ASSESSING DATA ACCESSIBILITY AND QUALITY: PREDICTING THE EFFECT OF IMPERVIOUS SURFACES ON WATER QUALITY IN URBAN CENTRES ON THE WEST COAST OF LAKE SUPERIOR

Nicole Anderson
Lakehead University
April 2020

ASSESSING DATA ACCESSIBILITY AND QUALITY: PREDICTING THE EFFECT OF IMPERVIOUS SURFACES ON WATER QUALITY IN URBAN CENTRES ON THE WEST COAST OF LAKE SUPERIOR

by

Nicole Anderson

An Undergraduate Thesis Submitted in
Partial Fulfillment of the Requirements for the
Degree of Honours Bachelor of Environmental Management

Faculty of Natural Resources Management

Lakehead University

April 2020

Alex Bilyk	Ryan Wilkie
Major Advisor	Second Reader

LIBRARY RIGHTS STATEMENT

In presenting this thesis in partial fulfillment of the requirements for the HBEM degree at Lakehead University in Thunder Bay, I agree that the University will make it freely available for inspection.

This thesis is made available by my authority solely for the purpose of private study and research and may not be copied or reproduced in whole or in part (except as permitted by the Copyright Laws) without my written authority.

Signature: _____

Date: ___April 16, 2020____

A CAUTION TO THE READER

This HBEM thesis has been through a semi-formal process of review and comment by at least two faculty members. It is made available for loan by the Faculty of Natural Resources Management for the purpose of advancing the practice of professional and scientific environmental management.

The reader should be aware that opinions and conclusions expressed in this document are those of the student and do not necessarily the opinions of the thesis supervisor, the faculty or Lakehead University.

CONTENTS

ABSTRACT

Anderson N.N. 2020. Assessing Data Accessibility and Quality: Predicting the Effect Of Impervious Surfaces On Water Quality In Urban Centres On The West Coast Of Lake Superior. 103pp.

Key Words: data accessibility, data quality, Duluth, MN, Impervious Surface Analysis Tool, impervious surface area, Thunder Bay, ON.

This research will critically examine the difference in data quality and accessibility between two cities through an analysis of predicting the effect of impervious surfaces on water quality. Through using the Impervious Surface Analysis Tool (ISAT) provided by the NOAA Office for Coastal Management, a basic analysis will be completed to compare data quality, accessibility and the application of the data.

The results of the analysis indicate that insufficient data quality hinders the understanding of effects of impervious surfaces on water quality. Thus, showing the requirement for producing and providing high quality data in Canada to allow water resource managers to make appropriate decisions to mitigate, minimize or avoid impacts on the local hydrology of the area.

TABLES

FIGURES

ACKNOWLEDGEMENTS

It is my sincere gratefulness to those organizations that provide freely accessible data online in order to allow scientific research to progress. With organizations providing data this study was made possible.

I am also thankful for Alex Bilyk (Thesis Supervisor) for allowing me to explore my potential by assisting as opposed to guiding me throughout this research. I have learned the importance of self-delegation and the need for persistence throughout the study.

In addition, I would like to personally thank Doug Lowman for providing the Ontario Land Cover data from the Ministry of Natural Resources, and Roman Augustyn from the Lakehead Region Conservation Authority for providing data that is unavailable to the public online. Without their help this study would not have happened to this extent.

Providing Organizations:

Ontario Ministry of Natural Resources
Government of Canada – Statistics Canada
Lakehead University – CARIS
Lakehead Region Conservation Authority
NOAA Office for Coastal Management Digital Coast
Multi-Resolution Land Characteristics (MRLC) Consortium
Minnesota Department of Natural Resources
United States Census Bureau
United States Census Bureau – American FactFinder

1.0 INTRODUCTION

As urban sprawl increases, hydrological modelling of impermeable surfaces is becoming an effective tool for watershed management. As impervious surface area (ISA) increases, models can be used by water resource managers to understand and predict new and complex systems of water flow. However, the quality and accessibility of required data may hinder research to examine altered hydrological systems. This research will critically examine the application of the Impervious Surface Analysis Tool (ISAT), provided by the National Oceanic and Atmospheric Administration Coastal Services Center, in two cities on the west coast of Lake Superior, Duluth, Minnesota, United States of America (USA) and Thunder Bay Ontario, Canada.

These cities were chosen as the quality and accessibility of surface imperviousness, land cover classification, population and watershed data vary from one another. When compared to the data readily available in the USA, Canada's data is limited in its availability. This study will compare the quality and accessibility of data as well examining the results of the ISAT applied to the two cities with varying scenarios.

1.1 OBJECTIVES

The purpose of this thesis is to demonstrate the requirement of providing good quality and accessible data for users, especially management planners concerned with the allocation of resources and land use planning. With data resources accessible, integrated water resource management decisions can be improved across Canada.

The goal of this thesis is to provide EcoSuperior, located in Thunder Bay, ON, Canada, with a resource to allow the organization to focus their de-paving initiative

efforts to sections of the city where there are greater effects of impervious surfaces on the local water quality. As a secondary goal, this thesis will provide Thunder Bay, ON, Canada a resource to assist land use zoning to ensure the effects of impervious surfaces are mitigated.

1.2 HYPOTHESIS

With good quality and accessible data available, researchers can run analyses to improve management decisions and progress science.

1. High quality data analysis has no effect on management decisions.
2. Data accessibility has no effect on management decisions.

2.0 LITERATURE REVIEW

2.1 IMPERVIOUS SURFACES AND WATER QUALITY RELATIONSHIP

It is collectively understood that higher population correlates with the increase of ISA, which has negative correlation of the effects on local hydrology (Wada et al. 2017; Brabec et al. 2002; Dwarkish et al. 2015; Zoppou 2000). With citizens residing mostly in urban catchments as opposed to rural areas, the effect of ISA on hydrology has become a greater concern. According to Statistics Canada, 84% of the 3.2 million people in Canada in 1861 lived in rural areas compared to in 2011, only 18.9% (2018a) of the 33.4 million people in Canada (2018b) lived in rural areas (2018a). With this increase of population in urban areas, ISA is being continuously implemented into urban catchments which may include roads, roofed buildings, sidewalks, parking lots, patios, bridges, and other impermeable anthropogenic structures.

The anthropogenic impact of converting vegetated land to impervious surfaces has profound impacts on the hydrological cycle including runoff and inflows to waterbodies (Wada 2017). Impervious surfaces prevent the natural infiltration of rainfall into the soil which filters many pollutants (Brabec et al. 2002). Without these natural filters, pollutants often end up directly in the water ways thus decreasing water quality (Brabec et al. 2002).

The effect of ISA on local streams is a calculable value which many researchers examine within studies. Tom Schueler proposed three thresholds that identified the quality of streams within a watershed based on the percentage of ISA (*Table 1*) (1994).

These thresholds became valuable to researchers to understand and predict the effect of impervious surfaces.

Table 1. Thresholds of stream quality based on percentage of impervious surface area.

Stream Quality Thresholds	Percentage of Impervious Surface Area
Stressed	1% - 10%
Impacted	11% - 25%
Degraded	26% - 100%

Adapted from Schueler 1994.

ISA has also become a significant concern for storm water managers due to the correlation of decreasing water quality (Ontario 2019). These concerns arise from the overall increase in surface water runoff due to impervious surfaces hindering natural hydrological processes. Through integrated water resource management, decisions to implement new impervious surfaces including roads, parking lots and buildings results in a trade off with decreasing the area of natural landscapes that naturally infiltrate rainfall (Brabec et al. 2002). Without integrated water resource management, decisions are often made by local planners who may not account for the potential impacts on the local hydrology (Brabec et al. 2002).

2.2 DATA ACCESSIBILITY

Accessible data for researchers remains an important question to many prior to commencement of research. All researchers alike require a form of data whether it needs to be produced or retrieved. The provision of accessible data for researchers can benefit in many ways and has been a discussion point for many years. According to a lecturer in epidemiology George Davey Smith stated key reasons why sharing data is beneficial as published in 1994 (Table 2) (Smith 1994).

Table 2. Smith's benefits of sharing data.

Benefits of Accessible Data
1. Reduces the cost of further research.
2. Replication increases robustness of data.
3. Assist in formulation of research questions.
4. Create new datasets with data linkage.
5. Build on results by combining data

Source: adapted from Smith 1994.

Despite numerous benefits of sharing data, many researchers choose not to share data whether it is raw or altered. The National Science Foundation-funded DataONE project conducted a survey about scientists' opinions and perceptions of data sharing (Tenopir et al. 2011). While the 1 329 respondents were global, the survey targeted researchers in North America (Table 3) (Tenopir et al. 2011).

Table 3. National Science Foundation-funded DataONE project respondents.

Respondents	Percentage of Distribution
North America	73%
U.S. South	36%
U.S. Midwest	26%
U.S. West	25%
U.S. Northeast	14%
Europe	15%
Asia/Oceanic	7.3%

Adapted from: Tenopir et al. 2011.

Key results from the survey indicated that the majority of people agree, when accessibility to data is limited it will impede the progress of scientific research (Figure 1) (Tenopir et al. 2011). However, when asked about how they share data, approximately one third of the respondents stated they do not make their data accessible (Tenopir et al. 2011). This revealed that most researchers surveyed understand the importance, yet many choose to keep their data in private databases.

	Agree Strongly	Agree Somewhat	Neither Agree Nor Disagree	Disagree Somewhat	Disagree Strongly
Lack of access to data generated by other researchers or institutions is a major impediment to progress in science.	353 (27.2%)	520 (40%)	230 (17.7%)	149 (11.5%)	48 (3.7%)
Lack of access to data generated by other researchers or institutions has restricted my ability to answer scientific questions.	228 (17.6%)	422 (32.5%)	297 (22.9%)	238 (18.4%)	112 (8.6%)
Data may be misinterpreted due to complexity of the data.	383 (29.6%)	590 (45.6%)	217 (16.8%)	77 (6%)	26 (2%)
Data may be misinterpreted due to poor quality of the data.	379 (29.4%)	540 (41.8%)	232 (18%)	107 (8.3%)	33 (2.6)
Data may be used in other ways than intended.	410 (31.8%)	539 (41.8%)	249 (19.3%)	68 (5.3%)	23 (1.8%)

doi:10.1371/journal.pone.0021101.t008

Source: Tenopir et al. 2011

Figure 1. National Science Foundation-funded DataONE project Data Reuse results.

Of those who stated they share data, 46% or greater of the people do not share data electronically (Figure 2). And less than 10% make all their data available electronically (Figure 2).

	None	Some	Most	All	Total
On My Organization's Website	495 (45.9%)	378 (35.1%)	143 (13.3%)	62 (5.8%)	**1078 (100%)**
On the Principal Investigator's Website	553 (56.7%)	303 (31.0%)	87 (8.9%)	33 (3.4%)	**976 (100%)**
Through a National Network	470 (46.4%)	331 (32.6%)	153 (15.1%)	60 (5.9%)	**1014 (100%)**
Through a Regional Network	579 (64.7%)	238 (26.6%)	58 (6.5%)	20 (2.2%)	**895 (100%)**
Through a Global Network	550 (57.6%)	242 (25.3%)	111 (11.6%)	52 (5.4%)	**955 (100%)**
On My Personal Website	668 (72.7%)	173 (18.8%)	49 (5.3%)	29 (3.2%)	**919 (100%)**
Other	370 (65.3%)	94 (16.6%)	47 (8.3%)	56 (9.9%)	**567 (100%)**

doi:10.1371/journal.pone.0021101.t010

Source: Tenopir et al. 2011

Figure 2. National Science Foundation-funded DataONE project Data Sharing Practices results.

Lastly, 75% of the respondents stated they share data with others, however only 36% of those researchers stated their data is easily or somewhat easily accessible (Figure 3) (Tenopir et al. 2011).

	Agree Strongly	Agree Somewhat	Neither Agree Nor Disagree	Disagree Somewhat	Disagree Strongly
I share my data with others.	418 (32.3%)	551 (42.6%)	199 (15.4%)	95 (7.3%)	30 (2.3%)
Others can access my data easily.	150 (11.6%)	317 (24.6%)	310 (24%)	307 (23.8%)	207 (16%)

doi:10.1371/journal.pone.0021101.t011

Source: Tenopir et al. 2011

Figure 3. National Science Foundation-funded DataONE project Data Sharing results.

This survey reveals that the majority of people understand the importance of accessible data. However, most do not make their data freely accessible for others despite their willingness to share data. When researches and institutions do not make their data accessible it further prohibits the progress of science. Therefore, in order to improve accessibility; initiative needs to be taken to ensure that data will be shared.

Data sharing is progressing through the establishment of data sharing centres around the world. They have become increasingly important for researchers to access data. For example, in West China the National Natural Science Foundation of China understood that data sharing would allow integrated study despite China's previous history of a weak data sharing culture (Li et al. 2011). This led to the establishment of the Environmental and Ecological Science Data Center for West China (Li et al. 2011). It was agreed upon that withholding scientific data as personal property is not the right course of action (Li et al. 2011). Thus, by creating the data centres, scientific researchers can reap the benefits of accessible data as stated in Table 1.

2.2.1 Data Accessibility in United States of America

In the USA, data accessibility is generally freely accessible online. The government of the United States has made it possible for the public to access most data. In fact, the Federal Data Strategy (FDS) outlines numerous Principles and

Practices to increase the value of publicly accessible Federal Government data (USAGov n.d.).

The FDS will ultimately facilitate usage of Federal Government data which allows data to be accessible for public use including education, private sector, government and more (USAGov n.d.). It will also improve data accessibility regarding file format, time of availability, maximal data availability, use of new technology, and allow access to sensitive data without compromising confidentiality (USAGov n.d.).

2.2.2 Data Accessibility in Canada

Accessible data in Canada remains more restricted and limited in comparison to the USA. The government of Canada has restricted data sharing where only select organizations can freely access statistical data (StatCan 2019b). These organizations receive data as they have active agreements that allow access for statistical uses only (StatCan 2019b). For researchers to access most data in Canada, inquiry and networking may be required. Some organizations may choose to share their data online. However, the decision to do so remains within the organization. In addition, the Canadian government has an Open Data initiative to provide the general public, researchers, students, and others accessible data (GovCan 2019). Similar to the USA FDS, the goal is to provide good quality data in an accessible format. Canadian data is becoming increasingly available which is prevalent throughout this study.

2.3 DATA QUALITY

While data may be accessible, it may not meet the requirements of the user's application. For example, data posted online may be processed, does not contain raw

data or the owner has withheld information for various reasons. This may hinder the quality of the data.

Data quality remains a biased and non-quantitative decision for researchers. Often the quality of data is measured on the following variables: completeness (Three Rivers District Council 2008; Pipino et al. 2002), security (Pipino et al. 2002), timeliness (Three Rivers District Council 2008; Pipino et al. 2002), relevance (Three Rivers District Council 2008; Pipino et al. 2002), validity (Three Rivers District Council 2008), accuracy (Three Rivers District Council 2008) and others depending on the application. Researchers must consider different aspects of data to deem its appropriateness. In some cases, data can be altered, cleaned or processed for the purposes of the researcher therefore access to all data may have more quality than reduced datasets.

2.4 HYDROLOGICAL MODELLING

Hydrological models are highly dependent on a large range and variation of parameters in order to develop a model that achieves accurate results. Research has proved that there are many models that have been developed with differing parameters. However, there is not one specific model that supports all types of research and studies.

Therefore, an ideal hydrological model will give rise to results with the least parameters and complexity that will predict the hydrological processes for its intended purpose (Gayathri et al. 2015). Key inputs into models often include rainfall data and drainage area. Other required data may be properties and characteristics of: soil, vegetation, topography and water (Gayathri et al. 2015). For the purposes of modelling

the effect of impervious surfaces, a model needs to obtain additional parameters to include data on ISA in order to accurately model hydrological processes within an urban catchment. ISA is a unique parameter to urban catchments as it impedes water flow causing increased surface runoff. The ratio of ISA is a critical parameter to be included in hydrological models for storm water runoff (Brabec et al. 2002). Therefore, models must consider human influences or assume error within the hydrological model.

For the purpose of this study, the assumption that the increase of impervious surfaces will have a direct influence on the decrease of water quality will allow a majority of the aforementioned parameters to be disregarded until further research is desired.

2.4.1 Impervious Surface Analysis Tool

The National Oceanic and Atmospheric Administration (NOAA) Coastal Services Center developed the Impervious Surface Analysis Tool (ISAT) to provide management planners with a resource that can be utilized to predict the impact of ISA on water quality (NOAA 2013). The analysis tool is used to predict results that may occur when implementing or removing infrastructure. This is done by applying impervious surface coefficients to land cover data which the tool uses to calculate the amount of ISA within defined analysis theme polygons (NOAA 2013).

The ISAT is accessible to the public on the NOAA Coastal Services Centre website where users can download the file folder containing tutorial data, scripts and a toolkit which is compatible with ArcGIS.

To utilize the ISAT there are four key datasets that are required. It is important to note that this tool is different from most hydrological models as it is a prediction tool as opposed to a hydrology analysis tool. Therefore, the required datasets are land cover, impervious surface coefficients, population data and an analysis theme polygon file (watershed units are used for this study) (NOAA 2013). Watershed units are selected to provide a spatial understanding of the impact of ISA face area on the local watersheds and to further predict the potential effects on water quality.

The impervious surface coefficients are a key component of the ISAT which is calculated and inputted into an excel file containing a coefficient for each land cover type and its associating high, medium and low population statistics (NOAA 2013). The ISAT then uses these coefficients to create a layer that displays the potential impact on water quality (NOAA 2013). The ISAT has a predetermined classification for the impact of ISA on water quality as Protected <10% ISA, Degraded 10% to 25% ISA and Impacted >25% ISA (NOAA 2013). These values are based on similar thresholds as Tom Schueler had proposed in 1994 (*Table 1*). It is important to recognize that the names of the thresholds differ. A comparison of the thresholds and classification of the effect of imperviousness is available in Table 4.

Table 4. Comparison of Tom Schueler and NOAA thresholds of the effects of impervious surfaces.

Threshold	Tom Schueler	Threshold	ISAT
1% - 10%	Stressed	<10%	Protected
11% - 25%	Impacted	10% – 25%	Degraded
26% - 100%	Degraded	>25%	Impacted

This research will apply the ISAT with Change Scenario script supplied in the Toolbox from NOAA Coastal Services Centre (NOAA 2013). The results of the analysis will provide a prediction of the effect of ISA on the local water quality and will provide a relatively simple tool for management planners to predict the effect of ISA.

2.5 RESEARCH DATA

Required data is collected from various resources online, networking and image processing.

2.5.1 Land Cover Classification

The land cover data is important in order to identify natural environment from developed environment. In this study the ISAT will use land cover classification to identify the imperviousness coefficients for each land cover classification for high, medium and low population (NOAA 2013).

Land cover classification data for Minnesota was obtained from NOAA Office for Coastal Management. This organization provides C-Cap Regional Land Cover and Change data which is updated every five years at a 30-metre resolution (NOAA 2019). The data was developed by using standardized data and practices to maintain consistency (NOAA 2019). The online download provided a Raster Dataset of the land cover classification and an XML Document describing the data. There is no metadata provided, however, the NOAA Office for Coastal Management website does provide some accompanying PDF Documents explaining the data.

The land cover classification of Ontario was obtained by contacting the Ministry of Natural Resources (MNR). During preliminary search for land cover classification

data, it was originally only accessible for the entirety of Canada from Natural Resources Canada. However, there was no data accompanying the ArcGIS compatible shapefile raster image. Alternatively, the Ontario Ministry of Natural Resources (OMNR) provided a shapefile that can be utilized and an accompanying document with an explanation of the file. Since the timing of the preliminary search Ontario Land Cover Compilation v.2.0 was added to the Ontario GeoHub data sharing site on October 15, 2019.

The Ontario land cover classification file is at a 15-metre resolution with 29 land cover classes (Ontario GeoHub 2019). The database was created by using three land cover databases in Ontario including the Far North Land Cover v1.4, Southern Ontario Land Resource Information System (SOLRIS) v1.2, and Provincial Land Cover 2000 Edition (Ontario GeoHub 2019). The three different land cover databases were combined to create a single cohesive database with 29 land cover classes which were described in the accompanying data specifications document (Ontario 2014).

When the two datasets are compared the Minnesota data has a coarser resolution and less land cover classes than the Ontario datasets. The resolution of the data may result in less accuracy for Minnesota. However, for the purpose of the tool the difference in resolution may not be critical to the user. The number of classes and type of classes should not impact the results, however it will impact the interpretation of the results and comparison between the two different datasets. For example, the land cover classification data for Minnesota provides urban areas with three levels of

classification based on the intensity of development. Whereas the data for Ontario urban areas are amalgamated into one classification.

Ideally, the user of the data will obtain data that includes classification of the development intensity.

2.5.2 Surface Imperviousness

Surface Imperviousness data is required for the analysis using the ISAT. The tutorial document directs users to the Multi-Resolution Land Characteristics Consortium which provides the *NLCD 2016 Percent Developed Imperviousness (CONUS)* data for surface imperviousness of the entirety of the USA (MRLC 2019). Through the partnership of multiple federal agencies, the National Land Cover Database is developed and open sourced (NOAA n.d.a.). This data displays the 30-metre pixels as percentage of surface imperviousness (MRLC 2019).

There is no surface imperviousness data available for Thunder Bay, ON, online or from the Ministry of Natural Resources.

2.5.3 Population Information

Two sets of data were required in relation to population information. These include census tracts and population statistics for the census tracts. More specifically population density, however population data can be used to calculate population density.

Census tracts for Canada and USA have different definitions, however they remain relatively similar which was deemed acceptable for both cities to use tracts for population statistics for adequately comparable results. Census tracts in Canada are

defined as areas that have relatively stable population that ranges between 2 500 and 8 000 people within metropolitan areas with populations greater than or equal to 50 000 (StatCan 2018c). Where census tracts in the USA are defined as a "relatively permanent statistical subdivisions of a county" and has a population size ranging from 1 200 to 8 000 people (optimally 4 000 people) (USCB 2012).

For both Canada and the United States, the data is available online, however the statistical population data was kept separately from the geographic data. This system is used to assist with confidentiality and privacy of the citizens.

Data for Canada was collected online from Statistics Canada where the Cartographic Boundary File for Census Tracts can be downloaded as a ArcGIS compatible file (StatCan 2019a) and the population densities were collected individually by Census Tract by searching the Geographic Code of the Census Tract of interest (StatCan 2018d).

Data for the United States was found online at two different websites. The TIGER/Line Shapefiles for the geographic boundaries of Census Tracts was found on the United States Census Bureau website (USCB n.d.) and the population statistics was obtained from American FactFinder by the United States Census Bureau where a Microsoft Excel file was downloaded (USCB 2010a) and linked to the shapefile by using the GEOIDs (USCB n.d.).

While the 2019 TIGER/Line® Shapefile is available online, the 2016 census data is utilized to remain consistent with the other data for Minnesota and is the same year as the census data collected for Ontario.

2.5.4 Analysis Theme Polygon

The analysis theme polygon is up to the discretion of the user who would likely base the analysis theme polygon on the purpose of their study. For the purpose of this thesis study the analysis theme polygon is the same one as used for the tutorial provided with the ISAT. This analysis theme polygon is watershed based as it will provide information on the effect of imperviousness within each watershed boundary polygon.

The watershed boundary data for Duluth, Minnesota was obtained online from the Minnesota Department of Natural Resources Geospatial Commons (MNDNR n.d.). This data provides multiple different levels of Watersheds classified by Hydrological Unit Codes (HUC) and catchments.

In the USA, hydrological units are given a code which identifies boundaries where all water flows to a single outlet on various scales (NRCS 2007). The respective HUC is then used to determine the imperviousness coefficients.

MNDNR also provided Metadata following the Minnesota Geographic Metadata Guidelines that includes Data Quality information which addresses consistency and completeness (MNDNR 2014).

In Canada there are 3 levels of watershed classification available. These include major drainage area, sub-drainage area and sub-sub-drainage area (StatCan 2017a) which are also known as Primary, Secondary and Tertiary watersheds respectively. The smallest sized watershed (tertiary) is too large to be used for the ISAT as the Dog Watershed encompasses the majority of the City of Thunder Bay (Figure 4).

Source: MNRF 2020

Figure 4. Tertiary Watersheds (indicated by green line borders) encompassing Thunder Bay, Ontario Canada.

However, the Lakehead Region Conservation Authority (LRCA) located in Thunder Bay, ON has refined watersheds for the region based on the Quaternary watersheds of Ontario.

3.0 MATERIALS

3.1 STUDY AREA

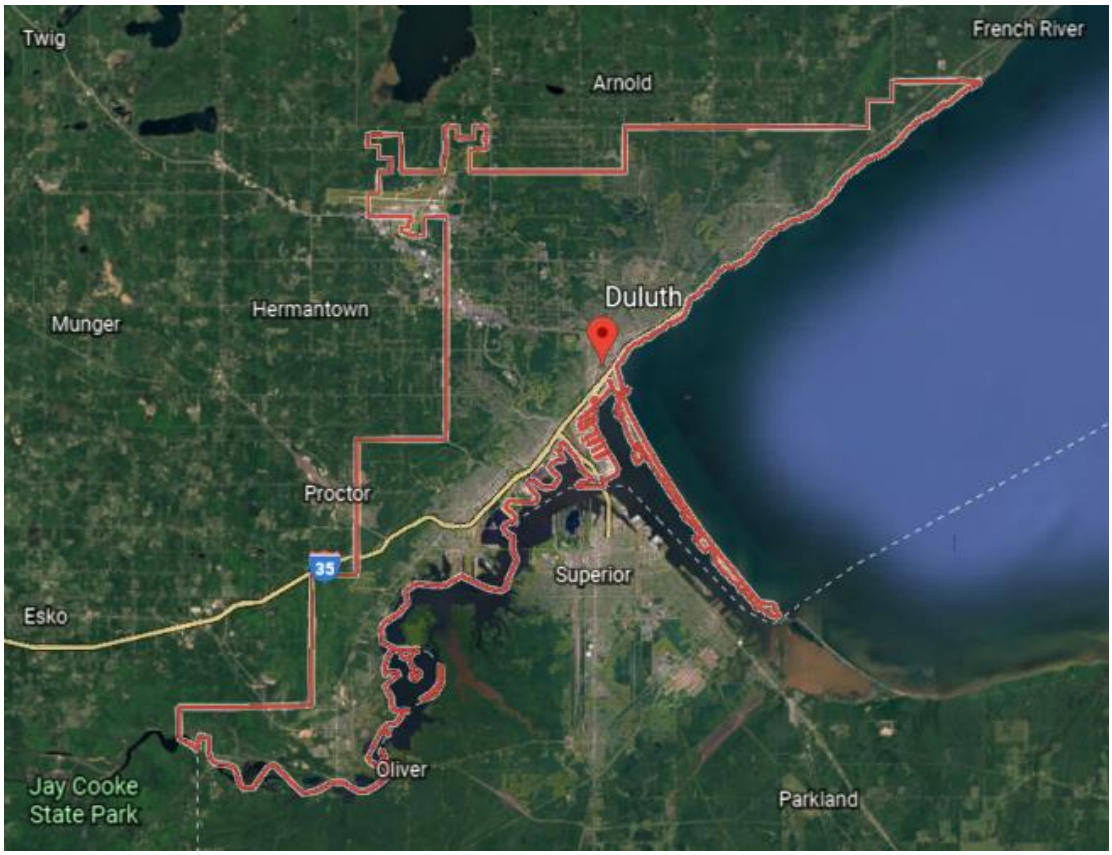
The purpose of the study is to assess data accessibility and quality using the ISAT tool on two different cities on the west coast of Lake Superior. The first city, Thunder Bay, ON, Canada (*Figure 5*), was selected because the study is conducted within the city at the Lakehead University institution. This city was also selected to provide EcoSuperior with a resource that they can utilize to determine where de-paving initiatives should occur within the city upon completion of the study.

t

(Google Earth n.d.a)

Figure 5. City of Thunder Bay, Ontario, Canada study area (48.38°N, 89.25°W).

Data accessibility and quality will be compared to data for Duluth, Minnesota, United States of America (*Figure 6*). This city was selected as it is located approximately 275 km southwest of Thunder Bay, ON within the United States of America (USA). This cross-border city allows the comparison of data accessibility and quality as different regulations and policies affect data sharing.



(Google Earth n.d.b)

Figure 6. City of Duluth, Minnesota, United States of America study area (46°N, 92.10°W).

These two cities are vastly different, however were selected as they are the largest urban catchments on the west coast of Lake Superior. Both cities have similar

aspects including population totals, inflow of people from neighbouring areas, and its hydrology drainage into Lake Superior.

4.2 DATA SELECTION

Data was selected based on the availability of the types of data and the direction provided by NOAA in the Tutorial information document for the ISAT tool. Required data includes the following: Analysis Theme shapefile of the area chosen to analyze (NOAA 2013); in this case watersheds are selected to represent the effect on water quality within the watersheds. Land Cover Grid (NOAA 2013) which constitutes the land cover data provided by NOAA Coastal Services Center Digital Coast website (NOAA 2013). Impervious Surface Coefficients which is developed by the user with impervious surface data provided by the Multi-Resolution Land Characteristics Consortium (MRLC) National Land Cover Database (NOAA 2013). Lastly, Population Data (NOAA 2013) which includes census boundaries and statistics. Therefore, the required datasets are:

1. Surface Imperviousness
2. Land Cover Classification
3. Population Statistics
4. Census Tract Boundaries
5. Watershed Boundaries

3.2 DATA RESOURCES

Data is collected from numerous sources including the federal government, provincial and state ministries, federal partnerships, educational institutions or

environmental agencies. The primary source of data was collected from online sources where providing organizations are most commonly governmental, or provincial or state ministries. Data that was not available online was accessed by contacting a member of a providing organization.

The data was provided in an array of formats as there is no policy indicating how data is to be publicly provided. Data differs in data type, acquisition, updates, coordinate system, and datum.

3.2.1 Thunder Bay, ON

Data sources for Thunder Bay, ON varies greatly and required some direct contact to providing organizations. The data used and its sources for Thunder Bay, ON, Canada are summarized in Table 5. Not all data required for the ISAT analysis was publicly accessible.

Table 5. Impervious Surface Analysis Tool data sources for Thunder Bay, Ontario, Canada.

Data Required	Providing Organization	Collection Method
Surface Imperviousness	*Lakehead University (secondary provider)	Contact with organization employee
Land Cover Classification	**Ministry of Natural Resources	Contact with organization employee
Population Statistics	Statistics Canada	Online data collection
Census Tract Boundaries	Statistics Canada	Online data download
Watershed Boundaries	Lakehead Region Conservation Authority	Contact with organization employee

*Data is developed from data provided from Lakehead University that is not the primary source of the data.
**Land cover database package became available online after data collection

There was no Surface Imperviousness data available for Thunder Bay, ON thus the data was developed using 'enhanced' Forest Resource Inventory (eFRI) imagery for the Thunder Bay region. The imagery used is sized at 5km square tiles and encompasses most of the City of Thunder Bay. The data development is outlined in image processing methodology.

Land Cover Classification data was obtained from the Ministry of Natural Resources (MNR) through contact with an employee. However, the data became available online through Ontario GeoHub throughout the study. The Ontario Land Cover Classification data provides 29 land cover classes at a resolution of 15 metres.

Population Statistics and Census Tract Boundaries are provided online by Statistics Canada. Obtaining Population Statistics had two methods where a large file could be downloaded and normalized or individually obtained by using the Statistics Canada Census Profile, 2016 Census search engine to search the applicable CTUID (Census Tract Unit ID) Geographic Code. The latter method was used in this study to develop a Microsoft Excel Sheet containing the CTUID and population density per square kilometer. The data accuracy is within 5 individuals per census tract.

Census Tract Boundaries were obtained online by downloading the Cartographic Boundary File, ArcGIS shapefile for Census Tracts. This file provides geographic polygons of the census tract boundaries and identification.

Watershed Boundary shapefiles were provided by the Lakehead Region Conservation Authority (LRCA) and obtained through contact with an organization employee. These shapefiles are refined versions of the Quaternary watershed

boundaries for the Thunder Bay region based on the 20m resolution DEM layer according to Roman Augustyn (in email, Mar. 6, 2020). These are the smallest sized watersheds for Ontario that is in an accessible format which is used to understand the regional watersheds.

The individual data file information including the exact name of the retrieved data file used, acquisition, data type and spatial information is provided in Table 6.

Table 6. Data information for Thunder Bay, ON, Canada ISAT analysis.

Data Required	File Name	Year Data Acquired	Data Type	Geographic Coordinate System	Datum
[1]Surface Imperviousness	[2]#.ecw	2008	File System Raster	North American 1983 UTM Zone 16N	North American 1983
Land Cover Classification	OLCC_V2	[3]2002	File Geodatabase Raster Dataset	NAD 1983 Lambert Conformal Conic	North American 1983
Population Statistics	[4]N/A	2016	Microsoft Excel Comma Seperated Values File	N/A	N/A
Census Tract Boundaries	Lct000b16a_e.shp	2016	Shapefile Feature Class	North American 1983	North American 1983
Watershed Boundaries	Watershed_Assessment.shp	unknown	Shapefile Feature Class	North American 1983 UTM Zone 16N	North American 1983

[1]Surface Imperviousness information is based on the imagery.
[2]Multiple image files that follow a standard format for naming images.
[3]Year acquired pertaining to the study area (regions acquired during different time periods).
[4]N/A: Not Applicable information as the data is developed by the user.

3.2.2 Duluth, MN

Data pertaining to Duluth, MN is provided by government sources, governmental partnerships and state ministries. These resources are displayed in Table 7.

Table 7. Impervious Surface Analysis Tool data sources for Duluth, Minnesota, United States of America

Data Required	Providing Organization	Collection Method
Surface Imperviousness	Multi-Resolution Land Characteristics Consortium	Online download
Land Cover Classification	NOAA Coastal Services Center	Online download
Population Statistics	American FactFinder	Online download
Census Tract Boundaries	United States Census Bureau	Online download
Watershed Boundaries	Minnesota Department of Natural Resources	Online download

The Surface Imperviousness data was obtained online from Multi-Resolution Land Characteristics Consortium which is a collaboration of federal agencies that provides updated imperviousness data of the entirety of the United States every five years (MRLC 2019). The data contains the percentage of imperviousness for every 30-meter pixel (MRLC 2019).

The Land Cover classification data was obtained from National Oceanic and Atmospheric Administration Office for Coastal Management. The organization provides updated data every five years and covers 23 different land classifications at 30-meter pixel resolution (NOAA 2019).

Population Statistics data was obtained from the United States Census Bureau (USCB) through the American FactFinder website. The census data for Duluth, MN is obtained for the St. Louis County which encompasses the study area and provides

various statistical information for each census tract including population density per square mile of land area.

Census Tracts Boundary data was obtained directly from the United States Census Bureau website which provides TIGER/Line Shapefiles of legal boundaries including census tracts and is updated yearly.

Watershed data was obtained from the Minnesota Department of Natural Resources database. The database provides multiple levels of watersheds based on HUCs (Hydrological Unit Codes). HUC level 7 was used for the study.

The individual data file information including the exact name of the retrieved data file used, acquisition, data type and spatial information is provided in *Table 8*.

Table 8. Data files information for Duluth, MN, United States of America ISAT analysis.

Data Required	File Name	Year Data Acquired	Data Type	Geographic Coordinate System	Datum
Surface Imperviousness	NLCD_2016_ Impervious_L48_ 20190405.img	2016	File System Raster	Albers Conical Equal Area	WGS 1984
Land Cover Classification	Mn_2016_ccap_ land_cover_ 20170410.img	2016	File System Raster	Albers Conical Equal Area	North American 1983
Population Statistics	DEC_10_SF1_ GCTPH1.CY07_ with_ann.csv	2010	Microsoft Excel Comma Separated Values File	N/A	N/A
Census Tract Boundaries	Tl_2016_27_tract .shp	2016	Shapefile Feature Class	GCS North American 1983	North American 1983
Watershed Boundaries	dnr_watersheds_ dnr_level_07_ minors	1979	Shapefile	North American 1983 UTM Zone 15N	North American 1983

3.3 SOFTWARE

Two major software products were used for the study, ArcGIS Desktop and eCognition. ArcGIS Desktop by Esri was primarily used for the ISAT analysis as well as for the development of the imperviousness layer for Thunder Bay, ON. Secondly, eCognition by Trimble Geospatial was used for classifying impervious surfaces for the imperviousness layer for Thunder Bay, ON.

4.0 METHODOLOGY

4.1 STUDY DESIGN

This study will use the five different data files for each city to ultimately utilize the ISAT analysis tool to predict the effect of imperviousness on water quality for each city. Through the process of obtaining, processing, and examining the data to utilize the tool, the accessibility and quality of the data for each city will also be determined. Data accessibility will be determined based on the ability to obtain the required files for the ISAT analysis and data quality will be based on the available metadata and accompanying information for each dataset.

4.3 DATA PREPARATION

Each data set must be prepared to ensure that the coordinate systems, datums, and measurements are consistent to properly compare the results. Since the ISAT results are based on percentages and results are shown in metres it is assumed that the difference between city measurements (E.g. Kilometres versus Miles) is negligible except for the population density value. However, within cities data needs to remain consistent.

For both cities, the data was transformed and/or projected as needed to the GCS North American 1983 coordinate system and the North American 1983 datum. This will ensure the consistency within the cities' measurements.

In addition, for both study areas the census tract data does not include population statistic data thus requiring combining of the datasets. A Joins was used to link the population statistics excel file to the census tract shapefile.

For Thunder Bay, ON, the data was collected and formatted appropriately to easily identify the field required for a Joins by linking the CTUID (Census Tract Unit ID).

For Duluth, MN, this was done by the normalization of the excel file to contain a 6-digit string of the census tract code which is derived from the Target Geo ID2 column and subtracting 27137000000 to obtain the Tract ID. Any strings with less than 6 digits, zeros are added to the beginning of the sequence and the cells are formatted as text. In addition, the population density value was converted from density per square mile to density per square kilometer.

Lastly, as there was no existing data available for the surface imperviousness of Thunder Bay, ON, imagery was processed to develop an imperviousness database as outlined in Image Processing.

4.2 IMAGE PROCESSING

Due to the unavailability of ISA data required for Thunder Bay, ON, image processing was required to develop an ISA database. Ideally, for comparison of data between Thunder Bay, ON and Duluth, MN the exact methodology for determining the ISA layer for the USA would be used. Determination of this methodology was not possible thus imagery was analyzed with the following methodology.

Step 1: Images were provided in an ECW file format which was converted to a TIFF file format through ArcCatalog Export Raster to Different Format.

Step 2: With eCognition, a sample TIFF image was used to create a standard ruleset to process each image in eCognition (Figure 7). As each image results

differently from the standard ruleset due to image features, variations of the ruleset were used to adequately classify impervious surfaces.

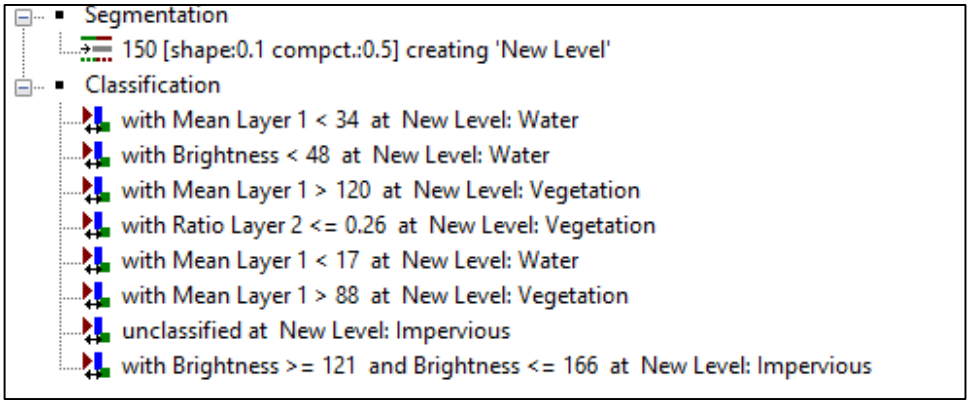


Figure 7. eCognition classification ruleset for impervious, water and vegetation surfaces of eFRI imagery.

Step 3: The resulting classification was exported as a shapefile with the total pixel area and class name (Impervious, Vegetation and Water).

Step 4: To develop a grid of the extent of the imagery a Polygon Shapefile was created from the boundaries of the TIFF imagery using ArcGIS through Reclassify, Raster to Polygon and Merge tools. A Field was added to record the percentage of imperviousness for each tile.

Step 5: The percentage of imperviousness for each tile was calculated by summarizing each classified shapefile by Pixel Area per Class and using the following equation $\frac{100}{250000000} * pixel\ area$. The value was then populated in the appropriate tile of the imperviousness grid.

Step 6: The symbology was adjusted to defined interval (1.00) and the colour map from the USA imperviousness data was imported. Importing of the same colour map allows a visual comparison of the data.

4.4 IMPERVIOUS SURFACE ANALYSIS TOOL

The method of utilizing the ISAT is outlined in the Tutorial for ArcGIS 10.x as provided by the NOAA (2013). This tool is to be used with the data provided for both Thunder Bay, ON and Duluth, MN. As not all data provided is available in an exact format for the ISAT, extra steps were required to set up the data for analysis as outlined in Data Preparation.

The ISAT Imperviousness Coefficients are based on population densities which are based on the following thresholds in square kilometres Low (<=500), Medium (>500 and <=2500), and High (>2500).

4.6 DATA ACCESSIBILITY ASSESSMENT

The data accessibility assessment was conducted throughout the process of obtaining the required data for the ISAT analysis. By the record of data acquisition process and its barriers, an assessment was made on data accessibility which is quantified on a scale of 0 – Not Available, 1 – Inaccessible, 2 – Accessible with Contact and Barriers, 3 – Inaccessible Online/Accessible with Contact, 4 – Accessible Online with Barriers, 5 – Freely Accessible Online.

4.7 DATA QUALITY ASSESSMENT

With each dataset, the following data quality characteristics were used to perform a data quality assessment: completeness, security, timeliness, relevance, validity, and accuracy. The definitions and parameters for the data quality analysis are outlined in Table 9. It is important to note that this assessment will remain accurate in terms of using metadata and/or accompanying information provided within the

dataset by the providing organization and the providing organizations' website. The quality analysis is biased in terms of the application of the data for the ISAT analysis.

Table 9. Data Quality characteristics and definitions.

Data Quality	Definitions
Completeness	The completeness of the data is subjective based on the end users need for the data. The completeness takes into consideration of the totality of the provided data. It asks if the provision of data is partial or whole.
Security	The security of the data considers the extent of availability to the public.
Timeliness	The timeliness of the data is dependent on timeline the data is captured/analysed, packaged, and made available. In some cases, the data may not be made available to users until a later date and rendered poor timeliness due to the inability to provide the most recent data.
Relevance	The relevance of the data is based on the usability of the format and the data provided. It asks if the data is relevant to the purpose of its use.
Validity	The validity is based on the compliance with rules and regulations set in place to ensure consistency of data provided. This may consider federal laws that impact data collection or organizational rules to maintain consistency. Validity may also be based on the source of the data.
Accuracy	The accuracy of the data is based on the detail provided by the organization. It considers how accurate the relevant data is in regards of its use.

Adapted from Pipino et al. 2002 and Three Rivers District Council 2008.

4.5 APPLICATION ANALYSIS

An analysis on the effect of data quality by executing a basic analysis with the ISAT tool 'ISAT with Change Scenarios' script provided by the NOAA ISAT Tutorial package. A basic analysis of both cities is conducted for High, Medium, and Low

Coefficients which is designed for comparing scenarios of different imperviousness coefficients.

The results of the Basic Analysis provide an output layer that predicts the impact of impervious surfaces on the water quality based on the percentage of imperviousness estimated from each analysis field (NOAA 2013). The output also indicates the mean and total area of imperviousness within the analysis field (watershed). This information will be used to compare the following:

1. The effect of data quality on the analysis.
2. The effect of data accessibility on the analysis.

5.0 RESULTS

5.1 DATA ACCESSIBILITY ANALYSIS

Data utilized for the ISAT were collected from various types of sources including online and direct contact sources. All data utilized for the Duluth, MN study area is obtained from online sources whereas Thunder Bay, ON data is obtained by contacting organization employees in addition to online sources. Table 10 indicates the type of data obtained and its accessibility on a scale 0 to 5.

0 – Not Available

1 – Inaccessible

2 – Accessible with Contact and Barriers

3 – Inaccessible Online/Accessible with Contact

4 – Accessible Online with Barriers

5 – Freely Accessible Online

Table 10. Data accessibility of acquired data for Thunder Bay, Ontario and Duluth, Minnesota Impervious Surface Analysis.

	Data File Type	How Data is Accessed	Accessibility Rating
Duluth, MN	Land Cover Classification	Online Database	5
	Surface Imperviousness	Online Database	5
	Population Statistics	Online Database	5
	Census Tract Boundaries	Online Database	5
	Watershed Boundaries	Online Database	5
Thunder Bay, ON	Land Cover Classification	MNR Employee + Online Database	*3 and 5
	Surface Imperviousness		0
	Population Statistics	Online Database	4
	Census Tract Boundaries	Online Database	5
	Watershed Boundaries	Lakehead Region Conservation Authority Employee	2

*Two accessibility ratings as accessibility to the data changed during the duration of the study

Duluth, MN data was fully accessible online and publicly available thus receiving a score of 5 for all required data. There were no barriers to obtaining data for the Duluth, MN study area.

Thunder Bay, ON data accessibility ranges from Not Available to Freely Accessible. The Surface Imperviousness data for Thunder Bay, ON did not exist according to a Ministry of Natural Resources employee (in email, Mar. 29, 2020).

The Population Statistics had a barrier in regard to data format. An excel file was available, however it contained extensive information which must be normalized to utilize appropriately. The alternative was to individually obtain the values by searching the individual census tract codes to obtain the population density.

The Thunder Bay, ON Land Cover Classification data had become increasingly accessible throughout the study. Primary contact to an MNR employee was required to obtain the data, however it became publicly accessible online at Ontario GeoHub at a later date. The data provided by the MNR employee is the same data package now available online. The first score represents the poor accessibility and requirement to directly contact the ministry as data was not publicly accessible thus given a score of 3. The score increased to 5 upon release of the data online at Ontario GeoHub.

The watershed data was from the LRCA, the provided database is of refined quaternary watersheds. This data was accessible through contact from a LRCA employee and was provided upon request. If one were not able to obtain this data an alternative option is the publicly available Quaternary Watershed database available on Ontario GeoHub.

Overall, required data for the ISAT analysis tool was fully available and freely accessible for Duluth, MN and was more restricted for Thunder Bay, ON.

5.2 DATA QUALITY ANALYSIS

Each dataset or database utilized for the ISAT analysis provided varying levels of information defining the Data Quality. Information utilized for the data quality analysis was typically found within the dataset. Some additional internet-based search was required to obtain the information.

5.2.1 Duluth, Minnesota Data Quality Analysis

5.2.1.1 Land Cover Classification

The land cover classification data for Duluth, MN was obtained from NOAA Office for Coastal Management. Information was used from the online sources and related documents. There was no metadata information available in the dataset that accounts for the quality of the data.

Table 11. Data quality analysis of land cover classification – Duluth, Minnesota

Data Quality	Minnesota Land Cover Classification
Completeness	The Federal Geographic Data Committee Management Plan ensured that the data is 'sufficiently complete' for the requirements of the United States government (FGDC n.d.). The data was provided by the state level to reduce the size of the datafiles.
Security	The land cover data was open sourced data provided by the NOAA Office for Coastal Management (NOAA 2019). There is no known security measures on the data.
Timeliness	According the XML file accompanied with the data file the Landsat imagery utilized for creating the land cover information was acquired either before or after the '2016-Era' (NOAA 2016). This is due to the inability to capture imagery for the whole extent within one year (NOAA n.d.b). The datasets were updated at 5-year intervals. According to NOAA within 5 years the most a region changes is 20% while vast areas are only changing by less than 6% (NOAA n.d.b).

Relevance	The relevance of the data is exact as the ISAT was developed with this dataset. In addition, the ISAT Tutorial directs users to the NOAA for data access (NOAA 2013).
Validity	The validity of the land cover data is complied with the National Geospatial Data Asset (NGDA) Management Plan in accordance with the OMB (Office of Management and Budget) policy and Administration direction to ensure the reliability of the data (FGDC n.d.)
Accuracy	The FGDC NGDA ensures the accuracy of the land cover data. The NGDA management plan provides a framework and processes for the management of the data (FGDC n.d.). The goal of the FGDC is to provide robust and accessible geospatial datasets (FGDC n.d.). This ensures that the accuracy of the data is consistent. The data itself was analyzed at a 30-metre resolution (NOAA 2019). When updating datasets, only areas that experienced change in the landscape are reclassified (NOAA n.d.b). The change was identified by using a change detection tool which allows better efficiency and data consistency and reduces the cost of developing the product (NOAA n.d.b). The overall accuracy of the C-CAP land cover data is 85% (NOAA n.d.b). With the development of the data there was significant quality control to ensure accuracy through training data that is edited by hand and is continuously being updated to improve the results (NOAA 2016).

5.2.1.2 Surface Imperviousness

The Surface Imperviousness data was obtained online from MRLC. The information used for the data quality analysis includes an XML file included in the dataset and the website that the data was obtained from.

Table 12. Data quality analysis of surface imperviousness – Duluth, Minnesota

Data Quality	Minnesota Surface Imperviousness
Completeness	The completeness of the data is not guaranteed by the U.S. Geological Survey as it is considered provisional (MRLC n.d.) The data will be classified as complete with the completion of a formal accuracy assessment (MRLC n.d.). The dataset provided imperviousness for the entirety of the contiguous states.
Security	The Surface Imperviousness data was open sourced and provided by the MRLC consortium consisting of federal agencies. There are no known security measures on this data.
Timeliness	The data is updated every 5 years by the MRLC.
Relevance	The relevance of the data for this data is exact as the ISAT was developed with this dataset. In addition, the ISAT Tutorial directs users to the MRLC website for data access (NOAA 2013).
Validity	According to the metadata available online the spatial information may be invalid due to the possible compilation of outside sources that may not meet the National Map Accuracy Standards (MRLC n.d.). USGS also indicates in the metadata that the interpretation of the results is up to the user and not the responsibility of USGS (MRLC 2016).
Accuracy	The accuracy of the imperviousness data was not formally assessed therefore the can is not guaranteed correct by the providing organization (MRLC 2016). However, the past results received an overall agreement between 71% and 97% thus proving the robustness of the data (MRLC 2016). The data itself is at a 30-metre pixel resolution which is used to calculate the surface imperviousness (MRLC 2019).

5.2.1.3 Population Statistics

The United States Population Statistics is available online at American FactFinder which provides open sourced census data. The information for data quality is provided through a Technical Document for the *2010 Census of Population and Housing* (USCB 2012).

Table 13. Data quality analysis of population statistics – Duluth, Minnesota

Data Quality	Minnesota Population Statistics
Completeness	Data is complete and available by category and region. Census data was available for all census tracts required and provides additional statistical information for each census tract including population density.
Security	The US census data was modified to protect confidentiality (USCB 2012:7-6). This wass achieved by: Title 13 U.S. Code, Disclosure Avoidance, and Data Swapping (USCB 2012:7-6). Security is ultimately achieved by disclosure avoidance by the method of data swapping where data with similar characteristics occur are swapped to protect the confidentiality of individuals (USCB 2012:7-6). The USCB has legislation set to protect individuals from release of confidential census data, this is the Title 13 U.S. Code (USCB 2012:7-6). The data available to the public is freely accessible however, security was maintained to provide modified data that is not identifiable to specific individuals.
Timeliness	The USA census is conducted every 10 years. The dataset itself is a compilation of extensive data collected which takes multiple years to complete. According to a USCB document of release dates the latest information was released was 2014 thus taking approximately 4-5 years to complete the dataset (USCB 2015).
Relevance	The relevance of the data for this dataset is exact as the ISAT was developed with this type of data. However, it is necessary to link the census data with the census tract data to be used for ISAT.
Validity	The data was prepared by the USCB which follows extensive methodologies and policies to develop the dataset thus validating the data.

Accuracy	Due to data swapping for security reasons the individual areas within a geographic area may not be accurate however according to the USCB there "no effect on the marginal totals for the geographic area with a small population" (USCB 2012:7-6). It is also expected that there are non-sampling errors that occur from nonresponses, respondent and enumerator error, and processing error (USCB 2012:7-7). There are methods used to reduce these errors however it does not result in a completely accurate census.

5.2.1.4 Census Tract Boundaries

The Census Tract Boundary data was obtained from United States Census Bureau online. The information on data quality was included within the dataset through the metadata and additional information on the data source website was used to identify the timeliness of the dataset.

Table 14. Data quality analysis of census tract boundaries – Duluth, Minnesota

Data Quality	Minnesota Census Tract Boundaries
Completeness	The dataset contains Census Tract boundaries for the St. Louis County which provides all census tracts for the study area. This file is an independent dataset from the greater Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) Database (MTDB) (USCB 2016).
Security	The TIGER/Line Shapefile for the census tracts was publicly available and not copyrighted however, it is a registered trademark of the USCB (USCB 2016). The data must be sourced to the USBC to be used in a product or publication (USCB 2016).
Timeliness	Upon review of available Census tract data online, it was determined that the data is updated regularly and usually yearly commencing on January 1st and typically released by the third quarter of the year (USCB n.d.)
Relevance	The relevance of the data is exact as the ISAT was developed with this type of dataset. However, data must be linked with the population statistics data.

Validity	The data was prepared by the USCB which follows extensive methodologies and policies to validate the data. Codes used by the dataset are taken from the "National Standard Codes (ANSI INICITS 38-2009), Federal Information Processing Series (FIPS) - States" (USCB 2016). The delineation of the census tracts was defined by local participants involved in the 2010 Census Participant Statistical Areas Program (USCB 2016).
Accuracy	According to the Shapefile metadata the positional accuracy is not to the standards that the six decimal places that the coordinate system suggests (USCB 2016). The boundaries are strictly for statistical purposes and does not contain any legal authority (USCB 2016).

5.2.1.5 Watershed Boundaries

The Minnesota watershed boundaries database contains multiple datasets which includes shapefiles of different levels of watersheds. The data quality was based on the selected watershed level metadata and the database metadata.

Table 15. Data quality analysis of watershed boundaries – Duluth, Minnesota

Data Quality	Minnesota Watershed Boundaries
Completeness	Tested for completeness (MNDNR 2009) and provides complete polygons for the study area.
Security	The data was unrestricted (MNDNR 2009). The publisher of the data includes a User Agreement where the data must be adequately sourced following the specific requirements (MNDNR 2009)
Timeliness	The data undergoes maintenance and updates yearly and represents the HUC level 7 from 1998 to the present (MNDNR 2009)
Relevance	The relevance of this data good. The ISAT tutorial guides users on how to use the tool with similar data.
Validity	The dataset was defined and referenced according to 103G.005 Subd. 17a and 103B.305 Subd. 10 Minnesota Statutes for the state watershed map production (MNDNR 2009).

Accuracy	There are no quantitative studies to determine the positional accuracy of the data (MNDNR 2009). However, the accuracy was estimated to be with 30-metres positional accuracy (MNDNR 2009).

5.3.1 Thunder Bay, Ontario Data Quality Analysis

5.3.1.1 Land Cover Classification

The Land Cover Classification data quality information for Thunder Bay, ON, was obtained from three different PDF's regarding each land cover database that is used to develop the provincial land cover database. This information is provided within the Ontario Land Cover dataset.

Table 16. Data quality analysis of land cover classification - Thunder Bay, Ontario

Data Quality	Ontario Land Cover Classification
Completeness	The land cover dataset was the result of combining three separate land cover databases to create a single dataset that covers the entirety of Ontario (OMNR 2014a).
Security	This data was recently made publicly available on Ontario GeoHub in the Autumn of 2019. Previously, the data was only made available to Ministry employees where the data was obtained upon request.
Timeliness	The land cover data is not recent and data dates to the mid-1990's where data was collected for the Southern Ontario Land Resource Information System Version 1.2 (Spectranalysis Inc. 2004). Each data set that is used also dates from a different point in time. Of which were combined into a single database. No data has been updated.
Relevance	The land cover data is relevant for the purpose of its intended use. However, the results may not be as detailed on the level of impact of imperviousness in urban areas as all anthropogenic impervious surface area was classified into one category.

Validity	The validity of the data is complex as the dataset is a compilation of three separate databases. However, the Far North Land Cover Data Specifications Version 1.4 indicates that the data was developed by the Ministry of Natural and Resources and Forestry and was classified as outlined by Spectranalysis Inc. (OMNRF 2014). It is unknown if the Southern Ontario Land Resource Information System (SOLRIS) was consistent with Spectranalysis as well. Therefore, the validity of the data cannot be accounted for.
Accuracy	Each database was re-sampled to pixel spacing of 15 metres, re-projected to NAD83 Lambert Conformal Conic, and re-classified to maintain consistency between the three different databases (OMNRF 2014). The data is inconsistent with date of data acquisition.

5.3.1.2 Surface Imperviousness

There was no previously developed impervious surface data available for Thunder Bay, ON. This was confirmed with an MNR employee in Thunder Bay, ON who works with GIS spatial systems. Data was developed for the study and the data quality analysis is based on the results of the data layer development and the provided imagery for the data development.

Table 17. Data quality analysis of surface imperviousness - Thunder Bay, Ontario.

Data Quality	Ontario Surface Imperviousness
Completeness	The provided imagery does not encompass the extent of the City of Thunder Bay thus the data is not complete.
Security	The imagery obtained was not publicly available. No security measures are placed on the developed dataset.
Timeliness	Imagery was collected in 2008 according to Alex Bilyk (pers. comm., March 30, 2020)
Relevance	The developed dataset has poor relevance however, it is satisfactory and can be used for the ISAT Basic Analysis.

Validity	No pre-defined procedures were followed.
Accuracy	The accuracy is based on a 5000-metre resolution. No accuracy assessment is completed.

5.3.1.3 Population Statistics

Population Statistics data was provided online by the Government of Canada through Statistics Canada. The relevant information was obtained from online Statistics Canada sources and through biased interpretation of the data. The data was collected individually by searching the CTUID thus resulting in a developed dataset and no provision of metadata.

Table 18. Data quality analysis of population statistics - Thunder Bay, Ontario.

Data Quality	Ontario Population Statistics
Completeness	The data is complete however individually separated. The completeness of the data was affected by security. The required data was provided online on individual site pages by searching each individual census tract unit ID number which must be identified by the user from another source.
Security	The data had security placed by altering the population value by 5 individuals for the total population in some census tracts (StatCan 2017c). Each census tract in the census division adds to the total amount in the census division (StatCan 2017c).
Timeliness	The Canadian census is conducted every 5 years (StatCan 2020). The most recent available census provided is 2016 (StatCan 2020).
Relevance	The relevance of the data format is poor. The data required was to be individually collected and inputted into a table to applied to the ISAT. The alternative was to normalize a large table with unnecessary data. However, the data relevance is good in terms of provided information.
Validity	The data was collected by the Canadian Government which follows extensive methodologies and policies to validate the data.

Accuracy	The Ontario census data's accuracy is based on the 98.6% response rate (StatCan 2017b). The accuracy of the data for the end user is within 5 individuals per census tract population value. The overall accuracy was impacted by the security of the data.

5.3.1.4 Census Tract Boundaries

Census Tract Boundaries are also provided online by the Government of Canada through Statistics Canada. Information regarding Canadas census tracts was provided in the Boundary Files, Reference Guide (StatCan 2017d).

Table 19. Data quality analysis of census tract boundaries - Thunder Bay Ontario.

Data Quality	Ontario Census Tract Boundaries
Completeness	In coastal areas there may be multiple polygons for a single geographic area due to the removal of coastal water polygons from the digital boundary file (StatCan 2017d). There were no other significant concerns.
Security	The data was publicly available on Statistics Canada.
Timeliness	The reference date for the Census Tract data is as of January 1, 2016, which is the same year the census was conducted.
Relevance	The relevance closely matches the preferred conditions for the census tract data. As tracts are based on the delineation of boundaries which are based on the political, demographic, natural, infrastructure, and population boundaries of the tracts are similar in size as used by the ISAT.
Validity	The data was developed for the purposes of both the National Geographic Database and Statistics Canada-Elections Canada initiative (StatCan 2017d). It is assumed that appropriate methods are carried out to validate the data however, it is not stated in the Boundary Files, Reference Guide.

Accuracy	The positional accuracy of the data is not precise and not intended to be used for legal purposes. The data involved was from multiple sources with different scales which were rescaled for the boundary files (StatCan 2017d). The attribute accuracy was verified against Spatial Data Infrastructure data and was deemed accurate (StatCan 2017d).

5.3.1.5 Watershed Boundaries

Watershed shapefile boundaries were obtained from Lakehead Region Conservation Authority through contact with an employee. The dataset does not provide any information, however the employee did indicate some information.

Table 20. Data quality analysis of watershed boundaries - Thunder Bay, Ontario.

Data Quality	Ontario Watershed Boundaries
Completeness	The data is incomplete as there is not watershed boundaries encompassing the entirety of the city.
Security	No available information.
Timeliness	No available information.
Relevance	The data is the most relevant and available data for a small scale watershed dataset to understand the effect of ISA.
Validity	The data is refined for the purposes of LRCA where watershed monitoring occurs. It is assumed that the refined watershed polygons are more appropriate for the City of Thunder Bay.
Accuracy	According to a LRCA employee the data based on a 20-metre resolution DEM layer (in email, Mar. 6, 2020).

5.3 APPLICATION ANALYSIS

For both Thunder Bay, ON, and Duluth, MN the Basic Analysis was executed using the city specific data and obtained the following results. The results show the impact of imperviousness on water quality within watershed boundaries based on the

developed High, Medium and Low Imperviousness Coefficients (Appendix B) created from the land cover classification data and population density (High, Medium and Low).

5.3.1 Thunder Bay, ON, ISAT Basic Analysis

The Imperviousness Coefficients (IC) that the Basic Analysis is based on was sourced from the population density for the City of Thunder Bay, ON. The city contains five census tracts where the population density is High in the North and South cores of the city as per the pre-defined thresholds. In addition the North and South cores of the City which results from the amalgamation of Fort William and Port Arthur was evident the form of population density distribution Figure 8. There are two clusters of medium-density and high-density tracts separated by low-density.

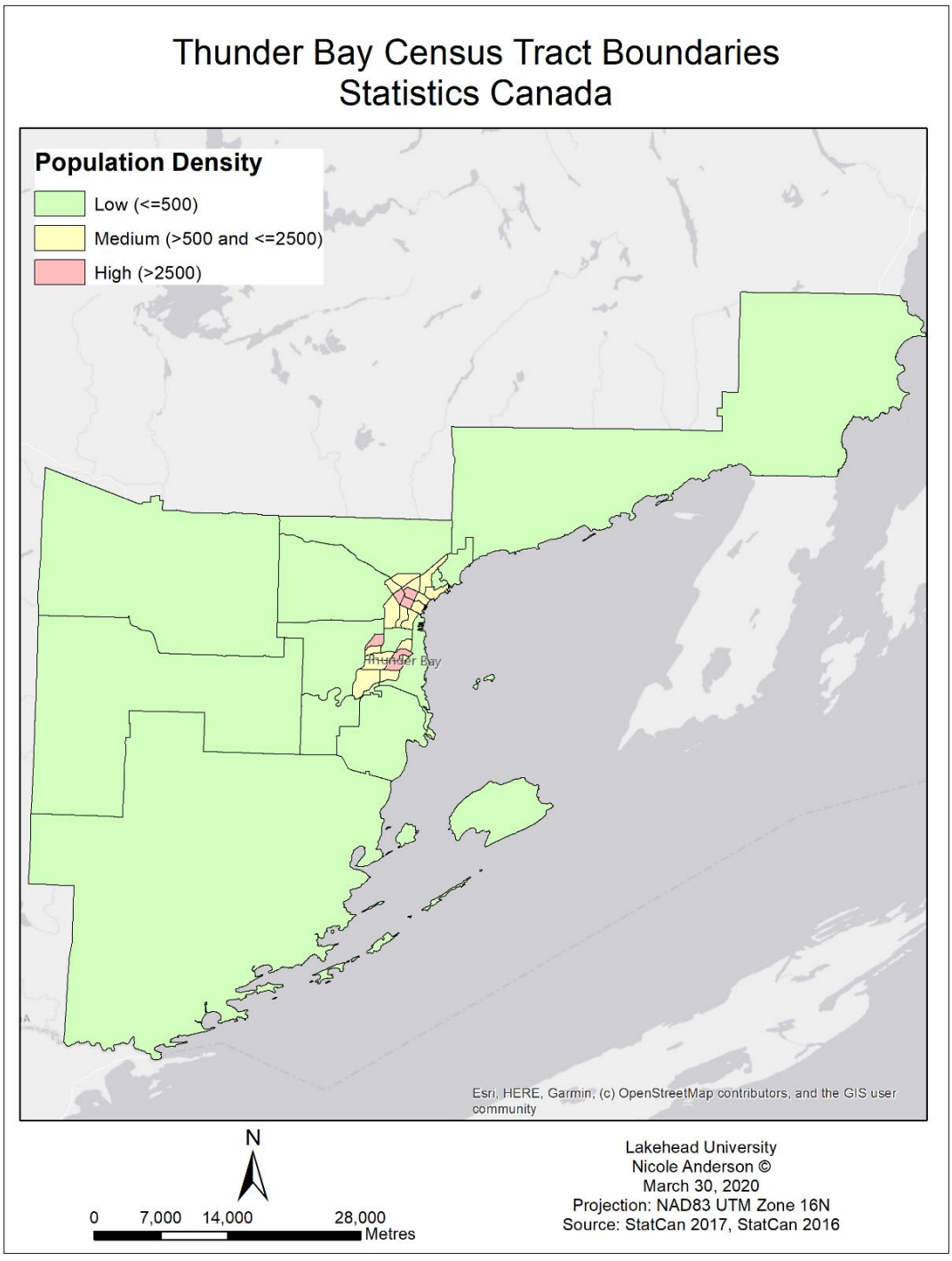


Figure 8. Map of Thunder Bay, ON Census Tracts and corresponding population density.

ICs were primarily identified for the following land classifications: Clear Open Water, Bog, Sparse Treed, Deciduous Treed, Mixed Treed, Coniferous Treed and Community Infrastructure (Table 21). In addition, the Low IC was also identified for Fen, Disturbance, and Agriculture and Undifferentiated Rural Land Use land classifications (Table 21). ICs in the High-density population census tracts are in the impacted (>25% ISA) and degraded (10-25%) thresholds of potential impact on water quality. Most Medium-density population census tracts are of the degraded threshold except for the Community/Infrastructure land classification which is impacted. Lastly, the Low-density population census tracts are of the protected threshold (<10%) except for Agriculture and Undifferentiated Rural Land Use land classification which is degraded.

Table 21. Thunder Bay, Ontario Impervious Surface Analysis Imperviousness Coefficients.

Value	Land Classification Name	High	Med	Low
0	Unclassified	0.00	0.00	0.00
1	Clear Open Water	33.47	17.25	2.54
2	Turbid Water	0.00	0.00	0.00
3	Shoreline	0.00	0.00	0.00
4	Mudflats	0.00	0.00	0.00
5	Marsh	0.00	0.00	0.00
6	Swamp	0.00	0.00	0.00
7	Fen	0.00	0.00	1.56
8	Bog	7.00	7.00	2.75
10	Heath	0.00	0.00	0.00
11	Sparse Treed	29.32	13.76	2.46
12	Treed Upland	0.00	0.00	0.00
13	Deciduous Treed	23.64	13.93	2.20
14	Mixed Treed	23.87	11.10	2.24
15	Coniferous Treed	25.20	11.73	3.73
16	Plantations - Treed Cultivated	0.00	0.00	0.00

17	Hedge Rows	0.00	0.00	0.00
18	Disturbance	0.00	0.00	1.35
19	Cliff and Talus	0.00	0.00	0.00
20	Alvar	0.00	0.00	0.00
21	Sand Barren and Dune	0.00	0.00	0.00
22	Open Tallgrass Prairie	0.00	0.00	0.00
23	Tall Grass Savannah	0.00	0.00	0.00
24	Tall Grass Woodland	0.00	0.00	0.00
25	Sand/Gravel/Mine Tailings/Extraction	0.00	0.00	5.02
26	Bedrock	0.00	0.00	0.00
27	Community/Infrastructure	33.50	25.86	16.70
28	Agriculture and Undifferentiated Rural Land Use	0.00	0.00	2.50

There was a significant difference in IC values where the difference results in different thresholds of effect of imperviousness. These differences are identified in Table 22. Most notably IC differences for Clear Open Water and Community/Infrastructure results in each population density consisting of a different threshold effect of imperviousness.

Table 22. Difference between Imperviousness Coefficients between population densities – Thunder Bay, Ontario.

Value	Land Classification Name	High	Medium to High Difference	Medium	Low to Medium Difference	Low
0	Unclassified	0	0	0	0	0
1	Clear Open Water	33.47	-16.22	17.25	-14.71	2.54
2	Turbid Water	0	0	0	0	0
3	Shoreline	0	0	0	0	0
4	Mudflats	0	0	0	0	0
5	Marsh	0	0	0	0	0
6	Swamp	0	0	0	0	0
7	Fen	0	0	0	1.56	1.56
8	Bog	7	0	7	-4.25	2.75
10	Heath	0	0	0	0	0

11	Sparse Treed	29.32	-15.56	13.76	-11.3	2.46
12	Treed Upland	0	0	0	0	0
13	Deciduous Treed	23.64	-9.71	13.93	-11.73	2.2
14	Mixed Treed	23.87	-12.77	11.1	-8.86	2.24
15	Coniferous Treed	25.2	-13.47	11.73	-8	3.73
16	Plantations - Treed Cultivated	0	0	0	0	0
17	Hedge Rows	0	0	0	0	0
18	Disturbance	0	0	0	1.35	1.35
19	Cliff and Talus	0	0	0	0	0
20	Alvar	0	0	0	0	0
21	Sand Barren and Dune	0	0	0	0	0
22	Open Tallgrass Prairie	0	0	0	0	0
23	Tall Grass Savannah	0	0	0	0	0
24	Tall Grass Woodland	0	0	0	0	0
25	Sand/Gravel/Mine Tailings/Extraction	0	0	0	5.02	5.02
26	Bedrock	0	0	0	0	0
27	Community/Infrastructure	33.5	-7.64	25.86	-9.16	16.7
28	Agriculture and Undifferentiated Rural Land Use	0	0	0	2.50	2.50

The ICs are evident in the results of the Basic Analysis. The Thunder Bay, ON Basic Analysis shows significant effect on water quality based on the level of ICs (Figure 9). The Low ICs indicated that the city water quality is relatively protected however with increase of population there is a significant increase of effect on water quality.

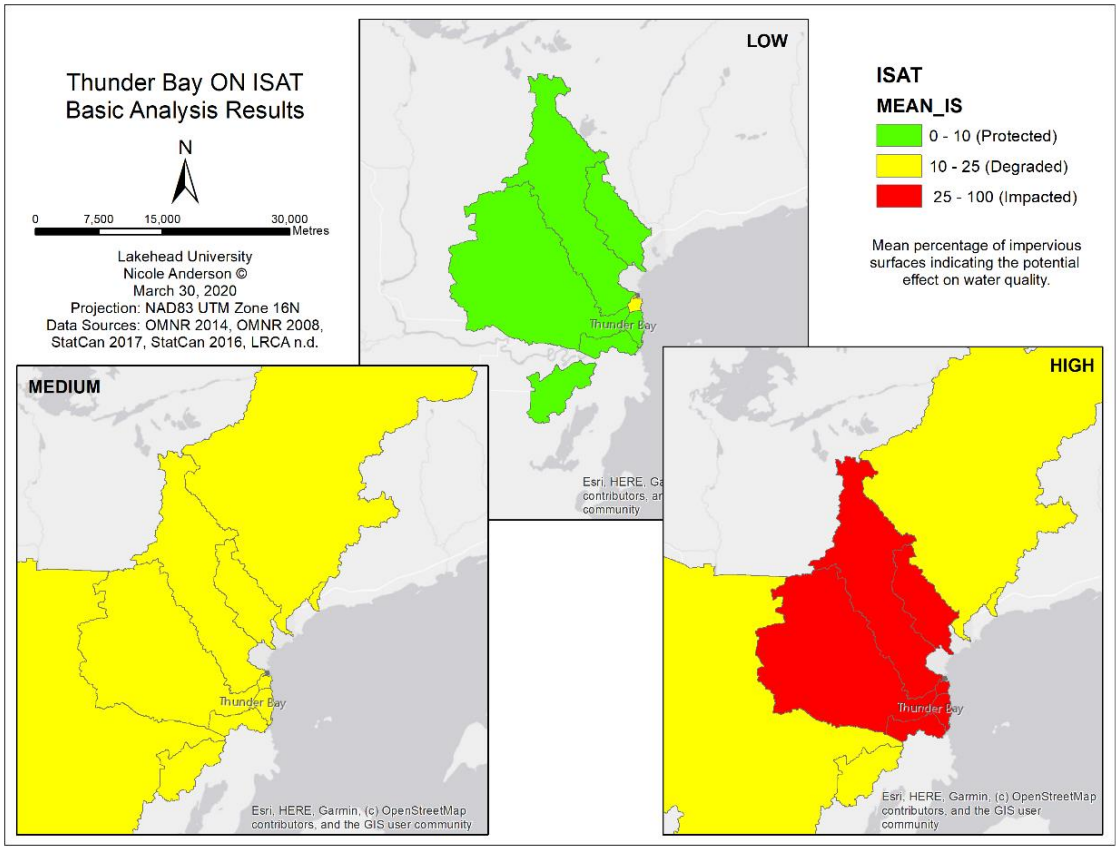


Figure 9. Map of Thunder Bay, Ontario Impervious Surface Analysis Tool Basic Analysis Results.

5.3.2 Duluth, MN, ISAT Basic Analysis

The Duluth, MN study area contains a greater number of census tracts and more census tracts with a population density greater than 500. There is one core of six high density census tracts with the remaining of the census tracts classified as medium along the shore and surrounding the core and low density is present inland. This is evident in Figure 10.

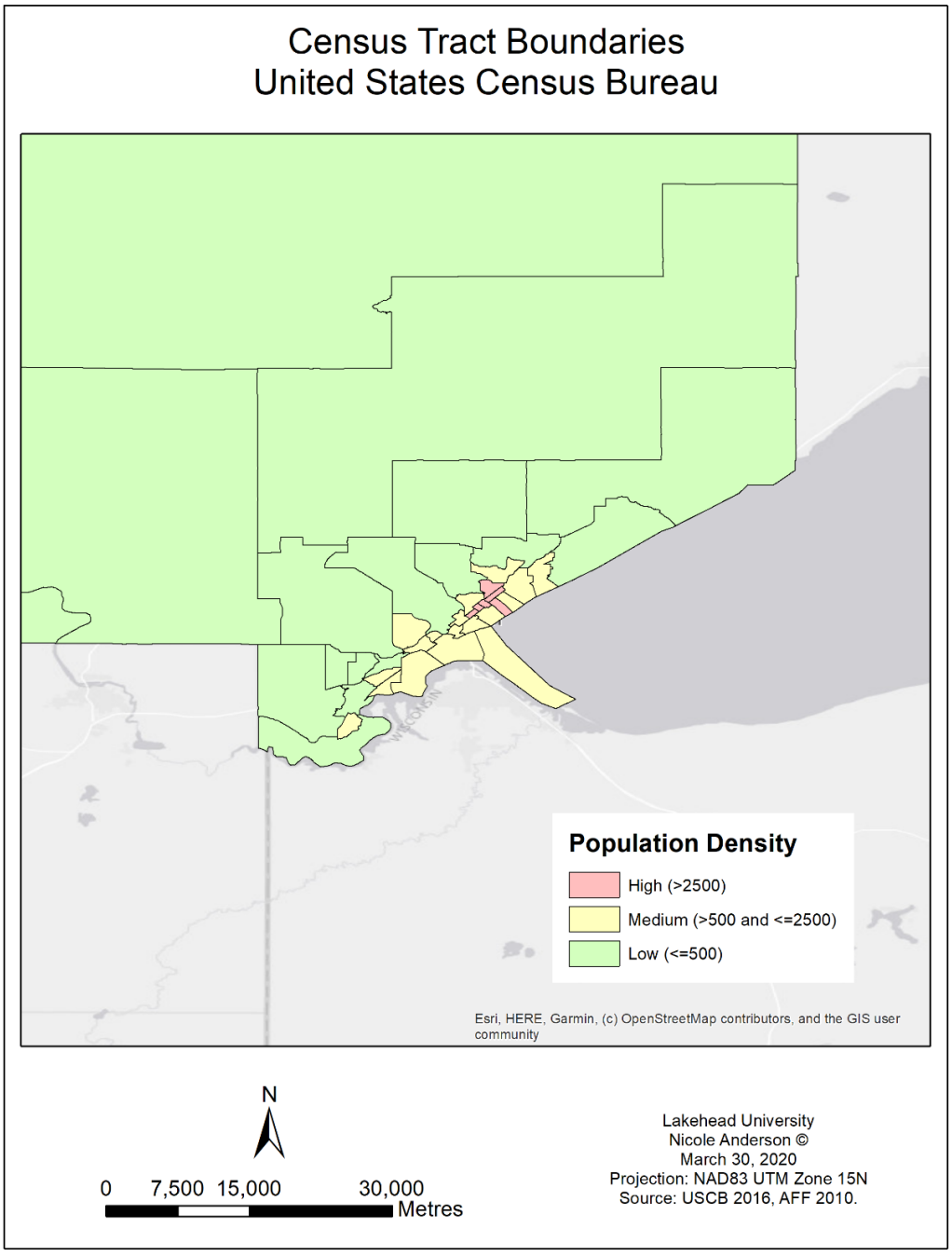


Figure 10. Map of Duluth, Minnesota Census Tracts and corresponding population density.

ICs were identified for all classifications except three which are of Estuarine wetland environments (Table 23). Only the following classifications have a significant IC greater than 1%: Developed High Intensity, Developed Medium Intensity, Developed Low Intensity, Developed Open Space, Unconsolidated Shore and Bare Land. Both Developed High Density and Developed Medium Density are in the Impacted threshold at each population density. Developed Low Intensity and Developed Open Spaces are both in the Degraded threshold while the rest of the land classifications are in the Protected threshold except for Low density Open Water land classification. Open Water uniquely has an Impacted threshold of effect of imperviousness.

Table 23. Duluth, Minnesota Impervious Surface Analysis Tool Imperviousness Coefficients.

Value	Land Classification Name	High	Med	Low
0	Unclassified	0.00	0.00	0.00
2	Developed, High Intensity	67.40	56.64	49.97
3	Developed, Medium Intensity	40.15	32.94	30.36
4	Developed, Low Intensity	24.34	22.15	21.59
5	Developed, Open Space	14.41	13.28	13.10
6	Cultivated Crops	0.34	0.40	0.45
7	Pasture/Hay	0.31	0.35	0.35
8	Grassland/Herbaceous	0.59	0.59	2.68
9	Deciduous Forest	0.15	0.15	0.99
10	Evergreen Forest	0.30	0.23	6.59
11	Mixed Forest	0.39	0.19	6.97
12	Scrub/Shrub	0.29	0.30	0.85
13	Palustrine Forested Wetland	0.06	0.05	0.11
14	Palustrine Scrub/Shrub Wetland	0.08	0.08	0.14
15	Palustrine Emergent Wetland	0.11	0.11	0.36
16	Estuarine Forested Wetland	0.00	0.00	0.00
17	Estuarine Scrub/Shrub Wetland	0.00	0.00	0.00
18	Estuarine Emergent Wetland	0.00	0.00	0.00
19	Unconsolidated Shore	7.93	0.52	1.17

20	Bare Land	6.59	2.34	2.30
21	Open Water	0.06	0.02	11.10
22	Palustrine Aquatic Bed	0.00	0.00	0.00

There is no significant difference between ICs for High, Medium and Low density population that results in a new classification effect of imperviousness on water quality. However, the Open Water land classification uniquely classifies as Degraded only in the Low density and higher ICs for Deciduous Forest and Evergreen Forest (Table 24). The effects of the ICs are evident in the ISAT Basic Analysis Results in Figure 11. *Map of Duluth, Minnesota, Impervious Surface Analysis Tool Basic Analysis Results*

Table 24. Difference between Imperviousness Coefficients between population densities – Duluth, Minnesota.

Value	Land Classification Name	High	Medium to High Difference	Medium	Low to Medium Difference	Low
0	Unclassified	0	0	0	0	0
2	Developed, High Intensity	67.40	-10.76	56.64	-6.67	49.97
3	Developed, Medium Intensity	40.15	-7.21	32.94	-2.58	30.36
4	Developed, Low Intensity	24.34	-2.19	22.15	-0.56	21.59
5	Developed, Open Space	14.41	-1.13	13.28	-0.18	13.10
6	Cultivated Crops	0.34	0.06	0.40	0.05	0.45
7	Pasture/Hay	0.31	0.04	0.35	0.0	0.35
8	Grassland/Herbaceous	0.59	0.00	0.59	2.09	2.68
9	Deciduous Forest	0.15	0.00	0.15	0.84	0.99
10	Evergreen Forest	0.30	-0.07	0.23	6.36	6.59
11	Mixed Forest	0.39	-0.20	0.19	6.78	6.97
12	Scrub/Shrub	0.29	0.01	0.30	0.55	0.85
13	Palustrine Forested Wetland	0.06	-0.01	0.05	0.06	0.11
14	Palustrine Scrub/Shrub Wetland	0.08	0	0.08	0.06	0.14
15	Palustrine Emergent Wetland	0.11	0	0.11	0.25	0.36
16	Estuarine Forested Wetland	0	0	0	0	0
17	Estuarine Scrub/Shrub Wetland	0	0	0	0	0

18	Estuarine Emergent Wetland	0	0	0	0	0
19	Unconsolidated Shore	7.93	-7.41	0.52	0.65	1.17
20	Bare Land	6.59	-4.25	2.34	-0.04	2.30
21	Open Water	0.06	-0.04	0.02	11.08	11.10
22	Palustrine Aquatic Bed	0	0	0	0	0

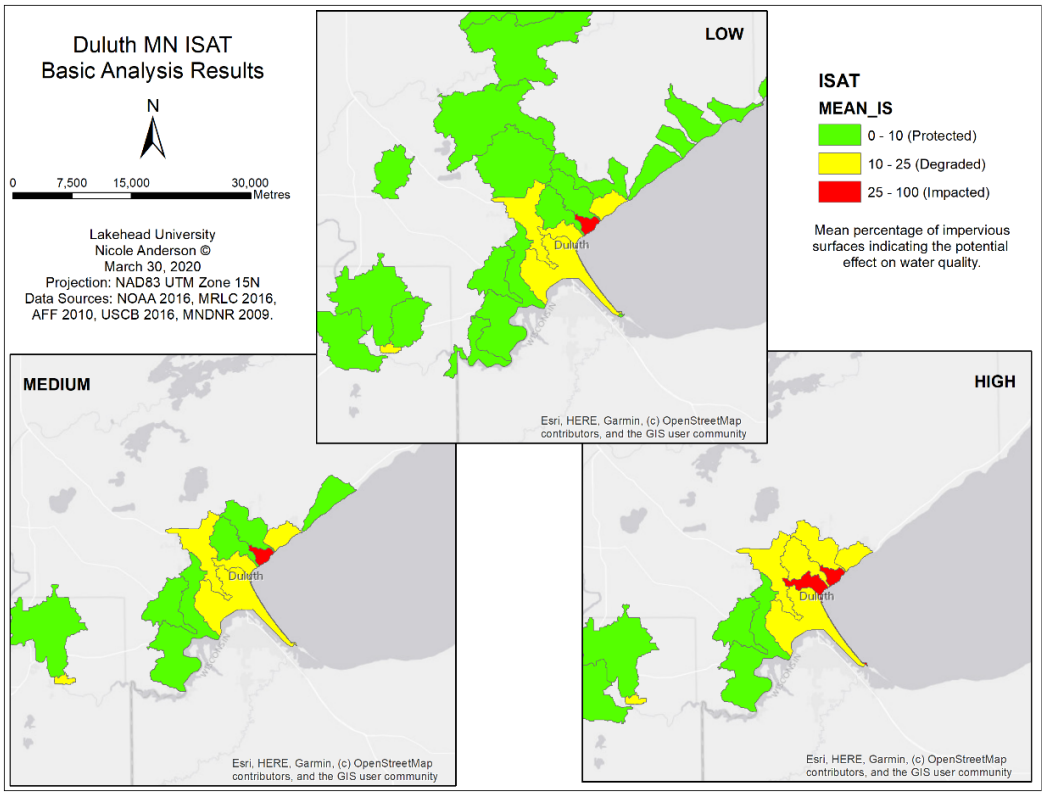


Figure 11. Map of Duluth, Minnesota, Impervious Surface Analysis Tool Basic Analysis Results.

6.0 DISCUSSION

6.1 DATA ACCESSIBILITY

Data accessibility is significantly different between Thunder Bay, ON and Duluth, MN. The USA provided data fully accessible online and available to the public. Canadas access was more restricted noting that data had become increasingly available.

While collecting data for Duluth, MN the only challenge posed to collection of data was the ability to navigate more complex databases and understanding differing terminology depicted by the providing organization.

The same barrier was posed for Thunder Bay, ON data collection. However, it was determined that data accessibility for Thunder Bay, ON was even more restricted. There is one dataset that was not developed, and two datasets where the providing organizations required direct contact to obtain the data. Fortunately, the data was provided freely upon request.

The Land Cover Classification data specifically highlighted the increasing accessibility of data resources online in Ontario. At the beginning of the study in September 2019 the Land Cover Classification data for Ontario was previously inaccessible online and was accessed by contacting a Ministry of Natural Resources employee. Shortly afterwards, in December 2019, it was determined that the data became accessible online through the Ontario GeoHub website. This shows the movement towards advancing science and conforming to Canadas new Open Data initiative.

The Surface Imperviousness data also highlighted the limitations of available data for Thunder Bay, ON. As the data could not be obtained, previously acquired imagery was used to develop the data. With this, two points that were brought to attention during development the dataset. First, the process had been streamlined for both surface imperviousness and the land cover classification on a five-year interval for the USA. Yet, Ontario and Canada have not developed a streamlined process. Secondly, accessibility to high quality imagery at a small resolution is a barrier to the development of high accuracy data. Yet again when the data acquisition had been streamlined for the USA.

Other restrictions to data accessibility for Thunder Bay, ON includes the Population Statistics data and the provisional format which required the individual search of each CTUID or the normalization of an extensive dataset. And lastly, the watershed boundaries were provincially limited to Quaternary Watersheds which are further divided into smaller sized watersheds by other organizations. The data used for this study was from the LRCA which derived their watersheds from the provincial dataset however, due to the organization specific purposes this data is incomplete. A trade-off with size and completion was made to use this data resource.

6.2 QUALITY ANALYSIS

The data quality analysis was largely based on the accompanying information provided by the organization and its application to the ISAT tool. The primary source of information was within the datasets whether it was in the form of metadata, XML file, or PDF file. However, not all datasets contained information pertaining to the quality of

the data. Some data quality information was obtained from the online source in the form of a document or directly on the providing organizations website.

Unfortunately, some datasets provided relatively poor and/or a lack of information to understand the quality of the data. This consequently resulted in the inability to successfully complete the quality analysis. As the quality analysis was strictly based on the provided information, contrast between the quality of information is significantly noticeable by the extent of information provided in the analysis.

It appears to be a general trend between Canadian and American data where the availability for data quality information was significantly poorer for Canadian data. This contrast was most significantly apparent for the land cover classification and population statistics data.

Within the quality analysis there was bias for both the Relevance and Completeness characteristics as it was based on the usability of the data in conjunction with the information provided by the organization. The remaining characteristics of data quality (Security, Timeliness, Validity, and Accuracy) remained relatively unbiased as they were based on the provided information. A side by side comparison of the data quality analysis for each type of data is available in Appendix A.

6.3.1 Relevance

The data quality analysis was ultimately biased as it is based on its intended use. Bias was most prominent for the Relevance characteristic as it determines its relevance to its use with the ISAT. The ISAT was developed in the USA thus using open

sourced data that is relevant to the USA specifically. This had an impact on the difference in relevance between Duluth, MN and Thunder Bay, ON.

In terms of the Duluth, MN analysis, all data was relevant and most particularly the Land Cover Classification and Surface Imperviousness data as the ISAT tutorial directs users to the providing organization website. The remaining data was also relevant; however some additional work is needed to link the population statistics excel file to the census tracts shapefile.

For Thunder Bay, ON, the relevance of the data was poorer which adds a greater challenge to utilize the ISAT. First and foremost, there was no pre-existing Surface Imperviousness data available for Thunder Bay, ON and the developed data is very poor in comparison.

Secondly, the land cover was good for its intended use, however will not produce desired results as land classifications are different for both cities. Most noticeably all urban areas were categorized into one classification as opposed to the land cover data for Duluth, MN which contains three classification for urban areas: high, medium, and low intensity development.

The strongest relevant data for Thunder Bay, ON was the Census Tract Boundaries and Population Statistics data as the information was based on similar census tract thresholds.

Lastly, Watershed data also has poor relevance as the watershed delineation was much larger than the recommended size as per the ISAT tutorial. While acceptable,

with smaller sized watershed polygons a stronger understanding of the impact of impervious surfaces can be derived.

6.3.2 Completeness

Bias was also noticeable in the Completeness of the data as it pertains to its completion of the data within the geographic coverage of the study area. Most of the datasets were a portion of a whole database which are separated by state or province. Land cover data for Duluth, MN and population statistics and census tracts for both Duluth, MN, and Thunder Bay, ON were all collected as a subset of the larger database that encompasses the whole country. This resulted in completion in terms of the study area, however if a user were analyzing areas on political boundaries multiple datasets would be required.

The Thunder Bay, ON Watershed Boundaries and Surface Imperviousness data was also not complete in terms of missing data. The provided Watershed Boundary data was not complete on the shore of Lake Superior in the core of the city. In regards to the Surface Imperviousness data developed for the City of Thunder Bay there was missing imagery that resulted in missing data tiles.

In addition to completeness, some providing organizations indicated that the data was not complete or may not be complete and was available under discretion of the user. For the Surface Imperviousness data for Duluth, MN, a formal accuracy assessment was required before being established as complete, however its past success accuracy assessments rendered the data adequate for use. The Thunder Bay, ON census tract data indicated that there may be error on coastal regions where there

may be multiple polygons for a single census tract due to the removal of coastal water polygons. No other completeness concerns were indicated by the provider.

6.3.2 Security

Highest security measures were placed on the Population Statistics data for both Duluth, MN and Thunder Bay, ON. Both include extensive measures to ensure the confidentiality of individuals. The main method for both countries was altering the number of individuals within the smaller sized census boundaries. Canada and USA both have different specifics on how to alter the data. For Canada, values were adjusted by 5 individuals and for America, values were swapped with other values that are based on similar characteristics. By changing the data values of smaller census boundaries public users are unable to successfully identify individuals thus maintaining confidentiality.

This security characteristics influenced the data quality characteristic for accuracy as described in section 6.3.3 Accuracy.

6.3.3 Timeliness

Timeliness had two components that impact the quality of data. First how often the dataset is updated with new data acquisition. Second was the timeliness from date of acquisition and its date of release.

For most datasets the data was updated regularly. The exception being the Ontario Land Cover Classification data which was acquired once. In addition, the acquisition dates for the southern, northern and Far North Ontario portions of the province were all collected at different points of time.

The only dataset that is updated more frequently for Thunder Bay, ON in comparison to Duluth, MN is the Population Statistics data from the Canadian census which is updated every 5 years as opposed to the American census occurring every 10 years.

Watershed data was a cohort for timeliness of data as the data is not continuously and rapidly changing. The Duluth, MN provincial dataset is updated yearly as required but does not undergo a re-evaluation of watershed boundaries. The Thunder Bay, ON data provider did not indicate whether the data is regularly updated, however it is likely only updated for the purposes of LRCA if or when boundaries are desired to be changed.

6.3.4 Validity

Validity was determined by collecting data from reputable sources. All datasets were valid except for the Ontario watershed and surface imperviousness data. The valid datasets were obtained directly from the federal government, provincial or state ministries, or federal partnership sources.

The data that falls under the exception were not identified as invalid, however it was developed for the purpose of the user. The watershed data was from the Lakehead Region Conservation Authority where water resource professionals delineated significant watersheds within the Thunder Bay region as per their needs and purposes. Secondly, the Thunder Bay, ON surface imperviousness data was created specifically for this study with the limitations set by the skills of the developer.

6.3.5 Accuracy

Accuracy of the datasets was a complex characteristic of data quality as each type of data has different requirements of accuracy depending on the type of data. In addition, different organizations provided different information to indicate the accuracy of the data.

With the information obtained the following was determined. Datasets completed by the federal government or through partnerships with the federal government had high data accuracy and high expectations to provide accurate and relevant data. The datasets that required the highest resolution of data is the Land Cover Classification and Surface Imperviousness data. This is provided at a 30 metre resolution for both datasets for Duluth, MN. The Thunder Bay, ON data provides a 15 metre Land Cover Classification and unfortunately the Surface Imperviousness data is at a 5000 metre resolution.

The accuracy of the Population Statistic data was affected by the security of the data. As previously mentioned, this data was altered however the impact was not significant and should not concern the user of the ISAT analysis when utilizing this data.

Uniquely the Census Tract Boundary data for both study areas indicated that the data was not geographically accurate and should not be used for statistical purposes. This did not significantly impact the study for either region.

6.3 APPLICATION ANALYSIS

The ISAT Basic Analysis applied the developed ICs to predict the effect of ISA on the local water quality. Each analysis contained different sources of data which affects

the quality of the data. This was evident when comparing the datasets visually in Appendix B. The quality of the Surface Imperviousness dataset has the greatest impact on the results as it was used to develop the ICs for the ISAT analysis.

Both studies had the presence of ICs for water-based land classifications. This was presently worse for the Thunder Bay, ON analysis, as Clear Open Water ICs were identified for High (33.47), Medium (17.25), and Low (2.54) densities. For Duluth, MN Open Water ICs were identified for High (0.06), Medium (0.02), and Low (11.10) population densities as well. It is undetermined as to the reason why the Low density areas had a significantly higher IC for Duluth, MN. However, the reasoning for ICs identified for water-based land classifications was the accuracy and resolution of the data. Surface Imperviousness data is an averaged value of imperviousness per pixel. The Thunder Bay, ON data was at a 5000-metre resolution which reasons as to why ICs are significantly large for the water-based land classification. It was expected with the Duluth, MN data at a 30-metre resolution, that the IC value would not exceed those from the Thunder Bay, ON results, however this was not the case for Duluth, MN Low density population. As this value is significantly larger than the other population densities this value was considered an anomaly.

The impact of the coarse Surface Imperviousness data for Thunder Bay, ON was likely the same for the remaining land classifications. The coarseness of the data did not represent the percentage of imperviousness correctly. This was present in tiles of the data where there is a small cluster of ISA among areas of pervious land and water

bodies. With improved data the impervious percentages can better represent the presence of ISA.

The impact of data quality and accessibility was also present for the Watershed Boundaries within the ISAT Basic Analysis Results. With the application of different watershed sizes, the importance of smaller watersheds was identified. The smaller sized Watershed Boundaries for Duluth, MN allowed greater interpretation of the results as the analysis displayed the results through effected polygons. Due to the large polygon sizes for Thunder Bay, ON the true impact of High, Medium, and Low densities could not be easily interpreted. This was due to the fact that the results were displayed in one large polygon as opposed to a selection of smaller polygons within the same area that is affected.

The High ICs for Thunder Bay, ON was more likely representative of the impact on water quality within the core of the city. However, outside of the core of the city the Low ICs result was likely more appropriate as the coarse Surface Imperviousness would affect the true results of imperviousness where there are clusters of imperviousness and the averaged percentage is not truly representative of the area.

Overall, the Duluth, MN results were adequate due to the ISAT tool design. The tutorial recommended users to specific resources to obtain the data which is available for the contiguous states. The Thunder Bay, ON results, however, has noticeable concerns that impact the legitimacy of the results.

With these complications the results of the application analysis for Thunder Bay, ON is rendered illegitimate until further improvement of quality data is made to more accurately utilize the ISAT tool.

7.0 CONCLUSION

This study sought to determine the effect of data quality and accessibility on the results of the ISAT analysis while comparing data quality and accessibility individually. The results determined that data for the most part is freely accessible in the USA and data is more restricted in Canada. This is evident from background research and with the data collected for the study. The American Federal Data Strategy moves towards improved and publicly accessible data which is evident given all data utilized is provided is of good quality and accessibility. With regards to Canadian data, barriers were evident early in the research. Despite the Canadian data restriction of specific organizations with data sharing agreements, this thesis showed a movement towards the Open Data initiative by an increase of data accessibility during the study.

This study also sought to predict the effect of ISA on water quality for both Thunder Bay, ON and Duluth, MN. The effect is determined to be relatively moderate for Duluth, MN with a greater impact for Thunder Bay, ON.

The results of the ISAT Basic Analysis indicated the requirement of high quality data that is accessible in order to advance the progress of science and research.

Upon completion of this study is it my recommendation for the City of Thunder Bay to invest in delineating higher quality data with a higher resolution to better understand the effect of imperviousness within the city. This will assist the water resource managers to appropriately zone the city and help EcoSuperior de-paving initiatives in Thunder Bay, ON.

LITERATURE CITED

[AFF] American FactFinder. 2010. Population, Housing Units, Area, and Density: 2010 - County - Census Tract. U.S. Census Bureau, 2010 Census. (Database).

Brabec E., S. Schulte and P.L. Richards. 2002. Impervious Surfaces and Water Quality: A Review of Current Literature and Its Implications for Watershed Planning. Journal of Planning Literature. 16(4):499:514.

[FGDC] Federal Geographic Data Committee. n.d. National Geospatial Data Asset (NGDA) Management Plan. The Federal Geographic Data Committee. https://www.fgdc.gov/initiatives/ngda-management-plan. Accessed February 8, 2020.

Gayathri K.D., B.P. Ganasri and G.S. Dwarakish. 2015. A Review on Hydrological Models. Science Direct. 4:1001-1007.

Google Earth n.d.a. Duluth Minnesota, USA. 46.79°N, 92.10°W, Camera 64km. NOAA. https://earth.google.com/web/@46.76498853,-92.11090395,182.60819275a,63696.24048192d,35y,0h,0t,0r/data=ChIaEAoIL20vMGgxazYYASABKAI. November 7, 2019.

Google Earth. n.d.b. Thunder Bay. 48.38°N, 89.25°W, Camera 62km. Maxar Technologies NOAA. https://earth.google.com/web/@48.39661822,-89.34199766,183.12604717a,66847.7625663d,35y,0h,0t,0r/data=ChMaEQoJL20vMDFnYl9wGAEgASgC. November 7, 2019.

[GovCan] Government of Canada. 2019. Open Data 101. https://open.canada.ca/en/open-data-principles. February 20, 2020.

[LRCA] Lakehead Region Conservation Authority. n.d. Watershed Assessment. (Database)

Li X., Z. Nan , G. Cheng , Y. Ding , L. Wu , L. Wang , J. Wang , Y. Ran , H. Li , X.Pan & Z. Zhu. 2011. Toward an improved data stewardship and service for environmental and ecological science data in West China. International Journal of Digital Earth. 4(4):347-359

[MNDNR] Minnesota Department of Natural Resources. 2009. MNDNR Watersheds - MN DNR Level 07 - HUC04. Minnesota DNR - Division of Waters - Watershed Delineation Project. Publication Date: 4/7/2009. (Database).

[MNDNR] Minnesota Geospatial Commons. n.d. MNDNR Watershed Suite. https://gisdata.mn.gov/dataset/geos-dnr-watersheds. January 18, 2020.

[MNRF] Ministry of Natural Resources and Forestry. 2020. Ontario Flow Assessment Tool. Queen's Printer for Ontario. https://www.gisapplication.lrc.gov.on.ca/OFAT/Index.html?site=OFAT&viewer=OFAT&locale=en-US. January 18, 2020

[MRLC] Multi-Resolution Land Characteristics Consortium. 2019. Data: Urban Imperviousness. https://www.mrlc.gov/data?f%5B0%5D=category%3Aurban%20imperviousness. October 21, 2019.

[MRLC] Multi-Resolution Land Characteristics Consortium. 2016. NLCD 2016 Percent Developed Imperviousness (CONUS). U.S. Geological Survey. (Database).

[MRLC] Multi-Resolution Land Characteristics Consortium. n.d. Metadata. https://www.mrlc.gov/downloads/sciweb1/shared/mrlc/metadata/NLCD_2016_Impervious_L48.xml. February 11, 2020.

[NOAA] National Oceanic and Atmospheric Administration, Office for Coastal Management. 2019. C-CAP Regional Landcover and Change. https://coast.noaa.gov/digitalcoast/data/ccapregional.html. October 31, 2019.

[NOAA] National Oceanic and Atmospheric Administration. 2016. NOAA's Coastal Change Analysis Program (C-CAP 2016 Regional Land Cover Data - Coastal United States. U.S. Department of of Commerce. (Database)

[NOAA] National Oceanic and Atmospheric Administration, Office for Coastal Management. 2013. Tutorial: Impervious Surface Analysis Tool (ISAT) for ArcGIS 10.x. 12pp. Retrieved from download https://coast.noaa.gov/digitalcoast/tools/isat.html. September 23, 2019.

[NOAA] National Oceanic and Atmospheric Administration, Office for Coastal Management. n.d.a. NLCD 2016 Percent Developed Imperviousness (CONUS). https://www.mrlc.gov/data/nlcd-2016-percent-developed-imperviousness-conus. December 16, 2019.

[NOAA] NOAA Office for Coastal Management. n.d.b. Frequent Questions: Coastal Change Analysis Program (C-CAP) Regional Land Cover. https://coast.noaa.gov/data/digitalcoast/pdf/ccap-faq-regional.pdf. Accessed February 9, 2020.

[NRCS] Natural Resources Conservation Service. 2007. Watersheds, Hydrologic Units, Hydrologic Unit Codes, Watershed Approach, and Rapid Watershed Assessments. United States Department of Agriculture. Retrieved from https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1042207.pdf November 4, 2019.

Ontario. 2019. Understanding Stormwater Management: An Introduction to Stormwater Management Planning and Design. Queen's Printer for Ontario. https://www.ontario.ca/page/understanding-stormwater-management-introduction-stormwater-management-planning-and-design. Feb 8, 2019.

Ontario. 2014. Land Cover Compilation Data Specifications Version 2.0. Queen's Printer for Ontario.

Ontario GeoHub. 2019. Ontario Land Cover Compilation v.2.0. Queen's Printer for Ontario. https://geohub.lio.gov.on.ca/datasets/ontario-land-cover-compilation-v-2-0. December 12, 2019.

[OMNR] Ontario Ministry of Natural Resources. 2014. Ontario Land Cover Compilation V.2. Science and Research Branch, Ministry of Natural Resources and Forestry. (Database)

[OMNR] Ontario Ministry of Natural Resources. 2014a. Ontario Land Cover Compilation Data Specifications Version 2.0. Queen's Printer for Ontario. 19pp.

[OMNR] Ontario Ministry of Natural Resources. 2008. Ontario Forest Resource Inventory - Term 1, cleaned imagery. (Database)

[OMNRF] Ontario Ministry of Natural Resources and Forestry. 2014. Far North Land Cover Data Specifications Version 1.4. Queen's Printer for Ontario. 33pp.

Pipino L.L., Y.W. Lee, and R.Y. Wang. 2002. Data Quality Assessment. Communications of the ACM. 45(4):211-218.

Schueler T. 1994. The Importance of Imperviousness. Watershed Protection Techniques. 1(3). Retrieved from http://pinelakedistrict.org/doc/resources/The%20Importance%20of%20Imperviousness.pdf. October 27, 2019.

Smith G.D. 1994. Increasing the accessibility of data: "See for yourself" should be the watchword. BMJ. 308:1519-1520.

SPECTRANALYSIS Inc. 2004. Introduction to the Ontario Land Cover Data Base, Second Edition (2000): Outline of Production Methodology and Description of 27 Land Cover Classes. Spectranalysis Inc. Remote Sensing and GIS Services. 34pp.

[StatCan] Statistics Canada. 2020. Census Program. https://www12.statcan.gc.ca/census-recensement/index-eng.cfm. February 26, 2020.

[StatCan] Statistics Canada. 2019a. Boundary Files. Retrieved from https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-eng.cfm. January 16, 2020.

[StatCan] Statistics Canada. 2019b. Data Sharing- Receiving Organizations. https://www.statcan.gc.ca/eng/about/accountability/receiving-organizations. October 25, 2019.

[StatCan] Statistics Canada. 2019c. Census Profile, 2016 Census. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CT&Code1=5950023.01&Geo2=PR&Code2=01&SearchText=5950023.01&SearchType=Begins&SearchPR=01&B1=All&TABID=3&type=0. January 10, 2020.

[StatCan] Statistics Canada. 2018a. Canada goes urban. https://www150.statcan.gc.ca/n1/pub/11-630-x/11-630-x2015004-eng.htm. October 27, 2019.

[StatCan] Statistics Canada. 2018b. The Canadian Population in 2011: Population Counts and Growth. https://www12.statcan.gc.ca/census-recensement/2011/as-sa/98-310-x/98-310-x2011001-eng.cfm. October 27, 2019.

[StatCan] Statistics Canada. 2018c. Census Tract (CT): Plain language definition. https://www150.statcan.gc.ca/n1/pub/92-195-x/2011001/geo/ct-sr/ct-sr-eng.htm. December 16, 2019.

[StatCan] Statistics Canada. 2018d. Census Profile, 2016 Census – Select Region 1. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/search-recherche/change-geo.cfm?Lang=E&Geo1=CT&Code1=5950023.01&Geo2=PR&Code2=01&SearchText=5950023.01&SearchType=Begins&SearchPR=01&B1=All&TABID=3&G=1&type=0. January 16, 2020.

[StatCan] Statistics Canada. 2017. Census Profile. 2016. Statistics Canada Catalogue no. 98-316-X2016001. Ottawa. Released November 29, 2017. (Database)

[StatCan] Statistics Canada. 2017a. Standard Drainage Area Classification (SDAC) 2003. https://www.statcan.gc.ca/eng/subjects/standard/sdac/sdacinfo1. January 18, 2020.

[StatCan]Statistics Canada. 2017b. 2016 Census of Population collection response rates. https://www12.statcan.gc.ca/census-recensement/2016/ref/response-rates-eng.cfm. February 29, 2020.

[StatCan] Statistics Canada. 2017c. 5950100.04 [Census tract], Ontario and Canada [Country] (table). Census Profile. 2016 Census. Statistics Canada Catalogue no. 98-316-X2016001. Ottawa. Released November 29, 2017.

[StatCan] Statistics Canada. 2017d. Boundary Files, Reference Guide, Second edition, 2016 Census. Statistics Canada Catalogue no. 92-160-G.

[StatCan] Statistics Canada. 2016. Census Tract Boundary File, 2016 Census. Statistics Canada Catalogue no. 92-168-X. (Database)

Tenopir C., S. Allard, K. Douglass, A. U. Aydinoglu, L. Wu, E. Read, M. Manoff, M. Frame. 2011. Data Sharing by Scientists: Practices and Perceptions. PLoS ONE. e21101. https://doi.org/10.1371/journal.pone.0021101.

Three Rivers District Council. 2008. Data Quality Strategy. Retrieved from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=20&ved=2ahUKEwizgu75vr3lAhUHVK0KHahxCdIQFjATegQIAxAC&url=https%3A%2F%2F

www.threerivers.gov.uk%2Fdownload%3Fid%3D30632&usg=AOvVaw0vJNdPer5vq406vdD5BvEr. October 27, 2019.

[USCB] United States Census Bureau. 2016. 2016 Tiger/Line Shapefiles: Census Tracts. US Census Bureau, Geography Division. (Database).

[USCB] United States Census Bureau. 2015. 2010 Census Data Products: United States - At a Glance (Version 2.8). 4pp.

[USCB] United States Census Bureau. 2012. 2010 Census Summary File 1 - Technical Documentation/prepared by the U.S. Census Bureau. 730pp.

[USCB] United States Census Bureau. 2010a. American Factfinder: Minnesota population data. https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml. October 16, 2019.

[USCB] United States Census Bureau. 2010b. 2010 TIGER/Line® Shapefile: Minnesota. https://www.census.gov/cgi-bin/geo/shapefiles/index.php. October 16, 2019.

[USCB] United States Census Bureau. n.d. Tiger/Line Shapefiles. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.2016.html. January 16, 2020.

[USAGov] United States Government. n.d. Federal Data Strategy: Leveraging Data as a Strategic Asset. https://strategy.data.gov/. October 21, 2019.

Wada Y., M.F.P. Bierkens, Ad de Roo, P.A. Dirmeyer, J.S. Famiglietti, N. Hanasaki, M. Konar, J. Liu, H.M. Schmied, T. Oki, Y. Pokhrel, M. Sivapalan, T.J. Troy, A.I.J.M. van Dijk, T. Emmerik, M.H.J. Van Huijgevoort, H.A.J. Van Lanen, C.J. Vörösmarty, N. Wanders, and H. Wheater. 2017. Human-water interface in hydrological modelling: current status and future directions. Hydrol. Earth Syst Sci. 21:4169-4193.

Zoppou C. 2000. Review of urban water models. Environmental Modelling & Software. 16:195-231.

APPENDIX A

Data Quality Characteristic	Minnesota Land Cover Classification	Ontario Land Cover Classification
Completeness	The Federal Geographic Data Committee Management Plan ensured that the data is 'sufficiently complete' for the requirements of the United States government (FGDC n.d.). The data was provided by the state level to reduce the size of the datafiles.	The land cover dataset was the result of combining three separate land cover databases to create a single dataset that covers the entirety of Ontario (OMNR 2014a).
Security	The land cover data was open sourced data provided by the NOAA Office for Coastal Management (NOAA 2019). There is no known security measures on the data.	This data was recently made publicly available on Ontario GeoHub in the Autumn of 2019. Previously, the data was only made available to Ministry employees where the data was obtained upon request.
Timeliness	According the XML file accompanied with the data file the Landsat imagery utilized for creating the land cover information was acquired either before or after the '2016-Era' (NOAA 2016). This is due to the inability to capture imagery for the whole extent within one year (NOAA n.d.b). The datasets were updated at 5-year intervals. According to NOAA within 5 years the most a region changes is 20% while vast areas are only changing by less than 6% (NOAA n.d.b).	The land cover data is not recent and data dates to the mid-1990's where data was collected for the Southern Ontario Land Resource Information System Version 1.2 (Spectranalysis Inc. 2004). Each data set that is used also dates from a different point in time. Of which were combined into a single database. No data has been updated.
Relevance	The relevance of the data is exact as the ISAT was developed with this dataset. In addition, the ISAT	The land cover data is relevant for the purpose of its intended use. However, the results may not be as detailed on the level of impact of imperviousness in

	Tutorial directs users to the NOAA for data access (NOAA 2013).	urban areas as all anthropogenic impervious surface area was classified into one category.
Validity	The validity of the land cover data is complied with the National Geospatial Data Asset (NGDA) Management Plan in accordance with the OMB (Office of Management and Budget) policy and Administration direction to ensure the reliability of the data (FGDC n.d.)	The validity of the data is complex as the dataset is a compilation of three separate databases. However, the Far North Land Cover Data Specifications Version 1.4 indicates that the data was developed by the Ministry of Natural and Resources and Forestry and was classified as outlined by Spectranalysis Inc. (OMNRF 2014). It is unknown if the Southern Ontario Land Resource Information System (SOLRIS) was consistent with Spectranalysis as well. Therefore, the validity of the data cannot be accounted for.
Accuracy	The FGDC NGDA ensures the accuracy of the land cover data. The NGDA management plan provides a framework and processes for the management of the data (FGDC n.d.). The goal of the FGDC is to provide robust and accessible geospatial datasets (FGDC n.d.). This ensures that the accuracy of the data is consistent. The data itself is analyzed at a 30-metre resolution (NOAA 2019). When updating datasets, only areas that experienced change in the landscape are reclassified (NOAA n.d.b). The change is identified by using a change detection tool which allows better efficiency and data consistency and reduces the cost of developing the product (NOAA n.d.b). The overall accuracy of the C-CAP land cover data is 85% (NOAA n.d.b). With the development of the data	Each database was re-sampled to pixel spacing of 15 metres, re-projected to NAD83 Lambert Conformal Conic, and re-classified to maintain consistency between the three different databases (OMNRF 2014). The data is inconsistent with date of data acquisition.

	there is significant quality control to ensure accuracy through training data that is edited by hand and is continuously being updated to improve the results (NOAA 2016).	

Data Quality Characteristic	Minnesota Surface Imperviousness	Ontario Surface Imperviousness
Completeness	The completeness of the data is not guaranteed by the U.S. Geological Survey as it is considered provisional (MRLC n.d.) The data will be classified as complete with the completion of a formal accuracy assessment (MRLC n.d.). The dataset provided imperviousness for the entirety of the contiguous states.	The provided imagery does not encompass the extent of the City of Thunder Bay thus the data is not complete.
Security	The Surface Imperviousness data was open sourced and provided by the MRLC consortium consisting of federal agencies. There are no known security measures on this data.	The imagery obtained was not publicly available. No security measures are placed on the developed dataset.
Timeliness	The data is updated every 5 years by the MRLC.	Imagery was collected in 2008 according to Alex Bilyk (pers. comm., March 30, 2020)
Relevance	The relevance of the data for this data is exact as the ISAT was developed with this dataset. In addition, the ISAT Tutorial directs users to the MRLC website for data access (NOAA 2013).	The developed dataset has poor relevance however, it is satisfactory and can be used for the ISAT Basic Analysis.
Validity	According to the metadata available online the spatial information may be invalid due to the possible compilation of outside sources that may	No pre-defined procedures were followed.

	not meet the National Map Accuracy Standards (MRLC n.d.). USGS also indicates in the metadata that the interpretation of the results is up to the user and not the responsibility of USGS (MRLC 2016).	
Accuracy	The accuracy of the imperviousness data was not formally assessed therefore the can is not guaranteed correct by the providing organization (MRLC 2016). However, the past results received an overall agreement between 71% and 97% thus proving the robustness of the data (MRLC 2016). The data itself is at a 30-metre pixel resolution which is used to calculate the surface imperviousness (MRLC 2019).	The accuracy is based on a 5000-metre resolution. No accuracy assessment is completed.

Data Quality Characteristic	Minnesota Population Statistics	Ontario Population Statistics
Completeness	Data is complete and available by category and region. Census data was available for all census tracts required and provides additional statistical information for each census tract including population density.	The data is complete however individually separated. The completeness of the data was affected by security. The required data was provided online on individual site pages by searching each individual census tract unit ID number which must be identified by the user from another source.
Security	The US census data was modified to protect confidentiality (USCB 2012:7-6). This wass achieved by: Title 13 U.S. Code, Disclosure Avoidance, and Data Swapping (USCB 2012:7-6).	The data had security placed by altering the population value by 5 individuals for the total population in some census tracts (StatCan 2017c).

	Security is ultimately achieved by disclosure avoidance by the method of data swapping where data with similar characteristics occur are swapped to protect the confidentiality of individuals (USCB 2012:7-6). The USCB has legislation set to protect individuals from release of confidential census data, this is the Title 13 U.S. Code (USCB 2012:7-6). The data available to the public is freely accessible however, security was maintained to provide modified data that is not identifiable to specific individuals.	Each census tract in the census division adds to the total amount in the census division (StatCan 2017c).
Timeliness	The USA census is conducted every 10 years. The dataset itself is a compilation of extensive data collected which takes multiple years to complete. According to a USCB document of release dates the latest information was released was 2014 thus taking approximately 4-5 years to complete the dataset (USCB 2015).	The Canadian census is conducted every 5 years (StatCan 2020). The most recent available census provided is 2016 (StatCan 2020).
Relevance	The relevance of the data for this dataset is exact as the ISAT was developed with this type of data. However, it is necessary to link the census data with the census tract data to be used for ISAT.	The relevance of the data format is poor. The data required was to be individually collected and inputted into a table to applied to the ISAT. The alternative was to normalize a large table with unnecessary data. However, the data relevance is good in terms of provided information.
Validity	The data was prepared by the USCB which follows extensive methodologies and policies to develop the dataset thus validating the data.	The data was collected by the Canadian Government which follows extensive methodologies and policies to validate the data.

Accuracy	Due to data swapping for security reasons the individual areas within a geographic area may not be accurate however according to the USCB there "no effect on the marginal totals for the geographic area with a small population" (USCB 2012:7-6). It is also expected that there are non-sampling errors that occur from nonresponses, respondent and enumerator error, and processing error (USCB 2012:7-7). There are methods used to reduce these errors however it does not result in a completely accurate census.	The Ontario census data's accuracy is based on the 98.6% response rate (StatCan 2017b). The accuracy of the data for the end user is within 5 individuals per census tract population value. The overall accuracy was impacted by the security of the data.

Data Quality Characteristic	Minnesota Census Tract Boundaries	Ontario Census Tract Boundaries
Completeness	The dataset contains Census Tract boundaries for the St. Louis County which provides all census tracts for the study area. This file is an independent dataset from the greater Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) Database (MTDB) (USCB 2016).	In coastal areas there may be multiple polygons for a single geographic area due to the removal of coastal water polygons from the digital boundary file (StatCan 2017d). There were no other significant concerns.
Security	The TIGER/Line Shapefile for the census tracts was publicly available and not copyrighted however, it is a registered trademark of the USCB (USCB 2016). The data must be sourced to the USBC to be used in a product or publication (USCB 2016).	The data was publicly available on Statistics Canada.

Timeliness	Upon review of available Census tract data online, it was determined that the data is updated regularly and usually yearly commencing on January 1st and typically released by the third quarter of the year (USCB n.d.)	The reference date for the Census Tract data is as of January 1, 2016, which is the same year the census was conducted.
Relevance	The relevance of the data is exact as the ISAT was developed with this type of dataset. However, data must be linked with the population statistics data.	The relevance closely matches the preferred conditions for the census tract data. As tracts are based on the delineation of boundaries which are based on the political, demographic, natural, infrastructure, and population boundaries of the tracts are similar in size as used by the ISAT.
Validity	The data was prepared by the USCB which follows extensive methodologies and policies to validate the data. Codes used by the dataset are taken from the "National Standard Codes (ANSI INICITS 38-2009), Federal Information Processing Series (FIPS) - States" (USCB 2016). The delineation of the census tracts was defined by local participants involved in the 2010 Census Participant Statistical Areas Program (USCB 2016).	The data was developed for the purposes of both the National Geographic Database and Statistics Canada-Elections Canada initiative (StatCan 2017d). It is assumed that appropriate methods are carried out to validate the data however, it is not stated in the Boundary Files, Reference Guide.
Accuracy	According to the Shapefile metadata the positional accuracy is not to the standards that the six decimal places that the coordinate system suggests (USCB 2016). The boundaries are strictly for statistical purposes and does not contain any legal authority (USCB 2016).	The positional accuracy of the data is not precise and not intended to be used for legal purposes. The data involved was from multiple sources with different scales which were rescaled for the boundary files (StatCan 2017d). The attribute accuracy was verified against Spatial Data Infrastructure data and was deemed accurate (StatCan 2017d).

Data Quality Characteristic	Minnesota Watershed Boundaries	Ontario Watershed Boundaries
Completeness	Tested for completeness (MNDNR 2009) and provides complete polygons for the study area.	The data is incomplete as there is not watershed boundaries encompassing the entirety of the city.
Security	The data was unrestricted (MNDNR 2009). The publisher of the data includes a User Agreement where the data must be adequately sourced following the specific requirements (MNDNR 2009)	No available information.
Timeliness	The data undergoes maintenance and updates yearly and represents the HUC level 7 from 1998 to the present (MNDNR 2009)	No available information.
Relevance	The relevance of this data good. The ISAT tutorial guides users on how to use the tool with similar data.	The data is the most relevant and available data for a small scale watershed dataset to understand the effect of ISA.
Validity	The dataset was defined and referenced according to 103G.005 Subd. 17a and 103B.305 Subd. 10 Minnesota Statutes for the state watershed map production (MNDNR 2009).	The data is refined for the purposes of LRCA where watershed monitoring occurs. It is assumed that the refined watershed polygons are more appropriate for the City of Thunder Bay.
Accuracy	There are no quantitative studies to determine the positional accuracy of the data (MNDNR 2009). However, the accuracy was estimated to be with 30-metres positional accuracy (MNDNR 2009).	According to a LRCA employee the data based on a 20-metre resolution DEM layer (in email, Mar. 6, 2020).

APPENDIX B

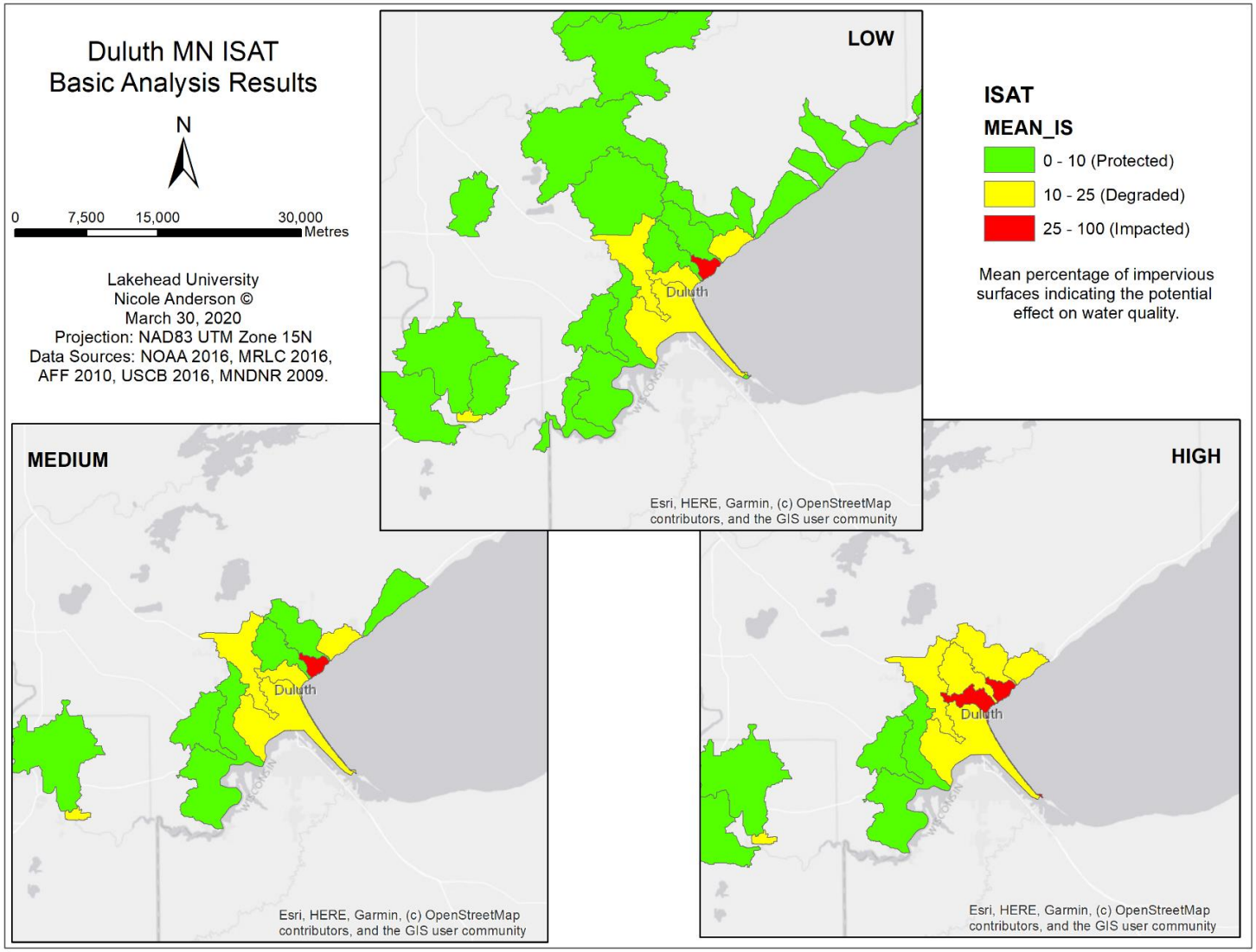
MAPS AND DATA

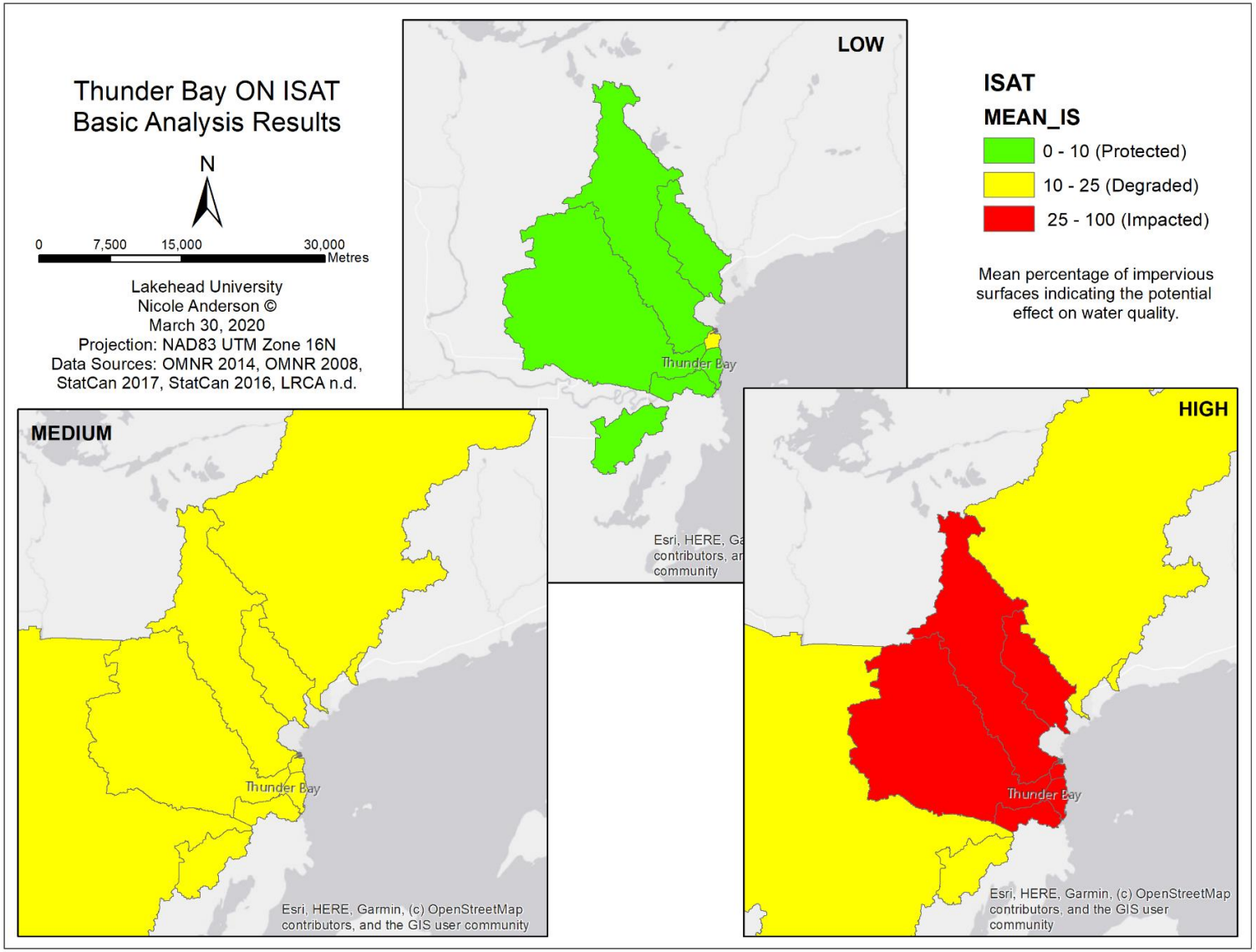
Thunder Bay ON, Land cover imperviousness coefficients based on high, medium and low population density.

Value	LCName	High	Med	Low
0	Unclassified	0.00	0.00	0.00
1	Clear Open Water	33.47	17.25	2.54
2	Turbid Water	0.00	0.00	0.00
3	Shoreline	0.00	0.00	0.00
4	Mudflats	0.00	0.00	0.00
5	Marsh	0.00	0.00	0.00
6	Swamp	0.00	0.00	0.00
7	Fen	0.00	0.00	1.56
8	Bog	7.00	7.00	2.75
10	Heath	0.00	0.00	0.00
11	Sparse Treed	29.32	13.76	2.46
12	Treed Upland	0.00	0.00	0.00
13	Deciduous Treed	23.64	13.93	2.20
14	Mixed Treed	23.87	11.10	2.24
15	Coniferous Treed	25.20	11.73	3.73
16	Plantations - Treed Cultivated	0.00	0.00	0.00
17	Hedge Rows	0.00	0.00	0.00
18	Disturbance	0.00	0.00	1.35
19	Cliff and Talus	0.00	0.00	0.00
20	Alvar	0.00	0.00	0.00
21	Sand Barren and Dune	0.00	0.00	0.00
22	Open Tallgrass Prarie	0.00	0.00	0.00
23	Tall Grass Savannah	0.00	0.00	0.00
24	Tall Grass Woodland	0.00	0.00	0.00
25	Sand/Gravel/Mine Tailings/Extraction	0.00	0.00	5.02
26	Bedrock	0.00	0.00	0.00
27	Community/Infrastructure	33.50	25.86	16.70
28	Agriculture and Undifferentiated Rural Land Use	0.00	0.00	2.50

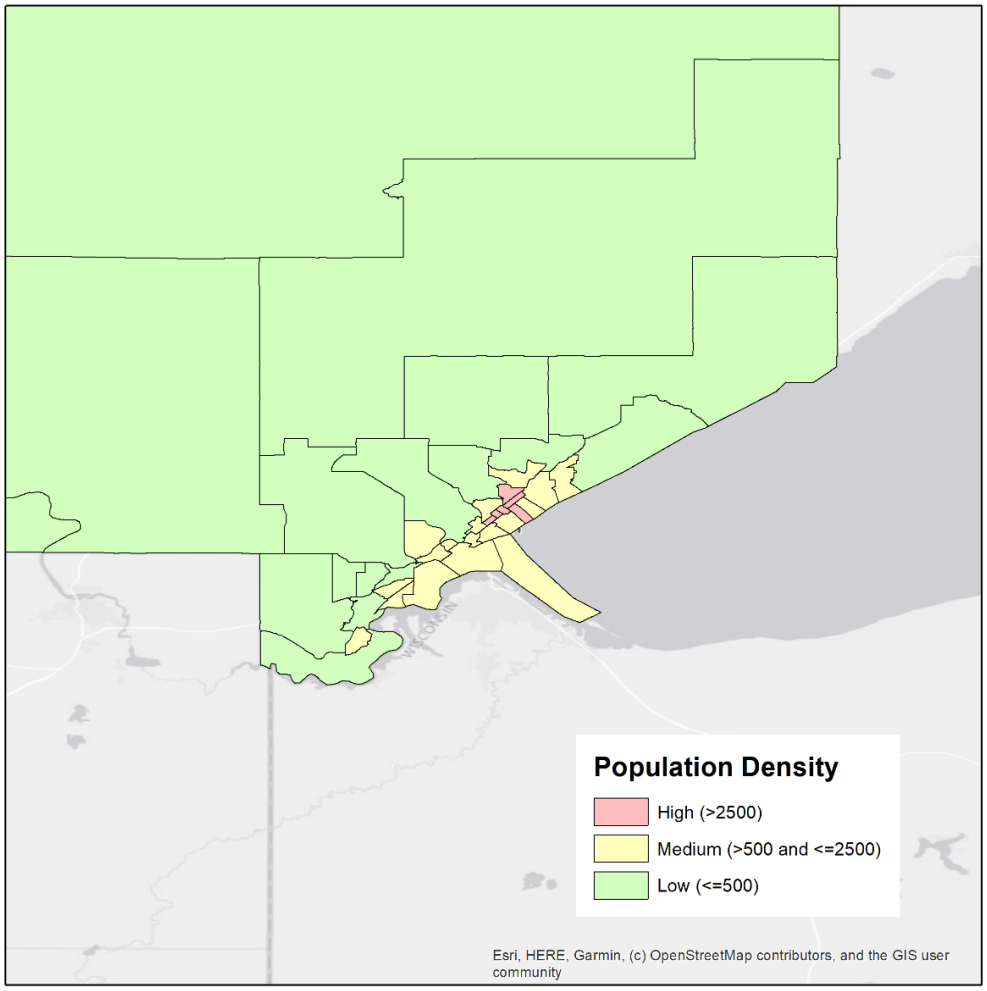
Duluth MN, Land cover imperviousness coefficients based on high, medium and low population density.

Value	LCName	High	Med	Low
0	Unclassified	0.00	0.00	0.00
2	Developed, High Intensity	0.80	56.64	49.97
3	Developed, Medium Intensity	67.40	32.94	30.36
4	Developed, Low Intensity	40.15	22.15	21.59
5	Developed, Open Space	24.34	13.28	13.10
6	Cultivated Crops	14.41	0.40	0.45
7	Pasture/Hay	0.34	0.35	0.35
8	Grassland/Herbaceous	0.31	0.59	2.68
9	Deciduous Forest	0.59	0.15	0.99
10	Evergreen Forest	0.15	0.23	6.59
11	Mixed Forest	0.30	0.19	6.97
12	Scrub/Shrub	0.39	0.30	0.85
13	Palustrine Forested Wetland	0.29	0.05	0.11
14	Palustrine Scrub/Shrub Wetland	0.06	0.08	0.14
15	Palustrine Emergent Wetland	0.08	0.11	0.36
16	Estuarine Forested Wetland	0.00	0.00	0.00
17	Estuarine Scrub/Shrub Wetland	0.00	0.00	0.00
18	Estuarine Emergent Wetland	0.00	0.00	0.00
19	Unconsolidated Shore	0.11	0.52	1.17
20	Bare Land	7.93	2.34	2.30
21	Open Water	6.59	0.02	11.10
22	Palustrine Aquatic Bed	0.06	0.00	0.00


Duluth MN ISAT
Basic Analysis Results
N
0
7,500
15,000
30,000
Metres
Lakehead University
Nicole Anderson ©
March 30, 2020
Projection: NAD83 UTM Zone 15N
Data Sources: NOAA 2016, MRLC 2016,
AFF 2010, USCB 2016, MNDNR 2009.
LOW
MEDIUM
HIGH
Duluth
Esri, HERE, Garmin, (c) OpenStreetMap
contributors, and the GIS user community
ISAT
MEAN_IS
0 - 10 (Protected)
10 - 25 (Degraded)
25 - 100 (Impacted)
Mean percentage of impervious
surfaces indicating the potential
effect on water quality.


Thunder Bay ON ISAT
Basic Analysis Results
N
0
7,500
15,000
30,000
Metres
Lakehead University
Nicole Anderson ©
March 30, 2020
Projection: NAD83 UTM Zone 16N
Data Sources: OMNR 2014, OMNR 2008, StatCan 2017, StatCan 2016, LRCA n.d.
LOW
Thunder Bay
Esri, HERE, Ga
contributors, ar
community
ISAT
MEAN_IS
0 - 10 (Protected)
10 - 25 (Degraded)
25 - 100 (Impacted)
Mean percentage of impervious surfaces indicating the potential effect on water quality.
MEDIUM
Thunder Bay
Esri, HERE, Garmin, (c) OpenStreetMap contributors, and the GIS user community
HIGH
Thunder Bay
Esri, HERE, Garmin, (c) OpenStreetMap contributors, and the GIS user community

Census Tract Boundaries
United States Census Bureau

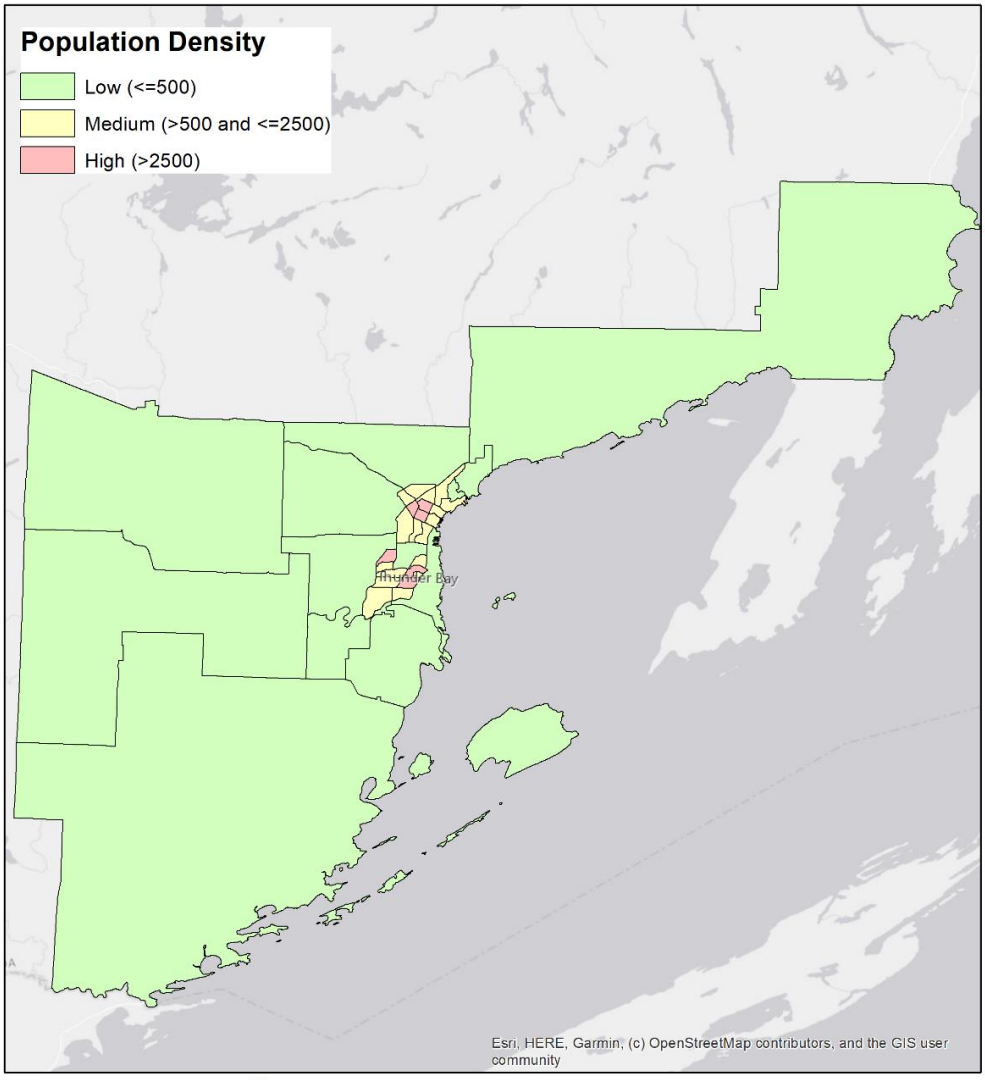


N

0 7,500 15,000 30,000
Metres

Lakehead University
Nicole Anderson ©
March 30, 2020
Projection: NAD83 UTM Zone 15N
Source: USCB 2016, AFF 2010.

Thunder Bay Census Tract Boundaries
Statistics Canada

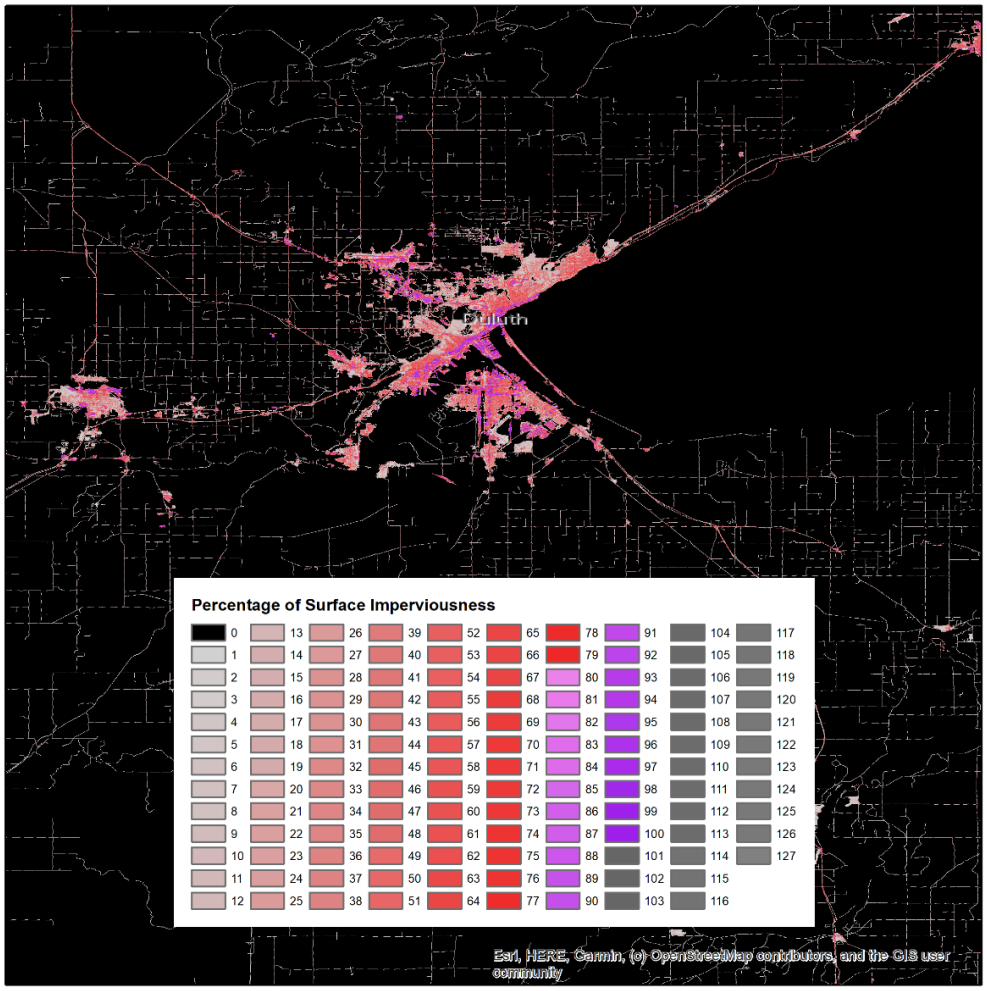


N

0 7,000 14,000 28,000
Metres

Lakehead University
Nicole Anderson ©
March 30, 2020
Projection: NAD83 UTM Zone 16N
Source: StatCan 2017, StatCan 2016

Duluth MN Surface Imperviousness
Multi-Resolution Land Characteristics Consortium

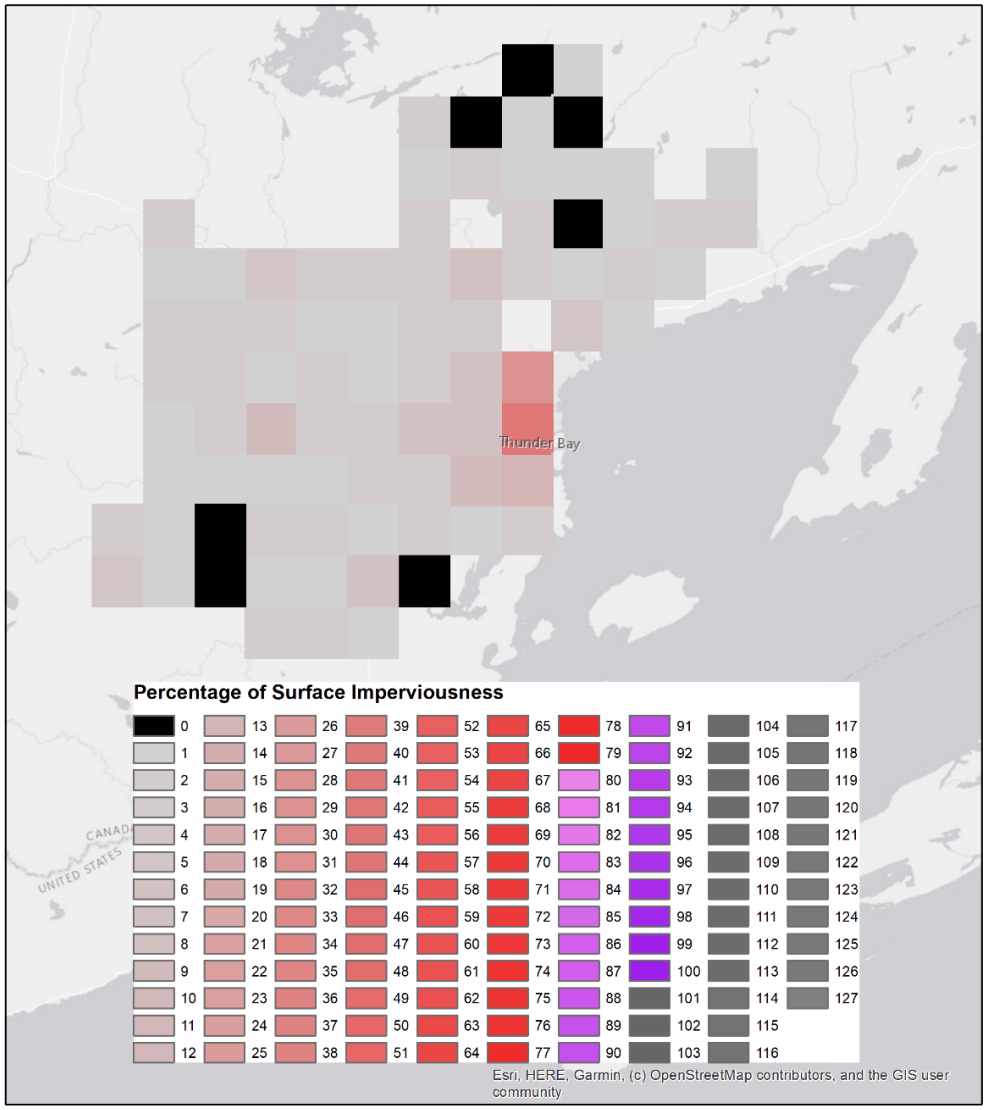


N

0 8,750 17,500 35,000
Metres

Lakehead University
Nicole Anderson ©
March 30, 2020
Projection: NAD83 UTM Zone 15N
Source: MRLC 2016

Thunder Bay Surface Imperviousness

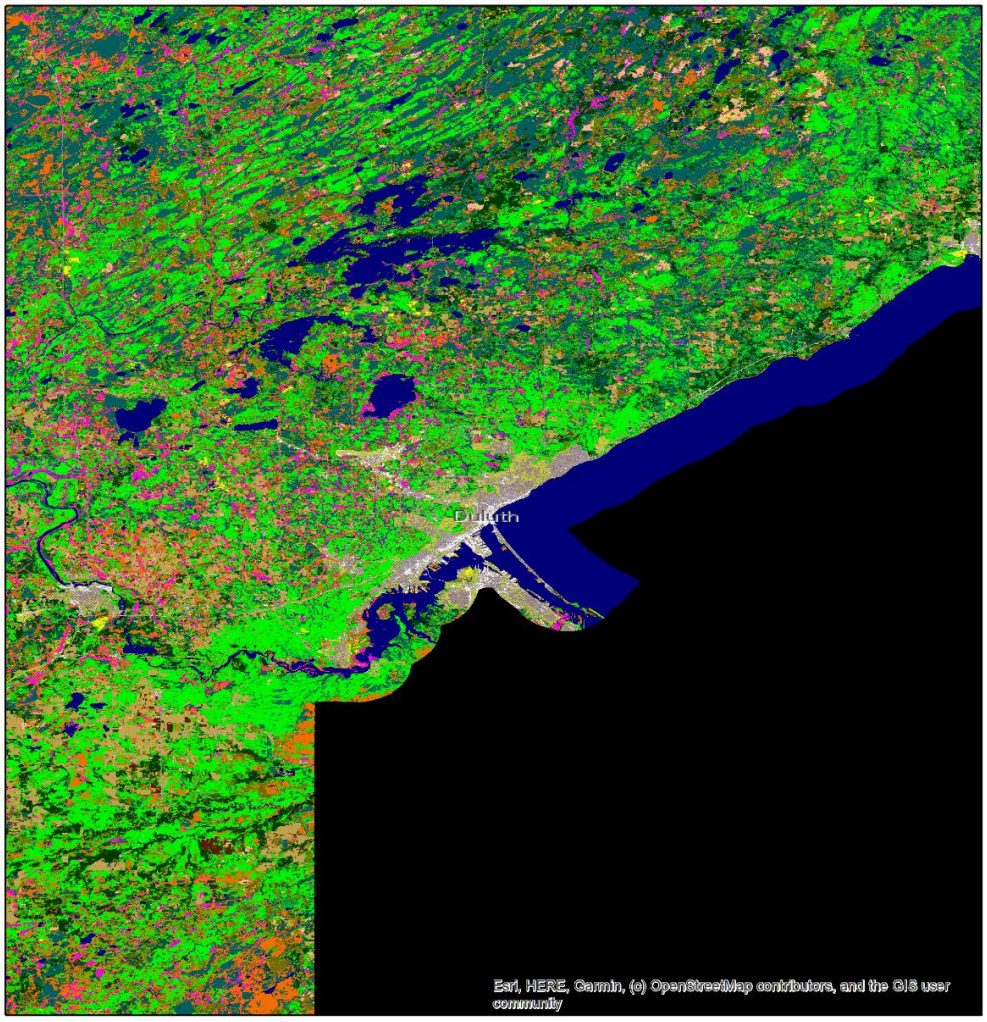


N

0 7,000 14,000 28,000
Metres

Lakehead University
Nicole Anderson ©
March 30, 2020
Projection: NAD83 UTM Zone 16N
Data Source: OMNR 2008

Land Cover Classification
National Oceanic and Atmospheric Association

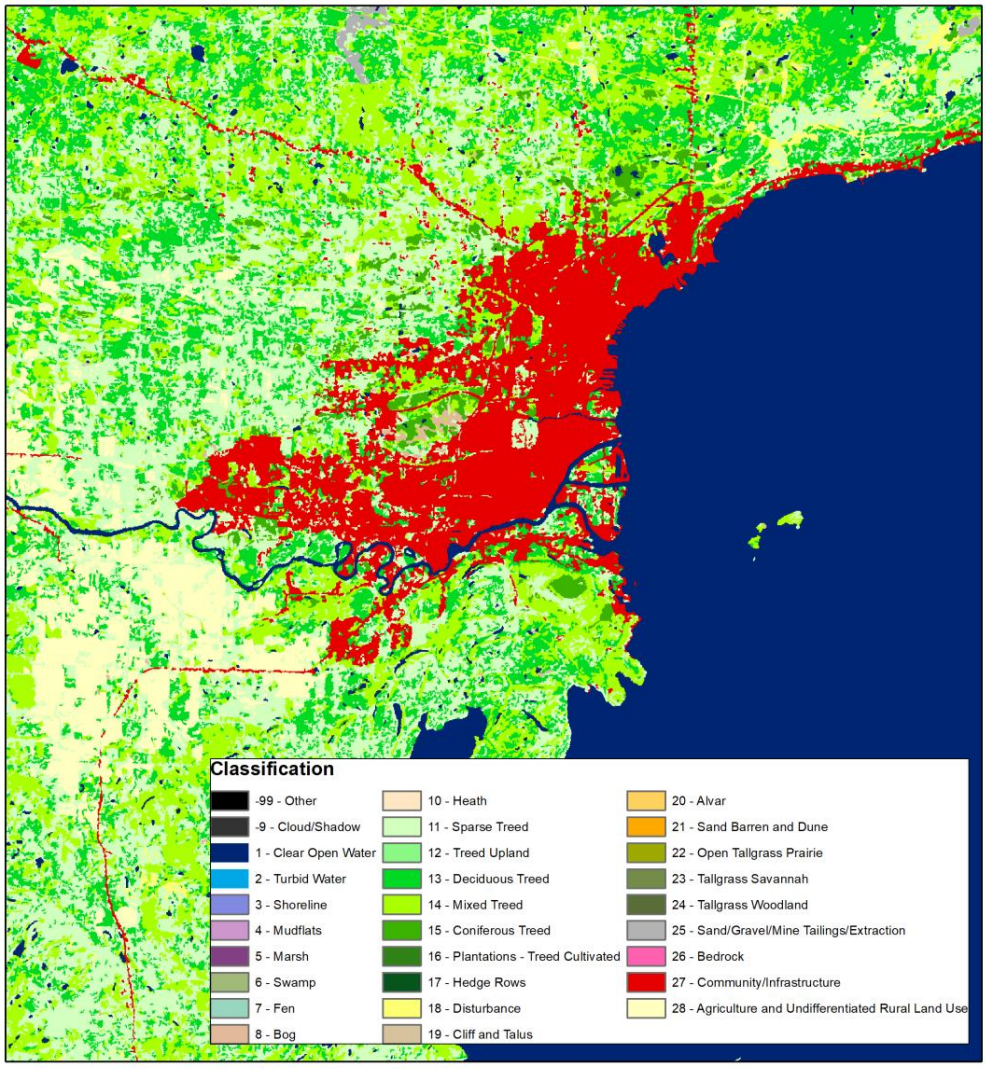


N

0 8,750 17,500 35,000
Metres

Lakehead University
Nicole Anderson ©
March 30, 2020
Projection: NAD83 UTM Zone 15N
Source: NOAA 2016

Thunder Bay Land Cover Classification
Ministry of Natural Resources

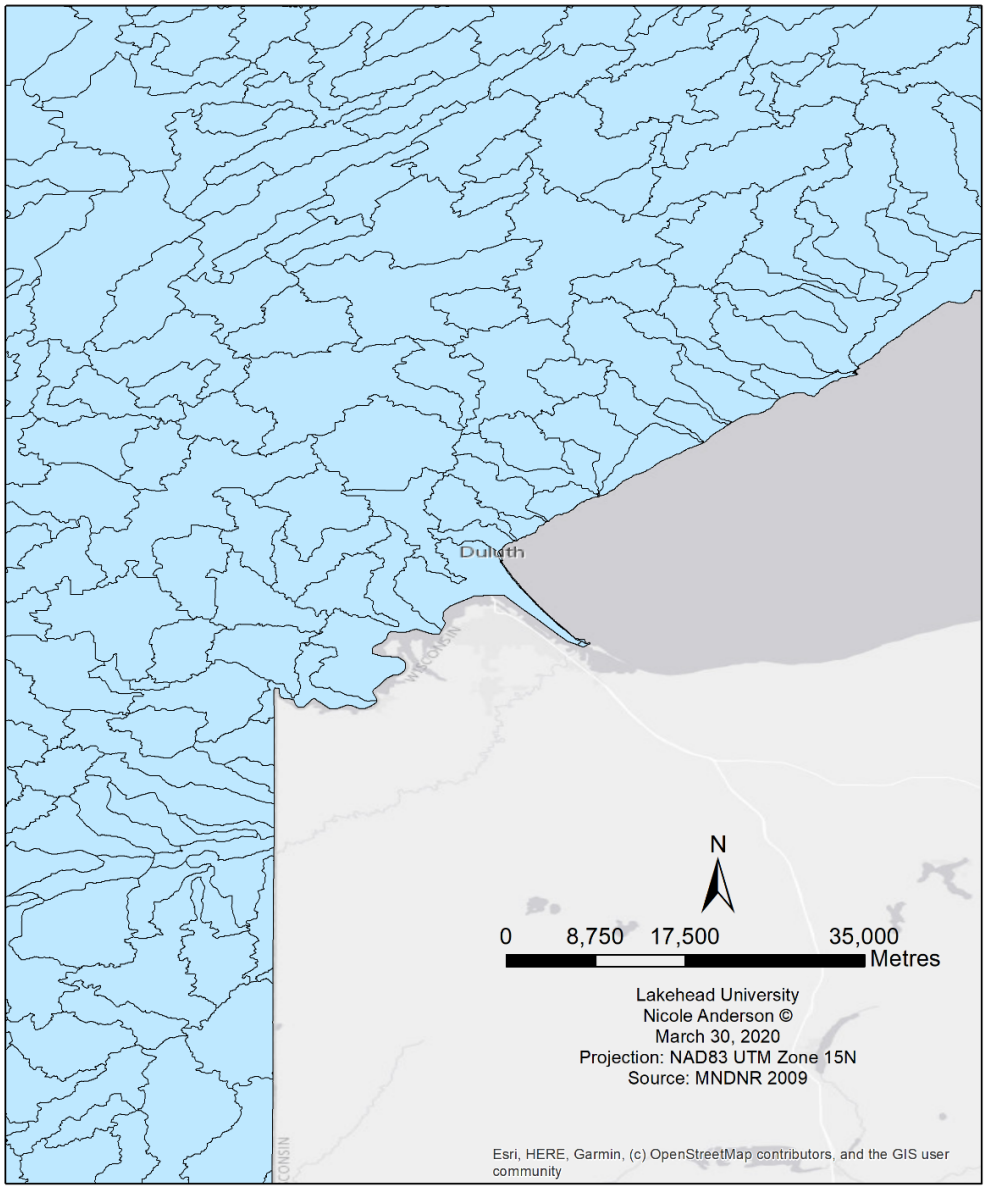


N

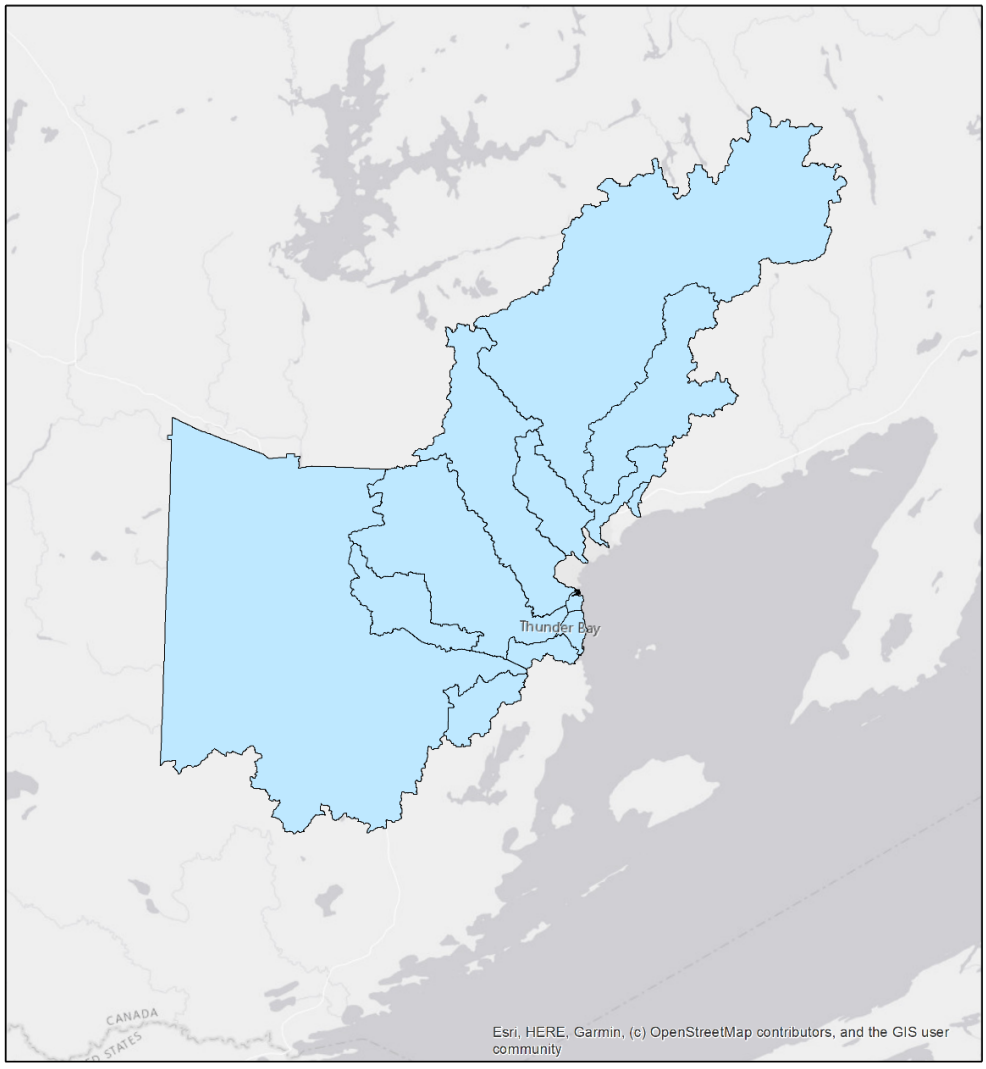
0 2,850 5,700 11,400
Metres

Lakehead University
Nicole Anderson ©
March 30, 2020
Projection: NAD83 UTM Zone 16N
Source: OMNR 2014

Duluth Minnesota Watershed Boundaries
Minnesota Department of Natural Resources



Thunder Bay Watershed Boundaries
Lakehead Region Conservation Authority



N

0 7,000 14,000 28,000
Metres

Lakehead University
Nicole Anderson ©
March 30, 2020
Projection: NAD83 UTM Zone 16N
Source: LRCA n.d.