

**Canadian Environmental Sustainability Indicators  
Water Quality Index Representivity Report**

April 8, 2009



**By: Mark Henry, François Souldard and Hugo Larocque  
Environment Accounts and Statistics Division  
Statistics Canada**

# Table of Contents

<b>1</b>	<b>Introduction.....</b>	<b>3</b>
<b>2</b>	<b>Delineation Datasets and Applications.....</b>	<b>3</b>
<b>3</b>	<b>Drainage Basin Delineation.....</b>	<b>3</b>
	3.1 Inputs.....	4
	3.2 Selecting upstream networks.....	4
	3.3 Creating Thiessen polygons and boundaries.....	4
	3.4 Extending boundaries beyond single NHN Work Unit.....	5
<b>4</b>	<b>Compiling the statistics.....</b>	<b>6</b>
	4.1 Drainage basin statistics.....	6
	4.2 Human activity - settlement.....	7
	4.3 Human activity - industry.....	7
	4.4 Human activity - agriculture.....	7
	4.5 Natural coverage.....	7
<b>5</b>	<b>Quality Control.....</b>	<b>7</b>
	5.1 Process quality control.....	7
	5.2 Quality control indicators.....	8
<b>6</b>	<b>Final Delivery.....</b>	<b>8</b>
<b>7</b>	<b>Future Research, Directions and Recommendations.....</b>	<b>8</b>
<b>8</b>	<b>Acknowledgement.....</b>	<b>8</b>
	<b>Appendix ‘A’ - Determining Representivity.....</b>	<b>9</b>

# 1 Introduction

This paper describes the methodology developed and implemented by Statistics Canada to create datasets that Environment Canada (EC) will use to evaluate the Canadian Environmental Sustainability Indicators (CESI) water quality network. This network is the source of data that is used to calculate the CESI Water Quality Indicator (WQI) and about which there remain some questions as to the degree to which it is representative of a national network. EC has therefore delegated to Statistics Canada (Statcan) authority for the enhancement of the CESI project by generating statistics, including geo-spatial statistics, that allow for the analysis of the network.

This project involved delineating upstream basins for each WQI site and then finding relevant, accessible and easily compiled socio-economic and bio-physical statistics to allow for the categorization of water quality sites. This work is intended to facilitate the efficient selection of a sub-set of sites that are minimally redundant and that represent a broad set of characteristics deemed representative of Canadian hydrology.

The goals of this analysis are to:

1. improve the scientific knowledge and reduce the limitations in the interpretation of the WQI and the monitoring network by characterizing monitoring sites by various statistics;
2. support development of sub-indices based upon different types of water quality parameters; and
3. support identification of important geo-spatial data gaps.

To meet these goals Statcan was contracted to do the following:

1. Review the representivity criteria (in collaboration with EC) and define a methodology for the geo-spatial analysis of 447 WQI sites;
2. Undertake a test using federal and federal-provincial sites (i.e. 77 sites);
3. Delineate the upstream drainage area for each site and overlay various spatial coverages; and
4. Assist in identifying geo-spatial data gaps where new sites could be located (according to criteria).

## 2 Delineation Datasets and Applications

### Data

Natural Resources Canada's National Hydro Network (NHN)

Linear flow layer – rivers and streams

Index Layer – NHN work unit and sub-sub drainage area boundaries

<http://www.geobase.ca/geobase/en/data/nhn/description.html>

WQI site layer – 459 sites

### Applications

ArcMap 9.3

SAS

MSAccess

## 3 Drainage Basin Delineation

The first and most time consuming step of this project was the development and implementation of a methodology for delineating the upstream catchment areas of 447 of the 459 sites identified in the WQI network. Somewhat constrained by time and fiscal limitations set out for this project, an efficient and repeatable methodology was required that allowed for the reasonable categorization of sites based upon an agreed set of criteria. More time intensive delineation methodologies were considered, such as those involving the use of digital elevation models, but were eliminated because of the time and resources required for only marginal improvements.

Twelve sites were deemed out of scope of the delineation process because they were located on the St. Lawrence, at the mouth of the Mackenzie or because there was no NHN flow data for their basin. It was anticipated that those on the St. Lawrence and at the mouth of the Mackenzie could be categorized without being processed.

The following steps outline the methodology developed to complete the delineation of the upstream watersheds portion of this project.

### 3.1 Inputs

The main datasets used to delineate the drainage basin boundaries were the flow and work unit layers from the NHN. These datasets were chosen because they were deemed to be part of the most current, accurate and comprehensive national hydrological dataset and they covered almost all areas required to delineate the upstream basins of EC's WQI network. The flow layer also contains directionality in its topology, an efficient vehicle for tracing upstream networks.

### 3.2 Selecting upstream networks

The first step in the delineation process involved selecting and tracing the upstream hydrological network for each site. To accomplish this we did an overlay of the WQI sites layer onto the NHN flow layer, using ArcMap, and selected the nearest arc on that network to serve as a starting point for tracing upstream. The name and description of the sites were also used to corroborate locations.

From this starting point, each upstream network was traced using a SAS application and the directionality contained in the topology of the NHN. These upstream networks became an input to the Thiessen polygon tool of ArcMap, used for delineating the basin boundaries.

In addition to the upstream network, the closest parts of the adjacent networks were selected and input into the same process to create Thiessen polygons between the upstream network and the adjacent networks. To do this, a buffer was created around the upstream network and a selection was made of the appropriate portions of the adjacent networks for input into the Thiessen polygon process. See Figures 1 and 2.

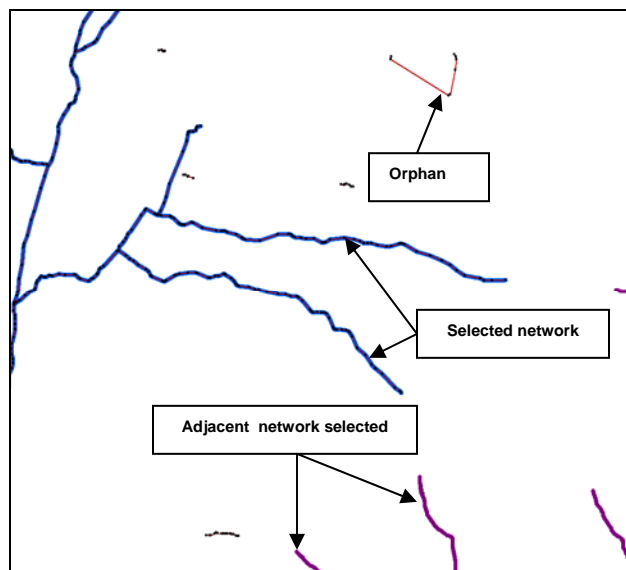
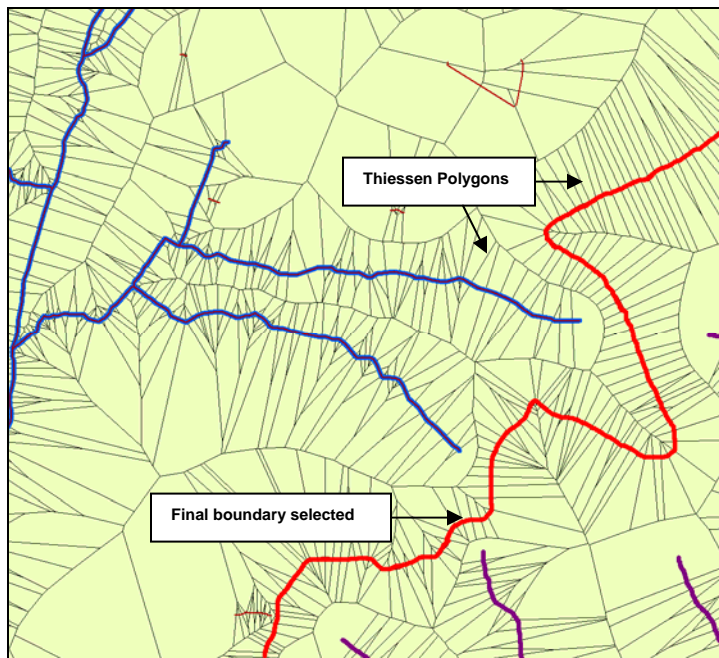


Figure 1 – The selection of an upstream network and its adjacent network

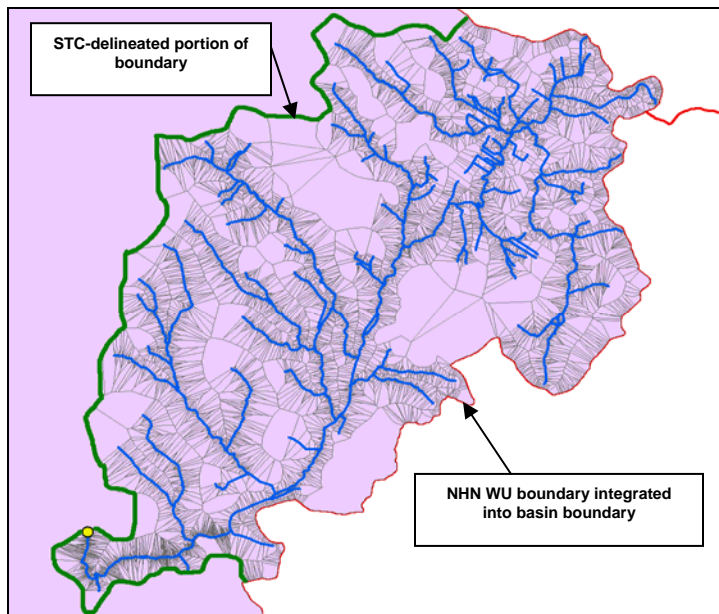
### 3.3 Creating Thiessen polygons and boundaries

Following step 3.2, points were created from the vertices of the upstream network and the adjacent network vectors. These points were then used as inputs into the ArcMap Thiessen Polygon function, as a batch process, to create Thiessen polygons between the upstream network and its adjacent networks. See Figure 2.



**Figure 2** – Creation of Thiessen polygons between networks

Once created, a path was selected between the networks by choosing all Thiessen polygons intersecting with the upstream network and then dissolving them to leave an outer boundary as the basin delineation. When the boundary intersected the NHN Work Unit (WU) boundary, the NHN WU boundary was selected and integrated as the basin boundary. See Figure 3.



**Figure 3** – Selection of Thiessen polygons to define basin boundary and integration of NHN WU boundary

### 3.4 Extending boundaries beyond single NHN Work Unit

Once the boundary was delineated from the site arc to the NHN boundary, it was often necessary to select adjacent NHN WU's, or sub-sub-drainage areas (SSDA), to build upstream basins that fully captured the upstream network and basin for that site. This was accomplished by locating the place where the water

feature crossed the boundary of a NHN WU and manually adding the upstream WU or SSDA. If time and resources permitted, one could have investigated the automation of this part of the process.

This latter step ensured that the upstream basin for each site was fully delineated, at least in so far as the scale and features of the NHN flow layer permitted.

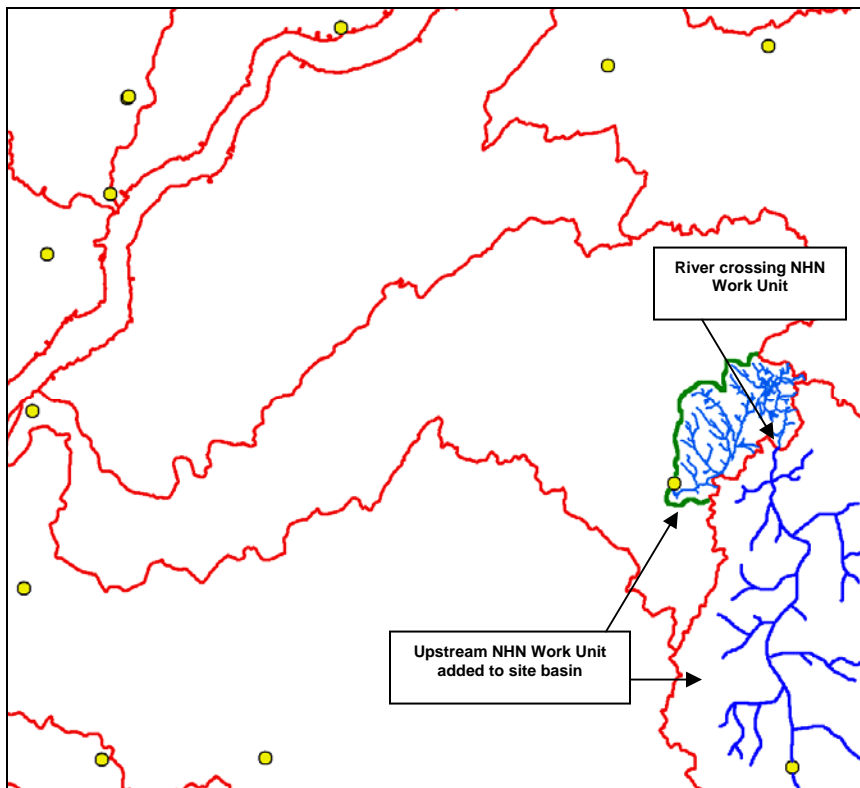


Figure 4 – Addition of NHN WU to complete basin

## 4 Compiling the statistics

The second phase of the project involved characterizing these upstream basins by various socio-economic and bio-physical variables, allowing for the categorization and comparison of sites within the network. In determining which variables to use for evaluating representivity, a list of desired characteristics was compiled in conjunction with EC. This list prioritized possibilities based on interest and level of difficulty in compilation. Appendix 'A' summarizes this selection process in chart form.

These statistics are intended to act as a general guide for characterizing these sites for the purposes of evaluating representivity. It was anticipated that some level of additional local knowledge could or would be added to enhance the verification and confirmation of the data compiled for each basin. Because of project constraints, variables that were chosen had to be readily accessible and ones that did not require a complex methodology to compile.

The following sections give some general information on each of the categories and their representative variables.

### 4.1 Drainage basin hydrological statistics

These statistics include information about the physical and hydrological characteristics of each upstream basin. They include basin area, river network length, network density and whether the site is northern or not, based on STC's north line delineation.

Also provided, was the long-term annual average runoff and mean annual flow, as estimated by Statcan's Water Yield Model. This data is being provided for the categorization and evaluation of these sites within the context of this project and are not intended for analysis beyond its framework.

## 4.2 Human activity - settlement

Human activity for the purposes of this project focused on population counts and population densities, with some additional information that could be distilled from the land cover summations.

## 4.3 Human activity - industry

Industrial information for this category was compiled from EC's National Pollutant Release Inventory and mining information from Natural Resource Canada's 2006 Census of Mines. The Business Registry of Statcan was not used at this time because of difficulties in geocoding individual businesses to these newly-delineated drainage areas. Statcan will be pursuing this work as a case study.

## 4.4 Human activity - agriculture

The variables used for this section were compiled using information generated from a joint project of Statcan and Agriculture and Agri-Food Canada, **Census of Agriculture: SLC V.3 and WSCSSDA V.5 Re-Allocation** which compiled agricultural information by sub-drainage area. For more information on this project go to:

[www.Statcan.gc.ca/cgi-bin/imdb/p2SV.pl?Function=getSurvey&SDDS=8012&lang=en&db=imdb&adm=8&dis=2](http://www.Statcan.gc.ca/cgi-bin/imdb/p2SV.pl?Function=getSurvey&SDDS=8012&lang=en&db=imdb&adm=8&dis=2)

## 4.5 Natural coverage

To represent land cover, both Natural Resource Canada's **250m Land Cover Map of Canada 2005 at 250m** and their 1km layer **Multi-Temporal Land Cover Maps of Canada using NOAA AVHRR 1-km data** were used in an effort to provide the broadest set of land cover possibilities for examining each basin. For more information on these datasets please go to [www.ccrs.nrcan.gc.ca/AD\EMS\Landcover2005](http://www.ccrs.nrcan.gc.ca/AD\EMS\Landcover2005) and [www.ccrs.nrcan.gc.ca/AD\EMS\Multi\\_temporal\\_Landcover1985\\_2000](http://www.ccrs.nrcan.gc.ca/AD\EMS\Multi_temporal_Landcover1985_2000).

## 5 Quality Control

For this project, issues of quality should be considered in relation to the size of the upstream basin, whether or not the basin is within a single NHN WU and how the basin size compares with the scale and units of the geospatial layers used to compile the various statistics. Generally speaking, if the basin is very small and particularly, if it is less than one NHN WU, the more uncertainty in some of the compiled statistics. It is advisable that metadata of layers used to compile statistics be consulted for additional information on quality.

For the purposes of this project, quality control (QC) is described in two ways; **process quality control** which was implemented during each phase of the project, and **quality control indicators** which were compiled and added to tables in order to assist users in conducting their analysis.

### 5.1 Process quality control

The unique nature of each basin delineation and the projects' scope dictated that manual observation would be the most efficient way for determining errors introduced during the delineation process. There were too many issues to consider automating QC of the 447 sites that were delineated. These issues included such things as complex directionality, zones with many orphan (unattached) vectors and unique and complex drainage patterns.

Process quality control was implemented by overlaying a skeletal network of rivers and a polygon layer containing bodies of water and main channels of larger rivers onto the NHN flow layer. This was done to provide additional information on the nature of flow patterns where questions arose. If judgements remained difficult after the overlay, other sources were consulted and occasionally screenshots passed on to EC for review.

The limitations of the QC were a result of time and the nature of the input layers. The delineation of boundaries for 447 sites took up the bulk of time and budget for this project. A larger investigation of the relationships (mostly topological) between the different layers of the NHN may have allowed for a robust semi-automated QC system to be implemented. However, the level of QC is considered to be more than reasonable for the objectives of this project.

Lastly, it is anticipated that these basin characterizations can leverage the expert knowledge of EC for final verification, especially where local knowledge is required (e.g. unique drainage patterns on the prairies).

## 5.2 Quality control indicators

The goal of these indicators was to give the user knowledge of the relationship between the basin size and the scale of the units from the various input layers used to characterize it. For example, when looking at block<sup>1</sup> aggregations used to create population counts and population density statistics, one must relate the size and number of blocks used to estimate these variables for each basin. A basin that has a population count and density that has been calculated by the overlay of parts of only one or a few blocks, contains more uncertainty than a basin that has a population based on a substantial number of blocks.

We have provided these relationships when we could efficiently do so and where it could not easily be derived from the base layer (i.e. the 1km landcover).

It is also important to restate that additional quality information can be extracted by referring to the metadata of datasets used in this study, particularly the dataset used to compile the agricultural variables.

## 6 Final delivery

The following are the contents of the package supplied to EC on DVD:

- Shapefile of 447 delineated basins
- 8 MSAccess databases containing subject matter detailed in section 4
- this report
- basin read me file

## 7 Future Research, Directions and Recommendations

The following findings are recommended for future improvements:

- orphan vectors, those not linked to the contiguous network, provided a challenge for delineating and understanding the hydrology of local networks. With more time and research, it may have been possible to better incorporate these into the delineation process;
- it would be beneficial to examine the relationship between the vector layers containing directionality and polygon layers representing water bodies and wider parts of rivers with the idea of creating upstream basins more efficiently;
- more time may have allowed for the construction of a more robust, detailed and semi-automated quality control process and/or more quality control indicators; and
- more diverse and specific industrial activity should be targeted for a more complete picture of this category.

## 8 Acknowledgment

Statcan would like to thank the NHN for its support.

<sup>1</sup> The block is the smallest geographic area for which population and dwelling counts are disseminated. It represents an area bounded on all sides by roads and/or boundaries of standard geographic areas. Blocks cover all the territory of Canada.

## Appendix 'A'

### Determining Representivity

Product Category	Product/Variable	Layer	Research Req'd	Difficulty (1 easy - 3 hard)	Completed as of report	EC Priority
<b>1) General statistics on each drainage basin</b>	Network length for each site	STC-D/NHN	n	1	y	2
	Watershed delineations	STC-D	n	3	y	3
	Drainage Area	STC-D	n	1	y	2
	Stream order	STC/NHN	y	3		1
	Network density	STC-D	n	2	y	1
	Northern	STC-GEO	n	1	y	2
	Long-term annual runoff	STC-EASD	n	1	y	1
	Mean annual water yield	STC-EASD	n	1	y	1
<b>2) WQI network related statistics</b>	Sites by Sub drainage area	STC-D/WQI	n	1	y	1
	Overlapping networks	STC-D/WQI	n	1	y	1
	Totals - area of each Pearse and Canada covered	STC-D/STC-Pearse	n	1		3
	Number of sites within distance (150km ?)	WQI	n	1		3
	Upstream distance between sites	STC-D	y	2		2
<b>3) Human Activity - Settlement</b>	Population	STC-D/STC-GEO	n	1	y	3
	Population density	STC-D/STC-GEO	n	1	y	1
	Population ecumene coverage	STC-D/STC-GEO	n	1		3
	Land cover	Land cover	n	2	y	1
	Settled area	STC-D/STC-EASD	y	2		3
	Road network length	STC-D/STC-GEO	n	2		3
	Road density	STC-D/STC-GEO	n	1		3
<b>4) Human Activity - Industry</b>	Mines and mills	NRCan	y	1	y	1
	Business register - business pattern	STC - BR	y	2		3
	NPRI	EC	y	2	y	1
<b>5) Human Activity - Agriculture</b>	Agricultural ecumene	STC-AG	n	1		3
	Agricultural variables (manure, crops etc..)	STC-AG	y	2	y	1
<b>6) Human Activity - Other</b>	Dams	NRCan	y	2		3
	Barriers from NHN	STC-D/NHN	y	2		3
	Hydro electricity generators - dams	STC	y	2		3
<b>7) Natural coverage</b>	Land cover	land cover	n	2	y	1
	Geology	NRCan	y	1		3
	Water yield	STC-EASD	y	3		3
	Slope	DEMs	y	3		3
	Precipitation	STC-EASD	y	2		3