Computers & Geosciences 176 (2023) 105361

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

Deep Kernel Dictionary Learning for detection of wave breaking features in Atmospheric Gravity Waves

Varanasi Satya Sreekanth [a,b,*], Karnam Raghunath [a], Deepak Mishra [b]

[a] National Atmospheric Research Laboratory, Department of Space, Gadanki, Pakala, 517112, Andhra Pradesh, India
[b] Indian Institute of Space Science and Technology, Department of Space, Valiyamala, Thiruvananthapuram, 695547, Kerela, India

ARTICLE INFO

Dataset link: https://daac.gsfc.nasa.gov, https://www.narl.gov.in, https://www.alpendac.eu

Keywords:
Dictionary learning
Deep learning
CLSTM
Kernel
Wave breaking

ABSTRACT

Atmospheric Gravity Waves play a significant role in the Middle Atmosphere Dynamics and breaking of Gravity Waves leads to turbulence. To confirm the breaking of Atmospheric Gravity Waves, we require Wind Velocity profiles with high accuracy and the problem in the existing scenario is the Wind velocity measuring instruments are not accurate at the altitude of interest and are very sparse. We came up with a solution by taking advantage of Dictionary Learning and Deep Learning approach for the detection of the Wave Breaking events from atmospheric temperature perturbations instead of looking for wind velocity profiles.

In the present work, we discuss incorporating Kernels into Deep Dictionary Learning. The Deep Dictionary Learning algorithm, introduced recently, shows an improvement in feature detection in comparison to k-means Singular Value Decomposition (KSVD) Dictionary Learning, Convolution Neural Network, and LSTM Auto-Encoders. Incorporating Kernels into Deep Dictionary Learning increases the effectiveness of the detection of the Wave Breaking events. The potential of the proposed method is demonstrated with a case study on the detection of atmospheric Gravity Wave Breaking event leading to turbulence in the atmosphere using satellite data (Aura-Microwave Limb Sounder) and is validated using data available from the ground-based instruments.

1. Introduction

A numerical study on the impact of Gravity Wave Breaking on the mean state is well explained by Liu et al. (1999a), considering the non-linear dynamics of the Gravity Waves. They also described the interaction between wave breaking and turbulence. Simultaneous interactions between wave breaking, turbulence, mean state, and the mesosphere are explained in their work. A set of thresholds for the breaking of Gravity Waves and multiple gravity wave breaking levels are discussed by Achatz (2007) and Lindzen (1985), where the dynamics of all the processes are explained within a framework that considers linear stability analysis to simulations. They also explained how Gravity Wave Breaking leads to turbulence. A theory on the effect of wave breaking on eddy diffusion, the relation of wavenumbers to wave periods, and mean stress are discussed by Lindzen (1984, 1981, 1973). The interaction vertically propagating Gravity Wave with zonal wind is explained in Lindzen and Holton (1968). The influence of Gravity Wave Breaking on the general circulation in the middle atmosphere is discussed by Holton (1983) and Garcia and Solomon (1985). They proved that mechanical dissipation by the breaking of the vertical component of the Gravity Wave is essential for the mean global circulation in the Mesosphere . Also, they showed that gravity wave drag and diffusion in the Mesosphere account for large departure from radiative equilibrium in summer and winter. This motivates us to detect wave-breaking events with high accuracy to better understand global circulation.

The linear wave theory of atmospheric Gravity Waves considers that waves vary only in the vertical direction from a stably stratified background state. The linear wave theory follows the rules of conservation of momentum, mass, and energy (Holton, 1973). The departure from linear wave theory to non-linear wave theory is to include dissipation via wave breaking, transience, localization, and variable environments. The deviations from the linear theory also occur because of the time dependence of the background flow and the spatial and temporal variations in wave activity. A detailed discussion on linear wave theory and its departure to non-linearity are explained in detail in Fritts and Alexander (2003).

The Gravity Waves in the atmosphere play a significant role in global circulation. The breaking of Gravity Waves triggers turbulence in the atmosphere. To study turbulence, we require wind velocity measurements with high accuracy and the drawback is the sparse availability of wind velocity measuring instruments with high accuracy. As the Gravity Wave excited at a low altitude and propagated upward,

* Corresponding author at: National Atmospheric Research Laboratory, Department of Space, Gadanki, Pakala, 517112, Andhra Pradesh, India.
E-mail address: varanasi_sreekanth@yahoo.com (V.S. Sreekanth).

https://doi.org/10.1016/j.cageo.2023.105361
Received 24 February 2022; Received in revised form 3 April 2023; Accepted 5 April 2023
Available online 14 April 2023

they got trapped where the horizontal phase speed of Gravity Waves was equal to the horizontal background wind in the direction of propagation. The Gravity Waves are superimposed with jet streams at an altitude between 15–25 km. The vertical displacement field (ζ) for a traveling wave due to perturbation is given by

$$\zeta(x,z,t) = A_{\zeta}\cos(k\mathbf{a_x} + m\mathbf{a_z} - \hat{\omega}t) \tag{1}$$

where k and m are the real-valued horizontal and vertical wavenumbers.

A Gravity Wave Breaking event occurs when the phase velocity (C_p) of the Waves is greater than the background horizontal wind ($\bar{\mathbf{U}}$). The Gravity Waves generated by the sources in the Troposphere propagate vertically with increasing amplitude in temperature perturbations with a decreasing density that induces convective instability ($N^2 < 0$, or $R_i < 0$) or shear instability ($R_i < 0.25$) at some height called 'breaking level' and is the height where the turbulence effects start prevailing. The Brunt Vaisala Frequency (N) and Richardson Number (R_i) are the indicators of instabilities in the atmosphere and are calculated using (2) and (5). Gravity Waves play an important role in the transport of energy and momentum vertically between the lower and upper atmosphere and are examined from the perturbed temperature profiles obtained ($T'(z,t)$) from Rayleigh lidar observations. The temperature perturbation $T'(z,t)$ is calculated as

$$T'(z,t) = T(z,t) - T_0(z,t) \tag{2}$$

$$N^2(z) = \frac{g}{T_0(z,t)}\left[\overline{\frac{dT_0(z,t)}{dz}} + \frac{g}{C_p}\right] \tag{3}$$

$$E_p(z) = \frac{1}{2}\left[\frac{g}{N(z)}\right]^2\left[\overline{\frac{T'_{break}(z,t)}{T_0(z,t)}}\right]^2 \tag{4}$$

$$R_i = \frac{N^2}{(\frac{\partial u}{\partial z})^2} \tag{5}$$

$$\frac{dT}{dz} = \frac{g}{T_0(z,t)}\left[\overline{\frac{dT'(z,t)}{dz}}\right] \tag{6}$$

where $N(z)$ is the Brunt-Vaisala frequency, $T'_{break}(z,t)$ is temperature perturbation at breaking altitude, $T_0(z,t)$ is mean of all temperature profiles at the respective altitude(z) over the duration of the observation, g is the acceleration due to gravity (9.8 $\mathrm{m\,s^{-2}}$), C_p is the dry air specific heat at constant pressure (1005 $\mathrm{J\,Kg^{-1}\,K^{-1}}$). The $\frac{dT'(z,t)}{dz}$ is calculated by using numerical differentiation over the 10 km height range that gives the change of rate of T_0 concerning to z. The following permutations of N^2 and R_i are given below:

- If $N^2 > 0$, and $R_i < 0.25$, represent refraction of internal Gravity Wave with the extraction of energy from the mean flow.
- If $N^2 > 0$ and $0.25 < R_i < 2$, represent refraction of Gravity Wave at Ducting height with a transfer of energy to the mean flow. The larger the value of R_i, the more the loss of energy to the mean flow.
- If $R_i > 2$, the wave rapidly transfers energy to the mean flow.

The interpretation of Wave Breaking events from Temperature perturbations and Wind profiles is made by identifying the presence of a large negative gradient (10 $\mathrm{km\,K^{-1}}$) of temperature and a strong wind shear of −40 $\mathrm{m\,s^{-1}\,km}$ indicate the presence of reduced convective and dynamic stability. The steepening of the potential temperature contours (i.e., contour lines become vertical) indicates reduced convective stability.

The atmospheric Gravity Waves play an important role in determining the general circulation in the middle atmosphere by the dynamic process caused by the breaking of the atmospheric Gravity Waves. The Gravity Waves get trapped at the altitude where the horizontal phase speed of the Gravity Wave becomes equal to the horizontal background wind speed in the direction of propagation. The wave cannot surpass this altitude and gets attenuated. The altitude at which the wave gets trapped is called the 'critical layer', and critical absorption is observed at this altitude. We particularly concentrate on the Mesosphere (50–90 km), where significant attenuation of Gravity Waves is observed due to non-linearity as reported in Tsuda (2014). The evidence of the detection of Gravity Wave breaking from lidar observations is given by Franke and Collins (2003). At this point, the proposed method, Deep Kernel Dictionary Learning (DKDL) has the advantage over the conventional methods, as discussed in detail in the following sections. The variations in the vertical wind profile coincide with the temperature perturbations. In the Mesosphere, the Gravity Waves have enough strength to disturb the background wind and temperature profile significantly and lead to low atmospheric stability. A detailed discussion on the behavior of Gravity Waves in the atmospheric layers in a general view is reported in Fritts (1984), Tsuda (2014), Booker and Bretherton (1967) and Lindzen (1984, 1981). Eqs. (5) and (6) quantify the stability at the respective altitude.

Atmospheric turbulence plays a significant role in the coupling of layers in the atmosphere. A detailed study on conditions under which breaking of Gravity Waves leads to turbulence is reported in Hines (1988). A numerical study on the breaking of Atmospheric Gravity Waves and its impact on turbulence is reported in Liu et al. (1999b). A detailed study on the impact of Gravity Wave Breaking on the middle atmosphere is reported in Fritts and Alexander (2003), Fritts et al. (2003) and Hines (1971).

Using Deep Learning and Dictionary Learning techniques, we extract the Wave Breaking features from temperature perturbations that are densely available for the middle atmosphere region. The methodology for the detection of features is explained in detail in the following sections.

The application of Deep Learning (DL) techniques in the field of atmospheric Remote Sensing is providing significant results for both feature detection and classification problems for spatial and temporal domain data. The wide variety of applications of Deep Learning in the field of Atmospheric Remote Sensing include the detection of turbulence (Vorontsov et al., 2020), estimation of atmospheric Gravity Wave parameters (Matsuoka et al., 2020), parameter profile retrieval techniques (Malmgren-Hansen et al., 2019), detection of atmospheric boundary layer height (Kumar et al., 2021), improvement of air quality prediction (Ma et al., 2019), and forecasting (Qiao et al., 2019).

The application of DL to non-linear problems is achieved through the introduction of kernels that map the data to higher dimensional feature space where the classification and regression tasks become easier. The DL with the introduction of kernels in sparse representations in the k-means Singular Value Decomposition (KSVD) called Kernel KSVD (KKSVD) is reported in Nguyen et al. (2012). In KKSVD, two Dictionaries are learned, the 'Base Dictionary' that acts as a feature map for all the signals, , and the 'coefficient Dictionary' is the Dictionary learned on updating the Dictionary.

In the Deep Kernel Dictionary Learning (DKDL), the feature-mapped Dictionary obtained in the present neural network is fed as input to the succeeding neural network layer, and the process is repeated for the number of the networks used in DKDL. The potential of the DKDL is verified in the following aspects.

- Improvement in Detection Rate of Wave Breaking events with the introduction of Kernel function into Deep Dictionary Learning (DDL).
- we compare the performance of DKDL with that of the existing feature extraction methods including Conventional Convolutional Long Short Term Memory (CLSTM) network and Kernel Convolutional LSTM (KCLSTM) (Agethen and Hsu, 2020).

In this work, we explain the potential of the proposed method with a case study on the detection of Atmospheric Gravity Wave Breaking events. We detect Atmospheric Gravity Wave Breaking events from the vertical temperature profiles obtained using Microwave Limb Sounder

(MLS) on the Earth Observation Satellite Aura. The detection of Wave Breaking events is verified using Temperature perturbations and wind velocities obtained using ground-based Rayleigh Lidar and Meteor Radar respectively.

The paper is organized in the following sections. Section 2 provides a brief introduction to classical Deep Learning and Dictionary Learning, and Deep Dictionary Learning. Section 3 is about the proposed method of Deep Kernel Dictionary Learning. Section 4 is on the Kernel Density Estimation for obtaining the optimal bandwidth of the probability distribution function. Section 5 demonstrates the potential of the proposed method with a case study on the detection of Wave Breaking in Atmospheric Gravity Wave Signatures. Section 6 is the conclusion about the performance of the method.

2. Dictionary Learning and Deep Learning

Any given data matrix can be represented as the product of the Dictionary matrix and Sparse matrix. The column vectors of the Dictionary matrix are called basis vectors or atoms. The column vectors forming the Sparse matrix are called sparse coefficients. The data in the sparse matrix are called Sparse coefficients. All the atoms are combined to form a Dictionary (Mallat, 1999; Candes and Donoho, 2002; Do and Vetterli, 2002). The most popular Dictionary learning (DL) algorithm is the KSVD which generalizes the k-means clustering and learns a Dictionary that specifies the input data (Aharon et al., 2006). Later Discriminative Dictionary Learning (Mairal et al., 2009) came for classification tasks by incorporating the classifier parameters in the optimization problem of DL and introduction of classifier parameters in the KSVD algorithm is proposed by Zhang and Li (2010) and is used for solving the optimization problem of the proposed method. In Cai et al. (2014) assigning weights for sparse vectors while learning the Dictionary is proposed. The detailed procedure for the inclusion of Kernels in Deep Neural Networks for feature recognition is reported in Agethen and Hsu (2020). The present section is on the background for combining the Deep Learning and the Dictionary Learning techniques (Tang et al., 2021). We also discuss Deep Dictionary Learning and Kernel Dictionary Learning.

2.1. Deep Learning

A CLSTM (Malmgren-Hansen et al., 2019; Courtney and Sreenivas, 2019) with a larger kernel can identify sudden changes. The hyper-parameter (Schratz et al., 2019)(weights, biases, drop rate, learning rate) optimization is done through the Grid search method. The problem of overfitting in the present work is suppressed by Batch Normalization (BN) and Dropout techniques. In the BN process, the inputs to the hidden layers are standardized with the mean (μ), and standard deviation (σ) of the previous layer using (7) and are applied before the activation function.

$$BN(u) = \frac{\gamma(u-\mu)}{\sigma} + \beta \tag{7}$$

where γ and β are rescaling and shifting parameters, and u represent an arbitrary variable. In the Dropout technique, the neurons along with their connections are removed from the network at probability (r) (called Drop rate) by multiplying its activation by zero and helps to reduce variance by combining multiple independent models. The network depth or width is increased to counter the decrease in the model capacity.

We employ the mini-batch (a subset of the training data set) based methods that involve updating the subset rather than the whole data, which decreases the time required for training.

2.2. Dictionary Learning

The DL aims at finding a Dictionary ($\mathbf{D} \in \mathbb{R}^{m\times k}$), that sparsifies ($\mathbf{X} \in \mathbb{R}^{k\times n}$) a set of input vector $\mathbf{Y} \in \mathbb{R}^{m\times n}$. The DL optimization problem is given by (8) and is solved by using the Method of Optimal Directions (MOD) (Engan et al., 1999) and KSVD (Aharon et al., 2006).

$$\min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_{\mathbf{F}}^{\mathbf{2}} + \lambda\|\mathbf{X}\|_{\mathbf{1}} \tag{8}$$

where F represent the Frobenius norm or l_2-norm, λ represent the regularization parameter. In general (8) gives information about the basic mathematical expression for DL that minimizes the residual by keeping constraints on the residual.

2.3. Deep Dictionary Learning

Multi-layer feature detection with Deep Learning combined with DL, leading to Deep Dictionary Learning (DDL), is proposed in Tariyal et al. (2016). For the output of each Neural Network, the Dictionary (**D**) and the Sparse matrix (**X**) are learned. The Sparse matrix obtained in the previous level is used as input data to the Neural Network. The performance of the Dictionary learned for feature detection from the total model of the neural network is not the same as the performance of the effective Dictionary learned from the individual layers. As the number of layers is being increased, the methodology is more prone to over-fitting. To overcome the over-fitting problem, we use the Batch Normalization (BN) technique at the end of each layer by the mean and standard deviation of the data obtained at the output of the present layer.

The DDL algorithm with multi-level DL in general with network layers having non-linear activation functions (Φ) is given by

$$X = D_1\Phi_1(D_2\Phi_2(..D_{N'}\Phi_{N'}Y)) \tag{9}$$

The multi-level DDL problem has the form as (10).

$$\min_{D_1,\ldots,D_{N'},X} \left\|Y - D_1\Phi\left(D_2\Phi(\ldots\Phi(D_{N'}X))\right)\right\|_F^2 + \lambda\|X\|_1 \tag{10}$$

where $1, 2, \ldots, N'$ represent the number of Neural Networks.

3. Methodology: Deep Kernel Dictionary Learning

Section 2 gave information about the relation between Deep Learning and Dictionary Learning, the introduction of Kernels into Dictionary Learning, and Deep Dictionary Learning techniques.

The training procedure for the implementation of the Deep Kernel Dictionary Learning method for any kind of activation and Kernel function is given in Algorithm 1. It can be noted that learning Dictionaries at multiple levels cannot be represented as a single dictionary even in the case of linear activation function as the DL is a bi-linear. For example: if the dimensionality of the sample is m, the first dictionary is of size $m \times n_1$ and the second dictionary is of size $n_2 \times n_3$, we are not able to get a single effective dictionary of size $m \times n_3$.

3.1. Kernel approximations

There are two types of methods for the inclusion of kernel functions. The first one is to use the Kernel Trick and the second is to explicitly use the feature maps. The usage of explicit feature maps has the advantage of significantly reducing the cost of learning when compared to the Kernel Trick with very large data sets. We select the Kernel Trick for Kernel approximation in the present work rather than the Nystroem method (Williams and Seeger, 2001; Yang et al., 2012a; Drineas et al., 2005).

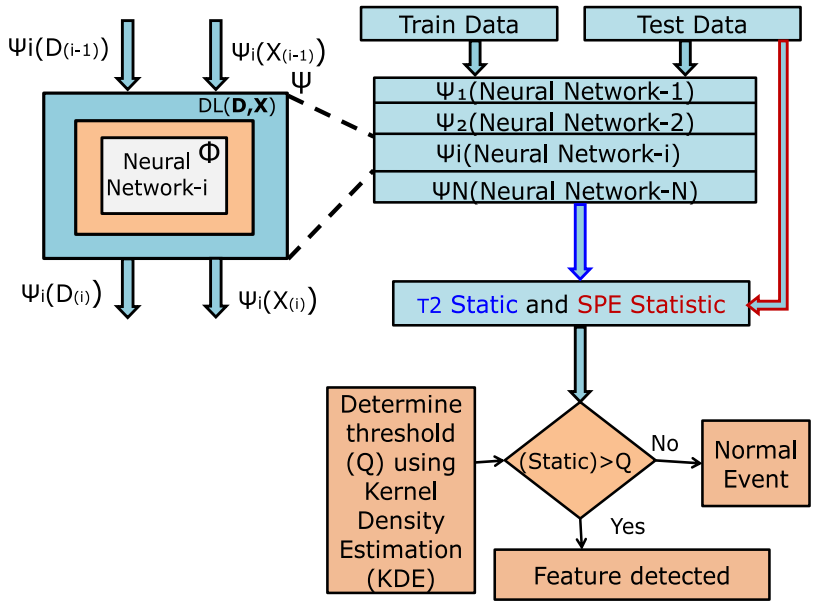

Fig. 1. Deep Kernel Dictionary Learning approach for feature detection. T^2 and SPE metrics greater than the threshold limit represent the detection of a feature. The highlighted block is the internal representation of incorporating kernels in the Deep Dictionary Learning technique.

3.2. Kernel Dictionary Learning

In this section, we discuss the inclusion of Kernels in DL (Nguyen et al., 2012), and the proposed method of Deep Kernel Dictionary Learning. Let $\Psi : \mathbb{R}^n \rightarrow \mathbb{F}$ represent a non-linear mapping from $\mathbb{R}^n$ to a higher dimensional space $\mathbb{F}$. At higher dimensional spaces, Mercer Kernels are commonly employed for carrying out the feature map explicitly. A Mercer Kernel is a function $\kappa(\mathbf{x}, \mathbf{y})$ that maps all the input data $\mathbf{y_i}$ that results in a positive semidefinite matrix $\kappa(\mathbf{y_i}, \mathbf{y_j})$. In this work, we use the following Dictionary model: $\mathbf{D} = \mathbf{BA}$, where $\mathbf{B}$ is a predefined basis, and $\mathbf{A}$ is the adaptively learned Dictionary. The input data $\mathbf{Y}$ mapped to a higher dimensional space is given by $\mathbf{\Psi(Y)}$ and the learned Dictionary in the feature space as $\mathbf{\Psi(D)}$. The learned Dictionary in Feature space is given by (17).

$$\mathbf{\Psi(D)} = \mathbf{\Psi(Y)A} \tag{11}$$

The optimization problem for kernel Dictionary Learning (KDL) is given by (6). It is reported in Nguyen et al. (2012) that non-linear DL can better Discriminate the features than the linear DL.

$$\underset{\mathbf{A},\mathbf{X}}{\operatorname{argmin}} \|\Psi(\mathbf{Y}) - \Psi(\mathbf{Y})\mathbf{AX}\|_F^2 \quad s.t. \quad \forall i = 1..n \quad \|\boldsymbol{X}_i\|_0 \leq T_0. \tag{12}$$

where T_0 represents the number of atoms selected. The classical KDL optimization problem (12) is solved by using MOD or KSVD and the objective function of the KDL can be represented in Kernel form as (13) and (14).

$$\|\Psi(\mathbf{Y}) - \Psi(\mathbf{Y})\mathbf{AX}\|_F^2 = \mathbf{tr}((\mathbf{I} - \mathbf{AX})^T \mathbb{K}(\mathbf{Y}, \mathbf{Y}) \ (\mathbf{I} - \mathbf{AX})) \tag{13}$$

$$[\mathbb{K}(\mathbf{Y}, \mathbf{Y})]_{ij} = \left[\langle \Psi(\mathbf{Y}), \Psi(\mathbf{Y}) \rangle\right]_{ij} = \kappa(y_i, y_j) \tag{14}$$

Implementation of (9), (10), (11), (13), and (14) in (12) make the proposed Deep Kernel Dictionary Learning method as represented in Algorithm 1 and the optimization problem is solved using the KSVD algorithm. In general, for the case of linear activation functions, the dictionaries at multiple levels can be collapsed to get a single effective Dictionary that isa product of Kernel function transformed Dictionaries obtained at each Neural Network layer. The expression for obtaining the effective dictionary and sparse coding phase is given in Algorithm 2 and features are extracted using l_1-minimization. In the case of non-linear activation functions, it is not possible to get a single effective dictionary for implementing the testing phase. In this case, the testing phase is implemented as per the procedure described in Algorithm 3.

3.3. Relevance to existing algorithms

In this section, we compare the proposed method with popular methods like Convolutional LSTM, Kernelized Convolution LSTM, and Auto-Encoder. In all the methods, the aim is to find weights and represent the given data. They differ from each other in the neural network layer configuration used. The methods CLSTM and KCLSTM differ from DKDL in Dictionary Learning, and the Kernel function used. A comparison of the proposed method with Auto-Encoder can be performed. The Dictionary Learning synthesis problem for the DKDL is given by Algorithm 1. The formulation of the optimization problem in Algorithm 1 is well known to all. There also exists another way of finding sparse representations where we aim to learn a dictionary that sparsifies the test data in the analysis domain and is equivalent to the denoising Auto-Encoder(AE) as reported in Tariyal et al. (2016) and is equivalent to hidden layer with a linear activation function. The mathematical expression for Dictionary Learning in the analysis domain is expressed in Algorithm 1.

$$\hat{Y} = \Psi_1(D_1)\left(\Psi_2(D_2)\left(\ldots \Psi_{N'}\left(D_{N'}X\right)\right)\right) = D_S X \tag{15}$$

$$\left\|Y - \hat{Y}\right\|_F^2 + \lambda \|X\|_1 \tag{16}$$

where D_S and D_A are the effective dictionaries in the synthesis domain and analysis domain for a linear activation function. The sparsity term is neglected from (16), and the modified optimization problem is given as (17) that represents a denoising Auto-Encoder (Salakhutdinov and Hinton, 2009).

$$\|Y - D_S D_A \hat{Y}\|_F^2 \tag{17}$$

3.4. Data independent Kernel maps

The Kernel maps are broadly classified into data-independent and data-dependent Kernel maps. The revolutionary idea of creating data-independent randomized kernel maps originated from the work reported in Rahimi and Recht (2008). In the present work, we use the Radial Basis Function (RBF) that belongs to the category of data-independent Kernel map generation. The Fourier Transform of the RBF kernel $\kappa(\Delta) = \exp(\frac{-\Delta}{2\sigma^2})$ is the Normal Probability distribution function $p_k(\omega) \sim \mathcal{N}(0, 1/\sigma^2)$, where $\Delta = \|x_i - x_j\|_F^2$. The features correspond to

Algorithm 1 Training algorithm for any activation and Kernel function

Input: Initialize $D_i \forall 1 \leq i \leq N'$

1. For the 1^{st} level, repeat until convergence.

$$X_1 \leftarrow \min_{X} \|Y - \Psi_1(\Phi_1(D_1 X))\|_F^2; \quad D_1 \leftarrow \min_{\Psi_1(D_1)} \|Y - \Psi_1(\Phi_1(D_1 X))\|_F^2$$

2. For the final level (ie., the effective Deep Kernel Dictionary Learning problem), repeat until convergence.

$$Y = \Psi_{N'}\left(Y_{N'}..\left(\Psi_2\left(Y_2(\Psi_1(Y_1))\right)\right)\right)$$

$$X'_N \leftarrow \min_{X'_N} \left\| Y - \Psi_1(D_1)\Phi_1\left(\Psi_2(D_2)\Phi_2\left(\ldots \Psi_{N'}\Phi_{N'}\left(D_{N'}X\right)\right)\right)\right\|_F^2 + \lambda \left\|X_{N'}\right\|_1$$

$$D'_N \leftarrow \min_{\Psi_{N'}(D_{N'})} \left\| Y - \Psi_1(D_1)\Phi_1\left(\Psi_2(D_2)\Phi_2\left(\ldots \Psi_{N'}\Phi_{N'}\left(D_{N'}X\right)\right)\right)\right\|_F^2 + \lambda \left\|X_{N'}\right\|_1$$

Algorithm 2 Testing algorithm for Linear Activation Function and any Kernel function

1. Collapse multiple levels of Dictionaries into a single effective one

$$D = \Psi_1(D_1)\Psi_2(D_2)...\Psi_{N'}(D_{N'})$$

2. Implement the Sparse coding for the test sample data(Y_{test})

$$X_{test} = \min_{X_{test}} ||Y_{test} - DX_{test}||_F^2 + \lambda ||X_{test}||_1$$

Algorithm 3 Testing Algorithm for non-linear activation

1. Generate features for the first level

$$X_1 \leftarrow \min_{X} \|Y_{test} - \Psi_1(\Phi_1(D_1 X))\|_F^2;$$

2. Generate features for the final level

$$X_{test} = \min_{X_{test}} \left\| Y_{test} - \Psi_1(D_1)\Phi_1\left(\Psi_2(D_2)\Phi_2\left(\ldots \Psi_{N'}\Phi_{N'}\left(D_{N'}X_{test}\right)\right)\right)\right\|_F^2 + \lambda \left\|X_{test}\right\|_1$$

making a matrix $\Omega \in \mathbb{R}^{k \times p}$ where $\omega_{i,j} \sim \mathcal{N}(0,1)$ and multiplying it with the signal(Y) and calculating the sin(.) and cos(.) of the each component. A relative comparison of advantages and disadvantages between data independent and data dependent Kernel maps is given by Yang et al. (2012b). The property of identification of transient events is from the Kernel function used and depends on the Euclidean distance between the Kernel approximation samples. If the distance is zero, it can better detect the transient events.

3.5. Determination of threshold limit

The T^2-Statistic, and Squared Prediction Error(SPE) defined in Chen et al. (2000) and Cui et al. (2012) by (18) and (19) are used for feature detection as represented in Fig. 1 and Algorithm 4. The T^2 metric is used to measure the minute variations in feature space. The SPE metric can recognize the small-scale fluctuations in the residual space and can be used for the detection of features based on the threshold limit. The two metrics provide information about Spatial changes that last for a short duration. The threshold limit is obtained from the Kernel Density Estimation (KDE) (Chen et al., 2000) technique by selecting the optimal bandwidth.

The T^2 Statistic at kth bin is defined as

$$T^2 = h_k^T \Theta^{-1} h_k \Theta = \frac{H^T H}{(n-1)} \tag{18}$$

where $H \in \mathbb{R}^{m \times n}$ the features extracted in encoding stage and Θ is the covariance matrix of the features.

The SPE metric is defined as

$$SPE = \|Y - \hat{Y}\|_2^2 \tag{19}$$

where Y and $\hat{Y}$ are the input sequence to the model and the predicted output of the model that is a solution to (11) given by $\hat{Y} = \hat{D}\hat{X}$ where $\hat{D}$ and $\hat{X}$ are effective Dictionaries and Sparse matrices of DKDL. When the T^2 and SPE statistic exceeds the threshold limit, a feature is detected.

3.6. Limitations and improvements

With the increase of the number of Neural Network Layers, the complexity in solving the effective optimization problem in Algorithm 1 increases and require explicit Kernel Approximation methods like the Nystroem method for incorporating kernels.

4. Kernel Density Estimation

The Kernel Density Estimation (KDE) (Chen et al., 2000) is a density estimation function of unknown numbers in the probability theory and belongs to the non-parametric test methods. Let 'Q' denote a Random Variable for the varying feature (here Wave Breaking events) 'q' at bin q, and the probability density function for the events as $P_Q(q)$ and defined as

$$P_Q(q) = \frac{1}{n\mathbf{W}} \sum_{i=1}^{n} \left(K(q - q_i)\right) \tag{20}$$

where q_i is a sample of a dataset, $\mathbf{W}$ is the bandwidth parameter, and $K(.)$ is the Kernel function that satisfies the following properties:

$$K(q) > 0; \sum_{0}^{\infty} K(q)dq = 1; \tag{21}$$

The proper selection of bandwidth (Turlach, 1993) is critical in the KDE for an accurate representation of the Probability Density Function

Algorithm 4 Deep Kernel Dictionary Learning for feature detection

Input: tolerence(ϵ), regularization parameter(λ), dropout, learning rate, Batch size, patch size, no. of. dense layers, no. of. cascaded Neural Networks, Kernel function, and sampling rate in Kernel function.
Output: feature detection of variable in both Spatial and Temporal domains
Data: training data, test data, and validation data set
Result: Spatial feature detection
if *(no. of. Networks*> 0*) and (all parameters*> 0*)* **then**

1. Calculate the effective Dictionary matrix and Sparse matrix using Algorithms 1, 2, and 3.
2. Calculate the T^2-metric and SPE metric using equations (18) and (19).
3. Determine the threshold(Q) values for the T^2-metric and SPE metric using Kernel Density Estimation.

if $(T^2 > Q)$ *and* $(SPE > Q)$ **then**

1. Detection of features.
2. Validate the detected features with data collected using in-situ instruments.
3. Calculate the Detection Rate and False Detection Rates.

else
No feature will be detected.
end
end
else

1. Change the NN's parameters to improve the detection efficiency.
2. Even if the feature is not detected, move to another data set.

.
end
end

(PDF) and the criteria defined in Bashtannyk and Hyndman (2001) is used for the derivation of optimal bandwidth ($\mathbf{W_{opt}}$) and $\mathbf{W} = \mathbf{W_{opt}}$.

$$W_{opt} = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{\frac{1}{5}} \approx 1.06\,\hat{\sigma}\,n^{-1/5} \tag{22}$$

where $\hat{\sigma}$ and n indicate the standard deviation of n data samples. The thresholding limit is calculated from KDE with a confidence level of 99%, representing a false alarm rate of 1% of T^2 and SPE. If the SPE statistic and T^2-Statistic are more than the threshold limit, the transient features (Wave Breaking events) will get detected. The detected feature has strong magnitudes of probability density at the location of the transient features. This makes us apply the proposed method to detect Gravity Wave Breaking event, where the energy of the Gravity Wave is transferred to the surrounding environment.

5. Experiments and results

The Wave Breaking event is considered a transient phenomenon as discussed in Liu et al. (1999b) and Farrell and Ioannou (1993). The breaking of Atmospheric Gravity Waves is observed from temperature data obtained by Microwave Limb Sounder (MLS) in Aura Satellite.

The Earth Observation Satellite (EOS) MLS-Aura (Schwartz et al., 2015) makes the measurements of atmospheric composition, temperature, humidity, and cloud ice through limb scanning. In limb sounding, the sounder scans the atmosphere at a tangent to the Earth, that is the Earth's limb (Waters et al., 2006). The MLS-Aura uses the limb scanning technique for data acquisition with a vertical spatial resolution of 1.5 km in the altitude range of 5–120 km and an horizontal effective resolution of 300 km (limb geometry).

The Rayleigh Lidar (Siva Kumar et al., 2003) system gives information about atmosphere temperature profiles (Sreekanth et al., 2020) with a spatial and temporal resolutions of 300 m and 4 min respectively. The Meteor Radar (MR) (Rao et al., 2014) provides information on wind profiles at the altitude between 70 km to 110 km with a spatial resolution of 2 km and a temporal resolution of 1 h. Whenever the Meteor Radar data is unavailable, the gap is filled with Horizontal Wind Model data (Drob et al., 2015). The cascaded network configuration used for the detection of Wave Breaking using DKDL is shown in Fig. 2.

The Wave Breaking features detected by using DKDL are validated using ground-based Rayleigh Lidar and Meteor Radar as shown in Table 10 and Fig. 5. The Wave Breaking patterns are also observed in the temperature perturbations obtained using Rayleigh Lidar at NARL, Gadanki (13.5° N, 79.2° E) and velocity perturbations obtained using Meteor Radar at Sri Venkateswara University, Tirupati (13.63° N, 79.4° E) on the reported days. The interpretation of Wave Breaking (Li et al., 2021; Wing et al., 2021) from Atmospheric density, Temperature, and Velocity perturbations and Wind profiles is made by identifying the presence of approaching large negative gradient (6) ($10\ \mathrm{K\,km^{-1}}$) of temperature, transfer of potential energy (4) to the surrounding environment, and a strong wind shear of $-40\ \mathrm{m\,s^{-1}\,km^{-1}}$ (indicate the presence of reduced convective and dynamic stability (3) and (4)). The steepening of the potential temperature contours (i.e., contour lines become vertical) indicates reduced convective stability. The Wave Breaking height detected by DKDL from temperature data without any prior information about Wind velocity profiles for various test data sets are consistent with the ground-based observations of Rayleigh Lidar and Meteor Radar observations and are discussed in detail in the following sections.

5.1. Dataset introduction

The MLS data consists of vertical temperature profiles varying from Atmospheric pressure of 316 hPa to 0.001 hPa, through which the corresponding altitude of the atmosphere is calculated.

The training data consists of all global information of Gravity Wave Breaking events in the form of temperature perturbations obtained using MLS in the year 2014 to 2019. The training data is of size 128×16000, where rows indicate the altitude varying from 25.1 km

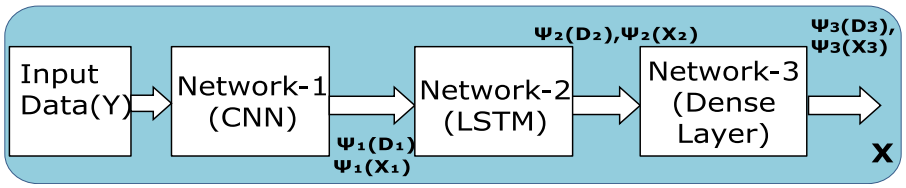

Fig. 2. Schematic diagram of Neural Networks used for the detection of Gravity Wave Breaking.

Table 1
Deep Learning Layers.

Network	Layers	Output shape	Parameters
CNN	Conv2D	(None, 128, 64, 32)	7520
	LeakyReLU	(None, 128, 64, 32)	0
	MaxPooling	(None, 64, 32, 32)	0
	Conv2D	(None, 64, 32, 64)	18496
	LeakyReLU	(None, 64, 32, 64)	0
	MaxPooling	(None, 32, 16, 64)	0
	Conv2D	(None, 32, 16, 128)	73856
	LeakyReLU	(None, 32, 16, 128)	0
	MaxPooling	(None, 16, 8, 128)	0
	Dense	(None, 16, 8, 128)	16512
	LeakyReLU	(None, 16, 8, 128)	0
LSTM	InputLayer	[(None, 128, 1)]	0
	lstm_1	(None, 128)	66560
	repeat_vector_1	(None, 128, 128)	0
	repeat_vector_2	(None, 127, 128)	0
	lstm_2	(None, 128, 128)	131584
	lstm_3	(None, 127, 128)	131584
	time_distributed_1	(None, 128, 1)	129
	time_distributed_2	(None, 127, 1)	129

to 80 km obtained after resampling the original data to have a spatial resolution of 0.3 km and the columns indicate the latitude variations (−82° S to +82 °N) of temperature profiles. The Wave Breaking events have been observed from temperature perturbations (A.2) and velocity perturbations obtained using Rayleigh Lidar and Meteor Radar respectively on 21 April 2014, 07 April 2014, 09 Jan 2015 and other data collected on different days are used as test data and for validating the proposed method. The test data is selected based on the satellite overpass ([13° N, 14° N], [79° E, 80° E]) over the Gadanki region. The proposed algorithm is applied individually to the selected days to determine the Wave Breaking altitude of Gravity Waves in the Atmosphere. The size of test data on each day is 128×256, where the rows determine the spatial resolution and columns indicate the latitude from 13° N to 14° N.

5.2. Training of Neural Network

The Neural Network (NN) training parameters are selected based on Grid Search Method and previous knowledge available during training. The prior information about the selection of Grid Search space for the hyper-parameter tuning is obtained from the previous literature relating to the present study. The learning rate is set between $[10^{-3}, 10^{-4}]$ and other hyper-parameters like batch size, dropout is selected appropriately. The Convolution Neural Networks are learned using linear activation function with an Input size of $(128, 64, 32)$, Kernel size of 16×16, α of 0.011 having an output size of 128×128. The detailed information about the layers implemented in each Neural Network is given in Table 1, and the specifications of the hyper-parameters used in the training of NN's is given in Table 2. The Kernel function used in the present analysis is the Radial Basis Function (RBF), and the regularization parameter (λ) used in Algorithm 1 is 0.01.

The other Kernel functions include the linear kernel function, polynomial Kernel Function, Gaussian Kernel Function, and RBF Kernel. The RBF Kernel is the most suitable algorithm for Kernelized Learning Algorithms and is more suitable for detection and classification problems. The RBF Kernel is popular as it is similar to the K-Nearest Neighborhood

Table 2
Hyper-parameters corresponding to top three regression models with least validation loss.

Hyper-parameters	DKDL	KCLSTM	CLSTM
Learning rate $(e-4)$	5.40	9.70	7.20
Droupout	0.80	0.80	0.80
Dense Layers	1	1	1
Batch size	3	3	3
Training loss	1.95E−4	1.60E−4	4.73E−4
Validation RMSE	1.82E−4	3.64E−5	1.97E−4

(KNN) Algorithm. Applying the KNN algorithm for regression and classification on RBF kernel is equivalent to applying our algorithm to the new infinite-dimensional data points. A hyper-plane in infinite dimensions acts as a strong classifier or regression after returning to original dimensions. The expansion of e^X in the RBF kernel gives a polynomial equation of infinite power, making it more suitable for sensitive information detection.

The neural network architectures of methodologies used for relative comparison regarding Detection Rate and False Detection Rate are given in Fig. 3

5.3. Evaluation and application of the proposed method

Generally, a large number of training epochs and loss values are used to stop network training. After reaching 10 epochs of training, the loss is stable as observed in Fig. 4. Even though, the training error has not reached the preset value, that requires a lot of time, and has a mild impact on the training of the network. So in the present experiment, the network will stop training after 20 epochs. The batch size used is 10, which means 10 samples are used for each iteration. The Kernel function used for Two indices called Detection Rate (DR) and False Detection Rate (FDR) is used as performance indicators. Here CLSTM and KCLSTM are used of comparison. CLSTM is widely used for feature detection. Although CLSTM and KCLSTM detect features well, they are unable to capture significant features. The original data cannot separate abnormal points. When CLSTM is combined with Kernel DL, can to detect all the features with a Detection Rate (DR) of 100%.

The Fig. 5 represents the SPE metric and T^2 metric for test data on 21 April 2014. Simultaneous observation of SPE and T^2 metrics above the threshold limit obtained using KDE. When both the metrics are above the threshold limit, a Wave Breaking event is detected. The Wave Breaking altitude is detected at 75 km. The above-mentioned data is only for an explanation of the method. The proposed method is tested with satellite data on all days between 2007 to 2018, where ground-based measurements are available.

5.4. Comparison with deep learning approaches

We first analyze the effect of increasing the number of layers on the proposed method in terms of detection accuracy. Upon working with all combinations of network layers, we find the neural network layers mentioned in Table 1 are suitable for the present work. The Deep Dictionary Learning offers improvement (Tariyal et al., 2016) in the detection accuracy than the shallow dictionary learning technique. Tables 3 and 4 give the performance comparison of the proposed method with the number of kernel approximation components in the sense of DR and training time. Generally, the increase in the number

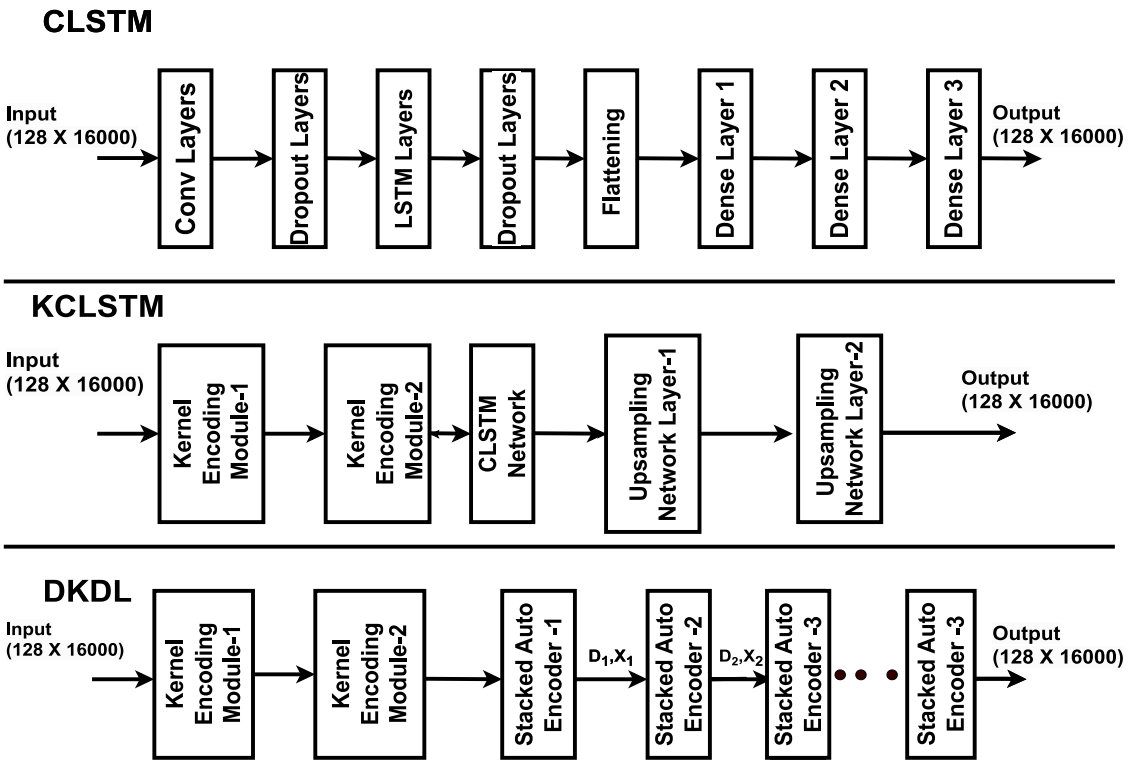

Fig. 3. Neural Network architecture of CLSM, KCLSTM, and DKDL.

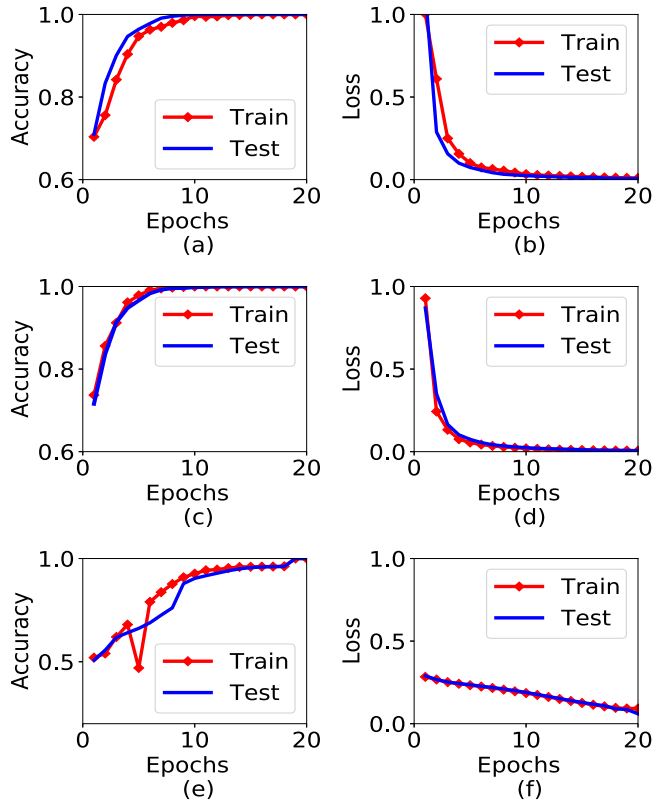

Fig. 4. Accuracy and Loss curves. (a), (c), (e) and (b), (d), (f) represent the Accuracy and Loss curves for Deep Kernel Dictionary Learning, Kernel Convolutional LSTM, and Conventional LSTM for the data on 21 April 2014.

of Kernel approximation components will improve the efficiency of feature detection. In the case when the number of Kernel approximation components is 128, it has the highest DR of 0.95 with a training time of 12.57 s. So in the present work, the number of Kernel approximation components is 128 and is the same in the assessment of DR and FDR.

Table 3
Comparison of Detection Rate with varying kernel Approximation components.

No. of. components	KCLSTM	DKDL
64	0.50	0.40
128	0.45	**0.95**
256	0.50	0.40
512	0.50	0.75

Table 4
Comparison of training time (seconds) with varying Kernel Approximation components.

No. of. components	KCLSTM	DKDL
64	8.14	11.19
128	8.44	**12.57**
256	11.47	13.73
512	10.08	15.44

The superiority of the proposed method with existing methods is explained using Detection Rate and False Detection Rate. Tables 5 and 6 present the DR and Threshold limit values of T^2 and SPE metric statistics of CLSTM, KCLSTM, and DKDL of three different data sets, respectively. In Table 3, the DR of T^2 and SPE statistics of CLSTM, and KCLSTM are low, which indicates that they cannot extract all the features in the dataset. The results show that the DKDL shows the best feature detection among all the methods and show that the addition of kernels in DDL improves the performance of feature learning and Fault Detection Rate (FDR). Table 6 represents the threshold limits of T^2 and SPE metrics obtained using KDE for the detection of Wave Breaking for the three different data sets. A higher value of T^2 threshold limit and a small value of SPE threshold limit indicate that DKDL has superior performance compared to the other two methods.

Table 7 shows the comparative results of the three methods. It should be noted that the network parameters used for training for the three methods remain unchanged and the overall average Fault Detection Rate is less than 6%. The CLSTM method has the largest training error and its DR based on SPE is the lowest in comparison to

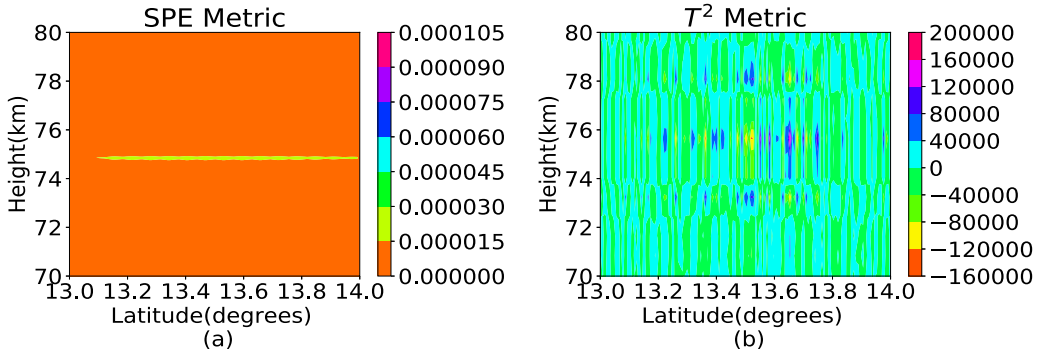

Fig. 5. Detection of Wave Breaking events. (a) SPE static and (b) T^2-static in the DKDL process of feature detection for the test data on 21 April 2014.

Table 5
Detection Rate (DR) of Gravity Wave Breaking based on T^2 metric (SPE metric)

Data	CLSTM	KCLSTM	DKDL
07 April, 2014	0.41 (0.47)	0.47 (0.76)	0.70 (0.95)
21 April, 2014	0.50 (0.95)	0.25 (0.50)	0.60 (1.00)
09 January, 2015	0.21 (0.89)	0.61 (0.80)	0.80 (0.90)
Average	0.37 (0.77)	0.44 (0.68)	0.70 (0.95)

Table 6
Threshold limits T^2 (SPE metric) for the detection of Gravity Wave Breaking.

Data	CLSTM	KCLSTM	DKDL
07 April, 2014	8E3 (12E-5)	8E3 (1E-5)	8E3 (15E-5)
21 April, 2014	8E3 (10E-5)	8E3 (1E-5)	20E3 (15E-5)
09 January, 2015	15E3 (25E-5)	20E3 (0.3E−5)	15E3 (16E-5)
Average	10.3E3 (15.6E−5)	12E3 (0.76E−5)	14.3E3 (15.33E−5)

Table 7
Training loss and Average Detection Rate of the three different methods.

Method	Loss	DR (T^2)	DR (SPE)
CLSTM	2%	40%	60%
KCLSTM	0.4%	40%	65%
DKDL	**0.027%**	**70%**	**90%**

Table 8
Fault Detection Rate (FDR) of Gravity Wave Breaking based on T^2 metric (SPE metric)

Data	CLSTM	KCLSTM	DKDL
07 April 2014	0.03 (0.04)	0.02 (0.05)	0.04 (0.04)
21 April, 2014	0.05 (0.02)	0.07 (0.05)	0.04 (0.04)
09 January 2015	0.05 (0.04)	0.04 (0.04)	0.04 (0.08)
Average	0.043 (0.033)	0.043 (0.046)	0.04 (0.053)

KCLSTM, and DKDL. Instead of having improvement in the DR based on T^2 for CLSTM and KCLSTM are the same as T^2 metric is calculated from encoder output and is not influenced by kernel the function. The results of CLSTM and KCLSTM are very poor compared to DKDL.

The performance of the proposed method against the false detection on average is less than 6% as observed in Table 8. From Tables 5 and 8, it is clear that the DR and FDR of T^2 and SPE static show superior performance than the other methods. These testing results demonstrate that the superiority of the proposed method in feature detection.

To demonstrate the efficiency of the proposed method, the method is tested with data collected on different days. The DR, FDR, of feature detection for various test data sets as shown in Table 9, is slightly higher than the average values mentioned in Table 5 and 8 respectively.

5.5. Validation of the proposed method

From Fig. 6(a), we obtain the information about the dominant vertical wavenumber and found as $1.046\ \mathrm{km}^{-1}$. There is a large negative temperature gradient of $13\ \mathrm{K\,km}^{-1}$ and a strong wind shear

Table 9
Detection of Wave Breaking region and their DR and FDR based on T^2 (SPE) metrics using Deep Kernel Dictionary Learning.

Data	Loss	DR	FDR	Threshold limit
07 February, 2007	0.023%	1 (0.97)	0.01 (0.01)	10E3 (14E-5)
22 November, 2017	0.043%	0.99 (0.99)	0.02 (0.01)	14E3 (15E-5)
08 December 2017	0.025%	0.83 (0.96)	0.15 (0.04)	16E3 (12E-5)
05 January, 2018	0.034%	0.95 (0.96)	0.05 (0.02)	15E3 (8E-5)

Table 10
Characteristics of Gravity Wave Breaking events observed using Rayleigh Lidar, Meteor Radar, and Horizontal Wind Model. The value of N^2 for all the data is less than 0.

Data	H (km)	R_i	$-dT/dz$ (K/km^{-1})	$log[E_p]$ (J/kg)
07 February, 2007	76	−1.80	13	3.00
21 April, 2014	75	−1.00	12	2.40
07 April, 2014	73	−1.50	14	2.65
09 January, 2015	72	−1.20	12	2.40
22 November, 2017	77	−2.00	14	2.50
08 December, 2017	75	−0.50	12	2.80
05 January, 2018	78	−0.20	11	3.30

Table 11
Confusion matrix for Wave Breaking event Detection.

	Actual positive	Actual negative
Predicted positive	0.97	0.02 (Type-II error)
Predicted negative	0.03 (Type-I error)	0.98

of $-40\ \mathrm{m\,s^{-1}\,km^{-1}}$ is observed in Fig. 6(b), 6(d), and 6(c) respectively. A huge transfer of Potential Energy(E_p) at an altitude of 75 km is observed in Fig. 6(e) and is a good indicator of Wave Breaking . The Wave Breaking creates an unstable atmosphere at the breaking altitude, and the instability can be verified by using (3), (5), and (6).

The inclusion of kernel function in the DDL method leading to DKDL improves the detection rate as compared to the conventional method as shown in Table 8. The feature detection algorithm, when applied to temperature perturbations obtained from Satellite data, is useful for the detection of wave breaking heights that are useful to study the middle atmosphere dynamics.

The overall performance of the proposed algorithm is explained in terms of the confusion matrix in Table 11. We consider the occurrence of Wave Breaking events as a positive condition, and no occurrence of Wave Breaking as a negative condition.

5.6. Applicability of the proposed method

The classification accuracy of the non-linear optimization problems is higher compared to their linear counterparts. Deep Dictionary Learning has the advantage of learning features adaptively at each layer and the formed effective Dictionary can detect the features very effectively. The proposed method is advantageous over the existing feature extraction techniques like CLSTM, KCLSTM, and Auto-Encoders if:

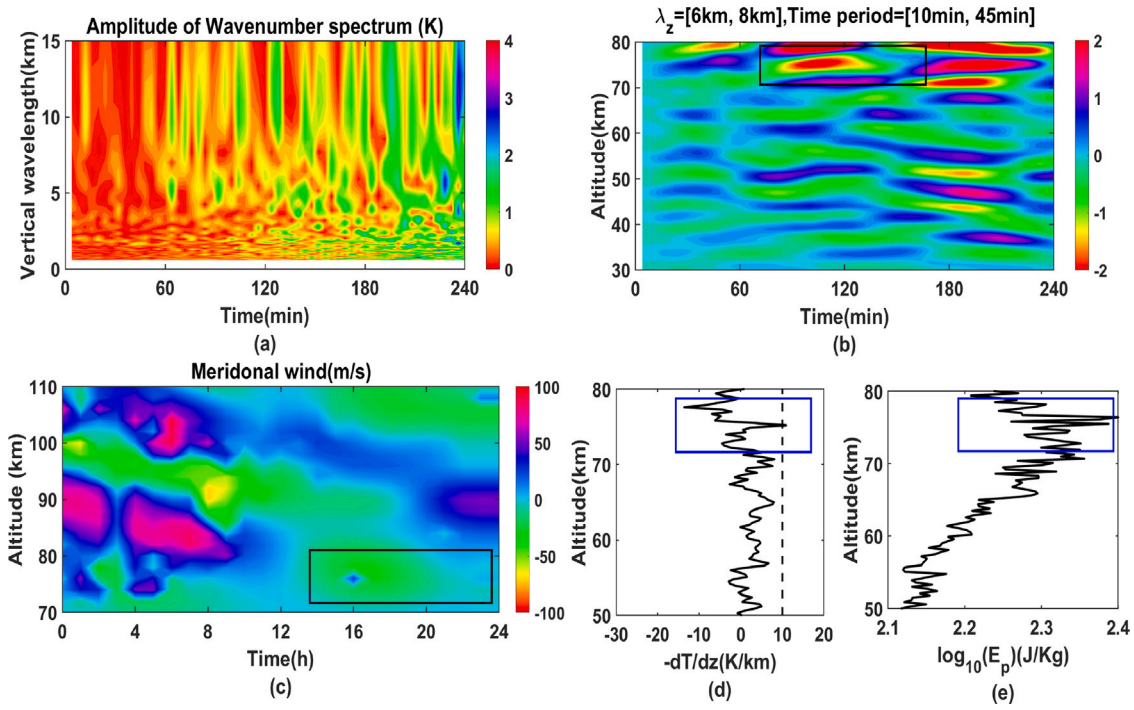

Fig. 6. Characteristics of Gravity Wave Breaking Events. (a) Dominant vertical wavenumber, (b) Wave progression, (c) Meridional Wind velocities obtained using Meteor Radar, (d) Negative gradient of temperature concerning altitude, and (e) Potential Energy (E_p) for the test data collected on 21 April 2014. The rectangles inside each figure represent the Wave Breaking region.

1. The learning problem is non-linear (as most of the optimization problems describing the natural phenomena are non-linear), and the requirement of high classification accuracy. The Deep Dictionary Learning technique improves Classification accuracy in solving optimization problems.
2. The requirement for detecting transient events is of utmost importance. The RBF kernel used in the proposed method can better detect sudden events when compared to other kernel functions.

5.7. Implications of proposed method

The DKDL-based method can identify spatial and temporal multivariate patterns that minimize the error in the detection of Wave Breaking events when compared to conventional methods. Integrating Deep Learning to detect Gravity Wave Breaking events will open new venues for understanding turbulence characteristics that define the dynamics of the Middle Atmosphere. The proposed methodology will open the venue to Atmosphere Physicists to explore the causes of instabilities in the Middle Atmosphere. In general, the proposed method can be applied for the detection of transient events like Wave Breaking, turbulence detection, and lightning event detection. Also, the proposed method can be applied to any data that has both spatial and temporal measurements for the identification of patterns in the data about a particular behavior of an atmospheric parameter or process.

As this is a new methodology, there is a need to address the issues associated (big data, more layers in the model) in the long-term period. Future research should explore the further integration of other AI solutions thus moving these technology fusions forward in the field of Gravity Waves.

6. Conclusion

In this work, we proposed a methodology for the detection of Gravity Wave Breaking events from Atmospheric Temperature profiles through the incorporation of kernels in the Deep Dictionary Learning technique, called "DKDL". The Kernel approximations are implemented using Fourier Feature generation. The features from one Neural Network layer are used as the input to the second Neural Network Layer giving better results than the cascaded combination. We have shown that the proposed method is the same as the denoising Auto-Encoder with inputs being Kernel function transformed for a linear activation function. The potential of the proposed method is demonstrated with different data sets in terms of Detection Rate and False Detection Rate of transient events with application to the detection of Atmospheric Gravity Wave Breaking events. The combination of hyper-parameters, selected based on the Grid-Search method gives a good Detection Rate. The inclusion of Kernel has improved the Detection Rate by 33% and 26% in comparison with CLSTM and KCLSTM respectively.

Inclusion of discriminative approaches like Fisher discrimination, and locality-constrained sparse coding in the Deep Kernel Dictionary Learning algorithm, and inclusion of supervised learning paradigm in Deep Kernel Dictionary Learning methodology may further improve the Detection Rate and reduce the false alarm rate for studying the Gravity Wave Breaking events more accurately.

CRediT authorship contribution statement

Varanasi Satya Sreekanth: Writing – review & editing, Conceptualization, Methodology, Coding, Visualization, Formal analysis. **Karnam Raghunath:** Writing – review & editing, Conceptualization, Supervision. **Deepak Mishra:** Writing – review & editing, Conceptualization, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The Temperature perturbations used for Training and Testing for Wave Breaking feature detection is available at https://daac.gsfc.nasa.gov. The Rayleigh Lidar and Meteor Radar data used in the present work is available at https://www.narl.gov.in. The data of the Horizontal Wind Model is available at https://www.alpendac.eu.

Acknowledgments

We are also thankful to Dr. M. Venkatratnam, ARTG, NARL, and Prof. Dr. S. Vijaya Bhaskara Rao, Department of Physics, Sri Venkateswara University, Tirupati, for helping us in the extraction of Gravity Wave Parameters and providing us the Meteor Radar data.

Funding

This research work was funded by the National Atmospheric Research Laboratory, Department of Space, Gadanki, Andhra Pradesh, India.

Computer code availability

- Name of code: Deep Kernel Dictionary Learning for Detection of Wave Breaking events.
- Hardware required: Code was tested on a modern PC with 8 GB RAM
- Software required: Matlab, Python 3.6+ with the numpy, scikitlearn, matplotlib, and scipy packages; Tensorflow, Keras, and Pytorch
- Program language: Python, and MATLAB.
- Details on how to access the source code: https://github.com/VARANASISATYASREEKANTH/Deep-Kernel-Dictionary-Learning .

References

Achatz, U., 2007. Gravity-wave breaking: Linear and primary nonlinear dynamics. Adv. Space Res. 40 (6), 719–733.

Agethen, S., Hsu, W.H., 2020. Deep multi-Kernel convolutional LSTM networks and an attention-based mechanism for videos. IEEE Trans. Multimed. 22 (3), 819–829. http://dx.doi.org/10.1109/TMM.2019.2932564.

Aharon, M., Elad, M., Bruckstein, A., 2006. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Signal Process. 54 (11), 4311–4322.

Bashtannyk, D.M., Hyndman, R.J., 2001. Bandwidth selection for kernel conditional density estimation. Comput. Statist. Data Anal. 36 (3), 279–298.

Booker, J.R., Bretherton, F.P., 1967. The critical layer for internal gravity waves in a shear flow. J. Fluid Mech. 27 (3), 513–539.

Cai, S., Zuo, W., Zhang, L., Feng, X., Wang, P., 2014. Support vector guided dictionary learning. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014. Springer International Publishing, Cham, pp. 624–639.

Candes, E.J., Donoho, D.L., 2002. Recovering edges in ill-posed inverse problems: Optimality of curvelet frames. Ann. Statist. 30 (3), 784–842.

Chen, Q., Wynne, R., Goulding, P., Sandoz, D., 2000. The application of principal component analysis and kernel density estimation to enhance process monitoring. Control Eng. Pract. 8 (5), 531–543.

Courtney, L., Sreenivas, R., 2019. Using deep convolutional LSTM networks for learning spatiotemporal features. In: Asian Conference on Pattern Recognition. Springer, pp. 307–320.

Cui, Y., Yang, J., Zhang, X., 2012. New CFAR target detector for SAR images based on kernel density estimation and mean square error distance. J. Syst. Eng. Electron. 23 (1), 40–46.

Do, M.N., Vetterli, M., 2002. Contourlets: A directional multiresolution image representation. In: Proceedings. International Conference on Image Processing, Vol. 1. IEEE, p. I.

Drineas, P., Mahoney, M.W., Cristianini, N., 2005. On the Nyström method for approximating a gram matrix for improved Kernel-based learning. J. Mach. Learn. Res. 6 (12).

Drob, D.P., Emmert, J.T., Meriwether, J.W., Makela, J.J., Doornbos, E., Conde, M., Hernandez, G., Noto, J., Zawdie, K.A., McDonald, S.E., Huba, J.D., Klenzing, J.H., 2015. An update to the horizontal wind model (HWM): The quiet time thermosphere. Earth Space Sci. 2 (7), 301–319.

Engan, K., Aase, S., Hakon Husoy, J., 1999. Method of optimal directions for frame design. In: 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258), Vol. 5. pp. 2443–2446. http://dx.doi.org/10.1109/ICASSP.1999.760624.

Farrell, B.F., Ioannou, P.J., 1993. Transient development of perturbations in stratified shear flow. J. Atmos. Sci. 50 (14), 2201–2214.

Franke, P.M., Collins, R.L., 2003. Evidence of gravity wave breaking in lidar data from the mesopause region. Geophys. Res. Lett. 30 (4).

Fritts, D.C., 1984. Gravity wave saturation in the middle atmosphere: A review of theory and observations. Rev. Geophys. 22 (3), 275–308.

Fritts, D.C., Alexander, M.J., 2003. Gravity wave dynamics and effects in the middle atmosphere. Rev. Geophys. 41 (1).

Fritts, D.C., Bizon, C., Werne, J.A., Meyer, C.K., 2003. Layering accompanying turbulence generation due to shear instability and gravity-wave breaking. J. Geophys. Res.: Atmos. 108 (D8).

Garcia, R.R., Solomon, S., 1985. The effect of breaking gravity waves on the dynamics and chemical composition of the mesosphere and lower thermosphere. J. Geophys. Res.: Atmos. 90 (D2), 3850–3868.

Hines, C., 1971. Generalizations of the Richardson criterion for the onset of atmospheric turbulence. Q. J. R. Meteorol. Soc. 97 (414), 429–439.

Hines, C.O., 1988. Generation of turbulence by atmospheric gravity waves. J. Atmos. Sci. 45 (7), 1269–1278.

Holton, J.R., 1973. An introduction to dynamic meteorology. Amer. J. Phys. 41 (5), 752–754.

Holton, J.R., 1983. The influence of gravity wave breaking on the general circulation of the middle atmosphere. J. Atmos. Sci. 40 (10), 2497–2507.

Kumar, N., Soni, K., Agarwal, R., 2021. Prediction of temporal atmospheric boundary layer height using long short-term memory network. Tellus A: Dyn. Meteorol. Oceanogr. 73 (1), 1–14.

Li, J., Collins, R., Lu, X., Williams, B., 2021. Lidar observations of instability and estimates of vertical eddy diffusivity induced by gravity wave breaking in the arctic mesosphere. J. Geophys. Res.: Atmos. 126 (4), e2020JD033450.

Lindzen, R.S., 1973. Wave-mean flow interactions in the upper atmosphere. Bound.-Lay. Meteorol. 4 (1), 327–343.

Lindzen, R.S., 1981. Turbulence and stress owing to gravity wave and tidal breakdown. J. Geophys. Res.: Oceans 86 (C10), 9707–9714.

Lindzen, R., 1984. Gravity waves in the mesosphere. Dyn. Middle Atmos. 3–18.

Lindzen, R.S., 1985. Multiple gravity-wave breaking levels. J. Atmos. Sci. 42 (3), 301–305.

Lindzen, R.S., Holton, J.R., 1968. A theory of the Quasi-Biennial oscillation. J. Atmos. Sci. 25 (6), 1095–1107.

Liu, H., Hays, P., Roble, R., 1999a. A numerical study of gravity wave breaking and impacts on turbulence and mean state. J. Atmos. Sci. 56 (13), 2152–2177.

Liu, H.L., Hays, P.B., Roble, R.G., 1999b. A numerical study of gravity wave breaking and impacts on turbulence and mean state. J. Atmos. Sci. 56 (13), 2152–2177.

Ma, J., Cheng, J.C., Lin, C., Tan, Y., Zhang, J., 2019. Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques. Atmos. Environ. 214, 116885.

Mairal, J., Ponce, J., Sapiro, G., Zisserman, A., Bach, F., 2009. Supervised dictionary learning. In: Koller, D., Schuurmans, D., Bengio, Y., Bottou, L. (Eds.), Advances in Neural Information Processing Systems, Vol. 21. Curran Associates, Inc..

Mallat, S., 1999. A Wavelet Tour of Signal Processing. Elsevier.

Malmgren-Hansen, D., Laparra, V., Nielsen, A.A., Camps-Valls, G., 2019. Statistical retrieval of atmospheric profiles with deep convolutional neural networks. ISPRS J. Photogramm. Remote Sens. 158, 231–240.

Matsuoka, D., Watanabe, S., Sato, K., Kawazoe, S., Yu, W., Easterbrook, S., 2020. Application of deep learning to estimate atmospheric gravity wave parameters in reanalysis data sets. Geophys. Res. Lett. 47 (19), e2020GL089436.

Nguyen, H.V., Patel, V.M., Nasrabadi, N.M., Chellappa, R., 2012. Kernel dictionary learning. In: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, pp. 2021–2024. http://dx.doi.org/10.1109/ICASSP.2012.6288305.

Qiao, W., Tian, W., Tian, Y., Yang, Q., Wang, Y., Zhang, J., 2019. The forecasting of PM2. 5 using a hybrid model based on wavelet transform and an improved deep learning algorithm. IEEE Access 7, 142814–142825.

Rahimi, A., Recht, B., 2008. Random features for large-scale kernel machines. In: Platt, J., Koller, D., Singer, Y., Roweis, S. (Eds.), Advances in Neural Information Processing Systems, Vol. 20. Curran Associates, Inc..

Rao, S.V.B., Eswaraiah, S., Venkat Ratnam, M., Kosalendra, E., Kishore Kumar, K., Sathish Kumar, S., Patil, P., Gurubaran, S., 2014. Advanced meteor radar installed at Tirupati: System details and comparison with different radars. J. Geophys. Res.: Atmos. 119 (21), 11–893.

Salakhutdinov, R., Hinton, G., 2009. Deep boltzmann machines. In: Artificial Intelligence and Statistics. PMLR, pp. 448–455.

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., Brenning, A., 2019. Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. Ecol. Model. 406, 109–120.

Schwartz, M., Livesey, N., Read, W., 2015. MLS/Aura Level 2 Temperature V004. Goddard Earth Sciences Data and Information Services Center GES DISC, Greenbelt, MD, USA.

Siva Kumar, V., Rao, P., Krishnaiah, M., 2003. Lidar measurements of stratosphere-mesosphere thermal structure at a low latitude: Comparison with satellite data and models. J. Geophys. Res.: Atmos. 108 (D11).

Sreekanth, V.S., Raghunath, K., Mishra, D., 2020. Dictionary learning technique and penalized maximum likelihood for extending measurement range of a Rayleigh lidar. J. Appl. Rem. Sens. 14 (3), 034529.

Tang, H., Liu, H., Xiao, W., Sebe, N., 2021. When dictionary learning meets deep learning: Deep dictionary learning and coding network for image recognition with limited data. IEEE Trans. Neural Netw. Learn. Syst. 32 (5), 2129–2141. http://dx.doi.org/10.1109/TNNLS.2020.2997289.

Tariyal, S., Majumdar, A., Singh, R., Vatsa, M., 2016. Deep dictionary learning. IEEE Access 4, 10096–10109.

Tsuda, T., 2014. Characteristics of atmospheric gravity waves observed using the MU (middle and upper atmosphere) radar and GPS (global positioning system) radio occultation. Proc. Jpn. Acad. Ser. B 90 (1), 12–27.

Turlach, B.A., 1993. Bandwidth selection in kernel density estimation: A review. In: CORE and Institut de Statistique. Citeseer.

Vorontsov, A.M., Vorontsov, M.A., Filimonov, G.A., Polnau, E., 2020. Atmospheric turbulence study with deep machine learning of intensity scintillation patterns. Appl. Sci. 10 (22), 8136.

Waters, J.W., Froidevaux, L., Harwood, R.S., Jarnot, R.F., Pickett, H.M., Read, W.G., Siegel, P.H., Cofield, R.E., Filipiak, M.J., Flower, D.A., et al., 2006. The earth observing system microwave limb sounder (EOS MLS) on the Aura satellite. IEEE Trans. Geosci. Remote Sens. 44 (5), 1075–1092.

Williams, C., Seeger, M., 2001. Using the Nyström method to speed up kernel machines. In: Proceedings of the 14th Annual Conference on Neural Information Processing Systems. (CONF), pp. 682–688.

Wing, R., Martic, M., Triplett, C., Hauchecorne, A., Porteneuve, J., Keckhut, P., Courcoux, Y., Yung, L., Retailleau, P., Cocuron, D., 2021. Gravity wave breaking associated with mesospheric inversion layers as measured by the ship-Borne BEM Monge Lidar and ICON-MIGHTI. Atmosphere 12 (11), 1386.

Yang, T., Li, Y.F., Mahdavi, M., Jin, R., Zhou, Z.H., 2012a. Nyström method vs random fourier features: A theoretical and empirical comparison. Adv. Neural Inf. Process. Syst. 25, 476–484.

Yang, T., Li, Y.f., Mahdavi, M., Jin, R., Zhou, Z.H., 2012b. Nyström method vs random Fourier features: A theoretical and empirical comparison. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems, Vol. 25. Curran Associates, Inc..

Zhang, Q., Li, B., 2010. Discriminative K-SVD for dictionary learning in face recognition. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 2691–2698. http://dx.doi.org/10.1109/CVPR.2010.5539989.