



# Applying GIS and Geospatial Data to Reference Condition Approach Stream Assessments

---

Guidance for Canadian Aquatic Biomonitoring Network Participants



Cat. No.: En84-337/2023E-PDF  
ISBN: 978-0-660-48781-6  
EC23118

Unless otherwise specified, you may not reproduce materials in this publication, in whole or in part, for the purposes of commercial redistribution without prior written permission from Environment and Climate Change Canada's copyright administrator. To obtain permission to reproduce Government of Canada materials for commercial purposes, apply for Crown Copyright Clearance by contacting:

Environment and Climate Change Canada  
Public Inquiries Centre  
12<sup>th</sup> Floor, Fontaine Building  
200 Sacré-Coeur Boulevard  
Gatineau QC K1A 0H3  
Telephone: 819-938-3860  
Toll Free: 1-800-668-6767 (in Canada only)  
Email: [enviroinfo@ec.gc.ca](mailto:enviroinfo@ec.gc.ca)

© His Majesty the King in Right of Canada, as represented by the Minister of Environment and Climate Change, 2023

Aussi disponible en français

**Applying GIS and Geospatial Data to Reference Condition  
Approach Stream Assessments: Guidance for the Canadian  
Aquatic Biomonitoring Network Participants**

**By: Adam G. Yates<sup>1</sup>, Edward M. Krynak<sup>1</sup> and Wendy A. Monk<sup>2</sup>**

<sup>1</sup>University of Waterloo, Department of Biology, Waterloo, Ontario

<sup>2</sup>Environment and Climate Change Canada, Fredericton, New Brunswick

## Contents

1	Document purpose .....	1
2	Background .....	1
2.1	What is GIS? .....	1
2.2	GIS software .....	2
2.3	Why a geospatial approach can benefit stream assessment.....	3
3	General considerations for using geospatial data in stream assessments .....	5
3.1	Locating and selecting data.....	5
3.1.1	Data sources.....	5
3.1.2	Data scale, resolution and coverage .....	8
3.1.3	Data gaps and incomplete coverages .....	9
3.1.4	Data resolution .....	10
3.1.5	Data temporal frame .....	10
3.2	Data management .....	11
3.2.1	Data QA/QC.....	11
3.2.2	Metadata.....	13
4	Applying GIS to stream assessments .....	13
4.1	Application #1 – Site description and site selection .....	13
4.1.1	Identification of reference sites.....	14
4.1.2	Locating and describing sites .....	14
4.1.3	Applications of a site’s landscape descriptions .....	15
4.1.4	Describing natural variation.....	17
4.1.5	Describing human activities using GIS .....	17
4.2	Application #2 – RCA model building.....	20
4.2.1	Identifying landscape attributes that discriminate community groups .....	20

4.2.2	Temporal frame considerations for GIS applications to RCA models.....	21
4.2.3	Use of standardized data for model building .....	21
4.3	Application #3 – Assessment interpretation.....	23
5	Summary .....	26
6	References .....	27
7	Glossary.....	i
8	Quick reference guide.....	iii
8.1	Finding geospatial data .....	iii
8.2	Selecting geospatial data .....	iii
8.3	Data QA/QC.....	iii
8.4	Metadata.....	iii
8.5	Assessment site selection and description .....	iv
8.6	Reference Condition Approach model building.....	iv
8.7	Assessment interpretation.....	iv
9	Additional resources .....	v
9.1	GIS resources.....	v
9.1.1	Introduction to GIS and GIS software - books .....	v
9.1.2	Introduction to GIS and GIS software - online resources .....	vi
9.1.3	GIS software and tools .....	vi
9.2	QA/QC .....	vii
ArcGIS Geodatabase Topology Rules.....	vii	
9.3	Metadata.....	vii
9.4	Inverse Distance Weighting (IDW) .....	viii

# 1 Document purpose

The purpose of this guidance document is to provide participants of the CABIN program with a broad overview of Geographic Information Systems (GIS) and geospatial data and how they can be applied to stream assessments in CABIN. Specifically, the document presents information related to two objectives. First, a series of general considerations are described related to selecting appropriate data and ensuring their utility through proper metadata generation and quality assurance and quality control protocols (QA/QC) for CABIN. Second, the document provides detailed guidance on how CABIN participants can apply geospatial data to meet three key stream assessment objectives: 1) selecting test sites and reference sites; 2) building assessment models; and 3) interpreting assessment results. Although this document provides general guidance for CABIN participants, it is not a how-to or step-by-step guide related to using GIS; rather, the following content is to arm participants with the knowledge of how GIS can support comprehensive, robust and informed stream assessments. However, this document outlines the basic process that the user would follow for each application, and detail key considerations when applying GIS to CABIN-based stream assessments. For more detailed instructions on how to use GIS and the different GIS software available, we encourage the user to seek out the resources listed later in this document (see GIS Resources).

## 2 Background

### 2.1 What is GIS?

GIS is a computer system for capturing, assembling, manipulating, analyzing, storing and displaying geographically referenced information, known as geospatial data (Tsihrintzis et al., 1996; Chrisman, 1999). GIS allows for the quantification of landscapes and climate longitudinally, laterally, vertically, and temporally, and thus allows streams to be put into context of the landscape. This, in turn, allows researchers and policy makers to visualize the stream and habitat hierarchy, landscape characteristics and patterns, and areas of influence for which policy decisions are being developed (Johnson & Gage, 1997). Indeed, the benefits of using GIS for the planning and implementation of aquatic assessments and monitoring have been recognized since the late 1980s (e.g., Osborne & Wiley, 1988; Tsihrintzis et al., 1996 and references therein).

Beyond the ability of GIS to help visualize a sampling site's place in the landscape, GIS can also increase the efficiency of identifying natural environmental and anthropogenic gradients along with the classification of reference sites, which are critical steps in the Reference Condition Approach (RCA) used in CABIN assessment and monitoring (Yates & Bailey, 2010a; Armanini et al., 2013). The remote identification of natural gradients and human activity has become easier and more efficient with GIS and the growing availability of high-quality landscape data (Yates & Bailey, 2010a). Natural environmental gradients such as climate, geology, and topography allow for the effective classification of streams and identification of potential sample sites with minimal time and cost commitment when compared to extensive and time-consuming pre-assessment sampling that would otherwise be necessary (Yates & Bailey, 2010a). In addition, GIS can help to reduce the subjectivity that often accompanies the selection of reference sites. Through the use of land-use data (e.g., row crops, urbanization) GIS supports the identification of reference sites and can place them within a regionally representative natural environment before field crews even leave the office (Yates & Bailey, 2010a).

Geospatial data offers the benefit of reducing the level of inter-operator variability that can affect assessment programs that are highly dependent on site-level descriptors such as habitat and water chemistry (Armanini et al., 2013). Field protocols for data collection tend to evolve through time and their interpretation can be different among agencies, or even between crews within an agency, complicating data sharing among programs. However, the collection of GIS data can be standardised and is reliant on GIS layers or source data that are highly regulated. Using GIS to identify site and landscape characteristics supports the creation of reliable datasets that can be shared among agencies and stakeholders, facilitating the management of stream systems. Furthermore, the characterization of landscape variables helps to identify and prioritize future studies and assessments, reducing costs and extending budgets (Ritters & Wickham, 1995).

## **2.2 GIS software**

The options for GIS software are numerous, ranging from commercial or proprietary software requiring purchase of a license to free and open-source software in which the original source code is available for distribution and modification. The intent of this document is not to recommend or review all the possible options, but only to provide a brief overview and resources for more information. The chosen software will likely depend on agency precedence, availability, funds, and user ability. If software choice is an option, the user should consider the current and future needs before making a decision. Software availability may differ among operating systems and some software may not meet

the requirements to perform all GIS-related tasks such as map creation, data analysis, data transformation, and data conflation (i.e., the integration of two or more datasets into a new dataset) (Samal et al., 2004; Steiniger & Hunter, 2013). Esri (Environmental Systems Research Institute), founded in 1969 shortly after the conception of GIS in the early 1960s, is one of the best-known and most-used commercial GIS softwares available ([www.esri.com](http://www.esri.com)) through ArcMap and ArcGIS. However, the availability and power of free and open-source software is growing and will continue to do so (Coetzee et al., 2020). QGIS ([www.qgis.org](http://www.qgis.org)) and GRASS GIS ([www.grass.osgeo.org](http://www.grass.osgeo.org)) are two examples of well-known open-source software with capabilities similar to that of proprietary software (Steiniger & Hunter, 2013; Coetzee et al., 2020). In addition, open-source data science languages, such as R, “a free software environment for statistical computing and graphics” (R Core Team, 2020), are becoming more common for the processing and analysis of geospatial data (Coetzee et al., 2020). Several R packages and tutorials have been developed explicitly for GIS analysis with R (See GIS Resources). Indeed, the R package ‘openSTARS’, the open-source version of the Spatial Tools for the Analysis of River Systems (STARS) ArcGIS toolset, allows users to prepare GIS data for use in a stream monitoring program using only free and open-source software (Kattwinkel et al., 2020). For a more complete review of free and open-source GIS software and their capabilities, participants should visit the Open-source Geospatial Foundation (OSGeo) website ([www.osgeo.org](http://www.osgeo.org)) and read Steiniger and Hunter (2013) and Coetzee et al. (2020).

### **2.3 Why a geospatial approach can benefit stream assessment**

A stream is intimately connected to the watershed in which it lies, and reflects its landscape through its physical features and biota (Hynes, 1975; Richards & Host, 1994). The potential impacts of rain on a stream system depend on where the raindrops first land. In an agricultural area, the raindrop has the potential to carry additional nutrients, pesticides, and sediment to streams. Impervious surfaces speed the passage of water and road waste to the stream, increasing conductivity and the flashiness of the hydrograph (Walsh et al., 2005). In contrast, a raindrop falling in a naturally vegetated landscape is more likely to percolate through the soil, slowing its path to the stream and collecting dissolved organic carbon (DOC), nutrients, sediments and other chemical constituents along the way (Allan, 2004; Burt & Pinay, 2005; Vidon et al., 2010). The path taken by rain across the landscape could be centimeters in length, or tens of kilometres before it reaches the stream. Stream assessment has historically used only site-level descriptors such as substrate, water quality, and biotic community with limited connection to



the bigger landscape (Thoms et al., 2018). However, we can see that we need to also consider the wider catchment when undertaking stream assessment.

The hierarchical nature of stream systems allows us to conceptualize the connection of a single point in a stream to the larger landscape (Figure 1, Frissell et al., 1986). A stream system consists of a series of nested levels of organization across different temporal and spatial scales with each level intimately connected to upper and lower levels.

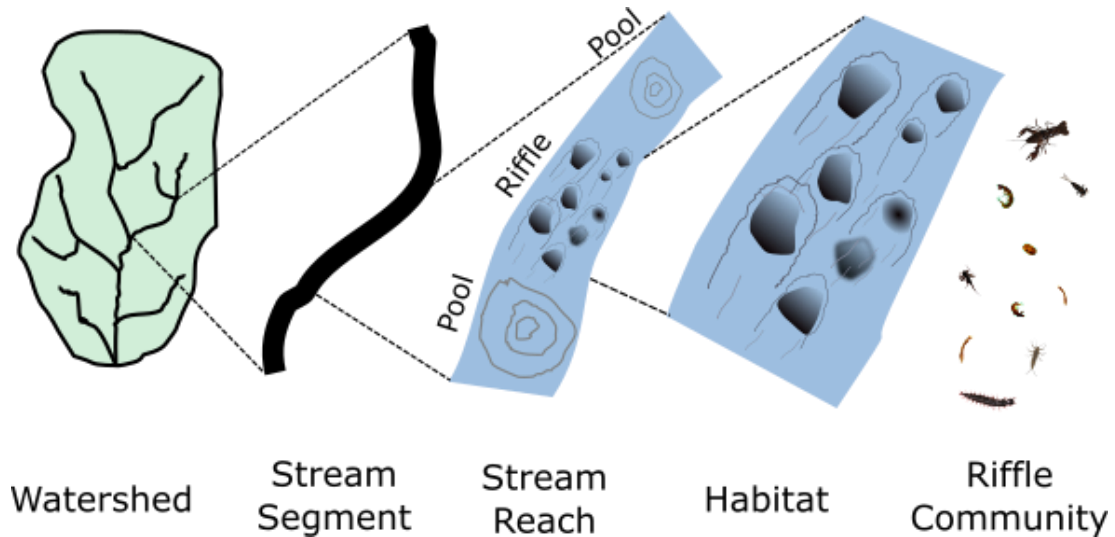


Figure 1. The hierarchical structure of a stream system

At the watershed scale, geology, topography, and climate are the drivers for each sequentially smaller scale through influences on hydrology, vegetation cover, channel morphology, and thermal regimes. Influences become progressively more localized across smaller temporal and spatial scales. For example, at the stream segment scale, surficial geology and land cover are reflected in variables such as channel morphology, groundwater input, and water chemistry (e.g., conductivity, alkalinity) (Figure 1). Likewise, sediment and nutrient inputs, as well as dominant substrate material, reflect surficial geology and land cover at the stream reach scale (Figure 1). Thus, by setting the template for habitats, landscape variables at the watershed scale indirectly determine the biota that reside in the stream, as only biota that are suited to the combination of habitats and water chemistry in a stream will successfully populate the system. Knowledge of the landscape conditions from which a stream arises can thus provide significant insight into the resident biological community. Consequently, stream bioassessments are increasingly incorporating collection and application of landscape information into assessment

procedures. GIS is an increasingly effective tool by which we can visualize and analyze the characteristics of a stream within its watershed and thereby enhance biological monitoring and assessment.

### 3 General considerations for using geospatial data in stream assessments

Prior to including geospatial data as part of a stream assessment, participants should carefully address a variety of considerations. This includes locating and selecting appropriate geospatial data, troubleshooting data, validating and verifying generated data, and properly annotating data generated using GIS. Care must also be taken to avoid misapplication of geospatial data. The following sub-sections address many of these considerations. Decisions surrounding these considerations must serve the purpose and conditions of the stream assessment to be undertaken.

#### 3.1 Locating and selecting data

##### 3.1.1 *Data sources*

Many geospatial data sources can be used in a GIS to assist with various aspects of stream assessment. Although a review of data sources is beyond the scope of this document, participants are directed to Yates et al. (2019) for a full review of geospatial data sources for use in stream assessments. Participants will almost certainly require a digital elevation model and stream network for all GIS applications related to stream assessment to delineate the watershed boundaries associated with their sampling sites. Moreover, depending upon the application, participants will also likely benefit from geospatial data describing aspects of geology, climate and land use/land cover. The utility and suitability of these and other more thematic geospatial data sources will be dependent upon the specific application and goals of the individual assessment at hand.

Numerous online geospatial data clearinghouses provide free access to Canadian GIS data (Table 1). This document focuses solely on well-known sources of geospatial data applicable to stream assessments that are open-access and requires little to no preprocessing for use in a GIS. Nearly all the provinces and territories have geospatial data available for free download, although some require registration for access. Large-scale data are distributed from federal agencies through online sources such as the Open Canada page ([open.canada.ca/](http://open.canada.ca/)), which has numerous geospatial data sources that are readily downloaded. These online sources provide many of the data that are routinely needed to

support stream assessments including stream network data, topography data in the form of digital elevation models (DEMs) and thematic landscape descriptions (e.g., land cover, geology, and climate).

Table 1. Geospatial data sources and associated websites at the national, provincial and territorial levels of Canada\*.

Jurisdiction	Provider	Webpage
Canada	Government of Canada	<a href="https://geogratias.gc.ca/">https://geogratias.gc.ca/</a>
Alberta	GeoDiscover Alberta	<a href="https://geodiscover.alberta.ca/">https://geodiscover.alberta.ca/</a>
British Columbia	DataBC	<a href="https://catalogue.data.gov.bc.ca/">https://catalogue.data.gov.bc.ca/</a>
Manitoba	Manitoba Land Initiative	<a href="https://mli2.gov.mb.ca/">https://mli2.gov.mb.ca/</a>
New Brunswick	GeoNB	<a href="http://www.snb.ca/geonb1/e/DC/catalogue-E.asp">http://www.snb.ca/geonb1/e/DC/catalogue-E.asp</a>
Newfoundland and Labrador	Fisheries, Forestry and Agriculture (FFA) GeoHub Environment, Climate Change and Municipalities	<a href="https://geohub-gnl.hub.arcgis.com/">https://geohub-gnl.hub.arcgis.com/</a>
		<a href="https://www.gov.nl.ca/eccm/">https://www.gov.nl.ca/eccm/</a>
Nova Scotia	Nova Scotia Government Open Data Portal	<a href="https://data.novascotia.ca/">https://data.novascotia.ca/</a>
Northwest Territories	NWT Centre for Geomatics Northwest Territories Geological Survey	<a href="https://www.geomatics.gov.nt.ca/">https://www.geomatics.gov.nt.ca/</a>
		<a href="https://ntgs-open-data-ntgs.hub.arcgis.com/">https://ntgs-open-data-ntgs.hub.arcgis.com/</a>
Ontario	Land Information Ontario	<a href="http://geohub.lio.gov.on.ca">http://geohub.lio.gov.on.ca</a>
Prince Edward Island	Prince Edward Island Government	<a href="http://www.gov.pe.ca/gis/">http://www.gov.pe.ca/gis/</a>
		<a href="https://data.princeedwardisland.ca/">https://data.princeedwardisland.ca/</a>
Québec	Québec Open Data Portal	<a href="https://www.donneesquebec.ca/">https://www.donneesquebec.ca/</a>
Saskatchewan	Saskatchewan GeoHub	<a href="https://geohub.saskatchewan.ca/">https://geohub.saskatchewan.ca/</a>
Yukon	Government of Yukon Open Data	<a href="https://open.yukon.ca/data/">https://open.yukon.ca/data/</a>

\* Currently there are no open access geospatial data available for Nunavut.

There are also useful regional datasets applicable to smaller scale assessments. CanadianGIS.com has an extensive list of links to such sources, as well as links to provincial and national government databases that the user may find helpful when searching for geospatial data. However, the user should be aware that the types of geospatial data available and the scales, resolutions, and coverages at which these data have been generated varies considerably depending upon the source.

The improvement of computing power, data acquisition and a recent push to make data freely available online has led to an overwhelming array of data sources for use in a GIS. To ensure that the most appropriate data are acquired, it is recommended that the user predetermine what geospatial data would be best suited to support each specific assessment's project goals prior to beginning a search to acquire data. In this way, the user can have a clear purpose and end goal to their data search at the outset of their project. Having a clear plan will better enable the user to avoid over-collection of data or settling for less appropriate data sources. However, the user should be aware that in many cases the "perfect" data source may not exist, and a data search plan may need to be modified as information on data availability is gathered through the search process.

### *3.1.2 Data scale, resolution and coverage*

Prior to selecting and analyzing geospatial data, it is critical that the most appropriate scale, resolution and coverage of data be determined. In this instance, scale is defined as the amount of reduction between the real world and its graphic representation. On paper maps, the scale helps relate the distance on the map to the distance in the real world. This is less important in a GIS, as the scale of the map is based on the display settings of the GIS (i.e., how zoomed in the map is) as geospatial data sources of greater resolution will thus better reflect changes in landscape properties that occur over small distances. Lastly, the coverage of the geospatial data is the total area that the data describes. For example, a spatial coverage could be the Province of Manitoba or North America. Coverage can also be considered from a temporal perspective, as all geospatial data will reflect a point in time or range of time. For example, a land cover layer could be based on satellite imagery from an individual year, whereas a geospatial climate layer could be based on averages from a 30-year time window.

There are several key considerations to ensure the selected geospatial data have the most appropriate scale, resolution and coverage for your project. First, these data characteristics should be considered in the context of the goals and context of the assessment project. Perhaps the easiest of the considerations surrounds coverage. Selected data should be of sufficient coverage to ensure that all

areas pertinent to the assessment are included. It is essential that the data user have a clear idea of the boundaries of their assessment area. For example, in many cases provincial/territorial-level data will offer the greatest balance between resolution and coverage for region-specific assessments within a province. Alternatively, when assessments encompass multiple provincial or territorial borders, data generated at the national level are likely to be most appropriate. However, a note of caution: the area of interest is often much larger in area than anticipated, as the cumulative drainage areas of rivers can extend large distances beyond the sampling sites and political boundaries of interest. The spatial extent of watersheds often has little relationship to the political boundaries that frequently represent the limits of data coverage for an individual data layer. As a result, participants may encounter situations where the region of interest crosses political boundaries.

Cross-boundary watersheds can pose difficulties for GIS coverage, as the data of choice may be unavailable on one side of the boundary or may be described at a different resolution or using a different classification system. In these instances, a less well-resolved data layer that has a larger extent may need to be used. For example, switching from a provincial layer to a national one may be necessary when a drainage area crosses a provincial boundary. In some circumstances, such as drainage areas that cross the international boundary of Canada and the USA, data of larger extent is not often available, and in such cases, participants may need to harmonize similar data layers that are available on both sides. “Stitching” data sources together in this manner should be used only as a last resort, as it can lead to improper classifications and increased levels of error. Fortunately, such instances will be rare for most participants, as most projects are likely to be contained within a province/territory or within a few provinces/territories.

### *3.1.3 Data gaps and incomplete coverages*

Users of this document should also be aware that even if a data layer has a spatial extent that is sufficient, there can be issues regarding coverage. Data layers can have gaps in coverage that are associated with unsurveyed areas (e.g., remote regions) or areas that could not be appropriately classified. Incomplete data coverage most often arises in data layers describing landscape attributes that vary at smaller time scales, such as land cover, and thus often have short data acquisition periods. For example, land cover layers are often generated from satellite-based imagery that may be hampered by cloud cover or shadows from steep slopes. These unclassified areas may result in significant portions of a watershed having no meaningful descriptions. Other cases of missing data occur when measurement stations are missing or too far apart to generate a continuous data coverage. Such situations are most

common in remote regions, where measurement stations, such as weather stations, may not be present in sufficient density to enable adequate interpolation of data to complete the geographical coverage.

Prior to using any geospatial data, participants should carefully examine all data for gaps and ensure that they read all available metadata so that classifications (e.g., cloud cover) are not misused. If it is determined that there are areas without data or that classifications include placeholder categories for areas where data could not be obtained, the user should clearly note in the metadata of any generated data that the data are incomplete. The user should carefully consider the prevalence of missing data prior to applying the data to their stream assessment; missing data could lead to erroneous conclusions about the state of the watershed. If large portions of the watershed are missing data, an alternative dataset may be more appropriate.

#### *3.1.4 Data resolution*

Appropriate data resolution as well as spatial extent of the study area should be selected to best answer the assessment questions. Landscape attributes that only vary significantly across larger distances (e.g., 10s to 100s of kilometres), such as climate and bedrock geology, can be described using data of very low resolution. In contrast, description of human activities may vary considerably across small distances (e.g., 10s to 100s of metres), and should therefore be described at much higher resolutions to ensure that spatial patterns are adequately captured. In many cases, the user may need to work with the data that are available for their study region regardless of the resolution. The data resolution challenge often applies to studies with very large spatial coverage (e.g., an entire province) or taking place in remote areas, where higher resolution data sources are less likely to be available.

#### *3.1.5 Data temporal frame*

GIS analyses are spatial in nature; however, some data are often associated with a specific temporal frame. Data that describes landscape attributes may represent extremely long time scales, such as geology, soils, and river network patterns. There is little concern as to when the data were generated as these features are highly unlikely to have changed. However, for descriptors of human activity and land cover, the temporal relevance of the data can be important. Even long-term climate normals must be used with caution given the rapidness of climate change in many parts of Canada. In such cases, the user should attempt to find geospatial data that most closely match the temporal period of relevance to the stream samples that have been or will be collected.

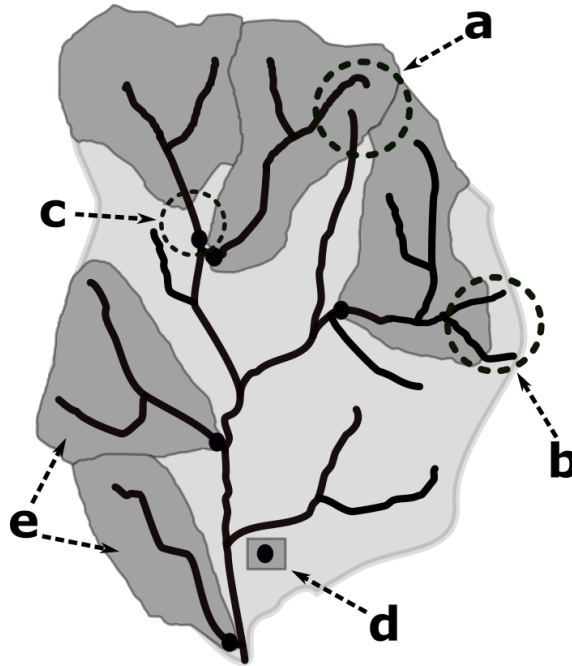
## 3.2 Data management

GIS gives the user the power to generate immense amounts of data in relatively short amounts of time. For example, simple applications such as intersections, spatial joins, as well as spatial calculations can generate data tables consisting of hundreds of rows and columns describing various attributes of a landscape even when starting with data layers consisting of only a small number of attributes. However, the user must take precautions to ensure that: 1) data generated are of appropriate quality; and 2) data are properly annotated with comprehensive metadata. Failure to address either of these issues can create significant problems for both the assessment for which the data have been generated as well as for future assessments which may benefit from using the generated data.

### 3.2.1 Data QA/QC

Data quality is essential for reliable and defensible stream assessments. GIS analysis will generate data based on the inputs and steps defined in the procedure. If there is a break in a stream line, an artificially flat spot in a DEM, or any other of a host of example issues, the GIS will still generate data. Errors in the layers used in the analysis will be multiplied in the data output. Layers describing stream networks and topography are used together to identify the boundaries of watersheds, and errors in any of the inputs can lead to watersheds that do not accurately reflect the landscape in a multitude of ways (Figure 2). Some errors, such as multiple outflows, may be easily identified using specific software functions, but other errors, such as incorrect truncation of the watershed or inaccurate placement of sampling sites, may only be identifiable through manual checks. Quality assurance and quality control (QA/QC) protocols must be in place at all phases of data generation to ensure quality and meaningful data. Data QA/QC processes are typically more time consuming than the actual analyses themselves. While some QA checks can be automated, there is no substitute for manual checks.





*Figure 2. Common errors in watershed generation. For example, errors in watershed boundaries (a), missing parts of the delineated watershed (b), or points not snapped correctly to the stream network resulting in either missing or incorrect boundaries (c & d). Correctly generated watersheds are noted as (e).*

Three general steps are recommended to ensure quality and meaningful data:

1. Check the completeness and accuracy of all data downloaded from Geospatial sources prior to using.
2. Conduct evaluations of the quality of the geospatial outputs at all steps of data processing to ensure that errors are corrected before they can propagate through multiple phases of analysis.
3. Evaluate data outputs/reports to ensure that parameter values (e.g., channel length, % forest cover, watershed area) are within expected ranges. For example, checking watershed area following watershed delineation can be an effective way of finding delineation errors, as extremely small values can indicate that a site's watershed did not fully delineate.

For comprehensive details and strategies for developing QA/QC protocols, it is suggested the user refer to the QA/QC Resources section of the Additional Resources Appendix.

### 3.2.2 *Metadata*

Stream assessment applications of GIS generate a wealth of highly valuable data. The CABIN database provides a secure and accessible data storage for GIS information. The ability to share geospatial data for stream assessments is invaluable as it allows participants to avoid duplication of efforts. As analysts across the country conduct analyses for stream assessment, vast amounts of data describing streams and rivers across Canada are being developed. The data must be accompanied by sufficient metadata that clearly articulates all aspects of the data to future participants to ensure their value and comparability. Metadata refers to “data about data”. Metadata provides descriptive information about temporal and spatial coverage, resolution, and data source. At a minimum, participants should identify the name and contact information of the GIS analyst, the source data used, coverage, resolution and temporal frame information, as well as short descriptions of the data fields in the output layers. The details of how to generate comprehensive and communicative metadata are outside the scope of this document. We suggest that the user refer to the Metadata Resources section of the Additional Resources Appendix.

## 4 Applying GIS to stream assessments

### 4.1 **Application #1 – Site description and site selection**

Incorporating a GIS and geospatial data describing landscapes into the site selection and description stages of a stream assessment is an enormous asset towards the development of monitoring plans as well as the identification and selection of appropriate reference sites. GIS allows the user to generate a comprehensive, large-scale description of the characteristics and conditions of a stream’s watershed area prior to field visits. Such data are crucial to ensuring that informed decisions can be made regarding the study plan and sampling.

Due to the current computing power, watershed areas of a large number of potential sampling points can be delineated and described in a relatively short amount of time. Enormous amounts of information about the landscape attributes can be gathered whether the study area is a single river basin (e.g., Thames River), an entire geographical region (e.g., East Slopes of the Rocky Mountains) or political jurisdiction (e.g., province of Nova Scotia). Patterns of landscape features (e.g., surficial geology and topography) and human activity (e.g., land use types) can be analyzed to identify sub-regions of

distinct natural or human character that may necessitate stratification of site selection and sampling. These data can be used to identify watersheds, stream segments or stream sites that are or are not exposed to human activity, and target sampling sites based on human influence or reference condition.

#### *4.1.1 Identification of reference sites*

Reference sites are sites that are minimally exposed to human disturbance. Matching test sites to an appropriate reference condition for biological assessment requires that the reference sites exhibit all the same natural attributes as the test sites to be assessed, but are minimally exposed to anthropogenic stressors, especially those that are of particular concern to the assessment (Bailey et al., 2004). Deviation from the matched reference site conditions is grounds for assuming that the test site has been affected by the human activities and associated stressors it has been exposed to. Selecting reference sites requires two things:

1. reference sites have as little exposure to human activity as possible, or at minimum, the amounts and types of activities should be fully understood, to allow for detection of impacts at test sites should those impacts exist; and
2. reference sites and test sites are as similar as possible in all attributes unrelated to human activity.

Proper application of a GIS and geospatial landscape data offer a great deal of utility in addressing both of these considerations.

GIS identifies the best available reference sites because statistical distributions of exposure to human activity can be generated from GIS-based descriptions. These distributions or “human activity gradients” are used to identify the required number of reference sites with full knowledge of the extent and intensity of the types of human activity present (Yates & Bailey, 2010a). Determining the level of detail to be used and generating strong objective criteria for what constitutes a reference site is critical prior to conducting GIS analyses.

#### *4.1.2 Locating and describing sites*

GIS is used to generate study area descriptions when the user: 1) already has a set of site locations identified; or 2) is looking to identify candidate sites (Figure 3). If sites have been pre-determined, the user only needs to have accurate geographic coordinates of each site’s location to

generate a point layer (i.e., representation of individual locations in an area). A stream network layer and a digital elevation model can be used with the point layer to generate the associated watersheds or other areas of interest (e.g., riparian corridor) for each site. Natural and anthropogenic attributes relevant to each defined watershed can be generated by conducting a series of spatial joins.

In cases where the user does not have pre-defined sites, candidate sites can be auto-generated for the entire study area and then evaluated based on watershed descriptions that meet with the objectives of the study (i.e., minimal human activity or targeted human disturbance). This process requires a high quality stream network layer to serve as a base for analysis. For example, points can be auto placed at the ends of stream segments throughout the network or in particular parts of the network (i.e., in stream segments of a desired stream order or at all stream-road crossings). Care should be taken to ensure that the stream layer is complete and that auto-generated points appear only where desired and not at artificial nodes or breaks. Once candidate sites have been identified, watersheds and associated landscape attributes can be generated as described in the previous paragraph.

#### *4.1.3 Applications of a site's landscape descriptions*

A regional description of landscape conditions allows participants to make informed decisions about site selection, sampling intensity, and study design, particularly across a large area. These data are very effective in informing random probabilistic designs where sampling sites are chosen at random, because they identify an essentially exhaustive pool of sites that represents all the possible types of landscape conditions present in the region. Similarly, such data are also very helpful for stratified designs aimed at capturing the range of natural and human conditions present in the region, as the existing range of variability can be identified and thus used to ensure that the selected sites are sufficient to capture regional conditions. The data are also useful for identifying particular landscape conditions for a more targeted selection of sites that meet the goals of the study. For example, the goal may be to find least disturbed reference sites.

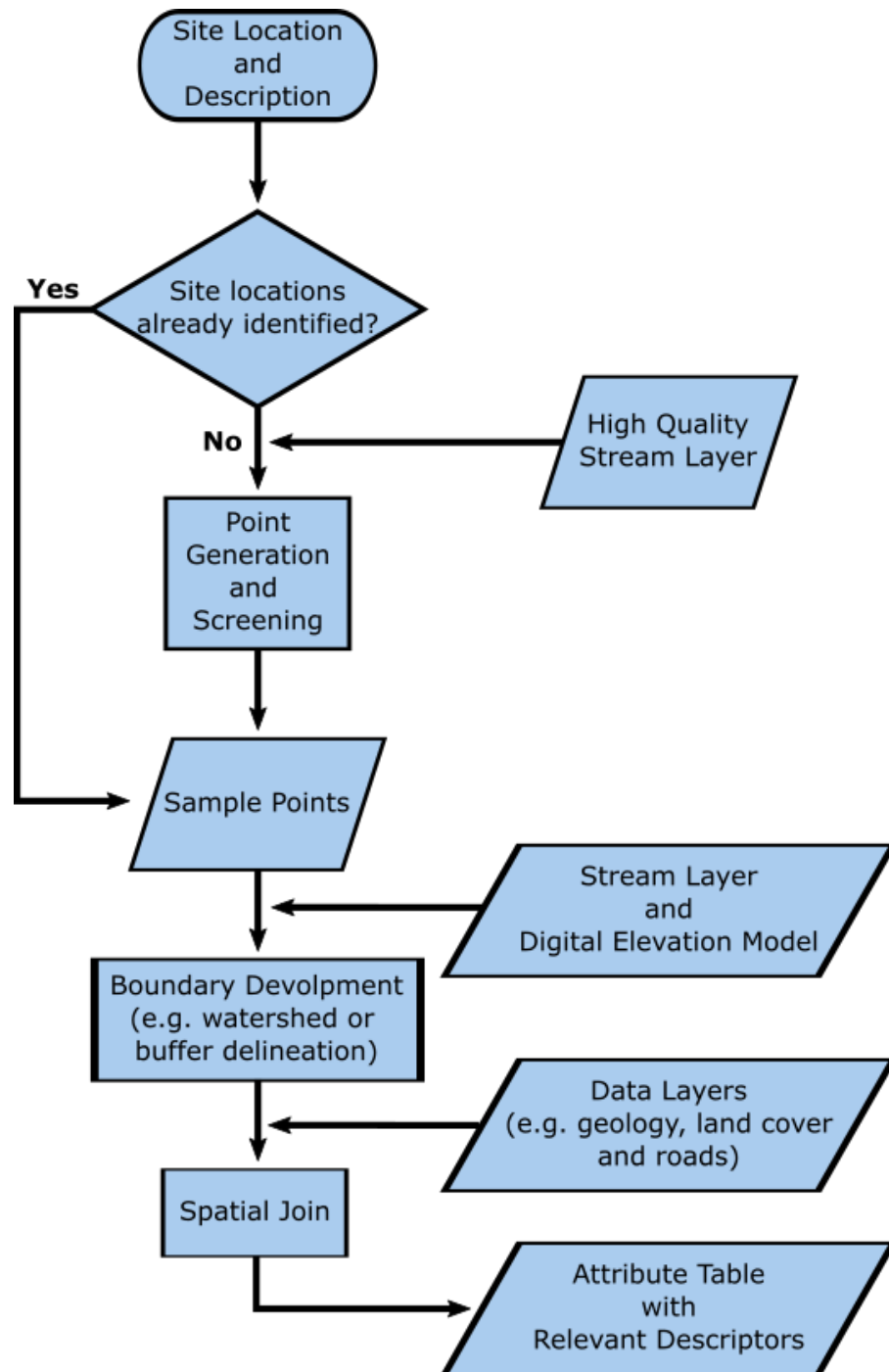


Figure 3. Steps in developing study area descriptors.

#### *4.1.4 Describing natural variation*

In order to effectively apply an RCA-based assessment such as that used by CABIN, reference sites must capture the same range of natural environmental attributes as the test sites that will be assessed. Ideally, the range of natural conditions present in the study area is identified and then reference sites are sampled throughout that range through some stratified sampling design. GIS can be used to characterize the natural environmental conditions in the study area as the first step in establishing reference sites. It is essential that only environmental attributes unaffected by human activity be used to describe the range of natural environmental conditions. For example, it is recommended that descriptions of natural environmental conditions are based on large-scale factors that are not impacted by human activities in the short-term, and that ultimately control small-scale stream conditions (e.g., channel gradient, substrate, water chemistry, stream flow) that directly impact biota, such as topography, surface and bedrock geology, as well as long-term climate norms. Descriptions of land cover, channel attributes (e.g., sinuosity) and other environmental attributes that can be altered by human activities should not be used to describe the range of natural environment conditions. Once the environmental attributes have been chosen, basic GIS processes (e.g., spatial joins and intersections) can generate descriptions for the candidate site watershed areas. Environmental conditions can be compared between test and reference sites to ensure that the full range of natural environmental conditions exhibited by test sites is also encompassed by reference sites.

#### *4.1.5 Describing human activities using GIS*

Descriptions of human activity for determining best available conditions can be as simple as describing the spatial extent (i.e., land area) of human activities using a land use layer or can be much more detailed and include descriptions of human activity using relatively complex methods of spatial analysis. A detailed discussion of the types of data and approaches that can be used to generate detailed descriptions of human activity using a GIS are beyond the scope of this document but can be found in Yates et al. (2019). For the purposes of this document, we will limit the discussion to an introduction of the key considerations of the human activity description process.

The ideal level of resolution for a given human activity will depend primarily on the nature of that activity and the spatial lens needed to observe it. For example, a general land use layer may suffice if the main goal of the assessment is to establish general condition of streams exposed to broad types of

human activity such as agriculture and urban land use. In contrast, a general land use layer will provide limited insight if the goal of an assessment is to determine the impacts of specific forestry operations, for example. In this case, detailed spatial information of the forestry activity describing the position and age of forest cut-blocks will be needed to ensure that the intensity and spatial extent descriptions are accurately meaningful.

Assessing the spatial configuration of human activities to which a stream is exposed within its watershed area can be approached in several ways. The simplest is to determine the proportion of area that an activity type covers within the whole watershed area. More spatially explicit descriptions may provide improved understanding of the exposure a stream has to a given activity. For example, it has been well established that activities taking place close to a stream or in areas where runoff is concentrated (a.k.a. hydrologically connected areas) cause a disproportionate amount of impact to streams compared to activities that are in upland areas of the watershed (Yates et al., 2014; Holmes et al., 2016; Grimstead et al., 2018). Several approaches can be taken to account for the disproportionate impact of proximity and hydrological connectedness. For example, the user can delineate additional zones within the watershed by generating buffers around the stream or segments of the stream (Figure 4 a-c). These additional zones can then be described in terms of the amount/intensity of human activity in each of the delineated zones.

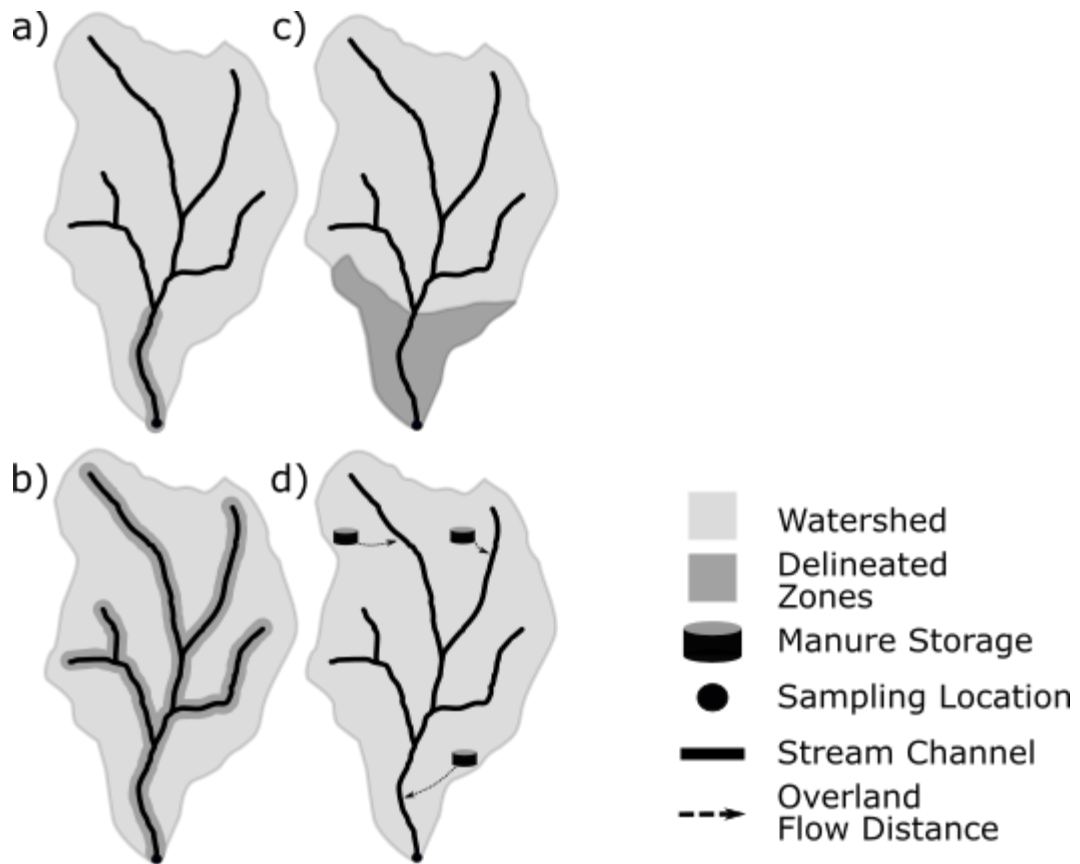


Figure 4. Approaches to account for the disproportionate impact of proximity and hydrological connectedness. For example, the user can delineate additional zones within a watershed such as a segment buffer (a), buffers among all the stream segments in the watershed (b) or a sub-watershed catchment (c). A more computationally extensive approach would be to use inverse distance approaches that take into consideration the flow distance from a pollution source to the sampling location (d).

A more computationally extensive approach to generate spatially explicit descriptions of human activities is through the use of inverse distance approaches (Figure 4 d). Although there are many methods to generate inverse weighted distances, the commonality is to establish the distances from the point in the stream where sampling is to take place within the watershed (most easily achieved using a rasterized depiction of the watershed). The inverse of the distances is then taken, thereby putting more weight on activities that are closest to the stream in explaining instream conditions. Although computationally more demanding, there is evidence that this approach is more likely to generate descriptions of human activity that best reflect the likely impacts on stream conditions (King et al., 2005; Walsh & Kunapo, 2009; Peterson et al., 2011; Yates et al., 2014). For further details on inverse distance approaches the user is directed to the following resources (see IDW Resources).



## 4.2 Application #2 – RCA model building

Stream assessments using the reference condition approach require a means of matching test sites to a group of reference sites that would be expected to have the same biological condition in the absence of impacts associated with human activities (as described in 4.1). CABIN, as well as many other RCA based biomonitoring programs, addresses the process of matching test and reference sites through the generation of a RCA model. The RCA modeling process CABIN-based assessments involves six steps (Reynoldson et al., 1995; Reynoldson et al., 1997; Reynoldson et al., 2001; Armanini et al., 2013; Strachan and Reynoldson, 2014).

1. Selection of reference sites (detailed in the previous section),
2. Collection of habitat data and benthic invertebrate data from reference sites,
3. Clustering methods to group reference sites based on similarities in benthic community composition,
4. Discriminant function analysis to identify habitat variables that can discriminate among the reference groups,
5. Prediction of the expected benthic community in each reference group, and
6. Assignment of test sites to an appropriate reference group using the predictive model

GIS plays an integral role in RCA model building through its use for the acquisition of landscape-level habitat variables describing reference sites (Step 2). The application of GIS to RCA model building is similar to the process described for the selection of candidate reference sites (section 4.1), and many of the same considerations apply, as do the processes for generating those descriptions.

### 4.2.1 *Identifying landscape attributes that discriminate community groups*

All RCA models involve the collection of habitat variables from multiple scales, such as stream slope at the stream segment scale and geology at the watershed scale. Ensuring that habitat attributes are described at the scale that is most relevant to an attribute of interest increases the likelihood that the role that attribute plays in controlling the local environmental conditions are captured. Improved descriptions of the landscape will in turn enhance the establishment of sets of environmental conditions that discriminate biological reference groups (see Step 4 above). To ensure that landscape attributes are described at the most appropriate scale, the user should develop conceptual models describing how a landscape descriptor, such as bedrock type, is expected to influence local stream conditions and ultimately the benthic community. For example, topography would be expected to influence the

gradient of the channel, thereby determining channel form and stream velocities, which in turn would influence the size distribution of stream substrate and thereby determine which biota are likely to be present. Based on this example conceptualization, topography of the stream segment upstream of the site, measured as slope, is likely to be a better predictor than the average slope of the entire watershed. Such conceptualizations should be completed for all potential descriptors prior to GIS analyses being conducted to ensure that the parameters outputted from GIS analyses meet the needs of the associated stream assessment. Additionally, using landscape-level habitat information to match test sites to a group of reference sites (see Step 6 above) is critical for all RCA models used by CABIN.

#### *4.2.2 Temporal frame considerations for GIS applications to RCA models*

Building on the considerations of spatial scale and influence of human activity described in the previous paragraphs, the temporal scale of geospatial data should also be considered when selecting variables. Geospatial data that has high temporal variability may not be appropriate for model construction. For example, percent canopy cover in temperate zones can be highly variable dependent on the season and may not properly reflect reference conditions. Likewise, land cover can change drastically over relatively short time periods (e.g., years) and thus raises issues of matching the temporal scope of the geospatial data with that of stream sampling. Thus, only geospatial data that is not inherently variable over short time (i.e., less than 5 years) periods be used for RCA models. Following this recommendation will eliminate the need to temporally match environmental descriptors with sampling dates and reduce concerns about whether test sites collected in the future can be matched with reference sites from the past. Geospatial data describing landscape attributes that vary over decades, such as long-term climate averages, can be more readily applied. Even following this advice on temporal frame does not completely eliminate the need to check and update all habitat attributes used in CABIN RCA models over time. Given that the detail and speed at which geospatial data are collected is continually increasing and landscape descriptions updates should be undertaken if more spatially or temporally appropriate geospatial data becomes available, thereby ensuring the best possible models are being used for stream assessment.

#### *4.2.3 Use of standardized data for model building*

There are potential advantages and disadvantages of using geospatial data layers that have national coverage for use by CABIN RCA models. The advantages of generating assessment models using a set of national level descriptors are primarily that the models generated from such descriptors can be

applied over larger spatial scales that are likely to cross provincial and territorial boundaries allowing for national consistency, an important component of CABIN. Models built in this way will, however, be more limited to a small suite of geospatial datasets, particularly, nationally-scaled surface and bedrock geology, climate, topography (from DEMs) and hydrologic network information that have nationally scaled data (Table 2). Although using such large-scale layers has advantages in terms of model transferability across political boundaries, there is the potential that using such layers will result in poorer model prediction as these larger scale layers have poor resolution compared to many provincial and regional datasets. It is likely that there will be more generalization of the landscape descriptions and a greater chance that classifications do not fully reflect the character of a stream's watershed. As with all GIS analyses, participants should be guided by their assessment goals in determining whether use of national layers is a good fit for their assessment.

Table 2. List of national level geospatial data sources available that could be used for generation of standardized reference condition assessment models.

Name	Descriptor	Scale or resolution	Source
Bedrock Geological Map of Canada	Bedrock	1:5,000,000	Natural Resources Canada – Open Canada Portal
Climate Normals (1980 – 2010)	Climate	7.5 km	Natural Resources Canada – available on request
National Hydro Network	Hydrology	1:50,000	Natural Resources Canada - Geogratis Canada
Shuttle Radar Topography Mission V. 3	Topography (DEM)	30 m	NASA - EarthData Search
Surficial Materials of Canada	Surficial Geology	1:5,000,000	Natural Resources Canada – Open Canada Portal

### 4.3 Application #3 – Assessment interpretation

Site assessments using an RCA predictive model provide an indication as to whether or not a test site’s biological condition deviates significantly from conditions observed at comparable reference sites. Current RCA approaches do not provide an indication of the probable cause of a site’s failure or deviation. Rather, it is up to the user to conduct follow-up analyses to determine the likely cause of a site’s deviation from reference so that appropriate stream management actions can be taken. Data derived from a GIS can help inform such analyses.

The ability of GIS to provide high quality descriptions of the amounts and types of human activities for the entire watershed area is a critical asset for assessment interpretation, as it allows for the completion of a comprehensive summary of the activities by which the stream biota may have been impacted. Interpreting assessment outcomes requires many of the same data used to select test and reference sites, and thus many of the same considerations regarding the resolution of geospatial data apply in addition to the spatial extent and resolution. However, the main difference stems from the goal of the application, which is to determine probable cause of a site’s deviation from reference condition. It is likely that the most appropriate data and the process of summarizing that data will differ from that used to conduct site selection procedures.

As part of the assessment interpretation process, we recommend that participants first generate a conceptual model that establishes the likely pathways through which human activities present in the study site's watershed area may be influencing biological conditions (Figure 5). Such a model is best started by reviewing the types and general locations of the various human activities present in the watershed using GIS. If the site was part of a site selection initiative then the data may already be generated; however, if a site selection process has not been undertaken, we recommend that the user review the information in the section above (GIS APPLICATION #1). To review, the site selection process is discovery-based with the aim to generate a broad description of the types and amounts of human activities present and the relative positions of these activities in the watershed. Basic land use/land cover data layers can be used to determine the general types of activities present (e.g., urban areas, agricultural lands, mining) in a study area. Once the different types of activities present in the watershed have been identified, the user can then seek out more detailed geospatial data regarding specific activity types (e.g., forest cut-block information, agricultural crop types) to better resolve the specific management practices being undertaken. In general, it is recommended that when aiming to interpret stream assessment results, it is better to have more detailed geospatial descriptions of the activities present in the watershed, as it will be possible to make more specific connections between the activity and the impact, leading to more targeted future management actions.

An increased understanding of the relative position of human activities in a stream's watershed will also help target management actions. Human activities are rarely evenly distributed across a landscape. Understanding the relative positions of different activities will also help determine the probable cause of a test site falling outside of reference conditions. Spatial configurations can be established in GIS as described in the sections above (see GIS application #1) and will allow the user to refine hypotheses about which pathways in their conceptual model are likely to be most important in determining the observed biological condition.

A detailed, geospatial description of the types, amount, nature, and position of human activity is a powerful complement to the site-level habitat descriptions generated as part of the CABIN sampling protocol (e.g., substrate size, water chemistry, channel shape). Many human activities influence stream environments through a set of specific pathways and site-level habitat data can be used to establish which human activities are most likely associated with the impact. For example, agricultural activity can be associated with channel straightening, which would be reflected in measures of channel sinuosity or variation in channel width and could be expected to lead to changes in biological condition (i.e., reduced

taxa richness). Connections made through a conceptual model process as described here are the best hypotheses of the likely causes. Further assessment of the biological stream condition would be needed to determine the cause of impairment.

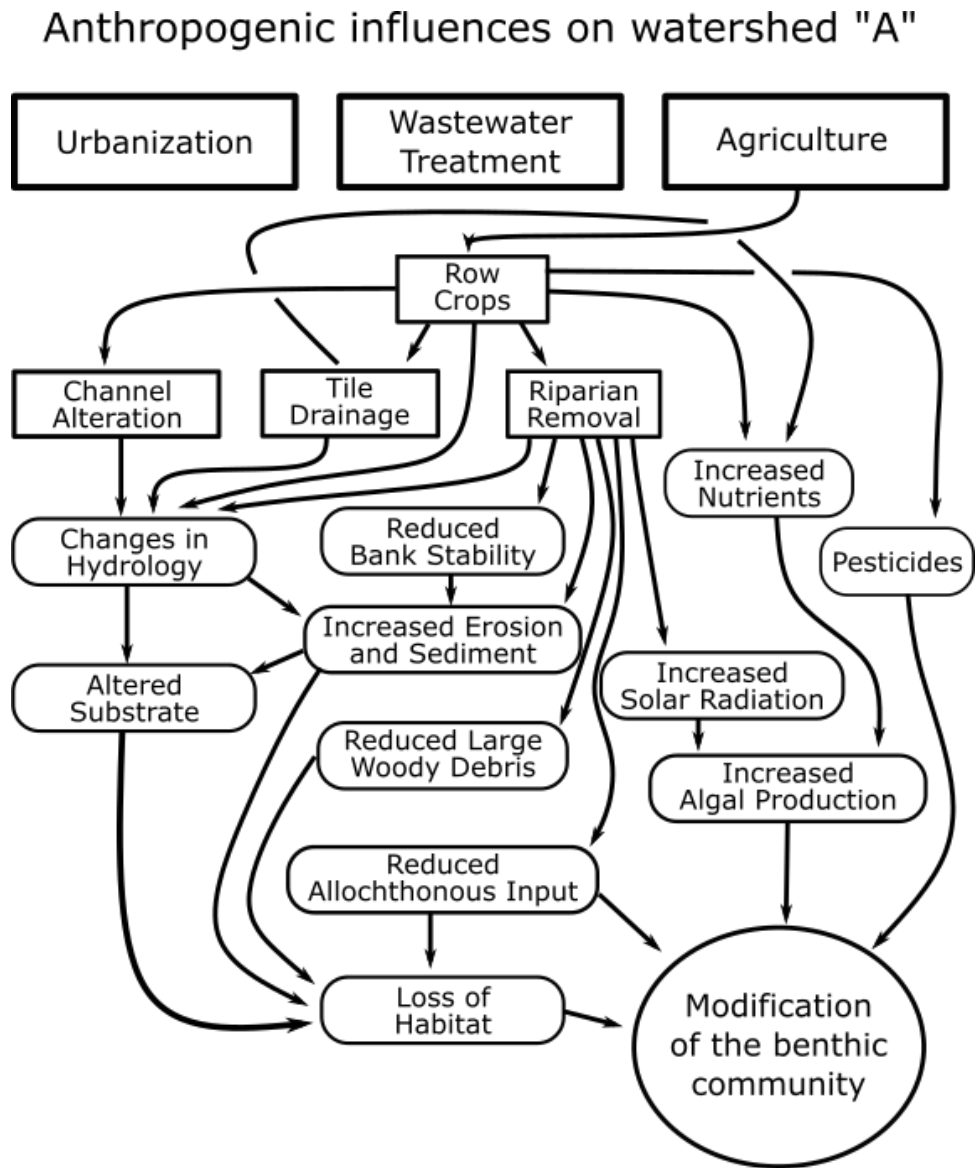


Figure 5. Example concept map representing the potential anthropogenic influences in watershed "A" with a focus on agricultural activities specific to the watershed and the connection to the benthic community.

## 5 Summary

GIS is a powerful tool that can greatly enhance several aspects of stream bioassessment as currently practiced as part of the CABIN protocols. This document and the guidance presented herein is intended to be used in coordination with the sampling, modelling and assessment protocols already in place as part of CABIN. In particular, this document has described how a GIS can be used to characterize and select sites for more powerful study designs and more objective definitions of reference sites. Likewise, GIS is an invaluable tool for generating descriptions of large-scale habitat attributes than can be used in RCA models to effectively match test sites with appropriate reference sites. Finally, GIS can strengthen post-assessment interpretation by providing nuanced information regarding the amount and location of human activities that may be affecting a site's condition.

Like all tools, GIS does have its limitations and the user should always be cognizant of these limits. First and foremost, the results of GIS analyses are only as good as the geospatial data used for the analysis, and often geospatial data availability is the most limiting factor in what information can be gathered using a GIS. Second, although many analyses can be conducted, and numerous output parameters can be generated, the user should ensure that all processes and generated parameters have a clear conceptual underpinning that links the output to the assessment questions at hand. As such, we strongly encourage all participants to construct basic conceptual models prior to conducting the GIS applications described above, to best ensure that the analyses and output are strongly linked to the goals of the assessment.

## 6 References

- Allan, J. D., 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems. *Annual Review of Ecology Evolution and Systematics* 35: 257–284.
- Armanini, D. G., W. A. Monk, L. Carter, D. Cote, & D. J. Baird, 2013. Towards generalised reference condition models for environmental assessment: a case study on rivers in Atlantic Canada. *Environmental Monitoring and Assessment* 185: 6247–6259.
- Burt, T. P., & G. Pinay, 2005. Linking hydrology and biogeochemistry in complex landscapes. *Progress in Physical Geography* 29: 297–316.
- Chrisman, N. R., 1999. What does 'GIS' mean? *Transactions in GIS* 3: 175–186.
- Coetzee, S., I. Ivánová, H. Mitasova, & M. A. Brovelli, 2020. Open geospatial software and data: a review of the current state and a perspective into the future. *ISPRS International Journal of Geo-Information Multidisciplinary Digital Publishing Institute* 9: 90.
- Frissell, C., W. Liss, C. Warren, & M. Hurley, 1986. A hierarchical framework for stream habitat classification - viewing streams in a watershed context. *Environmental Management* 10: 199–214.
- Grimstead, J. P., E. M. Krynak, & A. G. Yates, 2018. Scale-specific land cover thresholds for conservation of stream invertebrate communities in agricultural landscapes. *Landscape Ecology* 33: 2239–2252.
- Holmes, R., D. G. Armanini, & A. G. Yates, 2016. Effects of best management practice on ecological condition: does location matter? *Environmental Management* 57: 1062–1076.
- Hynes, H. B. N., 1975. The stream and its valley. *SIL Proceedings, 1922-2010 Informa UK Limited* 19: 1–15.
- Johnson, L., & S. Gage, 1997. Landscape approaches to the analysis of aquatic ecosystems. *Freshwater Biology* 37: 113–132.
- Kattwinkel, M., E. Szöcs, E. Peterson, & R. B. Schäfer, 2020. Preparing GIS data for analysis of stream monitoring data: The R package openSTARS. *PLOS ONE Public Library of Science* 15: e0239237.



- King, R. S., M. E. Baker, D. F. Whigham, D. E. Weller, T. E. Jordan, P. F. Kazyak, & M. K. Hurd, 2005. Spatial considerations for linking watershed land cover to ecological indicators in streams. *Ecological Applications* 15: 137–153.
- Osborne, L. L., & M. J. Wiley, 1988. Empirical relationships between land use/cover and stream water quality in an agricultural watershed. *Journal of environmental management* 26: 9–27.
- Peterson, E. E., F. Sheldon, R. Darnell, S. E. Bunn, & B. D. Harch, 2011. A comparison of spatially explicit landscape representation methods and their relationship to stream condition. *Freshwater Biology* 56: 590–610.
- R Core Team, 2020. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>.
- Reynoldson, T.B., R.C. Bailey, K.E. Day, & R.H. Norris, 1995. Biological guidelines for freshwater sediment based on Benthic Assessment of Sediment (the BEAST) using a multivariate approach for predicting biological state. *Australian Journal of Ecology* 20: 198-219.
- Reynoldson, T. B., R. H. Norris, V. H. Resh, K. E. Day, & D. M. Rosenberg, 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. *Journal of the North American Benthological Society* 16: 833–852.
- Reynoldson, T.B., D.M. Rosenberg, & V.H. Resh, 2001. Comparison of models predicting invertebrate assemblages for biomonitoring in the Fraser River catchment, British Columbia. *Canadian Journal of Fisheries and Aquatic Sciences* 58:1395-1410.
- Richards, C., & G. Host, 1994. Examining land use influences on stream habitats and macroinvertebrates: a GIS approach. *Journal of the American Water Resources Association* 30: 729–738.
- Ritters, K. H., & J. D. Wickham, 1995. A landscape atlas of the Chesapeake Bay watershed. Naval Research Lab Stennis Space Center MS Coupled Dynamic Processes Section, <https://apps.dtic.mil/sti/citations/ADA350198>.
- Samal, A., S. Seth, & K. Cueto, 2004. A feature-based approach to conflation of geospatial sources. *International Journal of Geographical Information Science* Taylor & Francis 18: 459–489.

- Steiniger, S., & A. J. S. Hunter, 2013. The 2012 free and open-source GIS software map - a guide to facilitate research, development, and adoption. *Computers Environment and Urban Systems* Elsevier Sci Ltd, Oxford 39: 136–150.
- Strachan, S.A., & T.B. Reynoldson, 2014. Performance of the Standard CABIN Method: Comparison of BEAST Models and Error Rates to Detect Simulated Degradation from Multiple Data Sets. *Freshwater Science* 33:1225-1237.
- Thoms, M., M. Scown, & J. Flotemersch, 2018. Characterization of river Networks: a GIS approach and Its applications. *Journal of the American Water Resources Association* 54: 899–913.
- Tsihrintzis, V. A., R. Hamid, & H. R. Fuentes, 1996. Use of Geographic Information Systems (GIS) in water resources: a review. *Water Resources Management* 10: 251–277.
- Vidon, P., C. Allan, D. Burns, T. P. Duval, N. Gurwick, S. Inamdar, R. Lowrance, J. Okay, D. Scott, & S. Sebestyen, 2010. Hot spots and hot moments in riparian zones: potential for improved water quality management. *JAWRA Journal of the American Water Resources Association* 46: 278–298.
- Walsh, C. J., & J. Kunapo, 2009. The importance of upland flow paths in determining urban effects on stream ecosystems. *Journal of the North American Benthological Society* 28: 977–990.
- Walsh, C. J., A. H. Roy, J. W. Feminella, P. D. Cottingham, P. M. Groffman, & R. P. Morgan, 2005. The urban stream syndrome: current knowledge and the search for a cure. *Journal of the North American Benthological Society* 24: 706–723.
- Yates, A., J. Culp, R. Bailey, & P. Chambers, 2019. Strengths and weaknesses of data sources for describing exposure of aquatic ecosystems to human activity In Hughes, R. M., D. M. Infante, K. Chen, L. Wang, & B. de F. Terra (eds), *Advances in Understanding Landscape Influences on Freshwater Habitats and Biological Assemblages*. American Fisheries Society, <https://doi.org/10.47886%2F9781934874561.ch2>.
- Yates, A. G., & R. C. Bailey, 2010a. Selecting objectively defined reference sites for stream bioassessment programs. *Environmental Monitoring and Assessment* 170: 129–140.
- Yates, A. G., & R. C. Bailey, 2010b. Improving the description of human activities potentially affecting rural stream ecosystems. *Landscape Ecology* 25: 371–382.

Yates, A. G., R. B. Brua, J. Corriveau, J. M. Culp, & P. A. Chambers, 2014. Seasonally driven variation in spatial relationships between agricultural land use and in-stream nutrient concentrations. *River Research and Applications* 30: 476–493.

## 7 Glossary

---

<b>Anthropogenic gradient</b>	The entire range of a given human activity (e.g., % agricultural land use) within an area of interest
<b>Attribute</b>	Non-spatial information about a feature stored in a table and linked to the feature
<b>Coverage</b>	The total area described by a specific geospatial dataset
<b>Data conflation</b>	The combining or reconciliation of two overlapping geospatial datasets
<b>Data Stratification</b>	Sorting data into distinct groups based on similarity of attributes
<b>Delineate</b>	To geospatially describe an attribute's boundary or border (e.g., watershed)
<b>Digital elevation model (DEM)</b>	The digital characterization of surface topography
<b>Geographic Information Systems (GIS)</b>	A computer system for capturing, assembling, manipulating, analyzing, storing and displaying geographically-referenced information
<b>Geospatial data</b>	Data that are associated with a specific geographic location
<b>Human activity gradient (HAG)</b>	A set of ecosystems (e.g., reaches, basins, or other geographical units of interest) that vary in their exposure to human activities (Yates & Bailey, 2010b)
<b>Hydrograph</b>	A graphical display of stream discharge over a set period of time.
<b>Inter-operator variability</b>	The amount of variation resulting from data collected by two different individuals.
<b>Intersection</b>	The selection of attributes from one data layer based on the location that features bisect in two or more data sources.
<b>Metadata</b>	Information about data. Metadata often contains, but is not limited to, information on the data origin, date of composition, quality, projection, scale, resolution and attribute descriptions.
<b>Natural gradient</b>	The entire range of a given natural attribute of the landscape (e.g., surface geology, stream size) within an area of interest
<b>Open-source</b>	Software that has the source code openly available so that the user can modify, copy, and share

---

<b>Probabilistic design</b>	Sampling sites are chosen at random throughout the area of interest
<b>Quality assurance</b>	A process or processes used to ensure data meet data-quality objectives and to prevent defects
<b>Quality control</b>	A process or processes used to ensure products meet overall quality goals and criteria
<b>Reach</b>	A length of stream in which stream assessment is to take place. In practice, any length of stream as defined by the user
<b>Reference site</b>	A site that is minimally exposed to human activities
<b>Resolution</b>	The smallest difference between adjacent positions that can be recorded. Higher resolution indicates more detail is detectable
<b>Scale</b>	The relationship between the distance on a map and the corresponding distance in the real world
<b>Segment</b>	A length of stream located longitudinally between two stream confluences
<b>Sinuosity</b>	The ratio of stream length to valley length
<b>Spatial join</b>	A function that appends attributes from one feature layer to another based on their spatial relationship
<b>Stratified design</b>	Sampling sites are placed in groups based on similar environmental attributes representing the entire range of natural and human conditions present in the region
<b>Stream order</b>	A method for indicating the size of stream segments starting with first order as the smallest designation and increasing in order as segments of equal size come together (e.g., two second order streams converge to become a third order stream)
<b>Test site</b>	A site that may be impacted by human activity and is the subject of a stream assessment
<b>Watershed</b>	An area of land bounded peripherally by a divide and draining ultimately to a particular watercourse

## 8 Quick reference guide

### 8.1 Finding geospatial data

1. Most provinces and territories have geospatial available for download as does the Federal government of Canada
2. Make a data search plan prior to looking for geospatial data based on what the ideal data are for the stream assessment goals
3. The perfect data may not be available but you will likely find data that fits the needs of your project

### 8.2 Selecting geospatial data

1. Ensure selected data have coverage of the entire spatial extent of the study area of interest
2. Ensure selected data have a temporal frame that is relevant to the period of study
3. The resolution of the selected data should be sufficient to portray differences in landscape character that are relevant to the stream assessment goals
4. Coarser resolution data may need to be used to balance needs between coverage and resolution particularly when study areas cross political boundaries

### 8.3 Data QA/QC

1. QA/QC protocols should be in place for all stages of data selection, use and reporting
2. Check the quality of the data prior to using – checks should assess that the dataset is complete, accurate and as advertised
3. QA/QC the data following each GIS process to ensure errors are corrected before they propagate
4. Review data outputs to ensure that parameter values meet expectations of the data being analysed

### 8.4 Metadata

1. Metadata should be generated for any GIS data produced for use in a stream assessment
2. Metadata should include information about the analyst who generated the data, the original source data (e.g., resolution, coverage, temporal frame) and descriptions of the new data (e.g., definition of field names)

## **8.5 Assessment site selection and description**

1. GIS can be used to auto identify candidate stream sites that meet specific criteria (e.g., at a stream-road crossing)
2. GIS can help ensure that the best available reference sites are identified because statistical distributions of exposure to human activity can be generated from GIS-based descriptions
3. The ideal level of resolution of description for a given human activity will depend primarily on the nature of that activity and the spatial lens that needed to observe it
4. Using spatially descriptive approaches (e.g., subwatershed zones, inverse weighted distance models) provide improved understanding of the likely exposure a stream has to a given activity.
5. Using GIS to generate comprehensive descriptions of natural environmental conditions is a powerful approach for establishment of a stratified sampling plan

## **8.6 Reference Condition Approach model building**

1. GIS is a practical means of acquiring large-scale descriptions of reference sites that are unaffected by human activity
2. Ensure that described geospatial attributes will not be affected by human activities present at test sites before using them in model building exercises
3. Using geospatial data that are not inherently variable over short time periods will eliminate the need to temporally match environmental descriptors with stream sampling dates and reduce the frequency with which models will need to be updated because of temporal changes in the geospatial predictors used
4. Model predictors derived from GIS should be regularly reviewed, and updated as necessary, to ensure that the data are still the most appropriate geospatial information available to meet the goals of the stream assessment
5. Using a set of national level geospatial data layers to generate RCA models has the advantage that resultant models can be applied over larger spatial scales and across political boundaries

## **8.7 Assessment interpretation**

1. GIS can provide high quality descriptions of the amounts and types of human activities allowing for completion of a comprehensive summary of the activities that the stream biota may be impacted by

2. A conceptual model establishing likely pathways through which human activities in the study site's watershed may be influencing biological conditions should inform which geospatial data to use to interpret stream assessment results
3. Geospatial data that provide the greatest amount of detail regarding human activities are most recommended to interpret likely causes of stream assessment results

## 9 Additional resources

### 9.1 GIS resources

#### 9.1.1 *Introduction to GIS and GIS software - books*

Bolstad, P., 2019. GIS Fundamentals: A First Text on Geographic Information Systems, NEW and UPDATED Sixth Edition. XanEdu Publishing Inc.

Brunsdon, C., & L. Comber, 2019. An Introduction to R for Spatial Analysis and Mapping. SAGE Publications Ltd, Thousand Oaks, CA.

Cutts, A., & A. Graser, 2018. Learn QGIS: Your step-by-step guide to the fundamental of QGIS 3.4, 4th Edition. Packt Publishing.

Kwast, H. V. D., & K. Menke, 2019. QGIS for Hydrological Applications: Recipes for Catchment Hydrology and Water Management. Locate Press.

Law, M., & A. Collins, 2021. Getting to Know ArcGIS Desktop 10.8. Esri Press.

MacLeod, C. D., 2015. GIS For Biologists: A Practical Introduction For Undergraduates. Pictish Beast Publications, Glasgow.

Shellito, B. A., 2017. Discovering GIS and ArcGIS. WHFreeman, New York, NY.

Wegmann, M. J. Schwalb-Willmann, & S. Dech, 2020. Introduction to Spatial Data Analysis. Pelagic Publishing.

Wegmann, M., B. Leutner, & S. Dech (eds), 2015. Remote Sensing and GIS for Ecologists: Using Open-source Software. Pelagic Publishing, Exeter.



### 9.1.2 Introduction to GIS and GIS software - online resources

Bivand, R., 2021. CRAN Task View: Analysis of Spatial Data. <https://CRAN.R-project.org/view=Spatial>

Campbell, J. E., & M. Shin, 2011. Essentials of Geographic Information Systems. Saylor Foundation. Open Textbook Library. <https://open.umn.edu/opentextbooks/textbooks/67>

Geospatial Analysis - spatial and GIS analysis techniques and GIS software. 2021. <https://spatialanalysisonline.com/index.html>

GRASS GIS. <https://grass.osgeo.org/>

Lansley, G., & J. Cheshire, 2016. An Introduction to Spatial Data Analysis and Visualisation in R | CDRC Data. <https://data.cdrc.ac.uk/dataset/introduction-spatial-data-analysis-and-visualisation-r>

QGIS Training Manual — QGIS Documentation. [https://docs.qgis.org/3.16/en/docs/training\\_manual/](https://docs.qgis.org/3.16/en/docs/training_manual/)

Sadler, J. Introduction to GIS with R. Jesse Sadler. <https://jessesadler.com/post/gis-with-r-intro/>

Spatial Data Science with R — R Spatial. <https://www.rspatial.org/>

Sutton, T., O. Dassau, & M. Sutton, 2021. A Gentle Introduction to GIS. [https://docs.qgis.org/3.16/en/docs/gentle\\_gis\\_introduction/index.html](https://docs.qgis.org/3.16/en/docs/gentle_gis_introduction/index.html)

### 9.1.3 GIS software and tools

Bivand, R., 2021. CRAN Task View: Analysis of Spatial Data. <https://CRAN.R-project.org/view=Spatial>

Esri - Environmental Systems Research Institute. <https://www.esri.com/en-us/home>

GRASS GIS. <https://grass.osgeo.org/>

Kattwinkel, M., & E. Szöcs, 2020. openSTARS: An open-source implementation of the “ArcGIS” Toolbox “STARS.” <https://CRAN.R-project.org/package=openSTARS>

OSGeo - The Open-source Geospatial Foundation. <https://www.osgeo.org/>

QGIS. <https://www.qgis.org/en/site/>

R: The R Project for Statistical Computing. <https://www.r-project.org/>

## 9.2 QA/QC

*ArcGIS Geodatabase Topology Rules.*

[https://help.arcgis.com/en/arcgisdesktop/10.0/help/001t/pdf/topology\\_rules\\_poster.pdf](https://help.arcgis.com/en/arcgisdesktop/10.0/help/001t/pdf/topology_rules_poster.pdf)

Bolstad, P., 2019. GIS Fundamentals: A First Text on Geographic Information Systems, NEW and UPDATED Sixth Edition. XanEdu Publishing Inc.

Johnson, M., & M. Mozingo. QA/QC for your GIS data. <https://www.esri.com/content/dam/esrisites/en-us/about/events/media/UC-2019/technical-workshops/tw-6391-1016.pdf>

Pascual, P. S., 2011. GIS Data: A Look at Accuracy, Precision, and Types of Errors. GIS Lounge. <https://www.gislounge.com/gis-data-a-look-at-accuracy-precision-and-types-of-errors/>

Rozenfeld, N., 2013. How to Check Your GIS Data. GIS Lounge. <https://www.gislounge.com/check-gis-data/>

Smith, S., & J. Cary, 2010. Developing a quality assurance plan. ArcGIS Blog. <https://www.esri.com/arcgis-blog/products/data-reviewer/data-management/developing-a-quality-assurance-plan/>

Srivastava, R. N., 2008. Spatial Data Quality: An Introduction. GIS Lounge. <https://www.gislounge.com/spatial-data-quality-an-introduction/>

United States Geological Survey. Data Management - Manage Quality. <https://www.usgs.gov/products/data-and-tools/data-management/manage-quality>

## 9.3 Metadata

Bolstad, P., 2019. GIS Fundamentals: A First Text on Geographic Information Systems, NEW and UPDATED Sixth Edition. XanEdu Publishing Inc.

Natural Resources Canada. Digital Geospatial Metadata. <https://www.nrcan.gc.ca/earth-sciences/geomatics/canadas-spatial-data-infrastructure/standards-policies/8912>

FGDC Technical Guidance — Federal Geographic Data Committee. <https://www.fgdc.gov/technical-guidance>

United States Geological Survey. Formal metadata: information and software.  
<https://geology.usgs.gov/tools/metadata/>

United States Geological Survey. Data Management - Metadata Creation.  
<https://www.usgs.gov/products/data-and-tools/data-management/metadata-creation>

#### **9.4 Inverse Distance Weighting (IDW)**

Peterson, E. E., F. Sheldon, R. Darnell, S. E. Bunn, & B. D. Harch, 2011. A comparison of spatially explicit landscape representation methods and their relationship to stream condition. *Freshwater Biology* 56: 590–610.

Staponites, L. R., V. Barták, M. Bílý, & O. P. Simon, 2019. Performance of landscape composition metrics for predicting water quality in headwater catchments. *Scientific Reports Nature Publishing Group* 9: 14405.

Yates, A. G., R. B. Brua, J. Corriveau, J. M. Culp, & P. A. Chambers, 2014. Seasonally driven variation in spatial relationships between agricultural land use and in-stream nutrient concentrations. *River Research and Applications* 30: 476–493.