Revealing the Hidden Writing of a Palimpsest using Hyperspectral Imaging Analyzed by Principal Component Analysis and Generalized Least Squares Weighting

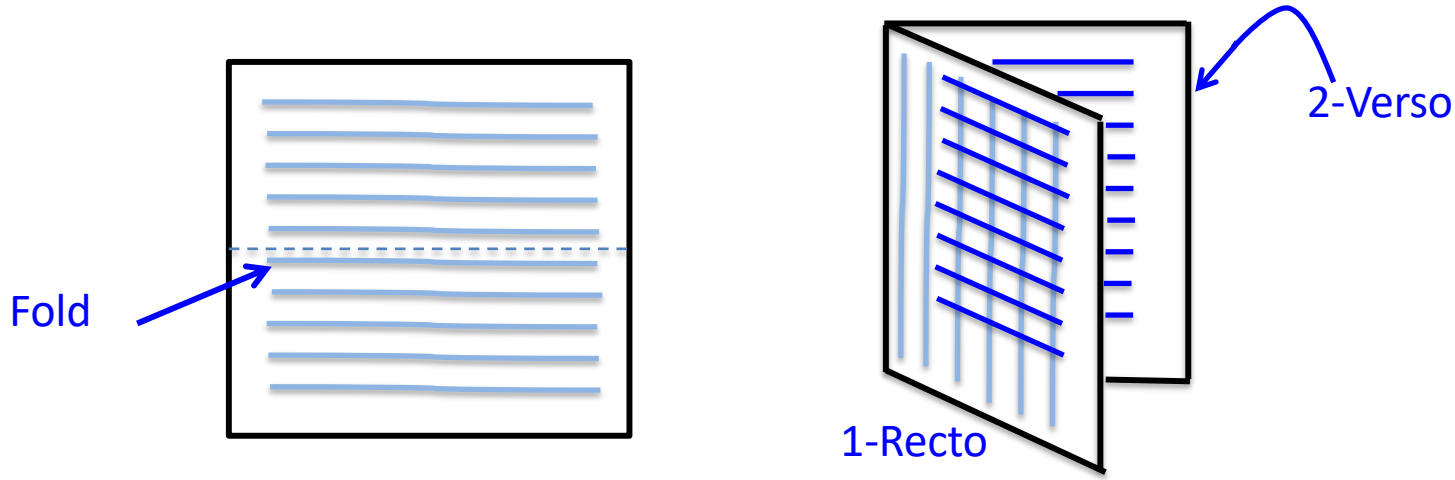
Donald Dahlberg – Lebanon Valley College

Meghan Wilson and Fenella France - Library of Congress

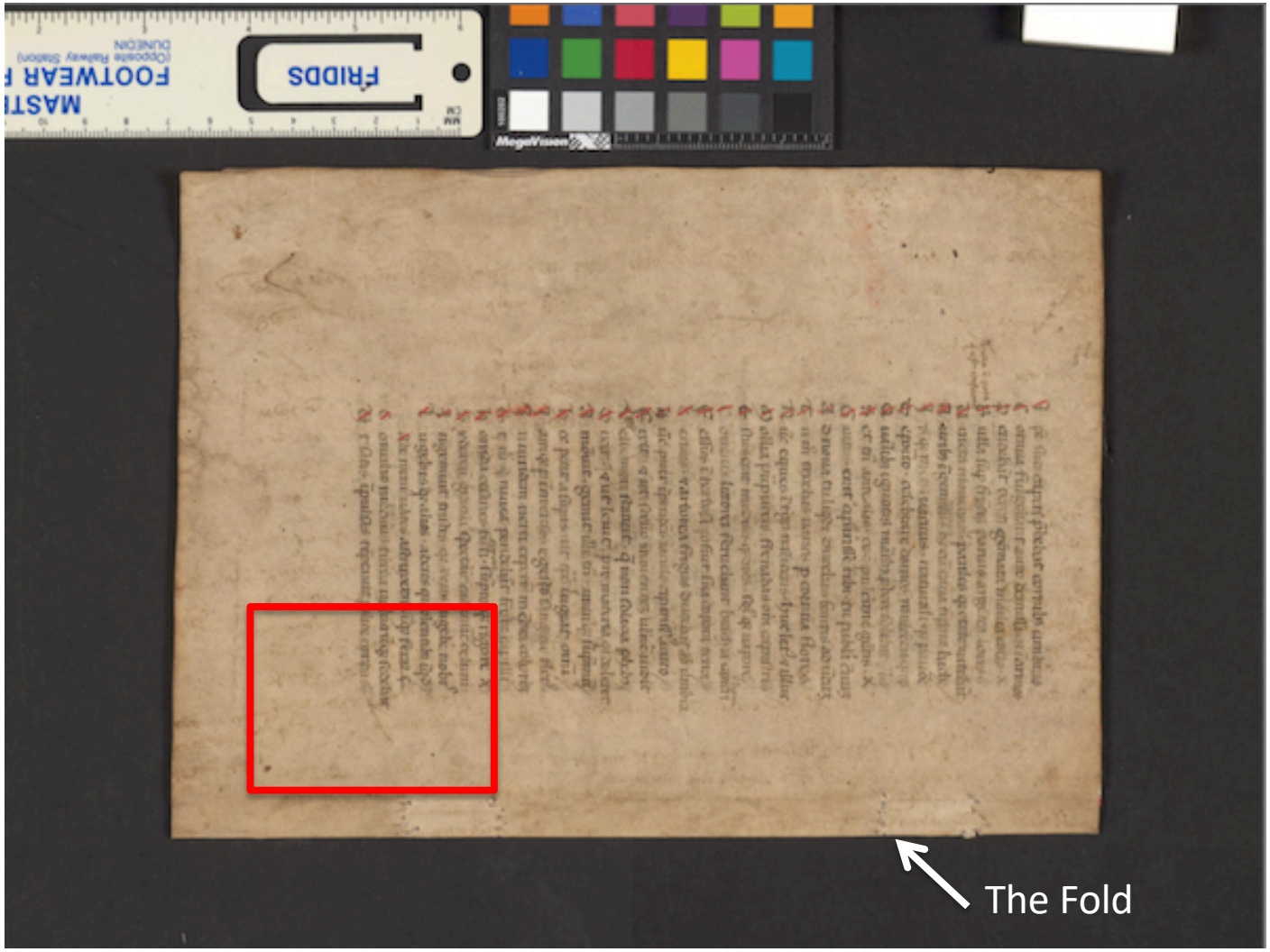
Neal Gallagher - Eigenvector Research

This 15th Century Document Originally Had Writing On One Side

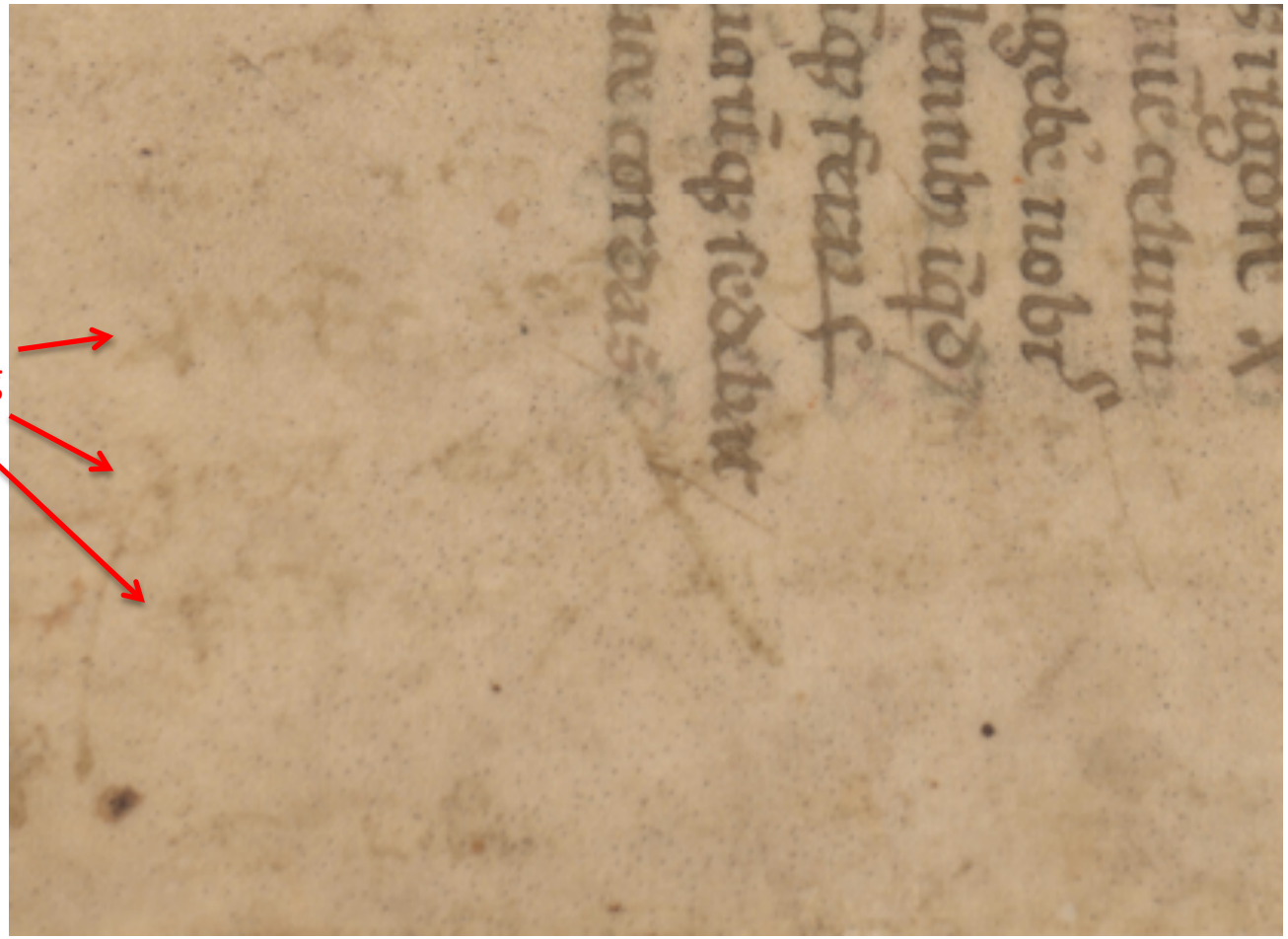
- The ink was washed off.
- The document was folded in half with the fold parallel to the original text.
- This yielded four sides for the new writing.
- The underwriting ended up on sides 1-Recto and 2-Verso, perpendicular to the overwriting.



RGB Image of 2-Verso of Document



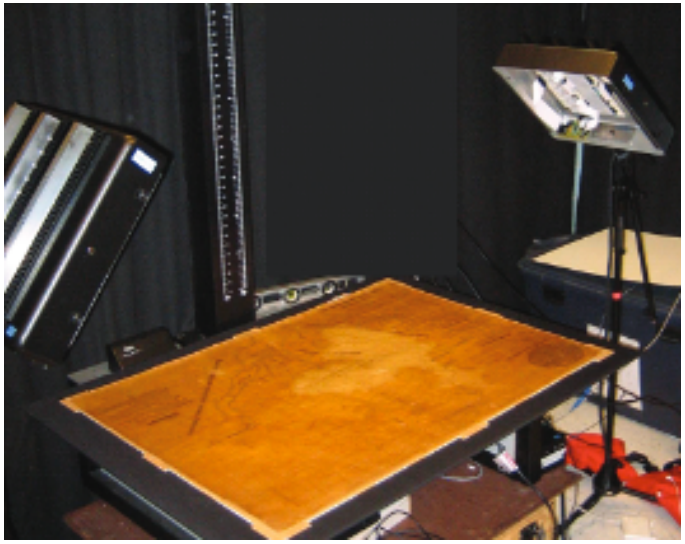
Zooming in Shows Underwriting



Faint
Underwriting

One of the Best Nondestructive Methods of Revealing the Underwriting is Hyperspectral Imaging

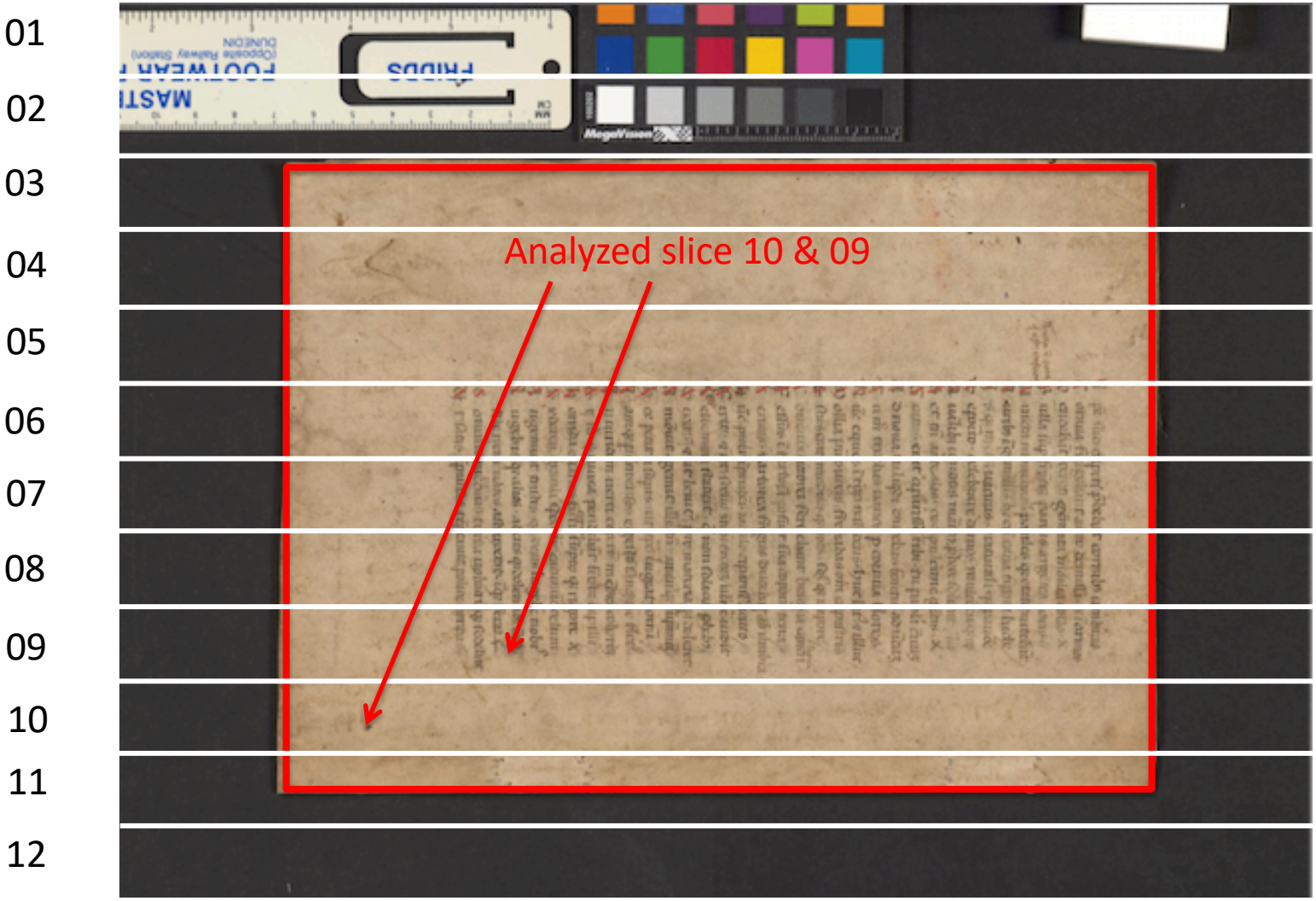
- Library of Congress, Preservation Research and Testing Division
- Two banks of LED illumination panels that can span 365-1000nm (6-30 nm band width)
- The document was photographed at 13 wavelengths (365, 450, 465, 505, 535, 592, 625, 638, 700, 735, 780, 870 and 940 nm)
- Mega Vision-Equipoise Eureka imaging system
- 39 megapixel monochrome camera E6 back
- APO-Digitar 5 6/120 lens



These Images Contain a Great Deal of Data

- Each of the 13 Images are 5412 X 7216 = 39,052,992 pixels/image. This is a total of 507,688,896 pieces of data.
- The chemometric analyses that we are going to do on this data are very calculation intensive.
- These images are too big for some computers to analyze in a timely manner.
- During the exploratory stage, breaking the image up into smaller sections allows us to analyze the data in a timely manner and makes it easier for the analyzer to made decisions and visualize the results.

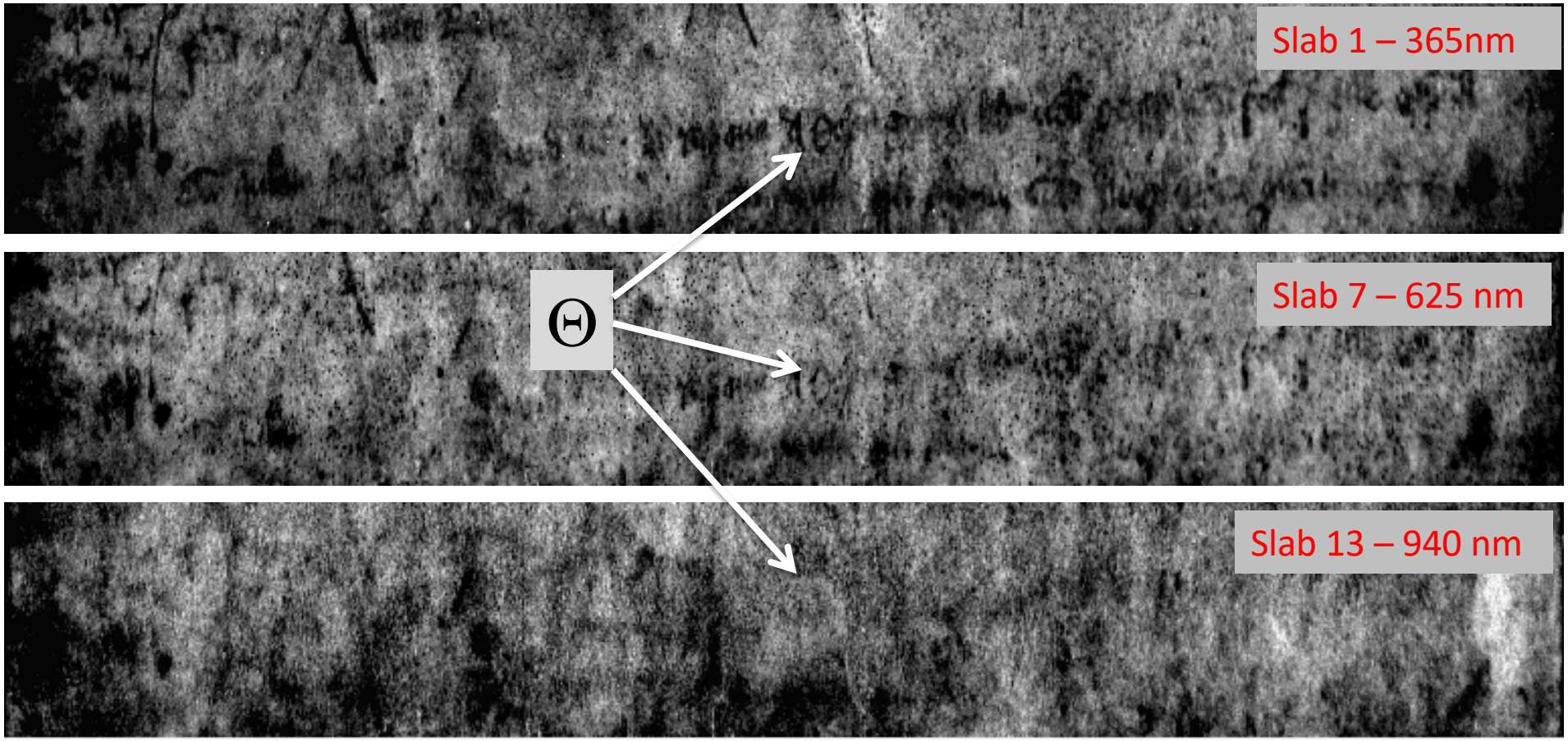
Slice the Image into 12 segments and Crop the Image



Let’s Start Out with Segment 10
451 x 5311 pixels x 13 channels

Examine the Individual Images

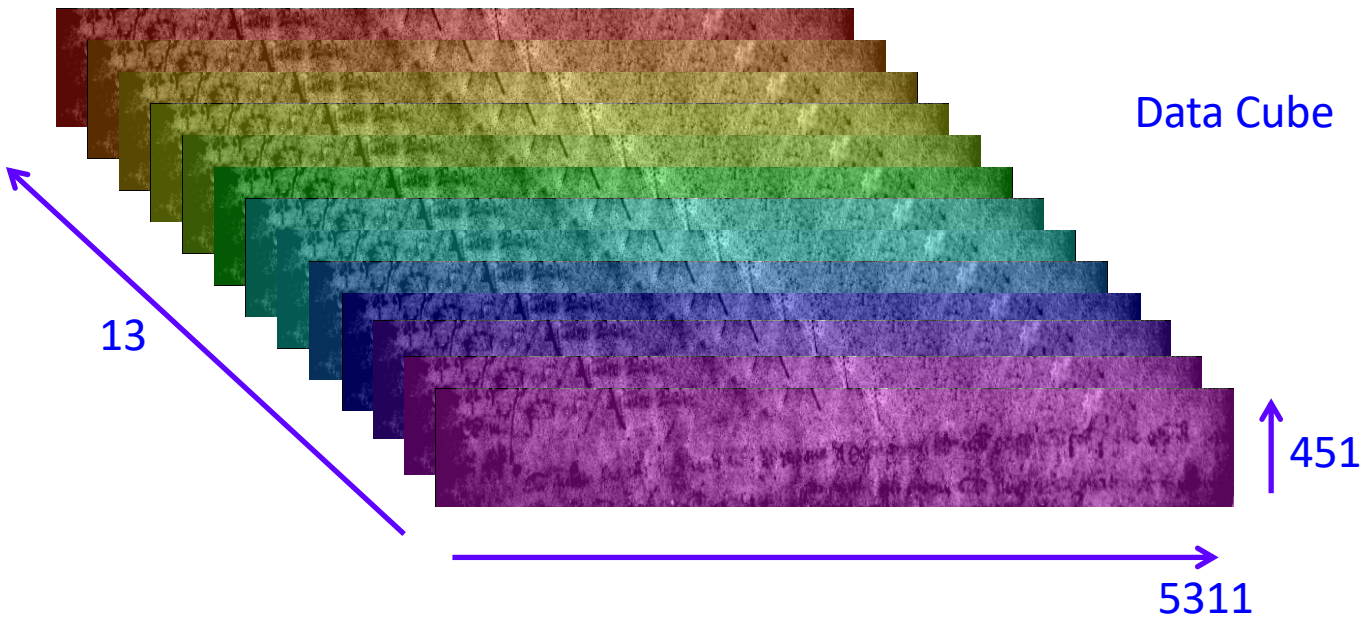
UV region most clearly reveals the underwriting



We Want to Be Able to Combine the Underwriting Information from All 13 Channels

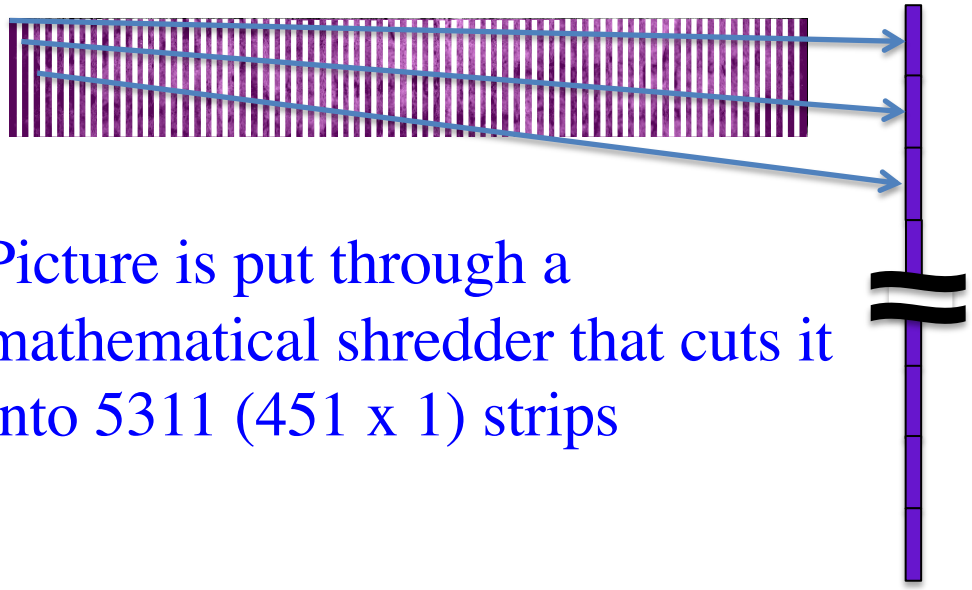
This is what Principal Component Analysis (PCA) was designed to do

Stack of 13 (451 X 5311) Grayscale Images



Before Performing a PCA We Have to Convert the Data to a Matrix – Matricize the Image

Slab 1 – 365 nm (451 x 5311)



Picture is put through a mathematical shredder that cuts it into 5311 (451 x 1) strips

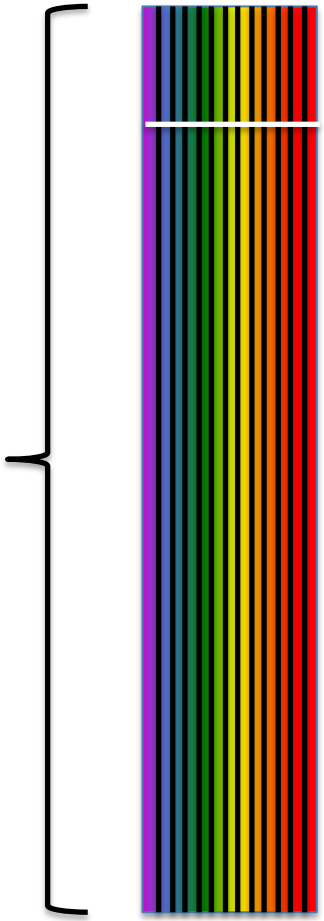
Strips are spliced together end to end to make a 2,395,261 x1 vector

Do the same thing to the other 12 slabs

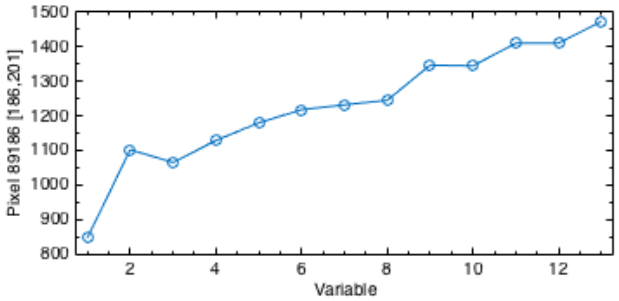
The Final 2,395,261 x 13 Matrix

13 channels (wavelengths)

Each Pixel of the original photograph is now an individual sample or row of the matrix



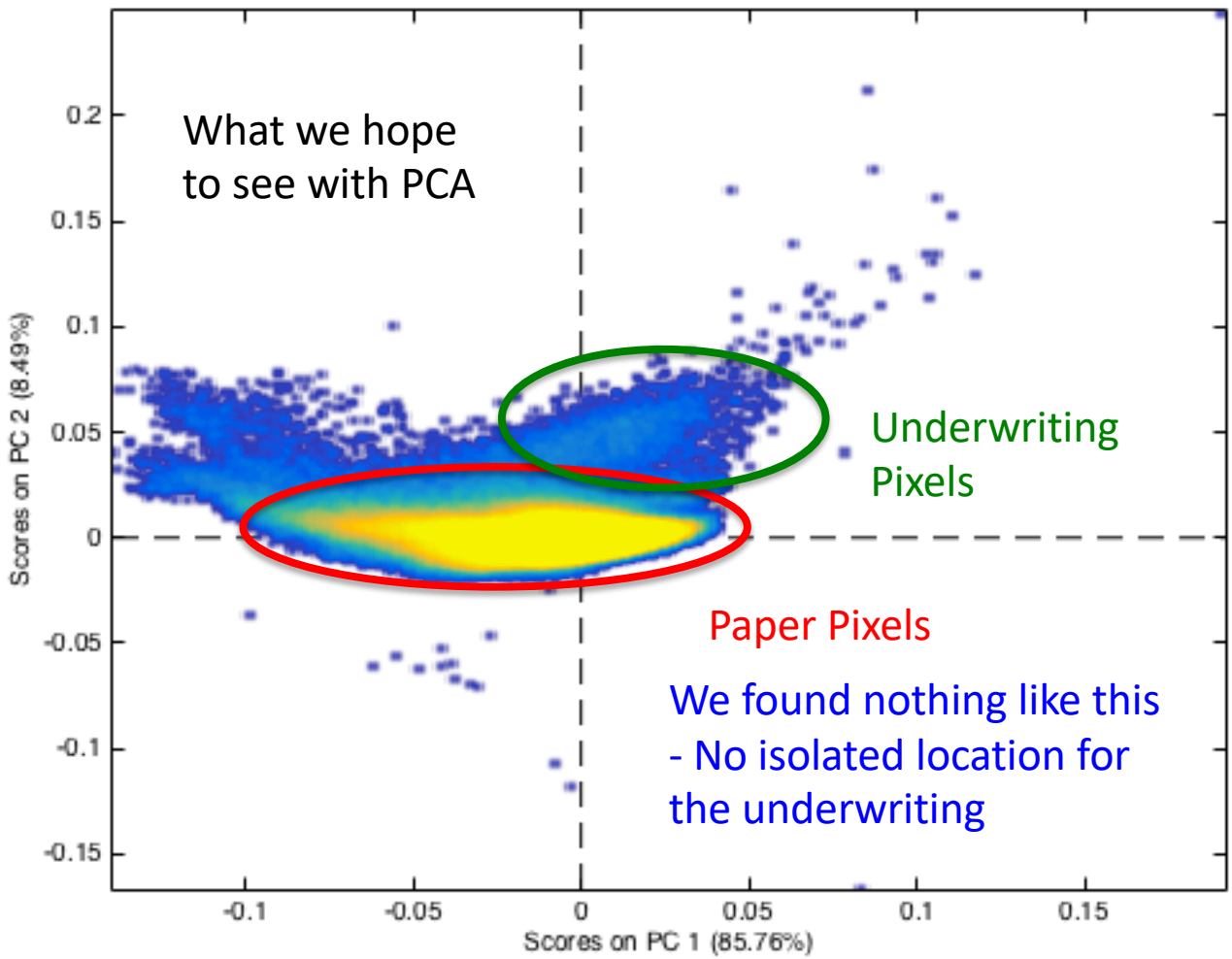
Each Row represents the Spectral Curve of that pixel



Now Perform a PCA of this matrix keeping track of where each sample (pixel) was located in the original photograph

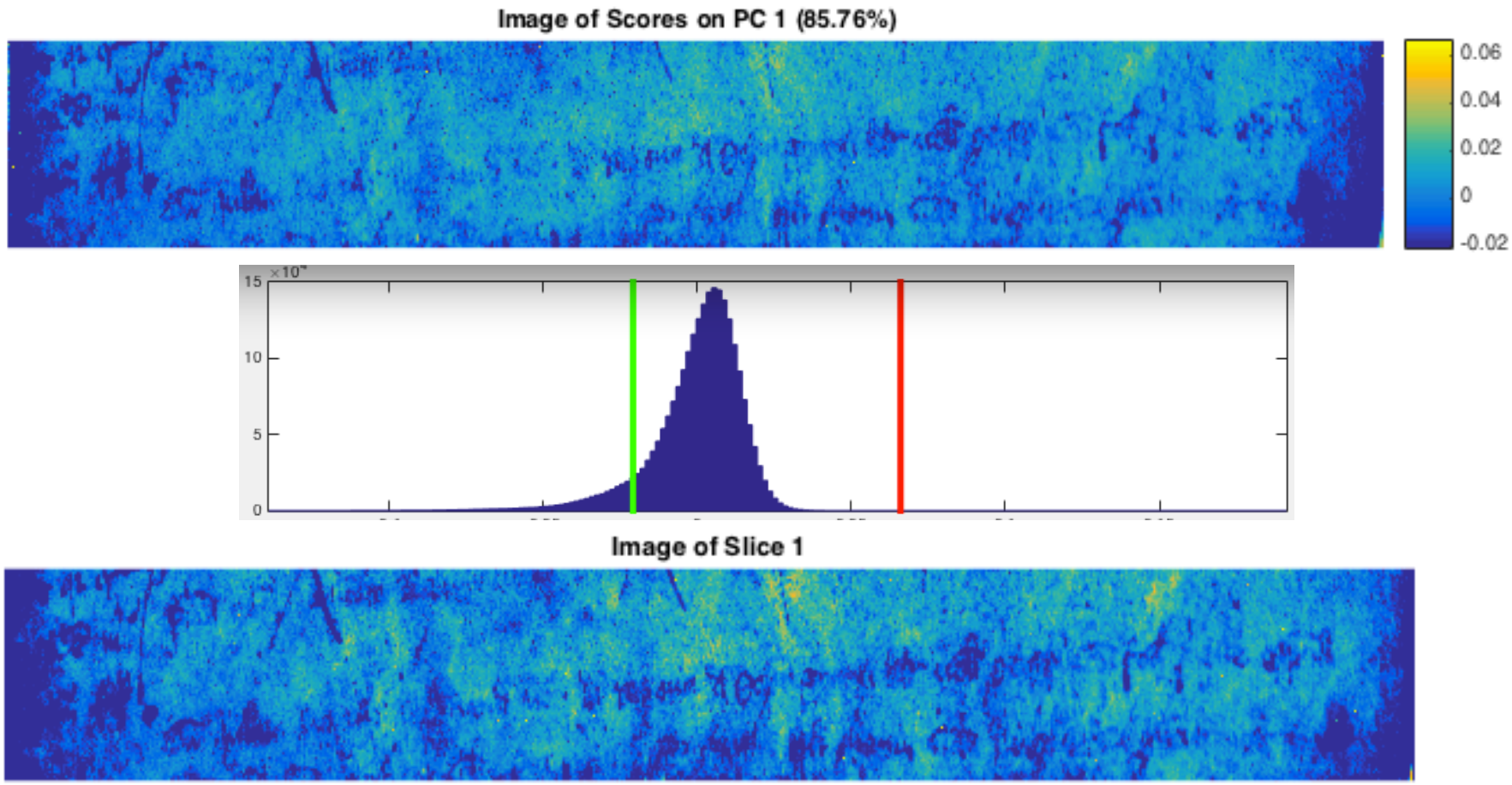
PCA Scores Plot of PC2 vs PC1

Similar Pixels fall nears each other in a scores plot



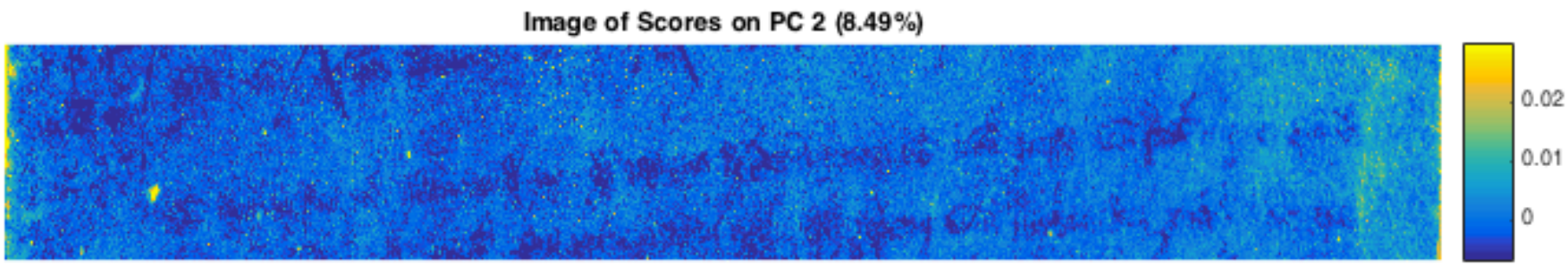
94.2% of variance in first 2 PCs

PCA - PC1 Scores Images



PC1 is not much of an improvement over Slice 1 - 365 nm

PC2 Scores Image Still Contains Information on Underwriting



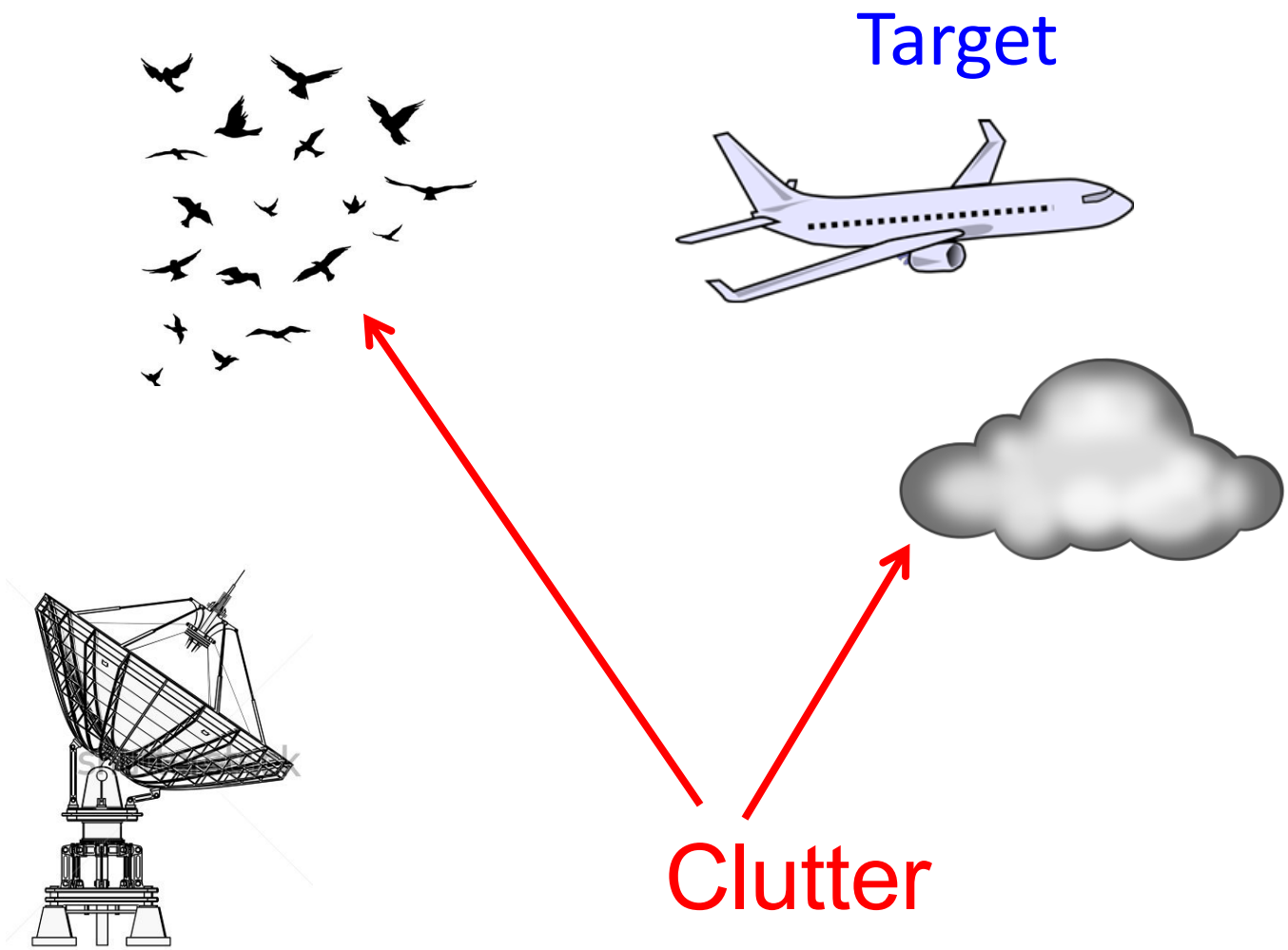
PCA alone is not doing a very good job of concentrating and segregating the underwriting from the paper.

It is at this point those new to Principal Component Analysis give up on the approach

When Washing Clothes
Sometimes Detergent is Not Enough
You may need Pretreatments

And You Need a Different Pretreatment for Different Stains

Using Radar to Track an Airplane



Generalized Least Squares Weighting (GLSW) is a pretreatment technique to help remove the Clutter

Generalize Least Squares Weighting

Select Regions of Clutter

Birds

Clouds

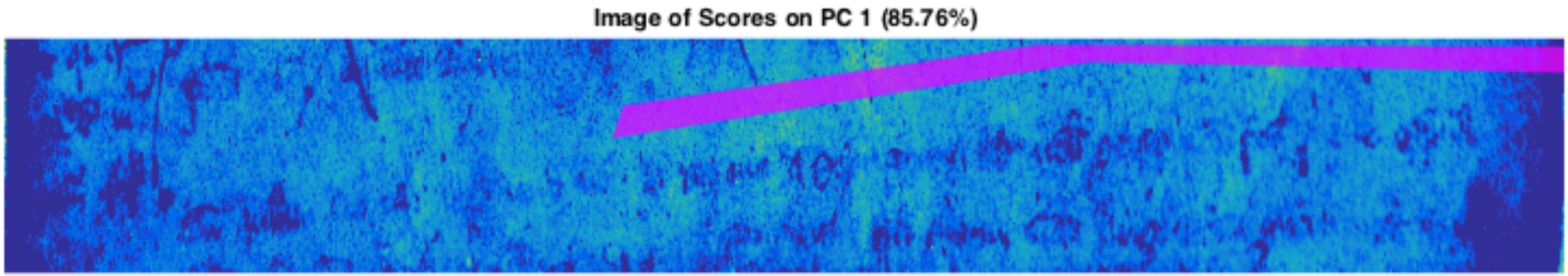
Create Classes for these Clutter Regions

GLSW creates a PCA model of the Clutter Classes

Deweights Variables Important to distinguishing the Clutter from the sky

Finally we build a PCA model of the whole image using these deweighted variables

Select a Region of Paper Where there is No Apparent Underwriting

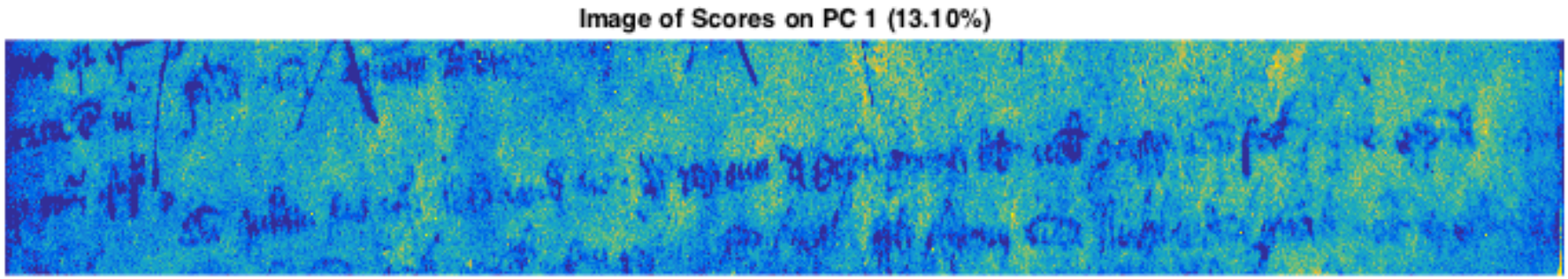


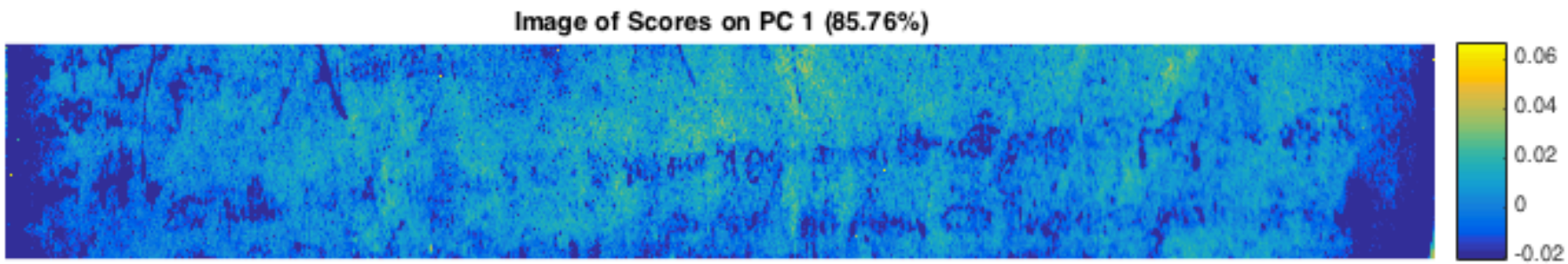
Define a Class “Paper”

GLSW will build a PCA model of the class, “Paper” and deweight the wavelengths important to the variance in the selected pixels.

A PCA model is then built for the entire image using these altered signals

PC1 - PCA Scores Image
(GLSW, MC)





PC1 - PCA Scores Image (1-Norm, MC)

Section 9 is More of a Challenge Because of the Overwriting

Image of Slice 1

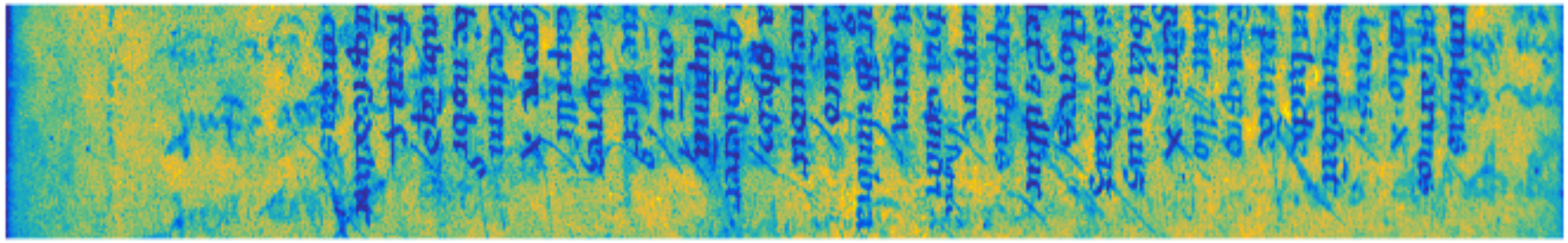


Image of Scores on PC 1 (88.78%)

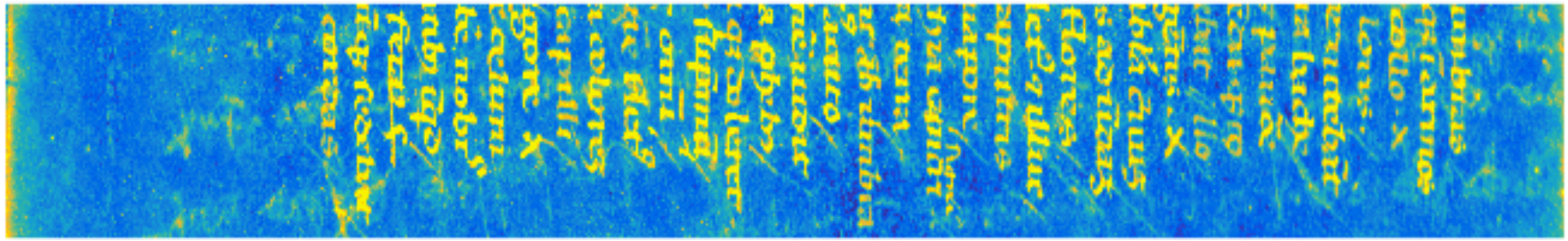
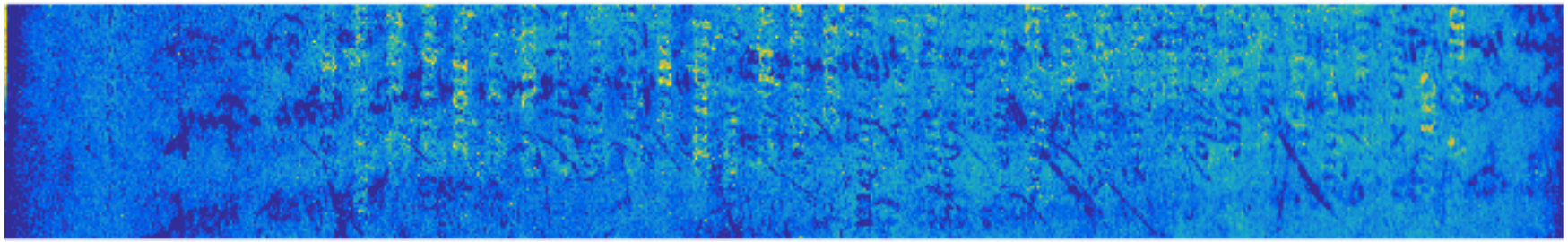
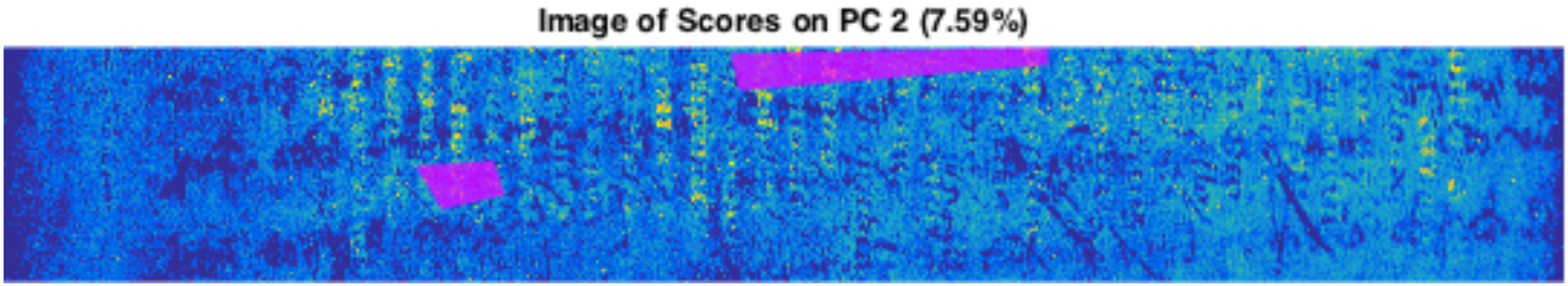


Image of Scores on PC 2 (7.59%)



PCA (1-Norm, MC)

Select Region Containing Paper and Overwriting

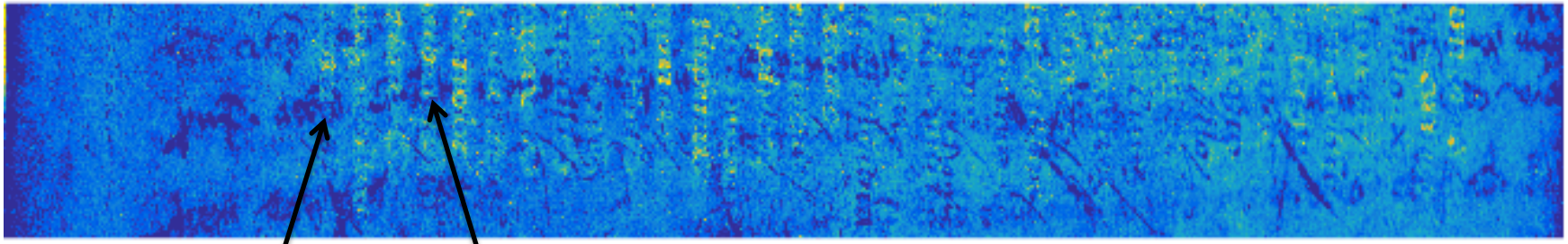


We are telling GLSW to deweight variables that important to distinguishing the overwriting from the paper.

Scores Images of Underwriting

PCA (1-Norm, MC)

Image of Scores on PC 2 (7.59%)

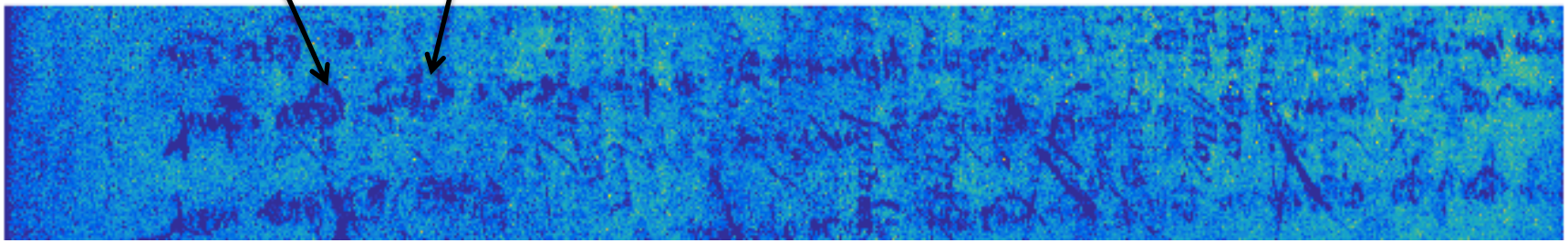


Other regions where overwriting is removed

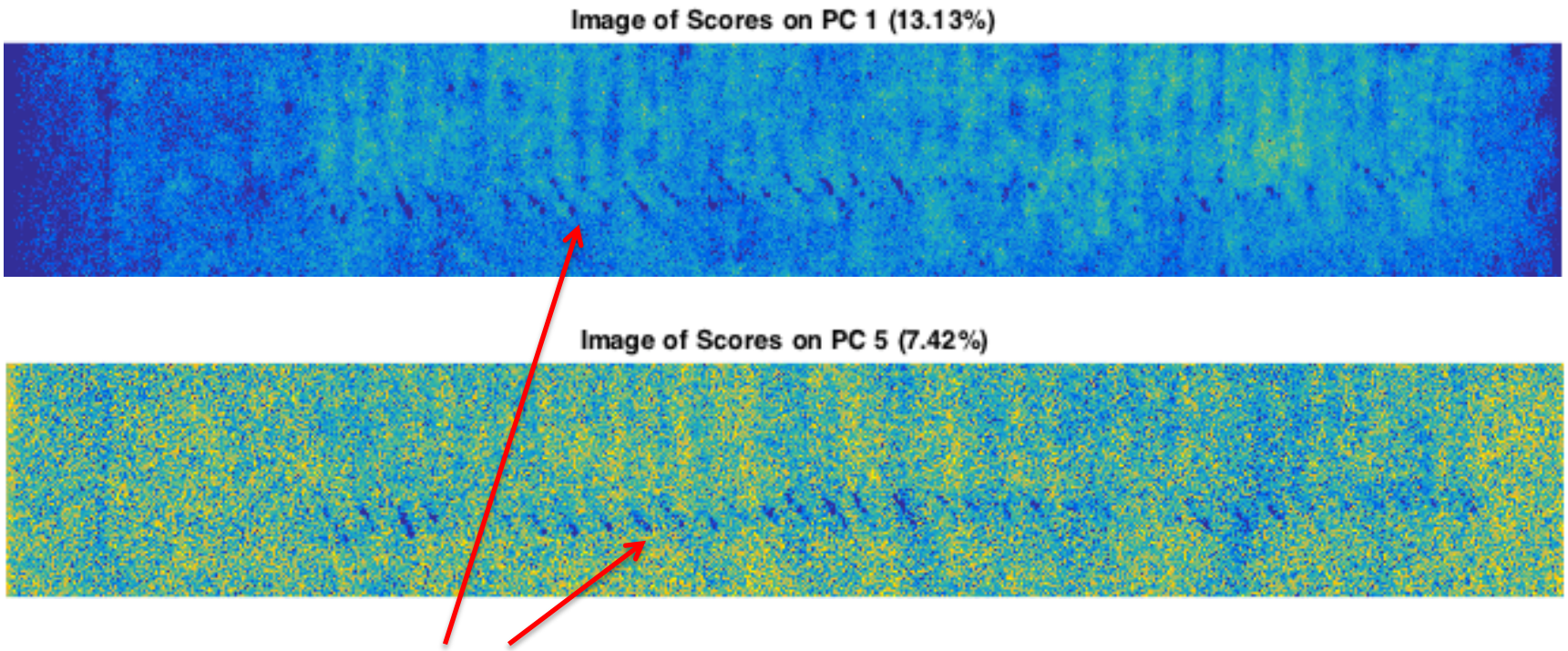
Regions where underwriting is revealed

PCA (1-GLSW, MC)

Image of Scores on PC 3 (9.27%)



PC1 and PC5 Reveal a Third Ink



There appears to be a third ink used to cross out some of the previous underwriting.

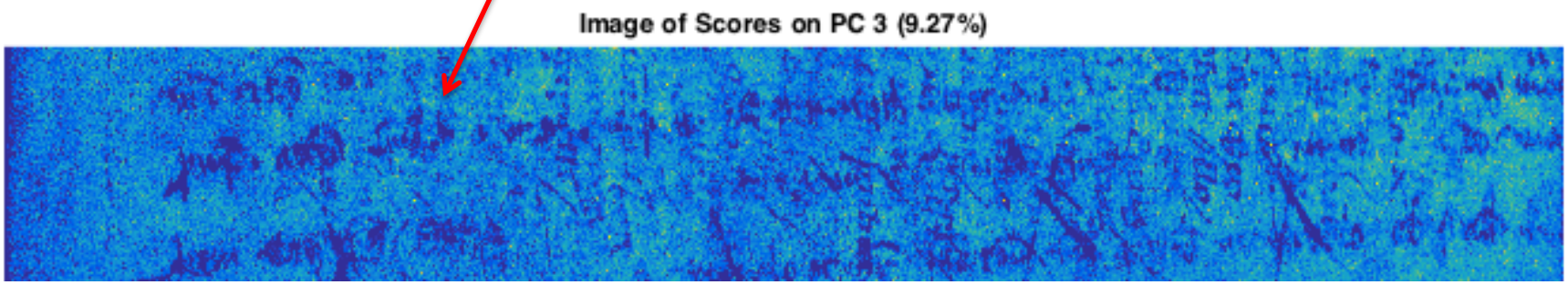
The overwriting has been moved to later PCs, because GLSW told the model to deweight those wavelengths.

GLSW Summary

- GLSW deweights variables (channels) that are important to variations in the clutter class – Paper.
- So the variables that are important to wrinkles, paper texture, uneven lighting (and overwriting) are being de-weighted.
- Hopefully this will leave variables important to the underwriting as more heavily weighted making it easier for the PCA to reveal the underwriting.

Electronic Spectra (UV-Vis region) will only provide so much information

Overwriting is masking the underwriting



Now that I know areas of interest, I would hit these areas with multispectral imaging using FTIR, NIR, Raman or even XRF, which has more chemical information that will detect small differences in the composition of the inks.

Neal and I would like to Thank Fenella France, Meghan Wilson and the Library of Congress for sharing this high quality and really neat data.

Lebanon Valley College

The LIBRARY of CONGRESS

EIGENVECTOR RESEARCH INCORPORATED