

Investigating Unsupervised Machine Learning for Seismic Facies Clusters using SOMs

Umang Nagpal*, Oil and Natural Gas Corporation Ltd., Ayush Kumar*, Indian Oil Corporation Ltd., Santosh Dhubia and Chintan Prajapati, Gujarat Energy Research and Management Institute

Nagpal_Umang@ongc.co.in, KumarA126@indianoil.in

Keywords

Machine Learning, Seismic Lithofacies, Principal Component Analysis (PCA), Self-Organizing Map, Seismic Attributes, Prototype Vectors (PVs), Best Matching Unit (BMU)

Abstract

To handle the increasing volume of seismic data, machine learning methods, specially unsupervised classification techniques, have emerged as a valuable tool. This research paper focuses on the application of unsupervised classification methods, specifically Principal Component Analysis (PCA) and Kohonen Self-organizing maps, for seismic facies analysis. Through case studies from the F3 Block of the Netherlands, the effectiveness and potential of these techniques are explored. Mathematically derived Seismic attributes were used as an input for clustering and recognizing seismic facies. By capturing the variability within seismic data, these methods reveal crucial insights into subsurface geological features. This paper highlights the importance of efficient data reduction algorithms for optimizing the analysis of seismic facies.

Introduction

Seismic interpretation is an essential component of exploration and production in the oil and gas industry, providing geoscientists with crucial insights for developing hydrocarbon prospects. The reflection seismic approach only provides indirect observations with constrained geographical and temporal resolutions, making the interpretation procedure labor and time-intensive. Despite advancements in data acquisition, processing, and software applications, reconstructing the geologic story from seismic surveys remains a significant challenge. Over the past few decades, machine learning techniques have become increasingly important in seismic analysis for oil & gas exploration. Supervised learning and unsupervised learning are the two categories under which these algorithms fall. Supervised learning techniques involve developing prediction models when labeled output data is available, discovering applications in geophysical issues such seismic interpretation, facies

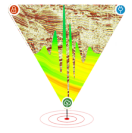
categorization, and petrophysical property forecasts. Unsupervised learning, on the other hand, concentrates on understanding patterns in input data without associated output labels.

The principal component analysis (PCA) methodology, a potent method for obtaining pertinent data from large datasets, is frequently linked to unsupervised learning. Data is projected onto a new coordinate system that is determined by the principal components with the biggest variances using the dimensionality reduction technique known as PCA. Moreover, the combination of PCA with clustering techniques has been explored in studies on seismic attributes analysis. Another notable method is the Kohonen Self-Organizing Map (SOM), an artificial neural network introduced by T. Kohonen. SOM, applying competitive learning, generates a low-dimensional picture of the input space while maintaining its characteristics.

This property is particularly valuable for visualization tasks, where similar clusters in the latent space are grouped together. SOM has been widely employed in seismic facies classification, with its connection to 1D and 2D colormaps, which facilitate seismic interpretation. In this study, we use a dataset obtained from the Netherlands Offshore F3 Block survey. By exploring the capabilities of unsupervised learning algorithms like PCA and SOM, we aim to enhance the understanding and interpretation of seismic data, contributing to the advancement of geophysical analysis in oil exploration.

Seismic Data

The seismic dataset utilized in this study is the publicly available Netherlands Offshore F3 Block 3D seismic survey. This survey is situated in the Dutch offshore region of the Central Graben basin within the North Sea. The dataset encompasses a total area



of 384 sq-kms and comprises time migrated 3D seismic data.

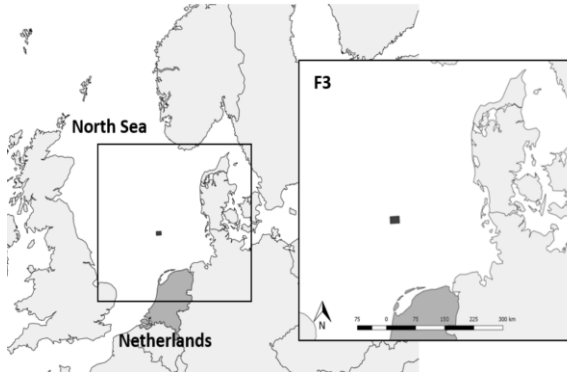


Figure 1: Location of the F3 3D survey in the North Sea, Netherlands offshore.(after Silva R et. Al. 2019)

It consists of 651 inlines and 951 crosslines, covering the offshore area of the Netherlands in the North Sea (Figure 1). The seismic data spans a time range of 1,848 milliseconds, with a sampling rate of 4 milliseconds and a bin size of 25 meters.

Methodology

The code developed for this process is based on Python. We start with installing desired Python packages i.e segysak, pykuwahara, susi, segyio, scipy, seaborn, etc. After the 3D volume is loaded, we visualize 2D seismic sections corresponding to inline, crossline and time slice (Figure2). The workflow used in this paper is illustrated in (Figure 3). Seismic attributes, derived mathematically from seismic data, are essential tools for analyzing seismic facies and capturing reflection characteristics that represent rock and sedimentary processes.

They serve as inputs for pattern recognition methods in seismic facies analysis. The selection of attributes plays a significant role in the precision of clustering and pattern identification algorithms. It is important to choose attributes with clear physical meanings, avoid redundancy, and ensure their contribution to accurately mapping the target facies.

In this paper, we calculate 14 different physical seismic attributes for input into the PCA algorithm in addition to T(t) and H(t), (Table 1).

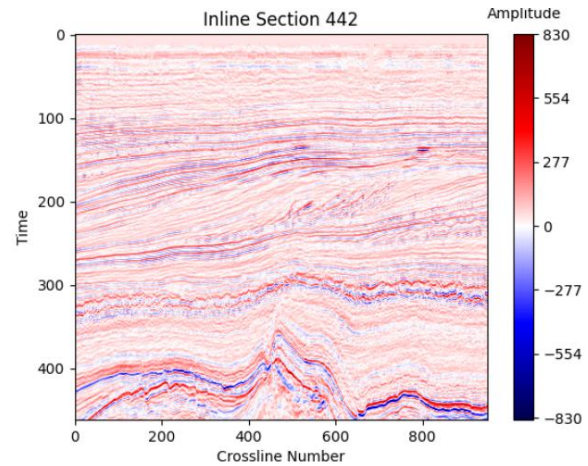


Figure 2: Inline section 442 of F3 Netherlands Block

The Kuwahara filter is applied on the generated seismic attributes to reduce the input noise. Processing which we convert the 2D data (for each calculated attributes) to 1D. The Data standardization and normalization is performed using the formula as:

$$x(\text{new}) = \frac{(x(\text{old}) - \mu)}{\sigma} \quad \text{-----} \quad 1$$

Such that, x here is attributes, new mean (μ') = 0; new Standard Deviation (σ') = 1

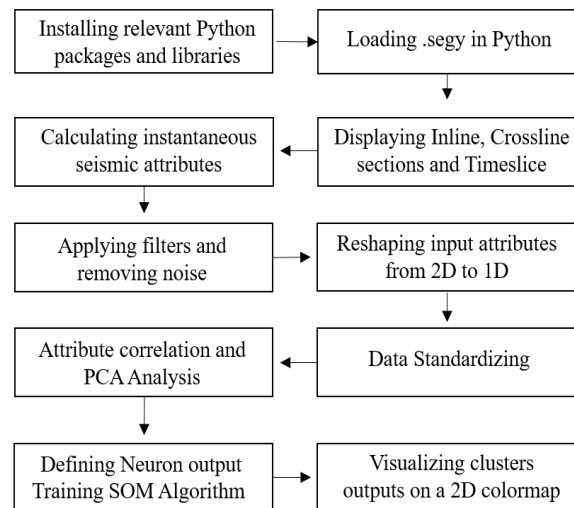
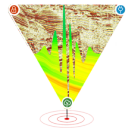


Figure 3: Workflow used in the paper



Before we send the attributes as input to our PCA model, it's essential to drop strongly correlating attributes hence we produce a HeatMap (Figure 4) to check for redundant attributes.

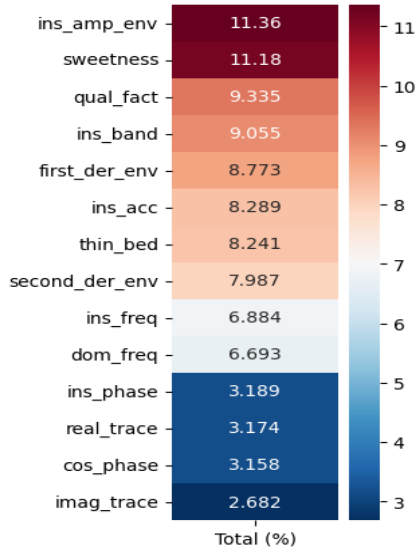


Figure 4: Correlating Attributes among themselves

The main idea about using PCA before SOM is to project the original dataset (i.e seismic attributes) into a new form with an effectively lower dimensionality. Principal components are projections that combine the original dataset in a linear fashion in the direction of highest variance. Out of 10 PCA vectors generated in our code, we select the top 4 PCAs which account for almost 80% of the original information. (Figure5)

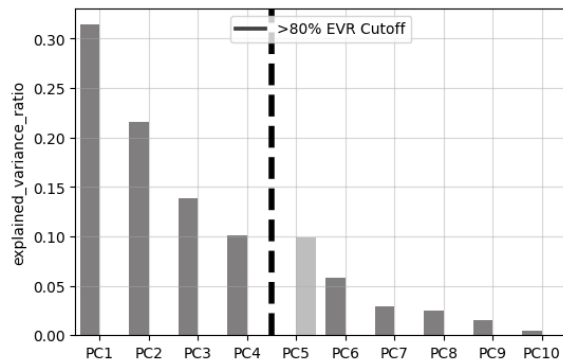


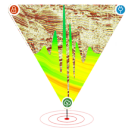
Figure 5: Trace Attributes important PCs

Instantaneous Amplitude	$E(t) = \sqrt{T^2 + H^2}$
Instantaneous Phase	$\emptyset(t) = \tan^{-1} H(t)/T(t)$
Instantaneous Frequency	$\dot{\omega} = d\emptyset(t)/dt$
First Derivative Amplitude	$DE = dE(t)/dt$
Second Derivative Amplitude	$DDE = d(DE)/dt$
Instantaneous Acceleration	$a(t) = d\dot{\omega}/dt$
Instantaneous Bandwidth	$B(t) = [d(E(t)/dt)/2\pi E(t)]$
Thin Bed	$TB(t) = \dot{\omega} - a(t)$
Instantaneous Quality Factor	$QF(t) = 0.5(\dot{\omega}/B(t))$
Dominant Frequency	$DF(t) = \dot{\omega}^2 + B^2$
Sweetness	$S(t) = E(t)/SQRT(DF(t))$
Cosine of Phase	$C(t) = \cos(\emptyset(t))$

Table 1: List of attributes used as input. Complex seismic trace $F(t)$ of a real seismic trace $T(t)$ can be written as: $F(t) = T(t) + iH(t)$, where $H(t) =$ Hilbert Transform of $(T(t))$.

The Self-Organizing Map (SOM) (Figure 6) organizes seismic attributes, represented as vectors in the space \mathbb{R}^n , $x=[x_1,x_2,\dots,x_n]$, into a geometric structure known as the SOM. Each unit of SOM, called a "Prototype Vector (PV)," is associated with its neighboring units, typically forming hexagonal or rectangular maps in two dimensions. For the map has P elements as assumed, then P n -dimensional prototype vectors m_i , $m_i=[m_{i1}, m_{i2},\dots, m_{in}]$, $i=1, 2,\dots, P$ will exist, where n is the quantity of input seismic properties. Through SOM training, these prototype vectors become a representative depiction of the input dataset. Furthermore, by employing SOM neighborhood training, these prototype vectors transform into clusters that signify different classes within the dataset. The output map will also show data that are close to one another in the input space. Since seismic data captures changes in the underlying geology and the SOM may be understood as n -dimensional input space is mapped onto a two-dimensional grid to achieve reduced intrinsic dimensionality while preserving the topological structure. An input vector is initialized during the SOM training process and is compared to all N -dimensional PVs on the 2D grid latent space. The prototype vector is then referred to as the "Winner" or the Best Matching Unit (BMU) since it is closest in Euclidean distance (d_j) to the input vector

$$d_j = \sqrt{\sum_{i=1}^N (x_i - w_{ji})^2} \quad \text{-----} \quad 2$$



here, x_i represents the i th sample vector in the dataset X , the w_{ji} represents the weights vector that joints input node and the j th output node

$$\|x - w_b\| = \min\|x - w_j\| \text{ ----- } 3$$

where $\|\cdot\|$ represent the Euclidean distance, w_b is the closest vector to x on the map. Small neighbourhood of PVs around the this BMUs are also updated. With successive iteration, the neighbourhood radius decreases and the cluster with a particular hue converges. As a result, the training causes the weight of the "winner" prototype vector to be closer to the input trace. Finally, we project the 64 trained PVs onto a 2D colour space for human visualization and enhanced interpretation.

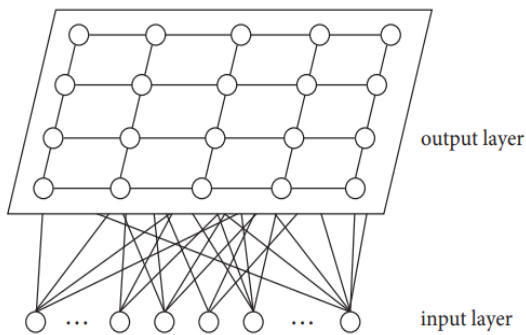


Figure 6: SOM Architecture (after Zhu Z, et. Al '22)

Results

In our seismic facies clustering study, we utilized unsupervised learning techniques to identify distinct patterns within a seismic section. By employing Self-Organizing Maps (SOM), we clustered the seismic data based on shared attributes and patterns. Our methodology involved enhancing the seismic image quality using the Kuwahara filter, data conditioning, and calculating additional physical attributes. The resulting noise-free seismic image served as input for the SOM algorithm, allowing us to divide the seismic section into multiple clusters (16, 64, 144, and 256) for inline section 442. Each cluster represented a unique seismic facies pattern, determined by grouping similar seismic amplitudes together (Figure 7 and Figure 8).

The Figure 9 illustrates the combination of 8 clusters, revealing additional seismic patterns. The different cluster sets were divided into eight sets labeled from 1 to 8, each represented by a different color as shown in the figure for inline section 200.

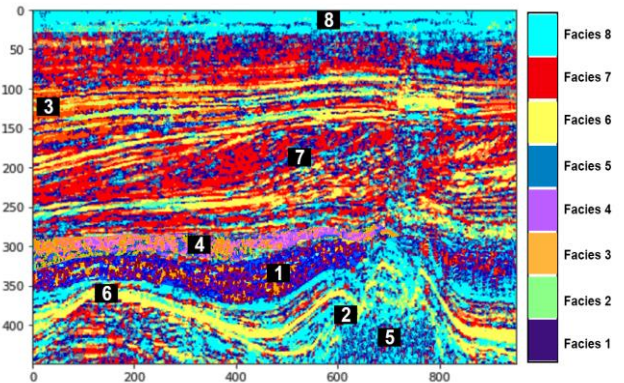


Figure 9: 8 Cluster Feature Map, Labelled Interpreted Seismic Facies.

Through the clustering analysis, we identified several distinct seismic facies clusters, including Low Coherency (facies 8) represented by a sky-blue color (cyan), High Amplitude (facies 6) represented by yellow, Horizontal Reflector (facies 3) represented by orange, Chaotic (facies 4) represented by baby pink color, Random (facies 5) represented by shades of blue, Low Coherency and Low Amplitude (facies 1) represented by shades of magenta, Low Amplitude (facies 7) represented by sharp red, and High Amplitude and Continuous (facies 2) represented by a light green. These seismic facies patterns provide indications of the depositional settings and consistent lithological properties. For instance, facies 8 suggests the presence of a weathered zone or noisy data, facies 6 indicates bright spots or gas pockets, facies 3 suggests parallel deposition of shale and sand, facies 4 may indicate slumped deposits, facies 5 could represent random signals from a salt dome, and facies 1 demonstrates deposition due to deformed structures.

Further investigation of the clusters revealed promising results in terms of higher resolution using more clusters say 64, 144, or 254 which can provide more informative insights for seismic data interpretation. Based on the calculated results, we can conclude that clustering based on

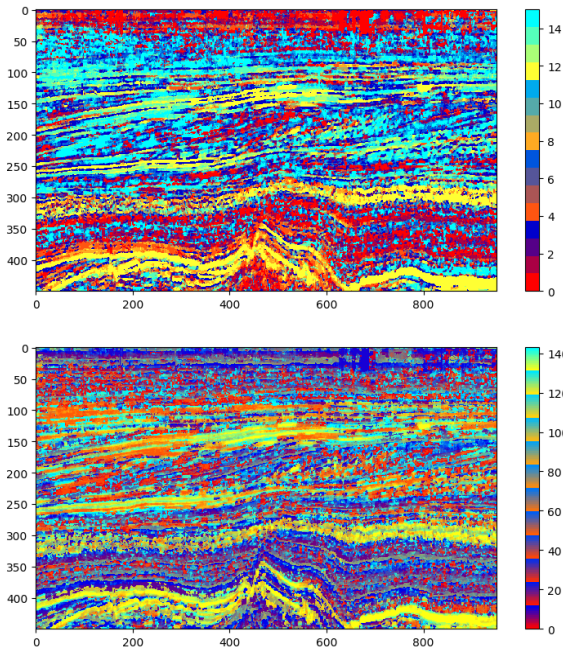
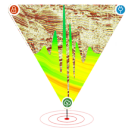


Figure 7: 16 (Top), 144 (Bottom) Cluster SOM Map

unsupervised learning serves as a supplementary tool for qualitative analysis

Conclusion and Validation

We successfully identified distinct seismic facies patterns within a seismic section and analyzed their characteristics. Through the utilization of unsupervised learning algorithms SOM, and data conditioning techniques, we achieved noise reduction and enhanced the quality of the seismic images. The results of our study demonstrated the effectiveness of unsupervised learning in clustering seismic data based on shared attributes and patterns.

Napoli et al. (2021) conducted a study titled 'Accelerating Multi-attribute Unsupervised Seismic Facies Analysis With RAPIDS' at the European Association of Geoscientists & Engineers, Subsurface Conference in Latin America (May 2021). Their research employed unsupervised learning, specifically k-means clustering, to divide the seismic section into eight clusters based on facies characteristics. Their findings align with our unsupervised learning approach, identifying distinct facies clusters: continuous, horizontal, low/high

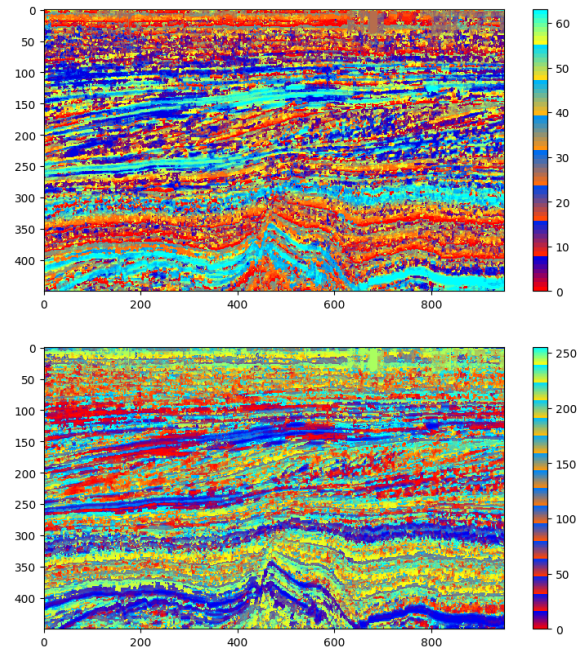
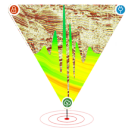


Figure 8: 64 (Top), 256 (Bottom) Cluster SOM Map

amplitude reflectors, and oblique areas. Our work extends these findings by utilizing a refined cluster selection approach, highlighting geological bodies through chaotic/random/low coherence seismic facies. This reveals gas-chimney (low-coherency extensional cluster) and salt dome (random facies clustered) structures, despite limited regional geological knowledge. This study demonstrates the effectiveness of unsupervised learning in enhancing seismic data interpretation for subsurface mapping and geological understanding.

On other hand, we also evaluated our unsupervised model alongside the supervised learning approach proposed by Silva et al. (2019) in their research paper 'Netherlands Dataset: A New Public Dataset for Machine Learning in Seismic Interpretation.' They trained deep learning models on the provided dataset for rock strata classification and seismic picture semantic segmentation. Their study focused on seismic reflection parameters, highlighting configuration patterns indicative of lithology, stratification, and depositional systems. Nine horizon intervals described the amplitude and continuity of seismic facies, serving as training data for their supervised deep learning model. Our results, depicted



in Figure 10, exhibit similar layer identification, reinforcing the efficacy of Self-Organizing Maps (SOM) for demarcating seismic sections into different facies classifications.

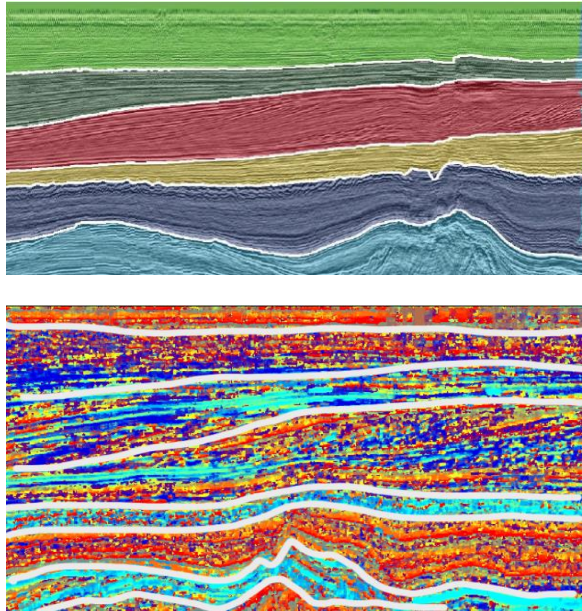


Figure 10: a) Result from supervised model used by Silva R et. al. 2019, b) SOM output by our algorithm, facies delineated with white marker.

Our study shows that SOM algorithm excels at clustering seismic facies, offering valuable insights into geological features and depositional settings. These results enhance seismic interpretation, subsurface mapping, and deepen geological understanding. Unsupervised learning improves interpretation and mapping, enabling enhanced analysis and high-resolution layer delineation. These advancements benefit geological mapping, reservoir characterization, and hydrocarbon exploration

Acknowledgment

The authors would like to Thank Rakesh R, Himanshu B, Akshay V, Anjali L, Jayanth B for their motivation and support. They would extend their gratefulness towards Petroleum Research Wing, GERMI for this project, especially P.H Rao. Authors are also thankful to ONGC, IOC for their resources to carry out the study. The views expressed in the paper

are of authors and may not necessarily be of ONGC, IOC or GERMI.

References

- [1] Marroquín I.D et al. A visual data-mining methodology for seismic facies analysis: Part 1 — Testing and comparison with other unsupervised clustering methods, 2017 GEOPHYSICS, VOL. 74.
- [2] Troccoli EB et al., K-means clustering using principal component analysis to automate label organization in multi-attribute seismic facies analysis, 2022 Journal of Applied Geophysics
- [3] Matos et al, Kurt J. Marfurt et al., Seismic Color Self-Organizing Maps, 2009 11th International Congress of the Brazilian Geophysical Society
- [4] Silva RM et. al, Netherlands Dataset: A New Public Dataset for Machine Learning in Seismic Interpretation, 2019 arXiv 1904.00770v1
- [5] Napoli O et. al, Accelerating Multi-attribute Unsupervised Seismic Facies Analysis With RAPIDS, 2020 arXiv:2007.15152v2
- [6] Zhaolin Zhu et. al, Seismic Facies Analysis Using the Multiattribute SOM-K-Means Clustering, 2022 Hindawi, 10.1155/2022/1688233
- [7] A Roy, M Matos, Kurt J.M, Automatic Seismic Facies Classification with Kohonen Self Organizing Maps - a Tutorial, Dec 2010/6 Geohorizons
- [8] Saraswat P, Mrinal K. Sen, Artificial immune-based self-organizing maps for seismic-facies analysis, 2012, Geophysics, Vol. 77
- [9] Chopra S, Kurt J. Marfurt, Seismic facies classification using some unsupervised machine learning methods, SEG International Exposition and 88th Annual Meeting 10.1190/segam2018-2997356.1
- [10] Coléou T et. al, Unsupervised seismic facies classification: A review and comparison of techniques and implementation, 2003 The Leading Edge
- [11] D.Subrahmanyam*,P.H.Rao, Seismic Attributes-A Review, 2008, SPG Conference.
- [12] Elmustafa S.Ali Ahmed et al. Median Filter Performance Based on Different Window Sizes for Salt and Pepper Noise Removal in Gray and RGB Images, 2015 10.14257/ijcip.2015.8.10.34
- [13] Riese, F.M., et. Al, 2020. Supervised and Semi-Supervised Self Organizing Maps for Regression and Classification Focusing on Hyperspectral Data. Remote Sensing 12(1). doi:10.3390/rs120100
- [14] SuSi Package Help Docs