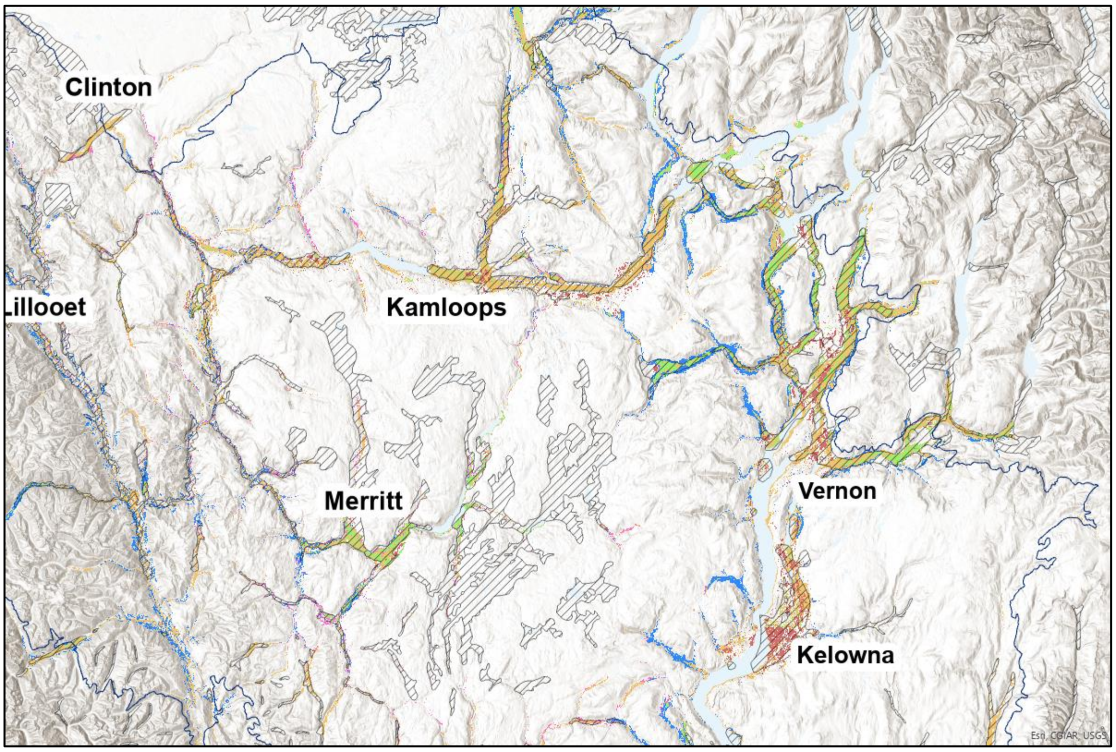
WATER SCIENCE SERIES

Surficial Mapping British Columbia's Landforms to Support Groundwater Understanding: Phase 2 Southern Interior Study Area Pilot

Elyse Sandl, Deepa Filatow and Jennifer Todd



December 2024

No. 2024-14

The **Water Science Series** are scientific technical reports relating to the understanding and management of B.C.'s water resources. The series communicates scientific knowledge gained through water science programs across B.C. government, as well as scientific partners working in collaboration with provincial staff. For additional information visit: http://www2.gov.bc.ca/gov/content/environment/air-land-water/water/water-science-data/water-science-series.

ISBN: 978-1-0399-0123-0

Citation:
Sandl, E., D. Filatow and J. Todd, 2024. Surficial Mapping British Columbia's Landforms to Support Groundwater Understanding, Phase 2: Southern Interior Study Area Pilot. Water Science Series, WSS2024-14. Province of British Columbia, Victoria.

Author's Affiliation:
Elyse M. Sandl, PGeo
Ministry of Water Land and Resource Stewardship
200 – 10428 153rd Street, Surrey BC, V3R 1E1

Deepa Filatow, PGeo
Ministry of Water Land and Resource Stewardship
200 – 1690 Powick Rd. Kelowna, BC, V1X 7G5

Jennifer Todd, PGeo
Ministry of Water Land and Resource Stewardship
4th Floor, 525 Superior Street, Victoria BC, V8V 0C5

Cover Image:
Elyse Sandl and Michael Chan. 2024. Predictive mapping results for Landform Subclass in the Okanagan area.

Acknowledgements

We'd like to send appreciation to thank Madrone Environmental for their support with developing the classification system and the training data set, to Hunter Gleeson and Alex Bevington for their R support, Michael Chan for GIS contributions, to Klaus Rathfelder and Cory McGregor for peer review and editorial support, and to the Groundwater Science fund for providing funding.

<u>**EXECUTIVE SUMMARY**</u>

Knowledge of British Columbia's (B.C.) complex surficial geology is crucial to understanding groundwater and surface water resources throughout the province. Information about landforms and surficial materials is collected in various inventory types including soil surveys, surficial geology, terrain mapping, well logs, and aquifer mapping. Provincially, these inventories have spatial gaps and overlaps and have been collected at various scales and purposes. B.C. currently lacks a consistent, and authoritative data layer representing surficial materials and landforms that covers the province at an appropriate scale to support current data needs. This study aims to fill a scale and information gap for provincial and regional level prioritization and decision making.

The overall intent of this project is to create a harmonized, categorical classification for substrate and landform information. In addition, to generate continuous spatial maps to increase the utility of substrate and landform information for decision making in B.C., we've focussed on identifying places with thick surficial deposits and accumulations of unconsolidated sediments with high aquifer potential, such as in valley bottoms.

During this second phase of the project, a classification system was developed that includes two levels of *Broad Landform* classes (Figure 1). The first level was developed specifically to meet the needs of the project which is intended to divide the landscape into generalized regions and to improve upon the provincial Quaternary Geology layer (BC Geological Survey 2023). The broad landform class currently includes two categories, *Eroding Upland* and *Valley Fill*. We describe *Eroding Upland* as continuous areas that are dominated by erosional, bedrock-controlled landforms, typically consisting of mountainous terrain and extensive uplands and plateaus. We describe *Valley Fill* as continuous areas of sediment accumulation in the landscape such as major valleys and lowlands dominated by constructional landforms such as terraces, planes, and hummocky glacial deposits. At the second level of the classification, the valley fill category is further divided into seven classes such as texture and aquifer type. Many of the classes were adapted from Howes and Kenk (1997).

The Southern Interior EcoProvince was selected as the study area for piloting a modelling approach to generating predictive maps. We took a simple modelling approach using a random forest model to generate preliminary maps of each broad landform class.

Training points were randomly generated within five mapsheet areas in the Clinton, Manning Park and Westwold areas (Figure 2). The training points were attributed by a terrain expert with a categorical value from each surficial material class, based on aerial imagery, elevation contours, information about the geological and glacial history of the area. The random forest models were used to identify the relationships between a surficial material class and 31 covariate raster layers. For each class, two models were developed, one with the original training points and a second where the number of training points within each category was synthetically balanced.

Internal model metrics provide an estimate of model performance within the five mapsheets with training points and do not provide any information about model performance in other parts of the study area. Overall, within the training point mapsheet areas, the models had high internal accuracy rates (98-99%) and balancing the training data consistently improved model performance (within the mapsheet areas). Each "balanced model" was used to create a predictive raster map of the study area. An independent data set is needed to validate the predictive maps. As the project progresses towards model optimization and model finalization during the next phase, it is recommended that both the training dataset and a validation dataset are developed across the study area.

The training point distribution did not sufficiently capture categories that have a small spatial extent (e.g. organic and fine texture material). These ‘minority’ classes, which are represented by a small sample of training points, are difficult for the models to interpret. Strategically capturing minority features during training point development would benefit predictive modelling. It is recommended that the next phase of training point development is designed to create more balanced training data. For example, the mapsheets did not cover areas with glaciolacustrine deposits, which were therefore not included in the training data or the models.

Outside of the five mapsheets with training data, across the rest of the study area, model performance was evaluated qualitatively by visually assessing the predicted classification distributions compared to expected characteristics of the surficial materials in that area. Predictions of valley fill over the study area agree with major valley systems. Most of the model predictions agree with aquifer and quaternary mapped layers and provide a greater level of detail in some areas. The predicted broad landform map is helpful in identifying areas with significant sediment accumulations, including areas with aquifer potential where more detailed aquifer mapping or exploration could be conducted. The eroding upland category did well overall, although it is less representative in the Thompson Plateau. In this area upland depositional landscapes are common, generally hummocky glacial deposits such as kame/esker, drumlinized till, and kame/kettle topography. A third Broad Landform category for upland depositional features is recommended for the next phase of the project.

As the terrain and landscape within the Southern Interior EcoProvince is variable, training data from the five mapsheets did not sufficiently capture all of the landscape characteristics of the study area. Some of the predictive maps did not match professional knowledge in areas that had different characteristics than the training point mapsheets. This modelling pilot revealed that when generating predictive maps, it is important to be mindful of capturing the variability of the covariate distribution with your training data and of extending predictions beyond the rule shed or areas represented by the training data (i.e. do not extend you predictions to areas with underlying characteristics that are inconsistent with the training areas).

Attribution of training points and random forest modelling is a cost-effective way to map simple discrete categories that are discernable from image interpretation. We recommend that future phases of the project adopt a more sophisticated modeling approach that includes K-fold cross validation, sensitivity analyses, and model optimization. As many of the covariate rasters are highly correlated, it is recommended that covariate reduction techniques such as variance inflation factor (VIF) and recursive feature elimination (RFE) are applied during future modelling.

Finally, in future phases of the project the classification and predictive mapping approach can be rolled out across the province to create consistent surficial material mapping layers for BC.

CONTENTS

EXECUTIVE SUMMARY II
1. BACKGROUND 1
   1.1 Key Findings from Phase 1 1
   1.2 Purpose and Objective of Phase 2 Study 2
2. METHODOLOGY 3
   2.1 Surficial Material Classification System 3
   2.2 Study Area and Training Point Development 5
   2.3 Co-variate Raster Stack Development 6
   2.4 Predictive Mapping with Machine Learning Model 7
      2.4.1 Training Classification Matrix 7
      2.4.2 Data Balancing 7
      2.4.3 Random Forest Modelling 7
3. RESULTS AND DISCUSSION 10
   3.1 Surficial Material Classification System 10
   3.2 Training Point Development 10
   3.3 Co-variate Raster Stack Development 12
   3.4 Predictive Mapping with Machine Learning Models 14
      3.4.1 Broad Landform 15
      3.4.2 Landform Subclass 20
      3.4.3 Surficial Material 24
      3.4.4 Surface Expression 27
      3.4.5 Aquifer Type 30
      3.4.6 Thickness Class 35
      3.4.7 Material Texture and Sorting Class 37
4. LEARNINGS AND RECOMMENDATIONS 40
5. CONCLUSIONS 42
REFERENCES 43
APPENDIX A: MADRONE METHODS SUMMARY 45
APPENDIX B: COVARIATE RASTER DESCRIPTIONS 46
APPENDIX C: DOMAIN TABLES 47

FIGURES

Figure 1: Summary of classes and categories within each class that were captured by training points in the Phase 2 study area. A full list of categories is included in Appendix C. 4
Figure 2: Phase 2 training points mapsheet areas and study area encompassing the Southern Interior EcoProvince. 5
Figure 3: Conceptual flow chart presenting the steps associated with predictive modelling and the progress that has been made in the Southern Interior to date. 9
Figure 4: Map of training points locations. 11
Figure 5: Visual representation of degree of correlation between covariate rasters. 12
Figure 6: Ranked predictor importance for the Broad Landform balanced model using the Mean Decrease in Accuracy and Mean Decrease in Gini metrics. 13
Figure 7: Broad Landform predictions across the Southern Interior. 16
Figure 8: Broad Landform predictions compared to Quaternary Geology polygons in the Thompson Plateau and northern Nicola Valley areas to the North and East of Merritt respectively. 17
Figure 9: Broad Landform predictions compared to Quaternary Geology polygons near Shuswap Lake and Armstrong, to the East of Kamloops. 18
Figure 10: Broad Landform predictions compared to Quaternary Geology polygons in the South Okanagan areas. 19
Figure 11: Landform Subclass predictions across the Southern Interior. 21
Figure 12: Landform Subclass predictions compared to Quaternary Geology polygons in the Kamloops and Shuswap Lake areas. 22
Figure 13: Landform Subclass predictions compared to Quaternary Geology polygons in the Kelowna and Vernon areas. 23
Figure 14: Surficial Material predictions across the Southern Interior. 25
Figure 15: Surficial Material predictions in the Kelowna area. 26
Figure 16: Surficial Expression predictions across the Southern Interior. 28
Figure 17: Surficial Expression predictions along the Fraser River and Thompson River valleys north of Lytton. 29
Figure 18: Summary of aquifer type and subtype classification results and modifications recommended for Phase 3 of the project. 30
Figure 19: Aquifer Type predictions across the Southern Interior. 32
Figure 20: Aquifer Type predictions compared to mapped aquifer polygons in the northwest portion of the study area, including the Thompson Plateau. 33
Figure 21: Aquifer Type predictions compared to mapped aquifer polygons near Kamloops, Shuswap Lake and Armstrong. 34
Figure 22: Material thickness predictions across the Southern Interior. 36
Figure 23: Sorting Class predictions across the Southern Interior. 38
Figure 24: Material Texture predictions across the Southern Interior 39
Figure 25: Distribution of Multiscale Topographic Position Index Covariate (MSTPI_WIDE) values across the study area (top) and across the training points (bottom). 42

TABLES

Table 1: Number of training points attributed in each category of a class. 10
Table 2: Summary of model outputs for balanced and unbalanced models of each class. 14

ACRONYMS

AHM	Annual Heat-Moisture Index
BCSIS	British Columbia Soil Information System
CAREA	Top-Down Flow Accumulation
CNBL	Channel Network Base Level
CPLAN	Plan Curvature
CPROF	Profile Curvature
DAH	Diurnal Anisotropic Heating
DEM	Digital Elevation Model
HACH	Height Above Channel
HILL	Hillshade
MAT	Mean Annual Temperature
MAP	Mean Annual Precipitation
MCMT	Mean Coldest Month Temperature
MRRTF	Multiresolution Index of Ridge Top Flatness
MRVBF	Multiresolution Index of Valley Bottom Flatness
MSP	Mean Annual Summer Precipitation
MSTPI	Multi-scale Topographic Position Index
MWMT	Mean Warmest Month Temperature
NOPEN	Negative Topographic Openness
OOB	Out-Of-Bag Error
PCC	Percent Correctly Classified
POPEN	Positive Topographic Openness
SHM	Summer Heat-Moisture Index
SMOTE	Synthetic Minority Oversampling Technique
TD	Temperature Difference
TOPOWET	SAGA Wetness Index
TPI	Topographic Position
TRI	Topographic Ruggedness Index
VIF	Variance Inflation Factor

1. BACKGROUND

British Columbia (B.C.) lacks mid-scale coverage of spatially accurate substrate and landform information, even though B.C. has the most variable geology, surficial materials, and landforms in Canada. This variety is a driver of the wealth of natural resource values in the province including forestry, mining, biodiversity, water resources, wildlife habitat, agriculture, as well as First Nations traditional land and water uses. Substrate and landform information is often neglected in analysis and planning due to the lack of provincial and regional coverage of consistent, reliable data.

This study aims to fill scale and information gaps for provincial and regional level prioritization and decision making. Current substrate and landform datasets that cover the province have poor spatial accuracy and polygon sizes that are too generalized for provincial and regional spatial analysis. For example, the spatial layer *Geology Quaternary Alluvium and Cover* in the BC Geographic Warehouse (the Quaternary Geology layer) was derived through bedrock mapping and provides a high-level representation of areas where thick quaternary sediments cover the bedrock. This data source is not detailed or spatially refined, yet it remains one of our most ubiquitous data sets on surficial sediments in B.C. This results in the use of these datasets outside their intended and appropriate use.

Existing detailed inventories of substrate and landform information have spatial gaps and overlaps, inconsistencies between datasets, and use a variety of classification schemes that together make the compilation of these inventories challenging. Open and complex legends are not conducive to GIS analysis and use outside of a small community of experts in the data and classification. The classification developed under this project provides a closed (categorical) legend that describes landforms and substrate (surficial material) complexes in B.C. by identifying repeating patterns of enduring features in the landscape and defining the surficial materials, landforms, and processes that dominate these complexes. The classification is therefore conducive to machine learning and GIS analysis.

The classification developed in this project also aims to find the “common ground” in the existing geomorphology related datasets in the province to support Provincial decision making and priority setting. The resulting datasets will be relevant to land-use planning and natural resource management relating to climate change strategy, water resources, environmental assessment, habitat management, inventory and gap filling, monitoring, cumulative effects, extreme events and natural disasters, and land use planning.

1.1 Key Findings from Phase 1

Phase 1 of this study (Todd and Filatow, 2024) yielded a harmonized surficial material legend, review of a variety of surficial geology related mapping products, and insights related to the challenges of consolidating multiple mapping projects into a single map. By considering and addressing the following limitations identified in Phase 1, it will be possible to improve surficial mapping results in future studies:

1. Creating a surficial geology map by overlaying available surficial material maps (e.g. terrain, soils, and surficial geology maps) proved to be challenging and inefficient due to disagreement in surficial materials, polygon boundaries, mapping conventions, and project purposes
2. Attempts at creating rules related to scale, year and purpose had significant exceptions and yielded unreliable and inconsistent results.
3. The simplified mapping methodology, which included producing new linework, proved slow and repetitive of mapping that had been completed in the past.

4. The Vanderhoof Watershed Group (the study area from Phase 1) was a challenging pilot area as the surficial materials consisted of thick valley fill with low relief, which can be difficult to map using air photo interpretation alone.
5. Using multiple and patchy datasets proved to be time consuming and inefficient.

1.2 Purpose and Objective of Phase 2 Study

The purpose of Phase 2 was to define a surficial materials classification that enables the prediction of discrete categories using expert interpreted points and a machine learning mapping approach. The overall goal was to produce consistent 25m x 25m raster predictive maps with reported accuracy for the Southern Interior Ecoprovince. The resulting data and maps fill the scale gap between physiographic regions and inventory polygons for soils, surficial geology, aquifers, and terrain mapping in B.C.

The overall intent of the classification and spatial layers developed in this work is to increase the utility of substrate and landform information for decision making in B.C. This was achieved through the use of reconnaissance provincial layers to prioritize areas for detailed mapping. The focus was on identifying places in the landscape with thick surficial deposits and accumulations of unconsolidated sediments with high aquifer potential, such as in valley bottoms.

A parallel initiative was to develop a set of training points in an initial study area, which were attributed with the surficial materials classification system by a geomorphologist with expertise in terrain mapping. The training points were then used to pilot the classification system through a machine learning model and predictive mapping workflow. The resulting classification and predictive mapping approach can be improved, validated, and rolled out across the province in subsequent phases of the project to create consistent surficial material mapping layers for B.C.

2. METHODOLOGY

2.1 Surficial Material Classification System

A classification system and domain names were outlined for mappers to use during training point interpretation (Appendix C). The classification system provides a hierarchy of landform and surficial material classes and properties that can be interpreted from images.

The classification system includes two levels of *Broad Landform* classes, defined in Appendix C. At the broadest level of the hierarchy, the landscape is divided into generalized regions. The *Broad Landform* class (Level 1, Figure 1) was developed specifically to meet the objectives of this project and is intended to improve upon the provincial quaternary geology layers (BC Geological Survey 2023). The broad landform class currently include two categories, *Eroding Upland* and *Valley Fill*.

> **Eroding Uplands** are continuous areas subject to net erosion and scour such as mountains, hills and plateaus dominated by bedrock, veneers and blankets. It consists of mountainous terrain characterized by erosional and bedrock-controlled landforms, and extensive uplands and plateaus that may be transitional between mountains and valleys. The eroding upland category represents high elevation and bedrock-controlled areas of the landscape. This category currently captures all areas that do not fit the definition of valley fill.
>
> **Valley Fill** are continuous areas subject to net sediment accumulation in the landscape such as major valleys and lowlands dominated by constructional landforms such as terraces, plains, and hummocky glacial deposits. It is defined as regions of thick, unconsolidated sediment accumulation in the landscape, and characterized by extensive depositional landforms. It is extensive in lowlands and, depressions within B.C. Valley fill is common at relatively lower elevations of most major watersheds and within major river valleys. Across B.C., these areas are dominated by alluvium (river deposits) and sediments deposited by glacial ice, glacial lakes, and glacial meltwater. They include thick materials reworked by gravity and wind such as eolian veneers, dunes and landslide deposits. Along valley side walls the valley fill category includes materials deposited in fans, cones, and aprons where steep slopes and tributary valleys join mainstem valley fill.

As this study is focussed on aquifers in surficial materials, the valley fill category is further characterized according to seven classes such as texture and aquifer type (Level 2, Figure 1). Many of the classes were adapted from Howes and Kenk (1997). Between three and thirteen categories were identified within each class, however not every class was captured by this pilot study. Initially, aquifer subtypes were also considered as a class. However, during training point attribution and model testing, challenges with differentiating some categories arose and the aquifer subtype class was not carried through to the end of Phase 2. This is discussed further in the aquifer type results section.

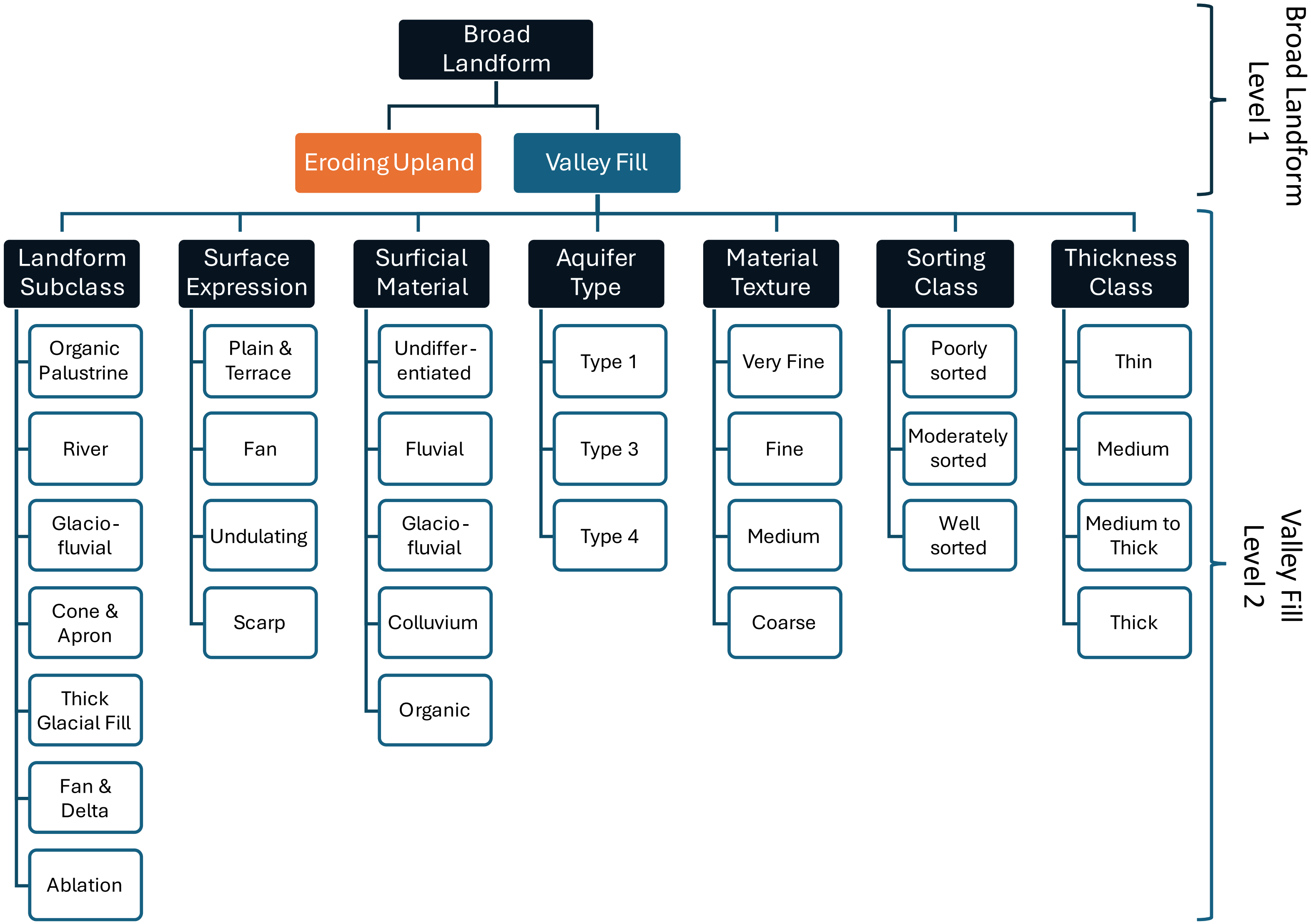


Figure 1: Summary of classes and categories within each class that were captured by training points in the Phase 2 study area. A full list of categories is included in Appendix C.

2.2 Study Area and Training Point Development

The Southern Interior EcoProvince (Southern Interior, Figure 2) was selected as the study area for piloting the modelling and predictive mapping methodology. Madrone Environmental Services Ltd. (Madrone) generated 500 random training points, with a minimum spacing of 150 metres, within five 1:20,000 grid map sheets in Manning Park, Westwold, and Clinton Areas (mapsheets 092H007, 092H006, 082L042, 092P002, and 092P003, referred to as Training Point Mapsheet Areas) as shown on Figure 2. These areas were chosen as there is a variety of topography and surficial materials to provide input to the predictive mapping model.

Using the classification, training point interpretation was conducted by Madrone in ArcGIS Online. A categorical value from each class was assigned to each point based on the aerial imagery, elevation contours, information about the geological and glacial history of the area, and existing mapping inventories. Information from GWELLS well logs and aquifer mapping were not used to attribute points in order to reduce “data leakage” between the training data and potential validation datasets (Johann de Jong, 2017).

Madrone conducted quality assurance throughout the mapping process. Point attribution was reviewed by qualified Registered Professional Geoscientists. The Professional Geoscientists signed-off on products and the final report. The final delivery of the products were in accordance with Terrestrial Ecosystem Information (TEI) digital data submission standards (Resources Information Standards Committee, 2023). While attributing the training points, notes were recorded to support the development of a classification key (Appendix A).

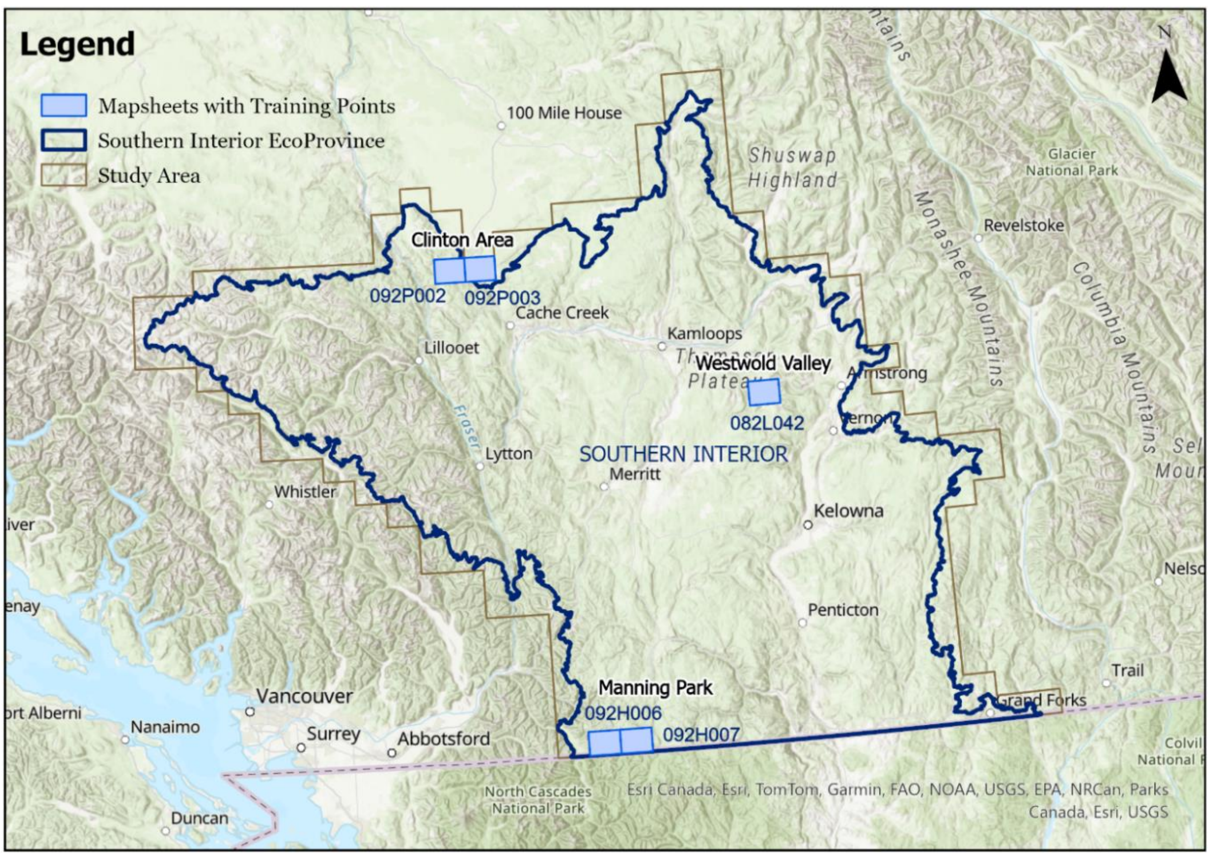


Figure 2: Phase 2 training points mapsheet areas and study area encompassing the Southern Interior EcoProvince.

Similar to Bulmer et al. (2016), it was identified that provincial maps should be developed as a mosaic of maps within subdivisions of the province. As the terrain, geology and geography of B.C. is highly variable, models are expected to make unique associations between terrain and climate variables and the surficial material class in different areas of the province. EcoProvinces (Dermarchi, 2011) were selected to use as subdivisions or "rule sheds" for the modelling as they describe areas with unique ecology, geography and climates where glacial, depositional and erosional environments are likely to be similar.

2.3 Co-variate Raster Stack Development

The covariate rasters were developed using a 25 m digital elevation model (DEM) raster layer in B.C. in Albers projection. The Albers projection was selected to facilitate development of continuous predictive maps across the entire province with a single projection. The study area was set to encompass the Southern Interior EcoProvince along the mapsheet grid (Figure 2) and the DEM was trimmed to the study area extent.

The DEM was smoothed using the focal function in R (Hijmans, 2024). This uses a moving window method to average across a 3 by 3 matrix, with equal weight on all cells. Using a stream vector shapefile from the B.C. Freshwater Atlas (Province of British Columbia n.d.), the stream network was then burned 5 m into the DEM surface to facilitate topographic analyses which assess the movement of water over the landscape by avoiding raster resolution issues.

The DEM was further processed to create 23 DEM derivative terrain rasters using the rsaga R package with SAGA GIS (Brenning, 2008), as listed below and described further in Appendix B. Rasters were also created for eight annual climate variables using ClimateBC (Wang et al. 2022). Climate rasters were generated using a 100 m scale DEM, created by averaging the 25 m DEM, and then subsampled back to the 25 m scale. The climate variables selected are listed below and are all directly calculated, rather than derivative parameters. All rasters were created with the same spatial extent and resolution, and then stacked to be used as covariates in the predictive modelling. The raster variables were selected based on those used in similar predictive mapping studies (Bulmer et al. 2016; Filatow et al. 2020).

Terrain Variables:

1. Elevation
2. Aspect
3. Top-Down Flow Accumulation (CAREA)
4. Plan Curvature (CPLAN)
5. Profile Curvature (CPROF)
6. Diurnal Anisotropic Heating (DAH)
7. Eastness
8. Northness
9. Topographic Openness – Negative and Positive (NOPEN, POPEN)
10. Height Above Channel (HACH)
11. Channel Network Base Level (CNBL)
12. Multiresolution Index of Valley Bottom Flatness (MRVBF)
13. Multiresolution Index of Ridge Top Flatness (MRRTF)
14. Analytical Hillshade (HILL)
15. Slope
16. SAGA Wetness Index (TOPOWET)
17. Topographic Ruggedness Index (TRI)
18. Topographic Position Index at four scales (TPI_500, TPI_1000, TPI_2000, TPI_5000)

19. Multi-scale Topographic Position Index at two scales (MSTPI, MSTPI_WIDE)

Climate Variables:

20. Mean Annual Temperature (MAT)
21. Mean Warmest Month Temperature (MWMT)
22. Mean Coldest Month Temperature (MCMT)
23. Temperature Difference Between MWMT and MCMT (TD)
24. Mean Annual Precipitation (MAP)
25. Mean Annual Summer Precipitation (MSP)
26. Annual Heat-Moisture Index (AHM)
27. Summer Heat-Moisture Index (SHM)

2.4 Predictive Mapping with Machine Learning Model

Preliminary modelling was conducted during Phase 2 to assess:

- the predictability of the classifications,
- the predictions generated with the produced training points, and
- the gaps in training data needed to improve predictions.

A simple modelling approach was taken using a random forest model to generate preliminary models and maps. Figure 3 presents a conceptual flow chart of the steps associated with predictive modelling and the progress that has been made in the Southern Interior to date.

2.4.1 Training Classification Matrix

The raster stack was sampled at the training point locations to create a covariate matrix that included the coordinates, surficial material classifications, and covariate values sampled from the rasters.

2.4.2 Data Balancing

The imbalance between the number of observations in each category within a class creates challenges for the model fit (e.g., Bulmer et al. 2016; More and Rana 2017; Bader-El-Den et al. 2018). This is a common challenge in data analysis aimed at predicting or understanding a rare but important event, such as scarps, landslides, undifferentiated, and organic deposits. Balanced datasets have been demonstrated to improve the predictive accuracy of minority classifications when using random forest models. Data balancing was performed on the training classification matrix with a combination of synthetic minority oversampling technique (SMOTE)(Chawla et al., 2002) and random undersampling. Each class was resampled such that all categories had a 1:1 balance except for the *eroding upland* category which was undersampled to have a 1:1, 2:1 or 5:1 ratio with the other categories depending on the number of samples available in the other categories. During Phase 2 of the project, modelling was only intended to assess methodologies and make recommendations for next steps, a sensitivity analysis was not performed to optimize data balancing.

2.4.3 Random Forest Modelling

A random forest model was used to identify the relationships between the categories for each class and the covariates. Random forest models are classification algorithms which ensemble the results of multiple decision trees. The models were run in R using functions from the ModelMap package (Freemen et al. 2022). Modelling was completed separately for each class.

Model performance was determined based on an internal error metric called out-of-bag error (OOB error), the percent correctly classified (PCC), and the kappa statistic. The OOB error indicates how well the algorithm performed when predicting held-out, or unseen, training data. A random sample of one third of the input training data, known as a bootstrap, is used to generate a fully-grown decision tree which is then compared with held out data to calculate error. Metrics from all other trees are combined to establish overall difference (Breiman, 2001). The kappa statistic is a chance corrected metric that reflects the difference between observed agreement and agreement expected by random chance. Variable importance was compared using the Mean Decrease in Accuracy and the Mean Decrease in Gini.

Each model was used to create a predictive raster map across the Southern Interior study area. However, model performance metrics such as OOB error and kappa, only provide an estimate of model performance within the Training Point Areas (five mapsheets). Across the rest of the study area, model performance was evaluated qualitatively by visually assessing the predicted classification distributions compared to expected characteristics of the surficial materials in that area. A separate validation data set is needed to conduct validation of predictive maps in future phases of the project.

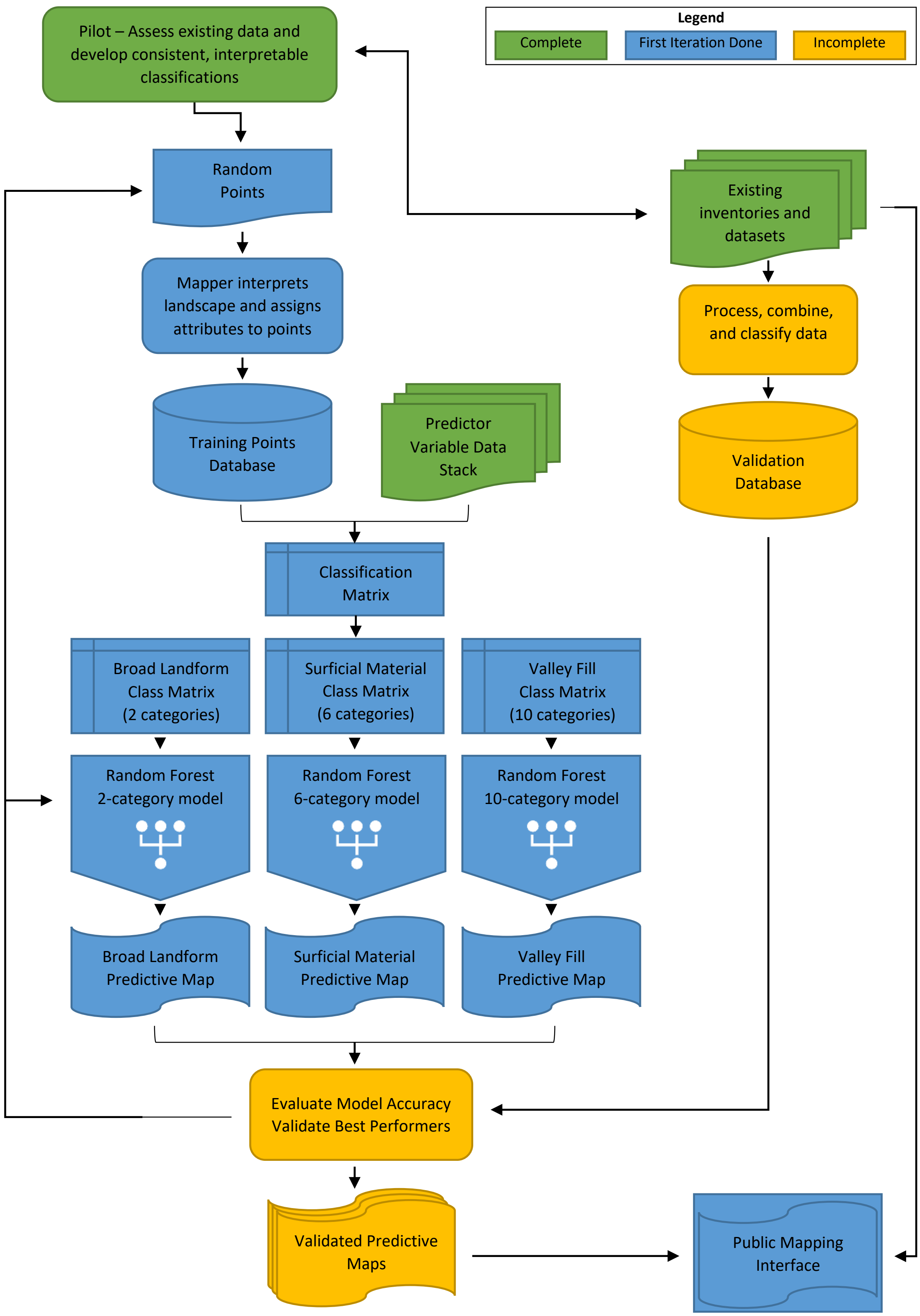


Figure 3: Conceptual flow chart presenting the steps associated with predictive modelling and the progress that has been made in the Southern Interior to date.

3. RESULTS AND DISCUSSION

3.1 Surficial Material Classification System

Appendix C outlines the broad landform classification domain tables defining each class and the included categories as summarized in Figure 1 above.

3.2 Training Point Development

Figure 4 shows where training points were attributed with classification categories. Table 1 presents the distribution of categories attributed to training points for each class after minor data processing. Not all categories identified in the classification were attributed in the training point set. Some categories do not occur in the Southern Interior, such as marine deposits, and others were not captured by the distribution of random points, such as lacustrine and glaciolacustrine deposits.

The imbalance between the eroding upland and valley fill categories in the broad landform class, creates high sample imbalance in the other classes, as the Level 2 categories are only attributed within the valley fill category.

Categories of very fine material texture and scarp landform subclass had to be merged with other categories as SMOTE technique requires a minimum of three points to create synthetic samples. Low sample numbers in some categories were challenging for balancing and modelling. To balance the data, synthetic samples of some categories were generated in numbers that were orders of magnitude greater than the original sample size and/or the majority categories was significantly undersampled to balance the data.

Table 1: Number of training points attributed in each category of a class.

Broad Landform	Blank 11	Eroding Upland 2200	Valley Fill 301				
Landform Subclass*	Blank 11	Eroding Upland 2200	Organic Palustrine 9 Thick Glacial Fill 9	River 136 Fan Delta 58	Glaciofluvial 61 Ablation 3	Cone Apron 25	
Surface Expression*	Blank 11	Eroding Upland 2200	Plain or Terrace 198	Fan 76	Undulating 27	Scarp 2	
Surficial Material*	Blank 11	Eroding Upland 2200	Undifferentiated 9	Fluvial 145	Glaciofluvial 85	Colluvium 53	Organic 9
Aquifer Type*	Blank 28	Eroding Upland 2200	Type 1 115	Type 3 85	Type 4 84		
Material Texture*	Blank 22	Eroding Upland 2200	Very Fine 1	Fine 157	Medium 126	Coarse 6	
Sorting Class*	Blank 20	Eroding Upland 2200	Poorly sorted 92	Moderately sorted 81	Well sorted 119		
Thickness Class*	Blank 178	Eroding Upland 2200	Thin 3	Medium 48	Medium to Thick 10	Thick 73	

*Eroding Upland training points were not classified during this Phase 2 assessment and are shown as "Eroding Upland".

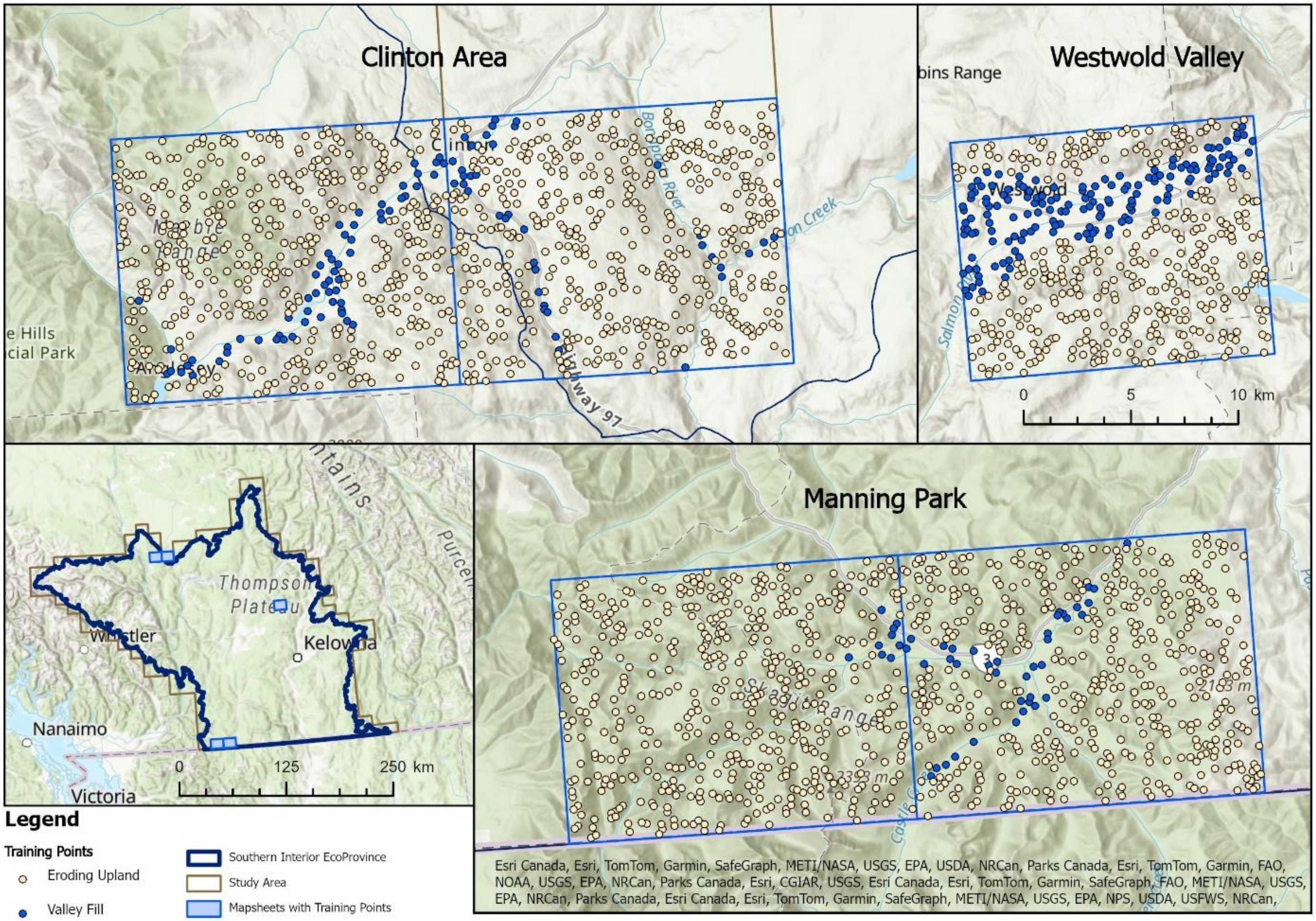


Figure 4: Map of training points locations.

3.3 Co-variate Raster Stack Development

As shown in the predictor correlation plot in Figure 5 below, many of the raster covariates were highly correlated. This is not unexpected, as many of the covariates were derived from the same DEM, and therefore the values are correlated to elevation. Multicollinearity creates challenges for machine learning models and can lead to inaccurate variable importance and overfitting. During the next phase of the modelling (Phase 3) use of covariate reduction methods such as variance inflation factor (VIF) and recursive feature elimination are recommended to improve model performance (Misra and Yadav 2020; Chan et al. 2022).

The relative importance of each covariate within a model was estimated by the Mean Decrease in Accuracy and the Mean Decrease in Gini (Freemen et al. 2022). The Mean Decrease Accuracy plot expresses how much accuracy the model loses by excluding each variable. The more the accuracy suffers, the more important the variable is for the successful classification. The mean decrease in Gini coefficient is a measure of how each variable contributes to the homogeneity of the nodes and leaves in the resulting random forest.

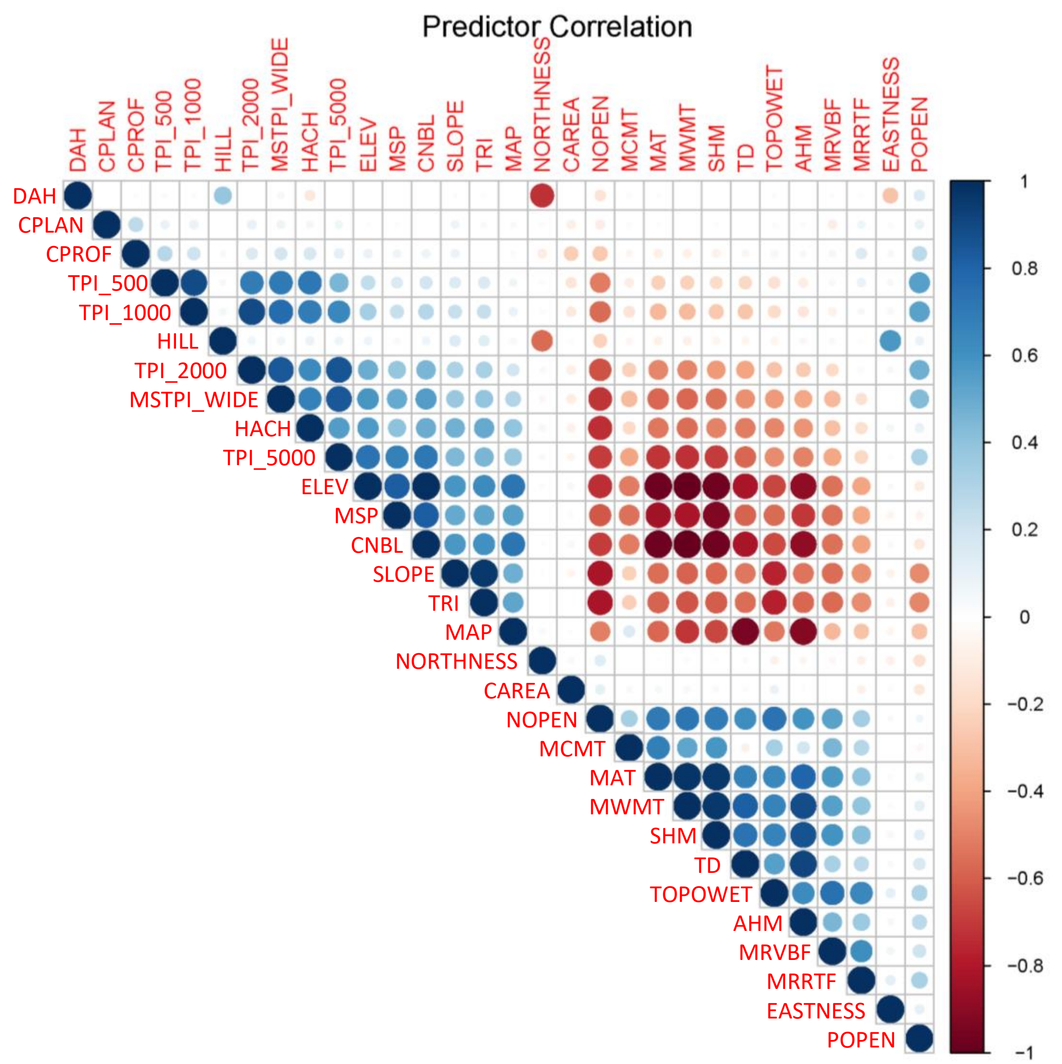


Figure 5: Visual representation of degree of correlation between covariate rasters.

An example of the ranked predictor importance for the broad landform balanced model is included in Figure 6 below. Results indicate that variables calculated with a higher radius of influence or multiple radiuses of influence, e.g., TPI_5000, MRVBF, had higher significance in the models. These covariates may help the model determine a cell's relative position in the landscape, such as whether it's located in a higher order valley. However, as mentioned above, multicollinearity of covariates may mask their true importance (Chan et al. 2022).

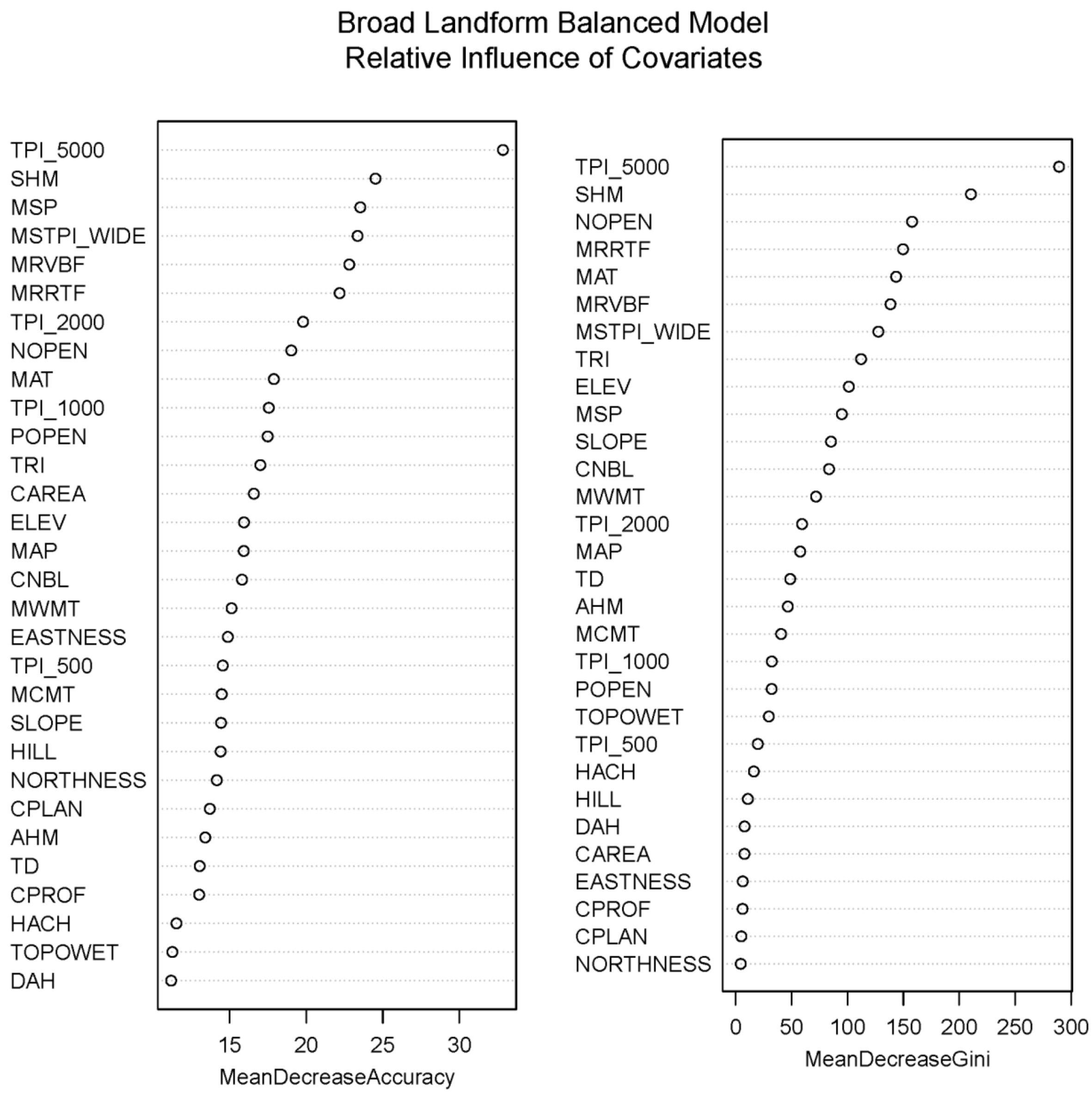


Figure 6: Ranked predictor importance for the Broad Landform balanced model using the Mean Decrease in Accuracy and Mean Decrease in Gini metrics. The variables are presented in order of descending importance.

3.4 Predictive Mapping with Machine Learning Models

For each class, two random forest models were run, one with the original training points and one with synthetically balanced training points. A summary of the model metrics discussed in section 2.4.3 for each model is included in Table 2 below. These metrics suggest that overall model performance, within the training point mapsheet areas was good and that data balancing improved model performance.

Without a validation dataset the model performance can only be evaluated quantitatively in Training Point Mapsheet Areas. Therefore, the values in Table 2 only represent model performance within the Manning Park, Westwold and Clinton mapsheets (5/>300 mapsheets in the study area).

Model performance within the rest of the study area was qualitatively assessed using a visual comparison of the model generated predictive maps with expert knowledge of the landscape and review of other available terrain, bioterrain, and soils mapping layers not provided to the models. The provincial Quaternary Geology layer is also presented with the mapping results below for visual comparison. However, we know that this dataset provides a rough presence-absence estimate of glacial sediments, which is not reliable in many locations.

A discussion of the results for each class is included in the following sections.

Table 2: Summary of model outputs for balanced and unbalanced models of each class.

Class	Number of Classes	Number of Predictors	Data Balancing?	Minority Class Size	Number of Training Points (balance*)	Out of Bag Error (%)	Kappa	Percent Correctly Classified (PCC)
Broad Landform*	2	30	Y	301	4400 (1:1)	0.9	0.98	99%
Broad Landform	2	30	N	301	2501	1.8	0.91	95%
Landform Subclass*	8	30	Y	3	4950 (1:2)	0.9	0.99	99%
Landform Subclass	8	30	N	3	2501	5.0	0.76	95%
Aquifer Type*	4	30	Y	84	2750 (1:2)	1.7	0.98	98%
Aquifer Type	4	30	N	84	2485	4.8	0.76	95%
Surficial Material*	6	30	Y	9	3080(1:2)	1.7	0.98	98%
Surficial Material	6	30	N	9	2501	4.8	0.77	95%
Surface Expression*	4	30	Y	27	5500 (1:2)	0.8	0.99	99%
Surface Expression	4	30	N	27	2501	4.0	0.81	96%
Thickness Class*	5	30	Y	3	1980 (1:5)	1.2	0.98	99%
Thickness Class*	5	30	N	3	2334	2.2	0.98	78%
Material Texture*	4	30	Y	6	2750 (1:2)	2.0	0.97	98%
Material Texture	4	30	N	6	2246	11.9	0.88	74%
Sorting Class*	4	30	Y	81	2750 (1:2)	1.8	0.98	97%
Sorting Class	4	30	N	81	2253	13.3	0.87	72%

* For balanced models, ratios within brackets indicate the balance between the eroding upland category and the other categories in the class.

3.4.1 Broad Landform

The broad landform models had 99% accuracy in the training point mapsheet areas. Visually, the broad landform map predicted throughout Southern Interior reflects the concept of the broad landform class, although performance cannot be quantified (Figure 7). The valley fill category is generally predicted in areas that were anticipated, although it may be slightly under-represented on the landscape. In areas with mountainous topography, the model predictions are more representative as the training points were developed in more classic examples of glacial valleys (Figure 9 and Figure 10).

Within the Okanagan-Thompson Plateau, where there were no training point mapsheets, the division of the landscape into eroding upland and valley fill is not as clear as in the mountain-valley landscapes where the definitions were developed. Figure 8 shows areas on the plateau which have been coarsely mapped as Quaternary deposits in the provincial mapping layer but are not within linear valley bottoms. Our expert knowledge of the area indicates that these landforms likely consist of thicker depositional sequences, however, the model prediction identified these areas as eroding upland. This suggests that a third Broad Landform category is needed to differentiate extensive depositional uplands and plateaus generally consisting of till plains, drumlins, and hummocky glacial deposits such as eskers, kames, and kettled topography, for example, in the northern Nicola Valley area east of Merritt (Figure 8).

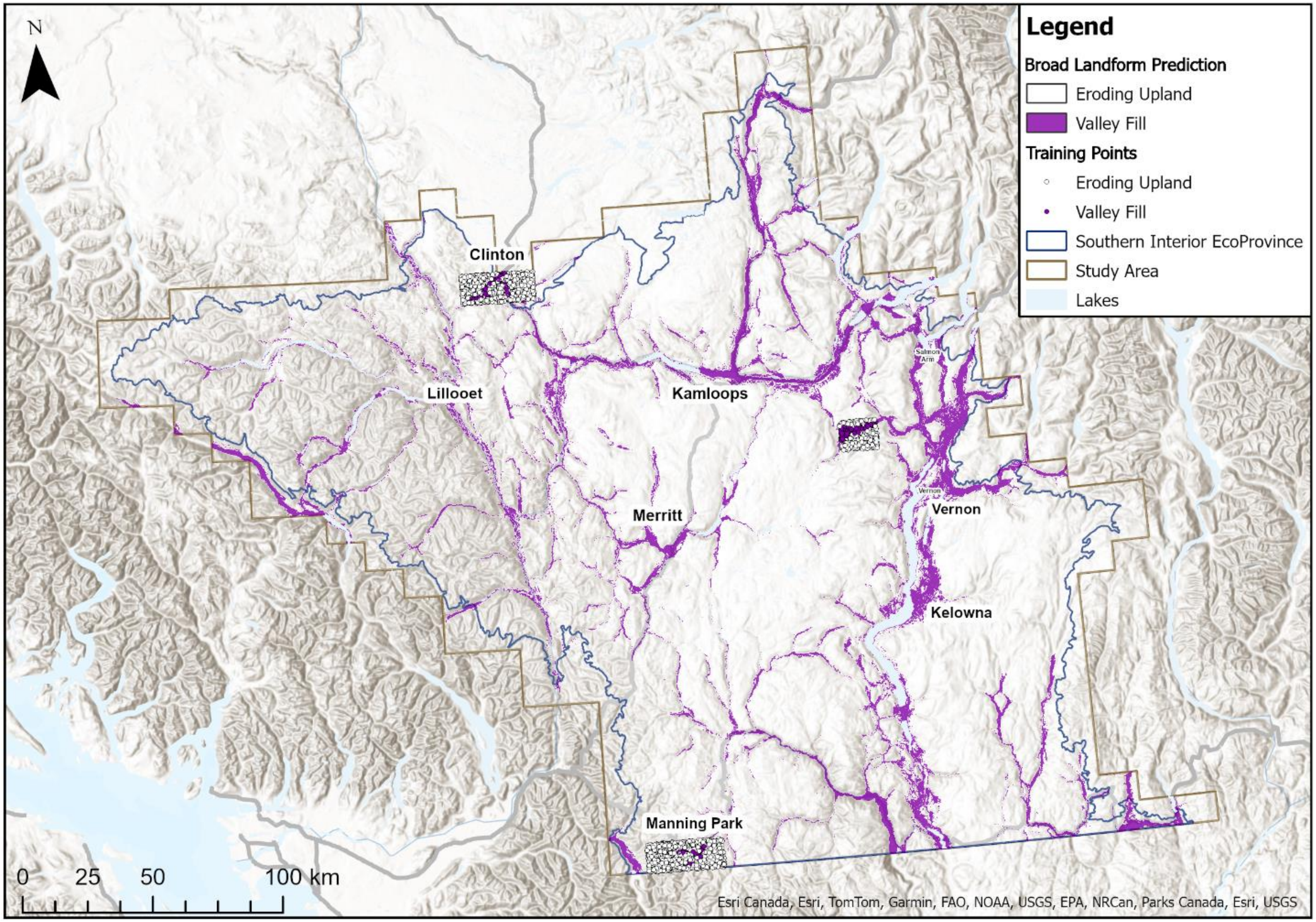


Figure 7: Broad Landform predictions across the Southern Interior.

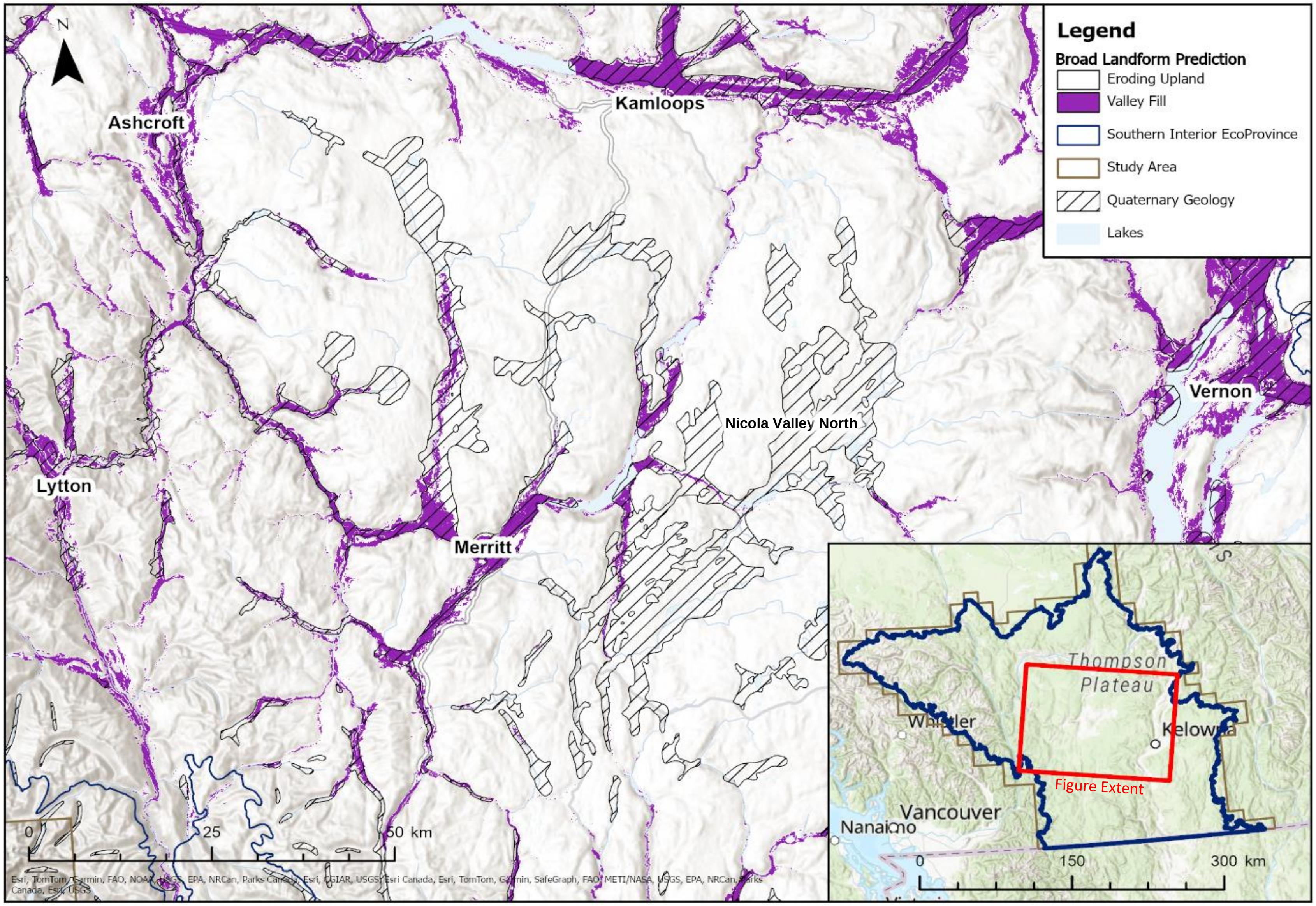


Figure 8: Broad Landform predictions compared to Quaternary Geology polygons in the Thompson Plateau and northern Nicola Valley areas to the North and East of Merritt respectively.

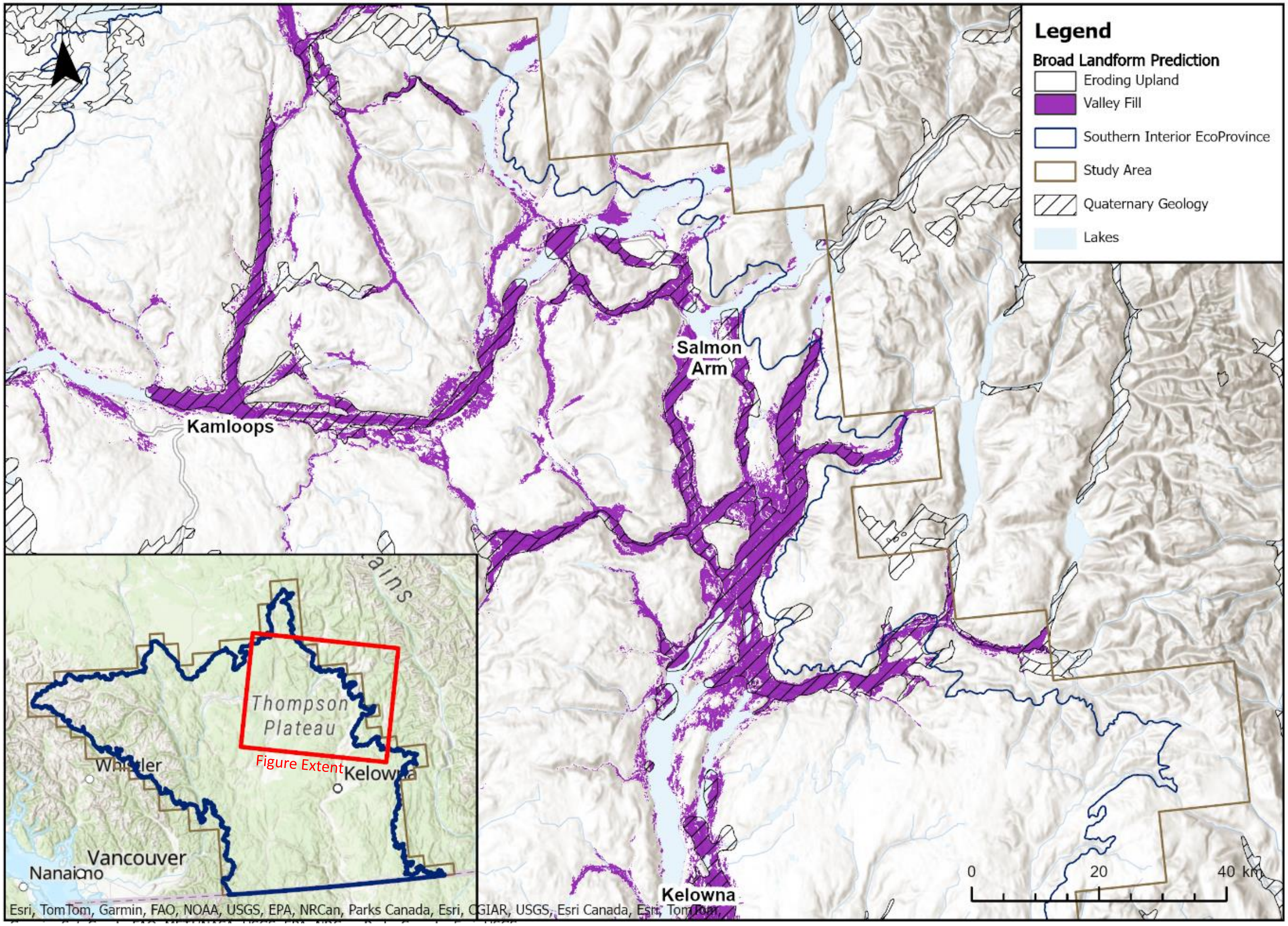


Figure 9: Broad Landform predictions compared to Quaternary Geology polygons near Shuswap Lake and Armstrong, to the East of Kamloops.

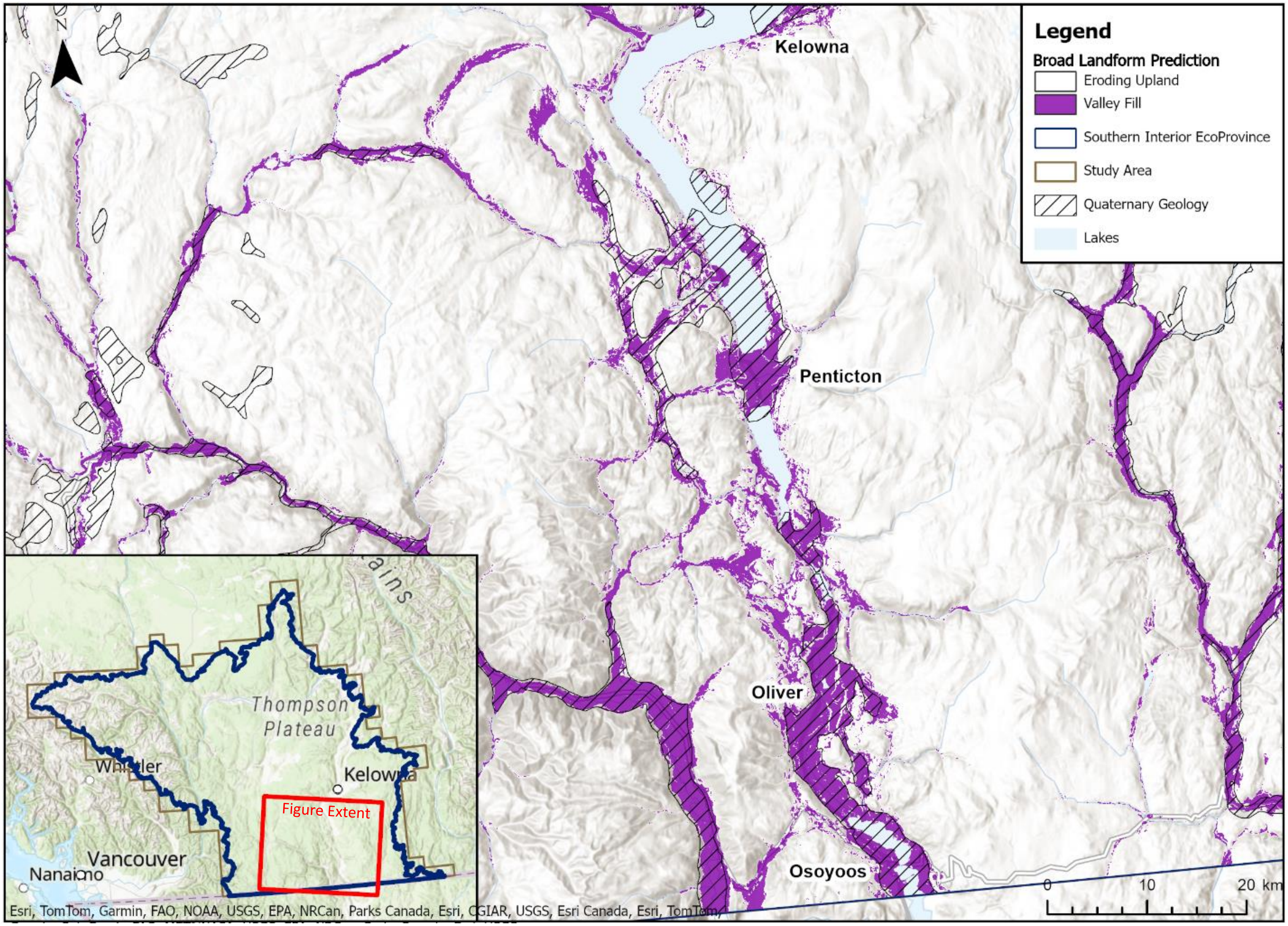


Figure 10: Broad Landform predictions compared to Quaternary Geology polygons in the South Okanagan areas.

3.4.2 Landform Subclass

The categories of the landform subclass that were attributed within the Southern Interior valley fill were ablation, cone and apron, fan and delta, river, glaciofluvial, organic palustrine and thick glacial fill (Figure 11). Glaciolacustrine and scarp landforms are also expected to be present within valleys in the study area, however they were not captured by the training points and therefore, were not included in the model.

The balanced model was 99% accurate within the training point mapsheet areas. Outside of the training point mapsheet areas, the predicted model results appear reasonable, but with room for improvement (Figure 12). For example, thick glacial fill sediments were likely over predicted in the Kelowna area based on our knowledge of the terrain and existing mapping inventories (Figure 13), perhaps because there are glaciolacustrine sediments present, which was not a class provided to the model. Despite the high number of categories in this class, the modelling and mapping results are promising and we expect that by improving and increasing the training data this class can be modelled with reliable accuracy. Model results demonstrate the diversity of landforms occurring within valley fill areas and provides more information than the Quaternary Geology layer.

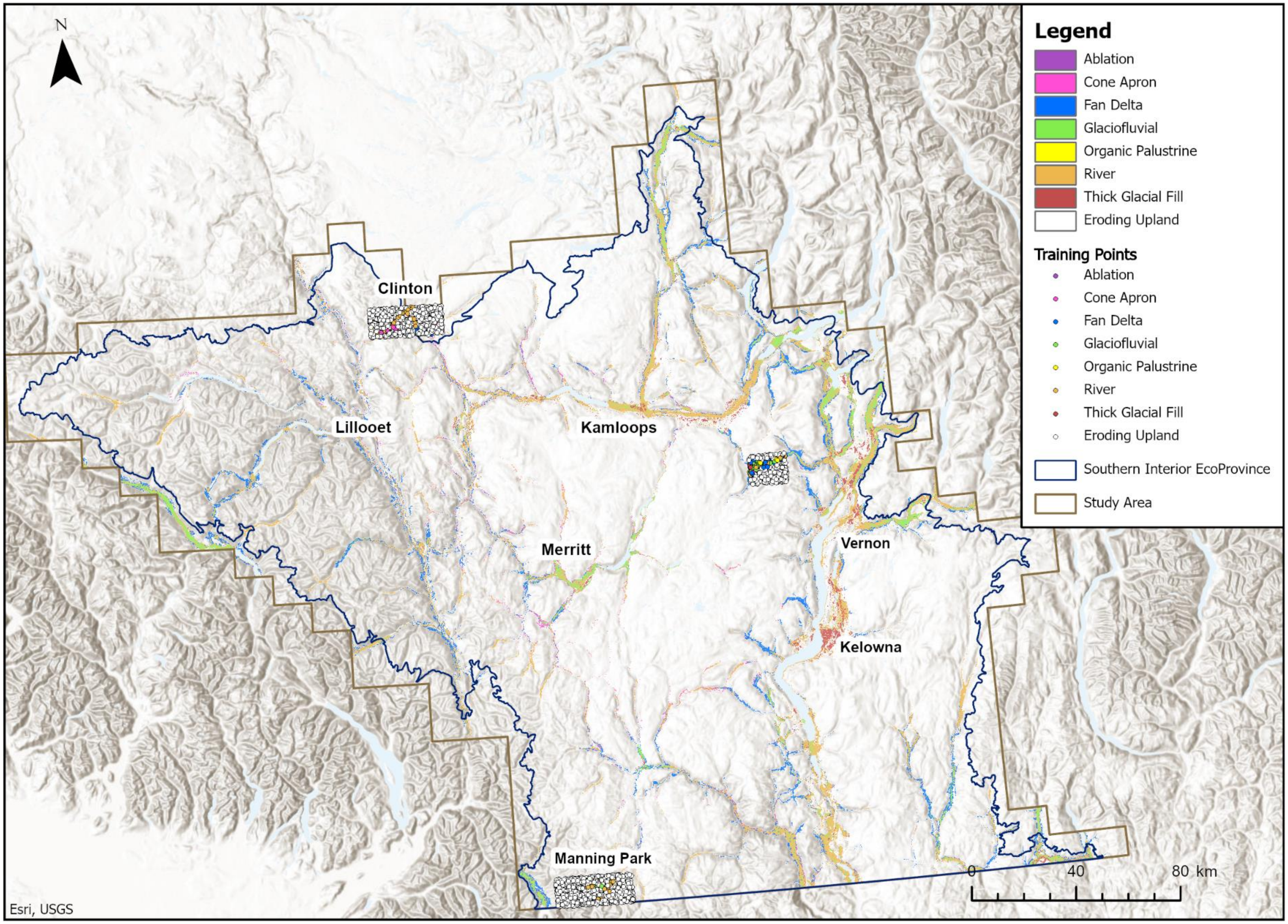


Figure 11: Landform Subclass predictions across the Southern Interior.

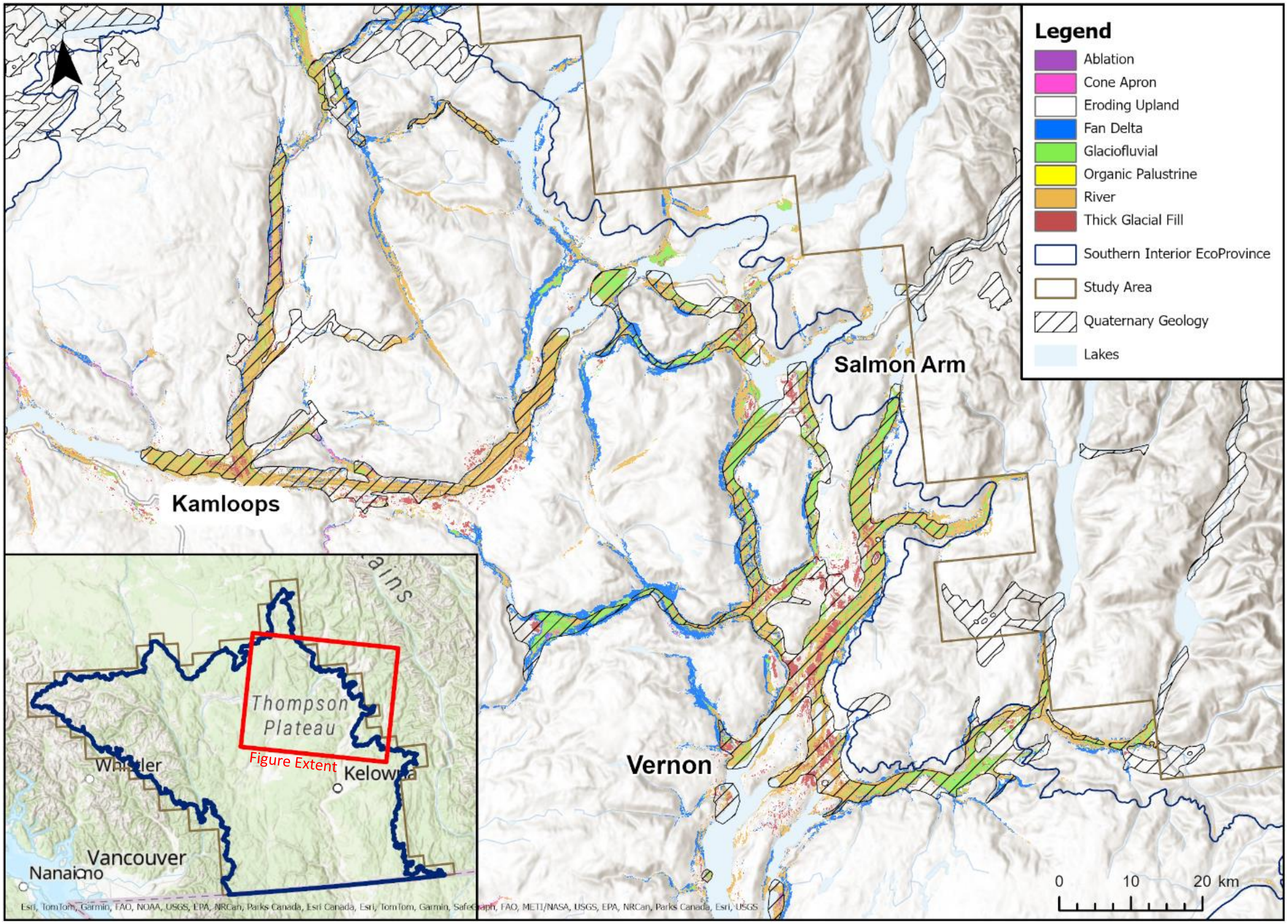


Figure 12: Landform Subclass predictions compared to Quaternary Geology polygons in the Kamloops and Shuswap Lake areas.

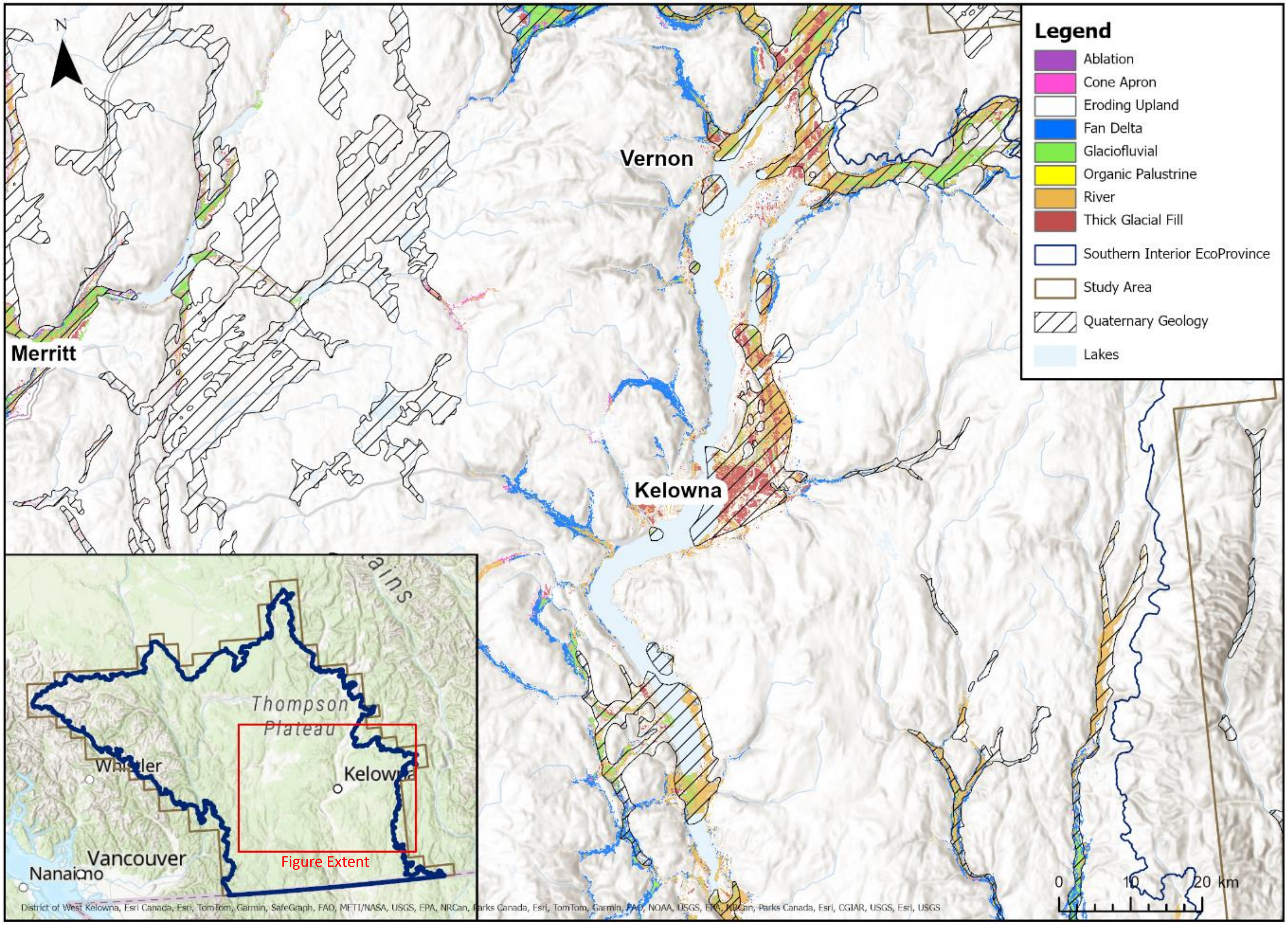


Figure 13: Landform Subclass predictions compared to Quaternary Geology polygons in the Kelowna and Vernon areas.

3.4.3 Surficial Material

After processing the training data, the surficial materials categories that were attributed within the valley fill were colluvium, fluvial, glaciofluvial, organic palustrine and undifferentiated. Glaciolacustrine and glacial till are other categories that we expect to be present within valleys, however they were not captured by the training points. Increasing the number of training points representing minority classes will be a focus during the next phase (Phase 3 of the project).

The balanced model predicted the surficial material categories within the training point mapsheet areas with 98% accuracy. In other parts of the Southern Interior study area the predictive maps look reasonable, but in some instances was not able to effectively discern between fluvial and glaciofluvial materials (Figure 14). For the surficial material class, the undifferentiated category was overpredicted in the Kelowna area (Figure 15). As the training points did not capture any lacustrine or glaciolacustrine materials, the predictive map does not include these material types, despite their known occurrence within valleys in the study area. During the next phase, capturing these materials when adding additional training points will improve model performance.

As modelled surficial material is only predicted within the valley fill, although this is an improvement on the Quaternary Geology layer, it’s not expected to be as useful to natural resource decision makers as a map that is fully attributed across the landscape. Bulmer et al. (2016) created a provincial soil parent material map at the 100 m scale using training data derived from existing soil mapping inventories and validation data from the B.C. Soil Information System (BCSIS) data. A similar approach to map surficial material at the 25 m scale may yield better results.

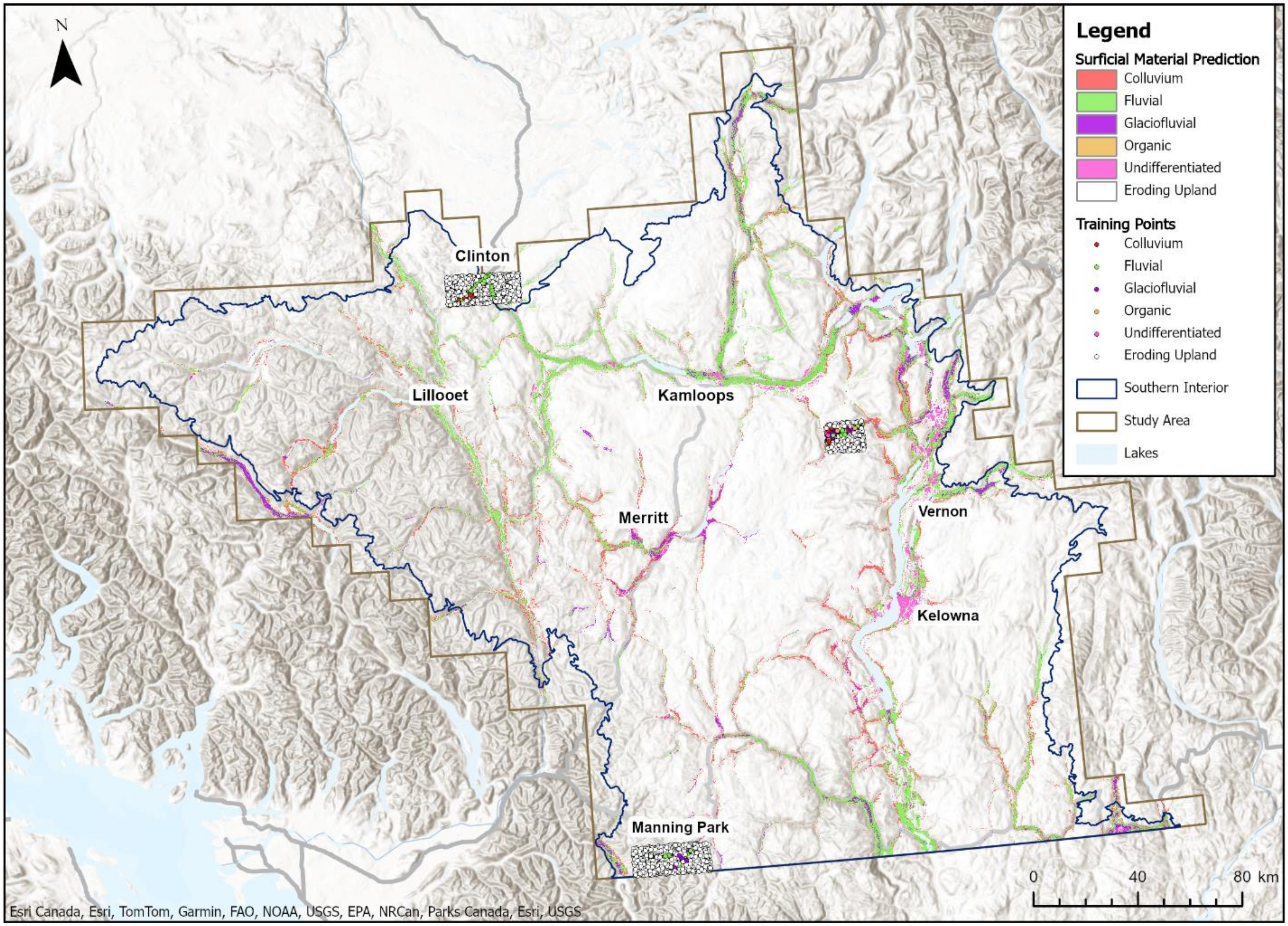


Figure 14: Surficial Material predictions across the Southern Interior.

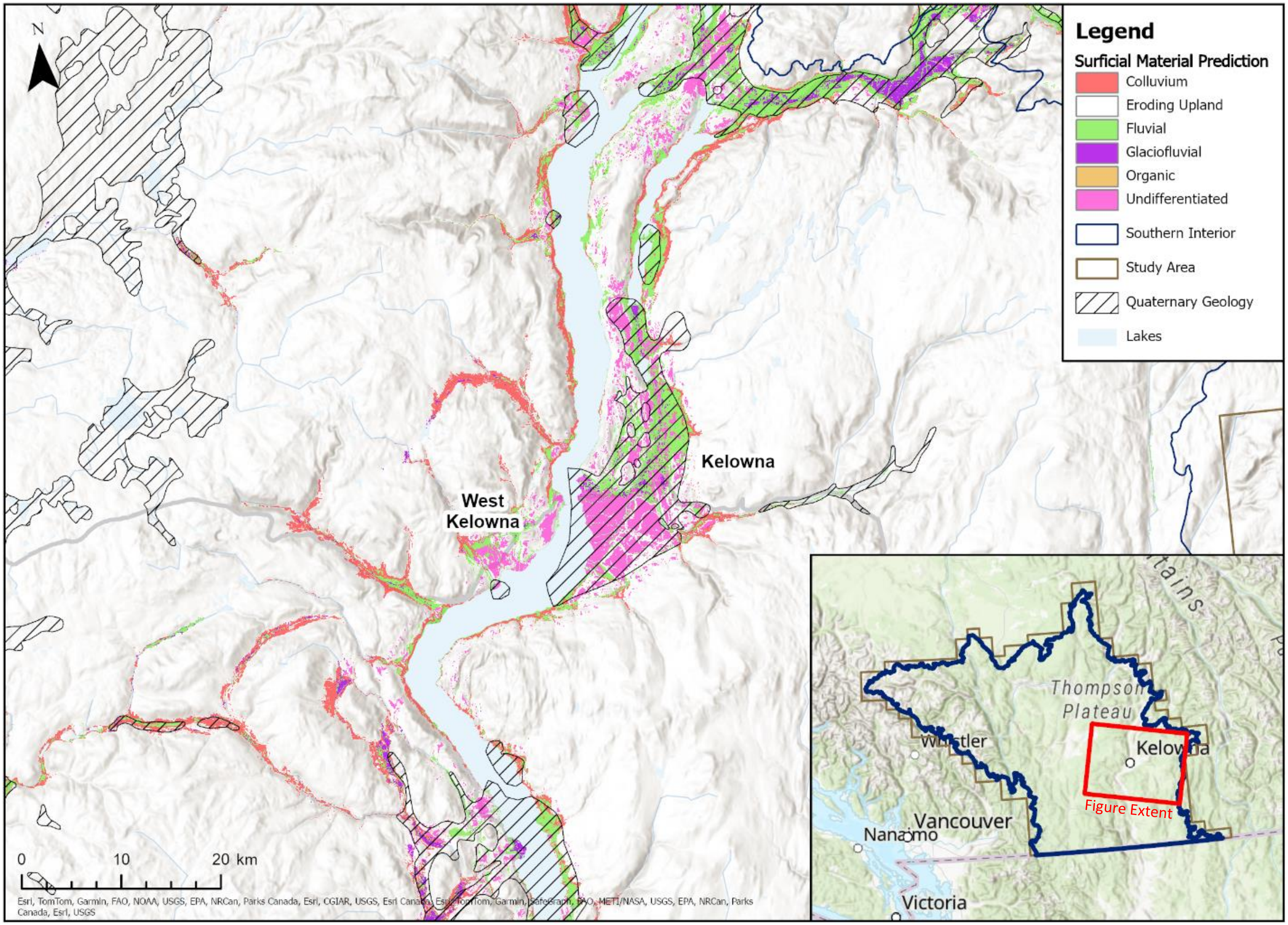


Figure 15: Surficial Material predictions in the Kelowna area.

3.4.4 *Surface Expression*

After processing the training data, the surface expression categories that were attributed within the Southern Interior valley fill were plain or terrace, fan, and undulating (Figure 16). Surface expressions such as hummocky or ridged, and cones, were used exclusively in the eroding upland and therefore were not included in the model. Two points were attributed as scarps within the valley fill, however the model requires a minimum of three points per category and therefore these points were not included in the current phase of modelling.

The balanced model was able to predict the surficial material categories within the training point mapsheet areas with 99% accuracy. Outside of the training point mapsheet areas the predictive map classified the landscape reasonably well. As shown in Figure 17 below, in some areas the model over predicted fans and underpredicted undulating expression.

Generally, a predictive map of surface expression which is only attributed within the valley fill areas may not be that useful for natural resource decision makers. It is likely more beneficial to attribute this class across the entire landscape. The categories within this class could be revised to reflect the typical surface expression standard from Howes and Kenk (1997) during a future phase.

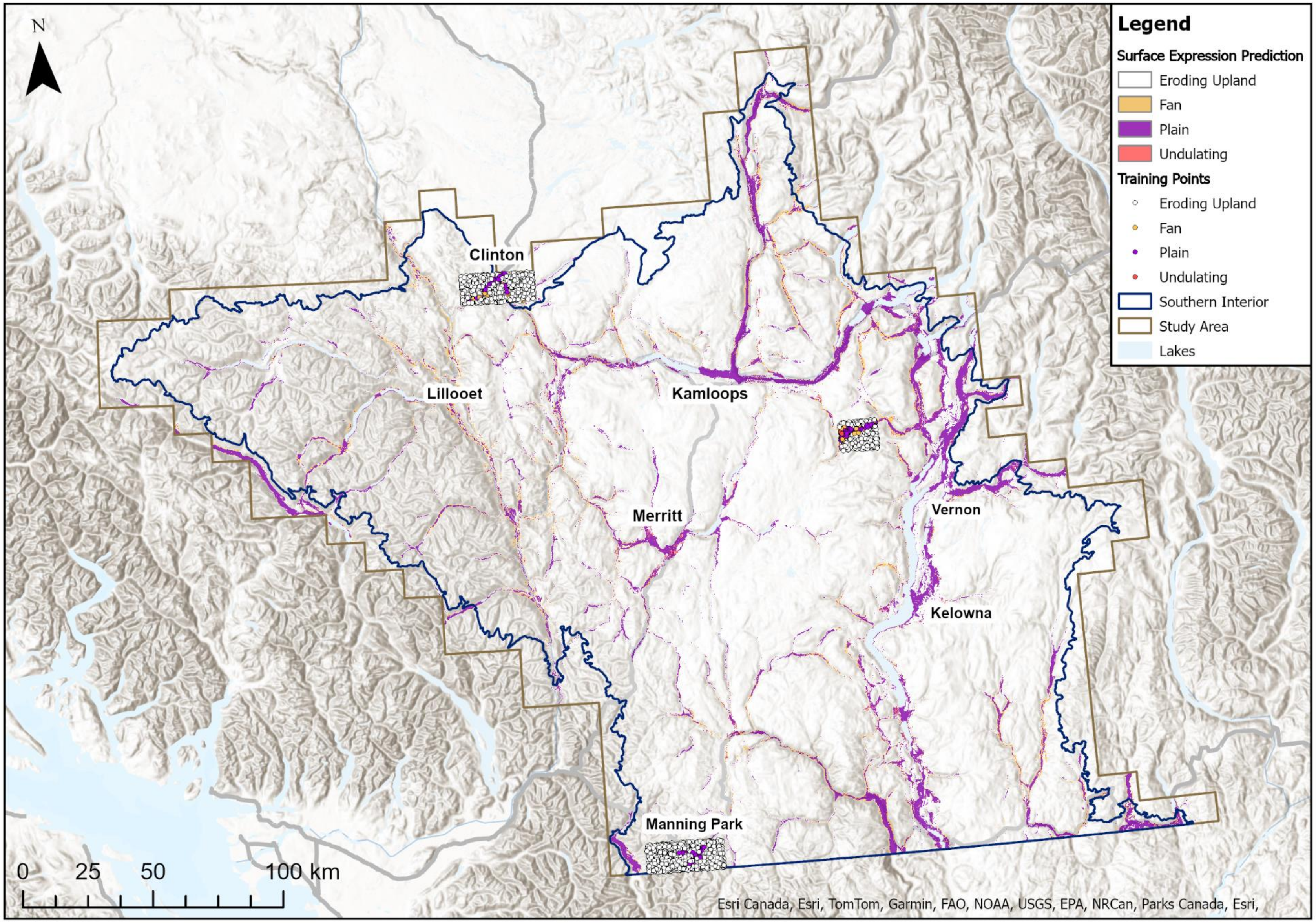


Figure 16: Surficial Expression predictions across the Southern Interior.

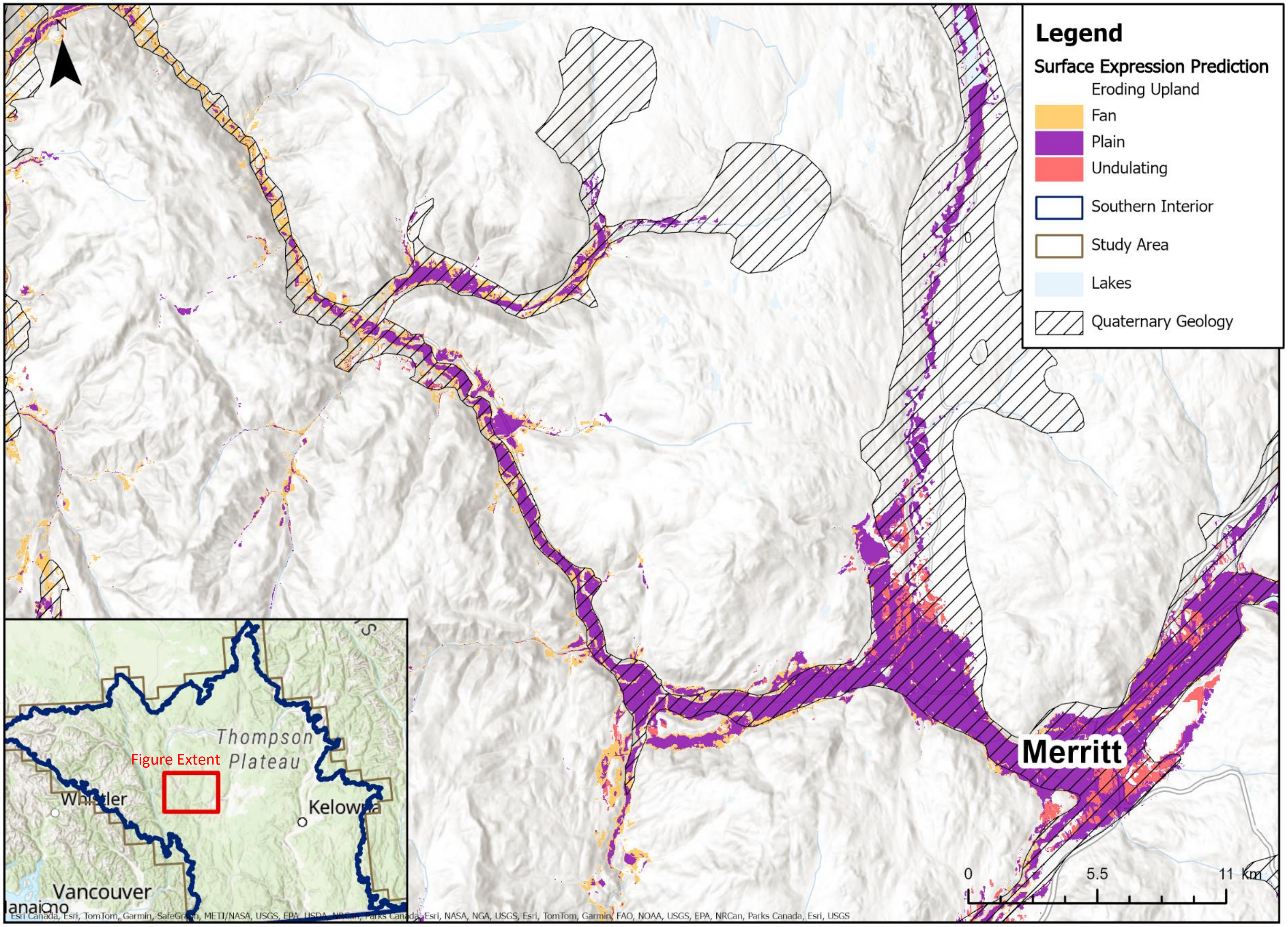


Figure 17: Surficial Expression predictions along the Fraser River and Thompson River valleys north of Lytton.

3.4.5 Aquifer Type

Initially, both aquifer type and aquifer subtype were included as classes under consideration. However, it proved difficult to differentiate between some aquifer subtypes during training point attribution and model testing. With the desktop approach it was particularly difficult to differentiate with high confidence between subtypes 4a, 4b and 4c, which are all aquifer types with glacial modes of deposition. After discussion within the working group, only the subtype 1a category was attributed for the aquifer subtype class, as the subtype 1a category could be consistently identified and attributed using the desktop approach. In future phases of the project the sub-type 1a category will be incorporated into the aquifer type class (Figure 18).

During Phase 2, aquifer type categories 1, 3, 4 and Not Mapped were attributed in the study area (Figure 19). All of the aquifers attributed as type 1, were also attributed as subtype 1a indicated that no subtype 1b or type 1c aquifers were captured by the training points. During Phase 3, subtype 1a and other type 1 aquifers will be strategically captured by the training points to differentiate them during modelling.

The balanced model classified the aquifer type categories with 99% internal accuracy for the training point data (applies to training mapsheets only). Outside of the training mapsheets, review of the predictive map indicates that the model is over predicting Type 3 aquifers across the landscape, particularly in the western portion of the study area shown in Figure 20.

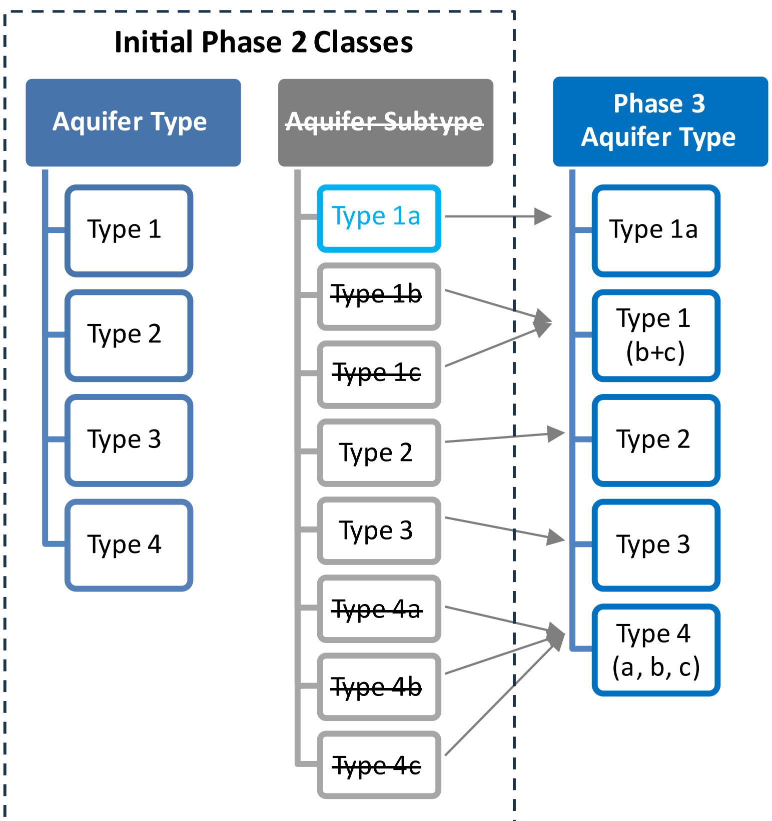


Figure 18: Summary of aquifer type and subtype classification results and modifications recommended for Phase 3 of the project.

The model also appears to be predicting type 4 (glacial aquifers) in areas where we would expect type 1 (fluvial aquifer) to be predicted. Figure 21 presents the eastern portion of the study area and shows that when compared to mapped aquifer polygons, the model does not consistently discern between type 1 and type 4 aquifers. Capturing the variability of type 1 aquifers in the Phase 3 training points (i.e. adding smaller subtype 1b and 1c aquifers, as only 1a was captured during Phase 2) is expected to improve model differentiation between type 1 and type 4 aquifers. It should also be noted that aquifer mapping is completed through hydrogeological interpretation often based on limited information or uneven distribution of borehole logs. Further, many valleys in B.C. are known to host multiple stacked aquifers. For example, a subtype 1b aquifer may overly a subtype 4b aquifer. Therefore, discrepancies between current aquifer mapping and model results may not indicate that the model is objectively incorrect in all cases.

During the modelling exercise it was noted that there was a definition discrepancy between the B.C. Aquifer Subtype descriptions and the aquifer categories applied by Madrone. The B.C. Aquifer Mapping captures type 2 deltaic aquifers along Shuswap Lake (Figure 21). However, Madrone reserved the type 2 aquifer category for deltas along the marine environment, attributing lake deltas as type 3 alluvial fan aquifers. This could be adjusted when extending this project to other areas of the province. Figure 21 demonstrates that the model struggled to predict aquifer types outside of the training areas as the fans/deltas along Shuswap Lake were predicted to be type 4 aquifers rather than type 2 or type 3.

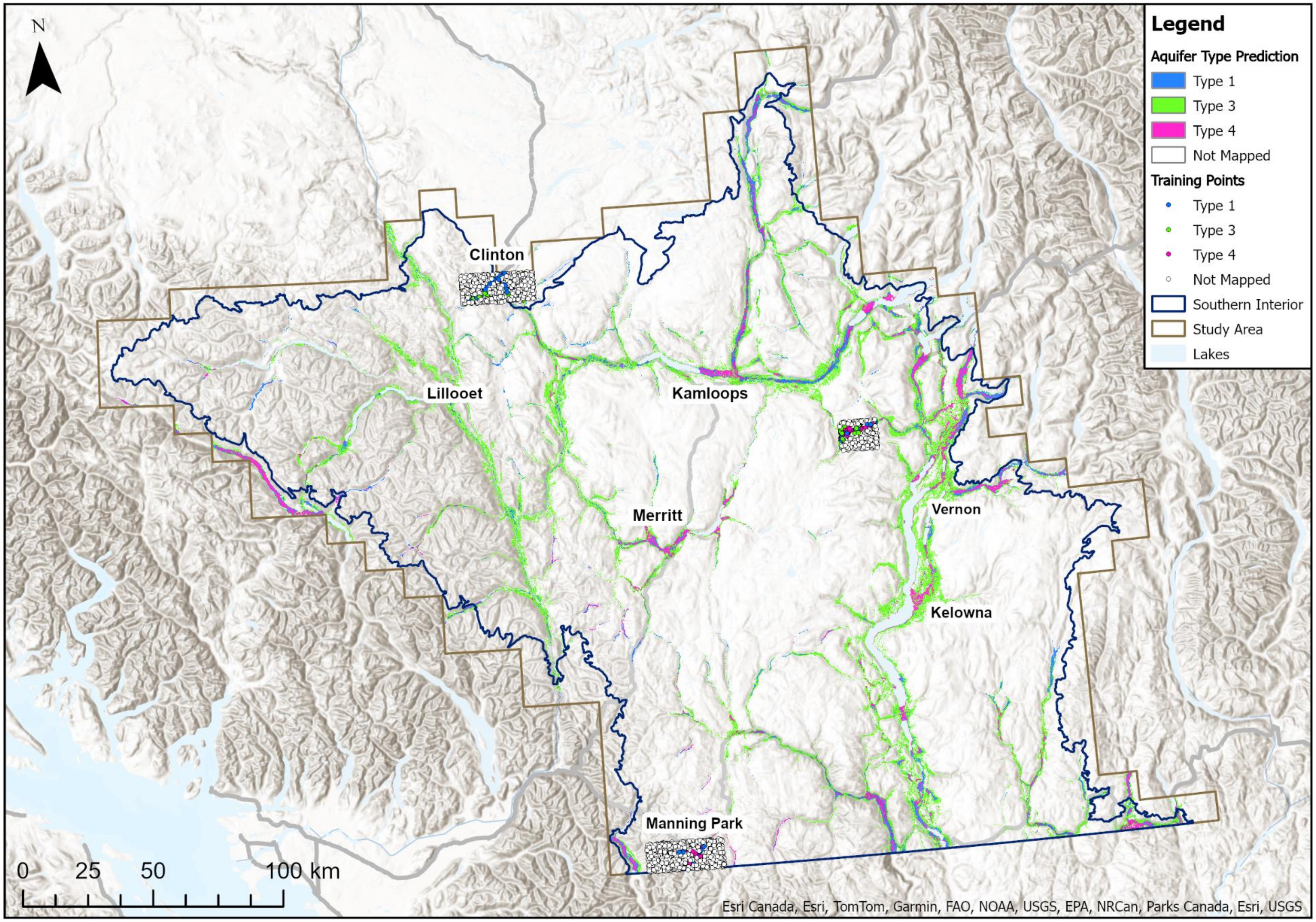


Figure 19: Aquifer Type predictions across the Southern Interior.

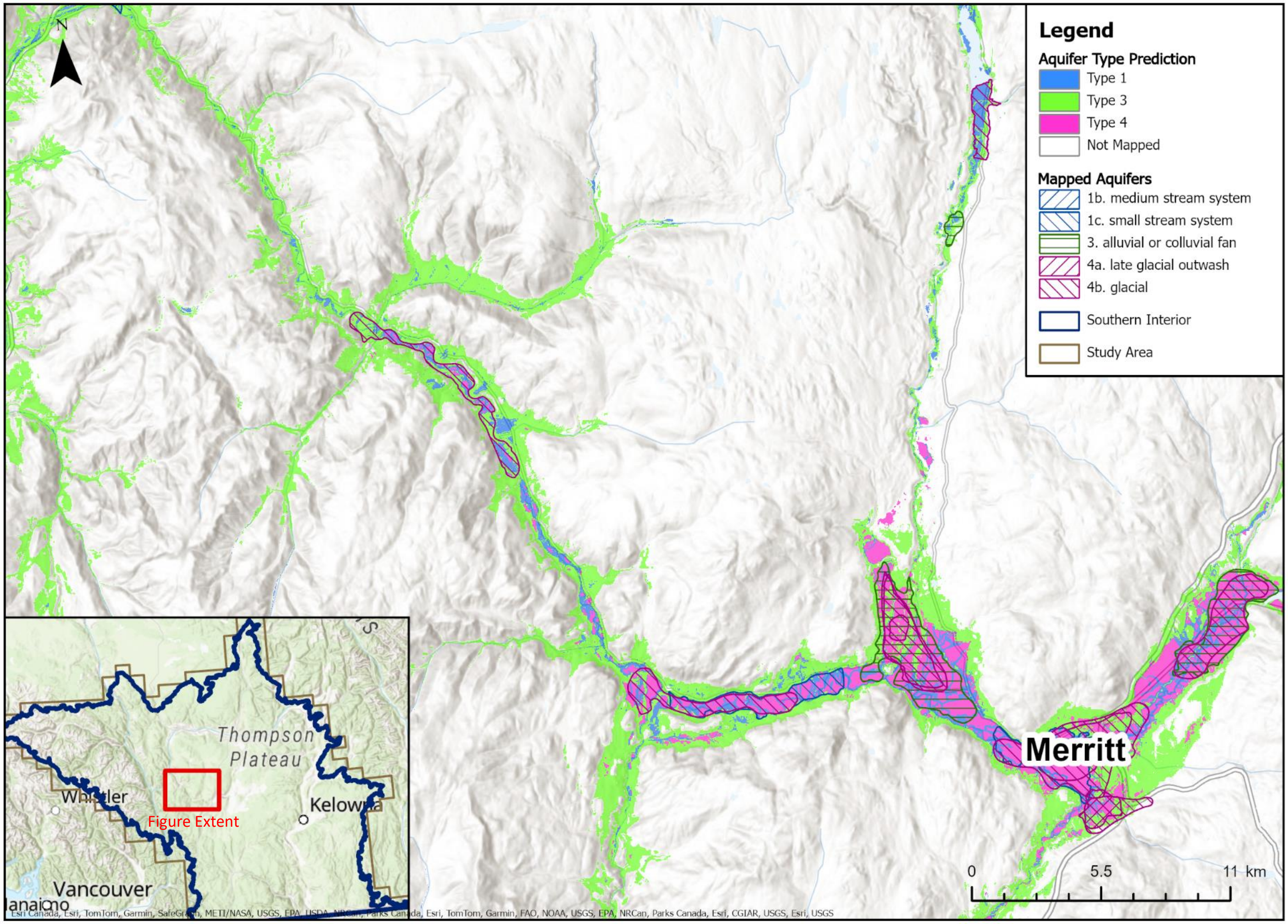


Figure 20: Aquifer Type predictions compared to mapped aquifer polygons in the northwest portion of the study area, including the Thompson Plateau.

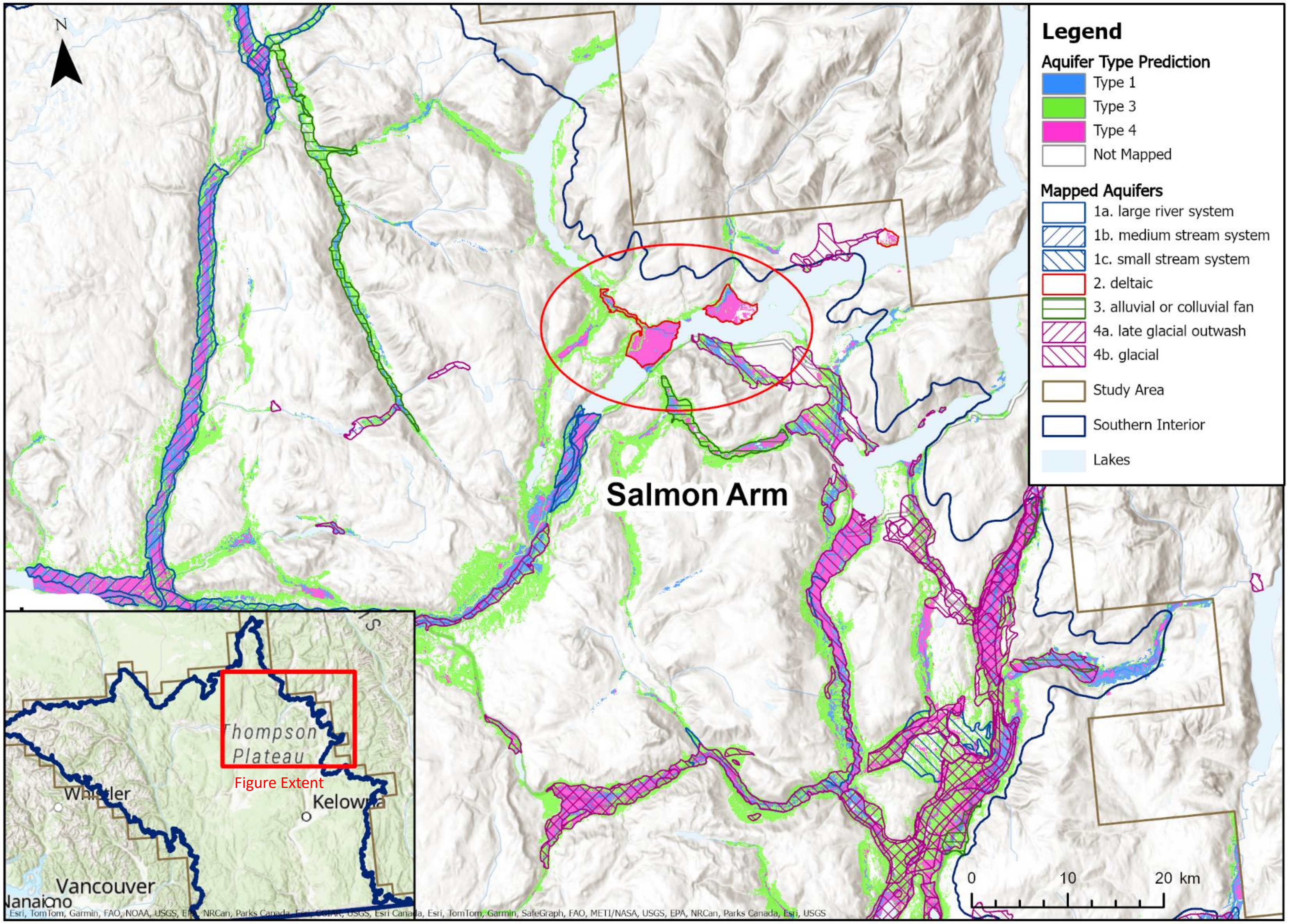


Figure 21: Aquifer Type predictions compared to mapped aquifer polygons near Kamloops, Shuswap Lake and Armstrong.

3.4.6 Thickness Class

It was identified by Madrone during training point attribution that it is challenging to accurately identify the valley fill thickness using a desktop approach. Madrone only attributed points in areas where they had sufficient confidence to describe the thickness and therefore only 150 points were attributed.

The internal accuracy of the balanced model was 99% for the mapsheet areas with training points. Outside of the training mapsheets, the predicted map does not appear to align with expert knowledge. Our regional knowledge indicates that valleys within the Okanagan area host relatively thick sediment deposits, however the model overpredicted the “medium thickness” category (1-3 m thickness) across much of this area (Figure 22).

This exercise indicated that material thickness may be effectively modelled by machine learning, when high quality training data is provided, but that generating enough training data with high confidence may be challenging with a desktop approach.

It is challenging to generate accurate thickness class estimates using a desktop approach as material thickness tends to vary according to localized patterns rather than regional scale trends. Including borehole and well log data into the training or validation data for this class may improve model results in the future.

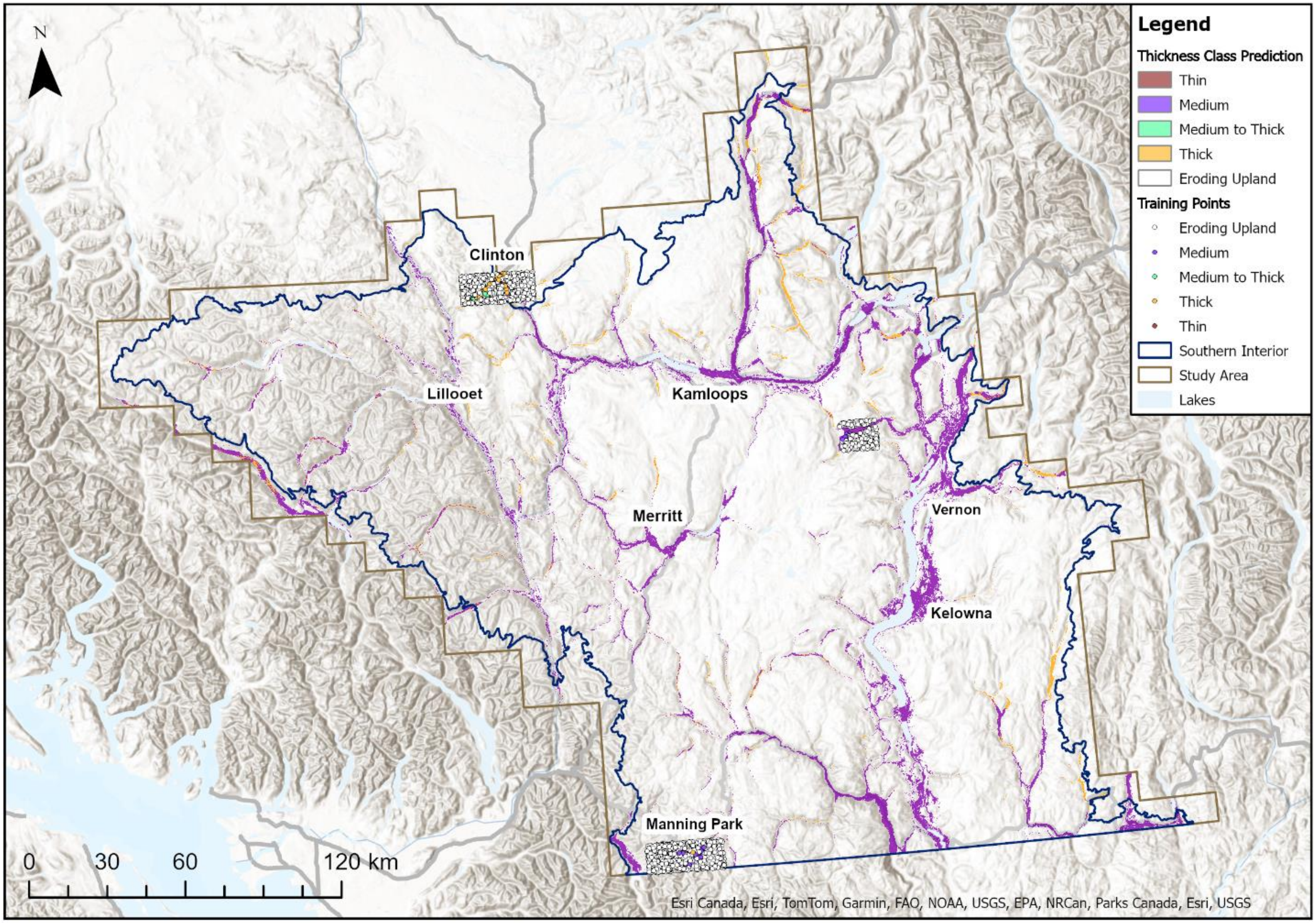


Figure 22: Material thickness predictions across the Southern Interior.

3.4.7 Material Texture and Sorting Class

Similar to thickness, material texture and sorting class were difficult to attribute through a desktop assessment. Approximately 300 points were attributed with each of these classes, typically in fluvial or glaciofluvial deposits. Predictive maps generated by the balanced models are included in Figure 23 and Figure 24 below. As only one point was attributed as very fine texture, that category could not be included in the model. As previously described the training data did not include glaciolacustrine deposits which are fine textured and well sorted. Predictive maps did not reflect the known distribution of these fine textured and well sorted materials.

Visual assessment of the predictive maps suggest that these classes have low accuracy outside of the training point mapsheet areas. For example, based on our expert knowledge, valley fill sediments south of Okanagan Falls have very coarse texture. However, the models predicted fine texture in this area. The training data provided insufficient for the model to discern between different valley bottom textures. The model also appears to over predict poorly sorted materials, as we understand that valley bottoms tend to have fluvial of glaciofluvial materials which are moderately to well sorted. Sediment textures and sorting tend to be landform level occurrences, rather than landscape level making it difficult for the model to identify patterns.

We recommend that attribution of these categories is not continued as it was difficult to develop high confidence training data, and these classes are expected to have relatively lower value to decision makers.

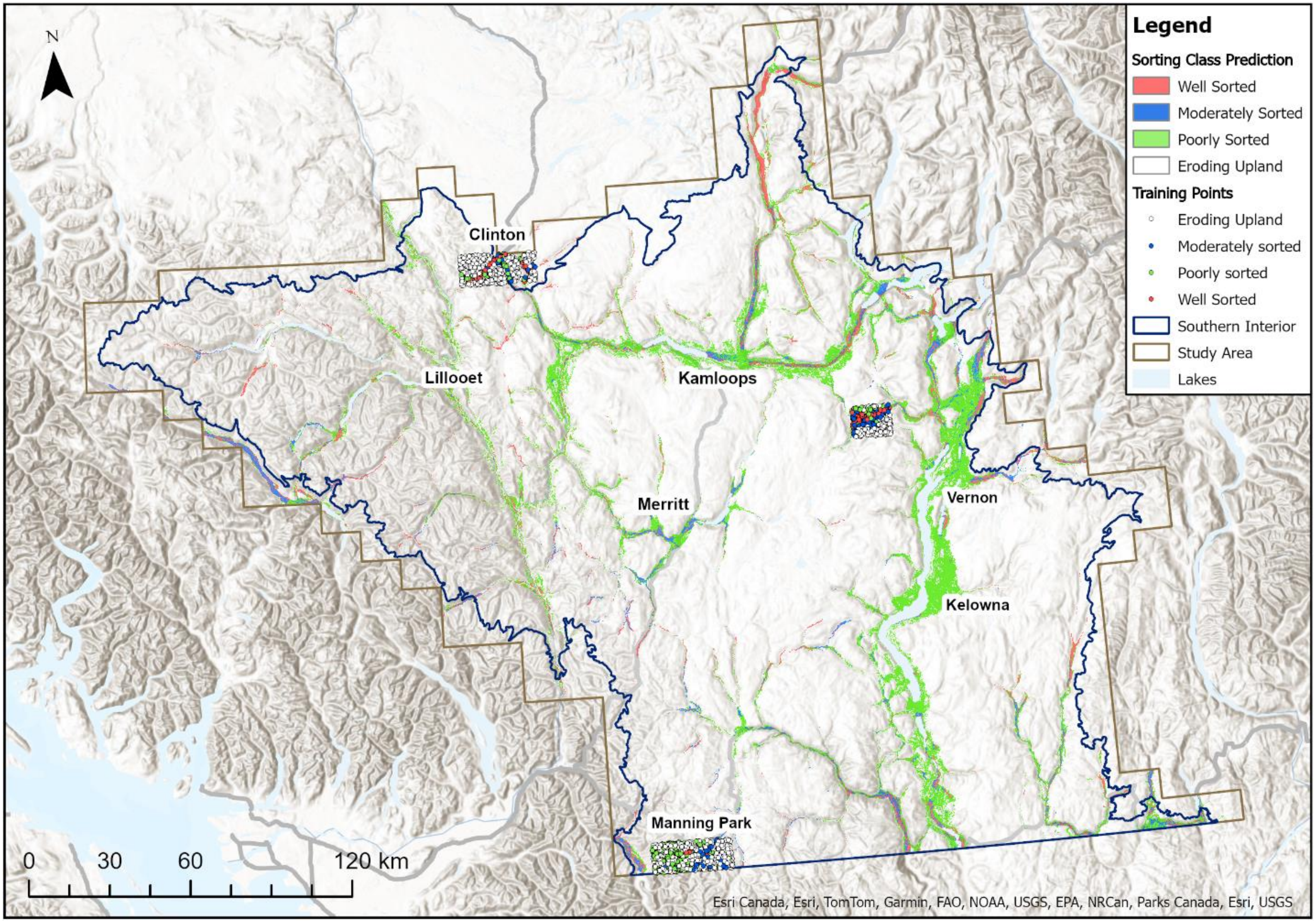


Figure 23: Sorting Class predictions across the Southern Interior.

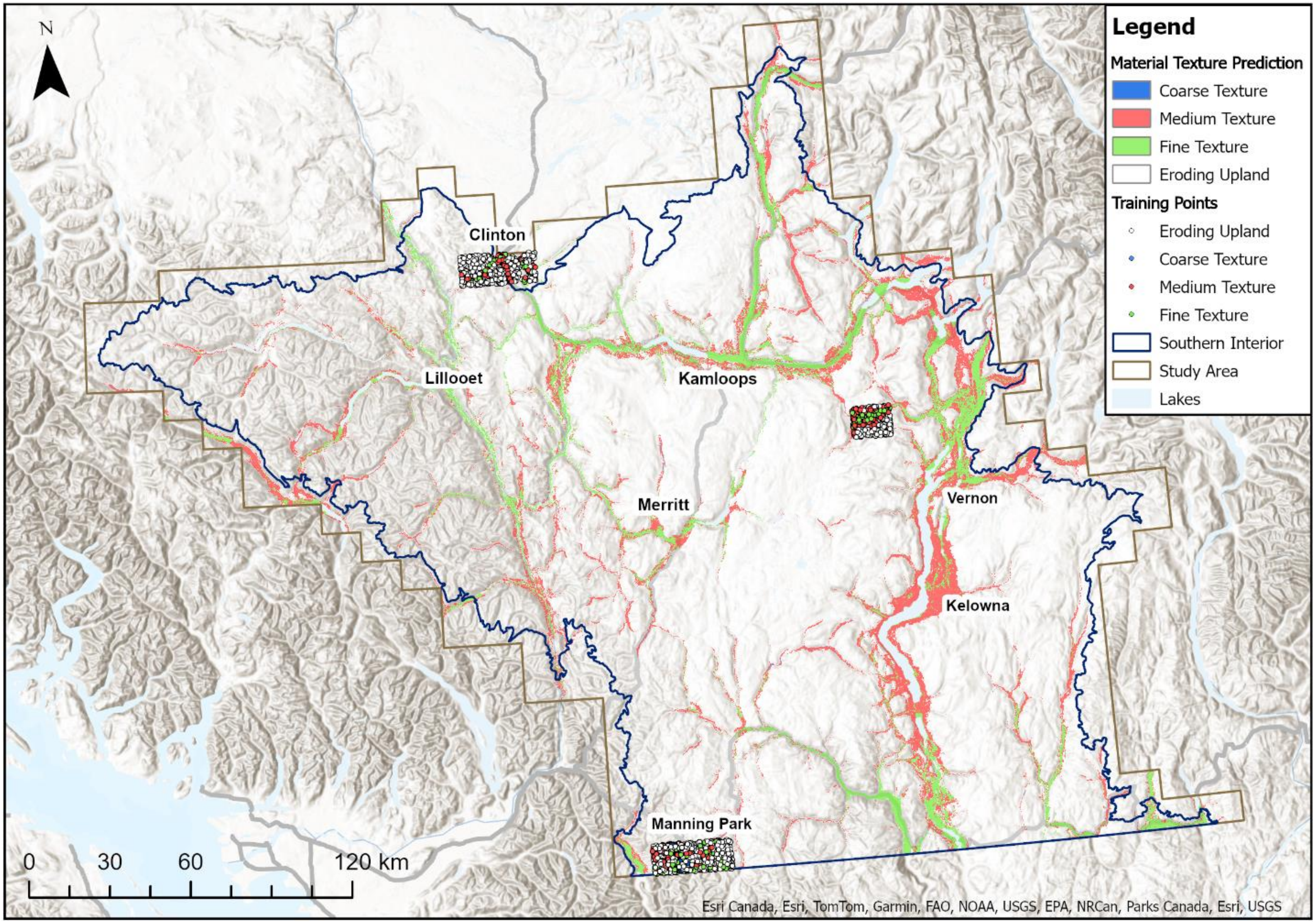


Figure 24: Material Texture predictions across the Southern Interior

4. LEARNINGS AND RECOMMENDATIONS

Classification:

- Predictions of valley fill over the study area presents reasonable patterns in the major valley systems. Much of the distribution is in agreement with aquifer and quaternary fill layers but provides a greater level of detail in some areas. The layer also highlights some areas of contiguous fill and connectivity that are missed by the quaternary fill layer.
- Areas of mountains and bedrock-controlled slopes are appropriately predicted in the eroding upland.
- Within the Okanagan-Thompson Plateau the division of the landscape into eroding upland and valley fill is less clear. A third broad landform category could be explored to capture large areas of generally hummocky glacial deposits such as kame/esker, drumlinzed till, and kame/kettle topography, such as the Lake of the Woods area east of Merrit. It is unclear if this is due to mapper confusion or model confusion. The mapping instructions for these areas was not provided and a third category of thick materials in the upland may lead to better mapping of these thick materials on upland surfaces.
- The surface expression and surficial material classes were only attributed within the areas of valley fill, and therefore predictive maps have limited attribution. Natural resource decision makers are expected to benefit from maps of these classes more extensively across the landscape. Therefore, expanding these classes to fully classify the landscape is recommended.
- Madrone identified that several aquifer subtypes classes were difficult to differentiate through desktop review, particularly the 4a and 4b subtypes which are glacially derived and highly variable in stratigraphy and depth. In the next phase, the aquifer type and subtype classes will be modified into a single class where only type 1 aquifers will be subdivided into subtype 1a and other type 1 aquifers, with subtypes 1b and 1c aquifers lumped together.
- Madrone also identified that the sorting class, material texture and material thickness classes were difficult to attribute with high certainty through desktop review alone. It is recommended that the sorting class and material texture classes are not carried through to the next phases of the project as they are not expected to provide high value to decision makers. Thickness class should be further evaluated to see if high confidence training data can be developed.
- As the landform subclasses have the most categories and the greatest sample imbalance between categories, it was the most challenging to model. A recommendation for the next phase of the project is to create more training points within the valley fill and to strategically sample the minority classes. The classification could also be reviewed to see if the number of categories is required.
- It was identified through the project that scale specific guidance and descriptions are important for consistent application of the classifications across landscapes and users. Specific instructions and examples are beneficial to support consistent mapping and classification.

Training Points:

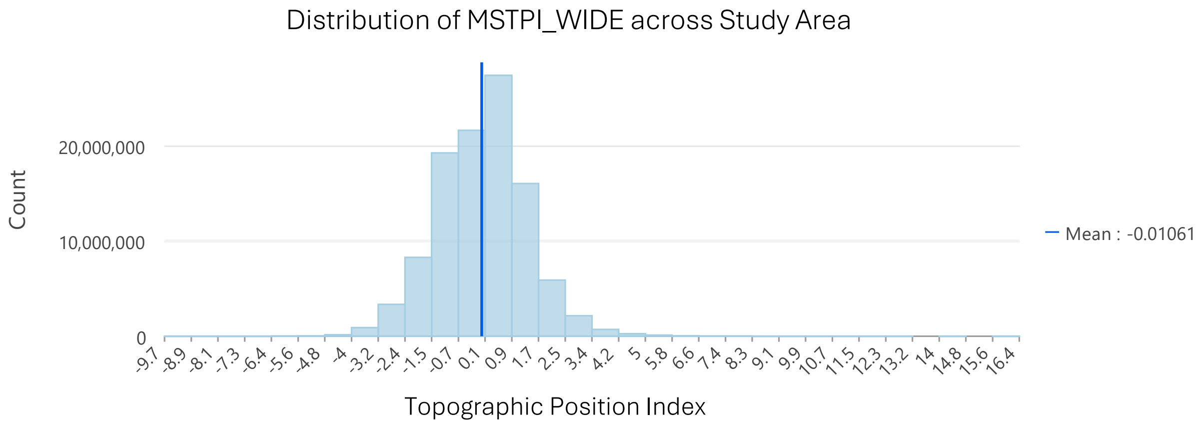
- The training points developed to date are located within mapsheets near the outer edges of the study area. To improve model predictions more training points across other areas of the Southern Interior are needed. Particularly, in the Thompson Plateau, which has different valley and upland patterns.
- The mapsheets did not cover areas that describe the distribution of glaciolacustrine deposits that accumulated in major glacial lakes in the Southern Interior, such as glacial Lake Penticton and glacial Lake Thompson (Fulton, 1969) and thus the presence of these materials were not

predicted, including well-known deposits in the Naramata bench, Kamloops, and Summerland areas. These deposits are fine textured and well sorted and their absence from the training dataset also impacted the results of the material texture and sorting classes.

- The imbalance between the eroding upland and valley fill categories in the broad landform class, creates high sample imbalance in the other classes, as the categories are only attributed within the valley fill class. Minority classes which are represented by a small sample of training points are difficult for models to interpret. Strategically capturing minority features during training point development would benefit predictive modelling. It is recommended that the next phase of training point development is designed to create a more balanced training data.

Modelling:

- The ModelMap R package, which was used for random forest modelling, was recently discontinued. It is recommended that future phases of the project adopt a more sophisticated modeling approach that includes K-fold cross validation, sensitivity analyses, and model optimization. The caret R package is an example of a tool that could accommodate that workflow.
- As many of the covariate rasters are highly correlated, it is recommended that covariate reduction techniques such as variance inflation factor (VIF) and recursive feature elimination are applied during future modelling.
- The deglaciation patterns and modes of deposition are variable throughout the Southern Interior. We did not provide the model any covariate layers with information about this glacial history. Adding data to the model that highlights glacial patterns may improve model performance in the future. Further, distributing training data across the study area rather than in dense mapsheets, may help the model identify these patterns.
- Predictive modelling assumes that landscape features will consistently share similar attributes and covariate values. Models will most accurately identify physiographic relationships for areas with similar landscapes, climate, glacial history, geology, ecosystems, etc. Models are expected to be more consistent within a watershed, Ecosection, Ecoprovince, or other related landscape units. Landscape units with similar covariate distributions are described as “rule sheds” (Bulmer et al, 2016). This modelling pilot revealed that when designing the sampling plan for training data it is important to be mindful of the rule sheds captured by the training points. For example, use of dense mapsheets was beneficial to capturing a variety of features within an area but was only representative of a handful of rule sheds within the study area. When generating predictive maps, it’s important to be mindful of not extending predictions beyond the rule sheds represented by your training point distribution. For example, Figure 25 compares the distribution of the MSTPI_WIDE covariate between the training point mapsheets and the entire study area. This demonstrates that the mapsheets selected for training data do not capture the variability of the study area as a whole.
- Without a validation dataset or training points throughout the study area, the model performance can only be evaluated qualitatively in the training mapsheets. As the project progresses towards model optimization and model finalization, it is recommended that both the training dataset and a validation dataset are developed across the study area.
- Model outputs and predictive maps should be reviewed by geomorphologists with expertise in the surficial geology and glacial history of the Southern Interior prior to finalizing.



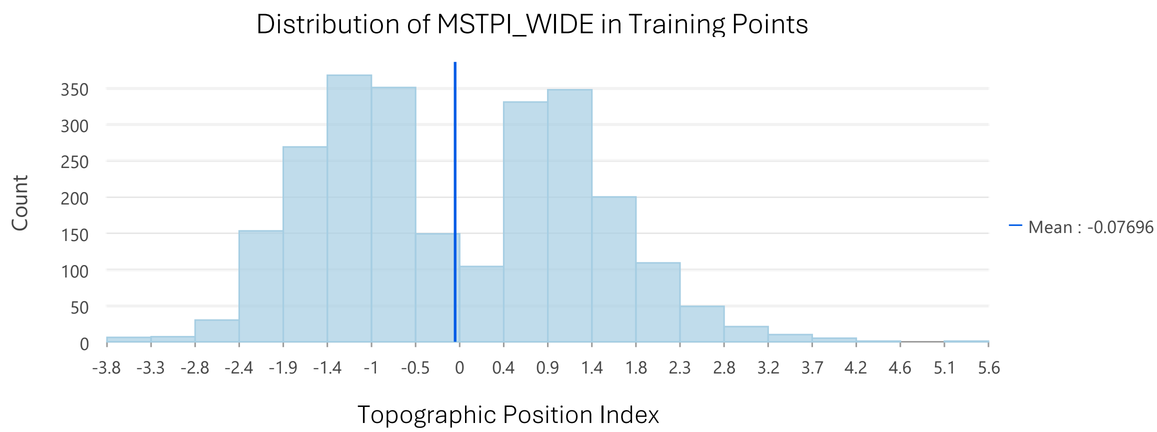


Figure 25: Distribution of Multiscale Topographic Position Index Covariate (MSTPI_WIDE) values across the study area (top) and across the training points (bottom).

5. <u>CONCLUSIONS</u>

The main conclusions of Phase 2 of this project were:

- Within a 1:20 000 NTS mapsheet with 500 training points, random forest predictions yield high internal model accurately rates (98-99%).
- Training and validation data distributed throughout the study area is needed to more effectively train models and validate results. Accuracy rates reflect predictions within the pilot map sheets and do not provide any information about accuracy in other parts of the study area. The random point distribution within a mapsheet poorly captured minority categories that have a small spatial extent (e.g. fine texture material).
- The combination of attributed training points and random forest modelling is a cost-effective way to map simple discrete categories that are discernable from image interpretation. The division of the landscape into two broad landform categories (eroding upland and valley fill) was effective except in upland depositional areas such as in the Thompson Plateau. Addition of a third Broad Landform category to upland depositional features is recommended for the next phase.

- Predictive maps of valley fill distribution are informative for identifying areas of the landscape with significant sediment accumulations, including the identification of areas with aquifer potential where more detailed aquifer mapping or exploration could be conducted.
- Our preliminary predictive maps indicate that within areas mapped by the provincial Quaternary Geology layer there is a diverse range of landforms, surficial materials, expressions, aquifer types, deposit thicknesses, and material characteristics. This limitation is important to note for decision makers who may require more detailed landscape classification.
- Predictive mapping results will likely support decision makers with identifying areas where field data collection or detailed mapping may be valuable.
- With appropriately distributed and detailed training data, models can effectively subdivide the valley fill areas into smaller landforms and potential aquifer types.
- When extending this mapping into other Ecoprovinces in B.C. other gaps in the mapping and modelling process may be identified.
- Model results and categories should be critically reviewed by geomorphologists for each Ecoprovince run.

REFERENCES

Bader-El-Den, M., E. Teitei, and T. Perry. 2018. Biased random forest for dealing with the class imbalance problem. *IEEE transactions on neural networks and learning systems*. 30(7):2163–2172.

BC Geologic Survey. 2023. Geology Quaternary Alluvium and Cover. Available online at: https://catalogue.data.gov.bc.ca/dataset/geology-quaternary-alluvium-and-cover.

B.C. Ministry of Forests and Range and B.C. Ministry of Environment, eds. 2010. *Field manual for describing terrestrial ecosystems*. 2nd ed. Province of British Columbia, Victoria. 1 p.

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

Brenning, A., 2008. Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models. In J. Boehner, T. Blaschke and L. Montanarella (eds.), SAGA - Seconds Out (=Hamburger Beitraege zur Physischen Geographie und Landschaftsoekologie, vol. 19), p. 23-32.

Bulmer, C., M. G. Schmidt, B. Heung, C. Scarpone, J. Zhang, D. Filatow, M. Finvers, S. Berch, and S. Smith. 2016. Improved Soil Mapping in British Columbia, Canada, with Legacy Soil Data and Random Forest. P. 291–303 in *Digital Soil Mapping Across Paradigms, Scales and Boundaries*, Springer Environmental Science and Engineering. Zhang, G.-L., D. Brus, F. Liu, X.-D. Song, and P. Lagacherie (eds.). Springer Singapore, Singapore. Available online at: http://link.springer.com/10.1007/978-981-10-0415-5_24; last accessed December 7, 2022.

Chan, J. Y., S. M. Leow, K. T. Bea, W. K. Cheng, S. W. Phoong, Z.-W. Hong, and Y.-L. Chen. 2022. Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. *Mathematics*. 10(8).

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-sampling Technique. *jair*. 16:321–357.

Dermarchi, D. 2011. *An Introduction to the Ecoregions of British Columbia*. British Columbia Ministry of Environment, Victoria, British Columbia. Available online at: https://www2.gov.bc.ca/assets/gov/environment/plants-animals-and-ecosystems/ecosystems/broad-ecosystem/an_introduction_to_the_ecoregions_of_british_columbia.pdf.

Filatow, D., G. Harvey, T. Carswell, and E. Cameron. 2020. *Predictive Wetland Mapping of the Williston Drainage Basin: Update*. Knowledge Management Branch, BC Ministry of Environment and Climate Change Strategy, Victoria, BC. Available online at: https://a100.gov.bc.ca/pub/acat/public/viewReport.do?reportId=58802.

Freemen, E. A., T. S. Frescino, and G. G. Moisen. 2022. ModelMap: an R Package for Model Creation and Map Production. Available online at: https://rdrr.io/cran/ModelMap/f/inst/doc/VModelMap.pdf.

Fulton, R. J. (1969). Glacial Lake history, southern interior plateau, British Columbia. Geological Survey of Canada, Paper, 69-37, 14. https://doi.org/10.4095/106455

Hijmans, R (2024). *terra: Spatial Data Analysis*. R package version 1.7-71, https://github.com/rspatial/terra.

Howes, D. E., and E. Kenk. 1997. *Terrain Classification System for British Columbia (Version 2).* Resources Inventory Branch Ministry of Environment, Lands and Parks., Victoria, BC.

Johann de Jong. 2017. Feature selection, cross-validation and data leakage. *Johann de Jong*. Available online at: https://johanndejong.wordpress.com/2017/08/06/feature-selection-cross-validation-and-data-leakage/; last accessed March 30, 2023.

Misra, P., and A. S. Yadav. 2020. Improving the classification accuracy using recursive feature elimination with cross-validation. *Int. J. Emerg. Technol*. 11(3):659–665.

More, A., and D. P. Rana. 2017. Review of random forest classification techniques to resolve data imbalance. P. 72–78 in *2017 1st International Conference on Intelligent Systems and Information Management (ICISIM)*, IEEE.

Province of British Columbia. No date. Freshwater Atlas Stream Network. *British Columbia Data Catalogue*. Available online at: https://catalogue.data.gov.bc.ca/dataset/92344413-8035-4c08-b996-65a9b3f62fca; last accessed November 23, 2021.

Province of British Columbia. 2016. Determining the Likelihood of Hydraulic Connection – Guidance for the Purpose of Apportioning Demand from Diversion of Groundwater on Streams. Version 1.0. Water Science Series, WSS2016-01. Prov. B.C., Victoria B.C. http://www2.gov.bc.ca/gov/content/environment/air-landwater/water/water-science-data/water-science-series.

Resources Information Standards Committee (RISC). 2023. Terrestrial Ecosystem Information Digital Data Submission Standard: Database and GIS Data Standards. Knowledge Management Branch, BC Ministry of Water, Land and Resource Stewardship. Victoria, BC. https://www2.gov.bc.ca/assets/gov/environment/natural-resource-stewardship/nr-laws-policy/risc/tei_digital_submission_standards_v3.pdf

Todd, J. and D. Filatow, 2024. Surficial Mapping British Columbia's Landforms to Support Groundwater Understanding, Phase 1: Vanderhoof Watershed Group. Water Science Series, WSS2024-01. Province of British Columbia., Victoria. https://a100.gov.bc.ca/pub/acat/documents/r62950/SurficialMappingPhase1_1710261481517_ED78912F0E.pdf

Wang, T., C. Mahony, D. Spittlehouse, and A. Hamann. 2022. *ClimateBC v7.30*. University of British Columbia. Available online at: https://climatebc.ca/.

Wei, M., D. Allen, A. Kohut, S. Grasby, K. Ronneseth and B. Turner. 2009. Understanding the types of aquifers in the Canadian Cordillera Hydrogeological Region to better manage and protect groundwater. *Streamline*, 13(1), p. 8.

APPENDIX A: MADRONE METHODS SUMMARY

MADRONE
environmental services ltd.

1081 Canada Ave
Duncan, BC V9L 1V2
p. 250.746.5545
f. 250.746.5850

#1 – 30435 Progressive Way
Abbotsford, BC V2T 6Z1
p. 604.504.1972
f. 604.504.1912

info@madrone.ca
www.madrone.ca

October 18, 2022

Deepa Filatow
Ministry of Environment and Climate Change Strategy (MoECCS)
Deepa.Filatow@gov.bc.ca

Dear Ms. Filatow,

RE: Surficial Material Mapping in Support of Groundwater Management – 2022 Update

Contract #: GS17JHQ092
WO#: WO21JHQ020
TEI SR#: 4

This letter is a summary of the method used to attribute the randomized points, including outline of how decisions were made during attribution. The following map sheets have been completed for the above forementioned contract, for a total of 2500 points (500 per map sheet).

- 092H007 & 006 – Manning Park
- 082L042 – Westwold
- 092P002 & 003 – Clinton

ArcGIS setup for Mappers

The following layers were available to review while mapping:

- *Groundwater Aquifers*
- *Groundwater Wells*
- *Geology Quaternary Alluvium and Cover*
- *Surficial Units* and *Surficial Linear Features*
- *BC Glacial Features*
- Previous maps at various scales

We found that the layers that were the most helpful were:

- *Groundwater Aquifers*
- *Groundwater Wells*
- *Geology Quaternary Alluvium and Cover*

Model Fields

We began this process with a list of specific fields that we thought would be helpful for the model:

- Broad Landform Type
- Surficial Material
- Surficial Expression
- Material Texture Class
- Sorting Class
- Material Thickness Class
- Layers and Bedding Class
- Aquifer Class
- Aquifer Subtype
- Aeolian Deposit
- Valley Fill Subclass
- Representative of Area
- Area of Representation
- Field Verified Radius of Representation
- Point Quality
- Sample Placement
- Sample Type

As the project progressed and discussions occurred, focus was spent on the following fields:

- Broad Landform Type
- Surficial Material
- Surficial Expression
- Aquifer Class
- Aquifer Subtype
- Valley Fill Subclass
- Representative of Area
- Area of Representation
- Point Quality

We found that there were issues with these fields:

- Material Texture Class – too assumptive at desktop scale
- Sorting Class – too assumptive at desktop scale
- Material Thickness Class – too assumptive at desktop scale
- Layers and Bedding Class – too assumptive at desktop scale
- Aeolian Deposit – location specific
- Field Verified Radius of Representation – desktop study, not applicable

We modified or partially used these fields:

- Aquifer Subtype – anything not 1A would be too assumptive at desktop scale; points were this was already attributed were changed to <Not Mapped>, or <Not Applicable> if their Aquifer Class was <Not Mapped>.
- Valley Fill Subclass – some categories could also work for Eroding Uplands or work better for Eroding Uplands rather than Valley Fill. Because we were focused on Valley Fill only, instead of blanket attribution across landscape, there was a period of adjusting required when considering how to attribute a given domain.

Attribution

Some of the decisions made for a given field require the [future] users to have a good understanding of glacial history, geomorphic processes, surficial material, hydrogeology, and surface water-groundwater interactions. As such, the steps provide insight on the decision-making process but do not make the actual point attribution.

1. Review the study area with only the imagery and contours on, making note of any significant valleys that are connected and may provide groundwater resources. Make note of any hanging valleys that traditionally would exclude a continuous groundwater resource.
2. Start by attributing *Broad Landform Type* field by differentiating between Valley Fill or Eroding Upland:
    a. Zoom to one map sheet, in the range of 1:50,000 to 1:100,000. Turn on the *Groundwater Aquifers* layer and the points symbolized by elevation groups. For points that are outside of the *Groundwater Aquifers* layer and at an elevation greater than 1,500 m, classification of *Broad Landform Type* as likely to be Eroding Upland. Exclude any points that occur in valleys with a direct connection to a groundwater resource, this can include valleys at upper elevations if the valley connection is contiguous.
        i. When using the *Groundwater Aquifers*, layer, use with a degree of caution, it is meant as a guideline, but it is not considered exact in its spatial accuracy. Your professional discretion on the extent of the valley fill will most likely override the *Groundwater Aquifers* layer.
        ii. Another layer that is useful in delineating the valley fill versus the eroding upland is the *Groundwater Wells* layer. This is a point file you can use in tandem with the *Groundwater Aquifers* layer; it is also worth looking at it separately and noting any clusters of wells outside of the aquifer extent layer.
        iii. *Geology Quaternary Alluvium and Cover* is another layer that can get you started, but we found it is less helpful than the *Groundwater Aquifers* layer.
    b. Once you have completed this exercise with the elevation greater than 1500 m, use the symbols for the 1200-to-1500-meter range. Pay close attention to avoid populating any fans that extend to the valley floor. Those must be considered separate at this point.

   c. Review any points that are outside of the *Groundwater Aquifers* layer and are present at an elevation less than 1200 m. Depending on the material and geomorphology, these points may not be considered in the eroding upland category. An example of points would be those on fans, terraces, or benches.
3. Once the majority of the sample points have been assigned as either Eroding Upland or Valley Fill, you can do a batch update of Eroding Upland points, to fill in *Aquifer Class* as 'Not Mapped' and *Aquifer Subtype* as 'Not Applicable'.
4. If you follow this process, there will likely be a distinctive band of sample points along the transition zone between Eroding Upland and Valley Fill. To determine if the points fall within either of those broad landform types you should zoom in to a scale of approximately 1:5,000 up to 1:25,000. At this scale in this region, you will likely encounter glacial terraces or fans (colluvium or alluvial). Depending on the geomorphic feature, you will attribute the point as either Eroding Upland or Valley Fill.
   a. In the case of fans, there may be multiple points within that unit, however the attribution of those points may vary depending on the location of the point relative to the fan's geometry. Fans produce an opportunity to add an opportunistic point.
5. Once all points have their Broad Landform Type attributed, further attribution can occur for those points identified as Valley Fill. While the intention is to fill in every data point across all fields, we found that accuracy of some of the fields was quite low for desktop analysis; unless we had high confidence in a data point's attribute, the value would be left as null.

<u>General Tendencies and Decisions regarding Attribution</u>

To visual what was being attributed, we created copies of the working file, symbolized multiple ways so we could toggle between seeing all Surficial Material calls at a glance, then to Broad Landform Type, to Aquifer Class. This helped with consistency across a map sheet and between regions (see **Figure 1**).

1. Surficial Material - Depending on your location within BC, surficial material found in the valleys tend to be fluvial, glaciofluvial, glaciolacustrine, or organic. Glacial till may also be present depending on the size of the valley. Colluvium within Valley Fill is rare and tends to be associated with significant mass movement into the valley. We rarely used 'undifferentiated' as a surficial material call, and 'not mapped' was reserved for either waterbodies or significant points of human disturbance such as large infrastructure.
2. Surficial expression – we utilized provincial standards to assign appropriate expressions based on the surficial material and the geomorphic feature the point is located on, i.e., fluvial plain rather than fluvial undulating. Surface expressions such as hummocky, scarp, and geomorphons, were used exclusively in the Eroding Uplands. While undulating is found predominantly within the Eroding Uplands, where it was used within the Valley Fill were in areas that were associated with mass movement or buried features.

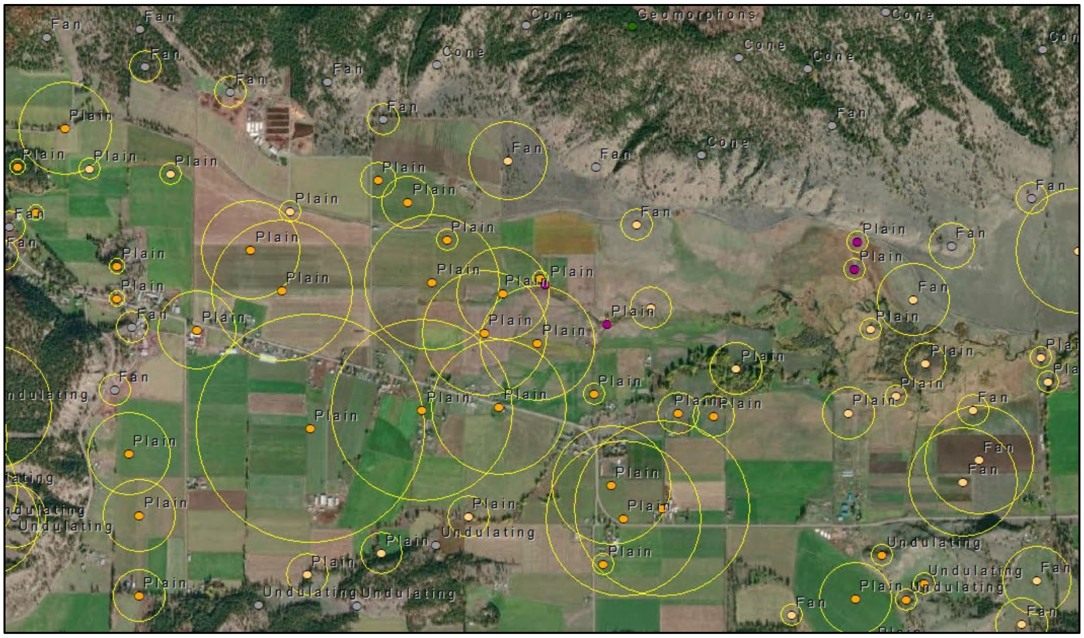


FIGURE 1: EXAMPLE OF SAME DATA SYMBOLIZED IN DIFFERENT WAYS CONCURRENTLY TO SHOW CONSISTENCY AND TRENDS ACROSS AN AREA; RADIUS OF REPRESENTATION (ORANGE BUBBLES), SURFICAL MATERIAL (COLOUR CODED POINTS), AND SURFICAL EXPRESSION (LABELLED).

3. Material Thickness - almost 150 data points were attributed with the material thickness. And when they were, they were in areas where the mapper felt confident in the call because of their experience in the area or a very distinct and recognizable feature.
4. Material Texture - just over 300 data points were attributed with the material texture; these points tend to be of fluvial or glacial fluvial origin where the texture could be confidently discerned from the desktop analysis of the features.
5. Sorting class - just over 300 data points were attributed with a sorting class; similar to material texture, the points tend to be of fluvial or glacial fluvial origin. Common attribute combinations used:
    a. Fluvial plains = well sorted
    b. Glaciofluvial plains and tills = moderately sorted
    c. Colluvium and fans = poorly sorted
6. Layers and Bedding – this field was attempted but abandoned after discussion within the working group; this attribute is best determined at either field scale or with site specific data. It also was not clear on how differentiating between layers vs bedding would be beneficial – this topic is on hold.
7. Aquifer Class was assigned for every Valley Fill data point. The surficial material and expression were cross referenced with the Aquifer Class to determine the most appropriate classification.
8. Aquifer Subtype - After discussion within the working group, the subtype was only attributed for Aquifer Class 1 where the subtype was 1A. All other values for subtype were attributed as not mapped or not attributed.

9. Radius of representation - utilizing a feature in ArcGIS that allows a layer's symbology to change based on a field's value, as well as a buffer layer in increments of 50 meters, all Valley Fill points were given a radius of representation with an approximate confidence up to 25 meters.
    a. If a data point was close to a feature where the point would no longer be 100% representative at a scale of ~1:25,000, the representative value is the maximum radius without overlapping that different feature.
    b. If the data point has radius of representation less than 10 m, the point was not considered representative.
    c. The radius of representation is unique to that data point and is not provided as a range.
    d. A point's radius could cross the radius of another point if the points were representative of the same attribution.

<u>Recommendations</u>

- When attributing a data point in the transition zone between points that are distinctly Valley Fill or Eroding Uplands, it is beneficial to work from the Valley Fill side towards the Eroding Upland, keeping in mind the relationship of the overall feature relative to an underlying aquifer. This is especially important if the feature is a colluvial fan that transitions to a cone as the slope increases towards the uplands.
- The *Groundwater Aquifer* layer was used primarily for its spatial extent rather than the aquifer type; the use of a deltaic aquifer in that layer does not align with the standard bioterrain unit for delta. Prior to moving into coastal areas of BC, a review of the delta definition in this framework is required.

<u>Future Considerations</u>

- While given the opportunity for opportunistic points, it was not fully exercised in these map sheets. After discussion in the working group, we feel like those opportunistic points can provide an opportunity for refining the model in complex terrain.
- Features that are long and thin (i.e., avalanche track, debris slides, lower order streams) can be difficult to attribute and provide little value to the model unless opportunistic points are added to delineate the feature to a larger spatial extent.
    - This is especially true when it is a lower order stream, with potential Aquifer Subclass of 1C.
- Surficial expressions can be limited in their value to the model depending on the scale of the feature.
    - For example, a glaciofluvial scarp may be present in a map sheet, however, to capture that in a point, the feature would have to have a radius of representation greater than 10 m (to be considered representative), and have additional opportunistic points added to show linear extent. However, a non representative point, or a long and thin linear feature

is not likely to represent significant source of groundwater at the model's scale, even though the feature may hold an ecological or resource value.

Yours truly,

MADRONE ENVIRONMENTAL SERVICES LTD.

Roberta Adams, M.Sc., P.Geo.
Senior Geoscientist

APPENDIX B: COVARIATE RASTER DESCRIPTIONS

All covariate rasters were generated using default parameters in SAGA GIS except for the following:

Height Above Channel (HACH)

A channel network raster was developed as the input to generate the height above channel and Channel Network Base Level. A simple sensitivity analysis was conducted to assess which channel network parameters generated the best height above channel raster for model performance. An initiation threshold of 20,000 was selected and a minimum segment length of 10 (map units, 250 meters) was selected.

Topographic Position Index (TPI)

TPI was calculated with multiple different radiuses, 500 m, 1000 m, 2000 m, 5000 m. To create TPI rasters with large radius, the resolution of the DEM was decreased to one hectare with a spatial averaging method. The TPI raster was created at the hectare scale and then sub-sampled back to 25-meter resolution with a nearest neighbours' method. This was done to reduce the computational requirement of these layers.

Multi-scale Topographic Position Index (MSTPI)

The implementation calculates the Topographic Position Index (TPI) for different scales and integrates these into one single grid. The hierarchical integration is achieved by starting with the standardized TPI values of the largest scale, then adding standardized values from smaller scales where the (absolute) values from the smaller scale exceed those from the larger scale.

Multi-scale Topographic Position Index (MSTPI) was generated by calculating TPI at three scales. Two MSTPI rasters were generated, the first had used scales between 8 and 40 grid cells and the second used scales between 40 and 200 grid cells.

APPENDIX C: DOMAIN TABLES

The domain tables used in this study are preliminary and does not represent a final classification. The domain tables will be further refined in future studies.

Broad Landform

Code	Short Name	Description
Eroding Upland	Eroding Upland	Continuous areas of erosion and scour such as mountains, hills and plateaus dominated by bedrock, veneers and blankets.
Valley Fill	Valley Fill	Regions of thick, unconsolidated sediment accumulation in the landscape, and characterized by extensive depositional landforms. Extensive in lowlands, depressions, and northern plains within B.C. In the Study Area, these areas are dominated by alluvium (river deposits) and sediments deposited by glacial ice, glacial lakes, and glacial meltwater. They include thick materials reworked by gravity and wind such as eolian veneers, dunes and landslide deposits. Along valley side walls the valley fill category includes materials deposited in fans, cones, and aprons where steep slopes and tributary valleys join mainstem valley fill.
Not Mapped	Not Mapped	

Developed for this study. Constructional landforms based on Howes and Kenk (1997). Subject to edits following model results and subsequent work.

Landform Subclass

Code	Short Name	Description
Coastal	Coastal Landforms	Wave, tidal and wind deposited and reworked sediments. Beaches, dunes, estuaries, cliffs typically below 100m elevation.
Large Landslide	Large Landslide Features	Large mass movement features such as landslides, rockslides, rock glaciers, slumps and other mass movement features. These include the head scarp, transportation zone and runout zone/deposit.
Scarp	Scarp Slopes	Steep scarps cut into thick surficial materials, often actively eroding and slumping. Often exposing sequences of deposits or relic environments. Can be found at the boundary of active geological processes.
River	River Landforms	River landforms, flat lying running water deposits with scarps and terraces formed in Holocene times. These are also called fluvial landforms and alluvium.

Code	Short Name	Description
Fan Delta	Fans and Deltas	Gently sloping landforms found where side valleys meet larger valleys or waterbodies. Formed in Holocene by flowing water and waves.
Cone Apron	Cones Aprons and Talus Slopes	Cones, aprons and talus slopes deposited from rapid mass movement and gully processes such as rock fall, debris flows and slides. Commonly coarse textured, bedrock related angular materials. Typically, on active Holocene landforms.
Organic Palustrine	Organic and Palustrine Depressions	Wetlands. Holocene organic and fine sediment accumulations in basins and lowlands. Typically, in areas of high-water table or with stagnant to sluggish water flow.
Ablation	Ablation and Meltwater Landforms	Areas of glacial ablation where stagnant ice and glacial meltwater formed hummocky coarse textured landforms. This includes kames, eskers, drumlins, and kettled topography. High energy and ice contact depositional environments.
Glaciofluvial	Glaciofluvial Landforms	Glaciofluvial meltwater deposited fans, terraces and planes deposited mainly by flowing water from glacial melt.
Glaciolacustrine	Glaciolacustrine Landforms	Stagnant water deposited fine textured flats and blankets.
Marine Flats	Marine Flats Landforms	Deeper water deposited glaciomarine sediments. Typically, low in coarse fragments content. Lower energy depositional environment. Found below the marine limit (isostatic rebound).
Glacial Till	Glacial Till Landforms	Thick, unsorted ice deposited sediments commonly undulating or gently sloping.
Thick Glacial Fill	Thick Glacial Fill	10s to 100s of meters of glacial infilling representing multiple glacial stages. Complex stratigraphy documented in the literature. Horizontally variable sediments from glacial reworking of deposits from proglacial and older glacial steads
Other	Missing Category	This category is temporary while the domain list is being finalized. Use this category when all others don't fit, and the domain table needs to be expanded. Use the comment field to elaborate.

Developed for this study. Adapted from Todd and Filatow (2024).

Surficial Material

Code	Short Name	Description
Till	Till	Materials deposited directly from glaciers.
Colluvium	Colluvium	Products of mass wasting.
Fluvial	Fluvial	River deposits.
Glaciofluvial	Glaciofluvial	Ice contact fluvial materials.
Glaciolacustrine	Glaciolacustrine	Ice contact lacustrine materials.
Marine	Marine	Marine sediments; includes wave deposits.
Eolian	Eolian	Material deposited by wind action.
Organic	Organic	Accumulation/decay of vegetative matter.
Undifferentiated	Undifferentiated	Layered sequence; three materials or more.
Not Mapped	Not Mapped	

Modified from Table 2.5 of Howes and Kenk (1997) for this study.

Surface Expression

Code	Short Name	Description
Fan	Fan	Sediment deposited in a fan or cone shape <25% slope.
Cone	Cone	Sediment deposited in cone shape >25% slope.
Plane	Plane or Terrace	Sediment with a unidirectional surface <5% slope.
Scarp	Scarp	Erosional slopes cut into thick sediments typically >50% slope.
Undulating	Undulating	Sediment deposited in multidirectional hillocks and hollows with <25% slopes.
Hummocky	Hummocky or Ridged	Sediment deposited in multidirectional hillocks and hollows with >25% slopes.
Not Mapped	Not Mapped	

Modified from British Columbia (2010), based on Howes and Kenk (1997).

Aquifer Type

Code	Short Name	Description
Type 1a	Type 1a	Predominantly unconfined fluvial or glacio-fluvial sand and gravel aquifers found along major rivers of higher stream order with the potential to be hydraulically influenced by the river.
Type 1	Type 1b and 1c	Predominantly unconfined fluvial or glacio-fluvial sand and gravel aquifers found along rivers of moderate stream order (1b) with the potential to be hydraulically influenced by the river (1b), or along lower order (< 3-4) streams in confined valleys with relatively undeveloped floodplains, where aquifer thickness and lateral extent are more limited (1c).
Type 2	Type 2	Predominantly unconfined deltaic sand and gravel aquifers are commonly found in deltas where a stream or smaller river flows into a standing body of water.
Type 3	Type 3	Alluvial or colluvial fan sand and gravel aquifers typically occur at or near the base of mountain slopes, either along the side of valley bottoms, or if formed during the last period of glaciation, raised above the valley bottoms.
Type 4	Type 4	Sand and gravel aquifers generally formed near or at the end of the last period of glaciation. This category includes both confined and unconfined aquifers. Landforms and materials associated with glacio-fluvial outwash, ice contact deposits and glacio-marine environments near the coast.
Not Applicable	Not Applicable	
Not Mapped	Not Mapped	Examples include ice or lakes

Adapted from Province of British Columbia (2016).

Thickness Class

Code	Short Name	Description
None	None: 0-10cm	None/ very thin veneer.
Thin	Thin: 10-100cm	Veneer.
Medium	Medium: 100-300cm	Blanket.
Thick	Thick: 3-30m	Thick blankets and shallow constructional landforms.
Very Thick	Very Thick: >30m	Constructional landforms such as terraces, planes, fans, hummocky glacial deposits and scarps in thick surficial materials.
Not Applicable	Not Applicable	
Not Mapped	Not Mapped	

Developed for this study. Adapted from Howes and Kenk (1997).

Material Texture

Code	Short Name	Description
Very Fine Texture	Very Fine Texture	Silts and clays with low coarse fragment content.
Fine Texture	Fine Texture	Silt and clay loams with low coarse fragment content.
Medium Texture	Medium Texture	Silt and clay loams with high coarse fragment content up to 70%.
Coarse Texture	Coarse Texture	Sands and gravels with <70% coarse fragments.
Very Coarse Texture	Very Coarse Texture	>70% coarse fragments.
Not Applicable	Not Applicable	
Not Mapped	Not Mapped	

Developed for this study. Existing texture classes for soils and rooting particle size are for field surveys and not easily mapped by image interpretation.

Sorting Class

Code	Short Name	Description
Well Sorted	Well Sorted	Particles of the same size. Commonly wind or water deposited (e.g., Eolian fine sand or Glaciolacustrine silt).
Moderately sorted	Moderately sorted	Particle size varies moderately. Typically deposited by higher energy water or from a consistent bedrock source (e.g., glaciofluvial sands and gravels, or till derived from granitic bedrock).
Poorly sorted	Poorly sorted	Particles size varies widely (e.g., diamicton glacial till ranging from clay to boulders).
Not Applicable	Not Applicable	
Not Mapped	Not Mapped	

Developed for this study.