

NTNU

TTK4255

Robotvision

Hyperspectral imaging

Mads Formo

April 8, 2020

Contents

1	Getting familiar with the data	1
1.1	Finding the spectral resolution	1
1.2	Relation to human color perception	1
1.3	Create a pseudo RGB image from the hyperspectral bands	1
1.4	Representative spectra for selected points	1
2	Classification & Bio-geophysical Parameter Retrieval	3
2.1	Can we predict where there is chlorophyll through classification?	3
2.2	How well can we directly estimate the chlorophyll content?	3
2.3	How can we estimate the reflectance from the surface of the ocean?	4
2.4	Compute chlorophyll concentration using atmosphere-corrected data	6
2.5	Classify land versus water	6
2.6	Other bio-geophysical parameters	7
2.7	Alternative atmospheric correction methods	7
3	Dimensionality Reduction & Noise Filtering	8
3.1	What is dimensionality reduction?	8
3.2	Principal Component Analysis (PCA)	9
3.3	How does dimensionality reduction via PCA affect classification?	10
3.4	Maximum Noise Fraction	10
3.5	Maximum Noise Fraction on HICO noisy	10
3.6	Discuss your results	10
3.7	How can we best use the subspace?	10
4	Fun but definitely hard problems	10
4.1	Deep learning	10
4.2	Multispectral-hyperspectral image fusion	10
4.3	Spatial-spectral methods	10
4.4	Locating methane emissions	10
	References	11

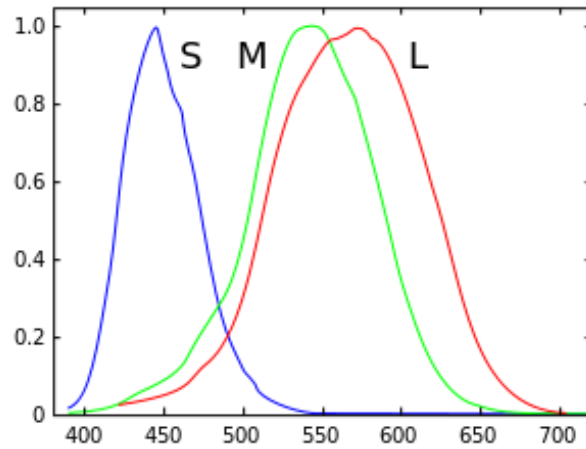


Figure 1: Graph for the human color sensitivity curves, according to Wikipedia [1]

1 Getting familiar with the data

1.1 Finding the spectral resolution

To find the spectral resolution of the dataset, we load the *hico_wl* array, which contains the wavelength corresponding to band i . We loop through the array and compare each wavelength i with the previous wavelength $i-1$ and we find that the average distance between the wavelengths is $5.728nm$, which seems to be constant between all wavelengths.

1.2 Relation to human color perception

The color sensitivity of the human eye is shown in fig. 1. As we can see, blue color has a peak around $450nm$ (*S*-curve), green peaks at $550nm$ (*M*-curve), and red at $600nm$ (*L*-curve).

1.3 Create a pseudo RGB image from the hyperspectral bands

From the *hico_wl* array, we find that Blue ($450nm$) is located at index $i = 8$, green ($550nm$) at $i = 25$, and finally red ($600nm$) at $i = 34$. We combine these indices from the HICO dataset and show it as an image to create a pseudo RGB image, shown in fig. 2.

1.4 Representative spectra for selected points

We want to look at the representative spectra of the points (20,20), (100,70) and (400,30), which is in deep water, shallow water and vegetation respectively. As we can see in fig. 3, we see that there is a clear difference in the spectra between water and vegetation. Both have amplitude peaks at the lower end of the spectra and then drop off in power as the wavelength increases. Vegetation however increases again in power at a wavelength of around $700nm$, while the water is still decreasing. The findings here seem to agree to the findings of Lucke et al [2] as the general shape of the curves matches those of figure 12 in that report.

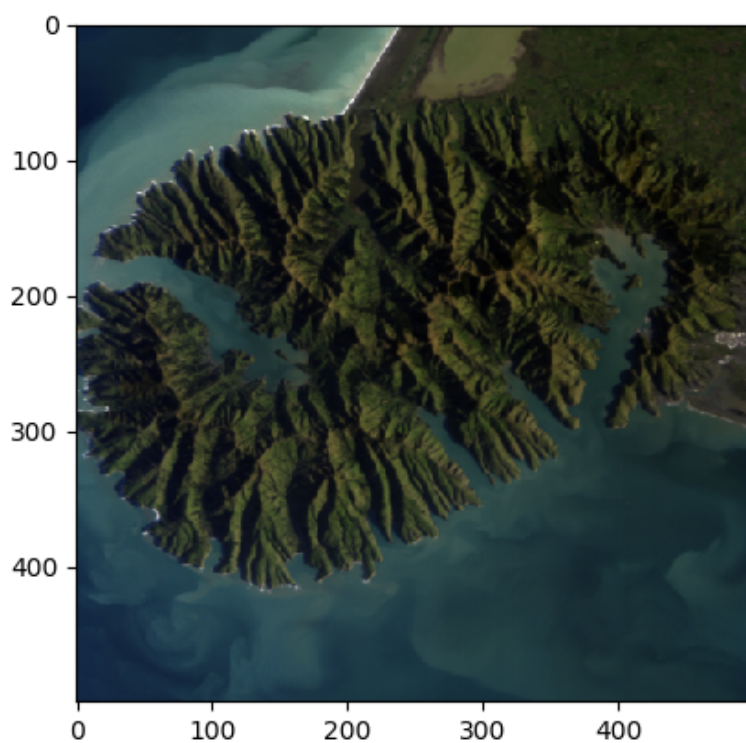


Figure 2: Pseudo RGB image, showing R (600nm), G (550nm), B (450nm)

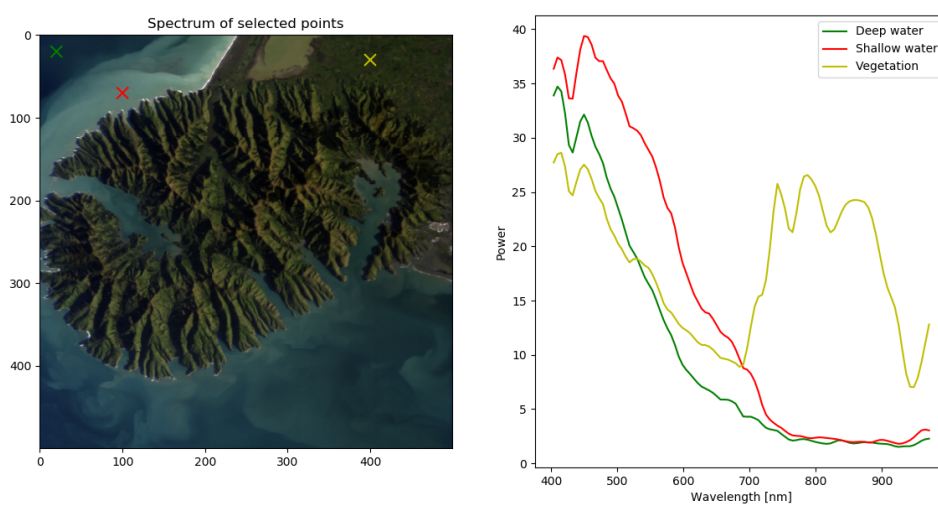


Figure 3: Representative spectra of specific points

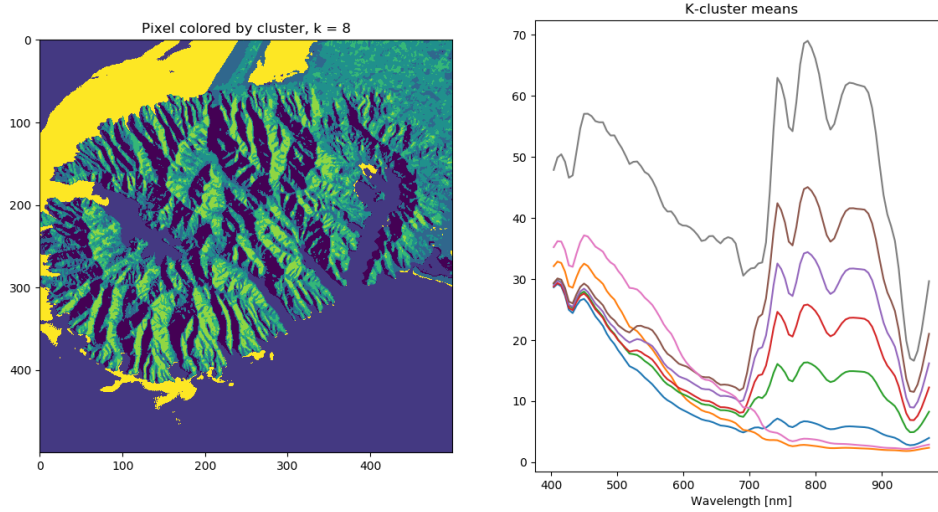


Figure 4: K-mean clusters of the image

2 Classification & Bio-geophysical Parameter Retrieval

2.1 Can we predict where there is chlorophyll through classification?

We will use *K-Means clustering* to classify the data. K-means clustering is a unsupervised learning algorithm that can be used to classify and cluster data into k different clusters. The data points are adjusted iteratively until all points are associated with the nearest cluster. We want to cluster each observation (pixels, with n spectral channels) into a specific cluster (environment class, ie. deep water, shallow water, vegetation).

As we can see from fig. 3, we know that those three different points have distinctly different spectra, thus it should be possible to classify them accordingly. The results of a K-mean clustering, run with Spectral Python's *kmeans* function [3], can be seen in fig. 4. We clearly see different classes for water, land, and vegetation, the latter containing lots of chlorophyll. We also see a very distinct class along the coast on the upper part of the image. This may very well be a collection of chlorophyll, but it might also just be shallow water, or more likely a combination of both.

2.2 How well can we directly estimate the chlorophyll content?

We use the NASA OBP algorithm, defined in equation 4 in the assignment [4], as well as the parameters given there, to try to visualize the chlorophyll contents. Using the closest available wavelengths in the dataset, $\lambda_{green} = 553$ ($i = 26$) and $\lambda_{blue} = [444, 490, 507]$ ($i = [7, 15, 18]$). The results can be seen in fig. 5. We can clearly see high concentrations on the north west coast (assuming north is at the top of the image), same place as in fig. 4, but now we also see quite a bit on the southern coast as well. Thus it seems that this algorithm performs better than the k-means clustering.

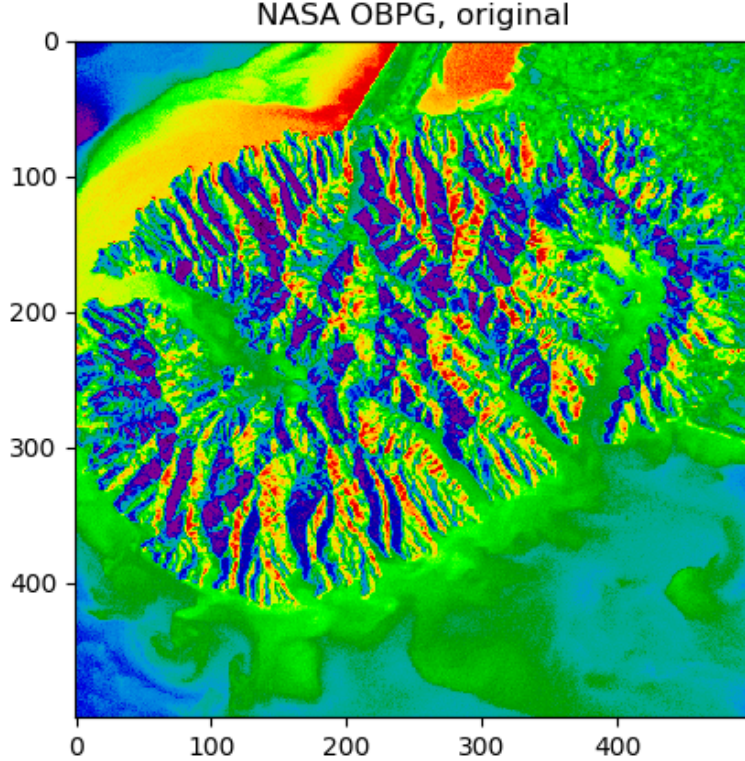


Figure 5: Results from the NASA OBPG algorithm, showing chlorophyll concentrations in the water

2.3 How can we estimate the reflectance from the surface of the ocean?

The data in the HICO dataset actually contains the measurement of radiance exiting the top of the atmosphere, and not the radiance of the water directly. Therefore we must recover the radiance R_{rs} of the water from the top of atmosphere (TOA) measurements. This is done with the empirical line (ELM) method, as described in [4], on the form eq. (1). Where a and b are terms that model the absorption of light in the atmosphere, which is to be estimated, and L is the measured value in the HICO dataset.

$$R_{rs}(\lambda) = \frac{L(\lambda) - b(\lambda)}{a(\lambda)} \quad (1)$$

After performing the atmospheric correction, the resulting image is shown in fig. 6. It is very difficult to see a clear difference between the two by only looking at the image, but if we look at the pixel values for the red, green and blue channels separately, we see that the blue color channel is only about 2% of the original, uncorrected pixel value, while the red and blue channels are about 9-10% of their original values. Thus it would seem that the atmospheric correction removes some of the blue color of the image.

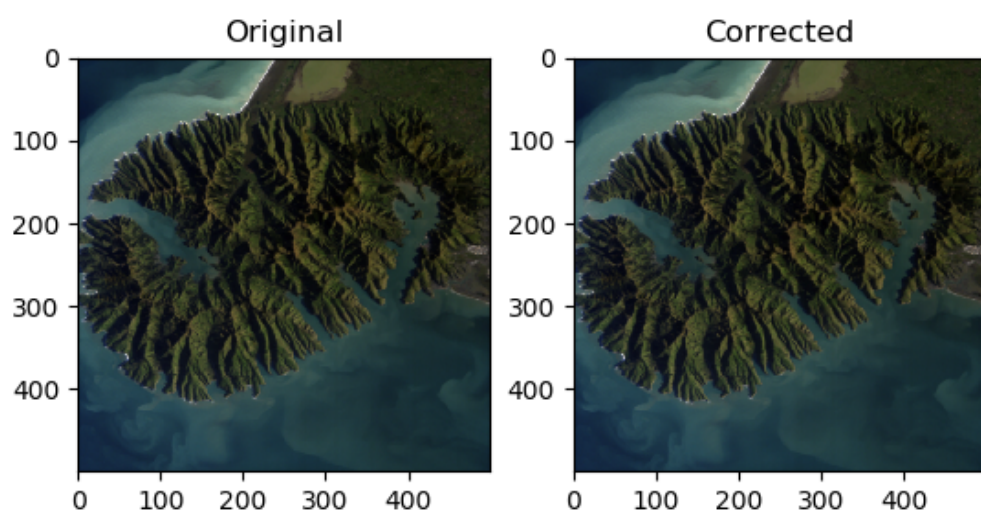


Figure 6: Pseudo RGB image of the atmosphere corrected image

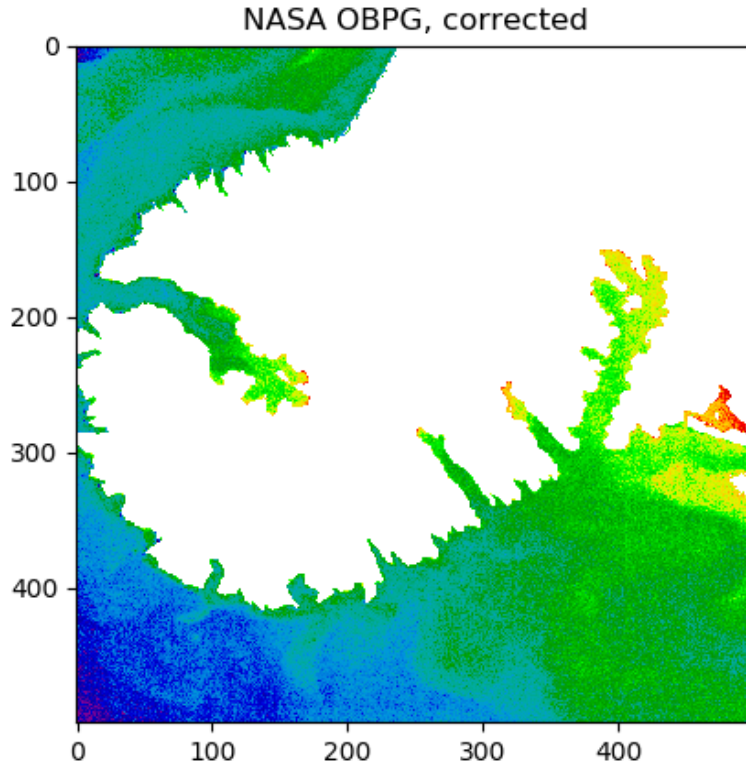


Figure 7: Results from the NASA OBPG algorithm using the atmospheric correction described in section 2.3, showing chlorophyll concentrations in the water

2.4 Compute chlorophyll concentration using atmosphere-corrected data

After performing the atmospheric correction, as described in section 2.3, we perform the NASA OBPG algorithm again on the corrected image cube. As we can see from the results in fig. 7, it is quite different from the original results from fig. 5. The strong concentration we found at the north west coast of the original (fig. 5), is no longer to be seen in the corrected image. Instead, we see a high concentration of chlorophyll on the south eastern coast, as well as a small patch in the bay on the western part of the landmass.

2.5 Classify land versus water

We want to classify only the chlorophyll content of the water, we're not interested in looking at the land. Therefore we want to mask out the landmass and show only the water. We again use K-means clustering to classify the data into different classes. By performing it multiple times for varying numbers of classes and comparing the result from fig. 4, fig. 8, as well as all other iterations from 1-10 classes on both corrected and original data, we find that performing k-means clustering with 10 clusters on the atmospherically corrected image cube gives the best performance regarding classifying

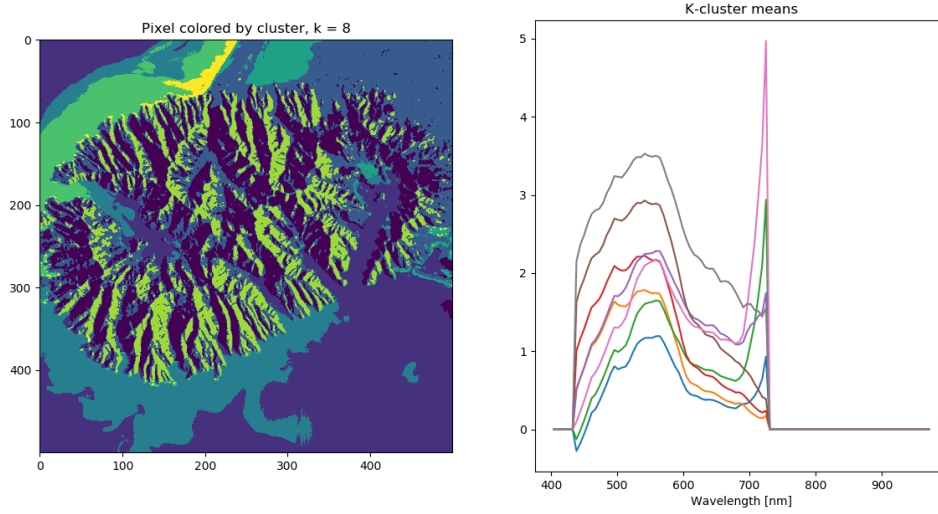


Figure 8: K-mean clustering performed on the atmospheric corrected image cube

water vs land. Setting the max number of iterations to 100, the algorithm converges after around 80 iterations, giving a good result.

After the k-means clustering is performed, we look at the spectral plot of the different classes. We know from section 1.4 roughly how the spectra of water looks compared to land/vegetation. We pick the spectral lines from the k-means result that most resemble the spectral lines of water, and we remove any lines that resemble those of land/vegetation. We then end up with a handful of different classes that represents water, and the rest is land. We set the pixel values of all points inside the water classes to 1, and all other pixels to 0 to create the mask. The mask itself is shown in fig. 10. When put over the image, we get fig. 9.

2.6 Other bio-geophysical parameters

In the available spectral measurements between 400-1000nm we may find several other kinds of bio-geophysical parameters. In general, bio-geophysical parameters include biological (plant species, interactions in the ecology, biotic productivity), geological (soil types, erosion), and physical (light, heat) [5]. Examples include detecting phytoplankton pigments and gelbstoff [6], or land degradation processes in semi-arid geographical areas [7].

2.7 Alternative atmospheric correction methods

One of the drawbacks of using the empirical line method (ELM) is that it requires at least one field, laboratory, or other reference spectrum, preferably 2 or more like the 2 reference spectra used in this report (shallow and deep water) to create the linear regression used in removing the solar- and atmospheric path radiance. If such reference spectra, for all required wavelengths, is not available for the region of interest, this method can not be used. [8]

There exists a whole lot of different atmospheric correction tools and algorithms. Harris Geospatial

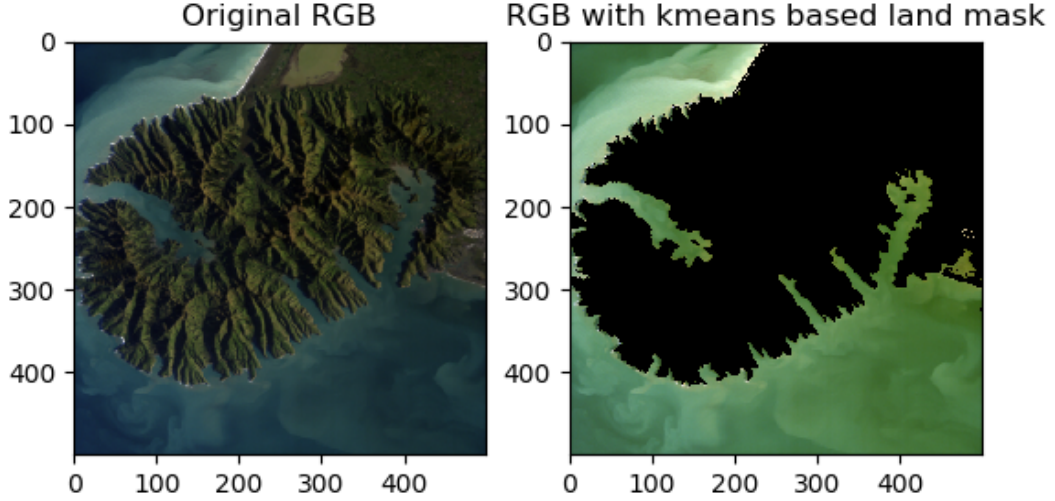


Figure 9: Masking away the land area

lists 9 different tools available for ENVI [9], including Dark Subtraction, Empirical Line, FLAASH, Flat Field, IAR, Log Residuals, QUAC, Thermal Atmospheric Correction, and Convert to Emissivity and Temperature.

For instance, the Flat Field Correction algorithm can be easily implemented by normalizing the images to an area of known "flat" reflectance, where the average spectrum from a region of interest can be used as reference.

3 Dimensionality Reduction & Noise Filtering

The HICO image cube is converted to a matrix on the form $L \times N$, with L spectral bands rows, and $N = WH$ columns.

3.1 What is dimensionality reduction?

Dimensionality reduction is the act of reducing the amount of random variables to consider in for instance machine learning and statistics, involving feature selection and feature extraction. Thus making the data smaller and making analyzing it faster and easier.

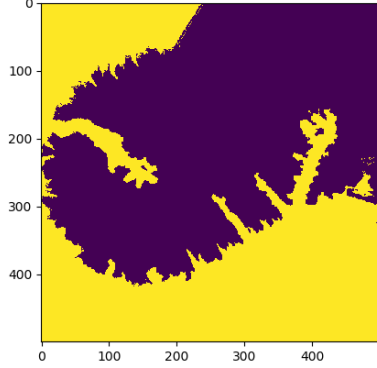


Figure 10: The mask found in section 2.5, based on kmeans clustering

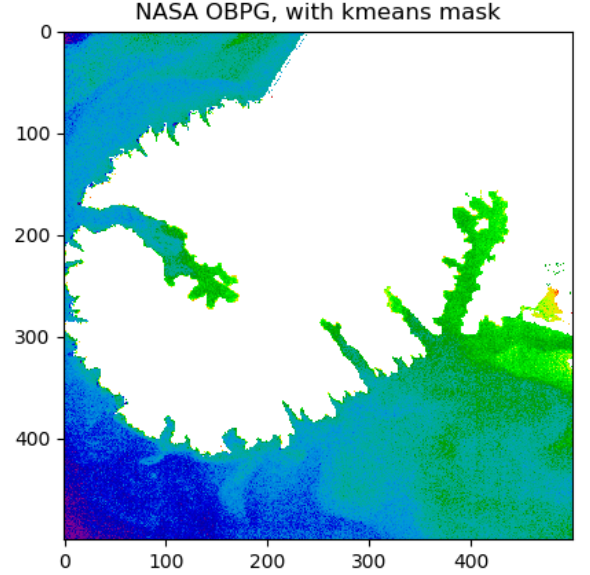


Figure 11: Running the OBPG algorithm on the masked image, using mask from fig. 10

Considering fig. 3, we see that there are quite a few similarities in the spectra of the various points. Looking at the two points in the water (shallow and deep), we see that the spectra are almost identical to each other for wavelengths between around 400 - 470nm, except for the clear difference in the amplitude. Even the spectra for the vegetation seem to share this characteristic. For wavelengths between 470 - 800nm all spectra seem to have unique signatures, until the two points in the water again seem to share quite similar spectra for wavelengths over 800nm, with only small variations between the two. An other thing to consider is the fact that the atmospheric correction algorithm used in section 2.3 is invalid for wavelengths outside the range 438 - 730nm, so these invalid bands can be dropped. Thus there seems to be some opportunities to reduce the dimensionality in the spectral dimensions for the aforementioned wavelengths.

In the spatial direction, we could for instance ignore all of the landmass, as we are only interested in looking at the water. This can be done as described in section 2.5, where all points on land are simply set to 0, thus reducing the spatial dimension significantly by removing a big chunk of the image. The spatial dimension can be reduced further by doing a nearest neighbor approach to cluster data together, for instance in the deep water in the upper left part of the image, fig. 2, where a portion of the ocean seem to be very similar in appearance. This can be done using for example k-means clustering, as described previously in section 2.1.

3.2 Principal Component Analysis (PCA)

PCA is a dimensionality reduction technique that transforms the columns of a dataset into a new set features, by finding a new set of directions that explain the maximum variability in the data [10]. These new coordinate axes/directions are known as the Principal Components (PCs). The dataset

columns contains the amount of variance of the data, computing the PCs will help explain the vast information in the original data in a fewer amount of columns.

Original data shape: (1000, 100) PCA dataframe shape: (1000, 100) PCA weights shape: (100, 100)

The weights in the PCA is the eigenvectors of X. Each principal component is the dot product of its weights and the mean-centered data (each column of X is subtracted from its own mean, so the mean of each column is zero).

$$PC_i = weights \cdot X_{meancentered} \quad (2)$$

So 100
weights per
column??
100 x 100
weights in
total??

3.3 How does dimensionality reduction via PCA affect classification?

3.4 Maximum Noise Fraction

3.5 Maximum Noise Fraction on HICO noisy

3.6 Discuss your results

3.7 How can we best use the subspace?

4 Fun but definitely hard problems

4.1 Deep learning

4.2 Multispectral-hyperspectral image fusion

4.3 Spatial-spectral methods

4.4 Locating methane emissions

References

- [1] Wikipedia. *Spectral sensitivity*. Jan. 2020. URL: https://en.wikipedia.org/wiki/Spectral_sensitivity.
- [2] Robert L. Lucke et al. “Hyperspectral Imager for the Coastal Ocean: instrument description and first images”. In: *Appl. Opt.* 50.11 (Apr. 2011), pp. 1501–1516. DOI: 10.1364/AO.50.001501. URL: <http://ao.osa.org/abstract.cfm?URI=ao-50-11-1501>.
- [3] Thomas Boggs. *SpectralPython*. Jan. 2014. URL: <https://www.spectralpython.net>.
- [4] Sivert Bakken, Joe Garret, and Simen Haugo. “Hyper Spectral Imaging Project”. In: *TTK4255 Robotic Vision, NTNU* (Feb. 2020). URL: https://ntnu.blackboard.com/bbcswebdav/pid-881058-dt-content-rid-25343529_1/xid-25343529_1.
- [5] ESA Sentinel Online. *Bio-geophysical Variable Mapping*. URL: <https://sentinel.esa.int/web/sentinel/thematic-areas/land-monitoring/bio-geophysical-variable-mapping>.
- [6] ZhongPing Lee, Kendall L. Carder, and Robert A. Arnone. “Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters”. In: *Appl. Opt.* 41.27 (Sept. 2002), pp. 5755–5772. DOI: 10.1364/AO.41.005755. URL: <http://ao.osa.org/abstract.cfm?URI=ao-41-27-5755>.
- [7] H. Kaufmann et al. “SAND - a hyperspectral sensor for the analysis of dryland degradation”. In: *IEEE International Geoscience and Remote Sensing Symposium*. Vol. 2. 2002, 986–988 vol.2.
- [8] Evanthia Karpouzli and Tim Malthus. “The empirical line method for the atmospheric correction of IKONOS imagery”. In: *International Journal of Remote Sensing - INT J REMOTE SENS* 24 (Mar. 2003), pp. 1143–1150. DOI: 10.1080/0143116021000026779.
- [9] Harris Geospatial Solutions. *Atmospheric Correction*. Jan. 2020. URL: <https://www.harrisgeospatial.com/docs/atmosphericcorrection.html>.
- [10] Selva Prabhakaran. *Principal Component Analysis (PCA) – Better Explained*. Mar. 2019. URL: <https://www.machinelearningplus.com/machine-learning/principal-components-analysis-pca-better-explained/>.