

### 3. Environmental Covariates for Digital Soil Mapping

Canadian Digital Soil Mapping Workshop, 2022

# About Us

**Brandon Heung**

Dalhousie University

[Brandon.Heung@dal.ca](mailto:Brandon.Heung@dal.ca)

**Daniel Saurette**

Ontario Ministry of Agri-Food and Rural Affairs

University of Guelph

[Daniel.Saurette@ontario.ca](mailto:Daniel.Saurette@ontario.ca)



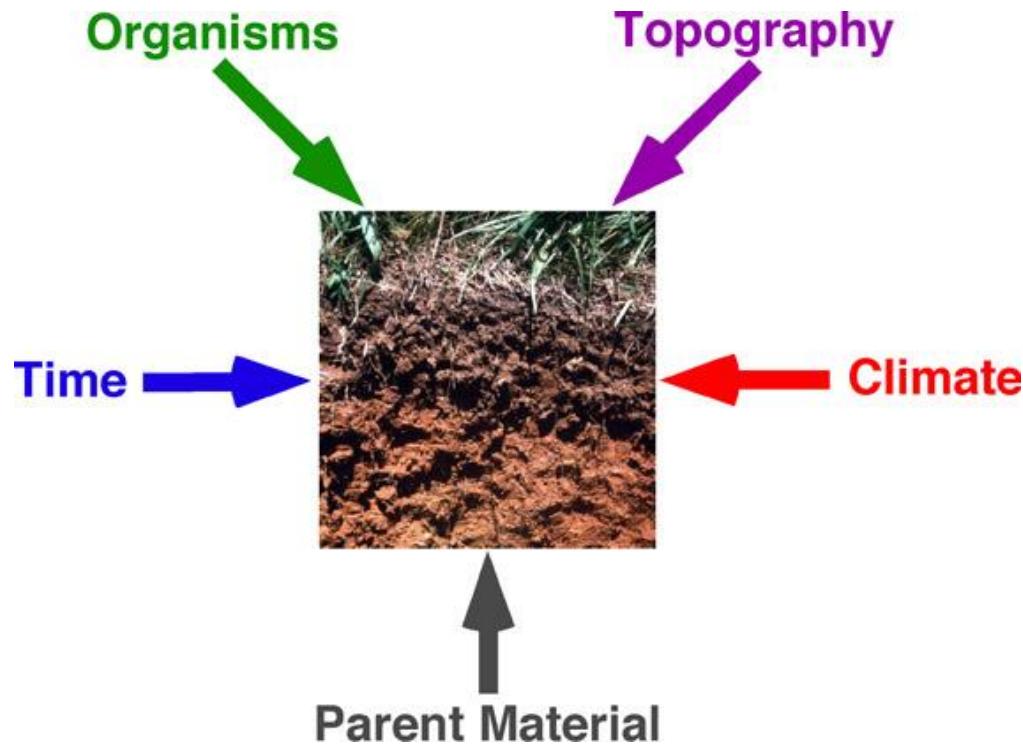
# Outline

---

1. Overview of Soil-Environmental Variables
2. Dealing with multi-collinearity



# Environmental Covariates



c = climate (temp, rainfall)

o = organisms (vegetation/land cover)

r = relief (topography, aspect)

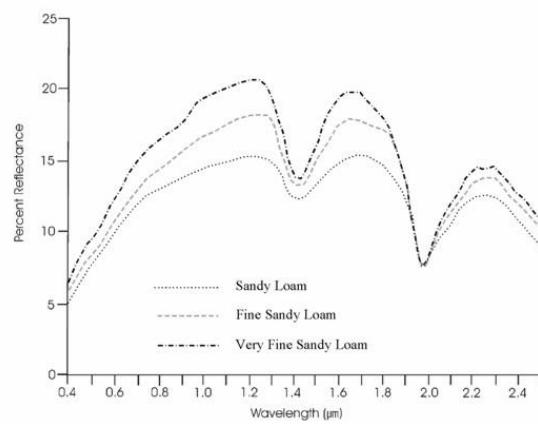
p = parent material (bedrock and surficial geology)

# Examples of Data Sources

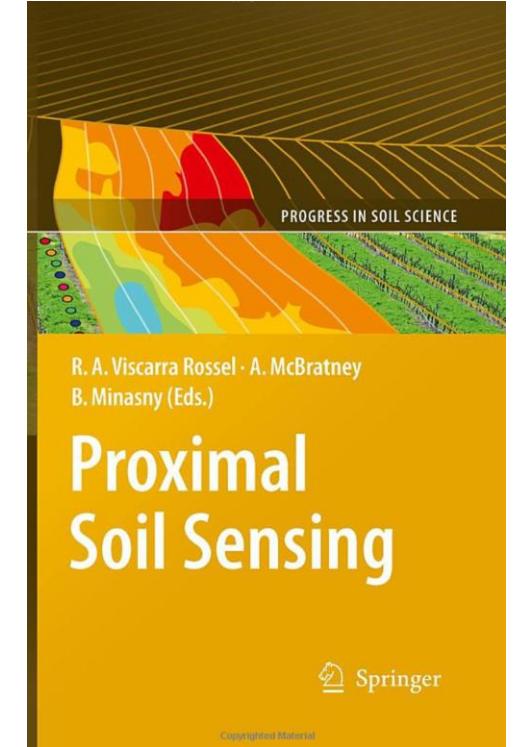
Proximal Soil Sensing	Remote Sensing	Other Geospatial Data
Electrical Conductivity Surveys	Satellite Imagery	AAFC Crop Inventory
Gamma Radiometric Surveys	Gamma Radiometric Surveys	Surficial Geology
Soil Resistivity	Digital Elevation Models	Bedrock Geology
Ground Penetrating Radar Surveys	LiDAR Surveys	Climate Model Data
Hyperspectral Surveys	Hyperspectral Surveys	Soil Surveys

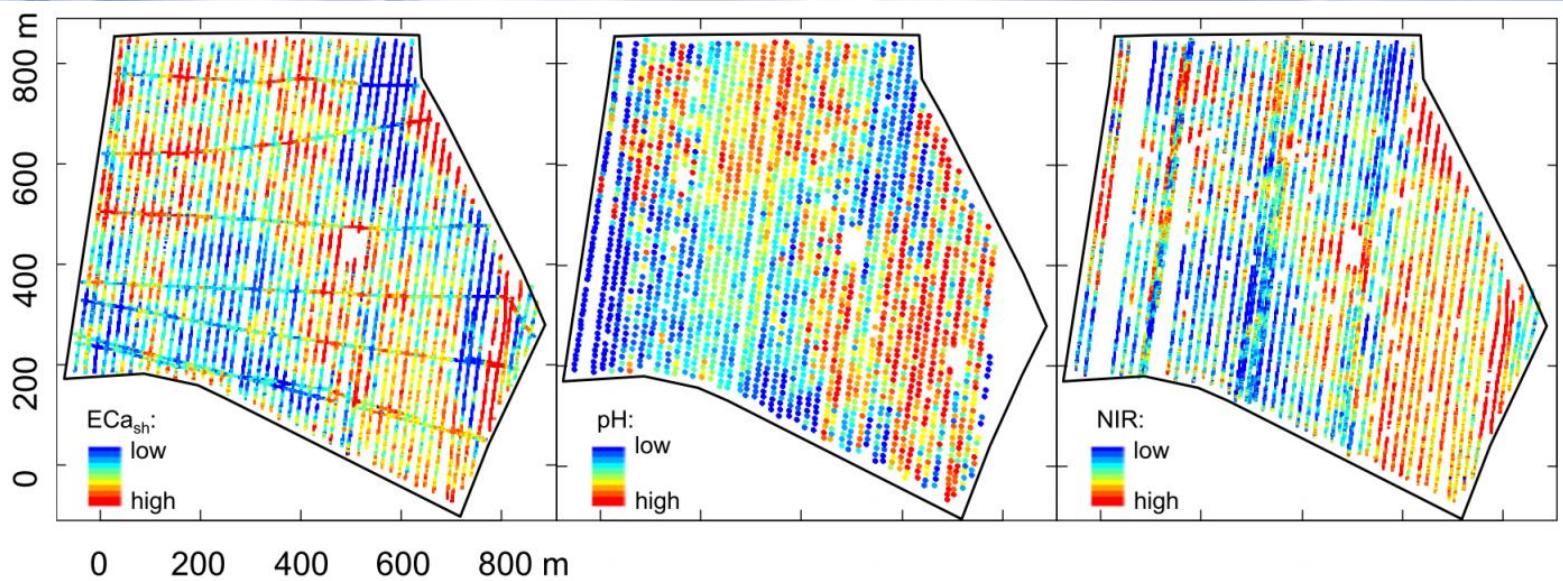


# Soil (*s*)



- Remote Sensing Data
- Proximal Soil Sensing Data

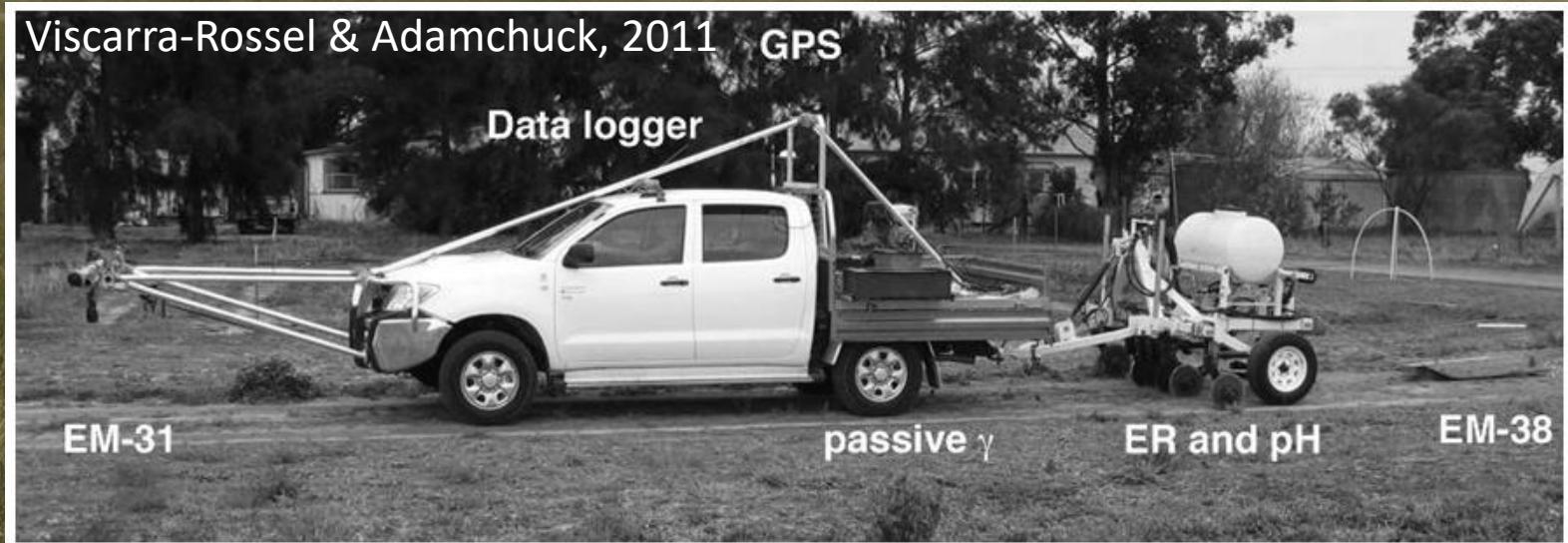




**Fig 3. Proximal sensor measurements.** Spatial locations of the sensor readings by the three-sensor platform.

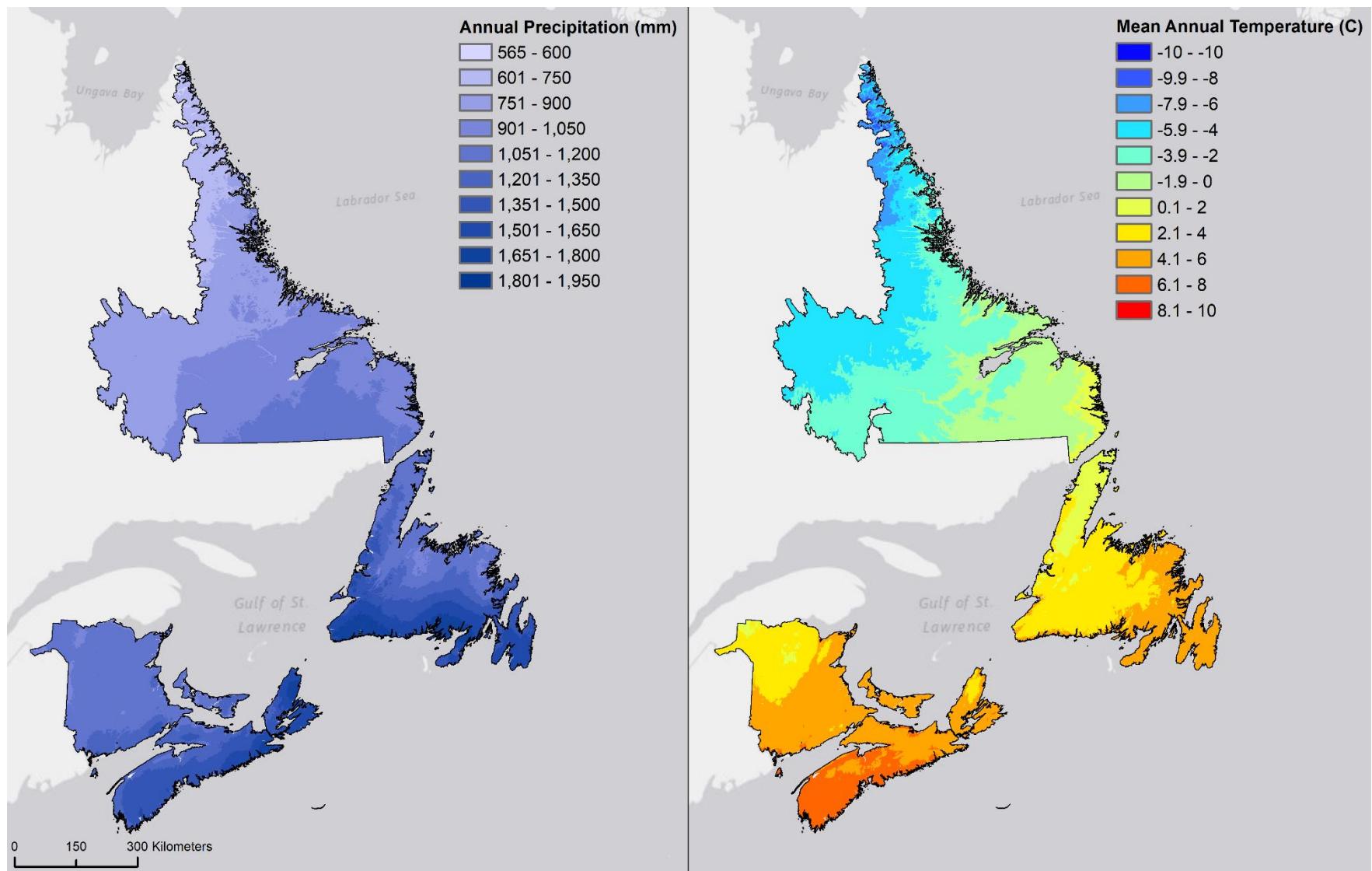
Apparent electrical conductivity (EC<sub>a</sub><sub>sh</sub>), pH, and light absorbance (NIR<sub>avg</sub>) in the topsoil are depicted from left to right.  
 Quantile classification scheme with 10 levels used.

Schirrmann et al., 2016



Multi-Platform Soil Sensor

# Climate (c)

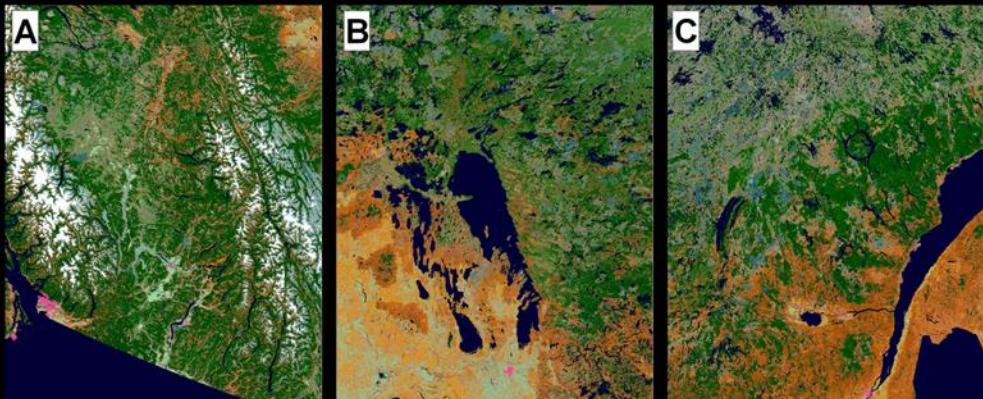
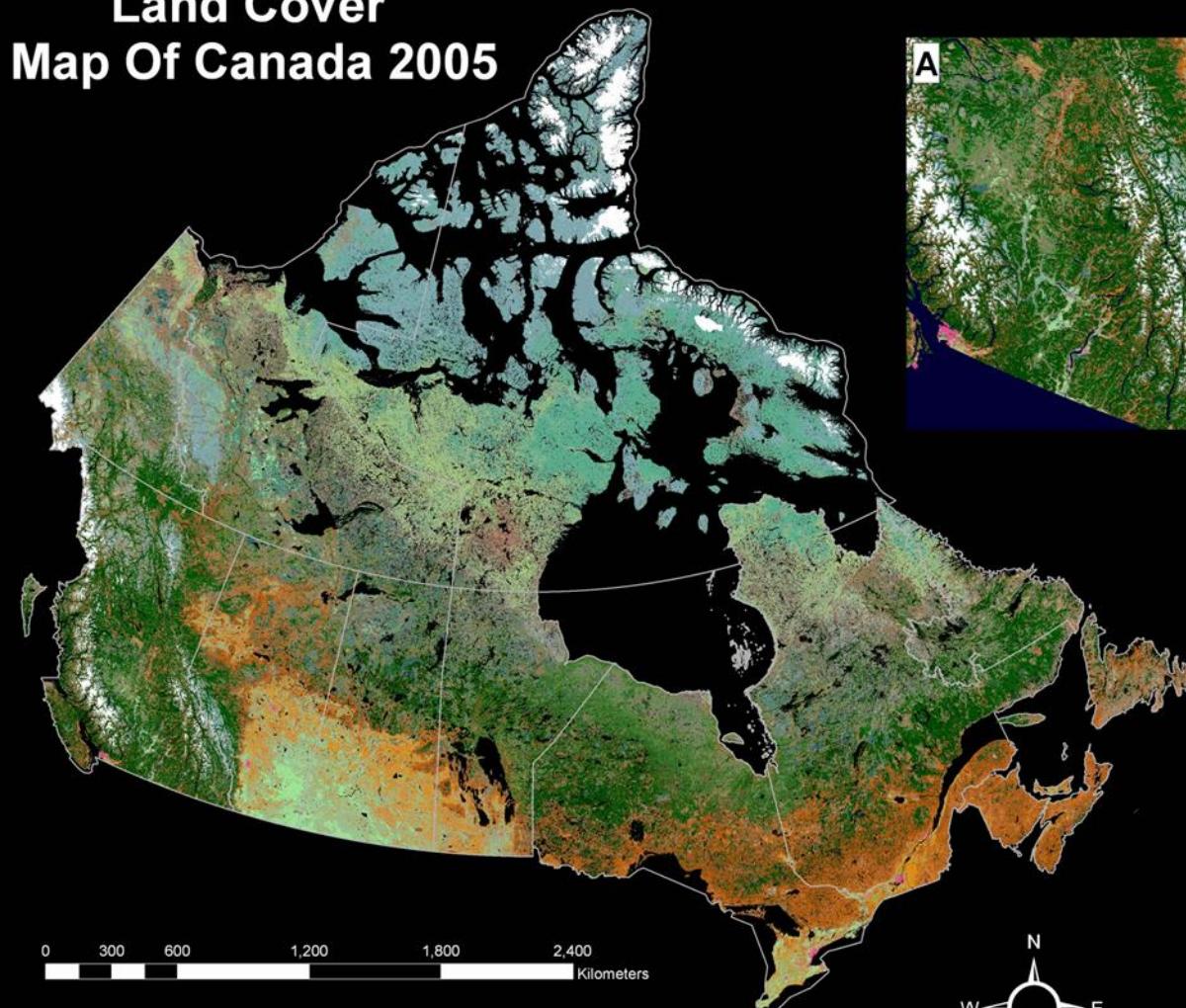


# Organisms (o)

Mainly derived from remote sensing data and are focused on vegetation



# Land Cover Map Of Canada 2005

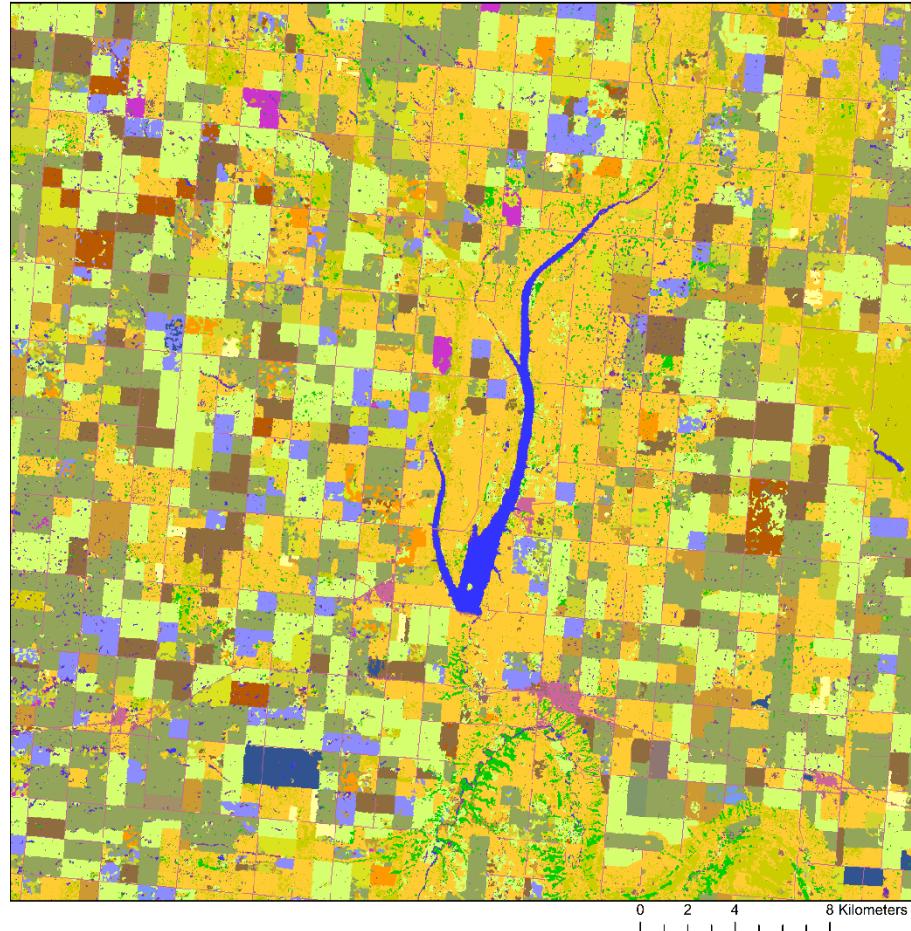


## Legend

- 1.Evergreen needleleaf forest / Close Canopy
- 2.Deciduous broadleaf forest /Close Canopy
- 3.Mixed evergreen-deciduous forest / Mature to Old Closed Canopy
- 4.Mixed evergreen-deciduous forest / Young Closed Canopy
- 5.Mixed evergreen-deciduous forest / Closed Canopy
- 6.Evergreen needleleaf forest / Medium crown density / Moss-Shrub Understory
- 7.Evergreen needleleaf forest / Medium crown density / Lichen-Shrub Understory
- 8.Evergreen needleleaf forest / Low crown density / Shrub-Moss Understory
- 9.Evergreen needleleaf forest / Low crown density / Lichen(Rock) Understory
- 10.Evergreen needleleaf forest / Low crown density / Poorly Drained
- 11.Deciduous broadleaf forest / Low to Medium Density
- 12.Deciduous broadleaf forest / Young Regenerating
- 13.Mixed evergreen-deciduous forest / Mixed coniferous / Low to Medium Density
- 14.Mixed evergreen-deciduous forest / Mixed deciduous / Low to Medium Density
- 15.Mixed evergreen-deciduous forest / Mixed deciduous / Low Regenerating Young Mixed Cover
- 16.Shrubland / High-Low Shrub Dominated
- 17.Herbaceous vegetation / Temperate or subpolar grassland / Grassland, Prairie Region
- 18.Herbaceous vegetation / Temperate or subpolar grassland /Herb-Shrub-Bare Cover
- 19.Herbaceous vegetation / Saturated temperate or subpolar grassland /Wetland
- 20.Herbaceous vegetation / Temperate or subpolar grassland with a sparse tree layer / Coniferous sparse
- 21.Herbaceous vegetation / Short sod polar grassland / Herb-Shrub
- 22.Herbaceous vegetation / Polar grassland with sparse shrub layer / Shrub-Herb-Lichen-Bare
- 23.Herbaceous vegetation / Polar grassland with sparse shrub layer / Herb-Shrub poorly drained
- 24.Herbaceous vegetation / Polar grassland with sparse dwarf-shrub layer / Lichen-Shrub-Herb, Bare Soil
- 25.Herbaceous vegetation / Polar grassland with sparse dwarf-shrub layer / Low vegetation cover
- 26.Annual graminoid or forb vegetation / Cropland-Woodland
- 27.Annual graminoid or forb vegetation / Temperate or subpolar annual grassland or forb vegetation / High Biomass Cropland
- 28.Annual graminoid or forb vegetation / Temperate or subpolar annual grassland or forb vegetation / Medium Biomass Cropland
- 29.Annual graminoid or forb vegetation / Temperate or subpolar annual grassland or forb vegetation / Low Biomass Cropland
- 30.Nonvascular Dominated / Temperate or subpolar lichen vegetation / Lichen-Barren
- 31.Nonvascular Dominated / Temperate or subpolar lichen vegetation / Lichen-sedges, moss low shrub wetland
- 32.Nonvascular Dominated / Temperate or subpolar lichen vegetation / Lichen-spruce bog
- 33.Vegetation Not Dominated / Consolidated rock sparse vegetation / Rock Outcrops
- 34.Recent Burns
- 35.Old Burns
- 36.Urban and Built-Up
- 37.Water Bodies
- 38.Mixes of Water and Land
- 39.Snowice

Canadian Land Cover Classification Map (circa 2000)

# Annual Crop Inventory (AAFC)



Water	Pasture / Forages	Triticale	Sunflower	Herbs
Exposed Land / Barren	Too Wet to be Seeded	Winter Wheat	Soybeans	Canaryseed
Urban / Developed	Fallow	Spring Wheat	Peas	Hemp
Greenhouses	Barley	Corn	Chickpeas	Coniferous
Shrubland	Millet	Canola / Rapeseed	Beans	Broadleaf
Wetland	Oat	Flaxseed	Fababeans	Mixedwood
Grassland	Rye	Mustard	Lentils	



1 m spatial resolution

15 m spatial resolution

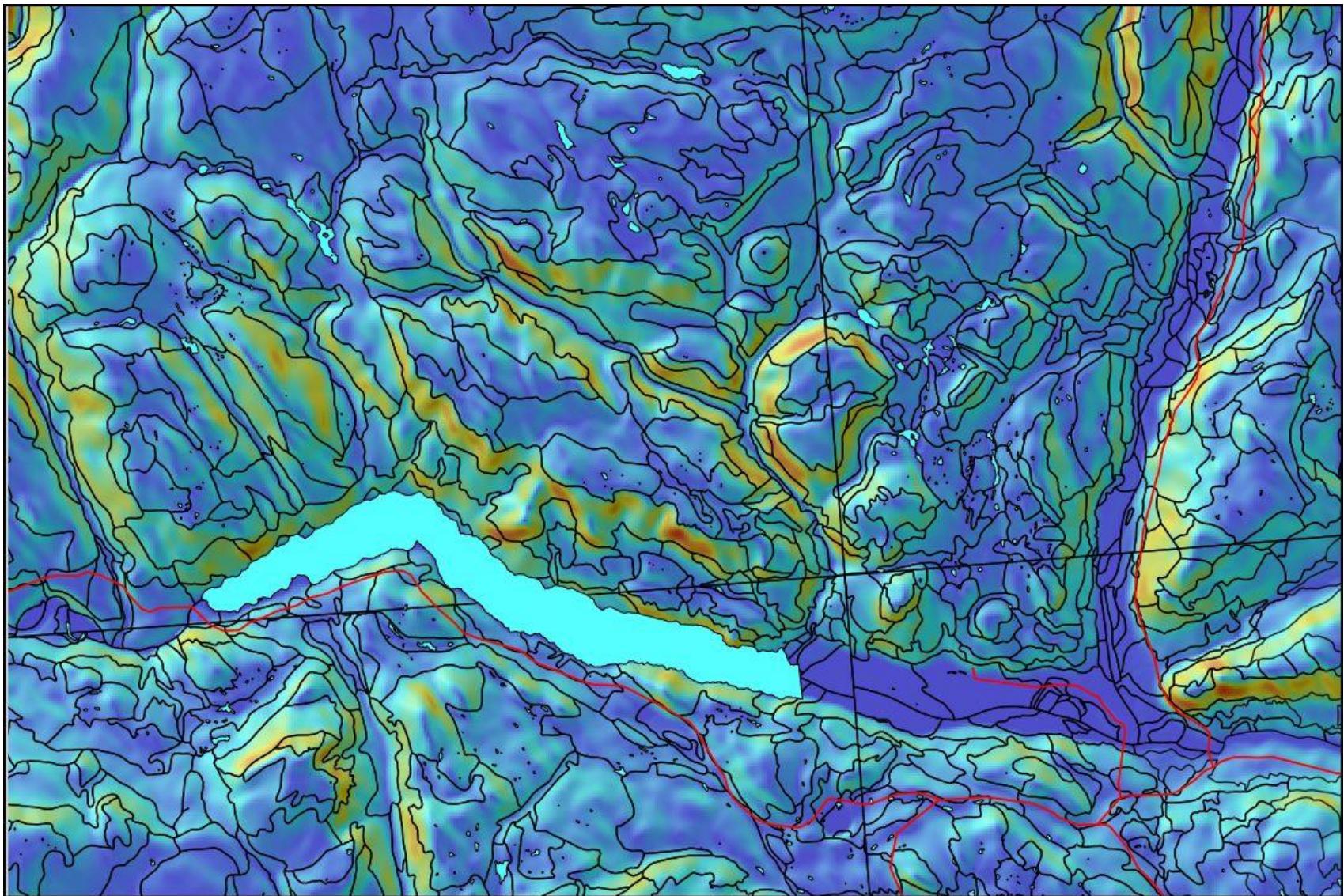
25 m spatial resolution

### Digital Elevation Model (DEM)

## Relief ( $r$ )

- Provides elevation data for each pixel with respect to coordinate system
- Readily available with extensive coverage
- Used as a base layer for nearly all digital soil mapping studies

# Relief ( $r$ )

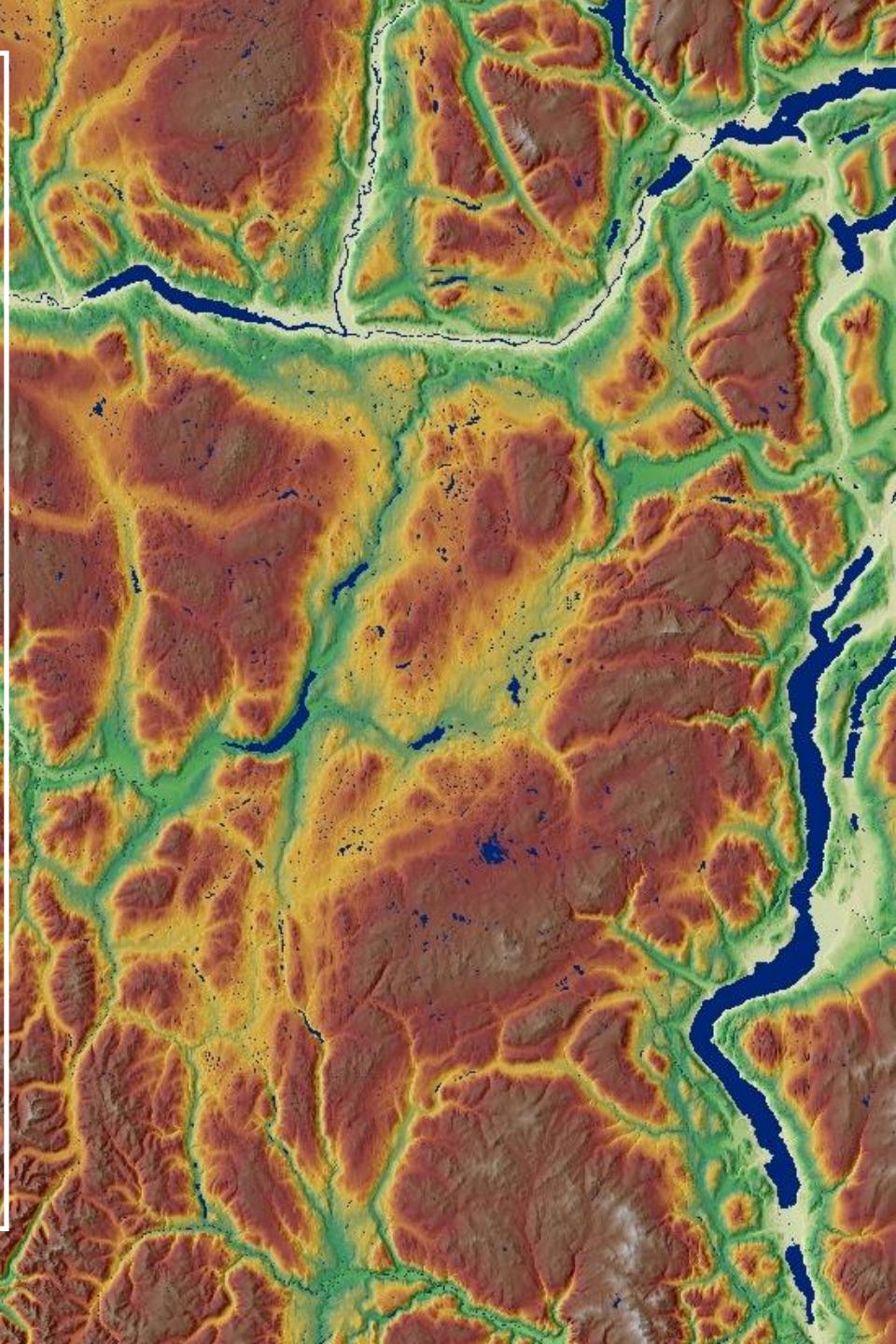


# Digital Elevation Model (DEM)

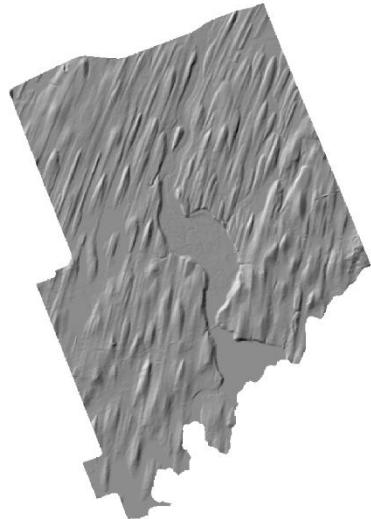
**Most commonly used dataset because it is widely available and cover large spatial extents**

**Used to derive many topographic derivatives:**

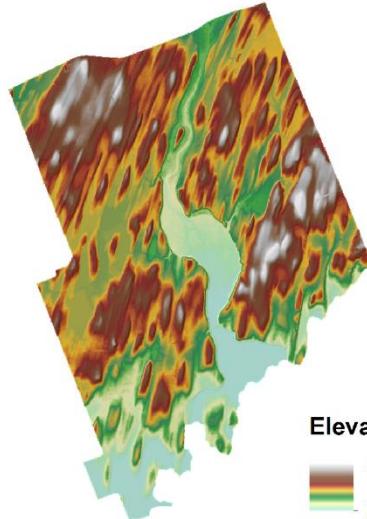
- Elevation, Slope, Aspect, Curvatures
- Solar Insolation, Skyview Factor
- Topographic Wetness Index, LS-Factor
- Landform Classifications
- Relative Slope Position
- Multi-Resolution Derivatives



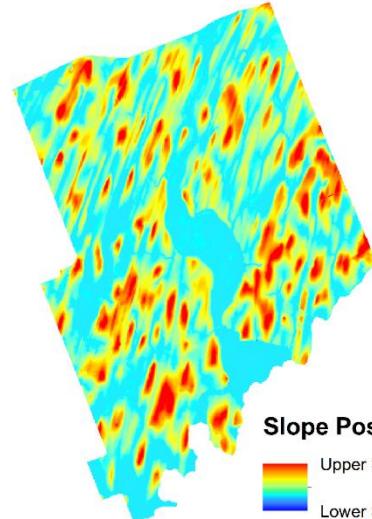
# Basic Terrain Derivatives



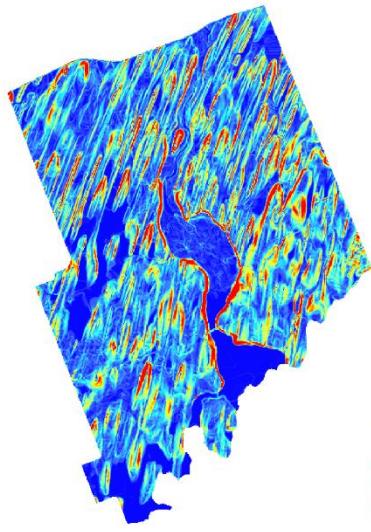
20 m Digital Elevation Model



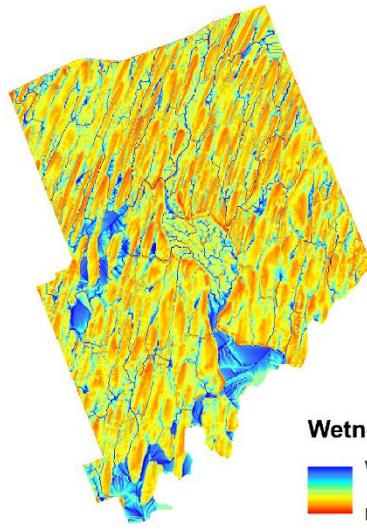
Elevation  
High Elevation  
Low Elevation



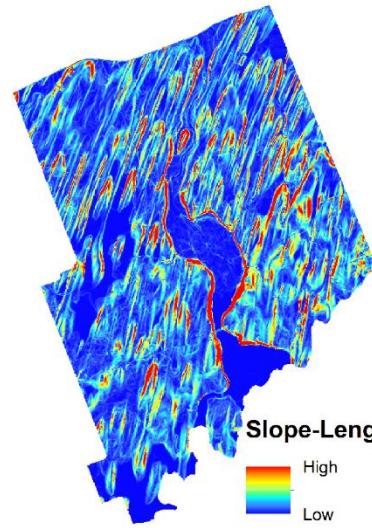
Slope Position  
Upper Slope Position  
Lower Slope Position



Slope  
Steep  
Level



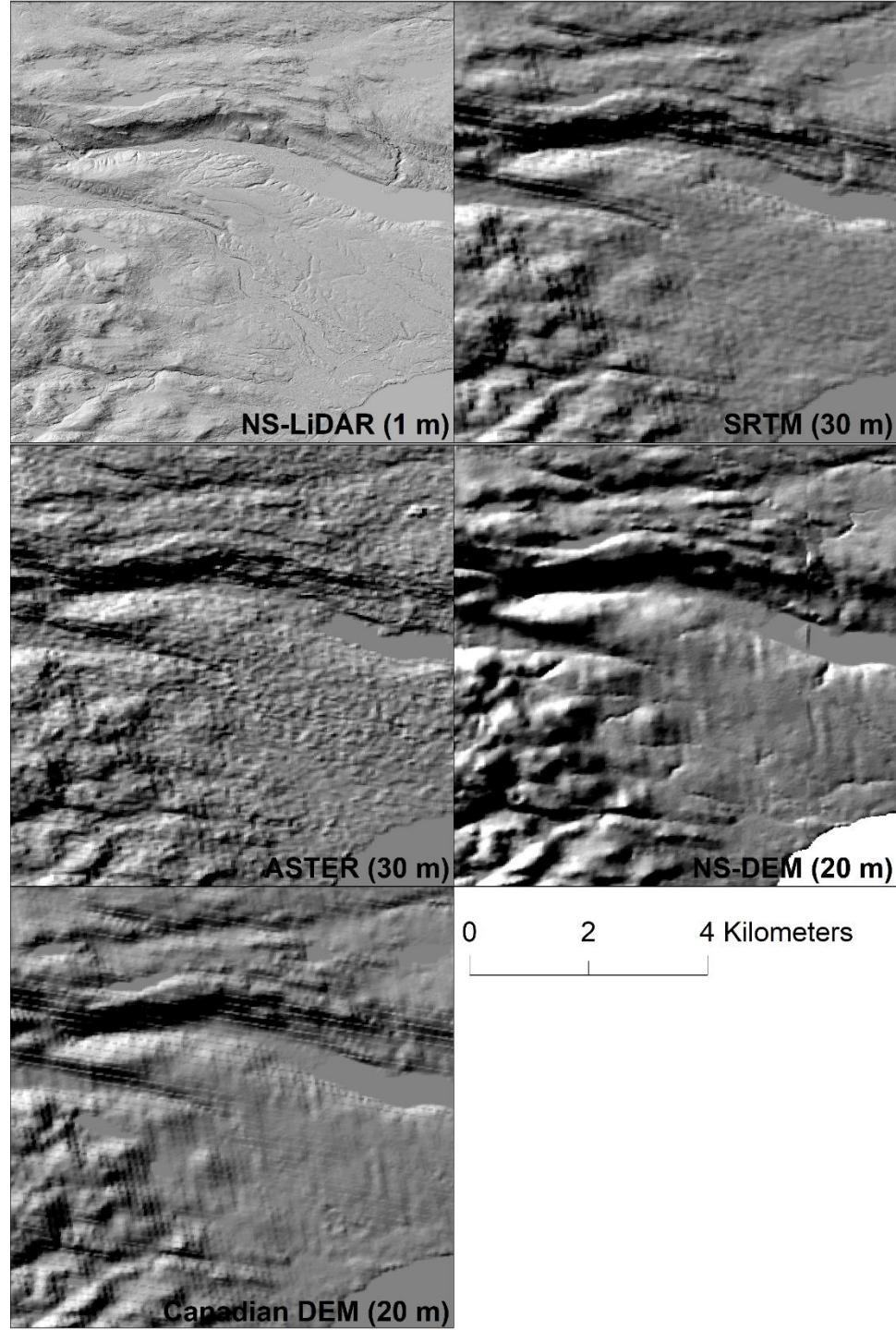
Wetness Index  
Wet  
Dry

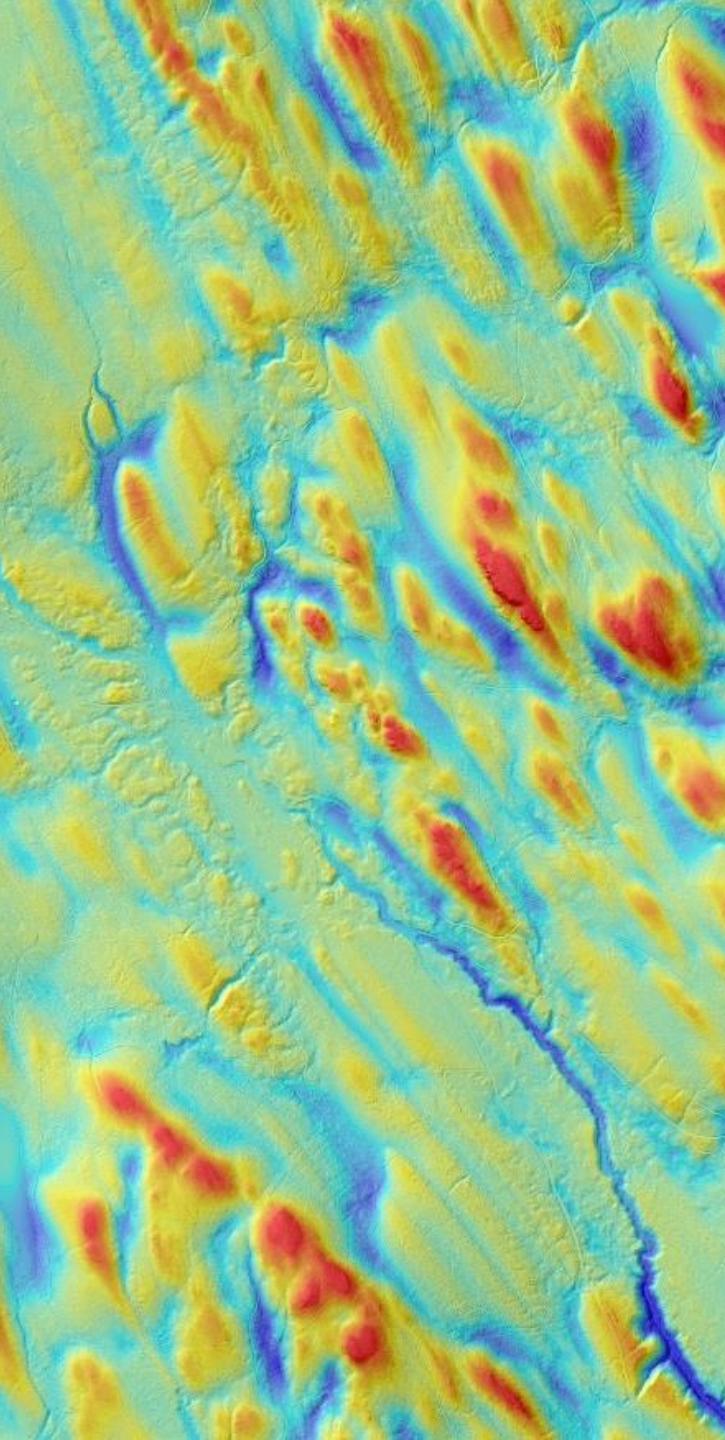


Slope-Length Factor  
High  
Low

# Some Problems

- Data anomalies – especially from satellite derived products.
- Increase in Resolution = Increase in Noise
- Influences of anthropogenic features
- Data storage
  - Float vs. Integer

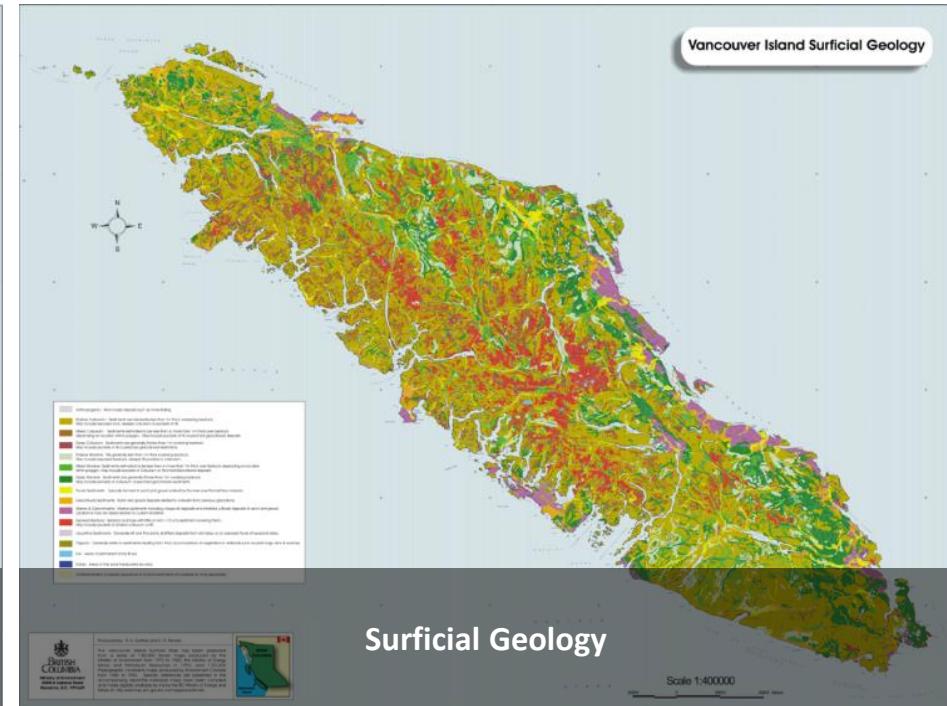
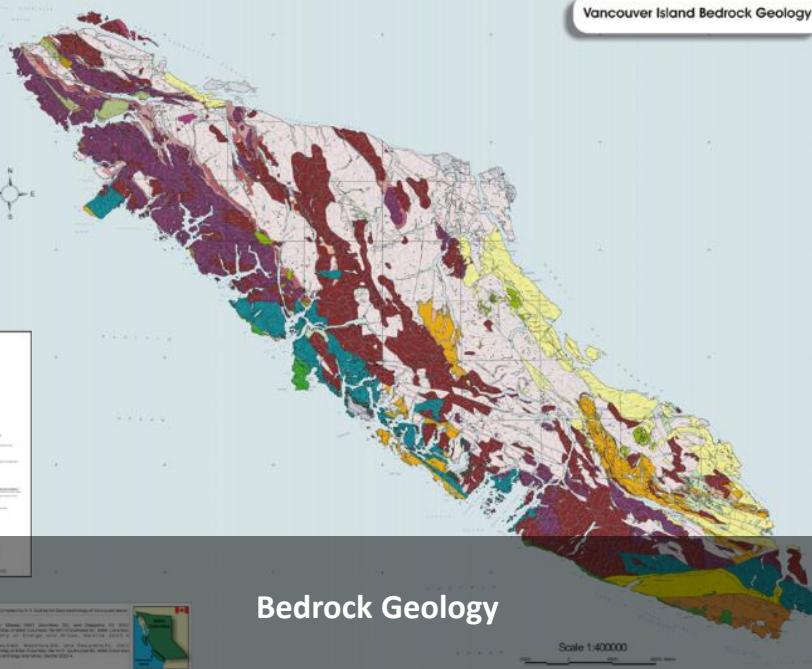




# Some Tips

---

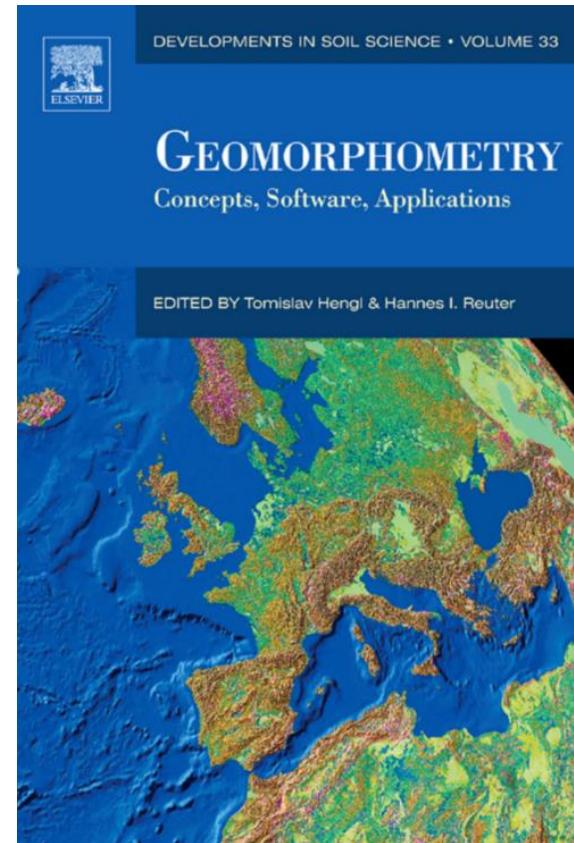
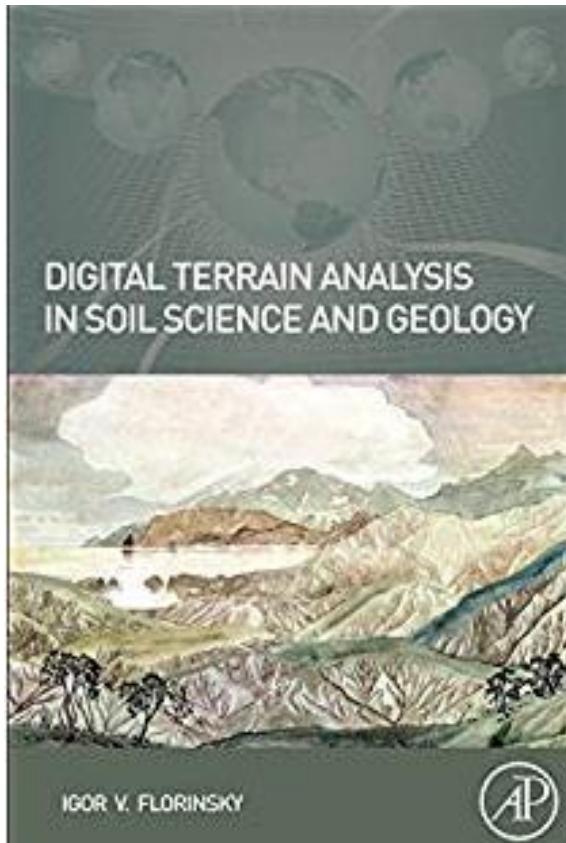
1. Always look at a hillshade model of the raw DEM.
2. Identify any anomalies or artifacts. These are usually the result of interpolation and data acquisition methods used for producing the DEM.
3. Apply pit removal algorithm and/or filtering.
4. Hydrological correction by reinforcing stream network.



# Parent Material ( $p$ )

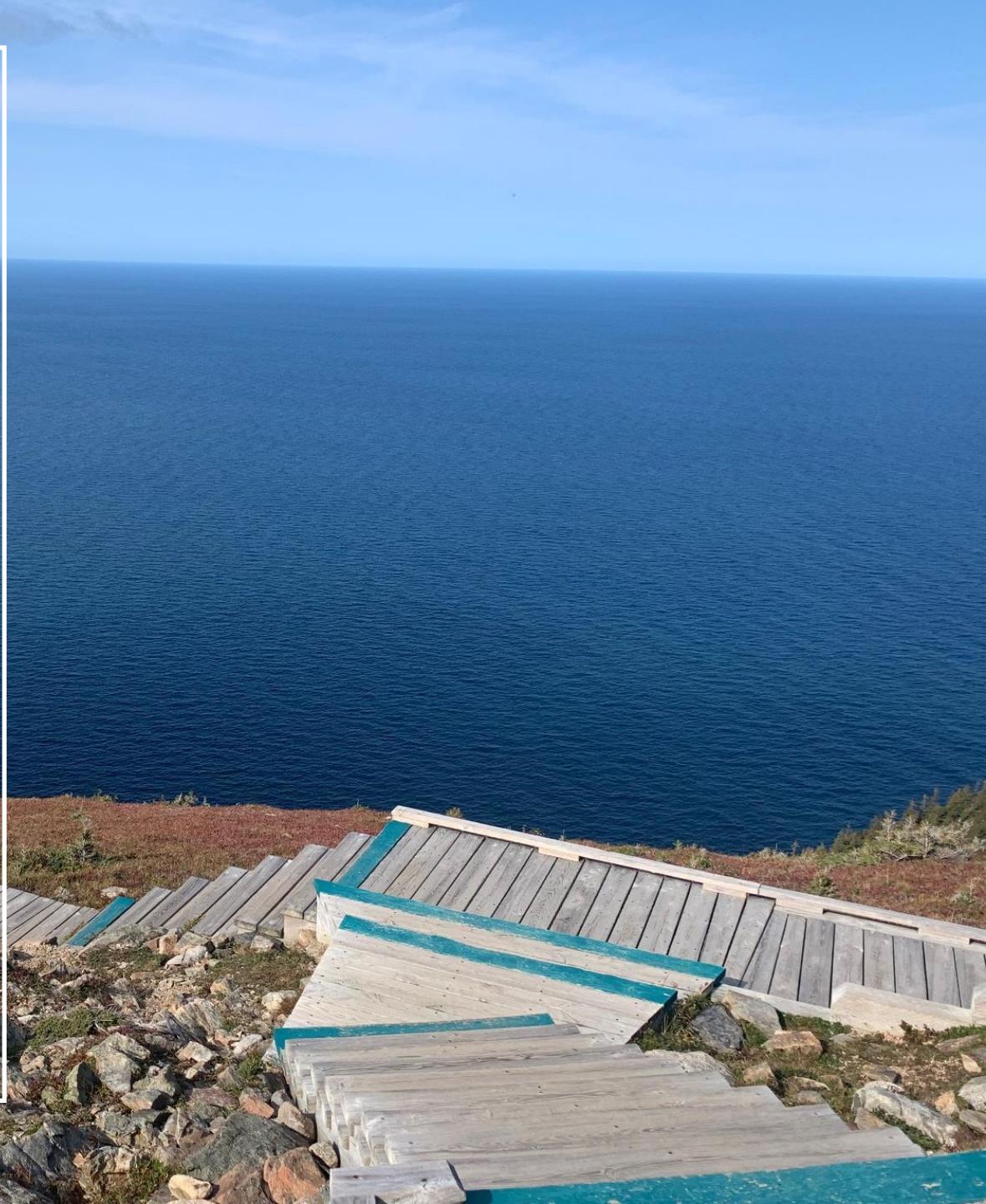
# Useful Resources

---



## Module 3.1: Creating Environmental Covariates

- Navigating the RSAGA package
- Geoprocessing with RSAGA
- Geoprocessing with Whitebox
- **Case Study:** A digital elevation model for Keene, ON will be used to generate a suite of topographic variables



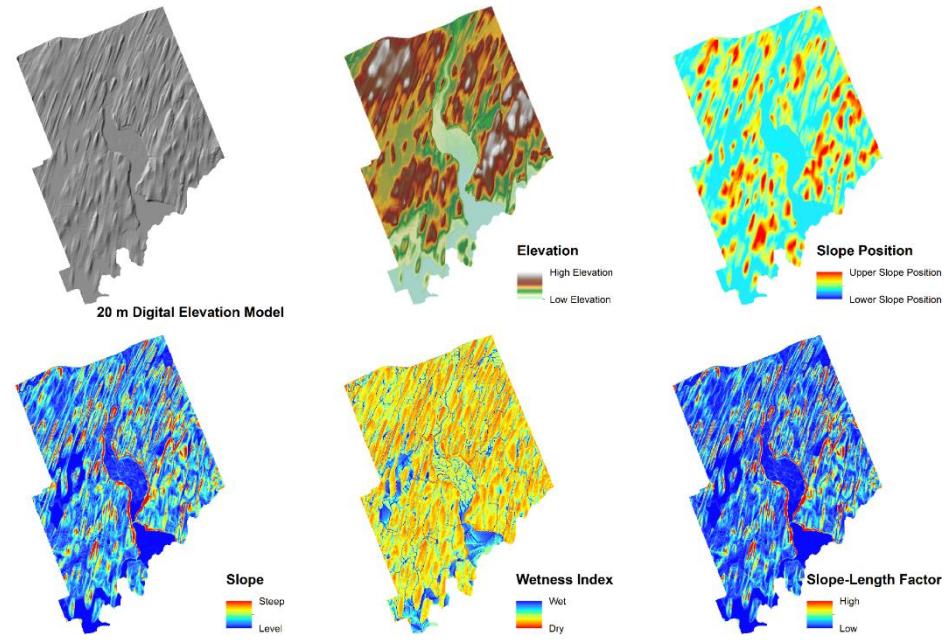
# Outline

1. Overview of Soil-Environmental Variables
2. Dealing with Multi-Collinearity



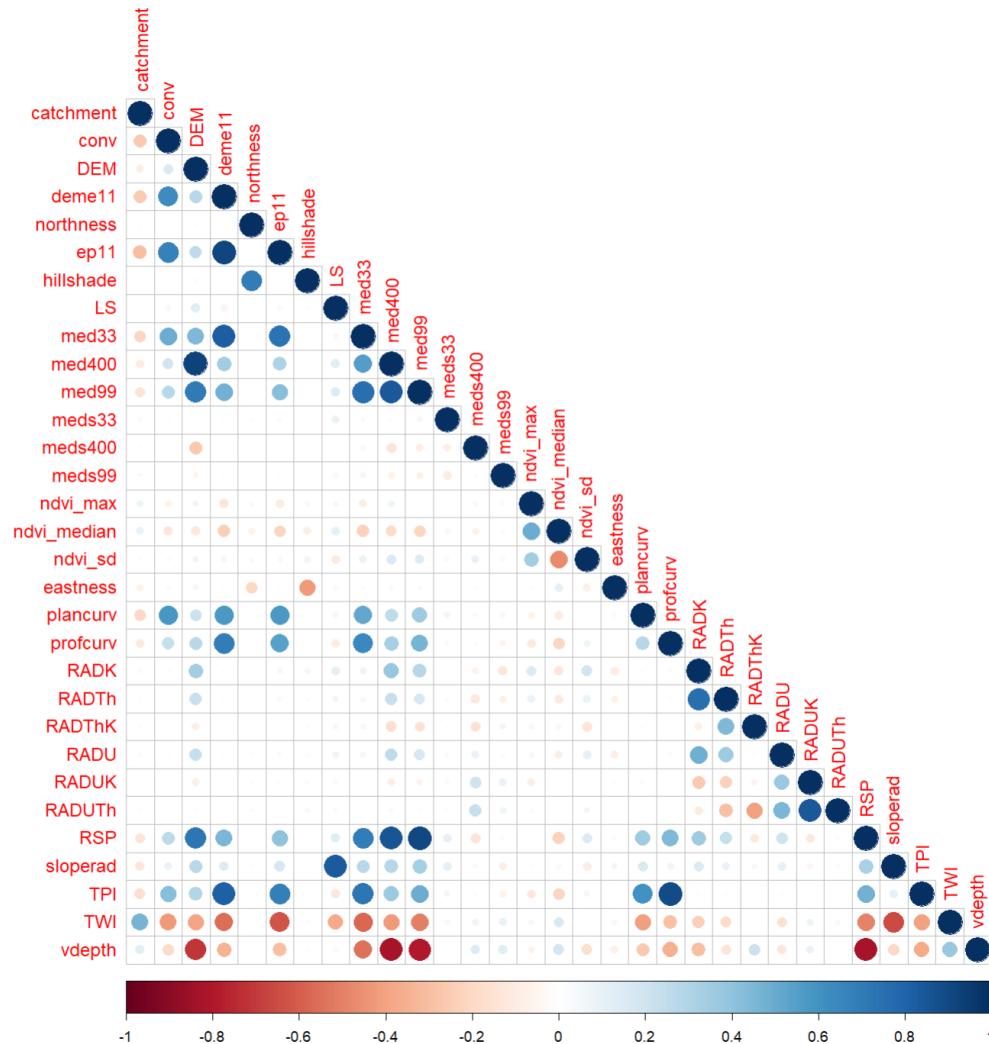
# On Multi-Collinearity—The Curse of Dimensionality

- Environmental covariates are often correlated.
  - Derivatives produced from the same DEM
  - Derivatives produced from multi-temporal remote sensing data
- Some models are sensitive to multi-collinearity.
- Difficult to interpret soil-environmental relationships.
- Reduces computational efficiency
- High-model complexity



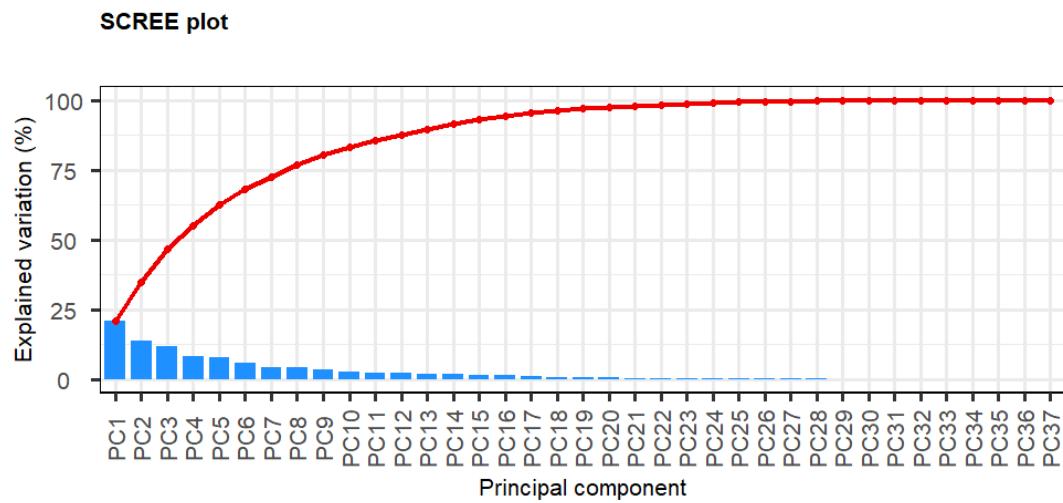
# On Multi-Collinearity—The Curse of Dimensionality

- Environmental covariates are often correlated.
  - Derivatives produced from the same DEM
  - Derivatives produced from multi-temporal remote sensing data
- Some models are sensitive to multi-collinearity.
- Difficult to interpret soil-environmental relationships.
- Reduces computational efficiency



# Solution 1: Principal Component Analysis

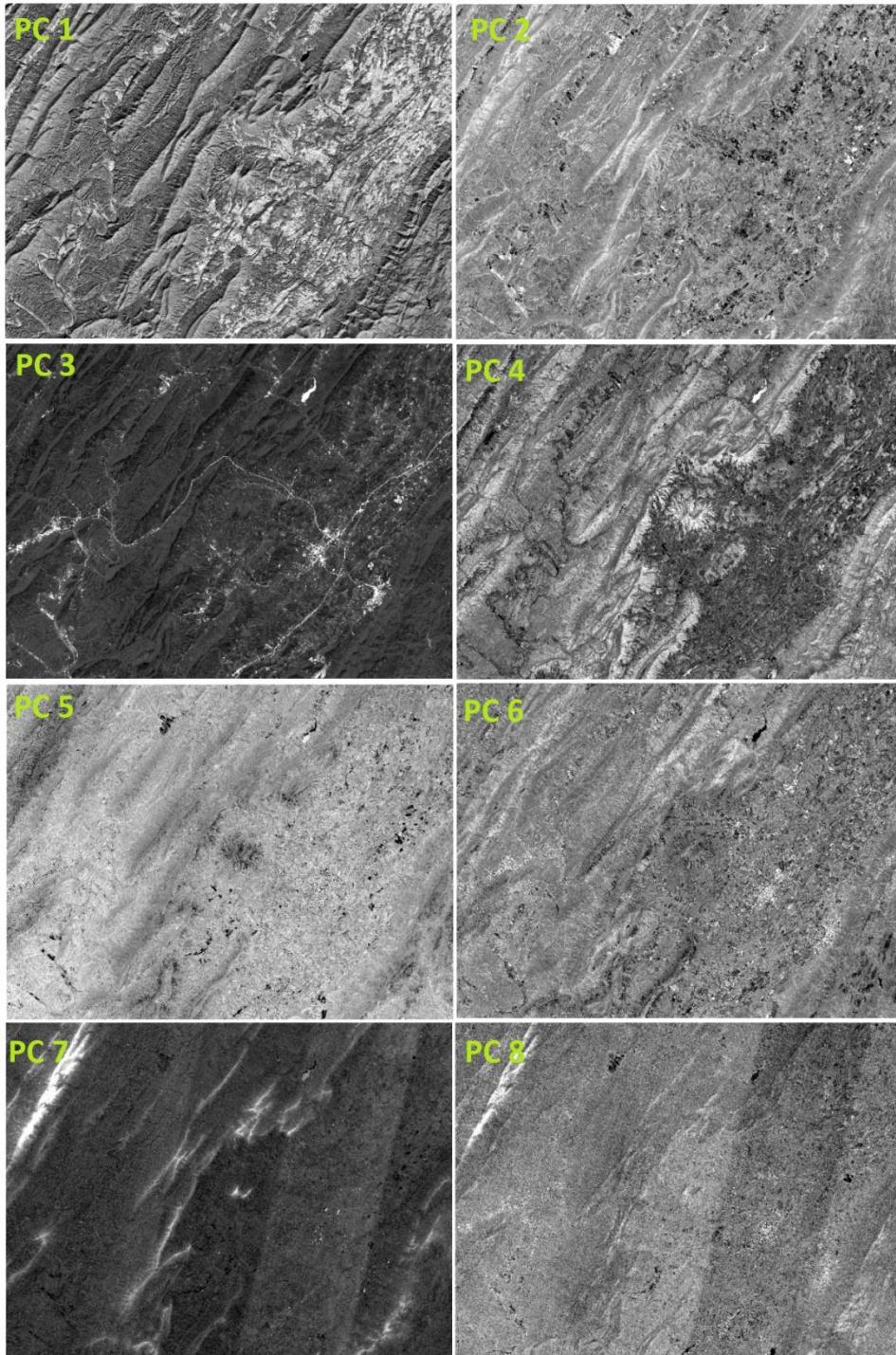
- Linear transformation of predictors into a series of non-correlated principal components.
- Each principal component accounts for some portion of the total variance.
- The principal components that account for 90-95% of the total variance are retained as predictors for modelling
- Higher PCs are noise



# Solution 1: Principal Component Analysis - Issues

---

- Difficult to interpret the soil-environmental relationships because all covariates are used to calculate each PC.
- Original covariates are lost.
- Could be computationally demanding for large datasets.



# Solution 2: Variable Inflation Factor Analysis

- **Step 1:** Run ordinary least squares regression whereby one predictor is fitted to all other predictors. Calculate  $R^2$ .
- **Step 2:** Calculate variable inflation factor for each variable.
- **Step 3:** Remove variable with highest VIF.
- **Step 4:** Repeat Steps 1-4 until all variables have a VIF of less than a threshold (e.g., VIF<5 or VIF<10)

$$X_1 = \alpha_0 + \alpha_2 X_2 + \alpha_3 X_3 + \cdots + \alpha_k X_k + e$$

$$\text{VIF}_i = \frac{1}{1 - R_i^2}$$

# Solution 2: Variable Inflation Factor Analysis

- The original covariates are retained.
- Easier to interpret soil-environmental relationships because the covariates are not transformed (in comparison to PCA).

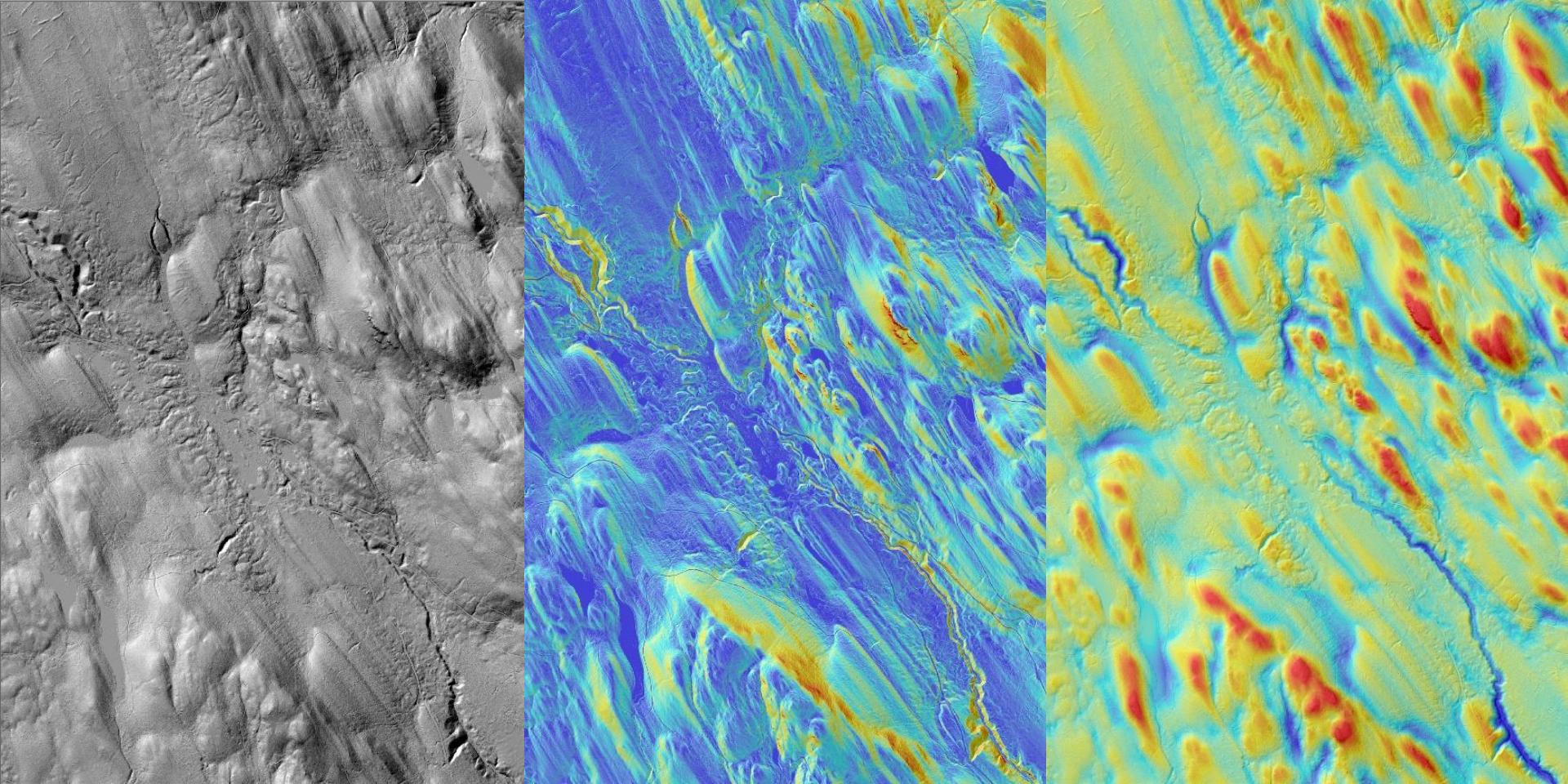
$$X_1 = \alpha_0 + \alpha_2 X_2 + \alpha_3 X_3 + \cdots + \alpha_k X_k + e$$

$$\text{VIF}_i = \frac{1}{1 - R_i^2}$$

## Module 3.2: Covariate Reduction

- **Case Study:** Variable inflation factor analysis will be carried out using environmental predictors for Keene, ON.





### 3. Environmental Covariates for Digital Soil Mapping

Canadian Digital Soil Mapping Workshop, 2022