Alternative Geomatic Approaches to Develop Spatially Disaggregate Data on Exurban Development

Namrata Shrestha and Tenley Conway*
Department of Geography, University of Toronto at Mississauga

Abstract

Exurban development is one of the leading anthropogenic causes of land conversion in North America. The unprecedented pace of such growth, attributed mainly to improvements in information technology and changing demographics, is impacting previously undisturbed areas including those with high conservation value. Though the significance of exurban development and its potential ecological consequences are recognized by many in the literature, there are limited studies pertaining to it. One of the major hindrances to such studies is the lack of spatially disaggregate data on exurban development that is indispensable for ecological inferences. In this study, we compared three specific geomatic approaches that have high potential to map exurban developments in our study area (Central Ontario, Canada). First, a remote sensing approach using relatively high resolution (10m) multispectral SPOT 5 imagery was employed to extract exurban built areas. Second, a less direct approach of using the road network as the surrogate of exurban intensity in the landscape was explored. Third, we incorporated census housing data and road data in a dasymetric mapping approach to capture exurban development in the landscape. An accuracy assessment was completed based on digital property parcel data and air photos. It was found that the indirect methods did have significant correlation over the study area with high strength of correlation over smaller spatial extent. The direct methods also showed potential but issue of commission error was prevalent as expected.

Introduction

Human activities in the landscape often result in conversion or loss of land cover types and fragmentation of remaining land cover into smaller, more isolated elements, thus threatening the biological conservation initiatives. Among various forms of anthropogenic causes, residential developments and associated land use is the leading causes of species imperilment. Various techniques have been adopted to conserve biodiversity, recent focus has been to take conservation initiatives beyond public and protected lands into private lands that support high levels of biodiversity and face equally high levels of threat from human activities.

In past half a century there has been a rapid growth of so called "exurban development" in the private lands of North America and, perhaps, elsewhere. Exurban development is defined as a kind of settlement pattern spanning the landscape between contiguous urban development and rural countryside (Nelson, 1992). Quantitatively, housing density measures have been used such as per unit parcel lot size of 4 to 16 ha (Brown et. al., 2005; Hansen et. al., 2005), 0.68 to 16 ha (or 8) (Theobald, 2005), 0.2 to 2 ha (McCauley and Goetz, 2004) and so on. In addition there are other combination of sprawl measures that have been used in urban literature to define exurban development.

Though the significance of exurban development for biological conservation have been recognized by many, there is limited work on studies pertaining to its ecological consequences. One of the major hindrances has been the lack of spatially explicit data, which are crucial for ecological inferences. The need to identify the extent and spread of disturbances, natural or anthropogenic, in the landscape has been emphasized time and again in literature. Without spatially explicit information on exurban development any subsequent study to examine its pattern, process, or consequences will be extremely difficult if not impossible.

Goals and Objectives

The overall goal of my research is to examine multiple geomatic approaches to develop spatially disaggregate data on exurban development at large spatial extent. Specifically, I assessed three geomatic approaches that I have categorized into two groups;

1. Indirect method
a. Road density as a surrogate for development
b. Dasymetric mapping using census and road data

2. Direct method (Remote Sensing)
a. SPOT 5 Multispectral Imagery (10m res.) supervised classification
b. Normalized Difference Vegetation Index recoding

Study Area

The area of interest for my research is located in the Central Ontario biome transition zone between Mixed Wood Forests in the South and Boreals to the North of the Canadian Shield border. This landscape has been recently named as "the Land Between" (TLB) (Fig. 1). It is roughly 240 km E-W by 20-40 km N-S and stretches over 8 counties. The surficial geologic core is granite barrens and limestone plains. TLB lies withiin the popular cottage country area of Ontario within the commuting distance from major urban centeres such as Greater Toronto Area. Thus, it seem to provide an ideal setting for exurban development.

In this phase of the research I have focused on a smaller subset of the area of interest, covering the county of Peterborough (4379 sq. km), which shares most of the characteristics of TLB described above.

Data

Reference data:
1. Property Parcel Data with lot size information (AutoCAD) - 2005 - City of Peterborough
2. Orthophotos (MrSID) - 2002 - Southern Ontario Orthophoto Inventory, OMNR

Major data:
3. Ontario Road Network (.shp) - 2004 - OMNR
4. Census Data at Census Block Level - 2001 - Statistics Canada
5. SPOT 5 Multispectral 10m Resolution Imageries (TIFF) - 2005 - TerraEngine

Other:
6. Ontario Land cover dataset (.img) - 2001 - OMNR
7. Ontario Parks dataset (.shp) - 2002 - OMNR
8. Census Urban Area Boundary File (.shp) - 2001 - Statistics Canada

Fig. 1: Study Area Location

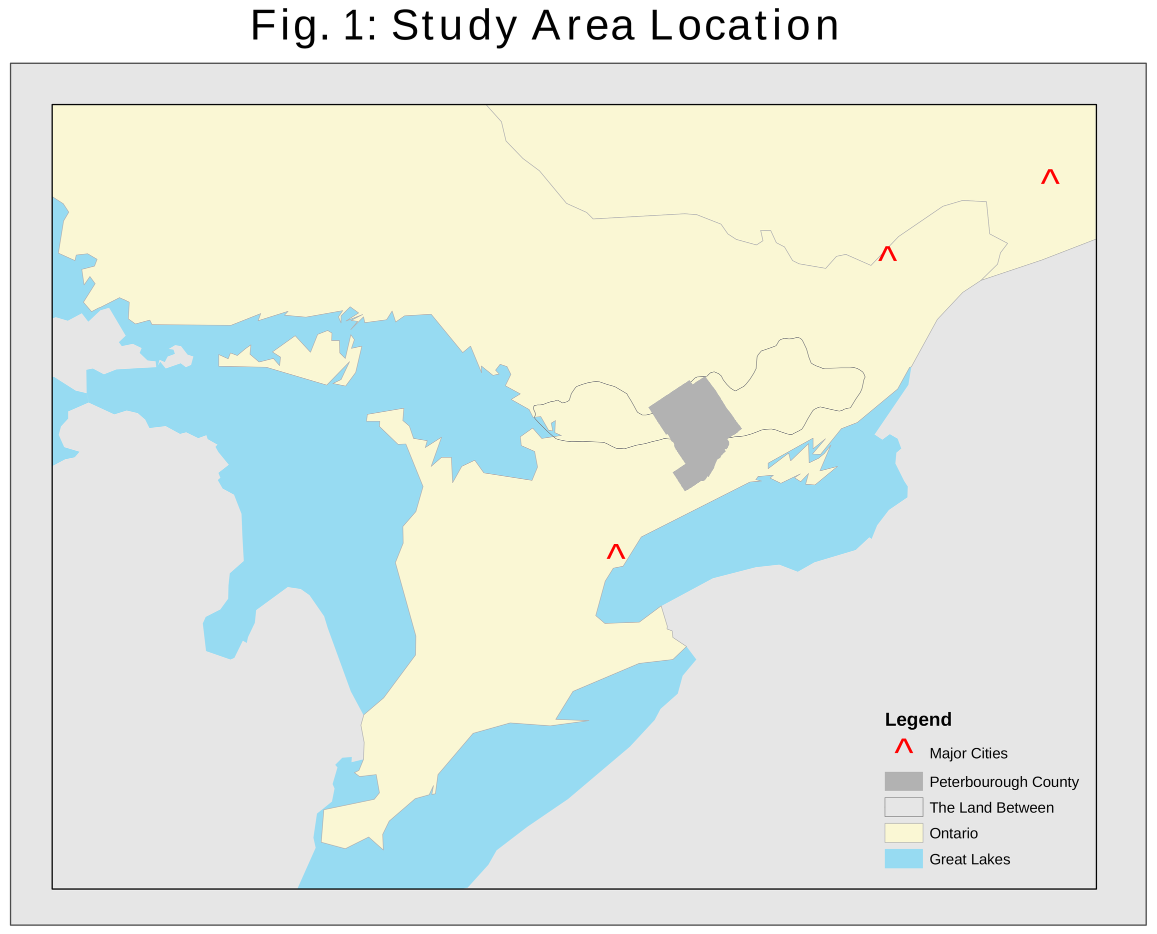
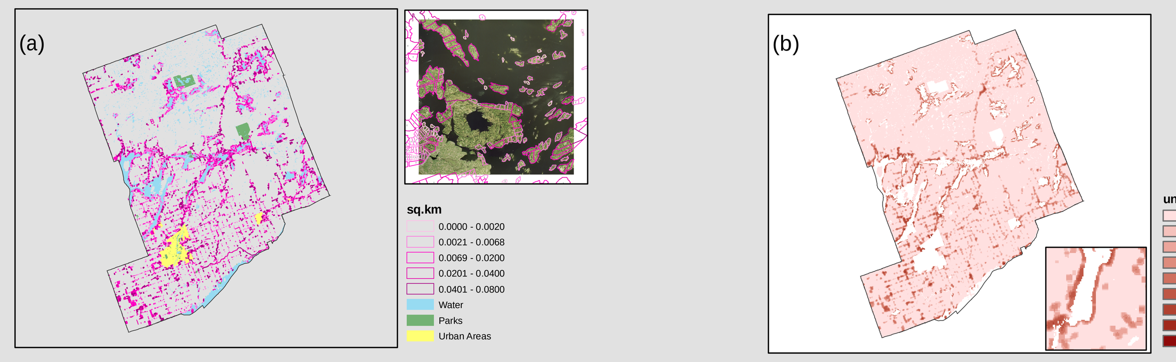
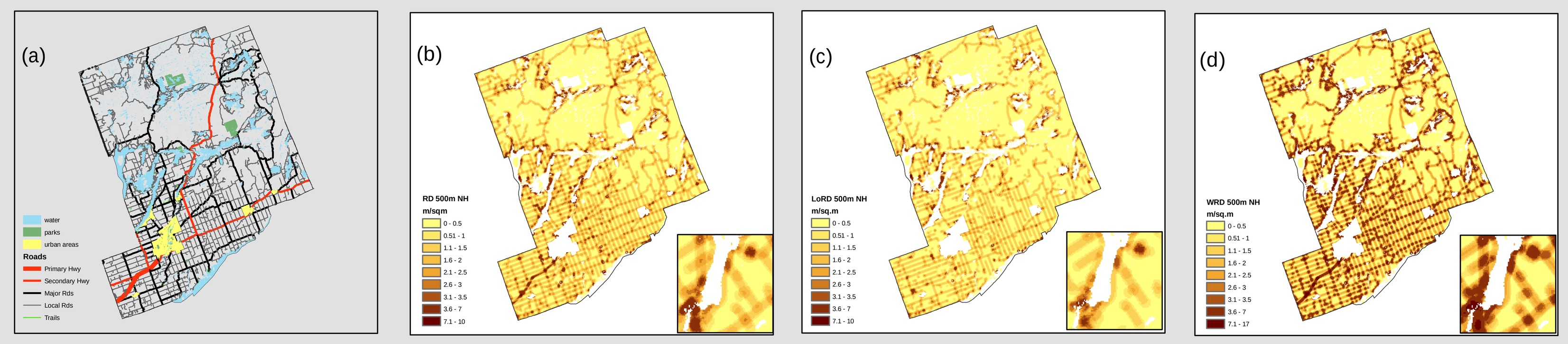


Fig. 2: Exurban parcel polygons & density map



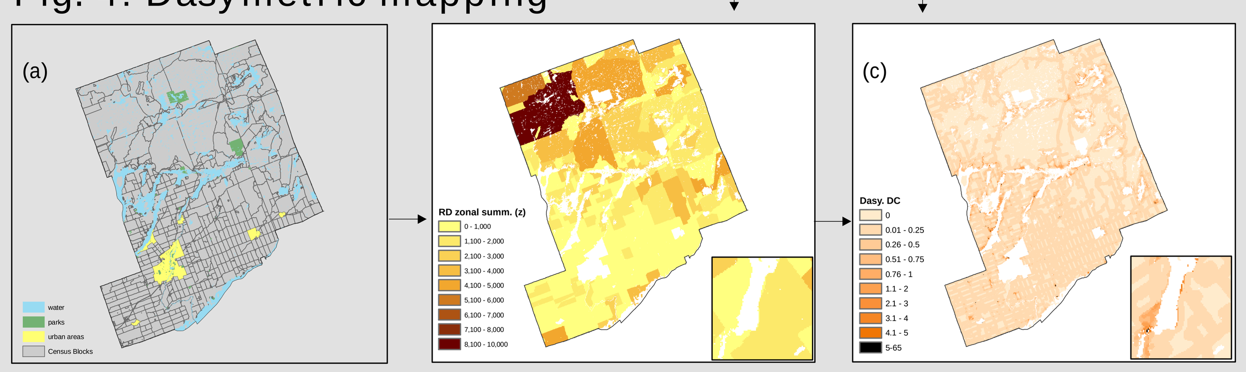
<= 8 ha property parcels (a) were used to calculate parcel density (500m NH) (b) to be used as reference data in indirect mapping. A mask created excluding all incorporated urban areas, public lands, and water were used to exclude the "non-exurban" areas from analysis.

Fig. 3: Road density rasters



Road vector data (a) was used to compute road densities with 500m NH and 1500m NH including all roads (b) local roads (c) weighted roads (d) for PB County. "Non-exurban" mask was applied to during all computation. Only 500m NH radii rasters are shown here.

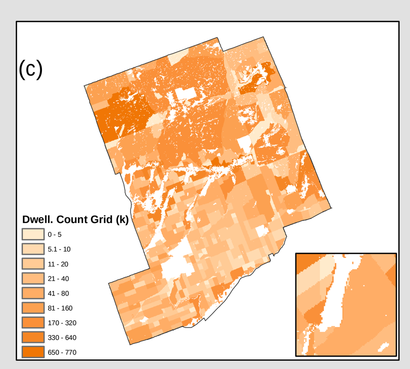
Fig. 4: Dasymetric mapping



2001 Census Block Boundary (CBB) Polygons were masked out for "non-exurban" areas (Fig. 4a).

Zonal summary of RD raster was computed with CB as zones and joined into CBB data along with aggregated dwelling counts.

CBB vector data with zonal road density and dwelling count fields was converted into 2 rasters; zonal road density raster, z (Fig. 4b) and dwelling count raster, k



Dwelling counts were redistributed preserving its pycnophylactic (mass preserving) property using simple map algebra, New Raster (y) = x/z * k (Fig. 4c).

500 m NH Radii	N	Pearson's Corr	Spearman's rho
Road Density*	153742	.451(**)	.558(**)
Local RD	153742	.034(**)	.075(**)
Weighted RD	153742	.051(**)	.100(**)
Dasymetric w RD**	153246	.018(**)	.092(**)
Dasymetric w LoRD	153246	.031(**)	.083(**)
Dasymetric w WRD	153246	.016(**)	.089(**)
1500m NH Radii			
Road Density	153616	.122(**)	.298(**)
Local RD	153616	.082(**)	.247(**)
Weighted RD	153616	.128(**)	.296(**)
Dasymetric w RD	153068	.037(**)	.236(**)
Dasymetric w LoRD	153068	.040(**)	.243(**)
Dasymetric w WRD	153068	.035(**)	.236(**)

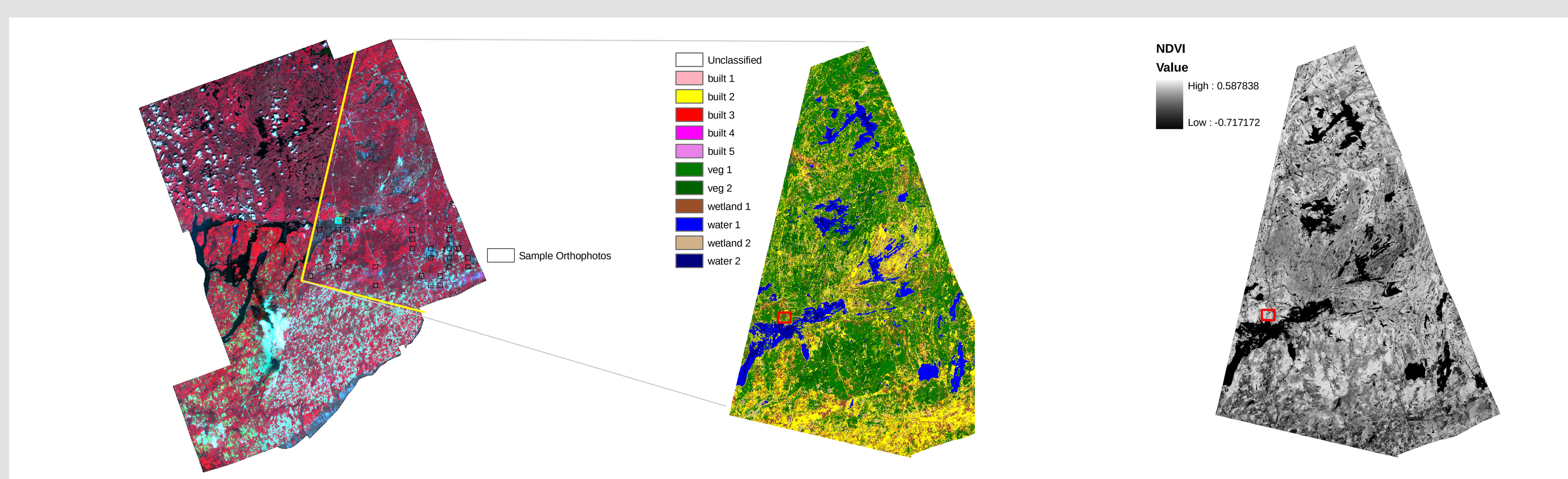
(**) Correlation is significant at the 0.01 level (2-tailed).

Table 1: Parametric and non-parametric correlation coefficients against parcel density data using 40% sample of PB County (excluding "non-exurban")

Geographic Extent		Road density*	Dasymetric**
Peterborough County		.558(**)	.092(**)
Ecodistricts	Shield (50%)	.516(**)	.052(**)
	Mid	.587(**)	.237(**)
	South	.542(**)	.097(**)
Sample Subsets	1	.575(**)	.575(**)
	2	.457(**)	.469(**)
	3	.688(**)	.614(**)
	4	.461(**)	.461(**)
	5	.530(**)	.495(**)
	6	.468(**)	.427(**)

Table 2: Non-parametric CC (Spearman's Rho) at multiple spatial extents for road density and dasymetric map (including all roads; 500mNH)

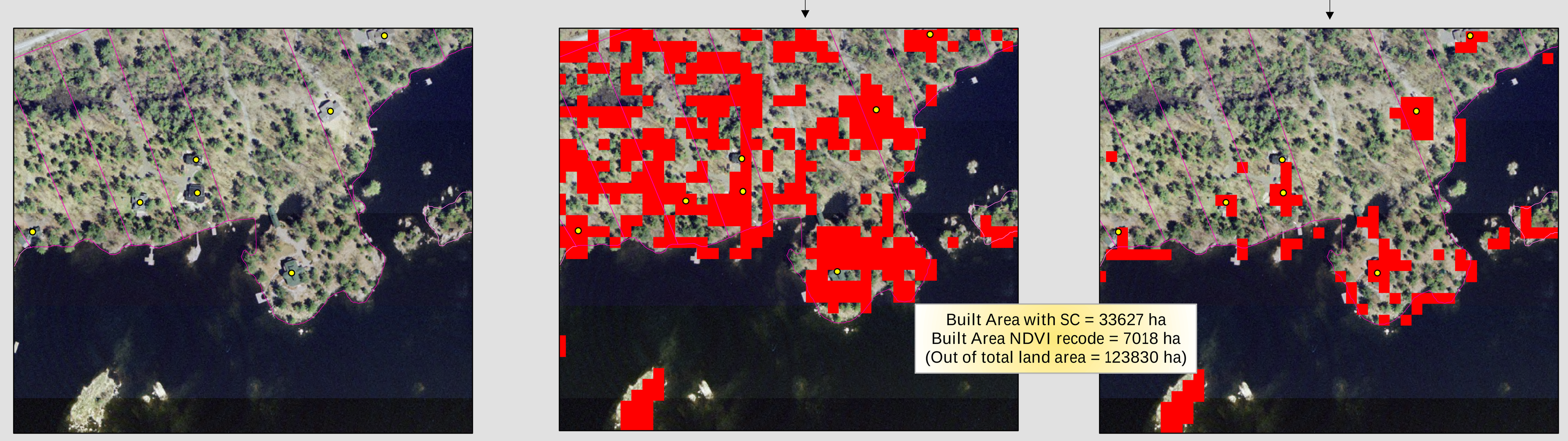
Fig. 5: SPOT 5 MS Classification



(a) Mosaic of five SPOT 5 MS 10m res. imageries with the outline of experimental scene, 617/260 subset.

(b) Supervised classification of sample scene subset

(c) Normalized Difference Vegetation Index (NDVI) of sample scene subset



(d) Orthophoto & exurban property parcels showing built areas in a sample area (red box)

(e) Binary recode of supervised classification overlaid on orthophoto and exurban parcels

(e) Binary recode of NDVI values (negative NDVI = built) with water bodies masked out

Methods

Reference data

Property Parcel Data (Fig. 2):
- All parcels <=16 ha parcels (based on literature) outside "non-exurban" areas (water, urban areas, parks) were extracted (N=35321).
- 5% sample (~50) from different parcel size ranges (<0.2, 0.2-2, 2-4, 4-8, 8-16 ha) were validated against orthophotos. Only those parcel size ranges that had high proportion of built status were included in further analysis as exurban parcels (<8 ha).

Orthophotos (Fig. 2 inset):
- 20 cm resolution orthophotos were the reference data during parcel validation (above) and training/testing during SPOT 5 MS classification.

1. Indirect Approach

a. Road density as the surrogate for development (Fig. 3)
- From ON Road Data (.shp) roads segments for area bit larger than PB county were extracted (to avoid boundary effect)
- Road Density (RD) rasters including (i) all roads (ii) local roads (iii) weighted roads (using speed) were computed using Spatial Analyst.
- Extent = PB county, Resolution = 100m, Neighbourhood Radii = 500m &1500m, Mask = "non-exurban" areas
- All rasters were exported to ASCII xyz format
- Correlation analysis was conducted against reference parcel density data using non-parametric test (Spearman's rho).

b. Dasymetric mapping using road and census data (Fig. 4)
- Aggregated 2001 Census Dwelling Counts (DC) data at Census Block (CB) level was redistributed preserving its pycnophylactic (mass preserving) property using road density as ancillary data.
- First, CB boundary (CBB) data was cleaned and joined with DC data.
- Second, RD rasters (x) were used to compute zonal RD, zones being the CB.The zonal averages were joined with CBB data with DC.
- Third, CBB vector data was converted into two separate rasters using (i) DC (k) and (ii) RD zonal summaries (z) as conversion attributes.
- Finally, DC was redistributed using simple map algebra, New Raster (y) = x/z * k
- Each dasymetric map was exported as ASCII xyz format and correlation analysis was conducted.

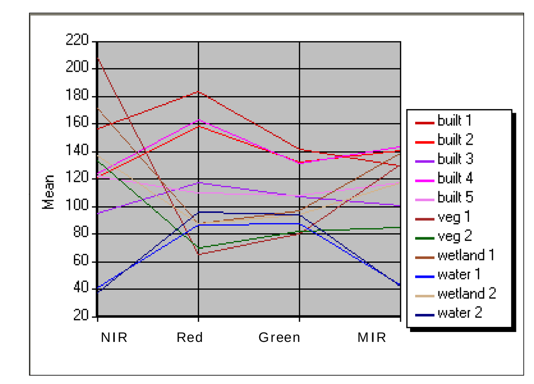
- Results are presented in Table 1.

2. Direct Approach

a. Supervised Classification of SPOT 5 MS (Fig. 5)
- Experimental sample scene = subset of 617/260 scene (within PB)
- Signature Collection (10 classes) = using combination of orthophotos, parcel data, visual interpretation
- Traditional per-pixel maximum likelihood classification algorithm
- Resulting classified image was recoded into binary classes; built vs. non-built

b. NDVI automated recoding
- Normalized difference vegetation index = (Near IR - Red) / (Near IR + Red)

- Healthy vegetation has high reflectance (thus high DN) in NIR and high absorption (thus lower DN) in red band. Non-vegetated areas show opposite trend as shown.
- Since, the target class is only built areas, the NDVI image was simply recoded such that positive NDVI = vegetated & negative NDVI = built.
- Water bodies were masked out with a water mask created from previous classification to avoid unnecessary confusion areas.



- Accuracy assessment for both methods were conducted using 399 built reference points collected from orthophotos and 100 points random sampling (equalized) of both classes

Preliminary Results

1. Indirect Approach

- Out of three types of road densities, RD that included all road types within smaller neighbourhood radius (500m) showed significant and strong positive correlation with parcel density data (Table 1).

- Though all RD and dasymetric maps for the study area showed significant correlation with parcel density data, the strength of correlation was weak.

- However, when correlation analysis was conducted over smaller extents (major ecodistricts and random subsets within the study area), the strength improved substantially, especially for those subsets with large water bodies north of the shield boundary.

2. Direct Approach

Accuracy Assessment of Supervised Classification and NDVI Binary Recode

Using 399 Reference Built Points

	Ref.	Supervised Classification				NDVI Recoding			
		Classifi	No.Corr	PA(%)	UA(%)	Classif	No.Corr	PA(%)	UA(%)
Built	399	321	321	80.45	100	130	130	32.58	100
NonBuilt	0	78	0	---	---	269	0	---	---
Totals	399	399	321			399	130		
OA (%)		**80.45**				**32.58**			

SC had higher Producer's Accuracy (low omission error) than for NDVI recode

But, here AA has been affected by the fact that the reference points were created through manual digitizing of 2D buildings as points from orthophotos.

Using 100 Random Points (Equalized) in Classified Image

	Supervised Classification					NDVI Recoding				
	Ref.	Classfi	No.Corr	PA(%)	UA(%)	Ref.	Classifi	No.Corr	PA(%)	UA(%)
Built	11	50	9	81.82	18	14	50	14	100	28
NonBuilt	89	50	48	53.93	96	86	50	50	58.14	100
Totals	100	100	57			100	100	130		
OA (%)	**57**					**68**				

- SC built class has lower User's Accuracy (higher commission error or error of inclusion). Since far more pixels were being classified as built by SC, its PA was higher in previous AA.
- NDVI recode built class = Higher PA and UA than SC.
- UA is low in classification results (as expected) mainly because of the large presence of barren lands (bare rocks and fallow fields).

Concluding Discussion

- Indirect methods' applicability is spatially dependent. In some areas the road network is not indicative of residential developments such as in southern region of the study area (where historically roads were laid in grid pattern regardless of level of development). However, in mid region and in areas around large lakes, roads are where the residential developments are so these indirect methods seem to work well.

- The road density computed using a smaller neighbourhood radii including all roads seem to reflect the built areas better than using larger neighbourhood radii, and including only local or weighted roads.

- The direct methods, both SC and NDVI binary recoding, that used only spectral attribute of medium resolution SPOT data, overestimated the exurban built areas. This was not surprising since there was a large presence of barren lands (bare rocks and fallow fields) in the study area.

- NDVI binary recoding, which is simpler than SC, showed higher PA and UA indicating better performance. Since UA was still below acceptable level further processing is needed before its result can be used as exurban built locations.

- I plan to use ancillary data captures structural and contextural difference between uninhabited bare areas and inhabited built areas (possibly using proximity measures from roads and large lakes).