Geography Working Paper Series
No. 1996-1

Wither Land Area? A Case for Retaining the Status Quo for the 1996 Census

by

Carolyn Weiss

Geography Division
Statistics Canada

December 1996

Catalogue no. 92F0138MIE, no. 1996-1

ISSN 1481-174X

Aussi disponible en français.

This paper represents the views of the authors and does not necessarily reflect the opinions of Statistics Canada. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by Statistics Canada.

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

The purpose of this document is to report on the reasoning for continuing to use the digital planimeter for calculating land area for the 1996 Census. Despite the fact that the data are subject to a number of errors and logical inconsistencies, the Geography Division does not yet have a digital database in place to support the automated calculation of land area.

This report first provides an historical background on the treatment of land area. It then presents the results of some land area tests based on 1991 data. The tests include the actual use of the digital planimeter, as well as logical consistency checks that compare land area with total area. Finally, a description of the 1996 Digital Cartographic File of enumeration areas (EA/DCF) demonstrates the reason why that file is not employed to calculate land area in an automated fashion.

1. INTRODUCTION

The purpose of this document is to report on the reasoning for continuing to use the digital planimeter for calculating land area for the 1996 Census. Despite the fact that the data are subject to a number of errors and logical inconsistencies, the Geography Division does not yet have a digital database in place to support the automated calculation of land area.

This report first provides an historical background on the treatment of land area. It then presents the results of some land area tests based on 1991 data. The tests include the actual use of the digital planimeter, as well as logical consistency checks that compare land area with total area. Finally, a description of the 1996 Digital Cartographic File of enumeration areas (EA/DCF) demonstrates the reason why that file is not employed to calculate land area in an automated fashion.

The scope of this report is not to be considered exhaustive on this topic. Further details, including additional tables, can be found in Caldwell, Lefebvre and Storey (1994), Storey (1994), and Storey and Tupper (1995).

2. HISTORICAL BACKGROUND

The Geography Division has traditionally employed manual methods for calculating land area. For the past several decades, that method relied on the planimeter. Land area measurements are done separately for census subdivisions (CSDs)[1], census tracts (CTs), designated places (DPLs) and urban areas (UAs). Land area is not calculated for enumeration areas (EAs) and federal electoral districts (FEDs).

The mechanics of the planimeter itself has changed over the years. Prior to the 1986 Census an analogue planimeter was used; this planimeter necessitated using a coefficient to input the scale of the map. For the 1986 Census onwards, a newer digital planimeter was employed. Users of this digital planimeter had to key in the scale of the map, and the instrument had to be calibrated every once and a while.

The process of measuring land area with a digital planimeter is a manual task. In order to control for measurement variation – and in some cases to "trap" measurement errors – the same geographic area is measured three times and then the three readings are averaged. The digital planimeter gives accurate readings for zones that are about 30 cm in diameter, as physically represented on the map (Placom Company, p. 7). Consequently, large geographic areas are subdivided into smaller ones and measured individually; the individual parts are then added together. Discernible water bodies as found on the maps are excluded. The digital planimeter, however, cannot handle extremely small polygons; in those cases a grid is used to measure small water bodies.

For CSDs, land areas are measured from 1:50,000 and 1:250,000 topographic maps. CTs are initially delineated on 1:50,000 topographic maps, and all land area measurements (and boundary revisions) are done on those map sheets. Only discernible bodies of water as found on the maps are excluded.

This manual land area calculation is a labour intensive process. It is very time consuming, expensive and prone to error. Until 1991, the amount and extent of these errors have not been researched or documented. For the 1991 Census a digital boundary file for EAs was created for the first time, thus making it possible not only to digitally aggregate EA limits to all higher geographic areas, but also to compare land area with total area.

There are, however, many "anecdotal" accounts about the possible sources of error. These include:

[1]The CSD measurements are 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.

- the accuracy of manually transcribing boundaries on maps.
- scale, accuracy and vintage of the source maps.
- incorrect use of the digital planimeter, including inputting the incorrect map scale.
- incorrect measurements using a grid for small polygons.
- taking measurements on paper maps. Paper maps stretch or shrink without strict temperature or humidity controls. As well, some paper maps might have been folded and unfolded many times.
- incorrect calculation of the mean (e.g. dividing by two instead of three).
- transcription errors, either on the coding forms or on the Geographic Attribute Data Base (GADB).
- methodology of recording land area changes. Land areas are not recalculated when boundary changes occur. Only land areas gained or lost due to boundary revisions are calculated, which are then added to and subtracted from the historically computed areas. The quality of such calculations, therefore, depends heavily on the first time the land area was calculated – and if incorrect, has allowed errors to propagate and accumulate over time.
- not subtracting the land areas lost from the original land areas for geographic areas undergoing boundary revisions. The GADB "edit" procedure is supposed to check that the land area gained and lost balances to zero (0).

Some of the above accounts are backed up by recent documentation. For example, Caldwell, Lefebvre and Storey (1994) remeasured three 1991 census tracts in Hamilton CMA using a digital planimeter. They show that two CTs were inversely coded in GADB and that one CT's original measurement was incorrect by 10 km^2. Another case pertains to two published 1991 land areas in Winnipeg CMA, brought to our attention by the Winnipeg Regional Office (Statistics Canada, 1996, p. 27). The published values were incorrect and were remeasured as follows:

- CT 0510.01 Published (1991) – 1.32 km^2; remeasured – 11.58 km^2
- CT 0510.02 Published (1991) – 18.42 km^2; remeasured – 7.10 km^2

If the remeasured land areas are added together (18.68 km^2), note that it almost equals the published figure for CT 0510.02 (18.42 km^2). Thus it appears that when the original CT (0510.00) was split, the split portion was not subtracted from it. We cannot explain why the GADB edit procedure did not catch this, nor can we explain the published value of 1.32 km^2.

One particular written procedure for calculating land area could potentially lead to errors. The procedure states that when measuring areas on a 1:50,000 map sheet, water bodies that are equal to or greater than 0.01 km^2 (about 2 mm x 2 mm) should be excluded. The document further instructs that if an area is calculated on a 1:250,000 map sheet, water bodies to be excluded should be determined and calculated on a 1:50,000 map sheet (Statistics Canada, 1994, p. 20). Consequently, working with two different map scales could possibly result in recording an incorrect scale for the planimeter measurements.

Undocumented evidence also abounds. For example, one individual informed this author about his calculations for CSD land areas in the early 1980s. He was advised that any new CSD land area calculations plus the unrevised CSD land areas must aggregate to the historic provincial land area. However, when the totals did not match the provincial level, then any excess land area from the newly calculated CSD was put in a large, northern CSD in the same province (or vice versa, i.e. any deficit was taken from a large, northern CSD).[2] Apparently, it was more important to be more accurate for land area (and thus population density) for the urbanized CSDs. Other situations also involved taking "short cuts". If there were too many small water bodies to measure with a grid, then he would estimate the number of lakes and their corresponding grid cells.

Other undocumented evidence pertains to "verbal" procedures, specifically regarding the three measurements. Staff are normally instructed to discard a measurement if it is significantly different than the other two, and to remeasure the zone. This approach, of course, avoids the situation of one very inaccurate

[2] There is no evidence that similar procedures were applied to the CSD/CMA level, or to ensure that the CT and CSD land area aggregations to the CMA/CA level were identical.

measurement throwing off the average. However, since this process is not a written procedure, it is possible that a supervisor could forget to inform his/her staff of this strategy.

3. LAND AREA TESTS

Two distinct tests were conducted using 1991 Census data, namely, the actual use of the digital planimeter and logical consistency checks. The logical consistency checks involve comparing land area with total area, and aggregating land areas for CSDs and CTs to the CMA/CA level.

3.1 Digital Planimeter

The purpose of this test was to determine the type and extent of errors in using the planimeter. The Winnipeg CMA was selected due to land area challenges for census tracts presented to the Geography Division from the Winnipeg Regional Office.

The number of census tracts (5) and test subjects (4) for the test is rather small, primarily because we anticipate that the 1996 Census represents the last census in which the planimeter will be used to calculate land area. In essence, it would have been counterproductive to conduct a large, manual and lengthy test when the future indicates that land area will be calculated using an automated method. As well, time and resources were rather scarce.

The five CTs were chosen to vary in terms of areal extent (small vs. large), configuration (smooth vs. irregular) and amount of water features contained within them. The census tract boundaries were transposed onto 1:50,000 topographic maps; the boundaries were transcribed twice (i.e. resulting in two sets of maps) so that the participants in the test were not slowed down in the measurement process.

Four individuals from the Operations and Integration Division (OID) participated in the test, based on their experience using a digital planimeter. They include one with no experience ("rookie"), one with limited experience ("novice") and two who have used the planimeter extensively in the past ("pro"); see Table 1. Each individual received written procedures on using the instrument, and were instructed to measure each CT three times and obtain an average for the three readings. As well, the persons with limited and no experience were given a chance to familiarize themselves with the planimeter prior to commencing the test.

During the early stages of the test, it was noticed that bizarre readings were obtained when measuring large areas. The two experienced users indicated that the planimeter gives accurate readings for only small areas. When measuring large areas, the planimeter display measurement overflows and then resets to zero (0). Experience obtained during the test showed that the largest area to be measured without questionable results was no more than about 10 cm x 10 cm, as physically represented on the map. As a result, the two large CTs were divided into small parts, each individual land area measured, the individual parts added together and then averaged. CT 0580.00 was divided into 9 parts and CT 0595.00 into 20 parts. The number of parts were the same on the two map sets, but their configuration was different. (This is quite unfortunate, since any testing process should keep the variables as constant as possible.) Another peculiarity of the planimeter is that at a precision of three decimal places it automatically rounds to multiples of 0.025.

The test results presented in Table 1 are not surprising. As expected, the averages are different than the published values, since different staff were involved in measuring the land areas. As well, it is not unusual for the same individual to obtain the same or close measurements for small land areas (see CTs 0023.00, 0102.02 and 0510.02). What cannot be explained, however, is why both the rookie and novice had identical readings for CT 0102.02.

Table 1. Digital Planimeter Test in Winnipeg CMA, 1991 Census

Person	CT Name	LAND AREA (KM^2) 1st Reading	2nd Reading	3rd Reading	Average	Published Value	Absolute Difference
Rookie	0023.00	0.425	0.425	0.475	0.442	0.52	-0.078
	0102.02	1.125	1.075	1.125	1.108	1.27	-0.162
	0510.02	7.275	7.525	7.325	7.375	***18.42**	0.275
	0580.00	207.175	206.825	206.800	206.933	214.92	-7.987
	0595.00	556.775	557.450	555.485	556.575	559.44	-2.865
Novice	0023.00	0.425	0.425	0.425	0.425	0.52	-0.095
	0102.02	1.125	1.075	1.125	1.108	1.27	-0.162
	0510.02	7.375	7.325	7.375	7.358	***18.42**	0.258
	0580.00	208.050	208.200	208.450	208.233	214.92	-6.687
	0595.00	333.525	333.025	333.325	333.292	559.44	-226.148
Pro	0023.00	0.425	0.425	0.425	0.425	0.52	-0.095
	0102.02	0.925	0.925	0.925	0.925	1.27	-0.345
	0510.02	8.675	8.625	8.625	8.641	***18.42**	1.541
	0580.00	209.775	209.800	209.775	209.783	214.92	-5.137
	0595.00	588.125	588.425	588.250	588.266	559.44	28.826
Pro	0023.00	0.450	0.450	0.500	0.466	0.52	-0.054
	0102.02	1.150	1.075	1.175	1.133	1.27	-0.137
	0510.02	7.375	7.600	7.425	7.466	***18.42**	0.366
	0580.00	213.975	214.125	213.725	213.725	214.92	-1.195
	0595.00	558.325	557.925	558.150	558.133	559.44	-1.307

***** The published land area of **18.42 km^2** for CT 0510.02 is incorrect. It was remeasured using a digital planimeter and should read 7.10 km^2; see Section 2. *The correct value of 7.10 km^2 is used to calculate the difference in the last column.*

For the first three CTs with relatively small land areas, the individual and average readings seem to lie within a natural or normal measurement variation when compared to the published value. However, much larger measurement variances between the average and published values occur for the two CTs that are subdivided into parts (CTs 0580.00 and 0595.00). In particular, note the novice's average measurement for CT 0595.00. He had aggregated the 20 parts into larger areas (perhaps to save time?). More than likely, a resulting overflow was not noted when the planimeter reset to zero. This aspect of the test demonstrates a significant error. Another significant error, of course, is the published land area for CT 0510.02 (see Table 1 footnote and Section 2). These types of errors are often undetected and are carried over into various census years. In some cases, perhaps they can be considered more serious than actually using the planimeter.

If one compares the results between the "rookie" and the first "pro", note that the latter produced a gross deviant value of nearly 29 km^2 for CT 0595.00, whereas the rookie did quite well overall. Thus we can speculate that it is not always experience that counts when using the planimeter.

One interesting comment by the second "pro" pertains to a particular methodology followed when using the planimeter. He noted that it is common practice to use the published CSD land area values when a CT(s) rolls-up to a CSD. This practice can either lead to another serious source of error propagation if the CSD land area value is incorrect, or conversely, can avoid the introduction of error if the CT land area measurement is incorrect.

3.2 Logical Consistency

The purpose of this test is to determine the logical consistency of land areas, specifically for census subdivisions and census tracts (the land areas for designated places and urban areas were not examined). Firstly, land area is compared with total area. Secondly, the land areas aggregations of CSDs and CTs to the CMA/CA level are compared.

Total area refers to all land and water areas within the official limits of the census geographic areas. The calculations are derived from the Digital Boundary File (DBF) for enumeration areas, and their aggregations to all higher level geographic areas. Total area is determined using the area calculation function of ARC/INFO® in the Lambert Conformal Conic projection.

Obviously, land area should be equal to or smaller than total area. However, if the situation is reversed (i.e. land area greater than total area), this signifies a logical error. It should be noted, however, that there could be errors even when land area is smaller than total area; this aspect of the problem is beyond the scope of this report. As well, the artifact of different mechanical instruments and methodologies result in discrepancies – land area measurements are made with a planimeter and total area measurements are derived from digital boundaries, which in turn were digitized using a cursor. Even the digital boundaries are not error-free; for example, there are known digitizing errors as well as linkage errors that generate boundary errors (Statistics Canada, 1996). Nevertheless, comparing land area with total area does give us a general “feel” of the situation.

Overall, the discrepancies between land area and total area appear quite serious. Nearly 20% of the CSDs have land areas greater than total areas and 31% for CTs (Table 2, columns 3 and 4). Practically, however, one has to accept some level of measurement variance which will generate logical inconsistencies of this nature – and perhaps incidences in the very low single-digit range should provide a comfort level for users. A few staff members of the Geography Division who have dealt extensively with land area issues contend that a serious discrepancy would apply to land areas greater than total area by over 1 km^2. By accepting this threshold, note that the rate of logical inconsistencies decreases substantially – to nearly 6% for CSDs and 2% for CTs (Table 2, columns 5 and 6).[3] However, as will be shown next, even these overall percentages can “mask” larger individual discrepancies.

Table 3 compares land area and total area for CSDs by province. For those CSDs having land area greater than total area by over 1 km^2, note that the overall percentage (5.6%) masks those provinces having higher, more critical, rates. Only Prince Edward Island, Nova Scotia and Saskatchewan appear to be within a tolerable range.

Table 4 shows the top 80 discrepancies between land area and total area for individual CSDs. The extent of the logical inconsistencies speak for themselves. Most of them occur in Quebec and British Columbia. Note especially the largest discrepancy for the CSD in the Yukon Territory, whereby its land area is greater than the total area by nearly 48,000 km^2!

Similar to the CSDs, for those census tracts having land area greater than total area by over 1 km^2, the overall percentage (1.7%) also masks the critical, large discrepancies (Table 5). Nearly 54% of the CMAs/CAs (21 out of 39) have CT land areas greater than total areas by over 1 km^2.

---

[3]It has been It has been pointed out that using an absolute threshold of greater than 1 km^2 is harsher for CSDs than CTs (Currie, 1996). For example, a 1 km^2 measurement variance for a CT that is 4 km^2 is a serious discrepancy, whereas the same measurement variance for a CSD that is 400 km^2 is not.

Table 2. Summary Statistics for CSDs and CTs, 1991 Census

	Total No. of Units	No. of Units Land > Total Area	Percent	No. of Units Land > Total Area by > 1 km^2	Percent
Census Subdivisions	6,006	1,196	19.9	337	5.6
Census Tracts	4,068	1,268	31.2	71	1.7

Table 3. Occurrences of Land Area Greater Than Total Area for CSDs, by Province, 1991 Census

Province	Total No. of CSDs	No. of CSDs Land > Total Area	Percent	No. of CSDs Land > Total Area by > 1 km^2	Percent
Newfoundland	404	62	15.3	15	3.7
Prince Edward Island	126	18	14.3	1	0.8
Nova Scotia	118	14	11.9	1	0.8
New Brunswick	287	62	21.6	37	12.9
Quebec	1,637	297	18.1	121	7.4
Ontario	951	141	14.8	46	4.8
Manitoba	293	38	13.0	14	4.8
Saskatchewan	953	204	21.4	16	1.7
Alberta	438	121	27.6	19	4.3
British Columbia	691	216	31.3	45	6.5
Yukon	36	13	36.1	13	36.1
Northwest Territories	72	10	13.9	9	12.5
Total	**6,006**	**1,196**	**19.9**	**337**	**5.6**

Table 4. Land Area and Total Area for CSDs (Top 80 Discrepancies), 1991 Census

Province	CSD Name and Type	KM²		
		Land Area	Total Area	Absolute Difference
Yukon Territory	Yukon, Unorganized, UNO	530,100.08	482,112.37	47,987.71
Quebec	Lac-Nilgaut, UNO	11,593.67	9,866.12	1,727.55
Alberta	Improvement District No. 12, ID	12,629.40	11,218.52	1,410.88
Quebec	Chute-des-Passes, UNO	18,336.56	17,040.00	1,296.56
Quebec	Lac-au-Brochet, UNO	11,228.82	10,306.48	922.34
Quebec	Mont-Valin, UNO	39,217.40	38,297.87	919.53
Quebec	Matchi-Manitou, UNO	16,760.57	15,954.46	806.11
Quebec	Rivière-Kipawa, UNO	13,384.59	12,775.29	609.30
British Columbia	East Kootenay, Subd. B, SRD	9,242.78	8,708.95	533.83
British Columbia	Central Kootenay, Subd. B, SRD	12,896.24	12,376.29	519.95
British Columbia	Fraser-Fort George, Subd. B, SRD	15,707.19	15,202.17	505.02
Ontario	**Head, Clara and Maria, TP***	800.07	304.80	495.27
Ontario	Matachewan, ID	1,081.04	600.54	480.50
British Columbia	Kitimat-Stikine, Subd. C, SRD	10,295.73	9,922.08	373.65
Quebec	Baie-Atibenne, UNO	978.87	637.68	341.19
British Columbia	Peace River, Subd. C, SRD	28,977.93	28,638.61	339.32
British Columbia	Fraser-Cheam, Subd. B, SRD	1,324.34	1,011.91	312.43
British Columbia	Columbia-Shuswap, Subd. A, SRD	13,677.29	13,372.47	304.82
British Columbia	Dewdney-Alouette, Subd. A, SRD	2,614.63	2,317.93	296.70
British Columbia	North Okanagan, Subd. A, SRD	2,085.55	1,836.47	249.08
British Columbia	Kitimat-Stikine, Subd. B, SRD	7,756.65	7,520.81	235.84
Quebec	Saint-Marc-de-Figuery, P	323.71	93.69	230.02
Quebec	Whapmagoostui, VC	308.46	121.77	186.69
Quebec	Rivière-Nouvelle, UNO	1,272.35	1,088.12	184.23
Quebec	Lac-Juillet, UNO	3,716.10	3,540.02	176.08
British Columbia	Alberni-Clayoquot, Subd. A, SRD	2,202.34	2,032.33	170.01
Newfoundland	Division No. 6, Subd. D, SUN	4,739.19	4,572.58	166.61
Ontario	Cochrane, Unorganized, North Part, UNO	136,679.33	136,515.44	163.89
Quebec	Lac-Quentin, UNO	495.03	341.27	153.76
Quebec	Lac-Boisbouscache, UNO	231.65	102.66	128.99
Quebec	Lac-Vacher, UNO	653.01	551.20	101.81
Yukon Territory	Beaver Creek, SET	123.13	25.67	97.46
Yukon Territory	Tagish, SET	97.50	4.74	92.76
Alberta	**Bighorn No. 8, MD***	1,282.35	1,195.84	86.51
Yukon Territory	Old Crow, SET	104.00	18.07	85.93
New Brunswick	Eldon, PAR	1,774.61	1,690.78	83.83
British Columbia	Squamish-Lillooet, Subd. A, SRD	7,444.09	7,362.67	81.42
Alberta	O'Chiese 203, R	138.77	69.09	69.68
Newfoundland	Division No. 6, Subd. A, SUN	5,983.89	5,919.77	64.12
Quebec	Saint-Guillaume-Nord, UNO	874.06	810.41	63.65

*** Head, Clara and Maria** (CSD 3547098) has a boundary error on the CSD/DBF, and **Bighorn No. 8** (CSD 4815015) has EA linkage errors on the EA/DBF that generates a boundary error on the CSD/DBF (see Statistics Canada, 1996, pp. 67-69).

Table 4. Land Area and Total Area for CSDs (Top 80 Discrepancies), 1991 Census (Cont'd)

Province	CSD Name and Type	KM²		
		Land Area	Total Area	Absolute Difference
New Brunswick	Drummond, PAR	1,075.23	1,017.75	57.48
Quebec	Lac-Despinassy, UNO	1,925.67	1,871.43	54.24
Saskatchewan	Kinistino No. 459, RM	1,019.46	965.66	53.80
British Columbia	Thompson-Nicola, Subd. B, SRD	4,602.57	4,549.22	53.35
Ontario	Iroquois Falls, T	689.94	639.42	50.52
New Brunswick	Northfield, PAR	352.94	304.10	48.84
Quebec	Waltham-et-Bryson, CU	451.43	403.33	48.10
Northwest Territories	Grise Fiord, HAM	153.63	108.09	45.54
Alberta	Hay Lake 209, R	167.83	123.18	44.65
British Columbia	East Kootenay, Subd. A, SRD	10,943.87	10,900.32	43.55
New Brunswick	Saint Marys, PAR	800.74	757.40	43.34
Quebec	Tourelle, SD	149.84	107.52	42.32
Quebec	Lac-De la Bidière, UNO	1,715.86	1,673.78	42.08
Northwest Territories	Wrigley, SET	142.97	101.03	41.94
Alberta	Stoney 142, 143, 144, R	449.70	408.74	40.96
Yukon Territory	Ross River, SET	65.63	27.25	38.38
Ontario	Schreiber, TP	74.49	36.69	37.80
Ontario	Red Rock, TP	102.70	66.86	35.84
Ontario	Golden, TP	564.14	530.43	33.71
Quebec	Eastmain, TR	151.45	118.00	33.45
Quebec	Lac-Chicobi, UNO	780.09	747.88	32.21
Quebec	Lac-Casault, UNO	1,502.09	1,472.85	29.24
New Brunswick	Blissfield, PAR	1,273.09	1,244.64	28.45
New Brunswick	Ludlow, PAR	1,051.82	1,023.95	27.87
Quebec	Cloutier, SD	128.98	101.87	27.11
Ontario	West Nissouri, TP	239.88	213.21	26.67
Manitoba	Split Lake 171, R	41.03	15.70	25.31
New Brunswick	Rogersville, PAR	351.75	326.93	24.82
Northwest Territories	Trout Lake, SET	118.55	94.69	23.86
Northwest Territories	Fort Liard, HAM	148.39	125.68	22.71
Quebec	Waswanipi, VC	265.34	243.06	22.28
Quebec	Saint-Alexis-des-Monts, P	521.18	499.07	22.11
Quebec	Saint-Zénon, P	512.86	491.04	21.82
Quebec	Lac-Montanier, UNO	283.91	262.63	21.28
Yukon Territory	Pelly Crossing, SET	35.13	14.18	20.95
New Brunswick	Saint-Léonard, PAR	366.04	345.10	20.94
Quebec	Alleyn-et-Cawood, CU	346.64	325.92	20.72
Alberta	Pincher Creek No. 9, MD	2,516.28	2,495.79	20.49
Yukon Territory	Johnson's Crossing, SET	43.93	23.49	20.44
Alberta	Leduc County No. 25, CM	2,732.29	2,712.21	20.08

Table 5. Occurrences of Land Area Greater Than Total Area for CTs, by CMA/CA, 1991 Census

CMA/CA Name	Total No. of CTs	No. of CTs Land > Total Area	Percent	No. of CTs Land > Total Area by > 1 km^2	Percent
Brantford	21	8	38.1	–	–
Calgary	153	59	38.6	5	3.3
Chicoutimi-Jonquière	35	5	14.3	1	2.9
Edmonton	195	61	31.3	10	5.1
Guelph	21	6	28.6	–	–
Halifax	75	10	13.3	–	–
Hamilton	163	79	48.5	1	0.6
Kamloops	22	11	50.0	2	9.1
Kelowna	26	5	19.2	–	–
Kingston	35	5	14.3	–	–
Kitchener	82	33	40.2	–	–
Lethbridge	21	12	57.1	–	–
London	88	28	31.8	1	1.1
Matsqui	29	19	65.5	10	34.5
Moncton	23	6	26.1	2	8.7
Montréal	749	219	29.2	4	0.5
North Bay	20	4	20.0	1	5.0
Oshawa	49	15	30.6	–	–
Ottawa-Hull	211	48	22.7	3	1.4
Peterborough	23	6	26.1	–	–
Prince George	23	5	21.7	1	4.3
Québec	152	40	26.3	1	0.7
Red Deer	16	8	50.0	1	6.3
Regina	49	13	26.5	–	–
Saint John	44	8	18.2	3	6.8
Sarnia-Clearwater	24	7	29.2	1	4.2
Saskatoon	49	11	22.4	–	–
Sault Ste. Marie	23	5	21.7	1	4.3
Sherbrooke	31	5	16.1	2	6.5
St. Catharines-Niagara	83	18	21.7	1	1.2
St. John’s	40	7	17.5	1	2.5
Sudbury	38	7	18.4	–	–
Thunder Bay	30	7	23.3	1	3.3
Toronto	812	327	40.3	7	0.9
Trois-Rivières	34	8	23.5	–	–
Vancouver	299	76	25.4	2	0.7
Victoria	65	8	12.3	1	1.5
Windsor	59	16	27.1	3	5.1
Winnipeg	156	53	34.0	5	3.2
Total	**4,068**	**1,268**	**31.2**	**71**	**1.7**

Table 6 compares land area with total area for individual CTs; only the top 45 discrepancies are shown. Some of the cases are rather serious, especially those differences that are double-digit. Note the number of times Edmonton and Matsqui are cited in the table; combined these CMAs/CAs represent 33% of the cases.

Matsqui represents a very interesting situation. Since Matsqui first entered into the CT program for the 1991 Census, it does not have the “history” of error propagation over time. Thus one should expect that the CT land areas be equal to or smaller than their respective total areas, and overall perform much better than some of the older tracted centres. However, out of a total of 29 CTs for Matsqui, 19 have land areas greater than total area, of which ten exceed 1 km^2 (Table 7). The latter figure represents 34% of the total number of CTs, a discrepancy rate that is overly high for a newly tracted centre. On the other hand, Red Deer – which like Matsqui entered into the CT program for the 1991 Census – fared much better. Out of a total of 16 CTs, 8 have land areas greater than total area, but only one CT is greater than 1 km^2.

So far we have examined discrepancies using absolute differences between land area and total area. However, there is another way of examining the data – namely, by calculating the percentage differences between land area and total area (Table 8). This approach takes into account that a small absolute variance for a small geographic area (e.g. 1 km^2 for a CT that is 2 km^2) can result in a very large percentage (50%), whereas a small absolute difference for a large geographic area (e.g. 1 km^2 for a CT that is 200 km^2) is not as significant (0.5%). This fact is illustrated in Table 8; the shaded rows indicate that the same CTs are also denoted in Table 6 (12 cases). Note, for example, CT 0103.00 for Matsqui. The absolute difference is nearly 2 km^2 (Table 6, row 42), but the percentage difference is 69% (Table 8, row 5). Conversely, the worst absolute difference is CT 0166.00 for Edmonton (Table 6), but it does not show up in the top 45 discrepancies using percentage differences (Table 8). Also note the number of times Montréal and Toronto are cited in Table 8; combined these CMAs represent 53% of the cases.

The second logical consistency test compares land area aggregations of CSDs and CTs to the CMA/CA level. Land area measurements for CSDs and CTs are, for the most part, done independently of each other. Thus one can expect the aggregations to differ (Table 9). Some of the absolute differences are negligible, and three CMAs/CAs even have no difference at all (Matsqui, Sarnia-Clearwater and Victoria). This is probably not a “coincidence”, as it is sometimes common practice to use the published CSD land areas as benchmark values when CTs are nested within CSDs. Apparently, if the component CT land areas do not add up to the CSD land area, then the CT land areas are “altered” until they do. This approach, of course, results in error propagation if the CSD land areas are incorrect.

Other absolute differences are more serious, such as Kelowna (49.05 km^2) and Calgary (36.04 km^2). When these differences are grouped into ranges (Table 10), one can see that 41% of the CMAs/CAs have differences less than 1 km^2. However, 58% of the cases exhibit differences of over 1 km^2 – a discrepancy rate that is rather high. Even if just the differences of greater than 5 km^2 are isolated, a discrepancy rate of 25% can still be considered excessive.

4. AUTOMATION

An automated method for calculating land area is a relatively simple GIS exercise. However, an appropriate national digital database is required to support it.

Consideration was given to using the newly created 1996 EA Digital Cartographic File (and the EA aggregations to higher level geographic areas) for calculating land area. The digital sources include: EA boundaries from the Digital Boundary File; shoreline from SNF; and shoreline from the National Atlas Information Service (NAIS). The scale of the NAIS shoreline is 1:2,000,000.

After considerable thought, we decided not to use the EA/DCF for the following reasons:

Table 6. Land Area and Total Area for CTs (Top 45 Discrepancies - Absolute Difference), 1991 Census

CMA/CA Name	CT Name	KM² Land Area	Total Area	Absolute Difference
Edmonton*	0166.00	1,533.79	1,369.40	164.39
Edmonton	0142.04	405.35	243.86	161.49
St. John’s	0300.00	100.84	9.57	91.27
Moncton	0016.01	375.62	303.36	72.26
Edmonton	0104.03	189.28	132.54	56.74
Sarnia-Clearwater*	0110.03	269.82	218.07	51.75
London	0130.00	239.88	213.21	26.67
Calgary	0038.19	84.31	60.96	23.35
Matsqui	0100.00	114.94	95.41	19.53
Edmonton	0142.01	605.31	592.00	13.31
Edmonton	0160.00	466.41	454.56	11.85
Kamloops	0019.00	55.74	44.38	11.36
Winnipeg*	0510.02	18.42	7.21	11.21
North Bay	0102.00	273.11	262.02	11.09
Sault Ste. Marie	0102.00	96.40	86.41	9.99
Saint John	0027.02	29.42	19.66	9.76
Windsor	0155.00	134.52	126.75	7.77
Winnipeg*	0580.00	214.92	207.97	6.95
Hamilton*	0085.02	7.88	1.71	6.17
Matsqui	0001.00	33.01	26.95	6.06
Toronto	0516.10	9.50	3.93	5.57
Victoria	0155.01	142.42	136.89	5.53
Edmonton	0163.00	37.50	32.24	5.26
Calgary	0002.03	54.90	49.71	5.19
Edmonton	0162.00	21.77	17.17	4.60
Montréal	0682.01	28.98	25.31	3.67
Edmonton	0079.00	67.78	64.31	3.47
Sherbrooke	0111.02	138.40	135.19	3.21
Matsqui	0002.00	11.58	8.48	3.10
Montréal	0415.01	31.52	28.54	2.98
Chicoutimi-Jonquière	0002.01	5.51	2.68	2.83
Toronto	0587.02	150.41	147.65	2.76
Vancouver*	0115.00	7.59	4.86	2.73
Montréal	0682.02	15.02	12.31	2.71
Montréal	0686.00	34.86	32.16	2.70
Saint John	0120.02	20.85	18.33	2.52
Edmonton	0104.02	106.53	104.29	2.24
Toronto	0412.04	9.18	6.95	2.23
Calgary	0002.01	64.90	62.74	2.16
Windsor	0140.00	114.81	112.93	1.88

Table 6. Land Area and Total Area for CTs (Top 45 Discrepancies - Absolute Difference), 1991 Census (Cont'd)

CMA/CA Name	CT Name	KM² Land Area	Total Area	Absolute Difference
Matsqui	0010.00	3.41	1.56	1.85
Matsqui	0103.00	2.62	0.80	1.82
Red Deer	0015.00	3.38	1.71	1.67
Québec	0520.00	110.83	109.20	1.63
Matsqui	0008.00	7.84	6.28	1.56

***** Edmonton, Sarnia-Clearwater and Vancouver have boundary errors on the CT/DBF, and Edmonton and Sarnia-Clearwater have EA/CT linkage errors on GADB (see Statistics Canada, 1996, pp. 67-68, 72). See Table 1 (and footnote) regarding Winnipeg. In Hamilton, the land areas for CTs 0085.02 and 0085.03 were coded inversely in GADB; the land area for CT 0085.02 should be 1.70 km^2 (see Caldwell, Lefebvre and Storey, 1994).

NOTE: The shaded rows indicate that the same CTs are also denoted in Table 8.

Table 7. Land Area and Total Area for CTs, Matsqui CA, 1991 Census

CT Name	KM² Land Area	Total Area	Absolute Difference
0100.00	114.94	95.41	19.53
0001.00	33.01	26.95	6.06
0002.00	11.58	8.48	3.10
0010.00	3.41	1.56	1.85
0103.00	2.62	0.80	1.82
0008.00	7.84	6.28	1.56
0009.00	8.52	7.06	1.46
0013.00	62.01	60.62	1.39
0007.00	2.70	1.58	1.12
0006.00	3.30	2.18	1.12
0102.00	2.20	1.34	0.86
0203.00	9.27	8.58	0.69
0004.00	2.62	1.95	0.67
0207.00	29.15	28.55	0.60
0003.00	2.61	2.13	0.48
0011.00	5.13	4.73	0.40
0200.00	6.05	5.88	0.17
0105.00	2.45	2.38	0.07
0005.00	3.14	3.10	0.04

Table 8. Land Area and Total Area for CTs (Top 45 Discrepancies - Percent Difference), 1991 Census

CMA/CA Name	CT Name	Land Area (KM^2)	Total Area (KM^2)	Percent Difference
St. John's	0300.00	100.84	9.57	90.51
Montréal	0057.00	0.78	0.14	82.05
Hamilton*	0085.02	7.88	1.71	78.30
Montréal	0153.00	0.26	0.07	73.08
Matsqui	0103.00	2.62	0.80	69.47
Toronto	0006.00	0.03	0.01	66.67
Montréal	0385.00	0.08	0.03	62.50
Winnipeg*	0510.02	18.42	7.21	60.86
Toronto	0412.06	2.60	1.07	58.85
Toronto	0516.10	9.50	3.93	58.63
Montréal	0235.00	0.26	0.11	57.69
Matsqui	0010.00	3.41	1.56	54.25
Montréal	0328.00	0.98	0.47	52.04
Chicoutimi-Jonquière	0002.01	5.51	2.68	51.36
Montréal	0080.00	0.16	0.08	50.00
Red Deer	0015.00	3.38	1.71	49.41
Guelph	0006.00	1.14	0.61	46.49
Vancouver	0161.02	3.39	1.83	46.02
Toronto	0316.04	0.81	0.44	45.68
Montréal	0417.02	0.67	0.37	44.78
Toronto	0019.00	0.44	0.25	43.18
Matsqui	0007.00	2.70	1.58	41.48
Edmonton	0142.04	405.35	243.86	39.84
Ottawa-Hull	0122.01	1.41	0.85	39.72
Matsqui	0102.00	2.20	1.34	39.09
Toronto	0050.01	2.25	1.40	37.78
Montréal	0857.01	2.35	1.50	36.17
Ottawa-Hull	0602.02	1.91	1.22	36.13
Vancouver*	0115.00	7.59	4.86	35.97
Toronto	0521.01	0.31	0.20	35.48
Matsqui	0006.00	3.30	2.18	33.94
Kamloops	0015.00	2.10	1.39	33.81
Montréal	0064.00	0.36	0.24	33.33
Saint John	0027.02	29.42	19.66	33.17
Toronto	0376.09	1.18	0.79	33.05
Montréal	0584.00	3.10	2.10	32.26
Toronto	0312.05	0.50	0.34	32.00
Toronto	0525.01	0.75	0.51	32.00
Sherbrooke	0010.00	2.59	1.77	31.66
Montréal	0094.02	2.86	1.97	31.12

Table 8. Land Area and Total Area for CTs (Top 45 Discrepancies - Percent Difference), 1991 Census (Cont'd)

CMA/CA Name	CT Name	Land Area (KM^2)	Total Area (KM^2)	Percent Difference
Toronto	0805.02	1.47	1.02	30.61
Toronto	0123.00	0.10	0.07	30.00
Edmonton	0104.03	189.28	132.54	29.98
Montréal	0104.00	0.54	0.38	29.63
Hamilton*	0084.05	0.88	0.62	29.55

***** Hamilton and Vancouver have boundary errors on the CT/DBF, and Hamilton has EA/CT linkage errors on GADB (see Statistics Canada, 1996, pp. 67-68, 72). See footnote in Table 1 regarding Winnipeg. In Hamilton, the land areas for CTs 0085.02 and 0085.03 were coded inversely in GADB; the land area for CT 0085.02 should be 1.70 km^2 (see Caldwell, Lefebvre and Storey, 1994).

NOTE: The shaded rows indicate that the same CTs are also denoted in Table 6.

Table 9. Aggregation of Component CSDs and CTs to CMA/CA Level, 1991 Census

CMA/CA Name	LAND AREA (KM^2) Aggregation of CSDs	Aggregation of CTs	Absolute Difference
Brantford	324.47	324.48	-0.01
Calgary	5,085.84	5,121.88	-36.04
Chicoutimi-Jonquière	1,723.31	1,722.86	0.45
Edmonton	9,532.48	9,532.51	-0.03
Guelph	375.41	374.54	0.87
Halifax	2,503.10	2,507.94	-4.84
Hamilton	1,358.50	1,359.44	-0.94
Kamloops	427.68	444.38	-16.70
Kelowna	3,006.66	2,957.61	49.05
Kingston	1,628.74	1,637.88	-9.14
Kitchener	823.64	827.57	-3.93
Lethbridge	119.90	121.53	-1.63
London	2,105.07	2,087.01	18.06
Matsqui	609.91	609.91	0.00
Moncton	1,719.23	1,692.20	27.03
Montréal	3,508.89	3,510.43	-1.54
North Bay	863.89	863.80	0.09
Oshawa	894.19	894.67	-0.48
Ottawa-Hull	5,138.34	5,142.02	-3.68
Peterborough	1,164.05	1,168.85	-4.80
Prince George	315.72	312.72	3.00
Québec	3,150.27	3,142.95	7.32
Red Deer	51.74	51.59	0.15
Regina	3,421.58	3,421.51	0.07
Saint John	2,904.80	2,908.04	-3.24
Sarnia-Clearwater	498.58	498.58	0.00
Saskatoon	4,749.35	4,749.17	0.18
Sault Ste. Marie	713.53	711.11	2.42
Sherbrooke	915.75	921.03	-5.28
St. Catharines-Niagara	1,399.80	1,398.09	1.71
St. John’s	1,129.99	1,126.27	3.72
Sudbury	2,612.11	2,611.20	0.91
Thunder Bay	2,202.55	2,206.73	-4.18
Toronto	5,583.51	5,567.34	16.17
Trois-Rivières	871.91	871.88	0.03
Vancouver	2,786.26	2,779.33	6.93
Victoria	633.44	633.44	0.00
Windsor	861.66	861.27	0.39
Winnipeg	3,294.82	3,296.03	-1.21

Table 10. Range Groupings of Absolute Differences of CSD and CT Land Area Aggregations to CMA/CA Level, 1991 Census

Ranges (KM^2)	Frequency	Percent
0 - < 1	16	41
1 - < 5	13	33
5 - <10	4	10
10 +	6	15
Total	**39**	**99**

- the file was created to support thematic mapping only. It is a "one shot deal" only, and is not meant to support other infrastructure tasks such as calculating land area (or even generating weighted and unweighted representative points, calculating spatial contiguity, producing collection and reference maps, etc.).
- the 1:2,000,000 shoreline from NAIS is not detailed enough to support the calculation of land area.
- the SNF shoreline can be quite angular and generalized.
- the file will be undergoing generalization processes, such as "moving" the NAIS shoreline when the EA representative points (which come from the EA/DBF) fall in water bodies, line smoothing, eliminating some lakes and islands if the file gets too large, etc.
- the reconciliation of the EA limits with a very small scale NAIS shoreline is very tricky, and will probably incur some incorrect judgements. Many decisions are rather straightforward in terms of determining the exact location of EA limits on or near water bodies. Other situations are not, and sometimes decisions will be made "on the fly" without consulting the original basemaps. Therefore, if incorrect judgements are made about adjusting EA boundaries, then the land areas for EAs and all higher level geographic areas will be wrong.
- the introduction of new arcs when DBF and NAIS do not match.
- ship EAs will become land-based EAs, resulting in portions of shorelines that are quite different from reality.
- EA boundaries representing apartment buildings and collective EAs are only very rough approximations of their areal extent (and location).
- the exaggeration of EA land areas if there is more than one EA in a single building. This exaggeration occurs because "stacked" EAs (i.e. one EA being a group of floors) are digitized side-by-side. Only the base of a building has land area – not, for example, an EA on floors 7-12 and another EA on floors 13-18.
- use of the file would introduce two methodological/longitudinal breaks in terms of land area calculation – one break in 1996 and another in 2001 (see below).

In essence, the EA/DCF may introduce land area errors more profound or serious than the digital planimeter measurements. As well, we believe that users of our land area data would be quite dissatisfied with two longitudinal breaks.

The plans for the 2001 Census involve creating a national digital database, with coverages emanating from the Street Network Files (SNFs) and the National Topographic Data Base (NTDB). It is anticipated that this database will support all the necessary infrastructure tasks, including land area calculation. All previous land areas measured by the digital planimeter, therefore, would be superseded by land areas calculated using the new digital file. In essence, that would represent a methodological break – and users would not be able to compare 2001 land area data with those published for any previous census. In fact, the U.S. Bureau of the Census followed this procedure for land area calculations using their new 1990 TIGER File (U.S. Bureau of the Census, 1992).

5. SUMMARY AND RECOMMENDATIONS

We have shown that much of our land area data are subject to a number of discrepancies, including propagation errors over time and logical inconsistencies. We have also demonstrated that the digital planimeter test in the Winnipeg CMA helps us understand some of the sources of error.

There are a number of reasons for keeping the *status quo* for the 1996 Census, and reserving methodological changes for the 2001 Census. Despite measurement and cumulative historic errors, overall the land area data are acceptable, and appear to fall within a tolerable measurement variation. For example, nearly 6% of the CSDs and 2% of the CTs have land areas greater than total area by over 1 km^2 (Table 2). It should be noted, however, that the land areas for designated places and urban areas were not examined.

It is recommended, however, that the significant discrepancies be corrected for the 1996 Census. The serious logical inconsistencies should be assessed individually; the threshold of "seriousness" should be based on the availability of time and resources. Quite pragmatically, it is simply too costly to correct all the known discrepancies, especially since the methodology will be different for the 2001 Census. In fact, the land areas could also be incorrect even when they are smaller than total area, but these errors would be very difficult to spot at best.

The 1996 EA/DCF was considered insufficient to calculate land area, mainly because the file was created to support small-scale thematic mapping only. Furthermore, since the file does not represent the 2001 infrastructure model, it makes more sense that only one clean break take place rather than two.

It is thus recommended that automation replace the historic manual method for the 2001 Census, whereby land areas are calculated using the new national digital database. There are a number of advantages of automation over manual methods, including:

- incongruities between land area vs. total area would be eliminated.
- aggregations to higher level geographic areas would be consistent (for example, CSD and CT aggregations to the CMA/CA level).
- error propagation over time would be eliminated by recalculating the entire area when boundaries change, as well as the area gained or lost associated with the change. The historic approach of only measuring the change itself, and then adding to or subtracting from the previously calculated values has allowed errors to accumulate over time.
- incorrect measurements using manual methods would be avoided.
- transcription errors on coding forms would be eliminated.
- more accurate derivatives of land area would be possible, such as deriving population density, edge cities or population ecumenes.
- automation would save time, resources and costs.

The accuracy of automating land area calculations, of course, will depend upon the accuracy of the digital boundaries, as well as the number (completeness) and spatial representation of double-line water bodies in the file. Concerning the latter issue, it is recommended that the double-line hydrographic features in the SNFs either be enhanced or replaced by the NTDB renditions, as they are currently very angular and generalized. It is also suggested that the land area calculations be based on an equal-area map projection rather than a conformal projection – since it is crucial that area, not shape, be preserved for more accurate results. Finally, a methodology needs to be developed for handling the "idiosyncrasies" of the digital world. For example, the land areas of geographic units can be altered even though conceptually there should be no change. This situation can arise when a geographic unit is split, but the new node(s) does not fall exactly on the old boundary. We suggest that this issue be addressed by the Geography Division's 2001 Data Model project.

ACKNOWLEDGEMENTS

Many thanks are extended to Henry Puderer, Stan Currie and Robert Parenteau for providing useful comments on this report, Rob Storey for supplying the land area data, Robert Perrault for aggregating the component CSDs and CTs to the CMA/CA level, and Alex Hays for coordinating the digital planimeter test.

REFERENCES

Caldwell, C., Lefebvre, S. and Storey, R. (1994). Automating Land Area Calculations, unpublished memo and attachments (Ottawa: Geography Division, Statistics Canada).

Currie, S. (1996). Personal communication, June 1996.

Hays, A. (1995). Winnipeg Census Tract Land Area Test, unpublished document (Ottawa: Geography Division, Statistics Canada).

Placom Company (n.d.). *Instruction Manual, Digital Planimeter KP-90*.

Storey, R. (1994). Census Tract Land Area Test Report, unpublished draft report (Ottawa: Geography Division, Statistics Canada).

Storey, R. and Tupper, A. (1995). Preliminary Summary Statistics for Land Area Analysis: Census Subdivisions and Census Tracts, unpublished draft report (Ottawa: Geography Division, Statistics Canada).

Statistics Canada (1994). High Level Delineation, unpublished report (Ottawa: Geography Division, Statistics Canada).

Statistics Canada (1996). Compendium of Updates: 1991 Geographic Reference, Data and Digital Products, unpublished report (Ottawa: Geography Division, Statistics Canada).

U.S. Bureau of the Census (1992). Area Measurement Information in the 1990 Census Data Products, unpublished report, Geography Division (Washington, DC: U.S. Bureau of the Census).

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)