Geography Working Paper Series
No. 1998-1

Part 1

Automated Land Area Tests for the 2001 Census: Preliminary Results Using the 1996 Digital Cartographic Files

by

Carolyn Weiss and Augustine Akuoko-Asibey

Geography Division
Statistics Canada

April 1998

Catalogue no. 92F0138MIE, no. 1998-1

ISSN 1481-174X

Aussi disponible en français.

Geography Working Paper Series

The Geography Working Paper Series is intended to stimulate discussion on a variety of topics covering conceptual, methodological or technical work to support the development and dissemination of the division's data, products and services. Readers of the series are encouraged to contact the authors with comments, criticisms and suggestions. A list of titles appears at the end of this paper.

A paper version, catalogue no. 92F0138MPE, is available for $10.00 per issue. Outside Canada the cost is US $10.00. All prices exclude sales tax. There is no charge for downloading the document yourself on the Internet (*http://www.statcan.ca*).

For inquiries about the working paper series, please contact:

Geo-Help
Geography Division
Statistics Canada
Jean Talon Building, 3rd floor
Ottawa, Ontario K1A 0T6

Telephone: (613) 951-3889
FAX: (613) 951-0569
Internet: geohelp@statcan.ca

TABLE OF CONTENTS

ABSTRACT

Prior to the 2001 Census, the Geography Division manually calculated land area using a planimeter. The manual approach was necessary since the division did not have a national digital database to support an automated method. The manual land area calculation was a labour intensive process – as well as very time consuming, expensive and very prone to error. With the anticipation of having a national digital base for the 2001 Census (to be called the National Geographic Base or NGB), the automated land area calculation becomes a simple GIS process.

The objectives of this paper are to show the effects of data aggregation and different map projections on automated land area calculation. The testing is a two-staged process; the first stage uses the 1996 Digital Cartographic Files (DCFs) and the second stage will use the NGB. Although the DCFs are not ideal for testing land area, this two-staged approach is necessary since the NGB will not be completed until Autumn 1998. This report presents results on the first stage using the 1996 DCFs.

The results reveal that the effects of data aggregation are minimal; thus land area can be stored at the census block level and then aggregated to higher level geographic entities without any serious ramifications for the 2001 Census. The results also indicate that an equal-area projection, specifically the Albers Equal-Area Conic projection, is more appropriate for calculating land area.

1. INTRODUCTION

Prior to the 2001 Census, the Geography Division manually calculated land area using a planimeter. The manual approach was necessary since the division did not have a national digital database to support an automated method. The manual land area calculation was a labour intensive process – as well as very time consuming, expensive and very prone to error (Weiss, 1996a,1996b). Land area measurements were done separately for census subdivisions (CSDs)[1], census tracts (CTs), designated places (DPLs) and urban areas (UAs).

With the anticipation of having a national digital base for the 2001 Census (to be called the National Geographic Base or NGB), the automated land area calculation becomes a simple GIS process.

The testing of automated land area calculation is a two-staged process; the first stage uses the 1996 Digital Cartographic Files (DCFs) and the second stage will use the NGB. Although the DCFs are not ideal for testing land area (see Section 2), this two-staged approach is necessary since the NGB will not be completed until Autumn 1998. This report presents results on the first stage using the 1996 DCFs; the tests involve determining the effects of data aggregation and different map projections on land area.

2. LIMITATIONS OF THE DIGITAL CARTOGRAPHIC FILES

Before describing the testing approaches, a short description of the Digital Cartographic Files (DCFs) is warranted in order to illustrate some of the limitations of those files. Obviously the accuracy of any automated area measurement is limited by the inaccuracy inherent in:

- the location and shape of the boundaries of the geographic areas.
- the Geographic Attribute Data Base (GADB) linkages of enumeration areas (EAs) to higher level geographies, since EAs are aggregated to form all administrative and statistical geographic areas.
- the location and shape of the digital coastline.
- the number of water polygons (lakes and double-line rivers) in the digital file.

The DCFs were produced to support small-scale thematic mapping only. It was a "one shot deal" only, and was not meant to support other infrastructure tasks such as calculating land area, generating representative points, calculating contiguity, etc.

The boundaries depicted on the Digital Boundary Files (DBF) were modified to follow the coastlines and shorelines at the perimeter of Canada's land mass. A separate map layer containing lakes, some double-line rivers and some estuaries was also created. The hydrographic features on the DCFs were obtained from the in-house Street Network Files (SNFs) and the Digital Hydrographic Base (DHB) from Natural Resources Canada (Statistics Canada, 1997, p. 5). Most of the map scales of the SNFs vary between 1:1,000 and 1:30,000 (though scales of 1:5,000, 1:10,000 and 1:25,000 are used most frequently); the scale of the DHB is 1:2,000,000. All DCF and SNF coordinates are stored in single precision, and the datum is NAD27.

In some cases shorelines were moved to ensure that all EA representative points fell on land, except for the legitimate cases of EAs entirely in water. Other modifications were made as well. Line generalization was performed on the DHB, ranging from minimal to extensive. Some polygons representing coastal islands, lakes, double-line rivers, and islands inside rivers were eliminated using various size thresholds (Statistics Canada, 1997, pp. 10-14).

[1]The CSD measurements were aggregated to obtain the land areas for the following higher level geographic areas: primary census metropolitan areas/primary census agglomerations (PCMAs/PCAs), census metropolitan areas/census agglomerations (CMAs/CAs), census consolidated subdivisions (CCSs), census divisions (CDs), economic regions (ERs), provinces/territories, and Canada.

All inland water bodies (except the Great Lakes and St. Lawrence River) are in a separate layer. Thus to calculate land area, both layers were merged, boundaries running through these inland water bodies were "dissolved", and new polygons formed.

It should be noted that the dissemination version of the province/territory DCF is a generalized file only. Thus an ungeneralized version was obtained from the product manager, and tests were performed on the ungeneralized file.

3. AGGREGATION TESTS

It is anticipated that data aggregations will affect the land area data such that the "sum of the parts might not equal the whole". For example, the land area for a single province might differ from aggregating the component CSD land areas for that same province. The purpose of the testing is to determine the magnitude. Two major geography processes will eventually be affected. The first is the aggregation of areas to higher level geographies. We expect that land area will be stored at the lowest level possible (e.g. at the census block level) and then sum the land areas of the blocks to any geographic level – that is, the land area measurement of any geographic entity is the sum of all census blocks comprising that entity. This approach is preferable to performing independent measurements for each geographic entity (U.S. Bureau of the Census, 1994, p. 15-10). The second process affected is the splitting of geographic areas when their limits change.

The aggregation tests include comparing provincial land areas to the sum of the component EA and CSD land areas, and comparing CSD land areas to the sum of the component EA land areas. The five provinces selected for the test – Newfoundland, Quebec, Ontario, British Columbia and Northwest Territories – are those that are considered large, and having intricate shorelines and many internal lakes. The detailed methodology, including the Arc/Info procedures, is presented in Appendix 1. Highlighted below, however, are some of the key methodological approaches:

- the Lambert Conformal Conic projection is used, utilizing national projection parameters for the five provinces. The effects of a particular map projection should not affect data aggregation.
- two levels of precision are examined, namely, four and two decimal places. The original eight decimals are individually rounded to four and two decimal places (i.e. four decimal places are not rounded to two). The Geography Division normally publishes land area data with a precision of two decimals. However, storing data in greater precision is a normal approach.
- the data are converted from square meters (the Arc/Info default output) to square kilometers. Land areas of the component geographic entities are first summed in their original output (square meters) and then converted to square kilometers. This procedure is considered more accurate than converting the land area of each component geographic entity to square kilometers and then summing (Janes, 1997).

Tables 1 and 2 present the results of aggregating the component EAs and CSDs to the provincial level. Numerous land polygons are created at the provincial level due to coastlines, islands, and inland lakes and double-line rivers. Consequently, it is necessary to aggregate the individual provincial land polygons to obtain the total provincial land area.[2] Note that the effects of EA and CSD aggregation are minimal. In fact, aggregating more geographic entities (EA) rather than aggregating fewer (CSD) does not necessarily imply poorer results.

The second test compares CSD land areas to the sum of their component EA land areas. The minimum, maximum, mean, and mean absolute differences are compared – as well as the standard deviation and the number of times (frequency) the differences are above, below or equal to zero (Table 3). Appendix 3, Part 3 contains the formulas for calculating the differences.

---

[2] A more accurate comparison would be comparing EA and CSD aggregations to one provincial land polygon, but that is not possible due to the nature of the DCFs.

Table 1. Comparison of Provincial Land Areas and Component EAs

PROVINCE			LAND AREA (KM^2)		
Name	No. of Land Polygons	No. of EAs*	Aggregation of Province Land Polygons	Aggregation of Province EAs	Absolute Difference
Newfoundland	437	1,231	375,371.8105 375,371.81	375,371.8216 375,371.82	0.0111 0.01
Quebec	531	11,682	1,395,857.0435 1,395,857.04	1,395,857.0379 1,395,857.04	0.0056 0.00
Ontario	581	16,465	938,845.0020 938,845.00	938,845.0423 938,845.04	0.0403 0.04
British Columbia	663	6,875	901,485.1423 901,485.14	901,485.1324 901,485.13	0.0099 0.01
Northwest Territories	387	169	3,131,172.8883 3,131,172.89	3,131,172.7648 3,131,172.76	0.1235 0.13

***Number includes ship EAs, but excludes unpopulated water EAs (see Appendix B in Statistics Canada, 1997, p. 71).

Table 2. Comparison of Provincial Land Areas and Component CSDs

PROVINCE			LAND AREA (KM^2)		
Name	No. of Land Polygons	No. of CSDs	Aggregation of Province Land Polygons	Aggregation of Province CSDs	Absolute Difference
Newfoundland	437	381	375,371.8105 375,371.81	375,371.8216 375,371.82	0.0111 0.01
Quebec	531	1,599	1,395,857.0435 1,395,857.04	1,395,857.0375 1,395,857.04	0.0060 0.00
Ontario	581	947	938,845.0020 938,845.00	938,845.0423 938,845.04	0.0403 0.04
British Columbia	663	713	901,485.1423 901,485.14	901,485.1323 901,485.13	0.0100 0.01
Northwest Territories	387	68	3,131,172.8883 3,131,172.89	3,131,172.7648 3,131,172.76	0.1235 0.13

Table 3. Comparison of CSD Land Areas and Component EAs

						NUMBER OF CSDs		
Province	**Minimum Difference**	**Maximum Difference**	**Mean Difference**	**Mean Absolute Difference**	**Standard Deviation**	**Diff. › 0**	**Diff. ‹ 0**	**Diff. = 0**
Nfld.	-0.0004	0.0051	0.0000	0.0000	0.0003	10	7	364
	0.00	0.01	0.00	0.00	0.00	1	0	380
Que.	-0.0027	0.0006	0.0000	0.0000	0.0001	11	14	1,574
	0.00	0.00	0.00	0.00	0.00	0	0	1,599
Ont.	-0.0020	0.0117	0.0000	0.0000	0.0004	8	12	927
	0.00	0.01	0.00	0.00	0.00	1	0	946
B.C.	-0.0048	0.0041	0.0000	0.0000	0.0003	120	115	478
	0.00	0.00	0.00	0.00	0.00	0	0	713
N.W.T.	-0.0115	0.0028	-0.0003	0.0004	0.0019	2	3	63
	-0.01	0.00	0.00	0.00	0.00	0	2	66

The results are indeed promising. Note that the differences are negligible. Also note that the frequency of differences above and below zero (Table 3, columns 7 and 8) are rather symmetrical, which means that the data are not skewed. In fact, most often there is no difference at all (column 9), regardless of the number of decimal places.

Thus we can safely recommend that land areas can be stored at the census block level and then aggregated to higher level geographic entities without any serious ramifications for the 2001 Census.

4. MAP PROJECTION TESTS

It is also anticipated that land area calculations will vary according to the map projection used. A map projection is both the process and result of transforming positions on the spherical surface of the earth onto a plane (flat) surface. That is, the curved, three-dimensional surface of the earth is transformed onto a flat, two-dimensional plane.

If we assume for simplicity that a globe, which is a sphere, can perfectly represent the surface of the earth, then all of the following characteristics must be true of that globe: areas are everywhere correctly represented; all distances are correctly represented; all angles are correctly represented; and the shape of any area is faithfully represented (Snyder and Voxland, 1989, p. 5).

When the sphere is projected onto a plane, the map will no longer have all of these characteristics or properties simultaneously. One way of classifying projections is through terms describing the extent to which the projection geometry preserves any of those properties. An **equal-area (equivalent)** map projection retains representation of areas so that all regions are shown in correct relative size. An **equidistant** map projection correctly represents distances along a line from only one or two points to any other point on the map. An **azimuthal** projection correctly shows directions from one or, at the most, two points. A projection that maintains correct angles at infinitely small locations is called **conformal**, meaning correct form or shape[3] (Hsu, 1981, pp. 158-163; Muehrcke and Muehrcke, 1992, pp. 568-571; Robinson et al., 1995, pp. 63-66; Snyder and Voxland, 1989, p. 5).

[3]A persistent misunderstanding of the conformal projection is that it preserves the shape of the *entire* mapped area. It does not; the shape of a very small area is nearly correct on a conformal projection (Hsu, 1981, p. 161).

Perhaps the most striking trade-off in map projection is between conformality and equivalence. No projection can be both conformal and equivalent. Not only are these properties mutually exclusive, but in parts of the map well removed from the standard parallels conformal maps severely exaggerate area and equal-area maps severely distort shape (Monmonier, 1996, p. 14).

Since the four basic spatial properties cannot all be held true simultaneously (if they could, there would be no "projection problem"), it is very important to select a projection having the characteristics that are best suited to the application at hand. The main goal of map projection selection is concerned with reducing distortion; selection should always lead to the map projection showing the least amount of distortion as compared to other projections suitable for the same application (De Genst and Canters, 1996, p. 146). In addition to the projection property requiring preservation, Snyder (1987, pp. 34-35) notes that the selection process is affected by other characteristics, such as: size of region (world, hemisphere, continent, or smaller region); directional extent of region (predominant east-west, predominant north-south); and general location of region (polar, equatorial, mid-latitude). For example, shapes are greatly distorted on conformal projections of the whole world (American Cartographic Association, 1988, p. 6).

Of the 46 map projections available in Arc/Info (ESRI, 1997), we selected four map projections for the test; two are conformal and two are equal-area. Two conformal projections, the **Lambert Conformal Conic** and the **Transverse Mercator**, are selected since they had, and still have, a significant role in Canadian mapping. The former projection is used extensively for general maps of Canada at small scales (including the *National Atlases of Canada*); the latter projection is the one used for Canada's National Topographic System (NTS) at 1:50,000 and 1:250,000 that is also superimposed by a UTM grid (Nicholson and Sebert, 1981, pp. 184-187). The two equal-area projections, **Lambert Azimuthal Equal-Area** and **Albers Equal-Area Conic** are obviously chosen since the equal-area property is indispensable for calculating land area. The oblique aspect for the Lambert Azimuthal is used (rather than the equatorial or polar aspects) since it is more appropriate for North America.

The five provinces used for examining aggregation effects are used in the projections tests (Newfoundland, Quebec, Ontario, British Columbia and Northwest Territories). The first test is based on the four projections mentioned above, using provincial projection parameters (Table 4). The detailed Arc/Info procedures are found in Appendix 1, Part 1. Appendix 2 presents a sample procedure for calculating provincial projection parameters and Appendix 3 lists the projection parameters for each province.

The most notable pattern evident in Table 4 is that the Transverse Mercator generates the largest land areas for all provinces. This indicates that in order to preserve conformality (shape), the areas are exaggerated – especially the Northwest Territories. The Lambert Conformal Conic generates the smallest land areas for three provinces (Quebec, Ontario and Northwest Territories) and the Lambert Azimuthal Equal-Area for two provinces (Newfoundland and British Columbia).

Table 4. Provincial Land Areas (Km^2) Using Provincial Map Projection Parameters

Province	Lambert Conformal Conic	Transverse Mercator	Lambert Azimuthal Equal-Area	Albers Equal-Area Conic
Newfoundland	382,905.9814	384,485.3814	382,646.9599	384,196.4789
Quebec	1,420,555.4876	1,431,067.3871	1,421,132.4146	1,426,940.6975
Ontario	941,045.0409	947,126.7424	941,750.7215	945,031.3903
British Columbia	930,476.3396	933,208.0995	927,429.6081	931,568.5517
Northwest Territories	3,197,345.0605	3,325,440.7977	3,255,346.8449	3,278,318.9003

Since the property of equal area is important in making land area computations, we eliminate the two conformal projections (Lambert Conformal Conic and Transverse Mercator). The Lambert Azimuthal Equal-Area is also rejected. Although this projection preserves area, it is useful for countries that have nearly equal east-west and north-south dimensions (Dent, 1990, p. 71; Robinson et al., 1995, p. 79). However, Canada's areal extent is not symmetrical; it has a much greater east-west extent than north-south. More importantly, Lambert Azimuthal Equal-Area is less accurate than Albers Equal-Area Conic, since the former projection uses only a sphere as the reference surface, whereas Albers is based on an ellipsoid.

This leaves the Albers Equal-Area Conic. This projection is used and recommended for mid-latitude regions having greater east-west than north-south extent, and obviously for maps requiring the preservation of area (Hsu, 1981, p. 171; Muehrcke and Muehrcke, 1992, p. 318; Robinson et al., 1995, p. 79; Snyder and Voxland, 1989, p. 100). However, the Arc/Info documentation also suggests that the total range in latitude from north to south should not exceed 30^o to 35^o for the Albers projection (ESRI, 1997). Even though Canada's north-south range is about 42^o, we do not believe that exceeding this range by 7^o is sufficiently serious.

The Albers projection using provincial projection parameters (Table 4, column 5) is next compared to the national projection parameters (Table 5). The differences are quite minor – with Newfoundland, Quebec and Northwest Territories having differences only after the decimal point, and Ontario and British Columbia having differences before the decimal point. We cannot explain the latter, even though the variations are only 8 km^2 and 6 km^2 respectively.

Table 5. Provincial Land Areas Using National Map Projection Parameters, Albers Equal-Area Conic Projection

Province	Land Area (Km^2)
Newfoundland	384,196.5874
Quebec	1,426,940.6870
Ontario	945,039.5079
British Columbia	931,562.5548
Northwest Territories	3,278,318.4552

National Parameters

1^{st} standard parallel: 49^o	Latitude of projection origin: 63^o 23' 26.43"
2^{nd} standard parallel: 77^o	False easting: 6,200,000
Central meridian: -91^o 52'	False northing: 3,000,000

In the early 1970s, Natural Resources Canada calculated land areas for all provinces. Table 6 shows their published land areas, only for those provinces used in our tests. Although the NRCan's methodology and map sources are completely different from ours[4] – and in a way is like comparing apples and oranges – we examined the differences merely out of "curiosity" (Tables 5 and 6). Except for the Northwest Territories, note that the land areas generated from the 1996 DCF are much greater than those from NRCan; this may indicate that the DCF does not contain a sufficient amount of water polygons. The reverse is true for the Northwest Territories, and perhaps this signifies the DCF does not contain enough Arctic islands.

[4]In the early 1970's, the areas of entire 1:50,000 and 1:250,000 NTS map sheets were calculated mathematically. A planimeter was then used to measure areas of freshwater, seawater and foreign territory in order to determine land area for Canada (Sebert and Munro, 1972). Later on a provincial breakdown was made, and in 1981 the measurement for Yukon Territory was adjusted (Gosson, 1997).

Table 6. Provincial Land Areas Published by Natural Resources Canada

Province	Land Area (Km^2)
Newfoundland	371,690
Quebec	1,356,790
Ontario	891,190
British Columbia	929,730
Northwest Territories	3,293,020

Source: Natural Resources Canada; published in Statistics Canada, 1996, p. 22.

Based on the findings of our map projection tests, we recommend that the Albers Equal-Area Conic be used for calculating land area. At this time, however, we are not certain whether national or provincial projection parameters should be used.

5. SUMMARY AND RECOMMENDATIONS

This document represents the first stage of testing land area calculation using the 1996 DCFs. As previously mentioned, although the DCFs are not ideal for testing land area, this approach is necessary since the NGB will be not completed until Autumn 1998.

We show that the aggregation effects are minimal. We therefore recommend that land areas be stored at the census block level and then aggregated to higher level geographic entities for the 2001 Census. Since aggregation effects are essentially independent of the quality of digital files used, it is not necessary to test aggregation using the NGB. We also demonstrate the effects of map projection have on land area calculation using two conformal and two equal-area projections. The Albers Equal-Area Conic projection appears to be the most appropriate for preserving the spatial property of area, although at this time we are not certain whether provincial or national projection parameters should be employed.

The second stage of testing land area will be based on the pre-2001 version of the NGB. The NGB will be far less generalized than the DCFs – and will be based on NAD83 (more precise than our current NAD27), and the coordinates will be stored in double precision (more precise than our current single precision). We recommend that the map projection tests be continued using the NGB.[5]

We also offer the following general observations:

- land areas must be calculated excluding <u>all</u> water bodies. This requires merging the census block boundaries with all coastlines, shorelines, islands and internal water bodies to determine land areas. No generalization should take place merely because the geographic boundaries and hydrography are stored in separate layers.
- land area values be calculated in double precision, stored in four decimal places, and disseminated in two decimal places.

[5]We recommend that Arc/Info's "describe" command be used to obtain the coverage limits in degrees, minutes and seconds. These coordinates are required to calculate the standard parallels, latitude of projection origin and central meridian. Since the ymin, ymax, xmin and xmax coordinates were determined manually using the *1996 CSD Reference Maps* (see note in Appendix 2), the parameters require recalculation for the map projection tests using NGB.

- accurate digital census boundaries and correct census block linkages to higher level geographies are more crucial now than ever with the automated approach for calculating land area.
- consideration be given to creating a special flag in the Geographic Attribute Data Base (GADB) to identify those census blocks that are single-address sub-blocks. Since these sub-blocks are to be digitally represented as trapezoids, it makes no sense to calculate land areas for these artificially sized symbols. This problem is compounded not only by "stacked" sub-blocks (i.e. an apartment building that contains more than one EA, with each EA being a group of floors), but also by single-address EAs occupying an entire block (e.g. the Civic Hospital). This requires that the trapezoids be dissolved into their surrounding block when calculating land area, since it is important to retain the morphology of the block for land area calculations.
- consideration be given to creating another special flag in GADB to identify those census blocks that are ships, oil tankers, oil rigs and houseboats, as well as those that are water blocks. These blocks should be excluded in land area calculations.

ACKNOWLEDGEMENTS

Many thanks are extended to Darryl Janes and Henry Puderer for providing useful comments on this report.

REFERENCES

American Cartographic Association (1988). *Choosing a World Map*, Special Publication No. 2 (Falls Church, VA: American Congress on Surveying and Mapping).

De Genst, William and Canters, Frank (1996). Development and Implementation of a Procedure for Automated Map Projection Selection, *Cartography and Geographic Information Systems* 23(3): 145-171.

Dent, Borden D. (1990). *Cartography: Thematic Map Design*, 2nd Edition (Dubuque, IA: Wm. C. Brown Publishers).

ESRI (1997). *Arc/Info Online Documentation*, Version 7.1 (Redlands, CA: Environmental Systems Research Institute).

Gosson, Claire (1997). Personal communication, December 1997.

Hsu, Mei-Ling (1981). The Role of Projections in Modern Map Design, *Cartographica* 18(2): 151-186.

Janes, Darryl (1997). Personal communication, November 1997.

Monmonier, Mark (1996). *How to Lie with Maps*, 2nd Edition (Chicago: University of Chicago Press).

Muehrcke, Phillip C. and Muehrcke, Juliana O. (1992). *Map Use: Reading, Analysis and Interpretation*, 3rd Edition (Madison, WI: JP Publications).

Nicholson, N.L. and Sebert, L.M. (1981). *The Maps of Canada: A Guide to Official Canadian Maps, Charts, Atlases and Gazetteers* (Hamden, CT: Archon Books).

Nyerges, Timothy L. and Jankowski, Piotr (1989). A Knowledge Base for Map Projection Selection, *The American Cartographer* 16(1): 29-38.

Robinson, Arthur H., Morrison, Joel L., Muehrcke, Phillip C., Kimerling, A. Jon and Guptill, Stephen C. (1995). *Elements of Cartography*, 6th Edition (New York: John Wiley & Sons).

Sebert, L.M. and Munro, M.R. (1972). *Dimensions and Areas of Maps of the National Topographic System of Canada*, Technical Report No. 72-1 (Ottawa: Department of Energy, Mines and Resources).

Snyder, John P. (1982). *Map Projections Used by the U.S. Geological Survey*, Geological Survey Bulletin 1532 (Washington, DC: U.S. Government Printing Office).

Snyder, John P. (1987). *Map Projections – A Working Manual*, U.S. Geological Survey Professional Paper 1395 (Washington, DC: U.S. Government Printing Office).

Snyder, John P. and Voxland, Philip M. (1989). *An Album of Map Projections*, U.S. Geological Survey Professional Paper 1453 (Washington, DC: U.S. Government Printing Office).

Statistics Canada (1996). *1997 Canada Year Book*, Catalogue No. 11-402-XPE/1997 (Ottawa: Industry Canada).

Statistics Canada (1997). *Digital Boundary Files and Digital Cartographic Files, 1996 Census*, Reference Guide, Catalogue No. 92F0029XDE, 92F0030XDE and 92F0032XDE to 92F0040XDE (Ottawa: Industry Canada).

U.S. Bureau of the Census (1994). Area Measurement/Water Classification, *Geographic Areas Reference Manual* (Washington, DC: U.S. Bureau of the Census), 15-1 to 15-11.

Weiss, Carolyn C. (1996a). *Wither Land Area? A Case for Retaining the Status Quo for the 1996 Census*, Geography Working Paper Series, No. 1996-1 (Ottawa: Geography Division, Statistics Canada).

Weiss, Carolyn C. (1996b). *Qu'adviendra-t-il du calcul de la superficie des terres? Justification du maintien du statu quo pour le recensement de 1996*, Série de document de travail de la géographie, N° 1996-1 (Ottawa: Division de la géographie, Statistique Canada).

METHODOLOGY FOR AGGREGATION TESTS

Part 1. Determining Province, CSD and EA Land Areas

1. Convert geographic coordinates to Lambert Conformal Conic Projection. This applies to all coverages: Province, CSD, EA and Water.
ARC: USAGE PROJECT
: PROJECT <COVER | GRID | FILE> <INPUT> <OUTPUT> <PROJECTION _FILE>
Project: OUTPUT
Project: PROJECTION LAMBERT
UNITS METERS
PARAMETERS
1st standard parallel: 49 00 00
2nd standard parallel: 77 00 00
Central meridian: -91 52 00
Latitude of projection origin: 63 23 26.43
False easting: 6200000
False northing: 3000000
END

2. Build polygons after converting from geographic coordinates to Lambert.
ARC: USAGE BUILD
BUILD <COVER> {POLY | LINE | POINT | NODE | ANNO. <subclass>}

3. Merge Province, CSD or EA coverage with the Water coverage using the IDENTITY command in ARC. This procedure maintains all input coverage features after the merger, and eliminates all identity coverage features overlapping the "in-cover".
ARC: USAGE IDENTITY
IDENTITY <IN_COVER> <IDENTITY_COVER> <OUT_COVER> {POLY | LINE | POINT} <FUZZY_TOLERANCE> {JOIN | NOJOIN}

4. Select the Province, CSD or EA land areas. In the 1996 DCF Reference Guide, Water has a value of "1" in ARC/INFO. The RESELECT command in ARC results in the selection of areas covered with water. In order to obtain the land area, issue the NSEL command.
ARC: RESELECT <IN_COVER> <OUT_COVER> < IN_FEATURE_CLASS>
RES WATER = 1
<return>
NO
YES
NSEL
N
N
END

5. Use STATISTICS in ARC to determine PRUID, CSDUID and EAUID frequencies, and individual land areas and total land area in square meters.
ARC: STATISTICS <IN_INFO_FILE> <OUT_INFO_FILE> <CASE_ITEM>
STATISTICS: SUM <TARGET_ITEM>
END

LIST <OUT_INFO_FILE>

6. Convert land area to square kilometers (in INFO) because the original areas are in square meters. Since 1,000,000 m^2 = 1 km^2, the original land area values are divided by 1,000,000. The data were originally stored in 3 decimals; this was changed to 8 decimals to ensure greater precision.

```
ENTER COMMAND >SELECT <INFO_FILE>
ENTER COMMAND >CALCULATE AREA = AREA / 1000000
ENTER COMMAND >LIST
```

After the conversion, the land area data were rounded from the original 8 decimal places to 4 and 2 decimal places.

```
ENTER COMMAND >SELECT <INFO_FILE>
ENTER COMMAND >ALTER
ITEM NAME>
ITEM OUTPUT WIDTH>
ITEM TYPE>
N. DECIMAL PLACES>

ENTER COMMAND >Q STOP
```

Part 2. Determining the Sum of EA Land Areas to CSD Level

Two data files are utilized:

A. A file from GADB, that denotes EA linkages to the CSD level.
B. An output file from Part 1, Step 5, listing EAs and their corresponding land areas in square meters.

1. The original data file from GADB denoting the EA linkage to the CSD level contained information for all the provinces and territories in Canada. Therefore, a SAS program was written to select data for the required province. In addition, all water EAs on the DCF and all water EAs on DBF but not on DCF were eliminated with the SAS program. The file was saved as an ASCII file.

2. Get the saved ASCII file into INFO.

```
ENTER COMMAND >DEFINE <IDENTIFIER>
ITEM NAME>
ITEM WIDTH>
ITEM OUTPUT WIDTH>
ITEM TYPE>
DECIMAL PLACES>
PROT. LEVEL>

ITEM NAME>
<RETURN>

ENTER COMMAND> GET < FILE_NAME> COPY ASCII
ENTER COMMAND> Q STOP
```

3. Join INFO file from Part 2, Step 2 and final output file from Part 1 (i.e. the EA land area output from Step 5). The reason for joining these files is to determine the number of EAs in each CSD and their corresponding total land area. Before joining these files, make sure the items have the same character format. Moreover, the link items must have common names. These can be done in INFO. Now proceed to join the two INFO files. Sort both data files by a common item in INFO before issuing the JOINITEM command.

```
ARC:  JOINITEM <IN_INFO_FILE> <JOIN_INFO_FILE> <OUT_INFO_FILE> <RELATE_ITEM>
<START_ITEM> [LINEAR|ORDERED|LINK]
```

List output file.

```
ARC: LIST <OUT_INFO_FILE>
```

4. Use STATISTICS in ARC to determine the number of EAs in each CSD and their corresponding areas.

```
ARC:  STATISTICS <IN_INFO_FILE> <OUT_INFO_FILE> <CASE_ITEM>
STATISTICS: SUM <TARGET_ITEM>
END
```

5. List output file.

```
ARC: LIST <OUT_INFO_ FILE>
```

Part 3. Comparing Sum of EA Land Areas to CSD Level and Original CSD Land Area

The purpose is to calculate the minimum difference, maximum difference, mean difference, mean absolute difference – as well as the standard deviation of the difference between the sum of EA land areas to CSD level (from Part 2) and the original CSD land areas (from Part 1).

1. List final output files generated from Parts 1 and 2 (Steps 6 and 5, respectively) to verify common items, characters, headings and decimal places. Sort both files by a common item.

2. Join the two output files.

```
ARC: JOINITEM <IN_INFO_FILE> <JOIN_INFO_FILE> <OUTPUT> <JOIN_ITEM> <START_ITEM> [LINEAR|ORDERED|LINK]
```

3. Calculate the difference between the constituent EA land area aggregations and original CSD land area. Before proceeding with the calculation, add an item to the attribute table of the output file in Part 3, Step 2 with the ADDITEM command in ARC.

```
ARC: ADDITEM <IN_INFO_FILE> <OUT_INFO_FILE> <ITEM_NAME> <ITEM_WIDTH> <OUTPUT_WIDTH><ITEM_TYPE> <DECIMAL_PLACES> <START_ITEM>
```

In INFO calculate the difference.

```
ENTER COMMAND >SELECT <INFO_FILE>
ENTER COMMAND: CALCULATE <DF_ITEM | SYSTEM_ITEM | SYSTEM_VARIABLE> = <ARITHMETIC EXPRESSION>
```

4. Use STATISTICS in ARC to calculate the minimum difference, maximum difference, mean difference as well as the standard deviation. The formula for calculating the mean difference is:

$$N^{-1}\sum_{i=1}^{N}\left(X_i - Y_i\right)$$

where N = number of CSDs
X_i = sum of EA land areas to CSD level
Y_i = original CSD land area

```
ARC:  STATISTICS <IN_INFO_FILE> <OUT_INFO_FILE> <CASE_ITEM>
STATISTICS: MEAN <TARGET_ITEM>
:MIN <TARGET_ITEM>
:MAX <TARGET_ITEM>
:STD <TARGET_ITEM>
:END
```

List output file.

ARC: LIST <OUTPUT_FILE>

5. Calculate the mean absolute difference using the INFO command. The formula is:

$$N^{-1}\sum_{i=1}^{N}\left|X_i - Y_i\right|$$

where N = number of CSDs
X_i = sum of EA land areas to CSD level
Y_i = original CSD land area

ENTER COMMAND >SELECT <INFO_FILE>
ENTER COMMAND >RES {difference less than zero}
ENTER COMMAND >CALCULATE {difference less than zero * -1}.

Multiply by -1 in order to obtain absolute values.

ENTER COMMAND >ASELECT
ENTER COMMAND >LIST
ENTER COMMAND >Q STOP

Issue the STATISTICS command in ARC and stipulate MEAN <TARGET_ITEM>.

6. In INFO use RESELECT to determine:
   A. Number of times difference greater than zero
   B. Number of times difference less than zero
   C. Number of times difference equal to zero

ENTER COMMAND >SELECT <INFO_FILE>
RESELECT {BY | FOR} <LOGICAL_EXPRESSION>

Appendix 2

SAMPLE PROCEDURE FOR CALCULATING PROVINCIAL PROJECTION PARAMETERS (ONTARIO)

1. Calculate the parameters for **Lambert Conformal Conic** and **Albers Equal-Area Conic** projections.

   A. 1st and 2nd Standard Parallels. The "one-sixth rule" is applied to calculate the 1st and 2nd standard parallels. The procedure is as follows:

      Most southerly latitude: 41° 30' 00"
      Most northerly latitude: 56° 43' 00"
      Range between north and south latitudes = 15° 13' 00"

      Applying the "one-sixth rule", divide the range by six (i.e. 15° 13' 00" ÷ 6 = 2° 32' 10"). Add the result to the south latitude to obtain the 1st standard parallel, and subtract from the north latitude to arrive at the 2nd standard parallel. Thus:
      1st standard parallel = 44° 02' 10"
      2nd standard parallel = 54° 10' 50"

   B. Central Meridian. The longitudes of the eastern-most and western-most parts of the province are summed up and divided by two, as follows:
      (-74° 35' 00" + -95° 10' 00") ÷ 2 = -84° 52' 30"

   C. Latitude of Projection Origin. This lies mid-way between the most northerly and most southerly latitudes of the province. It is obtained by dividing the range between the north and south latitude (15° 13' 00") by two, and either adding the result to the south latitude or subtracting the result from the north latitude. For example:
      15° 13' 00" ÷ 2 = 7° 36' 30"
      56° 43' 00" - 7° 36' 30" = 49° 06' 30"

   D. False Eastings and False Northings. Set to zero.

2. The **Transverse Mercator** employs the same projection parameters as the Lambert Conformal Conic and Albers Equal-Area Conic projections for the central meridian, latitude of origin and false northing. The scale factor at central meridian is set at 0.9996, and the false easting is set at 500,000 meters.

3. The **Lambert Azimuthal Equal-Area (Oblique Aspect)** employs the same projection parameters as the Lambert Conformal Conic and Albers Equal-Area Conic projections for the longitude of center of projection, latitude of center of projection, false easting and false northing. For this projection the default radius of sphere of reference is used (6,370,997 meters).

NOTE: For the five provinces included in the test, the most northerly and southerly latitude coordinates and the most easterly and westerly longitude coordinates were manually determined using the *1996 CSD Reference Maps*. The ymin/ymax latitude coordinates are required to calculate the standard parallels and latitude of projection origin; the xmin/xmax longitude coordinates are required to calculate the central meridian.

Appendix 3

PROVINCIAL MAP PROJECTION PARAMETERS

Newfoundland

Lambert Conformal Conic

1st standard parallel: 49° 02' 31"
2nd standard parallel: 58° 06' 15"
Central meridian: -60° 11' 43"
Latitude of projection origin: 53° 34' 23"
False easting: 0 (default)
False northing: 0 (default)

Albers Equal-Area Conic

1st standard parallel: 49° 02' 31"
2nd standard parallel: 58° 06' 15"
Central meridian: -60° 11' 43"
Latitude of projection origin: 53° 34' 23"
False easting: 0 (default)
False northing: 0 (default)

Transverse Mercator

Scale factor at central meridian: 0.9996
Longitude of central meridian: -60° 11' 43"
Latitude of origin: 53° 34' 23"
False easting: 500,000
False northing: 0 (default)

Lambert Azimuthal Equal-Area (Oblique Aspect)

Radius of sphere of reference: 6,370,997 (default)
Longitude of center of projection: -60° 11' 43"
Latitude of center of projection: 53° 34' 23"
False easting: 0 (default)
False northing: 0 (default)

Quebec

Lambert Conformal Conic

1st standard parallel: 47° 55' 33"
2nd standard parallel: 59° 37' 47"
Central meridian: -68° 24' 47"
Latitude of projection origin: 53° 46' 40"
False easting: 0 (default)
False northing: 0 (default)

Albers Equal-Area Conic

1st standard parallel: 47° 55' 33"
2nd standard parallel: 59° 37' 47"
Central meridian: -68° 24' 47"
Latitude of projection origin: 53° 46' 40"
False easting: 0 (default)
False northing: 0 (default)

Transverse Mercator

Scale factor at central meridian: 0.9996
Longitude of central meridian: -68° 24' 47"
Latitude of origin: 53° 46' 40"
False easting: 500,000
False northing: 0 (default)

Lambert Azimuthal Equal-Area (Oblique Aspect)

Radius of sphere of reference: 6,370,997 (default)
Longitude of center of projection: -68° 24' 47"
Latitude of center of projection: 53° 46' 40"
False easting: 0 (default)
False northing: 0 (default)

Ontario

Lambert Conformal Conic

1st standard parallel: 44° 02' 10"
2nd standard parallel: 54° 10' 50"
Central meridian: -84° 52' 30"
Latitude of projection origin: 49° 06' 30"
False easting: 0 (default)
False northing: 0 (default)

Albers Equal-Area Conic

1st standard parallel: 44° 02' 10"
2nd standard parallel: 54° 10' 50"
Central meridian: -84° 52' 30"
Latitude of projection origin: 49° 06' 30"
False easting: 0 (default)
False northing: 0 (default)

Ontario (cont'd)

Transverse Mercator
Scale factor at central meridian: 0.9996
Longitude of central meridian: -84° 52' 30"
Latitude of origin: 49° 06' 30"
False easting: 500,000
False northing: 0 (default)

Lambert Azimuthal Equal-Area (Oblique Aspect)
Radius of sphere of reference: 6,370,997 (default)
Longitude of center of projection: -84° 52' 30"
Latitude of center of projection: 49° 06' 30"
False easting: 0 (default)
False northing: 0 (default)

British Columbia

Lambert Conformal Conic
1st standard parallel: 50° 15' 28"
2nd standard parallel: 58° 03' 05"
Central meridian: -126° 31' 40"
Latitude of projection origin: 59° 04' 17"
False easting: 0 (default)
False northing: 0 (default)

Albers Equal-Area Conic
1st standard parallel: 50° 15' 28"
2nd standard parallel: 58° 03' 05"
Central meridian: -126° 31' 40"
Latitude of projection origin: 59° 04' 17"
False easting: 0 (default)
False northing: 0 (default)

Transverse Mercator
Scale factor at central meridian: 0.9996
Longitude of central meridian: -126° 31' 40"
Latitude of origin: 59° 04' 17"
False easting: 500,000
False northing: 0 (default)

Lambert Azimuthal Equal-Area (Oblique Aspect)
Radius of sphere of reference: 6,370,997 (default)
Longitude of center of projection: -126° 31' 40"
Latitude of center of projection: 59° 04' 17"
False easting: 0 (default)
False northing: 0 (default)

Northwest Territories

Lambert Conformal Conic
1st standard parallel: 56° 55' 33"
2nd standard parallel: 77° 57' 50"
Central meridian: -98° 38' 05"
Latitude of projection origin: 67° 26' 40"
False easting: 0 (default)
False northing: 0 (default)

Albers Equal-Area Conic
1st standard parallel: 56° 55' 33"
2nd standard parallel: 77° 57' 50"
Central meridian: -98° 38' 05"
Latitude of projection origin: 67° 26' 40"
False easting: 0 (default)
False northing: 0 (default)

Transverse Mercator
Scale factor at central meridian: 0.9996
Longitude of central meridian: -98° 38' 05"
Latitude of origin: 67° 26' 40"
False easting: 500,000
False northing: 0 (default)

Lambert Azimuthal Equal-Area (Oblique Aspect)
Radius of sphere of reference: 6,370,997 (default)
Longitude of center of projection: -98° 38' 05"
Latitude of center of projection: 67° 26' 40"
False easting: 0 (default)
False northing: 0 (default)

Geography Working Paper Series

<u>Number</u>

1993-1 ***A Comparison of Census Geographic Areas of Canada and the United States***, Carolyn Weiss, Michael Ratcliffe and Nancy Torrieri (November 1993)

1996-1 ***Wither Land Area? A Case for Retaining the Status Quo for the 1996 Census***, Carolyn Weiss (December 1996)

1998-1 ***Automated Land Area Tests for the 2001 Census: Preliminary Results Using the 1996 Digital Cartographic Files***, Carolyn Weiss and Augustine Akuoko-Asibey (April 1998)