Alberta Background Soil Quality System

Leveraging historical data and harnessing advances in data science

Project Status Update

Natalie Shelby-James (InnoTech Alberta) & Paul Fuellbrandt (Statvis)

CLRA – May 2022

statvis Trust Through Data

InnoTech ALBERTA
A SUBSIDIARY OF ALBERTA INNOVATES

Alberta Background Soil Quality System

Outline

- Problem
- Project Summary
- Benefits
- Data Acquisition
- Example Salinity Data Exploration
- Next Steps
- Participation Opportunities & Questions

Alberta Background Soil Quality System

The Problem

- Well known that numerous areas of Alberta have naturally elevated chemical parameter concentrations.
- No public scientifically vetted resource that accurately predicts background soil salinity and metals chemistry within the Province.
- High cost, schedule, and regulatory barriers to prove that elevated parameters are of natural origin.
- Efficiency challenges with the number of applications requiring detailed regulatory reviews.

Alberta Background Soil Quality System

Project Summary

The **Alberta Background Soil Quality System** is a **PTAC and InnoTech –** sponsored project to develop an interactive map of background soil salinity and metals parameters for use in environmental management.

Key activities

- Compile, clean and integrate existing soil salinity and metals data
- Analysis to fingerprint background and remove impacted samples
- Develop a predictive background soil mapping system through a phased approach
- Create an interactive web application to deploy the system

statvis Trust Through Data

InnoTech ALBERTA
A SUBSIDIARY OF ALBERTA INNOVATES

Alberta Background Soil Quality System - Benefits

Industry and Practitioners

- Reduce background data collection
- Empirical evidence of natural variability
- More accurate liability estimates
- Focus resources on managing risk to receptors
- Move stalled sites to closure

Government & Regulators

- Decreased review times
- More reclamation certificates issued
- Increased consistency in data presentation
- Less inactive wells

Albertans

- Open tool available to extrapolate data for multiple uses
- Less disturbance and disruption to the natural environment
- More sites being cleaned up & less orphan wells

statvis Trust Through Data

InnoTech ALBERTA
A SUBSIDIARY OF ALBERTA INNOVATES

Alberta Background Soil Quality System - Benefits

AB Background Soil Quality System

- Background fingerprinting based on pattern recognition **not** concentration
- Modern data science, dimensionality reduction now widely available
- Impacted samples included originally to learn differences between impacted and background

RESULT: Data-driven process removes potential for invisible bias

Other Background Databases

- Pre-determined concentration limits for background
- Simple statistics were the only tools widely available until recently
- Impacted samples removed based on assumptions, location, experience, etc.

RESULT: Potential for missing naturally elevated parameters

statvis Trust Through Data

InnoTech ALBERTA
A SUBSIDIARY OF ALBERTA INNOVATES

Alberta Background Soil Quality System

Data Acquisition

- Data Tiers
  1. Georeferenced datasets (point data) collected >2015
     - Standardized data collection & analytical methods
     - Less compilation efforts, comprehensive datasets (higher quality), additional metadata
  2. Non-georeferenced datasets (e.g., UWI info – 110 x 110 m)
     - Still of value, but lower resolution / data accuracy. Will be used to supplement primary dataset.

Alberta Background Soil Quality System

Data Acquisition

- Seven provided databases
- 50,000 salinity samples
- 14,000 metals samples
- > 1000 unique locations throughout Alberta

Example Salinity Exploration—Two Datasets

AER Audit Data and a Producer Dataset

- Total of **40,087** salinity samples
- Four distinct geographic areas activity
  - 39% in the region 1 (n=14,842)
  - 36.5% in the region 2 (n=13,840)
  - 6.5% in the region 3 (n=2,443)
  - 18% in the remaining areas (n=8,962)

Example Salinity Exploration—Two Datasets

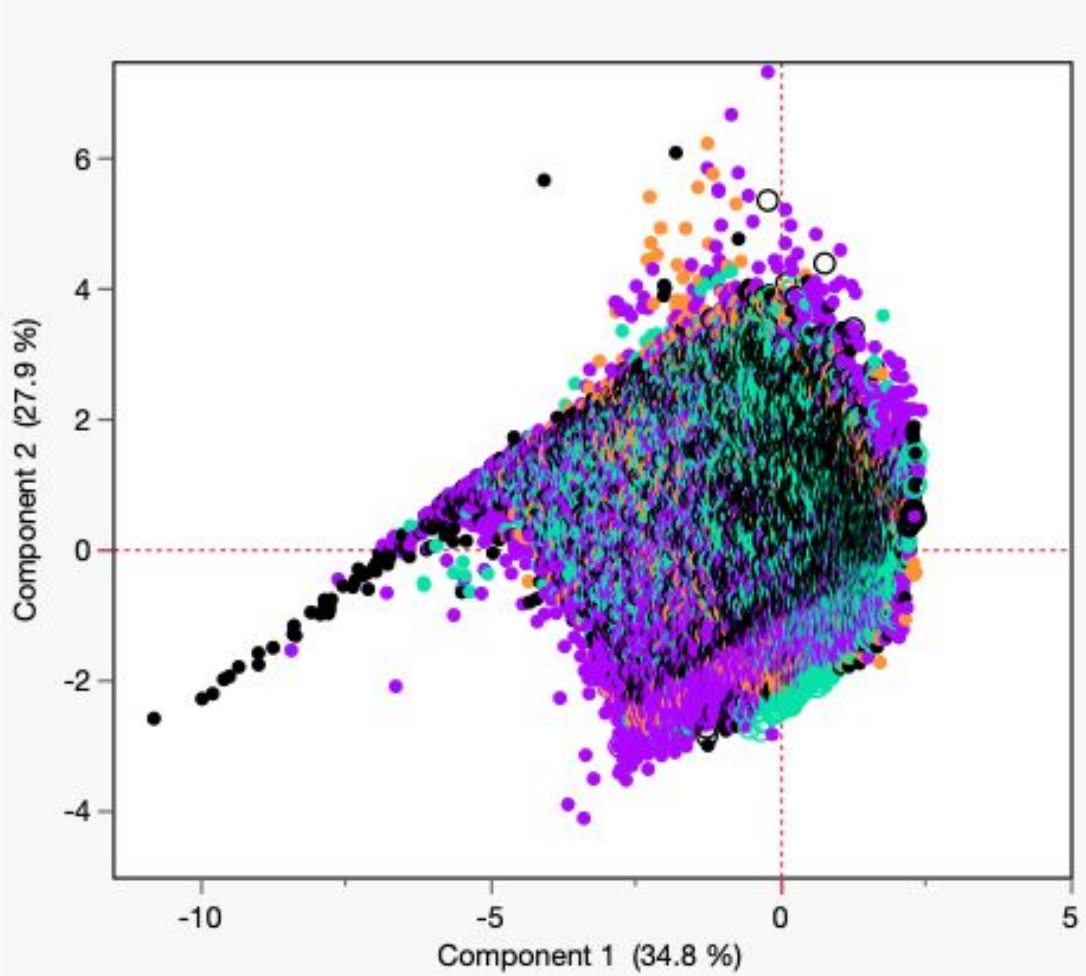
Data Exploration Process

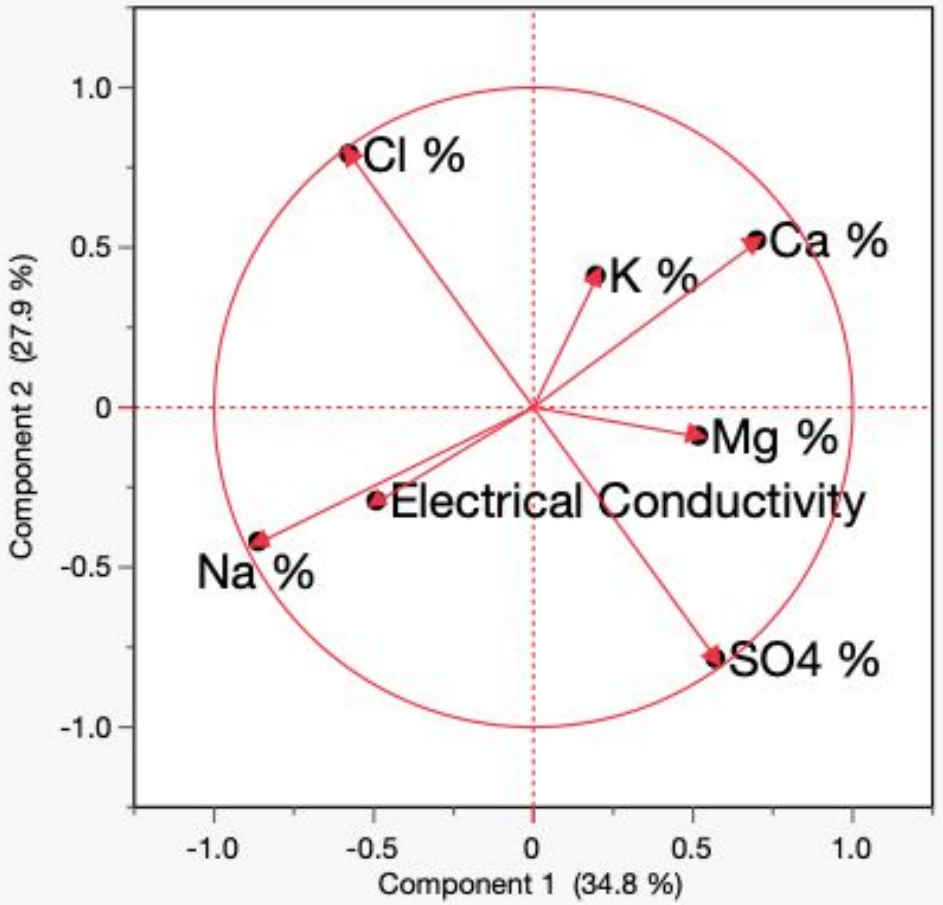
1. **Compilation–**creation of a single comprehensive database.
2. **Imputation–**Remove non-detect values using multiplicative lognormal replacement method
3. **Normalize–**Log normalize data to perform outlier tests and
4. **Dimensionality Reduction–**percent normalize data for pattern assessment (HCA/PCA/Radar plots/UMAP)
5. **Correlations–**Use meq/kg and mg/kg data for correlation plots

All techniques used together to create a multi-phased approach

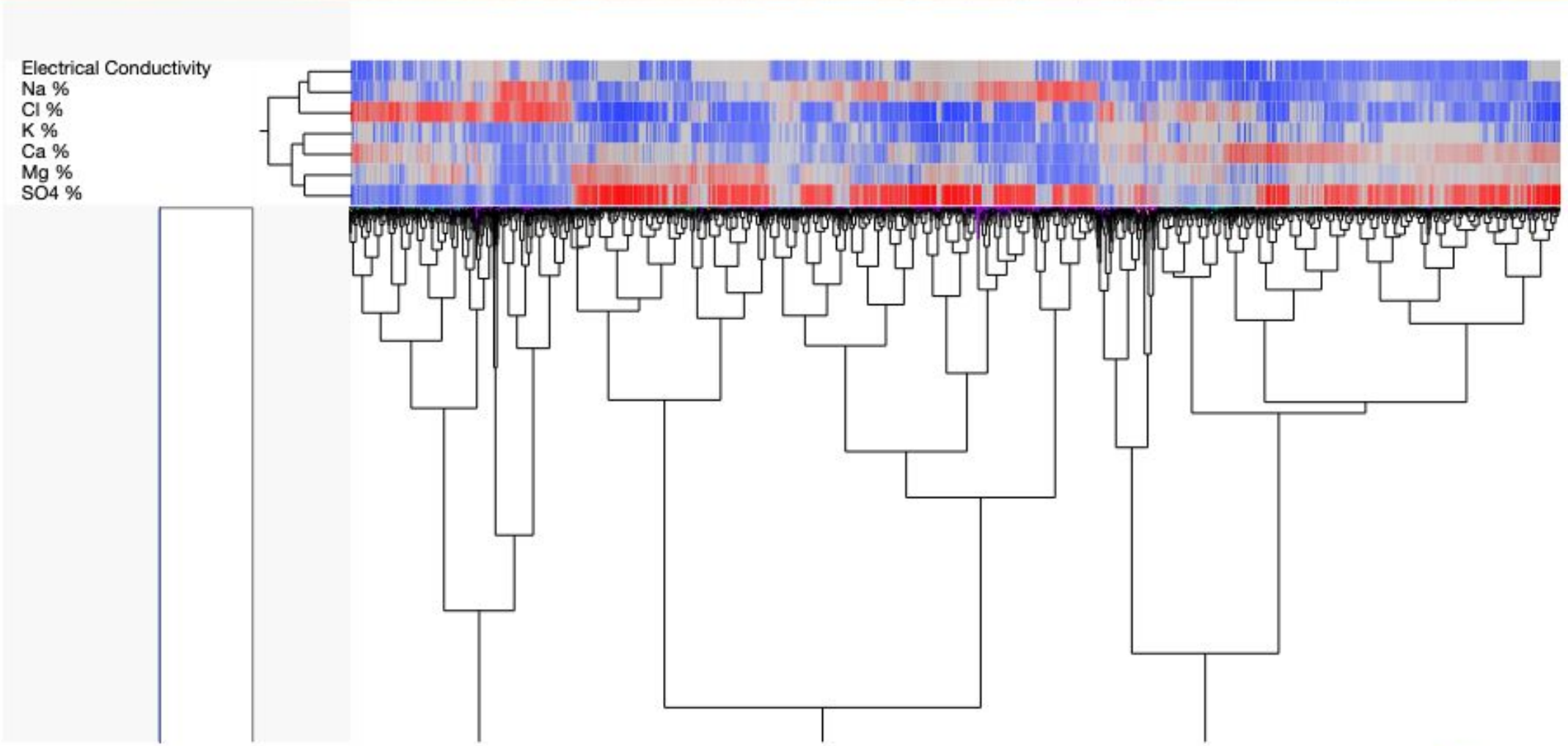
Data is **subset in various ways** (geospatially and by concentration) to identify trends in smaller subgroups

Example Salinity Exploration—Dimensionality Reduction





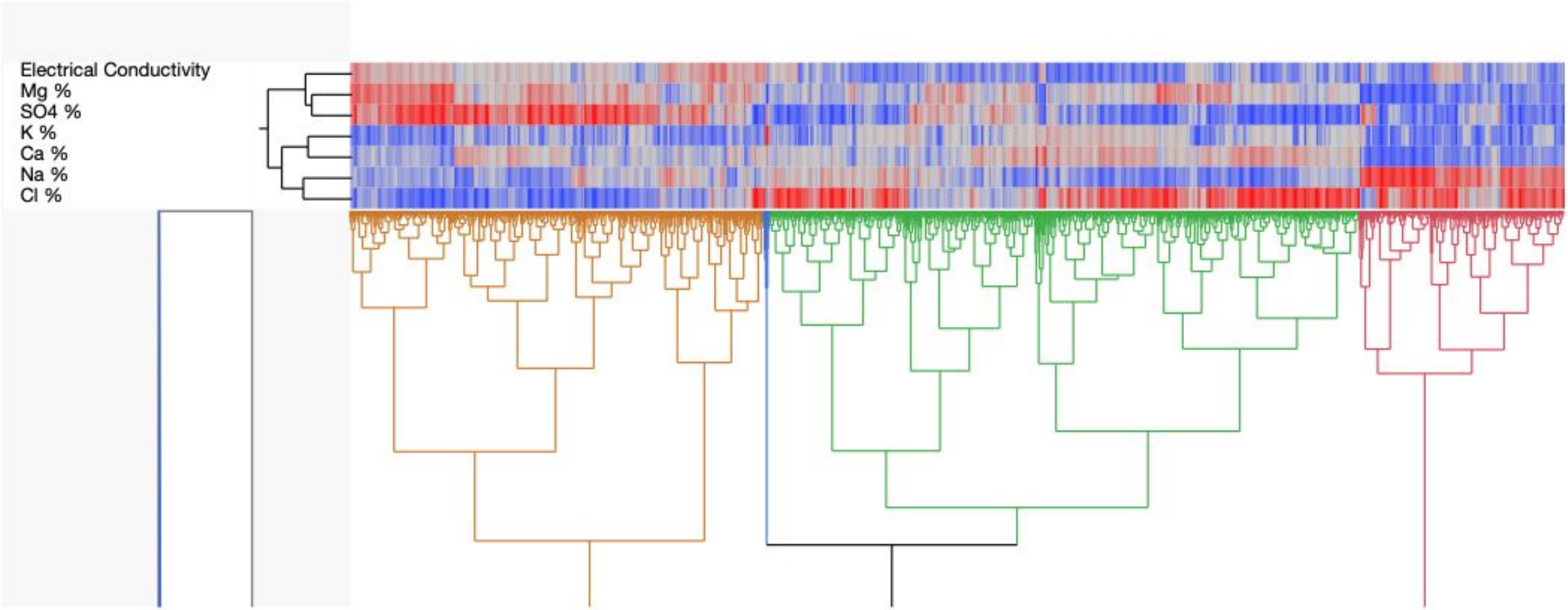
Example Salinity Exploration—Dimensionality Reduction



statvis Trust Through Data

InnoTech ALBERTA
A SUBSIDIARY OF ALBERTA INNOVATES

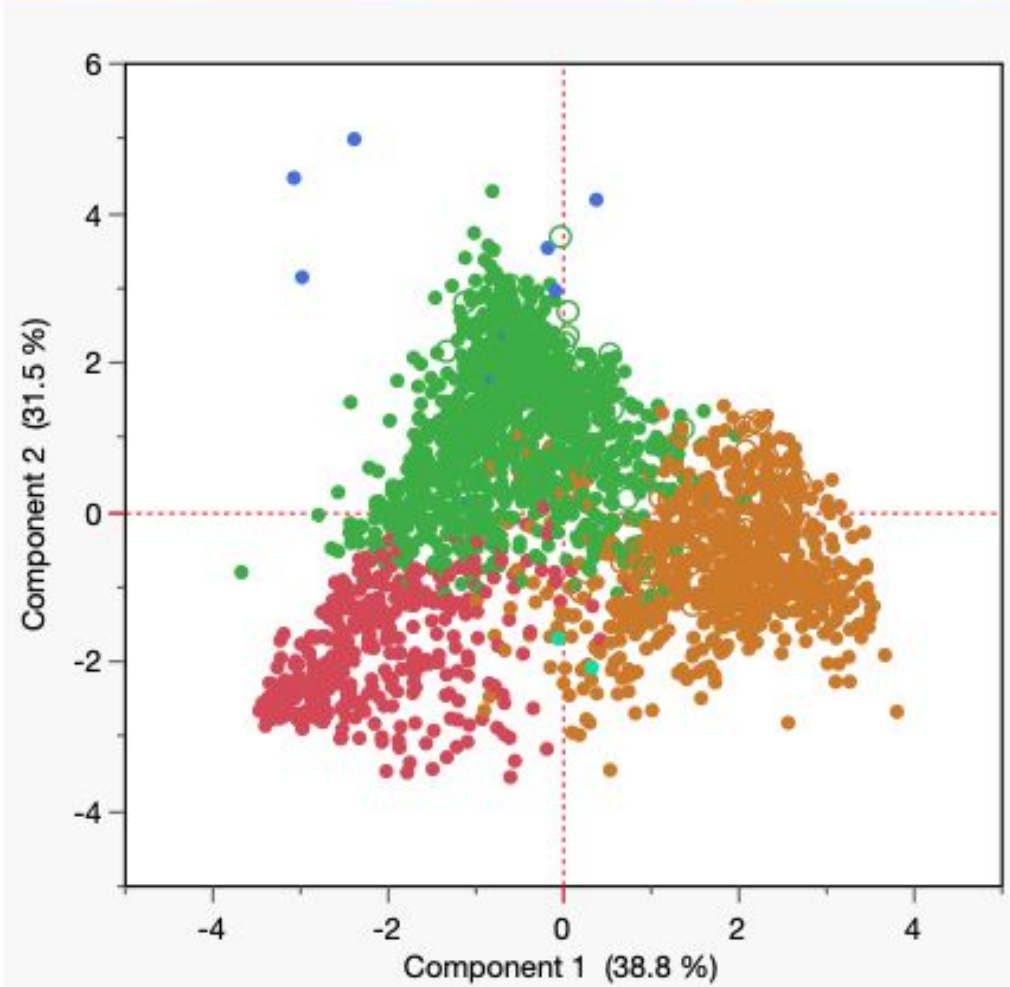
Example Salinity Exploration—Single Region

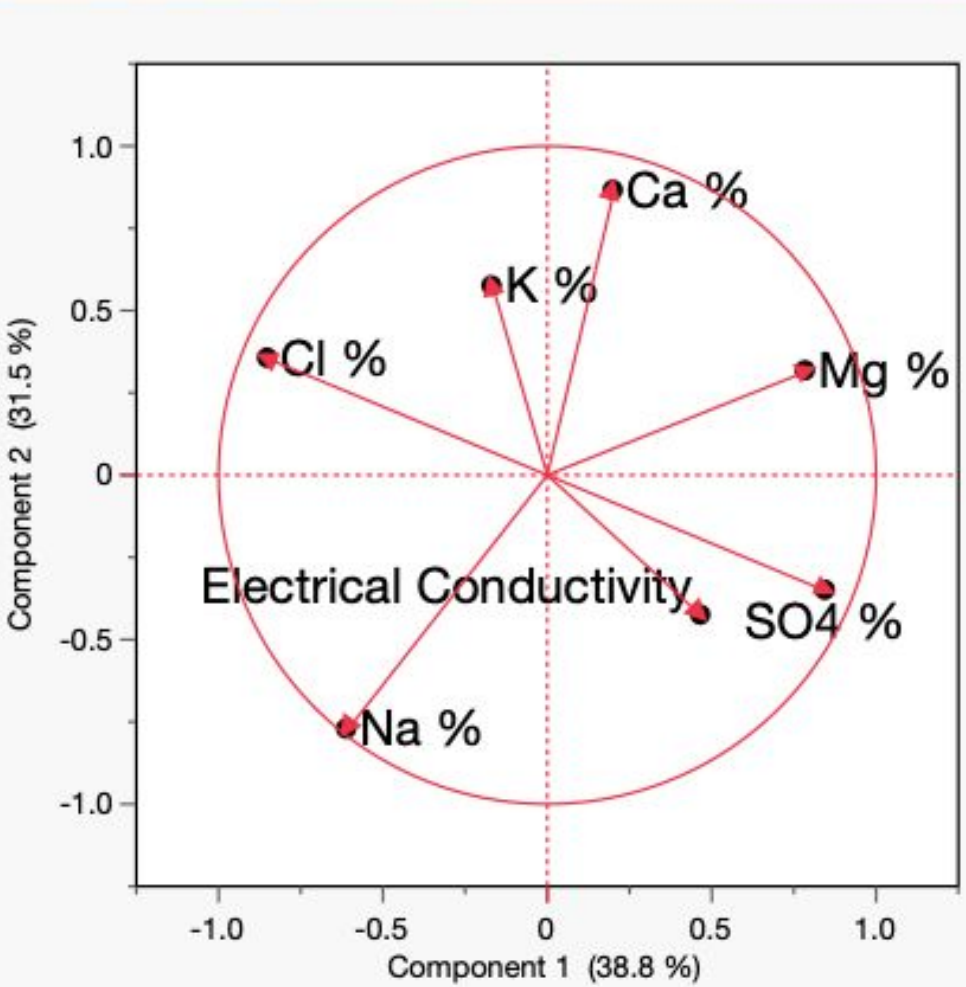


statvis Trust Through Data

InnoTech ALBERTA
A SUBSIDIARY OF ALBERTA INNOVATES

Example Salinity Exploration—Single Region

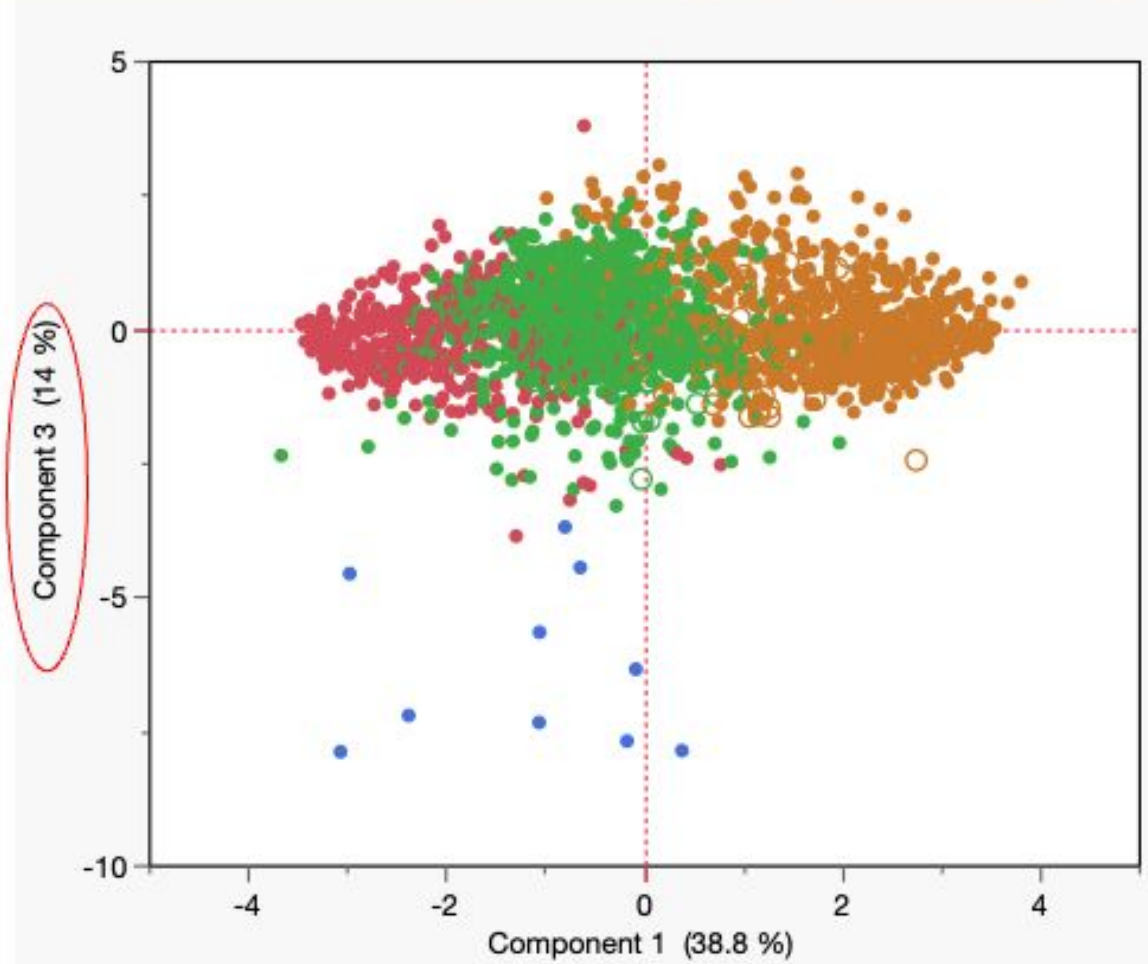


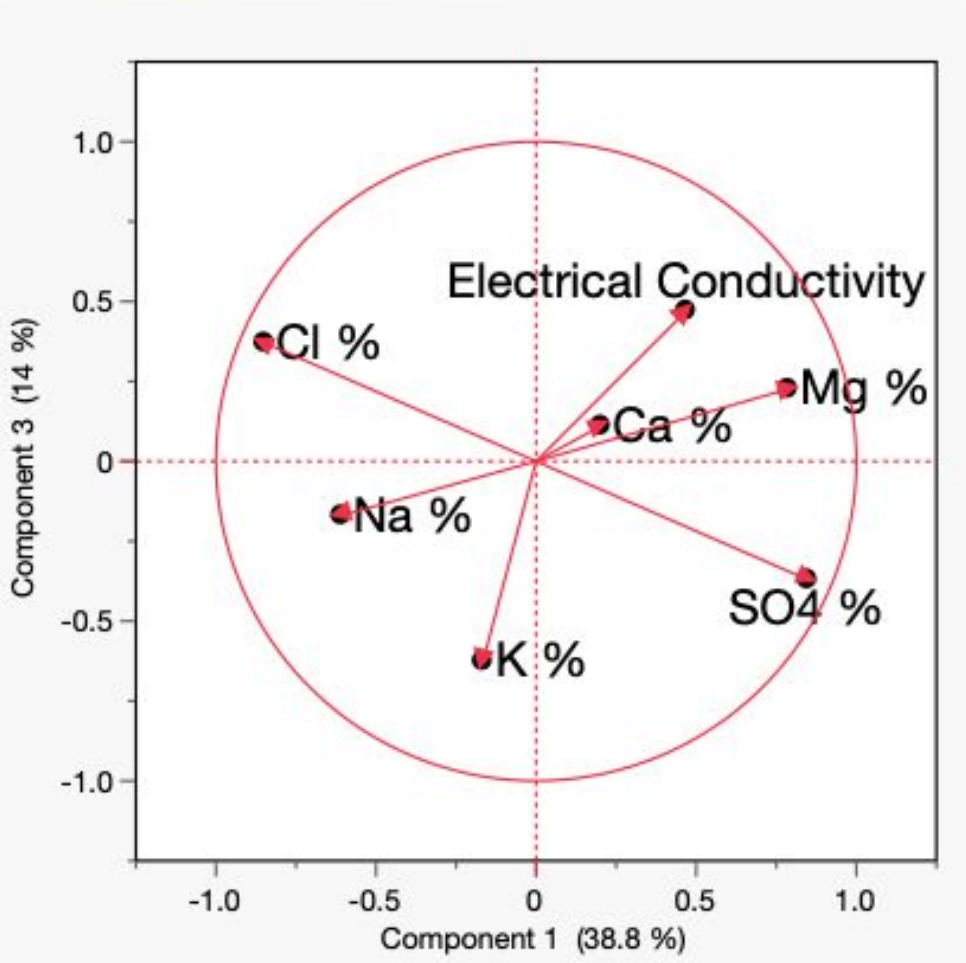


statvis Trust Through Data

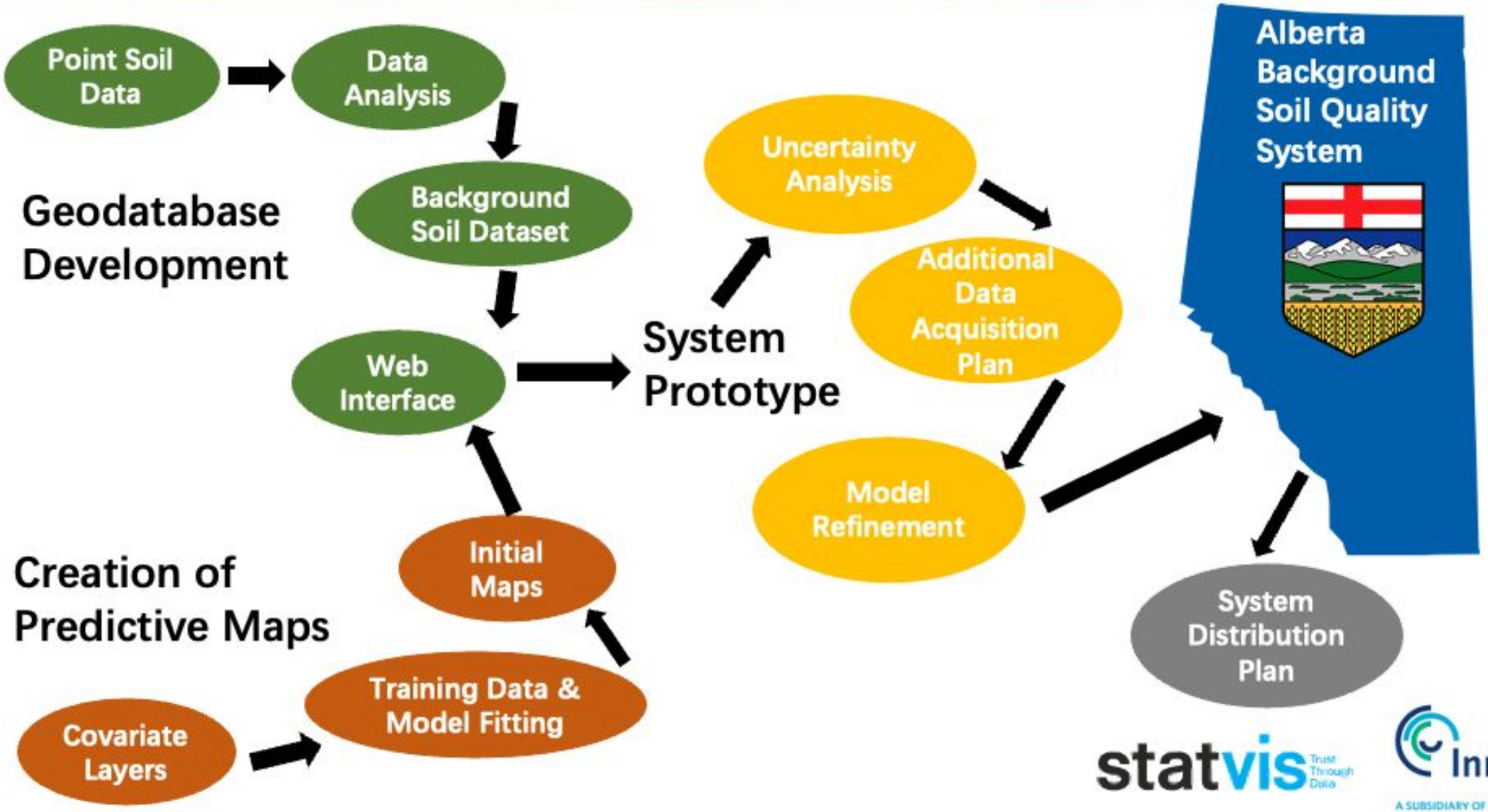
InnoTech ALBERTA
A SUBSIDIARY OF ALBERTA INNOVATES

Example Salinity Exploration—Single Region





Alberta Background Soil Quality System – Next Steps



statvis Trust Through Data

InnoTech ALBERTA
A SUBSIDIARY OF ALBERTA INNOVATES

Participation Opportunities

Data you want included?

Ideas about covariate data?

Other thoughts or musings?

Natalie Shelby-James (InnoTech Alberta) – e: : Natalie.Shelby-James@innotechalberta.ca
Paul Fuellbrandt (Statvis) – e: Paul@statvis.com
CLRA – May 2022

InnoTech ALBERTA
A SUBSIDIARY OF ALBERTA INNOVATES