

โครงข่ายประสาทเทียมสำหรับการทำนายปริมาณฝนรายวันในภาคใต้ฝั่งตะวันออกของประเทศไทย



วรชมน ภูสกุลขจร

## ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2552

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

ARTIFICIAL NEURAL NETWORK FOR DAILY RAINFALL  
PREDICTION IN EASTERN COAST OF SOUTHERN  
THAILAND



Miss Wassamon Phusakulkajorn

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements  
for the Degree of Master of Science Program in Computational Science

Department of Mathematics

Faculty of science

Chulalongkorn University

Academic Year 2009

Copyright of Chulalongkorn University

Thesis Title                      Artificial neural network for daily rainfall prediction in  
   Eastern Coast of Southern Thailand

By                                      Miss Wassamon Phusakulkajorn


Field of Study                      Computational science

Thesis Advisor                      Professor Chidchanok Lursinsap, Ph.D.


Thesis Co-advisor                      Associate Professor Jack Asavanant, Ph.D.


---


Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree.

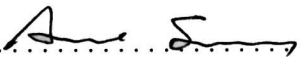
  
..... Dean of the Faculty of Science  
(Professor Supot Hannongbua, Ph.D.)


Thesis Committee

  
..... Chairman  
(Assistant Professor Vimolrat Ngamaramvaranggul, Ph.D.)

  
..... Thesis Advisor  
(Professor Chidchanok Lursinsap, Ph.D.)

  
..... Thesis Co-advisor  
(Associate Professor Jack Asavanant, Ph.D.)

  
..... Examiner  
(Assistant Professor Anond Sanidvongs, Ph.D.)

  
..... External Examiner  
(Chularat Tanprasert, Ph.D.)

วรรณมน กุ์สกุลขจร : โครงข่ายประสาทเทียมสำหรับการทำนายปริมาณฝนรายวันในภาคใต้ฝั่งตะวันออกของประเทศไทย. (ARTIFICIAL NEURAL NETWORK FOR DAILY RAIN-FALL PREDICTION IN EASTERN COAST OF SOUTHERN THAILAND) อ.ที่ปรึกษา  
 วิทยานิพนธ์หลัก : ศ.ดร. ชิดชนก เหลือสินทรัพย์, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม : รศ.ดร. จักรย์ อิศวานันท์, 62 หน้า.

การพยากรณ์น้ำฝนมีประโยชน์ในการดำเนินชีวิตของมนุษย์ในด้านการวางแผนต่างๆ เช่น การพยากรณ์การเกิดอุทกภัยและภัยแล้ง เป็นต้น ซึ่งปัญหาภัยธรรมชาติที่เกิดขึ้นก่อให้เกิดผลกระทบและความเสียหายต่อชีวิตและทรัพย์สินของประชาชนอย่างมากโดยเฉพาะพื้นที่เศรษฐกิจของประเทศไทย ดังนั้นจึงมีการศึกษา วิจัย คิดค้นพัฒนาวิธีการและเทคโนโลยีต่างๆ เพื่อนำมาใช้ในการพยากรณ์ปริมาณน้ำฝน แต่การพยากรณ์ให้มีความถูกต้องและทันต่อเหตุการณ์ทำได้ยาก เพราะการเปลี่ยนแปลงของสภาพอากาศเกิดขึ้นตลอดเวลาและเป็นแบบไม่คงที่ ดังนั้น ในงานวิจัยนี้จึงใช้วิธีการโครงข่ายประสาทเทียมซึ่งเป็นวิธีการหนึ่งที่มีประสิทธิภาพในการใช้ข้อมูลจากชุดการเรียนรู้ที่มีเพื่อใช้พยากรณ์ค่าที่มีความไม่แน่นอน

งานวิจัยนี้พัฒนาตัวแบบโดยใช้ Neural Network Toolbox ในโปรแกรม MatLab เพื่อทำการสร้างโครงข่ายประสาทเทียมแบบ Back-propagation และสอนโครงข่ายประสาทเทียมให้สามารถทำนายปริมาณฝนได้อย่างถูกต้อง แม่นยำ โดยใช้สองเทคนิควิธีในการเรียนรู้ของโครงข่ายประสาทเทียม เทคนิควิธีแรกได้แบ่งการเรียนรู้โครงข่ายประสาทเทียมตามช่วงปริมาณการตกของฝน – ช่วงฝนตกมากและช่วงฝนตกน้อย – โดยอาศัยข้อมูลทางอุตุนิยมวิทยาที่มีความสัมพันธ์กับปริมาณน้ำฝนเป็นข้อมูลเข้า ปัจจัยที่นำมาพิจารณา ได้แก่ ความกดอากาศ ความชื้น อุณหภูมิ ปริมาณเมฆ ความเร็วลม ทิศทางลม และเทคนิควิธีที่สองได้ใช้การแปลงเวฟเล็ต เพื่อแปลงข้อมูลที่ต้องการให้เป็นสัมพันธ์กับเวฟเล็ต จากนั้นจึงนำสัมพันธ์เวฟเล็ตที่ได้สอนให้โครงข่ายประสาทเทียมเรียนรู้โดยใช้ข้อมูลปริมาณน้ำฝนเท่านั้น ซึ่งเทคนิควิธีทั้งสองทดสอบในพื้นที่ศึกษาภาคใต้ฝั่งตะวันออกที่ประสบอุทกภัย ได้แก่ อำเภอท่าแซะ จังหวัดชุมพร อำเภอกาญจนดิษฐ์ จังหวัดสุราษฎร์ อำเภอเมือง จังหวัดนครศรีธรรมราช อำเภอเมือง จังหวัดพัทลุง และอำเภอหาดใหญ่ จังหวัดสงขลา ผลจากการทำนายพบว่าเทคนิควิธีโครงข่ายประสาทเทียมแบบใช้การแปลงเวฟเล็ต สามารถทำนายปริมาณน้ำฝนรายวันล่วงหน้าสี่วัน โดยมีความถูกต้องเท่ากับ  $R^2 = 0.8887$  และ  $RMSE = 4.2306$  มิลลิเมตร

ภาควิชา.....คณิตศาสตร์.....  
 สาขาวิชา.....วิทยาการคอมพิวเตอร์.....  
 ปีการศึกษา.....2552.....

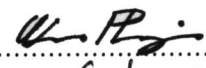
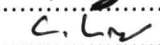
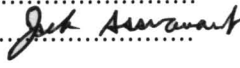
ลายมือชื่อนิติ.....  
 ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.....  
 ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม.....

# # 5072450023 : MAJOR COMPUTATIONAL SCIENCE

KEYWORDS : DAILY RAINFALL PREDICTION / ARTIFICIAL NEURAL NETWORK / WAVELET DECOMPOSITION / SOUTHERN THAILAND

WASSAMON PHUSAKULKAJORN : ARTIFICIAL NEURAL NETWORK FOR DAILY RAINFALL PREDICTION IN EASTERN COAST OF SOUTHERN THAILAND. THESIS ADVISOR : PROF. CHIDCHANOK LURSINSAP, Ph.D., THESIS CO-ADVISOR : ASSOC. PROF. JACK ASAVANANT, Ph.D., 62 pp.

Rainfall prediction generally requires reliable hydrological models as well as relevant information such as meteorological and geographical data. In this thesis, the prediction is, first, investigated by a conventional artificial neural network (ANN) model on accounts of the preceding events of rainfall data and climatological variables. Since precipitation behavior can be categorized into 2 time-periods; wet and dry period, ANN models for each time period are, then, implemented in order to improve network accuracy. A model based on artificial neural network (ANNs) and wavelet decomposition is another proposed learning tool predicting consecutive daily rainfalls on accounts of the preceding events of rainfall data only. In the combining model of ANN and wavelet decomposition, two sets of wavelet coefficients, for which one pattern represents detail information of rainfall data and the other acts as a smoothing filter, are extracted for the ANNs. A back-propagation neural network is used in the learning and knowledge extraction processes for all models. The methodologies are tested on rainfall data from five stations in the eastern coast of southern Thailand; Tha Sae district in Chumphon province, Kanchanadit district in Surat Thani province, Muang district in Nakhon Sri Thammarat province, Muang district in Phatthalung province and Hatyai district in Songkhla province. From the past historical records of Thai Meteorological Department and Royal Irrigation Department, these study area are vulnerable to flood disaster. The proposed model based on artificial neural network (ANNs) and wavelet decomposition is capable of forecasting daily rainfall up to 4 days in advance with accuracy of  $R^2 = 0.8887$  and  $RMSE = 4.2306$  mm.

Department : ....Mathematics..... Student's Signature : ..........  
 Field of Study : Computational science Advisor's Signature : ..........  
 Academic Year : .....2009..... Co-advisor's Signature : ..........

## ACKNOWLEDGEMENTS

I would like to express my deep appreciations to Professor Dr. Chidchanok Lursinsap, my advisor, and Associate Professor Dr. Jack Asavanant, my co-advisor, for their patience, help, comments and suggestions throughout the study. I am grateful to the advice given by my advisor and co-advisor who carefully reviewed the manuscript and suggested valuable improvements. Without their generosity, encouragement, and superb guidance, this thesis can not be achieved.

Special thanks are extended to the committees; Assistant Professor Dr. Vimolrat Ngamaramvaranggul, Assistant Professor Dr. Anond Snidvongs, Dr. Chularat Tanprasert, for their comments, constructive criticism and suggestions in their own ways in making the achievement of the thesis.

I am also indebted to the Development and Promotion of Science and Technology talents project (DPST), the CU Graduate School Thesis Grant, and National Research Council of Thailand for my financial support. The Advanced Virtual and Intelligence Computing Research Center (AVIC) is acknowledged for material supports in enabling me to accomplish this thesis. Furthermore, I would like to express my special gratefulness to Dr. Wattana Kanbua from Thai Meteorological Department and Mr. Ponchai Klinkajorn from Royal Irrigation Department for providing daily rainfall and climatological data in Southern Thailand.

I wish to thank friends of mine, Miss Paramitta Punwong, Miss Dhasida Sooksawat, and Miss Tassaneewan Laksanasophon, my best friends, for their emotional support, caring and encouragement. I have also had many generous assistances and entertainment from my colleagues of Computational Science discipline. Also, the members of AVIC center are thanked for great helping, effective cooperation and discussion.

Above all, I would like to express my gratitude to my family. The encouragement and many helpful suggestions from them were enable me get through the difficult times. For their unfailing love and support, the thesis is dedicated to them.

# Contents

|  |           |
|--|-----------|
| <b>Abstract (Thai)</b> . . . . .                     | <b>iv</b> |
| <b>Abstract (English)</b> . . . . .                  | <b>v</b>  |
| <b>ACKNOWLEDGEMENTS</b> . . . . .                    | <b>vi</b> |
| <b>List of Tables</b> . . . . .                      | <b>x</b>  |
| <b>List of Figures</b> . . . . .                     | <b>xi</b> |
| <b>I INTRODUCTION</b> . . . . .                      | <b>1</b>  |
| 1.1 Literature review . . . . .                      | 2         |
| 1.2 The objective . . . . .                          | 4         |
| 1.3 Outline . . . . .                                | 5         |
| <b>II THEORETICAL BACKGROUND</b> . . . . .           | <b>6</b>  |
| 2.1 Physical descriptions of precipitation . . . . . | 6         |
| 2.1.1 The Hydrological cycle . . . . .               | 6         |
| 2.1.2 Climatological Distribution . . . . .          | 7         |
| 2.1.3 Formation of precipitation . . . . .           | 8         |
| 2.2 Artificial Neural Network . . . . .              | 9         |
| 2.2.1 A Multilayer Feedforward Network . . . . .     | 10        |
| 2.2.2 Activation function . . . . .                  | 11        |
| 2.2.3 Learning algorithm . . . . .                   | 13        |
| 2.2.4 Resilient backpropagation . . . . .            | 15        |
| 2.3 Wavelet transform . . . . .                      | 15        |
| 2.3.1 Discrete wavelet transform . . . . .           | 16        |

|  |  |           |
|--|--|-----------|
| 2.3.2                                      | Convolution and Filter . . . . .               | 16        |
| 2.3.3                                      | Lowpass filter . . . . .                       | 16        |
| 2.3.4                                      | Highpass filter . . . . .                      | 17        |
| 2.3.5                                      | Multiresolution analysis . . . . .             | 17        |
| 2.3.6                                      | The Daubechies wavelet function . . . . .      | 18        |
| 2.3.7                                      | À trous wavelet transform . . . . .            | 20        |
| <b>III METHODOLOGY . . . . .</b>           |  | <b>22</b> |
| 3.1  | Study area . . . . .                           | 22        |
| 3.2  | Data refinement . . . . .                      | 24        |
| 3.3  | Data-preprocessing . . . . .                   | 24        |
| 3.3.1                                      | Wavelet decomposition . . . . .                | 24        |
| 3.3.2                                      | Linear Transformation . . . . .                | 27        |
| 3.4  | Data prediction . . . . .                      | 27        |
| 3.4.1                                      | Artificial Neural Network . . . . .            | 27        |
| 3.5  | Data post-processing . . . . .                 | 29        |
| 3.6  | Performance criteria . . . . .                 | 30        |
| 3.6.1                                      | Root Mean Squared Error : $RMSE$ . . . . .     | 30        |
| 3.6.2                                      | Coefficient of Determination : $R^2$ . . . . . | 31        |
| 3.6.3                                      | Correlation Coefficient : $C$ . . . . .        | 31        |
| <b>IV RESULTS AND DISCUSSION . . . . .</b> |  | <b>32</b> |
| 4.1  | Artificial neural network . . . . .            | 32        |
| 4.1.1                                      | Input and hidden layer configuration . . . . . | 32        |
| 4.1.2                                      | Activation function . . . . .                  | 33        |
| 4.1.3                                      | Period prediction . . . . .                    | 37        |



|                                      |   |           |
|--------------------------------------|---|-----------|
| 4.2                                  | Wavelet-transform based Artificial neural network . . . . . | 43        |
| 4.2.1                                | One-day forecasting . . . . .                               | 46        |
| 4.2.2                                | Many-day forecasting . . . . .                              | 48        |
| 4.3                                  | Discussion . . . . .  | 52        |
| <b>V CONCLUDING REMARK . . . . .</b> |   | <b>54</b> |
| <b>REFERENCES . . . . .</b>          |   | <b>56</b> |
| <b>APPENDIX . . . . .</b>            |   | <b>60</b> |
| 5.1                                  | Wavelet decomposition . . . . .                             | 60        |
| 5.2                                  | ANN implementation . . . . .                                | 60        |
| 5.3                                  | Wavelet recombination . . . . .                             | 61        |
| <b>VITAE . . . . .</b>               |   | <b>62</b> |

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

## List of Tables

|     |   |    |
|-----|---|----|
| 3.1 | Details of rainfall records in selected areas during the years of 1961 and 2006. . . . .  | 24 |
| 3.2 | An example of 3-hourly raw data in HTML format obtained from the Thai Meteorological department; temperature data at Hatyai station in the year 1997. . . . .   | 25 |
| 3.3 | An example of daily raw data in .DAT format obtained from the Royal Irrigation department; rainfall data at Hatyai station in the year 1997. . . . .            | 26 |
| 4.1 | ANN performance of daily Southern rainfall in selected areas with different number of input nodes and hidden nodes with various statistical evaluation. . . . . | 34 |
| 4.2 | Different activation function of ANN performance in the eastern coast of southern areas between the years 1995 and 2006. . . . .                                | 37 |
| 4.3 | The ANN performance of different climatological variable as rainfall predictors at Hatyai station between the years 1995 and 2006. . . . .                      | 38 |
| 4.4 | The ANN performance of different climatological variable as rainfall predictors at Hatyai station in wet period between the years 1995 and 2006. . . . .        | 43 |
| 4.5 | The ANN performance of different climatological variable as rainfall predictors at Hatyai station in dry period between the years 1995 and 2006. . . . .        | 43 |
| 4.6 | One-day forecasting of daily rainfall in southern provinces. . . . .  | 47 |
| 4.7 | The eastern coast of southern-area average daily rainfall for $n$ days prediction in selected provinces. . . . .  | 49 |
| 4.8 | Day 4 prediction of daily rainfall in the eastern coast of southern provinces. . . . .  | 51 |

## List of Figures

|      |   |    |
|------|---|----|
| 2.1  | Hydrological cycle [38]. . . . .  | 7  |
| 2.2  | Architectural graph of a neural network with one hidden layer. . .  | 10 |
| 2.3  | A supervised neural network. . . . .  | 10 |
| 2.4  | An unsupervised neural network. . . . .   | 11 |
| 2.5  | Architectural graph of a multilayer perceptron with one hidden layer.   | 11 |
| 2.6  | Piecewise-linear activation function. . . . .   | 12 |
| 2.7  | Gaussian activation function. . . . .   | 13 |
| 2.8  | Sigmoid activation function. . . . .  | 14 |
| 2.9  | The Daubechies scaling function [37]. . . . .   | 20 |
| 2.10 | The Daubechies wavelet function [37]. . . . .   | 20 |
| 2.11 | The recursive decomposition. . . . .  | 21 |
| 2.12 | The recursive reconstruction. . . . .   | 21 |
| 3.1  | The graphical view of southern Thailand [39]. . . . .   | 22 |
| 3.2  | Locations of Chumphon, Surat thani, Nakhon Srithamarat, Phatthalung<br>and Songkhla provinces in Southern Thailand. . . . .   | 23 |
| 3.3  | Artificial neural network related with time series. . . . .   | 28 |
| 3.4  | Creating training and testing set of the network. . . . .   | 29 |
| 3.5  | Diagram of wavelet-transform based artificial neural network. . . .   | 30 |
| 4.1  | Performance of ANN model in terms of $R^2$ with different number of<br>input nodes of daily rainfall prediction in Southern Thailand with<br>fixed hidden nodes at 200. . . . . | 33 |
| 4.2  | Performance of ANNs in term of RMSE with different number of<br>input nodes and hidden nodes. . . . .   | 35 |

|      |   |    |
|------|---|----|
| 4.3  | Scattered plots between observed and predicted daily rainfall in <i>mm</i> at Hatyai station with various input features: 4.3(a) using rainfall and humidity as input dimensions, 4.3(b) using rainfall and temperature as input dimensions, and 4.3(c) using rainfall and cloud amount as input dimensions, 4.3(d) using rainfall and pressure as input dimensions, 4.3(e) using rainfall and wind speed as input dimensions, 4.3(f) using rainfall and wind direction as input dimensions, 4.3(g) using rainfall, humidity and temperature as input dimensions, and 4.3(h) using rainfall, humidity, temperature, and cloud amount as input dimensions. . . . . | 36 |
| 4.4  | Daily rainfall prediction at Hatyai station using temperature, humidity and rainfall feature as input predictors. . . . .   | 39 |
| 4.5  | Rainfall distribution in Hatyai, Songkhla province. . . . .   | 40 |
| 4.6  | Scattered plots between observed and predicted daily rainfall in <i>mm</i> at Hatyai station with various input features in wet period: 4.6(a) using rainfall as the only input dimension, 4.6(b) using rainfall and humidity as input dimensions, 4.6(c) using rainfall and temperature as input dimensions, 4.6(d) using rainfall and cloud amount as input dimensions, 4.6(e) using rainfall, humidity and temperature as input dimensions, and 4.6(f) using rainfall, humidity, temperature, and cloud amount as input dimensions. . . . .  | 41 |
| 4.7  | Scattered plots between observed and predicted daily rainfall in <i>mm</i> at Hatyai station with various input features in dry period: 4.7(a) using rainfall as the only input dimension, 4.7(b) using rainfall and humidity as input dimensions, 4.7(c) using rainfall and temperature as input dimensions, 4.7(d) using rainfall and cloud amount as input dimensions, 4.7(e) using rainfall, humidity and temperature as input dimensions, and 4.7(f) using rainfall, humidity, temperature, and cloud amount as input dimensions. . . . .  | 42 |
| 4.8  | Daily rainfall at Hatyai station in wet period using rainfall, humidity and temperature variables as input features. . . . .  | 44 |
| 4.9  | Daily rainfall at Hatyai station in dry period using rainfall and humidity variables as input features. . . . .   | 44 |
| 4.10 | Graph of maximum-precipitation month at Hatyai station. . . . .   | 45 |
| 4.11 | Detail coefficients of daily rainfall series at different resolution level; (up) at the first resolution level and (down) at the second resolution level. . . . .   | 46 |
| 4.12 | Wavelet decomposition of daily rainfall series at different resolution level. . . . .   | 47 |

|      |  |    |
|------|--|----|
| 4.13 | Scattered plots between observed and predicted daily rainfall in <i>mm</i> in the eastern coast of southern provinces for one-day forecasting. .         | 48 |
| 4.14 | One-day daily rainfall prediction in the eastern coast of southern provinces. . . . .  | 48 |
| 4.15 | Day 2 rainfall prediction in the eastern coast of southern provinces.  | 49 |
| 4.16 | Day 3 rainfall prediction in the eastern coast of southern provinces.  | 50 |
| 4.17 | ANN performance for <i>n</i> –day daily rainfall prediction in the eastern coast of southern provinces. . . . .  | 50 |
| 4.18 | Scattered plots between observed and predicted daily rainfall in <i>mm</i> at the eastern coast of southern stations for day 4 forecasting. . .          | 51 |
| 4.19 | Day 4 daily rainfall prediction in the eastern coast of southern provinces. . . . .  | 52 |
| 4.20 | Performance comparison of southern daily rainfall among a conventional ANN model, a split-data ANN model, and wavelet-transform based ANN model. . . . . | 53 |

# CHAPTER I

## INTRODUCTION

Precipitation is one of the climatological key variables that plays an important role for human being. Water usage or consumption in regions around the world are related to rainfall in many direct or indirect ways – agricultural, industrial, ecosystem, transportation, and health. Over the coming decades, climate change has been recognized as one of the serious environmental issues. Many consequences of climate change are now being observed and investigated. For instance, rising intensity of rainstorms can increase the risk of flooding and agricultural hazards, while the spread of disease is another indirect-consequence of climate change on human health. Changes in precipitation are one of the expected impacts of global warming. Several regional precipitation trends, nowadays, are increasingly experiencing higher levels of precipitation, and some areas are witnessing reduced levels of precipitation and becoming arid. Thailand, which is one of the agricultural countries heavily depending on rains, has experienced flood and drought disaster for many years. Floods affect the northern and the central, as well as the southern regions of Thailand, while the north-eastern confronts with drought. Among the natural catastrophes, flooding is the most severe disaster that massively affects the residents and properties. According to the Royal Thai Government (RTG), the flooding destroyed homes, roads, bridges, and farmlands and led to the forced evacuation of more than ten-thousand of people. Among these flooding areas, the southern part of Thailand is a vulnerable area that has been influenced by uncertainties of tropical cyclones.

Southern Thailand is a tropical climate that is influenced by northeast and southwest monsoons. Most of the southern provinces experience not only these traditional monsoon rains but also tropical cyclones, which occasionally produce large amount of rainfall. According to the past historical records of Thai Meteorological Department and Royal Irrigation Department, the south of Thailand is vulnerable to cyclone disasters (reported in 1952, 1962, 1970, 1989, and 1997 events) with heavy rainfall that caused major flooding [34]. In November 2000, the eastern coast of southern Thailand was hit by torrential monsoon rains and floods creating one of the most dramatic natural hazards affecting at least 600,000 people in 10 southern provinces. Among these areas, provinces of Chumphon, Surat thani, Nakhon Si Thammarat, Phatthalung, and Songkhla were severely destroyed in terms of natural resources, infrastructures, and human lives.

For many decades, the forecasting of rainfall has been investigated in term of average quantity over some period of time in selected regions. One of the most important attributes to flood mitigation and water resources management is the accuracy of rainfall computation over a given area. At present, there have been

various models for monthly and seasonal rain computations with limited capacity to give satisfactory daily prediction. Many of these flooding areas are essential to social and economic growth of Southern Thailand, such as Hatyai district in Songkla, Muang district in Nakhon Si Thammarat, Kanchanadit district in Surat Thani, Muang district in Phatthalung, and Tha Sae district in Chumphon province. All of these areas are prone to flood hazard that requires serious investigation to prevent future damages and economic hardship of local residents. Therefore, knowledge of the variations of precipitation is important for water management application due to the significant effects on flooding.

## 1.1 Literature review

Rainfall prediction is one of the challenge problems in hydrology due to meteorological and geographical factors with uncertainties. The sophisticated nature of rainfall behavior makes it difficult to assess. In light of this, several dynamical forcings are related to rainfall's periodicity – climatological, topographical factors, and others. Therefore, most conventional rainfall modelings usually take these factors into account. Humidity, minimum and maximum air temperature were used as a rainfall predictors by many researchers ([20], [31], [25]), while the studies of Singhrattna et al. [21] and Chantasut [18] emphasized the large-scale ocean-atmospheric circulation variables – El Niño Southern Oscillation (ENSO), Sea Surface Temperature (SST), Southern Oscillation Index (SOI), as rainfall predictors in order to exhibit significant relationship to the climate change. Regardless of climatological factors, Toth et al. [10] investigated the capability of ANN in short-term rainfall forecasting using historical rainfall data as the only input information.

There has been a number of researches on rainfall forecasting in Thailand ([4], [18], [21], [20], [33], [34] [35]). Most of these studies have been contributed to the work in the region of central Thailand – Bangkok. Singhrattna, et al. [21] and Chantasut, et al. [18] contributed their works at stations in the West Central region and in the Chao Phraya River basin. Traditional linear regression and artificial neural network are the two significant tools for forecasting the large scale monthly and seasonal rainfall. For other parts of Thailand, Weesakul and Lowanichchai [35] employed the Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) methods to fit the time series of annual rainfall during 1951 to 1990 of 31 rainfall stations distributed in all regions of Thailand – Northern, Northeastern, Eastern, Southern, and Central areas. This model was proposed to predict the annual rainfall for agricultural water allocation planning management. The results showed that ARIMA and ARMA models were applicable for the purpose of agricultural water allocation planning. Among other parts of Thailand, Southern, revealed by Weesakul and Lowanichchai [35], provided less accuracy of forecast due to the influence of uncertain tropical cyclones which are the dominant cause of rainfall in that area. The influence was shown by the study of Vongvisessomjai [34], which established significant impacts of Typhoon Vae and Linda on heavy rainfall in southern Thailand. These cyclone disasters generated

heavy rainfall causing severe floods, high casualties and damages in many southern areas.

Reliable and accurate hydrological forecasting plays an important role in water management and flood warning. Therefore forecasting methods have been studied deeply in order to provide an efficient model. Both statistical and mathematical approaches are implemented to obtain such estimation. Most of the previous hydrological works considered hydrological forecasting as a deterministic approach. The work of Singhrattna, et al. [21] described the development of a statistical forecasting method for summer monsoon rainfall over Thailand with a linear regression and a local polynomial based non-parametric method. Solomatine et al. [9] indicated a problem of predicting surge water levels by using of linear autocorrelation and ARIMA models and non-linear methods. Among several rainfall forecasting techniques based on statistical or deterministic methods and computational approaches, there has been an excess of evidence in literature that artificial neural networks (ANNs) is increasingly used for hydrological modeling especially in rainfall prediction ([15], [18], [20], [25], [26], [31]). The ability of representing non-linear complex relation from a set of known input and output variables is the significant role of ANNs. Particularly, ANNs are non-linear modeling tools that do not require an explicit mathematical formulation of the physical relationship between variables. Among different kinds of ANNs, feedforward backpropagation has gained popularity in the investigation weather forecasting ([14], [20], [28]).

The majority of studies have proven that artificial neural networks are able to outperform traditional statistical techniques. Lee et al. [31] predicted daily rainfall data by employing divide and conquer technique. The whole region was divided into 4 sub-regions. Precipitation in two larger regions was predicted by radial basis function neural network, while the other two smaller regions were carried out by a regression model. The artificial neural networks performed well in comparison with the linear models. The result of better performance of artificial neural network over conventional methods has also revealed by many researchers ([6], [9], [14], [27], [30]). Results from the study of Solomatine et al. [9] revealed that there was still more local predictive information embedded into the attractor of the system so that the statistical models could not provide sufficient accuracy. A comparison between ANN models and traditional models has been made, as well, by Hsu et al. [11] who stated that the ANN approach would be more effective when explicit knowledge of the hydrological variables is not required.

Although the expert system of artificial neural network are capable of modeling non-linear relationships, its successful employment may be restricted due to the sophisticate nature of non-stationarity and non-linearity in the hydrological variables. The artificial neural network has been shown in many studies that the problem of extrapolation has certain limitations. In particular, the models were unable to estimate the peak signal ([5], [11]).

In recent years, the wavelet transform has been successfully applied to



data series in order to analyze the stochastic components. Geva [1] reported on an improvement in the results of multi-scale wavelet decomposition. The prediction accuracy was improved by using different scales of the time windows and frequencies bands, which supplied the networks with more information on the past. Soltani [32] and Renaud et al. [22] also proposed a method of predicting nonlinear time series with wavelet decomposition. Their works showed that the decomposition could simplify the main task of predicting a complex behavior. As a result, many applications of wavelet analysis have been consequently studied. Bi et al. [40] applied wavelet decomposition in the short-term load forecasting in Queensland, Australia. In hydrological approach, many researchers eliminated the nuances of the series, which are the noise of signal, in order to consider the deterministic part [33] and enhanced a network learning.

Due to the capability of wavelet decomposition in isolating the periodicity in a time series, the use of wavelet transform with several technique has been proposed. Both statistics and neural network are applied with wavelet decomposition. Mabrouk [2] presented wavelet decomposition and autoregression models for time series prediction. Jayawardena [4] predicted daily rainfall data by using wavelet technique and hidden markov model in Chao Phraya basin. A hybrid method which combines a deterministic model with stochastic model also presented by Cristea [23]. Many techniques with the combination of deterministic and stochastic model have been proposed. Tong [36] developed the combination of wavelet method, back-propagation neural network, and autoregressive moving average (ARMA) for data mining forecasting. All of these studies showed that the use of wavelet decomposition method could help reduce the empirical task and also improve a model accuracy.

## 1.2 The objective

The primary goal of this study is to predict daily rainfall in southern Thailand. The southern areas that is vulnerable to flood disaster are chosen as site studies. The selected southern provinces for this study are located in the eastern coast; Chumphon, Surat thani, Nakhon Si Thammarat, Phatthalung, and Songkhla. A 3-hourly rainfall and climatological data from monitoring stations in the given regions in the period of 1995-2006 are used as primary data. A feed-forward backpropagation ANN is employed to model and forecast southern daily rainfall data. According to the significant application of wavelet analysis on various purposes, a combining technique of wavelet decomposition and artificial neural network is proposed in this study to forecast  $n$ -day daily rainfall. Overall performance efficiency of predictions are summarized in terms of coefficient of determination, correlation coefficient and root mean squared error. Statistical evaluation of the rainfall models are also presented.

In summary, to fulfill the forecasting in southern part of Thailand, provinces of Chumphon, Surat thani, Nakhon Si Thammarat, Phatthalung, and Songkhla, that are vulnerable to flood disaster, are investigated in the study with deterministic

and stochastic models.

### 1.3 Outline

The thesis is organized as follows. Chapter II introduces the underlying theories related to the study – physical concept of precipitation, artificial neural network, and discrete wavelet analysis. Chapter III elaborates the relevant methodologies. Experimental results and statistical performance of the rainfall models are presented and discussed in chapter IV. Chapter V concludes the main results of this research.



## CHAPTER II

### THEORETICAL BACKGROUND

In this chapter, underlying mathematical concepts and techniques of the present study are introduced. Physical descriptions of precipitation are first elaborated in section 2.1. In section 2.2, technical overview of artificial neural network is presented. Wavelet theory, which plays an important role in this work, is described in the last section of this chapter.

#### 2.1 Physical descriptions of precipitation

Circulation and conservation of water in the hydrological cycle is one of the vital processes naturally operating within the global system. A major component of this cycle is the ocean due to its major regulation upon the flow of water in the system. The oceans influence, in particular, evaporation and the return of water to the atmosphere, and thereby controlling to a great extent flows of moisture in the atmosphere and rates of precipitation [3].

##### 2.1.1 The Hydrological cycle

The whole system of water movement has been termed the **hydrological cycle**. Fig. 2.1 illustrates main components of the system with inputs, outputs, flow regulators and storages. The main input to surface hydrological cycle is directly from precipitation. As a consequence, evaporation and transpiration constitute the system output. These inputs and outputs are linked by flows in the atmosphere, the oceans and rivers on the continents [7].

##### Inputs to the surface hydrological cycle

Precipitation is by far the most important input to the surface hydrological system. Precipitation occurs in a variety of forms – hail, snow, and rainfall. Throughout most of the world, the major input is in the form of rainfall. The distribution of this input across the world shows a marked relationship to the distribution of factors influencing precipitation, in particular, the incidence of storms, the atmospheric moisture content and the oceans. Characteristic of the rainfall has an important influence on what happens to the water after it has reached the ground. The effect upon the hydrological cycle, upon geomorphological processes, and, above all, upon man are almost always greater when precipitation is intense.

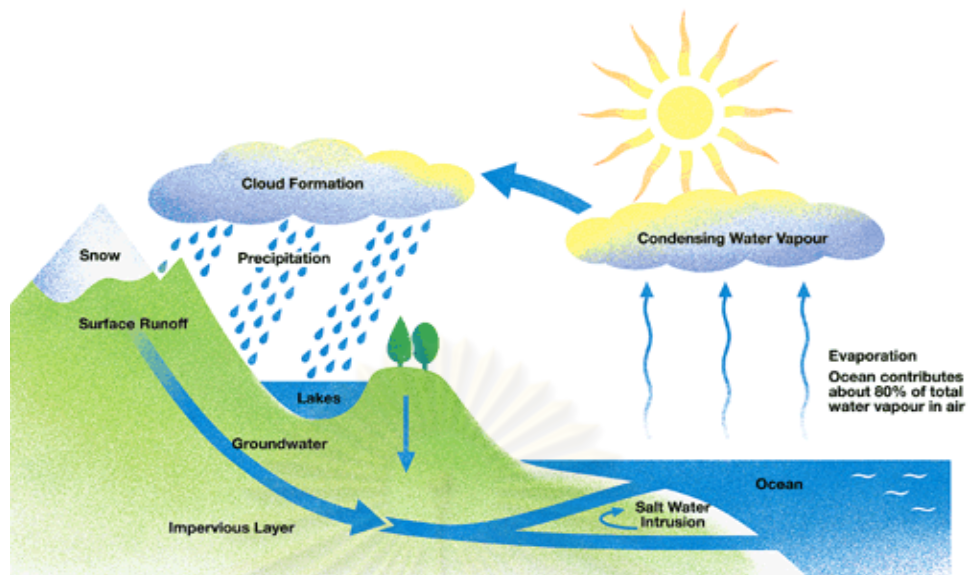


Figure 2.1: Hydrological cycle [38].

## Outputs from the surface hydrological system

Evaporation is one of the major outputs from the earth's surface. The rate of evaporation depends on various factors for which the supply of energy is the most important one. Evaporation process involves the conversion of water to water vapour that requires considerable inputs of energy. Another important factor is the availability of moisture at the surface. As the surface dries out and moisture becomes less available, the rates of evaporation tend to decline. In addition, evaporation is favoured by a moisture gradient between the surface and the air above, and thus evaporation rates decline when the atmosphere is moist. Finally, wind plays another important part by removing the moist air and maintaining a moisture gradient.

The inputs of moisture to the atmosphere are not everywhere in balance with outputs by precipitation. Over the oceans, for example, evaporation is high while precipitation is relatively low; the atmosphere gains more moisture than it loses. Over the continents, evaporation is less than precipitation, and the atmospheric moisture budget is negative. As with the oceans, therefore, horizontal flows of moisture must occur to maintain equilibrium.

### 2.1.2 Climatological Distribution

In order to understand the precipitation process, it is necessary to appreciate the factors affecting it. Intensity precipitation varies in time and over a catchment area. Wind plays an important effect in bringing moisture which has evaporated from exposed waters or transpired from surfaces. Wind also causes clouds to travel across the catchment. Precipitation can be generated if the tem-

perature of the cloud of water vapour drops below the dew point. Condensation is followed by precipitation. The cooling action may be caused by rising air; against mountains (orographic precipitation) due to cold fronts (frontal or cyclonic precipitation) or due to thermal currents (convictional precipitation). The latter gives rise to thunderstorms, which is an intense form of precipitation but often of relatively short duration [3].

### 2.1.3 Formation of precipitation

Precipitation forms differently depending on whether it is generated by warm or cold clouds. Warm clouds are defined as those that do not extend to levels where temperatures are below 32°F (0°C), while cold clouds exist at least in part at temperatures below 32°F (0°C). The formation of precipitation may occur at temperatures above or below freezing. Precipitation that is formed in temperatures entirely above freezing is called warm precipitation; cold precipitation involves ice at some stage of the process [3].

#### Warm Precipitation

Nearly all precipitation begins with condensation of water vapor of small particles in the air which is called cloud condensation nuclei. Condensation may occur at relative humidities less than 100 % for hygroscopic particles (those having an affinity for water) or may be delayed until the relative humidity exceeds 100 % if the particles are hydrophobic (lacking an affinity for water).

Saturation of air occurs when rising air currents cool adiabatically (that is, without loss of heat) by expansion. Because the saturation vapor pressure of water decreases exponentially with decreasing temperature, cooling of a moist air mass by lifting is an efficient mechanism for producing saturation and condensation.

The condensation processes are efficient in producing only cloud drops that are too small to have an appreciable fall velocity relative to the air. In order to produce precipitation particles that are heavy enough to fall to the surface, a cloud must increase its radius and volume. In the clouds with temperatures above the freezing point, the growth occurs by coalescence, which is simply the merging of colliding water drops. This merging is facilitated when an electric field is present.

#### Cold Precipitation

Whereas collision and coalescence are efficient means for producing precipitation in the warm, humid tropical regions, the formation of precipitation in middle latitudes usually involves ice. Because the vapor pressure at saturation is less over ice than over water, ice crystals will grow at the expense of water drops

when both exist together in a supercooled cloud, which contains liquid drops at temperatures below the freezing point.

Although most precipitation in the middle latitudes begins as snow at altitudes above the freezing level, the form of the precipitation reaching the surface depends on the temperature structure of the atmospheric layers through which the precipitation falls. If the temperature near the ground is warm enough, the snow has time to melt and reaches the ground as rain. Hail occurs when alternating strong updrafts and downdrafts cause ice crystals to pass repeatedly through layers that contain supercooled water. The frequent passage through these layers allows the water to freeze around the growing hailstone and to accumulate in one layer after another.

The distribution of precipitation is not uniform across the earth's surface, and varies with time of day, season and year. The lifting and cooling that produces precipitation can be caused by solar heating of the earth's surface, or by forced lifting of air over obstacles or when two different air masses converge. For these reasons, precipitation is generally heavy in the tropics and on the upwind side of tall mountain ranges.

## 2.2 Artificial Neural Network

Artificial Neural Networks (ANNs) is a non-linear mathematical structure which is capable of representing arbitrarily complex non-linear processes by relating inputs and outputs of such system. The network consists of three main parts: input layer, hidden layer and output layer. Each layer consists of neurons. The input layer constitutes with a set of sensory units. The network can have one or more hidden layers of computation nodes. Output nodes constitute an output layer [29]. Each layer is fully connected to the next one with a synaptic weight on each connection. Its architectural graph of neural network is shown in Fig. 2.2.

There are two major paradigms: supervised and unsupervised neural network.

### Supervised Neural Network

The supervised neural network consists of many pairs of input and output training patterns. The learning of the network benefits from output patterns (target) which act as assistances of the teacher to produce the estimations as close to the target as possible by using weight adjustment. Fig. 2.3 illustrates a supervised learning.

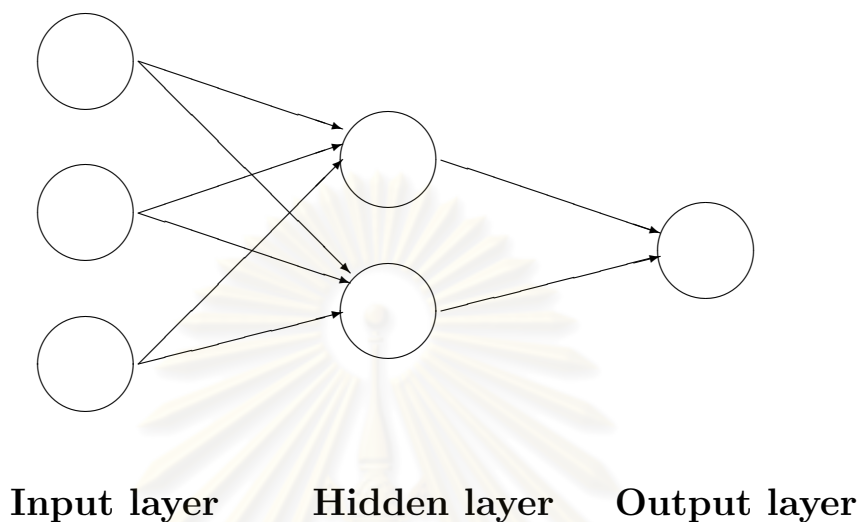


Figure 2.2: Architectural graph of a neural network with one hidden layer.

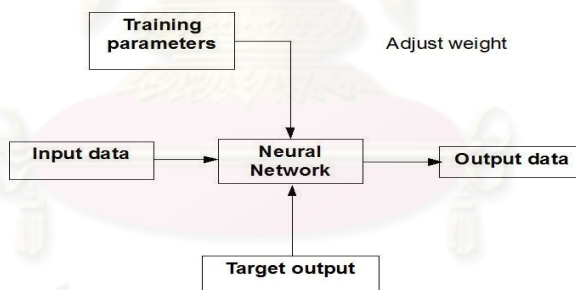


Figure 2.3: A supervised neural network.

## Unsupervised Neural Network

For unsupervised learning, the training set consists of input training patterns only. Therefore, the network is trained without benefit of any teacher. The network learns to adapt based on the experiences collected through the previous training patterns. A typical schema of an unsupervised system is shown in Fig. 2.4.

### 2.2.1 A Multilayer Feedforward Network

A feedforward is a network that the output values of each layer only move from one layer to the next; no values are fed back to earlier layers (a Recurrent

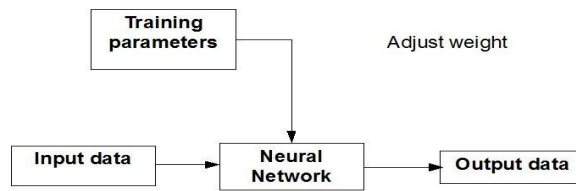


Figure 2.4: An unsupervised neural network.

Network allows values to be fed backward). A multilayer network is a class of a feedforward neural network that has one or more hidden layers.

At a neuron in the input layer in Fig. 2.5, the value from each node  $x^i$  is multiplied by a weight  $w^i$ , and the resulting weighted values are added together producing a combined value  $v_j = \sum_{i=1}^m x^i w^i$ . The weighted sum  $v_j$  is fed into a transfer function,  $\phi$ , which outputs a value  $y$ . The  $y$  value is the output of the network mathematically describes as (2.1).

$$\begin{aligned}
 y &= \phi(v_j) \\
 &= \phi\left(\sum_{i=1}^m x^i w^i\right).
 \end{aligned}
 \tag{2.1}$$

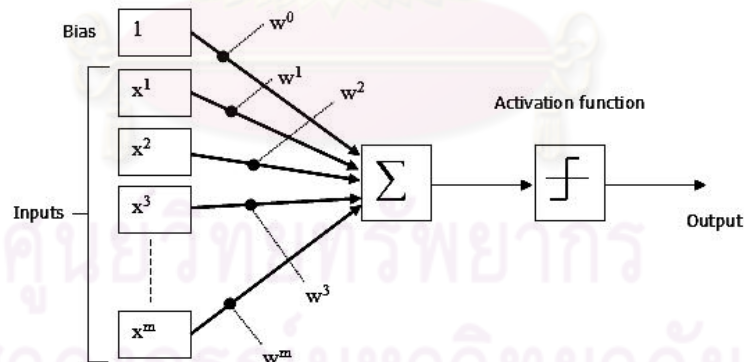


Figure 2.5: Architectural graph of a multilayer perceptron with one hidden layer.

## 2.2.2 Activation function

Activation function or transfer function is a function that introduces non-linearity into the network. It is possible for neural network to do non-linear mapping between inputs and outputs. Continuity of the functions implies that there are no sharp peaks or gaps, so that the function can be differentiated throughout, making it possible to implement the delta rule to adjust both input-hidden and hidden-output layer weights in backpropagation errors, this will be discussed later in detail.



Mathematical functions that are customarily used as activation functions of the network are the followings.

### Linear activation function

Linear function is a transfer function that is suitable for the unbounded output value. Fig. 2.6 illustrated the linear activation function which can mathematically describe as

$$\phi(v) = av + b. \quad (2.2)$$

where  $a$  is the slope parameter and  $b$  is an intersection of y-axis.

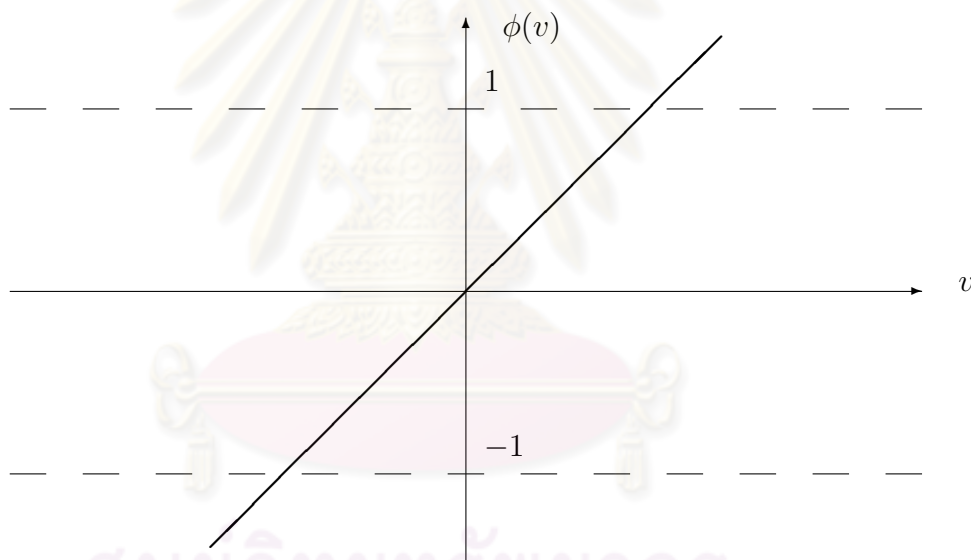


Figure 2.6: Piecewise-linear activation function.

### Gaussian activation function

The standard normal curve, shown in Fig. 2.7, has a symmetric bell shape and is commonly known as standard normal distribution. Its range is  $[0, 1]$ . The gaussian function is normally defined as

$$\phi(v) = e^{-av^2}. \quad (2.3)$$

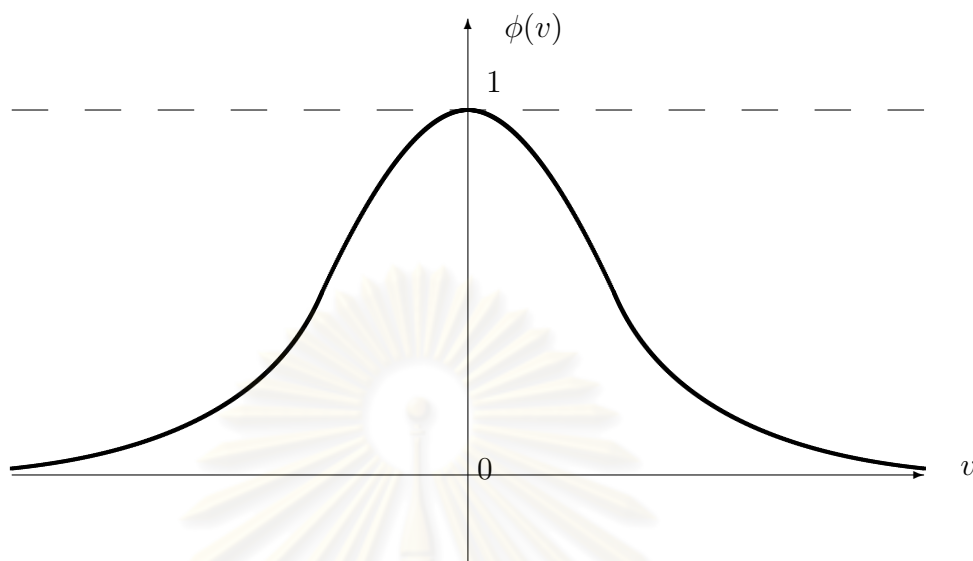


Figure 2.7: Gaussian activation function.

### Sigmoid activation function

The sigmoid function, whose graph is similar to a s-shape, is the most common form of activation function used in the construction of artificial neural networks with positive output. An example of the sigmoid function is the logistic function shown in Fig. 2.8 and defined by

$$\phi(v) = \frac{1}{1 + \exp^{-av}}, \quad (2.4)$$

where  $a$  is the slope parameter of the sigmoid function.

### 2.2.3 Learning algorithm

Basically, error back-propagation learning consists of two passes through the different layers of the network: a forward pass and a backward pass. In the forward pass, an input vector is applied to the sensory nodes of the network in an input layer, and its affect propagates through the network layer by layer.

In the forward pass, the synaptic weights of each node are adjusted in accordance with an error-correction rule as follows:

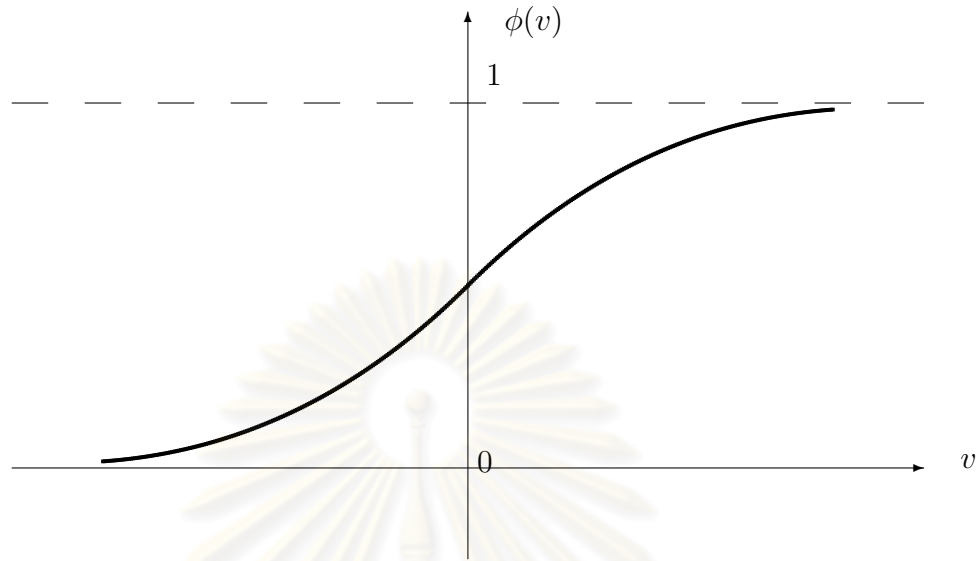


Figure 2.8: Sigmoid activation function.

$$E = \frac{1}{2} \sum_j (t^j - y^j)^2, \quad (2.5)$$

where

$t^j$  is the target output at unit  $j$ ,

$y^j$  is the predicted output at unit  $j$ .

Specifically, the actual response of the network is subtracted from a desired (target) response to produce against the direction of synaptic connections in the backward pass. Under the delta rule, the change in weight is

$$\Delta w_j^i(t) = \eta E_j(t) \phi_j'(v_j(t)) y_i(t), \quad (2.6)$$

where

$t$  represents the number of iteration,

$w_j^i$  refer to the weight of input from unit  $i$  to unit  $j$ ,

$\eta$  is the learning-rate parameter of the back-prop algorithm ,

$E_j$  is the error evaluated at unit  $j$ ,

$\phi_j'$  refers to the derivative of the associated activation function at the neuron  $j$  .

Thus, the synaptic weights are adjusted to make the actual response of the network move closer to the desired response in a statistical sense as

$$w_j^i(t+1) = w_j^i(t) + \Delta w_j^i(t). \quad (2.7)$$

## 2.2.4 Resilient backpropagation

The Resilient backpropagation training algorithm (Rprop) has been one of the advanced batch-training algorithms in weight adjustments for supervised learning in the field of ANN [17]. It is a well-established modification of the ordinary gradient descent. The basic idea is to adjust and hence eliminate the influence of the size of the partial derivative on the weight step. Rprop takes into account only the sign of the partial derivative over all patterns (not the magnitude), and acts independently on each weight. Only the sign of the derivative is considered to indicate the direction of the weight update. The size of the weight change is determined by a weight-specific,

$$\Delta w_j^i(t) = \begin{cases} -\Delta_j^i(t) & \text{if } \frac{\partial E(t)}{\partial w_j^i} > 0 \\ +\Delta_j^i(t) & \text{if } \frac{\partial E(t)}{\partial w_j^i} < 0 \\ 0 & \text{if } \frac{\partial E(t)}{\partial w_j^i} = 0, \end{cases} \quad (2.8)$$

where  $\frac{\partial E(t)}{\partial w_j^i}$  denoted the partial derivative with respect to each weight which referred to the sum gradient information over all patterns of the pattern set.

Next, the Rprop learning rules determine the new update value  $\Delta_j^i(t)$ . This is based on a sign-adaptation process as follows;

$$\Delta_j^i(t) = \begin{cases} \eta^+ * \Delta_j^i(t-1) & \text{if } \frac{\partial E(t-1)}{\partial w_j^i} * \frac{\partial E(t)}{\partial w_j^i} > 0 \\ \eta^- * \Delta_j^i(t-1) & \text{if } \frac{\partial E(t-1)}{\partial w_j^i} * \frac{\partial E(t)}{\partial w_j^i} < 0 \\ \Delta_j^i(t-1) & \text{if } \frac{\partial E(t-1)}{\partial w_j^i} * \frac{\partial E(t)}{\partial w_j^i} = 0, \end{cases} \quad (2.9)$$

where  $0 < \eta^- < 1 < \eta^+$ .

## 2.3 Wavelet transform

Wavelet analysis can be regarded as a transformation of time series from temporal domain to wavelet domain. The transform is particularly well adapted to characterize transient phenomena because it decomposes signals into building blocks that are well localized in space and frequency. There are two categories of wavelet transform: continuous wavelet transform (CWT) and discrete wavelet transform (DWT), which are a set of basis functions in Hilbert space  $L^2(\mathfrak{R})$ .

### 2.3.1 Discrete wavelet transform

The discrete wavelet transform (DWT) performs on a set of discrete input signal to provide the output of discrete wavelet transform as approximation coefficients and detail coefficients [13].

This section begins by introducing the definitions of inner product and discrete wavelet transformation.

**Definition 1. Inner product.**

Consider real or complex sequences in  $L^2(Z)$ ,  $x(n)$ ,  $n \in Z$ . The inner product is defined as [24]

$$\langle a(n), b(n) \rangle = \sum \overline{a(n)} b(n). \quad (2.10)$$

Using the transformation  $T$ , the wavelet representation of a function  $f(x)$  can be defined as follows:

**Definition 2. Discrete wavelet transform.**

$$\begin{aligned} f(x) &= \sum_{(j,k) \in \Lambda} C_{j,k} \phi_{j,k}(x) \\ &= \sum_{(j,k) \in \Lambda} \langle f, \phi_{j,k} \rangle \phi_{j,k}(x), \end{aligned} \quad (2.11)$$

where  $\Lambda$  is the Cartesian product  $\frac{1}{2}Z \times Z^2$ . The function  $\phi(x)$  enabling this decomposition is a wavelet and  $C_{j,k}$  are the associated coefficients [24].

### 2.3.2 Convolution and Filter

Convolution is a mathematical operation on two functions  $\mathbf{h}$  and  $\mathbf{g}$ , producing a third function that is typically viewed as a modified version of one of the original functions. The convolution of  $\mathbf{h}$  and  $\mathbf{g}$  is denoted by  $\mathbf{h} * \mathbf{g}$ . It is defined as the summation of the product of the two functions after being reversed and shifted.

**Definition 3. Convolution.**

Let  $\mathbf{h}$  and  $\mathbf{g}$  be two bi-infinite sequences. Then, the convolution product,  $\mathbf{y}$ , of  $\mathbf{h}$  and  $\mathbf{g}$ , denoted by  $\mathbf{h} * \mathbf{g}$ , is the bi-infinite sequence  $\mathbf{y} = \mathbf{h} * \mathbf{g}$ , whose the  $n$ th component is given by [13]

$$y_n = \sum_{k=-\infty}^{\infty} h_k g_{n-k} \quad (2.12)$$

### 2.3.3 Lowpass filter

**Definition 4. Lowpass filter.**

Let  $\mathbf{h}$  be some sequence,  $H(\omega)$  denote the Fourier series of  $\mathbf{h}$ , and  $0 < \omega_p \leq \omega_s < \pi$ .

Suppose that there exists  $0 < \delta < \frac{1}{2}$ , with  $1 - \delta \leq |H(\omega)| \leq 1 + \delta$  for  $0 \leq \omega \leq \omega_p$  and an  $0 < \lambda < \frac{1}{2}$ , so that for  $\omega_s \leq \omega \leq \pi$ ,  $|H(\omega)| \leq \lambda$ , then we call  $\mathbf{h}$  a lowpass filter [13].

A good working definition is to say that if  $\mathbf{h}$  is a lowpass filter, then  $|H(\omega)| \approx 1$  for  $0 \leq \omega \leq \omega_p$  for some  $0 < \omega_p < \pi$  and  $H(\omega) \approx 0$  for  $\omega_s \leq \omega < \pi$  where we require that  $\omega_p \leq \omega_s \leq \pi$ . That is, the lowpass filter have to satisfy the following condition:

$$|H(0)| = 1 \text{ and } H(\pi) = 0 \quad (2.13)$$

### 2.3.4 Highpass filter

**Definition 5. Highpass filter.**

Let  $\mathbf{g}$  be some sequence,  $G(\omega)$  denote the Fourier series of  $\mathbf{g}$ , and  $0 < \omega_p \leq \omega_s < \pi$ . Suppose that there exists  $0 < \lambda < \frac{1}{2}$  so that  $|G(\omega)| \leq \lambda$  for  $0 \leq \omega \leq \omega_p$  and a  $0 < \delta < \frac{1}{2}$  with  $1 - \delta \leq |G(\omega)| \leq 1 + \delta$  for  $\omega_s \leq \omega \leq \pi$ , then we call  $\mathbf{g}$  a highpass filter [13].

A good working definition is to say that if  $\mathbf{g}$  is a highpass filter, then  $|G(\omega)| \approx 0$  for  $0 \leq \omega \leq \omega_p$  for some  $0 < \omega_p < \pi$ , and  $|G(\omega)| \approx 1$  for  $\omega_s \leq \omega \leq \pi$  where we require that  $\omega_p \leq \omega_s < \pi$ . That is, the highpass filter have to satisfy the following condition:

$$G(0) = 0 \text{ and } |G(\pi)| = 1. \quad (2.14)$$

### 2.3.5 Multiresolution analysis

A multiresolution analysis is an increasing sequence of closed subspaces  $\{V_j\}_{j \in \mathbb{Z}}$  which approximates  $L^2(\mathfrak{R})$

$$0 \rightarrow \cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots \rightarrow L^2(\mathfrak{R}) \quad (2.15)$$

and satisfies the following properties:

$$f(x) \in V_j \iff f(2x) \in V_{j+1}. \quad (2.16)$$

There exists a function  $\phi(x)$  in  $V_0$  such that the set  $\{\phi(x - k)\}_{k \in \mathbb{Z}}$  is an orthonormal basis of  $V_0$ . The scaling function  $\phi(x)$  satisfies the well known two-scale difference equation with scale changes by any power of 2,

$$\phi(x) = 2^{\frac{i}{2}} \sum_{k=0}^{L^{(i)}-1} h^{(i)}(k) \phi(2^i x - k). \quad (2.17)$$

From (2.11), the functions  $\phi(x)$  and  $\phi^*(x)$  are not only a means for a wavelets definition. They actually define the multiresolution analysis notions. Consider a family of linear applications  $(A_j)_{j \in \frac{1}{2}Z}$  defined by

$$A_j(f)(x) = \sum_{-\infty}^{\infty} \langle f, \phi_{j,k}^* \rangle \phi_{j,k}(x). \quad (2.18)$$

These applications yeild continuous projectors which degrade the information contained in  $f(x)$  as  $j$  decreases, and which provide an increasingly better approximation of  $f(x)$  as  $j$  increases. According to the transformation  $T$  in (2.11),  $A_j(f)(x)$  can be interpreted as a version of  $f(x)$  at the scale of  $2^j$  ( $j \in \frac{1}{2}Z$ ).

The task of the wavelet transformation is therefore to extract the details lost between two consecutive scales. For every  $j$  in  $\frac{1}{2}Z$ , we have

$$A_{j+1}(f)(x) - A_j(f)(x) = \sum_{-\infty}^{\infty} \langle f, \psi_{j,k}^* \rangle \psi_{j,k}(x), \quad (2.19)$$

where  $\psi$  is a wavelet function and we retrieve formula (2.11), which is a consequence of the fact that the approximation improves as  $j$  increases.

The sequence of sampling coefficients  $\{S_k^j\}_{k \in Z} = \{2^{\frac{j}{2}} \langle f, \phi_k^j \rangle\}_{k \in Z}$  is the approximations and  $\{D_k^{j'}\}_{k \in Z} = \{2^{\frac{j'}{2}} \langle f, \psi_k^{j'} \rangle\}_{k \in Z}$  detail coefficients at lower scales. The sequence can be computed by using a convolution with the discrete filter ,  $h_n$ , followed by a decimation by a factor of two. The filer  $\{h_n\}_{n \in Z}$  can have finite impulsive response as shall be described in the next subsection.

### 2.3.6 The Daubechies wavelet function

**Definition 6. Daubechies wavelet function.** Let  $N$  be an even positive integer.

Then we define the daubechies wavelet transformation by the matrix

$$W_N = \begin{pmatrix} h_0 & h_1 & \cdots & h_{N-1} & 0 & 0 & \cdots & 0 \\ 0 & 0 & h_0 & h_1 & \cdots & h_{N-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ h_1 & \cdots & h_{N-1} & 0 & 0 & \cdots & 0 & h_0 \\ g_0 & g_1 & \cdots & g_{N-1} & 0 & 0 & \cdots & 0 \\ 0 & 0 & g_0 & g_1 & \cdots & g_{N-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ g_1 & \cdots & g_{N-1} & 0 & 0 & \cdots & 0 & g_0 \end{pmatrix}$$

The filter

$$\mathbf{h} = (h_0, h_1, \dots, h_{N-1})$$

is called the daubechies filter (scaling function) and we shall call

$$\mathbf{g} = (g_0, g_1, \dots, g_{N-1})$$

the daubechies wavelet filter [24].

Thus, the daubechies 2 matrix is

$$W_4 = \begin{pmatrix} h_0 & h_1 & h_2 & h_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & h_0 & h_1 & h_2 & h_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & h_0 & h_1 & h_2 & h_3 \\ h_2 & h_3 & 0 & 0 & 0 & 0 & h_0 & h_1 \\ g_0 & g_1 & g_2 & g_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & g_0 & g_1 & g_2 & g_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & g_0 & g_1 & g_2 & g_3 \\ g_2 & g_3 & 0 & 0 & 0 & 0 & g_0 & g_1 \end{pmatrix}$$

where the decomposition of lowpass coefficients is given by

$$\begin{aligned} h_0 &= \frac{1 + \sqrt{3}}{4\sqrt{2}} = 0.48296291314469025 \\ h_1 &= \frac{3 + \sqrt{3}}{4\sqrt{2}} = 0.83651630373746899 \\ h_2 &= \frac{3 - \sqrt{3}}{4\sqrt{2}} = 0.22414386804185735 \\ h_3 &= \frac{1 - \sqrt{3}}{4\sqrt{2}} = -0.12940952255092145 \end{aligned}$$

and the decomposition of highpass coefficients is given by

$$\begin{aligned} g_0 &= h_3 \\ g_1 &= -h_2 \\ g_2 &= h_1 \\ g_3 &= -h_0. \end{aligned}$$



Figures 2.9 and 2.10 illustrate the Daubechies 2 scaling and wavelet functions, respectively.

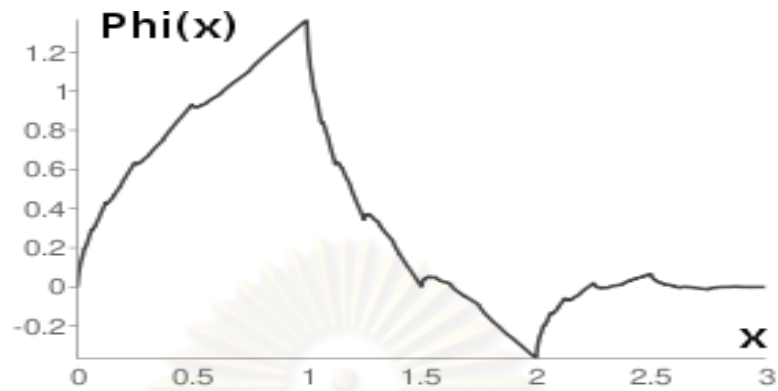


Figure 2.9: The Daubechies scaling function [37].

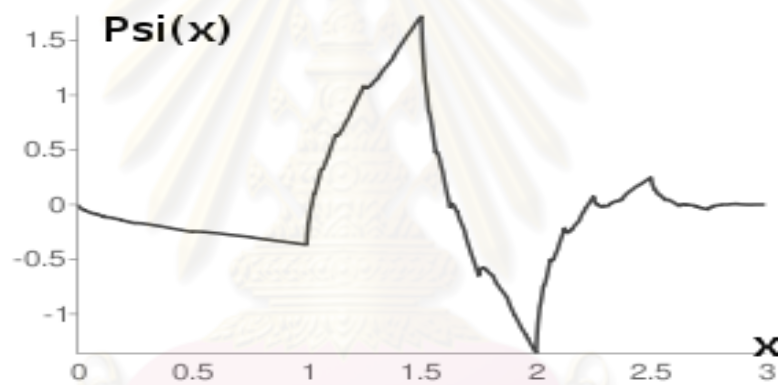


Figure 2.10: The Daubechies wavelet function [37].

### 2.3.7 À trous wavelet transform

À trous wavelet transform is a non-decimated wavelet transform which decomposes signal into coefficient series with the same length instead of decreasing by two as in the DWT case [22]. Two sets of wavelet coefficients, approximation and detail coefficients, can be obtained from the à trous wavelet algorithm as follows.

Given a time signal  $x(t)$ , the smoothed data or approximations at resolution level  $j$  at position  $k$ ,  $c_j(k)$ , is given by passing the signal  $x(t)$  through a series of low pass filters  $\mathbf{h}$  analyzed at each resolution level  $j$  at position  $k$  as

$$c_j(k) = \sum_{l=0}^N h(l)c_{j-1}(k + 2^{j-1}l). \quad (2.20)$$

$$\begin{array}{ccccccc}
 x & \rightarrow & c_1 & \rightarrow & c_2 & \rightarrow & \cdots & \rightarrow & c_n \\
 & & \searrow & & \searrow & & \searrow & & \searrow \\
 & & d_1 & & d_2 & & \cdots & & d_n.
 \end{array}$$

Figure 2.11: The recursive decomposition.

$$\begin{array}{ccccccc}
 x & \leftarrow & c_1 & \leftarrow & c_2 & \leftarrow & \cdots & \leftarrow & c_n \\
 & & \swarrow & & \swarrow & & \swarrow & & \swarrow \\
 & & d_1 & & d_2 & & \cdots & & d_n.
 \end{array}$$

Figure 2.12: The recursive reconstruction.

By recursively decomposed approximation coefficients at the previous level, approximation and detail coefficients at the next level are obtained. Fig. 2.11 illustrates the recursive decomposition.

The high frequency coefficients or details at resolution level  $j$  and at position  $k$ ,  $w_j(k)$ , can be calculated from the difference between consecutive resolutions of approximation series as

$$d_j(k) = c_{j-1}(k) - c_j(k). \quad (2.21)$$

Finally, the signal can be reconstructed by using the mathematical expression as

$$x(k) = c_n(k) + \sum_{j=1}^n d_j(k), \quad (2.22)$$

where  $n$  is a number of resolution level. Fig. 2.12 illustrates the recursive reconstruction. The combination of approximation and detail coefficients at a lower level reconstructs approximation coefficients at a higher level. By recursively reconstruction, the original signal  $x$  is obtained.

# CHAPTER III

## METHODOLOGY

### 3.1 Study area

The South of Thailand is a peninsula bounded by the Andaman Sea of the Indian Ocean to the West and the South China Sea of the Pacific Ocean to the East as shown in Fig. 3.1. Topography of this southern region is the peninsula with mountainous and basin areas for cultivation. It occupies the area of approximately  $70,715.2 \text{ km}^2$ . The southern weather regards as a tropical climate with two seasons: summer and rainy seasons. Most of southern peninsula's weather is influenced by seasonal monsoons.

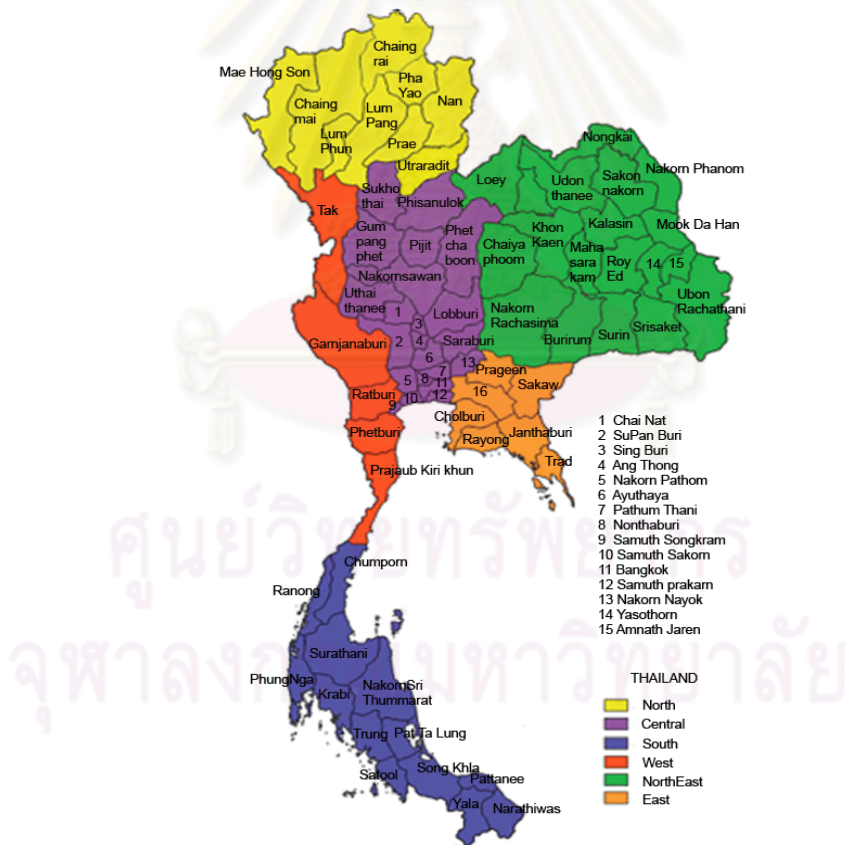


Figure 3.1: The graphical view of southern Thailand [39].

Chumphon, Surat Thani, Nakhon Si Thammarat, Phatthalung and Songkhla are provinces located in the east coast of the south of Thailand as illustrated in Fig. 3.2. From the past historical records of Thai Meteorological Department and

Royal Irrigation Department, these area are vulnerable to flood disaster.



Figure 3.2: Locations of Chumphon, Surat thani, Nakhon Srithamarat, Phatthalung and Songkhla provinces in Southern Thailand.

Table 3.1 shows the historical records of precipitation in selected provinces. Further information indicated that the month of maximum rainfall is November. Among these areas, the maximum precipitation record is 643.1 *mm* in Nakhon Si Thammarat province and the maximum mean number of wet days is 172 days in Chumphon province.

Certain districts of these provinces are prone to severe flood disaster. Therefore, 3-hourly rainfall data, together with climatological data of pressure, cloud density, temperature, humidity, wind speed and direction from the monitoring stations at Hat Yai district in Songkhla province and daily rainfall data from Tha Sae district in Chumphon province, Muang district in Nakhon Si Thammarat province, Kanchanadit district in Surat Thani province and Muang district in Phatthalung province during the period of 1995-2006 obtained from the Thai Meteorological department and the Royal Irrigation department are used as training and testing sets in this study.

Table 3.1: Details of rainfall records in selected areas during the years of 1961 and 2006.

| Province            | Area<br>( $km^2$ ) | Annual<br>rainfall<br>(mm) | Mean number<br>of wet days<br>(days) | Maximum<br>rain in<br>November |
|---------------------|--------------------|----------------------------|--------------------------------------|--------------------------------|
| Chumphon            | 6,010.85           | 1961.80                    | 172                                  | 380.4                          |
| Surat Thani         | 12,891.47          | 1635.50                    | 159                                  | 331.2                          |
| Nakhon Si Thammarat | 9,942.5            | 2381.30                    | 158                                  | 643.1                          |
| Phatthalung         | 3,424.47           | 1895                       | 143                                  | 465.8                          |
| Songkhla            | 7,393.89           | 2035.10                    | 153                                  | 587.9                          |

### 3.2 Data refinement

In this study, the eastern coast of southern rainfall data from the monitoring stations at HatYai district in Songkhla province, Tha Sae district in Chumphon province, Muang district in Nakhon Si Thammarat province, Kanchanadit district in Surat Thani province and Muang district in Phatthalung province are obtained from Thai Meteorological department and the Royal Irrigation department. The data obtained from Thai Meteorological department were recorded in 3-hourly raw data as shown in Table 3.2, while the data obtained from the Royal Irrigation department are in daily format as shown in Table 3.3. According to the daily rainfall prediction, the 3-hourly rainfall are accumulated into daily rainfall data. Due to missing records in the original rainfall data, the daily data of the 3-hourly missing rainfall are marked as missing data.

### 3.3 Data-preprocessing

Prior to the implementation of an ANN model, some preparation stages must be completed. These include data-filtering and data-scaling. In wavelet-transform based technique, rainfall series must be decomposed for filtering, as described in 3.3.1, into wavelet coefficients before the preprocessed normalization. ([2], [23], [33], [36]). In a conventional ANN technique, the data were also normalized before feeding to the network, as described in 3.3.2, due to the interval according to the output of the activation function ([5], [18]).

#### 3.3.1 Wavelet decomposition

The non-decimated wavelet decomposition is performed on the n-length of rainfall series to obtain n-length coefficients for each of the resolution levels [1]. The decomposition is accomplished by filtering the desired signal data with the chosen wavelet function at a number of decomposition level. Results of this filtering

Table 3.2: An example of 3-hourly raw data in HTML format obtained from the Thai Meteorological department; temperature data at Hatyai station in the year 1997.

อุณหภูมิคุ้มแห่ง(เซลเซียส)  
ราย 3 ชั่วโมง

| ที่ | รหัสสถานี-สถานี-จังหวัด | วันที่    | เวลาทำการตรวจ |      |      |      |      |      |      |      | เฉลี่ย |
|-----|-------------------------|-----------|---------------|------|------|------|------|------|------|------|--------|
|     |                         |           | 0100          | 0400 | 0700 | 1000 | 1300 | 1600 | 1900 | 2200 |        |
| 1   | 568502-หาดใหญ่ จ.สงขลา  | 1/1/1991  | 24.2          | 23.6 | 22.9 | 28.7 | 30.5 | 30.2 | 26.5 | 25.5 | 26.5   |
| 2   | 568502-หาดใหญ่ จ.สงขลา  | 2/1/1991  | 25.5          | 24.0 | 24.0 | 28.7 | 25.5 | 25.6 | 25.8 | 25.4 | 25.6   |
| 3   | 568502-หาดใหญ่ จ.สงขลา  | 3/1/1991  | 24.8          | 23.7 | 23.0 | 28.5 | 29.7 | 28.5 | 26.0 | 25.3 | 26.2   |
| 4   | 568502-หาดใหญ่ จ.สงขลา  | 4/1/1991  | 24.8          | 23.5 | 22.8 | 29.4 | 29.4 | 29.1 | 26.4 | 25.2 | 26.3   |
| 5   | 568502-หาดใหญ่ จ.สงขลา  | 5/1/1991  | 24.1          | 24.1 | 23.5 | 27.8 | 30.8 | 30.9 | 27.0 | 25.0 | 26.7   |
| 6   | 568502-หาดใหญ่ จ.สงขลา  | 6/1/1991  | 24.3          | 22.9 | 22.7 | 29.0 | 30.5 | 29.4 | 26.1 | 24.3 | 26.2   |
| 7   | 568502-หาดใหญ่ จ.สงขลา  | 7/1/1991  | 22.7          | 21.9 | 21.4 | 28.1 | 31.2 | 30.3 | 26.5 | 24.8 | 25.9   |
| 8   | 568502-หาดใหญ่ จ.สงขลา  | 8/1/1991  | 23.9          | 23.0 | 22.5 | 28.6 | 31.2 | 30.8 | 26.5 | 24.5 | 26.4   |
| 9   | 568502-หาดใหญ่ จ.สงขลา  | 9/1/1991  | 23.1          | 21.9 | 22.4 | 27.9 | 30.4 | 29.6 | 26.7 | 25.5 | 25.9   |
| 10  | 568502-หาดใหญ่ จ.สงขลา  | 10/1/1991 | 24.5          | 22.6 | 22.1 | 27.8 | 30.7 | 30.2 | 26.1 | 24.5 | 26.1   |
| 11  | 568502-หาดใหญ่ จ.สงขลา  | 11/1/1991 | 23.0          | 22.3 | 21.4 | 27.7 | 30.3 | 30.5 | 25.8 | 24.3 | 25.7   |
| 12  | 568502-หาดใหญ่ จ.สงขลา  | 12/1/1991 | 22.7          | 22.1 | 21.4 | 27.4 | 29.8 | 29.3 | 26.5 | 25.2 | 25.5   |
| 13  | 568502-หาดใหญ่ จ.สงขลา  | 13/1/1991 | 25.0          | 23.5 | 24.0 | 27.7 | 29.8 | 29.5 | 27.1 | 25.8 | 26.6   |
| 14  | 568502-หาดใหญ่ จ.สงขลา  | 14/1/1991 | 24.8          | 24.5 | 24.8 | 26.7 | 29.5 | 29.5 | 24.5 | 24.5 | 26.1   |
| 15  | 568502-หาดใหญ่ จ.สงขลา  | 15/1/1991 | 24.5          | 24.5 | 24.4 | 28.2 | 30.0 | 29.8 | 26.5 | 25.3 | 26.7   |
| 16  | 568502-หาดใหญ่ จ.สงขลา  | 16/1/1991 | 24.6          | 23.9 | 24.0 | 27.6 | 29.9 | 28.4 | 26.5 | 25.4 | 26.3   |
| 17  | 568502-หาดใหญ่ จ.สงขลา  | 17/1/1991 | 24.7          | 24.7 | 25.8 | 28.2 | 30.1 | 28.6 | 26.5 | 25.4 | 26.8   |
| 18  | 568502-หาดใหญ่ จ.สงขลา  | 18/1/1991 | 24.3          | 23.5 | 23.3 | 29.4 | 30.0 | 29.0 | 26.4 | 25.0 | 26.4   |
| 19  | 568502-หาดใหญ่ จ.สงขลา  | 19/1/1991 | 24.3          | 23.4 | 22.8 | 28.0 | 30.0 | 29.9 | 26.7 | 25.3 | 26.3   |
| 20  | 568502-หาดใหญ่ จ.สงขลา  | 20/1/1991 | 24.3          | 23.3 | 23.0 | 28.3 | 30.3 | 31.1 | 27.4 | 25.4 | 26.6   |
| 21  | 568502-หาดใหญ่ จ.สงขลา  | 21/1/1991 | 24.5          | 24.0 | 23.1 | 28.5 | 30.8 | 30.5 | 27.1 | 25.3 | 26.7   |
| 22  | 568502-หาดใหญ่ จ.สงขลา  | 22/1/1991 | 24.4          | 24.0 | 23.6 | 27.6 | 30.0 | 28.7 | 26.0 | 25.4 | 26.2   |
| 23  | 568502-หาดใหญ่ จ.สงขลา  | 23/1/1991 | 25.0          | 24.4 | 24.3 | 28.6 | 31.2 | 31.5 | 26.8 | 25.2 | 27.1   |
| 24  | 568502-หาดใหญ่ จ.สงขลา  | 24/1/1991 | 23.7          | 23.0 | 22.1 | 28.0 | 30.9 | 30.5 | 27.0 | 25.1 | 26.3   |
| 25  | 568502-หาดใหญ่ จ.สงขลา  | 25/1/1991 | 24.5          | 24.1 | 24.5 | 27.6 | 30.6 | 29.8 | 26.8 | 25.3 | 26.7   |



stage are the approximation (low frequency information) and detail (high frequency information) coefficient series for each resolution levels. Therefore, information at each resolution scale is directly related to each time point [22].

The number of resolution levels is experimentally chosen from the low frequency coefficient series (approximations) at the level such that the original data distribution is preserved. In this study, the resolution level is chosen to be 2. Among various choices of the mother wavelet, the Daubechies 2 wavelet is used in the present study for simplicity.

### 3.3.2 Linear Transformation

In the study, linear transformation is used to normalize the given data. The transformation is performed on both the input vectors and the targets. The outputs of the normalization corresponding with the logistic-sigmoidal function are real numbers between 0 and 1. The equation of linear transformation can be described as follows:

$$p_n = (b - a) \frac{p_0 - p_{min}}{p_{max} - p_{min}} + a, \quad (3.1)$$

where  $p_0$  is the observed data,  $p_n$  is the  $n^{th}$  scaled data,  $p_{max}$  and  $p_{min}$  are the maximum and minimum of the observed data, respectively, and  $[a, b]$  is the desired interval.

## 3.4 Data prediction

### 3.4.1 Artificial Neural Network

A multilayer feedforward neural network is applied in this study with one hidden layer. Resilient backpropagation is utilized to perform the weight adjustment, due to its fast and efficient computation [26]. The normalized data obtained from this linear transformation is then fed to the network in order to predict the rainfall data for  $n$  successive days. A set of feedforward backpropagation neural network with various activation functions is allocated to forecast the results.

#### Input node

In the thesis, rainfall data and climatological data of air temperature, humidity, pressure, wind speed, wind direction, and cloud amount are considered as features of the input layer. To predict the daily rainfall for  $n$  successive days, the number of preceding daily rainfalls and relevant climatological data are determined to find the significant correlation. The number of input nodes are, thus, experimentally found for the best ANN performance.



## Hidden node

Since the hidden layer introduces nonlinearity, the number of neurons in the layer are very important. If an inadequate number of neurons are used, the network will be unable to model complex data and the resulting fit will be poor. If too many neurons are used, the training time may become excessively long, and, worse, the network may over fit the data. When overfitting occurs, the network will begin to model random noise in the data. Therefore, the number of hidden nodes must be explored.

## Output node

The objective of the study is to predict the daily rainfall for  $n$  consecutive days. Therefore,  $n$  nodes of output are allocated with respect to each  $n$  successive days.

The prediction of time series using back-propagation neural network consists of teaching an ANN the historical data in a selected time and applying the taught information to the future data. Data from the past are provided to the inputs of neural network and from future to the output as the network prediction as shown in Fig. 3.3.

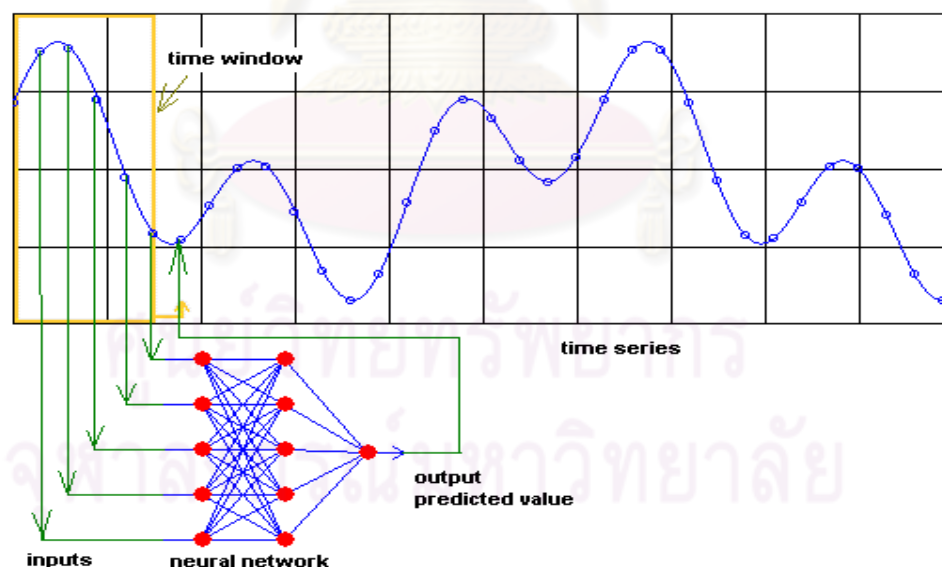


Figure 3.3: Artificial neural network related with time series.

By shifting the time-window over time series, the patterns of the network are made as shown in Fig. 3.4. These patterns can be adjusted for the needs of a particular neural network.

In this study, the network patterns are randomly separated into two parts; the training and the testing sets. The training set is used to train the neural network

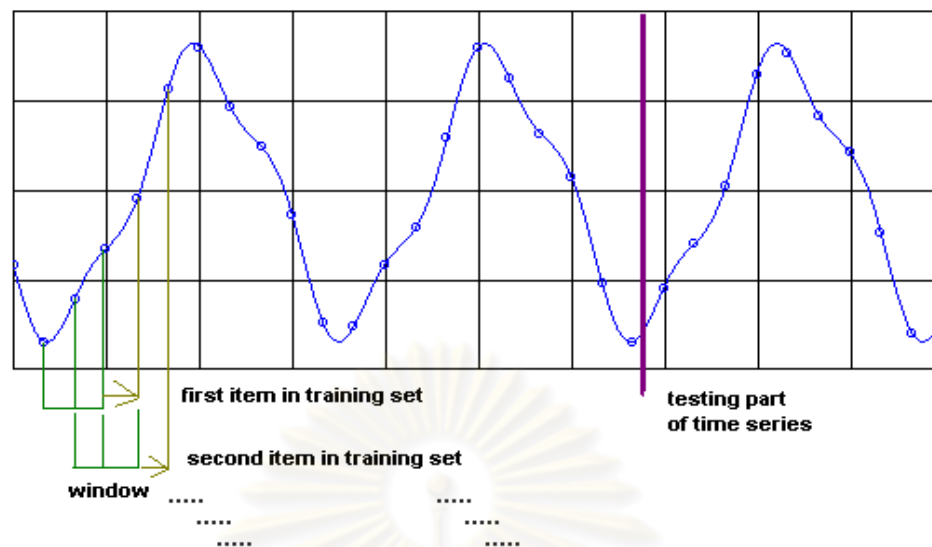


Figure 3.4: Creating training and testing set of the network.

model, whilst a testing set is used to verify the accuracy of the trained neural network model. The training patterns consist of 75% of the total dataset whereas the testing set consisting of the remaining 25%. The network pattern including the missing data are discarded.

In order to obtain the desired network, input patterns are feed to ANN with an appropriated number of input and hidden nodes. Training and learning functions are required in order to teach the network with a performance criteria. Training the network with the trainging set provides the desired weight associated with learning algorithm. Thus, the best ANN model is found for the prediction.

### 3.5 Data post-processing

In the wavelet-transform technique, rainfall data is decomposed into wavelet coefficients. Therefore, the wavelet predictions at different resolution levels obtained from a feedforward neural network are combined to reconstruct the original rainfall series. The reconstruction procedure of the original signal data is accomplished by recursively combining between an associated approximation and detail coefficients at a number of decomposition level. It reconstructs the original signal of rainfall series based on the multilevel wavelet decomposition.

The underlying idea of the wavelet-transform based ANN can be illustrated by Fig. 3.5.

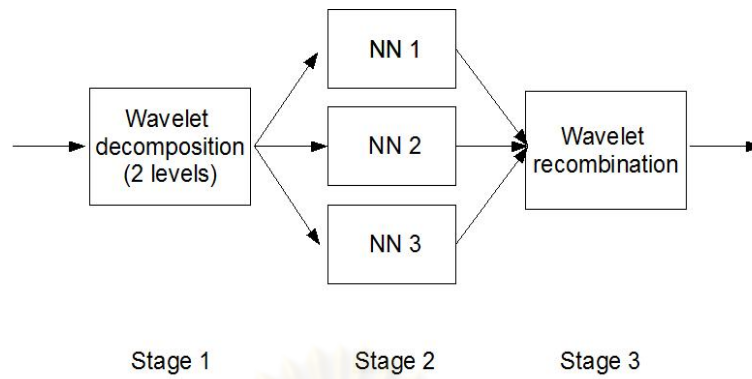


Figure 3.5: Diagram of wavelet-transform based artificial neural network.

### 3.6 Performance criteria

Performance of the network can be evaluated by various statistical measurements such as coefficient of determination ( $R^2$ ), root mean squared error ( $RMSE$ ), and correlation coefficient  $C$ . The notations used in each statistical formula are  $Q_{sim}(m)$ ,  $Q_{obs}(m)$ ,  $m$  and  $N$ , in which,  $Q_{sim}(m)$  and  $Q_{obs}(m)$  are the simulated and observed daily rainfall data at point  $m$ , respectively.  $N$  is the sample size.

#### 3.6.1 Root Mean Squared Error : $RMSE$

Root mean squared error (RMSE) is a statistical measure of the average magnitude over the verification sample of the squared values of the differences between the prediction and the corresponding observation. Since the errors are squared before averaging, the RMSE gives a relatively high weight to large errors. This implies that the RMSE is most useful when large errors are particularly undesirable. The formula for RMSE is

$$RMSE = \sqrt{\frac{1}{N} \sum_{m=1}^N (Q_{obs}(m) - Q_{sim}(m))^2}. \quad (3.2)$$

The RMSE can range from 0 to  $\infty$ . It is a negatively-oriented score – Lower values are better. The greater value of RMSE, the greater the variance in the individual errors in the sample. The RMSE is a statistical tool that give no direction of the prediction – underestimation or overestimation.

### 3.6.2 Coefficient of Determination : $R^2$

R-squared is one of the widely used statistics for the evaluation of model performance described as (3.3). The  $R^2$  ranges from zero to one; with zero indicating that the proposed model does not improve prediction over the predicted model and one indicating perfect prediction.

$$R^2 = 1 - \frac{\sum_{m=1}^N (Q_{obs}(m) - Q_{sim}(m))^2}{\sum_{m=1}^N (\bar{Q}_{obs} - Q_{obs}(m))^2}. \quad (3.3)$$

Note that, by this definition, the computational value of  $R^2$  can yield negative values which means the model failed to approximate the real data points.

### 3.6.3 Correlation Coefficient : $C$

Correlation coefficient measures the strength and the direction of a linear relationship between two variables. The mathematical formula for computing  $C$  is

$$C = \frac{N \sum Q_{obs}(m)Q_{sim}(m) - \sum Q_{obs}(m) \sum Q_{sim}(m)}{\sqrt{N(\sum Q_{obs}(m)^2) - (\sum Q_{obs}(m))^2} \sqrt{N(\sum Q_{sim}(m)^2) - (\sum Q_{sim}(m))^2}}. \quad (3.4)$$

The value of  $C$  lies between -1 and +1. The + and - signs are used for positive linear correlations and negative linear correlations, respectively. Positive values indicate a relationship between observed and predicted variables such that as values for observed data increases, values for the predictions also increase. On the other hands, negative values indicate a relationship such that as values for the observes increase, values for the predictions decrease. If there is no linear correlation or a weak linear correlation,  $C$  is close to 0.

# CHAPTER IV

## RESULTS AND DISCUSSION

This chapter will show results from the study. Daily rainfall in eastern coast of southern Thailand are predicted by two methods: artificial neural network and wavelet-transform based ANNs. In section 4.1, the predictions are obtained without any assistant of wavelet-transform based technique. The section 4.2 of this chapter shows an improvement of the prediction with wavelet-transform based artificial neural network. Experimental results and statistical evaluation of the rainfall models for performance comparison are presented and discussed in the last section.

### 4.1 Artificial neural network

Recently artificial neural network (ANN) as a non-linear inter-extrapolator is extensively used by hydrologists for rainfall modeling as well as other fields of hydrology ([20], [31]). In this study, a back-propagation ANN was applied in the prediction which is implemented in MATLAB 2007a. Daily rainfall data in southern Thailand during the years of 1995-2006 is randomly seperated into training and testing set. Resillient backpropagation (Rprop) learning algorithm is applied in the prediction in order to provide weight adjustments. Networks with only three layers (one hidden layer) were selected for all models.

#### 4.1.1 Input and hidden layer configuration

In this thesis, the input features of an artificial neural network (ANN) are determined from the available hydrological data – three-hourly rainfall and climatological data; pressure, air temperature, humidity, cloud amount, wind speed, and wind direction, at Hatyai station obtained from Thai Meteorological Department and daily rainfall data in Chumphon, Surat Thani, Nakhon Si Thammarat, and Phatthalung provinces obtained from Thai Royal Irrigation. To consider the appropriate input and hidden dimensions in the prediction, the number of nodes in the input and hidden layer are experimentally determined.

Improved performance of artificial neural network (ANN) is highly dependent on the selected input dimension. Fig. 4.1 shows the network accuracy of ANN model determined by using the constructive dimension of input nodes from 1 to 10. One layer of 200 hidden nodes with the output of next time-step daily rainfall in southern Thailand is used in the experiment. The graph shows that the network performance in term of  $R^2$  is optimal at 4 input-nodes. Using 9 and 10 input-nodes,

in this experiment, give negative values of  $R^2$ . An underlying reason is that these long-preceding rain events may take into account of overlapping between summer and rainy months. Different behaviors of rainfall such as the peak or zero precipitation are included in one input pattern which make some difficulties in a network learning. Therefore, the number of input nodes between 4 and 8 are proved to be the suitable input dimension in daily rainfall prediction.

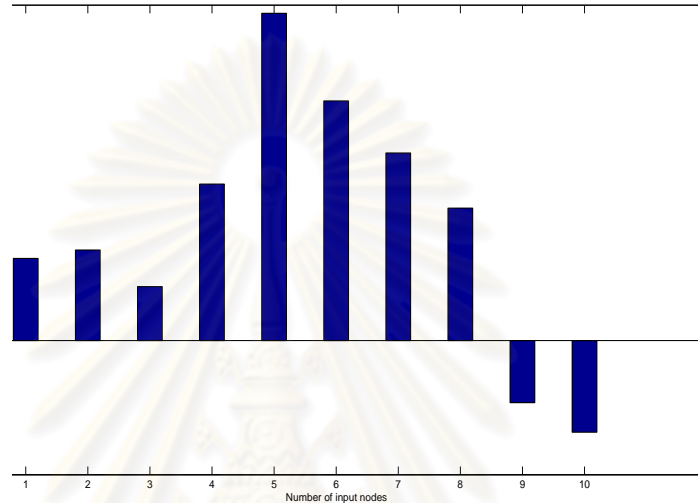


Figure 4.1: Performance of ANN model in terms of  $R^2$  with different number of input nodes of daily rainfall prediction in Southern Thailand with fixed hidden nodes at 200.

Table 4.1 illustrates the ANN performance of daily Southern rainfall in the selected areas with different number of input nodes and hidden nodes with various statistical evaluations. The input nodes are experimentally varied between 4 and 8 nodes for which the daily rainfall feature is used as the only predictor. The optimal size of the hidden node is found by increasing the number of hidden neurons by 100 between 100 and 500 of computation nodes. From the experiments on input and hidden nodes configuration, Fig. 4.2 shows that 5 computation nodes of input and 200 hidden nodes are suitable for network calibration and verification.

### 4.1.2 Activation function

In this section, various kind of activation functions are investigated with the 5-200-1 ANN architecture. Three kinds of activation function – logistic sigmoid, linear, and gaussian activation functions – are considered in both hidden and output layers. Before feeding data into the backpropagation neural network, the data needs to be normalized into the same range of output of the desired activation function.

From Table 4.2, ANN's performance on the training set is best when uti-

Table 4.1: ANN performance of daily Southern rainfall in selected areas with different number of input nodes and hidden nodes with various statistical evaluation.

(a) Root mean squared error:  $RMSE$

| Number of<br>input nodes | Number of hidden nodes |        |         |         |         |
|--------------------------|------------------------|--------|---------|---------|---------|
|                          | 100                    | 200    | 300     | 400     | 500     |
| 4                        | 10.4800                | 8.4215 | 8.4370  | 11.6410 | 9.4097  |
| 5                        | 10.3840                | 5.6147 | 7.7966  | 11.1920 | 9.5945  |
| 6                        | 8.8343                 | 7.5746 | 9.9639  | 10.7457 | 6.2259  |
| 7                        | 9.7059                 | 8.8129 | 8.9553  | 11.5114 | 10.7010 |
| 8                        | 8.4664                 | 7.8402 | 10.1142 | 9.0798  | 7.3126  |

(b) Correlation coefficient:  $C$

| Number of<br>input nodes | Number of hidden nodes |        |        |        |        |
|--------------------------|------------------------|--------|--------|--------|--------|
|                          | 100                    | 200    | 300    | 400    | 500    |
| 4                        | 0.5004                 | 0.4383 | 0.6620 | 0.5779 | 0.6377 |
| 5                        | 0.5135                 | 0.5651 | 0.6842 | 0.5371 | 0.1262 |
| 6                        | 0.2746                 | 0.5189 | 0.3956 | 0.2774 | 0.1028 |
| 7                        | 0.4773                 | 0.4650 | 0.6691 | 0.4052 | 0.5479 |
| 8                        | 0.3888                 | 0.4580 | 0.3085 | 0.4662 | 0.2600 |

(c) Coefficient of determination:  $R^2$

| Number of<br>input nodes | Number of hidden nodes |        |         |         |         |
|--------------------------|------------------------|--------|---------|---------|---------|
|                          | 100                    | 200    | 300     | 400     | 500     |
| 4                        | -0.5251                | 0.1166 | 0.0116  | -0.8817 | -0.2294 |
| 5                        | -0.3453                | 0.2439 | 0.1824  | -0.5626 | 0.0159  |
| 6                        | 0.0754                 | 0.1786 | 0.1564  | -0.1647 | -0.1595 |
| 7                        | -0.0664                | 0.1398 | 0.0921  | -0.5001 | -0.2963 |
| 8                        | -0.5776                | 0.0987 | -0.1527 | -0.8145 | -0.1769 |

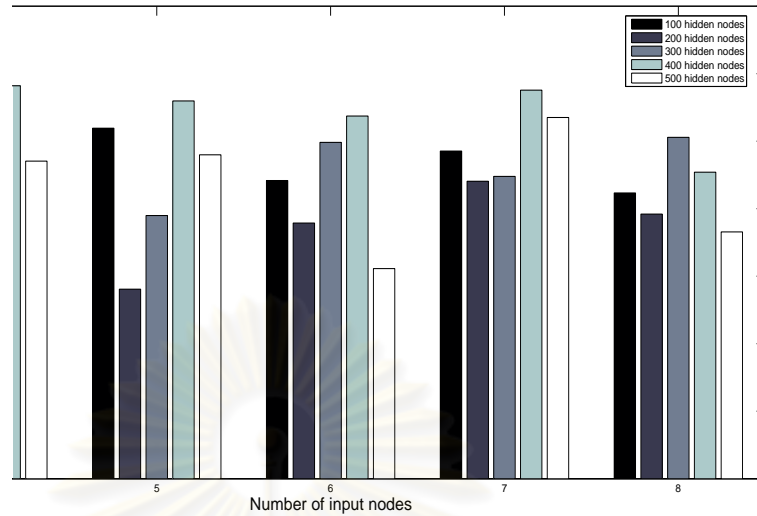


Figure 4.2: Performance of ANNs in term of RMSE with different number of input nodes and hidden nodes.

lizing Gaussian function in the hidden layer and linear function in the output layer. However, it fails to capture rainfall behavior of the testing set as indicated by the negative coefficient of determination ( $R^2$ ). Applying logistic sigmoid activation function in both hidden and output layers give satisfactory performance. Although the model provides less accuracy than the gaussian-linear functions in the network learning, the performance in the verification stage proves that the activation function of sigmoid-sigmoid model is capable to forecast the desired daily rainfall in selected area. Therefore, in this thesis, a back-propagation ANN with logistic sigmoid activation function in both hidden and output layers are applied for modeling and forecasting.

Relation between rainfall and climatological characteristics based on the physical evidence has been studied by some reseachers ([15], [18], [20], and [31]). Most of the studies could provide rainfall estimation with reasonable accuracy by taking into account climatological and topographical factors. If hydroclimatological characteristics indicate coincidental occurrence of rainfall events, the effects of enabling climatological variables as input features are evaluated in this section. These climatological variables are air temperature, humidity, pressure, wind speed, wind direction, and amount of cloud. ANN's performance of different climatological variables as rainfall predictors at Hatyai station between the years 1995 and 2006 is shown in Table 4.3. The model of daily rainfall, temperature and humidity with 15-100-1 architecture shows better accuracy than the others. The correlations between weather parameters and rainfall data, from Table 4.3, show that rainfall parameter exhibits a strong dependence on humidity parameter. The larger absolute value of significance test, the more relationship with rain the weather parameters are. Since wind direction variable has a small magnitude of significance, it gives the less correlation with rainfall data and provide less performance levels in forecasting.



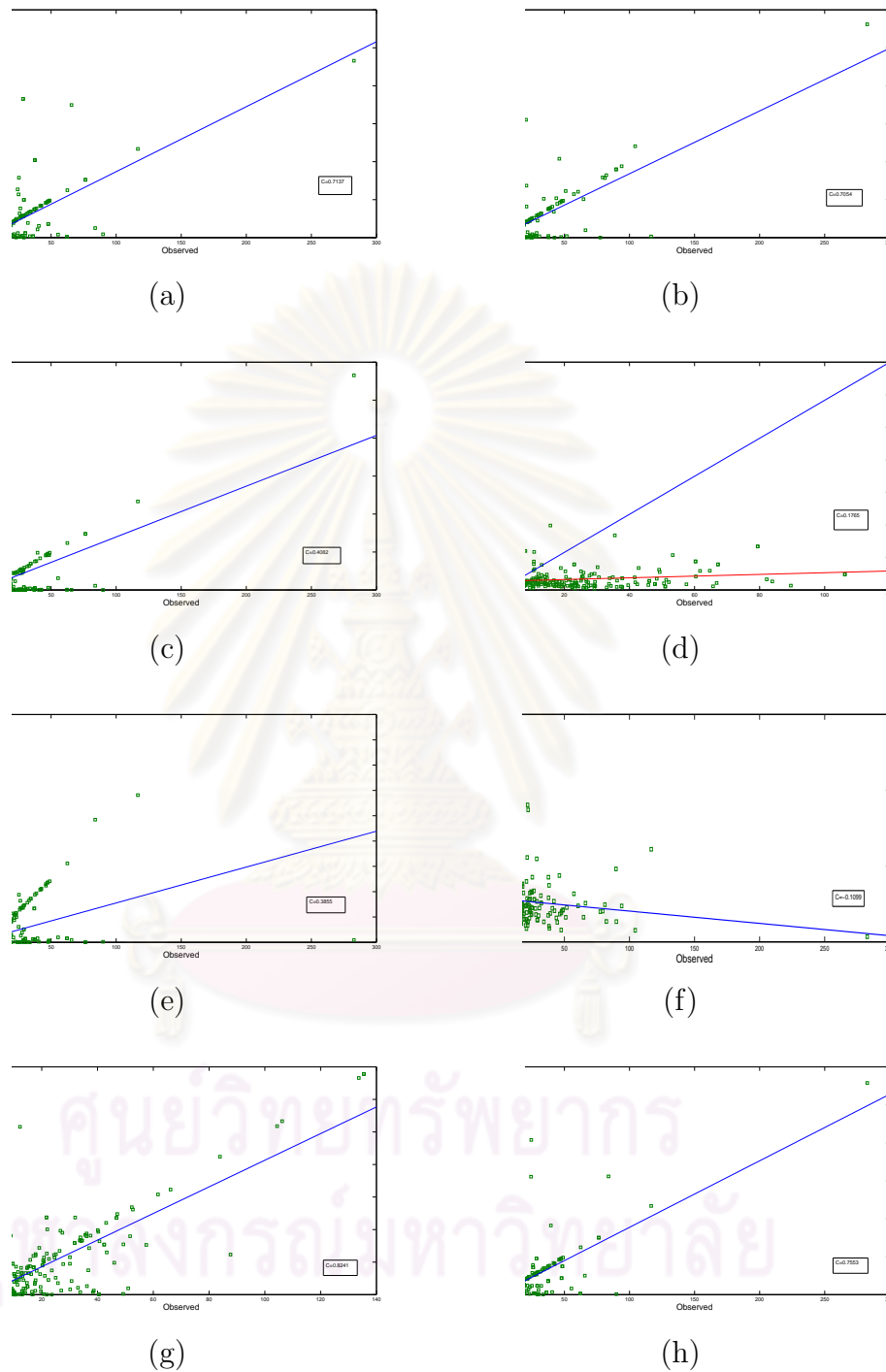


Figure 4.3: Scattered plots between observed and predicted daily rainfall in  $mm$  at Hatyai station with various input features: 4.3(a) using rainfall and humidity as input dimensions, 4.3(b) using rainfall and temperature as input dimensions, and 4.3(c) using rainfall and cloud amount as input dimensions, 4.3(d) using rainfall and pressure as input dimensions, 4.3(e) using rainfall and wind speed as input dimensions, 4.3(f) using rainfall and wind direction as input dimensions, 4.3(g) using rainfall, humidity and temperature as input dimensions, and 4.3(h) using rainfall, humidity, temperature, and cloud amount as input dimensions.

Table 4.2: Different activation function of ANN performance in the eastern coast of southern areas between the years 1995 and 2006.

| Act. function<br>in layer of |          | Training set |         |         | Testing set |        |         |
|------------------------------|----------|--------------|---------|---------|-------------|--------|---------|
| hidden                       | output   | RMSE         | C       | $R^2$   | RMSE        | C      | $R^2$   |
| sigmoid                      | sigmoid  | 5.5782       | 0.9104  | 0.8286  | 5.6147      | 0.5651 | 0.2439  |
| gaussian                     | sigmoid  | 6.5477       | 0.8882  | 0.7748  | 12.5767     | 0.3171 | -2.1878 |
| linear                       | sigmoid  | 13.5965      | 0.2252  | 0.0502  | 11.6474     | 0.3881 | -1.0266 |
| sigmoid                      | linear   | 7.2310       | 0.8761  | 0.7658  | 8.8617      | 0.3626 | -0.4498 |
| gaussian                     | linear   | 3.5366       | 0.9649  | 0.9311  | 10.7170     | 0.3039 | -1.1204 |
| linear                       | linear   | 12.8623      | 0.2975  | 0.0885  | 9.3376      | 0.3495 | 0.0060  |
| sigmoid                      | gaussian | 14.6649      | -0.0024 | -0.1749 | 8.9181      | 0.1205 | -0.1881 |
| gaussian                     | gaussian | 7.2856       | 0.8861  | 0.7714  | 17.5876     | 0.3304 | -5.2341 |
| linear                       | gaussian | 15.1991      | -0.1889 | -0.1869 | 10.0709     | 0.0582 | -0.1429 |

Fig. 4.3 illustrates scattered plots between observed and predicted daily rainfall in *mm* at Hatyai station with various input features of climatological variables. For the only rainfall predictor, air temperature and humidity have high achieved performance levels among others, as indicated by Figs. 4.3(a) and Fig. 4.3(b), respectively. This performance corresponds to the test of variable's significance toward rain events. Humidity and temperature variables produce good correlation coefficients at 0.5002 and -0.4002, respectively. These hydro-climatological features may link some occurrence in humidity and temperature with rain event. As shown by Fig. 4.3(g), the performance level is improved from that of Fig. 4.3(a) and Fig. 4.3(b). Then enabling of temperature and humidity together as part of the input dimensions results in considerable improvement of performance.

Nevertheless, the ANN model with employed humidity, temperature, and rainfall as input dimensions is still insufficient for daily rainfall prediction. The statistical evaluation in term of  $R^2$  is quite low although the correlation coefficient,  $C$ , which measuring the degree of correlation, is good at 0.8241. Fig. 4.4 exhibits the capability in capturing zero precipitation and some peak events. The difficulty in predicting moderate and heavy rain events, which oftens under and overfitting, proves to be the failure in this kind of prediction.

### 4.1.3 Period prediction

Distribution of rainfall in Southern Thailand has been influenced by seasonal monsoons. The Northeast and the Southwest monsoons play an essential role in generating rainfall in the south of Thailand. The period of October to February is referred to Northeast Monsoon season over the south peninsular of Thailand, while the months of June to September are referred to the Southwest monsoon.

Table 4.3: The ANN performance of different climatological variable as rainfall predictors at Hatyai station between the years 1995 and 2006.

| Input features | Correlation with rain | Architecture | Training |        |        | Testing |         |          |
|----------------|-----------------------|--------------|----------|--------|--------|---------|---------|----------|
|                |                       |              | RMSE     | C      | $R^2$  | RMSE    | C       | $R^2$    |
| R              | –                     | 5-200-1      | 5.5782   | 0.9104 | 0.8286 | 5.6147  | 0.5651  | 0.2439   |
| R              | 0.5002                | 8-200-1      | 2.5864   | 0.9755 | 0.9514 | 7.3516  | 0.7137  | 0.1809   |
| R              | -0.4002               | 10-200-1     | 2.6284   | 0.9798 | 0.9596 | 7.7043  | 0.7054  | 0.1969   |
| R              | 0.3188                | 8-100-1      | 2.6081   | 0.9752 | 0.9506 | 13.7487 | 0.4082  | -0.18648 |
| R              | -0.0467               | 10-500-1     | 3.2050   | 0.9764 | 0.9533 | 13.5714 | 0.1940  | -0.1706  |
| R              | -0.1715               | 8-200-1      | 3.7389   | 0.9484 | 0.8984 | 8.7341  | 0.3855  | -0.1561  |
| R              | -0.0355               | 10-500-1     | 3.0222   | 0.9731 | 0.9465 | 9.0571  | -0.1099 | -0.1765  |
| R              | –                     | 15-100-1     | 3.2939   | 0.9712 | 0.9429 | 4.6727  | 0.8241  | 0.6114   |
| R              | –                     | 16-200-1     | 2.5152   | 0.9768 | 0.9590 | 6.6164  | 0.7553  | 0.3366   |

Note that:

R refers to rainfall (mm),

H refers to humidity (percentage),

T refers to temperature (celcius),

C refers to cloud amount (deca),

P refers to pressure (hexto-pascal),

WS refers to wind speed (degree),

WD refers to wind direction (mile/hr).

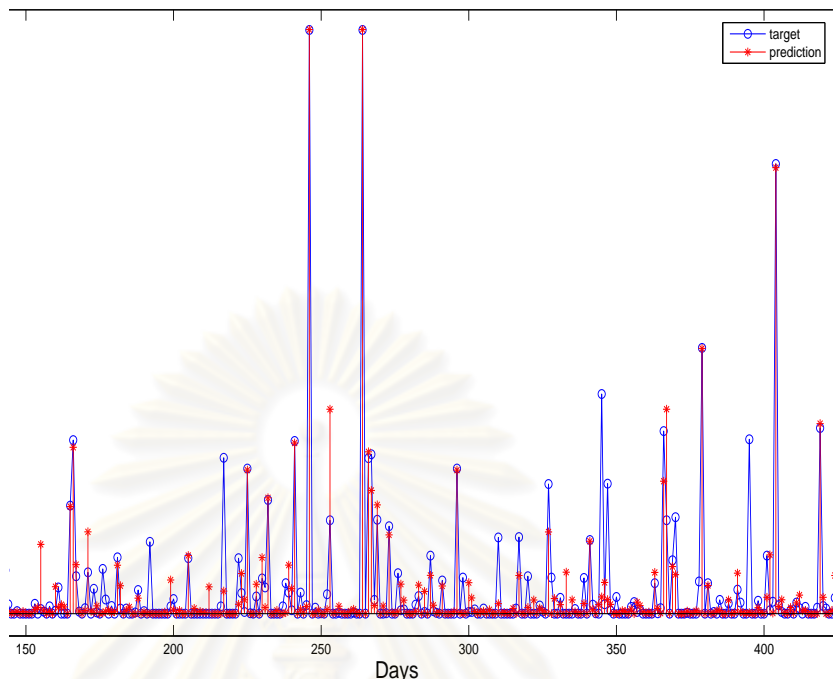


Figure 4.4: Daily rainfall prediction at Hatyai station using temperature, humidity and rainfall feature as input predictors.

Therefore, the rainfall distribution can be separately investigated into two periods – wet period and dry period, as shown in Fig. 4.5. The wet period is the period that Northeast and Southwest monsoons affect the southern weather. It begins in the month of June and ends by the month of February. The time interval between March and May is considered as a dry period.

In the split-data prediction, daily rainfall data together with climatological variables are split into wet and dry periods associated with the time-period of Northeast and Southwest monsoons. The back-propagation ANN with one hidden layer is implemented with logistic-sigmoid activation function in both layers. Resilient backpropagation (Rprop) learning algorithm is involved in the network's calibration.

According to Table 4.3, climatological variables of air temperature, humidity, and cloud amount are strongly related to rainfall data. Therefore, input features will be considered according to these climatological variables and rainfall data. Table 4.4 elucidates the ANNs' performance with various rainfall predictors of air temperature, humidity, and cloud amount in wet period, while those of dry period are shown in Table 4.5. The models are evaluated at Hatyai station during the year 1995 and 2006. It is worth seeing that the performances of the network calibration in both wet and dry periods achieved higher accuracy than that from models in the previous study. The improved performances prove that these kinds of

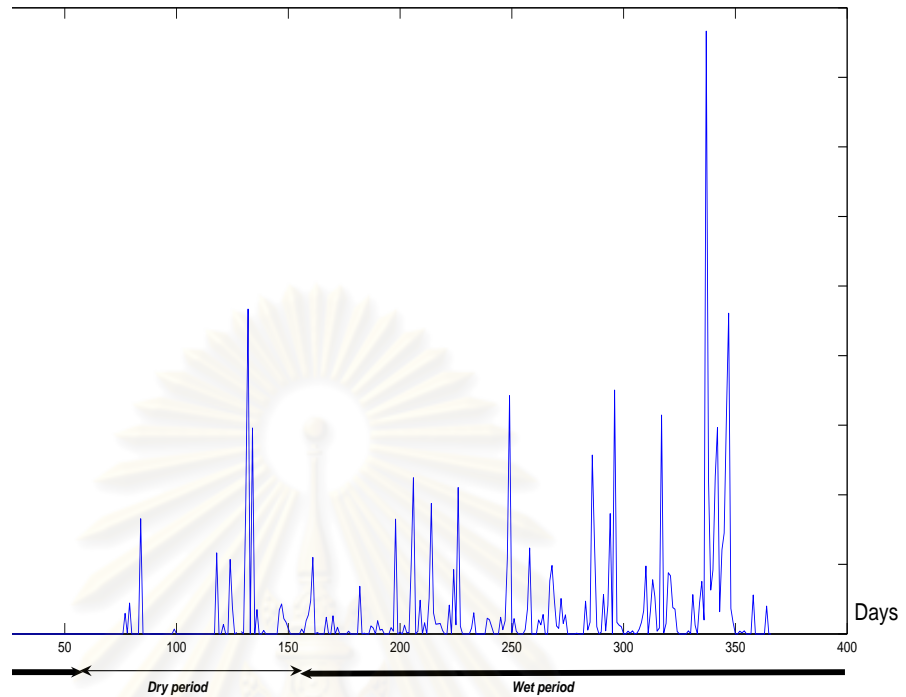


Figure 4.5: Rainfall distribution in Hatyai, Songkhla province.

network are capable of avoiding local extremum. Furthermore, ANN performances in testing set obtained from wet and dry period show an improvement over that of non-split model prediction. Using stronger related climatological variables as input feature, still, provided higher accuracy than weaker ones.

Fig. 4.6 illustrates the scattered plots in wet period prediction between observed and predicted daily rainfall in *mm* at Hatyai station with various input features of air temperature, humidity, cloud amount, and rainfall data itself. The model of rainfall, air temperature, and humidity shown in Fig. 4.6(e) gives good accuracy in term of  $R^2$  at 0.6295 and  $RMSE$  at 5.6443 mm. In dry period prediction, Fig. 4.7(b) shows that the model of rainfall and humidity data with 10-200-1 architecture outperforms from the other models with accuracy in terms of  $R^2$  and  $RMSE$  at 0.6075 and 3.7806 mm, respectively.

The daily rainfall at Hatyai station in wet period and that of dry period shown in Figs. 4.8 and 4.9 indicate that network's learning can not capture rainfall distribution in each time-period. The networks give under-estimations at peak events and over-estimate at some points. Moreover, the fitting graphs do not relate to the significance of the actual trend line of rainfall. It is possible that uncertainties monsoon may generate unexpected rain events in the periods. Fig. 4.10 shows that there are some shifting in rainfall. November usually should be the month that provide maximum precipitation, however, month of maximum precipitation varies and, sometimes, occurs in dry period. Therefore, these may be difficulties for

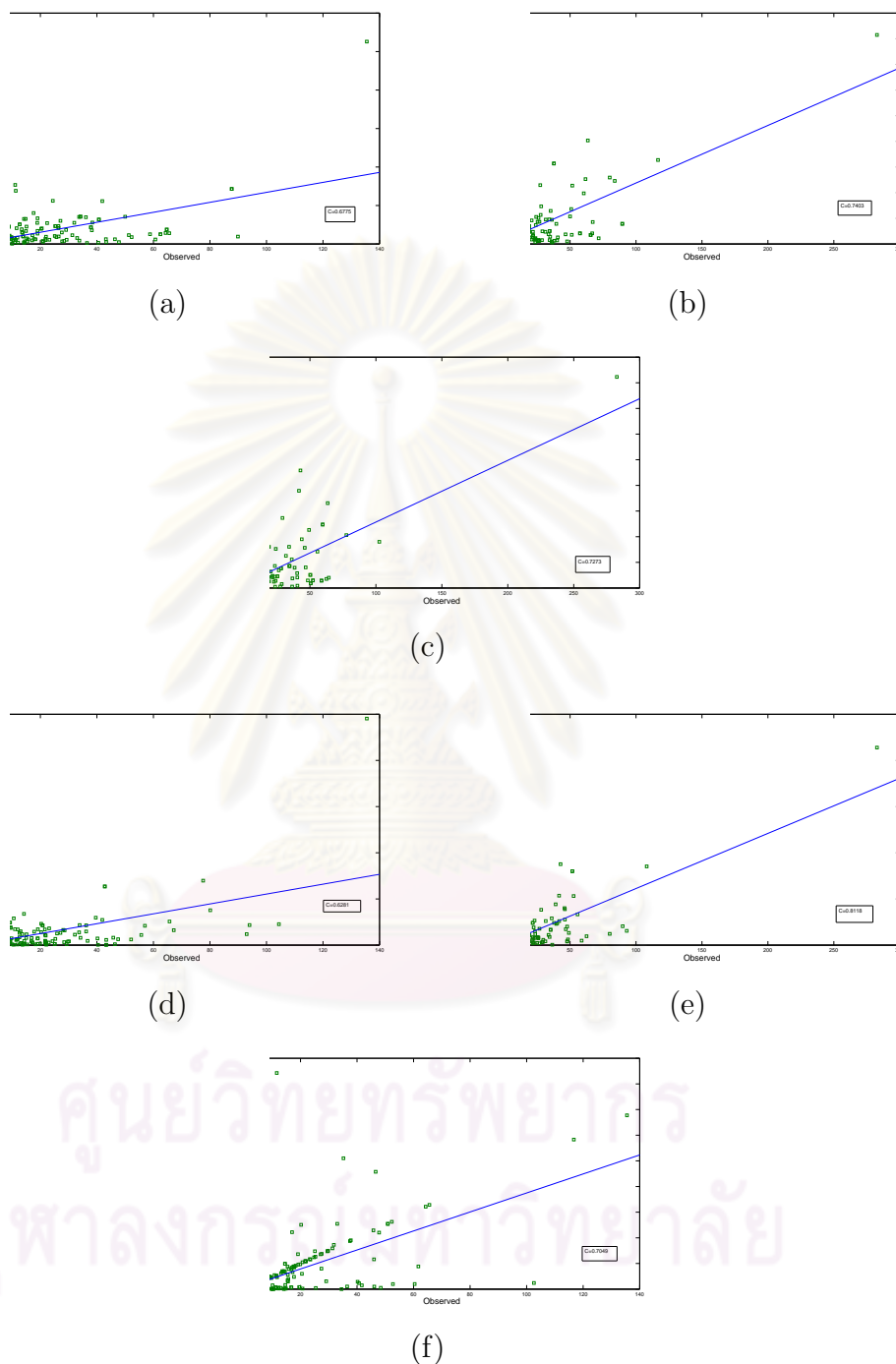


Figure 4.6: Scattered plots between observed and predicted daily rainfall in *mm* at Hatyai station with various input features in wet period: 4.6(a) using rainfall as the only input dimension, 4.6(b) using rainfall and humidity as input dimensions, 4.6(c) using rainfall and temperature as input dimensions, 4.6(d) using rainfall and cloud amount as input dimensions, 4.6(e) using rainfall, humidity and temperature as input dimensions, and 4.6(f) using rainfall, humidity, temperature, and cloud amount as input dimensions.

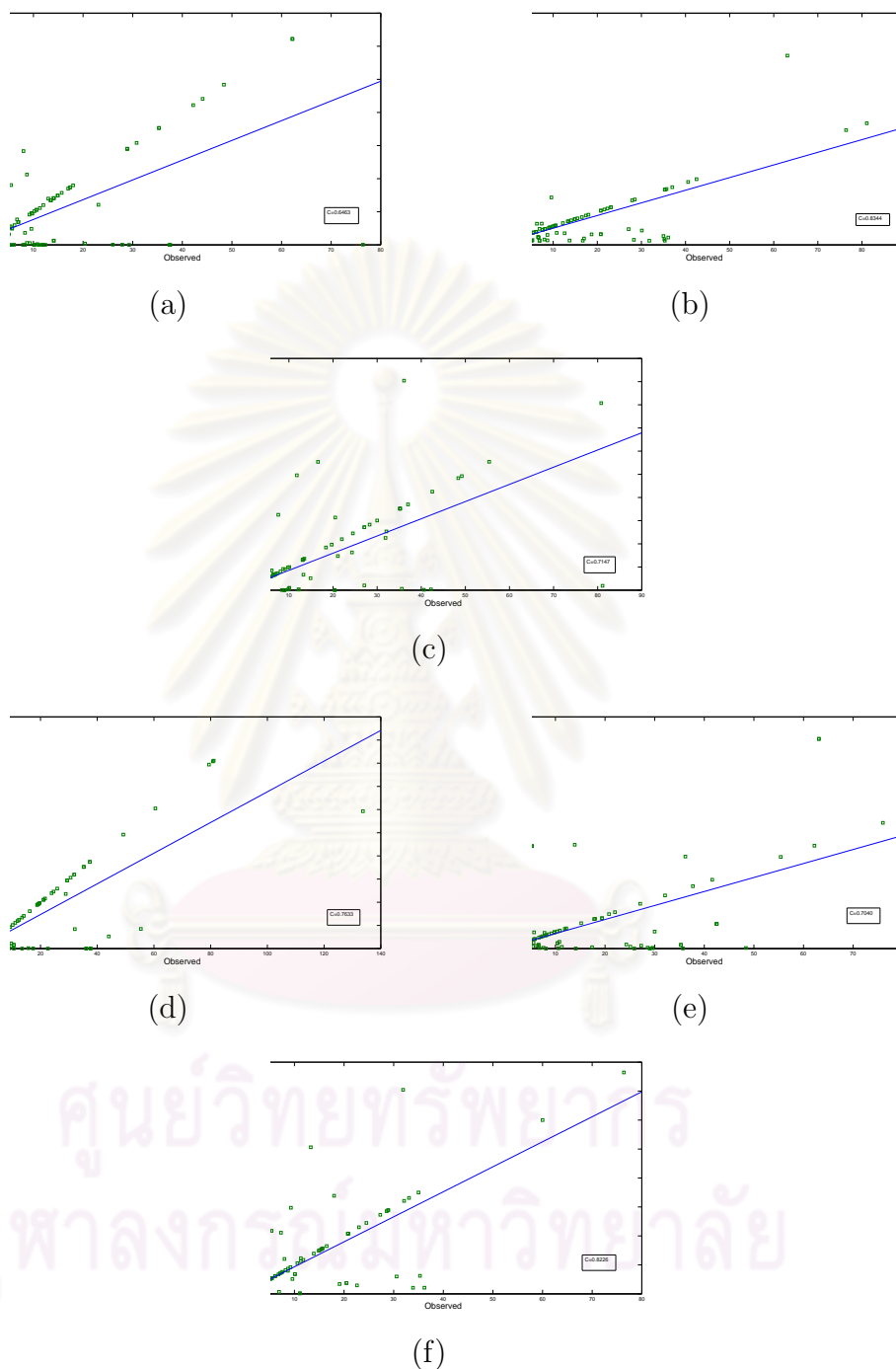


Figure 4.7: Scattered plots between observed and predicted daily rainfall in  $mm$  at Hatyai station with various input features in dry period: 4.7(a) using rainfall as the only input dimension, 4.7(b) using rainfall and humidity as input dimensions, 4.7(c) using rainfall and temperature as input dimensions, 4.7(d) using rainfall and cloud amount as input dimensions, 4.7(e) using rainfall, humidity and temperature as input dimensions, and 4.7(f) using rainfall, humidity, temperature, and cloud amount as input dimensions.

Table 4.4: The ANN performance of different climatological variable as rainfall predictors at Hatyai station in wet period between the years 1995 and 2006.

| Input features | Architecture | Training set |        |        | Testing set |        |        |
|----------------|--------------|--------------|--------|--------|-------------|--------|--------|
|                |              | RMSE         | C      | $R^2$  | RMSE        | C      | $R^2$  |
| R              | 5-200-1      | 1.4598       | 0.9906 | 0.9811 | 6.4068      | 0.6775 | 0.2780 |
| R,H            | 12-200-1     | 0.4559       | 0.9990 | 0.9980 | 6.9682      | 0.7403 | 0.5023 |
| R,T            | 10-100-1     | 0.7511       | 0.9973 | 0.9945 | 6.5034      | 0.7273 | 0.4929 |
| R,C            | 8-100-1      | 0.6276       | 0.9981 | 0.9962 | 6.4683      | 0.6281 | 0.2558 |
| R,H,T          | 18-100-1     | 2.1754       | 0.9744 | 0.9579 | 5.6443      | 0.8118 | 0.6295 |
| R,H,T,C        | 28-100-1     | 2.0291       | 0.9884 | 0.9769 | 6.0232      | 0.7049 | 0.2961 |

Table 4.5: The ANN performance of different climatological variable as rainfall predictors at Hatyai station in dry period between the years 1995 and 2006.

| Input features | Architecture | Training set |        |        | Testing set |        |        |
|----------------|--------------|--------------|--------|--------|-------------|--------|--------|
|                |              | RMSE         | C      | $R^2$  | RMSE        | C      | $R^2$  |
| R              | 6-400-1      | 1.2159       | 0.9956 | 0.9913 | 6.0487      | 0.7040 | 0.1325 |
| R,H            | 10-200-1     | 0.5488       | 0.9987 | 0.9973 | 3.7806      | 0.8344 | 0.6075 |
| R,T            | 12-100-1     | 0.7609       | 0.9975 | 0.9950 | 5.2120      | 0.7633 | 0.5173 |
| R,C            | 14-400-1     | 1.0038       | 0.9964 | 0.9928 | 5.5197      | 0.7147 | 0.3118 |
| R,H,T          | 18-200-1     | 0.5468       | 0.9988 | 0.9976 | 3.3118      | 0.8226 | 0.5684 |
| R,H,T,C        | 24-200-1     | 0.6743       | 0.9979 | 0.9959 | 5.1791      | 0.6463 | 0.2096 |

network learning in split-data prediction.

Based on the results from wet-period and dry-period predictions, the performance of models improves from that of the non-split data prediction. However, the improved accuracy shows that the models are not capable of forecasting daily rainfall in both wet period and dry period.

## 4.2 Wavelet-transform based Artificial neural network

From the previous section, the deterministic model – artificial neural network, failed to forecast daily rainfall data in the study areas. This may result from including of stochastic behavior in rainfall event. With the capability of separating a deterministic out of a stochastic part of wavelet decomposition, an artificial neural network is combined with the wavelet decomposition to improve models accuracy.



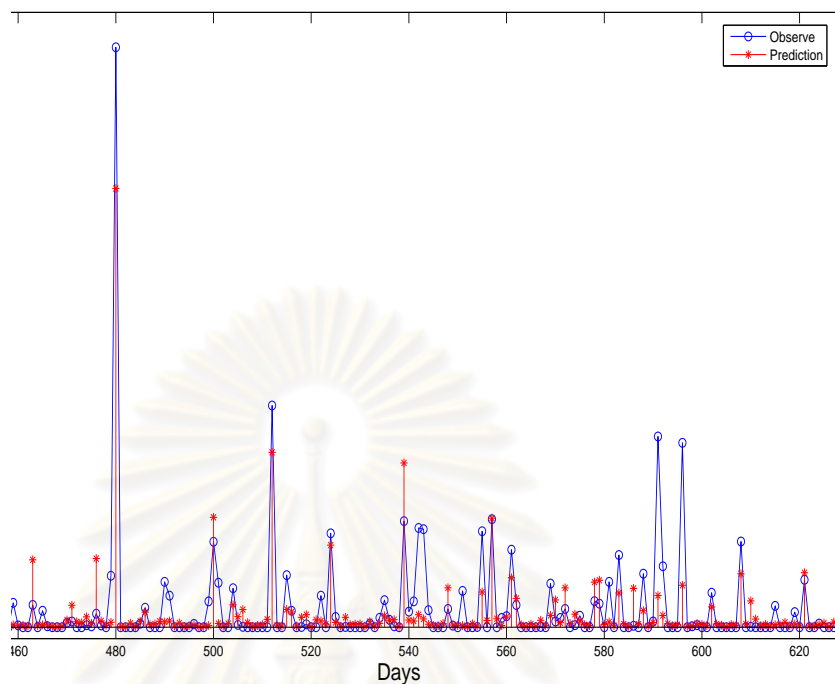


Figure 4.8: Daily rainfall at Hatyai station in wet period using rainfall, humidity and temperature variables as input features.

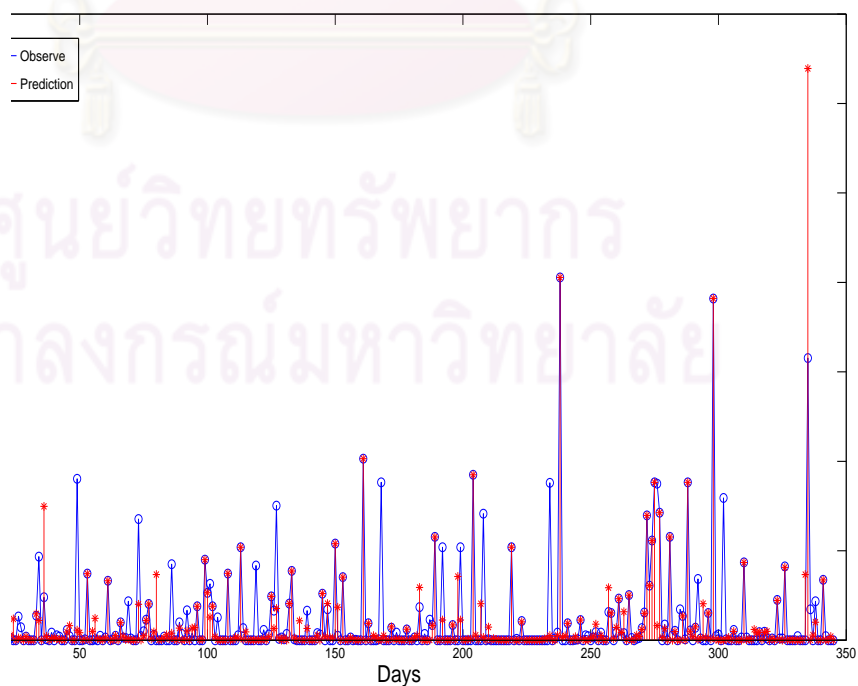


Figure 4.9: Daily rainfall at Hatyai station in dry period using rainfall and humidity variables as input features.

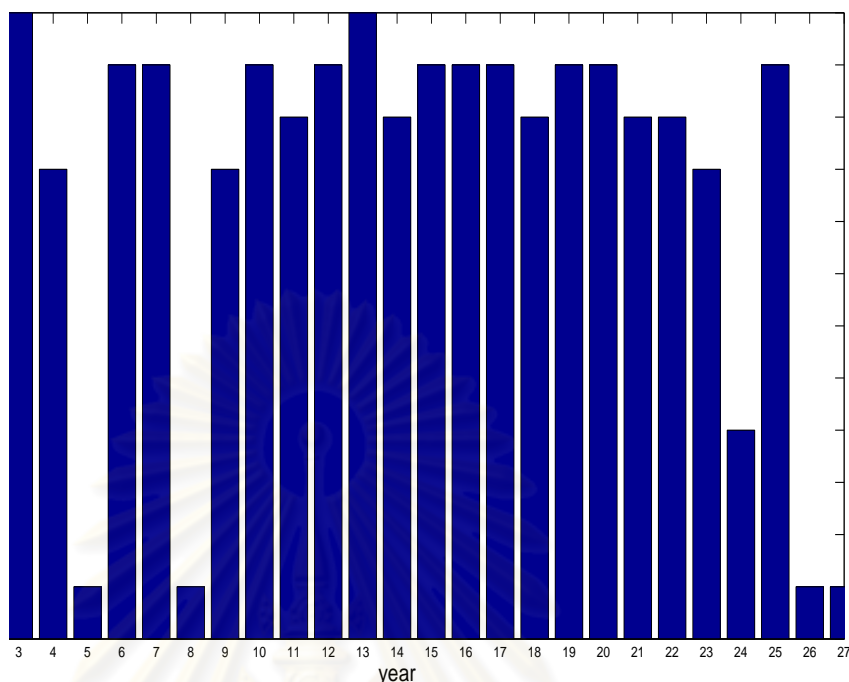


Figure 4.10: Graph of maximum-precipitation month at Hatyai station.

In section 4.1, implemented models considered various weather variables as input predictors in daily rainfall prediction. It might be the dependence of these factors in rainfall forecasting that affect the the accuracy of the prediction. This result associated with the study from French et. al [16]. The study concerns the uncertainties of hydrological variables that could affect the performances of both stochastic and deterministic rainfall prediction models. Therefore, in this section, accurate rainfall prediction based only on the collected historical data of rainfall is proposed.

Daily rainfall data at stations in the eastern coast of southern Thailand during the period of 1995-2006 are decomposed by the Daubechies 2 wavelet function at the second resolution level. Thus, two approximation and two detail coefficients are obtained. Figs 4.11 and 4.12 illustrate approximation and detail coefficients, respectively. As shown in Fig. 4.12, approximation coefficients at different level indicate daily rainfall series with different smooth fashion. The more resolution level, the more smoother daily rainfall series is. The wavelet coefficients of daily rainfall series at different resolution level during the period of 1995-2006 in southern stations are trained and tested with a number of different nodes of input and hidden neurons. Each daily rainfall data in the eastern coast of southern stations of years 1995 through 2003 are chosen to be the training set and those of years 2004 through 2006 are testing set. Before feeding data into this backpropagation neural network, the data is required to be normalized between 0 and 1 [12].

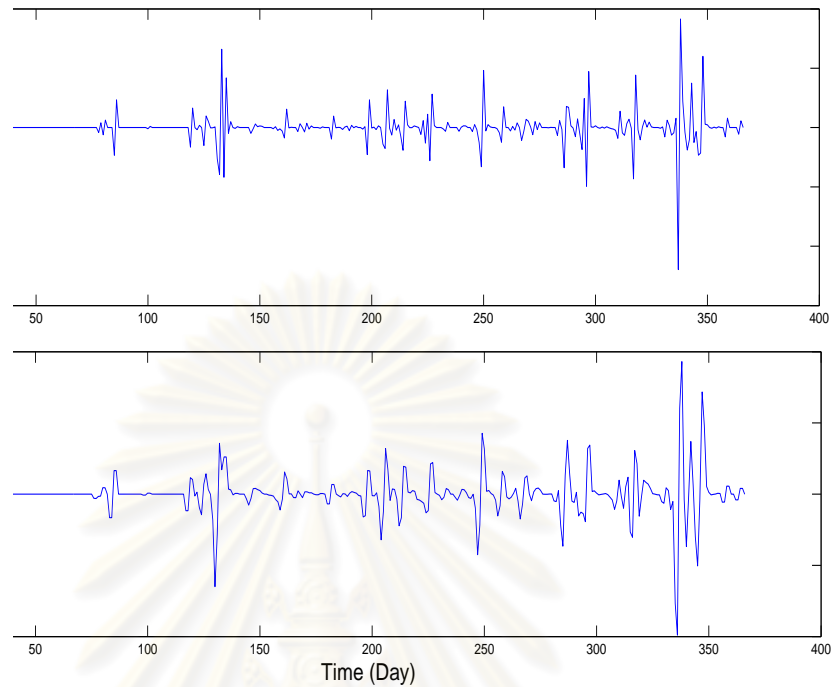


Figure 4.11: Detail coefficients of daily rainfall series at different resolution level; (up) at the first resolution level and (down) at the second resolution level.

### 4.2.1 One-day forecasting

An artificial neural network with wavelet decomposition is applied on daily rainfall data from five stations of the eastern coast of southern provinces. Only rainfall feature is considered as inputs of the network. For one-day forecasting, the output of the network is the next time-step daily rainfall. The network architecture of this one-day forecasting is 5–8–1 model which can be described mathematically as follows:

$$R_{t+1} = f(R_t + R_{t-1} + R_{t-2} + R_{t-3} + R_{t-4}), \quad (4.1)$$

where  $R_t$  represents daily rainfall of day  $t$ .

Table 4.6 shows a performance of wavelet based artificial neural network at different stations in the selected areas. The results show good accuracy with average  $R^2$  of 0.9946. One thing that should keep in mind is that  $R^2$ , in this study, regards as a measurement of a network's efficiency while  $RMSE$  measures a network's error. Moreover, performance in terms of  $R^2$  and  $RMSE$  might not be correspondent. The weight terms in their formula are responsible for this.  $RMSE$  is obtained from averaging the sum square error with sample size while  $R^2$  shows relative comparison between how far the prediction is from the observe and how far the observe is from its mean. Therefore, higher  $R^2$  could have higher  $RMSE$  than

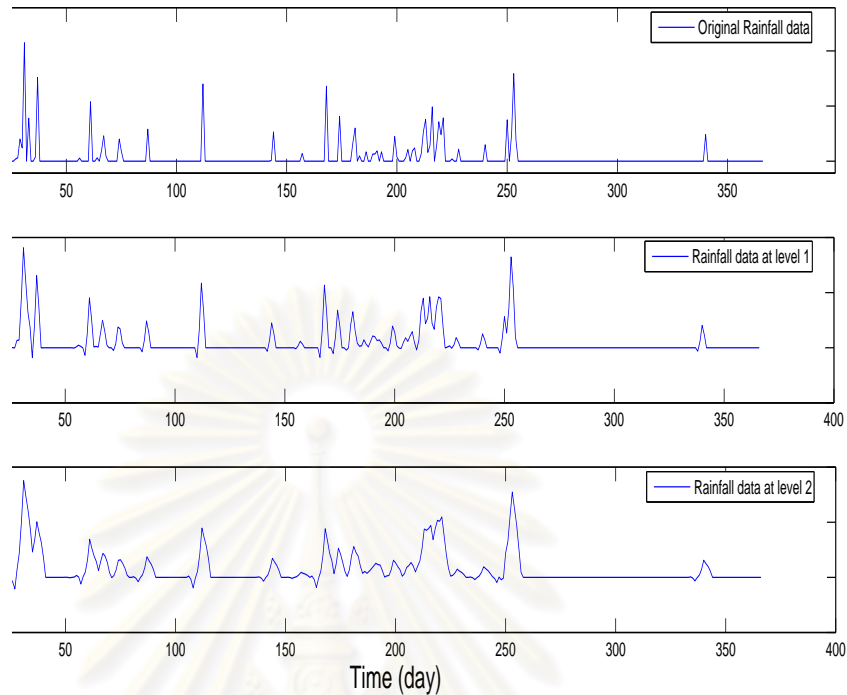


Figure 4.12: Wavelet decomposition of daily rainfall series at different resolution level.

lower  $R^2$ .

Table 4.6: One-day forecasting of daily rainfall in southern provinces.

| Station                       | Testing set |        |
|-------------------------------|-------------|--------|
|                               | RMSE        | $R^2$  |
| A. Kanchanadit, Surat Thani   | 0.8898      | 0.9946 |
| A. Tha Sae, Chumphon          | 0.7811      | 0.9939 |
| A. Muang, Nakhon Si Thammarat | 1.3964      | 0.9948 |
| A. Hatyai, Songkhla           | 0.8625      | 0.9950 |
| A. Muang, Phatthalung         | 0.9994      | 0.9947 |

An example of scattered plots of the simulation results based on daily rainfall data from the eastern coast of southern stations is illustrated in Figs. 4.13 and 4.14. Result of one-day daily rainfall output using a  $5 - 8 - 1$  ANN model shows that wavelet based artificial neural network is capable of forecasting daily rainfall at high accuracy with  $R^2 = 0.9950$  and  $RMSE = 0.8625$  mm.

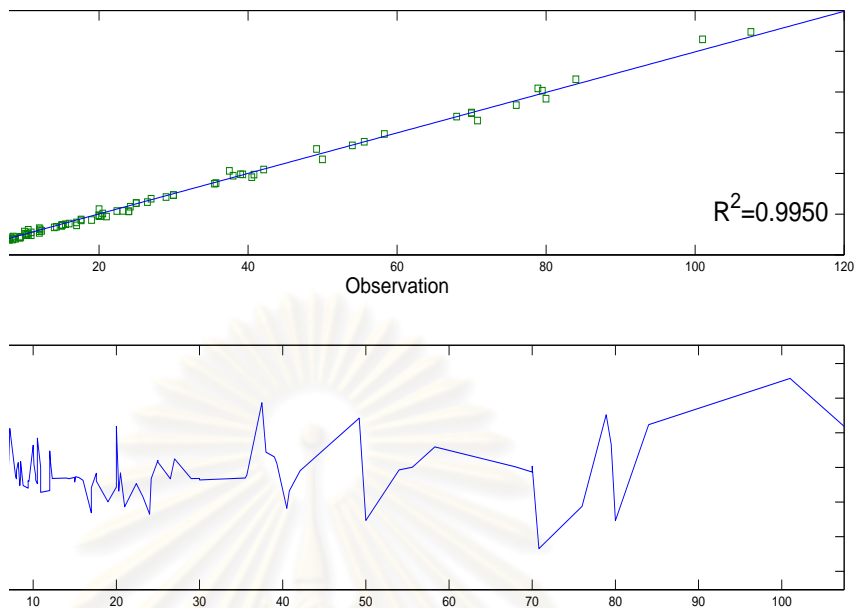


Figure 4.13: Scattered plots between observed and predicted daily rainfall in *mm* in the eastern coast of southern provinces for one-day forecasting.

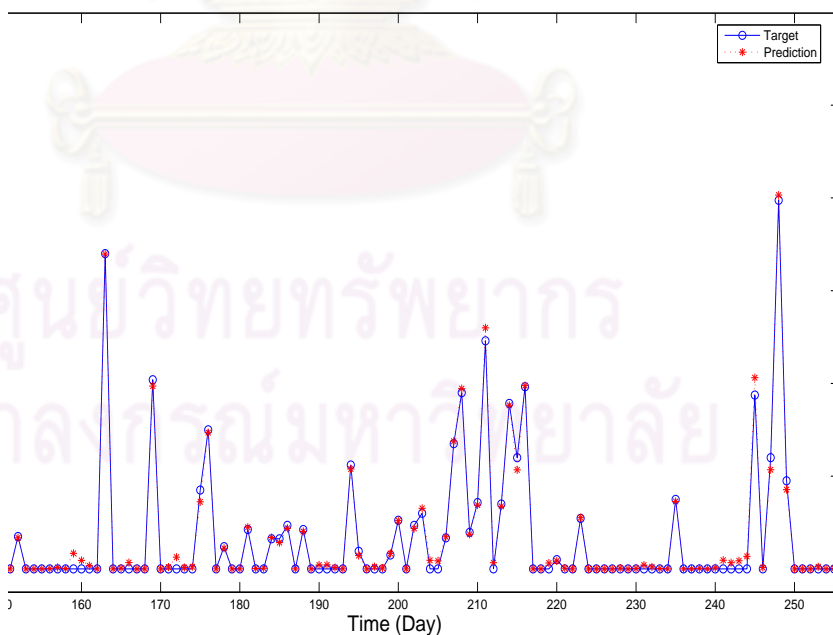


Figure 4.14: One-day daily rainfall prediction in the eastern coast of southern provinces.

## 4.2.2 Many-day forecasting

The capability of forecasting multi-step output of the wavelet-based artificial neural network is shown in Table 4.7. The daily rainfall forecasting in the

eastern coast of southern provinces for 2 and 3 days in advance gives a satisfactory accuracy. Simulation result of day 2 prediction shows accuracy with  $R^2 = 0.9739$  and  $RMSE = 2.0467 \text{ mm}$  while the network performance of day 3 prediction is  $R^2 = 0.9680$  and  $RMSE = 2.2699 \text{ mm}$ . Comparisons between the observed and simulated results in Fig. 4.15 and Fig. 4.16 show that extreme rainfall events can be detected for 2- and 3-day daily rainfall forecasting.

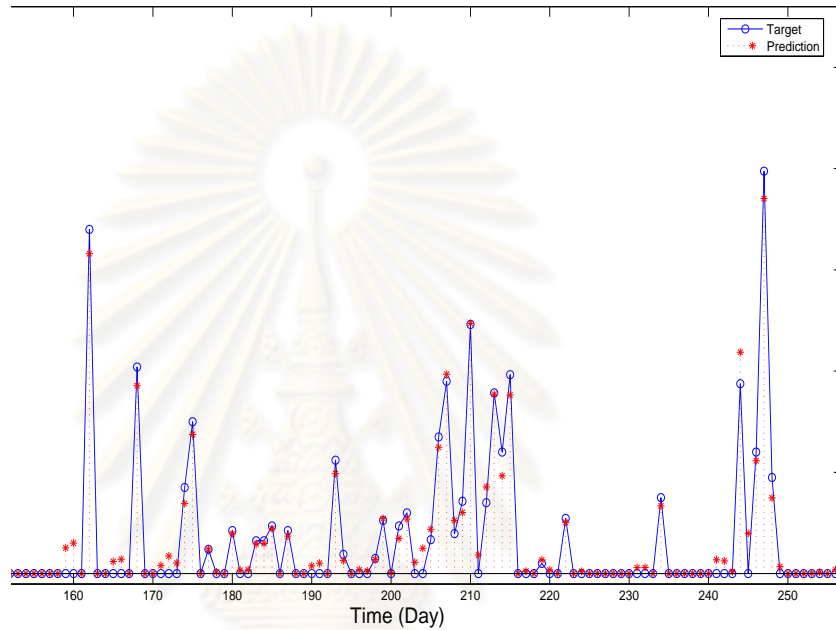


Figure 4.15: Day 2 rainfall prediction in the eastern coast of southern provinces.

The model can forecast daily rainfall up to 4 successive days with good accuracy. However, this reasonable fit of day 4 prediction gives the underestimate values of the peak events as shown in Fig. 4.19. When 5-day or more forecasting had been attempted, the performance of the network deteriorated as shown in Fig. 4.17.

Table 4.7: The eastern coast of southern-area average daily rainfall for  $n$  days prediction in selected provinces.

| Day | Training set |        | Testing set |        |
|-----|--------------|--------|-------------|--------|
|     | RMSE         | $R^2$  | RMSE        | $R^2$  |
| 1   | 1.4319       | 0.9929 | 0.9675      | 0.9942 |
| 2   | 2.7666       | 0.9737 | 2.0467      | 0.9739 |
| 3   | 3.1076       | 0.9668 | 2.2699      | 0.9680 |
| 4   | 5.4174       | 0.8990 | 4.2306      | 0.8887 |
| 5   | 11.1358      | 0.5735 | 8.7768      | 0.5206 |

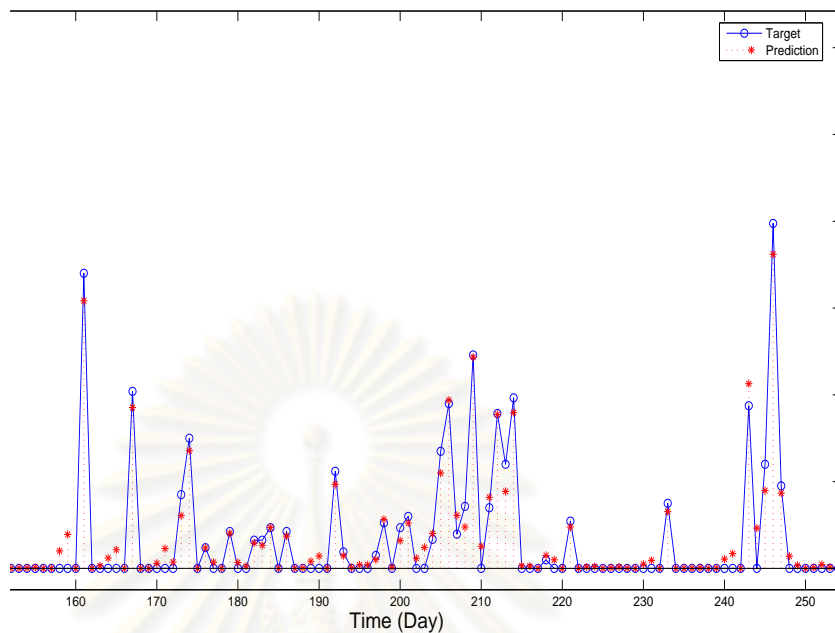


Figure 4.16: Day 3 rainfall prediction in the eastern coast of southern provinces.

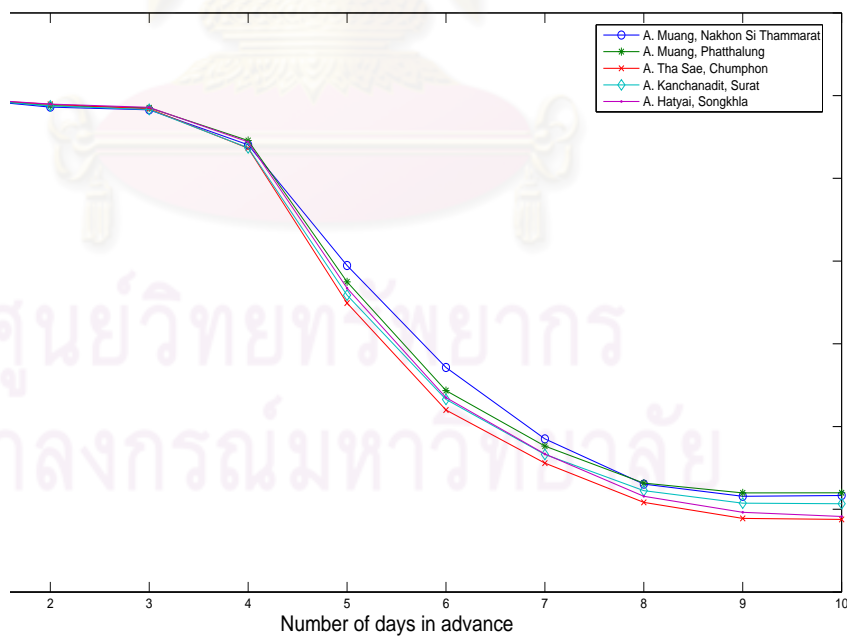


Figure 4.17: ANN performance for  $n$ -day daily rainfall prediction in the eastern coast of southern provinces.

Table 4.8 shows the performance of day 4 prediction of daily rainfall in selected provinces in the eastern coast of southern Thailand. The network architecture for this 4-day forecasting is  $5 - 19 - 4$ . The model input consists of 5 preceding

daily rainfall data to obtain predicted values of the next four daily rainfall as

$$(R_{t+1}, R_{t+2}, R_{t+3}, R_{t+4}) = f(R_t + R_{t-1} + R_{t-2} + R_{t-3} + R_{t-4}), \quad (4.2)$$

where  $R_t$  represents daily rainfall of day  $t$ .

Table 4.8: Day 4 prediction of daily rainfall in the eastern coast of southern provinces.

| Station                       | Testing set |        |
|-------------------------------|-------------|--------|
|                               | RMSE        | $R^2$  |
| A. Kanchanadit, Surat Thani   | 4.3135      | 0.8736 |
| A. Tha Sae, Chumphon          | 3.4144      | 0.8728 |
| A. Muang, Nakhon Si Thammarat | 6.6796      | 0.8812 |
| A. Hatyai, Songkhla           | 4.0671      | 0.8883 |
| A. Muang, Phatthalung         | 4.5271      | 0.8914 |

Fig. 4.18 shows another example of scattered plots between the observed and predicted daily rainfall data at the eastern coast of southern stations for day 4 prediction with accuracy of  $R^2 = 0.8883$ . Four-day daily rainfall forecasted at Hatyai station with  $RMSE = 4.0671$  mm is illustrated in Fig. 4.19.

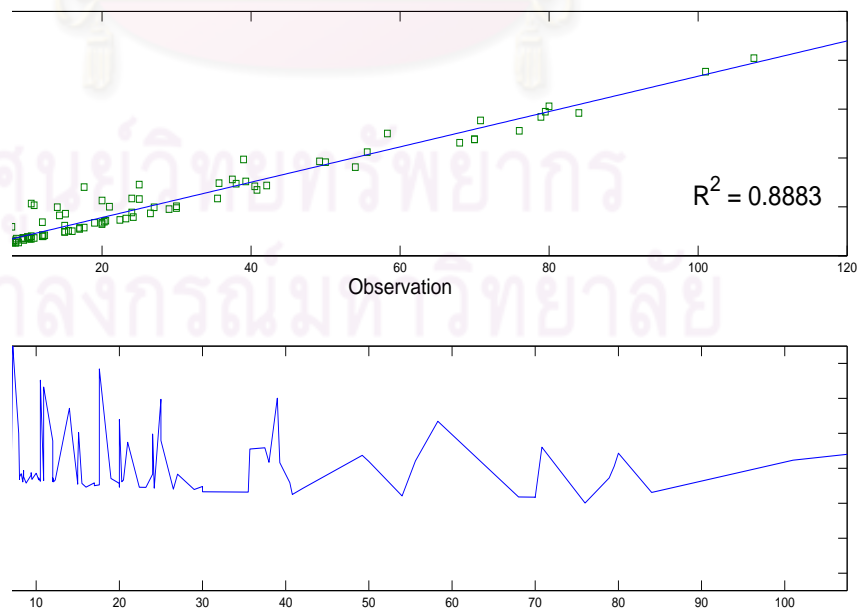


Figure 4.18: Scattered plots between observed and predicted daily rainfall in mm at the eastern coast of southern stations for day 4 forecasting.



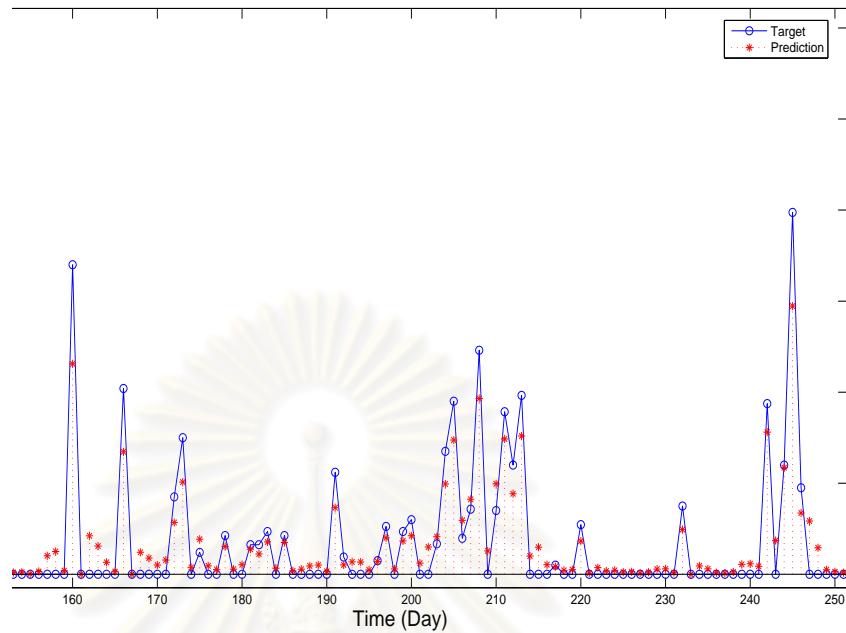


Figure 4.19: Day 4 daily rainfall prediction in the eastern coast of southern provinces.

### 4.3 Discussion

Fig. 4.20 illustrates daily rainfall models for performance comparison among a conventional ANN model, a split-data ANN model, and wavelet-transform based ANN model evaluated at selected provinces in the eastern coast of southern Thailand. Daily rainfall data are obtained by collecting distributed information in wet time-period. The graphs describe comparisons between observed and predicted rainfall data from various proposed models. The performance shows that the wavelet-transform based artificial neural network provides the most satisfactory performance with accuracy of  $R^2$  at 0.9948. The split-data model and the conventional ANN give accuracy of  $R^2$  at, approximately, 0.6295 which is insufficient to perform the southern daily rainfall prediction in this study.

The forecasting capability of wavelet decomposition technique outperforms the others. The wavelet based ANN model provides a good fit with the observed data, in particular for zero precipitation in the summer months, and for the peaks in the testing set of wet period. These results indicate that wavelet based ANN model estimations are significantly superior to those obtained by either the conventional ANN model or the split-data ANN model. Both models of non-wavelet decomposition technique are trapped in under- and overfitting. Although the split-data model is merely lightly better than the conventional ANN, all unnecessary information from summer monsoon uncertainties that may generate unexpected rain events on dry period is a difficulty in network learning. This shows that the wavelet transform can extract the chaotic components well from the original data

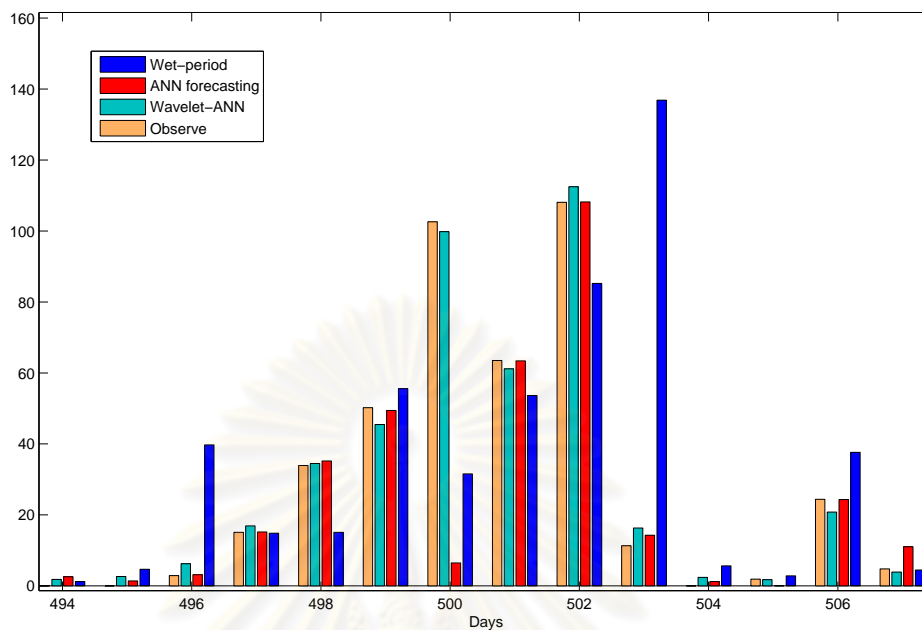


Figure 4.20: Performance comparison of southern daily rainfall among a conventional ANN model, a split-data ANN model, and wavelet-transform based ANN model.

for network learning.

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

## CHAPTER V

### CONCLUDING REMARK

Daily rainfall data obtained from various districts of provinces in the eastern coast of southern Thailand that are vulnerable to flood disaster: Tha Sae district in Chumphon province, Kanchanadit district in Surat Thani province, Muang district in Nakhon Si Thammarat province, Muang district in Phatthalung province and Hatyai district in Songkhla province, are investigated. A feedforward-backpropagation artificial neural network is applied to model and forecast southern daily rainfall data. Two conventional techniques of ANN model, e.g., split-data ANN model, and wavelet-transform based ANN, have been employed for comparison in this study. Statistical evaluation of the rainfall models for performance comparisons is summarized in terms of the coefficient of determination and the root mean squared error. Following previous works on forecasting the rainfall based on climatological variables, data of atmospheric pressure, air temperature, cloud density, humidity, wind speed, and wind direction are used as daily rainfall predictors in this thesis for the nondecomposed ANNs. Experimental results for nonsplit-data show that the 15-100-1 ANN model of rainfall, air temperature, and humidity variables, which is superior over the other models, is still trapped in under- and overfitting. The model accuracy in terms of  $R^2$  is 0.6114 and  $RMSE$  is 4.6727 mm. Split-data for two time-periods of wet and dry models also possesses some difficulties in forecasting accurate daily rainfall in southern Thailand. In the split-data prediction, rainfall data together with climatological variables are split into wet and dry periods associated with the time-period of Northeast and Southwest monsoon in order to implement models for each period. The best split-data model provides accuracy in term of  $R^2$  at 0.6295 and in term of  $RMSE$  at 5.6443 mm. The backpropagation artificial neural networks with one hidden layer shows that the models are not capable of forecasting daily rainfall in both wet period and dry period. It might be the dependence of climatological factors in rainfall forecasting and the complexity in rainfall behavior that may include some stochastic characteristics that affect the accuracy of the prediction. A hybrid technique using Wavelet transform and artificial neural network based only on the collected historical data of rainfall is proposed in order to predict accurate rainfall. In the wavelet-transform based ANN technique, prior to feeding input data of rainfall to ANN, information of rainfall series is extracted into two sets of wavelet coefficients – approximations and details. Results show that the neural network based on wavelet decomposition is preferable for daily rainfall prediction in the eastern coast of southern Thailand. Accuracy of the one-day daily rainfall prediction gives satisfactory prediction with  $R^2 = 0.9942$  and  $RMSE=0.9675$  mm. In addition, the network is also capable of forecasting up to 4 days in advance with reasonable accuracy of  $R^2 = 0.8887$  and  $RMSE = 4.2306$  mm.

Without a comprehensive technique of wavelet decomposition, the fore-

casting would not be offered valuable forecasting. For this reason, at the present time, wavelet decomposition plays an important role in extracting stochastic part out of deterministic part. Since the rainfall information is extracted when both high- and low-pass filters are applied at a specific resolution level, approximation and detail coefficients are the outputs which can be represented as rainfall trend and unexpected rain events, respectively. Multiresolution analysis allows trend behavior as the filters to extract noisy data in time series for the trained neural network. Therefore, this thesis demonstrates the crucial role of the wavelet-transform based artificial neural network as a practical tool for forecasting daily rainfall in the eastern coast of southern Thailand.



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

## REFERENCES

- [1] A. B. Geva, "ScaleNet-Multiscale neural network architecture for time series prediction," *IEEE Transaction on Neural Networks*, vol. 9, pp. 1471-1482, September 1998.
- [2] A. B. Mabrouk, N. B. Abdallah, and Z. Dhifaoui, "Wavelet decomposition and autoregressive model for time series prediction," *Applied Mathematics and Computation*, vol. 199, pp. 334-340, 2008.
- [3] A. Goudie, *The Nature of the Environment*, 4<sup>th</sup> Edition, Wiley-Blackwell, 2001.
- [4] A. W. Jayawardena, P. C. Xu, and F. L. Tsang, "Rainfall prediction by wavelet decomposition," *Hydrology and Earth Systems Science*, vol. 39, pp. 678-684, 1999.
- [5] A. W. Minns, and M. J. Hall, "Artificial neural networks as rainfall-runoff models," *J. Hydrol. Sci.*, vol. 41, pp. 399-417, 1996.
- [6] C. E. Imrie, S. Durucan, and A. Korre, "Riverflow prediction using artificial neural networks: generalisation beyond the calibration range," *J. Hydrol. Sci.*, vol. 233, pp. 138-153, 2000.
- [7] D. Briggs and P. Smithson, *Fundamentals of physical geography*, Rowman & LittleFeild, New Jersey, 1986.
- [8] D. E. Rumelhart, G. E. Hinto, and R. J. Williams, "Learning internal representations by error propagation," *Parallel Distributive Process*, vol. 1, pp. 218362, October 1986.
- [9] D. P. Solomatine, C. J. Rojas, S. Velickov, and J. C. Wüst, "Chaos theory in predicting surge water levels in the North Sea," 4<sup>th</sup> *International Conference on Hydroinformatics, Iowa, USA*, July 2000.
- [10] E. Toth, A. Brath, and A. Montanari, "Comparison of short-term rainfall prediction models for real-time flood forecasting," *J. Hydrol. Sci.*, vol. 239, pp. 132-147, September 2000.
- [11] K. L. Hsu, H. V. Gupta, S. Sorooshian, "Artificial neural networks modeling of the rainfall-runoff process," *water resources research*, vol. 31, no. 10, pp. 2517-2530, 1995.
- [12] L. Bodri and V. Čermák, "Prediction of extreme precipitation using a neural network: application to summer flood occurrence in Moravia," *Advance in Engineering Software*, vol. 31, pp. 331-321, 2000.
- [13] M. B. Ruskai, *Wavelets and their applications*, Jones and Bartlett Publishers, MA, 1992.
- [14] M. Campolo, P. Andreussi, and A. Soldati, "River flood forecasting with a neural network model," *Water resources research*, vol. 35(4), pp. 1191-1197, April 1999.
- [15] M. Nayebi, D. Khalili, S. Amin, and Sh. Zand-Parsa, "Daily stream flow prediction capability of artificial neural networks as influenced by minimum air temperature data," *Biosystems Engineering*, vol. 95(4), pp. 557-567, 2006.

- [16] M. N. French, W. F. Krajewski, and R. R. Cuykendall, "Rainfall forecasting in space and time using a neural network," *J. Hydrol.*, vol. 137, pp. 131, 1992.
- [17] M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: The RPROP algorithm," *Proceedings of the IEEE International Conference on Neural Network (ICNN)*, pp. 586-591, San Francisco, 1993.
- [18] N. Chantasut, C. Charoenjit, and C. Tanprasert, "Predictive Mining of Rainfall Predictions Using Artificial Neural Networks for Chao Phraya River," *The 4<sup>th</sup> International Conference of The Asian Federation of Information Technology in Agriculture*, August 2004.
- [19] N. J. de Vos, T. H. M. Rientjes, "Constraints of artificial neural networks for rainfall-runoff modeling: trade-offs in hydrological state representation and model evaluation," *Hydrology and Earth Systems Science.*, vol. 9, pp. 111126, 2005.
- [20] N. Q. Hung, M. S. Babel, S. Weesakul, and N. K. Tripathi, "An artificial neural network model for rainfall forecasting in Bangkok, Thailand," *Hydrology and Earth System Science*, vol. 5, pp. 183-218, 2008.
- [21] N. Singhrattna, B. Rajagopalan, M. Clark, and K. K. Kumar, "Seasonal forecasting of Thailand summer monsoon rainfall," *International Journal of Climatology*, November 2004.
- [22] O. Renaud, J. L. Starck, and F. Murtagh, "Wavelet-based combined signal filtering and prediction," *IEEE Transaction on System, Man, and Cybernetics*, vol. 35(6), 2005.
- [23] P. Cristea, R. Tuduce, and A. Cristea, "Time series prediction with wavelet neural networks," *5<sup>th</sup> Seminar on Neural Network application in Electrical Engineering*, September 2000.
- [24] P. J. Van Fleet, *Discrete wavelet transformations: An elementary approach with applications*, John Wiley & Sons, Inc, New Jersey, 2008.
- [25] P. S. Mathur, A. Kumar, and M. Chandra, "A feature based neural network model for weather forecasting," *International Journal of Computational Intelligence*, vol.3, no. 3, pp. 209-216,.
- [26] R. Bustami, N. Bessaih, C. Bong, and S. Suhaili, "Artificial neural network for precipitation and water level predictions of bedup river," *IAENG International Journal of Computer Science*, vol. 34:2, 2007.
- [27] R. Modarres, "Multi-criteria validation of artificial neural network rainfall-runoff modeling," *Hydrology and Earth Systems Science*, vol.13, pp. 411-421, March 2009.
- [28] R. VanRullen, "The power of the feed-forward sweep," *Advances in Cognitive Psychology*, vol. 3, no. 1-2, pp. 167-176, 2007.
- [29] S. Haykin, *Neural Network: A Comprehensive Foundation*, Prentice Hall International, New Jersey, 1999.
- [30] S. J. Cho and C. Hwang, "Statistical Method Development about the Spatial Relation between Climate and Land," *Proceeding: International Geoscience*

- and Remote Sensing Symposium (IGARSS)*, vol. 7, pp. 5137-5140, July 2005.
- [31] S. Lee, S. Cho, and P. M. Wong, "Rainfall prediction using artificial neural networks," *Journal of Geographic Informaiton and Decision Analysis*, vol. 2, pp. 233-242, 1998.
- [32] S. Soltani, "On the use of the wavelet decomposition for time series series prediction," *Neurocomputing*, vol. 48, pp. 267-277, 2002.
- [33] S. Tantanee, S. Patamakul, T. Oki, V. Sriboonlue, and T. Prempree, "Down-scaled rainfall prediction model (DRPM) using a unit disaggregation curve (UDC)," *Hydrol. Earth Sys. Sci. Discuss.*, vol. 2, pp. 543-568, 2005.
- [34] S. Vongvisessomjai, "Impacts of typhoon Vae and Linda on wind waves in the upper gulf of thailand and east coast," *Southern Thailand Environment Management Journal*, vol. 1, No. 1, 2009.
- [35] U. Weesakul, and S. Lowanichchai, "Rainfall forecast for agricultural water allocation planning in Thailnad," *Thammasat Int. J. Sc. Tech.*, vol. 10, No. 3, July-September 2005.
- [36] W. Tong, and Y. Li, "Wavelet Method Combining BP Networks and Time Series ARMA Modeling for Data Mining Forecasting," *IEEE Transaction on Neural Networks*, vol. 9, No. 4, pp. 345-356, 2001.
- [37] <http://www.math.ohiou.edu/shen/bookimages/book-figures1.html>
- [38] <http://www.sawater.com.au/SAWater/Education/OurWaterSystems/TheWater-Cycle.html>
- [39] <http://www.thailandtripbooking.com/thailand-map.html>
- [40] Y. Bi, J. Zhao, and D. Zhang, "Power load forecasting algorithm based on wavelet packet analysis," *International Conference on Power System Technology*, vol. 1, issue 21-24, pp. 987-990, November 2004.



## APPENDIX

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย



## APPENDIX

In this study, daily rainfall predictions are implemented in Matlab 2007a. The set of tools and facilities that help in the prediction are available in MaTLAB toolboxes. Many of these tools are in Wavelet toolbox and Neural Network toolbox. In Wavelet toolbox, the wavelet decomposition and recombination are introduced in sections 5.1 and 5.3, respectively. Section 5.2 describes ANNs' functions utilized in the prediction.

### 5.1 Wavelet decomposition

The decomposition is accomplished by using the function *swt* in wavelet toolbox of MATLAB. Its syntax is given as

$$[SWA, SWD] = swt(X, N, wname), \quad (5.1)$$

where

*SWA* refers to approximation of wavelet coefficients,  
*SWD* refers to detail of wavelet coefficients,  
*X* refers to signal data,  
*N* refers to number of decomposition level,  
*wname* refers to wavelet function.

### 5.2 ANN implementation

The neural network architecture is implemented via a *newff* function of neural network toolbox in MATLAB. Equation (5.2) is a syntax utilized to obtain the mentioned network.

$$net = newff(input\_pattern, [S_1 S_2], \{TF_1 TF_2\}, BTF, BLF, PF), \quad (5.2)$$

where

*S<sub>i</sub>* refers to number of hidden node in 1<sup>st</sup> layer and of output node in 2<sup>nd</sup> layer,  
*TF<sub>i</sub>* refers to the transfer function for *i*<sup>th</sup> layer,  
*BTF* refers to the training function,  
*BLF* refers to the learning function,  
*PF* refers to the performance criteria.

The following syntax describes the *train* function which used as a network training,

$$[net, tr, output, error] = train(net, Input\_patterns, Targets). \quad (5.3)$$

Equation (5.3) gives the trained network according to *net.trainParam* which is initially set as follows:

$$\begin{aligned} net.trainParam.epochs &= 100000, \\ net.trainParam.goal &= 0.000001. \end{aligned}$$

where,

*net.trainParam.epochs* represents the maximum epochs of the training,  
*net.trainParam.goal* represents the network performance goal in term  
of MSE for stopping condition.

The *train* function provides the following outputs: *net, tr, output*, and *error*, these correspond to the trained network, training information, network outputs, and network error, respectively.

### 5.3 Wavelet recombination

In the recombination procedure, the function *iswt* in wavelet toolbox, which is a inverse transform of *swt*, is utilized. The syntax of the function *iswt* is described as

$$X = iswt(SWA, SWD, wname), \quad (5.4)$$

where,

*SWA* refers to approximation wavelet coefficients,  
*SWD* refers to detail wavelet coefficients,  
*X* refers to original signal data,  
*N* refers to number of decomposition level,  
*wname* refers to wavelet function.

## VITAE

Miss Wassamon Phusakulkajorn was born in October 8, 1984, in Hatyai, Songkhla province. She received a bachelor degree (first-class honor) in Mathematics from Department of Mathematics, Faculty of Science, Prince of Songkla University, Thailand in 2006. She has been scholared by the Development and Promotion of Science and Technology talents project (DPST) to continue the study from bachelor degree to PhD degree.

## PUBLICATION

W. Phusakulkajorn, C. Lursinsap, and J. Asavanat, "Wavelet-transform based artificial neural network for daily rainfall prediction in Southern Thailand," *IEEE conference proceeding on Communications and Information Technologies 2009 (ISCIT2009), The 9<sup>th</sup> International Symposium on*, pp. 432 - 437, September, 2009.



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย