

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Raita-Hakola, A.-M.; Pölönen, I.

**Title:** Piecewise anomaly detection using minimal learning machine for hyperspectral images

**Year:** 2021

**Version:** Published version

**Copyright:** © Author(s) 2021

**Rights:** CC BY 4.0

**Rights url:** <https://creativecommons.org/licenses/by/4.0/>

**Please cite the original version:**

Raita-Hakola, A.-M., & Pölönen, I. (2021). Piecewise anomaly detection using minimal learning machine for hyperspectral images. In N. Paparoditis, C. Mallet, F. Lafarge, M. Y. Yang, J. Jiang, A. Shaker, H. Zhang, X. Liang, B. Osmanoglu, U. Soergel, E. Honkavaara, M. Scaioni, J. Zhang, A. Peled, L. Wu, R. Li, M. Yoshimura, K. Di, O. Altan, H. M. Abdulmuttalib, & F. S. Faruque (Eds.), XXIV ISPRS Congress Imaging today, foreseeing tomorrow, Commission III (V-3-2021, pp. 89-96). Copernicus Publications. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. <https://doi.org/10.5194/isprs-annals-V-3-2021-89-2021>

# PIECEWISE ANOMALY DETECTION USING MINIMAL LEARNING MACHINE FOR HYPERSPECTRAL IMAGES

A-M. Raita-Hakola<sup>a,\*</sup>, I. Pölönen<sup>a</sup>,

<sup>a</sup> Faculty of Information Technology, University of Jyväskylä, 40100, Jyväskylä, Finland -  
(anna.m.hakola, ilkka.polonen)@jyu.fi

Commission III, WG III/4

**KEY WORDS:** Hyperspectral imaging, Minimal Learning Machine, Piecewise approach, Anomaly detection, Real-time computation, Machine learning

## ABSTRACT:

Hyperspectral imaging, with its applications, offers promising tools for remote sensing and Earth observation. Recent development has increased the quality of the sensors. At the same time, the prices of the sensors are lowering. Anomaly detection is one of the popular remote sensing applications, which benefits from real-time solutions. A real-time solution has its limitations, for example, due to a large amount of hyperspectral data, platform's (drones or a cube satellite) constraints on payload and processing capability. Other examples are the limitations of available energy and the complexity of the machine learning models. When anomalies are detected in real-time from the hyperspectral images, one crucial factor is to utilise a computationally efficient method. The Minimal Learning Machine is a distance-based classification algorithm, which can be modified for anomaly detection. Earlier studies confirm that the Minimal Learning Machine (MLM) is capable of detecting efficiently global anomalies from the hyperspectral images with a false alarm rate of zero. In this study, we will show that by using a carefully selected lower threshold besides the higher threshold of the variance, it is possible to detect local and global anomalies with the MLM. The downside is that the improved method is highly sensitive with respect to the noise. Thus, the second aim of this study is to improve the MLM's robustness with respect to noise by introducing a novel approach, the piecewise MLM. With the new approach, the piecewise MLM can detect global and local anomalies, and the method is significantly more robust with respect to noise than the MLM. As a result, we have an interesting, easy to implement and computationally light method which is suitable for remote sensing applications.

## 1. INTRODUCTION

The Hyperspectral (HS) image typically consists of a stack of frames, where each frame represents the intensity of a different wavelength of light, and each pixel has its spectrum. The HS image anomaly detection is a process where the image is processed pixel-by-pixel. Each pixel spectrum is evaluated, and the aim is to detect pixels whose spectral signature differs from their surroundings. The high-dimensional HS data is suitable for the identification, characterisation and anomaly detection tasks of the targets with high accuracy and robustness (Camps-Valls and Bruzzone, 2005, Bioucas-Dias et al., 2013). The challenges of the spectral anomaly detection methods are usually combinations of the large amounts of data, platform's constraints on payload, processing capability, and restricted available energy with complex machine learning models (Haut et al., 2018, Caba et al., 2020).

The technical evolution of the earth observation instruments on the airborne and satellite platforms have raised the sensor capability of producing an almost continuous high-dimensional data stream (Chen et al., 2018). In remote sensing platforms, the exponentially growing high dimensional data challenges the real-time processing, the data analysis processes and the technical features (Chen et al., 2018, Bioucas-Dias et al., 2013). Despite the challenges, one of the main advantages of real-time processing is improved data quality since it is not compressed and transmitted to the processor (Chen et al., 2018). The higher precision raw data might increase the accuracy of the data analysis. Other advantages are the reduced need for the communication between the ground equipment and the platform, the reduced need for the data

processing on the ground and the possibility to get the real-time responses from the platform (Che et al., 2018).

The Minimal Learning Machine (MLM) is easy to implement, computationally efficient and fast machine learning method (de Souza et al., 2015) which is an effective alternative for detecting global anomalies from the HS images (Pölönen et al., 2020). MLM is a distance-based method, which utilises the mapping between the input and the output distances. Input distance is the distance between the training set and its subset  $R$ , representing the selected training points. The output distance is calculated from the label values of the training set  $X$  to the subset  $R$ 's label values. In this study, we will calculate a linear model between the distances and estimate whether a certain pixel spectrum is an anomaly by using threshold values (Pölönen et al., 2020). The approach we are using is an example of a semi-supervised learning method (Prasad et al., 2009).

This study is an independent continuation of the previous research (Pölönen et al., 2020). The study has been implemented using consistent data with the previous research and using the same methods accurately. The aim is to improve the previous results with a new approach by proving that the MLM can detect local and global anomalies. With the new approach, the method can be more robust for the noisiness of the data. The research differs from the previous with its two test setups. The first setup will concentrate on the local and global anomalies by examining the variance thresholds. The second setup will introduce the piecewise MLM approach and provide robustness for the noise.

Our hypothesis is that by implementing the MLM with a piecewise approach, class-by-class, we can significantly improve results of the previous MLM anomaly detection method (Pölönen

\*Corresponding author.

et al., 2020). We will show that the MLM anomaly detection method (Pölonen et al., 2020) is capable of detecting both local and global anomalies, by using the lower and higher variance threshold values, but the method is still highly sensitive with respect to the noise. The new piecewise MLM is capable of detecting the local and global anomalies being more robust with respect to the noise. The piecewise MLM increases the accuracy of the MLM anomaly detection method, but on the downside, it might not be as fast as the MLM anomaly detecting method.

The paper is organised as follows. Section 2. describes the methods and demonstration materials. The results are introduced in section 3.. The discussion is the fourth section, and the final conclusions can be found in section 5..

## 2. MATERIAL AND METHODS

### 2.1 Minimal Learning Machine

The Minimal Learning Machine (MLM) is a computationally cheap distance-based machine learning method (de Souza et al., 2015). With HS images, the MLM offers tools for classification (Hakola and Pölonen, 2020) and anomaly detection (Pölonen et al., 2020) applications. The basic idea of the MLM is to utilise linear mapping between the distances of input and output.

When we are implementing the MLM with HS images, the input distances are  $d(\mathbf{x}_i, \mathbf{m}_k)$  and output distances  $\delta(\mathbf{y}_i, \mathbf{t}_k)$ , where  $\mathbf{x}_i \in X \subset R^D$  are the training set with  $D$  wavebands and  $\mathbf{m}_k \in R$  are the randomly sampled subset of  $X$ . The labels of the training set  $Y$  and its subset  $\mathbf{t}_k \in T$  are correspondingly  $\mathbf{y}_i \in Y \subset R$  and  $\mathbf{t}_k \in T$ . The size of the training set  $X$  is  $N$  samples, and the size of the subset  $R$  is  $K$  samples.

By defining two matrices, based on these distances  $\Delta_y \in R^{N \times K}$  and  $\mathbf{D}_x \in R^{N \times K}$ , and assuming the linear mapping between these two distance matrices, we have a linear model

$$\Delta_y = \mathbf{D}_x \mathbf{B} + \mathbf{E}, \quad (1)$$

where  $\mathbf{B}$  are the coefficients and  $\mathbf{E}$  is the residual. Coefficients  $\mathbf{B}$  can be approximated using the ordinary least squares estimator

$$\hat{\mathbf{B}} = (\mathbf{D}_x^T \mathbf{D}_x)^{-1} \mathbf{D}_x^T \Delta_y. \quad (2)$$

As a result, the  $\hat{\mathbf{B}}$  is a linear model between the distances of  $\delta(\mathbf{y}_i, \mathbf{t}_k)$  and  $d(\mathbf{x}_i, \mathbf{m}_k)$ . Distances between a new spectrum  $\mathbf{x}_n$  and its label  $\mathbf{y}_n$  is

$$\delta(\mathbf{y}_n, T) = d(\mathbf{x}_n, R) \hat{\mathbf{B}}. \quad (3)$$

Outputs  $\mathbf{y}_n$  can be estimated by solving an optimisation problem

$$\min_{\mathbf{y}_n} \sum_{k=1}^K ((\mathbf{y}_n - \mathbf{t}_k)^T (\mathbf{y}_n - \mathbf{t}_k) - \delta^2(\mathbf{y}_n, T))^2. \quad (4)$$

The  $\mathbf{y}_n$  is the computationally most expensive part of the classification. On the anomaly detection version, we do not have to estimate it. If the new spectrum  $\mathbf{x}_n$  is inside of the training set, the label  $\mathbf{y}_n$  is nearby points in the subset  $T$ . Then, the distribution of estimated distances  $\delta(\mathbf{y}_n, T)$  should be relatively similar to training phase distances in  $\Delta_y$ . Whether the  $\mathbf{x}_n$  is an anomaly or outlier can be detected in  $\delta(\mathbf{y}_n, T)$  by studying its behaviour.

In this study, we use the variance

$$\text{Var}(\delta(\mathbf{y}_n, T)) = \frac{1}{K} \sum_{i=1}^K (\delta_i - \bar{\delta})^2 \quad (5)$$

to detect anomalies, where  $K$  is number of elements in a vector  $\delta$  and  $\bar{\delta}$  is average value of the  $\delta$ . For the variance, we set two threshold values that reveals the anomalous spectra from the dataset. The lower threshold is the key to the detection of the local anomalies and the higher threshold exists for the global anomalies.

The threshold values must be selected carefully. Therefore, it is a useful practice to use an optimiser. It can be a parameter search function that loops through different combinations of lower and higher threshold values, calculates and compares the accuracy rates and returns the highest score threshold values.

### 2.2 Piecewise MLM approach

In this study, we developed a piecewise MLM approach, which aims to improve the accuracy rate of the previous version of the MLM anomaly detection method (Pölonen et al., 2020) by including the detection of the local anomalies to the results being more robust for the noisiness of the data.

---

#### Algorithm 1: Training of the piecewise MLM

---

**Input:**  $X, R, Y, T$

Calculate distance  $\mathbf{D}_x$

**for each class  $i$  do**

    Re-label class  $i$  to 0 and other classes to 1

    Calculate distance  $\Delta_y$

    Calculate  $\hat{\mathbf{B}}_i = (\mathbf{D}_x^T \mathbf{D}_x)^{-1} \mathbf{D}_x^T \Delta_y$

**end**

**Return:** Piecewise MLM model  $\hat{\mathbf{B}}_{\text{pw}} = [\hat{\mathbf{B}}_1, \hat{\mathbf{B}}_2, \dots]$

---



---

#### Algorithm 2: Anomaly detection using the piecewise MLM model

---

**Input:** new data  $X_{\text{new}}, R, \hat{\mathbf{B}}_{\text{pw}}$ , upper and lower thresholds

$\gamma_l$  and  $\gamma_u$

Calculate distance  $d(X_{\text{new}}, R)$

**for each  $\hat{\mathbf{B}}_i$  do**

    Calculate  $\delta(\mathbf{y}_{\text{new}}, T)_i = d(X_n, R) \hat{\mathbf{B}}_i$

    Calculate  $\text{Var}(\delta(\mathbf{y}_{\text{new}}, T))_i$

**end**

Calculate  $\text{score} = \sum_i \text{Var}(\mathbf{y}_{\text{new}}, T)_i$

**if  $\gamma_l < \text{score}_n < \gamma_u$  then**

$A_n = 1$

**else**

$A_n = 0$

**end**

**Return:** Anomaly detection  $A_n$

---

The piecewise MLM approach means that instead of one MLM model, we are re-training class-by-class the linear model  $\hat{\mathbf{B}}_i$ . The training set  $X$  and the subset  $R$  remain unchanged. The labels  $Y$  and its subset  $T$  are updated so that every spectrum that does not belong to the class, regardless of whether it belongs to the anomalies or another classes are labelled as anomalies. The algorithm 1 shows the implementation for the piecewise MLM training. Fig. 1 explains the idea of the piecewise MLM implementation by visualising the class-wise re-labelled ground truth data  $Y$ .

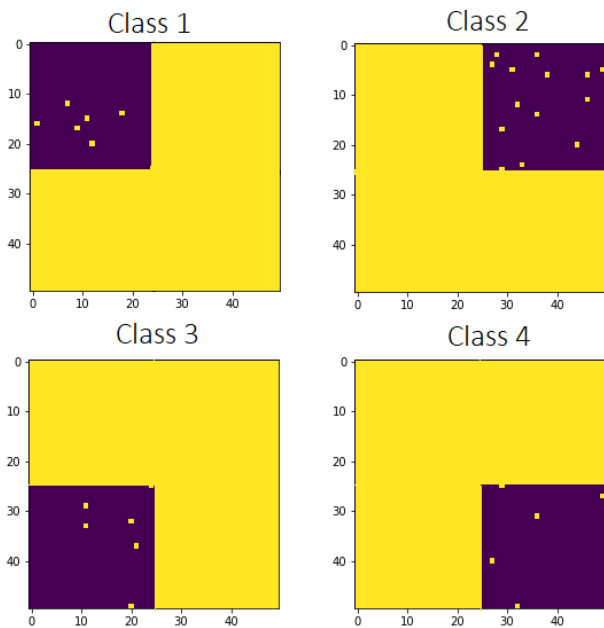


Figure 1: The piecewise MLM re-trains the MLM model class-by-class. Above is the visualisation of the re-labelled  $Y$ , where only the class we are expecting is labelled as zero (0). All the rest of the pixel spectrums, including the anomalies, are re-labelled with one (1), which means an anomaly. Zero (0): Purple, an expected spectrum, One (1): Yellow, an anomaly spectrum. The subset  $T$  is labelled similarly.

The detection of the actual anomalies is shown in the algorithm 2. The algorithm uses piecewise trained MLM model  $\hat{B}_i$  for each class  $i$ . Then the Equation 5 is calculated for each class. These variances are summed together and both of the lower and upper thresholds are used to detect anomalies. In our tests, we were using the Python libraries NumPy, scikit-learn and sciPy. The computations on this study were done with a Dell laptop (Intel Core i7-9850H and 16GB memory).

### 2.3 ColorChecker data

In this study, we used the same datasets with the previous study of the MLM anomaly detection method (Pölonen et al., 2020). This way, the results are comparable with the previous results. The first data is a HS image, captured using a visible and near-infrared HS imager, which is manufactured by the VTT (Saari et al., 2013). The dataset represents a four colours subset of a X-Rite's ColorChecker and it has 100 wavebands from 450nm to 750nm. The dataset is divided into training and testing portions. On the training set, we had 2500 pixels. Each class is represented with 625 pixels. The training set does not have any anomalous pixels.

The testing set represents the same colours as the training set, but it has a different spatial location. The testing set has a similar size and portions to the training set, but it contains randomly placed 30 anomalous pixels, which are captured from the different parts and colours of the ColorChecker. The first test setup with ColorChecker data was implemented without noise (both training and test data sets), on the second test setup, the measurements were done with different amounts of generated noise in the data. The noise drawn from the uniform distribution was added to each pixel. In the test, the level of the noise varied between 0-0.4.

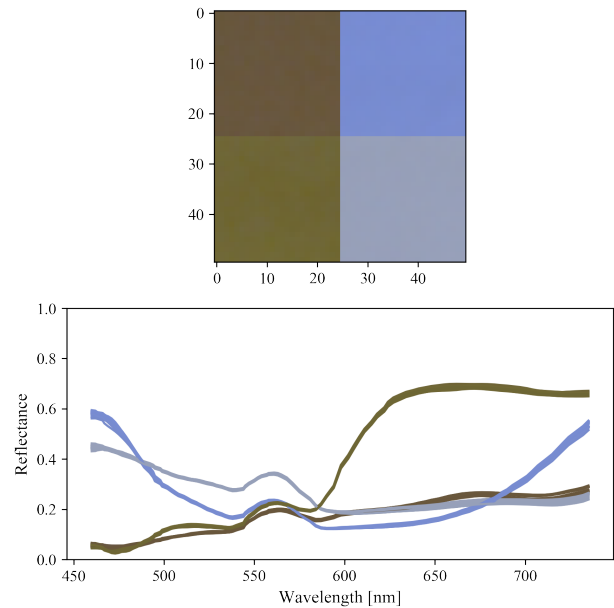


Figure 2: Illustration of the training dataset. Above is an "RGB" presentation of the ColorChecker dataset in the spatial dimension. Below are the samples of the reflectance spectra from each of the classes. Image is originally from (Pölonen et al., 2020).

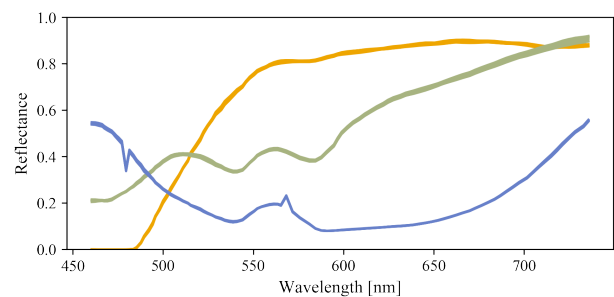


Figure 3: The spectra of the anomalous pixels on the ColorChecker data, which are randomly distributed to the test dataset. The test dataset has the same dimensions as the training set ( $50 \times 50 \times 100$ ). Image is originally from (Pölonen et al., 2020).

The training data and the ground truth data were randomised on the implementation phase and subset  $R$  with its 250-pixel spectra were selected randomly.

### 2.4 Forest data

The second dataset represents a Finnish forest. The tree species are mainly spruces, pines and birches. The dataset was originally captured for the tree species classification purposes (Nevalainen et al., 2017, Nezami et al., 2020, Polonen et al., 2018). The details of the dataset are described in (Nevalainen et al., 2017). The forest dataset consists of orthophoto mosaic images of the area (Honkavaara et al., 2013). The dataset was captured with a high spatial resolution (9 cm) ground sampling distance (GSD), and it has 38 spectral bands from 507 nm to 820 nm. In this study, we used a ( $1500 \times 1400 \times 38$ ) subset of the original data for the training. The selected training data is illustrated as a narrowband RGB image in Fig. 4, image A.

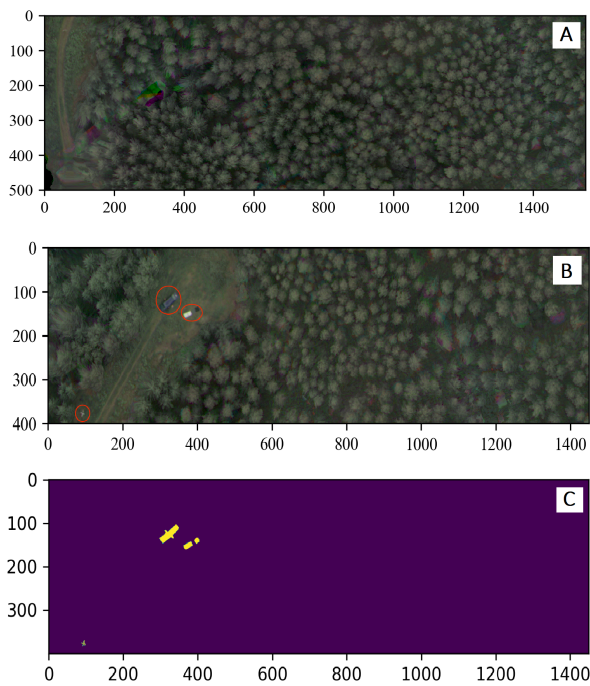


Figure 4: Image A: The training data for the anomaly detection from the forest. Image B: The forest dataset’s test data for the anomaly detection. Inside the red circulated areas are the anomalous objects (three reflectance panels and one black panel, one blue van, and a cross-shaped georeferencing signal). The third image is the hand-drawn ground truth image, for the test data. Images A and B are originally from (Pölonen et al., 2020).

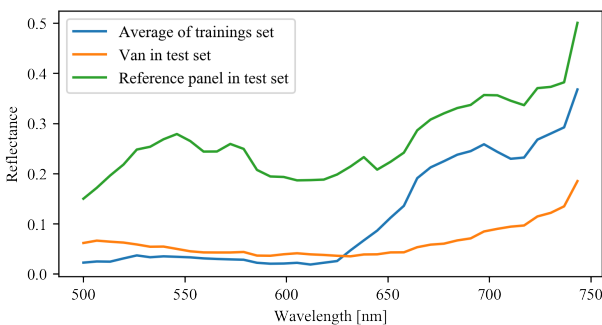


Figure 5: The mean spectra of the training dataset (forest data) and the samples from the van and the reflectance panel (nominal reflectivity of 0.5). Image is originally from (Pölonen et al., 2020).

The testing data is represented in Fig. 4, image B. It is a subset from the same remote sensing data as the training data, but it has some anomalies that are marked with red circles on the figure. The anomalies are three reflectance panels and one black panel, a blue van and a cross-shaped georeferencing signal. More details of these anomalies can be found from (Pölonen et al., 2020). Fig. 5 represents the average spectra of the training set compared to the example spectra of the anomalies.

The preprocessing of the forest data performed as follows. At first, the 100 000 pixel spectra samples were selected randomly from the training HS image for the training set  $X$ . The subset  $R$

(100 samples) were selected randomly from the training set  $X$ . We performed a k-means clustering ( $k=3$ ) (Pedregosa et al., 2011) for testing, which produced the needed ground truth labels for the HS image. The clustering time was excluded from the measured training and detecting times. We created the ground truth labels for the testing data from carefully pixel-by-pixel hand-drawn ground truth image. The ground truth image is visualised in Fig. 4, image C.

### 3. RESULTS

#### 3.1 Performance comparison

With the ColorChecker datasets, the MLM and the piecewise MLM were implemented using cosine distance metric (De Carvalho and Meneses, 2000, Yuhás et al., 1992) and variance with one (upper) or two (upper and lower) variance thresholds. The tests were repeated 20 times, and the reported results of the ColorChecker data are the mean values from the tests. The computationally interesting measures were the training and detection times. Accuracy was also calculated using the Scikit-learn metrics accuracy score method (Varoquaux et al., 2015)).

At first, we used only the upper threshold with the MLM, which was done similarly with the previous study (Pölonen et al., 2020). The threshold for the MLM’s higher variance was set to  $> 0.5$ . As a result, the method was not able to detect the local anomalies (Fig. 6). In the second phase, we used the MLM with two thresholds. The upper was set to  $> 0.5$  and the lower to  $< 0.12$ . After setting the thresholds, the accuracy improved significantly. With the piecewise MLM, the variance thresholds were set to  $> 0.7$  and  $< 0.4$ , respectively.

	A	B	C
Training time [s]	0.084	0.069	0.290
Testing time [s]	0.047	0.059	0.234
Accuracy [%]	80	98	100

Table 1: Comparison of the computation times for the ColorChecker dataset. A: MLM (cosine, variance, training set size 2500,  $R$  size 250), higher threshold:  $> 0.5$ . B: MLM (cosine, variance, training set size 2500,  $R$  size 250), lower threshold  $< 0.12$ , higher threshold:  $> 0.5$ . C: Piecewise MLM (cosine, variance, training set size 2500,  $R$  size 250), lower threshold  $> 0.7$ , higher threshold  $< 0.4$ .

Table 1 shows that the MLM is faster at training and detecting anomalies than the piecewise MLM. Using two carefully selected thresholds (the threshold values optimisation is described in 2.1), the MLM can detect both global and local anomalies with increased accuracy.

From Fig. 7 and Table 1, we can see that the accuracy rate is significantly improved with the second threshold and with the piecewise MLM approach. The MLM with the upper threshold reached 80% accuracy, leaving local anomalies undetected (red samples in Fig. 6). The MLM with two thresholds and the piecewise MLM could detect both the local and global anomalies. The accuracy of the MLM with two thresholds was 98%, and with the piecewise approach, it was 100%.

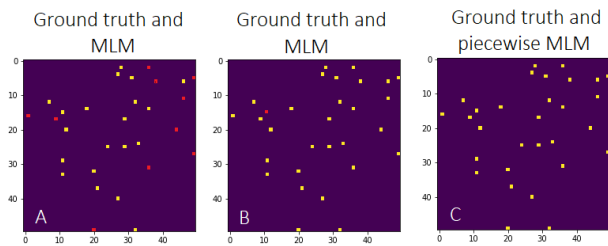


Figure 6: The joint anomaly detection maps of the ground truth and the MLM for the ColorChecker dataset. Yellow represents detected anomalies. Red spots are undetected local anomalies. A: MLM, only higher threshold, B: MLM with two thresholds, C: the piecewise MLM

### 3.2 Noise sensitivity test

Different amounts of uniformly distributed random noise were added to the training and test sets of the ColorChecker data. The accuracy of the anomaly detection from the data containing different amount of noise was evaluated with an area under curve (AUC) and by drawing the receiver operating characteristic curve (ROC) for the MLM and piecewise MLM setups.

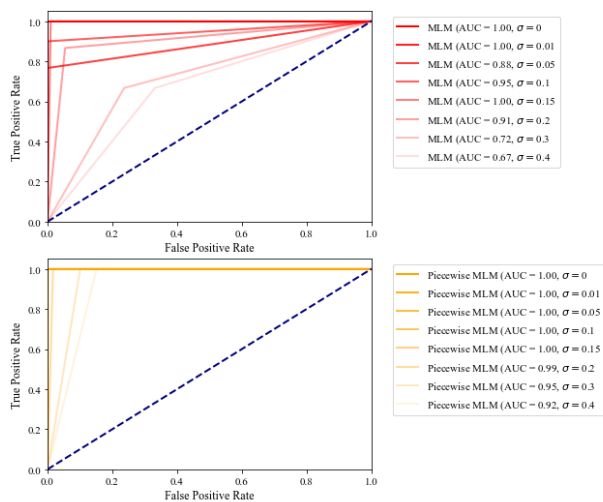


Figure 7: Comparison of the ROC-curves of the MLM and piecewise MLM with a different amount of noise in the data (the ColorChecker dataset). The noise drawn from the uniform distribution was added to each pixel. The level of the noise is denoted with  $\sigma$ . The ROC curves show that both approaches had no false positives with a low noise level. When the noise level increases, the piecewise MLM was significantly more robust with respect to the noise than the MLM.

The Fig. 7 shows that the piecewise approach is significantly more robust with respect to the noise than the MLM. Neither method had any false positives at first, but the number of false positives started to increase after adding noise. The effect of the noise on the performance is much more drastic for the MLM than it is for the piecewise MLM. The effect can be seen in how the increased noise affects the AUC. It ranges from 0.67 to 1.0 with the MLM and accordingly from 0.92 to 1.0 with the piecewise MLM.

The visualisation of the anomaly detection maps with the different noise levels in the data is shown in Fig. 8. Each pair is marked

with similar alphabets. The noise levels and the ground truth are shown in the figure. The effect of the decreasing accuracy, caused by the noise, can be seen from the MLM maps D-G, where the level of the noise was between 0.1 and 0.4 respectively. A similar effect can be seen with the piecewise MLM approach on F and G images (noise levels of 0.3 and 0.4).

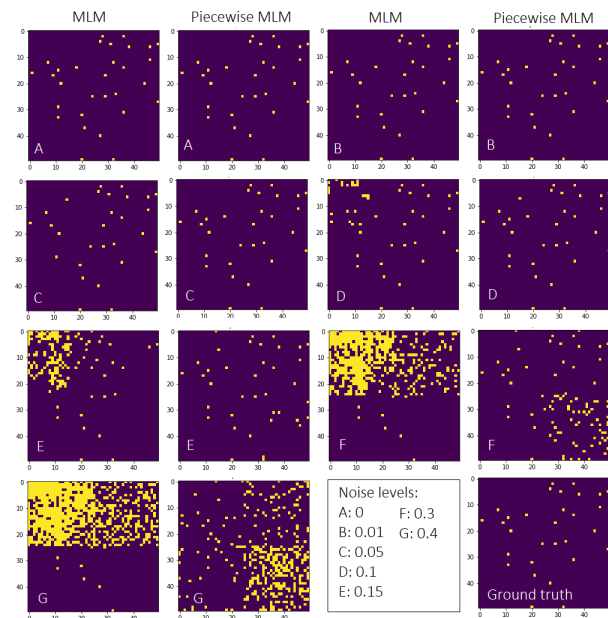


Figure 8: The ColorChecker dataset, the pairwise comparison of the MLM anomaly detection maps (on the left) and the piecewise MLM (on the right) with different amount of noise in the data. The level of the noise is marked with the alphabets.

### 3.3 Forest data

The visual evaluation of the forest data (Fig. 9) shows that both approaches can detect anomalies with a low false alarm rate. The most difficult anomaly to detect was the cross-shaped georeferencing signal on the left corner of the Fig. 4, image B. We tested the forest data with the MLM using different thresholds.

The MLM outperforms the piecewise approach in the forest data computation time comparison. It was the fastest method for both training and detecting tasks (Table 2).

Image on Fig.9	A-D	E-F
Training time [s]	0.390	1.238
Detecting time [s]	2.189	6.682

Table 2: The forest data, the comparison of the computation times for images on Fig. 9. A-D: MLM (cosine, variance, training set size 2500,  $R$  size 250). E-F: Piecewise MLM (cosine, variance, training set size 2500,  $R$  size 250).



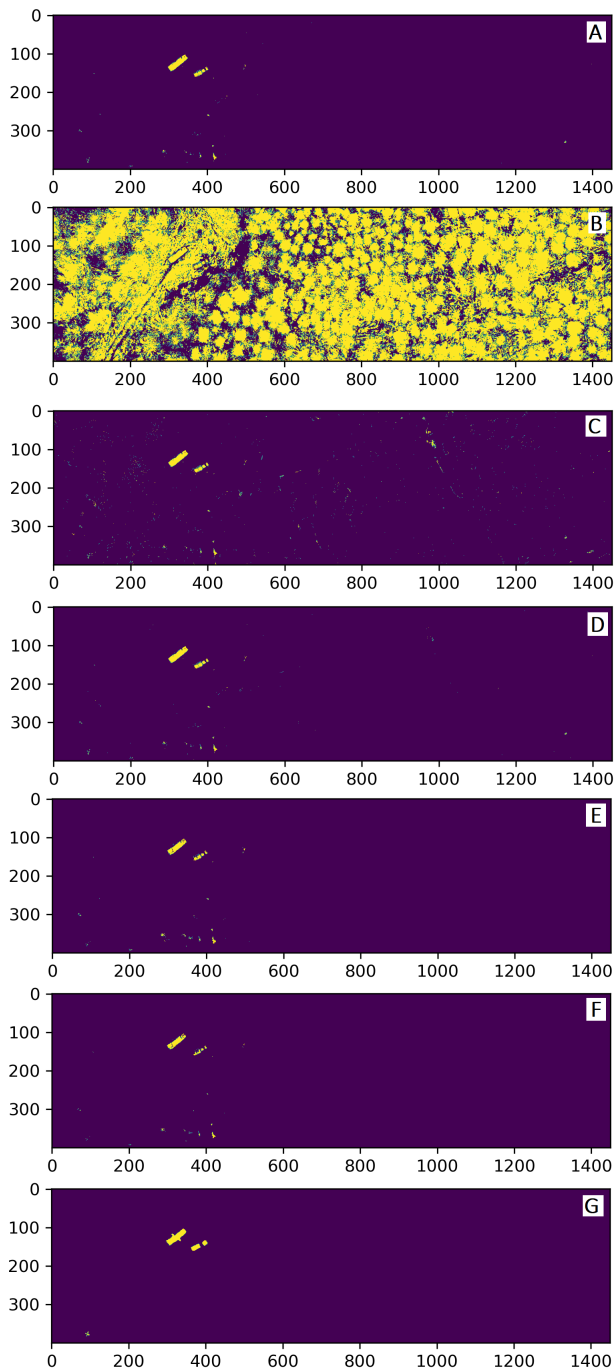


Figure 9: Comparison of the MLM (A-D) and the piecewise MLM (E-F) with the Ground truth (G) with the forest data. Thresholds and accuracy rates of the images are shown on Table 3.

Image on Fig.9	A	B	C
Higher threshold	0.1	0.1	0.1
Lower threshold	-	0.01	0.001
Accuracy rate [%]	99.86	30.10	99.57
Image on Fig.9	D	E	F
Higher threshold	0.1	1.2	1.0
Lower threshold	0.0001	0.07	0.07
Accuracy rate [%]	99.85	99.85	99.85

Table 3: The forest data, thresholds and accuracy rates for images on Fig. 9. A-C MLM, E-F Piecewise MLM.

The accuracy rate results show (Table 3) that if there is a need to use the lower limit, it must be chosen carefully. We can see from image B (Fig. 9) that if the lower limit is set too high, the false alarm rate will increase. Table 3 confirms that the range of the threshold is different between the methods, and the MLM lower threshold's sensitiveness is obvious.

The MLM performed the most accurately with only one threshold (image A in Fig. 9, Table 3), but images C, D and E show that the MLM can be implemented with two thresholds without compromising the accuracy rate significantly.

#### 4. DISCUSSION

This study confirms the previous observations (Pölonen et al., 2020) that the MLM is a fast and efficient method for detecting anomalies from the HS images. However, the method's weaknesses were the inability to detect local anomalies and the high sensitivity with respect to noise. This study proved that by using a carefully selected variance lower threshold value with the higher variance threshold value, it is possible to detect both local and global anomalies. The other main result of this study was the piecewise MLM, which seems to be a significantly more robust with respect to noise than the MLM. The ColorChecker noise comparison shows in Fig. 7 that the MLM's lowest AUC was 0.67, and with a similar amount of noise (0.4), the piecewise MLM approach reached 0.92 AUC.

With the MLM method, usually, size of the training data's subset  $R$  influences on the accuracy rate. By increasing the size of  $R$ , we can increase the accuracy of the results (de Souza et al., 2015, Hakola and Pölonen, 2020). In this study, we increased the accuracy without increasing the size of  $R$ , which was 250 samples with the ColorChecker data, containing randomly picked samples from all of the classes. The piecewise MLM was implemented by training the MLM model class-by-class and re-labelling  $Y$  and  $T$  so that there were only two classes. On the implementation, zero (0) represented the expected spectrum and one (1) represented anomalies, containing all of the rest of the ground truth classes and anomalies (Fig. 1). With the ColorChecker data and the piecewise MLM approach, we trained the MLM four times, which was the number of the ground truth classes. The class-by-class models seemed to be more accurate, even though the actual size of  $R$  remained unchanged. One of the reasons why the piecewise MLM is outperforming with the noisy data is that the approach provides more accurately trained models since there are stricter rules on the deviations and expected values.

The piecewise MLM is more robust with respect to noise than the MLM, and therefore it is more sensitive to detect the local anomalies, even if they are close to the classes of the expected spectrum. With the MLM, the lower threshold value must be set carefully, and only the small enough values will work, but with the piecewise MLM, the range of the possible threshold values seems to be wider.

The downside of the piecewise MLM approach can be seen in Table 2. The average training times of the MLM ranged from 0.069 to 0.084 seconds, where the piecewise approach took 0.290 seconds. The detecting times were similarly from 0.047 to 0.049 and 0.234 seconds.

The MLM has been performing well against other anomaly detection methods (Pölonen et al., 2020). Previous studies confirms that a one-class support vector machine (OC-SVM) is faster in training and anomaly detection than the MLM, but an Isolation forest is slower than the MLM. Global and local RX are slower

in the detection, but the training phase of the RX is faster than the MLM's training phase. The piecewise MLM is slower than the MLM in the training and anomaly detection, and it will probably lose against some of the mentioned reference methods. On the other hand, Fig. 10 shows promising results on the comparison of the anomaly maps. With the ColorChecker data, Fig. 10 shows that the piecewise MLM and the MLM are the most accurate methods and they performed without false positives.

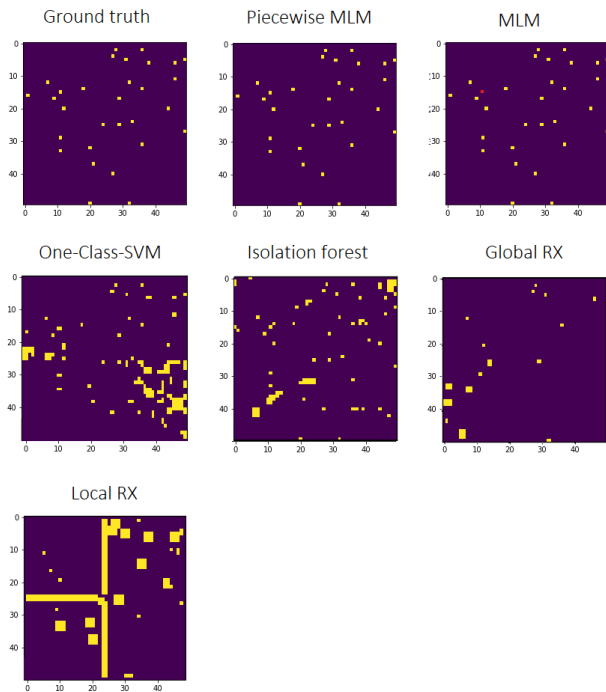


Figure 10: The anomaly detection maps of the ColorChecker dataset. The ground truth, the Piecewise MLM and the MLM: implemented with cosine distance metric, using variance and two thresholds (explained in subsection 3.1). The size of the  $X$  was 2500, and  $R$  was 250. The anomaly maps of the reference methods are originally from the (Pölonen et al., 2020).

The forest data comparison and the visual evaluation (Fig. 9) shows that there was no significant improvement on the anomaly maps between the different approaches of the MLM. The MLM performed faster than the piecewise MLM approach. There were no improvements on the piecewise MLM's results towards the MLM because the data does not necessarily contain enough local anomalies or noisiness for the results to be improved.

We used the k-means clustering method for creating the labels for the forest training data (chapter 2.4). It occurs that the initialization of the clustering might affect the final results and accuracy. If the randomly selected locations of the initial clusters are, for example, somehow overlapping or close to the other final classes, some of the labels might not be in the right classes after the clustering is performed. Another note for the results is that the accuracy rates of the forest comparison (Table 3) were based on the hand-drawn ground truth image, which affects accuracy. Especially the borders of the anomalous objects were slightly unclear when we were drawing the ground truth image.

The previous studies confirms that the MLM was outperforming against the reference methods with the forest data (Pölonen et al., 2020). Based on the visual evaluation (Fig. 11), this study confirms those findings. The MLM Forest data comparison (Table 3,

Fig. 9) shows that the threshold levels seem to be more sensitive with MLM than with the piecewise MLM. If the values are set too high or low, the accuracy will decrease. The range of the acceptable values was narrower on the MLM than in the piecewise MLM approach.

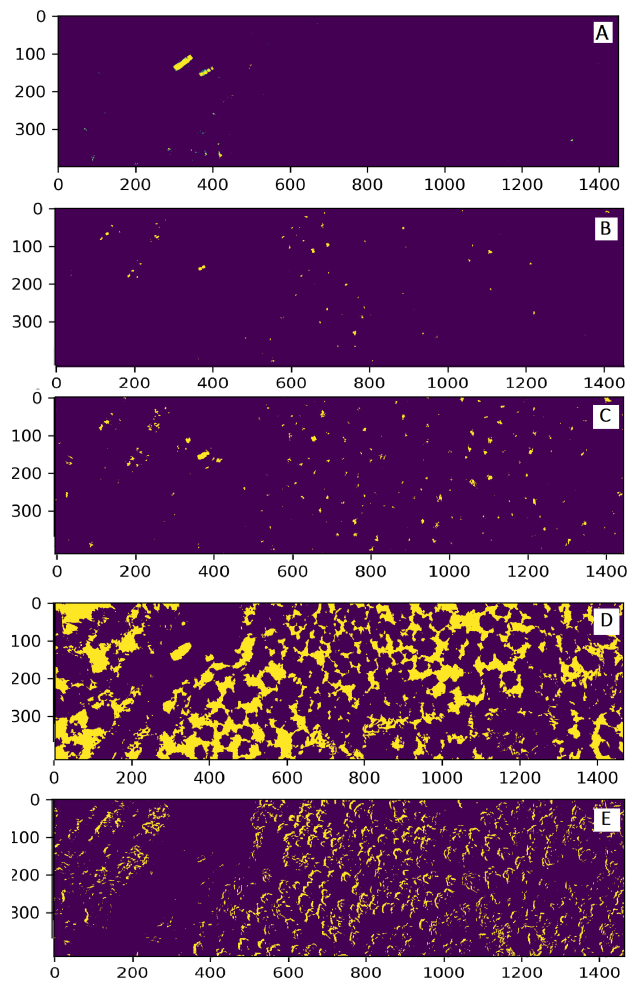


Figure 11: The anomaly detection maps of the forest data. A: MLM, implemented with cosine distance metric, using variance and one threshold (Table 3, image A). The size of the  $X$  was 2500, and  $R$  was 250. B: One-Class-SVM, C: Isolation forest, D: Global RX, E: Local RX. The anomaly maps of the reference methods are originally from the (Pölonen et al., 2020).

In this study, estimating whether the pixel is anomalous was carried out at individual pixel level. However, for the future, it is possible to improve these methods to include the pixel neighbours for the estimation. Another interesting idea to explore is the sensitivity analysis for anomalies that are smaller than one pixel. The subpixel level sensitivity studies could reveal the detection rates for anomalies that, for example, can be seen among the trees in Fig.9.

Since the MLM seems to perform well in the HS image anomaly detection, the answer to the question, which MLM version to select for the anomaly detection, depends on the data. If the HS image is large, moreover, if it contains both local and global anomalies, and the noise levels are unknown or high; the piecewise MLM would be the method to use. If the data is less noisy, the selection would be the MLM. The MLM can detect both local and global anomalies, and it is faster than the piecewise MLM,



but the piecewise MLM might be an easier method to use since it is not so sensitive with the threshold values and it is more robust with respect to noise.

## 5. CONCLUSIONS

The MLM anomaly detection method can be extended to detect both local and global anomalies from the hyperspectral images, but the method's accuracy is highly sensitive with respect to noise. The piecewise MLM is a new approach, which can perform similar anomaly detection tasks from the hyperspectral data, being significantly more robust with respect to the noise. The MLM or piecewise MLM approach should be selected depending on whether the data contains low or high noise levels. If the data is less noisy, the MLM would be a computationally effective solution for anomaly detection. If the data contains higher levels of noise, the choice would be the piecewise MLM. Both of these methods are easy to implement, and they can be used in real-time hyperspectral anomaly detection applications. The methods can be implemented with small, single-board computers.

## ACKNOWLEDGEMENTS

This study is partly funded by the Academy of Finland (Grant No. 327862). We thank our colleague Kimmo Riihiahio for assistance and comments that improved the manuscript.

## REFERENCES

Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N. M. and Chanussot, J., 2013. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine* 1(2), pp. 6–36.

Caba, J., Díaz, M., Barba, J., Guerra, R., de la Torre, J. A. and López, S., 2020. Fpga-based on-board hyperspectral imaging compression: Benchmarking performance and energy efficiency against gpu implementations. *Remote Sensing* 12(22), pp. 1–37.

Camps-Valls, G. and Bruzzone, L., 2005. Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* 43(6), pp. 1351–1362.

Che, W., Sun, L., Zhang, Q., Tan, W., Ye, D., Zhang, D. and Liu, Y., 2018. Pixel based bruise region extraction of apple using Vis-NIR hyperspectral imaging. *Computers and Electronics in Agriculture* 146(July 2017), pp. 12–21.

Chen, C., Li, W., Gao, L., Li, H. and Plaza, J., 2018. Special issue on advances in real-time image processing for remote sensing. *Journal of Real-Time Image Processing* 15(3), pp. 435–438.

De Carvalho, O. A. and Meneses, P. R., 2000. Spectral correlation mapper (scm): an improvement on the spectral angle mapper (sam). In: *Summaries of the 9th JPL Airborne Earth Science Workshop*, JPL Publication 00-18, Vol. 9, JPL Publication Pasadena, CA.

de Souza, A. H., Corona, F., Barreto, G. A., Miche, Y. and Lendasse, A., 2015. Minimal Learning Machine: A novel supervised distance-based approach for regression and classification. *Neurocomputing* 164(January), pp. 34–44.

Hakola, A.-M. and Pölönen, I., 2020. Minimal learning machine in hyperspectral imaging classification. In: L. B. Santi, F. Bovolo and Emanuele (eds), *Image and Signal Processing for Remote Sensing XXVI*, Vol. 11533, p. 26.

Haut, J. M., Paoletti, M. E., Plaza, J. and Plaza, A., 2018. Fast dimensionality reduction and classification of hyperspectral images with extreme learning machines. *Journal of Real-Time Image Processing* 15(3), pp. 439–462.

Honkavaara, E., Saari, H., Kaivosoja, J., Pölönen, I., Hakala, T., Litkey, P., Mäkynen, J. and Pesonen, L., 2013. Processing and assessment of spectrometric, stereoscopic imagery collected using a lightweight UAV spectral camera for precision agriculture. *Remote Sensing* 5(10), pp. 5006–5039.

Nevalainen, O., Honkavaara, E., Tuominen, S., Viljanen, N., Hakala, T., Yu, X., Hyyppä, J., Saari, H., Pölönen, I., Imai, N. N. and Tommaselli, A. M., 2017. Individual tree detection and classification with UAV-Based photogrammetric point clouds and hyperspectral imaging. *Remote Sensing*.

Nezami, S., Khoramshahi, E., Nevalainen, O., Pölönen, I. and Honkavaara, E., 2020. Tree species classification of drone hyperspectral and RGB imagery with deep learning convolutional neural networks. *Remote Sensing*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, pp. 2825–2830.

Polonen, I., Annala, L., Rahkonen, S., Nevalainen, O., Honkavaara, E., Tuominen, S., Viljanen, N. and Hakala, T., 2018. Tree Species Identification Using 3D Spectral Data and 3D Convolutional Neural Network. *Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing*.

Pölönen, I., Riihiahio, K., Hakola, A.-M. and Annala, L., 2020. Minimal learning machine in anomaly detection from hyperspectral images. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 43(B3), pp. 467–472.

Prasad, N. R., Almanza-Garcia, S. and Lu, T. T., 2009. Anomaly detection. *Computers, Materials and Continua* 14(1), pp. 1–22.

Saari, H., Pölönen, I., Salo, H., Honkavaara, E., Hakala, T., Holmlund, C., Mäkynen, J., Mannila, R., Antila, T. and Akujärvi, A., 2013. Miniaturized hyperspectral imager calibration and UAV flight campaigns. *Sensors, Systems, and Next-Generation Satellites XVII 8889*(October 2013), pp. 888910.

Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., Pedregosa, F. and Mueller, A., 2015. Scikit-learn. *GetMobile: Mobile Computing and Communications* 19(1), pp. 29–33.

Yuhas, R. H., Goetz, A. F. H. and Boardman, J. W., 1992. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm. In: *Summaries of the Third Annual JPL Airborne Geoscience Workshop*, Vol. 1, pp. 147–149.