Entropy Change as Influenced by Anthropogenic Impact on a Boreal Land Cover – A Case Study

J. Kumari¹*, A. Govind¹ and A. Govind²

¹Department of Geography, University of Toronto, #5047 Sydney Smith Hall, George Street, Toronto, ON M5S3G3, Canada
²Department of Geography, University of Saskatchewan, 9 Campus Drive, Saskatoon, SK S7N 5A5, Canada

ABSTRACT. Boreal forests are important terrestrial carbon sinks and hence routine monitoring for its vegetation dynamics is imperative. Remote sensing has proved to be the best option for detecting the land use and land cover changes at larger spatial scales, but it is often subjected to misinterpretation due to a variety of reasons such as sensor characteristics and heterogeneity of land cover. Landscape heterogeneity can be quantified based on the spectral heterogeneity. We hypothesize that this could be explained with the help of an entropy parameter, based on vegetation related spectral characteristics. This study demonstrates the spatio-temporal dynamics of landuse in a boreal landscape and the change in entropy as a result of increased heterogeneity in the upper Ottawa River basin, a region which faced dramatic landscape dynamics due to hydroelectric projects and other human activities. NDVI based strategy of land cover classification and the derivation of vegetation vigourosity, quantified the decrease in vegetation in the landscape and clarified that the urban areas have actually increased and the eclipsing effect created by slight increase in vegetation in the urban areas caused errors in landcover classification. Borrowing the idea from quantitative ecology, entropy based quantification of the vegetation diversity through the spectral signatures, specific to vegetation was developed by computing the Shannon’s entropy using the NDVI for two periods. The entropy has increased by a factor of 2.2 over the decadal period whereas there has been a general decrease in the vegetation along with a parallel increase in waterbodies. We hence conclude that the heterogeneity of the landscape has drastically increased. This boils down to the fact that the probability of getting vegetated surfaces has decreased in the landscape owing to anthropogenic influence and hence the measure of entropy will be used to augment our understanding about the landscape dynamics when studied from satellite platforms.

Keywords: Remote sensing, land use and land cover change, Shannon Entropy, vegetation vigor

1. Introduction

Boreal forests share a major portion of the terrestrial carbon sinks and much speculation has been attributed to the carbon sequestration sustainability of these ecosystems. Human alterations (positive as well as negative) on an ecosystem may perhaps not be accurately monitored, because it might have antagonistic or synergistic effects on the manner where those alterations are sensed from a higher spatial scale, e.g. landscape or regional. The continued logging of forests and increasing recreational activity has raised concern for many large and relatively pristine areas in the Canadian Boreal forests (Quinby et al., 2002). Such long term trends would lead to irreparable consequences because of fragmentation of forested landscape or because of the anthropogenic influences like dams and hydroelectric projects prevalent in these landscapes and the direct and indirect feedback effects which follow. Optical remote sensing sensors can discriminate spectral signatures from vegetated landscape and thereby give an idea about the heterogeneity of the land surface. Deforestation and Reforestation are the two contrasting forces, which could alter the spectral characteristics sensed by a satellite. This is of higher concern when opposing conditions such as optimal deforestation or increased reforestation on one hand and rapid urbanization on the other, occur in parallel. Spectral indices which are used to compress the information contained in several bands may not be representative of the real land surface characteristics because of the coarse nature of the remote sensor with spatio-spectral characteristics which are insufficient to quantify the heterogeneity at a landscape or watershed scale from a satellite bound platforms. The high spectral resolution sensors are an alternative, but its practical utility at large spatial scales warrants continued use of medium resolution sensors and account for a correction factor for the errors incurred. Vegetation indices are particular combination of spectral responses in different wavebands, which emphasize a particular feature of the land surface. There are several ratio vegetation indices which are derived from red and near infrared reflectance (Govind et al., 2005), which take the advantage of the negative correlation between the R and NIR, by chlorophyll absorption and by the canopy architecture of the leaf water content respectively (Galvao et al., 1999). Normalised difference vegetation index (NDVI) (Rouse et al., 1973) band ratio has been used indirectly to estimate photosynthetic capacity and net primary productivity (Field et al., 1983; Goward et al., 1985; Sellers, 1987). NDVI is related to green biomass and

* Corresponding author: jyothi.kumari@utoronto.ca
have been used to indirectly estimate photosynthetic capacity and net primary productivity (Goward et al., 1985; Sellers, 1987; Field et al., 1993). For broadband sensors such as Landsat Thematic Mapper, NDVI is still the most efficient general-purpose index for all conditions, particularly for qualitative study (Gibson and Power, 2000). This study demonstrates a simple and novel approach to quantify the heterogeneity of the vegetation of a landscape retrieved by satellite platforms using Total Information Content (Shannon Entropy) in a vegetation index, NDVI. The study is based on entropy theory, where Shannon’s entropy can be used to measure the degree of spatial concentration and dispersion exhibited by any geographical variable (Theil, 1967; Thomas, 1981). This could augment our understanding about the landscape vegetation status in addition to its heterogeneity. In ecology, this approach is widely used as a diversity index that was intended to measure the biodiversity of an ecosystem (Magurran, 1988; Roth et al., 1994; Rosenzweig, 1995; Begon et al., 1996; Beals et al., 1998; Gibbs et al., 1998). We applied this analogy to landscape level spectral signatures retrieved from a medium resolution sensor. This is thought to be a superior approach than the traditional statistical measures of dispersion, such as the standard deviation or coefficient of variation because this approach gives more information on the probability of a particular vegetation class to exist or not to exist and as such can be used as a parameter for correcting the errors involved in simulation models which are to characterize the landscape primary productivity by remote sensing approaches.

**Shannon’s Entropy:** The entropy expresses the expected information content or uncertainty of a probability distribution. Let \( E_i \) stand for an event or a geophysical variable (here NDVI) and \( p_i \) for the probability of event \( E_i \) to occur. Let there be \( n \) events \( E_1, E_2, \ldots, E_n \) with probabilities \( p_1, p_2, \ldots, p_n \) adding up to 1. Since the occurrence of events with smaller probability level yields more information (since these are least expected), a measure of information “\( h \)” should be a decreasing function of \( p_i \). Shannon (1948) proposed a logarithmic function to express information \( h(p_i) \):

\[
h(p_i) = \log_2 \left( \frac{1}{p_i} \right)
\]

(1)

It decreases from infinity to 0 for \( p_i \) ranging from 0 to 1. The function reflects the idea that the lower the probability of an event to occur, the higher the amount of information of a message stating that the event occurred.

From the \( n \) number of information values \( h(p_i) \), the expected information content of a probability distribution, called entropy, is derived by weighing the information values \( h(p_i) \) by their respective probabilities:

\[
H = \sum_{i=1}^{n} p_i \log_2 \left( \frac{1}{p_i} \right)
\]

(2)

where \( E_i \) is any variable, here NDVI of the \( i^{th} \) pixel.

**Study Site:** The study site is located at the border between the Canadian provinces of Ontario and Québec, which is on the Ottawa River basin, which lies approximately 1,130 kilometres (km), from its origins at Lake Capitnitchigama in Québec to the confluence at the St. Lawrence River (Legget, 1975), and has a vertical descent over its course of 365 metres (Telmer, 1996). It forms a natural provincial border between Ontario and Québec for approximately 580 km from Lake Temiscaming to Carillon (Chapman and Putnam, 1984; Haxton and Chubbuck, 2002). Such a river system, however, has undergone massive changes as both Ontario and Quebec have utilized the waterways for hydro generation. A complex system of dams, generating stations, and reservoirs was developed on both sides of the Ontario-Quebec border. A number of recent studies (Quinby et al., 2002) have revealed that severe biodiversity erosion has been occurring in this area. The studies conducted by Quinby et al. (2002) pertained mostly to the Temagami Management Unit (TMU), the Western part of the Lake Temiskaming, which were reported to have several environmental problems, like biodiversity erosion (Quinby, 1999; Quinby, 2000). It is quite probable that the Eastern section of this landscape also face similar biodiversity erosion, the intensity of which is yet to be quantified, since the landscape is continuous (Figure 1). Unfortunately, few studies have been undertaken to investigate the changes in land heterogeneity in this region, which very well exemplifies a boreal landscape. Although studies on landscape scale vegetation changes and carbon dynamics from a remote sensing frame work are very common, little importance have been given to the changes in the landscape heterogeneity, which could leads to unrealistic conclusions even though spectral indices based approaches are used. Only a few works has been done with this approach for monitoring land use land cover heterogeneity for natural resource management, in spite of its versatility.

**Satellite Data and Image Processing:** Radiometrically corrected Landsat TM data for 1991 and Enhanced Thematic Mapper data for 2002 was used in this study. This temporal span represents a decade, which the authors believe is reasonable to assess a major landscape dynamics that is pertaining to vegetation. Image to image geometric correction was carried out to ensure that the pixels in 1991 and 2002 imagery accurately match with each other. 1991 Landsat TM 28.5 m optical data (3 bands) were resampled to 30 m to match with the 2002 ETM data. In order to exploit the maximum information from all the bands (2, 3, 4, and 5) principal component analysis was carried out since first three principal components typically contain 98% of the variance in the data. A color composite (FCC) prepared by projecting the first principal component (PC1) in Red, PC2 in Green and PC3 in Blue thus contains most of the variance (Gibson and Power, 2000). Land cover classification and change detection was done subsequently.
using the first three PCs. The change detection was done as percentage area change (+ve or -ve) in each land cover class. The quantification of the vegetation over the landscape was undertaken with the Normalised Difference Vegetation Index NDVI (Rouse et al., 1973). A higher value is expected for the denser vegetation and a lower value is for a stressed, less dense or a deforested land surface. For a comprehensive understanding of spatio-temporal changes of vegetation for the landscape as a whole, NDVI was classified to 5 classes, which were named as “Very Good”, “Good”, “Moderate”, “Poor” and “Very Poor”. The pixel population in each NDVI class was used to get the area in that class in hectares. The arbitrary weightings were assigned viz. 10 for Very Poor, 20 for Poor, 30 for Moderate, 40 for Good and 50 for Very Good and a Composite Vegetation Index was computed, which turns into the summation of \(A_i \times W_i\) where \(A\) is the area of an NDVI class \(i\) and \(W\) is the weight assigned to that class. The index was computed for two years and their ratios were called Vigorosity \(\beta\), which assumed that a value above one means increased vegetation in 2002 or vice versa.

\[
\mu = \sum_{i=1}^{n} \left[ (A_1 \times W_1) + (A_2 \times W_2) + (A_3 \times W_3) + \ldots \ldots \ldots (A_n \times W_n) \right] \quad (4)
\]

where \(n\) is the Number of NDVI class. The \(\beta\) is given by:

\[
\text{Vigorosity} \quad \beta = \frac{\mu_{2002}}{\mu_{1991}} \quad (5)
\]

The entropy is implicitly a measure of probability of an entity, and as such, the value of \(H\) is non-negative. The minimum possible entropy value is zero corresponding to the case in which one event has 100% probability. Such an approach can hence be considered as a measure of uncertainty of that entity. In the present study, we quantified the heterogeneity of land cover using NDVI for two years, for a temporal span of almost a decade. We believe that this simple and novel approach is consistent with the diversity quantification strategy.

**Notes:** Lake Timinscaming (46°43’N, 79°05’W) separating Ontario and Quebec.

**Figure 1.** Location of the study site (Courtesy: http://www.algonquinnation.ca).
adopted by ecologists and hence it is conceptually a straightforward application. The algorithms were developed for the computation of a $p \times \log(1/p)$ imagery for a particular year, and the global summation of the raster values gave Shannon’s entropy of that year.

$$p_i = \frac{\text{NDVI}_i}{\sum_{i=1}^{n} \text{NDVI}_i}$$  \hspace{1cm} (6)

2. Results and Discussion

It is seen from the Figure 2 that there has been a slight increase in “Waterbodies” in the landscape from 1991 to 2002 (5020 ha). This may be due to reasons like creation of more wetlands/inundated areas, rapid snow ablation due to the polar amplification effect of the global warming in the recent years. This region also has tremendous harnessing of the water resources with the series of dams and the hydroelectric projects. According to a report prepared by the Algonquinn first nation secretariat (WESA, 2002), due to the construction of the dams and reservoirs, water levels have increased, lands have been flooded, erosion has occurred, water flow has been reversed, and some water courses have decreased, which justifies the hydroecological transition as ascertained by remote sensing.

The accuracy assessment of the classification showed that an overall accuracy of 78.85 and 84.23% were achieved in 1991 and 2002, respectively. The kappa statistics was fair in all the cases (land covers) with the highest for “Urban” (0.9067) followed by the “Water bodies” (0.8920) in 1991 while kappa was the highest in the “Deforested” (0.9156) that was followed by “Urban” (0.8884) in 2002. The lower kappa in 2002 may because a number of “Moraine” were misclassified as “Waterbodies” in the supervised classified imagery. “Waterbodies” showed the least accuracy in 2002 (0.7391) in which some of the pixels were misclassified as “Moraine”. In 1991, the least accuracy was given by “Moraine (vegetated)” where a number of pixels were misclassified as “Moraine”. The NDVI imagery also shows that there are some water bodies in the urban areas in and around Lake Timiskaming in 2002, which was absent in 1991. There has been a substantial increase in the “Moraine” (sparsely vegetated landcover) over the past years. We speculate that this may be because of the deforestation especially in the Northeastern parts of the study area (the right bank of Upper Ottawa River, Lac des Quinze, Winneway and around North of Lake Simard). The landcover, “Boreal forest” has decreased by 22,915 ha over the time period. This might be due to the increasing pressure on these landscapes from the lumbering industry or boreal fires, which have altered the pristine forest covers in various degrees. According to WESA (2002), the construction of the hydroelectric projects in this region has resulted in many changes to the physical environment. Consequently, the stakeholders have had to modify their cultures. The people were forced to relocate, which might hence provoke the increased spread of the anthropogenic impact. Land use for “Deforested/agriculture” has remained almost the same with respect to the magnitude of the area, but regionally wide variability has been noticed in terms of its spatial distribution. In and around the Lake Timiskaming, this land cover class has increased substantially while in northern and central zones of the study area, it was found to be lesser than the 1991 situation. The “Urban” areas have declined in the classified imagery. However, this may come as a surprise due to an eclipsing effect incurred due to the increased afforestation in urban areas, which is confirmed from the NDVI imagery. According to Haxton and Chubbuck (2002), this region has several hydroelectric projects which have directly or indirectly affected the wildlife and vegetation over the years. The classified NDVI imagery for the two years

![Image of land use and land cover in 1991 and 2002.](image-url)
Figure 3. NDVI Class distribution in 1991 and 2002.

Figure 4. Area in hectares for various landcovers: (a) NDVI, and (b) Classes for 1991 and 2002.
clearly reveals that over all vegetation has declined drastically. It is clear that severe deforestation has occurred on the Western side of Lake Timiskaming on the Ontario side (places like Cobalt), Western side of Lake Remigny, and Southeastern side of Lac Des Quinze. Like wise, Belle Valley, which is a part of the rich farm belt running north of the head of the Lake Timiskaming into Ontario and neighboring Quebec, along with Nedelec also shows a severe decline of vegetation vigor and increased “Water bodies”. This might probably be due to the increased farming activities and logging operations by private companies. According to a report by Vivre en ville (2002), there has been increased concern for the sustainability of the agro forestry communities through increased organizational activities in places like Nedelec. In an attempt to elucidate the trends in boreal forest fire cycle, Bridge (2001) highlights some keys differences between the Ontario and the Quebec landscapes. Since 1912, the area in the northwestern Quebec close to the Ontario border underwent extensive conversion to agriculture; while in Ontario, forestry remained the principle land use (McDermott, 1961). The unpublished data (Lefort et al., 1999) suggest that the logged area in Ontario had seven times less fires and half the area burned than the developing agricultural area in Quebec, which explains a higher deforestation trend in the Western side of Lake Timiskaming, which has been utilized by private companies as early as 1990, albeit severe deforestation in the Eastern side as well due to farming. The value of Vigorosity $\beta$ at landscape level was calculated as 0.8132 (using Equation 5), which clearly confirms that there has been a 20% reduction in vegetation in the landscape from 1991 to 2002. A keen perusal of the NDVI classes elicits more information. It can be understood from the Figure 5 that the land area with NDVI “V. Poor” and “Poor” has increased while “Medium”, “Good” and “V. Good” had decreased. The NDVI class, which has increased maximum, is “V. Poor” which corresponded roughly to the areas depicted as “Urban” in the classified imagery. The NDVI class, which has decreased maximum, is “Good”, which roughly corresponds to “Boreal forests” in classified imagery, revealing a severe deforestation problem. However, the NDVI class “V. Good” has decreased but to lesser magnitude as opposed to “Good”, revealing the fact that the dense vegetation was undisturbed over the time and only less dense vegetation was disturbed. Figure 6 shows the change in area in different NDVI classes from 1991 to 2002. The actual intensity of change can be better understood if we convert the area change into the percentage change (change from 1991 values).

It is hence clear that the NDVI class “V. Poor” was the most sensitive to change and had increased to nearly 420% from its 1991 value, revealing a rapid increase in poorly vegetated surfaces. The magnitude of change in “Good” and “V. Good” are comparatively low but significant (-30.5% and -33.5%, respectively). From this, the following conclusions could be drawn. The overall change in the landscape from 1991 to 2002 was primarily because of the significant increase in “V. Poor” NDVI class, i.e. the overall change on the land cover due to NDVI class “Good” and “V. Good” are almost similar and “Good” is less sensitive than “V. Good” clarifying that dense vegetation is declining more faster. But a contro-
versy still exists between the results of vigourosity and the acreage of urban areas in classified imagery. It is thought that the information derived from the radiance values of several bands using the first three principal components albeit a superior approach for land cover classification may not be an efficient method while mapping vegetation distribution and intensity. The NDVI based approaches utilising the Red and the Near Infrared bands are still the most promising procedure for conducting vegetation mapping. The vegetation vigourosity as well as the NDVI slicing approach reveals a clear anthropogenic influence on the landscape with time resulting in a complex landscape.

3. Shannon Entropy

The entropy calculated by the above mentioned procedure showed that there has been an increase of entropy by a factor of 2.2 from 1991 (17.92) to 2002 (30.50) (Figure 7). For a visual understanding of the distribution of the values, we displayed the entropy data in 256 grey shades (8 bit) although their inherent data type is 32 bit (visual analysis of which is extremely difficult because of the large number of possible grey shades). Although this indirect representation has a number of abstractions, it clearly reveals the heterogeneity in entropy between the two years. The entropy scatter plot shows that (Figure 6) the population of pixels with higher grey scale is more and less spread in 1991 while it is very much spread in all the grey scale in 2002, clearly showing an increased heterogeneity in the landscape.

Since the value corresponding to the grey scale is much dependent on $p$ rather than log$(1/p)$, a higher NDVI cluster (as in 1991) with lower variability means a grey scale distribution more localized in the grey scale space than with a lower NDVI values but with high variability (as in 2002). Entropy is a measure of randomness. The increased entropy can be interpreted as the increase in the heterogeneity of the land use and land cover in 2002 compared to 1991. This may seem contrary to the kappa statistics of image classification radiance values. In the case of entropy, a higher value signifies spectral heterogeneity on the land cover on one hand, while probability of vegetation to be lower on the other. In the case of kappa statistics; a higher accuracy means clearly separable spectral heterogeneity (2002). Hence the two trends are complementary to each other. The total Information Content or Entropy (Shannon, 1948) has been used for quantifying the species diversity by ecologists as well as biologists. In remote sensing based approach for ecosystem studies, it is still new and we believe that this approach is promising. Sahoo and Bhavana-rayana (2001) computed the entropy based on Band Ratio, DN values, NDVI and NDWI and concluded that entropy derived from DN had different values and not feature specific and TIC (Total Information Content) was proved to be an efficient tool not only for characterizing the natural features like soils, water bodies and vegetation but also for delineating boundaries of soil mapping units. Gardi et al. (2002) found promising relationships between vegetation and landform units in order to obtain an integrated classification of landscape units using Shannon’s Entropy approach. Entropy-based approaches are generally followed for studying the urban sprawl (Anthony, 1999) where an entropy increase corresponded to increase urban sprawl. Another information which we can derive from the Entropy imagery is that the probability of getting an urban land cover (with lesser NDVI) and probability of getting forest land cover is more in 1991 (urban areas getting lower grey scale and forests getting higher grey scale) but in 2002, the reverse happens, i.e. the probability of getting urban land cover is more than forest (urban areas getting higher grey scale and forests getting lower grey scale) and in both the years water bodies showed low probability owing to the lower values of NDVI associated with water. This is because of the strong influence of the probability than the reciprocal of the probability, which is governed by the logarithmic function in the Shannon’s Equation. The entropy is a description of randomness within a particular space of variability, hence a relative concept. The two-dimensional entropy space can be used to clearly differentiate various trends clearly revealing dynamics of vegetation at a landscape scale. The higher value of entropy ratio (2.2) over the vigourosity (0.81) suggests a higher sensitivity of this approach to define landscape complexities n/r heterogeneities.

4. Summary and Conclusions

A remote sensing based approach to understand the land use land cover changes was undertaken in the Ontario-Quebec region of Canada using multi-temporal Landsat data. The land cover classification showed an over all increase in “Moria” areas and a decrease of forests and “Urban” areas from 1991 to 2002. The NDVI aided vigourosity showed a value of 0.81, which signifies that the vegetation has decreased over the landscape by over 20%, even though land cover classification based approach showed a decrease in urban areas in 2002. We speculated and hypothesized that this is a misclassification owing to the increased complexity in the landscape due to increased spectral heterogeneity. Critical analyses of NDVI classes revealed that the dominant NDVI class to cause this change is “V. Poor”. The attempt to quantitatively describe the heterogeneity through using Shannon’s Entropy approach was successful showing an entropy increase of 2.2 in 2002. This increase in NDVI based entropy signifies a higher heterogeneity in 2002 than 1991 and implies the reduced probability of a landscape unit being vegetated in 2002 as opposed to 1991. We believe this is an important measure of diversity of spectral signatures quantifying the “intensity” term for a remotely sensed variable. So this approach could be used to quantify the anthropogenic influences when dealing with the multi-temporal data pertaining to a landscape for driving ecosystem models which has remote sensing derived inputs as the main component. This study, hence, demonstrates the anthropogenic impacts on vegetation trends and dynamics mediated through projects such as dam, hydroelectricity generation, farming and quantifying the heterogeneity it has created on the landscape. In a nutshell, deforestation and land use change occurring in boreal landscape has increased its heterogeneity in spite of its increased vegetation in the urban areas.
Figure 6. Entropy scatter plots in 1991 and 2002.

Figure 7. Shannon Entropy distributions in 1991 and 2002.
Acknowledgments. Many thanks to Late Dr. Ferenc Csillag, Prof., Dept of Geography, University of Toronto, whose ideas inspired the authors. The authors are also thankful to Global Land Cover Facility (http://glcf.umiacs.umd.edu/ for the Landsat data).

References


