

STUDY OF SPATIAL REGRESSION MODEL TO PREDICT SALINITY
IN SALINE SOIL AREAS OF NAKHON RATCHASIMA PROVINCE
USING REMOTE SENSING DATA AND GEOGRAPHIC INFORMATION
SYSTEM



A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of
Master of Engineering in Civil, Transportation and Geo-Resources Engineering
Suranaree University of Technology
Academic Year 2022

การศึกษาแบบจำลองถดถอยเชิงพื้นที่เพื่อคาดการณ์ปริมาณความเค็ม
ในพื้นที่ดินเค็มของจังหวัดนครราชสีมา
ด้วยการใช้ข้อมูลการรับรู้จากระยะไกลและระบบสารสนเทศภูมิศาสตร์

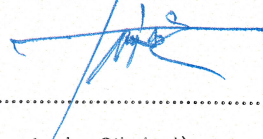


วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมโยธา ขนส่ง และทรัพยากรธรณี
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2565

STUDY OF SPATIAL REGRESSION MODELS TO PREDICT SALINITY IN
SALINE SOIL AREAS OF NAKHON RATCHASIMA PROVINCE USING REMOTE
SENSING DATA AND GEOGRAPHIC INFORMATION SYSTEM

Suranaree University of Technology has approved this thesis submitted in
partial fulfillment of the requirements for a Master's Degree.

Thesis Examining Committee



(Dr. Tawisak Silakul)

Chairperson



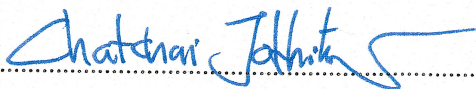
(Asst. Prof. Dr. Bantita Terakulsatit)

Member (Thesis Advisor)



(Asst. Prof. Dr. Prachya Tepnarong)

Member



(Assoc. Prof. Dr. Chatchai Jothityangkoon)

Vice Rector for Academic Affairs and
Quality Assurance



(Assoc. Prof. Dr. Pornsiri Jongkol)

Dean of Institute of Engineering

วชิรินทร์ เพิ่มผล: การศึกษาแบบจำลองถดถอยเชิงพื้นที่เพื่อคาดการณ์ปริมาณความเค็มในพื้นที่ดินเค็มของจังหวัดนครราชสีมาด้วยการใช้ข้อมูลการรับรู้จากระยะไกล และระบบสารสนเทศภูมิศาสตร์ (STUDY OF SPATIAL REGRESSION MODEL TO PREDICT SALINITY IN SALINE SOIL AREAS OF NAKHON RATCHASIMA PROVINCE USING REMOTE SENSING DATA AND GEOGRAPHIC INFORMATION SYSTEM) อาจารย์ที่ปรึกษา: ผู้ช่วยศาสตราจารย์ ดร. บัณฑิตา ธีรกุลสถิตย์, 140 หน้า.

คำสำคัญ: การประเมินศักยภาพของน้ำบาดาล/การถดถอยกำลังสองสูงที่สุด/การถดถอยแบบถ่วงน้ำหนักทางภูมิศาสตร์/น้ำบาดาลเค็ม/การวิเคราะห์ความสัมพันธ์เชิงพื้นที่

วัตถุประสงค์ของการศึกษานี้คือเพื่อประเมินการใช้แบบจำลองการถดถอยกำลังสองน้อยที่สุด (OLS) และแบบจำลองการถดถอยแบบถ่วงน้ำหนักทางภูมิศาสตร์ (GWR) ในการทำแผนที่ปริมาณความเค็ม ในเขตอำเภอโนนไทย อำเภอโนนสูง อำเภอโนนแดง อำเภอดง และอำเภอขามสะแกแสง จังหวัดนครราชสีมา ซึ่งพิจารณาถึงปัจจัยต่าง ๆ 13 ตัวแปร ได้แก่ ความลาดชัน ธรณีวิทยา ธรณีสัณฐานวิทยา ปริมาณน้ำฝน ชนิดดิน ความหนาแน่นของทางน้ำ สิ่งปกคลุมดินและการใช้ประโยชน์ที่ดิน ปริมาณคลอไรด์ ปริมาณของแข็งที่ละลายได้ทั้งหมด อุทกธรณีวิทยา ศักยภาพน้ำบาดาล ความหนาแน่นของโครงสร้างเชิงเส้น และดัชนีความแตกต่างของพืชพรรณ ผลการวิเคราะห์แสดงให้เห็นว่าแบบจำลองการถดถอยแบบถ่วงน้ำหนักทางภูมิศาสตร์ (GWR) สามารถอธิบายข้อมูลได้ดีกว่าแบบจำลองการถดถอยกำลังสองที่น้อยที่สุด (OLS) ซึ่งแสดงให้เห็นด้วยค่าสัมประสิทธิ์การตัดสินใจ (R^2) ที่ถูกปรับแก้แล้วเท่ากับ 0.995865 และค่าข้อสารสนเทศของอาโคเคะ (AICc) ที่ถูกปรับแก้แล้วเท่ากับ 9,801.877000 โดยพบว่าตัวแปรที่มีนัยสำคัญทางสถิติ จำนวน 11 ตัวแปร แบ่งเป็นตัวแปรที่มีความสัมพันธ์ทางลบ จำนวน 5 ตัวแปร ได้แก่ ดัชนีความแตกต่างของพืชพรรณ (NDVI) ศักยภาพน้ำบาดาล (GP) ความลาดชัน (S1) ธรณีสัณฐานวิทยา (Lf1) และสิ่งปกคลุมดินและการใช้ประโยชน์ที่ดิน (LULC1) และแบ่งเป็นตัวแปรที่มีความสัมพันธ์ทางลบ จำนวน 6 ตัวแปร ได้แก่ ความหนาแน่นของโครงสร้างเชิงเส้น (DD1) ปริมาณน้ำฝน (RF) ปริมาณคลอไรด์ (CC1) ปริมาณของแข็งที่ละลายได้ทั้งหมด (TDS1) ชนิดดิน (STC1) และธรณีวิทยา (Geol1)

สาขาวิชา เทคโนโลยีธรณี
ปีการศึกษา 2565

ลายมือชื่อนักศึกษา... วชิรินทร์ เพิ่มผล
ลายมือชื่ออาจารย์ที่ปรึกษา... บัณฑิตา ธีรกุลสถิตย์

WATCHARIN PHOEMPHON: STUDY OF SPATIAL REGRESSION MODEL TO PREDICT SALINITY IN SALINE SOIL AREAS OF NAKHON RATCHASIMA PROVINCE USING REMOTE SENSING DATA AND GEOGRAPHIC INFORMATION SYSTEM. THESIS ADVISOR: ASST. PROF. BANTITA TERAKULSATIT, Ph.D., 140 PP.

Keyword: GROUNDWATER POTENTIAL ASSESSMENT/ ORDINARY LEAST SQUARES REGRESSION (OLS)/ GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)/ GROUNDWATER SALINITY/ SPATIAL RELATIONSHIPS ANALYSIS

The main objective of this study is to assess the application of the ordinary least squares regression model and geographically weighted regression model for mapping the salinity content of Non Thai, Non Sung, Non Daeng, Khong, and Kham Sakae Saeng districts in Nakhon Ratchasima province. The analysis considers thirteen factors: slope, geology, landforms, annual rainfall, soil texture class, drainage density, land use/land cover, chloride content, total dissolved solids, hydrogeological unit, groundwater potential, lineament density, and normalized difference vegetation index. The results demonstrate the superiority of the GWR model over the OLS model, indicated by a higher adjusted R^2 (0.995865) and a lower corrected Akaike Information Criterion (AICc) (9,801.877000). The findings reveal significant associations between salinity content and various factors. Negative associations are five factors including normalized difference vegetation index (NDVI), groundwater potential (GP), soil texture class (Sl1), landforms (Lf1), and land use/land cover (LULC1). Positive associations are six factors including drainage density (DD1), rainfall (RF), clay content (CC1), total dissolved solids (TDS1), soil texture class (STC1), and geology (Geol1).

School of Geotechnology
Academic Year 2022

Student's Signature watcharin phoemphon
Advisor's Signature Bantita Terakulsatit

ACKNOWLEDGEMENTS

First and foremost, I am deeply grateful to my supervisor, Assistant Professor Dr. Bantita Terakulsatit, for her invaluable guidance, expertise, and unwavering support throughout this research endeavor.

I sincerely appreciate the Geo-Informatics and Space Technology Development Agency (GISTDA) for granting me access to the GISTDA Portal. This access played a vital role in acquiring crucial satellite imagery and geospatial data for my study.

I would like to express my heartfelt thanks to Dr. Siripon Kamontum for her valuable advice, insights, and assistance in the field of Geographic Information Systems (GIS).

Special thanks go to Ms. Wanida Thongwat and Mr. Thanet Phuphulb for their generosity in providing the discharge well data, which significantly contributed to the success of this research.

I would like to extend my appreciation to my family, friends, and colleagues for their unwavering support, encouragement, and understanding throughout this academic journey.

Finally, I am grateful to all the individuals and organizations whose direct or indirect contributions have made this research possible.

Watcharin Phoemphon

TABLE OF CONTENTS

	Page
ABSTRACT (THAI).....	I
ABSTRACT (ENGLISH).....	II
ACKNOWLEDGEMENTS.....	III
TABLE OF CONTENTS.....	IV
LIST OF TABLES.....	VIII
LIST OF FIGURES.....	IX
SYMBOLS AND ABBREVIATIONS.....	XI
CHAPTER	
I INTRODUCTION	
1.1 Background and Rationale.....	1
1.2 Research Objectives.....	2
1.3 Scope and Limitations.....	2
1.4 Research Methodology.....	3
1.4.1 Literature Review.....	4
1.4.2 Data Collection and Preparation.....	4
1.4.3 Regression Analysis.....	4
1.4.4 Comparison between OLS Regression and GWR Models.....	5
1.4.5 Discussions and Conclusions.....	5
1.4.6 Thesis Writing.....	5
1.5 Thesis Contents.....	6
II LITERATURE REVIEW	
2.1 Geology of Khorat Plateau.....	7

TABLE OF CONTENTS (Continued)

	Page
2.1.1 Lower Khorat Unit.....	8
2.1.2 Upper Khorat Unit.....	11
2.2 Lithostratigraphy of the Maha Sarakham Formation.....	15
2.2.1 Lower Member.....	15
2.2.2 Middle Member.....	17
2.2.3 Upper Member.....	18
2.3 Hydrogeology of Nakhon Ratchasima.....	19
2.3.1 Unconsolidated Aquifer.....	19
2.3.2 Consolidated Aquifer.....	20
2.4 Groundwater Quality of Nakhon Ratchasima.....	22
2.5 Saline Soil of Nakhon Ratchasima.....	24
2.6 Thematic Layers.....	26
2.6.1 Slope.....	26
2.6.2 Drainage Density.....	26
2.6.3 Lineament Density.....	26
2.6.4 Landform.....	27
2.6.5 Annual Rainfall.....	27
2.6.6 Geology.....	28
2.6.7 Hydrogeological Unit.....	28
2.6.8 Groundwater Potential.....	28
2.6.9 Soil Texture Class.....	29
2.6.10 Land Use/Land Cover.....	29
2.6.11 Normalized Difference Vegetation Index.....	29
2.6.12 Salinity Content.....	30
2.6.13 Chloride Content.....	30
2.6.14 Total Dissolve Solid.....	30

TABLE OF CONTENTS (Continued)

	Page
2.7 Regression Analysis.....	31
2.7.1 Ordinary Least Square Regression.....	32
2.7.2 Geographically Weighted Regression.....	34
2.7.3 Weighting Functions.....	35
2.7.4 Bandwidth Selection.....	36
2.7.5 Model Calibration and Diagnosis.....	37
III RESEARCH METHODOLOGY	
3.1 Study Area.....	40
3.2 Data Collection and Preparation.....	40
3.2.1 Dependent Variable.....	43
3.2.2 Independent Variable.....	43
3.2.3 Data Preparation.....	58
3.3 Regression Analysis Model Development.....	59
3.3.1 Exploratory Spatial Data Analysis (ESDA).....	59
3.3.2 Fit an OLS Model and Perform Diagnostics.....	62
3.3.3 Geographically Weighted Regression.....	64
3.3.4 Compare Model.....	65
IV RESULTS AND DISCUSSION	
4.1 Results of Global Model: Ordinary Least Squares (OLS) Regression Analysis.....	67
4.2 Results of Local Model: Geographically Weighted Regression (GWR) Analysis.....	73
4.3 Comparison of OLS and GWR Models.....	78
4.4 Discussion.....	81

TABLE OF CONTENTS (Continued)

	Page
V CONCLUSIONS AND RECOMMENDATION	
5.1 Conclusions.....	83
5.2 Recommendations for Further Research.....	84
REFERENCES.....	86
APPENDIX.....	96
APPENDIX A Chemical Properties.....	97
APPENDIX B Google Earth Engine Metadata and Scripts.....	101
APPENDIX C Histogram of Distribution of the Thematic Layers.....	108
APPENDIX D Results of Running Script.....	121
BIOGRAPHY.....	140

LIST OF TABLES

Table	Page
3.1 The data type and data sources of thematic layers.....	42
3.2 Classification of thematic layers.....	53
4.1 Ordinary least squares (OLS) regression coefficient estimate results.....	69
4.2 OLS model diagnostics.....	70
4.3 Summary Statistics of the local model.....	73
4.4 Geographically weighted regression (GWR) ANOVA results.....	78
4.5 Comparison of the global model (OLS) and local model (GWR).....	80
A1 Chemical properties of groundwater samples in May 2019.....	96
B1 Bands of Soil texture class (USDA system).....	100
B2 Bands Class Table of Soil texture class (USDA system).....	100
B3 Bands of Global SRTM Landforms.....	102
B4 Bands Class Table of Global SRTM Landforms.....	102
B5 Land Cover Class Sentinel-2 10m Land Use/Land Cover time-series.....	104

LIST OF FIGURES

Figure	Page
1.1 Flowchart of the research methodology.....	3
2.1 The Khorat-Ubol basin and the Udon-Sakon Nakhon basin (modified from El Tabakh et al., 1999).....	8
2.2 (A) Stratigraphy of the Khorat Basin, and (B) lithostratigraphy of the Maha Sarakham Formation (modified from El Tabakh et al. (1999) and Hansen et al. (2016)).....	13
2.3 Stratigraphic section of the High Terrace at Ban Phu Khao Thong, Mueang district, Nakhon Ratchasima. (After Tatong and Margane, 2004).....	14
2.4 Scatterplots: (A) positive relationship, (B) negative relationship, and (C) two variables are unrelated.....	32
3.1 The study area of saline soils area in Nakhon Ratchasima, Thailand.....	41
3.2 Salinity content.....	44
3.3 Chloride content.....	44
3.4 Total dissolve solid.....	45
3.5 Slope.....	45
3.6 Drainage density.....	47
3.7 Lineament density.....	47
3.8 Landforms.....	48
3.9 Soil texture class.....	48
3.10 Land use /land cover.....	49
3.11 Normalized difference vegetation index.....	50
3.12 Annual rainfall.....	51
3.13 Geology.....	52
3.14 Hydrogeological Unit.....	52

LIST OF FIGURES (Continued)

Figure	Page
3.15 Groundwater Potential.....	53
3.16 Grid cell size 1,000 m x 1,000 m.....	58
3.17 Converting polygon features to point features.....	59
3.18 Show original values are transformed to be closer to a normal distribution....	61
4.1 Standardized residual of the global model (OLS).....	71
4.2 Summary of global spatial autocorrelation (Moran's I).....	72
4.3 Standardized residual of the local model (GWR).....	74
4.4 Local R ² of the local model (GWR).....	76
4.5 Summary of local spatial autocorrelation (Moran's I).....	77
C1 Histogram plot of Salinity Content (SC) data and summary statistics.....	107
C2 Histogram plot of Chloride Content (CC) data and summary statistics.....	108
C3 Histogram plot of Total Solid Content (TDS) data and summary statistics.....	109
C4 Histogram plot of Slope (SI) data and summary statistics.....	110
C5 Histogram plot of Drainage Density (DD) data and summary statistics.....	111
C6 Histogram plot of Lineament Density (LD) data and summary statistics.....	112
C7 Histogram plot of Landforms (Lf) data and summary statistics.....	113
C8 Histogram plot of Soil Texture Class (STC) data and summary statistics.....	114
C9 Histogram plot of Land use/ Land cover (LULC) data and summary statistics.....	115
C10 Histogram plot of Normalized Differentiation Vegetation Index (NDVI) data and summary statistics.....	120
C11 Histogram plot of Mean Annual Rainfall (RF) data and summary statistics.....	116
C12 Histogram plot of Geology (Geol) data and summary statistics.....	117
C13 Histogram plot of Hydrogeological Unit (HU) data and summary statistics.....	122
C14 Histogram plot of Groundwater Potential (GP) data and summary statistics.....	118

SYMBOLS AND ABBREVIATIONS

$^{\circ}\text{C}$	=	Degrees Celsius
$d_i(u)$	=	Distance between observation i and the location u
h	=	Bandwidth
n	=	Number of observations
S	=	Hat matrix
T	=	Transpose of the matrix
$\text{tr}(S)$	=	Trace of the hat matrix
W	=	Square matrix
$W(u)$	=	Square matrix of weights at location u
$w_i(u)$	=	Geographical weight of observation i at location u
X	=	Design matrix that contains the values of the independent variables and a column of 1s
x'	=	Transformed value
x_i	=	Independent variable
y	=	Vector of observed values
\hat{y}_i	=	Predicted or fitted value for observation i
y_i	=	Dependent variable
$\hat{\beta}$	=	Vector of estimated parameters
β_i	=	Estimated parameters
$\beta_{i(u)}$	=	Estimated parameters at location u
ε_i	=	Error term
λ_1	=	Power (Exponent) parameter
λ_2	=	Shift parameter
$\hat{\sigma}$	=	Estimated standard deviation of the residuals
Adj.	=	Adjust

SYMBOLS AND ABBREVIATIONS (Continued)

API	=	Application Programming Interface
ASTER	=	The Advanced Spaceborne Thermal Emission and Reflection Radiometer
B.P.	=	Years Before Present
DF	=	Degrees of Freedom
E	=	East Coordinate
ESRI	=	The Environmental Systems Research Institute
F	=	F-Statistic
F-Stat	=	Joint F-Statistic
GDEM	=	Global Digital Elevation Model
iid	=	Independent and identically distributed
JB	=	Jarque-Bera Statistic
K(BP)	=	Koenke (BP) Statistic
K/Ar	=	Potassium-Argon Dating
MS	=	Mean Sum of Squares
m ³ /h	=	Cubic Meters Per Hour
N	=	North Coordinate
PPM	=	Parts-Per-Million
PPT	=	Parts-Per-Trillion
Pr/ Prob	=	Probability
SE	=	Standard Error
SS	=	Sum of Squares
SRTM	=	The Shuttle Radar Topography Mission
Std.	=	Standard
USDA	=	The United States Department of Agriculture
UTM	=	Universal Transverse Mercator
VIF	=	Variance Inflation Factor

SYMBOLS AND ABBREVIATIONS (Continued)

Wald	=	Joint Wald Statistic
WGS	=	World Geodetic System
XML	=	Extensible Markup Language



CHAPTER I

INTRODUCTION

1.1 Background and Rationale

Nakhon Ratchasima is located in northeastern Thailand, in the Khorat-Ubol Basin of the Khorat Plateau. Saline soil can be found in many parts of Thailand, and the source of the salinity might vary depending on the region's geological features. Human activity can contribute to soil and water salinity issues (Peck & Hatton, 2003). The Maha Sarakham Formation, which is mainly composed of interbedded rock salt layers, underlies most of northeast Thailand and provides the region's primary source of salinity (Wongsomsak, 1986). For regional groundwater exploration conventional methods involving geological, hydrogeological, and geophysical techniques, are costly due to the expenses associated with drilling, time-consuming procedures, and their complexity (Ndatuwong & Yadav, 2014). Meanwhile, Remote Sensing (RS) and Geographic Information Systems (GIS) are employed to identify the locations of potential zones. A Geographically Weighted Regression (GWR) model takes into account the spatial variability of independent and dependent variables (Fotheringham, Brunson, & Charlton, 2002) by evaluating correlations between observed points based on distance GWR models outperform the Ordinary Least Squares (OLS) Regression model in prediction for data with spatial heterogeneity (Brown et al., 2012; Pratt & Chang, 2012; Tu & Xia, 2008). The GWR has been widely used in hydrology to evaluate spatial correlations between various environmental characteristics, such as the relationship between land use categories and water quality (Pratt & Chang, 2012; Tu, 2011, 2013; Tu & Xia, 2008). Furthermore, groundwater salinity can be predicted using GWR in places where surface displacement measurements are unavailable or insufficient for groundwater quality mapping. The method is gaining popularity and has

been utilized in the research of economic, social, and environmental issues with geographical variability. In this study, GWR was used to identify a spatial regression model to examine correlations between groundwater salinity as a dependent factor and groundwater recharge potential and quality as an independent or influencing factor. Slope, geology, landforms, annual rainfall, soil texture class, drainage density, land use/land cover, chloride content, total dissolved solids, hydrogeological unit, groundwater potential, lineament density, and normalized difference vegetation index were among the independent factors studied.

1.2 Research Objectives

The objectives of this study are to 1) compare the performance of OLS regression and GWR models for salinity distribution in the Non Thai, Non Sung, Non Daeng, Khong, and Kham Sakae Saeng districts of Nakhon Ratchasima province, 2) investigate the spatial relationships between groundwater salinity and influencing factors.

1.3 Scope and Limitations

The scope and limitations of the research include as following:

- 1) The study area covered Non Thai, Non Sung, Non Daeng, Khong, and Kham Sakae Saeng districts in Nakhon Ratchasima Province.
- 2) The salinity content, chloride content, and total dissolved solids (TDS) are from the previous study by Thongwat (2018).
- 3) Slope, drainage density, lineament density, landforms, annual rainfall, geology, hydrogeological unit, groundwater potential, soil texture class, land use/land cover, normalized difference vegetation index, chloride content, and total dissolved solid were used as independent factors of groundwater discharge potential and quality.

4) ArcGIS software was used to run an exploratory regression model to determine the best model and create layer maps that help develop groundwater potential zoning maps.

5) The GWR4 software was used to determine and compare OLS regression and GWR models.

1.4 Research Methodology

Figure 1.1 provides a condensed representation of the research methodology, comprising six sequential stages: literature review, data collection and preparation, regression analysis, comparison of OLS regression and GWR models, discussions and conclusions, and thesis composition. Each step is as follows:

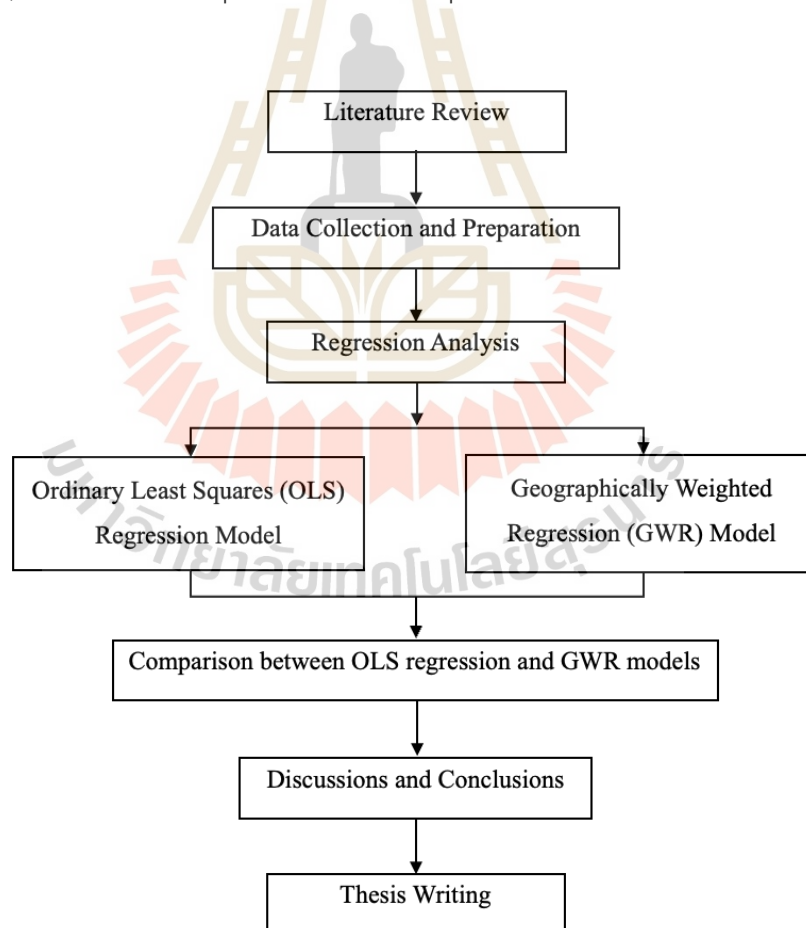


Figure 1.1 Flowchart of the research methodology.

1.4.1 Literature Review

Literature relevant to the research will be reviewed, summarized, and documented. The thesis will encompass a concise overview of the literature review, which will include the geology of the Khorat Plateau, the stratigraphy of the Maha Sarakham Formation, the hydrogeology of Nakhon Ratchasima, groundwater quality in Nakhon Ratchasima, saline soil in Nakhon Ratchasima, and regression analysis. The data was obtained from research articles, technical reports, and conference proceedings.

1.4.2 Data Collection and Preparation

The projections of each thematic layer were standardized, and the raster layers were converted into vectorized polygons. GDEM was used to generate the drainage density, lineament density, and slope in ArcGIS. Landforms, soil texture class, land use/land cover, and normalized difference vegetation index were obtained through the utilization of the Google Earth Engine and the Python API. Rainfall was gathered by the Royal Irrigation Department (RID). Geology, groundwater potential, and hydrogeological units were gathered from the Department of Mineral Resources (DMR) and digitized in ArcGIS. Salinity, chloride content, and total dissolved solids were collected from Thongwat (2018).

1.4.3 Regression Analysis

A statistical method for evaluating or estimating the association between a dependent variable and a group of independent explanatory variables.

1) Ordinary Least Squares (OLS) Regression

A global linear regression model predicts or models a dependent variable based on its correlations to a set of independent variables. A model is typically fitted using a process known as a multiple linear regression model to predict the dependent variable using a linear combination of the independent variables.

2) Geographically Weighted Regression (GWR)

A local form of spatial analysis was introduced in the geographical literature by Brunson, Fotheringham, and Charlton (1996); Brunson, Fotheringham, and Charlton (1998); Fotheringham et al. (2002). They utilized statistical methods for curve-fitting and smoothing to develop this approach. Geographically Weighted Regression (GWR) is an extension of Ordinary Least Squares (OLS) regression that considers spatial variation by examining the deviations of locally weighted regression coefficients from global coefficients.

1.4.4 Comparison between OLS Regression and GWR Models

The residuals and predictions of the dependent variables can be used to determine the model's goodness of fit. The coefficient of determination (R^2) or the adjusted coefficient of determination (adjusted R^2) value is the typical goodness of fit measure for the conventional global model. The corrected Akaike Information Criterion (AIC_c) is a measure of goodness of fit that utilize extensively in the GWR (Hurvich, Simonoff, & Tsai, 1998). The AIC_c can be used to compare models with different independent variable subsets, the global OLS, and a local GWR model.

1.4.5 Discussions and Conclusions

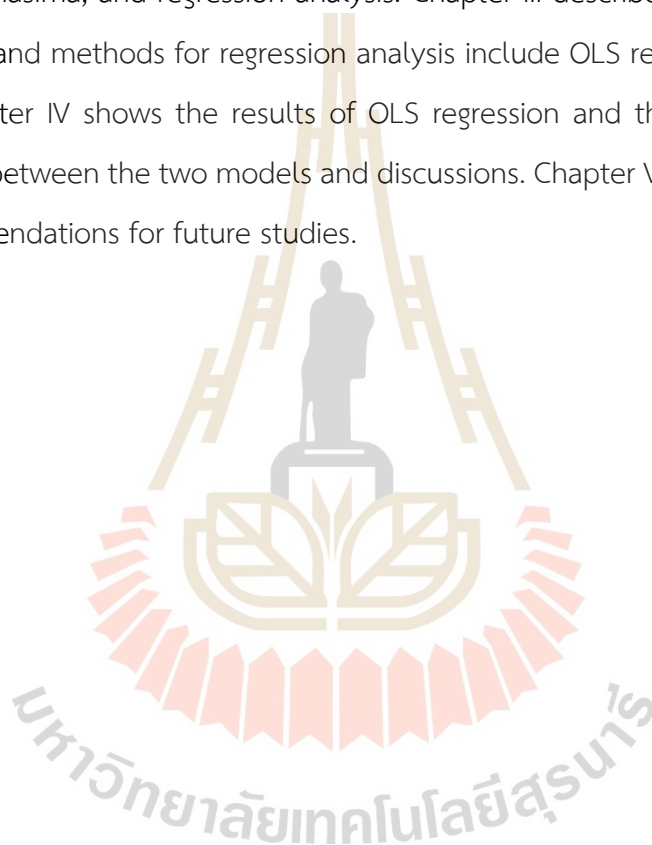
The reliability and adequacy of the utilized approaches are discussed, and areas for future research are identified. The research findings will be documented and published.

1.4.6 Thesis Writing

All research activities, methods, and results are documented and followed in the thesis.

1.5 Thesis Contents

This thesis is divided into five chapters. Chapter I includes background and rationale, research objectives, scope and limitations, and research methodology. Chapter II includes literature reviews of the previous studies related to the geology of the Khorat Plateau, the stratigraphy of the Maha Sarakham Formation, hydrogeology of Nakhon Ratchasima, groundwater quality of Nakhon Ratchasima, saline soil area of Nakhon Ratchasima, and regression analysis. Chapter III describes data collection and preparation, and methods for regression analysis include OLS regression and the GWR model. Chapter IV shows the results of OLS regression and the GWR model and a comparison between the two models and discussions. Chapter V presents conclusions and recommendations for future studies.



CHAPTER II

LITERATURE REVIEW

This chapter included a review of previous studies on the geology of the Khorat Plateau, the lithostratigraphy of the Maha Sarakham Formation, hydrogeology in Nakhon Ratchasima, groundwater quality in Nakhon Ratchasima, saline soil in Nakhon Ratchasima, and regression analysis. The results of the review of the literature are summarized below.

2.1 Geology of Khorat Plateau

The northeastern region of Thailand, referred to as the Khorat Plateau, is characterized by two extensive evaporate basins from the Cretaceous period. These basins, namely the Udon-Sakon Nakhon basin in the north and the Khorat-Ubol basin in the south, are geographically separated by the Phu Phan Range. The Khorat Plateau is primarily composed of Mesozoic-era continental sedimentary rocks belonging to the Khorat Group, predominantly from the Cretaceous period. The stratigraphy of the Khorat Plateau encompasses both Mesozoic and Cenozoic periods, as depicted in Figure 2.1 (El Tabakh, Utha-Aroon, & Schreiber, 1999).

The Khorat Group can be subdivided into two main units: the Lower Khorat Unit and the Upper Khorat Unit. The Lower Khorat Unit, there are several formations including the Huai Hin Lat Formation, Nam Phong Formation, Phu Kradung Formation, Phra Wihan Formation, Sao Khua Formation, Phu Phan Formation, and Khok Kruat Formation. On the other hand, the Upper Khorat unit consists of the Maha Sarakham Formation and Phu Tok Formation, as shown in Figure 2.2A. They are widely exposed throughout the northeastern part of Thailand, particularly on the edge of the Khorat

Plateau. The subsequent sections provide concise descriptions of the stratigraphic rock units, arranged in ascending order from the lower formations to the upper formations.

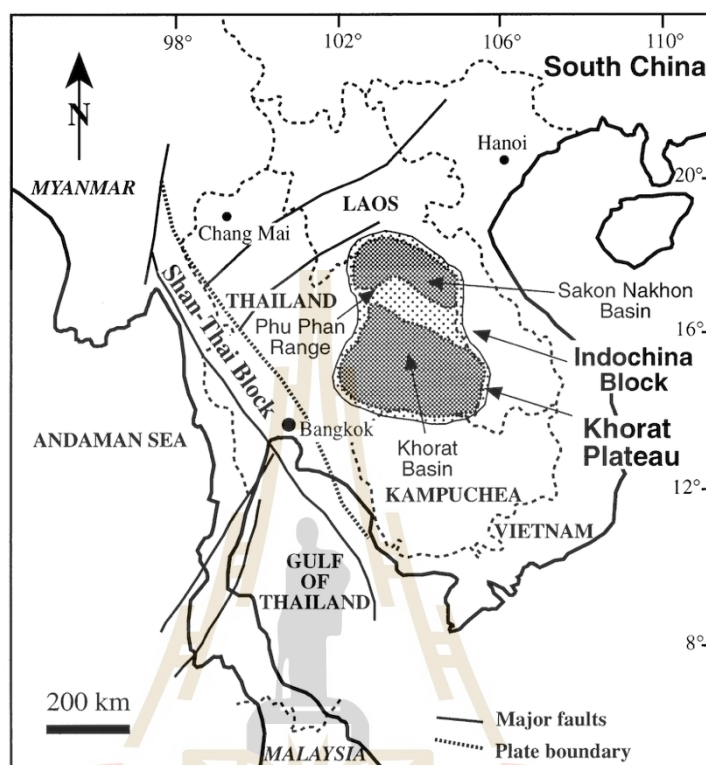


Figure 2.1 The Khorat-Ubol basin and the Udon-Sakon Nakhon basin (modified from El Tabakh et al. (1999)).

2.1.1 Lower Khorat Unit

1) Huai Hin Lat Formation

The lowest unit within the Khorat Group is the Huai Hin Lat Formation, characterized by a diverse lithology consisting of conglomerate, limestone conglomerate, sandstone ranging from grey to dark grey, siltstone, black shale, and marl containing plant fossils.

2) Nam Phong Formation

The Nam Phong Formation is divided into two parts. The upper part is mostly siltstone, with some sandstone and claystone and a thin layer of

limestone. Siltstone is a dusky red to reddish-brown to purple rock that is noncalcareous to slightly calcareous and contains lithoclastic claystone, limestone, and quartz grains. The lowest part consists of reddish-brown sandstone, siltstone, and claystone. Huai Hin Lat Formation is overlain by formation conformity. Its unconformity overlies Permian limestone in several regions.

3) Phu Kradung Formation

The Phu Kradung Formation is mainly composed of interbedded sandstone and siltstone, with occasional limestone lenses and concretions and a thin bed of claystone. Sandstone has a variety of colors, including light grey, light to medium brown, reddish-brown, and medium green to off-white with varicolored grains.

4) Phra Wihan Formation

The Phra Wihan Formation predominantly consists of light buff to grey sandstones with varying grain sizes, ranging from fine to coarse, accompanied by siltstones and mudstones to a lesser extent. Conglomerate occurrences are sporadic within the formation (Meesook, 2011). These sediments were deposited in a fluvial setting, specifically a shallow braided river system with high energy, along with the associated floodplain (Racey et al., 2009). The formation's thickness can vary between 100 and 250 meters.

5) Sao Khua Formation

The Sao Khua Formation is characterized by its reddish-brown conglomeratic sandstone, siltstone, and mudstone, occasionally accompanied by calcretes. Its thickness can vary between 100 and 600 meters. The formation is primarily deposited in a low-energy environment characterized by meandering channels and expansive floodplains.

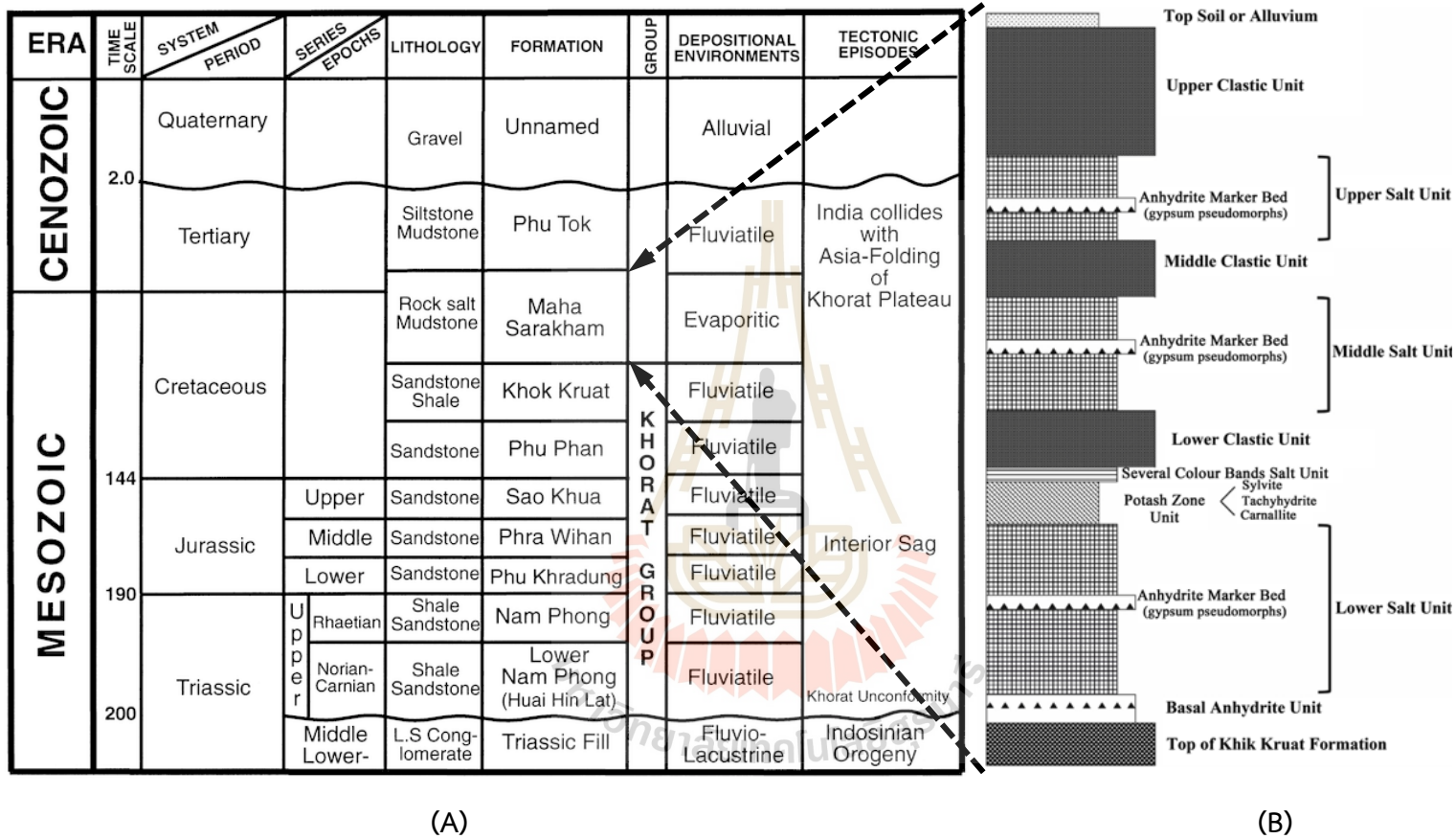


Figure 2.2 (A) Stratigraphy of the Khorat Basin, and (B) lithostratigraphy of the Maha Sarakham Formation (modified from El Tabakh et al. (1999) and Hansen, Wemmer, Eckhardt, Putthapiban, and Assavapatchara (2016)).

6) Phu Phan Formation

The Phu Phan Formation is characterized by greyish-white, medium to coarse-grained cross-bedded sandstones, accompanied by lacustrine grey siltstones and mudstones, with occasional conglomerates. Generally, the formation has a thickness ranging from 75 to 150 meters. The contact between the Phu Phan Formation and the underlying Sao Khua Formation varies, with local instances of erosive or distinctively sharp contact, but is mostly conformable with the overlying Khok Kruat Formation. The prevailing depositional environment of the Phu Phan Formation is interpreted as a high-energy, low-sinuosity braided river system with subordinate floodplains (Meesook, 2011; Racey et al., 2009).

7) Khok Kruat Formation

The Khok Kruat Formation consists primarily of reddish-brown sandstone, siltstone, shale, and mudstone, with occasional conglomerate layers (Meesook, 2011; Racey et al., 2009). A few meters below the Maha Sarakham Formation's underlying basal anhydrite, the reddish color gradually changes to greenish-grey. The presumed depositional setting was a meandering river system, characterized as less developed compared to the rivers that deposited the Sao Khua Formation (Meesook, 2011). The fauna present provides evidence of a riverine environment with potential marine influence (Racey et al., 2009).

2.1.2 Upper Khorat Unit

1) Maha Sarakham Formation

The Maha Sarakham Formation comprises four distinct salt units: the basal anhydrite and potash zone, lower salt and clastic bands, middle salt and clastic layers, and upper salt and clastic units (Figure 2.2B). Its thickness ranges from 250 to 1,100 meters across the basin, and its distribution varies in regions influenced by rock salt formations (Suwanich, 2007; Tabakh, Schreiber, Utha-Aroon, Coshell, & Warren, 1998). Each of the three units contains pseudomorphs of bottom-growth gypsum beds that have been replaced by halite. Due to the formation's origin in a

hyper-saline, inland lake within a dry continental desert, fossils are scarce (Booth & Sattayarak, 2011; Racey et al., 2009). Anticlines can be observed in different parts of the basin, including salt domes and salt formations. According to the descriptions provided by El Tabakh et al. (1999), a detailed description of all members, including the siliciclastic, is as follows. The rock salt found on the Khorat Plateau exhibits similar characteristics and correlations in both the Khorat and the Sakhon Nakhon Basins, suggesting that these two basins were once part of a single large basin during deposition until the Phu Phan Range uplifted in the Early Paleocene.

2) Phu Tok Formation

The Phu Tok Formation is located in the northern and central regions of the Khorat Plateau. It has a thickness ranging from 200 to 350 meters. The lower portion, approximately 100-200 meters, consists predominantly of brick-red to red-brown claystone with occasional siltstone and rare thin beds of fine-grained sandstone. However, the upper section primarily consists of thick layers of brick-red sandstone exhibiting cross-bedding. The lower part of the formation is believed to have been deposited in a relatively calm fluvial system, while the upper part is most likely indicative of an aeolian environment.

Tatong and Margane (2004), classified the sedimentary deposits in Nakhon Ratchasima into different categories, which include Residual Deposits (Qr), Colluvium Deposits (Qc), High Terrace Deposits (Qt), and Alluvial Deposits (Qa).

3) Residual Deposits (Qr)

Residual deposits within the Phu Tok Formation are formed through weathering processes occurring in situ. These deposits are primarily located in the northwestern region of Nakhon Ratchasima. The sequence of these deposits typically comprises layers of clayey silt or clayey fine sand, which overlay laterite and sapelite. Additionally, these deposits may exhibit some internal sedimentary structures.

4) Colluvium Deposits (Qc)

Colluvium Deposits were distinguished from the preexisting High Terrace Deposits by Chaimanee (2003). The sequence of Colluvium Deposits is characterized by its significant thickness, featuring a loose layer of reddish sand positioned above a gravelly laterite layer, which includes a few tektites. Beneath these two layers, there is an angular gravel bed.

5) High Terrace Deposits (Qt)

In the central region of Nakhon Ratchasima, one can encounter elongated hills known as High Terrace Deposits, stretching predominantly along the southern bank of the Lam Takhong River. These deposits, shaped by the forces of weathering, have transformed into mottled clay over time. At Ban Phu Khao Thong in Mueang district, these formations overlay the Khok Kruat Formation. Comprising mainly of fluvial sand and gravel, the High Terrace Deposits are further classified into two distinct sections: the lower, characterized as the older gravel beds, and the upper, representing the younger gravel beds.

Sataragsa (1987) has designated the lower section of the geological formation as the Phu Khao Thong Formation (see Figure 2.3). This segment primarily comprises partially consolidated bedded sand and gravel. The bedding of the Phu Khao Thong Formation exhibits a moderate inclination towards the north. The sediments display varying colors, ranging from shades of grey to yellowish-grey, with sporadic patches of red. The sandstone beds reveal conspicuous crossbedding features. In contrast, the gravel layers contain fragments primarily composed of black chert, quartz, and petrified wood, with poor sorting and varying degrees of roundness. The estimated age of the Phu Khao Thong Formation falls within the Late Tertiary period.

Moving to the upper part of the formation, we encounter layers of gravel and sand, overlain by laterite and eolian deposits, deposited approximately 2,000 to 3,500 years ago. The transition between the upper and lower parts of the formation is marked by an erosional contact. The gravel and sand within this segment

exhibit subrounded to rounded shapes and are well sorted. The fragments present in these layers are primarily composed of quartz, chert, and tektites. The age of the tektites has been determined using K/Ar isotopic and fission methods, indicating an age range of 0.68 to 0.73 million years before the present.

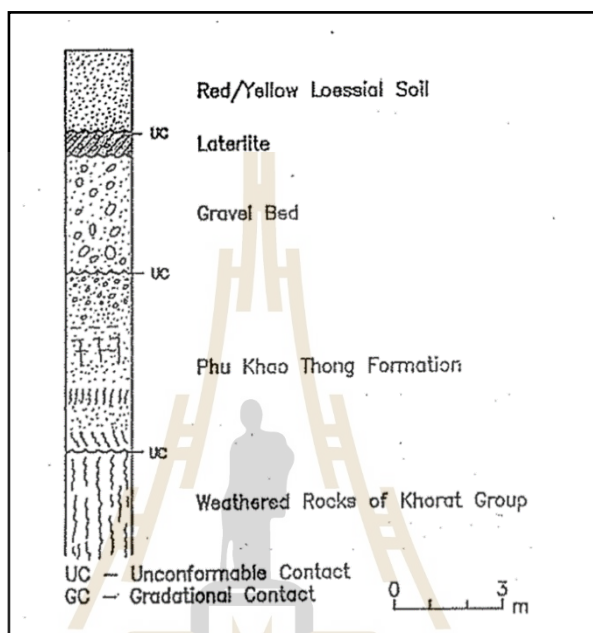


Figure 2.3 Stratigraphic section of the High Terrace at Ban Phu Khao Thong, Mueang district, Nakhon Ratchasima (After Tatong and Margane (2004)).

6) Alluvial Deposits (Qa)

The alluvial deposits found in the floodplains of the Mun, Lam Takhong, and Lam Chiang Krai rivers, particularly in the northern region of Nakhon Ratchasima, consist of various sediment types. These include silty clay, sandy clay, and clayey sand. Channel deposits within the floodplain are predominantly composed of sand and gravel. The color of these deposits ranges from grayish-brown to reddish-brown. Chaimanee (2003) classified the alluvial deposits into four units based on their sedimentary sequences and content, which are influenced by different sediment sources: Lam Takhong alluvial, Mae Nam Moon alluvial, Lam Chakkarat, and Alluvium Complex. The primary deposits within these sequences are fine sediments that

occasionally mix with sand and gravel. In general, as one moves towards the lower portion of the sequences, the sediments become coarser.

2.2 Lithostratigraphy of the Maha Sarakham Formation

The Maha Sarakham Formation, spanning from the Cretaceous to early Tertiary periods, consists of three evaporitic successions that are present in both the Khorat-Ubol basin and the Udon-Sakon Nakhon basin. These evaporites are characterized by extensive deposits of halite, anhydrite, and abundant potassic minerals such as sylvite and carnallite. Additionally, tachyhydrite and small amounts of borates can also be found within these formations (El Tabakh et al., 1999).

Based on the research by El Tabakh et al. (1999); Hansen et al. (2016), the stratigraphy of the Maha Sarakham Formation can be summarized as follows:

2.2.1 Lower Member

1) Basal Anhydrite Unit

The Basal Anhydrite Unit is the lowermost part of the Maha Sarakham Formation found in both the Khorat-Ubol and Udon-Sakon Nakhon basins. It is situated above the Khok Kruat Formation with a distinct disconformity, and it exhibits a stylolitic surfaces (El Tabakh et al., 1999; Tabakh et al., 1998). The most of anhydrite is formed directly or by the conversion of gypsum from beds and clusters with gradational contacts to enveloping sediments (Tabakh et al., 1998). Its thickness ranges between 0.7 and 6.2 meters (Utha-Aroon, 1993).

2) Lower Salt Unit

The Lower Salt Unit is situated below the Basal Anhydrite and the Potash Zone Unit. It mainly consists of coarse-crystalline, transparent to semitranslucent halite, with occasional thin layers of anhydrite Utha-Aroon (1993). The unit is divided into two sequences by an anhydrite marker bed known as the Halite L1 and Halite L2 Units, as named by El Tabakh et al. (1999). These marker beds, along with subsequent deposits, are primarily formed from seawater. The contact with the

Basal Anhydrite Unit is sharp and immediate. According to Utha-Aroon (1993), this could be attributed to either a sudden change in gypsum solubility as the brine entered the halite saturation field or a pressure-induced solution during burial. The thickness of this unit varies across the basin due to location and salt flowage, ranging from 30 to 500 meters (Suwanich, 2007).

3) Potash Zone Unit

The Potash Zone Unit is situated between the Lower Salt and the Several Color Bands Salt Units. It consists of layers containing potassium and magnesium minerals and is extensively distributed across the Khorat Plateau (Suwanich, 2007). The thickness and distribution of the Potash Zone Unit vary significantly, likely due to uneven dissolution, deformation, and minor variations in the deposition (El Tabakh et al., 1999). The Potash Zone Unit contains various potash minerals such as sylvite (KCl), carnallite ($\text{MgCl}_2 \cdot \text{KCl} \cdot 6\text{H}_2\text{O}$), and tachyhydrite ($\text{CaMgCl}_6 \cdot 12\text{H}_2\text{O}$) (Hite and Japakasetr (1979)). Sylvite is a secondary deposit that forms by leaching carnallite and occurs in intervals following the carnallite beds. On the other hand, potash minerals have a primary origin. The Potash Zone Unit is divided into three zones: a lower zone containing halite with small amounts of red-colored carnallite filling dissolution cavities, a middle zone consisting of pale red or colorless carnallite and halite, and an upper zone dominated by sylvite with traces of anhydrite and gypsum pseudomorphs (El Tabakh et al. (1999)). The boundaries between these zones exhibit gradual transitions. Borate minerals are commonly found in association with sylvite and halite, either as thin layers within the sylvite beds or as dispersed nodules and grains. Tachyhydrite is typically observed as a single bed situated between carnallite beds, with smooth contact. The presence of carnallite serves as an indicator of the lower boundary of the Potash Zone Unit. The unit's thickness typically varies from 10 to 80 meters, with an average thickness of 50 meters (El Tabakh et al., 1999; Suwanich, 2007).

4) Several Color Bands Salt Unit

The several color bands salt unit, which overlays the Potash Zone Unit, is a thin layer that exhibits varying colors such as grey, orange, red, honey, colorless, and smokey dark halite. It is likely that groundwater leaching played a role in its formation. The unit transitions into the underlying Potash Zone and contains deposits of carnallite and tachyhydrite (Suwanich, 2007).

5) Lower Clastic Unit

The Lower Clastic Unit is situated below the Potash Zone and above the Middle Salt Unit. The lithology of this unit closely resembles that of the Middle Clastic Unit, consisting of semi-consolidated to moist mudstones. It exhibits a dissolving contact with the underlying rock unit but lacks gypsum. The Lower Clastic Unit is predominantly dark red to reddish-brown in color, with occasional greenish-grey sections at the base. Within this unit, there are veins and veinlets of carnallite, which typically display a deep red-orange hue. The thickness of the Lower Clastic Unit can range from a few meters to up to 70 meters (Suwanich, 2007; Utha-Aroon, 1993).

2.2.2 Middle Member

1) Middle Salt Unit

The Middle Salt Unit is positioned above the Lower Clastic Unit and below the Middle Clastic Unit, with an average thickness ranging from 30 to 130 meters (Suwanich, 2007). Similar to the Lower Salt Unit, the Middle Salt Unit exhibits various depositional and diagenetic characteristics. El Tabakh et al. (1999) divided this halite unit into two sequences: the Halite M1 Unit and the Halite M2 Unit, which are separated by an anhydrite marker bed. The Halite M1 Unit consists of well-bedded, dark honey-colored halite layers interspersed with thin beds of anhydrite and gypsum pseudomorphs. Occasional occurrences of sylvite and carnallite spots have also been observed within this unit. On the other hand, the Halite M2 Unit features alternating dark honey-colored bedded halite and dark smoky-colored halite beds, along with anhydrite stringers and gypsum pseudomorphs.

2) Middle Clastic Unit

The Middle Clastic Unit is positioned above the Middle Salt Unit and below the Upper Salt Unit. It primarily consists of clay, displaying a dark reddish-brown color. Occasional occurrences of breccia strata have been identified within this rock layer. In certain locations, calcareous components, gypsum veins, and veinlets can be found near the base (Suwanich, 2007). The thickness of the Middle Clastic Unit varies between 20 and 70 meters (El Tabakh et al., 1999). In cases where the Upper Salt Unit is absent due to dissolution and salt flowage, the Upper Clastic Unit directly overlays the Middle Clastic Unit. The contact between the Middle Clastic Unit and the underlying Middle Salt Unit can exhibit either a sharp or gradational transition (Suwanich, 2007).

2.2.3 Upper Member

1) Upper Salt Unit

If not affected by salt structures or dissolution processes, the Upper Salt Unit is typically positioned above the Middle Clastic Unit and the underlying Upper Clastic Unit. The dissolution and erosion caused by salt flowage can result in the disappearance of Upper Salt ridges and domes near areas of structural uplift. The thickness of the unit varies significantly and can reach up to 20 meters. It predominantly consists of moderate to dark yellowish-brown halite, with occasional thin layers of anhydrite, dark smoky-colored halite bands, milky-white-colored halite, and rare, orange-colored halite grains. While the Upper Salt Unit shares similarities with the Middle Salt Unit in terms of its characteristics, geochemical methods utilizing K/Br content can be employed to distinguish between them (Suwanich, 2007). The Upper Salt Unit has been categorized into two sequences by various researchers, including Suwanich (2007) and El Tabakh et al. (1999). These sequences are referred to as the Halite U1 Unit and the Halite U2 Unit, with an anhydrite marker bed of around 1.5 meters separating them. The Halite U1 Unit consists of moderate to dark-honey or yellowish-brown halite, accompanied by minor anhydrite beds and occasional layers

of gypsum pseudomorphs. On the other hand, the Halite U2 Unit primarily comprises dark-honey halite, interspersed with dark smoky-colored halite and minor, orange-colored halite. Some localized occurrences of potash residues have also been observed in this unit.

2) Upper Clastic Unit

The Upper Clastic Unit, situated at the top of the Maha Sarakham Formation, predominantly consists of pale to moderate reddish-brown or brick-red sandstone, siltstone, and claystone (Suwanich, 2007). The presence of cross-beds and well-defined bedding is evident within this unit. Its thickness can vary significantly, reaching up to 680 meters (El Tabakh et al., 1999). Within the middle to lower parts of the unit, white clear gypsum veins and veinlets can be observed. The contact between the Upper Clastic Unit and the overlying Phu Tok Formation, which is also clastic in nature, is not clearly delineated. However, it is noteworthy that the clastic material within the Phu Tok Formation appears to be coarser compared to that of the Upper Clastic Unit.

2.3 Hydrogeology of Nakhon Ratchasima

According to Tatong and Margane (2004), the hydrogeological characteristics of Nakhon Ratchasima are primarily influenced by compacted formations consisting of Mesozoic sandstone, shale, and siltstone. Unconsolidated formations are limited to the vicinity of the Mun and Lam Takhong rivers. These formations can be classified into two distinct aquifer categories, as described below.

2.3.1 Unconsolidated Aquifer

The Unconsolidated Aquifer occurs in two distinct depositional settings, namely alluvial deposits and high terrace and colluvium deposits, which are described below.

1) Alluvial Deposits

Aquifers are present within these deposits, located adjacent to the Mun and Lam Takhong rivers. These aquifers form elongated strips that stretch from east to west. Groundwater is stored within sand and gravel layers at depths varying from 10 to 30 meters. These layers were formed by meandering streams and primarily consist of sand and gravel, with occasional thin layers of clay. It is important to note that groundwater within this layer is hydraulically interconnected.

2) High Terrace and Colluvial Deposits

These deposits form aquifers in the hilly region south of Nakhon Ratchasima and in the floodplain area where alluvial deposits are overlain. Groundwater is typically found at two distinct depth intervals in sand and gravel: 20 - 40 meters and 50 - 70 meters below the ground surface. A layer of fine-grained rock roughly 10 meters thick divides the two sand and gravel levels.

2.3.2 Consolidated Aquifer

The consolidated aquifer is divided into eight formations, which are described below: Nam Phong Formation, Phu Kradung Formation, Phra Wihan Formation, Sao Khua Formation, Phu Phan Formation, Khok Kruat Formation, Maha Sarakham Formation, and Phu Tok Formation.

1) Nam Phong Formation

This formation is composed of siltstone, sandstone, and conglomerates. The formation has a total thickness of 1,456 meters. The aquifer is supported by the Pre-Khorat erosional unit.

2) Phu Kradung Formation

This unit has a thickness of approximately 972 meters. Their outcrops and sub-outcrops were discovered across the Khorat Plateau. It is comprised of shale, siltstone, sandstone, and conglomerate.

3) Phra Wihan Formation

Its thickness ranges between 50 to 297 meters. This formation is composed of a massive white to pink, thick-bedded, well-sorted quartz sandstone with thin layers of laminated red siltstone. The groundwater was of high quality.

4) Sao Khua Formation

This unit is made up of sandstone and siltstone with thicknesses ranging from 400 to 720 meters. It was found in regions with flat to undulating terrain. The yield of groundwater from several boreholes in the Sao Khua aquifer is between 5 to 10 cubic meters per hour, with particularly high-quality water.

5) Phu Phan Formation

Massive coarse-grained quartzitic sandstone with occasional conglomerate defines the unit. Its thickness is between 100 to 400 meters. Geomorphologically, the landscape ranges from almost flat top hills to undulating terrain. The yield of groundwater ranges from 1 to 10 cubic meters per hour. This amount can be predicted from a drilling well that has penetrated the aquifer's fractured zones. Groundwater is generally of high quality, with significant iron amounts on occasion. Flowing artesian water was not discovered in the Phu Phan formation aquifer. Several flowing wells have been drilled, though.

6) Khok Kruat Formation

Groundwater is present mostly in the voids between fractures and bedding planes of sandstone, shale, and siltstone. The quality of the groundwater in this formation is generally good. However, saltwater can be found in places where the rock contacts the Maha Sarakham Formation.

7) Maha Sarakham Formation

The Maha Sarakham Formation is typically found at a relatively shallow depth of around 80 to 100 meters below the Earth's surface. The upper surface of the rock salt exhibits a smooth and moderate inclination towards the

northeast, as observed from seismic section profiles. Due to the limited presence of primary porosity, the formation predominantly acts as an aquitard, impeding the flow of groundwater. Groundwater can only accumulate within the formation when it comes into contact with porous rock units. Brine water, commonly extracted from aquifers associated with these porous rock units, serves as a major source for salt mining activities.

8) Phu Tok Formation

When compared to the underlying formations, it is not well cemented and slightly soft. It is mainly composed of claystone, siltstone, and sandstone. At outcrop, the topography is mainly flat to slightly undulating. The formation can form a suitable aquifer. Groundwater can be trapped in both primary and secondary pore spaces within the formation. However, the presence of underlying rock salt layers can contribute to the salinization of groundwater. The high concentration of salt chloride in the formation leads to poor-quality groundwater. In certain areas, local communities extract salty water and use it to produce table salt. However, regions with deeper rock salt layers or where freshwater forms a shallow lens over the saltwater in recharge areas may have low salinity in groundwater.

2.4 Groundwater Quality of Nakhon Ratchasima

Total Dissolved Solids (TDS) contain nitrate and iron, which may be harmful to human health, according to Tatong and Margane (2004). Groundwater with this high TDS content is unsuitable for human consumption. The TDS dissolved in groundwater of this region is mostly sodium chloride (NaCl).

Wannakomol (2012) identified shallow saline groundwater levels in Sikhio, Dan Khun Thot, Phra Thong Kham, Kham Sakae Saeng, Non Sung, Phimai, Kham Thale So, Muang Nakhon Ratchasima, and Chaloem Phra Kiat districts using an electrical survey in Nakhon Ratchasima province. The groundwater quality and hydrogeology of the study area were summarized as follows:

1) Non Thai district

The main rock types present in this district are shale, sandstone, siltstone, claystone, and rock salt, collectively known as the Maha Sarakham Formation. Rock salt serves as the primary contributor to the presence of saline and brackish groundwater. Aquifers containing groundwater have been identified within fractures in the rock formation. The yield of groundwater from these aquifers varies between 2 and 10 cubic meters per hour. On average, the aquifers have a depth of 15 to 30 meters, with the water level typically ranging from 3 to 8 meters. The groundwater in these aquifers is characterized by its salinity, brackishness, and hardness.

2) Non Sung district

The predominant rock type in the Non Sung district is rock salt from the Maha Sarakham Formation. Consequently, the groundwater in this district is characterized by its salinity, hardness, and elevated iron levels.

3) Non Daeng district

The geological composition of this district consists of shale, sandstone, siltstone, claystone, and rock salt derived from the Maha Sarakham Formation. Groundwater characteristics vary depending on the location: the groundwater in the river plain along the Sakae River is saline and brackish, while the groundwater on the hill is freshwater. The yield of groundwater ranges from 2 to 10 cubic meters per hour, with an average aquifer depth of 10 to 30 meters and a water level ranging from 3 to 8 meters.

4) Khong district

The geological composition of the district consists of various rock types such as shale, sandstone, siltstone, claystone, and rock salt from the Maha Sarakham Formation. The majority of groundwater in the area is characterized as saline or brackish, although freshwater sources have been identified in certain locations.

Groundwater yield ranges from 2 to 10 cubic meters per hour, with an average aquifer depth of 20 to 50 meters and a water level ranging from 3 to 8 meters. The quality of groundwater is generally high in terms of salinity, brackishness, and hardness, although some areas exhibit better quality.

5) Kham Sakae Saeng district

The Maha Sarakham Formation is composed of shale, sandstone, siltstone, claystone, and rock salt. The groundwater in the Kham Sakae Saeng district is predominantly characterized by high salinity, brackishness, and hardness, accounting for over 80% of the total groundwater quality. However, some regions exhibit medium to good quality. Freshwater areas are primarily located in the northern and northwestern parts of the district. The yield of groundwater ranges from 2 to 10 cubic meters per hour, with an average aquifer depth ranging from 10 to 35 meters and a water level varying from 3 to 8 meters.

2.5 Saline Soil of Nakhon Ratchasima

Saline soil contains a high concentration of dissolved salts. Saline soil areas are mainly devoid of vegetation or have only sparse vegetation. When dry, salt-affected soil areas can have a bright white, crusty color. The problem of soil salinity can be found all over the world, and the causes of the salinity might vary depending on the geological characteristics of the region. In both irrigated and non-irrigated agricultural areas, human activities can produce soil and water salinity concerns (Peck & Hatton, 2003).

According to DMR (2015), the Khorat Plateau covers approximately one-third of Thailand and was classified as saline soil terrains with three salinity levels: (1) extremely saline terrain, (2) moderately saline terrain, and (3) slightly saline terrain. Udon Thani, Sakon Nakhon, Nakhon Phanom, Khon Kaen, Kalasin, Chaiyaphum, Maha Sarakham, Roi Et, Yasothon, Amnat Charoen, Nakhon Ratchasima, Buri Ram, Surin, Si Sa Ket, and Ubon Ratchathani are the fifteen provinces affected by saline soils, which

are primarily caused by dissolved sodium chloride from deep level. During each dry season, salt is dispersed throughout the lowland ground surface.

Saline soil can be found in various regions of Thailand, and the salinity occurs from a variety of geologic conditions. The Maha Sarakham Formation, which is comprised of interbedded rock salt layers, underlies large areas of northeastern Thailand and is the region's main source of salinity. The Maha Sarakham Formation rock salt continues to cause serious difficulties with saline soils and saline water in the area (Wongsomsak, 1986). Topsoil contaminated with saltwater is typically found in low-lying areas with shallow groundwater levels and a low groundwater head gradient (Seeboonruang, 2013; Wannakomol, 2012).

The Rock Salt Members, with a potential thickness of 700 meters, play a significant role in causing salinity in the Khorat Plateau. Within the salt-bearing layer of the Upper Clastic Member of the Maha Sarakham Formation, salts such as gypsum, sulfate, and carbonate have accumulated. During the rainy season, a portion of the salty content is leached away, resulting in a reduction in salinity. Soil salinity has rapidly spread and reduced the quality of life for many people, producing various economic, social, and environmental problems.

One of the provinces affected by salinity is Nakhon Ratchasima. There are approximately 3,424 square kilometers of agricultural land with saline soil. Although salinity can contribute to the enhancement of soil structure, it can also exert adverse effects on plant growth and agricultural productivity. Northeastern Thailand's salt-affected soils are mainly sandy, with low infertility and high sodium and chloride concentration. Other observable indicators of saline soil encompass the presence of salt crystals in close proximity to an elevated soil moisture level at the uppermost layer.

2.6 Thematic Layers

2.6.1 Slope

The slope of a surface is the greatest rate of change in height across an area of the surface, and it is the primary topographical component influencing land stability (Manjare, 2014). Due to its impact on the infiltration of groundwater into the subsurface, the slope of the topography serves as a suitable indicator for assessing groundwater conditions (Sikakwe, Ntekim, Obi, & George, 2015). In regions with moderate slopes, surface runoff occurs gradually, allowing rainwater more time to seep into the ground. On the other hand, areas with steep slopes facilitate rapid runoff, reducing the residence time of rainwater and resulting in slightly lower infiltration (Malik & Rajeshwari, 2011). The slope map of the area was created using the ASTER Global Digital Elevation Model (GDEM) with a spatial resolution of 30 m.

2.6.2 Drainage Density

It is defined as stream network spacing and its relationship to surface runoff and permeability (Magesh, Chandrasekar, & Soundranayagam, 2012). It is an important factor for determining the groundwater potential zone because it is an inverse function of permeability. The interpretation of groundwater conditions in a specific region is influenced by the drainage characteristics of the basin, which in turn affect the subsurface hydrological condition. When the rock has low porosity, it results in a reduced rainfall infiltration (Hutti & Nijagunappa, 2011). The drainage density was generated in ArcMap using the spatial analyst tool, which was used to extract stream networks from the global digital elevation model (GDEM). The drainage density was determined from the derived stream network after the line density was prepared.

2.6.3 Lineament Density

Lineaments are linear topographic features that represent the underlying geological structure. Lineament refers to the weak zones of bedrock that are considered secondary aquifers in rock outcrops because of their higher permeability and porosity (Selvam et al., 2016). Lineaments are strong indications of

groundwater potential because they allow water to flow (Kumar, 1999; V.R, Aravindan, & Gopinath, 1998). A high lineament density results in a high groundwater potential, while a low lineament density results in a low groundwater potential.

2.6.4 Landform

Landform investigations are a critical component of assessing both surface and groundwater resources. Landform mapping assists in the identification and categorization of various landforms and structural characteristics that support groundwater occurrences (Malik & Rajeshwari, 2011). Analyzing the formation of landforms and studying the geological aspects in a comprehensive manner can aid in identifying the existence of areas that are permeable and have high porosity (Karanth & Seshu babu, 1987). An essential aspect in estimating groundwater resources is the classification of landforms based on their morphological and lithological characteristics, as it helps establish hydro-morphological features. The examination of landforms is vital for comprehending a range of factors that impact groundwater resources, including the distribution of groundwater, slope characteristics, relief patterns, weathering depths, the composition and thickness of deposited materials, and the spatial arrangement of diverse landforms. In their study, Theobald, Harrison-Atlas, Monahan, and Albano (2015) attributed both collective and individual weights to the landform data obtained from the SRTM Landform dataset, considering its impact on the occurrence of groundwater. The landform data underwent acquisition and processing through the utilization of the Google Earth engine and the Python API, and the values were subsequently organized using the attribute table in ArcGIS.

2.6.5 Annual Rainfall

Rainfall plays a critical role in the water cycle as it has a significant impact on groundwater potential. When rainfall is abundant, there is a greater likelihood of groundwater recharge, while scarce rainfall significantly reduces the potential for groundwater replenishment (Kumar, Gautam, & Kumar, 2014). The annual

rainfall data was collected by the Royal Irrigation Department (RID), and ArcGIS spatial analyst tools were employed to interpolate and classify the data.

2.6.6 Geology

The type of rock exposed to the surface has a significant impact on groundwater recharge, according to Shaban, Khawlie, and Abdallah (2006). Geological factors impact groundwater recharge by affecting the percolation of water. Geology data, obtained from Thailand's Department of Mineral Resources (DMR) at a scale of 1:50,000 using the Web Map Service (WMS), was digitized, and classified using ArcMap.

2.6.7 Hydrogeological Unit

An aquifer is a geological formation or zone with specific hydraulic properties that significantly impact groundwater storage and movement. It is considered a crucial dataset in assessing groundwater potential. An aquifer can consist of a single soil or rock unit or a combination of formations, providing abundant water to wells and springs due to its saturated and permeable characteristics (Lohman & Mentink, 1972). The hydrogeological unit data, obtained from Thailand's Department of Mineral Resources (DMR), was processed, and classified using ArcMap by digitizing and utilizing the WMS at a scale of 1:50,000.

2.6.8 Groundwater Potential

The reclassification of groundwater potential is influenced by two factors: 1) Specific yield, which plays a crucial role in land surface models for predicting the depth of the water table, thereby influencing the relationship between soil moisture and groundwater (Lv, Xu, Yang, Lu, & Lv, 2021). 2) Total dissolved solids (TDS), as defined by WHO. TDS levels of approximately 500 ppm are recommended, with a maximum of 1,500 ppm allowed. TDS concentrations in groundwater samples are high due to salt leaching from the soil, and domestic wastewater may percolate into the groundwater (Sarath Prasanth, Magesh, Jitheshlal, Chandrasekar, & Gangadhar, 2012). The salinization of soil is considerably aided by the quality and yield of groundwater

(Sattayarak, 1987). Specific yield and Total Dissolved Solids (TDS) were acquired at a 1:50,000 scale through the utilization of the Web Feature Service (WFS) provided by Thailand's Department of Mineral Resources (DMR). The data were digitized, classified, and processed using ArcMap.

2.6.9 Soil Texture Class

Soil, as a natural resource, holds significant importance in delineating zones of groundwater potential and is crucial for facilitating groundwater recharge (Radhakrishnan & Ramamoorthy, 2014). Sandy loam exhibits a high rate of infiltration, indicating areas with high groundwater potential. In contrast, clay has a lower infiltration rate and thus lower groundwater potential. The collection and processing of soil texture class data were conducted utilizing the Google Earth engine and the Python API.

2.6.10 Land Use /Land Cover

It plays a substantial role in influencing the runoff, infiltration, and groundwater recharge capacity of a watershed or subbasin (Guru, Seshan, & Bera, 2017). Additionally, it offers soil-related data such as soil moisture content, groundwater levels, surface water conditions, and an indicator of groundwater potential development (Ahmad et al., 2020). The land use and land cover datasets were collected from the Sentinel-2 Land Use and Land Cover map developed collaboratively by Esri, Microsoft, and Impact Observatory. These datasets were processed using the Google Earth Engine and Python API.

2.6.11 Normalized Difference Vegetation Index

The Normalized Difference Vegetation Index (NDVI) can be utilized to assess groundwater levels, specifically in naturally vegetated regions like forests or shrublands. The NDVI serves as an important indicator of water availability, which significantly impacts the growth and coverage of dense vegetation. In areas with shallow water tables, natural vegetation, and a lack of in-situ groundwater

observations, the NDVI can serve as a reliable indicator of groundwater storage conditions (Bhanja et al., 2019). The NDVI data were obtained from Sentinel-2 images and underwent processing using the Python API of the Google Earth Engine platform.

2.6.12 Salinity Content

Salinization is a significant environmental problem that has adverse effects on soil quality, water resources, agricultural productivity, and the balance of natural ecosystems. Elevated levels of different elements, including sodium, sulphate, boron, fluoride, selenium, arsenic, and increased radioactivity, are commonly linked to higher levels of groundwater salinity (Gopal, 2019). Soil salinization poses a significant risk to the desertification of land, particularly in regions that experience frequent droughts. In such areas, there is high evaporation from the soil, and surface water contains highly soluble salts. The movement of water and salt at a regional level is influenced by factors such as climate, topography, hydrogeology, and human activities, which can contribute to the occurrence of these conditions (Cui, Lu, Zheng, Liu, & Sai, 2019).

2.6.13 Chloride Content

Chloride is a prevalent anion found in tap water, often combining with calcium, magnesium, or sodium to form various salts due to its negative charge. Unlike elements, ions do not have a balanced charge. Sodium chloride (NaCl) is a common compound composed of sodium and chlorine. When dissolved in water, sodium chloride generates a positively charged ion and a negatively charged chloride anion. Chloride can be found in rainwater, streams, groundwater, seawater, wastewater, urban runoff, and geological formations.

2.6.14 Total Dissolve Solid

Total Dissolved Solids (TDS) is among the multiple techniques used to assess salinity levels. TDS quantifies the concentration of all dissolved substances in water, encompassing organic matter and suspended particles that can traverse a highly

refined filter. In laboratory settings, TDS measurements are conducted and reported in milligrams per liter (mg/L) ("Chemicals of Concern (COC) - Groundwater Information Sheets (Salinity)," 2017).

2.7 Regression Analysis

Regression analysis is widely acknowledged as the most commonly used statistical method for studying or estimating the relationship between a dependent variable and a set of independent predictor variables. By identifying and assessing these relationships, Insights into the dynamics of a given area are gained, predictions about future occurrences are made, and exploration of the underlying factors contributing to specific spatial patterns is conducted (M. Charlton & A. S. Fotheringham, 2009).

The most well-known regression approach is Ordinary Least Squares (OLS). It's also a useful place to begin any spatial regression analysis. It creates a single regression equation to represent the process and gives a global model of the variable or process you are attempting to understand or predict (M. Charlton & A. S. Fotheringham, 2009).

Geographically Weighted Regression (GWR) is a spatial regression approach gaining popularity. It fits a regression equation to each feature in the dataset, providing a localized modeling approach for a better understanding and prediction of the variable being investigated. The proximity characteristics, or bandwidth, are determined by user-defined parameters such as Kernel type, Bandwidth method, Distance, and Number of neighbors. GWR offers robust statistical tools for exploring and estimating linear relationships (M. Charlton & A. S. Fotheringham, 2009).

Positive and negative linear relationships exist, Figure 2.4 shows both positive and negative associations and situations in which no relationship exists between two variables (ESRI, n.d.).

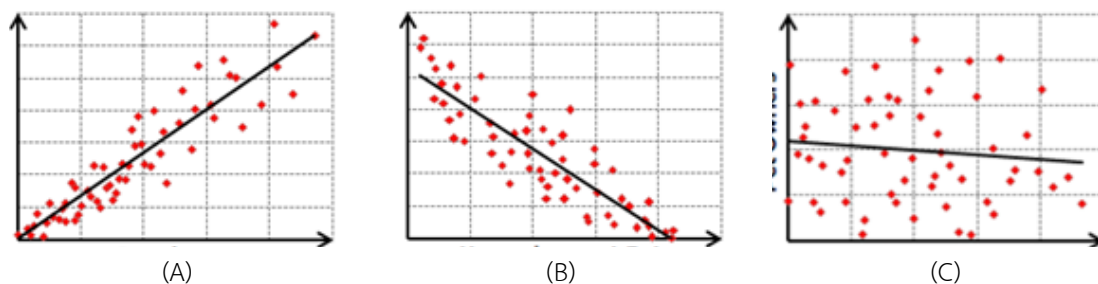


Figure 2.4 Scatterplots: (A) positive relationship, (B) negative relationship, and (C) two variables are unrelated (ESRI, n.d.).

2.7.1 Ordinary Least Square Regression

M. Charlton and S. Fotheringham (2009) stated that regression involves a range of methods employed to analyze the relationship between a dependent variable and one or more independent variables. The dependent variable, also known as the y-variable or response variable, is examined in conjunction with the independent variables, which can be referred to as x-variables, predictor variables, or regressors.

An equation represents a regression model. A linear regression model (Equation 2.1) can have the following form in its most basic version:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{for } i=1 \dots n \quad (2.1)$$

where y_i is the dependent variable that is measured at some location i ,

x_i is the independent variable,

ε_i is the error term, and

β_0 and β_1 are parameters that are to be estimated such that the value of

$\sum_{i=1}^n (y_i - \hat{y}_i)^2$ is minimized over the n observations in the dataset. The \hat{y}_i is the predicted or fitted value for its observation, given its value of x . The term $(y_i - \hat{y}_i)$ is known as the residual for its observation, it is important for the residuals to exhibit independence and conform to a Normal Distribution with a mean of zero. To accomplish this, an Ordinary Least Squares (OLS) Regression is commonly employed to fit such a model.

More generally, a multiple linear regression model may be written as Equation 2.2:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi} + \varepsilon_i \quad \text{for } i=1 \dots n \quad (2.2)$$

In the case where the predictions of the dependent variable are obtained by combining the independent variables linearly, the estimator for Ordinary Least Squares (OLS) takes the form as shown in Equation 2.3:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (2.3)$$

In the context of statistical analysis, the estimated parameters are represented by the vector $\hat{\beta}$, the design matrix X comprises the independent variables' values along with a column of 1s, the vector y represents the observed values, and the inverse of the variance-covariance matrix is denoted as $(X^T X)^{-1}$.

For example, weighting the observations in regression is sometimes advantageous to account for varying levels of data uncertainty. The square matrix W is enhanced by incorporating weights along its leading diagonal, thereby modifying the estimator (Equation 2.4) to incorporate these weights:

$$\hat{\beta} = (X^T W X)^{-1} X^T W y \quad (2.4)$$

The goodness of fit evaluates how well the model can replicate the observed y values. The measure of goodness of fit is often represented by the R^2 value, which falls between 0 and 1. It quantifies the proportion of variation in the observed y values that can be attributed to the variation in the model. However, since the R^2 value can be artificially increased by adding more variables, the adjusted R^2 is commonly used. The adjusted R^2 considers the number of independent variables in the model and reflects the model's simplicity.

In a regression model, it is important to assess if a parameter's value significantly deviates from zero, indicating that changes in the associated variable affect the predictions. To determine the meaningful contribution of variables in the model,

the estimated parameter for each variable is divided by its standard error. These statistics follow a t-distribution, and their comparison to critical values from the t-distribution helps evaluate their significance, considering the degrees of freedom in the model.

2.7.2 Geographically Weighted Regression

Brunsdon et al. (1996); Brunsdon et al. (1998); Fotheringham et al. (2002) introduced Geographically Weighted Regression (GWR) as a local form of spatial analysis. GWR is an extension of OLS regression which models relationships as they vary across space by evaluating which locally weighted regression coefficients deviate from global coefficients in the geographical literature. It draws from statistical approaches for curve-fitting and smoothing applications. In other words, GWR is used to recognize the various nonstationary in variable values; stationarity refers to the situation in which the mean, variance, and location dependency do not change throughout space. GWR, in effect, moves a weighted window all over the spatial data and estimates unbiased slope and intercept parameters from "fit points," which can then be compared to global estimated parameters from a distance matrix of all points in the dataset. The size of the bandwidth can be calibrated in a variety of ways, and data is weighted as a function of the distance from the fit point using a kernel density function. The regression equation is modified across geographic space to account for variation caused by changes in spatial autocorrelation structure across an environment. Although GWR explicitly models spatial variation, it is limited in that it violates the predictor independence assumption as the same location may be included in different local estimated parameters, and there are also known issues with determining the goodness of fit statistics. For an in-depth discussion of GWR.

According to M. Charlton and S. Fotheringham (2009), It is also assumed that the analyst possesses a dataset with a dependent variable y and a set of m independent variable(s) X_k , $k=1...m$, and that a suitable coordinate system provides a measure of the position for each of the n observations in the dataset.

One possible formulation (Equation 2.5) for the GWR variant of the OLS regression model's equation could be:

$$y_i(\mathbf{u}) = \beta_{0i}(\mathbf{u}) + \beta_{1i}(\mathbf{u})x_{1i} + \beta_{2i}(\mathbf{u})x_{2i} + \dots + \beta_{mi}(\mathbf{u})x_{mi} \quad (2.5)$$

The notation $\beta_{0i}(\mathbf{u})$ indicates that the parameter explains a relationship that is specific to location u . If measured values for the independent variables are available at location u , a prediction for the dependent variable could be made. Usually, the locations at which estimated coefficients are obtained correspond to the data collection points, although this may not always be true.

This appears to be an unusual claim, but the clarity emerges when considering the nature of geographical weighting.

2.7.3 Weighting Functions

The weighted function for parameter estimation was developed by Fotheringham et al. (2002). This model's estimator is identical to the WLS (weighted least squares) global model above, except that the weights are conditioned on the position u relative to the other observations in the dataset and so change for each location. The estimator (Equation 2.6) is written as follows:

$$\hat{\beta}(\mathbf{u}) = (X^T W(\mathbf{u}) X)^{-1} X^T W(\mathbf{u}) y \quad (2.6)$$

where $W(u)$ is a square matrix of weights according to location u in the study area; $X^T W(u) X$ is the geographically weighted variance-covariance matrix (the estimation requires its inverse); and y is the vector of dependent variable values.

The $W(u)$ matrix includes geographical weights in its lead diagonal and 0 in its off-diagonal components.

$$\begin{bmatrix} w_1(\mathbf{u}) & 0 & 0 & 0 \\ 0 & w_2(\mathbf{u}) & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & w_n(\mathbf{u}) \end{bmatrix}$$

2.7.4 Bandwidth Selection

The weights themselves are determined using a weighting technique known as a kernel. There are several kernels that can be used: A Gaussian form is usual (Equation 2.7):

$$w_i(\mathbf{u}) = e^{-0.5\left(\frac{d_i(\mathbf{u})}{h}\right)^2} \quad (2.7)$$

where $w_i(\mathbf{u})$ is the geographical weight of its observation in the dataset in relation to the location \mathbf{u} , $d_i(\mathbf{u})$ is a measurement of the distance between its observation and the location \mathbf{u} , and h is the bandwidth. When employing Cartesian coordinates, the distances typically employed are Euclidean distances, while spherical coordinates often utilize Great Circle distances. However, it is possible to use non-Euclidean distances, such as distances along a road network, without any limitations.

The bandwidth in the kernel is measured in the same units as the dataset's coordinates. As the bandwidth increases, the weights tend to converge towards unity, causing the local GWR model to approach the global OLS model.

As explained previously, the usual practice is to estimate parameters at the locations where the dataset's observations were collected. This allows for predicting the dependent variable and calculating the residuals, which are essential for assessing the model's goodness of fit. Further discussion on this topic will be provided below. In addition to the sample sites, parameters can also be estimated at non-sample sites within the study region, such as the points of a grid or the locations of observations in a validation dataset that shares the same dependent and independent variables as the calibration dataset.

The sample points refer to the specific locations where the calibration data was collected, while the regression points correspond to the locations where the parameter estimation takes place. The components of a local model consist of the geographically weighted estimator, kernel, and bandwidth.

Alternative kernels can be utilized in GWR, although their specific form is of less importance as long as they exhibit a "Gaussian-like" behavior. The selection of the bandwidth plays a more crucial role in determining the model's goodness of fit than the kernel's shape. When the sample points in the research area are appropriately distributed, a kernel with a fixed bandwidth is a suitable choice for modeling. However, if the sample points exhibit clustering or irregular spacing, it is preferable to adapt the kernel's size, accordingly, increasing it when the sample points are sparse and decreasing it when they are dense. One approach to achieve this adaptive bandwidth specification is to choose a kernel that ensures an equal number of sample points for each estimation. This can be achieved by sorting the distances of the sample points from the desired regression point and including only the first p observations, where the optimal value of p is determined from the data. The weight for each observation within the bandwidth can be determined using the specified kernel, setting the weight to zero for observations outside the bandwidth and excluding them from the local calibration. An example of such a kernel is the bi-square kernel (Equation 2.8).

$$w_i(\mathbf{u}) = \left(1 - \left(\frac{d_i(\mathbf{u})}{h}\right)^2\right)^2 \quad (2.8)$$

Where $d_i(\mathbf{u}) > h$, $w_i(\mathbf{u})$ is zero. This is a near-Gaussian function with the helpful property of having a zero weight at a finite distance. In ArcGIS, the fixed radius kernel is implemented as a Gaussian kernel, while the adaptive kernel utilizes the bi-square kernel.

2.7.5 Model Calibration and Diagnosis

When the sample and regression points overlap, the residuals and predictions of the dependent variables are provided, allowing for the assessment of the model's goodness of fit. The traditional global model typically utilizes the R^2 or adjusted R^2 value as a measure of goodness of fit, with the adjusted value being preferred in model comparisons as it considers the number of variables or parameters in the model. However, when utilizing GWR, the computation of a goodness-of-fit value

becomes more complex, requiring consideration of the effective number of parameters in the model.

The hat matrix, denoted as S , plays a significant role in regression modeling. By multiplying the observed y values by S , Upon obtaining the predicted or fitted values, represented by Equation 2.9, the process involves the following steps:

$$\hat{y} = \mathbf{S}y \quad (2.9)$$

In a global model, the count of parameters can be determined by the trace of the matrix S , which is the summation of values along the leading diagonal. This is denoted as $\text{tr}(S)$. For an OLS regression, the number of parameters in the model is equivalent to the trace of S . In the case of a GWR model, the effective number of parameters is calculated using the formula $2\text{tr}(S) - \text{tr}(S^T S)$. The effective number of parameters in the model is influenced by both the number of independent variables and the bandwidth. This value is often non-integer and can be relatively large. Consequently, it serves as a valuable metric for assessing the fit of the model.

The corrected Akaike Information Criterion (Hurvich et al., 1998) in Equation 2.10 is a widely used measure of goodness of fit in GWR. It is represented as follows:

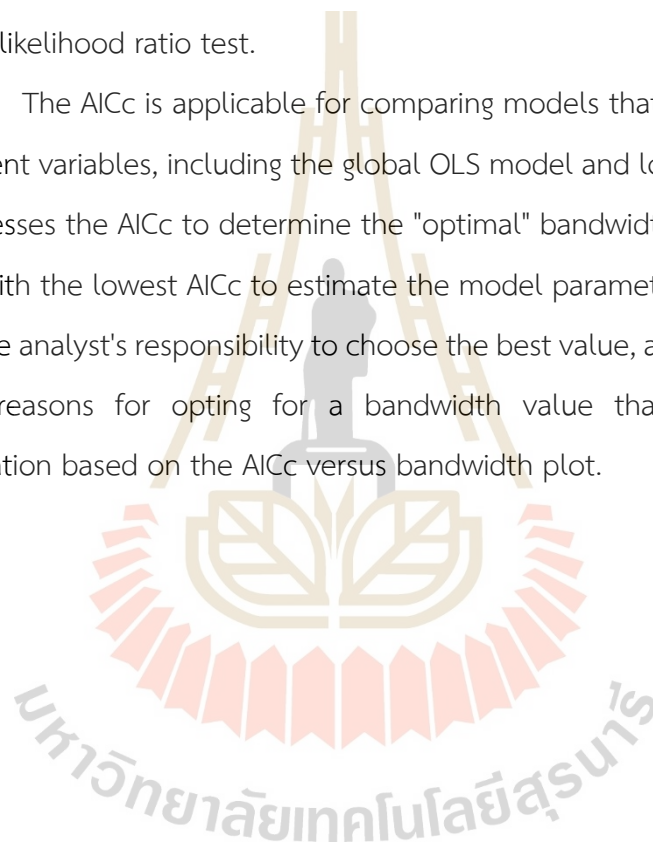
$$AIC_c = 2n \log_e(\hat{\sigma}) + n \log_e(2\pi) + n \left(\frac{n + \text{tr}(\mathbf{S})}{n - 2 - \text{tr}(\mathbf{S})} \right) \quad (2.10)$$

where the equation includes the variables n , $\hat{\sigma}$, and $\text{tr}(S)$, representing the number of observations, estimated standard deviation of the residuals, and the trace of the hat matrix, respectively. The AICc is a useful metric for comparing models that share the same dependent variable but have different independent variables. It considers the model's complexity, incorporating a penalty for degrees of freedom, especially when comparing models with significantly different right-hand sides.

The AICc is utilized to quantify the information discrepancy between the fitted model and the unknown true model. This discrepancy is referred to as the Kullback-Leibler information distance and is a relative measure rather than an absolute

one. According to a commonly accepted rule of thumb, two models are considered equivalent if the difference in their AICc values is less than 3, although a more cautious analyst may use a threshold of 4. The actual magnitudes of the AICc values displayed in the GWR output are not significant, as the focus lies on the differences between the AICc values. The AICc equation contains logarithmic terms, and it can be demonstrated that the disparity in AICc values for two models with identical degrees of freedom corresponds to the ratio of their likelihoods. However, it is important to note that the AICc is not a likelihood ratio test.

The AICc is applicable for comparing models that have distinct subsets of independent variables, including the global OLS model and local GWR models. The software assesses the AICc to determine the "optimal" bandwidth value, selecting the bandwidth with the lowest AICc to estimate the model parameters. Nevertheless, it is ultimately the analyst's responsibility to choose the best value, and there may be valid conceptual reasons for opting for a bandwidth value that deviates from the recommendation based on the AICc versus bandwidth plot.



CHAPTER III

RESEARCH METHODOLOGY

3.1 Study Area

The study site under examination is located in the northeastern part of Thailand, specifically in Nakhon Ratchasima Province. It spans from 818000E to 888000E and 1664000N to 1722000N, following the WGS 1984 UTM zone 47N coordinate system. The study area covers approximately 2,390 square kilometers and includes the districts of Non Thai, Non Sung, Non Daeng, Khong, and Kham Sakae Saeng within the province. This information is depicted in Figure 3.1. The region is among several areas impacted by the salinity of the Maha Sarakham Formation. In terms of climate, the average temperature in the area is 27.40 °C, with a relative humidity of 71% and an annual rainfall of 1,028.10 millimeters. The region experiences a distinct dry season from November to May, during which rainfall is limited. The dry season can be further divided into two parts: the cool season from November to February, characterized by chilly nights, and the hot season from March to May, with temperatures reaching up to 40 °C. The rainy season occurs from May to November, with the southwest monsoon prevailing and bringing the highest precipitation levels during this period. Rainfall during this season is often in the form of short showers lasting for an hour or two. As the rainy season progresses, rainfall intensifies, with peak levels typically observed in August and September.

3.2 Data Collection and Preparation

This study utilized secondary data for mapping groundwater potential. Fourteen thematic layers were obtained from reliable online sources (Table 3.1). Salinity, chloride, and total dissolved solids maps were sourced from Thongwat (2018).

The slope and lineament density maps were generated from ASTER GDEM data. The drainage density map was derived from stream networks using ASTER GDEM. The landforms map was acquired from the SRTM Landform dataset. The soil texture class map was obtained from the USDA system described by Hengl (2018). The land use/land cover map was sourced from the collaborative effort of Esri, Microsoft, and Impact Observatory for Sentinel-2 data. The normalized difference vegetation index map was derived from Sentinel-2 images. The annual rainfall map was obtained from the Royal Irrigation Department (RID). The geology, groundwater potential, and hydrogeological unit maps were obtained from the Department of Mineral Resources (DMR).

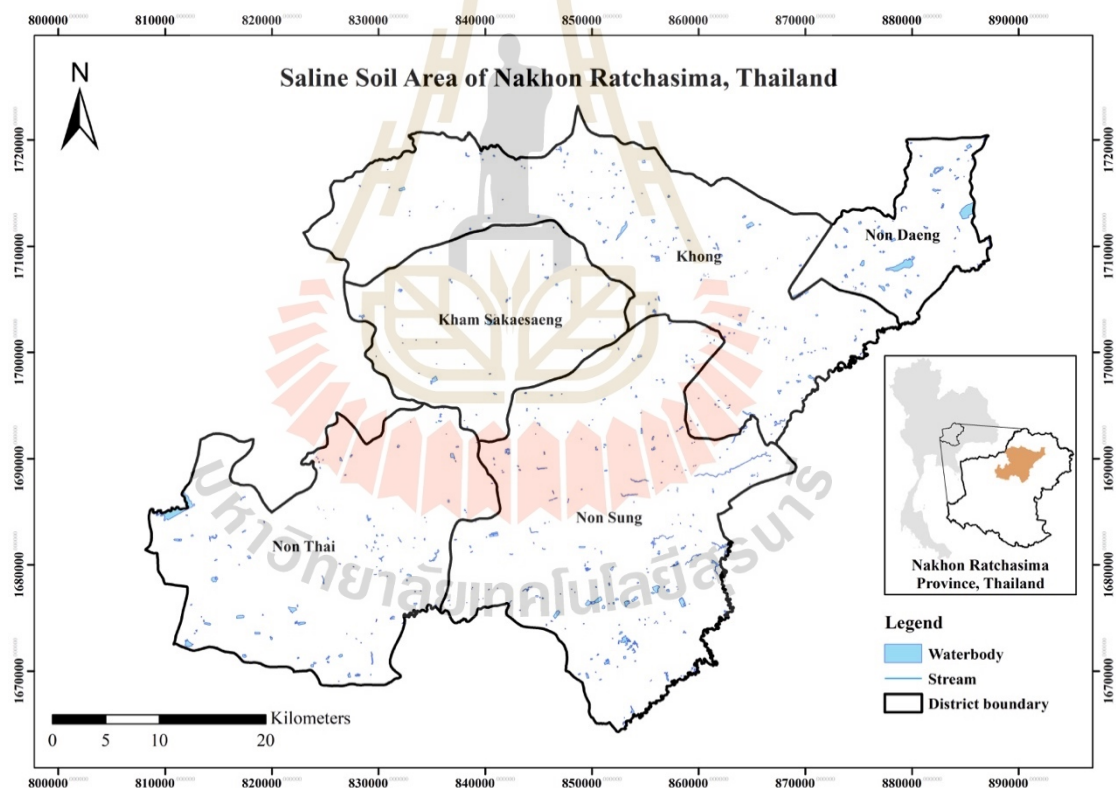


Figure 3.1 The study area of saline soils area in Nakhon Ratchasima province, Thailand.

Table 3.1 The data type and data sources of thematic layers.

No.	Thematic Layers	Data Type	Resolution	Source
1	Salinity content (SC)	XML	-	Thongwat (2018)
2	Chloride content (CC)	XML	-	Thongwat (2018)
3	Total dissolve solid (TDS)	XML	-	Thongwat (2018)
4	Slope (Sl)	Raster	30 m	ASTER GDEM
5	Drainage density (DD)	Raster	30 m	ASTER GDEM
6	Lineament density (LD)	Raster	30 m	ASTER GDEM
7	Landforms (Lf)	Raster	90 m	Theobald et al. (2015)
8	Soil texture class (STC)	Raster	10 m	Hengl (2018)
9	Land use /land cover (LULC)	Raster	10 m	Impact Observatory (2022)
10	Normalized difference vegetation index (NDVI)	Raster	10 m	Sentinel-2
11	Annual rainfall (Rf)	XML	-	RID
12	Geology (Geol)	WMS	1:50,000	DMR
13	Hydrogeological unit (HU)	WMS	1:50,000	DMR
14	Groundwater potential (GP)	WMS	1:50,000	DMR

Note: Information in the bracket is variable abbreviated names used in the analysis.

The presence of groundwater is governed by factors such as geology, structure, geomorphology, and drainage density, while the recharge of groundwater is influenced by land use, rainfall patterns, and the rate of infiltration (Manjare, 2014). The selection of factors used in groundwater studies may vary depending on the data availability for each researcher. Key factors include slope, drainage density, lineament density, landforms, annual rainfall, geology, hydrogeological units, groundwater potential, soil texture classes, land use/land cover, normalized difference vegetation index, salinity content, chloride content, and total dissolved solids. Each thematic layer was projected to the WGS_1984_UTM Zone_47N coordinate system for consistency.

In this study, all the thematic layers were employed to conduct an analysis of groundwater potential zones. The resulting classifications of these layers are presented in Table 3.2.

3.2.1 Dependent Variable

1) Salinity Content Map

The salinity content map was classified into five categories (“Understanding Salinity”, 2020) : 4.42 km² (0.19%) of Freshwater (< 0.50 ppt), 782.41 km² (32.76%) of Marginal water (0.50 – 1.00 ppt), 853.43 km² (35.73%) of Brackish water (1.00 – 2.00 ppt), 728.07 km² (30.48%) of Saline water (2.00 - 10.00 ppt), and 20.18 km² (0.84%) of Highly saline water (10.00 - 30.00 ppt) as shown in Table 3.2 and Figure 3.2.

3.2.2 Independent Variable

1) Chloride Content Map

The Chloride content map was classified into six classes according to Thongwat (2018): 0.22 km² (0.01%) of Freshwater (< 150.00 ppm), 4.36 km² (0.18%) of Fresh-brackish water (150 – 300 ppm), 613.72 km² (25.67%) of Brackish water (300 – 1,000 ppm), 1,639.48 km² (68.57%) of Brackish-salt water (1,000 – 10,000 ppm), 103.60 km² (4.33%) of Saltwater (10,000 – 20,000 ppm), and 29.52 km² (1.23%) of Hyper-halite water (> 20,000 ppm) as shown in Table 3.2 and Figure 3.3.

2) Total Dissolved Solid Map

The Total dissolved solid map was classified into three (Pradhan & Pirasteh, 2011): 623.77 km² (26.09%) of Freshwater (0 – 1,000 ppm), 1,747.19 km² (73.09%) of Brackish water (1,000 – 10,000 ppm), and 19.60 km² (0.82%) of Saline water (10,000 – 100,000 ppm) as shown in Table 3.2 and Figure 3.4.

3) Slope Map

The slope map was classified into two categories (Elewa & Qaddah, 2011): 2,368.69 km² (99.32%) of Flat (0 – 2 %), and 16.11 km² (0.68%) of Undulating (2 - 8 %) as shown in Table 3.2 and Figure 3.5.

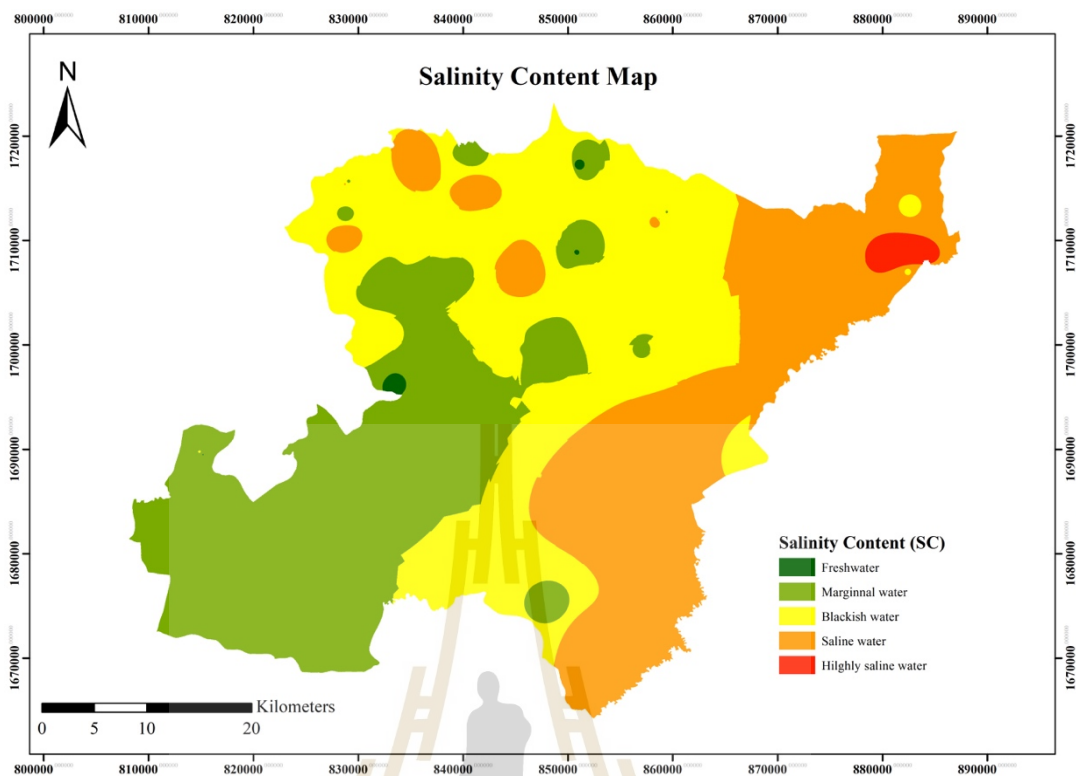


Figure 3.2 Salinity content.

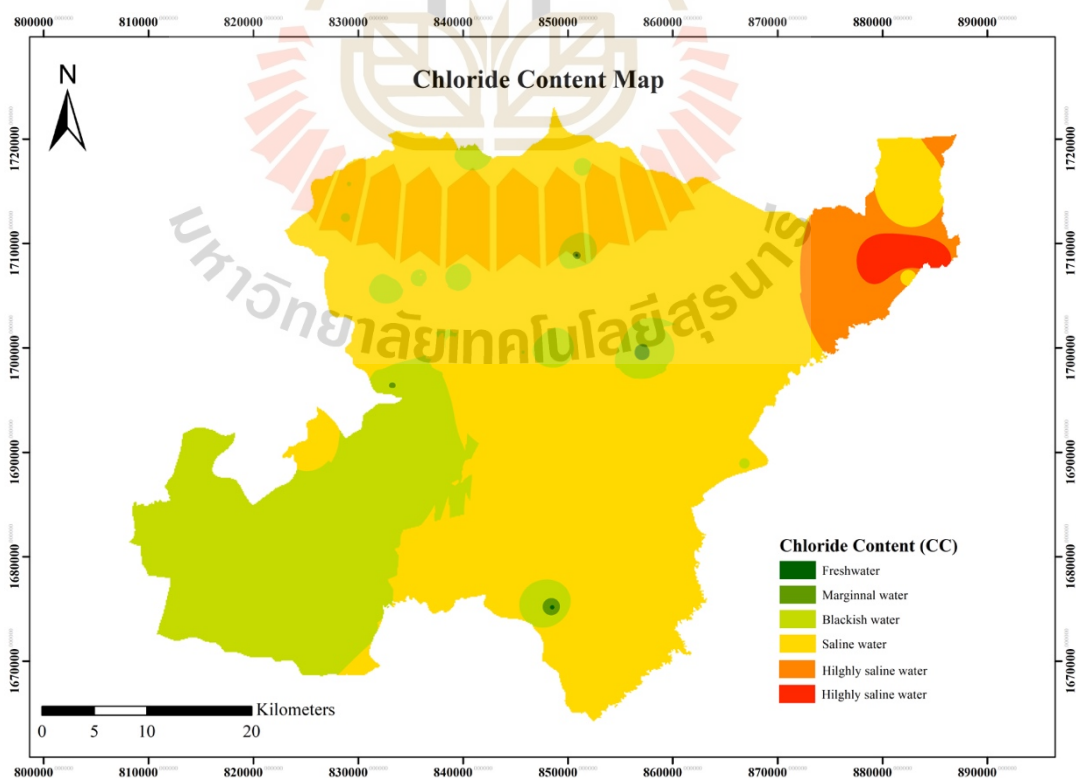


Figure 3.3 Chloride content.

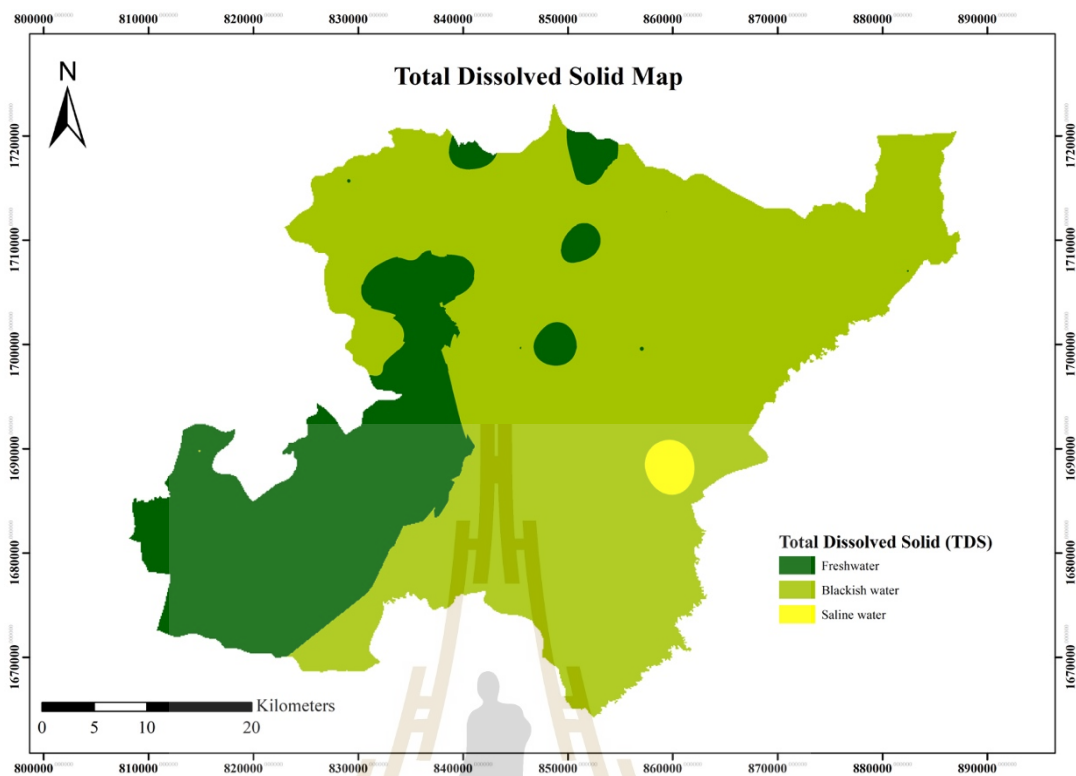


Figure 3.4 Total dissolved solids.

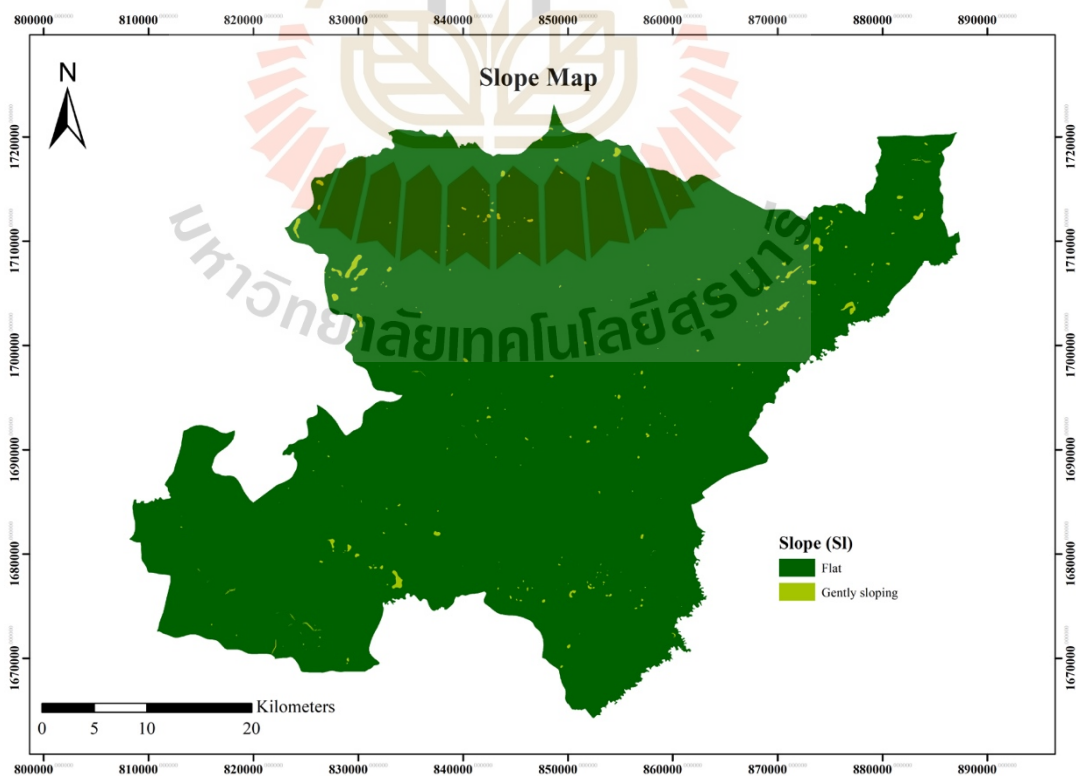


Figure 3.5 Slope.

4) Drainage Density Map

The drainage density map was classified into three density classes (Aldharab, Ali, Ikbal, & Ghareb, 2019): 1,805.11 km² (75.53%) of Very coarse drainage density (< 2 km/km²), and 584.73 km² (24.47%) of Coarse drainage density (2 – 4 km/km²) as shown in Table 3.2 and Figure 3.6.

5) Lineament Density Map

The lineament density map is reclassified into two classes: 2,279.81 km² (95.38%) of Absent, and 110.43 km² (4.62%) of Present as shown in Table 3.2 and Figure 3.7.

6) Landform Map

The landform data from the SRTM Landform dataset were collected and processed using the Google Earth engine and the Python API, and Theobald et al. (2015) assigned different weightage to the data based on its influence on groundwater occurrence. The landform map of the study area is covered by a variety of landform features 11.54 km² (0.48%) of upper slope (warm), 888.30 km² (37.17%) of upper slope (flat), 4.45 km² (0.19%) of lower slope (warm), 1,084.20 km² (45.37%) of lower slope (flat), 385.52 km² (16.13%) of the valley, and 15.62 km² (0.65%) of the valley (narrow) as shown in Table 3.2 and Figure 3.8.

7) Soil Texture Class Map

Hengl (2018) categorized the soil texture class into five distinct classes, with approximately 809.42 km² covering more than one-third of the area (33.87%) of Sandy clay loam, 714.10 km² (29.88%) of Loam, 527.55 km² (22.07%) of Sandy loam, 328.14 km² (13.73%) of Clayey loam, and 10.66 km² (0.45%) of Clay as shown in Table 3.2 and Figure 3.9.

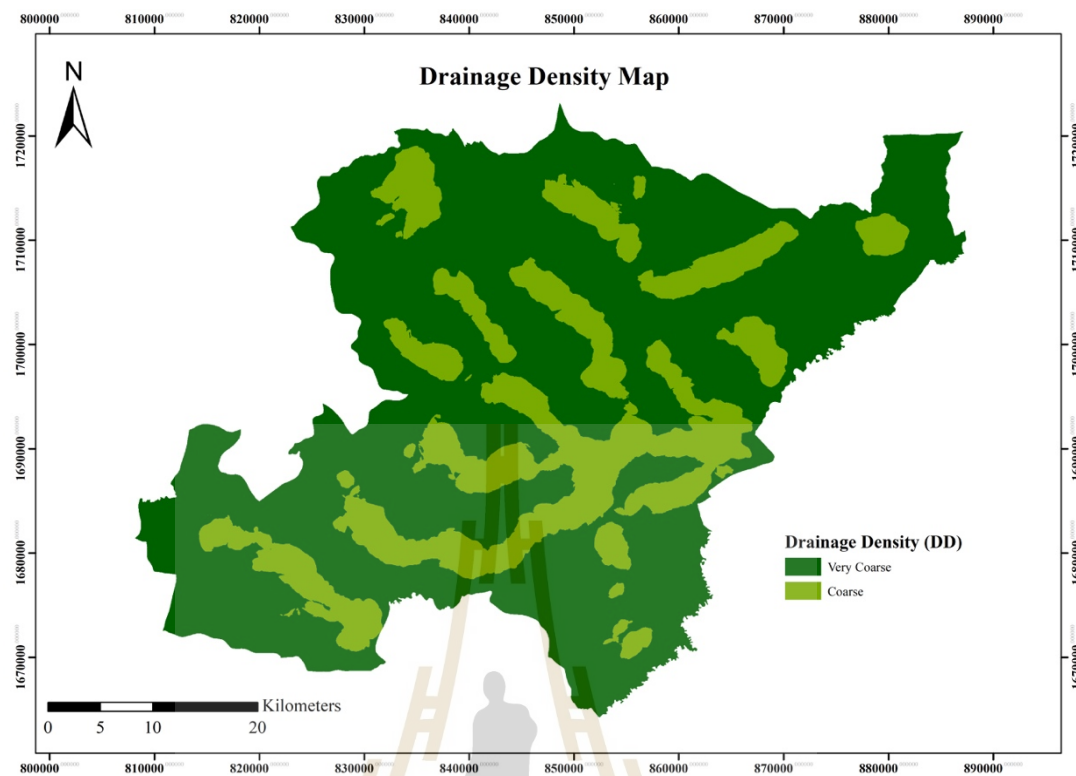


Figure 3.6 Drainage density.

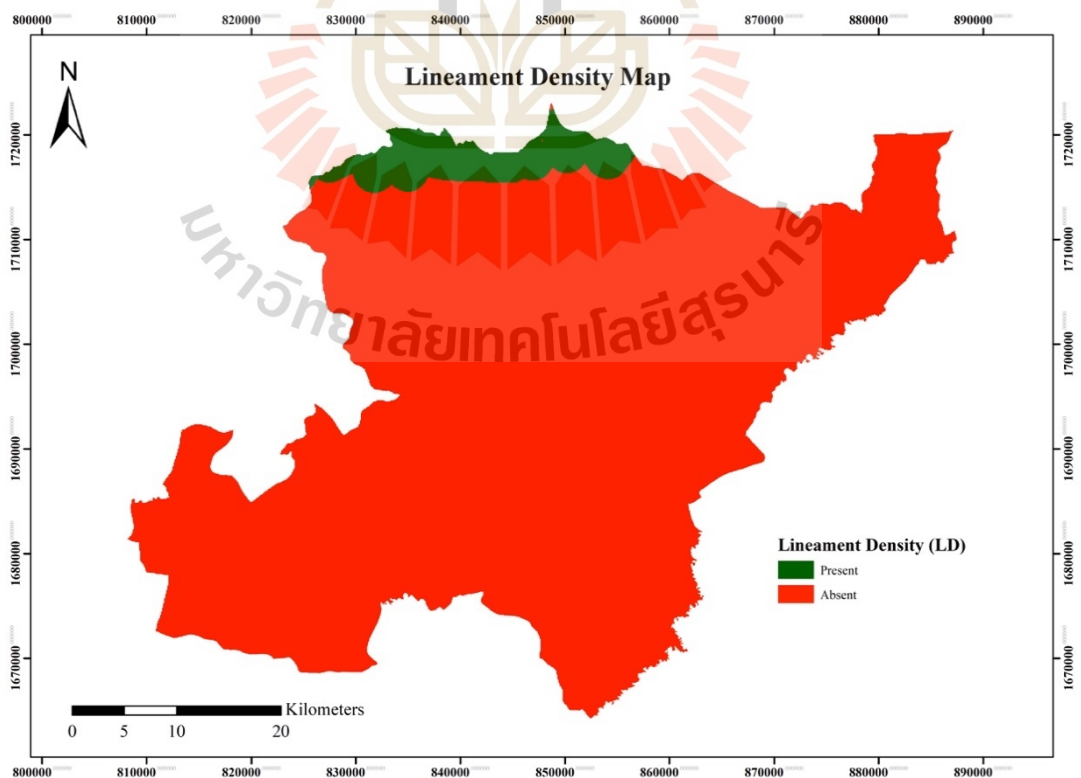


Figure 3.7 Lineament density.

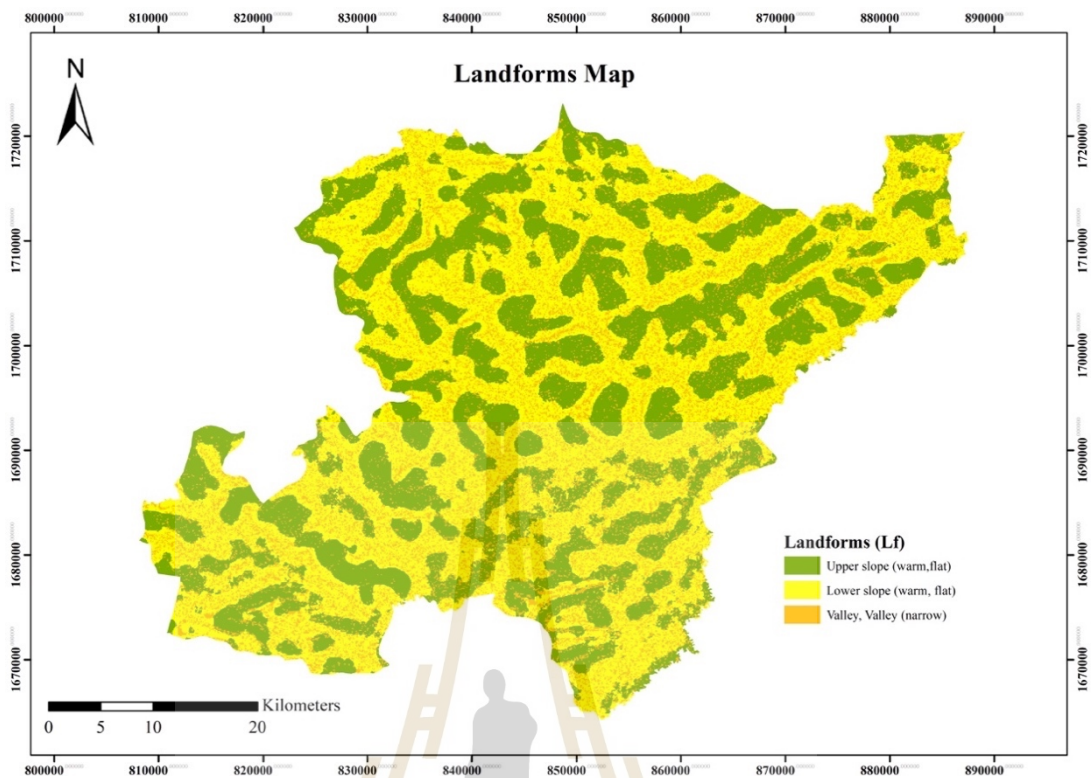


Figure 3.8 Landforms.

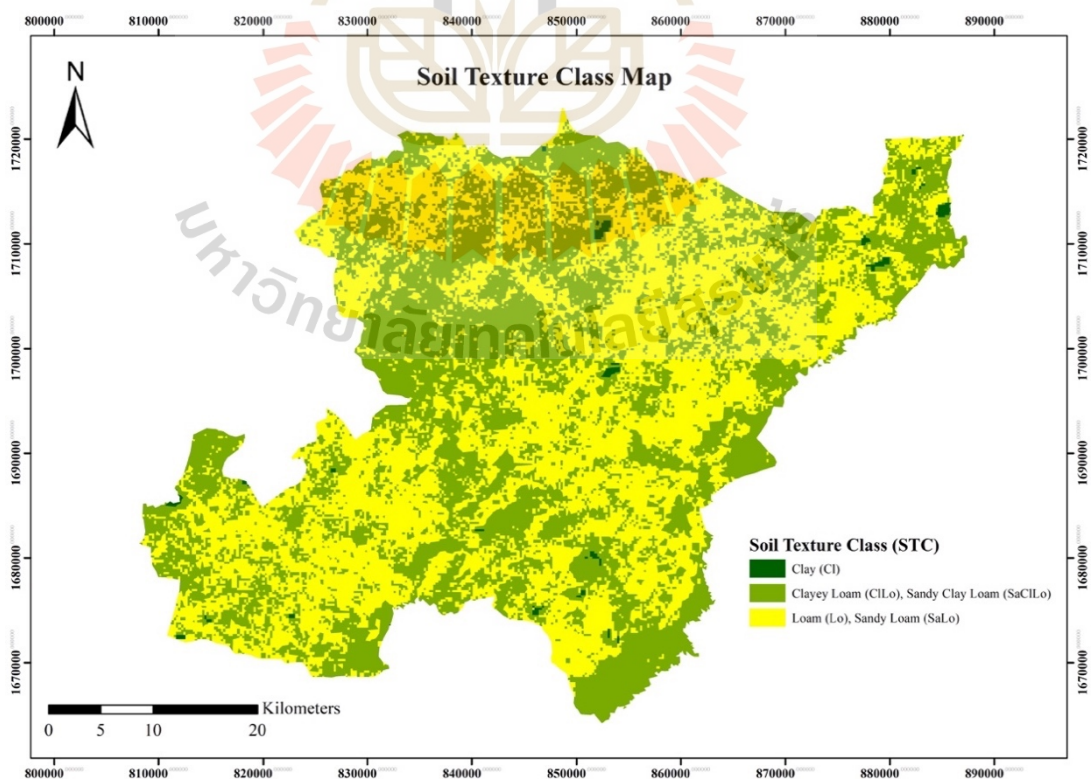


Figure 3.9 Soil texture class.

8) Land Use /Land Cover Map

Based on the Sentinel-2 10m Land Use/Land Cover Time Series dataset, the LULC map was categorized into seven classes according to Impact Observatory (2022), with 24.91 km² (1.04%) labeled as no data, 11.29 km² (0.47%) of water, 1.26 km² (0.05%) of grass, 1,928.06 km² (80.67%) of flooded vegetation, 221.71 km² (9.28%) of scrub/shrub, 1.42 km² (0.06%) of the build area, and 201.53 km² (8.43%) of cloud as shown in Table 3.2 and Figure 3.10.

9) Normalized Difference Vegetation Index Map

The NDVI was reclassified into five classes (Atun, Kalkan, & Gürsoy, 2020): 17.57 km² (0.74%) of water, 3.23 km² (0.14%) of bare soil, 2,106.69 km² (88.15%) of sparse vegetation, 262.24 km² (10.97%) of moderate vegetation, 0.05 km² (<0.01%) of dense vegetation as shown in Table 3.2 and Figure 3.11.

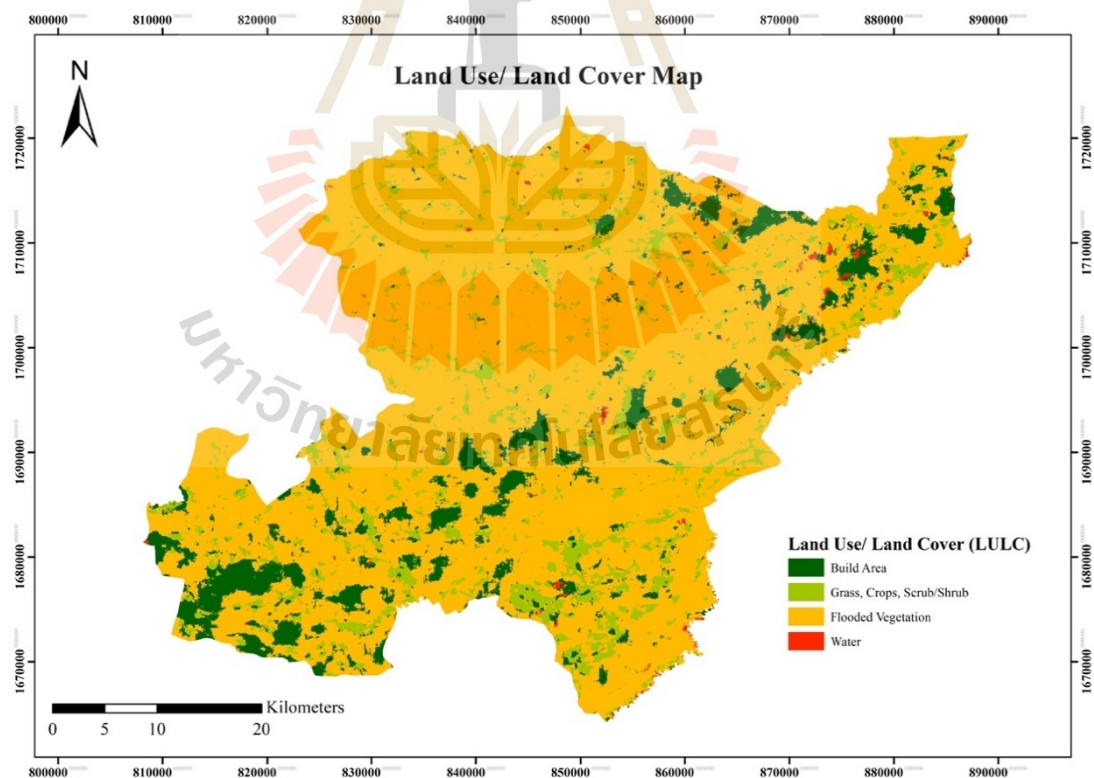


Figure 3.10 Land use /land cover.

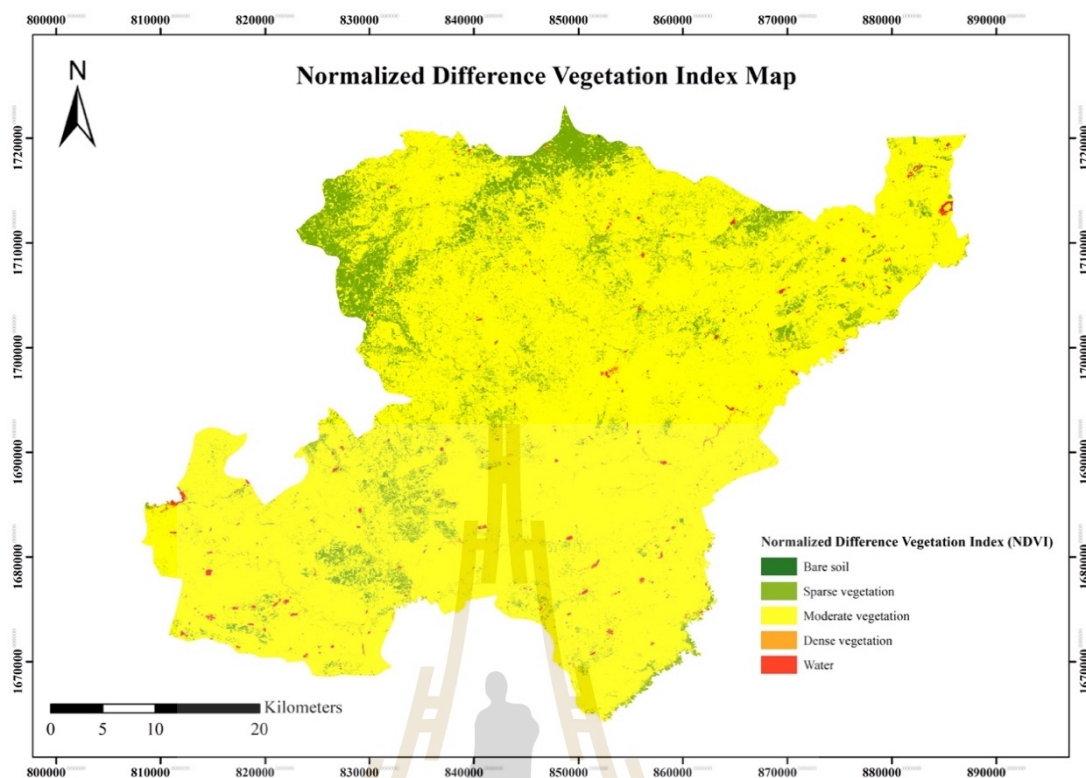


Figure 3.11 Normalized difference vegetation index

10) Annual Rainfall Map

The annual rainfall map was projected and divided into four using the natural breaks (Jenks) classification method: 154.45 km² (6.47%) of Low rainfall (186–193 millimeters), 722.25 km² (30.23%) of Medium rainfall (193–199 millimeters), 890.51 km² (37.28%) of High rainfall (199–204 millimeters), and 621.75 km² (26.03%) of Very High rainfall (204–206 millimeters) as shown in Table 3.2 and Figure 3.12.

11) Geology Map

The geology map was obtained from the Web Feature Service (WFS), digitized, and categorized into four geological units (Department of Mineral Resources, n.d.). These units consist of 190.24 km² (7.96%) of Pleistocene River terrace deposits (Q1), 1,266.95 km² (53.00%) of Quaternary Alluvial deposits (Q2), 10.57 km² (0.44%) of Khok Kruat Formation (Kk), and 922.80 km² (38.60%) of Maha Sarakham Formation (KTms) as shown in Table 3.2 and Figure 3.13.

12) Hydrogeological Unit Map

The hydrogeological unit map was obtained, digitized, and categorized into two units according to groundwater quality. Each unit was assigned distinct scores (Department of Mineral Resources, n.d.) : 93.11 km² (3.89%) of Floodplain Deposits Aquifer (Qfd), and 2,297.45 km² (96.11%) of Maha Sarakham Aquifer (Ms) as shown in Table 3.2 and Figure 3.14.

13) Groundwater Potential Map

The groundwater potential map was generated by specific yield (S_y) and TDS data. It was then classified into four categories based on guidelines provided by Department of Mineral Resources (n.d.): 1936.65 km² (81.01%) of G1 (S_y: < 2 m³/h, TDS: 500 - 1,500 ppm), 385.41 km² (16.12%) of G2 (S_y: 2 – 10 m³/h, and TDS: 500 - 1,500 ppm), 59.96 km² (2.51%) of R1 (S_y: < 2 m³/h, and TDS: > 1,500 ppm), and 8.54 km² (0.36%) of R2 (S_y: 2 – 10 m³/h, and TDS: > 1,500 ppm) (Table 3.2 and Figure 3.15).

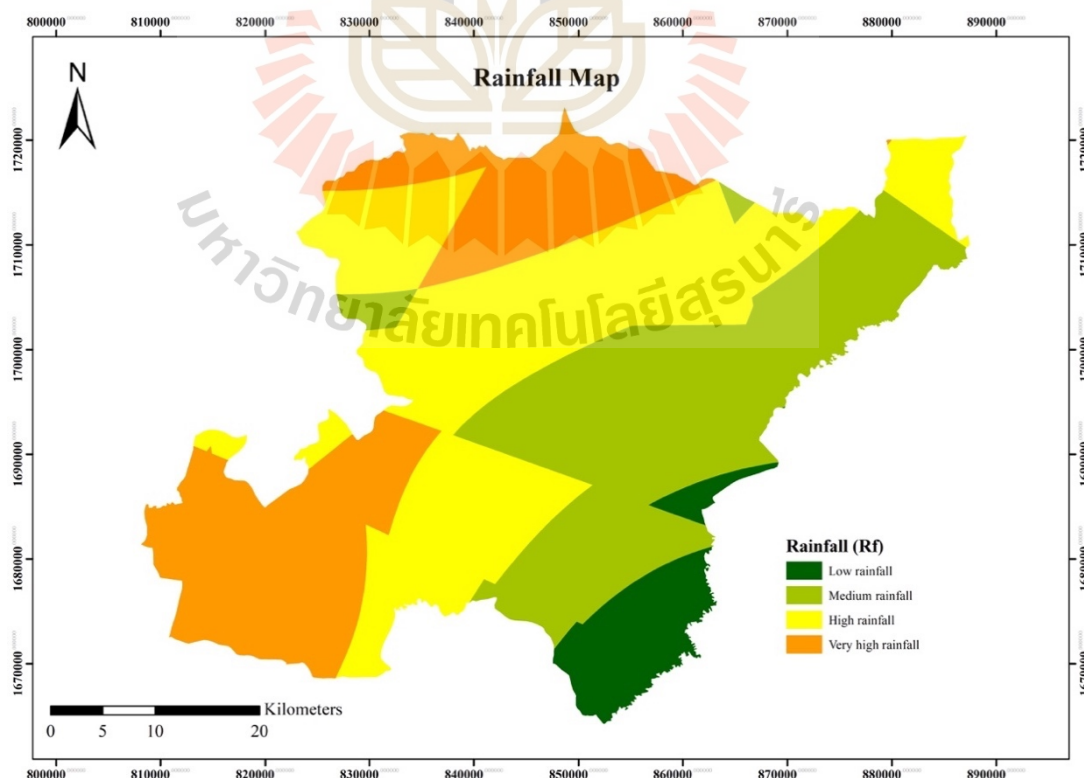


Figure 3.12 Annual rainfall

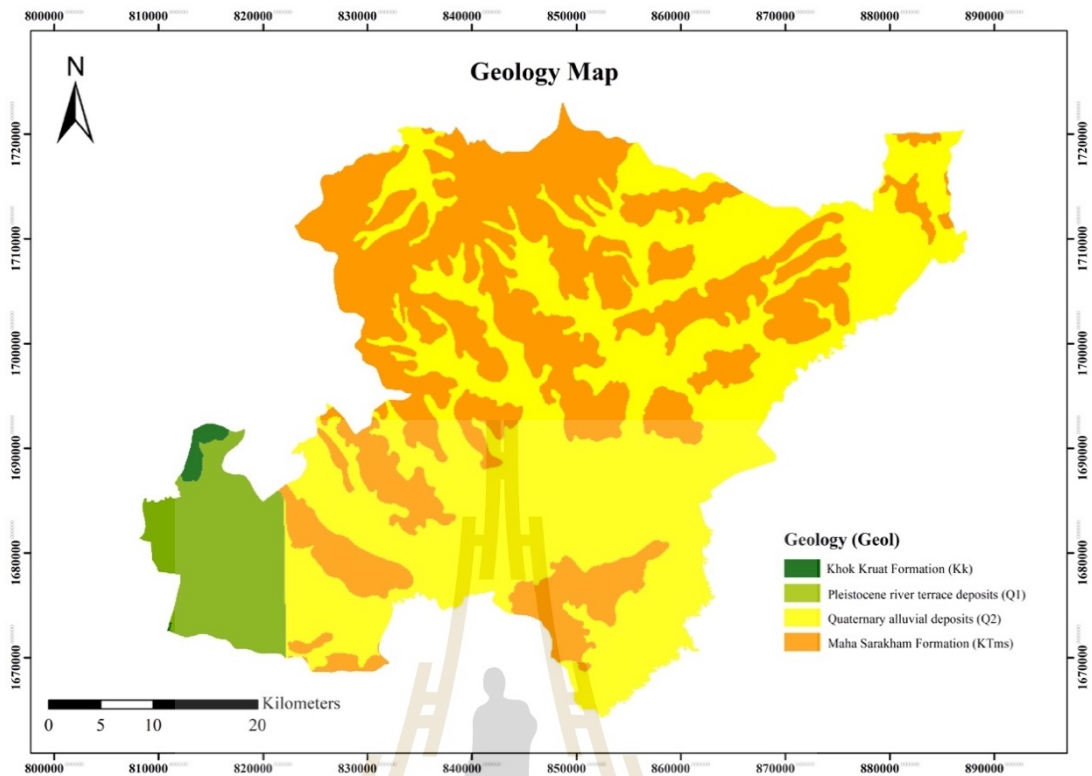


Figure 3.13 Geology.

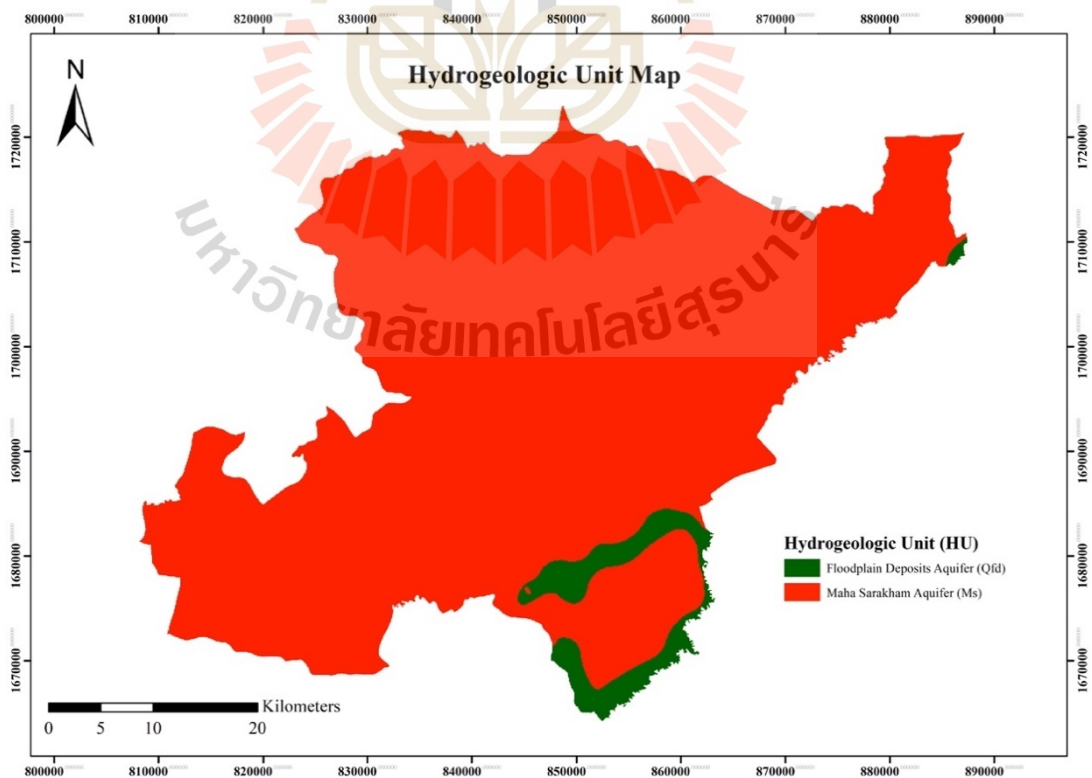


Figure 3.14 Hydrogeological unit.

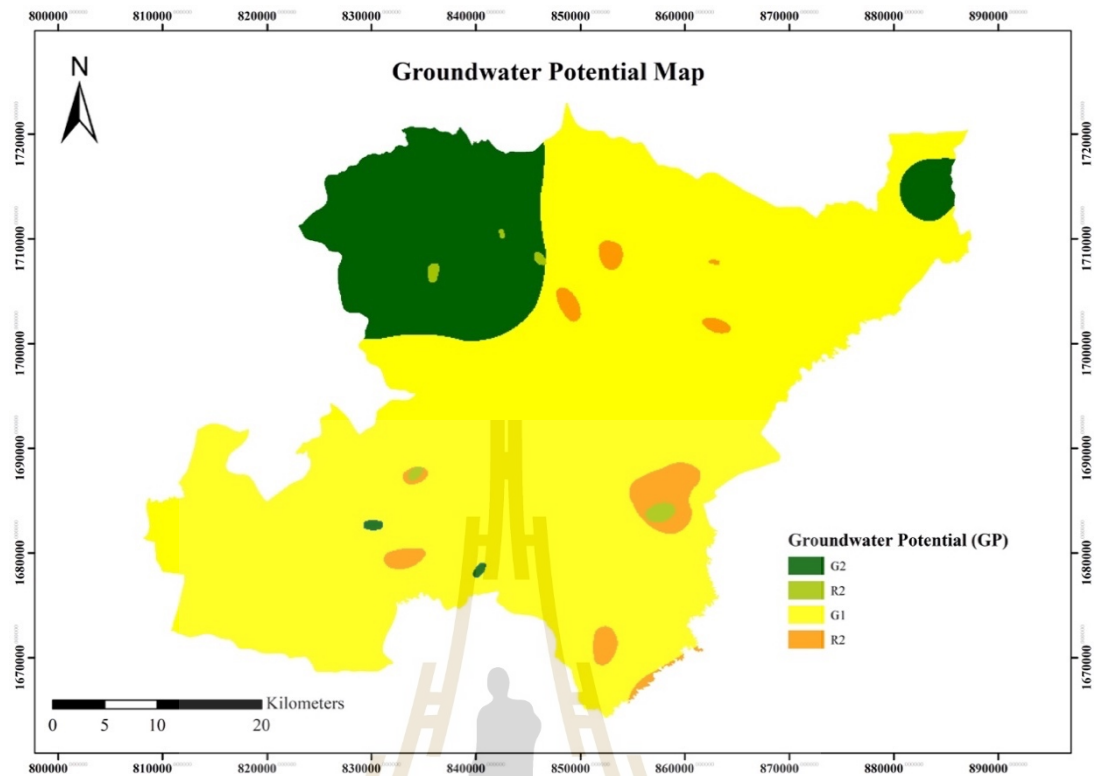


Figure 3.15 Groundwater potential.

Table 3.2 Classification of thematic layers.

No.	Thematic Layers	Feature classes	Rating	Area (%)
1	Salinity Content (ppt)	Freshwater	< 0.50	0.19
		Marginal water	0.50 - 1.00	32.76
		Brackish water	1.00 - 2.00	35.76
		Saline water	2.00 - 10.00	30.48
		Highly saline water	10.00 - 30.00	0.84
		Brine water	> 30.00	-

Table 3.2 Classification of thematic layers (Continued).

No.	Thematic Layers	Feature Classes	Rating	Area (%)
2	Chloride Content (ppm)	Freshwater	< 150	0.01
		Fresh-brackish water	150 – 300	0.18
		Brackish water	300 – 1,000	25.67
		Brackish-salt water	1,000 – 10,000	68.57
		Saltwater	10,000 – 20,000	4.33
		Hyper-halite water	> 20,000	1.23
3	Total Dissolved Solid (ppm)	Freshwater	0 – 1,000	26.09
		Brackish water	1,000 – 10,000	73.09
		Saline water	10,000 – 100,000	0.82
		Brine water	> 100,000	-
4	Slope (%)	Flat	0 - 2	99.32
		Undulating	2 - 8	0.68
		Rolling	8 - 15	-
		Moderately steep	15 - 30	-
5	Drainage density (km/km ²)	Very coarse	< 2	75.53
		Coarse	2 - 4	24.47
		Moderate	4 - 6	-
		Fine	6 - 8	-
		Very fine	> 8	-
6	Lineament Density	Absent	1	95.38
		Present	2	4.62

Table 3.2 Classification of thematic layers (Continued).

No.	Thematic Layers	Feature Classes	Rating	Area (%)
7	Landforms	Peak/ridge (warm)	11	-
		Peak/ridge	12	-
		Peak/ridge (cool)	13	-
		Mountain/divide	14	-
		Cliff	15	-
		Upper slope (warm)	21	0.48
		Upper slope	22	-
		Upper slope (cool)	23	-
		Upper slope (flat)	24	37.17
		Lower slope (warm)	31	0.19
		Lower slope	32	-
		Lower slope (cool)	33	-
		Lower slope (flat)	34	45.37
		Valley	41	16.13
Valley (narrow)	42	0.65		
8	Soil texture class	Clay	1	0.45
		Silty Clay	2	-
		Sandy Clay	3	-
		Clayey Loam	4	13.73
		Silty Clay loam	5	-
		Sandy Clay Loam	6	33.87

Table 3.2 Classification of thematic layers (Continued).

No.	Thematic Layers	Feature Classes	Rating	Area (%)
		Loam	7	29.88
		Silty Loam	8	-
		Sandy Loam	9	22.07
		Silt	10	-
		Loamy Sand	11	-
		Sand	12	-
9	Land use /land cover	No Data	1	1.04
		Water	2	0.47
		Trees	3	-
		Grass	4	0.05
		Flooded Vegetation	5	80.67
		Crops	6	-
		Scrub/Shrub	7	9.28
		Built Area	8	0.06
		Bare Area	9	-
		Snow/Ice	10	-
		Clouds	11	8.34
10	Normalized difference vegetation index	Water	< 0.00	0.74
		bare soil	0.00 – 0.03	0.14
		sparse vegetation	0.03 – 0.30	88.15
		moderate vegetation	0.30 – 0.50	10.97

Table 3.2 Classification of thematic layers (Continued).

No.	Thematic Layers	Feature Classes	Rating	Area (%)
		dense vegetation	> 0.50	< 0.01
11	Annual rainfall (mm)	Low rainfall	< 188	6.47
		Medium rainfall	188 – 196	30.23
		High rainfall	196 – 202	37.28
		Very High rainfall	202 – 216	26.03
12	Geology	Kk	1	0.44
		KTms	2	38.60
		Q1	3	7.96
		Q2	4	53.00
13	Hydrogeological unit	Qfd	1	3.89
		Ms	2	96.11
14	Groundwater potential (m ³ /h, ppm)	G2	S _y : 2 - 10 TDS: 500 - 1,500	16.12
		R2	S _y : 2 - 10 TDS: > 1,500	0.36
		G1	S _y : < 2 TDS: 500 - 1,500	81.01
		R1	S _y : < 2 TDS: > 1,500	2.51

3.2.3 Data Preparation

1) Fishnet tool

Grids are used to display coordinates or divide areas in a map. Create a grid to analyze the raster using the create fishnet tool in data management tools into 1,000 m x 1,000 m according to software limitation. The output feature class contains the fishnet of rectangular cells and clip using study area extent as shown in Figure 3.16.

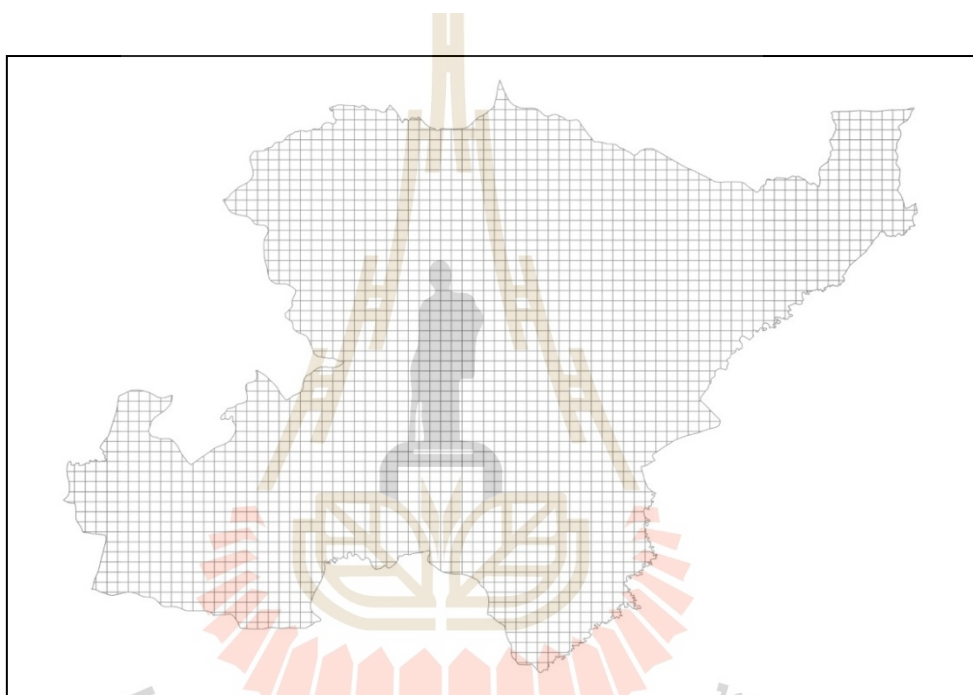


Figure 3.16 Grid cell size 1,000 m x 1,000 m.

2) Add Geometry Attributes Tool

Adds new attribute fields to the input features representing each feature's spatial or geometric characteristics and location, such as length or area and x-, y-, z-, and m-coordinates. This study uses the field (CENTROID_INSIDE) on Add Geometry Attributes tool in the data management tool.

3) Make XY Event Layer Tool

Creates a new point feature layer based on x- and y-coordinates defined in a table. If the source table contains z-coordinates (elevation values), that

field can also be specified in the creation of the event layer. The layer created by this tool is temporary. Make XY event layer tool in data management tool as shown in Figure 3.17 shows the inside centroid of each polygon.

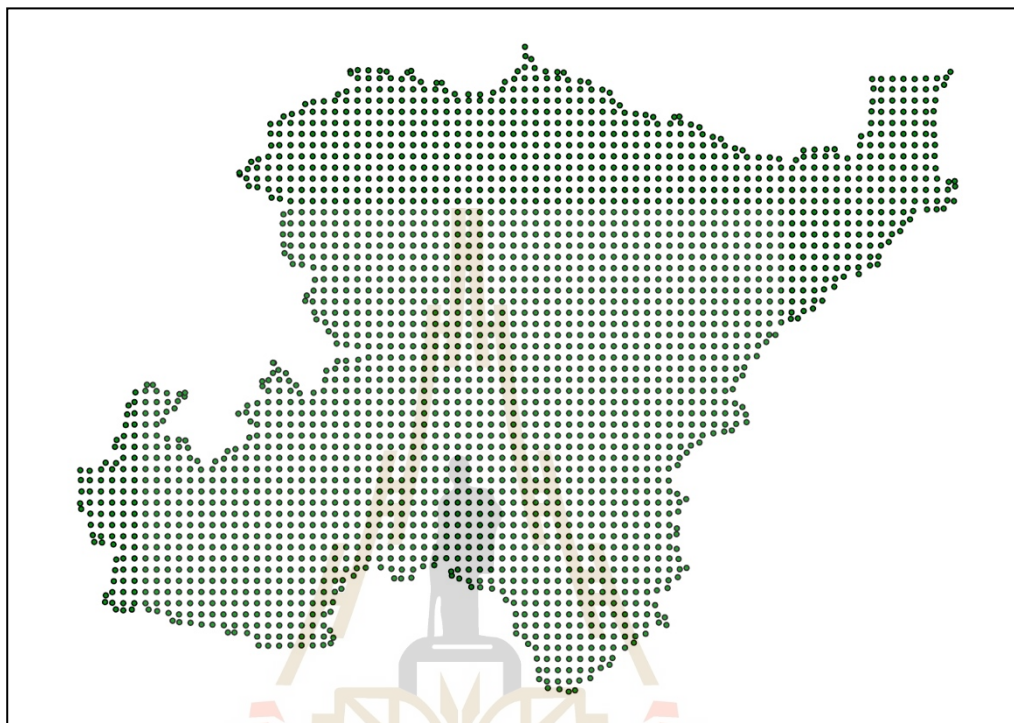


Figure 3.17 Converting polygon features to point features.

4) Extract Multi Values to Points Tools

Extract Multi Values to Points Tools in Spatial Analyst Tool can extract cell values at locations specified in a point feature class from one or more raster and record the values to the attribute table of the point feature class. If the field is populated with a zero or null value, that feature will be removed from the database.

3.3 Regression Analysis Model Development

3.3.1 Exploratory Spatial Data Analysis (ESDA)

Following the data mapping process, the next step in data exploration involves utilizing Exploratory Spatial Data Analysis (ESDA) tools. These tools provide a

quantitative approach to examining the data beyond mere mapping, enabling a more comprehensive understanding of the investigated phenomena. This deeper understanding aids in making informed decisions regarding the construction of the interpolation model. To assist in achieving these objectives, the ESDA tools provide various perspectives of the data. These perspectives can be manipulated and examined, creating an interconnected relationship between them and the data displayed in ArcMap through brushing and linking.

1) Histograms

The Histogram tool offers a way to examine data in a univariate manner. By displaying the frequency distribution and providing summary statistics, it allows you to gain insights into the dataset of interest.

The frequency distribution, depicted as a bar graph, illustrates the occurrence of values within specific intervals or classes. Determination of the number of classes with equal width to be utilized in the histogram is flexible. The height of each bar represents the relative proportion of data falling within each class.

In some cases, it may be necessary for the data to follow a normal distribution for certain analytical methods. If the data exhibits skewness, indicating an asymmetric distribution, it might be beneficial to apply data transformation techniques to achieve a normal distribution.

2) Transform Field Tool

The transformation process modifies the continuous values in one or multiple fields by employing mathematical functions on each value, thereby altering the distribution's shape. The available transformation methods encompass log, square root, Box-Cox, multiplicative inverse, square, exponential, and inverse Box-Cox.

Applying a transformation can effectively mitigate skewness within the distribution and yield a normal (Gaussian) distribution, as depicted in Figure 3.18.

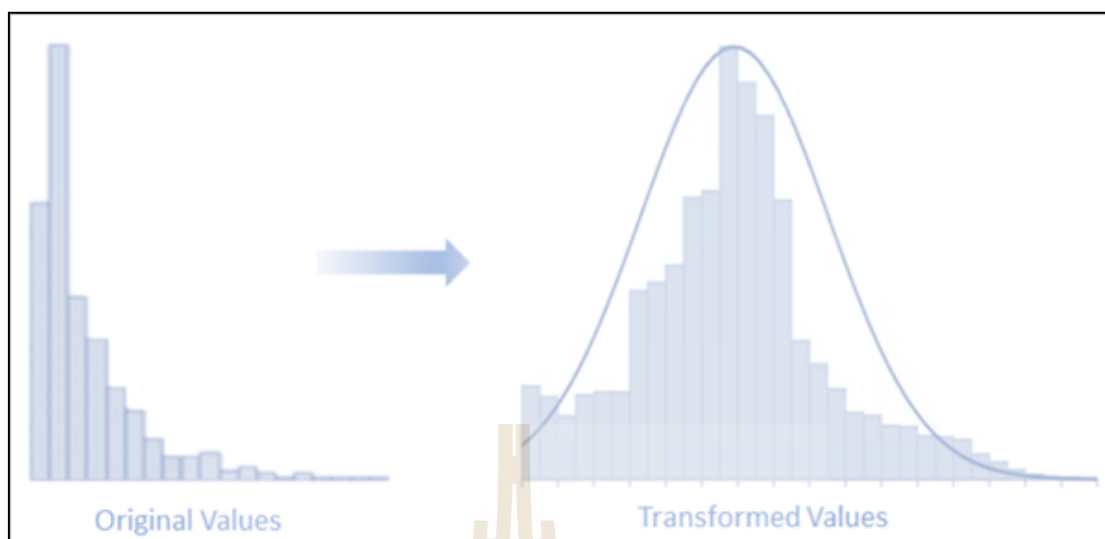


Figure 3.18 Shows original values are transformed to be closer to a normal distribution.

The Box-Cox transformation techniques utilize a specific power function, as expressed in equation 3.1, to achieve the normalization of the data within the chosen fields:

$$x' = \begin{cases} \frac{(x + \lambda_2)^{\lambda_1} - 1}{\lambda_1}, & \lambda_1 \neq 0 \\ \ln(x + \lambda_2), & \lambda_1 = 0 \end{cases} \quad (3.1)$$

In equation 3.1, the transformed value (x') is obtained from the original value (x) by applying the power (exponent) parameter (λ_1) and the shift parameter (λ_2)

The Box-Cox transformation is applicable only to positive values. In case the selected fields contain negative or zero values, a default shift is applied to make the values positive before the transformation. The default shift value is determined by the maximum absolute negative value in the field, ensuring nonnegative values. Additionally, a small infinitesimal value ($\sim 10^{-6}$) is added to make the transformed values nonzero. The Power parameter allows specifying the desired power value for the transformation, which can range from -5 to 5. If no specific value

is provided, the geoprocessing messages will display the best approximation of a normal distribution curve.

3.3.2 Fit an OLS Model and Perform Diagnostics

1) Calculate Distance Band from Neighbor Count Tool

The Calculate Distance Band from Neighbor Count Tool, which is one of the Spatial Statistics Tools, can be used to determine the minimum, maximum, and average distances required to ensure that each feature in a set has at least N neighbors. The input parameter N specifies the number of neighbors to be considered. The maximum distance output value from this tool indicates the distance that needs to be traveled away from each feature to ensure that every feature has at least N neighbors. On the other hand, the minimum distance output value is the distance from each feature to ensure that at least one feature has N neighbors. The average distance output value represents the average distance from each feature to find its N nearest neighbors.

This tool is useful for determining a neighborhood distance band or threshold distance value. The maximum distance output value obtained from this tool can be used as the distance band or threshold distance parameter value for other spatial analysis tools. This way, the distance band parameter can be set appropriately to ensure that the analysis is carried out at the appropriate spatial scale and that the results are reliable.

2) Incremental Spatial Autocorrelation

Spatial autocorrelation refers to the degree to which the values of a variable are correlated in space, that is, the extent to which nearby locations tend to have similar values of the variable being studied. This is an important concept in spatial analysis because it can provide insights into the underlying spatial processes that are responsible for the distribution of the variable. One tool that is commonly used to measure spatial autocorrelation is the Incremental Spatial Autocorrelation Tool.

This tool, which is part of the Spatial Statistics Tools in ArcGIS software, allows users to measure spatial autocorrelation for a series of distances. The tool also generates a line graph of those distances and their corresponding z-scores. Z-scores are a measure of the intensity of spatial clustering, with statistically significant peak z-scores indicating distances where spatial processes promoting clustering are most pronounced. These peak distances can be used as appropriate values for tools with a Distance Band or Distance Radius parameter.

The Incremental Spatial Autocorrelation Tool is particularly useful because it allows for the identification of specific distances at which spatial clustering is strongest. This information can be used to inform subsequent analyses and to gain a better understanding of the underlying spatial processes that are driving the observed patterns in the data. By using this tool, researchers can gain valuable insights into the spatial distribution of variables and can develop more accurate models to explain and predict spatial phenomena.

3) **Generate Spatial Weights Matrix**

In spatial analysis, a spatial weights matrix plays a crucial role in representing the spatial structure of data by quantifying the spatial relationships between the features in the dataset. The selection of an appropriate conceptualization for the spatial weights matrix is essential, as it reflects how the features interact with each other. The Generate Spatial Weights Matrix tool is used to create an SWM file that contains spatial relationship values using sparse matrix techniques, which reduces disk space, computer memory, and the number of calculations required. These values are utilized in various spatial statistics tools such as Spatial Autocorrelation, Hot Spot Analysis, and Cluster and Outlier Analysis. While the SWM file can store spatial relationships for all features in the dataset, in practice, it is advisable to relate each feature to only a few others. This is achieved through the sparse methodology, which stores only nonzero relationships, resulting in a smaller file size and faster calculations.

4) Exploratory Regression

The Exploratory Regression tool in the ArcGIS software is a useful technique to determine the best-fit model. This toolbox searches for the optimal model by evaluating all possible combinations of the explanatory variables based on various model performance indicators, such as R^2 , VIF, AIC, and JB p-values. The best-fit model identified by the Exploratory Regression tool is then validated using the stepwise variable selection method, which involves assessing changes in model diagnostics after adding or deleting variables. This method ensures that the selected model is robust and provides reliable results.

5) Ordinary Least Squares Regression

Regression analysis is a fundamental statistical technique used extensively in the social sciences to identify and measure relationships between two or more feature attributes. This method is commonly employed to better understand the dynamics of a particular location, predict future trends or events, and examine the causes of why things occur where they do. Ordinary Least Squares (OLS) is the most well-known and widely used regression technique, serving as a foundation for all spatial regression analyses. OLS is used to develop a global model that captures the overall behavior of the variable or process under investigation. The approach involves fitting a single regression equation to the entire dataset, thereby creating a unified representation of the process. This method provides a valuable starting point for spatial analysts looking to model complex phenomena and develop a deeper understanding of the underlying relationships.

3.3.3 Geographically Weighted Regression

Geographically Weighted Regression (GWR) is a spatial regression technique used in various disciplines, including geography. It aims to evaluate a local model of the variable or process of interest by fitting a regression equation to every feature in the dataset. This approach enables us to consider the spatial heterogeneity of the relationships between variables across a study area. In the construction of

distinct equations, Geographically Weighted Regression (GWR) takes into account both the dependent and explanatory variables of the features located within the vicinity of each target feature. The size and range of each analyzed neighborhood are determined by the parameters of Neighborhood Type and Neighborhood Selection Method. It is crucial to select these parameters thoughtfully to ensure the credibility and reliability of the analysis. It should be emphasized that the GWR technique is well-suited for datasets with a considerable number of features, typically several hundred. Therefore, it is not an appropriate method for small datasets. Additionally, GWR does not work with multipoint data, which should be preprocessed into point data before analysis. Overall, GWR is a powerful tool for analyzing spatially varying relationships, allowing researchers to gain a more comprehensive understanding of the data and make informed decisions.

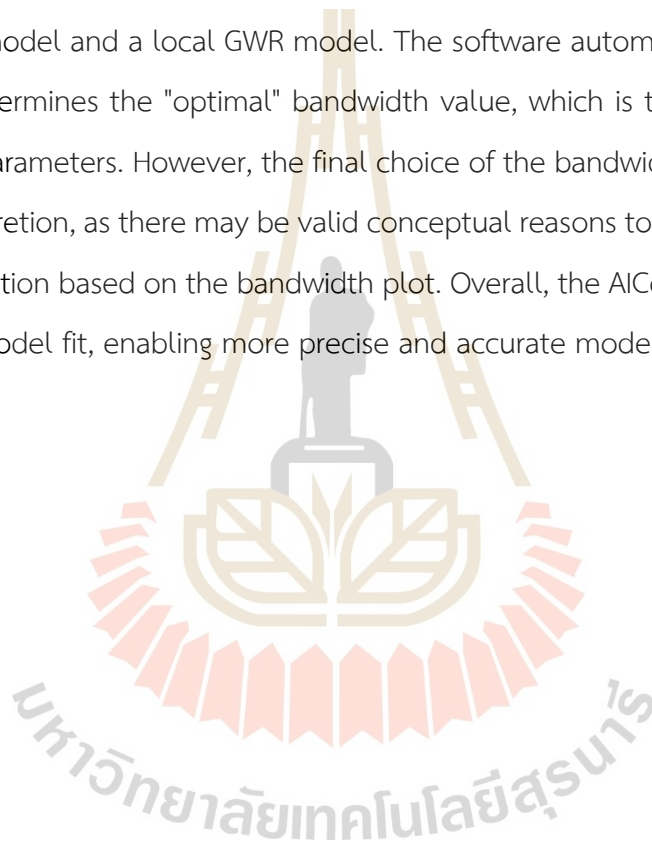
3.3.4 Compare Model

The alignment of sample and regression points enables the examination of residuals and predictions of the dependent variables, facilitating the evaluation of the model's goodness of fit. For traditional global models, the R^2 or adjusted R^2 value is commonly used as a measure of goodness of fit. When comparing different models, the adjusted R^2 is favored due to its consideration of the number of variables or parameters incorporated in the model. Models with more variables generally have higher R^2 values than those with fewer variables. Nevertheless, in the case of GWR, evaluating the goodness of fit becomes more intricate, necessitating the inclusion of the effective number of parameters in the model during the calculation of the goodness-of-fit measure.

The AICc, a statistical metric adjusted for small sample sizes, serves to compare the discrepancy between a fitted model and the underlying 'true' model. Termed the Kullback-Leibler information distance, this measure is not absolute in nature. As per a widely accepted guideline, if the disparity in AICc values between two distinct models is below 3 (or potentially 4 for more cautious analysts), they are

deemed equivalent. While the absolute values of AICc in GWR output may appear unexpectedly high or low, the differences in AICc values remain significant.

The AICc calculation involves logarithmic terms and can be shown that the disparity between AICc values for two models with equal degrees of freedom corresponds to the ratio of their likelihoods. Nevertheless, it is essential to emphasize that the AICc should not be considered as a likelihood ratio test. The AICc is commonly used to compare models with different sets of independent variables, including the global OLS model and a local GWR model. The software automatically calculates the AICc and determines the "optimal" bandwidth value, which is then used to estimate the model parameters. However, the final choice of the bandwidth value is left to the analyst's discretion, as there may be valid conceptual reasons to deviate from the AICc recommendation based on the bandwidth plot. Overall, the AICc is a valuable tool for evaluating model fit, enabling more precise and accurate model selection.



CHAPTER IV

RESULTS AND DISCUSSION

This chapter consists of 1) the comparative results of the OLS regression model and GWR model for salinity distribution in the Non Thai, Non Sung, Non Daeng, Khong, and Kham Sakae Saeng districts of Nakhon Ratchasima province, 2) investigate the spatial relationships between groundwater salinity and influencing factors.

4.1 Results of Global Model: Ordinary Least Squares (OLS) Regression Analysis

Based on the OLS model regression coefficient estimates, along with the Standard Error, t-Statistic, Probability, Robust_SE, Robust_t, Robust_Pr, and VIF for each estimated coefficient, the signs on the coefficient estimates are as expected and statistically significant. The t-Statistics has tested the hypothesis that the value of an individual coefficient estimate is not significantly different from zero, and the coefficient estimates are sufficiently large to assume that they are not zero in the population from which all sample data were drawn. Additionally, the VIF values are all reasonably small (less than 7.5), indicating no strong evidence of variable redundancy.

However, this OLS model was fitted with spatial data, so it is likely that there will be some structure in the residuals that cannot be accounted for by the model. This may contribute to its indifferent performance. The Jarque-Bera statistic suggests that the residuals are not normally distributed (Table 4.2). As a result, the OLS tool warns us to check whether the residuals are spatially autocorrelated.

In summary, the OLS model regression coefficient estimates, and regression diagnostics were specified as outputs and can be examined in Table 4.1 and Table 4.2, respectively.

The model diagnostics for this analysis can be found in Table 4.2, Table 4.5, and APPENDIX D. There are several measures of goodness-of-fit, including R^2 and adjusted R^2 . The R^2 assesses how much of the dependent variable variation can be attributed to the model variation. Generally, higher values indicate better predictive performance, but this is dependent on the number of variables in the model. The adjusted R^2 is preferred because it adjusts for more variables. This suggests that some variables may have been removed or that the model needs to be adjusted. For comparing different models, AICc can be used as a measure of goodness-of-fit. The AICc value was calculated as -5,461.8258. Generally, it is preferable to have a smaller AICc, but differences of less than 3 or 4 are considered equivalent in terms of explanatory power.

The information provided pertains to two questions: (a) identifying locations with unusually high or low residuals and (b) determining whether the residuals show any spatial autocorrelation. According to Figure 4.1, standardized residuals are displayed from the global model. As shown on the map, large positive residuals (StdResid > 2.5 Std. Dev.) are found in the northeast of the Non Sung district, which contains brackish water in terms of salinity, chloride, and total dissolved solids. According to the current model, this area has underestimated salinity levels because the high positive residuals indicate over-prediction and require further investigation to determine the possible causes.

Table 4.1 Ordinary least squares (OLS) regression coefficient estimate results.

Variable	Coefficient	Std. Error	t-Statistic	Probability	Robust_SE	Robust_t	Robust_Pr	VIF
Intercept	-35.605687	0.540819	-65.836591	0.000000*	0.527825	-67.457326	0.000000*	-
LD	0.355430	0.026170	13.581471	0.000000*	0.023625	15.044411	0.000000*	1.142010
NDVI	-0.107198	0.028275	-3.791310	0.000165*	0.028825	-3.718949	0.000217*	1.300689
RF	0.005731	0.000432	13.277345	0.000000*	0.000424	13.517349	0.000000*	3.237558
GP	-0.012500	0.003519	-3.552338	0.000404*	0.003888	-3.214989	0.001337*	1.107265
CC1	0.915191	0.007244	126.331509	0.000000*	0.009925	92.207515	0.000000*	2.714861
TDS1	15.420554	0.257127	59.972550	0.000000*	0.255014	60.469405	0.000000*	5.729843
SL1	-0.009607	0.003184	-3.017608	0.002585*	0.003355	-2.863518	0.004230*	1.189904
DD1	0.005698	0.002480	2.297442	0.021660*	0.002160	2.637592	0.008396*	1.247319
LF1	-0.006554	0.001616	-4.054770	0.000059*	0.001561	-4.198215	0.000033*	1.192347
STC1	0.000255	0.000064	4.011279	0.000070*	0.000064	4.012223	0.000070*	1.122606
LULC1	-0.002333	0.001162	-2.008506	0.044688*	0.000925	-2.522358	0.011708*	1.048693
GEOL1	0.000057	0.000021	2.663031	0.007790*	0.000022	2.653578	0.008010*	1.561706

Table 4.2 OLS model diagnostics.

Diagnostics	Diagnostics	Definition
Name	Value	
AIC	-5,461.994030	Akaike's Information Criterion (AIC) serves as a comparative measure of model performance, where a lower AIC value indicates a superior model.
AICc	-5,461.825896	Corrected Akaike's Information Criterion: second-order correction for small sample sizes.
R ²	0.975724	R-Squared, also known as the Coefficient of Determination, represents the percentage of variability in the dependent variable that can be accounted for by the model.
Adj. R ²	0.975608	Adjusted R-Squared is a modified version of R-Squared that takes into account the complexity of the model, specifically the number of variables, in relation to the data.
F-Stat	8,373.634298	The Joint F-Statistic Value is employed to evaluate the overall significance of the model.
F-Prob	0.000000*	Joint F-Statistic Probability (p-value): The probability that none of the explanatory variables influence the dependent variable.
Wald	117,383.495120	The Wald Statistic is employed to evaluate the overall significance and robustness of the model.
Wald-Prob	0.000000*	The Wald Statistic Probability (p-value) indicates the probability, based on robust standard errors, that none of the independent variables have an effect on the dependent variable.
K(BP)	156.086567	Koenker's studentized Breusch-Pagan Statistic is utilized to examine the accuracy of standard error estimates when there is heteroskedasticity, indicating non-constant variance.
K(BP)-Prob	0.000000*	The Koenker (BP) Statistic Probability (p-value) measures the likelihood that heteroskedasticity, which indicates non-constant variance, has not affected the reliability of standard errors.
JB	11,143.947084	The Jarque-Bera Statistic is employed to ascertain if the residuals depart from a normal distribution.
JB-Prob	0.000000*	The Jarque-Bera Probability (p-value) indicates the likelihood that the residuals follow a normal distribution.
Sigma ²	0.006622	Sigma-Squared refers to the Ordinary Least Squares (OLS) estimate of the variance of the residual term (error term).

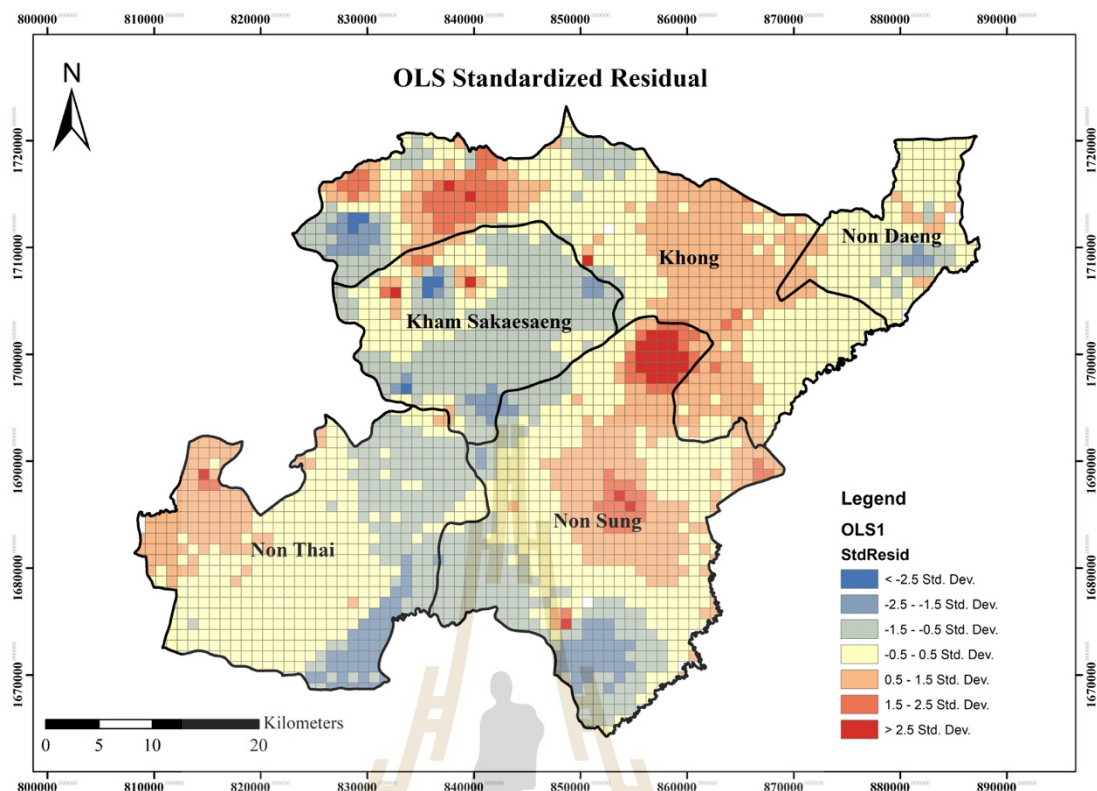


Figure 4.1 Standardized residual of the global model (OLS).

The distribution of the residuals was examined using global spatial autocorrelation, and Moran's I statistic was used for this purpose (Figure 4.2). According to the results, the OLS model had a dispersed pattern, and residuals were highly autocorrelated, indicating possible local variations. To confirm this assumption, however, a geographically weighted regression analysis was conducted. According to Moran's I value for the OLS model was -0.002194 with a p-value of 0.000000, which was statistically significant, implying that the residuals were not randomly distributed.

The study's results have identified the best model according to Equation 4.1, which describes the spatial relationships between groundwater salinity and the influencing factors in the saline soil area of Nakhon Ratchasima province.

$$Y = -35.605686 + 0.355430X_1 - 0.107198X_2 + 0.005731X_3 - 0.012500X_4 + 0.915191X_5 + 15.420554X_6 - 0.009607X_7 + 0.005698X_8 - 0.006554X_9 + 0.000255X_{10} - 0.002333X_{11} + 0.000057X_{12} \quad (4.1)$$

Where Y is Salinity Content (SC1), X_1 is Lineament Density (LD), X_2 is Normalized Difference Vegetation Index (NDVI), X_3 is Annual Rainfall (RF), X_4 is Groundwater Potential (GP), X_5 is Chloride Content (CC1), X_6 is Total solid Content (TDS1), X_7 is Slope (S1), X_8 is Drainage Density (DD1), X_9 is Landform (Lf1), X_{10} is Soil Texture Class (STC1), X_{11} is Land Use/ Land Cover (LULC1), and X_{12} is Geology (Geol1).

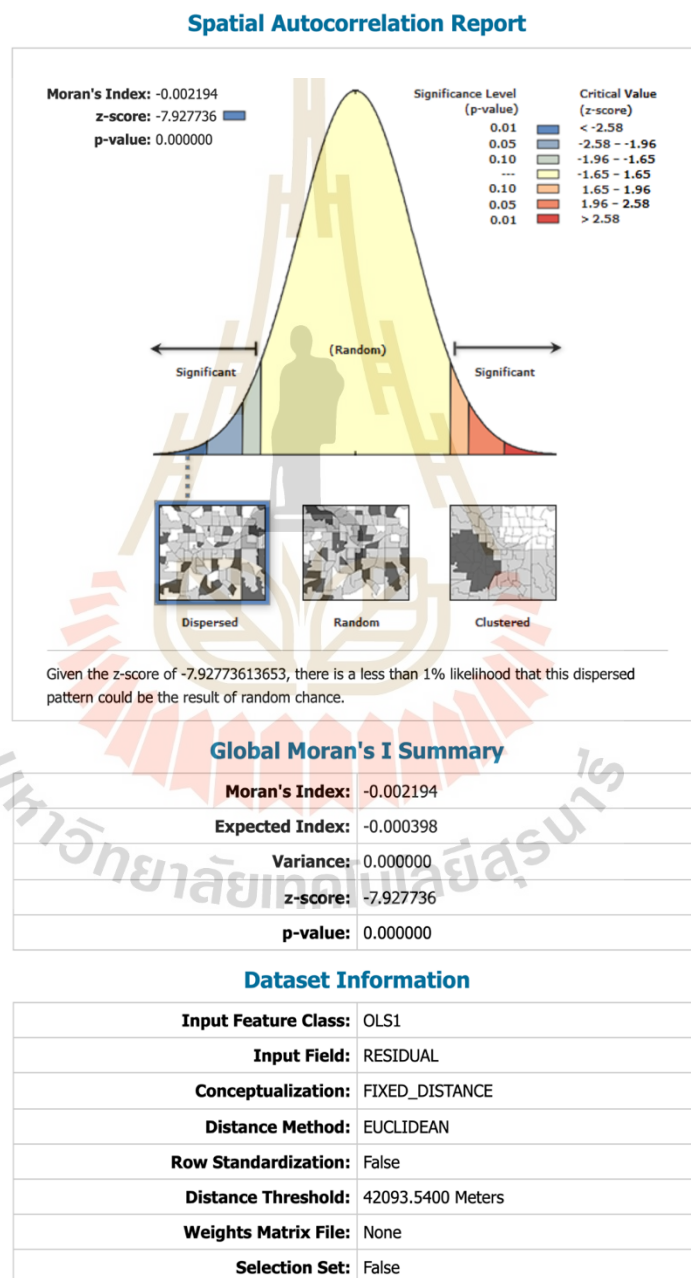


Figure 4.2 Summary of global spatial autocorrelation (Moran's I).

4.2 Results of Local Model: Geographically Weighted Regression (GWR) Analysis

The summary statistics of the local model are shown in Table 4.3 and the values of standardized residuals from the local model (GWR) are shown in Figure 4.3. A positive residual ($\text{StdResid} > 2.5 \text{ Standard Deviation}$) is mostly found in the northwest of the Khong district. Moreover, a positive residual can be found scattered throughout the Non Sung district, Kham Sakae Sang district, Non Thai district, and Non Daeng district in descending order respectively. The presence of positive residuals for a specific area suggests that there are significant discrepancies between the predicted values and the actual observed values of salinity content in that particular location. These positive residuals indicate that the model may be underestimating the salinity content in that area.

Table 4.3 Summary statistics of the local model.

Variable	Mean	Std. Deviation	Minimum	Maximum
Intercept	-23.170691	15.083501	-59.222814	44.050705
LD	0.137691	1.002525	-6.487336	1.948503
NDVI	-0.002584	0.109274	-0.332169	0.348690
RF	0.004080	0.007185	-0.016654	0.035667
GP	-0.013286	0.031139	-0.202453	0.097654
CC1	1.094047	0.393069	0.069137	1.893547
TDS1	8.697749	8.789318	-27.415645	29.465958
SL1	-0.003182	0.008078	-0.023830	0.023884
DD1	0.003086	0.012397	-0.028538	0.039837
Lf1	-0.001870	0.003895	-0.013114	0.009736
STC1	0.000076	0.000154	-0.000272	0.000663
LULC1	-0.000165	0.005207	-0.018065	0.012576
Geol1	0.000038	0.000125	-0.000311	0.000341

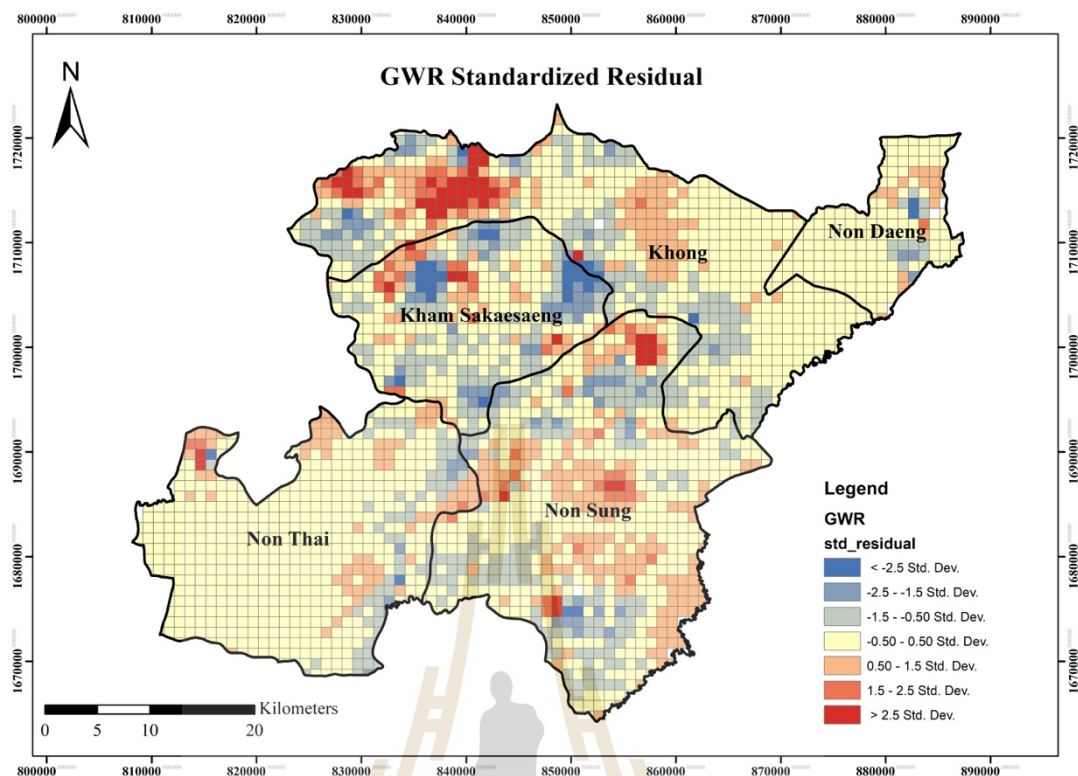


Figure 4.3 Standardized residual of the local model (GWR).

There could be several reasons for the occurrence of such positive residuals. Firstly, it is possible that there are influential variables or factors in that specific area that are not adequately captured or represented by the independent variables included in the model. These unaccounted factors may have a substantial impact on salinity content, leading to the observed discrepancies. Secondly, it is important to consider the limitations of the model itself. The GWR model assumes that the relationship between the independent variables and the dependent variable is spatially varying. However, if the model fails to accurately capture the spatial variations in the relationship, it may result in large positive residuals in certain areas. Furthermore, it is crucial to thoroughly examine the data quality and potential outliers in the dataset. Anomalies or errors in the collected data could contribute to the presence of large positive residuals. It is essential to verify the accuracy and reliability of the data to ensure that it is representative of the true salinity content in the study area.

To address this issue, further investigation and analysis are required. It may be necessary to consider additional variables or refine the existing ones to better capture the spatial variability of salinity content. Additionally, examining potential outliers and refining the data collection process can help improve the accuracy and reliability of the model.

Overall, the identification of positive residuals in the GWR model highlights the need for careful interpretation and potential refinement of the model to enhance its predictive performance and better understand the spatial patterns of salinity content in the specific area.

The model performance map or R^2 of the local model as shown in Figure 4.4 is mostly found in the northeast of the study area, especially the Non Daeng district, Non Daeng district Non Sung district, Khong district, Kham Sakae Sang district, and Non Thai district in descending order respectively. A high Local R^2 indicates that the spatial regression model captures a substantial portion of the spatial variation in the dependent variable (salinity content) within the study area. This suggests that the independent variables included in the model explain a significant proportion of the variability in salinity content at the local level. A high Local R^2 implies a good fit of the model to the data, indicating that it can effectively predict salinity content in specific regions of interest.

On the other hand, a low Local R^2 suggests that the spatial regression model explains only a limited amount of the spatial variation in the dependent variable. This could be due to various reasons. It is possible that the independent variables included in the model are not strong predictors of salinity content at the local level, or there may be influential factors that are not captured by the model. Additionally, if there is a high degree of spatial heterogeneity or complex spatial patterns in the relationship between the independent variables and salinity content, it can result in a lower Local R^2 as shown in Figure 4.4 on the northwest of the Non Thai district showing medium and low Local R^2 .

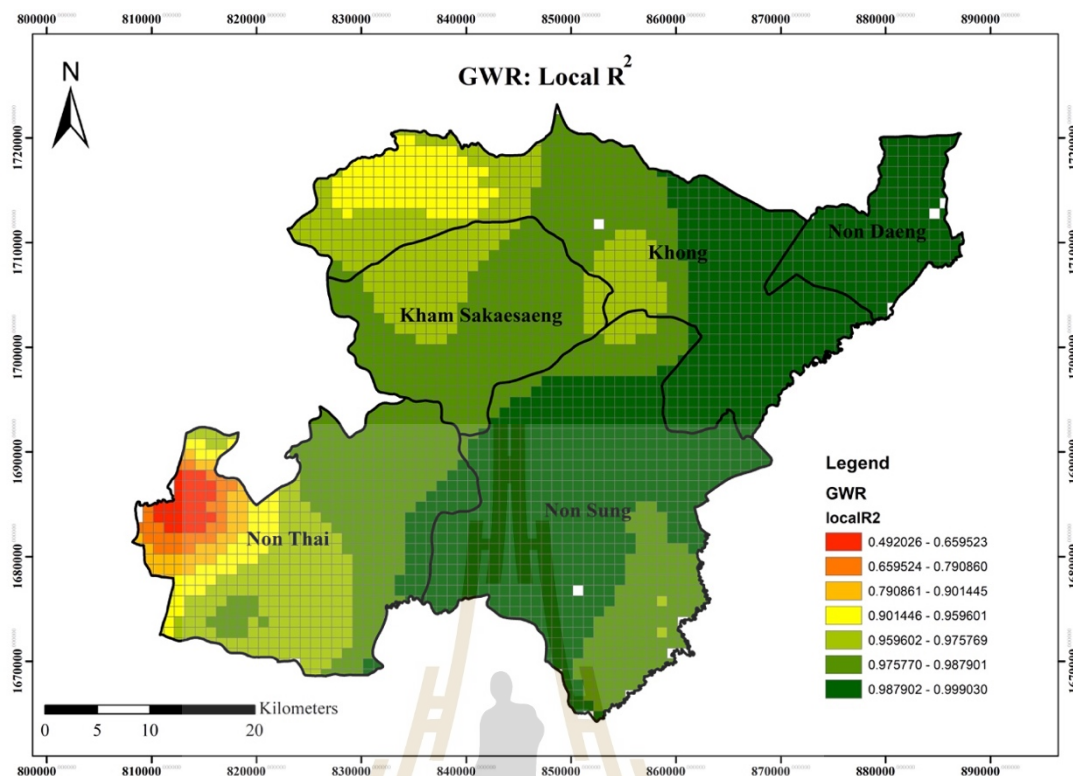


Figure 4.4 Local R^2 of the local model (GWR).

It is important to interpret the Local R^2 in conjunction with other diagnostic measures and the specific context of the study area. While a high Local R^2 indicates a strong relationship between the independent variables and salinity content in certain regions, it does not necessarily imply a causal relationship or generalizability to other areas. Local R^2 values should be interpreted and compared within the spatial context of the study to gain a comprehensive understanding of the model's performance at the local level.

Local spatial autocorrelation was utilized to assess the random distribution of the residuals, and Moran's I statistic was computed, as shown in Figure 4.5 and APPENDIX D. The findings demonstrate that the Moran's I value of the GWR model is 0.000299, and the p-value for testing the hypothesis that this value differs significantly from zero is 0.002104 (with a z-score of 3.075130). The pattern observed was "Clustered" suggesting the presence of significant local variations in the residuals. The

residuals from the GWR local regression model exhibited statistically significant autocorrelation.

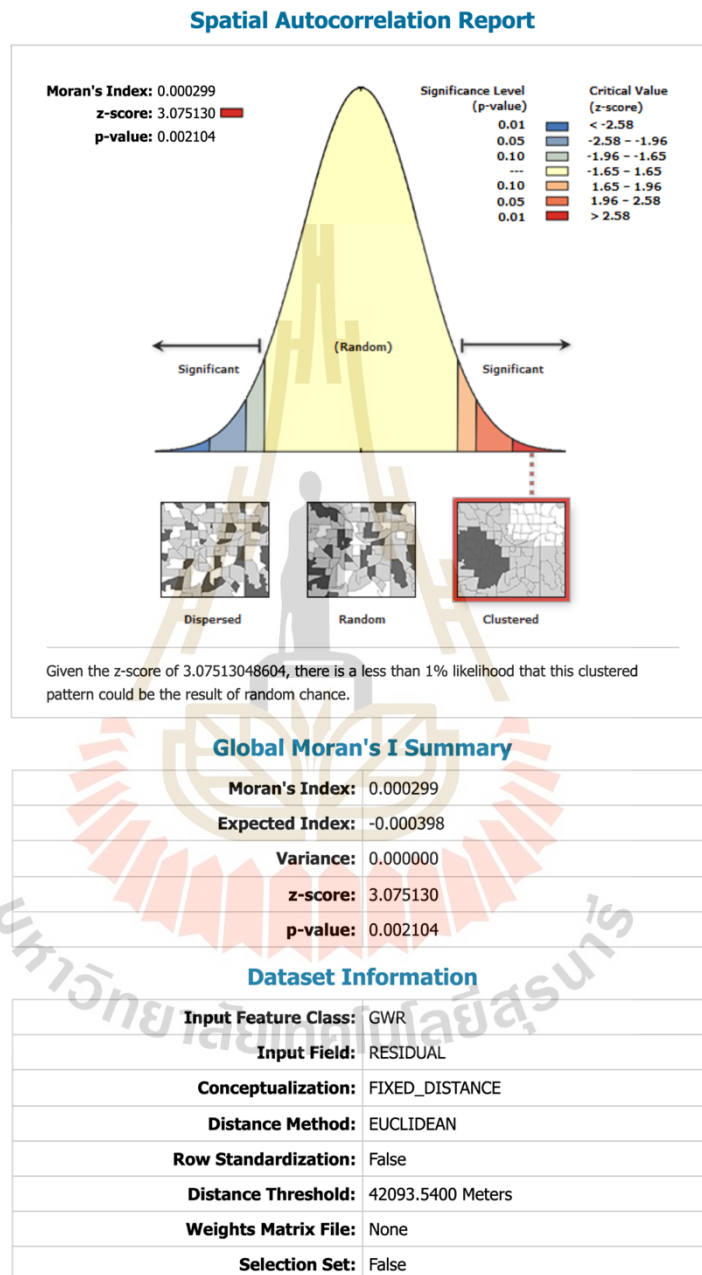


Figure 4.5 Summary of local spatial autocorrelation (Moran's I).

4.3 Comparison of OLS and GWR Models

According to Phoemphon and Terakulsatit (2023), the research aimed to assess groundwater potential zones in saline soil areas using remote sensing, GIS, and AHP techniques. Multiple influencing factors, including slope, rainfall, soil texture, and land use, were considered in the analysis. The weights of these factors were determined using the AHP technique, resulting in the classification of zones into five categories. Validation of the results using salinity data from groundwater wells showed an R^2 value of 0.7131. The findings revealed that the majority of the study area exhibited moderate to poor groundwater potential.

After comparing the fit of the global model (OLS) and local model (GWR) as presented in Table 4.5, it was observed that the locally adjusted R^2 (0.99) performed better than the global adjusted R^2 (0.97), indicating an improvement in model performance. The preferred measure of model fit, AICc, revealed that the local model (-9,801.87) had a significantly better fit than the global model (-5,461.82), with a difference of 4,340.05. These results were further supported by the statistical significance of Table 4.4 ($F(2249, 251) = 0.0178$, $p < 0.001$). Consequently, an examination of the impact of local coefficient estimates was necessary.

A model suitability test was performed using the F-Statistic, and the results are presented in the table provided.

Table 4.4 Geographically weighted regression (GWR) ANOVA results.

Source	SS	DF	MS	F
Global Residuals	16.555	2,500.000		
GWR Improvement	14.032	251.076	0.056	
GWR Residuals	2.523	2,248.924	0.001	49.813

OLS (Ordinary Least Squares) is a widely used regression technique that offers several advantages. One of its primary strengths is its simplicity, making it easy to understand and implement. As a baseline model, it provides a straightforward

approach to analyzing statistical relationships. OLS is computationally efficient and suitable for large datasets with a limited number of predictors, enabling efficient estimation of model parameters. Additionally, OLS allows for the interpretation of coefficient estimates, providing insights into the individual effects of independent variables on the dependent variable. This interpretability facilitates the comparison of predictor importance through standardized coefficients. Furthermore, OLS provides statistical inference measures such as p-values, confidence intervals, and hypothesis tests, supporting reliable statistical conclusions.

However, OLS also has certain limitations that need to be acknowledged. One significant drawback is its assumption of spatial homogeneity. OLS assumes that the relationship between independent variables and the dependent variable is constant across space, disregarding potential spatial variations. Consequently, when spatial heterogeneity exists, OLS may yield biased estimates. Another challenge is its sensitivity to outliers and influential observations, which can disproportionately influence the regression results, potentially distorting the estimated coefficients and overall model performance.

In contrast, GWR (Geographically Weighted Regression) addresses some of the limitations of OLS and offers distinctive advantages. GWR explicitly accounts for spatial variation in the relationship between independent variables and the dependent variable. It allows for different relationships to exist in different areas, capturing local nuances and spatial heterogeneity. This localized analysis is a major strength of GWR, providing insights into how the effects of predictors vary across the study area. Moreover, GWR can outperform OLS in terms of model performance when there are significant spatial variations in the relationships between variables. By capturing local patterns, GWR enhances predictive accuracy in areas with pronounced spatial non-stationarity.

However, GWR also has its own set of limitations. One notable drawback is its increased complexity compared to OLS. GWR requires additional computational resources and statistical techniques as it estimates regression parameters for each local

area. Consequently, GWR may be computationally demanding, particularly for large datasets. Another challenge lies in the interpretation of GWR results. Due to the localized nature of the model, interpreting coefficients and their significance becomes more challenging. The coefficients may vary across different locations, making it difficult to generalize the findings to the entire study area. Additionally, GWR has the potential for overfitting, particularly in areas with sparse data or extreme local variations, which may lead to less reliable predictions.

Considering these advantages and disadvantages, it is recommended to use GWR model. Since the objective is to investigate the relationship between groundwater salinity and independent variables in the soil salinity area of Nakhon Ratchasima, it is important to account for potential spatial variations. GWR's ability to capture local patterns and spatial heterogeneity makes it a suitable choice for this study. By using GWR, Insights into how the effects of predictors vary across different locations can enhance the understanding of salinity content patterns in the specific study area. However, it is crucial to carefully interpret the GWR results, considering the localized nature of the model and the challenges associated with coefficient interpretation.

Table 4.5 Comparison of the global model (OLS) and local model (GWR).

Diagnostic	Global Model (OLS)	Local Model (GWR)
Residual sum of squares	16.555494	2.523233
ML based sigma estimate	0.081166	0.031687
Unbiased sigma estimate	0.081377	0.033496
-2 log-likelihood	-5,489.994000	-10.217.392000
AICc	-5,461.825900	-9,801.877000
BIC/MDL	-5,380.384700	-8,715.614900
CV	0.006660	0.001267
R ²	0.975724	0.996300
Adjusted R ²	0.975598	0.995865

4.4 Discussion

The primary objective of this study was to examine the relationship between various influent factors and salinity content (SC) in the saline soil area of Nakhon Ratchasima province. The initial hypothesis proposed that higher levels of LD, RF, CC1, TDS1, DD1, STC1, and Geol1 would lead to an increase in SC. In contrast, lower levels of NDVI, GP, Sl1, Lf1, and LULC1 would decrease SC. Additionally, considering the expected spatial non-uniformity of the relationship, the implementation of a spatial regression model, specifically the Geographically Weighted Regression (GWR), was anticipated to provide superior predictions of SC compared to the non-spatial Ordinary Least Squares (OLS) model.

The obtained results from the saline soil area of the study area provide substantial support for the proposed hypothesis, both in terms of the impact of each individual variable and the overall performance of the spatial GWR model. There is a significant negative correlation between SC and NDVI, GP, Sl1, Lf1, and LULC1. Conversely, a positive association is observed between SC and LD, RF, CC1, TDS1, DD1, STC1, and Geol1. Based on these results, higher levels of these influencing factors are associated with an increase in salinity content, while lower levels are associated with a decrease.

Furthermore, the comparison between the local (GWR) model and the global (OLS) model reveals the superiority of the former in capturing the spatial variations within the data. The utilization of a global regression model, such as OLS, is valuable for assessing statistically significant relationships. However, the incorporation of a local regression model, like GWR, allows for the examination of regional variations and the identification of spatial patterns that may remain concealed by the global model. This finding aligns with the suggestions by Fotheringham et al. (2002) and highlights the enhanced model performance of the GWR model in the analysis of spatial data.

In the saline soil area of Nakhon Ratchasima province, where the model performance reaches its highest level, the results indicate that salinity content tends to decrease as the levels of NDVI, GP, Sl1, Lf1, and LULC1 increase. This suggests that

these influencing factors play a vital role in mitigating salinity content in the study area. Understanding and incorporating these factors into future land and water resource management strategies can contribute to effective measures to address salinity-related challenges.

The outcomes of this study have important implications for salinity management and highlight the significance of considering spatial variations when analyzing and modeling salinity content. By a combination of global and local regression models, researchers and policymakers can expand their understanding of the relationships between influencing factors and salinity content. This will result in more targeted and efficient interventions.

In summary, the findings of this study provide substantial evidence supporting the hypothesis regarding the impact of various influencing factors on salinity content in the saline soil area of Nakhon Ratchasima province. The results demonstrate a negative association of NDVI, GP, Sl1, Lf1, and LULC1 with salinity content. In contrast LD, RF, CC1, TDS1, DD1, STC1, and Geol1 exhibit a positive association. Further, the superior performance of the local (GWR) model compared to the global (OLS) model emphasizes the importance of incorporating spatial regression techniques to analyze salinity data. In addition to contributing to the current understanding of salinity management, these findings provide valuable insights that can be used to guide future research and decision-making in similar regions.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

This study used the global model (OLS) and local model (GWR) to investigate the spatial relationship between groundwater salinity and the independent variable in the soil salinity area of Nakhon Ratchasima province. It was found that the GWR model was more accurate than the OLS model based on an adjusted R^2 of 0.995865 and an AICc of -9,801.877000.

Based on the GWR model, the results showed that the salinity content is negatively associated with NDVI, GP, SL1, Lf1, and LULC1, but positively associated with LD, RF, CC1, TDS1, DD1, STC1, and Geol1, and that the spatial regression model of GWR generated the highest model performance when predicting salinity content. The standardized residuals reveal a clear spatial pattern, with the majority of high values concentrated in the northwest of the Khong district. These areas exhibit a significant positive deviation from the expected values based on the regression model. In contrast, positive residuals are scattered throughout the Non Sung district, Kham Sakae Sang district, Non Thai district, and Non Daeng district, with decreasing occurrence from the highest to the lowest. These scattered positive residuals indicate localized areas where the observed salinity content differs from the predicted values. The identification of these residual patterns can provide valuable insights into the spatial variability of salinity content and can guide further investigation into the underlying factors influencing salinity in these specific areas. In addition, the highest performance of the model is mostly found in the northeast of the study area, especially the Non Daeng district, Non Sung district, Khong district, Kham Sakae Sang district, and Non Thai district in descending order respectively. Furthermore, it highlights the need for similar

studies in multiple areas, to determine and conclude major variables impacting salinity content.

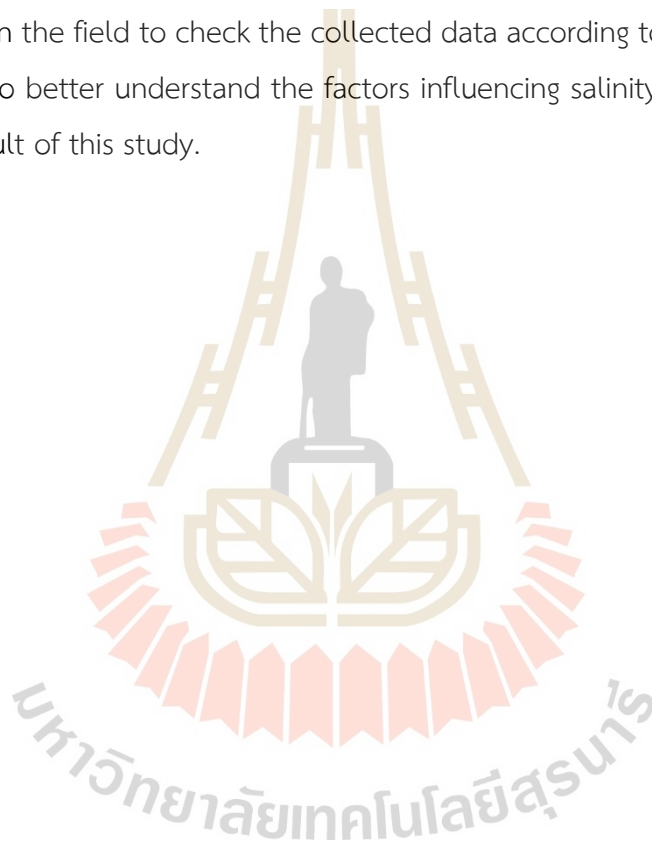
This study represents one of the studies to use chemical data along with spatial regression analysis to explore how NDVI, GP, S11, Lf1, LULC1, LD, RF, CC1, TDS1, DD1, STC1, and Geol1 affect salinity content (SC), a beneficial method in multiple ways. The study demonstrated the benefits of detailed chemical data, both with respect to model performance and by allowing spatial analysis of the data. Moreover, this study emphasizes the importance of using the global non-spatial analysis (OLS) model to determine statistically significant relationships and examining regional variations in the relationship between salinity content and independent variables using the local spatial analysis (GWR) model; revealing spatial patterns not found in global models. In addition, both OLS and GWR regression techniques have their respective advantages and disadvantages. OLS offers simplicity, computational efficiency, interpretability, and statistical inference measures. On the other hand, GWR accounts for spatial variation, captures local nuances, and can outperform OLS in the presence of significant spatial variations. However, GWR is more complex, computationally demanding, and challenging in coefficient interpretation, with the potential for overfitting. Consequently, this study contributes to the knowledge on salinity content and the use of the local spatial analyses (GWR) model in groundwater assessment studies. This knowledge can be applied to urban planning for the management and use of groundwater resources.

5.2 Recommendations for further research

In future studies, additional variables such as distance to the water body and remote sensing data can be considered to enhance the predictive performance of salinity content in Nakhon Ratchasima province. By including these factors, obtaining a more comprehensive understanding of the spatial relationship between groundwater salinity and relevant environmental factors would be possible.

Moreover, it is recommended that alternative statistical models be used in future research to test the prediction model's accuracy. An example of such a model would be a logistic regression model that would establish a functional relationship between an independent and dependent variable. Alternatively, other machine learning algorithms may improve prediction accuracy.

These suggestions can be incorporated into future research endeavors to provide a more robust and comprehensive analysis. In addition, it should be investigated in the field to check the collected data according to these models. It will be possible to better understand the factors influencing salinity content in the study area as a result of this study.





REFERENCES

มหาวิทยาลัยเทคโนโลยีสุรนารี

REFERENCES

- Ahmad, I., Dar, M. A., Teka, A. H., Teshome, M., Andualem, T. G., Teshome, A., & Shafi, T. (2020). GIS and fuzzy logic techniques-based demarcation of groundwater potential zones: A case study from Jemma River basin, Ethiopia. *Journal of African Earth Sciences*, *169*, 103860. doi:<https://doi.org/10.1016/j.jafrearsci.2020.103860>
- Aldharab, H., Ali, S., Ikbal, J., & Ghareb, S. (2019). Analysis of Basin Geometry in Ataq Region, Part of Shabwah Yemen: Using Remote Sensing and Geographic Information System Techniques. *Bulletin of Pure and Applied Sciences - Section F Geological Sciences*, *38*, 1-15. doi:10.5958/2320-3234.2019.00001.5
- Atun, R., Kalkan, K., & Gürsoy, Ö. (2020). Determining The Forest Fire Risk with Sentinel 2 Images. *1*, 22-26.
- Bhanja, S. N., Malakar, P., Mukherjee, A., Rodell, M., Mitra, P., & Sarkar, S. (2019). Using Satellite-Based Vegetation Cover as Indicator of Groundwater Storage in Natural Vegetation Areas. *Geophysical Research Letters*, *46*(14), 8082-8092. doi:<https://doi.org/10.1029/2019GL083015>
- Booth, J., & Sattayarak, N. (2011). Subsurface Carboniferous – Cretaceous geology of NE Thailand. In M. F. Ridd, A. J. Barber, & M. J. Crow (Eds.), *The Geology of Thailand* (pp. 0): Geological Society of London.
- Brown, S., Versace, V. L., Laurenson, L., Ierodiaconou, D., Fawcett, J., & Salzman, S. (2012). Assessment of Spatiotemporal Varying Relationships Between Rainfall, Land Cover and Surface Water Area Using Geographically Weighted Regression. *Environmental Modeling & Assessment*, *17*(3), 241-254. doi:10.1007/s10666-011-9289-8

- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, 28(4), 281-298. doi:<https://doi.org/10.1111/j.1538-4632.1996.tb00936.x>
- Brunsdon, C., Fotheringham, S., & Charlton, M. (1998). Geographically Weighted Regression-Modelling Spatial Non-Stationarity. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 47(3), 431-443. Retrieved from <http://www.jstor.org/stable/2988625>
- Chaimanee, N. (Cartographer). (2003). Geologic Map of The Khorat Greater City Area
- Charlton, M., & Fotheringham, A. S. (2009). *GEOGRAPHICALLY WEIGHTED REGRESSION WHITE PAPER*.
- Charlton, M., & Fotheringham, S. (2009). *Geographically Weighted Regression White Paper*.
- Chemicals of Concern (COC) - Groundwater Information Sheets (Salinity). (2017). Retrieved from https://www.waterboards.ca.gov/water_issues/programs/gama/docs/coc_salinity.pdf. from State Water Resources Control Board Division of Water Quality GAMA Program https://www.waterboards.ca.gov/water_issues/programs/gama/docs/coc_salinity.pdf
- Cui, G., Lu, Y., Zheng, C., Liu, Z., & Sai, J. (2019). Relationship between Soil Salinization and Groundwater Hydration in Yaoba Oasis, Northwest China. *Water*, 11(1), 175. Retrieved from <https://www.mdpi.com/2073-4441/11/1/175>
- Department of Mineral Resources, T. (n.d.). ROCK_UNIT_250K Web Map Service. Retrieved from https://gisportal.dmr.go.th/arcgis/rest/services/GEOL/ROCK_UNIT_250K/MapServer

- DMR. (2015). *Saline Soil Area Development in the Northeast: Geological approaches*. Retrieved from http://www.dmr.go.th/download/article/article_20151109134917.pdf
- El Tabakh, M., Utha-Aroon, C., & Schreiber, B. C. (1999). Sedimentology of the Cretaceous Maha Sarakham evaporites in the Khorat Plateau of northeastern Thailand. *Sedimentary Geology*, 123(1), 31-62. doi:[https://doi.org/10.1016/S0037-0738\(98\)00083-9](https://doi.org/10.1016/S0037-0738(98)00083-9)
- Elewa, H., & Qaddah, A. (2011). Groundwater potentiality mapping in the Sinai Peninsula, Egypt, using remote sensing and GIS-watershed-based modeling. *Hydrogeology Journal*, 19, 613-628. doi:10.1007/s10040-011-0703-8
- ESRI. (n.d.). Regression Analysis Basics - ArcGIS Desktop. Retrieved from <https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-statistics-toolbox/regression-analysis-basics.htm>
- Fotheringham, A. S., Brunson, C., & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*: John Wiley & Sons Ltd.
- Gopal, K. (2019). Groundwater Salinity. *Current World Environment*(14(2)).
- Guru, B., Seshan, K., & Bera, S. (2017). Frequency ratio model for groundwater potential mapping and its sustainable management in cold desert, India. *Journal of King Saud University - Science*, 29(3), 333-347. doi:<https://doi.org/10.1016/j.jksus.2016.08.003>
- Hansen, B., Wemmer, K., Eckhardt, M., Putthapiban, P., & Assavapatchara, S. (2016). Isotope Dating of the Potash and Rock Salt Deposit at Bamnet Narong, NE-Thailand. *Open Journal of Geology*, 06, 875-894. doi:10.4236/ojg.2016.68067
- Hengl, T. (2018). Soil texture classes (USDA system) for 6 soil depths (0, 10, 30, 60, 100 and 200 cm) at 250 m. doi:10.5281/ZENODO.2525817

- Hite, R. J., & Japakasetr, T. (1979). Potash deposits of the Khorat Plateau, Thailand and Laos. *Economic Geology*, 74(2), 448-458. doi:10.2113/gsecongeo.74.2.448
- Hurvich, C. M., Simonoff, J. S., & Tsai, C.-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2), 271-293. doi:<https://doi.org/10.1111/1467-9868.00125>
- Hutti, B., & Nijaḡunappa, R. (2011). Identification of Groundwater Potential Zone using Geoinformatics in Ghataprabha basin, North Karnataka, India. *INTERNATIONAL JOURNAL OF GEOMATICS AND GEOSCIENCES*, 2(1), 91-109. Retrieved from <http://www.ipublishing.co.in/jggsvol1no12010/voltwo/EIJGGS3009.pdf>
- Impact Observatory, M., and Esri. (2022). Sentinel-2 10m Land Use/Land Cover Time Series. Retrieved from <https://www.arcgis.com/home/item.html?id=d3da5dd386d140cf93fc9ecbf8da5e31>
- Karant, K. R., & Seshu babu, K. (1987). *Identification of major lineaments on satellite imagery and on aerial photographs for delineation for possible potential groundwater zones in Penukonda and Dharmavaram taluks of Anantapur district*. Paper presented at the Joint Indo-US Workshop on Remote Sensing of Water Resources (NRSA, Hyderabad),.
- Kumar, A. (1999). Sustainable Utilisation of Water Resource in Watershed Perspective - A Case Study in Alaunja Watershed, Hazaribagh, Bihar. *Journal of the Indian Society of Remote Sensing*, 27(1), 13-22. doi:10.1007/BF02990771
- Kumar, T., Gautam, A. K., & Kumar, T. (2014). Appraising the accuracy of GIS-based Multi-criteria decision making technique for delineation of Groundwater potential zones. *Water Resources Management*, 28(13), 4449-4466. doi:10.1007/s11269-014-0663-6

- Lohman, A. H. M., & Mentink, G. M. (1972). Some cortical connections of the tegu lizard (Tupinambis teguixin). *Brain Research*, 45(2), 325-344.
doi:[https://doi.org/10.1016/0006-8993\(72\)90466-0](https://doi.org/10.1016/0006-8993(72)90466-0)
- Lv, M., Xu, Z., Yang, Z.-L., Lu, H., & Lv, M. (2021). A Comprehensive Review of Specific Yield in Land Surface and Groundwater Studies. *Journal of Advances in Modeling Earth Systems*, 13(2), e2020MS002270.
doi:<https://doi.org/10.1029/2020MS002270>
- Magesh, N. S., Chandrasekar, N., & Soundranayagam, J. P. (2012). Delineation of groundwater potential zones in Theni district, Tamil Nadu, using remote sensing, GIS and MIF techniques. *Geoscience Frontiers*, 3(2), 189-196.
doi:<https://doi.org/10.1016/j.gsf.2011.10.007>
- Malik, S., & Rajeshwari, D. (2011). Delineation of groundwater potential zones in Mewat District. *INTERNATIONAL JOURNAL OF GEOMATICS AND GEOSCIENCES*, 2. Retrieved from https://www.researchgate.net/profile/Sitender_Malik4/publication/288837986_Delineation_of_groundwater_potential_zones_in_Mewat_District/links/5686884f08ae051f9af3faf9/Delineation-of-groundwater-potential-zones-in-Mewat-District.pdf
- Manjare, B. S. (2014). Identification of groundwater prospecting zones using Remote Sensing and GIS techniques in upper Vena river watersheds Nagpur district, Maharashtra, India. *15th Esri India User Conference 2014*, 14. Retrieved from <https://www.esri.in/~media/esri-india/files/pdfs/events/uc2014/proceedings/papers/UCP0072.pdf?la=en>
- Meesook, A. (2011). Cretaceous. In M. F. Ridd, A. J. Barber, & M. J. Crow (Eds.), *The Geology of Thailand* (pp. 0): Geological Society of London.
- Ndatuwong, L., & Yadav, G. (2014). Morphometric Analysis to Infer the Hydrogeological Behaviour in Part of Sonebhadra District, Uttar Pradesh, India, Using Remote Sensing and GIS Technique; Lazarus G. Ndatuwong and G. S. Yadav, *Int. Jl. of*

Remote Sensing and GIS, vol. 3 (1), pp. 1-7, 2014. *Int. Jl. of Remote Sensing and GIS*, 3, 1-7.

Peck, A. J., & Hatton, T. (2003). Salinity and the discharge of salts from catchments in Australia. *Journal of Hydrology*, 272(1), 191-202. doi:[https://doi.org/10.1016/S0022-1694\(02\)00264-0](https://doi.org/10.1016/S0022-1694(02)00264-0)

Phoemphon, W., & Terakulsatit, B. (2023). Assessment of groundwater potential zones and mapping using GIS/RS techniques and analytic hierarchy process: A case study on saline soil area, Nakhon Ratchasima, Thailand. *AIMS Geosciences*, 9(1), 49-67. doi:10.3934/geosci.2023004

Pradhan, B., & Pirasteh, S. (2011). Hydro-Chemical Analysis of the Ground Water of the Basaltic Catchments: Upper Bhatsai Region, Maharashtra. *Open Hydrology Journal*, 5, 51-57. doi:10.2174/1874378101105010051

Pratt, B., & Chang, H. (2012). Effects of land cover, topography, and built structure on seasonal water quality at multiple spatial scales. *Journal of Hazardous Materials*, 209-210, 48-58. doi:<https://doi.org/10.1016/j.jhazmat.2011.12.068>

Racey, A., Goodall, J. G. S., Buffetaut, E., Cuny, G., Loeuff, J. L., & Suteethorn, V. (2009). Palynology and stratigraphy of the Mesozoic Khorat Group red bed sequences from Thailand. In *Late Palaeozoic and Mesozoic Continental Ecosystems in SE Asia* (Vol. 315, pp. 0): Geological Society of London.

Radhakrishnan, D., & Ramamoorthy, P. (2014). Delineation of groundwater recharge potential zones in Mailam Block, Villupuram district, Using GIS. *International Journal of Water Research*, 2(2), 71-75.

Sarath Prasanth, S. V., Magesh, N. S., Jitheshlal, K. V., Chandrasekar, N., & Gangadhar, K. (2012). Evaluation of groundwater quality and its suitability for drinking and agricultural use in the coastal stretch of Alappuzha District, Kerala, India. *Applied Water Science*, 2(3), 165-175. doi:10.1007/s13201-012-0042-5

- Sataragsa, P. (1987). *Engineering Geology of Khorat City, Northeast Thailand*. (M.Sc.). Asian Institute of Technology, Bangkok,
- Sattayarak, N. e. a. (1987). *Influences of Rock Salt on Ground Water in The Northeast*. (In Thai). Paper presented at the The Seminar on Geology and Esarn Kieo Development, mperial Hotel, Bangkok.
- Seeboonruang, U. (2013). Relationship between groundwater properties and soil salinity at the Lower Nam Kam River Basin in Thailand. *Environmental Earth Sciences*, 69(6), 1803-1812. doi:10.1007/s12665-012-2012-5
- Selvam, S., Dar, F. A., Magesh, N. S., Singaraja, C., Venkatramanan, S., & Chung, S. Y. (2016). Application of remote sensing and GIS for delineating groundwater recharge potential zones of Kovilpatti Municipality, Tamil Nadu using IF technique. *Earth Science Informatics*, 9(2), 137-150. doi:10.1007/s12145-015-0242-2
- Shaban, A., Khawlie, M., & Abdallah, C. (2006). Use of remote sensing and GIS to determine recharge potential zones: the case of Occidental Lebanon. *Hydrogeology Journal*, 14(4), 433-443. doi:10.1007/s10040-005-0437-6
- Sikakwe, G., Ntekim, E., Obi, D., & George, A. (2015). Geohydrological study of weathered basement aquifers in Oban Massif and environs Southeastern Nigeria: using Remote Sensing and Geographic Information System Techniques.
- Suwanich, P. (2007, November 21-22, 2007). *Potash-evaporite deposits in Thailand*. Paper presented at the Geothai'2007, Proceedings of the International Conference on Geology of Thailand; Towards Sustainable Development and Sufficiency Economy, Bangkok, Thailand.
- Tabakh, M. E., Schreiber, B. C., Utha-Aroon, C., Coshell, L., & Warren, J. K. (1998). Diagenetic origin of Basal Anhydrite in the Cretaceous Maha Sarakham salt: Khorat Plateau, NE Thailand. *Sedimentology*, 45(3), 579-594. doi:<https://doi.org/10.1046/j.1365-3091.1998.00162.x>

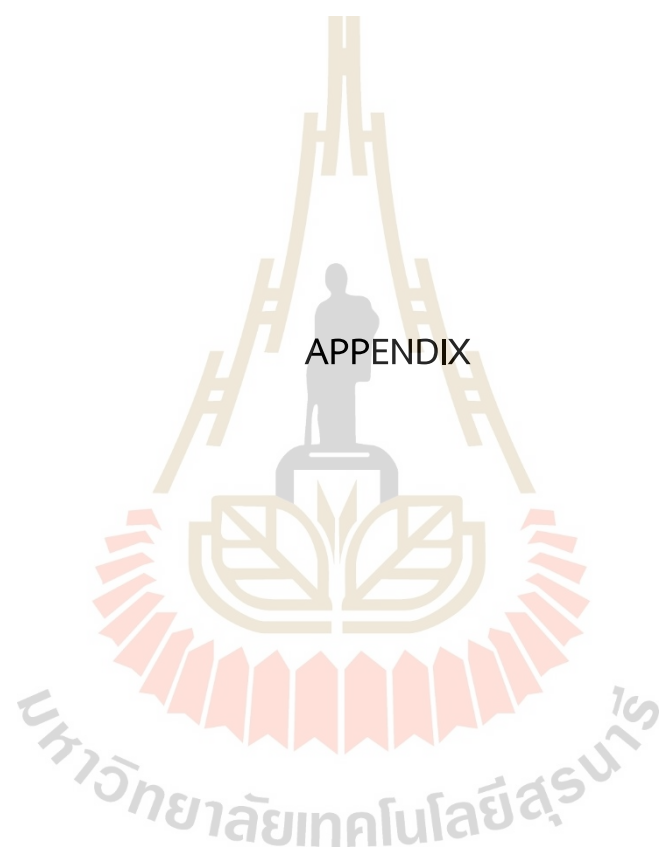
- Tatong, T., & Margane, A. (2004). *Hydrogeology of the Khorat Greater City Area and Investigation of the Causes and Effects of Groundwater and Soil Salinization*. Bangkok :: Department of Mineral Resources.
- Theobald, D. M., Harrison-Atlas, D., Monahan, W. B., & Albano, C. M. (2015). Ecologically-Relevant Maps of Landforms and Physiographic Diversity for Climate Adaptation Planning. *PLOS ONE*, 10(12), e0143619. doi:10.1371/journal.pone.0143619
- Thongwat, W. (2018). *Relationship between soil salinity and chloride content in groundwater in saline soil areas of Nakhon Ratchasima province*. (Master of Engineering). Suranaree University of Technology, Suranaree University of Technology Intellectual Repository. Retrieved from <http://sutir.sut.ac.th:8080/jspui/handle/123456789/8089> (161)
- Tu, J. (2011). Spatially varying relationships between land use and water quality across an urbanization gradient explored by geographically weighted regression. *Applied Geography*, 31(1), 376-392. doi:<https://doi.org/10.1016/j.apgeog.2010.08.001>
- Tu, J. (2013). Spatial Variations in the Relationships between Land Use and Water Quality across an Urbanization Gradient in the Watersheds of Northern Georgia, USA. *Environmental Management*, 51(1), 1-17. doi:10.1007/s00267-011-9738-9
- Tu, J., & Xia, Z.-G. (2008). Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluation. *Science of The Total Environment*, 407(1), 358-378. doi:<https://doi.org/10.1016/j.scitotenv.2008.09.031>
- Understanding Salinity. (2020). Retrieved from https://score.dnr.sc.gov/media/Understanding_Salinity.pdf
- Utha-Aroon, C. (1993). Continental origin of the Maha Sarakham evaporites, northeastern Thailand. *Journal of Southeast Asian Earth Sciences*, 8(1), 193-203. doi:[https://doi.org/10.1016/0743-9547\(93\)90021-G](https://doi.org/10.1016/0743-9547(93)90021-G)

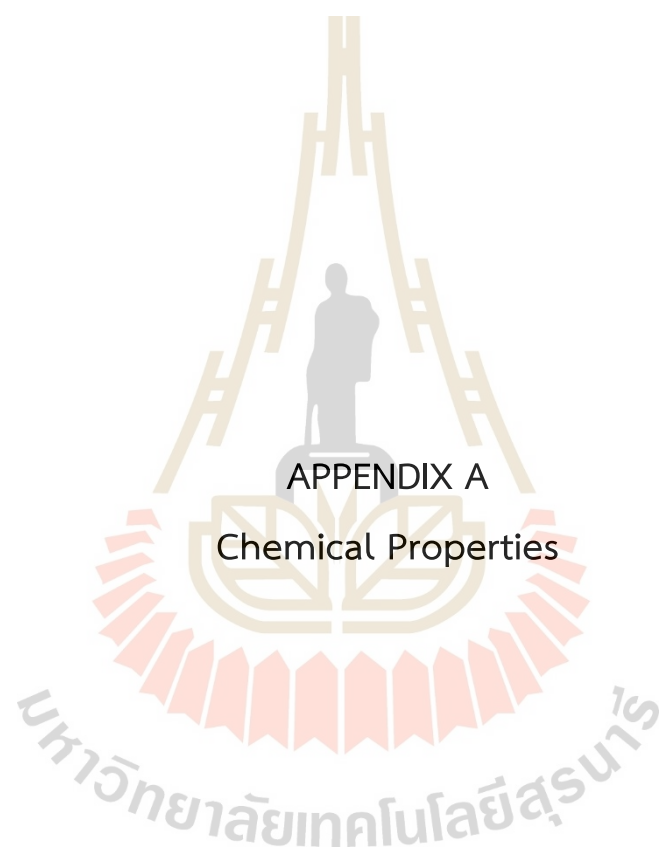
V.R, H., Aravindan, S., & Gopinath, G. (1998). Remote sensing and its application for ground water favorable area identification. *Journal of Geological Association and Research Center*, VI, 18-22. Retrieved from https://www.researchgate.net/profile/Sengalvarayan_Aravindan/publication/259478618_Remote_sensing_and_its_application_for_ground_water_favorable_area_identification/links/5ea081a992851c87d1ab2fef/Remote-sensing-and-its-application-for-ground-water-favorable-area-identification.pdf

Wannakomol, A. (2012). *Mapping of groundwater using electrical survey in the vicinity of Nakhon Ratchasima province*. Retrieved from <https://core.ac.uk/download/pdf/70938569.pdf>

Wongsomsak, S. (1986). Salinization in northeast Thailand. *Japanese Journal of Southeast Asian Studies*, 24(2), 133-153.







APPENDIX A

Chemical Properties

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table A1 Chemical properties of groundwater samples in May 2019.

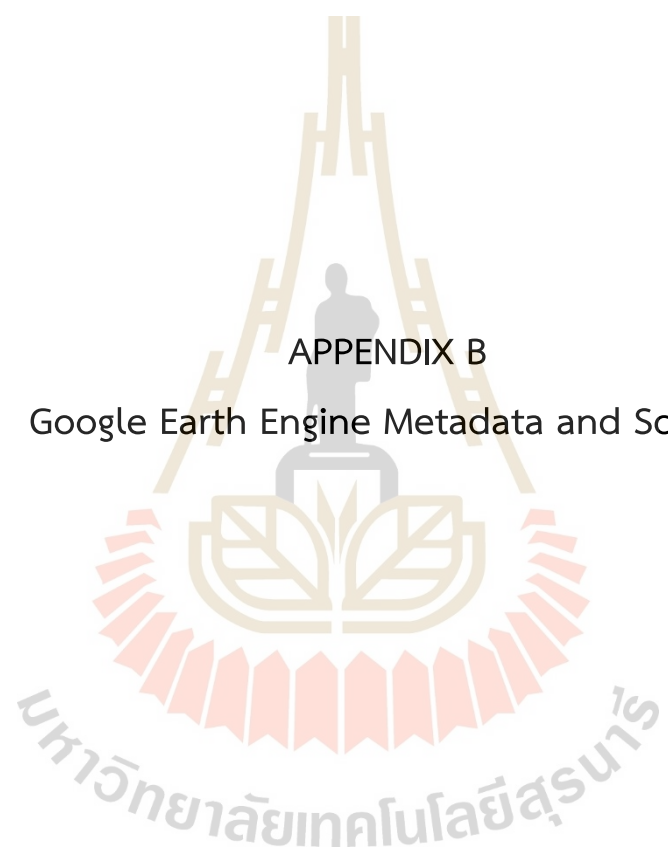
No	Station	Location	district	Sub-district	UTM E	UTM N	TDS	SC	CC
1	5705D031	Non Daeng Municipal School	Non Daeng	Non Daeng	238141	1706668	7,614.00	22.7	51,046.54
2	5405B007	Phu Wittaya School	Non Daeng	Non Daeng	238343	1705364	958.00	1.00	1,648.38
3	NR270	Ban Nong Ya Khao	Non Daeng	Samphaniang	238705	1711383	1,010.00	1.07	2,202.90
4	MG799	Ban Muang Kao School	Non Thai	Banlang	170377	1689973	1,074.00	1.08	1002.00
5	-	Ban Muang Kao 1	Non Thai	Banlang	170363	1690249	902.00	0.87	872.00
6	-	Ban Muang Kao 2	Non Thai	Banlang	170357	1688964	487.00	0.89	894.00
7	-	Ban Muang Kao 3	Non Thai	Banlang	170013	1689327	812.00	0.84	972.32
8	SC409	Ban Muang Kao Temple	Non Thai	Banlang	170683	1689675	483.00	0.48	885.43
9	-	Ban Sawai	Non Thai	Sai O	181412	1691223	952.00	0.97	1161.04
10	PW13862	Ban Nong Ta Maen	Non Thai	Thanon Pho	192879	1693814	607.00	0.71	899.21
11	MG1571	Nong Doom Health Promoting Hospital	Non Thai	Makha	187792	1689952	705.00	0.58	556.04
12	-	Ban Dong Plong	Non Sung	Lum Khao	215050	1687249	14,400.00	7.54	7,540.00
13	-	Ban Krok Kham	Non Sung	Lum Khao	209452	1684940	3,074.00	3.76	3,548.00
14	-	Ban Talad Kae	Non Sung	Than Prasat	222250	1687690	1,064.00	1.03	964.00
15	-	Ban Bu	Non Sung	Dan Khla	203592	1674430	1,045.00	0.51	138.00
16	D0516	Ban Non Makok	Non Sung	Tanot	206821	1671468	6,720.00	2.70	3,912.00
17	-	Ban Salao	Non Sung	Phon Sungkham	212794	1698603	995.00	0.96	203.00
18	MY610	Ban Non-Thong	Khong	Non Teng	206847	1708043	467.00	0.45	103.31
19	MY853	Ban Non-Thong Temple	Khong	Non Teng	206466	1707680	879.00	0.57	358.19
20	PW16	Ban Nong Bua Gra Jai School	Khong	Non Teng	207618	1709139	838.00	0.83	941.93

Table A1 Chemical properties of groundwater samples in May 2019 (Continued).

No	Station	Location	district	Sub-district	UTM E	UTM N	TDS	SC	CC
21	5805D004	Ban Don Klang School	Khong	Nong Manao	207201	1716400	247.00	0.25	358.54
22	5805A004	Ban Taluk Nam Khwang	Khong	Nong Manao	206169	1716165	1,746.00	1.98	2,958.42
23	MG679	Ban Nong Sakae School	Khong	Nong Bua	195999	1717793	867.00	0.86	948.01
24	AFD815	Ban Tha Yai 1	Khong	Nong Bua	197134	1717526	504.00	0.55	293.12
25	MY619	Ban Tha Yai 2	Khong	Nong Bua	197358	1717630	799.00	0.77	442.81
26	5505G052	Ban Ta Kim	Khong	Nong Bua	192508	1717165	2,006.00	3.30	3,942.48
27	25111	Ban Ba Dao Rueang	Khong	Nong Bua	197316	1714455	2,172.00	3.62	4,091.32
28	5405B019	Ban Don Du	Khong	Mueang Khong	214581	1714842	1,031.00	1.01	1,002.41
29	MG1014	Ban Kok Pet 1	Khong	Mueang Khong	214630	1710905	2,273.00	3.01	4,101.25
30	MG832	Ban Kok Pet 2	Khong	Mueang Khong	214775	1711005	921.00	0.91	1,049.26
31	MY327	Ban Kok Pet Temple	Khong	Mueang Khong	215532	1711669	997.00	0.98	1,101.94
32	PW20092	Ban Mai Don Tua Pap	Khong	Ban Prang	184962	1715211	1,784.00	2.05	2168.53
33	MG1017	Ban Don Tua	Khong	Ban Prang	185261	1715430	869.00	0.85	741.39
34	AFD806	Ban Huay Luek	Khong	Ban Prang	184921	1712189	1,551.00	0.58	673.84
35	5905H052	Ban Thap Ma Kham School	Khong	Ban Prang	184934	1710462	3,404.00	3.12	6,117.99
36	5805B035	Ban Non Mueang School	Kham Sakae Saeng	Non Mueang	191889	1706585	835.00	0.53	1,095.37
37	SC1145	Ban Taluk Hin	Kham Sakae Saeng	Non Mueang	188468	1705543	517.00	0.55	414.000
38	MY331	Ban Ngio School	Kham Sakae Saeng	Non Mueang	191207	1708251	1,074.00	1.60	1,968.94
39	MY337	Ban Khum Muang	Kham Sakae Saeng	Non Mueang	189356	1706970	988.00	0.90	1,076.16
40	MG320	Ban Sa Kruat	Kham Sakae Saeng	Non Mueang	191778	1706565	716.00	0.56	683.66

Table A1 Chemical properties of groundwater samples in May 2019 (Continued).

No	Station	Location	district	Sub-district	UTM E	UTM N	TDS	SC	CC
41	5605B031	Ban Khu Mueang School	Kham Sakae Saeng	Mueang Kaset	195665	1706447	537.00	0.59	437.54
42	5905D059	Chomcho Nong Hua Fan School	Kham Sakae Saeng	Nong Hua Fan	201683	1706461	4,297.00	4.38	7,305.28
43	MY228	Ban Non Makluea Temple	Kham Sakae Saeng	Nong Hua Fan	198289	1710131	1,719.00	1.43	2,175.90
44	5505C054	Ban Non Makluea School	Kham Sakae Saeng	Nong Hua Fan	198165	1710145	1,487.00	1.48	2,311.79
45	MG1637	Ban Jod School	Kham Sakae Saeng	Nong Hua Fan	207082	1705864	1,998.00	1.45	2,310.77
46	5905H029	Chiwuek 1	Kham Sakae Saeng	Chiwuek	188847	1695635	480.00	0.49	358.54
47	5805H044	Chiwuek 2	Kham Sakae Saeng	Chiwuek	187952	1700628	1,247.00	1.39	2,060.40
48	PW73	Ban Hua Bung	Kham Sakae Saeng	Chiwuek	188906	1696058	490.00	0.24	227.90
49	MY458	Ban Nong Pho	Kham Sakae Saeng	Chiwuek	187583	1697185	1,069.00	0.81	948.01
50	5805H049	Ban Non-Phak Chi School	Kham Sakae Saeng	Chiwuek	186671	1698911	1,487.00	1.43	2,217.99
51	MG1441	Ban Nook 1	Kham Sakae Saeng	Kham Sakae Saeng	201285	1698994	587.00	0.99	1,266.29
52	MG1442	Ban Nook 2	Kham Sakae Saeng	Kham Sakae Saeng	201393	1698950	2,050.00	0.88	805.20
53	MG1443	Ban Nook School	Kham Sakae Saeng	Kham Sakae Saeng	200904	1699036	1,602.00	1.31	1,777.51
54	MG318	Ban Namab	Kham Sakae Saeng	Kham Sakae Saeng	203193	1699010	600.00	0.59	734.55
55	AFD998	Ban Bu La Kro	Kham Sakae Saeng	Kham Sakae Saeng	196920	1694693	1,447.00	0.97	1,253.37
56	5705B005	Ban Sema school	Kham Sakae Saeng	Mueang Nat	208529	1701210	1,494.00	1.42	1,975.01
57	AA1655	Ban Nong Pho Namab School	Kham Sakae Saeng	Mueang Nat	204767	1699728	537.00	0.51	492.48



APPENDIX B

Google Earth Engine Metadata and Scripts

1. OpenLandMap Soil Texture Class (USDA System)

Soil texture classes (USDA system) for 6 soil depths (0, 10, 30, 60, 100, and 200 cm) at 250 m. Derived from predicted soil texture fractions using the soiltexture package in R.

Table B1 Bands of Soil texture class (USDA system).

Name	Description
b0	Soil texture class (USDA system) at 0 cm depth
b10	Soil texture class (USDA system) at 10 cm depth
b30	Soil texture class (USDA system) at 30 cm depth
b60	Soil texture class (USDA system) at 60 cm depth
b100	Soil texture class (USDA system) at 100 cm depth
b200	Soil texture class (USDA system) at 200 cm depth

Table B2 Bands Class Table of Soil texture class (USDA system).

Value	Symbol	Description
1	Cl	Clay
2	SiCl	Silty Clay
3	SaCl	Sandy Clay
4	CLLo	Clayey Loam
5	SiCLLo	Silty Clay Loam
6	SaCLLo	Sandy Clay Loam
7	Lo	Loam
8	SiLo	Silty Loam
9	SaLo	Sandy Loam
10	Si	Silt
11	LoSa	Loamy Sand
12	Sa	Sand

GEE Scripts

```

var dataset = ee.Image("OpenLandMap/SOL/SOL_TEXTURE-CLASS_USDA-TT_M/v02");

var visualization = {
  bands: ['b0'],
  min: 1.0,
  max: 12.0,
  palette: [
    "d5c36b","b96947","9d3706","ae868f","f86714","46d143",
    "368f20","3e5a14","ffd557","fff72e","ff5a9d","ff005b",
  ]
};

Map.centerObject(dataset);

Map.addLayer(dataset, visualization, "Soil texture class (USDA system)");

```

2. Global SRTM Landforms

The SRTM Landform dataset provides landform classes created by combining the Continuous Heat-Insolation Load Index (SRTM CHILI) and the multi-scale Topographic Position Index (SRTM mTPI) datasets. It is based on the 30m SRTM DEM (available in EE as USGS/SRTMGL1_003).

The Conservation Science Partners (CSP) Ecologically Relevant Geomorphology (ERGo) Datasets, Landforms, and Physiography contain detailed, multi-scale data on landforms and physiographic (aka land facet) patterns. Although there are many potential uses for these data, the original purpose for these data was to develop an ecologically relevant classification and map of landforms and physiographic classes that are suitable for climate adaptation planning. Because there is a large uncertainty associated with future climate conditions and even more uncertainty around ecological responses, providing information about what is unlikely to change offers a strong foundation for managers to build robust climate adaptation plans. The quantification

of these landscape features is sensitive to the resolution, The highest resolution possible is provided given the index's extent and characteristics.

Table B3 Bands of Global SRTM Landforms.

Name	Description
constant	SRTM-derived landform classes

Table B4 Bands Class Table of Global SRTM Landforms.

Value	Description
12	Peak/ridge
13	Peak/ridge (cool)
14	Mountain/divide
15	Cliff
21	Upper slope (warm)
22	Upper slope
23	Upper slope (cool)
24	Upper slope (flat)
31	Lower slope (warm)
32	Lower slope
33	Lower slope (cool)
34	Lower slope (flat)
41	Valley
42	Valley (narrow)

GEE Scripts

```

var dataset = ee.Image('CSP/ERGo/1_0/Global/SRTM_landforms');
var landforms = dataset.select('constant');
var landformsVis = {
  min: 11.0,
  max: 42.0,
  palette: [
    '141414', '383838', '808080', 'EBEB8F', 'F7D311', 'AA0000', 'D89382',
    'DDC9C9', 'DCCDCE', '1C6330', '68AA63', 'B5C98E', 'E1F0E5', 'a975ba',
    '6f198c'
  ],
};
Map.setCenter(-105.58, 40.5498, 11);
Map.addLayer(landforms, landformsVis, 'Landforms');

```

3. Sentinel-2 10m Land Use/Land Cover timeseries (2017-2021)

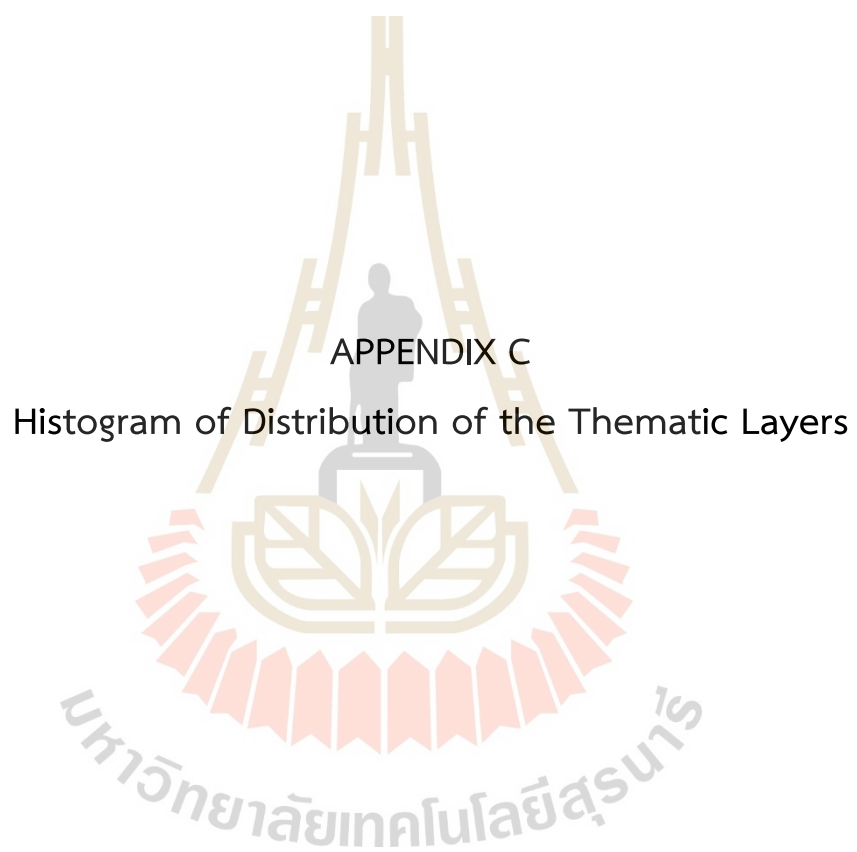
This application provides access to the 700 individual 10-meter resolution GeoTIFF files for each year of the Sentinel-2 Land Use/Land Cover map produced by Esri, Microsoft, and Impact Observatory. The map is derived from ESA Sentinel-2 imagery and is a composite of land use/land cover predictions for 9 classes for each year from 2017-2021.

Table B5 Land Cover Class Sentinel-2 10m Land Use/Land Cover time-series.

Category	Land Cover Class	Description
1	No Data	-
2	Water	Water Areas where water was predominantly present throughout the year; may not cover areas with sporadic or ephemeral water; contains little to no sparse vegetation, no rock outcrop nor built up features like docks; examples: rivers, ponds, lakes, oceans, flooded salt plains.
3	Trees	Trees Any significant clustering of tall (~15-m or higher) dense vegetation, typically with a closed or dense canopy; examples: wooded vegetation, clusters of dense tall vegetation within savannas, plantations, swamp or mangroves (dense/tall vegetation with ephemeral water or canopy too thick to detect water underneath).
4	Grass	Grass Open areas covered in homogeneous grasses with little to no taller vegetation; wild cereals and grasses with no obvious human plotting (i.e., not a plotted field); examples: natural meadows and fields with sparse to no tree cover, open savanna with few to no trees, parks/golf courses/lawns, pastures.
5	Flooded Vegetation	Flooded vegetation Areas of any type of vegetation with obvious intermixing of water throughout a majority of the year; seasonally flooded area that is a mix of grass/shrub/trees/bare ground; examples: flooded mangroves, emergent vegetation, rice paddies and other heavily irrigated and inundated agriculture.

Table B5 Land Cover Class Sentinel-2 10m Land Use/Land Cover time-series (Continued).

6	Crops	Crops Human planted/plotted cereals, grasses, and crops not at tree height; examples: corn, wheat, soy, fallow plots of structured land.
7	Scrub/Shrub	Scrub/shrub Mix of small clusters of plants or single plants dispersed on a landscape that shows exposed soil or rock; scrub-filled clearings within dense forests that are clearly not taller than trees; examples: moderate to sparse cover of bushes, shrubs and tufts of grass, savannas with very sparse grasses, trees or other plants
8	Built Area	Built Area Human made structures; major road and rail networks; large homogeneous impervious surfaces including parking structures, office buildings and residential housing; examples: houses, dense villages / towns / cities, paved roads, asphalt.
9	Bare Ground	Bare ground Areas of rock or soil with very sparse to no vegetation for the entire year; large areas of sand and deserts with no to little vegetation; examples: exposed rock or soil, desert and sand dunes, dry salt flats/pans, dried lake beds, mines.
10	Snow/Ice	Snow/Ice Large homogeneous areas of permanent snow or ice, typically only in mountain areas or highest latitudes; examples: glaciers, permanent snowpack, snow fields.
11	Clouds	Clouds No land cover information due to persistent cloud cover.



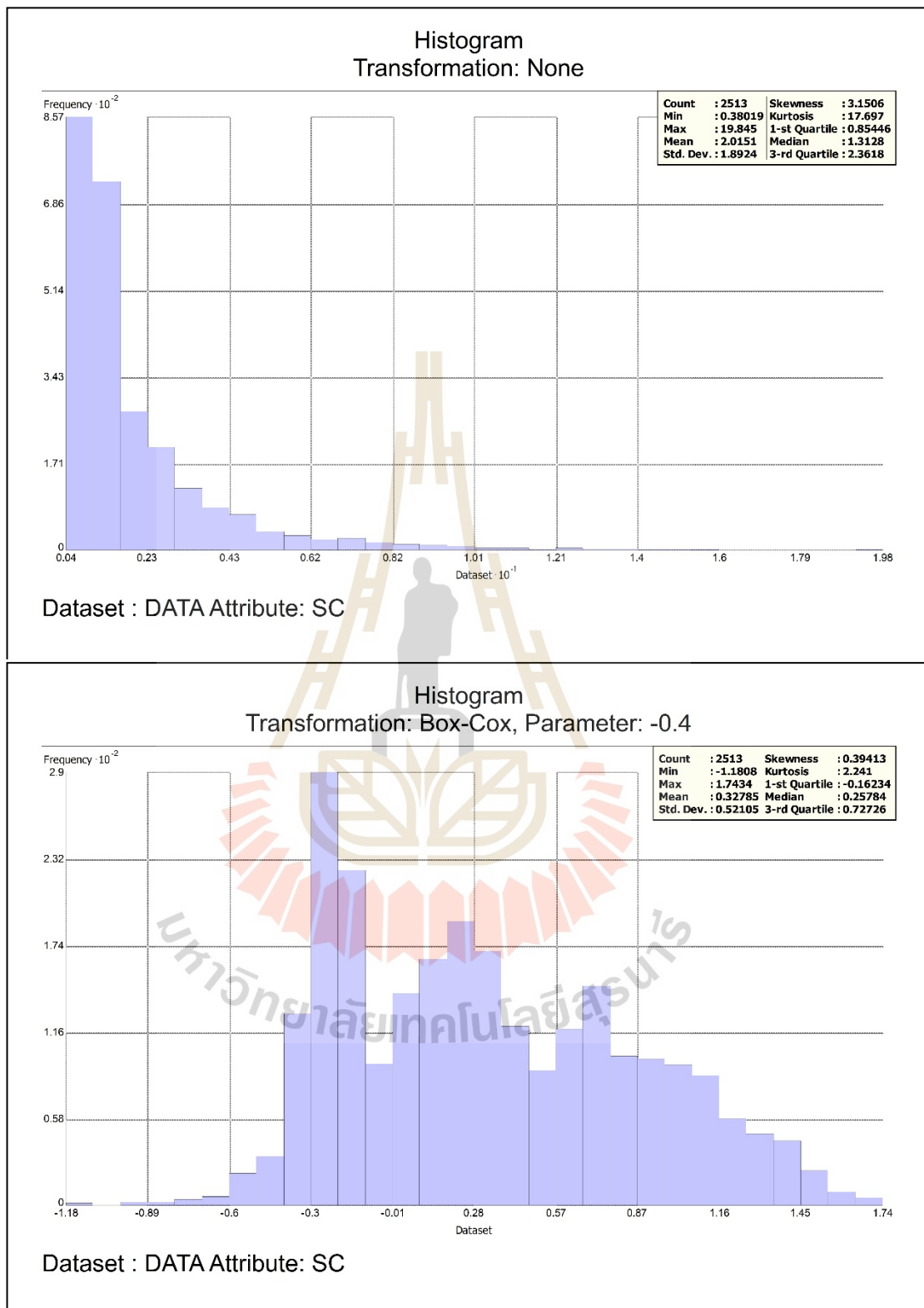


Figure C1 Histogram plot of Salinity Content (SC) data and summary statistics.

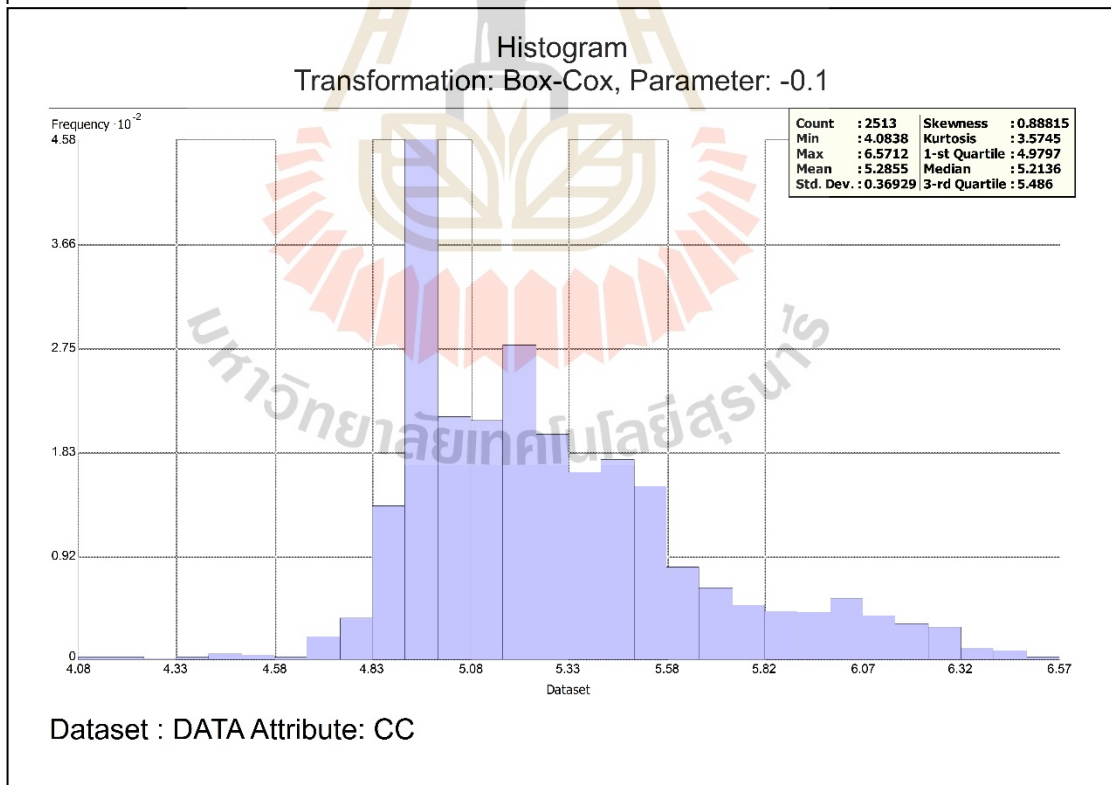
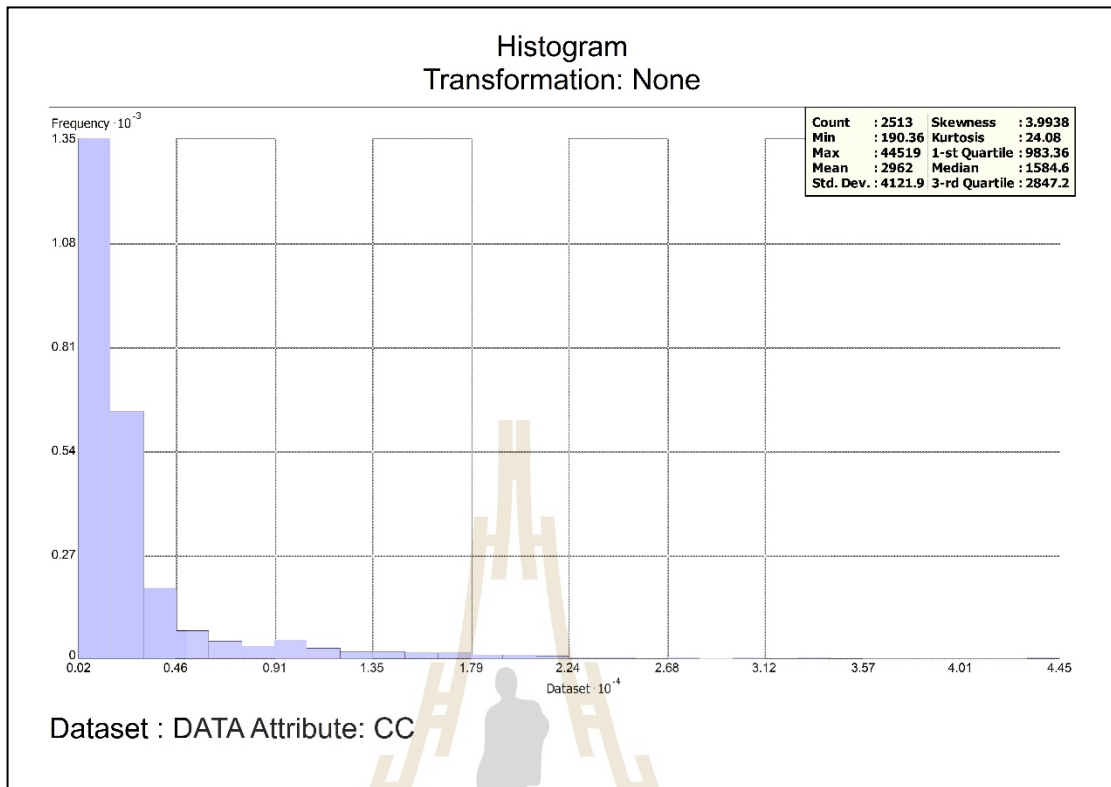


Figure C2 Histogram plot of Chloride Content (CC) data and summary statistics.

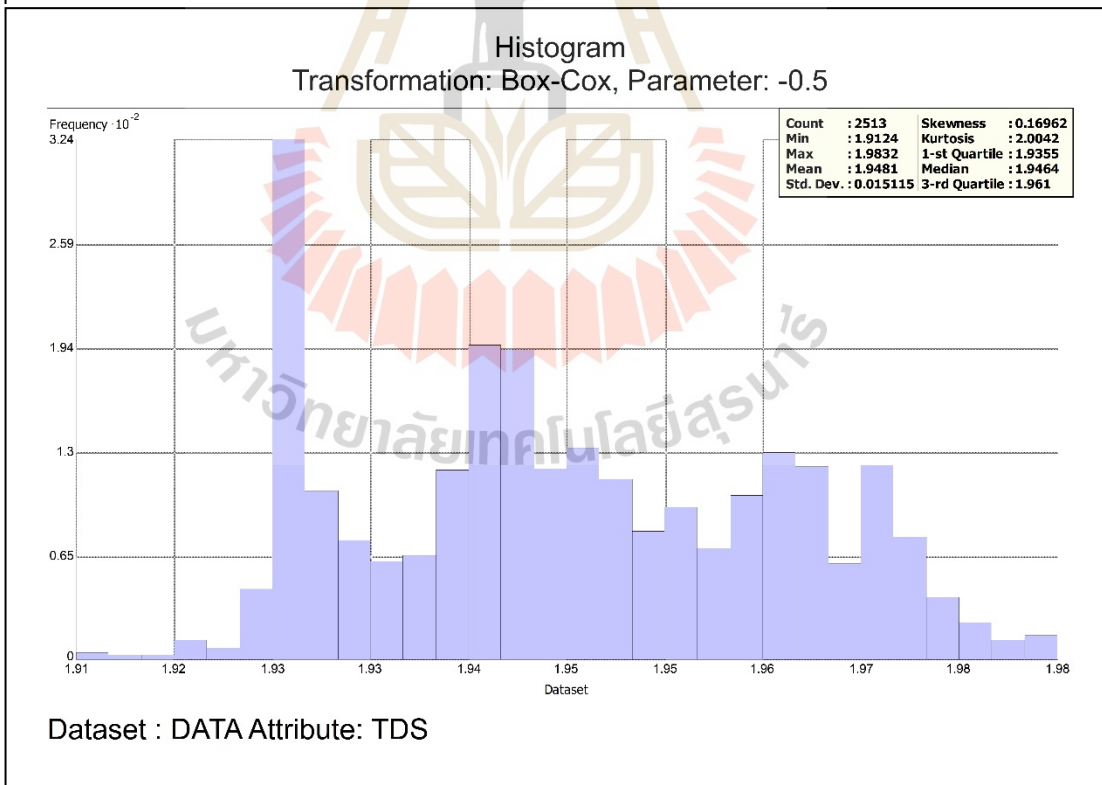
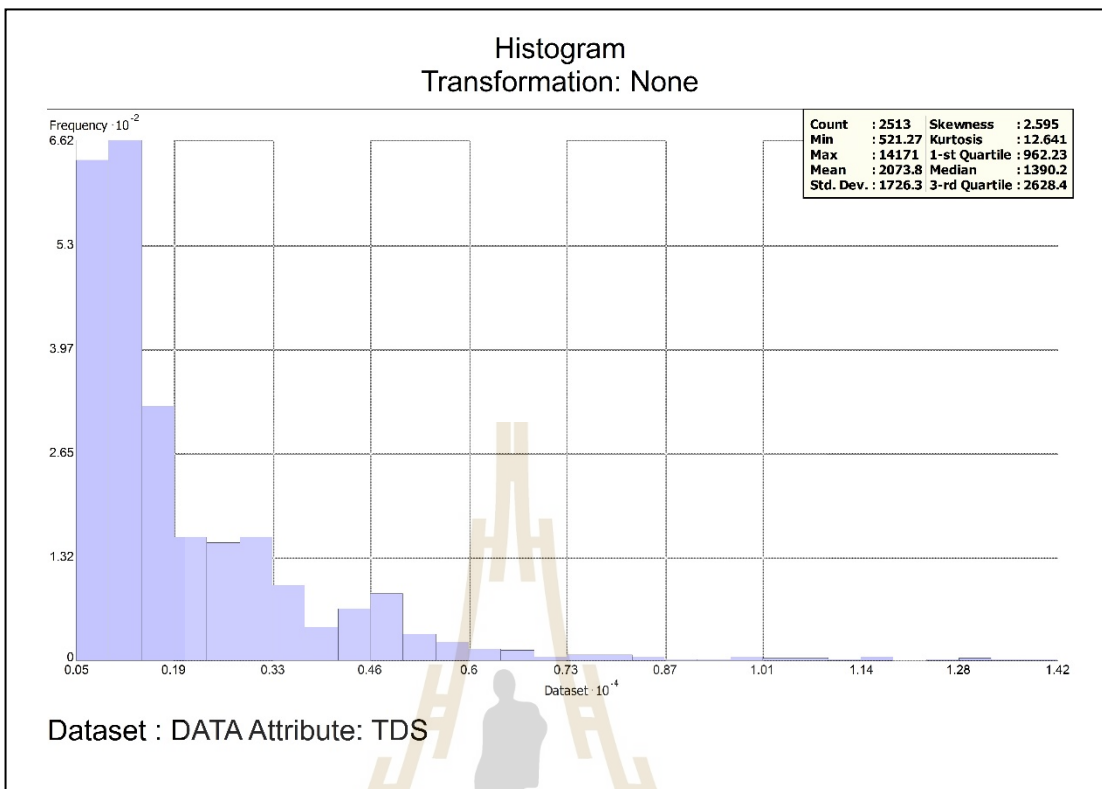


Figure C3 Histogram plot of Total Solid Content (TDS) data and summary statistics.

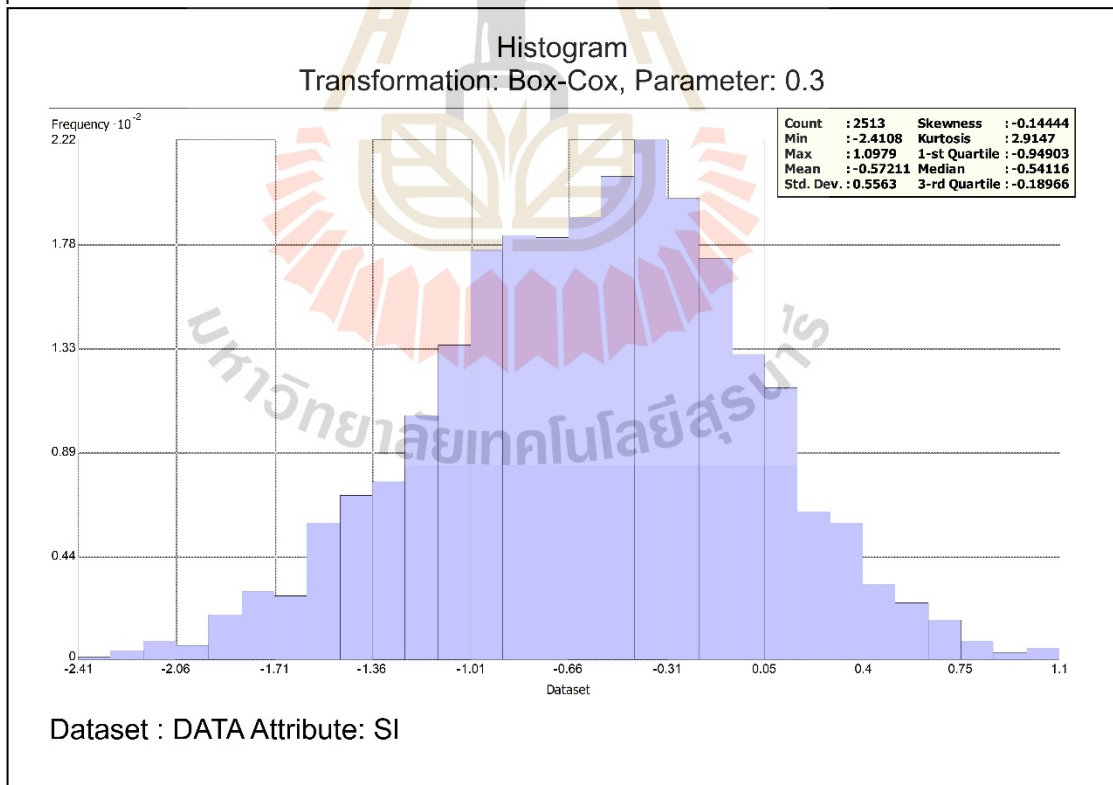
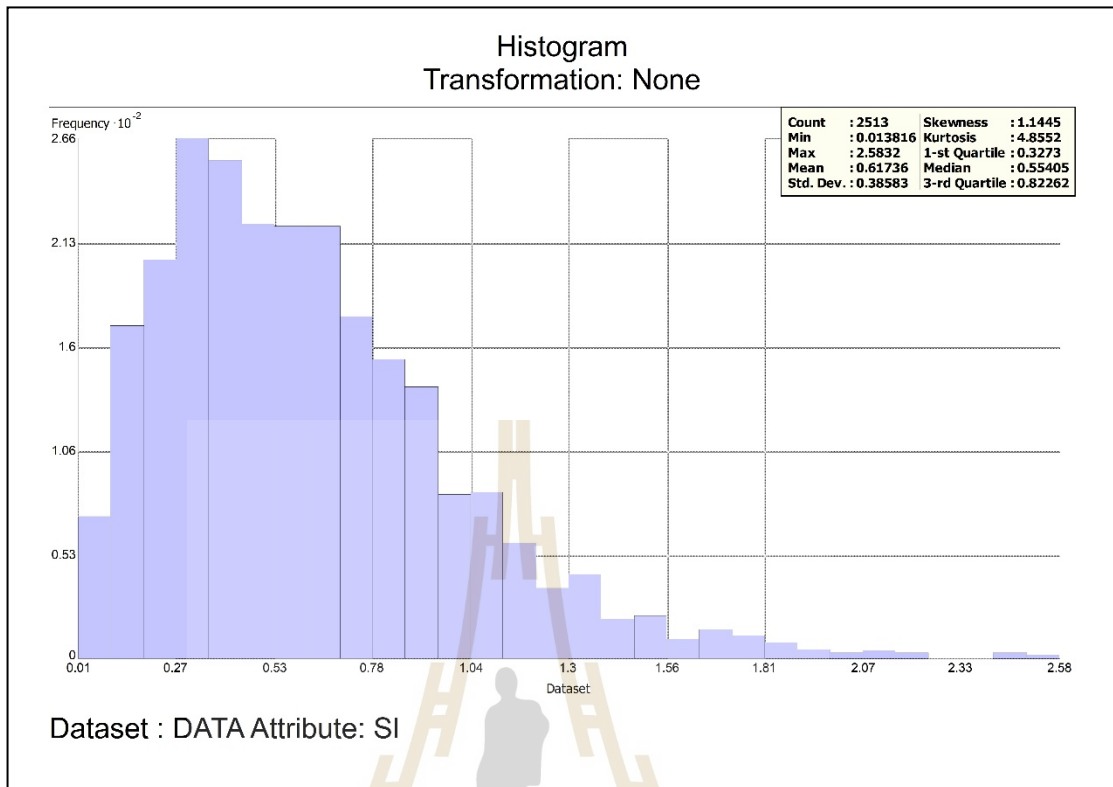


Figure C4 Histogram plot of Slope (SI) data and summary statistics.

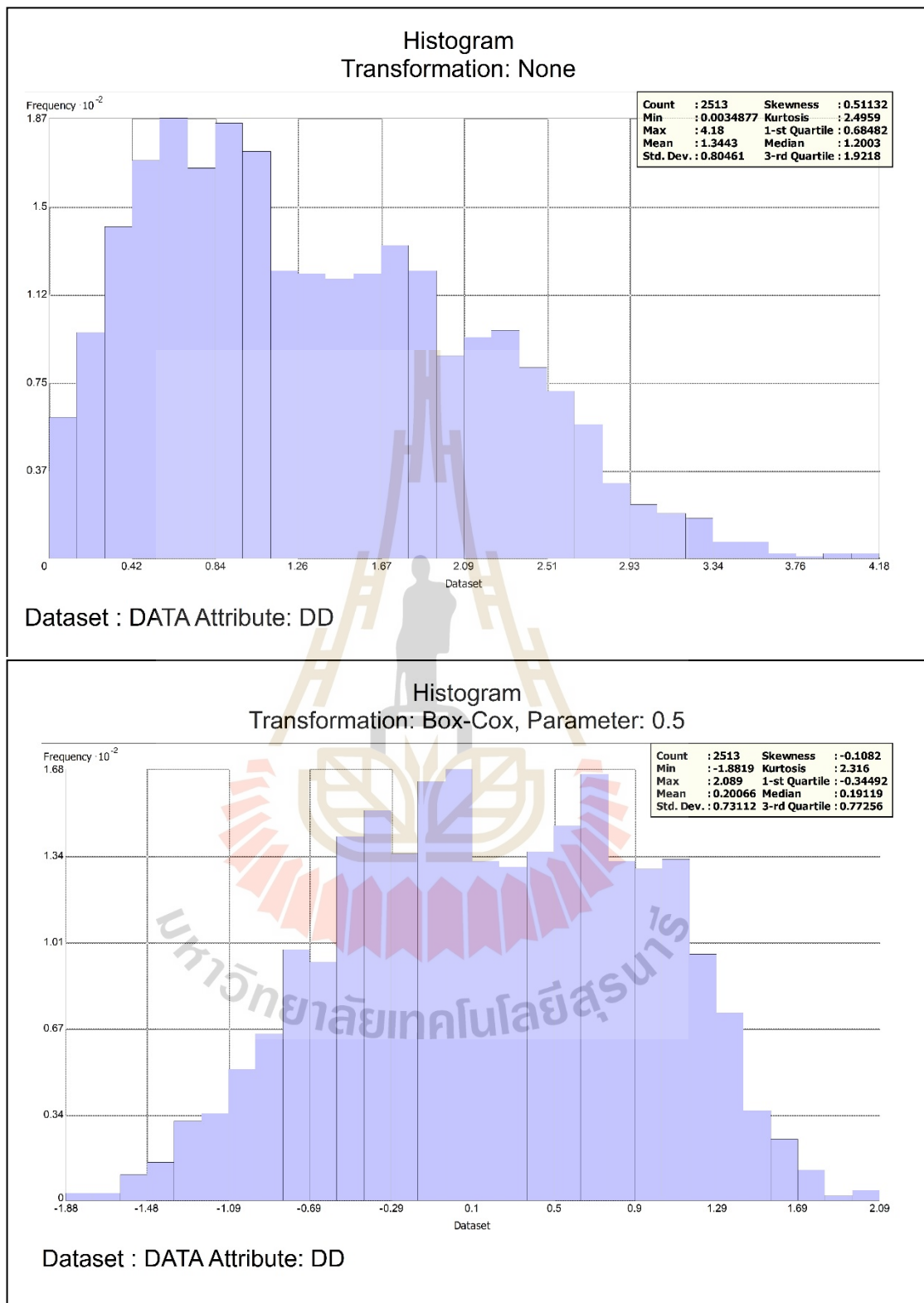


Figure C5 Histogram plot of Drainage Density (DD) data and summary statistics.

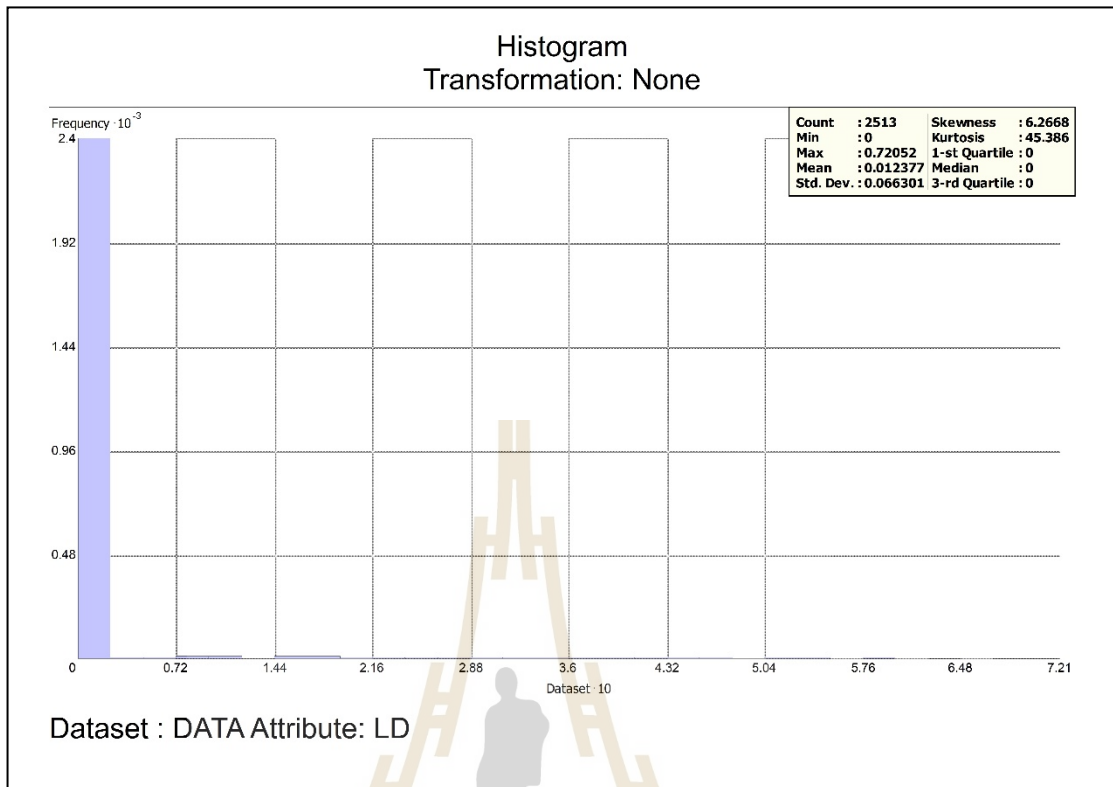


Figure C6 Histogram plot of Lineament Density (LD) data and summary statistics.



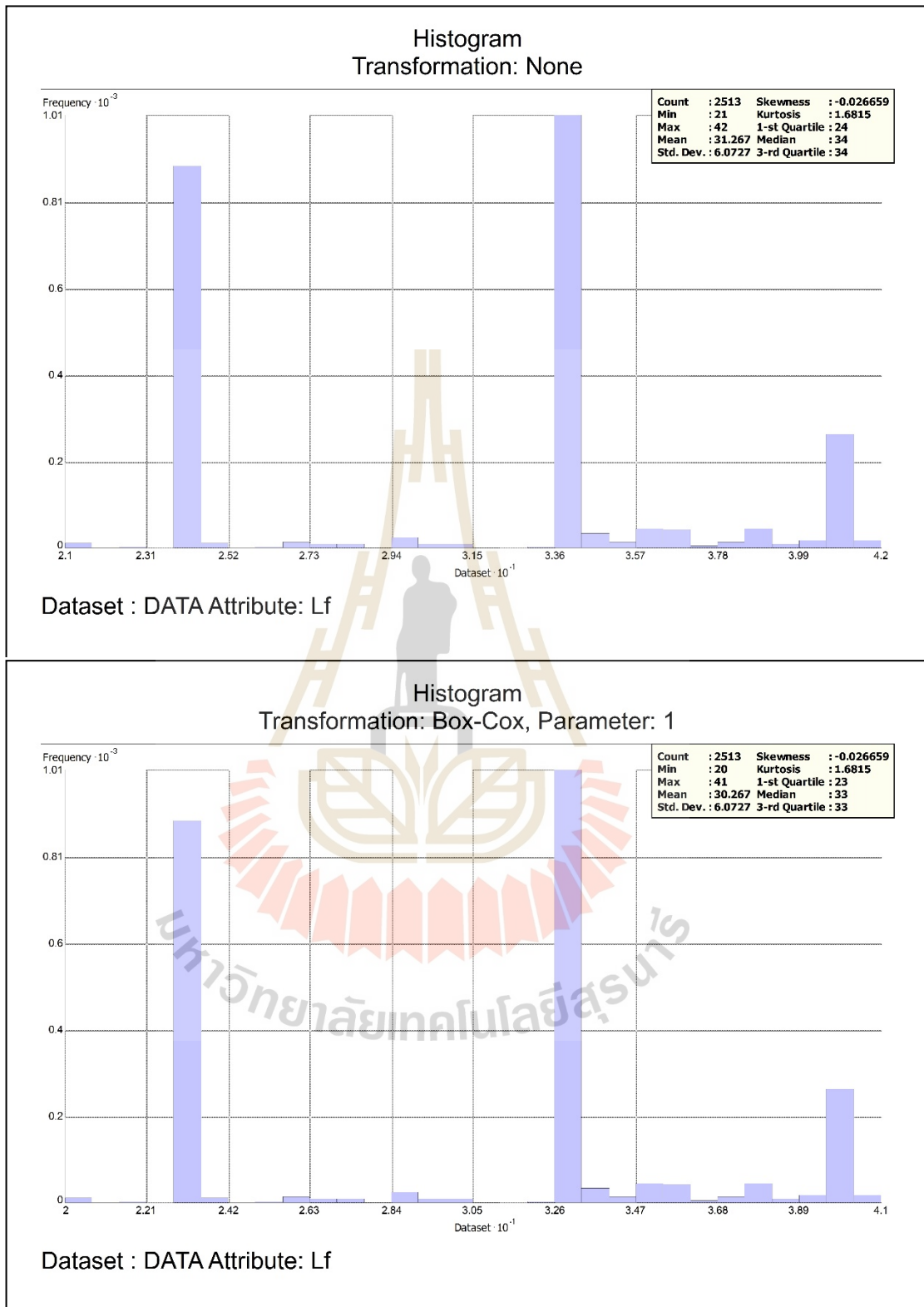


Figure C7 Histogram plot of Landforms (Lf) data and summary statistics.

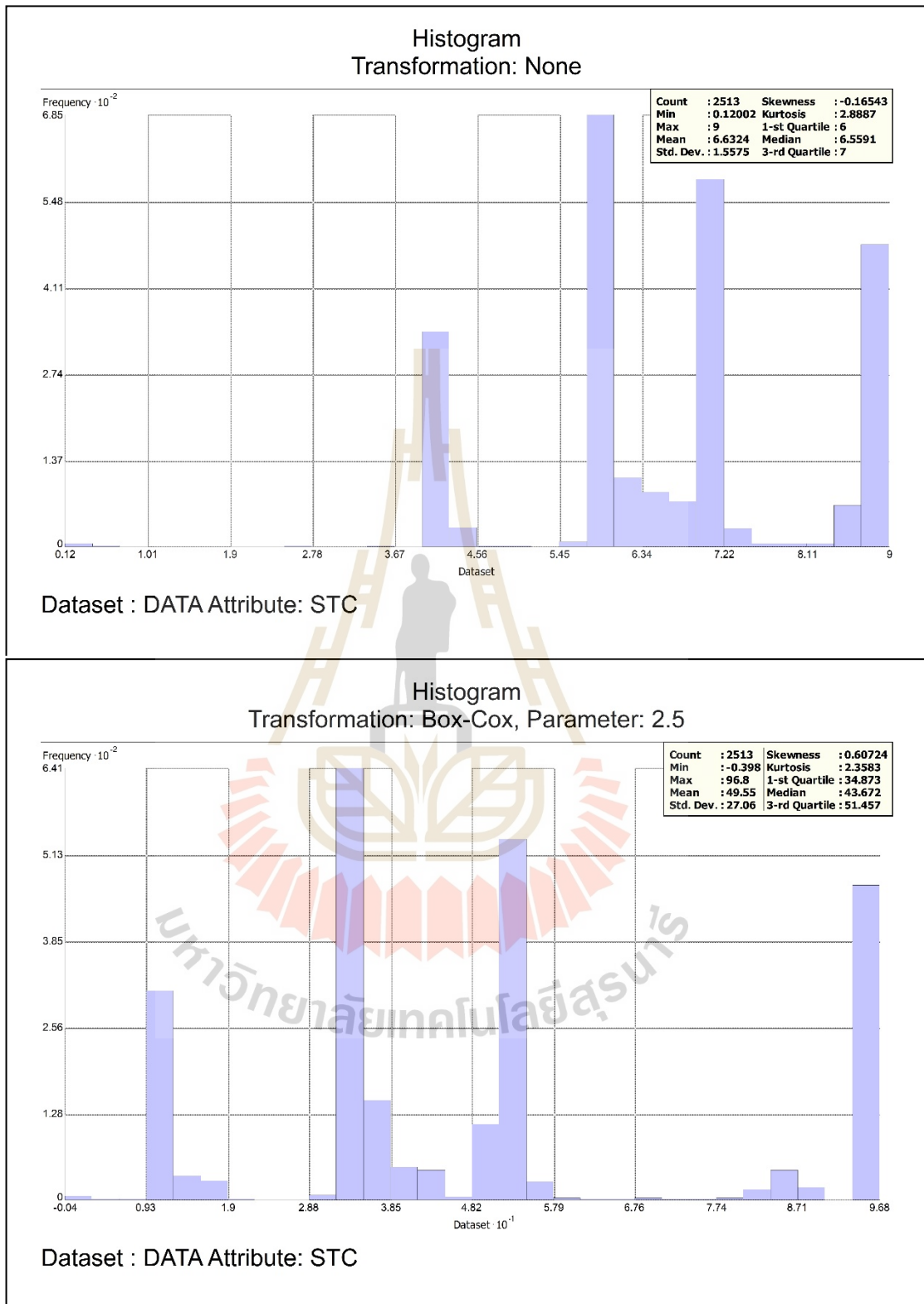


Figure C8 Histogram plot of Soil Texture Class (STC) data and summary statistics.

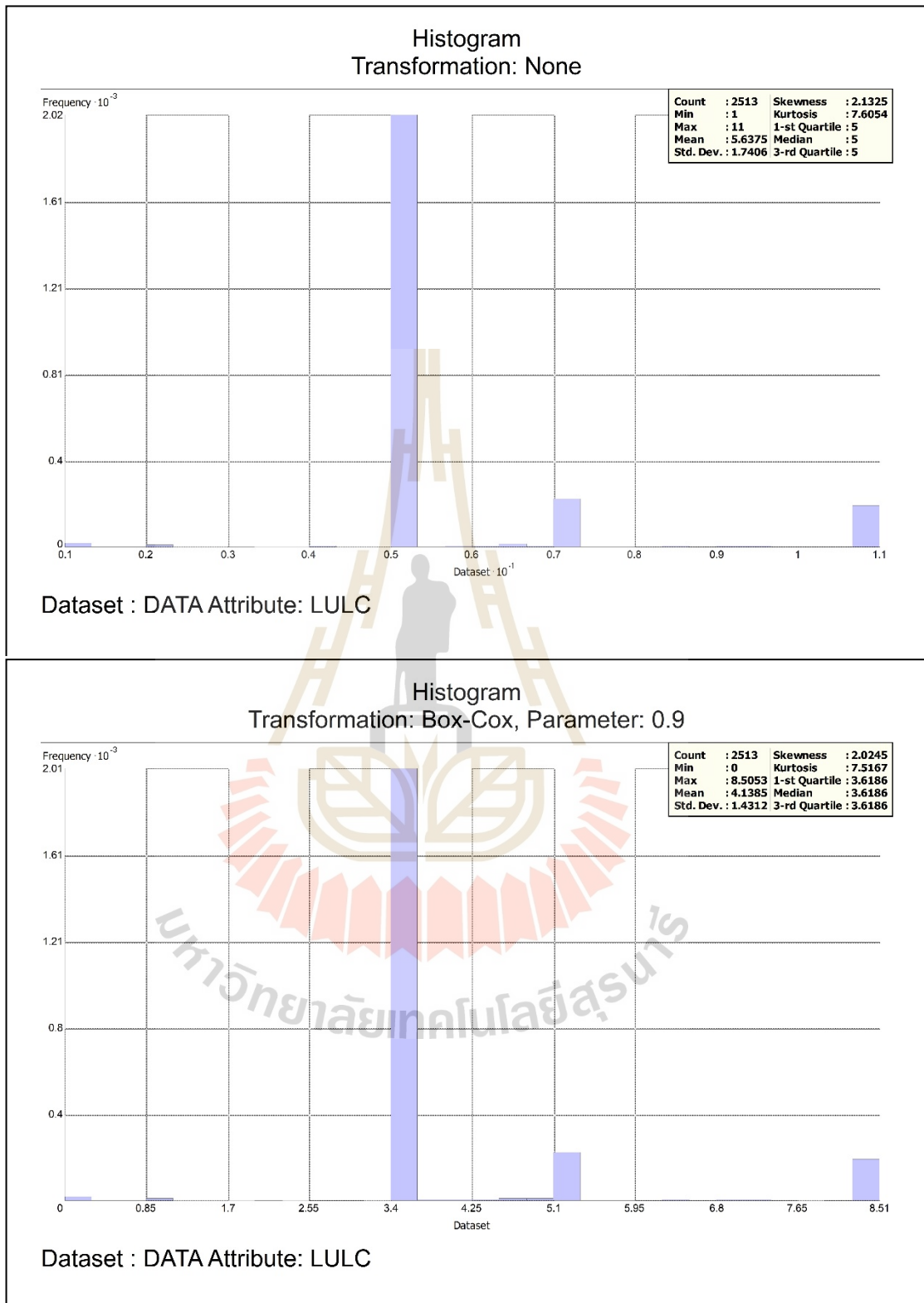


Figure C9 Histogram plot of Land use/ Land cover (LULC) data and summary statistics.

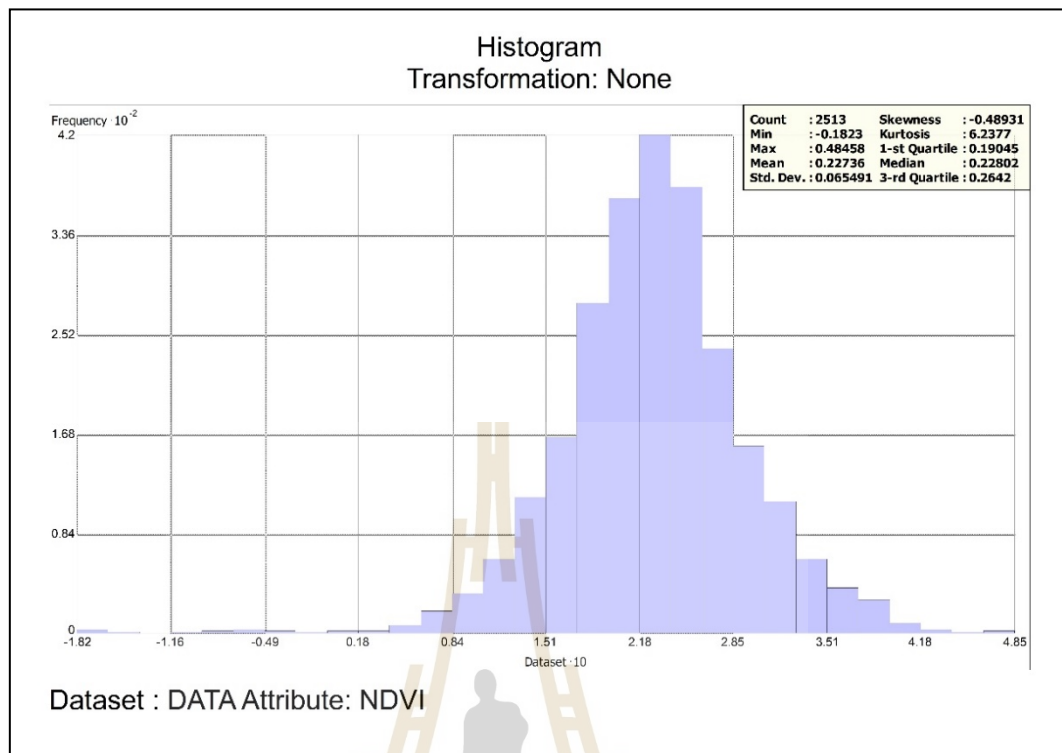


Figure C10 Histogram plot of Normalized Differentiation Vegetation Index (NDVI) data and summary statistics.

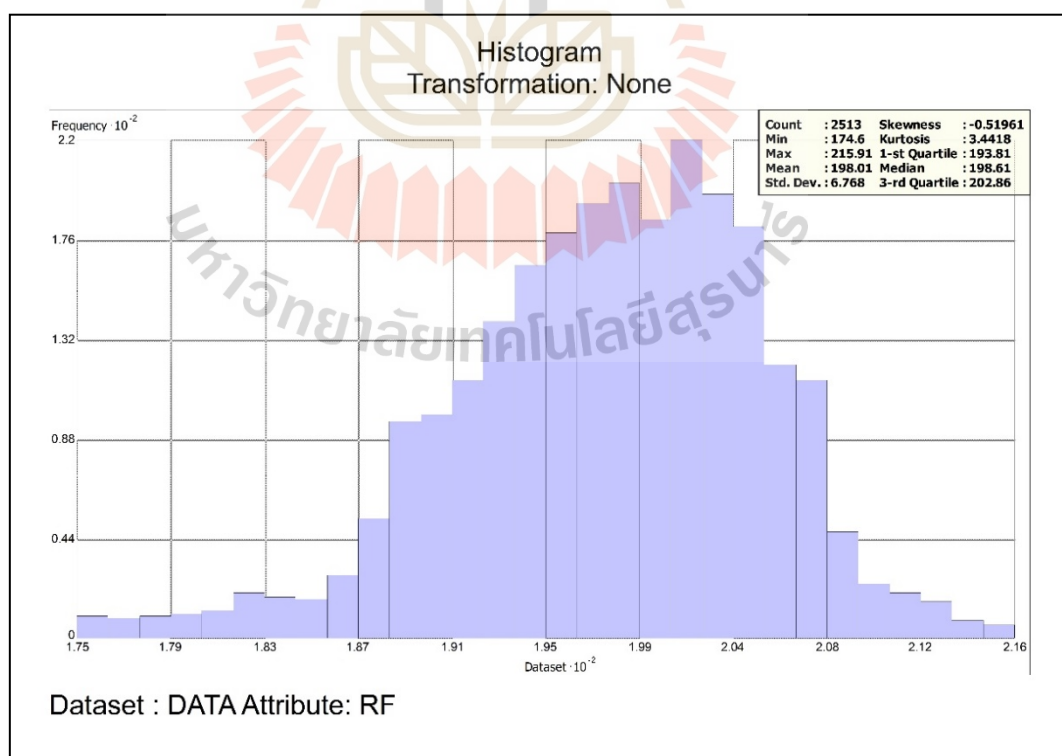


Figure C11 Histogram plot of Mean Annual Rainfall (RF) data and summary statistics.

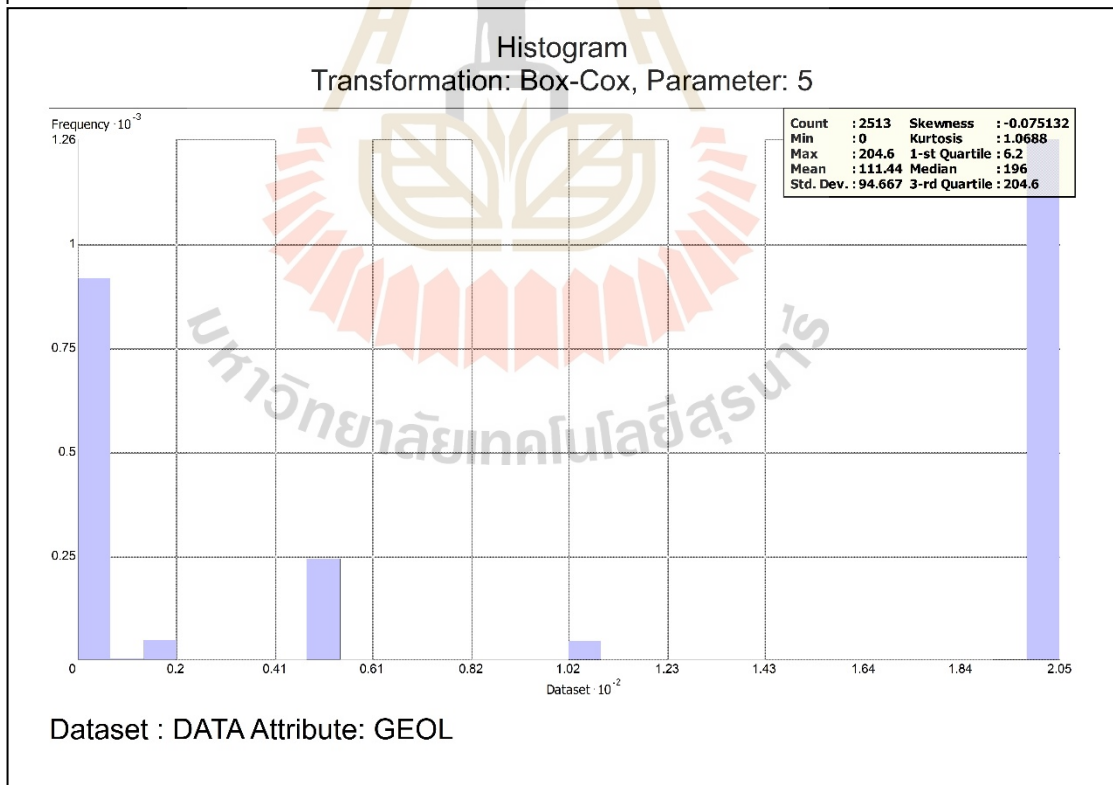
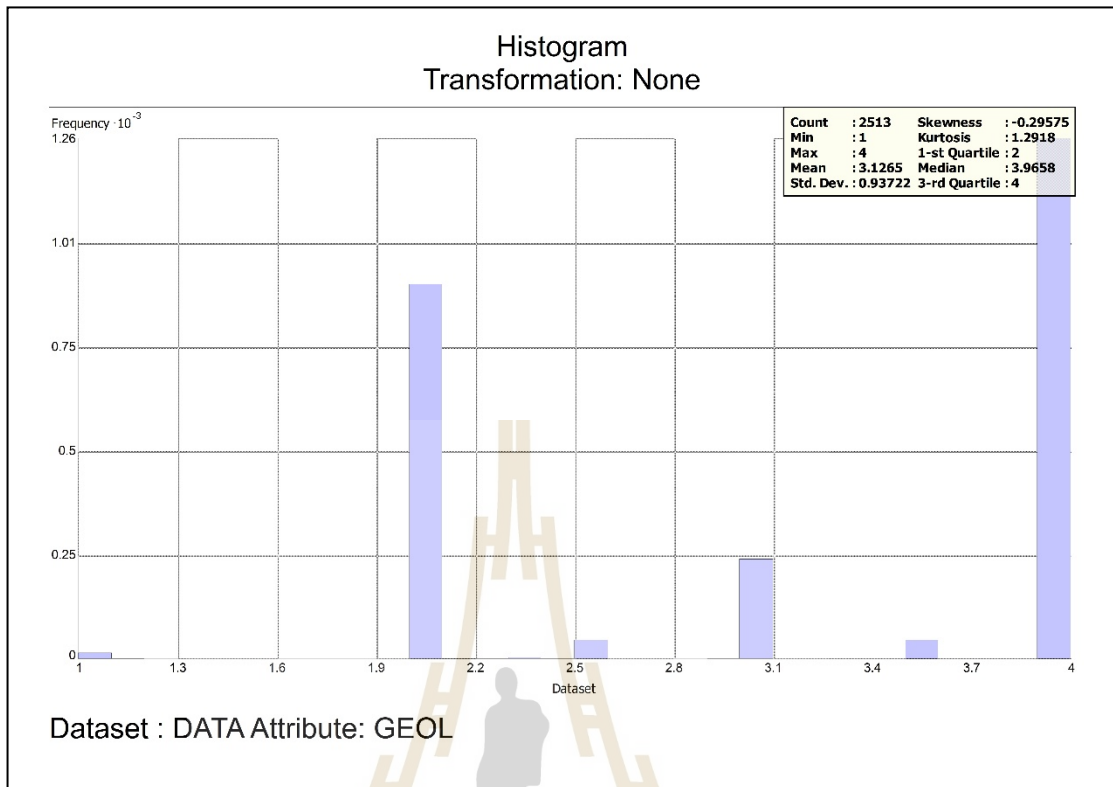


Figure C12 Histogram plot of Geology (Geol) data and summary statistics.

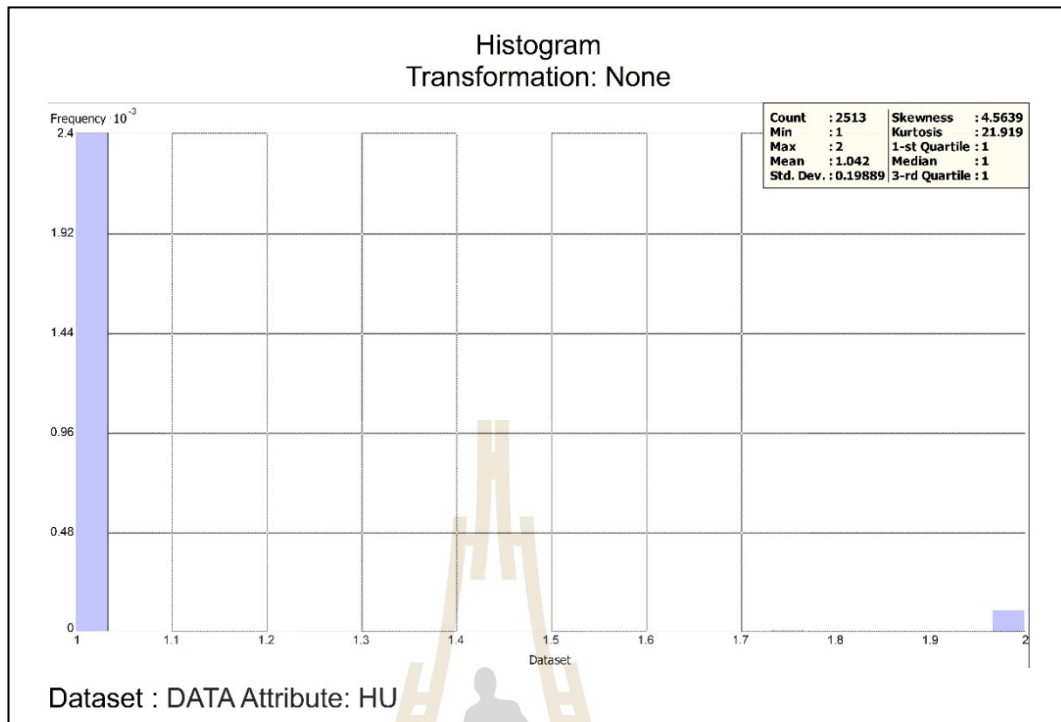


Figure C13 Histogram plot of Hydrogeological Unit (HU) data and summary statistics.

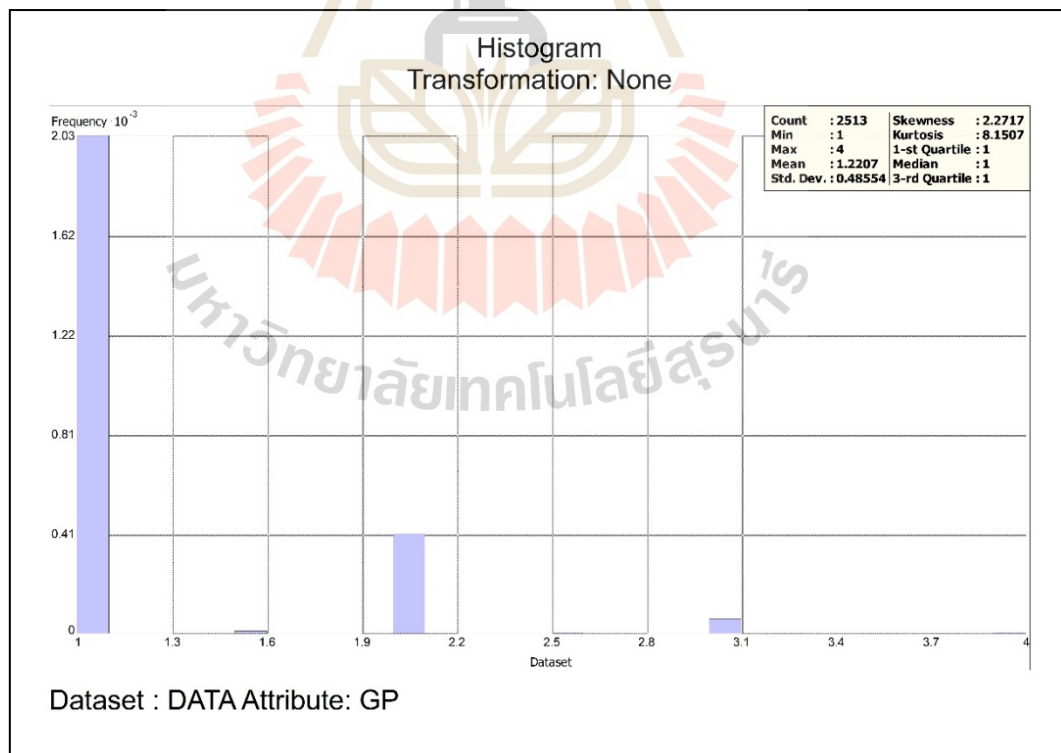
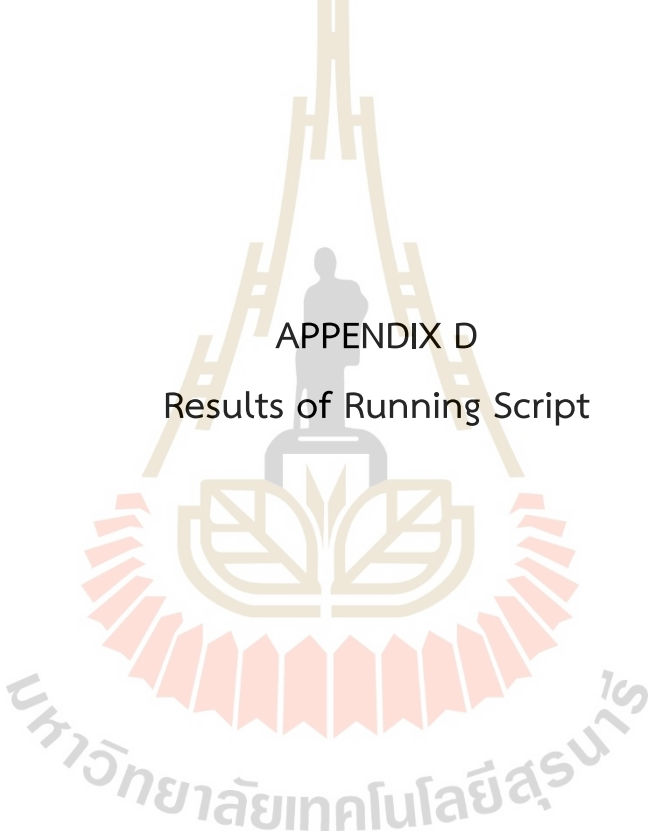


Figure C14 Histogram plot of Groundwater Potential (GP) data and summary statistics.

The logo of Sakon Nakhon Rajabhat University is a large, faint watermark in the background. It features a central figure of a person standing on a pedestal, flanked by two stylized figures. Above the central figure is a large, golden 'H' shape. The entire logo is set against a circular base with a scalloped edge. The text 'มหาวิทยาลัยเทคโนโลยีสุรนารี' is written in Thai script along the bottom curve of the logo.

APPENDIX D
Results of Running Script

Calculate Distance Band

Calculate Distance Band Summary

Minimum 8 neighbor distance: 1414.213562

Average 8 neighbor distance: 1504.429429

Maximum 8 neighbor distance: 2961.865937

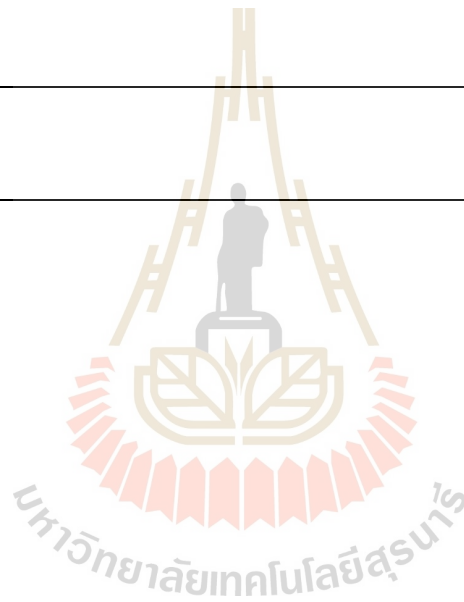
Distance measured in Meters

Incremental Spatial Autocorrelation

Running Moran's I at Varying Distances...

Global Moran's I Summary by Distance

Distance	Moran's Index	Expected Index	Variance	z-score	p-value
2961.87	0.908992	-0.000398	0.000035	153.681745	0.000000
4825.28	0.843329	-0.000398	0.000013	233.042488	0.000000
6688.69	0.776456	-0.000398	0.000007	298.015862	0.000000
8552.10	0.719583	-0.000398	0.000004	355.082777	0.000000
10415.52	0.672964	-0.000398	0.000003	399.550575	0.000000
12278.93	0.621863	-0.000398	0.000002	435.354580	0.000000
14142.34	0.571002	-0.000398	0.000002	456.921045	0.000000
16005.76	0.523503	-0.000398	0.000001	470.057625	0.000000
17869.17	0.473170	-0.000398	0.000001	476.003152	0.000000
19732.58	0.424022	-0.000398	0.000001	475.547757	0.000000
21595.99	0.380170	-0.000398	0.000001	472.813727	0.000000
23459.41	0.342374	-0.000398	0.000001	470.480305	0.000000
25322.82	0.308353	-0.000398	0.000000	466.686174	0.000000
27186.23	0.280652	-0.000398	0.000000	466.718332	0.000000



29049.65	0.256630	-0.000398	0.000000	468.862839	0.000000
30913.06	0.234645	-0.000398	0.000000	473.546832	0.000000
32776.47	0.214408	-0.000398	0.000000	478.731997	0.000000
34639.88	0.196319	-0.000398	0.000000	482.724570	0.000000
36503.30	0.179313	-0.000398	0.000000	489.458597	0.000000
38366.71	0.163570	-0.000398	0.000000	494.837220	0.000000
40230.12	0.148577	-0.000398	0.000000	497.457915	0.000000
42093.54	0.133480	-0.000398	0.000000	498.102712	0.000000
43956.95	0.119261	-0.000398	0.000000	495.594790	0.000000
45820.36	0.105141	-0.000398	0.000000	487.905864	0.000000
47683.77	0.092020	-0.000398	0.000000	478.245803	0.000000
49547.19	0.080608	-0.000398	0.000000	468.814601	0.000000
51410.60	0.070874	-0.000398	0.000000	461.628657	0.000000
53274.01	0.062312	-0.000398	0.000000	455.858820	0.000000
55137.43	0.055105	-0.000398	0.000000	449.531676	0.000000
57000.84	0.048607	-0.000398	0.000000	443.183179	0.000000

First Peak (Distance; Value): 17869.17; 476.003152

Max Peak (Distance; Value): 42093.54; 498.102712

Distance measured in Meters



Generate Spatial Weights Matrix

Constructing spatial weights based on distance criteria...

Spatial Weights Matrix Summary

Number of Features: 2513

Percentage of Spatial Connectivity: 26.22

Average Number of Neighbors:	658.92
Minimum Number of Neighbors:	170
Maximum Number of Neighbors:	999
Distance measured in Meters	

Exploratory Regression

Choose 10 of 13 Summary

Highest Adjusted R-Squared Results

AdjR2	AICc	JB	K(BP)	VIF	SA	Model
0.98	-5456.68	0.00	0.00	5.66	N/A	+LD*** -NDVI*** +RF*** -GP*** +CC1*** +TDS1*** -SL1*** -LF1*** +STC1*** +GEOL1***
0.98	-5454.96	0.00	0.00	5.71	N/A	+LD*** -NDVI*** +RF*** -GP*** +CC1*** +TDS1*** -SL1*** +DD1*** -LF1*** +STC1***
0.98	-5453.24	0.00	0.00	5.71	N/A	+LD*** -NDVI*** +RF*** -GP*** +CC1*** +TDS1*** +DD1*** -LF1*** +STC1*** +GEOL1***

Passing Models

AdjR2	AICc	JB	K(BP)	VIF	SA	Model
0.975538	-5456.682781	0.000000	0.000000	5.656793	N/A	+LD*** -NDVI*** +RF*** -GP*** +CC1*** +TDS1*** -SL1*** -LF1*** +STC1*** +GEOL1***
0.975521	-5454.962230	0.000000	0.000000	5.714087	N/A	+LD*** -NDVI*** +RF*** -GP*** +CC1*** +TDS1*** -SL1*** +DD1*** -LF1*** +STC1***
0.975505	-5453.244185	0.000000	0.000000	5.710472	N/A	+LD*** -NDVI*** +RF*** -GP*** +CC1*** +TDS1*** +DD1*** -LF1*** +STC1*** +GEOL1***
0.975490	-5451.753456	0.000000	0.000000	5.601404	N/A	+LD*** -NDVI*** +RF*** -HU*** +CC1*** +TDS1*** -SL1*** -LF1*** +STC1*** +GEOL1***
0.975438	-5446.425345	0.000000	0.000000	5.669943	N/A	+LD*** -NDVI*** +RF*** -HU*** -GP*** +CC1*** +TDS1*** -SL1*** -LF1*** +GEOL1***
0.975435	-5446.098378	0.000000	0.000000	5.685535	N/A	+LD*** +RF*** -GP*** +CC1*** +TDS1*** -SL1*** +DD1*** -LF1*** +STC1*** +GEOL1***
0.975426	-5445.218571	0.000000	0.000000	5.611921	N/A	+LD*** +RF*** -HU*** -GP*** +CC1*** +TDS1*** -SL1*** -LF1*** +STC1*** +GEOL1***
0.975409	-5443.491720	0.000000	0.000000	5.607336	N/A	+LD*** +RF*** -GP*** +CC1*** +TDS1*** -SL1*** -LF1*** +STC1*** -LULC1*** +GEOL1***
0.975405	-5443.037604	0.000000	0.000000	5.677566	N/A	+LD*** +RF*** -GP*** +CC1*** +TDS1*** -SL1*** +DD1*** -LF1*** +STC1*** -LULC1***
0.975380	-5440.544671	0.000000	0.000000	5.612963	N/A	+LD*** +RF*** -HU*** -GP*** +CC1*** +TDS1*** -SL1*** +STC1*** -LULC1*** +GEOL1***
0.975369	-5439.372101	0.000000	0.000000	5.614119	N/A	+LD*** +RF*** -HU*** -GP*** +CC1*** +TDS1*** -SL1*** -LF1*** -LULC1*** +GEOL1***

0.975360 -5438.416702 0.000000 0.000000 5.669143 N/A +LD*** +RF*** -GP*** +CC1*** +TDS1*** +DD1*** -LF1*** +STC1*** -LULC1*** +GEOL1***
 0.975341 -5436.544803 0.000000 0.000000 5.593307 N/A +LD*** +RF*** -HU*** -GP*** +CC1*** +TDS1*** -LF1*** +STC1*** -LULC1*** +GEOL1***

Choose 11 of 13 Summary

Highest Adjusted R-Squared Results

AdjR2	AICc	JB	K(BP)	VIF	SA	Model
0.98	-5459.84	0.00	0.00	5.67	N/A	+LD*** -NDVI*** +RF*** -HU*** -GP*** +CC1*** +TDS1*** -SL1*** -LF1*** +STC1*** +GEOL1***
0.98	-5459.80	0.00	0.00	5.72	N/A	+LD*** -NDVI*** +RF*** -GP*** +CC1*** +TDS1*** -SL1*** +DD1*** -LF1*** +STC1*** +GEOL1***
0.98	-5458.55	0.00	0.00	5.66	N/A	+LD*** -NDVI*** +RF*** -GP*** +CC1*** +TDS1*** -SL1*** -LF1*** +STC1*** -LULC1** +GEOL1***

Passing Models

AdjR2	AICc	JB	K(BP)	VIF	SA	Model
0.975578	-5459.796567	0.000000	0.000000	5.723712	N/A	+LD*** -NDVI*** +RF*** -GP*** +CC1*** +TDS1*** -SL1*** +DD1*** -LF1*** +STC1*** +GEOL1***
0.975477	-5449.440978	0.000000	0.000000	5.689441	N/A	+LD*** +RF*** -GP*** +CC1*** +TDS1*** -SL1*** +DD1*** -LF1*** +STC1*** -LULC1*** +GEOL1***
0.975466	-5448.254087	0.000000	0.000000	5.615859	N/A	+LD*** +RF*** -HU*** -GP*** +CC1*** +TDS1*** -SL1*** -LF1*** +STC1*** -LULC1*** +GEOL1***

Choose 12 of 13 Summary

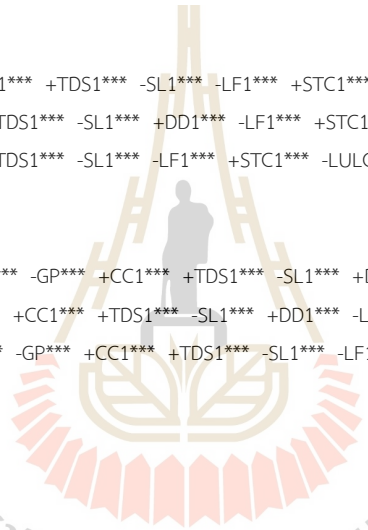
Highest Adjusted R-Squared Results

AdjR2	AICc	JB	K(BP)	VIF	SA	Model
0.98	-5461.83	0.00	0.00	5.73	N/A	+LD*** -NDVI*** +RF*** -GP*** +CC1*** +TDS1*** -SL1*** +DD1*** -LF1*** +STC1*** -LULC1** +GEOL1***
0.98	-5461.61	0.00	0.00	5.68	N/A	+LD*** -NDVI*** +RF*** -HU*** -GP*** +CC1*** +TDS1*** -SL1*** -LF1*** +STC1*** -LULC1** +GEOL1***
0.98	-5460.96	0.00	0.00	5.76	N/A	+LD*** -NDVI*** +RF*** -HU* -GP*** +CC1*** +TDS1*** -SL1*** +DD1** -LF1*** +STC1*** +GEOL1***

Passing Models

AdjR2	AICc	JB	K(BP)	VIF	SA	Model
-------	------	----	-------	-----	----	-------

Choose 13 of 13 Summary



Highest Adjusted R-Squared Results

AdjR2 AICc JB K(BP) VIF SA Model

0.98 -5462.88 0.00 0.00 5.76 N/A +LD*** -NDVI*** +RF*** -HU* -GP*** +CC1*** +TDS1*** -SL1*** +DD1** -LF1*** +STC1*** -LULC1** +GEOL1***

Passing Models

AdjR2 AICc JB K(BP) VIF SA Model

***** Exploratory Regression Global Summary (SC1) *****

Percentage of Search Criteria Passed

Search Criterion Cutoff Trials # Passed % Passed

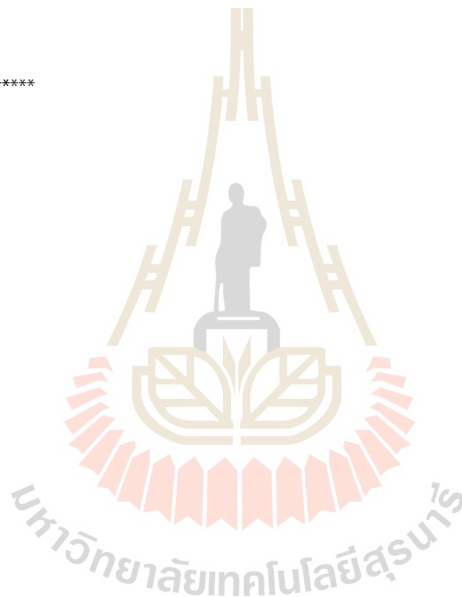
Min Adjusted R-Squared > 0.80 2380 2035 85.50

Max Coefficient p-value < 0.01 2380 116 4.87

Max VIF Value < 7.50 2380 2380 100.00

Min Jarque-Bera p-value > 0.00 2380 2380 100.00

Min Spatial Autocorrelation p-value > 0.00 122 122 100.00



Summary of Variable Significance

Variable % Significant % Negative % Positive

RF	100.00	35.44	64.56
CC1	100.00	0.00	100.00
TDS1	100.00	0.00	100.00
LD	98.87	0.00	100.00
DD1	73.39	34.68	65.32
HU	68.03	74.15	25.85
GEOL1	65.45	0.06	99.94

NDVI	62.17	75.66	24.34
SL1	59.27	66.96	33.04
STC1	36.89	29.45	70.55
LF1	35.31	84.87	15.13
GP	32.16	52.27	47.73
LULC1	24.53	92.50	7.50

Summary of Multicollinearity

Variable VIF Violations Covariates

LD	1.14	0	-----
NDVI	1.31	0	-----
RF	3.42	0	-----
HU	1.30	0	-----
GP	1.11	0	-----
CC1	2.78	0	-----
TDS1	5.76	0	-----
SL1	1.19	0	-----
DD1	1.32	0	-----
LF1	1.20	0	-----
STC1	1.13	0	-----
LULC1	1.05	0	-----
GEOL1	1.57	0	-----



Summary of Residual Normality (JB)

JB	AdjR2	AICc	K(BP)	VIF	SA	Model
0.000000	0.975246	-5428.846532	0.000000	5.609893	N/A	+LD*** -NDVI*** +RF*** -HU*** -GP*** +CC1*** +TDS1*** -LF1**
0.000000	0.975240	-5428.257111	0.000000	5.730310	N/A	+LD*** -NDVI*** +RF*** -HU** -GP*** +CC1*** +TDS1*** +DD1***
0.000000	0.975286	-5432.983783	0.000000	5.631842	N/A	+LD*** -NDVI*** +RF*** -HU*** -GP*** +CC1*** +TDS1*** -SL1***

Summary of Residual Spatial Autocorrelation (SA) (Not Applicable)

Table Abbreviations

AdjR2 Adjusted R-Squared

AICc Akaike's Information Criterion

JB Jarque-Bera p-value

K(BP) Koenker (BP) Statistic p-value

VIF Max Variance Inflation Factor

SA Global Moran's I p-value

Model Variable sign (+/-)

Model Variable significance (* = 0.10; ** = 0.05; *** = 0.01)



Ordinary Least Squares Regression

Summary of OLS Results

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	-35.605687	0.540819	-65.836591	0.000000*	0.527825	-67.457326	0.000000*	-----
LD	0.355430	0.026170	13.581471	0.000000*	0.023625	15.044411	0.000000*	1.142010

NDVI	-0.107198	0.028275	-3.791310	0.000165*	0.028825	-3.718949	0.000217*	1.300689
RF	0.005731	0.000432	13.277345	0.000000*	0.000424	13.517349	0.000000*	3.237558
GP	-0.012500	0.003519	-3.552338	0.000404*	0.003888	-3.214989	0.001337*	1.107265
CC1	0.915191	0.007244	126.331509	0.000000*	0.009925	92.207515	0.000000*	2.714861
TDS1	15.420554	0.257127	59.972550	0.000000*	0.255014	60.469405	0.000000*	5.729843
SL1	-0.009607	0.003184	-3.017608	0.002585*	0.003355	-2.863518	0.004230*	1.189904
DD1	0.005698	0.002480	2.297442	0.021660*	0.002160	2.637592	0.008396*	1.247319
LF1	-0.006554	0.001616	-4.054770	0.000059*	0.001561	-4.198215	0.000033*	1.192347
STC1	0.000255	0.000064	4.011279	0.000070*	0.000064	4.012223	0.000070*	1.122606
LULC1	-0.002333	0.001162	-2.008506	0.044688*	0.000925	-2.522358	0.011708*	1.048693
GEOL1	0.000057	0.000021	2.663031	0.007790*	0.000022	2.653578	0.008010*	1.561706

OLS Diagnostics

Input Features: DATA Dependent Variable: SC1

Number of Observations: 2513 Akaike's Information Criterion (AICc) [d]: -5461.825896

Multiple R-Squared [d]: Adjusted R-Squared [d]: 0.975608

Joint F-Statistic [e]: 8373.634298 Prob(>F), (12,2500) degrees of freedom: 0.000000*

Joint Wald Statistic [e]: 117383.495120 Prob(>chi-squared), (12) degrees of freedom: 0.000000*

Koenker (BP) Statistic [f]: 156.086567 Prob(>chi-squared), (12) degrees of freedom: 0.000000*

Jarque-Bera Statistic [g]: 11143.947084 Prob(>chi-squared), (2) degrees of freedom: 0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

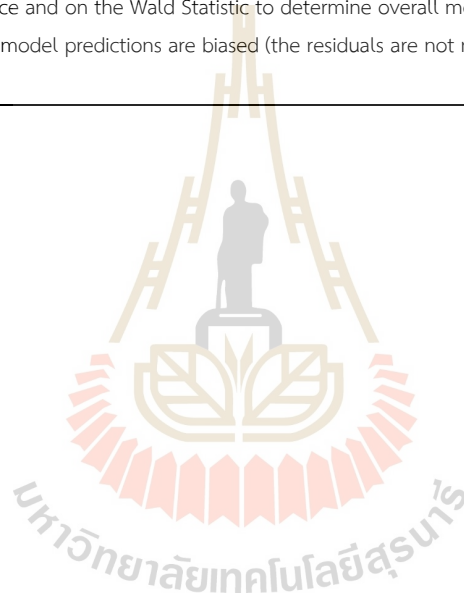
[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.01$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.01$) model predictions are biased (the residuals are not normally distributed).



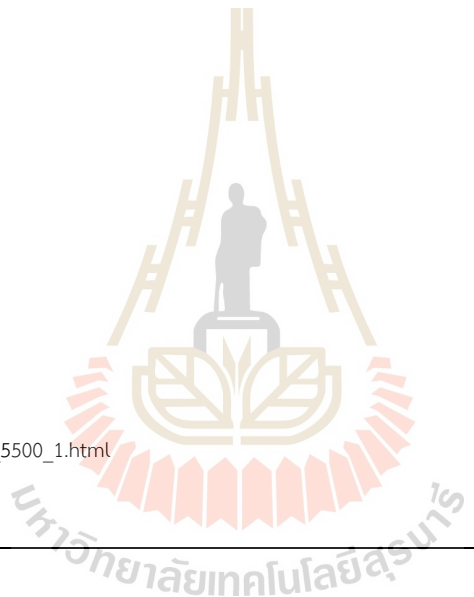
Spatial Autocorrelation

```
Executing: SpatialAutocorrelation OLS1 Residual GENERATE_REPORT INVERSE_DISTANCE EUCLIDEAN_DISTANCE NONE # #
Start Time: Fri Feb 10 00:56:17 2023
Running script SpatialAutocorrelation...
WARNING 000853: The default neighborhood search threshold was 1221.9316 Meters.

Global Moran's I Summary
Moran's Index: 0.879164
Expected Index: -0.000398
Variance: 0.000202
z-score: 61.864007
p-value: 0.000000

Distance measured in Meters

Writing html report...
C:\Users\WATCHA~1\AppData\Local\Temp\arc7490\MoransI_Result_3712_5500_1.html
Completed script SpatialAutocorrelation...
Succeeded at Fri Feb 10 00:56:19 2023 (Elapsed Time: 1.98 seconds)
```



Semiparametric Geographically Weighted Regression

```
*****
* Semiparametric Geographically Weighted Regression *
* Release 1.0.90 (GWR 4.0.90) *
* 12 May 2015 *
* (Originally coded by T. Nakaya: 1 Nov 2009) *
*
```

* Tomoki Nakaya(1), Martin Charlton(2), Chris Brunson (2) *
* Paul Lewis (2), Jing Yao (3), A Stewart Fotheringham (4) *
* (c) GWR4 development team *
* (1) Ritsumeikan University, (2) National University of Ireland, Maynooth, *
* (3) University of Glasgow, (4) Arizona State University *

Session:

Session control file: C:\Users\WATCHARINPHOEMPHON\OneDrive - Suranaree University of Technology\WatcharinP\01 Thesis\Thesis\GWR4\model\model1.ctl

Data filename: C:\Users\WATCHARINPHOEMPHON\OneDrive - Suranaree University of Technology\WatcharinP\01 Thesis\Thesis\GWR4\DATA.dbf

Number of areas/points: 2513

Model settings-----

Model type: Gaussian

Geographic kernel: adaptive Gaussian

Method for optimal bandwidth search: Golden section search

Criterion for optimal bandwidth: AICc

Number of varying coefficients: 13

Number of fixed coefficients: 0

Modelling options-----

Standardisation of independent variables: OFF

Testing geographical variability of local coefficients: OFF

Local to Global Variable selection: OFF

Global to Local Variable selection: OFF

Prediction at non-regression points: OFF

Variable settings-----



Area key: field5: ID
 Easting (x-coord): field3 : INSIDE_X
 Northing (y-coord): field4: INSIDE_Y
 Cartesian coordinates: Euclidean distance
 Dependent variable: field20: SC1
 Offset variable is not specified
 Intercept: varying (Local) intercept
 Independent variable with varying (Local) coefficient: field11: LD
 Independent variable with varying (Local) coefficient: field15: NDVI
 Independent variable with varying (Local) coefficient: field16: RF
 Independent variable with varying (Local) coefficient: field19: GP
 Independent variable with varying (Local) coefficient: field21: CC1
 Independent variable with varying (Local) coefficient: field22: TDS1
 Independent variable with varying (Local) coefficient: field23: S11
 Independent variable with varying (Local) coefficient: field24: DD1
 Independent variable with varying (Local) coefficient: field25: Lf1
 Independent variable with varying (Local) coefficient: field26: STC1
 Independent variable with varying (Local) coefficient: field27: LULC1
 Independent variable with varying (Local) coefficient: field28: Geol1



Global regression result

< Diagnostic information >

Residual sum of squares: 16.555494
 Number of parameters: 13

(Note: this num does not include an error variance term for a Gaussian model)

ML based global sigma estimate: 0.081166
 Unbiased global sigma estimate: 0.081377
 -2 log-likelihood: -5489.994000
 Classic AIC: -5461.994000
 AICc: -5461.825865
 BIC/MDL: -5380.384744
 CV: 0.006660
 R square: 0.975724
 Adjusted R square: 0.975598

Variable	Estimate	Standard Error	t(Est/SE)
Intercept	-35.605686	0.540819	-65.836591
LD	0.355430	0.026170	13.581472
NDVI	-0.107198	0.028275	-3.791310
RF	0.005731	0.000432	13.277343
GP	-0.012500	0.003519	-3.552337
CC1	0.915191	0.007244	126.331508
TDS1	15.420554	0.257127	59.972550
SI1	-0.009607	0.003184	-3.017607
DD1	0.005698	0.002480	2.297442
Lf1	-0.006554	0.001616	-4.054770
STC1	0.000255	0.000064	4.011278
LULC1	-0.002333	0.001162	-2.008506
Geol1	0.000057	0.000021	2.663031



 GWR (Geographically weighted regression) bandwidth selection

Bandwidth search <golden section search>

Limits: 66, 2513

Golden section search begins...

Initial values

pL Bandwidth: 66.000 Criterion: -9801.877

p1 Bandwidth: 118.087 Criterion: -8448.409

p2 Bandwidth: 150.279 Criterion: -7916.093

pU Bandwidth: 202.367 Criterion: -7424.522

iter 1 (p1) Bandwidth: 118.087 Criterion: -8448.409 Diff: 32.192

iter 2 (p1) Bandwidth: 98.192 Criterion: -8797.094 Diff: 19.896

iter 3 (p1) Bandwidth: 85.896 Criterion: -9218.393 Diff: 12.296

iter 4 (p1) Bandwidth: 78.296 Criterion: -9347.104 Diff: 7.599

iter 5 (p1) Bandwidth: 73.599 Criterion: -9387.270 Diff: 4.697

iter 6 (p1) Bandwidth: 70.697 Criterion: -9423.382 Diff: 2.903

iter 7 (p1) Bandwidth: 68.903 Criterion: -9776.327 Diff: 1.794

iter 8 (p1) Bandwidth: 67.794 Criterion: -9789.626 Diff: 1.109

The lower limit in your search has been selected as the optimal bandwidth size.

A new session is recommended to try with a smaller lowest limit of the bandwidth search.

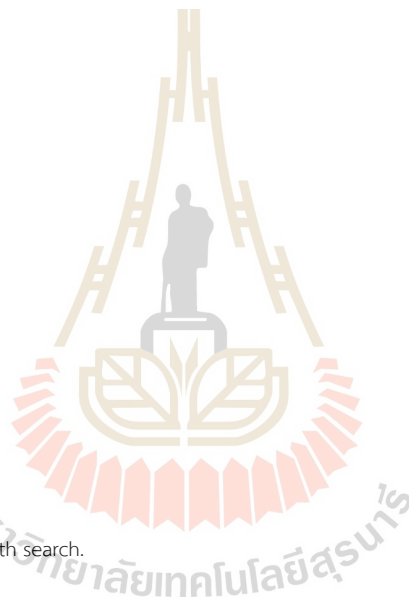
Best bandwidth size 66.000

Minimum AICc -9801.877

GWR (Geographically weighted regression) result

Bandwidth and geographic ranges

Bandwidth size: 66.000000



Coordinate	Min	Max	Range
X-coord	808794.569761	887280.238900	78485.669139
Y-coord	1664834.788730	1721720.160000	56885.371270

Diagnostic information

Residual sum of squares: 2.523233

Effective number of parameters (model: trace(S)): 190.816606

Effective number of parameters (variance: trace(S'S)): 117.556990

Degree of freedom (model: n - trace(S)): 2322.183394

Degree of freedom (residual: n - 2trace(S) + trace(S'S)): 2248.923778

ML based sigma estimate: 0.031687

Unbiased sigma estimate: 0.033496

-2 log-likelihood: -10217.391684

Classic AIC: -9833.758472

AICc: -9801.877003

BIC/MDL: -8715.614868

CV: 0.001267

R square: 0.996300

Adjusted R square: 0.995865



<< Geographically varying (Local) coefficients >>

Estimates of varying coefficients have been saved in the following file.

Listwise output file: C:\Users\WATCHARINPHOEMPHON\OneDrive - Suranaree University of Technology\WatcharinP\01 Thesis\Thesis\GWR4\model\model1_listwise.csv

Summary statistics for varying (Local) coefficients

Variable	Mean	STD	
Intercept	-23.170691	15.083501	
LD	0.137691	1.002525	
NDVI	-0.002584	0.109274	
RF	0.004080	0.007185	
GP	-0.013286	0.031139	
CC1	1.094047	0.393069	
TDS1	8.697749	8.789318	
SI1	-0.003182	0.008078	
DD1	0.003086	0.012397	
Lf1	-0.001870	0.003895	
STC1	0.000076	0.000154	
LULC1	-0.000165	0.005207	
Geol1	0.000038	0.000125	

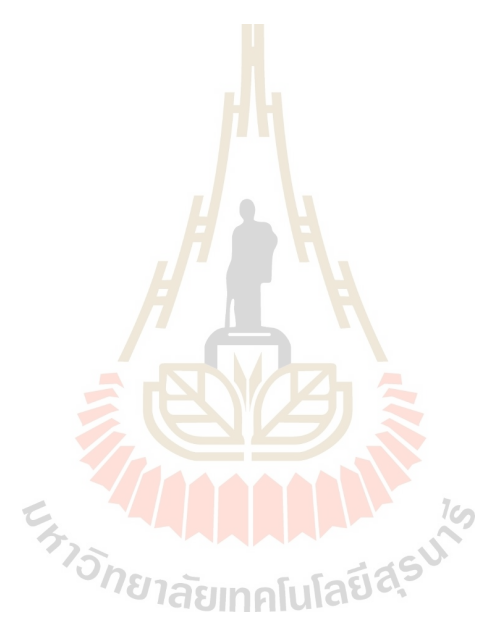
Variable	Min	Max	Range
Intercept	-59.222814	44.050705	103.273519
LD	-6.487336	1.948503	8.435839
NDVI	-0.332169	0.348690	0.680859
RF	-0.016654	0.035667	0.052320
GP	-0.202453	0.097654	0.300107
CC1	0.069137	1.893547	1.824410
TDS1	-27.415645	29.465958	56.881602
SI1	-0.023830	0.023884	0.047714
DD1	-0.028538	0.039837	0.068375
Lf1	-0.013114	0.009736	0.022850



STC1	-0.000272	0.000663	0.000935
LULC1	-0.018065	0.012576	0.030641
Geol1	-0.000311	0.000341	0.000652

Variable	Lwr Quartile	Median	Upr Quartile
Intercept	-32.299162	-23.644474	-11.777010
LD	-0.220619	0.340694	0.676377
NDVI	-0.062914	-0.009172	0.041435
RF	0.000088	0.003162	0.006744
GP	-0.021798	-0.011484	-0.000755
CC1	0.815744	1.172356	1.416128
TDS1	2.086056	9.422662	14.363044
SI1	-0.007567	-0.001739	0.001304
DD1	-0.005435	-0.000831	0.010716
Lf1	-0.004613	-0.000924	0.000740
STC1	-0.000043	0.000046	0.000167
LULC1	-0.001601	0.000351	0.002919
Geol1	-0.000025	0.000028	0.000121

Variable	Interquartile R	Robust STD
Intercept	20.522151	15.212862
LD	0.896996	0.664934
NDVI	0.104349	0.077353
RF	0.006656	0.004934
GP	0.021044	0.015599
CC1	0.600384	0.445059

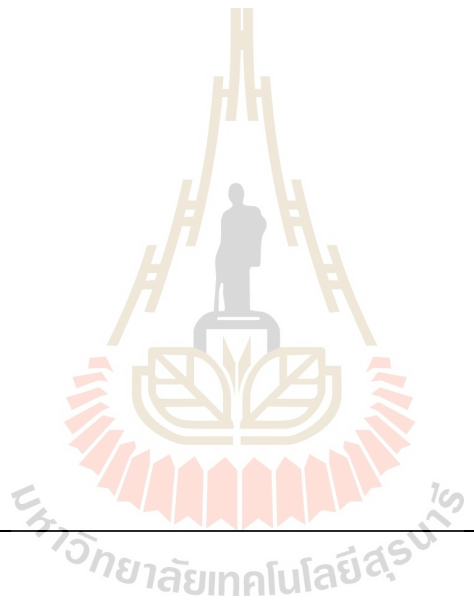


TDS1	12.276988	9.100807
Sl1	0.008872	0.006576
DD1	0.016152	0.011973
Lf1	0.005353	0.003968
STC1	0.000210	0.000156
LULC1	0.004520	0.003350
Geol1	0.000147	0.000109

(Note: Robust STD is given by (interquartile range / 1.349))

GWR ANOVA Table

Source	SS	DF	MS	F
Global Residuals	16.555	2500.000		
GWR Improvement	14.032	251.076		0.056
GWR Residuals	2.523	2248.924	0.001	49.812625



BIOGRAPHY

Mr. Watcharin Phoemphon was born on February 8, 1997, in Surin Province, Thailand, is a dedicated and passionate individual in the field of geotechnology. He completed his bachelor's degree in Petroleum Engineering and Geotechnology from the School of Geotechnology at the Institute of Engineering, Suranaree University of Technology in 2019. Driven by a strong desire for knowledge and academic growth, He pursued further studies and obtained a master's degree in Civil, Transportation, and Geo-resources Engineering from the same institution, further enhancing his expertise in the field.

