An investigation of the ‘Eye of Quebec’
by means of PCA, NDVI and Tasseled Cap Transformations
<table>
<thead>
<tr>
<th>Index</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview</td>
<td>1</td>
</tr>
<tr>
<td>Study Area</td>
<td>2</td>
</tr>
<tr>
<td>PCA Discussion</td>
<td>3-6</td>
</tr>
<tr>
<td>NDVI Discussion</td>
<td>6-9</td>
</tr>
<tr>
<td>Tasseled Cap Discussion</td>
<td>10-12</td>
</tr>
<tr>
<td>Conclusion</td>
<td>13</td>
</tr>
<tr>
<td>References</td>
<td>14</td>
</tr>
<tr>
<td>Appendix I: PCA Report</td>
<td>15</td>
</tr>
<tr>
<td>Appendix II: Image Channel Listing</td>
<td>16</td>
</tr>
</tbody>
</table>
Overview

An archived Landsat 5 TM image was attained from the Global Land Cover Facility (XXX) and three image enhancement transformations were explored; Principal Components Analysis (PCA), Normalized Difference Vegetation Index (NDVI) and Tasseled Cap. These transformations serve to alter an image in order to extract unique information or enhance particular characteristics. PCI Geomatica v10 was utilized to accomplish these transformations – the EASI environment performed the work while FOCUS was used to visualize the output.

PCA is a technique that transforms the original remotely sensed dataset into a substantially smaller and easier to interpret set of uncorrelated variables that represent most of the information present in the original dataset (Robinson, J.). In this case, six Landsat 5 TM bands were in the original dataset and following the extraction of unique information from each band this dataset was represented by three bands. A large reduction in file size results as well as simultaneous viewing capabilities of the uniqueness portrayed in each band.

NDVI is a simple mathematical formula which calculates the amount of biomass present. The amount of chlorophyll in plants is an indicator of species and health, which the short wave infrared (band 4) and visible red light (band 3) wavelengths are sensitive to. NDVI ratios band 3 and 4 in such a way as to reduce illumination differences, shadows, atmospheric attenuation and topographic variations (Jensen, J., 2000). The end result is an image highlighting biomass, with clear delineations from water, soil and urban areas.

Tasseled Cap, when performed on Landsat TM imagery, produces three bands containing an indicator of brightness, greenness, and wetness, respectively. These indices are indicative of a features reflectance angles, the amount of biomass present and the extent of moisture, respectively.

Each transformation is valuable on its own but when used in conjunction a wealth of unique information is represented.
**Study Area**

The area investigated is located in the remote wilderness near central Quebec, Canada. Of particular interest in this Landsat 5 TM scene is a large circular lake (cca. 70 km diameter), Manicouagan Reservoir. The formation occurred 215 million years ago by the collision of a 5 km asteroid, this is the fifth largest impact crater known on earth. This devastating impact shattered the bedrock and melted the asteroid into what is now called Ile-Rene-Levasseur. Over time the bedrock was carried away by moving ice while leaving the harder island material intact (The Canadian Encyclopedia).

![Figure 1. Study area as seen by NASA WordWind.](image1)

![Figure 2. Study area seen by Landsat 5 TM true colour composite.](image2)

The central region of the image (white areas in Figure 2) is virtually void of tall vegetation - short shrubbery, grass fields, rock and barren earth constitute this region. The green areas are comprised mainly of old boreal forest (coniferous) interspersed with small deciduous tree stands. There is only one major road captured in this image, highway 358, which travels along the east side of Manicouagan Reservoir and continues in a northerly direction.
PCA Discussion

Principal Components Analysis is a mathematical transformation technique used to minimize spectral redundancy through the extraction of unique information ((1)Milne, T., 2008, ). There is a tendency for adjacent bands in a multiband dataset to be correlated to each other, in that only subtle variations in DN values occur for the same location. PCA serves to decorrelate this information around multidimensional orthogonal axes. More specifically, PCA can be viewed as a rotation of the existing axes to new positions in the space defined by the original variables. In this new rotation, there will be no correlation between the new variables defined by the rotation. The first new variable contains the maximum amount of variation, the second new variable contains the maximum amount of variation unexplained by the first while remaining orthogonal to the first, and so on until the last axis accounts for the last amount of variation (Robinson, J.). This rotation is based on the orthogonal eigenvectors of the covariance matrix generated from a sample of image data from the input channels, creating an output of new image channels, sometimes referred to as eigenchannels (PCI).

A PCA was performed and a report generated (see Appendix I). This report can be investigated to help understand how the principal components were calculated. First, we can see that all six original Landsat 5 TM bands were used as input to compute the PCs (outputted to channels 7, 8 and 9, corresponding respectively to eigenchannels 1, 2 and 3). The mean and deviation of DN values are also displayed for the bands contained in the original dataset, useful to gain a broad understanding of the spread of values across bands. A covariance matrix is generated from the original bands, this shows the extent to which individual bands vary with each other. More specifically, if the DN value of one band increases and the same pixels DN value also increases in another band then their covariance will be positive. As the covariance value approaches zero the variables are increasingly independent of each other. More importantly, eigenvalues are listed for each band. Eigenvalues represent the amount of total variance that is explained by each principal component ((1)Milne, T., 2008) and a this amount is conveniently displayed as a percentage for easier interpretation as well. From this table, we can see that principal component channel 1 (eigenchannel 1) comprises 93.59 % of the total variance of the complete original dataset. If eigenchannel 2 (containing 4.47 % of total variance) and 3 (1.41 % variance) are added to eigenchannel 1’s variance we reach a total accounted for variance of 99.37 %. Simply put, this means that 99.37 % of the unique information contained in the original dataset can be captured and portrayed using the first three eigenchannels. Eigenchannels 4 through 6 contain less than 1 % of the unexplained variance and should not be included in the final compiled PCA output RGB image as they offer very little
unique information, of which the majority is likely noise which would negatively impact the quality of the resultant image.

The ‘eigenvectors of covariance matrix’ indicates the amount of variance each band contributes to each eigenchannel. In this table the rows represent eigenchannels while the columns represent the input bands. The contribution can be calculated by squaring a given coefficient, for example, squaring the value corresponding to band 1 and PC 1 (cell 1,1) reveals that 60.08% of the variance shown by PC1 is contributed by band 1:

$$ (0.77512)^2 = 0.600811 = 60.08\% $$

If this calculation is performed for the whole of PC1, it is found that 82.81% of the variance loaded in this eigenchannel is derived from bands 1, 3 and 4.

This claim is supported since the areas of highest DN values (brightest) in the true colour composite, shown in Figure 3, are areas of barren earth, very short vegetation and exposed rock (see Figure 5). These features appear the brightest in the true colour image according to their highly reflective nature in the visible spectrum. As clearly seen in Figure 4 the brightest feature corresponds to these same areas. Principal component 1 is mainly derived from bands 1, 3 and 4 and as such the barren areas are highlighted. Band 3 is particularly effective in delineating bare soil, rock and urban areas. Band 3 is also sensitive to the red chlorophyll absorption band of healthy green vegetation and thus contributes some variance useful for discriminating vegetation type. A clear water delineation can also be seen in Figure 4, characteristic of band 1 which encompasses the peak transmittance of clear water. In all, PC1 is heavily influenced by the visible spectrum but also shows some variance as detected through near-infrared wavelengths.
Investigating PC2 shows that 89.04 % of the total variance expressed is contributed by band 1 (19.59 %), band 4 (10.74 %) and band 5 (58.71 %). This eigenchannel is heavily influenced by the mid-infrared band.

As seen in Figure 6 the areas of water are well delineated. This is characteristic of infrared wavelengths since water absorbs nearly all of the incident radiation at those portions of the electromagnetic spectrum. The DN values of water in PC2 exceed 240 which is in agreement with the high delineation achieved with infrared wavelengths. Band 1 also contributes some variance in vegetation since it captures the peak of chlorophylls blue absorption band of healthy green vegetation. Areas of barren earth and sparse vegetation (as described on the previous page) appear very dark in the PC2, but this colour is misleading. When the DN values of the rocky shoreline of the islands in Figure 8 are investigated they are found to be 50 ± 5. This delineation is mainly a result of band 5’s sensitivity to rocks and minerals. Vegetation is also noticeably brighter in PC2 than PC1, this is due to the contribution of band 4 which encompasses vegetations peak reflectance.
The main contributors of variance for PC3 are bands 1, 3, and 4, explaining 80.09% of the total variance. Although these are the same three main bands that also comprise PC1 they do so with different percentages. Specifically, band 1 is the highest contributor for PC1 while band 3 is the highest in PC2. A greater importance is also emphasized on band 4 in PC2 in comparison to PC1. So, we would expect to see a greater influence of the characteristics prevalent from band 3 and band 4 in PC2. This is indeed the case, Figure 9 shows vegetation much brighter (higher DN values) than as seen in Figure 4 (PC1). This is a direct result of the greater contribution of band 4 (and band 3 to an extent) which is ideal for vegetation discrimination.

![Figure 9](image)

**Figure 9.** Principal component 3.

**NDVI Discussion**

Normalized difference vegetation index is one of a list of many vegetation indices. A vegetation index is a measure that represents the amount and quality of vegetation in an area ((2)Milne, T., 2008). The specific algorithm used to derive the NDVI is a simple one and it is as follows:

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)}$$

The theory behind this calculation involves the spectral properties of vegetation. The red band encompasses the peak of chlorophyll’s red absorption band of healthy green vegetation, thus, vegetation will readily absorb nearly all of the visible red light spectrum. Contrary to this, the nature of chlorophyll causes near-infrared wavelengths to be reflected in almost its entirety. By performing this
simple ratioing and division of these contradictory bands many negatively impacting variables in remotely sensed imagery are reduced. Namely, the NDVI inherently reduces the effect of shadow, illumination, topography, viewing angle and atmosphere by normalizing these effects. As previously stated, healthy green vegetation has high reflectance in the NIR, but as the health deteriorates and the leaves become yellow this reflectance typically decreases. The important factor to note is that there is a large difference in the amount of reflection of vegetation when comparing the red band and the near infrared band. On the same note, rocks, bare soil and urban areas tend to show little difference in their reflective properties across these same two bands. However, water, clouds and snow have higher reflectance in the visible red band than the infrared band. When all of these variables are considered the resultant DN value of pixel having undergone an NDVI transformation is indicative of the ground feature at that location. The manner in which the NDVI equation is set-up always results in an answer between -1 and +1. Healthy vegetation will result in positive values, approaching +1 if the coverage is extremely dense. Since rocks, bare soil and urban areas show little difference between the bands they will approximate the DN value 0. Furthermore, water, clouds and snow are inversely related to vegetation (in terms of reflectance across red and infrared bands) so their values tend to be negative.

The resultant NDVI image, as produced by PCI Geomatica, lacks appropriate colours for intuitive interpretation. The colours can be edited post-processing but some additional steps were performed to facilitate this process. Firstly, since the output value of the NDVI equation ranges from -1 to +1 the resultant image must be in floating point format. However, this format seemed to pose functionality problems with the software. To simplify computer processing and file size the NDVI was translated into an 8-bit format (see Figure 10).

![Figure 10. EASI modelling algorithm implemented to compute NDVI.](image)
This resulted in values ranging from 0 to 254. This also means that the zero value which is indicative of rocks and bare soil no lies at 127. Values above 127 are indicative of vegetation and values below represent water, clouds and snow.

Figure 11 shows the NDVI with edited colours and Figure 12 is a true colour composite given for comparison.

![Figure 11. Normalized Difference Vegetation Index (NDVI).](image1)

![Figure 12. True colour composite, bands 3, 2 and 1.](image2)

The NDVI has done a reasonable job at delineating vegetation from water, bare soil, rock and clouds. The general trend of these features as seen in the true colour composite is synonymous to the NDVI. However, the editing of colours may have been a source of visual misleading error, in that sample DN values were collected throughout known features in the original NDVI image and new colours were applied to these selected ranges. Therefore, of the three broad categories (water, rock/soil, and vegetation) there is surely some misclassification present in the imagery.

Unfortunately, this image does not contain a wide variety of features to investigate. There are no urban features whatsoever, apart from a single road travelling north alongside the east side of Manicouagan lake (see Figure 13). This road has DN values ranging from 121 to 132, which is expected for rock, bare soil and urban areas (remembering that the original range of -1 to +1 was scaled to 0 to 254). It is important to note the deviation of ± 6 arises from the 30 m spatial resolution of Landsat imagery resulting in spectrally mixed pixels incapable of accurately portraying a roughly 10 m wide road.
Vegetation was found to have DN ranges of 140 to 200. Again, this concurs what was theoretically expected. If a pixel belonging to this category approaches the DN value of 127 it is either unhealthy or covers the area only sparsely, or a combination of the two. Values at the extreme high range indicate healthy and dense vegetation (see Figure 14). Water was expected to have negative values (or in this case values below 127). Again, this was verified with the examination of DN values, which showed a very narrow range of 84 to 89. Figure 14 also shows the values associated with rock and bare soil. Again, these DN values approximate 127. This reservoir has characteristic rocky shorelines.
**Tasseled Cap Discussion**

A Tasseled Cap transformation (TCT) is similar to a PCA. When used on Landsat 5 TM imagery it will produce 3 image bands from the original 6 band dataset. It also serves to extract unique information from all inputted bands and portray this data in a condensed and reduced dataset. The three bands produced are indicative of brightness, greenness and wetness, respectively. As an example of how such an index is achieved, the wetness index contrasts the sum of the visible and near infrared bands with the longer infrared bands to determine the amount of moisture being held by vegetation or soil. The longer infrared bands are the most sensitive to soil and plant moisture, therefore, the contrast between these bands highlights moisture content (Lea, R., et al).

The brightness index, Figure 15, is responsive and indicative of the spectral properties portrayed by the visible spectrum. In that, features that appear bright (high DN values) in a true colour composite, such as rocks, bare soil and urban areas, will have the highest DN values in the brightness index. This is seen in Figure 15 since the central region (known to consist of barren earth and rock) encompasses the highest DN values of the entire scene (up to 221). Water is typically darker than vegetation in a true colour composite and this also true for the brightness index.

![Brightness index as calculated by a Tasseled Cap Transformation (TCT).](image)

**Figure 15.** Brightness index as calculated by a Tasseled Cap Transformation (TCT).
The greenness index is indicative of the biomass present. Figure 16 shows that vegetated areas are portrayed with the highest DN values, with a mean around 220. The next discernable broad category is water, having a DN mean of roughly 190. The central area of the image (exposed earth and rock) has a large spread of DN values, from 80 through to 180. This transformation clearly identifies areas of vegetation, furthermore, the density and health may also be inferred (much like the NDVI).

![Greened index as calculated by TCT.](image)

The third image band produced by TCT is the wetness index (Figure 17) which is indicative of the moisture content. One would expect water to be portrayed as being the most wet and having the highest moisture content, however, this is not the case. The DN values of Manicouagan reservoir are consistently 118 ± 2. The DN values of the central area in the image are significantly higher, ranging from 120 to 200+. This is most likely a result of greater importance being given to moisture content in vegetation and soil since water is obviously 100 % wet and inherently easy to delineate. This makes
sense since the barren earth would have greater moisture retention than the surrounding vegetated areas.

An interesting side note: The name “Tasseled Cap” comes from the fact that when greenness and brightness of a typical scene are plotted perpendicular to one another on a graph, the resulting plot usually looks like a cap.

The TCT was performed on 8-bit imagery and may produce results that not able to be stored in such a format. Thus, a scaling parameter was inputted into the algorithm and the results can be seen as ‘Scaling Information’ in Appendix I. Linear scaling was used to solve this issue, this is accomplished by performing two passes on the data. The first pass is used to determine the minimum and maximum values resulting from the transformation. In the second pass, these values are used to linearly scale the results to the full range of the output channel (PCI Geomatica).
Conclusion

The transformations performed in this investigation offer a wealth of information. Not only do they produce an image which is easier to interpret but they also allow all critical and unique information to be viewed simultaneously as an RGB composite using three bands rather than toggling back and forth between six bands. Furthermore, the resultant images are reduced in file size, allowing faster processing, storage and transferability. On top of this, the actual programming and processing time required to perform each of the PCA, NDVI and Tasseled Cap transformations usually falls under ten minutes – time well spent for an additional perspective of the original dataset.

These transformations are also useful if an image classification is to be performed. Supervised and unsupervised classifications can be decreased in quality when a large volume of redundant information is attempted to be processed. By transforming an image prior to such a classification the file size is reduced and the classification algorithm is allowed to concentrate on pertinent information, resulting in faster processing time and more accurate products.

These transformations are widely used, globally speaking, but some more than others. PCA is a valuable tool and has applications for many image processing tasks, especially when hyperspectral imagery is being investigated. NDVI is often applied to a global scale, computed from low resolution / large area sensors ((2) Milne, T., 2008). The MODIS sensor, for example, has the ability to produce a daily global NDVI, a valuable tool for monitoring day to day changes. Tasseled Cap transformations are not as frequently used as PCA and NDVI since easily accessible algorithms only exists for processing Landsat imagery. Nevertheless, the TCT provides useful information in certain scenarios.

Each transformation is valuable in its own respect but when all are used in conjunction they can also help to verify each others results. For example, the greenness index of the TCT may be used to assess the accuracy of an NDVI transformation. Considering the time spent in processing these transformations and the new perspective gained from the resultant imagery these are worthwhile steps to undertake in many remotely sensed investigations.
References


PCI Geomatica (v. 10.1). Geomatica Prime Help: search word TASSEL. PCI Geomatics, Ontario, Canada.


### Appendix I: PCA Report

#### PCA Principal Component Analysis

<table>
<thead>
<tr>
<th>Input Channels:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output Channels:</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eigenvectors</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Sampling Window:

**Input Channels:** 1 2 3 4 5 6  
**Output Channels:** 7 8 9  
**Eigenvectors:** 1 2 3  

#### Covariance matrix for input channels:

```
1 | 2 | 3 | 4 | 5 | 6
---|---|---|---|---|---
1 | 674.812 | | | | |
2 | 247.596 | 93.558 | | | |
3 | 261.851 | 101.618 | 115.401 | | |
4 | 202.802 | 114.239 | 126.265 | 156.195 | |
5 | 220.393 | 83.673 | 92.404 | 119.539 | 117.378 |
6 | 86.180 | 32.960 | 36.918 | 46.828 | 46.261 | 19.904 |
```

#### Eigenchannel Information:

```
Eigenchannel | Eigenvalue | Deviation | %Variance
-------------|------------|-----------|-----------
1           | 1101.7911 | 33.1932   | 93.59%    
2           | 52.5747   | 7.2508    | 4.47%     
3           | 15.4329   | 3.9285    | 1.31%     
4           | 6.2854    | 2.5071    | 0.53%     
5           | 0.7900    | 0.8888    | 0.07%     
6           | 0.3737    | 0.6113    | 0.03%     
```

#### Eigenvectors of covariance matrix:

```
0.77512 0.28914 0.31171 0.36092 0.27636 0.10858
0.44263 0.08569 -0.01005 -0.32769 -0.76623 -0.31961
0.40169 -0.19014 -0.61513 -0.51094 0.36853 0.16502
-0.08343 0.17126 0.62400 -0.68870 0.13876 0.28421
-0.01487 0.01519 0.13135 -0.15377 0.42505 -0.88203
0.18638 -0.91833 0.34311 0.05787 -0.02788 0.00861
```

#### Scaling Information:

<table>
<thead>
<tr>
<th>Eigen Channel</th>
<th>Output Channel</th>
<th>Min</th>
<th>Max</th>
<th>Range</th>
<th>Deviation</th>
<th>Midpoint</th>
<th>Scale Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>-39.280</td>
<td>293.076</td>
<td>2.00</td>
<td>127.500</td>
<td>1.928</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>-104.539</td>
<td>53.168</td>
<td>2.00</td>
<td>127.500</td>
<td>8.827</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>-102.149</td>
<td>40.969</td>
<td>2.00</td>
<td>127.500</td>
<td>16.291</td>
<td></td>
</tr>
</tbody>
</table>
Appendix II: Image Channel Listing

D:\PHILIPP S\aDiP\Quebec Eye TM\QuebecEye[S 13BIC 8389P 7433L]
20Mar2008

1 [ 8U] band1
2 [ 8U] band2
3 [ 8U] band3
4 [ 8U] band4
5 [ 8U] band4
6 [ 8U] band7
7 [ 8U] PCA :Eigen= 1 Inp: 1: 2: 3: 4: 5: 6:
8 [ 8U] PCA :Eigen= 2 Inp: 1: 2: 3: 4: 5: 6:
10 [ 8U] EASI Modeling Result
11 [ 8U] Brightness
12 [ 8U] Greeness
13 [ 8U] Wetness