# Remote Sensing 4113
# Lab 08: Filtering and Principal Components
# Mar. 28, 2018

In this lab we will explore Filtering and Principal Components analysis. We will again use the Aster data of the Como Bluffs region from lab 07. As before it contains the 3 bands of VNIR data which have 15-m resolution, and the 6 bands of SWIR data, resampled from 30-m to 15-m resolution to match the VNIR data. You should prepare a Word lab report as in past labs, copying it to the completed lab reports folder when you are finished.

Use the Como Bluffs image file contained `lab_07_data` zip file on the class website. As usual, use a temporary directory within `c:\tmp` on your local machine for all your files, and copy the `como_bluffs` image and `hdr` files to your local directory for use. Save any files you want to keep on your `Home Drive`.

## Part I: Filtering

We will explore various convolution filter operations similar to those described in Sabins on pages 267 – 275. First, open the Como Bluffs image (`File->Open Image File`), and load `Band 1` as a gray scale image. Enlarge the main image window enough that the `Scroll` window disappears. As was the case last week, you may need to try a couple times before the window actually is large enough.

1. **A low pass filter, with linked displays.**
   From the main ENVI window select `Filter->Convolutions and Morphology.` A `Convolutions and Morphology Tool` window will open. From its menu select `Convolutions->Low Pass.` Leave the two options `Kernel Size` at the default `3` and `Image Add Back` at the default `0%`. With the kernel now set select `Convolutions->User Defined.` The lower section of the window will now show the kernel which is about to be used – which should be a 3×3 matrix of 1s. This kernel will in effect average a 3×3 box, emphasizing the low spatial frequencies. Click `Quick Apply` which will open a window allowing you to select the data to filter. Select `Band 1` of the Como Bluffs data and click `OK`. A new display window will open with the filtered data. Once again, enlarge enough that the `Scroll` window disappears. Next we want to "Link" the original (Display #1) and filtered (`Display #2`) windows, so we can easily compare them. On the `Display #2` menu select `Tools->Link->Link Displays`. The default parameters should be set to link `Display #1` to `#2.` Leave the default 1's in all the `Offset` boxes as these tell them to compare images with no position offset. Click `OK`. Now when you drag the red zoom box in one window, it will also move in the other. Also, if you hold down the left mouse button in one window (outside the red box), or hold it down in the zoom window itself, the other display will momentarily appear, allowing you to easily compare then.

   Pick an interesting region of the image to display in the zoom box, and click the red + in the zoom box a few times to zoom in closer. (You may want to enlarge the zoom windows slightly to give a broader view.) Save a copy of both the original and filtered zoom box images for inclusion in your report. Describe in a sentence or two the difference between the original and filtered versions

2. `A High Pass filter`
   In the `Convolutions and Morphology Tool` window select `Convolutions->High Pass.` The kernel should change to be 8 in the center with -1s in the other locations. This filter will emphasize the higher spatial frequencies present in the data. Click `Quick Apply.` `Display #2` should change to an overall gray appearance, with bright or dark features where the original image has gradients. Save copies of the original and filtered zoom windows in your report. You may want to click the red – symbol to zoom out somewhat before doing this. Note that with most of these kernel operations we are skipping the last step employed by Sabins – adding the result of the convolution back on top of the image. Just this once, to see the result of that, do the equivalent by increasing the center number of the kernel by one. It should be changed from `8` to `9`, then press the return key to be sure it has been accepted, then click `Quick Apply`. Save a copy of this third zoom window in your report. In a few sentences discuss the differences abound the three images you should now have from this step.

3. `The Laplacian Filter`
   In the `Convolutions and Morphology Tool` window select `Convolutions->Laplacian` and compare the kernel to that shown in Sabins Figure 8-16. This time save full versions of both the original and filtered image for inclusion in your report.

4. **Directional filters**
   While ENVI has built-in directional filters, they are somewhat different than the ones defined in Sabins. However we can create filters like Sabins' by entering the kernel manually. You can stay in “laplacian mode” but just manually change the terms in the kernel. For the NE-SW edge and NW-SE edge filters shown in Sabins Figure 8-20, manually enter the kernel (press return to be sure the numbers are entered), then click `Quick Apply`, and save a copy of the resulting images. Discuss in a few sentences what features have been enhanced by these different kernels and also the Laplacian filter from step 3.

## Part II: Principal Components Analysis (PCA)

### Theory

As discussed in Sabins, Principal Components Analysis examines the data statistically to see if it can be "rotated" from axes made up of the original N bands to new axes (called principal components) where motion along axis 1 explains most of the data variation (for example the "bigness" axis in our weight vs. height plot shown in class), motion along axis 2 explains most of the remaining variation (the "heaviness" axis in our weight vs. height plot), and so forth up through the N'th principal component. The last components typically contain little of the coherent variations in our data and may be dominated by random noise.

As in the class example each pixel in the j'th component image is created out of a linear combination of the original i=1, 2, … N bands. For example in our Weight vs. Height plot, if we were using the right units so the original data points were scattered around a 45° line where Weight $\approx$ Height, then after computing principal components we would have

$$\begin{aligned}\text{Bigness} &= \text{Height} + \text{Weight}\\ \text{Heaviness} &= \text{Weight} - \text{Height}\end{aligned}$$

or after normalizing the equations and writing them in a standard order

$$\begin{aligned}\text{Bigness} &= 2^{-1/2}\,\text{Height} + 2^{-1/2}\,\text{Weight}\\ \text{Heaviness} &= -2^{-1/2}\,\text{Height} + 2^{-1/2}\,\text{Weight}\end{aligned}$$

To use a more generic form, if $(x_1, x_2, x_3, x_4, \ldots x_N)$ are the N different original band measurements for a given pixel, then after transformation the principal component values for that pixel are $(y_1, y_2, y_3, y_4, \ldots y_N)$ where

$$\begin{aligned}y_1 &= a_{11}\,x_1 + a_{12}\,x_2 + a_{13}\,x_3 + \ldots a_{1N}\,x_N\\ y_2 &= a_{21}\,x_1 + a_{22}\,x_2 + a_{23}\,x_3 + \ldots a_{2N}\,x_N\\ &\;\;...\\ y_N &= a_{N1}\,x_1 + a_{N2}\,x_2 + a_{N3}\,x_3 + \ldots a_{NN}\,x_n\end{aligned}$$

and $a_{ij}$ are the transformation coefficients.

| In our example | $X_1$ = Height | $X_2$ = Weight | are the original bands |
|---|---|---|---|
| | $Y_1$ = Bigness | $Y_2$ = Heaviness | are the principal components |

and the coefficients for the transformation are

$$\begin{aligned}a_{11} &= +2^{-1/2} & a_{12} &= +2^{-1/2}\\ a_{21} &= -2^{-1/2} & a_{22} &= +2^{-1/2}\end{aligned}$$

Since the system only understands statistics, not geology, it is up to the user to assign geological meaning to the new axes or components. Quite often the first component consists primarily of topography shading, with slopes facing towards the sun brighter. The second, third, fourth, etc. components often contain most of the geologically interesting information. If for example the scene consists primarily of two different types of material, then one of the early principle components will represent the difference between those two types of material.

**PCA Example**

Close all the ENVI windows except the `Available Bands List` and `Display #1.`

5. Load bands #3, #2, and #1 into `Display 1` as an RGB image. If necessary enlarge it enough that the `Scroll` window disappears. Finally, save it in your report for comparison with later results.

6. From the main ENVI menu select `TRANSFORM->PRINCIPAL COMPONENTS->FORWARD PC ROTATION->COMPUTE NEW STATISTICS AND ROTATE` to compute a PC transformation of all nine bands of your Como Bluff scene. In the window which opens select `como_bluffs` as the input file and accept the default settings of using the `full scene`, all `9 bands`, and `no mask`. Click `OK`. In the next window to open, `choose` the name of the `Output Stats File.` Adopt a name like `como_bluffs_pc.sta,` located in your own directory. Also `chose` the `Output Filename` (for the components), picking something like `como_bluffs_pc`. Leave the other options at their default values and click `OK`. The system will spend a short time calculating the principal components, and a 9 band PC image will appear in the available bands list. Also, the system will open a window showing a plot of "Eigenvalue" vs. Component number. The plot indicates how much of the total variation is due to each principal component or "Eigenvector". Most of the variation is due to the first one, a moderate amount the the second, decreasing towards zero for the ninth. Save the Eigenvalue plot for inclusion in your report.

7. From the `Available Bands List` window click `Display #1>` and create a `New Display` window. Load PC Band 1 into it in gray scale. Enlarge the main display window till the scroll window disappears, then link `Display #2` to `Display #1` using the commands given earlier. Go through loading each of the nine PC bands, saving copies of them for your report. For each component use the left mouse key to blink between the component image and the original IR color one. Discuss in a few sentences which of components show significant geological information and which are dominated by noise. At this point you don't need to discuss the detailed geological information for each band but provide an overview, for example: *Band A primarily shows topographic shading, Band B, E, and F show rock units, while band C shows the presence of water. Bands G, H, and I seem to mostly contain noise. (You would need to replace the letters by numbers, and the details may be different than those listed in this example.)*

8. We will assume that PC Band 1 primarily indicates topography and shadowing. If we examine PC Band 3, it seems to sense primarily water and vegetation. Water shows up as very bright, and vegetation as very dark. Therefore to examine the geological information we will ignore these two bands and load the remaining most significant PC bands. Load PC Bands 5, 4, and 2 into `Display #2` as an RGB image. Save a copy of the RGB image for your report. Discuss in one or two paragraphs how well the RGB image suppresses simple effects like topography, but highlights differences in the geological units. Compare the image to the geological map and discuss what units in the map you can recognize in the image.

9. Try loading a couple other combinations of PC component bands to see which remaining ones are most useful. Include at least one of those images in your report although with a sentence or two describing what additional information this image reveals.

10. To view the statistics information and the transformation equations used to create our Principal Components, from the main ENVI menu select **`Basic Tools->Statistics->View Statistics File`** and select the `.sta` file you created earlier. After clicking **`OK`** a window will open showing a plot in the upper pane and statistics information in the lower pane. To save the information in a form that can be used by other programs, from that new window's menu select **`File->Save->Save results to text file...`** then specify a name for that file, in your own directory.

    Scroll down in the statistics pane till you can see a matrix labeled Eigenvectors. That matrix contains the $\mathbf{a}_{ij}$ coefficients described above. For example, using the notation given above, to produce principal component 1 (called $Y_1$) from the original band values (called $X_1$, $X_2$, $X_3$, ... ) we would compute

    $$Y_1 = 0.56\ X_1 + 0.52\ X_2 + 0.24\ X_3 + 0.26\ X_4 + ...$$

    (I've rounded off the coefficients to two significant digits.)

    Note that all the coefficients for $\mathbf{Y_1}$ are positive. $\mathbf{Y_1}$ is in some sense a measure of the "average" brightness of a pixel, controlled partially by topography and shadowing.

11. We noted earlier that Component 3 seems to be strongly positive for water and strongly negative for vegetation. Row 3 tells us that

    $$Y_3 = 0.45\ X_1 - 0.18\ X_2 - 0.86\ X_3 - 0.079\ X_4 + \ldots$$

    Most of the coefficients are relatively small, so for simplicity lets ignore them. If we consider only the two largest terms

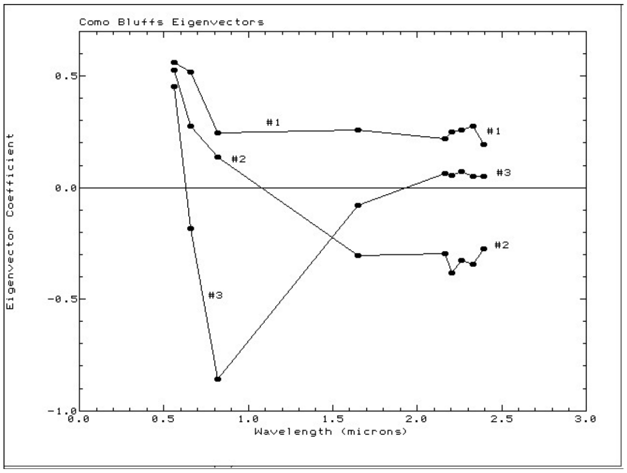
    $$Y_3 = 0.45 * X_1 - 0.86 * X_3$$

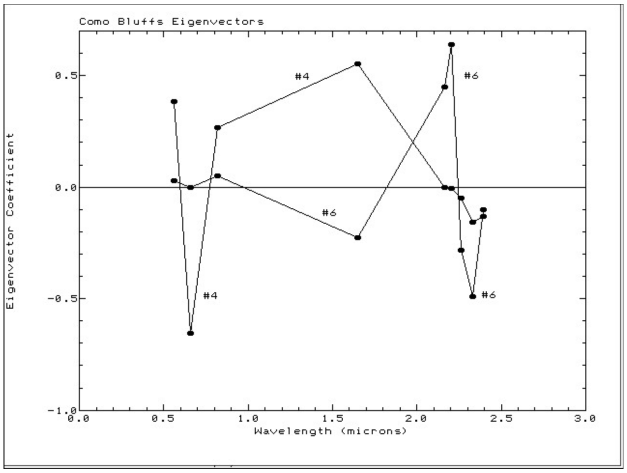
    where $\mathbf{X_1}$ is the green band and $\mathbf{X_3}$ is the Near-IR band.

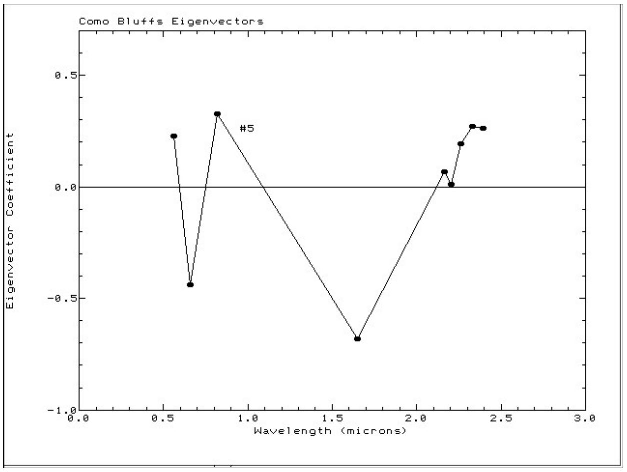
    Discuss briefly the reflectance of water and of vegetation in these bands, and explain in a couple sentences why those spectra would result in large positive $\mathbf{Y_3}$ for water and large negative $\mathbf{Y_3}$ for vegetation.

12. For your information plots of the first six transformation coefficients are shown versus wavelength on the following page. Most of them are more complicated than $\mathbf{Y_1}$ or $\mathbf{Y_3}$ but in some sense they have automatically been chosen to extract other features in the spectra. For example if most pixels were simple gray but a few showed a narrow absorption feature in one band, coefficients would be chosen to make one of the principal components roughly equivalent to a band-depth map for that absorption feature. In these plots remember coefficients close to zero mean a band is being ignored, but large positive or negative coefficients means the band contributes heavily to that principal component.

One component in particular has coefficients that are trying to extract information from variations in the 2.3 – 2.4 μm region. In a few sentences discuss which principal component this is and what class of minerals might have diagnostic features in this region. Finally, examine the image for this principal component. Which geological units are being separated by this principal component?







Summary:

Your report should contain:

Filtering:

1. A magnified portion of your original image and the result of the 3×3 box average filter, and a couple sentence discussion of the results.
2. A magnified portion of your original image and of the two high pass filter images, from steps 8 and 9, plus a couple sentence discussion of the results.
3. The original full image and the full image result of the Laplacian filter.
4. Copes of the full NE-SW edge and NW-SE edge filters, and a discussion of how those and the Laplacian filter enhance different features in your image.

Principal Components Analysis

5. The RGB image and the Eigenvalue plot showing the contribution to image variation from each eigenvector.
6. The nine eigenvector images, along with an overview discussion of which show geological information, and which noise. (Save the detailed discussion of the geology for later.)
7. An RGB version of PC Bands #5, #4, and #2, along with a discussion comparing that image to the geological map. Discuss what units you can recognize in the map.
8. At least one other RGB combination of PC Bands, and a sentence or two describing what additional information it shows.
9. A discussion of why water and vegetation have spectra which produce large and different effects on PC component 3, given the coefficients of $Y_3$.
10. A discussion of which component picks out minerals with features in the 2.3 – 2.4 μm region, what minerals these would be, and what geological units are being recognized or separated by this component.