Evaluating the Feasibility of using Artificial Neural Networks to Predict Hydrogeologic Units in Complex Glacial Deposits

by

Zidra Hammond

B.A.Sc. (Environmental Engineering), University of Waterloo, 2003

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

in the
Department of Earth Sciences
Faculty of Science

SIMON FRASER UNIVERSITY
Summer 2022

Declaration of Committee

Name:	**Zidra Hammond**
Degree:	**Master of Science (Earth Sciences)**
Title:	**Evaluating the Feasibility of using Artificial Neural Networks to Predict Hydrogeologic Units in Complex Glacial Deposits**
Committee:	**Chair: Shahin Dashtgard** Professor, Earth Sciences
	Diana Allen Supervisor Professor, Earth Sciences
	Brent Ward Committee Member Professor, Earth Sciences
	Michael Simpson Committee Member Senior Hydrogeologist City of Vancouver
	Jillian Sacré Examiner Principal/Senior Hydrogeologist WSP-Golder

Abstract

The feasibility of using Multilayer Perceptron (MLP) to predict hydrogeologic units (HGUs) in complex glacial deposits is evaluated. The study area includes the Fraser-Whatcom Basin. Material descriptions from boreholes logs are standardized into HGUs using natural language processing techniques to reduce subjectively and improve automation. Three data selection alternatives are considered to evaluate the training and prediction capabilities of MLP. Block-model representations of the subsurface are created and the best geologic realization is verified against predictions using the K-nearest neighbours algorithm and geologic cross-sections from independent studies. Validation results show MLP predictions are typically more generalized but produce similar subsurface trends and can recreate confining units contributing to local artesian conditions. MLP appears to be a promising algorithm to solve multi-class classification for geologic modelling purposes. The workflow developed has the added benefit of being stochastic with the potential to generate multiple geologic realizations to account for uncertainty in geologic structure.

Keywords: lithology material standardization; glacial deposits; Lower Mainland; block model, artificial neural networks; multi-class classification, hydrogeologic units

Dedication

For Mãe 

Acknowledgements

Many thanks to my supervisor, Diana Allen, for her support and guidance throughout this thesis. Thank-you to my committee members, Mike Simpson and Brent Ward, as well as Dirk Kirste and Jillian Sacré, for taking the time to review and provide comments on my thesis. I am thankful for the friendship and enthusiasm of the students I met at the SFU Groundwater Resources Research Group as well as the support from Lorena Munoz and the library staff. I am grateful to SFU Data Hub, WestGrid, and the open source community for making machine learning and coding so accessible. A special thanks to my loving family and friends for their support and encouragement.

Table of Contents

List of Tables

List of Figures

Chapter 1.

Introduction

1.1. Background

Canada is covered by complex and heterogeneous deposits influenced by glaciation. These deposits contain major aquifers that are important sources of water that are increasingly being relied upon as water scarcity increases and surface water availability decreases due to human activity and climate change (Vaccaro, 1992). Advancements in groundwater modelling software have resulted in more effective use of regional groundwater flow models as tools to support sustainable groundwater management (Pasanen and Okkonen, 2017); however, conceptualization of geologic architecture still presents the greatest uncertainty, particularly for glacial deposits (Anderson et al., 2015; Refsgaard et al., 2012). The continuity of geological units influences aquifer extents, where hydraulic interactions occur along the flow path (e.g. recharge and discharge areas), groundwater chemistry, and the development of hydraulic conditions (e.g. confined aquifers, artesian wells) (Bayless et al., 2017).

Geologic models are the backbone of every groundwater flow model and are critical to understanding groundwater flow within complex glacial aquifer systems (Pasanen and Okkonen, 2017). Traditional methods used to develop 3D geologic models for groundwater applications rely on manual interpretations that incorporate known geologic relationships (e.g. geologic cross-sections) and expert knowledge into a deterministic model; however, such models are typically time-consuming to create, difficult to update and do not adequately represent subsurface heterogeneity or account for uncertainty in geologic structure (Jørgensen et al., 2015; Kearsey et al., 2015). As a result, the end-product may have limited use or not be directly relevant to end-users like planners, engineers, and groundwater modellers (Kearsey et al., 2015). Geostatistical approaches have been applied using numerical data (e.g. geophysical surveys) to make spatial predictions in the subsurface but typically have limitations associated with the underlying assumption of stationarity (i.e., statistics used to describe data distribution that does not change throughout the spatial domain) and the inability to reproduce subsurface complexity (Bianchi et al., 2015).

An alternative to the methods above is to use a data-driven approach that is non-parametric (i.e., no assumptions on data distribution). Machine learning algorithms can use data to learn a function that maps inputs to outputs without any prior assumption on the data distribution. These algorithms can be used to solve nonlinear problems or when the modelled phenomena are very complex (Kanevski et al., 2009). They can be used to solve classification problems using categorical data (e.g. lithologic units) which is typically the most abundant type of subsurface data for hydrogeologic applications.

The main goal of this research is to leverage textual data from borehole lithology logs and apply machine learning to build a regional geologic model in an area with complex glacial deposits using publicly available information and open-source tools. This work will address the major source of uncertainty in groundwater modelling (e.g. geologic structure) and provide the foundation required to build future numerical models to better represent groundwater flow in complex glacial deposits.

1.2. Literature Review

1.2.1. Geologic Modelling

The capability to effectively and intuitively analyse geologic information is an important part of the groundwater modelling workflow, particularly in areas with glacial and post-glacial deposits (Russell et al,. 2013, Jørgensen et al., 2015; Kearsey et al., 2015; Pasanen and Okkonen, 2017). This is important because geologic structure is a major source of uncertainty in groundwater modelling. Bianchi et al. (2015) used an interpretive groundwater model (e.g. not calibrated) to investigate how uncertainty in lithofacies modelling impacts groundwater flux and hydraulic heads. Bianchi concluded that uncertainty in lithofacies modelling has a greater impact on groundwater flux predictions compared to simulated hydraulic heads. This highlights how the accuracy of geological information is important for reducing uncertainty in groundwater models that simulate flux-dependent processes, such as groundwater recharge to discharge pathways.

There are several examples in literature that provide recommendations for geologic model development specific to glacial environments. Kearsey et al. (2015) concluded that geologic models based on stratigraphy simplify the highly heterolithic

nature of glacial deposits and do not account for small scale lithological variability. They found that lithologies in boreholes were only predicted about 50% of the time and minor lithologies were found to represent a significant portion of the total volume of stratigraphic units delineated based on dominant lithology. Therefore, a lithofacies-based approach was advocated instead of stratigraphic layers for geologic modelling, especially in complex, heterogeneous deposits where it can be difficult to locate stratigraphic boundaries. The use of a more regular dataset and a finer grid was also recommended to improve prediction performance. Similarly, the 3D geologic model constructed by Jørgensen et al. (2015) used a combination of manual and stochastic methods to represent subsurface heterogeneity. Integration of the different methods into one 3D geologic model was identified as a major challenge.

While there are challenges and limitations associated with all sources of geologic data, the integration of spatially continuous or profile sources of information is important to understanding complex geologic structure (Frind et al,. 2014). Material descriptions of lithology (e.g. borehole log) provide the necessary vertical coverage but are likely insufficient for lateral characterization of important subsurface features such as buried bedrock valleys and buried stream channels (Cummings et al., 2012; Jørgensen et al,. 2015). Therefore, borehole logs are increasingly being supplemented with outcrop information and soft data (e.g. geophysical data) to construct comprehensive 3D geological models for groundwater applications (Koch et al., 2014; Pirot, 2017; Meyer et al., 2018; Morgan, 2018).

Borehole logs from public well records contain large amounts of textual data that can be used to classify similar materials and provide qualitative information about the subsurface. Examples in literature commonly complete this classification task using semi-automated methods that apply conditional rules based on expert knowledge (Russell et al., 1998; Allen et al., 2008; and Bayless et al., 2017). More recently, Natural Language Processing (NLP) techniques have been used in geoscience applications to explore geologic lexicon so that material classification is more automated and less subjective (Padarian and Fuentes, 2019; Fuentes et al., 2020). NLP is used in various fields of science that deal with textual descriptions of reality to provide a statistical approach to understanding language. Word counts and the occurrence of word combinations are examples of statistical outputs. These NLP techniques can be incorporated to provide justification where needed to reduce subjectivity. Statistical

results can also be used to understand uncertainty associated with the classification of material descriptions. Fuentes et al. (2020) provide one of the first geoscience applications where a numerical representation of textual data (word embedding) from borehole logs is used to spatially predict lithology using machine learning and 2.5D interpolation approach. However, the use of textual data to predict complex subsurface conditions directly in 3D was not found in literature.

1.2.2. Machine Learning

Although the popularity of machine learning is typically associated with artificial intelligence, it is commonly applied for hydrogeologic purposes. Nearest Neighbors and kriging are examples of algorithms that could be considered within the machine learning toolbox. While most machine learning applications in geoscience have focused on numerical data from geophysical methods (Cracknell and Reading, 2013; Baykan and Yilmaz, 2010), machine learning can also be used to solve classification problems using categorical data. **Artificial neural network (ANN)** is a widely used machine learning algorithm inspired by biological neural networks invented in the 1950s that can be used to solve classification problems. ANNs are nonparametric models trained to identify hidden patterns and structures without any assumptions on the data distribution. ANNs are suitable for geoscience applications because they can model complex nonlinear dependencies, are adaptive for managing nonstationary data, and allow the integration of contextual information (Kanevski et al., 2009).

There are limited studies that explore the use of ANNs to solve multi-class classification problems related to hydrogeology applications. **Multi-class classification** is a subset of classification problems where more than two classes (e.g. labels or categories) exist but only one can be predicted for each instance. Rizzo and Doughtery (1994) use ANN to characterize the 2D distribution of hydraulic conductivity based on three classes (low, medium, and high). ANN also proved to be a useful methodology for a basin-wide study characterizing complex and heterogenous lithology with a multi-layered aquifer system at the borehole level (Sahoo and Jha, 2017). The ANNs in the aforementioned studies included **self-organizing map (SOM)** and **multilayer perceptron (MLP)**.

Self-Organizing Maps

SOMs are a single-layer feedforward neural network used to produce a low dimensional representation of a dataset with a high dimensional space or multiple features (e.g. multiple inputs to describe each sample). SOMs have been combined with MLP to enhance pattern recognition in hydrogeologic applications (Rizzo and Doughtery, 1994; Sahoo and Jha, 2017). The SOM architecture is a fixed 2D grid (e.g. rectangular or hexagonal) of neurons referred to as a Kohonen map (Figure 1.1).

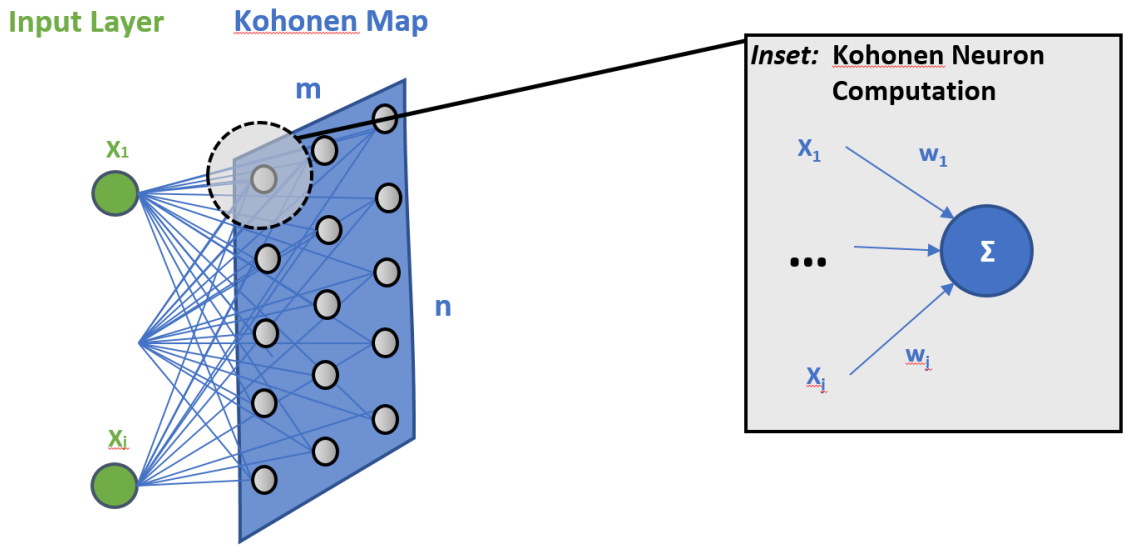


Figure 1.1 **Self-Organizing Map (SOM) conceptual architecture showing the input layer and a rectangular 2D grid (Kohonen Map). Every neuron of the input layer is connected to the neurons of the 2D grid by a weight.**

Every neuron on the map is connected to each node of the input layer with an initial weight randomly assigned. The weight is represented as a vector with the same dimensionality as the feature space of the input layer. SOM uses **unsupervised** learning, meaning that no prior knowledge is used to train the network. Training occurs in two stages using competitive and cooperative learning processes. For each record, the algorithm first tries to find the neuron on the map that is closest based on distance (e.g. Euclidian) using the input features and the initial weights (Equation (1)). This is the competitive step of the algorithm where the winning neuron becomes the 'best-matching-unit' (BMU). The weights of the BMU are updated to be closer to the record. Next, the cooperative process of the algorithm takes place where the weights of the neurons near the BMU (its neighborhood) are also updated using a neighborhood function (e.g.

Gaussian) to be closer to the record as a smoothing step. After all the records have been used to update the weights, the learning rate and neighborhood size are decreased.

$$Euclidian\ Distance = \sqrt{\sum_{i=0}^{j} (Xi - Wi)^2} \qquad (1)$$

The above process for the SOM is repeated based on the specified number of iterations. At first, the neighborhood function is broad and self-organization takes place at the global scale with large adjustments in weights. At the end, the neighborhood function shrinks to include fewer neurons and smaller adjustments in weights. SOM is similar to k-means clustering but instead of centroids moving to fit the data, the data moves to centroids that are constrained at neuron locations on the map (SuperDataScience, 2018). The resultant mapping provides input data that is assigned to a node and the coordinates of the node are used instead of the original features to reduce dimensionality and cluster data.

Multilayer perceptron (MLP)

MLP is a traditional ANN commonly referenced in the reviewed literature and can be used for multi-class classification tasks. The architecture of a MLP consists of an input layer, an output layer, and one or more hidden layers (Figure 1-1). Data flows in a forward direction from input to output (e.g. feedforward neural network). Every node in one layer is connected to every node in the next layer, resulting in a fully connected network. The input layer is the first layer and the number of nodes equals the input features (X_1 to X_j). For example, the input layer would have three nodes if the input features are based on geographical coordinates (e.g. northing, easting, elevation). The last layer is the output layer and the number of nodes equals the network outputs (Y). For multi-class classification, the output layer uses the softmax function to produce a vector of values that sum to 1.0 representing the probabilities of each label. The label with the highest probability is assigned as the predicted label.

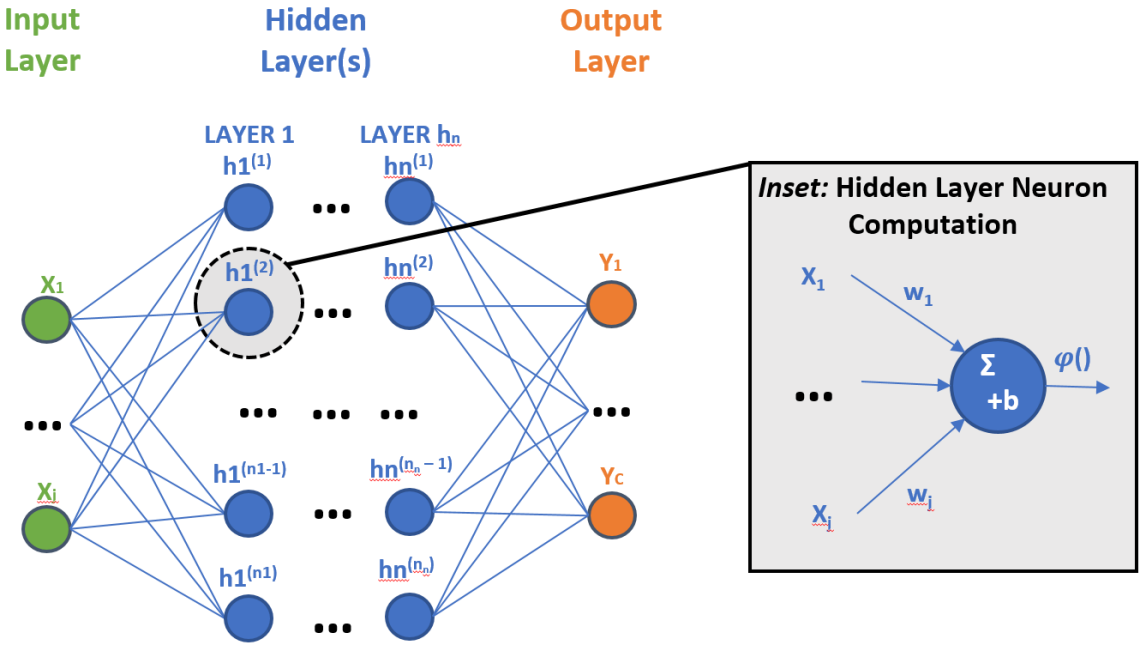


Figure 1.2 **Multilayer Perceptron (MLP) conceptual architecture showing input layer, hidden layer(s), and output layer. Every node in one layer is connected to every node in the next layer to form a fully connected network. The inset shows a neuron from the hidden layer and how it acts as a computational unit (see Equation 1).**

The layers between the input layer and output layer are known as the hidden layers. There can be one or multiple hidden layers that have the same or a different number of nodes. Each node in the hidden layer(s) is a neuron that functions as a computational unit. It takes the sum of the weighted input values, adds a bias (b), and passes it through an activation function (see inset of Figure 1.2 and Equation 2; Kanevski et al., 2009).

$$\varphi\left(\sum_{i=1}^{j}\omega_i x_i + b\right) \quad (2)$$

where:

φ	is the nonlinear activation function
w_i	are the weights
x_i	are the inputs to the neuron
b	is the bias

Activation functions are assigned to the hidden and output layers. For the hidden layer, the activation function transforms the summation and bias term into some form that can be used as input for the next layer. This is needed to control the amplitude of the output of the neuron (Kanevski et al., 2009) and to add non-linearity (higher degree of complexity) to the neural network (Jain, 2019). The output layer receives the values from the last hidden layer and transforms them into output values. For multi-class classification, Softmax is used as the activation function to convert inputs from the last hidden layer into a vector containing the probabilities that the sample belongs to each class. The output is the class with the highest probability.

The MLP must be trained before it can be used for predictions. This process requires data preparation (splitting, scaling), selection of hyperparameters, training, and testing. Data preparation first requires splitting the data into training and testing subsets. The training subset is used to establish the hyperparameters and weights of the MLP for the **training model**. A testing subset is withheld from training to evaluate the performance of the training model on predicting unseen data. MLP is sensitive to feature scaling (Buitinck et al., 2013), therefore input features should be scaled to have the same level of magnitude. The most common feature scaling techniques include normalization (transform data to range between 0 and 1) and standardization (transform data to have a zero mean and standard deviation of 1).

Hyperparameters are parameters that must be specified to configure the MLP model and whose value can not be estimated by the algorithm (Brownlee, 2020b) (e.g. number of hidden layers and neurons, iterations). There are often general heuristics or rules of thumb for configuring hyperparameters. Optimization methods can also be used to determine the combination of hyperparameters that achieves the best performance on a given dataset.

MLP uses supervised learning for training. **Supervised learning** means prior knowledge of what the network outputs should be for a given set of input features (e.g. existing lithology units at specific coordinates) is used to adjust weights. For multi-class classification tasks, this means that data points with known classes are used to establish the weights connecting the neurons in the network. One round of updating the weights using the entire training dataset is called an **epoch**.

Backward error propagation (back-propagation) is used by MLP as the supervised learning method to train the network. For multi-class classification, back-propagation is designed to iteratively minimize the error measured between the actual output of the neural network and the desired output. This is termed the **loss function** but can also be called cost function or error function. The initial weights are randomly assigned and then updated iteratively as training data is processed through the MLP. After each epoch, the output of the network is compared to the expected value and an error is calculated. This error is then propagated back through the network and the weights are adjusted based on the amount that they contribute to the error. Training is essentially adjusting weights to minimize a loss function and 'learning' occurs as long as the loss function changes.

Log-loss or cross-entropy is the default loss function used for multi-class classification (Brownlee, 2020d). Log-loss is closely related to entropy but represents how close the probability estimate is to the predicted class. It is calculated by taking the negative log of the probability estimate for the predicted class. The result is a positive number with zero indicating the most certainty. Higher entropy values indicate more uncertainty and occur when probability estimates are lower. These values can be summed and divided by the number of observations to get the log-loss (see Equation 2; Buitinck et al., 2013).

$$\log \text{loss} = -\frac{1}{N}\sum_{i=1}^{N}(\log(Pi)) \qquad (3)$$

Where: i is the label of the observation
$P(i)$ is the probability estimate of the label (e.g. HGU)
N number of observations

Once the training model has been validated and tested, then all data is scaled and used to establish the final weights of the **predictive model**. The scaling transformation becomes part of the predictive model since the same data preparation methods are required prior to running the predictive model. Further evaluation of the

predictive model is not required since the generalization capability of the modelling approach is determined during the final evaluation of the training model with the testing data.

1.2.3. Uncertainty Assessment in Geologic Models

Recent studies have recognized the importance of using automated, stochastic methods to develop geologic models for glacial depositional environments to account for complexity and to evaluate uncertainty (Refsgaard et al., 2012; Koch et al., 2014; Jørgensen et al., 2015; Toth et al., 2016). Stochastic is an adjective typically associated with randomness or probability. Both MLP and SOM are deterministic once the weights and the structure are fixed, and the algorithms are used to make predictions. However, randomness is part of training. Examples include random initialization of weights, the use of a stochastic gradient descent to optimize weights, the order of observations, or some chaotic behaviour due to nonlinearities (Brownlee, 2019b). Randomness during the process of training has the effect of fitting a different training model each time the algorithm is run on the same data. There are tactics to control randomness when training ANNs, such as defining a seed number so that the same sequence of random numbers is generated to improve reproducibility during training. There is also potential to take advantage of the stochastic nature of training ANNs to generate multiple geologic models although this is not explicitly explored in this thesis.

1.3. Purpose and Objectives

There is a need to effectively reproduce subsurface geologic complexity associated with glacial environments that can be used in groundwater models to support water resource management. The main purpose of this thesis is to develop a workflow to evaluate the performance of MLP on predicting hydrogeologic units (HGUs) for a region with complex glacial deposits. HGUs are proposed instead of stratigraphic units to better represent subsurface heterogeneity. There are limited examples in literature that use MLP to solve multi-class classification for geologic modelling of complex glacial deposits; therefore, this study evaluates different data selection alternatives and how this impacts the training and prediction capabilities of MLP. Research objectives include the following to address the purpose of this study:

- Advance semi-automated methods for standardizing material descriptions from borehole logs into HGUs by incorporating NLP techniques to provide a statistical approach to understanding geologic lexicon so that subjectivity is reduced, and uncertainty is better understood.

- Evaluate the feasibility of using MLP to predict HGUs using three data selection alternatives. This includes development of three workflows to process HGUs for geologic modelling purposes. Data preparation, tuning hyperparameters, and training/testing the MLP for each alternative is also required.

- Verify results by considering alternative analytical techniques, geologic cross-sections from independent studies, and hydrogeologic indicators within the region (e.g. artesian conditions).

This study explores the feasibility of using machine learning to interpret glacial deposits in the subsurface. Using MLP to create a geologic model has several advantages compared to more traditional interpolation methods in geoscience applications including the following:

- the capability of predicting HGUs (e.g. categorical data) based on training data derived from borehole logs which is typically the most abundant source of subsurface data for water resource applications (Russell et al. 2013);

- a 3D block-model approach can be applied to better represent geologic complexity instead of presenting the subsurface as a layered system;

- the stochastic nature of training MLP allows for multiple-geologic realizations that can be used to understand uncertainty; and

- MLP can be implemented using open-source resources with output files compatible with groundwater modelling software (e.g. MODFLOW).

1.4. Study Area

1.4.1. Location

The regional basin containing the Fraser Lowland in the southwest corner British Columbia, Canada and the northeast portion of Whatcom County, Washington, United States (Figure 1.3) was selected as the study area (herein referred to as Fraser-Whatcom Basin or study area). The study area was selected because aquifers exist within a complex sequence of glaciated materials (Halstead, 1986; Vaccaro et al., 1998). Recent groundwater licensing and stronger protection for ecosystems mandated under the *Water Sustainability Act* (SBC 2014, c 15) as well as more holistic approaches for land-use planning are also driving the need to better understand glacial aquifer systems within British Columbia.

The Fraser-Whatcom Basin covers an area of approximately 670 square kilometres (km^2). It includes gently rolling and flat-topped uplands (e.g. between 20 to 175 metres above sea level, masl) separated by wide, flat-bottomed lowlands (below 20 masl) bounded by mountainous terrain (e.g. greater than 175 masl) and the Strait of Georgia (Figure 1.3) (Halstead, 1986). Drainage is directed towards the Fraser and Nooksack rivers as well as local river systems (e.g. Nicomekl River, Serpentine River, Campbell River) that discharge directly to the sea.

The glacial deposits within the Fraser-Whatcom Basin have been well-studied. Historical investigations have been completed to advance the understanding of surficial deposits (Armstrong, 1977, 1976; Armstrong and Hicock, 1979, 1980; Armstrong, 1984; Washington Division of Geology and Earth Resources 2016), stratigraphic units (Easterbrook, 1963; Clague, 1976, 1986, 1989, 1991; Armstrong 1981; Clague and Luternauer, 1983b; Easterbrook, 1986; Cameron, 1989; Hamilton and Ricketts, 1994; Jones, 1999; Ward and Thomson, 2004; Eungar, 2014, Riedel, 2017), and groundwater resources (Halstead, 1986; Vaccaro et al., 1998; Ricketts, 2000; Scibek and Allen, 2006; Simpson, 2012; Bayless et al., 2017; Haj et al., 2018, Yager et al., 2019). Studies have been completed by national, provincial and municipal levels of government as well as by academic institutions and consultants.

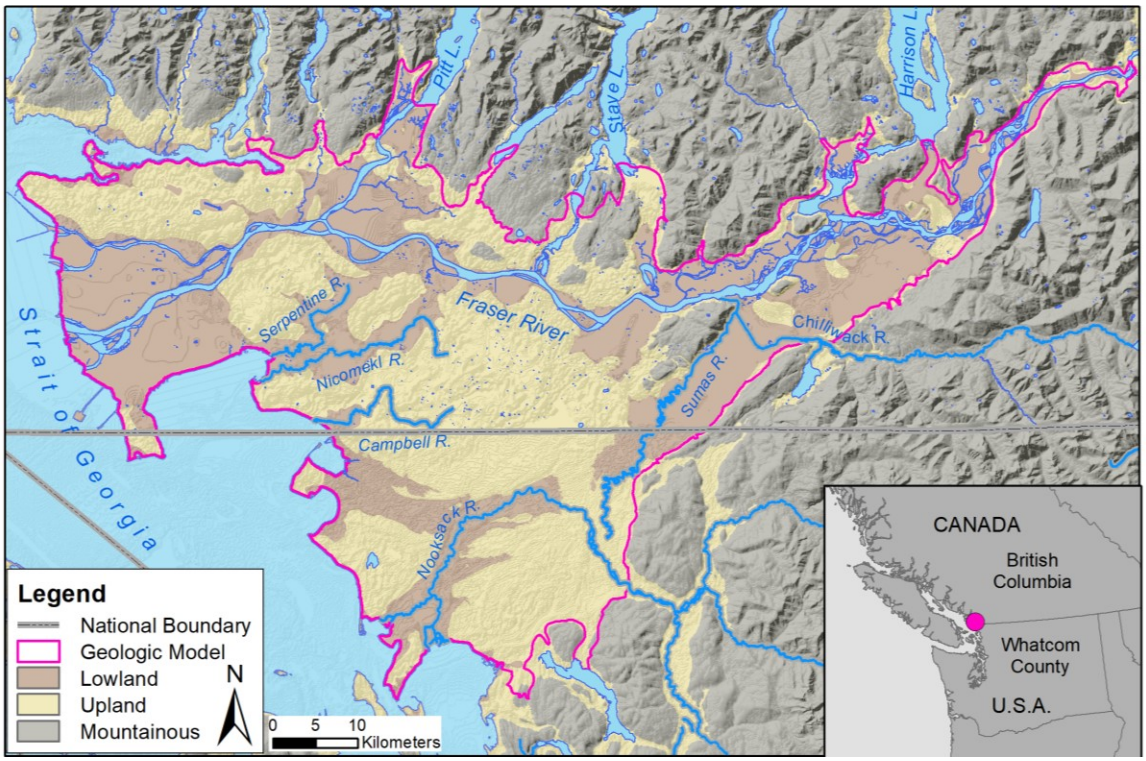


Figure 1.3 Fraser-Whatcom Basin location, topographical areas, and major drainage features. Source: Freshwater Atlas (BC) and National Hydrography Dataset 24k (Washington State) Digital Elevation Model (see Section 2.4.1).

The complex distribution of Quaternary sediments and the discontinuous nature of glacial units has resulted in a complex stratigraphic pattern. Geology is well mapped at the land surface but is generally less understood in the subsurface. Groundwater models have been developed for several municipalities that include geologic interpretations of the subsurface (Golder, 1997, 2005; Scibek and Allen, 2006; Advisian, 2018). Simpson (2012) developed one of the largest groundwater models for the region that focused on the central portion of the Fraser-Whatcom Basin. However, geology at the basin-scale has never been available to inform boundary conditions and to provide a greater understanding of groundwater flow.

This study is not intended as an exhaustive and detailed review of glacial sediments in the Fraser-Whatcom Basin. Some parts of the basin have limited subsurface data to support interpretations. However, a basin-wide perspective was used to allow an initial conceptualization based on a range of data densities. This may allow perspective on similarities and differences in distinct parts of the study area or conceptualization of how smaller, local deposits fit within the greater system.

1.4.2. Quaternary History

Multiple glaciations, fluctuations in sea level, and isostatic adjustments have contributed to the complex distribution of Quaternary deposits in the Fraser-Whatcom Basin. The stratigraphic framework is primarily based on Pleistocene deposits associated with the last glaciation (Fraser Glaciation) that started approximately 25 thousand years before present (ka ^{14}C BP, based on radiocarbon dating) and post-glacial deposits from the Holocene beginning approximately 10 ka ^{14}C BP (Figure 1.5, Figure 1.5). Quaternary deposits generally consist of nonglacial sediments and one to four regional drift sequences associated with repeated growth and decay of the Cordilleran ice sheet during the Fraser Glaciation as described below.

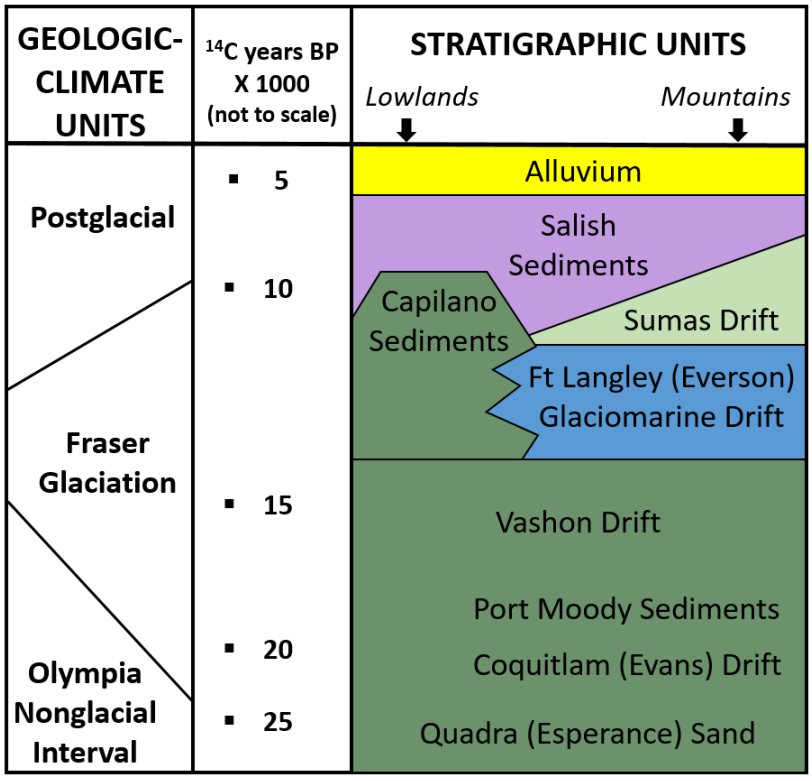


Figure 1.4 **Stratigraphic units for the Fraser-Whatcom Basin modified from Jones, 1999 (Figure 6A) and Ward and Thomson, 2004 (Fig. 2). Equivalent naming convention used in the United States are indicated in brackets.**

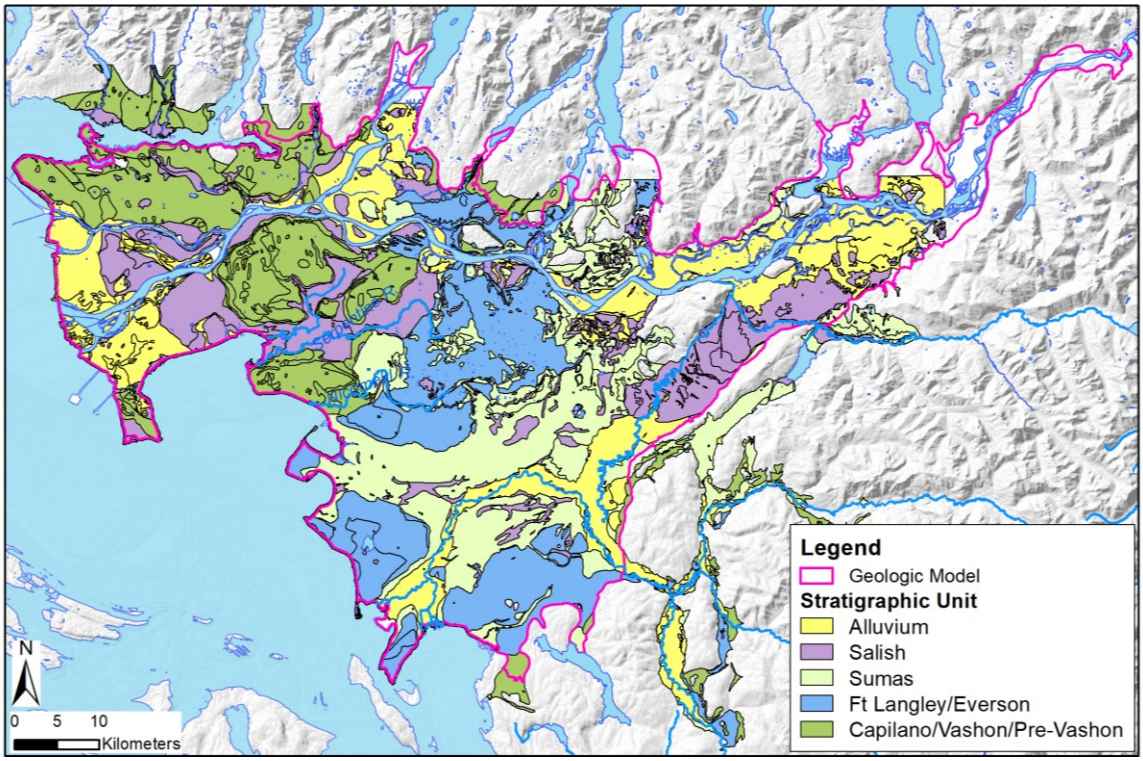


Figure 1.5 Stratigraphic units within the Fraser-Whatcom Basin modified from surficial mapping (Armstrong, 1976; Armstrong and Hicock, 1979, 1980; Armstrong, 1977; Washington Division of Geology and Earth Resources, 2016).

During the beginning of the Fraser Glaciation (approximately 26 to 22 ka ^{14}C BP), advancing outlet glaciers from the Cordilleran Ice Sheet flowing onto the coastal lowlands and down the Georgia Strait led to the deposition of thick, proglacial outwash called Quadra Sand (equivalent to Esperance Sand in the US). Quadra Sand consists of horizontally and cross-stratified, well sorted sand with minor silt and gravel (Clague, 1976). It is typically underlain by fluvial and marine sediments from the preceding Olympia nonglacial interval and overlain by glacial sediments from the Fraser Glaciation. It occurs below sea level but can also be found at elevations up to 100 masl as remnants of a formerly more extensive deposit that has been eroded by meltwater and glacial scour. Based on surficial mapping, Quadra Sand has been identified along the bottom slope of several upland areas (e.g. Vancouver Upland, Surrey Upland, White Rock Upland) but it is typically not exposed at the surface.

The Coquitlam Stade (equivalent to Evans Stade in the US) is associated with a climatic episode that occurred early in the Fraser Glaciation (21.3 to 18.7 ka ^{14}C BP). Glaciers from the Northshore mountains to the Pitt River valley flowed towards the

Fraser Lowland and coalesced with ice from the Strait of Georgia (Ward and Thomson, 2004). This early glacial advance only affected the northwest portion of the Fraser-Whatcom Basin and did not reach the Canadian-United States border. Glaciers of the Coquitlam Stade retreated and a nonglacial interval called the Port Moody Interstade occurred between 18.7 to 17.7 ka ^{14}C BP (Hicock and Armstrong, 1985; Hicock et al., 1999). Known deposits from the Coquitlam Stade and Port Moody Interstade are limited to the Pitt River Valley, Coquitlam River Valley, and in the Port Moody area.

By 17 ka ^{14}C BP, the Cordilleran Ice Sheet re-advanced as part of the Vashon Stade. The Vashon Stade was the most extensive glacier advance of the Fraser Glaciation. Glaciers from the southern Coast Mountains coalesced with glaciers from the Vancouver Island Ranges to produce two piedmont lobes. The Juan de Fuca lobe flowed west terminating on the continental shelf and blocked the Strait of Juan de Fuca. The larger Puget lobe flowed south into the North Cascade Range or terminated against the Olympic Mountains along the coastal lowland. Glaciers flowed along mountain fronts, impounding rivers and creating lakes in many of the mountain valleys (Thorson, 1980; Hicock and Lian, 1999). At the Vashon stadial maximum (~14.5 ka ^{14}C BP), the Fraser Lowland was completed covered by ice more than 1.5 km thick (Clague et al., 1997). Isostatic depression of the land mass occurred from the weight of the ice sheet.

The climate warmed after the glacial maximum, resulting in deglaciation of the Cordilleran ice sheet primarily by frontal retreat and downwasting (Armstrong, 1957). Once the Juan de Fuca lobe retreated and the Puget lobe thinned sufficiently to the north, marine waters inundated the Fraser Lowland and resulted in a calving embayment that induced rapid ablation of the glacier. Capilano sediments were deposited along the northwestern Fraser-Whatcom Basin during this period. This includes glaciomarine silt and clay containing stones dropped from melting icebergs and fossil shells. Large sand and gravel deltas formed at the mouths of many mountain valleys with major rivers that functioned as massive outwash channels.

Glacier retreat slowed and the ice margin stabilized in the central Fraser-Whatcom Basin where thick glaciomarine sediments were deposited (Clague et al., 1997). Ice from the Coast Mountains and BC interior continued to flow but was restricted to existing fjords and along the central Fraser Lowland. The glacier that persisted in the Fraser Lowland became land-based as isostatic rebound of the terrain started to occur

(Clague et al., 1997). This led to initial development of drainage networks on the emerged terrain while the sea level occupied the lower reaches of many of the paleovalleys (Kovanen, 2002).

At least four advances occurred during the Sumas Stade (14.5 to 11 ka BP). The most significant advance extended the glacier to the southwest across the Fraser Lowland and into deglaciated mountain valleys, blocking natural drainages (e.g. Chilliwack Valley) (Kovanen and Easterbrook, 2002). Interbedded glaciomarine and glacial sediments deposited in the central portion of the Fraser-Whatcom Basin are included in the Fort Langley Formation. Sumas till was deposited on top of glaciomarine sediments up to an elevation of 120 masl (Mathews et al., 1970). Sumas drift includes morainal and glaciofluvial (raised deltas, outwash floodplains) deposits related to the ice front itself or from deglaciation. This includes the formation of the Campbell River delta and lowland meltwater channel occupied by the modern Campbell River (Armstrong, 1981; Clague and Luternauer, 1983a).

The glacier retreated from the area by 10 ka ^{14}C BP (Armstrong 1981). Isostatic rebound of the land mass and sea level rise continued during the final retreat. Non-glacial sediments were redistributed by marine, fluvial, mass movement and aeolian processes. Aggradation of the floodplains and major deltas of the Fraser and Nooksack occurred and continues today. Peat accumulated in poorly drained areas in surface depressions or valley bottoms (e.g. Burns Bog, Nooksack floodplain). By 2 ka ^{14}C BP, the modern coastline position and sedimentary conditions were established (Clague, 1998).

1.5. Scope of Work

The scope of work for this study is outlined in Figure 1.6. It includes development of a geologic database, standardization of material descriptions from borehole lithology logs as HGUs, and creation of a 3D mesh for the study area. The geologic database includes information from well records and surficial geology mapping from both Canada and the United States. Material descriptions from borehole lithology logs are standardized using soil classification guidelines and incorporating NLP techniques to build on the semi-automated approach previously developed for the area (Allen et al., 2008). A mesh is required to facilitate processing of HGUs and to visualize results. It

provides a 3D representation of the study area, including a digital elevation model (DEM), that can be used to extract information (e.g. cell index, cell centroid coordinates) and to assign attributes.

The geologic database, HGUs, and mesh are used to develop three data selection alternatives for MLP training. Geologic realizations of the Fraser-Whatcom Basin are created using the MLP predictive model developed from each alternative. The geologic realizations are compared and the best outcome verified to further evaluate the performance of MLP at interpreting the distribution of glacial deposits. The methodology for each task is described further in Chapter 2.

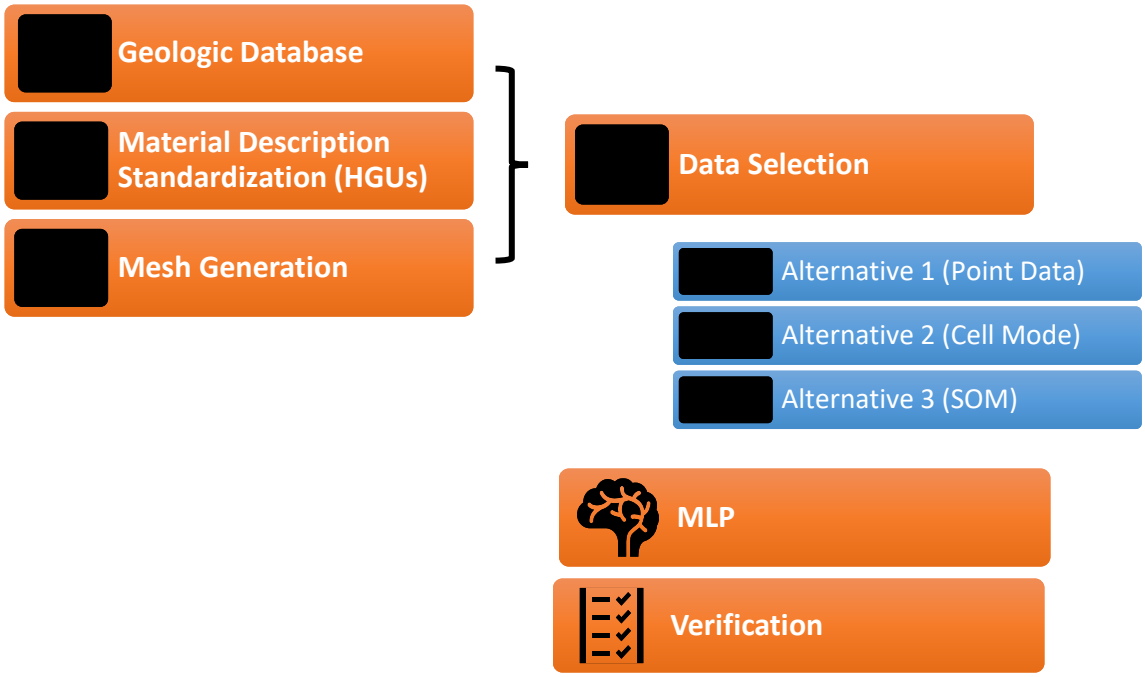


Figure 1.6 **Overview of scope of work. The first steps include development of a geologic database, material description standardization and mesh generation. This information is used to generate three data selection alternatives to develop the MLP algorithm with the best outcome verified to further evaluate MLP performance.**

1.6. Thesis Organization

This thesis is organized into the following chapters:

Chapter 1: Introduction – This section presents the purpose, objectives, and scope of work based on a review of literature. A description of the study area and its glacial history is also presented.

Chapter 2: Methodology – The methods used to complete the scope of work are presented including a description of software, development of the geology data, material description standardization, mesh generation, data selection, MLP training including performance metrics, and the verification approach.

Chapter 3: Geologic Model Results – The results of the MLP modelling using the three data selection alternatives are presented and compared to determine the best geologic realization for further verification.

Chapter 4: Verification Against Subsurface Interpretations and Hydraulic Indicators – The best geologic realization is verified against alternative analytical techniques (k-nearest neighbour), subsurface interpretations from independent studies, and hydraulic indicators for the region.

Chapter 5: Conclusions and Recommendations – Conclusions related to the feasibility of using MLP to interpret complex glacial deposits in the subsurface are provided. This chapter also includes recommendations for future research.

Chapter 2.

Methodology

2.1. Software

Open source resources were primarily used to undertake the study (Figure 2.1). The various packages were accessed in Jupyter notebooks (Kluyver et al. 2016) using python programming language with final visualization of results using Paraview. Jupyter notebooks facilitates interactive computing that captures development, documentation, execution of code, and visualization of results under one software platform. ArcGIS Desktop (Version 10.7) was also used to process geospatial data as needed.

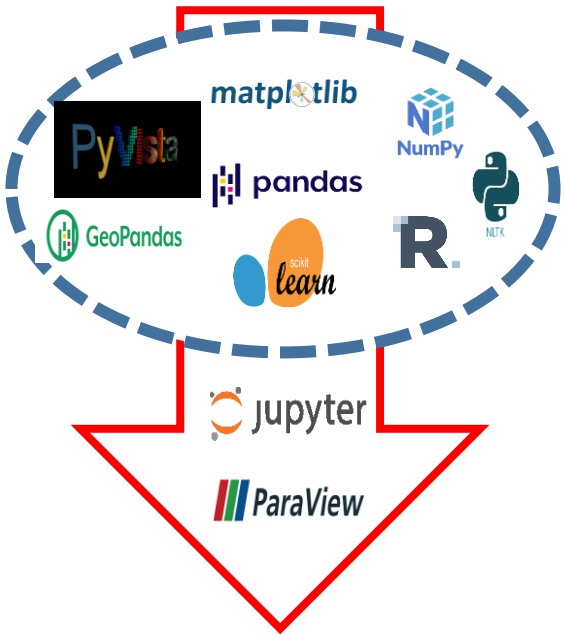


Figure 2.1 Open Source Resources.

Python packages include a collection of related modules that contain code to execute computing tasks. Packages facilitate modular programming which allows breaking-up a large programming task into separate, smaller, more manageable subtasks. The packages also provide access to code that can be easily reused, thus reducing coding effort. Key packages used in this study include the following: GeoPandas (v. 0.8.1) (Jordahl, 2020), Matplotlib (v. 3.3.4) (Hunter, 2007), NLTK (v. 3.5) (Bird, 2009), NumPy (v. 1.19.4) (Harris et al., 2020), Pandas (v. 1.2.2) (Pandas Development Team, 2020), Pyvista (v. 0.27.2)) (Sullivan and Kaszynksi, 2019), Scikit Learn (v. 1.0.1) (Pedregosa et al., 2011; Buitinck et al., 2013), Minisom (v. 2.3.0) (Vettigli, 2018), and Hyperopt (v. 0.2.5) (Bergstra, 2013). These packages are used for data processing, graph and table preparation, mesh generation, and algorithm development.

Paraview (v. 5.9.0) (Paraview Developers, 2022) was used for analysis and 3D visualization of the geologic realizations. Paraview is an open-source, multi-platform data analysis and visualization application developed to analyze large datasets.

2.2. Geologic Database

2.2.1. Data Sources

Data sources used to compile the geologic database include borehole data and surficial mapping. A summary of the borehole and surficial geology mapping data sources is provided below.

Borehole Data

Borehole data sources include publicly available data collections managed by government agencies in both British Columbia and Washington State (Table 2.1). Available borehole data varies in size, spatial density, completeness, accuracy and format. Borehole data was collected using different drilling equipment with different aims and objectives by individuals with various levels of geologic training.

Canadian data sources include legacy oil and gas wells from the British Columbia Oil and Gas Commission (BCOGC), borehole log information collected by the Environmental Management Branch (EMB) of the British Columbia Ministry of Environment and Climate Change Strategy (BCMEC), and water well records compiled by BCMEC. Legacy oil and gas wells in British Columbia (BCOGC) were searched using the Well Authority Number to find lithology data in the BCOGC eLibrary. EMB collects a representative borehole (BCBH) submitted by a third party for sites that have applied for a regulatory instrument under the *Contaminated Sites Regulation* (B.C. Reg. 375/96 O.C. 1480/96). Water well records from BCMEC include wells that have been voluntarily registered (wells constructed prior to 2005) and wells required to submit well reports under the BC *Groundwater Protection Regulation* (B.C. Reg. 39/2016 O.C. 113/2016). Water well records were accessed using the GWELLS data repository.

American data sources include legacy oil and gas wells compiled by the Washington Geological Survey (WGS), a subsurface database managed by the WGS, and well records accessed through the Washington State Department of Ecology (WAECY).

Table 2.1 Borehole data sources.

Agency	Application/ Product	Acronym	Original file format of Well Collar	Original file format of Lithology	Hyperlink
British Columbia Environmental Management Branch (EMB)	Borehole Log Lithology	BCBH	Shapefile	Shapefile	https://catalogue.data.gov.bc.ca/dataset/borehole-log-lithology-public-view accessed August 2021
British Columbia Oil & Gas Commission (BCOGC)	eLibrary	BCOG	Shapefile	PDF or image	https://catalogue.data.gov.bc.ca/dataset/well-bottom-hole-event accessed January 2019 https://www.bcogc.ca/energy-professionals/online-systems/elibrary/ accessed June 2021
British Columbia Ministry of Environment and Climate Change Strategy (BCMEC)	GWELLS	BCWR	Tabulated	Tabulated	https://apps.nrs.gov.bc.ca/gwells/ accessed April 2021
Washington Geological Survey (WGS)	Oil and Gas Wells	WSOG	Shapefile	PDF	http://www.dnr.wa.gov/publications/ger_portal_oil_gas_wells.zip accessed August 2018
Washington Geological Survey	Subsurface Database	WSSD	Shapefile	Tabulated	https://www.dnr.wa.gov/programs-and-services/geology/publications-and-data/gis-data-and-databases accessed June 2018
Washington State Department of Ecology (WAECY)	Well Log Viewer	WSWR	Shapefile	PDF or image	https://appswr.ecology.wa.gov/wellconstruction/map/WCLSWebMap/default.aspx accessed August 2020

Legacy oil and gas well data in Washington State (WSOG) were reviewed but interpretations of geophysical logs were not available or the documentation on file did not provide lithology information. The Washington State subsurface database (WSSD) is a compilation of data from various agencies (e.g. Washington State Departments of Transportation and Health, United States Geological Survey (USGS), local county and city governments, geotechnical firms) with a location precision of 30 m or better (Eungard, 2014). WAECY provides access to well records (WSWR) via a well log viewer application. Well records are centred within a section of the surveyed land system and do not represent the physical location of the well; actual locations differ by up to 300 m (Eungard, 2014). As such, multiple well reports from various wells can exist at the same location. Each well has a unique well tag ID that was used to search for lithology logs that were manually entered in the subsurface database.

Well records have been recognized as valuable data sets for regional studies despite their known data limitations (Russell et al., 1998; Allen et al., 2008; Arihood, 2009; Bayless et al., 2017). An understanding of data limitations is important to extract and standardize data in a manner that provides meaningful information (Russell et al., 1998). Common issues include inaccurate well locations, data duplication, variable sample collection methods and competency of logger that can influence the accuracy and detail of lithologic descriptions, vertical sampling bias (only drill until aquifer material/water is encountered), errors with data entry, and inconsistent geologic terminology. These databases are also continuously evolving with changes to format, reporting requirements, and/or submission methods over time.

Surficial Geology Mapping

Surficial geology mapping was used to improve the lateral distribution of HGUs at the surface where no boreholes exist and to provide additional horizontal/vertical delineation of bedrock where outcrops have been mapped. Digital surficial geology mapping published by the Geological Survey of Canada (GSC) (Dunn and Ricketts, 1994) and the Washington Division of Geology and Earth Resources (2016) were used. These data sources were converted to shapefile format and re-projected to NAD 1983 UTM zone 10. Manual addition of missing map unit labels was done based on comparison with original GSC maps. The re-projected data sources were combined and further processed to

remove map boundaries by merging adjacent polygons with common attributes, resulting in a total of approximately 1,500 polygons.

2.2.2. Well Data

General well details were compiled to form part of a relational database that links to lithology data. Data was processed to include the attributes summarized in Table 2.2 for wells within the study area. Data formats, coordinate systems, and units of measurement were standardized. Wells without lithology or with lithology errors (see Section 2.2.3) were removed from the dataset. Screen interval data is not included given limited data availability.

Table 2.2 Well attributes.

Attribute	Unit/Data Type	Description
Unique identification[1]	-	Combination of source and original identification
Source	Categorical	BCOG, BCBH, BCWR, WSSD, WSWR. See Table 2.1 for an explanation of the source abbreviations.
Original identification	-	Unique identifier in the original dataset
Well type	Categorical	Other (OTHER) Canadian - Oil and gas (OG), environmental (ENV), domestic (DOM), irrigation (IRR), drinking water system (DWS), commercial (COM), test well (TST), closed-loop geoexchange (OP_LP_GEO), BC Observation Well (OBS) USA – water (W), resource protection (R), decommissioning (D)
X	m	Easting, UTMNAD83 Zone 10
Y	m	Northing, UTMNAD83 Zone 10
Ground surface elevation	masl	Extracted from digital elevation model
Construction date	YYYY-MM-DD	Well construction date
Status		New, Alteration, Abandoned, Closure, Other
Lithology depth	mbgs	Maximum lithology depth
Water level	mbgs	Water level typically recorded after well construction
Artesian conditions	True/False	

Notes:
1) Required since unique identifiers were duplicated when the data sources were merged.
masl – metres above sea level
mbgs – metres below ground surface

Additional processing of well data was required for the BCBH and WSWR datasets. BCBH is formatted such that well details are repeated for each interval of lithology data. Well details were separated from the lithology data and grouped to provide a unique dataset for each well.

WSWR also required additional processing since multiple well reports from multiple wells can exist at a given location (Eungard, 2014). To improve efficiency, the well report having the deepest depth within a one (1) km grid spacing was selected for manual entry of lithology data. As such, the density of subsurface information in United States portion of the study area is much lower compared to the Canadian portion. The coordinates provided in the database (e.g. center coordinates of the section within the surveyed land system) were used instead of the physical location of the well.

The resultant well locations are shown in Figure 2.2 with a summary of data counts grouped by source provided in Table 2.3. Most lithology data are provided from BCWR. Boxplots showing well depth (based on the maximum lithology depth) for each data source are shown in Figure 2.3. BCOG wells provide the deepest subsurface information but there are only 12 wells within the dataset. BCBH and WSSD locations are relatively shallow (<20 mbgs) while both BCWR and WSWR have subsurface information that is generally less than 100 m deep.

Table 2.3 Well data key attribute counts.

Data Source	Well	Date	Water Level	Artesian Well
BCBH	780	780	460	0
BCOG[1]	12	12	0	0
BCWR	12,323	9,907	9,192	600
WSSD	24	24	0	0
WSWR	780	721	28	1

Note: See Table 2.1 for an explanation of the source abbreviations. Excludes wells without lithology or flagged as having lithology errors.

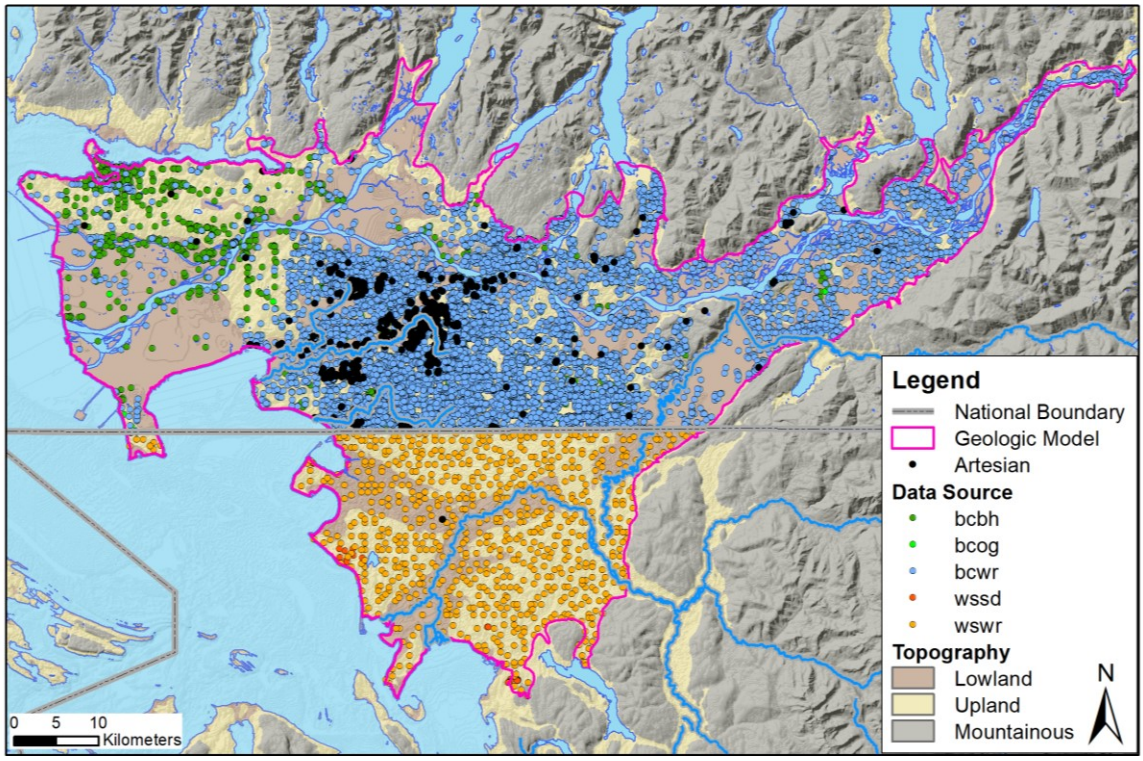


Figure 2.2 **Well locations based on data source. Excludes wells without lithology or flagged as having lithology errors. The majority of the artesian wells are located in the Nicomekl-Serpentine river valleys. See Table 2.1 for an explanation of the source abbreviations.**

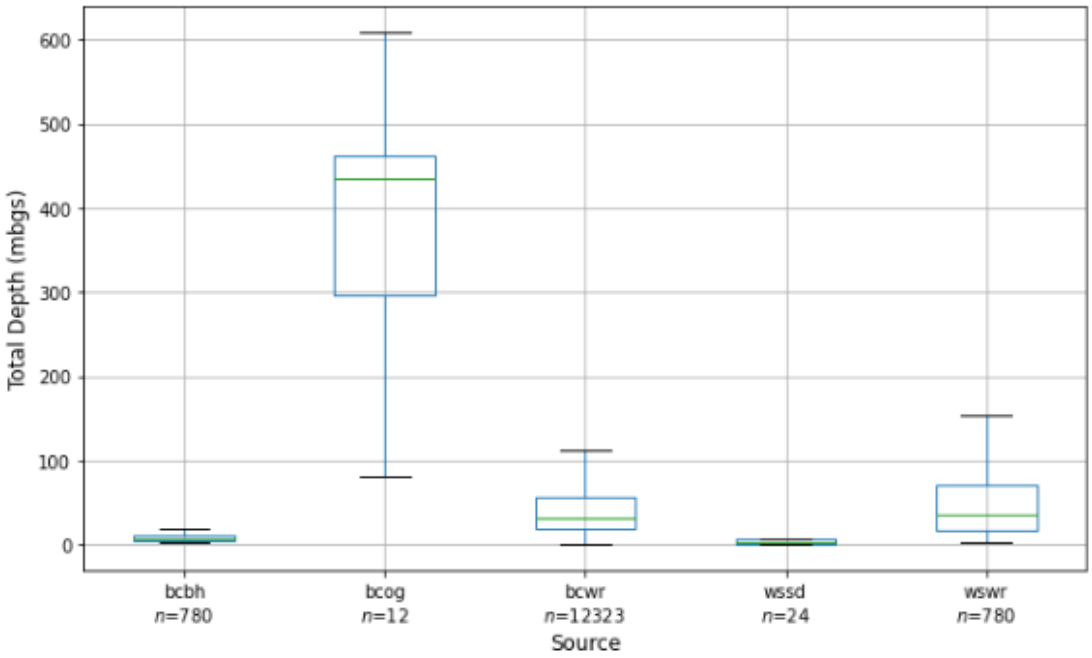


Figure 2.3 **Boxplot of well depth. See Table 2.1 for an explanation of the source abbreviations.**

2.2.3. Lithology Data

The lithology table provides interval data of material descriptions and hydrogeologic properties. Material descriptions describe the physical nature and state of the subsurface at discrete intervals during borehole drilling. Data was processed to include the attributes summarized in Table 2.4 for wells screened within the study area. Lithology processing steps included compiling lithology from different data sources, converting measurement units, combining data sources, cleaning interval data, and cleaning material descriptions (Figure 2.4).

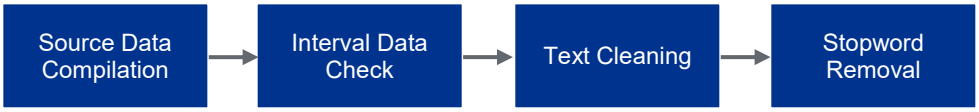


Figure 2.4 Process for lithology data-cleaning.

Table 2.4 Lithology attributes.

Attribute	Unit/Data Type	Description
Unique identification	-	Combination of source and original ID
Source	-	BCOG, BCBH, BCWR, WSSD, WSWR. See Table 2.1 for an explanation of the source abbreviations.
Original identification	-	Unique identifier in the original dataset
From	mbgs	Start of interval relative to ground surface
To	mbgs	End of interval relative to ground surface
Description	text	Material description
Hydraulic conductivity	m/s	Hydraulic conductivity
Yield	USGPM	Well yield estimate

Lithology data was compiled from various sources in different format types. GWELLS required additional compilation steps to avoid missing data and to remove problematic interval data. Firstly, categories for surficial material, bedrock material, colour, hardness, and water content were added in 2008 to well construction forms used to update GWELLS as an alternative to manual entry of material descriptions. As such, lithology information could be submitted by providing material descriptions or by selecting categorical terms. If a material description was not available, concatenated categorical terms were alternatively used. Secondly, it is common for older wells in GWELLS to have rows with intervals starting and ending in zero because of historical

limitations on character length for data entry. Additional rows exist below the original interval with these zero intervals to provide a continuation of text. The zero intervals were removed assuming primary materials are typically described at the beginning of material descriptions.

Lithology for BCOG and WSWR were manually entered since information was not available in an electronic format. Detailed information below the bedrock surface was not entered for WSWR wells. A cut-off limit was also used for very deep BCOG wells (>300m deep) given the lower model boundary for this study (See Section 2.4.2). For very deep BCOG wells, manual entry of lithology data was generally completed to the first low permeability unit 300 m below the surface.

Wells with any of the following interval errors were flagged and removed from the dataset as part of data cleaning:

- rows with null values for either the start or end of the interval;
- same value for the start and end of the interval;
- deeper start of the interval compared to the end; and
- end of the interval does not match the start of the interval below.

Material descriptions provide text data that require cleaning prior to use. There are various text cleaning methods that can be considered depending on how the data will be used. The primary objective of data cleaning for this study is to develop a meaningful geologic corpus that will inform the classification of HGUs. Text cleaning was completed by making all text lower case, removing text (punctuation, numbers, stopwords), correcting key spelling mistakes, as well as standardizing short-forms, abbreviations, acronyms, and regional naming conventions. Spelling corrections and standardization focused on words that occur the most frequently and were not intended to capture all occurrences. Secondary descriptors for grain size (e.g. fine, medium, coarse) were also removed since this level of detail was not necessary at the regional scale.

Summary statistics for the processed lithology dataset are shown in Table 2.5. The processed lithology dataset contains approximately 87,800 rows (originally contained 112,913 rows) with over 20,000 unique material descriptions. The BCBH includes hydraulic conductivity for targeted subsurface intervals, which is particularly

useful for characterizing hydrogeologic properties. Yield estimates associated with intervals from the lithology dataset that can be used to infer hydraulic properties are also recorded in the GWELLS; however, these are currently limited in number but will become a beneficial data source in the future.

Table 2.5. Summary statistics for key lithology attributes.

Data Source	Material Description Counts	K Count	Yield Count
BCBH	3,699	211	0
BCOG	103	0	0
BCWR	77,214	0	1,651
WSSD	245	0	0
WSWR	6,241	0	317
Total	87,502	211	1,968

Note: See Table 2.1 for an explanation of the source abbreviations.

Material descriptions contain short sentences with numerous words in various combinations based on a specialized geological lexicon. After text cleaning, fewer than five words are typically used, with two or three words being the most common (Figure 2.5). The top 20 words are shown in Figure 2.6. Further discussion is provided in Section 2.3.

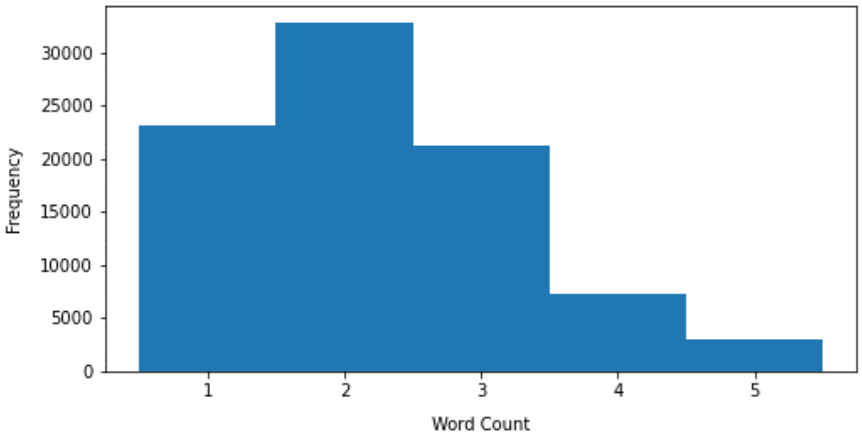


Figure 2.5 Word count histogram of the processed material descriptions.

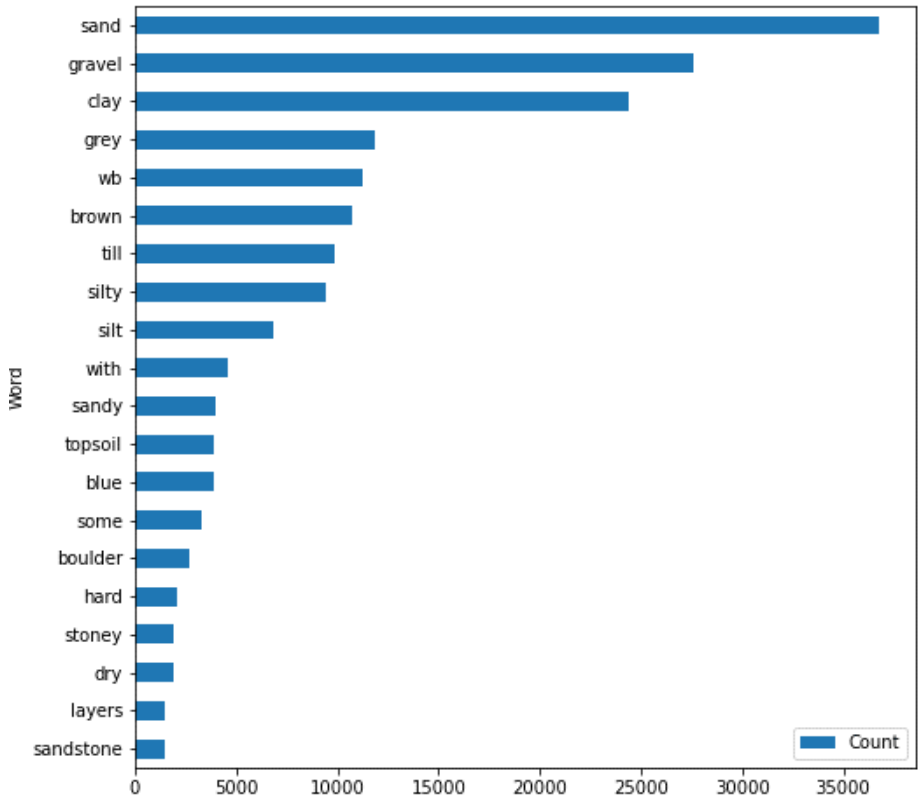


Figure 2.6 Top 20 words from the processed material descriptions.

2.2.4. Surficial Data Points

Digital surficial geology mapping was used to recreate the geomorphologic map prepared by Kovanen and Slaymaker (2015) so that surficial data points could be extracted using a single interpretation for the entire study area. Surficial data points were used to supplement subsurface data, particularly in areas where no borehole information was available. A summary of the methodology used to recreate the geomorphologic map and to determine surficial data points is provided below.

The twelve geomorphic units identified by Kovanen and Slaymaker (2015) for the study area were assigned based on the geologic period (Holocene or Pleistocene) and the sub-division of stratigraphic units contained as attributes in the surficial geology mapping shapefile, with some modification to match geomorphologic mapping particularly for Glacial Modified, Pleistocene units. All Pre-Vashon deposits were assigned as Glacial Modified, Pleistocene. Glaciomarine units were generally mapped below an elevation of 175 masl, which coincides with the marine limit inferred for the study area (Armstrong, 1957). The resultant geomorphology map for the study area is

shown in Figure 2.7. Minor differences exist primarily due to local modern drainage patterns/mass movement areas refined by Kovanen and Slaymaker (2015) that were not replicated given their limited surficial coverage and the regional focus of this study.

HGUs (see Section 2.3), including clay, coarse, fines, till and bedrock, were mapped primarily based on geomorphic units (e.g. glaciofluvial assigned as coarse) with consideration of material descriptions of select polygons consolidated from surficial geology mapping (e.g. fluvial assigned as coarse or fines depending on the material description of the polygon). A grid spacing of 1 km was used to generate point data for HGUs. This spacing is consistent with borehole data in the American portion of the study area. The northing and easting of the point data were based on the center of the grid. For bedrock outcrops, additional points were added from the ground surface and then every 5 m to a depth of 100 m to avoid categorical imbalance. The plan view distribution of this additional data is shown in Figure 2.8.

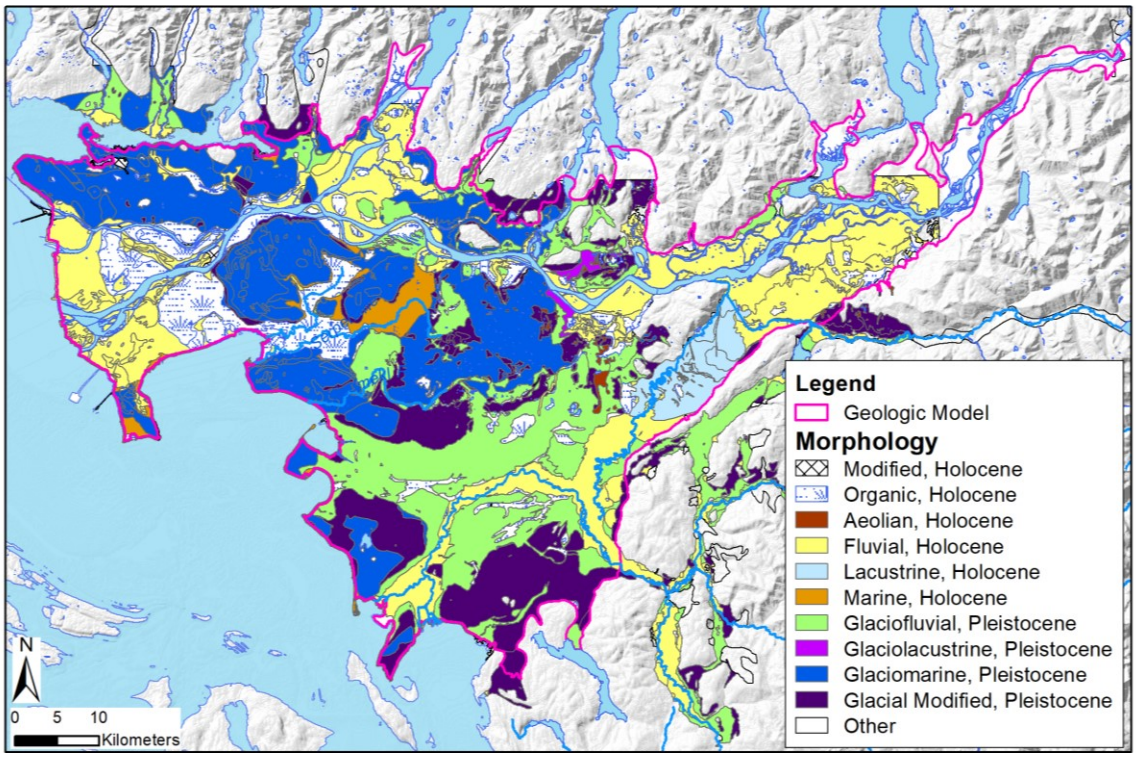


Figure 2.7 Geomorphologic mapping (modified from Kovanen and Slaymaker 2015).

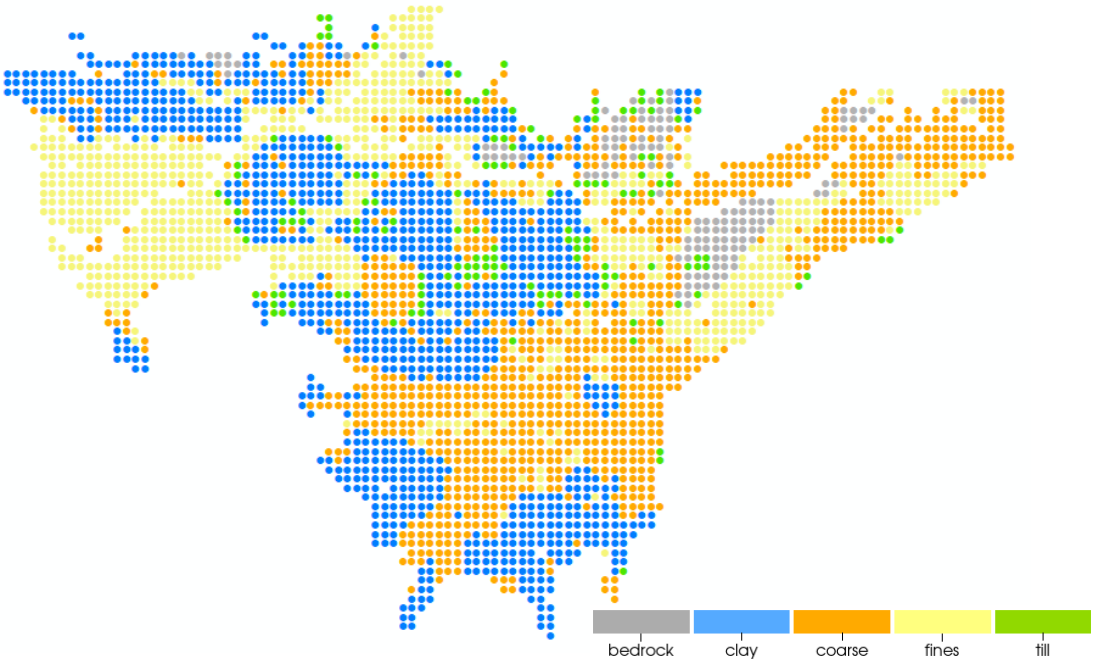


Figure 2.8. Plan view showing the distribution of additional surficial data points.

2.3. Material Description Standardization

2.3.1. General Approach

The lithology dataset contains large amounts of textual data that can be used to classify similar materials and provide qualitative information about the subsurface. Semi-automated methods that include Natural Language Processing (NLP) techniques are used in this study to classify material descriptors to support 3D geologic model development.

Classification is required to group material descriptions into HGUs for geologic model development. Classification schemes with detailed textural and structural information are not appropriate for databases when water well records are the single largest data source (Russel et al., 1998; Yager et al., 2018). This is because variations in geological lexicon can exist due to sample collection methods, material (soil) classification methods, experience of field personnel, drilling purpose, and local differences in geology; therefore, a generalized approach to classifying material descriptions that provides a region-wide perspective is more applicable. For the purpose of this study, material classification is primarily based on size and composition of grains using the American Association for Testing and Materials (ASTM) modified Unified Soil

Classification System (USCS) as a guide to lumping materials into more general lithologic units. Various soil classification systems exist but the ASTM USCS was chosen since it (or variations thereof) is commonly used in North America.

Given there are over 20,000 unique material descriptions, a three-step process was used to classify material descriptions as described below and schematically shown in Figure 2.9:

- The first step considers sentence structure and the relationship between words to identify descriptors that best describe material composition.
- The second step aggregates single and combinations of descriptors into material groupings to capture the most significant variations. Multiple descriptors were considered to improve the geological accuracy of classification results which is consistent with the general approach used in other regional studies (Russell et al., 1998; Allen et al., 2008; Bayless et al., 2017).
- The last step assigns HGUs to material groups, with consideration of hydrogeologic properties, to reduce complexity for modelling purposes.

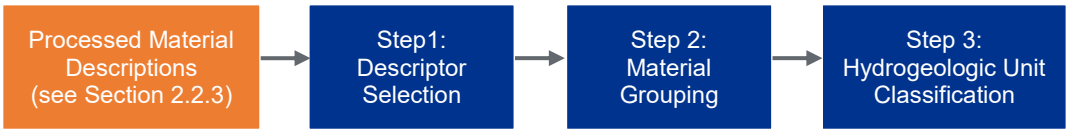


Figure 2.9 Process for material classification.

2.3.2. Descriptor Selection

The first step of the material classification process considers sentence structure and the relationship between words to identify descriptors that best describe material composition. Under the ASTM USCS, materials can be described using nouns for primary components (e.g. 50% by weight or more), descriptive adjectives ending in a ‘y’ (e.g. sandy, clayey, gravelly) for secondary components (greater than 30% by weight), and nouns following minor descriptive words (e.g. ‘with’, ‘trace’, ‘some’) for minor components (less than 30% by weight). For example, material described as ‘silty sand

and gravel with clay' is interpreted to contain both 'sand' and 'gravel' as primary components, silt as a secondary component, and clay as a minor component.

Minor Components – Do they impact classification?

Based on the ASTM USCS, minor components (less than 30% by weight) are typically described following minor descriptive words. In this study, minor descriptive words found in the dataset include: few, some, trace, odd, with, minor, very little, little, no, and occasional. Both 'with' and 'some' frequently occur in the dataset (Figure 2.6). Nouns following minor descriptive words could overemphasize their importance. Minor components are not considered important for regional modelling purposes. In this study, truncating material descriptions at minor descriptors did not significantly change word count distributions for the study area, suggesting minimal loss of textual information.

In addition, there are likely some records with various particle ranges following minor descriptors that are used to characterize diamicton (e.g. 'silty sand and gravel with clay and cobbles'). These records could be interpreted as till (glacial deposit), although other processes that deposit diamicton exist (e.g. landslides, debris flow). Direct interpretations of 'till' are typically provided in the lithology dataset instead of multiple words describing diamicton. This is based on the frequency of the word 'till' (Figure 2.6) and given that material descriptions are generally very short (Figure 2.5). Given the above, the word 'till' is used to categorize till. Material descriptions that include multiple descriptors to characterize diamicton are not considered for this study. Further textual analysis may be warranted to understand how material descriptions for diamicton can be processed in the future.

Based on the above, text following minor descriptive words in the processed material descriptions were ignored for classification purposes. The benefit of truncating processed material descriptions is that any remaining nouns used to describe composition can be considered as primary components with equal weighting. Alphabetical sorting can be done to reduce dimensionality (e.g. sand gravel is the same as gravel sand). It is interesting to note that sorting would not be required if word embeddings were used since the same numerical value would be obtained regardless of order.

Secondary Components – Are they important?

Descriptive adjectives ending in a ‘y’ (e.g. sandy, clayey, gravelly) apply to secondary components (greater than 30% by weight) based on the ASTM USCS. ‘Silty’, ‘sandy’, and ‘stoney’ are common secondary components in the dataset (Figure 2.6). In this study, a NLP technique called bigrams was used to review the significance of secondary components and to inform how they should be considered in the classification approach. Bigrams are two-word combinations that are typically more informative compared to individual words. The frequency of the 25 most common bigrams isolated from truncated material descriptions is shown in Figure 2.10.

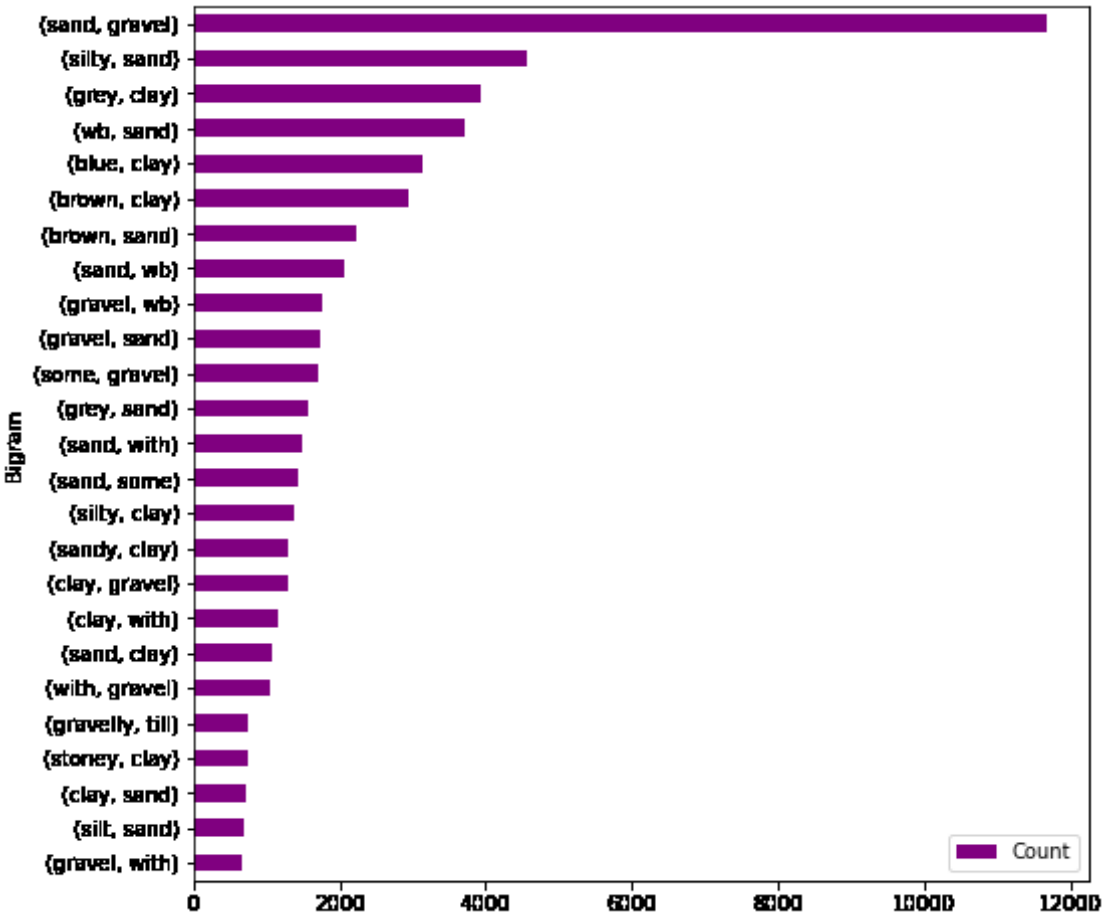


Figure 2.10 Most frequent bigrams (two word combinations) from the processed material descriptions.

The most common bigrams with primary and secondary components include the following in descending order: silty sand, silty clay, sandy clay, gravelly till, and stoney clay. ‘Silty’ occurs with multiple primary components (sand, clay). ‘Silty sand’ is the

second most common bigram with only ‘sand gravel’ occurring more frequently in the dataset. Clay is most frequently associated with a secondary component (‘silty clay’, ‘sandy clay’, ‘stoney clay’).

There may be similar deposits with different material descriptions that are not grouped together by ignoring secondary components. For example, ‘stoney clay’ and ‘gravelly till’ would be classified as clay and till, respectively, if secondary components are ignored. They could both be interpreted as glaciomarine or glacial deposits. Both ‘stoney’ or ‘gravelly’ are typical characteristics of Capilano Sediments or Vashon Drift deposits in the study area. On the other hand, ‘silty’ and ‘sandy’ are more general and not associated with a specific regional deposit.

From a hydrogeological perspective, secondary components could impact permeability, particularly if these components are fine grained. Both ‘stoney clay’ and ‘gravelly till’ would be assigned low permeabilities assuming glaciomarine or glacial deposition. However, the permeability of ‘silty sand’ is expected to be within the lower range of values for clean sands although there could be considerable overlap (Freeze and Cherry, 1979). Conversely, the permeability of ‘silty clay’ or ‘sandy clay’ is likely to be in the higher range of permeability values for clay. Clay is typically associated as an aquitard unit that restricts vertical movement of groundwater and provides protection to underlying aquifers from surface contamination.

To avoid over representing high and low permeability units, the word combinations of ‘silty sand’, ‘sandy clay’, and ‘silty clay’ were specifically included as part of the material classification approach. These word combinations frequently occur and have the potential to impact groundwater modelling results. Differentiation between ‘stoney clay’ and ‘gravelly till’ is not deemed critical for this study but may warrant further review to better represent stratigraphy in the subsurface.

Primary Components – Which ones are important?

Using ASTM USCS, primary component (e.g. 50% by weight or more) are described using nouns typically at the beginning of the description. The most common primary components were identified based on a review of the 100 most common words from the truncated material descriptions. Primary components are presented in Table 2.6 and include descriptors for both consolidated and unconsolidated materials.

Table 2.6 Most common primary components (nouns) from the processed material descriptions.

Subsurface Type	Primary Components
Bedrock	bedrock, coal, granite, sandstone, shale, siltstone
Quaternary	boulder, clay, cobble, fill, gravel, peat, sand, silt, till, topsoil

2.3.3. Material Grouping

Material groupings are used to aggregate single and combinations of descriptors to capture the most significant variations. Primary components identified in Table 2.6 and select combinations of primary and secondary components (e.g. 'silty sand', 'sandy clay', and 'silty clay') were considered as descriptors. Regular expressions were used to isolate descriptors from truncated material descriptions. Duplicate descriptors were removed (e.g. silt, sand, sand -> silt, sand) and records with multiple descriptors were alphabetically sorted to further reduce dimensionality (e.g. silt, sand -> sand, silt) prior to manual assignment of material groupings (Figure 2.11).

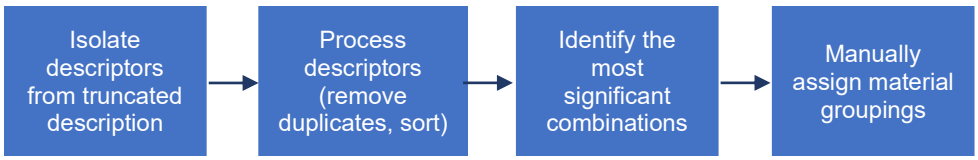


Figure 2.11 Process for material groupings

The dataset contains 274 combinations, including single and multiple descriptors. The 45 combinations with the highest cumulative thickness were selected since they capture the most significant variations (top 45 combinations). As shown in Figure 2.12, the cumulative thickness of the top 45 combinations is not evenly distributed. The cumulative thickness of the top five combinations (clay, sand, gravel sand, till, and gravel) accounts for approximately 65% of the total thickness of the lithology interval data. Some descriptors (topsoil, peat, fill) are common in the dataset (Figure 2.6), but they have a relatively small cumulative thickness even if other variations that include these words are considered (e.g. fill gravel sand). Conversely, 'bedrock' occurs less frequently compared to other descriptors but has a relatively high contribution to cumulative thickness.

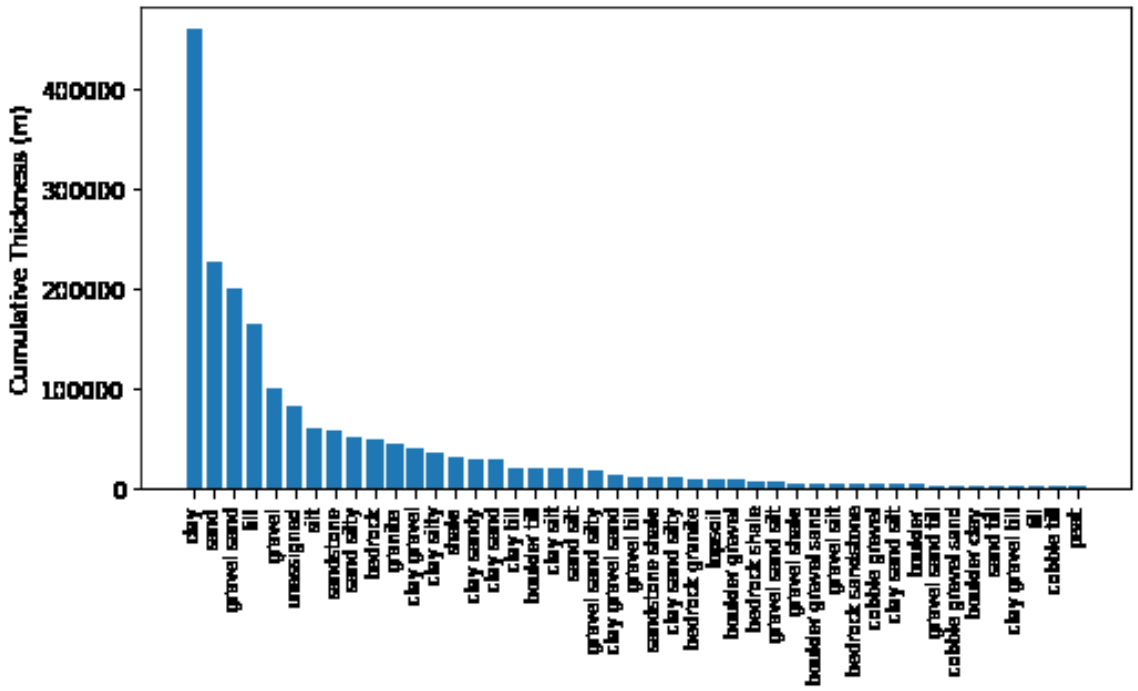


Figure 2.12 Top 45 descriptor combinations based on cumulative thickness.

Ten material groupings were used to aggregate the top 45 combinations as shown in

. Material groupings are loosely based on ASTM USCS but are generally more simplified (e.g. one grouping for silt) and include additions such as till and bedrock as well as consideration of boulder and cobble grain sizes. Fill, topsoil, and peat were grouped as unassigned (N) given their limited cumulative thickness from a regional perspective, although it is recognized that peat is significant in some parts of the study area (e.g. Burns Bog).

The cumulative thickness values of material groupings are shown in Figure 2.13. Material groupings are dominated by clay (C), sand (S), till (T), as well as sand and gravels (GS). Unassigned records (N) account for approximately 8% of the total thickness indicating most data has been captured.

Table 2.7 **Material classification overview. Material grouping code based on the Unified Soil Classification System. Hydrogeologic units have been assigned based on a review of hydraulic data (see Section 2.3.4).**

Step 1: Descriptor Combinations	Step 2: Material Grouping Code	Step 3: Hydrogeologic Unit	Hydraulic Conductivity
Boulder, boulder gravel, boulder gravel sand, cobble gravel, cobble gravel sand	GC	Coarse	High
Boulder clay, clay gravel, clay gravel sand, gravel sand silt, gravel sand silty, gravel silt	GF	Fines	Medium
Gravel, gravel sand	GS	Coarse	High
Sand	S	Coarse	High
Clay sand, clay sand silt, clay sand silty, clay sandy, sand silt, sand silty	SF	Fines	Medium
Clay silt, clay silty, silt	M	Fines	Medium
Clay	C	Clay	Low
Boulder till, clay gravel till, clay till, cobble till, gravel sand till, gravel till, gravel till, sand till, till	T	Till	Low
Bedrock, bedrock granite, bedrock sandstone, bedrock shale, granite, sandstone, sandstone shale, shale	B	Bedrock	Low
Unassigned (includes topsoil, peat, and fill)	N	Not applicable	Not applicable

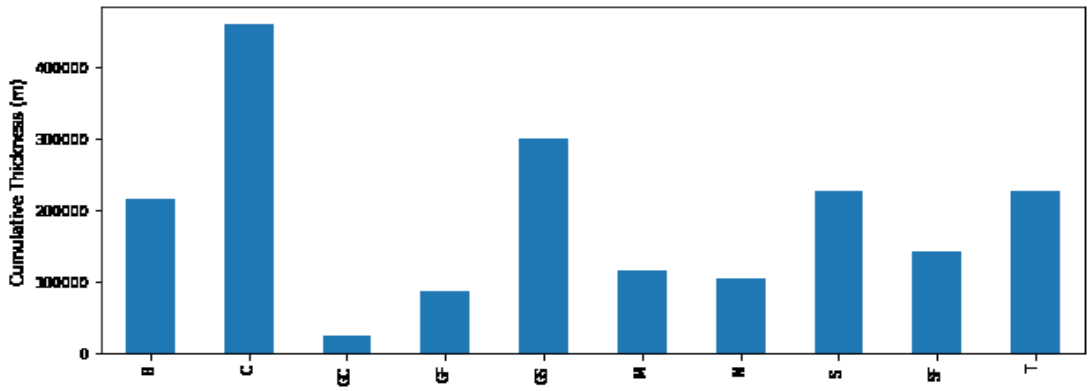


Figure 2.13 **Cumulative thickness of material groupings based on the ASTM Unified Soil Classification System. See for a description of material grouping codes.**

2.3.4. Hydrogeologic Unit Classification

Material groupings were classified into HGUs to reduce complexity for modelling purposes (Table 2.7). Hydraulic characteristics of the material groupings were used to inform the assignment of HGUs. Boxplots showing hydraulic conductivity (logK) and yields for material groupings are shown in Figure 2.14 and Figure 2.15, respectively.

LogK data is based on information from the BCBH database. Hydraulic conductivity (K) data is mostly obtained from shallow slug tests conducted during contaminated sites investigations rather than from pumping tests of aquifers used in groundwater supply. Although the data is biased towards water bearing materials, testing of lower permeability units to identify natural confining barriers is also included. For Figure 2.14, limited logK information is available for material groups GC and GF. There are fewer than five logK values for B and T; however, values appear to be consistent and within a narrow range. LogK values are generally lowest for B, C, and T groupings. The boxplot for M is like C but the range in values is more comparable to SF. GS has the highest logK mean with values varying over five orders of magnitude. The boxplot for S generally straddles the lower and upper boxplots of GS and SF, respectively.

Yield estimates available from water well records were reviewed to address data gaps in LogK data (e.g. limited LogK for GF and GC). Although numerical values are provided for yield, they were considered qualitative since they are typically cursory estimates made by the driller following completion of the well. Yield values presented in Figure 2.15 show a distinction between higher (GC, GS, and S) and lower (B, C, GF, M, and SF) ranges for material groupings. The interpretation of yield values for T are limited given the large variability and limited number of samples.

The following five HGUs were assigned to material groupings based on a review of logK and yield as shown in Table 2.7: coarse-grained materials (Coarse), fine-grained materials (Fine), Clay, Till, and Bedrock (capital letters used to distinguish HGU from primary component). GC and GF were classified as Coarse and Fine units, respectively, based on the yield review. Although available logK and yield values suggest similarities for Clay and Fines (C, M, SF), they were not lumped together to allow further spatial review during geologic model construction.

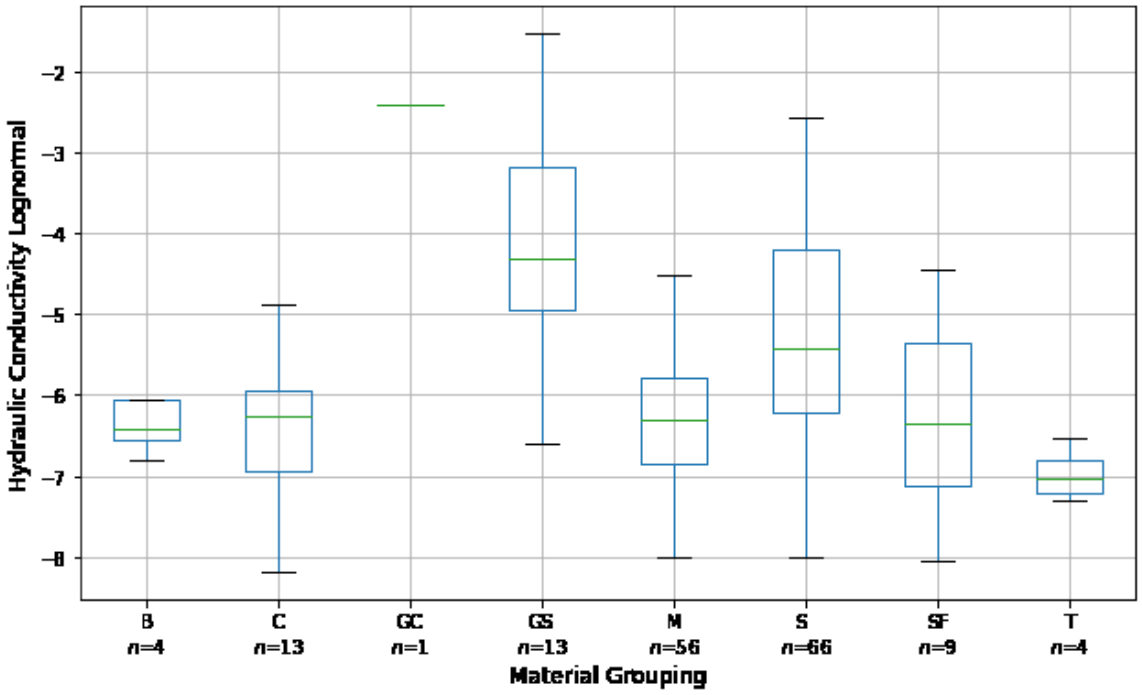


Figure 2.14 **Boxplots of Hydraulic Conductivity (Log K m/s). Refer to Table 2.7 for material group abbreviations.**

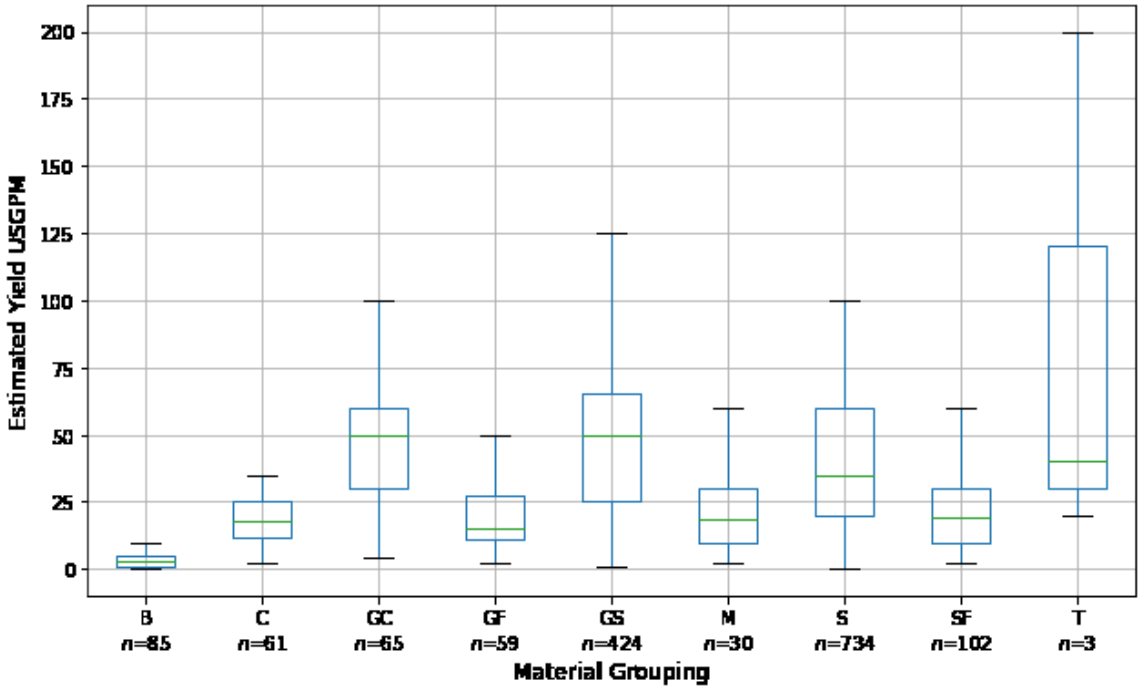


Figure 2.15 **Boxplots of Estimated Yield (USGPM). Refer to Table 2.7 for material group abbreviations.**

2.4. Mesh Generation

A mesh is required to facilitate processing of HGUs and to visualize results. The mesh includes cells that represent small volumes of the subsurface that are arranged in XYZ grid system. Attributes can be assigned to each cell, including both categorical (e.g. HGU, colour) and/or numerical data (e.g. elevation). For this study, the mesh is used to extract information and to visualize results. The coordinates of the cell centroids are used as input for predictive modelling of HGUs throughout the mesh extents.

Open source options are available to generate a mesh specific to groundwater modelling, such as GRIDGEN (Lien et al., 2015) and ModelMuse (Winston, 2019), both published by the USGS. However, Pyvista was selected because it provides coding that makes it relatively easy to create various types of mesh data structures, has various filtering methods to extract information (e.g. cell centroids, closest cell), and allows interactive visualization of results within notebooks. Pyvista provides a simplified interface to the functionality provided by Visualization Toolkit (VTK). GRIDGEN and ModelMuse do have the added functionality of mesh refinement which does not appear to be possible in Pyvista at the time it was used for this thesis but could be done directly using VTK (e.g. adaptive mesh refinement and level of detail rendering techniques). As such, mesh refinement is not considered as part of this study but may be of interest in future research to explore geology near areas of interest (e.g. along surface water features).

This section describes the information used to generate the mesh for the study area including development of a DEM to represent the ground surface for the top of the mesh, rationale for the mesh bottom, justification for the vertical spacing of each cell, as well as a summary of mesh details.

2.4.1. Top Elevation

A digital elevation model (DEM) was generated for the study area to represent the ground surface for the top of the mesh. A combination of topographic and bathymetric data from several publicly available data sources (Table 2.8) was processed using ArcMap®. Bathymetric data was required since topographic data sources apply a flatting method to assign a flat surface elevation to water bodies (DMTI, 2002).

Therefore, bathymetric data was used to provide better refinement of the bottom elevation of drainage features such as the Fraser River. A DEM for the Lower Mainland that combines more recent bathymetric data collected along the lower Fraser River and topographic data has been prepared as part of Phase 2 Flood Strategy Projects by the Fraser Basin Council (https://www.fraserbasin.bc.ca/Phase_2_Projects.html#dem) but was not available for public use at the time of preparing this thesis.

Table 2.8 Data sources for digital elevation model.

Agency	Application/ Product	Scale
Canadian Hydrographic Service	Non-Navigational (NONNA) Bathymetric Data	100 m cell spacing
DMTI Spatial Inc (1)	DMTI DEM	30 m cell spacing
U.S. Geological Survey	National Elevation Dataset (NED)	30 m cell spacing
Finlayson et al. (2000)	Puget Sound DEM	30 m cell spacing

(1) Accessed through SFU Library Data Services Dataverse

The Puget Sound DEM required conversion of the ASCII grid into raster format (Finalyson et al., 2000) and elevation units from integer decimeters (1/10 m) to metres. The dataset extents were adjusted as shown in Figure 2.16. The Canadian Hydrographic Service (CHS) Non-Navigational (NONNA) bathymetric data includes coverage along portions of the Fraser River and Georgia Strait with additional coverage supplemented by the Puget Sound DEM along the American coastline. The topographic information is based on the DMTI DEM for the Canadian portion and USGS National Elevation Dataset (NED) for the American portion of the study area.

All datasets were re-projected to NAD 1983 UTM Zone 10 with elevation units in metres above sea level and then combined using a cell spacing of 30 m and average functions with automatic filling of elevation voids (DEM30). DEM30 was then resampled to generate a DEM with 200 m cell spacing (DEM200).

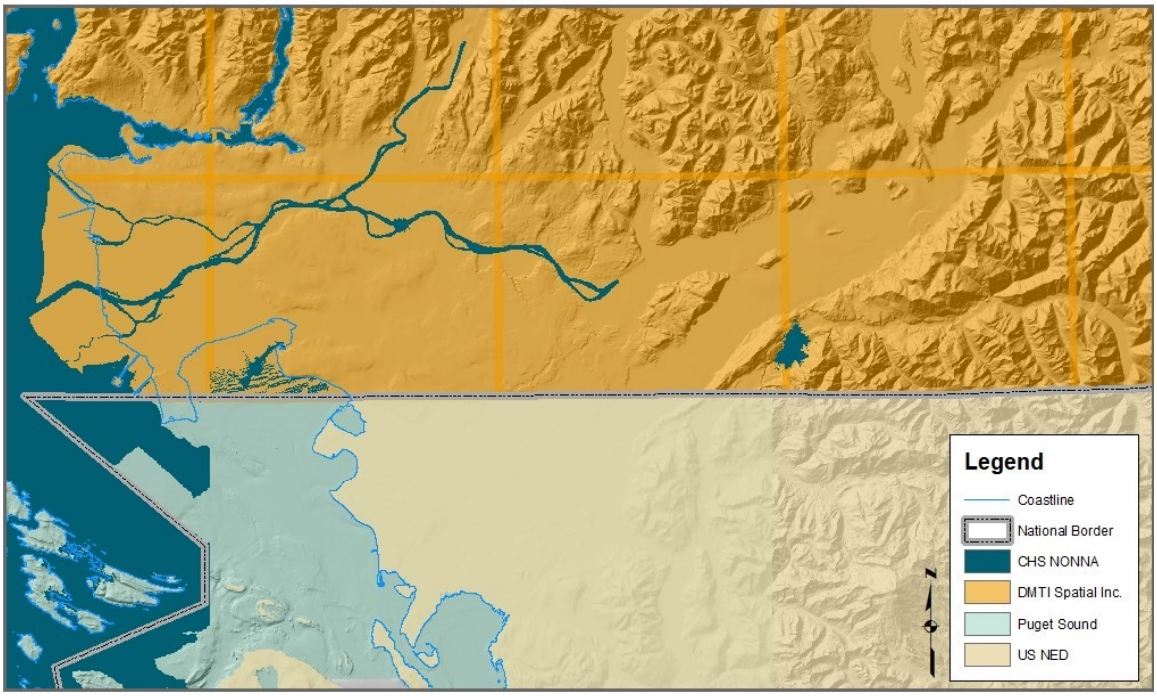


Figure 2.16 Spatial extent of data sources used for digital elevation model. Refer to Table 2.8 for data source references.

Sinks in the DEM200 were reviewed to evaluate the quality of DEM200. A sink is a cell or set of spatially connected cells whose surface flow direction cannot be assigned due to neighbouring cells with higher elevations or when two cells flow into each other (ArcGIS Help). Sinks in elevation data are commonly due to errors caused by sampling effects and the rounding of integer numbers. However, naturally occurring sinks in elevation can occur; therefore, an understanding of the morphology of the area is required to distinguish between sinks that may be naturally occurring and sinks that are errors in the data.

DEM200 has over 600 sinks with a total of 270 sinks occurring within the geologic model boundary as shown in Figure 2.17. Most sinks were either limited in spatial extent and/or less than 0.2 m deep. The most significant sink occurs within the Sumas Valley, south of Sumas Mountain (Sumas Sink). The Sumas Sink has an area of approximately 115 km^2 and a depth of 3.6 m. The historical occurrence of Sumas Lake suggests a sink could be naturally occurring in Sumas Valley and it is a significant topographic low. For this study, sinks have not been removed but may need additional consideration if surface flows are considered in the future.

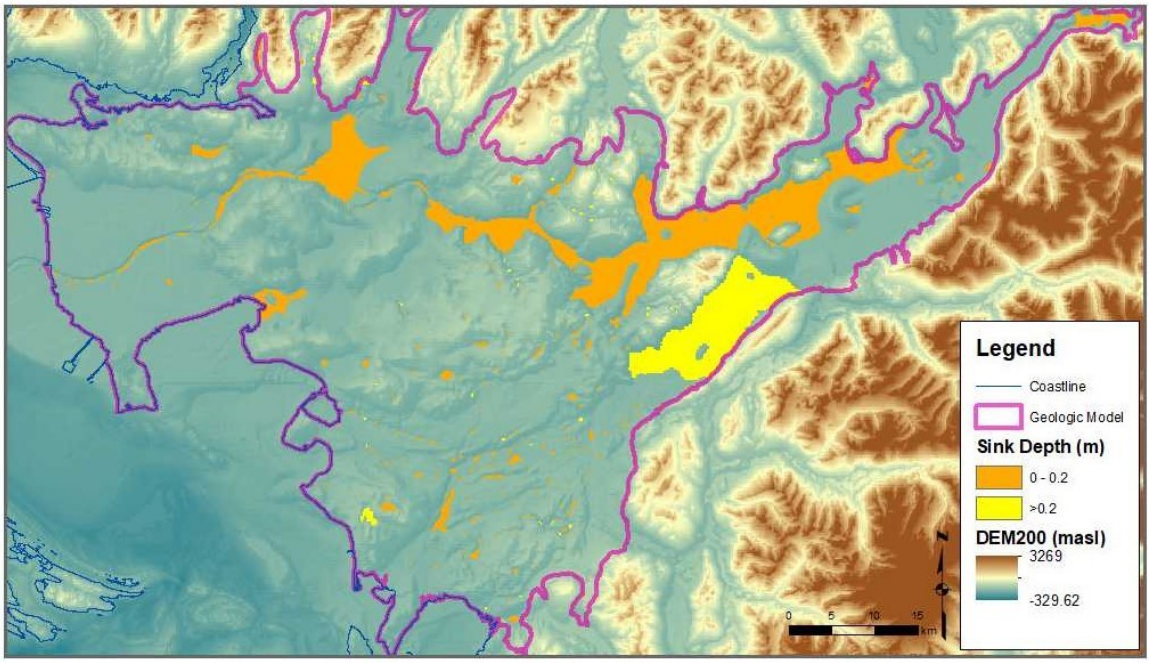


Figure 2.17 Digital elevation model (DEM200) sink locations.

2.4.2. Bottom Elevation

The use of a bedrock surface was initially planned as the bottom boundary of geologic model. The basement geology of the region has been described by Monger and Journeay (1994). Several bedrock surfaces have been prepared within the study area. Hamilton and Ricketts (1994) prepared bedrock contours for the Georgia Basin and Fraser Lowland using marine seismic surveys, bedrock outcrops and sparse data from borehole logs. The onshore portion of the map was intended as a generalized representation of the variable relief of the bedrock surface given data limitations.

Scibek (2005) used well data and bedrock outcrops in Washington State to expand bedrock contouring by Hamilton and Ricketts (1994). Valley wall profiles and extrapolated cross-sections were also used to refine bedrock depth estimates within the narrow Sumas Valley. Bedrock was extrapolated downward from the valley walls and shaped parabolically at the inferred valley bottom. The parabolic shape is characteristic of valleys modified by glacial erosion. High angle, normal faults along both margins of the valley that extend to the northeast (Sumas and Vedder faults) have also modified the bedrock surface in this area. Movement of the faults has caused a graben structure several hundred metres deep between Sumas and Vedder Mountains (Mustard and Rouse,1994).

A bedrock surface for northwest Washington State based on bedrock elevations from geologic maps, low confidence geologic cross-sections, and borehole logs was published by the USGS (Eungard, 2014). No bedrock elevations from the Georgia Basin or Fraser Lowland were used; as such, there are significant discrepancies (over 400 m) between bedrock contouring noted above for the Fraser Lowland, particularly along the international boundary.

The BC and Washington State oil and gas wells drilled in the 1960s include some of the deepest subsurface information, which has been used to interpret the bedrock surface by others (Ricketts et al., 1994; Scibek 2005), particularly where thick Quaternary deposits exist. However, the majority of oil and gas wells are limited in spatial extent, and available geophysical logging provides limited interpretations of bedrock units. Bedrock in the study area is generally correlated with the base of Pleistocene-aged deposits but this boundary is difficult to identify due to the rhythmic interbedding of strata and the general absence of horizons with distinct geophysical log signatures (Bustin, 1990). Although rare, there is the possibility of rafted bedrock interbedded with unconsolidated deposits that would make interpretations more complex.

Given the lack of deep subsurface information and the disproportionally large number of wells with relatively shallow depths compared to previously interpreted bedrock surfaces noted above, a different approach was used to define the base of the model. The spatial and vertical distribution of Coarse and Fine units were mapped. Coarse units at depth likely represent aquifers. The distribution of Fine units was also considered since they can act as aquifer units if surrounded by units with lower permeability. Using this approach, the bottom elevation of the model can be rationalized by accounting for the deepest lithologic units that could act as aquifers.

The spatial distributions of the deepest Coarse or Fine HGUs below -150 masl is shown in Figure 2.18. The lateral distribution of Coarse and Fine HGUs decreases with depth. Based on existing data, these HGUs are deepest in the Nicomekl-Serpentine river valleys where Aquifer No 58 has been mapped (see Figure 2.18). The bottom of the geologic model was extended to -150 masl given that limited information on Coarse and Fine HGUs is available beyond this elevation. This depth is similar to the deepest aquifer modelled by Golder in the Township of Langley area (Golder, 2005, pg 41). The

availability of subsurface data at this depth for basin-scale modelling purposes is limited outside of the Nicomekl-Serpentine river valleys which will contribute to uncertainty in geologic modelling results at depth.

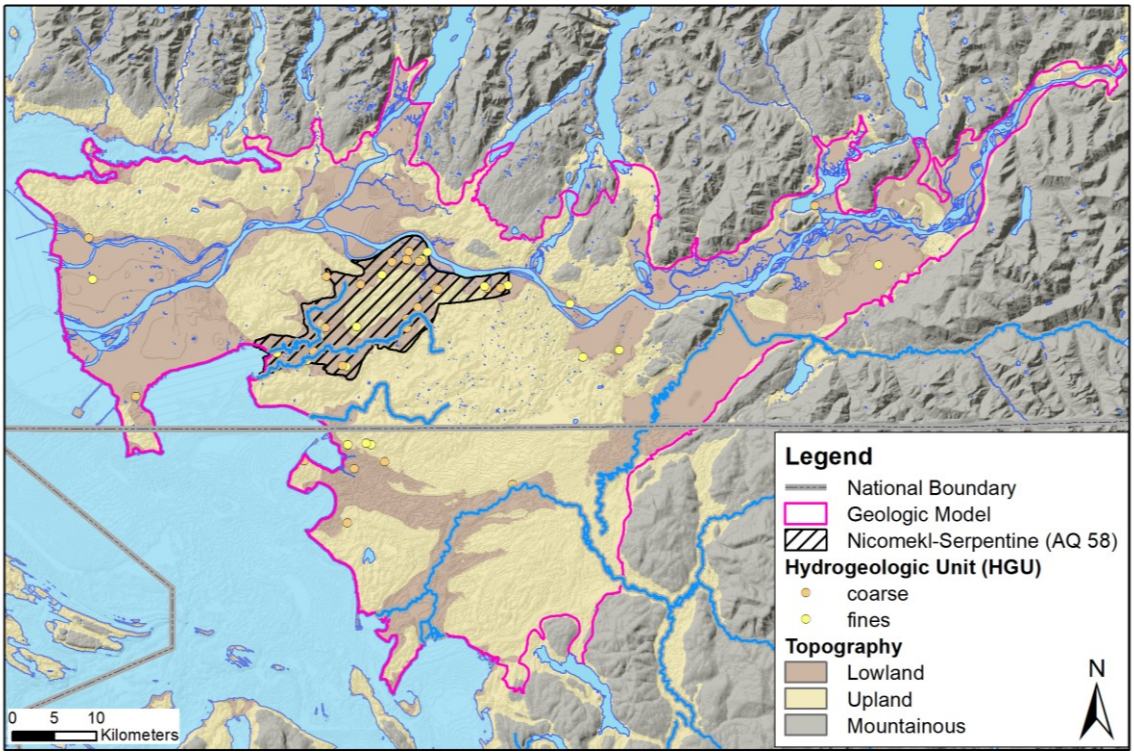


Figure 2.18 Spatial distribution of Coarse and Fine HGUs below -150 metres relative to sea level, N=46. The outline of the Nicomekl-Serpentine Aquifer (AQ 58) is shown.

2.4.3. Cell Height

The vertical cell height was based on a statistical review of HGU thickness and depth. HGUs were aggregated based on sequential groupings prior to calculating statistics. As such, statistics are based on grouping of HGUs and not on individual interval records to better represent processed data results. An example of processed lithology and the resultant grouping of HGU data is provided in Table 2.9 to illustrate the approach used.

Table 2.9 **Example of processed lithology and HGU grouping. Yellow highlights information on rows that have been grouped together since HGUs occur sequentially. The HGU interval is based on the minimum of the 'From' column and maximum of the 'To' column.**

Processed Material Description	From mbgs	To mbgs	usgs10	HGU	seq	HGU From mbgs	HGU To mbgs
sand brown	0.0	1.2	S	coarse	1	0	1.6
sand grey brown	1.2	1.6	S	coarse		MIN	MAX
clay dark grey	1.6	3.0	C	clay	2	1.6	3.0
silt sandy grey	3.0	4.6	M	fines	3	3.0	4.6
sand dark grey black	4.6	6.1	S	coarse	4	4.6	6.1

HGUs have an average thickness of at least 8 m with median values of at least 5 m as shown in Figure 2.19a. The histogram of HGU thickness shows highly skewed data biased towards shallow depths. The HGU thickness is commonly less than 5 m closer to the ground surface as shown in the 2D histogram (Figure 2.19c). The difference in thickness counts is less pronounced and fewer data are available with depth. Similarly, the most frequent information on thickness occurs between 0 to 100 masl with less frequent occurrence above and below this range. Thus, a uniform vertical cell height of 5 m was used to capture lithologic variability, recognizing the greatest generalization will likely occur near the surface. Any lenses less than 5 m thick are not captured, which was deemed acceptable given the regional scale of the study. A finer discretization or alternate approach to assigning HGUs at the surface may be warranted to better reflect recharge conditions for groundwater modelling purposes.

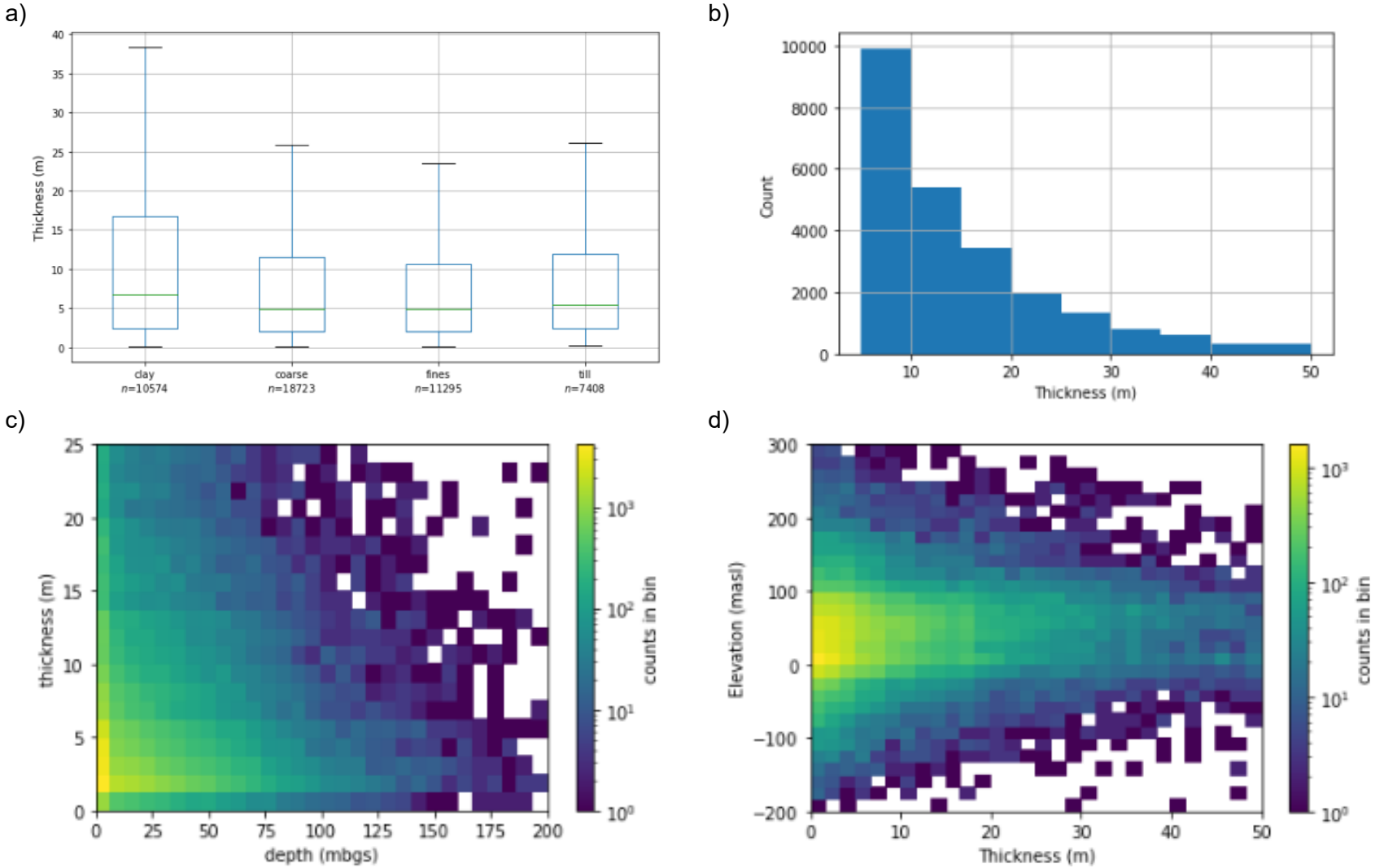


Figure 2.19 **HGU thickness shown as (a) boxplot b) histogram and c) 2D histogram of thickness relative to depth in metres below ground surface (mbgs), and d) 2D histogram of elevation relative to thickness.**

2.4.4. Mesh Extents

The mesh was generated using PyVista and python coding. The geologic model layout and extent of the mesh are shown in Figure 2.20. The length and width of the model are based on the extent of Quaternary mapping within the study area excluding the land north of the Burrard Inlet given this area is hydraulically separated from the rest of the Lower Mainland, and the intermountain valleys (e.g. Chilliwack River valley, and valleys containing upper tributaries of the Nooksack River) since information is not readily available to characterize the geologic structure in these areas. Cells outside of the lateral boundaries have been masked to generate a non-rectangular shape. Vertically, the bottom of the mesh was constrained by applying a filter to isolate cell centers with elevations greater than -150 masl. This did not result in a perfectly flat surface at the bottom of the model but is suitable for the purpose of this study.

A uniform cell size of 200 m wide and 200 m long was used. This was initially based on a visual review of the spatial resolution needed to capture the variation of mapped surficial geology units (Section 2.2.4). This lateral cell size would allow representation of Holocene deposits that thin along sections of the Fraser River within the study area. However, this level of detail is not expected from the model given that surficial geology points are spaced every 1000 m so that the geologic database is not overrepresented by surficial data. A uniform cell height of 5 m was applied based on the mean thickness of HGU units (Section 2.4.3).

Mesh details are summarized in Table 2.10. The grid is bounded along the X-axis between 481026 to 613026 m, the Y-axis between 5396343 to 5471743 m, and the Z-axis between 903 to -150 masl. This results in over three million cells covering a surface area of approximately 670 km^2.

Table 2.10 Mesh properties.

Property	Value
Number of Cells	3,336,175
X Bounds	481026, 613026
Y Bounds	5396343,5471743
Z Bounds	-150, 903
Surface Area	670 km^2

Notes: X and Y bounds are in UTM NAD 83 Zone 10 coordinate system.

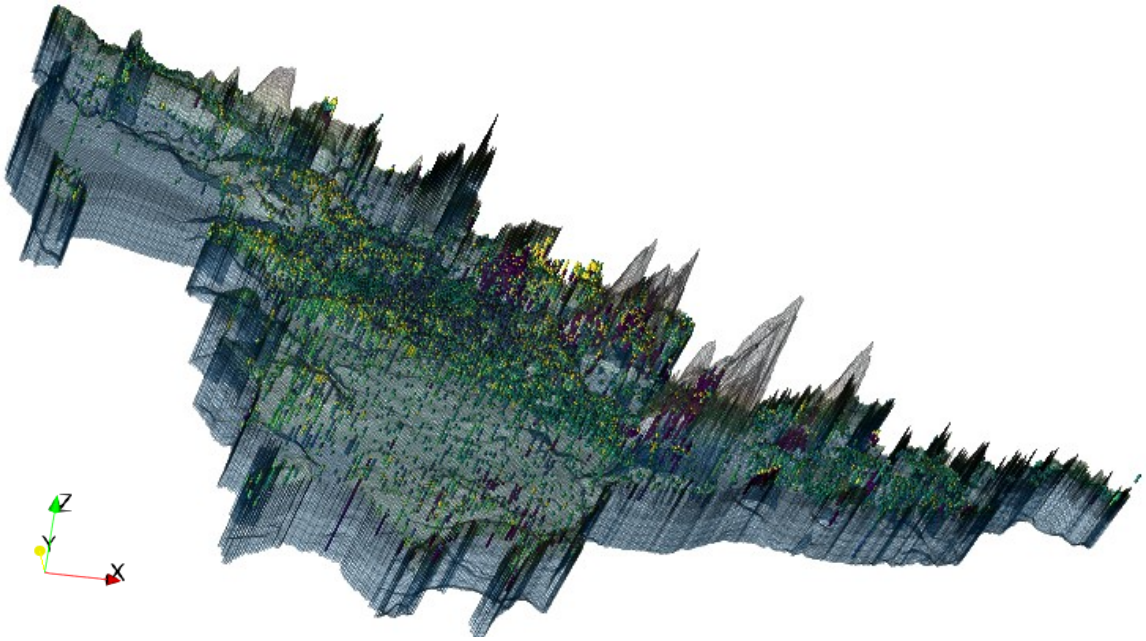


Figure 2.20 Mesh layout and extents.

2.5. Data Selection

2.5.1. Alternatives

Input features and targets are required for MLP training. The selection of inputs and targets for training purposes is an important task since MLPs learn from data with no prior assumption on data distribution. Three alternative methods are used for data selection to evaluate the impact on the training and prediction capabilities of MLP. Table 2.11 provides an overview of each data selection alternative, including the approach for input features and targets as well as the resultant number of samples. Alternative 1 has the lowest number of samples.

Table 2.11 Overview of data selection alternatives.

	Input Features	Target	Number of Samples
Alternative 1	Easting, northing, and elevation based on the borehole location and top elevation of each HGU interval.	HGU	54,225
Alternative 2	Easting, northing, and elevation based on the cell centroid	HGU mode	85,063
Alternative 3	X and Y coordinates of 2D mapping from SOM	HGU mode	85,063

Alternative 1 involves minimal processing of available data. The coordinates of the borehole and top elevation of the HGU interval (e.g. three features) are used as input while the HGU is used as the target for training purposes.

Alternative 2 uses the coordinates of the cell centroid (e.g. three features) and generalizes HGUs based on occurrence frequency (e.g. mode) within each cell. To determine the HGU mode, a point cloud was generated, spacing HGU data vertically every metre for each interval. The most frequently occurring HGU within each cell was then used as the target for training. This may potentially reduce noise in the data and change the performance of the MLP.

Like Alternative 2, Alternative 3 uses the HGU mode as a target for training the MLP but the coordinates of the cell centroids are spatially clustered using a self-organizing map (SOM). SOM maps the cell centroids onto the nodes of a 2D grid based on distance (e.g. Euclidian). The coordinates of the nodes (e.g. two features) are then used as input features for the MLP. This approach was considered to determine if spatial clustering enhances pattern recognition by MLP.

A conceptualization of the neural network architecture is shown in Figure 2.21 to highlight input differences between Alternative 1 and 2 compared to Alternative 3. Further SOM details are provided in Section 2.5.2.

2.5.2. Self-Organizing Map

Alternative 3 uses a hybrid ANN that combines SOM with MLP (Figure 2.21), which has been identified by others to enhance pattern recognition in hydrogeologic applications (Rizzo and Doughtery, 1994; Sahoo and Jha, 2017). The SOM is trained using two stages as recommended by Bacao et al. (2005). For Stage 1, the elevation of each cell centroid is set to zero with only the easting and northing coordinates used for mapping. For Stage 2, cell centroids coordinates (northing, easting, elevation) are then used to refine mapping of the samples onto the 2D grid. SOM is an unsupervised neural network algorithm; therefore, testing is not required. Figure 2.22 shows an overview of the method used to prepare the SOM with details provided in subsequent sections for data preparation, hyperparameter selection, and training of the SOM. The SOM algorithm was implemented using Minisom (v. 2.3.0).

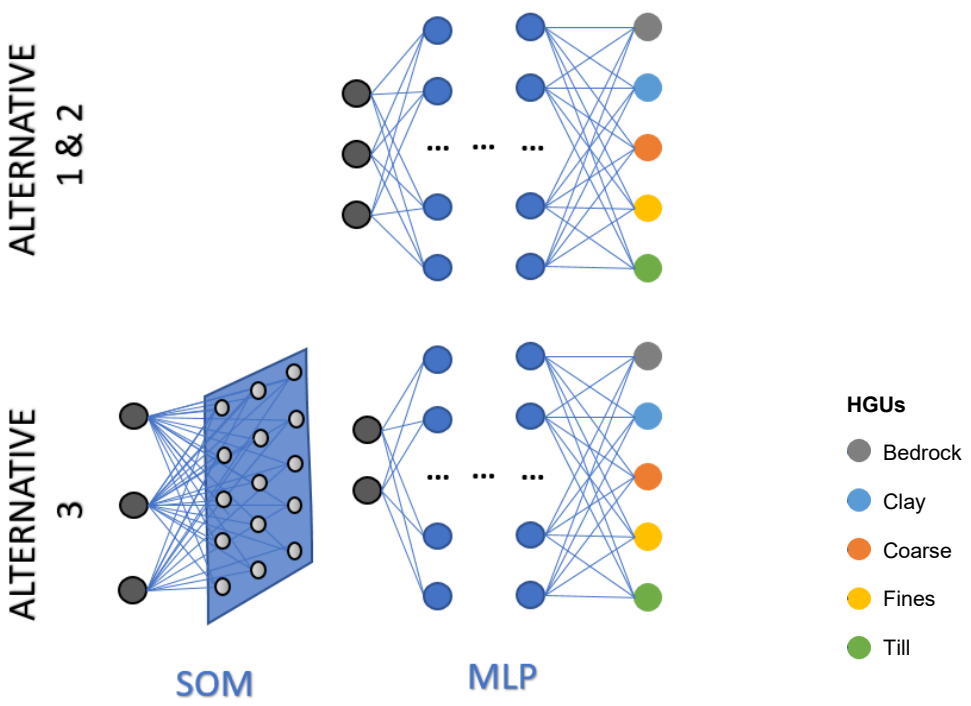


Figure 2.21 Neural network conceptualization for data selection alternatives.

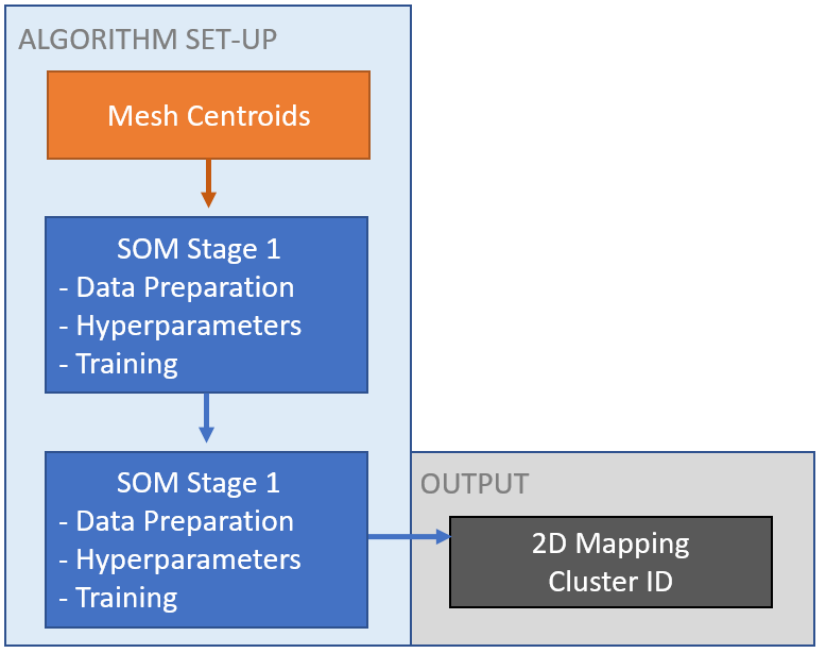


Figure 2.22 Overview of SOM methodology.

Data Preparation

The coordinates of the mesh centroids were used as input for the SOM. However, modifications to the dataset were necessary to lower memory requirements so that performance metrics (e.g. quantization error) could be calculated using available computer resources. This includes removing cell centroids below -100 masl and reducing data precision (e.g. float16 instead of float35).

Normalization was used to scale the coordinates of the cell centroids. For the first stage, the elevation of each cell centroid was set to zero; therefore, only the normalized eastings and northings were used to initially train the SOM. For the second stage, the normalized northing, easting, and elevation of the cell centroids were used to refine mapping of the samples onto the 2D grid.

Hyperparameters

The assignment approach for hyperparameters of the SOM algorithm are summarized in Table 2.12. The majority of hyperparameters were assigned using heuristics or default values.

The same initial weights were used during the development of each stage of the SOM. For Stage 1, initial weights were randomly generated using Minisom's random weight function. The weights from the trained Stage 1 SOM were then applied as initial weights for the Stage 2 SOM and used during Stage 2 hyperparameter optimization. This is important given that the performance of the SOM algorithm is dependent on the initial weights of the map (Vettigli, 2018).

Hyperopt (Bergstra et al., 2013) was used for hyperparameter optimization since the grid functionality provided in scikit-learn could not be directly applied. Hyperparameters selected for optimization include sigma and learning rate. Various sigma and learning rate values were considered to determine the best combination that would result in the lowest error.

The SOM was then configured with the optimized values for sigma and initial learning rate. To understand how training evolves and to establish the final number of iterations, the quantization error was reviewed for several iteration sizes. This was done to avoid overfitting the model.

Table 2.12 SOM Hyperparameter assignment.

Hyperparameter	Assignment Approach
x, y	A general rule of thumb for the dimension of a grid is that it should contain 5*sqrt(N) neurons where N is the number of samples in the dataset to analyse (minisom.py). Based on the number of cell centroids (3,336,175), the recommended number of nodes is 9,132. Therefore, the grid dimensions were set at x = 100 and y = 90 which results in 9,000 nodes.
input_len	The number of features in the input data. Three was used given the use of northing, easting, and elevation as features.
sigma	Sigma is the initial spread of the neighborhood function that is dependent on the dimensions of the map. A larger sigma value was used for Stage 1 given suggestions by Maimon and Rokach (2010, p435) to use an initial sigma value large enough to cover half of the nodes. The sigma value was lowered in Stage 2 to ensure a smaller spread of the neighbourhood function to support refinement of mapping of samples to the SOM grid.
learning_rate	The initial learning rate can be between 0 and 1. An initial learning rate of 1.0 was used for Stage 1 while a lower value of 0.5 (default) was used for Stage 2.
decay_function	Function that reduces the learning rate and sigma after each iteration. The default was used.
neighborhood_function	The default option 'Gaussian' was used.
topology	Possible values include rectangular or hexagonal. The default option 'rectangular' was used.
activation_distance	The default option 'Euclidean' was used.
random_seed	A value of 17 was arbitrarily selected for reproducibility.
initial_weights	Randomly generated for Stage 1. The initial weights for Stage 1 were saved and used during optimization and training of the SOM. The final weights from Stage 1 were used as initial weights for Stage 2. The precision of the weights was converted to float16 to reduce computer memory requirements to calculate performance metrics.
random_order	True was used to allow random sample selection.

Note: Hyperparameters highlighted in blue were considered as part of optimization.

Training

Once hyperparameters were selected for each stage of the SOM, training was completed using scaled data. Training establishes the weights between the nodes of the grid. Once complete, the coordinates of the winning neuron can be determined for each cell centroid.

SOM results were reviewed by visualizing weights (e.g. distance map or U-matrix), HGU assignment, and neuron activation frequencies. A distance map is used to display neurons as an array of cells coloured to represent the weights (based on Euclidian distance) between neighbouring neurons. Smaller weights indicate neighbouring neurons are similar while larger weights suggest greater differences. The

HGUs of samples with information are plotted on the distance map to show their distribution. Pie charts are also used to provide a better visualization of the frequency of HGUs mapped to each neuron. Lastly, a neuron activation map is plotted to show how often each neuron is chosen as the winning neuron.

Performance

For SOM to perform well, it must preserve the topology and neighborhoods of the input data. SOM performance metrics chosen for this study include quantization and topographic error (Forest et al., 2020). Quantization error (QE) is a clustering metric calculated based on the average distance between the data points and the map nodes to which they are mapped, with smaller values indicating a better fit. This is a basic quality measure that allows comparison between maps as part of training but is not a standalone assessment of quality.

Topographic error (TE) is a topographic metric that measures how well the structure of the input space is modelled by the map. TE is calculated by finding the best-matching and second-best-matching neuron on the map for each input and then evaluating positions. An error occurs if the input is not next to the best-matching neurons. The total number of errors divided by the total number of data points gives the topographic error of the map. A lower TE indicates a better performance of the SOM algorithm at preserving the topological features of the input space in the 2D grid.

2.6. Multi-Layer Perceptron

As described in Section 1.2.2, MLP is commonly used for classification because it can discover patterns in large or complex data sets. MLP is used for multi-class classification of HGUs at the cell centroids of the mesh to build 3D geologic realizations of the study area. The MLP algorithm was implemented using Scikit-Learn (v.0.24.2).

The MLP must be trained before it can be used for predictions. Training the MLP requires data preparation (splitting, scaling), selection of hyperparameters, and assignment of weights based on training and testing. After training, the coordinates of the cell centroids are scaled and passed through the MLP classifier to predict HGUs. The output includes predicted HGUs and probabilities throughout the study area that can be assigned as mesh attributes and then visualized in 3D. An overview of the method

used to prepare the MLP classifier is shown in Figure 2.23 with further details provided in subsequent sections.

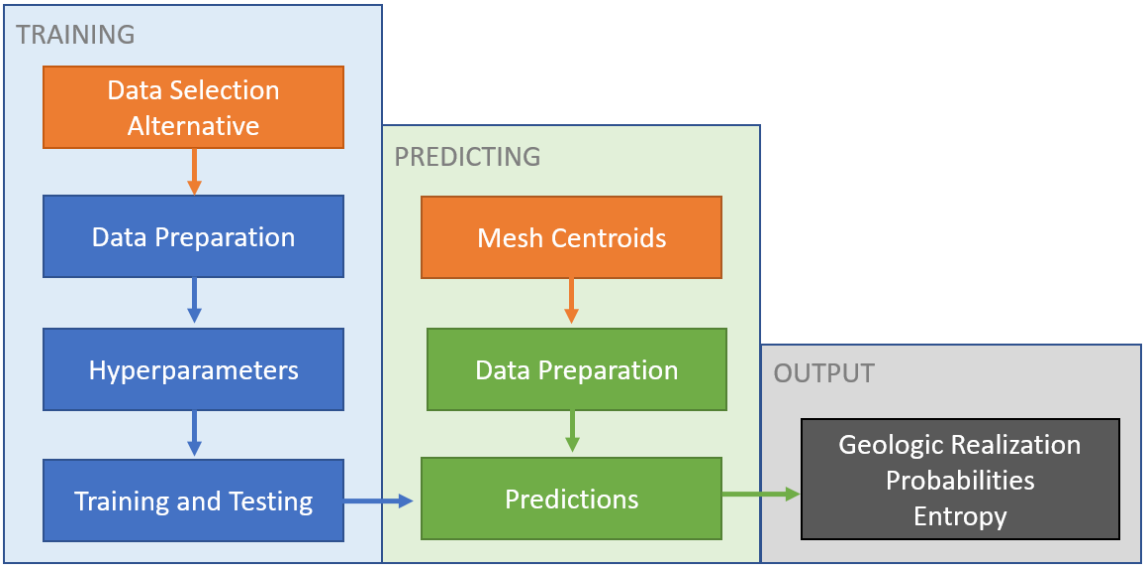


Figure 2.23 Overview of the MLP configuration used to investigate the impact of data selection alternatives on training and prediction performance.

2.6.1. Data Preparation

The first step in developing the MLP is data preparation (Figure 2.23). Data preparation includes splitting the data into training and testing subsets. Both subsets then need to be scaled to handle varying magnitudes.

The data from each alternative was split into training and testing subsets using a 80% and 20% split, respectively (Brownlee, 2020c). A stratified splitting approach was used to ensure each set contains approximately the same percentage of samples of each class (e.g., HGU) as the complete set. This means that the class ratios are preserved when splitting data. Datasets with a disproportionate ratio of observations in each class (imbalanced) can be problematic for MLP. The relative proportions of HGUs were reviewed for each alternative (Chapter 3) and performance metrics were selected to address this potential issue.

Data must be in a numerical format and scaled prior to being used in MLP. With MLP, categorical targets (HGUs) are automatically transformed to one-hot vectors. Feature scaling methods considered for each alternative include normalization

(transformed to range between 0 and 1) and standardization (transformed to have a zero mean and standard deviation of 1). The scaling method that resulted in the best MLP performance was used.

2.6.2. Hyperparameters

The selection of hyperparameters follows data preparation (Figure 2.23). Table 2.13 provides a summary of hyperparameters for the MLP algorithm. Four hyperparameters (hidden layer size, alpha, batch size, and initial learning rate) were selected for optimization. These hyperparameters were chosen because they appeared to have the greatest impact on MLP performance when using the ReLU activation function and Adam solver.

Table 2.13 MLP hyperparameter assignment.

Hyperparameter	Assignment Approach
hidden_layer_sizes	An adequate solution can typically be obtained with a network size that has more than three layers (Kanevski et al. 2001). Three hidden layers each having 100 neurons was initially used; resulting in a five layer network architecture.
activation	The default Rectified Linear Units (ReLU) activation function for hidden layers was used. It is one of the most popular activations functions, is non-linear, and is relativley fast for larger neural networks.
solver	Default solver Adam used for weight optimization. Softmax applied on the last layer by default.
alpha	Default value of 0.0001. Alpha can be increased/decreased to correct overfitting/underfitting respectively.
batch_size	Default value of 200.
learning_rate_init	Default value of 0.001.
max_iter	Default value of 200.
shuffle	Default True.
random_state	Seven (7) arbitrarily selected for reproducibility.
tol	Default value of 0.0001.
warm_start	Default False
early_stopping	Used when optimizing hyperparameters.
validation_fraction	Used when optimizing hyperparameters. Default value of 0.1
beta_1	Default value of 0.9.
beta_2	Default value of 0.999.
epsilon	Default value of 1e-8.
n_iter_no_change	Default value of 10.

Note: Hyperparameters highlighted in blue were considered as part of optimization.

The GridSearchCV optimizer in scikit-learn was used for hyperparameter optimization. GridSearchCV evaluates different combinations of hyperparameters using cross-validation to objectively choose a subset that achieves the best performance on the training data (Figure 2.24). Early stopping was applied to terminate training when validation scores stop improving.

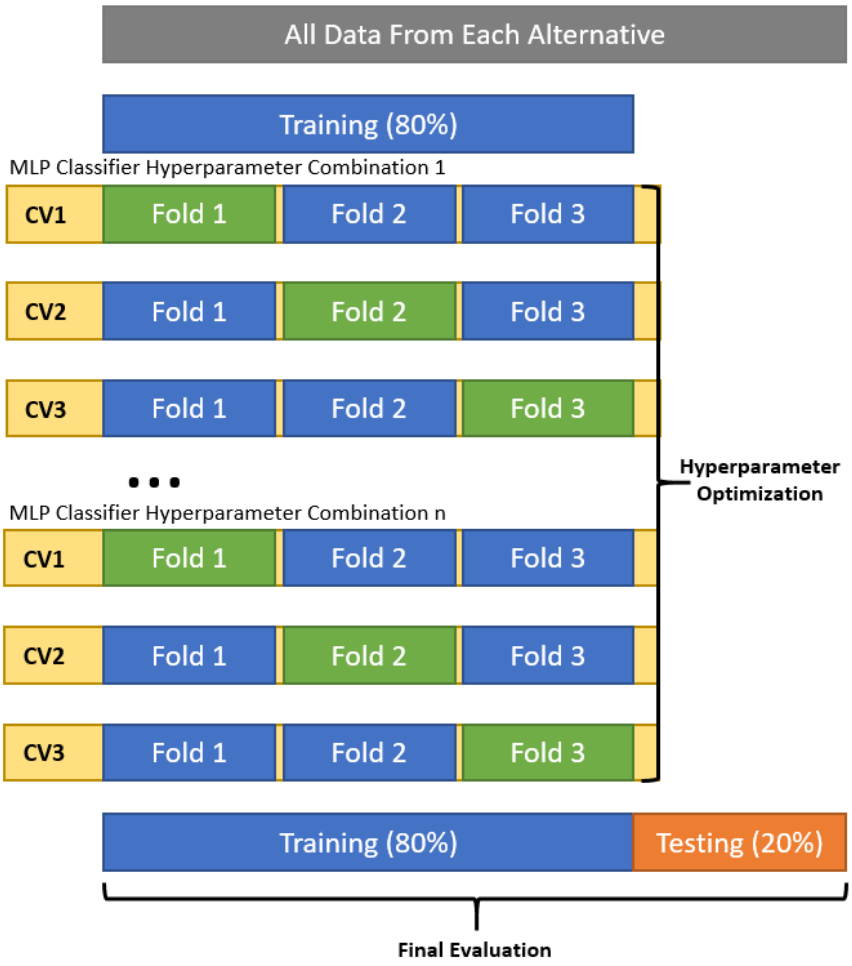


Figure 2.24 **Data splitting approach used for hyperparameter optimization. The data for each alterantive was split into training and testing subsets using a 80% / 20% split. Three cross-validation (CV) runs (CV1, CV2, CV3) were conducted on the training subset as part of hyperparameter optimization. For each CV, training was conducted using two folds (blue) and validated using one fold (green). The hyperparameter combination of the best performing model was then used for training and testing of the final model.**

Cross-validation involves dividing the training data into subsets (folds). One-fold is withheld from training and then used to evaluate the performance of model. This approach provides an indication of the models' ability to make predictions on unseen data and results in less biased or more conservative estimates of model generalization compared to other methods (e.g. simple train/test split) (Brownlee, 2020b). For this study, each combination of hyperparameters was evaluated using three-fold cross-validation. The same stratified approach for splitting data into training and testing subsets was used for splitting the training data into folds to ensure the relative proportions of HGUs were maintained (Figure 2.25).

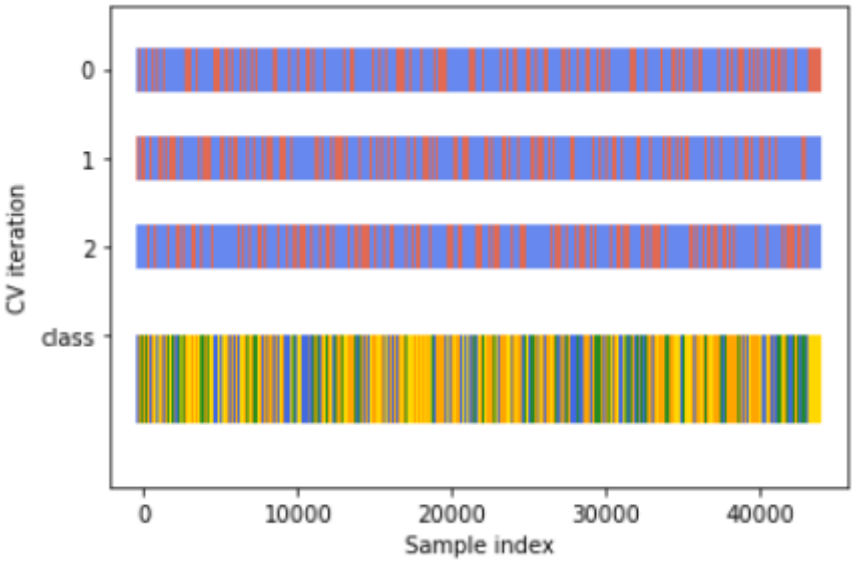


Figure 2.25 Stratified three-fold cross-validation example using Alternative 1. The red lines along the top blue bars represent the index value use to split the training data into three folds. The HGU classification for each index value is shown along the bottom bar (bedrock – grey, clay – blue, coarse – oranage, fines – yellow, till – green).

The combination of hyperparameters with the highest average validation score was selected for the training model. The same stratified three-fold cross-validation approach was then used on the training model with the optimized hyperparameters to determine the optimal number of iterations.

2.6.3. Training and Testing

Once the hyperparameters and iterations have been selected, the development of the training model proceeds with training and testing (Figure 2.23) as part of the final

evaluation. The MLP is trained using all of the training data. Training establishes the weights between the nodes of the neural network by optimizing the log-loss function using stochastic gradient descent. The testing data is then used to evaluate the generalization performance of the model when making predictions on unseen data.

2.6.4. Predictions

Once the training model has been validated and tested, all the data is used to train and establish the final weights of the predictive model. Further evaluation of the predictive model is not required since the generalization capability was determined during the final evaluation of the training model (Section 2.6.3).

The main purpose of the MLP is to take existing geologic data and make HGU predictions in the study area where information is not available. The coordinates of the cell centroids for the entire mesh are used as input into the predictive model. In the case of Alternative 3, the coordinates of the winning neuron mapped for each cell centroid using SOM are used as input instead.

The same data preparation methods are used to transform inputs prior to running the predictive model. Because a fixed random state is used in the predictive model to ensure reproducibility, each run provides one geologic realization of the HGUs at the cell centroids. Alternatively, the predictive model could be trained without a fixed random state which would result in different weights each time training is done and different predictive outcomes for each trained model (Brownlee, 2019b). A random process is used to initialize the weights, but this was controlled in this study by fixing a random state. Random state was not tuned as part of developing the predictive model.

The output from the predictive model includes HGU and probability predictions at each cell centroid. The output was used to calculate the probability of low permeable material and log-loss (e.g. entropy) at each cell. The probability of till and clay was added to calculate the probability of low permeable material. Areas with a high probability of low permeable material could act as aquitards and provide protection from surface contamination. The log-loss (e.g. entropy) was used to spatially quantify the uncertainty of the geologic realization. It was calculated by taking the negative log of

probability (between 0 and 1) for the predicted HGU in each cell. A lower entropy indicates less uncertainty while higher values suggest more uncertainty.

The model outputs were assigned as attributes to the mesh using PyVista and then exported as a voxel model for more advanced 3D viewing using Paraview. Select model outputs (e.g. HGU predictions, coarse HGU probabilities, low permeability probabilities, and entropy) were visualized using ParaView to compare the geologic realization from each alternative. Outputs were visualized in 3D cross-sectional view roughly positioned near the center of Aquifer 58, in the Nicomekl-Serpentine river valley (Figure 2.18).

2.6.5. Performance Metrics

The metrics used for MLP were selected with consideration of the potential for imbalanced classes. They include log-loss, balanced accuracy, and confusion matrix.

Log-loss (also called cross-entropy) is the loss function used by MLP to optimize results, meaning that the weights during training are adjusted with the objective of minimizing log-loss. To solve multiclass classification problems, the MLP algorithm predicts the probability of each HGU for every sample. The prediction probabilities and a binary representation of the target HGU are used to calculate entropy. During training, the weights are adjusted and the prediction probabilities change with the objective of minimizing log-loss. Log-loss quantifies how good or bad the prediction results are based on how far the predictions are from the actual values. An entropy value of 0 indicates a model with perfect skill. Higher values indicates poorer performance.

Balanced accuracy is the average recall obtained on each class based on the multiclass definition used in scikit-learn (Buitinck et al., 2013). Recall (also known as sensitivity) is the fraction of relevant instances that were correctly predicted (e.g. 10 Sand HGUs were predicted out of 15 possible instances). The balance accuracy score ranges from 0 to 1 with values closer to 1 indicating good accuracy.

A confusion matrix is a common technique used to summarize the performance of a classification algorithm. It provides a heat map of the number of correct and incorrect predictions for each label. It highlights the errors being made by the MLP or what is making the MLP ‘confused’ when making predictions.

2.7. Verification

Verification is the last step in the workflow to further evaluate the performance of MLP at predicting HGUs in the study area (Figure 2.23). The geologic realization of the alternative with the best prediction results was visually compared to prediction results using k-nearest neighbours (KNN), published interpretations of the subsurface, and hydrogeologic indicators within the region (e.g. artesian conditions).

KNN is an algorithm commonly used for classification. Predictions are made based on the most common HGU among its ‘k’ nearest neighbours. The KNN algorithm was implemented using Sklearn. Default hyperparameters were used except for setting the weight function to use distance. Inputs were normalized to avoid scale bias in the predicted results.

Cross-sectional views of subsurface interpretations provided by Golder (2005) for the Township of Langley were compared to predictions using MLP. In addition, the predicted geology within the Surrey area focusing on the Serpentine-Nicomekl river valley where Aquifer 58 is mapped was also reviewed to determine if confining units are represented by the geologic realization for artesian conditions to occur. Both of these locations correspond to areas with a relatively high density of well records.

Chapter 3.

Results

3.1. Data Selection Alternatives

3.1.1. Input Features and Targets

As discussed in Section 2.5, three data selection alternatives were developed by processing data from the geology database and mesh. Alternative 1 consists of HGUs spatially represented using well coordinates and the elevation at the top of each interval from the geology database. Alternative 2 uses the HGU mode within each cell of the mesh based on point cloud data and the coordinates of the cell centroid. Similar to Alternative 2, Alternative 3 uses the HGU mode within each cell but takes the coordinates of the winning neuron from mapping each cell centroid onto a 2D grid using SOM. SOM results are provided in Section 3.1.2

Normalized histograms showing the distribution of HGUs are shown in Figure 3.1. The histogram for Alternative 3 is not shown but is similar to Alternative 2. Coarse occurs as the most frequent HGU for all alternatives which may be reflective of a higher density of data where permeable deposits exist. For Alternative 2 and 3, the distribution of bedrock and clay increases after the interval data from Alternative 1 is processed to take the most common HGU in each mesh cell. The increase in bedrock and clay is partially attributed to typically thick intervals that are then represented in 5 m increments in those alternatives. The frequency of till is relatively low compared to the other unconsolidated HGUs (clay, coarse, fines) which may be attributed to till not being recognized as till during logging (e.g. logged as clay instead of till) or it has been substantially eroded within the study area during deglaciation.

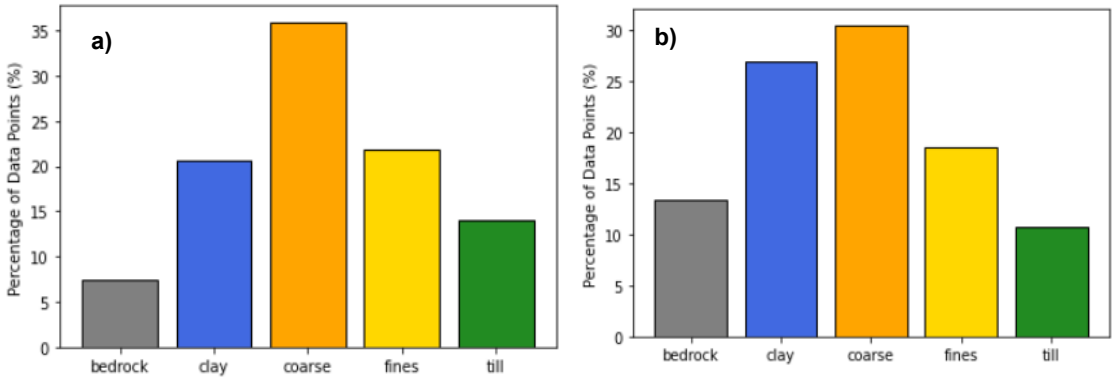


Figure 3.1 Normalized histograms showing HGU frequency for a) Alternative 1 and b) Alternative 2.

All three alternatives have imbalanced HGU distributions, meaning that the HGUs are not represented equally. Imbalance is common in classification datasets and ratios are typically used to quantify the degree of imbalance (Brownlee, 2020a). The greatest imbalance ratio is 1:5 for Alternative 1 (coarse compared to bedrock) and 1:3 for Alternative 2 and Alternative 3 (coarse compared to till). The imbalance in the HGU distribution is not considered extreme (e.g. 1:100) but may result in some bias predictive performance of the machine learning algorithm.

The spatial distribution of samples for each alternative is different (Figure 3.2). Alternative 2 has fewer lateral locations compared to Alternative 1 because cell centroids are used instead of well locations. However, there are more samples for Alternative 2 in the vertical direction since intermediate points are added to intervals greater than 5 m thick. For Alternative 3, the spatial distribution HGUs is now assigned using coordinates from 2D mapping. Data is grouped as clusters with multiple HGUs assigned to the same 2D coordinates.

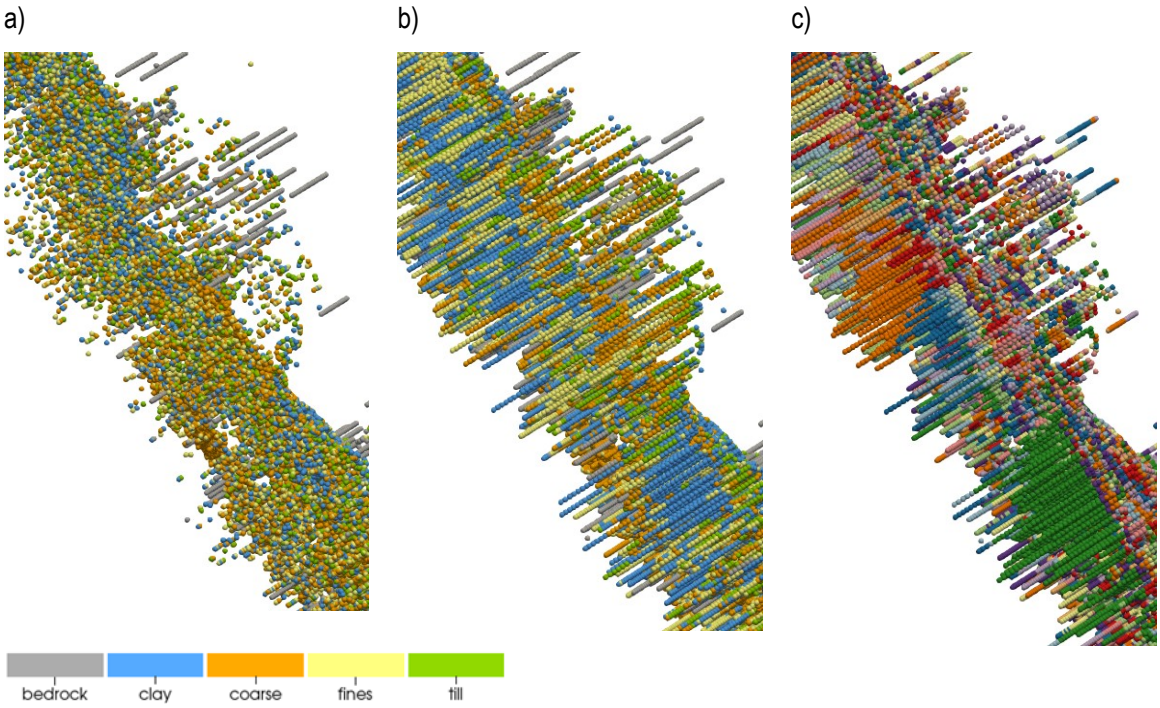


Figure 3.2 **Spatial difference in point data distribution between a) Alternative 1, b) Alternative 2, and c) Alternative 3. HGUs are shown for Alternative 1 and 2. Spatial clusters are shown for Alternative 3 to show grouping of data. Surficial geology points, including bedrock outcrops to depths of 50 m, are included for each alternative.**

3.1.2. SOM Training

As discussed in Section 2.5.2, hyperparameters selected for SOM optimization included sigma and learning rate. The search space, which represents the range of values considered, is shown in Table 3.1 for both stages. The search space for Stage 1 includes larger sigma and initial learning values compared to Stage 2. Smaller values are used in Stage 2 to limit the neighborhood distance and slow the learning rate to refine clustering from Stage 1. Optimized hyperparameters were selected based on the combination that resulted in the lowest QE. As shown in Table 3.1, the best hyperparameters include a sigma of 40 and learning rate of 0.75 for Stage 1 and a sigma of 2 and learning rate of 0.05 for Stage 2.

Table 3.1 Search space for SOM hyperparameter optimization.

Parameter	Search Space		
Stage 1			
Sigma	40	35	30
Learning Rate	1.0	0.75	0.5
Stage 2			
Sigma	10	5	2
Learning Rate	0.5	0.1	0.05

Notes: Optimized hyperparameters are highlighted in blue.

To establish the number of iterations, the QE was reviewed for iterations equal to 10,000, 15,000, 20,000, and the number of samples (85,063). For both SOM stages, the QE stopped decreasing after 15,000 iterations; therefore; the number of iterations was set at 15,000 for both SOM stages.

For each stage, the SOM was configured using the optimized hyperparameters and trained using the scaled data from Alternative 2. The QE and TE results for each stage are summarized in Table 3.2. The QE increased while the TE decreased from Stage 1 to Stage 2. TE is more representative of the performance of the SOM at preserving the topological features of the input space in the 2D grid.

Table 3.2 SOM Quantization and Topographic Error Results.

	SOM Stage 1	SOM Stage 2
Quantization Error	0.12	0.24
Topographic Error	0.14	0.06

The training results for both stages are visualized in Figure 3.3 to

Figure 3.5. Figure 3.3 shows the distance map (U-Matrix) where the neurons are represented as a grid cell and the greytone represents the weighted distance from the neighbouring neurons. If the distance is high, a darker colour is assigned to show the surrounding weights are very different. A lighter colour is assigned when the distance is low. The lighter portions typically represent clusters while the black portions represent the division between clusters.

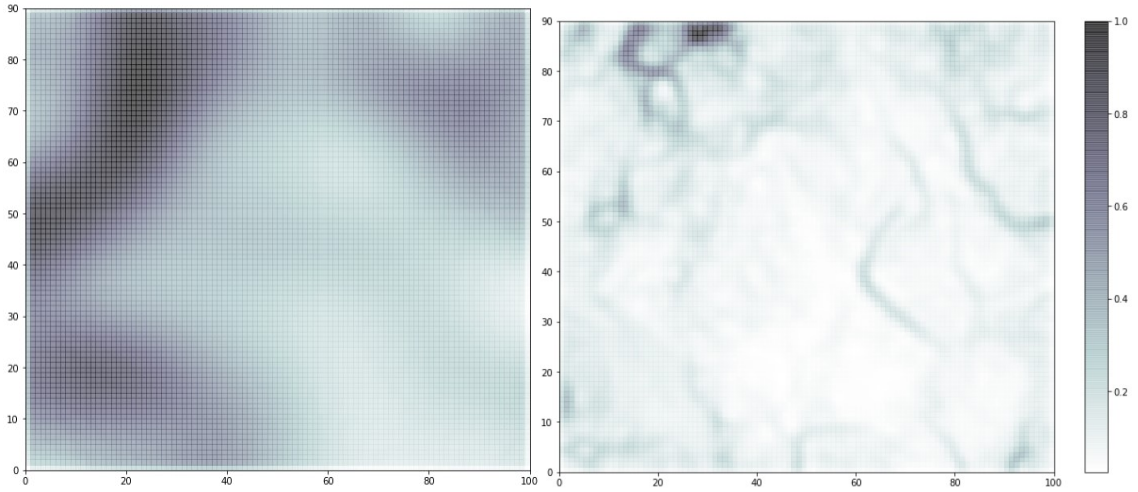


Figure 3.3 Distance map (U-matrix) for Stage 1 SOM (left) and Stage 2 SOM (right).

Figure 3.4 shows how often the neurons of the map were assigned as the winning neuron (e.g. activated). Initially, boundary effects are noted in the Stage 1 SOM based on the high number of neurons activated along the edges of the map. The final Stage 2 SOM shows a smoother distribution of activation throughout the map and less boundary effects.

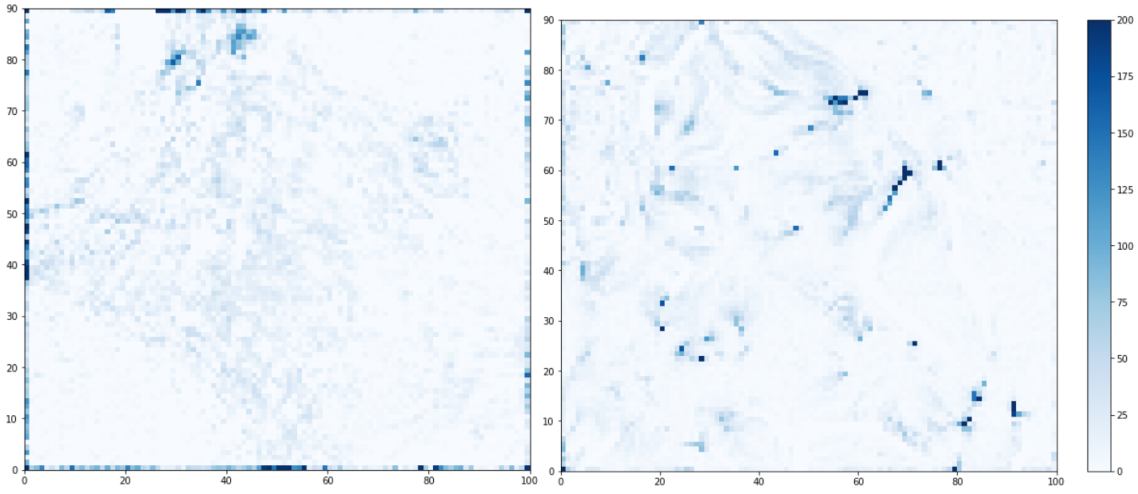


Figure 3.4 Neuron activation map for Stage 1 SOM (left) and Stage 2 SOM (right).

The proportion of HGUs at each neuron are shown in Figure 3.5. Each neuron typically has more than one HGU. This reflects the variability of HGUs at each cluster location. The potential impact of SOM architecture on MLP performance is further discussed in Section 3.2.4.

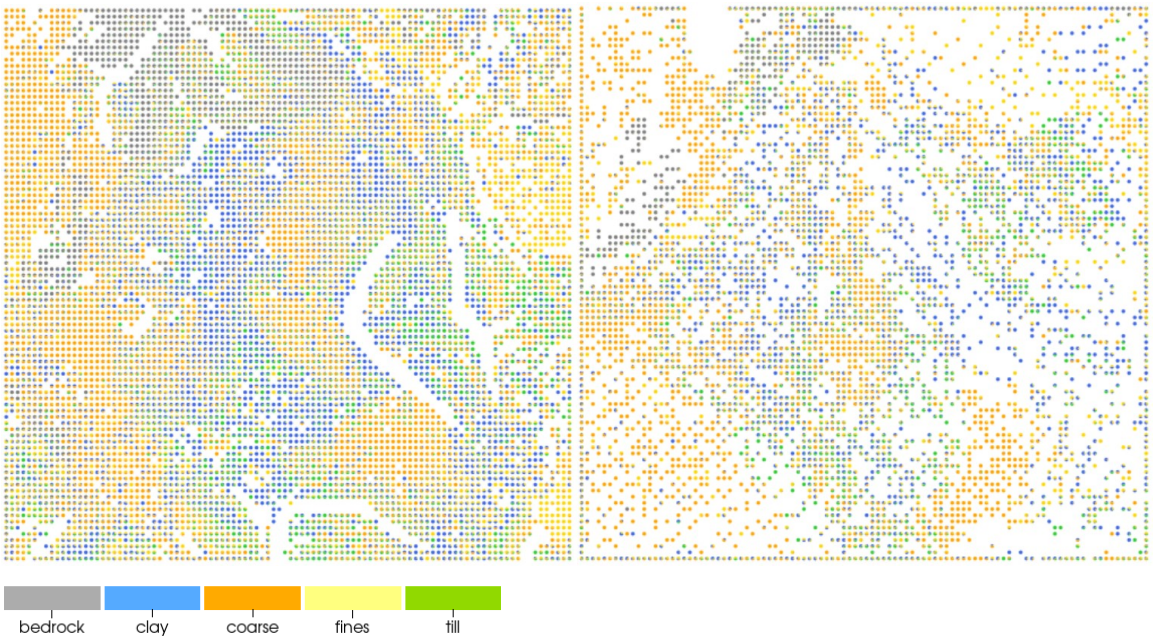


Figure 3.5 Pie plots showing the proportion of HGUs at each neuron for Stage 1 SOM (left) and Stage 2 SOM (right).

The trained SOM was used to map the co-ordinates of the cell centroids once they were transformed using the same standardization completed on the Alternative 3 dataset. This means that every cell centroid (over 3 million) was mapped to a neuron (7,578) on the 2D grid. Multiple centroids are mapped to the same neuron, resulting in spatial clusters when the data is viewed in 3D (Figure 3.2).

3.2. MLP

3.2.1. Data Preparation

Training and testing subsets were generated using 80% / 20% split of the data and a stratified sampling approach to ensure a consistent distribution of HGUs. Coordinates (northing, easting, and elevation for Alternative 1 and Alternative 2, X and Y for Alternative 3) were used as feature inputs, while the HGUs were assigned as targets to develop the MLP.

Standardization (-1 to 1) and normalization (0 to 1) was used to transform the training data which was then fit to a baseline MLP classifier to compare performance. The resultant balanced accuracy scores are shown in Table 3.3. The standardized input

data consistently resulted in better model performance; therefore, standardization was used to scale coordinates for all alternatives.

Table 3.3 MLP feature scaling balanced accuracy scores.

	Alternative 1	Alternative 2	Alternative 3
No Scaling	29	26	13
Standardized (-1 to 1)	39	46	42
Normalized (0 to 1)	35	43	36

3.2.2. Hyperparameter Optimization

Hyperparameters were optimized using a stratified 3-fold cross-validation grid search that considers all combinations of predefined hyperparameter values. Hyperparameters selected for optimization include hidden layer size, alpha, batch size, and initial learning rate as discussed in Section 2.6.2. The hyperparameter default values were used and modified by an order of magnitude to establish upper and lower limits of the search space as shown in Table 3.4. The best hyperparameters for each alternative based on grid search results are shown in Table 3.5.

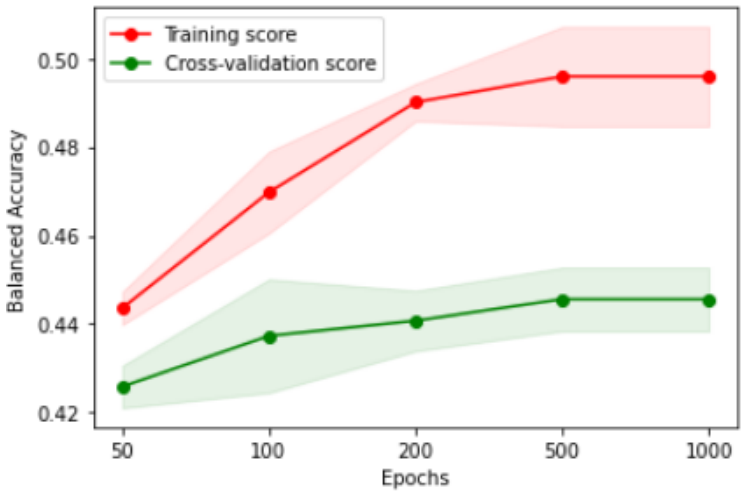
Table 3.4 Search space for MLP hyperparameter optimization.

Parameter	Search Space		
Hidden Layer Size	100x3	100x4	100x5
Alpha	0.00001	0.0001	0.001
Batch Size	20	200	2000
Initial Learning Rate	0.0001	0.001	0.01

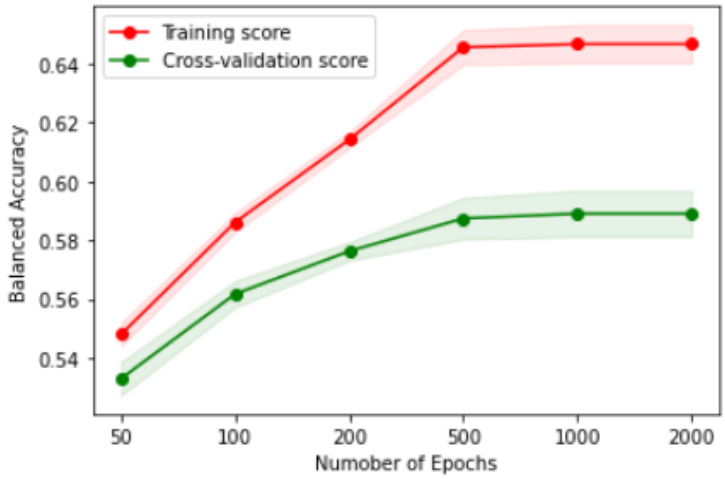
Table 3.5 Optimized hyperparameters for each MLP training model.

Hyperparameter	Alternative 1	Alternative 1	Alternative 3
Hidden Layer size	100x5	100x5	100x5
Alpha	0.0001	0.0001	0.0001
Batch Size	2000	200	200
Initial Learning Rate	0.01	0.001	0.001
Epochs			

a)



b)



c)

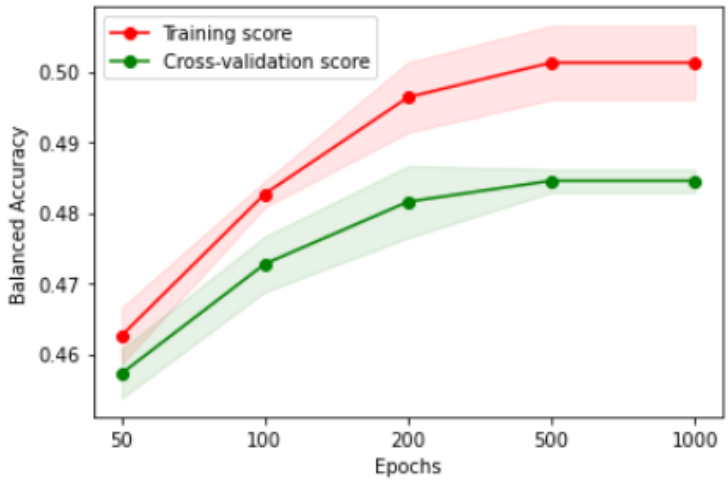


Figure 3.6 **Balanced accuracy versus number of epochs for a) Alternative 1, b) Alternative 2, and c) Alternative 3.**

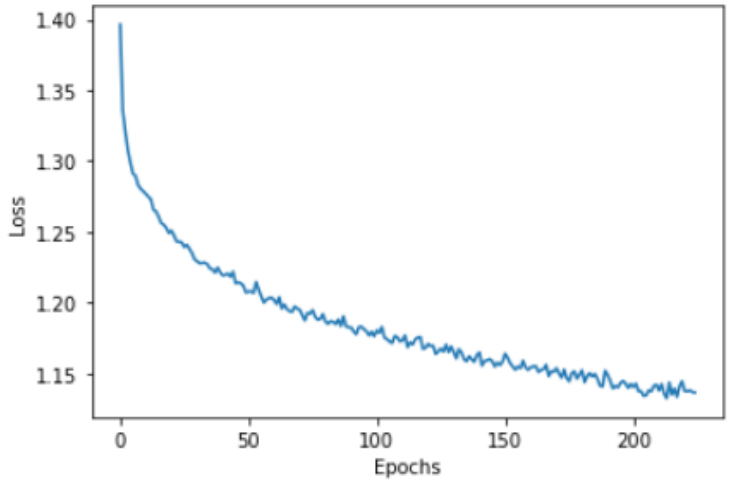
Balanced accuracy scores for various iterations are shown in Figure 3.6. Each epoch (X-axis) is equal to one pass of the entire data set through the algorithm. In general, balanced accuracy increases for both the training and validation datasets as the number of epochs increases, indicating an improvement in model performance and lower model bias. Balanced accuracy typically plateaus after 200 to 500 epochs depending on the alternative. Model variance (gap between training and validation scores) also increases as the number of epochs increases, indicating a trade-off with improvements in model bias. The number of epochs in the training model was updated from the default value of 200 to 500 for Alternative 2 and was not modified for Alternative 1 and Alternative 3 based on Figure 3.6.

3.2.3. Training and Testing

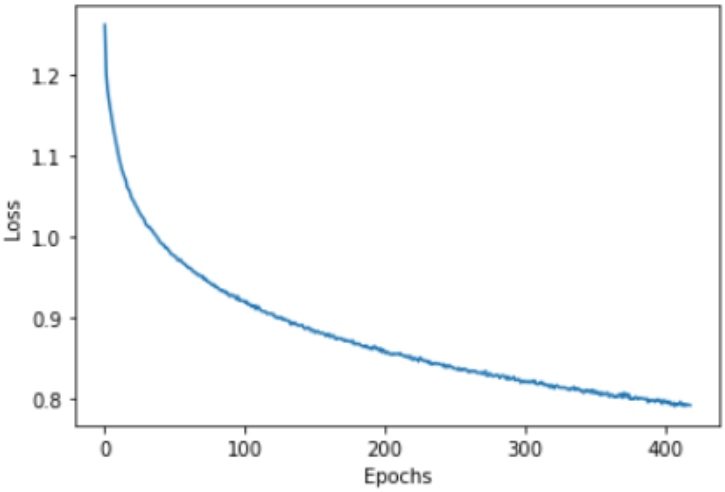
The model was configured using the optimized hyperparameters and iterations and then trained using the training data. The generalization performance of the model was then evaluated by using the testing data. Performance metrics (log-loss, learning curve, balanced accuracy, confusion matrix) were reviewed to evaluate model performance.

The log-loss curves from training are shown in Figure 3.7. This shows the performance of the MLP during training as the number of epochs increase. MLP tries to minimize log-loss using stochastic gradient descent by adjusting weights during training. At the end of training, Alternative 2 has the lowest log-loss at 0.8, followed by Alternative 3 at 1.10 and Alternative 1 at 1.15. For a naïve classification model, which assumes the same probability for each HGU, the log-loss score would be 1.3 (e.g. -log(0.2)). Using this as a baseline score, the MLP models for all three alternatives achieve a lower log-loss.

a)



b)



c)

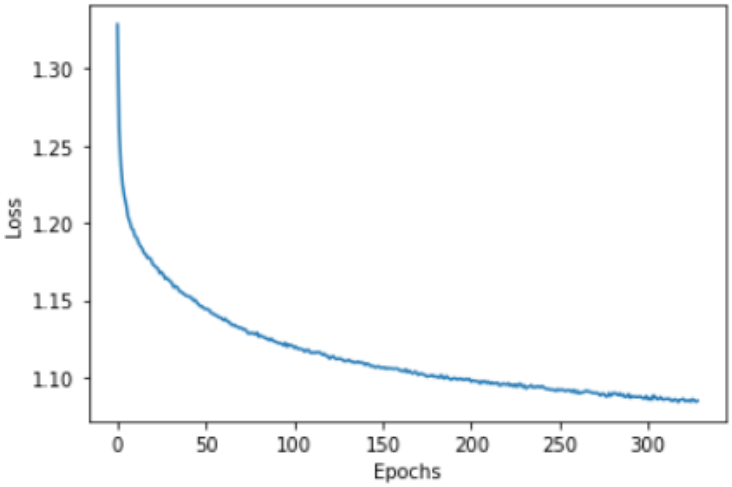


Figure 3.7 **Training model log-loss curves for a) Alternative 1, b) Alternative 2, and c) Alternative 3.**

Learning curves are shown in Figure 3.8. They show the training and validation scores for different training set sizes for each alternative. For Alternative 1 and Alternative 2, the training score is the highest with smaller number of samples and corresponds to the lowest validation scores because the model does a poor job predicting unseen data. As the number of training samples increases, the training score decreases but the model does a better job predicting unseen data based on higher validation scores. This is the normal behaviour for a learning curve (Yang, 2018). For Alternative 3, the balanced accuracy increases for both training and validation

a)

b)

c)

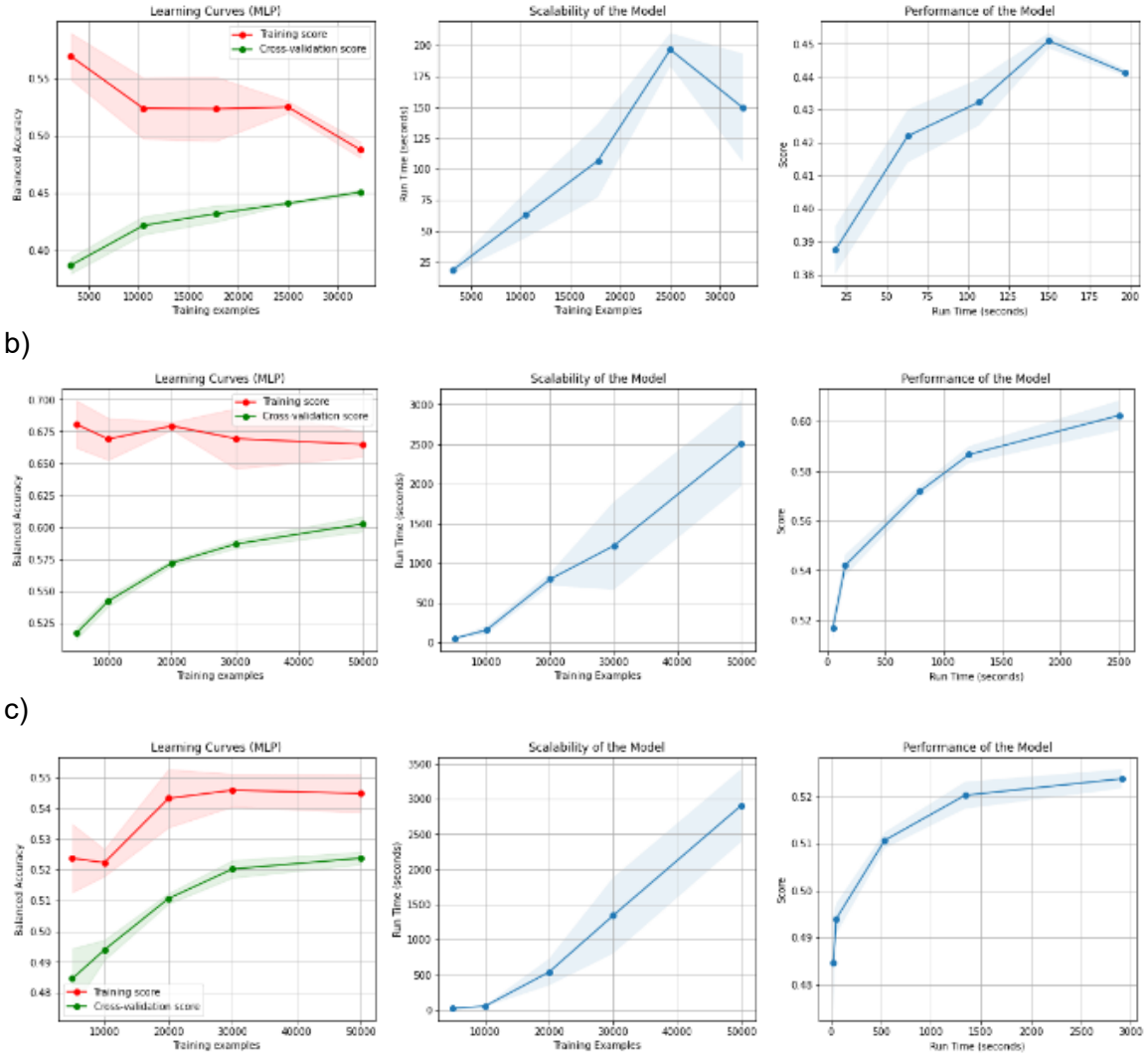


Figure 3.8 Training model learning curve, scalability based on the number of training examples, and model performance relative to run time for a) Alternative 1, b) Alternative 2, and c) Alternative 3.

scores as the number of samples increases. This may be attributed to multiple HGUs being assigned to the same winning neuron. As more samples are used for training, the model does a better job predicting the most frequently mapped HGU at each neuron.

Learning curves indicate models that show relatively high bias (based on relatively low balanced accuracy) and low variance (given the relatively small gap between training and validation scores). The models appear to be underfitting the training data. Adding more training examples is unlikely to lead to better training based on training curves that generally appear to plateau. This could be attributed to the variability in HGU distribution and the limitations of the algorithm in reproducing this level of subsurface complexity. Additional data may improve validation scores given that validation curves do not appear to have plateaued. The runtimes generally increase linearly as the number of training examples increase. Runtimes for Alternative 2 and Alternative 3 are similar and are generally an order of magnitude higher compared to Alternative 1. Model performance appears to plateau as runtimes increase.

Table 3.6 provides the balanced accuracy scores for training and testing of each MLP model. MLP tuning (e.g. hyperparameter optimization, number of epochs) improved training performance compared to the use of default MLP values for all alternatives. Tuning improved balanced accuracy scores by 9 to 17%. The alternatives for data selection used in the modelling also impacted scoring. Alternative 2 had the highest training and testing scores. Model variance (difference between training and testing score) is lowest for Alternative 3 and highest for Alternative 2 with values below 5%. The balanced accuracy scores indicate modelling results with a high bias and low variance. Typically, training and testing accuracy scores above 80% indicate good performance for other machine learning applications (machinelearningmastery.com accessed on April 11, 2022); however, this may not be achievable given the multi-class classification problem and variability in subsurface data for this study.

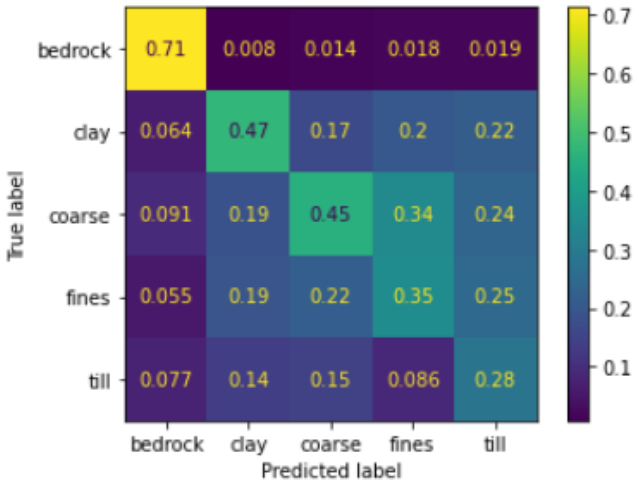
Table 3.6 Balanced accuracy scores for MLP training models developed using the data selection alternatives.

	Alternative 1	Alternative 2	Alternative 3
Training (default)	39	48	42
Training (tuned)	48	65	51
Testing	44	60	49

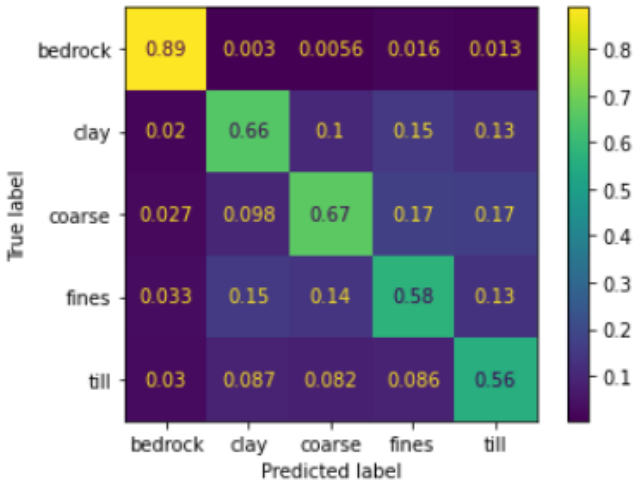
The confusion matrix from the testing results is shown in Figure 3.9 for each alternative. The scoring has been normalized to show the precision accuracy of the predicted HGUs for each alternative. For all alternatives, the MLP predicts bedrock the most accurately despite bedrock samples occurring less frequently in the datasets (Figure 3.1). This could be attributed to the continuity of bedrock once it is encountered. The prediction performance for unconsolidated HGUs follows the same trend for all alternatives where Coarse has the second highest precision with lower precision in descending order for Clay, Fines, and Till. This may be attributed to the distribution of unconsolidated HGUs (Figure 3.1) or could be associated with discontinuity of unconsolidated HGUs in the subsurface and the underfitting of the model to capture this level of complexity. From Figure 3.9, Clay and Till are most often confused with each other while Coarse seems to be confused most often with Till and Fines. Fines are commonly confused with Till and Clay.

Several ideas were considered to improve model performance but others exist that were not explored (Brownlee, 2019a). The regularization term for MLP (e.g. alpha) was reduced to address high bias and low variance with minimal success. Alternative 2 and Alternative 3 make changes to the original dataset and result in better model performance but the model still appears to underfit the data. Algorithm tuning (e.g. hyperparameter optimization, number of epochs) improved model performance but model bias remains relatively high based on performance metric standards for other machine learning applications. A literature search was conducted to compare performance metrics of MLP results for other geoscience applications, but none were found.

a)



b)



c)

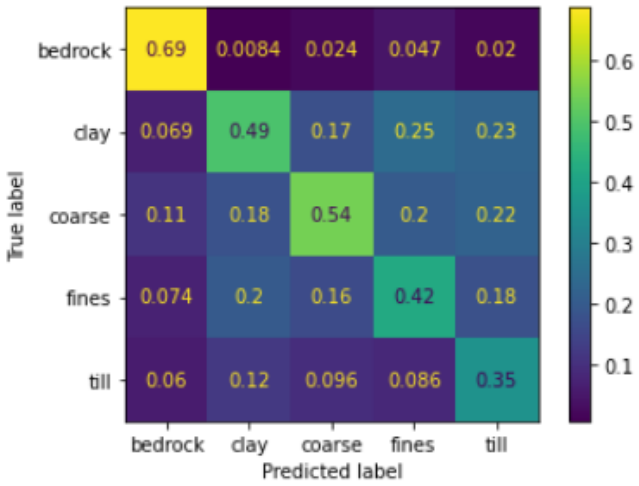


Figure 3.9 **Normalized confusion matrix over the predicted HGU based on testing results for a) Alternative 1, b) Alternative 2, and c) Alternative 3. The values along the diagonal represent the percentage of correctly classified samples. The percentage values off the diagonal show what the true HGU was confused with (e.g. 18% of the sample predicted as till should have been fines).**

3.2.4. Predictions

Before making predictions, all features for each alternative were scaled and used to train the MLP. Scaling was based on the coefficients calculated from the training data. Once trained, the MLP was used to predict HGUs at scaled coordinates representing the entire study area for each alternative. Cell centroid coordinates from the mesh were used for Alternative 1 and Alternative 2 while the coordinates from 2D SOM mapping were used for Alternative 3. This resulted in the prediction of HGUs at a total of 3,336,175 unique cell centroids and 7,578 unique SOM neurons.

Select model outputs (e.g. HGU predictions, coarse HGU probabilities, low permeability probabilities, and entropy) were visualized using ParaView to compare MLP results. Outputs were visualized in 3D cross-sectional view roughly positioned near the center of Aquifer 58 with additional 3D visualization of probabilities greater than 75%.

The cross-sectional view of predicted HGUs for each alternative is shown in Figure 3.10. Alternative 1 appears to be the most underfit which is expected given it has lowest balanced accuracy score. Alternative 2 shows more complexity in the subsurface compared to Alternative 1 and Alternative 2. Alternative 3 is more comparable to Alternative 2 but is more generalized. This suggests the dimensions of the 2D SOM could be inadequate to represent the subsurface complexity.

Figure 3.11 shows the probability distribution of the Coarse HGU for each alternative. Additional 3D visualization showing the distribution of Coarse probabilities above 75% is shown in Figure 3.12. Alternative 2 appears to have a larger coverage and higher probability values for Coarse HGU compared to the other alternatives.

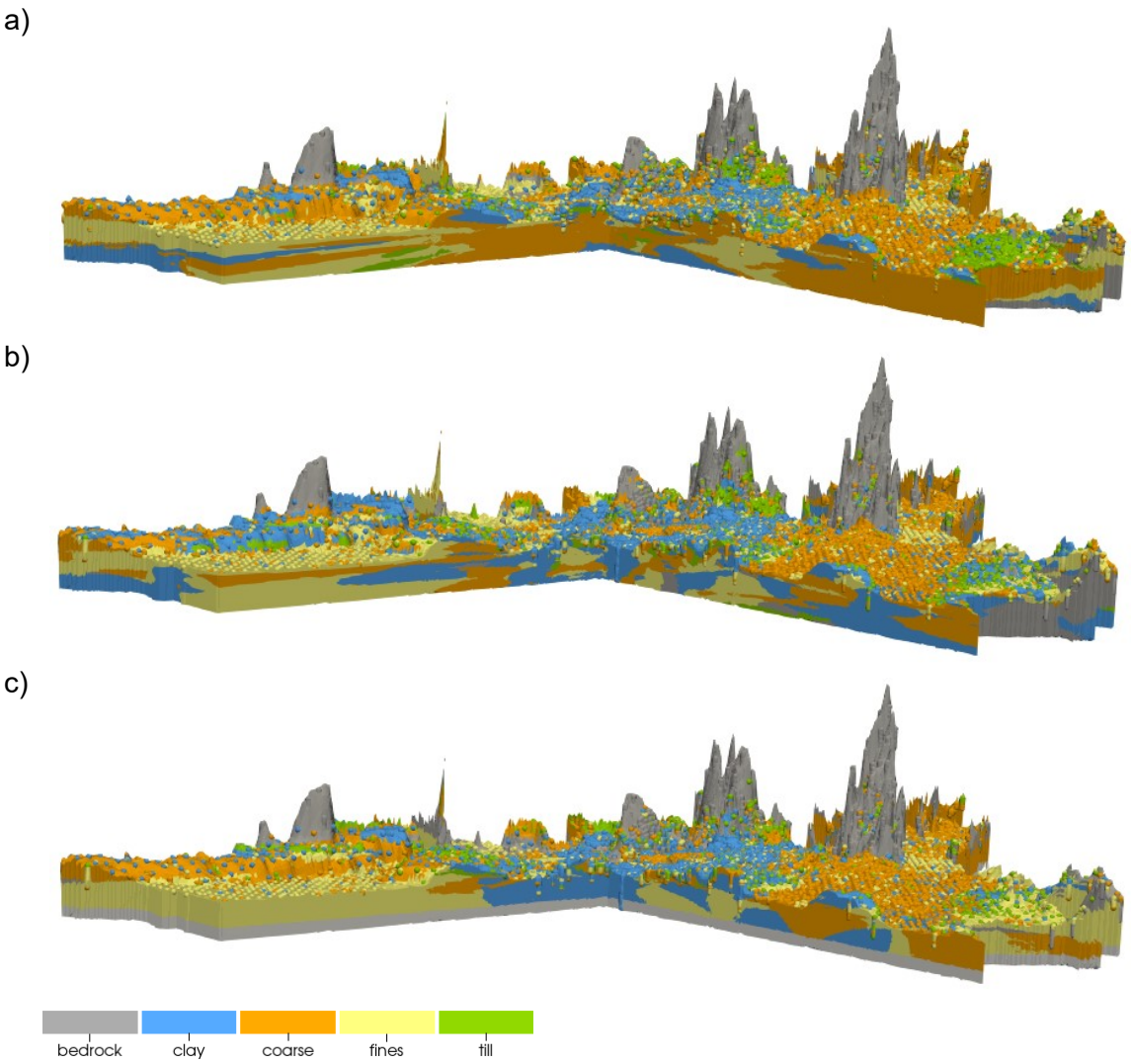


Figure 3.10 Cross-sectional view of HGUs from the MLP predictive model developed using a) Alternative 1, b) Alternative 2, and c) Alternative 3.

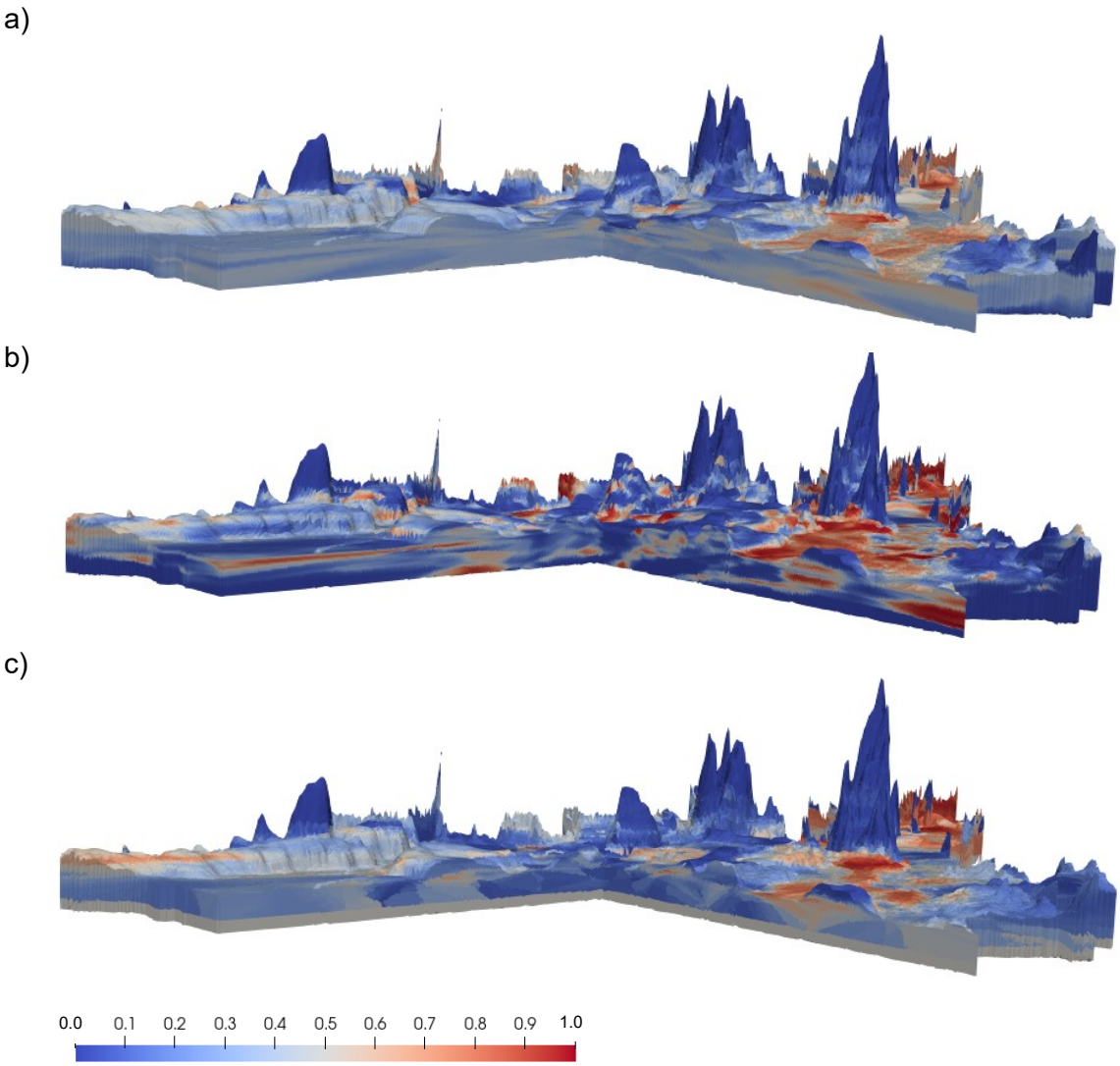


Figure 3.11 Cross-sectional view of probability for Coarse HGU from the MLP predictive model developed using a) Alternative 1, b) Alternative 2, and c) Alternative 3.

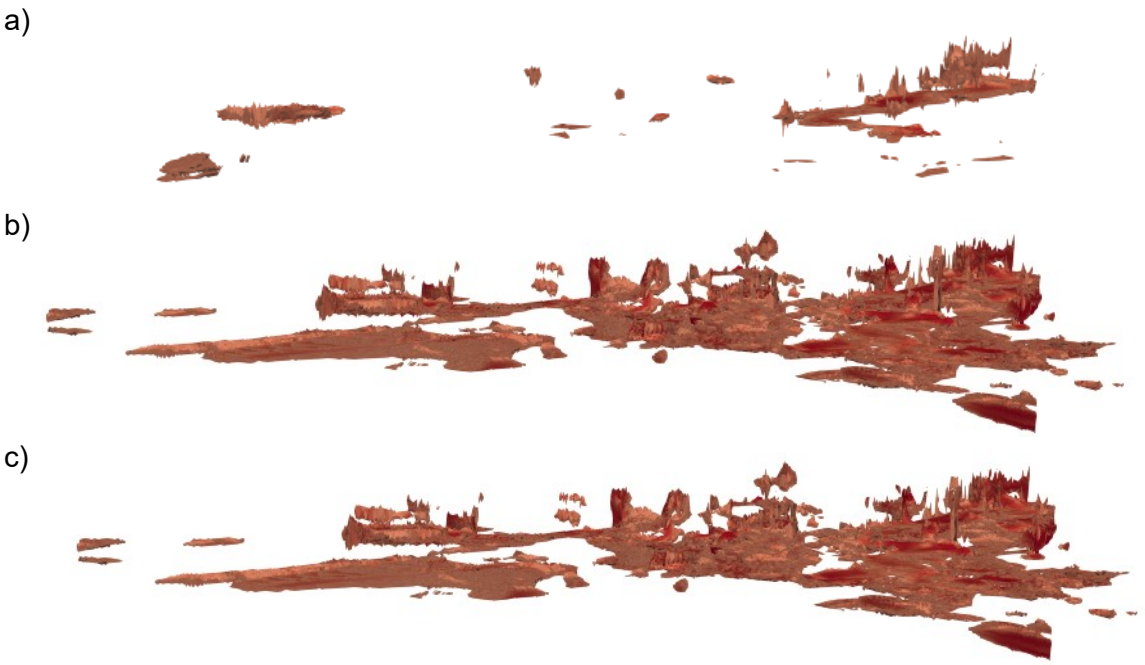


Figure 3.12 Cross-sectional view of probability above 75% for Coarse HGU from the MLP predictive model developed using a) Alternative 1, b) Alternative 2, and c) Alternative 3.

The probability distribution of low permeable material is shown in Figure 3.13. The probabilities of both Till and Clay HGUs were added to calculate the probability of low permeable material. Additional 3D visualization showing the probabilities above 75% for this material is shown in Figure 3.14. Similar to the Coarse HGU, Alternative 2 has the largest coverage of low permeability areas with relatively high probabilities.

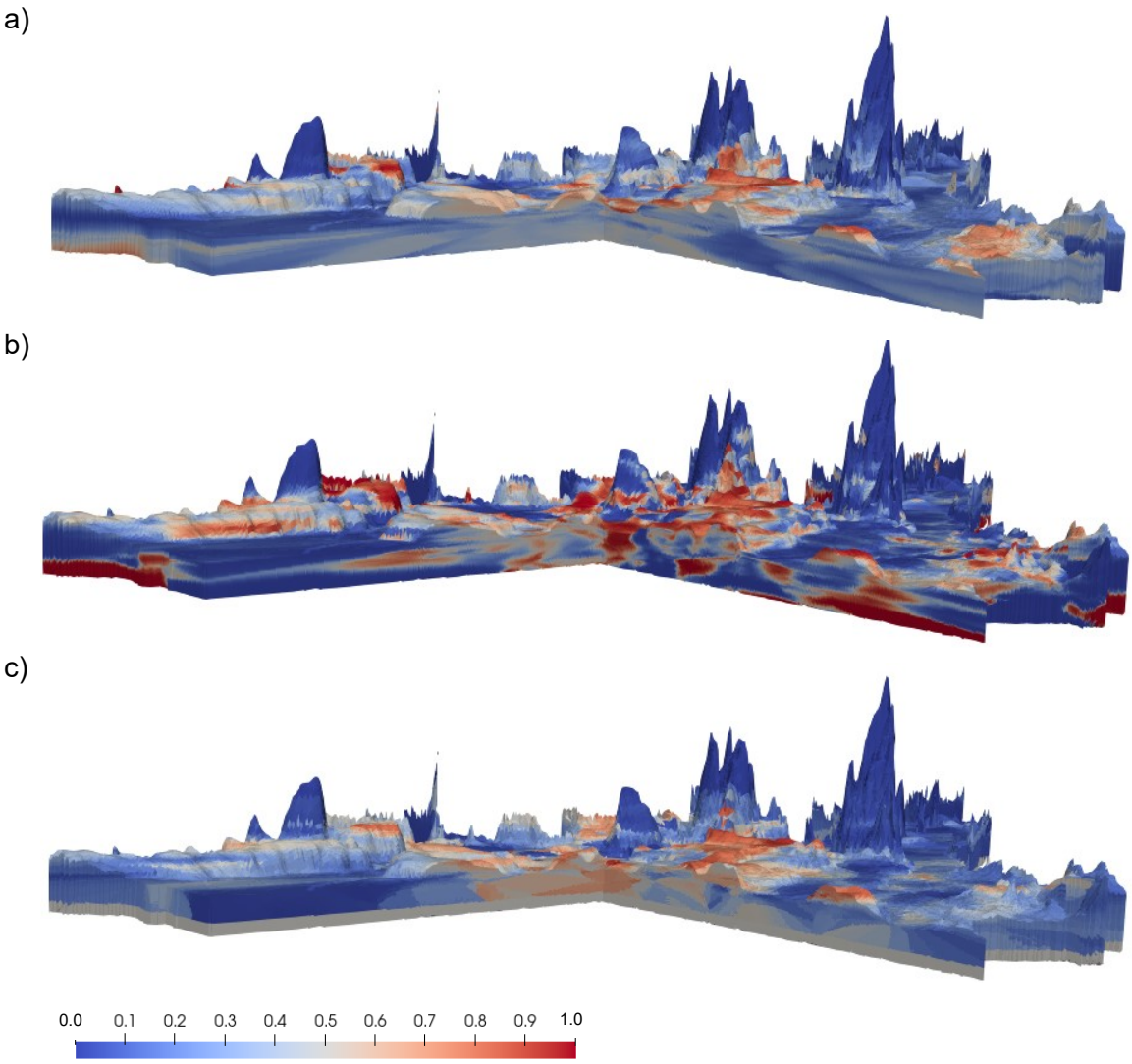


Figure 3.13 Cross-sectional view of probability for low permeability material (Till and Clay HGUs) from the MLP predictive model developed using a) Alternative 1, b) Alternative 2, and c) Alternative 3.

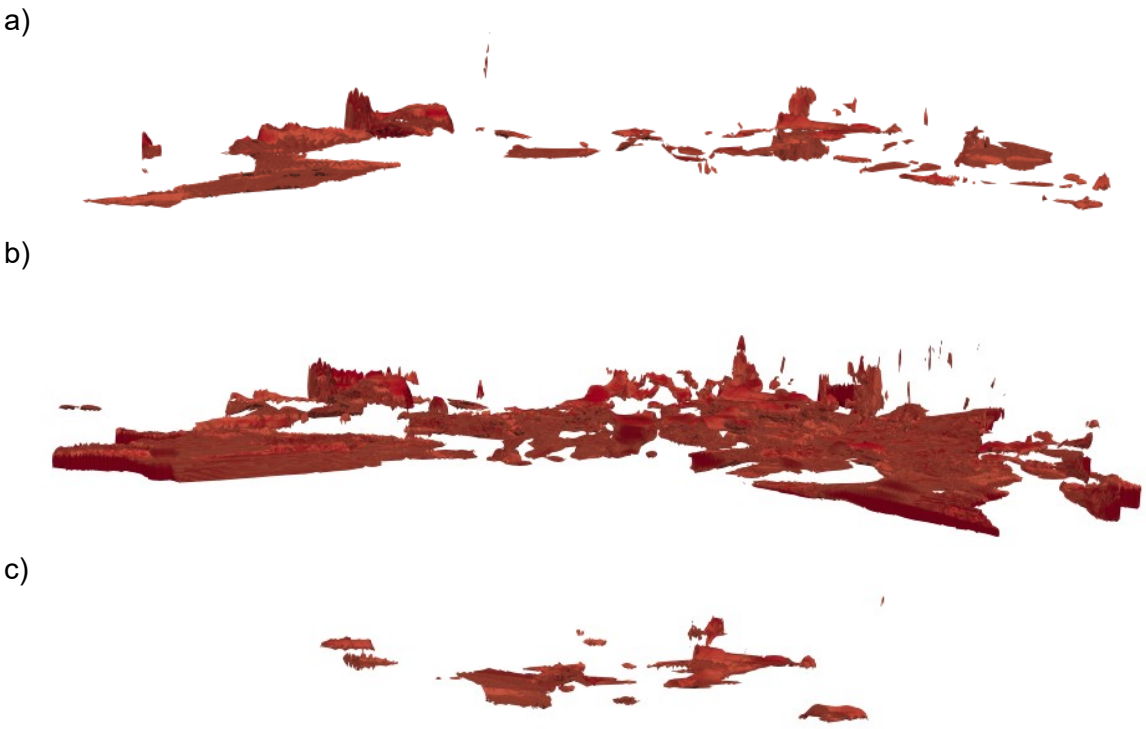


Figure 3.14 Cross-sectional view of probability above 75% for low permeability material (Till and Clay HGUs) from the MLP predictive model developed using a) Alternative 1, b) Alternative 2, and c) Alternative 3.

The last classification output is shown in Figure 3.15 and visualizes entropy for the three alternatives. A lower entropy value indicates more confidence in the predictions while higher values reflect greater uncertainty. The spatial coverage of higher entropy values is generally greater for Alternative 1 and Alternative 3 compared to Alternative 2. The average entropy values of 0.72, 0.28, and 0.64 for Alternatives 1, 2, and 3, respectively, indicate the most confidence in Alternative 2 results. The relatively small distribution of samples within the study area (e.g. less than 5% of the mesh cells have data), particularly at deeper elevations within the model, also contributes to uncertainty which does not appear to be reflected in entropy values. Extrapolation was required to make predictions given the distribution of available data for the region.

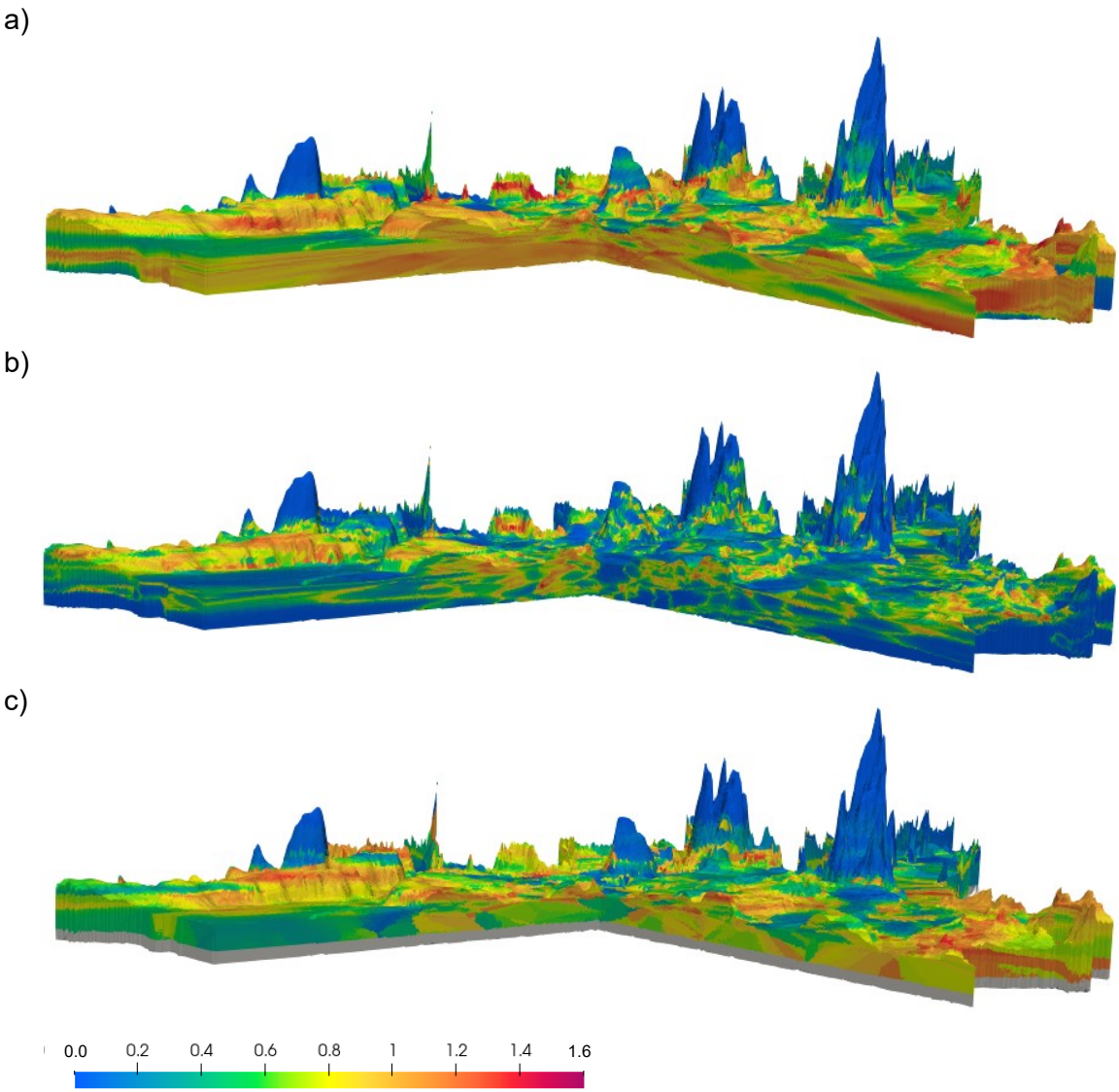


Figure 3.15 Cross-sectional view of entropy calculated based on the MLP predictive model developed using a) Alternative 1, b) Alternative 2, and c) Alternative 3.

Compared to Alternative 1 and Alternative 3, Alternative 2 had the lowest log-loss (Figure 3.7), highest balanced accuracy scores for both training and testing (Table 3.6), and highest precision for all HGUs (Figure 3.9). Based on the cross-sectional reviews, Alternative 2 shows more complexity in the subsurface compared to Alternative 1 and Alternative 3. Alternative 2 may have performed better because HGUs were averaged based on the cell mode which may have reduced some noise associated with the variability of HGU distribution for Alternative 1. The size of the 2D SOM for Alternative 3 may have been inadequate to represent the subsurface complexity. As a result,

Alternative 2 was selected as the preferred alternative based on better performance metrics and cross-sectional reviews that indicated more complexity in the subsurface compared to the other alternatives. Further verification of this geologic model is discussed in Chapter 4.

Chapter 4.

Verification against subsurface interpretations and hydrogeologic indicators

The geologic model results from MLP using Alternative 2 data were compared to interpolation using k-nearest neighbours (KNN) as well as published interpretations of the subsurface and hydrogeologic indicators within the region (e.g. artesian conditions). This verification is intended to determine how representative the predictive results are by considering alternative analytical techniques, independent studies, and additional hydrogeologic information.

4.1. K-Nearest Neighbours

KNN and MLP predictions of HGUs using Alternative 2 data were compared for verification purposes. This included a review of predicted HGU frequencies and geologic model outputs (e.g. HGU predictions, coarse HGU probabilities, low permeability probabilities, and entropy). The KNN algorithm was selected as an alternative analytical technique since it is commonly used to interpret categorical data (Fuentes et al. 2020).

In general, the frequency of predicted HGUs is comparable between MLP and KNN. As shown in Figure 4.1, both algorithms have a high percentage of cells predicted as bedrock, clay, coarse, and fines while the lowest percentage is till. There are some differences in HGU frequencies between the two interpolation methods but they are typically less than 5%.

Plan views of the predicted surficial results shown in Figure 4.2 are comparable to the distribution of surficial geology point data (Section 2.2.4). Both algorithms generally reproduce the spatial distribution of HGUs from surficial mapping. KNN is able to reproduce HGUs that are more localized and results in a more heterogeneous distribution overall.

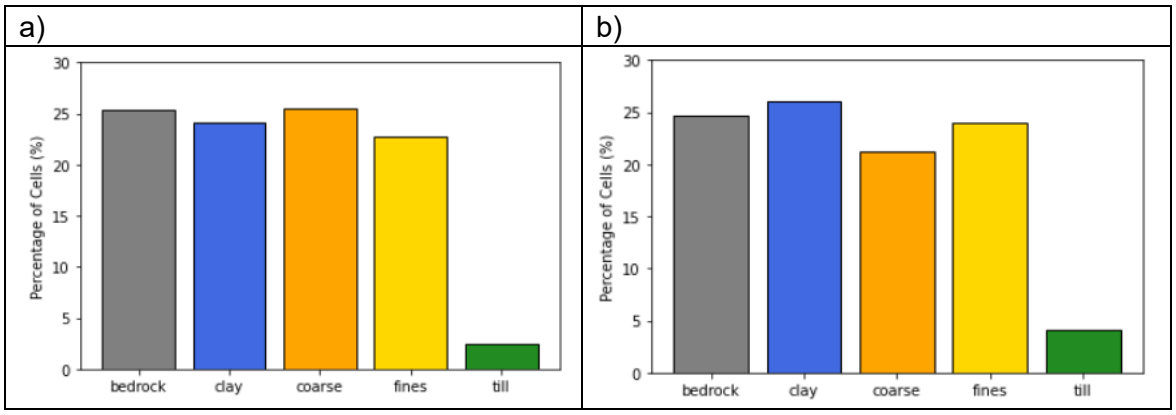


Figure 4.1 **Normalized frequencies of HGUs using Alternative 2 data and predictive models for a) MLP and b) KNN.**

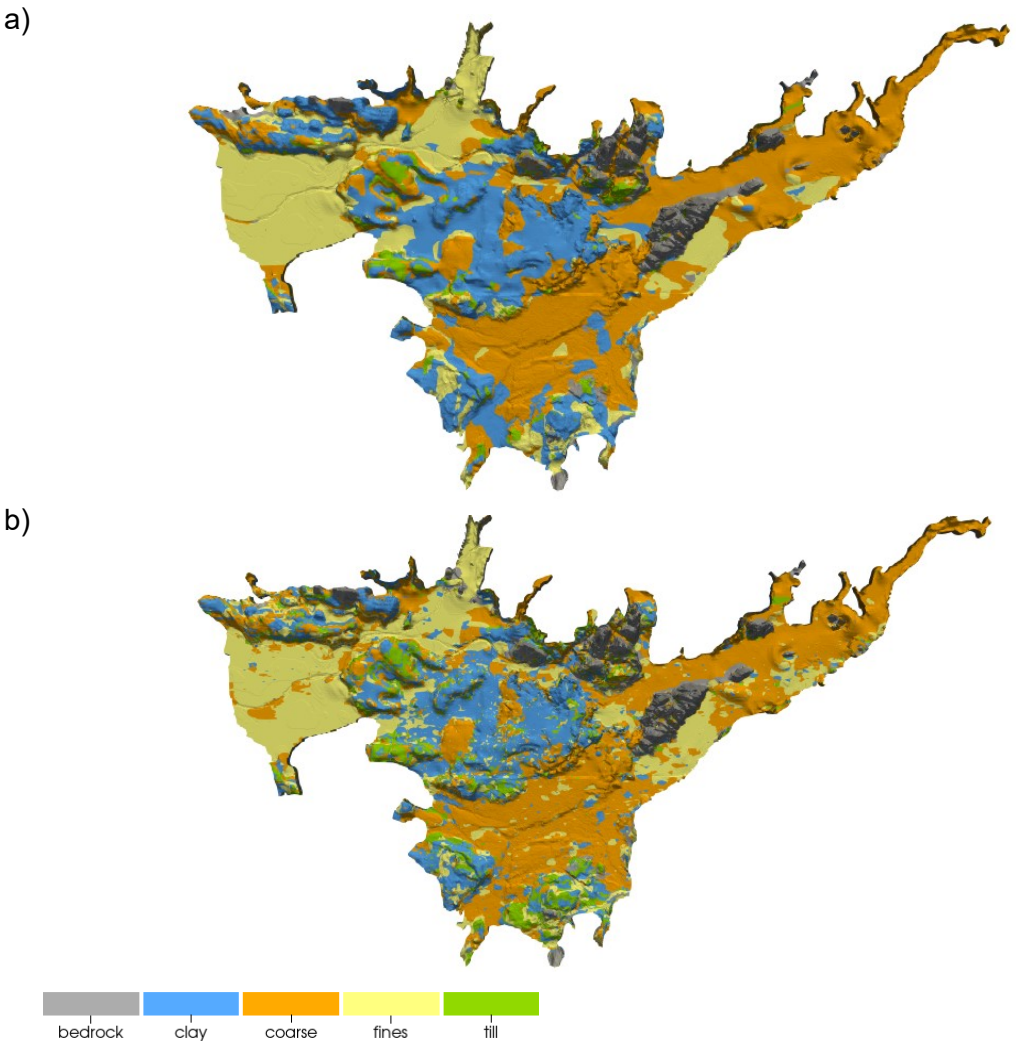


Figure 4.2 **Plan view of predicted HGUs based on a) MLP or b) KNN predictive models developed using Alternative 2 data.**

A cross-sectional view of predicted HGUs for each algorithm is shown in Figure 4.3. The HGU predictions made by KNN show more variation and are generally more discretized. The MLP predictions are more generalized but, as noted above, still have a similar frequency distribution and recreate the general subsurface trends throughout the study area similar to KNN. Both algorithms predict unconsolidated material below bedrock sections outside of the cross-section view; therefore, it may be best to incorporate bedrock directly into the geologic model and make predictions only on unconsolidated material.

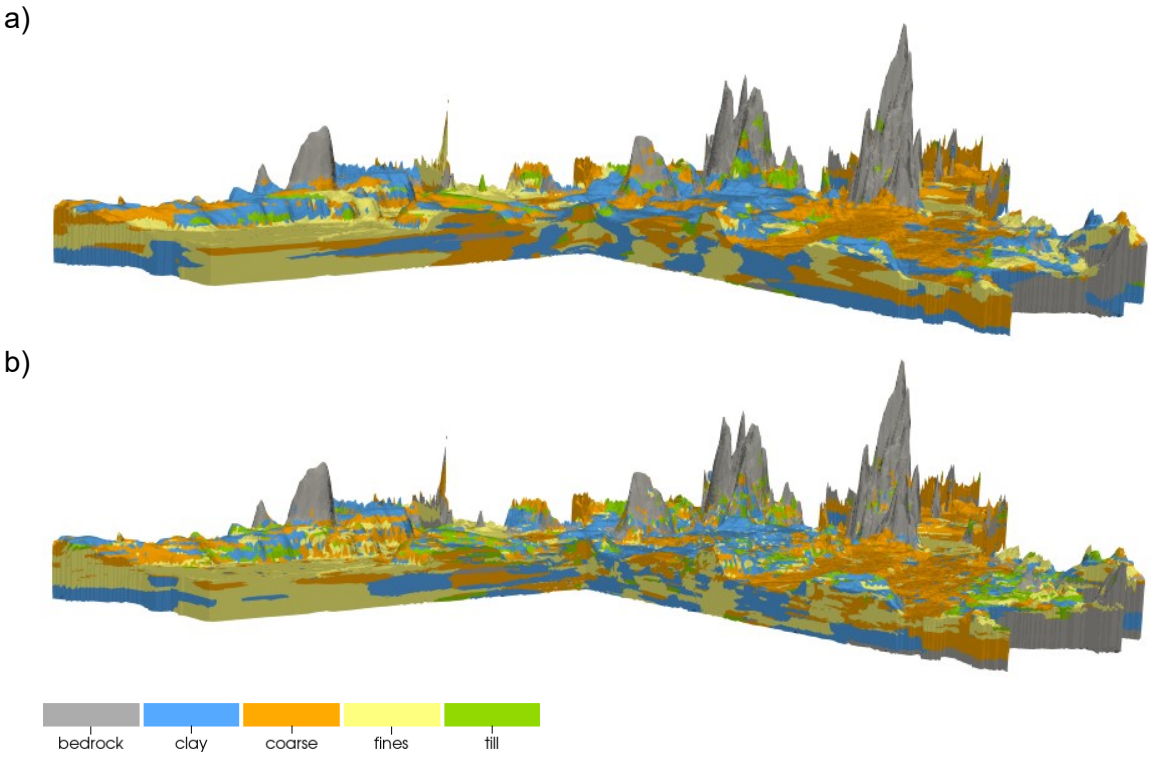


Figure 4.3 Cross-sectional view of predicted HGUs based on a) MLP or b) KNN predictive models developed using Alternative 2 data.

The probability distribution of Coarse HGU and low permeability material (combined probabilities of Till and Clay HGUs) for each algorithm are shown in Figure 4.4 and Figure 4.6, respectively, with additional 3D visualization of probabilities above 75% shown in Figure 4.5 and Figure 4.7. There are spatial similarities between the predictions from the two algorithms with a more generalized interpolation from MLP.

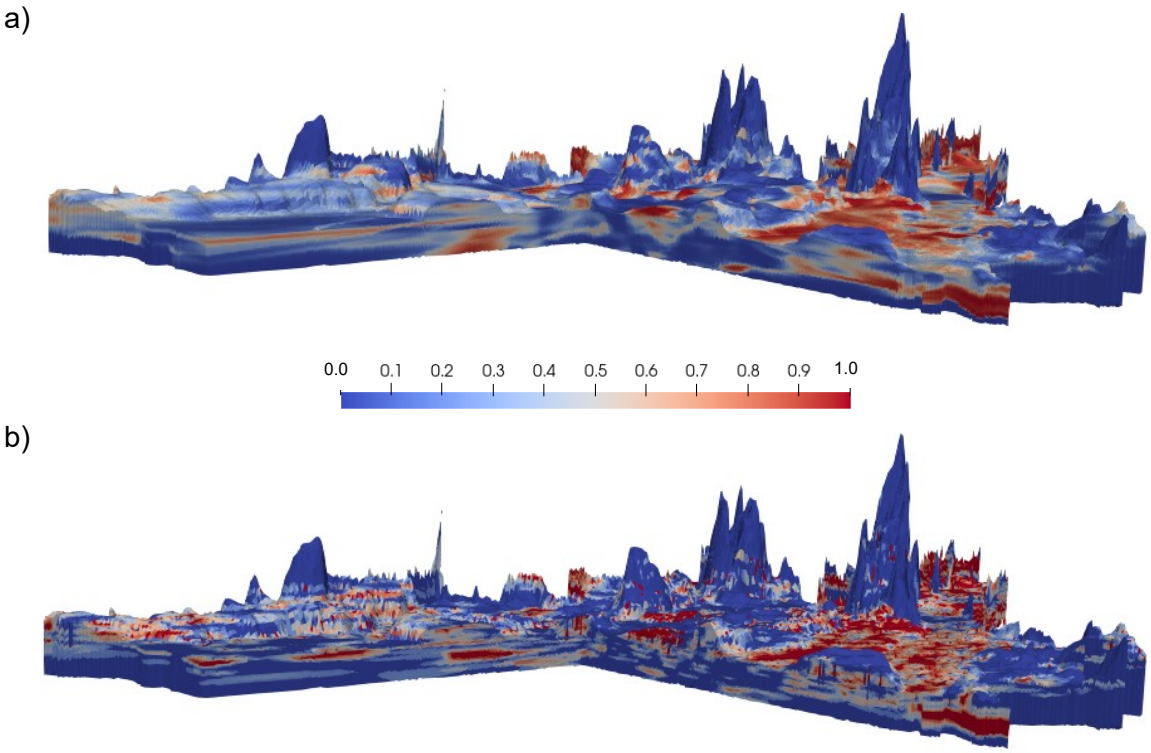


Figure 4.4 **Cross-sectional view of probability for Coarse HGU based on a) MLP or b) KNN predictive models developed using Alternative 2 data.**

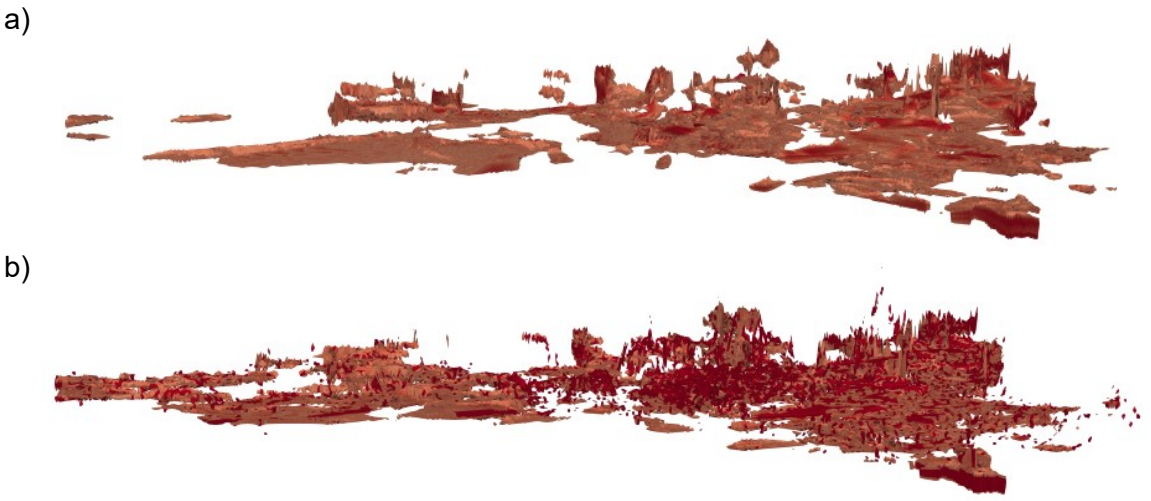
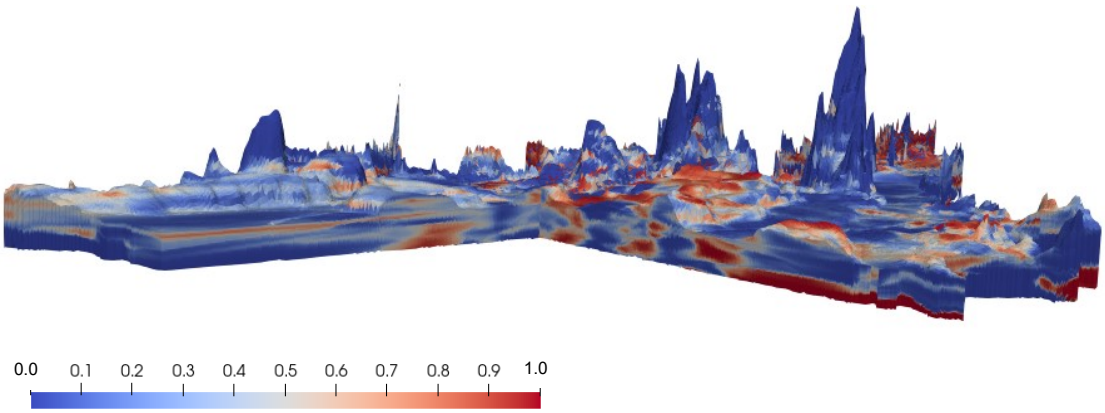


Figure 4.5 **Cross-sectional view of probability above 75% for Coarse HGU based on a) MLP or b) KNN predictive models developed using Alternative 2 data.**

a)



b)

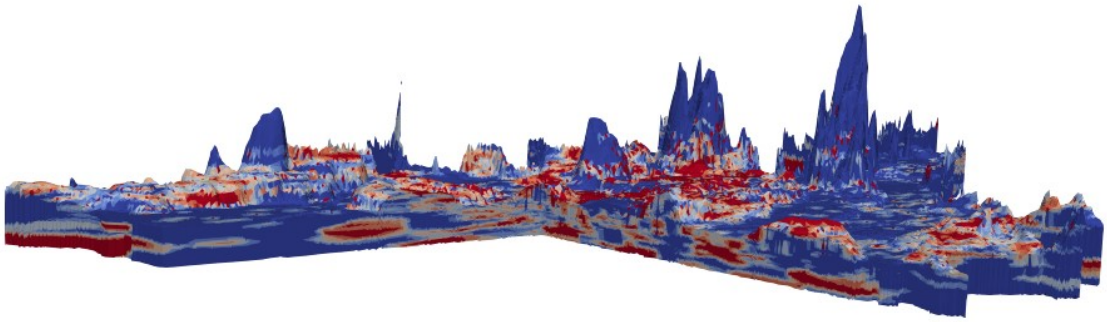


Figure 4.6 **Cross-sectional view of probability for low permeability material (Clay and Till HGUs(based on a) MLP or b) KNN predictive models developed using Alternative 2 data.**

a)

b)

Figure 4.7 **Cross-sectional view of probability above 75% for low permeability materail (Clay and Till HGUs) based on a) MLP or b) KNN predictive models developed using Alternative 2 data.**

Entropy values for HGU predictions from MLP and KNN are shown in Figure 4.8. Similar to the other modelling outcomes, entropy values generally have a smoother presentation and appear less blocky for MLP compared to KNN. MLP results in higher localized values of entropy but the overall average entropies are similar (0.28 for MLP versus 0.30 for KNN). The uncertainty associated with HGUs predicted at deeper elevations for both algorithms is likely underestimated using entropy given the limited distribution of subsurface data and extrapolation required to interpret the entire study area.

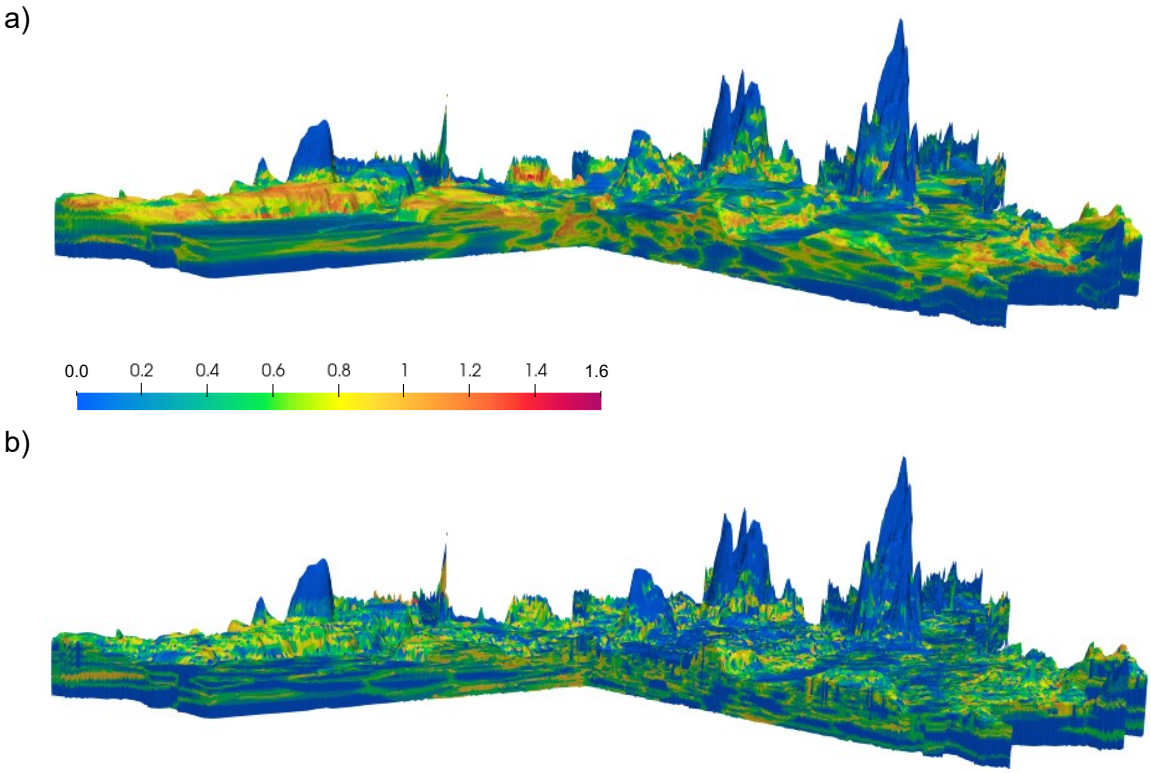


Figure 4.8 Cross-sectional view of calculated entropy based on a) MLP or b) KNN predictive models developed using Alternative 2 data.

Based on the above, predictions using MLP and KNN have comparable results. MLP provides a more generalized interpolation of outputs while the KNN appears to be more discretized and blocky in appearance.

4.2. Township of Langley

Geologic interpretations within the Township of Langley (TOL) published by Golder (2005) were reviewed to verify MLP predictions. This includes comparison of two geologic cross-sections and a 3D representation of major aquifers.

The representative geologic cross-sections include A-A' and B-B' aligned south-north and east-west, respectively, within the Langley Uplands (Figure 4.9). Cross-section A-A' is parallel to 240th Street beginning at 24th Avenue to 64th Avenue for 8.5 km. Cross-section B-B' is parallel to 31B Avenue and 31A Avenue starting at 260th Street to Lefeuvre Road for a distance of 4 km.

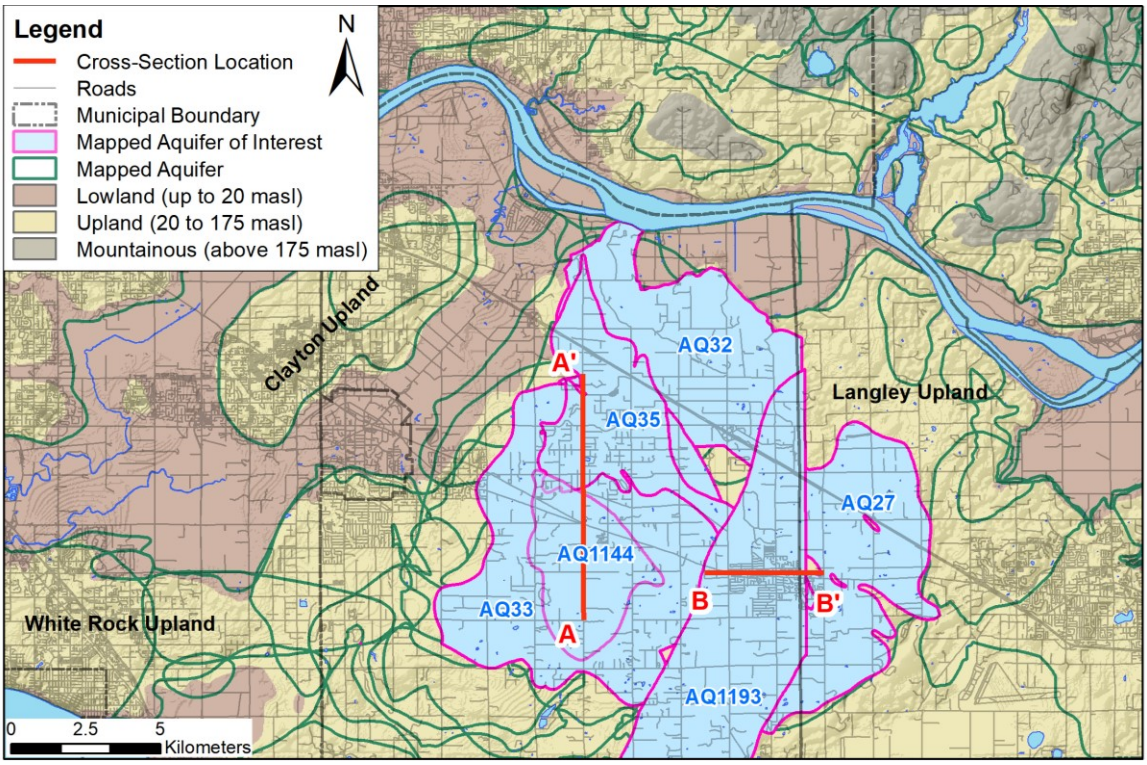


Figure 4.9 Approximate locations of Golder (2005) geologic cross-sections (A-A' and B-B') and provincial mapped aquifers of interest within the Township of Langley area.

Figure 4.10 and Figure 4.11 present cross-sections A-A' and B-B', respectively, for both interpretations from Golder and MLP. The MLP cross-section locations were prepared with a vertical exaggeration of 22 to be consistent with scaling used by Golder. MLP cross-sections include Alternative 2 point data (HGU mode at the cell centroid)

whereas the Golder cross-sections show borehole lithology but the well tag numbers are not included to identify the well.

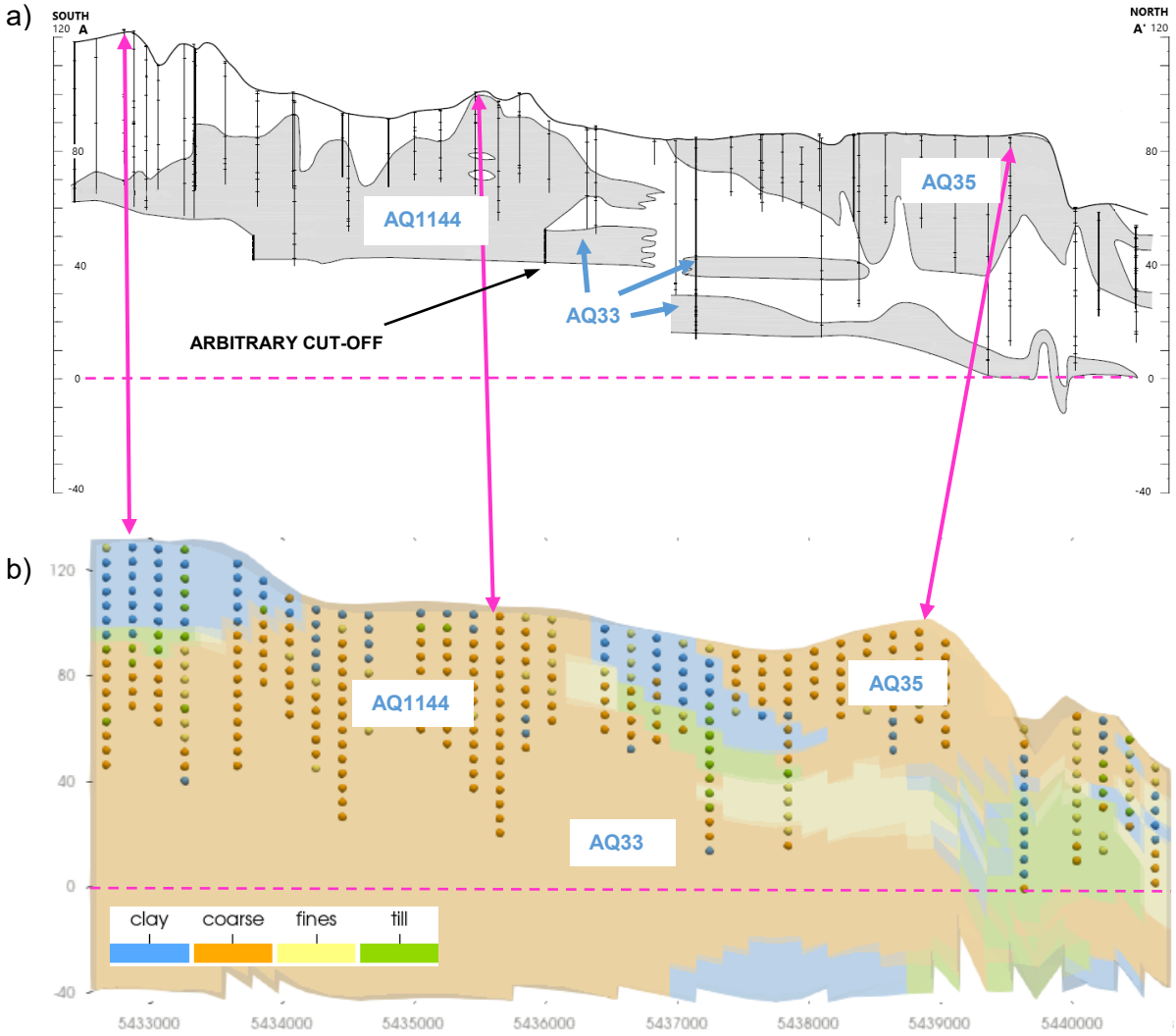


Figure 4.10 Geologic cross-section A-A' a) modified from Golder (2005) and b) from MLP predictions using Alternative 2 data. Provincial aquifers AQ1144, AQ33 and AQ35 are labeled and shaded grey in a. Pink arrows are possible tie lines.

Cross-section A-A' from the MLP geologic model shows AQ35 separately from AQ33 and AQ1144 consistent with the interpretation by Golder (2005). The distribution of Coarse HGU lumps AQ33 and AQ1144 into one aquifer unit which may be reasonable given arbitrary cut-offs were used by Golder to establish aquifer extents. Provincial mapping of AQ33 overlies AQ1144 although this is not shown on the Golder cross-section. There are interconnected areas between the three aquifers elsewhere in the

TOL area (Golder 2005). The Coarse HGU from MLP is interpolated at the surface where AQ1144 (confined aquifer) has been mapped and extends to deeper depths compared to interpretations by Golder. In general, the geologic model performs well at representing the subsurface in this area with the potential issue of extrapolation at depth.

Geologic cross-section B-B' showing interpretations from Golder (2005) and the MLP predictions are shown in Figure 4.11. There is more uncertainty in the alignment of this cross-section given the differences in subsurface data. Despite this, it appears AQ27 and AQ35 are lumped in the same Coarse HGU instead of being separate as interpreted by Golder. The original cross-section B-B' from Golder shows the aquifer material for AQ32 as sand and gravel but is described as 'a body of fine sands, sand and locally gravel and till' in the report. The MLP geologic model shows a relatively large Fines HGU approximately 40 m thick in this area. AQ33 and AQ1193 are lumped in the Coarse HGU from MLP predictions and have a larger extent compared to interpretations by Golder. The description for AQ1193 in Golder (2005) indicates the aquifer is located between +20 and -20 masl; therefore, it may be reasonable for the Coarse HGU to extend below 0 masl in this area. In general, the MLP cross-sections show a more generalized representation of aquifers and greater connectivity of permeable units compared to the Golder cross-sections.

Golder identified 18 major aquifers based on hydrostratgraphic interpretation of geologic units. For these major aquifers, permeable units that overlap by at least 10 percent and any aquitard between overlapping units less than 10 m thick were considered by Golder as 'well-connected hydraulically'. The 3D representation of the 18 major aquifers from Golder (Figure 4.12) was compared to MLP predictions (Figure 4.13). Coarse HGUs with a probability above 50% (50% Coarse) fom MLP modelling were arbitrarily selected and grouped by connectivity to represent geologic areas that are 'well-connected hydraulically'. This includes lithologic materials described as sand, sand and gravel, and gravel, but does not include fines such as 'silty sand' which may be important to this local area for deeper aquifer characterization purposes (Section 4.3). Grouping by connectivity assigns a connectivity zone for each group of connected cells. The elevation and connectivty of 50% Coarse from MLP modelling are shown in Figure 4.13 to provide a comparison to the 3D representation of major aquifers from Golder (2005).

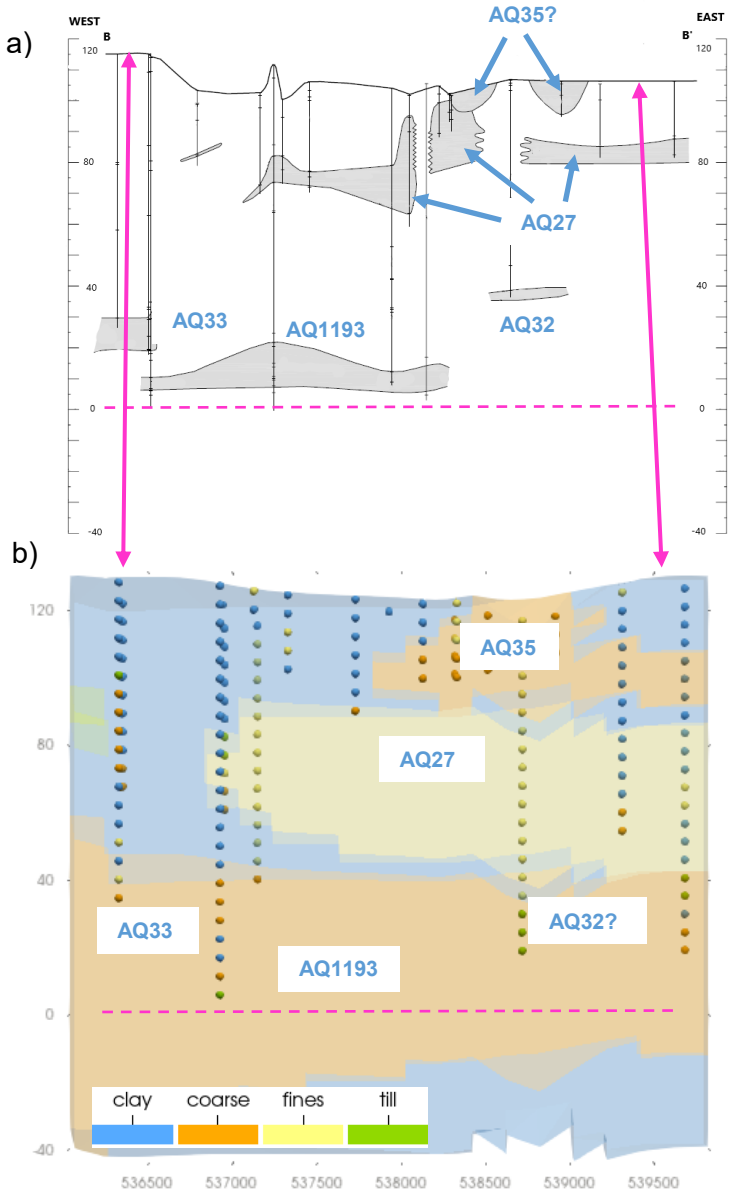


Figure 4.11 **Geologic cross-section B-B' a) modified from Golder (2005) and b) from MLP predictions using Alternative 2 data. Provincial aquifers AQ33, AQ1193, AQ27, AQ35 and AQ32 (possibly) are labeled and shaded grey in a. Pink arrows are possible tie lines.**

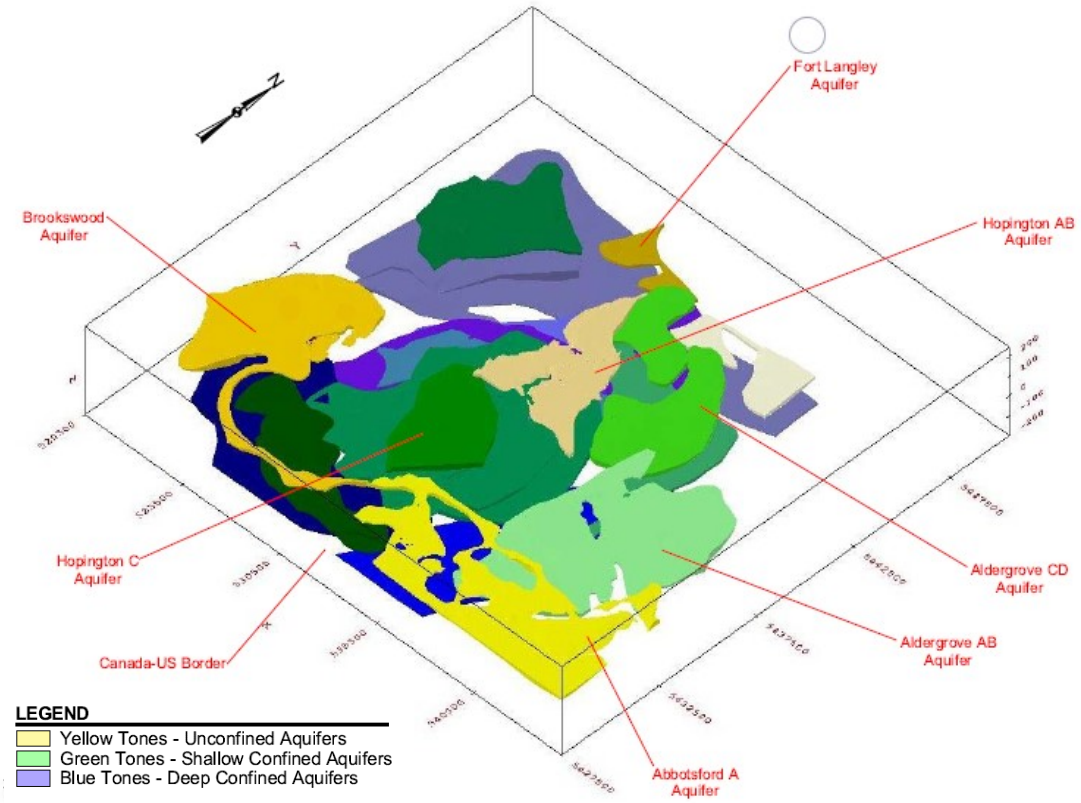


Figure 4.12 3D representation of major aquifers from Golder (2005).

The unconfined aquifers including the Brookswood Aquifer (AQ41), Fort Langley Aquifer (AQ41), Hopington AB Aquifer (AQ35), and Abbotsford A (AQ15) are represented in Figure 4.13; however, some of the linear features, interpreted as metlwater channels, extending from the main volume of the Brookswood (AQ41) and Abbotsford A (AQ15) aquifers by Golder (2005) are not reproduced using MLP. A finer mesh resolution and additional surficial data points could potentially be used to model this connection.

Figure 4.13b shows a large connected volume that consolidates several of the mapped aquifers in the area. This consolidated representation of aquifers may be plausible given the potential for interconnection noted by Golder (2005); however, the use of a higher probability as a cut-off limit may be more representative of aquifer extents which would be more conservative for water exploration but less conservative for water management purposes. Some of the deeper aquifers in Golder (2005) are only partially represented or not reproduced in Figure 4.13 likely due to aquifer materials with a greater fines content (e.g. silty sand) not being captured in the Coarse HGU but that

are locally significant as an aquifer unit. A more detailed review of data within this area is required to support conceptualization of intertill aquifers given the limited frequency of data points categorized as Till.

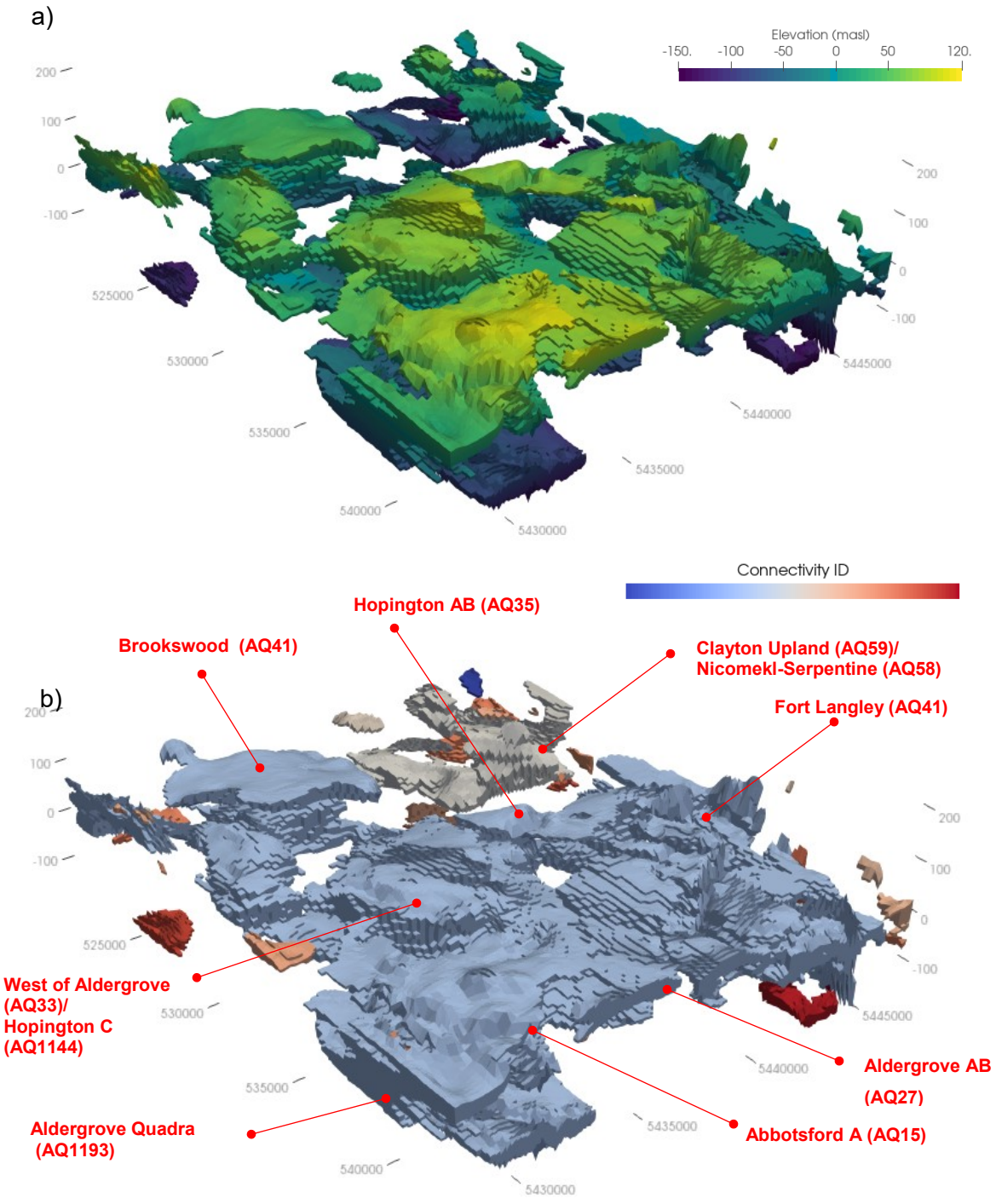


Figure 4.13 3D Representation of probability above 50% for Coarse HGU (50% Coarse HGU) showing a) elevation and b) connectivity zones.

4.3. Nicomekl-Serpentine Valley

Most flowing artesian wells within the study area are in the Surrey-Langley area, including the Nicomekl-Serpentine valley (Figure 4.14). Flowing artesian wells can occur when the aquifer is confined by low permeability materials or where there are large upward hydraulic gradients if the aquifer is unconfined. There are several provincially mapped aquifers within the area as shown in Figure 4.14. A well drilling advisory currently exists for Aquifer 58 (Nicomekl-Serpentine) due to the potential for flowing artesian conditions (FLNRORD, 2018). A recommendation to include a well drilling advisory for the western portion of Aquifer 33 (West of Aldergrove) has been made by Johnson et al. (2022).

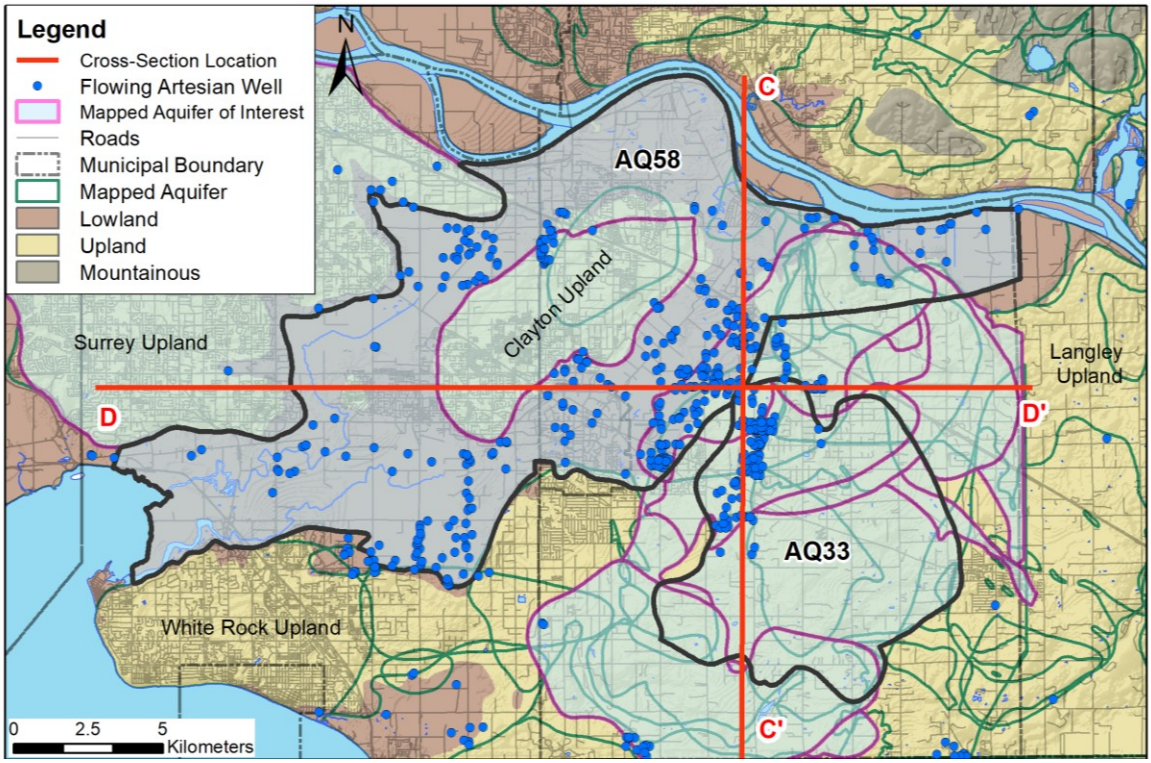


Figure 4.14 Location map of cross-section D-D' and C-C', flowing artesian wells in GWELLS, and Provincial mapped aquifers of interest within the Surrey-Langley area. Aquifer 58 and Aquifer 33 are outlined in black and labelled.

Aquifer 58 (Nicomekl-Serpentine) includes two permeable units consisting of a shallower unit generally occurring between -60 and -90 masl in the Clayton Upland area (Figure 4.14) and a deeper unit up to 20 m thick generally occurring between -120 and -150 masl that underlies the Clayton Upland but also extends along the northern portion

of TOL (Province of BC, 2016b). Aquifer 33 (West of Aldergrove) is described as an intertill aquifer consisting of two permeable units including a shallower unit between 5 to 15 m thick and a deeper unit up to 20 m thick, both sloping westward and merging along the western extent of the aquifer (Province of BC, 2016a).

Geologic cross-sections that intersect Aquifer 58 and Aquifer 33 are shown in Figure 4.15 based on MLP predictions trained using Alternative 2 data. The MLP predictions show upper and lower permeable units consistent with the description provided for Aquifer 58 and Aquifer 33. The continuity of the lower permeable unit for AQ58 is typically associated with the Fines HGU which includes subsurface materials described as silty sand. There are several overlapping confined aquifers in the Surrey-Langley area that were difficult to distinguish based on the MLP geologic model. Most flowing artesian wells appear to be screened in confined aquifers; however, several may be screened in unconfined Aquifer 35. The northward and westward sloping topography around unconfined Aquifer 35 likely contributes to flowing artesian conditions. A combination of topography and confining conditions along the western limit of Aquifer 33 may contribute to flowing artesian wells. The majority of flowing artesian conditions appear to be attributed to low permeability material overlying aquifer material and the topographical transition from upland to lowland as discussed below.

The distribution of clay in relation to flowing artesian wells in the Surrey-Langley area is shown in Figure 4.16. Till is not shown since the interpolated spatial extent is limited. Connectivity zones are assigned where clay cells are adjacent to each other. The majority of clay cells are connected in one zone (dark blue) that covers the majority of the area of interest with variability in both the vertical and horizontal directions. The rest of the connectivity zones for clay consist of a much smaller number of cells.

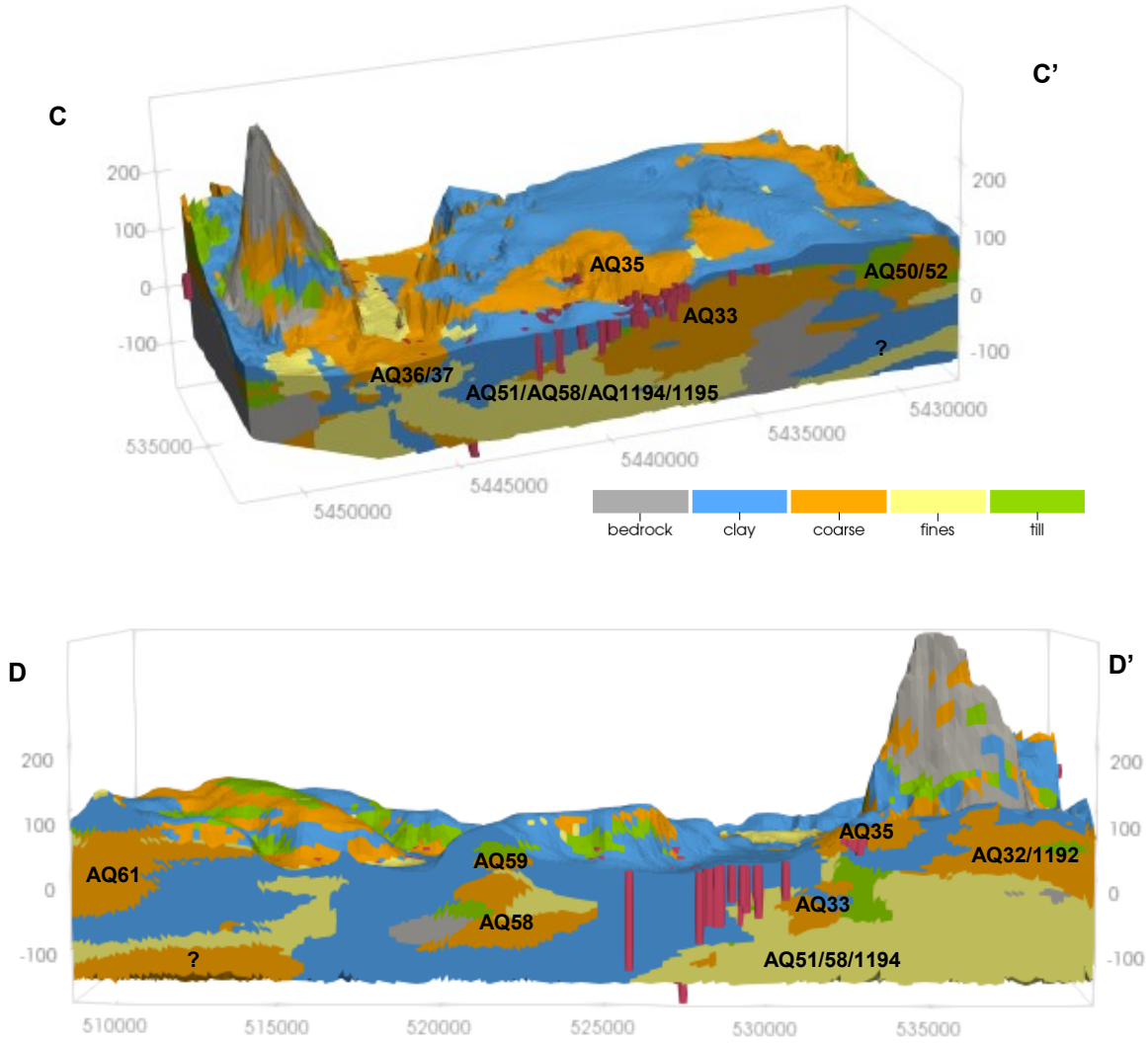


Figure 4.15 Geologic cross-section C-C' and D-D' showing predicted HGUs based on Alternative 2 data and MLP interpolation algorithm, artesian well locations (red tubes), and inferred provincial mapped aquifers. The lines of cross section are shown in Figure 4.14.

The top view in Figure 4.16 shows the clay connectivity zones intersecting the tops of most flowing artesian wells. The bottom view shows the bottom of most flowing wells extending through the clay. This suggests the model adequately represents the confining unit that contributes to flowing artesian conditions in the Surrey-Langley area. Aquifer material confined by the clay unit includes sand and gravel (e.g. standardized as Coarse HGU) but also includes finer grained material like silty sand (e.g. standardized as Fines HGU).

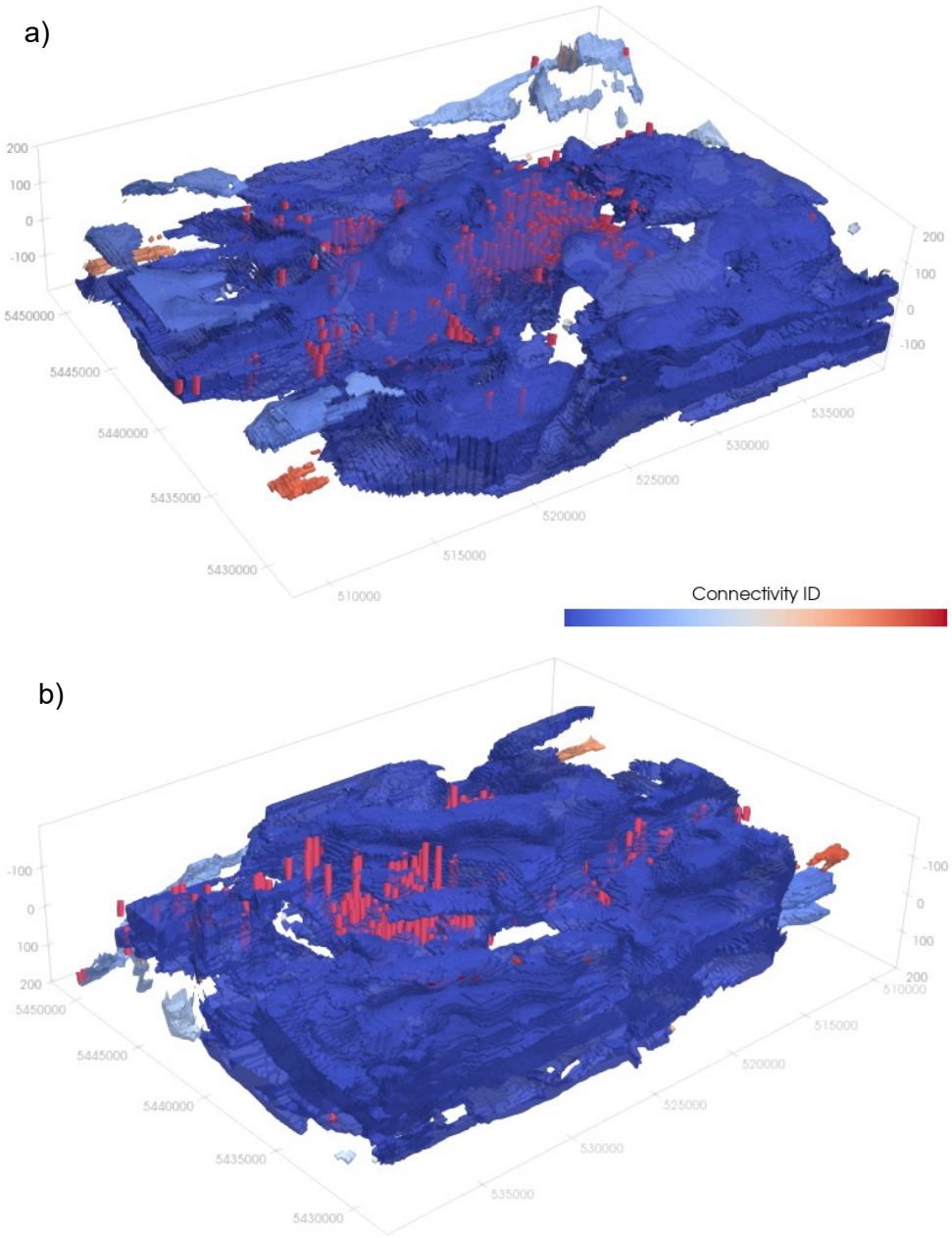


Figure 4.16 **Clay HGU connectivity zones (connectivity ID) and flowing artesian wells (red tubes) for a) top view and b) bottom view of the geologic model.**

Chapter 5.

Conclusions and Recommendations

5.1. Conclusions

The feasibility of using Multilayer Perceptron (MLP) to interpret glacial deposits in the subsurface of the Fraser-Whatcom basin was explored in this study. Three-dimensional block-model representations of the subsurface were constructed using three data processing methods to train MLP. This required development of a geologic database, standardization of geologic material descriptions, and creation of a 3D mesh. The performance of MLP at producing a geologic realization for each data selection alternative was compared and the best outcome verified to further evaluate the suitability of MLP predictions.

A geologic database was developed based on borehole data and surficial mapping. Borehole data sources include publicly available data collections managed by government agencies in both British Columbia and Washington State. This data source included approximately 13,900 wells, 87,500 material descriptions, 210 hydraulic conductivity values, and 2,000 well yield estimates. Several data processing steps were required to clean and standardize available data. Borehole data was augmented with approximately 40,000 data points based on surficial mapping. This point data was extracted using a grid spacing of one (1) km to improve the lateral distribution surficial information where no boreholes exist. Additional delineation of bedrock, both horizontally and vertically, was also provided where outcrops have been mapped.

Material descriptions from borehole data contain the largest source of subsurface information for the study area. Semi-automated methods were used to standardize material descriptions into hydrogeological units (HGUs) for geologic model development. A facies approach was used (instead of stratigraphic layers) to better capture the heterolithic nature and variability of glacial deposits. The ASTM USCS provided a guide to classify material descriptions into material groupings primarily based on grain size. Material groupings were then classified into HGUs based on hydraulic characteristics to reduce complexity for modelling purposes. The standardization method included Natural Language Processing (NLP) techniques to explore the geologic lexicon so that material

classification was more automated, less subjective, and the associated uncertainty better understood. The original 20,000 unique material descriptions were consolidated into the top 45 descriptor combinations and then grouped into 10 categories, capturing 92% of the total thickness from the processed borehole logs.

A mesh was created providing a 3D representation of the study area. It was used to extract information (e.g. cell index, cell centroid coordinates), to assign attributes, and to visualize results. The lateral boundaries of the mesh were based on the extent of Quaternary mapping, excluding the land north of Burrard Inlet and intermountain valleys. The areas outside of these lateral boundaries were masked to generate a non-rectangular shape. A cell size of 200 m wide by 200 m long and 5 m high was used to capture the major variability in HGUs for regional modelling purposes. A uniform vertical cell height of 5 m was established based on a statistical review of HGU thickness and depth. The greatest generalization using this approach occurs near the surface where HGUs tend to be less than 5 m thick. The top of the mesh was generated based on a digital elevation model (DEM) that combined topographic and bathymetric data. The bottom of the mesh was established at an elevation of -150 masl given the vertical extent of HGUs that could act as aquifers, which also corresponds to the deepest extent considered in groundwater modelling studies within the region (Golder 2005). However, there is a disproportionally large number of wells with relatively shallow depths and the deepest wells are generally located in the Nicomekl-Serpentine valleys; as such, the availability of deep subsurface information is limited. The resultant mesh has over 3.3 million cells with a surface area of approximately 670 km^2.

Three data selection alternatives were considered to evaluate the training and prediction capabilities of MLP. Alternative 1 consists of HGUs spatially represented using coordinates from the geologic database. Alternative 2 considers a 3D mesh and uses the cell centroid coordinates and the HGU mode within each cell to provide a more regularly spaced dataset. Like Alternative 2, Alternative 3 uses the HGU mode within each cell but takes the coordinates of the cell centroid and maps them on a 2D grid using Self-Organizing Map (SOM) before being used in the MLP algorithm. The main differences between the datasets includes level of effort to process data, number of samples, frequency distribution of HGUs, and spatial distribution of data. Alternative 1 involved the least amount of processing and has the lowest number of samples (approx. 53,800 samples) that typically under-represent bedrock and clay (due to thick intervals

indicated in the borehole logs) and are more irregularly spaced. Alternative 2 required more data processing effort but resulted in a larger dataset (approx. 85,000 samples) that was more regularly spaced. This dataset has fewer lateral locations but more samples in the vertical direction with a generalized representation of HGUs based on the mode. For Alternative 3, the dataset from Alternative 2 was lumped into spatial clusters with consideration of the entire mesh domain. The greatest amount of effort for data processing was required for Alternative 3 because of hyperparameter tuning and training with SOM prior to using MLP.

MLP was used to make predictions throughout the mesh representing the study area using each data selection alternative. Data preparation (splitting, scaling), hyperparameter tuning, and training was required prior to predictive modelling. A stratified three-fold cross-validation approach was used for hyperparameter tuning to reduce model bias. Once the hyperparameters were established, the MLP was trained using all the training data. The testing data was then used to evaluate the generalization performance of the model (e.g. MLP with optimized hyperparameters and weights) when making predictions on unseen data. The performance metrics used to evaluate the models include log-loss (aka cross-entropy), balanced accuracy, and the confusion matrix.

MLP tries to minimize log-loss using stochastic gradient descent by adjusting weights during training. At the end of training, Alternative 2 has the lowest log-loss at 0.8, followed by Alternative 3 at 1.10 and Alternative 1 at 1.15. The MLP models for all three alternatives achieve a lower log-loss compared to a naïve classification model which assumes the same probability for each HGU (1.3). Learning curves indicate all models have a relatively high bias (based on relatively low balanced accuracy scores ranging from 48% to 65% for training) and low variance (given the relatively small gap between training and validation scores). The models appear to be underfitting the training data. Adding more examples is unlikely to lead to better training scores. For Alternative 1 and Alternative 2, additional data may improve validation scores given that validation curves did not plateau. Runtimes for Alternative 2 and Alternative 3 are similar and generally an order of magnitude higher compared to Alternative 1.

MLP tuning (e.g. hyperparameter optimization, number of epochs) improved training performance compared to the use of default MLP values for all alternatives.

Tuning improved balanced accuracy scores by 9 to 17%. The alternatives for data selection used in by MLP also impacted scoring. Alternative 2 had the highest training and testing scores (65% and 60%, respectively) generally by 10% to 15% more compared to the other alternatives. Model variance (difference between training and testing score) is low, below 5% for all alternatives. In general, the balanced accuracy scores indicate modelling results with a high bias and low variance. Typically training and testing accuracy scores above 80% indicate good performance for other machine learning applications; however, this may not be achievable given the multi-class classification problem and variability in subsurface data for this study.

Based on the confusion matrix results, MLP predicts bedrock the most accurately despite bedrock samples occurring less frequently in the datasets for all alternatives. A high precision for bedrock may be associated with the continuity of bedrock once it is encountered. The prediction performance for unconsolidated HGUs follows the same trend for all alternatives where Coarse has the second highest precision with lower precision in descending order for Clay, Fines, and Till. This may be attributed to the distribution of unconsolidated HGUs or could be associated with discontinuity in the subsurface and the underfitting of the model to capture this complexity. Clay and Till are most often confused with each other while Coarse seems to be confused mainly with Till and Fines. Fines are most confused with Till and Clay.

Once trained, the MLP was used to predict HGUs at scaled coordinates representing the entire study area for each alternative. Cell centroid coordinates from the mesh were used for Alternative 1 and Alternative 2 while the coordinates from 2D SOM mapping were used for Alternative 3. This resulted in the prediction of HGUs at a total of 3,336,175 unique cell centroids and 7,578 unique SOM neurons. Select model outputs (e.g. HGU predictions, coarse HGU probabilities, low permeability probabilities, and entropy) were visualized is cross-sectional view to compare MLP results.

Based on cross-sectional reviews, Alternative 2 shows more complexity in the subsurface compared to Alternative 1 and Alternative 3. Alternative 3 is more comparable to Alternative 2 but is more generalized. This suggests the size of the 2D SOM could be inadequate to represent the subsurface complexity. Generally, Alternative 2 has the largest coverage of Coarse HGU and low permeability (Clay and Till HGUs) areas with probabilities above 75%. Entropy values are generally greater for Alternative

1 and Alternative 3 compared to Alternative 2. Alternative 2 data resulted in the best MLP performance based on better performance metrics and cross-sectional reviews that indicated more complexity in the subsurface.

Verification of MLP predictions using Alternative 2 data was completed by comparing results from other analytical techniques (k-nearest neighbours, KNN) as well as published interpretations of the subsurface and hydrogeologic indicators within the region (e.g. artesian conditions). Predictions using MLP and KNN have comparable results; however, MLP has the added benefit of being stochastic with the potential of generating multiple geologic realizations. The frequency of predicted HGUs is comparable between the two algorithms with differences typically less than 5%. Both algorithms generally reproduce the spatial distribution of HGUs from surficial mapping. KNN can reproduce HGUs that are more localized and results in a more heterogeneous distribution overall. Cross-sectional reviews indicate HGU predictions made by KNN show more variation and are generally more discretized, but MLP predictions recreate the general subsurface trends. MLP results in higher localized values of entropy, but the overall average entropies are similar (0.28 for MLP versus 0.30 for KNN). The uncertainty associated with HGUs predicted at deeper depths for both algorithms are likely underestimated using entropy given the limited distribution of deep subsurface data. In general, both algorithms reproduce similar subsurface trends, but MLP provides a more generalized interpolation while KNN appears to be more discretized and blockier in appearance.

Geologic interpretations within the Township of Langley published by Golder (2005) were also reviewed to verify MLP predictions. This includes comparison of two geologic cross-sections and a 3D representation of major aquifers in the Township of Langley. In general, MLP performs well at representing the subsurface in the areas of the cross-sections with the potential issue of extrapolation at depth. MLP interpolations show more generalized representations of aquifers and greater connectivity of permeable units compared to the Golder cross-sections. The 3D representation of the 18 major aquifers from Golder (2005) was compared to the distribution of Coarse HGU with probabilities above 50% from MLP predictions (50% Coarse HGU). The unconfined aquifers (Brookswood Aquifer [AQ41], Fort Langley Aquifer [AQ41], Hopington AB Aquifer [AQ35], and Abbotsford A [AQ15]) are represented by 50% Coarse HGU; however, some of the linear features extending from the main volume of the Brookswood

[AQ41] and Abbotsford A [AQ15] aquifers interpreted by Golder (2005) are not reproduced. The 50% Coarse HGU mapping shows a more consolidated representation of aquifers that may be plausible given the potential for interconnection noted by Golder (2005). A higher cut-off limit for the probability of Coarse HGU would reduce the extent and hydraulic connectivity of aquifers. Some of the deeper aquifers in Golder (2005) are only partially represented or not reproduced because these aquifer materials have a greater fines content (e.g. silty sand) that are not captured in the Coarse HGU but are locally significant as aquifer materials. The use of a more refined grid and additional surficial geology points may improve MLP performance in reproducing meltwater channels and subsurface complexity. A more detailed review of data within this area would be required to support conceptualization of intertill aquifers given the limited frequency of data points categorized as Till.

The last verification approach focused on the Nicomekl-Serpentine valley to determine if geologic interpretations support the development of known artesian conditions within this area. Geologic cross-sections show upper and lower permeable units consistent with the description provided for local confined aquifers (AQ58 and AQ33). The continuity of the lower permeable unit for AQ58 is typically associated with the Fines HGU which includes subsurface materials described as silty sand. Most flowing wells appear to be screened in confined aquifers; however, several may be screened in unconfined AQ35. The northward and westward sloping topography likely contributes to flowing artesian conditions around unconfined AQ35. A combination of topography and confining conditions may contribute to flowing artesian wells along the western limit of AQ33. Flowing artesian conditions for wells in the lowlands appears to occur because aquifer units are in valley bottoms surrounded by uplands and overlain by low permeability material. The geologic model shows a large clay unit in the Nicomekl-Serpentine valley with variability in both the vertical and horizontal directions. This clay unit intersects the top and bottom of most artesian wells. This suggests MLP predictions adequately represent the confining unit that contributes to flowing artesian conditions in the Nicomekl-Serpentine valley. Aquifer material confined by the clay unit includes sand and gravel (e.g. standardized as Coarse HGU) but also includes finer grained material like silty sand (e.g. standardized as Fines HGU).

Based on the results of this study, MLP appears to be a promising algorithm to solve multi-class classification problems related to modelling complex glacial deposits in

the subsurface. This has the benefit of interpreting subsurface conditions using categorical data instead of numerical information which is typically more readily available in the Fraser-Whatcom basin. This study showed additional processing effort to create a more regular dataset using the HGU mode (Alternative 2) produced better results compared to directly using borehole intervals (Alternative 1). The combination of SOM and MLP (Alternative 3) did not perform the best despite the enhanced pattern recognition anticipated using this approach. Heuristics used to size the SOM grid may have been insufficient to spatially cluster data in a manner to reproduce subsurface complexity.

5.2. Recommendations

Based on the results of this study, the following recommendations are provided:

- Bedrock should be interpreted separately from Quaternary sediments. An updated bedrock contour map for the Fraser-Whatcom Basin would be required, potentially building on work done by Hamilton and Ricketts (1994), Scibek (2005) and the USGS (Eungard 2014). Alternatively, a block-model of bedrock where it exists to a predefined depth (e.g. -100 masl) would be beneficial since insufficient data is available to interpret subsurface conditions overlying bedrock in most locations of the study area.

- This study provides an initial conceptualization of glacial sediments in the subsurface of the Fraser-Whatcom Basin. Additional Quaternary information could be incorporated into the geologic database that may improve predictions. This could include geologic cross-section interpretations from existing surficial mapping, fence diagrams in Halstead (1986), subsurface information from Ricketts (2000), and/or geologic interpretations from local studies. This would be particularly beneficial to constrain interpretations of the deeper units.

- Calculated entropy values likely underestimate the uncertainty of predictions where limited data is available (e.g. bottom of geologic realization). It may be beneficial to combine entropy with another metric

based on data density or proximity to a cell with data to better reflect extrapolation.

- A 3D groundwater numerical model could be developed using geologic realizations from MLP predictions to explore groundwater movement in the Fraser-Whatcom basin. Results could be compared with 3D groundwater numerical modelling results from Simpson (2012). Cross-sectional models could also be used to test concepts of vertical water movement and the hydrogeologic controls on that movement. This could facilitate development of a conceptual groundwater flow model for the region to support water management and to inform boundary conditions of smaller numerical models.

- Groundwater modelling using multiple geologic models is recommended to quantify the sensitivity of groundwater flow to geologic architecture (Poeter et al., 2005; Refsgaard et al., 2012; He et al., 2013, Lukjan, 2016); however the use of multiple geologic models is still not common practice due to a number of technical and economic challenges associated with current workflows (MacMillan, 2017). The suitability of MLP to generate multiple geologic models could be further explored.

- The workflow presented in this study may provide the opportunity to expand on the stratigraphic and geomorphologic understanding of the region by reviewing 3D representation of facies in the subsurface. This could focus on understanding the deposits in the Fraser River valley, extent of thick glaciomarine sediments associated with the Fort Langley Formation, deltas formed at the mouths of mountain valleys, or paleovalleys that have been filled and buried by periods of aggradation.

- Hydrochemical data can be an effective tool to improve the understanding of groundwater flow and possible hydraulic connections resulting from subsurface heterogeneity (Raiber et al., 2012). Cavalcanti de Albuquerque (2011) studied the hydrochemical evolution of groundwater in the Lower Fraser Valley, Canada. Chemistry data (e.g. major ion chemistry, arsenic, isotopes) could be used to fingerprint groundwater

flow systems or to identify where potential connections exist. This could be an additional input for MLP training which may reduce model bias and improve performance.

- Smaller scale studies within the Fraser-Whatcom basin could be used to further evaluate the performance of MLP. This may also help further refine the understanding of geological architecture and the subsequent hydrogeologic framework. The visualization methods used in this study could be used to define aquifer extents (e.g. cut-off limit for probability of HGU), provide scenarios to evaluate uncertainty for sustainable groundwater management (e.g. upper and lower limits for probability of aquifer units based on multimodels), or inform vulnerability mapping (e.g. distribution of low permeability units). Areas with the highest borehole density and supplemental studies to infer continuity in the subsurface would be the best candidates for this (e.g. TOL, Abbotsford).

References

Advisian (2018). City of White Rock Aquifer Protection Plan. Prepared for the City of White Rock. https://www.whiterockcity.ca/DocumentCenter/View/2017/2018-Aquifer-Protection-Plan-PDF?bidId=

Allen, D. M., Schuurman, N., Deshpande, A., & Scibek, J. (2008). Data integration and standardization in cross-border hydrogeological studies: a novel approach to hydrostratigraphic model development. *Environmental Geology*, 53(7), 1441-1453.

ArcGIS Help (n.d.). ArcMap 10.8. Sink, Spatial Analyst toolbox. https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/sink.htm

Anderson, W., Hunt, Woessner, W. W., & Hunt, R. J. (2015). Applied groundwater modeling: Simulation of flow and advective transport. (Second ed.). Academic Press. ISBN 978-0-12-058103-0

Arihood, L. D., (2009). Processing, analysis, and general evaluation of well-driller records for estimating hydrogeologic parameters of the glacial sediments in a ground-water flow model of the Lake Michigan Basin. U.S. Department of the Interior and U.S. Geological Survey, Scientific Investigations Report 2008-5184. ISBN 1-411-32302-5

Armstrong, J. E., (1957). Surficial geology of New Westminister map-area, British Columbia. Geological Survey of Canada, Paper 57-5.

Armstrong, J. E., (1976). Surficial geology, Mission, British Columbia. Geological Survey of Canada, Map 1485A, 1 Sheet, scale 1:50,000.

Armstrong, J. E., (1977). Surficial geology, Chilliwack, British Columbia. Geological Survey of Canada, Map 1487A, 1 Sheet, scale 1:50,000.

Armstrong, J. E. (1981). Post-Vashon Wisconsin Glaciation, Fraser Lowland, British Columbia. Geological Survey of Canada Bulletin. Vol. 322.

Armstrong, J. E. (1984). Environmental and Engineering Applications of the Surficial Geology of the Fraser Lowland, British Columbia. Geological Survey of Canada, Paper 83-23. http://geogratis.ca/download/part6/ess_pubs/210/210364/of_3723.pdf.

Armstrong, J. E., and Hicock, S.R., (1980). Surficial geology, New Westminister, British Columbia. Geological Survey of Canada, Map 1484A, 1 Sheet, scale 1:50,000.

Armstrong, J. E., Hicock, S.R., (1979). Surficial geology, Vancouver, British Columbia. Geological Survey of Canada, Map 1486A, 1 Sheet, scale 1:50,000.

Bacao, F., Lobo, V., and Painho, M. (2005). GEO-SOM and it's integration with Geographic Information Systems. WSOM, Paris. https://www.novaims.unl.pt/labnt/geosom/Public/3-2_16_lobo05_GeoSOM_GIS.pdf

Baykan, N. A., and Yilmaz, N. (2010). Mineral identification using color spaces and artificial neural networks. *Computers & Geosciences* 36, 91-97. doi:10.1016/j.cageo.2009.04.009

Bayless, E.R., Arihood, L. D., Reeves, H. W., Sperl, B.J.S, Qi, S.L., Stipe, V.E., and Bunch, A. R. (2017). Maps and Grids of Hydrogeologic Information Created from Standardized Water-Well Drillers' Records of the Glaciated United States. U.S. Geological Survey Scientific Investigations Report 2015–5105. https://doi.org/doi.org/10.3133/sir20155105.

Bianchi, M., Kearsey, T., and Kingdon, A. (2015). Integrating Deterministic Lithostratigraphic Models in Stochastic Realizations of Subsurface Heterogeneity. Impact on Predictions of Lithology, Hydraulic Heads and Groundwater Fluxes. *Journal of Hydrology* 531. Elsevier B.V.: 557–73. https://doi.org/10.1016/j.jhydrol.2015.10.072.

Bird, S., Loper, E. and Klein, E. (2009). *Natural Language Processing with Python*. O'Reilly Media Inc.

Bergstra, J., Yamins, D., Cox, D. D. (2013). Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. To appear in Proc. of the 30th International Conference on Machine Learning (ICML 2013).

Brownlee, J. (2019a, August 6). How to Improve Deep Learning Performance. machinelearningmastery.com. https://machinelearningmastery.com/improve-deep-learning-performance/

Brownlee, J. (2019b, August 12). Embrace Randomness in Machine Learning. machinelearningmastery.com. https://machinelearningmastery.com/randomness-in-machine-learning/

Brownlee, J. (2020a, January 14). A Gentle Introduction to Imbalanced Classification. machinelearningmastery.com. https://machinelearningmastery.com/what-is-imbalanced-classification/

Brownlee, J. (2020b, March 12). Hyperparameter Optimization with Random Search and Grid Search. machinelearningmastery.com. https://machinelearningmastery.com/hyperparameter-optimization-with-random-search-and-grid-search/

Brownlee, J. (2020c, July 24). Train-Test Split for Evaluating Machine Learning Algorithms. machinelearningmastery.com. https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/

Brownlee, J. (2020d, December 22). A Gentle Introduction to Cross-Entropy for Machine Learning. machinelearningmastery.com. https://machinelearningmastery.com/cross-entropy-for-machine-learning/

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., and Varoquaux, G. (2013). API design for machine learning software: experiences from the scikit-learn project. ECML PKDD Workshop: Languages for Data Mining and Machine Learning, 108-122.

Bustin, R. M. (1990). Stratigraphy, sedimentology, and petroleum source rock potential of the Georgia Basin, southwest British Columbia and northwest Washington State. In Geological Survey of Canada, Current research, Part F: Geological Survey of Canada Paper 90-lF, p. 103-108.

Cavalcanti de Albuquerque, R. (2011). Hydrogeochemical evoluation and arsenic mobilization in confined aquifers formed within glaciomarine sediments [Master's thesis, Simon Fraser University]. http://summit.sfu.ca/item/11698

Cameron, V.J. (1989). The late Quaternary geomorphic history of the Sumas Valley [Master's thesis, Simon Fraser University]. http://summit.sfu.ca/item/4924

Clague, J. J. (1976). Quadra Sand and it relation to the late Wisconsin glaciation of southwest British Columbia. *Canadian Journal of Earth Sciences*, 13, 803-8015.

Clague, J. J. (1986). The Quaternary stratigraphic record of British Columbia - evidence for episodic sedimentation and erosion controlled by glaciation. Canadian Journal of Earth Sciences, 23, 885-894.

Clague, J. J. (1989). Quaternary geology of the Canadian Cordillera. Quaternary geology of Canada and Greenland. Edited by Fulton, R.J., Geological Survey of Canada, no.1, Chap. 1, pp 17-83.

Clague, J. J. (1991). Quaternary Stratigraphy and History of South-Coastal British Columbia. Geology and geological hazards of the Vancouver region, southwestern British Columbia, by Monger, J. W. H. (ed.), Geological Survey of Canada Bulletin 481, 181–92.

Clague, J. J. (2000). Recognizing Order in Chaotic Sequences of Quaternary Sediments in the Canadian Cordillera. *Quaternary International* 67–71, 29–38. https://doi.org/10.1016/S1040-6182(00)00006-9.

Clague, J. J., Mathewes, R. W., Guilbault, J. P., Hutchinson, I., and Ricketts. B. D. (1997). Pre-Younger Dryas resurgence of the southwestern margin of the Cordilleran ice sheet, British Columbia, Canada. *Boreas*, 26, 261 278. Oslo. ISSN 0300-9483.

Clague, J. J., Luternauer, J. L. (1983a). Field trip guidebook: late Quaternary geology of southwestern Brtish Columbia. Geological Association of Canada, Mineralogical Association of Canada - Canadian Geophysical Union Joint Annual Meeting.

Clague, J. J., Luternauer, J. L., and Hebda, R. J. (1983b). Sedimentary environments and postglacial history of the Fraser Delta and lower Fraser Valley, British Columbia. *Canadian Journal of Earth Sciences*, 20, 1314-1326.

Cracknell, M. J., and Reading, A. M. (2014). Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computer & Geosciences*, 63, 22-33. http://dx.doi.org/10.1016/j.cageo.2013.10.008

Cummings, Don I., Russell, H. A. J., and Sharpe, D.R. (2012). Buried-Valley Aquifers in the Canadian Prairies: Geology, Hydrogeology, and Origin. Earth Science Sector (ESS) Contribution 20120131. *Canadian Journal of Earth Sciences,* 49(9), 987–1004. https://doi.org/10.1139/e2012-041.

DMTI Spatial Inc. (2002). Digital Elevation Model (DEM). NAD83, UTM projection, 30 m resolution from 1:50,000 scale digital mapping. Accessed through SFU Library.

Dunn, D. and Ricketts, B. (1994). Surficial geology of Fraser Lowlands digitized from GSC Maps 1484A, 1485A, 1486A, and 1487A (92 G/1, 2, 3, 6, 7; 92 H/4). Geological Survey of Canada, Open File 2894. https://doi.org/10.4095/194084

Easterbrook, D. J. (1963). Late Pleistocene Glacial Events and Relative Sea-Level Changes in the Northern Puget Lowland, Washington. Geological Society of America Bulletin (74), 1465-1484).

Easterbrook, D. J. (1986). Stratigrpahy and chronology of Quaternary deposits of the Puget Lowland and Olympic Mountains of Washington and the Cascade Mountains of Washington and Oregon. *Quaternary Science Reviews* (5), 183-196.

Eungard, D. W. (2014). Models of Bedrock Elevation and Unconsolidated Sediment Thickness in Puget Lowland, Washington. Open-File Report 2014-04.

Finlayson, D. P., Haugerud, R. A., Greenberg, H. and Logsdon, M.G. (2000). Puget Sound Digital Elevation Model. University of Washington https://www.ocean.washington.edu/data/pugetsound/asciigrid.html

Forest, F., Lebbah, M., Azzag, H., and Lacaille, J. (2020). A survey and implementation of performance metrics for Self-Organizing Maps. arXiv:2011.05847v1 [cs.NE] https://doi.org/10.48550/arXiv.2011.05847

Freeze, R. A. and Cherry, J. A. (1979). Groundwater. Englewood Cliffs, N.J: Prentice-Hall. http://hydrogeologistswithoutborders.org/wordpress/original-groundwater-by-freeze-and-cherry-1979-now-available-online/

Frind, E. O., Molson, J. W., Sousa, M. R., and Martin, P. J. (2014). Insights from Four Decades of Model Development on the Waterloo Moraine: A Review. *Canadian Water Resources Journal* 39 (2). 149–66. https://doi.org/10.1080/07011784.2014.914799.

Fuentes, I., Padarian J., Iwanaga, T., and Vervoort, R.W. (2020). 3D lithological mapping of borehole descriptions using word embeddings. *Computer & Geosciences* 141 104516. https://doi.org/10.1016/j.cageo.2020.104516

Golder Associates (Golder) (2005, June). Comprehensive Groundwater Modelling Assignment, Township of Langley. Prepared for the Township of Langley.

Golder Associates (Golder) (1997, November). Groundwater Protection Plan District of Chilliwack. Prepared for the District of Chilliwack.

Haj, A. E., Soller, D. R., Reddy, J. E., Kauffman, L. J., Yager, R. M., and Buchwald, C. A.(2018). Hydrogeologic framework for characterization and occurrence of confined and unconfined aquifers in quaternary sediments in the glaciated conterminous United States—A digital map compilation and database. U.S. Geological Survey Data Series 1090. https://doi.org/10.3133/ds1090

Halstead, E. C. (1986). Ground water supply - Fraser Lowland, British Columbia. Scientific series, Canada Inland Waters Directorate, no. 145. NHRI paper no. 26.

Hamilton, T. S. and Ricketts, B. D. (1994). Contour Map of the Sub-Quaternary Bedrock Surface, Strait of Georgia and Fraser Lowland. Bulletin - Geological Survey of Canada, 193–96.

Harris, C. R., Millman, K. J., van der Walt, S. J.,…Oliphant, T. E. (2020). *Array programming with NumPy*. Nature 585, 357–362 (2020). DOI: 10.1038/s41586-020-2649-2

He, X., Sonnenborg, T. O., Jørgensen, F., Høyer, A. S., Møller, R. R., and Jensen, K. H. (2013). Analyzing the effects of geological and parameter uncertainty on prediction of groundwater head and travel time. *Hydrology and Earth System Sciences,* 17, 3245-3260. doi:10.5194/hess-17-3245-2013

Hicock, S. R., and Armstrong, J. E. (1985). Vashon Drift: definition of the formation in the Georgia Depression southwest British Columbia. Canadian Journal of Earth Sciences, 22, 748-757.

Hicock, S. R. and Lian, O. B. (1999). Cordilleran Ice Sheet lobal interactions and glaciotectonic superposition through stadial maxima along a mountain front in southwestern British Columbia, Canada. *Boreas*, 28, 531-542.

Hicock, S. R., Lian, O. B., and Mathews, R. W. (1999). 'Bound cycles' recorded in terrestrial Pleistocene sediments of southwestern British Columbia, Canada. *Journal of Quaternary Science*, 14, 443-449.

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3), pp. 90-95

Johnson, B., Allen, D. M., and Wei, M. (2022). Mapping the Likelihood of Flowing Artesian Conditions in the Okanagan Basin and Fraser Valley, British Columbia. Water Science Series. Province of British Columbia, Victoria.

Jones, M.A. (1999). Geologic Framework for the Puget Sound Aquifer System, Washington and British Columbia. U.S. Geological Survey. http://www.getcited.org/pub/100328685.

Jordahl K., Van den Bossche J., Fleischmann M., Wasserman J., McBride J., Gerard, J., … Leblanc, F. (2020, July 15). geopandas/geopandas: v0.8.1 (Version v0.8.1). Zenodo. http://doi.org/10.5281/zenodo.3946761

Jørgensen, F., Høyer, A. S., Sandersen, P.B.E, He, X., and Foged, N. (2015). Combining 3D Geological Modelling Techniques to Address Variations in Geology, Data Type and Density – An Example from Southern Denmark. *Computers & Geosciences*, 81, 53–63. https://doi.org/10.1016/J.CAGEO.2015.04.010.

Kanevski, M., Pozdnoukhov, A., & Timonin, V (1st Edition). (2009). *Machine Learning for Spatial Environmental Data*. EPFL Press. https://doi.org/10.1201/9781439808085

Kearsey, T., Williams, J., Finlayson, A., Williamson, P., Dobbs, M., Marchant, B., Kingdon, M., and Campbell, D. (2015). Testing the Application and Limitation of Stochastic Simulations to Predict the Lithology of Glacial and Fluvial Deposits in Central Glasgow, UK. *Engineering Geology,* 187, 98–112. https://doi.org/10.1016/j.enggeo.2014.12.017

Koch, J., He, X., Jensen, K. H., and Refsgaard, J. C. (2014). Challenges in Conditioning a Stochastic Geological Model of a Heterogeneous Glacial Aquifer to a Comprehensive Soft Data Set. Hydrology and Earth System Sciences, 18, 2907–23. https://doi.org/10.5194/hess-18-2907-2014.

Kovanen, D. J. (2002). Morphologic and stratigraphic evidence for Allerod and Younger Dryas age glacier fluctuations of the Cordilleran Ice Sheet, British Columbia, Canada and Northwest Washington, U.S.A. *Boreas*, 31, 163-184. Oslo. ISSN 0300-9483.

Kovanen, D. J., and Easterbrook, D. J. (2002). Timing and extend of Allerod and Younger Dryas Age (ca. 12,500-10,000 14C yr B.P.) Oscillations of the Cordilleran Ice Sheet in the Fraser Lowland, Western North America. *Quaternary Research*, 57, 208-224.

Kovanen, D. J. and Slaymaker O. (2015). The paraglacial geomorphology of the Fraser Lowland, southwest British Columbia and northwest Washington. *Geomorphology*, 232, 78-93. http://dx.doi.org/10.1016/j.geomorph.2014.12.021

Lien, J. M., Liu, G., and Langevin, C. D. (2015). GRIDGEN version 1.0 -- A computer program for generating unstructured finite-volume grids. U.S. Geological Survey Open-File Report 2014-1109. http://dx.doi.org/10.3133/ofr20141109.

Lukjan, A., Swasdi, S., Chalermyanont, T. (2016). Importance of Alternative Conceptual Model for Sustainable Groundwater Management of the Hat Yai Basin, Thailand. *Procedia Engineering,* 154, 308–16. https://doi.org/10.1016/j.proeng.2016.07.480

Maimon, O. and Rokach, L. (2010). Data Mining and Knowledge Discovery Handbook (2nd Edition). Springer Science+Business Media. DOI 10.1007/978-0-387-09823-4

Mathews, W. H., Fyles, J. G., and Nasmith, H. W. (1970). Postglacial crustal movements in southwestern British Columbia and adjacent Washington State. *Canadian Journal of Earth Sciences*, 7, 690-702.

MacMillan, G. J. (2017). Potential Use of Multimodels in Consulting to Improve Model Acceptance and Decision Making. *Groundwater,* 55(5), 635–40. https://doi.org/10.1111/gwat.12559

Meyer, R., Engesgaard, P., Høyer, H. S., Jørgensen, F., Vignoli, G. and Sonnenborg, T. O. (2018). Regional Flow in a Complex Coastal Aquifer System: Combining Voxel Geological Modelling with Regularized Calibration. *Journal of Hydrology,* 562, 544–63. https://doi.org/10.1016/j.jhydrol.2018.05.020

Ministry of Forests, Lands, Natural Resource Operations and Rural Development (FLNRORD). (2018, April). Well Drilling Advisory Flowing Artesian Conditions: Surrey and Langely, BC.

Monger, J. W. H., and Journeay, J. M. (1994). Basement Geology and Tectonic Evolution of the Vancouver Region. Geology and geologic hazards of the Vancouver region, southwestern British Columbia; by Monger, J.W.H (ed.), Geological Survey of Canada, Bulletin 481, pp 3-25. https://doi.org/10.4095/203245

Morgan, S. E. (2018). Investigating the role of buried valley aquifer systems in the regional hydrogeology of the Central Peace Region in Northeast British Columbia [Master's thesis, Simon Fraser University]. http://summit.sfu.ca/item/17906

Mustard, P. S. and Rouse, G. E. (1994). Stratigraphy and evolution of Tertiary Georgia Basin and subjacent Upper Cretaceous sedimentary rocks, southwestern British Columbia and northwestern Washington State. Geology and geologic hazards of the Vancouver region, southwestern British Columbia; by Monger, J.W.H (ed.), Geological Survey of Canada, Bulletin 481, 97-169.

Padarian, J. and Fuentes, I. (2019). Word embeddings for application in geosciences: development, evaluation, and examples of soil-related concepts. *Soil*, 5, 177-187. https://doi.org/10.5194/soil-5-177-2019

Pandas Development Team (2020, February). Pandas-dev/pandas: Pandas. Zenodo. DOI: 10.5281/zenodo.3509134

ParaView Developers (2022). ParaView Users Guide Documentation (Release 5.10.1). https://docs.paraview.org/_/downloads/en/v5.10.1/pdf/

Pasanen, A. H. and Okkonen, J. S. (2017). 3D Geological Models to Groundwater Flow Models: Data Integration between GSI3D and Groundwater Flow Modelling Software GMS and FeFlow®. Geological Society, London, Special Publications 408 (1): 71–87. https://doi.org/10.1144/SP408.15.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research (12), 2825-2830.

Pirot, G. (2017). Using Training Images to Build Model Ensembles with Structural Variability. *Groundwater,* 55(5), 656–59. https://doi.org/10.1111/gwat.12556.

Poeter, E. and Anderson, D. (2005). Multimodel Ranking and Inference in Ground Water Modeling. *Ground Water,* 43(4), 597–605. https://doi.org/10.1111/j.1745-6584.2005.0061.x

Province of British Columbia (2016a, July 29). Aquifer Classification Worksheet (Aquifer Reference Number: 33). https://s3.ca-central-1.amazonaws.com/aquifer-docs/00000/AQ_00033_Aquifer_Mapping_Report.pdf

Province of British Columbia (2016b, July 29). Aquifer Classification Worksheet (Aquifer Reference Number: 58). https://s3.ca-central-1.amazonaws.com/aquifer-docs/00000/AQ_00058_Aquifer_Mapping_Report.pdf

Raiber, M., White, P. A., Daughney, C. J., Tschritter, C., Davidson, P., and Bainbridge, S. E. (2012). Three-Dimensional Geological Modelling and Multivariate Statistical Analysis of Water Chemistry Data to Analyse and Visualise Aquifer Structure and Groundwater Composition in the Wairau Plain, Marlborough District, New Zealand. Journal of Hydrology, 436–437, 13–34. doi:10.1016/j.jhydrol.2012.01.045

Refsgaard, J. C., Christensen, S., Sonnenborg, T. O., Seifert, D., Højberg, A. L. and Troldborg, L. (2012). Review of Strategies for Handling Geological Uncertainty in Groundwater Flow and Transport Modeling. *Advances in Water Resources,* 36, 36–50. doi.org/10.1016/j.advwatres.2011.04.006

Ricketts, B. D. (Edited by) (2000). Mapping, Geophysics, and Groundwater Modelling in Aquifer Delineation, Fraser Lowland and Delta, British Columbia. Geological Survey of Canada, Bulletin 552.

Riedel, J. L. (2017). Deglaciation of the North Cascade Range, Washington and British Columbia, from the Last Glacial Maximum to the Holocene. *Cuadernos de Investigación Geográfica,* 43(2), 467-496. https://doi.org/10.18172/cig.3236.

Rizzo, D. M. and Dougherty, D. E. (1994). Characterization of aquifer properties using artificial neural networks: Neural kriging. *Water Resources Research*, 30(2), 483-497. https://doi.org/10.1029/93WR02477

Russell, H. A. J., Brennand, T. A., Logan, C., and Sharpe, D. R. (1998). Standardization and assessment of geological descriptors from water well records, Greater Toronto and Oak Ridges Moraine Areas, Southern Ontario. Current Research 1998-E, Geological Survey of Canada, p. 89-102.

Russell, H. A. J., Crow, H., Hinton, M., Oldenborger, G., Paradis, D., Pugin, A. J. M., and Sharpe, D. R. (2013). Three Dimensional Mapping for Groundwater Applications At the Geological Survey of Canada. 2011- 2013 Developments, 2011–13.

Sahoo, S. and Jha, M. K. (2017). Pattern recognition in lithology classification: modeling using neural networks, self-organizing maps and genetic algorithms. *Hydrogeology Journal*, 25, 311-330.

Scibek, J. (2005). Modelling the impacts of climate change on groundater: A comparative study of two unconfined aquifers in southern British Columbia and northern Washington State [Master's thesis, Simon Fraser University]. http://summit.sfu.ca/item/10289

Scibek, J., and Allen, D. (2006, February). Groundwater Sensitivity to Climate Change : Abbotsford-Sumas Aquifer in British Columbia , Canada and Washington State , US. Prepared for BC Ministry of Water, Land and Air Protection and Environment Canada.

Simpson, M.W.M. (2012). Assessing Risk to Groundwater Quality Using an Integrated Risk Framework [Master's thesis, Simon Fraser University]. http://summit.sfu.ca/item/12333

Standard, A.S.T.M. (2011) D2487-11 Standard Practice for Classification of Soils for Engineering Purposes (Unified Soil Classification System). ASTM International, West Conshohocken, PA.

Sullivan, C. B., and Kaszynski, A. (2019). PyVista: 3D plotting and mesh analysis through a streamlined interface for the Visualization Toolkit (VTK). *Journal of Open Source Software*, 4(37), 1450, https://doi.org/10.21105/joss.01450

SuperDataScience https://www.superdatascience.com/blogs/the-ultimate-guide-to-self-organizing-maps-soms accessed January 20, 2022.

Thorson, R. M. (1980). Ice-sheet glaciation of the Puget Lowland, Washington, during the Vashon Stade (Late Pleistocene). *Quaternary Research*, 13, 303-321.

Toth, A., Havril, T., Simon, S., Galsa, A., Santos, F. A. M., Muller, I., and Madl-Szonyi, J. (2016). Groundwater Flow Pattern and Related Environmental Phenomena in Complex Geologic Setting Based on Integrated Model Construction." *Journal of Hydrology,* 539, 330–44. https://doi.org/10.1016/j.hydrol.2016.05.038

Vaccaro, J. J. (1992). Plan of Study for the Puget-Willamette Lowland Regional Aquifer System Analysis, Western Washington and Western Oregon. U.S. Geological Survey Water-Resources Investigations Report 91-4189.

Vaccaro, J. J., Hansen, A. J., and Jones, M. A. (1998). Hydrogeologic Framework of the Puget Sound Aquifer System. Professional Paper. Vol. 1424–D. https://pubs.usgs.gov/pp/1424d/report.pdf.

Vettigli, G. (2018). MiniSom: minimalistic and NumPy-based implementation of the Self Organizing Map. https://github.com/JustGlowing/minisom/

Ward, B. C. and Thomson, B. (2004). Late Pleistocene stratigraphy and chronology of lower Chehalis River valley, southwestern British Columbia: evidence for a restricted Coquitlam Stade. *Canadian Journal of Earth Sciences*, 41, 881-895.

Washington Division of Geology and Earth Resources (2016). Surface geology, 1:100,000--GIS data. Washington Division of Geology and Earth Resources Digital Data Series DS-18, version 3.1, previously released June 2010.

Winston, R. B. (2019). ModelMuse (version 4.1).U.S. Geological Survey Software Release, 11 October 2019.

Yager, R. M., Kauffman, L. J., Soller, D. R., Haj, A. E., Heisig, P. M., Buchwald, C. A., Westenbroek, S. M., and Reddy, J. E. (2019). Characterization and occurrence of confined and unconfined aquifers in Quaternary sediments in the glaciated conterminous United States (ver. 1.1). U.S. Geological Survey Scientific Investigations Report 2018–5091. https://doi.org/10.3133/sir20185091

Yang, A. (2018, January 3). Tutorial: Learning Curves for Machine Learning in Python. Dataquest.io. https://www.dataquest.io/blog/learning-curves-machine-learning/