AUTOMATED URBAN LAND USE AND LAND COVER CLASSIFICATION FOR MESOSCALE ATMOSPHERIC MODELING OVER CANADIAN CITIES

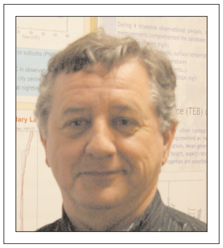
A. Leroux and J.-P. Gauthier, Environment Canada, Canadian Meteorological Centre, Dorval, Québec
A. Lemonsu, S. Bélair and J. Mailhot, Environment Canada
Meteorological Research Division, Dorval, Québec

An automated geospatial database processing approach has been developed to characterize the urban areas of major Canadian cities for use in mesoscale atmospheric modeling. Mesoscale atmospheric numerical models, including urban canopy models such as the Town Energy Balance (TEB) model, require surface characteristics to represent surface processes that occur in cities. The methodology developed in this study uses the following pan-Canadian databases: the National Topographic Data Base (NTDB) vector data for land use and land cover (LULC) characterization, the Shuttle Radar Topography Mission (SRTM-DEM) and the Canadian Digital Elevation Data (CDED1) digital elevation models (DEM) for building height assessment, and census data for characteristics of residential districts. These databases are jointly processed to automatically generate a high-resolution urban LULC classification for Canadian cities. The main benefits of this approach are (a) Canada-wide applicability with available continuous databases, and (b) complete automation, with the exception of optional post-processing.

Pour les besoins de la modélisation méso-échelle utilisée dans l'étude de l'atmosphère, un système automatisé de traitement de base de données géospatiales a été développé pour définir les zones urbaines des grandes villes canadiennes. À l'instar des modèles de canopée urbaine tel que le modèle Bilan d'énergie terrestre (ou Town Energy Balance (TEB*), les modèles numériques méso-échelle utilisés dans l'étude de l'atmosphère requièrent les caractéristiques de surface pour représenter les données de surface des zones urbaines. Les méthodes mises au point dans la présente étude se servent de bases de données pancanadiennes telles que la Base nationale de données topographiques (BNDT), composée de données numériques vectorielles qui servent à définir l'utilisation et la couverture des terres (*land use and land cover*), la Mission topographique radar de la navette spatiale (MTRNS), les Données numériques d'élévation du Canada (DNEC1), constituées de modèles numériques d'élévation utilisés pour calculer la hauteur des édifices, et les données de recensement pour la délimitation des zones résidentielles. Le traitement conjoint de ces bases de données génère automatiquement une classification de l'utilisation et la couverture des terres (*land use and land cover*) à haute résolution spatiale des villes canadiennes, offrant des avantages importants tels que (a) des bases de données disponibles sur l'ensemble du territoire canadien, et (b) une automatisation complète, exception faite d'un post-traitement optionnel.*


A. Leroux


S. Bélair


J. Mailhot

1. Introduction

Mesoscale atmospheric numerical models, with horizontal resolution ranging from a few hundred meters to a few kilometres, require detailed characterization of the earth's surface to drive sophisticated surface parameterization schemes. This requirement is even more important for complex environments such as urban areas. The integration of urban characterization is a new issue which previous Canadian atmospheric numerical models did not take into account. Recently, the Meteorological Service of Canada (MSC) included in their atmospheric numerical models a specific urban canopy model, the Town Energy Balance (TEB) model [*Masson* 2000], to improve the representation of urban covers and their impact on the local meteorology. The purpose of this study is to develop and provide an urban land use and land cover (LULC) database for mesoscale atmospheric modeling of Canadian cities, specifically for the TEB model.

Several urban LULC classifications exist that cover large portions of Canada. These include the Canada Land Inventory (CLI), the Canadian Urban Land Use Survey (CUrLUS) and the Canada Land Use Monitoring Program (CLUMP) [*Cihlar et al.* 2003; *Guindon et al*. 2004; *Zhang and Guindon* 2005]. Recent global land cover classifications include the GLC2000 [*Bartholome and Belward* 2005] and the GlobCover LandCover v2.1

[*Defourny et al.* 2006] initiatives. Existing classifications, however, suffer from at least one of the following elements: (a) they are outdated, (b) the thematic classes do not correspond to the ones required for atmospheric modeling, (c) their spatial resolution is too low, (d) they require significant effort to generate over a large area or (e) human interpretation is strongly involved. The classes generated in this study are used to specify the input parameters of TEB. These are the mean geometric parameters describing the urban canopy and the radiative and thermal properties for roofs, walls and roads.

The only other classification methodology for urban LULC that is appropriate for mesoscale atmospheric modeling with the TEB model for Canadian cities is the satellite imagery classification approach described in *Lemonsu et al*. [2006; 2008]. The new approach proposed here is predicted to replace the previous one and provides several significant practical benefits. The new system involves less human intervention, offers a better spatial resolution, improves the number and determination of classes and is season independent (crops and trees). This method also avoids problems related to satellite imagery such as cloud coverage, the shadow of tall buildings in central business districts and boats over water areas. Furthermore, the proposed scripted workflow for the automated spatial data processing can be directly useful to other applications requiring LULC data over Canadian territory, especially since the National Topographic Data Base (NTDB) and CanVec datasets are now publicly available.

Section 2 details the source data used for the classification method, the data processing and the associated workflow. Section 3 presents the results and discusses the main benefits and limitations of the approach, along with foreseen improvements. Section 4 summarizes the main conclusions.

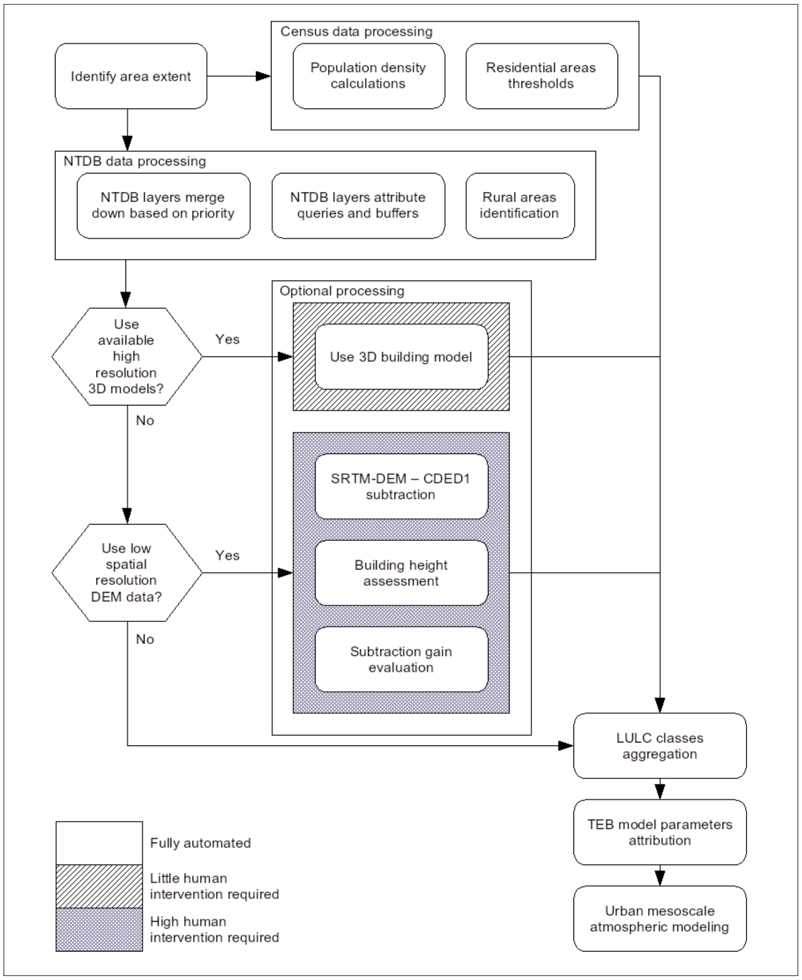


Figure 1: Data processing workflow.

2. Data Processing

The automated LULC classification method presented here has been developed for databases available Canada-wide. Three types of data were employed: vector topographic database, census data, and digital elevation models (DEM). Metadata for each of these datasets is available.

Figure 1 presents the general data processing workflow. It shows the main steps of the data processing and indicates which steps require human intervention. After the area of interest was identified, the core processing of the NTDB database was completed, along with processing of census data. Optionally, a building height assessment is undertaken, which requires human intervention. Ultimately, the new LULC classes were aggregated and readied for urban mesoscale atmospheric modeling.

2.1 Spatial Data Accuracy and Georeference

This approach relies mostly on the NTDB to produce the urban LULC classification. The NTDB was developed by Geomatics Canada. It covers the entire Canadian landmass and contains features normally found on topographic maps at scales of 1:50 000 and 1:250 000.

NTDB vector data was converted to raster format at the beginning of the processing flow for simplicity and speed. DEM databases are provided in raster format and the final output for atmospheric modeling was also required to be in raster format.

The spatial resolution for the rasterization of the 1:50 000 NTDB data and of the resulting urban classification was set to 5 m per pixel. This spatial resolution constitutes a compromise in order to appropriately represent small or narrow entities (e.g., trails, railways, navigable canals) and take into account the data scale and spatial resolution accuracy. Points layers occupy a minimum of one pixel and lines layers are at least 5 m thick. Some layers were buffered to reflect their average physical geometry (see Section 2.3).

Even though absolute accuracy of georeferencing is desirable, it is not crucial to atmospheric modeling which uses much lower spatial resolutions. Indeed, the TEB model ultimately runs with grid sizes varying from a few hundred meters to several kilometres, using coverage fractions of urban classes at each grid point.

2.2 NTDB Layer Priority Assessment

NTDB data is composed of thematic vector layers represented by georeferenced points, lines and polygons. The NTDB provides up to 110 thematic layers (e.g. buildings, roads, waterbodies, built-up areas, vegetated areas) generally accompanied by attributes. Each NTDB National Topographic System (NTS) sheet comes with metadata. Their spatial accuracy targets are 10 m for urban areas, 25 m for rural areas and 125 m for isolated areas (Geomatics Canada 1997). While some thematic layers overlap, certain areas have no associated information, especially rural areas and building vicinities (see Section 2.4). Additionally, vegetation information in the NTDB database is limited and does not specify the type of vegetation, which is important information for atmospheric numerical models (see Sections 2.7 and 3.4). Table 1 provides the list of NTDB layers retained to produce our classification.

Raw 1:50 000 NTDB layers were rasterized and hierarchically ordered by priority to manage overlapping layers (e.g., bridges have priority over water bodies). The priorities attributed to layers are detailed in Table 1. About 30 NTDB layers were excluded, such as entities that do not have a physical representation (e.g., ferry routes) or do not significantly contribute to surface LULC (e.g., tunnels and underground reservoirs). These choices were motivated by the application type for the LULC classification.

2.3 Use of Layer Attributes and Buffer Zones

The distinction of classes for a layer is determined by the attributes available for that specific layer. Two general cases are found: (a) a layer generates multiple classes based on attribute values (for example, a road splits non-exclusively in paved and unpaved roads) and (b) some features are ignored based on attribute values (for instance, railways are ignored when in an underground tunnel).

Buffer zones were applied to some layers based on attribute values, such as when a road attribute indicates the road feature is a highway. Some categories of buildings represented as points were also buffered. Buffer zones may overlap only with specified classes. This processing adjusts specific point and linear features closer to their real world physical size.

NTDB data is composed of thematic vector layers represented by georeferenced points, lines and polygons.

2.4 Rural Areas Identification

The NTDB provides no information for rural and grass areas and for building vicinity areas. This is one of the main limitations of the new approach. Post-processing is required to differentiate these two types of cover. This post-processing consists of generating buffers for certain types of buildings to produce a building vicinity class. It is assumed that building vicinity areas include artificial surfaces (e.g., parking lots, concrete areas).

To obtain a buffer proportional to the building span, selected buildings represented as points were granted a 25 m buffer and selected polygonal buildings a 100 m buffer. The selection was attribute-based. For example, the buffer area of barns, grain conveyors and greenhouses were left in the 'Grass and fields' class, while the buffer area of commercial buildings was associated with the 'Building vicinity' class. In this post-processing, we assumed that areas not associated to any NTDB layer were either grass and fields or building vicinity areas. Even though several urban areas are misclassified by this method, it is still the best automated technique found by the authors to segregate rural areas from commercial building vicinity using the limited information available from the NTDB. Section 3.4 explains how we planned to circumvent this limitation.

2.5 Residential Areas Classification

Residential areas occupy a significant proportion of the urban landscape. The housing types and the presence of vegetation in residential areas have an impact on urban mesocale atmospheric models. In the NTDB, residential areas are associated with the built-up areas layer (Table 1, layer priority #59), which provides little feature attribute information about the type of buildings. Therefore, it is important to use another source of information for residential areas. Census data provides a geographic portrait of the social and economic situation of Canada's population. Statistics Canada collects and compiles census data for Canada every five years.

Table 1: NTDB layer priority attribution.
P: Priority, T: Geometry type (P: Points, L: Lines, A: Area)
Italics: Post-processed layers which have their priorities adjusted later in the process

P	Layer	T	P	Layer	T	P	Layer	T
1	*Building*	P	35	Crane	L	69	Fish pound	A
2	*Building (industrial-commercial)*	P	36	Wind-operated device	P	70	Liquids depot/dump	A
3	*Building (day-night 24/7)*	P	37	Trail	L	71	Wetland	A
4	Tower	P	38	Dam	P	72	Vegetation	A
5	Chimney	P	39	Marina/yacht club	P	73	Transformer station	A
6	*Runway*	P	40	Hazard to air navigation	P	74	Rocky ledge/reef	A
7	*Runway (unpaved)*	*P+A*	41	Navigable canal	L	75	Waterbody	A
8	Liquids depot/dump	P	42	Solids depot/dump	P	76	Drive-in theater	A
9	Transformer station	P	43	*Sports track*	*L+A*	77	Slip	A
10	Ski centre	P	44	*Sports track (paved)*	*L+A*	78	Lumber yard	A
11	Tank	P	45	Lock gate	L	79	Auto wrecker	A
12	Cemetery	P	46	Wharf	L	80	Gas and oil facilities	A
13	Drive-in theater	P	47	*Building*	*A*	81	Exhibition ground	A
14	Picnic site	P	48	*Building (industrial-commercial)*	*A*	82	Ruins	A
15	Camp	P	49	*Building (day-night 24/7)*	*A*	83	Fort	A
16	Campground	P	50	*Railway*	*L*	84	Mining	A
17	Silo	P	51	Conduit	L	85	Stockyard	A
18	Ruins	P	52	Conveyor	L	86	Peat cutting	A
19	Parabolic antenna	P	53	Hazard to air navigation	L	87	Golf course	A
20	Gas and oil facilities	P	54	Dam	L	88	Golf driving range	A
21	Lookout	P	55	Pipeline	L	89	Campground	A
22	Historic site/point of interest	P	56	Wall/fence	L	90	Amusement park	A
23	*Mining (surface)*	*P*	57	Ford	L	91	Park/sport field	A
24	*Mining (underground)*	*P*	58	Watercourse	L	92	Picnic site	A
25	Boat ramp	P	59	*Built-up area*	*A*	93	Zoo	A
26	*Seaplane base*	*P*	60	Navigable canal	A	94	Botanical garden	A
27	Ford	P	61	Seawall	L	95	Cemetery	A
28	Lock gate	P	62	Dike/levee	L	96	Sand	A
29	Footbridge	L	63	Esker	L	97	Dam	A
30	*Bridge*	*L*	64	Breakwater	L	98	Solids depot/dump	A
31	*Road*	*L*	65	*Runway*	*A*	99	Cut	A
32	*Highway*	*L*	66	*Stadium*	*A*	100	Embankment	A
33	*Road (unpaved)*	*L*	67	Tank	A	101	Dry river bed	A
34	Limited-use road	L	68	Lookout	A	102	Permanent snow and ice	A

The census data provides information on population density and allows the distinction of residential districts at the dissemination area (DA) level (which generally corresponds to one city block or more) [*Statistics Canada* 2002]. The DA boundary layer is spatially compatible with the NTDB.

Population densities were calculated after excluding water bodies from the original DA polygons. Population density thresholds were established in order to generate five classes of residential areas. Some residential areas associated with high building areas were later reclassified into building classes. Since buildings in the central business districts may be composed of both building and built-up NTDB layers, the reclassification of the built-up layer where high elevation occurs was required to identify all tall buildings (see Section 2.6).

2.6 Building Height Assessment

Elevation information was used to complement the NTDB for building height assessment because the urban microclimate largely depends on the three-dimensional configuration of the urban canopy. Two DEM sources were used, one for ground-level elevation and one for the total height of features over a given area. The first DEM used was the Canadian Digital Elevation Data, Level 1 (CDED1), a free Canada-wide ground elevation DEM at the 1:50 000 scale distributed by Natural

Resources Canada. The CDED1 DEM was extracted from the NTDB or from various scaled positional data. The second DEM used, based on NASA's Shuttle Radar Topography Mission (SRTM) in February 2000, covers most of the world with a 3 arc-seconds spatial resolution. This free database provides elevation at the top of features, such as the top of buildings and the top of the vegetation canopy. Because the SRTM-DEM database only provides information up to a latitude of 60° north, the city of Whitehorse is the only major Canadian city not covered by it. The SRTM-DEM database constitutes the lowest spatial resolution data (about 90 m per pixel) included in this approach.

Identification of new classes from NTDB classes can be accomplished by using information on urban canopy height. The general idea of this building height assessment method is to subtract CDED1 elevation from SRTM-DEM elevation. CDED1 data represent ground elevation values, while SRTM-DEM data represent elevation at the top of features, including vegetation and buildings [*Walker et al.* 2007]. This method is discussed in detail in *Lemonsu et al.* [2008]. A similar technique was employed by *Kellndorfer et al.* [2004] for vegetation height assessment. This method is far from perfect but nonetheless allows successful identification of tall building areas in central business districts and moderate building elevations in densely populated built-up areas using freely available Canada-wide data. The automation is not considered complete because the calculation of SRTM-DEM minus CDED1 requires human intervention. This operation can however be optional, since without this input data the classification still provides valuable classes over urban areas. SRTM-DEM data has to be systematically evaluated and corrected to replace missing data and abnormal values.

The subtraction was completed at the CDED1 spatial resolution, after the SRTM-DEM data was interpolated bilinearly. An elevation gain was obtained by calculating the subtraction for grass and field areas, which are likely to be flat surfaces. This gain was then removed from the result of the subtraction to compensate for any present offset, a technique similar to that used by *Kellndorfer et al.* [2004]. Building height assessment values were preserved only for the central business district and its surroundings in order to limit incorrect height attribution. Height data was ultimately considered only for 2D buildings, including the forts and stadium NTDB layers, in addition to built-up residential areas over a specified height threshold.

Mesoscale atmospheric models do not require individual building geometry but an estimation of mean urban canopy height. The absolute height value accuracy for individual buildings is poor, mainly due to the spatial resolution of the SRTM-DEM database. But the height assessment is good enough to distinguish classes of building height within a city with appropriate thresholds. According to our assessment and to *Gamba et al.* [2002], the building height derived from STRM-DEM is underestimated. Section 3.4 describes how this whole subtraction approach could be replaced by more appropriate data and automated processing.

2.7 LULC Classes Aggregation

Classes generated by this processing were aggregated into 44 classes (identified in Table 2). These were defined based on physical similarity in accordance with the classes typically used in TEB. Three types of building usage (i.e. industrial-commercial, open 24/7, others and unspecified) were distinguished following attributes within the NTDB. These three types of buildings were useful to help specify anthropogenic fluxes and for other applications related to atmospheric modeling. Four thresholds of building heights (less than 10 m, 10 to 20 m, 20 to 30 m, and taller than 30 m) were specified in order to establish mean geometric parameters at the model spatial scale. The LULC classes could then be associated with TEB input parameters, i.e., geometric parameters as well as radiative and thermal properties [*Masson* 2000]. The new classes were somewhat different from those generated by the satellite imagery approach described in *Lemonsu et al.* [2006; 2008] and currently used to provide TEB input parameters. An additional difficulty was related to the fact that some classes are precise thematic classes, such as paved roads, whereas others are of uncertain and potentially diversified composition, such as zoos and ruins. Thus, some classes had parameters directly associated with them while surface composition statistics had to be obtained for others.

Classes generated by this processing were aggregated into 44 classes. These were defined based on physical similarity in accordance with the classes typically used in TEB.

A vegetation mask was generated in order to reduce the impact of discontinuities at the interface (located far from the studied urban area) between adjacent LULC classes that were specified from data sources with significantly different scales, such as what is done in mesoscale atmospheric models to provide information for a large domain. This mask consists of large and relatively continuous vegetated regions. For meteorological modeling, masked areas from the urban classification will be replaced by lower resolution data which is continuous over the whole modeled area. Only natural cover classes, excluding water, are removed by the use of the vegetation mask. Since the NTDB provides no information on the vegetation types, the benefit of this

Table 2: LULC classes.

	Buildings		Roads and transportation network
🟦	1D buildings	🟥	Highway (including runways)
🟦	1D buildings (industrial-commercial)	🟥	Paved road
🟦	1D buildings (day-night 24/7)	🟧	Unpaved roads and paths
🟦	Polyg. Very Low < 10 m	⬛	Railway
🟦	Polyg. Very Low < 10 m (industrial-commercial)	🟥	Bridge
🟦	Polyg. Very Low < 10 m (day-night 24/7)	🟥	*Other*
🟦	Polyg. Low 10-20 m		**Industrial and other constructions**
🟦	Polyg. Low 10-20 m (industrial-commercial)	🟪	Industrial-commercial (building excluded)
🟦	Polyg. Low 10-20 m (day-night 24/7)	🟪	Towers
🟦	Resid. Mid 10-20 m	🟪	Various conduits
🟦	Polyg. Mid 20-30 m	🟪	*Water-related*
🟦	Polyg. Mid 20-30 m (industrial-commercial)	🟥	*Other*
🟦	Polyg. Mid 20-30 m (day-night 24/7)		**Mixed covers**
🟦	Resid. Mid 20-30 m	▨	Building vicinity
⬜	Polyg. High > 30 m	🟩	Green areas
⬜	Polyg. High > 30 m (industrial-commercial)	🟩	*Other*
⬜	Polyg. High > 30 m (day-night 24/7)		**Natural covers**
⬜	Resid. High > 30 m	🟩	Trees
	Residential areas	🟩	Grass and fields
🟨	Very Low pop density (< 2000 persons/km2)	🟫	Soils and rock
🟨	Low pop density (2000-5000 persons/km2)	🟫	Wetlands
🟨	Medium pop density (5000-15000 persons/km2)	🟦	Water
🟧	High pop density (15000-25000 persons/km2)	⬜	Permanent snow and ice
🟧	Very high pop density (> 25000 persons/km2)	🟨	Sand

mask is the use of other vegetation classifications of lower spatial resolutions that are already used and evaluated for meteorological modeling.

2.8 Scripted Spatial Data Processing

In order to automate the whole process over any Canadian area, a flexible script was developed based on the Tool Command Language (TCL), with Application Programming Interface (API) extensions. This code uses previously developed in-house code libraries and the community-developed Geospatial Data Abstraction Library (GDAL) and OGR Simple Feature Library open source code libraries [*Walter et al*. 2002]. The script allows the automated identification of files and datasets, reads and writes multiple formats including custom meteorological formats, processes any recognized projection and datum, and takes into account the spatial resolution and the scale of the input data. The script proceeds to many spatial data geoprocessing tasks, including cropping, subtraction (cookie cutting), buffering, rasterizing, performing Structured Query Language (SQL) queries on attributes, flattening multiple layers (merge down), performing basic spatial queries, look-up table value attributions and much more. For a user-submitted urban area bounding box, all the required data sources are automatically identified along with the geographic parameters of the area. This scripting approach is essential to the automation of the urban classification and constitutes a significant benefit.

3. Results and Discussion

3.1 Results and Discussion

As a first step, the classification method was applied to the cities of Montreal and Vancouver. The classification for these cities is a matrix of about 10 000 x 10 000 pixels of 5 m spatial resolution. Figures 2 to 5 show the results of the LULC classification for Montreal and Vancouver. The results show homogeneity in the urbanization layout of cities. Several urban features can be observed in Figures 2 and 3 over the Montreal area: the industrial-commercial area surrounding the international airport, south-west Montreal Island (Figure 2), the petrochemical industrial area, north-east Montreal Island (Figure 2) and the central business district east of Mount Royal Park (Figure 3). Distinct neighbourhoods can be distinguished by their population density, such as the adjacent Town of Mount Royal and Parc Extension districts (Figure 2).

Results were also generated for the following major Canadian cities (not shown here) and without the optional building height assessment processing: Calgary, Edmonton, Halifax, Ottawa, Quebec, Regina, Toronto, Victoria and Winnipeg.

A systematic evaluation of the quality of source data and of all processing took place during the development of the procedure. Aerial photographs and other available sources of ground truthing were used, such as VanMap for Vancouver, Navigateur urbain for Montreal and additional minor validation sources such as the Google Maps and Microsoft Virtual Earth tools, for additional high-resolution imagery.

An exhaustive qualitative accuracy assessment of the classification results was completed for Montreal and Vancouver. The classification results were largely derived directly from the NTDB dataset. Consequently, the NTDB is directly accountable for the quality of the classification. No quantitative accuracy assessment was completed since this was not judged essential due to the thorough validation conducted on the NTDB by the data provider [*Geomatics Canada* 1997 and section 2.1]. Additionally, all NTDB layers are spatially coherent among themselves and census data is spatially compatible with NTDB layers. The resulting classification is of much higher spatial resolution and quality, in its ‘fitness for use’ sense [*Mostafavi et al*. 2004], than is required for mesoscale atmospheric modeling.

3.2 Benefits

Based on the urban LULC classification described above, the input parameters required by TEB, or by any other urban surface schemes used in atmospheric models for that matter, can be described in a much more accurate manner compared with results from the satellite imagery approach described in *Lemonsu et al.* [2006; 2008]. In addition, this new classification is in general

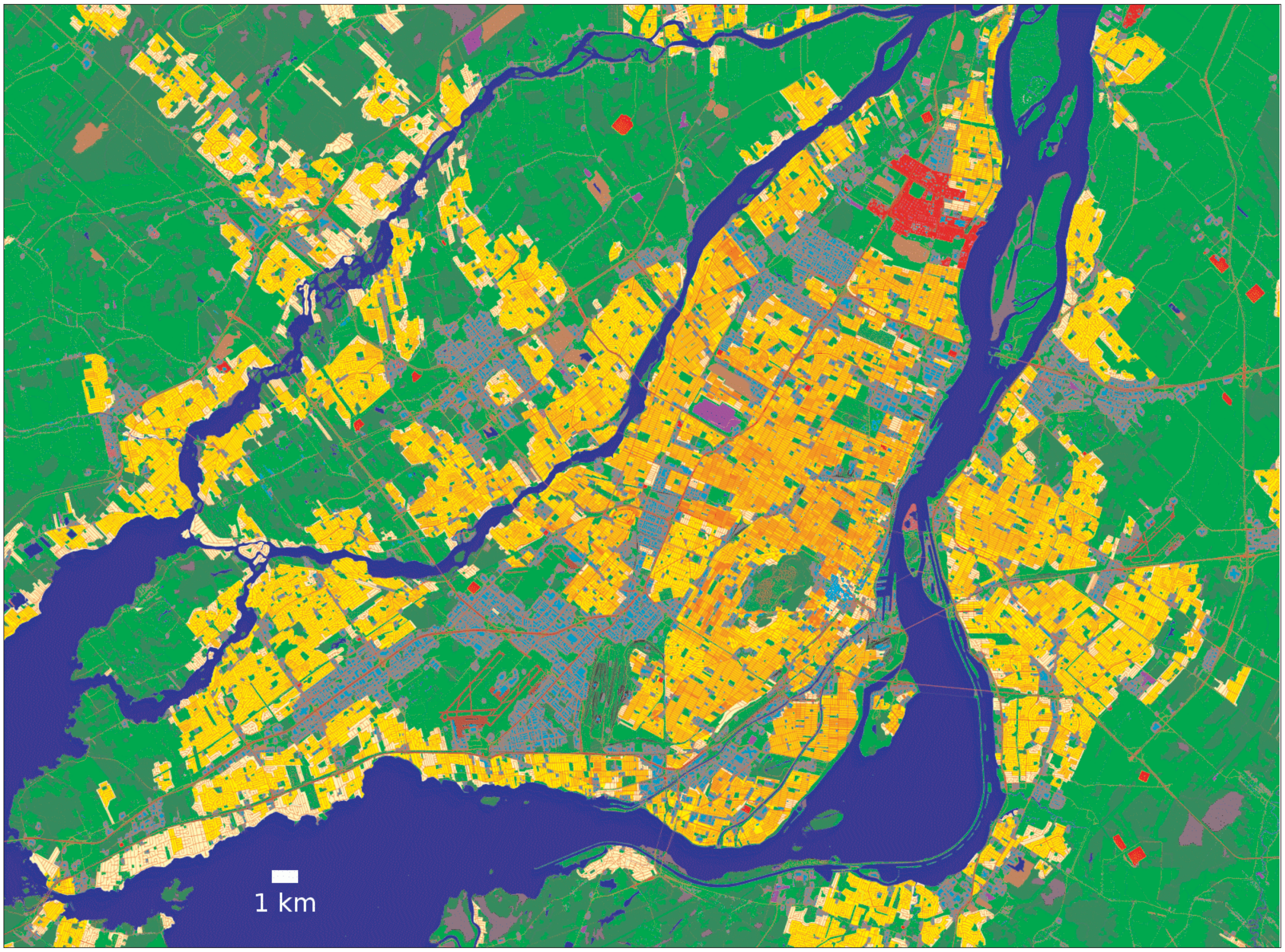


Figure 2: LULC classification over Montreal, overview. Colour coding for the various classes are given in Table 2.

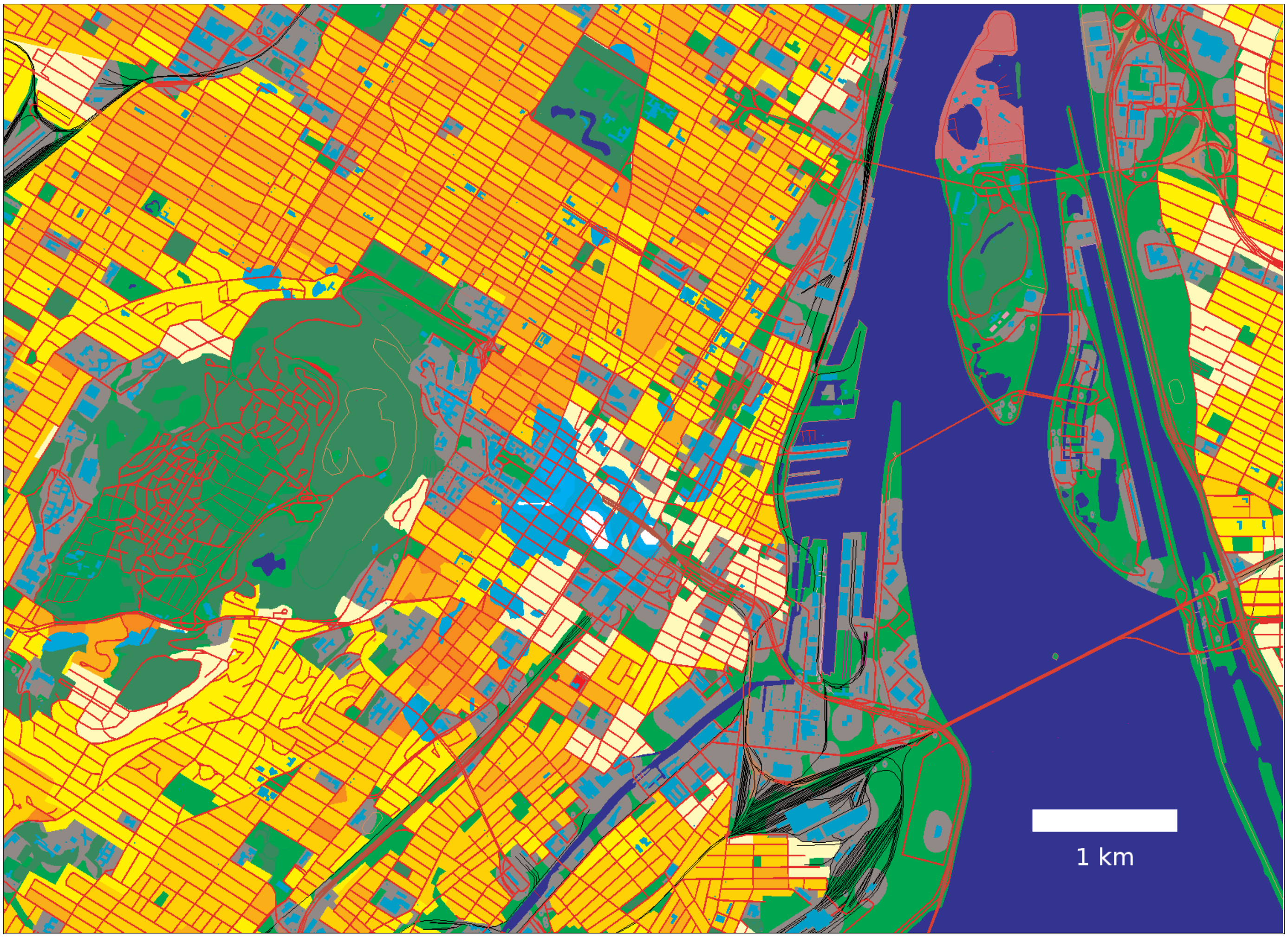


Figure 3: LULC classification over Montreal, central business district. Colour coding for the various classes are given in Table 2.

agreement with *Cihlar et al*. [1998], who identified the qualities of a good classification; i.e., accurate, reproducible by others given the same input data, robust (not sensitive to small changes in the input data) yet able to fully exploit the information content of the data, uniformly applicable over the whole domain of interest, and objective (not dependent on the analyst's decisions).

The two main benefits of our methodology are (1) its Canada-wide applicability, using the same data across Canada with a common data format, available at limited or no cost within the Canadian federal government (NTDB, CDED1, census data) or from on-line sources (SRTM-DEM), and (2) the mostly complete automation of the process, with only the bounding box of the area of interest required for the classification to be computed. Among other benefits, the approach is simple and flexible, the classification involves nearly no interpretation, the computations are fast (from three minutes for Victoria (BC) to 40 minutes for Toronto (ON) on a personal workstation), and the developed code is directly reusable for other data processing projects. When compared with the satellite imagery classification described in *Lemonsu et al*. [2006; 2008], the new approach involves less human intervention, offers a better spatial resolution, and improves the number and determination of classes. The new approach is also season independent (crops and trees) and avoids problems related to satellite imagery such as cloud coverage, the shadow of tall buildings in central business districts and boats over water areas. However, the accuracy of the generated classification is limited by the timeliness of updates to the databases used.

3.3 Limitations

This urban classification approach also has a few important limitations: (1) NTDB data is based on aerial imagery sometimes years old and is not up to date. Buildings are missing in the central business districts of some cities, and recent suburban developments may be missing (these missing suburbs may be present in census data and recent satellite imagery used for ground truthing). (2) In spite of the numerous thematic layers of the NTDB, some LULC features are not represented. For instance, there is no direct way to distinguish paved surfaces, such as parking lots, from other non-vegetated uninhabited areas in urban and rural areas. Post-processing is required to circumvent this absence of thematic representation (see Section 2.4). (3) Little information is directly provided by the NTDB data over residential areas, in particular regarding the built and vegetation ratio and the diversity of urban residential areas. Census data allows the categorization of residential types, but does not inform on vegetation presence. (4) The building height assessment methodology presented here allows the distinction between categories of building heights, but only at a low horizontal resolution with poor vertical accuracy. (5) The NTDB provides little information regarding vegetation types.

3.4 Foreseen Improvements

Since the need for urban classifications supporting atmospheric modeling and environmental modeling in general is increasing, together with the spatial resolution and sensitivity of numerical atmospheric models themselves, several improvements to the approach are anticipated. Among these: (1) the processing will be adapted to include appropriate newly available high-resolution Canada-wide databases such as the CanVec database, the Earth Observation for Sustainable Development of Forests (EOSD) database [*Wulder et al*. 2003] and the most recent census data. The use of the EOSD database, which provides 22 vegetation classes at a spatial resolution of 25 m over Canada, will also render obsolete the need to generate vegetation masks mentioned in Section 2.7. (2) 3D building vectors will be used, when available, to replace the building height assessment calculated through the STRM-DEM minus CDED1 subtraction. 3D building data has been used for the city of Ottawa and has significantly improved the urban classification, especially the building height classes. In this case, data with the lowest spatial resolution is replaced with data of high spatial resolution. In addition, using such data significantly

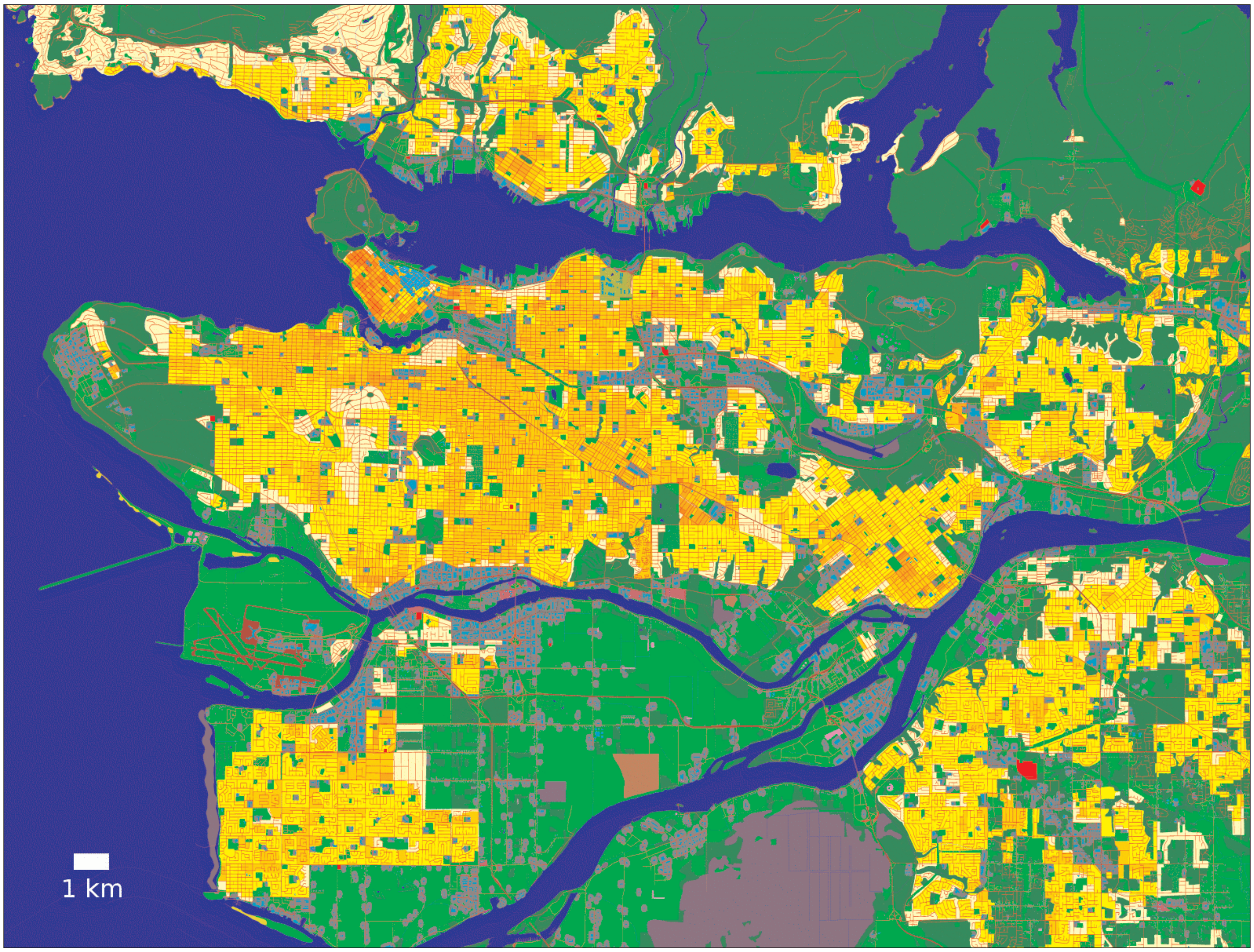


Figure 4: LULC classification over Vancouver, overview. Colour coding for the various classes are given in Table 2.

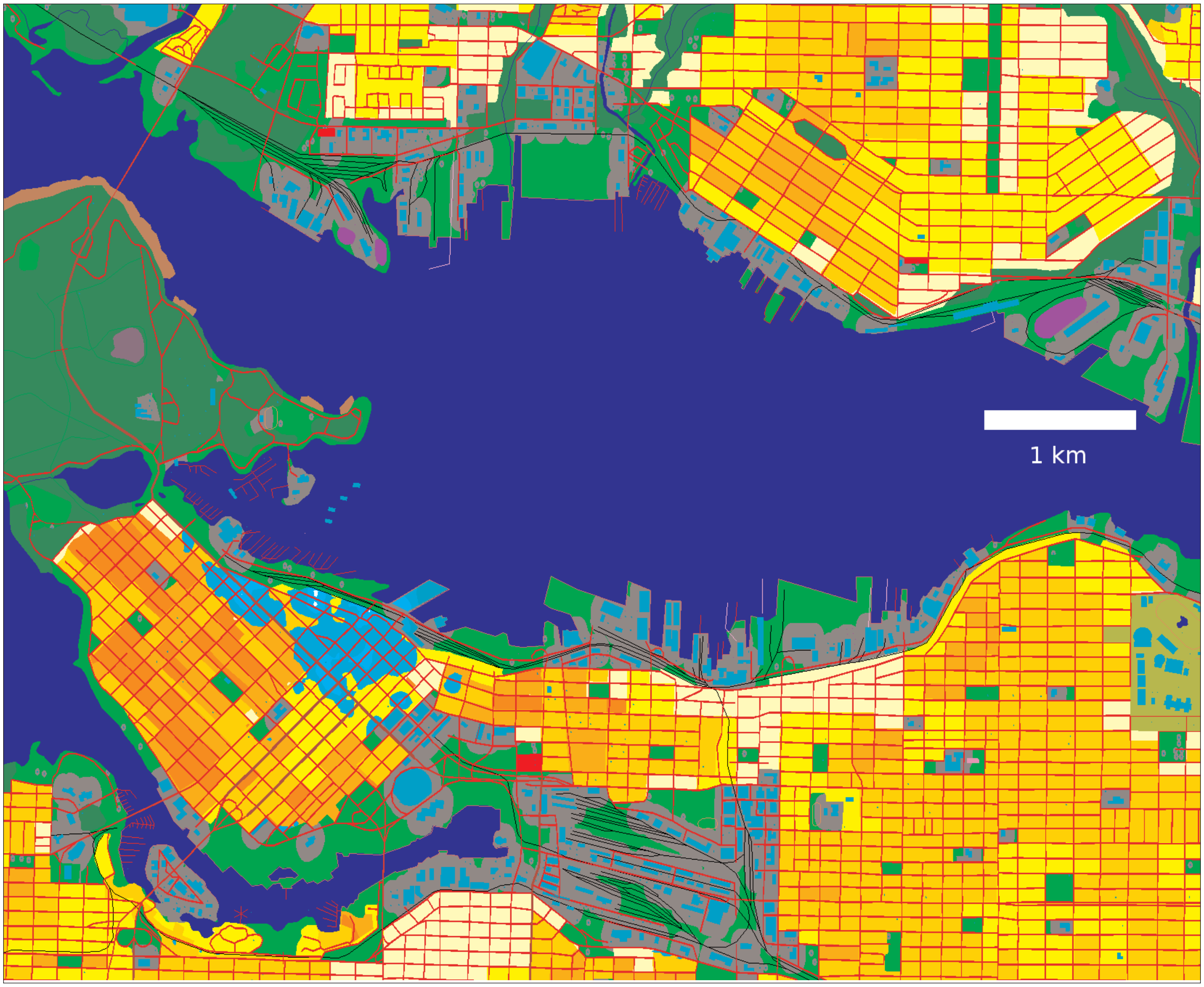


Figure 5: LULC classification over Vancouver, central business district. Colour coding for the various classes are given in Table 2.

reduces the amount of human intervention required for the optional building height assessment. Such high-resolution 3D building data, typically used for computational fluid dynamics (CFD) modeling, would allow a more direct specification of building geometric parameters required by the TEB model. (3) A complete analysis will be conducted to evaluate TEB sensitivity to the urban classification in order to evaluate the impact of spatial resolution and quality of input data on atmospheric modeling. (4) Finally, efforts will be made to increase the use of all available information from the databases, such as the number of residences in census data and additional attributes in the NTDB.

4. Conclusion

A new automated urban LULC classification approach is presented in this study. The main objective is to improve input parameters for mesoscale atmospheric modeling, particularly for the TEB surface scheme, over Canadian cities. The resulting classification significantly improves the urban classification and other pan-Canadian LULC classifications used today at MSC. The methodology requires readily available Canada-wide geospatial databases, namely the NTDB for LULC characterization, census data for residential characterization, and SRTM-DEM together with CDED1 DEM for building height assessment. To the knowledge of the authors, no other automated LULC classification methodology for urban areas in Canada exists at a similar spatial resolution.

The data processing is fully automated with the exception of the optional building height assessment. The proposed methodology is flexible and extensible. Even if the actual method is tailored to the requirements of mesoscale atmospheric modeling, several other applications could benefit from this approach for the generation of Canada-wide LULC classifications, especially since NTDB and CanVec datasets became publicly available in April 2007.

The limitations of the resulting LULC classification are discussed, along with foreseen improvements which will significantly enhance the overall accuracy of the LULC classes. More research is required to address two elements of the TEB scheme which are not considered in the present work: the urban vegetation characterization and the identification of radiative and thermal properties of urban materials.

Acknowledgments

This work was funded by the Chemical, Biological, Radiological and Nuclear (CBRN) Research and Technology Initiative (CRTI) (Project #02-0093RD).

References

Bartholome, E. and A.S. Belward. 2005. GLC2000: a new approach to global land cover mapping from Earth observation data. *Int. J. Remote Sensing*, 26(9), pp. 1959-1977.

Cihlar, J., B. Guindon, J. Beaubien, R. Latifovic, D. Peddle, M. Wulder, R. Fernandes, and J. Ken. 2003. From need to product: a methodology for completing a land cover map of Canada with Landsat data. *Canadian Journal of Remote Sensing*, 29(2), pp. 171-186.

Cihlar, J., Q. Xiao, J. Chen, J. Beaubien, K. Fung, and R. Latifovic. 1998. Classification by progressive generalization: a new automated methodology for remote sensing multichannel data. *Int. J. Remote Sensing*, 19(14), pp. 2685-2704.

Defourny, P., C. Vancutsem, P. Bicheron, C. Brockmann, F. Nino, L. Schouten, and M. Leroy. 2006. GlobCover: a 300 m global land cover product for 2005 using Envisat MERIS time series. ISPRS Commission VII Mid-term Symposium "Remote Sensing: From Pixels to Processes." Enschede, The Netherlands, pp. 59-62.

Gamba, P., F. Dell'acqua, and B. Houshmand. 2002. SRTM data characterization in urban areas. ISPRS Commission III Symposium on Photogrammetric Computer Vision, Graz, Austria, XXXIV(3B), pp. 55-58.

Geomatics Canada. 1997. Standards and specifications of the National Topographic Data Base – Edition 3.1. Minister of Supply and Services Canada, 32 p.

Guindon, B., Y. Zhang, and C. Dillabaugh. 2004. Landsat urban mapping based on a combined spectral-spatial methodology. Remote Sensing of Environment, No. 92, pp. 218-232.

Kellndorfer, J.M., W.S. Walker, L.E. Pierce, M.C. Dobson, J. Fites, C. Hunsker, J. Vona, and M. Clutter. 2004. Vegetation height derivation from Shuttle Radar Topography Mission and National Elevation data sets. Remote Sensing of Environment, 93 (3), pp. 339-358.

Lemonsu, A., A. Leroux, S. Belair, S. Trudel, and J. Mailhot. 2006. Methodology of urban cover classification for atmospheric modeling. 6th Symposium on the Urban Environment, January 30 – February 2, 2006, Atlanta, GA, American Meteorological Society, 5.5.

Lemonsu, A., A. Leroux, S. Belair, S. Trudel, and J. Mailhot. 2008 (submitted). A general methodology of urban land cover type classification for atmospheric modelling. *Journal of Applied Meteorology and Climatology*.

Masson, V. 2000. A physically-based scheme for the urban energy budget in atmospheric models. Boundary-Layer Meteorology, No. 94, pp. 357-397.

Mostafavi, M.A., G. Edwards, R. Jeansoulin. 2004. An ontology-based method for quality assessment of spatial data bases. ISSDQ 2004, Brunck-Leitha, Austria, 14 p.

Statistics Canada. 2002. 2001 Census Dictionary. Statistics Canada Catalogue no. 92-378-XIE, Ottawa, Canada, 382 p.

Walker, W.S., J.M. Kellndorfer, and L.E. Pierce. 2007. Quality assessment of SRTM C- and X-band interferometric data: Implications for the retrieval of vegetation canopy height. Remote Sensing of Environment, No. 106, pp. 428-448.

Walter, G., F. Warmerdam, and P. Farris-Manning. 2002. An open source tool for geospatial image exploitation. Geoscience and Remote Sensing Symposium, IGARSS '02, June 2002, Vol. 6, pp. 3522-3524.

Wulder, M.A., J.A. Dechka, M.A. Gillies, J.E. Luther, R.J. Hall, A. Beaudoin, and S.E. Franklin. 2003. Operational mapping of the land cover of the forested area of Canada with Landsat data: EOSD land cover program. *Forestry Chronicle*, 79(6), pp. 1075-1083.

Zhang, Y. and B. Guindon. 2005. Using Landsat data to assess land use conversion impacts arising from urbanization: the Canadian context. 5th International Symposium on Remote Sensing of Urban Areas (URS 2005), Tempe, Az, USA, 4 p.

MS rec'd 08/04/24
Revised MS rec'd 08/11/07

Authors

Alexandre Leroux is an engineer specializing in geomatics and remote sensing at the Canadian Meteorological Centre of Environment Canada. He works within the Environmental Emergency

Response Section on projects related to urban geospatial databases and land use land cover classifications for mesoscale meteorological modeling and CFD modeling in Canada, on worldwide geospatial databases for meteorological models and on geospatial data diffusion. Before joining the Canadian federal government, he worked in the private sector on geomatics and hyperspectral remote sensing projects. He is also co-founder of *Slashgeo.org*, a popular community-driven website for geospatial news and discussions.

Stéphane Bélair is a research scientist in the Meteorological Research Division of Environment Canada since 1997. Dr. Bélair is mainly working on improving the representation of physical processes in local, regional, and global numerical prediction weather systems, with a focus on clouds, precipitation, boundary-layer turbulence, and land surface processes. Dr. Bélair was the scientific leader of several major operational implementations at the Canadian Meteorological Centre (CMC) in the last decade. He was head of the global modeling group of the Numerical Prediction Section from 2001 to 2006, and has been since head of the land surface modeling and assimilation group. He is also a member of several international committees and initiatives.

Dr. Jocelyn Mailhot is a Senior Research Scientist and Head of the Mesoscale Modeling Group in the Meteorological Research Division of Environment Canada's S&T Branch. He has led the development of mesoscale models and physics parameterizations. He has been the PI or a co-investigator in several major field projects related to the atmospheric boundary layer and clouds, and has collaborated to several international modeling programs. He is one of the scientific leaders for the development of the Canadian urban forecast prototype system. He is currently leading the development of a very-high-resolution operational forecast system to be used at the different competition venues during the Vancouver 2010 Winter Olympics. ❑