Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

Trym Woldstad

Self-Organizing Maps for Fusion of Spectral Images

Master's thesis in Industriell Kybernetikk
Supervisor: Joseph Garrett
June 2023

NTNU
Norwegian University of Science and Technology

Trym Woldstad

Self-Organizing Maps for Fusion of Spectral Images

Master's thesis in Industriell Kybernetikk
Supervisor: Joseph Garrett
June 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

NTNU
Norwegian University of
Science and Technology

Self-Organizing Maps for Fusion of Spectral Images

Trym Woldstad

June 2023

Master Thesis

Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics
Norwegian University of Science and Technology

Supervisor: Joseph Garrett

Preface

This master thesis is the culmination of my studies in Industrial Cybernetics at NTNU, focusing particularly on Machine Learning and AI in my final year. It has been undertaken as part of the HYPSO mission, a journey that lasted from 1st February to 16th June. Special appreciation is directed towards Joseph Garrett, whose guidance and counsel during the feasibility project and the master thesis were invaluable.

This thesis encapsulates the knowledge and experience I gained during my time at NTNU, and I look forward to its impact in the future, especially towards AI. Trondheim, June 16, 2023

TRYM WOLDSTAD

Remark:

Special thanks to Professor Emeritus Marvin Rausand for creating the LaTeX template used in this thesis.

Summary

This thesis explores a novel application of Self-Organizing Maps (SOMs) for the purpose of multimodal image fusion in remote sensing. The research builds upon the work of the HYPSO satellite project and the SeaBee project, both aiming to enhance the monitoring and mapping of coastal environments for ecological preservation.

The thesis addresses a key challenge in the field: the processing of vast amounts of high-dimensional data collected by different imaging modalities, including RGB, multispectral, and hyperspectral cameras. The motivation behind the task of fusing images is to keep the from the advantages of both hyperspectral and RGB images. While hyperspectral images possess high spectral resolution, facilitating precise identification of materials, they often suffer from low spatial resolution. Conversely, RGB images possess high spatial resolution but lack in spectral information.Traditional techniques for data fusion often struggle with real-world complexities, resulting in inefficiencies in time and computational power.

To overcome these limitations, the thesis proposes the use of SOMs, a type of shallow neural network. SOMs are advantageous due to their simplicity, reliability, and computational efficiency. With a unique capability of visualizing uncertainties, a deeper understanding of the underlying patterns and errors in the data set can be obtained. Key features of SOMs that are explored include their topological preservation property, which retains distances from the high-dimensional space in the outputted 2-dimensional space.

The thesis discusses potential challenges in implementing SOMs in the application of image fusion. The study includes testing and evaluation on both simulated and real data of the proposed SOM. The results demonstrate that the fusion of HSI and RGB data enables the SOM to project and infer spectral information beyond the limitations of the HSI, where only RGB data is available. The results and their potential impacts are expected to contribute significantly to the ongoing efforts of NIVA and NTNU, as well as the broader field of remote sensing.

Sammendrag

Denne oppgaven utforsker en ny anvendelse av selvorganiserende kart (SOM) for bildefusjon i satelitt-relaterte applikasjoner. Forskningen bygger på arbeidet til HYPSO-satellittprosjektet og SeaBee-prosjektet, som begge tar sikte på å forbedre overvåking og kartlegging av kystmiljøer for økologisk bevaring.

Oppgaven forsøker å løse en viktig problemstilling i forskningsfeltet: behandlingen av enorme mengder høydimensjonale data samlet inn av forskjellige sensorer, inkludert RGB, multispektrale og hyperspektrale kameraer. Motivasjonen bak oppgaven med å sette sammen bilder kommer fra de distinkte fordelene med både hyperspektrale og RGB-bilder. Mens hyperspektrale bilder har høy spektral oppløsning, noe som muliggjør presis identifikasjon av materialer, har de ofte lav optisk oppløsning. Omvendt har RGB-bilder høy optisk oppløsning, men mangler spektral informasjon. Tradisjonelle teknikker for datafusjon sliter ofte med kompleksiteten i den virkelige verden, noe som resulterer i ineffektivitet i både tid og beregningskraft.

For å overvinne disse begrensningene foreslår oppgaven bruk av SOM-er, en type grunt nevrale nettverk. SOM-er er fordelaktige på grunn av deres enkelhet, pålitelighet og beregningseffektivitet. Med en unik evne til å visualisere usikkerheter, kan en dypere forståelse av de underliggende mønstrene og avvikene i datasettet oppnås. En viktig egenskap ved SOM-er som utforskes inkluderer deres topologiske bevaringsegenskap, som bevarer avstander fra det høydimensjonale rommet i "input" til det 2-dimensjonale rommet i "output".

Oppgaven diskuterer potensielle utfordringer ved implementering av SOM-er i anvendelse av bildefusjon. Studien inkluderer testing og evaluering på både simulerte og ekte data fra SOM-en. Resultatene viser at sammenslåingen av HSI- og RGB-data gjør det mulig for SOM å projisere og utlede spektral informasjon utover begrensningene til HSI, der bare RGB-data er tilgjengelig. Resultatene og deres potensielle virkninger forventes å bidra til den pågående innsatsen til NIVA og NTNU, samt det bredere feltet fjernmåling.

Abbreviations

Table 1: List of Abbreviations

Abbreviation	Full Form
SOM	Self-Organizing Map
ML	Machine Learning
ANN	Artificial Neural Network
HSI	Hyper Spectral Image
RGBI	Red Green Blue Image
PCA	Principal Component Analysis
BMU	Best Matching Unit
HS	Hyperspectral
MS	Multispectral
DR	Dimensionality Reduction
SSL	Semi-supervised Learning
H-R	High-resolution

Contents

1 Introduction

1.1 Motivation

The coastal environment contains many ecologically important species such as kelp and seagrass. Monitoring and mapping the coast can be a tool for ensuring the health and prosperity of these species. For this purpose NIVA and the research infrastructure SeaBee, in collaboration with NTNU, is developing coastal mapping systems using drones an various imagers, such as an RGB-, a multispectral-, and a hyperspectral camera. The hyperspectral camera captures the most data, while the RGB covers a larger area with higher resolution.

Data fusion is an important research area within the field of remote sensing. This process describes the science of combining data from different signals or sources, with the goal of increasing the result. One example is earth-observation data, which can be acquired with many types of technologies, leading to regular aerial photos, GPS data, LiDAR, multi- or hyperspectral imagery (HSI) and so on. These types of data can display great diversity in the information it contains. This is the main reasoning behind why there is such a big interest in researching the field of data fusion for remote sensing.

1.2 Spatial and spectral resolution

RGB images(RGBI) and multispectral images(MSI) are typically photographed using drones or other aerial vehicles, flying at relatively low height with a high spatial resolution.This refers to the smallest distinguishable detail in the image and is often represented by the size of the pixels in the image. A higher spatial resolution will be able to better separate closely spaced objects in a scene. The spectral resolution refers the ability to distinguish between different wavelengths on the electromagnetic spectrum. RGB images are usually represented by three color channels, red, green and blue, representing a range of wavelengths on the visible spectrum from around 450 to 750 nm. HSI captures a much higher spectral resolution by dividing the spectrum into many narrow bands, ranging from the ultraviolet to near the infrared regions of the spectrum.

When working with data from a single source there is usually an overweight towards either of these resolutions. In the case were ones working with an RGB image, there spatial resolution is often high, and when working with an HSI, the spectral resolution is high. Fusing data from these two sources can improve the overall spatial resolution of the image, while maintaining the spectral resolution. The challenge lies in achieving a balance between these two resolutions.

1.3 Thesis Contributions

This thesis contributes to the project in the following way:
1. A novel way to fuse real data from HSI and RGBI using SOMs.
2. Evaluation of the models performance with different parameters.
3. Evaluation of the models performance on different HS scenes.
4. Strengths and weaknesses of fusion with simulated and real images.
5. Evaluation of the HS information obtained in areas with only RGB data available.

1.4 Thesis Outline

This project will focus on creating a straightforward multimodal image fusion technique, which can then be applied to the vast data sets (100s of GB) already collected by the SeaBee project. One possible fusion technique self-organizing maps (SOMs), a shallow neural network. The goal of the project is to fuse the data collected by these sensors to improve the strengths of each modality. The advantage of using SOMs for an application like this is its simplicity, reliability, as well as the ability to visualize uncertainties. Due to the fact that it is hard to acquire the ground truth for HSIs, this way of extracting patterns is useful when working with such high-dimensional data. HSI are often very large and computational expensive, and SOMs are therefore useful in both the pre-processing and the evaluation of these images. The possibility of training a SOM with fused data from both HS and MS sources to gain information where only MS data is present, is to be investigated in the thesis.

2 Theory

2.1 Remote sensing

Remote sensing has become a vital technique for collecting data about the Earth's surface and atmosphere without the need for direct contact, as described in various studies. Analyzing images taken across multiple spectral bands is a core component of remote sensing. RGB (red, green, blue) imaging, which creates a true-color representation akin to human vision, is the simplest form [1].
In order to obtain in-depth information, researchers use multispectral and hyperspectral imaging methods. Multispectral imaging gathers data from a few discrete spectral bands, usually extending from visible to near-infrared wavelengths. This allows experts to assess factors like vegetation condition and moisture content [2]. On the other hand, hyperspectral imaging acquires data across hundreds of closely spaced, narrow spectral bands, yielding a detailed spectral signature for each pixel. This extensive dataset enables sophisticated analyses such as categorization of surface materials, contributing to a deeper comprehension of Earth's ecosystems and the consequences of human interaction with the environment [3].

2.1.1 Previous work in Field Report by NIVA

This project is one of the Norwegian Environment Agency's pilot projects for remote sensing, mapping, and environmental monitoring. The aim is to explore the potential for developing more efficient methods for mapping marine habitats compared to traditional and manual monitoring techniques, using high-resolution multispectral drone and satellite imagery. The project is divided into four activities, with fieldwork and post-processing of drone imagery carried out in the fall of 2022, while image analysis, validation, and upscaling are planned for 2023. The Field Report suggests that the work was successful, covering larger areas despite challenges. Drone flights at 120 meters balanced coverage and resolution. Weather analysis highlights the importance of flexibility in future drone campaigns in similar areas. While chosen areas for mapping kelp forests were suitable, the southern area's sparse islands and deep water made orthomosaic creation difficult. Future study area selection should involve expert consultation, and the collected data provides a strong foundation for further marine vegetation analysis and habitat mapping [4].

2.1.2 Hyperspectral imaging

Hyperspectral images (HSI) capture a continuous range of spectral bands from the light spectrum, rather than just the three primary RGB colors. Spectral resolution is defined by the bandwidth that a sensor records in a single image band. Multispectral images (MSI), like those from

Landsat, have a narrow bandwidth of about 50-100 nm, while HSI have a finer resolution of 5-10 nm. HSI is particularly useful for coastal mapping, as it can distinguish between various land cover components. This can be minerals and vegetation, and with this additional information each image pixel transforms into a hypercube data, adding a third dimension of spectral values to the existing two-dimensional spatial image. This can be illustrated in the continuous spectrum in Figure 1. The increased data complexity and volume can lead to challenges in collecting, processing and analysis, including increased computational requirements [5].

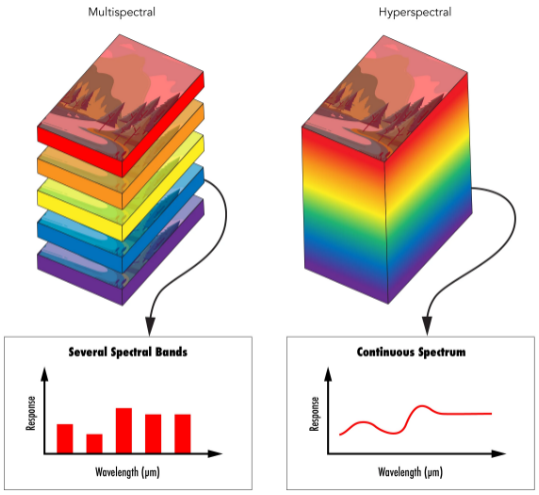


Figure 1: A comparison of multispectral imaging and hyperspectral imaging, where images are acquired across a much broader range of spectral bands [5].

2.2 Image fusion

Image fusion is the field of integrating information from one or more data sources to enhance the understanding of an environment. With applications in remote sensing, computer vision, medical imaging, and robotics, the field of fusion aims to improve decision-making and improve the performance of data analysis. Fusion techniques can combine different modalities, like images with high spatial or high spectral resolutions, with the goal of providing a richer representation of the data. This can enhance analyses and supports various applications in environmental monitoring, urban planning, healthcare and many more [6].

2.2.1 Techniques for image fusion

While it is possible to compare pan-sharpening methods like Wavelet Decomposition, Intensit-Hue-Saturation and Principal Component Analysis quantitatively, different algorithms work well for different images and sensors. One way to get an overview over the different methods that is to be presented is to group them by their function. In table 3, the techniques are categorized into

the three groups "Edge-based", "Component Substitution", and "Other". Edge-based image fusion methods emphasize on preserving edges and detailed features within the images. These methods are often sensitive to noise and misalignment. On the other hand, Component-based (or Component Substitution) methods operate in a transformed domain, merging or substituting components to generate the fused image. Typically, these methods are more efficient at maintaining good spectral quality, but may lack in preserving spatial details. The choice between the two types of methods depends on the specific requirements of the fusion task, as well as the data one is working with [7].

Table 2: Classification of the variety of the image fusion techniques

Method	Edge	Component	Other
Convolutional Neural Networks (CNNs)			X
Wavelet-based methods	X		
Maximum A Posteriori (MAP)			X
Random Forest			X
Intensity-Hue-Saturation (IHS)		X	
Laplacian Pyramid	X	X	
Coupled Nonnegative Matrix Factorization (CNMF)			X

Pan-Sharpening

Numerous studies have explored sharpening algorithms, mainly for MS data but also HS data, to fuse with high-resolution panchromatic images. This is known as pan sharpening, where the goal is to create a high-resolution color image, by rescaling the color image multispectral color image to match the spatial resolution of the panchromatic image. Several pan-sharpening algorithms exists and can be implemented through libraries like `ERDAS` and `ENVI`. The most successful techniques were discovered in a comptetion held in 2006, published in Alparone et al. [8]. Wavelet transform and Laplacian pyramid were considerably the best ones. The Laplacian Pyramid functions by decomposing an image into different frequency bands to preserve details and edges. This creates a pyramid structure with each level providing a different resolution, with the output being the most detailed. As these methods focused on the low-pass components, many of them lead to some degree of spectral distortion [8].

One of the first algorithms for HS and MS data fusion were a Wavelet-based-method (WBM), adapted from pan sharpening. Its performance was significantly influenced by the spectral resampling method. This made it challenging to enhance the spatial resolution from all the hyperspectral bands. The maximum a posteori method (MAP) is a mixing approach to optimize

estimated HS data relative to observed HS and MS data. A mixing approach refers to the combination of different image fusion techniques to enhance the output. MAP can be seen as a mixing approach because it combines two sources of information: the likelihood $P(X|\Theta)$ and the prior $P(\Theta)$, which in the context of image fusion is how well a fused image fits the observed source images and the prior knowledge of the fused image [9]. Another mixing method is spectral mixture analysis, which involves unmixing low-resolution HS data into endmember spectra and abundances, and ultimately fusing abundance maps with high-resolution panchromatic data. This method has limitations, but is also proven to be a viable and effective strategy which will be explored later [10].

The next subsection will take a closer look at existing literature on two specific pan-sharpening algorithms, IHS and Wavelet based methods. *Intensity-Hue-Saturation*
The IHS technique started appearing in literature related to remote sensing in the 1980's and 1990's. As in many other image fusion algorithms, there are three main steps in the process: decomposition, substitution and recomposition. Decomposition involves transforming the color space of the image from red, green and blue to intensity, hue and saturation. The intensity component represents the brightness while the two others represents the color information. In the substitution part, the intensity component of the multispectral image substitutes the panchromatic image. The last step is to recompose the the modified IHS image back to the RGB color space. The technique has gained popularity due to its fast computing capability for image fusion, and being able to merge massive volumes of resampled MS data. However, the fast IHS fusion struggles with efficiency when merging large data sets from satellite images. One of the downsides are color distortions, that apppears after fusioning. Choi suggests a new approach to deal with the trade-off between the spatial and the spectral resolution of the image [11]. Another technique that preserves the spectral characteristics better is using a wavelet based method.

Wavelet Based Methods (WBM)

Even though WBM are better at preserving spectral characteristics, they usually offer less spatial information. Just like IHS fusion, WBM processes consists of decomposition, applying the fusion function and then recomposition. Wavelets are mathematical functions used to decompose signals and data to different frequency components. The advantage is the possibility to analyze and manipulate the details of the images because it can be fused in the wavelet domain. WBM performs the wavelet transformation on each of the images to obtain a series of coefficients at different resolutions. These coefficients can later be combined, averaging and finding maximum values of the desirable parts of each image, capturing details while keeping the spectral resolution [12]. The Morlet wavelet in the time domain is presented in the formula:

$$\Psi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{\Psi}(\omega) e^{i\omega t} d\omega \tag{1}$$

where $\Psi(t)$ is the wavelet function in the time domain, and $\hat{\Psi}(\omega)$ is the Fourier transform of the wavelet function in the frequency domain. The wavelet function is the inverse of its own Fourier transform.

The limitations of the method are often related to reduced quality of the spatial information and complex computations. The latter is specifically relevant when working with large volumes of data, which affects the efficiency working with HSI. For MS data there are several scientific papers indicating potential usefulness for sharpening images [13].

Random Forest

Random Forest (RF) is a machine learning algorithm that constructs numerous decision trees during training, and outputs its final decision based on either the majority voting (classification) or averaging (regression) of the decisions from all trees. In image fusion this can be utilized to independently analyze the input images and integrate features from multiple sources. These features can be color, texture and intensity, and will be the input to training the Random Forest Algorithm. Hänsch and Hellwich [14] takes on a comparative analysis of different multimodal sources such as Multispectral LiDAR, Hyperspectral and RGB data- The applications are mainly within land cover classification, utilizing both RF and deep neural networks. Some of the main findings in the paper illustrate that combining data yields proficient results, and that RF in a few cases outperformed more complex approaches. It is worth highlighting that the LiDAR data processed with RF performed almost as good as many of the multiple data source models [14].

Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are types of deep learning models typically used for image processing and computer vision. CNNs uses convolutional layers, applying learnable filters to the input. These filters enables the network to detect and learn patterns in the input image, like edges and textures. Furthermore, CNNs deals with complexity in a great matter by having the ability to stack multiple layers on top of each other, deepening the network. The advantages the CNN holds towards image processing is that it is designed to preserve and utilize the spatial information from the input, by having the convolutional layers to identify that inputs that are close together in the original image is likely to be related [15].

A further advantage of CNNs in the realm of image processing pertains to their inherent property of translation invariance. This characteristic enables CNNs to identify and learn patterns independent of their spatial location within the data. For image fusion, CNNs can be use-

ful because they are able to output a single image that incorporates useful information from several inputs. One approach is to transform the input images to a common feature space, e.g. training a CNN to extract features from a high-resolution RGB and a lower-resolution MSI with the goal of obtaining a higher-resolution HSI. Within medical fields, CNNs has proven to be successful in fusioning data, and the potential is even bigger [16]. The results indicates that several methods exhibit state-of-the-art performance in fusioning images from Computed Tomography (CT), Magnetic Resonance (MR) and Single-photon emission computed tomography (SPECT). Also within the field of remote sensing CNNs has proven its usefulness. In "this paper", a novel fusion framework for remote sensing is proposed. The method extracts features from PAN and MS images, and concludes with being superior to both classical fusion methods and other deep learning-based methods [1].

In Zhang et al. [17], both GANs and CNNs are used to perform sharpening tasks on a set of hyperspectral images from 31 bands, fused with either MSI or panchromatic images of higher resolution. As CNN requires supervision, the methods are trained on simulated data. This involves implementing spectral and spatial degradation of the original HSI, to learn the model hyperspectral sharpening. Despite challenges, the article claims to be successful at hyperspectral sharpening, and makes the usecase for remote sensing monitoring as a pre-processing part for such tasks [17].

However, there a couple of disadvantages of the CNN approaches, depending on what the application is. CNN is a supervised machine learning technique and typically requires large labelled datasets to perform well. Training a deep CNN is also computationally expensive, especially when receiving high-dimensional data as input, such as HSI with many bands. Even though CNNs often can model complex patterns and relationships, understanding how this works and visualizing the spatial errors can be difficult [18].

2.2.2 State-of-the-art literature

Coupled Nonnegative Matrix Factorization (CNMF)
In recent years many hyperspectral unmixing techniques has evolved as an option for image fusion. The process involves two main steps: extracting endmember spectra and then computing their abundance maps. Some of the commonly used approaches are convex-geometry-based, which assumes at least one pure pixel for each endmember. This means that the algorithm for every distinct material present in the image, there exists at least one pixel that contains only that material. Nonnegative matrix factorization (NMF) has recently gained popularity due to its efficiency. This process involves factoring a nonnegative hyperspectral data matrix into two

nonnegative matrices, identifying endmember spectra while at the same time estimating their abundances without the need of pure pixels [6].

This paper covers the use of CNMF as an algorithm for fusion of HS and MS data. The method is unsupervised and considers sensor obervation models. It generates high-spatial-resolution HS data by integreating the endmember matrix from the HS with the high-spatial-resolution abundance matrix from the MS data. Through three simulations the algorithm's effectiveness is tested on synthetic data from actual airborne HS data. The algorithm can be used alongside Japan's next-generation spaceborne radiometer, Hyperspectral Imager Suite (HISUI), which is able to captures both HS and MS data over the same region under the same atmospheric and illumination conditions. This is a powerful tool for image fusion. The paper concludes that the CNMF's strengths include its simplicity and its ability to handle spectrally varying scenes. Its performance is considered comparable or superior to the advanced MAP/SMM algorithm in simulations with diverse image datasets [6]. The method resembles similarities with the method explained in this thesis.

Other reviews

Hyperspectral and Multispectral Data Fusion is a comparative review of recent literature, by Yokoya, Grohnfeldt and Chanussot from 2017. This article compares ten hyperspectral-multispectral (HS-MS) fusion methods for enhancing spatial resolution in HS imagery. Using diverse data sets, the review evaluates fusion techniques, such as component substitution and Bayesian probability, for performance, robustness, and versatility. The study also assesses the impact of spatial resolution enhancement on classification and suggests potential future directions for HS-MS fusion development. The paper can be summarized in four main observations.

The study found that the fusion of Hyperspectral and Multispectral (HS-MS) images significantly improved classification accuracies compared to solely using the HS-MS images independently. This was exemplified in the HySure classification maps which demonstrated enhanced spatial details and superior material discrimination potential. HySure is a total variation regularization, effective in preserving edges while smoothing out noise [19]. The classification performance was largely consistent with the fusion quality assessment results. For instance, HySure yielded excellent classification results, while other methods, like Gram-Schmidt adaptive (GSA), showed higher potential for classification despite not being competitive in the quantitative tests. GSA is an improved component substitution pan-sharpening method. Some of the fused images even surpassed the reference images in the classification tasks. This was primarily due to a denoising effect that were introduced by the HS-MS fusion. Using classification-based valida-

tion does not reflect the absolute quality of fused data because the ground-truth information is typically available only for a small spatial subset of the data [19].

2.3 Self-Organizing Maps

2.3.1 Overview

A Self-organizing Map (SOM) is a type of artificial neural network (ANN) invented by Professor Teuvo Kohonen in the early 1980s. It maps multidimensional data onto lower-dimensional subspaces, often using a rectangular grid of nodes. Each node in the SOM represents a spectra, which could be randomly initialized or chosen by Principal Component Analysis (PCA). The input, which in this case is an image with a spectra of intensities, is then compared to each node's weight to determine the Best Matching Unit (BMU) based on similarity. The BMU and its closest neighbors are then moved closer to the input, updating the grid of the SOM to better represent the input data's distribution. SOMs preserves the geometric relationships between points, only with fewer data points than the original dataset, making it a valuable tool for data visualization and analysis.

2.3.2 Machine Learning, classification and clustering

Machine Learning (ML) is the use of computer algorithms to automatically discover regularities in data, and is commonly divided into two types of learning. SOMs are often used in unsupervised learning, as opposed to supervised learning. Supervised learning is often used for tasks like regression and *classification*. The latter is where samples of the input data are assigned to a predefined set of discrete classes or labels, such as $L = \{l_1, l_2, \ldots, l_k\}$[20]. The objective is to build a model that can process unseen data and assign to one of these predefined classes. The number of classes can either be two, which yields a binary classification problem, or more which gives a multiclass. Some of the most common methods are decision trees, support vector machines (SVM), and neural networks. SOMs can also be used for classification [21].

Confusion matrix

A confusion or error matrix is a common way to visualize the performance of a classification problem as the one mentioned above. The rows of the matrix represent the instances of the predicted classes, while the columns represent the actual labels, often referred to as the ground truth labels. This visualization is a powerful tool because it provides four key metrics in the same figure, displaying the correct prediction on the main diagonal as seen in. The usefulness expands to being able to quantify the error of the classification task in many statistical ways,

such as accuracy, precision, F1-score and recall, which can further enhance the understanding of the model.

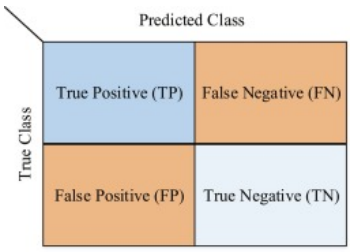


Figure 2: The scheme of a confusion matrix of a binary classification problem [22] .

Clustering however, refers to an unsupervised technique which discovers groups of data points that can be aggregated together because of similarities. As apposed to unsupervised applications they do not require *a-priori* knowledge, also known as the ground truth. Each input vector of data is projected onto a lower-dimensional space, and is often useful in preprocessing steps. It reduces the computational cost while transforming the data, before potentially extracting useful features. Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of n data points where each point $x_i = (x_{1,i}, x_{2,i}, \ldots, x_{d,i})^T$ consists of d features. Then a clustering algorithm is a function f that takes the input data X and a set of hyperparameters λ such that

$$f(X, \lambda) = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \tag{2}$$

where $y_i \in \{1, 2, \ldots, c\}$ is the assigned label to a data point i, c is the number of clusters to separate the data into. There are several widely used clustering methods, and one of the most popular ones is *k-means*, due to its simplicity and computational efficiency.

2.3.3 K-means

The *K-means* algorithm specifies a number of k centroids, an imaginary or real location that represents the center of the cluster. The *means* derives from the averaging of data, which is the process of finding the centroids. Let M be the set of centroids, such that: $M = \{\mu_1, \mu_2, \ldots, \mu_c\}$ where $\mu_i \in \mathbb{R}^d$ such that each centroid is a prototype or mean member of a cluster. The objective is to find M and the cluster assignments $C = \{C_1, C_2, \ldots, C_c\}$ that minimizes the sum of distances from points in the cluster to the center:

$$\min_{M,C} \sum_{j=1}^{c} \sum_{x_i \in C_j} k(x_i, \mu_j), \tag{3}$$

where k is the Euclidean distance [23]. The point here is to allocate every data point to the nearest cluster and at the same time keep the centroids as small as possible [24]. Figure 3 shows how

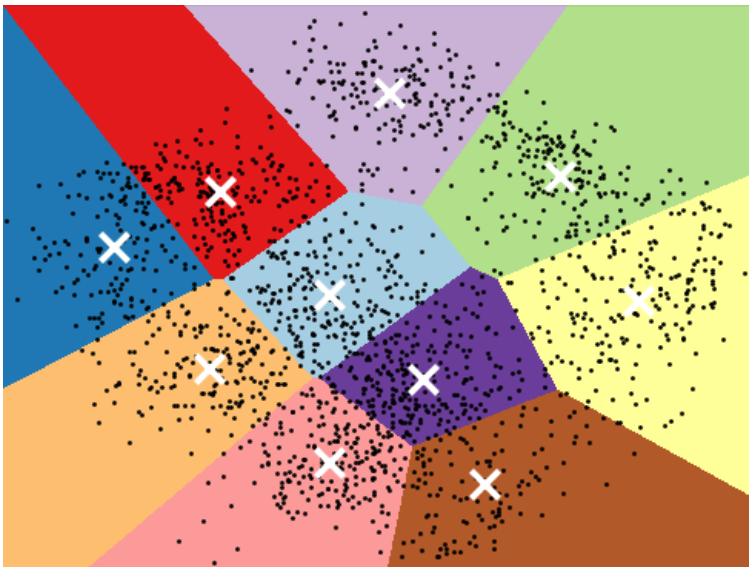


Figure 3: K-means clustering on a PCA-reduced dataset, where the colors are the clusters and the white crosses are the centroids [25].

the K-means algorithm partitions data into clusters based on its inherent patterns and relationship in the data. SOMs shares several similarities with K-means, and these will be explored in the next part introducing the mathematical formulation of the SOM [25].

2.3.4 Weight initialization

The weight initialization of a SOM is significant and influences the learning dynamics and the quality of the final map. In this thesis the three initialization approaches that will be discussed are: normal, random, and principal component analysis (PCA). Normal initialization sets weights using small, normally distributed random values, usually centered around zero. It's a simple approach, but it doesn't consider the specific traits of the input data. This can lead to slower learning or subpar final results, as it may not provide an ideal start for the learning process.

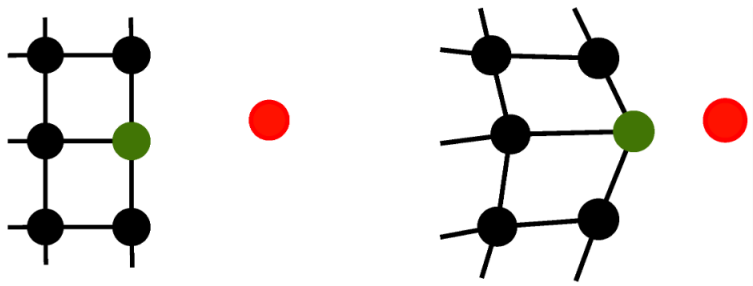
Random initialization sets weights by randomly selecting from the input data. This limits the initial weights within the data's input space, and does often lead to faster convergence compared to normal initialization. However, this method lacks a system as it entirely depends on random data point selection. The PCA initialization method, on the other hand, is a more data-driven approach. This statistical method performs principal component analysis on the input data, and the first two principal components are used to determine the initial weights. This ini-

tialization takes advantage of the underlying linear structure of the data and provides a good approximation of the data's main features. This tends to provide a better starting point for the SOM learning process, often leading to faster convergence and superior map quality. However, it assumes that the primary variance in the data is the most important feature for the map, which might not always be the case. The specific choice often depends on the nature of the input data and the specific requirements of the analysis [26].

2.3.5 Mathematical background of the SOM

In the SOM, each node is represented by a vector z with the same dimension as the input data. The SOM's function can be divided into two parts, the training and the labeling. Although these two can be combined, for simplicity measures these two will be discussed separately. Both parts involve finding the Best Matching Unit (BMU) for an input vector. This is the node closest to the input vector under a given distance metric, typically Euclidean. One can denote the following as: $X = [x_1, x_2, \ldots, x_n]$, $x_i \in \mathbb{R}^d$ as the input data, and $Z = [z_{1,1}, \ldots, z_{m,m}]$, $z_i \in \mathbb{R}^d$ are the $m \times m$ grid of SOM nodes, where k is the distance metric.

A neighborhood function of the BMU can be computed during the training. As the training iterates, nearby nodes are adjusted to match the input vector. In the next step, the input vector is labeled with the lattice coordinates of the BMU as seen in Figure 4.



(**a**) Before training (**b**) After training

Figure 4: Training the SOM with a new input sample (represented in red) primarily impacts the best-matching unit (illustrated in green). Neighboring nodes experience less influence. The similarity in the data space is proportional to the distance between the nodes [27].

Once the SOM is trained, the model evaluates which two points are the closest together with respect to the distance metric k. The points are then mapped to spatially close nodes in the SOM grid. Thus, the topology of the input space is preserved in the grid. This is beneficial for several reasons, such as interpretability, noise reduction and non-linear relationships [28]. The great interpretability of the SOM is due to the fact that the data points in the original input

space remains close in the lower-dimensional representation, such that underlying patterns and structures in the data is still there. For the majority of ANNs this does not exist, and its a powerful tool to enhance the understanding of the underlying dataset [29].

A reduction of the noise can be obtained by focusing on the relationships between data points rather than absolute positions, which ultimately can provide a more robust representation of the data.Another benefit is the ability to capture some of the non-linear relationship, which is quite common in real-world datasets and can be complex to work with. This is due to the competitive learning and the above-mentioned factors [23].

The training process of the SOM is quite similar to gradient descent. They are both interative, woth a learning rate and a criteria for the convergence. Because an objective function for the original SOM is not found, the method differs from gradient descent. $u_i^* = \text{BMU}(x_i)$ denotes the index on the SOM of the BMU for input vector x_i from eq. (4). The update step s for each SOM node z_j is given by:

$$z_{s+1,j} = z_{s,j} + \alpha\beta(t, u_j, u^* i)k(x_i, zs, j) \tag{4}$$

where $\alpha > 0$ is the learning rate for the iteration, $\beta(t, u_j, u_i^*) > 0$ is the neighborhood function, t denotes the epoch of training, and s denotes the iteration within the epoch. In this thesis the neighborhood function is not defined, rather a gaussian is used.

The learning rate α determines how quickly the map adapts to new input data. The neighborhood function $\beta(t, u_j, u_i^*)$ preserves the topology by training nodes that are spatially far from u_i^* with a lower magnitude. The training process also includes an update radius, commonly denoted $\sigma(t)$, which gradually decreases over time. This allows the algorithm to initially capture the overall distribution topology and, as the training progresses, tune the local regions of the distribution for a more accurate representation [28].

2.3.6 Preprocessing for SOMs

Preprocessing is an important part of image processing and fusion, especially when working with high-dimensional data like HSI. Some of the most important reasons why preprocessing is important are normalization, noise reduction, dimensionality reduction (DR) and mitigating alignment and effects in the atmosphere. To enhance the performance of the model (the SOM), spatial preprocessing can be used to smooth the image and remove noise. Normalization can be useful in the context of image fusion because images from different sensors can have varying intensity ranges. In order to get them on a comparable scale, the values might need rescaling.

Another aspect in image fusion is alignment of the images, and preprocessing can help to correct for geometric inaccuracies. In remote sensing, different conditions in the atmosphere and illumination can result in worse and varying image quality, and preprocessing can be used to mitigate the effects of bad detectors, inter-detector miscalibration, and intermittent anomalies [30].

Simplifying or making the data easier to handle makes DR an important step in image processing and even more important when dealing with HSI. To reduce and limit computational cost it can often be useful to reduce the dimensions of the the HS data while still preserving the most important features. There are several ways to do this, one of them being Principal Component Analysis (PCA).

2.3.7 Principal Component Analysis

Principal Component Analysis (PCA) is a statistical procedure extensively used in various fields such as population genetics, microbiome studies, and atmospheric science. Particularly effective with large datasets featuring a high number of dimensions or features, it simplifies the data's complexity by reducing its dimensionality. Through a linear transformation, PCA converts the original, possibly correlated variables into a set of linearly uncorrelated variables, or principal components. Each component accounts for a portion of the total variance, with the first capturing the most and the rest following in descending order. This method enhances data interpretability while preserving as much original information as possible. By primarily employing the first two principal components, it's possible to visually represent multidimensional data in two dimensions, which can be seen in Figure 5, thereby aiding in the identification of data clusters and patterns [31].

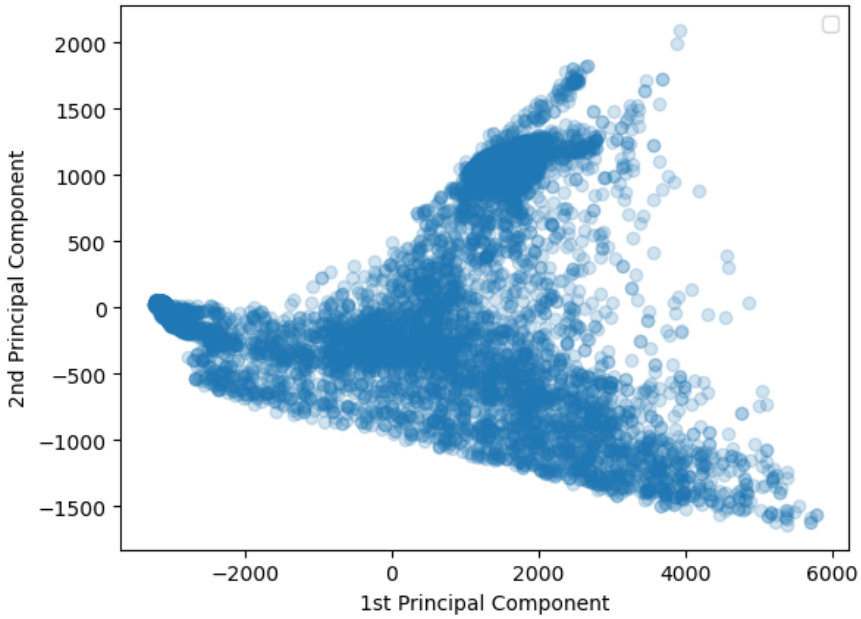


Figure 5: Plot showing PCA of the Samson dataset, with the first two principal components along the x- and y-axis.

Although PCA is fundamentally not designed for future data prediction, it can significantly aid in handling large data sets as the statistical method it is. It is however worth mentioning that opposed to SOMs, PCA seeks primarily linear features of the original data rather than capturing complex nonlinear relationships. In some sense the two techniques do complement each other well when used in conjunction [32]. Rodarmel and Shan [33] presents PCA as a useful and adaptive tool for data analysis, particularly for handling large datasets. By creating new uncorrelated variables that maximize variance, PCA can reduce the dimensionality while minimizing the loss of information. PCA's adaptability is highlighted in its numerous applications for different data types and structures. The article also mentions PCA's direct role in other statistical methods, further showcasing its broad utility [33].

Jolliffe and Cadima [31] emphasizes the usefulness of PCA as a preprocessing technique for classification of HSI. The study finds that the significant information contained in an entire HSI data set can be largely condensed into the first few PCA bands, with later bands containing mostly noise and highly correlated bands. Employing the most significant PCA bands yields classification rates comparable to those obtained using the full hyperspectral data, suggesting an efficient way to reduce computational complexity while maintaining decent accuracy. PCA-induced information may however cause some misclassifications, particularly at feature class borders [31].

Hasan and Abdulazeez [32] reviews PCA's role as the most commonly used approach for dimen-

sionality reduction in big data analysis. It emphasizes PCA's ability to condense high-dimensional data into fewer dimensions, facilitating faster processing and storage. PCA is used as a technique that preserves a significant amount of information and eliminates redundant noise, making data more interpretable. The authors also remind readers of the vast literature on PCA, suggesting that the technique's potential may extend beyond what is covered in the paper [32].

2.4 SOMs for HSI applications

In this section, a variety of applications of Self-Organizing Maps (SOMs) will be presented. The different studies each have relevancy for this thesis, and an overview can be seen in table 3.

Table 3: Applications of Self-Organizing Maps (SOMs) in Remote Sensing

Applications	Year	Author
Multisource Data Fusion with MSOMs	1999	Wan and Fraser [34]
Unsupervised Classification of HSI	2001	Martínez et al. [35]
Pattern recognition in satellite imagery	2003	Richardson et al. [36]
SSL on M-R and MS Remote Sensed Images	2008	Giacco et al. [37]
SSL using fusion of thermal HS- and H-R VIS-data	2014	Michaelsen [38]
Regression Problems on High-Dimensional Data	2018	Riese and Keller [39]
Supervised and SSL for Regression and Classification on HSI	2018	Riese et al. [29]
Clustering HSI On-Board Cubesat	2021	Danielsen et al. [27]

Self-Organizing Maps (SOMs) gained attention in the field of remote sensing in the late 1990's and early 2000's, especially in regards of HSI. The good applicability is due to the promising extraction of patterns from large and complex satellite data sets. The study emphasizes the usefulness of SOMs for different oceanographic data sets, being able to visualize the underlying patterns. Another advantage discovered in the study was the robustness when encountering missing data, as the SOMs work well estimating missing input data compared to other multivariate techniques. One observed difficulty using the SOM is to select the dimensions of the SOM. It requires a tradeoff between capturing relevant patterns, avoiding overfitting, ensuring computational efficiency, and maintaining visual interpretability [36].

Back in 2008, Giacco and Scarpetta investigated the Application of Self Organizing Maps to multi-resolution and multi-spectral remote sensed images. This involved two main steps, training the SOM on one large unlabelled set of data and then learn on one more limited labelled set. The approach is semi-supervised, focusing on cluster detection and land-cover classification. In regards of uncertainty in remote sensing, the study mentions mixed pixels, correlation among training patterns from the same area, definition of classes. The unsupervised part of the paper

concludes with good capability to separate 7 different classes on a two-dimensional output grid [37].

In 2014, Michaelsen [38] were using a large amount of un-labeled data and a smaller amount of labeled data. The HSI was of lower resolution, while the visual data had a higher spatial resolution by a factor of five. The results ended up showing a small gain in the fused data, with a comparably large computational cost. The HS data set contained 84 bands [38].
In 1999 Fraser and Wan looked at Multisource Data Fusion with Multiple Self-Organizing Maps MSOMs. The main area of interest in this research was the use of MSOM for classification and clustering tasks [34]. The paper uses a stack-vector approach, where all sources of spectral, spatial and geographic data are stacked into a single vector. This approach is somewhat similar to the SOM method implemented in this thesis.

Wan and Fraser [34] looks at relevant challenges with hyperspectral data, and concludes best accuracy is obtained with a low number of iterations for the SOM, and an increase in number of neurons, and this is probably due to high dimensionality. When increasing the number of neurons, there is also a need to increase the initial reference distance. The focus is the compromise between topology preservations, accuracy and the update radius (sigma function). The accuracy is measured based on the topology of the confusion matrix [35]. Section 2 of this paper explains well how SOMs work on unlabelled data, so this could be used to explain the theory when going into how the SOM is implemented. It also has a short paragraph about hyperparameters, mainly iterations and size of the SOM. This article presents the idea of using two SOMs to solve regression problems on high-dimensional input data. One SOM is used for unsupervised clustering of the input data, while the second SOM is a supervised approach, where the BMU acts as the target or labelled data. The output SOM adjusts its weights based on the BMUs of the input SOM. The SOM framework is compared to a random forest and a Support Vector Machine(SVM) approach, and is marginally outperforming the two of them [39].

2.4.1 Uncertainty in data fusion

Although there are many advantages with working with simulated data, it inherently introduces certain assumptions which does not necessary hold in the real world. A few of these uncertainties are related to the capturing of the HS data from drones or other unmanned aerial behicles (UAVs). These typically include things related to either the Global Navigation Satellite Systems (GNSS) or the Inertial Measurement Units (IMU), which are used for accurate positioning. These measurements can be influenced by factors such as atmospheric conditions, sensor biases and satellite geometry. Drone movements is also a factor, as spatial distortions can be introduced due to turbulence and wind alternating the drone's yaw, pitch or roll. Vibrations in the drone

platform might also cause errors like blurring and misalignments. Uneven terrain may affect how the images are georeferenced [40]. The Mean Absolute Error (MAE) is a measure used to see how accurate the georeferencing is done. It is calculated by taking the average magnitude of the differences between thue locations and the georeferenced locations. For the hyperspectral image collected by SeaBee used in this thesis, the MAE was found to be 8.5 cm for the image later referenced to as The Remøy Scene [41].

All the above-mentioned concepts can be sources of uncertainty, that simulated datasets often disregard or lack the complexity to model them. This can lead to overly optimistic results if these real-world errors are not quantified and reflected in the results. Using real data instead of simulated can help ensure less assumptions and keep the complexity intact [42]. The SOM method implemented in this thesis will be tested on a real dataset, as well as a simulated one.

3 Method

In this part of the thesis the methodology for processing the data will be described, as well as an overview of the SOM-architecture on a high-level.

3.1 Data Description

In the process of developing and testing the framework of the Self-Organizing Maps (SOM), the methodology was first applied to a smaller and lower dimensional dataset, the Samson scene. This strategy has several benefits.

Firstly, the computational expensiveness and time required for training the SOM are significantly reduced when dealing with a smaller dataset. This process allows for quicker iterations and adjustments to the model parameters and settings, leading to faster and more fine-tuned SOM in a shorter period of time.

Secondly, the lower complexity of a smaller and 2-dimensional representation makes it easier to visually inspect the output and evaluate the performance of the SOM through the spatial error and the nodes of the SOM. Visualizing errors and assessing the quality of the map become less complicated with simpler data.

This approach of starting with the Samson scene, before progressing to the more complex and high-dimensional Remøy scene, helps ensure the robustness of the SOM implementation. The first part will go through the preprocessing of the images, while the second part will be looking at the fusion process for the Samson scene before moving to the Remøy scene.

3.1.1 Samson Scene

Initial testing was carried out on a rather small HSI, the Samson image which is 95x95 pixels with 156 different bands in the range 401 to 889 nm. There are three endmembers in the image, Soil, Tree and Water. One of the HS bands towards the infrared spectra can be seen in Figure 6. The different bands of the Samson dataset displays great variety in intensity, which is something important to consider when working with a selection of the bands. Since the amount of information and bands are so vast, and many of the bands are heavily correlated, one often wishes to select the bands capturing the most important features. Figure 7 displays a comparison between two widely separated bands of the HSI.

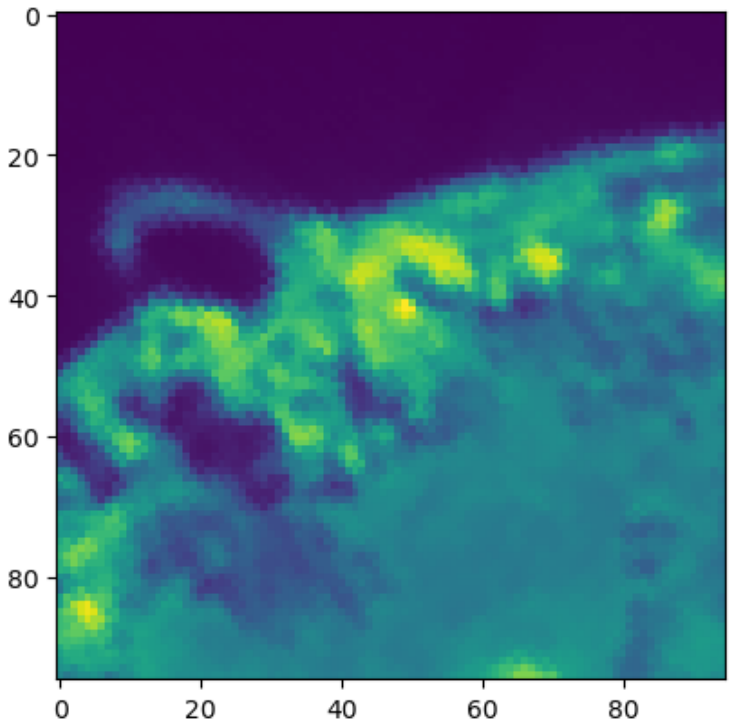


Figure 6: Image of the 120th band of the Samson HSI.

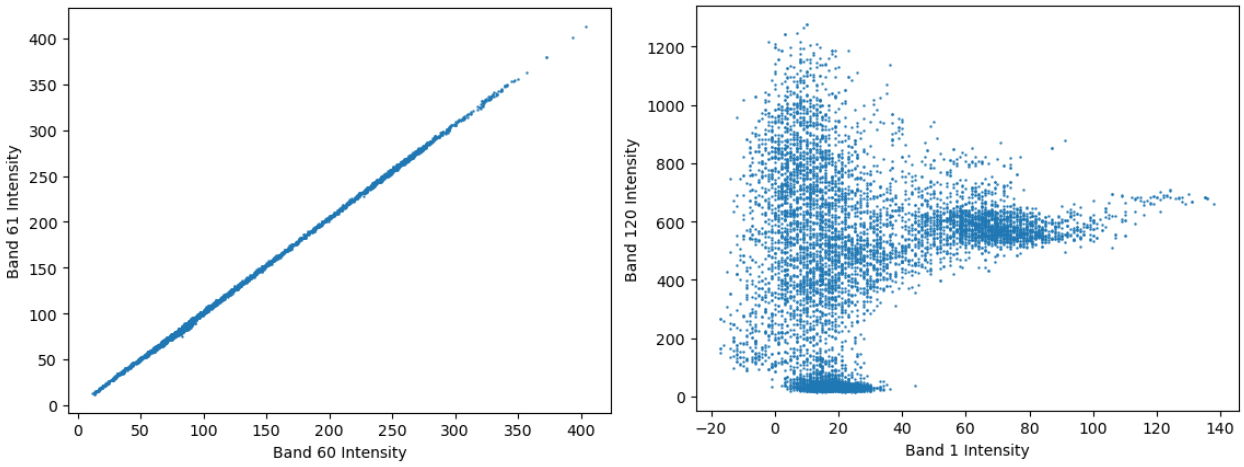


Figure 7: A comparison of the scatter plots of the intensities of two pairs of bands from the Samson image, two neighbouring bands in the middle of the spectra (60, 61) on the left, and one at each end (1, 120) on the right.

3.1.2 Remøy Hyperspectral Scene

After getting familiar with the SOM working with the Samson scene, a larger HSI was introduced. This is an HSI mapping a larger area in Remøy, from now on referred to as the Remøy scene.The

whole area of interest, mapped by SeaBee can be displayed in Figure 8. The yellow square is the area where both the HSI and the RGB used for analysis is taken from.

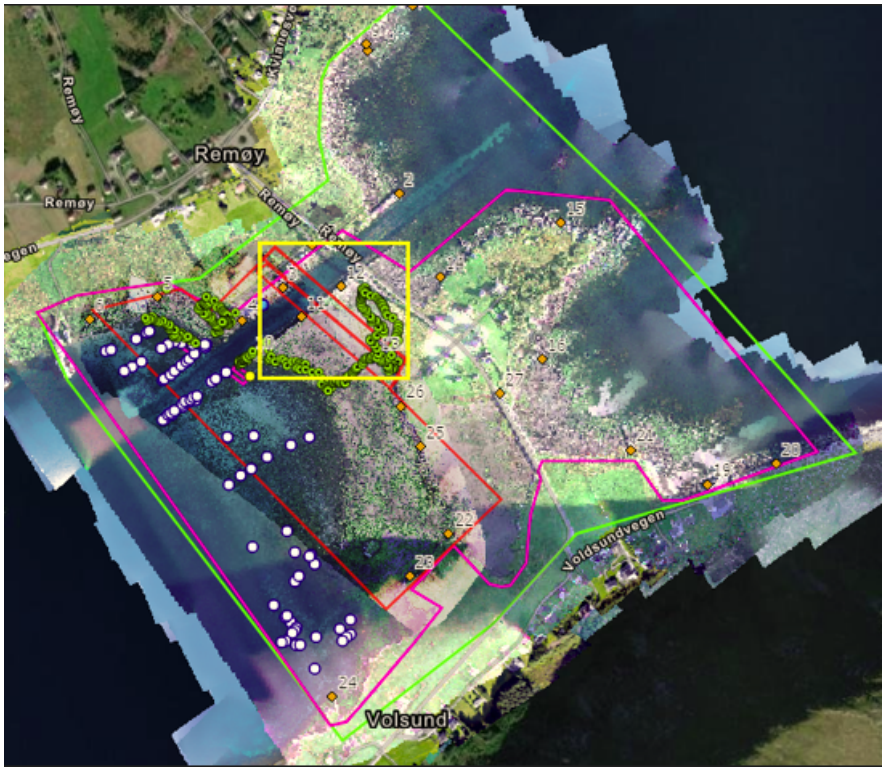


Figure 8: Image of the mapped area of Remøy, provided by the SeaBee project. The red lines encloses the hyperspectral data, the pink lines the multispectral, and the green lines is the RGB. The yellow square is the area of interest.

The HSI is significantly larger than the Samson scene with even more bands, being 2982x2544 pixels and 224 bands. The 120th band is shown in Figure 9. In the image one can observe that there are many dark areas around the actual area of interest. This will be dealt with in the pre-processing part.

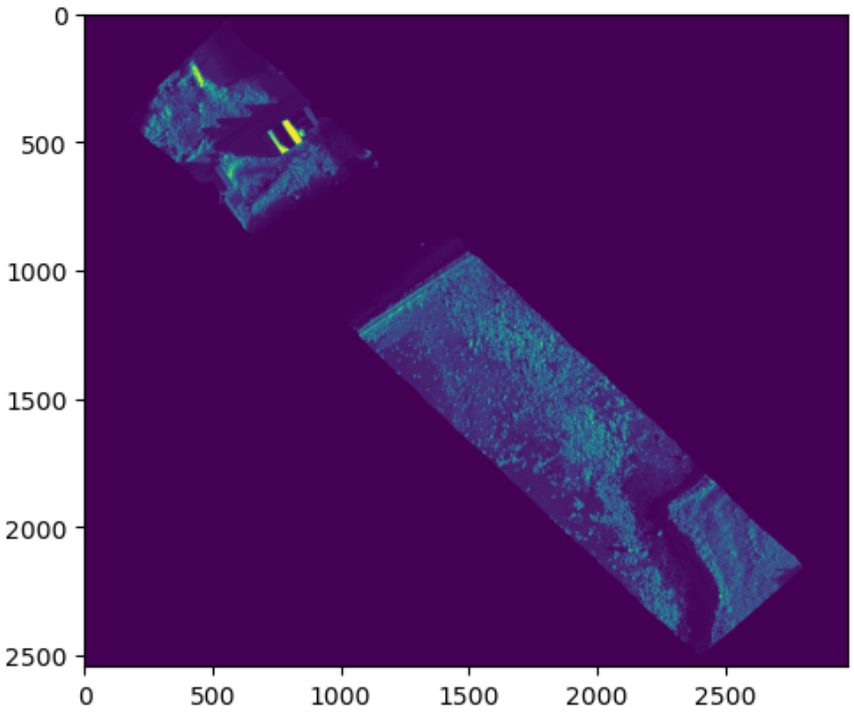


Figure 9: Image of the 120th band of the HSI, displaying the dark areas on each sides.

3.1.3 The high-resolution RGB mosaic

The RGB data used to fuse with the Remøy scene in this thesis is a mosaic image collected from multiple drone flights. This resulted in a large number of individual images that required post-processing through various steps before image analysis could be performed. First, the images were quality-checked through orthorectification and photogrammetry (a process where images are converted to a format suitable for map production by removing geometric distortions related to the drone, sensor, terrain, and other factors). Next, the individual images were stitched together to create a composite image mosaic seen in Figure 10, which was then georeferenced. This was all done in advance by the SeaBee project and is not apart of the thesis.

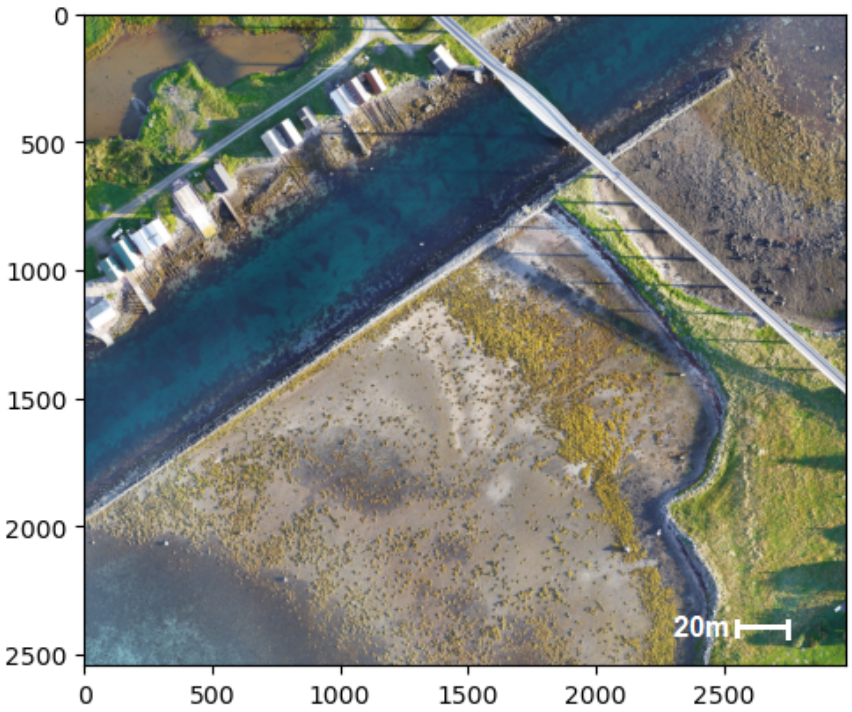


Figure 10: RGB image of the mapped area of Remøy, displaying the same geographical area as the HSI.

3.2 Preprocessing

Different preprocessing steps were used to prepare the data for further analysis. In the first section the preprocessing of the Samson scene will be included, and next section will be about the Remøy scene. It is worth noticing that there are numerous preprocessing steps such as georeferencing and noise reduction performed on the data already, but in this part only the steps done with regards to image fusion is included.

3.2.1 Preprocessing Samson Scene

For the implementation of just the HSI, very few steps was needed working with the Samson scene. When initializing the SOM, the image needs to be a 2D-array and hence the data is flattened. This procedure follows for all data used with the SOM.

Even though its not usually an issue with this scene, experiments with dimensionality reduction (DR) using Principal Component Analysis (PCA) was also carried out as a part of the preprocessing. This mainly involved using `sklearn.decomposition` integrated PCA function, and the results will be presented in the results chapter [25].

3.2.2 Preprocessing Remøy Scene

In the process of preparing the Remøy scene for analysis several steps were included. The first thing to do was to match the orientation of the HSI with the scene by flipping it. As seen in Figure 9, the next step is to remove the dark pixels by adding a mask. This is to prevent the model to focus on irrelevant parts of the image. PCA was also experimented with as the image is quite high-dimensional. The final step is to initialize the SOM with a given lattice size and update radius, and a selected sub-sample of the image to reduce the size and computational time. In this thesis the chosen sample size is every 1000th pixel for the Remøy scene, as this comprised in a large dataset without an overwhelming computational cost. This resulted in 1677 pixels.

3.3 Image Fusion

Fusion was done with both the Samson and the Remøy scene. The biggest difference was that for the Samson scene the RGB bands were simulated using HSI data, while Remøy used real high-resolution RGB data. The fusion process is similar, but some of the pre-processing steps are different. Both approaches stacks the vectors from each input image into one singel vector, as mentioned in [34]. This is done using the `numpy.concatenate` function.

3.3.1 Fusion of the Samson scene

To introduce fusion to the Samson scene, a simulated approach is to be implemented. Simulation is, as described in the theory part, quite common in the world of image fusion for remote sensing applications. For the fusion of the Samson image, we would simulate to have one HSI with low spatial resolution, and one RGB with high spatial resolution. This involves two main steps; spatially downsampling the 156 bands of Samson using a Gaussian filter, and spectrally blur three selected bands, ultimately adding them at the end of the HSI bands. Figure x shows the fusion process (to be added). The selection of the three bands was carried out in order to capture red, blue and green wavelengths. The approach used in this thesis was to select three different intervals within the range of each of these color, and then finding the mean intensity value. The representation with scaled intensity can be seen in Figure 11.

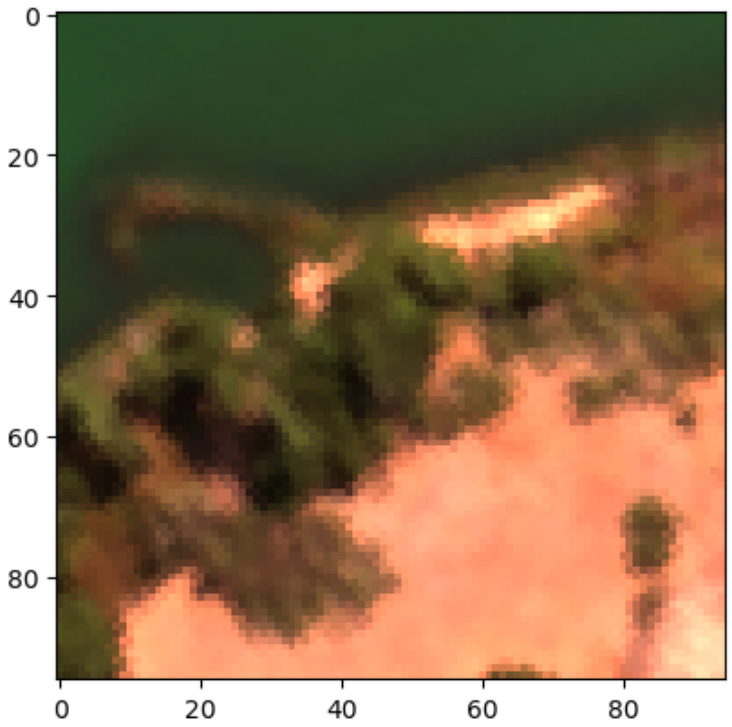


Figure 11: The three simulated RGB bands of the Samson image, normalized intensity to be between 0 and 1.

After this step, the simulated RGB bands are multiplied with a constant to illustrate the intensity difference between RGBs and HSIs. This is to qualitatively examine how it affects the SOM's performance. A various selection of constants were used (0.1-1000) before selecting one. Lastly, the bands were concatenated along with the downsampled HSI. The difference between the intensity of the different bands can be seen in . The stapled lines are the simulated RGB bands.

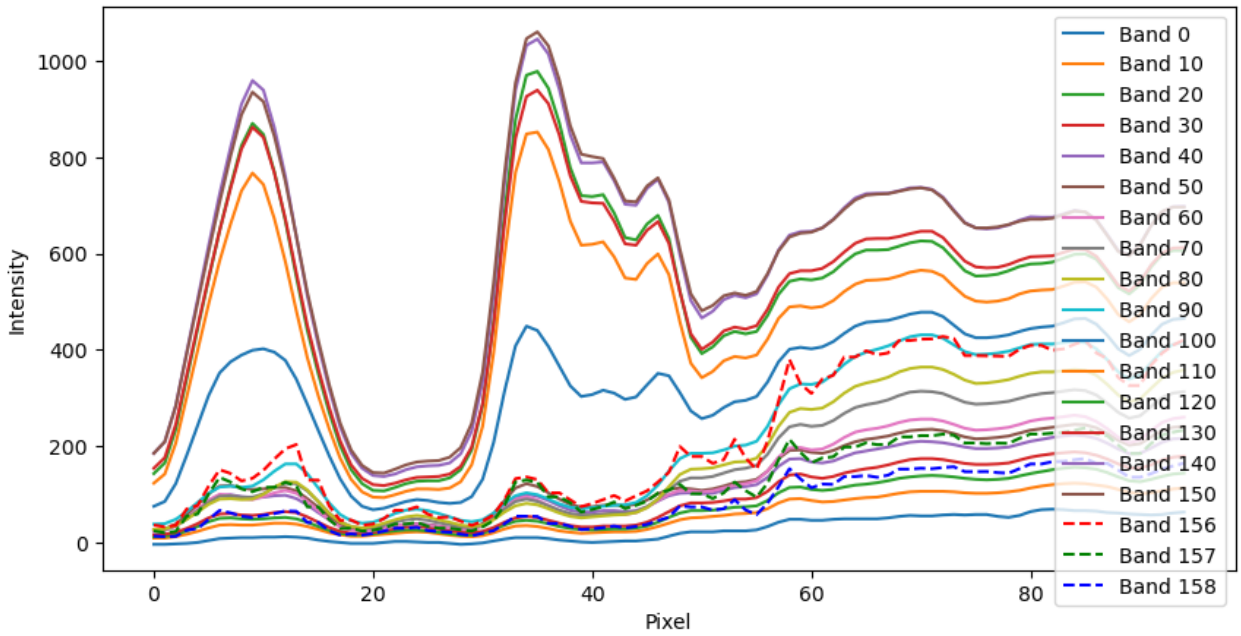


Figure 12: Intensity plot over the 50th row of the Samson image. Every 10th band as well as the last three simulated RGB bands are plotted.

Figure 13 illustrates the methodology of pre-processing Samson scene, from input to fusion.

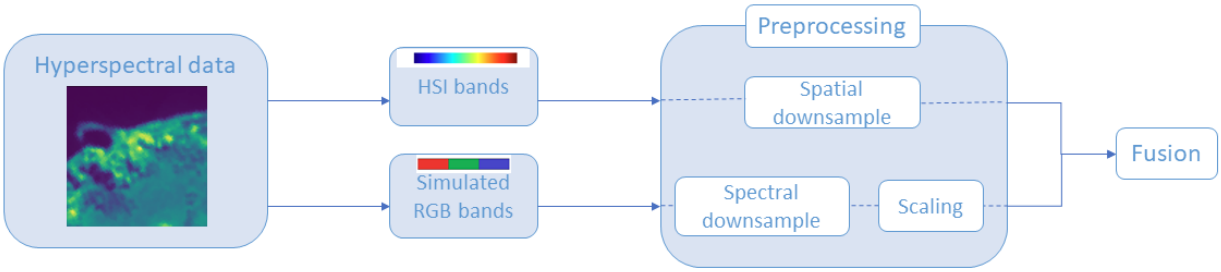


Figure 13: A simplified illustration of how the Samson image is processed, from the hyperspectral image, to separating the bands, downsampling and eventually going into the fusion process, which will be illustrated next.

Figure 14 shows the rest of the methodology, after the image has been processed and how the fusion and training works.

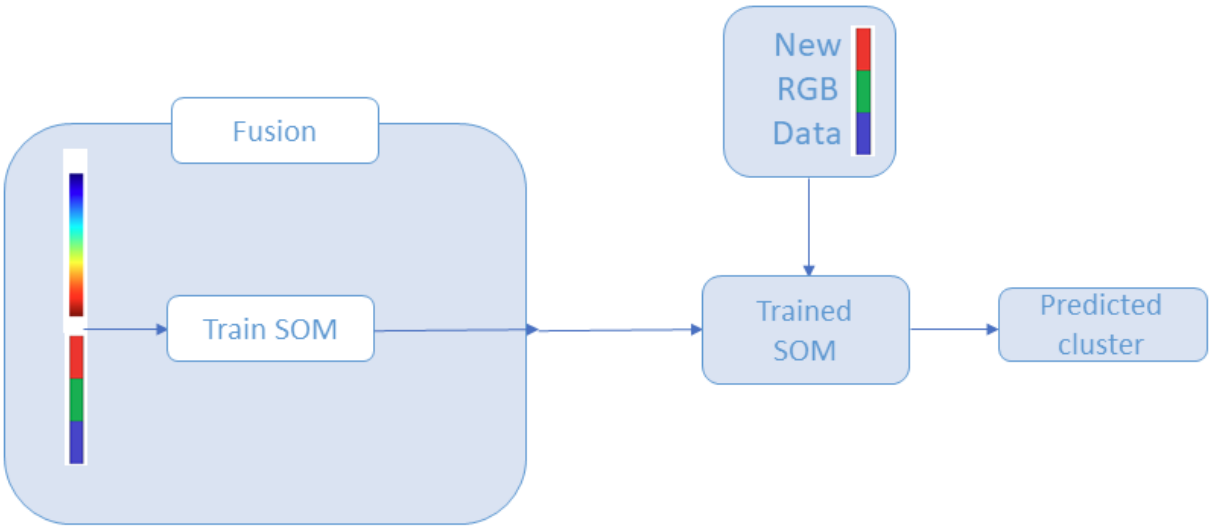


Figure 14: A simplified illustration of how the Samson image is fusioned, and how the process of predicting the spectral information in the areas with only RGB information works.

3.3.2 Fusion of the Remøy scene

After working with the HSIs alone, it was time to introduce the RGBI spanning a larger area of Remøy with higher resolution, seen in Figure 8. The fusion process of the Remøy hyperspectral image involved three main steps:

1. Aligning the pixels of the two images.
2. Concatenating the pixels (stack into a single vector).
3. Training the SOM with the stacked vector.

Firstly, the `osgeo` and the `pyproj` packages were used to get the latitudes and longitudes of the HSI, before using `georaster` to select the right section in the RGBI. These 3 RGB bands were then concatinated using the `SingleBandRaster` function. Since the size and resolution is different for the two images, downsampling of the RGBI was needed in order to match the bands of the HSI. This included reshaping and adding a gaussian blur to the RGB, to prevent aliasing artifacts. The final product can be seen in Figure 10. Additionally, HSI pixels usually have a higher intensity than RGB pixels. In this case, the maximum pixel intensity was 65535 in the HSI, in contrast to only 255 in the RGB image. To ensure that the SOM model appropriately considers the contributions of both RGB and HSI pixels, a scaling process was implemented. A scaling factor was chosen based on an estimated ratio between the two images' intensities.

This scaling factor was adjusted to maintain a balance, ensuring that both types of pixels had a reasonable influence on the final image. This involved visually examining the output to con-

firm that the end result accurately reflected the input data from both image types.

3.4 Configurating the SOM

3.4.1 Lattice size

Secondly, there is the lattice size, which is the distance metric between each node in the SOM. According to theory, an increased lattice size should provide more granularity and a better representation of the input data. Though, this is a trade-off as the computational complexity and training time increases.

3.4.2 Update radius

The update radius is a parameter that determines the extent of influence the Best Matching Unit (BMU) has on its neighboring neurons during the process where the weights are updated. The BMU is commonly known as the neuron with the smallest distance to the current input vector, informally the most similar one, as mentioned in the Theory. The ideal UR is dependent on the complexity of the input data, and can be selected manually.

3.5 Analysis of SOMs

Lastly, the SOMs were analyzed through different metrics and tools. The most common metrics used in the thesis is to be presented in this section.

3.5.1 Spatial error

One useful metric is the measure of how well the SOM preserves the topology of the input image. This can be displayed in a spatial error plot, where the accuracy of the relative relationships between data points in the input space are represented in the SOM's output space. With a low spatial error, the nodes in the SOM is effectively preserving the original image's topology and geometric properties. This can be seen as bright areas, while dark areas represent large distances between the original data and the SOM. This provides a insightful visualization of the model performance.

3.5.2 The relative quantization error

The evaluation metrics in the thesis are mostly the per-pixel relative quantization, relative to the magnitude of the intensity in the input image, which can be calculated using eq. (5).

$$q_i = \frac{\|\mathbf{x}_i - \mathbf{B}(\mathbf{x}_i)\|^2}{\|\mathbf{x}_i\|^2}, \tag{5}$$

where **B** is the function that maps each spectrum to its BMU on the SOM, and q_i is the relative quantization error for a single pixel [27]. The relative error, a proportion of the true value, can help better represent uncertainty in scenarios where values vary greatly in magnitude, as it adjusts for scale. In contrast, Root Mean Squared Error (RMSE) gives higher weight to large errors and can thus be more suitable in other cases. The logarithmic error is another used metric, calculated using eq. (6).

$$\text{The logarithmic error} = \frac{\log(\text{SOM.errors})}{\log(10)} \tag{6}$$

3.5.3 U-matrix

The Unified Distance Matrix, commonly referred to as the U-matrix, is a representation that captures the distance between the neurons in the SOM grid. The purpose is to visualize the topology of high-dimensional input data projected onto the two-dimensional SOM. By calculating the average distance between each neuron and its neighbours in the grid, the dissimilarities between the neurons' weight vectors are presented. Essentially, the clustering and the separation of the input data is visualized in these plots and can help illustrate the borders between clustered neurons. The mathematical definition can be seen in

$$|N(i)| \sum_{j \in N(i)} ||w_i - w_j|| \tag{7}$$

Here, $||w_i - w_j||$ is the high-dimensional distance (usually Euclidean) between the weight vectors of neuron i and its neighbor j, and $|N(i)|$ is the number of neighboring neurons for neuron i.

4 Results

In this section of the report, the results from three different scenes are presented. First is the initial testing analyzing the Samson HSI, then next is the HSI from Remøy, and lastly the fused RGB mosaic and HSI are evaluated. Most of the results presented in this section use the relative quantization error or logarithmic error.

4.1 The Samson scene

Several experiments were carried out training with the Samson HSI. The most important ones to emphasize the different characteristics of the SOM are presented in this segment.

4.1.1 Weight initialization

In the Self-Organizing maps framework, there are several ways to initialize the weights of the SOM. The convergence of the error for three configuration,s `normal`, `random` and `PCA`, were tested and the results can be compared in Figure 15. The SOMs are trained for 15 iterations over all the data, defined on the x-axis as the "Epoch".

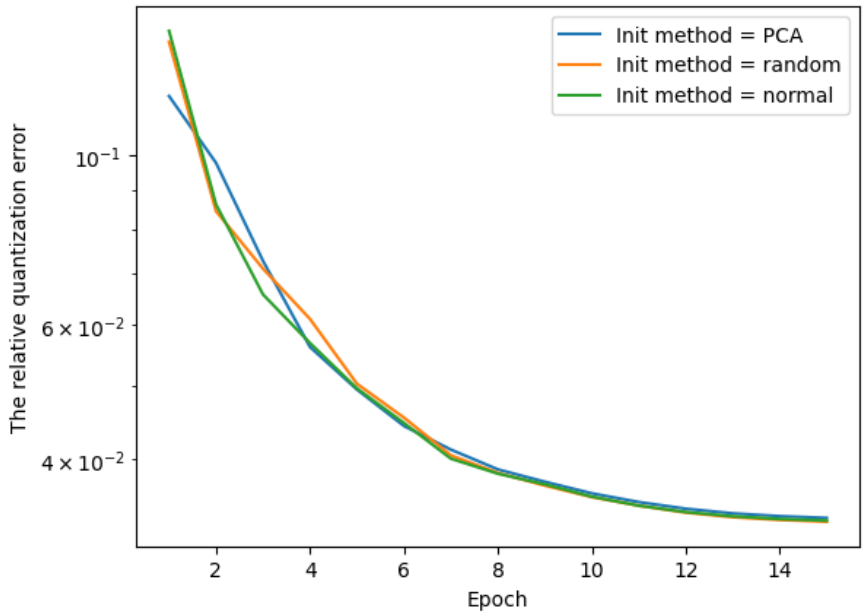


Figure 15: Plot showing 15 epochs of training with the Samson image, lattice size 16, update radius 4 evaluated over all bands.

As PCA yields the fastest convergence and lowest error, this initialization will be used for the rest of the testing.

4.1.2 Lattice size

The different lattice sizes of the SOM was compared against the training time. Figure 16 shows the evaluation for 4 common lattice sizes, plotted against the training time and the error for 10 epochs.

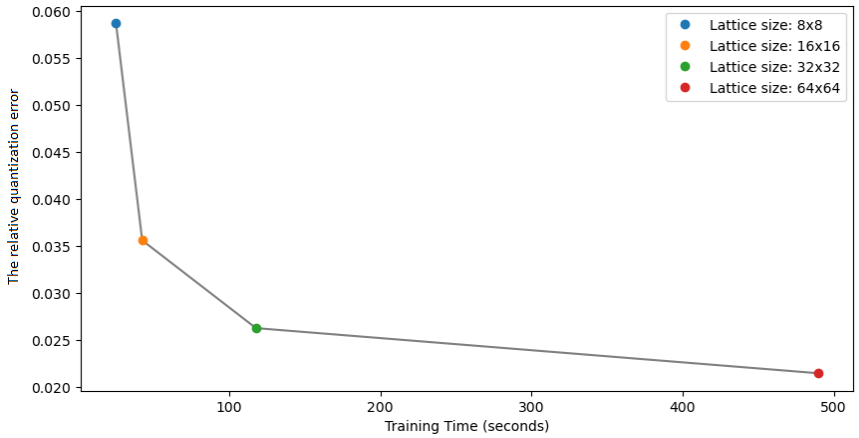


Figure 16: The relative quantization error versus the training time for different lattice sizes of the SOM trained for 10 iterations.

4.1.3 The Unified Distance Matrix

The Unified Distance Matrix was compared for different lattice sizes of the SOM. The comparison of them can be seen in Figure 17.

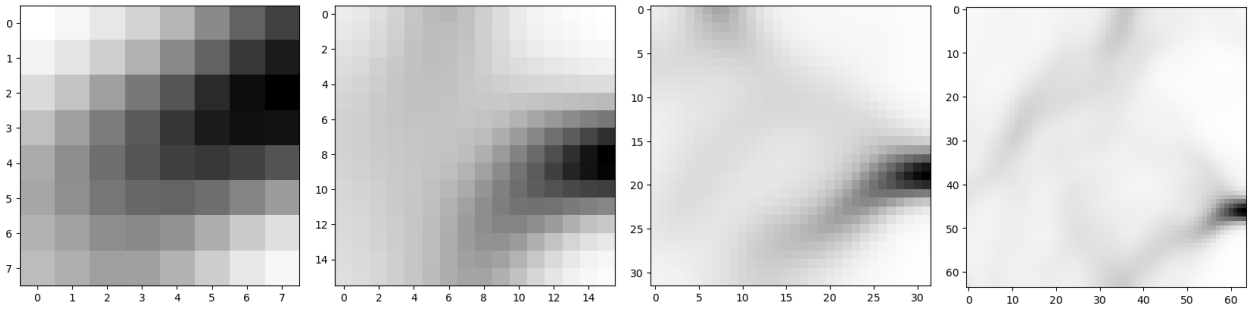


Figure 17: Plot showing the U-matrix for the different lattice sizes of the SOM, with an update radius of 4 trained for 10 epochs. Leftmost image is lattice size 8, rightmost is 64.

4.1.4 Epochs and convergence

The advantages with training the model for more iterations are many, among them improved accuracy and better convergence. The weights of the SOM are adjusted for every iteration, making it learn patterns in the data better and improving the performance. Ultimately this will lead

to convergence towards a state were further training won't improve the model significantly. In Figure 18 the error converges after around 15 epochs.

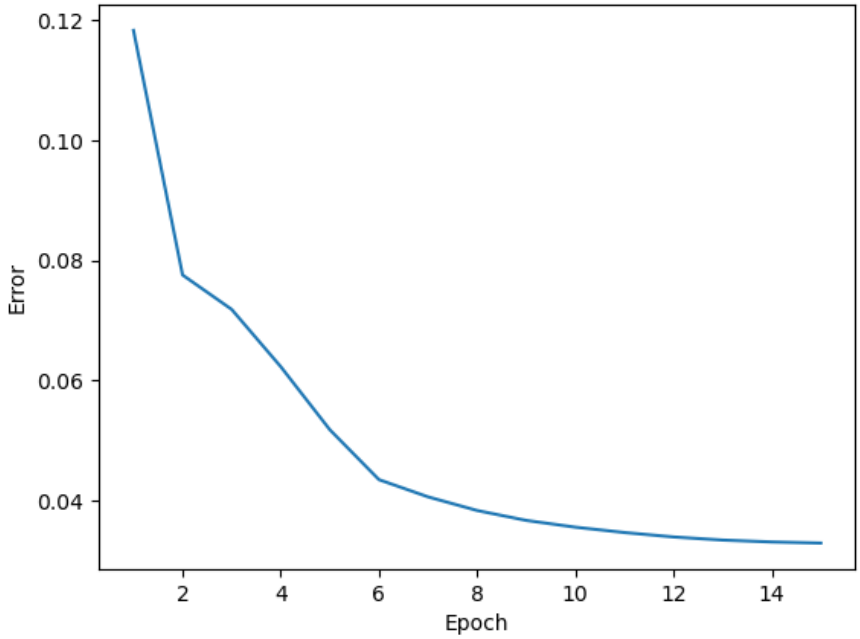


Figure 18: Plot showing 15 epochs of training with the Samson image, lattice size 16, update radius 4, evaluated over all bands.

4.1.5 Update radius

In this section the effect the update radius (UR) has on the error was investigated. Figure 19 illustrates the relationship between the error and the update radius of a SOM with L=16 evaluated on all bands. Note that the difference is larger on the y-axis in the plot on the left side, than on the right after 15 iterations.

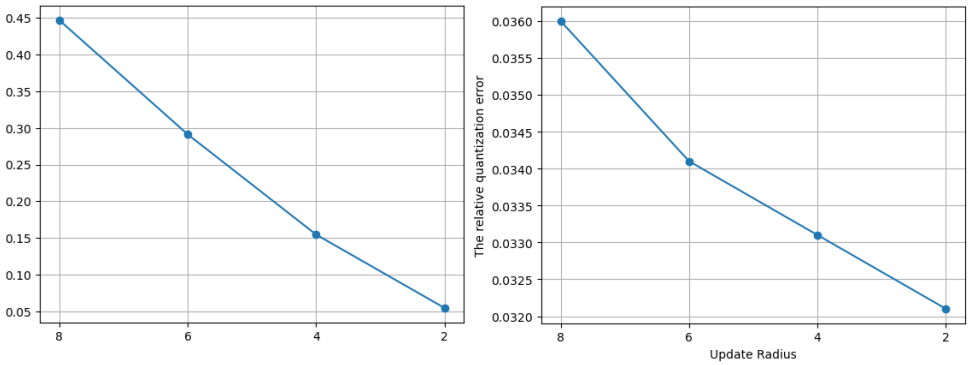


Figure 19: The relationship between the relative quantization error and the update radius of the SOM. The plot on the left is for one iteration, while the plot on the right is after 15 epochs.

Another comparison can be shown plotting the spatial error for each update radius of the SOM. This comparison can be seen in Figure 19, revealing the logarithmic error.

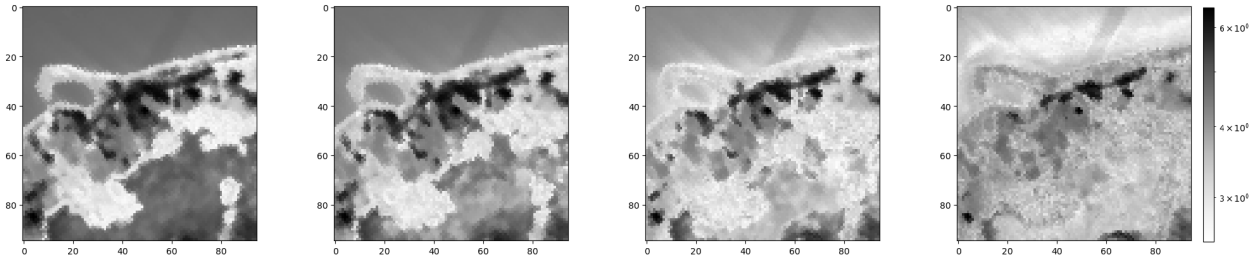


Figure 20: The logarithmic spatial error for the different configuration. From left to right the update radii are 8, 6, 4 and 2.

To incorporate multiple of the introduced metrics, one can observe the progression of the error during training. This process begins with a larger UR and subsequently reduces it across further iterations. This is in order to learn the rough distribution topology first, before finetuning local areas of distribution later. The spatial errors relative to the magnitude of the input for the five epochs can be seen in Figure 21.

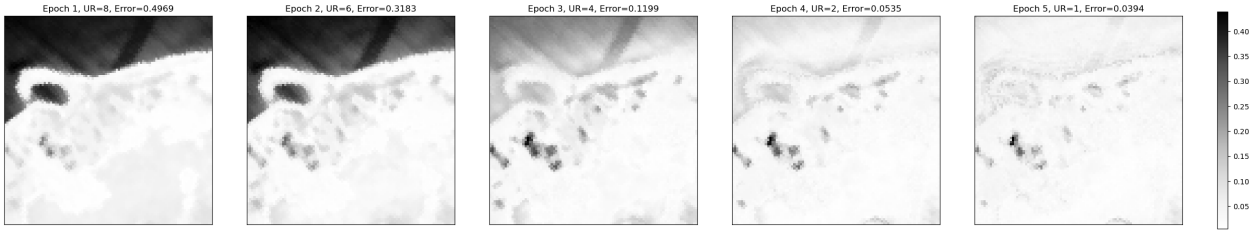


Figure 21: The spatial errors relative to the magnitude for the five iterations with decreasing update radius for the 16x16x156 SOM.

4.1.6 Dimensionality reduction using PCA

Using PCA to reduce the dimensionality was done, and ended up reducing the number of bands from 156 to 12, while maintaining 0.9999 of the variance. This can be seen in the Scree plot in Figure 22.

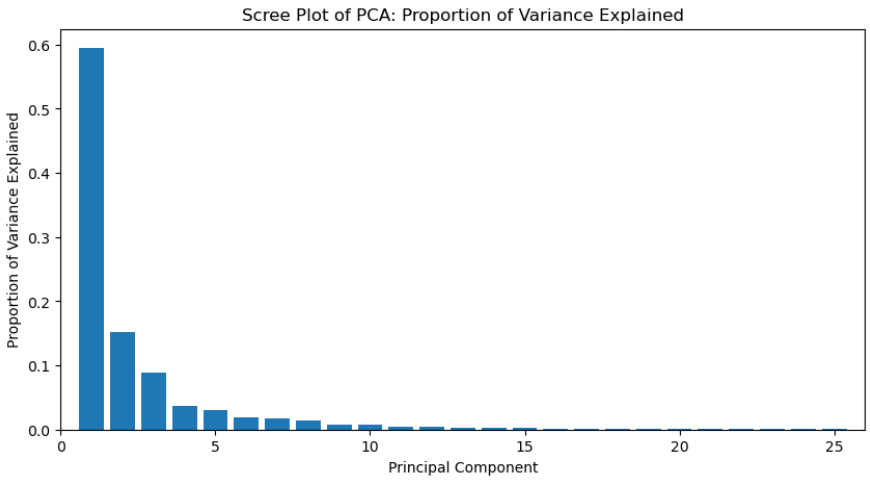


Figure 22: Scree plot showing the eigenvalues of the components derived from principal component analysis. Each pole in the plot represents an individual component, plotted in descending order of their eigenvalues.

The spatial error in the original and the dimensionality reduced Samson can be seen in Figure 23.

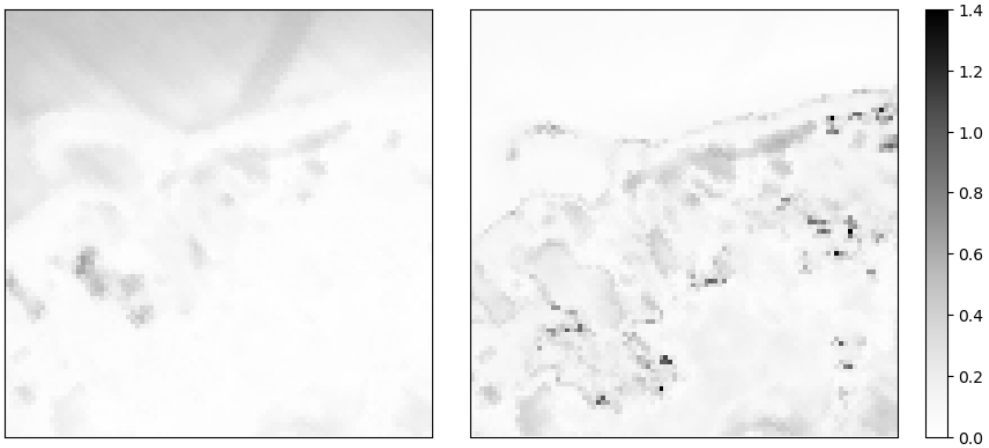


Figure 23: A comparison between the spatial error in the original and the PCA DR version of the Samson.

4.2 The fused Samson scene

4.2.1 Lattice size and epochs

Tests with different lattice sizes were carried out with the same amount of iterations over all the training data, 15. The plot shows how the error decreases for each epoch, with each line representing a SOM with a given lattice size, Figure 24.

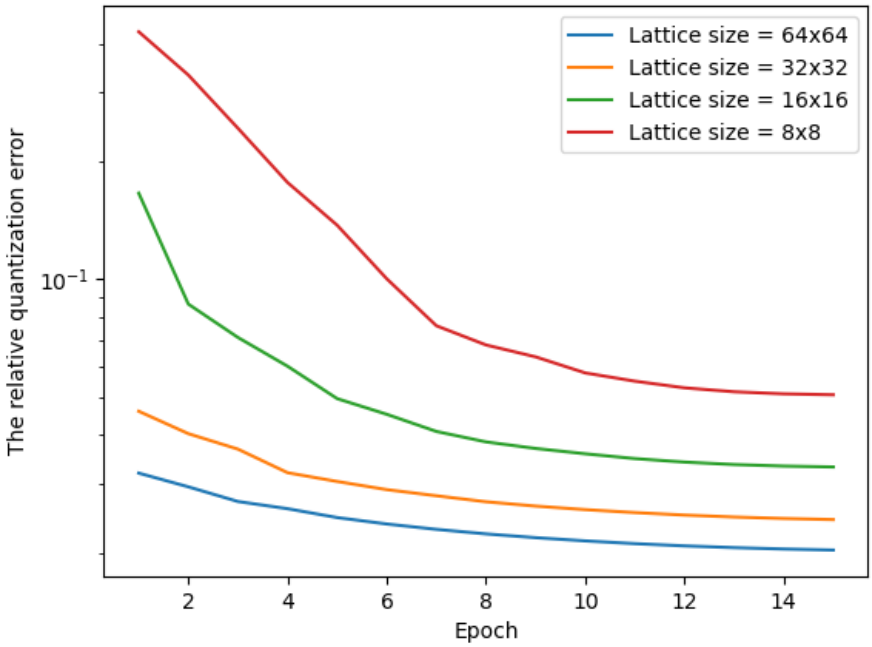


Figure 24: Plot showing 15 epochs of training with the Samson image, SOM with lattice sizes 8, 16, 32, 64, update radius 4 evaluated over all bands. The y-axis is a logarithmic scale.

4.2.2 Scaling the RGB bands

In this the results from the fused Samson scene will be presented. Some of the same metrics from the last section will be used again, as well as a few new ones, like classification. Firstly, the different configurations of scaling the RGB image will be presented. The four different constants that were applied to the RGB bands were 0.1, 1, 10 and 100. A comparison of the spatial error when evaluating all bands can be seen in Figure 25.

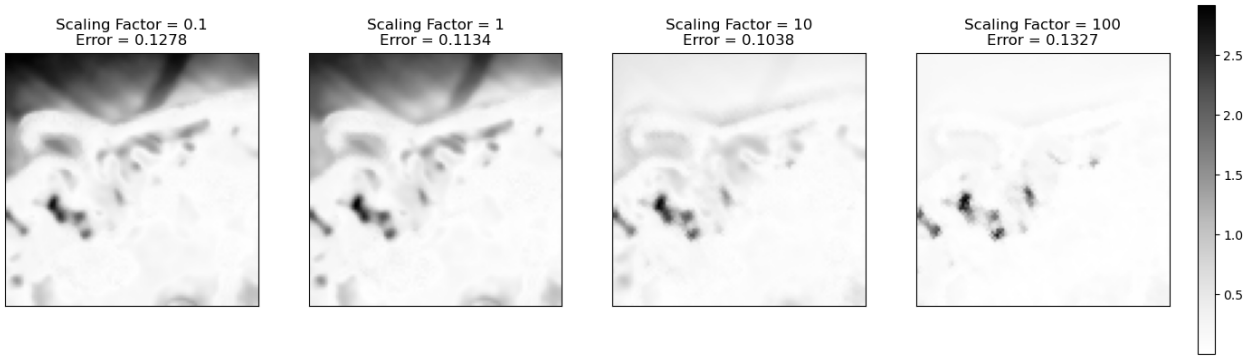


Figure 25: The spatial errors relative to the magnitude for the for different scaling factors, on the 16x16x12 SOM, with an update radius of 4.

For different selections of bands the error changes significantly. In Figure 26, the evaluation is conducted for the simulated RGB bands. The lowest error, 0.0036 is obtained with a scale of

100.

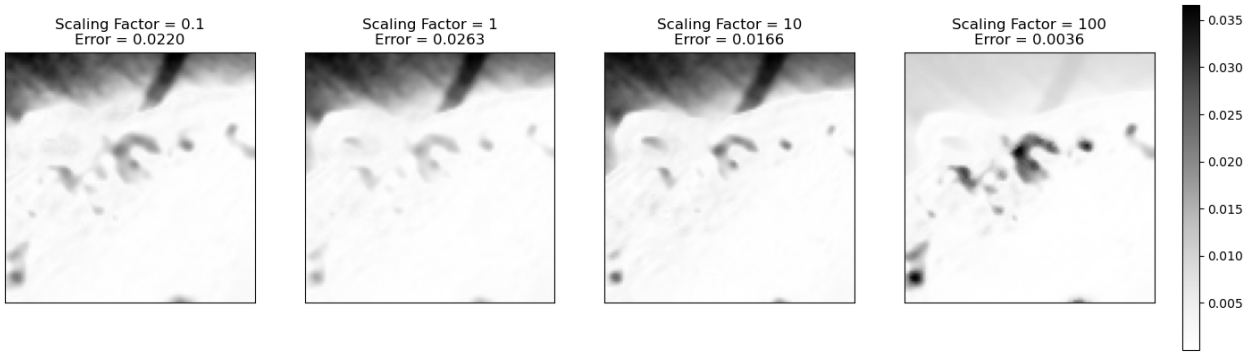


Figure 26: The spatial errors relative to the magnitude for the different scaling factors, evaluating on the three simulated bands. All tests on the 16x16x12 SOM, with an update radius of 4.

When evaluating three arbitrary bands across the image, 60th, 100th and 140th, the lowest error, 0.0030, is also obtained scaling the RGB bands with 100.

The U-matrix can also be plotted for the different scaling factors of the SOM. The four different configuration as well as their error for the 30th, 60th and 90th band can be seen in Figure 27.

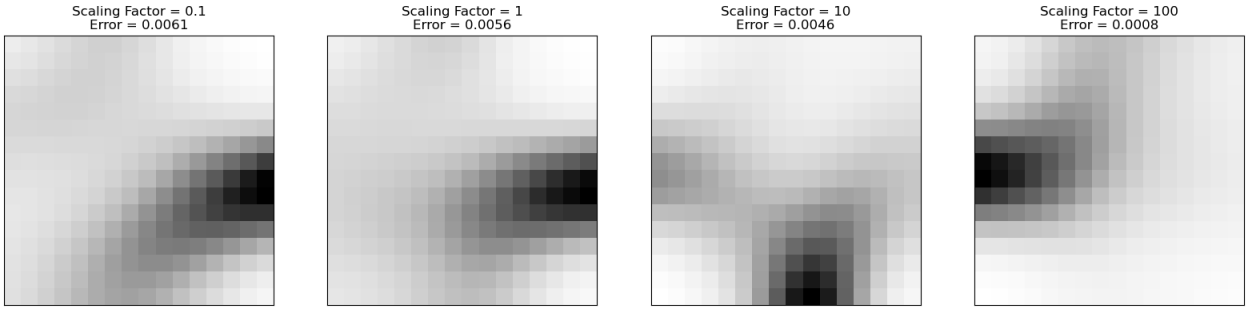


Figure 27: The U-matrix for the different scaling factors, with the relative quantization error plotted on a grey scaled image.

The clustering of the pixels in the Samson scene is shown in Figure 28, with a) being column number and b) the row number. c) shows the relative quantization error.

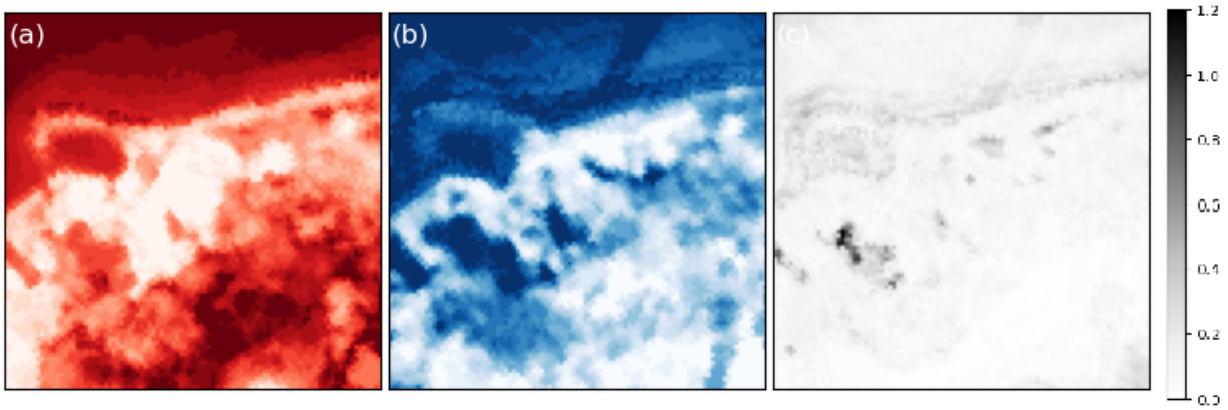


Figure 28: The Samson scene classified according to a 16x16 SOM. (a) is the column and (b) is the row of the BMU index. (c) is the relative quantization error.

4.2.3 Classification

Classification was done using the fused Samson image. The image contains three labels: soil, trees and water, which can be illustrated in Figure 29. The yellow part represents the soil class, the green trees, and the dark blue represents the water.

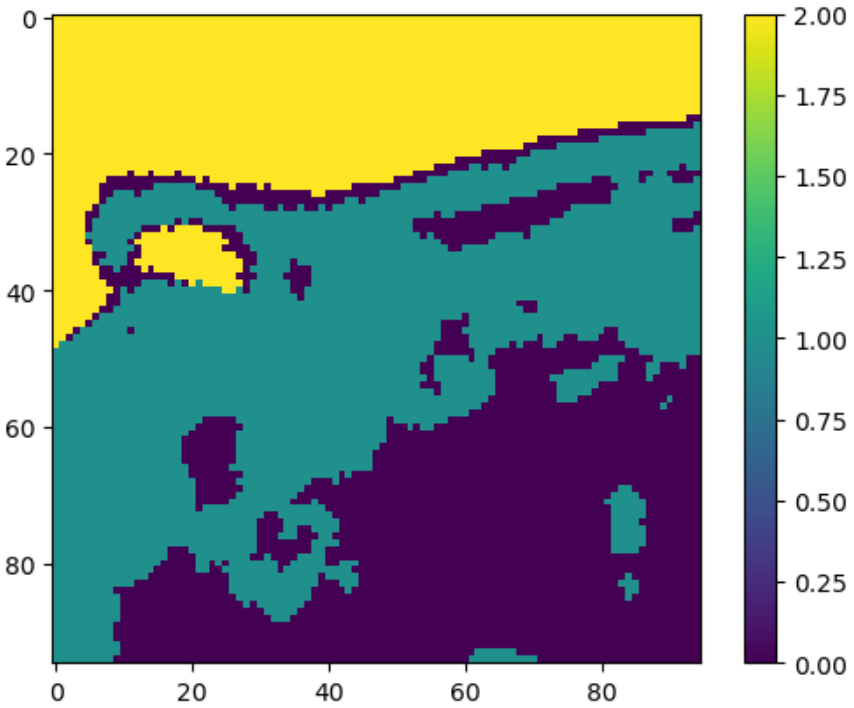


Figure 29: The ground truth labels displayed in the Samson scene.

The distribution of the three classes in the SOM trained on the fused Samson scene can be seen in Figure 30.

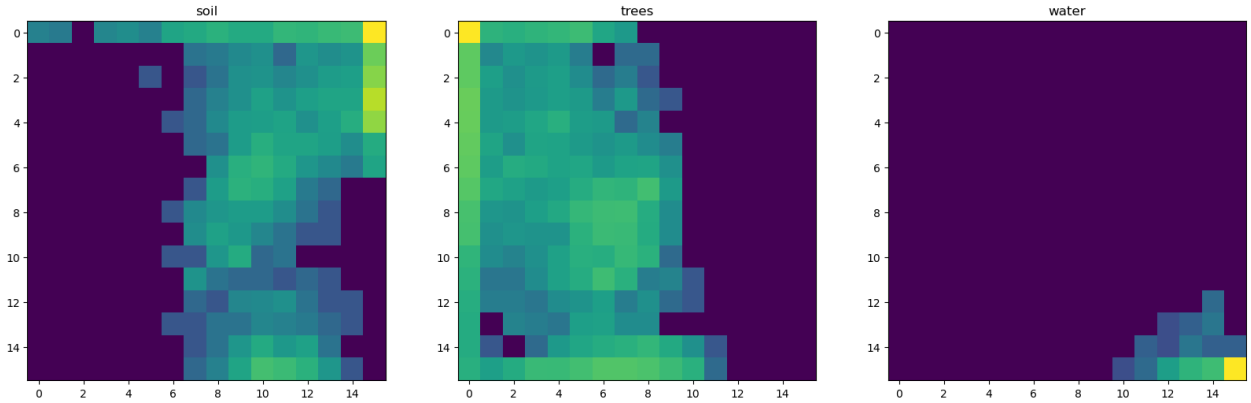


Figure 30: The visualized relationships between the different classes, where the number of pixels of a given class are grouped into a given node in the SOM.

The results of the classification can be compared with the true labels in a confusion matrix. This matrix displays the false and the true labels for the three classes, and a comparison between the original Samson and the fused Samson with 2 principal components can be seen in Figure 31.

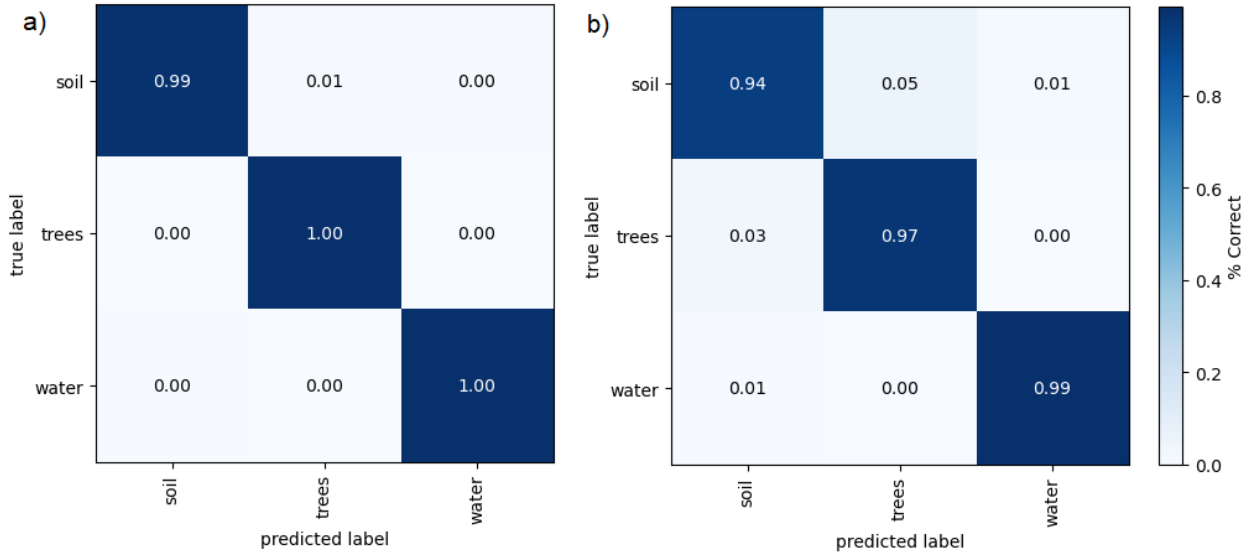


Figure 31: The confusion matrices for the a) original Samson trained with a lattice size of 64, and b) the fused Samson scene trained with a lattice size of 16, dimensions reduced with PCA to two components.

4.3 The Remøy scene

In this section, a visualization of how the SOM represents the RGB Remøy image is presented. This is a way to see how the nodes captures the features (colors) and the importance of them

(number of nodes in that color). Figure 32 shows the Remøy RGB (left) and the 32x32 SOM representing the nodes of the input image (right).

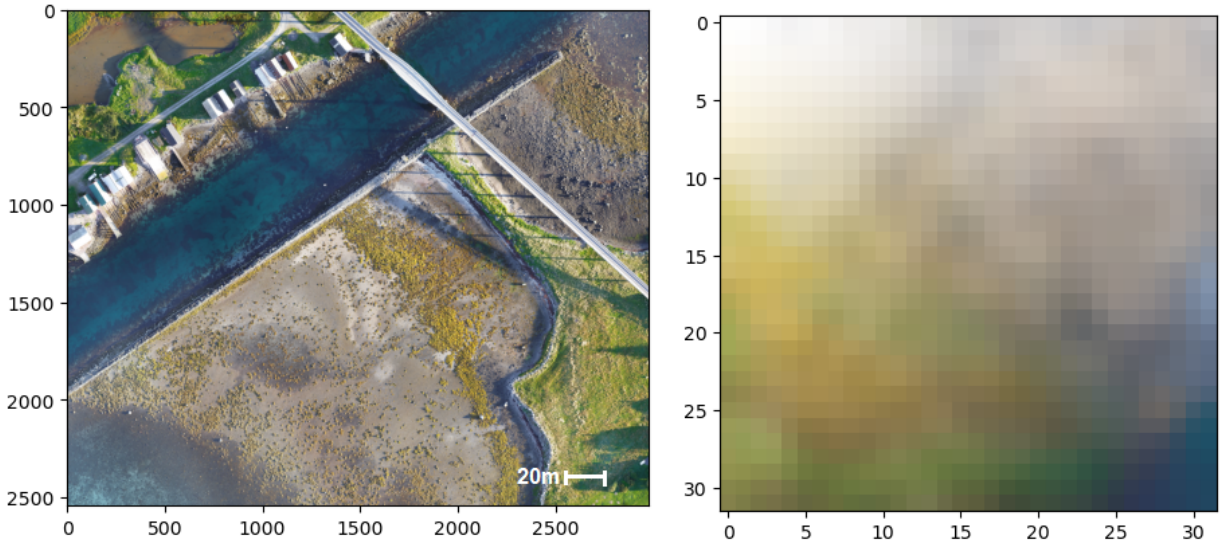


Figure 32: Remøy RGB with the color visualization of the 32x32 SOM trained for 3 epochs.

4.3.1 Dimensionality reduction using PCA

An early test of using PCA to reduce the dimensionality was done, and ended up reducing the number of bands from 224 to 10, while still maintaining 0.999 of the variance, this is more detailed above.

4.4 The fused Remøy scene

In this section, the results of the experiments using the SOM to analyze the fused Remøy scene is presented. The tests involved evaluating the performance of the SOM on various combinations of HSI bands, specifically examining three bands at a time, as well as on the three RGB bands alone. Furthermore, the analysis was extended to evaluate a new region within the RGBI where no corresponding HSI data was used in the training. The results include a comparison of different SOMs for the different tests, emphasizing the error of the nodes.

4.4.1 Dimensionality reduction with PCA

DR was also tested for the fused Remøy image. In Figure 33 the proportions of variance explained are plotted against the number of principal components.

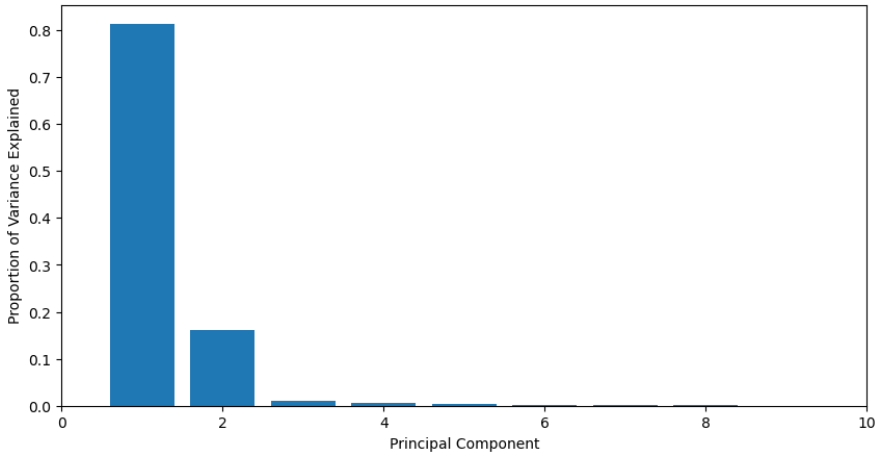


Figure 33: Scree plot showing the eigenvalues of the components derived from principal component analysis. Each pole in the plot represents an individual component, plotted in descending order of their eigenvalues.

4.5 Evaluation on the fused image

4.5.1 The spatial errors

In this section the approach for training and evaluating the SOM was slightly different. A decreasing update radius over three epochs was chosen, with an increase in the number of pixels used for training. Epoch 1 had UR = 4 with every 1000th pixel, epoch 2 UR = 2 with every 500th pixel, and epoch 3 with UR = 1 and every 200th pixel.

In Figure 34 the results for the 16x16 SOM of the fused Remøy is illustrated. The error is the relative quantization error, evaluated on all bands. The darker areas display a higher distance between the spatial information in the original image and the nodes in the SOM.

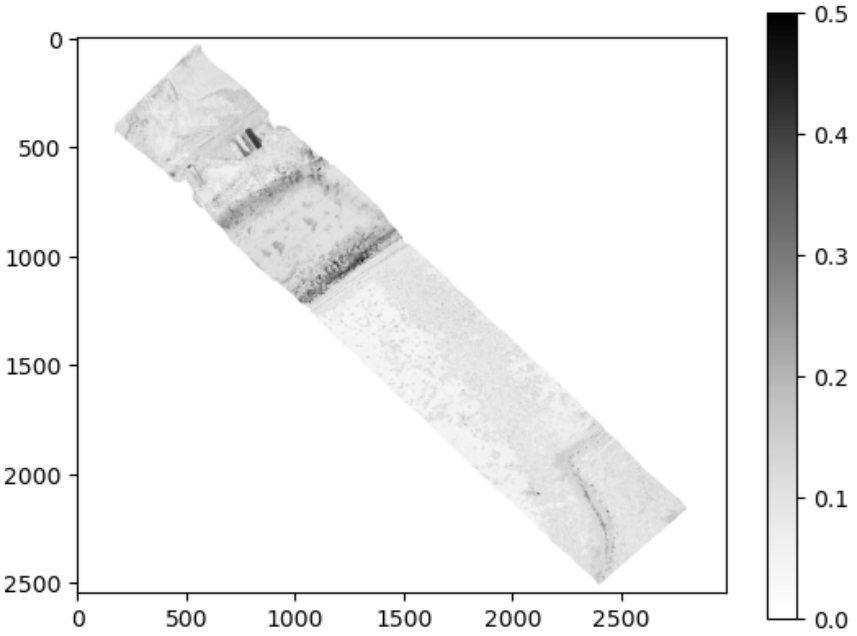


Figure 34: The spatial error for the 16x16 SOM of the fused Remøy. The SOM is trained for three iterations.

The clustering of the Remøy image can also be displayed in terms of the row and column number of the BMU, beside the relative quantization error. This can be shown in Figure 35.

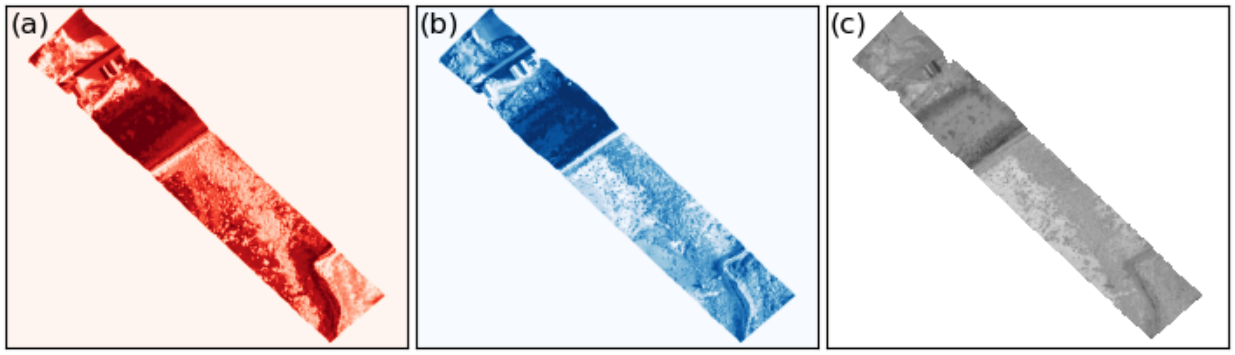


Figure 35: The image classified according to the 16x16 SOM. The SOM is trained for three iterations. The (a) column and (b) row of the BMU index. (c) the logarithmic error.

4.5.2 The Unified Distance Matrix

The U-matrix can help visualize the borders between the clustered neurons. To display the difference in the fused Remøy scene, a comparison between two SOMs are shown in Figure 36.

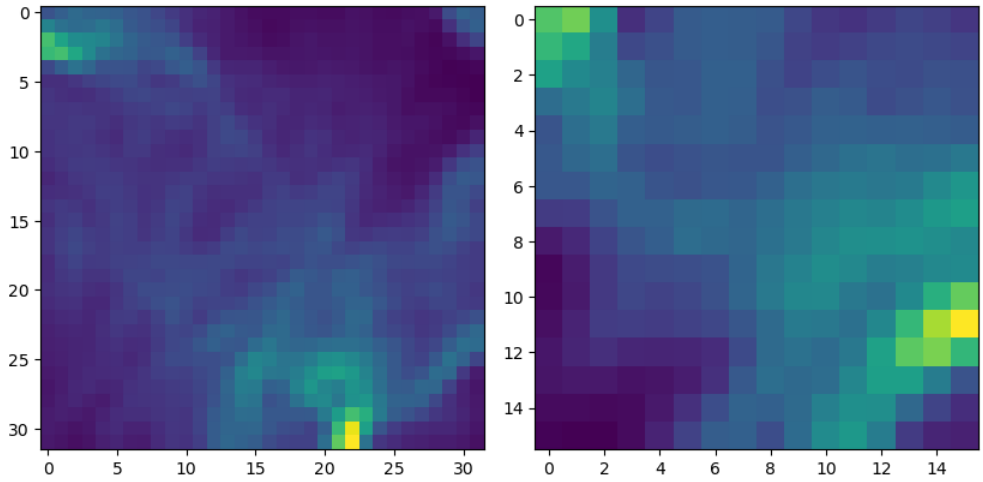


Figure 36: A comparison of the U-matrices of the fused Remøy image, using a 32x32 (left) and a 16x16 (right) SOM.

4.6 Evaluation of the RGB bands

4.6.1 The spatial errors

So far the evaluations have been looking at the errors and labels over all the data trained with the SOM. In this section the metrics presented are obtained by evaluating only the RGB part of the fused image on the SOM trained on the fused scene, as illustrated in Figure 13. Firstly, the 16x16 SOM was analyzed using the same plot as earlier, in Figure 37.

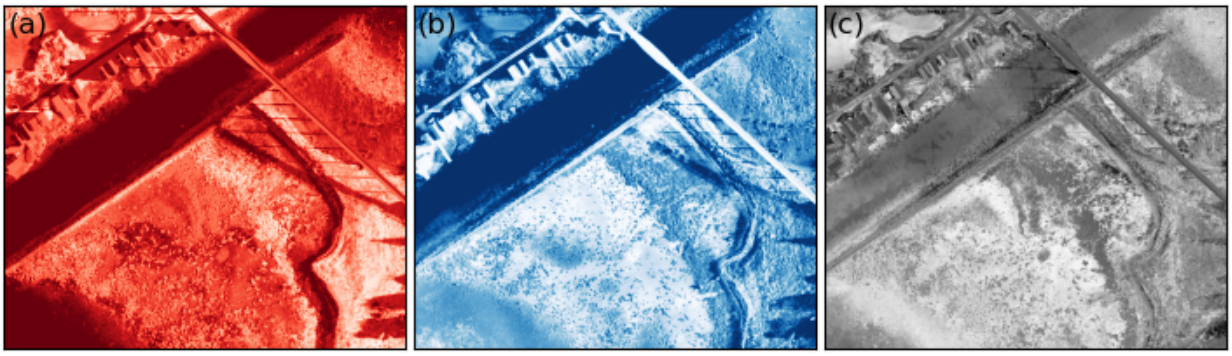


Figure 37: The image classified according to the 16x16 SOM. The SOM is trained for 3 iterations with decreasing update radius and more pixels for each epoch. The (a) column and (b) row of the BMU index. (c) the logarithmic quantization error.

The logarithmic error alone with a scale can be seen in Figure 38, with the 32x32 SOM.

Figure 38: The spatial logarithmic error for the 32x32 SOM of the fused Remøy. The SOM is trained for three iterations.

The comparison between the SOM evaluated on the fused scene and the latter one evaluated on the RGB bands can be highlighted by using the exact same training scheme. The one described in section 4.6 was used for a 64x64 SOM for both evaluations. The comparison between the two can be seen in Figure 39.

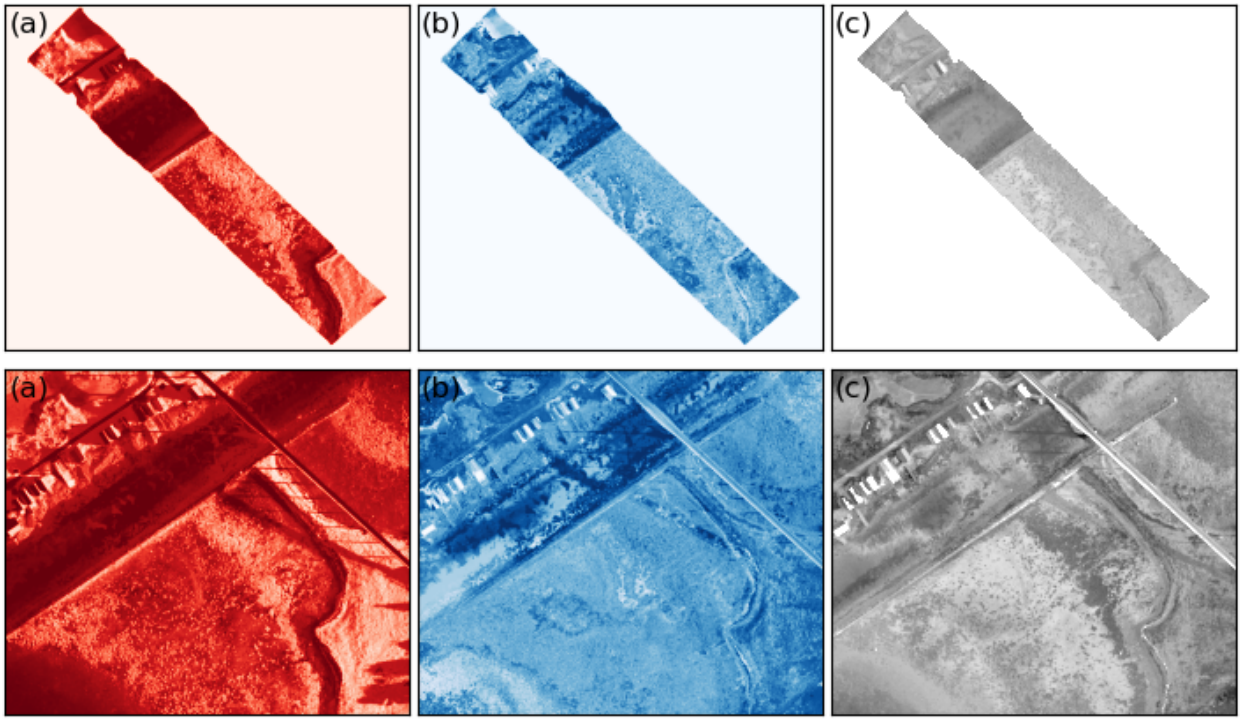


Figure 39: The image classified according to the 64x64 SOM. Same conditions as described earlier, with evaluation for fused image on top and RGB in the bottom part.

4.6.2 U-matrix

A comparison of the U-matrix for the SOM evaluated on the fused image and on only the RGB was made. Figure 40 illustrates differences in the borders of the clusters between the two evaluations.

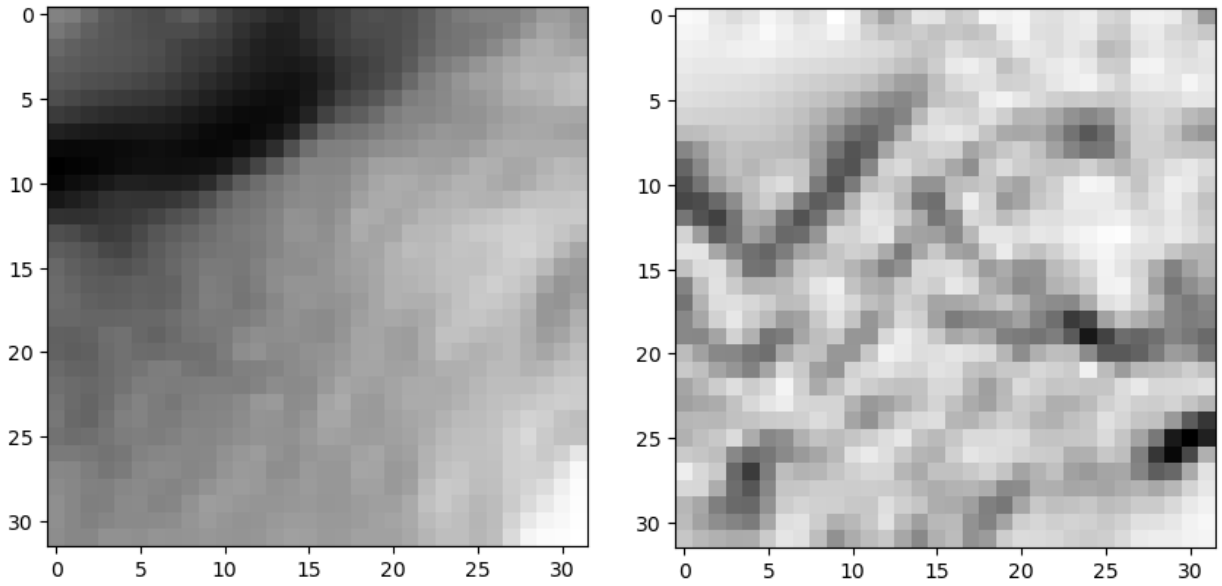


Figure 40: A comparison of the U-matrices of the fused Remøy image, using a 32x32 evaluating on the fused(left) and the RGB image (right).

5 Discussion

The section discusses the results from the previous section with the intention of evaluating the SOM's usefulness with the new fusion technique, as well as how its performance can be improved for future use.

5.1 The Samson scene

5.1.1 Pre-processing and Dimensionality reduction

This part discusses how the various pre-processing steps influence how the model is trained. These factors include, masking, downsampling, DR and more. Next it will discuss the trade-offs between the different settings of the SOM and relevant hyperparameters, with focus on reducing the time and complexity while maintaining most of the relevant information in the data.

5.1.2 Weight initialization

Three weight initialization strategies were tested to inspect its influence on the model performance: 'normal', 'random', and 'PCA'. The results demonstrated that PCA-based initialization provided a faster initial convergence, with an error approximately 0.06 lower than the others after the first epoch. This shows that PCA can immediately take advantage of the hyperspectral data's inherent linear structure. However, by the fourth epoch, the 'normal' and 'random' strategies caught up, converging to errors comparable to the PCA initialization. This might be related to the complexity and structure of the Samson image, as this is relatively low-dimensional HSI. This suggests that, although PCA-based initialization offers faster initial learning, the advantage diminishes when the iterations increases. Depending on the use of the SOM, how many iterations its going to be trained for, and the amount computational power available, as PCA requires more calculations.

5.1.3 Lattice size

The choice of lattice size in SOMs introduces a trade-off between computational complexity and the granularity of the learned representation, which Figure 16 illustrates. A significant decrease in error when the lattice size is increased from 8 to 16, indicating that the increase in computational complexity is well compensated by the gain in the model's ability to learn finer patterns in the data. However, as the lattice size continues to increase beyond 16, the reduction in error starts to flatten out. This suggests that while a larger lattice size may provide some additional gains in terms of the model's accuracy, the increase in computational complexity might not justify these marginal improvements. As the lattice size increases, the SOM gains the ability

to capture more detailed patterns in the data, thereby increasing granularity. However, this increase in lattice size leads to a higher computational cost, as each node requires calculation and adjustment during the training process.

When working with smaller datasets like the Samson HSI, a lattice size of 16 could be a reasonable choice. It provides a good balance between computational cost and granularity, ensuring a detailed enough representation without overly taxing computational resources. However, as the complexity and size of the dataset increase, it might be beneficial to experiment with larger lattice sizes to ensure that the SOM can capture the more intricate patterns in the data. However, one must be aware that as the lattice size increases, the model could become more prone to overfitting, as it may start to learn noise in the data along with the actual patterns. It's crucial to have a validation strategy in place to monitor the model's performance and prevent overfitting when working with larger lattice sizes. Thus, while considering the lattice size, both the nature of the dataset and the computational resources at hand must be considered.

5.1.4 The Unified Distance Matrix

The U-Matrix is a key tool for visualizing and interpreting a SOM's structure - High U-Matrix values signal large differences, boundaries between clusters in the data, while low values suggest similarities, pointing to areas where data points are closely related, e.g. lies in the same cluster. The U-Matrix's value comes into particular focus when exploring varying lattice sizes. A smaller lattice offers a more macroscopic view of the data, capturing broad patterns but potentially missing finer details. On the other hand, a larger lattice brings these subtleties into view, offering a more granular look at the data.

5.1.5 Update radius

In the initial stages of training, a larger update radius is usually beneficial. In the case where the SOM's weights are initialized randomly, the map won't represent the input space as well for the first iteration. Thus, a larger update radius allows the SOM to quickly learn the general structure of the input space. In our case, the weights are initialized using PCA, which makes the weights align with the direction of the greatest variance in the data. Ultimately this provides a better starting point for the learning process, and might also result in faster convergence. As training progresses, the SOM's representation of the input space improves, it is generally a good idea to reduce the update radius. This allows the SOM to fine-tune its representation of the input space, as only the nodes very close to the best match are updated. Essentially, a decreasing update radius shifts the SOM's learning from a global structure to more local details.

However, the best schedule for decreasing the update radius can vary depending on the complexity and nature of the data. Decreasing update radius will not always yield better results than a constant one. If the data is noisy, or if the complexity of the data requires a large radius for a proper mapping of global structure, a smaller radius could limit the SOM's ability to learn. In contrast, if the data is relatively simple, like in the Samson case, or the initial radius is too large, then a decreasing radius could help with fine tuning and achieve better results.

5.1.6 Epochs and convergence

The number of iterations significantly affects the SOM's accuracy and convergence. The findings shows the disparity in convergence speed between the lattice size 8 and 64 was notable. As expected, the smaller lattice converged faster due to fewer nodes to update and thus a smaller error to minimize. On the contrary, the larger lattice required more iterations to converge due to the higher granularity and larger initial error.

Regarding the selection of 15 epochs, it's a balance between giving the SOM enough iterations to learn enough of the data patterns while minimizing the computational cost. For this dataset, 15 epochs seems to be more than a enough, as the error does not decrease significantly after 10 epochs. Often one is mostly interested in the result after convergence, but for this application and the relative simplicity of the Samson image, 15 epochs seemed too time consuming for all the tests.

5.1.7 Dimensionality reduction using PCA

It is worth to keep in mind that while PCA preserves a vast majority of the variance, it might not strictly maintain the original spatial correlations within the data. This can be observed in the spatial error distribution between the original fused image and the PCA-reduced data. This is something that was not further investigated in this thesis, but could be interesting for future work.

5.1.8 Relative scaling of the RGB bands

Scaling the RGB bands provides an understanding of how the SOM is effected, and can be a good foundation for comparative analysis, as well as qualitative visualization and numeric evaluation. The relative scaling is implemented to simulate the difference in intensities when working with images from different modalities. Applying varied scaling constants to the RGB bands in the fused Samson scene introduces interesting dynamics in the performance of the Self-Organizing Map (SOM) and the ensuing spatial error. Testing with constants 0.1, 1, 10, and 100 demonstrated that the band selection significantly influences the error. The minimum error

was achieved with a scaling factor of 100 across both the three simulated RGB bands and three arbitrary bands (60th, 100th, and 140th).

An observable visual difference is apparent when contrasting the evaluations based on all bands and just a subset of bands. The spatial error across all bands is largely influenced by the scaling of the RGB bands. Conversely, the error evaluation based on a subset of bands shows a different pattern, reflecting the specific spectral accuracies rather than overarching RGB distortions. Thus, a careful band selection for evaluation, considering the spectral features of interest and the intended use of the hyperspectral data, is key. Additionally, awareness of the RGB bias in the all-band evaluation due to the scaling factor is crucial.

5.1.9 Comparison between the original and the fusioned Samson scene

The Self-Organizing Map's (SOM) performance on the original and fused images showcases minimal difference, both qualitatively, as visualized through spatial error, and quantitatively. Remarkably, when dimensionality reduction was employed using Principal Component Analysis (PCA), the performance seemed to favor the fused scene. The fusion process, while a transformation, did not hinder the model's performance. Rather, it appeared to have a slight positive effect, which aligns with expectations given the additional information from the fused dataset.

The results further suggests that due to the simplicity and low-dimensionality of the Samson scene the necessity for a complex model might not be that big. Even with only two principal components, the SOM performs well. This competent performance can be largely attributed to the simplicity of the scene and the distinctive nature of its labels. Thus, it underscores that in scenarios with less complex data, simpler models may suffice. This was also backed up by the confusion matrices, which only displayed a marginal difference between the most complex Samson SOM (64x64x156) and the fused simplified SOM (16x16x2).

5.2 The Remøy scene

5.2.1 Mapped nodes of the SOM from the RGB Remøy

In Figure 32, the nodes of the 32x32 SOM of the RGB were normalized and plotted to resemble the input image. One can observe that the different features of the image are displayed through the nodes, with the upper left part being the bright features in the houses and the bridge, and the bottom right the blue in water. Green and yellow features were observed in the left part, representing the algae, which is one of the features of interest for the SeaBee project, and the

grass. This is an indicator that the SOM is able to represent the information from the input in the weights of the nodes.

5.2.2 Reducing the dimensionality with PCA

The scree plot showing the variance for different number of principal components, Figure 33, suggested that quite few PC were needed to retain most of the information in the image. This was previously mentioned and discussed in the literature, to be a successful way to get rid of high numbers of correlated bands. As this neither affected the training time or the quantization error significantly, it was not something that was deeply investigated, and should be something to look into in future work.

5.2.3 Evaluation of the fused Remøy

In this section, a more thorough training procedure were implemented to the model. This was mostly due to findings in the preliminary results found in the Samson scene regarding decreasing update radius and sampling size of the data. Because of the simplicity of the Samson and the increased complexity in the Remøy scene, the method chosen was to use three iterations of training. Starting with a bigger UR and less data points, one could achieve more refined details in the final iteration, learning the general and less detailed patterns first. This was discussed earlier in the thesis.

The errors in the scene seems to be mostly centred in areas with complex features and in the water. In Figure 35 one can observe that the border between the water and the land, as well as the beach and the rock-edge in the bottom, seem to have the highest error. The algae, as mentioned earlier, does also demonstrate complexity as these areas appear darker.

5.2.4 Evaluation of the RGB bands

This section will discuss how the SOM is projecting the spectral information from training on the fused Remøy scene onto an evaluation with just the RGB Remøy. In the initial SOM with a lattice size of 16x16, one can observe that it struggles to distinguish between different parts of the water, seen in Figure 37. In b), it manages to differentiate between the shallow water in the bottom left and the canal, but not in a). Generally, the plots suggests that the feature's complexity determines how large the error is. For the next plot, Figure 38, one can observe more details in the water part, as well as a significantly lower error overall. The most interesting part can be observed in the water where the HSI is present. There appears to be a clear border where

the HSI information disappears. The bottom right corner, as well as the algae, seems to be the hardest features for the SOM to project.

5.2.5 Comparison between evaluation with fused scene and RGB scene

A comparison of the 64x64 SOM was done by evaluating on the fused data and then replacing the nodes of the SOM with only the three RGB bands, as described in the methodology. In Figure 39, one can observe the same phenomena as earlier with a brighter area appearing in the water of the spatial error. The differences can be observed throughout the image, but might still be the most apparent for the hardest features, such as algae and bounds of clusters. The method for projecting the spectral information using only the RGB bands appears to be working, which can be visually observed in these images.

5.2.6 Difference in fusion with simulated and real data

As mentioned in the the theory when giving a brief overview of the existing literature on image fusion, most of the techniques work with simulated data. One of the strengths with simulated is the ability to control the various conditions, such that the analysis can focus on the metrics and the model in different controlled environments. This can eliminate time-consuming work, but the downside is the lack of realism. Simulations can lose accuracy due to simplifications that are done to be able to deal with the complexity in real-life data. When using data from multiple modalities the information is authentic and represents the use-cases such that the results are directly transferable to the real applications.

In this study, both simulated and real data was used for the application of image fusion with the SOM. Unfortunately, the results for the two scenes are not comparable as they differ heavily in complexity and features. Future work could involve quantification of the difference in the error between a simulated couple of image fusion and one containing real data.

5.2.7 Future work

For future work, as mentioned in different parts of the discussion, there are several areas to explore further:
1. The investigation of how real data actually affects the model compared to simulated data. This also involves relative scaling of the real RGB pixels.
2. How PCA affects the fusion when projecting the hyperspectral information onto the RGB image.
3. Introducing multiple images from other modalities, such as other MS sources. 4. Further investigations of the effect of model parameters in the fusion process.

5. Apply and compare the fusion on more datasets.

HYPSO

The HYPSO satellite is a small research satellite developed at the NTNU SmallSat Lab equipped with a hyperspectral camera. Its mission is to scan different water such as lakes and oceans from 500 km the Earth's surface, in order to detect harmful algal blooms. For future work a possibility is to apply the same SOM framework to data here, with the goal of enhancing and improving the coastal mapping and research in the field.

6 Conclusion

The application of the Self-Organizing Map (SOM) trained on fused data from the Remøy scene, incorporating real data from both a hyperspectral image (HSI) and a high-resolution RGB provides a straightforward and effective method for combining the strengths of each modality. This resulted in enhancing the spectral information of the scene beyond the boundaries of the HSI. Secondly, extensive evaluations of the SOM model's performance have been conducted, by exploring different parameters and configurations of the model. This analysis has provided insights into the impact of various factors on the fusion process, aiding in improving the model's performance and applicability.

Thirdly, the performance of the SOM model has been assessed on two different hyperspectral scenes, showcasing its usefulness across one simple and one complex dataset. This evaluation demonstrates the potential of this SOM method for more than just one application. Additionally, a comparison between fusion with simulated and real images has been conducted, highlighting the strengths and weaknesses of each approach. While simulated data allows for controlled analysis, real data offers a more realistic representation of the actual application scenarios. This is the biggest difference between the approach presented in this thesis and previous literature on image fusion.

Finally, an important contribution of this research is the evaluation of the SOM model's ability to extract hyperspectral information in areas where only RGB data is available. The results demonstrate that the fusion of HSI and RGB data enables the SOM to project spectral information beyond the boundaries of the HSI, possibly providing valuable insights and enhancing the overall analysis. This is particularly useful in applications such as coastal mapping, where HSI is limited and expensive.

References

[1] Zhenfeng Shao and Jiajun Cai. Remote Sensing Image Fusion With Deep Convolutional Neural Network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(5):1656–1669, May 2018. ISSN 2151-1535. doi: 10.1109/JSTARS.2018.2805923.

[2] Prasad S. Thenkabail and John G. Lyon. *Hyperspectral Remote Sensing of Vegetation*. CRC Press, April 2016. ISBN 978-1-4398-4538-7. Google-Books-ID: B0LNBQAAQBAJ.

[3] Paul M. Mather and Magaly Koch. *Computer Processing of Remotely-Sensed Images: An Introduction*. John Wiley & Sons, July 2011. ISBN 978-1-119-95640-2. Google-Books-ID: GWhvDMNh1hAC.

[4] NIVA. KELPMAP-Vega: Metodeutvikling for kartlegging av tareskog ved bruk av droner og satellittbilder - Miljødirektoratet. URL https://www.miljodirektoratet.no/publikasjoner/2022/desember/kelpmap-vega-metodeutvikling-for-kartlegging-av-tareskog-ved-bruk-av-droner-og-sate

[5] Hyperspectral and Multispectral Imaging | Edmund Optics. URL https://www.edmundoptics.eu/knowledge-center/application-notes/imaging/hyperspectral-and-multispectral-imaging/.

[6] Naoto Yokoya, Takehisa Yairi, and Akira Iwasaki. Coupled Nonnegative Matrix Factorization Unmixing for Hyperspectral and Multispectral Data Fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 50(2):528–537, February 2012. ISSN 1558-0644. doi: 10.1109/TGRS.2011.2161320.

[7] Yun Zhang and Rakesh K. Mishra. A review and comparison of commercially available pan-sharpening techniques for high resolution satellite image fusion. In *2012 IEEE International Geoscience and Remote Sensing Symposium*, pages 182–185, July 2012. doi: 10.1109/IGARSS.2012.6351607.

[8] Luciano Alparone, Lucien Wald, Jocelyn Chanussot, Claire Thomas, Paolo Gamba, and Lori Mann Bruce. Comparison of Pansharpening Algorithms: Outcome of the 2006 GRS-S Data-Fusion Contest. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10):3012–3021, October 2007. ISSN 1558-0644. doi: 10.1109/TGRS.2007.904923.

[9] EOS. Panchromatic Imagery And Its Band Combinations In Use, September 2021. URL https://eos.com/make-an-analysis/panchromatic/.

[10] Yi Li, Junli Zhao, Zhihan Lv, and Zhenkuan Pan. Multimodal Medical Supervised Image Fusion Method by CNN. *Frontiers in Neuroscience*, 15, 2021. ISSN 1662-453X. URL https://www.frontiersin.org/articles/10.3389/fnins.2021.638976.

[11] Myungjin Choi. A new intensity-hue-saturation fusion approach to image fusion with a tradeoff parameter. *IEEE Transactions on Geoscience and Remote Sensing*, 44(6):1672–1682, June 2006. ISSN 1558-0644. doi: 10.1109/TGRS.2006.869923.

[12] R.L. King and Jianwen Wang. A wavelet based algorithm for pan sharpening Landsat 7 imagery. In *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*, volume 2, pages 849–851 vol.2, July 2001. doi: 10.1109/IGARSS.2001.976657.

[13] Anke Meyer-Baese. Image Fusion - an overview \textbar ScienceDirect Topics. URL https://www.sciencedirect.com/topics/engineering/image-fusion.

[14] Ronny Hänsch and Olaf Hellwich. Fusion of Multispectral LiDAR, Hyperspectral, and RGB Data for Urban Land Cover Classification. *IEEE Geoscience and Remote Sensing Letters*, 18 (2):366–370, February 2021. ISSN 1558-0571. doi: 10.1109/LGRS.2020.2972955.

[15] Mayank Mishra. Convolutional Neural Networks, Explained \textbar by Mayank Mishra \textbar Towards Data Science. URL https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939.

[16] Yi Li, Junli Zhao, Zhihan Lv, and Zhenkuan Pan. Multimodal Medical Supervised Image Fusion Method by CNN. *Frontiers in Neuroscience*, 15, 2021. ISSN 1662-453X. URL https://www.frontiersin.org/articles/10.3389/fnins.2021.638976.

[17] Hao Zhang, Han Xu, Xin Tian, Junjun Jiang, and Jiayi Ma. Image fusion meets deep learning: A survey and perspective. *Information Fusion*, 76:323–336, December 2021. ISSN 1566-2535. doi: 10.1016/j.inffus.2021.06.008. URL https://www.sciencedirect.com/science/article/pii/S1566253521001342.

[18] Yanan Luo, Jie Zou, Chengfei Yao, Xiaosong Zhao, Tao Li, and Gang Bai. HSI-CNN: A Novel Convolution Neural Network for Hyperspectral Image. In *2018 International Conference on Audio, Language and Image Processing (ICALIP)*, pages 464–469, July 2018. doi: 10.1109/ICALIP.2018.8455251.

[19] Naoto Yokoya, Claas Grohnfeldt, and Jocelyn Chanussot. Hyperspectral and Multispectral Data Fusion: A comparative review of the recent literature. *IEEE Geoscience and Remote Sensing Magazine*, 5(2):29–56, June 2017. ISSN 2168-6831. doi: 10.1109/MGRS.2016.2637824.

[20] Tom Michael Mitchell. *Machine Learning*. McGraw-Hill Education, March 1997. ISBN 978-0-07-042807-2. Google-Books-ID: xOGAngEACAAJ.

[21] Bishop. Pattern Recognition and Machine Learning | SpringerLink. URL https://link.springer.com/book/9780387310732.

[22] Ajay Kulkarni. Confusion Matrix - an overview | ScienceDirect Topics. URL https://www.sciencedirect.com/topics/engineering/confusion-matrix.

[23] J. Vesanto and E. Alhoniemi. Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 11(3):586–600, May 2000. ISSN 1941-0093. doi: 10.1109/72.846731. Conference Name: IEEE Transactions on Neural Networks.

[24] U. Shruthi, V. Nagaveni, and B.K. Raghavendra. A Review on Machine Learning Classification Techniques for Plant Disease Detection. In *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, pages 281–284, March 2019. doi: 10.1109/ICACCS.2019.8728415. ISSN: 2575-7288.

[25] scikit learn. 2.3. Clustering. URL https://scikit-learn/stable/modules/clustering.html.

[26] Mathias Seuret, Michele Alberti, Marcus Liwicki, and Rolf Ingold. PCA-Initialized Deep Neural Networks Applied to Document Image Analysis. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 01, pages 877–882, November 2017. doi: 10.1109/ICDAR.2017.148. ISSN: 2379-2140.

[27] Aksel S. Danielsen, Tor Arne Johansen, and Joseph L. Garrett. Self-Organizing Maps for Clustering Hyperspectral Images On-Board a CubeSat. *Remote Sensing*, 13(20):4174, January 2021. ISSN 2072-4292. doi: 10.3390/rs13204174. URL https://www.mdpi.com/2072-4292/13/20/4174.

[28] Kadim Tasdemir and ErzsÉbet Merenyi. Exploiting Data Topology in Visualization and Clustering of Self-Organizing Maps. *IEEE Transactions on Neural Networks*, 20(4):549–562, April 2009. ISSN 1941-0093. doi: 10.1109/TNN.2008.2005409. Conference Name: IEEE Transactions on Neural Networks.

[29] Felix M. Riese, Sina Keller, and Stefan Hinz. Supervised and Semi-Supervised Self-Organizing Maps for Regression and Classification Focusing on Hyperspectral Data. *Remote Sensing*, 12(1):7, January 2020. ISSN 2072-4292. doi: 10.3390/rs12010007. URL https://www.mdpi.com/2072-4292/12/1/7. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

[30] M Graña. Hyperspectral Remote Sensing Scenes - Grupo de Inteligencia Computacional (GIC). URL https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

[31] Ian T. Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, April 2016. doi: 10.1098/rsta.2015.0202. URL https://royalsocietypublishing.org/doi/full/10.1098/rsta.2015.0202.

[32] Basna Mohammed Salih Hasan and Adnan Mohsin Abdulazeez. A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. *Journal of Soft Computing and Data Mining*, 2(1):20–30, April 2021. ISSN 2716-621X. URL https://publisher.uthm.edu.my/ojs/index.php/jscdm/article/view/8032. Number: 1.

[33] Craig Rodarmel and Jie Shan. Principal Component Analysis for Hyperspectral Image Classification. *Surv Land inf Syst*, 62, January 2002.

[34] Weijian Wan and D. Fraser. Multisource data fusion with multiple self-organizing maps. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1344–1349, May 1999. ISSN 1558-0644. doi: 10.1109/36.763298.

[35] P Martínez, J A Gualtieri, P L Aguilar, R M Pérez, M Linaje, J C Preciado, and A Plaza. HYPERSPECTRAL IMAGE CLASSIFICATION USING A SELF-ORGANIZING MAP.

[36] A. J Richardson, C Risien, and F. A Shillington. Using self-organizing maps to identify patterns in satellite imagery. *Progress in Oceanography*, 59(2):223–239, October 2003. ISSN 0079-6611. doi: 10.1016/j.pocean.2003.07.006. URL https://www.sciencedirect.com/science/article/pii/S007966110300171X.

[37] Ferdinando Giacco, Silvia Scarpetta, Luca Pugliese, Maria Marinaro, and Christian Thiel. Application of Self Organizing Maps to multi-resolution and multi-spectral remote sensed images. volume 193, pages 245–253, January 2008. doi: 10.3233/978-1-58603-984-4-245.

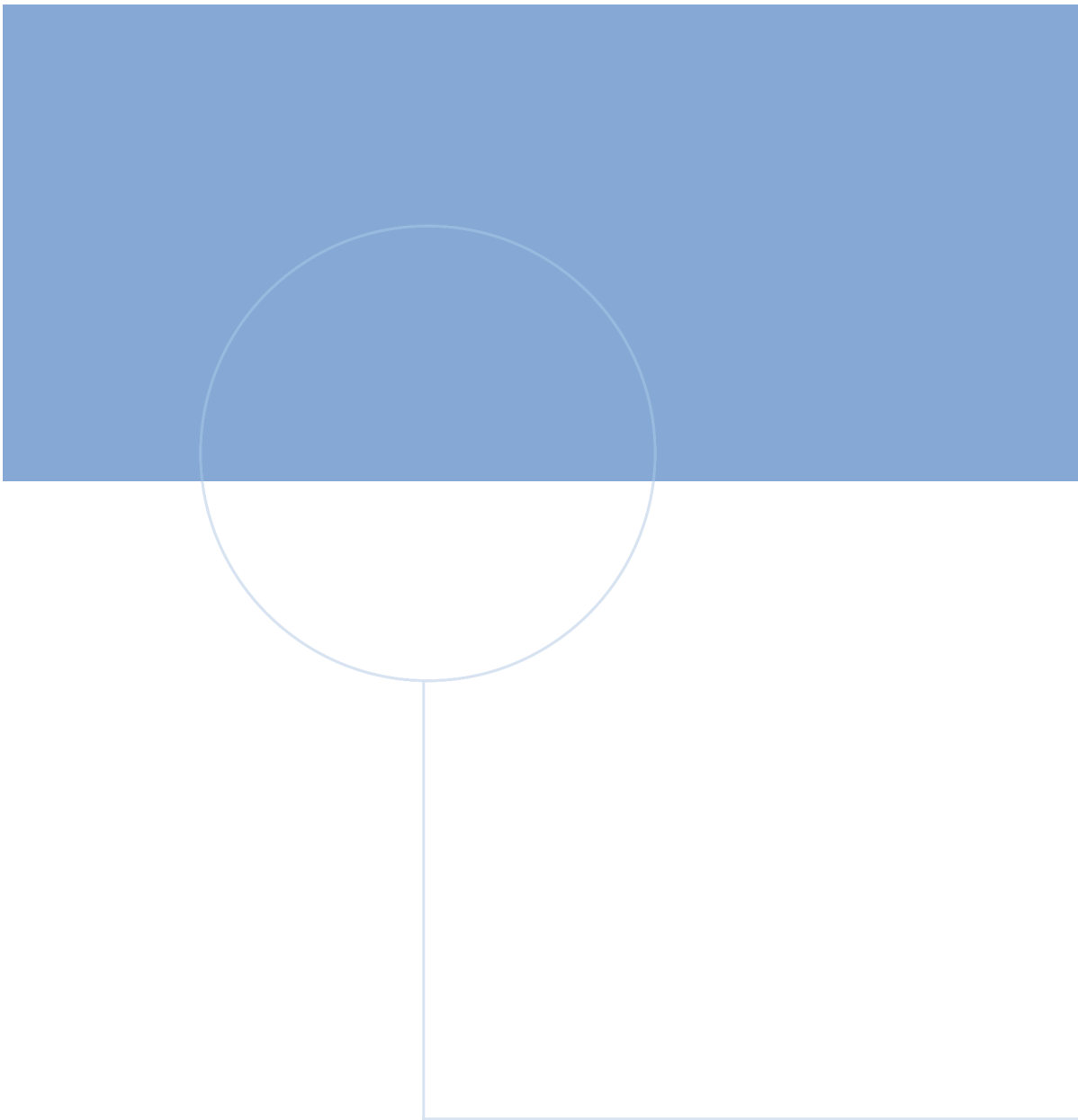
[38] Eckart Michaelsen. Self-organizing maps for fusion of thermal hyperspectral- with high-resolution VIS-data. In *2014 8th IAPR Workshop on Pattern Reconition in Remote Sensing*, pages 1–4, August 2014. doi: 10.1109/PRRS.2014.6914281.

[39] Felix M. Riese and Sina Keller. Introducing a Framework of Self-Organizing Maps for Regression of Soil Moisture with Hyperspectral Data. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 6151–6154, July 2018. doi: 10.1109/IGARSS.2018.8517812.

[40] Duccio Rocchini, Giles M. Foody, Harini Nagendra, Carlo Ricotta, Madhur Anand, Kate S. He, Valerio Amici, Birgit Kleinschmit, Michael Förster, Sebastian Schmidtlein, Hannes Feilhauer, Anne Ghisla, Markus Metz, and Markus Neteler. Uncertainty in ecosystem mapping by remote sensing. *Computers & Geosciences*, 50:128–135, January 2013. ISSN 0098-3004. doi: 10.1016/j.cageo.2012.05.022. URL https://www.sciencedirect.com/science/article/pii/S0098300412001781.

[41] NIVA. KELPMAP-Vega: Metodeutvikling for kartlegging av tareskog ved bruk av droner og satellittbilder - Miljødirektoratet. URL https://www.miljodirektoratet.no/publikasjoner/2022/desember/kelpmap-vega-metodeutvikling-for-kartlegging-av-tareskog-ved-bruk-av-droner-og-sate

[42] Curtis E. Woodcock, Alan H. Strahler, and David L. B. Jupp. The use of variograms in remote sensing: I. Scene models and simulated images. *Remote Sensing of Environment*, 25(3): 323–348, August 1988. ISSN 0034-4257. doi: 10.1016/0034-4257(88)90108-3. URL https://www.sciencedirect.com/science/article/pii/0034425788901083.



NTNU

Norwegian University of
Science and Technology