Assessing and Mapping Groundwater Vulnerability to Bacteria in Alberta

Tamara Van Staden
Supervisors: Dr. Edwin Cey and Dr. Cathy Ryan

A thesis submitted for Bachelor of Science honours degree

April 2017

Abstract

Aquifer vulnerability index methods are commonly used for assessing groundwater vulnerability to surface contaminants. However, the methods have primarily been developed for dissolved contaminants. Microbial contaminants have unique characteristics that result in different transport behavior in the subsurface, and thus different tools need to be designed. Key vulnerability factors specific to microbial sources and subsurface transport mechanisms were identified in this study and incorporated into a model using an ArcGIS framework to create provincial-scale maps of groundwater vulnerability, specific to *E. coli*, in Alberta for the year 2012. Examples of these factors include: soil texture based on grain size, soil organic matter, hydrogeologic properties, depth to aquifer, and meteorological conditions. These factors were combined from individual GIS layers to create an intrinsic vulnerability map, demonstrating where aquifers were more vulnerable to bacterial contamination if a source became present. Maps were created for the growing season and cold season, and attempts were made to test the model with *E. coli* detection data. The results of these statistics were not significant enough for this model to be used for predictive purposes, but this could be caused by the presence or lack of risk (i.e., source of contaminants), as opposed to real differences in aquifer vulnerability. This project helped inform which factors should be considered when making a vulnerability map for bacterial contaminants, most notably temporal factors such as precipitation and soil moisture. The developed map provided insights as to where shallow aquifers in Alberta are intrinsically vulnerable to bacterial contamination.

Table of Contents

Table of Figures

Table of Tables

Table of Equations

1.0 Introduction

Groundwater is an important drinking water resource, especially for those who live far away from surface water sources. There are many private residences in Canada, especially in rural areas, that use groundwater as their primary drinking source. 30% of Canadians, and 90% of rural households in Alberta use groundwater as a drinking water resource [*Environment and Climate Change Canada*, 2013; *Alberta Agriculture and Forestry*, 2017]. The misconception is that groundwater is safe from contamination because of the geology that protects it. However, infiltration of water contaminated with bacteria is possible and these contaminants present a significant risk. An example of this is the Walkerton groundwater contamination tragedy, an incident in which 2,300 residents became ill and seven people died as a result of contamination of well water by microbes after a heavy rainfall, and the disinfection system failures that followed [*Hrudey et al*., 2003]. Health and safety concerns are associated with vulnerable groundwater, so it is important to protect.

The impacts of indicator bacteria presence in groundwater are important to monitor due to the number of private well users in Alberta. Private well users are more exposed to the risks of groundwater contamination, as they drink their water from the source rather than receiving it via a water treatment plant. The Canadian drinking water guideline for *E. coli* is none detectable per 100mL [*Health Canada*, 2017], but this is not regulated for private wells. When tested, bacterial contaminants are frequently found to exceed water quality guidelines [*van der Kamp and Grove*, 2001]. The decision to treat groundwater for consumption is up to the individuals who own or drink from the wells. There are more than 500,000 water wells across Alberta, which are often in rural or agriculture-dominated regions [*Alberta Agriculture and Forestry*, 2017]. The concern with wells being present in agricultural areas is that health and wellness risks increase with exposure to animal and human sources that contain pathogens such as manure fertilizer and septic systems [*Bradford et al*., 2013].

The sensitivity of a groundwater source to contamination by any substance or species is known as aquifer vulnerability [*National Research Council*, 1993]. Vulnerability is based on aquifer characteristics such as the depth to the water table, the material overlying and within the aquifer, and the fate and transport of the contaminant itself. The faster and easier it is for water and contaminants to travel through the subsurface, the greater the intrinsic vulnerability of that aquifer [*Dixon and Uddameri*, 2015]. Vulnerability risk assessment of aquifers has been a prevalent method for determining the health and relative susceptibility of aquifers to contaminants. The assessment can be represented as a map, providing water managers and other stakeholders with a visual indicator of aquifer vulnerability. Most commonly the assessments have been targeted towards dissolved contaminants such as nitrate, chloride, and pesticides. The lack of microbial risk assessment can be attributed to the complex biological and physiochemical characteristics of pathogens that convolute otherwise typical fate and transport mechanisms.

Models that currently exist to evaluate aquifer vulnerability use overlay weight of evidence (WoE) methods to determine the relative vulnerability of a region. They incorporate certain factors that are determined to be the most influential on the aquifer's susceptibility. One prominent model of this type is DRASTIC, and it considers seven factors in its weights of evidence formula: Depth to water (D), net Recharge (R), Aquifer media (A), Soils (S), Topography (T), Impact of vadose zone (I), and hydraulic Conductivity (C) [*Aller et al*. 1987]. DRASTIC was developed with the intention of assessing aquifer vulnerability to pesticides, which have different fate and transport processes than microbes. Bacteria are affected by processes such as sorption, advection, dispersion,

and diffusion that are typical of dissolved contaminants, but they have unique characteristics as colloidal particles and living organisms that influence their transport and fate in the subsurface. The need to survive is a unique characteristic of this type of contaminant. If bacteria are no longer living, they no longer have the potential of harming their receptors. Bacteria can also move themselves using their flagella towards areas with higher food supply (chemotaxis), another unique characteristic [*Corapcioglu and Haridas*, 1984]. This project aims to make a similar WoE model that has been altered to include vulnerability factors that specifically affect bacterial fate and transport, as there are very few models that do this currently. One example of a vulnerability map that was created for bacterial contaminants was made by *Dixon and Uddameri* [2016], which combined intrinsic vulnerability (DRASTIC methods) with specific vulnerability (factors such as soil moisture and organic matter).

Three mechanisms influence bacterial fate and transport: attachment to aquifer medium, bacterial survival, and bacterial movement through pores. Each of these mechanisms are influenced by other subsurface processes. Attachment of particles to the subsurface material, bacterial or otherwise, will be affected by pH and the soil material [*Corapcioglu and Haridas*, 1984; *Bradford et al*., 2013]. Bacterial survival will depend on the temperature of the aquifer, the amount of organic matter present, the amount of moisture in the soil, and the pH of the soil [*Corapcioglu and Haridas*, 1984; *Crane and Moore*, 1984; *Sjogren*, 1994; *Conner and Kotrola*, 1995; *Wang et al*., 2004; *Tufenkji*, 2007; *Bradford et al*., 2013]. The movement of bacteria into the soil and through the subsurface is affected by the amount of water in the soil, the amount of precipitation that occurs in the area, the depth to the water table, and the hydraulic conductivity of the material that bacteria will be moving through to get to the water table [*Conboy and Goss*, 2000; *Curriero et al*., 2001; and *Bradford et al*., 2013]. Details and sources regarding the impacts of the above factors are discussed in Table 1. Converting these mechanisms into a spatial format and using them in an aquifer vulnerability model would allow for the production of an intrinsic aquifer vulnerability map, which indicates where groundwater would be vulnerable to bacteria-specific contamination.

Validation of vulnerability maps is an important but difficult process. The majority of vulnerability maps and methodologies are not validated after they are produced [*Leal and Castillo*, 2003]. There is also no standardized validation method for aquifer vulnerability models [*Neukum et al.*, 2008]. Validation can be attempted by comparing the map to field data for the contaminant of concern. Links between existing bacterial data and the map will potentially illustrate the effectiveness of the developed map. However, the detectability of pathogens and pathogen indicators is more complicated than that of dissolved contaminants, and can be sporadic and non-representative of overall groundwater conditions. Typically, there is a lack of reporting or submitting regular samples, which is necessary for bacterial detection, and the decreased reliability of these samples makes the results of monitoring bacteria a less accurate representation of bacterial presence [*Batterman et al*., 2009]. The most reliable data comes from repeated and temporal sampling, which can be rare. Another option would be to test the developed model by comparing each map with the disease outbreak distributions from that year.

The protection of drinking water wells is important because groundwater contamination can lead to severe health risks [*Hrudey et al*., 2003]. Intrinsic vulnerability maps can aid in determining where aquifers are susceptible to contamination and help decision makers decide where drinking water wells should be constructed. The primary goal of this project is to fill a gap in vulnerability mapping methodology by creating a model that can determine where there are shallow aquifers that are intrinsically vulnerable to bacterial contamination. Key vulnerability

factors specific to microbial sources and subsurface transport mechanisms were identified. Factors that were appropriate, and had available data, were selected and converted into ArcGIS format. Climatic data were also included to generate maps for the growing season and cold season of 2012. The final output was an intrinsic bacterial vulnerability map for shallow aquifers in the province of Alberta. The suitability of the map was tested by comparing the map with *E. coli* detection data.

2.0 Methods

2.1 Site and Data Description

2.1.1 Site Description

The study area was focused on the "white zone" of the province of Alberta, which comprises the settled portions of the province where the majority of the population lives (northern, southern, central, foothills, and Peace River regions), and coincides with the agricultural region of the province where appropriate soil property information is available. Regions are highlighted in Figure 1, with labels indicating the name each region will be referred to throughout this document.

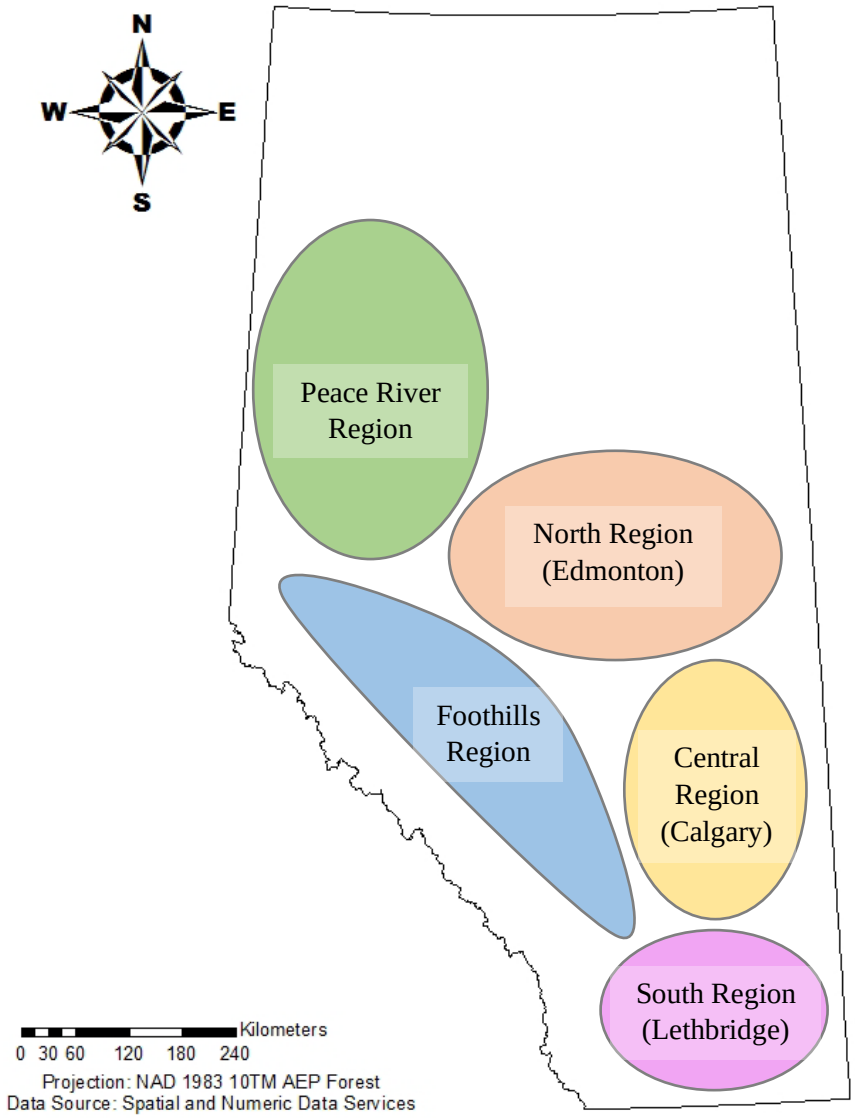

Figure 1. Names assigned to each region covered by the vulnerability map.

2.1.2 Data Description

The selected vulnerability factors were: soil moisture, pH, soil texture, organic matter, depth to water table and hydraulic conditions, and precipitation. Information and sources for the determination of these factors are provided in Table 1. Soil temperature was excluded from this list because of the complexities that came from determining temperature at depth, and the low likelihood of large subsurface temperature variability across Alberta. Although temperature influences bacterial survival, the temperature of groundwater is likely not significantly influenced by short-term air temperature fluctuations. There are other temperature-related factors, such as climate and weather, that are more influential on bacteria than the actual soil temperature. Instead, this study focussed on the general "hot" and "cold" seasonal patterns. The literature suggests that bacteria survival is higher in warmer temperatures and that there are more outbreaks in the summer [*Goss et al*., 1998; *Charron et al.*, 2005; *Valeo et al.*, 2016]. In the cold season – defined by Alberta Agriculture and Forestry (AAF) – temperatures freeze the ground and prevent infiltration. The cold season and growing season are defined by the AAF as follows: the cold season is October 1 to March 31 and the growing season is April 1 to September 30.

The ionic strength of groundwater was determined to be an important factor as well, but could not be included because of the lack of available data. Topography was also considered, but its overall influence on the transport and fate of bacteria has been observed to be minimal [*Adesiyun et al*., 1983; *Cui et al*., 2016].

The temporal (precipitation and soil moisture) data collected were for the year 2012, which was selected because it is the most recent year before the last major flood in the province. Flood impacts on water quality have been studied and have been found to be important [*Wade et al*., 2004]. However, this was not the focus of this study.

To simplify researching the impact of different factors on bacterial vulnerability, the type of bacteria was limited to *E. coli* (EC) because it is a very commonly used and tractable indicator of fecal contamination, which is the main parameter analyzed when considering bacterial groundwater contamination [*Invik*, 2015].

The data for organic matter, soil texture, and pH layers were obtained from the Alberta Detailed Soils geodatabase provided by Agriculture and Agri-Food Canada (AAFC), which was provided in a GIS polygon shapefile. The remaining data were obtained in an Excel format from various sources. The depth to the water table/hydraulic conductivity layer was obtained by using the geological information, such as lithology and the corresponding depths, from well reports in the Alberta Water Well Information Database (AWWID). The precipitation and soil moisture data were obtained from the Alberta Climate Information Service (ACIS) through Alberta Agriculture and Forestry (AAF).

Table 1. List of all the key vulnerability factors, their effects, and cited sources.

Factor	Condition	Effect	Source
Soil Moisture	Low	• Low moisture increases filtration	*Corapcioglu and Haridas* [1984]
	High	• More moisture led to more total coliforms	*Mellor and Cey* [2015]
Depth to Water Table	Shallow	• Shallow aquifers are more vulnerable	*Pandey et al.* [2014], *Cui et al.* [2016]
	Deep	• Deeper wells have longer travel times for bacteria and lower chance of contamination due to die-off	*Conboy and Goss* [2000]
pH	Low	• Bacteria are negatively charged, so low pH would mean more adsorption	*Corapcioglu and Haridas* [1984], *Bradford et al.* [2013]
		• Lower survival rate of bacteria in lower pH	*Sjogren* [1994]
	Neutral	• Quartz has a negative surface charge, metal oxides have a positive charge, bacteria attracted to positive charges	*Ginn et al.* [2002]
	High	• Less microbe attachment to suspended particles	*Guber et al.* [2009]
		• Can mobilize biocolloids in saturated systems	*DeNovio et al.* [2004]
		• Biocolloid mobilization is independent of pH	*Grolimund and Borkovec* [1999]
		• Lower collision efficiency and less attachment	*Schijven et al.* [2006]
Soil Properties	Generally	• Particle size distribution and clay content influence microbial migration	*Butler et al.* [1954]
		• Found no significant difference in transport rates between clay or sand soils in saturated or unsaturated conditions	*Safadoust et al.* [2011]
	Clay	• Macropores can form, especially in shrinking clay, but otherwise good filtration	*Safadoust et al.* [2011]
		• Clay can act as protection for bacteria from UV radiation and antibiotics	*Marshall* [1980]
		• Bacteria survived longer with at least 25% clay in the soil	*Burton et al*., [1987]
	Sand	• Bacteria + sand have optimum retention in saturated conditions. Weathered sand soil had less transport than weathered clay	*Invik* [2015]
		• Sand might have a higher inactivation of *E.coli* than areas without sand	*John and Rose* [2005]
		• Sandy soil has been found to provide some protection from contamination	*Conboy and Goss* [2000]
Organic Matter		• Bacteria that can grow outside of organisms can grow on organic matter, which can also be a source of nutrients which affects survival	*Corapcioglu and Haridas* [1984]
		• Organic matter also competes for adsorption space	*Bradford et al.* [2013], *Guber et al.* [2004], *Schijven et al.* [2006]
		• The above found for viruses as well	*Sobsey et al.* [1980]
		• High concentrations of available carbon can hinder irreversible sorption of bacteria	*Marshall et al.* [1971]
Seasonal Variations	Generally	• Seasonal changes occur in a triangular wave	*Schijven et al.* [2006]
		• August and September have higher risk of *E. coli* contamination	*Invik* [2015]

	Growing Season	• More exfiltration	*Schijven et al*. [2006]
	Cold Season	• Colder, higher water table (better for pathogens) • High water table is more vulnerable	*Schijven et al*. [2006] *Elçi* [2012]
Hydraulic Conditions	Generally	• Microbe transport is heavily influenced by aquifer flow	*Unc and Goss* [2004]
		• Flow of groundwater has been observed to be an important consideration for well susceptibility	*Conboy and Goss* [2000]
	Low flow	• Better retention esp. with small grain size • No flow can act as storage for biocolloids	*Bradford et al*. [2006] *Keller and Auset* [2007]
	High flow	• Larger grain size leads to higher hydraulic conductivity → "velocity enhancement"	*Ginn et al.* [2002]
		• Macroporosity increases flow rate which increases pathogen transport	*Guber et al*. [2004]
		• High velocity had higher output concentration even with small grain size, because it decreases straining	*Bradford et al*. [2006]
		• "Velocity enhancement" decreases with distance	*Keller and Auset* [2007]
Precipitation	Low	• Low precipitation would lead to dry conditions, increasing the number of air-water interfaces, there would be more filtration	*Corapcioglu and Haridas* [1984]
	Intermittent	• Moving of air-water interfaces increases mobilization	*DeNovio et al*. [2004]
	High	• More rain → more pathogens in groundwater, increase turbidity and ionic strength	*Crane and Moore* [1983]
		• Heavy rainfall will lead to biocolloids reaching the water table fast	*Bradford et al*. [2013], *Curriero et al*. [2001], *Cey et al*. [2009]

2.1.3 GIS Methods

Producing a vulnerability map requires selecting important factors and creating a layer for each one based on the influence it has on aquifer vulnerability. For example, data for pH can be represented spatially by using corresponding coordinate data and importing into GIS. One layer of information can then be produced and re-interpreted to reflect the impact different pH levels can have on bacterial survival and movement. Vulnerability indexes (VIs) are ranges of values specific to each factor that represent the relative influence that factor has on the overall aquifer vulnerability. VIs were selected based on information from literature, data distributions, and existing regulations. Using pH as an example, a low VI value would be used for pH conditions that are to harmful bacteria because they would not be able to survive (1 in this case). A high VI value would be used for pH conditions that are optimal for surviving (4 in this case). The data were reclassified into a new data layer that represents areas of high and low vulnerability based on these VIs. Each selected factor was converted into a GIS format and given VIs (i.e. ratings) based on their impact on bacterial fate and transport as described above. A vulnerability layer was created for each factor, all of which were overlaid on top of each other to produce an intrinsic bacterial vulnerability map for shallow aquifers in the Province of Alberta. An example of the layering system of index methods is shown in Figure 2.

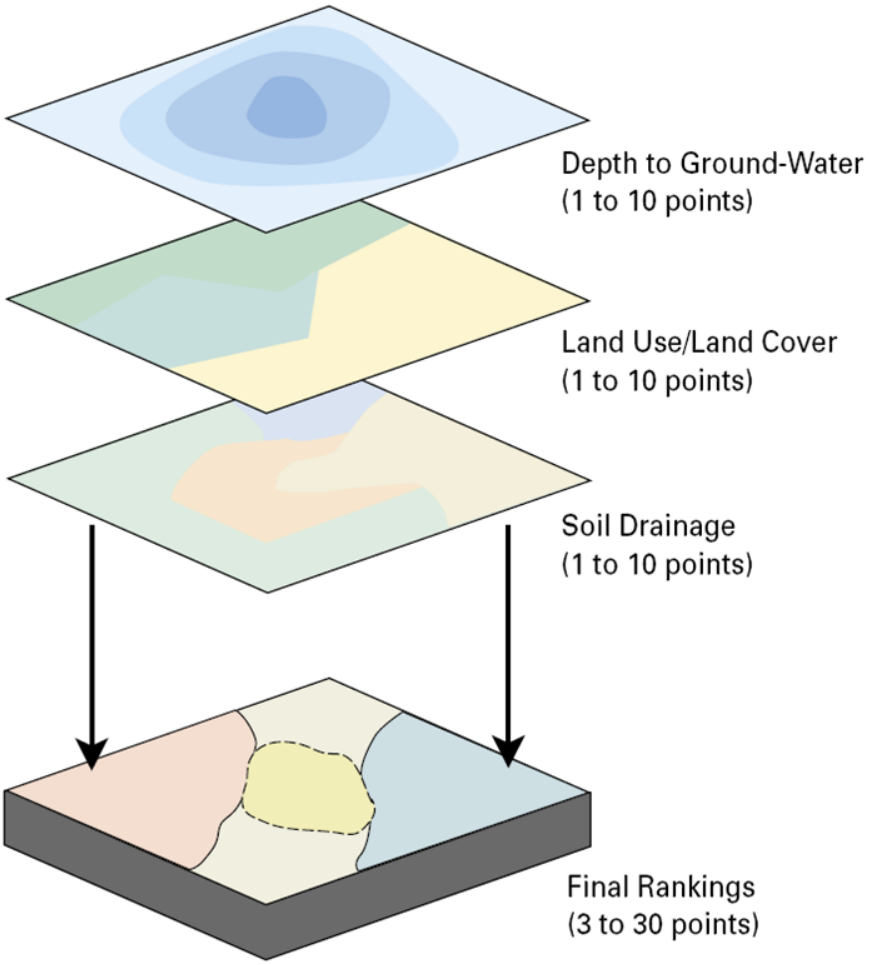

Figure 2. An example of GIS layering for determining aquifer vulnerability using index methods. Retrieved from https://pubs.usgs.gov/circ/2002/circ1224/html/new.htm [*Focazio et al.*, 1984].

Data that were not already in GIS format were loaded into GIS with their corresponding spatial information and turned into a point shapefile. This shapefile was converted into a raster format using kriging. The result of the kriging was then reclassified into the VIs that were defined, and this was clipped to the shapefile of Alberta. Data that were obtained already in GIS format were also reclassified into the determined VIs in raster format.

The projection selected for the map was NAD 1983 10TM AEP Forest. This is a common projection used for the entire province of Alberta. Data that were imported from Excel that didn't have a coordinate system were set to NAD 1983 (2011). The data were imported into a file geodatabase set to the 10TM AEP Forest projection. The final output of the map was in a raster format. The raster resolution selected was 100m by 100m, as recommended by *Elçi* [2012] for groundwater vulnerability assessment.

2.2 Climatic Data

2.2.1 Precipitation

The size of individual precipitation events is important when considering EC contamination of aquifers, as large rainfall events are connected to increased disease outbreaks [*Curriero et al*., 2001]. The average event size and the variance of the event size together at one weather station both represent the potential for that location to have large precipitation events. To demonstrate the effect of precipitation event size in a spatial environment, a new precipitation variable was developed. The variable took the average amount of precipitation per event into account (total precipitation per season/number of events in that season), as well as the variance of the precipitation on days when there was a rain event. The variable (P_v) was calculated separately for each weather station for both the cold season and the growing season by multiplying the mm/event (P_e) and the variance of all the events (V_e). The formula for P_v (mm^3/event) is:

$$P_v = P_e \times V_e \qquad \text{(Equation 1)}$$

where P_e represents the average amount of precipitation per event in that season (mm/event), and V_e represents the variance in event size (mm^2). The final units of P_v are mm^3/event. The values for all weather stations in both seasons are available in Appendix A in the "Precipitation" sheet.

Precipitation VI values were selected using literature, combined with patterns observed within the data (Table 2). Disease outbreaks caused by waterborne pathogens tend to be preceded by precipitation events that are in the 80^{th} and 90^{th} percentile in size [*Curriero et al*., 2001]. This information was used to determine the ranges of the P_v that would lead to a higher aquifer vulnerability. For cold and growing seasons separately, the 80^{th} and 90^{th} percentile values of P_e and V_e for that season were calculated. These values were multiplied together (respectively) using Equation 1 and represented the 80^{th} and 90^{th} percentile P_v values for both seasons, and were used as VI values. The lower bounding range limits were determined by using the median values of the P_v data for both seasons. The median value coincided with the peak of the density graphs for either season, which represented the most common event size and not extreme events. The lower vulnerability indexes represent data from the cold season, and the higher indexes represent data from the growing season, but all indexes can influence data from either season.

Table 2. List of vulnerability index ranges for the calculated precipitation variable (units mm^3/event) for both the cold and the growing season, and their given vulnerability indexes.

Precipitation Variable (mm^3/event)	Vulnerability Index
0 – 5.27	1
5.27 – 15.70	2
15.70 – 29.53	3
29.53 – 186.18	4
186.18 – 364.01	5
364.01 – 492.64	6
492.64+	7

2.2.2 Soil Moisture

Soil moisture data were obtained as volumetric moisture content (VMC) measured from locations across the province. The VMC is calculated using the Versatile Soil Moisture model, and provides an estimate of the soil moisture contained in the soil profile from groundwater surface to a specific depth for both spring wheat (mm/120cm) and pasture (mm/60cm). The land use (wheat or pasture) selected for the VMC model depended on the dominant land use at the time of measurement. VMC was not directly measured, but calculated from both land use measurement types and estimations made from field data [*Hayashi et al*., 2012]. The soil moisture data provided by AAF included spatial point data for each township within the "white zone" for every day of 2012. The seasonal average was calculated for each data point, so a single layer could be produced in GIS. After the point data for each season were imported into ArcGIS, a layer was created by kriging the point data.

The soil moisture VIs were determined using the quartiles that came from using the volumetric moisture content (VMC) data for both the cold season and the growing season (Table 3), with increasing soil moisture corresponding to an increase in aquifer vulnerability.

Table 3. List of determined volumetric moisture content (VMC) ranges, which have been converted to a percentage, for both the cold and the growing season, and their given vulnerability indexes.

VMC Range (%)	Vulnerability Index
11.28 – 13.52	1
13.52 – 15.22	2
15.22 – 17.67	3
17.67 – 22.96	4

2.3 Geologic Data

2.3.1 Hydraulic Resistance

One model that has been developed for the purposes of understanding aquifer vulnerability is the Aquifer Vulnerability Index (AVI) method, which uses the thickness of each geological layer above the water table and the corresponding hydraulic conductivity to calculate hydraulic resistance [*Van Stempvoort et al*., 1993]. Hydraulic resistance represents an estimate of protection that geology overlying the water table provides for aquifers. Hydraulic resistance was the only value used in the AVI method to assess groundwater vulnerability [*Van Stempvoort et al*., 1993; *Simpson et al*., 2014]. In this study, it will be one of many layers. To account for depth to the water table and hydraulic conductivity, well log data from the Alberta Water Well Information Database (AWWID) were used in order to analyze the subsurface geologic properties, as well as the depth

to the shallowest aquifer. These data were combined to calculate a hydraulic resistance value (s) for each well with the following formula:

$$R = \sum(^{d}/_{K}) \quad \text{(Equation 2)}$$

where d (m) is the thickness of each geologic layer above the water table, and K (m/s) is the hydraulic conductivity assigned to that geologic material.

Hydraulic resistance was calculated for each groundwater well that was in a shallow aquifer in the dataset. Shallow aquifers were the focus of this study because they are more vulnerable, and because bacteria have a finite lifespan. Knowing the water level in the well was critical for the hydraulic resistance calculation because in shallow aquifers, it indicated where the water table was, and therefore the total depth over which hydraulic resistance was calculated. The well log data did not include water table information, so the well screen mid-point or static water level was used instead. The shallower depth was used if the well report had both of these values, but in most cases the only available information was the static water level. Each analyzed well needed to be shallow or unconfined to ensure that the static water level was representative of the top of the water table. A 30m limit was established as the maximum depth for a shallow aquifer, and wells that had a water level deeper than 30m were removed. The factor layer, and therefore the vulnerability map, will only be applicable to shallow aquifers.

Each well report contained layered geological data. The hydraulic resistance was calculated for each layer above the water table and added together to determine total hydraulic resistance. Well reports that had incomplete or non-descriptive data were removed from the dataset. A hydraulic conductivity was assigned to every material type listed in the well logs. Hydraulic conductivities were determined by taking the midpoint of existing hydraulic conductivity ranges for specific materials [*Clapp and Hornberger* 1978; *Freeze and Cherry*, 1979; *Rehm et al*., 1980; *Lee et al*., 2001]. After the calculation for each well, the data were imported into ArcGIS and interpolated using kriging to make a raster layer that would be used in further analysis. More details regarding the removal of wells for quality assurance, and the assigning of hydraulic conductivity values are provided in Appendix B.

The VI ranges of hydraulic resistance were determined by using a combination of natural bins in the data, the lower four deciles of the data, and a regulation for wellhead protection zones (Table 4) [*Moore*, 1993]. Resistance was calculated with units of seconds, and was log-transformed prior to classification. Larger values of hydraulic resistance represented more aquifer protection. The upper VI range was the most important to define because any resistance value within that range would indicate that bacteria would not survive. The maximum value was determined by using the maximum time-related capture zone determined by a study conducted in the United States for the Environmental Protection Agency. The 400-day capture zone represents the steady state zone of influence if a well is being pumped at 650 gallons per minute. The study also recommends a 200-day time-of-transport limit for protection against pathogen contamination, which is considering the lower bacterial survival rate in tortuous conditions [*Moore*, 1993]. However, the 400-day limit was used instead for the purposes of remaining conservative. The lower limits were defined using natural bins that occurred within the data. After classifying the upper limit, the remaining data were comprised of the lower four deciles, which were then used to create the remaining VIs.

Table 4. List of log(resistance) ranges for the dataset obtained from the Alberta Water Well Information Database (AWWID), and the vulnerability index value given to each range.

Time (days)	Corresponding resistance value (log_{10}[seconds])	Vulnerability Index
Minimum – 0.25	**-0.52 – 4.33**	4
0.25 – 10.15	**4.33 – 5.94**	3
10.15 – 400	**5.94 – 7.54**	2
400+	**7.54+**	1

2.3.2 Soil Texture

Soil texture is important because it can affect bacterial attachment and movement in the subsurface. For information about how soil texture data was incorporated into GIS, see Appendix B. For the soil texture vulnerability factor classification, similar classifications used in other vulnerability models were examined [*Aller et al*., 1987; *Dixon and Uddameri*, 2016].

Another category was created additionally to consider the influence of macropores. Macropores are cavities in the soil that can become preferential flow pathways that have the potential to direct contaminated water to aquifers at a faster rate [*Cey et al*. 2009]. The presence of macropores would therefore increase the aquifer vulnerability. Clay-rich soil can form macropores as the clay expands and shrinks with changing moisture conditions. The VI values were changed from *Dixon and Uddameri* [2016] to increase the risk of clay. More changes were made to sand index values which were lowered because of cases where sand in soil has been found to increase protection from contamination or inactivate/immobilize bacteria [*Conboy and Goss* 2000; *John and Rose* 2005; *Invik* 2015]. Values for soil texture identifiers other than sand and clay were determined from previous DRASTIC values from *Aller et al*., [1987] and altered based on increasing/decreasing presence of sand and clay. The values determined from the alternative impacts of sand and clay were found to have no correlation with the EC detection maps, so the *Dixon and Uddameri* [2016] values were used instead (Table 5).

Table 5. List of soil texture identifiers and corresponding vulnerability indexes from *Dixon and Uddameri* [2016] along with an alternative vulnerability index developed based on potential interactions of EC with sand, and macropores in clay.

Soil Texture	*Dixon and Uddameri* [2016]	Alternative Vulnerability Index
Sandy Loam	**8**	1
Loamy Sand	**9**	1
Sand	**10**	2
Silt Loam	**3**	3
Sandy Clay Loam	**4**	4
Silty Clay Loam	**2**	5
Loam	**5**	5
Clay Loam	**2**	6
Silty Clay	**1**	6
Clay	**1**	7
NA	**NoData**	NoData

2.4 Geochemical Data

2.4.1 Organic Matter

Dixon and Uddameri [2016] also used soil organic matter percent (OM%) values as a layer in their vulnerability map. Organic matter presence is beneficial to bacteria; increasing OM% increases vulnerability because bacteria are more likely to survive in these conditions [*Corapcioglu and Haridas*, 1984]. The ranges selected by *Dixon and Uddameri* [2016] combined with the four quantiles from the data obtained were used to set the VI ranges (Table 6).

Table 6. List of organic matter content, and the vulnerability index values assigned.

Organic Matter (%)	Vulnerability Index
<1.25	1
1.25 – 2.99	2
3.00 – 3.99	3
4.00 – 54.5	4

2.4.2 pH

The pH conditions of the aquifer will affect bacterial mobility and survival, because their ability to grow in population size and survive is highly impacted by pH. The optimum range for EC growth is at a pH of 7 [*Madigan et al*., 2012], so a range of pH 6.9 – 7.1 was selected as the most vulnerable range (Table 7). Another range of pH for optimum survival was observed to be 6.0 – 8.3 based on high survival time and low die-off rates [*Cuthbert et al*., 1955; *Sjogren*, 1994]. Two ranges of 6.0 – 6.9 and 7.1 – 8.3 were selected for the second highest vulnerability. The lower range (0 – 6.0) was determined to be most hostile for bacteria, and the upper range (8.3 – 14.0) was determined to be moderate for bacteria. The higher extreme pH values were determined to be more vulnerable than the lower extreme because of more observed EC growth at high-neutral pH [*Rudolfs and Ragotzkie*, 1950; *Conner and Kotrola*, 1995].

Table 7. List of determined pH ranges and the given vulnerability indexes.

pH Range	Effect	Vulnerability Index
0 – 5.5	Dangerous for bacteria	1
5.5 – 6.9	Safe for bacteria	3
6.9 – 7.1	Optimal for bacteria	4
7.1 – 8.3	Safe for bacteria	3
8.3 – 14.0	Moderate for bacteria	2

2.5 Comparison to Microbial Detection Data

Once all the VIs were assigned and the layers were created, the VIs were normalized to give each factor equal weight. The model was compared to provincial EC detection data to determine which factor had a stronger correlation to EC presence in the subsurface. If a factor had a strong correlation, it would receive a higher weight value in the final model. The EC detection data used were provided by an Alberta Health Services program called Provincial Laboratory for Public Health (ProvLab). ProvLab receives groundwater samples from Alberta well owners for bacterial testing and monitoring purposes. The results are given back to the owners as either a detection of EC (1) or a non-detection (0). Total coliform (TC) results were provided as well, but were not used in this study. All of the samples were submitted by individuals, so there was no control over where and when the sample was taken, or how often the sampling occurred for one

location. The ProvLab dataset was used for the year 2012, and included EC testing on a total of 8610 samples, of which 155 tested positive for the presence of EC.

Detection and non-detection point data for the year 2012 were used for the comparison, initially in an aggregated format. By comparing each layer to detections of EC for the year 2012, a weight could be assigned based on the significance of the correlation (See Appendix C, Figure C4 and Figure C5 for aggregated maps of the detection data). However, the relationships were unexpectedly negative after having done an ordinary least squares test for aggregated EC detection compared to each layer's vulnerability values, so another comparison was done for point detection data. Kruskal-Wallis tests were used to compare the vulnerability values of EC detection points and non-detect points. Since the data are ordinal (i.e. a ranking), a non-parametric test was the most appropriate.

The value of vulnerability at each point of EC detection and non-detection was extracted in each separate factor layer using the Extract by Points tool in GIS. The outcome of the extraction was the vulnerability value at each point of detection, and each point of non-detection. The vulnerability values of the points were compared for each factor using a Kruskal-Wallis test to see if the factor had a higher vulnerability value at the points of detection. If there was a strong EC detection and high vulnerability correlation, the factor would be given a large weighting in the final map. Based on the results, further factor adjustment was not warranted. A similar Kruskal-Wallis test was then conducted for the overall vulnerability maps to observe whether the final result had accurate predictive properties for determining where bacterial contamination would occur.

3.0 Results and Discussion

An aquifer vulnerability map was created by determining which factors affect bacterial fate and transport, establishing the degree of the effect caused by each one, and combining them in an ArcGIS framework. Six factors were selected based on their influence on bacteria and data availability (Table 8). The layers produced for each factor had their own spatial distributions and were combined to produce two final vulnerability maps, one for the growing season (Figure 3) and one for the cold season (Figure 4).

3.1 Aquifer Vulnerability Maps

The results from the final vulnerability maps gave an indication of where shallow aquifers could be intrinsically vulnerable to bacterial contamination. The overall vulnerability maps had similar patterns for both the growing season (Figure 3) and the cold season (Figure 4). The foothills region of Alberta had the highest vulnerability, while there was lower vulnerability in the central and southern regions (Figure 1). The higher vulnerability in the foothills region is likely driven by the high vulnerability values of both soil moisture and soil organic matter in that region. The ranges of vulnerability for each map are slightly different, with the cold season having the lowest value of 1.49, and the growing season having the highest value of 5.26. The growing season had more vulnerable areas than the cold season, although the spatial distribution of high vulnerability regions was similar for both maps.

The growing season had an overall higher vulnerability than the cold season, which is driven by the higher vulnerability values of precipitation in the growing season (Figure 5b). The literature states that more disease outbreaks and higher densities EC detections occur in the summer [*Goss et al*., 1998; *Charron et al*., 2005; *Valeo et al*., 2016]. Severe weather conditions have been

noted as a catalyst for gastrointestinal disease outbreaks, and also occur more frequently in summer in temperate regions such as Alberta [*Curriero et al*. 2001; *Charron et al.*, 2005].

3.2 Notable Layers

The precipitation (P_v), soil moisture, and organic matter layers were important to focus on because they are unique layers that are not used in vulnerability maps for dissolved contaminants. Each of these three layers are naturally linked to each other, and gave some insight into what might be causing the patterns visible in the final vulnerability maps. The remaining layers are available in Appendix C (Figure C1 - C3).

3.2.1 Precipitation

Patterns for the precipitation layers were very different between the cold season (Figure 5a) and the growing season (Figure 5b). The cold season vulnerability was highest in the mountains and foothills area, whereas the growing season had the highest vulnerability in the Peace River region. It is important to note that the highest vulnerability in the cold season was equivalent to the lowest vulnerability index in the growing season. The data values in the cold season only covered the lower four vulnerability indexes, and the data values in the growing season only covered the higher four.

The growing season precipitation layer had an interesting spatial pattern (Figure 5b). A large portion of the Peace River region displayed a high vulnerability. The high values in this area indicate that there was a higher average precipitation event size, and a larger variance in the event sizes, meaning that there was potential for very large events. The convective weather that occurs during a period of the growing season from May to September is what causes these large events. The summer heat results in convective storms that primarily focus on a region that extends from Peace River to Rocky Mountain House [*Vickers et al*., 2001]. Summer storms have the potential to occur frequently, and can be large events, which would increase the vulnerability of this area. The potential for gastrointestinal disease outbreaks increases significantly after a rain event that exceeds the 80^{th} and 90^{th} percentile [*Curriero et al*., 2001], therefore the large precipitation events in the Peace River region of Alberta present concerns for the vulnerability of the aquifers there. This high vulnerability region does not align with the agricultural region and thus is not well represented in the final map, so it is important to make a separate note of this area as a potential region of higher vulnerability.

The cold season precipitation layer demonstrated a separate distribution of vulnerability, with a lower overall vulnerability than in the growing season (Figure 5a). Most of the precipitation was focussed in the mountain ranges, which is due to the increased snowfall in the mountains during the winter. Snow precipitation is not likely to increase the vulnerability of shallow aquifers because of the cold temperature, and the frozen ground decreasing infiltration. The impact of snow on EC can be lethal due to the damage caused by ice crystals during freezing [*Parker et al*., 2000]. However, if bacteria already exist in the groundwater, the snow could act as an insulator, protecting the bacteria from harsh winter weather. In the spring the snow could cause an increase in vulnerability because of the runoff and rapid infiltration during snowmelt. Regardless, the impact of snow on EC survival is harmful, especially in freeze-thaw chinook conditions that frequently occur in Alberta.

3.2.2 Soil Moisture

Soil moisture had similar distributions for both seasons, but there was more overall soil moisture in the growing season (Figure 6b). Both seasons had the same vulnerability range of 1 – 4 (before normalizing), and the highest soil moisture regions occurred in the mountains and foothills for both layers. The soil moisture vulnerability in the cold season is overall much lower than in the growing season, which is a result of the ground freezing in the winter (Figure 6a).

Agriculture and Agri-Food Canada [2005] determined that the dominant soil type which is in the high moisture area is Gray Luvisols (Appendix C Figure C6). These soils have a LFH horizon (Litter-Fermented-Humic) that responds quite rapidly to precipitation, increasing soil moisture [*Howitt and Pawluk*, 1985]. The observed pattern of soil moisture could therefore be a result of the soil type (and resulting soil properties) that exists in Alberta.

3.2.3 Organic Matter

The organic matter layer had a large portion with high organic matter percent in the north and foothills regions in the province (Figure 7). The central and south regions of the province had less organic matter, and therefore lower vulnerability.

Presence of organic matter within the soil is strongly related to the soil type, which is closely related to ecozones. Both the soil moisture and soil organic matter layers are influenced by climatic and environmental conditions that cause the different ecozones. According to a map produced by Agriculture and Agri-Food Canada, the main ecozones in Alberta are the Prairies, the Boreal Plains, the Taiga Plains, and the Montane Cordillera (Appendix C, Figure C7). Comparing the ecozones to the soil moisture and organic matter layers reveals that the Montane Cordillera and the Boreal Plains are the primary areas where there is high organic matter and high soil moisture. The Prairie ecozone corresponds reasonably well with the regions of lower organic matter (Figure 7). The conditions within each ecozone influence and are influenced by the amount of organic matter, soil moisture, and type of climate in the area. These factors are closely intertwined, and affect each other [*Schultz*, 2005].

3.3 Statistical Comparison to EC Detection

Extracted vulnerability values from each factor layer were compared to EC detection and non-detection point data using a Kruskal-Wallis statistical test (Table 9). The aim was to test the relationship between the vulnerability factor values and the distribution of groundwater EC detections and non-detections across the province in the year 2012. The same process was repeated for the final maps, comparing the final vulnerability values of the locations with EC detection and non-detection (Table 10).

The Kruskal-Wallis tests only showed significant results ($p < 0.05$) for the cold season soil moisture, growing season precipitation, and growing season overall. However, only the cold season soil moisture results had a significantly larger vulnerability for the detections ($\chi^2 = 9.4637$, df = 3, p-value = 0.02372). The growing season precipitation layer and overall vulnerability map for the growing season had significantly larger vulnerabilities for the non-detections (precipitation: $\chi^2 = 9.6345$, df = 3, p-value = 0.02194; final map: $\chi^2 = 57.3914$, df = 39, p-value = 0.02896). Other results did not display any statistically significant relationships.

The only significantly larger results for EC detections were for the cold season soil moisture layer, which suggests that the locations of detections of EC were accurately predicted by the soil

moisture vulnerability values. The other significant result was for the growing season precipitation layer, but the non-detection had the larger vulnerability, the opposite of what was expected. A larger vulnerability at points of non-detection suggests that large rainfall events decrease the intrinsic vulnerability. It is known that rainfall events increase the risk of gastrointestinal disease [*Curriero et al*., 2001], so it is unlikely that regions with more rainfall are less vulnerable.

A similar phenomenon of high vulnerability at non-detect points was also observed in the final map for the growing season. Regions with non-detects of EC had a significantly higher vulnerability value than the locations with detections. The other factors did not have statistically significant relationships, although cold season precipitation had a p-value of ~0.06, approaching significance for a larger vulnerability for EC detection. The stronger statistical relationships observed in the cold season present the idea that the cold season is more predictable, or perhaps there is a factor missing in the growing season that does not influence the cold season. However, considering the overwhelming lack of significant relationships, and unexpected negative relationships, there is likely a lack of reliability in detection measurements.

3.4 Sources of Error

The EC detection data did not correlate well with the map vulnerability, but this could be attributed to several things. Bacterial detection data are not inherently reliable due to the common lack of appropriately frequent sampling rates [*Batterman et al*., 2009]. The samples tested by ProvLab are primarily obtained from individuals who collect and send samples from their own private water supply well for monitoring purposes. Some individuals for the year 2012 took regular samples in different months, while others only took samples once that year. Other possible reasons for the poor correlation could be: the potential lack of a contaminant source at a sampling location; the distribution of samples, as most of the samples were collected in regions that happened to have a high vulnerability.

Most samples were collected from the foothills and north regions of the province, areas with high organic matter and soil moisture. Samples submitted for EC analysis in the ProvLab data were generally focussed in regions of high vulnerability, biasing the relationship between aquifer vulnerability and EC detections. A more even distribution of samples across the province might have made the comparison more representative, including detections and non-detections in regions of high and low vulnerability. There was also a large difference in the number of detections and non-detections; the number of non-detections was nearly two orders of magnitude larger than the number of detections. This caused a significant skew in the data and potentially affected the results. Finally, the actual presence of a source of microbial contamination would affect the correlation, and could be the main cause of the discrepancy. If there is no EC source near aquifers in regions with high vulnerability, then the aquifer will not become contaminated, and there will not be a EC detection. A new layer that represents potential sources of contamination could be created by acquiring data that shows the locations and magnitudes of various sources (e.g. manure, septic systems, wastewater discharge). However, the risk of contaminant presence was not considered in this study because the map is intended to represent intrinsic vulnerability.

The lack of spatially continuous data resulted in gaps in each layer that had to be filled in with interpolation, increasing the uncertainty of the results. Most of the data for layers constructed with point data were estimated through the kriging process. If this methodology was used in a smaller area with more thorough data, it could yield more accurate results. There were disclaimers from most sources that were used to collect data stating that their data would not be helpful in large

modeling studies. Most of the data were not measured directly, but estimated by experts using data that were more easily collectible. A more ideal situation would be having data that have been measured directly, more accurately representing the conditions of the subsurface. If better data are collected in the future, this map could be recreated using that information.

3.5 Future Studies

The goal of the statistical analysis was to determine the degree of influence of each factor and to assign a weight to each layer for the final map. The model was originally intended to be a weights of evidence (WoE) model, but the mixed statistical results made this impossible without arbitrarily assigning weights based on perceived impacts of the different factors. This led to the decision to not assign weights and normalize the vulnerability of each layer. One potential method of determining weights would be to compare the factor layers to gastrointestinal disease outbreak distributions. Outbreak data would provide a different perspective on the impact of EC contaminated water on humans and would be an interesting continuation for this research. If any future studies determine the degree of influence of the selected factors, those values could be applied to this model.

3.6 Model Testing

Aquifer vulnerability map validation can be a complicated and convoluted process, especially if the map covers a large area. Many factors contribute to the location and detection of contaminants, including time lags caused by the travel time of the contaminant, contamination source presence, and degree of field and lab work required to collect contaminant detection data [*Neukum et al*., 2008]. Validation usually occurs after the model has been calibrated. However, vulnerability maps are rarely, if ever, validated. In this study, a validation was attempted on the unweighted map, but this yielded uncertain and unexpected negative results (Table 10). Regardless, the attempt to validate the model improves the knowledge of how useful vulnerability maps can be. Literature exists that shows some successful model validations after some recalibration, but this is not done as often as it needs to be [*Leal and Castillo*, 2003; *Neukum et al*., 2008]. At the very least, the importance of incorporating risk data (i.e. contaminant source locations) was made apparent, and could help improve this map. The map on its own is useful for understanding the spatial distribution of influence of different important factors on bacteria. The verification of the overall map only indicates its predictive properties, and how useful it might be for decision-makers [*de Marsily et al*., 1992].

3.7 Is the Vulnerability Map Useful?

Ultimately, the map produced in this study could not be verified using EC detection data, and therefore could not be used with certainty for decision making purposes. However, this is the case with several existing vulnerability maps [*Eilers and Buckley*, 2002]. Verification of vulnerability maps is not often required as they are used as general guides by decision makers as an indication of relative vulnerability. These kinds of maps should not be considered hard evidence of true vulnerability, but rather a relative vulnerability of that region in comparison with surrounding regions. Aquifer vulnerability maps are intended to determine relative intrinsic vulnerability, and should be used as guides that are followed up with a site-specific investigation. The idea that vulnerability maps are not true representations of exact vulnerability also carries through with all kinds of models; as noted in *Box and Draper* [1987]: “Essentially, all models are wrong, but some are useful.” The map developed in this study is useful in the sense that it can be used to estimate relative intrinsic vulnerability of shallow aquifers to bacterial contamination in the Province of Alberta.

Table 8. List of aquifer vulnerability factors and the influence each one would have on bacterial fate and transport.

Category	Vulnerability Factor	Effect on Bacteria
Climatic	Precipitation	Mobilization (release from manure and infiltration)
	Soil Moisture	Survival and filtration effects
Geochemical	Organic Matter	Attachment and survival
	pH	Attachment and survival
Geologic	Hydraulic Resistance	Movement through the subsurface, survival, distance to travel in vadose zone
	Soil Texture	Specifically, clay/silt/sand effects on attachment and filtration

Table 9. Results of the Kruskal-Wallis tests comparing EC detection (+) and EC non-detection (-) point data using spatially corresponding vulnerability values for all factors (p = 0.05).

Factor	p-value	Larger Mean (+ or - EC Detection)	Larger Median (+ or - EC Detection)	Significant + or -
Growing Season Precipitation	0.022	-	-	Yes (-)
Cold Season Soil Moisture	0.024	+	Equal	Yes (+)
Cold Season Precipitation	0.062	+	+	No (+)
Soil Texture	0.117	-	Equal	No (-)
Growing Season Soil Moisture	0.189	-	-	No (-)
Hydraulic Resistance	0.476	+	Equal	No (+)
pH	0.489	+	Equal	No (+)
Organic Matter	0.552	-	Equal	No (-)

Table 10. Results of the Kruskal-Wallis tests comparing EC detection (+) and non-EC detection (-) point data using spatially corresponding vulnerability values for the final vulnerability maps (p = 0.05).

Season	p-value	Larger Mean (+ or - EC Detection)	Larger Median (+ or - EC Detection)	Significant + or -
Growing Season Overall	0.029	-	-	Yes (-)
Cold Season Overall	0.604	+	+	No (+)

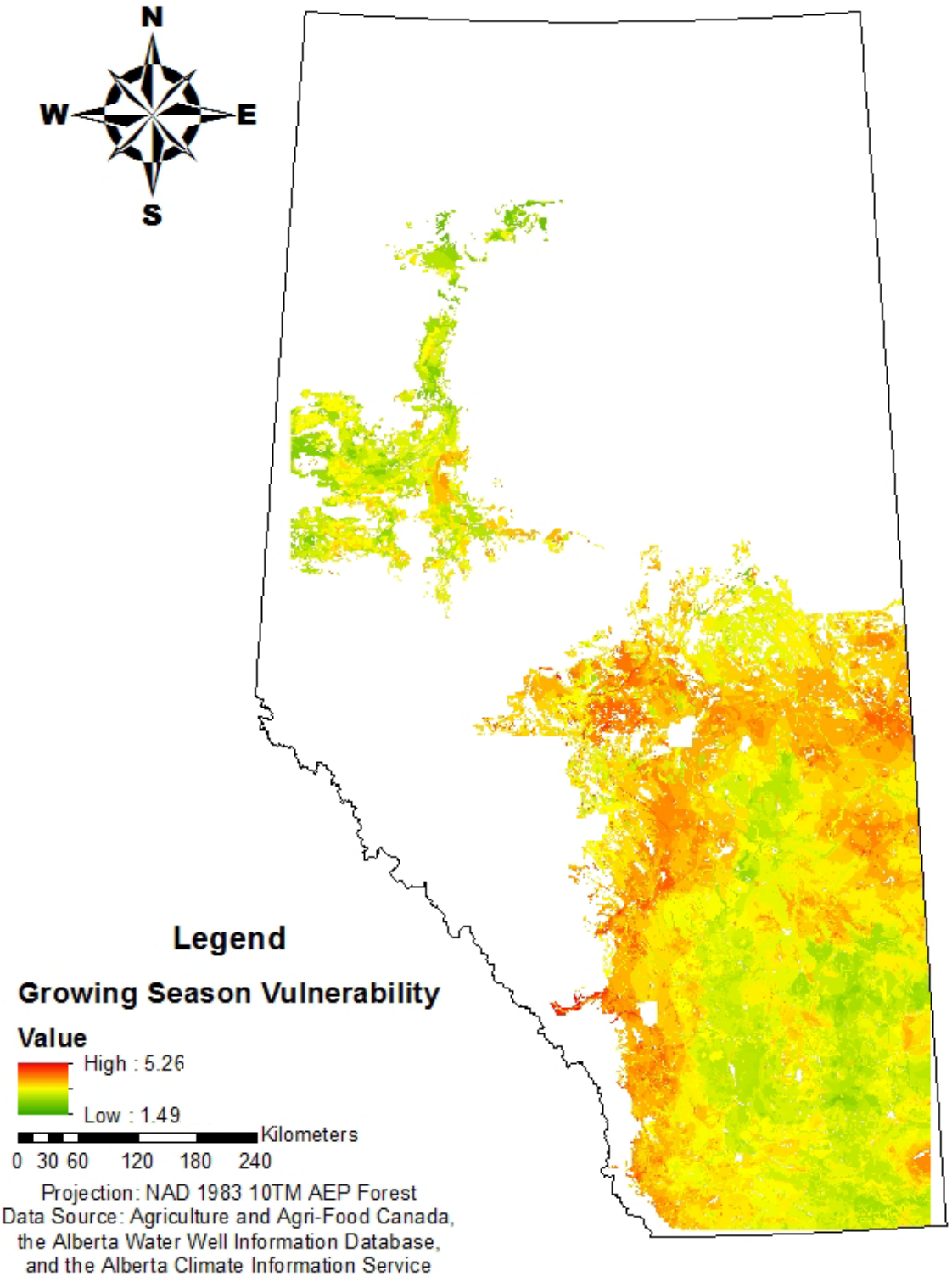

Figure 3. Growing season groundwater bacterial vulnerability map for Alberta, Canada in the year 2012.

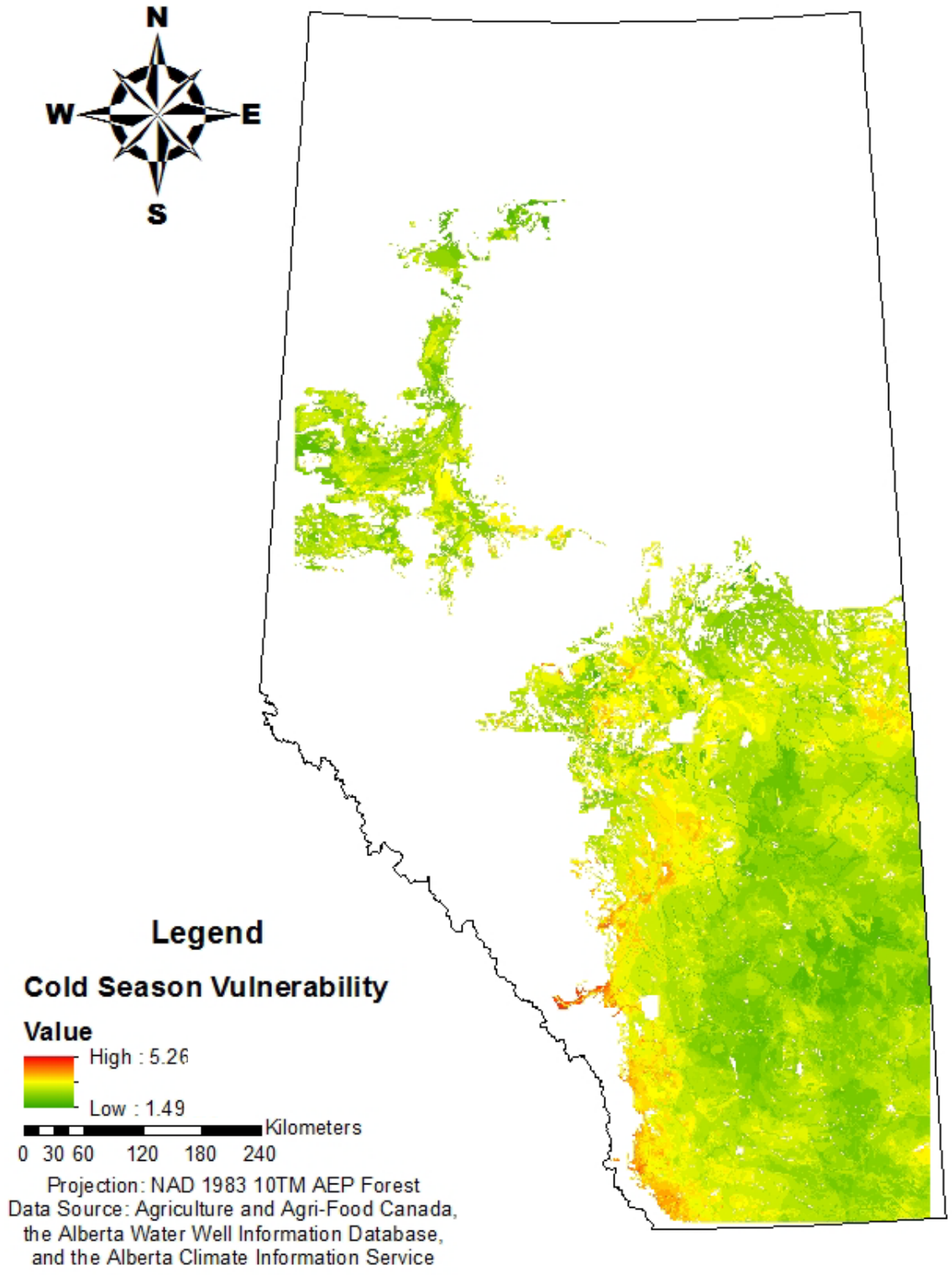

Figure 4. Cold season groundwater bacterial vulnerability map for Alberta, Canada in the year 2012.

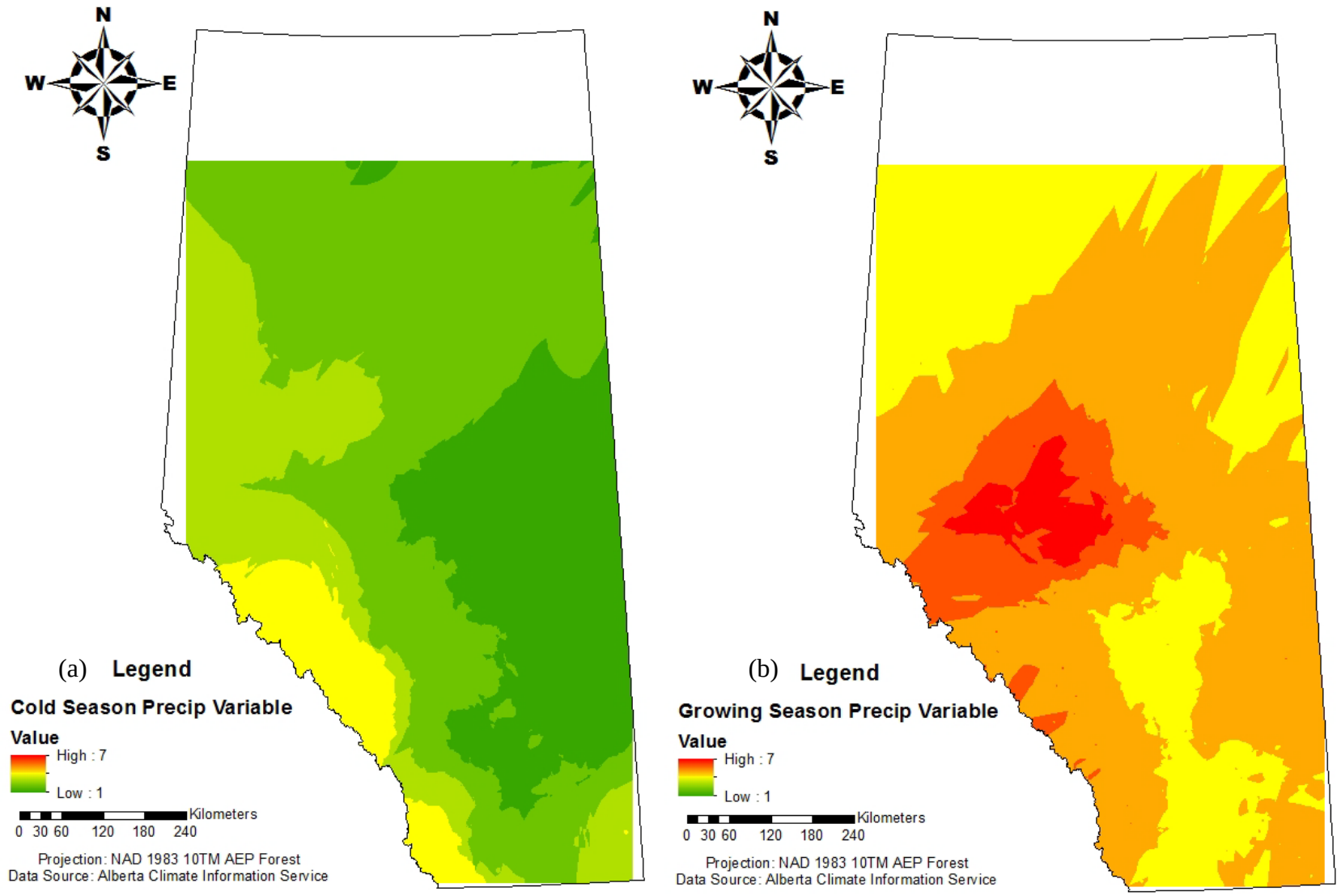

Figure 5. The classified and non-normalized layer for the precipitation variable P_v (mm^2/event) in the cold season (a) and the growing season (b) in Alberta, Canada in the year 2012.

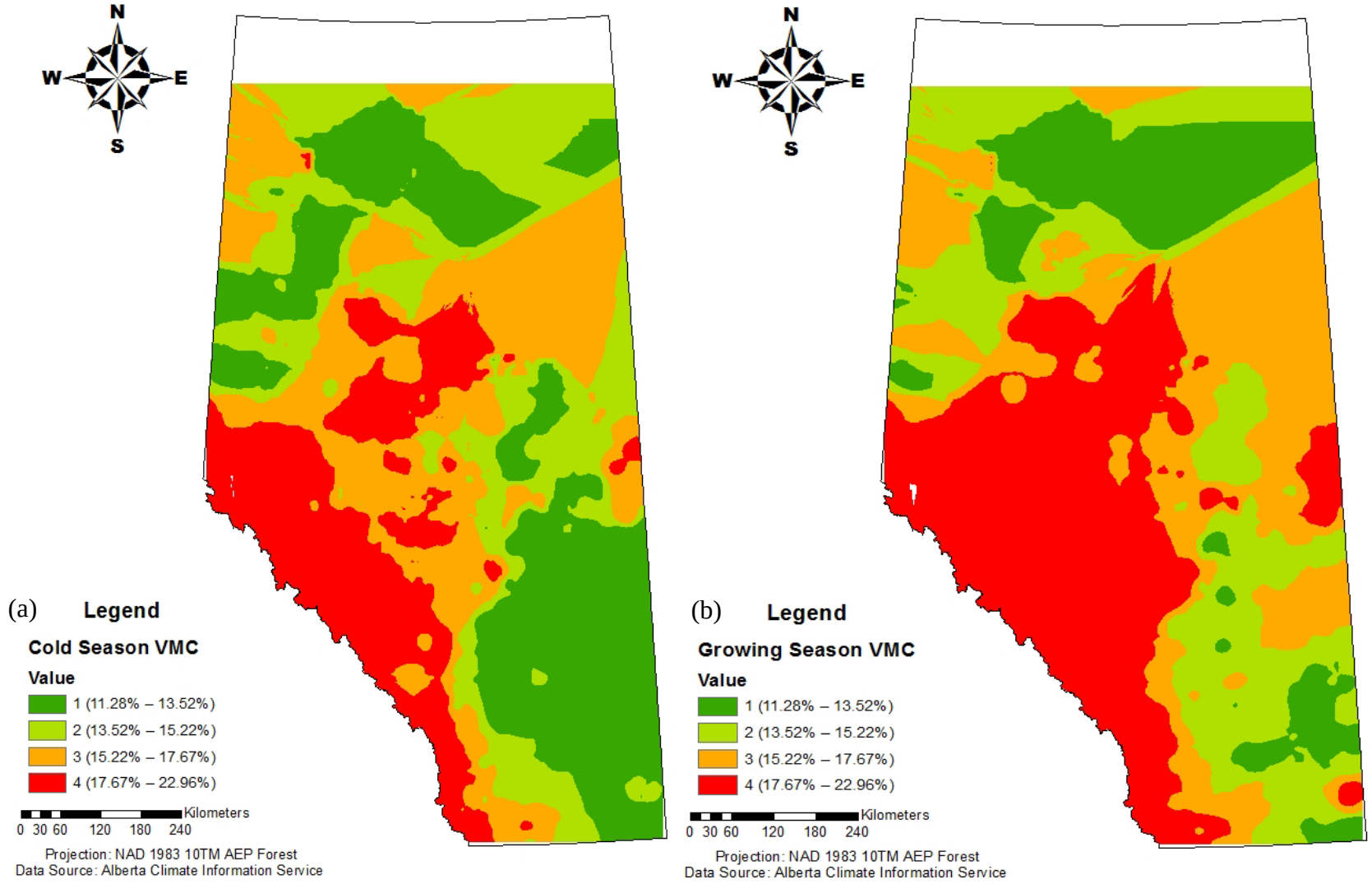

Figure 6. The classified and non-normalized layer for soil moisture as volumetric moisture content (VMC%) in the cold season (a) and the growing season (b) in Alberta, Canada in the year 2012.

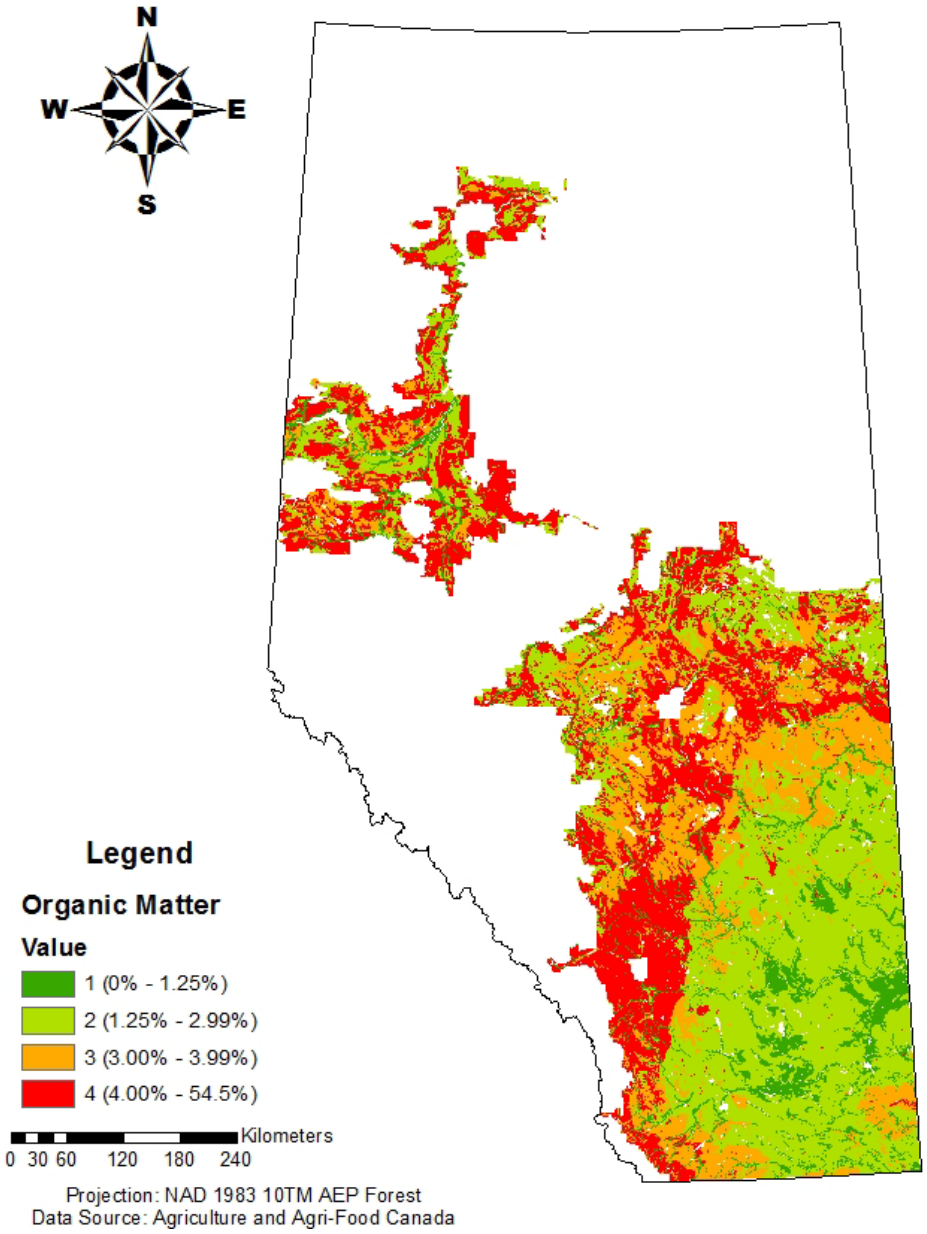

Figure 7. The classified and non-normalized layer for percent organic matter for both seasons in Alberta, Canada.

4.0 Conclusions

An intrinsic aquifer vulnerability map for bacterial contaminants was developed by including climatic, geochemical, and geologic characteristics as vulnerability factors. All of the factors were converted to GIS layers and combined to produce an intrinsic vulnerability map for the Province of Alberta for the year 2012.

Climatic temporal factors such as weather events affect bacterial survival and infiltration rates, and therefore affect bacterial presence in an aquifer. Climatic factors influence soil conditions such as soil moisture and organic matter, which also affect bacteria survival. These factors are not commonly used in vulnerability maps, but are important when considering the particular fate and transport properties of bacterial contaminants. The inclusion of these factors make this vulnerability map unique.

Although six factors that were thought to be the most relevant for bacterial vulnerability were successfully assembled in this study (including soil moisture, hydraulic resistance, pH, soil texture, organic matter, and precipitation), only two were found to have a significant relationship to bacterial detection. However, more tests need to be done that incorporate contaminant sources with the intrinsic vulnerability to further understand the relationship between vulnerability and bacterial detection.

There was a lack of statistically significant relationships when the model was tested, which could be due to the unreliable nature of bacterial data, or the presence/absence of a contaminant source. The statistical analysis of the final map indicated that the model does not have predictive properties, but it could be used to get a basic understanding of intrinsic characteristics and vulnerability.

5.0 Acknowledgements

Many thanks to Dr. Ed Cey and Dr. Cathy Ryan for their ongoing support and helpful knowledge. A special thank you to Jesse Invik for all his time in developing the *E. coli* detection maps, and a thank you to Dr. Tak Fung for his mathematical advice and data organization skills, and all his help in developing the precipitation variable.

Literature Cited

Adesiyun, B. Y. A. A., J. Adekeye, and J. U. Umoh (1983), Studies on well water and possible health risks in Katsia, Nigeria, *J. Hyg. (Lond).*, *90*(2), 199–205.

Agriculture and Agri-Food Canada (2005), *Soil Groups of Alberta*.

Agriculture and Agri-Food Canada (2013), National Ecological Framework, *Natl. Soil Database*, 1. Available from: http://sis.agr.gc.ca/cansis/nsdb/ecostrat/index.html (Accessed 14 March 2017)

Alberta Agriculture and Forestry (2017), Understanding Groundwater, Available from: http://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/wwg406 (Accessed 4 May 2017)

Aller, L., T. Bennett, J. H. Lehr, R. J. Petty, and G. Hackett (1987), DRASTIC: A Standardized Method for Evaluating Ground Water Pollution Potential Using Hydrogeologic Settings, *J. Geol. Soc. India*, *29*(1), 38–57.

Batterman, S. et al. (2009), Sustainable control of water-related infectious diseases: A review and proposal for interdisciplinary health-based systems research, *Environ. Health Perspect.*, *117*(7), 1023–1032, doi:10.1289/ehp.0800423.

Box, G. E. P., and N. R. Draper (1987), *Empirical Model-Building and Response Surfaces*, Wiley, New York.

Bradford, S. A., J. Simunek, and S. L. Walker (2006), Transport and straining of *E. coli* O157:H7 in saturated porous media, *Water Resour. Res.*, *42*(12), 1–12, doi:10.1029/2005WR004805.

Bradford, S. A., V. L. Morales, W. Zhang, R. W. Harvey, A. I. Packman, A. Mohanram, and C. Welty (2013), Transport and fate of microbial pathogens in agricultural settings, *Crit. Rev. Environ. Sci. Technol.*, *43*(8), 775–893, doi:10.1080/10643389.2012.710449.

Burton, G. A., D. Gunnison, and G. R. Lanzal (1987), Survival of Pathogenic Bacteria in Various Freshwater Sediments, *Appl. Environ. Microbiol.*, 633–638.

Butler, R. G., G. T. Orlob, and P. H. Mcgauhey (1954), Underground Movement of Bacterial and Chemical Pollutants, *Source J. (American Water Work. Assoc.*, *46*(2), 97–111.

Cey, E. E., D. L. Rudolph, and J. Passmore (2009), Influence of macroporosity on preferential solute and colloid transport in unsaturated field soils, *J. Contam. Hydrol.*, *107*(1–2), 45–57, doi:10.1016/j.jconhyd.2009.03.004.

Charron, D. F., T. Edge, M. D. Fleury, W. Galatianos, D. Gillis, R. Kent, A. R. Maarouf, C. Neudoerffer, C. J. Schuster, M. K. Thomas, J. Valcour, and D. Waltner-Toews (2005), *Links Between Climate, Water And Waterborne Illness, And Projected Impacts Of Climate Change*.

Clapp, R. B., and G. M. Hornberger (1978), Empirical equations for some soil hydraulic properties, *Water Resour. Res.*, *14*(4), 601–604, doi:10.1029/WR014i004p00601.

Conboy, M. J., and M. J. Goss (2000), Natural protection of groundwater against bacteria of fecal origin, *J. Contam. Hydrol.*, *43*(1), 1–24, doi:10.1016/S0169-7722(99)00100-X.

Conner, D., and J. Kotrola (1995), Growth and survival of Escherichia coli O157: H7 under acidic conditions., *Appl. Environ. Microbiol.*, *61*(1), 382–385.

Corapcioglu, M. Y., and A. Haridas (1984), Transport and Fate of Microorganisms in Porous Media: A Theoretical Investigation, *J. Hydrol.*, *72*, 149–169.

Crane, S. R., and J. A. Moore (1984), Bacterial pollution of groundwater: A review, *Water. Air. Soil Pollut.*, *22*(1), 67–83.

Cui, C., W. Zhou, and M. Geza (2016), GIS-based nitrogen removal model for assessing Florida's surficial aquifer vulnerability, *Environ. Earth Sci.*, *75*(6), 526, doi:10.1007/s12665-015-5213-x.

Curriero, F. C., J. A. Patz, J. B. Rose, and S. Lele (2001), The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948-1994, *Am. J. Public Health*, *91*(8), 1194–1199, doi:10.2105/AJPH.91.8.1194.

Cuthbert, W. A., J. J. Panes, and E. C. Hill (1955), Survival of Bacterium coli Type I and Streptococcus Faecalis in Soil, *J. Appl. Microbiol.*, *18*(3), 408–414.

de Marsily, G., P. Combes, and P. Goblet (1992), Comment on `Ground-water models cannot be validated', by L.F. Konikow & J.D. Bredehoeft, *Adv. Water Resour.*, *15*, 367–369.

DeNovio, N. M., J. E. Saiers, and J. N. Ryan (2004), Colloid Movement in Unsaturated Porous Media: Recent Advances and Future Directions, *Vadose Zo. J.*, *3*(2), 338–351, doi:10.2113/3.2.338.

Dixon, B., and V. Uddameri (2015), *GIS and Geocomputation for Water Resource Science and Engineering*.

Eilers, R. G., and K. E. Buckley (2002), *A Methodology for Evaluating Soils, Landscapes and Geology for Nutrient Management Planning in the Prairie Landscape*.

Elçi, A. (2012), *Advances in GIS-Based Approaches to Groundwater Vulnerability Assessment: Overview and Applications*.

Environment and Climate Change Canada (2013), Groundwater, *Gov. Canada*. Available from: https://www.ec.gc.ca/eau-water/default.asp?lang=En&n=300688DC-1 (Accessed 4 April 2016)

Focazio, M. J., T. E. Reilly, M. G. Rupert, and D. R. Helsel (1984), Assessing Ground-Water Vulnerability to Contamination: Providing Scientifically Defensible Information for Decision Makers, *U.S. Geol. Surv. Circ. 1224*. Available from: https://pubs.usgs.gov/circ/2002/circ1224/html/new.htm (Accessed 31 March 2017)

Freeze, A. R., and J. A. Cherry (1979), *Groundwater*, Prentice-Hall, Englewood Cliffs, NJ.

Ginn, T. R., B. D. Wood, K. E. Nelson, T. D. Scheibe, E. M. Murphy, and T. P. Clement (2002), Processes in microbial transport in the natural subsurface, *Adv. Water Resour.*, *25*(8–12), 1017–1042, doi:10.1016/S0309-1708(02)00046-5.

Goss, M. J., D. A. J. Barry, and D. L. Rudolph (1998), Contamination in Ontario farmstead domestic wells and its association with agriculture: 1. Results from drinking water wells, *J. Contam. Hydrol.*, *32*(3–4), 267–293, doi:10.1016/S0169-7722(98)00054-0.

Grolimund, D., and M. Borkovec (1999), Long term release kinetics of colloidal particles from natural porous media, *Environ. Sci. Technol.*, *33*(22), 4054–4060.

Guber, a. K., D. R. Shelton, and Y. a. Pachepsky (2004), Transport and Retention of Manure-Borne Coliforms in Soil, *Vadose Zo. J.*, *4*(3), 828–837, doi:10.2136/vzj2004.0097.

Guber, A. K., Y. a. Pachepsky, D. R. Shelton, and O. Yu (2009), Association of Fecal Coliforms With Soil Aggregates: Effect of Water Content and Bovine Manure Application, *Soil Sci.*, *174*(10), 543–548, doi:10.1097/SS.0b013e3181bccc85.

Hayashi, M., G. A. Mohammed, K. Harrer, and C. R. Farrow (2012), Performance Evaluation and Improvement of the Versatile Soil Moisture Budget (VSMB) Model Prepared for Alberta Agriculture and Rural Development, 45.

Health Canada (2017), Guidelines for Canadian Drinking Water Quality – Summary Table, *Water and Air Quality Bureau, Healthy Environments and Consumer Safety Branch*, Health Canada, Ottawa, Ontario.

Howitt, R. W., and and S. Pawluk (1985), The Genesis of a Gray Luvisol Within the Boreal Forest Region. II. Dynamic Pedology, *Can. J. Soil Sci.*, *65*, 9–19.

Hrudey, S. E., P. Payment, P. M. Huck, R. W. Gillham, and E. J. Hrudey (2003), A fatal waterborne disease epidemic in Walkerton, Ontario: Comparison with other waterborne outbreaks in the developed world, *Water Sci. Technol.*, *47*(3), 7–14.

Invik, J. (2015), Total Coliform and Escherichia coli Positivity in Rural Well Water in Alberta, Canada: Spatiotemporal Analysis and Risk Factor Assessment.

John, D. E., and J. B. Rose (2005), Critical Review Review of Factors Affecting Microbial Survival in Groundwater, *Environ. Sci. Technol.*, *39*(19), 7345–7356.

Keller, A. A., and M. Auset (2007), A review of visualization techniques of biocolloid transport processes at the pore scale under saturated and unsaturated conditions, *Adv. Water Resour.*, *30*(6–7), 1392–1407, doi:10.1016/j.advwatres.2006.05.013.

Leal, J. A. R., and R. R. Castillo (2003), Aquifer Vulnerability Mapping in the Turbio River Valley, Mexico: A Validation Study, *Geofísica Int.*, *42*(1), 141–156.

Lee, J. O., W. J. Cho, C. H. Kang, and K. S. Chun (2001), Swelling and Hydraulic Properties of Ca-Bentonite for the Buffer of a Waste Repository, *Technol. Manag. Radioact. Waste from Nucl. Power Plants Back End Nucl. Fuel Cycle Act.*, 1–10.

Madigan, M. T., D. P. Clark, J. M. Martinko, D. A. Stahl, K. Bender, and D. H. Buckley (2012), *Brock Biology of Microorganisms 13th Edition*.

Marshall, K. C. (1980), Adsorption of microorganisms to soils and sediments, in *Adsorption of Microorganisms to Surfaces*, pp. 317–329, Wiley, Toronto.

Marshall, K. C., R. Stout, and R. Mitchell (1971), Mechanism of the Initial Events in the Sorption of Marine Bacteria to Surfaces, *Microbiology*, *68*(3), 337–348, doi:10.1099/00221287-68-3-337.

Mellor, A. F. P., and E. E. Cey (2015), Using generalized additive mixed models to assess spatial, temporal, and hydrologic controls on bacteria and nitrate in a vulnerable agricultural aquifer, *J. Contam. Hydrol.*, *182*, 104–116, doi:10.1016/j.jconhyd.2015.08.010.

Moore, B. A. (1993), *Case Studies in Wellhead Protection Area Delineation and Monitoring*.

National Research Council (1993), *Ground Water Vulnerability Assessment: Predicting Relative Contamination Potential under Conditions of Uncertainty*.

Natural Resources Conservation Service (N.D.), Soil Texture Calculator | NRCS Soils, Nrcs.usda.gov. Available from: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcs142p2_054167 (Accessed 15 March 2017).

Neukum, C., H. Hötzl, T. Himmelsbach, H. Hötzl, and T. Himmelsbach (2008), Validation of vulnerability mapping methods by field investigations and numerical modelling, *Hydrogeol. J.*, *16*, 641–658, doi:10.1007/s10040-007-0249-y.

Pandey, P. K., P. H. Kass, M. L. Soupir, S. Biswas, and V. P. Singh (2014), Contamination of water resources by pathogenic bacteria., *AMB Express*, *4*(1), 16, doi:10.1186/s13568-014-0051-x.

Parker, L. V, M. L. Yushak, C. J. Martel, and C. M. Reynolds (2000), *Bacterial Survival in Snow Made from Wastewater*.

Rehm, B. W., G. H. Groenewold, and K. A. Morin (1980), Hydraulic Properties of Coal and Related Materials, Northern Great Plains, *Ground Water*, *18*(6), 551–561.

Rudolfs, W., L. L. Falk, and R. A. Ragotzkie (1950), Literature Review on the Occurrence and Survival of Enteric, Pathogenic, and Relative Organisms in Soil, Water, Sewage, and Sludges, and on Vegetation: I. Bacterial and Virus Diseases, *Water Environ. Fed.*, *22*(10), 1261–1281, doi:10.1016/S0262-1762(99)80122-9.

Safadoust, A., A. A. Mahboubi, B. Gharabaghi, M. R. Mosaddeghi, P. Voroney, A. Unc, and G. Sayyad (2011), Bacterial filtration rates in repacked and weathered soil columns, *Geoderma*, *167–168*, 204–213, doi:10.1016/j.geoderma.2011.08.014.

Sällfors, G., and A. L. Öberg-Högsta (2002), Determination of hydraulic conductivity of sand-bentonite mixtures for engineering purposes, *Geotech. Geol. Eng.*, *20*(1), 65–80, doi:10.1023/A:1013857823676.

Schijven, J. F., J. H. C. Mülschlegel, S. M. Hassanizadeh, P. E. M. Teunis, and A. M. de Roda Husman (2006), Determination of protection zones for Dutch groundwater wells against virus contamination - Uncertainty and sensitivity analysis, *J. Water Health*, *4*(3), 297–312, doi:10.2166/wh.2006.012.

Schultz, J. (2005), *The Ecozones of the World*, Springer Berlin Heidelberg.

Simpson, M. W. M., D. M. Allen, and M. M. Journeay (2014), Assessing risk to groundwater quality using an integrated risk framework, *Environ. Earth Sci.*, *71*(11), 4939–4956, doi:10.1007/s12665-013-2886-x.

Sjogren, R. E. (1994), Prolonged Survival of an Environmental Escherchia Coli in Laboratory Soil Microcosms, *Water. Air. Soil Pollut.*, *75*(3), 389–403.

Sobsey, M. D., C. H. Dean, M. E. Knuckles, and R. A. Wagner (1980), Interactions and survival of enteric viruses in soil materials, *Appl. Environ. Microbiol.*, *40*(1), 92–101.

Tufenkji, N. (2007), Modeling microbial transport in porous media: Traditional approaches and recent developments, *Adv. Water Resour.*, *30*(6–7), 1455–1469, doi:10.1016/j.advwatres.2006.05.014.

Unc, A., and M. J. Goss (2004), Transport of bacteria from manure and protection of water resources, *Appl. Soil Ecol.*, *25*(1), 1–18, doi:10.1016/j.apsoil.2003.08.007.

van der Kamp, G., and G. Grove (2001), Well water quality in Canada: An overview, in *An Earth Odyssey. Proceedings of the 54th Canadian Geotechnical Conference*, pp. 45–49, Calgary.

Van Stempvoort, D., L. Ewert, and L. Wassenaar (1993), Aquifer Vulnerability Index: a Gis - Compatible Method for Groundwater Vulnerability Mapping, *Can. Water Resour. J.*, *18*(1), 25–37, doi:10.4296/cwrj1801025.

Valeo, C., S. Checkley, J. He, and N. Neumann (2016), Rainfall and Microbial Contamination in Alberta Well Water, University of Victoria.

Vickers, G., S. Buzza, D. Schmidt, and J. Mullock (2001), Chapter 3: Weather Patterns of the Prairies, in *The Weather of the Canadian Prairies Graphic Area Forecast 32*, pp. 47–67, NAV CANADA and Meteorological Service of Canada, Ottawa.

Wade, T. J., S. K. Sandhu, D. Levy, S. Lee, M. W. LeChevallier, L. Katz, and J. M. Colford (2004), Did a Severe Flood in the Midwest Cause an Increase in the Incidence of Gastrointestinal Symptoms?, *Am. J. Epidemiol.*, *159*(4), 398–405, doi:10.1093/aje/kwh050.

Wang, L., K. R. Mankin, and G. L. Marchin (2004), Survival of Fecal Bacteria in Dairy Cow Manure, *Am. Soc. Agric. Eng.*, *47*(4), 1239–1246.

Appendix A – Data Values

For all data used for the creation of GIS layers, please refer to the additional document titled “Data for Assessing and Mapping Groundwater Vulnerability to Bacteria in Alberta.xlsx”.

Sheet names in order: Soil Moisture, pH, Soil Texture, Organic Matter, Hydraulic Resistance, and Precipitation.

Descriptions are as follows:

Sheet A1. Soil moisture data including spatial coordinates and the seasonal averages of volumetric moisture content percentage (VMC%).

Sheet A2. pH data including object ID of the polygon in the Agriculture and Agri-Food Canada shapefile and the corresponding pH assigned to that polygon.

Sheet A3. Soil texture classification data including object ID of the polygon in the Agriculture and Agri-Food Canada shapefile and the corresponding texture type assigned to that polygon.

Sheet A4. Organic matter percent data including object ID of the polygon in the Agriculture and Agri-Food Canada shapefile and the corresponding percentage of organic matter assigned to that polygon.

Sheet A5. Hydraulic resistance data including spatial coordinates and the log transformation of the resistance values.

Sheet A6. Precipitation data including the weather station name, the spatial coordinates, the average precipitation per event (P_e), the variance in the event size (V_e), and the calculated precipitation variable (P_v).

Appendix B – Extended Methodology

1. ArcGIS Methods

Soil texture data required additional management before it could be used in ArcGIS. The attribute tables needed to be joined with the geochemical and locational data based on the "POLY_ID" attribute. First, the SLT shapefile must be joined to the CMP shapefile using the "Soil_ID" attribute, then the joined CMP value can be joined to the Alberta Soils shapefile using the "POLY_ID" attribute. This shapefile contained data for clay (TCLAY), silt (TSILT), and sand (TSAND) percentages, pH (PH2), and organic matter (ORGCARB). The shapefile was converted to a raster format using the Polygon to Raster tool (Conversion Tools → To Raster). The desired feature was selected and converted to a new raster layer for that feature (e.g. ORGCARB for organic matter). For the soil texture raster, the percentages of clay, silt, and sand were extracted to Excel and used to determine the soil type with a "Soil Type Calculator" spreadsheet found online [*Natural Resources Conservation Service*, N.D.]. However, this spreadsheet only calculated for one manual entry at a time, so the formulas used in the tool were moved to a different spreadsheet that would allow for mass calculation. Example of the formula for sand: "=IF(AND(I4>=7,I4<27,G4>=28,G4<50,F4<=52),"Loam","")", where 'F' refers to sand, 'G' refers to silt, and 'I' refers to clay. This formula indicates that if the three percentages meet these requirements (e.g. sand is < or = 52%), then that data point receives the "Loam" classification, otherwise the cell is left blank. This was repeated with the formulas for each soil classification in a new column, and the results were then combined into a new column using the "CONCATENATE" function. These data were then imported into ArcGIS and joined to the Alberta Soils shapefile using the "POLY_ID" attribute.

Table B1. Formulas used to determine soil classification based on clay, silt, and sand percentages. From https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcs142p2_054167.

Excel Formula	Soil Classification
((silt + 1.5*clay) < 15)	Sand
((silt + 1.5*clay >= 15) && (silt + 2*clay < 30))	Loamy Sand
((clay >= 7 && clay < 20) && (sand > 52) && ((silt + 2*clay) >= 30) || (clay < 7 && silt < 50 && (silt+2*clay)>=30))	Sandy Loam
((clay >= 7 && clay < 27) && (silt >= 28 && silt < 50) && (sand <= 52))	Loam
((silt >= 50 && (clay >= 12 && clay < 27)) || ((silt >= 50 && silt < 80) && clay < 12))	Silt Loam
(silt >= 80 && clay < 12)	Silt
((clay >= 20 && clay < 35) && (silt < 28) && (sand > 45))	Sandy Clay Loam
((clay >= 27 && clay < 40) && (sand > 20 && sand <= 45))	Clay Loam
((clay >= 27 && clay < 40) && (sand <= 20))	Silty Clay Loam
(clay >= 35 && sand > 45)	Sandy Clay
(clay >= 40 && silt >= 40)	Silty Clay
(clay >= 40 && sand <= 45 && silt < 40)	Clay

Normalization was done by dividing each layer by the highest vulnerability value in that layer. For example, the pH layer had a maximum vulnerability of 4, the entire layer was divided by 4 using the raster calculator. Formula: ("ProSoilpHRas" / 4.0000000000). The additional zeros were necessary to prevent rounding.

The normalized layers were added together to create the final map. Final map raster calculator formula for the growing season: ("NormGrowSM" * 10.0000000000) + ("NormpH") + ("NormOM" * 6.0000000000) + ("NormLogR" * 2.0000000000) + ("NormGrowP" * 4.0000000000) + ("NormSTex" * 2.0000000000).

2. Hydraulic Resistance Extended Methods

The well report data was received in Microsoft Access format. To get the desired data, the necessary values (e.g. core material, depth, location) were related to each other based on well ID and well report ID. The data were organized in terms of descending well ID and descending depths (recorded in ft, converted to m). The depths and the material attached to each depth (for estimating hydraulic conductivity) was used to calculate hydraulic resistance (s). Some wells needed to be removed because of a lack of information.

Shallow aquifers were focussed on for the purposes of simplifying the map. The water table depth was limited to 30m, and only new wells were focussed on. The hopes were that this would mean that the static water level would be indicative of water table depth.

Water levels were necessary to determine within each well for the purposes of calculating the hydraulic resistance above the water table. If one well reported two water depths, the well report was examined to find the listed water depth. If the well report could not be found in the AWWID, the first listed depth was removed. In the cases where the well had screen depth information, the deeper value between the static water level and the top of the screen was taken. If screen depth data were not available, it was assumed that the static water level represented the top of the aquifer.

Layers that had >5m of "Unknown", "Unreadable", "Blind", "See comments", "Predrilled", "Old Well", or "Lost Circulation" were removed from the dataset. If listed static water level is >30m, well is removed. Removed wells with static water level at zero (or left blank) because of the possibility that the water level was not recorded in those cases. If a layer that did not have a hydraulic conductivity assigned to it (such as "See Comments" or "Unknown") was above the water level, that well was removed from the dataset.

The material identifiers for several layers were convoluted (e.g. Sand and Boulders, Sand and Clay, etc.). They were sorted into groups, in this example, Sand + (Sand and Boulders) and Sand – (Sand and Clay). Some generalizations for hydraulic conductivity selections were made (e.g. cap rock was given the hydraulic conductivity of unfractured igneous rock). Fill was generalized to be loam. If a layer was named "Formation", the well log was examined for further information, if none was available, the material was assumed to be sandstone.

There were large ranges of hydraulic conductivities for generic labels of soil/rock types, so the midpoint value of the range was used. Bentonite was labelled as a layer in some wells. It is a fine-grained clay that is prone to swelling [*Sällfors and Öberg-Högsta*, 2002], and was assigned the conductivity value that was found by *Lee et al.*, [2001]. Coal and bedrock conductivities were taken from *Rehm et al*. [1980], this paper also notes that coal's conductivity will only be caused by fractures. Coal's hydraulic conductivity ranges from $10^{-5.02}$ to $10^{-5.89}$ (m/s) with a midpoint at 5.419×10^{-6} m/s.

3. Additional Information

It is important to note that the average soil moisture value that was used to determine the vulnerability does not pick up the change in soil moisture from spring to summer, where evapotranspiration and evaporation would decrease the soil moisture in the summer from the wet snow-melt springs [*Howitt and Pawluk*, 1985].

This method is very similar to DRASTIC and is limited to the accuracy of the data and the oversights that were necessary to take to complete the simulation (E.g. generalized polygons and missing data from AAFC, interpolated values).

4. Missing Information

- The distribution of manure application and location of irrigation zones were not used in this study because they are indicators of risk, and not intrinsic vulnerability.
- Solar radiation is considered dangerous for bacteria, but was not considered to be a die-off factor because most of the processes would be happening underground.

Appendix C – Additional Figures

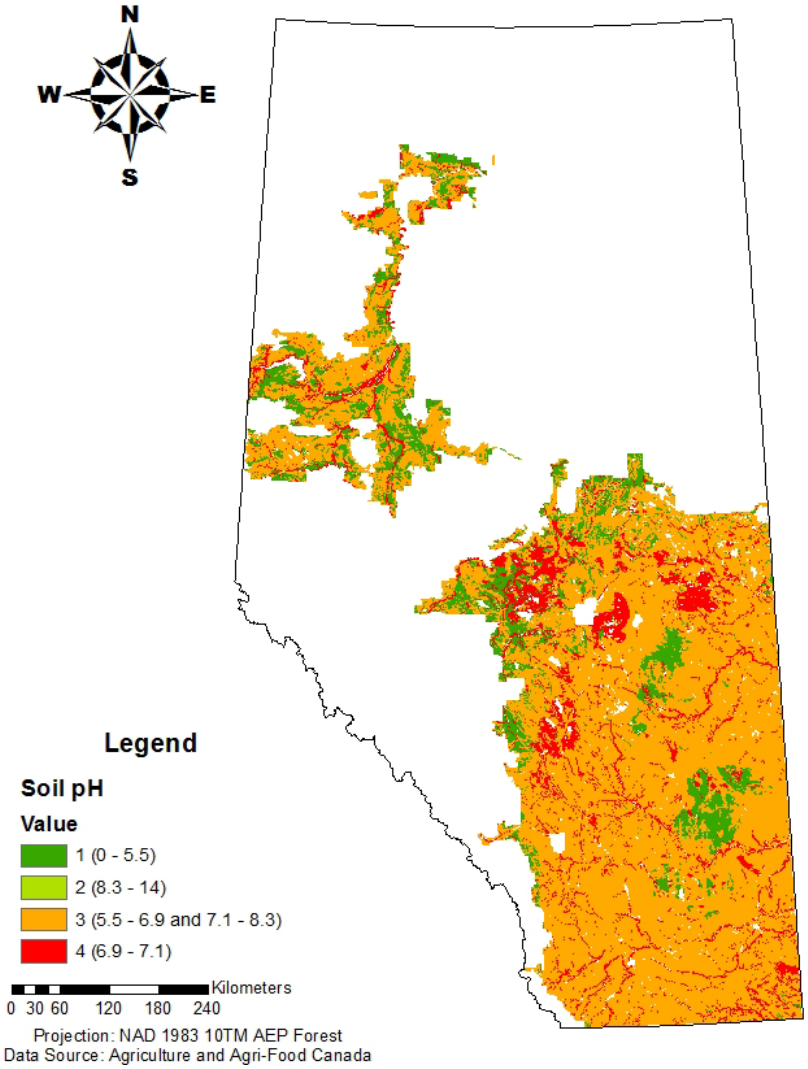

Figure C1. The classified and non-normalized layer for pH for both seasons in Alberta, Canada.

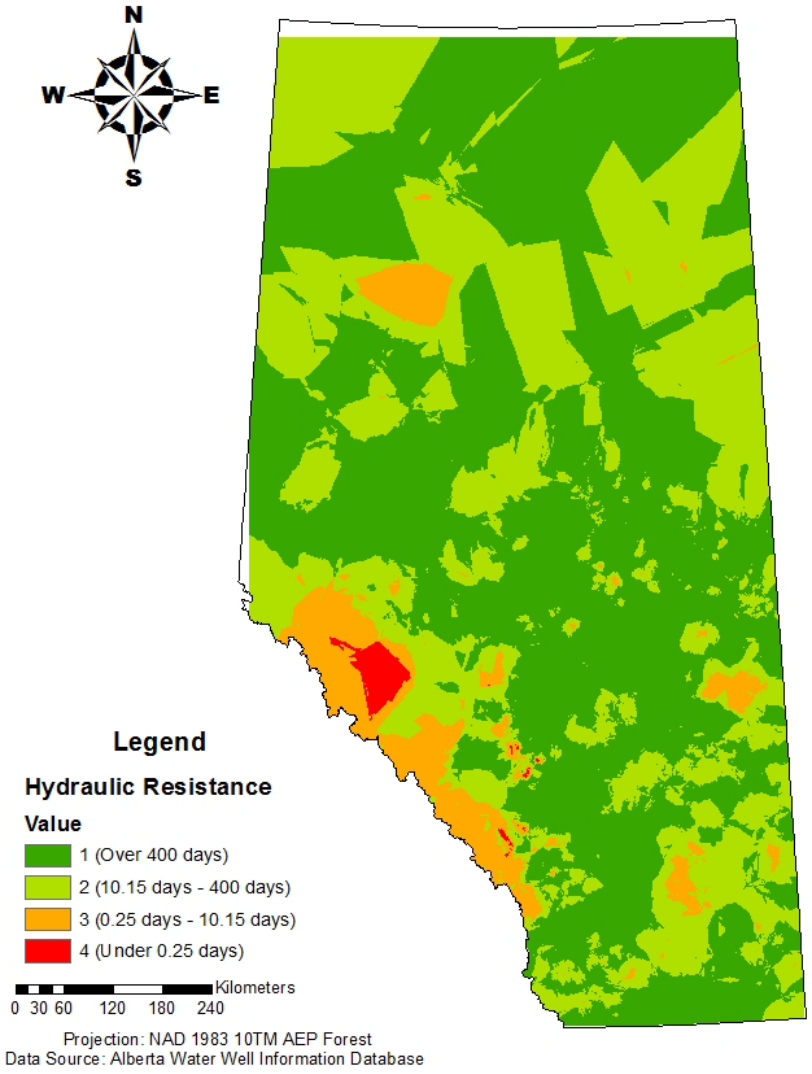

Figure C2. The classified and non-normalized layer for hydraulic resistance (in log(m/s)) for both seasons in Alberta, Canada.

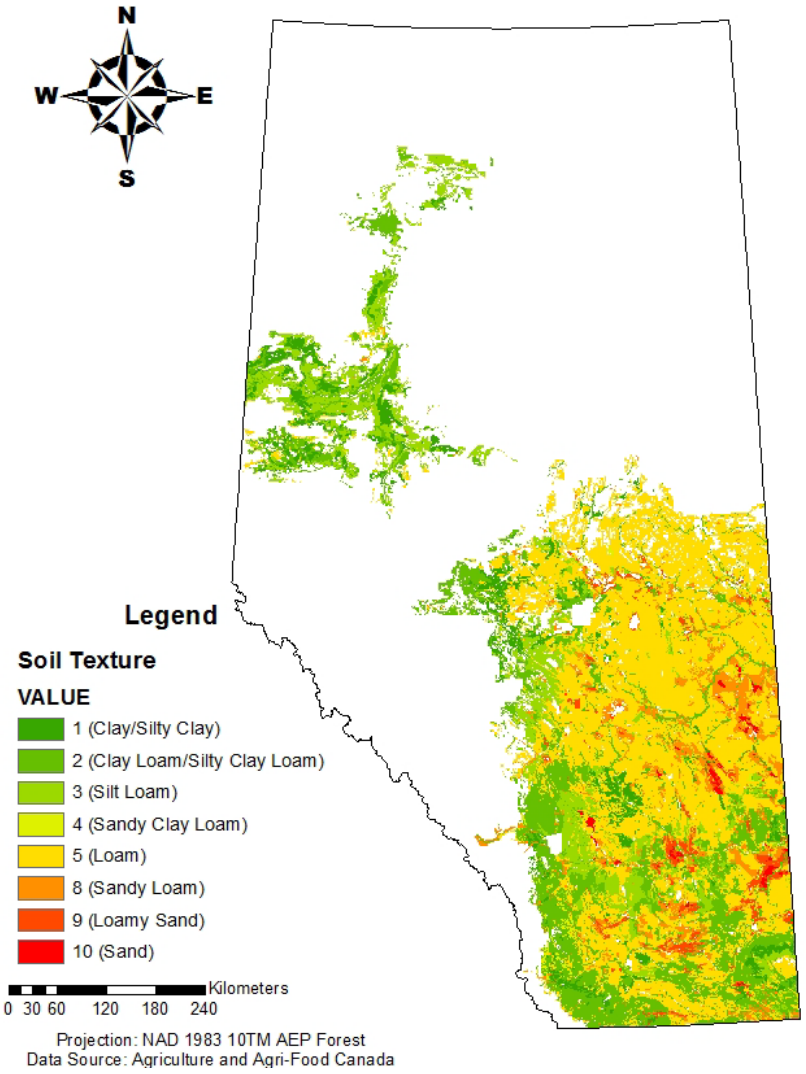

Figure C3. The classified and non-normalized layer for soil texture for both seasons in Alberta, Canada.

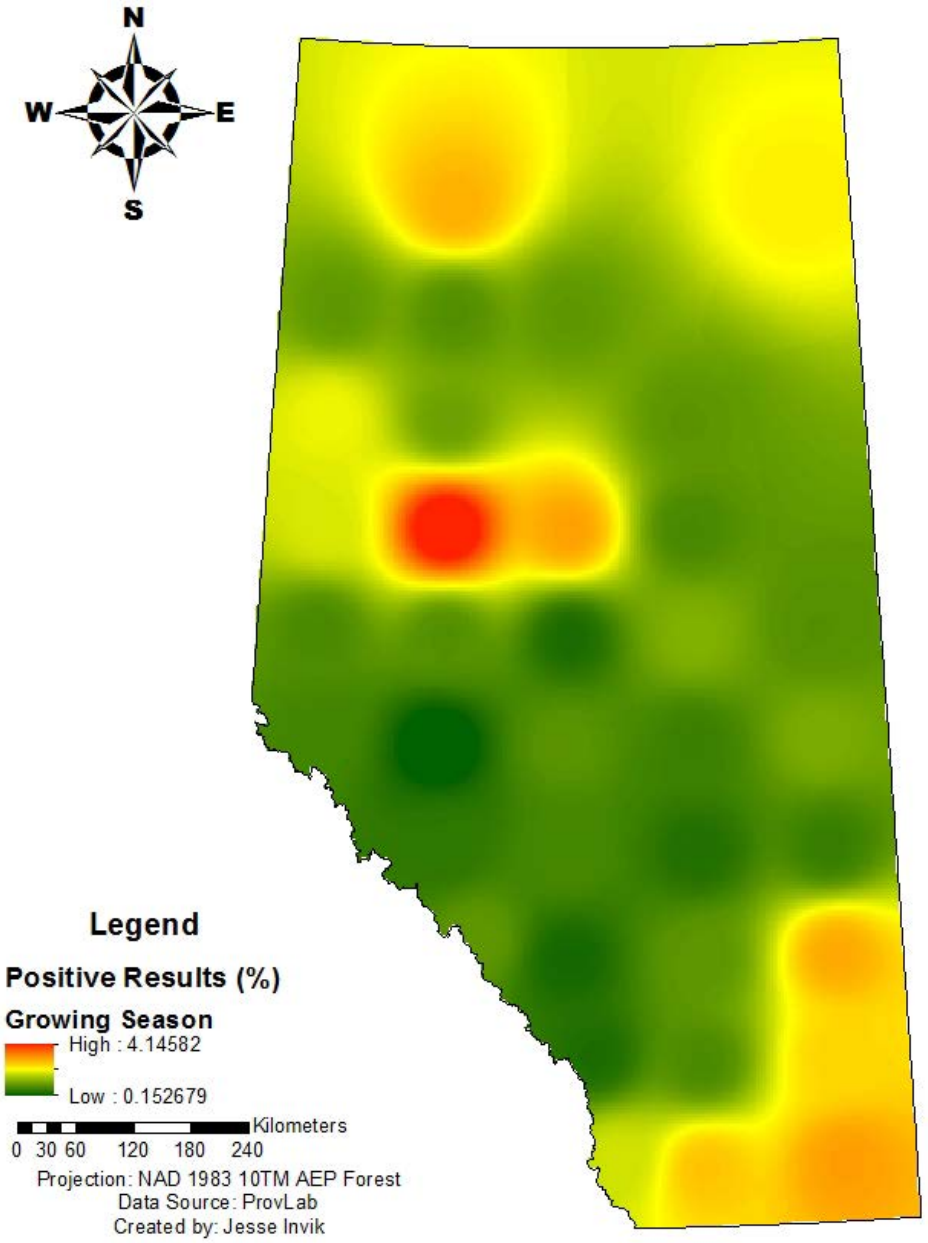

Figure C4. Relative risk map for Alberta for total coliform and EC in the growing season of 2012. Produced by Jesse Invik with similar methods outlined in *Invik* [2015].

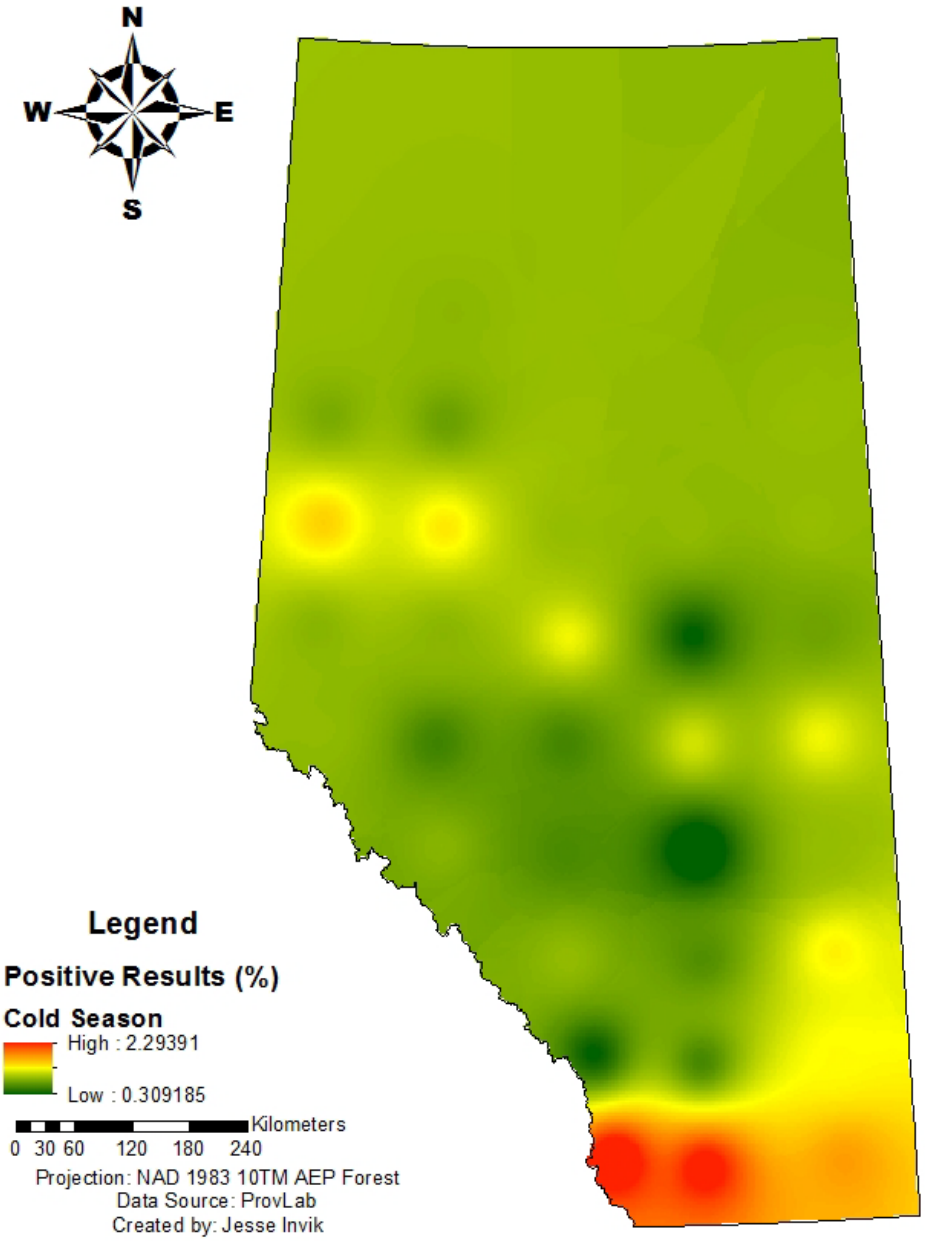

Figure C5. Relative risk map for Alberta for total coliform and EC in the cold season of 2012. Produced by Jesse Invik with similar methods outlined in *Invik* [2015].

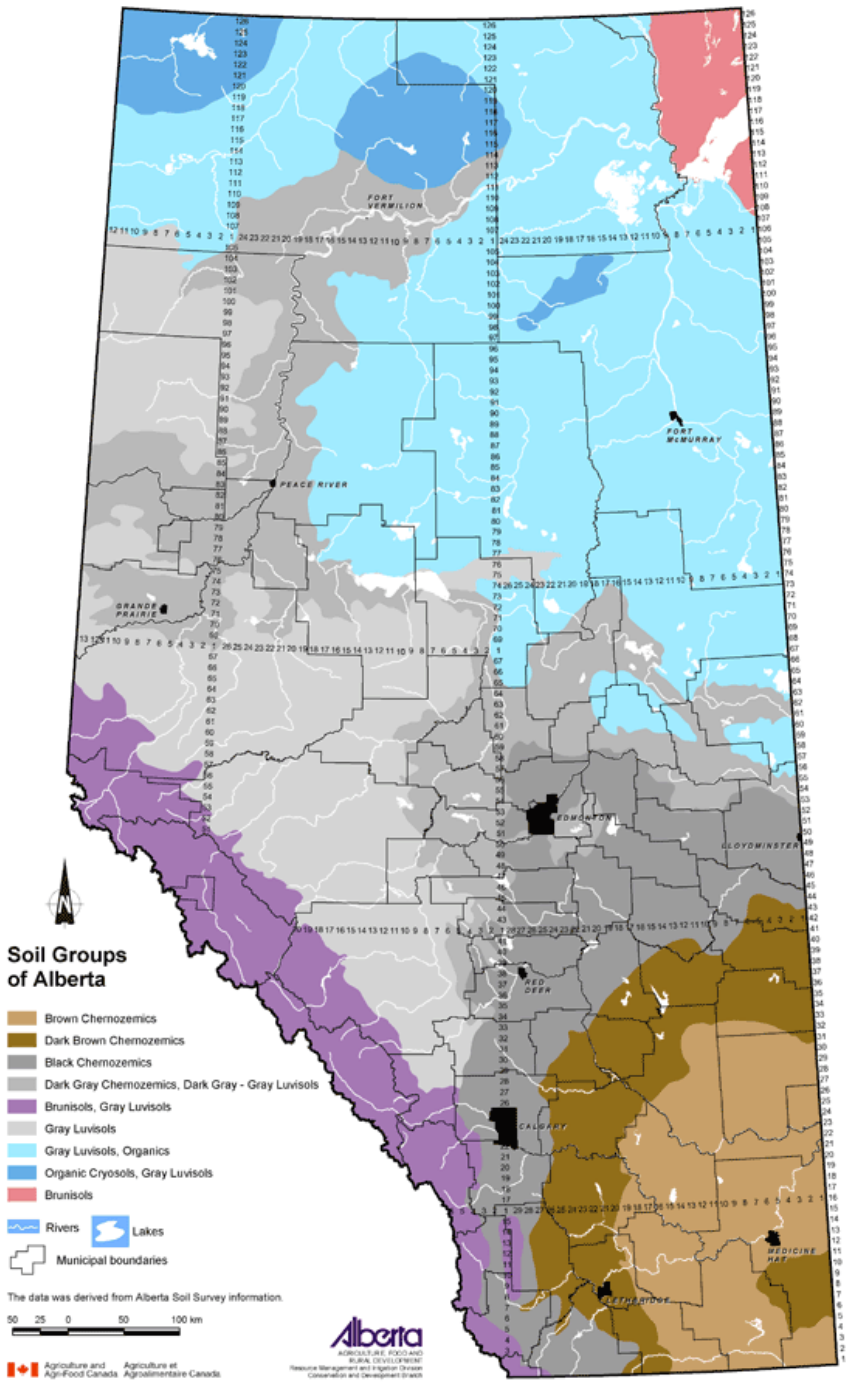

Figure C6. Dominant soil types found within the Province of Alberta. Obtained from *Agriculture and Agri-Food Canada* [2005].

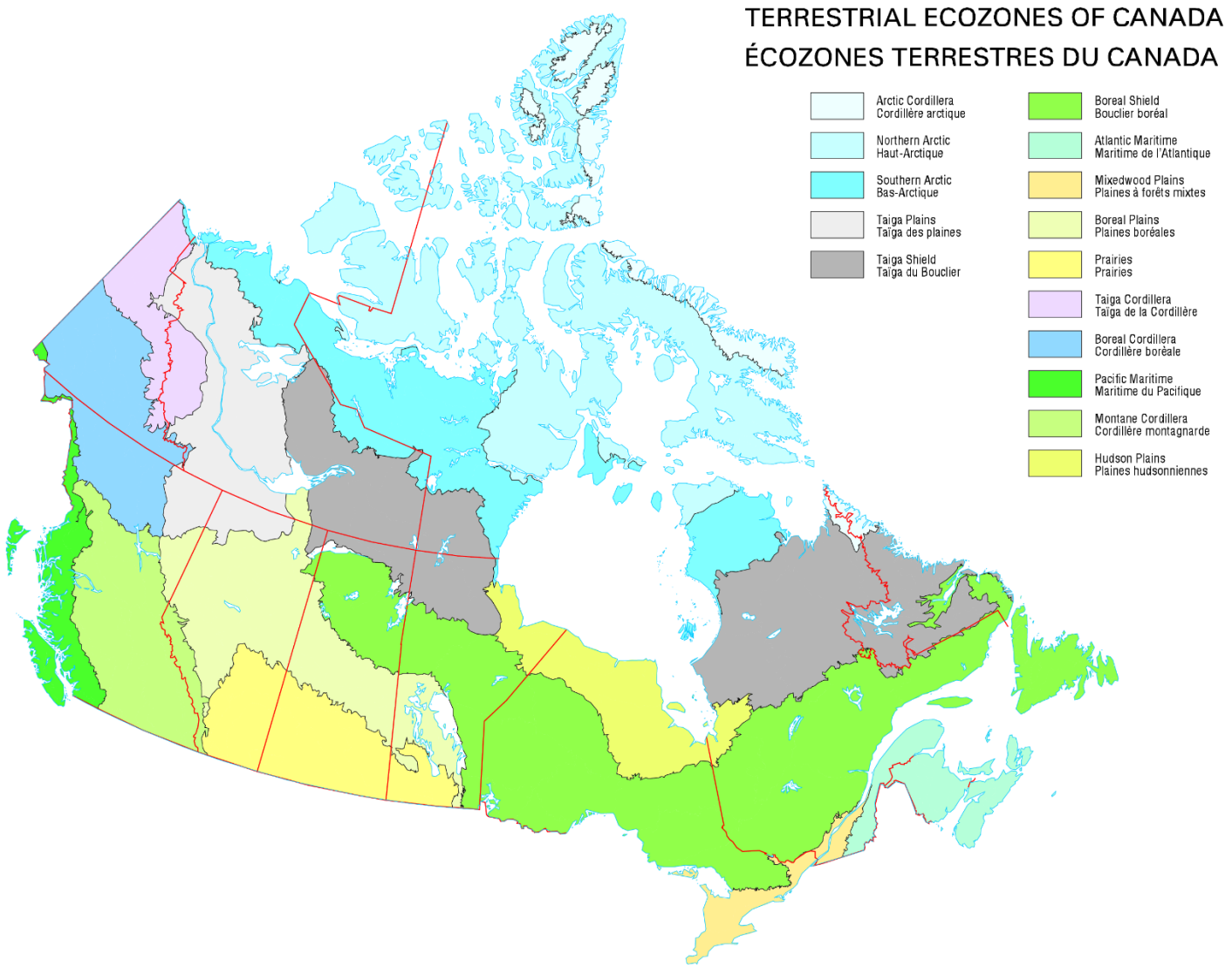

Figure C7. Image of the terrestrial ecozones of Canada, obtained from the National Ecological Framework as a part of *Agriculture and Agri-Food Canada* [2013]. Link: http://sis.agr.gc.ca/cansis/nsdb/ecostrat/index.html.