Mineral Resource Prediction Using Advanced Data Analytics and Machine Learning of the QUEST-South Stream Sediment Geochemical Data, Southwestern British Columbia

Eric Grunsky[1] & Dennis Arne[2]

[1]Department of Earth and Environmental Sciences, University of Waterloo, egrunsky@gmail.com [2]Telemark Geosciences, Canada/Australia, dennis.arne@telemarkgeosciences.com

Introduction

Geoscience BC conducted an infill stream sediment sampling program in the QUEST-South project area in 2009 and reanalysed archived regional geochemical survey (RGS) samples using ICP-MS in 2010 (Figure 1). Catchments were determined for these samples in 2011 and a preliminary interpretation of the geochemical data undertaken using the dominant rock type in the catchments to level the data for the effects of variable background. In this new Geoscience BC project we apply multivariate statistical methods, including the random forests classification method, to interpret the data from 8545 samples. Data for 35 elements were levelled for laboratory analytical effects and values below the lower limit of detection imputed prior to a centred log ratio transformation to moderate the effects of geochemical closure. Multivariate methods were applied to the clr-transformed data for the purposes of discovering patterns and features that potentially describe geochemical, geological, geophysical and the effects of gravitational processes (Grunsky et al., 2010). These methods included principal-component analysis (PCA) and t-distributed stochastic neighbour embedding (t-SNE) (van Maaten and Hinton, 2008). Each sample was also attributed with the closest MINFILE occurrence, excluding anomalies and showings, within 2.5 km of the sample site (Figure 3). MINFILE occurrences were grouped based on similarities in BCGS mineral deposit models and geochemical signatures for training purposes (Figure 4). A training data set of 474 samples, including 100 samples not attributed with a MINFILE occurrence and the most significant principal components (Figure 5), was used to generate random forests prediction model from which posterior probabilities were estimated for the remaining 8071 samples.

Conclusions

Two approaches using the random forests procedure have been trialled: principal component analysis (PCA) and t-SNE with 9 factors. The posterior probabilities for various grouped mineral deposit models have been used to generate kriged images to test for geospatial coherence in the predictions. These have been compared to kriged images for porphyry Cu-Au-Mo deposits generated using raw data and data corrected for the effects of catchment bedrock type (Figures 6 - 9). The t-SNE9 posterior probabilities provide a better visual fit to the distribution of known mineral occurrences and have a slightly higher level of accuracy compared to the posterior probabilities obtained using PCA (Figures 10 & 11). Catchment polygons have been thematically coded using the t-SNE9 posterior probabilities to provide maps of exploration potential for 13 of the grouped mineral deposit types for which measurable accuracies in prediction were obtained. The use of random forests provides predictions of mineral occurrences that are better than those obtained by manual data analysis but does require some optimisation in its use.

References

Arne, D., Mackie, R., Pennimpede, C., Grunsky, E. and Bodnar, M. (2018): Integrated assessment of regional stream-sediment geochemistry for metallic deposits in northwestern British Columbia (parts of NTS 093, 094, 103, 104), Canada; Geoscience BC, Report 2018-14.

Breiman, L. (2001): Random Forests; Machine Learning, v. 45, p. 5–32.

Grunsky, E.C. (2010): The interpretation of geochemical survey data; Geochemistry, Exploration, Environment Analysis, v. 10, p. 27–74.

Grunsky, E.C. and Arne, D.C. (2020): Mineral-Resource Prediction Using Advanced Data Analytics and Machine Learning of the QUEST-South Stream-Sediment Geochemical Data, Southwestern British Columbia (Parts of NTS 082, 092) , Geoscience BC Report 2020-06.

Grunsky, E.C. and de Caritat, P. (2019): State-of-the-Art Analysis of Geochemical Data for Mineral Exploration, Geochemistry, Exploration Environment Analysis, Special Issue from Exploration 17, October, 2017, Toronto, Canada

van der Maaten, L.J.P. and Hinton, G.E. (2008): Visualizing data using t-SNE; Journal of Machine Learning Research, v. 9, p. 2579–2605.

Project Area

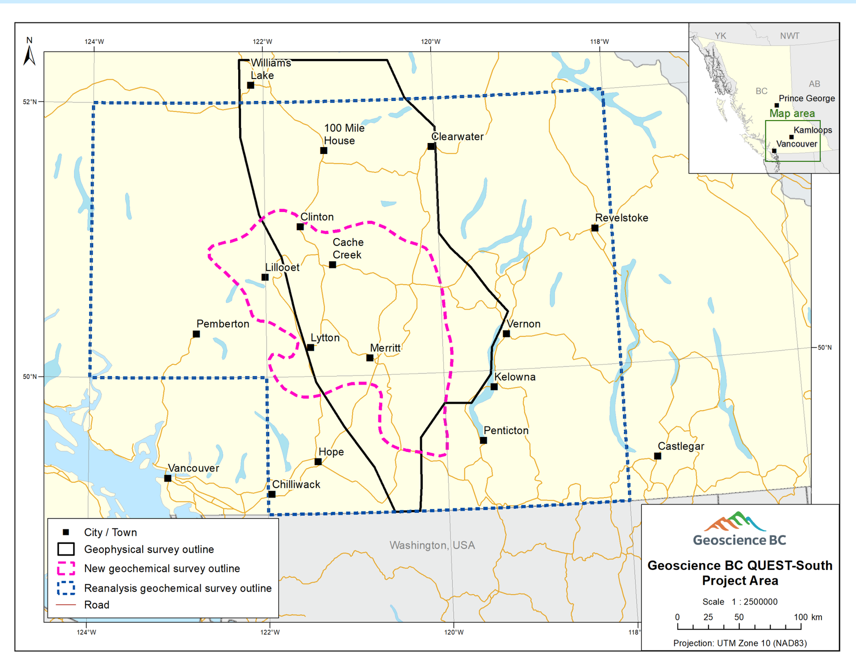

Figure 1: Location of the QUEST-South project area.

Distance to MINFILE Site

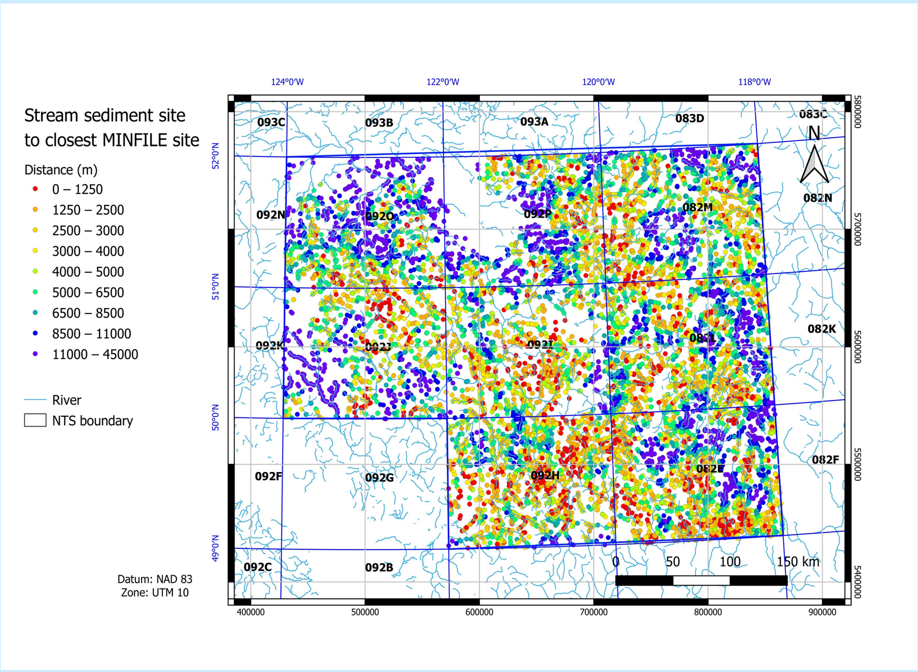

Figure 3: Geographic distribution of the distance measures between stream-sediment sites and the closest MINFILE site .

Thematic Catchment Maps

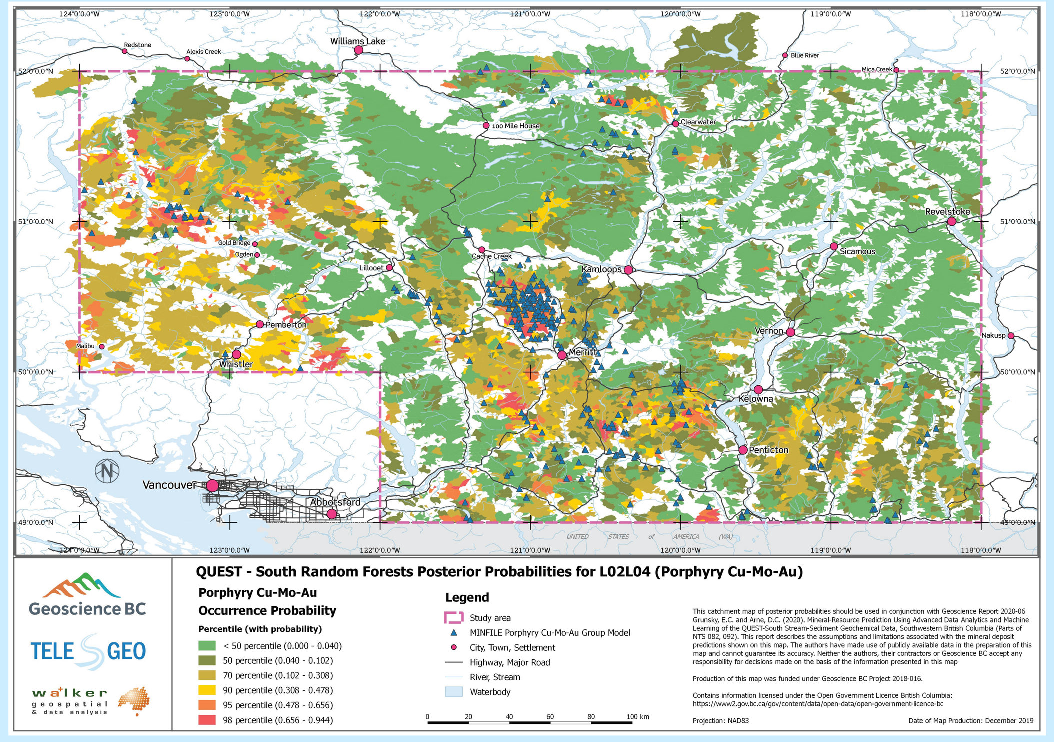

Figure 2: Example of a prospectivity map product for porphyry Cu-Au-Mo deposits in which catchments have been thematically codes based on random forests t-SNE9 posterior probabilities.

Legend

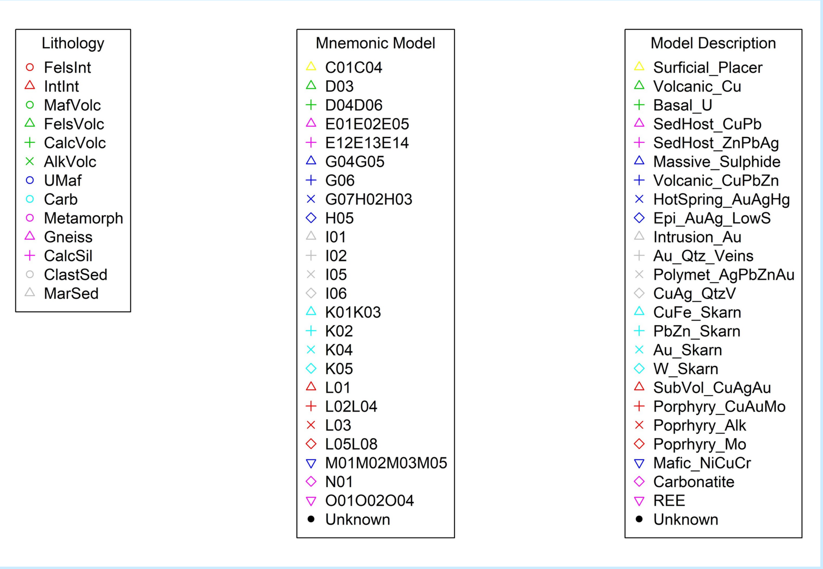

Figure 4: Legends showing colours and symbols for lithology (left) mineral deposit types as BCGS Mineral Deposit Model mnemonics (centre) and short descriptions of the respective BCGS Mineral Deposit Models (right).

Principal Components

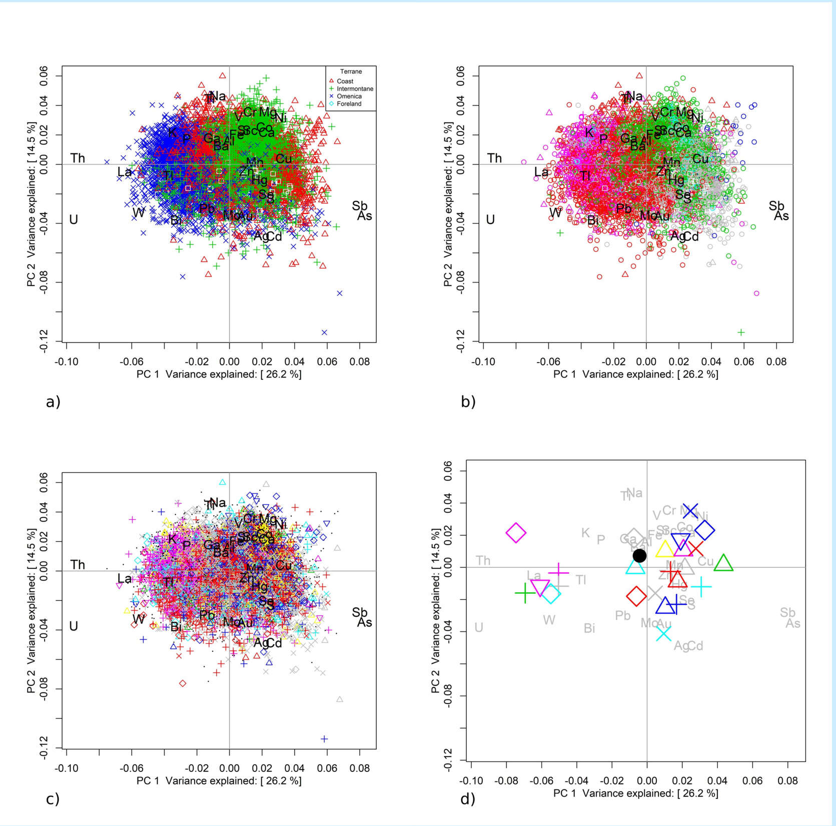

Figure 5: a) Biplot of PC1-PC2 showing the relative relationships between the principal terranes in the Quest-South region. The Omineca terrane shows relative enrichment in Th-U-La while the Coast terrane shows relative enrichment of Sb-As indicating provenance with chalcophile-rich rocks and relative enrichment in Th-U-La. The Intermontane terrane shows a mixture of lithophile, siderophile and chalcophile elements; **b)** Biplot of PC1-PC2 showing the relative relationships between the generalized lithologies of the region. Felsic intrusive rocks show relative enrichment in Th-U-La and sedimentary rocks show relative enrichment in Sb-As. Volcanic rocks show a mixture of lithophile and siderophile elements; **c)** Biplot of PC1-PC2 showing the relative relationships between the GroupModels as defined from the proximity of stream sediment sites and MINFILE sites. See Figure 8d for an interpretation of the patterns; **d)** Biplot of PC1-PC2 showing the mean values of the PC1-PC2 scores for each of the GroupModel classes. Relative relationships between the GroupModels are defined from the proximity of stream-sediment sites and MINFILE sites. Epithermal Au-Ag deposits show relative enrichment with siderophile elements. Porphyry deposits show relative enrichment with chalcophile elements. Carbonatite, REE, basal U, W skarn and sediment-hosted deposits show relative enrichment with U-Th-La-W-Tl. See Figure 4 for the legend of colours and symbols. Note that the scaling of the mean values has been changed to enhance the separation. The relative positions of the GroupModel icons do not match the scales of the biplot axes.

Random Forests Compared with Conventional Approaches

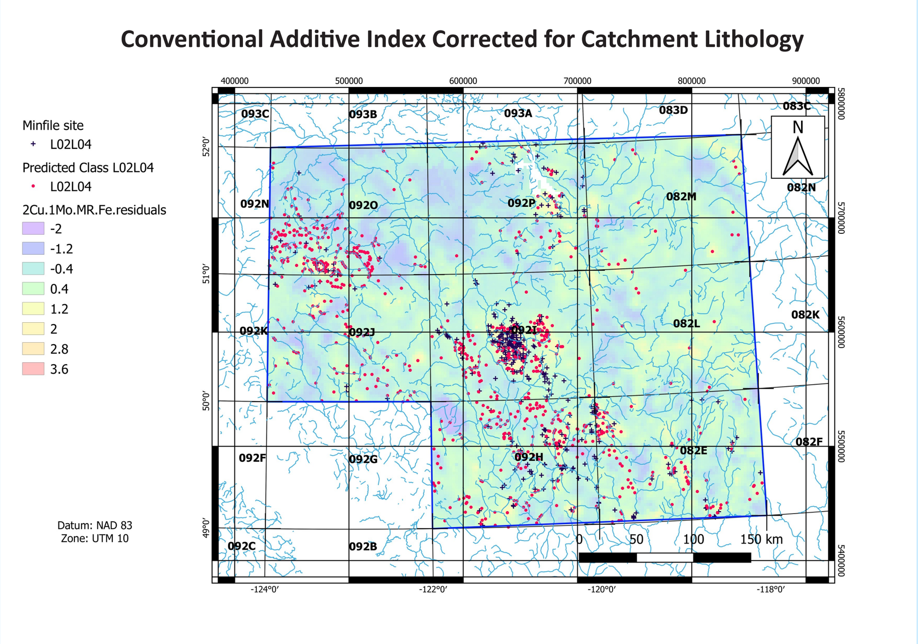

Figure 6: An additive model of 2 Log_{10} Cu multiple regression residuals regressed against Log_{10} Fe plus 1 Log_{10} Mo multiple regression residuals regressed against Log_{10} Fe shown with MINFILE porphyry Cu-Au-Mo occurrences (in black symbols) and random forests class predictions for porphyry Cu-Au-Mo (in red symbols). Areas of increased potential for L02L04 deposits are shown by colour shading of the kriged image.

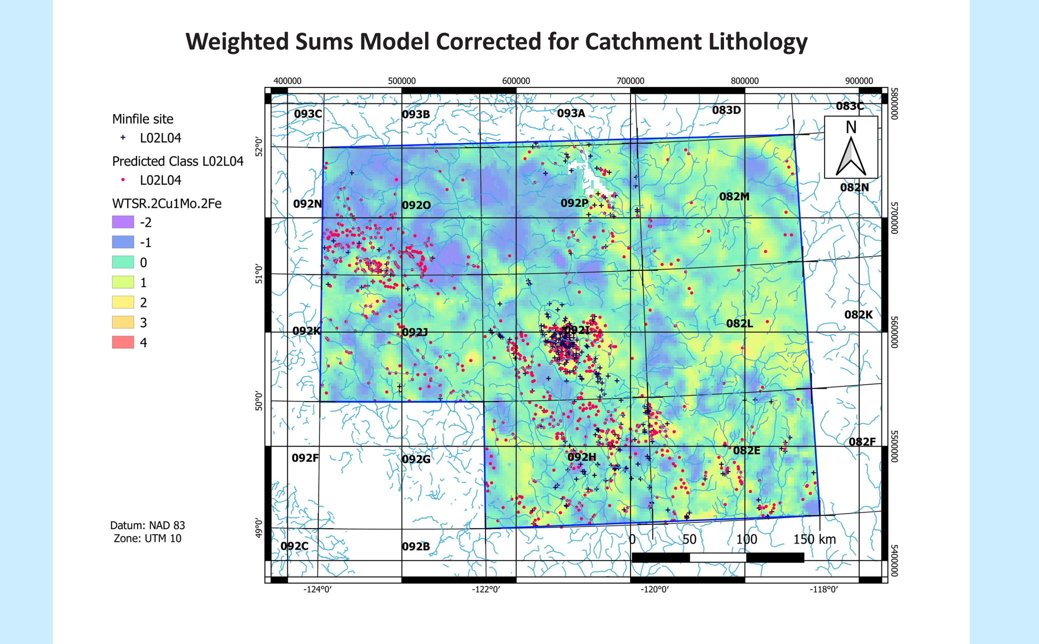

Figure 7: Weighted sums model consisting of Log_{10} Cu multiple regression residuals (2), Log_{10} Mo multiple regression residuals (1) and Log_{10} Fe (-2) shown with MINFILE porphyry Cu-Au-Mo occurrences (in black symbols) and random forests class predictions for porphyry Cu-Au-Mo (in red symbols) . Areas of increased potential for L02L04 deposits are shown by colour shading of the kriged image.

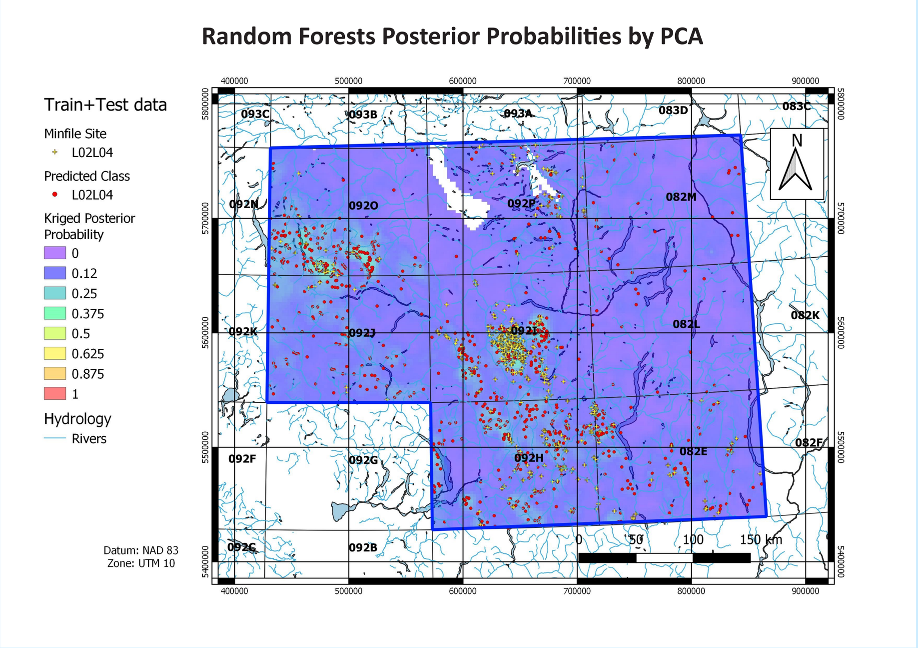

Figure 8: Geographic distribution of individual sites for GroupModel L02L04 overlain on a kriged image of the posterior probabilities for porphyry Cu-Au-Mo (L02L04) prediction using random forests and the PCA metric, based on the train + test data and a distance threshold of 2500 m. MINFILE sites tagged as L02L04 are shown as yellow crosses. Stream-sediment sites identified as class L02L04 by random forests are shown as red dots. Areas of increased potential for L02L04 deposits are shown by colour shading of the kriged image.

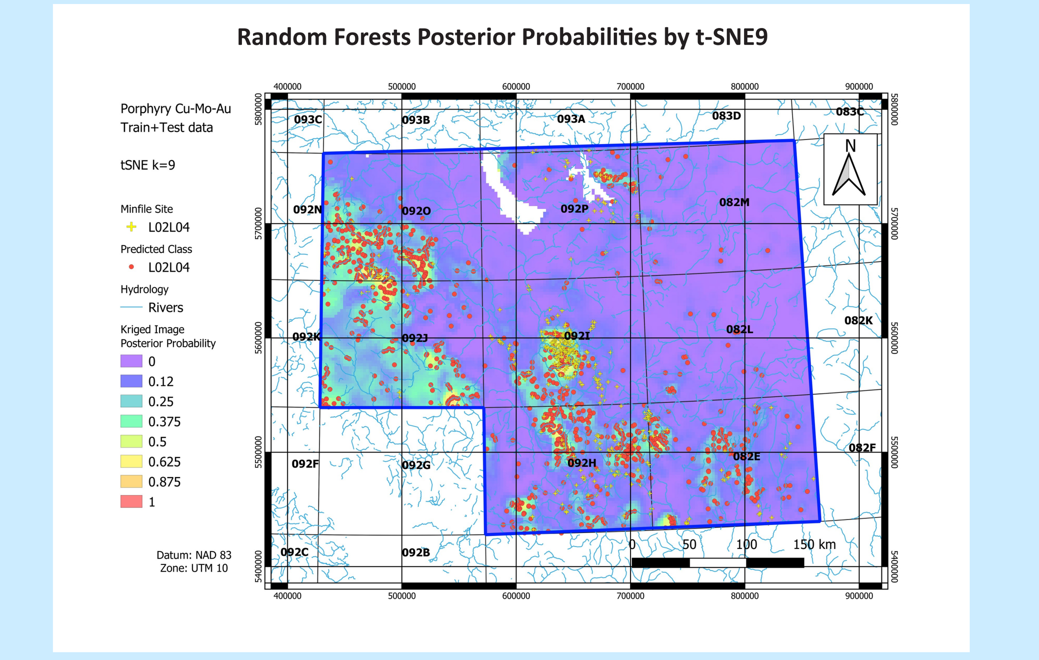

Figure 9: Geographic distribution of individual sites for GroupModel L02L04 overlain on a kriged image of the posterior probabilities for porphyry Cu-Mo-Au (L02L04) prediction using random forests and the t-SNE9 metric, based on the train + test data and a distance threshold of 2500 m. MINFILE sites tagged as L02L04 are shown in yellow crosses. Stream-sediment sites identified as class L02L04 by random forests are shown in red dots. Areas of increased potential for L02L04 deposits are shown by map colour shading of the kriged image

PCA vs tSNE9 Metrics

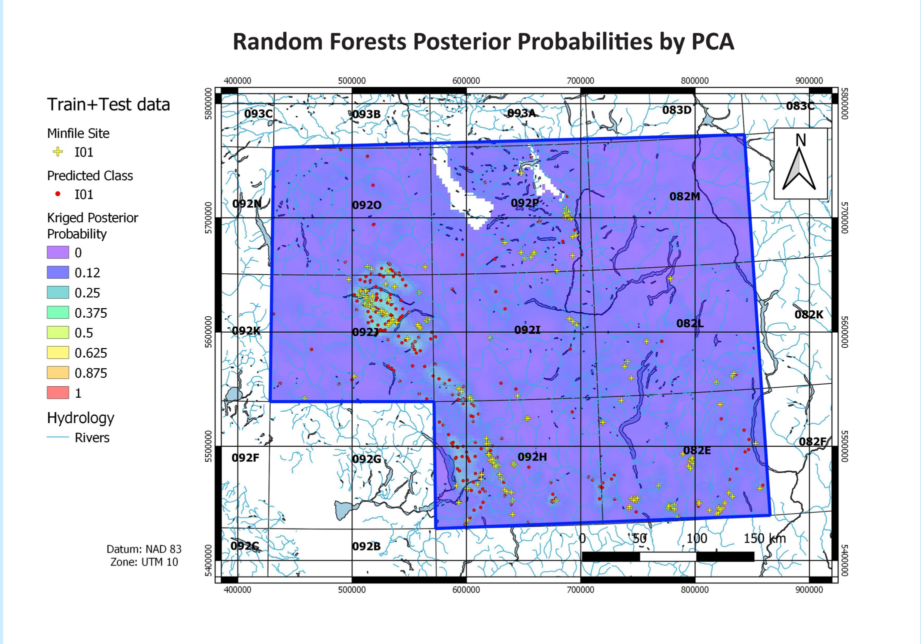

Figure 10: Geographical distribution of individual sites predicted for GroupModel I01 overlain on a kriged image of posterior probabilities for Au quartz veins (I01) prediction using random forests and the PCA metric, based on the train + test data and a distance threshold of 2500m. MINFILE sites tagged as I01 are shown in yellow crosses. Stream-sediment sites identified as class I01 by random forests are shown in red dots.

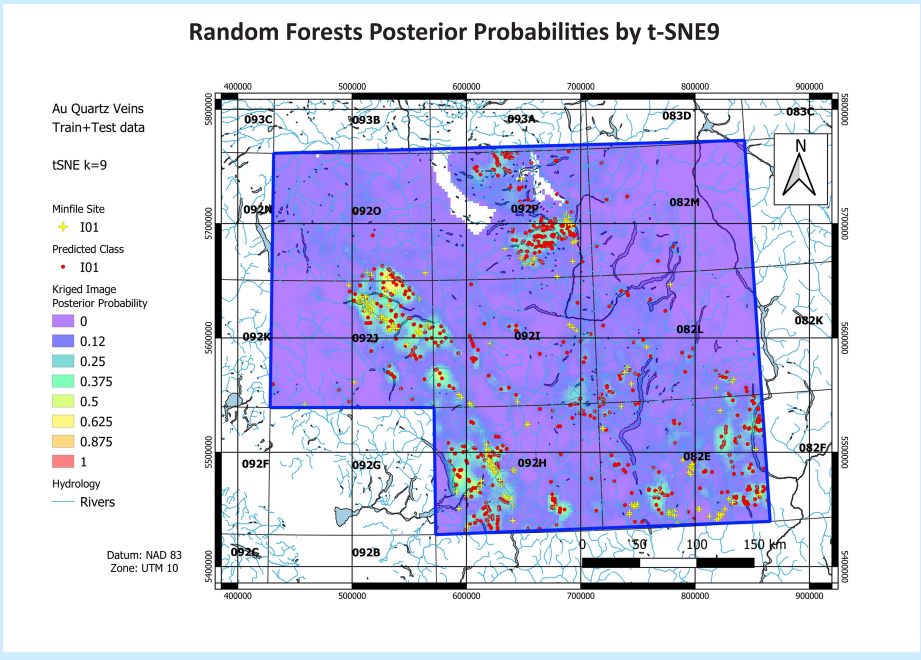

Figure 11: Geographic distribution of individual sites for GroupModel I01 overlain on a kriged image of the posterior probabilities for Au quartz veins (I01) prediction using random forests and the t-SNE9 metric, based on the train + test data and a distance threshold of 2500 m. MINFILE sites tagged as I01 are shown in yellow crosses. Stream-sediment sites identified as class I01 by random forests are shown in red dots. Areas of increased potential for I01 deposits are shown by colour shading on the kriged image.