly runs the risk of both over- and under-estimation. Therefore, the unpredictability of the climate and global systems should be a concern in further studies (Mitchell and Hulme, 1999). Moreover, the regional patterns of temperature change in the Arctic and Australia were classified by factor analysis and cluster analysis. This procedure offers the occasion to define regions with the variable of interest having similar values (Mahlstein and Knutti, 2009), so this method can be applied to identify spatial correlation, combining small surface areas of a region into larger sub-regional groups. This statistical technique was also used in other studies as mentioned in Unal *et al*., (2003), Mahlatein and Knutti (2009) and McNeil and Chooprateep (2014), but the results from the classification grouping were different because of the different time periods and regions studied.

5.3 Recommendations and Further study

The recommendations and further study are as follows:

5.3.1. The data sets of this study have many missing data, whether the Arctic region or Australia. Although the grid-boxes were combined to reduce these missing data and missing data were also interpolated by a multiple linear regression model, the data sets still lost information and may be not a good representation. Therefore, the evaluate performance of other methods for missing data estimation such as artificial neural network (ANN), optimized regression analysis and cubic spline methods (CDP) should be considered in future studies.

5.3.2 In this study, identification of the order of the autoregressive model (AR) was only considered from an auto correlation function plot (ACF) which was insufficient to confirm the order of the autoregressive model (AR). In some cases, the auto correlations were significant for a large number of lags. Perhaps the auto correlations at lag 2 and above are merely due to the propagation of the auto correlation at lag 1. Therefore, the order of autoregressive model (AR) is needed to correct any auto correlation that remains in the differenced series by looking at both the auto correlation function plot (ACF) and the partial auto correlation function plot (PACF) of the differenced series in further study.

5.3.3 For the study of spatial correlation, factor analysis was used to group similar patterns in the region into larger regional groupings by considering factor loading values. Some cases may have had similar factor loading values in each factor which were not easy to group in these regions. Although cluster analysis was used to help create geographic groupings in this study, there are many other methods that could be applied to compare and efficiently classify groupings such as principle component analysis, T-mode PCA with k-mean clustering and Ward's method, empirical orthogonal function analysis and functional analysis.

5.3.4 In this study, the data showed a periodic pattern of temperature change. A polynomial regression model was fitted to explain trends and patterns of temperature change, but this model is not the best model for forecasting. Although the quartic trend model combined with a 3^{nd}order time lag was fitted well with the data but it could not forecast for longer periods. However, there are many statistical methods for forecasting periodic patterns of temperature change such as periodogram analysis, sine and cosine function, harmonic function, spline regression models and locally weight regression model to compare the results with further study.

5.3.5 The results of the present research provide comprehensive information about temperature changes on land and sea surfaces in the Arctic region and Australia which are the most important areas of global warming. This study focused on only temperature data without considering other factors affecting the fluctuation of temperature. Thus, additional factors affecting climate such as distance from the sea, ocean currents, and direction of prevailing winds, the EI Niño phenomenon and human activities should be investigated. The methodological approach presented in this research can be applied to the studies of other factors such as rainfall, wind, and solar radiation in future research.

References

ACIA. 2004. Impacts at a Warming Arctic: Arctic Climate Impact Assessment, ACIA Overview Report, Cambridge University.

Alexander, L. V., Hope, P., Collins, D., Trewin, B., Lynch, A., and Nicholls, N. 2007. Trends in Australia's climate means and extremes: a global context. Australian Meteorology Magazine. 56: 1-18.

Anisimov, O. A., Lobanov, V. A., and Reneva, S. A. 2007. Analysis of changes in air temperature in Russia and Empirical forecast for the first quarter of the 21st century. Russian Meteorology and Hydrology. 32(10): 620-626.

Benestad, R. 2003. What can present climate models tell us about climate change?. Climatic change. 59: 311-331.

BOM. 2013. Bureau of Meteorology, http://www.bom.gov.au. (Accessed 1 July 2013).

Box, J. E., Yang, L., Bromwich, D. H., and Bai, L. S. 2009. Greenland ice sheet surface air temperature variability: 1840-2007. America Meteorological Society. 22: 4029-4049.

Brockwell, P. J., and Davis, R. A. 2002. Introduction to Time Series and Forecasting, Springer, USA.

Brohan, P., Kennedy, I. J., Harris, I., Tett, S. F. B., and Jones, P. D. 2006. Uncertainty estimate in regional and global observed temperature changes: A new data set from 1850. Geophysical research. 11: D12106. Doi: 10.1029.

Chatfield, C. 1996. The Analysis of Time series, Melbourne: Chapman & Hall.

Chooprateep, S., and McNeil, N. 2014. Temperature change of Southeast Asia: 1973-2008. Chiang Mai University Journal of National Science. 13(2): 105-116.

Collins, D. 2000. Annual temperature summary: Australia records warmest decade. Climate Change Newsletter. 12(6).

Collins, D. A., Della-Marta, P. M., Plummer, N., and Trewin, B. C. 2000. Trends in annual frequencies of extreme temperature events in Australia. Australian Meteorology Magazine. 49: 277-292.

CRU. 2009. Climate Research Unit, http://www.cru.uea.ac.uk/cru/data/temperature. (Accessed April 1, 2010).

Cryer, J. D. 1986. Time series analysis, United States. R. R., Donnelley & Sons Company, USA.

Delworth, T. L., and Knutson, T. R. 2000. Simulation of the early 20th century global warming. Science. 287: 2246-2250.

Deo, R. C., Syktus, J. I, McAlpine, C. A., Lawrence, P. J., McGowan, H. A., and Phinn, S. R. 2009. Impact of historical land cover change on daily indices of climate extremes including droughts in eastern Australia. Geophysical Research Letters. 36: L08705

Donaldson, G. C., Keatinge, W. R., and Nayha, S. 2002. Change in summer temperature and heat-related mortality since 1971 in North Carolina South Finland and Southeast England. Environmental Research. 91: 1-7.

Grainger, S., and Frederiksen, C. S. 2008. Estimating the potential predictability of Australia surface maximum and minimum temperature. Center for Australia Weather and Climate Research. 32: 443-445.

Griffiths, G. M., Chambers, L. E., Haylock, M. R., Manton, M. J., Nicholls, N., Baek, H. J., Choi, Y., Della-Marta, P. M., Gosai, A., Iga, N., Lata, R., Laurent, V., Maitrepierre, L., Nakamigawa, H., Ouprasitwong, N., Solofa, D., Tahani, L., Thuy, D. T., Tibig, L., Trewin, B., Vediapan, K., and Zhai, P. 2005. Change in mean temperature as a predictor of extreme temperature change in the Asia-Pacific region. Journal of Climatology. 25: 1301-1330.

Hair, J. F., Anderson, R. E., Tatham, R. L., and Black, W. C. 1998. Multivariate Data Analysis, Prentice Hall International (UK) Limited, London.

Hartmann, B., and Wendler, G. 2005. The significance of the 1976 Pacific climate shift in the climatology of Alaska. Journal of Climatology. 18: 4824-4839.

Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., Van der Linden, P. J., Dai, X., Maskell, K., and Johnson, C. A. 2001. Climate Change 2001: The Scientific Basic, Cambridge University Press, United Kingdom.

Hughes, L. 2003. Climate change and Australia: trend, projections and impacts. Austral Ecology. 28: 423-443.

Hughes, G. L., Rao, S. S., and Rao, T. S. 2006. Statistical analysis and time-series models for minimum/maximum temperatures in the Antarctic Peninsula. The Royal Society. 463: 241-259.

Husson, F., Le, S., and Pages, J. 2011. Exploratory multivariate analysis by example using R, Chapman & Hall/CRC, London.

IPCC. 2007. Intergovernmental Panel on Climate Change, http://www.ipcc.ch. (Accessed 1 July, 2013).

Johannessen, O. M., Bengtsson, L., Miles, M. W., Kuzmina, S. I., Semenov, V. A., Alekseev, G. V., Nagurnyi, A. P., Zakharov, V. F., Bobylev, L. P., Pettersson, L. H., Hasselmann, K., and Cattle, H. P. 2003. Arctic climate change: observe and modeled temperature and sea-ice variability. Tellus. 56: 328-341.

Johnson, D. E. 1998. Applied Multivariate Methods for Data Analysts, Duxbury, U.S.A.

Jones, D. A. 1999. Characteristics of Australia land surface temperature variability. Theoretical and Appied Climatology. 63: 11-31.

Jones, P. D., New, M., Parker, D. E., Martin, S., and Rijor, I. G. 1999. Surface air temperature and its changes over the past 150 years. Reviews of Geophysics. 37: 173-199.

Karoly, D. J., and Braganza, K. 2004. Attribution of Recent Temperature changes in the Australian Region. American Meteorological Society. 18: 457-464.

Kedem, B., and Fokianos, K. 2002. Regression Model for Time series Analysis, Willy, Canada.

Krupnik, I., and Jolly, D. 2002. The Earth is Faster Now. Arctic Research Consortium, Fairbanks, AK.

Lean, J .L., and Rind, D. H. 2009. How will Earth's surface temperature change in future decades?. Geophysical Research Letters. 36: L15708.

Mahlstein, I., and knutti, R. 2009. Regional climate change patterns identified by cluster analysis. Climate Dynamics. 35: 587-600.

Maiyza, I. A., Said, M. A., and Kamel, M. S. 2009. Sea surface temperature anomalies in the south eastern Mediterranean sea. Journal of the Faculty of Marine Science. 21: 151-159.

Mann, M. E. 2000. Observed and simulated multidecadal variability in the Northern Hemisphere. Climate Dynamics. 16: 661-676.

McNeil, N., and Chooprateep, S. 2014. Modeling sea surface temperatures of the North Atlantic ocean. Theoretical and Applied Climatology. 116: 11-17.

Mitchell, T. D., and Hulme, M. 1999. Predicting regional climate change: living with uncertainty. Progress in Physical Geography. 23: 57-78.

Montgomery, C. M., Jennings, C. L., and Kulahci, M. 2008. Introduction to Time Series Analysis and Forecasting, John Willey and Sons, New Jersey.

Murphy, B. F., and Timbal, B. 2007. A review of recent climate variability and climate change in south eastern Australia. International Journal of Climatology. 28: 859-879.

NASA Goddard Institute for Space Studies. 2010. GISS Surface Temperature Analysis. (Accessed November 9, 2013).

National Academies. 2008. Understanding and Responding to Climate Change. Highlights of National Academies Report.

Neter, J., Kuther, M. H., Nachtsheim, C. J., and Wasserman, W. 1996. Applied Linear Statistical Models, Times Mirror Higher Education Group, U.S.A.

Nicholls, N. 2004. The changing nature of Australian droughts. Climatic change. 63: 323-336

Nicholls, N., Della-Marta, D., and Collins, D. 2004. 20th century changes in temperature and rainfall in New South Wales. Australian Meteorology Magazine. 53: 263-268.

NOAA. 2014. National Oceanic and Atmospheric Administration, http://icoads.noaa.gov. (Accessed November 9, 2013).

Overpeck, J., Hughen, K., Hardy, D., Bradley, R., Case, R., Douglas, M., Finney, B., Gajewski, K., Jacoby, G., Jennings, A., Lamoureux, S., Lasca, A., MacDonale, G., Moore, J., Retelle, M., Smith, S., Wolfe, A., and Zielinski, G. 1997. Arctic environmental change of the last four centuries. Science. 278: 1251-1256.

Polyakov, V. I., Brkryaev, R. V., Bhatt, G. V. U. S., Colony, R.L., Johnson, M. A., Maskshtas, A. P., Walash, D. 2002. Variability and Trends of Air Temperature and Pressure in the Maritime Arctic, 1875-2000. American Meteorological Society. 16 : 2067-2076.

Przybylak, R. 1996. Spatial and temporal changes in extreme air temperatures in the Arctic over the period 1951-1990. International Journal of Climatology. 17: 615-634.

R Development Core Team. 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Rigor, I. G., Colony, R. L., and Martin, S. 1999. Variation in surface air temperature observations in the Arctic, 1979-97. Journal of Climate. 13: 896-914.

Semenov, V. A. 2007. Structure of temperature variability in the high latitudes of the northern hemisphere. Atmospheric and Oceanic Physics. 43: 744-753.

Suppiah, R., Collins, D., and Della-Marta, P. 2001. Observed changes in Australia climate, http://www.dar.csiro.au/publications/Suppiah_2001a.pdf. (Accessed November 9, 2013).

Taniguchi, M., Uemura, T., and Jago-on, K. 2007. Combined effect of urbanization and global warming on subsurface temperature in four Asian cities. Vadose Zone Journal. 6: 591-596.

Torok, S. J., and Nicholls, N. 1996. A historical annual temperature dataset for Australia. Bureau of Meteorology Research Center. 45: 251-260.

Unal, Y., Kinda, T., and Karaca, M. 2003. Redefining the climate zone of turkey using cluster analysis. International Journal of Climaology. 23: 1045-1055.

Venables, W. N., and Ripley, B. D. 2002. Modern Applied Statistics with S, Springer, Queenland.

Wanishsakpong, W., Luo, K., and Tongkumchum, P. 2014. Earth surface temperature change above latitude 45 degrees North from 1973 to 2008. Chiang Mai University Journal of Natural Science. 13(3): 247-257.

Wanishsakpong, W., and McNeil, N. 2015. Modeling of daily maximum temperatures over Australia from 1970 to 2012. Metrological Application. (Accepted from publication).

Wanishsakpong, W., Notodiputro, K. N., and McNeil, N. 2015. Clustering and Forecasting Maximum Temperature of Australia. Chiang Mai University Journal of Science. (Accepted from publication).

WMO. 2011. WMO statement on the status of the global climate in 2010. World Meteorological Organization, Switzerland.

Wang, X., Key, J., Fowler, C., Maslanik, J., and Tschudi, M. 2012. Arctic Climate Variability and Trends from Satellite Observations. Advances in Meteorology. DOI: 10.1155/2012/505613.

Appendix 1

Regression coefficients of 67 sub-regions.

Sub-region	Latitude and longitude	b_{1i}	P-value
3	75-85^{o}N – 135-90^{o}W	0.673	< 0.05
4	75-85^{o}N – 90-45^{o}W	0.717	< 0.05
5	75-85^{o}N – 45-0^{o}W	0.447	< 0.05
6	75-85^{o}N – 0-45^{o}E	0.272	< 0.05
7	75-85^{o}N – 45-90^{o}E	0.670	< 0.05
8	75-85^{o}N – 90-135^{o}E	0.418	0.001
9	75-85^{o}N – 135-180^{o}E	0.406	< 0.05
10	65-75^{o}N – 180-150^{o}W	0.560	< 0.05
11	65-75^{o}N – 150-120^{o}W	0.733	< 0.05
12	65-75^{o}N – 120-90^{o}W	0.412	< 0.05
13	65-75^{o}N – 90-60^{o}W	0.391	0.001
14	65-75^{o}N – 60-30^{o}W	0.231	< 0.05
15	65-75^{o}N – 30-0^{o}W	0.312	< 0.05
16	65-75^{o}N – 0-30^{o}E	0.288	< 0.05
17	65-75^{o}N – 30-60^{o}E	0.209	< 0.05
18	65-75^{o}N – 60-90^{o}E	0.472	< 0.05
19	65-75^{o}N – 90-120^{o}E	0.535	< 0.05
20	65-75^{o}N – 120-150^{o}E	0.441	< 0.05
21	65-75^{o}N – 150-180^{o}E	0.464	< 0.05
22	55-65^{o}N – 180-165^{o}W	0.200	< 0.05
23	55-65^{o}N – 165-150^{o}W	0.165	0.019
24	55-65^{o}N – 150-135^{o}W	0.124	0.031
25	55-65^{o}N – 135-120^{o}W	0.167	0.116
26	55-65^{o}N – 120-105^{o}W	0.246	0.065
27	55-65^{o}N – 105-90^{o}W	0.362	0.002
28	55-65^{o}N – 90-75^{o}W	0.478	< 0.05
29	55-65^{o}N – 75-60^{o}W	0.349	< 0.05
30	55-65^{o}N – 60-45^{o}W	0.251	< 0.05
31	55-65^{o}N – 45-30^{o}W	0.353	< 0.05
32	55-65^{o}N – 30-15^{o}W	0.338	< 0.05
33	55-65^{o}N – 15-0^{o}W	0.239	< 0.05
34	55-65^{o}N – 0-15^{o}E	0.349	< 0.05
35	55-65^{o}N – 15-30^{o}E	0.401	< 0.05

Sub-region	Latitude and longitude	b_{1i}	P-value
36	55-65°N – 30-45°E	0.402	0.01
37	55-65°N – 45-60°E	0.405	0.001
38	55-65°N – 60-75°E	0.384	0.003
39	55-65°N – 75-90°E	0.492	0.001
40	55-65°N – 90-105°E	0.448	< 0.05
41	55-65°N – 105-120°E	0.441	0.001
42	55-65°N – 120-135°E	0.478	< 0.05
43	55-65°N – 135-150°E	0.305	< 0.05
44	55-65°N – 150-165°E	0.228	< 0.05
45	55-65°N – 165-180°E	0.322	< 0.05
46	45-55°N – 175-160°W	0.149	< 0.05
47	45-55°N – 160-145°W	0.133	< 0.05
48	45-55°N – 145-130°W	0.128	< 0.05
49	45-55°N – 130-115°W	0.176	0.001
50	45-55°N – 115-100°W	0.232	0.051
51	45-55°N – 100-85°W	0.259	0.003
52	45-55°N – 85-70°W	0.023	0.764
53	45-55°N – 70-55°W	0.372	< 0.05
54	45-55°N – 55-40°W	0.321	< 0.05
55	45-55°N – 40-25°W	0.314	< 0.05
56	45-55°N – 25-10°W	0.303	< 0.05
57	45-55°N – 10-5°E	0.373	< 0.05
58	45-55°N – 5-20°W	0.448	< 0.05
59	45-55°N – 20-35°E	0.356	< 0.05
60	45-55°N – 35-50°E	0.368	0.001
61	45-55°N – 50-65°E	0.411	0.001
62	45-55°N – 65-80°E	0.413	< 0.05
63	45-55°N – 80-95°E	0.461	< 0.05
64	45-55°N – 95-110°W	0.505	< 0.05
65	45-55°N – 110-125°E	0.488	< 0.05
66	45-55°N – 125-140°E	0.356	< 0.05
67	45-55°N – 140-155°E	0.210	< 0.05
68	45-55°N – 155-170°E	0.170	< 0.05
69	45-55°N – 170-175°E	0.159	< 0.05

DOI: 10.12982/CMUJNS.2014.0034

CMU J. Nat. Sci. (2014) Vol. 13(3)

Appendix 2

Earth Surface Temperature Changes above Latitude 45 Degrees North from 1973 to 2008

Wandee Wanishsakpong[1*], Kehui Luo[2] and Phattrawan Tongkumchum[1]

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

[2]*Department of Statistics, Macquarie University, NSW 2109, Australia.*

**Corresponding author. E-mail: one_d7@hotmail.com*

ABSTRACT

In this study, we examined monthly temperature variation from 1973 to 2008 on grid regions of the earth surface above latitude 45˚North, covering the Arctic Ocean, northern areas of the Atlantic and Pacific Oceans, and the Asian and European continent. Linear modelling was used to investigate the trends and patterns of temperature changes and account for auto-correlations of the temperature changes over time. Factor analysis was then used to model filtered residuals, i.e., the residuals after removing time trends and auto-correlation, providing a basis for identifying and classifying regions with similar temperature change. Twelve large regions, each having similar temperature change patterns,

were identified. Of the 69 sub-regions considered in the study, 64 sub-regions experienced significant increase in temperature, 2 sub- regions had insufficient data, and only 3 sub-regions remained unchanged. High temperature increases (0.200˚Cto 0.320˚C) occurred in the North Pacific Ocean, Alaska and Eastern Siberia. Moderate temperature increases (0.130˚C to 0.199˚C) occurred in north Canada, Greenland, Iceland, Norway, Sweden and Finland. The north of Siberia and part of the North Atlantic had low increases (0.090˚C to 0.129˚C) while north east Canada and its surrounding seas did not show evidence of warming.

Key words: Latitude, Climate change, Time series analysis, Correlation, Auto-correlation, Factor analysis.

INTRODUCTION

Earth surface temperature change is one of the most important issues the world faces today. Global surface temperature has changed over the past 150 years with a slightly higher rate of warming in the 20th century (Jones et al., 1999). This warming is associated with change in sea levels, destruction of ecosystems, shrinkage of mountain glaciers, reduction of ice cover (National Academies Report, 2008) and altered ocean circulation patterns (Houghton et al., 2001).

Increased surface temperature in the Arctic Ocean has been a major topic of international reviews and indigenous observations during last decade (Krupnik and Jolly, 2012). According to the study by Overpeck et al., (1997), the average temperature in the Arctic increased about 0.6˚C from the beginning of the 20th century and the maximum temperatures there have increased approximately 1.2˚C since 1945. The temperature change was about +1˚C per decade (increase) in the eastern Arctic Ocean and -1˚C per decade (decrease) in the western Arctic Ocean during the winter season. Furthermore, significant warming in spring has been detected across most of the Arctic region (Rigor et al., 1999).

Various studies have assessed temperature changes over the Earth's surface. For example, Box et al., (2007) studied temperature changes in Greenland over period 1840-2007. The annual warming trend in 1919-1932 was 33% greater than in 1994-2007 and the recent warming was high in western Greenland during autumn and southern Greenland in winter. In addition, Anisimov et al., (2007) investigated the changes of air temperature in Russia over the periods of 1900-2004. The trend of annual average temperature was 0.5˚C per 100 years in the north of European Russia, 1.4˚C-1.6˚C per 100 years in the south of Ural Siberia and the Far East. On average for the entire territory of Russia the trend was 1.1˚C per 100 years.

It is not easy to understand and explain the variability and trends of change in Earth surface temperatures or quantify the size and speed of the rate of change. Many scientists and researchers have studied patterns of Earth surface temperature changes in different regions of the world using various methodologies, including computer simulation models (Johannessen et al., 2003), empirical orthogonal functions (Semenov, 2007) and statistical techniques, such as multiple linear regression (Lean and Rind, 2009), multiple regression with non-Gaussian correlated errors and linear autoregressive moving average (ARMA) model (Hughes et al., 2006) and linear spline functions and factor analysis (McNeil and Chooprateep, 2013).

An investigation of temperature changes a region covering northern Atlantic Ocean was carried out in a more recent study by McNeil and Chooprateep (2013). The present study examines the region centered around middle part of the Northern hemisphere on Earth surface for temperature changes, using a combination of statistical techniques including linear regression allowing for auto-correlation (Venables and Ripley, 2002) and factor analysis (Johnson, 1998).We examined and quantified the trend and patterns of the temperature changes across the Earth surface above 45 degrees north from 1973 to 2008, based on monthly temperature data. The area studied includes both land and sea.

MATERIALS AND METHODS

Temperature data from 1973 to 2008 for the study area were obtained from the Climatic Research Unit at the University of East Anglia, UK (CRU, 2009). They include 432 monthly average temperature anomalies (defined as excesses over monthly averages for 1961-1990; CRU, 2009) for all 5˚ by 5˚ latitude- longitude grid-boxes on the Earth's surface studied, and were collected from weather stations, ships, and more recently satellites. This area comprises 648 5˚by 5˚grid-boxes above latitude 45 degrees North, including the Arctic Ocean, the northern areas of the Atlantic and Pacific Oceans, the North America, Asian and European continents. However temperature data were missing in a number of those grid-boxes, particularly in the polar zones. Therefore, the 648 grid-boxes were combined into 69 sub-regions

CMU J. Nat. Sci. (2014) Vol. 13(3)

of similar size (approximately 0.45 km^2) for each region, following a design in the shape of an igloo. These 69 sub-regions include one covering up to 5˚ below the North Pole (85˚N-90˚N), eight in the 75˚N-85˚N band, 12 in the 65˚N-75˚N band, 24 in the 55˚N-65˚N band and another 24 in the 45˚N-55˚N band (Figure 1). A number of those sub-regions have missing data, as shown in brackets in Figure 1.

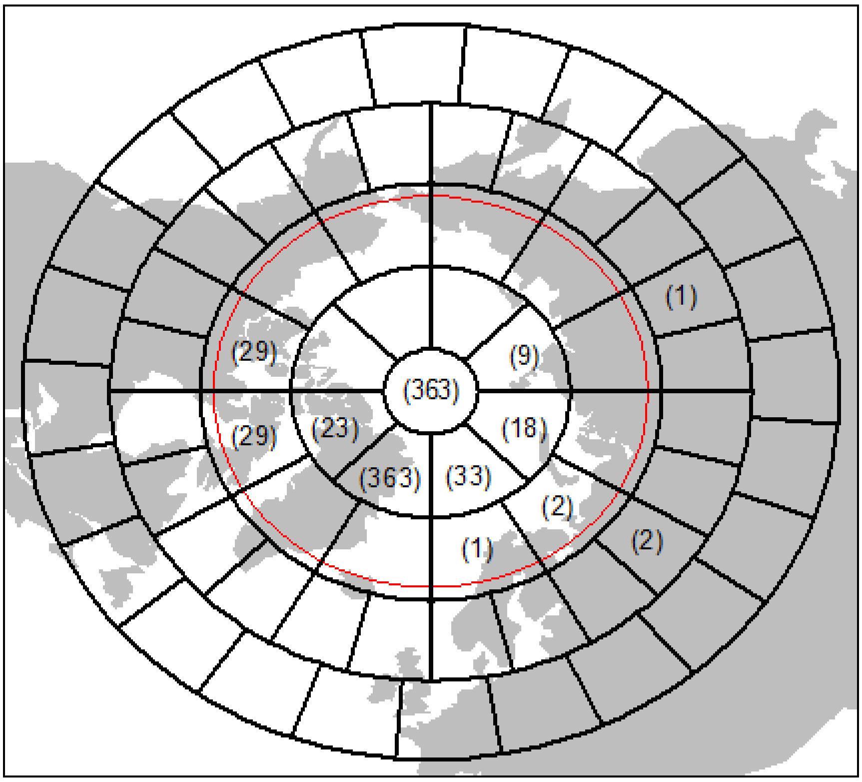


Figure 1: Map of the 69 sub-regions above latitude 45 degrees North. Numbers in brackets are the number of months with missing data for 1973-2008.

Statistical methods

A separate linear regression model was first fitted to the average monthly temperature anomalies for each of the 69 sub-regions. The model takes the following form:

$$y_{it} = b_{0i} + b_{1i}t + \in_{it} \quad \text{for } i = 1,2,...,69 \text{ and } t = 1,2,...,432 \qquad (1)$$

where y_{it} denotes the average monthly temperature anomaly in sub-region i for month t, b_{1i}corresponds to the temperature anomaly change per month in sub-region i and $b_{0i}$is the intercept (not of direct interest to our study) and $\in_{it}$ is the error assumed to follow a normal distribution with constant variance.

Residuals(γ_{it}) from the linear model (1) were analyzed using an auto-regressive model AR(2). Filtered average monthly temperature anomalies (z_{it}) were then obtained by removing auto-correlations at lags 1 and 2 months using the equation (Chatfield, 1996) as follows.

$$\gamma_{it} = y_{it} - \hat{y}_{it}, \qquad (2),$$

$$z_{it} = \gamma_{it} - a_1\gamma_{i,t-1} - a_2\gamma_{i,t-2} \quad \text{for } i = 1,2,...,69 \text{ and } t = 3,4,...,432, \quad (3)$$

In these equations γ_{it} is the difference between the average monthly temperature anomaly in sub-regions i for month $t(y_{it})$ and the corresponding fitted value ($\hat{y}_{it}$), z_{it} is the filtered average monthly temperature anomaly in sub-region $i$and at month t, and a_1, a_2 are the average coefficients of the fitted auto-regressive models AR(2) across all 69 sub-regions.

Factor analysis (Johnson, 1998) was used to model the filtered average monthly temperature anomalies of the 67 sub-regions, not including the two sub-regions with insufficient data (see Figure 2 in next section), having 84% data missing. The factor analysis model with m factors (m < 67), denoted by $f_1, f_2, \ldots, f_m$, takes the form

$$z_{it} = \mu_i + \sum_{k=1}^{m} \lambda_{ik} f_k \quad \text{for } i = 1,2,\ldots,67,\ t = 3,4,\ldots,432 \text{ and } k = 1,2,\ldots,m \qquad (4)$$

where z_{it} is the filtered average monthly temperature anomaly in sub-region i at month t, μ_i is the average across 432 months for sub-region i, λ_{ik} is the factor loadings at the i^{th} sub-region on the k^{th} factor and f_k is the k^{th} common factor.

As a result, inter-correlated sub-regions among the 67 sub-regions were identified. The highest inter-correlated sub-regions, i.e., those with similar filtered average monthly temperature anomalies, were classified and regrouped to form a larger region (factor) by maximizing the likelihood of the covariance matrix and minimizing the correlation between the factors for a specified number of factors, which needed to be specified in advance. The loadings (usually between -1 and 1) were controlled by rotating the factors, using the Promax method, to make these loadings as close as possible to 0 or 1. Ideally, each variable (i.e., a sub-region in our data) is correlated with only one factor, so that the correlation matrix of the residuals from this model is close to a diagonal matrix. For each of the large regions identified in the factor analysis, change in (filtered) temperature was estimated by fitting a linear model. All analyses and graphical displays were carried out using R (R Development Core Team, 2009).

RESULTS

Based on the estimated temperature change per month, obtained from the simple linear regression of time for each of the 69 sub-regions, temperature change per decade was derived and mapped in Figure 2, where red is for a significant increase (p-value < 0.05), orange if no evidence of change and white for insufficient data in the sub-region. Of the 69 sub-regions, average monthly temperature increased in all but 5 sub-regions. There is no evidence of change in three and insufficient data in two of these five sub-regions. The two sub regions with insufficient data were omitted in further analysis.

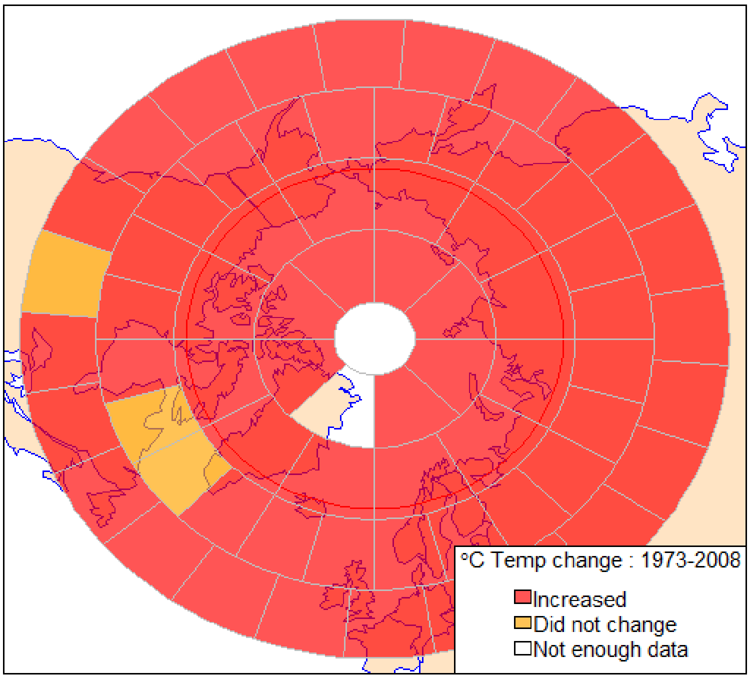


Figure 2: Map of temperature changes of the 69 sub-regions from 1973 to 2008 (before removing auto-correlations).

The time series plots of the three sub-regions which had no evidence of temperature change from 1973 to 2008 are presented in Figure 3. These sub-regions were latitude 55˚N–65˚N longitude 135˚W–120˚W (p-value = 0.116), latitude 55˚N–65˚N longitude 120˚W–105˚W (p-value = 0.065) and latitude 45^{o} N–55^{o} N longitude 85˚W–70˚W (p-value = 0.764).

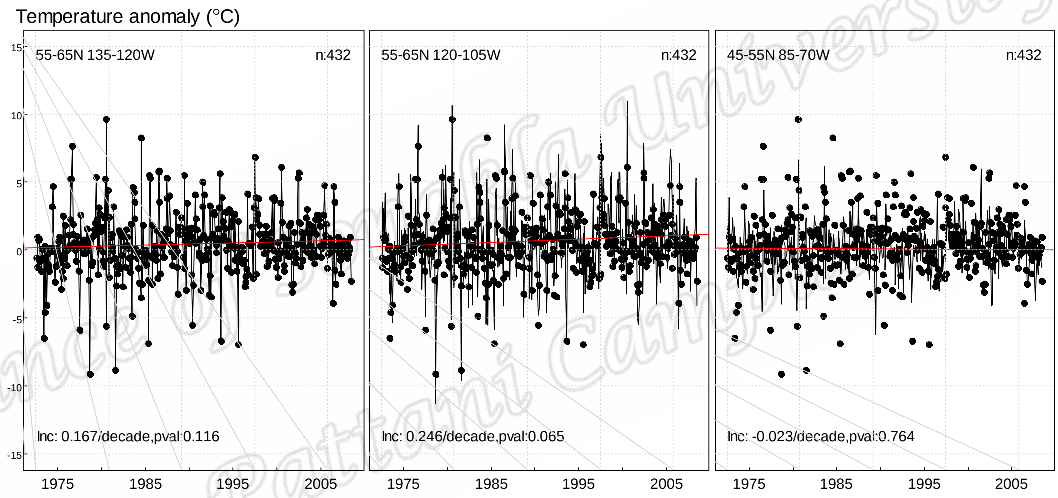


Figure 3: The time series plot of filtered average monthly temperatures anomalies in the three sub-regions with no evidence of change.

The 95% confidence intervals of temperature change per decade for each of the 67 sub-regions were estimated and plotted in Figure 4, ordered by latitude N and longitude from 180 W. A 95% confidence interval covering zero implies no significant temperature change.It confirms the results shown in Figures 2 and 3.

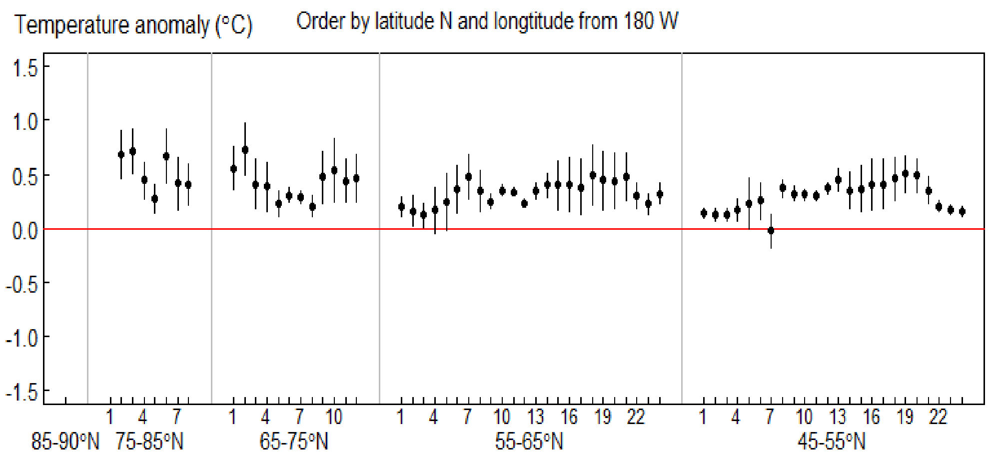


Figure 4: 95% confidence intervals of monthly temperature anomalies changes, ordered by latitude N and longitude from 180W. The horizontal axis shows sub-regions in each group of latitudes.

Residuals after removing a time trend for each region were used to assess possible auto-correlations over time. These residual time series are assumed to be stationary. Figure 5 is a graph of auto-correlation functions (ACF) of the residuals for two of the 67 sub-regions. Coefficients a_1 and a_2 were obtained by fitting an auto-regressive model with two parameters to the data in each region.

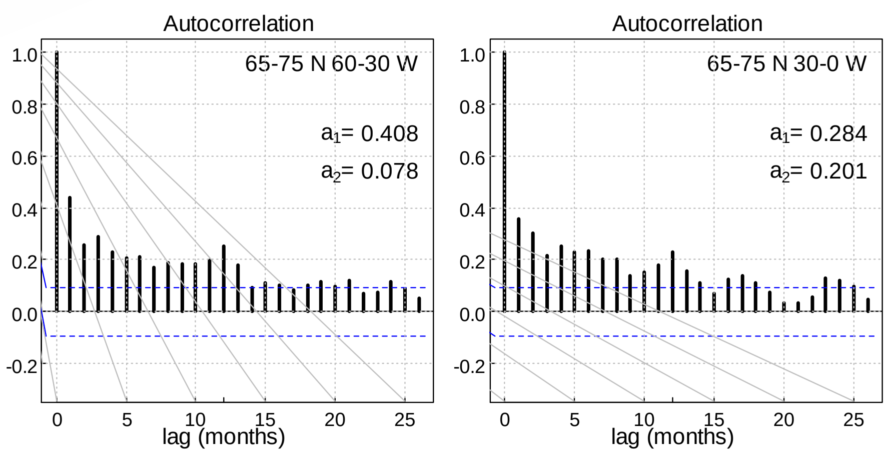


Figure 5: Auto-correlation function plots of the residuals for two of the 67sub-regions. The dotted line represents the 95% confidence interval of a zero correlation.

The average coefficients, $\bar{a}_1$ and $\bar{a}_2$ of the fitted auto-regressive models AR(2) across all the 67 sub-regions, are 0.333 and 0.04, respectively. Filtered residuals were then obtained after removing the auto-correlation structure using those average coefficients. Figure 6 contains the ACF plots of the filtered residuals for the same two of the 67 sub-regions illustrated in Figure 5. It shows that those filtering residuals are reasonably free of any auto-correlation.

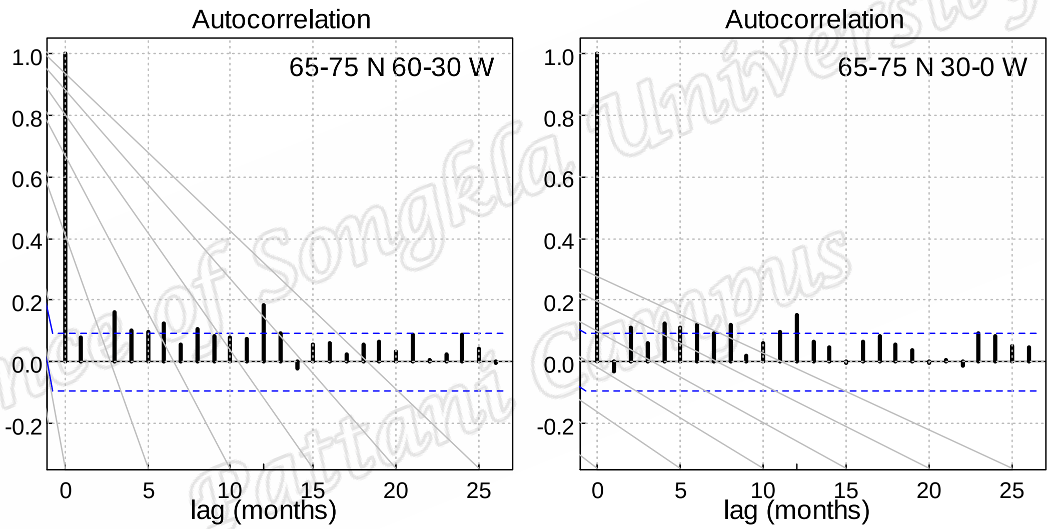


Figure 6: Auto-correlation function plots of the filtered residuals for the two sub-regions illustrated in Figure5. The dotted line represents the 95% confidence interval for a zero correlation.

Factor analysis of those filtered residuals was applied to the 67 sub-regions and resulted in 12 factors. Each factor with a unique pattern of temperature change represents a large region that consists of two to nine adjoining sub-regions (coded from 1 to 12 in Figure 7). Five (in cream in Figure 7) of the 67 sub-regions included in the factor analysis had high uniqueness.

CMU J. Nat. Sci. (2014) Vol. 13(3)

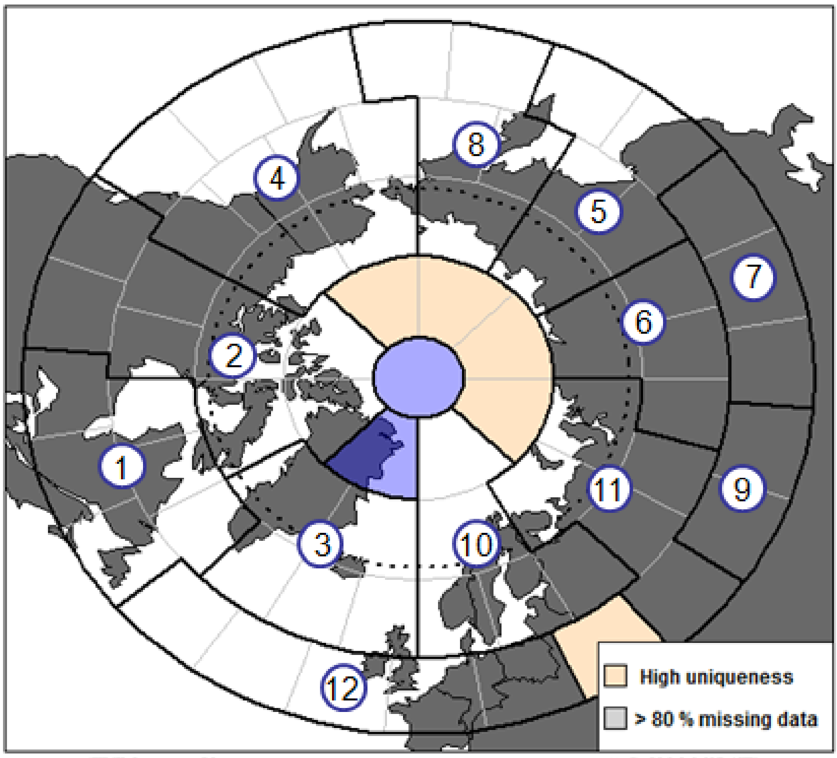


Figure 7: Twelve large regions (coded 1-12) identified from factor analysis.

The correlation matrix of the filtered residuals between pairs of the sub-regions within each of the 12 large regions is displayed as dots in the left panel of Figure 8. It is clear that sub-regions within each large region are highly positively correlated (showing large black dots) with each other, but there is little correlation (small faded dots) between pair of sub-regions from different large regions. The circles in the right panel of Figure 8 is based on residuals derived from the factor model, which shows very little correlation between pairs of sub regions even within the same large region, as expected. This indicates that the factor model has largely accounted for the correlation between those sub-regions considered, suggesting that the 12 factors explain the correlations between sub-regions reasonably well.

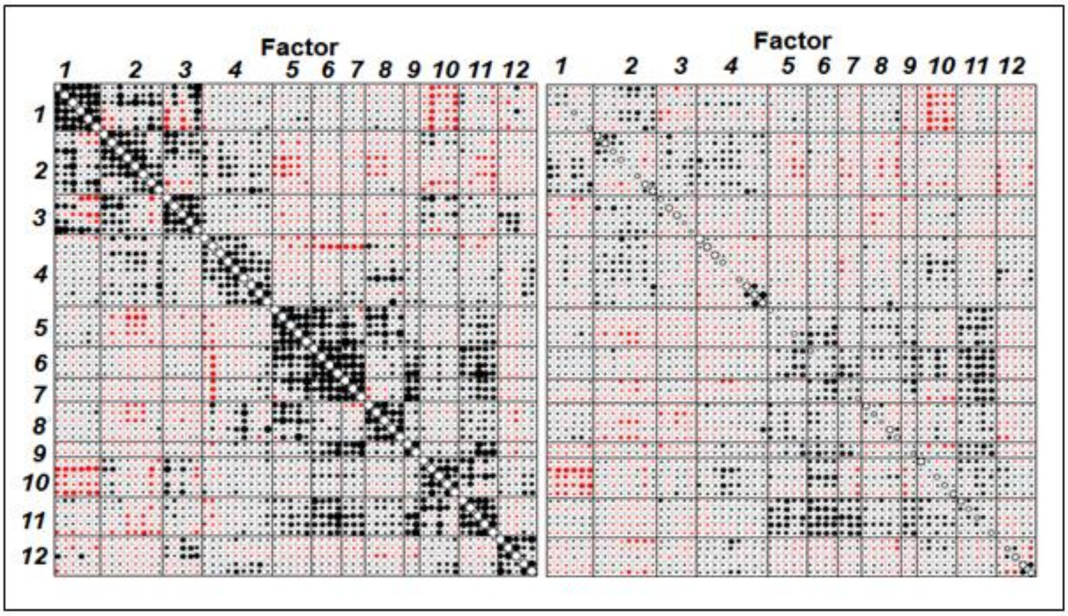


Figure 8: Bubble plots matrix of correlations between filtered monthly temperature anomalies in regions before (left panel) and after (right panel) fitting the factor model. Positive correlations are shown as black dots and negative correlation are shown as red dots. The size of each dot is proportional to the absolute value of the correlation coefficient.

The average changes with 95% confidence intervals (CI) of filtered monthly temperature anomalies per decade over 1973-2008 were estimated for each of the 12 large regions, as well as the five sub-regions with high uniqueness in the factor analysis, and are mapped in Figure 9. Eleven of the 12 large regions and all the five sub-regions showed a significant increase in temperature (p-values< 0.05). Each such region was re-classified into three levels according to the lower bound of the 95% CI of its temperature change per decade: small if between 0.090˚C and 0.129˚C, moderate (0.130˚C-0.199˚C) and large (0.200˚C-0.320˚C) temperature increase. Three colours, yellow, orange and red, were used to indicate the three different levels in Figure 9.

Clearly, large temperature increases occurred in the north Pacific Ocean, Alaska and Eastern Siberia (large regions 4, 8 and 9). Increases in north Canada, Greenland, Iceland, Norway, Sweden and Finland (large regions 2, 3, 10 and 11) were moderate. Small increases occurred in northern Siberia and part of the north Atlantic (large regions 5, 6, 7 and 12), whereas north-east Canada and its surrounding seas (large region 1) did not experience much change (coloured green in Figure 9). Four out of the five sub-regions (coloured cream in Figure7) experienced a small temperature increase per decade, but the other sub-region, covering Severanya Zemlya, had a large temperature increase. Note that the two sub-regions with more than 80% missing data are coloured blue in Figure 9.

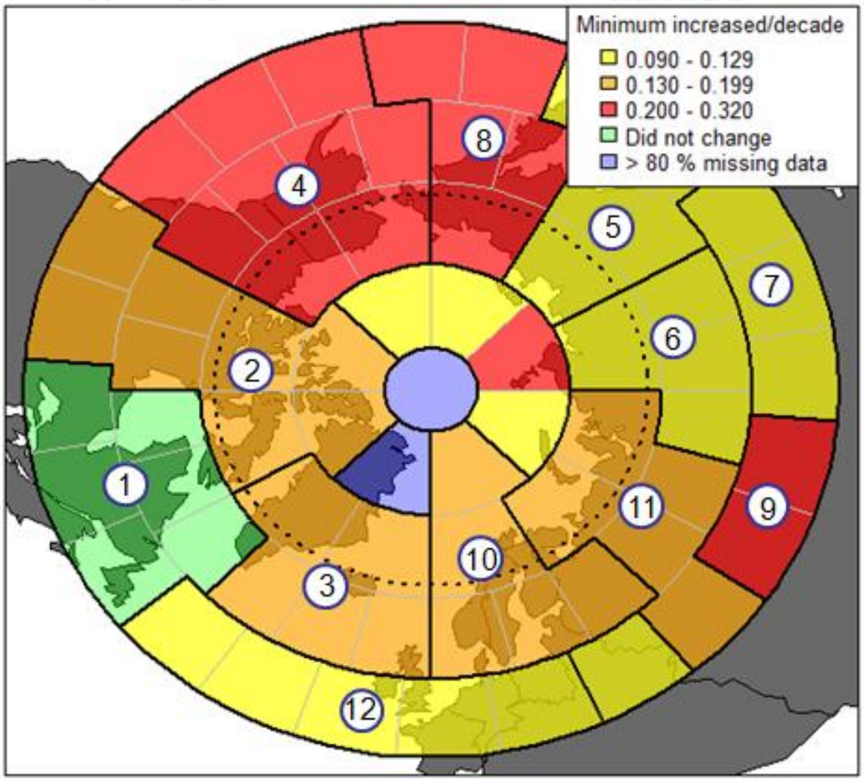


Figure 9: The minimum temperature increased per decade in regions

DISCUSSION AND CONCLUSIONS

Twelve large regions covering 62 of the 69 sub regions studied, each with different warming patterns, were identified from our study. Some had relatively large temperature increases per decade, including the north Pacific Ocean, Alaska and Eastern Siberia. A large temperature change was also detected in Severanya Zemlya. Except for north-east Canada and its surrounding seas, almost all regions included in the study experienced some level of temperature increase.

An earlier study by Jones et al., (1999) suggested that temperature in the Northern hemisphere rose on average by 0.37˚C in 1925-1944 and 0.32˚C over 1978-1997. As in Jones's study, our study showed a temperature increase almost everywhere. Temperature increase detected in Arctic in the present study is consistent with that from the studies of the same region over a similar period by Polyakov (2012) and Rigor et al., (1999). In addition, this study showed that rising temperature in the northern areas (land and sea) of the Atlantic, similar to that on the surface of the northern Atlantic Ocean reported recently by McNeil and Chooprateep (2013).

The results from the present research provide comprehensive information about temperature changes on the Earth surface close to Arctic region (land and sea), which has been one of the most important areas for global warming research. Any change in the Arctic may result in a substantial change in climate globally, and by the same token global change may have pronounced effects on the Arctic region. The methodological approach used in the present study can be applied to similar studies in other regions of the Earth surface. Furthermore, there are other statistical methods or techniques, such as linear spline model, T-mode PCA with K-means clustering and polynomial functions, which could be utilised to extend the approach used in the current study. Although the present study only considered temperature, the approach presented here can be applied to the studies of other factors, such as solar radiation, wind, and rainfall in future studies.

ACKNOWLEDGEMENTS

The authors are grateful to Emeritus Prof.Don McNeil, Department of Statistics, Macquarie University, Australia for supervising our research.

REFERENCES

Anisimov, O.A.,V.A. Lobanov, and S.A. Reneva. 2007. Analysis of changes in air temperature in Russia and Empirical forecast for the first quarter of the 21st century. Russian Meteorology and Hydrology 32: 620-626.

Box, J.E., L. Yang, D.H. Bromwich, and L.S. Bai. 2009. Greenland Ice sheet Surface Air Temperature Variability: 1840-2007. America Meteorological Society 22: 4029-4049.

Chatfield, C. 1996. The Analysis of Time Series, Chapman & Hall, Melbourne.

Hughes, G.L., S.S. Rao, and T.S. Rao. 2006. Statistical analysis and time-series models for minimum/maximum temperatures in the Antarctic Peninsula. The royal society 463:241-259. Doi:10.1098/rspa.2006.1766.

Houghton, J.T., Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, and D.Xiaosu. 2001. Climate change 2001: The Scientific Basic. Cambridge University Press, 944 pp.

Johannessen, O.M., L. Bengtsson, M.W. Miles, S.I. Kuzmina, V.A. Semenov, G.V. Alekseev, A.P. Nagurnyi, V.F. Zakharov, L.P. Bobylev, L.H. Pettersson, K. Hasselmann, and H.P. Cattle. 2004. Arctic climate change: observe and modeled temperature and sea-ice variability. Tellus 56A:328-341.

Johnson, D.E. 1998. Applied Multivariate Methods for Data Analysts. Duxbury.

Jones, P.D., M. New, D.E. Parker, S. Martin, and I.G. Rijor. 1999. Surface air temperature and its changes over the past 150 years. Reviews of Geophysics 37: 173-199.

Krupnik, I., and D.Jolly. 2002. The Earth is Faster Now. Arctic Research Consortium, Fairbanks, AK.

Lean, J.L., and D.H. Rind. 2009. How will Earth's surface temperature change in future decades?. Geophysical research letters 36: L15708, Doi:10.1029/2009Glo03892.

McNeil, N., and C.Chooprateep.2013. Modeling sea surface temperatures of the North Atlantic Ocean. Theoretical and Applied Climatology, Doi:10.1007/s00704-013-0930-0.

National Academies Report. 2008. Understanding and Responding to Climate Change. Highlights of National Academies Report.

Overpick, J., K. Hughen, D. Hardy, R. Bradley, R. Case, M. Douglas, B. Finney, K. Gajewski, G. Jacoby, A. Jennings, S. Lamoureux, A. Lasca, G. MacDonale, J. Moore, M. Retelle, S. Smith, A. Wolfe, and G. Zielinski. 1997. Arctic environmental change of the last four centuries. Science 278:1251-1256.

Polyakov, V.I., R.V. Brkryaev, G.V. U.S. Bhatt, R.L. Colony, M.A. Johnson, A.P. Maskshtas, and D. Walash. 2002. Variability and Trends of Air Temperature and Pressure in the Maritime Arctic, 1875-2000. American Metrological Society 16:2067-2076.

R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2009.Available at URL: http://www.R-project.org. [Accessed "1 November 2009"].

Rigor, I.G., R.L. Colony, and S. Martin. 1999. Variation in Surface Air Temperature Observations in the Arctic, 1979-97. Climate 13: 896-914.

Semenov, V.A. 2007. Structure of Temperature Variability in the High Latitudes of the Northern Hemisphere. Atmospheric and Oceanic physics 43: 744-753.

Venables, W.N., and B.D. Ripley. 2002. Modern Applied Statistics with S. Springer.

Appendix3

Modeling of daily maximum temperatures over Australia from 1970 to 2012

Wandee Wanishsakpong[a*] Nittaya McNeil[a]

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

Abstract

The variations of daily maximum temperatures over consecutive 5-day periods from 1970 to 2012 in Australia are described. The data were obtained from the Australia Bureau of Meteorology comprising measurements from 85 weather stations. Linear regression was initially used to model the seasonally adjusted daily maximum temperatures. Factor analysis was used to classify the temperatures from the 85 stations into seven factors which correspond to seven geographical regions. The average maximum annual temperature in these seven regions ranged from 23^{o}C to 36^{o}C. A sixth order polynomial regression model was then fit to investigate the trend and pattern of the temperatures in these seven regions. Similar increasing temperature trends occurred in the central, eastern, southern, and southeastern parts of the country.

Key words Australia; climate change; time series analysis; linear regression; factor analysis; polynomial regression model

1. Introduction

Australia has become the most interesting country for conducting research on global atmospheric systems because its climate has geographical variety and it is also surrounded by the Indian and Pacific Oceans. The majority of Australia is desert or semi-arid land. The southeastern and southwestern corners have a temperate climate. The northern part of the country has a tropical climate and contains tropical rainforests, grasslands as well as desert areas. The climate and geographic characteristics could also be a related factor to natural disasters, such as floods, droughts, tropical cyclones, rising sea levels, and the EI Nino southern oscillation (ENSO) (Hughes, 2003). The temperature in Australia has been increasing over the recent decades (Murphy and Timbal, 2007). The average temperature has increased by approximately 0.8^{o}C since 1910 (Collins, 2000), and has been increasing by about 0.32^{o}C from 1981 to 2005 (Collins *et al.*, 2000). An increase in temperature by 0.1^{o}C - 0.2^{o}C per decade was found over most of Australia since1951. From 1910 to 2006, the temperatures have increased by approximately 0.19^{o}C per decade (Murphy and Timbal, 2007). During the period 1910-1929, the seas around Australia have experienced warmer temperatures by around 0.7^{o}C with waters warming faster in the southwestern and southeastern areas (Poloczanska *et al.*, 2009).

Many studies have assessed the trend and pattern of temperature change using various methodologies including computer simulation models (Johannessen *et al.*, 2003), empirical orthogonal functions (Semenov, 2007) and various statistical techniques. Griffiths *et al.* (2005) used linear regression models and Pearson correlation to study the relationship between mean and extreme temperatures in the Asia Pacific region. Jones (1999) used principal component analysis to investigate the characteristics and some mechanisms of Australian temperature variability. More recently, McNeil and Chooprateep (2014) used linear regression models, spline linear regression models and factor analysis to examine the trend and pattern of sea surface temperature in the North Atlantic Ocean.

Increases in extreme temperatures are consistent with warming of the climate (IPCC, 2007). It also associated with the variations of maximum temperature in part of Australia during the 20th century (Nicholls *et al*., 2004). In southeastern Australia the temperature records during 1958-2008 revealed increasingly hot summers since the 1990s, with daily maximum temperatures reaching 10^{o}C above average (Richman and Leslie, 2014). Therefore, this study the maximum temperature over consecutive 5-day periods from 1970 to 2012 in Australia are investigated by using linear regression models, factor analysis and polynomial regression models to assess the trends and patterns of temperature change.

2. Data management and study areas

The daily maximum temperatures were downloaded from the Australian Bureau of Meteorology (BOM) website at http://www.bom.gov.au for over 700 weather stations. These stations have different periods of observation, some records starting earlier or have incomplete records due to weather phenomena and changing requirements (BOM, 2013). An area-based sample of 85 stations was selected consistent with data from Australia’s Reference Climate Station Network (RCS) and the Australian Climate Observations Reference Network-Surface Air Temperature (ACORN-SAT). These sampled stations are spread approximately evenly over Australia as shown in Figure 1. Moreover, they have 80% complete data from 1970 to 2012.

Daily maximum temperature data was collected at each station. Observations on February 29 were omitted to have equal numbers of observations for each year, totaling 15,695 observations including missing data over the 43-year period. The maximum temperature over consecutive 5-day periods was used for analysis to reduce the number of data to 73 periods in each year, as well as reducing correlation between successive observations. Missing data was imputed using linear interpolation.

3. Methods

The data were seasonally adjusted to remove variation in the maximum temperature over consecutive 5-day periods by subtracting the average of the maximum temperatures in each period and then adding back the overall mean for all 85 stations. The equation takes the form

$$y_{ij} = x_{ij} - \bar{x}_{.j} + \bar{x}_{..} \qquad (1)$$

where y_{ij} is the seasonally adjusted maximum temperature over consecutive 5-day periods at year i period j, x_{ij} is the maximum temperature over consecutive 5-day periods at year i period j, $\bar{x}_{.j}$ is the average of period j over 43 years and $\bar{x}_{..}$ is overall mean.

A linear regression model was fitted to the seasonally adjusted maximum temperatures for each of the 85 stations (Venables and Ripley, 2002). The model takes the form

$$y_{ijk} = s_k + a_{ik} + b_{jk} \qquad (2)$$

where y_{ijk} denotes the seasonally adjusted maximum temperature over consecutive 5-day periods at year i for period j and station k, s_k is a constant for station k and the terms a_{ik} and b_{jk} are effects for year i and period j at station k, repectively.

The dependences among the seasonally adjusted maximum temperature data were reduced by removing the auto-correlations at lag 1 term (equivalent to 5-day period) (Chatfield, 1996). Factor analysis (Johnson, 1998) was used to model the maximum temperatures over consecutive 5-day periods in the 85 stations. The factor analysis model with *m* factors (*m*<85), denoted by $f_{1,} f_2, \ldots, f_m$, takes the form

$$y_{ijk} = \mu_{ijk} + \sum_{p=1}^{m} \gamma_{kp} f_{kp} \quad (3)$$

where y_{ijk} is the seasonally adjusted maximum temperature over consecutive 5-day periods at year i for period j of station k, μ_{ijk} is the average across all times for station k, γ_{kp} is the factor loading at station k for factor p, and f_{kp} is the p^{th} common factor of station k.

The highest inter-correlated stations, i.e., those with similar seasonally adjusted maximum temperature over consecutive 5-day periods, were classified into factors by maximizing the likelihood of the covariance matrix. These factors were identified using factor loadings. The loadings were controlled by rotating the factors using the Promax method (Venables and Ripley, 2002). The cut-off value for factor loadings was 0.3 (Hair *et al*., 1998). Stations with high uniqueness, that is, those stations which have variations in temperatures that are not similar to the other stations, were omitted from further analysis.

Since the temperature displayed a curvilinear pattern, polynomial regression was used to model this pattern. A sixth order polynomial regression model with consecutive 5-day periods as the sole predictor variable was fitted to account for the trend of maximum temperatures over consecutive 5-day periods in each factor. The model takes the form:

$$y_{jm} = b_{0m} + b_{1m}j + b_{2m}j^2 + b_{3m}j^3 + b_{4m}j^4 + b_{5m}j^5 + b_{6m}j^6 \quad (4)$$

where y_{jm} is the maximum temperature over consecutive 5-day periods at period j of factor m , b_{0m} represents the mean response of y_j when j = 0, and $b_{1m}, b_{2m,\dots,}b_{6m}$ are effect coefficients of factor m.

All data analysis and graphical displays were carried out using R (R Development Core Team, 2008).

4. Results

Figure 2 shows the pattern of 5-day maximum temperatures for five selected stations. There is a noticeable seasonal periodicity pattern.

The data were seasonally adjusted and the auto-correlation at lag 1was removed; the normality distributed assumption was also verified. A linear regression model (model 2) was then fitted to the seasonally adjusted, decorrelated data and the results are shown in Figure 3. The quantile-quantile (Q_Q) plot shown in Figure 4 indicates that the residuals from the model are normally distributed. The coefficients for the periods (b_{jk}), which represent the temperature trends and p-values, are also displayed in Table I. The most coefficients were significantly different from zero (p-value<0.05), indicating that the temperatures for these stations increased but one station (a1013) had decreased in temperature as negative coefficient. Overall, the trend increased in 74 stations, decreased in one station and remained the same in 10 stations that was not significant.

The spatial correlation of the seasonally adjusted maximum temperatures was also investigated. The correlations of the seasonally adjusted temperature in the 85 stations ranged from 0.00007 to 0.925, with a median of 0.167. A factor model was used to reduce these correlations by classifying the 85 stations into groups (factors) by considering the value of the factor loadings. Table II shows the factor loadings ordered from highest to lowest within each factor (excluding the mixed factor loading), and uniqueness values. Figure 5 shows a map of Australia and the seven regions as determined by the factor model. The circles represent the 85 stations. The size of each circle is proportional to the factor loading for each station. Thirty stations with mixed factor loadings were omitted in further analyses as shown in superimposed circles. One station (Lockhart River Airport; a28008) in northern Australia (hollow circle in Figure 5) had a uniqueness value above 0.85, thus there is no dominating factor for this station.

In each factor, the average of maximum temperatures over consecutive 5-day periods ranged from 23^{0}C to 36^{0}C over the 43 years as shown in Figure 6. The data shows distinct periodic variation. A sixth order polynomial regression model was fitted with these trends and is shown as curves in Figure 6(a)-6(g). The coefficients of the polynomial regression models up to the sixth order were all significantly different from zero (p-value <0.05) as shown in Table III.

Overall, temperatures increased with some decreases occurring during some periods. It can be seen that factor 2 represents the desert area of the northwestern area with the highest average temperature of approximately 36 ^{0}C in 2002. The trend and pattern of temperature change in the central (factor3), eastern (factor 4), southern (factor 6) and southeastern (factor 7) regions were grouped with similar patterns as shown in Figure 6(h). The maximum temperatures increased after 1974, steadily decreased around 1984, then gradually increased from 2000 until 2005, and finally decreased. Figure 6(i) shows the three regions with different temperature trends, namely factor 1(northern), factor 2 (northwestern) and factor 5 (southwestern).

5. Conclusion and discussion

The seasonally adjusted maximum temperatures over consecutive 5-day periods were analyzed using factor analysis. Seven factors corresponding to seven geographical regions, namely the northern, northwestern, central, eastern, southwestern, southern and southeastern regions, were identified each having different patterns of temperature changes. The average maximum temperatures in all regions between 1970 and 2012 ranged from 23^{o}C to 36^{o}C.

The data from each region showed distinct periodic variations and were investigated using a sixth order polynomial regression model. Four geographic regions, namely the central, eastern, southern and southeastern regions, had similar temperature patterns. The temperatures of these regions are influenced by rainfall and EI Nino southern oscillation events (Nicholls *et al*., 2004; Alexander *et al*., 2007; Murphy and Timbal, 2007; Deo *et al*., 2009). The increase in temperature and decrease in rainfall has direct consequences such as increasing occurrence and duration of drought**s** (Deo *et al.*, 2009). In the southeastern region, droughts have been occurring since 1973 with increased temperatures during 2002 and 2003 (Nicholls, 2004).The rainfall in central New South Wales and northern Victoria decreased by about 15% and the number of dry days increased 3-5 days per year for the period 1951-2003. Moreover, eastern Australia has experienced significant increased temperatures and decreased rainfall in the summer periods during 1982-

1983 and 2002-2003. The temperature increased by around 2^{o}C to 2.5^{o}C with rainfall deficit in this period (Deo *et al*., 2009).

The variability of temperature can be explained by a simple linear regression model, which is used in the first step to identify trends (Nicholls *et al*., 2004) but may not be the best way to estimate trends for shorter periods (Benestad, 2003). In the same way, a sixth order polynomial regression model was considered to account for detailed description of the trend within the calibration interval (Benestad, 2003). However, this method is not suitable for forecasting trends and there is no conclusive evidence that the trend could be predicted (Woodward and Gray, 1992; Benestad, 2003). It is clear that Australia has a periodic pattern of temperature change. Therefore, other methods for projection of temperature trends, such as periodogram analysis and use of harmonic functions could be used to investigate the temperature trend in future.

Acknowledgement

The authors are grateful to Emeritus Prof Don McNeil for supervising our research.

References

Alexander LV, Hope P, Collins D, Trewin B, Lynch A, Nicholls N. 2007. Trends in Australia's climate means and extremes: a global context. Aust Met Mag 56:1-18

Benestad R. 2003. What can present climate models tell us about climate change?. Climatic change 59:311-331

BOM. 2013. Bureau of Meteorology, http://www.bom.gov.au. Accessed 1 July 2013

Chatfield C. 1996. The Analysis of Time series, Chapman & Hall, Melbourne

Collins D. 2000. Annual temperature summary: Australia records warmest decade. Climate Change Newsl 12(6)

Collins DA, Della-Marta PM, Plummer N, Trewin BC. 2000. Trend in annual frequencies of extreme temperature events in Australia. Aus Met Mag 49:277-292

Deo RC, Syktus JI, McAlpine CA, Lawrence PJ, McGowan HA, Phinn SR. 2009. Impact of historical land cover change on daily indices of climate extremes including droughts in eastern Australia. Geophys Res Lett 36:L08705

Griffiths GM, Chambers LE, Haylock MR, Manton MJ, Nicholls N, Baek HJ, Choi Y, Della-Marta PM, Gosai A, Iga N, Lata R, Laurent V, Maitrepierre L, Nakamigawa H, Ouprasitwong N, Solofa D, Tahani L, Thuy DT, Tibig L, Trewin B, Vediapan K, Zhai P. 2005. Change in mean temperature as a predictor of extreme temperature change in the Asia-Pacific region. Int J Climatol 25:1301-1330

Hair JF, Anderson RE, Tatham RL, Black WC. 1998. Multivariate Data Analysis, Prentice Hall International (UK) Limited, London

Hughes L. 2003. Climate change and Australia: trend, projections and impacts. Austral Ecology 28:423-443

IPCC. 2007. Intergovernmental Panel on Climate Change, http://www.ipcc.ch. Accessed 1 July 2013

Johannessen OM, Bengtsson L, Miles MW, Kuzmina SI, Semenov VA, Alekseev GV, Nagurnyi AP, Zakharov VF, Bobylev LP, Pettersson LH, Hasselmann K, Cattle HP. 2003. Arctic climate change: observe and modeled temperature and sea-ice variability. Tellus 56:328-341

Johnson DE. 1998. Applied Multivariate Methods for Data Analysis, Duxbury, USA

Jones DA. 1999. Characteristics of Australian land surface temperature variability. Theor Appl Climatol 63:11-31

McNeil N, Chooprateep S. 2014. Modeling sea surface temperatures of the North Atlantic ocean. Theor Appl Climatol 116:11-17

Murphy BF, Timbal B. 2007. A review of recent climate variability climate change in southeastern Australia. Int J Climatol 28:859-879

Nicholls N. 2003. Continued anomalous warming in Australia. Geophys Res Lett 30:1370

Nicholls N. 2004. The changing nature of Australian droughts. Climatic change 63:323-336

Nicholls N, Della-Marta D, Collins D. 2004. 20th century changes in temperature and rainfall in New South Wales. Aust Met Mag 53:263-268

Poloczanska ES, Hobday AJ, Richardson AJ. 2009. Report Card of Marine Climate Change for Australia. NCCARF publication 05/09. ISBN 978-1-921609-03-9

R development Core Team. 2008. A language and environment for statistical computing. R Foundation for Statistical computing, Vienna

Richman MB, Leslie LM. 2014. Attribution and Prediction of Maximum Temperature Extremes in SE Australia. Procedia Computer Science 36:612-617

Semenov VA. 2007. Structure of temperature variability in the high latitudes of the northern hemisphere. Atmos Oceanic Phys 43:744-753

Venables WN, Ripley BD. 2002. Modern Applied Statistics with S. Springer, Queenland

Woodward W, Gray HL. 1992. Global warming and the problem of testing for trend in time series data. J Climate 6:953-962

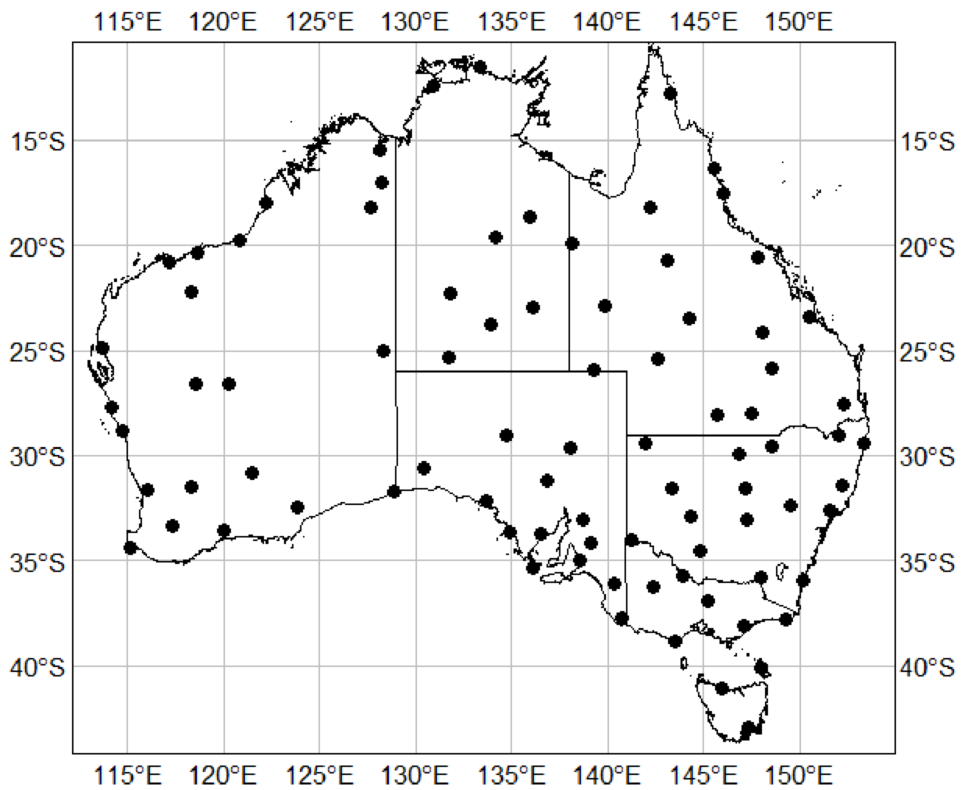


Figure 1. Station locations where daily maximum temperature data were collected. The state boundaries in Australia are also given.

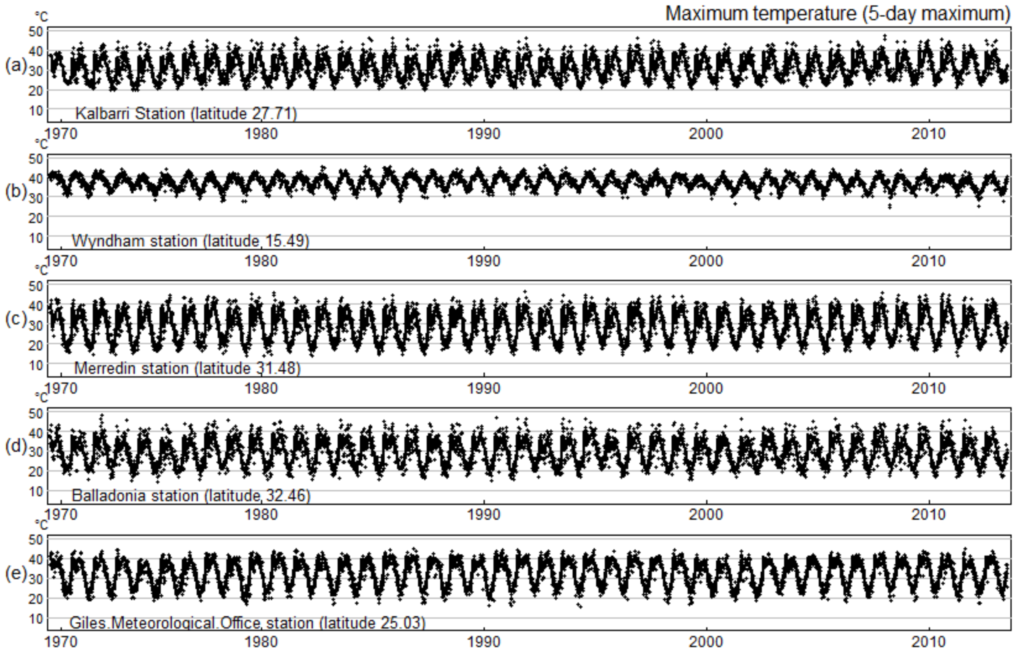


Figure 2. Maximum temperatures over consecutive 5-day periods of 5 selected stations (a) Kalbarri station (b) Wyndham station (c) Merredin station (d) Balladonia station and (e) Giles Meteorological station. The dots represent the data and the curve represents their fitted values using model (2).

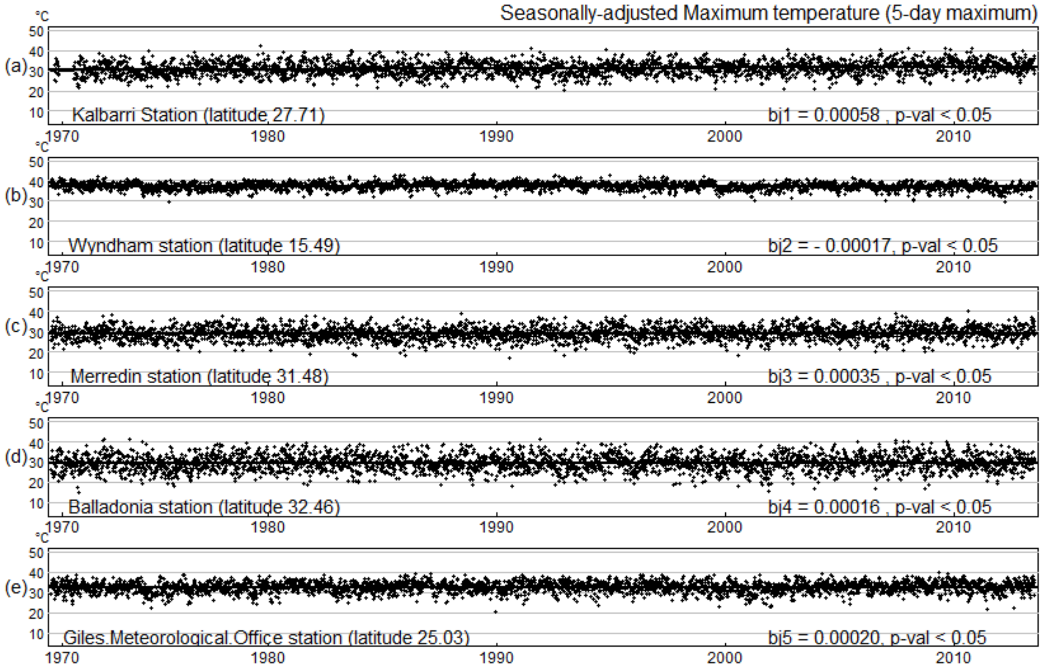


Figure 3. Seasonally adjusted maximum temperatures over consecutive 5-day periods of 5 selected stations (a) Kalbarri station (b) Wyndham station (c) Merredin station (d) Balladonia station and (e) Giles Meteorological station. The dots represent the data and the lines represents temperature trend using model (2).

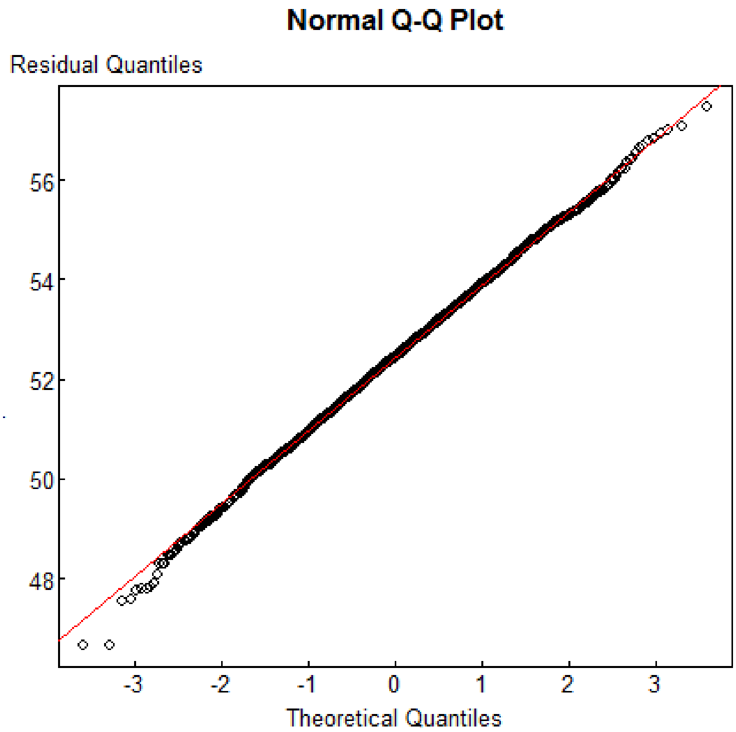


Figure 4. Quantile-quantile (Q-Q) plot of residuals from model (2).

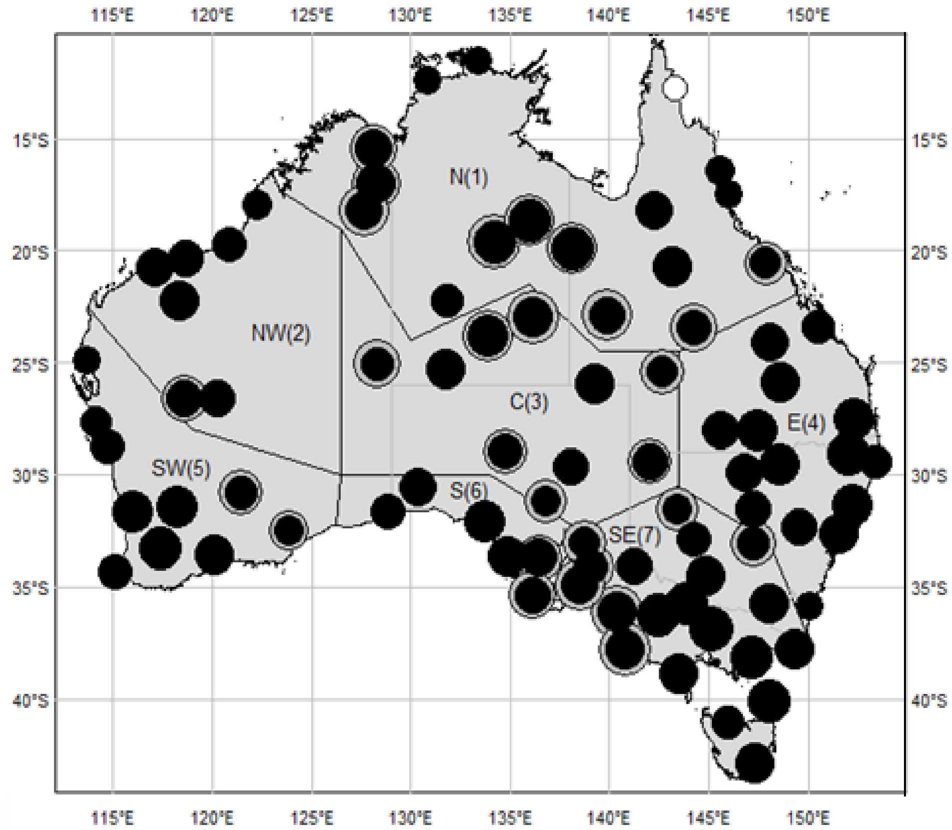


Figure 5. Factor analysis divides the stations into 7 geographical groups. Numbers in brackets are the number of factors. E = east, SE = southeast, S = south, SW = southwest, NW = northwest, N = north, C = central.

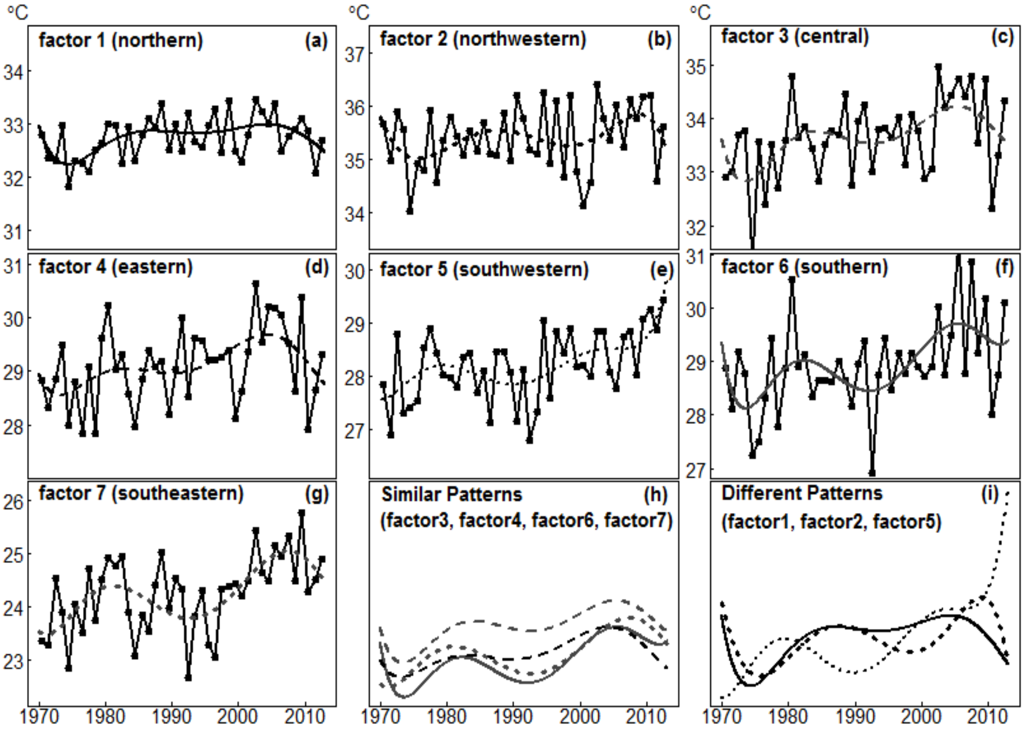


Figure 6. The patterns of temperature change in seven factors corresponding to seven geographical (a)-(g). The dots are mean of maximum temperature over consecutive 5-day periods based on model (2). The curved lines are trend estimation based on model (4). The bottom centre panel shows regions with similar patterns, while the bottom right panel shows regions with dissimilar patterns.

Stations	b_{jk}	p-value	Stations	bjk	p-value	Stations	bjk	p-value
a8251	0.00058	< 0.0005	**a1013**	-0.00018	< 0.0005	**a99005**	0.00029	< 0.0005
a11003	0.00038	0.00028	**a4035**	0.00018	0.00017	**a75031**	0.00037	< 0.0005
a6011	0.00030	0.00001	**a2012**	-0.00001	0.82617	**a90015**	0.00012	0.14892
a48031	0.00022	0.00017	**a62013**	0.00044	< 0.0005	**a15085**	0.00013	0.00451
a10647	0.00034	< 0.0005	**a37010**	0.00009	0.06904	**a33013**	0.00039	< 0.0005
a44026	0.00032	< 0.0005	**a44010**	0.00016	0.00681	**a38002.38**	0.00040	< 0.0005
a15528	0.00054	< 0.0005	**a15511**	0.00043	< 0.0005	**a85072**	0.00034	0.00001
a49019	0.00060	< 0.0005	**a43015**	0.00030	< 0.0005	**a69018**	0.00015	0.02255
a10633	0.00016	0.02764	**a18069**	0.00046	< 0.0005	**a32025**	0.00028	< 0.0005
a5026	0.00016	0.00026	**a7045**	0.00026	< 0.0005	**a4032**	0.00018	0.00029
a18012	0.00031	0.00121	**a17031**	0.00014	0.02027	**a50052**	0.00040	< 0.0005
a46043	0.00016	0.01519	**a94029**	0.00026	0.00010	**a51039**	0.00023	0.00009
a28008	0.00015	< 0.0005	**a4019**	0.00010	0.06615	**a23034**	0.00027	0.00049
a88109	0.00038	< 0.0005	**a9518**	0.00022	0.00001	**a80023**	0.00048	< 0.0005
a2032	0.00006	0.14374	**a13017**	0.00020	0.00012	**a26021**	0.00031	0.00011
a48015	0.00015	0.01769	**a78077**	0.00032	0.00001	**a56032**	0.00005	0.29877
a15135	0.00011	0.01879	**a35065**	0.00011	0.02424	**a24511**	0.00048	< 0.0005
a3003	-0.00008	0.08989	**a15602**	0.00014	0.01823	**a14401**	0.00006	0.00346
a38024	0.00035	< 0.0005	**a18115**	0.00021	< 0.0005	**a38003**	0.00025	< 0.0005
a14015	0.00008	0.00013	**a16001**	0.00048	< 0.0005	**a10092**	0.00036	< 0.0005
a29012	-0.00005	0.11128	**a40082**	0.00053	< 0.0005	**a13012**	0.00031	< 0.0005
a36031	0.00016	0.00139	**a18110.18**	0.00036	0.00004	**a84070**	0.00039	< 0.0005
a30045	-0.00004	0.35443	**a16007.16**	0.00012	0.08696	**a39083**	0.00030	< 0.0005
a25507	0.00023	0.00252	**a47016**	0.00049	< 0.0005	**a11017**	0.00017	0.03703
a72043	0.00050	< 0.0005	**a19062**	0.00034	< 0.0005	**a18014**	0.00045	< 0.0005
a61250	0.00039	< 0.0005	**a9053**	0.00044	< 0.0005	**a8051**	0.00025	0.00069
a31037	0.00015	< 0.0005	**a60085**	0.00018	0.00173	**a58012**	0.00037	< 0.0005
a46037	0.00050	< 0.0005	**a15590**	0.00039	< 0.0005			
a12038	0.00032	< 0.0005	**a91009**	0.00010	0.00688			

Table I. Regression coefficients of 85 stations.

Stations	Factor Loading F1	F2	F3	F4	F5	F6	F7	Uniqueness
a88109	**0.93**							0.23
a80023	**0.87**							0.17
a99005	**0.87**		0.11	-0.17				0.42
a85072	**0.87**			-0.14				0.36
a78077	**0.85**				0.12			0.13
a75031	**0.75**	0.17		0.15				0.16
a90015	**0.75**	-0.10	0.11	-0.16	0.24			0.38
a84070	**0.75**						-0.13	0.48
a94029	**0.73**	-0.12	0.11	-0.15	0.15			0.55
a72043	**0.73**	0.22						0.36
a47016	**0.63**			0.13	0.21			0.31
a49019	**0.55**	0.25		0.26				0.29
a91009	**0.52**							0.79
a46043	**0.43**	0.28		**0.43**				0.19
a19062	**0.47**			0.27	**0.41**			0.16
a50052	**0.49**	**0.46**	-0.13	0.17	-0.11			0.24
a24511	**0.56**			0.14	**0.51**			0.11
a25507	**0.70**	-0.10			**0.56**	-0.13		0.10
a26021	**0.70**	-0.16	0.12	-0.12	**0.55**	-0.15		0.19
a56032		**0.83**						0.35
a40082	-0.16	**0.81**		-0.18	0.11	0.10		0.39
a60085		**0.80**	-0.12	-0.14		0.16		0.39
a48031		**0.79**						0.28
a44010		**0.77**		0.18				0.22
a43015	-0.11	**0.75**	0.24		0.10			0.31
a61250	0.20	**0.71**	-0.16		-0.10	0.16	-0.11	0.39
a48015	0.16	**0.66**		0.25				0.28
a35065	-0.14	**0.65**	0.25		0.12			0.36
a62013	0.28	**0.63**	-0.15					0.27
a51039	0.29	**0.60**	-0.16	0.22				0.23
a39083	-0.21	**0.53**	0.12	-0.18				0.50
a58012	-0.18	**0.50**		-0.12		0.11		0.72
a44026		**0.69**		**0.33**		-0.11		0.21
a69018	**0.30**	**0.38**			-0.12	0.16	-0.13	0.71
a30045		0.27	**0.70**			-0.18		0.35
a29012			**0.69**					0.51
a15528			**0.52**	0.11		0.23		0.38
a31037			**0.40**	-0.12				0.75
a14401			**0.39**	-0.12		0.17		0.84
a14015	0.12		**0.36**			0.24		0.82
a32025			**0.36**	-0.14				0.79
a15135		-0.13	**0.86**	**0.45**				0.18
a15085	0.11		**0.85**	**0.35**				0.27
a37010			**0.82**	**0.39**			0.11	0.29
a2032	0.12	-0.13	**0.73**	0.13		**0.41**		0.32
a2012		-0.15	**0.69**	0.21		**0.47**		0.28
a1013		-0.15	**0.67**			**0.40**		0.43
a38003		0.22	**0.63**	**0.55**		-0.22	0.12	0.23
a36031		**0.49**	**0.47**	0.17		-0.20		0.26
a33013	-0.12	**0.39**	**0.47**	-0.16				0.52

Stations	Factor Loading F1	F2	F3	F4	F5	F6	F7	Uniqueness
a38002.38		0.24	0.28	**0.75**		-0.13		0.22
a15511	-0.14		0.16	**0.70**	0.14	0.29	-0.15	0.27
a17031	0.20	0.11		**0.60**	0.17			0.21
a15590			**0.42**	**0.76**		0.11		0.21
a15602			**0.54**	**0.75**				0.22
a46037	0.22	**0.35**		**0.59**				0.18
a16007.16				**0.56**	**0.34**	0.16		0.27
a13017	-0.18		0.13	**0.53**	0.18	**0.50**	-0.13	0.30
a38024		**0.46**	**0.41**	**0.48**		-0.19	0.11	0.22
a16001	0.26			**0.43**	**0.41**			0.18
a18012	0.12		-0.12		**0.76**	0.11		0.24
a18069	0.20				**0.75**			0.29
a18110.18			-0.12		**0.64**	0.25		0.37
a11003					**0.57**	0.22		0.54
a23034	**0.52**				**0.65**			0.12
a18014	**0.36**				**0.64**			0.24
a18115	**0.43**			-0.12	**0.61**			0.37
a5026						**0.74**		0.39
a4035						**0.68**		0.51
a4032			0.13			**0.62**		0.57
a13012	-0.12		-0.13		0.21	**0.61**	0.22	0.39
a4019			0.15			**0.57**		0.64
a3003			0.16	-0.11		**0.41**	0.14	0.80
a7045			-0.13		0.13	**0.60**	**0.35**	0.37
a10647			0.12				**0.86**	0.26
a10092						0.24	**0.80**	0.20
a9053					-0.11		**0.77**	0.42
a10633					0.11	0.16	**0.75**	0.35
a8051					-0.22		**0.56**	0.64
a9518		-0.12					**0.56**	0.68
a8251				-0.12	-0.22		**0.48**	0.71
a6011				-0.14	-0.18		**0.37**	0.80
a12038					0.28	**0.41**	**0.49**	0.35
a11017					**0.33**	0.29	**0.39**	0.52
a28008			0.27	-0.11			-0.12	**0.86**

Table II. The factor loading of 85 stations from the factor analysis.

Factor 1	Coefficients	p-value
b_{01}	32.99	< 0.0005
b_{11}	-0.005900	< 0.0005
b_{21}	0.000015	< 0.0005
b_{31}	-1.61E-08	0.00001
b_{41}	7.98E-12	0.00021
b_{51}	-1.89E-15	0.00150
b_{61}	1.71E-19	0.00657

Factor 2	Coefficients	p-value
b_{02}	35.8	< 0.0005
b_{12}	-0.006000	0.03730
b_{22}	0.000012	< 0.0005
b_{32}	-1.04E-08	< 0.0005
b_{42}	3.63E-12	< 0.0005
b_{52}	-4.50E-16	< 0.0005
b_{62}	-1.14E-19	< 0.0005

Factor 3	Coefficients	p-value
b_{03}	33.63	< 0.0005
b_{13}	-0.00780	0.00869
b_{23}	0.000024	0.00292
b_{33}	-2.87E-08	0.00394
b_{43}	1.54E-11	0.00743
b_{53}	-3.88E-15	0.01496
b_{63}	3.68E-19	0.02908

Factor4	Coefficients	p-value
b_{04}	28.97	< 0.0005
b_{14}	-0.004330	0.0490
b_{24}	0.000014	0.0202
b_{34}	-1.84E-08	0.0155
b_{44}	1.06E-11	0.0166
b_{54}	-2.80E-15	0.0219
b_{64}	2.76E-19	0.0326

Factor 5	Coefficients	p-value
b_{05}	27.58	< 0.0005
b_{15}	0.004700	0.0052
b_{25}	-8.25E-06	0.0135
b_{35}	5.94E-09	0.0275
b_{45}	9.98E-12	0.0480
b_{55}	-3.10E-15	0.0312
b_{65}	3.53E-19	0.0205

Factor 6	Coefficients	p-value
b_{06}	29.39	< 0.0005
b_{16}	-0.011960	0.01187
b_{26}	0.000037	0.00441
b_{36}	-4.56E-08	0.00383
b_{46}	2.57E-11	0.00491
b_{56}	-6.74E-15	0.00757
b_{66}	6.68E-19	0.01238

Factor 7	Coefficients	p-value
b_{07}	23.55	< 0.0005
b_{17}	-0.001810	< 0.0005
b_{27}	0.000012	< 0.0005
b_{37}	-1.88E-08	0.04850
b_{47}	1.11E-11	0.04830
b_{57}	-2.88E-15	0.04020
b_{67}	2.73E-19	0.04420

Table III. Sixth order polynomial regression coefficients of 7 factors.

Appendix4

Chiang Mai J. Sci. 2015; 42(X) :1-11
http://epg.science.cmu.ac.th/ejournal/
Contributed Paper

Clustering and Forecasting Maximum Temperature of Australia

Wandee Wanishsakpong*[a], Khairil Anwar Notodiputro [b] and Nittaya McNeil [c]
[a] Department of Statistics, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand.
[b] Department of Statistics, Faculty of Mathematics and Science, Bogor Agricultural University, Bogor 16668, Indonesia.
[c] Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Pattani 94000, Thailand.
*Author for correspondence; e-mail: w.wandee43@gmail.com; khairilnotodiputro@gmail.com

Received: 8 April 2015
Accepted: 20 August 2015

ABSTRACT

The objectives of this paper were to investigate the trend and patterns of maximum temperature data in Australia as well as to forecast maximum temperature during the period of 2013-2015. Data obtained from the Australia Bureau of Meteorology (BOM) involved maximum monthly temperature from 85 of 700 stations during the period of 1970 to 2012. These 85 stations were randomly chosen in the same way as was previously done by Australia's Reference Climate Station Network, which took into account some stations had incomplete records due to weather phenomena and changing reporting requirements. Moreover, these 85 stations were spread approximately evenly over Australia and had more complete data than other stations during this period. Missing values in the data were estimated using a regression model accounting for information from the nearest stations as well as the time periods. The dimensions of the dataset, i.e. number of stations and time periods were of considerable magnitude, hence factor analysis was utilized to reduce the dimensions of the dataset. The resulting factor loadings were used as the input for cluster analysis based on complete linkage methods using Euclidian distances. Cluster analysis produced four clusters of stations. For each cluster, a quartic trend model with 3rd order time lag was fitted. It was demonstrated in this paper that the forecasted maximum monthly temperature during the period of 2013-2015 was decreasing. This model is most effective in forecasting maximum temperatures over relatively short multiple-year periods. Future research should focus techniques for more effectively modeling longer-term time periods.

Keywords: Australia, Cluster Analysis**,** Factor analysis, Linear Regression Model, Time-lags, Autocorrelation Function

1. INTRODUCTION

The characteristics of climate change are determined by the variation of globally averaged surface temperature [1] resulting from rainfall, drought, raise of sea level, melting of snow and ice, the El Niño Southern Oscillation [2, 3] and the increase of carbon dioxide (CO_2) from human activities [2]. The climate of various regions in the world has changed in different ways and has presented distinctive patterns of changes [4].

Australia is a continent which has remarkable geographical differences. Climate change in different regions of the continent may also reflect these geographical differences. Since 1910, the average temperature on this continent has increased by approximately 0.8^0C [5] and about 0.32^0C from 1981 to 2005 [6]. The mean of maximum temperature has increased 0.9^0C since 1950. However, the north west has exhibited a trend toward cooling during the summer, and a cooling trend in the south coast of western Australia has been evident in most seasons [7]. Moreover, the mean of minimum temperatures has increased over almost all of Australia except in some parts in the north west as mentioned in Nicholls [7].

In large regions, the possibility of having incomplete data is higher than in small regions since different temperatures are averaged together [4]. There have been studies carried out by many researchers demonstrating methods to investigate the characteristics and classification of temperature variability including computer simulation modeling [8], Empirical orthogonal function [9] and various statistical techniques such as principal component analysis [10] and factor analysis [11, 12]. In addition, cluster analysis is another method suitable for grouping regions together based on similarity of their climate signals [4, 13].

In this study, the limitation of factor analysis is that some of the stations are not clearly separated. Therefore, this study was conducted using similar approaches as the above mentioned studies, combining factor analysis and cluster analysis prior to the modeling of monthly maximum temperatures being carried out. We redefined the climate division, which is usually based on geographic positions, based on the factor loadings in combination with cluster analysis measuring the similarity of the loadings. Furthermore, the trends and patterns of maximum temperature data were investigated as well as forecasting of the temperature for the period of 2013-2015 within these clusters by polynomial trend models with time lags. The forecasting method was based on historical data. Thus, there was no other usage of weather parameters in this study. This method has been known as a univariate time series technique.

2. DATA MANAGEMENT AND STUDY AREAS

Daily maximum temperatures were obtained from the website http://www.bom.gov.au of the Australia Bureau of Meteorology (BOM) for over 700 weather stations across the country. These stations have different periods of observation. In this study, an area-based sample of 85 stations was randomly selected and these stations are distributed across Australia. The selection of these 85 stations was in line with Australia's Reference Climate Station Network (RCS) and the Australia Climate Observations Reference Network Surface Air Temperature (ARORN-SAT). These stations have approximately 80% complete data from 1970 to 2012. Prior to the analysis, missing data were estimated using linear regression models [14] and then the maximum monthly temperatures were defined as the highest daily temperature in a particular month for analysis in this study. (Figure 1) shows the distribution of 85 sample stations across the country. This figure also indicates the percentage of missing data in each station.

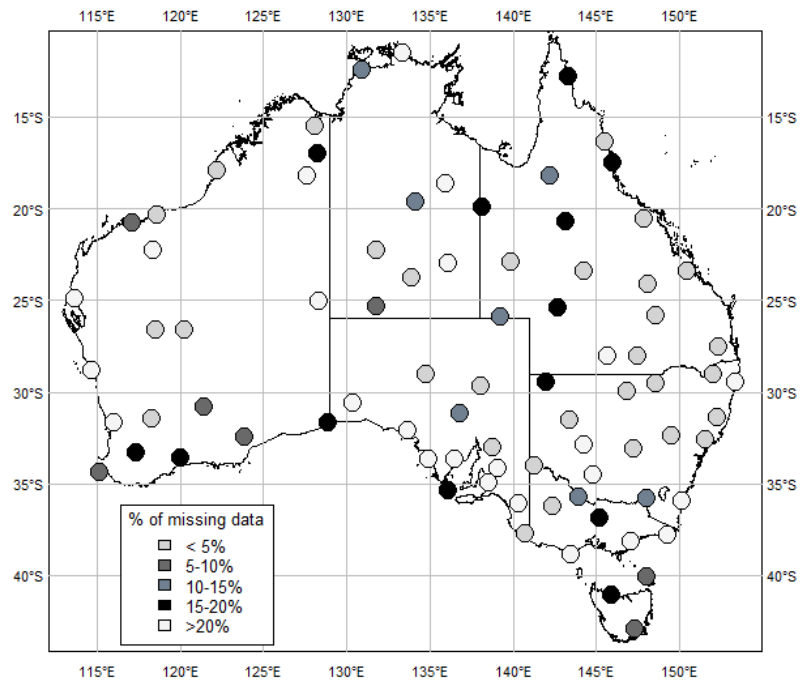


Figure 1. An area-based sample of 85 stations with indication of percentage missing data in each station. The state boundaries of Australia are also shown.

3. METHODS

As shown earlier there were missing data in several stations (see Figure 1). A linear regression model was fitted to estimate the missing data in 85 stations. Let *t* denote index for time and *k* denote index for station, the model takes the form

$$Y_{tk} = \tau_0 + \tau_1 Y_{t,k-} + \tau_2 Y_{t,k+} + \tau_3 Y_{t-1,k} + \tau_4 Y_{t+1,k} + \epsilon_{tk} \qquad (1)$$

for t =1, 2, … , 516 and k = 1, 2, …, 85, where Y_{tk} denotes the maximum temperature at month t of station k, $Y_{t,k-}$ is a maximum temperature at month *t* for the nearest station from station k, $Y_{t,k+}$ is the maximum temperature at month t for the second nearest station from station k, $Y_{t-1,k}$ is the maximum temperature at month *t*-1 for station *k*, and $Y_{t+1,k}$ is the maximum temperature at month *t*+1for station *k*.

Random samples of 10 stations were used to estimate missing values. The details of configuration for Y_{tk}, $Y_{t,k-1}$ and $Y_{t,k+1}$ are displayed in (Figure 2).

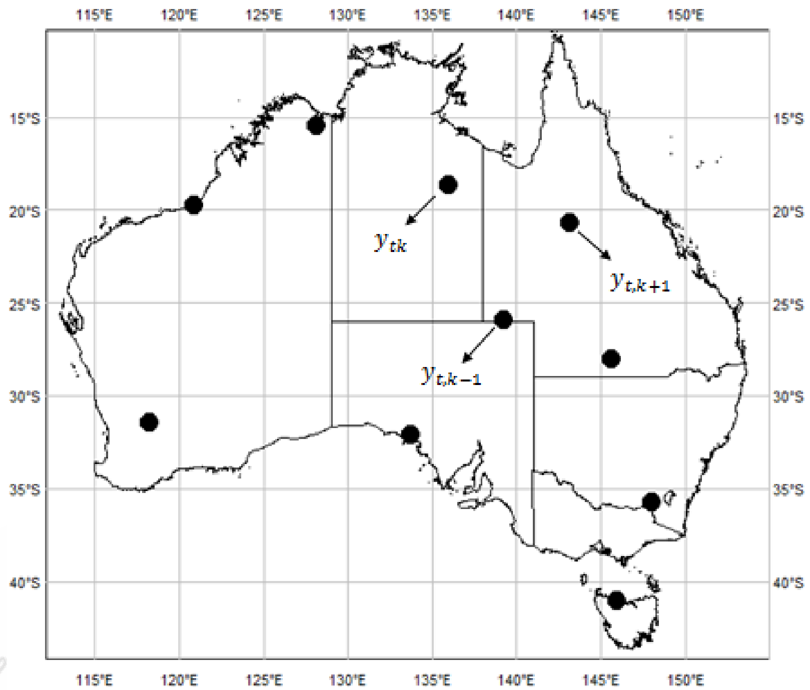


Figure 2. A random sample of 10 stations and example of configuration of Y_{tk}, $Y_{t,k-1}$ and $Y_{t,k+1}$.

The corresponding regression parameters of the models are denoted by τ_0, τ_1, τ_2, τ_3, and τ_4, whereas the error terms are denoted by ϵ_{tk}. This model assumes that the maximum monthly temperature at a particular month and station can be predicted using the previous and following month temperatures as well as the first and second nearest station temperatures at the predefined month. The model combines time series and spatial information to estimate the missing data. Adjustment has to be taken for the initial and end of times as well as for the edges of station locations [15].

A random sample of 10 stations was utilized to fit a model that relates temperature at month t and temperatures of neighboring stations as well as temperatures at month

(t-1) and (t+1). Using this model the missing values in all 85 stations were estimated with accuracy measured by adjusted R^2.

The dimensions of the data, i.e. number of stations and time period were quite large, hence factor analysis was applied to reduce the dimension. A factor analysis model was employed with m factors (m<85), denoted by $f_{1,} f_2, \dots, f_m$, and takes the form

$$y_{tk} = \mu_k + \sum_{p=1}^{m} \gamma_{kp} f_{kp} \quad (2)$$

where y_{tk} is maximum temperature at month t of station k, μ_k is the average across all times for station k, γ_{kp} is the factor loading at station k for factor p, and f_p is the p^{th} common factor [16].

The highest inter-correlated stations, i.e., those with similar maximum temperature, were grouped into factors by maximizing the likelihood of the covariance matrix. These factors were identified using factor loadings. The loadings were controlled by rotating the factors using the varimax method [17]. The cut-off value for factor loadings was 0.3 [18].

Factor analysis provides factor loadings which are correlation between stations and time periods. Based on the resulting factor loadings, cluster analysis was applied to cluster the 85 stations into several geographic groupings. The number of clusters was determined by the level of similarity among stations. This technique is concerned with exploring factor loadings to assess whether or not they can be summarized meaningfully in terms of a relatively small number of groups of stations which resemble each other and which are different in some respects from stations in other clusters. The complete linkage method was used in the analysis based on similarity measures [19].

For each cluster of stations, a quartic trend model combined with the 3rd order time lag was fitted to account for the trend and to predict the maximum monthly temperature. The model takes the following form,

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \delta_1 Y_{t-1} + \delta_2 Y_{t-2} + \delta_3 Y_{t-3} + \epsilon_t \quad (3)$$

for t =1, 2, … , 516, where Y_t denotes the maximum monthly temperature at month t, and $\beta_0, \beta_1, \beta_2, \beta_{3,} \beta_4, \delta_1, \delta_2, \delta_3$ are the parameters of the model, whereas ϵ_t is the error term. All data analysis and graphical displayed were carried out using R [20].

4. RESULTS AND DISCUSSION

Maximum monthly temperature data for 85 Australian stations were studied. As mentioned earlier there were missing data in some stations. A linear regression model (1) was used to estimate these missing data by using a random sample of 10 stations out of the 85 stations. The resulting regression is as follows:

$$\hat{Y}_{tk} = -\,2.13 \; + 0.286Y_{t,k-} + 0.297Y_{t,k+} + 0.247\,Y_{t-1,k} + 0.240Y_{t+1,k}$$

This equation was used to provide estimates of all missing data in all stations. Accuracy of these estimates was indicated by the fact that all coefficients of the regression were significant (P-value < 0.05) and adjusted R-squared was 94.2%.

After the missing data were estimated, in (Figure 3), examples are shown of the patterns of maximum monthly temperatures in four selected stations out of 85 stations. These stations were located in the northern, southern, eastern and western region of Australia, respectively. The observed data were plotted by dots whereas the fitted values were drawn using curve lines from the regression equation (3). It is clearly demonstrated that the equation can estimate the patterns and trends of the time series data as shown by the adjusted R-squared for these stations. In addition, the model applies equally well to the other stations. In general, the pattern showed polynomial pattern with range was between 20^{0}C and 50^{0}C. This result was similar with Wanishsakpong and McNeil's study [21] in term of the general pattern but it was different in the range of maximum temperatures which was between 23^{0}C to 36^{0}C. These differences could be due to different times used in this study.

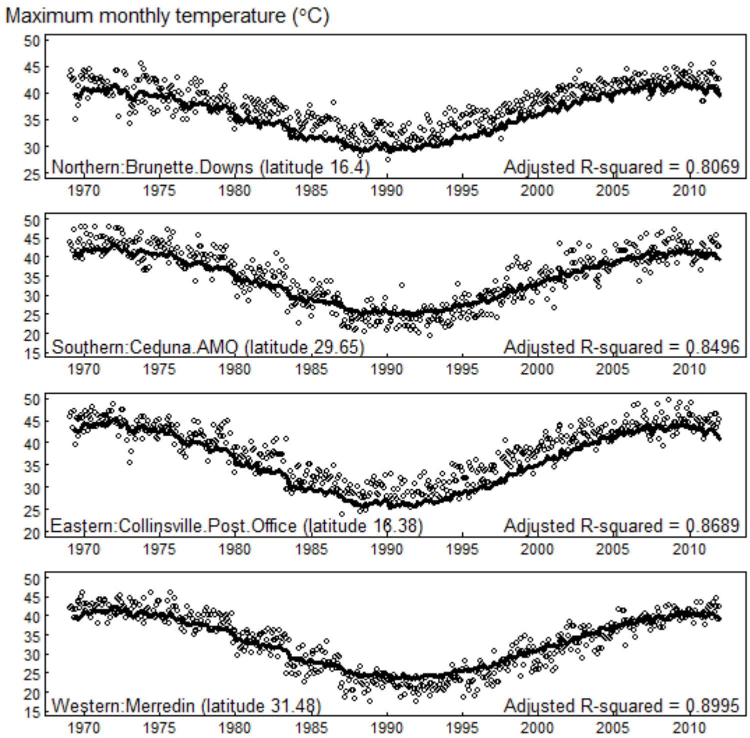


Figure 3. Observed maximum monthly temperatures with fitted values using the regression equation (3) for four stations. The horizontal axis is the year from 1970 to 2012.

Since there has been a similarity of patterns and trends of temperatures for the 85 stations, these stations were then placed into several clusters as follows. The data structure was arranged into a matrix of 85 stations as the rows and 516 times measurements as the columns. First the dimensions of the data were reduced using factor analysis as mentioned in section 2. Three factors were obtained with a cumulative variance of 90.4%. Factor loadings are presented in (Table 1). The factor loading clearly show that there are several stations having similar absolute factor loadings such as stations a5026, a4023, a13017, a15511, a48015, a31037, a44026, a43015, a48031, and a56032. In this situation it is not easy to group the 85 stations based on the results of factor analysis directly without using cluster analysis techniques.

Hence, the continuing use of cluster analysis was based on the resulting factor loading of 85x3 dimensions as the input for cluster analysis. To construct a hierarchical tree, the distance or dissimilarity between groups of stations was defined using the complete linkage and it was based on the Euclidian distance [19]. In general if you consider two groups of individuals A and B then the complete linkage between A and B is the greatest distance between an element in A and an element in B.

Stations	F1	F2	F3	Stations	F1	F2	F3	Stations	F1	F2	F3
a9053	0.88	-0.35	-0.22	a47016	0.73	-0.42	-0.48	a14015	-0.06	-0.78	0.00
a8251	0.87	-0.33	-0.22	a18069	0.73	-0.44	-0.43	a14401	0.37	-0.76	-0.21
a8051	0.86	-0.37	-0.22	a18012	0.72	-0.46	-0.45	a3003	0.32	-0.76	-0.09
a10647	0.85	-0.39	-0.26	a84070	0.71	-0.35	-0.49	a30045	0.41	-0.75	-0.42
a26021	0.83	-0.34	-0.38	a18110.1	0.71	-0.48	-0.42	a37010	0.44	-0.75	-0.40
a10633	0.83	-0.40	-0.29	a4035	0.71	-0.58	-0.31	a15135	0.49	-0.74	-0.39
a6011	0.83	-0.30	-0.19	a11003	0.71	-0.45	-0.38	a36031	0.43	-0.71	-0.49
a91009	0.82	-0.26	-0.35	a49019	0.70	-0.45	-0.51	a33013	0.46	-0.70	-0.42
a10092	0.82	-0.43	-0.31	a50052	0.70	-0.43	-0.55	a38003	0.48	-0.69	-0.48
a9518	0.81	-0.18	0.02	a5026	0.69	-0.59	-0.35	a15602	0.51	-0.67	-0.48
a18115	0.81	-0.25	-0.38	a62013	0.69	-0.42	-0.56	a15528	0.56	-0.66	-0.43
a99005	0.81	-0.27	-0.40	a4032	0.68	-0.59	-0.31	a38024	0.51	-0.64	-0.53
a25507	0.80	-0.39	-0.42	a17031	0.68	-0.50	-0.51	a15590	0.55	-0.64	-0.48
a88109	0.79	-0.37	-0.43	a46043	0.68	-0.48	-0.53	a35065	0.49	-0.64	-0.51
a72043	0.79	-0.36	-0.46	a16001	0.68	-0.49	-0.51	a4019	0.59	-0.64	-0.29
a90015	0.79	-0.31	-0.41	a51039	0.66	-0.46	-0.56	a39083	0.48	-0.64	-0.46
a78077	0.79	-0.40	-0.44	a13017	0.65	-0.60	-0.41	a38002.3	0.54	-0.62	-0.53
a23034	0.78	-0.40	-0.44	a46037	0.64	-0.51	-0.55	a28008	0.54	-0.60	-0.29
a94029	0.78	-0.33	-0.40	a16007.1	0.64	-0.53	-0.50	a43015	0.52	-0.57	-0.56
a85072	0.78	-0.34	-0.46	a15511	0.62	-0.60	-0.45	a60085	0.51	-0.51	-0.62
a24511	0.77	-0.40	-0.46	a48015	0.61	-0.50	-0.58	a61250	0.59	-0.47	-0.61
a12038	0.76	-0.47	-0.36	a31037	0.61	-0.59	-0.34	a40082	0.48	-0.55	-0.59
a80023	0.76	-0.40	-0.47	a32025	0.60	-0.50	-0.26	a44010	0.54	-0.55	-0.59
a18014	0.76	-0.41	-0.46	a44026	0.58	-0.55	-0.57	a48031	0.58	-0.51	-0.59
a7045	0.76	-0.51	-0.35	a29012	0.28	-0.85	-0.32	a56032	0.58	-0.49	-0.59
a11017	0.76	-0.46	-0.35	a1013	0.36	-0.84	-0.27	a69018	0.49	-0.39	-0.58
a19062	0.75	-0.43	-0.47	a2032	0.39	-0.82	-0.32	a58012	0.26	-0.53	-0.56
a13012	0.74	-0.52	-0.36	a15085	0.42	-0.78	-0.36	**Var**	36.39	24.48	16.01
a75031	0.74	-0.42	-0.50	a2012	0.45	-0.78	-0.35	**% Var**	0.428	0.288	0.188

Table 1. The first three factor loadings of 85 stations with cumulative variances of 90.14%

Based on the similarity of the factor loadings, the results of cluster analysis showed the 85 stations were grouped into four clusters. The description of these clusters is shown in (Table 2).

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster 1	39	0.6486	0.1183	0.2541
Cluster 2	14	0.2448	0.1165	0.3322
Cluster 3	12	0.1615	0.1083	0.2489
Cluster 4	20	0.3055	0.1121	0.2578

Table 2. Description of resulting clusters based on factor loadings

(Figure 4) provides a visualization of the clusters of the 85 stations on the map of Australia. It can be seen in the figure that the corresponding geographical regions for the four clusters are north-west with some parts of the central and south regions, the south-eastern region including some part of the northern region (Cluster 1); the lower boundary of the south-western and, southern region including Tasmania (Cluster 2); the north (Cluster 3) and some parts of the central and eastern regions (Cluster 4).

Factor analysis and cluster analysis were similar methods used for classification of the data. This statistical technique was used in other studies as mentioned in Mahlstein and Knutti [4], Wanishsakpong et al.[12] and Unal et al.[13] but the result from classified group were different because of different time period and region studied. Moreover, time series data for each cluster were studied separately and forecasting models were developed for future temperatures in each cluster.

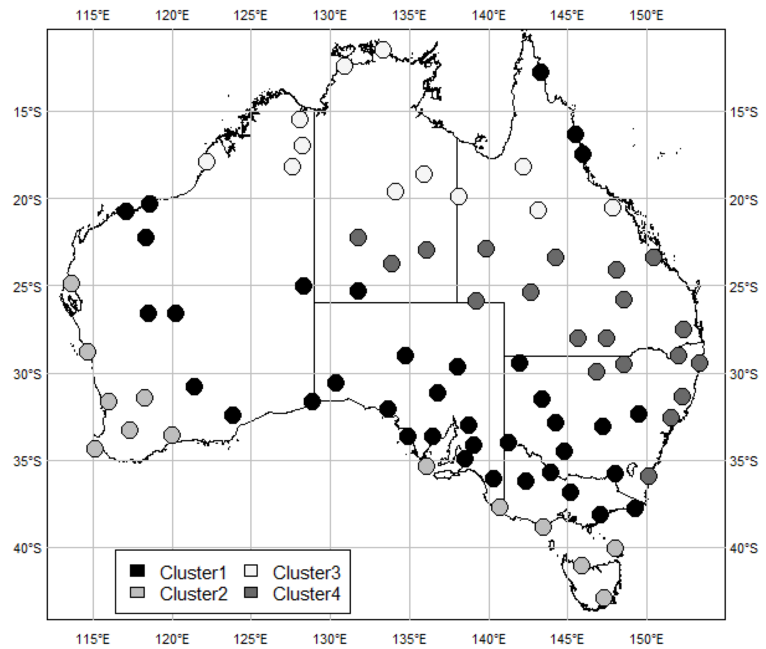


Figure 4. Cluster analysis divides the stations into 4 groups

In each cluster, the data showed distinct periodic variation. The maximum monthly temperatures ranged from 20^0C to 45^0C over the 43 years. Temperatures decreased around from 1970 to 1990 and gradually have increased until year 2012. A quartic trend model with 3rd order time lag was fitted to the time series data as shown in (Figure 5), the goodness of fit was high. The regression equation for each cluster is shown in (Table 3). The quantile-quantile (Q-Q) plot shown in (Figure 6) indicates that the residuals from the model are normally distributed and the residuals are independent as shown in (Figure 7).

	Coefficients			
	Cluster1	**Cluster2**	**Cluster3**	**Cluster4**
Intercept	17.9900	16.8900	15.5100	19.7000
t	0.0647	0.0663	0.0293	0.0498
t^2	-0.0010	-0.0010	-0.0004	-0.0009
t^3	3.36×10^6	3.04×10^6	1.51×10^6	2.98×10^6
t^4	-3.19×10^9	-2.78×10^9	-1.49×10^9	-2.91×10^9
y_{t-1}	0.1842	0.1545	0.1499	0.1631
y_{t-2}	0.1932	0.1850	0.2597	0.2208
y_{t-3}	0.1615	0.1777	0.1800	0.1000
Adjusted R^2	0.9287	0.9251	0.8694	0.8924

Table 3. Regression coefficients for each cluster

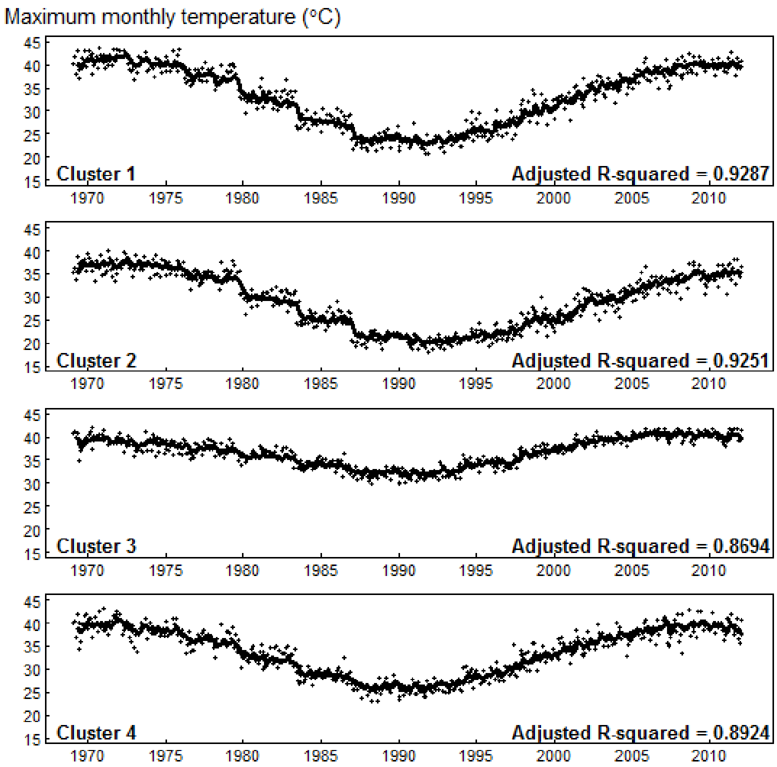


Figure 5. The fitted values of maximum monthly temperatures in each cluster based on model (3). The horizontal axis is the year 1970 to 2012.

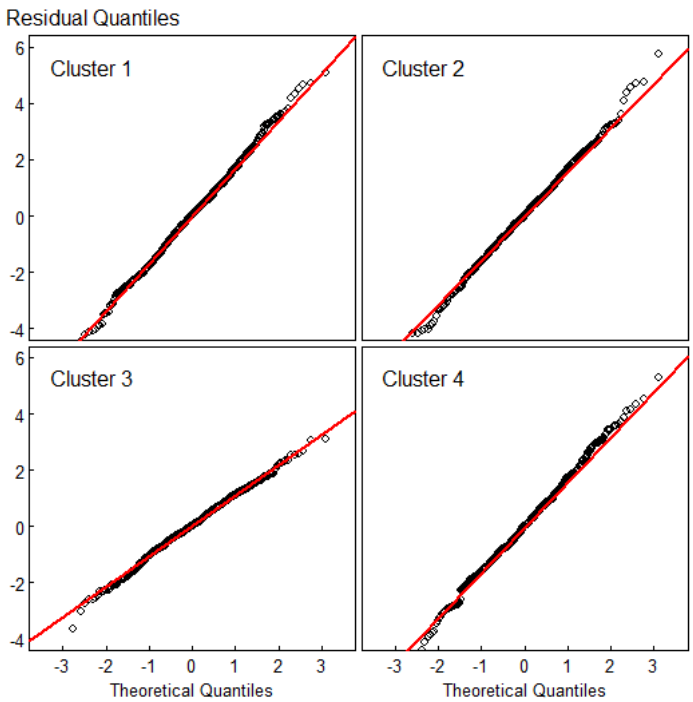


Figure 6. Residual Q-Q plots (quantiles plots) for each cluster

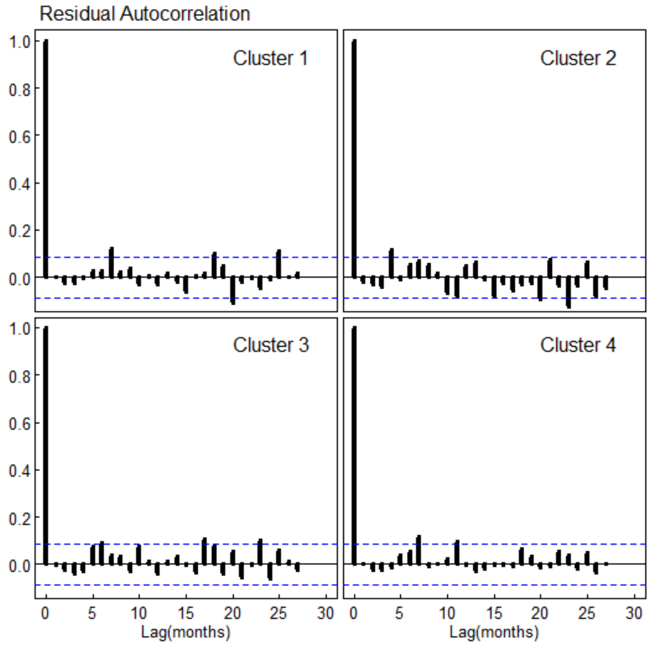


Figure 7. Residual auto-correlation function (ACF) plots for each cluster

The trend of temperature changes in each cluster was found to occur over 10 year periods. Therefore, the most accurate forecast that could be done by the model should be a short term forecast. In this case, the forecast of monthly maximum temperatures is shown from year 2013 up to 2015 or 36 months. The forecast have gradually decreased over 2013-2015 in each clusters as shown in the curved lines highlighted in (Figure 8). The highlight showed 95% confidence intervals (CI) of the maximum monthly temperatures predictions for each cluster. The 95% CI ranged from about 27^0C-44^0C, 26^0C-40^0C, 31^0C-42^0C and 26^0C-42^0C in cluster 1, 2, 3 and 4, respectively. Clearly, the confidence interval of cluster 3 is narrower than other clusters and the temperature changes are higher than for other clusters (ranging from 32^0C-43^0C) which occurred in the north. In addition, the highest temperatures approximately 45^0C occurred in cluster 1 which is in a desert area.

The polynomial model should be used carefully to forecast for long period of times [22, 23]. Benestad [22] suggested that the polynomial regression model can be considered to account for the detailed description of trends but should be used carefully to forecast for long periods of time due to the uncertainty of other factors in the future.

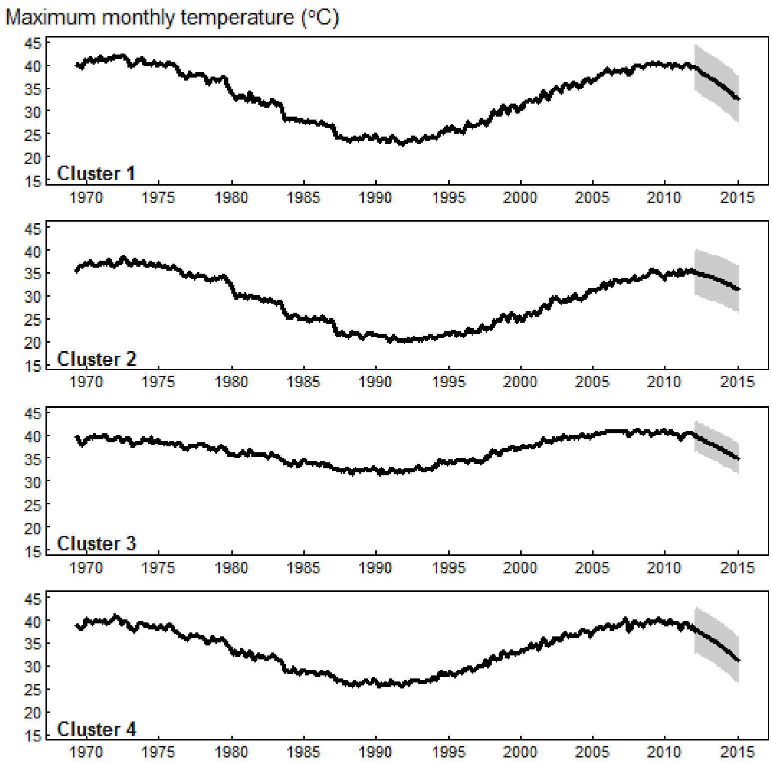


Figure 8. Forecasting the maximum monthly temperature from 2013 to 2015. The curved lines are trend estimation base on model (3). The highlighted lines are 95% confidence interval of prediction.

5. CONCLUSION

The trends and patterns of monthly maximum temperature data for 85 Australian stations from 1910 to 2012 were investigated. A linear regression model was used to estimate missing values in all 85 stations by using the information of neighboring stations and previous as well as following year temperatures of the same month. Factor analysis and cluster analysis were combined to group the number of station from 85 stations into 4 clusters corresponding to geographical regions. Cluster 1 consisted of 39 stations, most of them located in the southern part of the continent, but a few of them located in the northern and central regions of Australia. Cluster 2 consisted of 14 stations, distributed at the boundary of the south-west and south of Australia. Cluster 3 consisted of 12 stations which were located in northern Australia whereas cluster 4 consisted of 20 stations and distributed in parts of the central and eastern region of Australia.

The patterns of monthly maximum temperature for the four clusters were found to be a polynomial pattern. The overall temperature trend was a decrease between 1970 and 1990, afterwards continuously increasing until 2012. A quartic trend model combined with 3^{nd}order time lag has reasonably well been fitted to the data in each cluster. This model was used to forecast the maximum temperatures from 2013 to 2015. The result of forecasting showed a 95% confidence interval (CI) of the maximum monthly temperature prediction ranged from about 27^0C-44^0C, 26^0C-40^0C, 31^0C-42^0C and 26^0C-42^0C in cluster 1, 2, 3 and 4, respectively.

Future work should evaluate the performance of other models for missing data estimation. Furthermore, we should develop models that enable us to forecast for much longer periods. Other models such as sine and cosine models need to be considered.

REFERENCES

[1] Jones P.D., New M., Parker D.E., Martin S., and Rijor I.G., Surface air temperature and its changes over the past 150 years. *Review of Geophysics*., 1999; **37**:173-199.

[2] CSIRO., Climate Change in Australia: observed changes and projection; Available at:htpp://www.climatechangeinaustralia.gov.au. Accessed 1November 2014.

[3] Hughes L., Climate change and Australia: trend, projections and impacts, *Austral Ecology.*, 2003; **28**:423- 443.

[4] Mahlstein I., and knutti R., Regional climate change patterns identified by cluster analysis, *Clim Dyn.,*2009; **35**:587-600.

[5] Collins D., Annual temperature summary: Australia records warmest decade, *Climate Change Newsl.,* 2000; **12**:6.

[6] Collins D.A., Della-Marta P.M., Plummer N., and Trewin B.C., Trend in annual frequencies of extreme temperature events in Australia, *Aus Met Mag*., 2000; **49**:277-292.

[7] Nicholls N., Detecting and attributing Australian climate change: a review, *Aust Met Mag.,* 2006; **55**:199-221.

[8] Johannessen O.M., Bengtsson L., Miles M.N., Kuzmina S.I., Semenov V.A., Alekseev G.V., Nagurnyi A.P., Zakharov V.F., Bobylev L.P., Pettersson L.H., Hasselmann K., and Cattle H.P., Arctic climate change: observe and modeled temperature and sea-ice variability, *Tellus*., 2003; **56**:328-341.

[9] Semenov V.A., Structure of temperature variability in the high latitudes of the northern hemisphere, *Atmos Oceanic Phys*., 2007; **43**:744-753.

[10] Jones D.A., Characteristics of Australian land surface temperature variability, *Theor Appl Climatol*.,1999; **63**:11-31.

[11] McNeil N., and Chooprateep S., Modeling sea surface temperatures of the North Atlantic ocean, *Theor Appl Climatol.,* 2014; **116**:11-17.

[12] Wanishsakpong W., Luo K., and Tongkumchum P., Earth surface temperature change above latitude 45 degrees North from 1973 to 2008, *CMU J Nat Sci.,* 2014; **13**:247-257.

[13] Unal Y., Kinda T., and Karaca M., Redefining the climate zone of turkey using cluster analysis, *Int.J.Climatol*., 2003; **23**:1045-1055.

[14] Yaffee R., *Introduction to time series analysis and forecasting with applications of SAS and SPSS*: Brooklyn, New York, 1999.

[15] Kedem B., and Fokianos K., *Regression Model for Time series Analysis*: Willy, Canada, 2002.

[16] Johnson D.E., Applied Multivariate Methods for Data Analysis: Duxbury, USA, 1998.

[17] Venables W.N., and Ripley B.D., *Modern Applied Statistics with S*, Springer, Queensland, 2002.

[18] Hair J.F., Anderson R.E., Tatham R.L., and Black W.C., *Multivariate Data Analysis*: Prentice Hall International (UK) Limited, London, 1998.

[19] Husson F., Le S., and Pages J., *Exploratory multivariate analysis by example using R*: Chapman & Hall/CRC, London, 2011.

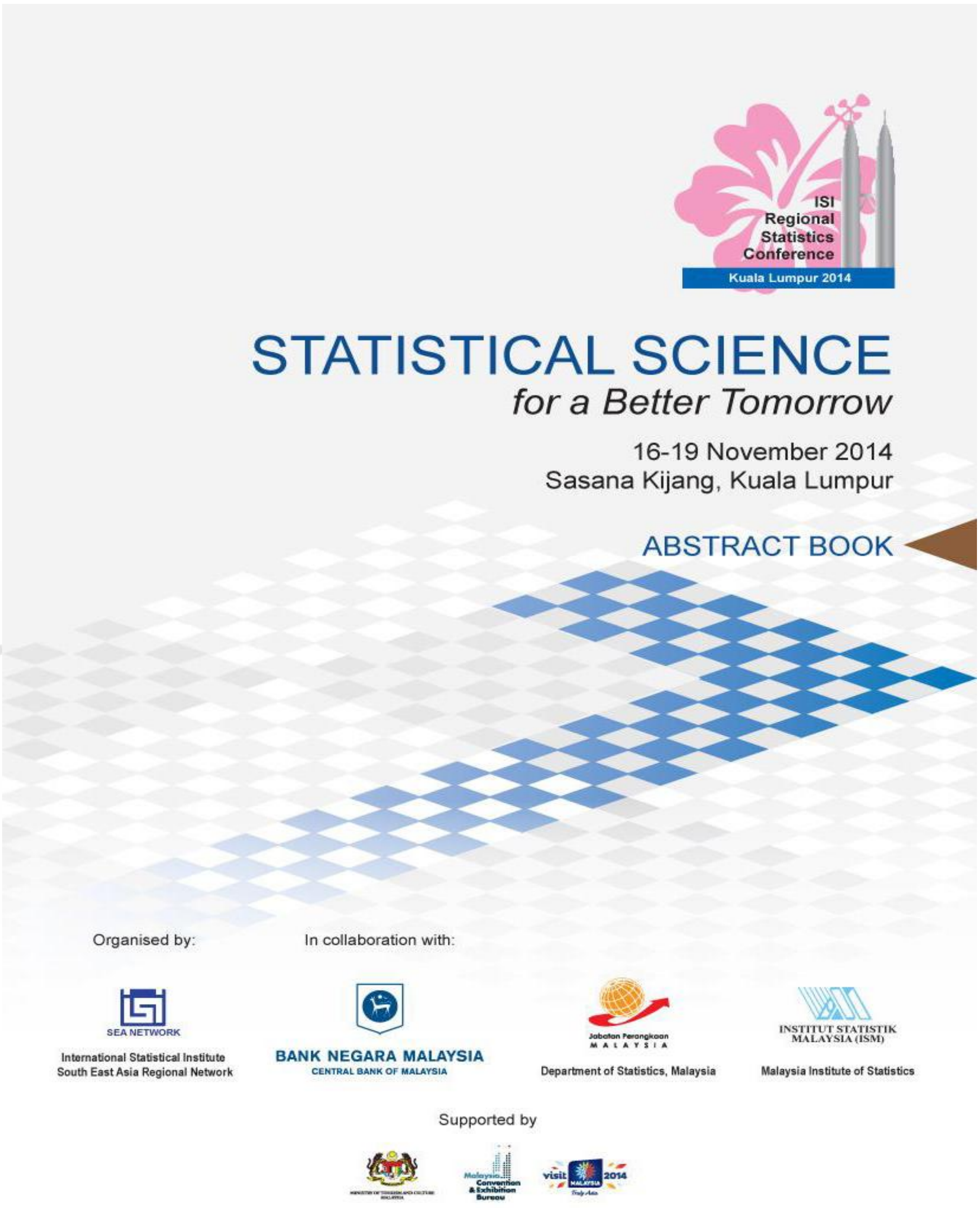
[20] R development Core Team., R: A language and environment for statistical computing. R Foundation for Statistical computing; Available at: htpp://www.R-project.org. Accessed 1 November 2009.

[21] Wanishsakpong W., and McNeil N., Modeling of daily maximum temperatures over Australia from 1970 to 2012, *Met apps*, 2015 (Accepted from publication)

[22] Benestad R., What can present climate models tell us about climate change ?, *Climate chang*., 2003; **59**:311-331.

[23] Woodward W., and Gray H.L., Global warming and the problem of testing for trend in time series data, *J climate*., 1992; **6**:953-962.

Appendix 5


ISI
Regional
Statistics
Conference
Kuala Lumpur 2014
STATISTICAL SCIENCE
for a Better Tomorrow
16-19 November 2014
Sasana Kijang, Kuala Lumpur
ABSTRACT BOOK
Organised by:
In collaboration with:
SEA NETWORK
International Statistical Institute
South East Asia Regional Network
BANK NEGARA MALAYSIA
CENTRAL BANK OF MALAYSIA
Jabatan Perangkaan
MALAYSIA
Department of Statistics, Malaysia
INSTITUT STATISTIK
MALAYSIA (ISM)
Malaysia Institute of Statistics
Supported by
Malaysia Convention & Exhibition Bureau
visit MALAYSIA 2014

ISI-RSC 2014 ABSTRACT BOOK

Pattern Classification for Earth Surface Temperature Changes in the Arctic

Wandee Wanishsakpong*

Prince of Songkla University, Pattani, Thailand : One_D7@Hotmail.Com

Nittaya Mcneil

Prince of Songkla University, Pattani, Thailand: Nchirtki@Gmail.Com

The trends and patterns of monthly seasonally adjusted temperatures in the arctic zone were examined using statistical methods. The data were obtained from temperatures recorded between 1973 and 2013 from the climate research unit. The data was filtered with a second order autoregressive process to remove autocorrelations. Factor analysis was used to classify regions with similar temperature changes and identified twelve factors. The temperature in the twelve factors showed a significant increase. Large temperature increases (at least 0.19oc) were found in the North Siberia, South Siberia and part of the Arctic Ocean. Increases in North Canada, Alaska, North Pacific Ocean, East Siberia, Iceland, Norway and Sweden were moderate (0.15oc to 0.16oc). The North East Canada, Greenland and its surrounding North Atlantic Ocean and Arctic Ocean had increases below 0.15oc.

Key Words: Time Series Analysis, Autocorrelation, Linear Regression Model, Factor Analysis

ISI Regional Statistics Conference Kuala Lumpur 2014

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

Wandee Wanishsakpong

Contributed to the International Statistical Institute
Regional Statistics Conference 2014

Marzunisham Omar
Chairman, Local Organising Committee

Organised by: SEA NETWORK International Statistical Institute South East Asia Regional Network

In collaboration with: BANK NEGARA MALAYSIA CENTRAL BANK OF MALAYSIA; Jabatan Perangkaan MALAYSIA; MALAYSIA INSTITUTE OF STATISTICS

ISI Regional Statistics Conference Kuala Lumpur 2014

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

Wandee Wanishsakpong

Attended the International Statistical Institute
Regional Statistics Conference 2014

Marzunisham Omar
Chairman, Local Organising Committee

Organised by: SEA NETWORK International Statistical Institute South East Asia Regional Network

In collaboration with: BANK NEGARA MALAYSIA CENTRAL BANK OF MALAYSIA; Jabatan Perangkaan MALAYSIA; MALAYSIA INSTITUTE OF STATISTICS

Vitae

Name: Miss. Wandee Wanishsakpong

Student ID: 5520330014

Educational Attainment:

Degree	Name of institution	Year of Graduation
B.Ed. (Mathematics)	Srinakharinwirot University (SWU)	2003
M.S. (Applied Statistics)	National Institute of Development Administration (NIDA)	2005

Scholarship Awards during Enrolment

The Domestic Doctoral Scholarship, Faculty of Science, Kasetsart University, Thailand

Graduate School Research Support Funding for thesis, Prince of Songkla University, Thailand

Work-Position and address:

Lecturer

Department of Statistics, Faculty of Science

Kasetsart University, Thailand

List of Publication:

Wanishsakpong, W., Luo, K., and Tongkumchum, P. 2014. Earth Surface Temperature Changes above Latitude 45 Degrees North from 1973 to 2008. Chiang Mai University Journal of Natural Science. 13(3):247-257.

Wanishsakpong, W., and McNeil, N. 2015. Modeling of daily maximum temperatures over Australia from 1970 to 2012. Metrological Application (Accepted from publication).

Wanishsakpong, W., Notodiputro, K. N., and McNeil, N. 2015. Clustering and Forecasting Maximum Temperature of Australia. Chiang Mai University Journal of Science (Accepted from publication).

Wanishsakpong, W., McNeil, N., and Notodiputro, K. N. 2015. Trend and Pattern Classification of Surface Air Temperature Change in the Arctic Zone. Atmospheric Science Letter (Submitted to publication).

Proceedings:

Wanishsakpong, W., and McNeil, N. 2014. Pattern classification for Earth Surface Temperature changes in the Arctic. ISI Regional Statistics conference. Kuala Lumpur, Malaysia, November 16-19.