



**DEPARTMENT OF MINING GEODESY AND ENVIRONMENTAL ENGINEERING**

**Theme:** “Hyperspectral processing with ENVI”

**Author:** Bartosz Augustyn, Tymoteusz Maj, Michał Walencik, Gabriela Bebak, Wiktoria Gądek

**Field of study:** Remote Sensing and GIS

Kraków, 2024

## Table of contents:

<b>Project assumptions</b>	<b>3</b>
<b>Data loading</b>	<b>3</b>
<b>Working with data</b>	<b>4</b>
Recording data as NTIF	4
Hyperion data preprocessing	5
Removal of artifacts from hyperspectral data	5
Data format conversion: BSQ to BIL	5
Atmospheric correction: FLAASH	6
Atmospheric correction: QUAC	8
Smiling removal	8
Principal Component Analysis (PCA)	10
Channel selection without interference	10
<b>Classification</b>	<b>12</b>
Supervised classification	12
Unsupervised classification	14
Summary and conclusions	15
<b>Environmental indicators</b>	<b>16</b>
Vegetation Index Calculator	16
ENHANCED VEGETATION INDEX	17
NDWI	17
MOISTURE STRESS INDEX	18
Band Math	18
Intensity	18
<b>Forest Health</b>	<b>19</b>
<b>Global spatial statistics</b>	<b>20</b>
<b>Object detection using spectral curves</b>	<b>21</b>
Topaz	21
Asbestos	23
Construction Concrete	23
<b>Submission</b>	<b>25</b>

## Project assumptions

The aim of the project was to learn about ENVI software and methods for processing hyperspectral data. The project used hyperspectral data from the Hyperion satellite, covering the area of the steel mill in Krakow and adjacent areas. The processing focused on exploring the functions of the ENVI program, with particular emphasis on the built-in spectral library. In addition, a basic detection analysis was carried out to identify characteristic surface features in the analyzed area.

## Data loading

The data in question was received in .enp, .hdr and no extension formats, which turned out to be in .dat format by default, which is the standard data format for ENVI software. The .enp format is the spectral data file, the .hdr contains metadata about the image, such as size and resolution, and the .dat format is the actual data file that stores information about the pixel values of the hyperspectral image.

The screenshot shows the ENVI Data Manager window. At the top, there is a file list with three entries:

File Name	Created	Type	Size
Hyperion_25_06_2006_Huta	13.11.2024 13:34	Plik	61 952 KB
Hyperion_25_06_2006_Huta.enp	26.11.2024 22:27	Plik ENP	15 779 KB
Hyperion_25_06_2006_Huta.hdr	13.11.2024 13:34	Plik HDR	20 KB

Below the file list is a tree view of the dataset structure:

- Hyperion\_25\_06\_2006\_Huta
  - (256x242x7205):EO1H1880252006176110KY.L1R: Band 1 (355.5900)
  - (256x242x7205):EO1H1880252006176110KY.L1R: Band 2 (365.7600)
  - (256x242x7205):EO1H1880252006176110KY.L1R: Band 3 (375.9400)
  - (256x242x7205):EO1H1880252006176110KY.L1R: Band 4 (386.1100)
  - (256x242x7205):EO1H1880252006176110KY.L1R: Band 5 (396.2900)
  - (256x242x7205):EO1H1880252006176110KY.L1R: Band 6 (406.4600)
  - (256x242x7205):EO1H1880252006176110KY.L1R: Band 7 (416.6400)
  - (256x242x7205):EO1H1880252006176110KY.L1R: Band 8 (426.8200)
  - (256x242x7205):EO1H1880252006176110KY.L1R: Band 9 (436.9900)
  - (256x242x7205):EO1H1880252006176110KY.L1R: Band 10 (447.1700)

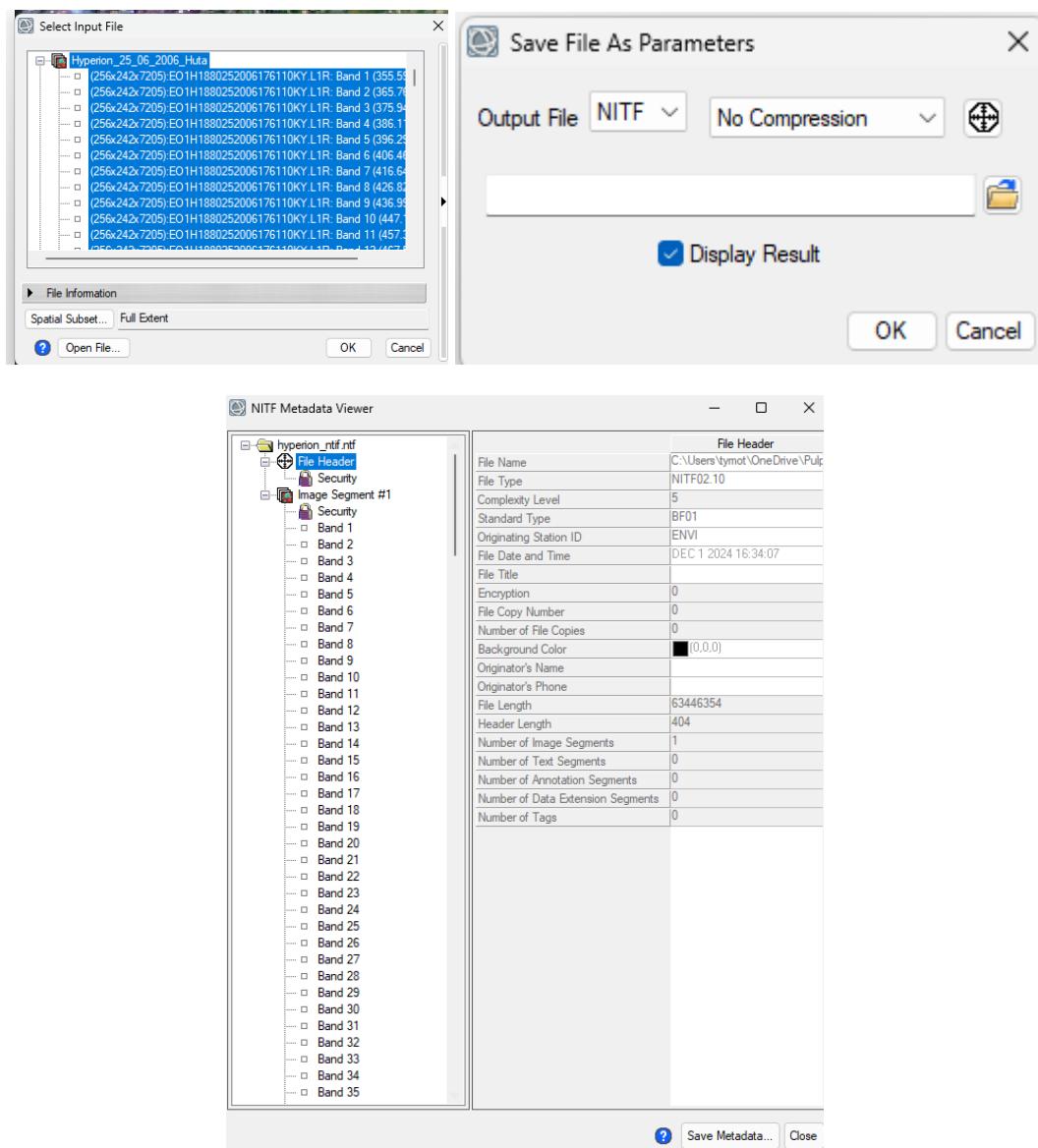
At the bottom of the window, there are buttons for "File Information" and "Band Selection", and a "Load Data" button.

For unknown reasons, the supplied file in .dat format caused difficulties during work, so we began our work with hyperspectral data by re-saving the file in the default ENVI format, which enabled further processing and analysis of the data without the problems encountered earlier.

# Working with data

## Recording data as NTIF

The first operation we decided to perform was to test saving the file to a format other than the default ENVI, so we decided to save our file in NITF (National Imagery Transmission Format). NITF is a format used primarily in military and government applications for storing satellite imagery and geospatial data, which allows both imagery and metadata to be stored in a single file. After saving the file in this format, we decided to read the image metadata, which was previously unavailable. Unfortunately, the available metadata did not deepen our understanding of the data.



## Hyperion data preprocessing

Pre-processing of Hyperion hyperspectral data was carried out to eliminate artifacts related to the quality of the raw data and prepare it for further analysis. This stage included a series of steps to improve data quality, enable atmospheric correction and prepare for spectral analysis.

### Removal of artifacts from hyperspectral data

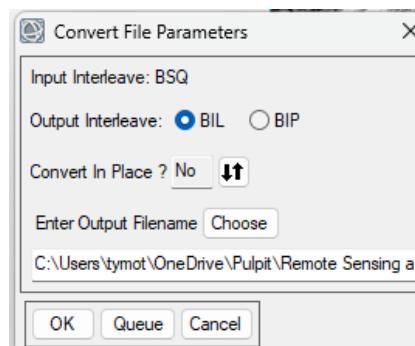
The first step was to remove noise generated by damaged detectors, such as the striping effect. To do this, the Destripe algorithm was used, which corrects for differences in pixel intensity between neighboring detectors.

The striping effect was identified as a significant disturbance affecting the quality of Hyperion data, and its elimination improved image homogeneity, which was crucial for further analysis.

### Data format conversion: BSQ to BIL

Hyperion's input data was in Band Sequential (BSQ) format, which records all the data for one band before moving on to the next band. Due to the requirements of the FLASSH program, which operates on the Band Interleaved by Line (BIL) format, it was necessary to change the data interleav.

The conversion from BSQ to BIL was done to allow for atmospheric correction. The BIL format provides better processing performance when analyzing multiband data, which is crucial for atmospheric correction and subsequent spectral analysis.

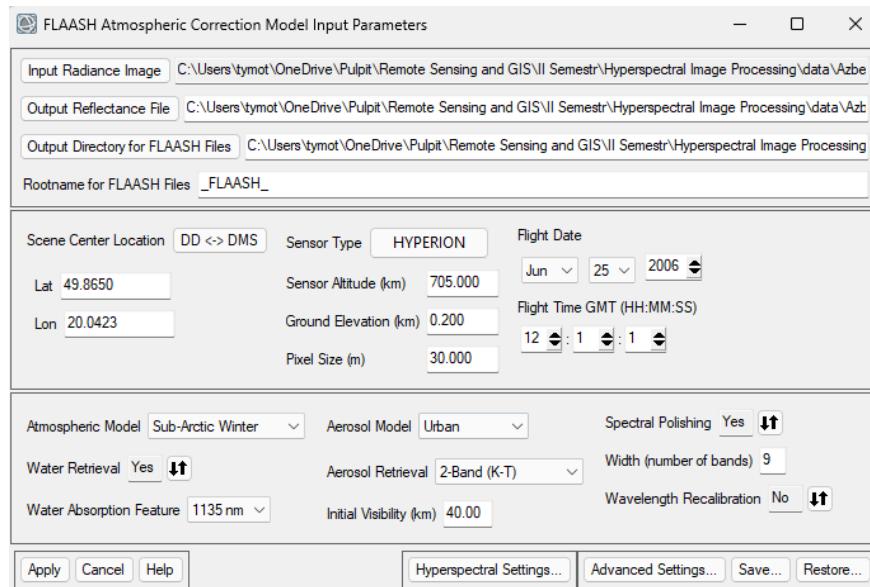


## Atmospheric correction: FLAASH

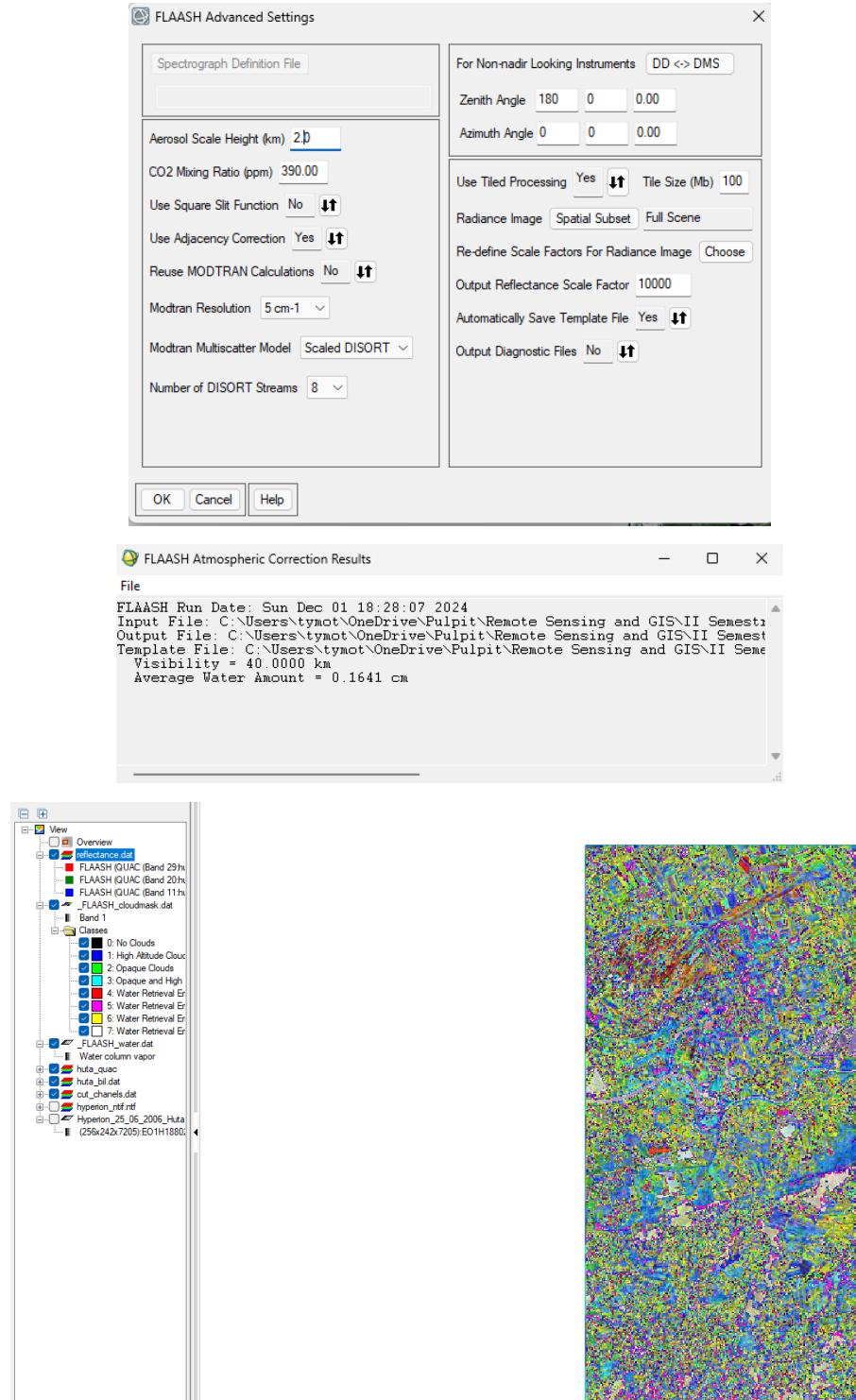
Atmospheric correction was performed using the FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes) algorithm available at ENVI. The parameters used in this process were taken directly from the dissertation of the presenter, which ensured the accuracy and reliability of the operations performed.

Tab 14. Parametry wykorzystane do korekcji atmosferycznej w programie FLAASH.

Data rejestracji	25 czerwiec 2006
Współrzędne centrum sceny	20° 02' 34.08, 49° 51' 54.01
Wysokość orbity satelity	705 km
Rozdzielcość przestrzenna obrazu	30 m
Wysokość terenu nad poziomem morza	200 m
Widoczność	40 km
Model aerozolu	Urban
Model atmosferyczny	<i>Sub-Arctic Summer</i>
Efektywna wysokość profilu pionowego areozolu	2 km
Kanał absorpcji wody	1135 nm
Stężenie CO <sub>2</sub>	390 ppm
Modtran resolution	1 cm <sup>-1</sup>



## Hyperspectral processing with ENVI



The result of the atmospheric correction was data in units of surface reflectance (reflectance), which is the basis for analyzing the physical properties of objects on the Earth's surface.

## Atmospheric correction: QUAC

In addition to atmospheric correction using the FLASSH algorithm, an alternative method, QUAC (Quick Atmospheric Correction), was also used. QUAC is a fast algorithm that automatically estimates the influence of the atmosphere based on the image spectrum, without the need to enter atmospheric parameters.

## Smiling removal

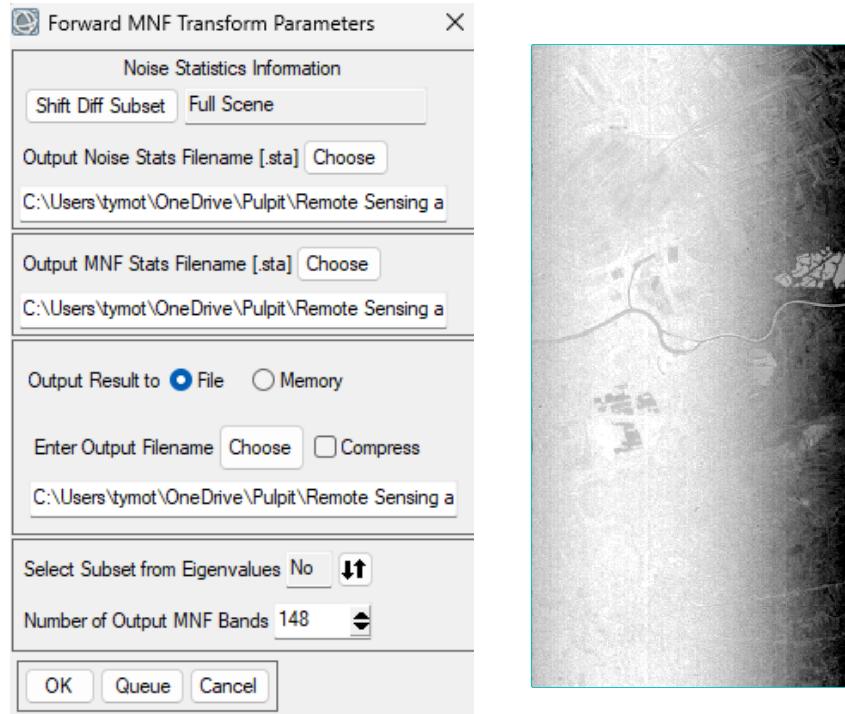
The smiling effect is a phenomenon occurring in hyperspectral data that manifests itself as a nonlinear spectral shift along the spatial axis of the image. This causes distortions in spectral analysis, especially for high-spectral resolution data such as that from the Hyperion instrument. To ensure the correctness of subsequent analyses, it was necessary to eliminate this artifact.

The process of removing the smiling effect was carried out in two steps, using the Minimum Noise Fraction (MNF) transformation and its inverse, the Inverse Minimum Noise Fraction (IMNF):

- **Visual analysis of MNF channels**

In the first step, the MNF transformation was performed on the hyperspectral data. This transformation sorted the image channels by signal-to-noise ratio (SNR), allowing easier identification of interference.

Channels with low signal-to-noise ratios were visually analyzed to identify those that contained smiling artifacts. These channels were characterized by nonlinear data distortions that could interfere with subsequent spectral analysis processes.



- **IMNF inverse transformation without channels with smiling distortion**

Once the channels with distortions were identified, an IMNF inverse transformation was performed, omitting these channels. This process allowed the image data to be reconstructed in its original form, but without the interference caused by the smiling effect.

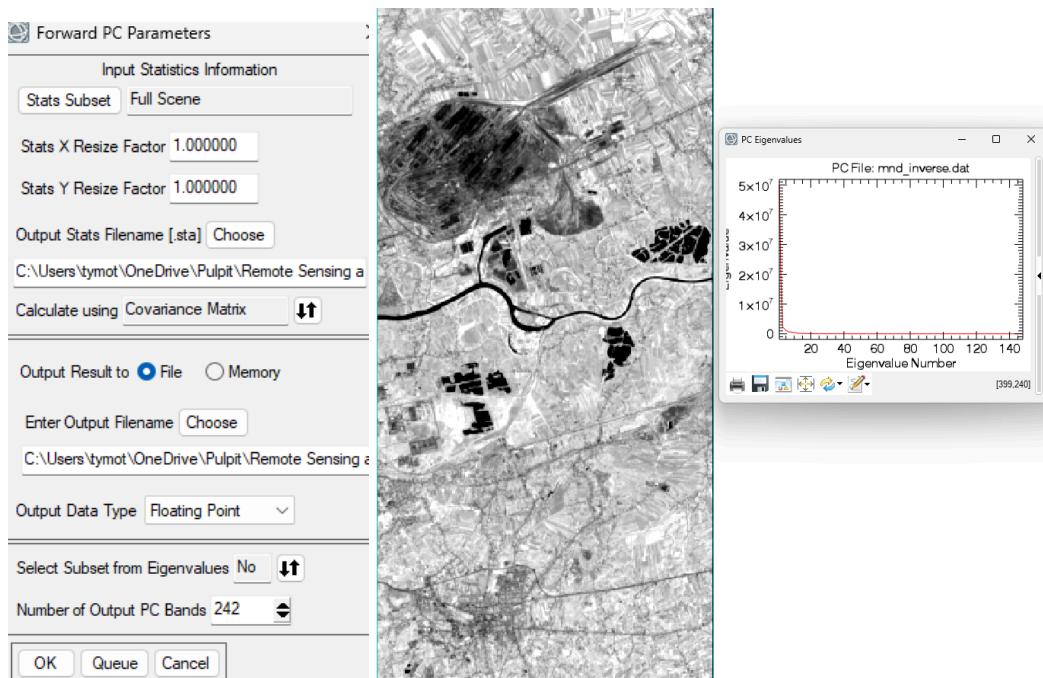


## Principal Component Analysis (PCA)

To further process the data, principal component analysis (PCA) was performed. PCA made it possible to reduce the dimensionality of the data and identify and potentially eliminate distortions such as the smile effect.

The results of the PCA analysis enabled:

- Identify major sources of variance in the data, including artifacts.
- Removing the component corresponding to distortions (e.g., smile effect) and reconstructing the data based on the remaining components.



## Channel selection without interference

The first operation we performed when saving the file was to apply the spectral subset command, designed to remove hyperspectral channels with interference. Distorted channels in Hyperion data typically occur in certain spectral ranges, such as the UV range and part of the SWIR range, where there is a low signal-to-noise ratio (SNR). Channels at the edges of the spectrum, including the transition between VNIR and SWIR, also have high noise. In addition, artifacts may appear in the data due to atmospheric absorption (e.g., water vapor) in certain ranges.

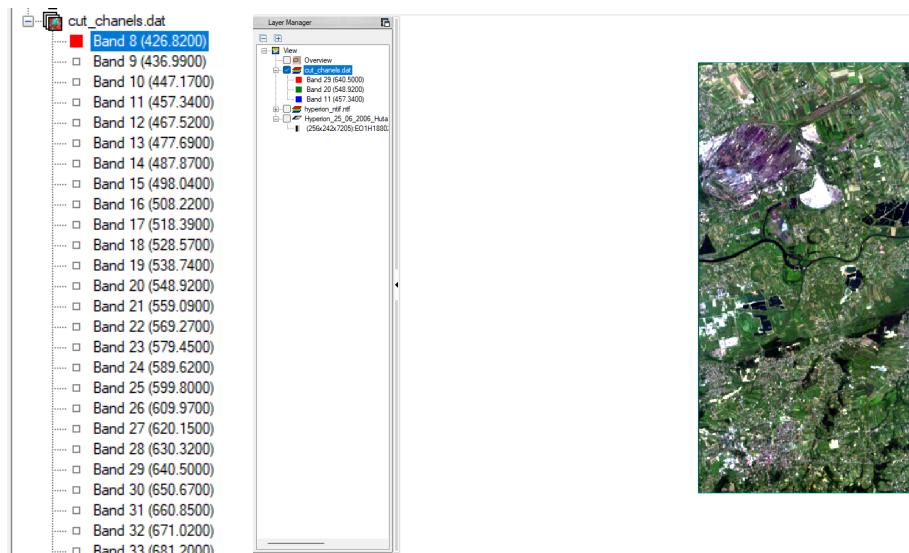
Based on the technical description of the Hyperion sensor and visual analysis of the channels, we identified the following ranges that contained interference:

- VNIR range (400-1000 nm): Channels 1-7, which show high interference in the UV range, and channels 58-78, which are in the transition between VNIR and SWIR and have high interference.
- SWIR range (1000-2500 nm): Channels 225-242, which have high IR interference and low SNR, and channels 193-198 and 207-213, which are affected by water artifacts and atmospheric absorption.

Based on these observations, we removed the disturbed channels, leaving the useful bands. After the selection operation, we finally obtained about 198 bands that are free of interference:

- VNIR channels: 8-57
- SWIR channels: 79-192, 199-206.

With this operation, we prepared the data for further analysis, eliminating disturbed bands that could affect the accuracy of the results.



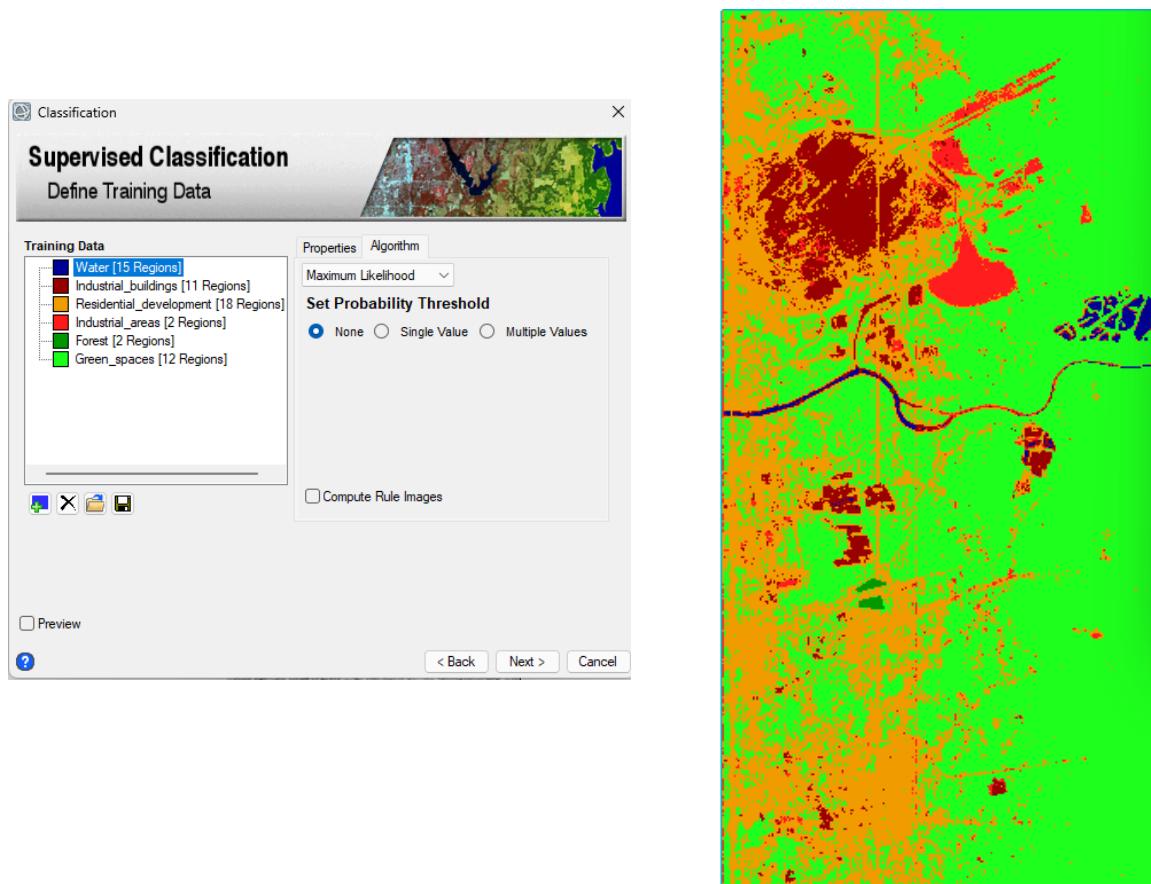
This operation made it possible to seamlessly load the base image into the RGB visualization, which is now done automatically and without difficulty. By removing the disturbed channels, the image became clearer and ready for further analysis. At the same time, we are aware that the manual selection may have missed some channels containing interference, which may be due to human error, but we tried to analyze the data as carefully as possible.

# Classification

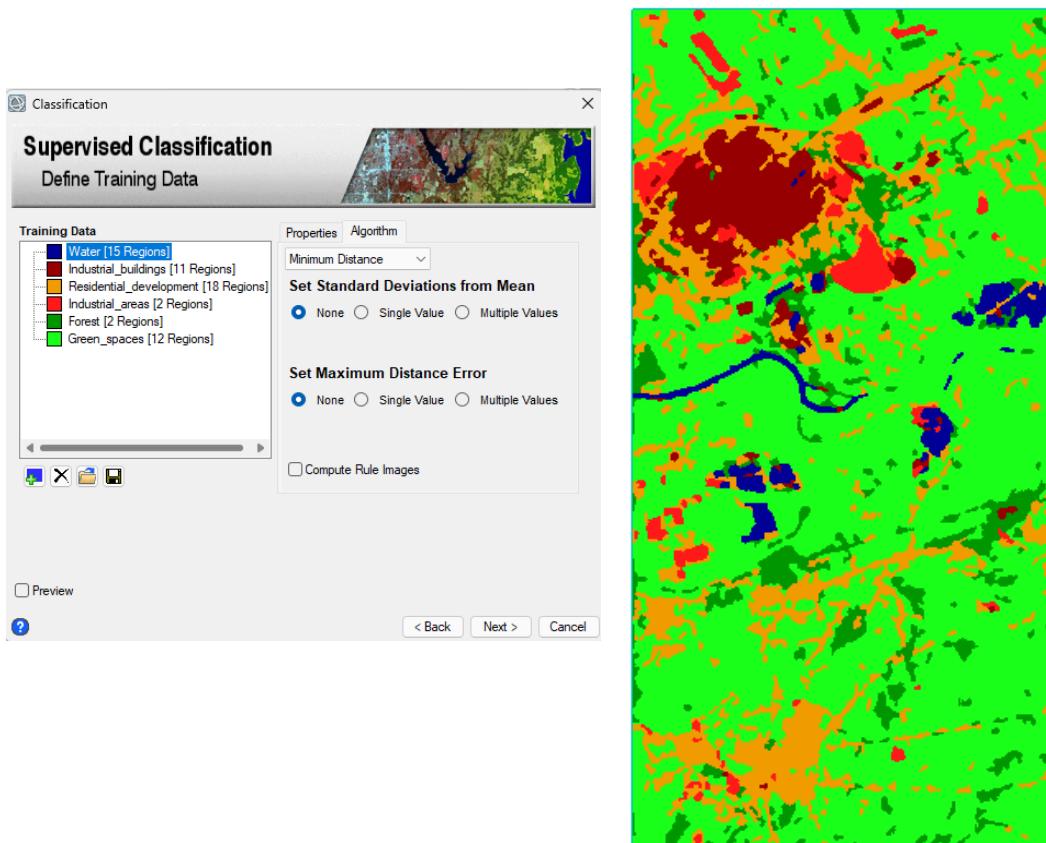
As part of the processing of hyperspectral data from the Hyperion instrument, image classification was performed with the aim of assigning individual pixels to appropriate subject classes. Two approaches were used: supervised classification and unsupervised classification.

## Supervised classification

The supervised classification process was based on the manual delineation of training polygons that reflected known subject classes. These polygons were drawn based on visual interpretation of the image, based on what could be identified “at a glance.” Training sets were created for each of the six identified subject classes, which were used for classification using the Maximum Likelihood Classification (MLC) algorithm.



To further compare the classification results, an additional analysis was performed using the Minimum Distance Classification method. This classification method is based on calculating the distance between the average feature vectors for each class and assigning pixels to the class whose feature vector is closest. The results from the Minimum Distance method turned out to be superior, and we were more satisfied with the outcome, as it provided more accurate and consistent classification compared to the other methods.

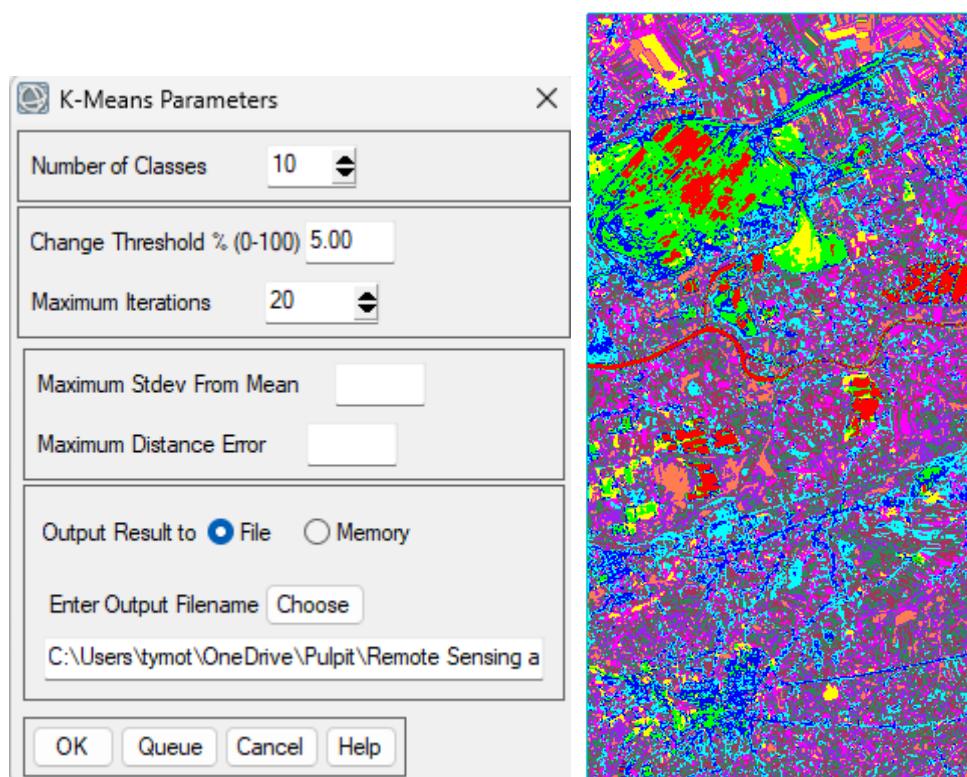


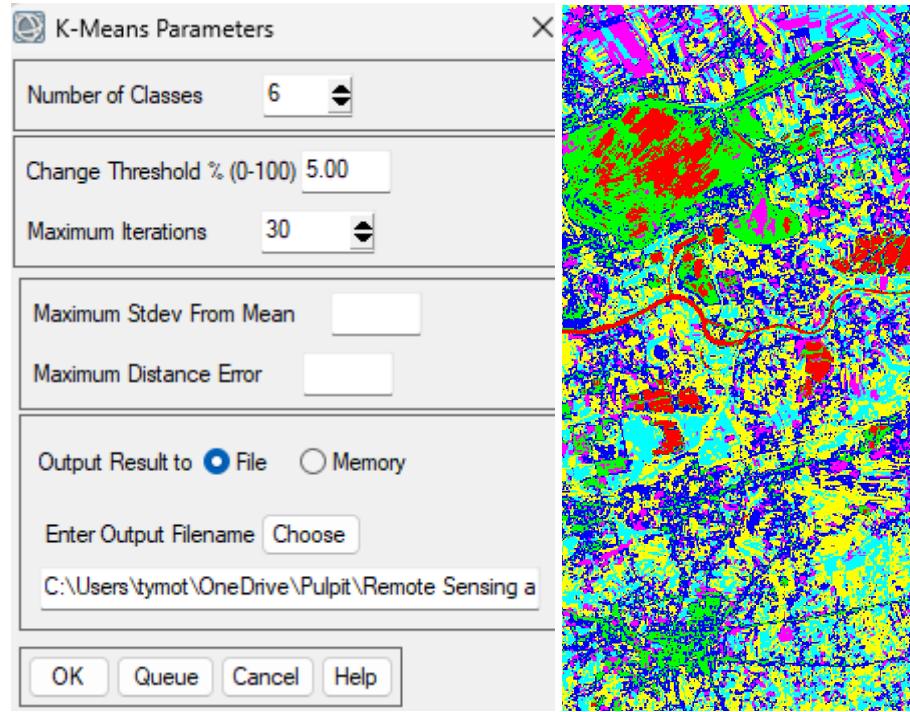
The results obtained with this method turned out to be almost identical, which suggests that the classification process in our case is very stable and independent of the chosen method. Therefore, we can conclude that possible classification errors are most likely due to the quality of the input data, particularly the quality of the hyperspectral images, as well as the limited representativeness of the training fields. Weak training fields, which do not fully reflect the variability in the data, may have contributed to inaccurate classifications.

## Unsupervised classification

To compare the results, unsupervised classification was also carried out, initially generating a large number of classes (e.g., 10) to capture the spectral diversity of the data. The classes were then aggregated, reducing the number of classes to six - in line with the number of classes defined in the supervised classification.

To improve the quality of the unsupervised classification, a higher number of iterations of the k-means clustering algorithm was used. Increasing the number of iterations allowed the algorithm to converge more precisely, which is due to the fact that in each iteration the boundaries between classes in the spectral space are optimized. As a result, the algorithm had more opportunities to match the classes to the actual distribution of the data, which significantly improved the final results.





## Summary and conclusions

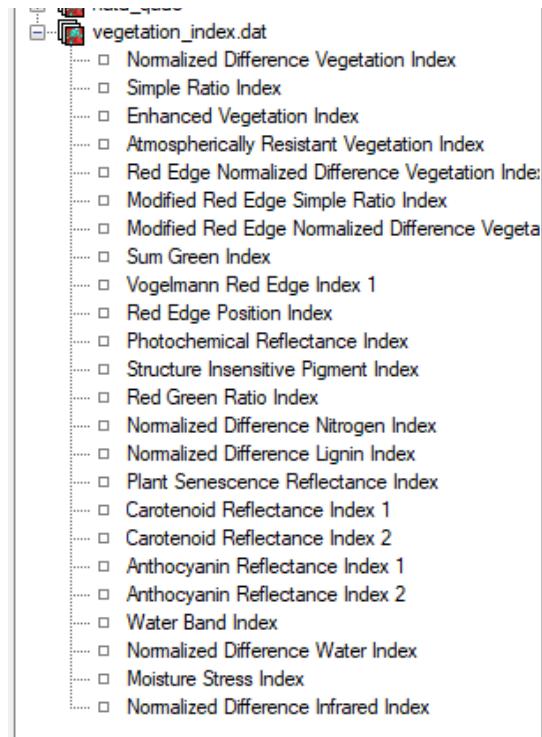
While supervised classification, particularly the Minimum Distance method, proved to be the most effective and gave the best results, unsupervised classification surprised us with its relative accuracy. The unsupervised method, although it did not initially produce results as clear as the supervised classification, showed some potentials that could be better exploited with further parameter tuning. We noted that increasing the number of iterations significantly improves the quality of classification, suggesting that further tuning of this parameter could contribute to even better results. While supervised classification remains the preferred method, unsupervised classification has proven that it can be a useful tool, especially in cases where we do not have access to sufficient training data. If this process could be optimized, the results could improve even further.

## Environmental indicators

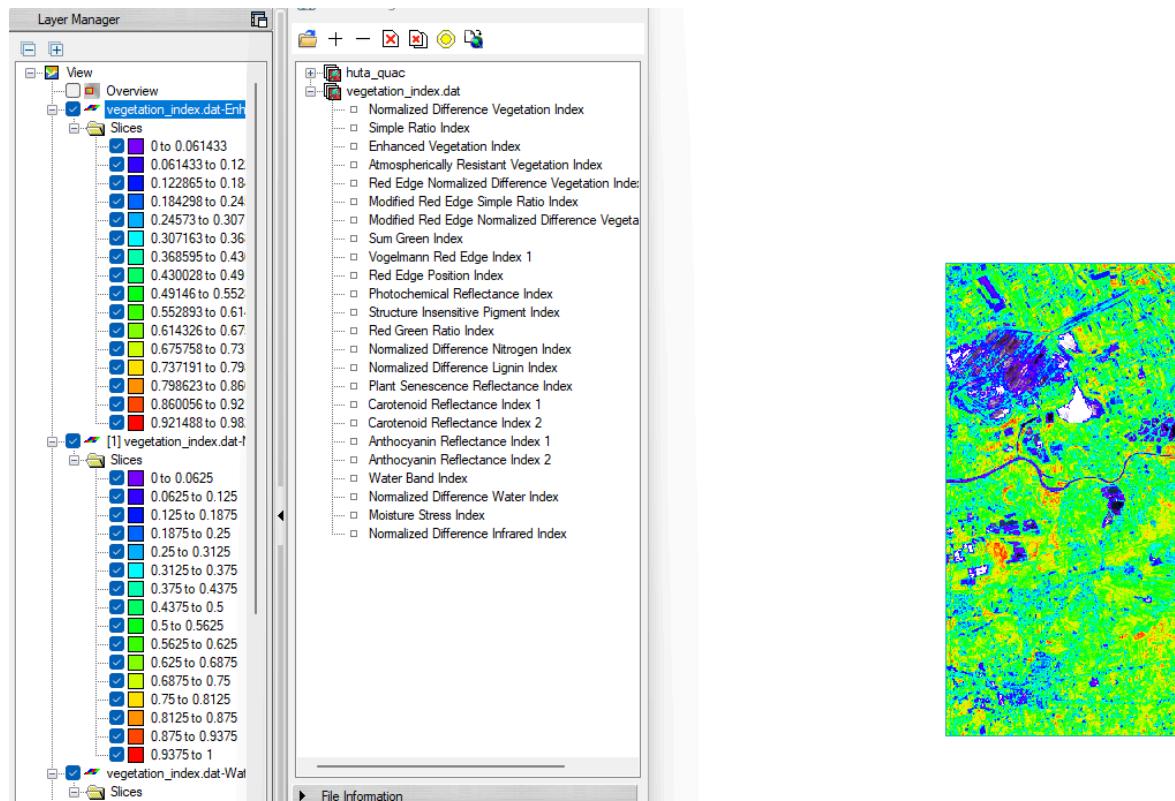
Environmental coefficients play a key role when working with hyperspectral data, as they allow for a better understanding and interpretation of surface characteristics such as vegetation, soil moisture or the presence of water. Therefore, we decided to include them in our analysis, although we did not focus too much on their impact on the final results. To simplify the calculation process, we used the Vegetation Index Calculator tool, which automatically calculates a number of environmental factors, such as vegetation indices, from hyperspectral data. In the screenshots below, we show an excerpt of the calculation results obtained by this method.

In addition, for comparison, we decided to calculate one of these coefficients manually using the Band Maths method, which allowed for a more controlled and precise calculation. The results of the calculations using the two methods are presented below, allowing us to compare them and evaluate the differences in the results.

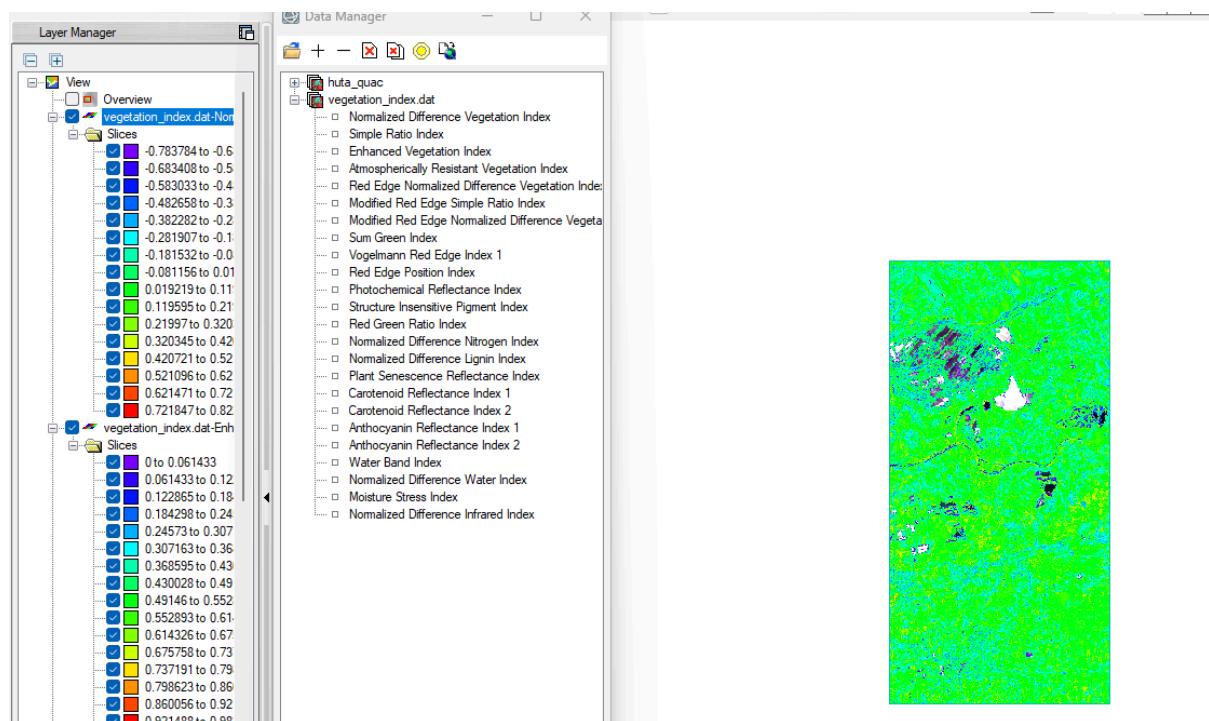
### Vegetation Index Calculator



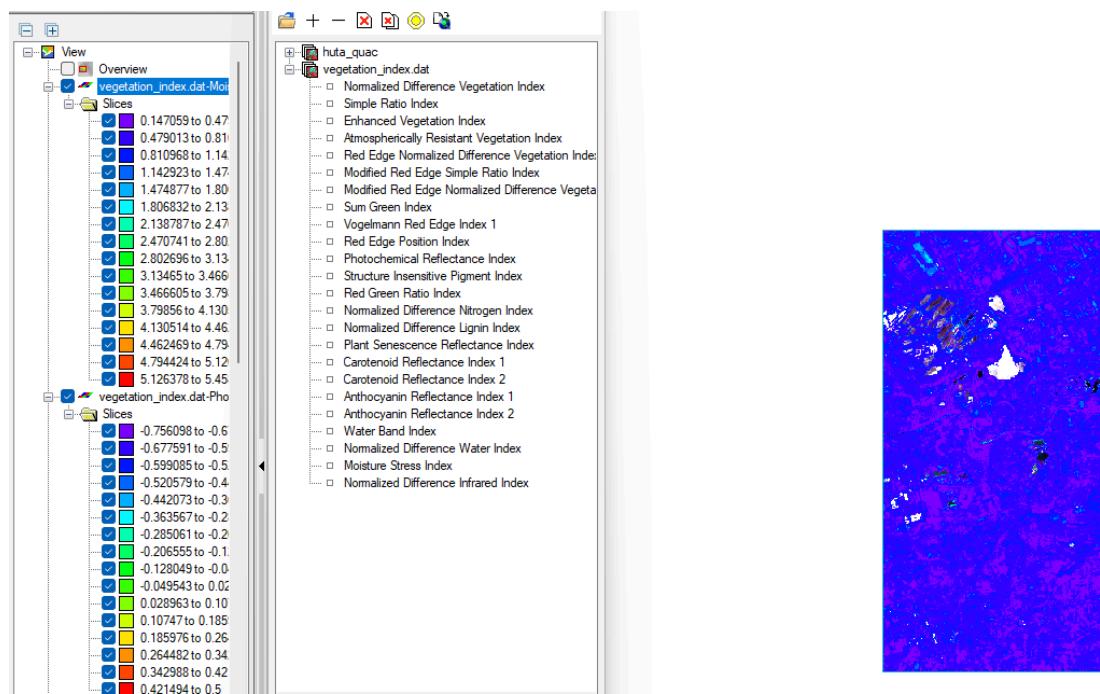
## ENHANCED VEGETATION INDEX



## NDWI



## MOISTURE STRESS INDEX

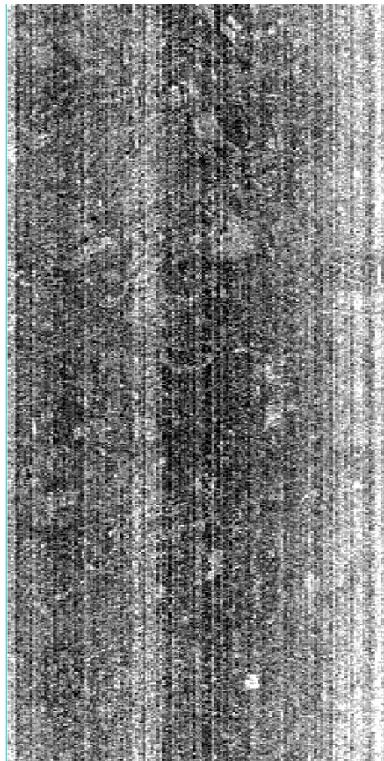
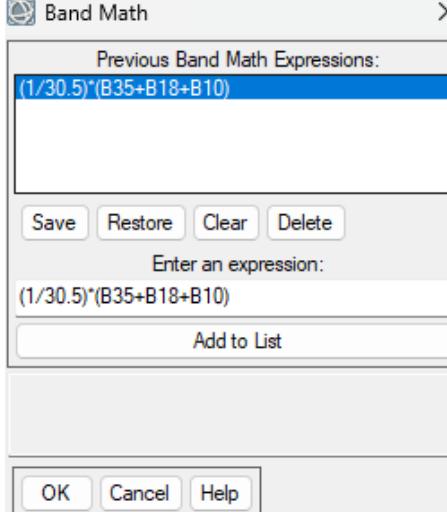


## Band Math

### Intensity

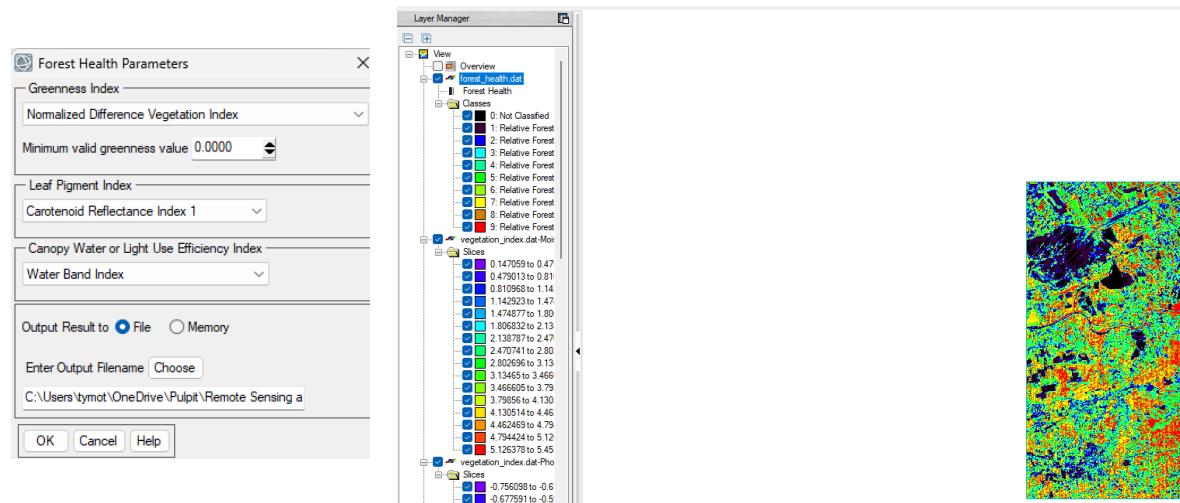
$$\left(\frac{1}{30.5}\right) (R + G + B)$$

$$\left(\frac{1}{30.5}\right) (B35 + B18 + B10)$$



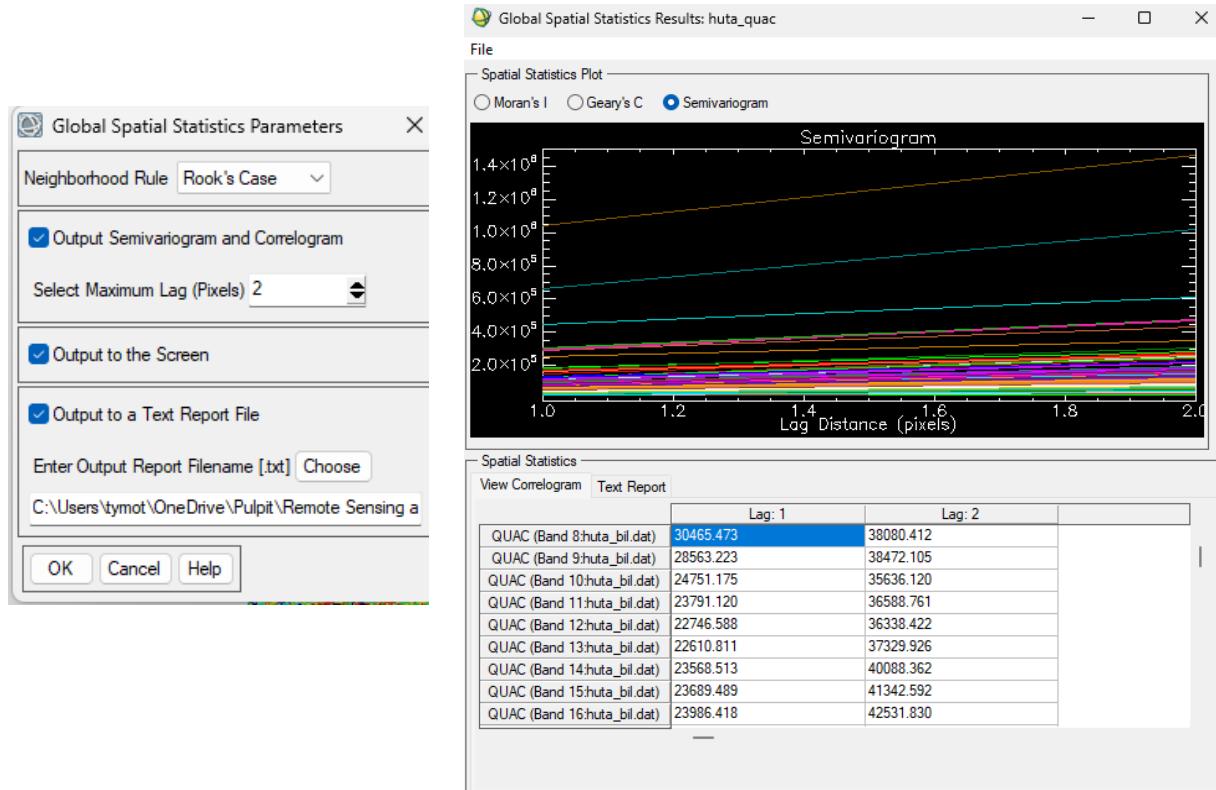
## Forest Health

The next step, which links directly to the calculation of environmental coefficients, was to test the Forest Health function against the Normalized Difference Vegetation Index (NDVI). This index, which is one of the most popular coefficients used in vegetation analysis, allows the assessment of vegetation health, including the identification of areas subject to environmental stress, disease or changes in soil quality.



## Global spatial statistics

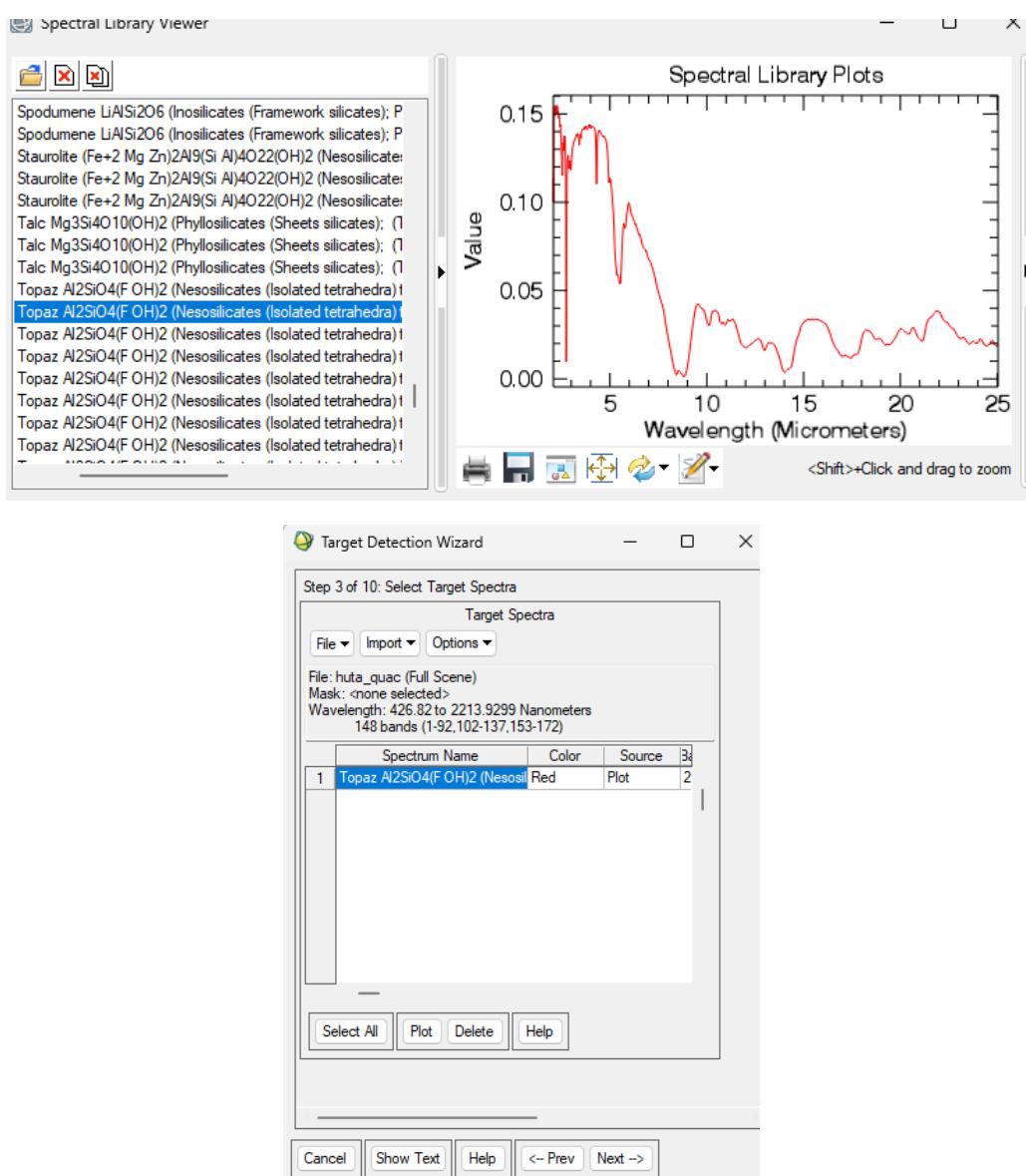
Finally, we conducted testing of global spatial statistics methods, which we performed for data just after atmospheric correction. As part of this process, we calculated a semivariogram that allowed us to assess spatial variability and understand the relationship between pixels at different scales.



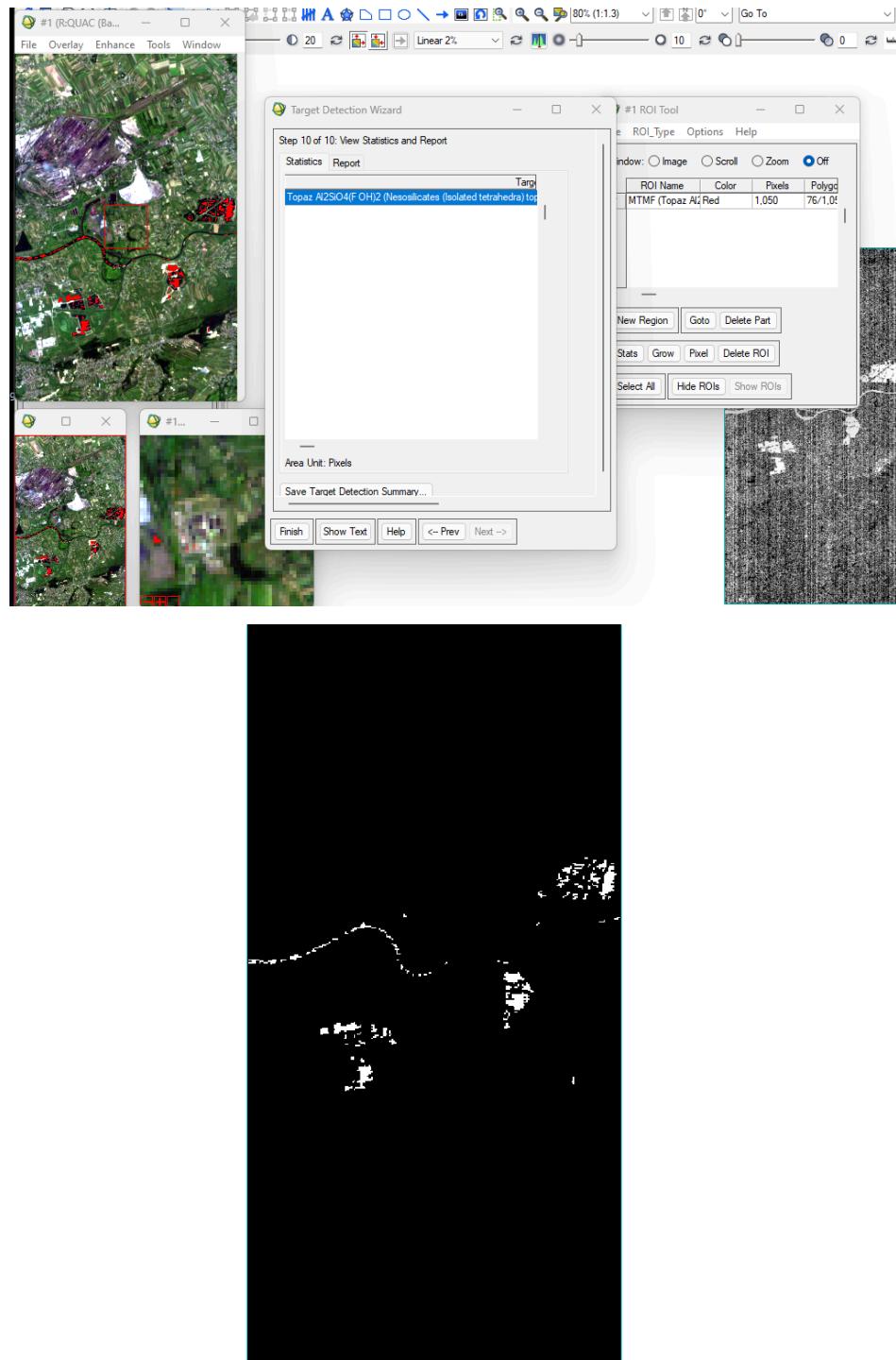
## Object detection using spectral curves

The last thing we wanted to test was target detection, based on spectral curves available in the spectral library at ENVI. We got the idea from an article (<https://www.nv5geospatialsoftware.com/Support/Maintenance-Detail/hyperspectral-analytics-in-envi-target-detection-and-spectral-mapping-methods>) that outlines the methodology and application of this type of analysis in the context of hyperspectral data. Using the available spectra from the library, we performed a target detection process to identify specific objects or features in the image based on their characteristic spectra.

### Topaz



## Hyperspectral processing with ENVI



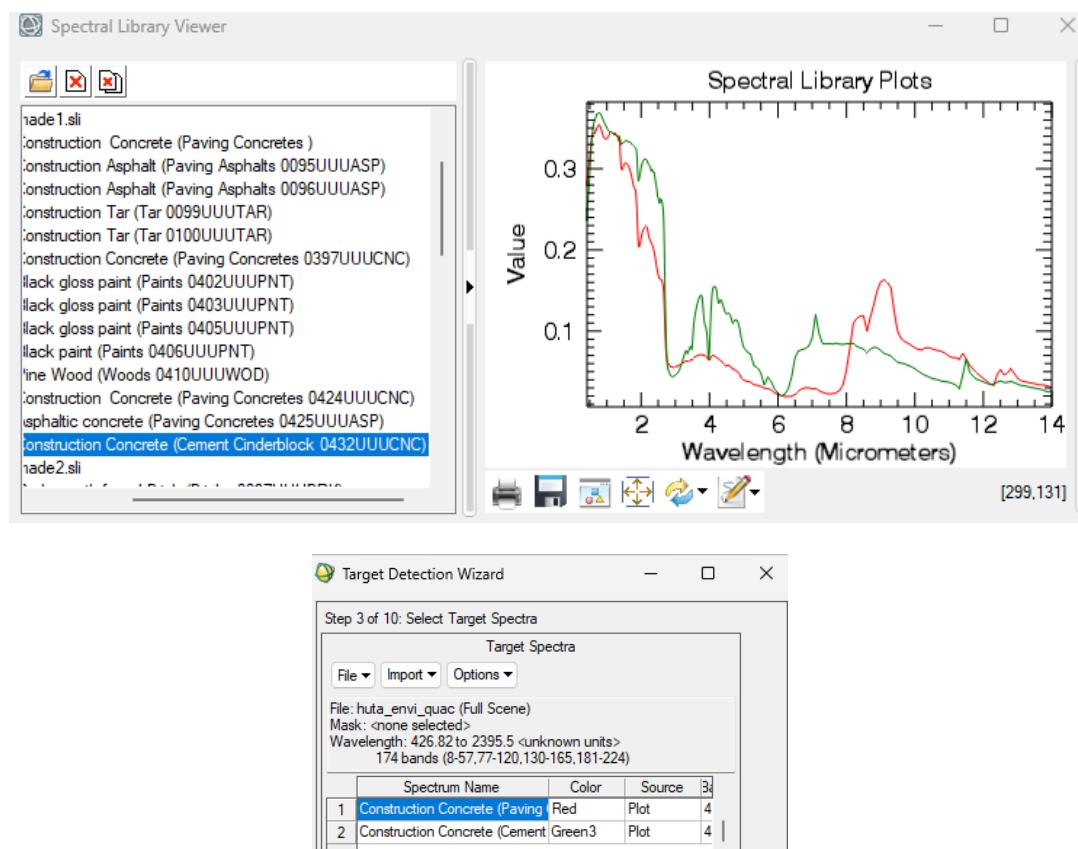
ENVI's built-in target detection function has identified Topaz in Krakow's water areas. This error may be due to several factors. First and foremost, water in hyperspectral data may exhibit similar spectral properties to those of materials detected as Topaz, especially if reflections or interferences in certain spectral bands are present. In addition, it is possible that the spectral library that was used in the detection process did not include all the spectral variables characteristic of different aqueous materials, which may have led to misclassification.

## Asbestos

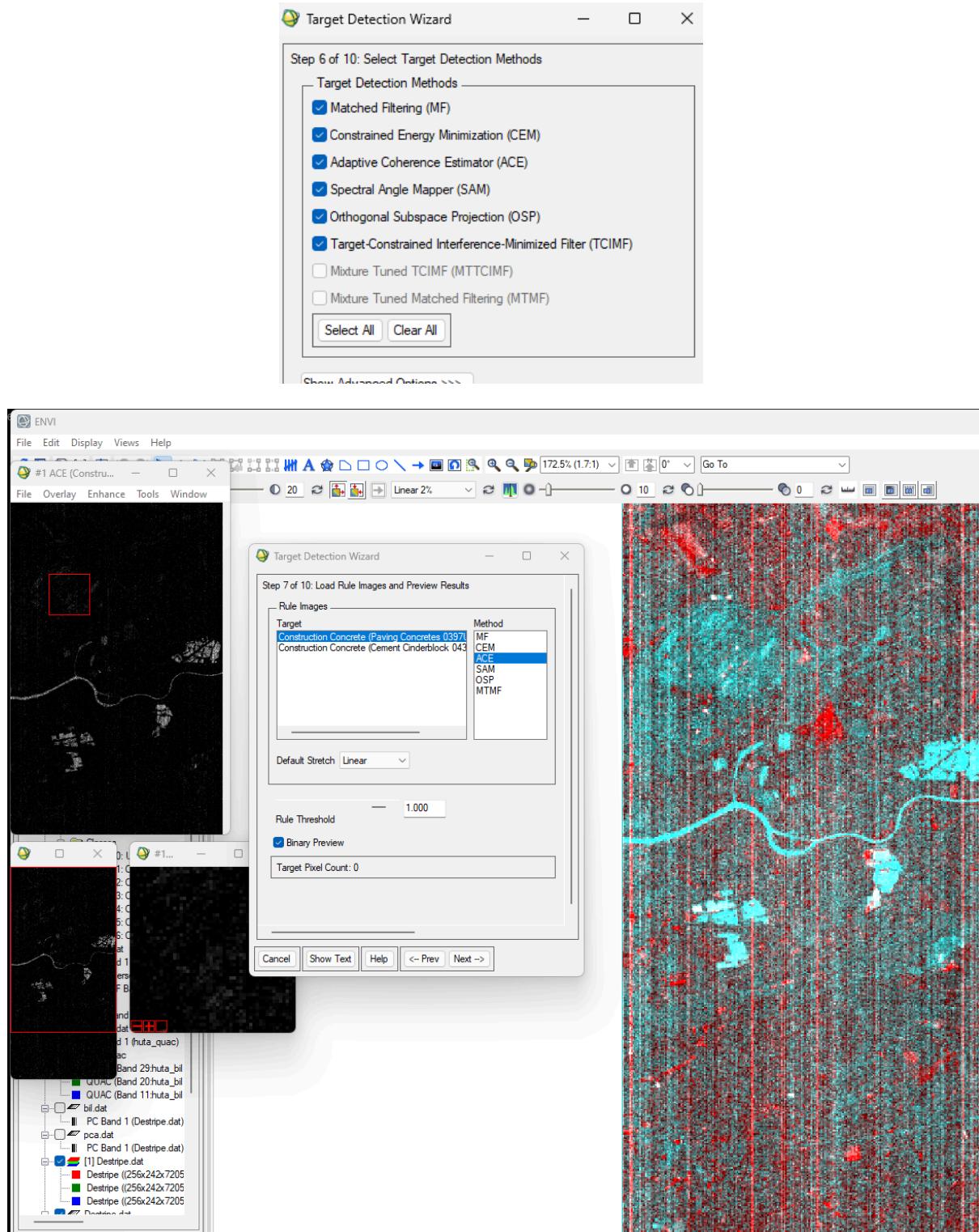
According to the agreement at the university, our goal was to detect asbestos. After reviewing the literature, we determined that we should focus on the spectral curve of Chrysotile ( $Mg_3(Si_2O_5)(OH)_4$ ), which is the most common type of asbestos. However, we encountered technical difficulties - the spectral curve of Chrysotile did not want to load on any of the computers in our group, which made it impossible to continue with this task.

## Construction Concrete

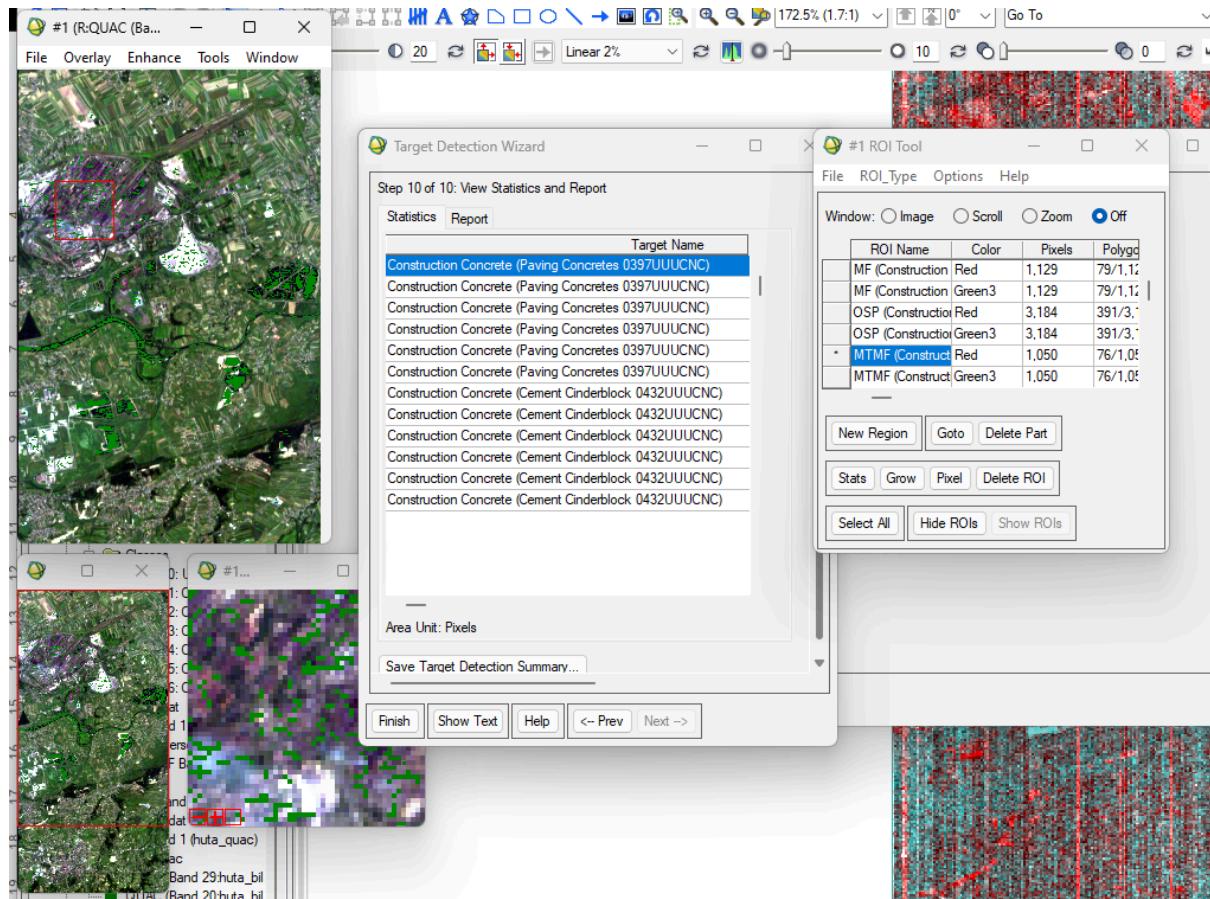
Finally, we decided to perform detection for Construction Concrete (Cement Cinderblock) and Construction Concrete (Paving Concretes) to try to get more reliable material detection results.



## Hyperspectral processing with ENVI



## Hyperspectral processing with ENVI



Despite the analysis, the data still classifies concrete as water, even though the classification appears to be partially correct at the smelter site. This result may be due to several factors. First, concrete may exhibit similar spectral properties to water in some spectral bands, particularly in the near-infrared, where water and concrete may have similar reflectance values. Water tends to strongly absorb radiation in this range, and concrete, especially wet concrete, may also show increased absorption, which can lead to misclassification.

## Submission

As part of the project, we deepened our knowledge of ENVI software and hyperspectral data processing methods. We understood the operation of tools in the software, such as atmospheric correction, classification and spectral analysis. We performed a number of operations, including FLAASH and QUAC atmospheric correction, data conversion and classification. Although the classification results were satisfactory, we noted the need for further work to refine object detection.