1. This section describes the approaches, results, and assessments of the data to support the goal of locating and identifying invasive animal species. While there are many potential uses and considerations as to the overall value of the data, all assessments and supporting evidence are related to the invasive species mission.

2. Approach. The first step is to select pixels that represent an animal from the full sensor-sweep. *Figure 1* is a graphical representation of how the HSI data is organized to reflect ground truth. In this case, the Red-Green-Blue spectra are presented across a line representing the sweep of the internal mirror across the Field of View (FOV). Various targets can be "seen" as color-points within the sweep. In this case, green represents vegetation, yellow represents a hog, and the reddish color represents the fencing material. To get to this point:

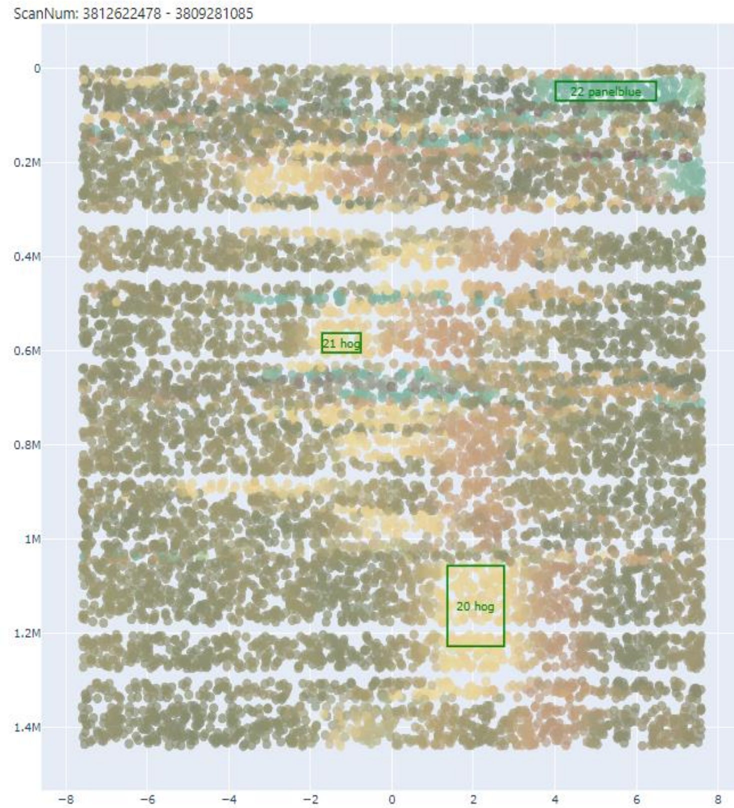


Figure 1 - RGB depiction of Hyperspectral Data over target materials

2.1. Data was standardized by subtracting the mean of each spectra, and then dividing by the standard deviation.

2.2. The spectra from each mission were then transformed into approximate RGB color, and then used to create a 2D scatter plot of all scans with the sensor mirror angle on the X-axis, and time on the Y-axis as shown in figure 11.

2.3. The 2D scatter plots were correlated with photos to help identify ranges of mirror sweeps and timestamps that contain a material of interest (shown in the green boxes in Figure 11).

2.4. Models were then configured and/or trained on a set of materials identified in the previous step. The two models demonstrated are described in section 1.3

2.5. Once the models were prepared, they were then used to classify the full mission scan set to validate their performance. This was done as the "ground truth" was easier to determine.

2.6 After the models were validated they were used to classify various flights, and the data was then visually inspected for validation. This is discussed in more detail in the following sections.

3. Observations

3.1. Spectral Similarity. The spectra of some materials from 330nm to 1017nm are shown below - this plot is a diagnostic used in the compute node to show the median signals of different 'materials'. In the first plot *(figure 2)* we show the spectra of 4 different 'cow' materials extracted from a capture of cows in an enclosure. The vertical bars are used to represent the approximate visual color of light corresponding to wavelength in the spectra.

To illustrate the issue in spectral similarity, the blue cow signal (corresponding to darker fur) has been hidden, and signals for ‘hog’, ‘goat’, ‘horse’ and ‘plywood’ have been added.

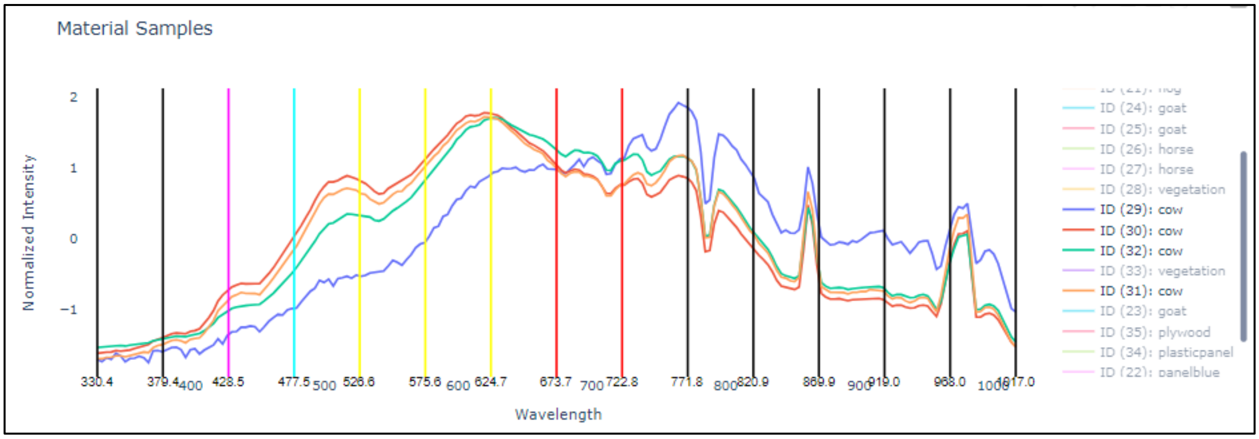


Figure 2 - Spectral plot depicting differing cow signatures

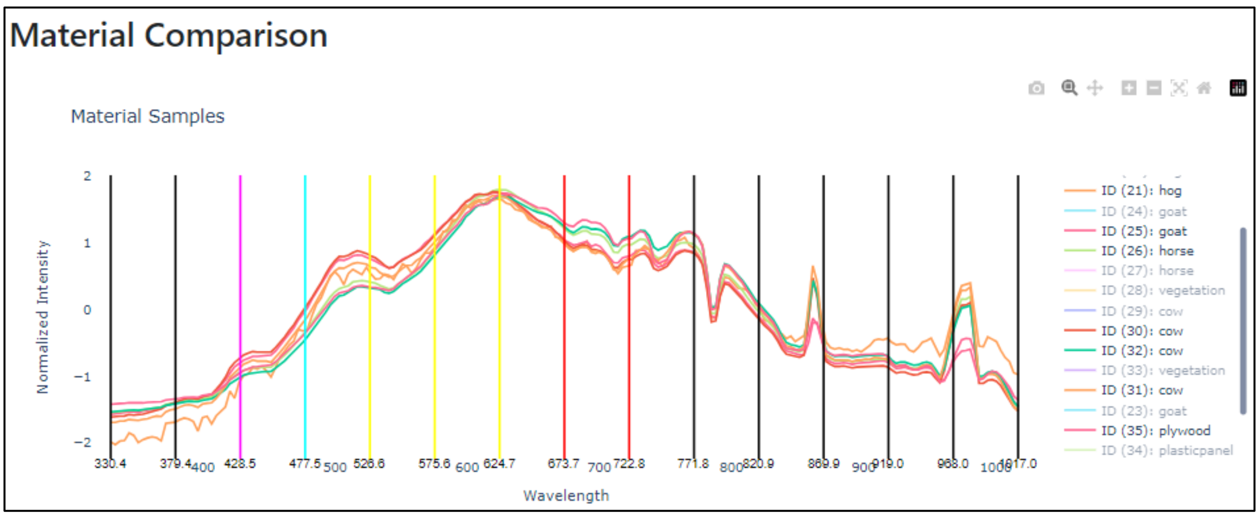


Figure 3 - Spectral plot depicting different animal species

The spectral similarity shown above *(figure 3)* is a **significant challenge**, especially when considering the variation within the cow spectra samples can be as significant as the variation from other materials, even plywood.

3.2. Signal to Noise ratio. Shown below *(figure 4)* are the spectra for 2 hog samples, 2 piglet samples, and one sample of vegetation. The vegetation signal is shown in orange and is used to illustrate the difference between a good and ‘noisy’ (or dimly lit) signal. The piglet signals in red and green demonstrate a noisy signal where the electrical biases of the sensor are having a stronger effect on the signal.

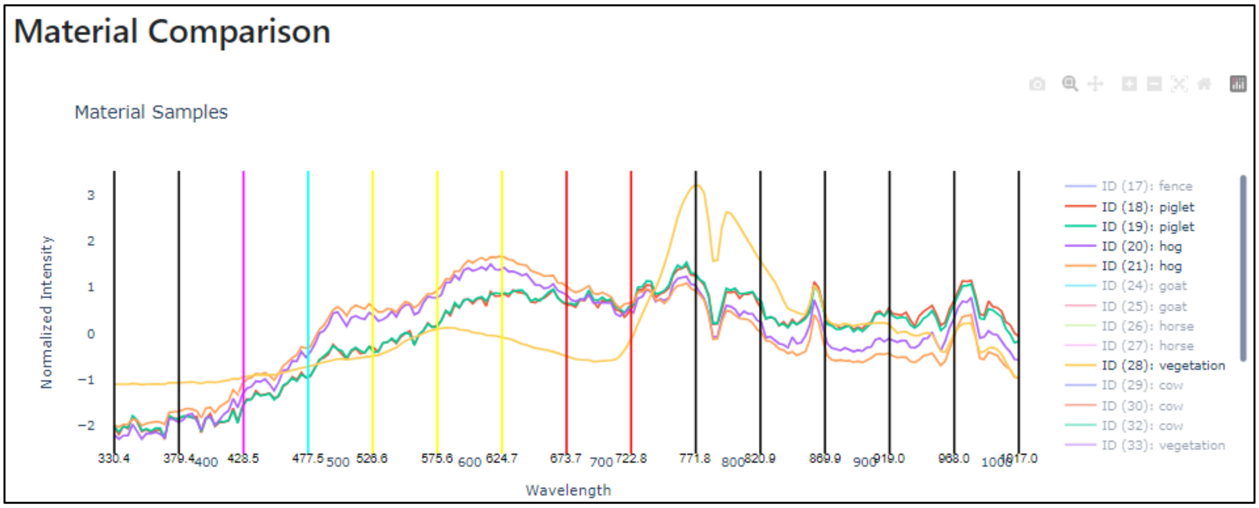


Figure 4 - Spectral plot of noise and animal signal

A separate diagnostic page is available to show a number of individual scans *(figure 5)* (red transparent), and the median scan which we viewed above (blue)

- Spectral mixing leading to misclassification
    - Demonstrate cow/panel mixup
    - Caused by dominance of spectra by color
    - Other features may be of interest but the Signal/Noise ratio is too poor, and insufficient samples of animals to create a statistically significant analysis

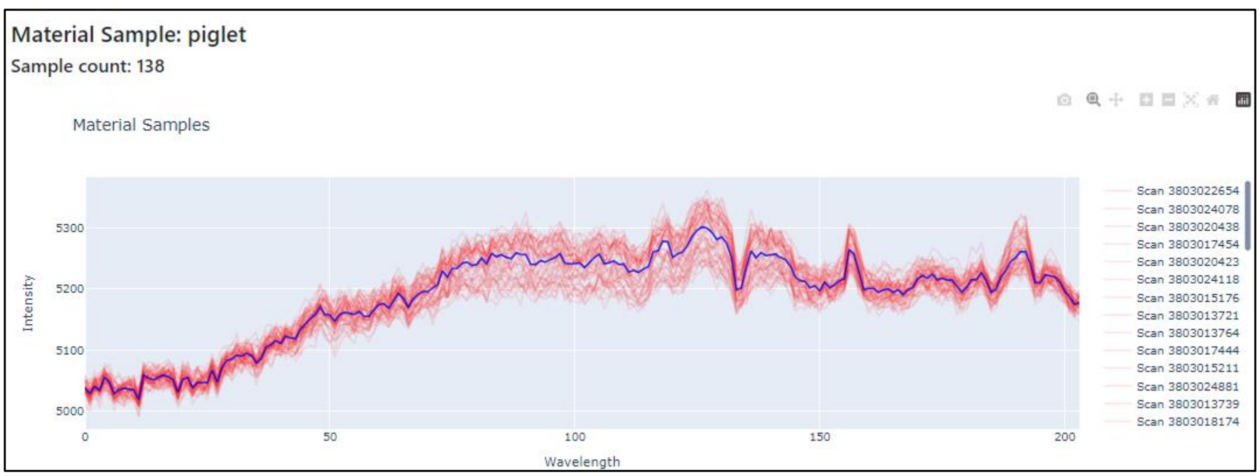


Figure 5 - Determining average scan values from all input scans

- We're rarely classifying a matching signal. We are usually trying to correctly identify a mixture of signals which may contain our signal of interest (in some unknown ratio)

3.3. Material Catalog

There were 21 materials *(figure 6)* identified in preparation for analysis. Materials were identified using a combination of mirror-angle range and scan number range. The table of materials are shown below. In the pseudo-color ranges of scans, if any material is defined within the range of scans being shown, it is rendered in a green box.

MATERIAL ID	SCANNUM START	SCANNUM END	GALVO START	GALVO END
plasticpanel	3856091655	3856114077	31428	37981
plywood	3856149491	3856185223	27089	31085
plasticpanel	3854040947	3854091704	21201	27594
vegetation	3819005121	3819039490	29335	41312
cow	3820209717	3820252233	47458	56080
cow	3819081737	3819109630	37283	40107
cow	3819492441	3819539040	58420	63740
cow	3819455190	3819532955	28690	38659
vegetation	3817915529	3818022364	56963	62236
horse	3817410018	3817565847	45025	51054
horse	3817914013	3818018547	23028	31735
goat	3814207562	3814238277	50781	57496
goat	3814286725	3814339255	30931	39522
goat	3813727981	3813795019	53391	61529
panelblue	3809347311	3809441066	52963	61458
hog	3810587028	3810682759	33247	36511
hog	3811733305	3812129990	43827	48612
piglet	3803013003	3803026302	29629	33008
piglet	3803027073	3803042735	29470	32545
fence	3803028704	3803044363	21578	24729
panelblue	3802848792	3802902317	25000	30000

21 materials

Figure 6 - Classification materials list for testing

4. Algorithms Considered. To overcome the data challenges referenced above, we took three distinct approaches to data analysis to further identify signal characteristics, analytic tools suitable for a small compute node at the edge, and accuracy characteristics that could lead to further study.

4.1. Siamese networks. These networks differ from a common classification model in that they do not produce an output label, but instead produce "similar" or "dissimilar" notation.
To use this network, we provide 'reference' spectra which we then use the model to be similar or dissimilar to a given input spectra from the sensor.

4.1.1. Training Data Preparation: Training data is prepared for the network as follows:

- Material Scans are sampled in batches to produce an equal number of median signals (reducing noise)
- The training data is then created by pairing two signals together and providing a target similarity value (0 for dissimilar, 1 for similar). 'Similar' signals were created by pairing all median signals of the same animal species, including those of different colors. 'Dissimilar'

signals were created by pairing any different animal/material species.
This encourages the model to recognize similarities between cow signatures even when the color of the fur is different while also recognizing when the species is different.

- The training data was then split 75/25 into a training and 'validation' dataset. The training dataset is used for training the model, while the validation dataset is kept separate and 'unseen' to give a better measurement of classification performance.

4.1.2. Training Results: The model quickly performed well and produced an accuracy of 91.8% on the training data, and 97.2% on the validation data.

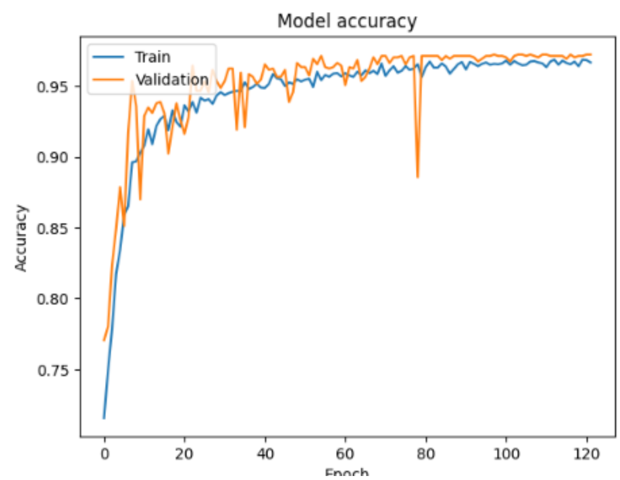


Figure 17 - Model Training and performance plot

4.2. RElative Spectral Differential ANalysis (RESDAN) is an algorithm developed during prior phase II efforts with the sensor. It is based on statistical analysis involving distance metrics between spectra, and cosine similarities (similar to spectral angle mapping). It utilizes a combination of spectral clustering and signal deltas between sequential measurements to perform a type of spectral angle mapping.

- Training: There is not necessarily any training involved. There is an option to select K most significant wavelengths from the spectra, which reduces the data dimensionality (~210 wavelengths reduced to K wavelengths). Reducing dimensionality may speed up the analysis and improve accuracy but this is dependent on the quality of information within the signal.

4.3. N-Class classifier. This classifier was intended to be evaluated but due to concerns raised during preliminary analysis (spectral similarity and signal to noise ratio) it was not. Earlier efforts during Phase II development of the sensor produced several versions of such models which exhibited poor performance when used on data beyond what was included in training datasets.

5. Algorithm Selection Criteria

5.1. Classification performance:

5.1.1. Correct classification is measured by how much of a <u>known material</u> (animal species, fence, vegetation etc) is correctly identified - this would be true-positive identification. This must be done using the training data (from collections of species in animal pens) to support their being a reasonably known material.

5.1.2. Incorrect classification is measured on a case-by-case basis as many factors may be involved and understanding these factors is important to understand the degree of the misclassification. For example, spectral mixing and spectral similarity of materials of interest.

5.1.3. Speed: Although speed is important for aspects such as real-time analytics, there are many ways that a processing bottleneck can be overcome, so this is not a factor for algorithm selection.

5.1.4. Result interface

The result rendering interface is shown below.

- **Top left panel** allows for Mission and Classifier selections to render the result in the scatter plot.
- **Bottom left panel** shows an interactive scatter plot of the results. Here scans are color coded by their classified label, and positioned by the timestamp and mirror angle. When hovering the mouse over the points, a popup box is shown which provides additional details of the scan and additional information about the classification. When a point is clicked, the top right and bottom right panels are updated.
- **Top right panel** shows a line plot of the scan point that was clicked (blue line with points and dots), and other material spectra (transparent lines).
- **Bottom right panel** shows the closest camera image to the point in time of the clicked scan.

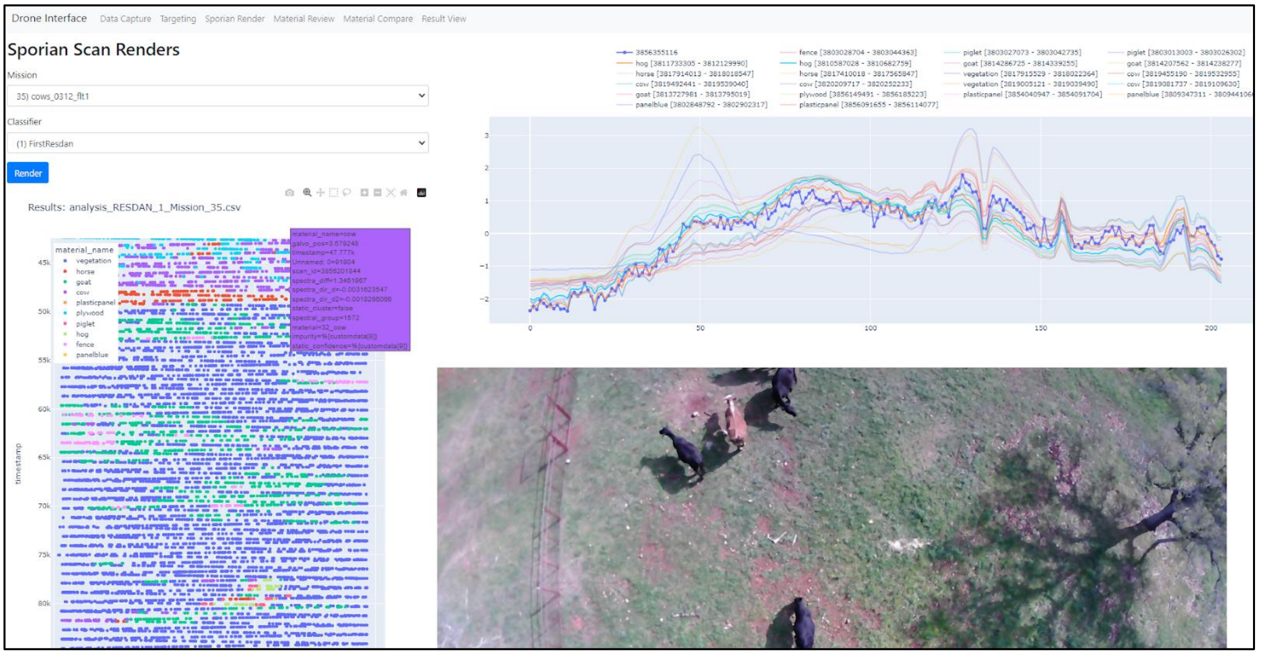
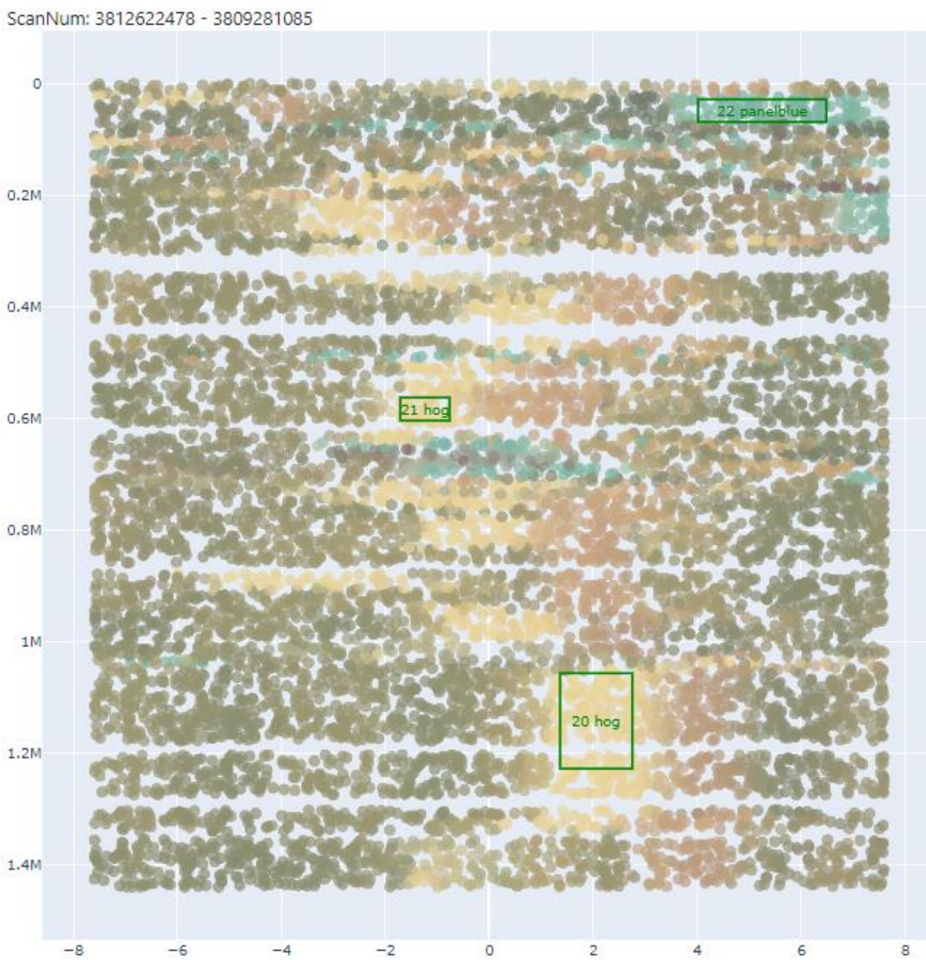


Figure 1 - Spectral data visualization depicting results

5.2. Comparison of outputs

5.2.1. Hog Capture

- Shown to the right is the reference photo from pen capture showing a hog against a fence. The figure below shows the pseudo-color plot of spectra which is used to determine the correctness of outputs from the Siamese and RESDAN analysis.



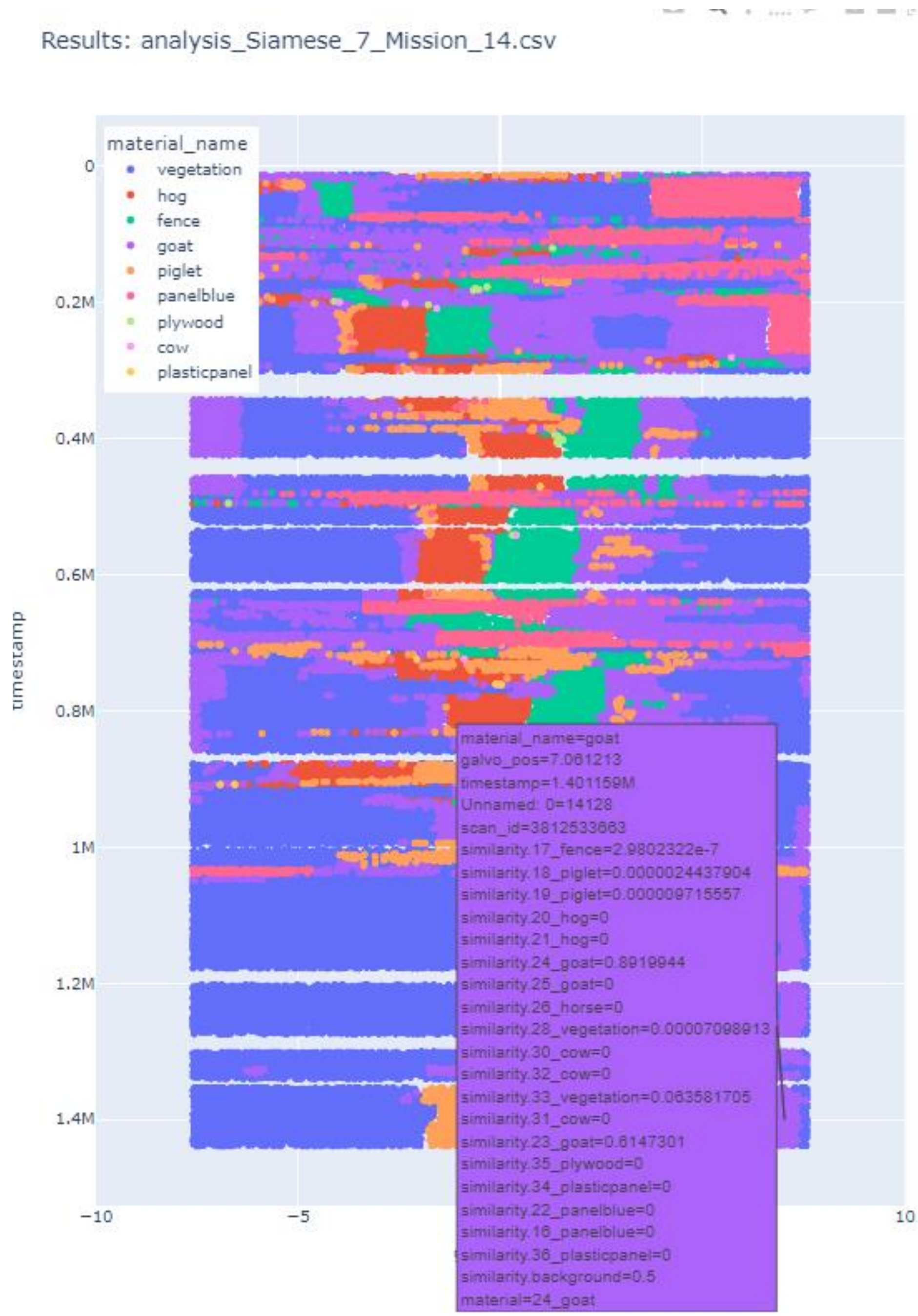
5.2.1.1. Siamese Network (All materials)

- The Siamese analysis is shown below.
- The network was configured to identify matches to all 21 materials defined in section 1.3.3

Result preview
The result of the Siamese classification for the Hog pen capture is shown below.

It is immediately clear that the purple ‘horse’ spectra is being identified in regions where there was only vegetation

5.2.1.1. RESDAN Network (All materials)

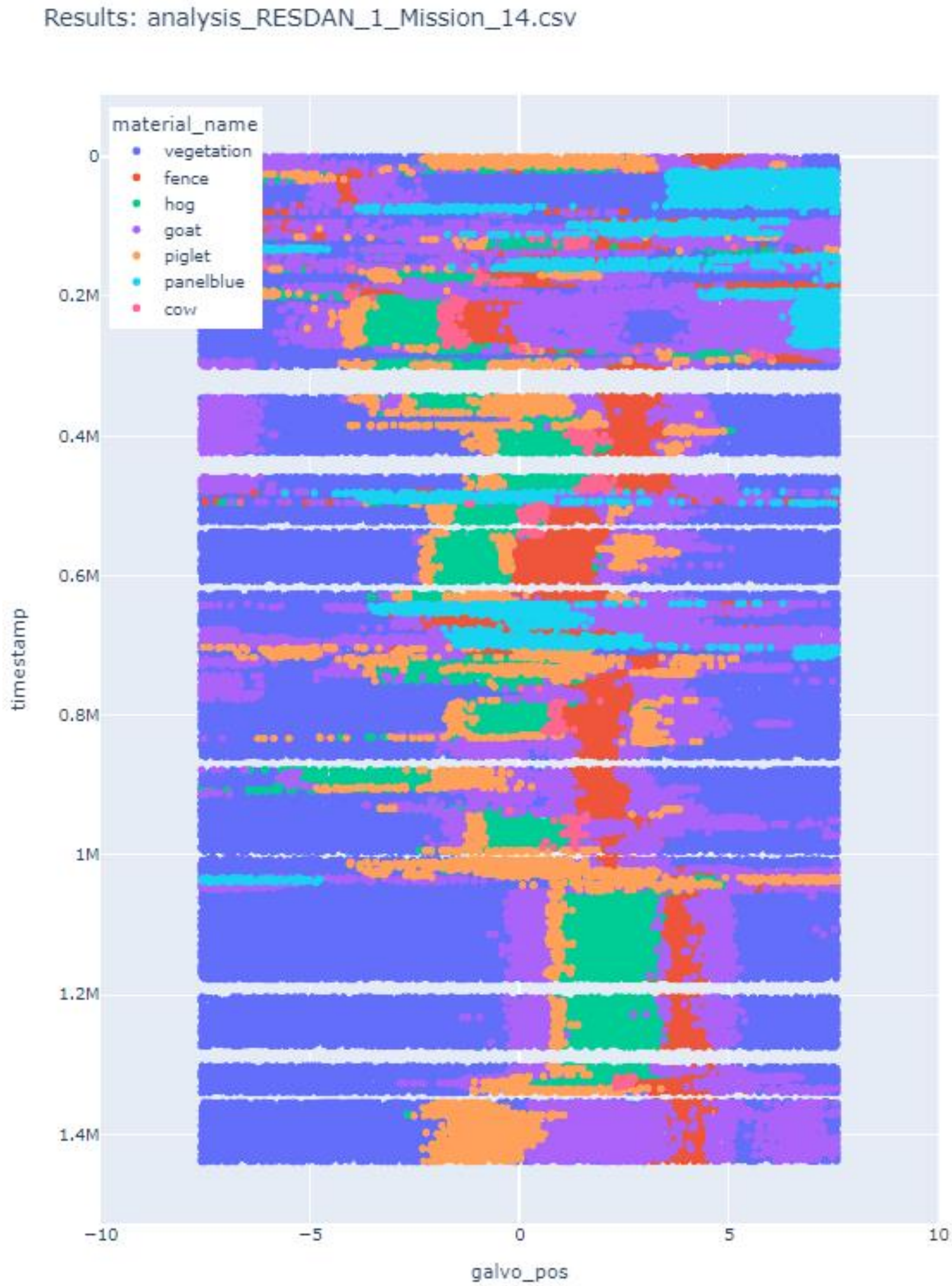


The RESDAN algorithm was configured to identify matches to all 21 materials defined in section 1.3.3

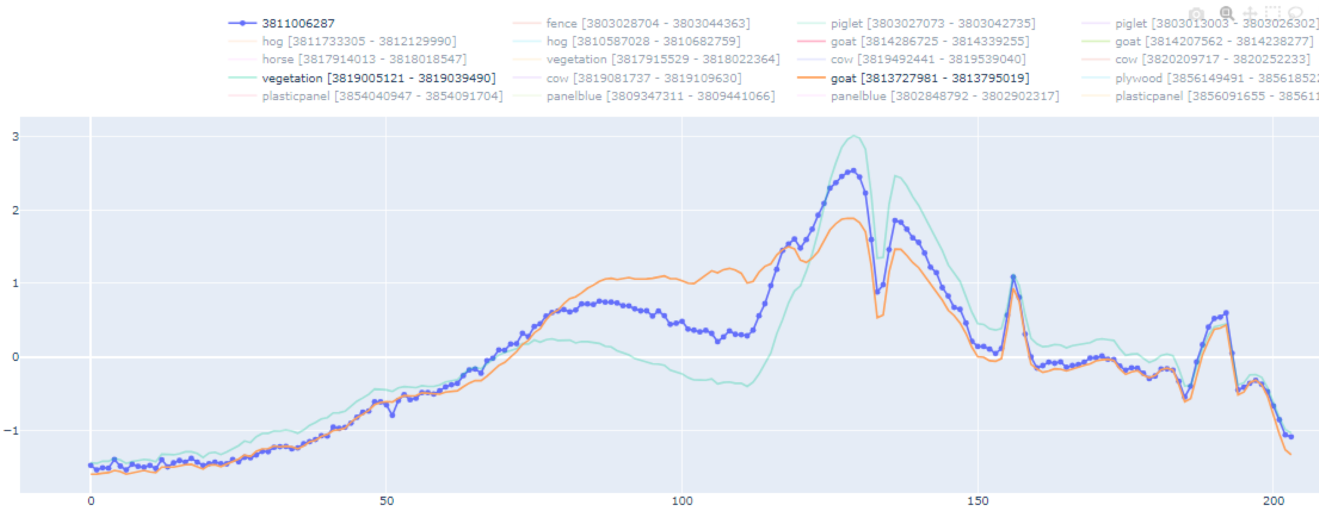
Result preview
The output classifications from the RESDAN algorithm are shown on the right. The color codes can be seen in the top right.

Correct classifications
The hog material (green) is found in the correct locations as far as can be determined from the pseudo-color plot shown earlier in section **5.2.1**. The fence (red) and panelblue (cyan) materials are also correctly identified.

Incorrect classifications
The goat (purple), piglet (orange), and cow (pink) materials are incorrectly identified frequently. Each of these are further investigated below, but unfortunately in each case, it is clear that the spectra being measured are closer to the incorrect materials than expected.



Incorrect classifications - Goat
By clicking on any of the points in the scatterplot, a plot of the spectra are shown with all known materials. Below is a screenshot of this plot, showing the scan itself (blue dots and line) with the goat and vegetation reference spectra (orange and cyan respectively).



It is clear from the plot that the spectra being viewed is somewhere between vegetation and goat spectra.

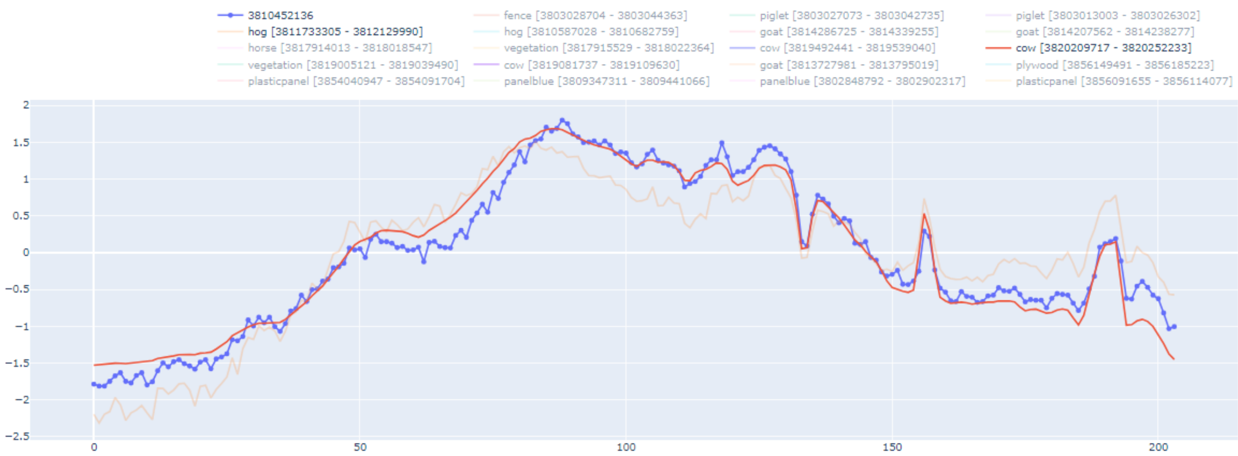
Incorrect classifications - Piglet

We obtained a screenshot of the scan spectra, piglet and hog reference signals in the same way - shown below.

Although difficult to see the light orange line for a hog spectra, the scan spectra (blue) is again between the expected signal (hog) and the unexpected piglet spectra - in this case, a little more similar to the piglet signal.

Incorrect classifications - Cow

We obtained a screenshot of the scan spectra where cow material was detected (pink regions between hog and fence), cow and hog reference signals in the same way - shown below.



It appears that a spectral mixture of <u>hog</u> and <u>fence</u> spectra may have a result that is similar to the spectra of a <u>cow</u>.

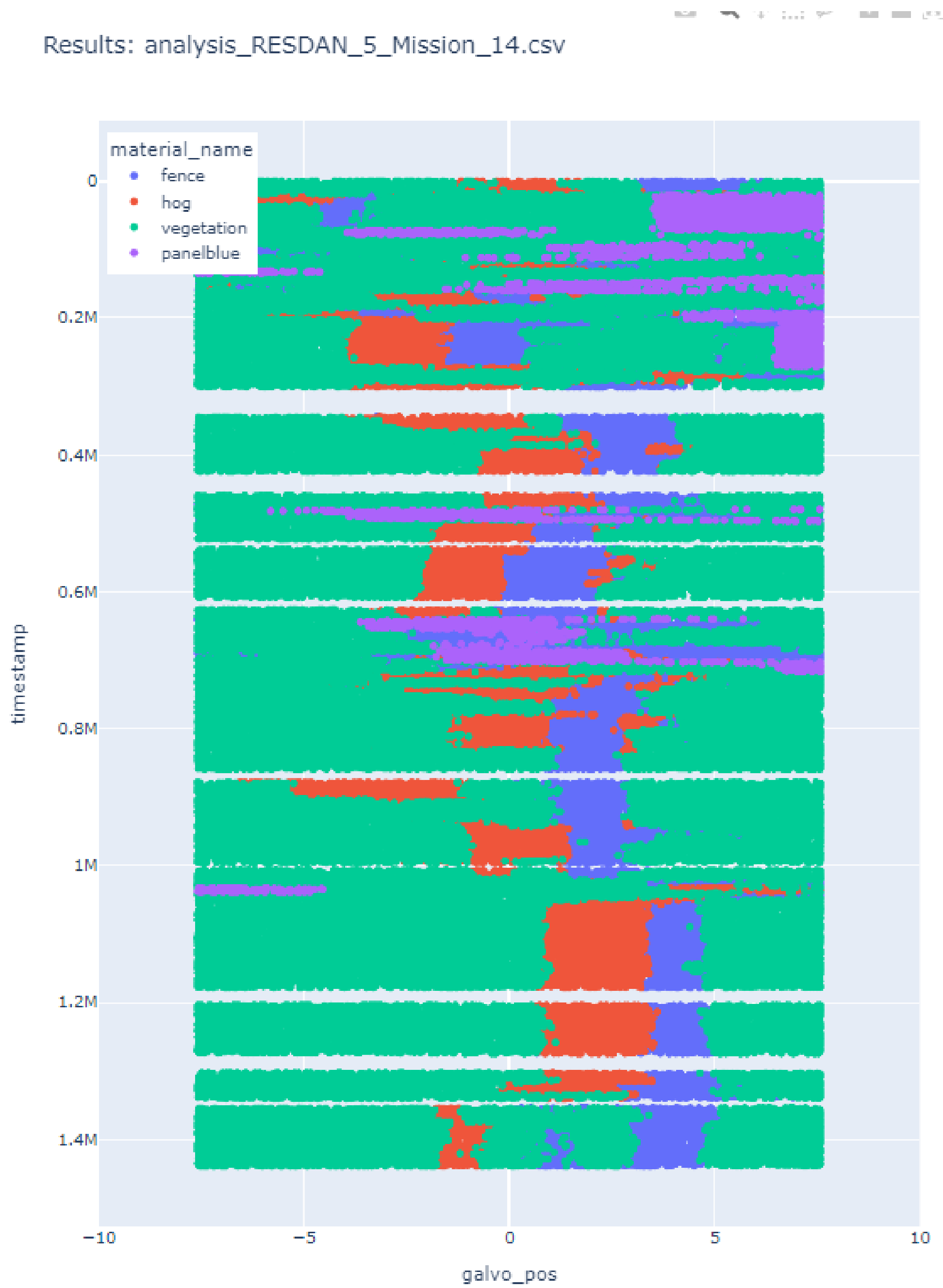
Result interpretation:
The RESDAN algorithm in its current state is not sufficient for the task of correctly identifying animal species using their fur spectra alone. There is not sufficient clarity of unique features in the fur samples that would be greater than the variance between different animals of the same species, and spectral mixing makes it very difficult to discern a true from false detection.

5.2.1.3. RESDAN Network (Fence, Hog, Vegetation, Blue Panel materials only)
The RESDAN algorithm was configured to identify matches to only 4 materials (Fence, hog, vegetation, blue panel) as defined in section 1.3.3

Result preview
The output classifications from the RESDAN algorithm are shown on the right. The color codes can be seen in the top right.

Correct classifications
The hog material (red) is found in the correct locations as far as can be determined from the pseudo-color plot shown earlier in section **5.2.1**.
The fence (blue) and panelblue (cyan) materials are also correctly identified.

Incorrect classifications
By reducing the available species spectra in the classification problem, we have mitigated the incorrect classifications. It is suspected that piglet spectra were at least partially visible on the right side of the fence so the 'correctness' of not classifying them more consistently (even as hog) is not clear.



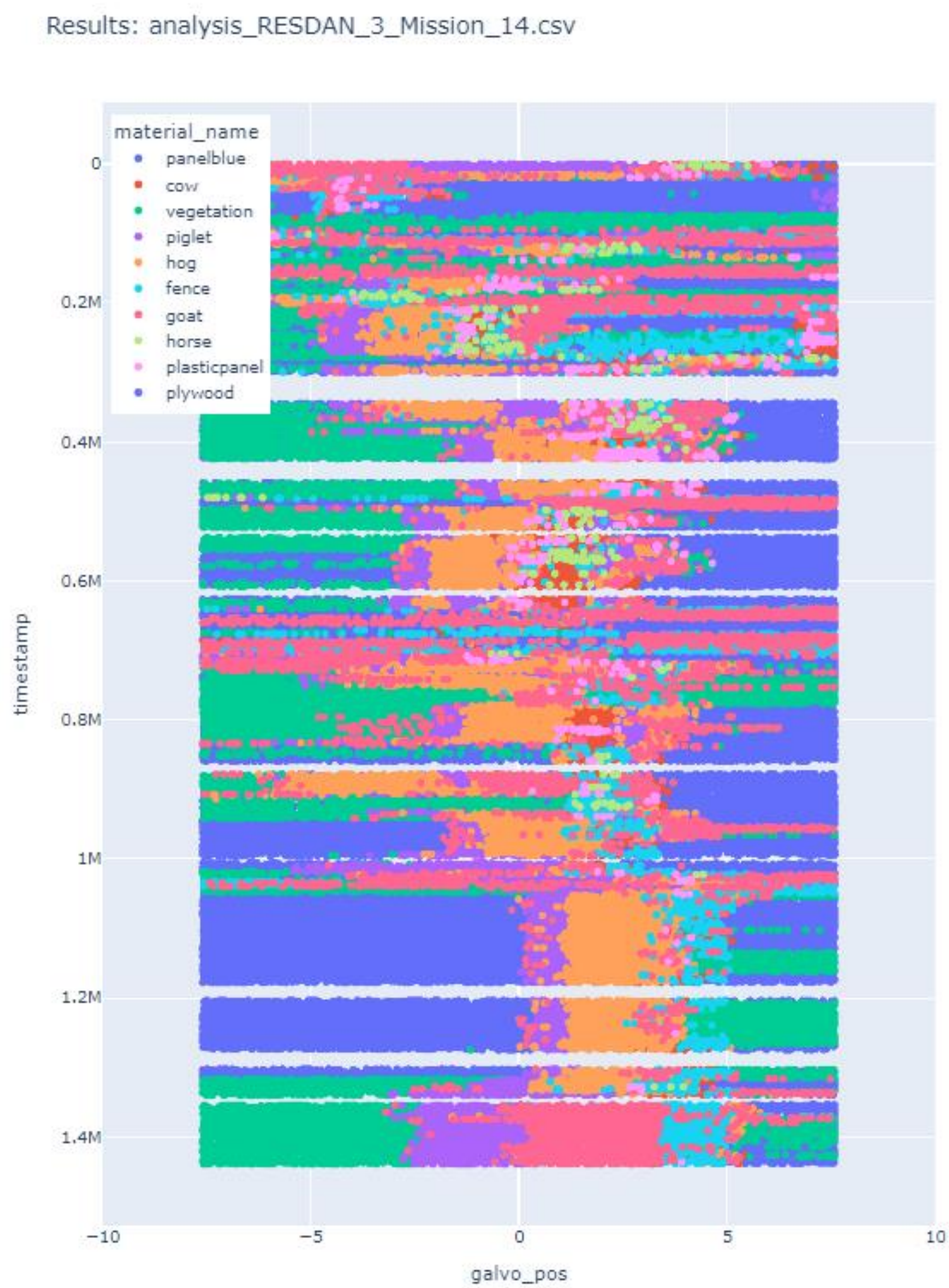
5.2.1.4. RESDAN Network (All materials - 25 components)

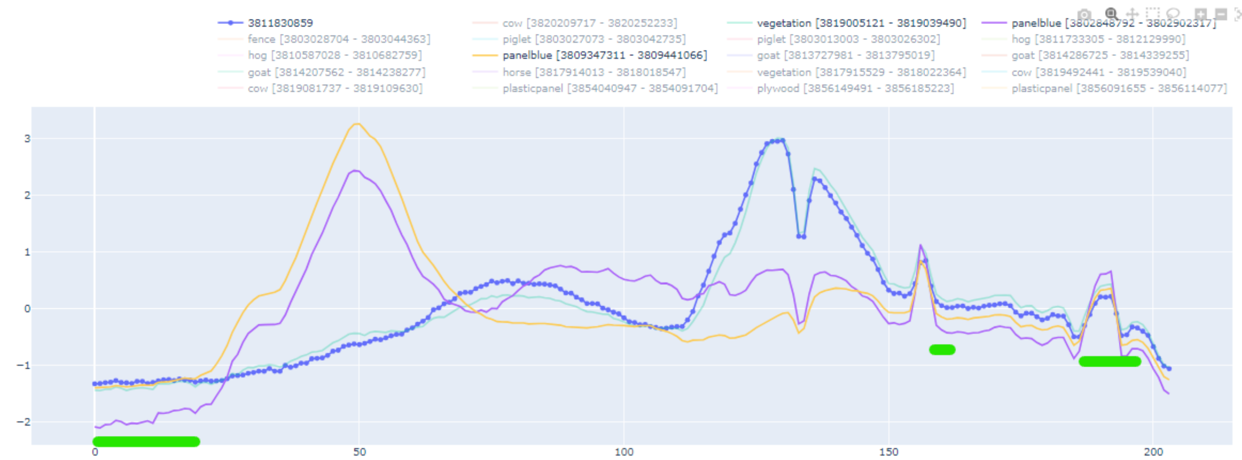
The RESDAN algorithm was configured to identify matches to all materials as defined in section 1.3.3, but using only 25 of the available wavelengths. The intention was to reduce dimensionality and improve clarity in the output.

Result preview
The output classifications from the RESDAN algorithm are shown on the right. The color codes can be seen in the top right.

Correct classifications
The hog material (orange) is found in the correct locations. However, the confusion of other spectra nearby is much worse.

Incorrect classifications
There are large swathes of 'blue panel' classifications where there should be vegetation. After inspecting the spectra, it appears possible that the top 25 wavelengths were biased by the other samples that aren't influenced by blue pigment. A plot of the spectra comparison showing the scan, vegetation, and bluepanel is shown below, with a green line under the spectral indices that were considered most influential.





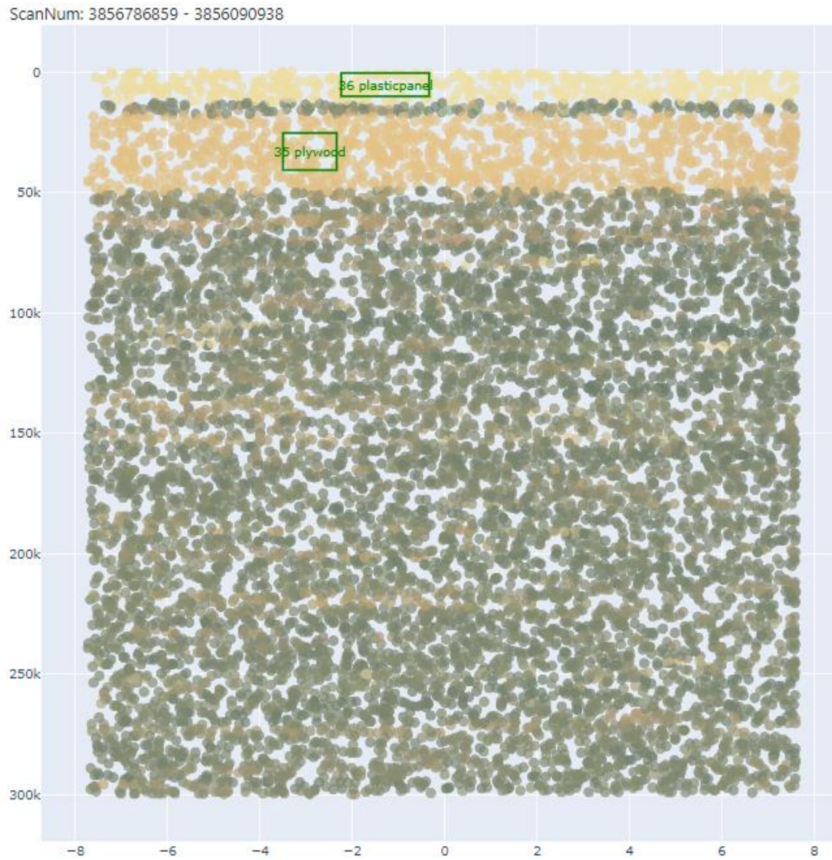
The result of selecting K best indices is disappointing and is supporting even stronger confusion between spectrally dissimilar materials, so it was not used any further given the time available for analysis.
The methods used here are **SelectKBest**, and **mutual_info_regression** from the scikit-learn python package. Other signal transformations such as LDA and PCA were considered but early experiments yielded poor results, and so further investigation was postponed.

5.2.2. Cow flight capture

The first flight to be analyzed was the cow-flight mission, where the drone was flown over a small herd of cows in a pen at different altitudes.
<more Detail on flight>

The figure to the right shows the pseudo-color plot of the spectra from the 5 minute recording which is used to help determine the correctness of outputs from the Siamese and RESDAN analysis.

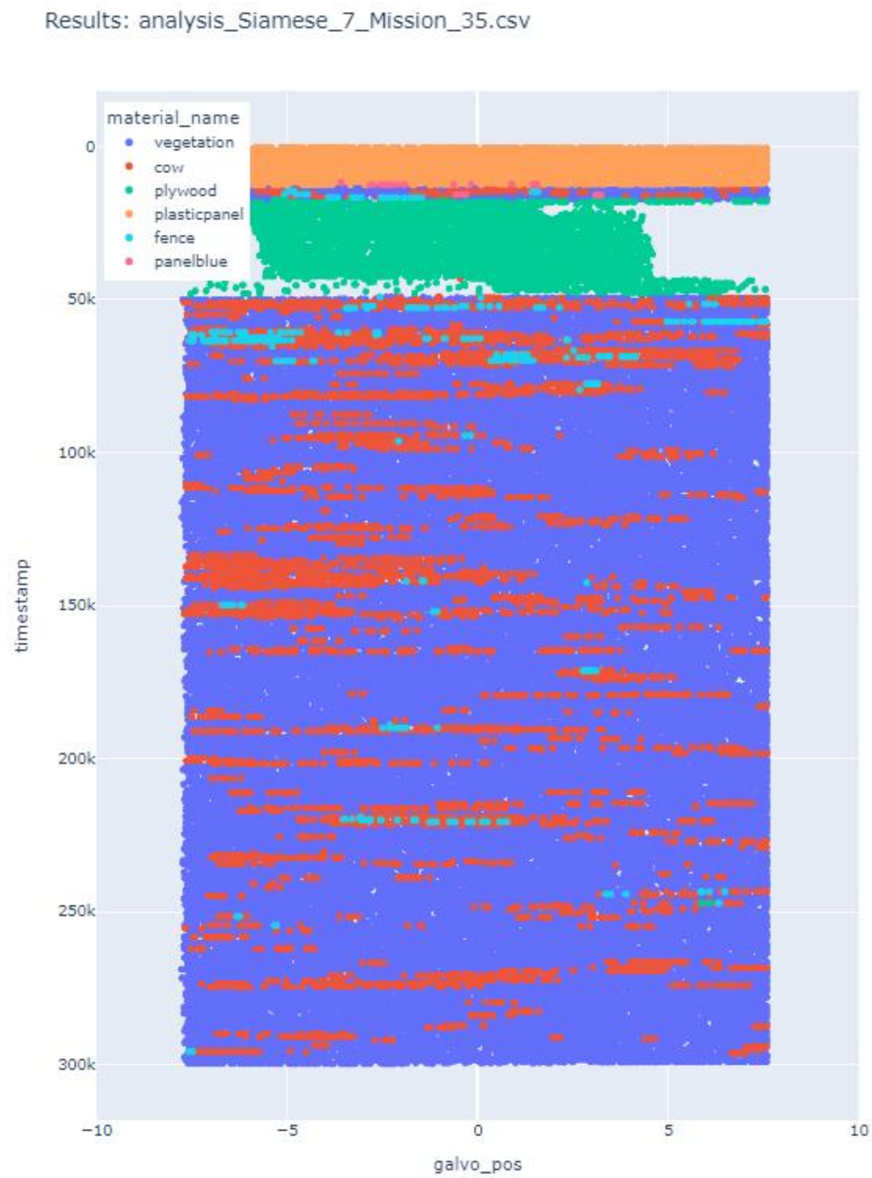
Camera images are also available when any of classification spectra points are clicked, some of these will be shown in analysis.



5.2.1. Siamese Network (Fence, Vegetation, Cow, Plywood, Plastic Panel, Blue Panel)

The Siamese analysis is shown to the right.
The network was configured to identify matches to a subset of the materials defined in section 1.3.3. Although the results look encouraging, there is no other animal species being tested with the classification.

Result preview
The result of the Siamese classification…

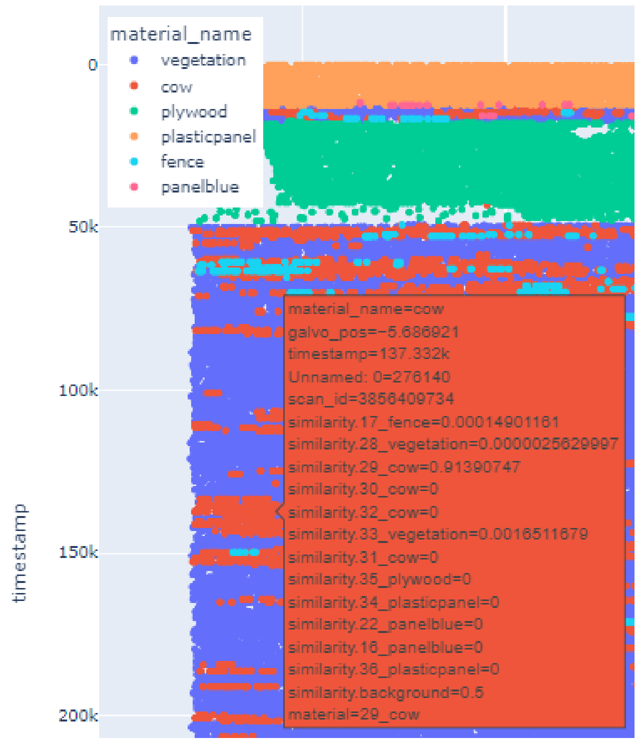


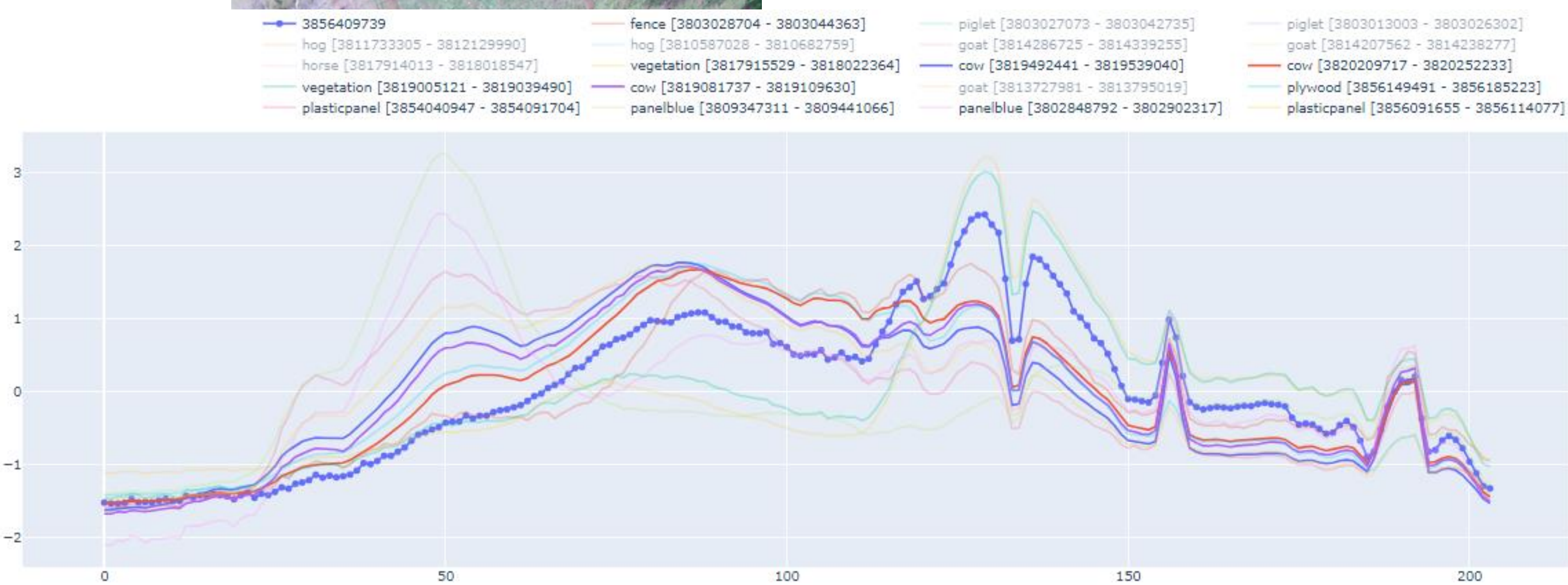
Inspection of Cow classification
When a point in the scatterplot is hovered, the similarity of all tested materials is shown (the red box). When a point is clicked on, the spectra and closest photo are shown.

For the point shown on the right, we see a red 'cow' classification. It can be seen from the red popup card that the cow similarity was 0.9139 (91.39%).

The closest photo to this point on the scatter plot is shown below followed by its associated spectra plot (materials not tested by the classifier are not rendered).

This classification seems reasonably accurate, but precise projection of where this point is on the photo requires additional data. This is a challenge in validating the flight capture classifications.





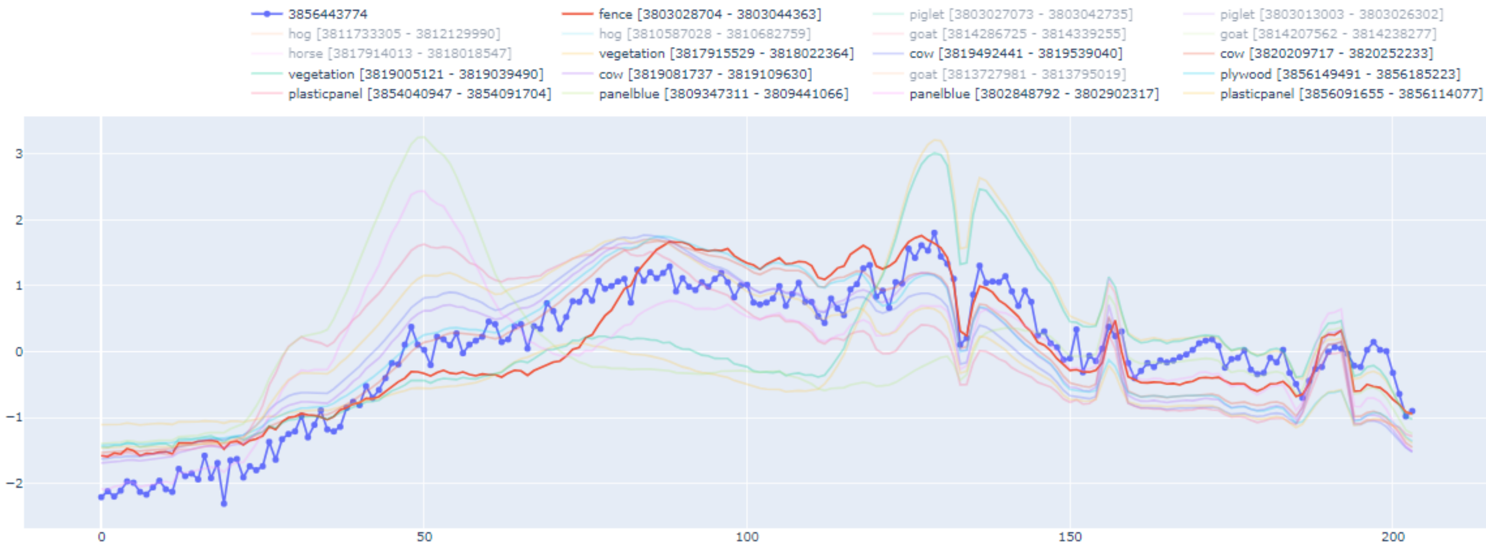
Fence misclassification

If you refer back to the preview of the full classification map, you can see there are two cyan dots ('fence' classifications) at approximately 145k on the timestamp axis.

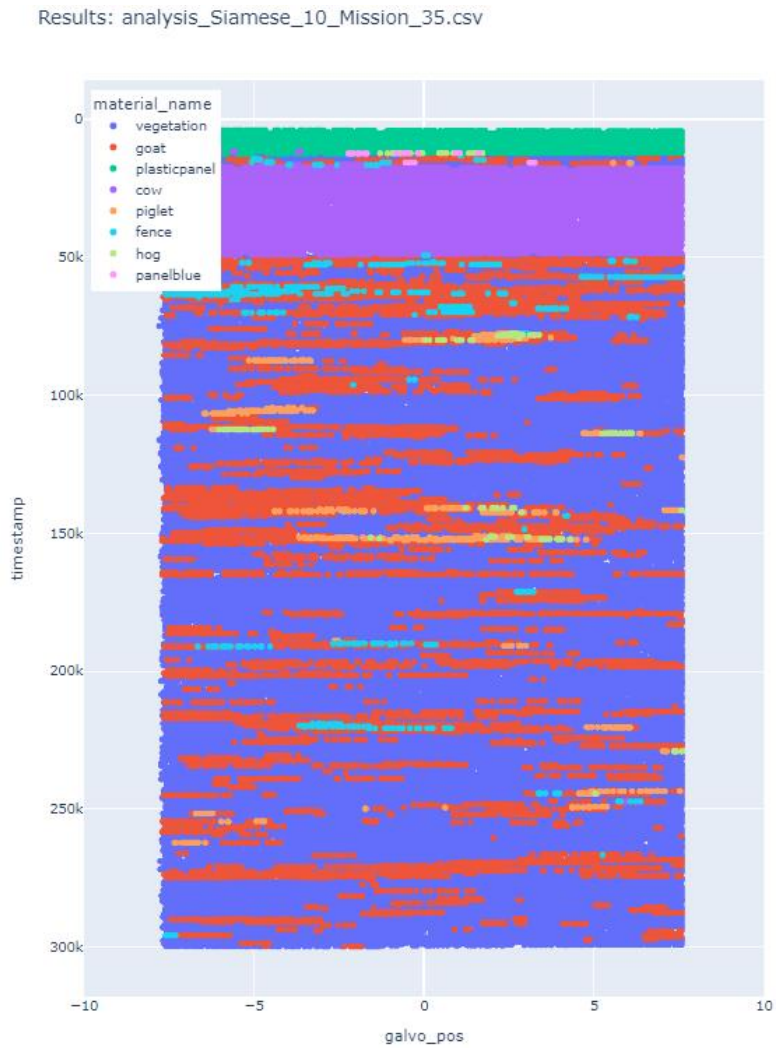
A zoomed in view of this with the mouse hovering over a 'fence' point is shown on the right.
This point was chosen as it was near the center of the galvo mirror angle, and should therefore be somewhere in the center of the corresponding camera image.

When reviewing the camera image below, we can see there are only cows present and no fence.

When reviewing the spectra plot shown below, we can see the signal appears to be quite noisy and this is likely a misclassification due to low light signal (likely in the shadow of a cow).



5.2.1.2 Siamese Network (All materials)
The Siamese analysis is shown to the right. The network was configured to identify matches to a subset of the materials defined in section 1.3.3. Although the results look encouraging, there is no other animal species being tested with the classification.

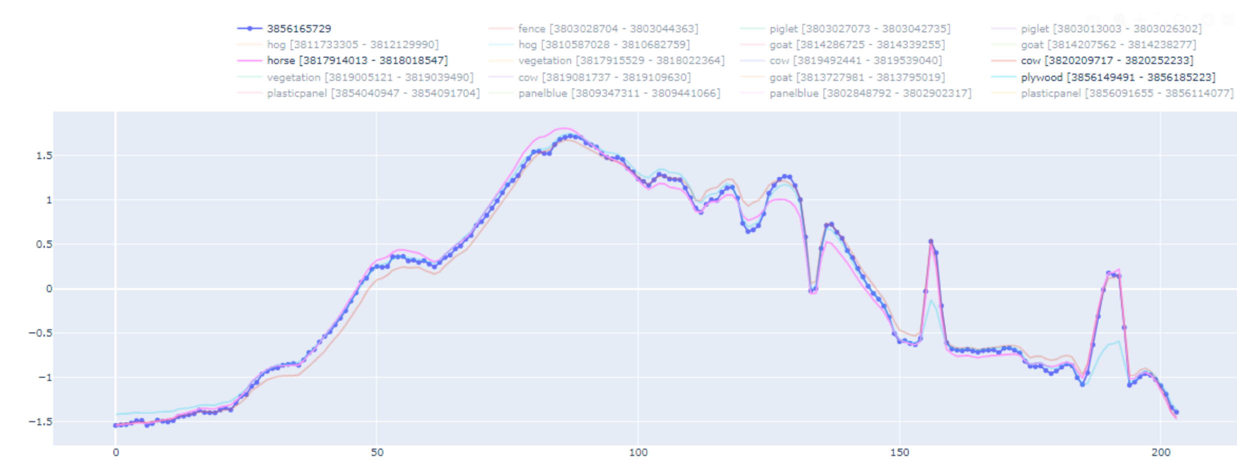


Result preview
The result of the Siamese classification immediately shows a majority of misclassifications involving ‘goat’ signals (red), and ‘cow’ signals (purple) where plywood was present.

This was particularly interesting. When referencing the classification preview in section **5.2.1.1,** there were regions of the ‘cow’ classification which were not classified due to a lack of certainty.
When hovering the cursor over the ‘cow’ classification we can also see that plywood material had a similarity of about 97% with cow being higher at approx 99%.

The root cause is unclear for this, but one hypothesis is that training the classifier on a broader range of materials may have led to a “preference” for wavelengths similar to what was seen in section **5.2.1.4.** This requires further investigation but could indicate the model is being trained to recognize artifacts

of particular data collections (peaks at approximately 810nm and 980nm) rather than generalizing toward recognizing fur of species.
Suggested steps to improve this are discussed in the conclusion section.



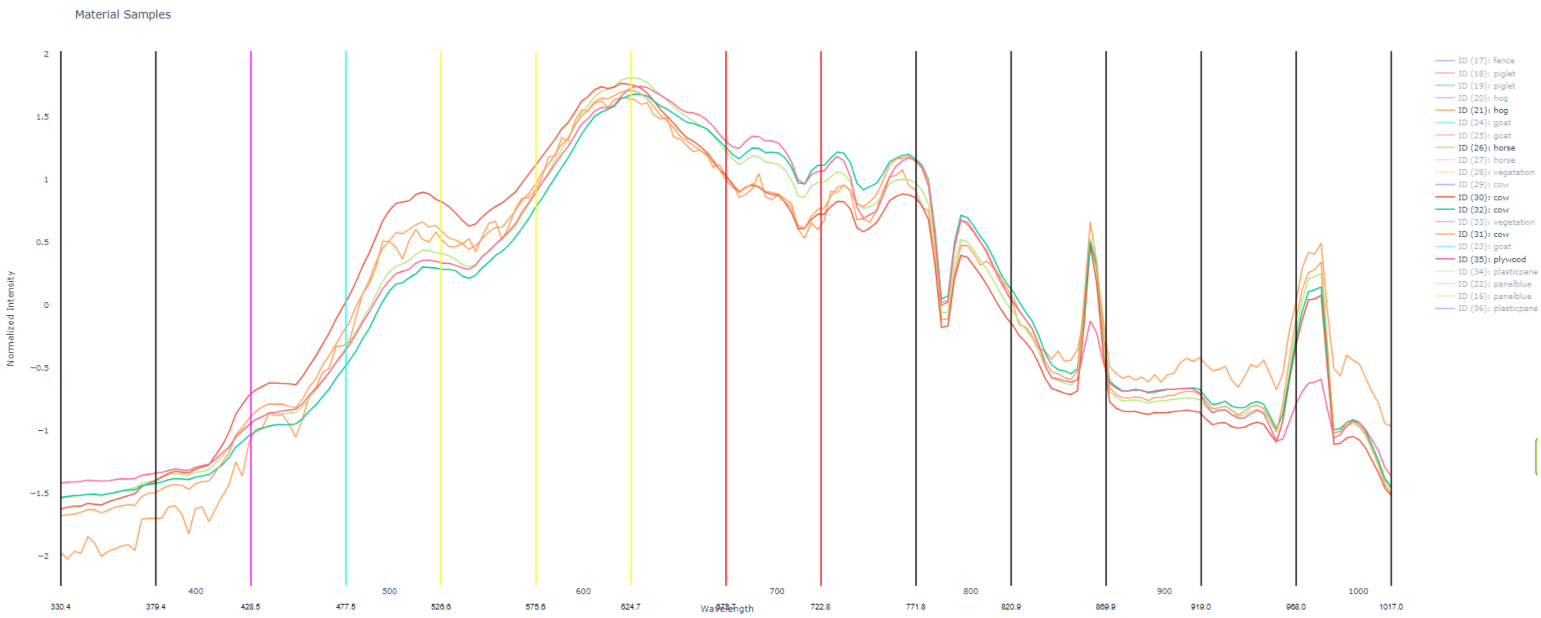
Results: analysis_RESDAN_1_Mission_35.csv

○


Results: analysis_RESDAN_1_Mission_35.csv
material_name
vegetation
plywood
cow
goat
piglet
horse
plasticpanel
hog
fence
panelblue
130k
135k
140k
145k
150k
155k
160k
165k
timestamp
−10
−5
0
5
10
galvo_pos

○

Material Comparison



○

-
- Breakdown and analysis of key regions

4 Machine Learning Vs. Statistical Analysis

- Preprocessing
- Noise analysis
- Metrics vs real performance
- Mixing concerns