FLOOD AREA SIZE PREDICTION USING ARTIFICIAL NEURAL NETWORK: A CASE STUDY OF SUPHANBURI PROVINCE

SIRIPONG WERANANTAWAT

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
(TECHNOLOGY OF INFORMATION SYSTEM MANAGEMENT)
FACULTY OF GRADUATE STUDIES
2013

Thesis
entitled

FLOOD AREA SIZE PREDICTION USING ARTIFICIAL NEURAL NETWORK: A CASE STUDY OF SUPHANBURI PROVINCE

ศิริพงศ์ วีระอนันตวัฒน์
....
Mr. Siripong Weranantawat
Candidate

....
Asst.Prof. Bunlur Emaruchi,
Ph.D. (Environmental Systems Engineering)
Major advisor

Waranyu Wongseree
....
Lect. Waranyu Wongseree,
Ph.D. (Electrical Engineering)
Co-advisor

....
Lect. Supaporn Kiattisin,
Ph.D.(Electrical and Computer Engineering)
Co-advisor

B. Mahai
....
Prof. Banchong Mahaisavariya,
M.D., Dip Thai Board of Orthopedics
Dean
Faculty of Graduate Studies
Mahidol University

....
Lect. Supaporn Kiattisin,
Ph.D.(Electrical and Computer Engineering)
Program Director
Master of Science Program in Technology of Information System Management
Faculty of Engineering
Mahidol University

Thesis
entitled

FLOOD AREA SIZE PREDICTION USING ARTIFICIAL NEURAL NETWORK: A CASE STUDY OF SUPHANBURI PROVINCE

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Science
(Technology of Information System Management)
on
February 18, 2013

....
Mr. Siripong Weranantawat
Candidate

Chanattha Thongsuk.
....
Lect. Chanattha Thongsuk,
Ph.D. (Information Technology)
Chair

....
Lect. Supaporn Kiattisin,
Ph.D.(Electrical and Computer Engineering)
Member

....
Asst.Prof. Bunlur Emaruchi,
Ph.D. (Environmental Systems Engineering)
Member

Worasit
....
Asst.Prof. Worasit Choochaiwattana,
Ph.D. (Information Science)
Member

Waranyu Wongseree
....
Lect. Waranyu Wongseree,
Ph.D. (Electrical Engineering)
Member

....
Prof. Banchong Mahaisavariya,
M.D., Dip Thai Board of Orthopedics
Dean
Faculty of Graduate Studies
Mahidol University

....
Lect. Worawit Israngkul,
M.S. (Technical Management)
Dean
Faculty of Engineering
Mahidol University

ACKNOWLEDGEMENTS

The success of this thesis can be attributed to the extensive support and assistance from my major advisor, Asst.Prof. Bunlur Emaruchi, and my co–advisor, Lect. Waranyu Wongseree. and Lect. Supaporn Kiattisin. I deeply thank them for their valuable advice and guidance in this research.

I would like to acknowledge Royal Irrigation Department for providing the hydrological data and Geo-Informatics and Space Technology Development Agency for providing the flood shape file data.

My thanks also go to all the lectures of Technology Information System Management and Computer engineering in Faculty of Engineering, Mahidol University for their valuable advice.

Finally, I am grateful to my family for their financials support and love.

Siripong Weranantawat

FLOOD AREA SIZE PREDICTION USING ARTIFICIAL NEURAL NETWORK: A CASE STUDY OF SUPHANBURI PROVINCE

SIRIPONG WERANANTAWAT 5137555 EGTI/M

M.Sc. (TECHNOLOGY OF INFORMATION SYSTEM MANAGEMENT)

THESIS ADVISORY COMMITTEE: BUNLUR EMARUCHI, Ph.D., WARANYU WONGSEREE, Ph.D., SUPAPORN KIATTISIN, Ph.D.

ABSTRACT

Flooding has damaged many areas in Thailand every year. In order to evaluate the damages and tendency of flooding, this study used water level and flood shape files to predict the size of flood area occurrence focusing on Suphanburi Province. This paper presents the development of an artificial neural network for flood area size prediction. The Bayesian regularization training function was used for model development. The features, which are the water level data from each gauge station and floodgate, were used as input. The flood shape files were analyzed through GIS method to find the size of the flood area, and were used as quantitative output. In addition, the correlation coefficient was used for the input selection. The results indicated the performance of the artificial neural network suit predicted the dataset one day in advance with the selected input. Moreover, we have proven that the artificial neural network outperforms the conventional model.

KEY WORDS: FLOOD / PREDICTION / ARTIFICIAL NEURAL NETWORK

86 pages

การทำนายขนาดพื้นที่น้ำท่วมด้วยวิธีการโครงข่ายประสาทเทียม กรณีศึกษาจังหวัดสุพรรณบุรี
FLOOD AREA SIZE PREDICTION USING ARTIFICIAL NEURAL NETWORK: A CASE STUDY OF SUPHANBURI PROVINCE

ศิริพงศ์ วีระอนันตวัฒน์ 5137555 EGTI/M

วท.ม. (เทคโนโลยีการจัดการระบบสารสนเทศ)

คณะกรรมการที่ปรึกษาวิทยานิพนธ์ : บันลือ เอมะรุจิ, Ph.D., วรัญญู วงษ์เสรี, Ph.D., สุภาภรณ์ เกียรติสิน, Ph.D.

บทคัดย่อ

ภัยน้ำท่วมได้สร้างความเสียหายในหลายๆพื้นที่ของประเทศไทยทุกๆปี เพื่อประเมินความเสียหายและแนวโน้มของภัยน้ำท่วม ในการศึกษานี้ได้ใช้ข้อมูลระดับน้ำและข้อมูล Shape file ของพื้นที่น้ำท่วมมาประยุกต์ใช้ในการทำนายขนาดพื้นที่น้ำท่วมที่เกิดขึ้นในจังหวัดสุพรรณบุรี งานวิจัยนี้ได้นำเสนอการพัฒนาโมเดล Artificial neural network สำหรับการทำนายขนาดพื้นที่น้ำท่วม นอกจากนี้ Bayesian regularization training function ได้ถูกนำมาใช้ในการพัฒนาโมเดล ตัวแปรที่ใช้ในการพัฒนาโมเดลซึ่งได้แก่ข้อมูลระดับน้ำของแต่ละสถานีวัดระดับน้ำและประตูระบายน้ำตามจุดต่างๆถูกนำมาใช้เป็นตัวแปรนำเข้า ส่วนข้อมูล Shape fileของพื้นที่น้ำท่วมได้ถูกวิเคราะห์ผ่านวิธีการทางด้าน GIS เพื่อหาขนาดพื้นที่น้ำท่วมซึ่งถูกนำมาใช้เป็นผลลัพธ์ของการทำนาย นอกจากนี้วิธีการ Correlation coefficient ได้ถูกนำมาใช้สำหรับในการเลือกตัวแปรนำเข้า ผลลัพธ์ได้แสดงให้เห็นว่าประสิทธิภาพของ Artificial neural network เหมาะกับชุดข้อมูลในการทำนาย 1 วันล่วงหน้าที่ได้ทำการเลือกตัวแปรนำเข้าแล้ว นอกจากนี้เราได้พิสูจน์ประสิทธิภาพของ Artificial neural network ที่ได้พัฒนา ซึ่งได้ให้ผลลัพธ์ที่ดีกว่าโมเดลอย่างง่ายเช่น Regression model

86 หน้า

CONTENTS

			Page
ACKNOWLEDGEMENTS			**iii**
ABSTRACT (ENGLISH)			**iv**
ABSTRACT (THAI)			**vi**
LIST OF TABLES			**ix**
LIST OF FIGURES			**x**
CHAPTER I	**INTRODUCTION**		**1**
	1.1	Statement of problems	1
	1.2	Objective	2
	1.3	Scope of study	2
CHAPTER II	**LITERATURE REVIEW**		**4**
	2.1	Study area	4
		2.1.1 Location	4
		2.1.2 Topography	4
		2.1.3 Climate	5
		2.1.4 Water resource	5
	2.2	Definition and flooding characteristic	5
	2.3	Geography Information System(GIS)	6
	2.4	Artificial Neural Network	6
		2.4.1 Back propagation Neural Network(BPNN)	7
		2.4.2 The model development and BPNN procedures	8
	2.5	The ANN application for flood prediction	10
	2.6	The research related flood extent areas	13
CHAPTER III	**RESEARCH METHODOLOGY**		**15**
	3.1	Data collection	16
		3.1.1 Water level	16
		3.1.2 Flood shape files	16

CONTENTS (cont.)

Page

		3.3.1 Dataset	16
	3.2	Model development	17
		3.2.1 The network structure	17
		3.2.2 The activation function	18
		3.2.3 Training function	18
	3.3	Evaluation of model performance	19
		3.3.1 Mean Absolute Error (MAE)	19
		3.3.2 Correlation Coefficient (r)	19
		3.3.3 Mean Absolute Percentage error (MAPE)	20
	3.4	Material and development tools	20
	3.5	Schedule/Time plan	21
CHAPTER IV	**RESULTS AND DISCUSSION**		**22**
	4.1	The results of model development	22
		4.1.1 The results of 1 day in advance	22
		4.1.2 The results of 2 days in advance	23
	4.2	The results of model development with the feature selection	24
		4.2.1 The results of 1 day in advance (selected features)	24
		4.2.2 The results of 2 days in advance (selected features)	26
	4.3	The comparison of model performance	27
	4.4	The examples of predicted result	29
	4.5	The discussion of model development	35
CHARPTER V	**CONCLUSION AND RECOMMENDATIONS**		**37**
	5.1	Conclusion	37
	5.2	Recommendation and Future Work	38

CONTENTS (cont.)

	Page
REFERENCES	**39**
APPENDICES	**42**
Appendix A	43
Appendix B	48
BIOGRAPHY	**86**

LIST OF TABLES

Table		Page
3.1	List of gauge stations and floodgates	16
3.2	Grant chart: Steps of work	21
4.1	The results of 1 day in advance	23
4.2	The results of 2 days in advance	24
4.3	The results of 1 day in advance (selected inputs)	25
4.4	The results of paired samples statistics of 1 day in advance dataset	25
4.5	The results of paired samples test of 1 day in advance dataset	26
4.6	The results of 2 days in advance (selected inputs)	26
4.7	The results of paired samples statistics of 2 day in advance dataset	27
4.8	The results of paired samples test of 2 day in advance dataset	27
4.9	The comparison between neural network and linear regression model	28
4.10	The results of paired samples statistics of model	28
4.11	The results of paired samples test of model	28
4.12	The examples of predicted result	29
4.13	The summary of proper dataset for prediction	35

LIST OF FIGURES

Figure		Page
2.1	Map of Suphanburi province	4
2.2	Structure of Multilayer Preceptron	7
2.3	The characteristic of Back Propagation	8
2.4	The process chart of model development	8
2.5	The chart of back propagation steps	9
3.1	Flow Chart of Research Methodology	15
3.2	The location of gauge stations and floodgates	17
3.3	The network structure	18
4.1	Flood area size 475.02 km^2 for the first example	30
4.2	Flood area size 475.02 km^2 for the second example	30
4.3	Flood area size 513.88 km^2 for the third example	31
4.4	Flood area size 341.65 km^2 for the fourth example	31
4.5	Flood area size 702.04 km^2 for the fifth example	32
4.6	Flood area size 527.06 km^2 for the sixth example	32
4.7	Flood area size 634.93 km^2 for the seventh example	33
4.8	Flood area size 737.03 km^2 for the eighth example	33
4.9	Flood area size 1758.17 km^2 for the ninth example	34
4.10	Flood area size 1904.96 km^2 for the tenth example	34
4.11	Flood area size 1577.71 km^2 for the eleventh example	35

CHAPTER I
INTRODUCTION

1.1 Statement of problems

Flooding is natural phenomena that occur everywhere. In each year, flood disaster has more serious and the values of damages amount many billion baht, which government agencies charged in the restoration of country. Flood disaster not only affecting human' life and properties, but also natural resource damage and leading to environmental problems.

The major flood occurred in year 2011. Thailand had been damaged by flooding, which is widely spread in many provinces, especially in the central region of the country, which cause from tropical storms during the monsoon season, causing heavy rains in northern and north-east regions and happen flooding in many provinces. While the rain is still heavily fall. In a few time, the inundation spread to the lower part of the country. When the Chao Phraya River receives mass water from branches of river and spread to other branches that resulting of flooding for many provinces of the central plain areas. Due to the large amount of water overflow river that exceed the capacity of river and rush down to the lower part of the country.

In addition, the expansion of city and the invasion of land in the river area, making physical areas changed such as an increase in the impermeable area, which affecting the more run-off, while infiltration of water into ground are decreased. In this situation will cause the flooding is large areas until the water level in the main rivers will decrease that level of water can drain out.

Although the mitigation of flood problems are helped both short and long terms from relevant agencies. The flooding extent map is also important for providing information to many aspects such as the analysis of flooding risk areas, warning and evacuate.

At present, there are provide flood area information that displayed the satellite maps and tracking the daily flood situation. Geo-Informatics and Space

Technology Development Agency (Public Organization) is also known as GISTDA that provides geospatial information using satellite image. The satellite information will be used to prepare a map showing the flooding and mudslide areas along with the damage have occurred. This information can be analyzed to assess the damage from flood disaster and support for people who are be in trouble.

Furthermore, there are the related researches that analysis to display the flood extent areas by using hydrological model with geographic information system. Nevertheless, the most of study areas are valley which cannot be applied to plain areas because of the height lines of terrain (contour line) are inadequate resolution and causing mistake for display flood extent areas and unreliable.

In year 1980, Artificial Neural Networks (ANNs) are applied to solve science and engineering that based on historical data for training and testing process. In general, the advantage of ANN model is a minimum requirement of data than hydrological models. Moreover, the user don't need to knowledge for hydrological process. As mentioned above, ANN model is most popular applied in many problems and especially water resource problem.

Therefore, in this study developed ANN model based on flood map information from the GISTDA agency and information of water level at gauge stations and floodgates from the Royal Irrigation Department to predict size of flood area in advance at floodplain area and to be useful for providing information or use together with damage assessment or the tendency of flood areas.

1.2 Objective

To develop model for the flood area size prediction at floodplain area by using Artificial Neural Network (ANN) method.

1.3 Scope of study

1) In this study selected the Basin of Tha Chin River in Suphan Buri province.

2) Use the daily water level data in flood event of each gauge stations and floodgates that located in the Tha Chin River and Irrigation canals that is representing the flood area.

3) Use shape files of flood event from the GISTDA agency that used to determine the size of the flood areas in the study area.

4) For flood area size prediction, using the water level 1-2 previous days in each gauge stations for in advance prediction where flood area size obtained from shape files for developing the model.

CHARPTER II
LITERATURE REVIEW

2.1 Study area

Figure 2.1 Map of Suphanburi province.

2.1.1 Location

Suphanburi province is one of the west central regions of the country**,** which located on floodplain of Tha Chin River or another call Suphanburi River. The Tha Chin River flows along the length of the province from north to south. Suphanburi province is located between latitude 14^{o} 4'-15^{o} 5'N and between longitude 99^{o} 17' - 100^{o} 16'E, height of 3-10 meters above mean sea level and total areas of 5358.01 square kilometers, representing 5.2 percent of the central region.

2.1.2 Topography

Suphanburi province has mainly plain areas. There are some areas of the plateau with a slope of 0-3 percent, which is located along the west side of the province from north to south. The lowest area is in the southeast that located above mean sea level about of 3 meters and 10 meters for the northern of province. The most of areas is a rice field, canal, marsh everywhere. The major river is Tha Chin river or Suphanburi river that flow through from north to south of province.

2.1.3 Climate

Generally, Suphanburi province has similarity of climate to other provinces in the central region. The summer season has been influenced by the monsoon from the southeast of the South China Sea during February to mid-May. The rainy season has been influenced by the monsoon from the southwest of the Indian Ocean during May to mid-October. The winter season has been influenced by the northeast monsoon from the October to mid-February.

2.1.4 Water resource

Suphanburi province has consisted various river. Tha Chin river or Suphanburi river is a large river that the most important and beneficial to the living of the population and economy. Kaseaw Dam is a major branch of Tha Chin River and the others are small rivers, which is mainly flow to Tha Chin River.

2.2 Definition and Flooding characteristic

Flooding means the water flow in river with water level is higher than normal level and overflow the banks, as result from runoff caused by rainfall that exceed the flow in the river. This may cause damage or not, and needn’t occur every year.

The meaning of flooding is defined by Meteorological Department [1] that is the water conditions overflow the banks and occurrence of flood on areas or the drainage of flood couldn’t drain out and the areas are covered by water. The divisions of flooding characteristics are two types.

1. Drainage floods are the occurred flooding condition due to drainage system is not efficient and often occurs at floodplain area and large urban area. The most of drainage floods occurred downstream and spread over wide areas.

2. Flash floods are flooding condition that occurred suddenly and appeared on more slope areas and less catchment. Flash floods are usually occurs after heavy rain, within 6 hours.

The cause of flooding that occurred in Thailand is divided into two types. The first type is flooding by natural such as rainfall intensity, flash flood from hills, water overflow the bank of river, the flooding of estuary occurred due to high tide. The another type is flooding by man-made such as an inappropriate land use, the deforestation on the hill, the intrusion of water, the settlements in the flood risk areas and the construction of roads that obstruct waterway and the policy of reservoir management that is mistake. [2]

2.3 Geography Information System (GIS)

GIS is the process of spatial data with the computer system by determination the attribute data and information. Besides, it is able to manage data from various sources that including satellite and aerial imagery by storage, classification, analysis and display of spatial data to the objectives of geographical application. Particularly, spatial data have format and relationship of the all spatial data and it can be analyzed by GIS to provide a meaningful change in relation to the period. [3], [4]

Goodchild *et al* (1996) said that the GIS is a powerful tool for data collection, data analysis and display. At present, the GIS are applied to flooding management such as displaying the flooding and damage areas that occurred and provide information to the hydrological model. [5]

This study applied GIS as tool to analyze the size of flood area for preparing the data to the model development.

2.4 Artificial neural network

Artificial neural networks (ANNs) were used to more the water resource prediction. ANNs are considered in a group of black box model which is not attention on hydraulics processes, it can also accurately predict. Unlike other model groups that based on information and relationship of the amount of large data that caused the black box model has been used to predict flooding increasingly. [6]

In general, the advantages of ANN application are the ability to solve problems that not require a clear definition of the mathematical problems. On the other hand, ANN has solved by compiling the sample set of problems and answers. The network tries to learn the nonlinear relationships of input and output that leads to an accurate answer. [6]

2.4.1 Back propagation Neural Network (BPNN) [7]

The BPNN has a very popular structure that is the Multilayer Perceptrons structure which is model consist of three layers such as input layer, hidden layer and output layer as shown in Figure 2.2. The process of learning is used by data feature of available input and output. The input data is fed into the unit of the first layer then each input is multiplied by the weights, only link between the layers. The default setting of the weights may be assigned by random. The product will be used together to transform a value by activation function, and then the results will be compared with actual data, lead to the weight adaptation and minimize errors for each iteration. The weight adaptation will continue until the error is less and acceptable and then the process is terminated, which the characteristic of the Back Propagation (BP) as shown in Figure 2.3.

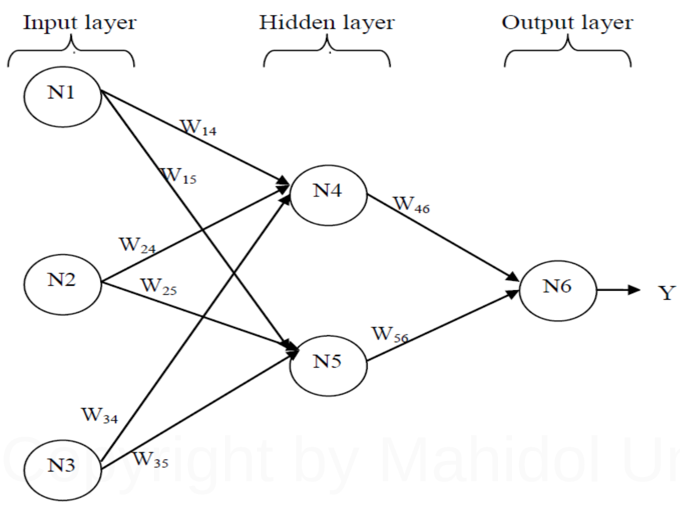

Figure 2.2 Structure of Multilayer Preceptron

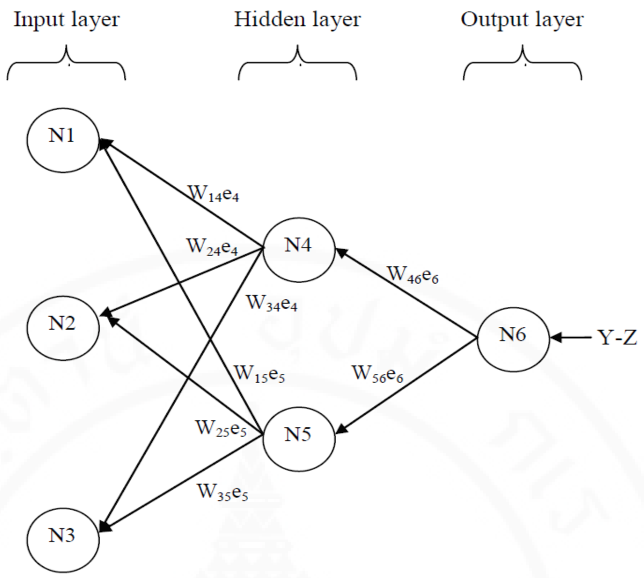

Figure 2.3 The characteristic of Back Propagation

2.4.2 The model development and BPNN procedures [7]

In model development, the dataset is learned and tested that through process of learning with Back propagation and testing. The process chart of model development is shown in Figure 2.4.

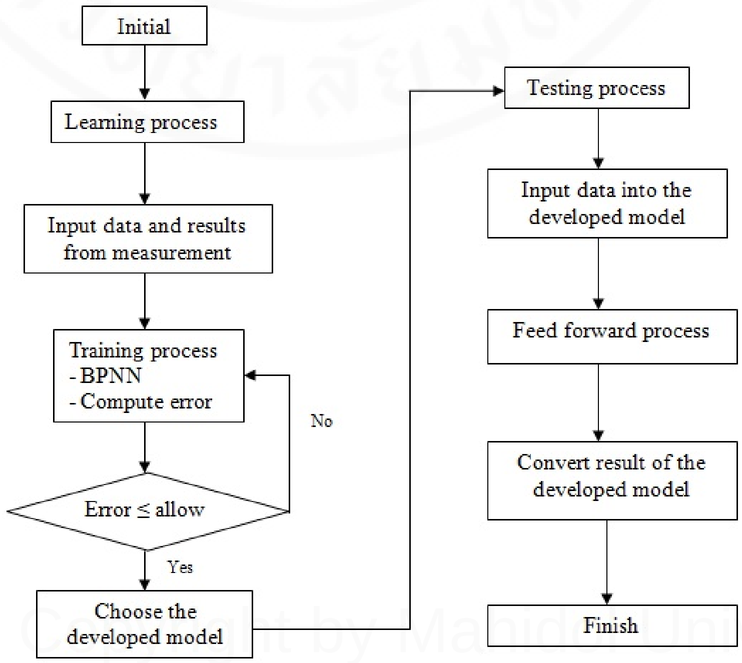

Figure 2.4 The process chart of model development

2.4.2.1 The learning procedure of ANN [7]

1. The assignment consists of ANN structure, initial conditions of the weights and bias, besides included assignment of parameters and activation function.

2. The learning process had calculated weights and bias that the results of calculation similar to the results of actual measurement. The process of the BP can be divided into two parts such as a forward step (forward pass) and a backward step (backward pass), which is shown the chart of back propagation steps as figure 2.5.

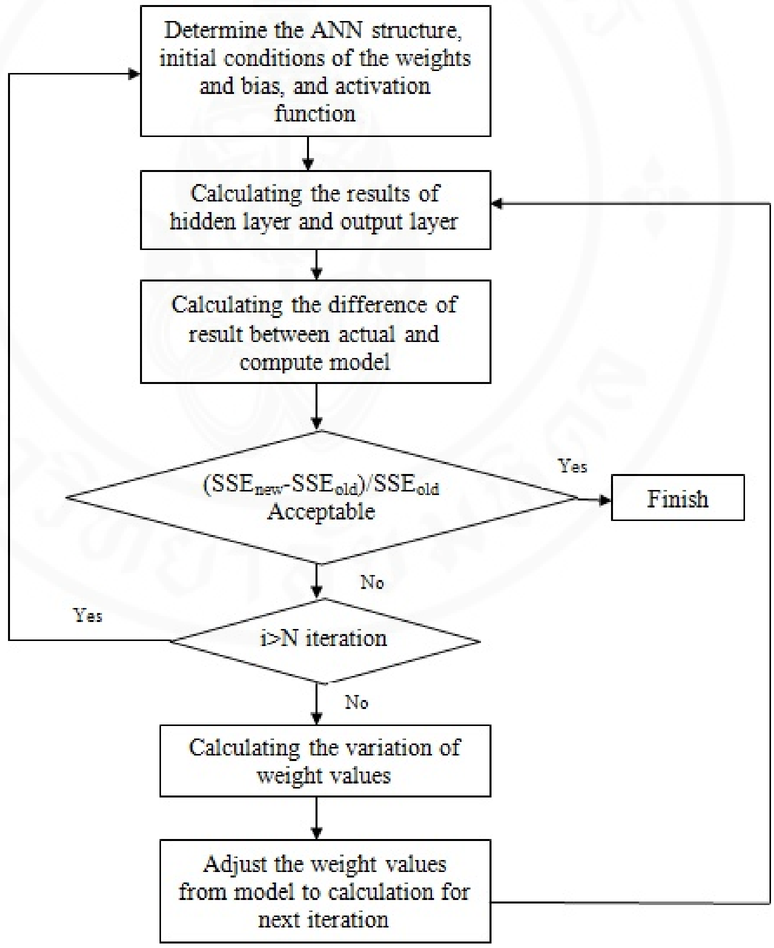

Figure 2.5 The chart of back propagation steps

However, this process may cause more learning or overfitting that is drawback of ANN model. Because the network tries to learn the relationship of input and output data, which is the recognitions of some data that may include noise data. Therefore, the network will provide good results for training process but the results is not good for testing. The avoidance method of overtraining is cross training, which it making the network learn more efficiency. Thus, the learning process needs to calculate for weight and bias values properly to learning that making the errors are decrease trend and stable, which it will make the appropriate iteration for stopping the learning. If calculation in the next more iteration, although errors are fixed or decrease, the model may be learning too much that given not good results of test.

This study applied ANN for model development by using the hydrological data such as water level data and flood data that obtain from flood shape files. The developed model can predict flood area size where the water level data of each gauge station as inputs and the flood data were analyzed through ArcGIS to obtain flood size area as output for model. The developed model can be applied the analysis of flood trend or assessment of damage.

2.5 The ANN application for flood prediction

In this section showed the examples of previous researches that applied ANN for hydrological model development.

Mandal et al. (2005) had used ANN model for flood management, where prediction of flood plays a crucial role. The ANN model uses five variables such as temperature, humidity, water level, precipitation, and wind speed as input. The relative importance of these input variables in the prediction of occurrence and non-occurrence of flood by ANN model indices that water level is the key variable related to flood while temperature is the least important variable among those five variables. From the results, it is clear that the proposed model is statistically significant. The overall prediction accuracy is 97.33%. [8]

Adnan et al. (2012) had proposed ANN model to predict the flood water level. A supervised neural network with the back propagation model is adopted in this study. The water level is only the variable in this Back Propagation Neural Network model. The performances of BPNN model have been evaluated but yet the result is not that satisfactory. Hence, Extend Kalman Filter is proposed to improve the results. In conclusion, Extend Kalman Filter is able to provide better result compared with actual result without application of Extend Kalman Filter. The results are improved as the overshoot is reduced and pattern of the predicted and actual water level between 0-6 hours is nearly. [10]

Campolo et al. (1999) had addressed the problem of flood forecast using ANN on the basis of rainfall and water level data. They attempted to set up a model which would be able to forecast the water level of the river on the basis of rainfall information and previous water level with a sufficient lead time in order to reduce the consequences of floods. Predictions of the model for 1-hour forecasting are accurate, with a minimal error of the order 4%. When trying to use the model to forecast water level with a longer time horizon, performance decreased rapidly. [9]

Leahy et al. (2008) had studied water level at the flood location and the two available upstream monitoring stations were found to be the most useful predictor of levels 5h ahead at the flood point. The other input vectors such as the recent change in river level at the flood location, sea level pressure or the change in sea level pressure did not improve prediction accuracy to the point where the extra network complexity added by their introduction was justified by an associated improvement in prediction performance. Optimization of weights alone increases the ANN performance to a value close to that of the fixed architecture but with performance penalty relative to a fixed architecture trained using conventional back propagation only. In this study, they had demonstrated that a global optimization methodology for ANN architecture and weights can be employed successfully to a river level prediction problem. [14]

Cruz et al. (2011) had developed ANN model of a 40-year time series of river water-level observations was used to reconstruct floodplain water levels in a study area of the lower Cuiaba River floodplain of the northern Pantanal. The ANN yielded a prediction error not greater than 9 cm in 95% of occurrences and it can contribute to the management of hydroelectric reservoirs upstream of the floodplain to minimize their adverse impacts on inundation regimes. [11]

Sulaiman et al. (2011) had developed an ANN based model with Zoning Matching Approach (ZMA) for real-time flood forecasting at Rantau Panjang station, specifically to monitor high water level events. The forecasting models developed using ZMA are able to forecast water levels above 8000mm with high accuracy with a lead-time of up to 3 h, and with satisfactory performance at 6 h. [12]

Bustami et al. (2006) had developed an ANN to predict daily flood based only on antecedent rainfall and water level data, which a part of flood management. Sungai Bedup station was selected. The sensitivity of the network performance to the content and length of the calibration data was using various training data set. It was found that to obtain satisfactory results, the neural network using 4 days of antecedent data performed very well in simulating the daily water level. [13]

Shongsak (2003) had conducted a daily flood forecasting in Lam Phachi basin with MIKE 11 and ANN models. For ANN, the data of input layer consist of the rainfall gauge station (K.25A), the K.17 of the malaria tropical of research (Ratchaburi) and the K.59 of runoff gauge station (*kanchanaburi*). The MIKE 11 model includes the Unit Hydrograph and Hydrodynamic models, which require the information of soil type, land use, physical conditions of the area, and characteristics of river, runoff and rainfall. While the neural network model used runoff and rainfall for runoff prediction with Feed Forward Back Propagation method and using 5-5-1 structure. The results of comparing the suitable model with EI showed that the performance of MIKE 11 model is equal to 64.4 percents for learning and 18.0 percents for testing. For ANN, The performance is equal to 99.8 percents for learning and 59.0 percents for testing. In this study, it had concluded that the of neural network model are suitable to be used to forecasting the flooding of the Lam Phachi river more than MIKE 11 model. [6]

The most of these studies had predicted the water level at gauge station of basin that it can be predicted in advance or warning to the application of flood management. In addition, the application of ANN model to forecasting the results that is satisfactory for previous researchers, especially when compared to other hydrological models. From these studies, the most of variable for flood prediction that used is water level data, so this study applied the water level data for flood area prediction.

2.6 The research related flood extent areas

In addition, the related researches had analyzed the flooding extent areas by hydrological models integrated with and Geographic Information System (GIS) to apply the assessment of risk areas or warning. The researches had shown examples as follows.

Rangsimapirat (2005) had developed a mathematical model to predict the expected flooding areas of the lower Mae Klong river basin. The model calculates flood elevation and constructed the Digital Elevation Model (DEM) from aerial photographs to create a flood map by ArcView and compare to the LANDSAT5 satellite photos to evaluate the error of model. [15]

Sansena (2004) had studied runoff of The Mae Klong River and develop GIS for supporting a Hydraulic model to simulate and create flood map based on the hydrological and hydraulic approach. The process involved runoff frequency analysis by 6 probability distributions and statistics to determine a suitable probability for designing runoff frequency and developing GIS data for generated Digital Terrain Modeling (DTM) and integration of Hydrologic Engineering Centers River Analysis System (HEC-RAS) with DTM to develop a regional model for floodplain determination and a model for simulation of the flood prevention, warning and forecasting. [16]

Mangsilp (2007) had studied evaluation of flood area in Mae Taeng Watershed, Chiang Mai province using the GIS integrated with potential surface analysis and hydrological model. In addition, the appropriate model compared with real flooding that obtained from satellite photos. [17]

Arayawongwarn (2001) had studied the flood phenomena in the Yom basin, Phare province by using MIKE11 and the Geographic Information System (GIS). MIKE11 is used for forecasting flooding at three periods of 10 years, 25 years and 100 years. The GIS is advantageous for displaying spatial data and it is applied for estimating flood areas and generating flood maps. The map are analyzed considering physical factors and land use change, which are also considered in this research, in order to suggest guidelines for management of future flood areas in the Yom basin, Phare province. [18]

Wongwitaya (2001) had developed and modified Arcview GIS, which is a geographic information system (GIS) to make flood map in the Yom Basin but in this study, the focus was on Phare and Sukhothai Province. In this investigation, the using some functions in Arcview along with the developed program. The program has the capability to develop the flood area maps by integrating maximum flood water level data from MIKE11 along with geological information data. The result of research is the application, which can make GIS flood map boundary similar to that of a manual flood map in shape and area, users can observe the preliminary spread out of flooding area. However the accuracy of the result depends on the scale of the contour line input. [19]

However, researches have been mentioned above. The most of study areas is a valley, which can't be applied to the floodplain area because of the contour line are inadequate resolution and causing mistake for display flood extent areas and unreliable. Therefore, this study attempts to study the topography of floodplain area and predict sizes of the flooding area with taking the advantage of ANN technique for model development, to be useful to provide information or use together with damage assessment or the tendency of flood areas.

CHARPTER III
RESEARCH METHODOLOGY

This study only used water level for flood area size prediction due to the case study area was floodplain area. The characteristic of flooding had expanded to wide area when water level overflow bank that unnecessary to using the other variables such as rainfall data, which it related water level directly. Therefore, this study used daily water level of each gauge station and floodgate to find relation of flood area size for prediction.

The concept of model development is based on ANN technique, and the water level was measured at gauge stations and floodgates of the Tha Chin River and Irrigation canals in Suphan Buri province together with the flood event data that obtained from flood shape files. Flood event data were processed by GIS to obtain the flood area sizes for prediction that showed in appendix to find flood area size. The methodology are shown in Figure 3.1

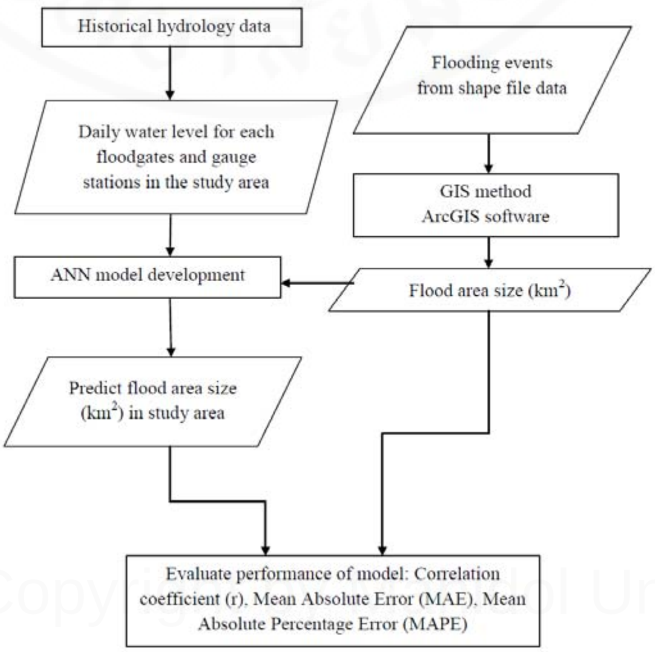

Figure 3.1 Flow Chart of Research Methodology

3.1 Data collection

Hydrological information used water level that obtain from the gauge stations and floodgates of the Royal Irrigation Department that located in the Tha Chin River and Irrigation canals of Suphan Buri province. In addition, the shape files of flood area obtain from Geo-Informatics and Space Technology Development Agency (GISTDA) to be used in the analysis of flood area size that is processed by ArcGIS software.

3.1.1 Water level data

Daily water level data obtained from gauge stations since in 2002-2012. In addition, the water level data obtained from the floodgates since the year 2002-2012 that used in this study and shown list of gauge stations and floodgates as Table3.1. The location of gauge stations and floodgates are shown in Figure 3.2

Table 3.1 List of gauge stations and floodgates

Type of station	**ID station**
Gauge station	T10,T13
floodgate	SC01,SC02,SC03,SC04,SC05,PP01,PP02,PP03,PP04,PP05,PP06, PP07,PP08,PP09,PP10,PP11,PP12,PP13,PP14,PP15,PP16,PP17, PP18,PP19,DJ20,DJ23

3.1.2 Flood shape files

Flood shape files obtained since in September-December of 2006-2011 whereas the information occurred for some days. Because the information obtained from satellite images that may be not record images every time.

3.1.3 Dataset

In creating the dataset for the model development, the flood area sizes of present day were compared with the 1 and 2 previous days of water level to create 1 day in advance dataset and 2 day in advance dataset of prediction. The water level data of each station and floodgate were inputs of model where as flood area size was the output of model.

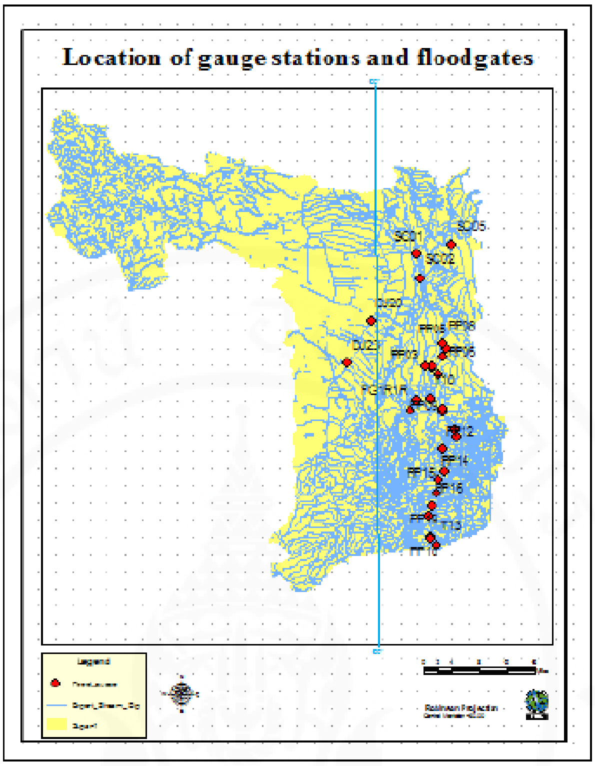

Figure 3.2 The location of gauge stations and floodgates

3.2 Model development

In this study, MATLAB 2010a software is used as a tool in the development of the artificial neural network models. The model is developed by using the water level dataset that has been divided by K-fold cross validation to find the most of appropriate model for dataset. This method is resampling with divide into many groups of dataset and partially tested. The best result of cross validation will be chosen for the model.

3.2.1 The network structure

In this study, the structure of the Multi layer perceptron is used, which is the most popular structure are shown in Figure 3.3.

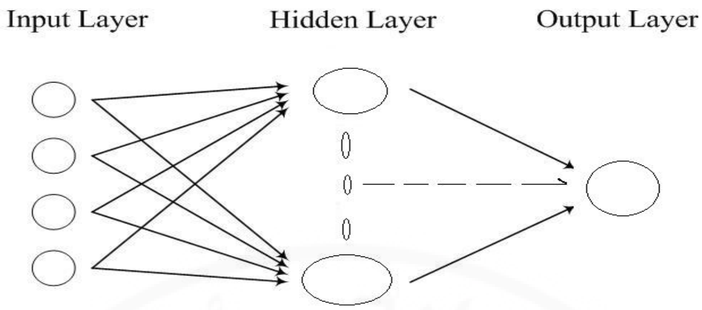

Figure 3.3 The network structure

From figure 3.3, structure of the network consists of 3 layers included input layer, hidden layer and output layer where each layer consists of several neurons and each neuron is connected to other neuron of the layer next. The first layer is input layer, which input data as water level data of each station. The hidden layer determined only one layer. Number of neurons in hidden layer obtains from experiment by changing the number of neurons in the hidden layer from 1…N neurons to obtain the optimal network structure and then selects the number of hidden neurons of ANN model that provides the best statistical. Output layer has only one neuron is a flood area size.

3.2.2 The activation function

Typically, each neuron has activation function that transform input to output of function. The sigmoid function is widely used such as logistic and hyperbolic tangent functions. However, this study used hyperbolic tangent sigmoid functions in hidden layer and linear function in output layer for model development that follow predicting the regression pattern. [21]

3.2.3 Training Function

In the ANNs model development, it often encounter overfitting problem that affect to the developed model can't available (generalize). An alternative method for improving generalization is regularization. Regularization can be done automatically by using the Bayesian regularization training function. The advantage of this function is not require a validation data set that be separated from the training data set and provides better generalization performance than early stopping when the data set is small and be suitable for training function approximation networks. [21]

In addition, the research of Jeong (2004) compared the effectiveness of the training and learning functions. The result showed that the Bayesian Regularization training function outperformed the other 12 training functions while the result of experiment for 14 learning functions showed similar performance. [20]. Therefore, in this study chose the Bayesian Regularization training function.

3.3 Evaluation of model performance

After the developed model has been completed, the test dataset will be used to test. The comparison of results that obtained from calculated model and measured data by using statistical methods, including Mean Absolute Error (MAE), Correlation Coefficient (r) and Mean Absolute Percentage Error (MAPE).

3.3.1 Mean Absolute Error (MAE)

MAE is a statistical method that shows the absolute error between the calculated model and actual of flood area size that should be close to zero. The equation that shown the calculation following.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \qquad (3\text{-}1)$$

Where Y_i = actual value

$\hat{Y}_i$ = predicted value

n = number of data

3.3.2 Correlation Coefficient (r)

The Correlation coefficient represents the relationship between the data obtained from the model and measurement that is linear. If the value of one indicates that the relationship is linear at best (perfect positive linear association) and a zero that has no relationship (no linear relation) by the equation shown as follow.

$$r\,(Y\,,\hat{Y}\,) = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y}_o)(\hat{Y}_i - \overline{Y}_p)}{\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y}_o)^2 \cdot \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y}_p)^2}} \quad (3\text{-}2)$$

Where Y_i = actual value

$\hat{Y}_i$ = predicted value

$\overline{Y}_o$ = mean of actual value

$\overline{Y}_p$ = mean of predicted value

n = number of data

3.3.3 Mean Absolute Percentage Error (MAPE)

MAPE is a statistical method that shows the percent of absolute error between the calculated model and actual of flood area size that should be close to zero. The equation that shown the calculation following.

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|Y_i - \hat{Y}_i\right|}{Y_i} \times 100\,\% \quad (3\text{-}3)$$

Where Y_i = actual value

$\hat{Y}_i$ = predicted value

n = number of data

3.4 Material and development tools

Hardware: 1) Notebook Intel(R) Core2(TM) Duo 2.53GHz RAM 4.00GB

Software: 1) Microsoft Windows7

2) ArcGIS 9.3

3) Matlab R2010a

4) SPSS 16.0

3.5 Schedule/ Time plan

In this section, step of work and time periods that shown in Table3.2 as below.

Table 3.2 Grant chart: Steps of work

Steps of work	2012										2013	
	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
1. Study area of Supanburi province and collect the secondry data of water level and flood area from the Royal Irrigation Department and GISTDA.												
2. Analyze and select water level data of each station together with the flood area to create a dataset.												
3. Study and analyze for model development to predict size of flood area.												
4. Study and evaluate for model performance.												
5. Discussion and conclusion of research.												

CHAPTER IV
RESULTS AND DISCUSSION

The result of flood area size prediction in Suphanburi province: a comparative study of dataset (1 and 2 days in advance) with the neural network technique to find the proper dataset. Moreover, this study had the comparison of 2 models between neural network and linear regression and discusses the result for prediction.

In construct dataset, the number of samples is 104 samples. The inputs of model consist of 31 inputs of daily water level data from floodgates and gauge stations. The output is flood area size from daily shape file data of 104 events whereas flood events are discontinuous data but daily water level data is continuous data. Therefore the dataset were constructed by the reference of daily flood event to create dataset for 1-2 days in advance prediction.

In the development of model, the dataset were divided as 10-fold cross validation. The Bayesian regularization training function is used, and changing the number of neurons in the hidden layer from 1…N neurons, number of training epochs and activation functions to obtain the optimal network structure. The result of network structure is chosen by the best statistical performance.

4.1 The results of model development

4.1.1 The results of 1 day in advance

In this section, the results of 1 day in advance dataset for model development are showed in Table 4.1.

Table 4.1 The results of 1 day in advance

Structure	MAE	Correlation coefficient(r)	MAPE (%)
31-1-1	208.55	0.88	55.96
31-2-1	187.57	0.90	44.49
31-3-1	179.13	0.89	35.89
31-4-1	160.47	0.90	36.67
31-5-1	192.78	0.87	39.83
31-6-1	164.02	0.90	36.33
31-7-1	**155.99**	**0.90**	**34.88**
31-8-1	179.21	0.88	42.87
31-9-1	166.37	0.91	36.73
31-10-1	179.92	0.86	43.56
31-11-1	168.28	0.89	38.07
31-12-1	167.44	0.90	38.95
31-13-1	189.32	0.86	41.52

Table 4.1 shows the results of 1 day in advance. In experiment, the optimal of structure is 31 inputs, 7 hidden neurons and 1 output. The results of evaluating 3 statistical performances are 155.99 (MAE), 0.90 (Correlation coefficient) and 34.88 (MAPE) respectively. In addition, the optimal activation function of this structure is hyperbolic tangent sigmoid function (tansig) and linear function (purelin) for hidden layer and output layer respectively.

4.1.2 The results of 2 days in advance

In this section, the results of 2 days in advance dataset for model development are showed in Table 4.2.

Table 4.2 The results of 2 days in advance

Structure	MAE	Correlation coefficient(r)	MAPE (%)
31-16-1	166.37	0.91	40.64
31-17-1	163.07	0.91	34.33
31-18-1	159.23	0.91	31.02
31-19-1	156.09	0.92	36.62
31-20-1	**151.74**	**0.92**	**32.46**
31-21-1	162.01	0.91	37.13
31-22-1	186.43	0.89	43.13
31-23-1	208.14	0.87	51.06
31-24-1	165.30	0.91	36.18
31-25-1	175.04	0.91	42.05

Table 4.2 shows the results of 2 days in advance. In experiment, the optimal of structure is 31 inputs, 20 hidden neurons and 1 output. The results of evaluating 3 statistical performances are 151.74 (MAE), 0.92 (Correlation coefficient) and 32.46 (MAPE) respectively.

4.2 The results of model development with the input selection

From previous section, the results of 1 and 2 days in advance for model development that given high error. In this section, we test 1 and 2 days in advance dataset with the input selection. The correlation coefficient is used for the input selection to find the relation between water level data of each gauge station and flood event. In this study, the correlation criterion was selected 0.85 up. The 10 inputs of each gauge stations are chosen for 1-2 days in advance dataset such as T13, PP10, PP11, PP12, PP14, PP15, PP16, PP17, PP18 and PP19.

4.2.1 The results of 1 day in advance (selected inputs)

In this section, the results of 1 day in advance dataset that selected inputs for model development are showed in Table 4.3.

Table 4.3 The results of 1 day in advance (selected inputs)

Structure	MAE	Correlation coefficient(r)	MAPE (%)
10-6-1	165.41	0.91	33.25
10-7-1	162.30	0.91	33.35
10-8-1	156.68	0.90	29.09
10-9-1	163.53	0.89	34.73
10-10-1	161.71	0.89	30.58
10-11-1	154.36	0.91	30.06
10-12-1	**139.55**	**0.92**	**23.13**
10-13-1	149.43	0.91	27.45
10-14-1	170.68	0.88	33.13
10-15-1	150.87	0.90	26.96
10-16-1	181.53	0.87	43.55
10-17-1	164.46	0.89	33.94
10-18-1	171.59	0.88	37.19

Table 4.3 shows the results of 1 day in advance (selected inputs). In experiment, the optimal of structure is 10 inputs, 12 hidden neurons and 1 output. The results of evaluating 3 statistical performances are 139.55 (MAE), 0.92 (Correlation coefficient) and 23.13 (MAPE) respectively.

In addition, the error results of 1 day in advance dataset (selected inputs) were compared with the error results of 1 day in advance dataset to investigate the difference of 2 dataset. The paired T-test method was used for this study. The results are showed in Table 4.4 and 4.5. Table 4.4 showed the preliminary statistical results of 2 dataset, which the statistical results given the difference of each dataset. Table 4.5 showed the results of paired samples test between unselected and selected input.

Table 4.4 The results of paired samples statistics of 1 day in advance dataset

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	unselected	155.99	104	183.83	18.03
	selected	139.55	104	173.55	17.02

Table 4.5 The results of paired samples test of 1 day in advance dataset

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	unselected - selected	16.45	184.75	18.12	-19.48	52.38	0.91	103	0.37

From this study found that the results of paired samples test of 1 day in advance dataset between unselected and selected input was not statistically significant of difference.

4.2.2 The results of 2 days in advance (selected inputs)

In this section, the results of 2 days in advance dataset that selected inputs for model development are showed in Table 4.6.

Table 4.6 The results of 2 days in advance (selected inputs)

Structure	MAE	Correlation coefficient(r)	MAPE (%)
10-28-1	164.97	0.91	33.31
10-29-1	166.65	0.90	33.02
10-30-1	162.88	0.90	34.89
10-31-1	154.11	0.92	31.39
10-32-1	**147.63**	**0.92**	**31.26**
10-33-1	155.96	0.91	34.68
10-34-1	160.29	0.90	36.95
10-35-1	168.85	0.90	35.96
10-36-1	165.77	0.90	33.69

Table 4.6 shows the results of 2 days in advance (selected inputs). In experiment, the optimal of structure is 10 inputs, 32 hidden neurons and 1 output. The results of evaluating 3 statistical performances are 147.63 (MAE), 0.92 (Correlation coefficient) and 31.26 (MAPE) respectively.

In addition, the error results of 2 day in advance dataset (selected inputs) were compared with the error results of 2 day in advance dataset to investigate the difference of 2 dataset. The paired T-test method was used for this study. The results are showed in Table 4.7 and 4.8 respectively. Table 4.7 showed the preliminary statistical results of 2 dataset, which the statistical results given the difference of each dataset. Table 4.8 showed the results of paired samples test between unselected and selected input.

Table 4.7 The results of paired samples statistics of 2 day in advance dataset

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	unselected	151.74	104	158.94	15.59
	selected	147.63	104	161.09	15.80

Table 4.8 The results of paired samples test of 2 day in advance dataset

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	unselected - selected	4.10	135.31	13.27	-22.21	30.42	0.31	103	0.76

From this study found that the results of paired samples test of 2 day in advance dataset between unselected and selected input was not statistically significant of difference.

4.3 The comparison of model performance

This section compared the model performance between neural network and conventional model to examine the performance of neural network that better than the conventional model or not. In this experimental used linear regression model as the conventional model with 1 day in advance dataset (selected input) and compare with the neural network. The results are showed in Table 4.9.

Table 4.9 The comparison between neural network and linear regression model

Model	**MAE**	**Correlation coefficient(r)**	**MAPE (%)**
Linear regression	198.48	0.89	45.10
Neural Network	**139.55**	**0.92**	**23.13**

Table 4.9 shows the comparison between neural network and linear regression model which the result found that the neural network outperforms than the linear regression. Moreover, this study investigated the difference of 2 models error with the paired T-test method. The results are showed in Table 4.10 and 4.11 respectively.

Table 4.10 shows the preliminary statistical results of 2 models error, which the statistical results given the difference of each model. Table 4.11 showed the results of paired samples test between neural network and linear regression, which the results are difference of the paired samples test of significance. These tables are showed as follows.

Table 4.10 The results of paired samples statistics of model

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	NN	139.55	104	173.55	17.02
	LR	198.48	104	158.66	15.56

Table 4.11 The results of paired samples test of model

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	NN - LR	-58.93	185.73	18.21	-95.05	-22.81	-3.24	103	.002

4.4 The examples of predicted result

In this section, the examples of predicted result are showed after the model was developed. The samples of dataset are randomized and training the dataset with 90% and 10% of testing the dataset by using the 1 day in advance dataset (selected input) and the optimized structure of ANN. The performance results were 66.86, 0.99 and 9.83 of MAE, correlation coefficient and MAPE respectively after testing the 11 examples. The predicted results were showed in Table 4.12.

Table 4.12 The examples of predicted result

Number of examples	Input of gauge stations (msl.)										Predicted result(km^2)
	T13	PP10	PP11	PP12	PP14	PP15	PP16	PP17	PP18	PP19	
1	2.37	2.73	2.49	2.51	2.53	2.55	2.51	2.51	2.18	2.15	469.83
2	2.35	2.9	2.69	2.57	2.62	2.6	2.58	2.51	2.17	2.13	479.52
3	2.35	2.93	2.73	2.59	2.64	2.6	2.57	2.51	2.16	2.13	504.62
4	2.19	2.88	2.75	2.42	2.52	2.42	2.42	2.39	2	1.96	318.55
5	2.6	3.21	3.06	2.88	2.98	2.86	2.84	2.82	2.39	2.35	692.42
6	2.57	2.94	2.83	2.56	2.69	2.75	2.72	2.8	2.33	2.3	536.27
7	3.08	3.57	3.6	3.33	3.41	3.34	3.33	3.3	2.87	2.85	630.63
8	2.85	3.19	3.2	3.11	3.1	3.03	3.04	3.09	2.61	2.59	738.37
9	3.8	4.45	4.52	3.79	3.77	3.78	3.78	3.73	3.57	3.69	1763.14
10	3.89	4.42	4.5	3.84	3.83	3.85	3.86	3.83	3.67	3.79	1880.64
11	3.82	4.23	4.28	3.85	3.85	3.87	3.89	3.84	3.65	3.73	1555.68

From the predicted results in Table 4.12, the flood area of each example are showed to clearly indicate the results whereas the show of flood area chose the flood area size that similar to the predicted result manually. The flood areas obtain from shape file that gathered from the GISTDA agency. The flood areas are showed to following the predicted result of each example.

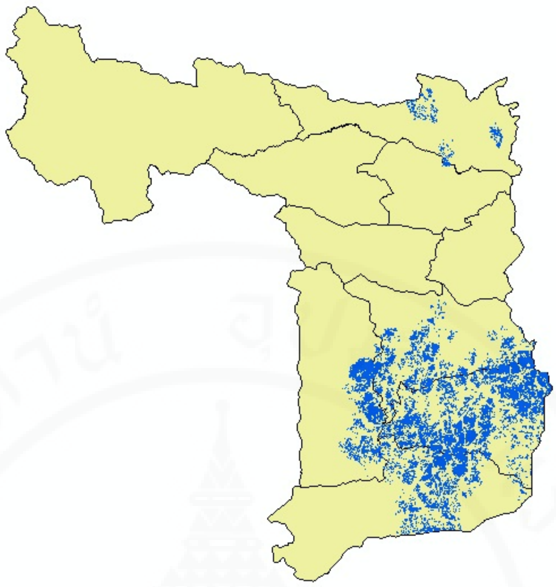

Figure 4.1 Flood area size 475.02 km^2 for the first example

From the figure 4.1, the flood area size was showed that more than the result of prediction 1.1%.

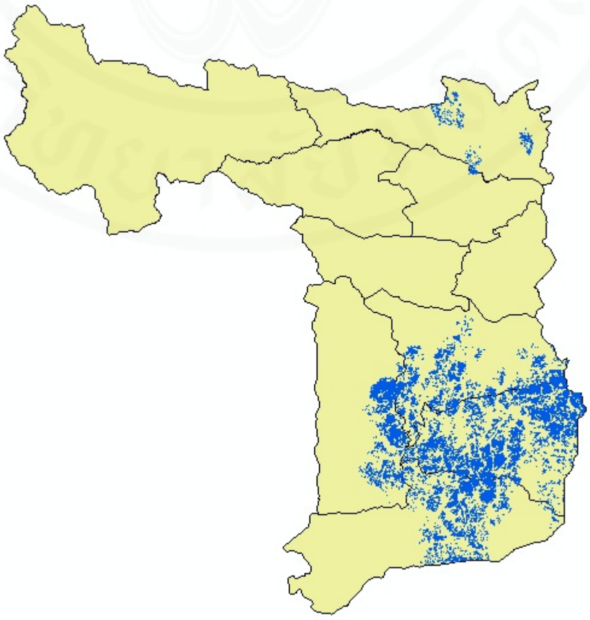

Figure 4.2 Flood area size 475.02 km^2 for the second example

From the figure 4.2, the flood area size was showed that less than the result of prediction 0.94%.

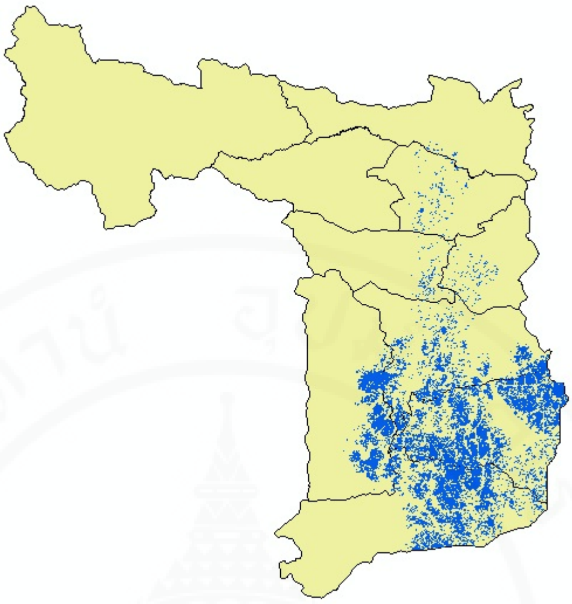

Figure 4.3 Flood area size 513.88 km^2 for the third example

From the figure 4.3, the flood area size was showed that more than the result of prediction 1.84%.

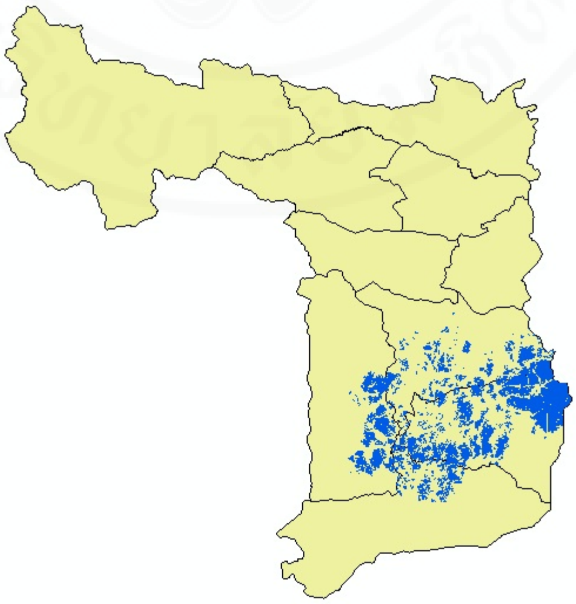

Figure 4.4 Flood area size 341.65 km^2 for the fourth example

From the figure 4.4, the flood area size was showed that more than the result of prediction 7.25%.

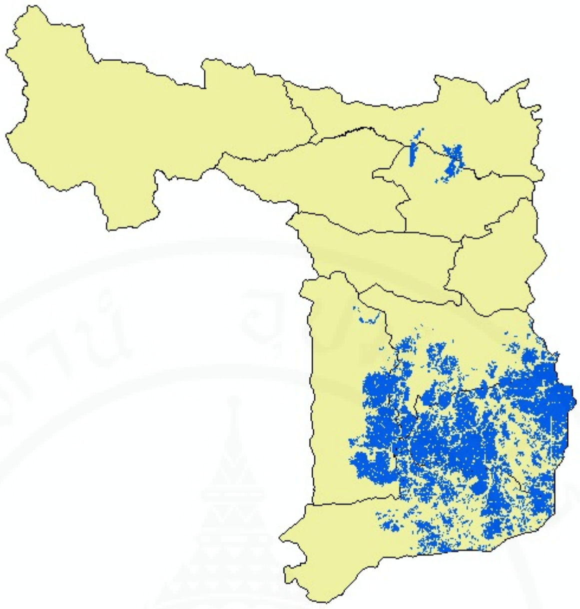

Figure 4.5 Flood area size 702.04 km^2 for the fifth example

From the figure 4.5, the flood area size was showed that more than the result of prediction 1.39%.

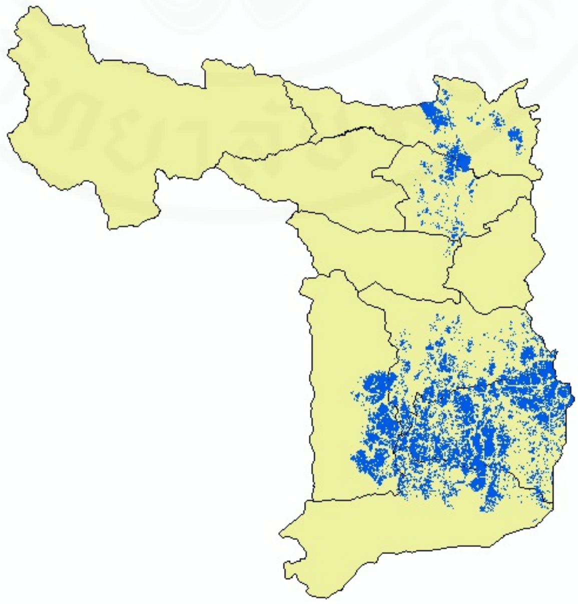

Figure 4.6 Flood area size 527.06 km^2 for the sixth example

From the figure 4.6, the flood area size was showed that less than the result of prediction 1.72%.

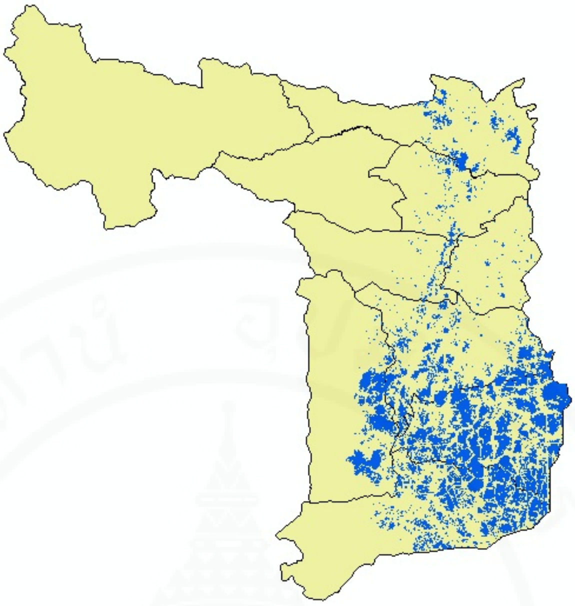

Figure 4.7 Flood area size 634.93 km^2 for the seventh example

From the figure 4.7, the flood area size was showed that more than the result of prediction 0.68%.

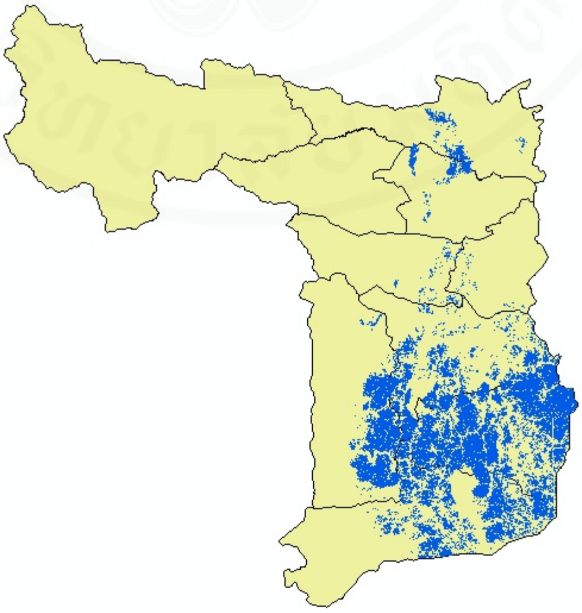

Figure 4.8 Flood area size 737.03 km^2 for the eighth example

From the figure 4.8, the flood area size was showed that less than the result of prediction 0.18%.

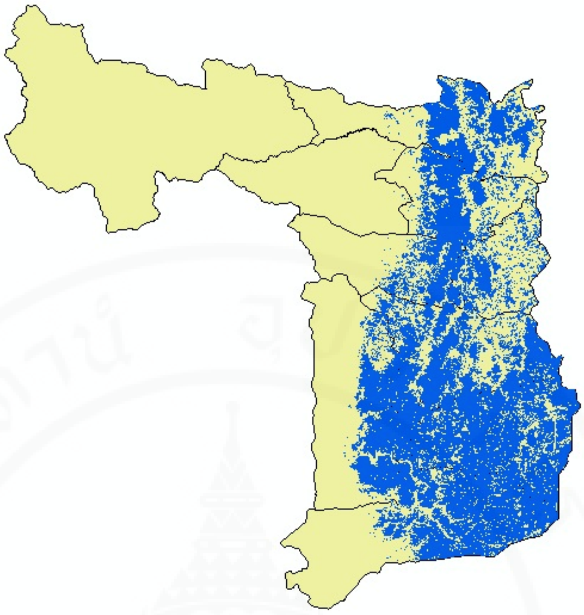

Figure 4.9 Flood area size 1758.17 km^2 for the ninth example

From the figure 4.9, the flood area size was showed that less than the result of prediction 0.28%.

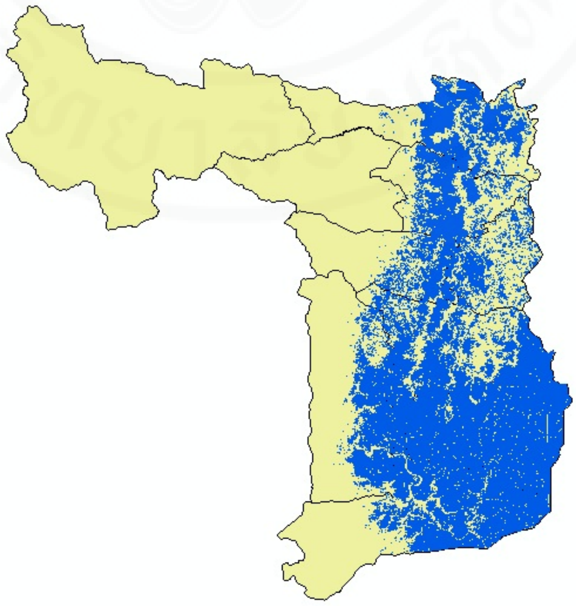

Figure 4.10 Flood area size 1904.96 km^2 for the tenth example

From the figure 4.10, the flood area size was showed that more than the result of prediction 1.29%.

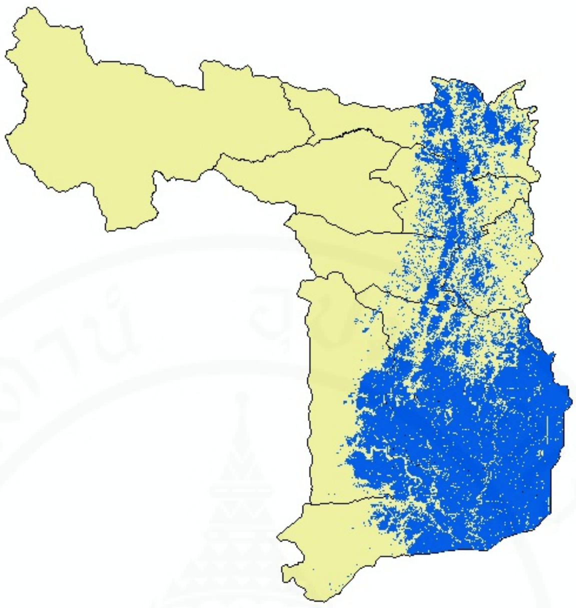

Figure 4.11 Flood area size 1577.71 km^2 for the eleventh example

From the figure 4.11, the flood area size was showed that more than the result of prediction 1.42%.

4.5 The discussion of model development

From the experiment of all dataset, it found that the 1 day in advance with the selected input outperforms than the other dataset that summarize as follow in Table 4.13.

Table 4.13 The summary of proper dataset for prediction

Dataset	MAE	Correlation coefficient(r)	MAPE (%)
1 day in advance	155.99	0.90	34.88
2 day in advance	151.74	0.92	32.46
1 day in advance (selected input)	**139.54**	**0.92**	**23.13**
2 day in advance (selected input)	147.63	0.92	31.26

Furthermore, it shows that the dataset of selected input is outperform when consider dataset between the selected input and the unselected input. The dataset of unselected input is highly possible the overfitting which the results are more error. Therefore, we need to delete the unnecessary inputs to find each input that extremely relates the actual result of dataset. In this way, the correlation coefficient is used for input selection to decrease the complexity of relation and reduce duration for training. Although, the result of error prediction between the selected input and unselected input was not statistically significant of difference both 1 day and 2 days in advance dataset. Therefore, the selected input dataset suited for the model development because using the necessary input for the model development and outperformance. Furthermore, the linear regression models are compared to verify that the neural network outperforms than the linear regression method that showed in Table 4.9. The results of error are examined by paired t-test method, which the results had significant difference and showed in Table 4.10 and 4.11 respectively. Due to the basis of neural network is the repeated calculation to adjusting the parameters for each variable but regression method is once calculate the parameters, so the neural network outperforms than the regression method. Although, this study used few samples due to the limited of flood area samples and the characteristic of flood area had high vary samples, which the results of model development had given the high average error but the correlation coefficient and MAPE values are satisfactory. In addition, the predicted results had given a better performance because the samples were randomized from the old dataset and the displays of flood area were chosen manually.

CHAPTER V
CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

This study has an objective to develop the neural network model for flood area size prediction of Suphanburi province. Due to the Suphanburi province is a large floodplain area and encountering the wide spread of flood. The previous related research had predicted by using water level and showing the extent of flood area with GIS method. However, this method can't show the extent of floodplain area due to the resolution of contour line is not enough. In addition, the previous study of neural network application of the hydrology research had given a good result of prediction. Therefore, the neural network was applied for flood area size prediction by using water level of each gauge station and flood gate that located in the Suphanburi province, which used as predictor, and using the shape file of flood area with analysis through GIS method to find the flood area size, which used as result of 1-2 days in advance prediction.

In the development of model, 10-fold cross validation was divided as dataset. The Bayesian regularization training function is used, changing the number of neurons in the hidden layer from 1…N neurons, number of training epochs and activation functions to obtain the optimal network structure. The result of network structure is chosen by the best statistical performance.

The results of model development, the 1 day in advance and 2 days in advance dataset were tested that the results given high average error, so the inputs of the 1 and 2 days in advance dataset were selected with correlation coefficient method to improve the result of model development. The result demonstrates that the selected input dataset outperform the unselected input dataset and it found that the 1 day in advance (selected input) outperformed other dataset. Moreover, this study had the comparison of 2 models between the neural network and regression model to examine the performance of neural network that better than the regression model or not. The result demonstrates that the neural network outperforms regression model and the models demonstrate the result of error that significant difference. Therefore, the development of neural network model had given satisfactory result. Although, this study used few samples due to the limited of flood area samples and the characteristic of flood area had high vary samples, which the results of model development had given the high average error but the correlation coefficient and MAPE values are satisfactory. In addition, the predicted results had given a good performance but the displays of flood area were chosen manually.

5.2 Recommendation and Future Work

The following recommendations were recommendations for the improvements of prediction are below:

1. In model development, it should receive more samples of data to improve the efficiency of prediction.

2. It should study the use of parameters adjustment, which they can be advantage too, for example, the parameters adjustment of the mu value, which they are in the Bayesian regularization algorithm function for training stop condition.

3. It should study more the efficient model such as the Support Vector Machine (SVM) to better results for prediction.

Furthermore, the following future works were recommendations for the next research are below:

1. The study of comparison between the neural network and the SVM.

2. The improvement of results that can display the flood area automatically.

3. The system development of the real time prediction.

REFERENCES

1. กรมอุตุนิยมวิทยา. (2522). นิยามศัพท์อุตุนิยมวิทยา.*กรุงเทพมหานคร: โรงพิมพ์กรมอุตุวิทยา.*

2. เล็ก จินดาสงวน. (2546). วิกฤตการณ์น้ำท่วมและภัยแล้งในประเทศไทย**.** *ฉบับวันชูชาติ:* 10-127.

3. Heywood,I.(1998). An Introduction to Geographical Information System. *New York: Longman.*

4. Worboys, M.F. and M. Duckham. (2004). GIS: A Computing Perspective. *2nd ed. Florida: CRC Press.*

5. Goodchild, M.F., Steyaert, L.T., Park, B.O., Johnston, C., Maidment, D., Crane, M., and Glendinning, S. (1996). GIS and Environmental Modeling: Progress and Research Issues. *Fort Collins: GIS World Inc.*

6. Songsak Pattrawutichai. (2003). Flood forecasting in Lam Phachi River Basin by MIKE11 Model and Artificial Neural Networks. *M.Eng. Thesis in Irrigation Engineering. Bangkok: faculty of Engineering, Kasetsart University.*

7. Terayut Jaitrong. (2005). Flood forecasting in Ubol Ratana Reservoir by Artificial Neural Network for flood management. *M.Eng. Thesis in Irrigation Engineering. Bangkok: faculty of Engineering, Kasetsart University.*

8. Mandal, S., Saha, D., & Banerjee, T. (2005). A Neural Network Based Prediction Model for Flood in a Disater Management System with Sensor Networks. *Proceedings of ICSIP 2005*, 4-7 Jan, 78-82.

9. Campolo, M., Andreussi, P., & Soldati, A. (1999). River flood forecasting with a neural network model. *water resources research,35,*1191-1197.

10. Adnan, R., Ruslan, F.A., Samad, A.M., & Zain, Z.M. (2012). Flood Water Level Modelling and Prediction Using Artificial Neural Network : Case Study of Sungai Batu Pahat in Johor. *IEEE Control and System Graduate Research Colloquium*, 16-17 July, 22-25.

11. Cruz, I.F., Pedrollo, O., Castro, N.M.R., Girard, P., Zeilhofer, P., & Hamilton, S.K. (2011). Historical reconstruction of floodplain inundation in the Pantanal (Brazil) using neural networks. *Journal of Hydrology*, *399*, 376-384.

12. Sulaiman, M., Shafie, A.E., Karim, O., & Basri, H. (2011). Real-time flood forecasting by employing artificial neural network based model with zoning matching approach. *Hydrology and Earth System Sciences Discussions, 8,* 9357-9393.

13. Bustami, R.A., Bessaih, N., & Muhamud, M.S. (2006). Artificial Neural Network for Daily Water Level Estimation. *Engineering e-Transaction, University of Malaya, 1*, 1 March , 7-12.

14. Leahy, P., Kiely, G., & Corcoran, G. (2008). Structural optimization and input selection of an artificial neural network for river level prediction. *Journal of Hydrology*, *335*, 192-201.

15. Waleerat Rangsimapirat. (2005). A Mathematical Model For Locating Flooding Area In The Lower Maeklong River Basin. *M.Sc. Thesis in Technology of Information system Management. Bangkok: Faculty of Graduate Studies, Mahidol University*.

16. Thudchai Sansena. (2004). The Application of Hydraulic Model With Geographic Information System For Obtaining Data To Create Flood Risk Mapping In The Upper Part Of Maeklong River. *M.Sc. Thesis in Technology of Information system Management. Bangkok: Faculty of Graduate Studies, Mahidol University.*

17. Ruthairat Mangsilp. (2007). GIS And Hydrological Model For Flood Risk Area Assessment In Maetaeng Watershed, Chiang Mai Province. *M.Sc. Thesis in Information Management on Environment and Resources. Bangkok: Faculty of Graduate Studies, Mahidol University*.

18. Panvadee Arayawongwarn. (2001). Mike 11 and geographic information system for estimating the flood area in the Yom Basin, Phare province. *M.Sc. Thesis in Information Management on Environment and Resources. Bangkok: Faculty of Graduate Studies, Mahidol University.*

19. Chutima Wongwitaya. (2001). A New View of an Old Foe: Using GIS and Hydrology Software to Produce 3D Animation of Potential Flooding

Events. *M.Sc. Thesis in Technology of Information system Management. Bangkok: Faculty of Graduate Studies, Mahidol University.*

20. Jeong, R.H.(2004). The Prediction of Bus Arrival time Using Automatic Vehicle Location Systems Data. *A Ph.D. Dissertation in Civil Engineering. Texas: Office of Graduate Studies A&M University.*

21. Beale, M.H., Hagan, H.T. and Demuth, H.B. (2010). Neural Network Toolbox™ User's Guide. Retrieved December 12, 2010, from MathWorks Website: www.mathworks.com/help/pdf_doc/nnet/nnet_ug.pdf

APPENDICES

APPENDIX A
HOW TO THE ANALYSIS OF FLOOD AREA SIZE WITH ARCGIS 9.3

The analysis of flood area size had steps as follow

1. Open the shape file of flood area and Suphanburi province are shown in Figure A.1.

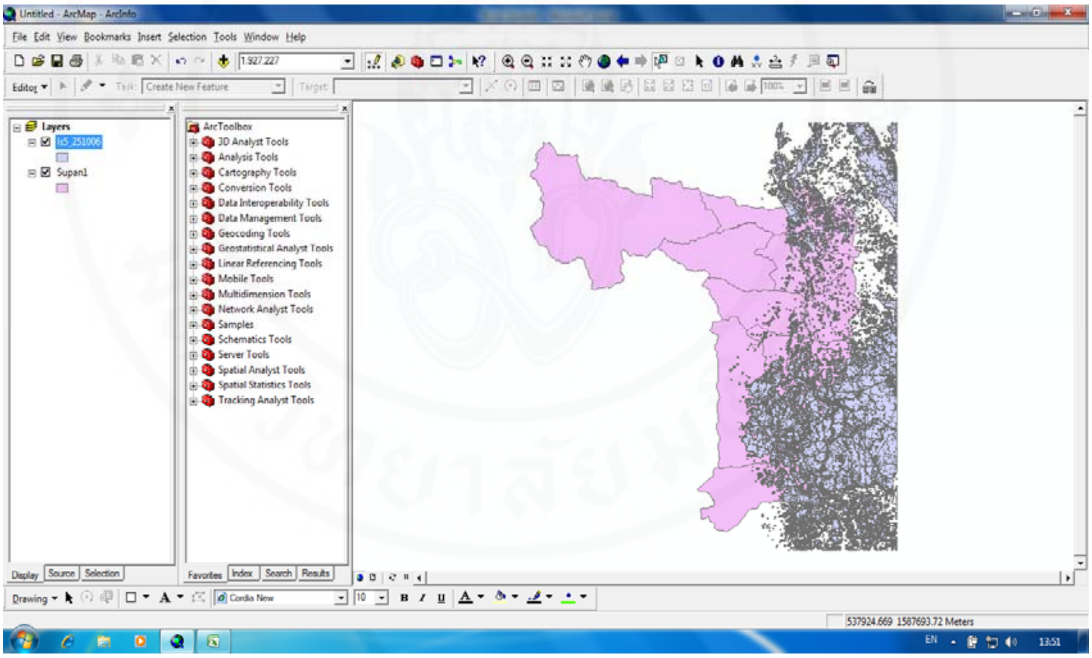

Figure A.1 Open shape file

2. Open the Clip function for analysis of flood area which in the area of Suphanburi province as showed in Figure A.2.

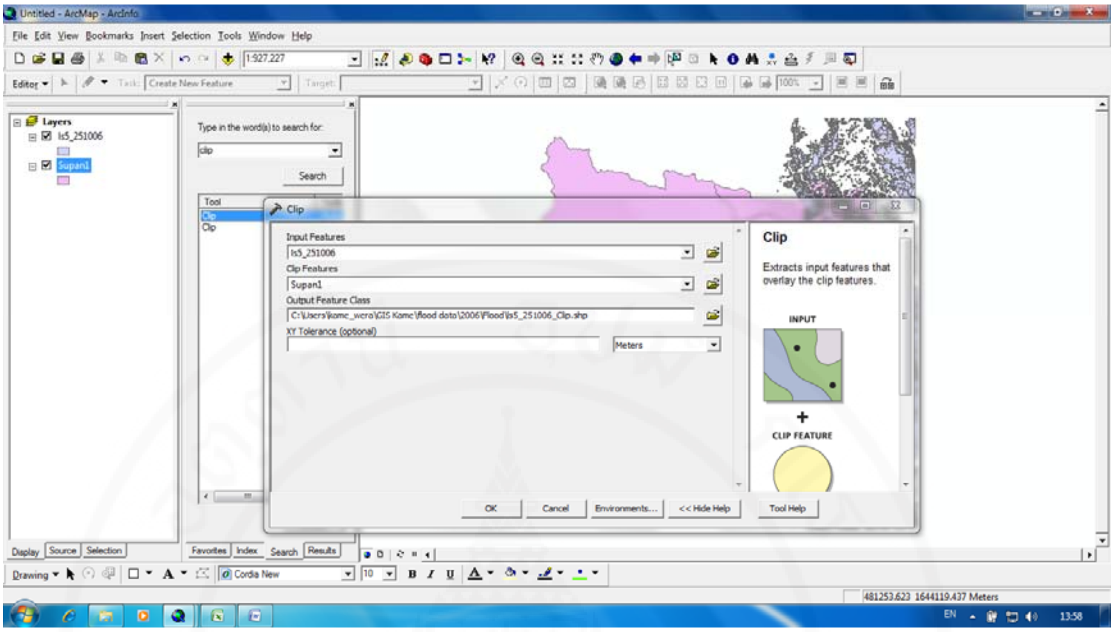

Figure A.2 Open the Clip function for analysis of flood area

3. After the analysis of flood area with the Clip function, the flood displayed in the area of Suphanburi province as showed in Figure A.3.

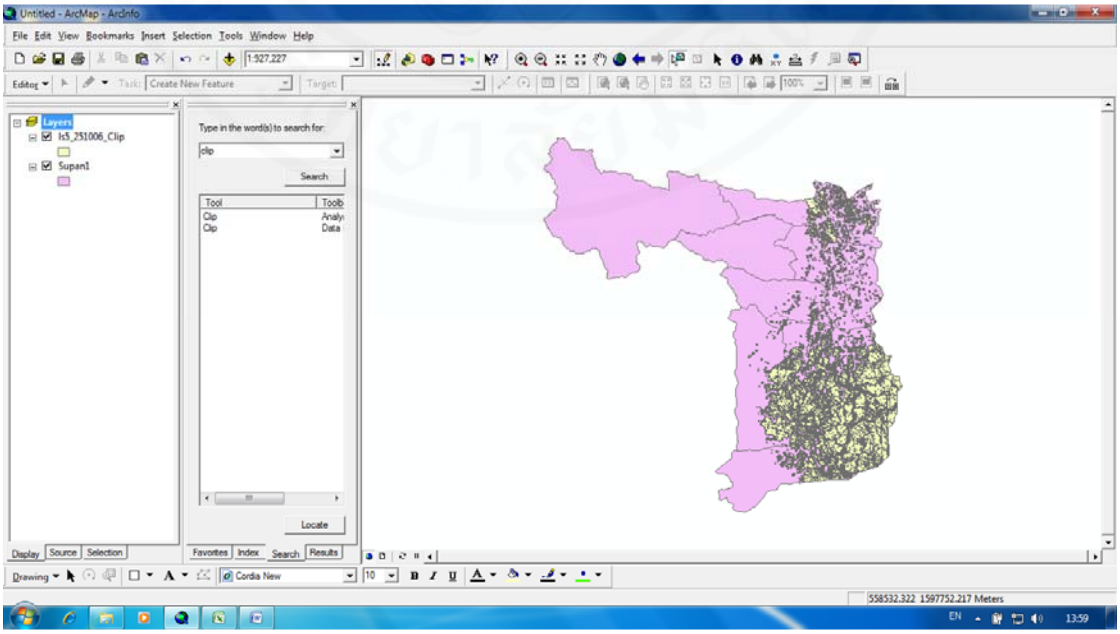

Figure A.3 Display the flood in the area of Suphanburi province

4. Open the Attribute Table of flood in the area of Suphanburi province by the right click to open the table as showed in Figure A.4.

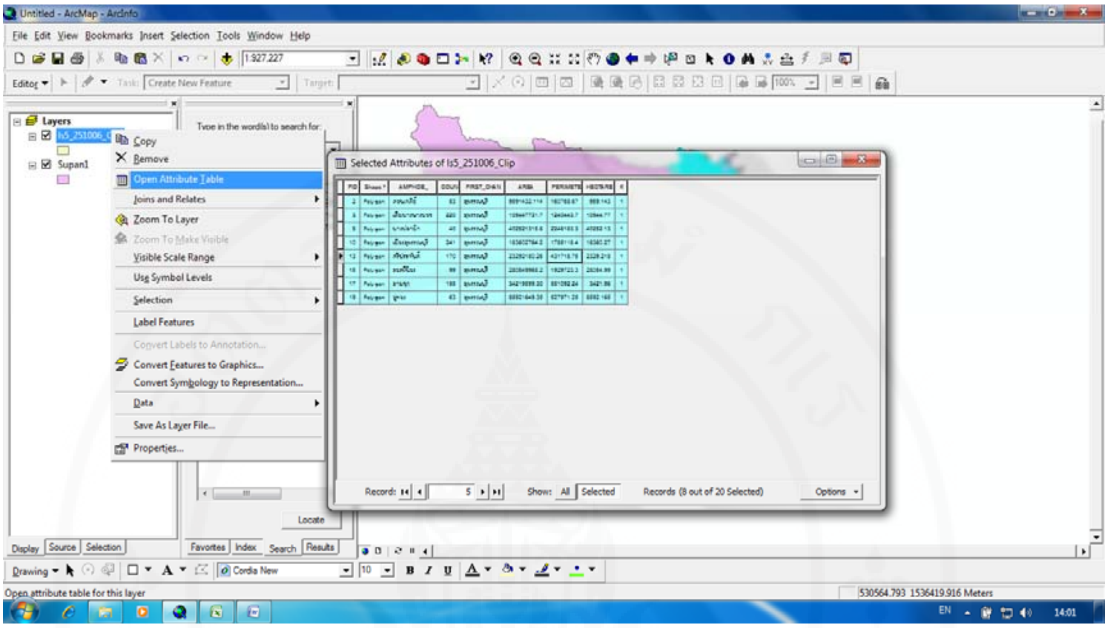

Figure A.4 Display the attribute table of flood area

5. Create the new attribute in table by click the options button and choosing the Add Field after that create the attribute name is Flood_area that are shown in Figure A.5.

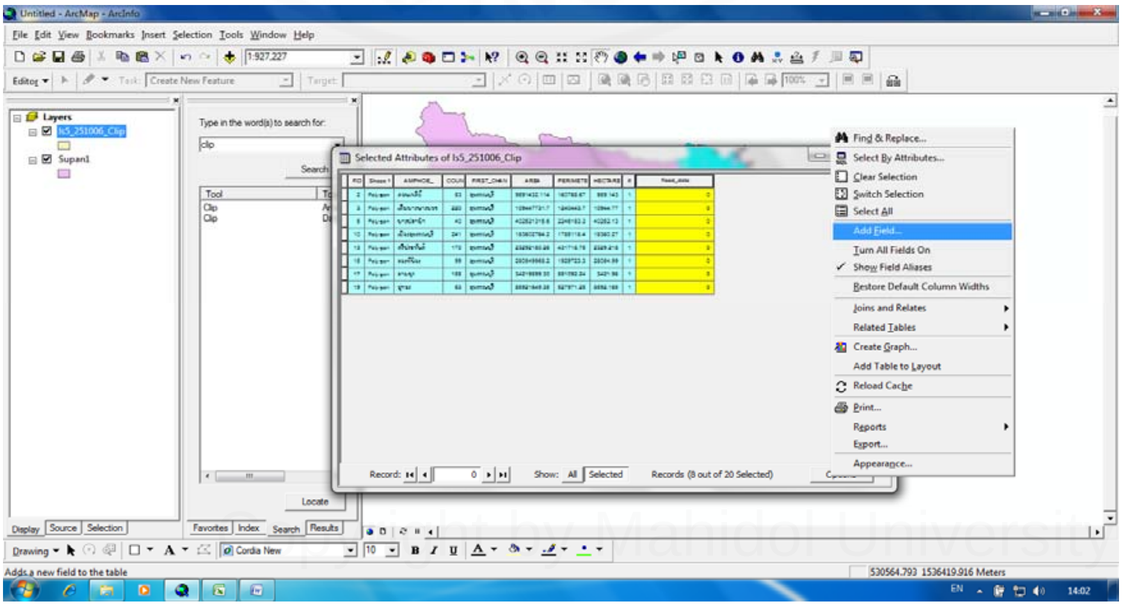

Figure A.5 Create the new attribute in table

6. Right click the new attribute and choosing the Calculate Geometry as showed in Figure A.6.

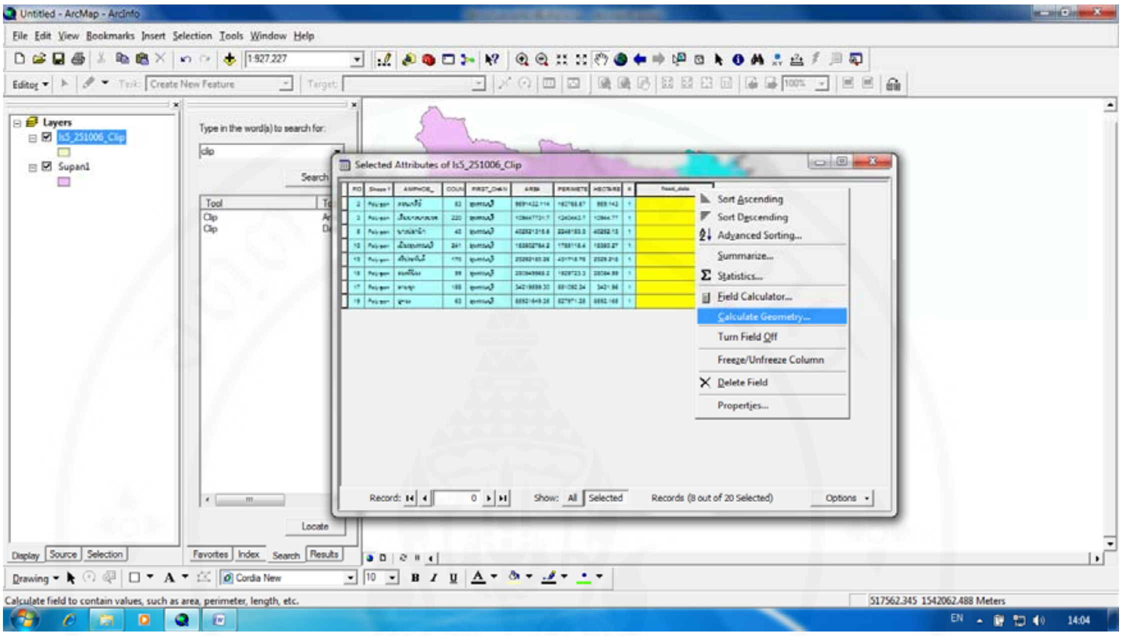

Figure A.6 Calculate the geometry for flood area size

7. Set the unit of flood area size by choosing at Units and the selection of unit as square kilometer that showed in Figure A.7.

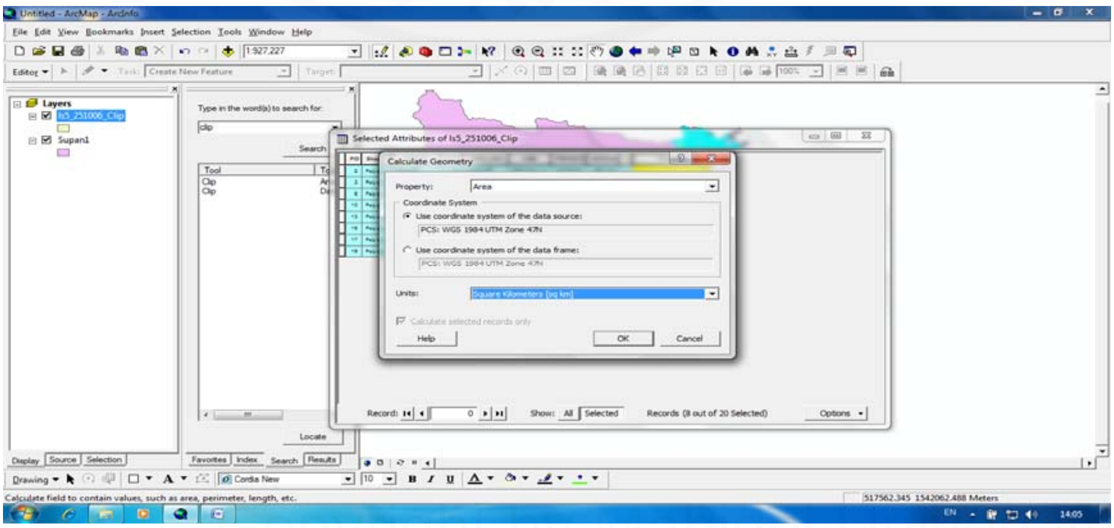

Figure A.7 Set the unit of flood area size

8. Click the OK button, the program will calculate the flood area and display in the new attribute that showed in Figure A.8.

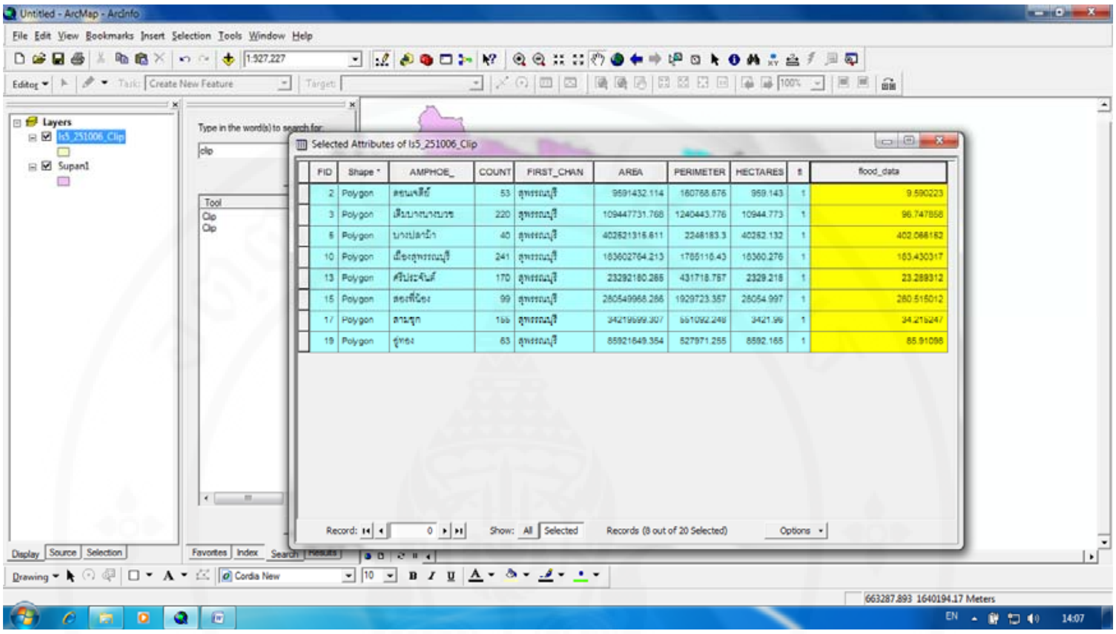

Figure A.8 Display the results of flood area size

9. Calculate the summation of flood area size of all Suphanburi province by the right click at the new attribute and choosing the Statistics. The result showed the summation of flood area size. The Figure A.9 was shown below.

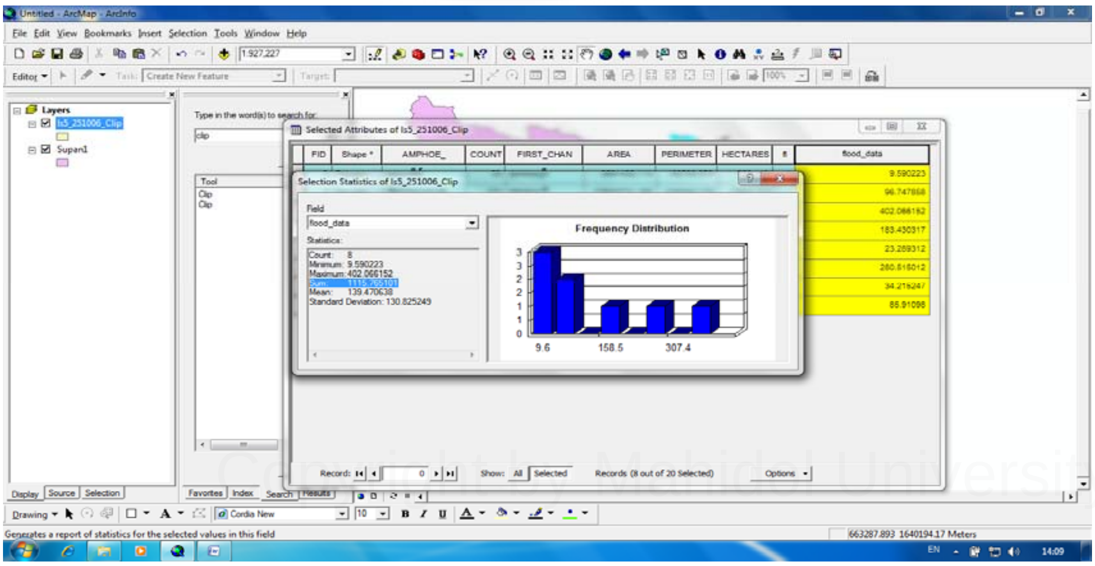

Figure A.9 the summation of flood area size

APPENDIX B
THE SAMPLES OF FLOOD AREA SIZE IN SUPHANBURI PROVINCE

Figure B.1 Flood area size 50.95 km^2

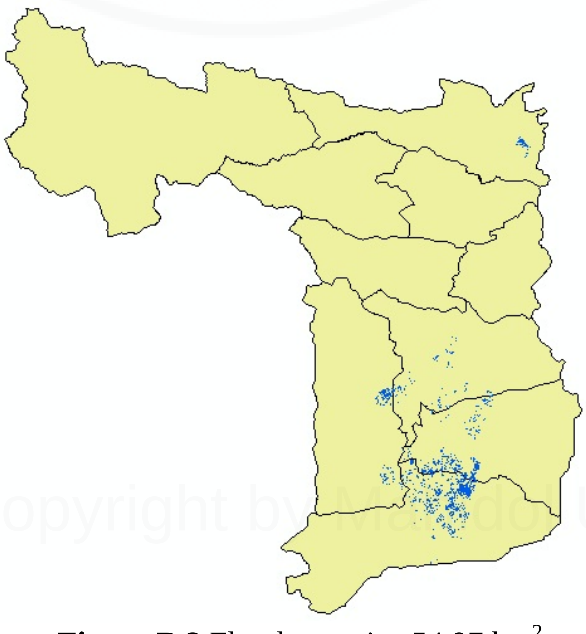

Figure B.2 Flood area size 54.97 km^2

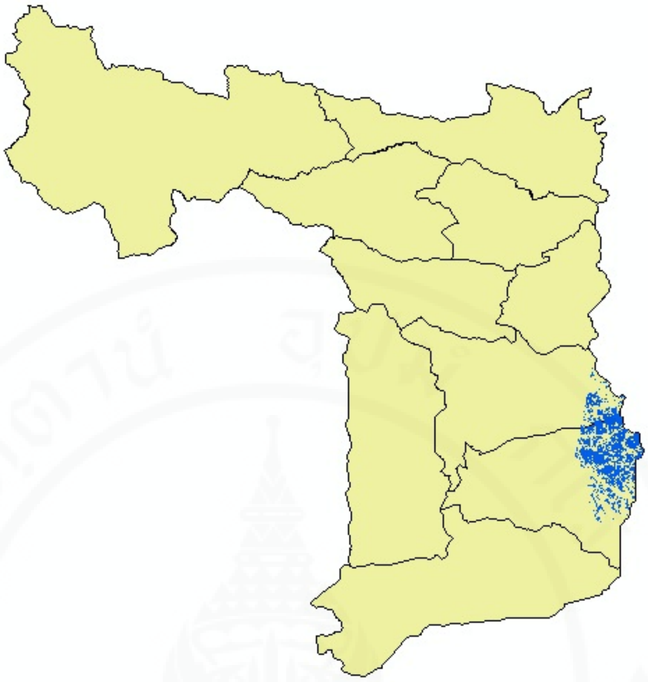

Figure B.3 Flood area size 77.75 km^2

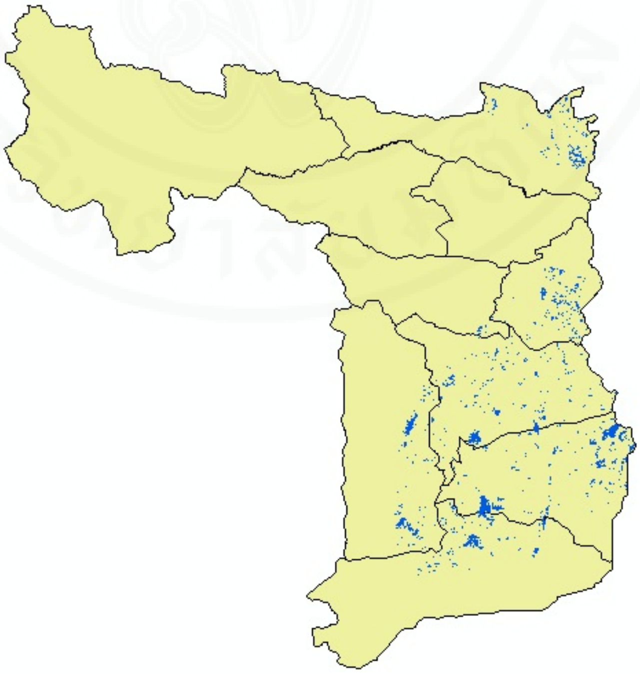

Figure B.4 Flood area size 83.63 km^2

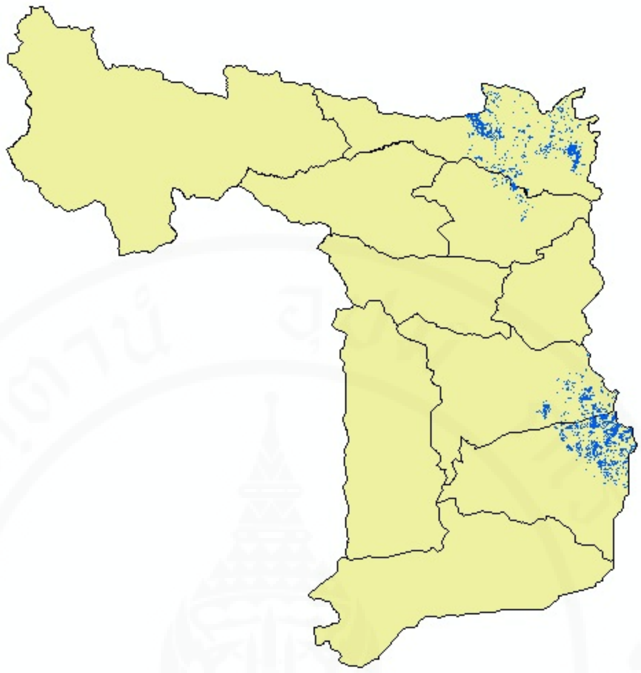

Figure B.5 Flood area size 84.03 km^2

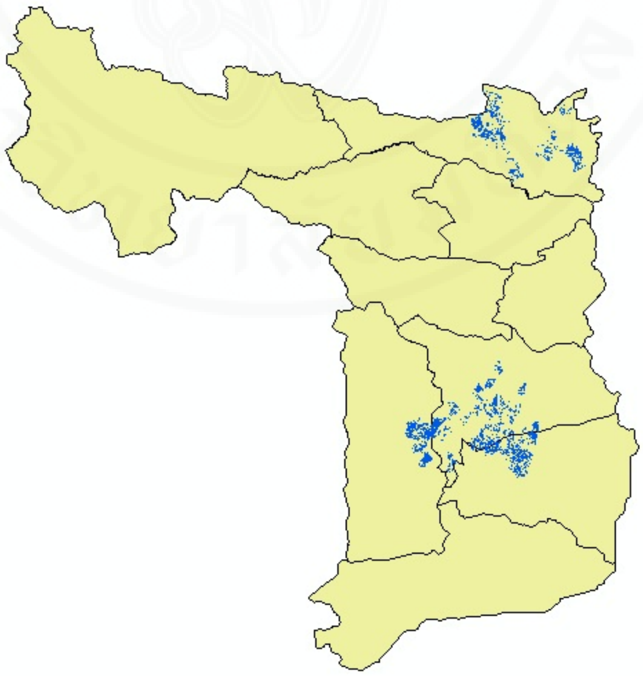

Figure B.6 Flood area size 91.17 km^2

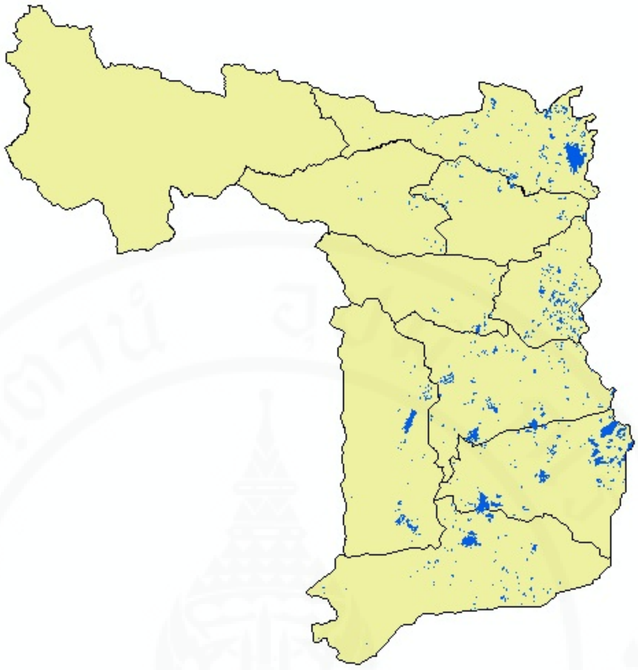

Figure B.7 Flood area size 120.48 km^2

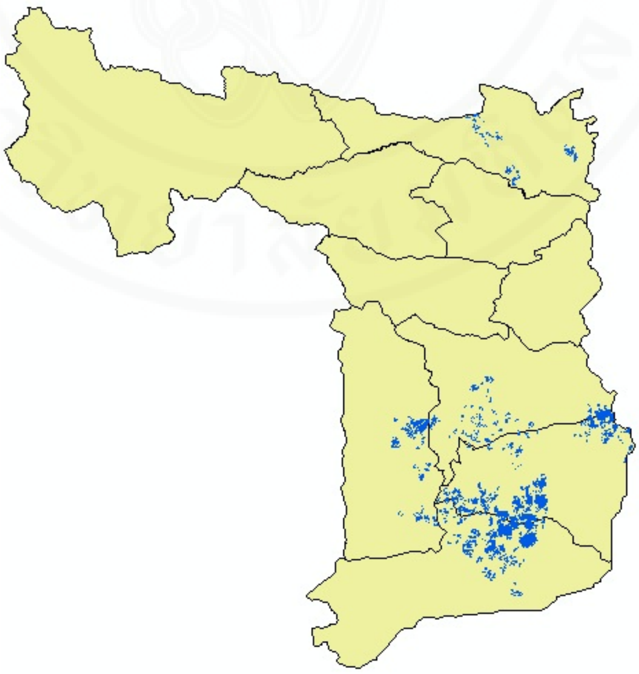

Figure B.8 Flood area size 130.1 km^2

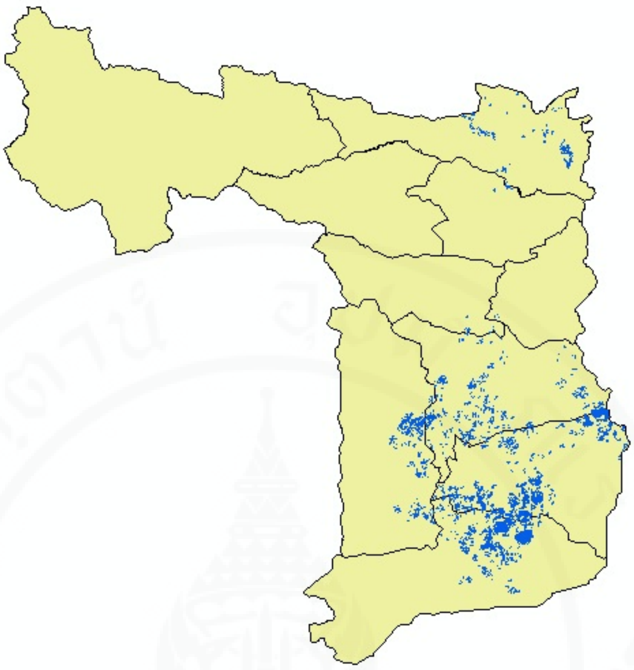

Figure B.9 Flood area size 169.66 km^2

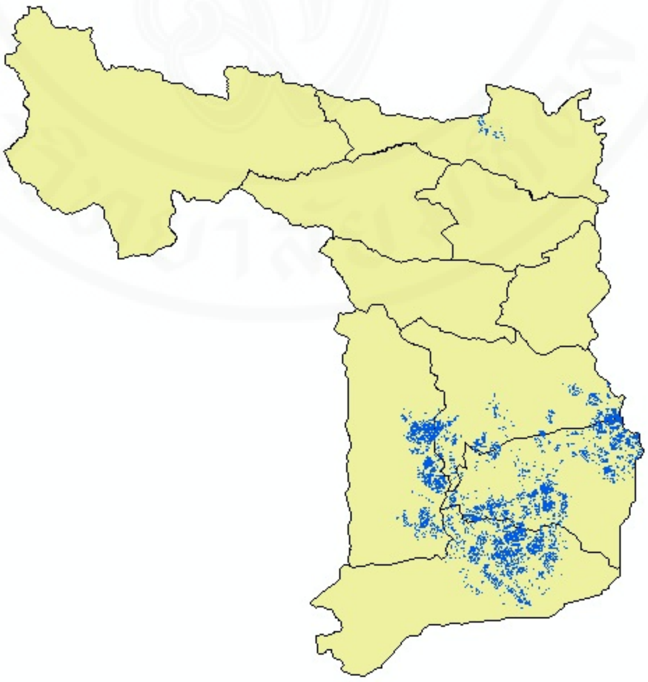

Figure B.10 Flood area size 203.29 km^2

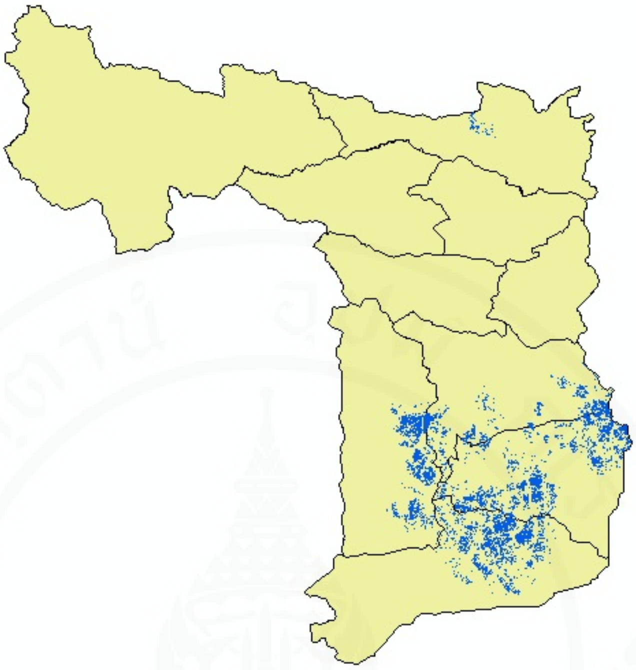

Figure B.11 Flood area size 203.83 km^2

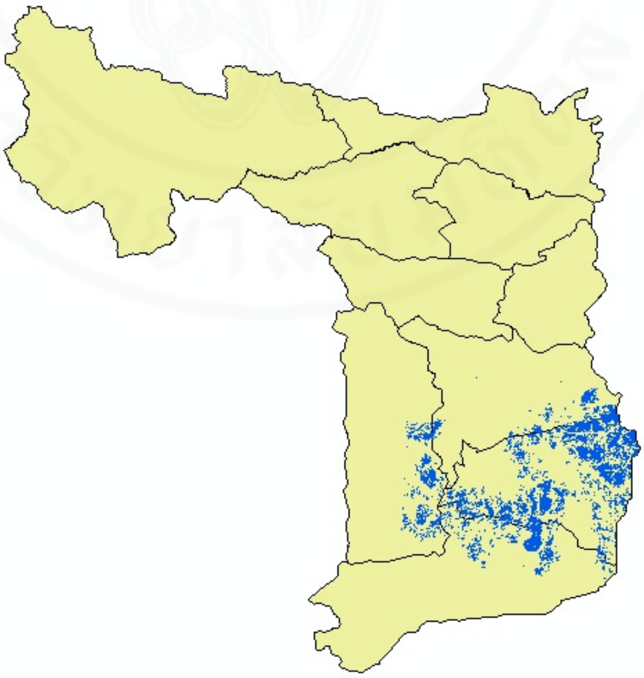

Figure B.12 Flood area size 228.13 km^2

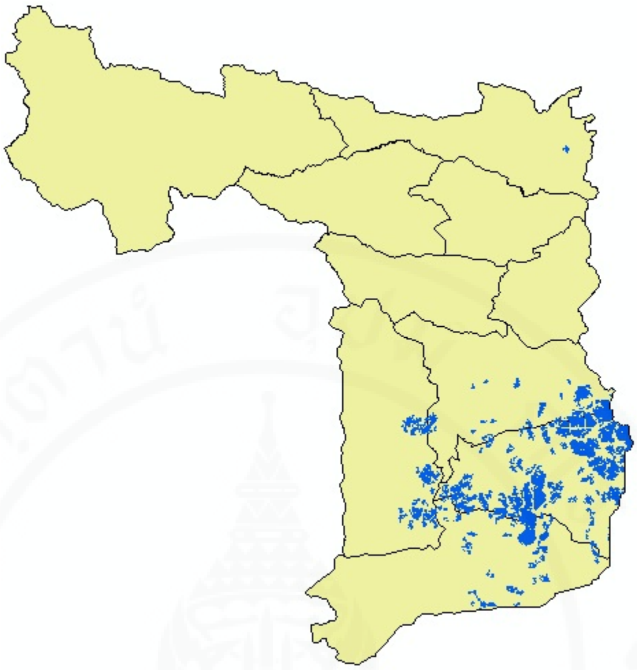

Figure B.13 Flood area size 238.97 km^2

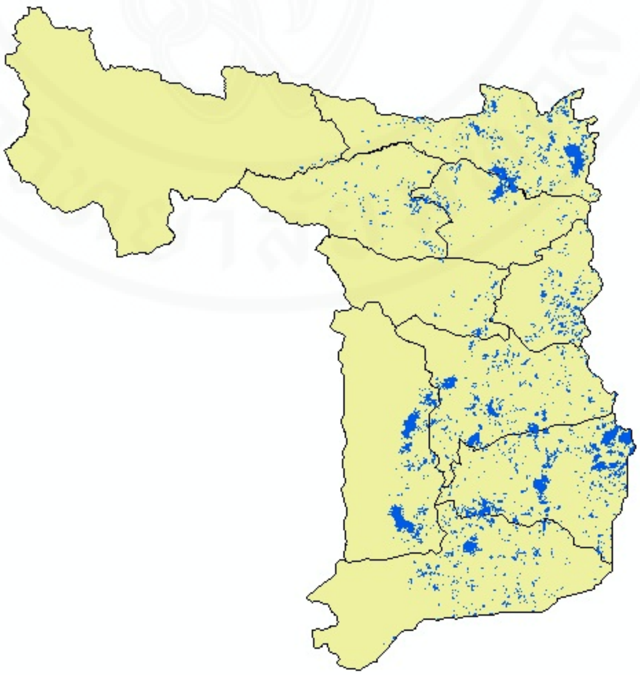

Figure B.14 Flood area size 265.11 km^2

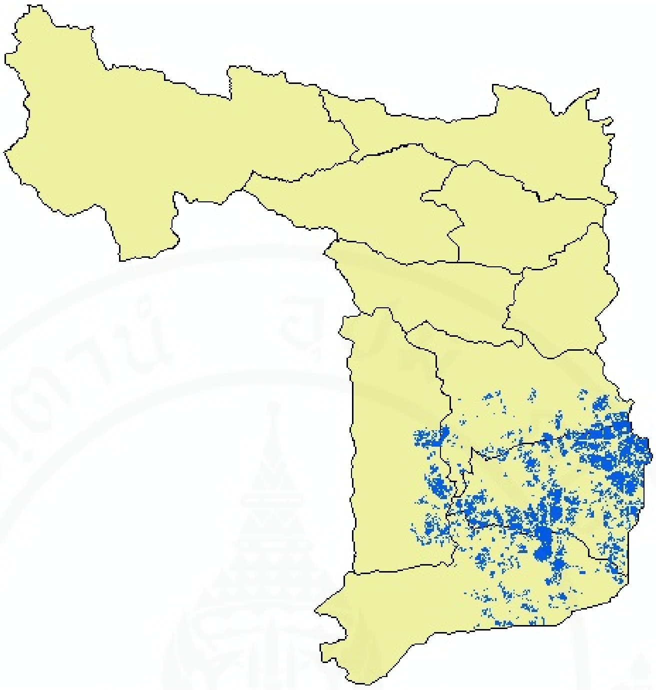

Figure B.15 Flood area size 282.01 km^2

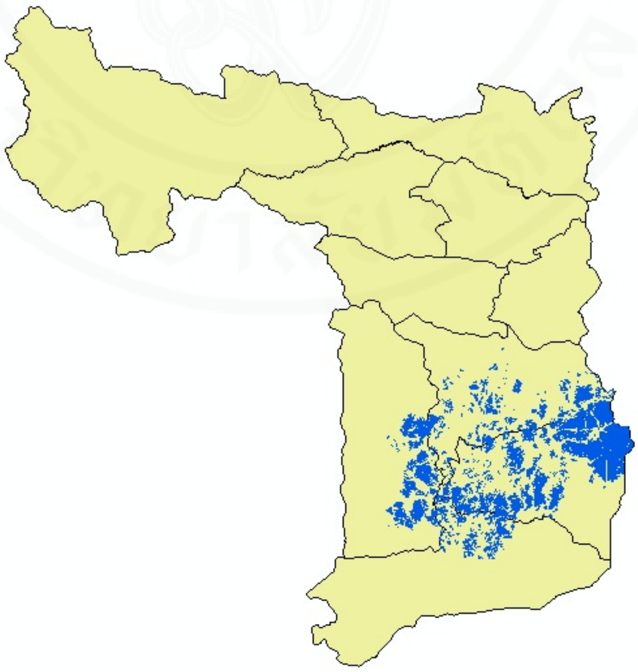

Figure B.16 Flood area size 341.65 km^2

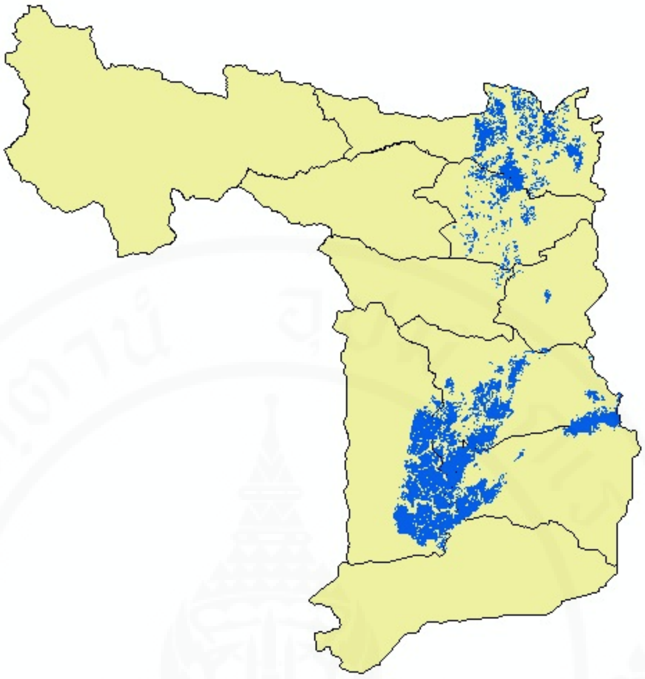

Figure B.17 Flood area size 409.79 km^2

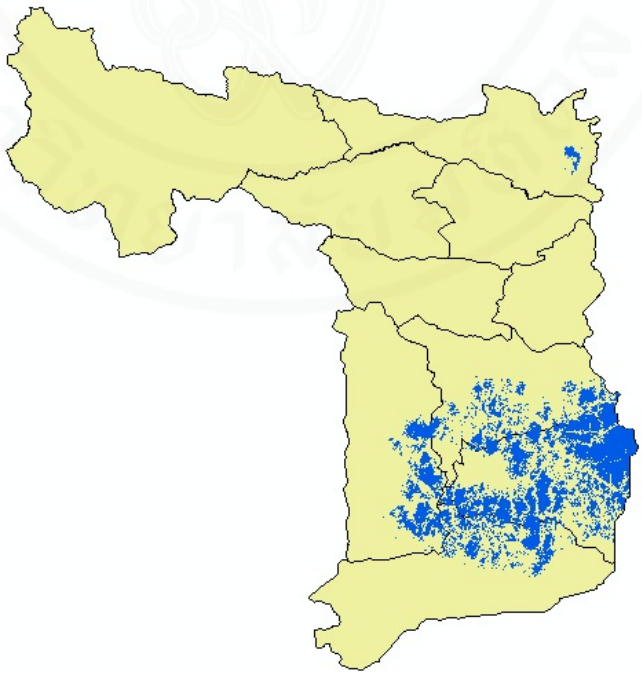

Figure B.18 Flood area size 439.77 km^2

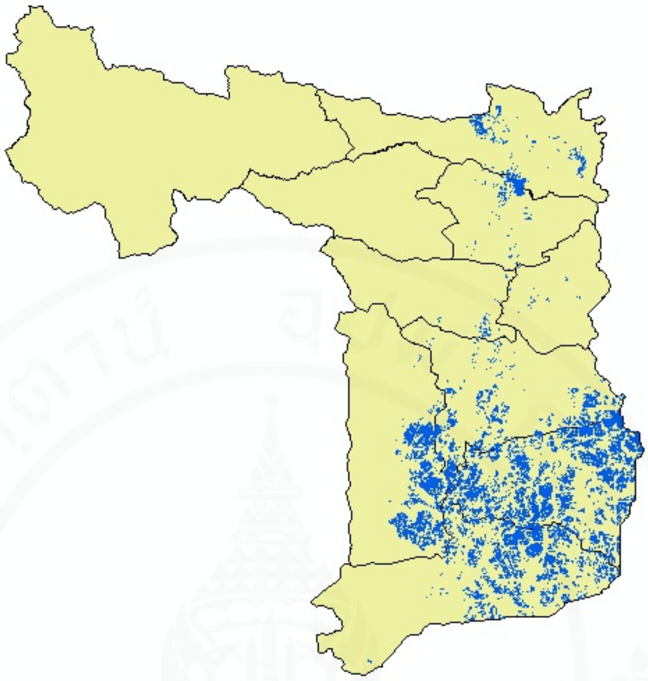

Figure B.19 Flood area size 454.91 km^2

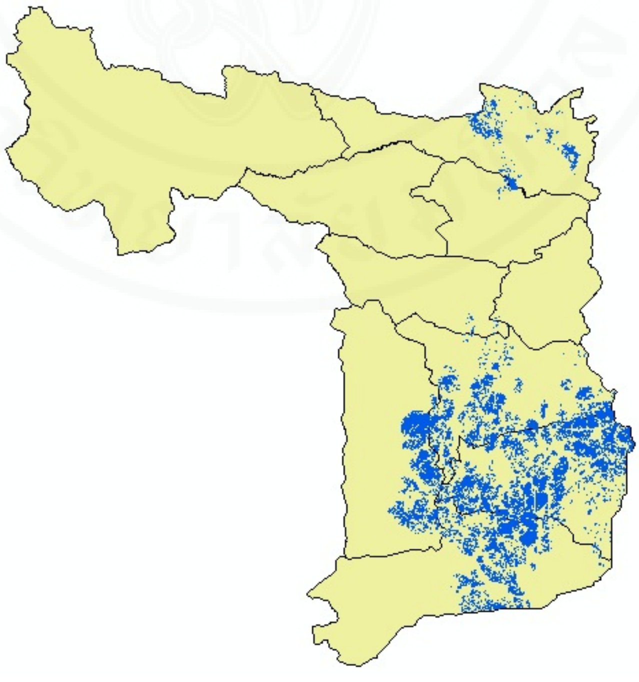

Figure B.20 Flood area size 461.6 km^2

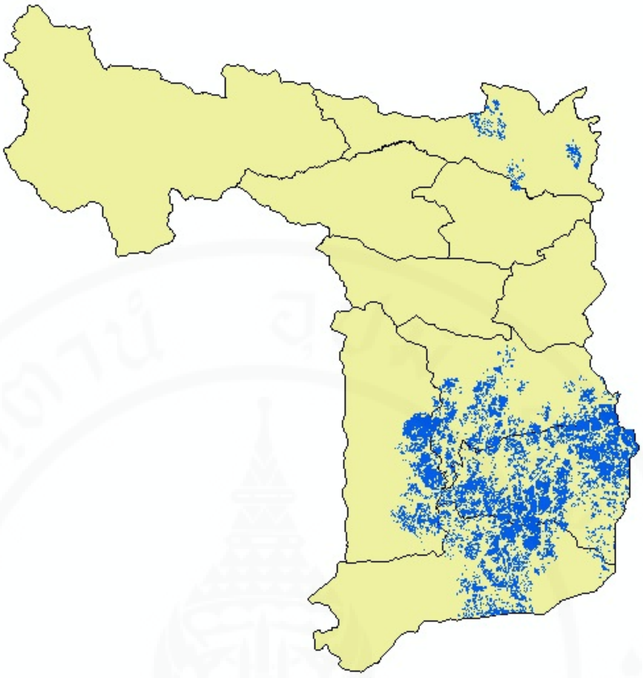

Figure B.21 Flood area size 475.02 km^2

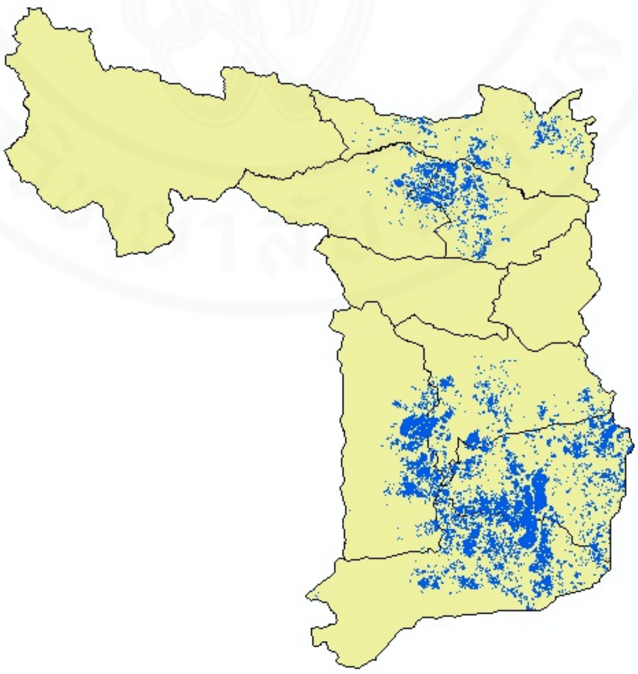

Figure B.22 Flood area size 489.43 km^2

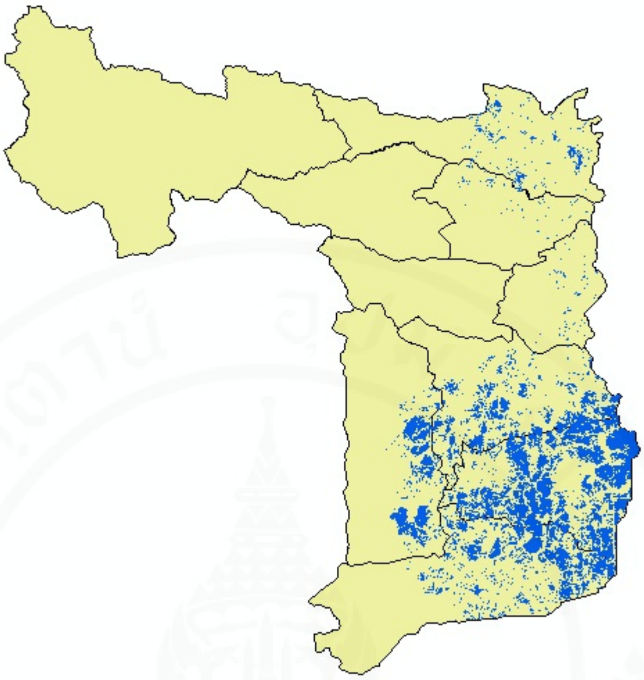

Figure B.23 Flood area size 497.42 km^2

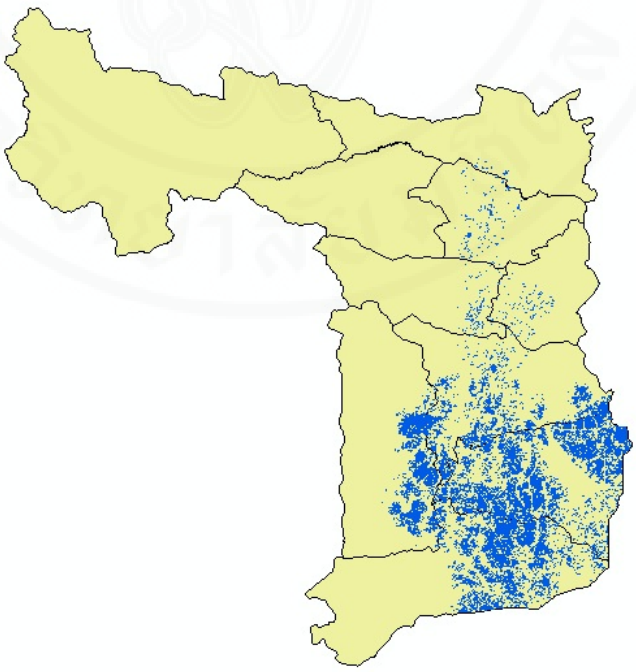

Figure B.24 Flood area size 513.88 km^2

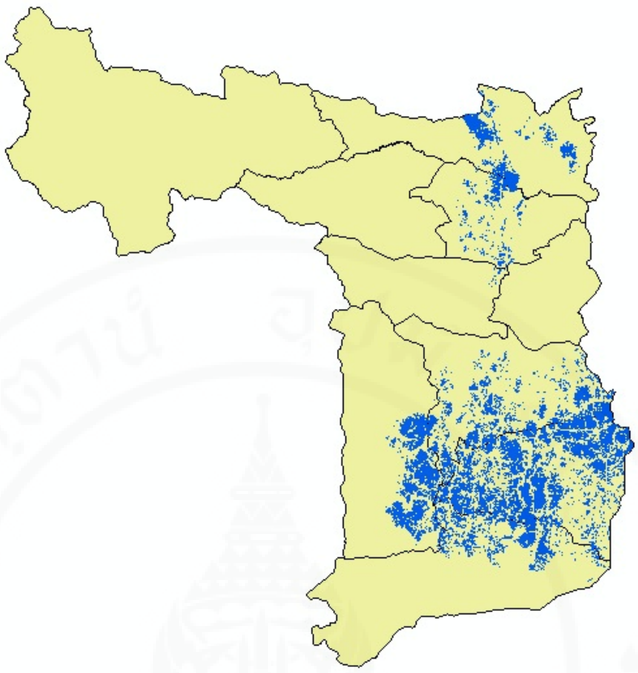

Figure B.25 Flood area size 527.06 km^2

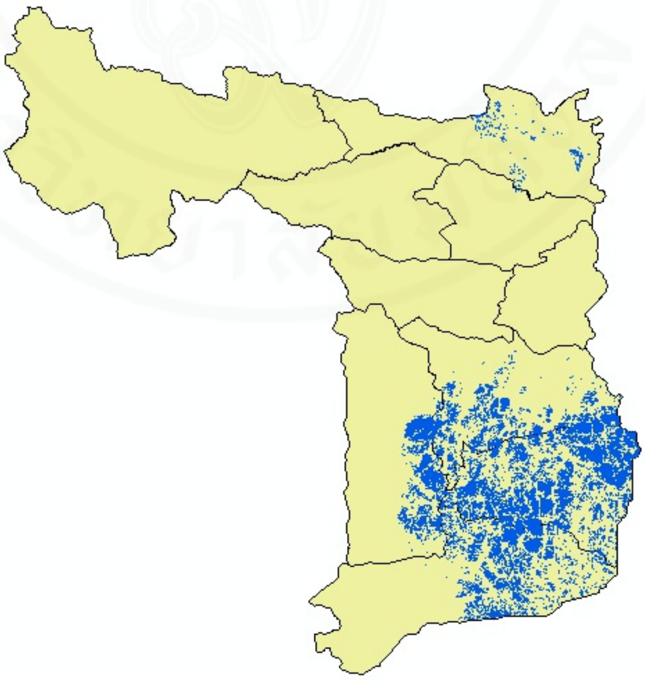

Figure B.26 Flood area size 546.97 km^2

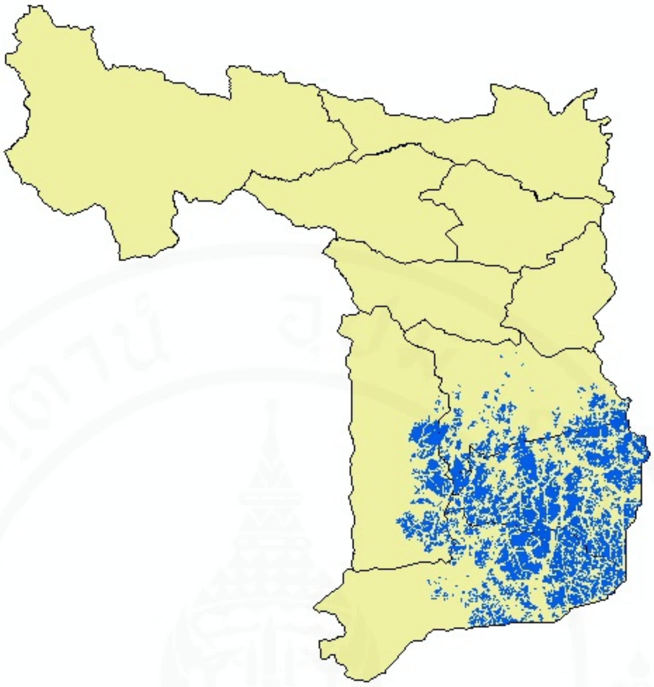

Figure B.27 Flood area size 552.01 km^2

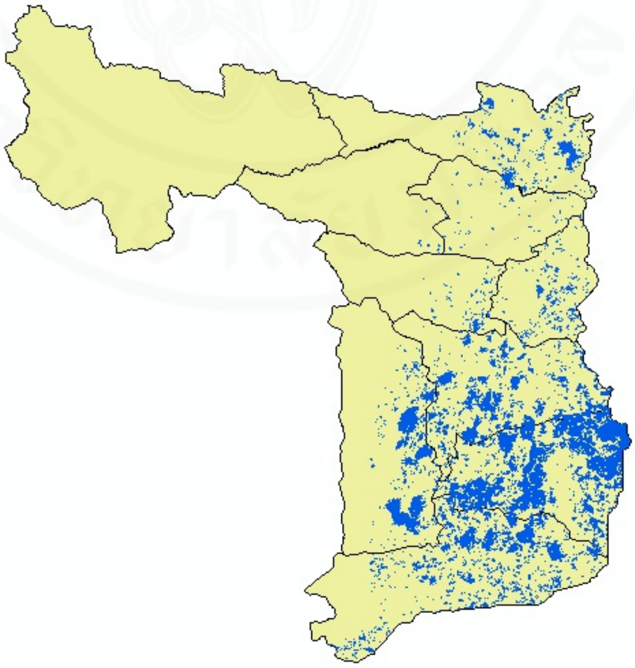

Figure B.28 Flood area size 590.24 km^2

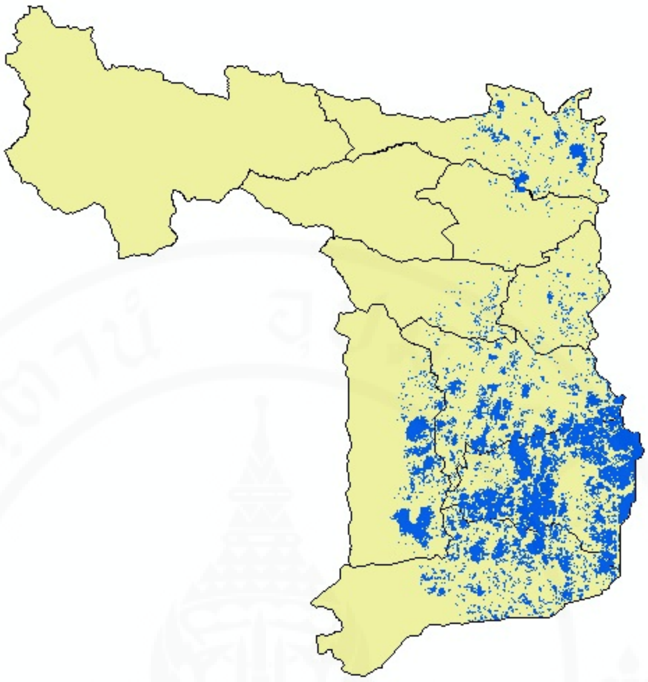

Figure B.29 Flood area size 599.75 km^2

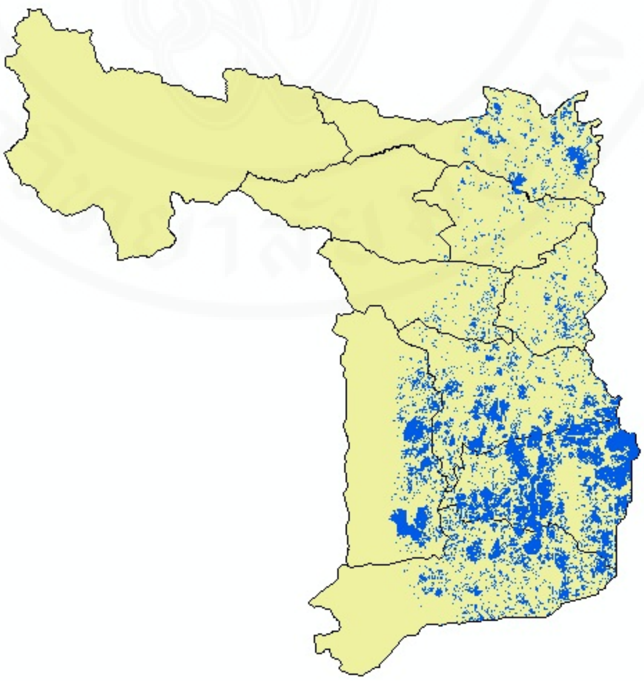

Figure B.30 Flood area size 599.79 km^2

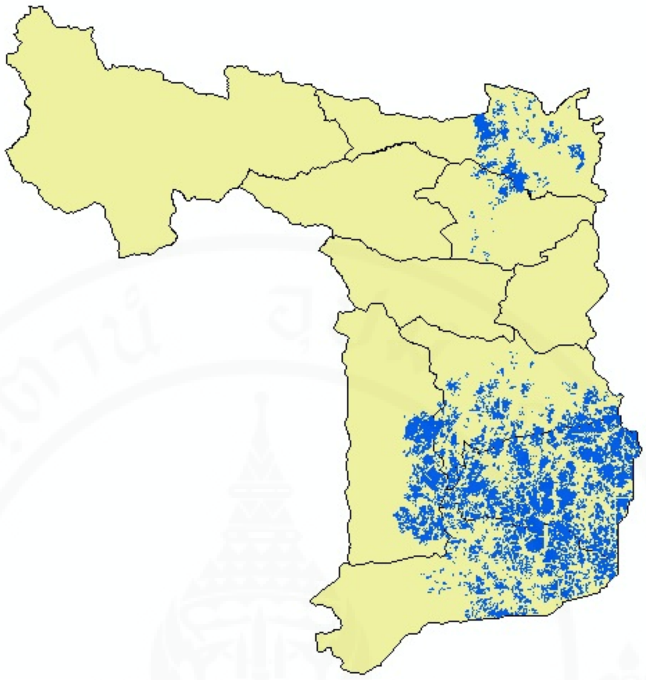

Figure B.31 Flood area size 613.19 km^2

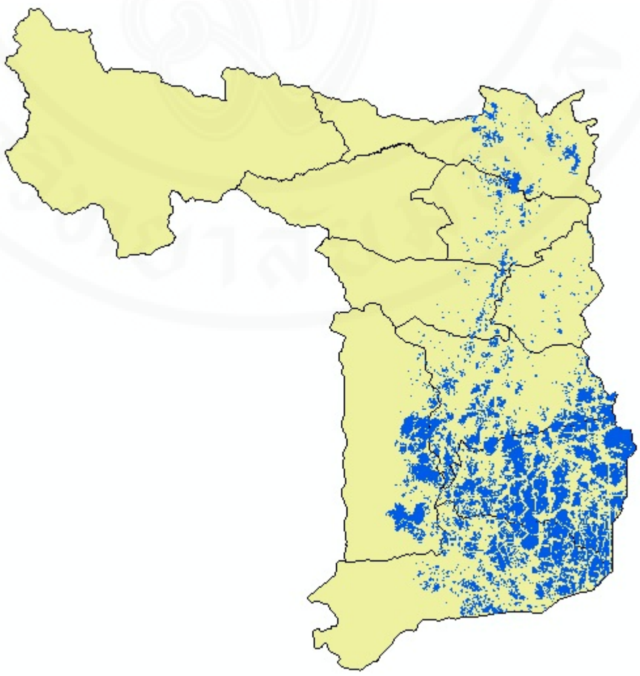

Figure B.32 Flood area size 634.93 km^2

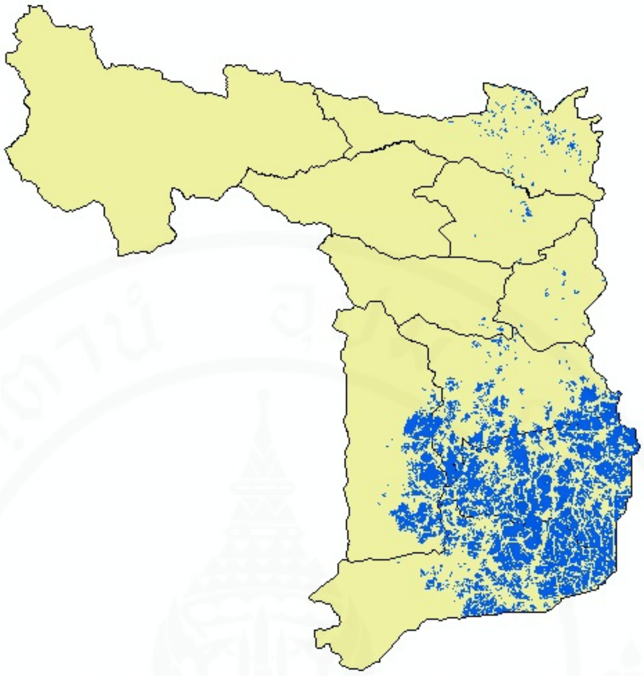

Figure B.33 Flood area size 665.67 km^2

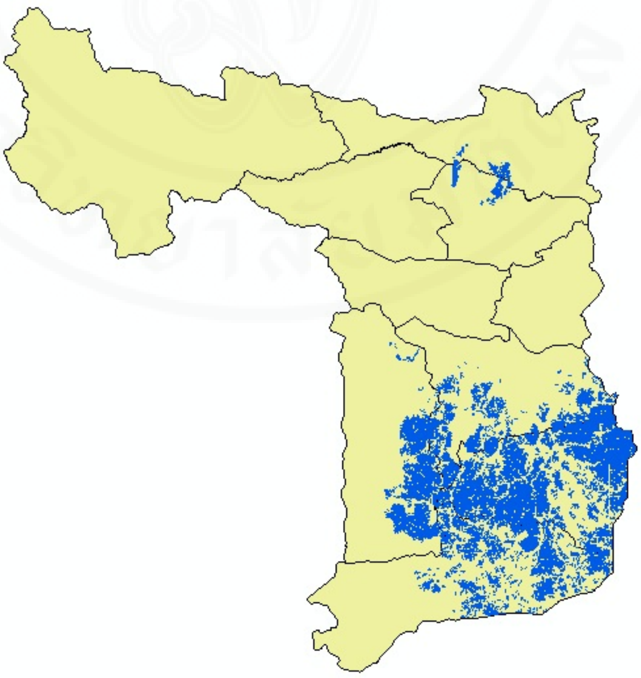

Figure B.34 Flood area size 702.04 km^2

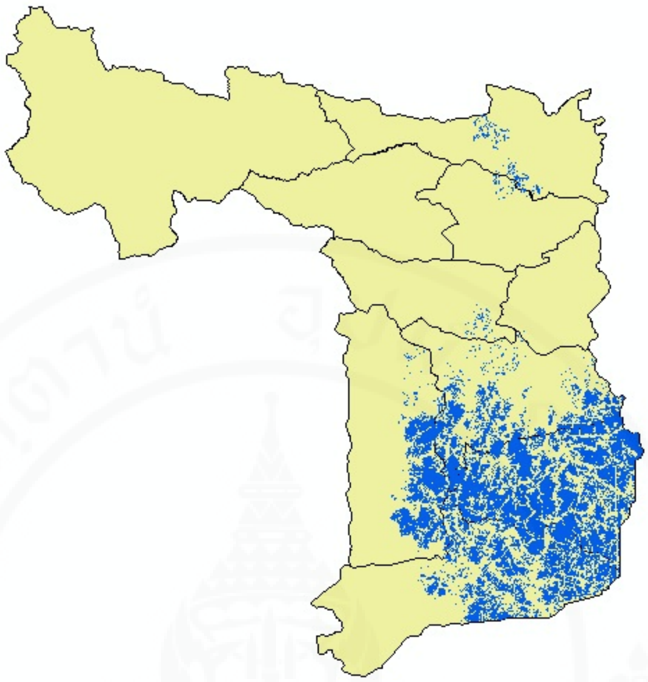

Figure B.35 Flood area size 714.25 km^2

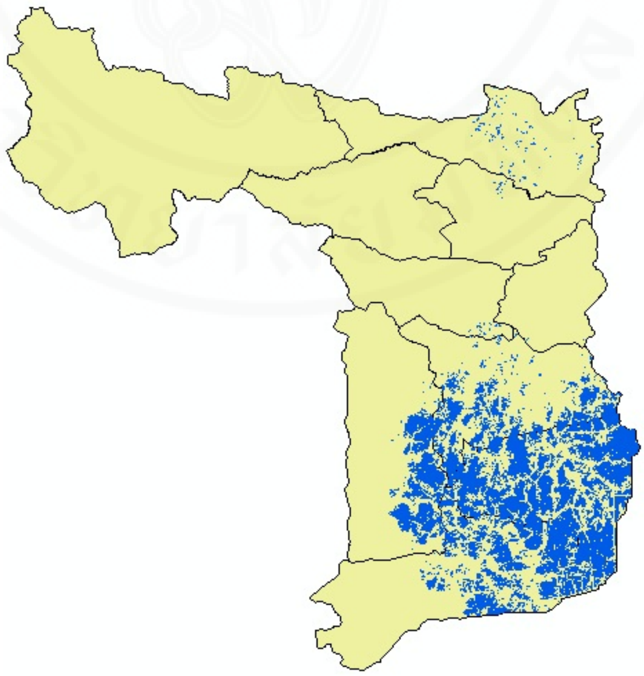

Figure B.36 Flood area size 720.52 km^2

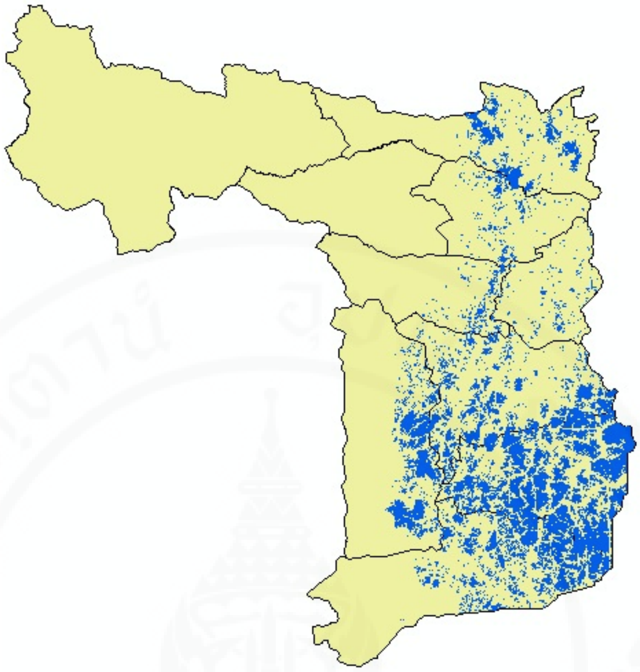

Figure B.37 Flood area size 722.3 km^2

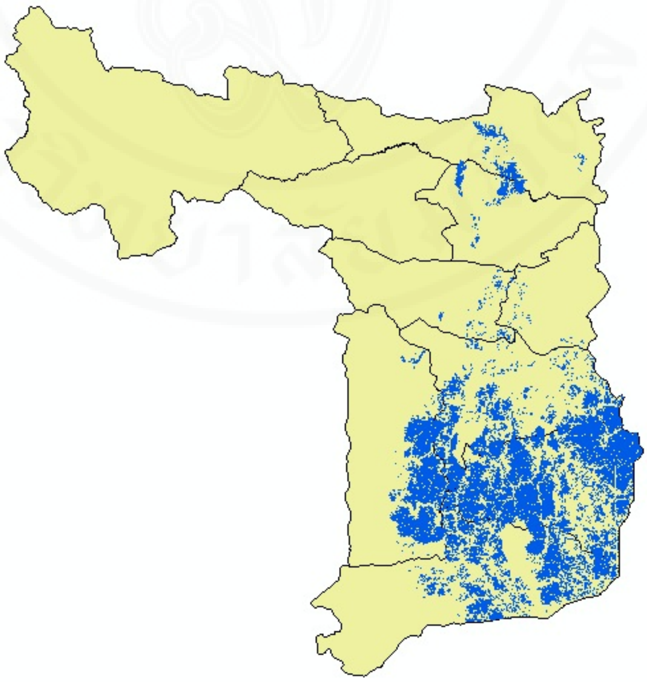

Figure B.38 Flood area size 737.03 km^2

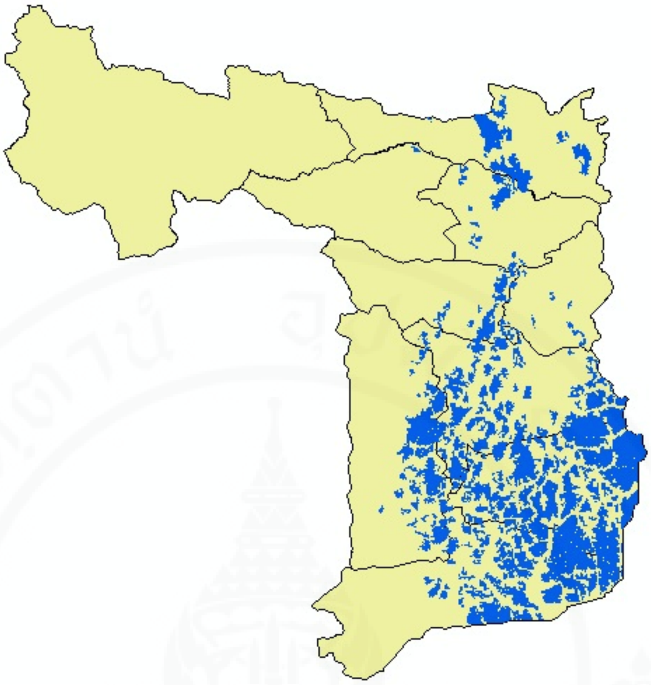

Figure B.39 Flood area size 816.19 km^2

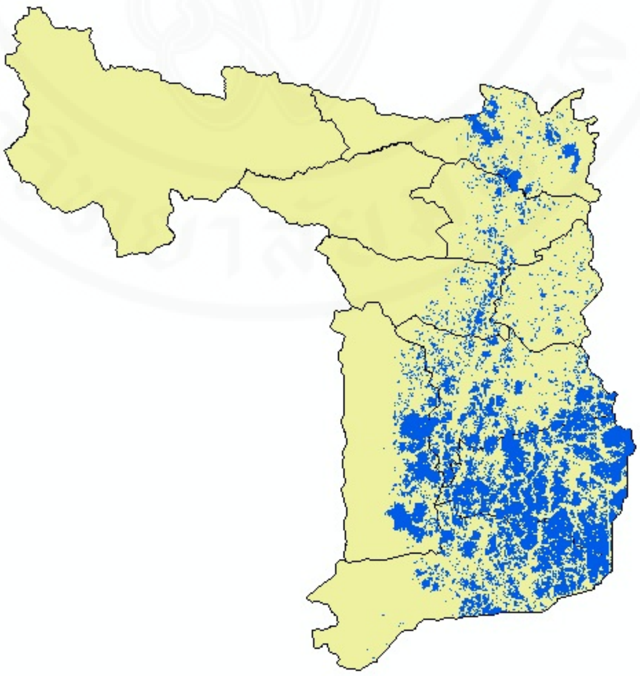

Figure B.40 Flood area size 842.35 km^2

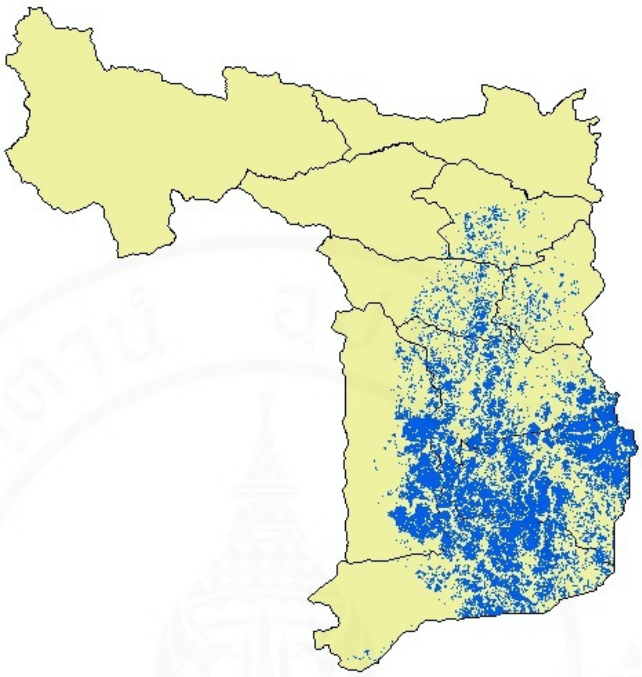

Figure B.41 Flood area size 860.59 km^2

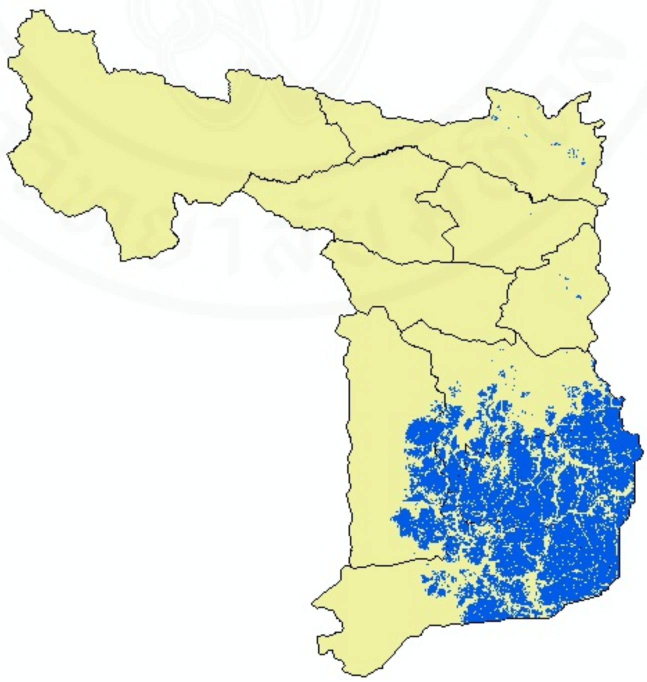

Figure B.42 Flood area size 872.92 km^2

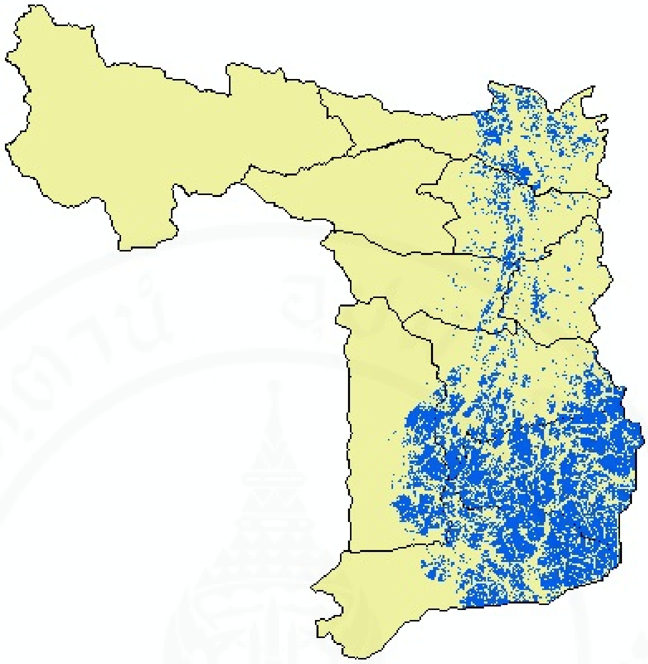

Figure B.43 Flood area size 900.84 km^2

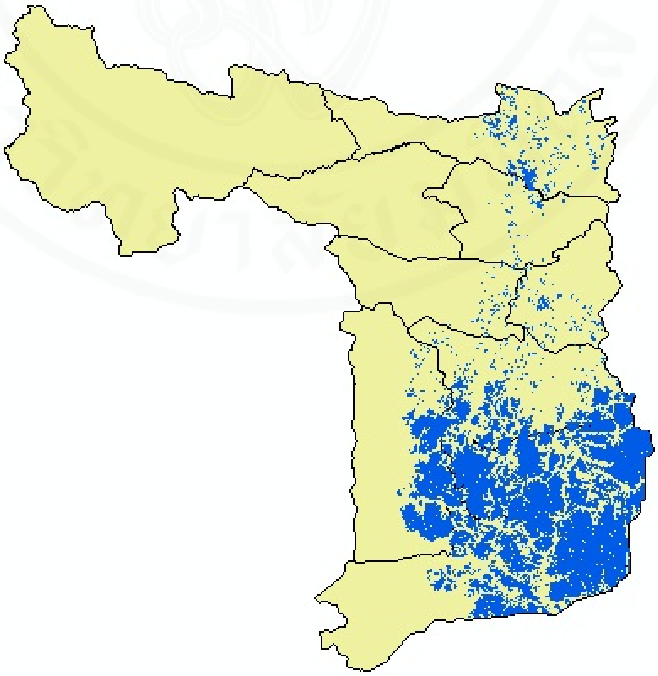

Figure B.44 Flood area size 908.28 km^2

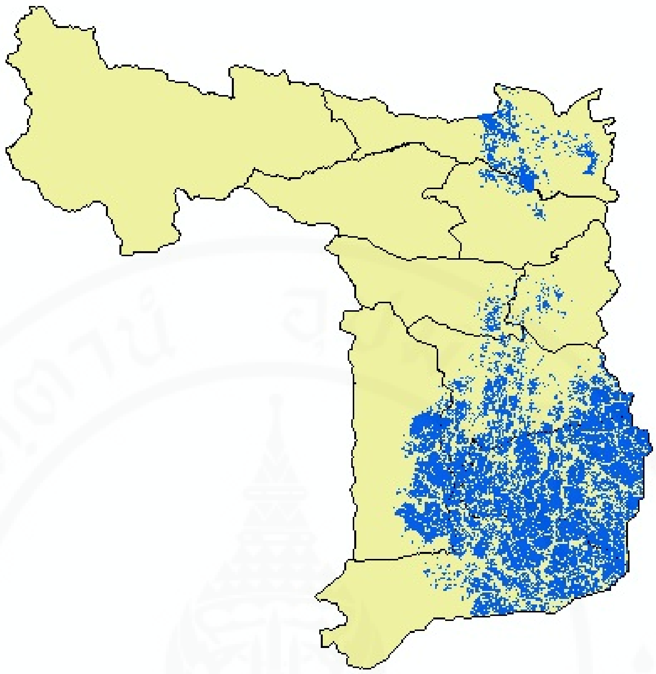

Figure B.45 Flood area size 912.63 km^2

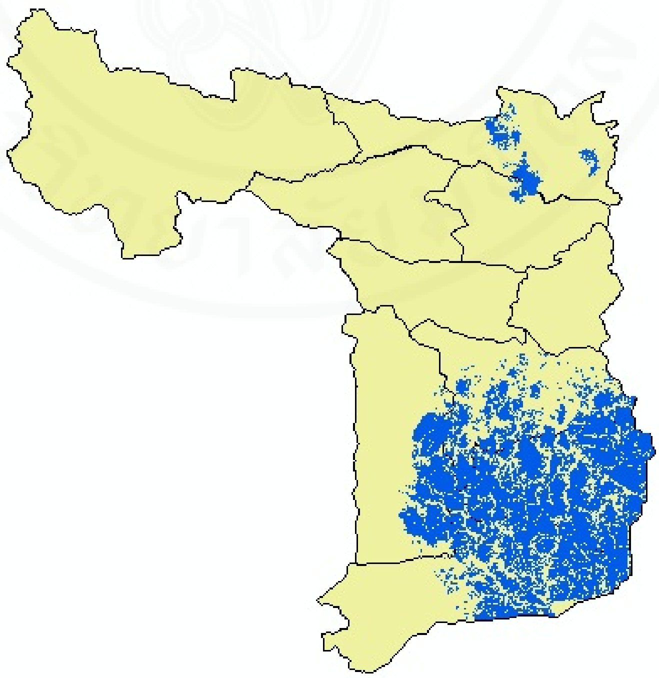

Figure B.46 Flood area size 942.32 km^2

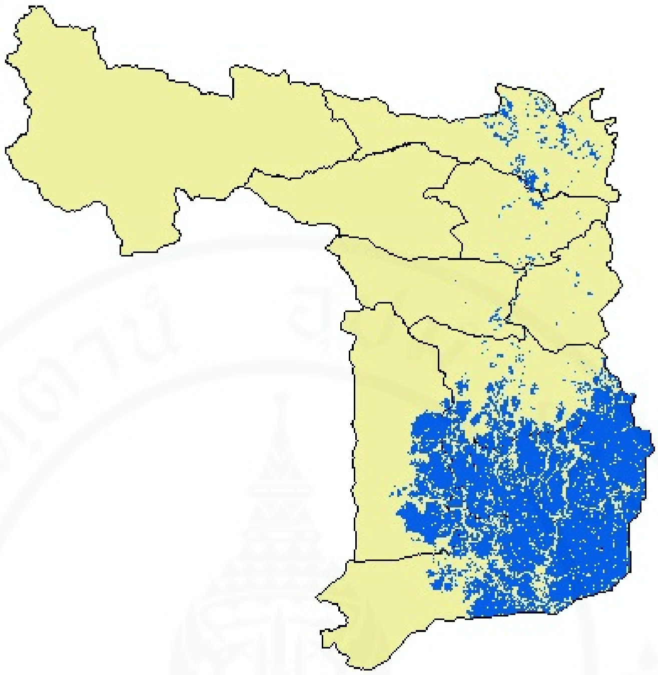

Figure B.47 Flood area size 1009.48 km^2

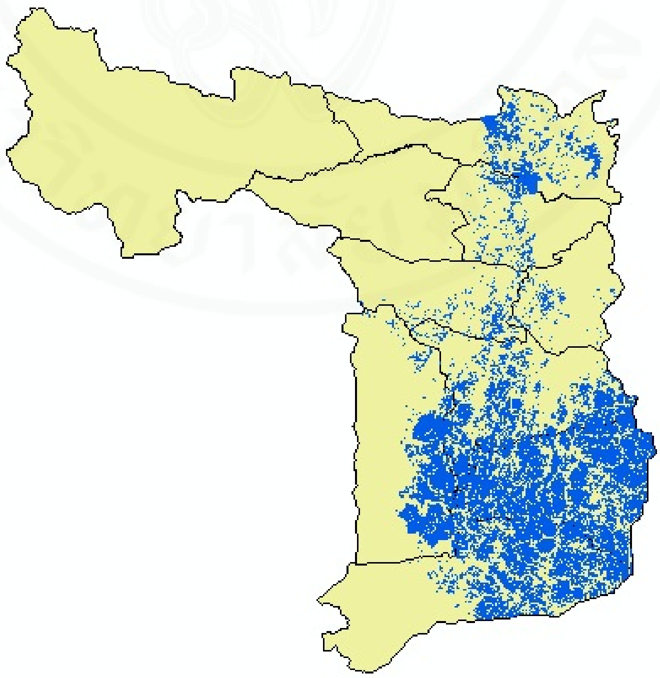

Figure B.48 Flood area size 1026.94 km^2

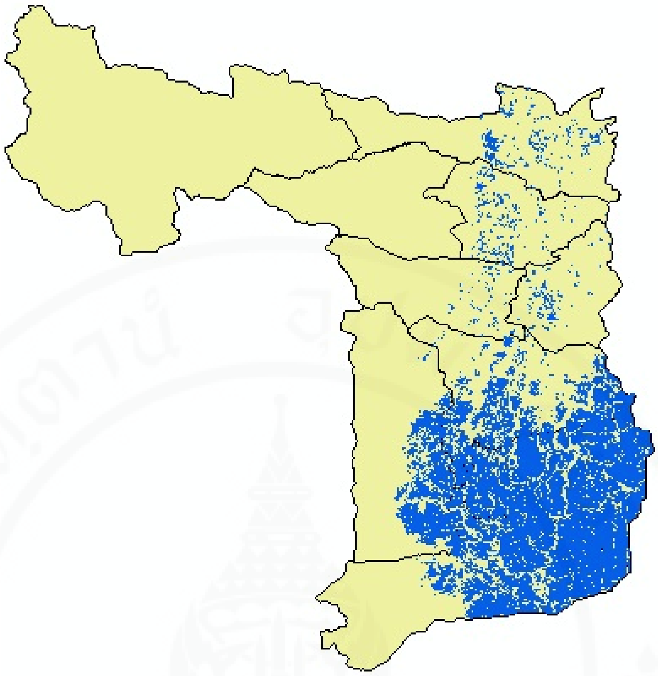

Figure B.49 Flood area size 1094.14 km^2

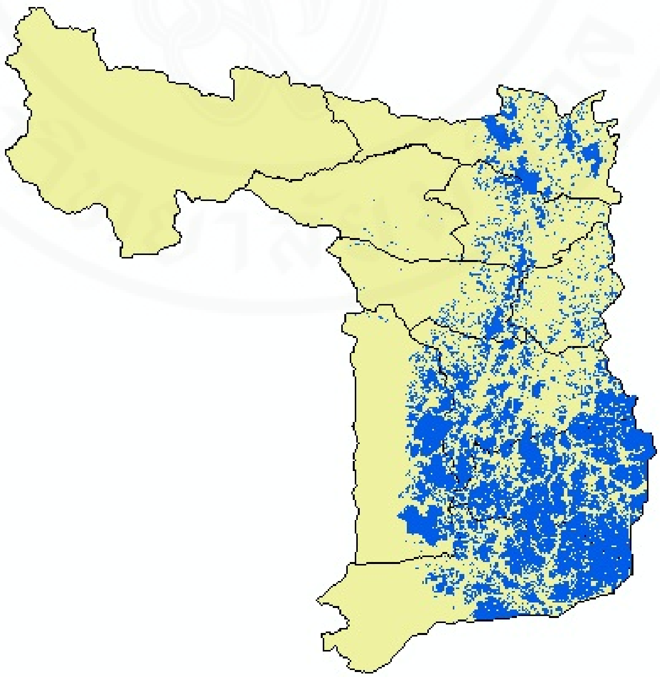

Figure B.50 Flood area size 1103 km^2

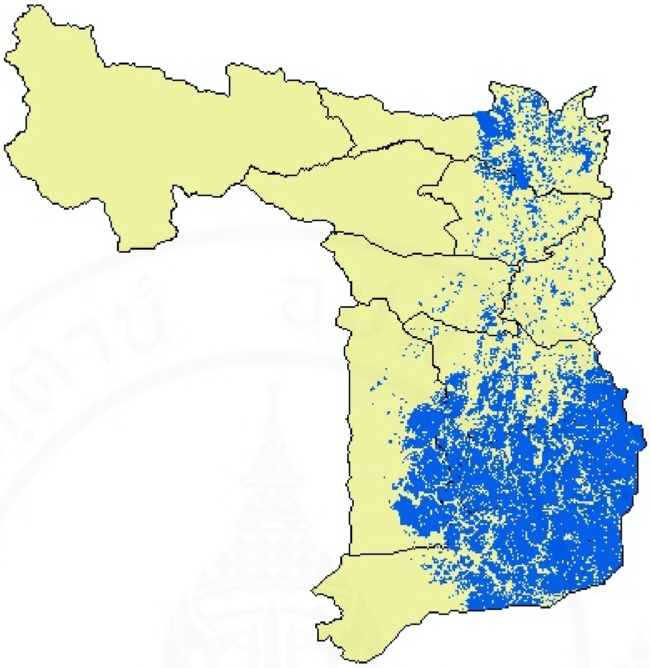

Figure B.51 Flood area size 1115.77 km^2

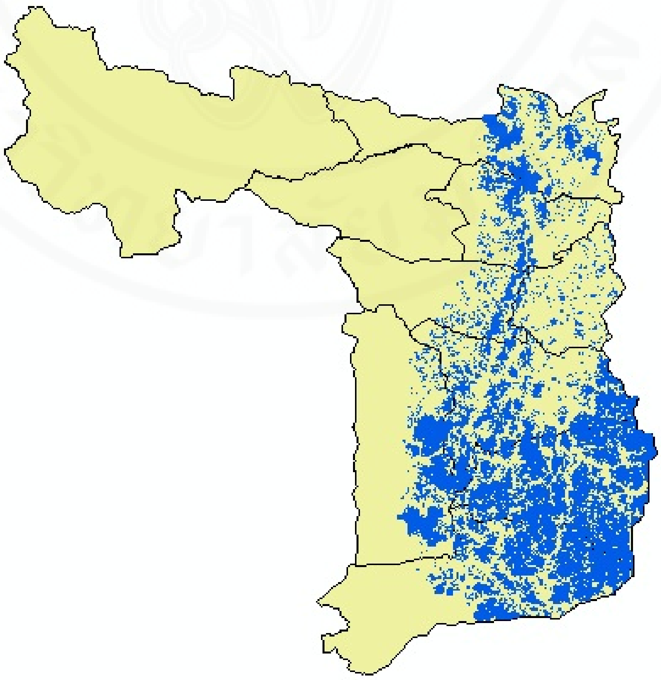

Figure B.52 Flood area size 1122.73 km^2

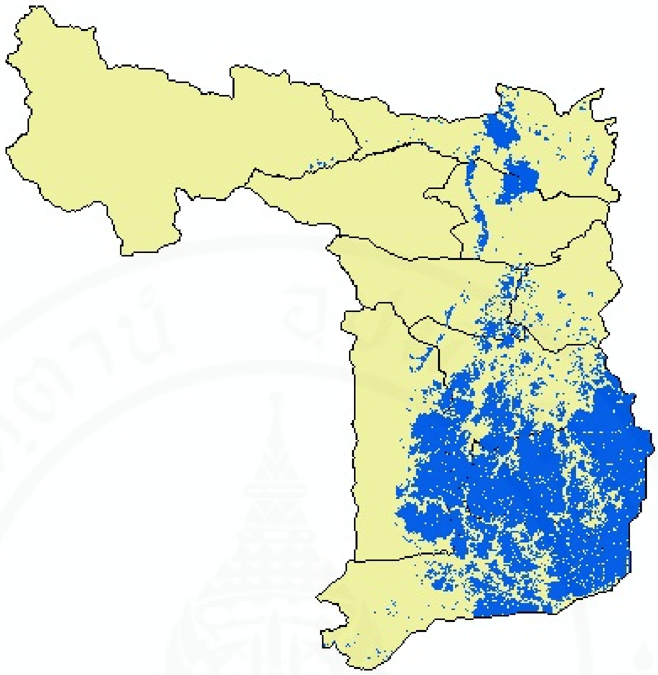

Figure B.53 Flood area size 1153.87 km^2

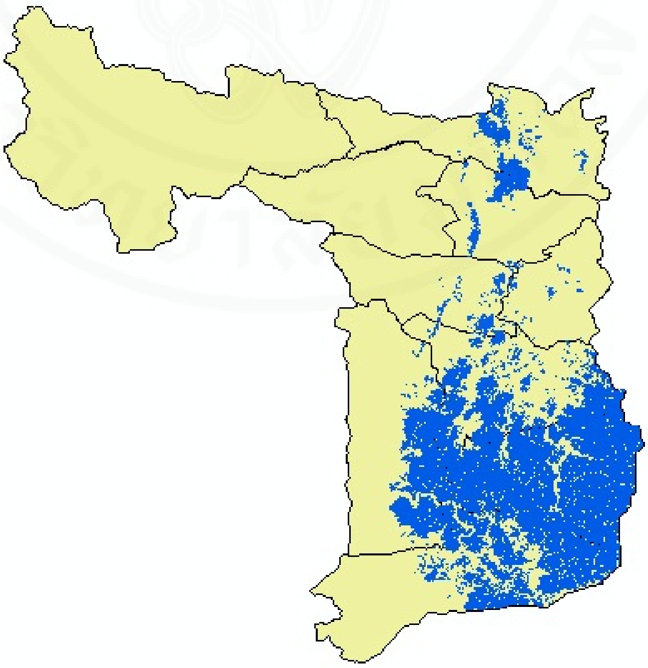

Figure B.54 Flood area size 1166.25 km^2

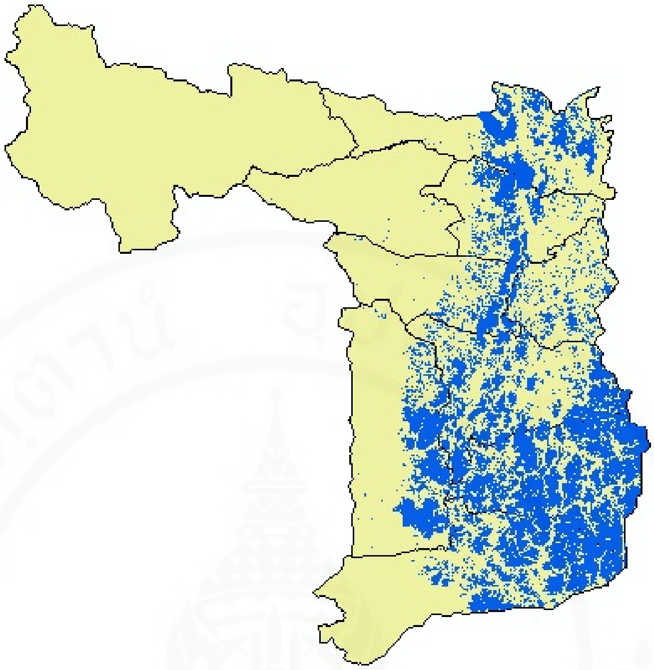

Figure B.55 Flood area size 1195.35 km^2

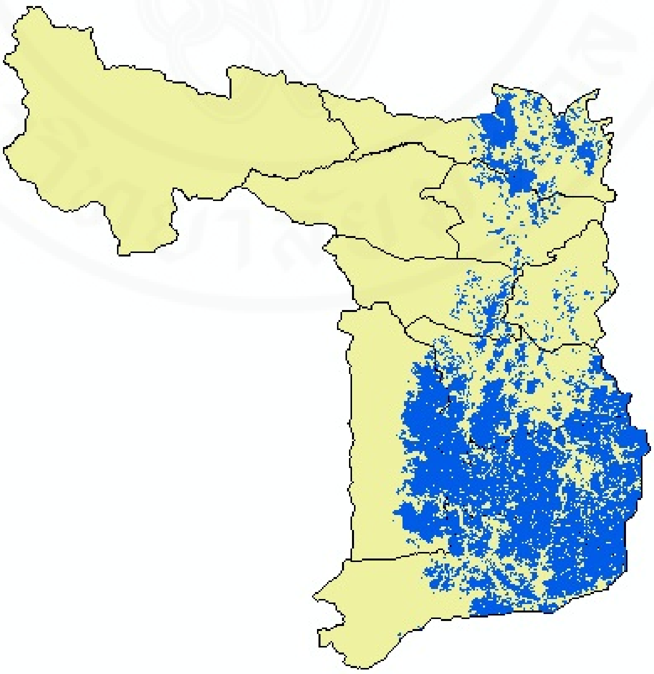

Figure B.56 Flood area size 1222.15 km^2

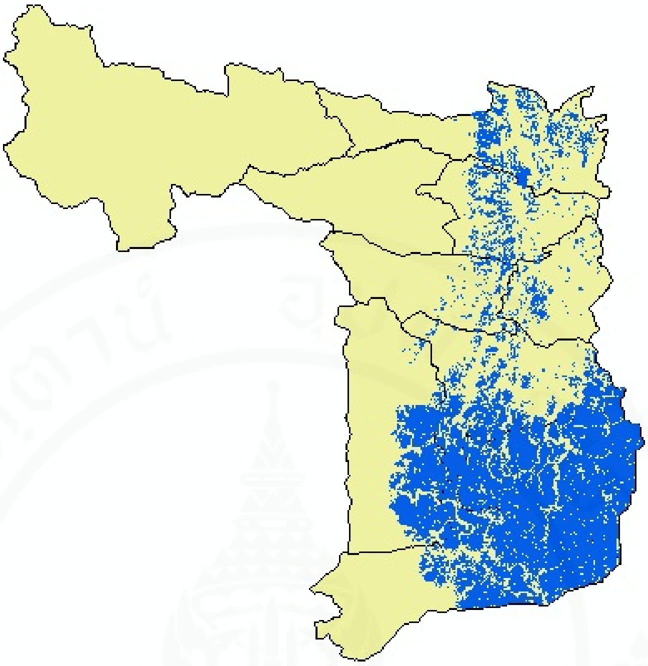

Figure B.57 Flood area size 1262.1 km^2

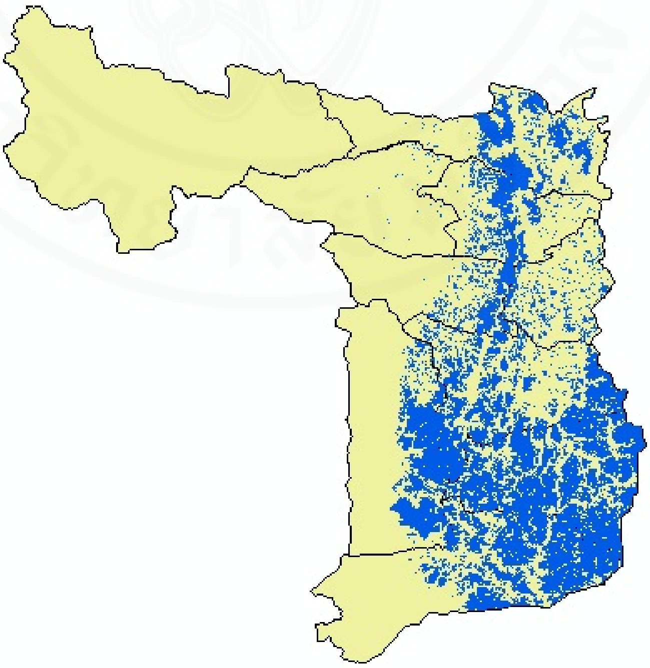

Figure B.58 Flood area size 1265.24 km^2

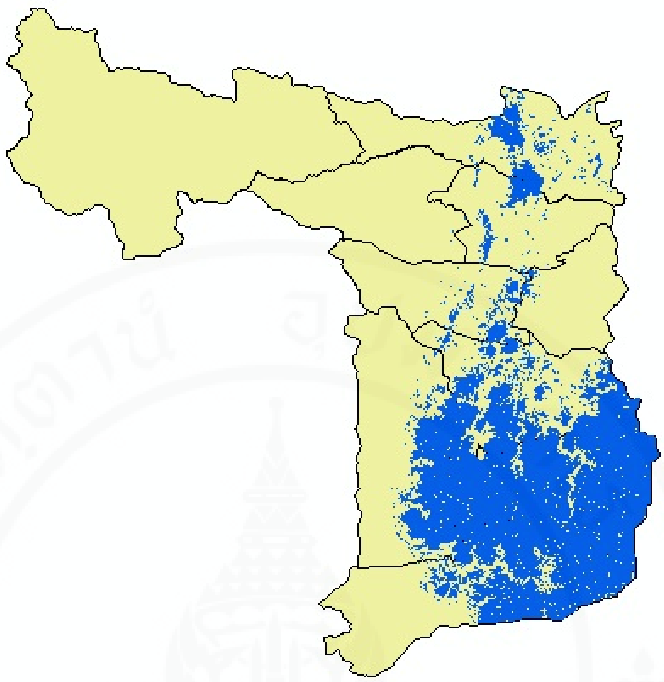

Figure B.59 Flood area size 1326.19 km^2

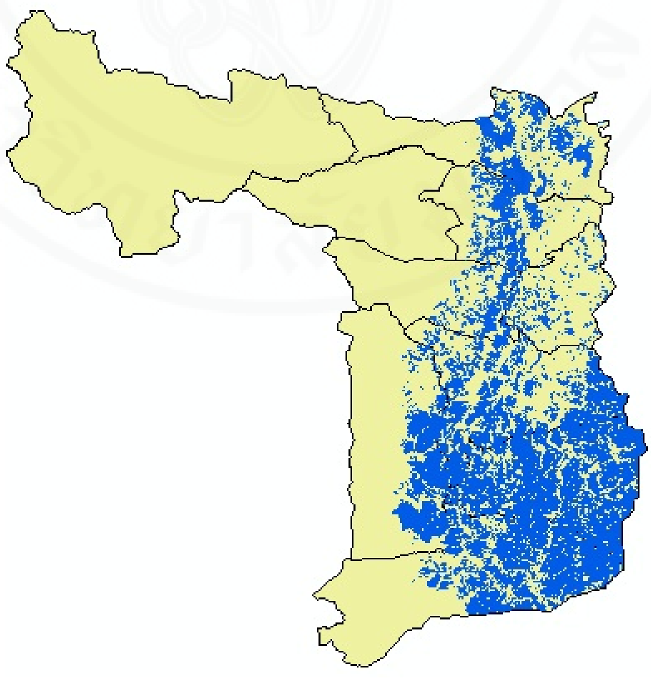

Figure B.60 Flood area size 1346.98 km^2

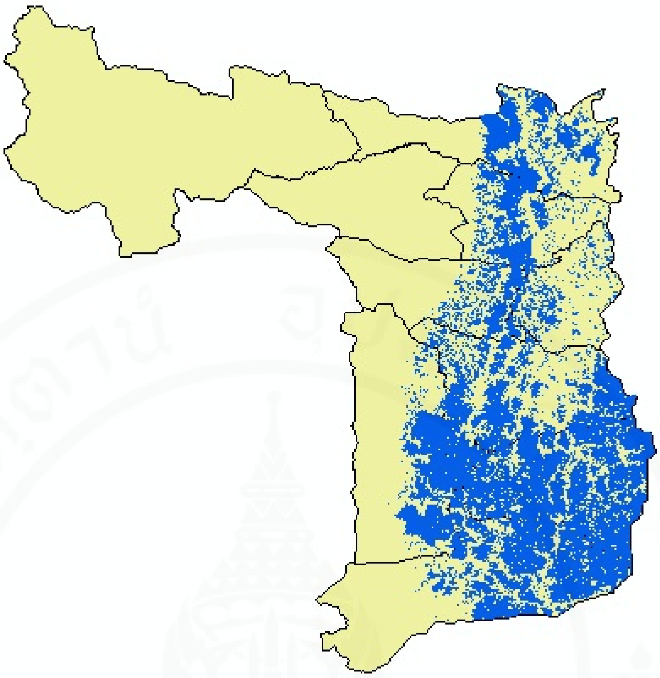

Figure B.61 Flood area size 1421.1 km^2

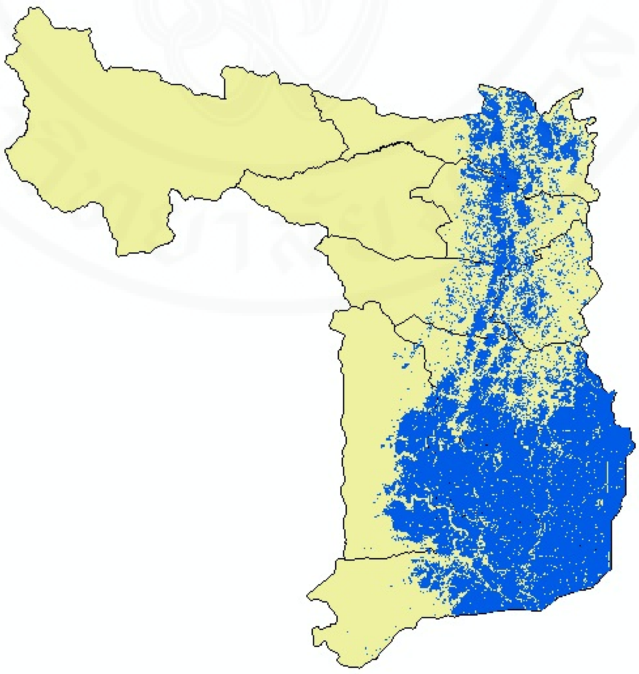

Figure B.62 Flood area size 1577.71 km^2

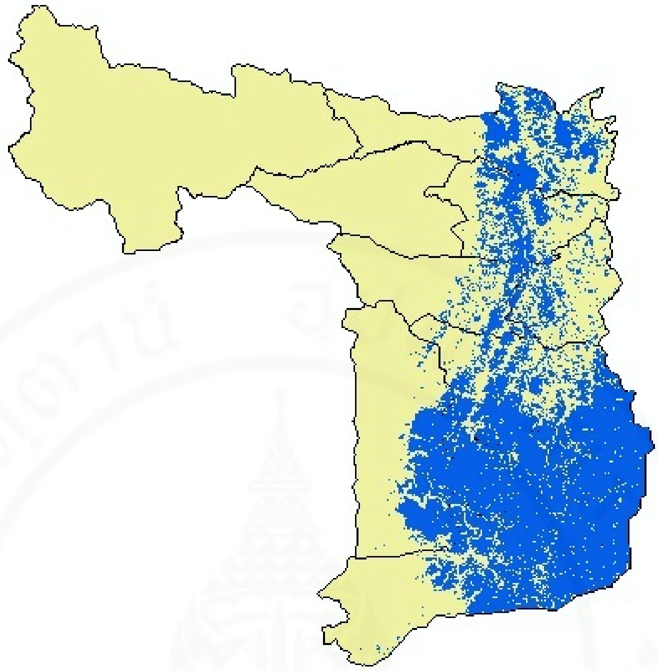

Figure B.63 Flood area size 1599.6 km^2

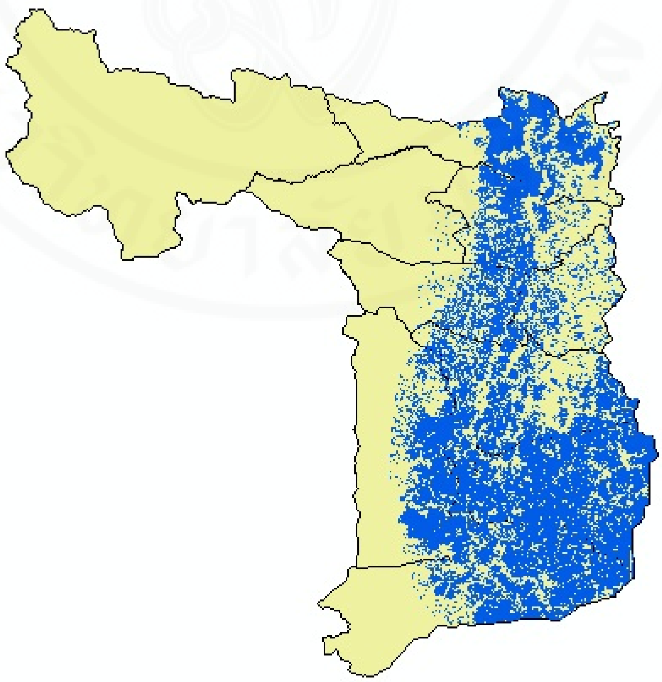

Figure B.64 Flood area size 1635.7 km^2

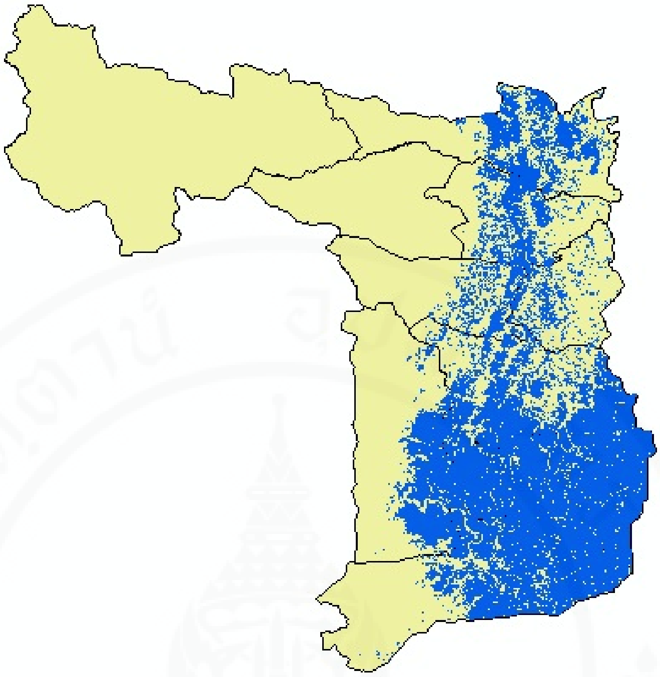

Figure B.65 Flood area size 1659.44 km^2

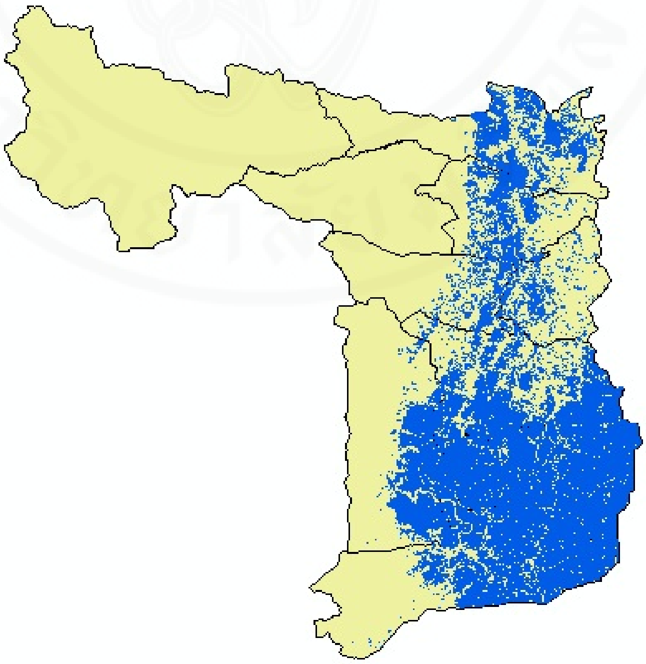

Figure B.66 Flood area size 1666.54 km^2

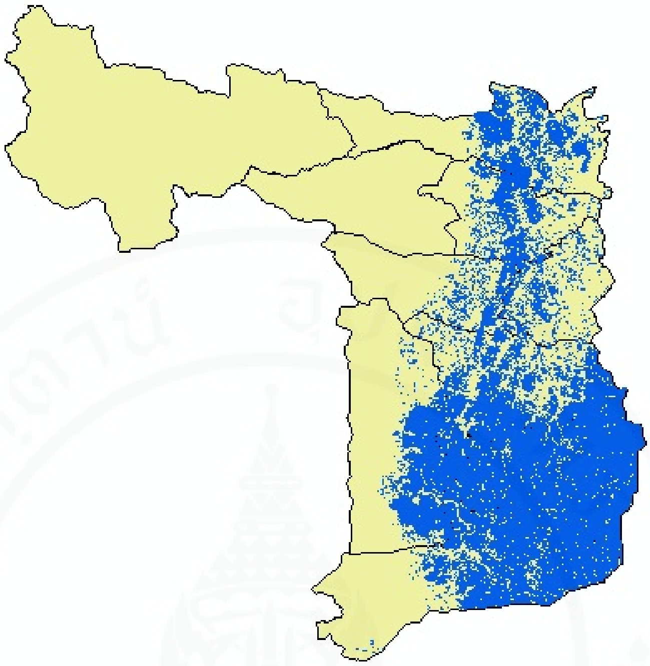

Figure B.67 Flood area size 1691.03 km^2

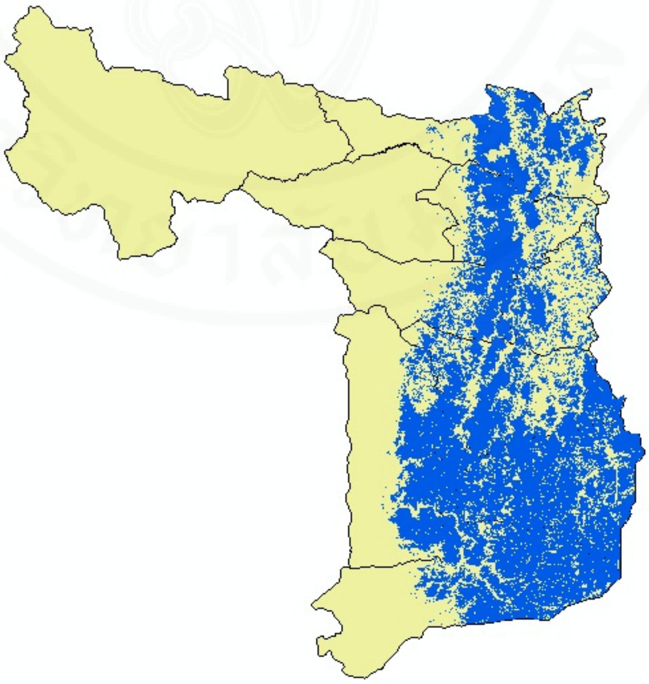

Figure B.68 Flood area size 1758.17 km^2

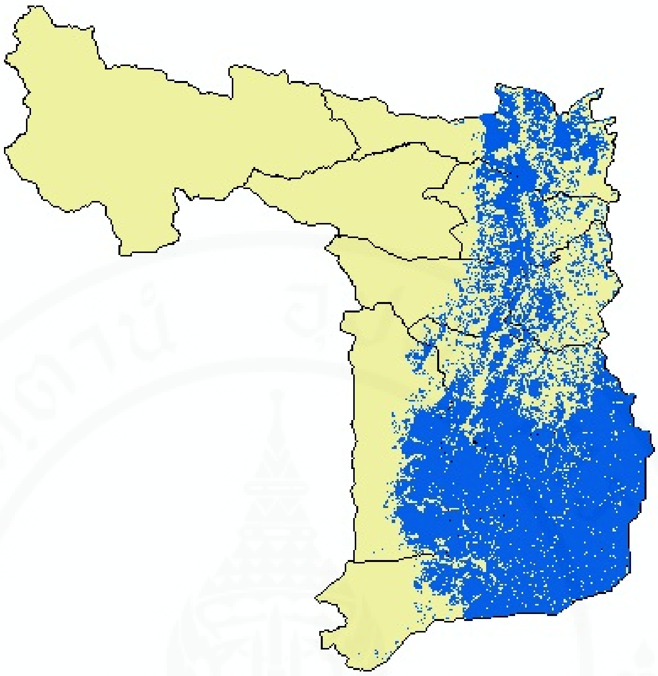

Figure B.69 Flood area size 1771.36 km^2

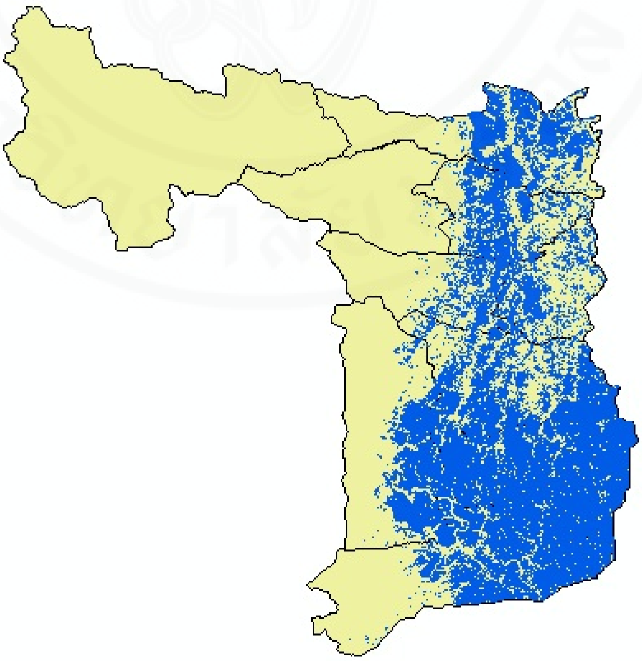

Figure B.70 Flood area size 1806.3 km^2

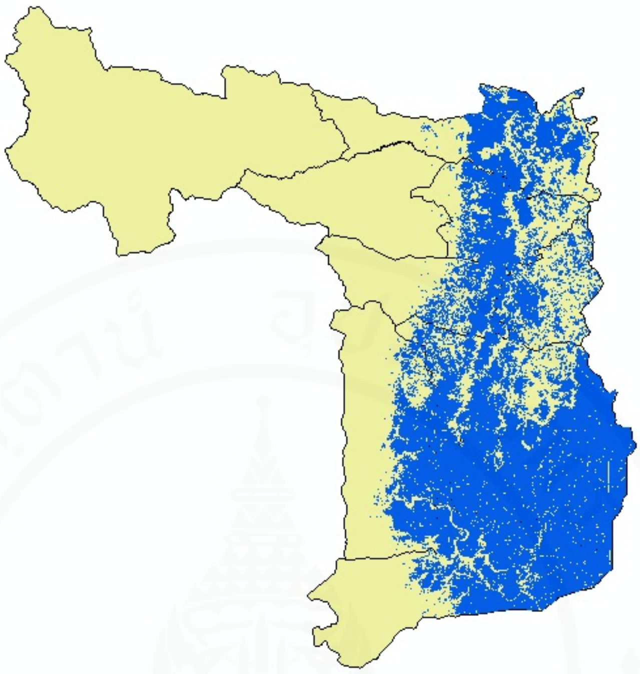

Figure B.71 Flood area size 1904.96 km^2

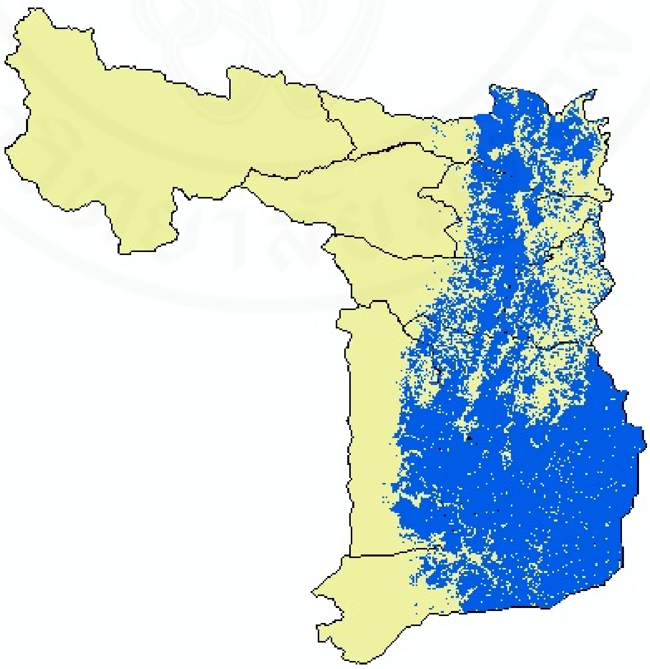

Figure B.72 Flood area size 1914.82 km^2

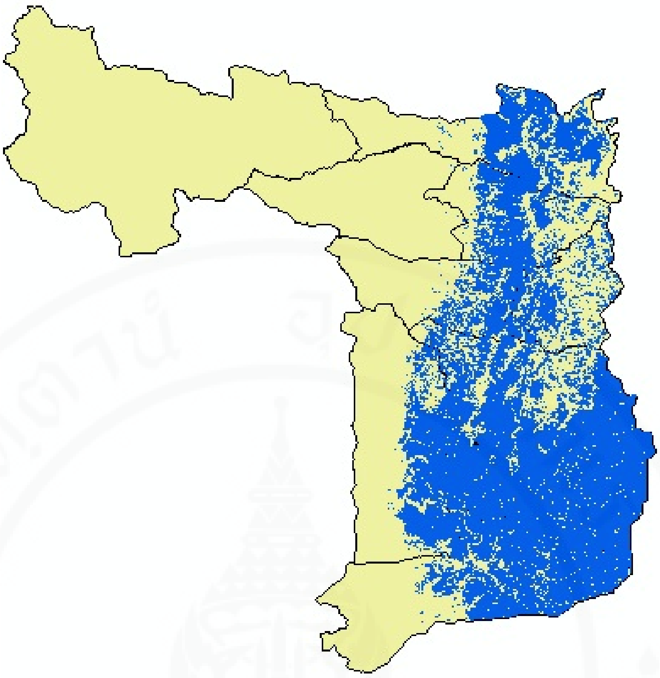

Figure B.73 Flood area size 1916.57 km^2

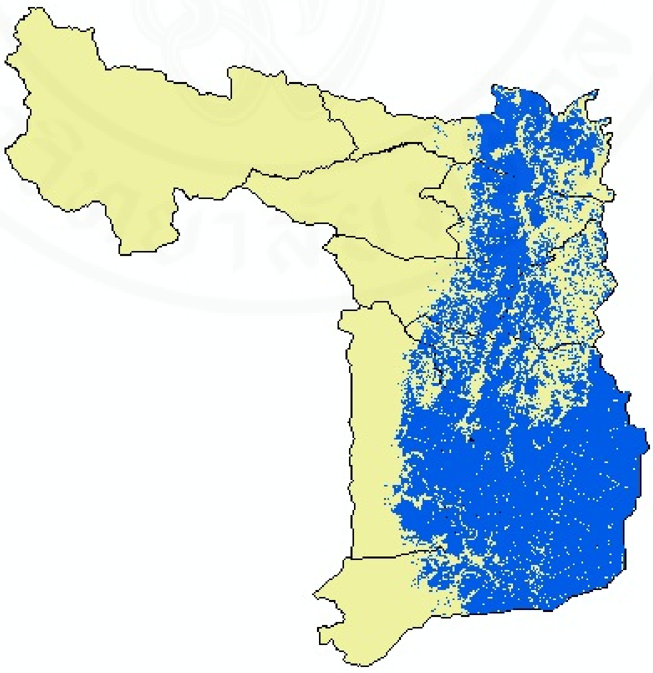

Figure B.74 Flood area size 1918.31 km^2

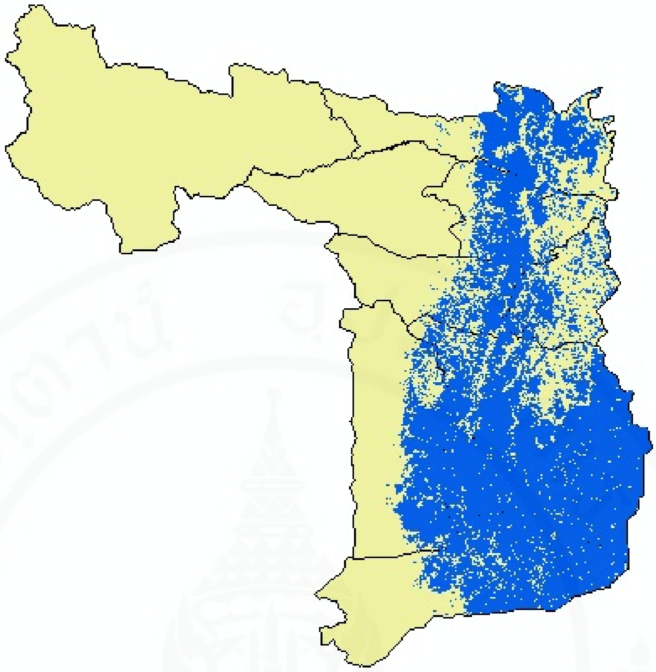

Figure B.75 Flood area size 1937.83 km^2

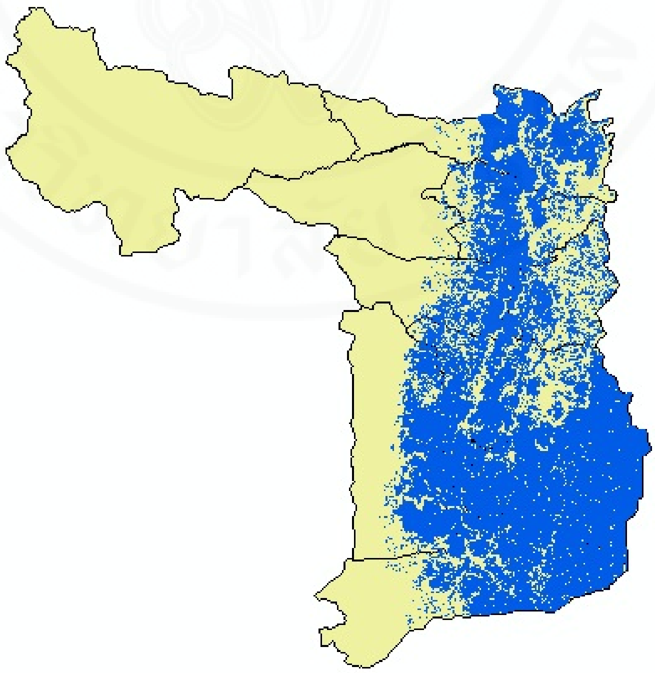

Figure B.76 Flood area size 2017.92 km^2

BIOGRAPHY

NAME	Mr. Siripong Weranantawat
DATE OF BIRTH	11 September 1984
PLACE OF BIRTH	Bangkok, Thailand
INSTITUTIONS ATTENDED	Rangsit University, 2003-2007
	Bachelor of Engineering
	(Computer Engineer)
	Mahidol University, 2008-2012
	Master of Science (Technology of Information System Management)
HOME ADDRESS	100/285 soi Senanikom 1 Phaholyothin Road, Ladprao, Bangkok 10230
	Tel 087-094-2600
	Email: siripong_wera@hotmail.com