2015-09-25

Geovisualization for Association Rule Mining in CHOPS Well Data

Sun, Xiaodong

UNIVERSITY OF CALGARY

Geovisualization for Association Rule Mining in CHOPS Well Data

by

Xiaodong Sun

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN GEOMATICS ENGINEERING

CALGARY, ALBERTA

SEPTEMBER, 2015

Abstract

Association rule mining has recently been applied to improve the oil recovery of CHOPS by discovering the association rules between reservoir properties and oil production from CHOPS well data. However, it leaves reservoir engineers with big challenging tasks to find interesting rules, understand the rules by the distribution patterns of relevant wells and make subsequent predictions by the application areas of the rules.

In this thesis, three kinds of rule filters are developed to find out the interesting rules. Moreover, point-based and surface-based geovisualization methods are proposed to display the distribution patterns of relevant wells, build and represent potentially applicable areas for the rules on the map. A system prototype, containing association rule mining with filters, geovisualization functions, is developed. A case study has been carried out on a real CHOPS well dataset in western Alberta, Canada. The findings in the case study illustrate the feasibility of the proposed methods.

Acknowledgements

I would like to express my gratitude to the attention, patience, encouragement, knowledge and friendship that my supervisor Dr. Xin Wang has given me during my MSc program. Moreover, this research would not have been possible without the oil well data of CHOPS from Divestco Inc. Thanks to Dr. Yongxiang Cai for her support on data preprocessing and sharing of her rich experience and expertise in reservoir engineering. I would also like to give thanks to Dr. Danielle J. Marceau and her thoughtful advices on my geovisualization course project in her class. I would like to thank all the help that the group members in the Intelligent Geospatial Data Mining Laboratory have given me.

I am very grateful to my mother, Mrs. Liqin Wang, for her unfailing love and belief in me during my master's study. I am also extremely grateful to my brother and sister-in-law, Mr. Wugang Sun and Mrs. Yi Shi, my sister, Mrs. Xinfeng Xu for their constant encouragement and support during the three years in my life.

Table of Contents

List of Tables

List of Figures and Illustrations

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
API	Application Programming Interface
ARM	Association Rule Mining
CHOPS	Cold Heavy Oil Production with Sand
GeoViz	Geovisualization
GUI	Graphical User Interface
IDW	Inverse Distance Weighting
RMSE	Root Mean Squared Error

Chapter One: **Introduction**

1.1 Background

Cold Heavy Oil Production with Sand (CHOPS) is an oil extraction process for producing heavy crude oil. CHOPS technology has been widely used in heavy oil production in the Western Canadian Sedimentary Basin, since the mid-1990s (Sawatzky et al., 2002). Crude oil is produced along with oil sand, water, and gas. One of the biggest challenges for reservoir engineers is to understand the factors impacting oil recovery for CHOPS.

Data mining refers to a knowledge discovery process by which interesting, implicit and unknown patterns could be found in large databases (Frawley et al., 1992). It also integrates statistics and database systems, machine learning and artificial intelligence (Chakrabarti et al., 2006). With large amount of collected oil and gas data in petroleum industry, data mining has enormous potentials in explaining the complex underground geological and reservoir conditions affecting oil production.

Association rule mining (ARM) is one of the most popular data mining methods and it was first proposed for mining causal structures, patterns, or correlations from transaction data or other data repositories (Agrawal et al., 1993). Recently, it has been successfully utilized in reservoir analysis and modeling. Aulia et al. (2010) used association rule mining to increase oil recovery, by discovering the association rules between field parameters (e.g., bottom hole pressure, surface rate at each injection well) and oil recovery from data of general oil wells. Because the oil recovery of CHOPS wells

is mainly influenced by reservoir properties, Cai et. al. (2014) applied association rule mining on historical CHOPS well data to improve the oil recovery of CHOPS. CHOPS well data generally includes well locations and other non-spatial attributes, such as reservoir property parameters and production performance parameters. By applying association rule mining algorithms such as Apriori (Agrawal and Srikant, 1994) on CHOPS well data, the quantitative relationships between reservoir properties influencing oil recovery and oil production performance could be discovered (Cai et al., 2014).

1.2 Problem Statement

Although association rule mining is a promising approach for understanding and improving oil recovery of CHOPS, some unsolved problems cause that association rule mining has not gotten a wide use.

1.2.1 The Problem of Discovering Interesting Rules in CHOPS Well Data

Mining association rules in the data of CHOPS wells inevitably results in a very large number of association rules. Users are always interested in a subset of interesting association rules containing specific items and/or matching some wells. Traditionally, item constraints that are some expressions stating some conditions on the items of output association rules were used a post processing step of mining algorithms to generate interesting association rules with specific items. For example, boolean expressions (Ramakrishnan et al., 1997) were applied to control the absence or presence of some items of output association rules. But classic item constraints were developed for the problem of discovering interesting association rules in transaction data; thus, they did not take the geospatial constraints of spatial objects related to association rules into account.

Whether an association rule in CHOPS well data is interesting, is not only influenced by support and confidence value of the rule, but also determined by other important factors.

As for the CHOPS data, the interesting rules must be the ones that are able to represent the influence of reservoir properties on the oil production performance since the main objective of association rule mining in CHOPS well data is to discover the relationships between reservoir properties and oil production. In other words, it indicates that the reservoir property indicators must be the antecedent of an interesting association rule, while the oil production performance indicators must be the consequence. On top of it, whether an association rule is interesting depends on the information or location of relevant wells. For example, after users obtain the rules between reservoir properties and oil production, they may only want a subset of the rules, which match the wells with certain unique well identifiers or match a group of nearby wells within a specific geospatial area. Thus, the problem of extracting interesting rules needs to be solved according to the characteristics of CHOPS well data.

1.2.2 Visualization of Interesting Rules Regarding Wells

After interesting rules are discovered from CHOPS well data, the locations of the wells that match the rules need to be visualized in an intuitive way. Compared with the customer objects in transaction data, the well objects in CHOPS well data have geospatial attributes, e.g., longitude and latitude. It causes that the valuable information of the interesting rules in CHOPS well data do not only include the patterns between the non-spatial attributes of wells, i.e., the relationships between reservoir properties and oil production, but also contain the spatial distribution patterns of matching wells hidden behind the rules. Visualization is a possible solution and has a long history of making

large data and hidden patterns within the data accessible. Several visualization methods were proposed for association rules, such as scatter plots, graphs plots and parallel coordinate plots (Buono and Costabile, 2005; Ertek and Demiriz, 2006; Hahsler and Chelluboina, 2011; Klemettinen et al., 1994; Rainsford and Roddick, 2000; Unwin et al., 2001; Yang, 2003), to make the association rules accessible and hidden patterns within large number of the rules be identified. However, the existing visualization methods for association rules are limited on representing the geospatial attributes of the rules in CHOPS well data. The methods are designed for association rules discovered in transaction data or other non-spatial data; thus, they inevitably focus on the visualization of the rule content characteristics instead of the spatial distribution of data objects related to the rules.

1.2.3 Visualization of Interesting Rules Regarding Application Areas

In addition to relating well locations with interesting association rules, building and visualization of possible application areas for the interesting rules are also worthy of research. Application areas of an association rule refer to continuous surfaces where the association rule may happen or be applied. In practice, the application areas are very valuable for reservoir engineers to make decisions or predictions based on the pattern included by the rule. The locations of CHOPS wells are generally an irregular array of discrete geospatial points. It is impossible to directly construct continuous areas by CHOPS well data for each attribute appearing in an association rule. We need to firstly investigate proper data spatialization techniques that can fill in data between oil wells for attributes appearing in the rule. Secondly, on the basis of the built layers containing continuous surfaces of the attributes, proper visualization schemes need to be designed in

order to finally generate application areas for the whole rule in the form of prediction maps.

1.2.4 GIS Prototype for CHOPS Well Data

With the growing volume of oil and gas data that have been collected, increasingly petroleum companies as well as oil and gas data companies have implemented geographic information systems (GIS) to visualize, analyze and study the large amount of oil and gas data. Combined with association rule mining and visualization of found rules a GIS prototype will make the big oil and gas data managed more efficiently. However, there is no existing GIS prototype integrating mining and visualization of association rules and special for CHOPS well data. It is necessary to develop a system prototype that can efficiently manage collected CHOPS well data, extract and visualize the interesting association rules from the data.

1.3 Research Objectives

The overall objective of the research is to promote the application of association rule mining on CHOPS well data by proposing new methods for extracting interesting association rules and new visualization methods for the interesting rules on the basis of previously related research. Specifically, the problems stated in Section 1.2 lead to the following research objectives:

1. Propose association rule filters to facilitate the discovery of interesting rules from the large sets of generated rules in CHOPS well data;

2. Propose a new visualization method to display the distribution patterns of oil wells relevant to the extracted interesting rules;

3. Propose a new visualization method to build and represent application areas for the extracted interesting rules;

4. Develop a system prototype integrating management, association rule mining and visualization on CHOPS well data.

1.4 Research Contribution

The main contributions of the thesis can be summarized as follows.

1. This thesis introduces three new association rule filters that help reservoir engineers discover interesting rules efficiently. The antecedent-consequence filter produces the interesting rules with some specific antecedents and consequences. As options, the well identifier filter further finds out the interesting rules by information of wells such as the unique well identifiers and the well location filter directly extracts the interesting rules on the map;

2. A visualization method for association rules in CHOPS well data, point-based geovisualization, is proposed for linking interesting association rules with wells. The method uses different symbols to highlight the locations of the wells whose records in the database satisfy some interesting rule at various extents. Reservoir engineers can better understand the rules that they are interested in and find possible distribution patterns of the wells by the visualized well locations on the map;

3. Another novel visualization method, surface-based geovisualization, builds and represents the areas on the map for an interesting rule where the rule may be applicable. The method is based on spatial interpolation and it can be used on the premises that spatial dependence basically exists in all attributes appearing in an association rule. The surface-based geovisualization can assist reservoir engineers in making decisions or predictions based on the patterns included by the discovered interesting rules;

4. A system prototype, named CHOPSData-GeoViz, is developed for association rule mining and visualization of found rules in CHOPS well data. The system prototype integrates association rule mining with proposed new filters, point-based and surface-based visualization;

5. A case study was conducted on a real CHOPS well dataset from the Lloydminster heavy oil block in Alberta, Canada. The case study validates the feasibility of the proposed association rule filters and visualization methods for association rules in CHOPS well data, as well as the usefulness of the built system prototype.

1.5 Thesis Outline

Chapter Two gives a literature review of association rule mining, existing visualization techniques for association rules, as well as geovisualization techniques. Chapter Three introduces the detailed methodology of the proposed novel rule filters and visualization methods, and the system prototype for association rule mining and visualization of found rules in CHOPS well data. In Chapter Four, a case study was conducted on a real CHOPS well dataset from the Lloydminster heavy oil block in Alberta, Canada. Chapter Five draws conclusions and states future work of the thesis.

Chapter Two: **Related Work**

This chapter presents the literature study in the following areas. First, association rule mining and related research are reviewed. Second, previous works on visualization methods for association rules are presented. Third, a review of geovisualization is given.

2.1 Association Rule Mining

Association rule mining is one of the popular and well-developed data mining techniques for discovering correlations among variables in large datasets. It is introduced in this thesis to analyze the relationships between reservoir properties and oil production performance from real historical CHOPS well data.

2.1.1 Overview

Association rule mining (ARM) was first introduced by Agrawal and Srikant to facilitate in analyzing transactional databases and derive association rules (Han and Kamber, 2006; Wu X., 2007). A typical example comes from the market basket analysis. Association rule mining analyzes customer consumption behaviors and habits by finding associations between different items that customers purchase. For example, if the customer buys bread, how likely does the customer buy milk meanwhile? Such information can be helpful in improving sale assignments such as shelf space placement. Association rule mining has been applied to many domains including marketing (Sohn and Kim, 2008; Jiao and Zhang, 2005), bioinformatics (Creighton and Hanash, 2003) and reservoir analysis (Aulia et al., 2010). With the definition of the association rule in Agrawal et al. (1993), let D be

the set of all items, and X and Y be two subsets of D such that $X, Y \subset D$. An association rule with respect to X and Y can be in the following form of:

$$X \Rightarrow Y (\text{or IF } \{X\} \text{ THEN } \{Y\}), \text{ such that } X, Y \subset D,\ X \cap Y = \varnothing \text{ and } X, Y \neq \varnothing \quad (2.1)$$

where X is called the antecedent and Y is called the consequence. The two concepts are important in defining the interestingness (interest degree) of an association rule (Han J. and Kamber M. 2006), support and confidence. The support of rule $X \Rightarrow Y$ is defined to be the percentage of transaction records including $X \cup Y$ to the total number of transaction records in a database.

$$\text{support } (X \Rightarrow Y) = P(X \cup Y) \quad (2.2)$$

The confidence of rule $X \Rightarrow Y$ is the percentage of transactions containing X and Y to the number of transaction records only containing X.

$$\text{confidence } (X \Rightarrow Y) = P(Y \mid X) = \frac{P(X \cup Y)}{P(X)} \quad (2.3)$$

Association rules satisfied with the support and confidence threshold values are recognized to be strong. One of main objectives of association rule mining is the generalization of all interesting rules satisfying both minimum support and confidence thresholds from some transaction database.

The general process of ARM can be divided into two steps. Frequent itemsets satisfying the minsup threshold are firstly found. Then all the interesting rules are generated from the frequent itemsets. Detailed process will be introduced in the section 2.1.2.

2.1.2 Mining Association Rules with Item Constraints

The problem of discovering association rules has received considerable research attention and several constraints for mining association rules have been developed using the Apriori algorithm as the basis.

Apriori is a classic association rule mining algorithm which was proposed by Agrawal and Srikant (1994) for mining frequent itemsets and associations for a transactional dataset. Apriori is a seminal algorithm, using a level-wise search mechanism to find all the frequent itemsets. The algorithm starts by identifying the frequent 1-itemset through scanning the whole dataset and computing the support of each item. Then, the frequent 1-itemsets are used to find the frequent 2-itemsets and the frequent 2-itemsets are used to find frequent 3-itemsets. The whole process goes on until frequent itemsets cannot be found any more. The search for the itemsets of any frequent level needs to do a full scan in the dataset. The Apriori property is introduced for reducing the searching space and improving the searching efficiency of the level-wise frequent itemset. The Apriori property refers to "all nonempty subsets of a frequent itemset must also be frequent" (Agrawal and Srikant, 1994). The basis is the observation that a super itemset of a non-frequent itemset is still non-frequent. For instance, assuming itemset X is not frequent, $\sup(X) < minsup$. If item Y is added to itemset X, then the obtaining itemset, $X \cup Y$, cannot occur more frequently than the itemset X; thus, $X \cup Y$ is not frequent, either. On the basis of the Apriori property, the Apriori algorithm is summarized as follows. Let k-itemset denote an itemset including k items and F_k and C_k denote the collections of frequent k-itemsets and candidate k-itemsets, respectively. The Apriori algorithm firstly scans the dataset, computes the presence time of each item and

determines 1-frequent itemsets, denoted by F_1. The subsequent scans include two procedures. In the first procedure, F_k found in the k-th scan are used to create the C_{k+1}. C_{k+1} is a superset of F_k and all the subsets are considered as frequent. In the second step, the Apriori algorithm scans the dataset again to compute the support of each candidate in $C_{k+1;}$ and, the ones with support of less than the minimum support threshold are removed. This process ends when F_k is empty. After all frequent itemsets are found, all nonempty subsets of every frequent itemset, h, will be enumerated to generate interesting rules. For each subset of h, r = subset(h), a rule is generated with the form of r=>h-r, if its confidence is larger than the minimum confidence threshold (Han and Kamber, 2006).

Applying association rule mining algorithms such as the Apriori on transaction data often results in a very large number of association rules. In practice, users are often only interested in a subset of association rules and they may only need the rules that contain specific items.

As a solution, many researchers proposed and applied item constraints as a "post-processing" step of the mining algorithms. For example, Ramakrishnan et al. (1997) solved the problem in the presence of constraints that were boolean expressions, allowing users to specify the subset of rules that they were interested in. The boolean expressions were used to control the presence or absence of some items. Such constraints could be used in the process of finding frequent itemsets or generating candidates. For example, if the boolean expression constraints were applied when finding frequent itemsets, only the frequent itemsets that satisfied the boolean expression were found, instead of all the potentially frequent itemsets. Item constraints in the form of expressions were most popular ways to efficiently extract a subset of rules from large amounts of generated rules.

But the item constraints were developed for the problem of discovering interesting association rules in transaction data; thus, they did not take the geospatial constraints of spatial objects related to association rules into account. Thus, the common item constraints are not suitable for the stated problems of discovering interesting rules in CHOPS well data. The goal of association rule mining in CHOPS well data is to find relationships between reservoir properties and oil production. We should further limit that reservoir properties can only be in the antecedents and oil production can only be in the consequences of generated rules. In other words, antecedent and consequence constraints should be considered in the constraints. Moreover, the objects in CHOPS well data are the wells owning geospatial attributes and other information. The common constraints will fall short on discovering interesting rules that matching the wells located in a specific area or having special properties. The information and location of spatial objects related to association rules should also be included into the constraints for association rule mining in CHOPS well data.

2.2 Visualization of Association Rules

After interesting rules are extracted, the distribution patterns of oil wells and application areas of the interesting rules need to be represented to reservoir engineers in a proper way. As a possible solution, data visualization and it refers to a set of techniques applied to encode and represent the data in the form of visual elements (e.g., points, lines, icons or bars) in plots, and its main goal is to represent the data information clearly and intuitively to users (Friedman, 2008). We can often visualization techniques to discover the hidden patterns within the raw data. Recently, several visualization techniques for

association rule mining in transaction data have been recently proposed to help users analyze the association rules. In this section, we briefly introduce several representative visualization methods for association rules including scatter, parallel coordinate and graph plots and then analyze their limitations on the stated problems.

2.2.1 Traditional Visualization Methods for Association Rules

Scatter plots visualize association rules as scatter points on two-dimensional or higher coordinate systems. They are intuitive visualizations of association rules. For example, in a two-key plot, the coordinate axes *x* and *y* represent the support and confidence values of the rules, and the color of the scatter points represents another attributes of the rules (Unwin et al., 2001). Figure 2-1 shows an example of visualization of 5668 rules. It uses two interesting measures, support and confidence on the x-axe and y-axe. Additionally, a third measure lift is used as the color (gray level) of the points. A color key from 0 to 20 is provided to the right side of the plot chart. From the figure we can see that scatter plots like two-key work well for very large sets of association rules.

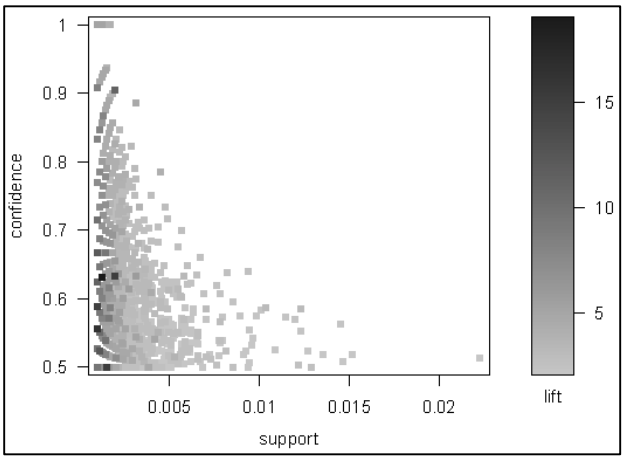


Figure 2-1 An example of rule visualization by a scatter plot (Hahsler and Chelluboina, 2011)

In a parallel coordinate plot, association rules are represented as polygonal lines on a coordinate system with shared *x* and *y* axes (Yang, 2003). The antecedent and consequence items of the rules are used for one coordinate axis, and the other axis is used to represent the corresponding positions of antecedent and consequence items in the rules. Unlike scatter plots, parallel coordinate plots more emphasize the visualization of structure characteristics of antecedent and consequence items within the rules. Through this method, the item composition of the rules and common patterns of the structure characteristics can be observable. Figure 2-2 shows a parallel coordinates plot for 10 rules.

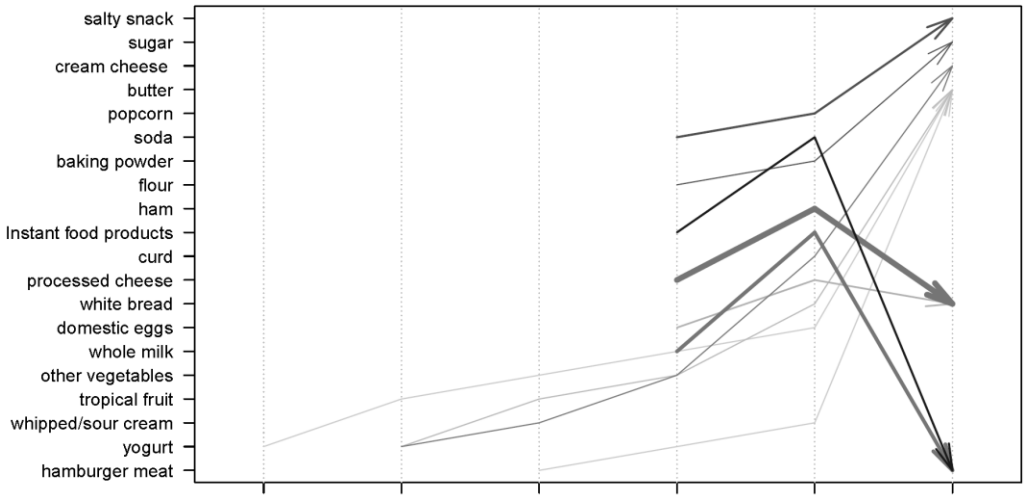


Figure 2-2 An example of rule visualization by a parallel coordinate plot (Hahsler and Chelluboina, 2011)

Graph plots represent association rules as figures with vertices and edges (Klemettinen et al., 1994; Rainsford and Roddick, 2000; Buono and Costabile, 2005; Ertek and Demiriz, 2006). The vertices are used for the antecedent and consequence items of the rules. The relationship of the items of one association rule is shown by the

connected edges of the items. Figure 2-3 shows a graph plot for an association rule IF {soda, popcorn} THEN {salty snack}. Graph plots can clearly show the internal relation of the antecedent and consequence items within the rules.

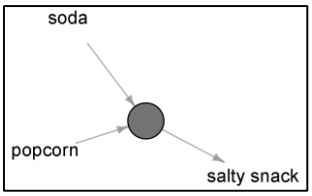


Figure 2-3 An example of rule visualization by a graph plot (Hahsler and Chelluboina, 2011)

2.2.2 Limitations of Traditional Methods in CHOPS Well Data

The goals of the classical visualization methods for association rules are the discovery of the interesting rules in transaction data as well as the patterns that cannot be directly identified from the interesting rules, such as distributions of measurements (scatter plots), sequences of items (parallel coordinate plots), and relationships between items (graph plots). The classical visualization methods were proposed for the association rules in transaction data and they did not consider the spatial attributes of the objects relevant to the association rules. In CHOPS well data, well objects have geospatial locations. For the interesting rules in CHOPS well data, their valuable information or patterns are hidden distribution patterns of the wells matching the rules and the application areas where the rules may happen. If the hidden knowledge are represented in a proper way, it will be very helpful for the reservoir engineers to make predictions based on the patterns included by the rules. Traditional visualization methods for association rules are undoubtedly suitable for the interesting association rules in CHOPS well data. To

overcome challenges, geovisualization that is special for geospatial data analysis and its application will be reviewed in the next section.

2.3 Geovisualization and Applications

On the basis of scientific visualization, Geovisualization (Geographic Visualization) integrates GIS, cartography to explore geographic data and communicate geographic information in support of geospatial analysis (MacEachren and Kraak, 1997). Geovisualization mainly helps identify, compare, and interpret features within geographic data (MacEachren et al. 1999).

By Geovisualization, the geographic data can be displayed by a map interface or high dimensional coordinate, and unexpected geospatial trends and patterns hidden behind the data can be discovered. Geovisualization has recently become widely employed in many scientific disciplines. Recently, geovisualization is combined with symbology and spatialization techniques to mapping and predicting potential patterns within geospatial data. Most representatives of them are the geovisualziation methods proposed by Aoidh et al. (2013) and Gienko and Terry (2012).

Aoidh et al. (2013) proposed a geovisualization method where symbology was explored for communicating the landscape genetics data in an intuitive way. Landscape genetics, considering genetic population structure represented by spatially referenced parameters in ambient landscape were proved to be important for wildlife management. However, there is no effective visualization method to communicate the usable information within the landscape genetics data with stakeholders in a suitable format. Aoidh et al. (2013) in response proposed a geovisualization method to help stakeholders

without any GIS or genetic expertise learn about landscape genetics and managing the wildlife. To display the landscape genetic information and represent the spatial distribution behind the landscape genetics data, they explored appropriate symbology such as the symbols with different colors and sizes to show categories and amount of sampling data, through an accessible user friendly interface, as shown in Figure 2-4.

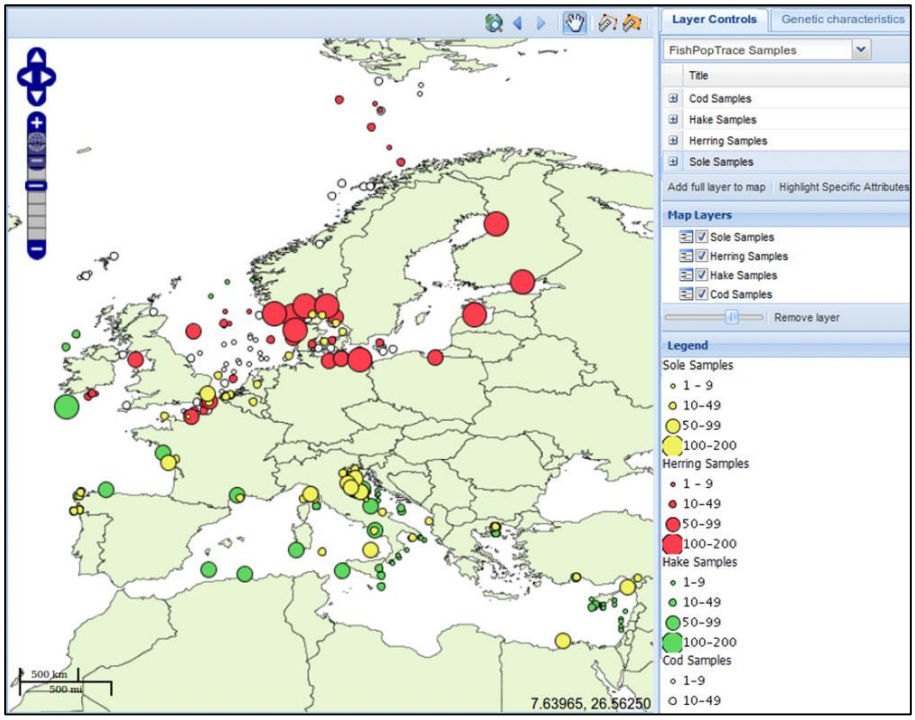


Figure 2-4 A geovisualization tool interface showing the sampling locations and quantities collected for genetic analysis (Aoidh et al., 2013).

Moreover, Gienko and Terry (2012) introduced a geovisualization method for representing and predicting cyclone behaviors, where several spatial interpolation techniques were successfully combined with geovisualization for identification and analysis of cyclone behavior features. They illustrated and discussed the value of geovisualization combined with spatialization techniques when it was used to make

analysis of spatial structures and characteristics of tropical cyclones behaviours in the South Pacific Ocean. Figure 2-5 illustrates a spatialization plot of cyclone duration using the inverse distance weighted spatial interpolation. The discovery of certain patterns and dependencies in cyclone behaviour were found by the proposed geovisualization method which helped to prepare further strategy in advanced data analysis using data mining methods. Most importantly, the research illustrated and provided preliminary exploration of spatial interpolation techniques that may be of enormous value for the geovisualization of geospatial data.

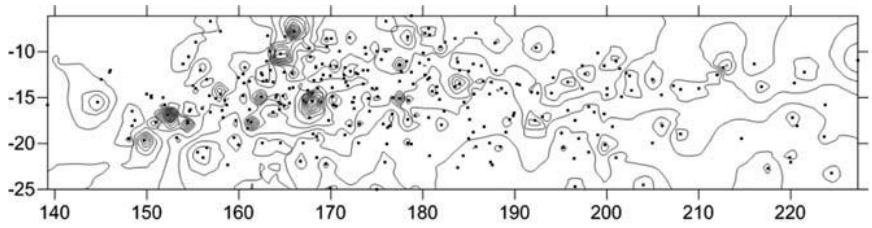


Figure 2-5 Geovisualization of tropical cyclone duration based on inverse distance weighted interpolation (Gienko and Terry, 2012).

Although the above geovisualization methods cannot be utilized to represent the distribution patterns of relevant wells and application areas of the association rules, they can be used as the basic ideas of geovisualization of the rules in CHOPS well data. The well locations can be highlighted with the different symbols on the map depending on their relationship with association rules in CHOPS well data. The application areas of association rules can start by applying spatial interpolation and generating the continuous surfaces from the attribute data of discretely sample wells on the map.

Chapter Three: **Mining and Geovisualization of Association Rules in CHOPS Well Data**

This chapter starts by introducing association rule mining in CHOPS Well Data, including data transformation and three association rule filters for extracting interesting rules for association rule mining in CHOPS well data. Next, two geovisualization methods are proposed to visualize the selected interesting rules on the map. Finally, a system prototype, named CHOPSData-GeoViz, is implemented for mining, filtering and visualizing association rules in CHOPS well data.

3.1 Data Transformation for Association Rule Mining

In the CHOPS well dataset, each record contains reservoir properties and oil production performance of each well. Moreover, the location of each well is denoted in longitude and latitude coordinates. In advance of association rule mining, the essential preprocessing work is the transformation of the values of the studied reservoir property and oil production performance parameters into a set of sub-ranges through the use of discretization schemes. Common algorithms of association rule mining, such as Apriori (Agrawal and Srikant, 1994), handle data better with discretized attributes for two reasons. Association rule mining emphasizes discovering unknown and useful patterns instead of over accurate or even trivial patterns (Marco and Valentina, 2004). Also, the reduction of detail in the source data can make the mining process more efficient and found patterns more accessible (Cai et al., 2014). Therefore, the continuous values of the reservoir property and oil production performance parameters in the source CHOPS well data need to be discretized before association rule mining. For example, a numeric value

(73.2) of the reservoir property attribute, cumulative pore volume, can be transformed into discretized value “1”, which represented a range of values from 59.2 to 94.4. Then the association rule mining algorithm such as Apriori can be applied to the preprocessed data. A more detailed discussion about data preprocessing work for association rule mining in CHOPS well data is provided in Cai et al (2014).

3.2 Mining Interesting Association Rules with Filters

Association rule mining on CHOPS data usually generates many association rules. However, not every association rule is interesting. Users have to look though every generated rule to select the ones that they are interested in. Item constraints in the form of expressions are the most popular ways to efficiently extract a subset of rules from large amounts of generated rules. As previously discussed in the Chapter Two, the common item constraints are not suitable for discovering interesting rules in CHOPS well data. On the basis of item constraints, the following three filters are introduced to specially facilitate mining interesting rules in CHOPS well data.

3.2.1 Antecedent-Consequence Filter

As previously mentioned, an interesting association rule in CHOPS well data should represent the influence of the reservoir property on oil production. Therefore, the antecedent-consequence filter is developed. It is applied during the process of association rule mining in the CHOPS well data. Specifically, after all the frequent itemsets that satisfy the minimum support threshold are found, only the association rules whose antecedent items are reservoir property parameters and consequence items are oil production performance parameters will be generated, rather than generating all the

possible association rules. An example of the association rules is IF {cumulative pore volume = 168.8 ~ 291.2} THEN {effective yield = 808.2 m^3 ~ 1452.9 m^3}, i.e., if the reservoir property's cumulative pore volume is between 168.8 and 291.2, then the effective yield is in the range of 808.2 m^3 and 1452.9 m^3. This type of rules is of interest to reservoir engineers who want to find the relationship between reservoir properties and oil production.

3.2.2 Well Identifier Filter

In addition to specific antecedents and consequences, whether an association rule is interesting also relates to the wells that the rule matches. The well identifier filter is developed and is applied after the process of association rule mining in the CHOPS well data. On top of the antecedent-consequence filter, the well identifier filter can be used to find interesting rules by well identifiers from the ones whose antecedent items are reservoir properties and consequence items are oil production performance parameters.

3.2.3 Well location Filter

Different from the objects in traditional transaction data, CHOPS wells have locations. As mentioned in the problem statement, whether an association rule is interesting also depends on the locations of the wells that the rule matches. Reservoir engineers might want to further search some of the association rules between reservoir properties and oil production, based on the well locations. Well location filter is developed based on this requirement. It provides the function for finding interesting rules by interactively selecting one or multiple wells in an area on the map.

In summary, the entire process of association rule mining with filters on CHOPS Well Data includes the following steps, as shown in Figure 3-1. First, the source CHOPS well data are discretized and then the Apriori algorithm is applied to the processed CHOPS data to find all the frequent itemsets in the processed data. Next, the association rules whose antecedents are reservoir property parameters and consequences are oil production parameters are generated, by applying the antecedent-consequence filter on the found frequent itemsets. As an option, users can continue to apply well identifier filter or location filter to search for the interesting rules by well information or locations. Finally, all interesting rules are outputted by the requirements of users.

3.3 Geovisualization of Interesting Association Rules

To achieve the objectives of visualization the visualization of the association rules in CHOPS well data, two novel geovisualization methods, point-based and surface-based geovisualization are proposed and introduced in this section. Point-based geovisualization conceptualizes an association rule with regards to well locations, and surface-based geovisualization builds the applicable areas for an association rule based on spatial interpolation techniques and then effectively represents the areas on the map.

3.3.1 Point-based Geovisualization Method

Point-based geovisualization uses different symbols to show the locations (points) of wells, depending on their relationship with association rules in CHOPS well data. The geospatial distribution of the wells associated with one association rule was visualized on the map with the following steps.

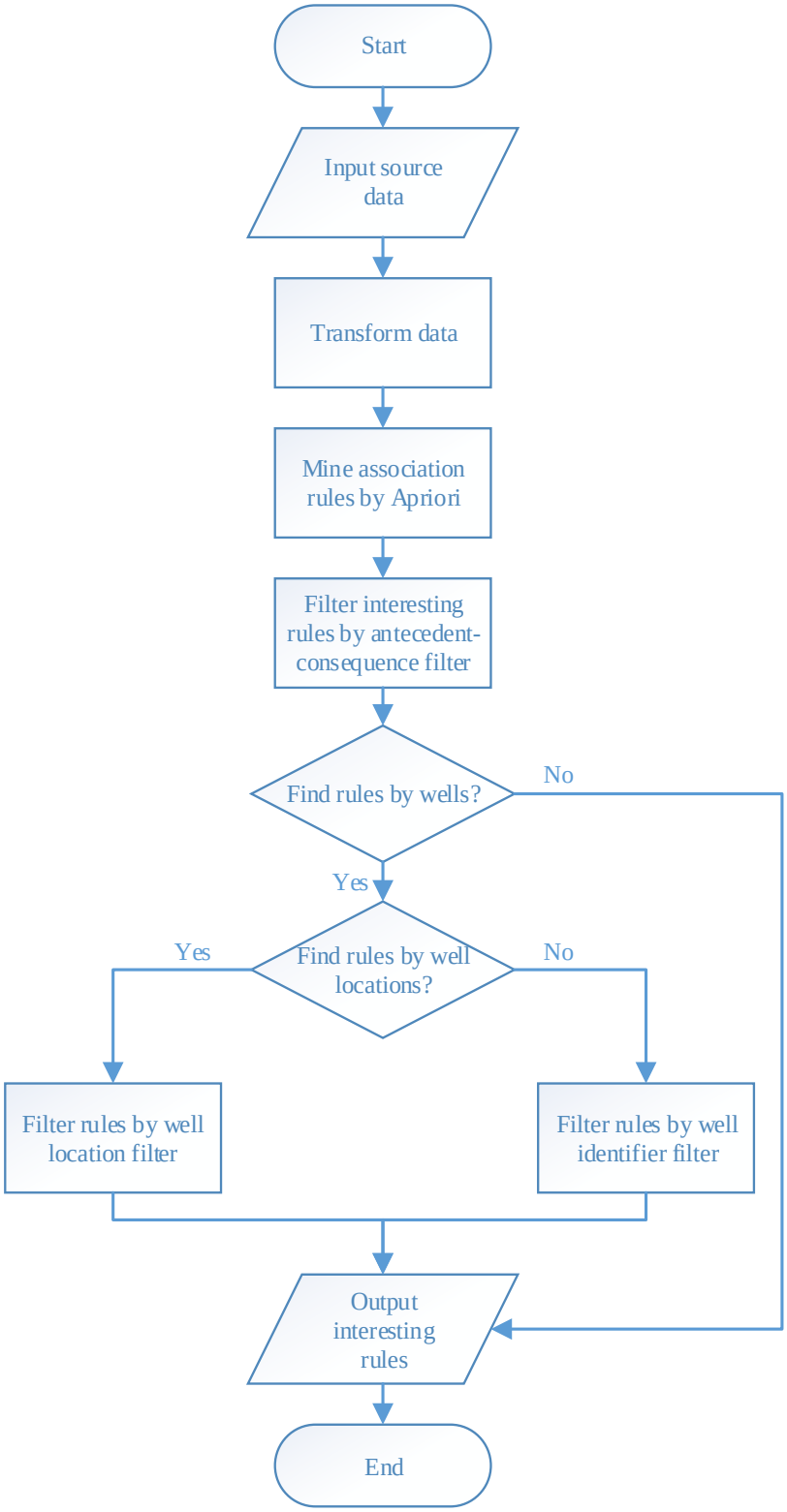


Figure 3-1 The flow chart of association rule mining process in CHOPS well data

First, all CHOPS oil wells were categorized into three groups according to the extent of satisfaction of the rule: the wells with the same ranges of reservoir properties and production performances described in the rule (i.e., satisfied both antecedents and consequences of the rule); the wells with the same range of reservoir properties described in the rule, but a different range of production performance (i.e., satisfied the antecedents but not the consequences of the rule); and, the wells with different reservoir properties and production performances than described in the rule (i.e., neither satisfied the rule). After categorization, the well locations of the three different groups were represented on the map using different symbols.

Compared to traditional visualization methods for association rules, point-based visualization emphases connecting the discovered rules with the well locations. Users (i.e., reservoir engineers) can easily identify the wells satisfying the rule and can reversely find rules associated with the well. Moreover, this categorization scheme can reflect the two traditional interestingness measurements (support and confidence) on the map. For the association rules in CHOPS well data, the support is the percentage of the wells in the whole dataset satisfy this rule while the confidence is the percentage of the wells among all wells that satisfy the reservoir property conditions in the antecedent. Therefore, the comparison and analysis of the geospatial distributions of the three groups of categorized wells on the map can help users in learning about the support and confidence of each discovered rule. Specifically, by comparing the wells that satisfy the association rule and those that do not satisfy the association rule, support of the rule can be understood and the wells that support the rule can be located on the map. Similarly, the confidence of the rule and the locations of the wells giving the confidence of the rule

can be identified by comparing the wells that satisfy the rule and those that only satisfy the antecedents and not the consequences of the rule.

3.3.2 Surface-based Geovisualization Method

Surface-based geovisualization extends the conceptualization of the rule from discrete points to continuous surfaces, with the aim of generating and representing areas where the rule may be applicable.

As shown in Figure 3-2, surface-based geovisualization of an association rule mainly includes the following steps. First, the spatial dependence of all the attributes appearing in the antecedents and consequences of the rule are examined. If the spatial dependence does not exist in the attributes, the surface-based geovisualization method is not applicable for the rule. Otherwise, continuous surfaces of the attributes are generated by applying a deterministic or stochastic spatial interpolation method on the well data. Next, the corresponding application areas of antecedents and consequences of the rule are extracted from the continuous surfaces and indicated by different colors. Finally, a prediction map of the rule is obtained by overlaying the application areas of the antecedents and consequences of the rule.

The existence of spatial dependence in the attributes in the studied area is the precondition for the use of spatial interpolation. The spatial dependence of each antecedent or consequence attribute of the association rule should be examined before using spatial interpolation. The spatial dependence of attributes can be checked by semi-variogram clouds. If spatial dependence of a reservoir property attribute exists in the

studied reservoir area, the semi-variance decreases with increasing spatial distance in the semi-variogram cloud of the attribute.

Figure 3-3 shows a semi-variogram cloud of the cumulative pore volume attribute in an area. Since the semi-variance increases as the spatial distance increases in the cloud, the semi-variogram cloud suggests that spatial dependence appears in the cumulative pore volume to some degree. Pairs of sample wells that are closer in distance have more similar values for the cumulative pore volume than well pairs that are farther apart.

Note that directional influences also need to be considered when generating semi-variogram clouds. The spatial dependence of an attribute can be stronger in specific directions. The directional influences may come from geological structures or a variety of other more complex processes. The directional influence of spatial dependence needs to be incorporated into the spatial dependence validation of each antecedent or consequence attribute of the association rule.

If spatial dependence of the antecedent and consequence attributes of the rule exists in the studied area, deterministic or stochastic interpolation methods are used to generate continuous surfaces for the attributes.

Under the assumption that the estimated value of an interpolation point should be influenced more by nearby control points than distant control points, spatial interpolation can fill in data between sample points. Spatial interpolation methods can be categorized into stochastic and deterministic methods.

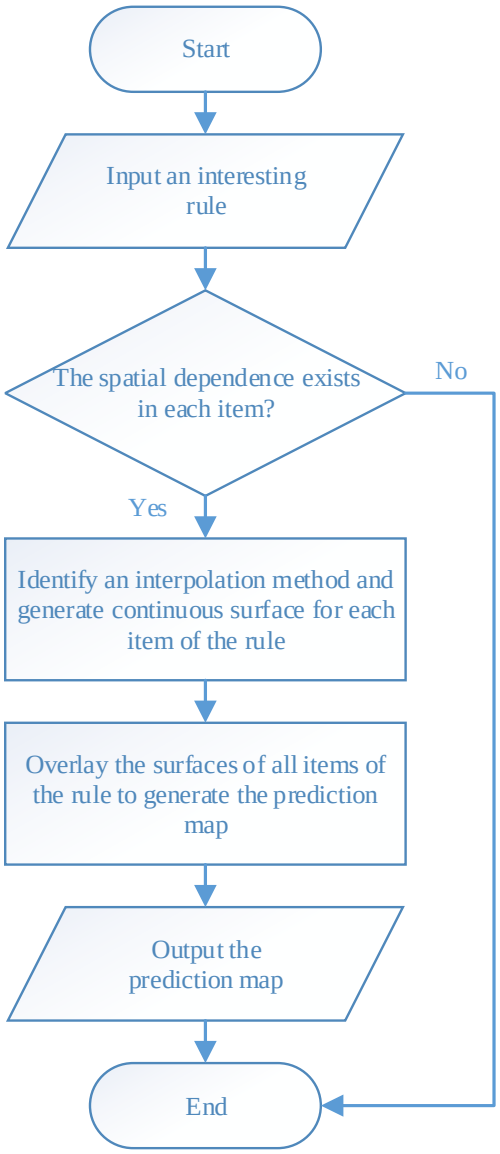


Figure 3-2 The flow chart of surface-based geovisualization

A stochastic interpolation method provides assessment of prediction errors by estimated variances in the form of prediction standard errors with interpolated values. Kriging interpolation is one of most common spatial stochastic interpolation methods. Not only it can interpolate the value of a certain attribute for an unknown (interpolation)

point with the known attribute values of its neighbor points, but it can also offer prediction errors with estimated values to assess the quality of the interpolation. In Kriging, the spatial variation of an attribute to be interpolated may be composed of a spatially correlated component and a drift component. The former represents the variation of the localized variable and the latter represents a trend and random error.

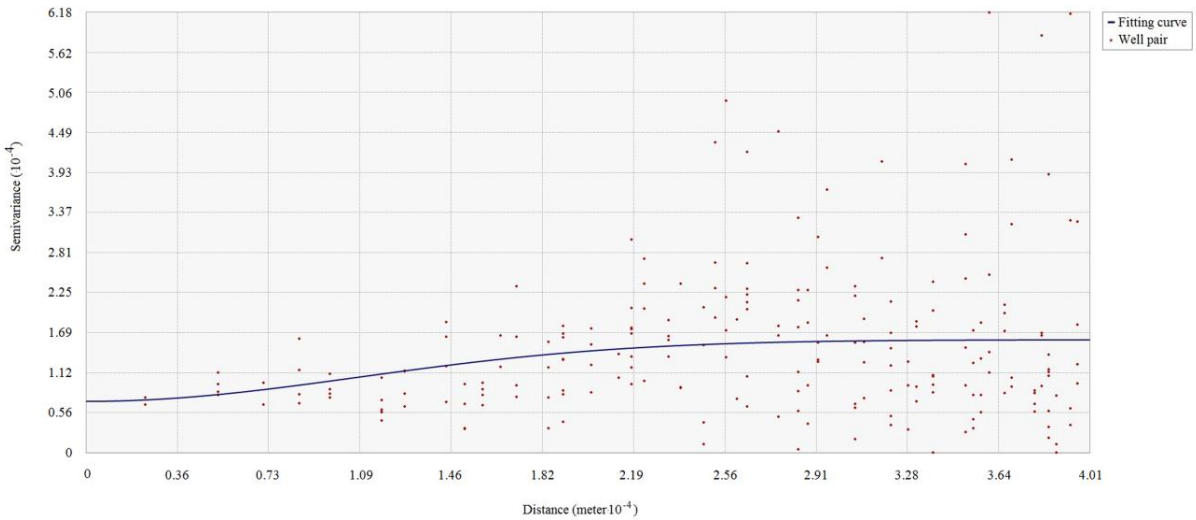


Figure 3-3 Semi-variogram clouds of the cumulative pore volume in an area

A deterministic interpolation method does not involve probability theory, thereby offering no assessment of errors with predicted values. Spline interpolation estimates values on the basis of a mathematical function where overall surface curvature was minimized, ending up with a smoother statistical surface. The surface passes exactly by the control points. In Inverse Distance Weighted (IDW) interpolation, a weight is assigned to each neighborhood point within a predefined radius for an interpolation point. The weight decreases as the distance from the interpolation point to its neighborhood points increases. The estimated value of the interpolation point is the weighted average of

its neighborhood points. Trend surface interpolation estimates the unknown values of interpolation points with a polynomial equation. The order of the polynomial equation can be adjusted according to the complexity of the specific situation.

CHOPS wells are represented as the point features discretely distributed on the map. Spatial interpolation then fills in the missing data between the wells based on the attribute values of the wells, i.e., the value of the attribute at a location with no recorded data can be estimated using the corresponding known value of the attribute of nearby sample CHOPS wells.

In terms of the format of the geospatial data, the spatial interpolation generates a raster layer with estimates made for all cells for each antecedent or consequence attribute of an association rule from a vector layer containing oil wells, where the value of each attribute is known. After this step, each attribute appearing in the association rule to be visualized will have an interpolated continuous surface.

The applicable areas of the antecedents and consequences of the rule are then extracted and rendered from the surface of each attribute. For instance, the applicable areas of an antecedent of an association rule (e.g., cumulative pore volume = 59.2% ~ 94.4%) can be gained by extracting the cells whose interpolated values belong to the range from 59.2% to 94.4% from the interpolated continuous surfaces of the cumulative pore volume.

Finally, a prediction map of the applicable areas of the rule is obtained by overlaying all interpolated continuous surfaces of the attributes appearing in the rule.

3.4 CHOPSData-GeoViz System Prototype

To efficiently manage, mine and visualize association rules in CHOPS well data, a system prototype, called CHOPSData-GeoViz, was developed. In the following, the architecture of CHOPSData-GeoViz prototype and different components in the architecture are introduced.

3.4.1 CHOPSData-GeoViz Prototype Architecture

CHOPSData-GeoViz prototype includes four main components (Figure 3-4): a well database, association rule mining function, geovisualization function, and graphical user interface (GUI). The GUI, association rule mining, filtering, geovisualization were developed using C# programming language with integration of the ESRI ArcObjects API. The CHOPSData-GeoViz prototype also supports diverse data formats (*.mxd map file, *.lyr layer file, *.shp shape file, *.mdb geodatabase file) and displays them on the map. The following section describes the main components of the prototype in detail.

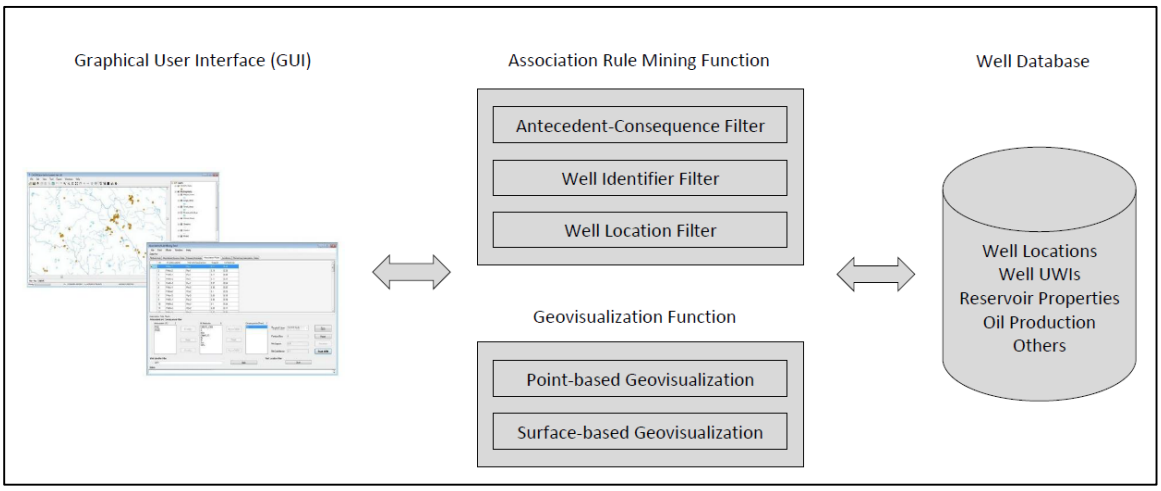


Figure 3-4 Architecture of the CHOPSData-GeoViz prototype

Well Database: The well database in the CHOPSData-GeoViz prototype was implemented with Microsoft SQL Server 2005 and the ArcObject's GeoDatabase Library. The well database contains both the spatial and non-spatial data. The spatial data include the well locations (i.e., longitude and latitude) and spatial objects representing wells (i.e., points). The non-spatial data include the unique well identifier (UWI), reservoir properties and oil production data. The UWI is unique to each well and is used as the primary key in the database. The non-spatial data are connected with the spatial data using the primary key of UWI.

Association Rule Mining Function: The association rule mining function includes the association rule mining and filtering sub-functions. The continuous values of the reservoir property and oil production performance parameters in the source CHOPS well data are firstly discretized before association rule mining by data transformation function. The data transformation function transforms the data format from numerical to nominal. The current version of CHOPSData-GeoViz prototype implements the Apriori algorithm. A group of attributes from selected wells can be assigned to the Apriori analysis. After transforming the data, Apriori analyzes and presents the discovered association rules among the attributes. The association rule filtering function in the prototype also includes three filters introduced in the previous sections: the antecedent-consequence filter, well identifier filter and well location filter.

Geovisualization Function: The interesting rules can be visualized on the map by the geovisualization function. The geovisualization function include point- and surface-based sub-functions. In point-based sub-function, for each selected interesting association rule IF $\{X\}$ THEN $\{Y\}$, the system queries the database to find the wells

whose records satisfy both the antecedent *X* and the consequence *Y*, and the wells whose records satisfy only the antecedent *X*. Then the system highlights the wells on the map. The application areas of each interesting rule are generated by the surface-based visualization function. All of these functions are implemented by calling APIs from ESRI ArcObjects.

Graphical User Interface (GUI): Figure 3-5 shows the main GUI of CHOPSData-GeoViz system prototype. The main interface consists of map display area, table of layer content, eagle eye window, menu and tool bar. The map display area in the middle of the interface shows the visualization of interesting rules with the designated map scale and coordinates. The layer table on the right of the interface shows the map layers. The eagle eye window shows a global view of the current map. The top of the interface contains the menu and tool bar where association rule mining and geovisualization functions can be accessed.

Through the main interface, the user can execute the association rule mining and geovisualization by clicking the Association Rule Mining and Geovisualization Function button on the tool bar. Figure 3-6 shows the user interface of the association rule mining and geovisualization. From top to bottom, there are the Menu Bar, Data & Result Viewer, Association Rule Mining, Rule Filters, and Message Box. Geovisualization function including point- and surface-based methods can be launched under the Menu Bar. In the filters, users can screen out interesting rules from all of the resulting association rules generated and listed in the Data & Result Viewer, by setting or selecting the antecedents and the consequences, unique well identifiers, and well locations. Next, users can select point- or surface-based methods under Geovisualization and then click each individual

interesting rule in the Data & Result Viewer to obtain the geovisualization results on the map in the main interface in Figure 3-5.

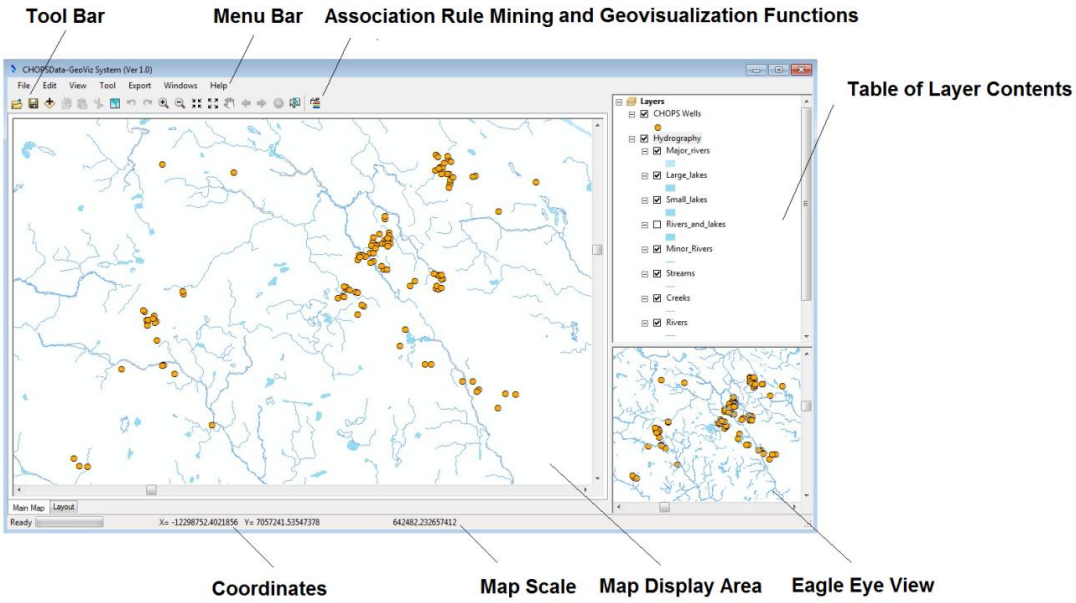


Figure 3-5 The main GUI of the CHOPSData-GeoViz prototype

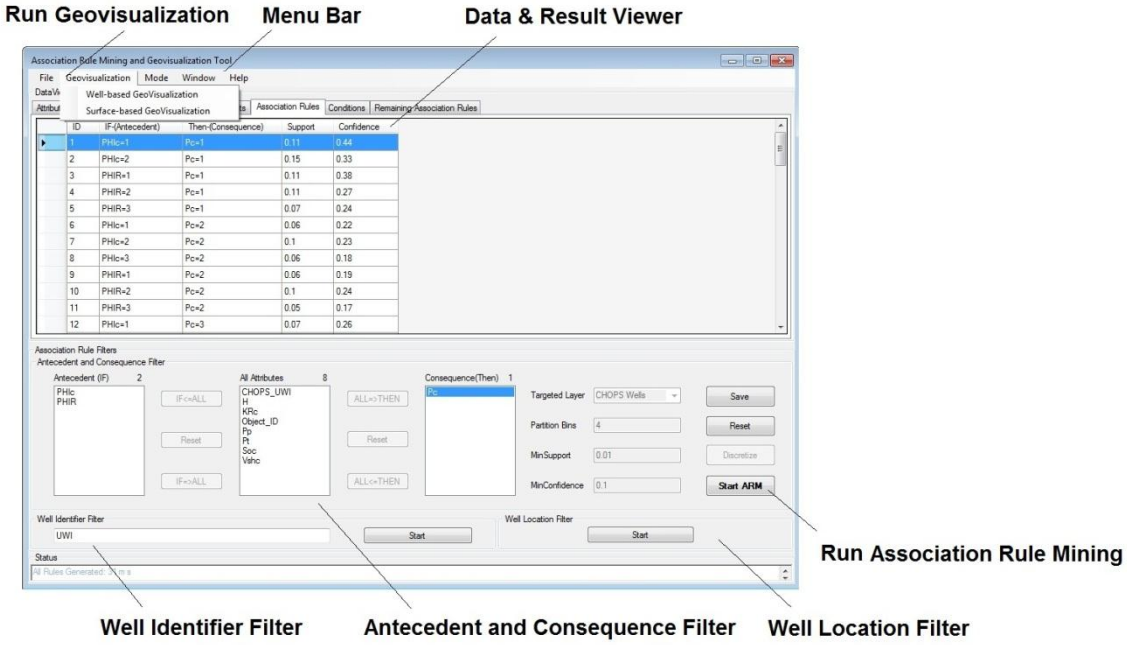


Figure 3-6 The GUI of the association rule mining and geovisualization functions

3.5 Summary

In this Chapter, several rule filters are firstly introduced to help reservoir engineers effectively discover interesting rules from association rules generated in the CHOPS well data. By the rule filters, the interesting rules representing the influence of the reservoir properties on production and matching some specific wells will be extracted. On top of rule filters, two geovisualization methods – point-based and surface-based – are proposed for the interesting rules. The point-based geovisualization method links association rules with well locations. The method uses different symbols to show the well locations that satisfy the rules. Reservoir engineers can understand and study rules that they are interested in from the map. The surface-based geovisualization method generates the areas where an interesting association rule may be applicable based on spatial interpolation. The applicable areas on the map generated by surface-based geovisualization can assist reservoir engineers in making predictions based on the patterns represented by the relevant rules. Finally, a system prototype called CHOPSData-GeoViz is developed and its architecture and user interfaces of CHOPSData-GeoViz prototype are introduced in the chapter. The system prototype integrates association rule mining, filtering and visualization to efficiently mine and visualize association rules in CHOPS well data.

Chapter Four: **A Case Study**

In this chapter, a case study was carried out on real CHOPS well data for the Lloydminster heavy oil block in Alberta, Canada.

4.1 Data Collection and Description

The Lloydminster heavy oil block is a large reservoir zone located in the central eastern part of the province of Alberta in Canada. In the block, more than 3000 wells have been drilled. One hundred and eighteen CHOPS wells were selected from the block based on the following selection criteria: (1) drilling date between 1992 and 2005; (2) vertical well; and, (3) one perforation formation. The distribution of the studied 118 CHOPS wells of the area is shown in Figure 4-1.

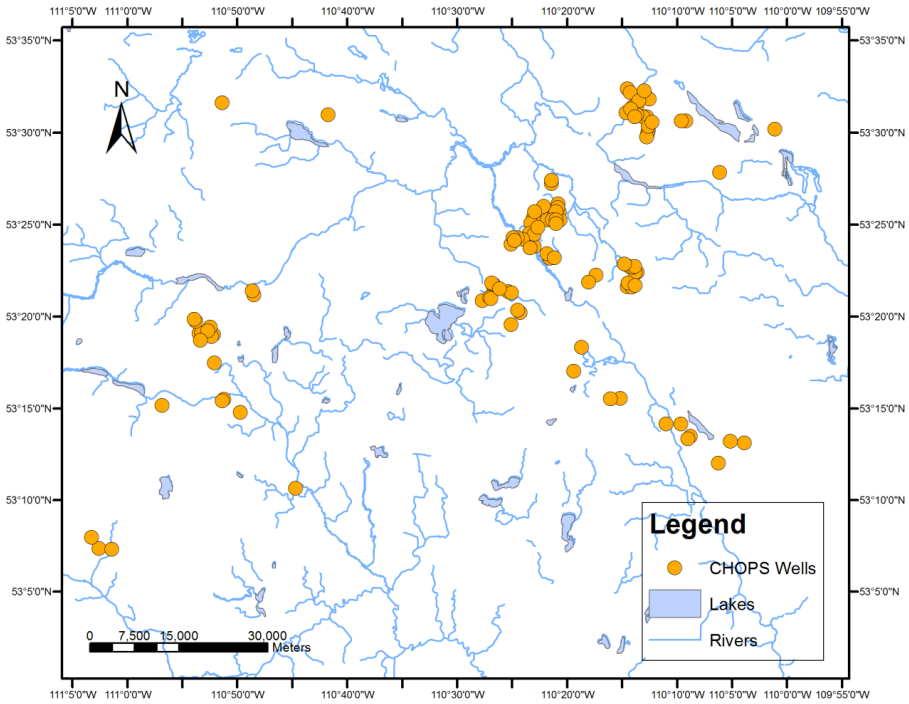


Figure 4-1 Geospatial distribution of the studied 118 wells in Alberta, Canada

The source data of the studied CHOPS wells in the case study were collected and transformed from the Canadian Geo-service company Divestco Inc. Each well record in the data contained reservoir property and oil production performance parameters of the well, and each well had its geospatial location in longitude and latitude coordinates. Table 4-1 shows the some sample records of well locations. Table 4-2 shows some sample reservoir property and production records from the source dataset.

Table 4-1 Sample records of well locations in source dataset

UWI	Latitude	Longitude	UTM-*x*	UTM-*y*
00011404808W40	53.13296	111.0551	496316.3317	5887063.037
00012705103W40	53.42623	110.3526	543015.0584	5919881.409
00012805202W40	53.51453	110.2314	550963.1614	5929785.506
00020305203W40	53.45412	110.3575	542665.2424	5922981.317
00021604901W40	53.22049	110.0867	560980.9436	5897187.764
00021605103W40	53.39740	110.3839	540966.8985	5916655.758

The reservoir property parameters used for association rule mining were basic petro-physical parameters of a reservoir that can be used to identify to describe the characteristics of a reservoir: cumulative porosity, cumulative pore volume, cumulative shale content, cumulative oil saturation, cumulative fluid mobility factor, and effective thickness.

Table 4-2 Sample records of reservoir property and oil production in the source dataset

UWI	Cumulative porosity (%)	Cumulative pore volume	Cumulative shale content (%)	Cumulative oil saturation (%)	Cumulative fluid mobility factor	Effective thickness (m)	Effective yield (m^3)	Peak value (m^3/d)	Effective life cycle (days)
100011404808W400	8.398741362	96.23524378	2.071810184	21.25729863	38.18490774	29	0	3.131025	0
100012705103W400	17.433315	94.3797535	4.080026238	43.56617361	142.7010683	59	793.42	21.31411	64
100012805202W400	15.08544664	90.72664589	3.943412344	38.56685157	80.59311485	48	92.06	7.159461	8
100020305203W400	12.78071559	84.06041891	1.520339702	27.75209989	88.32543179	40	107.58	4.313461	0
100021604901W400	12.27264157	101.9466699	1.85604121	30.88900622	107.6644539	40	18.33	6.389804	5
100021605103W400	20.91777291	111.4668523	1.402918955	43.37780694	128.3114458	65	194.89	8.286773	27

The porosity influences the amount of trapped fluids and the rate at which the fluids flow during main production. A reservoir with higher porosity has more pore space to hold more fluid. The pore volume represents the portion of the pore volume of the main contributor and plays a major role in production performance. The shale content (volume of shale) is an indicator of reservoir quality. A lower shale content usually indicates a better caliber of reservoir.

The oil saturation and effective thickness are also important parameters that contribute directly to reservoir reserves. In permeable formations, mud filtrate invades the formation and forms a zone during well drilling, which results in changes of resistivity along the horizontal zone. The fluid mobility factor is a parameter that denotes the variation of the resistivity. The cumulative porosity, cumulative pore volume, cumulative shale content, cumulative oil saturation, and fluid mobility factor can be calculated from the basic petro-physical parameters of samples in the net pay. Pay is a reservoir or portion of a reservoir that includes economically producible hydrocarbons, and the smaller portions of the gross pay that meet local criteria for pay are considered to be the net pay (Cai et al., 2014).

Oil production is generally described by cumulative production and daily production. Due to different lengths in production performance, daily oil production data were used in this study. However, the daily production varied over time, increasing once production started, reaching and stabilizing at the peak production rate, and then declining to uneconomic levels.

To analyze the main features of production performance, a polynomial fitting curve of the daily production data that denoted production performance of the well was created. Three production parameters were used to characterize the fitting curve: peak value, effective life cycle and effective yield (Cai et al., 2014). As shown in Figure 4-2, the peak value is the highest value point of the curve. The effective life cycle is the time of the production fitting curve above a certain threshold, usually determined by a combination of factors including technical, political and economic factors. The effective yield is the cumulative production in the whole effective life cycle. In this paper, these three parameters are used to describe production performance for a well. The depth of pay zone was between 400 m and 700 m, and 4.5m^3/d was used as the production rate threshold value.

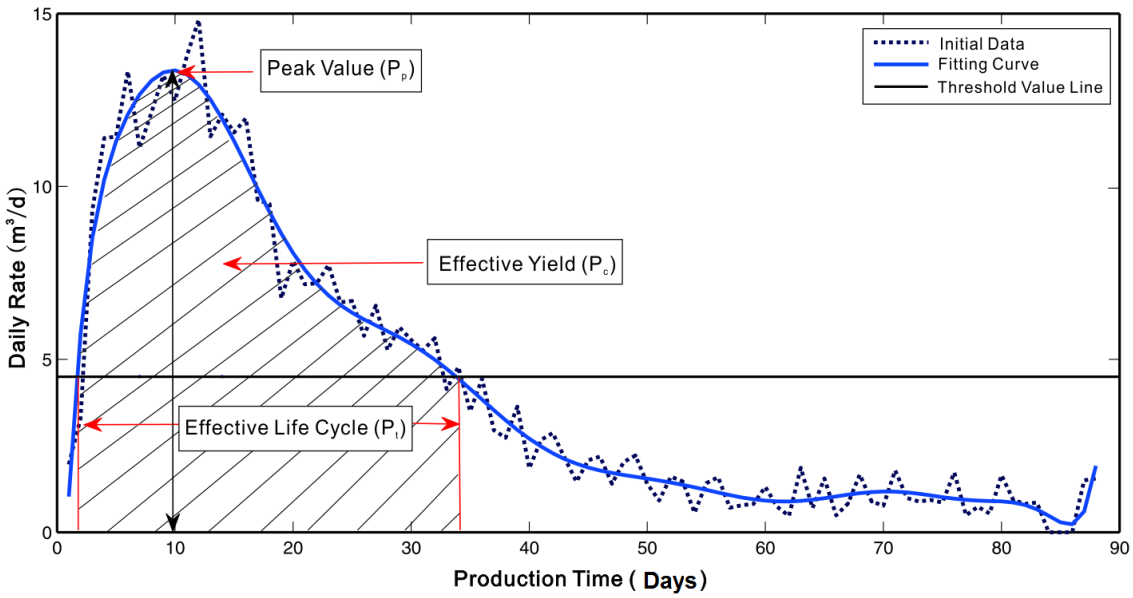


Figure 4-2 Three parameters used to characterize oil production performance of CHOPS wells

4.2 Data Transformation

The essential preprocessing work before association rule mining is the transformation of the values of reservoir property and oil production performance parameters into a set of sub-ranges through discretization schemes. The reduction of detail in data results can make the mining process more efficient and the patterns more accessible.

The values of the three yield performance parameters were clustered into four categories with the *k*-means algorithm. The clustering results are listed in Table 4-3. A one-dimensional self-organizing map was used to discretize the values of six other reservoir property parameters into four categories, as shown in Table 4-4. We use the self-organizing map and k-means techniques because they keep the distribution of the attributes and found rules can be more intuitive (Marco and Valentina, 2004). Therefore, in this case study, the self-organizing map was used to discretize the selected variables of each reservoir properties; and, the *k*-means technique was utilized for clustering the production data so that the clustered results could be explained easily by the cluster centroids. The *k*-means method is mainly influenced to the value of k. So it must be fixed in addition of the computation. For the self-organizing map, only the maximum number of desired intervals must be confirmed (Marco and Valentina, 2004).

4.3 Association Rule Mining and Interesting Rule Filtering

After data transformation, the association rule mining algorithm Apriori was applied on the data. This procedure resulted in rules that had support and confidence values greater than the specified thresholds. In this case study, the minimum support and confidence

values were set at 2% and 30%, respectively. The preprocessed data generated 3812 association rules. Table 4-5 listed some of the discovered association rules. Reservoir engineers may not be interested in every association rule because it is more interesting for them to understand how reservoir properties influence the oil production. For example, the first rule in the table IF { Effective yield = 1 ($0m^3$ ~ $104.3m^3$) } THEN { Effective thickness = 1 (0.4m ~ 3.5m) } meant that if the effective yield was between $0m^3$ ~ $104.3m^3$, then the effective thickness would be 0.4m ~ 3.5m. However, usually the oil production performance is caused by different factors. Effective thickness is one of the reservoir properties, which is from the well log analysis. So, the rule did not represent the influence of reservoir properties on oil production and thus should be removed.

Table 4-3 Discretization results of the oil production performance parameters

Cluster No.	Cluster centroids of k-means clustering			No. of wells	Discretized value
	Effective yield	Peak value	Effective life cycle		
Cluster 1	$0m^3$ ~ $104.3m^3$	$0.8m^3/d$ ~ $6.7m^3/d$	0day ~ 8days	28	1
Cluster 2	$104.3m^3$ ~ $467.3m^3$	$6.7m^3/d$ ~ $8.2m^3/d$	8days ~ 25days	30	2
Cluster 3	$467.3m^3$ ~ $808.2m^3$	$8.2m^3/d$ ~ $14.2m^3/d$	25days ~ 50days	40	3
Cluster 4	$808.2m^3$ ~ $1452.9m^3$	$14.2m^3/d$ ~ $37.8m^3/d$	50days ~ 144days	20	4

Table 4-4 Discretization results of the reservoir property parameters

Cumulative porosity		Cumulative pore volume	
Range	Discretized value	Range	Discretized value
1.1% ~ 10.8%	1	11.9 ~ 57.5	1
10.9% ~ 17.4%	2	59.2 ~ 94.4	2
17.5% ~ 25.2%	3	95.9 ~ 154.3	3
25.3% ~ 38.5%	4	168.8 ~ 291.2	4
Cumulative fluid mobility factor		Cumulative oil saturation	
Range	Discretized value	Range	Discretized value
7.5 ~ 61.4	1	2.8% ~ 20.6%	1
63.7 ~ 92.5	2	21.3% ~ 33.1%	2
96.4 ~ 128.6	3	33.9% ~ 49.5%	3
132.3 ~ 187.0	4	55.0% ~ 85.6%	4
Cumulative shale content		Effective thickness	
Range	Discretized value	Range	Discretized value
0.5% ~ 1.9%	1	0.4m ~ 3.5m	1
1.9% ~ 2.9%	2	3.6m ~ 5.2m	2
3.0% ~ 4.1%	3	5.3m ~ 7.1m	3
4.2% ~ 6.2%	4	7.8m ~ 11.0m	4

The next task for reservoir engineers is to find out the possibly interesting rules satisfying certain antecedent and consequence conditions from the original 3812 association rules. As mentioned in the problem statement section, one of the objectives of

applying association rule mining is to discover the relationship between reservoir properties and production. So the interesting rules must at least represent the influence of the former on the latter. For example, the second association rule in Table 4-5 is interesting and valuable; because it shows the relationship between reservoir properties and oil production, that is, if cumulative pore volume is between 11.9 and 57.5 then the effective yield would be $0m^3$~$104.3m^3$. This kind of interesting rules with specific antecedents and consequences was screened out by the antecedent-consequence filter.

For another example, we supposed a reservoir engineer who wants to study the influence of the reservoir property parameter, cumulative pore volume, on the oil production parameter, effective yield. By respectively setting cumulative pore volume and effective yield as the antecedent and consequence, the antecedent-consequence filter finds the interesting rules. Some of them were sorted and listed in Table 4-6. The rules reflect the quantitative relationship of cumulative pore volume and effective yield. The effective yield increases as the as the cumulative pore volume grows. The antecedent-consequence filter undoubtedly effectively assisted in filtering out the interesting rules between reservoir properties and oil production and discovering quantitative relationship between the two and some hidden trends would be possibly discovered with domain knowledge from the interesting rule selected by the antecedent-consequence filter.

Table 4-5 Some discovered association rules

No.	Antecedent (IF)	Consequence (THEN)	Support	Confidence
1	Effective yield = 1 ($0m^3$~$104.3m^3$)	Effective thickness = 1 (0.4m ~ 3.5m)	11% (13/118)	83% (13/15)
2	Cumulative pore volume = 1 (11.9~57.5)	Effective yield = 1 ($0m^3$~$104.3m^3$)	11% (13/118)	83% (13/15)
3	Effective yield = 3 ($467.3m^3$ ~ $808.2m^3$)	Effective thickness = 3 (0.4m ~ 3.5m)	11% (13/118)	39% (13/33)
4	Cumulative oil saturation = 2 (21.3% ~ 33.1%)	Effective yield = 1 ($0m^3$~$104.3m^3$)	15% (18/118)	37% (18/49)
5	Cumulative pore volume = 1 (11.9~57.5)	Effective yield = 1 ($0m^3$~$104.3m^3$)	11% (13/118)	83% (13/15)
6	Effective yield = 4 ($808.2m^3$ ~ $1452.9m^3$)	Cumulative mobility factor = 4 (132.3 ~ 187.0)	10% (12/118)	34% (12/35)
7	Cumulative pore volume = 4 (168.8~291.2)	Effective yield = 4 ($808.2m^3$~$1452.9m^3$)	6% (7/118)	58% (7/12)
8	Cumulative shale content = 1 (0.5% ~ 1.9%)	Effective yield = 1($0m^3$~$104.3m^3$)	9% (11/118)	26% (9/35)
9	Cumulative pore volume = 2 (59.2~94.4)	Effective yield = 2 ($104.3m^3$~$467.3m^3$)	10% (12/118)	64% (12/19)
10	Cumulative porosity = 3 (17.5% ~ 25.2%) and Cumulative oil saturation = 3 (33.9% ~ 49.5%)	Effective yield = 3 ($467.3m^3$ ~ $808.2m^3$)	11% (13/118)	38% (13/34)

Table 4-6 Some interesting rules related to cumulative pore volume and effective yield

No.	Antecedent (IF)	Consequence (THEN)	Support	Confidence
1	Cumulative pore volume = 1 (11.9~57.5)	Effective yield = 1 ($0m^3$~$104.3m^3$)	11% (13/118)	83% (13/15)
2	Cumulative pore volume = 2 (59.2~94.4)	Effective yield = 2 ($104.3m^3$~$467.3m^3$)	10% (12/118)	64% (12/19)
3	Cumulative pore volume = 3 (95.9~154.3)	Effective yield = 3 ($467.3m^3$~$808.2m^3$)	9% (10/118)	82% (10/12)
4	Cumulative pore volume = 4 (168.8~291.2)	Effective yield = 4 ($808.2m^3$~$1452.9m^3$)	6% (7/118)	58% (7/12)

On top of the antecedent-consequence filter, the well identifier filter further screened out the association rules by the identifiers of wells from the rules representing between reservoir properties and oil production. For example, with a well identifier (UWI) “100023604806W400”, the well identifier filter discovered the association rule IF { cumulative pore volume = 4 (168.8 ~ 291.2) } THEN { effective yield = 4 ($808.2m^3$ ~ $1452.9m^3$) } from the rules representing the quantitative relationship of cumulative pore volume and effective yield such as the ones in Table 4-6. The records of the well in the database satisfied the rule. Combined with point-based geovisualization, the filter could help users find other rules that they were interested in. For instance, users could use the point-based geovisualization function on this rule to see where the other wells whose records in the database also satisfied this rule located and what the identifiers of the wells

were. Then the identifiers of the wells could be used iteratively as the inputs of the well identifier filter and resulted in more interesting rules.

In addition to well identifier filter, the interesting rules could be extracted by selecting wells on the map with the well location filter. Suppose a reservoir engineer was interested in the association rules matching the wells located in the area shown in Figure 4-3. There were 19 wells in the selected area. The well location filter found out two interesting rules (listed in Table 4-7) from the rules representing the quantitative relationship of cumulative pore volume and effective yield. The first rule IF {cumulative pore volume = 2 (59.2~94.4) } THEN { effective yield = 2 (104.3m^3 ~ 467.3m^3) } matched the wells, 100012805202W400 and 102082805202W400, in the selected area. The other rule IF { cumulative pore volume = 3 (95.9~154.3) } THEN { effective yield = 3 (467.3m^3 ~ 808.2m^3) } whereas matched the wells, 102022205202W400 and 102152805202W400 in the selected area. By the antecedent-consequence filter, the well location filter helped discover two association rules from the association rules listed in Table 4-7 representing the relationship between cumulative pore volume and effective yield discovered. On the other hand, the identifiers of the four wells discovered by the well location filter could also be iteratively reused as the inputs of the well identifier filter to find other interesting rules. The user could also further run point-based geovisualization on the rules discovered by the well identifier filter, to obtain the locations and identifiers of the other wells that also satisfied the rules. Applying the well identifier filter on identifiers of the wells would result in other possible interesting rules.

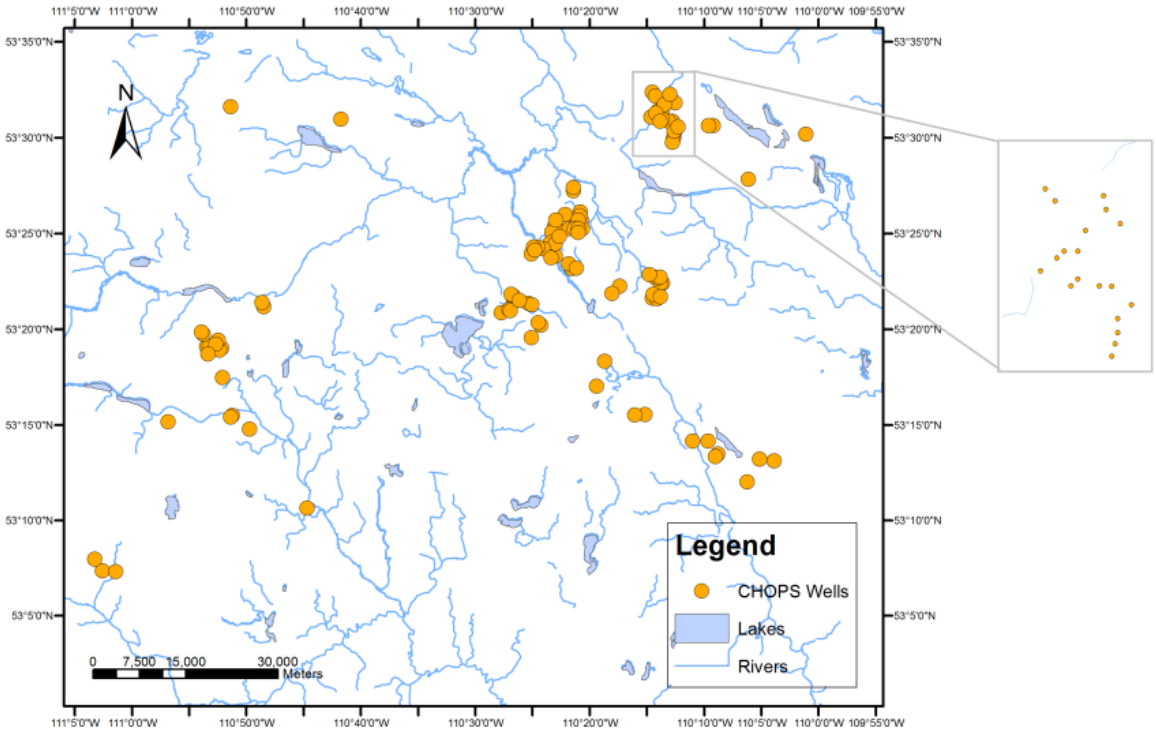


Figure 4-3 An area selected for running well location filter

Table 4-7 Some interesting association rules discovered by the specific well locations

UWIs	Association rules	
	Antecedent (IF)	Consequence (THEN)
100012805202W400 102082805202W400	Cumulative pore volume = 2 (59.2~94.4)	Effective yield = 2 ($104.3m^3$~$467.3m^3$)
102022205202W400 102152805202W400	Cumulative pore volume = 3 (95.9~154.3)	Effective yield = 3 ($467.3m^3$~$808.2m^3$)

4.4 Geovisualization Results of Some Association Rules

In this section, some interesting association rules discovered were visualized on the map by the proposed point- and surface-based geovisualization methods.

4.4.1 Point-based Geovisualization Results and Interpretations

Through an accessible map, the point-based geovisualization method for interesting association rules offers effective communication on the rules and the spatial distribution of the wells. For example, by association rule filters, reservoir engineers discovered the interesting association rule in Table 4-6:

IF {cumulative pore volume = 168.8 ~ 291.2} THEN {effective yield = 808.2 m^3 ~ 1452.9 m^3}.

This association rule represented the relationship between the cumulative pore volume and the effective yield and indicated that high cumulative pore volume might cause high effective yield. Point-based geovisualization represented the relationship in terms of the wells, by demonstrating the association rule that suggested the relationship on the map with the locations of the wells.

Firstly, the selected 118 CHOPS wells were classified into three classes based on the extent of satisfaction with the association rule: wells that had cumulative pore volumes and effective yields that satisfied the rule (totally satisfying the rule); wells that had cumulative pore volumes that satisfied the rule but had effective yields that did not satisfy the rule (partially satisfying the rule); and, wells that had cumulative pore volumes and effective yields that did not satisfy the rule (not satisfying the rule).

The corresponding visualization result in Figure 4-4 of the above association rule (i.e., the rule in Table 4-6) can be interpreted as follows. The locations of the three well classes were highlighted on the map using circle symbols with different colors: red, yellow and green respectively represented totally satisfying, partially satisfying and not satisfying the rule classifications.

The point-based visualization result of the rule in Figure 4-4 shows that there were seven wells that totally satisfy the association rule, i.e., the cumulative pore volume values of the wells were in the range of 169.8 to 291.2 and the effective yield values of the wells fell into the range of 808.2 m^3 to 1452.9 m^3. Also, the wells were mainly distributed in the central east of the studied reservoir area (as shown in Figure 4-4 (a)). The point-based geovisualization bridged the association rule and the geospatial distribution patterns of the wells satisfying the rule. Note that the seven wells satisfying the rule were located close together (Figure 4-4 (a)). Reservoir engineers may further discover some other interesting rules satisfying the seven wells or nearby wells by inputting the identifiers of the wells into well identifier filter. It may also lead to further study of the reason for the geospatial distribution patterns of wells satisfying the rule.

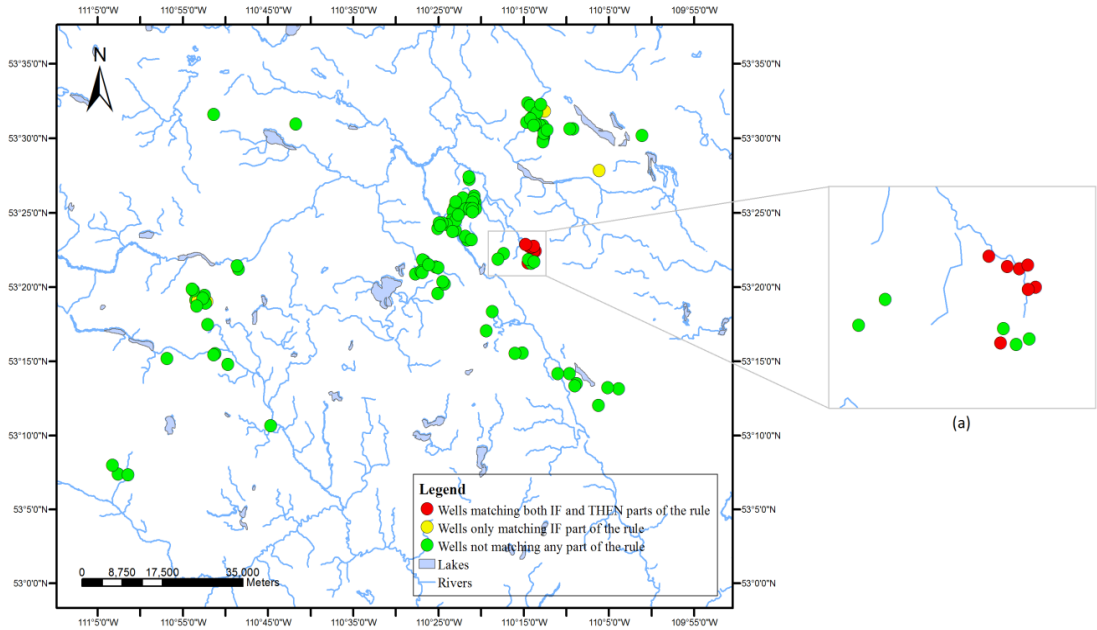


Figure 4-4 Visualization of rule IF {cumulative pore volume = 168.8 ~ 291.2} THEN {effective yield = 808.2 m^3 ~ 1452.9 m^3} generated by point-based geovisualization

4.4.2 Surface-based Geovisualization Results and Interpretations

The potential areas where an interesting association rule may happen are very valuable for making predictions based on the patterns within the rule. Surface-based geovisualization can be used to predict and visualize the applicable areas for association rules discovered from the CHOPS well data.

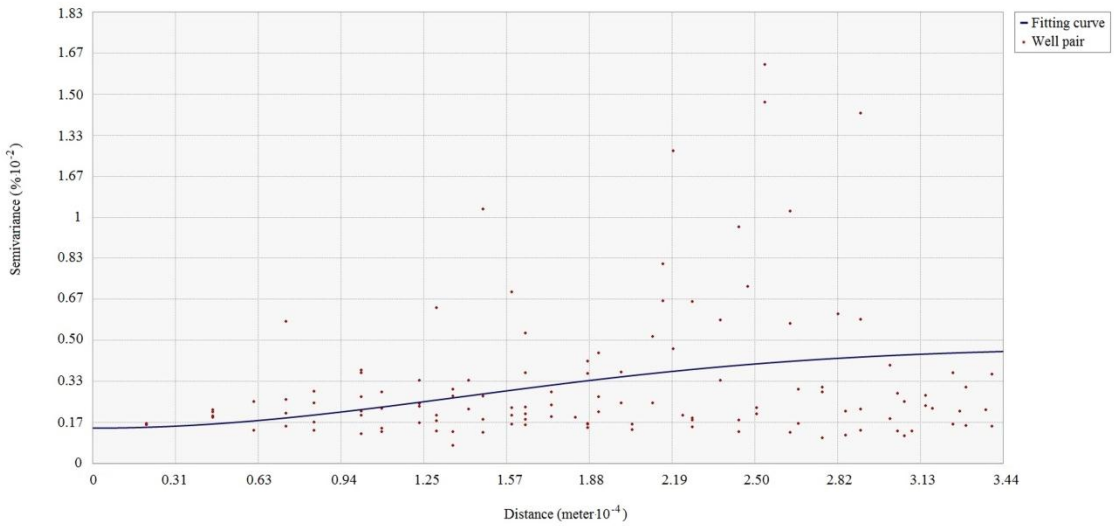
The following rule is used as an example to illustrate the process of the surface-based geovisualization:

IF {cumulative porosity = 25.3% ~ 38.5%} THEN {effective yield = 808.2 m^3 ~ 1452.9 m^3}.

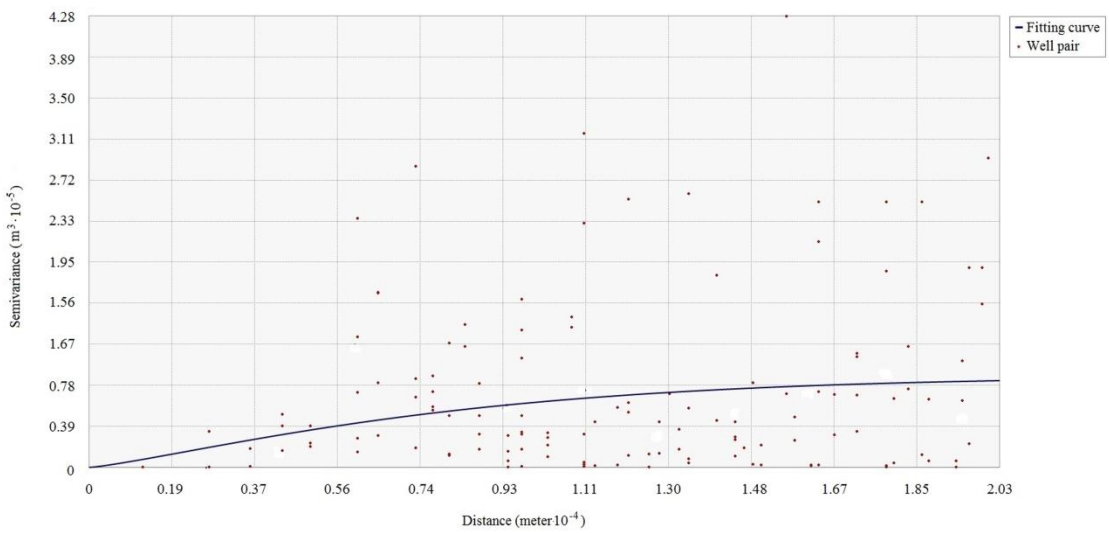
4.4.2.1 Spatial Dependence Analysis for Source CHOPS Data

The first step of the surface-based geovisualization requires that the spatial dependence of cumulative porosity and effective yield in the rule be examined. Figure 4-5 shows the semi-variogram clouds of these two attributes in the CHOPS well data used by the case study. The semi-variogram clouds were generated in the directions of 37.62^{o} and 91.25^{o}, where spatial dependence of the attributes was the strongest.

The semi-variogram clouds suggested that spatial dependence existed in the cumulative porosity and effective yield attributes of the 118 sample CHOPS wells, since the semi-variance increased as the distance increased, i.e., closer well pairs had more similar values for the cumulative porosity and effective yield than well pairs that were farther apart.



(a)



(b)

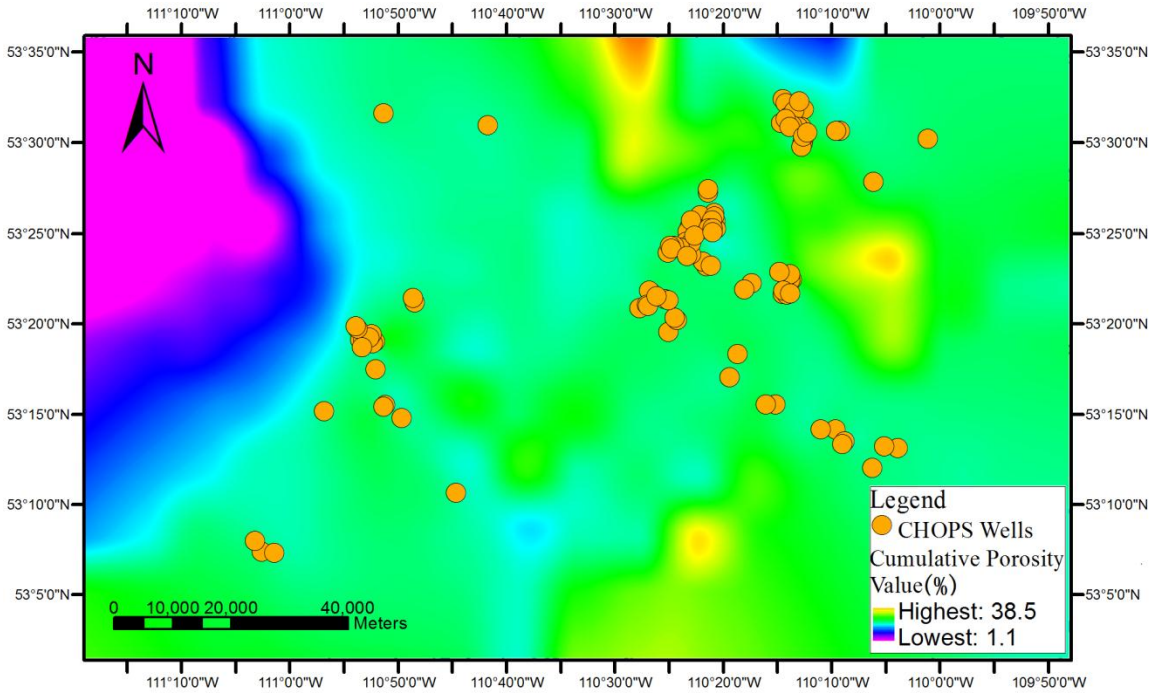
Figure 4-5 Semi-variogram clouds of (a) cumulative porosity and (b) effective yield of the CHOPS data

4.4.2.2 Surface-based Geovisualization based on Deterministic Spatial Interpolation

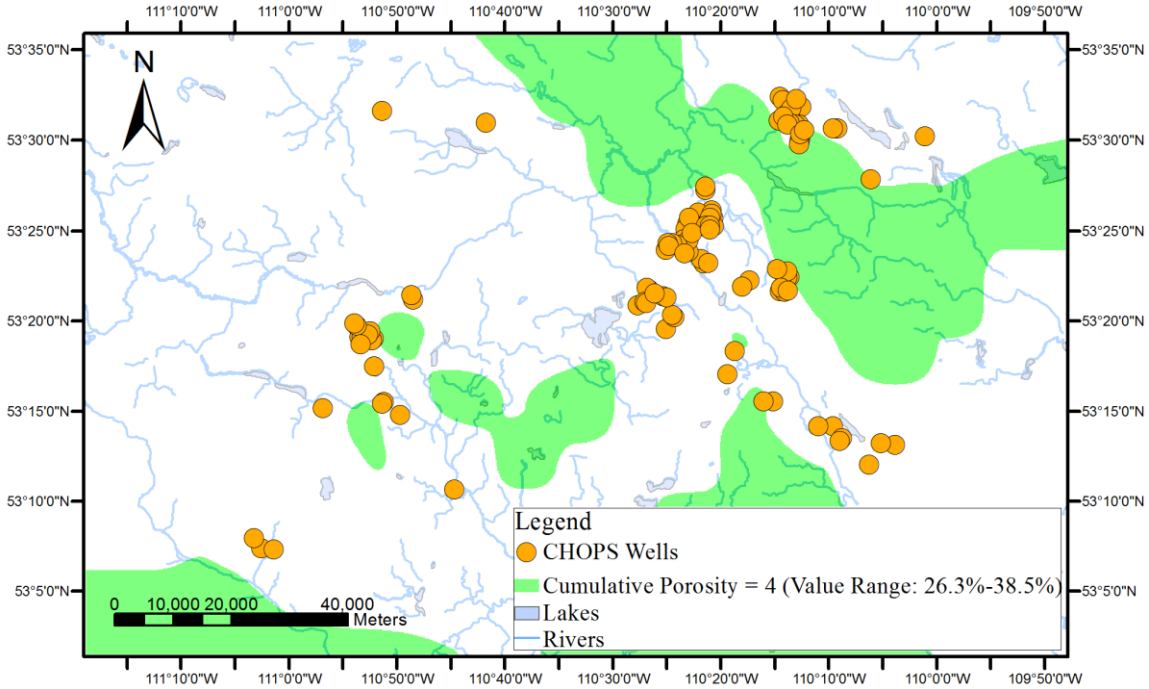
Since the spatial dependence of the cumulative porosity and effective yield attributes held for the area of 118 CHOPS wells, deterministic spatial interpolation methods were used to build continuous applicable areas for association rules.

The values of the cumulative porosity for the whole study area can be estimated using the corresponding known attribute values at nearby CHOPS wells by interpolation. Figure 4-6 (a) shows the gradient map of the cumulative porosity generated by applying the Spline interpolation on the source CHOPS well data. The applicable areas of the antecedent of the rule, i.e., cumulative porosity = 25.3% ~ 38.5%, were extracted from the gradient map according to the corresponding discretization results in Table 4-4 and indicated using the green color as shown in Figure 4-6 (b). Similarly, the gradient map of the effective yield and the applicable areas of the consequence of the rule, i.e., effective yield = 808.2 m^3 ~ 1452.9 m^3, were extracted and indicated by the red color, as shown in Figure 4-6 (c) and Figure 4-6 (d).

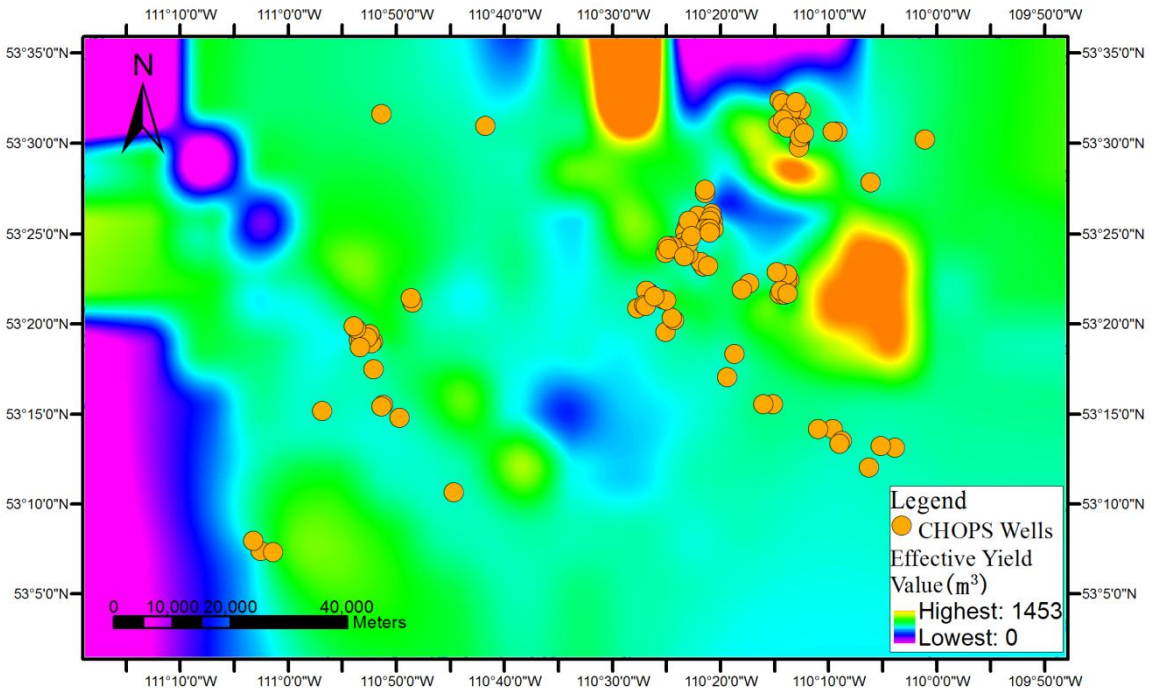
Finally, the prediction map of the rule can be obtained by overlaying the applicable areas of cumulative porosity = 25.3% ~ 38.5% and effective yield = 808.2 m^3 ~ 1452.9m^3, as shown in Figure 4-6 (e). The predicted areas where the rule may be applied or occur are located in the central northeast of the studied reservoir area within the black barrier denoted by oblique lines.



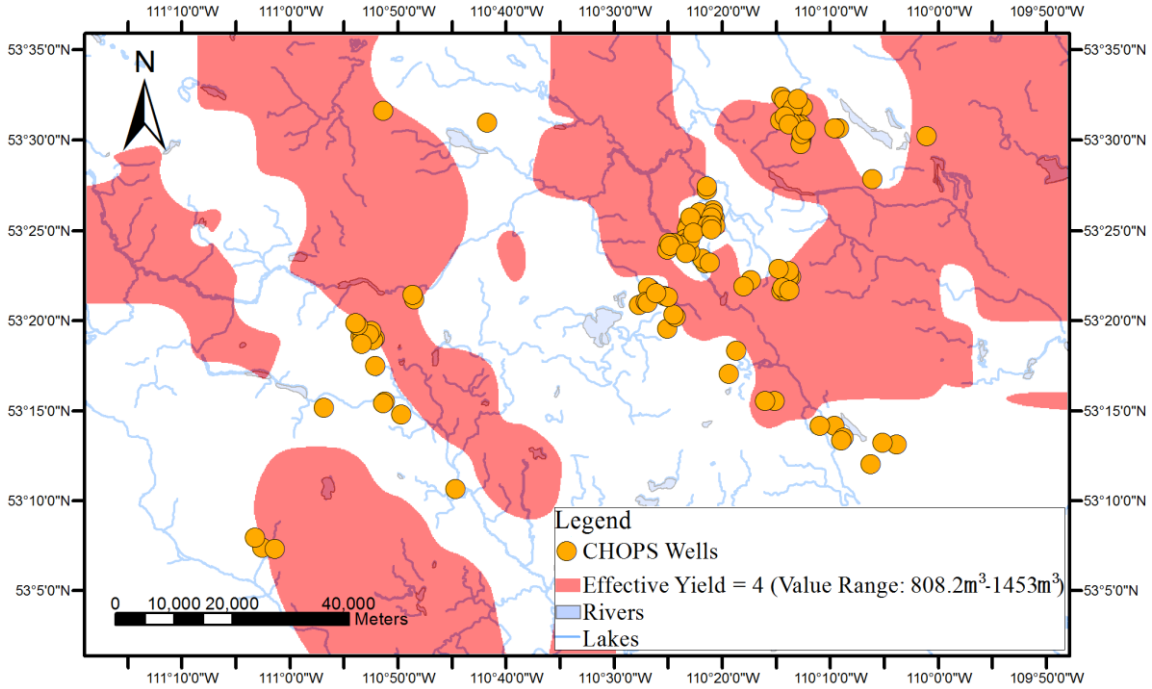
(a) Gradient map of cumulative porosity with Spline interpolation



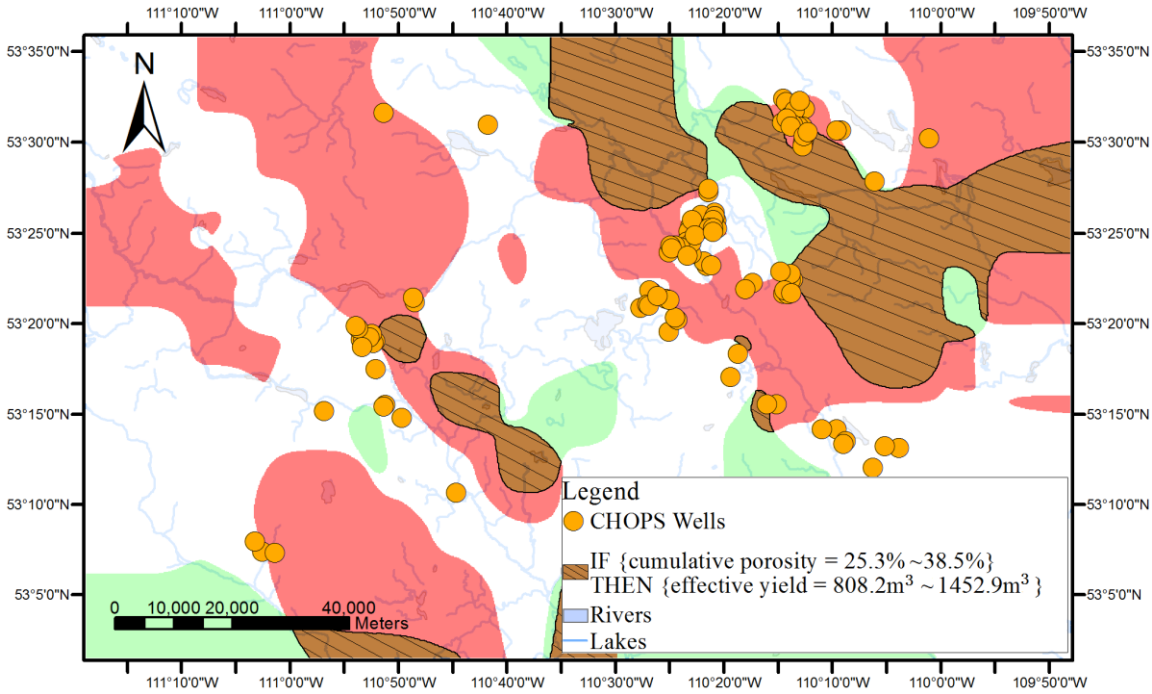
(b) Prediction map of cumulative porosity from 25.3% ~ 38.5%



(c) Gradient map of effective yield with Spline interpolation



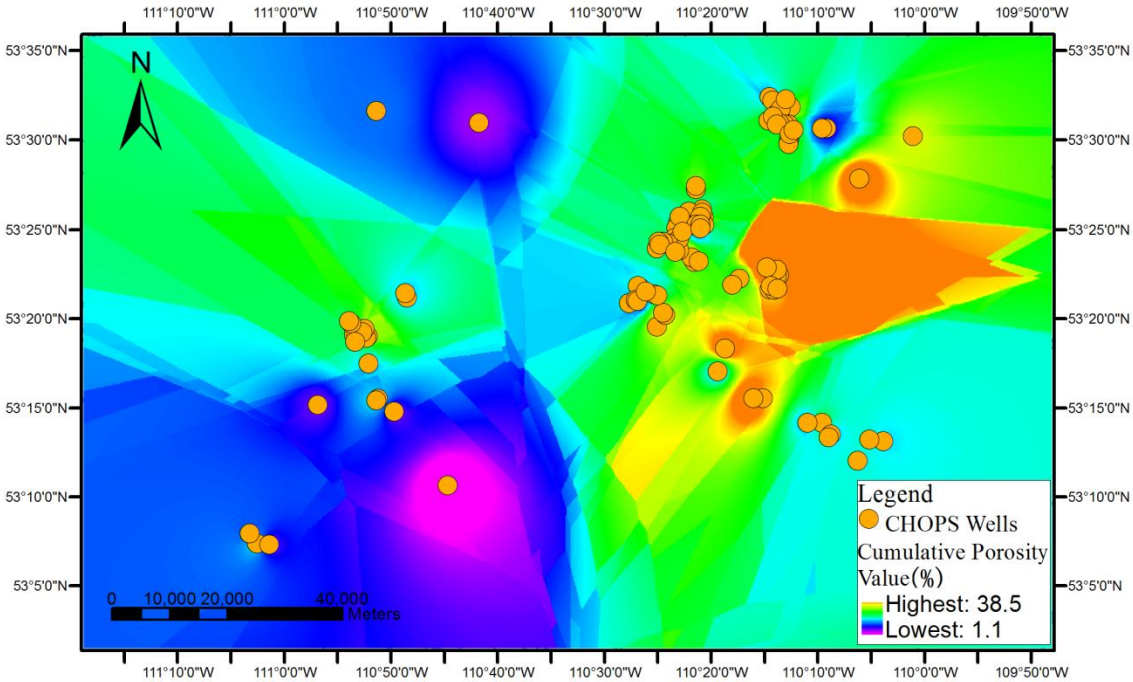
(d) Prediction map of effective yield from 808.2 m^3 ~ 1452.9 m^3



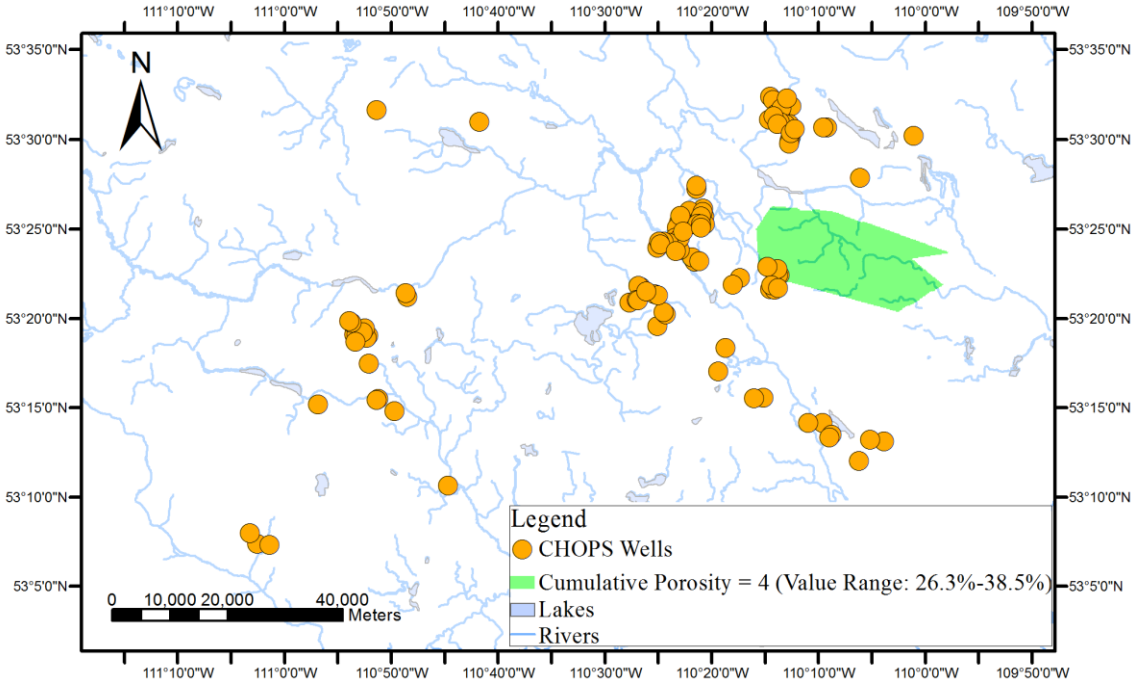
(e) Final prediction map of the rule

Figure 4-6 Surface-based geovisualization based on Spline interpolation of IF {cumulative porosity = 25.3% ~ 38.5%} THEN {effective yield = 808.2 m^3 ~ 1452.9 m^3}.

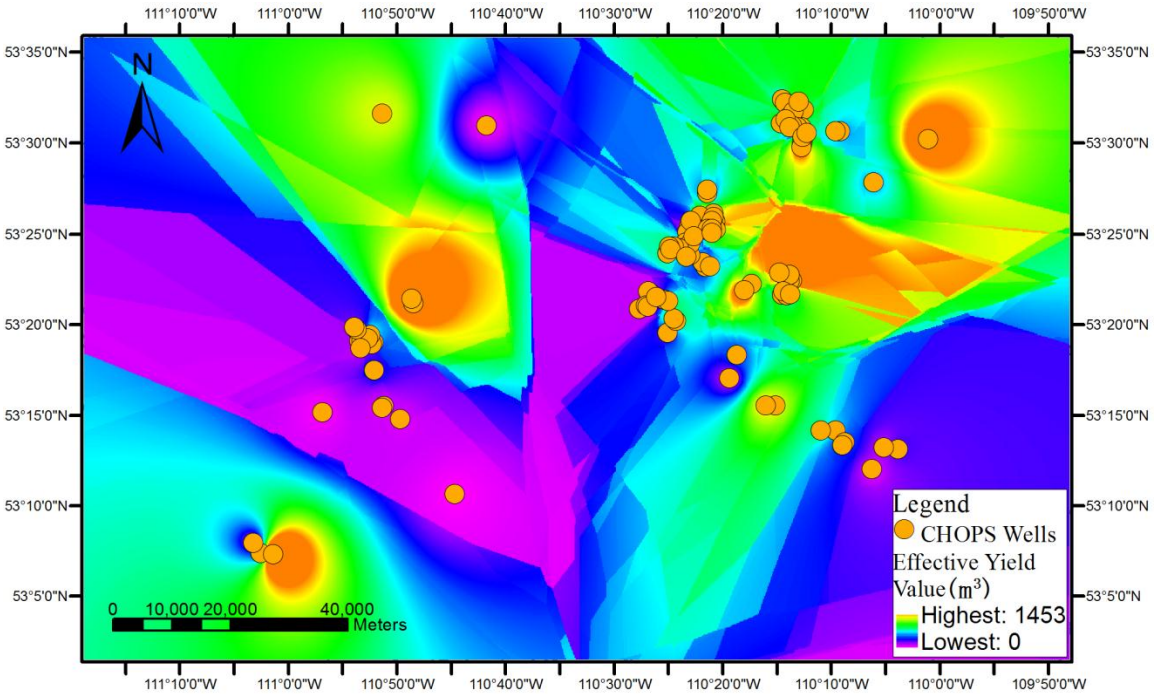
Figures 4-7 and Figures 4-8 show the process of surface-based geovisualization of the same rule IF {cumulative porosity = 25.3% ~ 38.5%} THEN {effective yield = 808.2 m^3 ~ 1452.9 m^3}, based on two other deterministic spatial interpolation methods – Inverse Distance Weighting (IDW) and Trend methods.



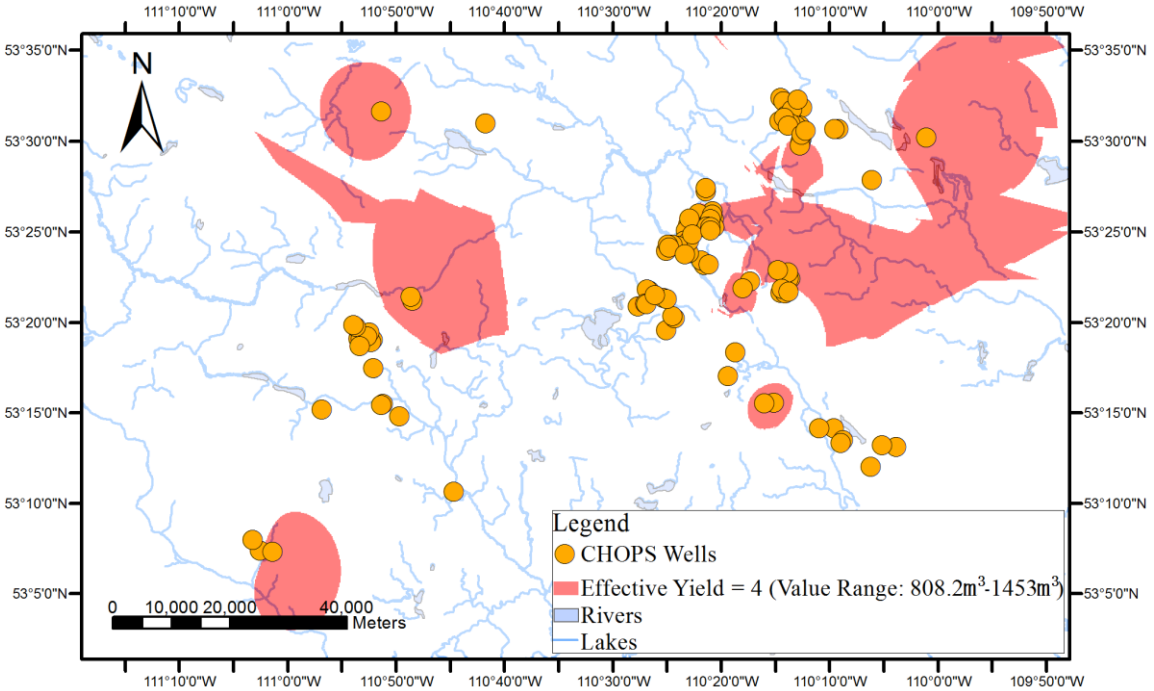
(a) Gradient map of cumulative porosity with IDW interpolation



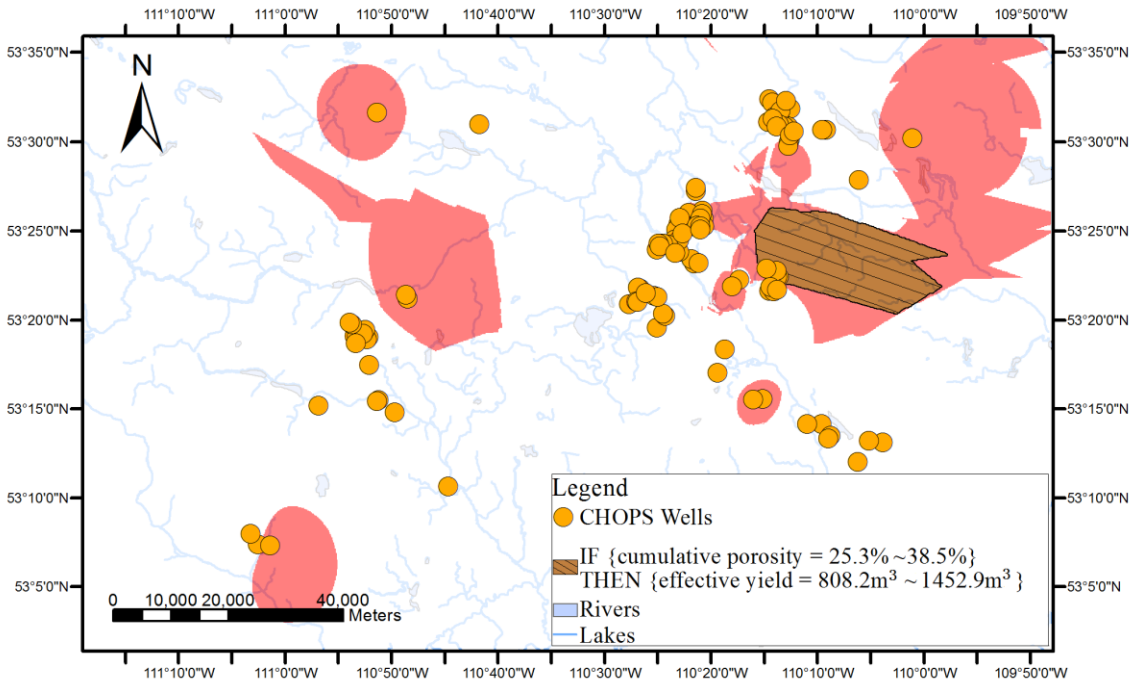
(b) Prediction map of cumulative porosity from 25.3% ~ 38.5%



(c) Gradient map of effective yield with IDW interpolation

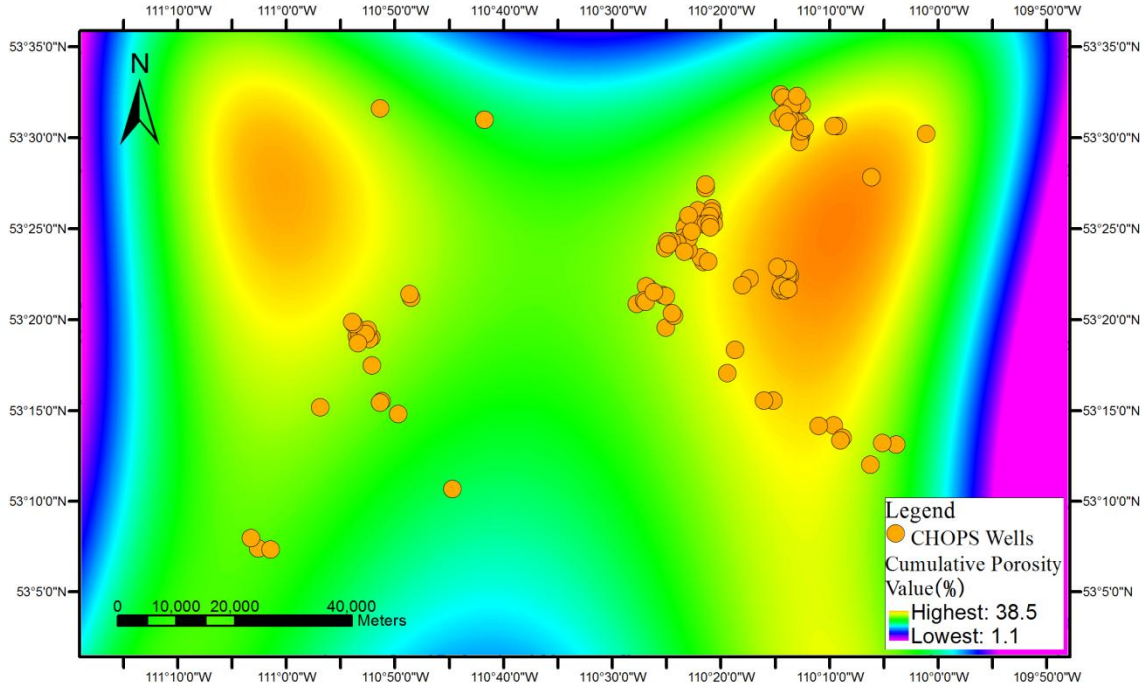


(d) Prediction map of effective yield from 808.2 m^3 ~ 1452.9 m^3

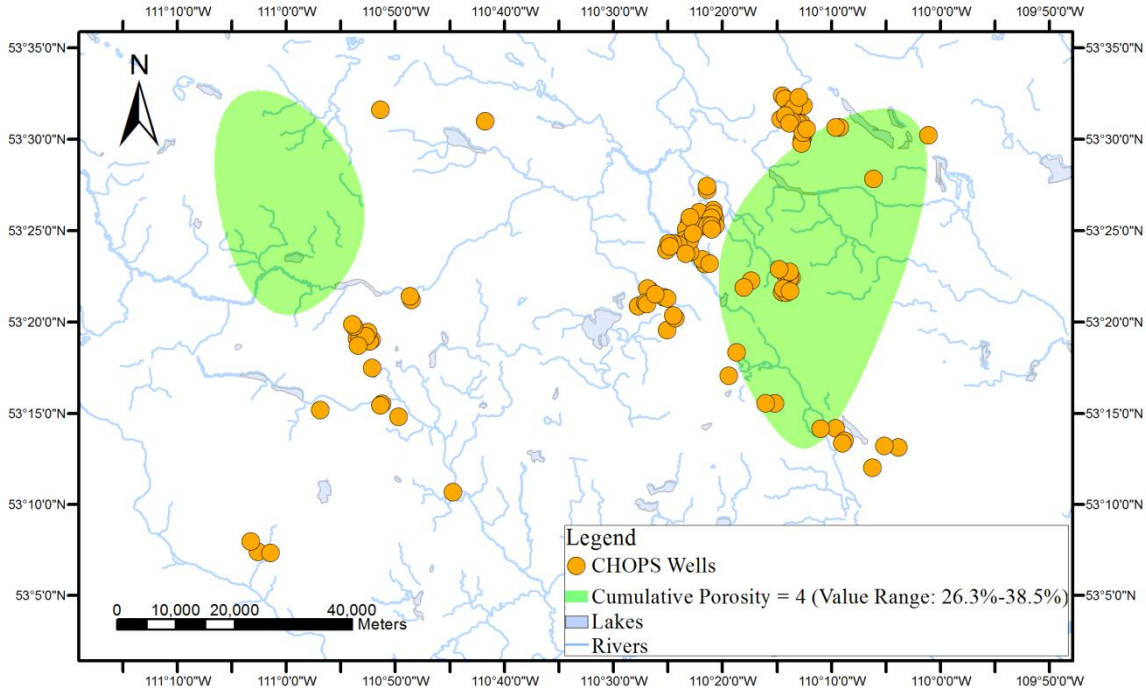


(e) Final prediction map of the rule

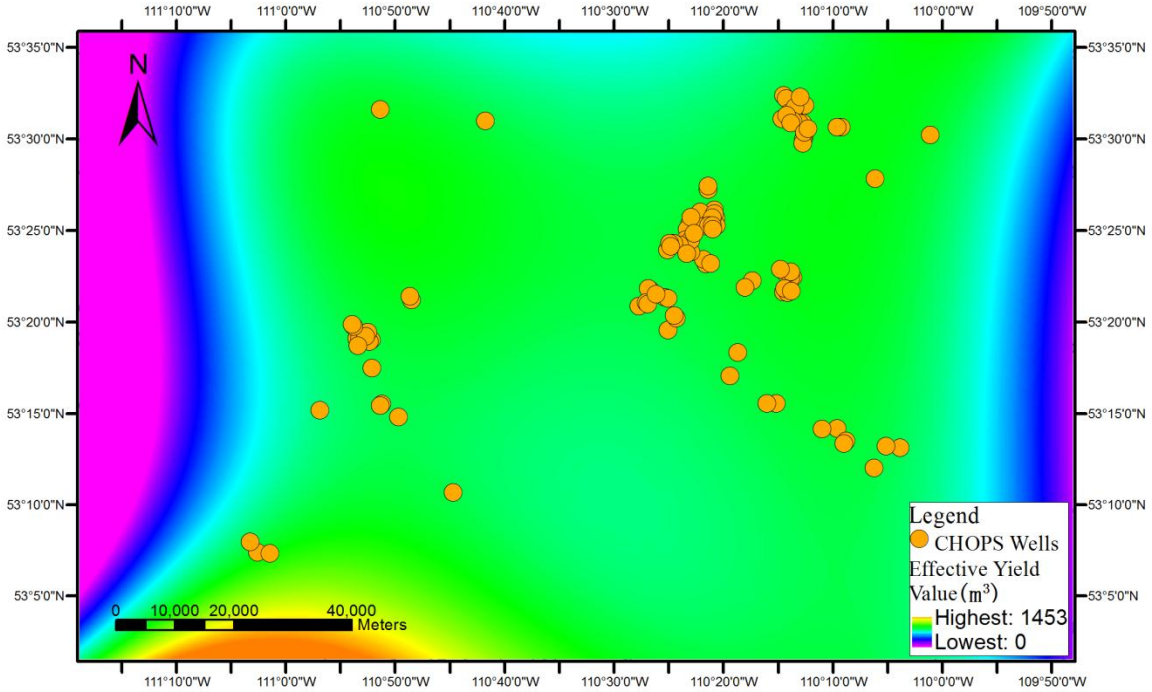
Figure 4-7 Surface-based geovisualization based on IDW interpolation of IF {cumulative porosity = 25.3% ~ 38.5%} THEN {effective yield = 808.2 m^3 ~ 1452.9 m^3}.



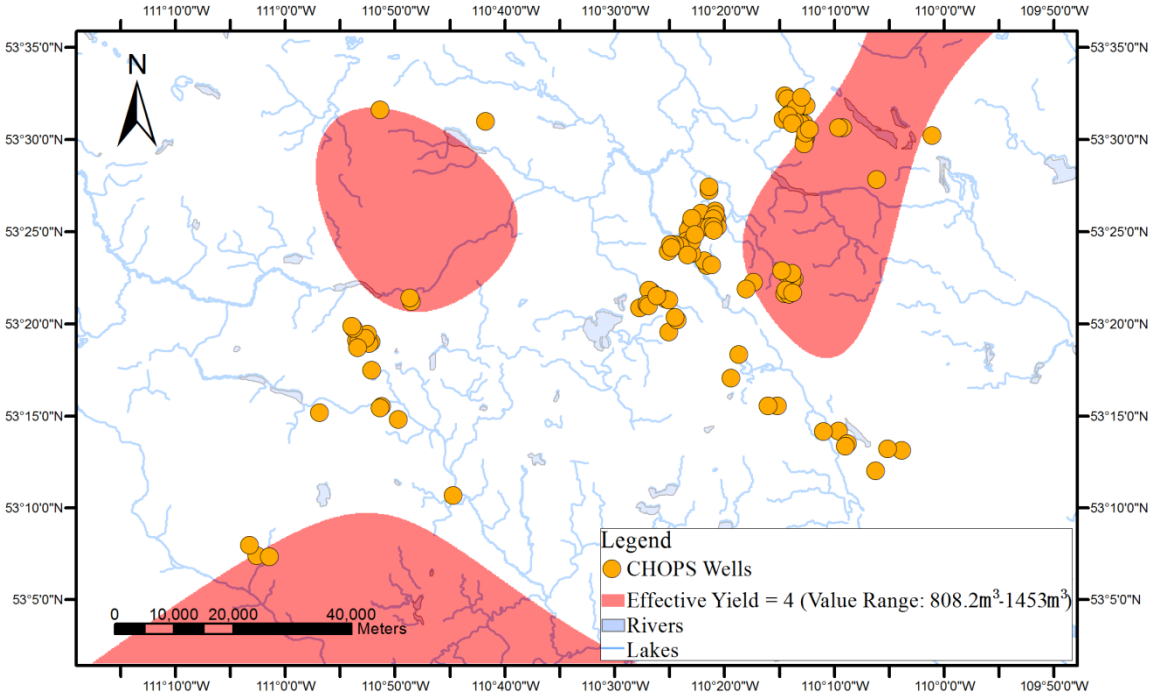
(a) Gradient map of cumulative porosity with Trend interpolation



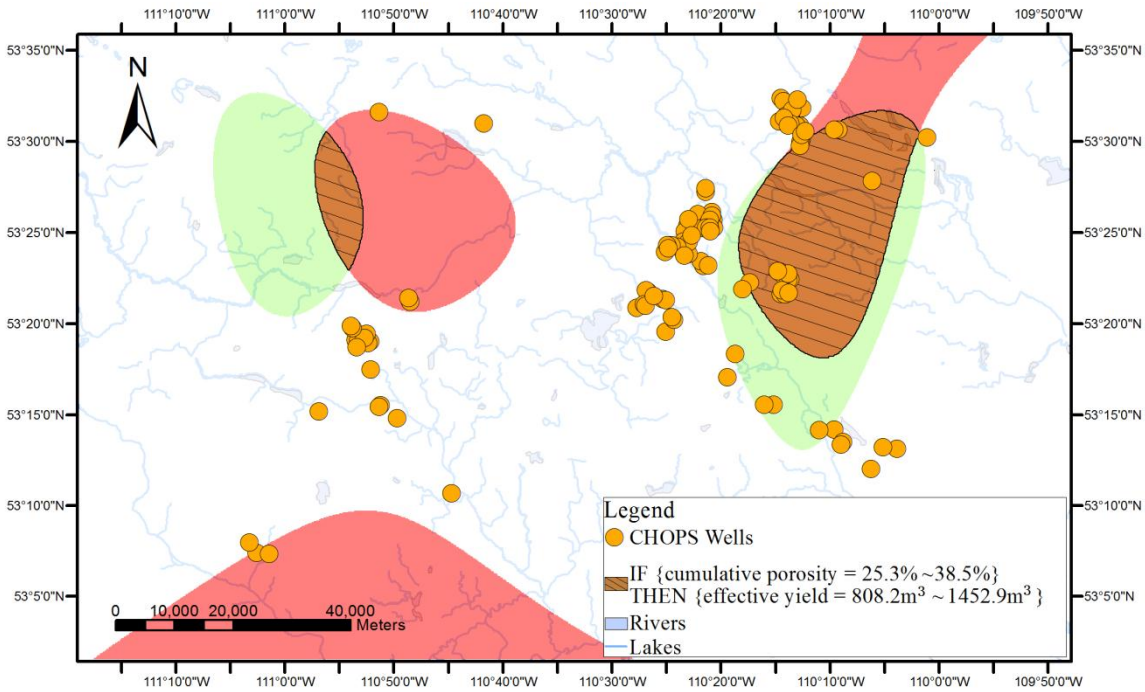
(b) Prediction map of cumulative porosity from 25.3% ~ 38.5%



(c) Gradient map of effective yield with Trend interpolation



(d) Prediction map of effective yield from 808.2 m^3 ~ 1452.9 m^3



(e) Final prediction map of the rule

Figure 4-8 Surface-based geovisualization based on Trend interpolation of IF {cumulative porosity = 25.3% ~ 38.5%} THEN {effective yield = 808.2 m^3 ~ 1452.9 m^3}.

All the maps clearly displayed the locations of the continuous applicable areas of the association rule. However, a comparison of the prediction maps generated by the three different deterministic spatial interpolation methods shown in Figures 4-6(e), Figure 4-7(e) and Figure 4-8(e) indicated that the three techniques built different applicable areas for the rule based on the CHOPS well data. Therefore, the quality of deterministic spatial interpolation results in the prediction maps was assessed in a follow-up validation procedure. In the procedure, the estimated attribute values were compared with the

residual attribute values of sampling CHOPS wells to validate the precision of interpolation.

Cross-validation is one of most common validation approaches. During the cross-validation for the interpolation results of an attribute in the CHOPS well data, one of the 118 sample wells was left out each time. An estimated value of the attribute for this well was derived using the values of the same attribute of all the other sample wells. This procedure was repeated until a value was estimated for all of the original sample wells.

Table 4-8 lists the interpolation qualities of the three used methods. According to the CHOPS well data in the case study, the cumulative porosity ranged from 1.1% to 35.0%; and, the effective yield ranged from 0 m^3 to 1452.9 m^3. IDW provided the most accurate interpolation of the two attributes based on the validation method. As shown in Table 4-8, the root mean squared error (RMSE) values of the two attributes using the IDW method were the lowest, i.e., 4.8% and 325.4 m^3, respectively. Please note that, in the Trend method, the estimated value of a location directly decided by the values of its neighbor control points was fixed and exact.

Cross-validation could only assess the overall the interpolation quality for the whole studied area. Thus, if the locations to be estimated were in data-poor areas (e.g., the central part between 110°40'W and 110°30'W of the studied areas in Figure 5), the accuracy of their estimated values was difficult to determine with the cross-validation results.

One feasible option is the use of stochastic interpolation methods, in which all the interpolated values can be evaluated by the errors with estimated values. The quality of

the interpolated points, especially the ones in data-poor areas, can be assessed by the evaluation methods provided by stochastic interpolation methods.

Table 4-8 Cross-validation of the deterministic interpolation results

Items / Interpolation	Cumulative porosity (1.1%~35.0%)		Effective yield ($0m^3$~$1452.9m^3$)	
	Estimated value range	RMSE	Estimated value range	RMSE
Spline	5.9%~39.6%	6.9%	$2.0m^3$~$1459.6m^3$	$520.4m^3$
IDW	6.8%~32.3%	4.8%	$7.2m^3$~$969.1m^3$	$325.4m^3$
Trend	4.5%~40.9%	11.6%	$5.6m^3$~$1390.5m^3$	$406.3m^3$

4.4.2.3 Surface-based Geovisualization based on Stochastic Spatial Interpolation

The same example was used to illustrate the process of surface-based geovisualization based on a stochastic spatial interpolation method, i.e., Kriging interpolation. The process of generating continuous applicable areas for the rule by stochastic spatial interpolation is similar to that using deterministic interpolation.

The gradient maps of the cumulative porosity and effective yield attributes were first generated by the Kriging method, as the spatial dependence of the two attributes was found to hold for the studied area in Section 4.4.2.1. The values of the attributes for the whole study area were also estimated with interpolation using the corresponding known attribute values at nearby CHOPS wells.

The applicable areas with cumulative porosity between 25.3% and 38.5% and effective yield between 808.2 m^3 and 1452.9 m^3 were extracted from the gradient cumulative porosity and effective yield maps and indicated using green and red colors. Through the overlaying of the applicable areas of the two attributes, the final prediction

map of the association rule, i.e., IF {cumulative porosity = 25.3% ~ 38.5%} THEN {effective yield = 808.2 m^3 ~ 1452.9 m^3}, was obtained, as shown in Figure 4-9.

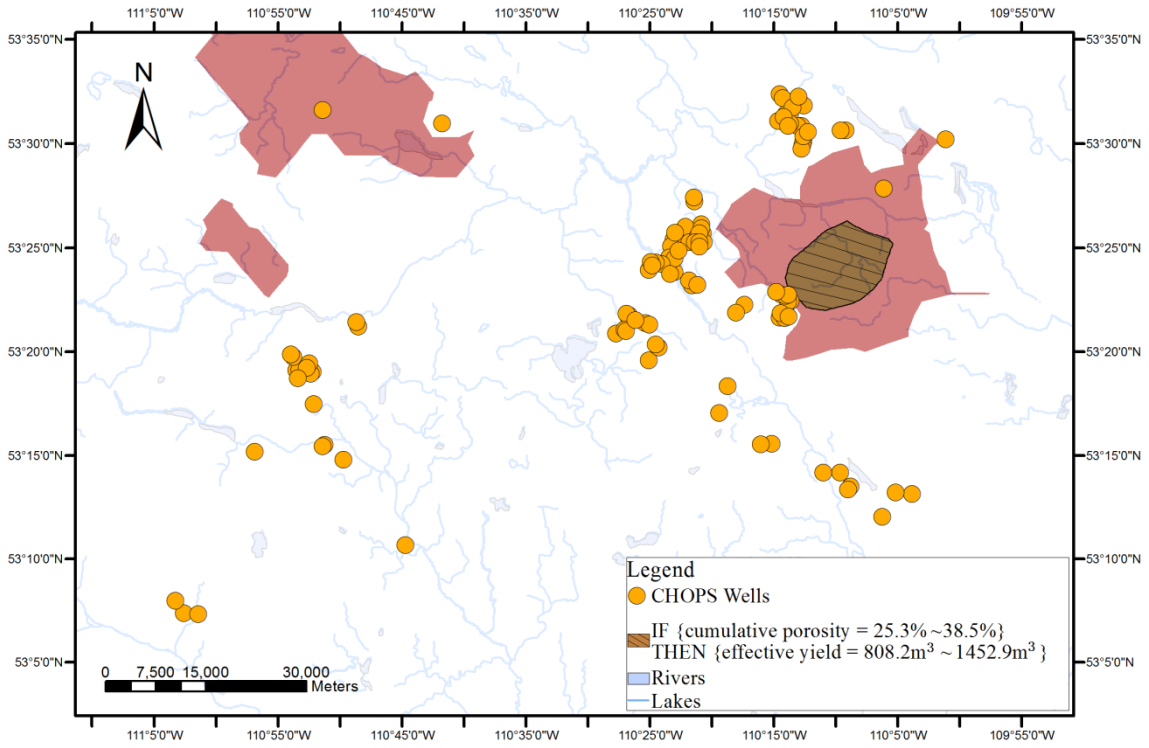
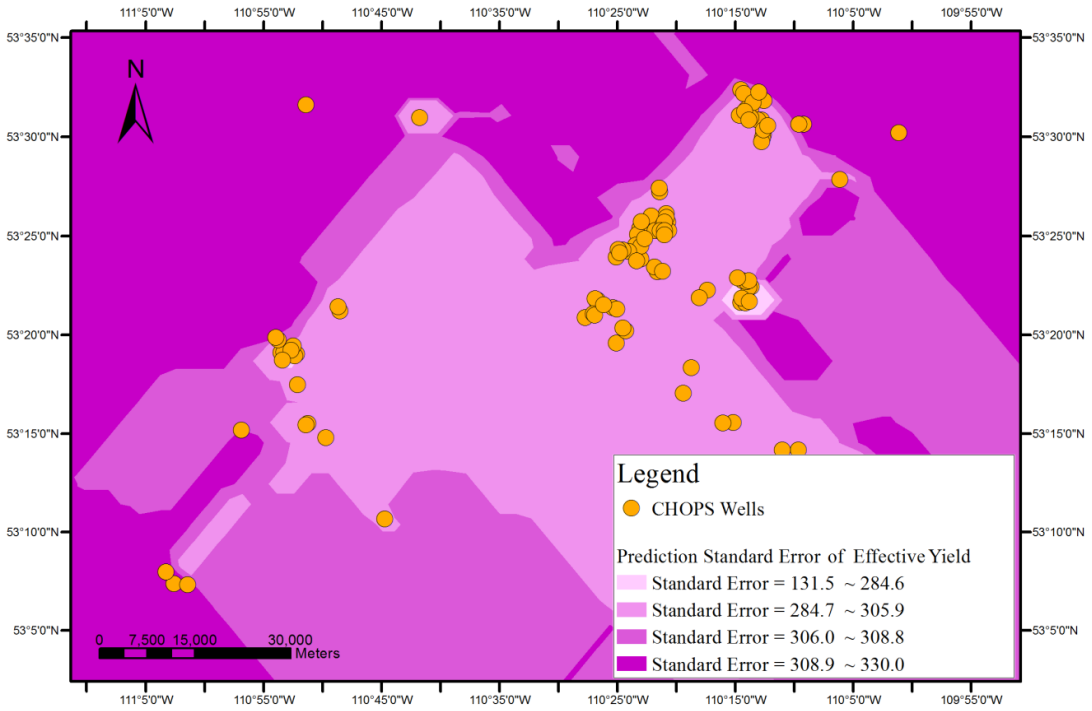


Figure 4-9 Prediction map generated by surface-based geovisualization based on Kriging interpolation

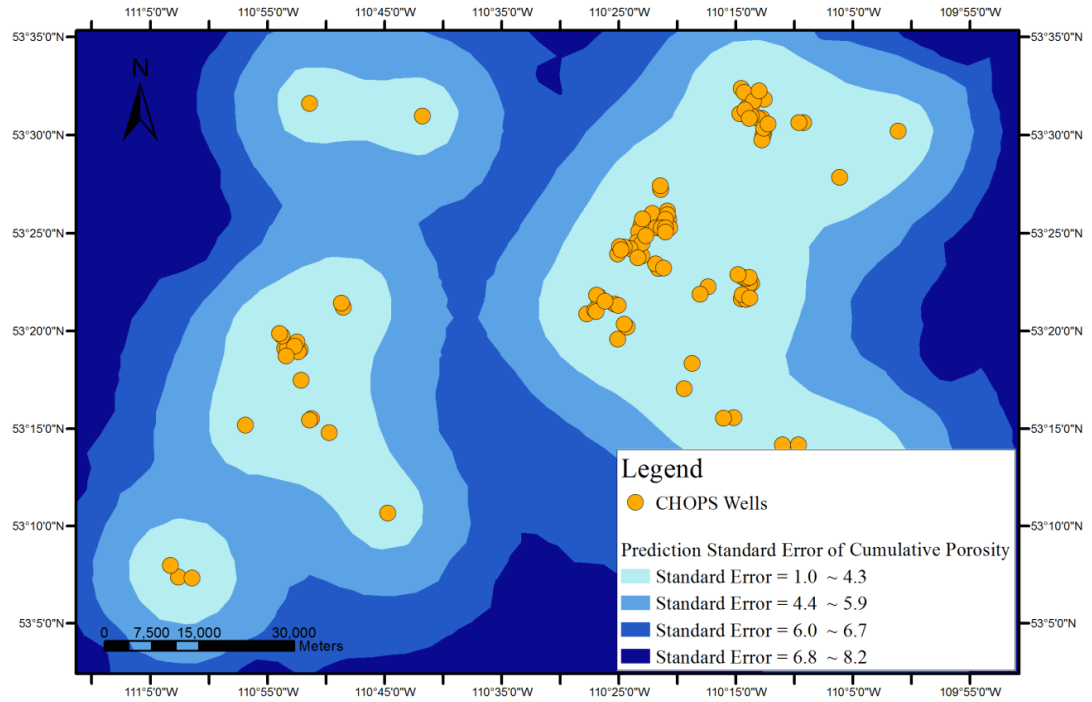
Geovisualization results based on stochastic spatial interpolation can be assessed using errors of the estimated values. The Kriging method provides evaluation methods for the estimated values in the form of prediction standard errors. Figure 4-10 shows the prediction error maps for the interpolation results of the cumulative porosity and effective yield attributes. Unlike cross-validation with deterministic spatial interpolation, the prediction error maps from stochastic spatial interpolation can be utilized to evaluate the reliability of the interpolation (or geovisualization) results of the attributes at any

location. For example, the regions with the deepest blue color in Figure 4-10(a) represent the areas where the highest prediction error range of the cumulative porosity (5.87 ~ 7.52) occurred. It is also easy to observe that the locations were mainly in the central areas of the map, due to deficiency of the sample data.

Furthermore, when the two prediction error maps in Figure 4-10 were combined with the prediction map of the association rule in Figure 4-9, it could be determined that the predicted applicable areas of the association rule (central east part of the studied area) were relatively reliable, since the two error prediction maps together suggested that prediction errors of the cumulative porosity and effective yield attributes were both relatively low in the predicted applicable areas. Such exploratory information from surface-based geovisualization can guide reservoir engineers or other users in making predictions.



(a)



(b)

Figure 4-10 Prediction error maps of (a) the antecedent item of the rule (cumulative porosity) and (b) the consequence item of the rule (effective yield)

4.5 Summary

This chapter presented a case study conducted on a real CHOPS well dataset for the Lloydminster heavy oil block in Alberta, Canada. It starts by selecting CHOPS wells from the Lloydminster heavy oil block in Alberta and collecting the data of the wells from Divestco. Then in Section 4.2 the values of the reservoir property and oil production performance parameters in the collected source data were discretized to make the process of association rule mining efficient and association rules more accessible. In Section 4.3, by antecedent, consequence and wells conditions, the proposed association rule filters successfully helped in discovering interesting rules from a large number of rules after association rule mining.

In Section 4.4, the proposed point- and surface-based geovisualization methods used on the association rules were described. To illustrate the process of the geovisualization methods, some interesting rules discovered by the rule filters in Section 4.3 were selected as examples. The preliminary findings are very encouraging and show the feasibility of the visualization methods. The visualization of an interesting rule by the point-based geovisualization method shows that the method can effectively allow the interesting rule to be connected with the wells on the map and hidden distribution patterns of the wells related to the rule to be identified. In addition, surface-based geovisualization was carried out on another interesting rule respectively based on deterministic and stochastic spatial interpolation. All the visualizations (in form of prediction maps) of the association rule clearly displayed the locations of the continuous applicable areas of the rule. The prediction maps generated by the three different

deterministic spatial interpolation methods were compared and analyzed. The quality of deterministic spatial interpolation results in the prediction maps was assessed in a validation procedure. However, the accuracy of their estimated values was difficult to determine in data poor area. Surface-based geovisualization method by stochastic interpolation solved this problem by providing errors of the estimated values. Prediction error maps assessed the quality of the interpolated points (wells), especially the ones in data-poor areas.

Chapter Five: **Conclusions and Future Work**

This chapter draws conclusions from this thesis and provides suggestions for future work.

5.1 Conclusions

The thesis mainly investigates association rule mining with filters and visualizing interesting rules in CHOPS well data.

Through the discovery of association rules between reservoir properties and oil production performance from CHOPS well data over the years, association rule mining is an effective approach to understand and enhance oil recovery. However, mining association rules in CHOPS well data always generates a large number of found rules so that reservoir engineers have to go through all the rules and manually discover the rules that they are interested in. As a solution, three association rule filters are proposed according to the characteristics of interesting rules in CHOPS well data.

The association rules between reservoir properties and oil production are of interest to reservoir engineers who wish to study how reservoir properties contribute the oil recovery of CHOPS wells. The antecedent-consequence filter screens out this type of interesting rules by limiting the antecedent as reservoir property parameters and the consequence as oil production parameters before the association rules are generated from frequent itemsets. It improves the traditional item constraints in the form of regular expressions by further limiting the internal structure (antecedent and consequence) of generated association rules.

Another factor determining the interestingness of association rules in CHOPS well data is the wells related to the rules. In practice, reservoir engineers may only want a subset of the association rules between reservoir properties and oil production, which match some wells. Therefore, on top of the antecedent-consequence filter, we provide well identifier filter and well location filter for specially extracting association rules by oil wells. After the association rules between reservoir properties and oil production are discovered, as optional filters, the well identifier filter facilitates the search of the rules by well identifiers; whereas the well location filter outputs the rules by interactively selecting oil wells in the map interface. Compared with the traditional regular expression constraints, the well identifier and well location filter take spatial objects into account and extract interesting rules by restricting external spatial objects (oil wells) relevant to association rules.

Visualization can make interesting association rules more accessible and allow hidden knowledge behind the rules to be identified. In spite of the enormous advances that have been made in visualizing the patterns included by association rules over recent years, a neglected aspect so far is the spatial objects such as oil wells related to the rules.

In response, this thesis proposes two geovisualization methods, point- and surface-based geovisualization, which aim to bridge the gap between the interesting rules and wells and to predict the applicable areas for the rules. The association rules mined in CHOPS well data are different with the rule in transaction data, because their valuable information is not only reflected by the included patterns, but also hidden behind the locations of the wells related to them. The point-based geovisualization method helps identify the geospatial distribution patterns of the wells associated with association rules

by using different symbols to show the locations of wells depending on their relationship with association rules in CHOPS well data.

Combined with spatial interpolation, the surface-based geovisualization constructs and visualizes the application areas for association rules. The spatial interpolation is only meaningful when it comes to the data with spatial dependence. Thus the premise of surface-based geovisualization is that spatial dependence exists in the attributes appearing in the target association rule. The thesis illustrates and provides preliminary exploration of deterministic or stochastic spatial interpolation methods that may be of enormous value for predicting and visualizing the application areas of association rule items. A prediction map of interesting rule obtained by overlaying the built application areas allows the uncovering of certain application areas and helps reservoir engineers outline further strategy in advanced data analysis.

The proposed methods were implemented in the form of a system prototype named CHOPSData-GeoViz to efficiently mine and visualize association rules in CHOPS well data. Using an available dataset in western Alberta, Canada, a case study was carried out and the first results have been very encouraging. The outputted association rules in CHOPS well data could be effectively customized according to the requirements of users. The hidden distribution patterns of wells and application areas related to the interesting rules were visualized on the map and they were identified clearly and easily without GIS or other expertise.

5.2 Future Work

Several extensions to this thesis are suggested and listed as follows.

1. The data will be extended to larger datasets with more reservoir properties, such as pressure and fluid property, and operational records, such as oil sand production records;

2. The proposed point- and surface-based geovisualization methods will be further improved. For point-based geovisualization, more appropriate and sophisticated symbology will be explored for representing the geospatial distribution of the wells related to the association rules. In surface-based geovisualization, the influences of spatial direction and distance on the process of spatial dependence examination of larger oil datasets will be studied. Also, the proposed point- and surface-based geovisualization methods can currently only be applied to a single rule; therefore, the methods will be extended to multiple rules;

3. The graphical user interface of the system prototype, CHOPSData-GeoViz, will be improved by using more user-friendly visualization options such as the point- and surface-based geovisualization functions for multiple rules. Also, the CHOPSData-GeoViz system of web version will be developed using ArcGIS API for JavaScript and ArcGIS Server 10.0.

REFERENCES

Agrawal R., and Srikant R. (1994). “Fast Algorithms for Mining Association Rules.” In *Proceedings of the 20th Very Large Data Bases Conference*, Santiago, Chile, pp. 487-499.

Agrawal, R., Imielinski, T., and Swami, A. N. (1993). “Mining Association Rules between Sets of Items in Large Databases.” In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Washington D.C., pp. 207-216.

Aminian K., and Ameri S. (2005). "Application of artificial neural networks for reservoir characterization with limited data." *Journal of Petroleum Science and Engineering*, 49(3-4), 212-222.

Aoidh, E. M., Martinsohn, J. T., Maes, G. E., Cariani, A., and Nielsen E. E. (2013). “Exploring Geovisualization Symbology for Landscape Genetics.” *Transactions in GIS*, 17 (2), pp. 267-281.

Aulia A., Keat T. B., Maulut M. S., El-Khatib N., and Jasamai M. (2010). “Smart Oilfield Data Mining for Reservoir Analysis.” *International Journal of Engineering and Technology*, 10(6), pp.78-88.

Bayardo, Jr., R. J. and Agrawal, R. (1999). “Mining the most interesting rules.” In *KDD ’99: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp. 145–154.

Berrado, A. and Runger, G. C. (2007). “Using metarules to organize and group discovered association rules,” *Data Mining and Knowledge Discovery*, 14, pp.409–431.

Brin, S., Motwani, R., Ullman, J. D., and Tsur, S. (1997). “Dynamic Itemset Counting and Implication Rules for Market Basket Data,” In *SIGMOD 1997, Proceedings ACM SIGMOD International Conference on Management of Data*, Tucson, Arizona, USA, pp. 255–264.

Bruzzese, D. and Davino, C. (2008). “Visual Mining of Association Rules,” In *Visual Data Mining: Theory, Techniques and Tools for Visual Analytics*, Springer-Verlag, pp. 103–122.

Buono, P. and Costabile, M. F. (2005). “Visualizing Association Rules in a Framework for Visual Data Mining.” In *From Integrated Publication and Information Systems to Virtual Information and Knowledge Environments*, pp. 221-231.

Cai, Y., Hu, K., and Wang, X. (2011). “Finding Relationships between Reservoir Characteristics and Oil Production for the Cold Production.” *The 32nd Annual Symposium & Workshop of IEA Collaborative Project on Enhanced Oil Recovery*, Vienna, Austria.

Cai, Y., Wang, X., Hu, K., and Dong, M. (2014). “A Data Mining Approach to Finding Relationships between Reservoir Properties and Oil Production for CHOPS.” *Computers and Geoscience*. Vol. 73, December 2014, pp. 37-47.

Chakrabarti S., Ester M., Fayyad U., Gehrke J., Han J., Morishita S., Piatetsky-Shapiro G. and Wang W. (2006) “Data mining curriculum: A proposal.” *ACM SIGKDD.*

Chugh S., Baker R., Telesford A., Zhang E. (2000). “Mainstream options for heavy oil: Part I-Cold Production.” *Journal of Canadian Petroleum Technology*, 39(4), pp.31-39.

Creighton C., and Hanash S. (2003). “Mining gene expression databases for association rules.” *Bioinformatics*, 19(1), pp.79-86.

Divestco data. ⟨http://www.divestco.com/getdoc/14af7505-8813-406e-b8b7-757f9a751ff5/Add-On-Datasets⟩.aspx (accessed 1.06.15.).

Dumanoir J. L., Hall J. D., Jones J. M. (1972). “Rxo/rt methods for wellsite interpretation.” *The Log Analyst*, XIII (5), 20-27.

El Ouahed A.K., Tiab D., and Mazouzi A. (2005). "Application of artificial intelligence to characterize naturally fractured zones in Hassi Messaoud Oil Field, Algeria." *Journal of Petroleum Science and Engineering*, 49(3-4), 122-141.

Ertek, G. and Demiriz, A. (2006). “A Framework for Visualizing Association Mining Results.” In: *Proceedings of ISCIS*, pp. 593-602.

Frawley W., Piatetsky-Shapiro G., and Matheus C. (1992). “Knowledge discovery in database: An overview.” AI Magazine, 13(3), 57-70.

Gienko, G. A., and Terry, J. P. (2012). “Geovisualization of Tropical Cyclone Behavior in the South Pacific.” Geological Society Special Publications. 361 (1), pp. 195-208.

Guo, D. (2009). "Multivariate Spatial Clustering and Geovisualization", In Geographic Data Mining and Knowledge Discovery, edited by H. J. Miller and J. Han. London and New York, Taylor & Francis, pp. 325-345.

Gupta, G., Strehl, A., and Ghosh, J. (1999), “Distance Based Clustering of Association Rules,” in *Intelligent Engineering Systems Through Artificial Neural Networks*, ASME Press, pp. 759–764.

Hahsler, M., and Chelluboina, S., 2011. “Visualizing Association Rules in Hierarchical Groups”. In Proceedings of *the 42nd Symposium on the Interface: Statistical, Machine Learning and Visualization Algorithms* (Interface 2011).

Hahsler, M., Grun, B., and Hornik, K. (2005), “arules – A Computational Environment for Mining Association Rules and Frequent Item Sets,” *Journal of Statistical Software*, Vol.14, pp.1–25.

Hahsler, M., Chelluboina, S., (2011), “Visualizing Association Rules: Introduction to the R-extension Package arulesViz,” *R project module*, 2011, pp.223-238.

Han J., and Kamber M. (2006). "Data mining concepts and techniques." *Morgan Kaufmann*.

Hartigan, J. A. and Wong, M. A. (1979), "A K-means clustering algorithm," *Applied Statistics*, Vol.28, pp.100–108.

Hofmann H., Seibes A., Wilhelm AFX. (2000). "Visualization Association Rules with Interactive Mosaic Plots." In *KDD* (Knowledge Discovery in Databases) 2000, pp. 227-235.

Istchenko C. M., Gates I. D. (2012). "The well-wormhole model of CHOPS: history match and validation." In: *SPE Heavy Oil Conference*, Calgary, Alberta, Canada, pp.1-9.

Jiao J., and Zhang Y. (2005). "Product portfolio identification based on association rule mining." *Computer-Aided Design*, Vol.37, pp.149-172.

Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H., and Verkamo, A. I. (1994). "Finding Interesting Rules from Large Sets of Discovered Association Rules." In: *Proceedings of CIKM*, pp. 401-407.

Lam, N. S. (1983). "Spatial Interpolation Methods: A Review." *The American Cartographer*, Vol. 10, pp. 129-149.

Lin Y., Cunningham III G. A., Coggeshall S. V. (1996). "Input variable identification - fuzzy curves and fuzzy surfaces." *Fuzzy Sets and Systems*. Vol.82, pp. 65-71.

Liu X., Zhao G. (2005). "A fractal wormhole model for cold heavy oil production." *Journal of Canadian Petroleum Technology*, 44(9), pp. 31-36.

MacEachren, A.M., and Kraak, M. J. (1997). “Exploratory Cartographic Visualization: Advancing the Agend,” *Computers and Geoscience*. Vol. 23, pp. 335-343.

MacEachen, A.M., Wachowicz, M., Edsall, R., Haug, D. (1999). “Constructing knowledge from multivariate spatiotemporal data: Integrating geographic visualization with knowledge discovery in database methods.” *International Journal of Geographic Information Science*, Vol.13, pp. 311-334.

Mai A., Bryan J., Goodarzi N., Kantzas A. (2009). “Insights into non-thermal recovery of heavy oil.” *Journal of Canadian Petroleum Technology*, 48(3), pp.27-35.

Marco V., and Valentina C. (2004). “Meaningful Discretization of Continuous Features for Association Rules Mining by means of a SOM.” In *Proceedings of: European Symposium on Artificial Neural Networks* (ESANN '04), Bruges. pp. 489-494.

Marroquin I.D., Brault J.J., and Hart B.S. (2009). "A visual data mining methodology for seismic facies analysis: Part2-Application to 3D seismic data." *Geophysics*, 74(1), 13-23.

Mohaghegh, S., Arefi, R., Ameri, S., Aminiand, K., and Nutter, R. (1996). "Petroleum reservoir characterization with the aid of artificial neural networks." *Journal of Petroleum Science and Engineering*, 16(4), 263-274.

Rainsford, C. P. and Roddick, J. F. (2000). “Visualization of Temporal Interval Association Rules.” In IDEAL '00: *Proceedings of the 2nd International Conference on Intelligent Data Engineering and Automated Learning, Data*

Mining, Financial Engineering, and Intelligent Agents, Springer-Verlag, pp. 91-96.

Ramakrishnan S., Quoc V., Rakesh A. (1997). “Mining Association Rules with Item Constraints”, In *proceedings of the KDD conference*, pp. 67-73.

Rivero J. A., Coskuner G., Asghari K., David H., Pearce A., Newman R., Birchwood R., Zhao J., Ingham J. (2010). “Modeling CHOPS using a coupled flow-geomechanics simulator with none quilibrium foamy-oil reactions: a multiwell history matching study.” In: *SPE Annual Technical Conference*, Florence, Italy, pp. 1-21.

Sawatzky R. P., Lillico D. A, London M. J., Tremblay B. R., and Coates R. M. (2002). “Tracking Cold Production Footprints.” In: *Proceedings of the Canadian International Petroleum Conference*, Calgary, Alberta, Canada, pp.1-16.

Smith G. E. (1988). Fluid flow and sand production in heavy-oil reservoirs under solution-gas drive. *SPE Production Engineering*, 3(2), 169-180.

Strecker U., and Uden R. (2010). "Data mining of 3D poststack seismic attribute volumes using Kohonen self-organizing maps." *The Leading Edge*, 21, 1032-1037.

Sohn S. Y., Kim Y. (2008). “Searching customer patterns of mobile service using clustering and quantitative association rule.” *Expert System with Applications*, 34(2), pp.1070-1077.

Tremblay B. (2009). “Cold flow: a multi-well cold production (CHOPS) model.” *Journal of Canadian Petroleum Technology*, 48(2), pp. 22-28.

Unwin, A., Hofmann, H., and Bernt, K. (2001). “The Two-Key Plot for Multiple Association Rules Control.” In PKDD '01: *Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery*, Springer-Verlag, pp. 472-483.

Vitaly Friedman (2008) "Data Visualization and Infographics" In: *Graphics, Monday Inspiration*, January 14th, 2008.

Wang, B.J., Wang, X., and Chen, Z.X. (2013). “A Hybrid Framework for Reservoir Characterization Using Fuzzy Ranking and an Artificial Neural Network.” *Computers and Geosciences*, Vol. 57, pp. 1-10.

Wu X., Kumar V., Quinlan J.R., Ghosh J., Yang Q., Motoda H., Mclachlan G.J., Ng A., Liu B., Yu P.S., Zhou Z.H., Steinbach M., Hand D.J., and Steinberg D. (2007). “Top ten algorithms in data mining.” *Knowledge and Information System*, 14(1), pp.1-37.

Yang, L. (2003). “Visualizing Frequent Itemsets, Association Rules, and Sequential Patterns in Parallel Coordinates.” In: *Proceedings of Computational Science and Its Applications* (ICCSA 2003). Lecture Notes in Computer Science, pp. 21-30.