The Automated Seismologist: An Application of Deep Clustering to Detect Events and Low Energy Non -Volcanic Tremor in Japan

LUKE CARDEW MACPHERSON

The Automated Seismologist

An Application of Deep Clustering to Detect Events and Low Energy Non - Volcanic Tremor in Japan

Luke Macpherson

Submitted in fulfilment for the degree Master of Science by Research (MScR) in Geological Sciences

Department of Earth Sciences

Durham University

December 2025

Preliminary

Abstract

The deep learning model, Deep Embedded Clustering (DEC), has been applied to the Japanese HiNet network to test its ability to identify and cluster different features in a seismic signal. The main focus of this work is to test whether DEC could (1) detect unknown singular patterns that may arise specifically before large or intermediate earthquakes; (2) identify low energy non-volcanic tremor (NVT) more efficiently than direct observation or statistical analysis of the seismic signal (e.g. root mean square or RMS) and (3) produce coherent clusters corresponding to other sources (anthropic activity, weather conditions). Training and testing on high passed data was successful at recognising day night perturbations globally across stations. For the NVT dataset, two clusters identified NVT: one could be verified visually, and the other correlated strongly with the RMS. Building on a previous application of DEC aimed at detecting and classifying anthropic activities in local array data, this thesis has broken new ground in adapting the method to the large scale HiNet network, and in testing DEC potential for seismic event detection and NVT discovery.

Contents

Declaration

Declaration: The model architecture directly follows on from research conducted by Dylan Snover at University of California San Diego and co-authors (Snover et al., 2021).

Statement of copyright

Acknowledgements

I would firstly like to thank my supervisor Professor Stefan Nielsen for supervision, and my PhD supervisor Dr Craig Warren for assistance during correction stages. I express my sincere thanks to my informal reviewers, Dr Andrew Valentine and Dr Jenny Jenkins. In addition, I would like to thank Dr Jenny Jenkins who was my undergraduate academic mentor and provided pastoral support throughout my time here in Durham. Further special thanks go to my girlfriend, Freya Alldred and Professor Darren Gröcke for their continued support and encouragement in research. Furthermore, Mr Fernando Andreu has been of invaluable help with all things Linux and Hamilton8 which has helped save me crucial time and stress – Fernando thank you! Finally, I would like to thank to my reviewers Dr Augustin Marignier and Dr Andrew Valentine.

Aim of the thesis

The automated model is the self-supervised Deep Embedded Clustering (DEC) model put forward by Xie at al., (2017) and expanded into earthquake source seismology (ESS) by Snover et al., (2021) and Mousavi et al., (2019). The model has been tested on seismic data in the time-frequency domain (spectrograms). This thesis aims to:

- Effectively detect unusual signals seen before a mega-thrust earthquake.
- Understand the nature of anthropogenic seismic signals.
- Detect low energy non-volcanic tremor (NVT) and determine whether the DEC is more effective than statistical or visual analysis.

Chapter 1 – Introduction

1.1 An introduction to earthquake source seismology and machine learning

The field of earthquake source seismology (ESS) aims to study the mechanics and occurrence of earthquakes. Understanding earthquake mechanics is reliant on observing seismic energy that is emitted before, during and after an earthquake in a seismic timeseries. Seismic timeseries contain multiple overlapping signals that are difficult to separate. Features include NVT, a certain form of human noise, or more unusual signals. Decomposing these timeseries down into features among different stations is difficult. However, machine learning (ML) can break down the seismic timeseries into distinct variables, especially when those variables are not known to the user.

This thesis presents results on the use of ML to analyse seismic timeseries to reveal more information about events and unusual signal types prior to earthquakes and during non-volcanic tremor (NVT) episodes. Presented is an application of Artificial Neural Networks (ANNs), a subset of ML, which mimics the architecture of a primitive brain. ANNs have been successful in completing ML tasks for several years in a range of fields, from computer vision to natural language processing (Bergen at al., 2019). Only recently has machine learning been routinely applied to ESS problems. There is a vast selection of algorithms in Table 1.1 that have been used to solve problems in ESS. ML problems broadly fall into categories of regression (separating 2 or more independent variables; output of a real number), classification (determining the discrete class of a label), clustering (separating observations into groups), noise removal and dimensionality reduction (reducing the number of variables an object can be described by). In seismology, picking events can be cumbersome (Hulbert et al., 2022), and understanding underlying physical mechanisms during NVT episodes is a challenge (Obara, 2002). This study will focus on two ML problems: clustering and dimensionality reduction. First, seismic data in the time-frequency domain will be compressed (dimension reduced) into a "latent vector" that is representative of relevant features within the spectrograms. Within the latent space, each spectrogram will be clustered for its various features. These two processes make up the model: Deep Embedded Clustering (DEC). This raises the question of whether the learned latent space provides a meaningful representation of the seismic data space, as a true seismic space only contains variables sensitive to features in the seismic timeseries (Valentine and Trampert, 2011).

1.2 Is deep learning a necessary application to earthquake source seismology (ESS)?

Data constraints

Seismic timeseries contains significant amounts of data. This can result in unrealistic operator time when processing and analysing datasets. In this study, data preceding the Iwate EQ exceeded 50Gb with the representation in the time-frequency domain exceeding 200Gb. The DEC model presented here compresses 5440 dimensional spectrograms into 14–16-dimensional latent space, *i.e.,* each instance is represented by 14-16 variables – a 99.7% compression in the number of variables. This is known as feature learning (FL). Mathematically, we compress the dataspace $\mathbb{D}$ into the latent space $\mathbb{L}$. As a result, the reduced data size is more accessible to additional ML. The memory and time needed to perform calculations in the latent space is a fraction of the dataspace allowing for deep clustering.

Ensuring data is clear and meaningful

Seismic datasets can contain a lot of noise including random, coherent and instrument noise (Saad et al., 2022). ML can aid denoising the data such that the signal to noise ratio is increased. The DEC model used in this thesis could serve as a denoiser if the spectrograms are converted back into the seismic timeseries.

Addressing subjectivity in seismic interpretation

As Ian Goodfellow stated in the book "Deep Learning", the advent of ML will "tackle problems involving knowledge of the real world and make decisions that appear subjective." Predicting earthquakes and attempting to find features in seismograms can be hard. An experienced seismologist will often learn features from a seismogram and use the learned experience to make interpretations of unseen instances (Valentine and Trampert, 2011). This is unfeasible for rapidly growing seismic networks and risks human error. A second option is to automate event detection. Some methods are deterministic: locating megathrust earthquakes or human induced events. However, tasks such as phase picking, precursor detection, ground motion and human induced noise are more subjective as these events are hard to label in seismic timeseries.

ANNs make subjective decisions based upon their experience and training. This usually results in a black box type model. However, this may not always be the case: Deep Neural Networks (DNNs) can adopt the architecture similar to the workflow of an inversion problem whilst each CL in a CNN can focus on a particular feature of an image (Bergen et al., 2019). Whether the steps taken by an autoencoder can be explained is unknown, but each embedded latent variable can be investigated in the time domain.

Knowledge space

There are many phenomena in earth sciences for which the physical mechanisms are poorly understood such as identifying earthquake precursors, and understanding the presence and characteristics of NVT (Bergen et al., 2019). A precursor is an event or accumulation of events that indicates an earthquake will immediately happen and can be easily disguised among other features. The field of ML, especially DNNs, presents the opportunity to extend the knowledge and function space beyond human capacity. Bergen et al., (2019) describe ML as the "automation to perform a complex prediction task that cannot easily be described by a set of explicit commands" whilst opening the opportunity to "reveal new and often unanticipated patterns, structures, or relationships." This is especially applicable to ESS, where many signals are convoluted upon one another (e.g. Snover et al., 2021) and more unusual signals may exist. In particular, the unsupervised training scheme helps avoid any preexisting biases that may arise when labelling data.

Furthermore, physical mechanisms, especially regarding NVT, are not entirely understood. The relationship between NVT and slip is still not known (Hulbert et al., 2022). Current non-ML methods may not detect low energy NVT that get disguised in the noise, which could benefit further analysis into the physical mechanisms, such as chatter along the Cascadia subduction zone identified by Rout-Leduc et al., (2019).

Lack of ground truth

The field of ESS contains a lack of ground truth and is reliant on seismic waves, geodetic data and remote sensing to interpret behaviour along faults (Obara, 2010). Seismic data is affected by many features which could be decomposed during ML.

Speed and efficiency

Many problems in geophysics, such as geophysical inversion, probabilistic forecasting for early warning (EEW) systems require large computational power due to inefficient algorithmic time complexities and urgency in hazard prediction respectively (Kauful et al., 2016). Once trained, ML algorithms are extremely efficient: the prediction of a NN can take seconds. However, NN training is long. In this study, average training times were approximately 200 CPU hours of a 256Gb RAM Hamilton 8 node. Likewise, Valentine et al., (2013)'s application of an AE into classifying seamounts took several hundred CPU hours (Valentine and Kalnins, 2016). In summary, an NN is appropriate to solve problems with large dataspaces.

1.3 Progress with Machine Learning in Earthquake Source Seismology

A large number of ML studies exist to overcome problems encountered in 1.2. There have been many attempts to identify and cluster features in the seismic timeseries. As presented in Table 1.1, all the studies hitherto known to the author have been summarised by the model and ML task. There are two common themes: the models and learning schemes used are diverse, yet they are dominated by the desire to cluster and classify events with respect to time over other ML problems. Methods such as dimensionality reduction and noise removal usually complement classification/clustering problems (e.g. Ong et al., 2020; Snover et al., 2021; Johnson et al., 2020).

Here, studies that apply ML to ESS are separated into the 3 different types of learning: supervised (the model is given a human labelled dataset to test itself in training), un-/self-supervised learning (the model is not given any labels in training) and transfer learning (a model trained in one dataspace and tested in another dataspace *ex situ*).

Supervised learning models

The supervised learning models rely on ground truth or the "assumption" of ground truth. Rouet-Leduc et al., (2019) managed to train a supervised random forest algorithm to predict GPS movement from seismic data as a result of NVT. Ong et al., (2020) applied CNNs to classify seismic timeseries as precursory upon the assumption that timeseries before and after represents noise and precursor activity respectively – it was able to identify a low energy signal around 2Hz that was statistically precursory. Additionally, there have been applications of recurrent neural network (RNN) – networks that utilise repetition among neurons such as an LSTM model which was used to discriminate explosions from earthquakes (Magana & Ruppert, 2017). As seen, supervised ANNs have been applied in many diverse areas of earth science.

Self/un supervised models

Self-supervised learning is a subset of unsupervised learning. Snover et al., (2020) applied a DEC model, trained using self-supervised learning, to cluster anthropogenic noise in California (see Section 1.5). The architecture of this model forms the basis for this research. Completely unsupervised models such as that by Johnson (2020) successfully clusters/classifies noise of 1 second waveforms applying statistical parameters in the waveform to KMeans clustering. It was successful in identifying anthropogenic noise (day/night perturbations).

Transfer learning

Neural networks, once trained, can be applied to different datasets known as transfer learning. For instance, Jozinovic et al., (2022) utilised transfer learning to detect ground motion intensity during Mw > 3

earthquakes using CNNs in Italy. Due to a lack of data they managed to apply a pre trained CNN from *ex situ* data to their current study location and found that it improved detection in comparison to the CNN that was trained on *in situ* data. Wang et al., (2021) predicted fault slip from transfer learning. Additionally, Wang et al., (2021) utilised synthetic laboratory acoustic emission data and used that to train a CAE to predict evolutions of stress. They then applied this to real laboratory data with successful results. In this research, the model was trained on data from the HiNet network in Japan. It is hoped that any follow-on research will implement the trained models in other regions.

	Supervised Learning		Unsupervised learning				
Class of seismic data / ML model	CNN	MDS	AE	(C)VAE	KMC	DEC	Other models
Miscellaneous features						Macpherson et al., (2024) (this study) Jenkins et al., (2021)	Steinmann et al., (2022).
NVT	Hulbert et al., (2022)					Macpherson et al., (2024) (This study)	
Anthropogenic					Johnson (2020)	Snover et al., (2020)	
"Precursor" detection	Ong (2020)						Seydoux et al., (2020)[1]
Teleseism detection						Mousavi et al., (2019)	
First motion polarity						Mousavi et al., (2019)	
Ground motion	Jozonovic et al., 2021			Ning & Xie (2023)			
Microearthquake /earthquake detection		Dzwinel et al., (2005)					
Fault slip			Wang et al., (2021)				
Denoising			Saad et al., (2022)				

Table 1.1: [1]Including GMM with Convolutional layers,
[2]hierachical clustering within a deep scattering network
Red = classification task, Blue = clustering task, Green = dimensionality reduction, Purple = regression.
These are interpreted from written hypotheses, NOT the algorithm they use. Italics are experimental or theoretical studies.

1.4 How this study will take the research forward

In the last 60 years, efforts to find precursors in Japan have been poor with a good proportion of investment and effort going into strengthening and expanding networks. The scale of mega-thrust EQ predictions existing is in the order of the decade - usually informed by the EQ cycle (Uyeda, 2013). Uyeda (2013) suggests that effort should be made into finding non-seismic precursors, this thesis stresses the importance of utilising AI to expand the knowledge space of seismic timeseries and investigate whether there is a statistical

link between known and unknown phenomena (such as NVT seen by Hulbert et al., 2022), EQ and NVT. If not, additional discovery of events could aid future regression tasks that incorporate seismic and non-seismic phenomena to establish precursory criteria, such as finding low energy NVT. This thesis will test the DEC's ability to cluster events in the HiNet dataset. HiNet provides good coverage over one of the most tectonically active areas in the world which contains large sources of noise from urban metropolises. This provides a perfect mix to test the clustering model in the hope it will discover known and unknown features whilst highlighting varying sources of noise.

Overview of DEC usage

The DEC is a form of self-supervised learning model where, instead of being provided with human labelled data, creates its own labels during learning from the reconstructed dataspace. This is not to be confused with semi-supervised learning, where the model is provided with a subset of human labelled data in training. As stated, the model has been successful in a range of tasks: Snover et al., (2021) tested it on an extremely dense dataset on Long beach California to cluster noise, whereas Mousavi et al., (2019) identified teleseisms and first motion polarities. It has also been applied in bioacoustics of when analysing whale calls with additional Gaussian Mixture Models for clustering (Ozanich et al., 2021). Its ability to cluster impulsive environmental features along an ice shelf was investigated by Jenkins et al., (2021) who demonstrated the DECs ability to cluster tidal and weather signals. These studies proved that the DEC can identify a diverse range of complex acoustic phenomena, which suggests it may be able to extract more sophisticated features from the HiNet dataset. This thesis will test the first 31.25 days before an earthquake off the coast of Iwate in Japan, and it will also test whether the model can be scaled up to a network less dense than in the Snover et al., (2021) study. In addition, inspired from studies into identifying and investigating NVT using ML (Johnson, 2020; Hulbert et al., 2022), the DEC will be trained on known NVT episodes in SW Japan (Obara, 2002). NVT, which is identified by the processed RMS signal, is continuous in time. In comparison to classification methods used in Hulbert at al. (2022) etc., this thesis will test whether clusters, plotted over a moving average in the temporal domain, can be cross correlated with the RMS signal - the first time this has been done with the DEC.

1.5 The study this thesis is based upon

This study is based upon the original study from Snover et al., (2021) that used the DEC to cluster anthropogenic signals at Long Beach California.

Their study used 5200 vertical component geophones and spectrograms from 6 days of data in March 2011. The first 10 minutes of each hour were converted into 4 second slices with 3 seconds overlap. The spectrograms were originally cropped from 5200 features to 4800 features to make them compatible for training. Training consisted of two parallel processes: training the convolutional autoencoder model by mapping to lower dimensional latent representation whilst training the DEC: the clustering layer that separates each spectrogram into their respective classes.

Their model was able to identify known classes of noise, such as the vibriosis truck (linear sweep in the spectrograms), and other forms of noise not previously known. For instance, some clusters were able to identify activities such as nearby industrial rotating machinery, and nearby oil extraction. These signals are clearly distinguished in the time-frequency domain.

It was noted that other forms of urban noise, such as highway traffic, caused constructive interference in the spectrograms. The DEC could not identify these distinctive signals in the latent space and that KMeans clustering initialisation was limited by varying sample sizes among clusters and different variance. In this thesis, the variance and sample sizes for the clusters are not known *a priori*. Therefore, this study will remain with KMeans clustering.

Snover at al., (2021) stresses that the choice of window length is meta parameter that can be varied to prioritise different features. Given this thesis aims identify features of varying lengths, window size has been made a hyperparameter.

The use of multi station data helps identify trends across the network. This model has been chosen as it works on the principle that there are core seismic components in the network which exist in the "latent space", and the rest of the components are noise. These seismic variables are known but also not known to the user, which makes an unsupervised learning regime particularly promising to solve the problem at hand as it avoids any preexisting biases that would be incorporated into the model. Additionally, using windows of data in the time-frequency domain allows the model to distinguish between classes of seismic signal that may vary as a function of frequency, as seen in this thesis with identifying unknown events and NVT.

Chapter 2 – Theory and methodology

2.1 The mathematics behind ANNs

In simple terms, neural networks or artificial neural networks (ANN) comprise of nodes or neurons that reside in interconnected layers. A neuron corresponds to a vector of numbers. In a fully connected neural network, each neuron in a layer is connected to all the neurons before and after the layer. The input layer of a neural network must fit the shape of input data whilst the shape of the output layer can vary upon the ML task. An ANN becomes a deep ANN (hereafter DNN) when there are multiple hidden layers in-between the input and output. They are "hidden" because the layers are not given information about the inputs and outputs. The network presented in this thesis is deep, corresponding to more than one hidden layer. Each neuron is bound by a weight and bias (commonly known as "parameters", Equations 2.1, 2.2, and 2.4) which are altered during training (section 2.7). The resulting number is mapped into an activation function (here ReLU) to simulate a neuron "firing". In a neural network, this corresponds to the transfer of data simulating a somewhat-binary "on" "off" state which adds non-linearity helping the DNN to learn complex patterns adding a level of subjectivity to its predictions. This informs the weights and biases of the neurons succeeding (forward propagation training) and preceding it (back propagation training).

$$\boldsymbol{w}.\mathbf{x}_{\boldsymbol{n}} + b \rightarrow f_v \tag{2.1}$$

A neuron mapped onto another neuron. X is the neuron of a layer (n), w corresponds to the associated weights and b corresponds to the neuron's internal bias (constant). The resultant scalar is mapped into the activation function f_v and combined with all the other values from neurons which are fed into the next activation function (for a fully connected network (FCNN)) demonstrated in equation (2.1).

$$\bar{\mathbf{x}}^{(n)} = f\left(\begin{bmatrix} w_{0,0} & \cdots & w_{0,n} \\ \vdots & \ddots & \vdots \\ w_{k,0} & \cdots & w_{k,n} \end{bmatrix}.\begin{bmatrix} \mathbf{x}_0 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}\right) \rightarrow \bar{\mathbf{x}}^{(n+1)} \tag{2.2}$$

A layer with n neurons: Expanded form of equation along with the corresponding weights and associated bias in each layer for a neuron of dimension k.

Convolutional neural network

A CNN is an expansion upon a typical ANN. Used typically for image classification, CNNs utilise the process of convolution in subsampling parts of the image or matrix. If every sample was processed independently of the image, the amount of time to train the network would be an order of magnitude more. Hence, convolution is utilised to extract features from the image by applying learned filters. Sub sampling this sample helps the network become invariant to translation such that a feature in the image can always be recognised regardless of its position. The size of the sub sample is known as the receptive field. The kernel filter (KF) extracts data within the receptive field. This filter varies in size. In the 2-dimensional domain, it is usually a 3x3 shape. This parameter is a hyperparameter balancing detail (small size) with breadth (large size) in data. This kernel (ϕ) performs convolution, defined in the discrete real domain in the bound [-a, a] as:

$$g(x) = [f * \phi] = \sum_{x=-k}^{x=\mathrm{k}} \phi(\Delta x) f(x + \Delta x);\ x, a \in \mathbb{R} \tag{2.3}$$

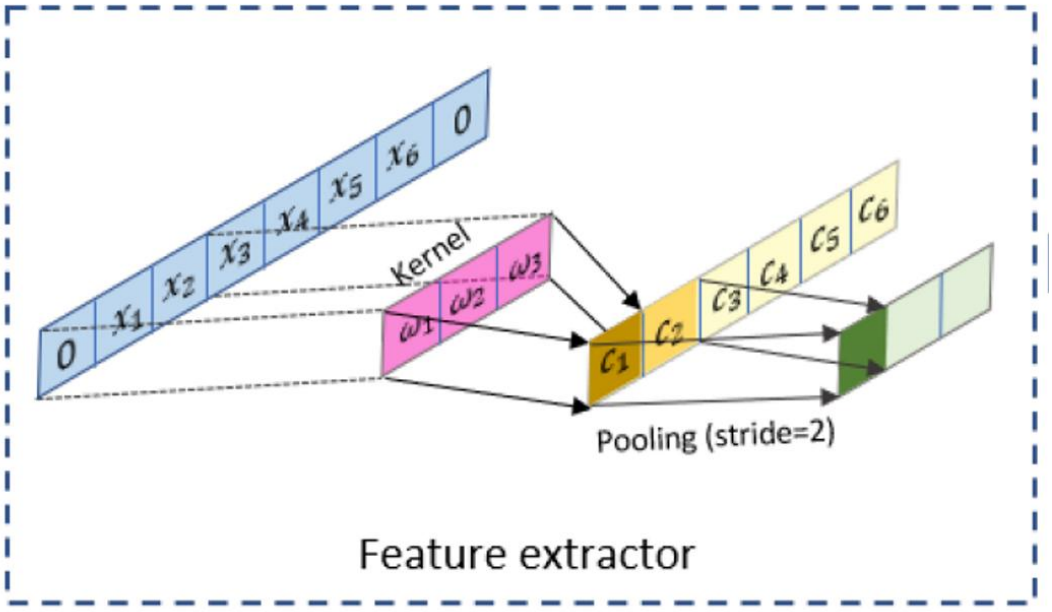


Figure 2.1: A 1-dimensional representation from Ong et al., (2020). In this figure, zero padding has been applied.

For vector, **x,** padding can be applied at the boundaries to allow the kernel to span the entire vector. Without padding, the kernel would not be able to cover to the edges of the RF.

The stride of the kernel determines the gap in which the KF spans the RF. A large stride allows a greater coverage in the neural network but conversely risks eliminating fine details, whilst a smaller stride risks overfitting. Here, a stride of 3 has been chosen to optimise catching details whilst avoiding overfitting.

Pooling / feature layers

Convolutional neural networks utilise feature pooling between the hidden layers as a dimension reduction tool. In theory, each layer of a convolutional neural network should correspond to a feature in the image. Usually, each successive layer targets more complex features in the dataset. For instance, the first layer represents simple images such as edges and the last layer represents minute details.

How much the pooling layer reduces the dimensions of the input section is usually determined by hyperparameter depth. Pooled data (pooling function = ρ) is used to create a feature map that is mapped onto the next hidden layer such that for depth = d,

$$f\left(\begin{bmatrix} w_{0,0} & \cdots & w_{0,n} \\ \vdots & \ddots & \vdots \\ w_{k,0} & \cdots & w_{k,n} \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}\right) \rightarrow \rho \rightarrow f\left(\begin{bmatrix} w_{0,0} & \cdots & w_{0,\frac{n}{d}} \\ \vdots & \ddots & \vdots \\ w_{\frac{k}{d},0} & \cdots & w_{\frac{k}{d},\frac{n}{d}} \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{\frac{n}{d}} \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_{\frac{n}{d}} \end{bmatrix}\right) \rightarrow \rho \qquad (2.4)$$

Arbitrary hidden layer (n) Arbitrary hidden layer (n+1)

The pooling function, ρ, can vary based upon the problem one is trying to solve. Examples are average pooling, such that the average layer is taken across the receptive field or max pooling such that the maximum value is taken across the receptive field. The former allows a generalisation of certain large-scale features whereas the latter allows the network to pinpoint on minor details.

2.2 The deep convolutional autoencoder (DCAE)

A deep neural network is one that consists of more than two hidden layers. Here, instead of classifying images, a CNN (comprised of 2d convolutional layers) is used to compress and reconstruct image data into a n-dimensional latent space. The process of convolution reduces the number of parameters that are needed. A dimension is the extent to which something can be measured. Here, we define the dimension of input data as each sample or pixel (Valentine and Trampert, 2012).

When decoding, a convolutional transpose neural network (CTNN) is used to reconstruct the image data. A CTNN is essentially the inverse of a CNN: it utilises deconvolution in up-sampling the data. It is still defined by usual parameters such as padding or stride.

In our example here, we compress images comprising of 140 frequency bins and 40–41-time bins totalling 5440 dimensions (corresponding to each bin) into 14-16 latent dimensions.

2.3 KMeans clustering

KMeans clustering is an unsupervised clustering algorithm, used here to initialise clusters in the latent space produced by the DCAE. As demonstrated in Figure 3.2, it is a simple yet effective process of clustering data in n-dimensional space.

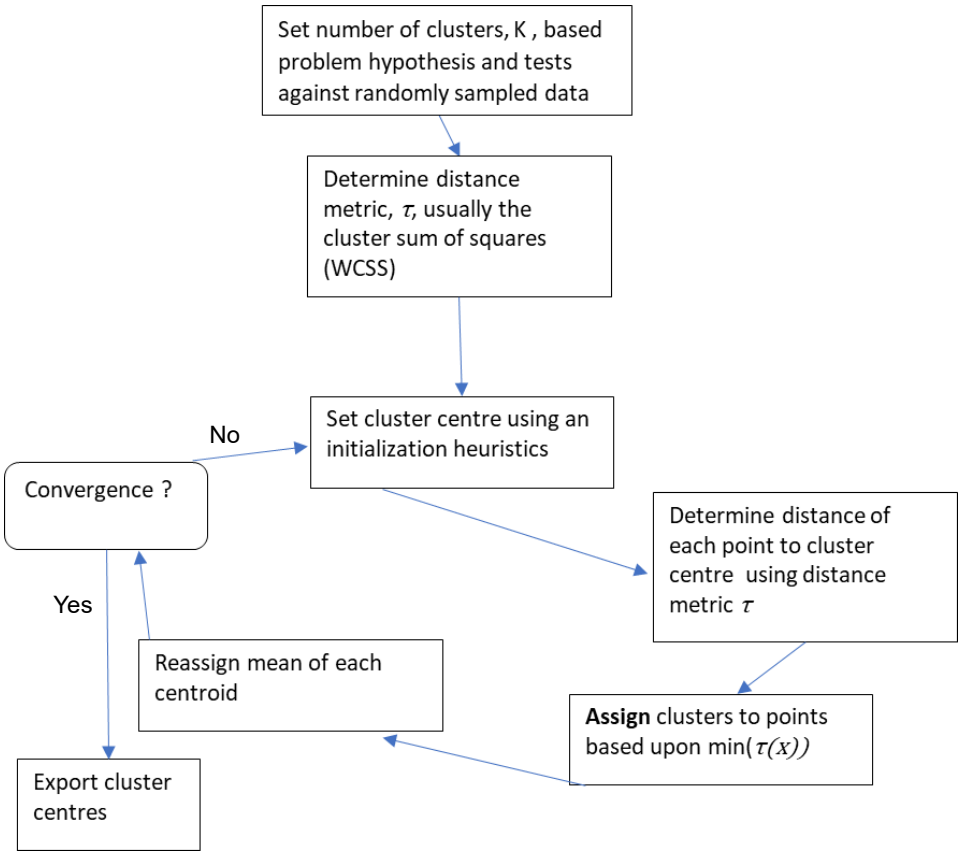


Figure 2.2: simplified KMeans clustering process. It demonstrates the iterative process of moving cluster centres from minimising the distance of data points to cluster.

2.4 Combining deep autoencoder and KMeans clustering: Deep Embedded Clustering

The DEC was proposed by Xie et al. in 2016. It combines DL and clustering algorithms to perform unsupervised ML tasks. Deep clustering was chosen because it is an unsupervised method, which is essential

for discovering unknown patterns in seismic data that lacks pre-existing, human-provided labels. It comprises of a DCAE (returning a reconstructed image) with an additional clustering layer with cluster centres initialized by KMeans for hyperparameter: number of clusters (*K)* (Figure 2.3). KMeans assigns each point a definitive cluster – known as "hard clustering". It also assumes the variance to be the same in each dataset cluster (Snover et al., 2021). It would prove highly unstable if there were different data variances in each cluster and there were clusters that were highly mixed in the latent space. Additionally, it will stop local convergence in the 14-16 dimensional latent space used in this thesis. Instead, we can use KMeans to initialize clusters and then iteratively assign "soft labels" to each datapoint such that each point is given a probability of being assigned to a cluster centroid to allow for greater exploration of the latent space. Gaining inspiration from VAEs, each point in the latent space is mapped to a soft label within the Student's T generating soft assignments q_{ij} (Equation 2.5)

$$q_{ij} = \frac{\left(1 + \left\|\boldsymbol{z}_i - \boldsymbol{c}_j\right\|^2\right)^{-1}}{\sum_j^K \left(1 + \left\|\boldsymbol{z}_i - \boldsymbol{c}_j\right\|^2\right)^{-1}} \tag{2.5}$$

Zi is the point in the latent space, q_{ij} represents the probability of z belonging cluster c_j.

This is then compared with the target distribution, p_{ij} with f_j is the soft cluster frequency:

$$f_j = \sum_i^{\mathrm{n}} q_{ij} \tag{2.6}$$

$$p_{ij} = \frac{q_{ij}^{\,2}}{f_j} \Bigg/ \sum_j^{\mathrm{K}} \left(\frac{q_{ij}^{\,2}}{f_j}\right) \tag{2.7}$$

The soft assignment represents, for each cluster *j*, a probability of each point being assigned to the cluster. Conversely, the target distribution Q represents the probability of each cluster j being assigned to each point. This outputs two distributions: the target distribution, P and the soft assignment Q. In an ideal and stable system, $q_{ij} = p_{ij}$ (Equation 2.6 and 2.7). However, due to complexities of samples in the latent space, this is unlikely. Kullback-Leibler Divergence (KLDiv) describes how two different probability distributions (here distribution P sampled by p_{ij} and Q sampled by q_{ij}) differ (Equation 2.7).

$$\mathrm{KL}(P||Q) = \sum_i^n \sum_j^{\mathrm{K}} p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{2.8}$$

The KLDiv equals 0, it is a high confidence sample. The KL divergence is used to adjust the cluster centroids in training. However, each point will be given a probability distribution of its chances of being assigned to each centroid.

In summary, KMeans clustering initially assigns each spectrogram to a cluster. Soft clustering then adjusts these initial clusters during the training process.

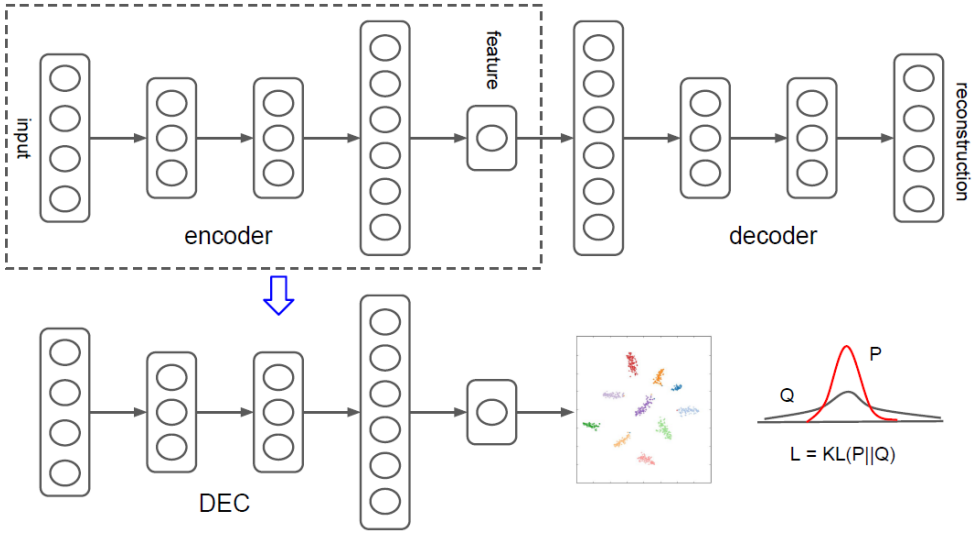


Figure 2.3 : CAE and DEC architecture given in Xifeng et al., (2017). Both the CAE and DEC are trained simultaneously to gain spectrogram reconstructions and cluster centroids.

2.5 The network parameters

Hyperparameter tuning

Hyperparameters were varied to optimise the network. From a literature review and consideration of time constraints, the following hyperparameters and core features in the network were kept same as the Snover et al., (2021) study (Table 2.1).

Hyperparameter	Symbol	Set to	Justification
Strides	s	3	A balance between detail and breadth. Same as Snover et al., (2020).
Depth (for pooling)	d	Powers of 2	The dimensions of the spectrograms being used are the same.
Epochs	E	400	Loss/cost become negligible after a number of iterations.
Learning rate	l	1.0001	The learning rate has been experimentally analysed across multiple studies: Snover et al., (2021),
Padding	P	Either "valid" or "same"	This architecture was successful in the Snover et al., (2021) study.
Time bins	x_t	136	Successful parameters used in Snover et al., (2021). Too many parameters risks the model underfitting and potentially causes the problem to become ill-posed.
Frequency bins	x_f	40	Same as above
Loss adjustment parameter	λ	1.1	Proved successful in Mousavi et al., (2019) and Snover et al., (2021)

Table 2.1: A description of the core hyperparameters kept constant in this thesis to optimise the DEC. Note: "valid" is without padding such that a dimension reduction is applied at the convolutional limits, "same" is with padding such that 0s applied at convolutional limit.

2.6 Training specific to the DEC network

An ML algorithm is defined to learn from experience E**,** with respect to a tasks **T** and performance measure **P,** if its performance in tasks **T**, measured by **P**, improves with experience **E** (Mitchell, 1997). The training of the DEC model must satisfy this criterion with the class of tasks, **T** consisting of 1) compress and reconstruct and 2) cluster data. The performance metric **P** is described by a loss function which combines the difference (MSE) in 1) compressing and reconstructing spectrograms and the 2) KLDiv from clustering spectrograms (summarised in Table 2.2). The lambda hyperparameter determines the influence of 1) and 2) in the total loss function, described below:

$$L_R = \frac{1}{n}\sum_{i}^{n}(\hat{x_i} - x_i) \tag{2.9}$$

$$L = (1-\lambda)L_R + \lambda L_c \; ; \; \lambda \in [0,1] \tag{2.10}$$

In consideration of memory, neural networks are usually trained in batches. An epoch is when all batches of the training dataset have been processed by the neural network.

2.7 The training process

Forward propagation

The weights and biases are randomly initialised in a neural network.

Forward propagation, is a process by which the inputs of a neural network are converted into outputs, as introduced in the first section, and can be summarised in simple notation as:

$$\mathbf{x}^{(n)} = f\left(W\mathbf{x}^{(n-1)} + b\right) \tag{2.11}$$

Backward propagation

The loss or cost (hereafter loss) is a measure of the performance of the network summarised as a difference between two quantities. In back propagation, gradient descent is often used to adjust weights and biases defined as -∇L where ∇L is the gradient of the loss function with respect to each parameter. The DEC utilises the ADAM optimiser. ADAM is an extension on standard gradient descent: it combines momentum (which accelerates descent in the relevant direction) and RMSProp (which adapts learning rates based on the variance of the gradients) to navigate the loss landscape more effectively and avoid stagnation in local minima.

The autoencoder

Autoencoders are a dimensionality reduction tool used to compress and find patterns in a dataset. First an encoder, *E,* performs a dimensionality reduction on a given dataset contained in dataspace $\mathbb{D}$, preserving key features and patterns in dimensionally reduced latent space $\mathbb{L}$ comprising of latent vectors (*l(x)*) of size *m*. From this latent space, it will then reconstruct the original vector or tensor data using the decoder, *D*. For vector $x \in \mathbb{R}^n \in \mathbb{D}$ of input size *k* and reconstruction size *k',* along with trained parameters, [*w, b*] $\in \theta$, an autoencoder is defined:

$$x \rightarrow E(x) \rightarrow l(x) \rightarrow D(x) \rightarrow \hat{x} \quad (2.12)$$

In an ideal autoencoder, it applies back propagation to adjust parameters θ to minimise the reconstruction error, such that the output reconstruction spectrogram approximates the input spectrogram. This does not always happen however and the compression of data of in an autoencoder essentially falls to 3 classes in which the data is lossless, acceptably lossy, or unusable (Valentine & Trampert ,2012). The size of the latent space, between 14 and 16 dimensions, was guided by Snover et al., (2021) to be within the lossless to acceptable lossy range.

In summary,

Class	Parameter
E	Units of "epoch"
T	1) Reconstruction of spectrograms 2) Clustering of spectrograms
P	Loss function 1) MSE 2) KLDiv

Table 2.2: outlining the fundamental ML tasks, experience and performance metrics in this thesis.

2.8 A note on hardware

All computation has been processed and stored on the high performance computer Hamilton8 with 128CPU cores and 256Gb of random-access memory per node.

2.9 The HiNet Network of Japan

The HiNet network is a dense 3 component array in Japan formed following the disastrous Kobe Earthquake in 1996 to improve detection of micro seismic activity (Obara and Hirose, 2006). Station separation and depth 15-20km and 100m on average. Hi-Net uses a moving coil velocity type seismometer (https://seisman.github.io/HinetPy/appendix/response.html). The sampling frequency is 100Hz, meaning the highest frequency signal identifiable is 50Hz, according to the Nyquist-Shannon theorem.

2.10 Preprocessing

Preprocessing (Figure 1e) involves preparing the data before training and testing the network. It is an essential component in ML to ensure the dataspace is fully explored.

Distribution

The aim of unsupervised ML is to automate detection of unlabelled events/objects/patterns. Some patterns are known and could be detected already by traditional processing (but exist as a singularity in large data sets), or they could be unknown or too subtle for easy detection. Therefore, data is evenly distributed to the network to avoid bias. This is done by randomly shuffling the time chunks in the data set. To maximise randomness, data was shuffled twice. From Snover et al., (2021), data is randomly ordered into chunks with 80% being used for training whilst the other 20% used for testing.

Within each chunk, data is randomly ordered into spectrogram lengths of either 120 and 240 seconds.

Filtering

In certain trained models, discussed in Chapters 2 and 3, the dataset was filtered. A digital Butterworth filter was used:

	Value	Justification
Filter	Butterworth filter	Butterworth filters are commonly used as they have a flat pass with a monotonic decreasing gain after the cutoff frequency, meaning they are less likely to introduce artefacts and ringing into the data.
Order	3	Eliminates the possibility of ringing around the 2Hz mark especially given a the high passed data is dominated in the low band.
Pass	High	We want to eliminate the dominant low frequency band.
Cutoff	2Hz	There is a high-amplitude low frequency (<1Hz) background signal.

Table 2.3: Justifications for parameters in the Butterworth filter.

Scaling

Each point in the spectrogram is normalised between [-1,1]. Normalisation ensures that high amplitude signals are dampened and prevents the network becoming biased toward them. Therefore, the network is able to detect low amplitude features, such as NVT that is disguised among noise. Maximum normalisation was first experimented with and then l2 normalisation was applied which allows fairer representation of the signal and was applied to each chunk. L2 and Max normalisation, for a 1-dimensional signal, are respectfully described as:

$$\|x\| = \frac{x}{\sqrt{\sum_{i=1}^{k} \mathbf{x}_i^2}}; x \in \mathbb{R}^k \tag{2.13}$$

$$\|x\| = \frac{x - \bigwedge x}{\bigvee x - \bigwedge x}; x \in \mathbb{R}^k \tag{2.14}$$

Where $\vee$ and $\wedge$ are maximum and minimum values respectively, k is the number of data points, x is value being normalised, x is the vector, and $\|x\|$ is the normalised value.

Representation

Spectrograms are created to represent the data in the time-frequency domain. The scikit learn Kaiser spectrogram has been used with the same parameters that proved successful in the Snover et al., (2021) paper with a high overlap of 88% utilising a Kaiser window with a beta value of 5.6. The high overlap ensures that different features are evenly distributed throughout the spectrogram and utilises a CNNs invariance to translation (Ong et al., 2020).

Standardisation

Each spectrogram is standardised by their mean and standard deviation using Tensorflow Image Data Generator class. This prevents bias in the network.

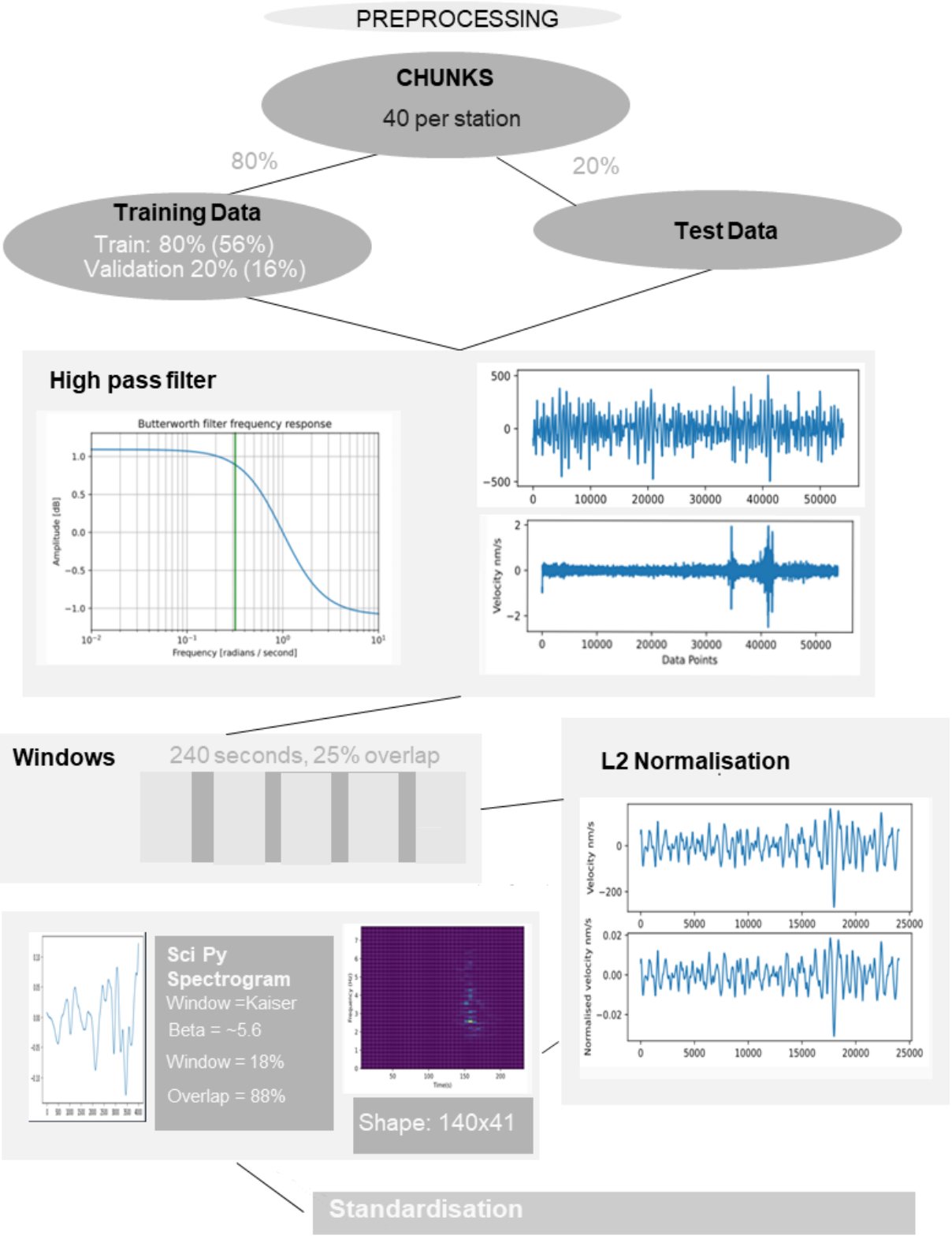


Figure 2.4: summary of preprocessing. The high pass filter was applied in the latter half of chapter 2. Note the Sci Py spectrogram uses a Kaiser window with beta of 5.6%, window parameter 18% and window overlap of 88%.

The following hyperparameters were varied:

Hyperparameter	Symbol	Justification
Latent space size	$L^n; n \in [14,16]$	Latent space of 14 used in Snover et al., (2021). Expanded to 16 in this

		study to test the effect of additional variables.
Duration of the spectrogram	P	Either 120 seconds or 240 seconds to capture a range of frequencies.
Number of clusters	K	Between 4 and 6 dependant on the clustering class.
Normalisation	$\hbar$	Experimented with l2 and maximum normalisation.

Table 2.4: varied parameters with the name of the hyperparameter, the symbol assign and justification as to why these hyperparameters are varied over others.

2.12 Analysis of results (methodology)

Reconstructions

Each reconstruction is visually analysed in frequency and time. This is a quality control measure to ascertain what is modelled as either "anomalous" and "noise" in each spectrogram and gives a good indication of DEC performance.

Cluster analysis

Each cluster is analysed for specific features in the time domain. Cross correlation metrics are calculated between suspected features (e.g. day/night) and cluster occurrences and tests whether the signal has a similar wavelength and constant phase difference with a feature, such as the number of recorded earthquakes of a certain magnitude in that time. The Pearson Correlation Moment (PCM) has been chosen:

$$\rho_{X,Y} = \frac{cov(X.Y)}{\tilde{n}\,\sigma_X \sigma_Y} \tag{2.15}$$

Where X and Y are two random variables given as a function of time. They are standardised by their mean values and normalised by their maximum value. $\rho_{X,Y}$ is calculated from the dot product of X and Y normalised by the product of sample size $\tilde{n}$ and their respective standard deviations (σ).

This allows a singular value to be returned for each cross correlation. If it is negative, the signal is out of phase. If it is 1, it is in phase and -1, completely out of phase. This allows one to conduct a statistical test to reject the null hypothesis that the cluster and selected feature do not correlate. The selected confidence interval will be a two-tailed (to account for in phase and out of phase signal), 5% as it allows room for noise in the cluster occurrence but ensures statistical rigor that is needed in ML (Table 2.5).

df \ α	0.2	0.1	0.05	0.02	0.01	0.001
1	0.951057	0.987688	0.996917	0.999507	0.999877	0.999999
2	0.800000	0.900000	0.950000	0.980000	0.990000	0.999000
3	0.687049	0.805384	0.878339	0.934333	0.958735	0.991139
4	0.608400	0.729299	0.811401	0.882194	0.917200	0.974068
5	0.550863	0.669439	0.754492	0.832874	0.874526	0.950883
6	0.506727	0.621489	0.706734	0.788720	0.834342	0.924904
7	0.471589	0.582206	0.666384	0.749776	0.797681	0.898260
8	0.442796	0.549357	0.631897	0.715459	0.764592	0.872115
9	0.418662	0.521404	0.602069	0.685095	0.734786	0.847047
10	0.398062	0.497265	0.575983	0.658070	0.707888	0.823305
11	0.380216	0.476156	0.552943	0.633863	0.683528	0.800962
12	0.364562	0.457500	0.532413	0.612047	0.661376	0.779998
13	0.350688	0.440861	0.513977	0.592270	0.641145	0.760351
14	0.338282	0.425902	0.497309	0.574245	0.622591	0.741934
15	0.327101	0.412360	0.482146	0.557737	0.605506	0.724657
16	0.316958	0.400027	0.468277	0.542548	0.589714	0.708429
17	0.307702	0.388733	0.455531	0.528517	0.575067	0.693163
18	0.299210	0.378341	0.443763	0.515505	0.561435	0.678781
19	0.291384	0.368737	0.432858	0.503397	0.548711	0.665208
20	0.284140	0.359827	0.422714	0.492094	0.536800	0.652378
21	0.277411	0.351531	0.413247	0.481512	0.525620	0.640230
22	0.271137	0.343783	0.404386	0.471579	0.515101	0.628710
23	0.265270	0.336524	0.396070	0.462231	0.505182	0.617768
24	0.259768	0.329705	0.388244	0.453413	0.495808	0.607360
25	0.254594	0.323283	0.380863	0.445078	0.486932	0.597446
26	0.249717	0.317223	0.373886	0.437184	0.478511	0.587988
27	0.245110	0.311490	0.367278	0.429693	0.470509	0.578956
28	0.240749	0.306057	0.361007	0.422572	0.462892	0.570317
29	0.236612	0.300898	0.355046	0.415792	0.455631	0.562047
30	0.232681	0.295991	0.349370	0.409327	0.448699	0.554119

df \ α	0.2	0.1	0.05	0.02	0.01	0.001
35	0.215598	0.274611	0.324573	0.380976	0.418211	0.518898
40	0.201796	0.257278	0.304396	0.357787	0.393174	0.489570
45	0.190345	0.242859	0.287563	0.338367	0.372142	0.464673
50	0.180644	0.230620	0.273243	0.321796	0.354153	0.443201
60	0.164997	0.210832	0.250035	0.294846	0.324818	0.407865
70	0.152818	0.195394	0.231883	0.273695	0.301734	0.379799
80	0.142990	0.182916	0.217185	0.256525	0.282958	0.356816
90	0.134844	0.172558	0.204968	0.242227	0.267298	0.337549
100	0.127947	0.163782	0.194604	0.230079	0.253979	0.321095
125	0.114477	0.146617	0.174308	0.206245	0.227807	0.288602
150	0.104525	0.133919	0.159273	0.188552	0.208349	0.264316
175	0.096787	0.124036	0.147558	0.174749	0.193153	0.245280
200	0.090546	0.116060	0.138098	0.163592	0.180860	0.229840
250	0.081000	0.103852	0.123607	0.146483	0.161994	0.206079
300	0.073951	0.094831	0.112891	0.133819	0.148019	0.188431
350	0.068470	0.087814	0.104552	0.123957	0.137131	0.174657
400	0.064052	0.082155	0.097824	0.115997	0.128339	0.163520
450	0.060391	0.077466	0.092248	0.109397	0.121046	0.154273
500	0.057294	0.073497	0.087528	0.103808	0.114870	0.146436
600	0.052305	0.067103	0.079920	0.094798	0.104911	0.133787
700	0.048427	0.062132	0.074004	0.087789	0.097161	0.123935
800	0.045301	0.058123	0.069234	0.082135	0.090909	0.115981
900	0.042711	0.054802	0.065281	0.077450	0.085727	0.109385
1000	0.040520	0.051993	0.061935	0.073484	0.081340	0.103800
1500	0.033086	0.042458	0.050582	0.060022	0.066445	0.084822
2000	0.028654	0.036772	0.043811	0.051990	0.057557	0.073488
3000	0.023397	0.030027	0.035775	0.042457	0.047006	0.060027
4000	0.020262	0.026005	0.030984	0.036773	0.040713	0.051996
5000	0.018123	0.023260	0.027714	0.032892	0.036417	0.046512

Table 2.5: significant PCM values from real-statistics.com for two tailed tests of a certain confidence (horizontal row). The vertical column represents the number of samples.

t-SNE analysis

t-distributed stochastic neighbour embedding (t-SNE) analysis determines whether the labels in the encoded latent space are statistically "different". t-SNE is an unsupervised dimensionality reducing ML technique, used here to test the quality of deep clustering in Jenkins et al., (2021) and Snover et al., (2021). It assigns points probabilities of being "related" to each other in both the high dimensional feature space (here between 14 and 16 dimensions) and presents them in a low dimensional representation space (here 2 dimensions) such that the clusters in the high dimensional space are preserved. With both these distributions, the cost function (KLDiv) is minimised.

Perplexity is a hyperparameter corresponding to the number of neighbours each point is modelled to have. Therefore, a higher perplexity results in more global training of the t-SNE model.

Latent space analysis – decoding

Each K-means initialisation was decoded to gain a general understanding of each cluster centre to ascertain whether clustering is catching distinct features in the seismograms. This informs the size of the latent space, and the value of K. Chapter two will evaluate how well the latent space represents the seismic space, as each latent variable is plotted with respect to time.

With this technical and methodological framework established, the following chapter will apply this DEC model to its first major case study: a 31.25-day seismic dataset preceding the 2015 Mw 6.7 Iwate earthquake.

Chapter 3 – Clustering 30.25 days prior to Iwate EQ

3.1 Introduction

Following Snover et al. (2021)'s efforts to cluster anthropogenic noise and Mousavi et al. (2019)'s research on teleseism and local earthquakes, this chapter will test the DEC for its ability to cluster seismic data 31.25 days prior to a Mw 6.7 megathrust earthquake (hereafter EQ) off the coast of Iwate on the 16th February 2015 (Figure 3.1). This will test its ability to detect events with respect to station and time, and uncover additional information in the seismic timeseries like anthropogenic activity. The hypocentre exists in the Tohoku rupture zone where the Pacific Plate subducts under the Okhotsk Plate. It is a shallow earthquake (11.8km) in respect to other earthquakes in the region.

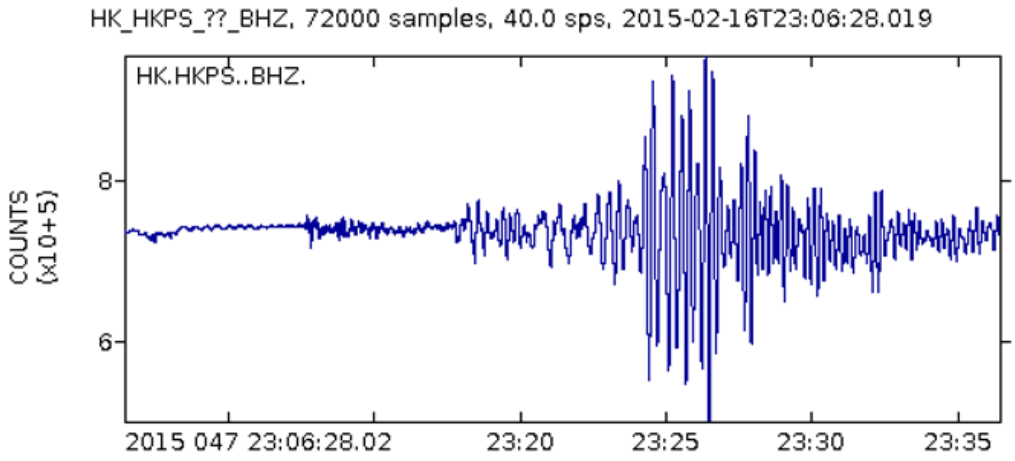


Figure 3.1: the seismogram of the earthquake following the first P-wave arrival as 23:06.28 on the 16th February 2015. This was a 6.7 Mw earthquake, and its high energy it can be seen in the resulting surface waves at 23:24.

Here, 10 HiNet stations all off the coastline have been used. Their close vicinity allows for comparison of local tectonic signals in addition to higher energy signals such as Mw>3 (Figure 3.2).

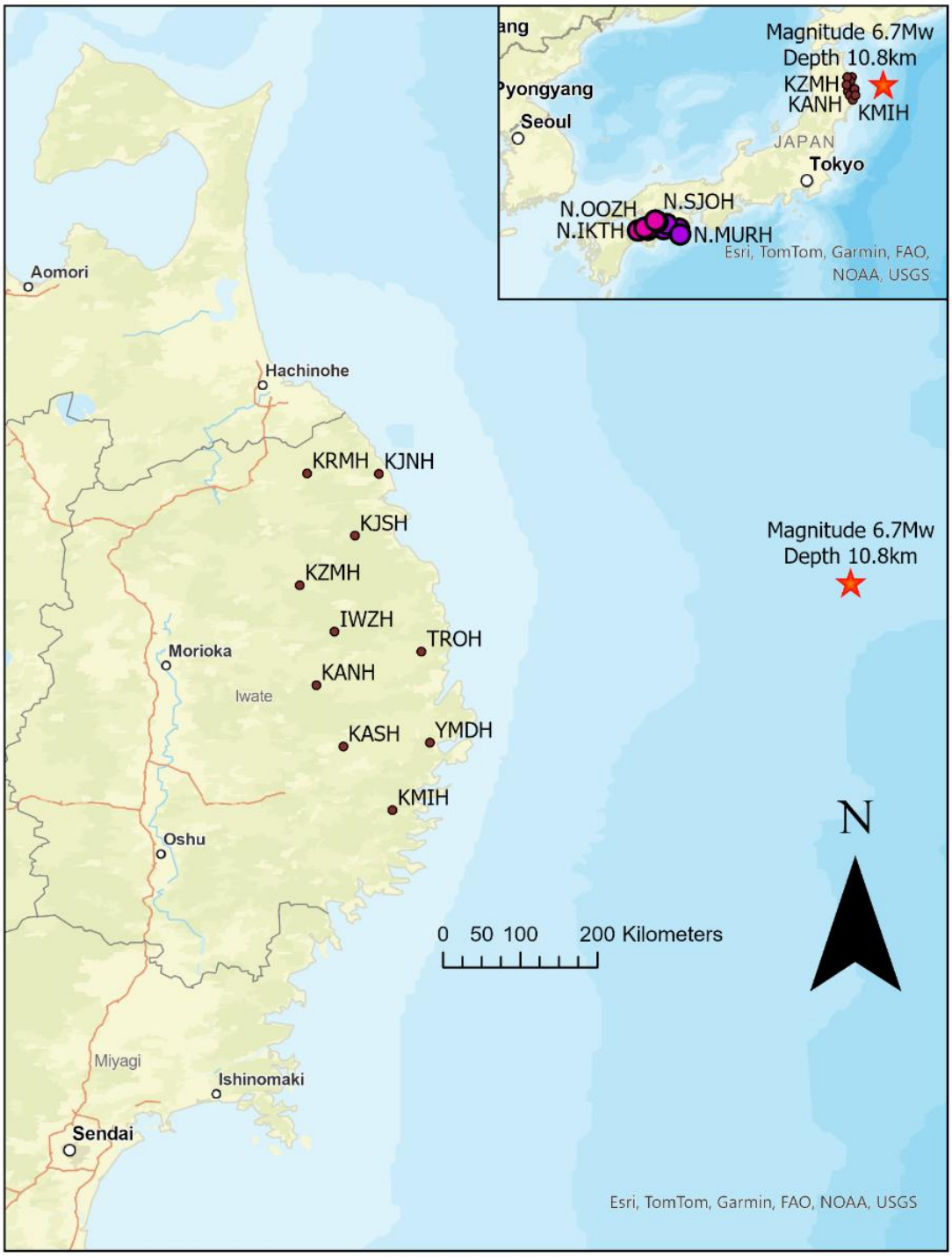


Figure 3.2: location of seismic stations in Japan used (in Red) along with the location of the EQ (star). The reference weather point is Morioka. The top right figure shows the location of these stations in respect to the stations used for the resulting NVT dataset.

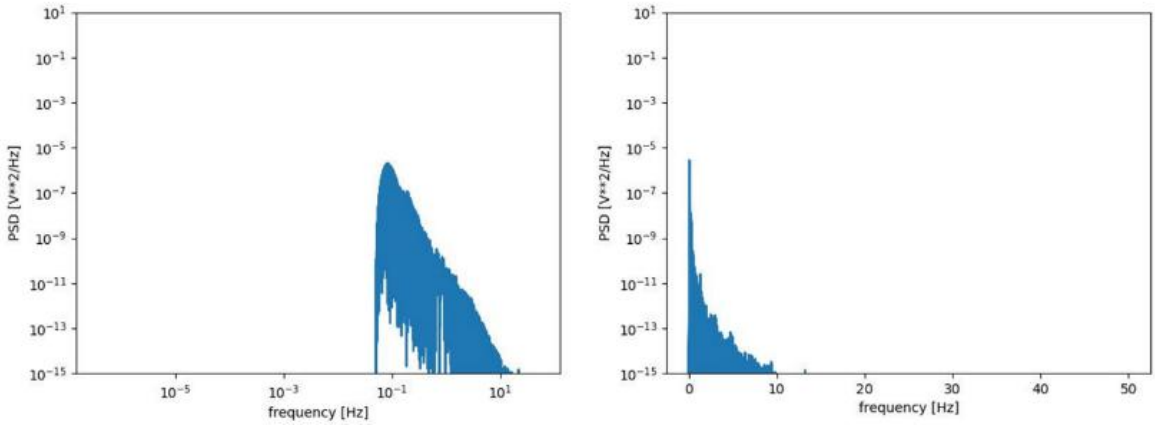


Figure 3.3: large window periodograms of the data set with logarithmic scale (left) and linear scale (right) highlighting that low frequency dominated the seismic signal.

The data is dominated by a low frequency signal which could represent background noise from ocean waves (Figure 3.3).

Additionally, the lowest frequency in the data is approximately 0.05Hz. This could be due to the limits of the seismometer, which is not broadband. However, signals below 0.05Hz will not be analysed in this thesis as it would require a large spectrogram window which would obscure high frequency signals.

3.2 Methodology specific to this chapter

Preprocessing was conducted alongside the methodology set out in the previous chapter. Each dataset experimented with is summarised in Table 3.1.

Savgol Filter

The Savitzky–Golay (Savgol) filter is a smoothing technique that performs a local polynomial regression within a sliding window of size m, replacing each data point with the value predicted by an n-th order polynomial fit. When applied to cluster assignments in the time domain, it reduces high-frequency fluctuations while maintaining the overall structure and relative prominence of local variations. This provides a balance between suppressing noise and retaining meaningful temporal patterns. In this study, a low-order Savgol filter is used (first-order), where $\tilde{n}$ denotes the total number of time instances in the cluster time series. For an independent time variable *t* and dependent cluster density *y(t)*, the convolution coefficients *C* are obtained by minimizing the sum of squared errors of the polynomial fit within each window, *m*,

$$Y_j = \sum_{i=\frac{1-m}{2}}^{\frac{1}{2}(m-1)} C_i y_{j+i} \; ; \frac{m+1}{2} \leq j \leq \tilde{n} - \frac{m-1}{2} \tag{3.1}$$

Correlation analysis

As described in chapter one, PCM analysis will be conducted to see whether each cluster signal (as a function of time) correlates with day/night signals and the cumulative number of earthquakes of Mw>3, Mw>4 and Mw>5. This is conducted for the whole set (signals are summed) and individual stations. The DEC successfully clustered some environmental noise in Jenkins et al., (2021). Here, the signals will be clustered to see if they monitor environmental / metrological noise from hourly data for Morioka from visualcrossing.com.

Summary of results

Many iterations of training the dataset with the raw data were made. The hyperparameter decisions are summarised in Table 3.1. After some preliminary analysis, it was decided P=240 second spectrogram lengths were the best. The frequencies and time bins will remain consistent with Snover et al., (2021), so having a longer time window captures lower frequencies. The number of clusters were varied between K = 4 and 8. Both low and high pass clusters of $K \neq 6$ were experimented with, but these did not produce anything meaningful in comparison. Below presents a table of all the DEC models analysed in this chapter. The models are trained independently of each other. There is no transfer learning.

Run	***Training Instances***	***Validation Instances***	***Testing Instance***	***Stations***	***P, Window length (seconds)***	***Window Overlap***	***K, number of cluster***	***Latent Space Size, $\mathbb{L}$***	***Fitted to***
1	107776	*26944*	*0*	*All*	*240*	*80*	*6*	*14*	*TrD*
2	*215808*	*53952*			*120*	*40*		*15*	
3								*16*	

Table 3.1: models that will be analysed in this chapter. The main differences are the number of samples, with an increased amount for runs 2 and 3 to compensate for larger latent spaces.

3.3 Run 1: High pass analysis ($\mathbb{L}_{14}$ Space)

Introduction

The DEC will be trained on a high pass of dataset as the low frequency tidal signal masks information in the spectrogram and little in an initial run of unfiltered data. Furthermore, training with high frequency waveforms was helped identify features in the Snover et al., (2021) study. High passing the data eliminates the high amplitude <1Hz signal seen in the unfiltered data which could mask low amplitude features. As previously stated, the high pass Butterworth filter of order 3 will be used to filter the seismic timeseries.

Loss function

The loss function is attached for the first 200 epochs (Figure 3.4). As can be seen the MSE training/validation losses are correlated implying the model is not overfitting. Nonetheless, the overall loss is high taking a number of epochs to eventually learn features from the training and validation sets. The overall loss (around 1.55) is significantly higher than the unfiltered data and the Snover et al., (2021) study. This could imply that the network treats the spectrograms as noise which is observed.

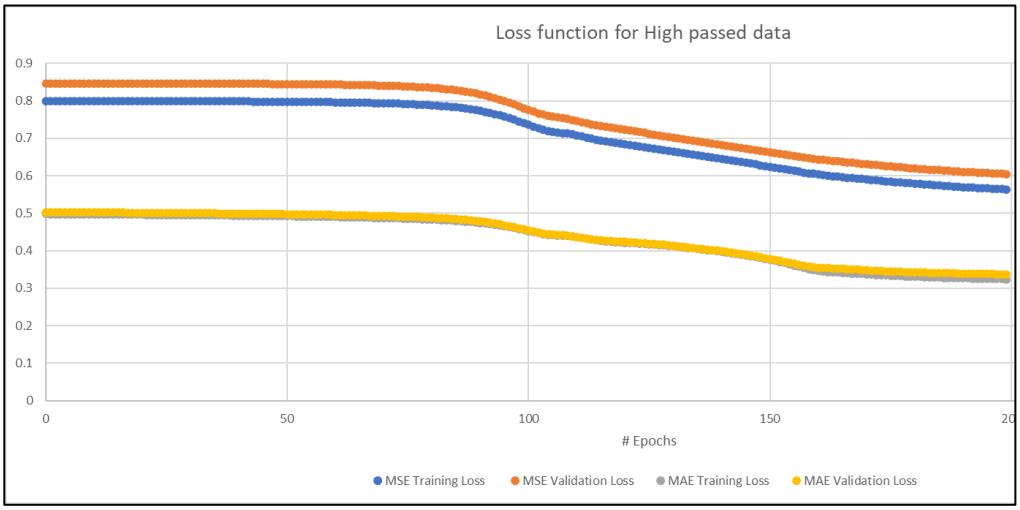


Figure 3.4: loss function for high pass dataset showing MSE and MAE curves for both the training and validation datasets.

Reconstructions of various spectrograms and reconstructions of cluster centres.

The low dimensional latent space eliminates substantial detail in the reconstructed spectrograms, although a good proportion of the spectrograms are noisy. Nevertheless, high amplitude impulsive features are preserved (Figure 3.5, more figures in Appendix). The latent space was hard to interpret for individual spectrograms. Reconstructions that occupy a large amount of the spectrogram stimulate a greater number of latent variables. Clearly, the DEC is regarding a good proportion of the spectrograms as signal. To test whether this is noise or not, the spectrograms would be inverse Fourier transformed into the time domain. As the Fourier coefficients are not known this cannot happen. Therefore, a higher dimensional latent space will be experimented with to see if features are preserved or eliminated in spectrograms.

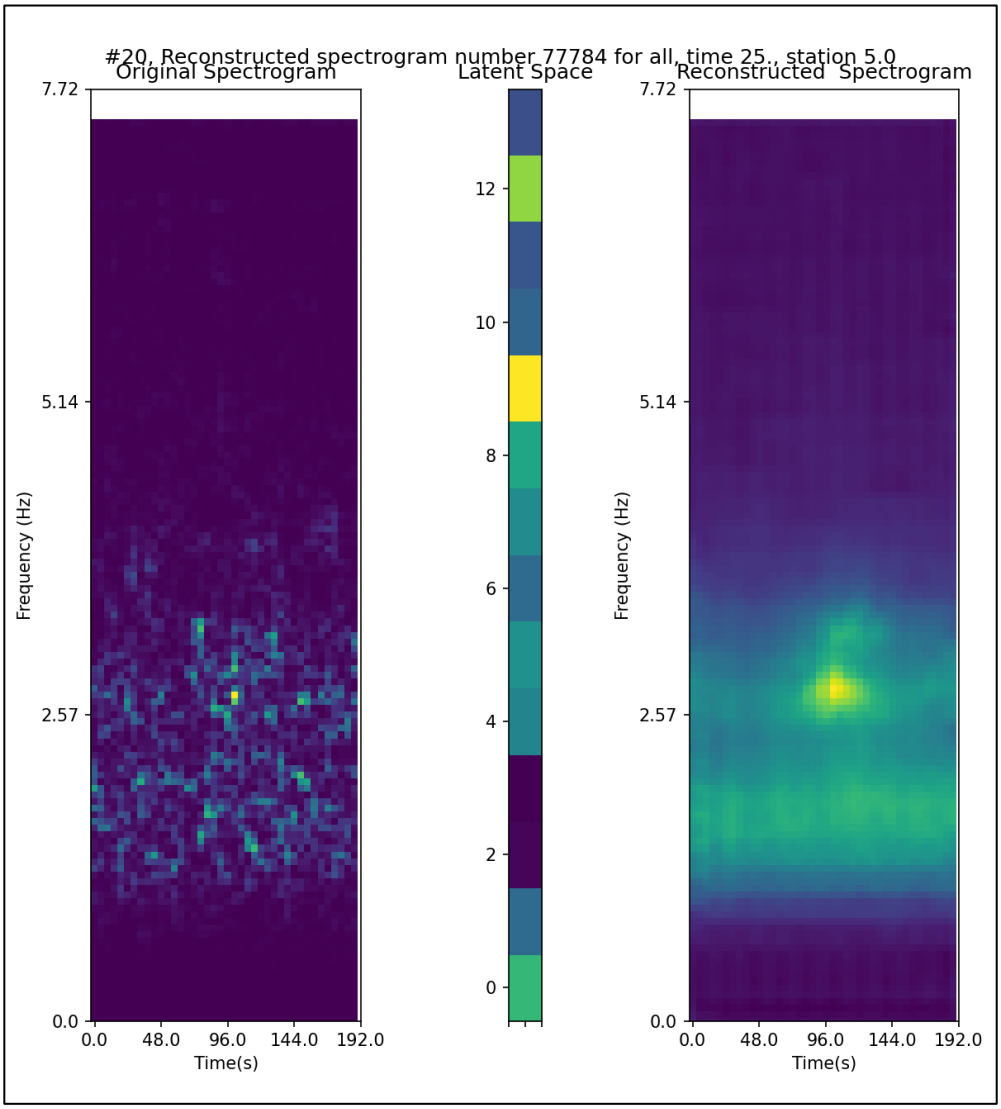
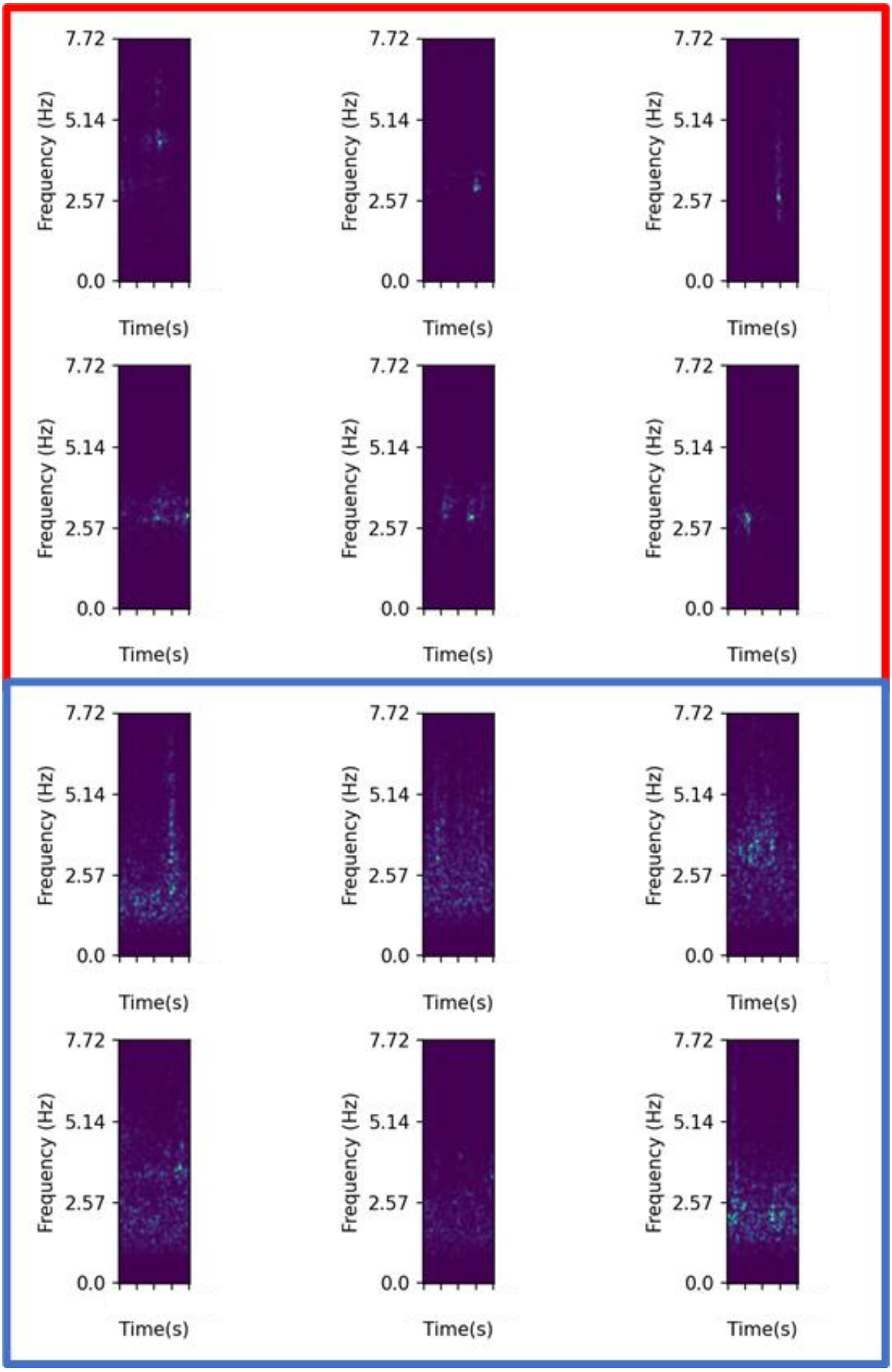


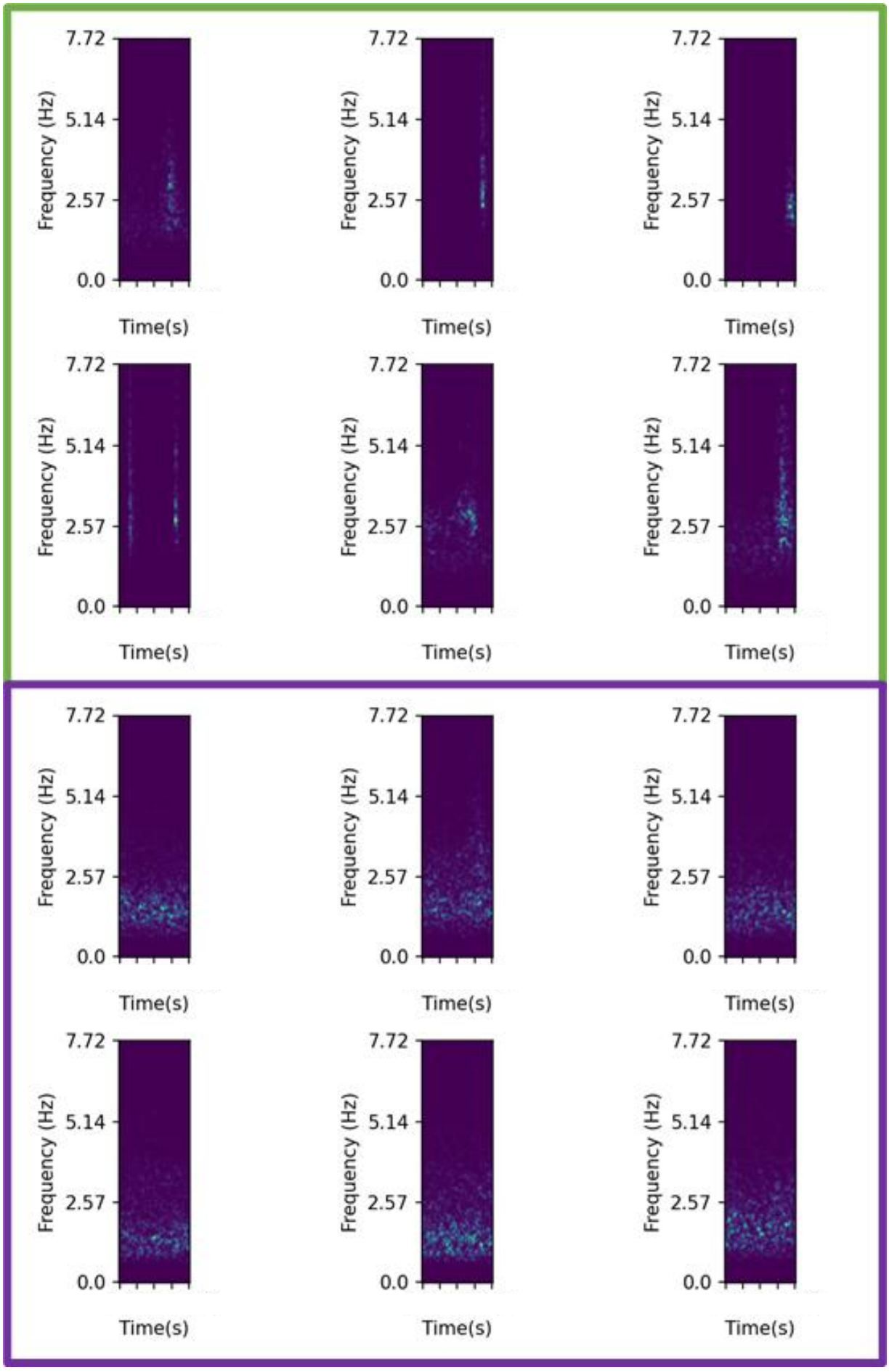
Figure 3.5: This presents an example reconstruction for spectrogram in 14-dimensional latent space with the original spectrogram on the left (normalised), the latent space in the middle (same scale as the colours in the spectrograms), and the reconstructed spectrogram on the right. A good proportion of the essential details are omitted with a large amplitude feature preserved. It suggests that substantial the noise in the spectrogram is preserved and grouped into one, effectively dimming potential detail.

Clusters overall

The high pass data was experimented for clusters 4,6,8 and k=6 was the only cluster in which a meaningful correlation was shown.

Clusters for k = 6


7.72
5.14
2.57
0.0
Frequency (Hz)
Time(s)


7.72
5.14
2.57
0.0
Frequency (Hz)
Time(s)

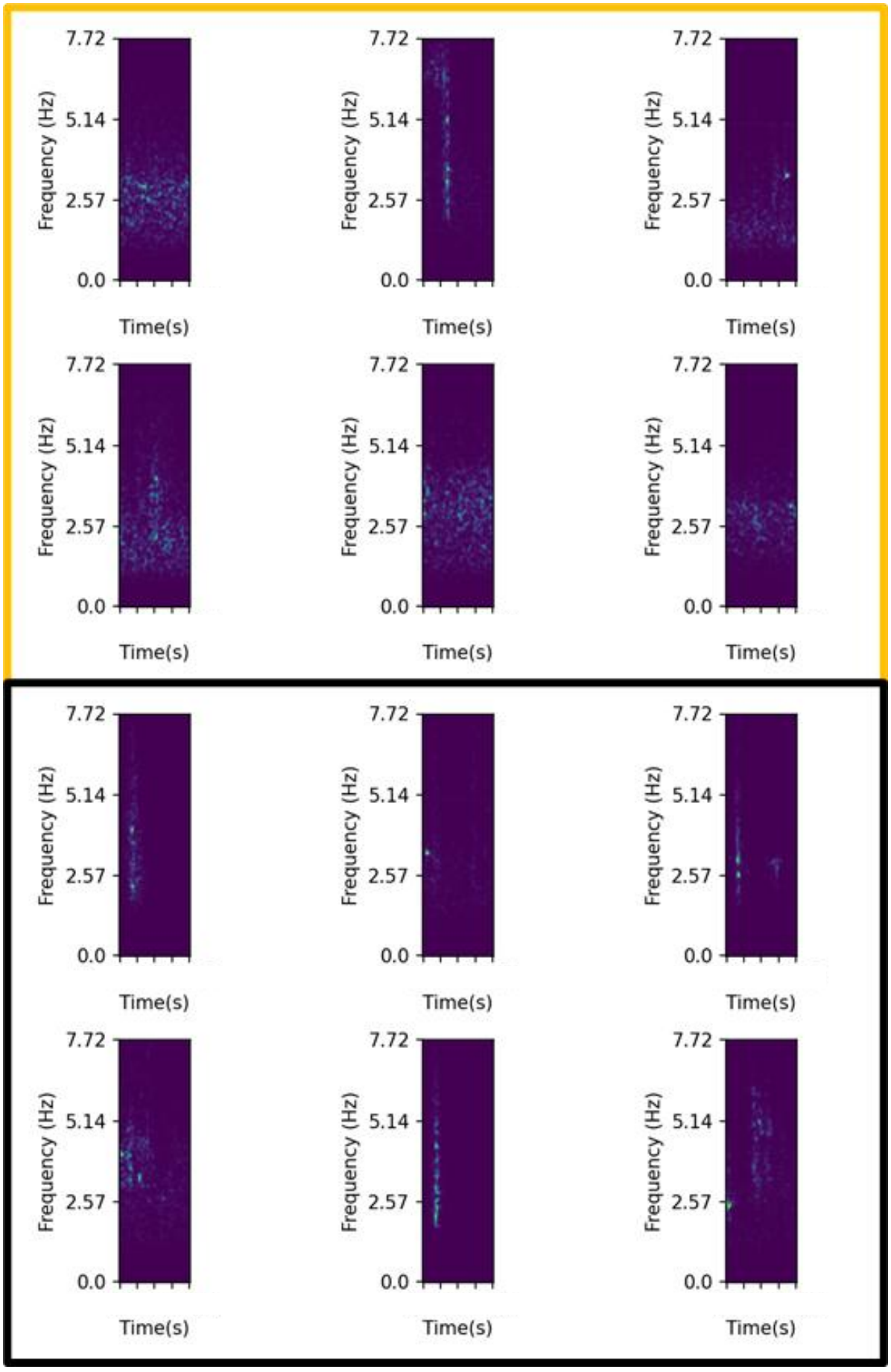


Figure 3.6: high passed cluster for k = 6 in order from 0 (top) to 5 (bottom) for constant spectrogram length of 240 seconds. The border corresponds to the assigned colour the cluster which will be used in further

analysis in this chapter. There are two main differences: C1,2,4 number of the clusters, have quite short impulsive features, whereas others (C0,3,5) have more-impulsive longer more noisy features.

Clusters space and time HIGH PASSED data, k = 6

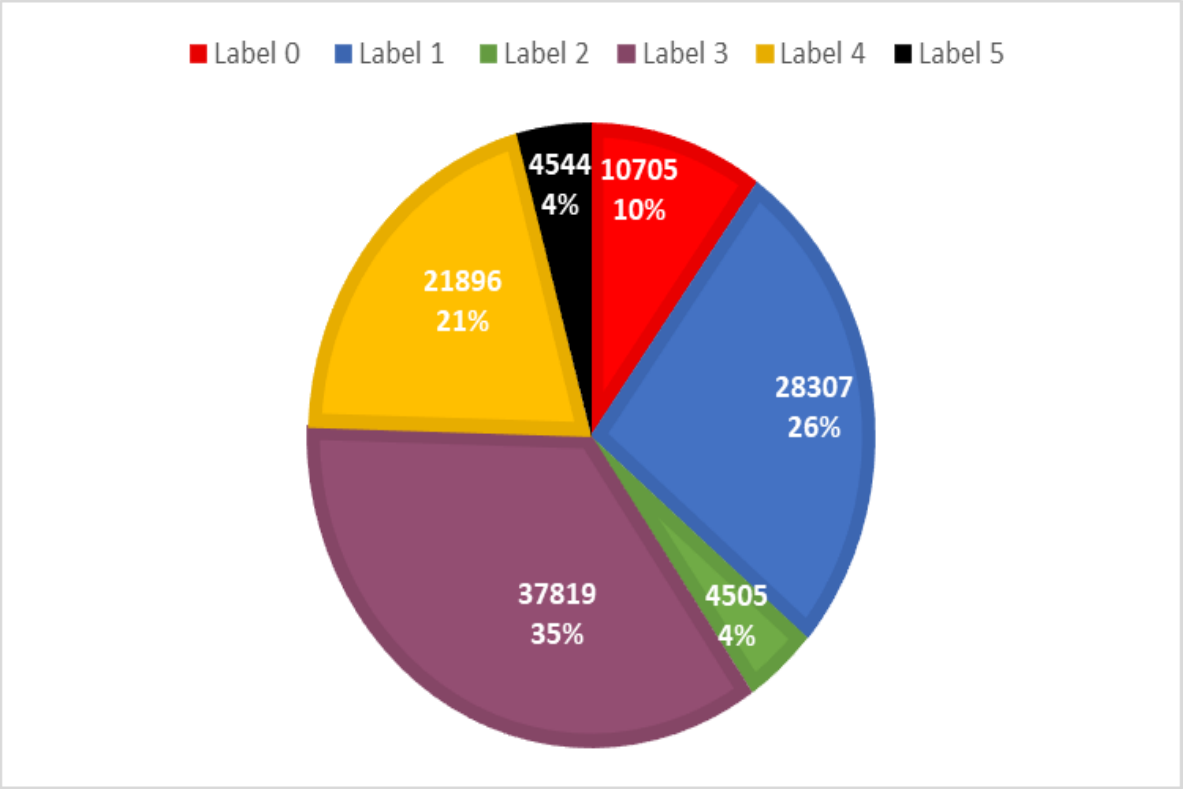


Figure 3.7, the number of spectrograms assigned to each cluster. C4,1,3 are somewhat similar occurrence rates, whilst C0,2,5 are less common in occurrence but have similar occurrence rates.

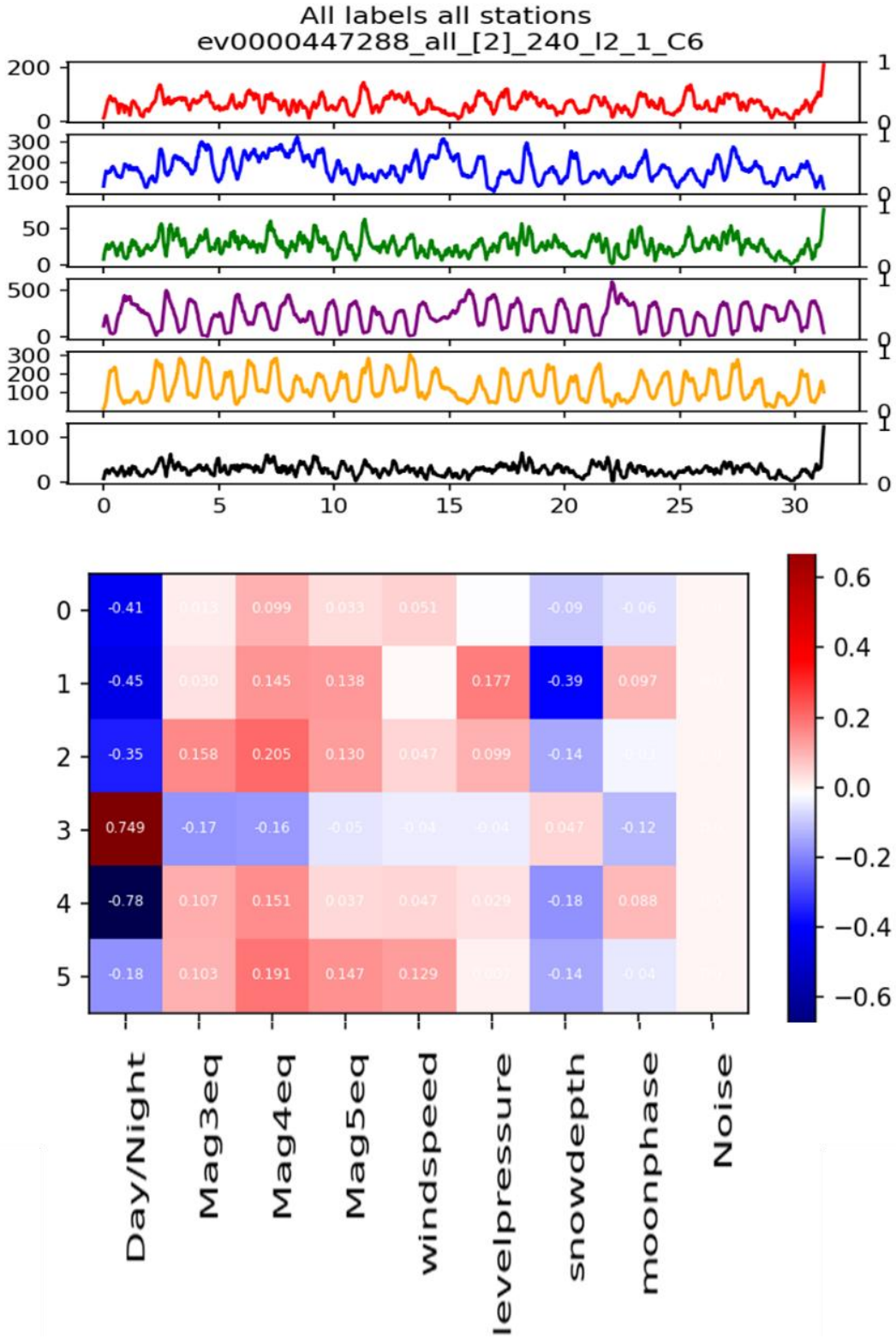


Figure 3.8 I and II: cluster analysis over the course of the 31.25 days. Each cluster has been normalised between 0 and 1 (right hand axis) to allow for comparison as the number of samples vary between cluster (left hand y axis). The bottom figure demonstrates the PCM values for each cluster and respective relationship with strong correlations between day and night, especially in C3 and C5.

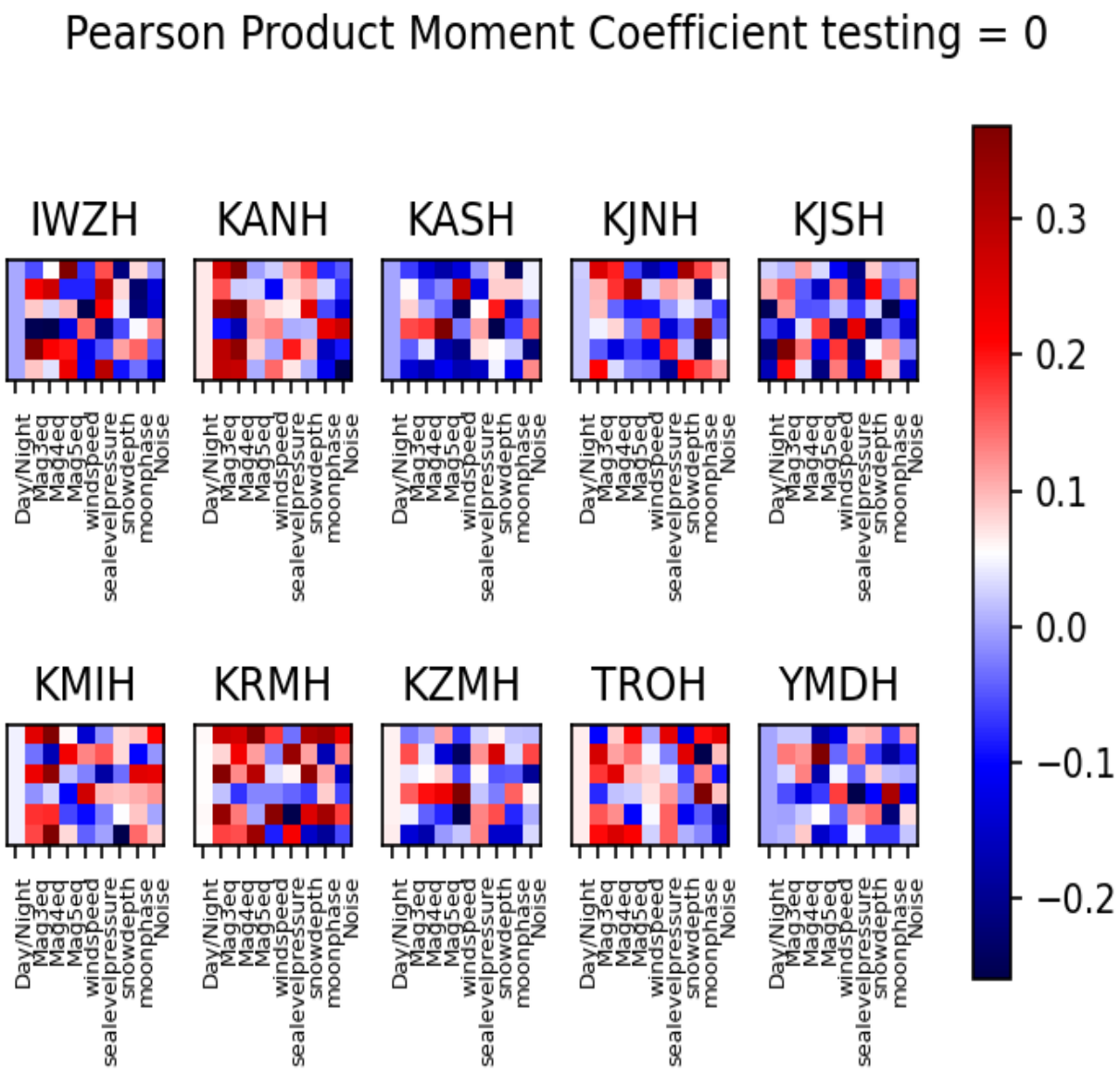


Figure 3.9: PCM values for each individual station demonstrating that that the day /night signal is low, almost 0 for all clusters at the station.

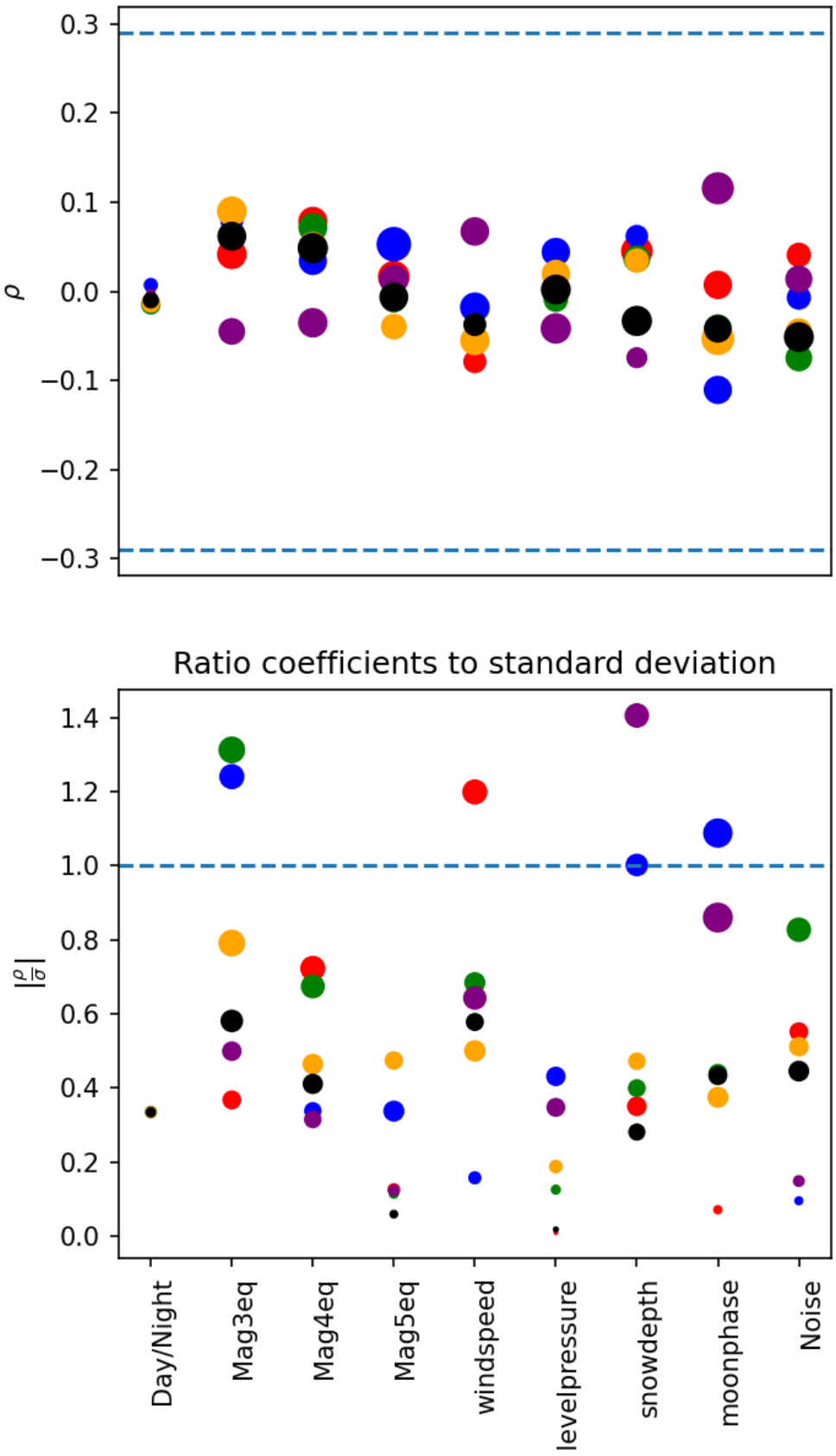


Figure 3.10: Average cluster coefficients for each potential relationship, along with the ratio of coefficients in comparison to the standard deviation of coefficients between each station to ascertain whether the cluster correspond to a comprehensive relationship between all stations.

Observations of the all-stations set

C3 and C4 have a distinct day / night correlations with C3 picking up day time signals and C4 picking up night time signals (Figure 3.8II). C3 is dominated by 1-2.57Hz signal with a noticeable band amongst all spectrograms whilst C4 is dominated by more infrequent signals with some reaching as high as 5Hz (Figure 3.7I). This, however, is not shown for individual station occurrence as shown in the instance of station 9 below (Figure 3.11). This suggests that it picks infrequent day / night signals spread amongst all stations that average a day / night signal. For C [0,2,5] there is a sudden increase at the end of the timeseries prior to the earthquake. These clusters oscillate at a high frequency. This suggests the events they pick could be impulsive or infrequent which is demonstrated in Figure 3.6 for C0 and C2 but not so much for C5. For all other characteristics, the only strong PCM correlation is an out of phase correlation with C1 and snow depth. For the Mag > 3 correlation metric, there is a large ratio of PCM to standard deviation, which suggests the network has potentially learnt features of the Mag > 3 earthquakes, irrespective of the variations between stations.

For the singular station set, there is no common theme for potential relationships among clusters with all falling out outside of the 2 tailed criterion of 1.39 for the given number of samples in the earthquake series (n=4)

From these initial results, is clear that when high pass filtering the data, a large number of stations should be used to effectively cluster and identify relevant features across the entire network.

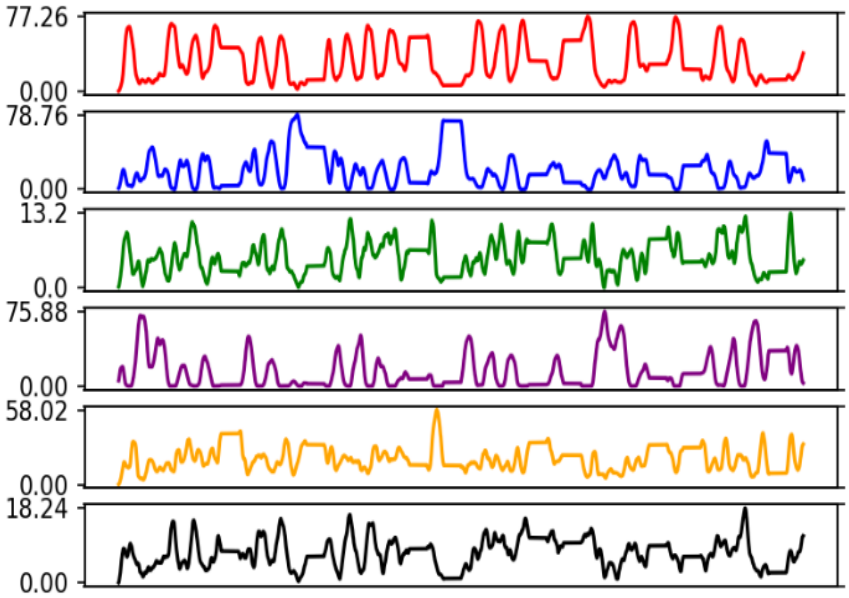


Figure 3.11: representation of clusters for station YMDH showing only mild day night perturbations with respect to 31.25 days of time although some of these perturbations are broken in places.

3.4 A note on the latent space

Due to a significant loss of information for latent space size = 14, the latent space hyperparameter has been altered. Latent values of 15 and 16 have been chosen to compare how much is retained in the reconstructed spectrograms. With every increase in dimension comes an order of magnitude increase in volume, resulting in the number of latent samples per unit of "latent volume" becoming an order of magnitude less. This is known as the curse of dimensionality (COD). Values higher than this have not been chosen in consideration

of COD. Increasing the latent space shows little difference and only features from the L14 space. The findings of this are contained in Appendix II.

3.5 Comparison of Latent space dimensions

This comparison shows that while increasing the latent space dimension from 14 to 15 (and halving the window size) improved the detail of reconstructions, it also negatively impacted the model's performance. The model became more sensitive to the day/night signal (biasing 5 of 6 clusters) and produced less distinct cluster centres (KMCR) and poorer cluster separation (t-SNE). This suggests that L=14 was the most effective run, and that L=15 and L=16 may have led to overfitting to anthropogenic signals (Table 3.2). The findings for L=15 and L=16 are contained in Appendix II.

Metric	**Run 1: L=14**	**Run 2: L=15**	**Run 3: L=16**
Window Length (P)	240 seconds	120 seconds	120 seconds
Training Instances	107,776	215,808	215,808
Reconstruction Quality (Figure 3.14 and Appendix for examples)	*Poor* Eliminates significant detail. Oversmoothed, indicating information loss.	*Better* Features are finer and more detailed. Noise removal is still strong.	*Poor* Reconstructions are similar to L15. Features have a large, blurred features, suggesting uncertainty.
Day/Night Signal	**2 clusters** (C3, C4) show strong correlation. *Quantified:* PCM $\rho \approx$ -0.75 (C3) and -0.78 (C4).	**4 clusters** (C0, C1, C4, C5) show correlation. *Quantified:* PCM $\rho \approx$ -0.65 (C0), -0.42 (C1), -0.59 (C4), 0.45 (C5).	**5 clusters** show correlation. *Quantified:* PCM $\rho \approx$ 0.61 (C0), -0.66 (C2), -0.54 (C4), -0.39 (C5).
"Spike" (EQ Event)	Detected in clusters C0, C2, and C5.	Detected. The spike appears "sharper" in the cluster time series.	Detected. The spike is sharpest in this run.
KMeans Centre (KMCC)	*Good* Shows promising difference and heterogeneity.	*Poor* Shows a less diverse array of cluster centres.	*Poor* Also shows a less diverse array than L14.
Cluster Separation (t-SNE)	*Poor* Shows slightly more structure than unfiltered data. No distinct groups uncovered.	*Poor* Shows no great distinctions between samples.	*Poor* Also shows no great distinctions. Fails to show clear whitespace between clusters.

Table 3.2: Comparison of latent space sizes with respect to different performance metrics.

3.6 Overall Discussion

Day / night signal

Prominent industrial activity is not present at night which allows more non-anthropogenic signal (tectonic or not) to appear which could explain the discrepancies in frequency range. Nevertheless, the nighttime cluster C4 does not look different in comparison to other spectrograms although it is the most diverse in containing noisy, quiet and impulsive signals (Figure 3.6). This could suggest the DEC was successful in recognising

more complex patterns amongst the nighttime signal. The neural network in general picks up on a day/night signal but not in individual locations. This suggests that the anthropogenic noise is not local. Originally, it was thought it could be due to solid earth tidal signal, but this would be very low frequency and unlikely to be picked up. Low rates of cluster occurrence at individual stations, with some maximum occurrences under 30, are not able uncover significant anthropogenic signal that could be effectively correlated with the PCM value (Figure 3.8), as they contain a greater amount of noise relative to signal.

Investigation of the spike

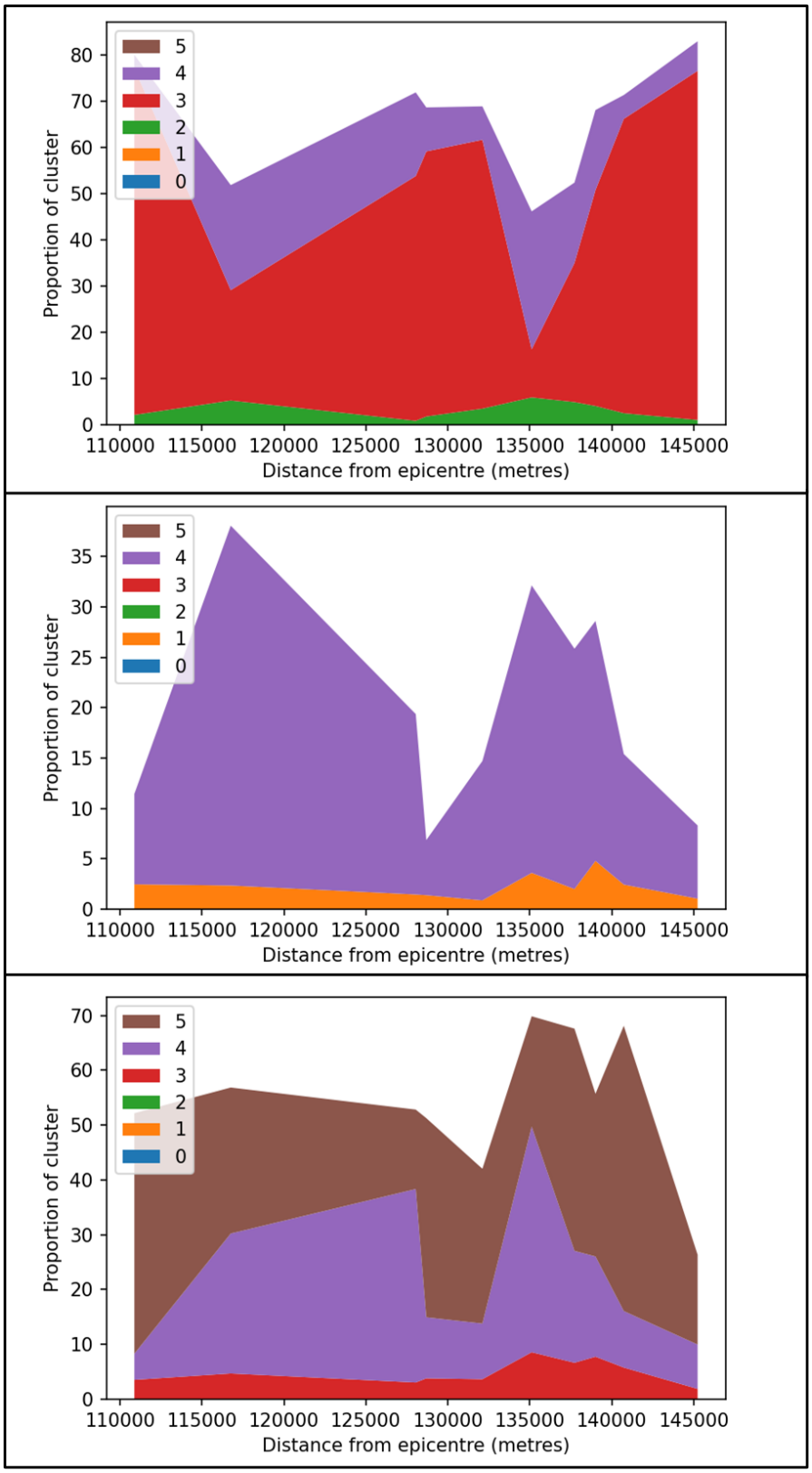


Figure 3.12: investigation of the spike (L14 top to L16 bottom) in respect to distance from epicentre amongst clusters that show a spike at the end of the timeseries suggesting that there is no immediate attenuation in energy. This are plotted for the last 1000 data points (12.5 hours for L_{14}, 6.25 hours for $L_{15\text{-}16}$).

The spike, that featured in both L14, 15 and 16 space, was investigated to see whether it is correlated with distance from the epicentre of the earthquake as there was a large discrepancy seen between the stations – as some stations had no spike. There is no attenuation shown suggesting the spike feature is not a function of the distance from the epicentre (Figure 3.12). Then again, all stations are a large distance away from the

epicentre, so if there was a precursory activity, it most likely was attenuated prior or existed to this distance. The spike is from a high energy signal – most likely the result of potential foreshocks prior to the EQ.

After some investigation, it was discovered that the 6.7Mw event might have been included in the timeseries, approximately an hour before. The reason this sudden burst of seismic activity is included is unknown to the author as the dataset is provided by an external source. Nevertheless, the earthquake only accounted for approximately 15 minutes of data (Figure 3.1). This suggests that the DEC can detect megathrust earthquakes and events around it, as demonstrated by impulsive events in Figure 3.14– potentially showing foreshock activity. There were a few Mb=3 events around the vicinity of the earthquake so these shall be investigated. Figure 3.12 shows the proportion of the Mw>3 EQs in fact decreases at the end of the series. This suggests that this spike either represents the earthquake or various precursory phenomena surrounding it. The only way to tell would be to rerun the model with the earthquake omitted. Interestingly, this spike in in the clusters appears only when the data has been high passed, suggesting that the phenomenon exist in the high frequency range. Figure 3.14 show spectrogram instances for clusters that show the spike in the time period that the spike is presented. A good proportion of the spectrograms are impulsive and noisy potentially representing EQ activity. These spectrograms have been standardised but not normalised. Further research should normalise the spectrograms in addition to standardisation to assess if this causes any changes in clustering.

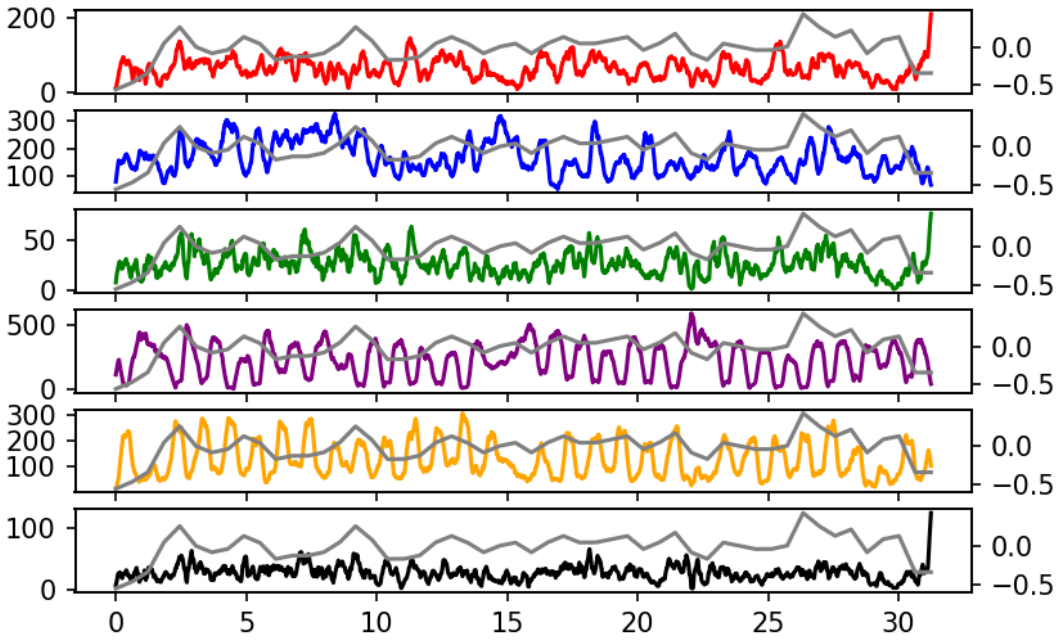


Figure 3.13: All labels over plotted for the number of Mw>3 earthquakes seen (grey line) and occurrence counts on the left- hand axis for all stations in L14 space.

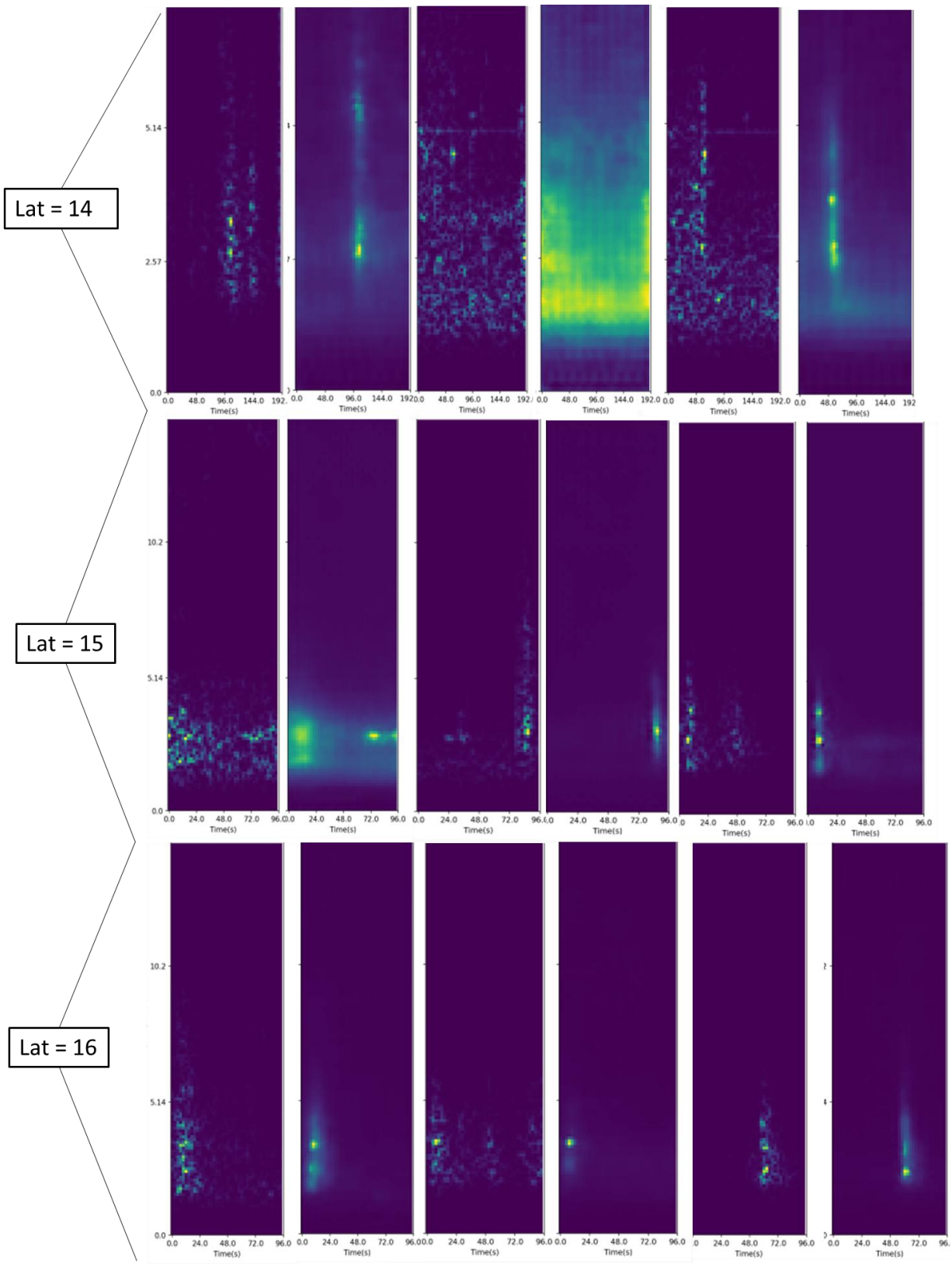


Figure 3.14: pairs of original and reconstructed spectrograms within the last 10 hours for identified clusters that have a "spike" for $L_{14\text{-}16}$ demonstrating a level of foreshock activity that is reconstructed regardless of the size of the latent space.

Comparison of latent variables.

To understand what is influencing the clustering, the latent variables are plotted (Figure 3.14) as a function of time to gain a understanding of each dimension in the latent space.

In plotting, a rolling mean of window size 150 was taken along with decimation.

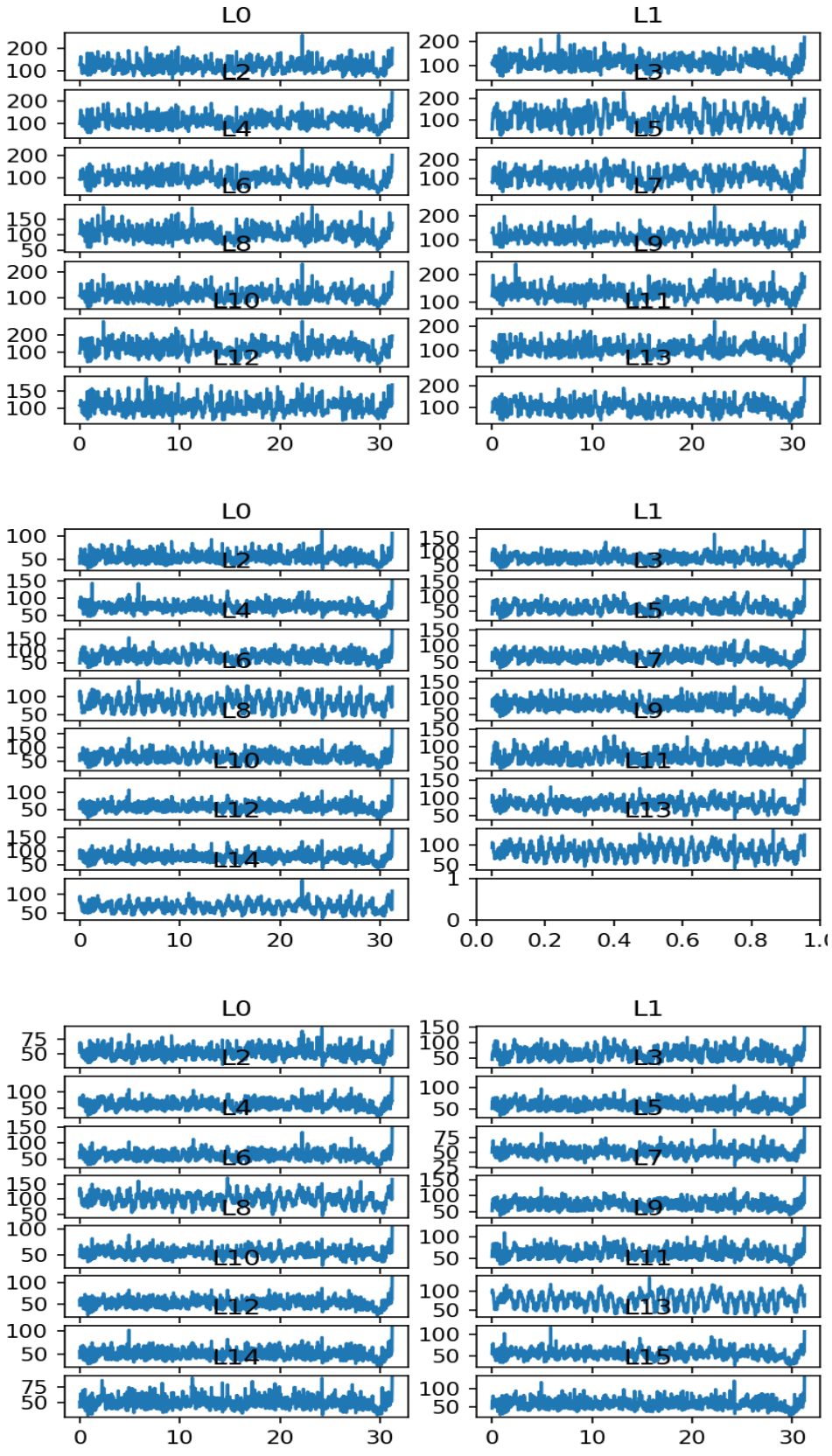


Figure 3.15 I-III (top to bottom): $L_{14\text{-}16}$ variables plotted in ascending latent dimension against time (given in days). They have been averaged over every 150 values and hence decimated.

Variables sensitive to the event (the "spike") at the end of the timeseries demonstrated sharper spike with each increasing latent dimension. This suggests the effect of the spike becomes more apparent in variables for increasing latent dimensions. Additionally, oscillatory activity indicating a day/night signal is seen in more

latent variables as the size of the latent space increases which explains why many of the clusters are sensitive to the day/night perturbations.

There is still a large amount of uncertainty as to whether the encoded latent space is representative of known and unknown seismic features. The K-means initialization is vital in representing the seismic space fully.

To test the effectiveness of the K-means initialization, the K-means cluster centres (hereafter KMCC) can be reconstructed to test whether they are unique and have a semantic meaning (Figure 3.16). Ideally, heterogeneity is needed in the cluster centres as that makes it more likely that they represent the seismic space.

Unsurprisingly, the unfiltered data KMCC are all similar (Figure 3.16). This suggests the DEC was not able to distinguish distinct clusters which is unsurprising given the small number of bins they occupy. Reconstructing the high passed KMCC (third figure down) shows a promising difference in the spectrograms. KMCC0 demonstrates no high pass activity at all, whilst KMCC1 demonstrates a large section of high amplitude energy with a prominent blurred feature seen in previous reconstructions. KMCC2-4 show narrow frequency spikes whilst KMCC5 demonstrated a large high frequency impulsive spike, suggesting it represents impulsive activity, which indicate the spike seen in Figure 3.15. For $L_{15\text{-}16}$, the KMCR surprisingly shows a less diverse array of cluster centres than L_{14}. This could most likely be due to the higher dimensional space not capturing features that cannot be represented in the seismic space $\mathbb{S}$ whilst also being unable to effectively train the latent space due to lower number of samples. It must, however, be noted that due to 120 second long spectrograms, the dominant signals will occupy fewer frequency bins. L_{16} contains one impulsive feature at CC1 whereas L_{15} contains 3 at KMCC0, KMCC2 and KMCC3, these counteract clusters for where the spike is seen at the end of series. Interpretation of the cluster centres suggest that using a higher dimensional latent space is not appropriate for this seismic timeseries and number of samples given.

To test the effectiveness of increased latent dimensions on the cluster assignments instead of the reconstructions, t-SNE analysis will be conducted on random samples of the latent space (varying sample sizes) to test whether each sample can be distinguished. These samples have been assigned a cluster post soft clustering (KL divergence) from the DEC model. This was a successful technique used in Snover et al., (2021) and Ozanich et al., (2021) to observe the latent space and to test the quality DEC clustering. It is obvious that for the low passed data there are no distinct clusters. Despite that, C5 is not seen, however, this could be due to the low sample size used.

No distinctive groups were uncovered in t-SNE analysis (Figure 3.17). t-SNE analysis for L_{14} shows slightly more structure with a slight spike encountered at the top of the graph and has been clustered with C4 which demonstrates a nighttime signal. Nevertheless, the K-means assignment has grouped the data into more distinct sets than the unfiltered dataset. $L_{15\text{-}16}$ show that there are no great distinctions in each sample – there are no distinct clusters although they do demonstrate spikes that are assigned to C1 (L_{15}) and C3 (L_{16}) with both demonstrating day/night perturbation. However, slight gaps in the representation could be seen for C4 (L_{15}) and C5 and C2 (L_{16}) – these clusters also demonstrated day / night perturbations and have greater sample sizes. Gaps in the data could be the result of the small sample size used in the t-SNE analysis. Here, a large amount of overlap in each cluster and the sizes of a good proportion of the clusters are not consistent. A requirement of the K-means clustering is that the variance is equal between the clusters (Ong et al., 2020).

Most of the clusters are of a similar size with some that occupy an small space (Figure 3.17). t-SNE approximates the higher dimensional latent space, and is a non-deterministic ML algorithm so these comments must be treated with caution. However, comparing this to the Snover et al., (2021) study (Figure 3.18), all the t-SNE plots are poor, as there is clear whitespace between each of their clusters. The DEC t-SNE here is comparable with their K-means initialisation t-SNE and not with their final post soft-clustering t-SNE. The method of redefined cluster centroids with soft assignments has not been entirely successful and is invariant to the latent dimension suggesting DEC is not as successful at clustering for the data in this set. Other studies often compared the DEC with gaussian mixture modelling (GMM) with some demonstrating that GMM is a more effective clustering technique (e.g. Ozanich et al., 2021) which could also be a method to test the effectiveness of the DEC. A disadvantage of the DEC model is that it uses the student's t distribution meaning there are fewer contrasts in the soft labels (an array of probabilities) assigned to each datapoint. This potentially causes less of a discrepancy to be made in the latent space but does not explain why these results deviate from Snover et al., (2021)'s study.

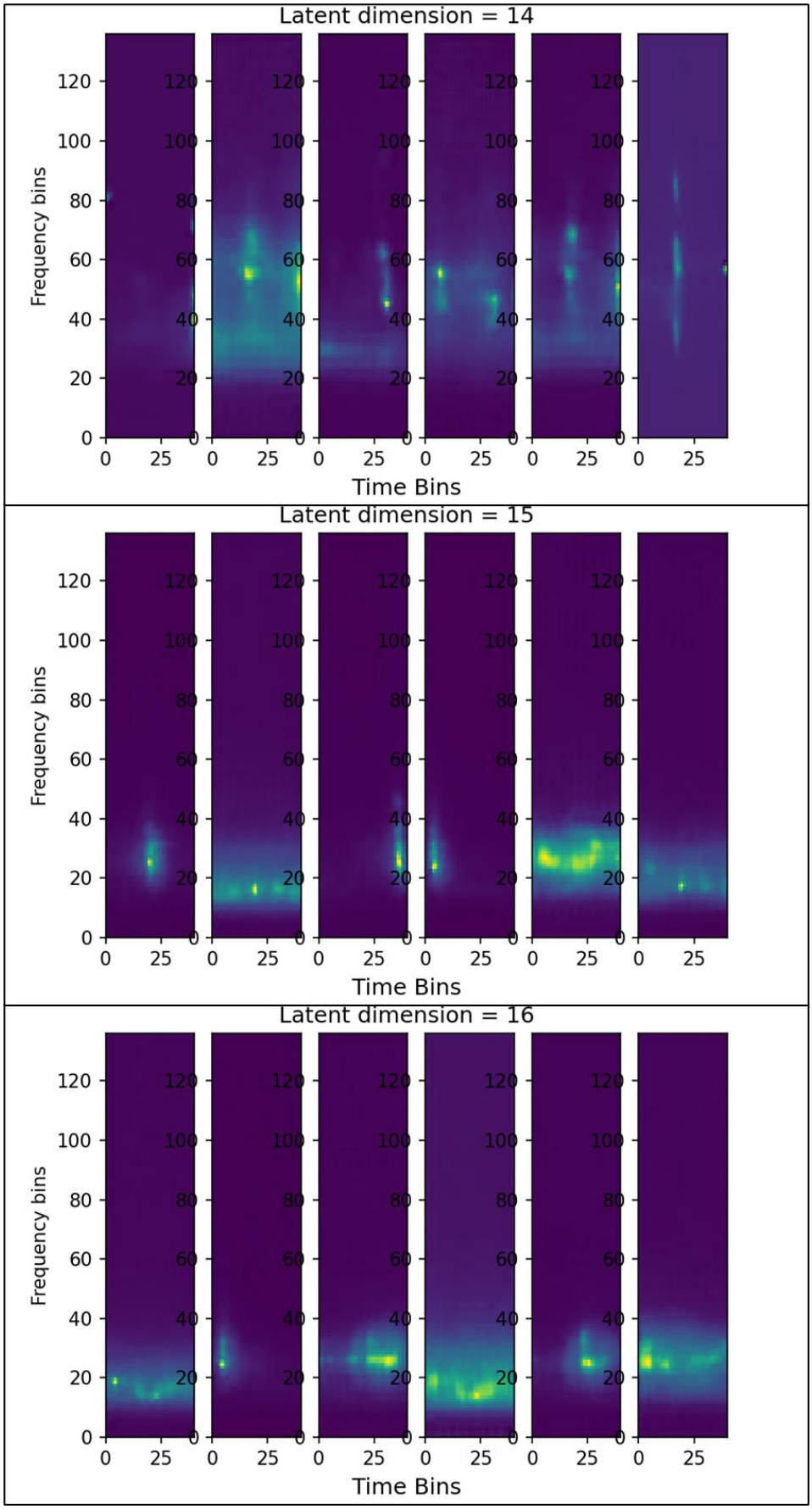


Figure 3.16: K-means initialisation reconstructed cluster centres (KMCR) over time bins (41). Top is unfiltered, below is high pass L_{16}, below that is high pass L_{14} and bottom is high pass L_{15}. For each set of KMCR, the spectrograms are labelled CC[0-6] from left to right.

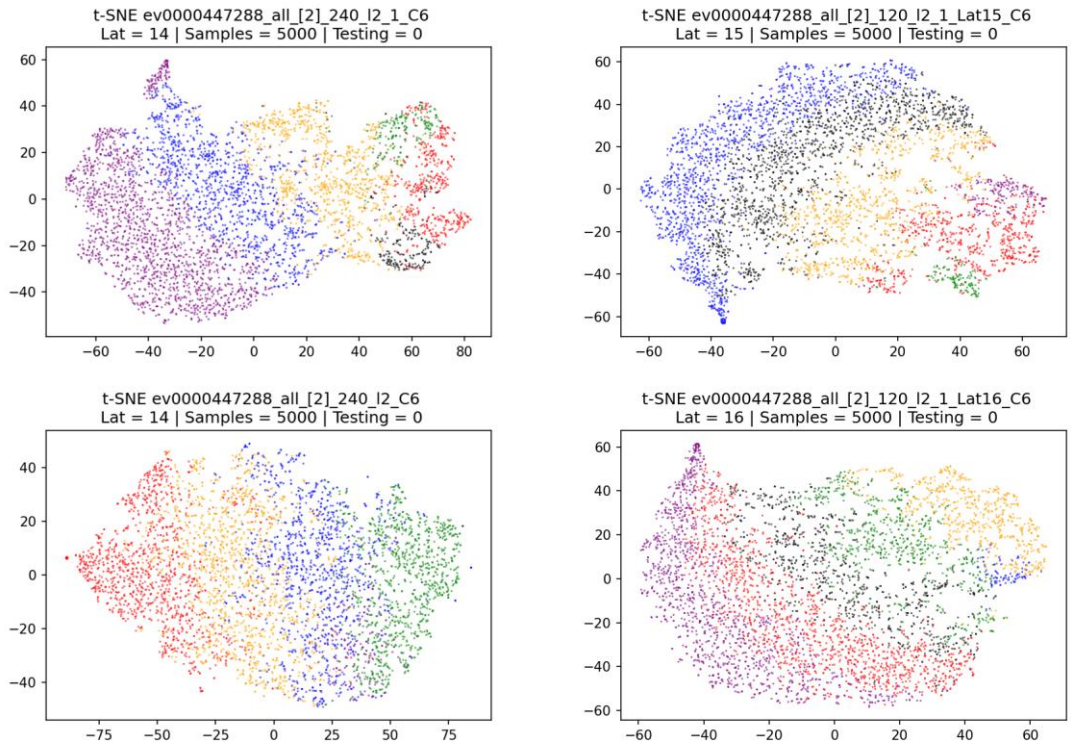


Figure 3.17: t-SNE for all models respective latent space along with final labels from soft clustering. There are no distinctive groups amongst the data.

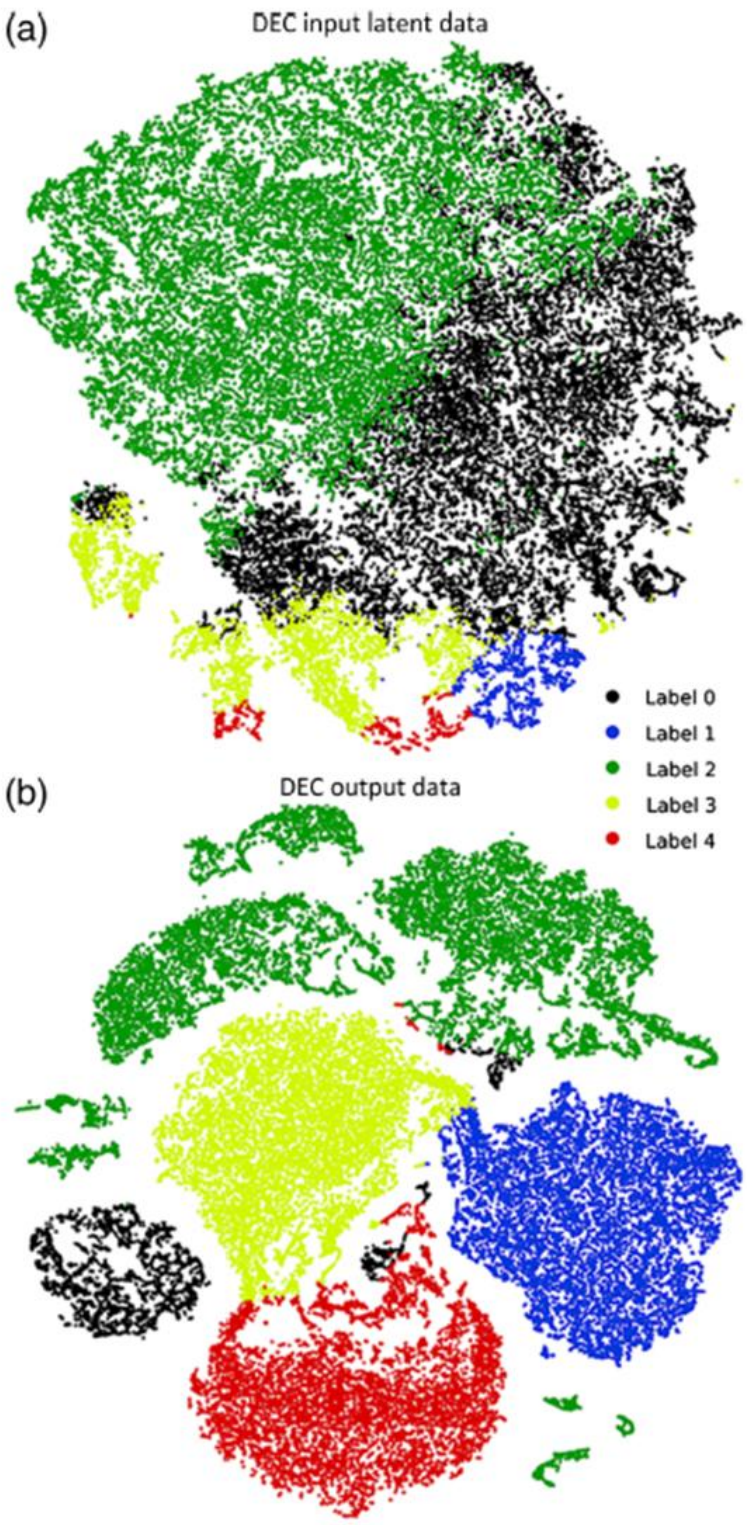


Figure 3.18: t-SNE plots from Snover et al., (2021)'s study into using the DEC to cluster anthropogenic noise. The top graph shows the t-SNE reduced latent space into the representation space, the bottom graph shows the result following application of the DEC model showing clear white space between many of the clusters.

3.7 Conclusion for chapter 3

This chapter set out to uncover commonly occurring distinctive features prior to earthquakes, whether they can be clearly uncovered in the spectrograms or not. The data must be high passed in order to return distinctive short-term features, in spectrograms prior to earthquakes. Here, 2 clear features are identified, a prominent day/night/anthropogenic signal and a signal corresponding to a sudden increase (the “spike”) at the end of the series whilst poorly detecting known features (EQs) prior to earthquakes

The anthropogenic signals were not recognised at individual stations yet were identified in the whole station set. This was possibly due to the low sample rate among singular stations along with the DEC identifying commonality in anthropogenic activity among all stations.

This is either a precursory activity or fore and main shocks of the earthquake. The DEC reconstructions are mostly impulsive with many representing normal seismic activity although it does not attenuate with respect to epicentral distance. This suggests the DEC has learnt that these spectrograms correspond to the EQ, despite the occurrence of Mw>3 EQs decreasing at this point. To the user, they do not look any different from impulsive activity earlier in the series. A full interpretation has not been established, but the results/data show likely represent the 6.7Mw EQ being mistakenly represented at the end of the cluster but could also correspond to lower energy impulsive activity around the time of the EQ. Nonetheless, the DEC has uncovered unknown signals around the time of the EQ which could indicate precursory activity.

Due to uncertainty in reconstructions, the size of the latent space was altered to test if it would uncover additional detail. Increasing the latent space for the high passed data only increases cluster sensitivity to day/night perturbations and the EQs at the end of the series whilst not uncovering further information in reconstructions. The t-SNE analysis was poor in comparison to previous studies which suggested training had not been optimised.

Overall, high passing the data only presents some additional features in the seismic timeseries prior to earthquakes whilst increasing the latent dimensions prevents other features from being represented. This chapter demonstrates a step forward in using deep clustering to uncover features prior to earthquakes.

This chapter demonstrated the DEC's capability to identify and cluster high-energy, impulsive signals (the Iwate earthquake) and globally coherent anthropogenic noise. However, the t-SNE analysis revealed poor cluster separation, suggesting the model struggled to distinguish more subtle, unknown features from the background noise.

The next chapter will therefore test the model against a more challenging target: the low-energy, emergent signals of NVT. This NVT dataset is a crucial benchmark, allowing the model's performance to be quantitatively assessed by correlating its clusters against the processed RMS signal—the established proxy for NVT activity.

Chapter 4 - Testing the DEC on Non-volcanic Tremor

4.1 Introduction

What is NVT

Slow EQs (SEQ) are EQs that slip over a longer period than a regular EQ. They are known as slow slip (SS) when aseismic, or non-volcanic tremor (NVT) when seismic. A NVT is defined as many low frequency (LFEQ) and very low frequency earthquakes (VLFEQ) superimposed in time. An NVT activity encompasses many VLFEQs and LFEQs at a particular location. LFEQs are said to be analogous to repeating earthquakes, whereby they are a summation of minor earthquakes along a fault (Nadeau & Johnson, 1999). Long period NVT activity is initiated by a high amplitude body wave of 1-2Hz and followed by a wave-train with a frequency varying between 1 and 10Hz (Obara, 2002).

Reason for applying ML/DL to detect NVT.

NVT is a scarcely investigated phenomenon with the first instance of NVT recorded in Japan in 2002 (Obara, 2002). Since then, there have been many ML applications for additional NVT detection. For instance, Hulbert et al., (2022) used CNN attribution – comparing the CNN inputs in comparison to CNN predictions – to identify slow earthquakes and NVTs which successfully located more NVTs in comparison to the PNSN catalogue, highlighting that CNNs can serve as an effective NVT detection tool.

NVT is emergent and subjective, which makes it difficult for an automatic picking tool to detect (Obara, 2002). Therefore, this chapter will test the DECs ability to capture emergent events. Its emergent nature in comparison to impulsive earthquakes is demonstrated in the time and time-frequency domains in Figure 4.3 (Brudazinki et al., 2010)

Creating an effective NVT detection tool can help assist with scientific questions around the mechanics of NVT. There's uncertainty around whether there is a relation between NVT and megathrust earthquakes. For instance, NVT succeeded the Central Shizuoka earthquake (Mw = 4.8) yet preceded the Western Aichi (Mw = 4.1) earthquakes in Tokai, a region slightly north of Shikoku (Obara, 2002). Likewise, neural networks have been trained on both seismic and remotely sensed data to understand the relationship between slow slip and NVT (McCausland et al., 2005).

The DEC has already demonstrated that it can identify anthropogenic activity and potentially other unknown low energy features prior to earthquakes in Chapter 2. Here, it will be tested against the NVT data to see if it can detect NVT and potentially different forms of NVT.

It is clear from the previous chapter that the lack of ground truth inhibited hyperparameter tuning of the model. NVT provides a benchmark dataset (Bergen et al., 2019), allowing us to test the model in a supervised fashion – similar to Chapter 1. NVT is detected from a pre-processed continuous RMS signal (Figure 4.1). VLFEs and LFEs are emergent with no distinct P and S first arrival and therefore it is extremely difficult to locate NVT hypocentres. As a result, the NVT spectrogram signature is often represented by a 2-8Hz emergent band of varying energy (Figure 3.3). Therefore, the data will be band-passed within this frequency range prior to preparing the RMS signal. This signal will be cross correlated with the encoded/latent cluster signal C[k] (where k is an arbitrary cluster). NVT can range in energy and amplitude (Figure 4.1) and a purpose of the DEC is to detect further NVTs that cannot be seen with the naked eye.

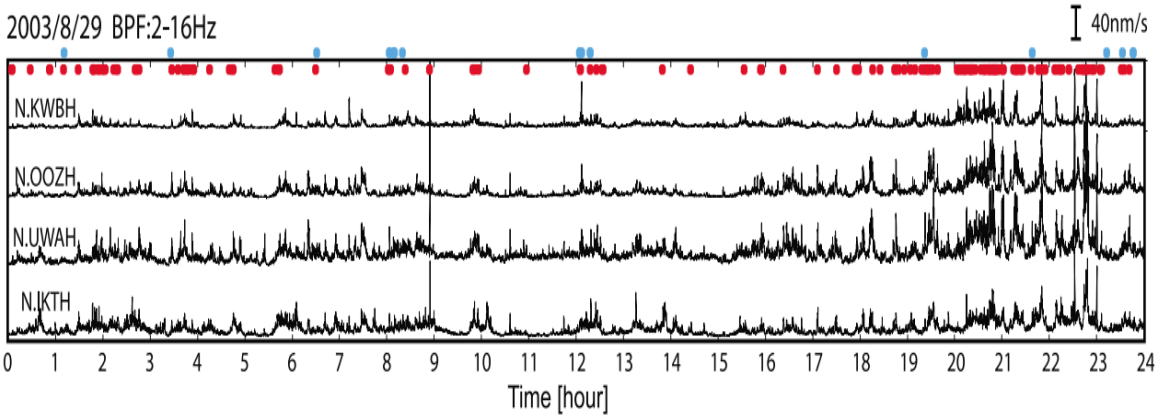


Figure 4.1 from Obara (2010): examples of LFEs (blue dots) and NVT (red dots) derived, and no NVT filtered such that the 1Hz frequency band is omitted. High amplitude NVT can be noted in the later hours for all stations whilst NVT in the earlier hours is hard to spot with the naked eye.

In southwest Japan, NVT often occurs in the plate interface and between the seismic and free slip zones (Obara, 2010). It always occurs in the shallow and deep regions of the interlocked zone (Vannuchi, 2022) which implies it exists in a place where friction is highly variable. For the regions where deep NVT is observed, the hypocentres of deep low frequency earthquakes (DLFs) were estimated to be approximately 30km - near the Mohorovicic Discontinuity. These tend to have more distinct P and S first arrivals (Obara, 2006) and have a dominant frequency of 2Hz. Long period NVT episodes are said to migrate along the strike at approximately 13km/day (Obara, 2002). This chapter will investigate NVT around Shikoku as it has been studied an intense amount by Obara (2002) and hence contains a good amount of ground truth.

Aim of the chapter

This chapter will aim to test the DECs ability to identify further NVT events, among others, in the in the Shikoku region of Japan. It is hypothesized that the DEC may be able to uncover further features about NVT by assigning spectrograms automatically to one or more cluster classes. Analysis will be conducted in the period identified by Obara (2010):

- 8 days from 9th April 2005

The DEC has been trained on 6 instances of NVT episodes from:

- A day in March 2005
- 8 days from 9th April 2005
- 13 days from 6th May 2005
- A day in January 2007
- A day in March 2007
- A day in January 2008
- A day in March 2008

Using different dates this could allow the network to learn about NVT over a longer period and prioritise pattern recognition in NVT over other features (such as short-term anthropogenic activity).

The following stations, shown in Figure 4.4 identified from NVT episodes (Obara, 2010), are analysed:

- GHKH
- INOH
- KTGH
- MTYH

- MURH

The following stations were trained on but not analysed:

- IKTH
- KWBH
- OOZH
- SJOH
- TBRH

To the best of the author's knowledge, this is the first time the DEC model has been applied to detect and understand further NVT.

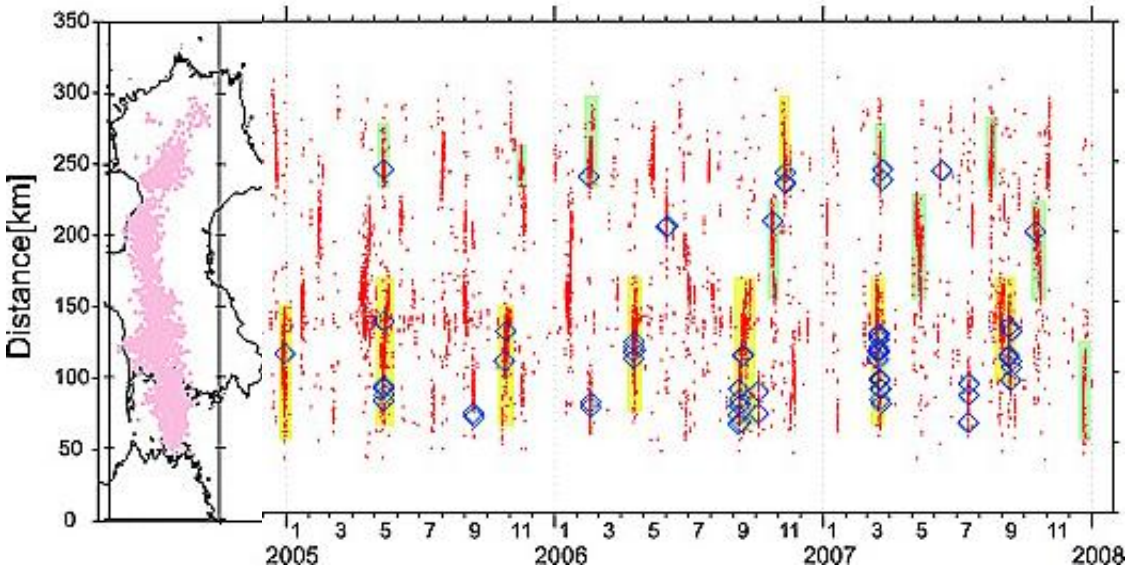


Figure 4.2. Different NVT episodes in the Shikoku area used to extract NVT from Obara (2010) highlighted by the blue boxes. It can be noted that NVT episodes can exist at various lengths along the strike although there is a level of consistency in some places, especially northern Shikoku.The original image was cropped for the time period being considered here.

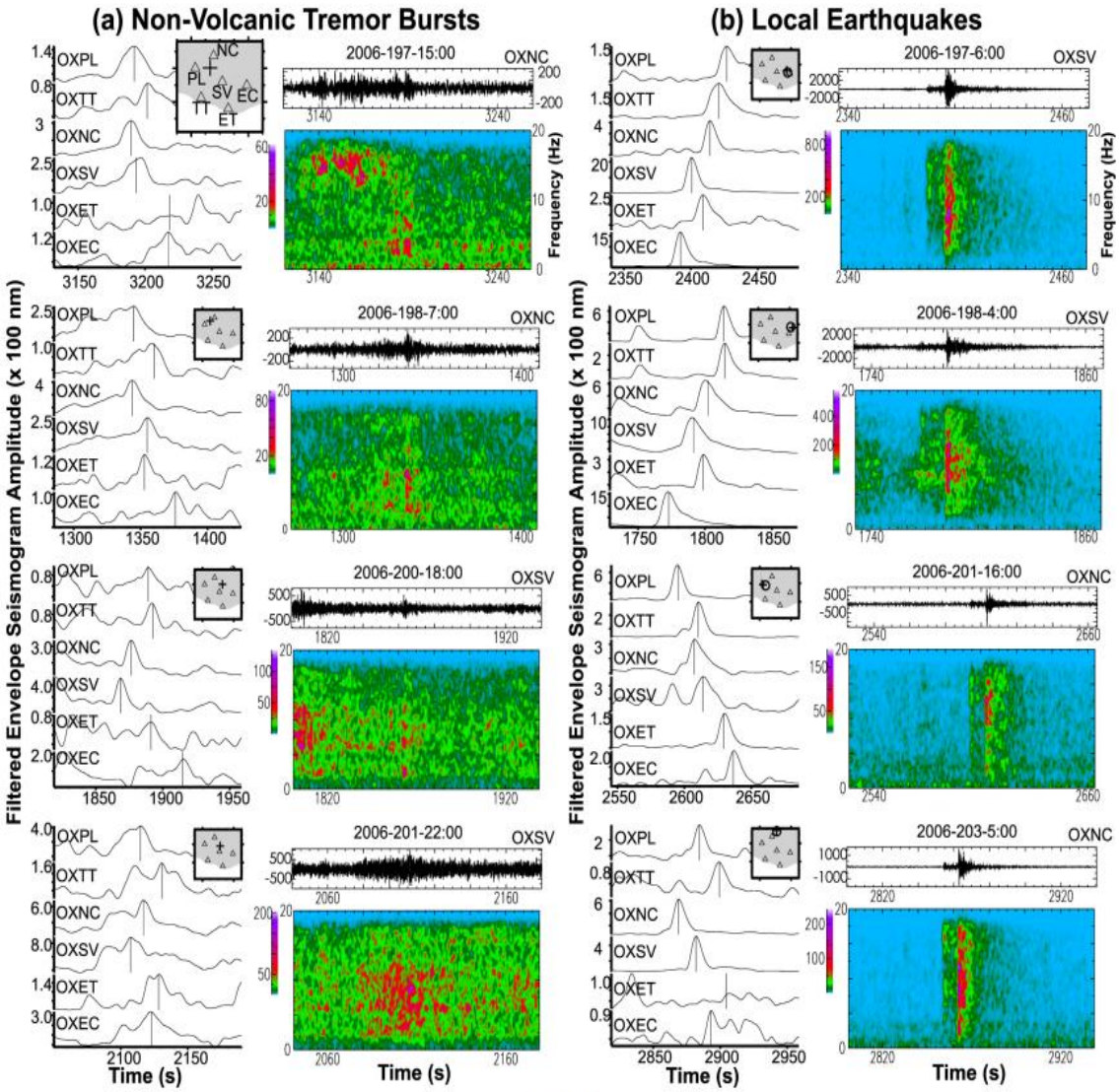


Figure 3

Figure 4.3: left are examples of NVT and right examples of earthquakes from Brudzinksi et al., (2010)'s research into NVT along the middle America subduction zone. NVT is observed here to be emergent and exists over longer time ranges with a good proportion existing in the same frequency range as EQs.

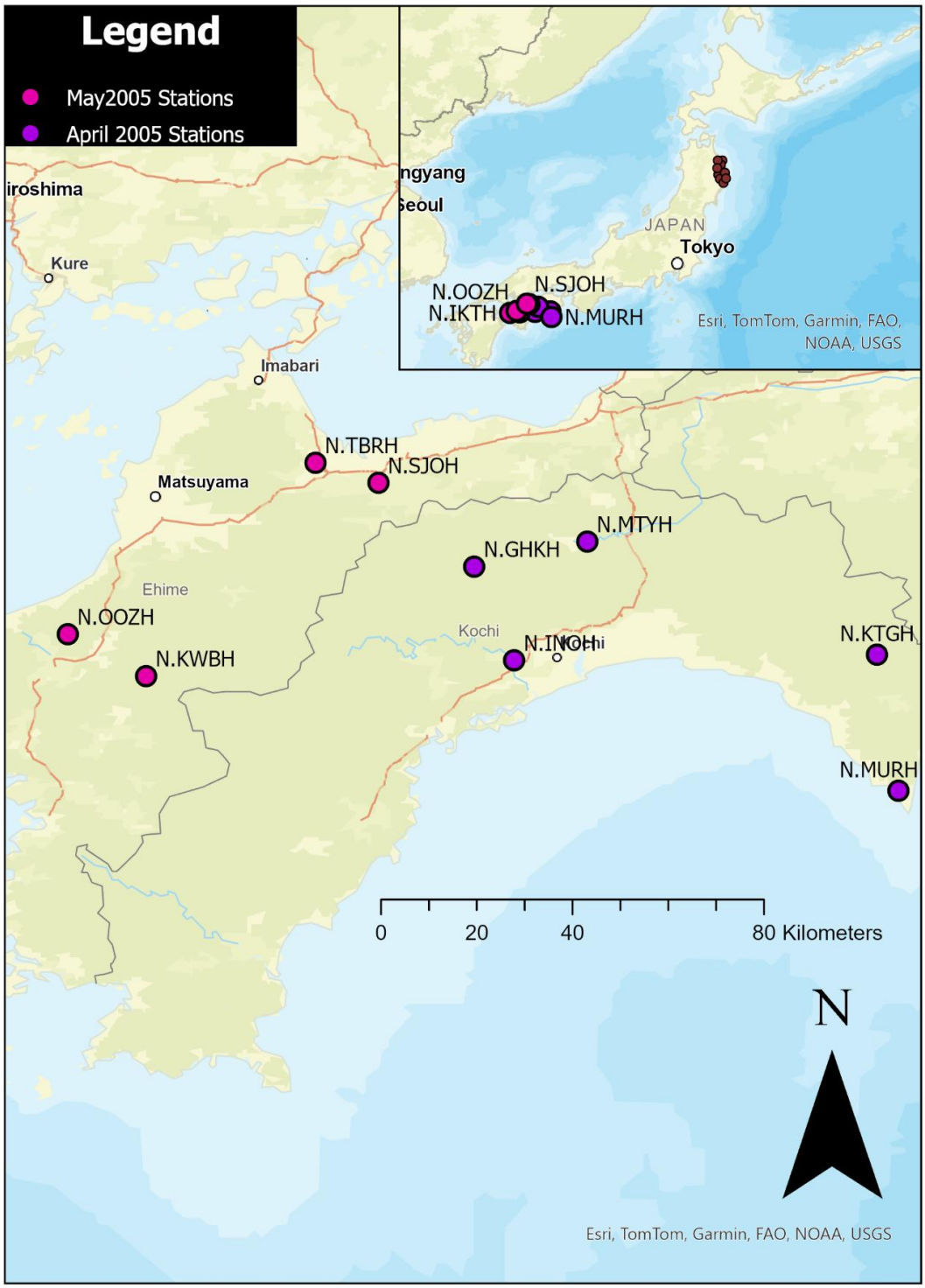


Figure 4.4 : a map of the stations used in April 2005 and May 2005. The top right map shows this study location in comparison to the location used in Chapter 2.

4.2 NVT specific methodology

As seen, NVT was first identified by calculating the envelopes of the root mean squared (RMS) error amplitude (Obara, 2002).

RMS is used is to detect emergent signals and is representative of energy, a good proxy for NVT. However, some NVT is low energy, which means RMS may not always be the appropriate method for detecting NVT. It is hypothesised that the DEC will be able to distinguish NVT episodes of varying energies.

Data was manually extracted for Shikoku using the HiNet Python 3.0 Client() class. Datasets that did not contain digits or not satisfy the expected length were omitted. The omission rate was below 5% for most stations.

This section covers methodology specific to this chapter (Figure 4.5). It mostly takes inspiration from Obara (2002)'s preprocessing. There are two signals derived in preprocessing from the seismic timeseries: the original signal (converted into spectrograms and used to train the DEC model) and the supervised signal (NOT used to train the DEC model and instead used to test the clusters from the DEC).

The original signal is sent to preprocessing and is high passed using the Butterworth Filter (same parameters as Chapter 2), allowing effective comparison with Chapter 2 plus allowing the DEC to capture the 1-2Hz body wave. This frequency range omits the dominant ~1Hz signal seen in Chapter 2 and then cropped down to 8Hz to satisfy the number of frequency bins needed.

RMS (Root Mean Squared)

During DEC training, the supervised signal is then detrended and band-passed between 4 and 12 Hz and then the RMS function is applied. This gains an average of total amplitude over a given rolling window. The RMS of a discrete signal $\mathcal{V}(t)$ is given by:

$$\mathcal{V}_{RMS}(t) = \left| \sqrt{\frac{1}{N}\sum_{i=1}^{N} \mathcal{V}^2(t)} \right| \tag{4.1}$$

Where N is window size which is equal to the sample rate (f_s) multiplied by the duration P , $\mathcal{V}$ is wave amplitude, t is time.

The signal is then decimated to 120 second windows to save data as the presence of NVT diminishes after 12Hz. This does not remove "spikes" in the data, which could correspond to EQ or impulsive events / noise which could potentially mask low energy NVT. Initially, a low pass filter was applied to the data but this removed a large proportion of the 4-12Hz component. To avoid any removal of potential NVT information, the peaks were manually removed using an inbuilt SciPy module where a peak is defined as any point exceeding the maximum height of the signal prior to it. As impulsive events exist in a small time period of less than 120 seconds, every peak is assigned to the datapoint before it returning a flat line.

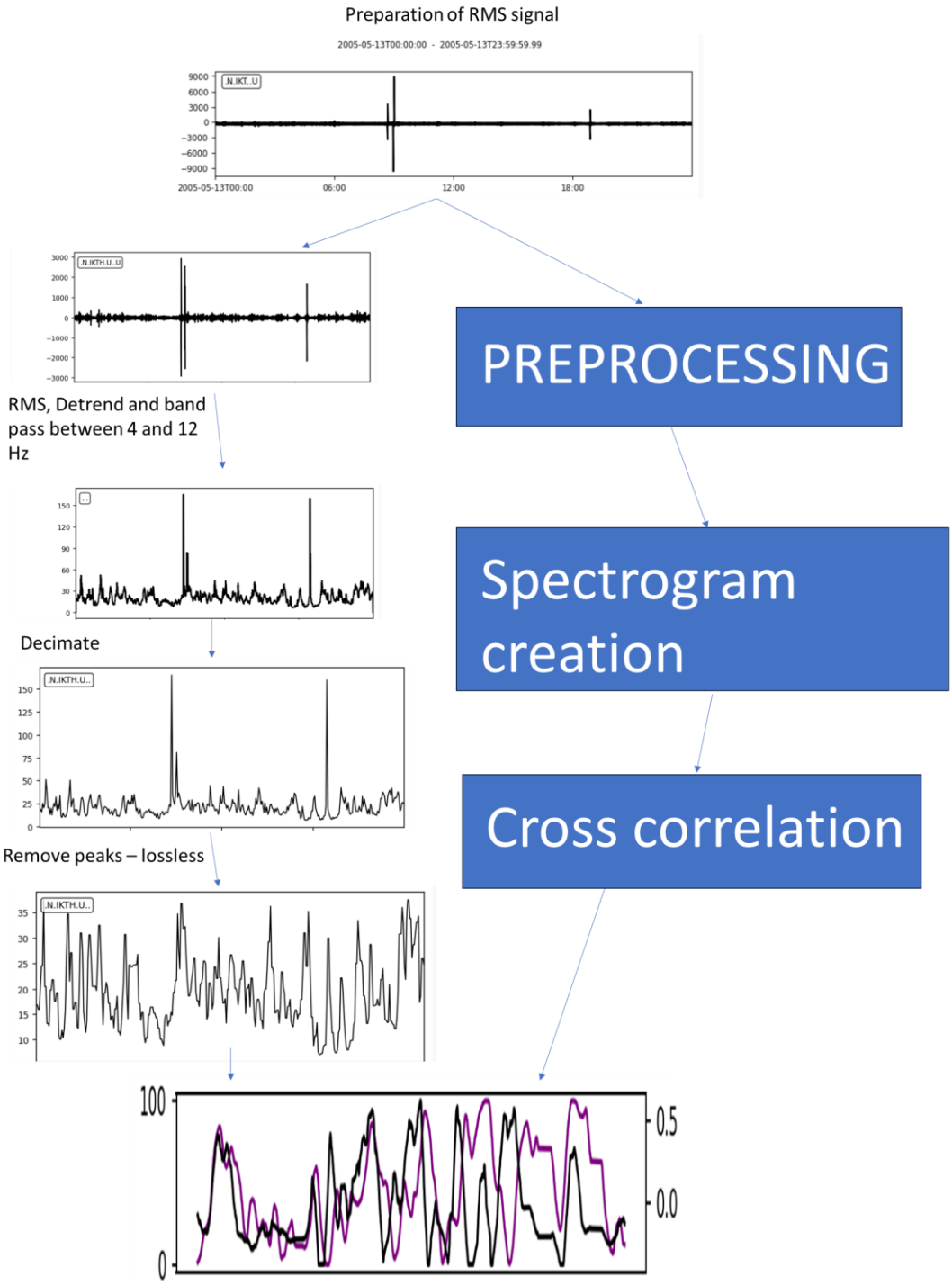


Figure 4.5: preprocessing of the RMS signal. This describes two processes on the original data. 1) the RMS signal – sampled and converted into a spectrogram - trains the DEC 2) the RMS signal undergoes detrending, decimation, and peak removal which is a method that is traditionally used to visually identify NVT. The bottom picture demonstrates a visual comparison between the cluster in time and the RMS signal.

McCausland et al., (2005) plotted the envelope of the RMS when investigating NVT. This method is omitted as it cannot be easily correlated with the cluster signal in PCM analysis. Instead, the Savgol filter was applied to make the signal more compliant with the cluster signal which also received Savgol smoothing (see Chapter 2 for a description of Savgol smoothing).

PCM analysis

PCM analysis is conducted similarly to Chapter One. However, as NVT is often attenuated between stations, PCM analysis for NVT will be conducted at individual stations only. As the RMS signal is sampled for 120 second windows, for 8 days there will be 5760 samples. So, for a two-tailed test n_{hat} = 5760, meaning the PCM criterion is now 1.114 for a 1% significance. However, due to Savgol smoothing with a window of 100, this reduces n_hat to 57 resulting in a critical PCM value of ±1.250 for a 5% confidence.

As the DEC is being used as a NVT detection tool that could potentially unveil further information about the NVT, a smaller number of clusters were used here. Therefore, K was set to 4 (Table 4.1) with the colours used for each cluster shown in Table 4.2.

Results will be presented for the following dataset:

Total instances	Training instances	Validation instances	Stations	P, Window	Window overlap	K, number of clusters	Latent space $\mathbb{L}$ size	Fitted to
120000	**92641**	**30000**	**9 (4 in April)**	**120 seconds**	**40 second**	**4**	**14**	**TrD**

Table 4.1: information about the data the model is trained on and analysed with.

C0	C1	C2	C3

Table 4.2: cluster acronyms and their respective colours for this chapter

4.3 Results

Training – loss/cost function

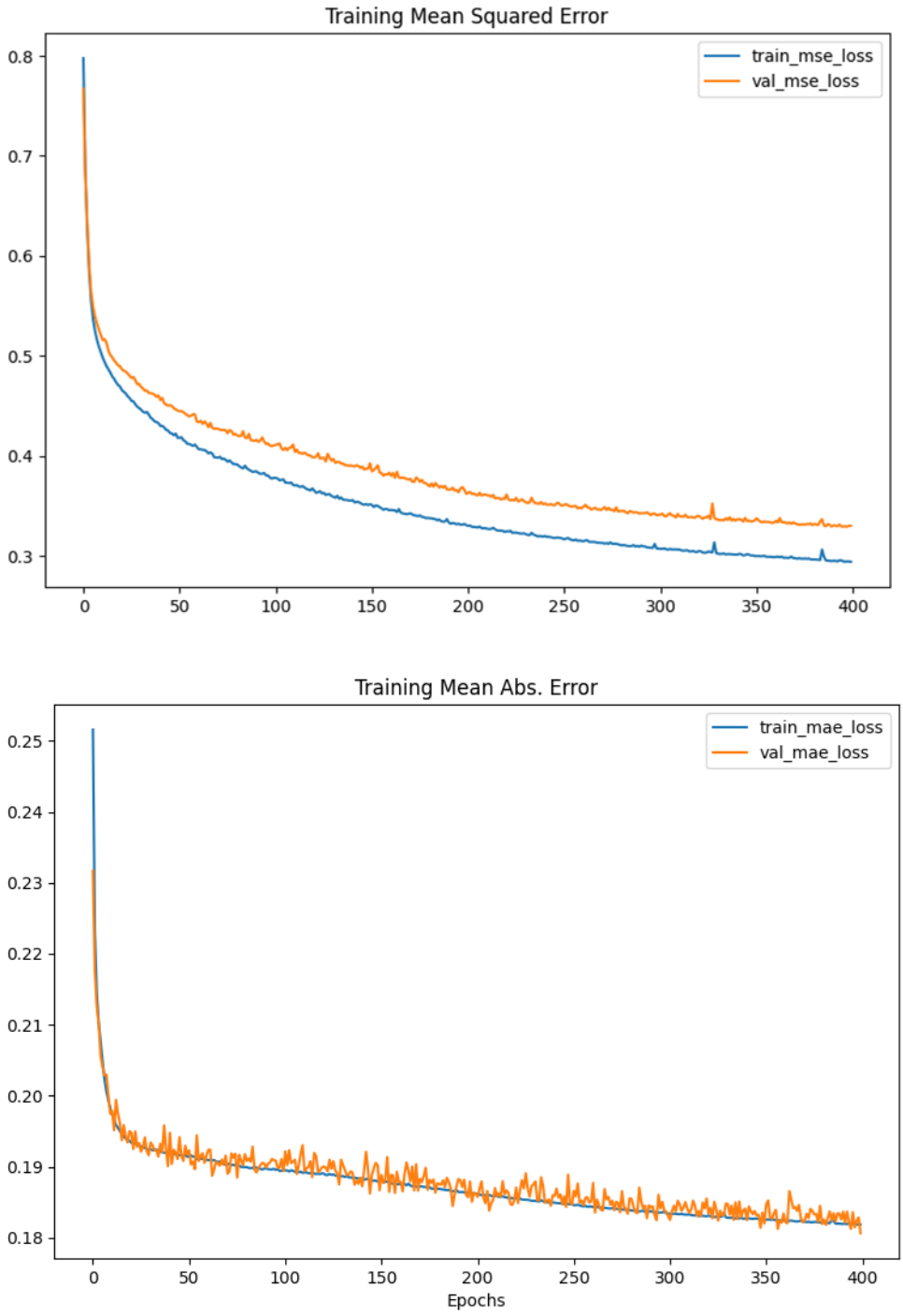


Figure 4.6 (top): Total reconstruction loss because of MSE, ii (bottom) training mean absolute error (MAE). The MSE and MAE for both training and validation datasets broadly aligns which suggests that the DEC learning is comprehensive.

The total loss was around 1.3 for the training dataset whilst the validation dataset's loss was 1.05 larger (Figure 4.6). This is more compliant with the Snover et al., (2021) study. The difference in validation and training loss is representative of a small generalization gap but is not representative of overfitting (Hao et al.,

2024). Perturbations in the validation MAE is consistent with what was seen in the Snover et al., (2021) study and could be the result of random difference between validation and training datasets.

Reconstructions

There is greater variability in the reconstruction of original spectrograms in comparison to Chapter 2. Not only are high amplitude features preserved, but some reconstructions contain impulsive features that look both natural (Figure 4.7) and non-natural (Figure 4.9) impulsive features, whereas others reconstruct noise to multiple bins, that were originally scattered and are combined into a singular feature (Figure 4.8).

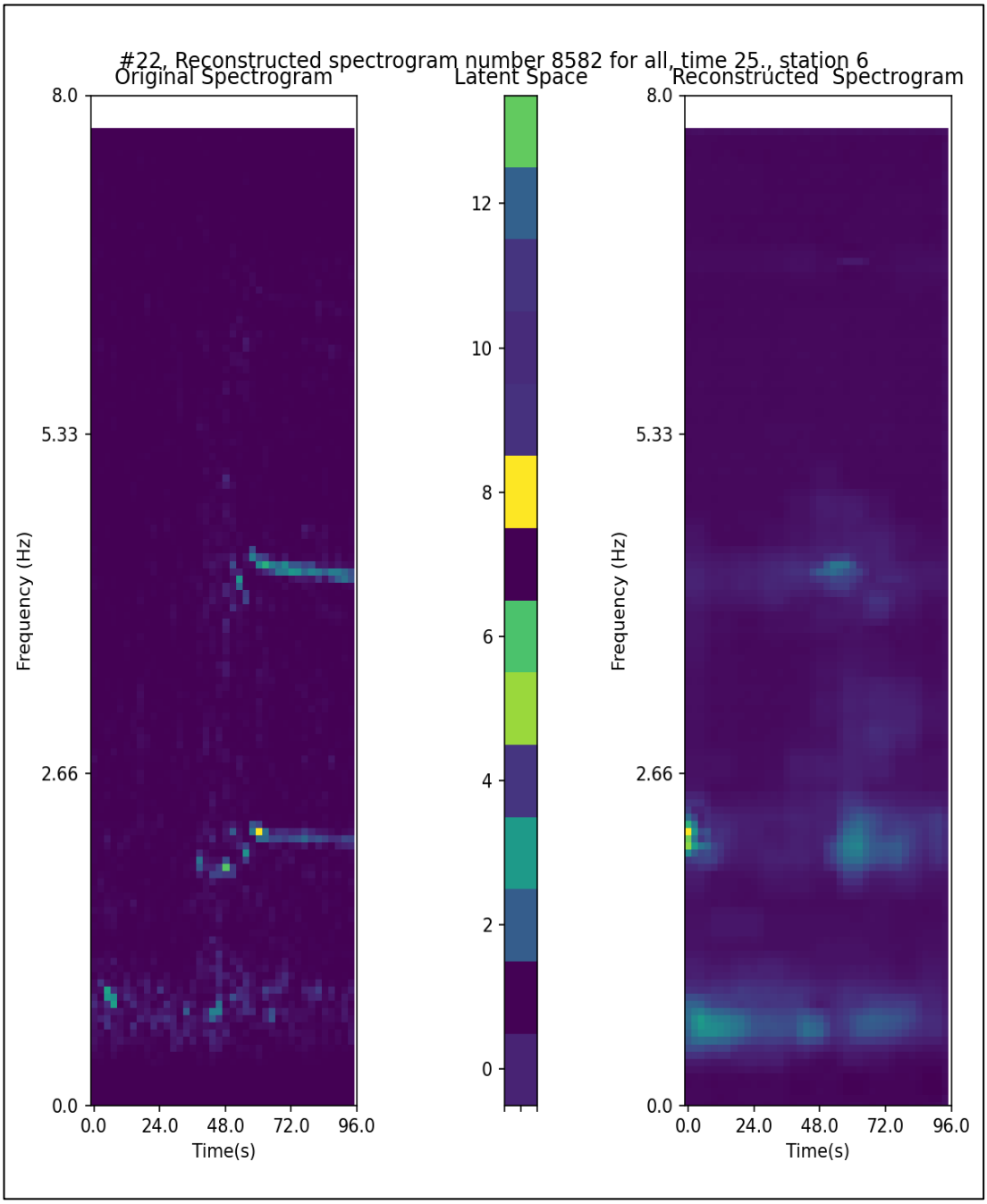


Figure 4.7: Reconstruction demonstrating impulsive features, which could represent an earthquake in this instance.

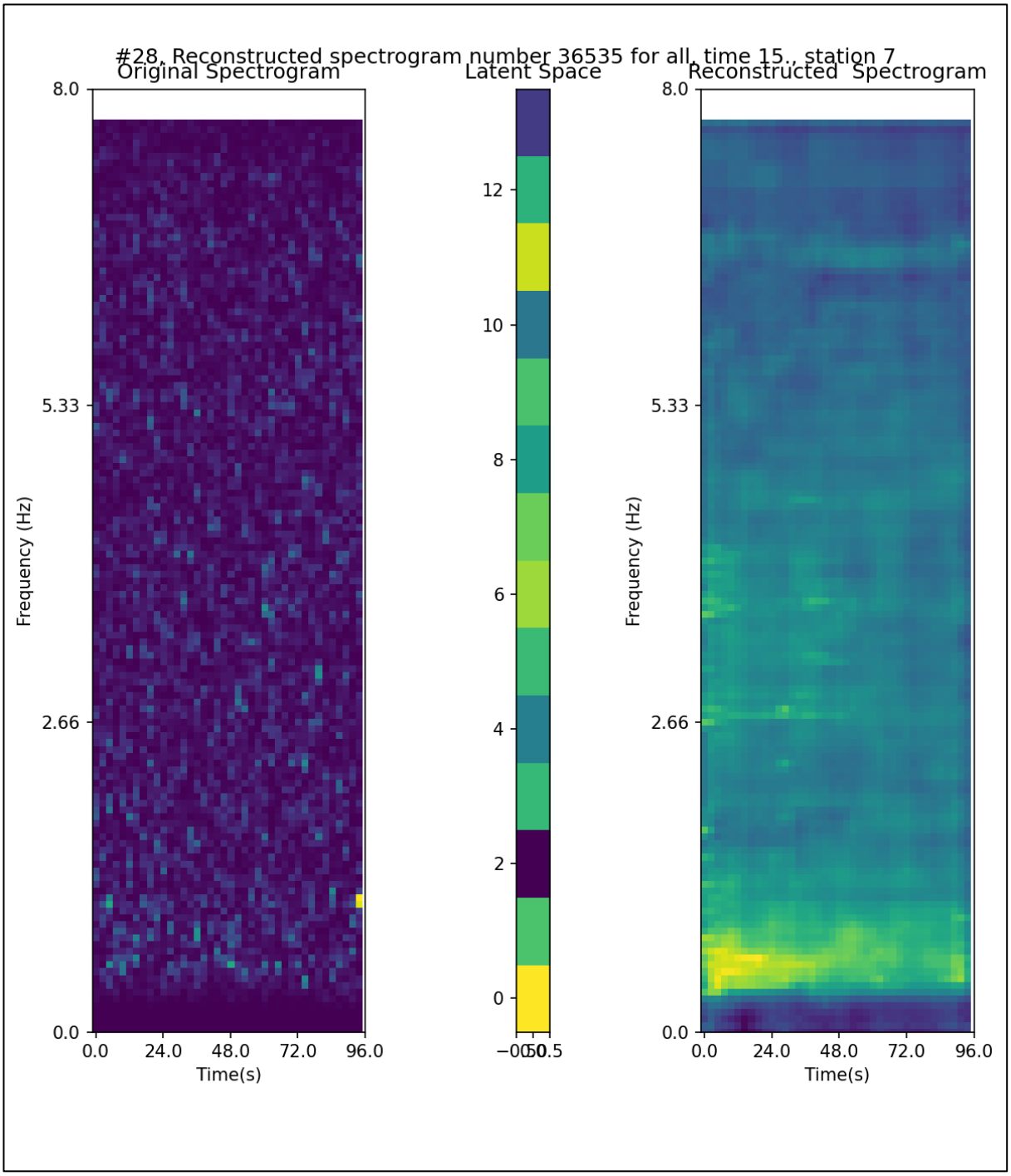


Figure 4.8: Reconstruction of a noisy spectrogram, which could represent NVT.

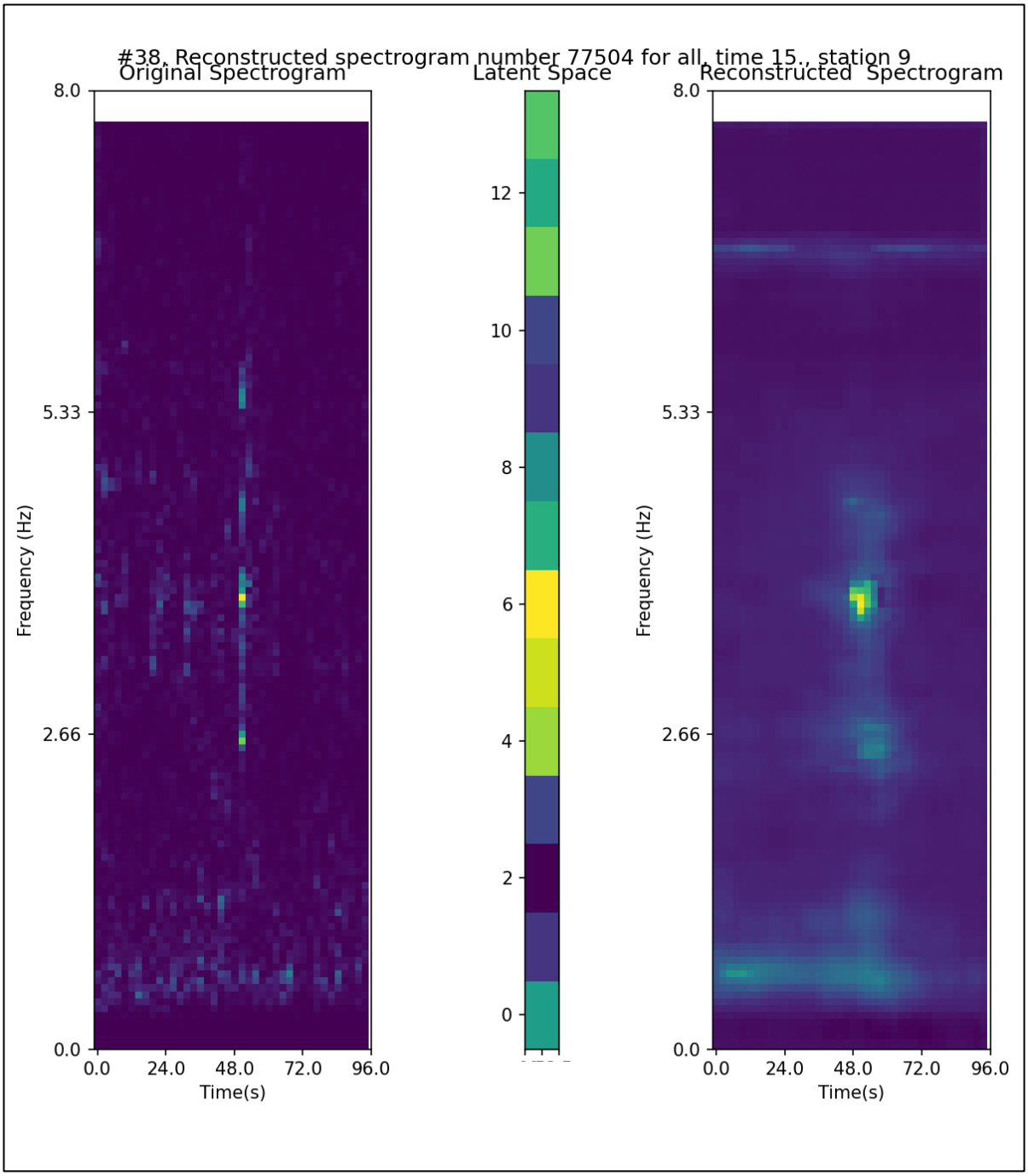


Figure 4.9: Reconstruction of a highly impulsive feature, most likely the cause of a non-natural impulsive source.

Observations from ReSp

Latent Vector LV_0 and LV_{11} are sensitive to spectrograms that have been reconstructed into a block-like shape (Figure 4.8). For these spectrograms, their original frequency range (which is between 2-8Hz) and non-impulsive nature is consistent with NVT. A good proportion of the data is an unnatural looking flatline sinusoid with harmonic multiple (see Appendix). This artificial or erroneous signal was created by ObsPy or electrical noise at the station or could represent a physical feature of unknown origin. There were several impulsive events that were successfully reconstructed. At this stage, it is not known if they are EQs or impulsive noise. A dominant low band signal remains to some extent in some reconstructions. The filter's order is 3 to stop ringing in the data therefore resulting in preservation of the dominant <1Hz signal. This band is often reconstructed.

Clusters

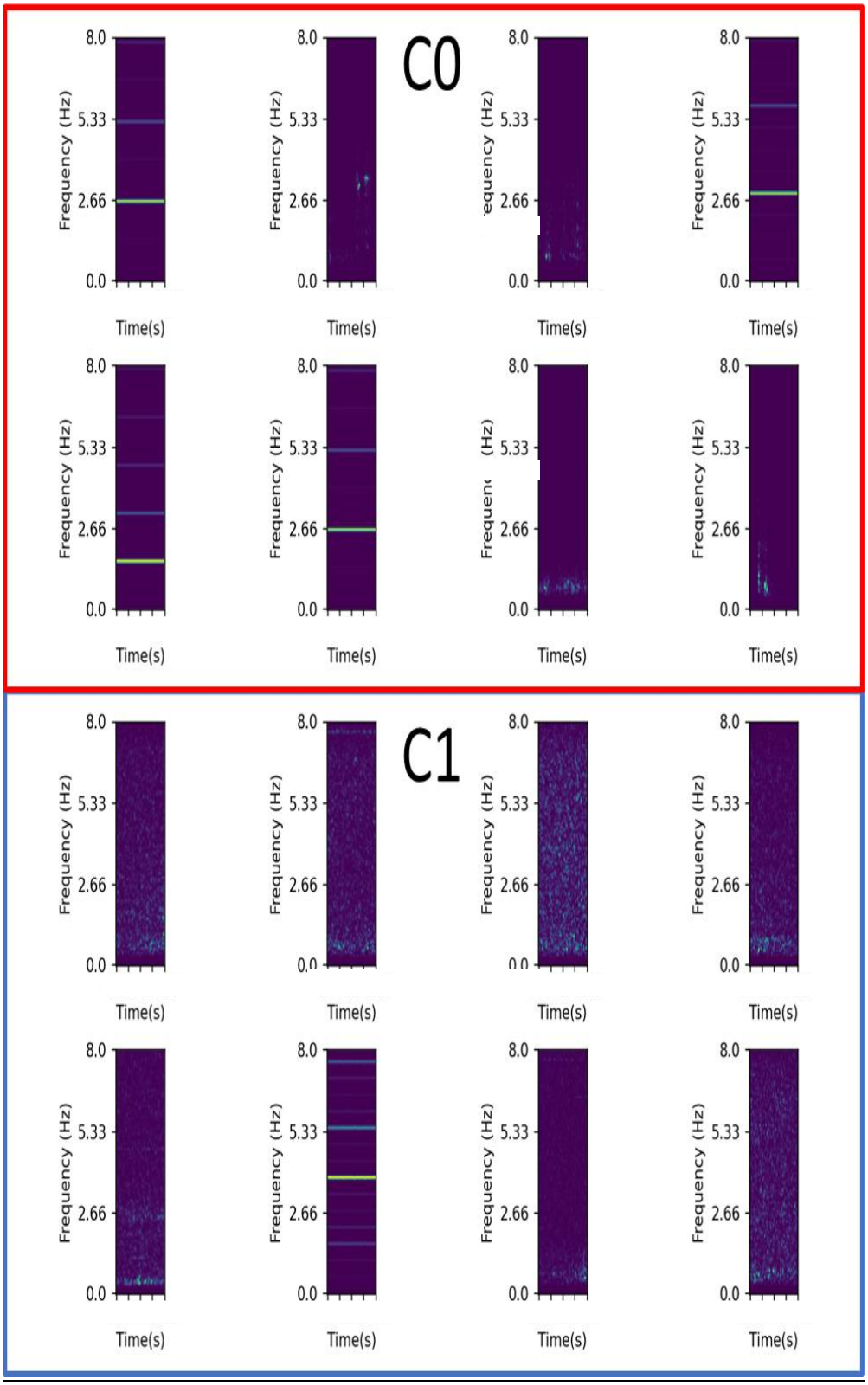


Figure 4.10 I-II: Example instances of clusters 0 and 1 contained within the training dataset.

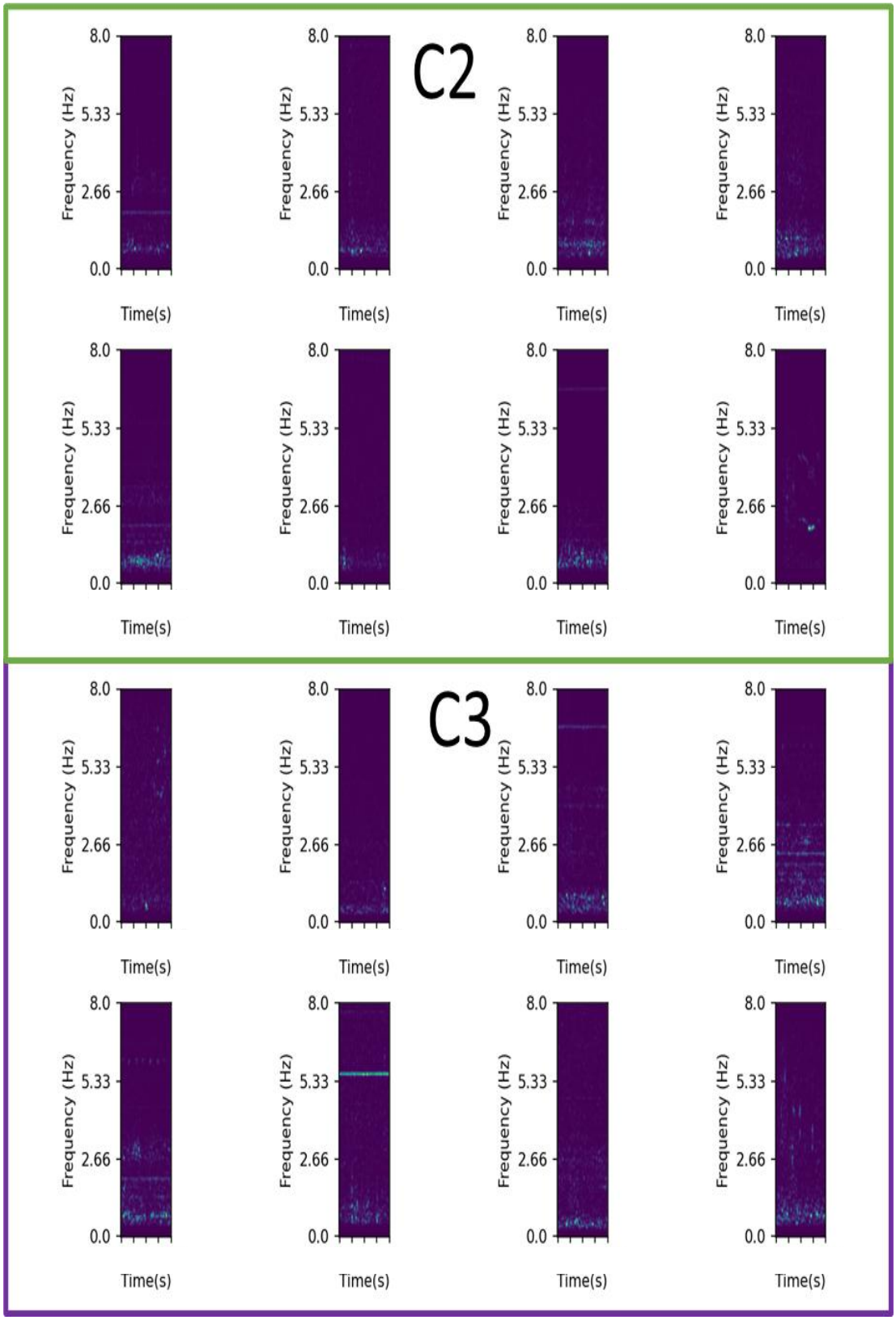

C2
C3
Frequency (Hz)
8.0
5.33
2.66
0.0
Time(s)

Figure 4.10 III-IV: Example instances of clusters 2 and 3 contained within training dataset. Length of spectrograms are at a constant time length of 240 seconds. See Figure 3.6 for more details.

Observations from clusters

The table below summarises results from the clusters in Figure 4.10. Some of the clusters identified visible tremor whilst others did not. It must be noted that there was a distinct difference between impulsive features (C0 and C2 for no tremor) and the frequency of the features seen (C1 and C3 for indicators of tremor).

Cluster	NVT observed? (Clearly/intermittent/no)	General comment
0	No	Mostly impulsive and artificial signals.
1	Mostly	Lots of strong high amplitude noisy data up to 8Hz.
2	No	Mostly lower frequency signal shown with a few isolated patches which could represent noise. This cluster represents low energy activity.
3	Intermittent	Some higher energy signal that could be dominated by NVT. There is noticeable medium amplitude sinusoidal signals seen in the dataset. Some of them exist with noisy data suggesting the issue could arise with the station and not the preprocessing of the data. There are some impulsive signals seen which could indicate earthquake activity.

Table 4.3: a summary table describing whether is observed in the clusters, and the nature of the signals seen in the spectrograms.

Clusters in space and time

ALL stations

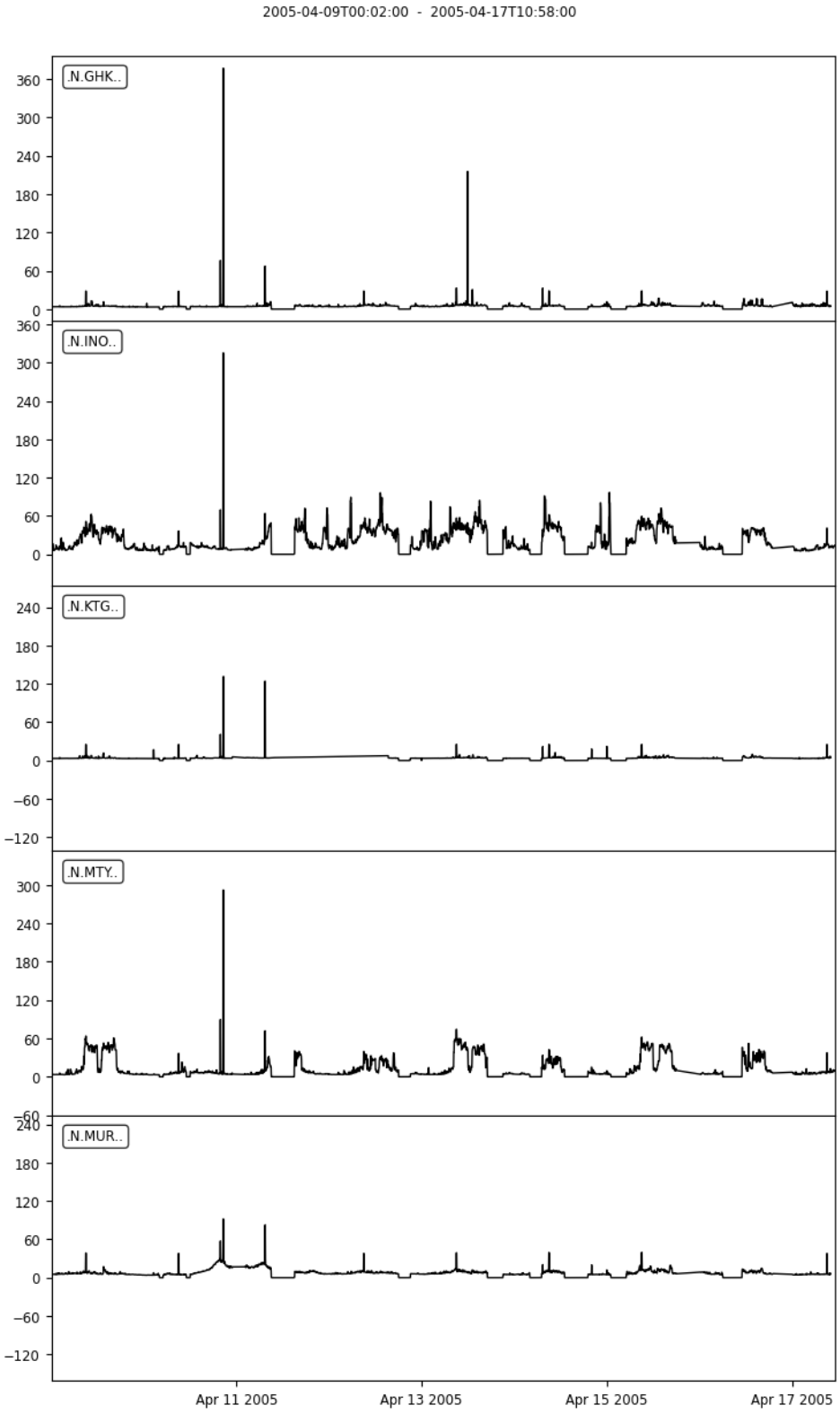
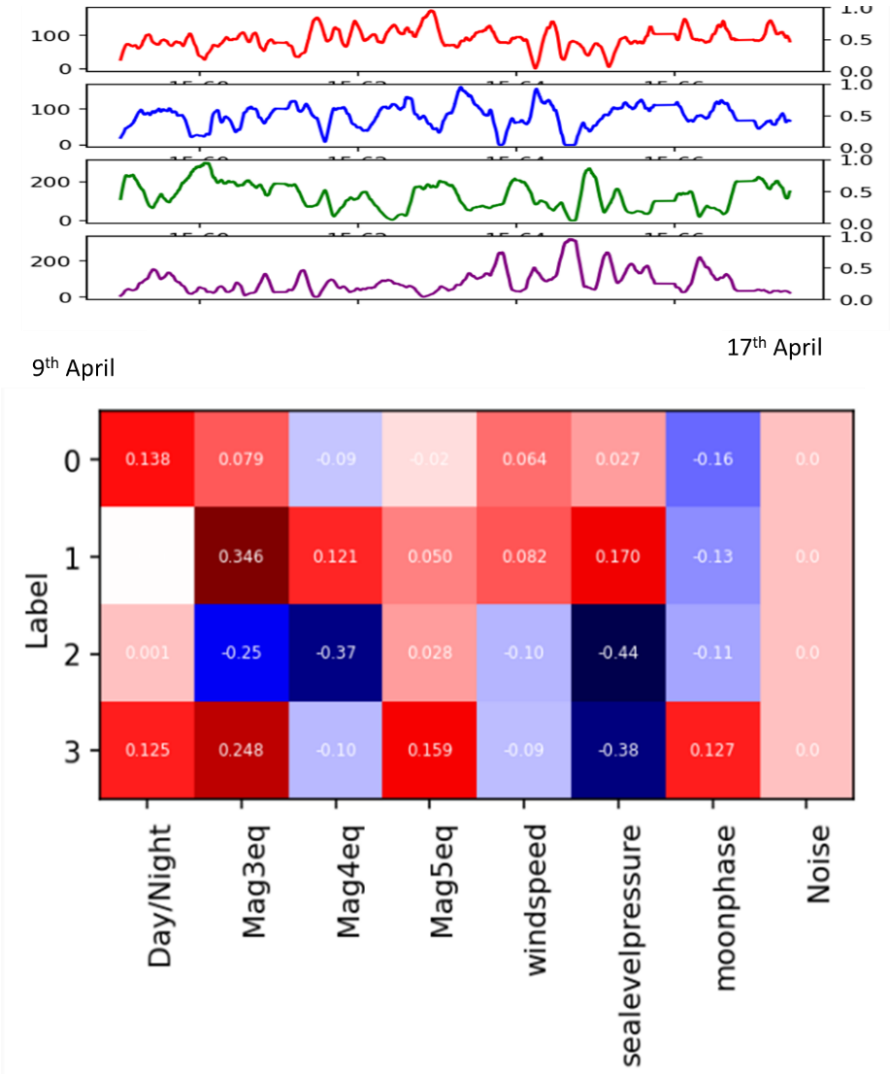


Figure 4.11: RMS signals across ALL the stations for April, GHKH (top), INOH, KTGH, MTYH, MURH (bottom), which shows quite some variability in the magnitude of NVT signals among different station. Note the flat lines observed which could potentially contribute to the artificial signal seen and the peak removal has not been applied here.



Figures 4.12 I, II (top to bottom): Cluster occurrences across the period from the 9th April to the end of the 17th April which demonstrate a level of incoherent periodicity along with flat lines corresponding potentially to artificial signals. Note that the left-hand y axis demonstrates occurrences. The bottom figure represents PCM values for features against label.

At first glance, the result of all stations (Figure 4.12 I-II) does not suggest a major correlation with any feature, especially day and night. There is potential activity around the 15th of April which excites C[2,3]. For C2 there is a an out of phase correlation for Mw>4 EQs whilst C3 suggest an in-phase correlation. C1 however shows the strongest correlation across the entire series with significant PCM value of 1.346 for Mw>3. Clusters 2 and 3 show a strong anticorrelation with sea level pressure. And C1 showed a moderately strong correlation with the number of Mw>3 or more EQs potentially suggesting that this band could also represent EQ and NVT activity.

Singular Station Set

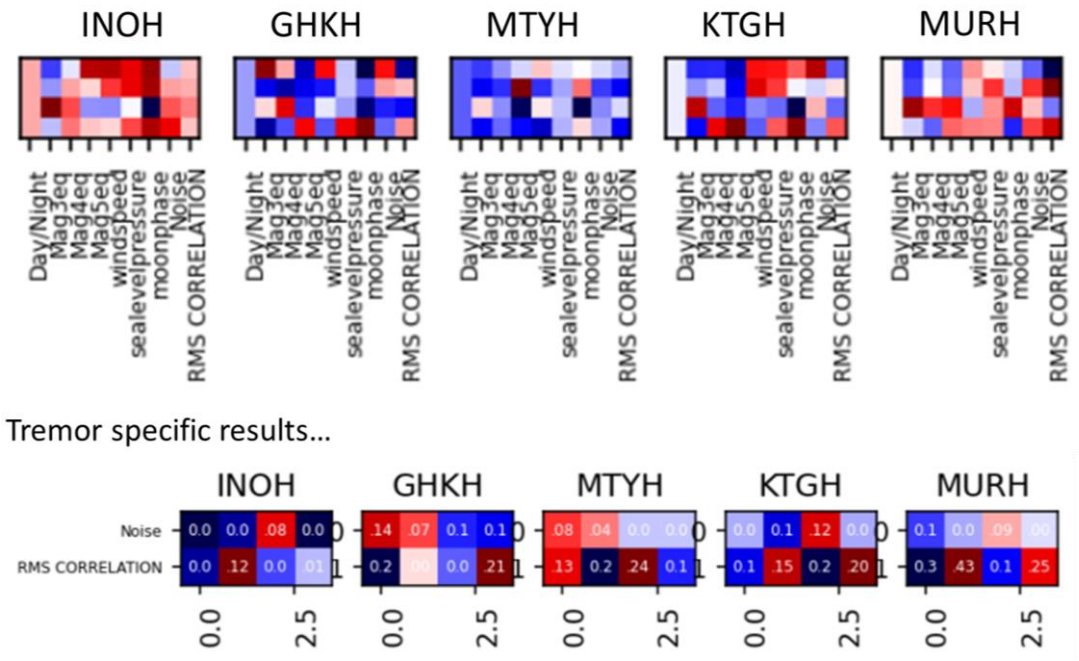


Figure 4.13 I-II: top shows PCM values for the singular station set, bottom (II) shows PCM values for the singular station set for only RMS and noise (rows) and for the clusters 0 (left) to 3 (right) (columns).

Observations

The 90,621 training instances (Table 4.1) were cross correlated for 5 stations using the supervised RMS as seen in Figure 4.12 II. The peak removal was not perfect. Therefore, in analysing the results, these peaks were manually removed to allow the signals to be normalised in amplitude for PCM analysis. Stations INOH and MTYH show recognisable NVT (Figure 4.13 II). They have emergent signals consistent with that seen in Obara (2002), whilst there is potential NVT in the other 3 stations.

There is a large variation among all the stations similar to Chapter 2. There is no day / night signal, this could be due to the lack of day / night instances on a training dataset that spans a longer time period. Despite RMS corelating well amongst C2 and C3, there are no specific clusters that correlates well with RMS, potentially indicating NVT is station dependent. There are 8 (40%) clusters that surpasses the 2 tailed criterion for 5% confidence of ~1.210 (Table 4.4):

Station	C0	C1	C2	C3
1 (GHKH)	-1.21			1.21
2 (MTYH)		-1.2	1.24	
3 (KTGH)		-1.2		
4 (MURH)	-1.3	1.43		1.25

Table 4.4: cross correlations between each cluster and the RMS signal.

Station: 0 (INOH)

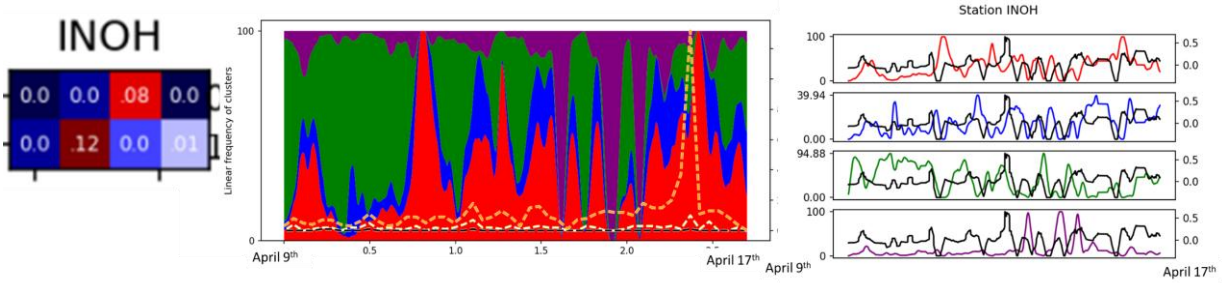


Figure 4.14. I, II, III: The left figure represents PCM for noise (top row) and RMS (bottom row). The middle figure is a stack plot wherein the dash lines correspond to earthquakes with orange, white and black corresponding to Mw > 3,4,5 respectively. Right figure are cluster occurrences for the 8 day period with the RMS in black. This scheme will remain the same among these figures.

Observations

NVT exists at this station during this time period in addition to a number of periods that represent a flat line (Figure 4.14 I). There is an extreme perturbation in C3 around April 15th. There is a strong correlation between C1 and RMS and its frequency is moderate in comparison to other clusters. C2 visually correlates with a flat lined RMS around April 11th.

Station 1 (GHKH)

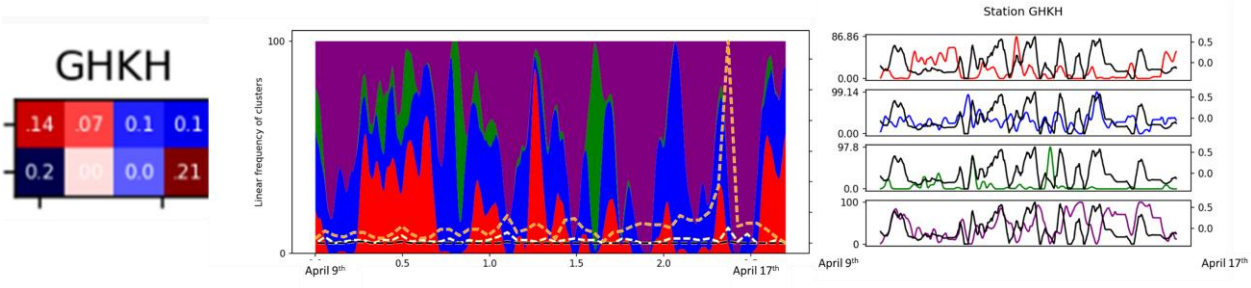


Figure 4.15 I-III: for station 1 (GHKH)

This station demonstrates a number of spikes around a low amplitude signal in Figure 4.15. There are large perturbations in the cluster signals. C0 is out of phase with RMS signal. C3 has a larger visual correlation with the RMS than its moderate PCM value of 1.21. Although, there is a phase difference that becomes apparent between April 14th and 17th – this could be due to the off shift caused by the moving average in C3 and disturbance from perturbations around April 15th.

Station 2 (MTYH)

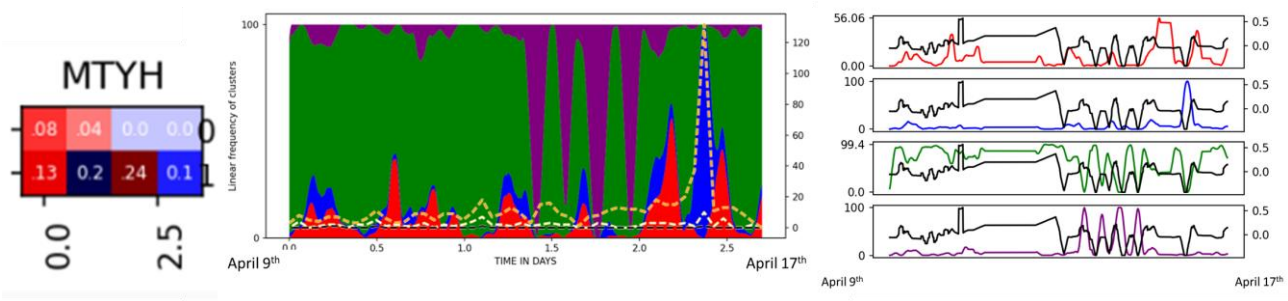


Figure 4.16 I-III : station MTYH

Observations

This station demonstrates NVT in Figure 4.16. The first half of this signal is dominated by C2 picking up on the artefact in the RMS signal. This station is the likely root cause of the resonant sinusoid signal. There are extreme out of phase perturbations between C2 and C3 seen around April 14th eliminating C1 and C1. C1 is sensitive to increased Mw>3 seismicity around April 16th. C3 is in phase with oscillatory RMS.

Station 3 (KTGH)

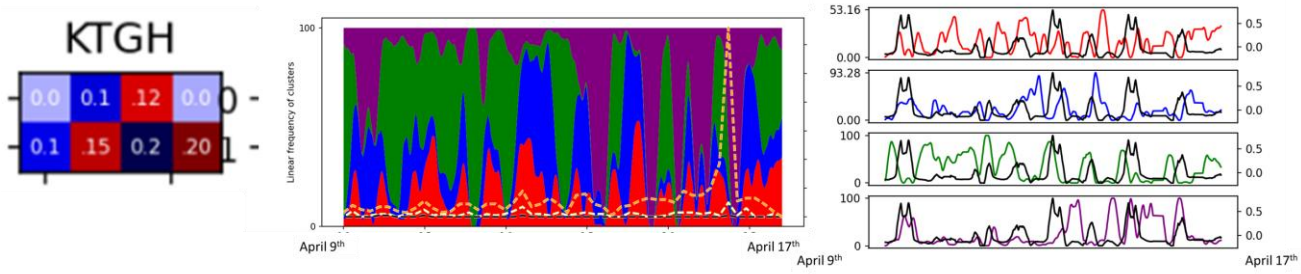


Figure 4.17 I-III: Station 3 KTGH

Observations

C3 demonstrated some activity that could represent NVT in Figure 4.17. C3 correlates with RMS but continues to become out of phase in the second half of the series due to perturbations. C2 is somewhat out of phase with the RMS signal.

Station: 4 (MURH)

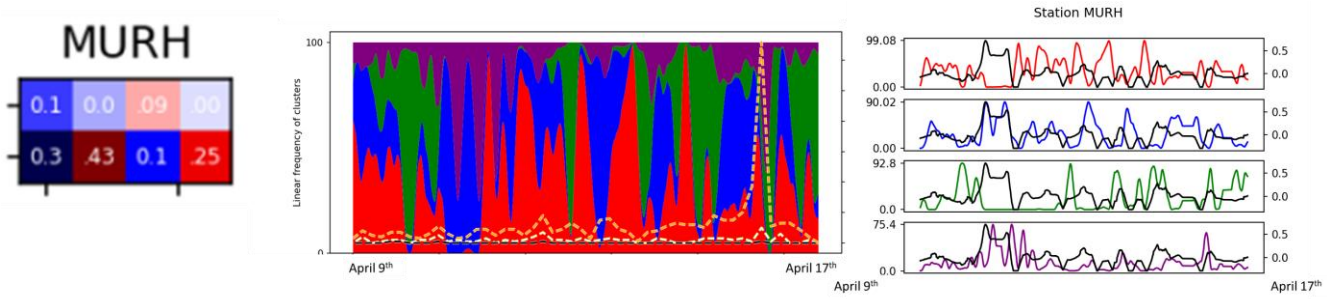


Figure 4.18 I-III station 4 MURH

Observations

There is a large out of phase correlation of C0 with RMS and a large correlation with C1 with RMS at 1.43, which picks up the high energy event at the beginning of the series. There is a good visual correlation with C4 which is at the first half of the series picking up on a large burst of energy. The last half of the series demonstrates lower energy perturbations that do not correlate with RMS, although demonstrate a strong in-phase PCM value of 0.25.

t-SNE analysis

t-SNE analysis on the encoded latent space (Figure 4.19) revealed a greater distinction than clusters in Chapter Two. There was a distinct cluster at the bottom that was dominated by C2 and C1. Both of these clusters demonstrated a stronger correlation with day and night but the PCM values were still below 1.2. Given these two clusters had poor correlations with the RMS, this cluster at the bottom could be

representative of data that does not represent NVT. The dataset entirely consists of NVT episodes, so a good proportion of the spectrograms will contain NVT. C2 demonstrates a strong correlation with erroneous/flat lined RMS signal for Stations 0 (INOH) and 2 (MTYH), suggesting this area could comprise of erroneous spectrograms. There are a few smaller clusters at the top of the dataset.

The Scikit learn default perplexity is 31. Increasing this parameter made little difference to the number of clusters, whereas decreasing it eliminated the southern island seen in Figure 4.19. This suggests the latent space can be represented by a random sample of 3% of all instances. A higher perplexity along with a greater sample size represents two clusters only and reduces the distance of the northern specs to the main body (Figure 4.20). This southern island in the t-SNE data is hypothesised to be the "null" spectrogram set consisting mostly of spectrograms that do not represent NVT. Both C4 and C1 demonstrate NVT and correlate with the RMS, they also exist close to each other in the t-SNE analysis suggesting NVT dominates many components in the embedded latent space.

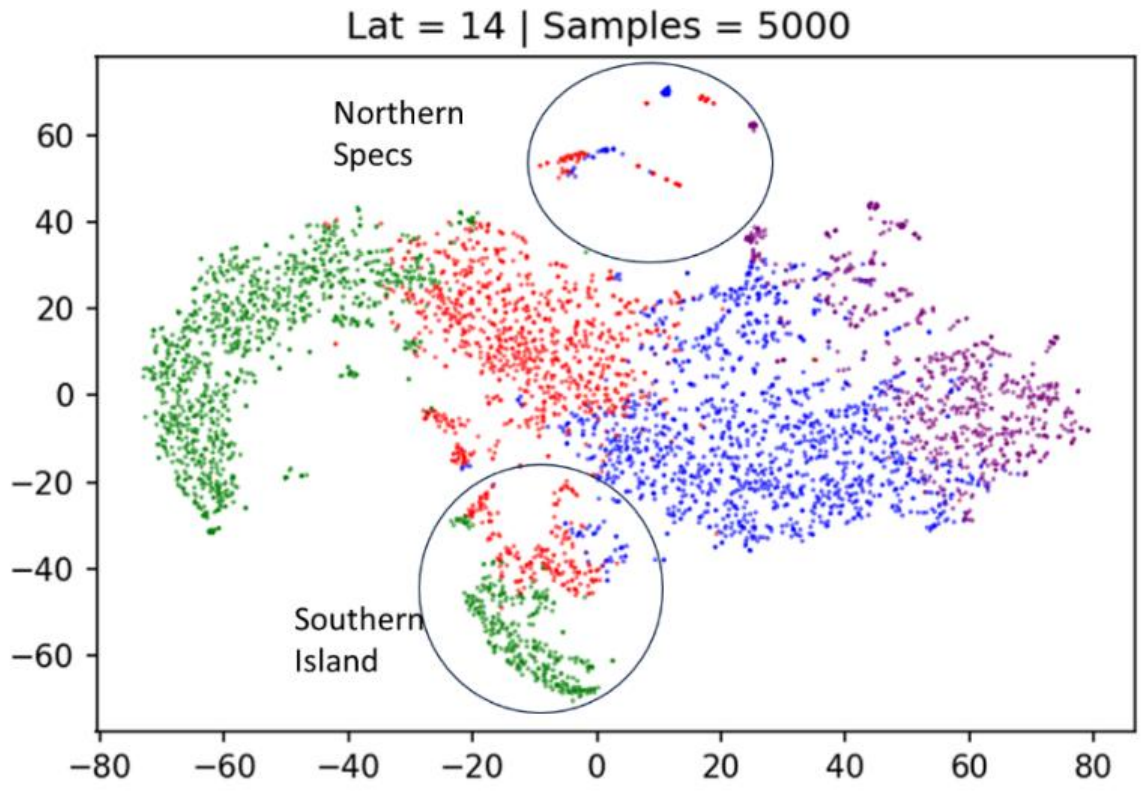


Figure 4.19 : t-SNE analysis on the dataset of a random sample of 5000 points in the encoded latent space coloured with their cluster assignments.

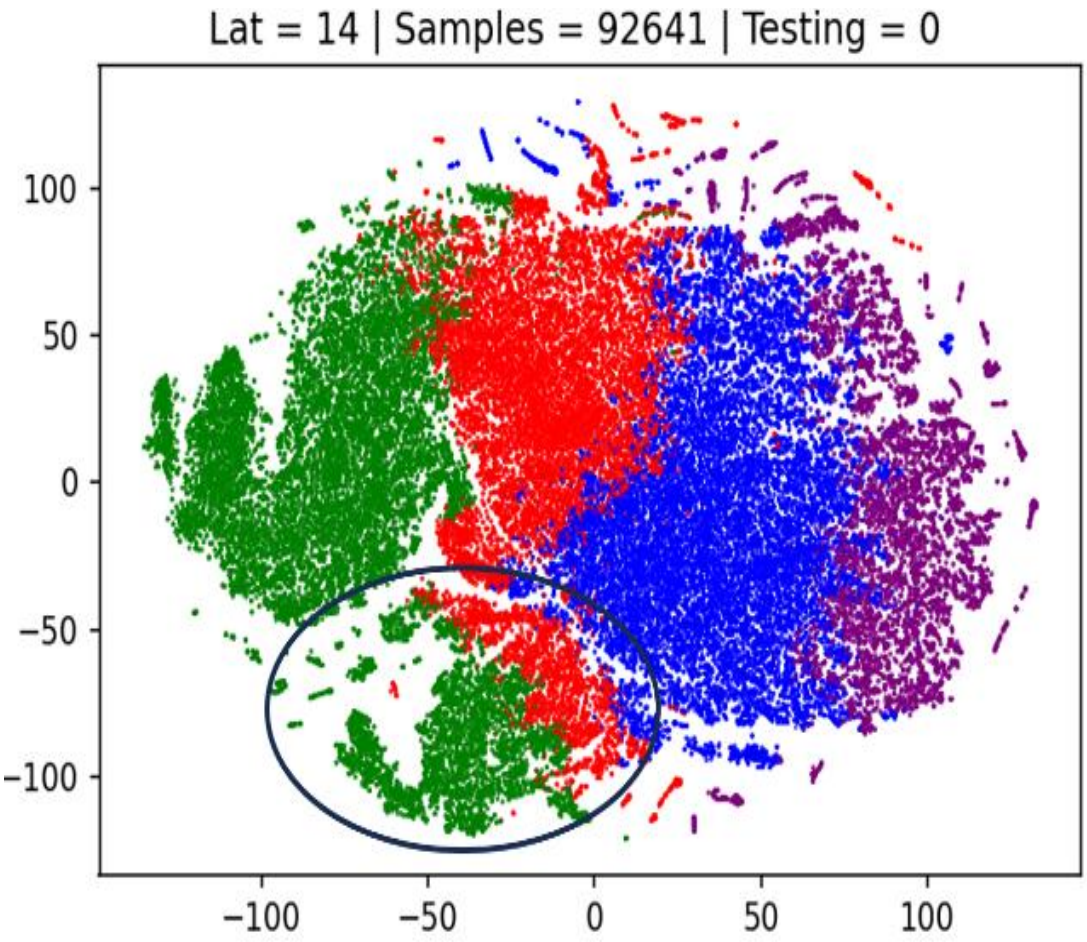


Figure 4.20: t-SNE on all instances of the NVT training set.

Station / cluster	C0	C1	C2	C3
Is NVT observed from cluster instances (Table 4.1)	No	Mostly	No	Intermittent
1: INOH		IP (Uncertain)		EP 15/04
2: GHKH	OP			IP
3: MTYH			IP (picks up erroneous set)	IP (EP 14/04)
4: KTGH				IP (with phase shift) (EP 14-15/04)
5: MURH	OP	IP		IP (EP 11/04)
Cluster indicates	NO NVT	LOW FREQUENCY NVT	UNKOWN / RETRIEVES ERRORNOUS SIGNAL	BROADBAND NVT

Table 4.5. A table demonstrating whether each cluster indicates NVT or not. OP = out of phase (visual), IP = in phase (visual). Highlighted station clearly indicates NVT in supervised RMS plot, outlined red and blue box surpass the PCM two tailed criterion for IP and OP respectively. EP = extreme perturbation.

4.4 Discussion

Reconstructions

There have been some positive reconstructions for spectrograms that most likely contained NVT. Nevertheless, there is an erroneous signal that was reconstructed. All the impulsive signals were reconstructed whilst some low amplitude noise signals were not. The reconstructions demonstrate that the DEC is moderately good at distinguishing noise from NVT.

The lack of day/night signal amongst the entire dataset

For the entire dataset, including the all-stations and singular-station sets, there is a clear lack of day/night signal, despite the training set being high passed with the Butterworth filter. As a result of the set being trained on ~20 days over the course of several years, there is likely to be no or a very inconsistent day night signal in the dataset. Unlike previous datasets, the NVT dataset is trained on tremor signals whose dominant frequency range coincides with that of the day/night anthropogenic signal. Chapter 2 highlighted a night signal is sporadic up to 5Hz whereas the day signal dominates the lower frequencies between 1-2.57Hz. Day/night signals are a summation of human activities so they are likely to be emergent, low amplitude and sporadic in time/frequency, resulting in them not featuring at all in the spectrogram. Additionally, the DEC is not given a sufficient amount of diurnal variation to learn about features within days and night. In preprocessing, the timeseries has been normalised and the spectrogram only standardised. Therefore, the dominant signal, here the NVT, will dominate over lower amplitude day/night signals.

The event on April 16th

For Station 2 (MTYH), C1 was particularly sensitive to increased Mw>3 activity on April 16th (Figure 4.16). Its PCM value is around 1.4 for Mw>5 but has also been signalled as out of phase for Mw>4 suggesting it does not correlate for EQ activity other than April 15th. It is likely that some of the EQ activity around April 15th could accompany NVT and slow slip. After all, C1 correlates well with RMS for some stations and visually contains NVT. However, C1 has an out of phase RMS correlation with a PCM of -1.2.

In the t-SNE diagrams, there were isolated C1 patches which could potentially represent the increased Mw>3 activity. C1 is less sensitive to the increased activity at other stations. This suggests that the DEC has spotted something unique and unknown at MTYH that correlates with the increased Mw>3 activity. Wilber, a website that records EQs worldwide, reports that the increased Mw>3 seismic activity around this period is evenly distributed around Japan suggesting that C1's correlation with increased seismic activity on the April 15th could be random and most likely the result of extreme perturbations in the RMS signal (Figure 4.21).

MTYH was a problematic station: the RMS signal is erroneous for the first half of the series which activated C2 causing that to become dominant. Therefore, there is doubt as to whether C1 is sensitive to EQ activity as only a fraction of non-erroneous data can be compared at this station.

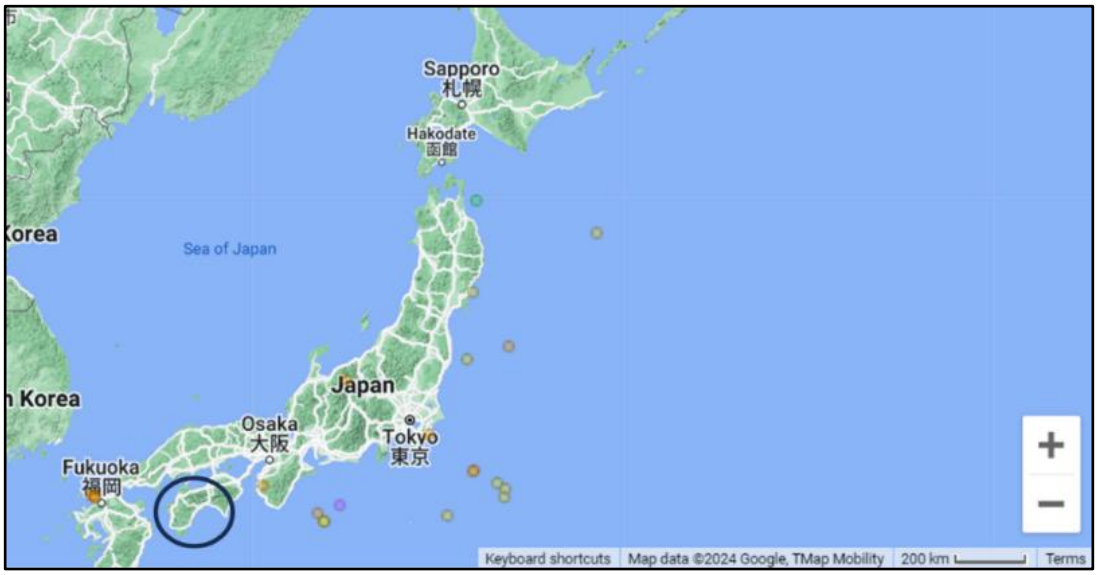


Figure 4.21: events from the 14th – 16th April 2005 with the blue circle corresponding to the area of study (derived from Wilber). This demonstrates that a good proportion of seismic activity is evenly distributed around Japan at this time.

The effect of peak removal

Peak removal is not unusual among papers that investigate NVT. Rouet-Leduc et al., (2019) clipped their waveforms to identify NVT along the Cascadian subduction zone but claimed that it eliminated some of the information in the seismic timeseries. Peak removal here was only conducted to cross correlate each cluster with the RMS signal. As can be seen, peak removal along with Savgol smoothing was successful at removing EQ activity on April 15th and aided PCM analysis.

Perturbations in C3 and C2 on April 15th

C3 demonstrates extreme perturbation between the 14th and 17th April for stations 0 (INOH), 2 (MTYH) and 3 (KTGH). Whatever this is, it is a feature that exists at different points in time for different stations. C3 is plotted in time for each station (Figure 4.22 II). According to Yube et al., (2014), Shikoku and the Kii channel have larger attenuation factors than the Cascadia subduction zone; attenuation for NVT in this region is strong. Figure 4.22 presents a direct comparison of C3 in time.

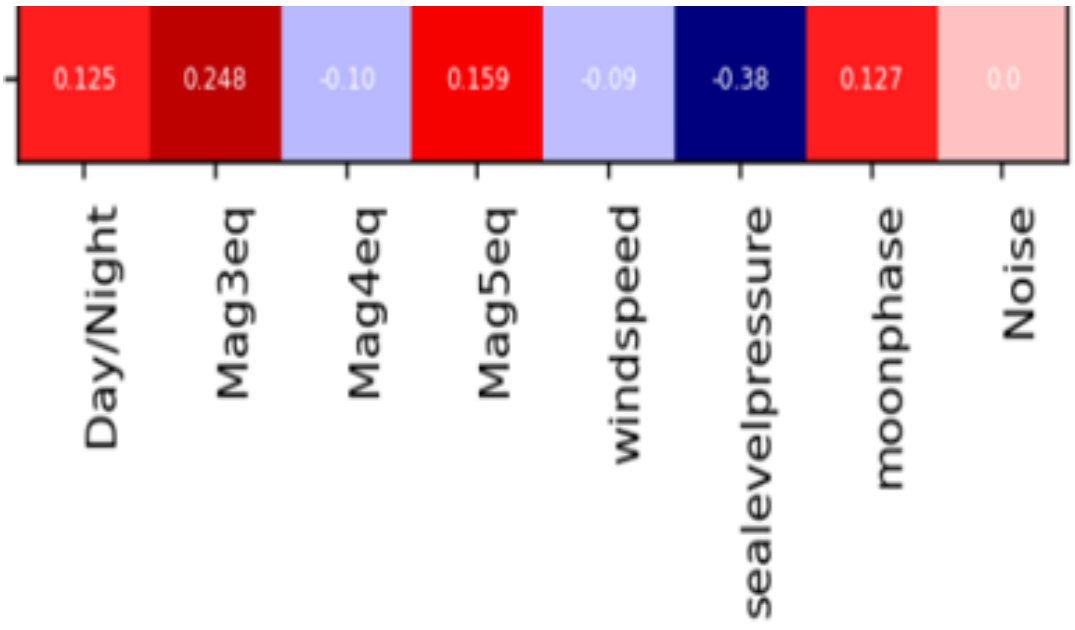


Figure 4.22:a reminder of the PCM values from the whole station set demonstrating that C3 only anti correlates with sea level pressure.

In Figure 4.22, large perturbations arise from GHKH, then KTGH then MTYH at April 15th. They all exist inland (Figure 4.4) in comparison to INOH and MURH. There is a significant anticorrelation with C3 and sea level pressure in the whole station set, which could demonstrate that lowering in sea pressure causes a change in the nature of the seismic signal or it relieves stress along faults causing the nature of NVT to change. C1 demonstrates NVT as a clearly large band of frequencies whilst demonstrating a very poor correlation with sea level pressure. The effect of sea level pressure is not known/certain and should be the driver for future research.

For stations which exhibited a delayed perturbation later on the cluster series, Station 0 (INOH), Station 2 (MTYH) and Station 3 (KTGH). Station 3 (KTGH) precedes Station 2 (MTYH) at the first part on the 15th April but then Station 2 (MTYH) then precedes Station 3 (KTGH) around April 15th. There is delayed onset of Station 0 (INOH) in comparison to Stations 2 and 3. What is most fascinating is that all events align for two oscillations around April 15th (black dashed box in Figure 4.24) additional contribution from GHKH and no contribution from MURH. INOH is approximately 40 kilometres from MTYH whilst MTYH is approximately 80 kilometres from KTGH. Something happens for this cluster at April 15th at these 3 stations that activates C3.

When observing the spectrograms, the reconstructions in Figure 4.23 are dominated by an anomalous signal. It exists across 4 stations which are close to each other: GHKH, KTGH, INOH and MTYH. It is unlikely to be electrical noise as electrical noise would likely exist beyond 50Hz. Some spectrograms have a dominant high amplitude signal and then lower amplitude harmonics at 1.75 Hz. This could symbolise resonant activity wherein a particular frequency causes a significant increase in the amplitude because of NVT or EQ activity, although there are no major EQs that exist around this period – it could perhaps symbolise geological resonances. Therefore, it could be a systematic error in preprocessing. Others spectrograms just have a fundamental harmonic and no other features. In Figure 4.12, GHKH, INOH, KTGH represent distinct flat lines around this period. The Fourier transform it amplified ambient electrical noise as a result of a quiet period and the signal was amplified in the normalisation of the spectrograms.

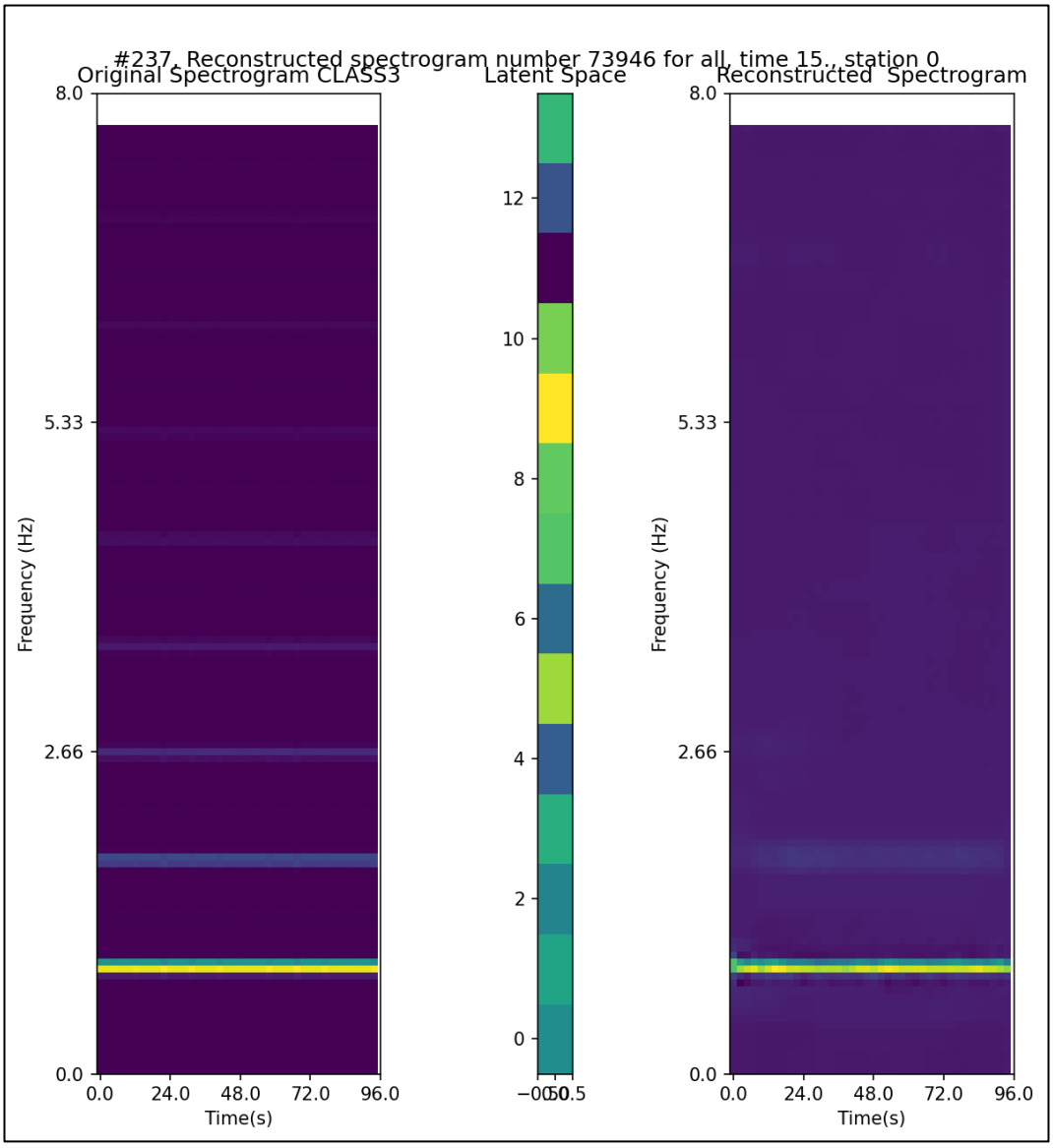


Figure 4.23: 14th April 15th an example reconstruction containing the anomalous signal. There is a dominant low frequency signal at around ~0.7Hz followed by a number of multiples at lower amplitudes and higher frequencies at equal intervals which suggests they are harmonic oscillations. All other example spectrograms are contained in the appendix.

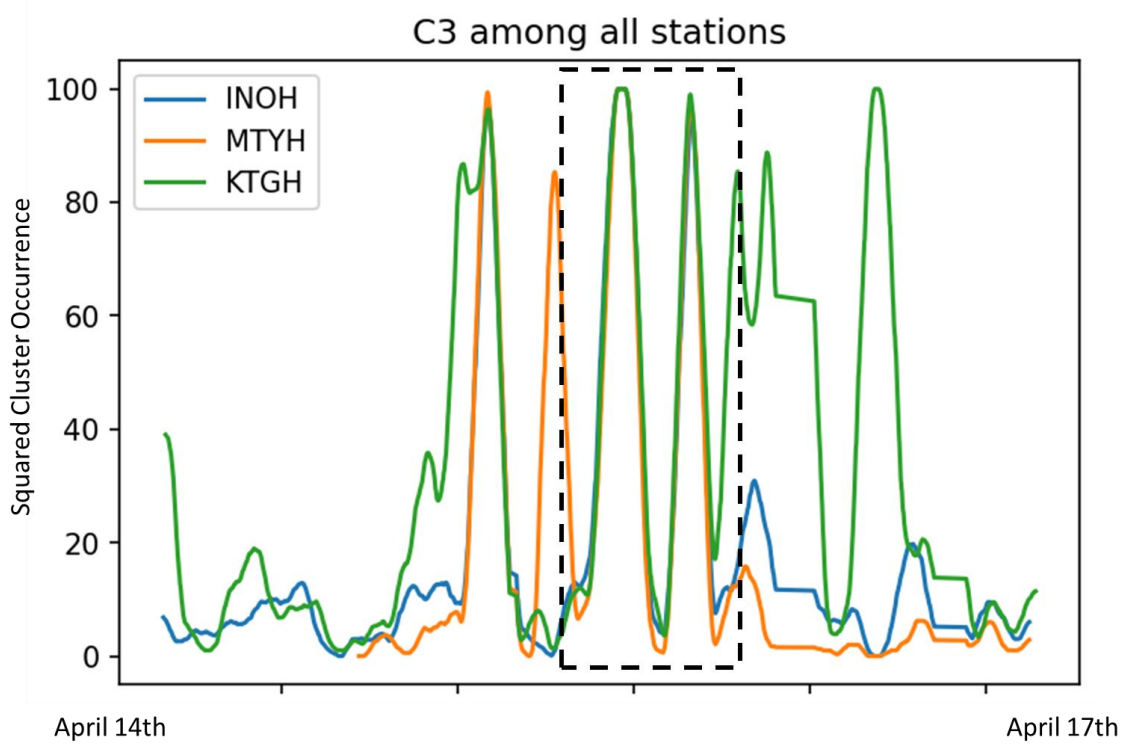


Figure 4.24: for April 14th to 17th for clusters that demonstrate perturbations only at April 15th demonstrating coherency between stations INOH, MTYH and KTGH.

Is there a distinction in the depth of NVT?

As previously stated, DLFS are deep and have a prominent frequency of around 2Hz (Obara, 2006). C3 contains a dominant 2Hz signal in places. Shikoku is known for having DLFs (Obara, 2006). There is a larger frequency range in C1 in comparison to C3, which has a small frequency range at lower frequencies which could demonstrate the DEC may be able to distinguish depth of NVT. It must be noted that in the t-SNE representation, C1 and C3 are in close vicinity. Several latent dimensions could correspond to the depth and frequency of NVT and should be the driver for future research.

4.5 Overall Discussion

The NVT dataset outperforms the dataset from Chapter 2 in clustering spectrograms for physical attributes. There was great variation between the stations which inhibited analysis of the RMS signal at each station. Instead of correlations with day/night oscillations, there is a distinct correlation with the RMS signal in C3 which could correspond to NVT activity but unfortunately an anomalous feature in the dataset causing C3 to become out of phase with the RMS for many stations. C1 presented variable low-frequency NVT visually with strong reconstructions but did not correlate well with the RMS (Figure 4.13 II) which could suggest that it identifies low energy NVT. C0 provides a good anticorrelation to NVT, and C2 often correlated well with RMS signal. Integrating clusters in the latent space such as C1 and C2 creating a new super cluster with a distinct cluster centre or simply reducing the number of clusters should be investigated in the future. Without anomalous data and with a sufficient and wide-ranging dataset, it is hypothesised that the DEC does recognise NVT and potentially distinguishes different features in NVT based upon bandwidth.

Chapter 5 – Conclusion

A DEC has been applied to data from the Japanese HiNet network, setting the groundwork for future research into its use in identifying and finding events.

Iwate EQ

When using high passed data to eliminate background signals, the DEC is identifies global features for day / night perturbations and 6.7Mw event whilst uncovering low energy impulsive signals that could represent precursory activity. Increasing the size of the latent space does not improve the network but makes it more sensitive to day/night perturbations.

Non-volcanic tremor

Reconstructions of the dataset were positive and the DEC successfully denoised a large proportion of the spectrograms that contained non-seismic noise. Two clusters indicated NVT although this is variable for each station. There was a discrepancy between those clusters which correlated with the RMS and those which visually contained NVT, suggesting that the DEC is able to separate clusters based upon tremor energy levels. Spectrograms with artificial looking signals of constant frequency dominated certain time periods and caused disturbances in clustering. This is hypothesised to be electrical noise.

Overall, this thesis indicates that the DEC is a good reconstructor in the time-frequency domain. Its clustering ability was strong for anthropogenic signals and NVT over a global set of stations but poor for singular stations and impulsive events such as low magnitude EQs whilst uncovering some unusual signals prior to earthquakes. This thesis lays the groundwork for future research into deep clustering for unusual events and NVT.

Recommendations for additional testing of the DEC model

The following suggestions has been made by the author when working adapting and testing the DEC to classify seismic events:

1. A good test for the DEC denoising capability could be to store the real and imaginary components of the STFT and inverse the reconstructed spectrograms into the seismic timeseries. This proved successful in Saad et al., (2022) and will test this thesis' hypothesis that the DEC is a good denoiser from directly comparing seismic timeseries.
2. Seismic data should be combined with additional remotely sensed data for slow-slip and the DEC is trained in a supervised fashion to detect ETS.
3. The convolutional kernel should be adjusted to prioritise a great range in frequency over time (a matrix that has more rows than columns) which could potentially improve impulsive event and EQ detection.
4. The DEC is given a small, labelled subset of varying forms (deep shallow) NVT and trained in a semi-supervised way providing a known dataset of NVT to the DEC and then training it on data that precedes an EQ as per Chapter 2 and in places where the existence of NVT is unknown.
5. Test a smaller latent space or a bigger sample size similar to that of Snover et al., (2021).

6. Experiment with the loss adjustment parameter to prioritise KLDiv over MSE in the network to optimise clustering.

7. Test the capability of the DEC to identify long-term seasonal changes seen in parameters such as moon phase and sea level temperature. A longer period could be chosen to train the network, along with additional subsampling of the dataset to save memory.

8. Test to see whether NVT is detectable in the Kii Channel (Obara, 2002). The qualitative nature of the study alongside attenuation of the NVT could result in NVT being missed in addition to the lack of coverage around the Kii channel. Due to the emergent nature of VLFs and VLFEs, and lack of P waves, NVT could simply have been missed. Of course, all the seismic stations studied exist along Shikoku. Therefore, future studies could incorporate seismic stations from inland Japan and around the channel utilising a trained model on seen NVT.

References

Bergen, K. J., Johnson, P. A., De Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. In Science (Vol. 363, Issue 6433). American Association for the Advancement of Science. https://doi.org/11.1126/science.aau0323

Brudzinski, M. R., Hinojosa-Prieto, H. R., Schlanser, K. M., Cabral-Cano, E., Arciniega-Ceballos, A., Diaz-Molina, O., & DeMets, C. (2010). Nonvolcanic NVT along the Oaxaca segment of the Middle America subduction zone. Journal of Geophysical Research: Solid Earth, 115(8). https://doi.org/11.1029/2008JB006061

Dzwinel, W., Yuen, D. A., Boryczko, K., Ben-Zion, Y., Yoshioka, S., & Ito, T. (2005). Nonlinear multidimensional scaling and visualization of earthquake clusters over space, time and feature space (Vol. 12).

Guo, X., Liu, X., Zhu, E., & Yin, J. (2017). Deep Clustering with Convolutional Autoencoders. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10635 LNCS, 373–382. https://doi.org/11.1007/978-3-319-70096-0_39

Hulbert, C., Jolivet, R., Gardonio, B., Johnson, P. A., Ren, C. X., & Rouet-Leduc, B. (2022). NVT Waveform Extraction and Automatic Location With Neural Network Interpretation. IEEE Transactions on Geoscience and Remote Sensing, 61. https://doi.org/11.1109/TGRS.2022.3156125

Jenkins, W. F., Gerstoft, P., Bianco, M. J., & Bromirski, P. D. (2021). Unsupervised Deep Clustering of Seismic Data: Monitoring the Ross Ice Shelf, Antarctica. Journal of Geophysical Research: Solid Earth, 126(9). https://doi.org/11.1029/2021JB021716

Johnson, C. W., Ben-Zion, Y., Meng, H., & Vernon, F. (2020). Identifying Different Classes of Seismic Noise Signals Using Unsupervised Learning. Geophysical Research Letters, 47(15). https://doi.org/11.1029/2020GL088353

Jozinović, D., Lomax, A., Štajduhar, I., & Michelini, A. (2022). Transfer learning: improving neural network based prediction of earthquake ground shaking for an area with insufficient training data. Geophysical Journal International, 229(1), 704–718. https://doi.org/11.1093/gji/ggab488

Käufl, P., Valentine, A. P., & Trampert, J. (2016). Probabilistic point source inversion of strong-motion data in 3-D media using pattern recognition: A case study for the 2008 Mw 5.4 Chino Hills earthquake. Geophysical Research Letters, 43(16), 8492–8498. https://doi.org/11.1002/2016GL069887

Lye, V., Ong, S., Nielsen, S., Giani, S., Johnson, P. A., Ong, V., & Johnson, P. (2022). Temporal earthquake forecasting. doi: 11.1002/essoar.10511752.1

Magana-Zook, S. A., & Ruppert, S. D. (2017). Explosion Monitoring with Machine Learning: A LSTM Approach to Seismic Event Discrimination.

McCausland, W., Malone S., Johnson D. (2005). Temporal and spatial occurrence of deep non-volcanic NVT: From Washington to northern California. Geophysical Research Letters. doi:10.1029/2005GL024349

Mousavi, S. M., Zhu, W., Ellsworth, W., & Beroza, G. (2019). Unsupervised Clustering of Seismic Signals Using Deep Convolutional Autoencoders. IEEE Geoscience and Remote Sensing Letters, 16(11), 1693–1697. https://doi.org/11.1109/LGRS.2019.2909218

Mitchell, T (1997). Machine Learning, McGraw-Hill Education.

Ning, C., & Xie, Y. (2024). Convolutional variational autoencoder for ground motion classification and generation toward efficient seismic fragility assessment. Computer-Aided Civil and Infrastructure Engineering, 39(2), 165–185. https://doi.org/11.1111/mice.13061

Obara, K. (2002). Nonvolcanic Deep NVT Associated with Subduction in Southwest Japan. In SCIENCE (Vol. 296). www.sciencemag.org

Obara, K. (2010). Phenomenology of deep slow earthquake family in southwest Japan: Spatiotemporal characteristics and segmentation. Journal of Geophysical Research: Solid Earth, 115(8). https://doi.org/11.1029/2008JB006048

Obara, K., & Hirose, H. (2006). Non-volcanic deep low-frequency NVTs accompanying slow slips in the southwest Japan subduction zone. Tectonophysics, 417(1–2), 33–51. https://doi.org/11.1016/j.tecto.2005.04.013

Ong, V., & Lye, S. (2020). Durham E-Theses Articial intelligence to detect and forecast earthquakes. http://etheses.dur.ac.uk

Ozanich, E., Thode, A., Gerstoft, P., Freeman, L. A., & Freeman, S. (2021). Deep embedded clustering of coral reef bioacoustics. The Journal of the Acoustical Society of America, 149(4), 2587–2601. https://doi.org/11.1121/11.0004221

Rouet-Leduc, B., Hulbert, C., & Johnson, P. A. (2019). Continuous chatter of the Cascadia subduction zone revealed by machine learning. Nature Geoscience, 12(1), 75–79. https://doi.org/11.1038/s41561-018-0274-6

Saad, O. M., Chen, Y., Savvaidis, A., Chen, W., Zhang, F., & Chen, Y. (2022). Unsupervised Deep Learning for Single-Channel Earthquake Data Denoising and Its Applications in Event Detection and Fully Automatic Location. IEEE Transactions on Geoscience and Remote Sensing, 61. https://doi.org/11.1109/TGRS.2022.3209932

Seydoux, L., Balestriero, R., Poli, P., Hoop, M. de, Campillo, M., & Baraniuk, R. (2020). Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning. Nature Communications, 11(1). https://doi.org/11.1038/s41467-020-17841-x

Snover, D., Johnson, C. W., Bianco, M. J., & Gerstoft, P. (2021). Deep clustering to identify sources of urban seismic noise in long beach, California. Seismological Research Letters, 92(2), 1011–1022. https://doi.org/11.1785/0220200164

Steinmann, R., Seydoux, L., Beaucé, É., & Campillo, M. (2022). Hierarchical Exploration of Continuous Seismograms With Unsupervised Learning. Journal of Geophysical Research: Solid Earth, 127(1). https://doi.org/11.1029/2021JB022455

Uyeda, S. (2013). On earthquake prediction in Japan. Proceedings of the Japan Academy Series B: Physical and Biological Sciences, 89(9), 391–401. https://doi.org/11.2183/pjab.89.391

Valentine, A. P., Kalnins, L. M., & Trampert, J. (2013). Discovery and analysis of topographic features using learning algorithms: A seamount case study. Geophysical Research Letters, 40(12), 3048–3054. https://doi.org/11.1002/grl.50615

Valentine, A. P., & Trampert, J. (2012). Data space reduction, quality assessment and searching of seismograms: Autoencoder networks for waveform data. Geophysical Journal International, 189(2), 1183–1202. https://doi.org/11.1111/j.1365-246X.2012.05429.x

Vannucchi, P., Clarke, A., de Montserrat, A., Ougier-Simonin, A., Aldega, L., & Morgan, J. P. (2022). A strength inversion origin for non-volcanic NVT. Nature Communications, 13(1). https://doi.org/11.1038/s41467-022-29944-8

Wang, K., Johnson, C. W., Bennett, K. C., & Johnson, P. A. (2021). Predicting fault slip via transfer learning. Nature Communications, 12(1). https://doi.org/11.1038/s41467-021-27553-5

Xie, J., Girshick, R., & Farhadi, A. (2016). Unsupervised Deep Embedding for Clustering Analysis. https://github.com/piiswrong/dec.

Yabe, S., Baltay, A. S., Ide, S., & Beroza, G. C. (2014). Seismic-wave attenuation determined from tectonic NVT in multiple subduction zones. Bulletin of the Seismological Society of America, 104(4), 2043–2059. https://doi.org/11.1785/0120140032

Chapter 6 – Appendices

Appendix I: List of abbreviations

Machine Learning and Artificial Intelligence		
Acronym	**Definition**	**Contained in (if applicable)**
ADAM	Adaptive moment estimation algorithm	
AE	Autoencoder	
ANN	Artificial Neural network	
CAE	Convolutional autoencoder	$\in$ AE
CL[(n)]	(N)th convolutional layer	
C[*k*]	Cluster k .	
CC[k']	Cluster centre k'.	$\in \mathbb{L}$
CNN	Convolutional neural network	
COD	Curse of dimensionality	
CrL	Cropping Layer	
CAE	Convolutional autoencoder	
CTL[(n)]	(N)th Convolutional transpose layer	
DCAE	Deep convolutional autoencoder	
DEC	Deep embedded clustering	
DL	Deep Learning	
DNN	Deep Neural network	$\in$ ANN
FL	Feature learning	
FCNN	Fully connected neural network	
GMM	Guassian mixture modelling	
KF/RF	Kernel filter/receptive field	
KLDiv	Kullback-Leibler Divergence	
KMC	K-means clustering	
KMCC	K-means Cluster Centre	$\in \mathbb{D}'$
ML	Machine Learning	
MAE	Mean Absolute Error	
MSE	Mean Squared Error	
RBM	Restricted Boltzmann machines	
ReLU	Rectified Linear Unit Function	
SAE	Sparse autoencoder	
SMOTE	Synthetic Minority Oversampling Technique	
TrD	Training dataset (spectrograms)	$\in \mathbb{D}$
TeD	Testing dataset (spectrograms)	$\in \mathbb{D}$
VaD	Validation dataset (spectrograms)	$\in \mathbb{D}$
ReSp	Reconstructed dataset (spectrograms)	$\in \mathbb{D}'$

EnD	Encoded dataset (n dimensional vectors)	$\in \mathbb{L}$
VAE	Variational Autoencoder	
WCSS	Within-cluster sum of squares	
t-SNE	t-distributed stochastic neighbour embedding	$\in \mathbb{L}$

Signal Processing	
ECM	Envelope Correlation method
RMS	Root mean squared
SNR	Signal to Noise Ratio
STFT	Short Time Fourier transform

Earthquake Source Seismology	
DLF	Deep low frequency earthquakes (2Hz, 30km depth)
NVT	Non-volcanic tremor
EQ	Earthquake
ESS	Earthquake Source Seismology
SEQ	Slow earthquakes
SS	Slow slip
ETS $\in$ SEQ	Episodic NVT and slip
LFEQ	Low frequency earthquake (2Hz dom freq)
LFT	Low frequency NVT (1-10Hz dom freq)
VLFEQ	Very low frequency earthquake (<1Hz Dom freq)
VLFT	Very low frequency NVT (<1Hz dom freq)
GPS	Global positioning system
NIED	National Research Institute for Earth Science and Disaster Prevention (Japan)
GR	Gutenberg-Ritcher Law
Mw	Moment magnitude of an earthquake (W phase)
PCM	Pearson correlation moment
PCM/SD	Pearson correlation moment to standard deviation ratio
GSN	Global seismic network

Mathematical symbols	
Symbol	**Definition**

$\mathbf{x}$	A neuron: a vector.
$\bar{\mathbf{x}}^{(n)}$	Arbitrary layer *n* comprising of neurons
$\mathbf{x_t}$; $\mathbf{x_f}$	Time bin; frequency bin
$\|\mathbf{x_t}\|$; $\|\mathbf{x_f}\|$	Number of time bins; number of frequency bins
x	A given point of $\mathbf{x}$
w	Neuron weight
t	Time
b	Associated bias in a neuron
f_s	Sampling frequency (seismometer)
n	Total number of samples
$\tilde{n}$	Total number of samples in cluster occurrence/sample size in Pearson correlation
P	Duration / window size
q_{ij}	Soft assignments
p_{ij}	Target distribution
c_j	Cluster centroid
θ	Trainable parameters combining w and b .
f	Activation function of a given neuron.
K	Number of clusters assigned.
$\mathbb{D}$	Dataspace
$\mathbb{D}'$	Reconstructed dataspace
$\mathbb{L}$ / $\mathrm{L_m}$	Latent space / Latent space dimension
$\mathbb{S}$	Potential hypothetical seismic space here?
$\mathcal{V}$	Amplitude signal in the time domain
g	The result of convolution between activation function.
ϕ	Convolutional kernel
ρ	Dimension reducing pooling function
E	Encoder Function in Autoencoder
D	Decoder Function in Autoencoder
(n)	Arbitrary layer (n)
k	Dimension size of autoencoder and neuron
m	Size of latent space dimension
k'	Size of reconstructed latent space
λ	Loss adjustment parameter
ℓ	Learning rate
$\mathcal{P}$	Padding
P	Window size
$\mathcal{F}$	High pass filter

Appendix II: Exploration of higher latent spaces

High pass analysis ($\mathbb{L}_{15}$ Space)

The latent space will be varied to investigate the effect of being able to derive more features in the seismic timeseries. Due to this window size (P) is set to 120 second to double the number of instances. This should make little difference to the high-passed spectrograms whilst doubling the number of instances in the latent space. It will also increase the frequency resolution, which may help in analysing more complex features in the spectrograms.

Reconstructions

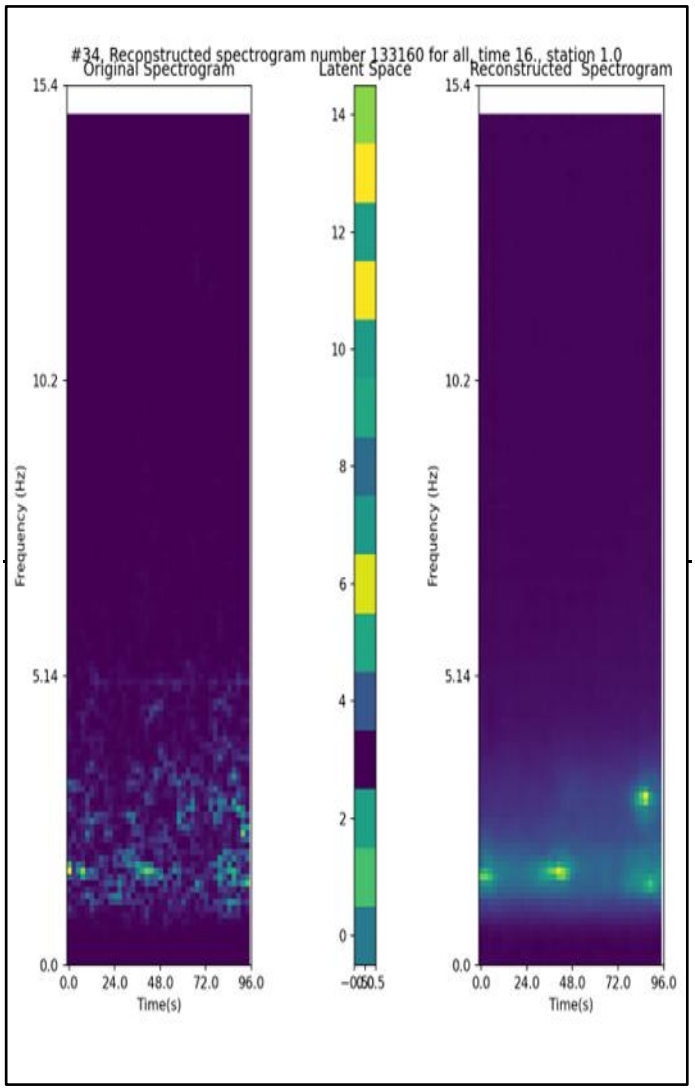


Figure II.1 Reconstructions for 15-dimensional latent space. More detailed features in the spectrogram are reconstructed. The rest of the reconstructed spectrograms for this feature can be found in the appendix.

These reconstructed spectrograms continue to demonstrate a large amount of noise removal and the reconstructed features are finer and more detailed. The surrounding noise, as determined by the DEC, is more dimmed in the reconstruction, which is a sign that the DEC is learning about a greater number of features.

Clusters in space and time for Lat = 15

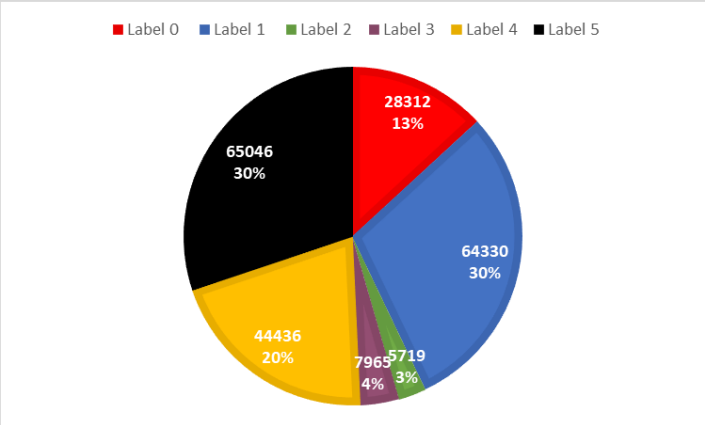


Figure II.2 the number of clusters assigned in L15 space demonstrating consistency form the L14 space.

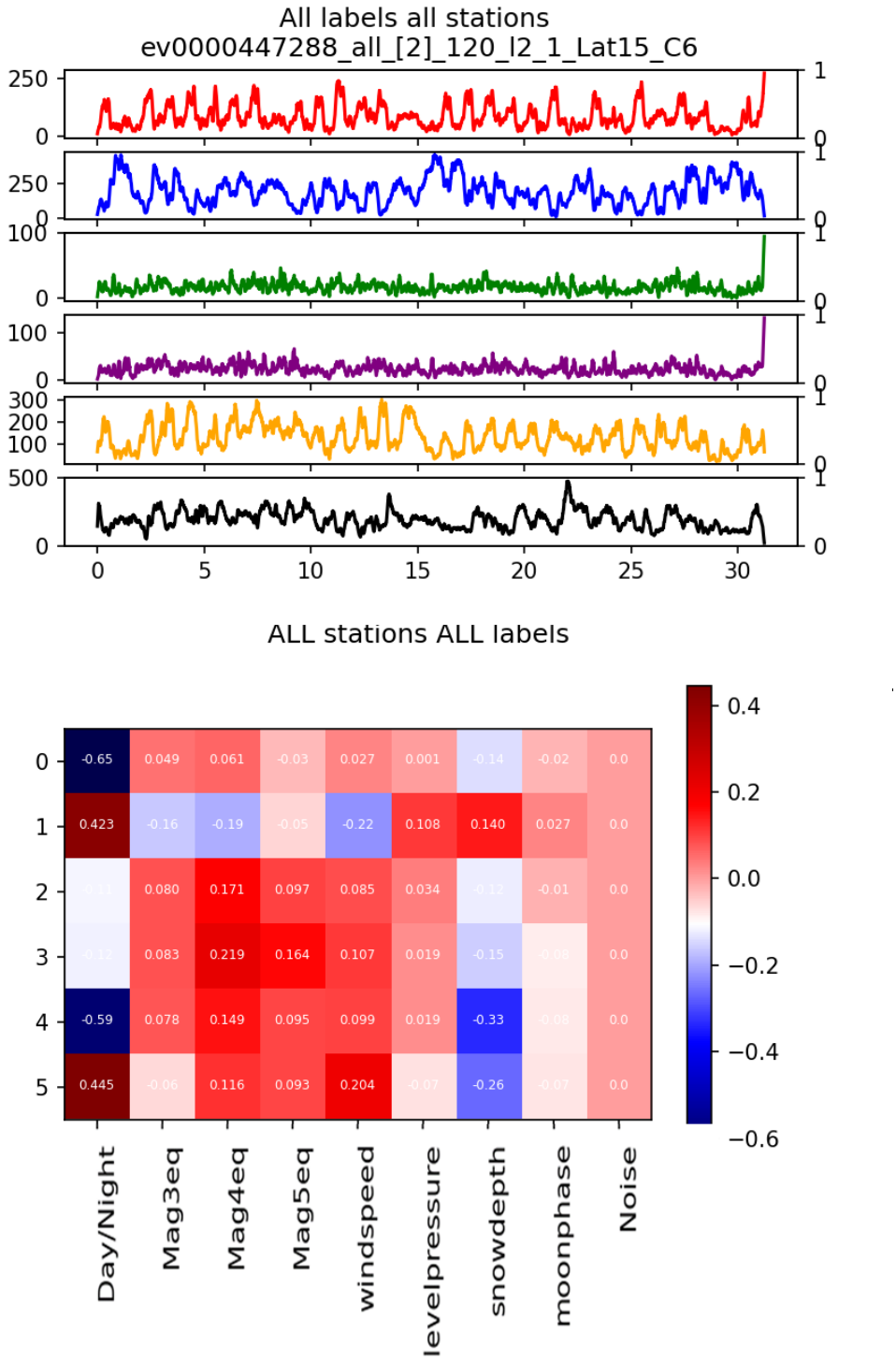


Figure II.3: The clusters (above) along with PCM values for the L15 space showing day night perturbations and a strong impulse at the end of the series for C0,1,2. The correlations between day and night are equally strong among 4 clusters 0,1,4,5 whilst some demonstrate next to no correlation. There is no recognisable correlation among other potential features.

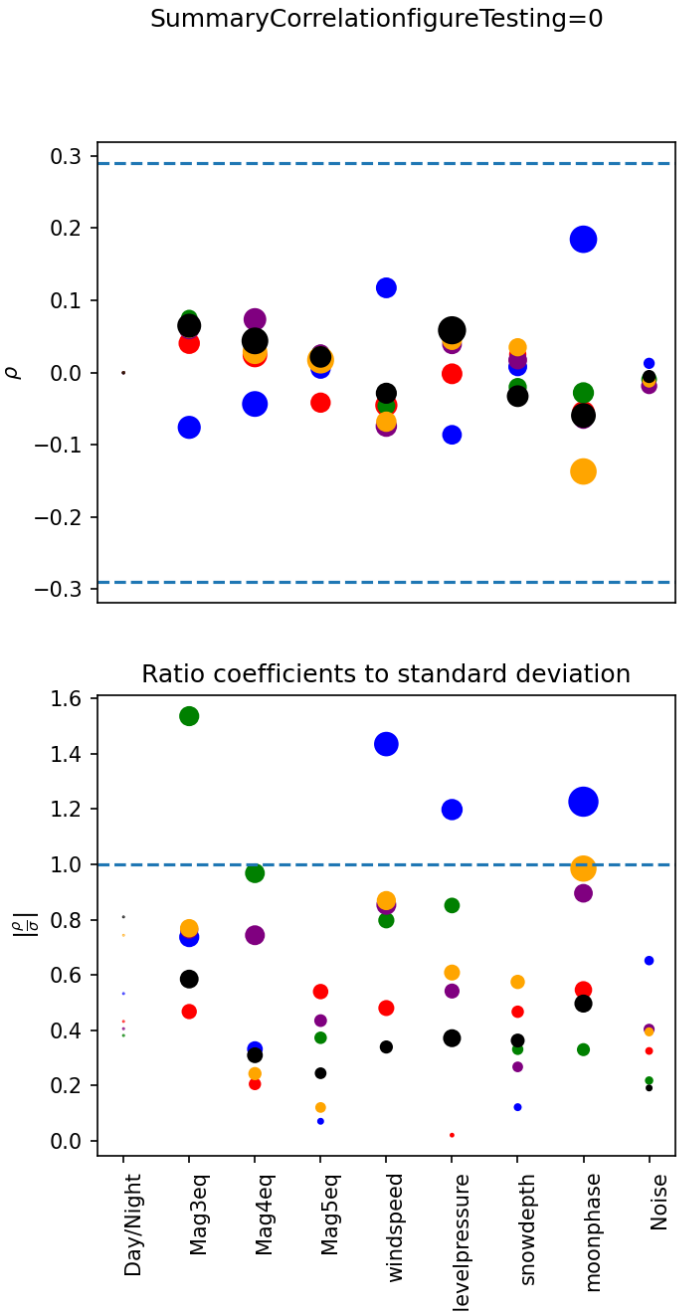


Figure II.4 I-V Clusters for 15-dimensional latent space demonstrating slightly more variability than the L14 space.

For the whole station set, 4 clusters (C[0,1,4,5]) are correlated with space and time with the other two corresponding to the spike. The spike is sharper here than seen in 14d latent space. Furthermore, the anthropogenic signal is less frequent among clusters with lower PCM values than the L14 space.

High pass analysis ($\mathbb{L}_{16}$ Space)

Reconstruction

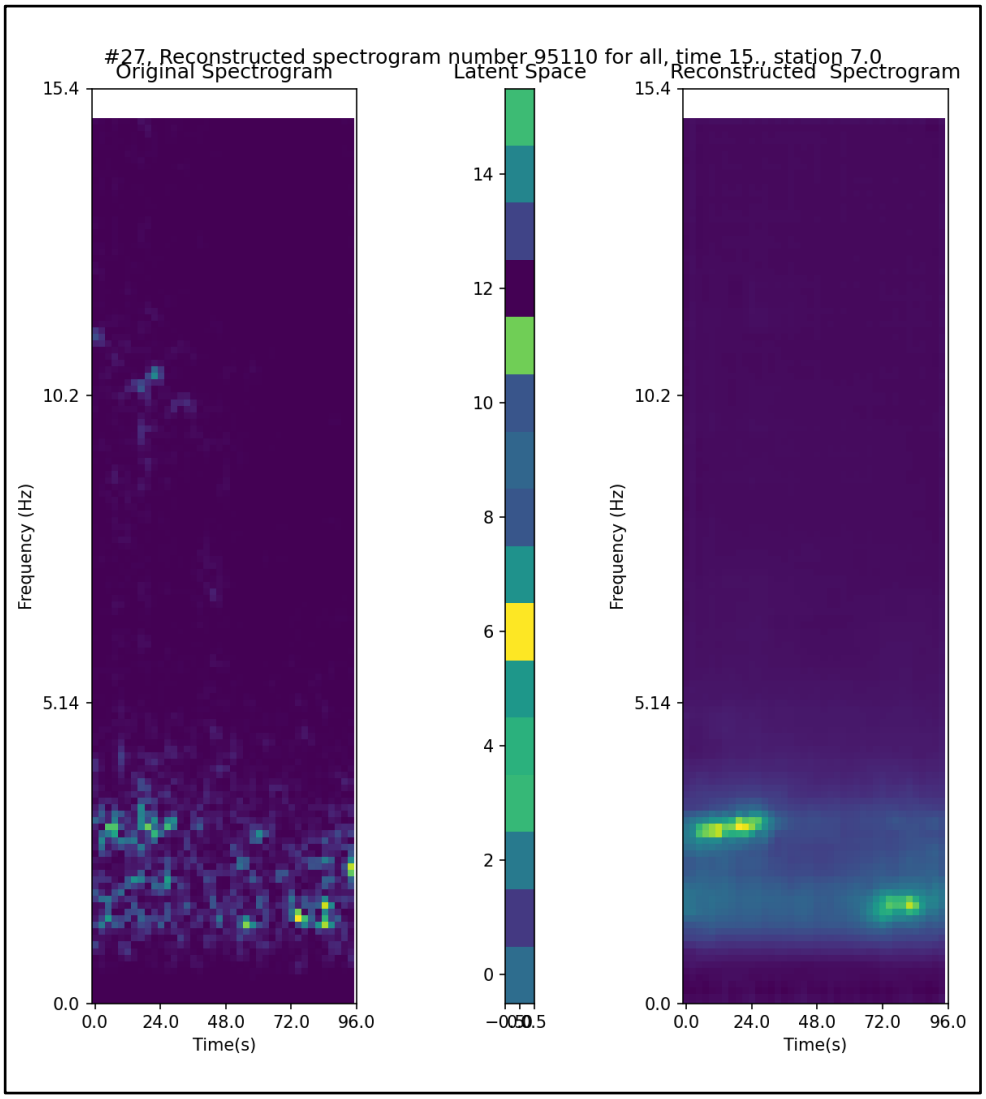


Figure II.5: Reconstruction for 16-dimensional latent space with additional reconstructions in the appendix.

Reconstructions are similar to 15-dimensional latent space although the retained features tend to have a blurred feature around them. This suggests a level of uncertainty in the reconstruction of the spectrograms and potentially demonstrates noise that is seeping through.

Cluster in space and time

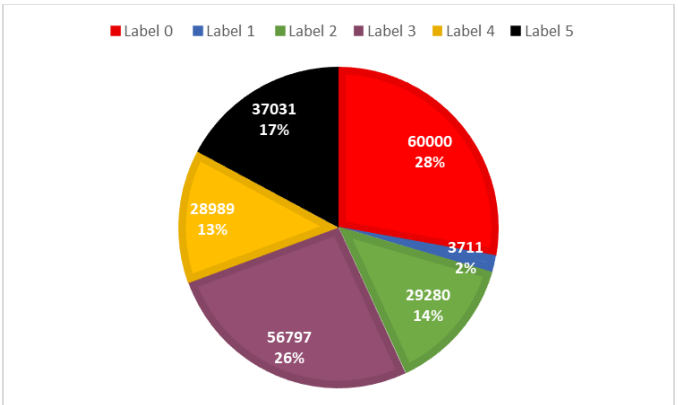
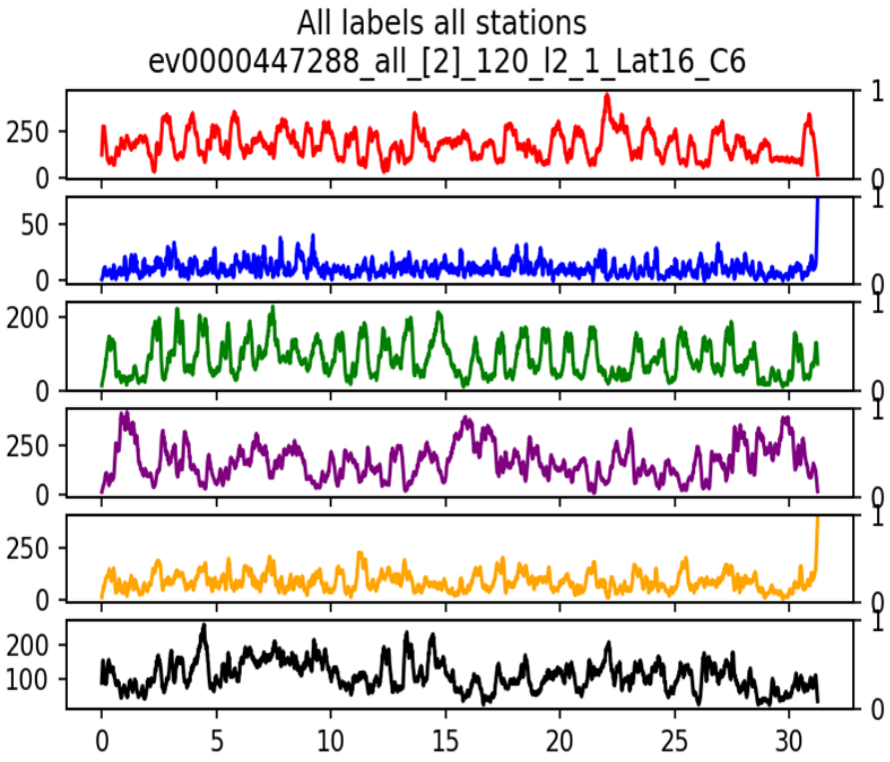


Figure II.6: There is less variance among the cluster numbers in the L16 space.



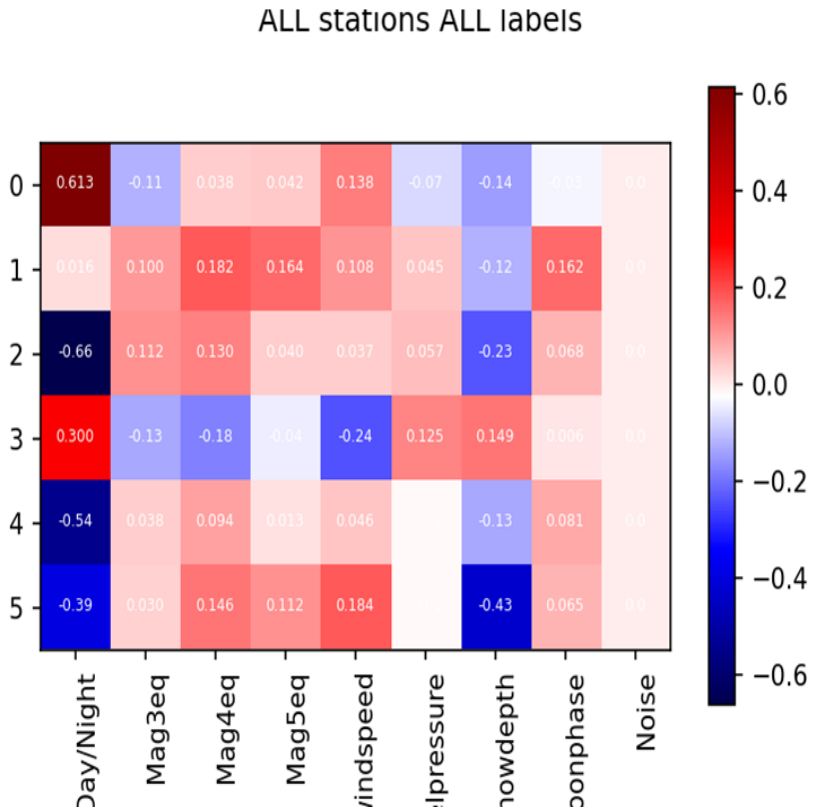


Figure II.7: cluster occurrences in space and time and PCM values demonstrating that there is still a prominent day/night signal and a spike at the end of the series. Both PCM values, PCM vs standard deviation among all stations are contained in the appendix as they are similar to the L15 space.

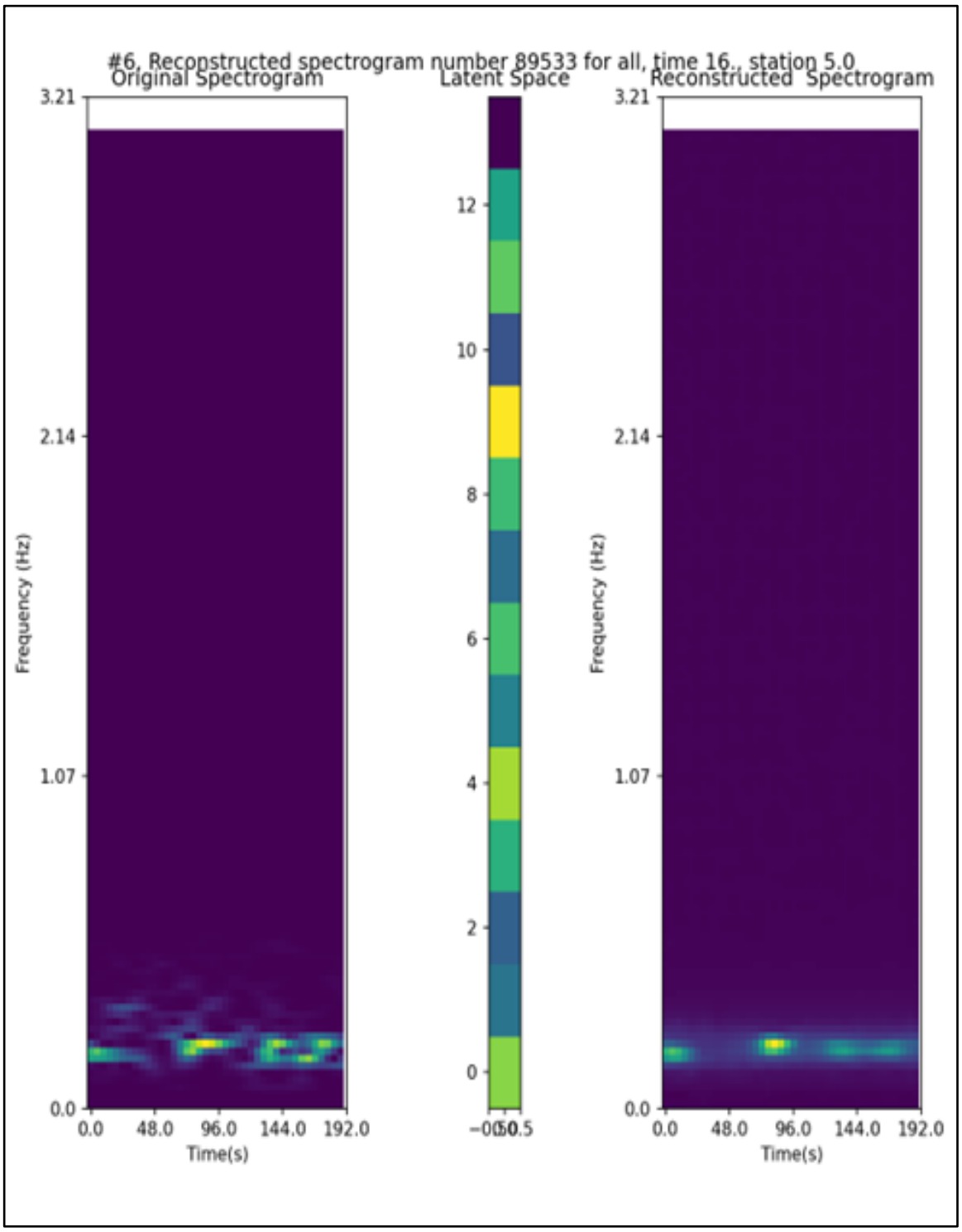
There are now 5 clusters that correlate with day/night (Figure 2uI) which indicates an increase in the latent space dimensions causes the clustering to become more sensitive to day/night and hence anthropogenic perturbations.

The PCM values are consistent along with L15 space with there still being an outlier for Mag > 3 ratio of PCM to standard deviation across stations and is contained in the appendix. This shows that the DEC capacity to identify Mag > 3 earthquakes is invariant to the latent dimension.

On visual inspection, the clusters appear to show short period noise perturbations which could be a result of increased latent dimension.

Appendix III: Supplementary figures from Chapter 2

Example of an unfiltered reconstruction with most of the energy in the low end of the spectrogram.



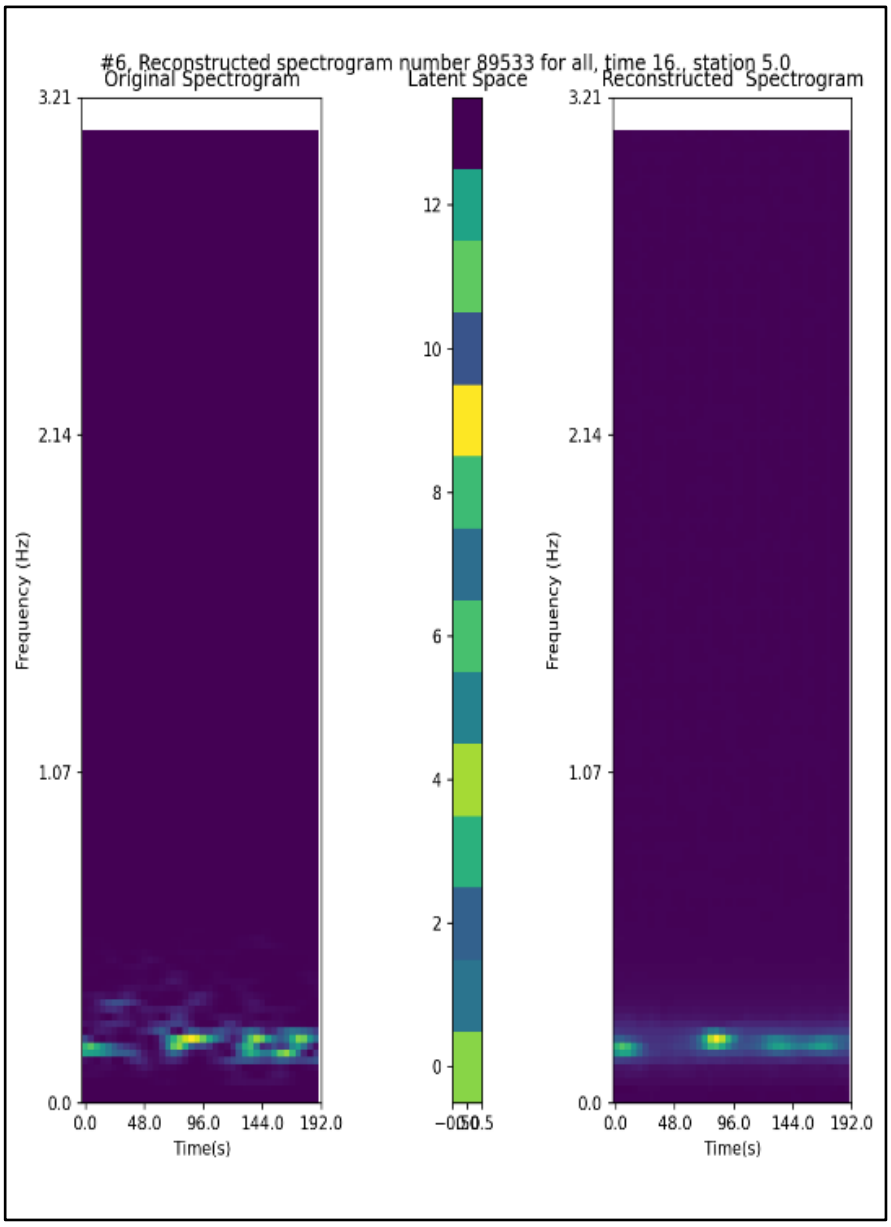


Figure III.1 Examples of reconstructions from the unfiltered data.

L14 space. Supplementary spectrogram.

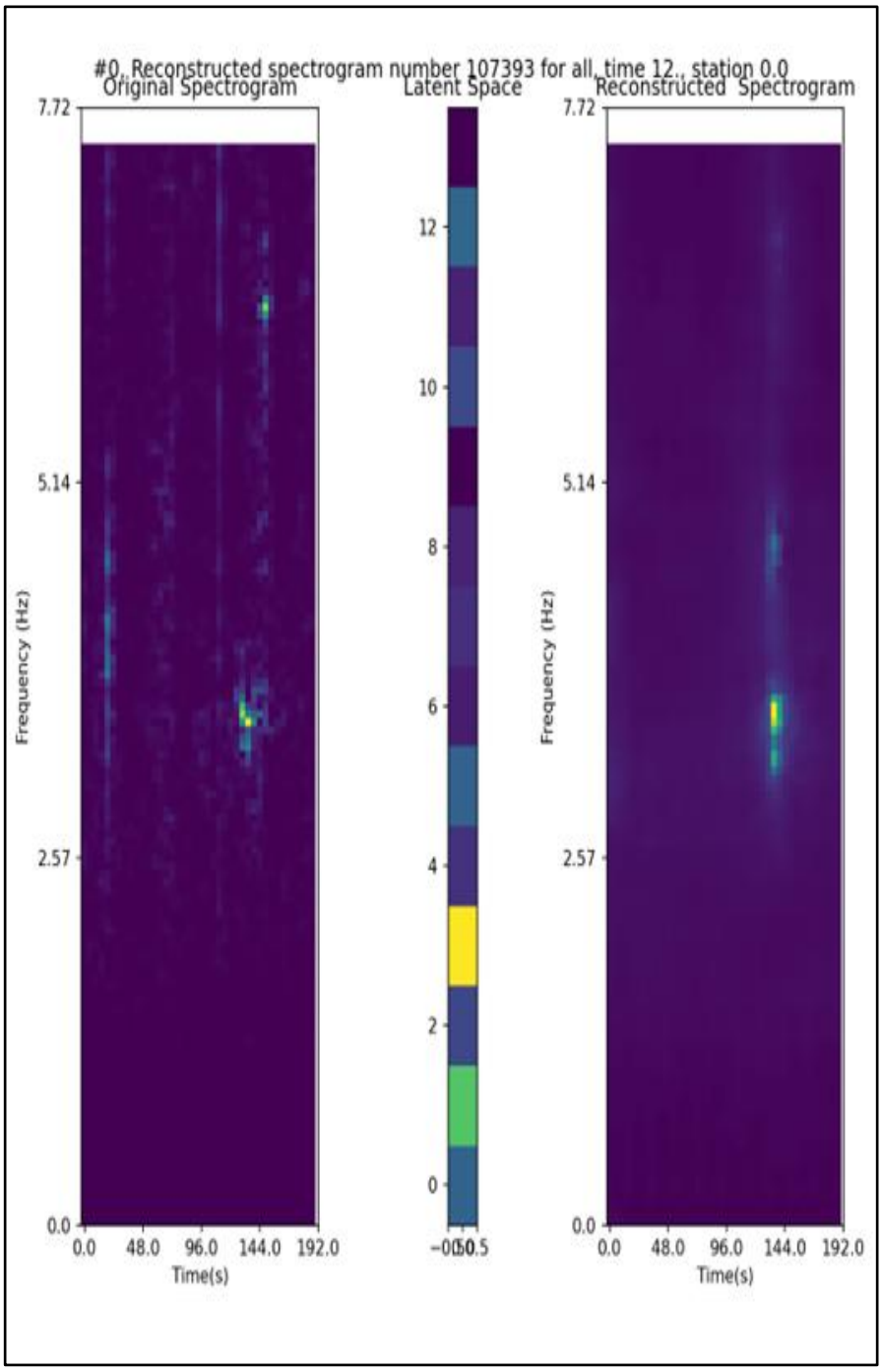


Figure III.2 Examples of reconstructions from the L14 space.

L15 space Supplementary Spectrograms

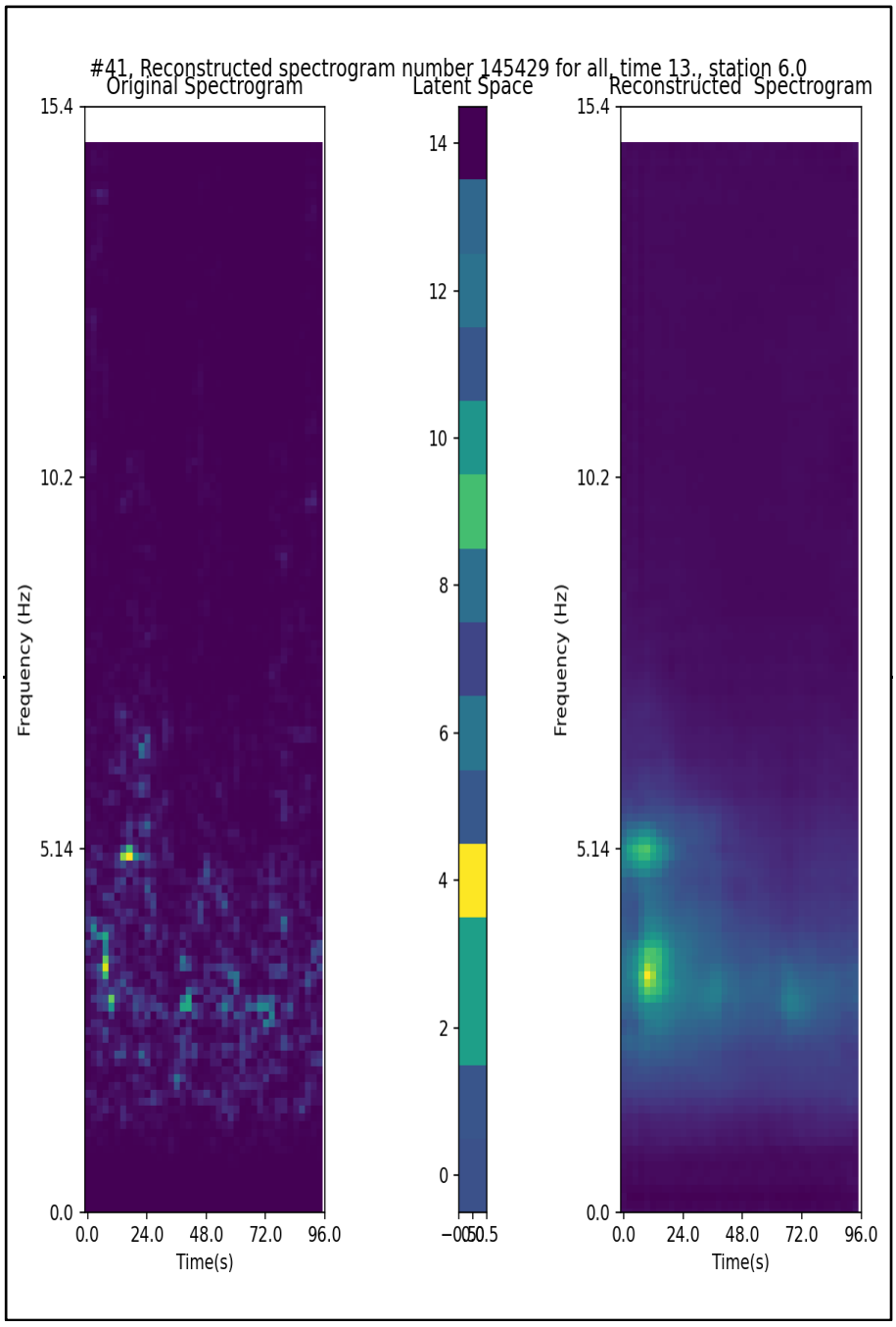


Figure III.2 Examples of reconstructions from the L15 space.

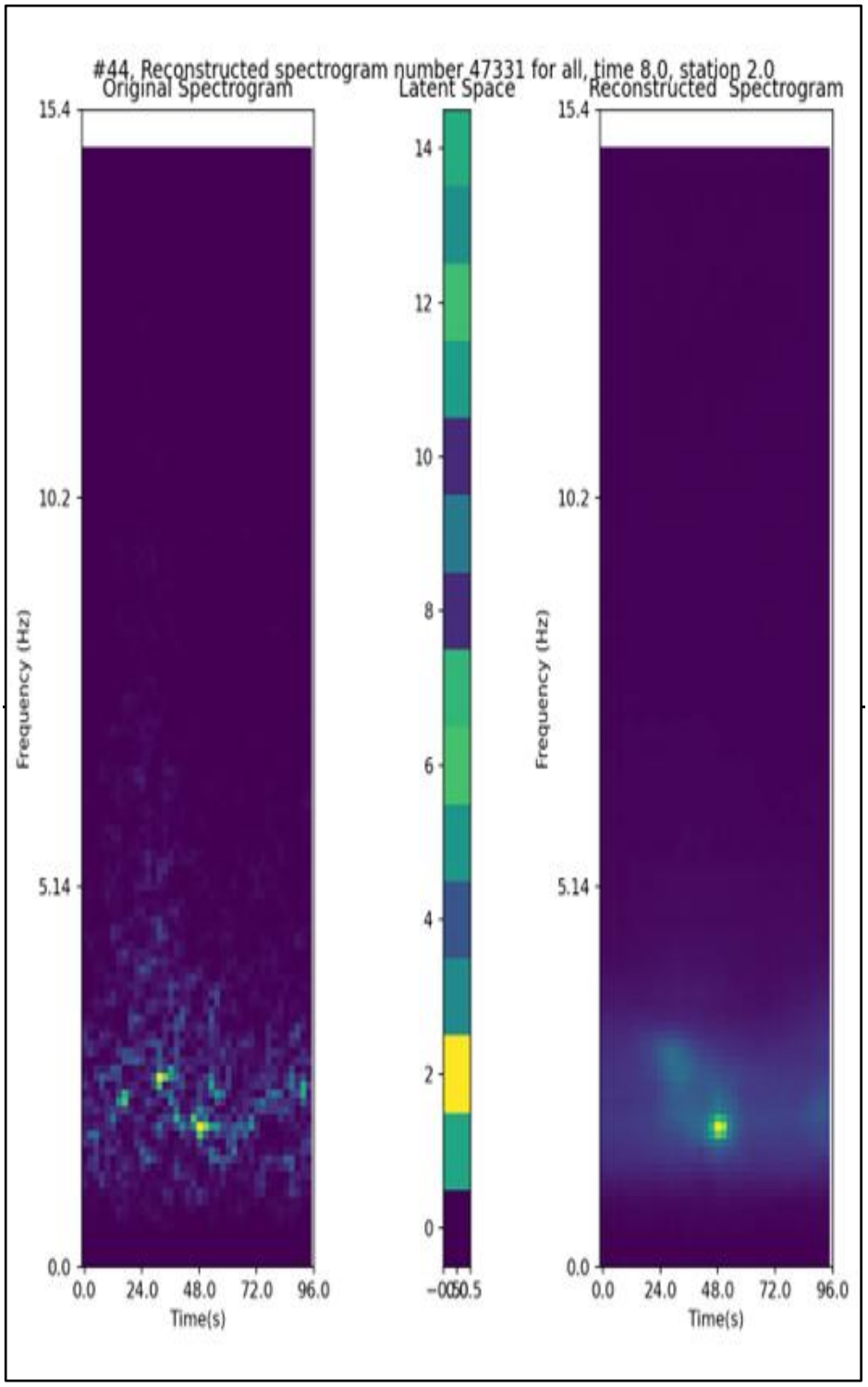


Figure III.3 Examples of reconstructions from the L15 space.

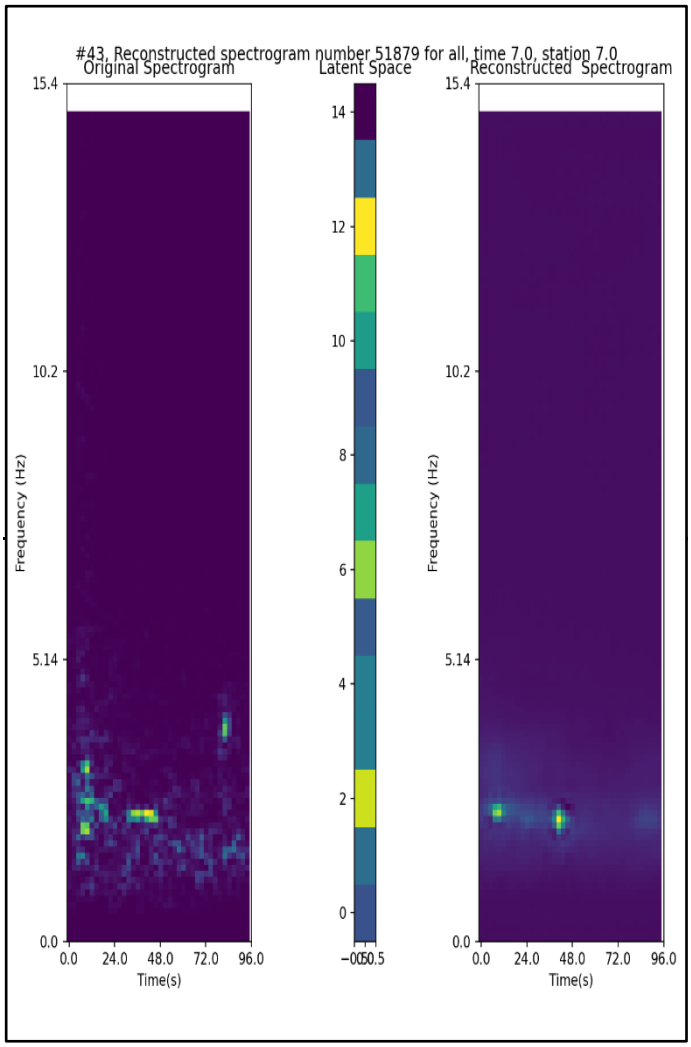


Figure III.4: reconstructed spectrograms I-XIV (left top middle top right top next row left etc etc)

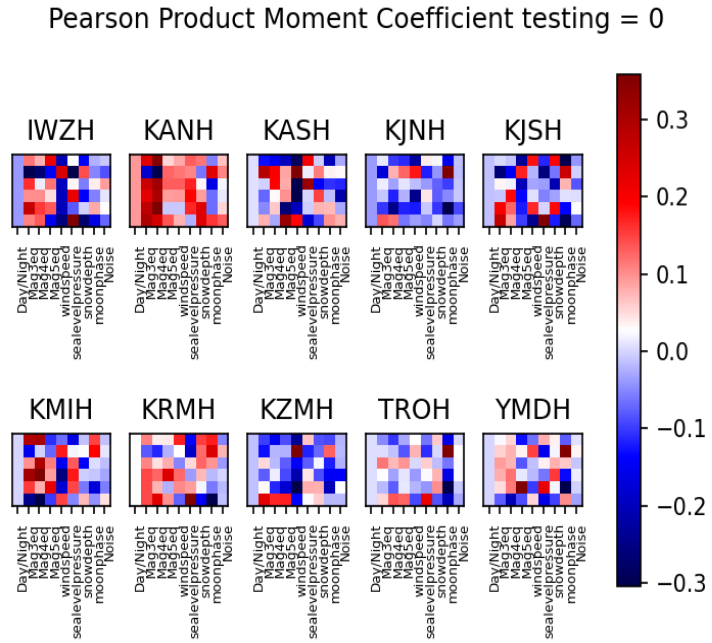
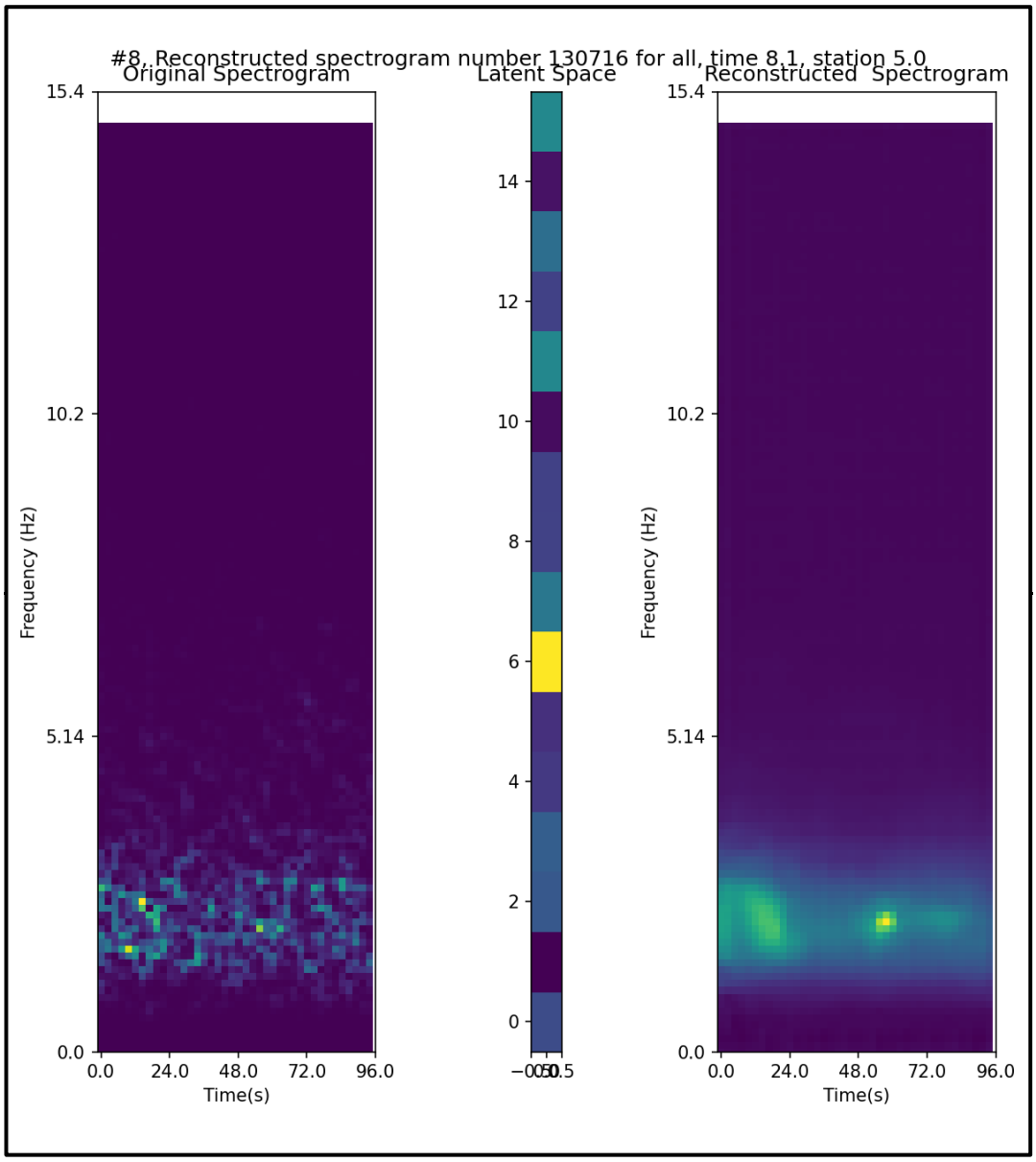
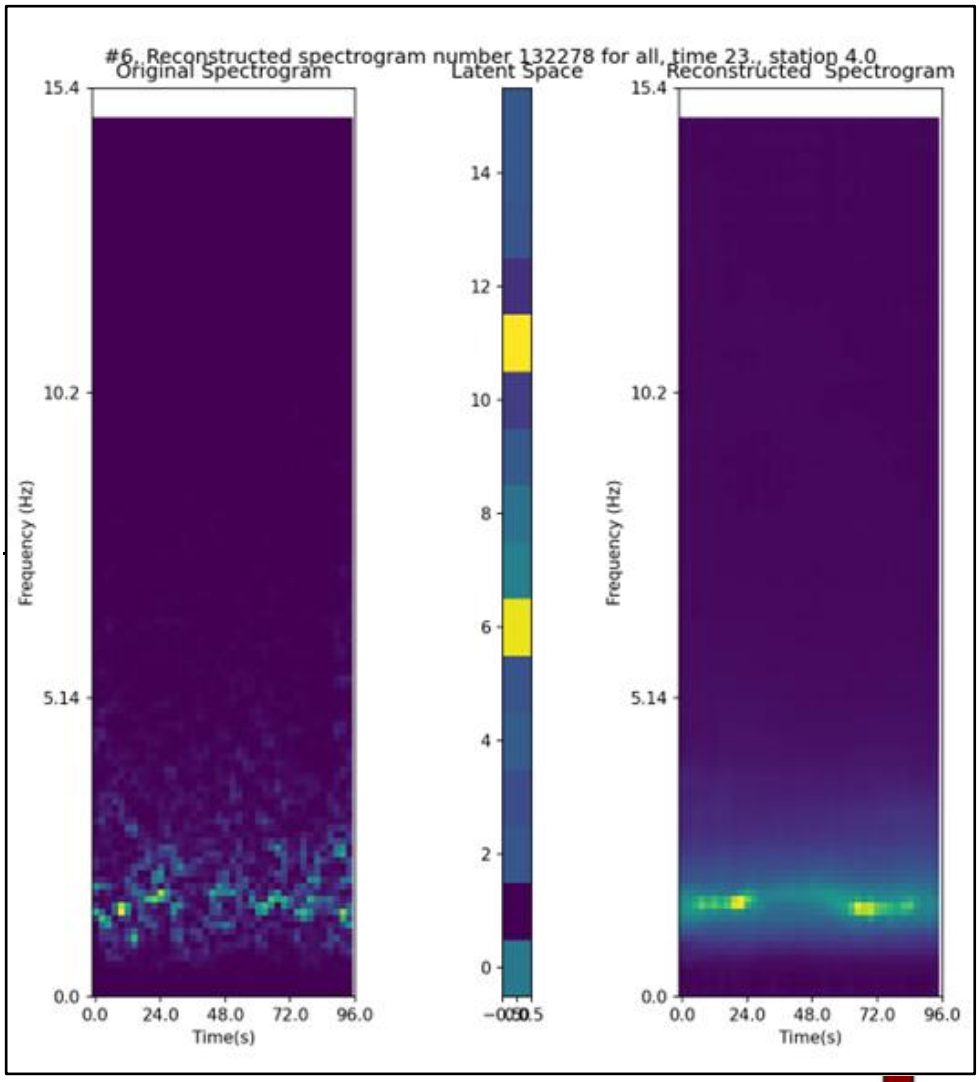
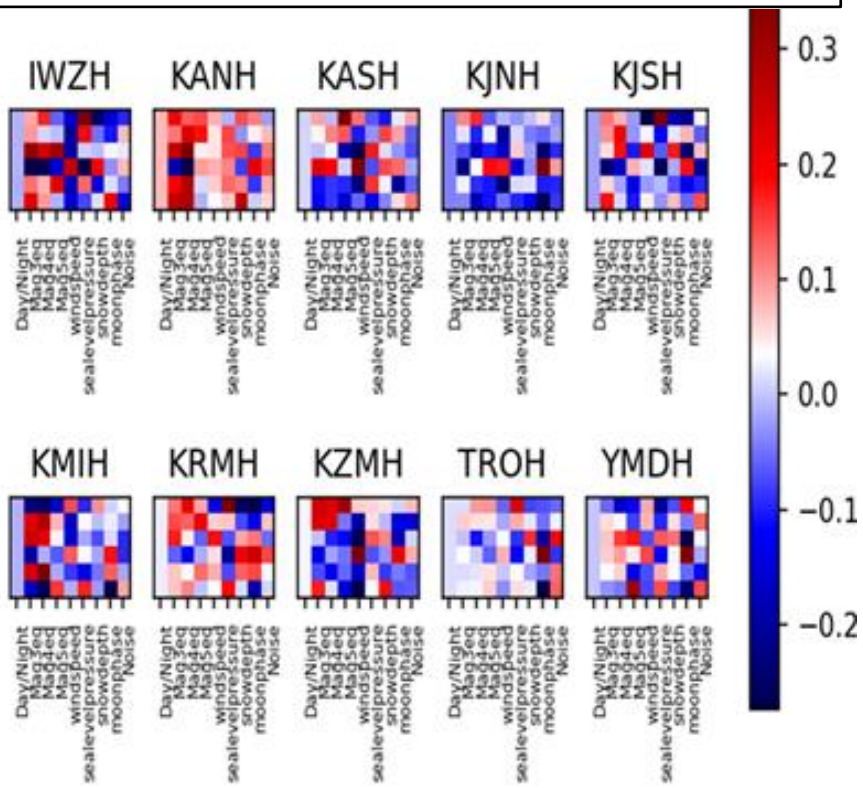


Figure III.5: PCM values for each station and characteristic feature for the L15 space.

L16 Supplementary spectrograms




#6, Reconstructed spectrogram number 132278 for all, time 23., station 4.0
Original Spectrogram
Latent Space
Reconstructed Spectrogram
Frequency (Hz)
Time(s)
15.4
10.2
5.14
0.0
0.0
24.0
48.0
72.0
96.0
14
12
10
8
6
4
2
0


IWZH
KANH
KASH
KJNH
KJSH
KMIH
KRMH
KZMH
TROH
YMDH
Day/Night
Mag3eq
Mag4eq
Mag5eq
windspeed
sealevelpressure
snowdepth
moonphase
Noise
0.3
0.2
0.1
0.0
−0.1
−0.2

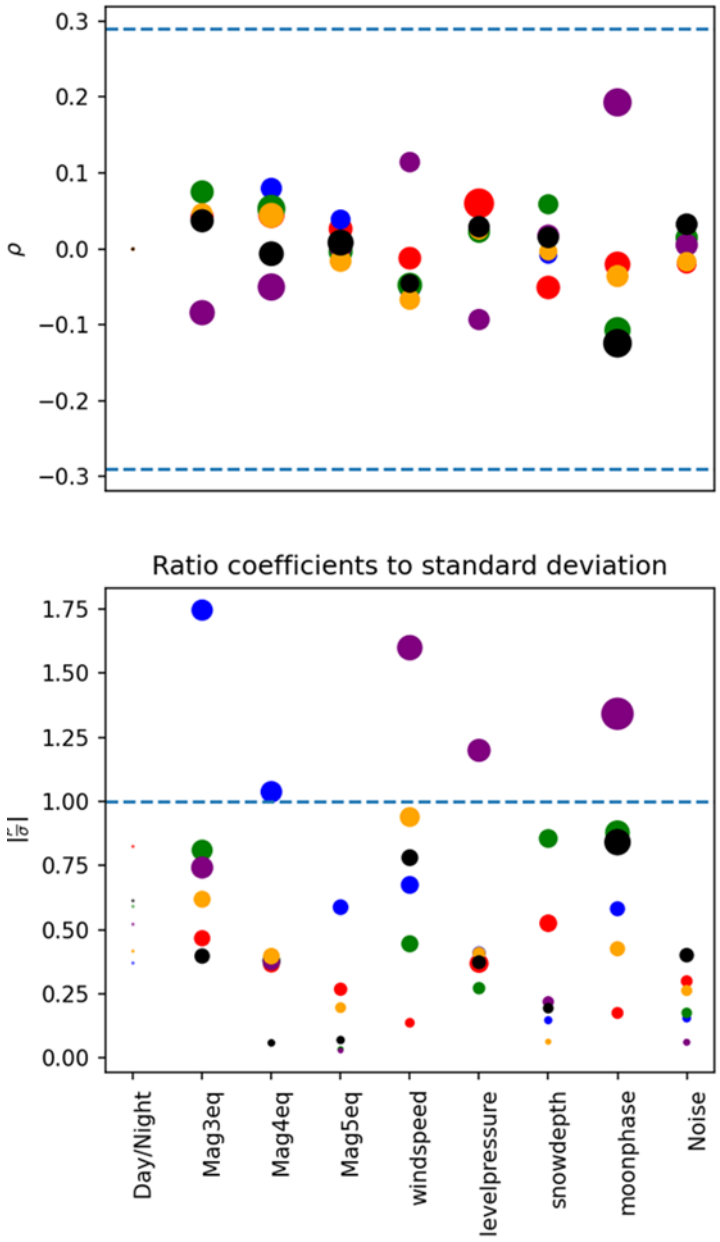
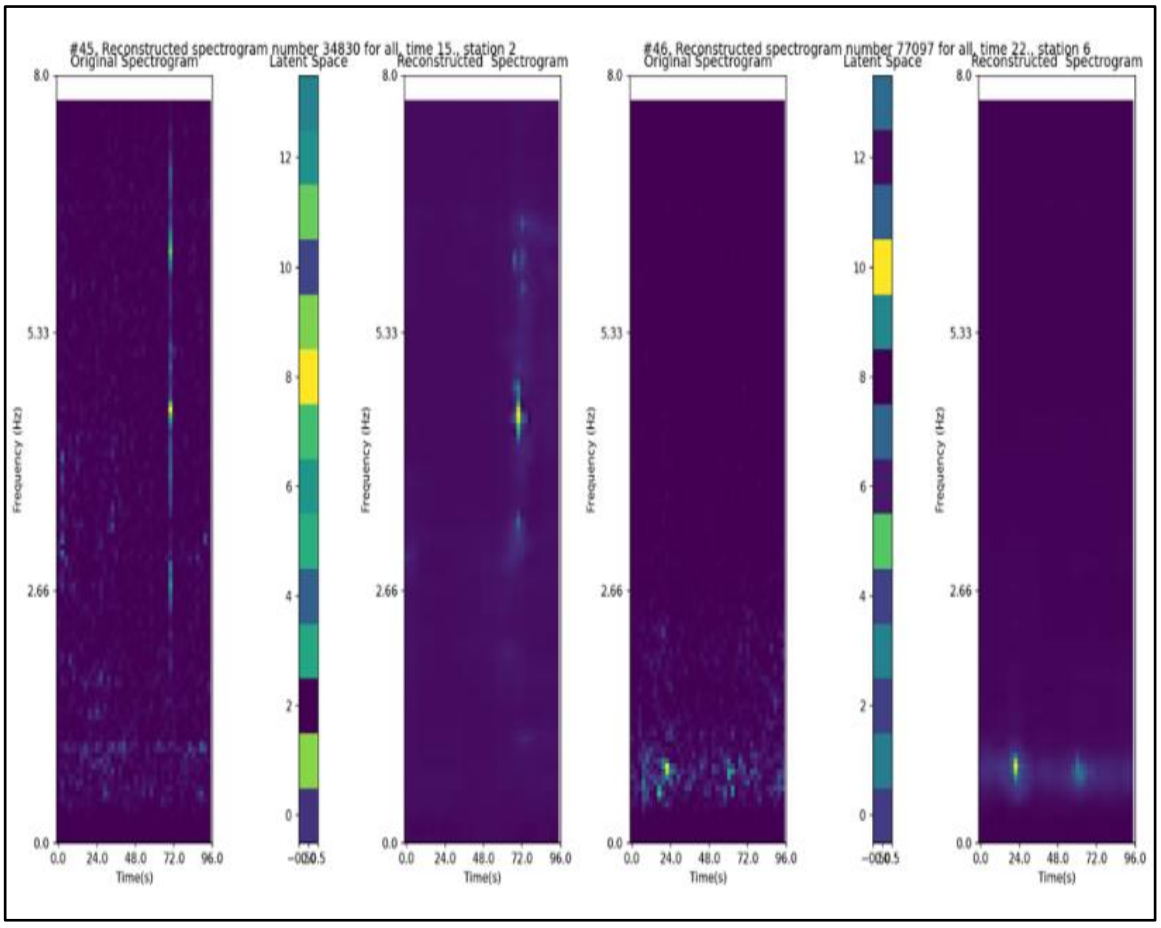
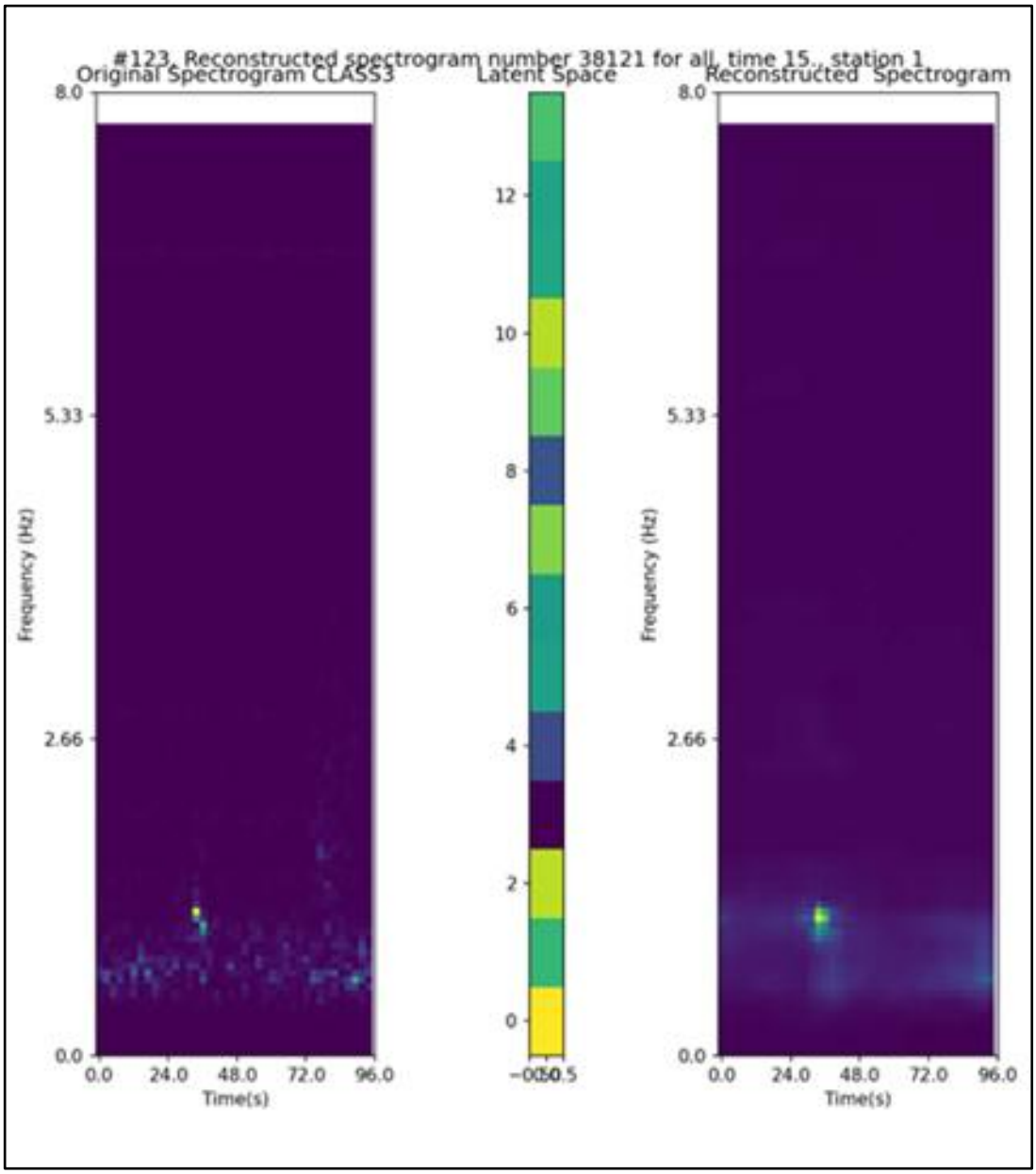
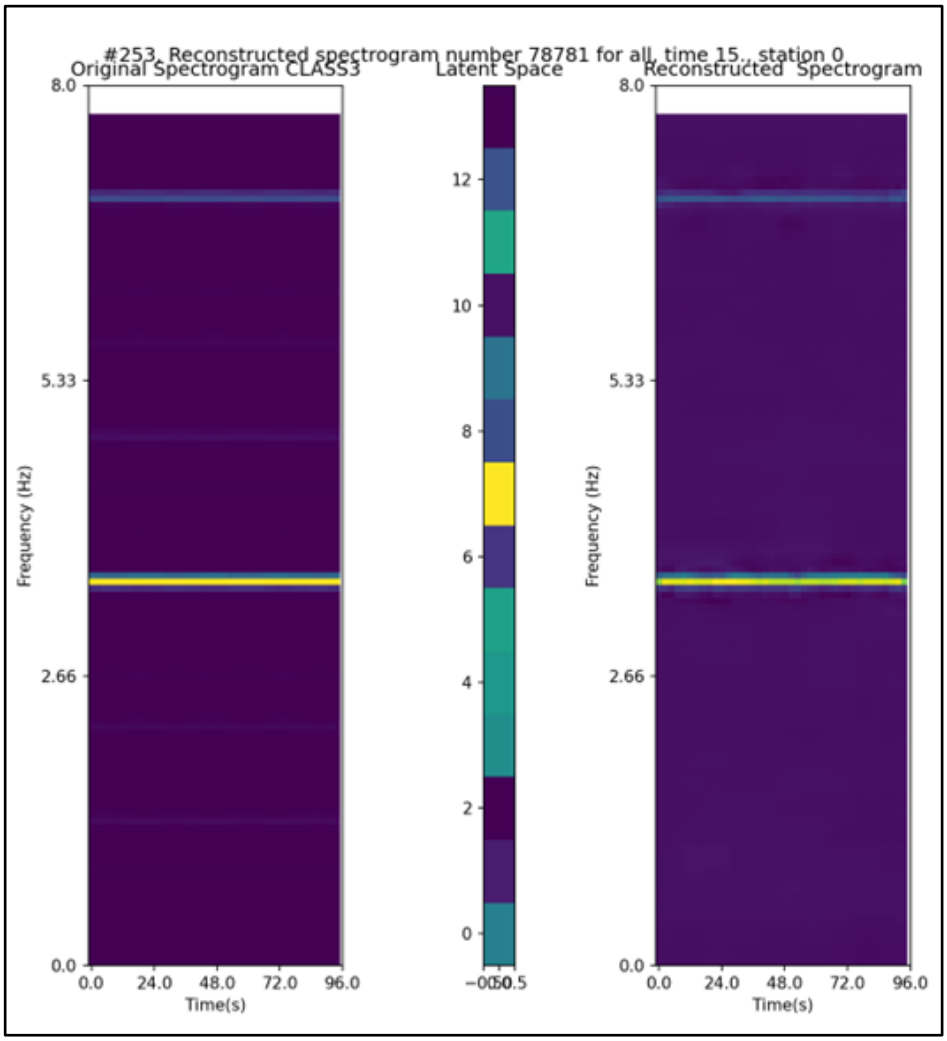


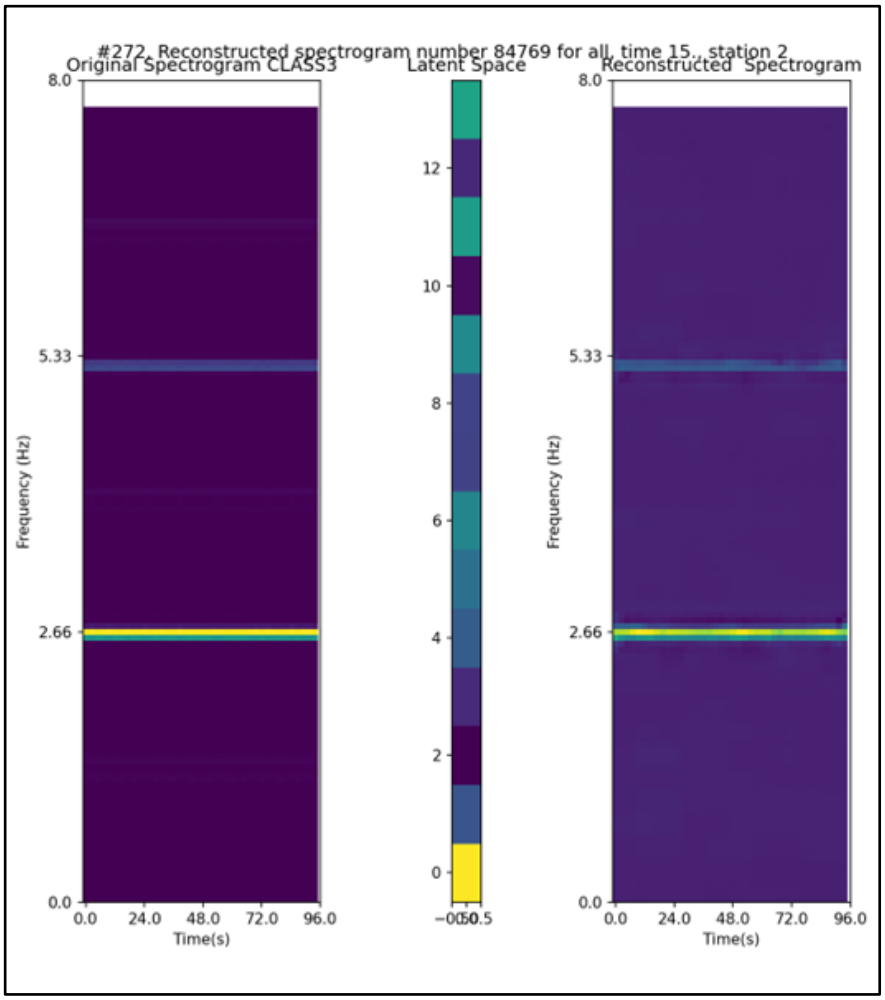
Figure III.6: PCM values for each station and characteristic feature for the L16 space.

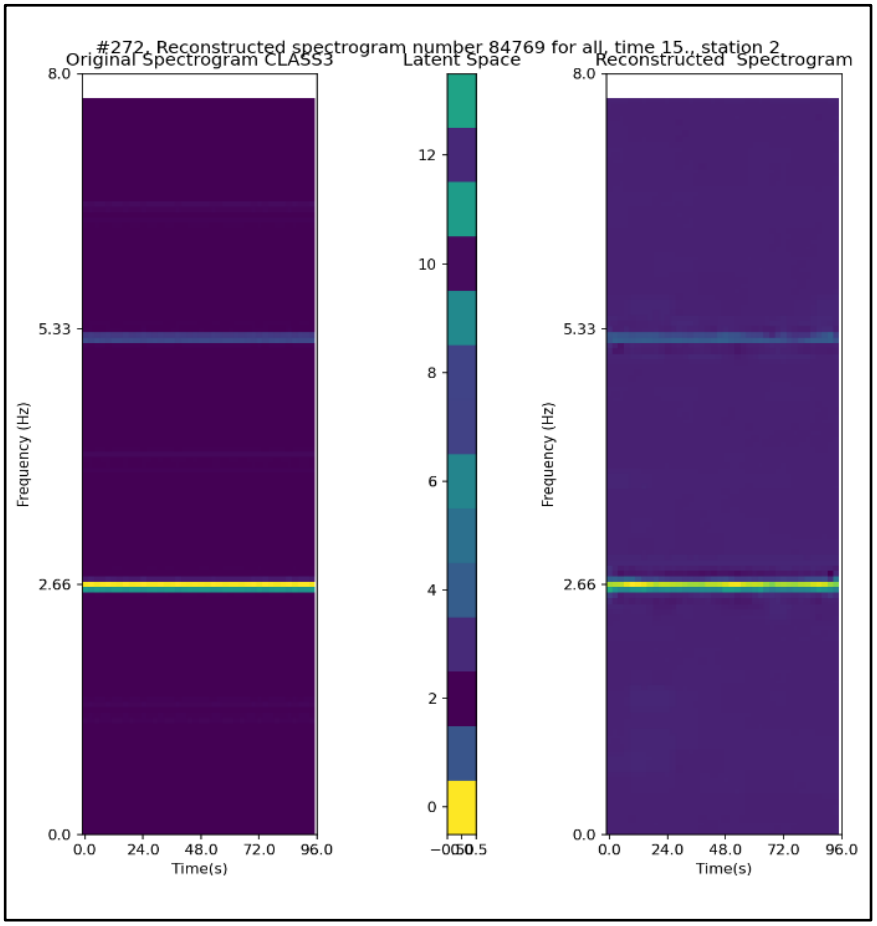
Appendix IV: Supplementary Figures from Chapter 3

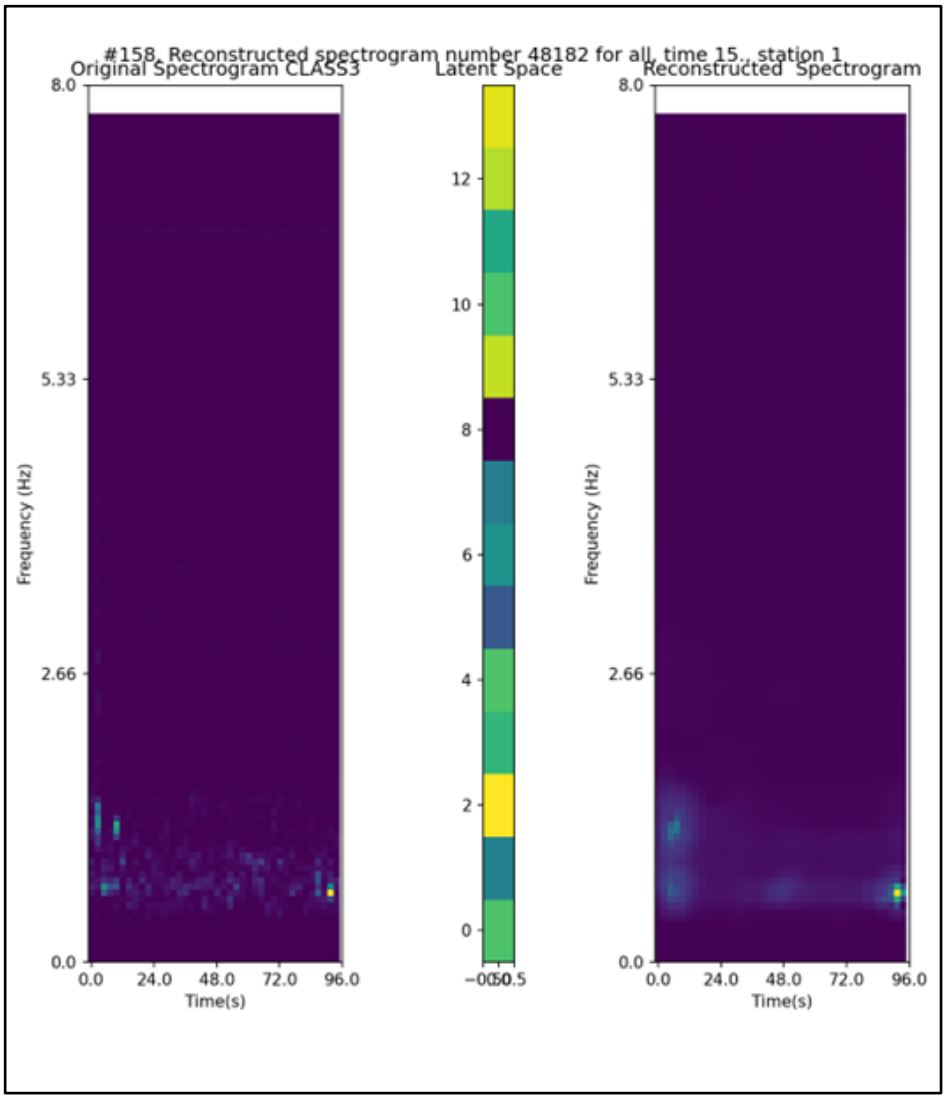


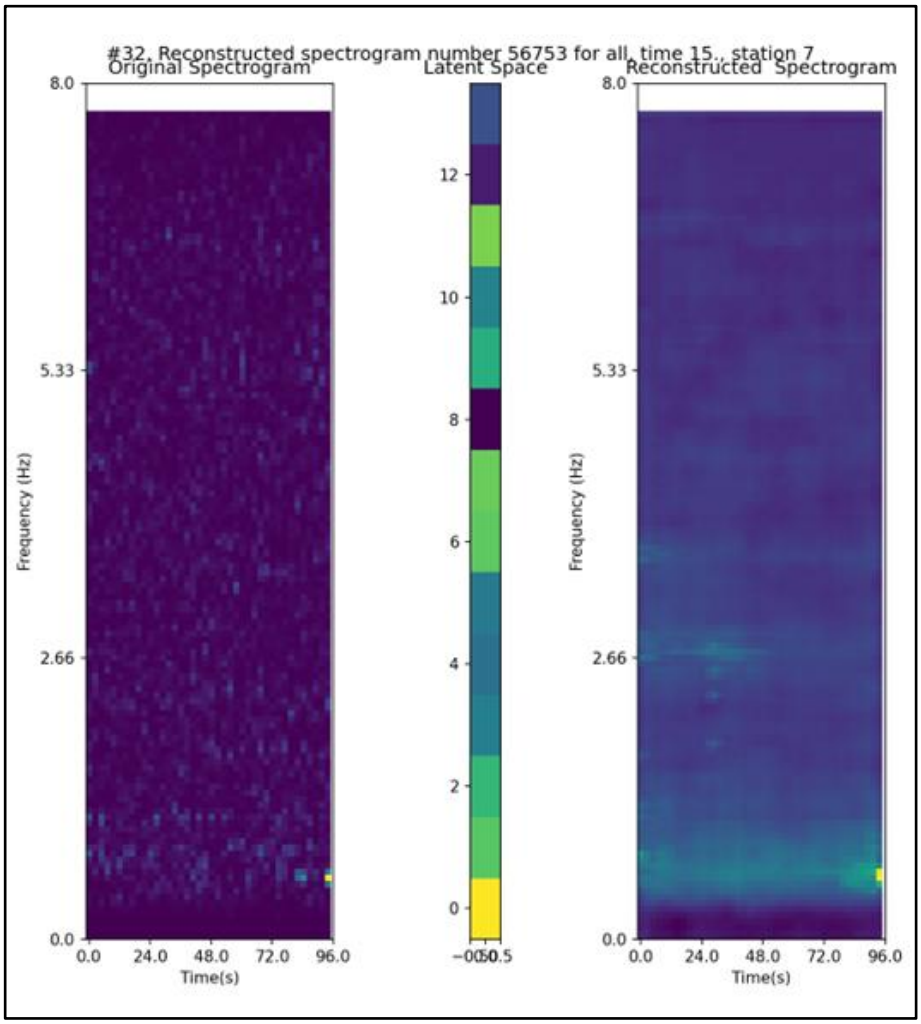

#123, Reconstructed spectrogram number 38121 for all, time 15., station 1
Original Spectrogram CLASS3
Latent Space
Reconstructed Spectrogram
Frequency (Hz)
Time(s)
8.0
5.33
2.66
0.0
0.0
24.0
48.0
72.0
96.0
12
10
8
6
4
2
0


#253, Reconstructed spectrogram number 78781 for all, time 15., station 0
Original Spectrogram CLASS3
Latent Space
Reconstructed Spectrogram
Frequency (Hz)
Time(s)
8.0
5.33
2.66
0.0
0.0
24.0
48.0
72.0
96.0
12
10
8
6
4
2
0


#272, Reconstructed spectrogram number 84769 for all, time 15., station 2
Original Spectrogram CLASS3
Latent Space
Reconstructed Spectrogram
Frequency (Hz)
Time(s)
8.0
5.33
2.66
0.0
0.0
24.0
48.0
72.0
96.0
12
10
8
6
4
2
0
−0.5
0.0
0.5


#272, Reconstructed spectrogram number 84769 for all, time 15., station 2
Original Spectrogram CLASS3
Latent Space
Reconstructed Spectrogram
Frequency (Hz)
Time(s)
8.0
5.33
2.66
0.0
0.0
24.0
48.0
72.0
96.0
12
10
8
6
4
2
0
−0.5
0.0
0.5


#158, Reconstructed spectrogram number 48182 for all, time 15., station 1
Original Spectrogram CLASS3
Latent Space
Reconstructed Spectrogram
8.0
5.33
2.66
0.0
Frequency (Hz)
0.0
24.0
48.0
72.0
96.0
Time(s)
12
10
8
6
4
2
0
−0.5
0.5



Figures IV.1: supplementary spectrograms from the NVT dataset.