Feasibility of Using Automated Generalisation Techniques to Process the National Road Network

Rupert Brooks, Rhian Evans
National Atlas Information Network
October, 1999

Table of Contents

Introduction

In September 1999, the National Atlas of Canada investigated the possible use of automated generalisation techniques for processing the National Road Network (NRN). This report documents the results achieved. These results are experimental in nature, but indicate that the semi-automated production of generalised products from the NRN is feasible.

A Brief Overview Of Theory

Ultimately, the process of database generalisation may be viewed as simply a sorting process. Elements in a set are ranked in some order of importance. At a given level of detail, only a certain percentage of the elements will be used. For this view to operate, we should make the assumption that this is a well-ordering. That is, that if an element disappears at level of detail **a** it will not reappear at level of detail **b**, if **b < a**.

This ordering method could be applied globally, or at varying degrees of locality. For instance, all roads in a dataset could be ranked together, or all roads in each defined area (urban, rural, etc) could be ranked and then the rankings related based on some property of the areas.

We believe that best results are obtained by carefully choosing which objects to order. Simply ordering the arcs in a network by length, or by class will not produce a good generalised result. For this reason, algorithms are first applied to allow the software to perceive meaningful units within the dataset. Then these meaningful units are ranked.

In the case of the road network, we have used the technique of 'perceptual strokes' [Thomson and Richardson, 1999] as the core algorithm used by the software. The results of this technique are then refined by the application of shortest path spanning trees (SPST) [Richardson and Thomson, 1996] and direct intervention by the user.

The concept of a perceptual stroke is based on the theories of gestalt psychology. It is theorised that when perceiving an image, the human visual cortex applies a number of simple rules to break the visual field into a number of perceptual objects. Strokes are based on the principle of good continuation, which states “that elements that appear to follow in the same direction (as in a straight line or simple curve) tend to be grouped together.” [Coren and Ward, 1989]

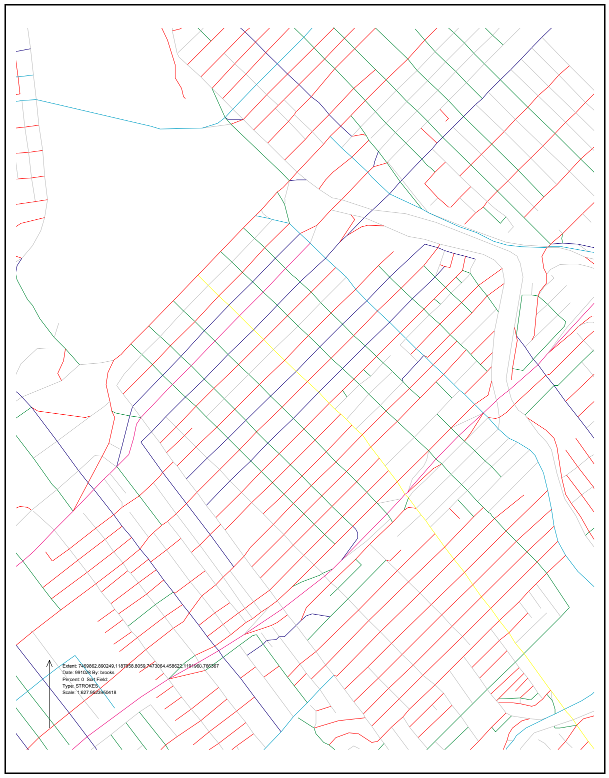


Figure 1: Perceptual strokes in the downtown Ottawa road network. Colorized so that no two intersecting strokes have the same color

Given a set of nodes in an arc-node graph which are considered important, a shortest path spanning tree (SPST) is the subset of the graph which contains all the points of importance, while minimising the total length. There are a number of variations on this concept, which are collectively known as Steiner optimisation problems [Balakrishnan, 1997].

Requirements For Data Preparation

The generalisation of roads operates on the creation of perceptual strokes, which are an approximation of what the human eye perceives when looking at an image of the data. Success of this process hinges on the machine's ability to build sensible strokes from the data. Clearly, the more accurate the data, and the more information which is available the better the result at this stage.

Currently, the system builds strokes by taking into account the road class, the angle of intersection and the name of the road when deciding how to build strokes across intersections. More information could be added as needed. It is of course, ideal that if class or name information is present, that it be accurate and well documented. However, we have found the algorithm to be fairly robust, even using erroneous data.

When this experiment was conducted, the software was limited in the size of the dataset that it could process to something with less than 24000 nodes and 99999 arcs. This limit has recently been removed but it is unlikely that processing the entire dataset at once would be practical. However, it is in principle possible to partition the datasets, and produce a reliable generalisation across partitions [Thomson, 1999]. This would require some additional software development. Furthermore, the character of the dataset is such that generalisation to a level of detail below about 0.5 – 1% is not appropriate. To achieve such high levels of generalisation, a two-stage process would be recommended.

In the short term, it is probably realistic to process map sheet products individually to scales down to 1 – 2 million. Estimates of the resources required to do this are detailed in the next section.

Resources Required For Processing

The typical processing of a dataset is as follows:

1. Back up the original
2. If the class structure of the dataset has not been previously processed, then the rule base should be updated
3. The class information is applied
4. The turntable is updated
5. Strokes are built
6. One or more ordering methods are applied to the strokes
7. A generalisation level is selected, which amounts to a percentage of strokes to keep.
8. The result is inspected, and post processing adjustments are applied
9. The output is used for whatever purpose.

Steps 5 and 6 require significant amounts of computer time, and steps 7 and 8 require significant amounts of skilled operator time.

The processing is moderately computationally intensive on medium size datasets, and is expected to grow roughly linearly as the size of the dataset increases. Currently processing a dataset with ~28000 arcs, the Ottawa area, takes several hours on an UltraSparc. This processing can be expected to get more efficient as the software is advanced from the experimental to the production stage. To be safe, however, we will estimate 6 hours of dedicated computer processing time per map sheet.

After processing some interactive activity is required to take the generalised result and make it acceptable for publication. This interactive activity would probably be on the order of hours of person-time per map sheet. For instance, the examples prepared show the improvement after much less than a person-day of interactive work. Once again, software and procedural refinement can be expected to improve the efficiency of this process. However, preparation of a product for publication would require more editing than our example datasets, and we will therefore estimate 3 person-days. As the computer time required is much less than the interactive time estimated, we expect that the datasets could be processed in the off-hours, and be ready in advance for the interactive processing.

It is our intention that the interactive editing will be stored with the data, so that if the database is updated, then only those areas affected by the updates will need to be reviewed interactively. Therefore over time, the use of such a system will result in savings of time and effort

The estimated size of the road network [Wall, 1999] is minimum 800 000 arcs. We suspect that this estimate is too small, (the Ottawa - Toronto area alone has 200 000 arcs) so we assume for these calculations that the NRN will contain 1 500 000 arcs. If the average size of a practical dataset is 20 000 arcs (it is probably possible to process 100 000 arcs or more at once but the dataset must be broken into meaningful sheets, rather than arbitrarily) this would give us 75 partitions to deal with.

Thus, to generalise the entire NRN after breaking it into approximately 75 map sheets would require 75 x 6 = 450 hours of processing time on an UltraSparc-class machine, along with 225 person days of interactive editing time, approximately one person-year. It would be wise to add an administrative overhead to this estimate to take into account the cost of organizing and managing such a project. There may also be a training cost, if new staff are assigned to the project. Finally, this estimate does not take into account any movement of vertices, or any fitting to other map layers that would likely be necessary to produce a publishable map product.

We were not able to accurately estimate the time required to do manual generalisation of this dataset. However, we expect that it would be of a similar order of magnitude. The greatest savings due to using an automated process will occur in 2nd generation and later products, when the regeneralisation of edited data, or the generalisation to a different scale will take much less time and effort.

Further Software Development

Currently the software used for these experiments is in the prototype stage. Before production could begin, the software would have to be made more efficient and robust. We estimate the required effort at two person-months. As the Atlas staff are currently fully loaded, this might not be begun for several months.

In addition to software refinement, it would be worth refining the rule-base used by the software. Currently, this set of rules was created by guesswork and estimation with only an incomplete knowledge of the classes present in the road network and their meanings. Ideally, using a complete specification for the NRN and with the assistance of someone experienced in how it should be visualised, a better set of rules could be developed. This task would take approximately a person-week.

Example Generalisations

We have prepared samples from the dataset which was provided for experimental purposes as well as from various sections of the NRN beta version provided to us in Spring 1999. In each case, we have compared the results with existing road maps. Improvement can be expected from the results of this work as this work was done with prototype software with known limitations.

Testing was performed on 3 coverages. The Fort St. John - Dawson Creek test area provided by Mr. Wall, the National Capital Region area, and the Western Ontario area. For each test area we have prepared plots of the ungeneralised data, and we have attempted to generalise the data to the approximate level of off the shelf road maps, to see how close the generalisation process will get to a human-generalised map.

After the automatic calculations were done, the feature selection tends to be close but fragmented. At this point, the approximate level of generalisation was chosen, and then the dataset was post processed. In each case, we limited the human intervention so that the result would reflect what could be done with limited time spent on it. The results are in the attached Appendices. Each set of examples shows the raw data, the data with only automated generalisation applied, and the data with the stated amount of post-processing done on it. The raw and final examples are symbolised by road class, as shown in the legend below, while the automated example shows those roads which have been dropped out in grey, and the remainder in blue or black.

Table 1: LEGEND These are the internal classes used by the software, we developed a mapping between the actual classes and these categories, but this may require revision.		
Major Highway	Very important road, part of road structure at provincial or national level	
Highway	Multilane, divided	
Primary road	Road of high importance	
Secondary road	Road of medium importance, collector road	
Tertiary road	Road of lesser – average importance	
Urban	Urban roads have the special property that they can connect to any other type of road.	
Minor road	Seasonal, etc	
Ramps	A road will connect to a ramp as a last resort	
Cart track, trail		
Not a valid road	Abandoned, bicycle path, etc	
Bridges	A bridge could be part of any road system	

Conclusions

The NRN could be generalised using semi-automatic techniques with a reasonable amount of time and resources. The effort of making adjustments to the resulting generalised dataset could be reapplied to the next generation of data, resulting in further savings in future. The estimated resource requirements are summarised in the following table:

Table 2 : Summary of resources required

Task	Resource /Units	Quantity	Total
Software Revision	2 Person months		2 Person Months
Rule-base revison	1 Person Week		1 Person Week
Generalisation computer processing	6 hr per mapsheet	75	450 hours
Generalisation interactive editing	3 person days per mapsheet	75	225 person days
Mgmt overhead	????		Not estimated
Training	????		Not estimated
Other map development	????		Not estimated

Our experience with hydrology generalisation to date suggests that it may be possible to find synergies between the preparation for automated generalisation and the necessary quality control procedures. Not having seen these procedures, we can only speculate at this point.

References

AllMaps. 1996. *British Columbia Provincial Map.* Rand McNally / Allmaps Canada Markham, ON.
AllMaps. 1997. *Prince George, British Columbia City Map.* Rand McNally / Allmaps Canada, Markham, ON.

Balakrishnan, V. K. 1997. *Schaum's Outline of Theory and Problems of Graph Theory.* McGraw-Hill, Toronto.

Capital Publishers. 1999. *Ottawa Tourist Map.* Capital Publishers. See www.tourottawa.org.

Cohen, S. and Ward, L. M. 1989. *Sensation and Perception, 3rd Ed.* Harcourt, Brace and Jovanovich, Orlando, FL.

Mapart. 199?. *Ontario Road Map.* MapArt, Oshawa, Ontario.

Pathfinder Maps. 1998. *National Capital Region.* Map Book published by Pathfinder maps Carp, Ontario.

Richardson, D. E. and Thomson, R. C. 1996. *Integrating Thematic, Geometric and Topologic Information in the Generalization of Road Networks.*

Thomson, R. C. and Richardson, D. E., 1999. *The 'Good Continuation' Principle of Perceptual Organization Applied to the Generalization of Road Networks.* Proceedings ICA 1999, Ottawa.

Thomson, R. C., 1999. Technical Memoranda delivered to GeoAccess Division while acting as advisor to generalisation project.

Wall, D. 1999. Personal Communication.

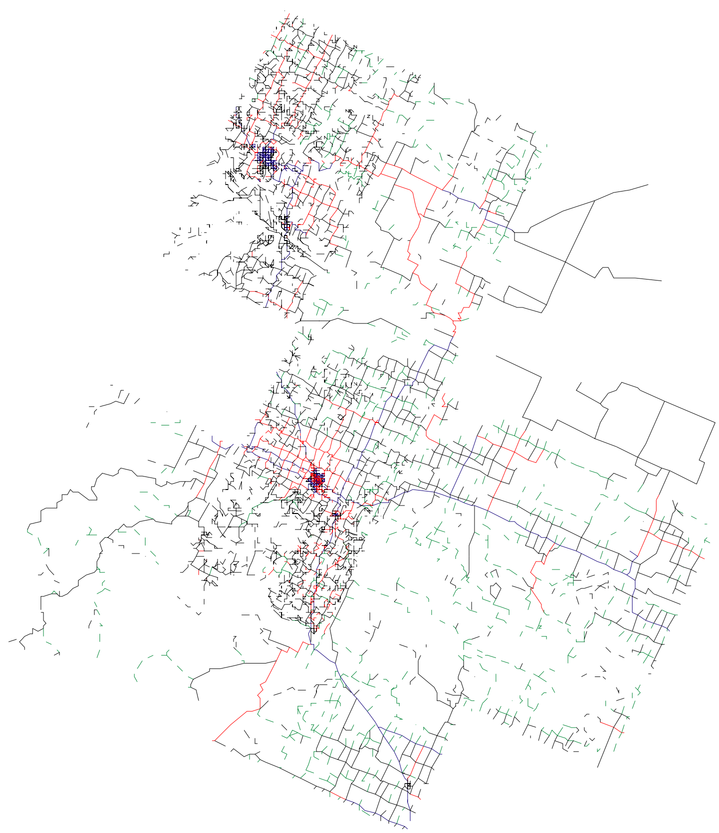
Appendix I: Fort St. John - Dawson Creek Test Dataset

This dataset consists mainly of two small towns with the Alaska Highway running through them.
ARCS: 10323
NODES: 8483
SPECIFIC PROBLEMS
The Peace river and several census boundaries were located in the road cover, classed as 'BO'. Although it is clear to us now that BO represents boundaries, there is a SNF code for BO - Bourg. In the beginning, therefore, we assumed that it was some sort of urban road and left it in the dataset. Naturally the results were confusing

Furthermore the Alaska highway is fragmented into many tiny, twisting roads of various classes. It is our impression from examining road maps that in places the Alaska highway is actually missing from the dataset, and only its parallel access roads are present. In any case, the majority of interactive editing on this dataset was required to force this highway to remain ungeneralised.

Finally there was a notch in the side of the dataset, and this clipped a piece of the Alaska highway out. This resulted in the impression that this was not a major road, but two small roads that went nowhere. Again, interactive editing was required to adjust that.

Fort St. John – Dawson Creek Test Dataset.
Unprocessed



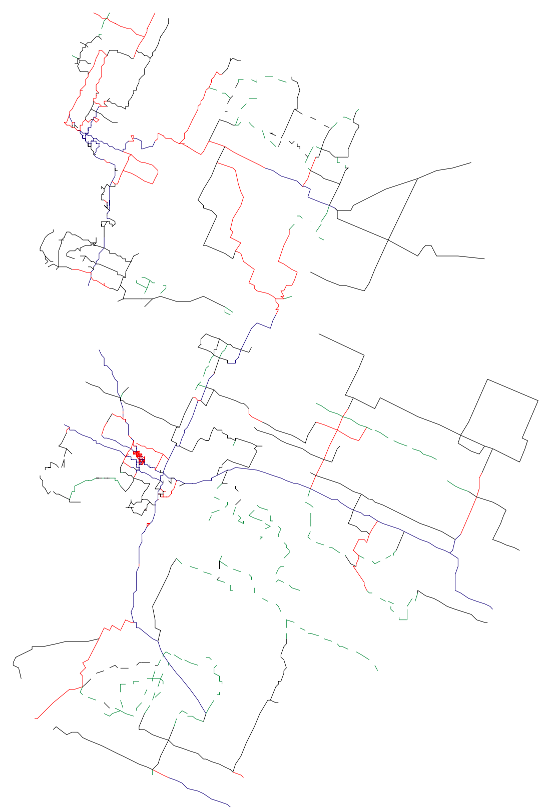
Extent: 4445140.968454,2477795.958863,4598756.716006,2656172.905685
Date: 991008 By: brooks
Percent: 15 Sort Field:
Type: CLASS
Scale: 1:1500000

Fort St. John – Dawson Creek Test Dataset.
Automated Generalization Only

Extent: 4445140.968454,2477795.958863,4598756.716006,2656172.905685
Date: 991008 By: brooks
Percent: 5 Sort Field:
Type: SLIDER
Scale: 1:1500000

Fort St. John – Dawson Creek Test Dataset.
Automated Generalization and 5 hours Post-Processing

Note: 5 hours work produced all 3 results in this series, and includes reprocessing after spurious boundary arcs were detected.

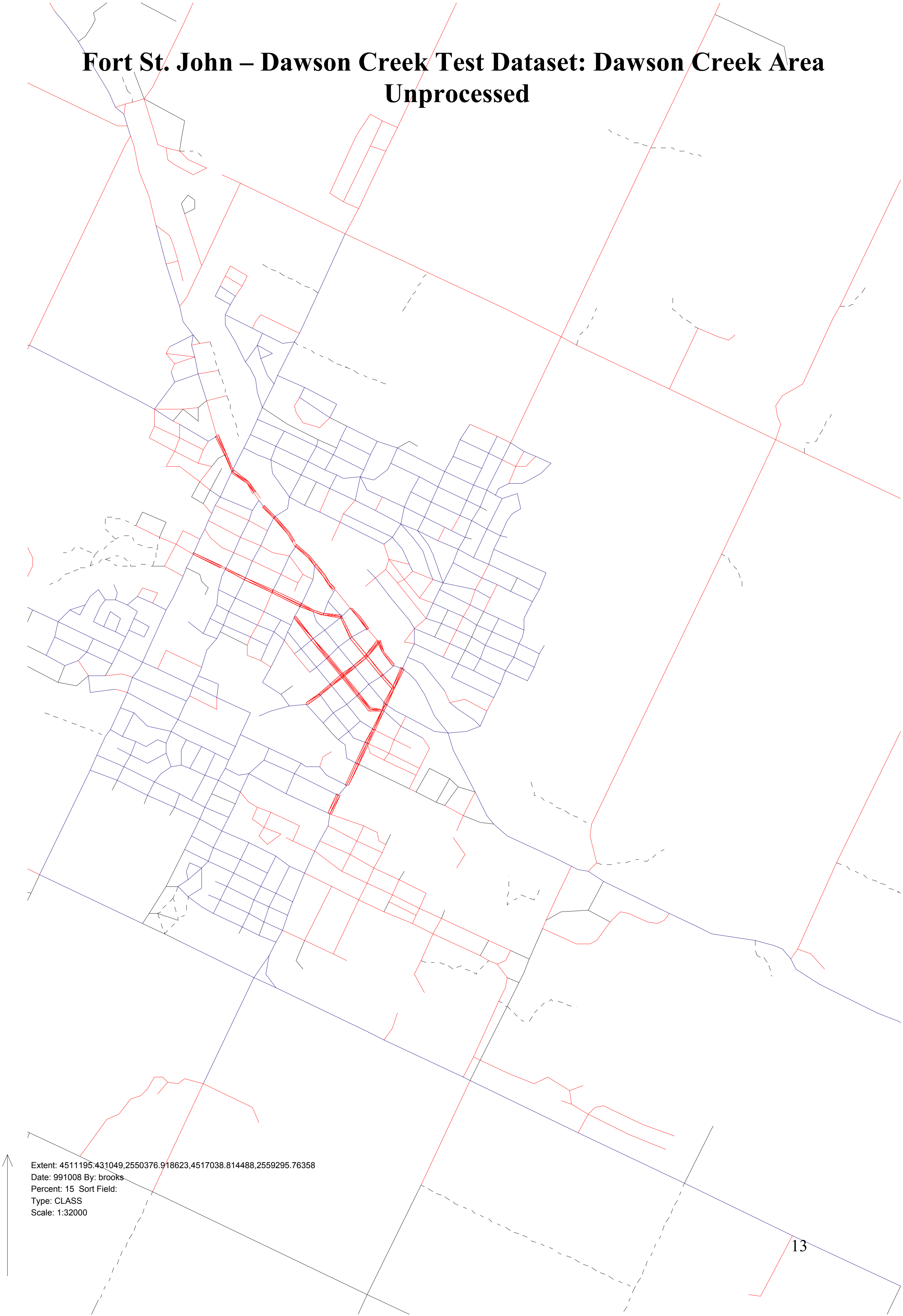


Extent: 4445140.968454,2477795.958863,4598756.716006,2656172.905685
Date: 990924 By: revans
Percent: 5 Sort Field:
Type: FINAL

Fort St. John – Dawson Creek Test Dataset.
Road map of Local Area [AllMaps, 1996]

Please note that this map is in a different projection than the generalised samples

Fort St. John – Dawson Creek Test Dataset: Dawson Creek Area Unprocessed

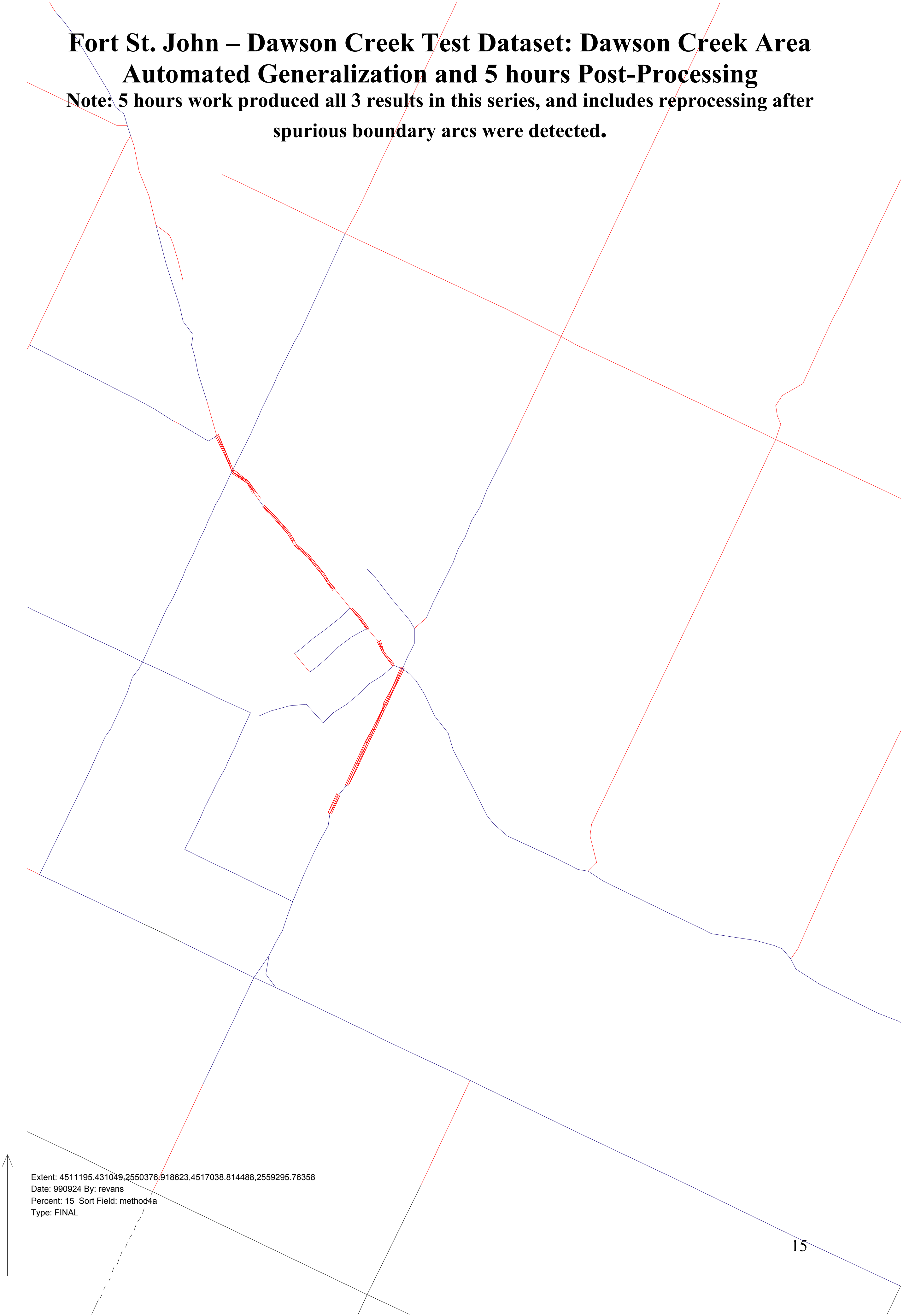


Fort St. John – Dawson Creek Test Dataset: Dawson Creek Area
Automated Generalization Only

Extent: 4511195.431049,2550376.918623,4517038.814488,2559295.76358
Date: 991008 By: brooks
Percent: 15 Sort Field:
Type: SLIDER
Scale: 1:32000

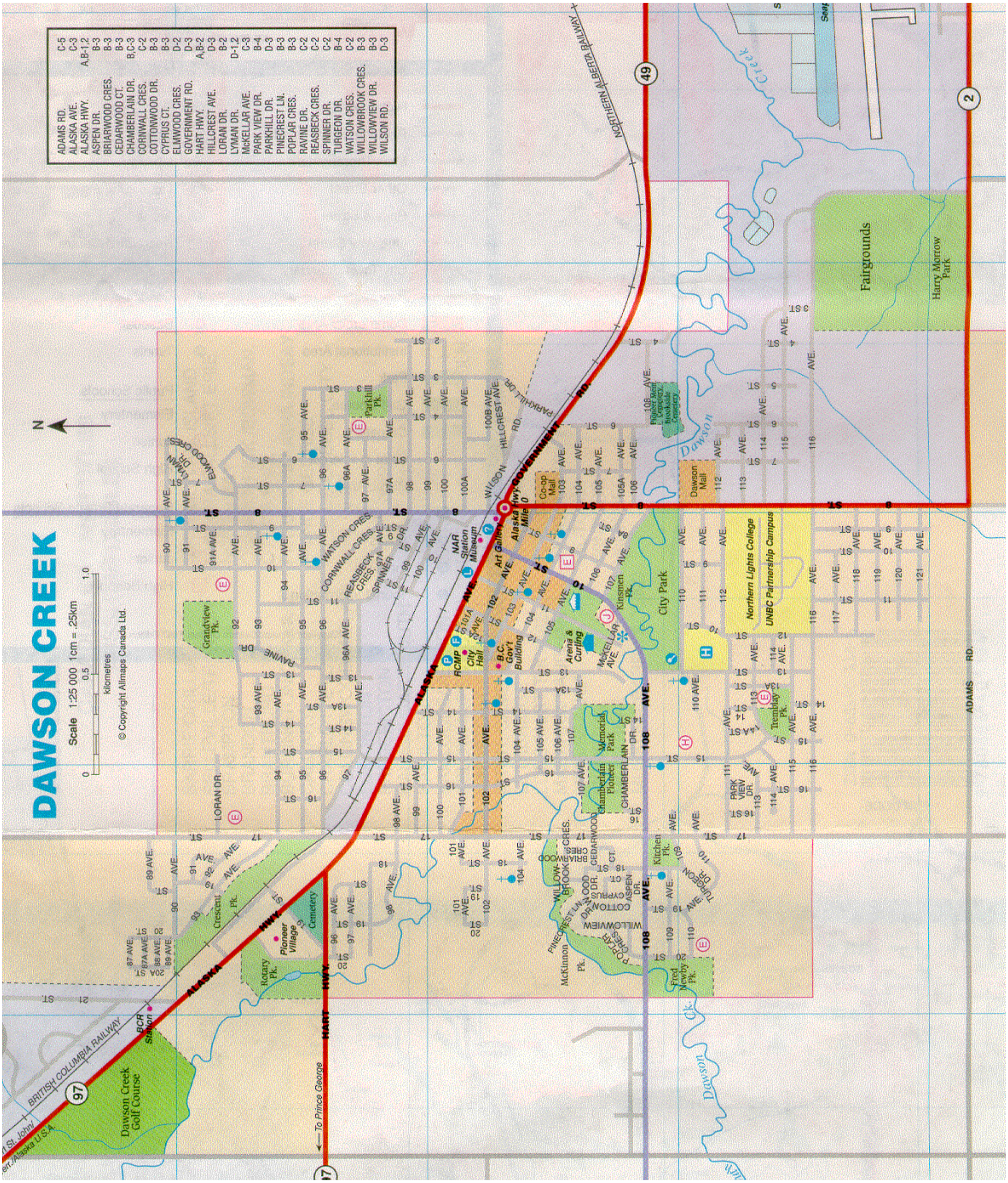
Fort St. John – Dawson Creek Test Dataset: Dawson Creek Area Automated Generalization and 5 hours Post-Processing

Note: 5 hours work produced all 3 results in this series, and includes reprocessing after spurious boundary arcs were detected.

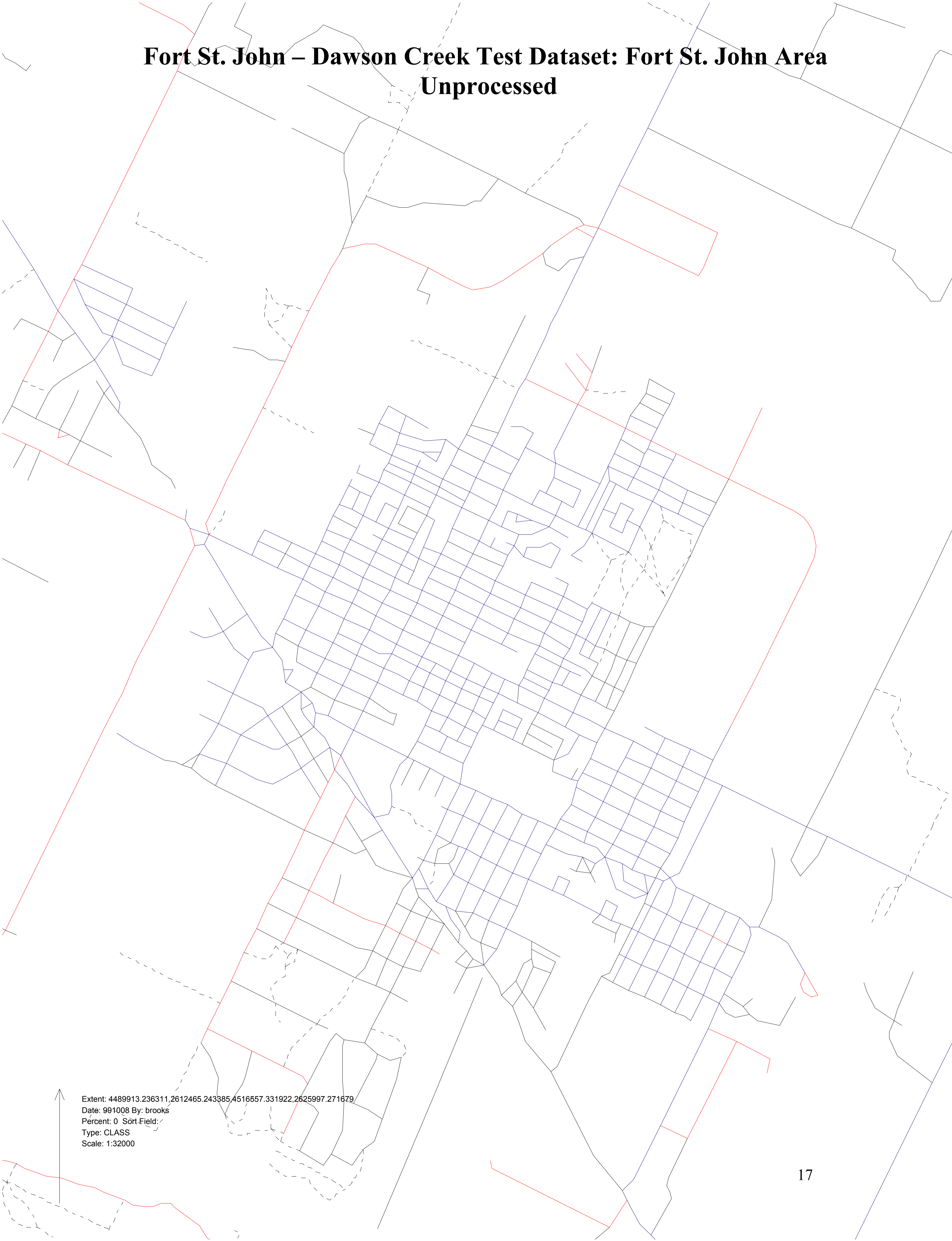


Fort St. John – Dawson Creek Test Dataset: Dawson Creek Area
Road map of Local Area [AllMaps, 1997]

Please note that this map is in a different projection than the generalised samples



Fort St. John – Dawson Creek Test Dataset: Fort St. John Area Unprocessed



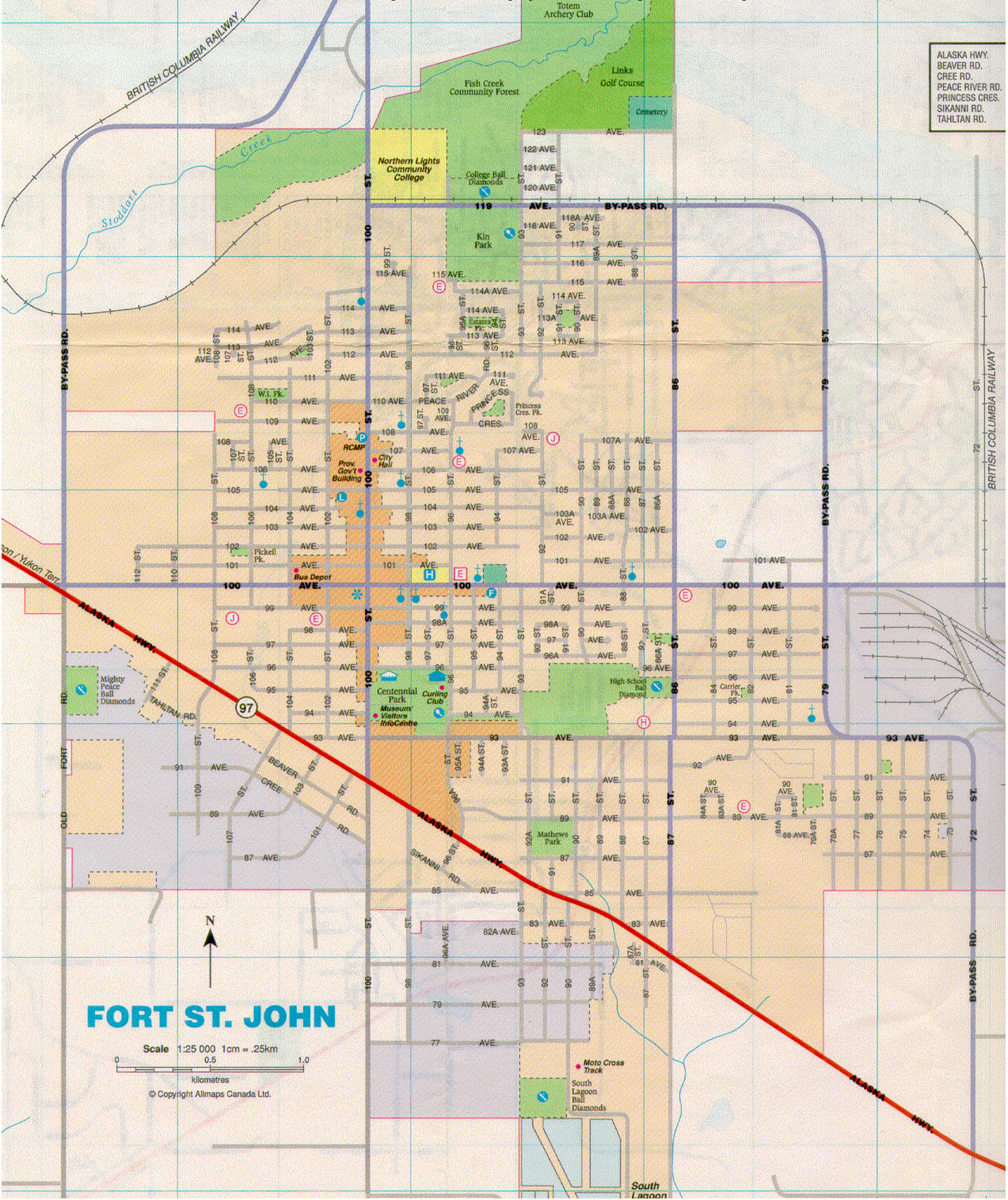
Fort St. John – Dawson Creek Test Dataset: Fort St. John Area
Automated Generalization Only

Fort St. John – Dawson Creek Test Dataset: Fort St. John Area
Automated Generalization and 5 hours Post-Processing

Note: 5 hours work produced all 3 results in this series, and includes reprocessing after spurious boundary arcs were detected.

Fort St. John – Dawson Creek Test Dataset: Fort St. John Area
Road map of Local Area [AllMaps, 1997]

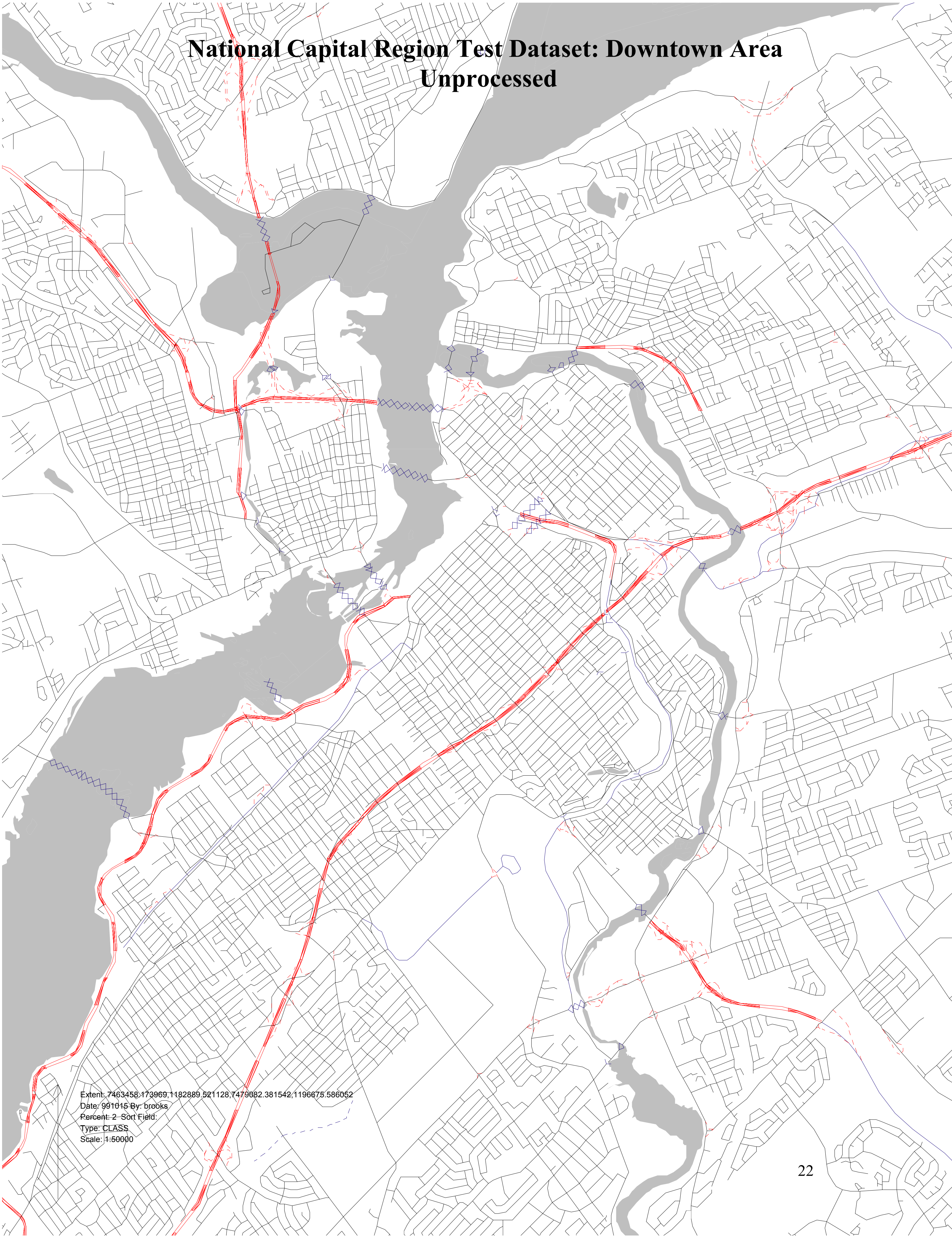
Please note that this map is in a different projection than the generalised samples



Appendix II: National Capital Region Test Dataset

ARCS: 28841
NODES: 20796
SPECIFIC PROBLEMS:
There was extremely limited class information available for this area. The Queensway (rte 417) was identified as a highway, and some bridges, ramps were indicated. The great majority of roads were simply unclassed.

National Capital Region Test Dataset: Downtown Area
Unprocessed



National Capital Region Test Dataset: Downtown Area
Automated Generalization Only

National Capital Region Test Dataset: Downtown Area

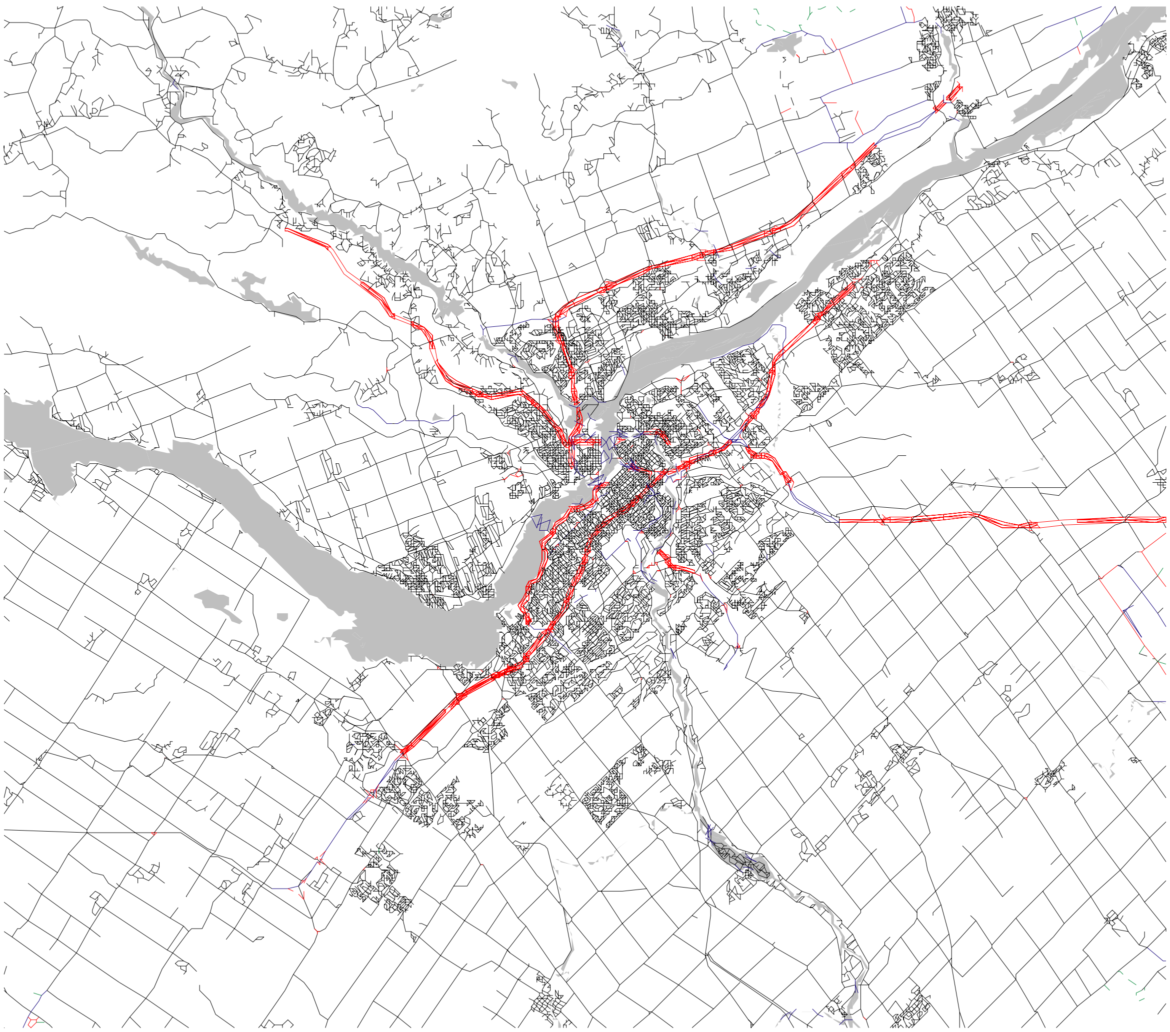
Automated Generalization and 3 hours Post-Processing

National Capital Region Test Dataset: Downtown Area

Road Map of Local Area [Capital Publishers, 1999]

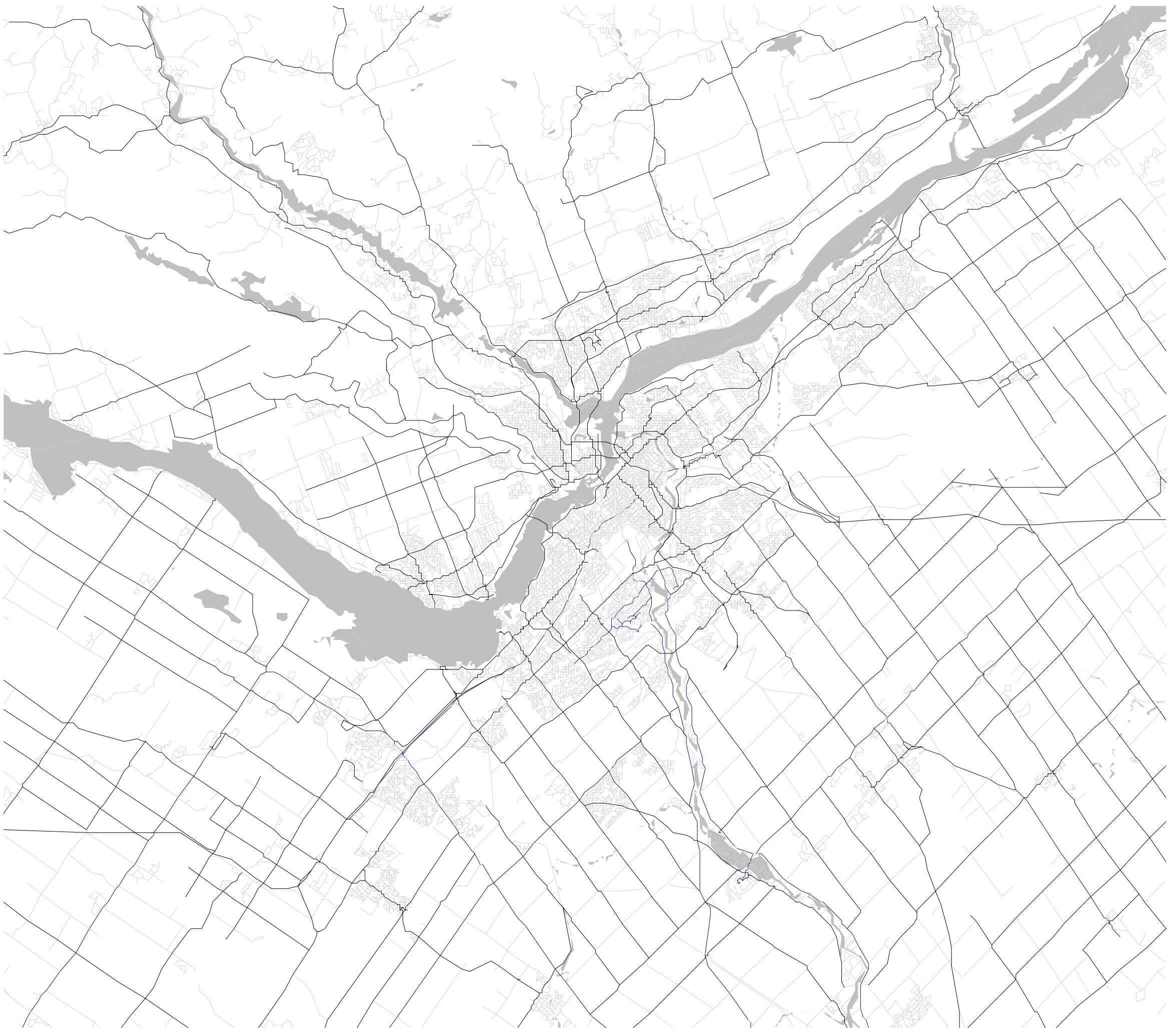
Please note that this map is in a different projection than the generalised samples

National Capital Region Test Dataset: Entire Area
Unprocessed



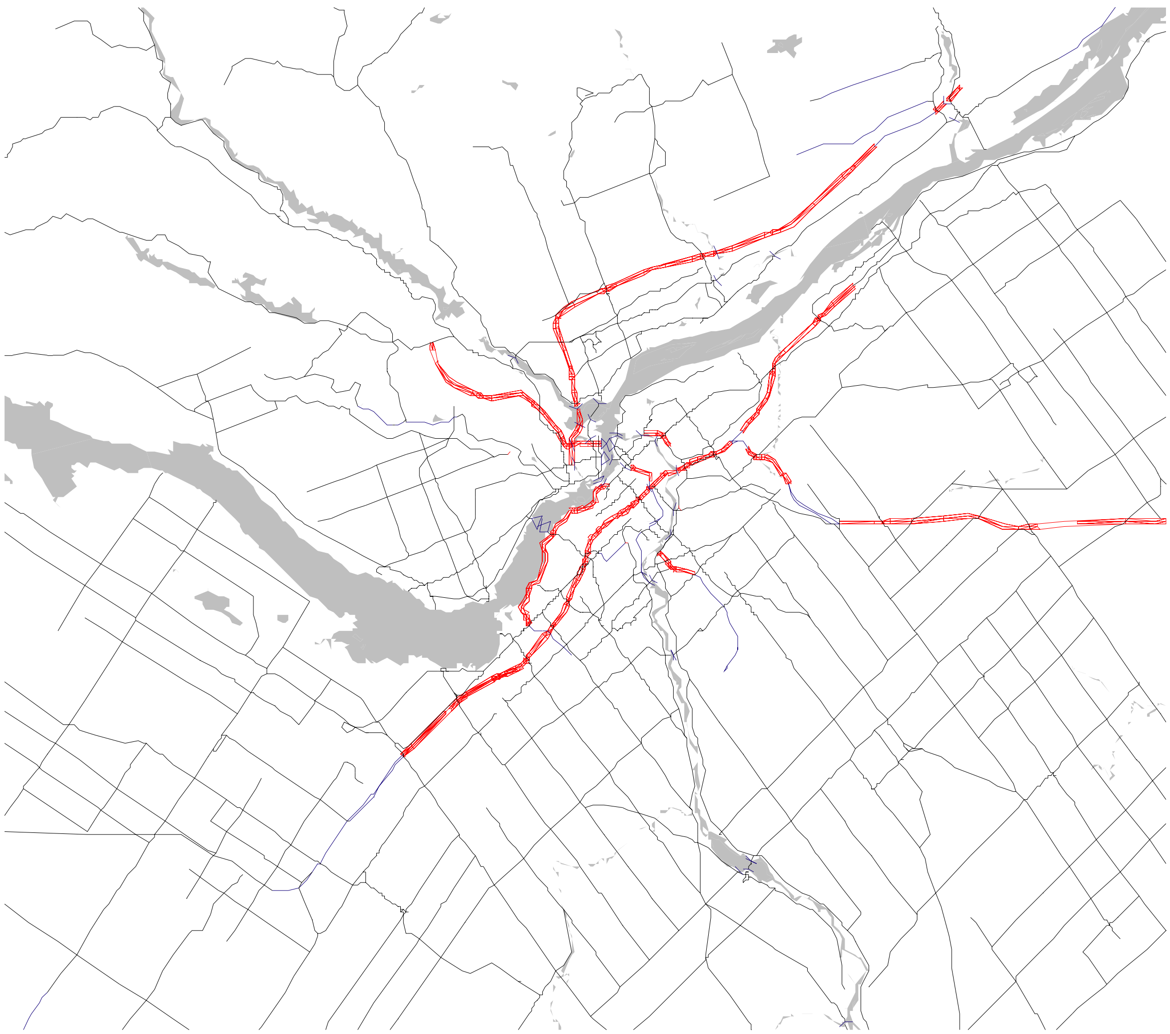
Extent: 7435466.701089,1158687.320461,7502886.997361,1217916.344237
Date: 991015 By: brooks
Percent: 2 Sort Field:
Type: CLASS
Scale: 1:400000

National Capital Region Test Dataset: Entire Area
Automated Generalization Only



Extent: 7435466.701089,1158687.320461,7502886.997361,1217916.344237
Date: 991015 By: brooks
Percent: 2 Sort Field:
Type: SLIDER
Scale: 1:400000

National Capital Region Test Dataset: Entire Area
Automated Generalization and 1.5 hours Post-Processing

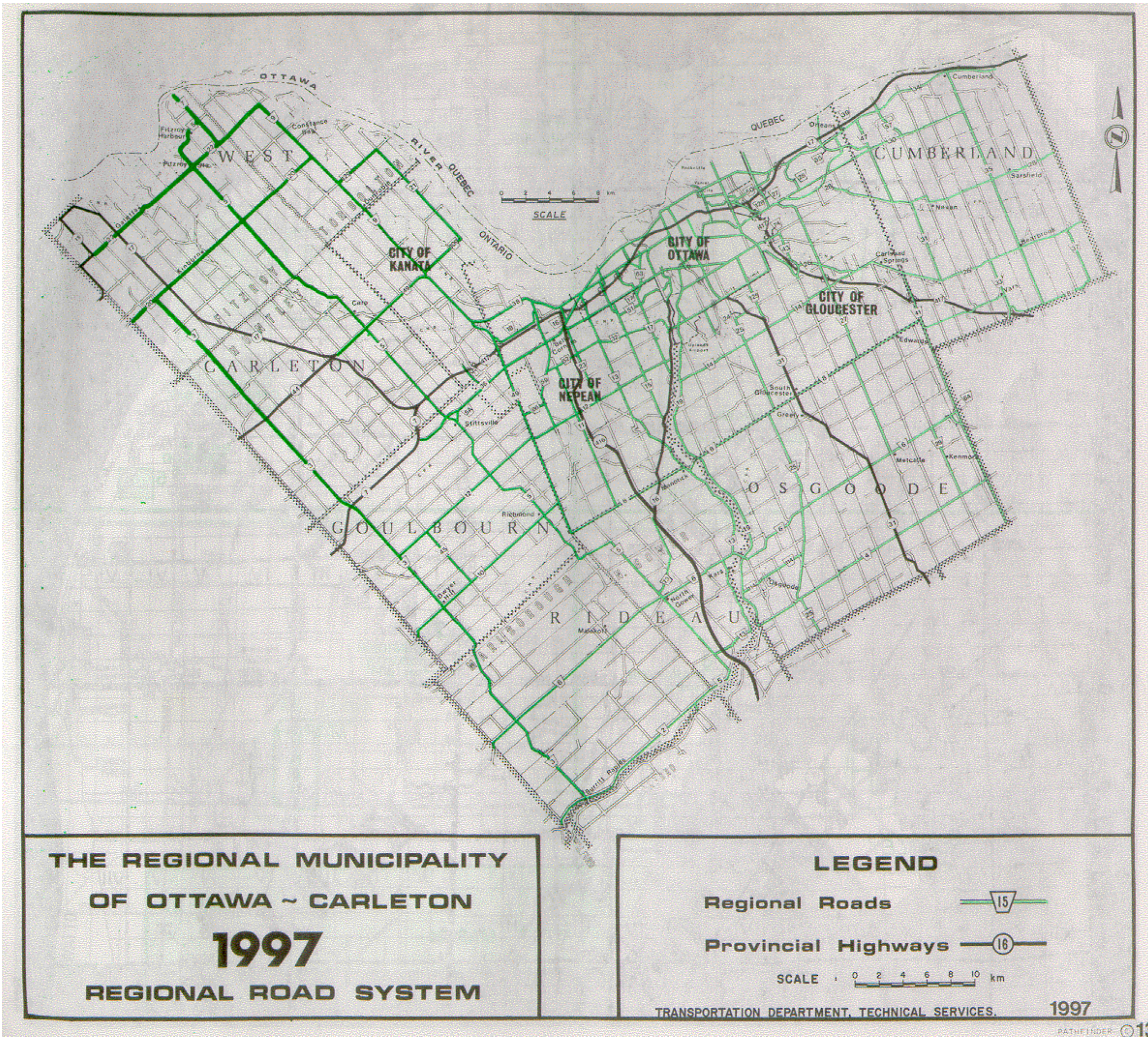


Extent: 7435466.701089,1158687.320461,7502886.973613,1217916.344237
Date: 990923 By: revans
Percent: 2 Sort Field:
Type: FINAL

National Capital Region Test Dataset: Entire Area

Road Map of Local Area [Pathfinder Maps, 1998]

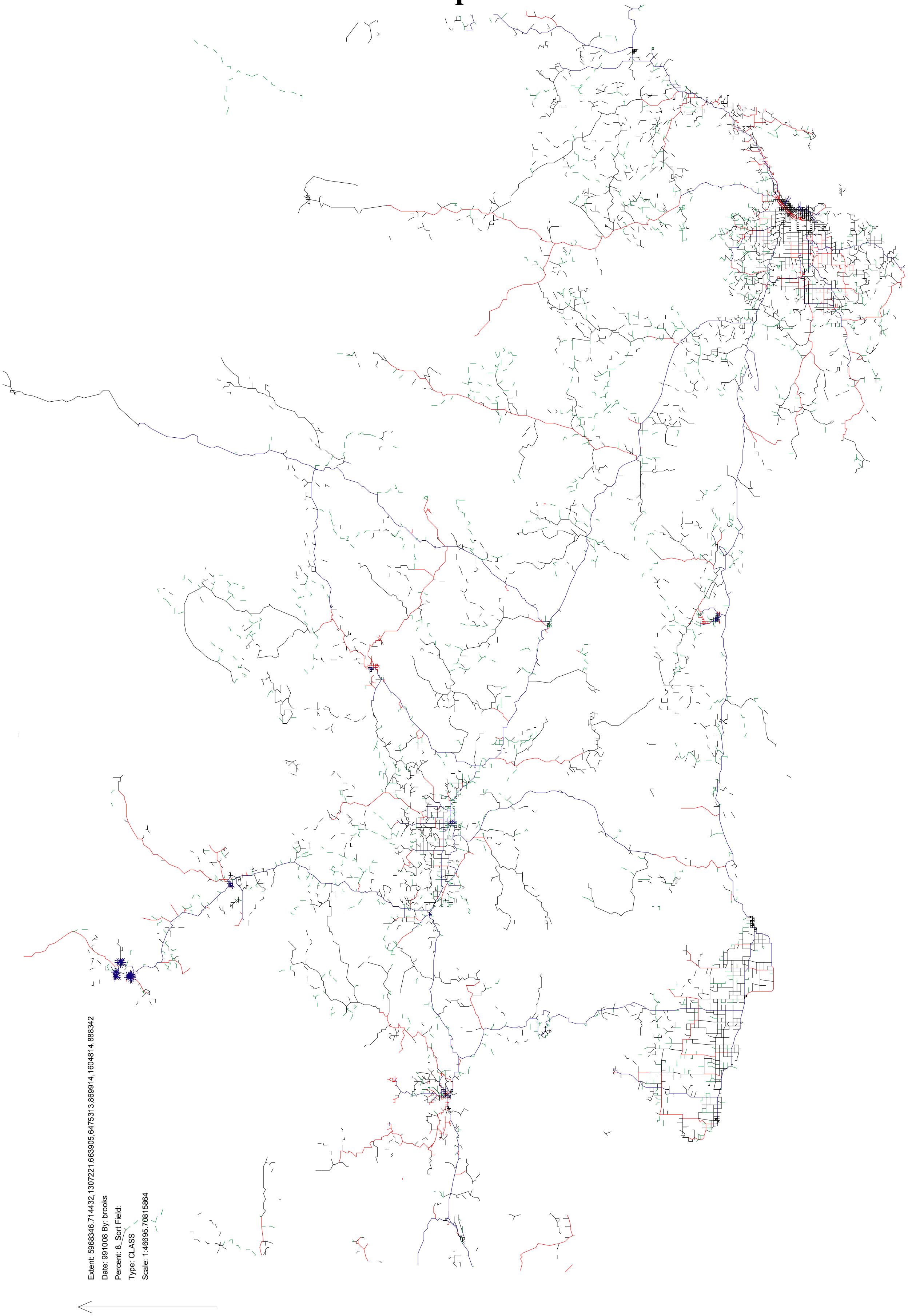
Please note that this map is in a different projection than the generalised samples



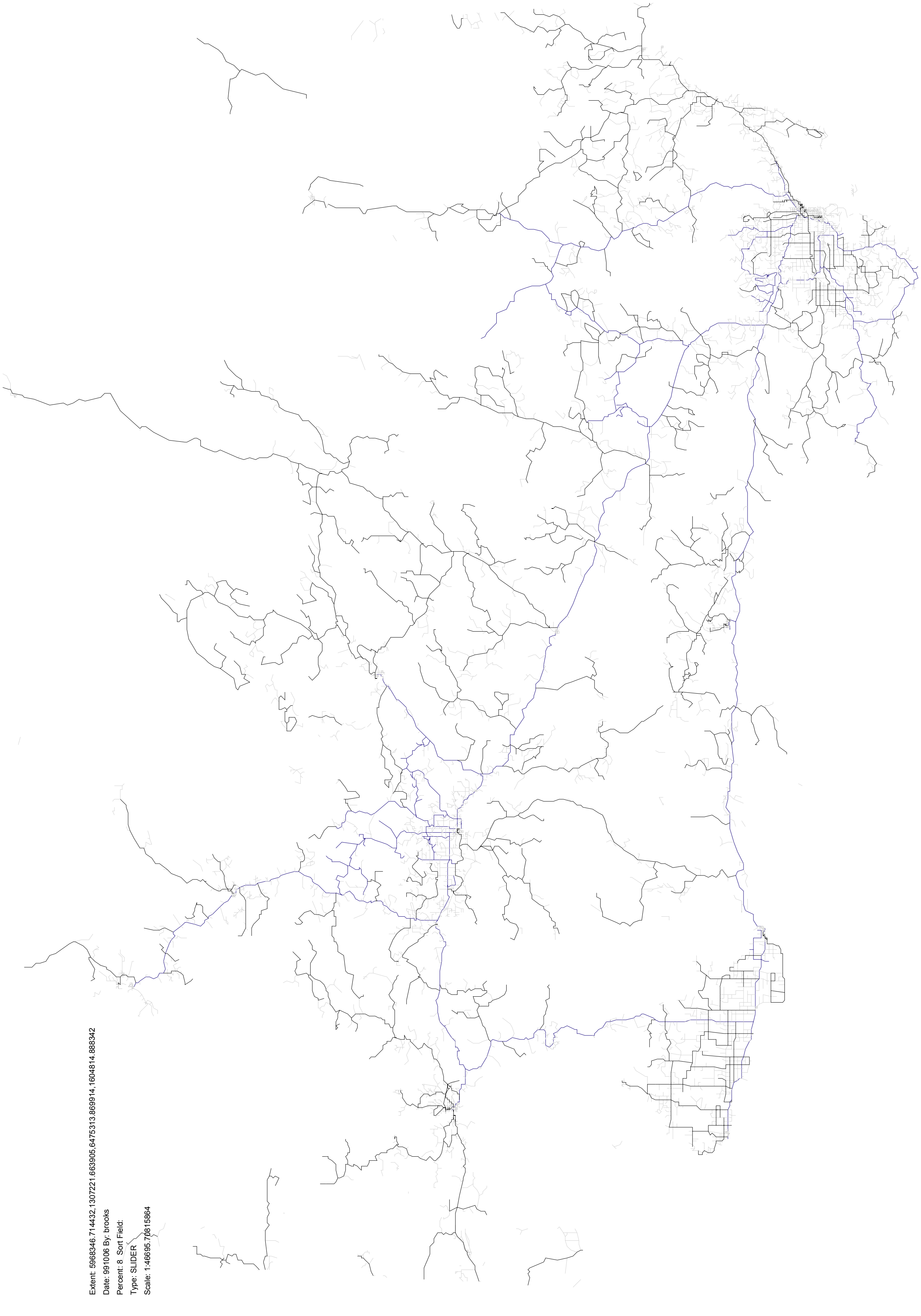
Appendix III: Western Ontario Test Dataset

ARCS: 16830
NODES: 13412
SPECIFIC PROBLEMS: None

Western Ontario Test Dataset
Unprocessed

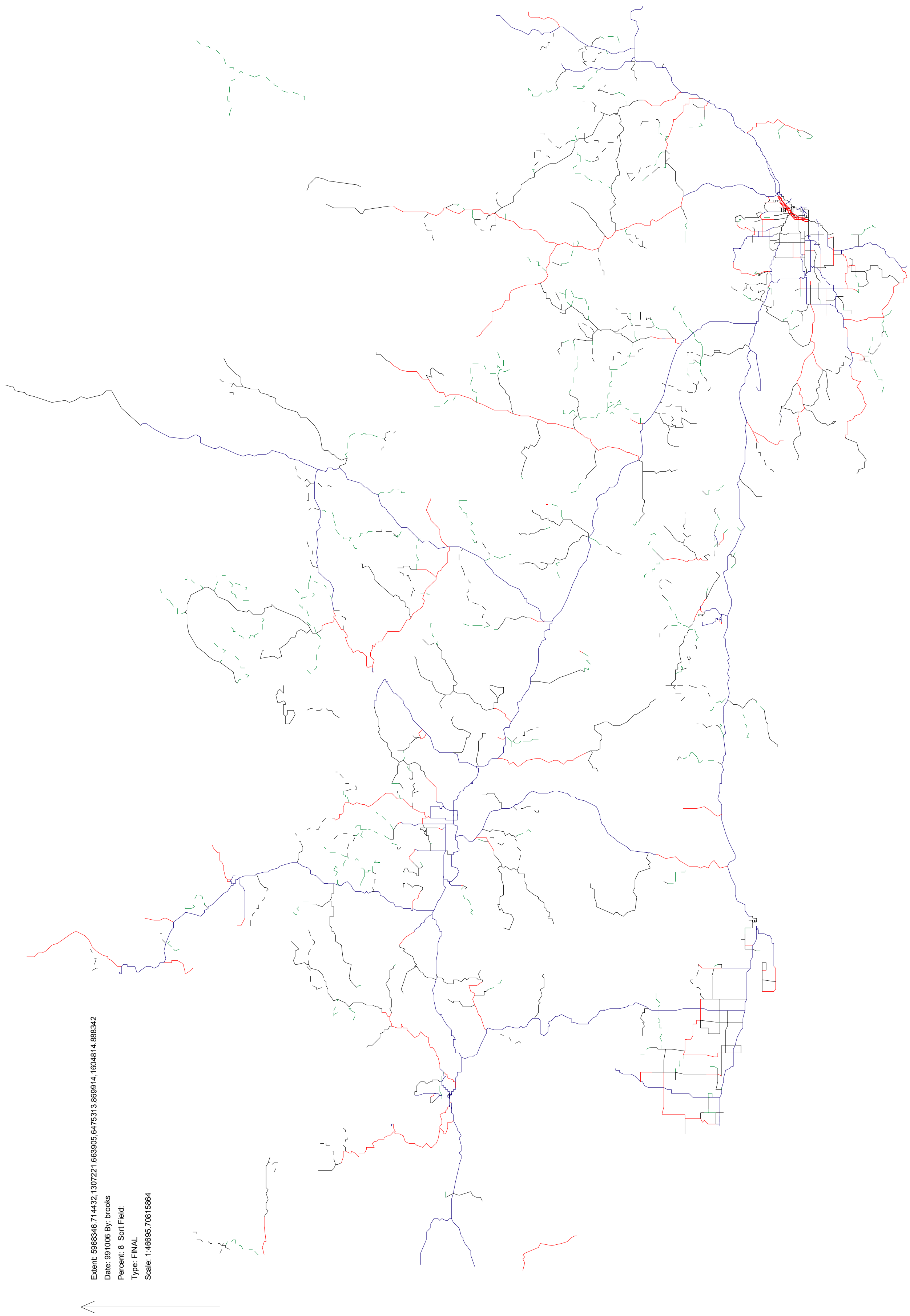


Western Ontario Test Dataset
Automated Generalization Only



Western Ontario Test Dataset

Automated Generalization and 3 hours Post-Processing



Western Ontario Test Dataset

Road Map of Local Area [MapArt, 199?]

Please note that this map is in a different projection than the generalised samples