5.5 A GENERAL METHODOLOGY OF URBAN COVER CLASSIFICATION FOR ATMOSPHERIC MODELLING

Aude Lemonsu*
Environment Canada, Meteorological Service of Canada, Meteorological Research Branch

Alexandre Leroux
Environment Canada, Meteorological Service of Canada, Environmental & Emergency Response Division

Stéphane Bélair
Environment Canada, Meteorological Service of Canada, Meteorological Research Branch

Serge Trudel
Environment Canada, Meteorological Service of Canada, Environmental & Emergency Response Division

Jocelyn Mailhot
Environment Canada, Meteorological Service of Canada, Meteorological Research Branch

1. INTRODUCTION

To represent in a realistic way the urban areas, the Town Energy Balance (TEB) (Masson 2000) urban surface scheme was recently implemented in the physics package of the Global Environmental Multiscale (GEM) and the Mesoscale Compressible Community (MC2) Canadian models. Detailed data describing the urban landscapes are required to initialize input parameters used by TEB. These parameters are usually obtained from land-use and land-cover (LULC) databases that characterize the spatial variability of the surface properties. At the moment, no such appropriate database is currently available for Canadian cities. Based on the global land-cover characteristics (GLCC) database (Loveland et al. 2000), GEM and MC2 currently use twenty-six classes including water bodies, ice, and various kinds of soils and vegetation covers. The database only considers one urban class defined from the Digital Chart of the World (DCW) (Danko, 1992). At present, without urban parameterization, this class is considered by GEM and MC2 as sand with large roughness.

Because of the lack of data for Canada and the recent needs for urban modelling applications, our main purpose in this study is to develop a general methodology for producing urban LULC classifications in a semi-automatic way for major Canadian cities. This method is based on the joint analysis of satellite imagery and digital elevation models (DEMs), and the application of a decision tree model to identify the urban classes (Fig. 1). Considering the targeted large spatial coverage and availability, we opted for medium-resolution data: advanced spaceborne thermal emission and reflection (ASTER) and Landsat-7 satellite imagery, shuttle radar topography mission (SRTM-DEM), national elevation dataset (NED), and Canadian digital elevation data level 1 (CDED1) DEMs. By coupling satellite imagery and DEMs, the description of the urban covers is improved because both surface properties and geometric characteristics of the urban canopy are taken into account.

Two application cases are conducted for Oklahoma City (OKC) (OK, United States) and Montreal (MTL) (QC, Canada) in order to develop and assess the methodology (Lemonsu et al., 2005). A first classification is produced from a data analysis carried out using no predetermined classes (neither for the number of classes nor for their characteristics). This approach is then refined for MTL in order to establish a flexible general methodology adapted to the entire North America.

2. SOURCE DATA

2.1. ASTER and Landsat-7 imagery processing

The ASTER satellite image of 21 July 2001 is processed on the large urban area of OKC. The nine VNIR (15-m resolution) and SWIR (30-m resolution) bands of the ASTER image are used, by disaggregating the SWIR bands on the same 15-m grid than that of the VNIR bands through nearest neighbor resampling.

---

* *Corresponding author address*: Aude Lemonsu, Environment Canada, Meteorological Service of Canada, Meteorological Research Branch, 2121 TransCanada Highway, Dorval (QC) H9P 1J3, Canada. E-mail: aude.lemonsu@ec.gc.ca

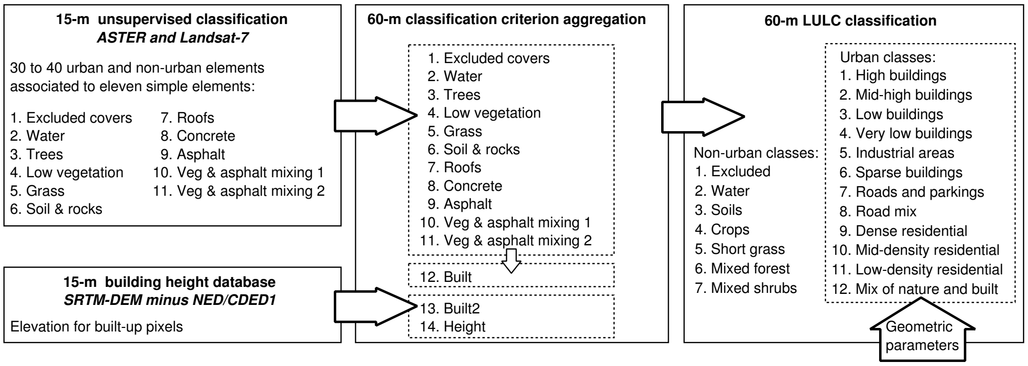

Fig.1 - General description of the methodology.

For the MTL's region, the Landsat-7 image of 8 June 2001 is chosen. By using the 15-m panchromatic (PAN) band, a PAN-sharpening algorithm is applied to the 30-m VNIR-SWIR bands of the Landsat-7 image to disaggregate the multi-spectral data at a 15-m resolution.

Both images are processed to produce an unsupervised isodata classification of single elements at a 15-m horizontal resolution. The algorithm is performed for 30 and 40 elements for OKC and MTL, respectively. Some of these elements are grouped together since their signatures are really close and they represent very similar types of covers. As summarized in the first panel of Fig. 1, eleven elements are finally defined, i.e., (1) *excluded covers* (2) *water,* (3) *trees,* (4) *low vegetation,* (5) *grass,* (6) *bare soil and rocks,* (7) *roofs,* (8) *concrete,* (9) *asphalt,* (10) *vegetation and asphalt mixing 1,* and (11) *vegetation and asphalt mixing 2*.

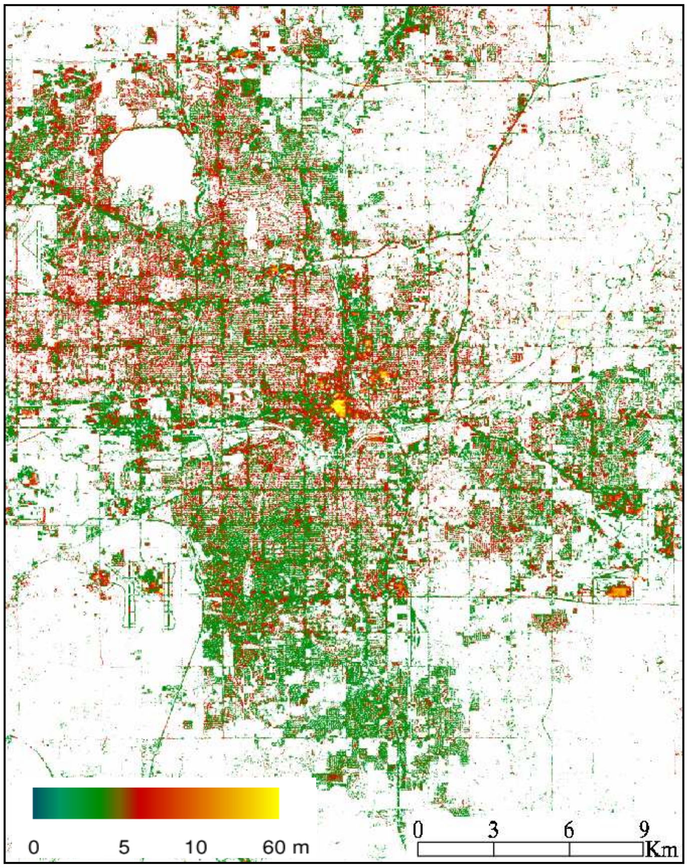

Fig. 2 - Estimation of building heights resulting from SRTM-DEM minus NED/CDED1 databases for OKC.

2.2. Building height database

The methodology for the estimation of building heights consists in combining databases of total elevation (i.e. the sum of the bald Earth's topography and of the obstacles) and of bald Earth's topography. By computing the differences between these elevation datasets, we obtain an estimation of the obstacles above the ground level (i.e., the height of trees and buildings).

The field of total elevation results from the quasi-global STRM-DEM database (post-2000) for both OKC and MTL with a spatial resolution of 1 arc-sec (i.e. 30 m). The bald Earth's topography comes from NED (1998 and 2001) for OKC (1/3 arc sec) and from CDED1 (1995-1999) for MTL (0.75-3.00 arc-sec). Since the spatial resolution of SRTM-DEM is lower than those of NED and CDED1, the SRTM-DEM database has to be disaggregated on the grids of NED and CDED1. The elevation field is finally defined at 15-m resolution for both OKC and MTL. As our objective is the analysis of urban covers, information on vegetation height is not taken into account. Only the results of SRTM-DEM minus NED/CDED1 for pixels correspond-ding to built-up areas (identified by the ASTER and Landsat-7 classifications) are kept.

Figure 2 shows the field of building heights obtained for OKC. Most of the building heights are lower than 10 m, due to the predominance in OKC of residential areas composed of single- or multi-family housings which do not exceed 3 stories. The city business centre can be associated with the core of very large buildings heights reaching 60 m observed in the centre of the study domain.

3. URBAN CLASSIFICATION

The eleven single elements are now used as classification criteria for the identification of the urban classes. We assume that each kind of urban landscape can be described as a combination of these classification criteria.

3.1. Aggregation process

The urban classes are obtained by the aggregation of 15-m pixels on a lower resolution target grid of 60-m resolution (i.e. 4×4 pixels).

The fractions of the eleven single elements are computed for the new target grid. The fractions of vegetated and urban elements are calculated relative to the fraction of ground (i.e., without taking *excluded covers* and *water* into account). The fraction of total built-up areas (referred as *built*) is also calculated, as the sum of *roofs, concrete, asphalt, vegetation and asphalt mixing 1,* and *vegetation and asphalt mixing 2*.

The same aggregation is performed for the building height databases. The mesh fraction of built-up areas and the mean building height are computed and referred to as *built2* and *height*, respectively. It should be noted that *height* only includes pixels with built-up areas within this fraction.

3.2 Statistical methods of classification

In order to identify a limited number of urban classes, a decision tree model (see Fig. 3) is applied to the aggregated fields. For each test of the decision tree model, a qualitative validation is performed by comparing the classification results to aerial photographs of the study domain.

The criterion *built* is used to differentiate urban and non-urban covers. It is assumed that the covers are purely natural for aggregated pixels with under 10% of built-up areas. Thus, the lower branch of the decision tree (Fig. 3) classifies the covers of natural soil and vegetation. It is processed in a simple way by defining four pure classes of *trees*, *low vegetation*, *grass*, and *bare soil and rocks*, and four classes mixing the different kinds of vegetation. For more consistency, they are associated with the most similar classes found in the global classification used by GEM and MC2, i.e., *crops*, *short grass and forbs*, *mixed shrubs*, or *mixed forests*.

The upper branch of the decision tree is dedicated to the identification of the urban classes. The first test applies a threshold of 80% to the fraction of *asphalt*, which is most of the time associated with roads or parking lots. Since the large dark roofs have sometimes the same signature as asphalt, a complementary test for *built2* is done. If *built2* is less than 20%, the aggregated pixels are identified as *roads and parkings*. Otherwise, the pixels correspond to districts mostly composed of buildings.

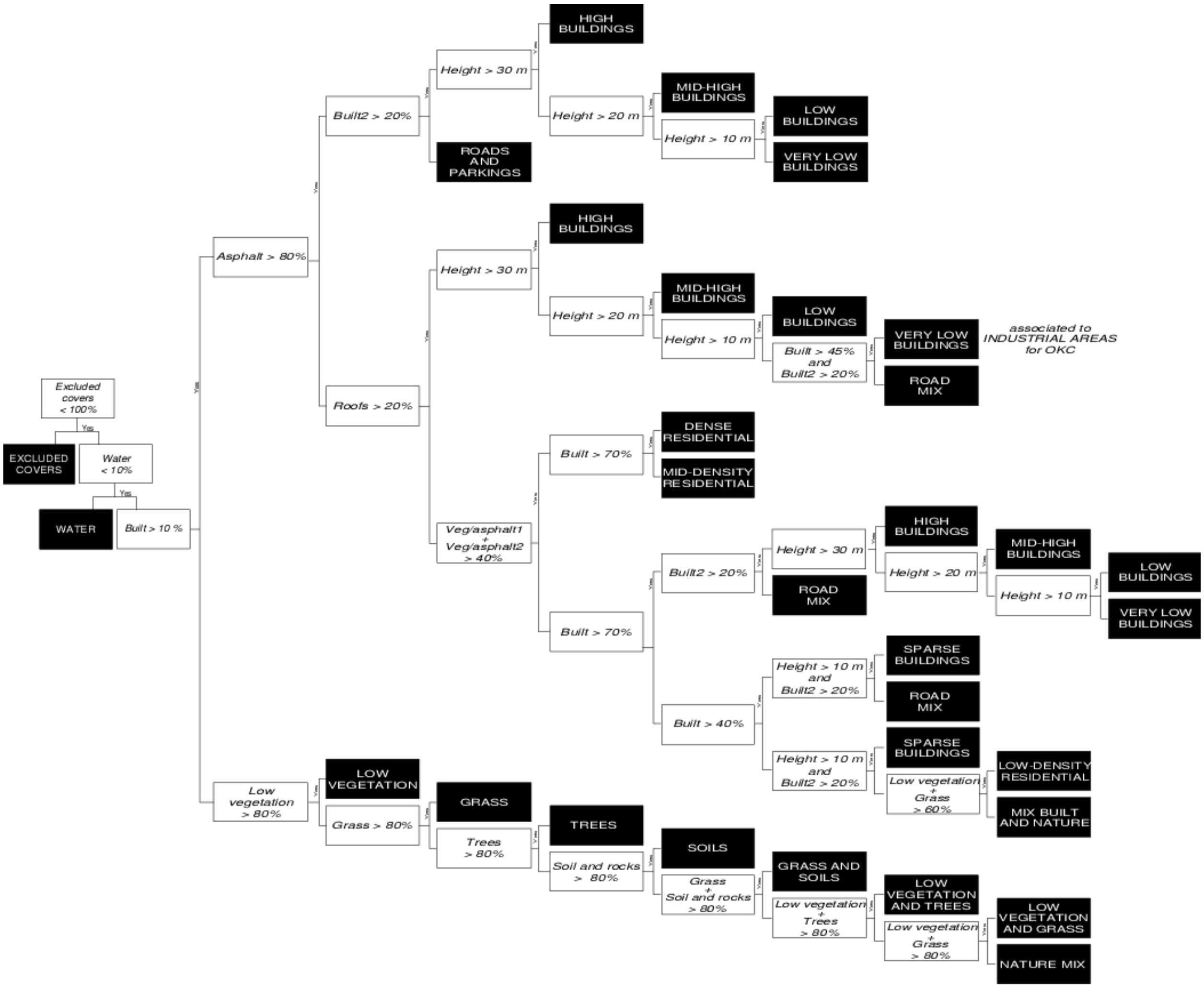

Fig. 3 - Decision tree model applied to the 60-m simple element databases by using the fourteen selection criteria.

Three successive thresholds of 30, 20 and 10 m, applied to *height*, lead to three urban classes called *high buildings*, *mid-high buildings* and *low buildings*, respectively. They are characterized by urban areas of large commercial buildings or multi-family housings.

For fractions of *asphalt* lower than 80%, the decision tree tests next the fraction of *roofs*, which is very useful to identify large buildings. Consequently, when the fraction is greater than 20%, the aggregated pixels are associated with the urban classes of buildings already defined previously. When *height* is lower than 10 m, the aggregated pixels are classified as *very low buildings* (i.e., one or two-storey buildings) if *built* and *built2* are greater than 45% and 20%, respectively. Otherwise, they are grouped in the class *road mix*.

The single elements *vegetation and asphalt mixing 1* and *vegetation and asphalt mixing 2* are often associated and quite representative of the residential areas. They are jointly used in the decision tree to identify the residential districts. According to the built-up density, they are subdivided in two classes *dense residential* and *mid-density residential*.

The last branch of the decision tree gathers pixels for which the roofs are not successfully identified by the unsupervised classification. In this case, the successive tests are based on the fractions of *built* and *built2* and on *height*. For large built-up and building densities, they lead to the four classes of buildings. For lower built-up and building densities, the aggregated pixels are classified as *sparse buildings* or *low-density residential* depending on the building height.

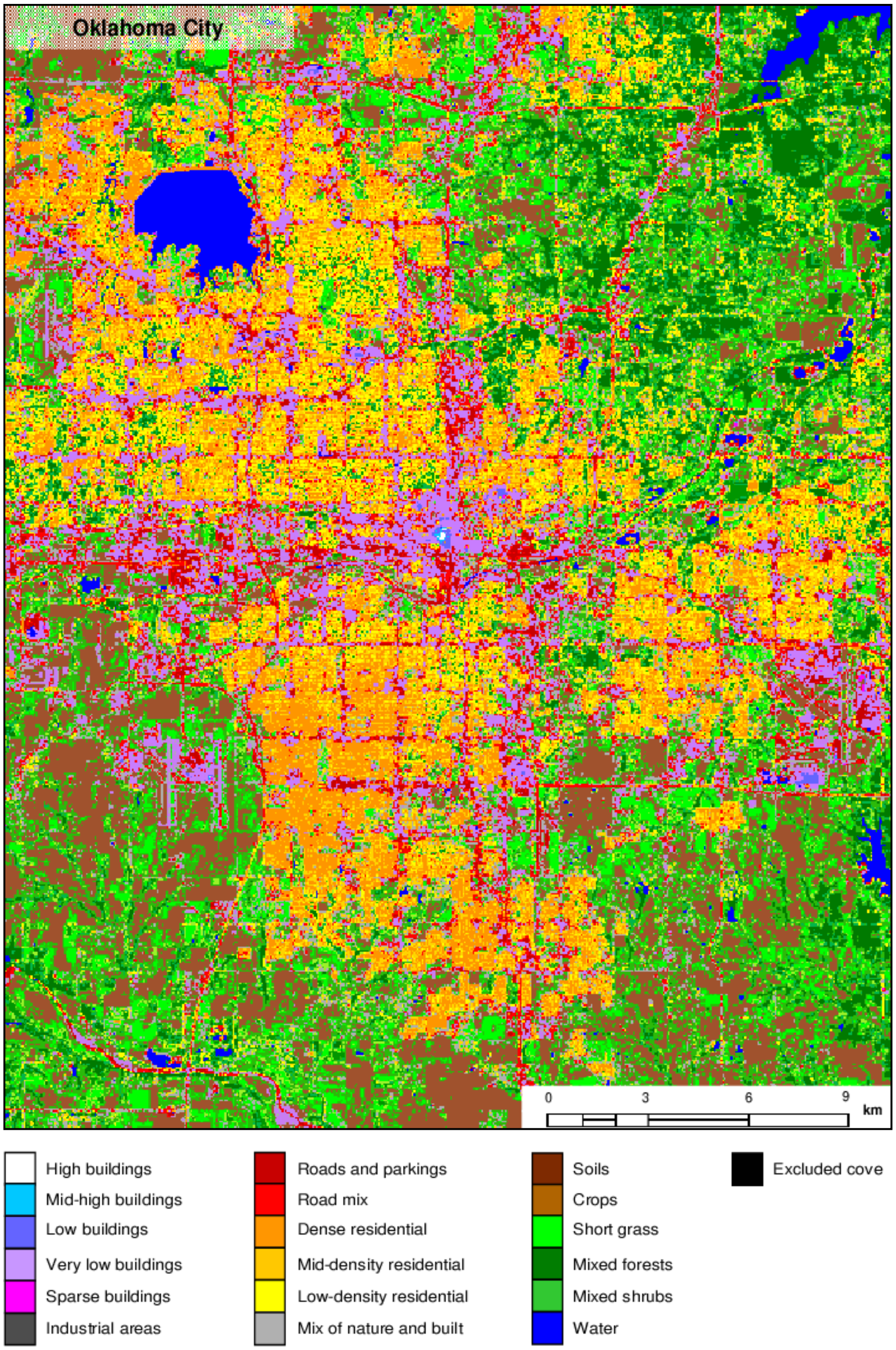

Fig. 4 - Final LULC 60-m classification of OKC including five classes of natural covers already defined in the global classification, and twelve new urban classes.

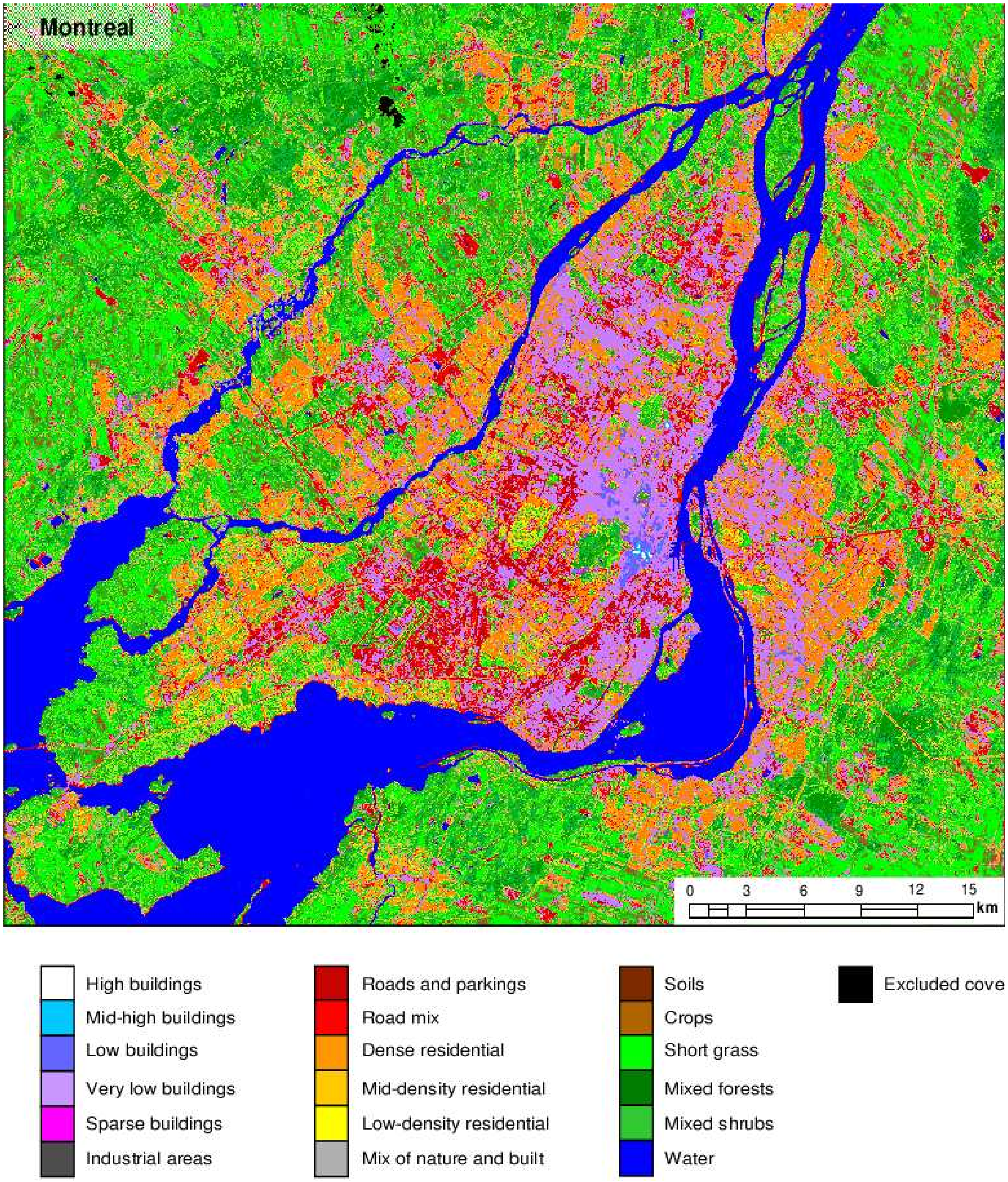

Fig. 5 - Final LULC 60-m classification of MTL including five classes of natural covers already defined in the global classification, and twelve new urban classes.

In summary, five classes of natural covers (including one class of water bodies) and twelve new classes of urban areas are identified in the final version of the classification. The 60-m resolution classifications for OKC and MTL are shown in Figs. 4 and 5, respectively. The three first urban classes essentially correspond to the city cores of OKC and MTL, and present a similar spatial distribution, i.e., the *high buildings* class is directly associated with the city business centre mostly composed of skyscrapers and is surrounded first by *mid-high buildings*, and then by *low buildings*. However, the rest of the classification is very different for the two cities. In the case of OKC, industrial and commercial areas (associated with *mid-high buildings*, *very-low buildings* and *industrial areas* classes) are concentrated along the main roads, whereas elsewhere the residential districts largely dominate. MTL is much more densely built up. A large part of the east island and of the southern shore is occupied by *low buildings* corresponding to contiguous multi-family housings. Elsewhere, the *dense residential* districts are predominant, whereas the classes *mid-density residential* and *low-density residential* are negligible.

4. CONCLUSION

Because of choices that were done for the source data and for the semi-automatic processing, the general methodology presented in this study makes the production of urban LULC classifications possible and practical for any North American city. The originality of this approach, aimed at mesoscale atmospheric modelling, rests on the joint analysis of satellite imagery and DEMs in order to take both surface properties and three-dimensional characteristics of the urban canopy into account.

The application of a decision tree model results in the identification of twelve urban LULC classes and eight non-urban or natural cover classes. Even though each type of data has some limitations, their coupling in the decision tree makes it possible to improve the quality of the analysis results and to better characterize the urban covers. The methodology is automated as much as possible but some manual pre-processing is still necessary.

It should be noted that the present decision tree is well adapted for North American cities, i.e., the combinations of classification criteria are fitted to the particularities of these cities. In the future, other versions of the decision tree could be developed to better fit other types of cities, such as European cities.

Acknowledgments

The authors are grateful to Jean-Philippe Gauthier and Louise Blais for their assistance. This work is funded by the Chemical, Biological, Radiological and Nuclear (CBRN) Research and Technology Initiative (CRTI) (Project #02-0093RD).

References

Danko, D. M., 1992: The digital chart of the world. *GeoInfo Systems*, **2**, 29--36.

Lemonsu, A., Leroux, A., Bélair, S., Trudel, S., and Mailhot, J., 2005: A general methodology of urban cover classification for atmospheric modeling. *Remote Sensing of Environment*. Submitted.

Loveland, T.R., Reed, B.C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., and Merchant, J. W., 2000: Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *Int. J. Remote Sensing*, **21**(6--7), 1303--1330.

Masson, V., 2000: A physically-based scheme for the urban energy budget in atmospheric models. *Boundary-Layer Meteorol.*, **94**, 357--397.