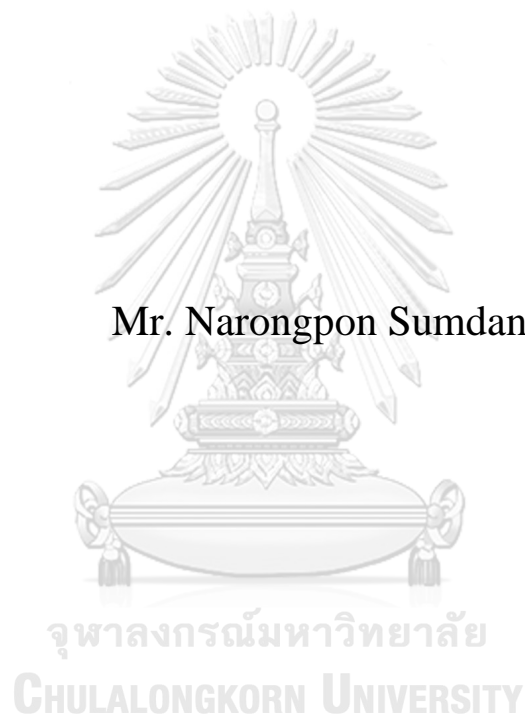


**PREDICTION OF ARSENIC CONTAMINATION IN
RAYONG GROUNDWATER BASIN USING MACHINE
LEARNING BASED APPROACHES**

Mr. Narongpon Sumdang



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Hazardous Substance and
Environmental Management
Inter-Department of Environmental Management
GRADUATE SCHOOL
Chulalongkorn University
Academic Year 2020
Copyright of Chulalongkorn University

การทำนายการปนเปื้อนสารหนูในชั้นน้ำบาดาลระยอง โดยประยุกต์ใช้การเรียนรู้ของเครื่อง



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาการจัดการสารอันตรายและสิ่งแวดล้อม สหสาขาวิชาการจัดการสิ่งแวดล้อม

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2563

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Thesis Title	PREDICTION OF ARSENIC CONTAMINATION IN RAYONG GROUNDWATER BASIN USING MACHINE LEARNING BASED APPROACHES
By	Mr. Narongpon Sumdang
Field of Study	Hazardous Substance and Environmental Management
Thesis Advisor	Associate Professor SRILERT CHOTPANTARAT, Ph.D.
Thesis Co Advisor	Associate Professor Kyung Hwa Cho, Ph.D.

Accepted by the GRADUATE SCHOOL, Chulalongkorn University in
Partial Fulfillment of the Requirement for the Master of Science

..... Dean of the GRADUATE
SCHOOL
(Associate Professor THUMNOON NHUJAK, Ph.D.)

THESIS COMMITTEE

..... Chairman
(Associate Professor EKAWAN LUEPROMCHAI,
Ph.D.)

..... Thesis Advisor
(Associate Professor SRILERT CHOTPANTARAT,
Ph.D.)

..... Thesis Co-Advisor
(Associate Professor Kyung Hwa Cho, Ph.D.)

..... Examiner
(Assistant Professor PENRADEE CHANPIWAT, Ph.D.)

..... External Examiner
(Assistant Professor Pensiri Akkajit, Ph.D.)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ณรงค์พันธ์ สำแดง : การทำนายการปนเปื้อนสารหนูในชั้นน้ำบาดาลระยอง โดยประยุกต์ใช้การเรียนรู้ของเครื่อง.
 (PREDICTION OF ARSENIC CONTAMINATION IN RAYONG
 GROUNDWATER BASIN USING MACHINE LEARNING BASED
 APPROACHES) อ.ที่ปรึกษาหลัก : รศ. ดร.ศรีเลิศ โชติพันธรัตน์, อ.ที่ปรึกษาร่วม : รศ. ดร.หญิง
 หวา โข

งานวิจัยฉบับนี้จัดทำขึ้นเพื่อระบุพื้นที่เสี่ยงต่อการปนเปื้อนจากสารหนูในแอ่งบาดาลชายฝั่งระยอง โดยใช้อัลกอริทึมสามประเภทประกอบด้วย Support Vector Machine (SVM), Random Forest (RF) และ Artificial Neuron Network (ANN) งานวิจัยประกอบด้วยสามส่วนในการศึกษา คือ 1) การเลือกตัวแปรที่เหมาะสมสำหรับสร้างแบบจำลอง 2) การเลือกอัลกอริทึมแบบจำลองที่เหมาะสมและ 3) การสร้างแบบจำลองสำหรับแสดงแผนที่ความเสี่ยงเพื่อทำการสร้างแบบจำลองมีประสิทธิภาพมากขึ้นตัวแปรที่นำมาสร้างแบบจำลองผ่านการคัดเลือก โดยดูจากความสัมพันธ์ระหว่างตัวแปรทางอุทกเคมีและสารหนู ซึ่งสามารถอธิบายกลไกการปนเปื้อนของสารหนูในน้ำบาดาลได้และเนื่องจากข้อมูลส่วนใหญ่จัดเป็นประเภทโมโนโทนิก (Monotonic) และไม่มีการจัดเรียงตัวของข้อมูลในรูปแบบการกระจายตัวแบบปกติ ดังนั้นจำเป็นต้องมีการใช้ Spearman's correlation เพื่อเป็นเครื่องมือคัดกรองและนำเอาตัวแปรที่ไม่มีความจำเป็นออกไปจากชุดข้อมูล โดยผลลัพธ์แสดงให้เห็นว่าตัวแปรที่ได้รับการคัดสรรสามารถนำมาใช้ในการอธิบายการปนเปื้อนของสารหนูในน้ำบาดาลได้ โดยกลไกหลักๆที่มีอิทธิพลต่อสารหนูในพื้นที่คือสภาพแวดล้อมแบบรีดิเวจจึงสรุปได้ว่าการหาความสัมพันธ์ตัวแปรต่างๆที่ส่งผลกระทบต่อความเข้มข้นของสารหนู โดยใช้ Spearman's correlation นั้นจะสามารถใช้เพื่อคัดกรองตัวแปรที่มีความสำคัญต่อสารหนู ก่อนนำไปใช้ในการสร้างแบบจำลองได้ และเพื่อคัดเลือกอัลกอริทึมของแบบจำลองที่เหมาะสมสำหรับการนำไปสร้างแผนที่ความเสี่ยงต่อการปนเปื้อนสารหนู จึงจำเป็นต้องมีการประเมินแบบจำลองโดยใช้ ค่าประสิทธิภาพในการทำและค่าความไม่แน่นอน ในการประเมินแบบจำลองที่เหมาะสม ผลการวิจัยระบุว่าค่าประสิทธิภาพในการทำของแบบจำลองของ RF นั้นมีประสิทธิภาพดีกว่าของ SVM และ ANN นอกจากนี้ค่าความไม่แน่นอนของแต่ละแบบจำลองสรุปได้ว่าแบบจำลอง RF นั้นมีค่าความไม่แน่นอนน้อยที่สุดเมื่อเทียบกับแบบจำลองอื่น นอกจากนี้เพื่อเป็นการยืนยันผลจากการประเมินประสิทธิภาพของแบบจำลอง ค่าการทำนายของสารหนูในแต่ละแบบจำลองถูกนำมาตรวจสอบโดยใช้ปริมาณสารหนูที่ได้จากการสำรวจในภาคสนามของการศึกษานี้ ผลลัพธ์ที่ได้ยืนยันได้ว่า RF เป็นแบบจำลองที่มีประสิทธิภาพดีที่สุดเทียบกับอีกสองแบบจำลอง ผลลัพธ์จากแผนที่แสดงความเสี่ยงการปนเปื้อนของสารหนูโดยใช้แบบจำลอง RF ระบุได้ว่า ประชากรในพื้นที่บริเวณทางตอนเหนือของแอ่งบาดาลระยองนั้นมีความเสี่ยงที่จะได้รับผลกระทบจากการปนเปื้อนของสารหนูจากการใช้น้ำบาดาลจากบ่อระดับลึก ในทางกลับกันสำหรับประชากรในพื้นที่ทางตอนใต้ของแอ่งน้ำบาดาลระยองอาจได้รับผลกระทบจากการปนเปื้อนของสารหนูจากการใช้น้ำบาดาลจากบ่อระดับตื้น ซึ่งจากการศึกษาจากงานต่างๆ ในพื้นที่สาเหตูอาจมาจากพื้นที่บ่อเก็บขยะและพื้นที่อุตสาหกรรมภายในแอ่งน้ำบาดาล สุดท้ายนี้ผลลัพธ์จากงานวิจัยฉบับนี้สามารถนำไปใช้ประโยชน์ได้โดยหน่วยงานรัฐบาล หรือองค์กรภาคเอกชนเพื่อช่วยในการจัดการสิ่งแวดล้อม และสามารถนำไปต่อยอดในระดับสากลได้

CHULALONGKORN UNIVERSITY

สาขาวิชา	การจัดการสารสนเทศรายและสิ่งแวดล้อม	ลายมือชื่อนิสิต
ปีการศึกษา	2563	ลายมือชื่อ อ.ที่ปรึกษาหลัก
		ลายมือชื่อ อ.ที่ปรึกษาร่วม

6187520120 : MAJOR HAZARDOUS SUBSTANCE AND ENVIRONMENTAL
MANAGEMENT

KEYWORD: Machine Learning, Groundwater contamination, Spatial probability model,
Groundwater risk assessment

Narongpon Sumdang : PREDICTION OF ARSENIC CONTAMINATION IN
RAYONG GROUNDWATER BASIN USING MACHINE LEARNING BASED
APPROACHES . Advisor: Assoc. Prof. SRILERT CHOTPANTARAT, Ph.D.
Co-advisor: Assoc. Prof. Kyung Hwa Cho, Ph.D.

The present study used three algorithms consisting of Support Vector Machine (SVM), Random Forest (RF), and Artificial Neuron Network (ANN) to locate risk area of arsenic (As)contamination in Rayong coastal aquifers, Thailand. There were three parts in this study consisting of 1) selecting the proper parameters 2) selecting the appropriate model, and 3) constructing the risk map of As. To perform models efficiently, the parameters used to generate the models have to be selected based on the correlation of each hydrochemical parameter with As concentration, which could explain the mechanisms of As release in groundwater. Due to major parameters in the dataset were monotonic and not presented by the normal distribution, thus, Spearman's correlation was conducted to screen the suitable parameters. The results showed that parameters correlated with As mostly supported by the mechanism of As release in groundwater, which is dominantly controlled by the reducing condition. Spearman's correlation technique would help to select the crucial parameters in the further modeling process. To select an appropriate model to generate the risk map, the model's performance has to be measured by the prediction performance and uncertainty of each model. The prediction performance indicated that the RF algorithm has the highest performance as compared to those in SVM and ANN. In addition, the uncertainty of each model confirmed that the RF algorithm has the lowest uncertainty. Moreover, to confirm the performance of the models, the actual As concentration in field data were used to validate the prediction result of each model. The result, also confirms that the RF model was the best performance model compared with the other two models. Therefore, the RF was the appropriate algorithm that can generate the probability map to locate the areas of As contamination in groundwater. The result of the risk map obtained from the RF model indicated that the deep aquifer (granite aquifer, Gr), in the northern part of the Rayong basin has a higher risk for people who have used groundwater to expose to As. In contrast, the shallow aquifer revealed that the southern part of the Rayong basin has a higher risk for people who use groundwater, which is also supported by the location of the landfill and industrial estates in the Mueang District. The outcome of this study can be useful for the government and other organizations for groundwater resource management and environmental protection. Furthermore, the novelty of this research can be used to further study other groundwater aquifers contaminated with As in the world.

Field of Study:	Hazardous Substance and Environmental Management	Student's Signature
Academic Year:	2020	Advisor's Signature
		Co-advisor's Signature

ACKNOWLEDGEMENTS

The author thankfully acknowledges the support of the International Postgraduate Program in Hazardous Substance and Environmental Management, Graduate School, Chulalongkorn University for their invaluable supports in terms of facilities and scientific equipment. I would like to thank my advisor Associate Professor Srilert Chotpantarat, Ph.D., and co-advisor Associate Professor Kyung Hwa Cho, Ph.D. for giving many crucial suggestions for this study. I also would like to thank Mr. Pongsathorn Thunyawatcharakul and Ms. Satika Boonkaewwan for support and help in the samples field collected. This research was supported by the scholarship from the Graduate School, Chulalongkorn University to commemorate the 72nd anniversary of his Majesty King Bhumibol Aduladej, the 90th Anniversary Chulalongkorn University Fund (Ratchadaphiseksomphot Endowment Fund), and National Research Council of Thailand (NRCT): NRCT5-RSA63001-06.

Narongpon Sumdang

TABLE OF CONTENTS

	Page
.....	iii
ABSTRACT (THAI)	iii
.....	iv
ABSTRACT (ENGLISH).....	iv
ACKNOWLEDGEMENTS.....	v
TABLE OF CONTENTS.....	vi
LIST OF TABLES.....	ix
LIST OF FIGURES	xi
Chapter 1 Introduction.....	1
1.1 Introduction.....	1
1.2 Objectives	4
1.3 Hypothesis	5
1.4 Scope of the study.....	5
Chapter 2 Literature Review.....	6
2.1 Study area	6
2.1.1 Geology characteristics	7
2.1.2 Soil characteristics.....	7
2.1.3 Hydrogeologically characteristics	8
2.1.4 Land use	9
2.1.5 Meteorological.....	10
2.2 Arsenic	11
2.2.1 Arsenic toxicity	11
2.2.2 Arsenic sources and behavior.....	12
2.2.2.1 Arsenic source	12
2.2.2.2Mechanisms of Arsenic Release to Groundwater	13

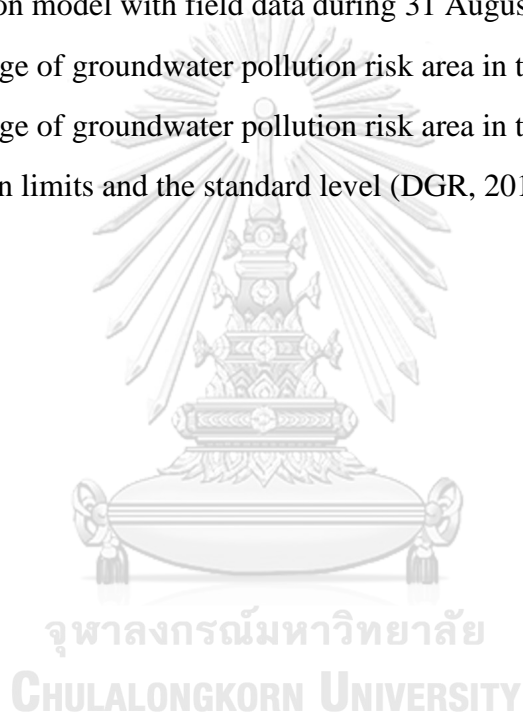
2.2.2.3 Redox condition	14
2.3 Machine learning	16
2.3.1 RF (Random forest).....	16
2.3.2 SVM (Support vector machine)	17
2.3.3 ANN (Artificial neural network).....	18
2.3.4 Summarize all algorithms.....	19
2.4 Risk and probability occurred maps	20
2.4.1 Probability map	20
2.4.2 Risk map.....	21
2.5 Spearman's correlation.....	21
2.6 IDW	23
2.7 Quartile regression	24
Chapter 3 Methodology	27
3.1 Framework of research	27
3.2 Data preparation and field-collected data	28
3.2.1 Data preparation	28
3.2.1.1 Physical data.....	29
3.2.1.2 Hydrochemical data.....	29
3.2.2 Field collected data.....	30
3.2.2.1 Data measuring in the field.....	31
3.2.2.2 Total arsenic analysis	32
3.3 Arsenic contamination probability map.....	32
3.3.1 Framework of the arsenic contamination probability map.....	32
3.3.2 Spearman's correlation.....	33
3.3.3 IDW interpolation.....	33
3.3.4 Model probability	34
3.4 Probability map validation.....	35
3.4.1 Evaluating prediction performance	35

3.4.2 Evaluate models uncertainty	36
3.4.3 Validation with field data	36
3.5 Groundwater pollution risk map	36
3.5.1 Probability cutoff value	36
3.5.2 Risk map for As.....	37
Chapter 4 Results	38
4.1 Groundwater wells data	38
4.2 Field data	42
4.3 Spearman's correlation	44
4.4 Probability map.....	50
4.5 Performance of models	56
4.5.1 Prediction performance	56
4.5.2 Models uncertainty	57
4.5.3 Validation with field data	58
4.6 Groundwater pollution risk map.....	59
Chapter 5 Discussions	61
5.1 Parameters associated with As mechanisms in groundwater.....	61
5.2 Evaluating prediction performance.....	64
5.3 Uncertainty assessment.....	66
5.4 Risk map assessment	67
Chapter 6 Conclusions and recommendations	69
6.1 Conclusion	69
6.2 Recommendation	70
APPENDIX.....	72
REFERENCES	82
VITA.....	90

LIST OF TABLES

	Page
Table 1 Soil types classification based on soil texture and proportion of soil types in the study area	8
Table 2 Summarize the advantage and disadvantage of RF, SVM, ANN	20
Table 3 The occurrence of the event based on probability level (Garvey 2001)	21
Table 4 Source of physical data	29
Table 5 Groundwater sampling aquifer	31
Table 6 Description statistics in the deep aquifer	40
Table 7 Description statistics in the shallow aquifer	41
Table 8 Arsenic concentrations in groundwater samples collected from Gr,Qa and Qcl aquifers during 31 August – 1 September, 2019	43
Table 9 Descriptive statistics of in-situ measured parameters and total As concentration in groundwater during 31 August – 1 September 2019	44
Table 10 Descriptive statistics of total As concentrations in Gr, Qa, and Qcl aquifers during 31 August – 1 September 2019	44
Table 11 Spearman’s correlation of hydrochemical parameters in the deep aquifer..	46
Table 12 Spearman’s correlation of hydrochemical with shallow aquifer	47
Table 13 Spearman’s correlation of physical parameters in the deep aquifer	48
Table 14 Spearman’s correlation of physical parameters in the shallow aquifer	49
Table 15 Spearman’s correlation with physical and hydrochemical parameters in both deep and shallow aquifers	50
Table 16 Percentage probability of each district derived from SVM model in the deep aquifer	53
Table 17 Percentage probability of each district derived from RF model in the deep aquifer	53
Table 18 Percentage probability of each district derived from ANN model in the deep aquifer	53
Table 19 Percentage probability of each district derived from SVM model in the shallow aquifer.....	55

Table 20 Percentage probability of each district derived from RF model in the shallow aquifer.....	55
Table 21 Percentage probability of each district derived from ANN model in the shallow aquifer.....	55
Table 22 Performances of SVM, RF, and ANN in the deep aquifer	56
Table 23 Performances of SVM, RF, and ANN in the shallow aquifer	57
Table 24 Uncertainty analysis of SVM, RF, and ANN of the deep aquifer	58
Table 25 Uncertainty analysis of SVM, RF, and ANN of the shallow aquifer	58
Table 26 Validation model with field data during 31 August – 1 September 2019 ...	59
Table 27 Percentage of groundwater pollution risk area in the deep aquifer	60
Table 28 Percentage of groundwater pollution risk area in the shallow aquifer	60
Table 29 Detection limits and the standard level (DGR, 2017).....	81



LIST OF FIGURES

	Page
Figure 1 Study area	6
Figure 2 Geologic map.....	7
Figure 3 Soil map.....	8
Figure 4 Hydrological map	9
Figure 5 Land use map.....	10
Figure 6 Natural geochemical processes that release As into groundwater (Herath, Vithanage et al. 2016)	14
Figure 7 Illustration of the mechanisms of mobilization and redox transformation of As in aquifer sediments (Herath, Vithanage et al. 2016).....	15
Figure 8 DNN diagram	19
Figure 9 monotonic and non-monotonic relationships diagrams.....	22
Figure 10 Inverse Distance Weight (IDW) Interpolation	24
Figure 11 Predictions for quantiles	25
Figure 12 Framework of the present study	28
Figure 13 Groundwater sampling map	30
Figure 14 Framework of constructing the probability map of As	33
Figure 15 Framework of constructing the probability model of As	35
Figure 16 The probability map in the deep aquifer derived from a) RF, b) SVM, c) ANN and the probability map in the shallow aquifer derived from d) RF, e) SVM and f) ANN	52
Figure 17 Taylor's diagram for a) the deep aquifer and b) the shallow aquifer	57
Figure 18 Groundwater pollution risk map of As contamination in a) the deep aquifer and b) the shallow aquifer.....	68
Figure 19 J1 groundwater well	72
Figure 20 J2 groundwater well	72
Figure 21 J3 groundwater well.....	73
Figure 22 J4 groundwater well	73

Figure 23 J6 groundwater well	73
Figure 24 J7 groundwater well.....	74
Figure 25 J8 groundwater well.....	74
Figure 26 J9 groundwater well.....	74
Figure 27 J10 groundwater well	74
Figure 28 J11 groundwater well	75
Figure 29 J13 groundwater well	75
Figure 30 J14 groundwater well	75
Figure 31 G1 groundwater well	76
Figure 32 G2 groundwater well	76
Figure 33 G3 groundwater well	76
Figure 34 G4 groundwater well	77
Figure 35 G7 groundwater well	77
Figure 36 G8 groundwater well	77
Figure 37 G9 groundwater well	78
Figure 38 G10 groundwater well	78
Figure 39 G11 groundwater well	78
Figure 40 G12 groundwater well	79
Figure 41 G13 groundwater well	79
Figure 42 G14 groundwater well	79
Figure 43 G15 groundwater well	80
Figure 44 G17 groundwater well	80
Figure 45 G19 groundwater well	80

Chapter 1 Introduction

1.1 Introduction

Groundwater is the most valuable resource of drinking water especially, in South East Asia and developing countries (Winkel, Berg et al. 2008, Cho, Sthiannopkao et al. 2011). However, arsenic (As) contamination in groundwater is a major problem in many countries and regions because As is one of the major carcinogenic elements, which mostly presents in a toxic form as an inorganic species in natural water systems including groundwater (WHO 2018). Arsenic can be considered as a toxic element to humans in several forms, especially arsenite (As(III)), arsenate (As(V)), and organic As compounds. A lethal dose in humans is 1.5 mg/kg of body weight (WHO 2018). The acute intoxication symptoms include vomiting, abdominal pain, muscular pain, diarrhea, and weakness, with flushing of the skin, and chronic intoxication symptoms, including dermal lesions such as hypopigmentation and hyperpigmentation, skin cancer, peripheral neuropathy, lung cancers, bladder and peripheral vascular disease (WHO 2018). Due to the As toxicity, the monitoring strategies methods to observe As contamination need to be improved to quantify and predict As concentrations in groundwater. The proper method might be used to provide necessary information for better assessment and manage public health (Cho, Sthiannopkao et al. 2011). However, due to lacking equipment and human resource, dealing with local As contamination problems in regional areas remains problematic (DGR 2017). Therefore, modeling approaches for As concentrations using geological and on-site measurement data can be an alternative to characterize and measure the As contamination potential, as well as to provide

predictive information for better public health management (Winkel, Berg et al. 2008, Cho, Sthiannopkao et al. 2011, Bindal and Singh 2019). The prediction information in terms of the probability and risk map, that generated from the model, will help to support groundwater management and to plan to install a monitoring well system for the local government agencies. The probability map usually uses in many types of environmental science studies to inform the general information about a percent chance to encounter some study's element. On the other hand, the risk map can define an area in which it will encounter the study's element (Sajedi Hosseini, Malekian et al. 2018). The prediction information in terms of the probability and risk map, that generated from the model, can help to support groundwater management and to plan to install groundwater monitoring wells system for the local government agencies. The probability map usually uses in many types of environmental science studies to inform the general information about a percent chance to encounter some study's focused elements. On the other hand, the risk map can define an area in which it will encounter the study's focused element (Sajedi Hosseini, Malekian et al. 2018).

Currently, machine learning (ML) has been applied in several fields in environmental scientist's study. The MLs power comes from their powerful nonlinear modeling capability, which usually uses for assessment in environmental science aspects. To study groundwater contamination, machine learning (ML) has been applied for the prediction of several risk assessments in groundwater resources (Winkel, Berg et al. 2008, Sajedi Hosseini, Malekian et al. 2018, Bindal and Singh 2019, Podgorski, Wu et al. 2020). There are several algorithms such as random forest (RF), support vector machine (SVM), and artificial neural network (ANN). Random Forest (RF) classifier for interpretations of the land cover shows that this algorithm is

very fast, and has satisfactory results with the limited data set (Havryliuk, Korol et al. 2018, Podgorski, Wu et al. 2020). The algorithm could provide better detection of the variability in apartment values and predicts them more effectively than multiple regression (Marjan, Kilibarda et al. 2018). Support vector machine (SVM) was used to determine the risk of nitrate contamination in groundwater (Sajedi Hosseini, Malekian et al. 2018), and it is good to classify data and trends to be resistant to the overfitting problem. An artificial neural network (ANN) has been applied to evaluate the As contamination in groundwater. Furthermore, combine with the PCA technique, ANN algorithms provide a significant result to determine the As contamination in groundwater in Cambodia, Laos, and Thailand areas (Cho, Sthiannopkao et al. 2011). All three algorithms are suitable to handle a large amount of data to generate a model, which have many hydrochemical and physical variables to analyze and predict the groundwater contamination.

Rayong groundwater basin is located between Chonburi and Rayong provinces and contact with the gulf of Thailand coastal. Groundwater in the Rayong groundwater basin has been found an As contamination problem in groundwater (Kerdthep, Tongyonk et al. 2009, Boonkhao, Phanprasit et al. 2017, Pipattanajaroenkul, Sonthiphand et al. 2018, Boonkaewwan, Sonthiphand et al. 2020). Besides, the study area is the part of the Eastern Economic Corridor (EEC), where is the project for the economic development of Thailand's Eastern Seaboard and the government has been launching measures to support the economic growth in EEC. In the future, the EEC can grow into a new trade center in Asia (Ootsahkarn 2018). Thus, the demand for the groundwater resource in this region will be dramatically

increased; thus, the groundwater quality of groundwater has to be considered as the priority before pumping groundwater to supply for each sector.

However, the parameters that use in the implementation of ML approaches of groundwater pollution risk usually used only physical parameters (Winkel, Berg et al. 2008, Sajedi Hosseini, Malekian et al. 2018). It is limited understanding in integrating physical and hydrochemical parameters of groundwater to assess contamination in groundwater, particularly As contamination in the urbanized coastal aquifer (Zubair, Begum et al. 2015). As mentioned, to fulfill this research gap, this study attempted to apply ML algorithms, including RF, SVM, and ANN to investigate and predict the As contamination in Rayong coastal basins. Thus, the main objectives of the current study are: (i) to evaluate the machine learning algorithms that suitable to predict the As contamination in groundwater, (ii) to investigate the environmental factors (e.g., hydrochemical characteristics, soil types, land use/landcover) influencing on As contamination in groundwater. This study provided the appreciative predictive information for better public health management and groundwater quality control to support the EEC project in this area.

1.2 Objectives

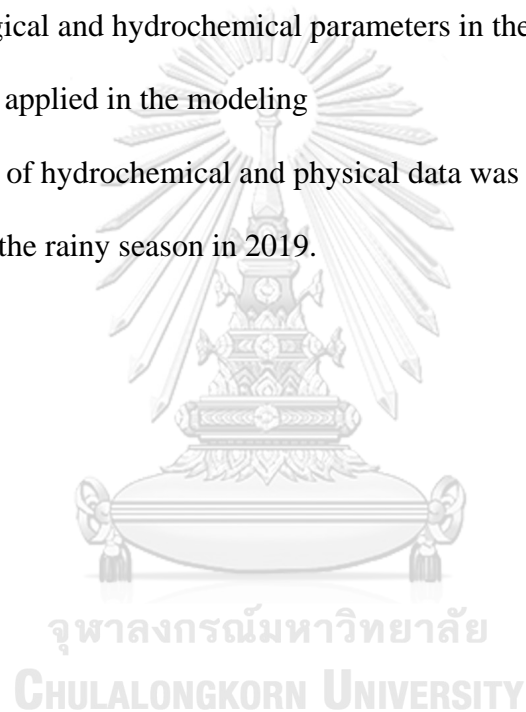
1. To evaluate the machine learning algorithms suitable to assess the As contamination in groundwater.
2. To investigate the environmental factors influencing an As contamination in groundwater.

1.3 Hypothesis

1. The machine learning algorithms can provide risk areas of As contamination in groundwater.
2. The hydrochemical parameters mainly are the influencing environmental factors on As contamination in groundwater.

1.4 Scope of the study

1. The geological and hydrochemical parameters in the Rayong groundwater basin were applied in the modeling
2. The period of hydrochemical and physical data was from the dry season in 2012 until the rainy season in 2019.



Chapter 2 Literature Review

2.1 Study area

The study area is located between Chonburi and Rayong provinces, combined with 9 sub-district, Nong yai, Bang lamung, Sathip, Si racha, Ban chang, Mueang Rayong, Pluak deang and King amphoe nikhom phatthana, covering Rayong groundwater basin, which has faced the problem of As contamination in groundwater. The study area has a mountain in the west area and a large plain in the middle part of the basin (DGR 2012, DGR 2017), covering areas approximately 2,236 km², as shown in Figure 1.

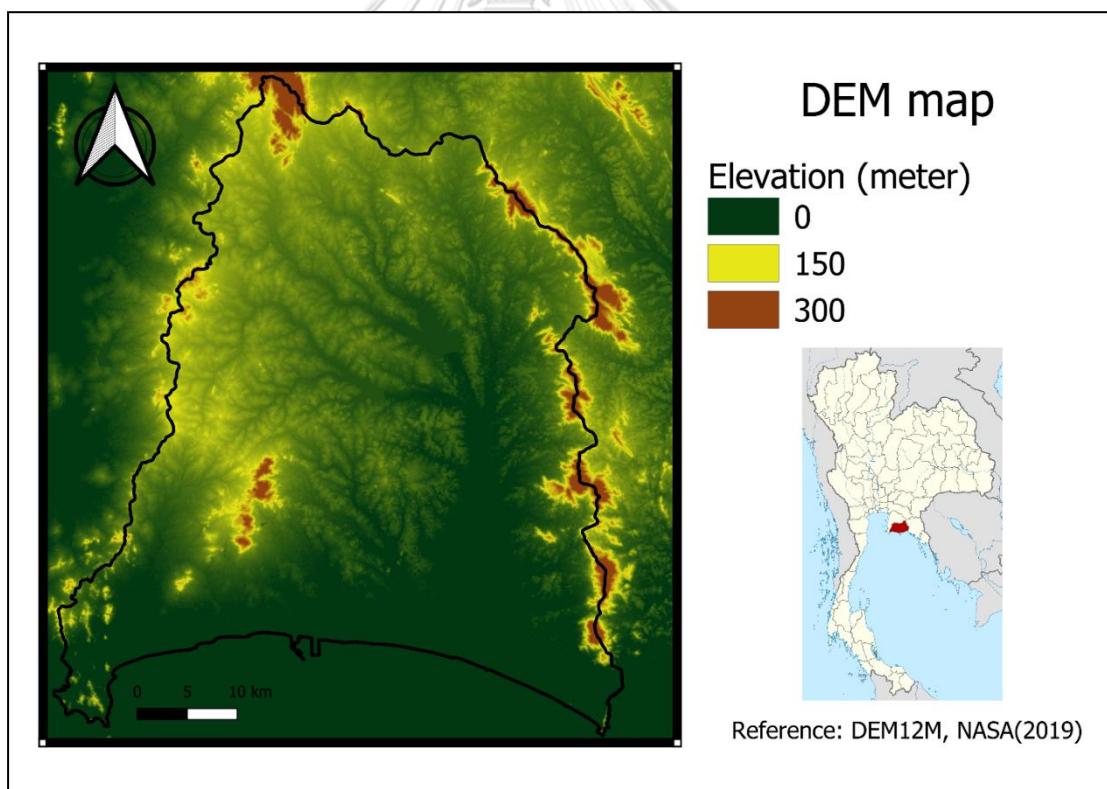


Figure 1 Study area

2.1.1 Geology characteristics

Biotite granite in Carboniferous age rocks is located in the Northern and Western part of Rayong groundwater basin as shown in Figure 2. Alluvia and Terrace's deposits are in the middle part of the Rayong groundwater basin (DMR 2007) as shown in Figure 2. There is a source of the organic component, affecting the reducing condition in the groundwater environment.

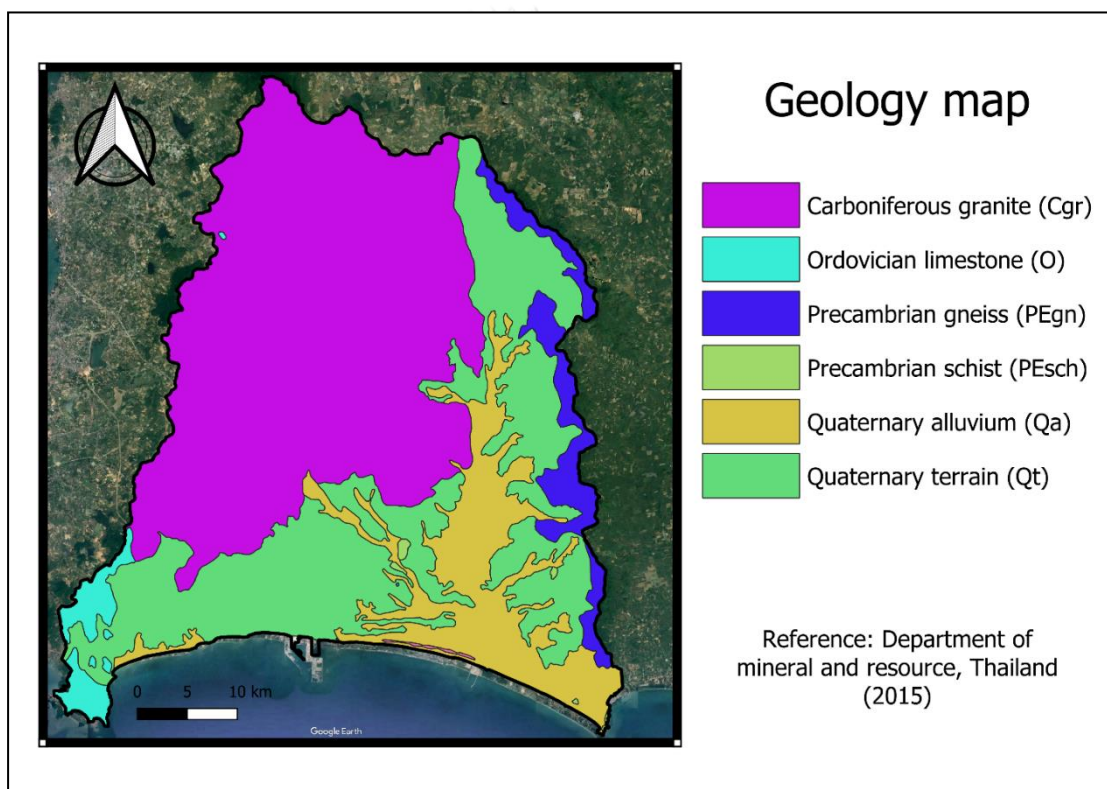


Figure 2 Geologic map

2.1.2 Soil characteristics

The soil type in the study area is derived from the Land Development Department (LDD) (LDD 2016). A soil map was classified by each soil type in the study area based on soil textures. The new classification can be separated into 7 types as shown in Table 1 and Figure 3.

Table 1 Soil types classification based on soil texture and proportion of soil types in the study area

Soil texture	Percentage in the study area (%)
Gravel, Gravel loam	6.75
Gravel loam, Sand	0.70
Sand, Sandy Loam	31.52
Sandy loam, Silt Loam	35.03
Silt loam, Silt	2.53
Silt, Clay loam	2.47
Clay	21.01

Note: rock land and slope complex are not included in a calculation.

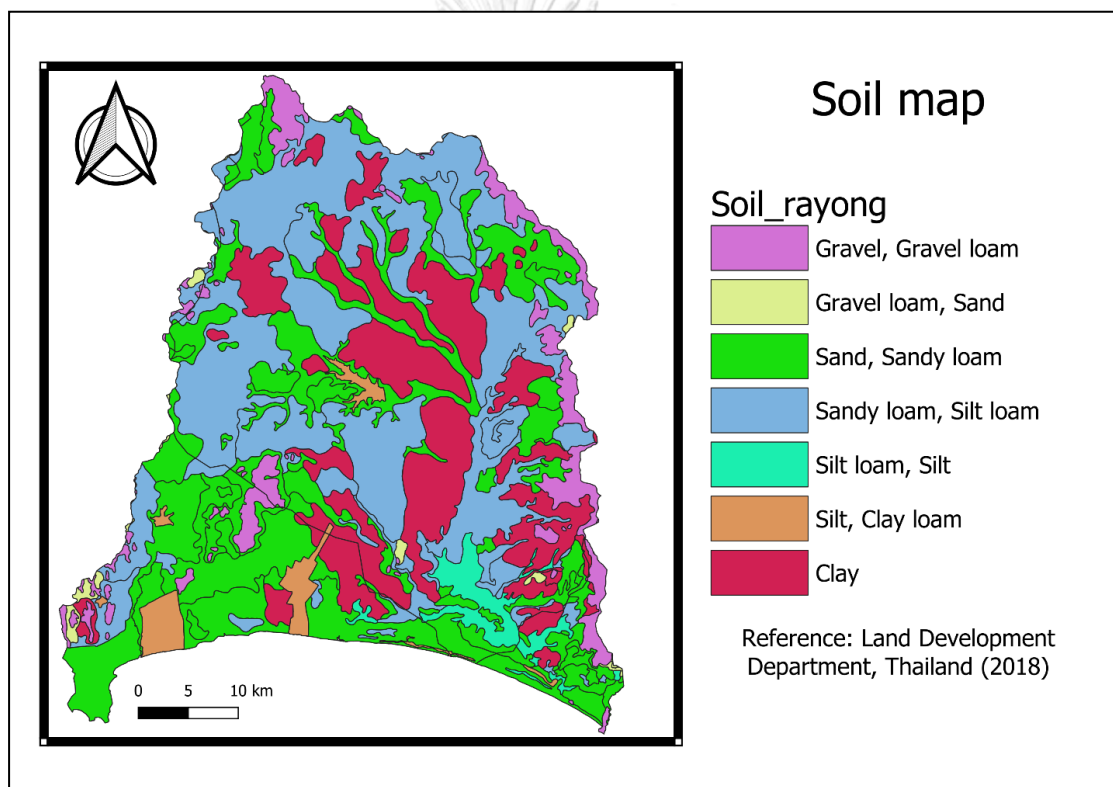


Figure 3 Soil map

2.1.3 Hydrogeologically characteristics

Colluvium sediment (Qc) in the quaternary age; has the largest aquifer area as compared to other hydrologic units. Another main hydrologic unit is alluvium sediment in quaternary age, which locate alongside the main river in the study area. Both aquifer characteristics have rich organic matter, which can provide a reducing

condition in groundwater. The reducing condition in groundwater is suitable for the release of arsenide, which is the most severe form of in As species (Jacks 2017). The other two aquifers are the consolidated aquifers, including granite and carbonate aquifers, which are located in the mountainous areas distributed around the groundwater basin.

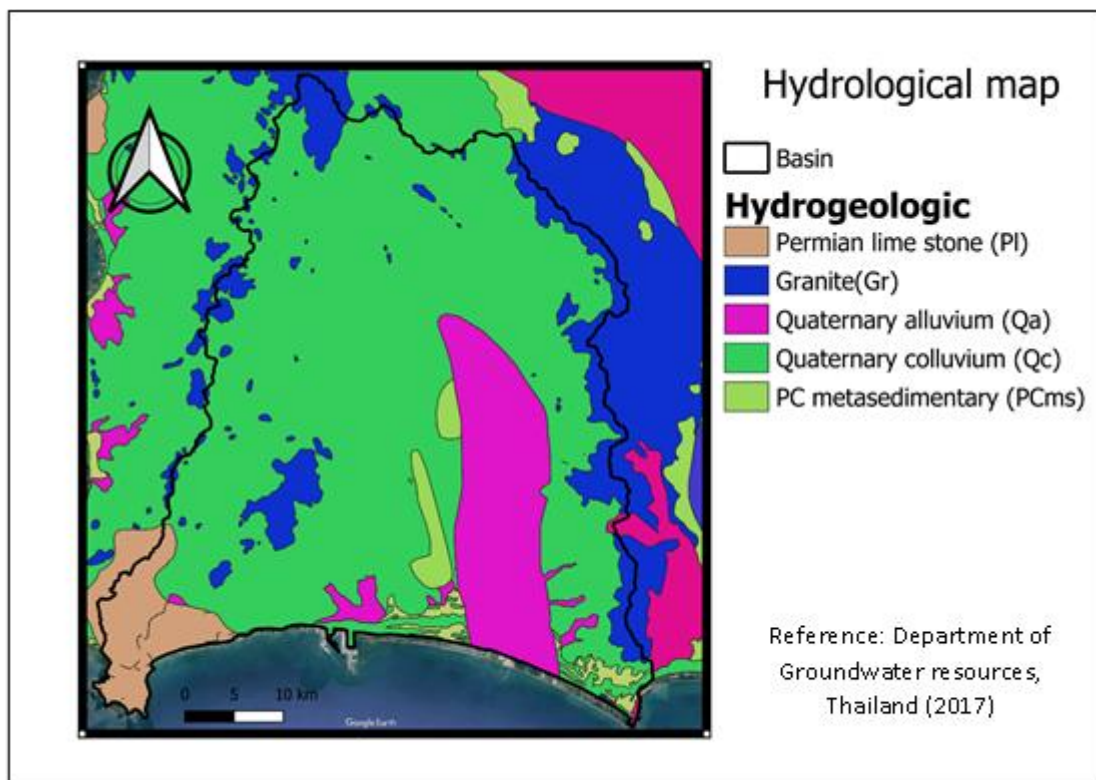


Figure 4 Hydrological map

2.1.4 Land use

In the study area, the majority of land use is dominant by agricultural areas average around 80% in the total area following by Forest (8%), Urban (5%), Mining (4%), Irrigation (2%), and water (>1%) (LDD, 2016). The agricultural area majority by rice fields and cassava. For mining, the majority is sand mining following by granite and limestone mining (DMR 2007).

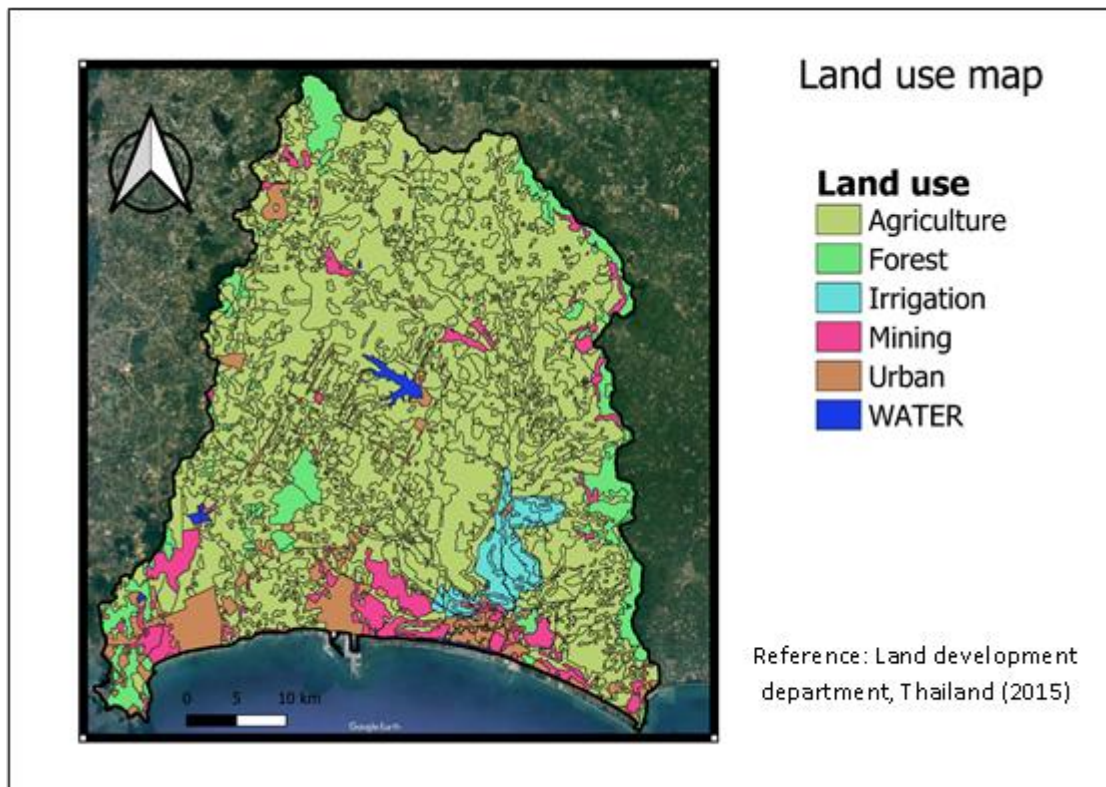


Figure 5 Land use map

2.1.5 Meteorological

Thailand usually experiences dry weather in winter because of the northeast monsoon which is the main cause that controls the climate of this region (TMD 2015). The later period, summer, is characterized by gradually increasing rainfall with thunderstorms. The onset of the southwest monsoon leads to intensive rainfall from mid-May until early October. Rainfall peak is in August or September which some areas are probably flooded. However, dry spells are commonly occurred for 1 to 2 weeks or more from June to early July due to the northward movement of the ITCZ to southern China. According to a general annual rainfall pattern, most areas of the country receive 1,200- 1,600 mm a year (TMD 2015). In Rayong province, Rainfall annual average around 1,500 mm/year, which average rate for rain annually in Thailand (TMD 2015).

2.2 Arsenic

2.2.1 Arsenic toxicity

As is a natural component of the earth's crust, which is widely distributed in the environment throughout water and land. As is very toxic in inorganic form, which mostly in Arsenate and Arsenide form. People are exposed to inorganic As in many several ways such as: through drinking contaminated water, using as food preparation and irrigation of crops, eating contaminated food, and industrial process. Long-term exposure to inorganic As, mainly through drinking-water and food, can lead to chronic symptoms such as Skin lesions and skin cancer. As in form of inorganic is a confirmed as a carcinogen and significant contaminant in drinking-water in worldwide. As can also occur in an organic form. Inorganic As compounds are highly toxic while organic As compounds are less harmful to health.

Acute effects: The symptoms of acute As poisoning include abdominal pain, vomiting, and diarrhea. These are usually followed by tingling, numbness, and muscle cramping, and death (WHO 2018).

Long-term effects: Skin cancer is the first symptom, which usually is observed in long-term exposure with high levels of inorganic As case. Following by pigmentation changes and skin lesions (hyperkeratosis). These might be occurred with a minimum exposure of five years with As poisoning (WHO 2018). As contamination in drinking water is globally and there are several numbers regions where As contamination in drinking water is very significant. In currently circumstance is recognized that at least 140 million people in 50 countries have been consuming

contaminated water, which As is above 10 $\mu\text{g/L}$ that excess the WHO provisional guideline value.

2.2.2 Arsenic sources and behavior

2.2.2.1 Arsenic source

Arsenic is a common element that occurs in the environment. Generally, As can find in igneous and sediment rocks, where it is a higher level in fine-grained sediment and marine sediment (Naidu and Bhattacharya 2009). The arsenic compound in rock can enter As groundwater system mainly through the weathering process (Maity, Kar et al. 2011). As species in groundwater is linked to As occurring in mineral, soil, and water phases. There are many minerals that As rich mineral, including olivenite ($\text{Cu}_2\text{OHAsO}_4$), proustite (Ag_3AsS_3), orpiment (As_2S_3), realgar (As_4S_4), and tennantite ($\text{Cu}_6[\text{Cu}_4(\text{Fe,Zn})_2]\text{As}_4\text{S}_{13}$), cobaltite (CoAsS), enargite (Cu_3AsS_4), arsenolite (As_2O_3), and FeAsS (Francesconi et al., 2002). Soil also consider potential As sources, because As is more highly concentrated in soils than rock (Meharg and Rahman 2003). Generally, the most As toxic form in the soil is usually found in inorganic form. The average concentration in soil of As around 3 to 4 mg/L (Mukherjee, Bhattacharya et al. 2009). Alluvial and organic soil types contained higher As concentration, exceed the standard level (Smith, Naidu et al. 2001). An As source in water is usually found in low concentration. Arsenate As(III) and arsenide As(V), also are the most abundant As species in water. A redox potential, presence of adsorbents, humid substances, pH, dissolved organic matter, and clay minerals are the factors that affected the As in natural water (Bissen and Frimmel 2003).

2.2.2.2 Mechanisms of Arsenic Release to Groundwater

The release mechanisms of As to groundwater is related to aquifer and sediment. Normally, the As concentration is very low and varies around 1-2 mg/L (Taylor and McLennan 1985). However, with four major geochemical processes such as reductive dissolution, alkali desorption, sulfide oxidation, and geothermal processes (Figure 6), additional As concentration can release into groundwater. The reductive dissolution process is a common geochemical process that releases additional As from aquifer sediments into the aqueous system (Bauer and Blodau 2006). Iron hydroxides $\text{Fe}(\text{OH})_3$ are the main component associated with this reaction. It is affected in high pH conditions, the desorption process becomes stronger, leading to high levels of As in groundwaters (Welch, Westjohn et al. 2000). Alkaline desorption usually occurs, where pH is high and low oxygen and leads to be an anaerobic environment. It directly affects the release of As in groundwater by changing the subsurface environment to a reduction condition (Sanjrani et al., 2019). The oxidation of As sulfides is a source of As and. Sulfide oxidation minerals can release As in the aquifer. The Oxidation of As-bearing sulfides is recognized as an important cause of As contamination in groundwater and produce producing of acid drainage, containing toxic inorganic pollution (Nriagu, Bhattacharya et al. 2007). Geothermal is also the main geochemical process that triggers the As release process

in groundwater and mobile it to cold aquifers (Bundschuh and Maity 2015).

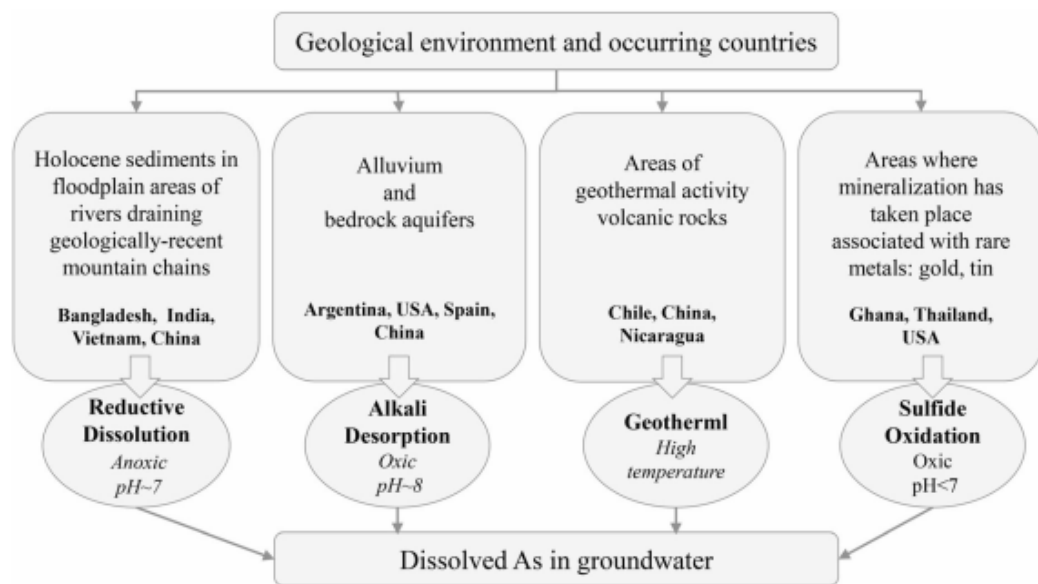


Figure 6 Natural geochemical processes that release As into groundwater (Herath, Vithanage et al. 2016)

2.2.2.3 Redox condition

The redox reaction is the chemical reaction that important for As distribution in the environment including groundwater. Reduction and oxidation reaction can provide different conditions in the groundwater and can affect As mobilizing as shown in Figure 7. Reduction condition usually occurs in less oxygen environment for example in alluvial, deltaic sediments aquifer and also fine-grain sediment aquifer. Thus, created an anaerobic environmental condition. As adsorbed, which play an important role of As mobilizing, is affected by reduction condition. Some anaerobic bacteria and dissolved organic carbon (DOC) also play a crucial role to an As mobilizing in groundwater with a reduction condition by dissolved As absorbents (Cummings, Caccavo et al. 1999). Therefore, The reducing reactions under anaerobic conditions can cause a result of As concentrations in groundwaters going higher

(Ahmed, Bhattacharya et al. 2004). Oxidation reaction occurs in the environment, where has a high ration of oxygen. In this condition arsenite As(III) is unstable; thus, it will be oxidized to arsenate As(V). However, this process is very slow if it has only oxygen to process. Therefore, ferric iron Fe(III), manganese oxides MnO₂, clay minerals, and some microorganisms have to play the vital role to intensively increase the rate of As(III) oxidation converting into the less toxic As(V) form this reaction is thermodynamically feasible over a wide range of pH values (Scott and Morgan 1995). Moreover, clay minerals and some microorganisms, including *Pseudomonas arsenitoxidans*, *Alcaligenes faecalis*, *Cenibacterium arsenoxidans*, *Thermus sp.*, *Thermus thermophilus*, and *Agrobacterium tumefaciens*, also can oxidize As(III) (Lin and Puls 2000, Valenzuela, Campos et al. 2009).

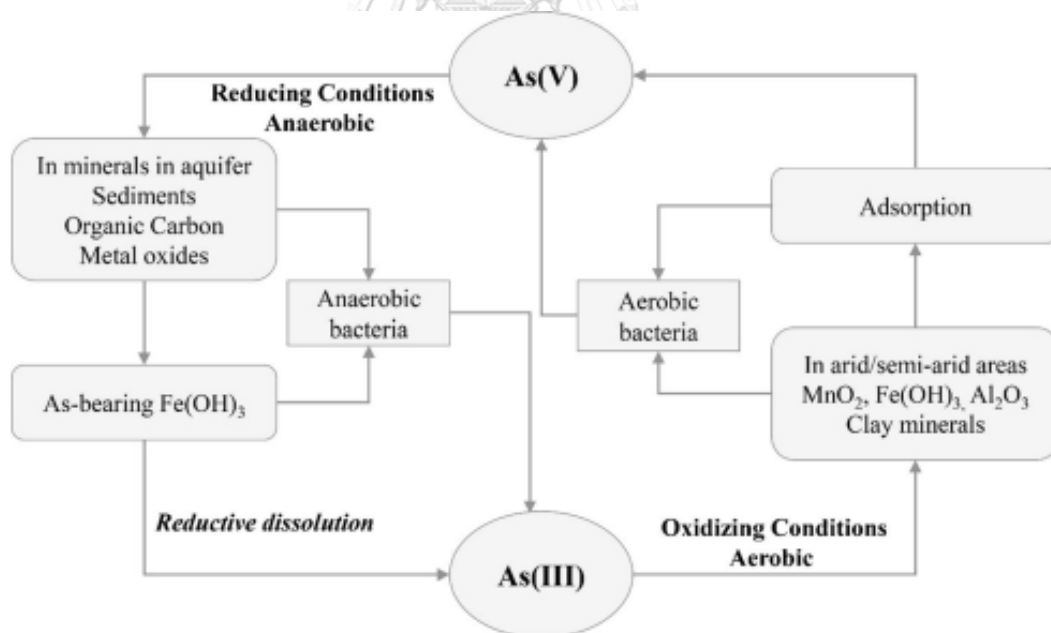


Figure 7 Illustration of the mechanisms of mobilization and redox transformation of As in aquifer sediments (Herath, Vithanage et al. 2016)

2.3 Machine learning

2.3.1 RF (Random forest)

Random forest algorithm is a modification of bagging that built a large collection of decision trees and average them. Therefore, random forests are popular and are implemented in a variety of packages (Breiman 2001). The bagging can demonstrate, given a train set $X = X_1, \dots, X_n$ with dependent $Y = Y_1, \dots, Y_n$, bagging repeat (A times) selected a random data from the training set and fits trees as this sample: $a = 1, \dots, A$:

1. Ample, with replacement, n training, X, Y ; will be X_a, Y_a .
2. Train a classification or regression tree f_a on X_a, Y_a .

After training, predictions for samples, the prediction from all the individual regression trees on X' can be averaging as an equation 1

$$f = \frac{\sum_{a=1}^A f_a (X')}{A} \text{-----Eq. 1}$$

This process can lead to better model performance, because it decreases the number of variances in the model, and didn't increase the bias. This means using a single tree is highly sensitive to noise and error, however, the average of many trees is minimum noise and error. Simply training many trees on a single training set would give strongly correlated. Moreover, to estimate the uncertainty of the prediction results, the predictions from all the individual regression trees on x' can be made as a standard deviation as an equation 2.

$$\sigma = \sqrt{\frac{\sum_{a=1}^A (f_a(x') - f)^2}{A-1}} \text{-----Eq. 2}$$

A is an optimal number of free parameters. Generally, multiple trees are used, depending on the size and nature of the training set. Using cross-validation or observing the out-of-bag error can be found an optimal number of trees A. The training and test error tends to level off after some number of trees have been fit.

2.3.2 SVM (Support vector machine)

Support vector machine (SVM) is one of the most usually used algorithms for prediction methods based on risk evaluation and classification (Evgeniou, Pontil et al. 2000). The SVM can simplify formulation is the linear equation as equation 3, where the hyperplane lies on the space of the input data x .

$$f(x) = w \cdot x + b. \text{----- Eq. 3}$$

In their general formula of SVM, a hyperplane is a feature space induced by a kernel K (the kernel defines a dot product in that space. The hypothesis space of the kernel K is defined as a set of “hyperplanes” in feature space. This can be also set of Reproducing kernel Hilbert space (RKHS). Also, SVM is a subset of hyperplanes, which can be formally written as Eq. 4

$$(f:|f|_k^2 > \infty) \text{-----Eq. 4}$$

Where K is the kernel, and $|f|_k^2$ is the RKHS norm of the function (Wahba, 1990). For example, for the linear case mentioned above, K is the kernel $K(x_1, x_2) = x_1 \cdot x_2$, the functions considered are of the form $f(x) = w \cdot x + b$, and the RKHS norm of these functions is simply the norm of w , namely $|f|_k^2 = |w|^2$. In fact SVM consider subsets of this space can be written as Eq. 5

$$(f:|f|_k^2 > A^2) \text{-----Eq. 5}$$

The goal of SVM is to find the solution with the "optimal" RKHS norm, that is, to find the optimal A . This SVM search for the optimal A has been discussed in the literature (Evgeniou, Pontil et al. 2000). SVM is learning machine algorithms that minimize the empirical error while taking into account the complexity of the hypothesis space used by minimizing the RKHS norm of the $|f|_k^2$. SVM in practice minimizes a tradeoff between empirical error and complexity of hypothesis space.

SVM classification

$$\text{Min } |f|_k^2 + C \sum_{i=1}^i |1 - yf(x)| \text{ -----Eq.6}$$

SVM regression

$$\text{Min } |f|_k^2 + C \sum_{i=1}^i |y - f(x)| \text{ -----Eq.7}$$

C is called "regularization parameter", which controls the tradeoff between empirical error and complexity of the hypothesis space.

2.3.3 ANN (Artificial neural network)

Artificial Neural Network is a computing system inspired by a biological neural network that constitutes an animal brain (Gupta, Akinola et al. 2019). Such systems "learn" to perform tasks by considering examples, generally without being programmed with any task-specific rules. The Neural Network is constructed from 3 types of layers: An input layer, Hidden layers, and Output layer, which produce the result for given inputs, as shown in Figure 8.

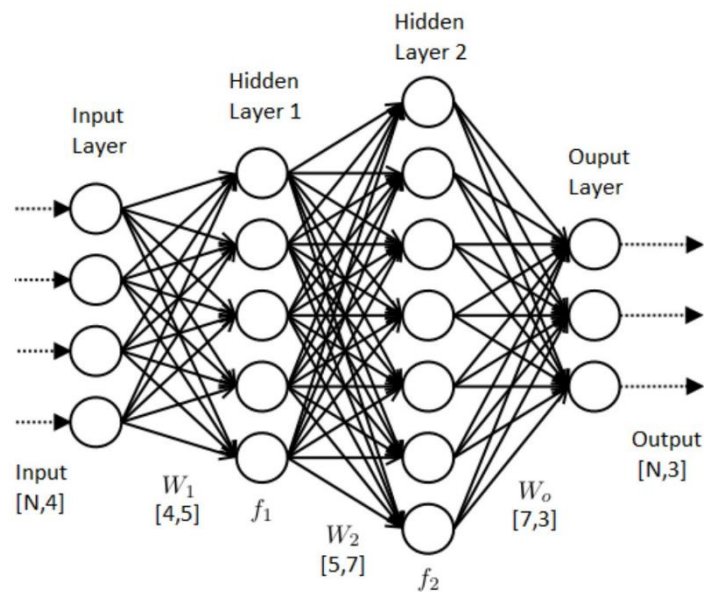


Figure 8 DNN diagram

The 3 yellow circles on the image above. They represent the input layer and usually are noted as vector X . There are 4 blue and 4 green circles that represent the hidden layers. These circles represent the “activation” nodes and usually are noted as W or θ . The red circle is the output layer or the predicted value. Each node is connected with each node from the next layer and each connection has a particular weight. Weight can be seen as the impact that that node has on the node from the next layer.

2.3.4 Summarize all algorithms

Every model generating algorithm, such as RF, SVM, and ANN has its advantage and disadvantage points as shown in Table 2. Random Forest (RF) is very fast and has satisfactory results with the limited data set (Havryliuk, Korol et al. 2018). However, there are many disadvantages of RF algorithms, e.g. for very large data sets, the size of the trees can take up a lot of memories. Thus, this algorithm needs a lot of computation resources for calculation. Moreover, this algorithm tends to overfit the data, so the model can't analyze a new data set (Kho 2018). Support vector machine (SVM) is good to classify data and trends to be resistant to overfitting

problems (K 2019). Nevertheless, SVM still has some disadvantages such as it does not perform very well when the data set has more noises i.e., target classes are overlapping. In case the number of features for each data point exceeds the number of the training data sample, the SVM will underperform (K 2019). Artificial neural networks (ANN) can learn and model complex and non-linear relationships, which very crucial, many of the relationships between inputs and outputs are non-linear and complex. Moreover, ANN can infer and analyze the unseen relationships between data (Mahanta 2017). A disadvantage of ANN is the unexplained behavior of the network: when ANN produces a probing solution, it does not give a clue as to why and how (Mijwil 2018). All three algorithms are suitable to handle a large amount of data, which have many variables to analyze and predict the model.

Table 2 Summarize the advantage and disadvantage of RF, SVM, ANN

	RF	SVM	ANN
Advantages	<ul style="list-style-type: none"> • Fast process • Work well with limited data 	<ul style="list-style-type: none"> • Resist overfitting 	<ul style="list-style-type: none"> • Learning ability • Analyze unseen relationships
Disadvantages	<ul style="list-style-type: none"> • Use a lot of computation resource • Trended to overfitting 	<ul style="list-style-type: none"> • Not perform well with noising data • Not work with data that has variable more than the sample 	<ul style="list-style-type: none"> • Unexpected behavior of the network • Overfitting

2.4 Risk and probability occurred maps

2.4.1 Probability map

A probability occurred map is a map that defines the probability distribution of occurred event. The distribution of occurred events bases on the probability level can

be arranged as shown in Table 3. The probability level starts from 0 to 1 or 0-100%, when the probability level enhances the chance of the event to occur will increase as well. Many researchers used a probability to show the risk assessment in groundwater to clearly understand the distribution of a pollutant (Winkel, Berg et al. 2008, Sajedi Hosseini, Malekian et al. 2018).

Table 3 The occurrence of the event based on probability level (Garvey 2001)

Probability of Occurrence		
0 - 10%	or	Very unlikely to occur
11 - 40%	or	Unlikely to occur
41 - 60%	or	May occur about half of the time
61 - 90%	or	Likely to occur
91 - 100%	or	Very likely to occur

2.4.2 Risk map

Generally, a risk map is also known as a data map visualization tool for communicating specific risks. A risk map helps a decision maker identify and prioritize the risks associated with their concerning topic. Risk pollution occurred map in groundwater is widely used in groundwater contamination assessment to determine and recognize areas that are more trended to contaminate than others (Winkel, Berg et al. 2008, Sajedi Hosseini, Malekian et al. 2018).

2.5 Spearman's correlation

Spearman's rank-order correlation is the nonparametric version of the Pearson product-moment correlation. Spearman's correlation coefficient, (ρ , also signified by r_s) measures the strength and direction of the association between two ranked variables. Spearman's correlation determines the strength and direction of the monotonic relationship between your two variables rather than the strength and direction of the linear relationship between your two variables, which is what Pearson's correlation determines. A monotonic relationship is a relationship that does

one of the following: (1) as the value of one variable increases, so does the value of the other variable; or (2) as the value of one variable increases, the other variable value decreases. Examples of monotonic and non-monotonic relationships are presented in the diagram below:

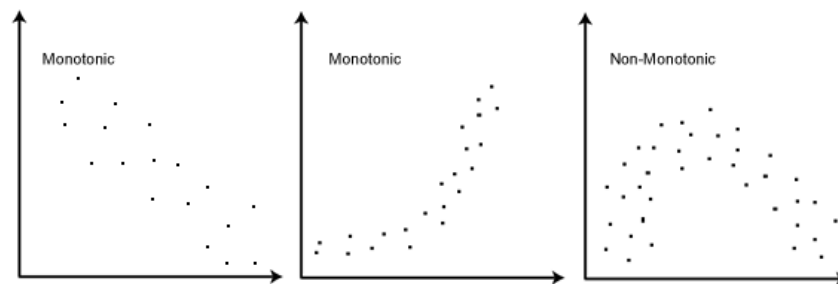


Figure 9 monotonic and non-monotonic relationships diagrams

There are two methods to calculate Spearman's correlation depending on whether: (1) your data does not have tied ranks or (2) your data has tied ranks. The formula for when there are no tied ranks is:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2-1)} \text{----- Eq. 8}$$

where d_i = difference in paired ranks and n = number of cases. The formula to use when there are tied ranks is:

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \text{----- Eq. 9}$$

where i = paired score. The Spearman correlation coefficient, r_s , can take values from +1 to -1. A r_s of +1 indicates a perfect association of ranks, a r_s of zero indicates no association between ranks and a r_s of -1 indicates a perfect negative association of ranks. The closer r_s is to zero, the weaker the association between the ranks. It is important to realize that statistical significance does not indicate the strength of Spearman's correlation. The statistical significance testing of the Spearman

correlation does not provide you with *any* information about the strength of the relationship. Thus, achieving a value of $p = 0.001$, for example, does not mean that the relationship is stronger than if you achieved a value of $p = 0.04$. This is because the significance test is investigating whether you can reject or fail to reject the null hypothesis. If you set $\alpha = 0.05$, achieving a statistically significant Spearman rank-order correlation means that you can be sure that there is less than a 5% chance that the strength of the relationship you found (your ρ coefficient) happened by chance if the null hypothesis were true (Laerd 2018).

2.6 IDW

Interpolation is a method to predict an unknown from known values. From the definition, we need some known values to do an interpolation using any interpolation method. The known values which are commonly called sampling points can be gathered from some measurements and site investigations like drilling, surveying, etc. Using the known value from some locations, we are trying to predict the value of other neighborhood location that is close to the known location. There are many interpolation methods available including Inverse Distance Weighting (IDW). Inverse Distance Weighted interpolation is a deterministic spatial interpolation approach to estimate an unknown value at a location using some known values with corresponding weighted values. The basic IDW interpolation formula can be seen in equation 10. Where x^* is the unknown value at a location to be determined, w is the weight, and x is the known point value. The weight is the inverse distance of a point to each known point value that is used in the calculation. Simply the weight can be calculated using equation 11.

$$x^* = \frac{w_1x_1 + w_2x_2 + \dots + w_nx_n}{w_1 + w_2 + \dots + w_n} \text{ ----- Eq. 10}$$

$$W_1 = \frac{1}{d_{1x}^p} \text{ ----- Eq. 11}$$

Figure 10 gives an illustration of how the IDW interpolation works. As can be seen in the figure, a value at position x will be determined from sampling points 1, 2, and 3, with the distances to x point, are d_{1x} , d_{2x} , and d_{3x} . Using equation 11, each respective weight will be calculated and then the value at position x will be determined using equation 10 (Geomatics 2019).

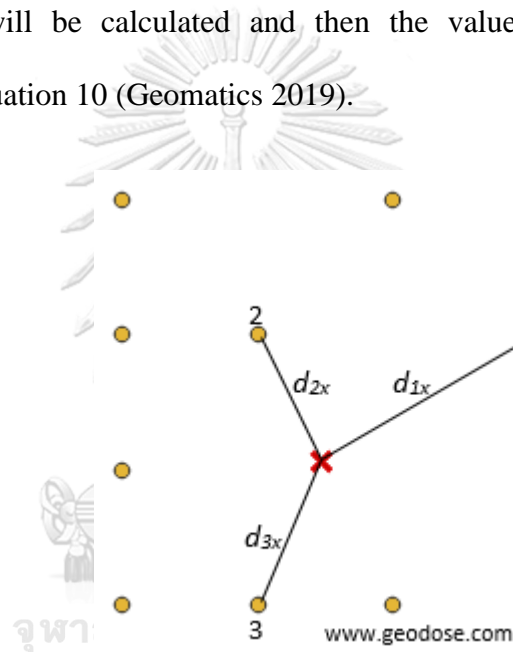


Figure 10 Inverse Distance Weight (IDW) Interpolation

2.7 Quartile regression

Usually, the studies to evaluated models' performance are mostly evaluated the prediction performance of the model and disregarded models' uncertainties. It is well known that uncertainty is inherent in modeling (Solomatine and Shrestha 2009), thus it is crucial to report the uncertainty of the model to make it transparent in decision-support tools (Uusitalo, Lehikoinen et al. 2015). To evaluate the uncertainty of the models, commonly using prediction intervals. One way of generating the prediction

interval is through quantile regression. As opposed to linear regression which needs to estimate the conditional mean of the response variable given certain values of the predictor variables, quantile regression aims at estimating conditional quantiles (typically, median) of the response variable. quantile regression is not limited to just finding the median, but can calculate any quantile (percentage) for a particular value in the feature's variables. the quantile regression model equation for the τ^{th} quantile is

$$Q_{\tau}(y_i) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \dots + \beta_p(\tau)x_{ip} \quad i = 1, \dots, n \text{ -----Eq.12}$$

To create a prediction interval, we can now use other quantile values. For example, in the image below we have 0.977 and 0.023 percentiles (Figure 11). This gives a prediction interval with a 0.95 probability of having the true value within its bounds.

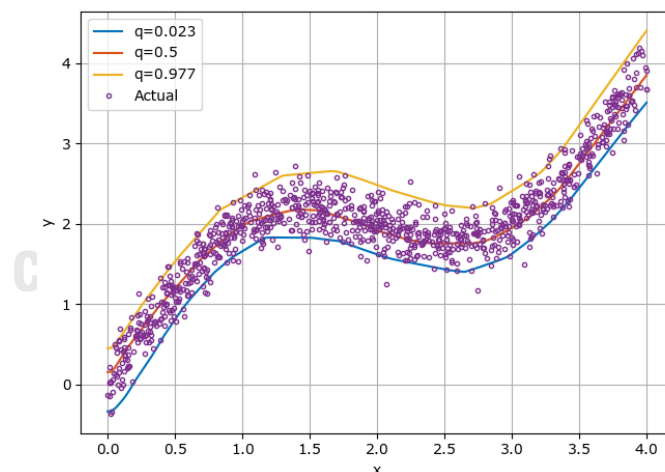


Figure 11 Predictions for quantiles

There are several statistical measures of uncertainty such as mean prediction interval (MPI) and prediction interval coverage probability (PICP), which were used as suggested by Shrestha and Solomatine (Shrestha and Solomatine 2006). MPI is the

average of the widths of the prediction intervals, where the lower values of MPI indicate lower uncertainty. The best method MPI and PICP values are calculated as:

$$MPI = \frac{1}{n} \sum_{t=1}^n (PL_t^{upper} - PL_t^{lower}) \text{ ----- Eq. 13}$$

$$PICP = \frac{1}{n} \sum_{t=1}^n C, C = (1, PL_t^{lower} < y_t < PL_t^{upper}, 0, otherwise) \text{ ----- Eq. 14}$$

where y_t is the observed value, PL_t^{lower} and PL_t^{upper} are lower and upper prediction limits respectively. The PICP is the more important measurement of uncertainty as it indicates the number of observations that fall within the estimated interval (Dogulu et al., 2015). Therefore, MPI is used as a supplementary metric: between models with similar PICP values, the one with a lower MPI is regarded as the better model (Muthusamy, Godiksen et al. 2016).

Chapter 3 Methodology

3.1 Framework of research

The overall framework of this study is shown in Figure 12. The preparing data process was conducted by collecting data such as groundwater chemical data, Soil type, land use, geologic, aquifer, etc. from various departments. The data were categorized as deep and shallow aquifers. The arsenic concentration parameter was classified by standard arsenic in drinking water to using them in the modeling process. Other parameters were screened by spearman's correlation technique to screen out unnecessary parameters. In the modeling process, the spatial modeling uses a different algorithm to generate different probability models such as Support vector machine (SVM), Random Forest (RF), and Artificial neural networks (ANNs) these models will produce a probability map that can locate the As potential area.

The probability map will be compared with each other in the validation process. There are three aspects that we used in the validation process to measure the models such as prediction performance, uncertainty evaluation, and Validation with field data. The best probability map was classified between risk and non-risk areas by the Cut-off value technique. Then, the map was calculated with the population density and water consumption data in the study area to generate the risk maps.

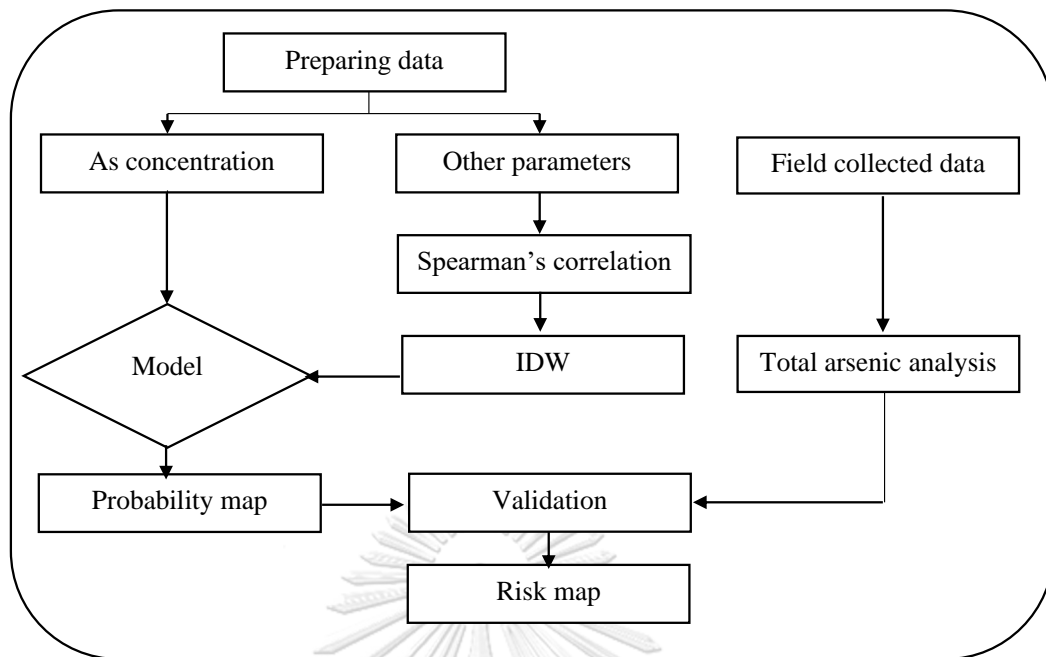


Figure 12 Framework of the present study

3.2 Data preparation and field-collected data

3.2.1 Data preparation

The data were collected from various government departments such as the Department of mineral and resource (DMR), Department of Groundwater Resources (DGR), and Land Development Department (LDD). These data can be classified into two groups physical and hydrochemical data. The period of hydrochemical data was collected around 2011-2012, 2017-2018, and 2019 in dry and rainy seasons around the study area (DMR 2007, DGR 2012, LDD 2016, DGR 2017). The missing data of each hydrochemical parameters are replaced with limit detection of the measure tool. The purpose of the study want the risk map to do not confuse the user, so we attempt to merge data from dry and rainy seasons every year together. Thus, this will make the risk map simply to use and increase the number of data to generate a model, which increases the performance of the model.

3.2.1.1 Physical data

The physical data were collected from various government departments and the form of the data categorized as raster data, which ready to use in the modeling process. The data combined with various aspects such as geological characteristics, soil properties, soil texture, aquifer characteristics, groundwater level. The date of these data is not over 10 years as shown in Table 4.

Table 4 Source of physical data

Data	Data source	Year
Geological characteristics	Ministry of Natural Resources	2015
Soil properties	Land Development Department	2011
Aquifer characteristics	Department of Groundwater Resources	2017
Land use	Land Development Department	2015
Elevation	USGS	2019
Population density	National Statistical Office	2019
Groundwater consumption	Department of Groundwater Resources	2019

3.2.1.2 Hydrochemical data

Hydrochemical data is collected from the Department of Groundwater Resources (DGR) and categorize as the point data that aren't ready to use in the modeling process. Hydrochemical data, including a concentration of Calcium (Ca), Magnesium (Mg), Sodium (Na), Potassium(K), Iron (Fe), Chloride (Cl), Fluoride(F⁻), Carbonate (CO₃²⁻), Bicarbonate (HCO₃²⁻), Sulfate (SO₄²⁻), Nitrate (NO₃⁻), Phosphate (PO₄⁻), Cadmium (Cd), Chromium (Cr⁶⁺), Copper (Cu), Mercury (Hg), Manganese (Mn), Nickel(Ni), Lead(Pb), Selenium(Se), Zine(Zn), and other hydrochemical

parameters, such as Electrical Conductivity(EC), Temperature, pH, Total dissolved solids (TDS) and Total hardness (TTH).

3.2.2 Field collected data

The field data were collected during 31 August – 1 September 2019 from twenty-seventh wells around the Rayong groundwater basin shown in Figure 13 and Table 5. The collecting methods will follow the groundwater sampling guidelines for Superfund and RCRA Project Managers (USEPA 2002).

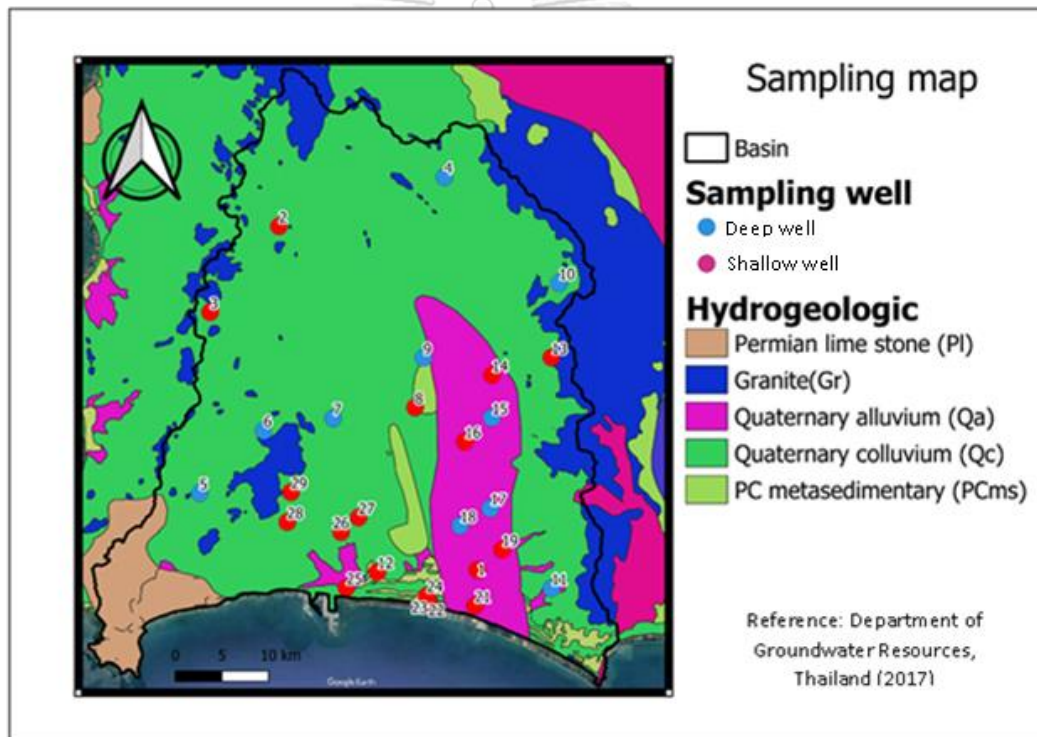


Figure 13 Groundwater sampling map

Table 5 Groundwater sampling aquifer

No	Name	Latitude	Longitude	Well type	Aquifer
1	G10	-12° -41' -46.748" S	-101° -17' -54.635" W	Deep	Gr
2	G7	-12° -45' -19.991" S	-101° -18' -42.622" W	Deep	Gr
3	G8	-12° -44' -20.841" S	-101° -16' -55.013" W	Deep	Gr
4	G9	-12° -42' -56.31" S	-101° -19' -22.872" W	Deep	Gr
5	J11	-12° -58' -20.979" S	-101° -22' -54.267" W	Deep	Gr
6	J2	-13° -1' -46.882" S	-101° -6' -19.05" W	Deep	Gr
7	J3	-12° -56' -50.222" S	-101° -2' -11.284" W	Deep	Gr
8	J4	-13° -4' -32.647" S	-101° -16' -9.835" W	Deep	Gr
9	J8	-12° -50' -37.698" S	-101° -9' -26.373" W	Deep	Gr
18	G3	-12° -50' -34.81" S	-101° -18' -50.928" W	Deep	Gr
19	J10	-12° -54' -5.369" S	-101° -14' -49.394" W	Deep	Gr
20	J6	-12° -46' -21.475" S	-101° -1' -28.79" W	Deep	Gr
21	J7	-12° -49' -57.246" S	-101° -5' -18.295" W	Deep	Gr
10	J1	-12° -41' -46.748" S	-101° -17' -54.635" W	Shallow	Qa
11	G1	-12° -54' -2.419" S	-101° -22' -24.755" W	Shallow	Qcl
12	G11	-12° -39' -41.817" S	-101° -17' -44.834" W	Shallow	Qcl
13	G12	-12° -39' -58.382" S	-101° -15' -0.359" W	Shallow	Qcl
14	G13	-12° -40' -7.653" S	-101° -14' -56.597" W	Shallow	Qcl
15	G14	-12° -40' -18.213" S	-101° -14' -54.271" W	Shallow	Qcl
16	G15	-12° -40' -48.242" S	-101° -10' -8.196" W	Shallow	Qcl
17	G2	-12° -53' -2.933" S	-101° -18' -53.016" W	Shallow	Qcl
22	J13	-12° -40' -40.22" S	-101° -22' -17.722" W	Shallow	Qa
23	J14	-12° -41' -42.018" S	-101° -11' -59.16" W	Shallow	Qa
24	G17	-12° -44' -52.257" S	-101° -10' -53.785" W	Shallow	Qcl
25	G19	-12° -46' -21.952" S	-101° -6' -53.877" W	Shallow	Qcl
26	G4	-12° -49' -12.899" S	-101° -17' -14.244" W	Shallow	Qcl
27	J9	-12° -51' -12.963" S	-101° -14' -17.753" W	Shallow	Qcl

3.2.2.1 Data measuring in the field

Groundwater level

All groundwater levels will be measured from the reference point by the use of a weighted steel tape and chalk or an electric tape (USEPA 2002).

Hydrochemical parameters measured on-site

Hydrochemical parameters, consisting of EC, ORP, Temperature, pH, will be measured on-site by a multi-parameter meter (USEPA 2002).

3.2.2.2 Total arsenic analysis

The collecting methods are according to Ground-Water Sampling Guidelines for Superfund and The Resource Conservation and Recovery Act (RCRA) Project Managers (USEPA 2002). The collected samples will be fixed by HCl(con) 1 cc. and contained under 4°C. Then the samples will be sent to the UAE laboratory to analyse the amount of total arsenic in water samples by the Hydride Generation AAS technique with a detection limit of 0.3 µg/L.

3.3 Arsenic contamination probability map

3.3.1 Framework of the arsenic contamination probability map

The dataset will be divided into arsenic concentration and other parameters. The Arsenic concentration will use to conduct the probability model. On the other hand, other parameters will be screened by Spearman's correlation method to eliminate unnecessary parameters, separately between shallow and deep aquifers. After that, the selected parameters will be interpolated by Inverse Distance Weight (IDW) interpolation process that was used to create the arsenic probability map as shown in Figure 14.

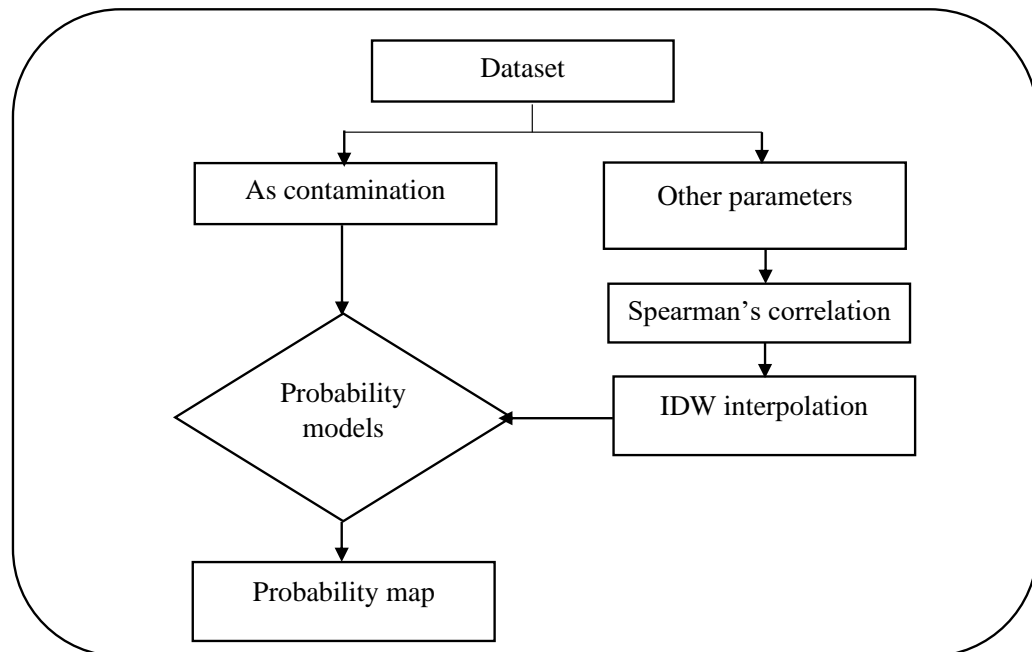


Figure 14 Framework of constructing the probability map of As

3.3.2 Spearman's correlation

Nonparametric Spearman's rank correlation was calculated using the SPSS software to study the monotonic relationships between arsenic and physical parameters and hydrochemical parameter, to draw inferences about mechanisms of arsenic release. A correlation was considered significant when the correlation coefficient (r_s) excess 0.1.

3.3.3 IDW interpolation

To make the points data (Hydrochemical parameter) ready to use in the modeling process, the data will be interpreted from point data to raster data by using the IDW interpolation package in R-studio software. IDW interpolation is the interpretation process that usually use to generate raster data. It can generate the point data (concentration) into area data (concentration area) by averaging the point data.

3.3.4 Model probability

The framework of model probability was showed in Figure 15. Star with, Arsenic concentrations parameter which obtained from the dataset was be used to provide a groundwater probability map in the Rayong groundwater basin. According to the World health organization, a threshold of arsenic ($10 \mu\text{g/l}$) was to be used to classify the polluted and non-polluted wells. Groundwater wells were classified into 2 types as follows: the polluted type when arsenic concentration is above $10 \mu\text{g/l}$, and the non-polluted type when arsenic concentration is below $10 \mu\text{g/l}$. After that, the datasets were randomly divided into a training dataset (70% of the dataset) and a testing dataset (30% of the dataset). The training set will be used to provide the probability models along with three algorithms, consisting of SVM, RF, and ANN through coding by SDM package in R-studio software (Naimi and Araújo 2016). The probability models had measured their performance. If their performance is above 60% of the area under the curve (AUC), the model will be acceptable. In contrast, if the model performance is below 60% of AUC, the model had to be calibrated and provide the new models until it reaches 60% AUC (Havryliuk, Korol et al. 2018). When finishing the model, the arsenic probability map will be generated using the data from IDW interpolation using the SDM package in R-studio software (Naimi and Araújo 2016).

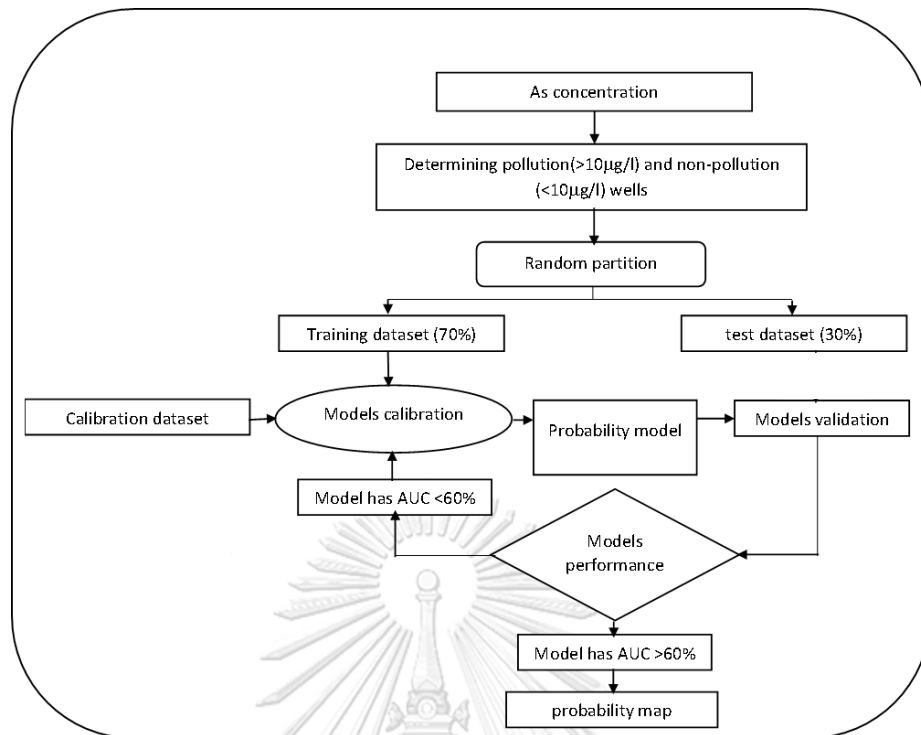


Figure 15 Framework of constructing the probability model of As

3.4 Probability map validation

3.4.1 Evaluating prediction performance

To measure the performance of the model, the models were validated with 30% of cases that were not used for training. Model accuracy was evaluated using an area under curve (AUC) and Root mean square (RMSE). The Receiver operating characteristic (ROC) is a probability curve to measurement a model's performance for classification problems at various threshold settings. It uses area under curve (AUC) representing a degree or measure of separability. It tells how much the model is capable of distinguishing between classes (Narkhede 2018). Moreover, a graphical comparison was conducted using Taylor diagrams (Taylor 2001), which enable visualization of the models' performances using correlation coefficients, RMSE, and standard deviations (SD) (Choubin, Malekian et al. 2017).

3.4.2 Evaluate models uncertainty

To make sure the models are work properly and reliably, the quantile regression (QR) was used to assess the predictive uncertainty of the models. These methods evaluate the model residuals and consider all sources of uncertainty, which is in contrast to the classic methods (such as Monte Carlo-based methods) in which the estimate usually regards only one source of uncertainty (Solomatine and Shrestha 2009).

3.4.3 Validation with field data

To evaluate the performance of models, The As concentration collected in the field were used to compare between the actual value and prediction value from probability map in each model using Root mean square as a measurement indicator.

3.5 Groundwater pollution risk map

3.5.1 Probability cutoff value

The best model selected through comparing every model using prediction performance and uncertainty to select them was used to create the risk map. But, before the probability map can be used to create the risk map. It had to classify by cut-off value first. The cutoff value was calculated through the Caret package in R-studio software. The cutoff value was used to determine the risk area and non-risk area in the probability map (Bindal and Singh 2019). The cut-off value was used to classify between Arsenic probability risk area and non-risk area in the probability map. When Arsenic probability greater than the cutoff value the area will be classified as the risk area. In contrast, when the area had a probability below the cutoff value, this area will be classified as a non-risk area (Winkel, Berg et al. 2008).

3.5.2 Risk map for As

After producing the probability map which already classifies by cut-off value. The risk map was created by using Qgis software using data of population density and water consumption in Rayong groundwater basin base on equation 13. The population density data derived from The Bureau Of Registration Administration in 2020 (BRA 2020) and the water consumption came from the Department of Groundwater resource 2020 (DGR 2020). The water consumption data created by using the pumping rate of groundwater well selected the well that used in domestic and agricultural use. The pumping rate from the selected wells was normalized in the range from 0-1, and the interpolation was carried out with the IDW technique to generate water consumption map.

$$\text{Risk} = \text{Probability map} \times \text{Population density} \times \text{Water consumption} \text{ -----Eq. 13}$$

The unit of As risk map is the people who use groundwater that might expose to As per square kilometer. To make them more simple, the map unit will be classified into five risk levels such as Very High, High, Moderate, Low, and Very low using equal interval mode to classify (Bindal and Singh 2019).

Chapter 4 Results

4.1 Groundwater wells data

The groundwater quality parameters were obtained from the Department of Groundwater Resources (DGR). The secondary data was separated into 2 datasets consisting of a deep well and shallow well groups, which represented hydrochemical data in the deep aquifer and the shallow aquifer. The descriptive statistics of groundwater quality parameters in the deep aquifer were shown in Table 6. There are twenty-four hydrochemical parameters.

Generally, according to the field measurement parameters, pH ranged from 4.10 to 8.30 with an average of 6.93 indicating groundwater was slightly acidic. The electrical conductivity (EC) values ranged from 0.0 to 31,000 $\mu\text{s}/\text{cm}$, with an average of 754.36 $\mu\text{s}/\text{cm}$. Similarly, the total dissolved solids (TDS) values were 12.26-19855.00 mg/l. Lastly, total hardness values were in the range of 3-7,100 mg/l, with an average value of 309.27 mg/l.

Generally, Ca and Mg range from 0.10-1400 mg/l and 0-1,100 mg/l, respectively. Na and Cl had range from 0.01-6,000 mg/l and 1.5-1,1000 mg/l similarly with K and Fe with range between 0.87-240 mg/l and 0.01-270 mg/l respectively. In addition, for other anions, F, HCO_3 , SO_4 , NO_3 also had range from 0.01-16.60 mg/l, 1-494 mg/l, 0.1-2800 mg/l and 0.01-80.6 mg/l, respectively. Arsenic was in the range between 0.30-280 $\mu\text{g}/\text{l}$ with an average of 13.85 $\mu\text{g}/\text{l}$, which is higher than the groundwater drinking standard the other heavy metals in the deep aquifer had a wide range concentration. For example, Cd, Cr, Cu, Hg, Mn, Ni, Pb, Se, Zn had ranges from 0.4-1.5 $\mu\text{g}/\text{l}$, 2.40-20 $\mu\text{g}/\text{l}$, 3-400 $\mu\text{g}/\text{l}$, 0.1-10 $\mu\text{g}/\text{l}$, 5-22000 $\mu\text{g}/\text{l}$, 1-30 $\mu\text{g}/\text{l}$, 0.70-27.7 $\mu\text{g}/\text{l}$, 0.3-160 $\mu\text{g}/\text{l}$, and 5-150000 $\mu\text{g}/\text{l}$ respectively.

The metals which were detected to be higher than groundwater standard was Hg, Mn, Ni, Pb, Se, and Zn.

The descriptive statistics of groundwater quality parameters in the shallow aquifer shown in Table 7. According to the field measurement parameters, pH ranged from 3.10-8.30 with an average of 6.77, indicating more slightly acidic than that in the deep aquifer. The electrical conductivity (EC) values ranged from 2.28 to 44,400 $\mu\text{s}/\text{cm}$, with an average of 1644.5 $\mu\text{s}/\text{cm}$. Besides, TDS values were in the range of 13-27,600 mg/l. Lastly, TDS ranged from 1-6,600 mg/l with average value 277.83 mg/l.

Similar to those parameters in the deep aquifer, Ca and Mg were in the range of 0.05-620 mg/l and 0-1,400 mg/l respectively. Na and Cl had a range from 0.005-7,000 mg/l and 0.20-13,166 mg/l, which was similar to K and Fe with a range between 0.01-710 mg/l and 0-70 mg/l respectively. In addition, for other anions, F, HCO_3 , SO_4 , NO_3 also had widely range from 0.01-11.20, 0-5,700, 0.1-1,800, and 0.01-240 mg/l, respectively.

Arsenic was in the range of 0.3-500 $\mu\text{g}/\text{l}$ with an average of 12.85 $\mu\text{g}/\text{l}$, which was higher than the groundwater drinking standard. The other heavy metals in the shallow aquifer had a wide range of concentrations. Cadmium, Cr, Cu, Hg, Mn, Ni, Pb, Se, Zn had a wide range from 0.4-0.6 $\mu\text{g}/\text{l}$, 2.40-40 $\mu\text{g}/\text{l}$, 3-500 $\mu\text{g}/\text{l}$, 0.1-700 $\mu\text{g}/\text{l}$, 5-18000 $\mu\text{g}/\text{l}$, 1-180 $\mu\text{g}/\text{l}$, 0.70-60 $\mu\text{g}/\text{l}$, 0.3-160 $\mu\text{g}/\text{l}$, and 5-1900 $\mu\text{g}/\text{l}$ respectively. The metals which were detected to be higher than the groundwater standard was Cu, Mn, Ni, Pb, and Se.

To know the distribution of data, Skewness and Kurtosis measurement were used to describe groundwater quality parameters. In particular, groundwater quality

parameters in the deep aquifer were explained by normal distribution were just only pH, HCO₃, Cr, and Ni with Skewness and Kurtosis -0.89/2.25, 0.29/-0.74, -0.53/1.37, 1.90/5.13 respectively. As compared to those in the deep aquifer, groundwater quality parameters in the shallow aquifer which close to be normal distribution were only pH and Cr with Skewness and Kurtosis around -0.89/2.25, 1.90/5.13 respectively.

Table 6 Description statistics in the deep aquifer

	Unit	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis	Detection limit
EC	µs/cm	236	0.00	31000.00	754.36	2884.50	9.52	94.85	-
pH	-	236	4.10	8.30	6.93	0.64	-0.89	2.25	1-14
TDS	mg/l	236	12.60	19855.00	740.39	2640.95	6.52	43.49	5.00
Total hard	mg/l	236	100.00	7100.00	309.27	917.80	5.79	34.75	100.00
Ca	mg/l	236	0.10	1400.00	71.27	164.99	5.21	30.66	0.05
Mg	mg/l	236	0.00	1100.00	25.38	105.42	7.44	61.41	0.00
Na	mg/l	236	0.01	6000.00	144.14	634.69	6.80	49.69	0.005
K	mg/l	236	0.87	240.00	9.30	18.98	8.74	96.20	0.005
Fe	mg/l	236	0.01	270.00	3.30	18.19	13.62	198.73	0.005
Cl	mg/l	236	1.50	11000.00	259.85	1329.45	7.08	50.97	0.2
F.	mg/l	236	0.01	16.60	1.08	1.95	4.56	27.32	0.01
HCO ₃	mg/l	236	1.00	494.00	175.87	105.05	0.29	-0.74	0
SO ₄	mg/l	236	0.10	2800.00	62.50	317.56	7.08	52.54	0.10
NO ₃	mg/l	236	0.01	80.60	4.51	9.06	4.38	26.81	0.01
As	µg/l	236	0.30	280.00	13.85	34.56	5.87	37.72	0.30
Cd	µg/l	236	0.40	1.50	0.40	0.07	15.36	236.00	0.40
Cr⁶⁺	µg/l	236	10	20.00	8.58	3.42	-0.53	1.37	10
Cu	µg/l	236	3.00	400.00	7.08	29.36	10.97	139.18	3.00
Hg	µg/l	236	0.10	10.00	0.16	0.64	15.27	234.08	0.10
Mn	µg/l	236	5.00	22000.00	784.49	2239.41	6.73	53.86	5.00
Ni	µg/l	236	1.00	30.00	3.59	4.40	1.90	5.13	1.00
Pb	µg/l	236	0.70	21.70	1.37	2.67	5.08	28.65	0.70
Se	µg/l	236	0.30	160.00	2.14	11.32	12.15	163.98	0.30
Zn	µg/l	236	5.00	15000.00	2189.70	12845.44	9.91	104.47	5.00

*Bold characters using µg/l unit

Table 7 Description statistics in the shallow aquifer

	Unit	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis	Detection limit
EC	μs/cm	417	2.28	44400.00	1644.50	5470.32	5.14	27.59	-
pH	-	417	3.10	8.30	6.77	0.81	-1.27	3.53	1-14
TDS	mg/l	417	13.00	27600.00	1081.84	3537.66	4.99	25.75	5.00
Total hard	mg/l	417	100.00	6600.00	277.83	753.74	5.11	28.24	100.00
Ca	mg/l	417	0.05	620.00	49.41	88.08	4.10	18.21	0.05
Mg	mg/l	417	0.00	1400.00	35.49	132.07	6.11	43.34	0.00
Na	mg/l	417	0.005	7000.00	237.40	930.57	5.53	31.25	0.005
K	mg/l	417	0.01	710.00	20.30	66.33	7.44	62.81	0.005
Fe	mg/l	417	0.0005	70.00	3.64	9.10	4.29	22.02	0.005
Cl	mg/l	417	0.20	13166.00	433.07	1741.68	5.37	30.00	0.2
F.	mg/l	417	0.01	11.20	0.45	0.99	6.16	51.80	0.01
HCO ₃	mg/l	417	0.00	5700.00	180.75	356.96	10.37	143.88	0
SO ₄	mg/l	417	0.10	1800.00	77.16	254.08	4.79	23.31	0.10
NO ₃	mg/l	417	0.01	240.00	12.48	27.56	4.88	30.33	0.01
As	μg/l	417	0.30	500.00	12.85	41.81	7.97	77.05	0.30
Cd	μg/l	417	0.40	0.60	0.40	0.01	20.42	417.00	0.40
Cr⁶⁺	μg/l	417	2.40	40.00	8.79	3.71	1.58	16.43	10
Cu	μg/l	417	3.00	500.00	14.51	38.30	6.10	62.71	3.00
Hg	μg/l	417	0.10	700.00	1.92	34.35	20.27	412.77	0.10
Mn	μg/l	417	5.00	18000.00	700.29	1483.19	5.58	49.26	5.00
Ni	μg/l	417	1.00	180.00	3.18	9.54	15.52	284.81	1.00
Pb	μg/l	417	0.70	60.00	1.66	4.85	7.66	69.89	0.70
Se	μg/l	417	0.30	160.00	3.85	17.17	7.14	54.84	0.30
Zn	μg/l	417	5.00	1900.00	66.14	159.11	8.11	85.59	5.00

*Bold characters using μg/l unit

4.2 Field data

The results of field groundwater wells were as shown in Table 8. Ten groundwater wells out of twenty-seven wells found total As concentrations exceeding the standard of drinking water of 10 µg/l. Total As concentrations were in the range of 0.8-52.6 µg/l with an average value of 18.11 µg/l as shown in Table 9.

The in-situ measurement parameters were shown in Tables 8 and 9. The range of pH was in the range of 3.41-7.79 with an average value of 6.83 indicating a slightly acidic condition. Depth to groundwater table was in the range of 1.4-14.18 m. with an average of 5.46 m. Furthermore, the EC value ranged from 78.1 to 32,000 µs/cm. (average = 2677.35 µs/cm) and ORP was in the range between 1.40-288.4 mV (average = 5.64 mV).

For total As concentration, the average total As concentrations of Gr, Qa, and Qcl were 16.69, 27.5, and 17.19 µg/l, respectively, which were higher than the groundwater drinking standard. The maximum total As concentration was found at 52.60 µg/l in the Qc aquifer, following by 44.40 and 34.40 µg/l in Gr and Qa, respectively (Table 10).

Table 8 Arsenic concentrations in groundwater samples collected from Gr, Qa and Qcl aquifers during 31 August – 1 September, 2019

No	Name	Latitude	Longitude	Well type	Aquifer	Water table (m.)	Temp.(C)	pH	ORP	EC	As(μ g/l)
1	G10	-12° -41' -46.748" S	-101° -17' -54.635" W	Deep	Gr	2.1	29.3	7.04	-83.50	538.00	ND
2	G7	-12° -45' -19.991" S	-101° -18' -42.622" W	Deep	Gr	3.04	30	7.26	-42.50	11400.00	ND
3	G8	-12° -44' -20.841" S	-101° -16' -55.013" W	Deep	Gr	6.83	31	7.76	-78.50	6000.00	ND
4	G9	-12° -42' -56.31" S	-101° -19' -22.872" W	Deep	Gr	1.4	29.6	7.71	-165.20	402.00	ND
5	J11	-12° -58' -20.979" S	-101° -22' -54.267" W	Deep	Gr	7.65	28.1	6.669	-66.4	96.1	4.9
6	J2	-13° -1' -46.882" S	-101° -6' -19.05" W	Deep	Gr	2.55	30.6	6.912	199.3	196	5.1
7	J3	-12° -56' -50.222" S	-101° -2' -11.284" W	Deep	Gr	2.6	29	6.409	75.5	271	5.1
8	J4	-13° -4' -32.647" S	-101° -16' -9.835" W	Deep	Gr	5.63	29.1	6.367	-158.7	78.1	ND
9	J8	-12° -50' -37.698" S	-101° -9' -26.373" W	Deep	Gr	4.16	29.5	5.207	288.4	137.3	1.3
18	G3	-12° -50' -34.81" S	-101° -18' -50.928" W	Deep	Gr	3.64	29.6	7.79	-108.10	351.00	44.400
19	J10	-12° -54' -5.369" S	-101° -14' -49.394" W	Deep	Gr	4.17	29.7	7.74	-162.9	252	30.7
20	J6	-12° -46' -21.475" S	-101° -1' -28.79" W	Deep	Gr	14	29.3	6.053	126.5	103	20
21	J7	-12° -49' -57.246" S	-101° -5' -18.295" W	Deep	Gr	4.6	29.2	5.675	167.2	154.4	22
10	J1	-12° -41' -46.748" S	-101° -17' -54.635" W	Shallow	Qa	10.9	30	6.14	88.9	232	ND
22	J13	-12° -40' -40.22" S	-101° -22' -17.722" W	Shallow	Qa	1.8	29.2	7.215	249.1	177.5	20.6
23	J14	-12° -41' -42.018" S	-101° -11' -59.16" W	Shallow	Qa	3.27	30.6	7.777	212.9	1625	34.4
11	G1	-12° -54' -2.419" S	-101° -22' -24.755" W	Shallow	Qcl	14	28.8	6.36	172.10	301.00	ND
12	G11	-12° -39' -41.817" S	-101° -17' -44.834" W	Shallow	Qcl	4.6	30	7.78	-246.60	679.00	ND
13	G12	-12° -39' -58.382" S	-101° -15' -0.359" W	Shallow	Qcl	2.58	29	7.38	-225.30	461.00	3.000
14	G13	-12° -40' -7.653" S	-101° -14' -56.597" W	Shallow	Qcl	4.85	30.5	7.17	-111.90	262.00	0.800
15	G14	-12° -40' -18.213" S	-101° -14' -54.271" W	Shallow	Qcl	3.8	30	7.40	-143.40	6000.00	ND
16	G15	-12° -40' -48.242" S	-101° -10' -8.196" W	Shallow	Qcl	2.76	31.6	3.41	-104.30	32000.00	3.900
17	G2	-12° -53' -2.933" S	-101° -18' -53.016" W	Shallow	Qcl	5.58	29.9	7.17	-17.90	735.00	7.700
24	G17	-12° -44' -52.257" S	-101° -10' -53.785" W	Shallow	Qcl	5.2	29.5	7.60	-206.10	594.00	15.400
25	G19	-12° -46' -21.952" S	-101° -6' -53.877" W	Shallow	Qcl	5.2	29	7.50	-247.10	8240.00	40.800
26	G4	-12° -49' -12.899" S	-101° -17' -14.244" W	Shallow	Qcl	14.18	28.8	6.58	-137.00	659.00	13.300
27	J9	-12° -51' -12.963" S	-101° -14' -17.753" W	Shallow	Qcl	6.27	29.6	6.41	-281.7	344	52.6

Table 9 Descriptive statistics of in-situ measured parameters and total As concentration in groundwater during 31 August – 1 September 2019

	Max	Min	Mean	Std.
Water table (m.)	14.18	1.40	5.46	3.62
Temp.(C)	31.60	28.10	29.65	0.74
pH	7.79	3.41	6.83	0.96
ORP	288.40	1.40	5.64	3.73
EC	32000.00	78.10	2677.35	6400.63
As($\mu\text{g/l}$)	52.60	0.80	18.11	15.83

Table 10 Descriptive statistics of total As concentrations in Gr, Qa, and Qc aquifers during 31 August – 1 September 2019

	Max	Min	Mean	Std. Deviation
Gr	44.40	1.30	16.69	14.35
Qa	34.40	20.60	27.50	6.90
Qc	52.60	0.80	17.19	17.91

4.3 Spearman's correlation

In this study, Spearman's correlation was used to examine the correlation between parameters. In the deep aquifer, hydrochemical parameters were correlated as shown in Table 11. The significant correlations between total As with pH, TDS, Ca, Mg, F and HCO₃ at 0.01 level with positive correlation were at $r=0.222$, $r=0.230$, $r=0.170$, $r=0.171$, $r=0.183$, $r=0.291$, respectively and shown a significant correlation at 0.05 level with TTH, ($r=0.165$).

The hydrochemical parameters in the shallow aquifer were correlated at 0.01 level between total As and EC, pH, TDS, TTH (total hardness), Ca, Mg, Na, K, Fe, Cl, F, HCO₃, SO₄, NO₃, Cr, Cu, Hg, Mn, Ni, Pb and Se with $r=0.368, 0.255, 0.385, 0.281, 0.253, 0.288, 0.420, 0.144, 0.450, 0.190, 0.375, 0.383, 0.159, -0.230, -0.292, -0.230, 0.327, 0.415, 0.240, 0.155, 0.421$, respectively. In the deep aquifer, the

Spearman's correlation of physical parameters was correlated as shown in Table 13. Furthermore, total As has a significant correlation with Granite ($r=-0.172$) at the 0.01 level, as well as correlated with DEM Forest, Sandy soil, Quaternary colluvium aquifer at $r=0.161$, $r=-0.145$, $r=-0.150$ and $r=0.155$ at 0.05 level, respectively.

In shallow aquifer, the Spearman's correlation of physical parameters was correlated as shown in Table 14. Total As had a significant correlation with Granite ($r=-0.172$) at the 0.01 level, as well as correlated with DEM, Forest, Sandy soil, Quaternary colluvium aquifer at $r=0.161$, $r=-0.145$, $r=-0.150$, and $r=0.155$ at 0.05 level, respectively. Total As had a significant correlation at 0.01 level with DEM, Sandy soil, Clay soil, Granite, Qmc with $r=-0.193$, $r=-0.127$, $r=0.143$, $r=0.206$, $r=-0.152$, respectively, and had a significant correlation at 0.05 level with sand, Agricultural area (Field), and Quaternary colluvium at $r=0.097$, $r=-0.124$, $r=-0.111$, respectively.

In conclusion, the parameters that meet a significant correlation both in 0.05 and 0.01 levels were selected as the selected parameters for further modeling process as shown in Table 15.

Table 11 Spearman's correlation of hydrochemical parameters in the deep aquifer

	As	EC	pH	TDS	TTH	Ca	Mg	Na	K	Fe	Cl	F	HCO32	SO42	NO3	Cd	C6+	Cu	Hg	Mn	Ni	Pb	Se	Zn
As	1.000	.152**	.222**	.230**	.165**	.170**	.171**	0.096	0.008	0.093	-0.054	.183**	.291**	0.076	-0.109	-0.022	-0.089	-0.055	0.115	0.118	0.086	0.011	0.107	-0.064
EC	.152**	1.000	.517**	.612**	.731**	.710**	.595**	.642**	.281**	.190**	.526**	.279**	.676**	.428**	-0.058	-0.074	0.050	0.064	-0.006	.433**	-0.081	-0.034	.190**	-0.161*
pH	.222**	.517**	1.000	.358**	.395**	.429**	.234**	.373**	0.011	-0.083	0.116	.398**	.603**	.202**	0.091	0.001	-0.217**	-0.066	.232**	0.013	-0.108	.170**	.188**	-0.255**
TDS	.230**	.612**	.358**	1.000	.710**	.672**	.611**	.546**	.327**	0.042	0.500**	.213**	.621**	.350**	-0.157*	-0.080	0.104	0.003	-0.067	.313**	0.039	-0.087	0.061	-0.130*
TTH	.165**	.731**	.395**	.710**	1.000	.944**	.789**	.465**	.294**	0.096	.561**	.171**	.761**	.512**	-0.079	-0.079	0.127	0.025	-0.102	.514**	0.082	-0.084	0.093	-0.137*
Ca	.170**	.710**	.429**	.672**	.944**	1.000	.672**	.464**	.269**	0.065	.482**	.206**	.744**	.482**	-0.089	-0.079	0.086	0.007	-0.055	.486**	0.055	-0.017	.135**	-0.199**
Mg	.171**	.595**	.234**	.611**	.789**	.672**	1.000	.466**	.362**	.164**	.559**	0.037	.534**	.468**	-0.007	-0.033	0.110	0.009	-0.084	.452**	0.008	-0.094	0.094	-0.124
Na	0.096	.642**	.373**	.546**	.465**	.464**	.466**	1.000	.470**	0.115	.571**	.421**	.510**	.317**	-0.115	-0.034	0.071	0.102	-0.036	.262**	-0.045	-0.108	.138**	-0.069
K	0.008	.281**	0.011	.327**	.294**	.269**	.362**	.470**	1.000	0.041	.430**	0.032	0.089	.181**	0.117	0.015	0.020	0.125	0.003	.243**	0.128	-0.118	.129**	-0.052
Fe	0.093	.190**	-0.083	0.042	0.096	0.065	.164**	0.115	0.041	1.000	0.016	-0.090	0.029	-0.062	-0.235**	-0.002	0.116	0.110	-0.122	.508**	-0.132*	-0.132*	0.011	.306**
Cl	-0.054	.526**	0.116	.500**	.561**	.482**	.559**	.571**	.430**	0.016	1.000	0.069	.288**	.497**	0.070	-0.062	0.049	0.067	-0.057	.303**	0.058	-0.149*	.139**	-0.133*
F	.183**	.279**	.398**	.213**	.171**	.206**	0.032	.421**	0.032	-0.090	0.069	1.000	.442**	.118	-0.108	-0.036	-0.029	0.068	0.066	-0.045	-0.038	-0.029	-0.029	-0.042
HCO32	.291**	.676**	.603**	.621**	.761**	.744**	.534**	.510**	0.089	0.029	.288**	.442**	1.000	.334**	-0.144*	-0.066	0.103	-0.011	-0.089	.304**	-0.087	-0.066	-0.021	-0.154*
SO42	0.076	.428**	.202**	.428**	.428**	.428**	.428**	.428**	.428**	.428**	.428**	.428**	.428**	1.000	0.075	-0.078	0.047	-0.027	-0.025	.270**	-0.008	-0.119	0.070	-0.228**
NO3	-0.109	-0.058	0.091	-0.074	-0.034	-0.034	-0.034	-0.034	-0.034	-0.034	-0.034	-0.034	-0.034	0.075	1.000	0.088	-0.270**	0.092	.235**	-0.191**	0.106	.140**	.256**	-0.147*
Cd	-0.022	-0.074	0.001	-0.080	-0.079	-0.079	-0.079	-0.079	-0.079	-0.079	-0.079	-0.079	-0.079	-0.078	0.088	1.000	-0.124	-0.015	0.123	-0.028	0.042	-0.026	0.097	-0.085
Cr6+	-0.089	0.050	-0.217**	0.104	0.127	0.086	0.110	0.071	0.020	0.116	0.049	-0.029	0.103	0.047	-0.270**	-0.124	1.000	0.107	-0.951**	.328**	-0.204**	-0.551**	-0.769**	.366**
Cu	-0.055	0.064	-0.066	0.003	0.025	0.007	0.009	0.102	0.125	0.110	0.067	0.068	-0.011	-0.027	0.092	-0.015	0.107	1.000	-0.121	.150**	0.031	-0.091	-0.083	.228**
Hg	0.115	-0.006	.232**	-0.067	-0.102	-0.055	-0.084	-0.036	0.003	-0.122	-0.057	0.066	-0.089	-0.025	.235**	0.123	-0.951**	-0.121	1.000	.305**	.221**	.554**	.804**	-0.362**
Mn	0.118	.433**	0.013	.313**	.514**	.486**	.452**	.262**	.243**	.508**	.303**	-0.045	.304**	.270**	-0.191**	-0.028	.328**	.150**	-0.305**	1.000	-0.036	-0.199**	-0.089	0.117
Ni	0.086	-0.081	-0.108	0.039	0.082	0.055	0.008	-0.045	0.128	-0.132*	0.058	-0.038	-0.087	-0.008	0.106	0.042	-0.204**	0.031	.221**	1.000	.192**	.189**	.189**	0.049
Pb	0.011	-0.034	.170**	-0.087	-0.084	-0.017	-0.094	-0.108	-0.118	-0.132*	-0.149**	-0.029	-0.066	-0.119	.140**	-0.026	-0.551**	-0.091	.554**	-0.199**	.192**	1.000	.421**	-0.206**
Se	0.107	.190**	.188**	0.061	0.093	.135**	0.094	.138**	.129**	0.011	.139**	-0.029	-0.021	0.070	.256**	0.097	-0.769**	-0.083	.804**	-0.089	.189**	.421**	1.000	-0.311**
Zn	-0.064	-0.161*	-0.255**	-0.130*	-0.137*	-0.199**	-0.124	-0.069	-0.052	.306**	-0.133*	-0.042	-0.154*	-0.228**	-0.147*	-0.085	.366**	.228**	-0.362**	0.117	0.049	-0.206**	-0.311**	1.000

*. Correlation is significant at the 0.05 level (2-tailed).

** . Correlation is significant at the 0.01 level (2-tailed).

Table 12 Spearman's correlation of hydrochemical with shallow aquifer

	As	EC	pH	TDS	TTH	Ca	Mg	Na	K	Fe	Cl	F	HCO32	SO42	NO3	Cd	Cr6+	Cu	Hg	Mn	Ni	Pb	Se	Zn
As	1.000	.368**	.255**	.385**	.281**	.253**	.288**	.420**	.144**	.450**	.190**	.375**	.383**	.159**	-.230**	0.011	-.292**	-.230**	.327**	.415**	.240**	.155**	.421**	0.081
EC	.368**	1.000	.494**	.957**	.843**	.786**	.692**	.817**	.541**	.357**	.696**	.443**	.740**	.623**	-.146**	0.007	0.039	-.145**	0.005	.443**	.203**	-.0045	.274**	0.056
pH	.255**	.494**	1.000	.505**	.531**	.514**	.348**	.383**	.121**	.158**	.168**	.438**	.672**	.225**	-.179**	-.083	-.108**	-.116**	.127**	.185**	0.069	0.071	.191**	0.029
TDS	.385**	.957**	.505**	1.000	.860**	.799**	.710**	.839**	.560**	.362**	.720**	.452**	.749**	.647**	-.143**	0.008	0.038	-.142**	0.007	.459**	.176**	-.0051	.274**	0.047
TTH	.281**	.843**	.531**	.860**	1.000	.932**	.793**	.634**	.475**	.302**	.599**	.358**	.771**	.608**	-.114**	-.0054	0.034	-.141**	0.004	.471**	.187**	-.0049	.242**	0.034
Ca	.253**	.786**	.514**	.799**	.932**	1.000	.650**	.593**	.440**	.243**	.515**	.325**	.790**	.563**	-.042	0.003	0.003	-.110**	0.023	.453**	.154**	-.0026	.207**	0.032
Mg	.288**	.692**	.348**	.710**	.793**	.650**	1.000	.630**	.506**	.320**	.633**	.277**	.543**	.582**	-.0018	-.0054	0.030	-.142**	0.006	.404**	.188**	-.0014	.258**	0.044
Na	.420**	.817**	.383**	.839**	.634**	.593**	.630**	1.000	.549**	.387**	.761**	.437**	.617**	.597**	-.156**	-.0006	-.008	-.105**	0.052	.401**	.169**	-.0044	.313**	0.059
K	.144**	.541**	.121**	.560**	.475**	.440**	.506**	.549**	1.000	.240**	.641**	.132**	.284**	.463**	-.0040	-.0040	0.056	-.105**	0.021	.241**	.100**	-.148**	.239**	0.025
Fe	.450**	.357**	.158**	.362**	.302**	.243**	.320**	.387**	.240**	1.000	.301**	.203**	.304**	.191**	-.230**	-.010	-.121**	-.0092	.165**	.671**	.092	-.0054	.285**	.182**
Cl	.190**	.696**	.168**	.720**	.599**	.515**	.633**	.761**	.641**	.301**	1.000	.224**	.369**	.645**	-.0009	0.012	0.021	-.0035	0.026	.329**	.139**	-.0070	.291**	0.013
F	.375**	.443**	.438**	.452**	.358**	.325**	.277**	.437**	.132**	.203**	1.000	.224**	.369**	.188**	-.261**	-.0001	-.207**	-.201**	.214**	.211**	.184**	.163**	.290**	0.072
HCO32	.383**	.740**	.672**	.749**	.771**	.790**	.543**	.617**	.284**	.304**	.369**	.515**	1.000	.373**	-.246**	-.0083	-.044	-.143**	0.072	.423**	.104**	0.027	.175**	0.052
SO42	.159**	.623**	.225**	.647**	.608**	.563**	.582**	.597**	.463**	.191**	.645**	.188**	.373**	1.000	-.086	0.057	0.051	-.0014	-.023	.294**	.111**	-.103**	.210**	0.004
NO3	-.230**	-.146**	-.179**	-.143**	-.114**	-.116**	-.018	-.156**	.177**	-.230**	-.009	-.261**	-.246**	-.086	1.000	0.005	-.0057	.125**	0.050	-.232**	-.0055	-.0018	-.003	-0.061
Cd	0.011	0.007	-0.083	0.008	-0.054	-0.042	-0.054	-0.006	-0.040	-0.010	0.012	-0.001	-0.083	0.057	0.005	1.000	-0.081	-0.017	.098**	0.056	.102**	.139**	0.092	0.052
Cr6+	-.292**	-.145**	-.108**	-.142**	-.141**	-.110**	-.142**	-.105**	0.056	-.121**	0.021	-0.207**	-0.044	0.051	-0.057	-0.081	1.000	.137**	-.953**	0.026	-.311**	-.590**	-.689**	.097**
Cu	-.230**	-.145**	-.116**	-.142**	-.141**	-.110**	-.142**	-.105**	0.021	-0.092	-0.035	-0.201**	-.143**	0.051	-0.014	-.017	.137**	1.000	-.115**	-.141**	-.111**	-.120**	-.133**	0.024
Hg	.327**	0.005	.127**	0.007	0.004	0.023	0.006	0.052	-0.011	.165**	0.026	.214**	0.072	-.0023	0.050	.098**	-.953**	-.115**	1.000	-.0005	.363**	.576**	.762**	-.097**
Mn	.415**	.443**	.185**	.459**	.471**	.453**	.404**	.401**	.241**	.671**	.329**	.211**	.423**	.294**	-.232**	0.056	0.026	-.141**	-.0005	1.000	0.091	-.103**	.110**	.126**
Ni	.240**	.203**	0.069	.176**	.187**	.154**	.188**	.169**	.100**	0.092	.139**	.184**	.104**	.111**	-.0055	.102**	-.311**	-.111**	.363**	0.091	1.000	.317**	.456**	.157**
Pb	.155**	-.0045	0.071	-0.051	-0.049	-0.026	-0.014	-0.044	-.148**	-0.054	-0.070	.163**	0.027	-.103**	-.0018	.139**	-.590**	-.120**	.576**	-.103**	.317**	1.000	.408**	-0.007
Se	.421**	.274**	.191**	.274**	.242**	.207**	.258**	.313**	.239**	.285**	.291**	.290**	.175**	.210**	-.0003	0.092	-.689**	-.133**	.762**	.110**	.456**	.408**	1.000	-0.042
Zn	0.081	0.056	0.029	0.047	0.034	0.032	0.044	0.059	0.025	.182**	0.013	0.072	0.052	0.004	-0.061	0.052	.097**	0.024	-.097**	.126**	-.157**	-.007	-0.042	1.000

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Table 13 Spearman's correlation of physical parameters in the deep aquifer

	As	DEM	Gravel	Sand	Sandy	Silt	Clay	Urban	Forest	Mine	Water	Arg(Tree)	Arg(Field)	Cpms	Gr	Qa	Qis	Qcl	Qmc
As	1.000	.161*	-0.052	0.111	-0.145*	0.088	0.013	-0.087	-0.150*	0.114	0.034	0.107	0.015	0.059	-0.172**	-0.045	-0.041	.155*	-0.030
DEM	.161*	1.000	0.091	-0.036	.217**	-0.301**	-0.109	-0.073	-0.082	-0.136*	0.054	-0.258**	.239**	-0.533**	-0.365**	0.009	0.019	.773**	-0.104
Gravel	-0.052	0.091	1.000	-0.141*	-0.118	-0.034	-0.069	-0.075	-0.021	-0.032	-0.018	-0.032	0.102	-0.075	-0.115	.768**	-0.024	-0.136*	-0.015
Sand	0.111	-0.036	-0.141*	1.000	-0.637**	-0.184**	-0.370**	.131*	0.018	0.005	0.054	-0.174**	-0.058	-0.115	.251**	-0.099	-0.128*	-0.058	-0.081
Sandy	-0.145*	.217**	-0.118	-0.637**	1.000	-0.153*	-0.310**	0.009	-0.096	-0.145*	-0.083	.226**	0.006	-0.037	-0.216**	-0.065	-0.107	.270**	0.127
Silt	0.088	-0.301**	-0.034	-0.184**	-0.153*	1.000	-0.089	-0.098	-0.028	.507**	-0.024	0.068	-0.147*	.342**	-0.060	-0.044	-0.031	-0.177**	-0.019
Clay	0.013	-0.109	-0.069	-0.370**	-0.310**	-0.089	1.000	-0.105	0.127	-0.084	0.057	-0.084	0.110	0.049	0.024	-0.089	.347**	-0.118	-0.039
Urban	-0.087	-0.073	-0.075	.131*	0.009	-0.098	-0.105	1.000	-0.061	-0.093	-0.053	-0.093	-0.739**	-0.072	.163*	0.122	-0.068	-0.122	-0.043
Forest	-0.150*	-0.082	-0.021	0.018	-0.096	-0.028	0.127	-0.061	1.000	-0.026	-0.015	-0.026	-0.209**	-0.061	0.046	-0.028	.437**	-0.111	-0.012
Mine	0.114	-0.136*	-0.032	0.005	-0.145*	.507**	-0.084	-0.093	-0.026	1.000	-0.023	-0.040	-0.316**	.254**	-0.141*	-0.042	-0.029	-0.033	-0.018
Water	0.034	0.054	-0.018	0.054	-0.083	-0.024	0.057	-0.053	-0.015	-0.023	1.000	-0.023	-0.180**	-0.053	-0.080	-0.024	-0.017	.135*	-0.010
Arg(Tree)	0.107	-0.258**	-0.032	-0.174**	.226**	0.068	-0.084	-0.093	-0.026	-0.040	-0.023	1.000	-0.316**	.428**	-0.141*	-0.042	-0.029	-0.168**	-0.018
Arg(Field)	0.015	.239**	0.102	-0.058	0.006	-0.147*	0.110	-0.739**	-0.209**	-0.316**	-0.180**	-0.316**	1.000	-0.198**	-0.011	-0.054	-0.038	.187**	0.058
Qpms	0.059	-0.533**	-0.075	-0.115	-0.037	.342**	0.049	-0.072	-0.061	.254**	-0.053	.428**	-0.198**	1.000	-0.330**	-0.098	-0.068	-0.392**	-0.043
Gr	-0.172**	-0.365**	-0.115	.251**	-0.216**	-0.060	0.024	.163*	0.046	-0.141*	-0.080	-0.141*	-0.011	-0.330**	1.000	-0.149*	-0.104	-0.598**	-0.066
Qa	-0.045	0.009	.768**	-0.099	-0.065	-0.044	-0.089	0.122	-0.028	-0.042	-0.024	-0.042	-0.054	-0.098	-0.149*	1.000	-0.031	-0.177**	-0.019
Qis	-0.041	0.019	-0.024	-0.128*	-0.107	-0.031	.347**	-0.068	.437**	-0.029	-0.017	-0.029	-0.038	-0.068	-0.104	-0.031	1.000	-0.124	-0.014
Qcl	.155*	.773**	-0.136*	-0.058	.270**	-0.177**	-0.118	-0.122	-0.111	-0.033	.135*	-0.168**	.187**	-0.392**	-0.598**	-0.177**	-0.124	1.000	-0.078
Qmc	-0.030	-0.104	-0.015	-0.081	0.127	-0.019	-0.039	-0.043	-0.012	-0.018	-0.010	-0.018	0.058	-0.043	-0.066	-0.019	-0.014	-0.078	1.000

*. Correlation is significant at the 0.05 level (2-tailed).

** . Correlation is significant at the 0.01 level (2-tailed).

Table 14 Spearman's correlation of physical parameters in the shallow aquifer

	As	DEM	Gravel	Sand	Sandy	Silt	Clay	Urban	Forest	Mine	Water	Arg(Tree)	Arg(Field)	Cpms	Gr	Qa	Qis	Qcl	Qmc
As	1.000	-.193**	-0.049	.097*	-.127**	0.066	.143**	-0.064	0.082	0.016	0.059	0.047	-.124*	0.002	.206**	-0.064	-0.036	-.111*	-.152**
DEM	-.193**	1.000	0.063	-.107*	.167**	-.308**	-0.054	.126**	-.223**	-0.043	-.267**	-.264**	.445**	-.622**	-.144**	-0.034	0.044	.116*	.653**
Gravel	-0.049	0.063	1.000	-.139**	-.102*	-0.037	-0.029	-0.067	-0.078	.192**	-0.034	-0.034	0.038	-0.090	-.109*	-0.011	.911**	-0.019	-.122*
Sand	.097*	-.107*	-.139**	1.000	-.575**	-.209**	-.165**	-.377**	.164**	0.002	0.085	-.193**	-0.093	0.087	-0.007	0.079	-.152**	-.107*	-0.003
Sandy	-.127**	.167**	-.139**	-.575**	1.000	-.154**	-.121*	-.277**	-0.062	-.097*	-0.092	-0.042	.143**	-0.056	-0.039	-0.045	-0.081	.098*	.098*
Silt	0.066	-.308**	-0.037	-.209**	-.154**	1.000	-0.044	-0.100*	-.118*	-0.035	.206**	.720**	-.304**	.362**	-.118*	-0.016	-0.041	-0.029	-.184**
Clay	.143**	-0.054	-0.029	-.165**	-.121*	-0.044	1.000	-0.079	.107*	-0.028	-0.041	-0.041	-0.047	-.107*	.240**	-0.013	0.048	-0.023	-.146**
Urban	-0.064	.126**	-0.067	-.223**	-.078	-.079	-0.079	1.000	-0.096	0.074	-0.093	-0.093	.143**	-.182**	0.059	-0.030	-0.073	0.060	.120*
Forest	0.082	-.223**	0.091	-.223**	1.000	-.118*	.107*	-0.096	1.000	-0.074	-.109*	-.109*	-.741**	0.091	.127**	.139**	-0.014	-0.060	-.204**
Mine	0.016	-0.043	.192**	0.002	-.097*	-0.035	-0.028	0.074	-0.074	1.000	-0.032	-0.032	-.221**	-0.085	0.073	-0.010	.172**	.259**	-.116*
Water	0.059	-.267**	0.059	0.085	-.109*	-.109*	-0.041	-0.093	-.109*	-0.032	1.000	-0.048	-.325**	.328**	-.103*	-0.015	-0.038	-0.026	-.171**
Arg(Tree)	0.047	-.264**	0.047	-.264**	1.000	-.325**	-0.041	-0.093	-.109*	-0.032	-0.048	1.000	-.325**	.381**	-.152**	-0.015	-0.038	-0.026	-.171**
Arg(Field)	-.124*	.445**	0.038	-0.093	.143**	-.304**	-0.047	.143**	-.741**	-.221**	-.325**	-.325**	1.000	-.371**	-0.018	-.103*	-0.008	-0.006	.366**
Qpms	0.002	-.622**	-0.090	0.087	-0.056	.362**	-.107*	-.182**	0.091	-0.085	.328**	.381**	-.371**	1.000	-.398**	-0.040	-.099*	-0.069	-.447**
Gr	.206**	-.144**	-0.007	-0.007	-0.039	-.118*	.240**	0.059	.127**	0.073	-0.103*	-.152**	-0.018	-.398**	1.000	-0.048	-0.120*	-0.084	-.543**
Qa	-0.064	-0.034	-0.011	0.079	-0.045	-0.016	-0.013	-0.030	.139**	-0.010	-0.015	-0.015	-.103*	-0.040	-0.048	1.000	-0.012	-0.008	-0.054
Qis	-0.036	0.044	.911**	-.152**	-0.081	-0.041	0.048	-0.073	-0.014	.172**	-0.038	-0.038	-0.008	-0.099*	-.120*	-0.012	1.000	-0.021	-.134**
Qcl	-.111*	.116*	-0.019	-.107*	.098*	-0.029	-0.023	0.060	-0.060	.259**	-0.026	-0.026	-0.006	-0.069	-0.084	-0.008	-0.021	1.000	-0.094
Qmc	-.152**	.653**	-.122*	-0.003	.098*	-.184**	-.146**	.120*	-.204**	-.116*	-.171**	-.171**	.366**	-.447**	-.543**	-0.054	-0.021	-0.094	1.000

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Table 15 Spearman's correlation with physical and hydrochemical parameters in both deep and shallow aquifers

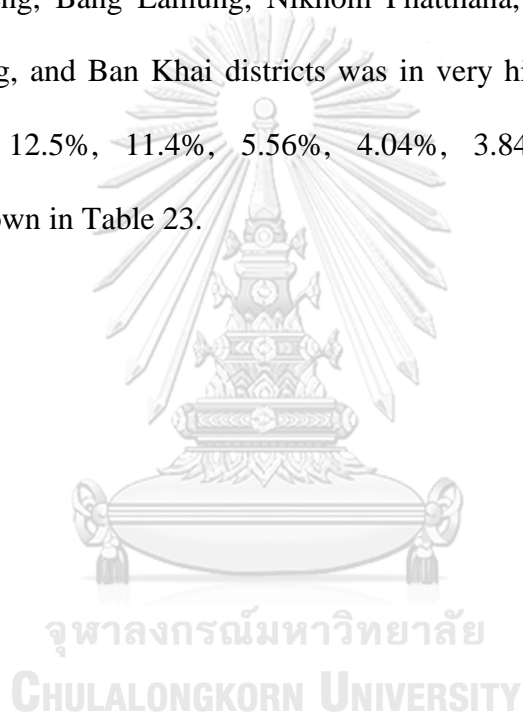
Parameters	Deep aquifer	Shallow aquifer
Physical parameters	DEM, Silt loam, Forest, Quaternary, Granite	DEM, Sandy loam, Silt loam, Clay loam, Agricultural(fields), Gr, Qc, Qmc
Chemical parameters	EC, pH, TDS, TTH, Ca, Mg, F, HCO ₃	EC, pH, TDS, TTH, Ca, Mg, Na, K, Fe, Cl, F, HCO ₃ , SO ₄ , NO ₃ , Cd, Cr, Cu, Hg, Mn, Ni, Pb, Se

4.4 Probability map

The selected models including RF, SVM, ANN were used to generate a probability map with the deep and shallow aquifers by the sdm and neutral net packages in R-studio software. The probability map shows the probability area of total As concentration exceeding the standard of drinking water (Figure 17). The result of the probability map showed RF had fine distribution probability areas. Also, the results of the SVM and ANN map had roughly probability distribution areas compared to RF. To evaluate a high probability area, each probability map was classified into 4 probability levels as follows: very high, high, moderates, and low.

For the deep aquifer, the SVM probability map showed district areas, consisting of Si Racha, Bang Lamung, Pluak Daeng, Nikhom Phatthana, and Sattahip, was classified in the very high probability area of 48.5%, 35.21%, 11.54%, 2.84%, and 1.49%, respectively, as shown in Table 21. Moreover, other district areas such as Ban Khai, Nong Yai, Ban Chang, Ban Bueng, and Mueng Rayong were classified in the high-level areas with 12.82%, 4.0%, 3.97%, 3.69%, and 2.86%, respectively. The RF probability map, in the deep aquifer, showed district areas, which are Pluak

Daeng, Bang Lamung, and Si Racha in the very high-level probability areas of 40%, 40%, and 20%, respectively. Other district areas including Ban Khai, Nikhom Phatthana, Ban Bueng, Mueang Rayong, Ban Chang were in the high-level areas of 29.76%, 9.01%, 4.14%, 0.23%, and 0.06%, respectively as shown in Table 22. Lastly, Sattahip and Nong Yai districts were in the moderate level areas of 5.36% and 2.76%, respectively. The ANN probability map in the deep aquifer of Mueang Rayonh, Si Racha, Pluak Daeng, Bang Lamung, Nikhom Phatthana, Sattahip, Nong Yai, Ban Chang, Ban Bueng, and Ban Khai districts was in very high-level areas of 21.06%, 17.9%, 17.69%, 12.5%, 11.4%, 5.56%, 4.04%, 3.84%, 2.98%, and 2.15%, respectively as shown in Table 23.



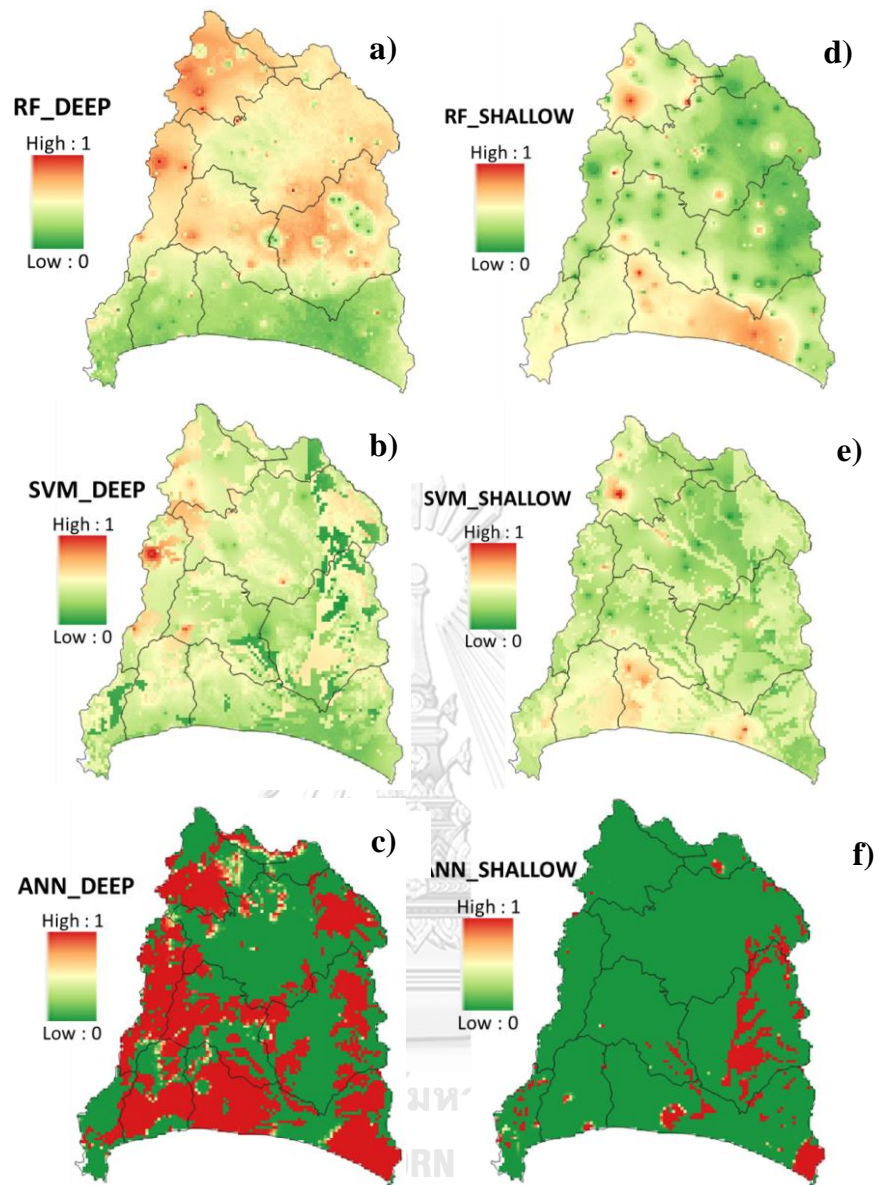


Figure 16 The probability map in the deep aquifer derived from a) RF, b) SVM, c) ANN and the probability map in the shallow aquifer derived from d) RF, e) SVM and f) ANN

Table 16 Percentage probability of each district derived from SVM model in the deep aquifer

Districts	Very high (%)	High (%)	Moderate (%)	Low (%)
Pluak Daeng	11.54	26.1	24.57	8.45
Mueang Rayong	0	2.86	22.47	43.51
Nong Yai	0	4.09	1.71	0
Ban Chang	0	3.97	9.69	3.44
Ban Khai	0	12.82	18.57	28.79
Sattahip	1.49	1.76	5.05	10.96
Si Racha	48.5	17.58	3.18	0
Ban Bueng	0	3.69	0.41	0
Nikhom Phatthana	2.87	10.19	11.93	4.29
Bang Lamung	35.21	16.29	1.54	0

Table 17 Percentage probability of each district derived from RF model in the deep aquifer

Districts	Very high (%)	High (%)	Moderate (%)	Low (%)
Pluak Daeng	40	12.38	25.98	0
Mueang Rayong	0	0.23	22.02	0
Nong Yai	0	0	2.76	0
Ban Chang	0	0.06	8.96	0
Ban Khai	0	29.76	13.82	0
Sattahip	0	0	5.36	0
Si Racha	20	31.37	3.17	0
Ban Bueng	0	4.14	0.67	0
Nikhom Phatthana	0	9.01	10.99	0
Bang Lamung	40	12	5.55	0

Table 18 Percentage probability of each district derived from ANN model in the deep aquifer

Districts	Very high (%)	High (%)	Moderate (%)	Low (%)
Pluak Daeng	17.69	36.97	29.96	25.16
Mueang Rayong	21.06	24.51	30.66	15.65
Nong Yai	4.04	2.96	0.25	1.55
Ban Chang	3.84	16.3	11.23	8.27
Ban Khai	2.15	4.97	12.08	23.51
Sattahip	5.56	6.18	1.66	3.98
Si Racha	17.9	2.18	1.15	5.19
Ban Bueng	2.98	0	0.11	0.72
Nikhom Phatthana	11.4	3.94	8.58	10.5
Bang Lamung	12.5	1.57	4.24	4.67

For the shallow aquifer, the SVM probability map consisting of Mueang Rayong, Si Racha, and Nikhom Phatthana districts were classified in very high-level areas of 57.69%, 38.46%, and 3.85%, respectively, as shown in Table 24. Other districts including Sattahip, Ban Khai, Bang Lamung, and Pluak Deang were categorized in the high-level area of 3.69%, 0.58%, 0.38%, and 0.37%, respectively. Besides, Ban Chang, Nong Yai, and Ban Bueng districts were classified in the moderate level of 8.42%, 2.08%, and 1.33%, respectively.

In Table 25, based on the RF probability map in the shallow aquifer, Mueang Rayong, Si Racha, Pluak Deang, Nikhom Phatthana, Ban Khai, Bang Lamung were defined as the very high-level area, which was of 31.62%, 30.65%, 12.67%, 12.41%, 6.33%, 6.33%, respectively. Furthermore, other areas including Ban Chang and Ban Bueng were grouped in the high-level area of 0.38% and 0.16%, respectively. The moderate level including Sattahip and Nong Yai districts is 5.08% and 2.18%, respectively. The ANN probability map in the shallow aquifer was shown in Table 26, which shows districts with the very high-level area consisting of Mueang Rayong, Pluak Daeng, Ban Khai, Si Racha, Sattahip, Ban Chang, Bang Lamung, Nikhom Phatthana, Nong Yai, and Ban Bueng districts of 25.81%, 20.37%, 16.26%, 10.66%, 6.79%, 5.09%, 4.87%, 4.44%, 3.1%, and 1.92%, respectively.

Table 19 Percentage probability of each district derived from SVM model in the shallow aquifer

Districts	Very high (%)	High (%)	Moderate (%)	Low (%)
Pluak Daeng	0	0.38	21.1	51.07
Mueang Rayong	57.69	83.35	15.65	14.43
Nong Yai	0	0	2.08	4.68
Ban Chang	0	0	8.42	0.27
Ban Khai	0	0.58	16.95	19.85
Sattahip	0	3.69	4.8	0.88
Si Racha	38.46	11.06	9.33	1.5
Ban Bueng	0	0	1.33	2
Nikhom Phatthana	3.85	0.37	11.94	2.87
Bang Lamung	0	0.37	7.58	2.3

Table 20 Percentage probability of each district derived from RF model in the shallow aquifer

Districts	Very high (%)	High (%)	Moderate (%)	Low (%)
Pluak Daeng	12.67	2.16	26.16	37.74
Mueang Rayong	31.62	66.6	10.97	4.77
Nong Yai	0	0	2.18	17.83
Ban Chang	0	0.38	8.45	0.01
Ban Khai	6.33	2.45	19	15.64
Sattahip	0	0	5.08	0
Si Racha	30.65	19.57	7.05	5.91
Ban Bueng	0	0.16	1.48	2.33
Nikhom Phatthana	12.41	8.13	11.08	4.43
Bang Lamung	6.33	0	7.69	11.31

Table 21 Percentage probability of each district derived from ANN model in the shallow aquifer

Districts	Very high (%)	High (%)	Moderate (%)	Low (%)
Pluak Daeng	20.37	8.67	13.75	27.28
Mueang Rayong	25.81	0.89	1.48	11.72
Nong Yai	3.1	10.2	6.3	0.92
Ban Chang	5.09	2.66	5.23	9.6
Ban Khai	16.26	12.3	12.74	17.73
Sattahip	6.79	6.84	2.19	2.08
Si Racha	10.66	28.1	12.17	5.95
Ban Bueng	1.92	0	0	0.89
Nikhom Phatthana	4.44	14.92	21.25	15.59
Bang Lamung	4.87	12.72	24.75	7.35

4.5 Performance of models

4.5.1 Prediction performance

The prediction performance of the models was evaluated using AUC and RMSE measurements which were shown in Tables 22 and 23. The goodness of fitting (Training dataset) for the deep aquifer revealed that. an ANN model had AUC=1 and RMSE=5E-5, following by SVM (AUC=0.73, RMSE=0.48) and RF (AUC=0.93, RMSE=0.34). In the shallow aquifer, the goodness of fitting from ANN had AUC=1 and RMSE=7.3E-5, following by SVM (AUC=0.84, RMSE=0.41) and RF (AUC=0.93, RMSE=0.33).

Furthermore, to evaluate the appropriate model, the predictive performance (Test dataset) needs to be mainly considered (Bindal and Singh 2019). In deep aquifer, RF (AUC=0.72, RMSE=0.48) model was the best model as compared with SVM (AUC=0.69, RMSE=0.49) and ANN (AUC=0.65, RMSE=0.62). In addition, in the shallow aquifer, RF (AUC=0.81, RMSE=0.42) also was the best model following by SVM (AUC=0.79, RMSE=0.44) and ANN (AUC=0.78, RMSE=0.47). The visualization of the models' performance was supported by using the Taylor diagram (Taylor 2001) as shown in Figure 16. The RF had a higher correlation with observed total As probability and had lower RMSE compared to those of SVM and ANN in both the deep and shallow aquifers.

Table 22 Performances of SVM, RF, and ANN in the deep aquifer

Model	Train		Test	
	RMSE	AUC	RMSE	AUC
SVM	0.48	0.73	0.49	0.69
RF	0.34	0.93	0.48	0.72
ANN	5E-05	1.00	0.62	0.65

Table 23 Performances of SVM, RF, and ANN in the shallow aquifer

Model	Train		Test	
	RMSE	AUC	RMSE	AUC
SVM	0.41	0.84	0.44	0.79
RF	0.33	0.93	0.42	0.81
ANN	7.3E-05	1	0.53	0.75

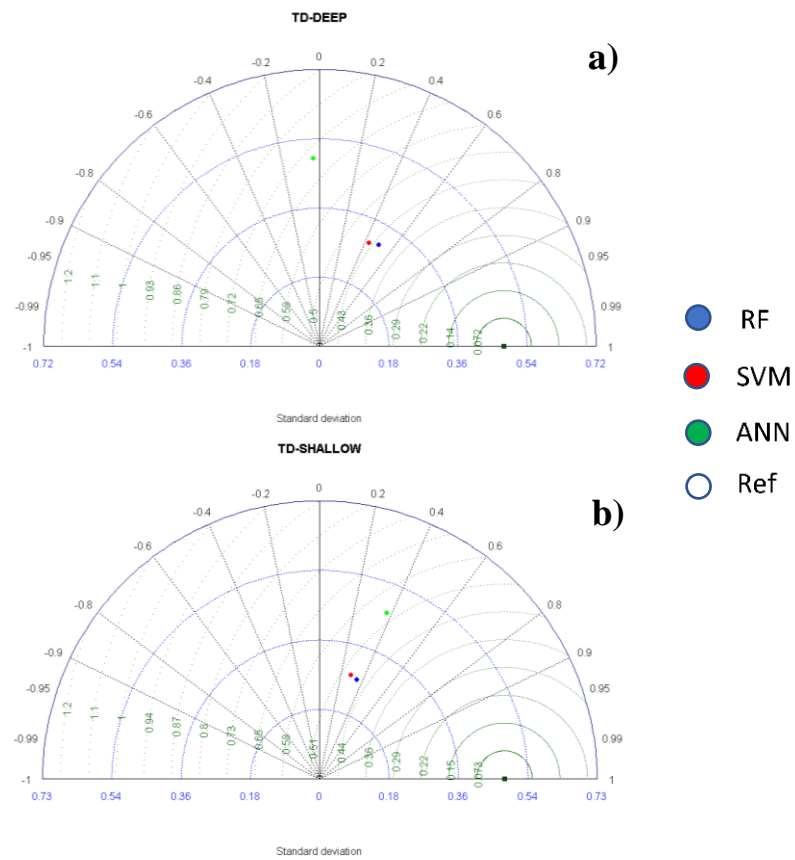


Figure 17 Taylor's diagram for a) the deep aquifer and b) the shallow aquifer

4.5.2 Models uncertainty

The uncertainty results of training and test groups in the deep aquifer were shown in Table 24. The training groups showed uncertainty values of three models, which were measured by PICP and MPI indicators, revealing the uncertainty of the individual model in the descending order as follows: SVM (PICP=0.27, MPI=0.16),

RF (PICP=0.53, MPI=0.49), and ANN (PICP=0.69, MPI=1.3E-4). For testing data, the lowest uncertainty model was RF (PICP=0.20, MPI=0.13), following by SVM (PICP=0.16, MPI=0.15), and ANN (PICP=0.05, MPI=6.075E-10).

The uncertainty results of training and test groups in the shallow aquifer were shown in Table 25. Again, the training group showed uncertainty values of three models, which were measured by PICP and MPI indicators, indicating the uncertainty of the individual model in the descending order as follows: SVM (PICP=0.50, MPI=0.37), RF (PICP=0.58, MPI=0.44) and ANN (PICP=0.86, MPI=8.7E-5). For the testing data, the lowest uncertainty model was RF (PICP=0.34, MPI=0.27) following by ANN (PICP=0.25, MPI=1.63E-10), and SVM (PICP=0.23, MPI=0.39).

Table 24 Uncertainty analysis of SVM, RF, and ANN of the deep aquifer

Models	Train		Test	
	PICP	MPI	PICP	MPI
SVM	0.27	0.16	0.16	0.15
RF	0.53	0.49	0.20	0.13
ANN	0.69	1.3E-4	0.05	6.075E-10

Table 25 Uncertainty analysis of SVM, RF, and ANN of the shallow aquifer

Models	Train		Test	
	PICP	MPI	PICP	MPI
SVM	0.50	0.37	0.23	0.39
RF	0.58	0.44	0.34	0.27
ANN	0.86	8.7E-05	0.25	1.63E-10

4.5.3 Validation with field data

To evaluate performance between the deep and shallow aquifers of probability map, total As concentration collected in the field during 31 August – 1 September 2019 were used to compare between the actual value and the prediction value from the probability map. The validation results were shown in Table 26, which revealed RMSE values for probability map of deep and shallow aquifer. The deep aquifer

showed that RF had the lowest RMSE (0.48) compared with SVM (RMSE=0.73) and ANN (RMSE=0.67). Similarly, RF also had the lowest RMSE (0.75) compared to SVM (RMSE=0.76) and ANN (RMSE=0.84).

Table 26 Validation model with field data during 31 August – 1 September 2019

	RMSE		
	SVM	RF	ANN
Deep	0.73	0.48	0.67
Shallow	0.76	0.75	0.84

4.6 Groundwater pollution risk map

The groundwater pollution risk map can be generated by using the RF probability map (cut-off value for deep and shallow aquifers were 0.589 and 0.596, respectively) multiply with the population density and water consumption map in the study area. Then, the risk maps were classified by an equal interval method into five categories as follows: very low, low, moderate, high, and very high. Total As shown in Table 27, the percentage of groundwater pollution risk area in the deep aquifer had areas in very high-level including Sattahip and Ban Khai districts of 71.9% and 28.1%, respectively. Ban Chang district was classified as in the high-level area (15.6%). The district classified as the moderate level included Si Racha, Bang Lamung, Mueang Rayong, Pluak Daeng, Nikhom Phatthana of 73.89%, 16.87%, 3.48%, 0.57%, and 0.13%, respectively. The very low-level districts were Nong Yai and Ban Bueng (2.77% and 1.52%, respectively).

The Percentage of groundwater pollution risk area in the deep aquifer was shown in Table 28, revealing that Mueang Rayong and Ban Khai districts were classified in the very high-risk level of 99.27% and 0.73%, respectively. For the high-

level area, there was only Nikhom Phatthana covering 3.65%. There are two districts classified as the moderate level consisting of Si Racha and Pluak Daeng of 41.73% and 0.29%, respectively. Ban Chang district was grouped in the low-level area (0.94%). Lastly, Bang Lamung, Sattahip, Nong Yai, and Ban Bueng districts areas were classified in the very low-level of 8.07%, 4.99%, 2.54%, and 1.40%, respectively.

Table 27 Percentage of groundwater pollution risk area in the deep aquifer

Districts	Very high	High	Moderate	Low	Very low
Pluak Daeng	0.00	0.00	0.57	32.76	22.57
Mueang Rayong	0.00	0.00	3.48	0.10	22.31
Nong Yai	0.00	0.00	0.00	0.00	2.77
Ban Chang	0.00	15.58	0.00	0.79	9.14
Ban Khai	28.13	77.26	5.06	26.14	14.89
Sattahip	71.87	7.16	0.00	0.02	5.03
Si Racha	0.00	0.00	73.89	5.94	7.44
Ban Bueng	0.00	0.00	0.00	0.00	1.52
Nikhom Phatthana	0.00	0.00	0.13	33.39	6.23
Bang Lamung	0.00	0.00	16.87	0.87	8.10

Table 28 Percentage of groundwater pollution risk area in the shallow aquifer

Districts	Very high	High	Moderate	Low	Very low
Pluak Daeng	0.00	0.00	0.29	9.29	26.52
Mueang Rayong	99.27	91.62	55.22	29.76	10.66
Nong Yai	0.00	0.00	0.00	0.00	2.54
Ban Chang	0.00	0.00	0.00	0.94	8.54
Ban Khai	0.73	4.73	1.22	0.94	18.80
Sattahip	0.00	0.00	0.00	0.00	4.99
Si Racha	0.00	0.00	41.73	16.95	8.35
Ban Bueng	0.00	0.00	0.00	0.00	1.40
Nikhom Phatthana	0.00	3.65	1.55	42.12	11.23
Bang Lamung	0.00	0.00	0.00	0.00	8.07

Chapter 5 Discussions

5.1 Parameters associated with As mechanisms in groundwater

The As concentration that exceeds the drinking water standard of 10 $\mu\text{g/l}$ (WHO 2018) is a common problem in groundwater supply wells. The natural As sources in groundwater mostly come from the weathering of certain rock types in the region and release into groundwater. For example, in a Quaternary aquifer such as Qc1 and Qa have As concentration around 27.50 and 17.19 mg/l, respectively. For granite base rock, it has As a concentration approximately 16.69 mg/l (Acharyya, Shah et al. 2005).

The other sources of As in groundwater can be from anthropogenic activities such as precipitation and water infiltrating through municipal waste in landfills that was contaminated with various organic and inorganic substances from the municipal waste and it contains As concentration around 0.004 mg/l (Wexler and Maus 1988). The result of the previous studies indicates that the sources of As in groundwater probably came from both natural and anthropogenic sources (Garelick, Jones et al. 2008, Shankar, Shanker et al. 2014, Boonkaewwan, Sonthiphand et al. 2020)

Therefore, to screening unnecessary parameters before using them in the modeling process, the spearman's correlation was used in this process. For making sure the spearman's correlation selected useful parameters, the correlation between other parameters and total As concentration in groundwater had to be toughly considered. In the deep aquifer, the strong positive of total As correlated with pH and bicarbonate (HCO_3) indicates the reducing environment in the groundwater system. This was in agreement with previous studies which revealed high total As

concentrations in groundwater were positively correlated with high bicarbonate (HCO_3) concentrations in the reducing condition in the groundwater environment (Nickson, McArthur et al. 2000, Charlet, Chakraborty et al. 2007). Ca had a strong positive correlation with Mg ($r=0.672$) and HCO_3 ($r=0.744$) implying the dissolution of calcite from geological formations into groundwater. Also, the good correlation between Mg and Na ($r=0.466$) indicates that ion exchange happened in groundwater (Sae-Ju, Chotpantararat et al. 2019). TDS was strongly correlate with EC ($r=0.612$), Na ($r=0.546$), Cl ($r=0.50$), SO_4 ($r=0.350$), Mg ($r=0.611$) and Ca ($r=0.672$). Similarly, the EC was positive correlation with Na ($r=0.642$), Cl ($r=0.526$), SO_4 (0.428), Mg ($r=0.595$) and Ca ($r=0.710$). These good correlations with the EC indicate that the increase in salinity was caused by groundwater mineralization and seawater intrusion (Sae-Ju, Chotpantararat et al. 2019). These results can confirm that the deep aquifer in the Rayong groundwater basin was influenced by seawater intrusion.

For the shallow aquifer, Ca also had a strong positive correlation with Mg ($r=0.650$) and HCO_3^- ($r=0.790$) indicating the dissolution of calcite in aquifers. Also, the correlation between Mg and Na ($r=0.630$) indicates that ion exchange was occurring during seawater intrusion. TDS was strongly correlate with EC ($r=0.957$), Na ($r=0.839$), Cl ($r=0.720$), SO_4 ($r=0.647$), Mg ($r=0.710$) and Ca ($r=0.799$). Similarly, the EC was positive correlation with Na ($r=0.817$), Cl ($r=0.696$), SO_4 (0.623), Mg ($r=0.692$) and Ca ($r=0.786$). These good correlations with the EC indicate that the increase in salinity was caused by groundwater mineralization and seawater intrusion (Sae-Ju, Chotpantararat et al. 2019). These results can confirm that the shallow aquifer in the Rayong groundwater basin also influent by seawater intrusion. Furthermore, arsenic had a positive correlation with pH and a negative correlation with NO_3 ,

implying that reducing conditions might cause the release of As in the shallow aquifer in the Rayong groundwater basin. From the previous study, the reduction environment was mainly caused by the hydrogeological characteristic, and through a denitrification process, make As released into groundwater (Boonkaewwan, Sonthiphand et al. 2020). Usually, the majority of As in groundwater was in oxidation form including As-rich Fe oxyhydroxide (FeOH) and As-bearing pyrite that exists as a coating on grain particle in sedimentary grain in the oxidation process (Thornton 1996, Bowen, Alpers et al. 2014). However, in the reducing environment which was derived by microbial degradation of organic matter in groundwater consumed O₂ and NO₃ in the process while release As into groundwater, this was the major event that impacts directly to As concentration in groundwater (Nickson, McArthur et al. 2000).

However, anthropogenic activities, such as agriculture, urbanized wastewater, livestock, industrial estates, and municipal landfill sites, some organic pollutants and/or organic acid from the ground surface leaches into the groundwater environment, enhancing the reducing conditions with a high correlation with SO₄ and NO₃. Therefore, from the correlation between SO₄ and NO₃ with negative correlation ($r = -0.086$) which mean when SO₄ release from organic matter was increased, the consumption of NO₃ by microorganism was also increased this can imply the source of reducing condition in the shallow aquifer came from anthropogenic activities (Boonkhao, Phanprasit et al. 2017).

The physical parameters influencing total As concentrations in groundwater are mainly from hydrogeological characteristics, soil types, and agricultural area (Tables 14 and 15). For examples, in shallow aquifer total As was mainly associated to sandy soil ($r = -0.127$), clay($r = 0.143$), granite($r = 0.206$) and quaternary alluvium

($r = -0.152$). From the correlation, we can indicate that bedrock and soil were the major sources of As release into the groundwater environment. Moreover, the correlation with the agricultural ($r = -0.124$) area might indicate the anthropogenic activities from agriculture in the area. However, although the physical parameters can be screened by this technique, it doesn't help much in terms of explaining the mechanism of As release in groundwater. Thus, as mention above we can conclude that Spearman's correlation can be used as a screening parameter tool, which is efficient to enhance the modeling process by eliminating unnecessary parameters out of datasets. Furthermore, the influence factors that affected an As mechanism in the study area might come from both natural and anthropogenic sources.

5.2 Evaluating prediction performance

The goodness-of-fit and predictive performance of the models were also quantified using RMSE and AUC measurements. In the training step of the deep aquifer, the ANN model produced an overfitting goodness-of-fit of total As probability with RMSE=5E-5 and AUC=1.00, following by SVM (RMSE=0.48, AUC=0.73) and RF (RMSE=0.34, AUC=0.93). This also happened in the shallow aquifer as shown in the descending order: ANN (RMSE=7.3E-5, AUC=1), SVM (RMSE=0.41, AUC=0.84), RF (RMSE=0.33, AUC=0.93). Based on the goodness-of-fit result, ANN shows the sign of overfitting with very small RMSE and high AUC compared to those of other models. Commonly, the overfitting in the ANN model is found when ANN algorithms work with a small dataset (Rao, Prasad et al. 2018). Thus, we can indicate that ANN was not an appropriate model using for the modeling process in this study.

The predictive performance of the model as shown in Tables 17 and 18 had the result of RF with RMSE=0.42 and AUC=0.72 following by SVM (RMSE=0.44, AUC=0.79) and ANN (RMSE=0.53, AUC=0.75). The goodness-of-fit of the model shows how well the model fits the training dataset. The prediction and generalization abilities of the model cannot be evaluated using only the goodness-of-fit of the model because it was measured by the data that were used to calibrate the model (Henseler and Sarstedt 2013). Therefore, the predictive performance (the accuracy of the model in the testing step) reflects the ability of the model in the accurate prediction. The result of prediction performance indicated that the RF had the best performance compared to the performances of SVM and ANN in both deep and shallow aquifers.

Furthermore, from a previous study using Taylor's diagrams can describe a model's performance in visualization form (Rahmati, Choubin et al. 2019). The visualization of the models' performance using the Taylor diagram confirmed that the RF was the appropriate model to predict total As concentrations in groundwater in this study area as presented in Figure 16. According to the Taylor criteria (i.e., correlation, standard deviation, and RMSE), the RF had the highest correlation with the observed total As probability and had the lowest RMSE as compared to those values of the SVM and ANN models (Taylor 2001). The validation with actual field data showed that RF had the lowest RMSE in both deep and shallow aquifers (RMSE =0.48 and 0.75, respectively) compared to those of SVM (RMSE = 0.73 and 0.76) and ANN (RMSE =0.67 and 0.84). Thus, it could confirm that the RF model was the appropriate model in the prediction groundwater pollution risk map of total As than SVM and ANN. (Cutler, Edwards et al. 2007, Podgorski, Wu et al. 2020).

5.3 Uncertainty assessment

The uncertainty for each ML model was determined using QR methods. To determine the uncertainty of ML models, we applied only the testing data to carry out. In this study, based on uncertainty assessment, two statistics are consisting of mean prediction interval (MPI) and prediction interval coverage probability (PICP), which were used as suggested by Rahmati (Rahmati, Choubin et al. 2019). MPI was the average of the widths of the prediction intervals, where the lower values of MPI indicates the lower uncertainty. PICP was the probability that the observed values are within the prediction intervals.

The result of PICP in the deep aquifer showed that RF had the lowest uncertainty (PICP=0.20) as compared to those of SVM (PICP=0.16) and ANN (PICP=0.05) (Table 9). In the shallow aquifer, RF also had the lowest uncertainty (PICP=0.34) as compared to those of SVM (PICP=0.23) and ANN (PICP=0.25) (Table 10). Since the PICP measurements for the three models are very different, there was no need to compare the MPI value (Rahmati et al., 2019). The PICP was the more important measurement of uncertainty as it indicates the number of observations that fall within the estimated interval (Dogulu, Lopez Lopez et al. 2015). Therefore, MPI was used as a supplementary metric between models with similar PICP values, the one with a lower MPI was regarded as the better model (Muthusamy, Godiksen et al. 2016). Thus, the QR calculated that the RF model had the lowest uncertainty compared to the uncertainty values of SVM and ANN in both shallow and deep aquifers. As mentioned above, based on the prediction performance and uncertainty assessment, we can conclude that the RF was the appropriate model to predict the total As risk area in both deep and shallow aquifers.

5.4 Risk map assessment

Many of the literature in the study area (Pipattanajaroenkul, Sonthiphand et al. 2018, Boonkaewwan, Sonthiphand et al. 2020), which study the As mechanism in Rayong groundwater basin found that As contamination in Rayong groundwater basin appeared to be a potential risk and need a proper method to locate risk area. Therefore, the present study used risk maps that output from the RF model multiplied with population density and water consumption to indicate the districts that potentially had people to expose to As contaminated in groundwater. The result showed that the districts that may have people expose to As in the deep groundwater aquifer were mostly in the north part of the Rayong groundwater basin such as Sattahip and Ban Khai districts, which had very-highly risk level, following by Ban Chang district which had a high-risk level of exposure. Other districts that might have people expose to As contaminated in the deep aquifer were Si Racha, Bang Lamung, Mueang Rayong, Pluak Daeng, and Nikhom Phatthana.

Based on the exposure of local people with As contaminated in the shallow groundwater aquifer, it mostly affected in the south path of Rayong groundwater basins such as Mueang Rayong and Ban Khai districts had a very-highly risk level, following by Nikhom Phatthana district (high-risk exposure) as well as Si Racha and Pluak Daeng districts also had a moderate risk level.

In summary, the risk map in the deep aquifer appeared to had more risk in the northern part of the Rayong groundwater basin whereas in the shallow aquifer it appears to have high risk in the southern part of the Rayong groundwater basin as shown in Figure 18. Furthermore, based on various types of land uses in Rayong

groundwater basin including industrial estate, agricultural areas, and urban areas, particularly, two landfills in the Mueang District that is close to the estuary of the groundwater basin (Boonkaewwan, Sonthiphand et al. 2020). Also, the risk areas of the shallow aquifer might be affected by anthropogenic activity in the study area such as landfill, agricultural, and industrial areas.

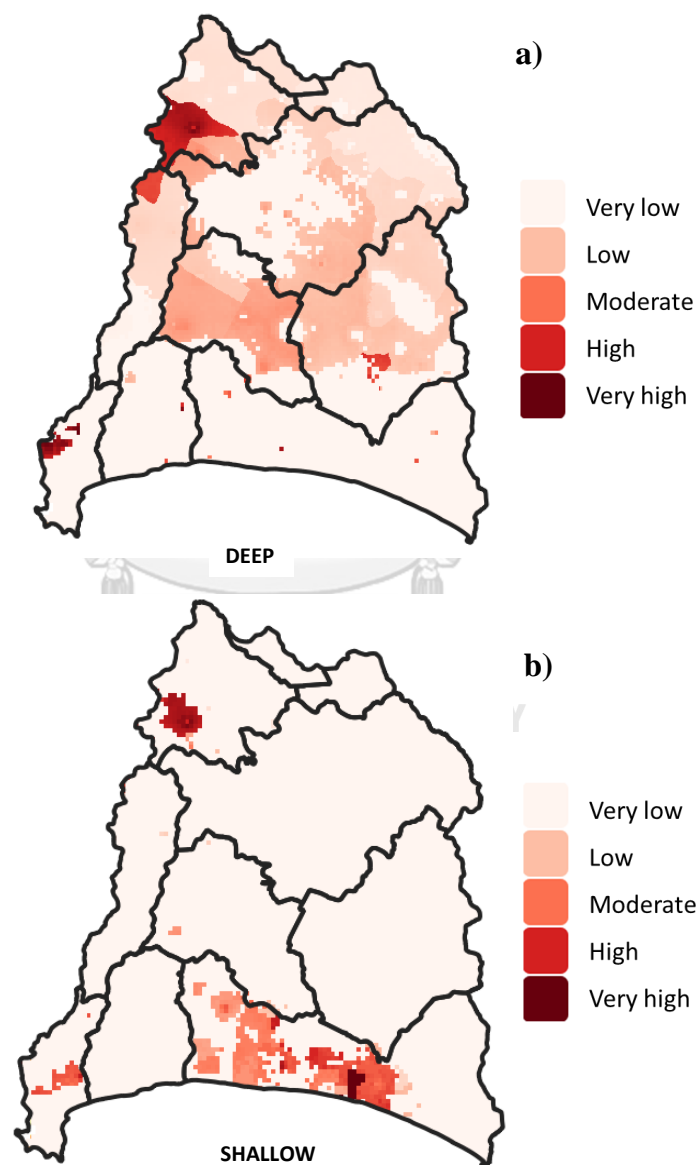


Figure 18 Groundwater pollution risk map of As contamination in a) the deep aquifer and b) the shallow aquifer

Chapter 6 Conclusions and recommendations

6.1 Conclusion

The increasing demand for groundwater has caused concern for groundwater risk assessment, especially on the EEC area project in eastern, Thailand. In this research, the machine learning algorithms had been applied to predict As contamination in Rayong coastal aquifers. Based on the present study, the majority of water quality parameters were not a normal distribution. Thus, Spearman's correlation was used in this study. From the correlation analysis, hydrochemical parameters were correlated with each other, including As concentration at the significant level, explaining As mechanisms in the groundwater environment. Therefore, using Spearman's correlation to screen unnecessary parameters before modeling might be served as an alternative way to select suitable parameters apart from the PCA method. The correlation result indicated that the point source of As in this study area was caused by both natural and anthropogenic sources.

In the modeling process, ANN showed a sign of overfitting in the goodness-of-fit, which was usually found in ANN algorithms with a small dataset. In prediction performance and uncertainty evaluation, the results indicated that the RF model had the highest prediction performance and the lowest uncertainty compare with the SVM model. Thus, it can be concluded that the RF model was the most suitable model for predicting As contamination in the Rayong groundwater basin compared to SVM and ANN models. The result of the risk map from the RF model indicates that the deep aquifer, the northern part of Rayong groundwater basin had a higher risk for people to expose to As contaminated in groundwater. In contrast, the shallow aquifer indicates

that the southern part of the Rayong groundwater basin had a higher chance for people to expose to As in groundwater, which might be related to anthropogenic activities in the Mueang District that is close to the estuary of the Rayong groundwater basin.

The outcome of this study can be used to help decision-makers, i.e. DGR and other organizations, to manage groundwater resources and protect the environment in the study region. For example, the risk area where suitable to install a filter tank for remediation and treat groundwater can be located. Moreover, base on the risk map, the decision-maker can propose a suitable policy to sustainably manage groundwater consumption in the study area.

6.2 Recommendation

Since the difference of hydrochemical facies between rainy and dry seasons, it would be further considered this effect. Furthermore, in this study, the contributing factors to the modeling process in deep and shallow aquifers are different and give significant information. Therefore, in future studies, an attempt to study groundwater modeling integrated with ML techniques should be considered under different seasons and in both aquifers.

In addition, the study of As contamination in groundwater using a hydrogeochemical model should consider the speciation of As, As(III), and As(IV) in groundwater. This was directly correlated to the oxidizing or reducing conditions in this area. When considering As(III) and As(IV), we can use this information to classify the potential risk area better than the use of total As concentration.

Moreover, in the selected parameter process by the Spearman's correlation, it would work well with hydrochemical parameters, which can describe As mechanisms

in groundwater. However, for physical parameters, this technique doesn't help much to screen unnecessary parameters. Thus, the use of hydrochemical parameters only in the modeling process was recommended in terms of time-saving.

In summary, the limitation of this study showed in many aspects as follows: the seasonal effect may relate to the chemical mechanism of As in aquifer, the total As parameter might not be enough to determine the risk area affected by As contamination. All of these aspects might affect the accuracy and reliability of predicting As risk area.

Furthermore, the methodology applied in this study can be used to study other types of contaminants in polluted groundwater. This will be useful in the investigation of groundwater contamination and remediation in other areas. Also, the study can be applied using only conventional on-site measurement parameters to predict pollution in groundwater that might decrease the cost to analyze other parameters.

APPENDIX



Figure 19 J1 groundwater well



Figure 20 J2 groundwater well



Figure 21 J3 groundwater well



Figure 22 J4 groundwater well



Figure 23 J6 groundwater well



Figure 24 J7 groundwater well



Figure 25 J8 groundwater well



Figure 26 J9 groundwater well



Figure 27 J10 groundwater well



Figure 28 J11 groundwater well



Figure 29 J13 groundwater well



Figure 30 J14 groundwater well



Figure 31 G1 groundwater well



Figure 32 G2 groundwater well



Figure 33 G3 groundwater well



Figure 34 G4 groundwater well



Figure 35 G7 groundwater well

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY



Figure 36 G8 groundwater well



Figure 37 G9 groundwater well



Figure 38 G10 groundwater well



Figure 39 G11 groundwater well



Figure 40 G12 groundwater well



Figure 41 G13 groundwater well



Figure 42 G14 groundwater well



Figure 43 G15 groundwater well



Figure 44 G17 groundwater well

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY



Figure 45 G19 groundwater well

Table 29 Detection limits and the standard level (DGR, 2017)

Parameters	Unit	Analysis techniques	Detection Limit	Standard level
Physical				
Color	Pt-Co	Spectrophotometric-Single Wavelength (2120C)	0	5
Turbidity	NTU	Nephelometric(2013B)	0.01	5
pH	-	-	1.0-14.0	7-8.5
Chemical				
TDS	mg/l	Dried at 103-105C	5	600
Total Hardness	mg/l	EDTA Titrimetric (2340C)	100	300
Non-Carbonate Hardness	mg/l	Calculation/EDTA Titrimetric (2340C)	0	200
Ca	mg/l	EDTA Titrimetric (2340C)	0.05	-
Mg	mg/l	Calculation	0	-
Na	mg/l	Direct Air-Ace tylene Flame (3111B)	0.005	-
K	mg/l	Direct Air-Ace tylene Flame (3111B)	0.005	-
Fe	mg/l	Direct Air-Ace tylene Flame (3111B)	0.005	0.5
Cl	mg/l	Argentometric (4500-FD)	0.2	250
F	mg/l	SP ANDS (4500-FD)	0.01	0.7
CO ₂	mg/l	SPANDS (4500-FD)	0	-
HCO ₃	mg/l	Titration(2320B)	0	-
SO ₄	mg/l	Titration(2320B)	0.1	200
NO ₂	mg/l	Turbidimetric (4500 -SO ₄ E)	0.01	45
PO ₄	mg/l	Cadmium Reduction (4500-NOE)	0.0001	-
Heavy metal				
As	mg/l	Hydride Generation AAS (3114B)	0.0003	0.01
Cd	mg/l	Direct Air-Ace tylene Flame (3111B)	0.0004	0.003
Cr	mg/l	Colorimetric(3111B)	0.01	0.05
Cu	mg/l	Direct Air-Ace tylene Flame (3111B)	0.003	1
Hg	mg/l	Hydride Generation AAS (3114B)	0.0001	0.001
Mn	mg/l	Direct Air-Ace tylene Flame (3111B)	0.005	0.5
Ni	mg/l	Direct Air-Ace tylene Flame (3111B)	0.001	0.02
Pb	mg/l	Direct Air-Ace tylene Flame (3111B)	0.007	0.01
Se	mg/l	Hydride Generation AAS (3114B)	0.003	0.01
Zn	mg/l	Direct Air-Ace tylene Flame (3111B)	0.005	5

REFERENCES



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Uncategorized References

Acharyya, S., et al. (2005). "Arsenic contamination in groundwater from parts of Ambagarh-Chowki block, Chhattisgarh, India: Source and release mechanism." Environmental Geology **49**: 148-158.

Ahmed, K. M., et al. (2004). "Arsenic Enrichment in Groundwater of the Alluvial Aquifers in Bangladesh: An Overview." Applied Geochemistry **19**: 181-200.

Bauer, M. and C. Blodau (2006). "Mobilization of arsenic by dissolved organic matter from iron oxides, soils and sediments." The Science of the total environment **354**: 179-190.

Bindal, S. and C. Singh (2019). "Predicting groundwater arsenic contamination: Regions at risk in highest populated state of India." Water Research **159**.

Bissen, M. and F. J. A. H. E. H. Frimmel (2003). "Arsenic a Review. Part I: Occurrence, Toxicity, Speciation, Mobility." **31**: 9-18.

Boonkaewwan, S., et al. (2020). "Mechanisms of Arsenic Contamination Associated with Hydrochemical Characteristics in Coastal Alluvial Aquifers Using Multivariate Statistical Technique and Hydrogeochemical Modeling: A Case Study in Rayong Province, Eastern Thailand." Environmental Geochemistry and Health.

Boonkhao, L., et al. (2017). "Arsenic exposure levels of petrochemical workers in three workplace settings in Rayong Province, Thailand." Human and Ecological Risk Assessment: An International Journal **23**: 00-00.

Bowell, R., et al. (2014). The Environmental Geochemistry of Arsenic -- An Overview. **79**: 1-16.

BRA (2020). Household registration population. B. O. R. Administration. Thailand.

Breiman, L. (2001). "Random Forests." Machine Learning **45**(1): 5-32.

Bundschuh, J. and J. P. Maity (2015). "Geothermal arsenic: Occurrence, mobility and environmental implications." Renewable and Sustainable Energy Reviews **42**: 1214-1222.

Charlet, L., et al. (2007). "Chemodynamics of an arsenic "hotspot" in a West Bengal aquifer: A field and reactive transport modeling study." Applied Geochemistry **22**: 1273-1292.

Cho, K. H., et al. (2011). "Prediction of contamination potential of groundwater arsenic in Cambodia, Laos, and Thailand using artificial neural network." Water Research **45**(17): 5535-5544.

Choubin, B., et al. (2017). "An Ensemble Forecast of Semiarid Rainfall Using Large-Scale Climate Predictors." Meteorological Applications **24**.

Cummings, D. E., et al. (1999). "Arsenic Mobilization by the Dissimilatory Fe(III)-Reducing Bacterium *Shewanella alga* BrY." Environmental Science & Technology **33**(5): 723-729.

Cutler, D., et al. (2007). "Random Forests for Classification in Ecology." Ecology **88**: 2783-2792.

DGR (2012). Project for exploration and study of heavy metals in groundwater Rayong and Chonburi Groundwater Basin. D. o. G. Resources. Thailand, Department of Groundwater Resources.

DGR (2017). Project of exploration and study of heavy metals in groundwater in the central and eastern regions of Thailand. D. o. G. Resources. Thailand, Department of Groundwater Resources

DGR (2020). Groundwater well data Provinces/District/Subdistrict. D. o. G. Resources. Thailand, Ministry of Natural Resources and Environment.

DMR (2007). Mineral resource in Rayong province, Thailand. D. o. m. a. resource. Thailand, Department of mineral and resource.

Dogulu, N., et al. (2015). "Estimation of predictive hydrologic uncertainty using the quantile regression and UNEEC methods and their comparison on contrasting catchments." Hydrology and Earth System Sciences **19**: 3181–3201.

Evgeniou, T., et al. (2000). "Regularization Networks and Support Vector Machines." Adv. Comput. Math. **13**: 1-50.

Garelick, H., et al. (2008). "Arsenic Pollution Sources." Reviews of environmental contamination and toxicology **197**: 17-60.

Garvey, P. J. I. (2001). "Track 2: Implementing a Risk Management Process for a Large Scale Information System Upgrade@ A Case Study." **4**: 15-22.

Geomatics (2019). "Spatial Interpolation with Inverse Distance Weighting (IDW) Method Explained." *GEODOSE* <https://www.geodose.com/2019/03/spatial-interpolation-inverse-distance-weighting-idw.html> 2017.

Gupta, A., et al. (2019). Artificial Neural Networks: Its Techniques and Applications to Forecasting.

Havryliuk, S., et al. (2018). Using the Random Forest Classification for Land Cover Interpretation of Landsat Images in the Prykarpattya Region of Ukraine.

Henseler, J. and M. Sarstedt (2013). "Goodness-of-Fit Indices for Partial Least Squares Path Modeling." *Computational Statistics* **28**: 565-580.

Herath, I., et al. (2016). "Natural Arsenic in Global Groundwaters: Distribution and Geochemical Triggers for Mobilization." *Current Pollution Reports* **2**(1): 68-89.

Jacks, G. (2017). Redox Reactions in Groundwater with Health Implications, Redox - Principles and Advanced Applications. Mohammed Awad Ali Khalid, IntechOpen.

K, D. (2019). "Top 4 advantages and disadvantages of Support Vector Machine or SVM." from <https://medium.com/@dhiraj8899/top-4-advantages-and-disadvantages-of-support-vector-machine-or-svm-a3c06a2b107>.

Kerdthep, P., et al. (2009). "Concentrations of cadmium and arsenic in seafood from Muang District, Rayong Province." *J. Health Res.* **23**.

Kho, J. (2018). "Why Random Forest is My Favorite Machine Learning Model." from <https://towardsdatascience.com/why-random-forest-is-my-favorite-machine-learning-model-b97651fa3706>.

Laerd (2018). "Spearman's Rank-Order Correlation." from <https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide.php>.

LDD (2016). Land use summary in Rayong province, Thailand. L. d. department. Thailand, Land development department

Lin, Z. and R. W. Puls (2000). "Adsorption, desorption and oxidation of arsenic affected by clay minerals and aging process." Environmental Geology **39**(7): 753-759.

Mahanta, J. (2017). "Introduction to Neural Networks, Advantages and Applications." from <https://towardsdatascience.com/introduction-to-neural-networks-advantages-and-applications-96851bd1a207>.

Maity, J., et al. (2011). "The potential for reductive mobilization of arsenic [As(V) to As(III)] by OSBH2 (*Pseudomonas stutzeri*) and OSBH5 (*Bacillus cereus*) in an oil-contaminated site." Journal of Environmental Science and Health Part A **46**: 1239–1246.

Marjan, Č., et al. (2018). "Estimating the Performance of Random Forest versus Multiple Regression for Predicting Prices of the Apartments." ISPRS International Journal of Geo-Information **7**: 168.

Meharg, A. A. and M. M. Rahman (2003). "Arsenic Contamination of Bangladesh Paddy Field Soils: Implications for Rice Contribution to Arsenic Consumption." Environmental Science & Technology **37**(2): 229-234.

Mijwil, M. M. (2018). "Artificial Neural Networks Advantages and Disadvantages." from <https://www.linkedin.com/pulse/artificial-neural-networks-advantages-disadvantages-maad-m-mijwil>.

Mukherjee, A., et al. (2009). "Chemical evolution in the high arsenic groundwater of the Huhhot basin (Inner Mongolia, PR China) and its difference from the western Bengal basin (India)." Applied Geochemistry **24**(10): 1835-1851.

Muthusamy, M., et al. (2016). "Comparison of Different Configurations of Quantile Regression in Estimating Predictive Hydrological Uncertainty." Procedia Engineering **154**: 513-520.

Naidu, R. and P. Bhattacharya (2009). "Arsenic in the environment—risks and management strategies." Environmental Geochemistry and Health **31**(1): 1-8.

Naimi, B. and M. Araújo (2016). "sdm: A reproducible and extensible R platform for species distribution modelling." Ecography **39**: n/a-n/a.

Narkhede, S. (2018). "Understanding AUC - ROC Curve." from <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>.

Nickson, R., et al. (2000). "Mechanism of Arsenic Release to Groundwater, Bangladesh and West Bengal." Nickson, R. and McArthur, J.M. and Ravenscroft, P. and Burgess, W.G. and Ahmed, M. (2000) Mechanism of arsenic release to groundwater, Bangladesh and West Bengal. Applied Geochemistry, 15 (4). pp. 403-413. ISSN 08832927 **15**.

Nriagu, J., et al. (2007). "Arsenic in soil and groundwater: an overview." Trace Metals and other Contaminants in the Environment **9**: 3-60.

Ootsahkarn, K. (2018). "The Eastern Economic Corridor, Thailand."

Pipattanajaroenkul, P., et al. (2018). "Detection of arsenite-oxidizing bacteria in groundwater with low arsenic concentration in Rayong province, Thailand." MATEC Web of Conferences **192**: 03036.

Podgorski, J., et al. (2020). "Groundwater Arsenic Distribution in India by Machine Learning Geospatial Modeling." International Journal of Environmental Research and Public Health **17**: 7119.

Rahmati, O., et al. (2019). "Predicting uncertainty of machine learning models for modelling nitrate pollution of groundwater using quantile regression and UNEEC methods." Science of The Total Environment **688**.

Rao, M. R., et al. (2018). "A Survey on Prevention of Overfitting in Convolution Neural Networks Using Machine Learning Techniques." International Journal of Engineering and Technology(UAE) **7**: 177-180.

Sae-Ju, J., et al. (2019). "Hydrochemical, geophysical and multivariate statistical investigation of the seawater intrusion in the coastal aquifer at Prachuap-Khiri-Khan Province, Thailand." Journal of Asian Earth Sciences **191**: 104165.

Sajedi Hosseini, F., et al. (2018). "A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination." Science of The Total Environment **644**.

Scott, M. J. and J. J. Morgan (1995). "Reactions at Oxide Surfaces. 1. Oxidation of As(III) by Synthetic Birnessite." Environmental Science & Technology **29**(8): 1898-1905.

Shankar, S., et al. (2014). "Arsenic Contamination of Groundwater: A Review of Sources, Prevalence, Health Risks, and Strategies for Mitigation." The Scientific World Journal **2014**: 304524.

Shrestha, D. and D. Solomatine (2006). "Machine learning approaches for estimation of prediction interval for the model output." Neural networks : the official journal of the International Neural Network Society **19**: 225-235.

Smith, E., et al. (2001). "Chemistry of Inorganic Arsenic in Soils: II. Effect of Phosphorus, Sodium, and Calcium on Arsenic Sorption." Journal of environmental quality **31**: 557-563.

Solomatine, D. and D. Shrestha (2009). "A novel method to estimate model uncertainty using machine learning techniques." Water Resources Research, **45**, 2009 ; doi:10.1029/2008WR006839 **45**.

Taylor, K. (2001). "Summarizing multiple aspects of model performance in a single diagram." Journal of Geophysical Research **106**: 7183-7192.

Taylor, S. R. and S. McLennan (1985). The continental crust : its composition and evolution : an examination of the geochemical record preserved in sedimentary rocks.

Thornton, I. (1996). "Sources and pathways of arsenic in the geochemical environment: Health implications." Geological Society, London, Special Publications **113**: 153-161.

TMD (2015). Thai Climate. M. Department. Thailand, Thai Meteorological Department.

USEPA (2002). Groundwater Sampling Guidelines for Superfund and RCRA Project Managers. U. S. E. P. Agency. U.S. Environmental Protection Agency, Region 10

1200 Sixth Avenue

Seattle, Washington 98101, Environmental Protection Agency. **EPA542-S-02-001**.

Uusitalo, L., et al. (2015). "An overview of methods to evaluate uncertainty of deterministic models in decision support." Environmental Modelling and Software **63**: 24-31.

Valenzuela, C., et al. (2009). "Isolation of Arsenite-Oxidizing Bacteria from Arsenic-Enriched Sediments from Camarones River, Northern Chile." Bulletin of environmental contamination and toxicology **82**: 593-596.

Welch, A. H., et al. (2000). "Arsenic in ground water of the United States: occurrence and geochemistry." Ground Water **38**(4): 589-604.

Wexler, E. J. and P. E. Maus (1988). Ground-water flow and solute transport at a municipal landfill site on Long Island, New York; Part 2, Simulation of ground-water flow. Water-Resources Investigations Report.

WHO (2018). "Arsenic." from <https://www.who.int/news-room/fact-sheets/detail/arsenic>.

Winkel, L., et al. (2008). "Predicting groundwater arsenic contamination in Southeast Asia from surface parameters." Nature Geoscience - NAT GEOSCI 1: 536-542.

Zubair, A., et al. (2015). "Contamination of Arsenic in Sea, Surface, and Ground water in the coastal aquifers of Sindh, Pakistan." Mitteilungen Klosterneuburg: 163-178.



VITA

NAME Narongpon Sumdang

DATE OF BIRTH 9 Jan 1995

PLACE OF BIRTH Udon thani, Thailand

INSTITUTIONS ATTENDED Chulalongkorn University

HOME ADDRESS 27 Village 1, Khao Chiak, Mueang Phatthalung District, Phatthalung, Thailand 93000.

PUBLICATION Sumdang, N. & Chotpantararat, S. (2020). PROBABILITY MAPPING OF ARSENIC CONTAMINATION USING ARTIFICIAL NEURAL NETWORK. Oral presented at the TSHE 3rd National Conference on Environment : Healthy Environment in the Next Decade (2021 - 2030 A.D.).

AWARD RECEIVED The scholarship from the Graduate School, Chulalongkorn University to commemorate the 72nd anniversary of his Majesty King Bhumibol Aduladej.
The 90th Anniversary Chulalongkorn University Fund (Ratchadaphiseksomphot Endowment Fund).