

## Technical paper

Environment Accounts and Statistics Analytical and  
Technical Paper Series

# Introducing a New Concept and Methodology for Delineating Settlement Boundaries: A Research Project on Canadian Settlements



Environment Accounts and Statistics Division



Statistics  
Canada

Statistique  
Canada

Canada

## How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website at [www.statcan.gc.ca](http://www.statcan.gc.ca), e-mail us at [infostats@statcan.gc.ca](mailto:infostats@statcan.gc.ca), or telephone us, Monday to Friday from 8:30 a.m. to 4:30 p.m., at the following numbers:

### Statistics Canada's National Contact Centre

Toll-free telephone (Canada and the United States):

Inquiries line	1-800-263-1136
National telecommunications device for the hearing impaired	1-800-363-7629
Fax line	1-877-287-4369

Local or international calls:

Inquiries line	1-613-951-8116
Fax line	1-613-951-0581

### Depository Services Program

Inquiries line	1-800-635-7943
Fax line	1-800-565-7757

## To access this product

This product, Catalogue no. 16-001-M, is available free in electronic format. To obtain a single issue, visit our website at [www.statcan.gc.ca](http://www.statcan.gc.ca) and select "Publications."

## Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed *standards of service* that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on [www.statcan.gc.ca](http://www.statcan.gc.ca) under "About us" > "Providing services to Canadians."

# Introducing a New Concept and Methodology for Delineating Settlement Boundaries: A Research Project on Canadian Settlements

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2010

All rights reserved. The content of this electronic publication may be reproduced, in whole or in part, and by any means, without further permission from Statistics Canada, subject to the following conditions: that it be done solely for the purposes of private study, research, criticism, review or newspaper summary, and/or for non-commercial purposes; and that Statistics Canada be fully acknowledged as follows: Source (or "Adapted from", if appropriate): Statistics Canada, year of publication, name of product, catalogue number, volume and issue numbers, reference period and page(s). Otherwise, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form, by any means—electronic, mechanical or photocopy—or for any purposes without prior written permission of Licensing Services, Client Services Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

February 2010

Catalogue no. 16-001-M, no. 11

ISSN 1917-9693

ISBN 978-1-100-14126-8

Frequency: Occasional

Ottawa

Cette publication est également disponible en français.

---

#### **Note of appreciation**

*Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.*

---

## ***User information***

### **Symbols**

The following standard symbols are used in Statistics Canada publications:

- . not available for any reference period
- .. not available for a specific reference period
- ... not applicable
- 0 true zero or a value rounded to zero
- 0<sup>s</sup> value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
- p preliminary
- r revised
- x suppressed to meet the confidentiality requirements of the *Statistics Act*
- E use with caution
- F too unreliable to be published

---

## **Acknowledgements**

This report was prepared by the Environment Accounts and Statistics Division under the direction of Rowena Orok, Acting Director and Pat Adams, Acting Chief, Natural Resource Stock Accounts Section.

The Settlements Project team consisted of the following:

Nancy Hofmann, project manager  
Akmal Elgarawany, technical leader (Research and Methodology)  
Hugo Larocque  
Giuseppe Filoso  
Tim Dennis

Thank you to the following people for the contributions in the areas of editing, reviewing, proofreading, dissemination and technical development:

Jennie Wang  
Peter Murphy  
Michelle Tait  
Monique Deschambault  
Laurie Jong  
Doug Trant  
Smart Publishing Unit team

The project team gratefully acknowledges the funding provided for this project by the Policy Research Data Group (PRDG),<sup>1</sup> which was only possible with the support of Infrastructure Canada. We acknowledge the support of our project-specific PRDG steering committee consisting of Michel Ouellet (committee chairperson), Mary Herbert-Copley (Infrastructure Canada), Judy Watling (Policy Research Initiative) and Rowena Orok (Statistics Canada).

We also appreciate the funding previously received by the Canadian Space Agency's *Government Related Initiatives Program* (GRIP),<sup>2</sup> which funded research that assisted in the development of the settlements methodology. We especially thank our GRIP partners for all their contributions including data sharing and expert advice. These individuals include Mark McGovern, Environment Canada; Ted Huffmann, Agriculture and Agri-food Canada, and Don Leckie, Natural Resources Canada.

---

1. For more information please visit [http://www.policyresearch.gc.ca/page.asp?pagenm=inf\\_prdg\\_index](http://www.policyresearch.gc.ca/page.asp?pagenm=inf_prdg_index).

2. For more information please visit <http://www.asc-csa.gc.ca/eng/programs/grip/about.asp>.

---

---

## ***Table of contents***

<b>Abstract</b>	<b>5</b>
<b>Introducing a New Concept and Methodology for Delineating Settlement Boundaries: A Research Project on Canadian Settlements</b>	
1 Introduction	6
2 Overview: The settlements project	6
3 The methodology: research component	9
4 The methodology: processing the data	16
5 Initial research results and future directions	22

---

---

## **Abstract**

The purpose of this paper is to introduce a new concept and methodology developed by Statistics Canada to delineate or map boundaries for Canada's settlements. This research project was funded by the Policy Research Data Group (PRDG) to meet the needs of Infrastructure Canada and other clients.

Settlements, for the purposes of this research, are defined as tracts or blocks of land where humans have altered the physical environment by constructing residential, commercial, industrial, institutional and other installations/buildings. Settlements include cities, towns, villages and other concentrations of human populations that inhabit a given area of the environment.

The purpose of the settlements research project is to provide detailed, harmonized and comparable data sets that will allow for a more complete national analysis of settlements including their physical form and growth patterns. The methodology is applied to Census of Population data sets for 2001 and 2006.

The methodology is based primarily on data available by a statistical unit called a 'dissemination block' and an associated series of standardized rules, created by rigorous statistical and spatial analysis. A database referred to as the 'Settlements Earth Observation Inventory' (SEOI) was created to provide supplementary information for each block, to assist in the development of thresholds and subsequent rules, to conduct data quality and accuracy activities and to improve estimates.

Initial results for 2006 suggested that approximately 20,000 square kilometres of Canada's land area was occupied by settlements.

---

---

## ***Introducing a New Concept and Methodology for Delineating Settlement Boundaries: A Research Project on Canadian Settlements***

### **1 Introduction**

The size, structure and form of settlements over time can have a variety of socio-economic and environmental implications. The expansion of cities, for instance, requires infrastructure spending for new roads, sewers and water supply lines. Potential environmental impacts include loss of wildlife habitat and/or high quality agricultural land, increased air pollution and greenhouse gas emissions, and the contamination of rivers, lakes and aquifers. The type or form of expansion is also significant. Low density expansion consumes more agricultural and forested land and results in more infrastructure costs per capita than high density expansion.

The nature and growth of settlements are of particular importance for policy research related to cities, particularly metropolitan areas. The new methodology introduced in this paper, addresses data gaps and limitations with respect to current measures of Canada's built-up or settled areas. This methodology also aims to produce spatial data sets that are coherent through time and space.

### **2 Overview: The settlements project**

Coherent national analyses of the physical form and growth patterns of settlements and the resulting socio-economic and environmental impacts requires data sets that are comparable over time and space. At the same time, these data sets must track changes in the physical form of settlements as precisely as possible.

The research phase of this project included an in depth analysis of existing settlements-related data sets and a review of the relevant literature. While many useful national data sets on settlements were found, not all were produced at regular intervals and most, owing to their underlying methodology, yielded settlement boundary results that were not comparable over time and space. For example, in the past some boundaries were created using coarse data that did not necessarily reflect the true physical form of settlements; oftentimes this resulted in the inclusion of other land uses, such as agriculture, within the boundaries of settlements.

Datasets are produced for different reasons. For instance, census geographic boundaries at Statistics Canada were not developed to provide precise measures of settled areas or to conduct detailed time series analyses on the physical form of Canada's cities, towns and villages. These 'census geographies' are often designed according to administrative information. As a result, they tend to have limited utility when examining issues such as land use change because the boundaries do not always reflect areas that have been physically altered for settlement purposes.

The new approach presented in this research paper addresses current data gaps and allows the coherent mapping of settlements across the country.<sup>1</sup> The new methodology takes advantage of more refined census boundaries and satellite-based data sets to help address the challenges noted above.

There is strong policy and research interest for this research project. For example, Infrastructure Canada will use the results for research and analysis activities related to of Canada's infrastructure needs and programs. The

---

1. For more detailed information about this methodology, contact the Information Officer (613-951-0297; [environ@statcan.gc.ca](mailto:environ@statcan.gc.ca)), Environment Accounts and Statistics Division.

methodology also fills an important data gap in Statistics Canada's Land Accounts. It allows the creation of a new land use/land cover change matrix, which is used to track changes in land use and land cover over time.

### **The Land Accounts**

The Land Accounts are a major component of the Canadian System of Environmental and Resource Accounts (CSERA). Based on the compilation of various environmental data sets, the Land Accounts offer an integrated view of land and its use by Canadians. These accounts have a national coverage and are compatible with the economic accounts, supporting the coherent analysis of land issues in Canada.

For more information:

Statistics Canada, 1997, Concepts, Sources and Methods of the Canadian System of Environmental and Resource Accounts, Catalogue no. 16-505-G, Ottawa.

## **2.1 Key project goals**

The primary goal is to develop a new methodology and produce corresponding data sets that more specifically delineate or map the physical form of settlements across the country and through time. The resulting spatial data sets will reflect more accurately where people live and work. They will also ensure that the amount of unsettled land<sup>2</sup> within the new settlements boundaries kept to minimum and is easily measurable; the existence of these areas can impact calculations related to settlement growth and form.

The methodology described in this study is based on Geographic Information System (GIS) technology.<sup>3</sup> The application of the methodology results in a data set that contains digital boundaries depicting the limits of settlements.

## **2.2 What are settlements?**

Settlements reflect areas where humans live and work. They are tracts or blocks<sup>4</sup> of land where humans have altered the physical environment by constructing residential, commercial, industrial, institutional and other installations/buildings. They include cities, towns, villages and other concentrations of human populations that inhabit a given area of the environment.<sup>5</sup>

### **Challenges in the creation of a national dataset**

There are several challenges inherent in the creation of a national data set on settlements including the diverse nature of settlements in terms of size, geography, land use and form.

Some settlements consist of a small collection of houses, whereas others are very large cities with millions of inhabitants. The intensity of land use often increases with the size of settlements. For instance, large skyscraper buildings are typically found in the largest cities and not in small cities. The largest cities also tend to have denser road networks and higher population densities than remote or isolated towns. Even within a very large city, there are areas of high intensity, with skyscraper office buildings and apartment complexes, and areas of lower intensity such as strip malls or houses on large lots. There is variability in all settlements, but this variability typically increases with the size of settlement.

2. For the purposes of this research, 'unsettled' means that the area or blocks are not considered to be settlements as defined in Section 2.2 and in the methodology found in Section 4 "Methodology: processing the data."

3. Geographic Information Systems are computer systems that are capable of storing, analyzing and displaying geographically referenced data—data with a location identifier.

4. As shown in 3.2, settlements are based on a geo-statistical unit called a 'dissemination block'.

5. Adapted from [www.eionet.europa.eu/gemet/concept?cp=4062](http://www.eionet.europa.eu/gemet/concept?cp=4062).

### 2.3 The need for a new data set

While some detailed data sets on settlements exist for Canada's largest cities, they are not necessarily comparable with one another and do not provide for a complete national picture of settlements. Extensive research and analysis found that national data sets do exist but have certain limitations.

For instance, the National Topographic Database (NTDB) includes a built-up layer, but it has not been consistently or regularly updated. Using the NTDB to accurately track land use/cover change over time is therefore quite a challenge. As well, various types of satellite data are available, but the time, expertise and financial resources required to develop a national coverage based only on high-resolution data make such an approach unrealistic. Many of these national datasets do not include attribute or other supplementary data useful for research.

The most common data set of national data from which national and regional settlement-type analysis or indicators are derived is Statistics Canada's Census of Population. Although used for planning and other purposes, census geographies were not developed to provide precise measures of settled areas or to conduct detailed time series analyses on the physical form of Canada's cities, towns and villages. That is, these boundaries were often designed according to administrative information. As a result, these boundaries tend to have limited utility when examining issues such as land use change, because the boundaries do not always reflect areas that have been physically altered for settlement purposes.<sup>6</sup> For instance, these boundaries may include agriculture or forest land. As an illustration of this issue, consider the CMA and settlement boundaries (based on the methodology outlined in this report) of Regina in comparison to satellite imagery outlined in Figure 1.<sup>7</sup>

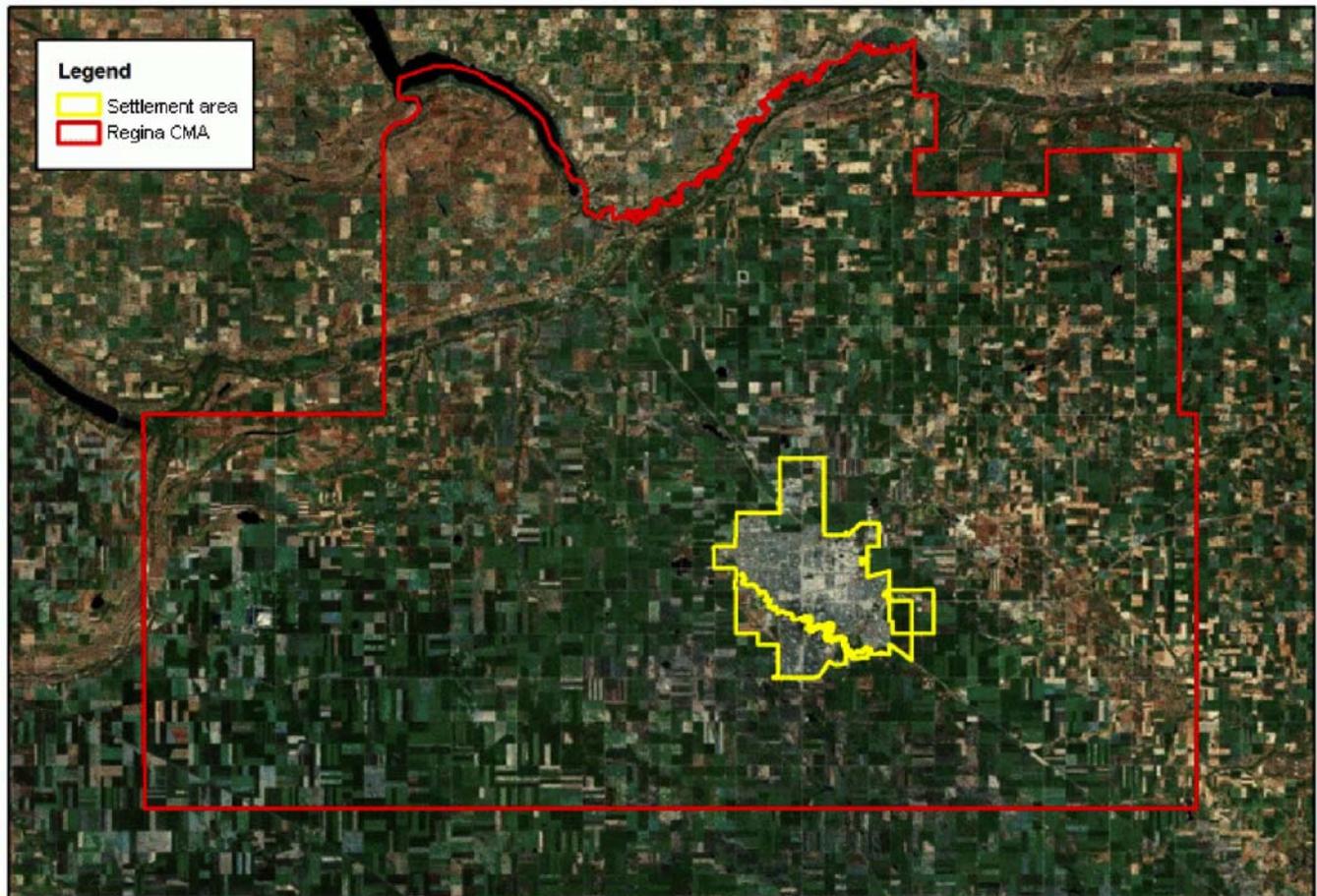
The development of a new data set is particularly timely given recent advances in Geographic Information Systems, satellite technology and the creation of the census geographical unit—the dissemination block.

---

6. All census geographies include some variation in land use. The block also includes some variation or mixed use of land; see Section 3.2 for more details.

7. Although commonly used to describe Canadian cities, census metropolitan area (CMA) boundaries were designed to reflect regions with a high degree of integration and were not designed to represent the nature, extent and change of the physical form of settlements in any detail. Many users will continue to use CMAs and other existing census geographies given their specific needs and the large volume of historic data available.

Figure 1

**Regina: census metropolitan area and settlement boundaries**

Source(s): Google Earth Pro ©2009.  
Statistics Canada, Environment Accounts and Statistics Division.

**3 The methodology: research component**

Methodology development for the project began with a preliminary research and analysis phase. This included the identification and examination of available data sets, followed by in-depth research into settlements accomplished through the creation of the Settlements Earth Observation Inventory (SEOI). The SEOI was instrumental in the following key research tasks: identifying settled features and their location (for example, residential, employment, and recreational areas, etc.); analyzing spatial structure and relationships; and establishing rules and thresholds based on spatial and statistical analysis.

### 3.1 Available data sets: identification and assessment

Several geographic data sets<sup>8</sup> were considered as potential foundations for the project, including:

- Statistics Canada geographic boundaries (dissemination blocks, block face, urban areas, designated places, census tracts, census subdivisions and localities);
- the National Topographic Database (NTDB);
- National Land and Water Information System (NLWIS) land cover (2000);
- Natural Resources Canada's National Road Network (NRN);
- Canada Centre for Remote Sensing MODIS (moderate resolution imaging spectrometer) product ('built-up area'); and
- various types of satellite data, for example, Landsat, Google Earth.

Each of these data sets had certain limitations for use in the settlements project. For instance, some of the data were produced only for a given year or did not follow a regular update cycle and therefore did not permit time series analysis. Other considerations included issues related to scale or resolution, cost of product, differing representation or definition of settlements, inclusion of unsettled land, lack of attribute data and others.

Upon consideration of these existing data sources, the dissemination block geography from the Census of Population was chosen as the mapping unit for the settlements project. This data set was chosen due to the small size of the geographic units<sup>9</sup> and the variety of attribute data available such as dwelling, population and employed labour force (ELF) counts (see Text box: **Key census variables and density measures for definitions**). The data are also regularly updated every five years; this supports the need for consistency through time and space.

### 3.2 The dissemination block and census variables

The mapping unit for this methodology is the 'block'. Known more formally as the 'dissemination block', the block is an area equivalent to a city block bounded by intersecting streets. More precisely, blocks may be bounded on all sides by roads and/or boundaries of standard geographic areas.<sup>10</sup> Blocks cover all of Canada. In less densely populated areas, blocks are typically larger than in cities and towns due to the sparse road network.

The block is the smallest geographic unit available for creating census geographies. As such, it is possible to aggregate dissemination block data to other standard census geographies. Since the block structure was created in 2001, the time series for this study is limited to 2001 and onwards.

While the block is a suitable mapping unit for this project, the following issues related to the unit and certain associated data are present:

- As the mapping unit of the settlements methodology, blocks are not split or otherwise changed. Since blocks can contain mixed uses, some blocks may be classified as "settled" due to high population density counts, but may nonetheless include portions of agriculture, forest, wetlands or other "unsettled" land.<sup>11</sup>

---

8. For information on these data sets, please visit the following internet sites: <http://www12.statcan.gc.ca/english/census06/reference/dictionary/geotoc.cfm>, <http://www.geogratis.cgdi.gc.ca/geogratis/en/product/search.do?id=8147>, <http://www4.agr.gc.ca/AAFC-AAC/display-afficher.do?id=1227635802316&lang=eng>, <http://www.geobase.ca/geobase/en/data/nrm/index.html>, <http://article.pubs.nrc-cnrc.gc.ca/RPAS/rpv?hm=Hlnit&journal=cjrs&volume=31&afpf=m05-019.pdf>, <http://landsat.gsfc.nasa.gov/>, and [http://www.google.com/enterprise/earthmaps/earth\\_pro.html](http://www.google.com/enterprise/earthmaps/earth_pro.html).

9. Blocks vary in size. In areas with concentrations of population, blocks are small in size. Almost three in every four blocks were smaller than 1 km<sup>2</sup> in 2001.

10. Please refer to <http://www12.statcan.gc.ca/english/census06/reference/dictionary/geo014.cfm> for examples of standard geographies and further details.

11. The Settlements Earth Observation Inventory (SEOI) codes (1, 2 and 3) could be used to improve this limitation. Please see section 3.3.

- Because dissemination blocks are a function of the road network, the number of blocks created is a function of the timeliness of updates of the road network files prior to a given census. Highway medians, ramp areas and other irregular shapes may also form blocks on their own due to road networks. There are also changes or updates in block boundaries between censuses that are not due to actual change, but rather, to improvements in the block geography. A synchronization process in the settlement methodology is used to reduce the impact of these changes to blocks. In the data preparation phase of this study (see Section 4.1), some of these issues were addressed.
- The Census of Population is distributed to all households in the country; information for all households is obtained regarding dwelling and population characteristics. However, more detailed questions (for example, the long form) were sent to 20% of Canadian households. These questions included information about one's place of work and employment. This sampled approach has implications on data quality.<sup>12</sup>
- Certain census information are geo-coded (that is, located physically on a map) by Statistics Canada. Depending on the census variable, geo-coding is not always conducted at the block-face<sup>13</sup> or block level. For example, the workplace location of people working outside CMAs and census agglomerations (CAs) are coded to census subdivisions (CSDs), and not the block.

### 3.3 The Settlements Earth Observation Inventory (SEOI)

The SEOI is an inventory of geographic information and codes based on an overlay and analysis of the blocks in comparison to satellite imagery. The comparison includes, for example, estimates of the proportion of land that is covered by settlement features in each individual block. To produce the SEOI, imagery for about 260,000 blocks were visually assessed and coded. As a result, in addition to data from the census for each individual block, extra information is now available for each individual block based on satellite imagery.

The creation of the SEOI by Statistics Canada fulfilled a number of important needs for this project, including the following:

#### 1. **Characteristics of settled blocks**

Information and improved understanding about the characteristics of settled blocks was needed in order to develop the new methodology. This included information about densities, physical form, land uses, etc. For example, it helped answer the question: at what population density is a block considered settled?

#### 2. **Analysis of spatial relationships leading to standardized rules for the methodology**

The SEOI provided data for analyzing and understanding the spatial structure, distribution and relationships of settled blocks. The analysis of the SEOI led to the concepts and threshold-setting for rules related to population, employed labour force (ELF) and dwelling densities, as well as buffers. A 'buffer' is a zone around a block or cluster of blocks; for example, a block cluster with a buffer of 1 km would involve a zone 1 km wide around the periphery of the block cluster boundary. (see Figure 5 for an illustration of the buffering exercise).

These concepts and rules were then used to establish an automated or rule-based process for delineating settlement boundaries across the country.

12. For more information on sampling and its impact on data quality, please visit the following location:  
<http://www12.statcan.gc.ca/english/census06/reference/dictionary/app002.cfm>.

13. A block-face is one side of a street between two consecutive intersections. For example, a block-face can be one side of a city block.

### 3. Data enhancement/improvement

(i) SEOI data were also used to improve overall estimates of total settlement area. Even with well-established rules, some blocks that should have been included were initially omitted, due to limitations of the ELF data, geocoding,<sup>14</sup> geographic variability or other reasons. The SEOI allowed settled blocks that would have otherwise been missed to be added. The reverse is also true: some unsettled blocks that were initially included were excluded based on SEOI analysis.

(ii) The supplementary data in the SEOI allows for improved calculations and flexibility. Limitations related to mixed-use settlement blocks, where settlements are not the dominant feature in a block, can be addressed through use of the SEOI. For example, estimates for settlement area can be adjusted to exclude blocks with limited settlement activity.

### 4. Data quality and accuracy assessment activities

The SEOI provided an avenue for conducting data accuracy assessments for the rule-based standardized process (see Section 4). This product can also be used to assess the data quality of other products.

#### 3.3.1 Creating the SEOI

Using satellite data, analysts verified whether or not a given block was settled and determined the proportion of the block that was settled. In doing so, the assessed blocks were given specific codes based on a visual assessment of the block data in comparison to the satellite imagery.

The satellite data consisted of either high-resolution satellite data (QuickBird) or medium-resolution data (Landsat), usually from Google Earth Pro.<sup>15</sup> Approximately 90% of the imagery used was high resolution, while the remaining 10% consisted of medium resolution data. All census subdivisions (CSDs) with a population of 1,000 or more were analysed, representing an overall total of about 260,000 blocks.

During the coding process, blocks were assigned the following codes:

- Code 1—over 50% settled<sup>16</sup>
- Code 2—between 25% and 50% settled
- Code 3—between 10% and 25% settled
- Code 4—under 10% settled
- Code 0—unsettled.

Other information was also tracked, including the year of the satellite imagery, the type of imagery (high or medium resolution), and general notes including observations on dominant land use.

Figure 2 outlines a sample block of the SEOI in Google Earth Pro as well as the accompanying data for a single block.

Since the SEOI contains satellite imagery from about 2005 to 2008, it is partnered with data for the 2006 Census. To benefit further from the work of the SEOI, rules were created to include information from the 2006 SEOI in the 2001 settlements data set. The rules were primarily based on census and land cover data.

---

14. Geocoding is the process of converting street addresses into spatial data that can be displayed on a map.

15. For more information, please go to [http://www.google.com/enterprise/earthmaps/earth\\_pro.html](http://www.google.com/enterprise/earthmaps/earth_pro.html).

16. "Settled" refers to areas where humans live and work. The imagery would depict residential, commercial, industrial, and institutional areas. In the high-resolution data, for instance, houses are clearly visible in residential areas.

Figure 2

## Illustration of Settlements Earth Observation Inventory (SEOI)



**Source(s):** Google Earth Pro ©2009.  
 Statistics Canada, Environment Accounts and Statistics Division.

**Figure 2: Additional details**

The SEOI provides supplementary data, which can be added to census attribute data. That is, in addition to census data, information based on the satellite imagery is added to the data set; collectively referred to as a 'geodatabase'.

The data set includes census related information by block (for example, shape length, ID #, population, employment, dwelling, etc.). It also includes census subdivision information (for example, ID, CSD type, population, dwelling, etc.). SEOI information is also available including codes (that is, 1, 2, 3, 4, 0), information on the rules applied, and land use type.

**3.4 Results of SEOI analysis: spatial structure and relationships**

Analysis of the SEOI shows that there is great variety in settlement patterns across Canada. While some settlements are isolated and remote, others are located in regions with a higher concentration of settlements, such as in southern

Ontario. Settlements themselves vary in size and scale, from small groupings of houses to large metropolitan areas with millions of inhabitants.

There is also much variation within settlements, with some areas densely populated and other areas more sparsely populated. Within a given settlement, there may be areas with high employment counts coexisting with areas of no employment activity whatsoever.

The use of the SEOI and GIS technology permitted the spatial analysis of these complex patterns and relationships. Given this complexity, it was clear that one simple rule or variable was insufficient to classify settlement area with a high degree of certainty.

The spatial analysis was conducted using CSDs as a frame. The CSDs were divided into seven categories based on their population.<sup>17</sup> This analysis of the spatial nature of settlements formed the basis of the thresholds and resulting rules established for the methodology of this project. Rules were initially developed and tested; testing frequently required the modification of rules or the addition of new rules. The final set of rules is presented in Section 4 "The methodology: processing the data."

### 3.4.1 Core areas and neighbours

Overall, settlements vary considerably in their form and size as a result of a number of factors including geographic location, historical evolution, socio-economic bases, etc. However, there are similar patterns with regards to their use of space.

Settlements contain different, distinct areas or zones. Relationships exist *between* these zones. Most settlements have concentrated areas or 'core areas'<sup>18</sup> where the use of land appears to be more intense.<sup>19</sup> Core areas can be located throughout a given settlement, not just in areas that would casually be referred to as "downtown". These core areas are usually surrounded by or neighbored by areas where land use is less intense, but that are still settled or partially settled. Typically, there is a dependent relationship between the more intensely used core zones and their less intensely used neighbours.

Analysis using the SEOI found that approximately 80% of blocks coded 1, 2 and 3 had a population and/or employment density of at least 400 people per square kilometre. Similarly it was discovered that over 85% of blocks with a population density of 400 people or employees per square kilometre were at least 50% settled (Code 1). Core areas also typically had a more concentrated road network due to the high degree of settlement activity. As a result, core area blocks were typically smaller in size compared to less intensely used zones. These statistics helped formulate the rules for core areas based on density and block size as outlined in Section 4 "The methodology: processing the data."

Less intensely settled blocks were located adjacent to core blocks. Intense settlement activity appeared to encourage additional settlement activity around its borders, but at less intense levels. For instance, over 80% of the less intensely settled blocks were located completely within 500 m of core areas. These blocks were considered 'first order neighbours' (Figure 3). A key attribute of these first order neighbour blocks is their relatively small size and the high frequency of certain specific land uses such as parks, golf courses, parking and areas of growth. This analysis helped formulate rules related to buffers and block size for first order neighbours (see Section 4 "The methodology: processing the data").

---

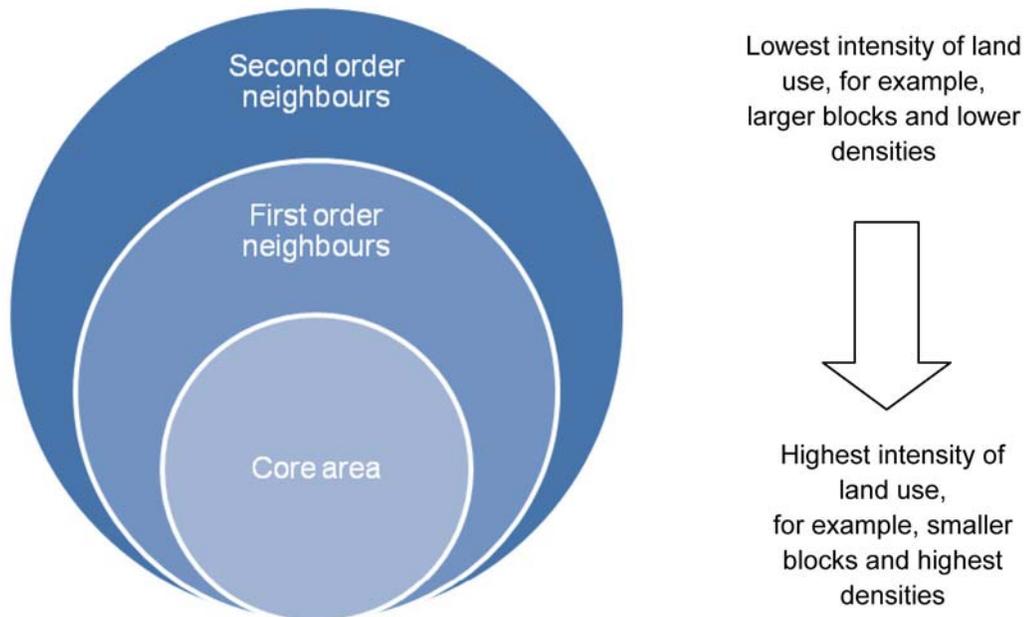
17. The seven categories consisted of the following: (i) 1,000 to 4,999, (ii) 5,000 to 9,999, (iii) 10,000 to 19,999, (iv) 20,000 to 49,999, (v) 50,000 to 99,999, (vi) 100,000 to 199,999, (vii) greater than 200,000.

18. The use of term 'core areas' in this research project should not be confused with the term "core" as it may be used elsewhere in Statistics Canada; for example, the term 'Urban Core', defined as a large urban area around which a CMA or a CA is delineated, may be used in context of Census of Population statistics.

19. 'Intensity' refers to the degree of settlement activity taking place. Typically intense areas include more buildings and higher population, employment and dwelling densities. They could include areas with apartment buildings and office towers, but can also include areas with single detached dwellings where lots are small. Less intense areas include areas that are covered by fewer structures; examples include areas with single detached dwellings on larger lots or partially settled blocks.

Figure 3

Distinct zones within settlements



Source(s): Environment Accounts and Statistics Division, Statistics Canada.

Another distinct zone was discovered beyond the first order neighbours. These 'second order neighbours' (Figure 3), typically consisting of blocks larger than the first order neighbours, were often characterized by less intense settlement activity. About 94%<sup>20</sup> of these lower density settled blocks had their centre of gravity or centroid<sup>21</sup> within 1,000 m of a core block. This statistic, and the testing activities noted above, led to the development of rules related to buffers, as well as densities and block size for second order neighbours (see Section 4 "The methodology: processing the data").

20. This statistic is derived from an analysis of CSDs having a population greater than 5,000. This percentage decreases slightly for CSDs having less than 5,000 inhabitants.

21. The block centroid or centre of gravity is represented by a point and illustrates the centre of the block.

## Thresholds and rules

### ***Thresholds and rules for settlements methodology***

Thresholds and rules were established to ensure that only those blocks that should be classified as settlement were actually classified as such. The thresholds for minimum of population, employment and dwelling densities and maximum block areas were based on median population, dwelling and employment densities and areas in the SEOI. The median densities and areas were calculated for all census subdivisions (CSDs) with a population of over 1,000. However, the thresholds for this methodology were based on data for all CSDs with a population over 5,000 in 2006.

To determine the thresholds for population and employment densities, one standard deviation (STD) was subtracted from the median. Two standard deviations were added to the median area to create the maximum block area limit. For the dwelling density thresholds, the minimum density of settled blocks was visually assessed with the SEOI. In some cases, such as the population density threshold of 150 people per km<sup>2</sup>, the thresholds were rounded.

### ***Thresholds and rules in other methodologies***

The methodology introduced in the paper uses a population density of 400 people per square kilometre, this density has been a long-standing threshold used by Statistics Canada's urban area program.<sup>22</sup> The United States Census Bureau has also used a similar threshold for defining core census block groups. Its threshold of 1,000 people per square mile is equivalent to 396 people per square kilometre.<sup>23</sup>

While this threshold has been used previously, a key difference between this and other projects is that the new settlements project methodology uses this threshold at the block level, not at the higher settlement level.

In finding second order neighbours, this methodology uses a population density threshold of at least 150 people per km<sup>2</sup>. The same threshold is used in Statistics Canada's designated places program, which has used this minimum density threshold for many decades.<sup>24</sup> Similarly, the Organisation for Economic Co-operation and Development (OECD) has created a definition for rural areas for member countries to follow. It defines areas with a population density under 150 people per km<sup>2</sup> as rural<sup>25</sup>— non-rural areas therefore have a population density of more than 150 people per km<sup>2</sup>.<sup>26</sup> Other research, including the NEWRUR (Urban Pressure on rural areas) project also used a threshold of 150 people per km<sup>2</sup>.

Other methodologies have recognized the spatial nature of settlements and the core and lower density neighbour relationship. For example, the United States Census Bureau also includes areas with lower densities within their Urban Area and Urban Clusters definition. Initially all blocks with a density of 1,000 people per square mile are chosen, followed by surrounding census blocks that have an overall density of at least 500 people per square mile (193 people per km<sup>2</sup>).<sup>27</sup> A buffer approach has also been used for the delineation of urban areas in England and Wales, where it was found to minimize the misclassification of rural areas.<sup>28</sup>

## 4 The methodology: processing the data

There are four key stages to the methodology (Figure 4):

- data preparation;
- the rule-based selection process, which identified core areas and first and second order neighbours;
- data quality and accuracy assessment activities; and
- data enhancement activities.

22. See the following webpage for more information <http://www12.statcan.gc.ca/english/census06/reference/dictionary/geo049a.cfm>.

23. See the following for more information [http://www.census.gov/geo/www/ua/ua\\_2k.html](http://www.census.gov/geo/www/ua/ua_2k.html).

24. See <http://www12.statcan.ca/english/census06/reference/dictionary/geo018a.cfm> for more information.

25. See [www.oecd.org/dataoecd/42/16/15181756.doc](http://www.oecd.org/dataoecd/42/16/15181756.doc).

26. The NEWRUR project was a collaborative effort between five European research centres, see for more information: [http://newrur.grenoble.cemagref.fr/project/wp\\_1\\_d12.htm](http://newrur.grenoble.cemagref.fr/project/wp_1_d12.htm).

27. See [http://www.census.gov/geo/www/ua/ua\\_2k.html](http://www.census.gov/geo/www/ua/ua_2k.html).

28. See [www.statistics.gov.uk/census2001/pdfs/urban\\_area\\_defn.pdf](http://www.statistics.gov.uk/census2001/pdfs/urban_area_defn.pdf).

Key concepts used in this methodology include the use of census-based densities measures (see Text box: **Key census variables and density measures**), buffers<sup>29</sup> (see Figure 5), and block size limitations (see Text box: **Thresholds and rules**).

### **Key census variables and density measures**

#### **Key census variables**

Three key census variables were used in the methodology to identify settlements: population, dwelling and employment counts. The research and analysis portion of this project concluded that all three variables were required, given the complex nature of settlements.

#### **Use of density measures**

Density measures, such as population counts divided by block area, were used as indicators of the concentration or intensity of human activity. By dividing by the total area of the block, the population, dwelling or employment data were normalized or standardized, making them more readily comparable.

A drawback of this approach is that the standardization process may obscure the results of blocks with mixed uses, such as blocks where one corner contains a small hamlet and the rest is occupied by forest. To address this drawback, the SEOI was used. That is, the coding provided by the inventory provides details on the dominant land use and the proportion of the block that was settled. In summary, if a settled block was not classified by the density parameters set out in the rule-based selection process, the SEOI could be used to include this area.

The variables and their density measures are outlined below:

#### **Population and population density**

Population data consists of the number of permanent inhabitants in a given area. Population density is the number of people per square kilometre (total population divided by land area). For identifying settlements, population density helps locate concentrations of people, such as residential areas.

#### **Dwelling density**

Dwellings refer to a set of living quarters in which a person or a group of people resides or could reside, such as single detached dwellings, duplexes, townhouses, apartments, mobile homes, etc. For this project, occupied and unoccupied private dwellings were included. Collective dwellings such as hospitals or hotels were excluded. Dwelling density is the number of dwellings per square kilometre (total dwellings divided by land area). These data were useful for locating residential areas.

#### **Employed labour force**

Based on **Place of Work** data, the employed labour force (ELF) was defined as the population 15 years of age and over, excluding institutional residents, who worked during the week prior to census day. ELF density was calculated as the number of employees per square kilometre. These data were used to locate areas of employment.

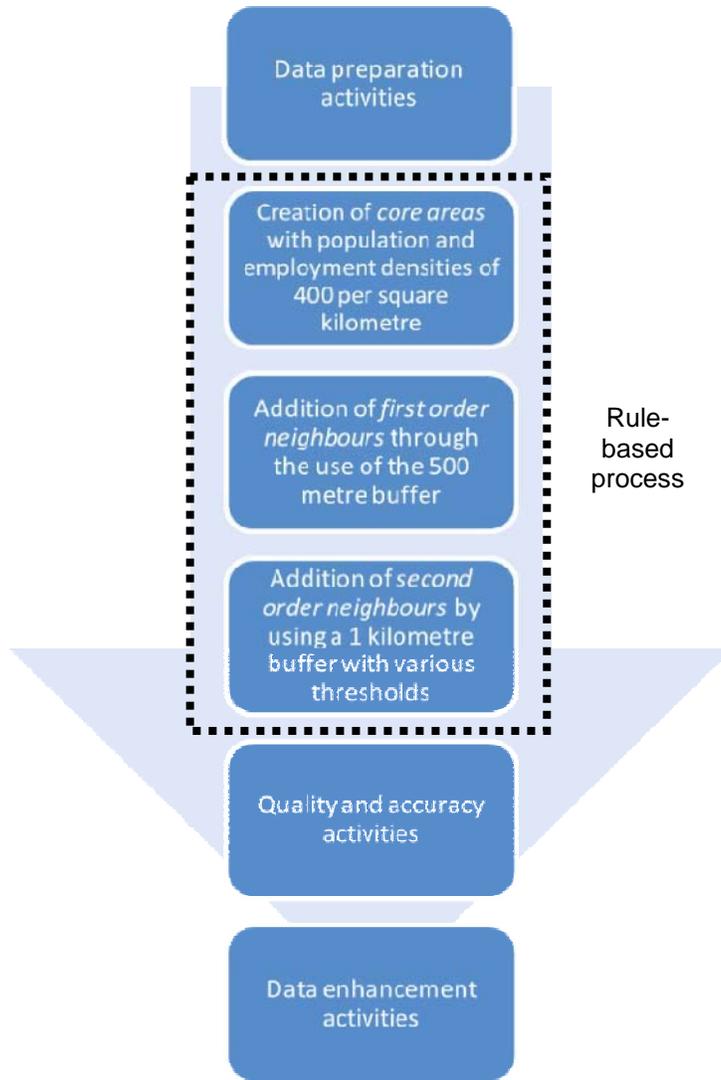
**Note(s):** This text box refers only to census variables. Other variables were used in this research project, please visit Section 3 "The methodology: research component" for more information on these non-census variables.

For more information about census variables please visit the Census of Population dictionary: <http://www12.statcan.gc.ca/english/census06/reference/dictionary/atoz.cfm>.

29. A "buffer" is a zone around a block or cluster of blocks; for example, a block with a buffer of 1 km would involve a zone of 1 km around the periphery of the block boundary.

Figure 4

Key steps of the methodology



**Note(s)** : The creation of core areas and first order neighbours involved rules regarding the size of blocks. That is, to be a core area, blocks had to have a maximum size of 0.3 km<sup>2</sup> for population blocks, and 0.7 km<sup>2</sup> for employment blocks. First order neighbours had a maximum size of 0.5 km<sup>2</sup>. Second order neighbours had various thresholds as described below.

**Source(s)**: Environment Accounts and Statistics Division, Statistics Canada

## 4.1 Data preparation

Data preparation involved a number of different activities. Geospatial data preparation activities included map projection changes,<sup>30</sup> the integration of a new hydrology layer,<sup>31</sup> the creation of a central geodatabase, and other activities.<sup>32</sup> If this research were to be updated, the SEOI would be updated and analyzed in the data preparation phase.

The geodatabase consists of many attribute tables as well as spatial features (for example, census geographies) and raster data (for example, thematic layers and mosaics derived from satellite data). The attribute tables include information such as block identification codes, census counts (population, dwelling and employment), area calculations, linear and surface densities of census variables, correspondence codes, data codes and information from the SEOI and geospatial statistics based on the analysis by block of various national land cover products. This geodatabase is the key connecting foundation of this project; it connects census data with other geospatial datasets.

Another component of the data preparation activity was a synchronization process. Because of some changes in blocks between 2001 and 2006, a synchronization process was needed to ensure data integrity as well as temporal and spatial consistency. Essentially, population, dwelling and employment data from 2006 blocks were transferred to 2001 blocks if certain criteria were met.<sup>33</sup>

## 4.2 The rule-based selection process

An automated and standardized rule-based selection process was responsible for classifying the greatest number of blocks. Overall, 98% of blocks that were classified as settled were classified during the rule-based process.

The process applied rules to all blocks in Canada in an automated fashion within ArcGIS. The key elements of the rule-based process included the selection of core areas, the creation of the distance buffers around core areas and the application of thresholds or rules related to population, dwelling and employment densities of blocks for each of the two buffers. These rules were a result of analysis of the SEOI. Please refer to Text Box: **Thresholds and rules**, for more details regarding thresholds and rules selection.

An automated process is desirable for two principle reasons. Firstly, it is more cost effective than a process that requires a great deal of manual intervention. Secondly, an automated process helps ensure consistency over time and across the country. This is important for ensuring comparability, particularly for analytical purposes.

### 4.2.1 Selection of core areas

The first step of the rule-based process was to select blocks that had a population or employment density of at least 400 people per square kilometre and that did not exceed a certain size limitation; that is, in terms of population-related blocks, only blocks with a maximum area of 0.3 km<sup>2</sup> were selected, while for employment-related blocks, the blocks had to have a maximum area of 0.7 km<sup>2</sup>. This first rule accounted for almost 80% of the total number of the blocks classified as settlements.

30. The map projection used was *Albers Equal Area*. This projection is well-known for ensuring the most accurate area estimates.

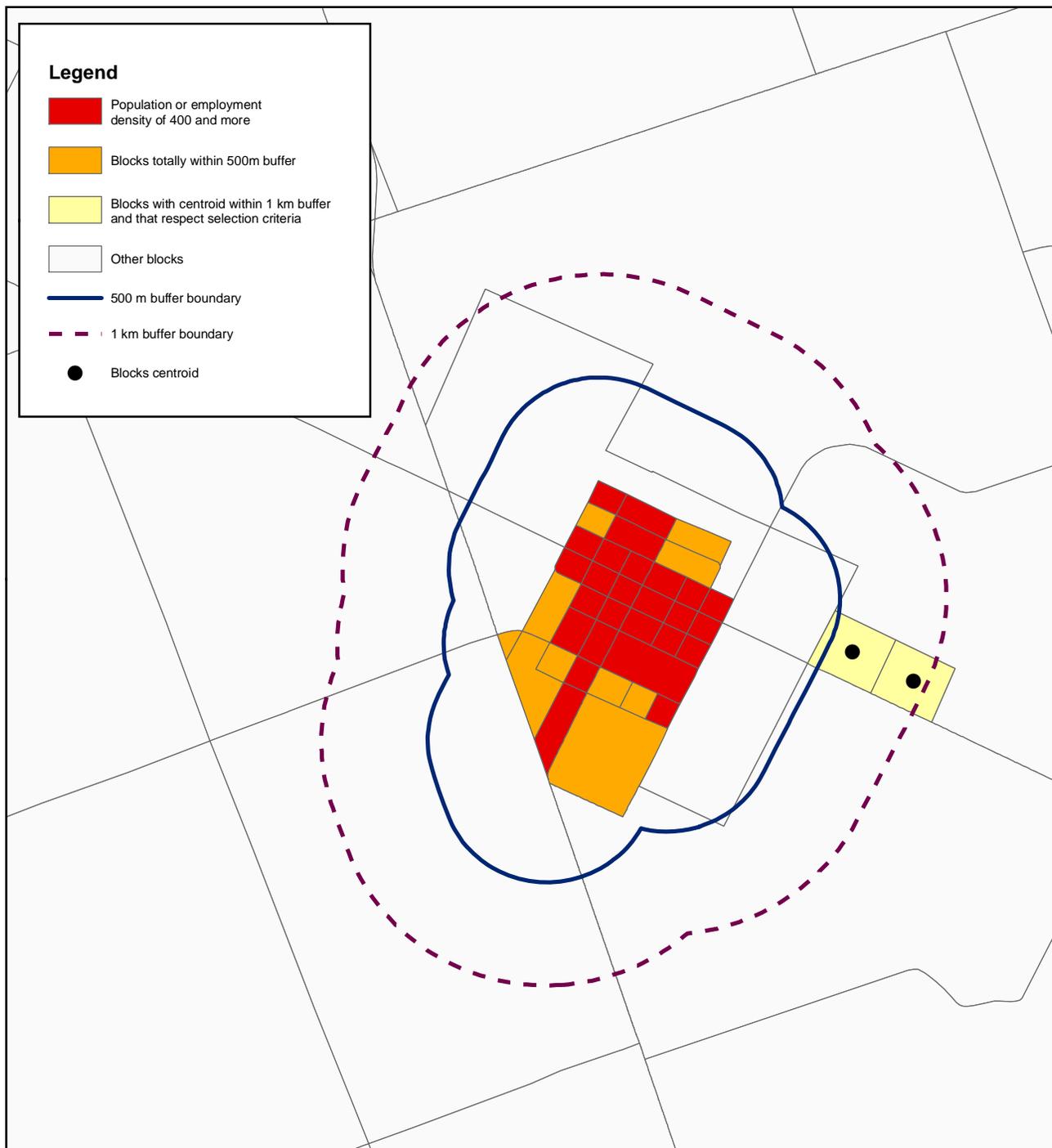
31. The hydrology incorporated was from the Statistics Canada National Geographic Database. For more information see: Statistics Canada, Catalogue no. 92-160-G.

32. In addition to ArcGIS, this work was accomplished through the use of Microsoft Access and Structure Query Language (SQL).

33. Changes occurred in the block geography between 2001 and 2006. In 2001, there were a total of 478,707 blocks, compared to 478,831 blocks in 2006. The vast majority of blocks, about 80%, remained geometrically unchanged between the two periods. The *Dissemination Block Correspondence File* was used to match blocks from the two censuses. This file provides information on which blocks remained the same and which blocks changed. For more information about this file, please see Statistics Canada, Catalogue no. 92-156-G.

Figure 5

Illustration of the buffering exercise



Note(s): The blocks with a population or employment density of 400 or more (with size limitations as noted) are considered core areas. Blocks totally within the 500 m buffer that are less than 0.5 km<sup>2</sup> are referred to as 'first order neighbours'. Lastly, blocks having a centroid within the 1 km buffer that also meet other selection criteria are known as 'second order neighbours'.

Source(s): Statistics Canada, Environment Accounts and Statistics Division.

#### 4.2.2 Selection of first order neighbours

Once the 'core blocks' (Figure 5) were selected, a 500 m buffer<sup>34</sup> was generated around these blocks (Figure 5). The use of buffers reflects the spatial relationship (proximity, adjacency, etc.) between the different spatial features of settled blocks. The effectiveness (that is, accuracy) of the buffer was evaluated and changes in buffer width were made in order to reduce the incidence of unsettled blocks being classified as settled.<sup>35</sup>

Only blocks less than 0.5 km<sup>2</sup> that fall completely within this buffer were also classified as settled (Figure 5). This method ensured that only small blocks were included (for more information, please see Text box: **Thresholds and rules**).

#### 4.2.3 Selection of second order neighbours

The selection of the second order neighbours was more complex and required a range of rules or thresholds based on settlement structure, distribution of densities and block size. These thresholds ensured that only the most appropriate blocks were classified as settled. For instance, the rules ensured that large blocks with limited settlement activity were not included.

A 1,000 metre buffer was generated around all core blocks (Figure 5). All blocks within the 1,000 metre buffer that met the following criteria were selected as settlement blocks:

- (a) The block centroid<sup>36</sup> must be found within the 1,000 m buffer (the entire block does not have to be within the buffer). If the centroid is outside the buffer, the block is not classified as settled at this time.
- (b) All blocks with a centroid falling within the 1,000 m buffer were examined for the following criteria; if the following criteria were met, they were classified as settled:
  - a minimum population density of 150 people/km<sup>2</sup> and minimum dwelling density of 50 dwellings/km<sup>2</sup> and a maximum area of 0.9 km<sup>2</sup> or
  - a minimum dwelling density of 55 dwellings/km<sup>2</sup> and a maximum area of 0.9 km<sup>2</sup>
  - a minimum ELF density of 100 employees/km<sup>2</sup> and a maximum area of 2.9 km<sup>2</sup>

### 4.3 Quality and accuracy activities

Various verification activities were undertaken to ensure quality and accuracy. These activities were limited primarily to thematic accuracy checks, for example, determining if the settlements are actually settled.

The settlement boundaries were converted into a format that could be viewed in Google Earth Pro. GIS specialists then compared all settlement boundaries with satellite imagery. All areas that appeared settled, but that were not included by the rule-based selection process were flagged and a follow-up was conducted. Similarly, areas that were identified as settled but that did not appear settled were also flagged for follow-up. Over 250 areas received follow-up examination and all were found to follow the rules created in this methodology and no changes were made.

A data accuracy assessment was conducted<sup>37</sup> by comparing the results of the rule-based selection to the SEOI for 2006. If the rules classified the block as settled and the SEOI results also showed that the block was settled it was deemed a success.

Three types of accuracy are considered:

- overall accuracy;
- producer accuracy (or omission error); and

34. The buffer was calculated outward from the boundary of the block, which serves as the mapping unit for this study. For CSDs with a population under 1,000, a buffer of 300 m was used.

35. Evaluations were conducted for all seven CSD categories discussed earlier. For example, a 500 m buffer had an accuracy or "fitness" of 95% in the CSDs with a population greater than 100,000, whereas a buffer of 500 m had an accuracy or "fitness" of 88% for the CSDs with a population less than 5,000 inhabitants.

36. The block centroid or centre of gravity is represented by a point and illustrates the centre of the block.

37. The assessment was limited to CSDs with a population greater than 1,000.

- user accuracy (or commission error).

In terms of overall accuracy, there was a success rate of 97%.<sup>38</sup> When the results of the rule-based selection process were combined with additions made from the SEOI, a success rate of around 99% was achieved.

Producer accuracy, or omission error, examines the proportion of settled blocks from the SEOI that should have been captured by the rules-based selection process, but were not. The number of blocks omitted was 7,443, for an omission error rate of 2.8%.<sup>39</sup>

User accuracy, also known as commission error, is the portion of unsettled blocks that were captured by the methodology but that, according to the SEOI, were not settled. The number of unsettled blocks captured by the rule-based was 2,141 blocks for a commission error of 0.8% or a user accuracy of 99.2%.<sup>40</sup>

## 4.4 Data enhancement activities

The settlements data were primarily improved by the inclusion of the SEOI data and by the cleaning and merging processes.

### 4.4.1 Inclusion of the Settlements Earth Observation Inventory (SEOI)

The flexibility of the SEOI allows users to include or exclude specific settlement blocks. In order to address various limitations and improve data quality, some SEOI data were included.

### 4.4.2 Cleaning and merging processes

The final stage of the methodology involved cleaning the settlements data selected by the rule-based process using the SEOI and merging settlement blocks into larger groupings.

The purpose of the cleaning process was to exclude areas that had been classified in error during the rule-based process, using the SEOI information. If a block was coded as forest or agricultural land in the SEOI, for example, but classified as settlement by the rules-based process, it was removed from the database.

The merging process involved dissolving the polygons; that is, merging individual settlement blocks together to create actual settlements.

## 5 Initial research results and future directions

### 5.1 Initial research results

Approximately 20,000 square kilometres of Canada's land was considered to be settled in 2006.<sup>41</sup> Map 1 illustrates the location of all settlements found through this project. Maps 2 to 4 provide a closer look at settlements for specific areas of the country. Concentrations of settlements occur in certain regions, particularly in central Canada.

---

38. The rule-based selection process captured 254,769 blocks and the number of settled blocks found in the SEOI was 262,212.

39. Reasons for this omission include: limitations of the employment data as a result of its sampled approach; the existence of blocks that do not follow the rules created (for example, blocks may be located just outside of a buffer); and omissions due to geocoding issues (for example, large airports that cover many blocks but whose employees are assigned to one block).

40. Reasons for this commission include: geocoding issues; irregular/atypical settlement patterns (small unsettled blocks surrounded by settled blocks); and the presence of blocks subdivided for future development.

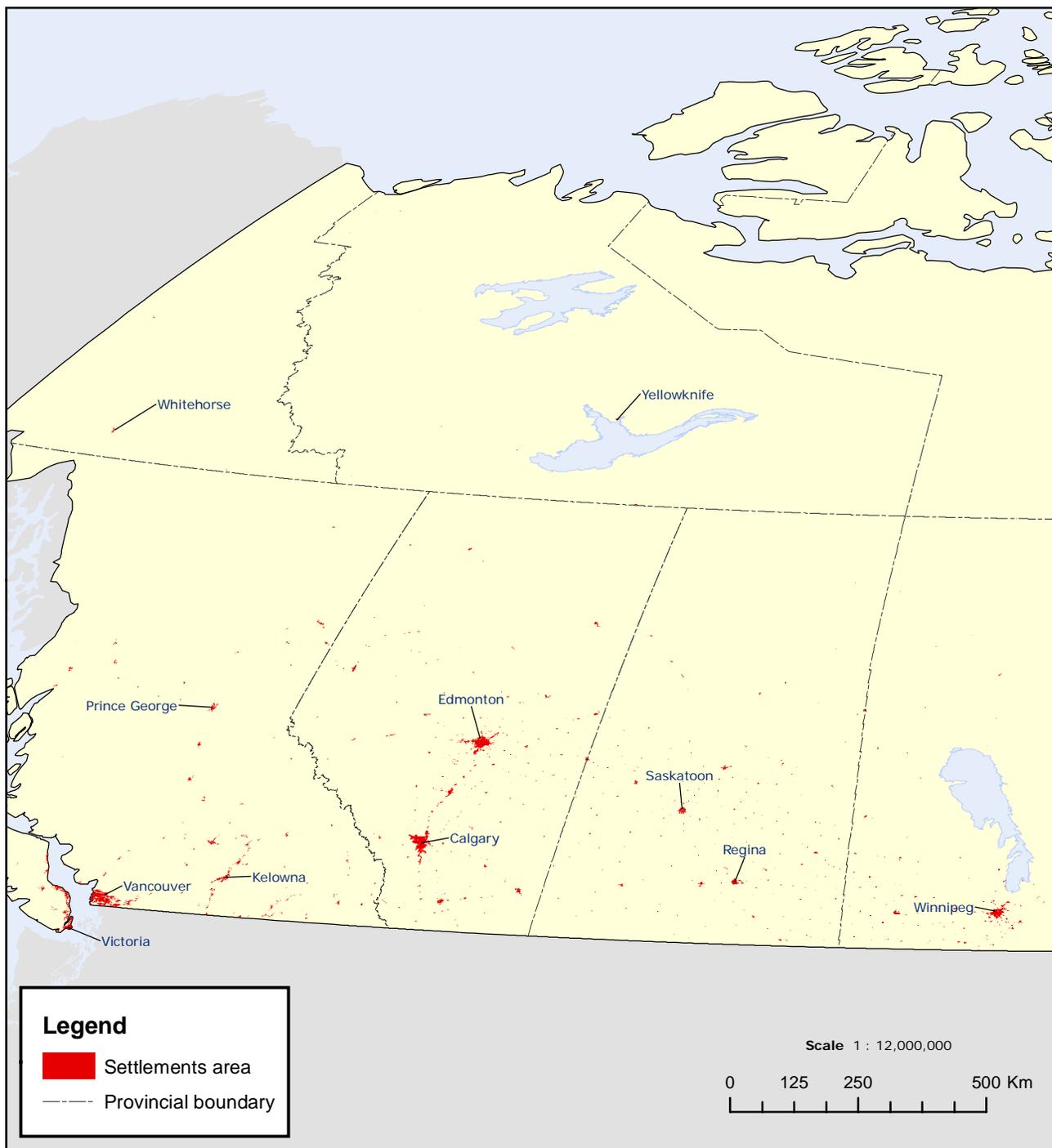
41. This figure includes adjustments made to the total area of blocks classified as settled by using the Earth Observation Inventory.

**Map 1**  
**Initial research results - settlements of Canada, 2006**



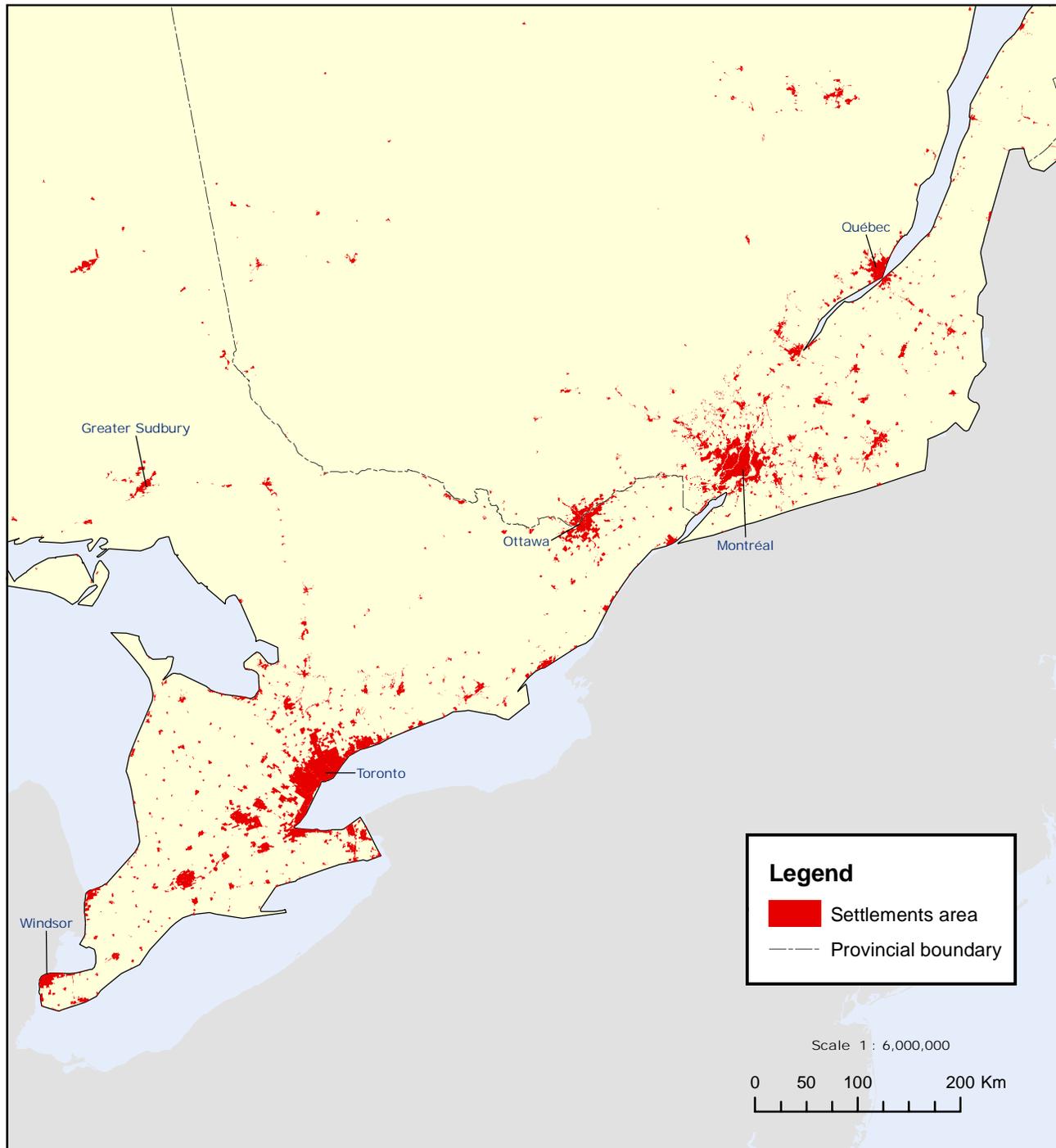
Source(s): Statistics Canada, Environment Accounts and Statistics Division.

Map 2  
Initial research results - settlements in western Canada, 2006



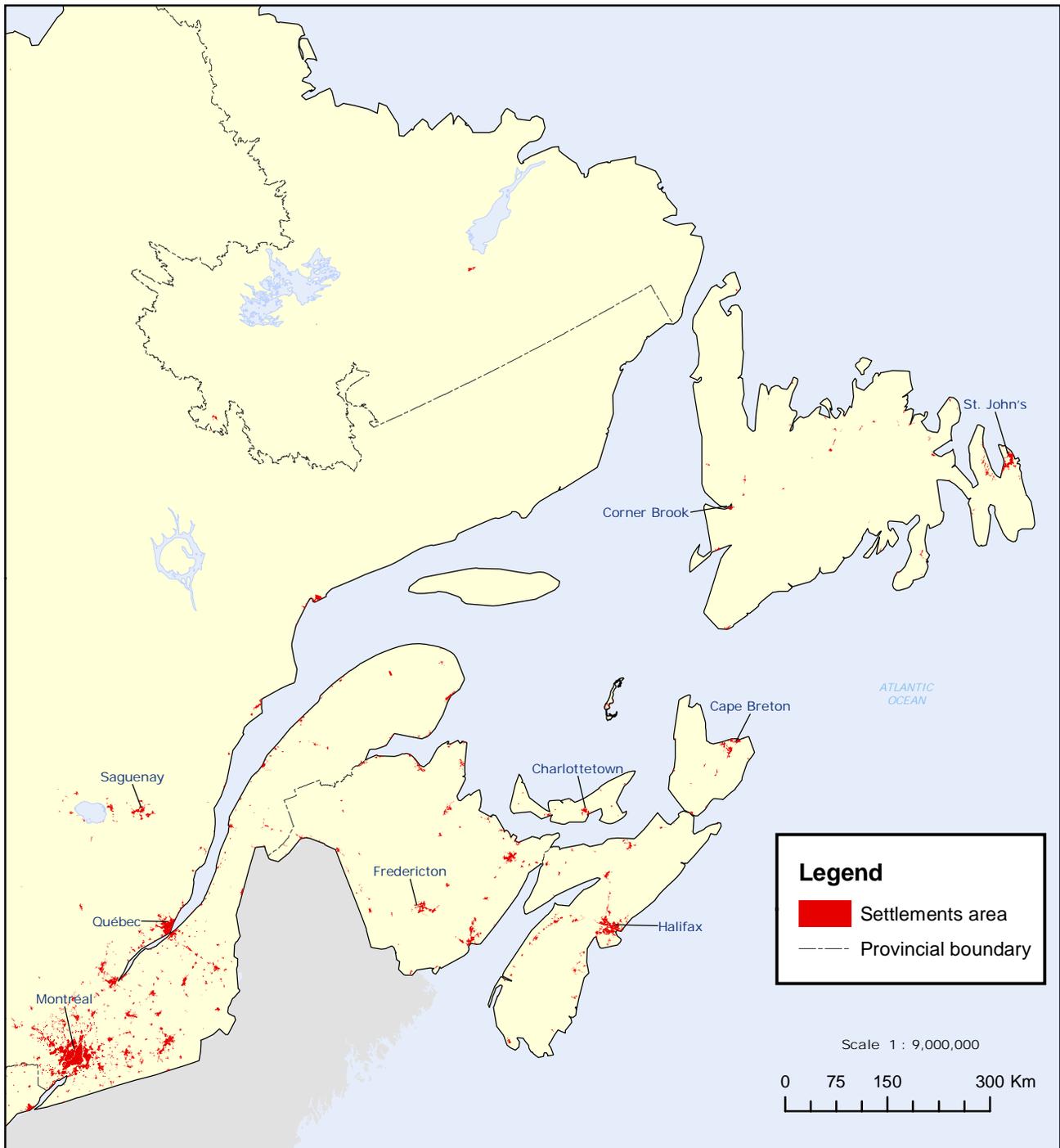
Source(s): Statistics Canada, Environment Accounts and Statistics Division.

Map 3  
Initial research results - settlements in central Canada, 2006



Source(s): Statistics Canada, Environment Accounts and Statistics Division.

Map 4  
Initial research results - settlements in eastern Canada, 2006



Source(s): Statistics Canada, Environment Accounts and Statistics Division.

## 5.2 Future directions

Using the methodology and data sources described in this study, a series of indicators are planned to help illuminate settlement-related issues in Canada. Several key themes have been proposed for further study. Proposed initial indicators may cover:

- land converted to settlements;
- density of settled areas;
- compactness/dispersion; and
- diversity of settlements.

The indicators will be based on two broad categories: settlements with a population over 500 and Canada's largest settlements.