Dictionary Learning Approach To Lidar Sensing Of The Atmosphere

A thesis submitted
in partial fulfillment for the award of the degree of

Doctor of Philosophy

by

Varanasi Satya Sreekanth

Department of Avionics
Indian Institute of Space Science and Technology
Thiruvananthapuram, India

December 2023

Certificate

This is to certify that the thesis titled ***Dictionary Learning Approach To Lidar Sensing Of The Atmosphere*** submitted by **Varanasi Satya Sreekanth**, to the Indian Institute of Space Science and Technology, Thiruvananthapuram, in partial fulfillment for the award of the degree of **Doctor of Philosophy** is a bonafide record of the original work carried out by him/her under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Deepak Mishra
(Research Supervisor)
Professor
Department Of Avionics
IIST

Dr. K Raghunath
(Research Supervisor)
Scientist/Engineer-SG
NARL

Dr. N. Selvaganesan
Professor and Head,
Department of Avionics
IIST

Place: Thiruvananthapuram
Date: December 2023

Declaration

I declare that this thesis titled ***Dictionary Learning Approach To Lidar Sensing Of The Atmosphere*** submitted in partial fulfillment for the award of the degree of **Doctor of Philosophy** is a record of the original work carried out by me under the supervision of **Dr. K. Raghunath**, and **Dr. Deepak Mishra**, and has not formed the basis for the award of any degree, diploma, associateship, fellowship, or other titles in this or any other Institution or University of higher learning. In keeping with the ethical practice in reporting scientific information, due acknowledgments have been made wherever the findings of others have been cited.

Place: Thiruvananthapuram
Date: December 2023

Varanasi Satya Sreekanth
(SC18D018)

To my parents Uma Devi and Sree Rama Krishna, To my wife Suchitra
To my Supervisors Deepak Mishra, K Raghunath

Acknowledgements

I acknowledge that this thesis represents not only my work at the keyboard, it is a milestone in more than 4 years of work at the National Atmospheric Research Laboratory(NARL) and Indian Institute of Space Science and Technology (IIST) and specifically within the Rayleigh Lidar Building and Department of Avionics respectively. As I am with Microwave and Communication Engineering specialization at the graduate level, the suggested course work at IIST helped me more in carrying out the research work in a smooth fashion.

First and foremost I wish to thank my advisors, Scientist/Engineer-SG, **K Raghunath**, Group Head, Atmospheric Structure and Dynamics Group, NARL and Professor, **Deepak Mishra**, Department of Avionics, IIST, for their supportiveness since the days I began working on the origins of Dictionary Learning applied to lidar remote sensing. Ever since Deepak Mishra and Raghunath have supported me not only by providing a research assistantship over almost five years but also academically and emotionally through the rough road to finish this thesis. I particularly thank **Deepak Mishra** for helping me to come up with the thesis topic, for his suggestions during the preparation of the manuscript, his quick decision, and his guidance to me over almost five years.

My Doctoral committee at IIST and academic committee at NARL guided me through all these years. I am thankful to A. K. Patra, T. V. C. Sarma, M. Venkatratnam, S. Sridharan at NARL and B. S. Manoj, Rajeevan, Nidamanuri Rama Rao and Sumitra S. Nair at IIST and Rakesh Kumar Singh at IIT-BHU for their valuable suggestions during the annual review meeting. I am also thankful to T. Rajendra Prasad, Urvashi Jinwal, G. Sandeep, and the past technical assistants of Lidar Operations namely Nizamuddin, Sivakumar, Nagarjuna, and P. Venkatramana from whom I learned the Lidar Operations.

Varanasi Satya Sreekanth

Abstract

Extending temperature measurement range, and sustaining the system performance is an important requirement for Rayleigh lidar remote sensing of the atmosphere. Power aperture product enhancement is one way to achieve the desired measurement range. The second way is the Dictionary Learning approach. This work aims to retrieve atmospheric temperature profiles using denoised photon counts using the penalized maximum likelihood Dictionary Learning technique. The proposed algorithm for temperature retrieval provides an acceptable level of temperatures at higher altitudes and reduces the **Standard Error(SE)** in temperatures. In the proposed method, uncertainties in temperature profiles are estimated using Monte Carlo Simulations. The observational data for evaluating the proposed algorithm was acquired with the help of a Rayleigh Lidar System at the National Atmospheric Research Laboratory, Gadanki, India. It has been measured that the SE of the temperatures retrieved get reduced by **5K** at the altitude of 84 km and the measurement range has been extended by 6 km.

The Atmospheric Gravity Waves are studied based on temperature perturbations obtained from the measurements of Rayleigh Lidar. Background sources must be removed when studying gravity waves. In the analysis of Gravity Waves from temperature perturbations derived from Rayleigh Lidar, the presence of wind in the horizontal plane causes the frequency spectrum to shift and power spectral density get change than the actual. A methodology named Multi-Resolution Dictionary Learning combined with Instantaneous Frequency Estimation is proposed for removing background wind effects in both the temporal and spatial domain for analyzing the behavior of Atmospheric Gravity Waves accurately. An in-depth study on the seasonal changes of frequency shift due to background wind is discussed.

The temperature perturbations without the presence of background wind are used to detect Gravity Wave breaking. Middle atmospheric circulation is heavily influenced by the Atmospheric Gravity Waves. Gravity Waves get trapped at the altitude where their horizontal phase speed is equal to the horizontal background wind speed, causing a critical layer to form. We focus on the **mesosphere(50-90km)**, where Gravity Waves get attenuated due to non-linearity. To enhance the study of Global Circulation, Kernel functions are included for the first time in Deep Dictionary Learning techniques. As a result of the introduction

of **Kernel functions in Deep Dictionary Learning,** it is possible to detect sudden transient events, like the breaking of a Gravity Wave. It discusses how Deep Kernel Dictionary Learning techniques can detect Gravity wave-breaking events and the false detection rate of those events.

Contents

List of Figures

List of Tables

List of Algorithms

Abbreviations

ADC	Analog to Digital Converter
AGW	Atmospheric Gravity Wave
ADMM	Alternating Direction Method of Multipliers
AE	Auto-Encoder
BN	Batch Normalization
BOMP	Batch Orthogonal Matching Pursuit
CNN	Convolutional Neural Network
DL	Dictionary Learning
DDL	Deep Dictionary Learning
DKDL	Deep Kernel Dictionary Learning
DKDL	Deep Kernel Dictionary Learning
DR	Detection Rate
DWT	Discrete Wavelet Transform
EOS	Earth Observation Satellite
EMD	Emperical Mode Decomposition
FDR	False Detection Rate
GPS	Global Positioning System
GS	Grid Search
HWM	Horizontal Wind Model
HLR	High Level Rayleigh
IQR	Inter Quartile Range
IF	Instantaneous Frequency
IFE	Instantaneous Frequency Estimation
IST	Indian Standard Time
KKT	Karush Kuhn Tucker
KSVD	k-means Singular Value Decomposition
KLD	Kullback Leibler Divergence

KDE	Kernel Density Estimation
KNN	K-Nearest Neighbour
KCLSTM	kernel Convolution LSTM
KF	Kalman Filtering
KDL	Kernel Dictionary Learning
KKSVD	Kernel K-Singular Value Decomposition
LLR	Low Level Rayleigh
LSC	Laplacian Sparse Coding
LASSO	Least Absolute Shrinkage and Selection Operator
LS-SVM	Least Squares Support Vector Machine
LSTM	Long Short Term Memory
MCS	Multi Channel Scaler
MLS	Microwave Limb Sounder
MR	Meteor Radar
MC	Mutual Coherence
MOD	Method of Optimal Directions
MRA	Multi Resolution Analysis
MSE	Mean Square Error
MMSE	Minimum Mean Square Error
MA	Moving Averages
MLE	Maximum Likelihood Estimate
MODWT	Maximum Overlap Discrete Wavelet Transform
MRDL	Multi-Resolution Dictionary Learning
NARL	National Atmospheric Research Laboratory
NSP	Null Space Property
OMP	Orthogonal Matching Pursuit
OEM	Optimal Estimation Method
PMT	Photo Multiplier Tube
PE	Peak Error
PCL	Purple Crow Lidar
PC	Photon Counting
RLS	Rayleigh Lidar System
RS	Radiosonde
RIP	Restricted Isometric Property
RSP	Range Space Property

ReLU	Rectified Linear Unit
RMSE	Root Mean Square Error
RE	Reconstruction Error
RBF	Radial Basis Function
RKHS	Reproducing Kernel Hilbert Space
SNR	Signal to Noise Ratio
SABER	Sounding of the Atmosphere using Broadband Emission Radiometry
SC	Sparse Coding
SPE	Squared Prediction Error
SE	Standard Error
SKLD	Skewed Kullback Leibler Divergence
STFT	Short Time Fourier Transform
SSE	Sum of Squares due to error
TFD	Time Frequency Distribution
VMD	Variational Mode Decomposition
WT	Wavelet Transform
WPD	Wavelet Packet Decomposition
WB	Wave Breaking

Nomenclature

$\mathbf{Y}$:Observed data matrix	ρ:density of Atmosphere
$\mathbf{D}$:Dictionary matrix	C: Lidar system constant($= 2.5e18$)
$\mathbf{X}$:Sparse Coefficient matrix	n:Photon count data in range bin
$\mathbf{V}'$: Noise matrix	B: Background Photon count
N: Number of range bins	X': arbitrary variable
i,j: variables	M_{air}: Mass of dry air
$\hat{\mathbf{Y}}$: Denoised data matrix	R: Universal gas constant
K': arbitrary length	T: Temperature
m: number of profiles	L: Discrete integral
δ: error tolerance	z_0: Initial height in Downward integration
$\Vert.\Vert_2$: square root of sum of squares	P: Pressure
$\Vert.\Vert_0$: logic either or	g: acceleration due to gravity
$\mathbf{U}$: Upper trianglar matrix	z: variable height
Δ: Eigen matrix	$\Re$: Real value
$\mathbf{W}$: Lower triangular matrix	$\odot$: Element wise multiplication
$\mathbf{E_{k'}}$: Error matrix	d': atoms of dictionary
T_0:logic 1 (True)	$\hat{S}$: Penalized maximum likelihood function
T': Transpose of a matrix	J: arbitrary function
$\omega_{k'}$: arbitrary variable	γ: Regularization parameter
σ' variance of noise:	$\hat{\mathbf{y}}$: denoised photon count profile
σ'_1 variance of significant coefficients:	$\tilde{\mathbf{y}}$: arbitrary matrix
σ'_2 variance of insignificant coefficients:	S:Likelihood function
p'' Prior for denoising :	P_0: seed pressure
z' :arbitrary variable	$P_{0_{true}}$: Change in seed presure to seasonal variations
$\mathcal{N}$: Normal distribution	Δz: range resolution
$\mathbf{I}$: Identity matrix	p': arbitrary variable
a:constant value	P^*: The power of back scattered signal
$\tau(.)$: Shrinkage factor	α: Temperature dependant parameter
P': Poisson distributed signal count	$\bar{\mathbf{y}}$: arbitrary vector
Q': Poisson distributed noise counts	$E'(.)$: arbitrary function(Expectation)
d':Lidar system dependent function	$F(X)$: non linear operator

$\zeta(x,z,t)$: Electric Field Intensity	λ_z: Vertical Wave length
λ_x: Horizontal Wavelength	m: Vertical Wavenumber
k: Horizontal Wavenumber	
$E_p(z)$	$T'(z,t)$
$T_0(z,t)$	$T(z,t)$
N^2	R_i
$T'_{break}(z,t)$	f_d
$f_{intrinsic}$	$\mathbf{K}$:Wavenumber vector
U:Wind velocity vector	$\mathbf{K}$

Chapter 1

Introduction

The present thesis is on the development of Dictionary Learning Algorithms for solving problems in the Lidar Remote Sensing of the Atmosphere. In this work, we will take the advantage of Dictionary Learning, Deep Learning, Kernel approximation, and Adaptive signal processing techniques to study and analyze the Gravity Wave Breaking events leading to turbulence from the temperature perturbations obtained using the Rayleigh Lidar System. As a first step before analysing wave breaking, one must eliminate all the contributions of other sources of wave energy such as horizontal background wind and planetary waves, Rossby waves, etc. In the present work, we formulate the diagnostic metrics that ensure background wind is successfully removed. We eliminate other wave sources present in the temperature perturbations using the filtering techniques or by using the wavelet decomposition using the Multi-Resolution Analysis technique. The Dictionary Learning approach addressess problems faced during the analysis of the Wave Breaking. The purpose of this paper is to examine the Rayleigh Lidar temperature profiles for solving the research problem. We also use ground based observational data from GPS Radiosonde, Meteor Radar(MR), and also data from Earth Observation Satellites for validation in order to end the existing research gaps.

The present work attracts readers from the areas of Applied Machine Learning, Adaptive Signal Processing, Instrumentation, Remote Sensing, and Atmospheric Science.

1.1 Research Problem

Following below are identified research gaps in the existing scenario and are solved using the Dictionary Learning approach.

- **Problem 1:** The acceptable measurement range of the Rayleigh Lidar System has gone down from 80km at the time of installation in 1991 to 75km in the present year. Adaptive signal denoising algorithms and new methods for retrieval of temperatures are needed to achieve a high Signal to Noise Ratio (SNR) and a low value of uncertainty at higher altitudes respectively.

- **Problem 2:** It is essential to remove all sources present in the background in the analysis of Atmospheric Gravity Waves. The presence of the horizontal background wind in the analysis of Atmospheric Gravity Waves has the effect of triggering instabilities and these instabilities are interpreted as Atmospheric Gravity Waves. The presence of background horizontal wind makes Gravity Wave's intrinsic frequency to shift(Doppler Shift) more from the actual. To address the mentioned problem, we proposed the “Multi-Resolution Dictionary Learning" to address the mentioned problem to separate the background wind effect by promoting sparsity and by using the Instantaneous Frequency Estimation technique.
- **Problem 3:** To enhance the study of turbulence, we must study the breaking of Gravity Waves. For the analysis of wave breaking, we require both zonal and meridional wind velocity components, where the measurements are sparse between altitude 30km and 80km . Here we propose the “Deep Kernel Dictionary Learning technique" that will be capable of identifying the wave breaking features directly from the temperature perturbations without the need for the Wind velocity components.

In overall solving all the problems using the Dictionary Learning approach enhance the study by eliminating background effects such as Gravity Wave breaking, and Turbulence that play a major role in Global Circulation.

1.2 Dictionary Learning Approach to Research Problems

The Dictionary Learning approach has addressed the above mentioned problems. The detailed procedure on how Dictionary Learning can solve the problems is explained in the individual chapters. The approach to solving the mentioned problems is explained below:

1. **Solution to Problem 1:**The first task is to denoise the photon-count profiles using Dictionary Learning via thresholding. A temperature retrieval algorithm using Penalized Maximum Likelihood Estimation is proposed that has less uncertainty at higher altitudes. The contribution of various sources to uncertainty in the proposed temperature retrieval algorithm is estimated using synthetic data and Monte-Carlo Simulations. The interesting result is that we have got an extension in the acceptable measurement range of the Rayleigh Lidar System.

2. **Solution to Problem 2:** The solution to the problem is to formulate a methodology that eliminates the impact of other sources from the analysis for gravity wave data. The temperature perturbations are decomposed into different resolutions using wavelets. Our objective is to reduce the impact of the background horizontal wind component of the Gravity Wave frequency spectrum and wavenumber spectrum. Here we employ multi-resolution dictionary learning to promote sparsity to remove the effect of background horizontal wind on the wavenumber spectra and the Instantaneous Frequency Estimation (IFE) technique to reduce it in the frequency domain.

3. **Solution to Problem 3:** After eliminating ambient horizontal wind component from temperature perturbations, they are the most appropriate for the analysis of Wave breaking events that lead to the turbulent atmosphere. Here we propose an algorithm suitable to detect wave breaking features from temperature perturbations without the actual need for Wind velocity profiles. Detection of wave breaking events is based on incorporating Kernel functions into Deep Dictionary Learning. This algorithm's potential is explained in comparison with existing feature extraction algorithms and validated by using Rayleigh Lidar and meteor radar.

1.3 Outline of Thesis

The present chapter gives brief information about the existing research problems in Lidar remote sensing of the atmosphere, especially the middle atmosphere (30km to 80km) region.

Chapter 2 gives detailed information on the instruments and data used in the thesis, information on the scattering of light by the atmosphere, mathematical formulation of received back-scattered power density, and its dependency on laser pulse energy. It also defines the standard definition for the calculation of Signal to Noise Ratio, and uncertainty in the temperature retrieved in using the algorithm Chapter 2 also describes methodologies for the calculation of temperature perturbations and interpretation of characteristics of the Atmospheric Gravity Waves from density, temperature and velocity perturbations.

Chapter 3 gives a detailed explanation of the implementation of Dictionary Learning and Deep Neural Networks. It gives information on the formulation of the optimization problem for the given task by adding constraints for single and multi-level parameters. The various performance metrics for the evaluation of proposed algorithms are defined clearly in this chapter. Discussions regarding the selection of optimal bandwidth for kernel Density Estimates is clearly defined.

Chapter 4 provides information about the application of Dictionary Learning with Penalized Maximum Likelihood for reducing the uncertainty in the temperatures retrieved at higher altitudes. The method used for temperature retrieval and its uncertainty budget in the temperature profile, which is discussed in this chapter, are presented. This chapter discusses how to improve the acceptable measurement range.

Chapter 5 is about the temporal and spatial representation of temperature profiles in multi-resolutions. It discusses both the temporal and spatial methodology for removing background winds. This chapter also discusses seasonal variations in the frequency of Gravity Waves caused by Doppler shift. The proposed methodology of removing the background wind effect helps to investigate and study Global Circulation more effectively.

Chapter 6 is about the introduction of kernels to Deep Dictionary Learning for the detection of sudden transient events. The chapter discusses the application of Deep Kernel Dictionary Learning to the detection of Atmospheric Gravity Wave breaking events. A case study on the detection of Gravity Wave breaking events is used to discuss performance metrics, detection rate, and false detection rate.

Chapter 7 is on the detailed implications of the Dictionary Learning technique on improvement in the acceptable measurement range, removal of background wind effects, and detecting Gravity Waves breaking.

1.4 Summary

1. A clear road map to the existing research problems in Rayleigh Lidar remote sensing of the Atmosphere is discussed.

2. All the research problems are explained and approaches to solve them using Deep Learning, Dictionary Learning, and Kernel theory is explained.

3. The approach to solution of research problems is explained in detail in each Chapter.

Chapter 2

Instrumentation and Data

2.1 Light Scattering in the Atmosphere

The scattering of light is classified into Elastic and In-elastic scattering. Elastic scattering processes conserve kinetic energy of particles in their center-of-mass frames, while changing scattering direction. In elastic scattering, Linear Energy Transfer (LET) occurs when incident particles achieve the same speed and energy as their surroundings. While in the In-elastic scattering, the total Kinetic energy is not conserved. When the incident particle scatters inelastically, some of its energy can be lost or increased. Photons that undergo inelastic scattering lead to Raman scattering . The phenomena of in-elastic scattering is observed if the interaction is among electron and a photon. In addition, the electron gives energy to a photon when it collides with an electron with relativistic energy. Irreversible Compton scattering is the result of this process.

2.1.1 Rayleigh Scattering and Mie Scattering

Interaction of laser pulses with the atmosphere can be treated as Rayleigh scattering if the atmospheric height is between 25kms to 80kms and Mie-Scattering if the atmospheric height is between 0kms to 25kms.

Rayleigh Scattering: In atmospheric science, the Rayleigh scattering is elastic scattering of light by atoms or molecules where the wavelength is much shorter than that wavelength of the radiation. Electric polarizability of particles causes Rayleigh scattering, which is a parametric process. Radiation from the polarized particles appears as scattered light because they act as small radiating dipoles.

$$x = 2\pi r/\lambda \tag{2.1}$$

The size of the particle is often characterized by equation (2.1) where x=size of the particle;r=characteristic length(radius);λ=wave length of light. For Rayleigh scattering the value x <1.

Mie-Scattering($x \geq 1$): In Mie theory, the ratio of diameter of particle to wavelength are greater than unity in case of scattering by clouds and haze. Mie scattering in the atmosphere is caused by smoke, pollen, dust, and microscopic water droplets. As particles get larger in size at lower altitudes, Mie scattering occurs. When particle diameters are approximately equal to wavelengths, Mie scattering is dominant at altitudes between 4.5 and 25 km in the atmosphere. For large particles, Mie scattering theory converges to geometric optics' limit.

2.2 Lidar Sensing of Atmosphere

Light Detection and Ranging(Lidar) works similarly to Radio Wave Detection and Ranging(Radar). The transmitted Electromagnetic (EM) signals are reflected back, attenuated, and recorded in terms of photon counts by a variety of atmospheric components, such as gaseous molecules, aerosols, water vapor, and ice crystals. The received photon counts are used for interpreting the atmospheric changes, and for characterizing the atmospheric waves. In addition, Lidar system studies various aspects of the atmosphere based on the wavelength. This is the primary advantage of Lidar systems over radar systems.

In an elastic scattering process, back-scattered photons have the same wavelength and energy as incident photons. The wavelength selection is dependent on probing the required altitude of the atmosphere, the molecules, and the aerosols. Up to 25 km altitude, Mie scattering is dominant, and it is associated with atmospheric aerosols. The elastic scattering phenomena are dominant in the atmosphere. As a result of the interaction between a photon and an electron in an elastic scattering process, the electron gets excited to a high energy unstable state when it interacts with a photon. In order to return to its initial state, the excited electron emits a photon that has a wavelength equal to that of the incident photon.

2.2.1 Lidar System

As shown in Figure 2.1, a Lidar system is briefly described, and in particular the Lidar system at National Atmospheric Research Laboratory, Gadanki, India as shown in Fig. 2.2. In contrast to detection systems like Radar, Sodar and Lidar use high-energy short laser pulses that are collimated together to have a good spatial resolution. In Lidar, there are three parts: the transmitter, the receiver, and the detector. Into the atmosphere, the transmitter emits a pulsed beam of high intensity from a laser. As the laser pulse hits molecules or particles in the atmosphere, it gets back-scattered. Using a telescope, the receiver collects the back-scattered signal and focuses it on the Photo Multiplier Tube (PMT), which transforms it into an electrical signal.

Alternatively, there is a photon counting method that uses a photon counting mode. Signal to Noise Ratio (SNR) decreases as a function of the square root of altitude. A near-field spatial resolution is maximized in an analog system while a far-field spatial resolution is minimized. Comparatively, to analog mode, the returning photons are recorded over long periods of time using photon counting mode as described in [1]. The photon counting technique is very effective if we want to detect signal from higher altitudes that have low signal strength. Through pulse height discriminators, signals are separated from noise in the photon counting technique, allowing high-precision measurements having a Signal-to-Noise Ratio (SNR) higher than in analog mode. Digitizers convert analog signals into digital signals, which are recorded by the photon counting data acquisition system based on distance from the Lidar.

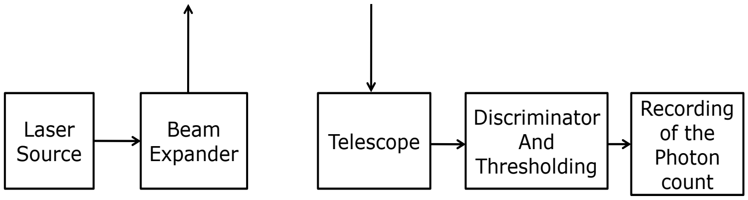


Figure 2.1: Schematic block diagram of Light Detection and Ranging System

2.2.2 The Rayleigh Lidar System at NARL

The Rayleigh Lidar at National Atmospheric Research Laboratory (NARL), Gadanki (13.5^0N, 79.2^0E), India operates with a wavelength of 532nm at 600mJ of energy per pulse with Pulse Repetition Rate (PRR) at 50Hz. The receiving system consists of a Newtonian-type telescope with a diameter of $750mm$ and a Field of View of $1mrad$. A Photo Multiplier Tube (PMT) is connected to the focal point of the telescope to convert light into electric signals. PMT output is fed to the pulse discriminators. The output of the pulse discriminator is fed to electronic boards(Multi-Channel Scaler) for acquiring data by counting photons. During data acquisition, the PMT used is Hamamatsu R3234-01. The photon-counting technique is implemented by a Multi-Channel Scaler with a maximum counting rate of 100MHz. The Rayleigh Lidar system includes two channels, namely a high sensitivity channel and a low sensitivity channel for extending the dynamic range of the system. The low sensitivity channel is meant for collecting back-scattered signals from lower altitudes ($< 50km$) and the high sensitivity channel is for higher heights ($> 50km$). The two channels have two PMTs with the same specifications and their gains are in the ratio of $9 : 1$ respectively. In the present analysis, temperature profiles of the two channels were merged in the height region between 45km to 55km.

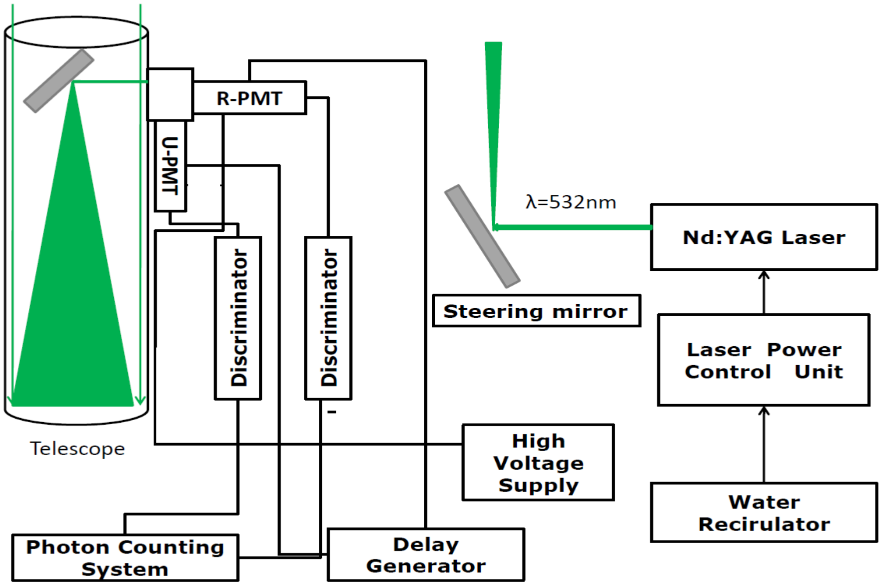


Figure 2.2: Configuration Of The Rayleigh Lidar System at National Atmospheric Research Laboratory, Gadanki, India.

2.2.3 Lidar Equation

In the present section, the relationship between the back-scattered photons and atmospheric density is derived. The expression is derived based on the following three assumptions:

1. Each photon has only one scattering event.
2. The pulse width is shorter than the trigger pulse used to trigger the data acquisition system.
3. There is a constant number of back-scattered molecules or atoms in the atmosphere per range bin (Δz).

Table 2.1: Transmitter and Receiver specifications of Rayleigh Lidar System at NARL, Gadanki

Transmitter Specifications		
Quantity	***Value***	***Units***
Energy Per Pulse	600	mJ
Pulse Width(t)	7	nsec
PRF(1/T)	50	Hz
Wavelength(λ)	532	nm
Resolution(Δz)	300	m
Beam Divergence	0.1	mrad
Receiver Specifications		
Beam Divergence	1	mrad
Telescope Geometry	Newtonian	
Diameter	750	mm
Field of View	1.0	mrad
Data Acquisition System		
Type	Single Photon Counting	
Model	EG & G Ortec,MCS-Plus	
Maximum Counting Rate	100	MHz
Dwell Time	2	μsec

$$N(z_i) = N_0 A K_0 K_q \frac{\tau^2(z_0, z_i)}{4\pi(z_i - z_0)^2} \left[n_r(z_i)\beta_r + n_m(z_i)\beta_m(z_i) \right] \Delta z + B(z_i) \tag{2.2}$$

Based on these assumptions, the mathematical expression for Lidar back-scattered signal is given by (2.2), where z_i is the i^{th} altitude, $N(z_i)$ represents photons detected from back-scattered light, ΔZ is the altitude distribution, z_i, N_0 is the photon count per laser pulse, 'A' represents aperture area, K_0 denote optical efficiency, K_q represent the quantum efficiency of the Photo-Multiplier Tube, $\tau(z_0, z_i)$ transmission coefficient of atmosphere, $n_r(z_i)$ and $n_m(z_i)$ are the air molecule and the aerosol concentrations respectively, β_r and $\beta_m(z_i)$ represent the Rayleigh and Mie back-scatter cross-sections respectively, and $B(z_i)$ is the background count.

2.2.4 Ascertaining Atmospheric Density Profile Using Lidar Counts

The equation(2.2) is a generalization of the relationship between back-scattered photons and atmospheric density applicable for all types of scattering phenomena, and can be reduced to simpler form depending on scattering process. When Rayleigh scattering occurs, the transmittance parameter is constant, whereas when Raman scattering occurs, the transmittance varies by the wavelength of the back-scattered photon. Dark counts and background noise, are considered to be measured between 120 km and 147 km altitude.

$$\rho(z_i) = C\Big(N(z_i) - B(z_i)\Big)(z_i - z_0)^2 \tag{2.3}$$

After subtracting the noise counts, the Lidar equation (2.2) reduces to equation (2.3), considering all other atmospheric parameters to be constant. The atmospheric density (ρ) as a function of altitude is given by equation 2.3, where z_i is the i^{th} altitude, $N(z_i)$ represents photons detected from back-scattered light, $B(z_i)$ represents background photon-count at i^{th} altitude, C represent a constant that considers all parameters of Lidar system, and z_0 is the seed altitude from where the downward integration of density profile will be initiated.

The atmospheric pressure profiles are obtained using equation 2.3 and the ideal gas law. From the pressure profiles, temperature profiles can be derived. As density and pressure terms occur in ratios, all that is required for retrieval is a relative density profile.

2.2.5 Photon Count Uncertainty

The Rayleigh Lidar System measurements consist of two types of uncertainties: systematic and random. The systematic uncertainties include assumptions in defining parameters such as the accuracy of acceleration due to gravity with altitude, using hydro-static equilibrium conditions or an atmosphere behaving as an ideal gas. In addition to systematic uncertainties, signal counts left after subtracting the background component from Lidar measurements and the selection of backscatter-to-extinction ratios are also considered systematic uncertainties. The Poisson distribution characterizes the random uncertainties associated with the photon counting process. In this section, we address the systematic uncertainties associated with photon counts.

There are two main uncertainties in the measured photon count profile: nonlinear count correction and random uncertainty. At the bottom of each channel, non-linear correction is large where the count rate is high. The SNR is low at higher altitude as the atmospheric density exponentially decreases with altitude. The time integration techniques (Pulse averaging or co-adding the pulses) are often used to improve the SNR. The time integrated profiles have large SNR. The time integration technique has the draw back of providing low spatial resolution. In Lidar measurement systems, where the back-scattered collection of photons is through photon counting mode, the statistical uncertainty is dominant and follows Poisson distribution.

$$\sigma = \frac{\sqrt{N}}{N} \tag{2.4}$$

where σ is the standard deviation of the random process, N is the number of counts. Taking the number of counts as a measure, we can determine the standard deviation using (2.4). From the equation 2.4, it can be inferred that when N is higher, the statistical uncertainty decreases and can be analyzed with respect to photon count profile.

2.2.5.1 Problems in retrieving the temperature profiles

Temperature retrieval is limited by the seed pressure value in the HC method as reported in [2]. For the ideal gas law to be applied, first an absolute pressure that is hydro-statistically balanced must be generated. In spite of the uncertainty in seed pressure, changes in seed pressure will shift pressure profiles parallel by ΔP. The temperatures profiles retrieved at higher altitude has most uncertainty. It was reported in [2] that, when there is an uncertainty of 15% for seed pressure, the uncertainty for retrieved temperatures falls below 2%, 15 km lower than the top altitude, and less than 1%, 5 km lower than the top altitude. At high altitudes, this problem arises irrespective of altitude. In using the HC method [2] data must be discarded even when working with the middle atmosphere (20km-50km) or upper atmosphere (50km-100km). At higher altitudes, another issue arises due to the increasing rarefaction of the atmosphere. A seed temperature (T_{seed})value can also be determined at the top of the atmosphere, in order to initiate the algorithm described in [2].

2.3 Signal To Noise Ratio

To calculate SNR for this study, we used equation 2.5.

$$SNR = \frac{S}{\sqrt{S + S_{Background}}} \tag{2.5}$$

where S is the number of back-scattered photons, and $S_{Background}$ represent the background noise photon count. The background count is the average of all photon counts at 150km to 300km altitude. The various sources that contribute to the noise photon count are given in Table. 2.2.

Table 2.2: Kinds of noise related to Lidar Measurement

Kinds of Noise	Physical Mechanism
Quantum Noise	Statistical Fluctuations of signal radiation
Radiation Noise	Statistical Fluctuations of background radiation
Dark current Noise[1]	Thermal generation of carriers without optical signal
Thermal Noise	Thermal agitation of current carriers

2.3.1 Coadding Photon Count Profiles To Increase SNR

The data with $SNR < 3$ are rejected during raw count correction as a preprocessing step. As a result of co-adding spatial and temporal data, the established method sacrifices some spatial and temporal resolution to increase SNR.

$$\mu(X_1 + X_2 + ...X_n) = \mu(X_1) + \mu(X_2)....\mu X_n \tag{2.6}$$

$$\sigma(X_1 + X_2 + ...X_n) = \sigma(X_1) + \sigma(X_2)....\sigma X_n \tag{2.7}$$

$$SNR = \frac{mean}{\text{Standard Deviation}} = \frac{\mu}{\sigma} \tag{2.8}$$

Assuming that there are **M** samples per night, the SNR will be given by (2.9).

$$SNR = \frac{M\mu}{\sigma\sqrt{M}} \tag{2.9}$$

Integration of photon profiles improves the SNR as represented in equation (2.8). In statistical theory, the mean(μ) and variance(σ^2) of the sum of 'n' randomly distributed variables given by 2.6 and 2.7 respectively. Depending on the tolerance limits, the SNR limit varies. Thus the SNR increases by $\sqrt{M}$, and for improvement in Signal to Noise Ratio, co-adding data are generally used in conjunction with conventional data retrieval methods.

2.3.2 Upper Altitude Limit of the Rayleigh Lidar System

A long time ago, Rayleigh Lidar was believed to be able to measure the temperature profiles of the atmosphere up to a radius of about 90km or 100km. However, above this region, the technique becomes invalid since atmospheric composition changes affect the basic assumptions that underpin Rayleigh Lidar, causing it to fail. The unique capability of Rayleigh Lidars to measure the backscatter from well above 100km makes it particularly important to study how atmospheric composition affects the derived Rayleigh Lidar temperature profiles. An NRLMSISE-00 model was used to simulate Lidar signal (photon-count) profiles, incorporating changing atmospheric composition, to analyze biases and errors associated with extending Rayleigh Lidar temperature measurements above 100km. The effect of altitude on atmospheric composition has also been investigated for biases.

Based on simulations in [3], Rayleigh Lidar's upper altitude limit appears to be more dependent on seed temperatures and pressures than on atmospheric composition. The standard Rayleigh Lidar technique is relatively insensitive to atmospheric composition changes with altitude if its temperature retrieval algorithm is started at 110 km or below (bias typically 0.1 K). Initializing the Rayleigh Lidar algorithm with a small seed value is the limit to extending Rayleigh Lidar temperature measurements. Seeding temperature uncertainties are likely to be much greater than uncertainty in atmospheric composition uncertainties unless there are independent, coincident, accurate measurements at the integration start altitude.

2.3.3 Spatial and Temporal Resolution

A spatial resolution is a measure of how much spatial detail an observation contains, whereas a temporal resolution is the amount of time it would take to revisit and acquire the same data. NARL's Rayleigh Lidar system was used to acquire the data used for this study. The data was acquired with a spatial resolution of 300m, and a temporal resolution of four minutes.

2.4 The Vertical Structure Of Atmosphere

Gases, solids, and liquid particles make up the Earth's atmosphere. There is no clear limit to the upper limit of the atmosphere, but gaseous materials extend several hundred kilometers above the earth's surface. Although the upper limit of the atmosphere is not well defined, gaseous materials can be found several hundred kilometers above the ground. More than 99 percent of the earth's atmosphere is contained within the first 80 km of the atmosphere.

Atmospheric properties include thermal properties, chemical composition, electrical attributes, and density which are often divided into vertical layers. A change in one of these physical properties defines a new layer in the atmosphere, which determines the vertical structure. The stratosphere, troposphere, mesosphere, and thermosphere are the four layers of the atmosphere. A tropopause, a stratopause, a mesopause, and a thermopause are each named after the top altitudes of these layers, as shown in Figure 2.3.

Atoms and molecules in the atmosphere are measured in terms of their density per unit volume. There is no fixed arrangement of molecules in air, and they can move at random. A unit of mass can contain any amount of mass, so air density can fluctuate. Altitude and air density are not linearly related because atmospheric gases are compressible. When air molecules are subjected to gravitational acceleration over a surface, the pressure in the atmosphere is determined.

Generally, sea-level pressure on Earth is $1011hPa$ without considering the local variability. The pressure decreases with increasing height in the same manner as the density of dry air.A pressure of 500 hPa corresponds to approximately half of the atmosphere below 5.5km. Nearly 90% of the atmosphere lies below 16km and has a pressure of 100 hPa. In terms of thermal characteristics, the explanation of the structure of the atmosphere is more complex. Interactions between atmospheric gases and radiant energy from the Sun and Earth determine the vertical temperature profile. The troposphere extends to about 12km from the surface and is the first layer of the atmosphere. Radiant energy exchanges with the underlying surface determine the troposphere's temperature. As a result, the troposphere's temperature decreases with height at an average rate of $6.5^oC/km$, a measurement referred to as a normal lapse rate. There is an isothermal layer at the top of the troposphere, the Tropopause, where the temperature does not change with altitude. According to latitude and season, the tropopause occurs at heights ranging from 9km to 16km. The stratosphere lies above the tropopause. The isothermal layer in the stratosphere allows temperatures to increase with height. A temperature rise is caused by an increase in the concentration of ozone molecules (O_3) at altitudes between 15km and 30km above the surface of the planet. Stratosphere temperature is increased by UV solar radiation absorbed by the ozone molecules. At an altitude of 50km, the Stratopause, the transition point from Stratosphere to Mesosphere, ends the stratospheric warming. The Mesosphere experiences the coolest temperatures, which range from 140K to 183K at 80 km above the Earth's surface. Temperatures in the Mesosphere rapidly decrease with height. A mesopause, which divides the Mesosphere and Thermosphere, is where temperatures are lowest.

Due to the small amount of oxygen molecules that absorb solar energy (O_2), the temperature of Earth's outer layer increases substantially with height. There is no significant amount of heat at the surface, despite temperatures reaching thousands of Kelvin, as atmospheric gases have a very low density. Ions and free electrons are also found in the

Thermosphere (and upper mesosphere) sublayer. In the region between 80km and 400km, charged particles from the sun interact with these ionized molecules to produce spectacular light displays known as aurora borealis in the Northern Hemisphere and aurora australis in the Southern Hemisphere.

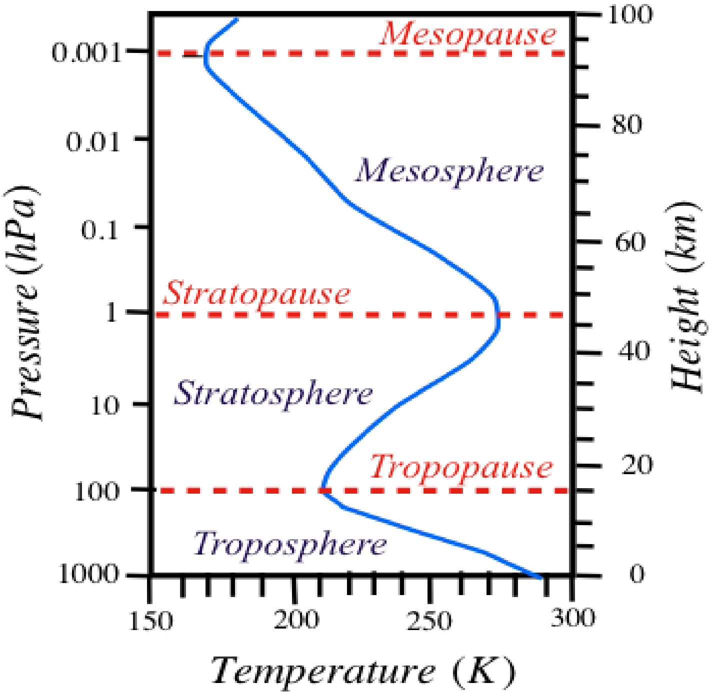


Figure 2.3: Vertical Temperature Structure Of Atmosphere

2.5 Methodologies For Atmospheric Temperature Retrieval

There are several algorithms that allow a temperature profile from photon-count data to be determined. There are two basic assumptions made by all retrieval algorithms based on fundamental physics principles.

1. Atmospheres behave like ideal gases, and

2. Hydrostatic equilibrium exists in the atmosphere.

2.5.1 Downward Integration Method

The Downward Integration (DI) [2] method assumes the Hydro-Stastic Equilibrium (HSE) condition. Lidar photon count and temperature at each altitude in the Lidar range are related using the HSE, the ideal gas law, and the reduced Lidar equation. Hauchecorne and Chanin first proposed this method in [2].

$$\rho(z_i) = C(n(z_i) - B(z_i))(z_i - \Delta z/2)^2 \tag{2.10}$$

$$ln[P(z_i - \Delta z/2) - P(z_i + \Delta z/2)] = \rho(z_i)g(z_i)\Delta(z) \tag{2.11}$$

$$T(z_i) = (Mg(z_i)\Delta(z))/R\ln\left[P(z_i - \Delta z/2) - P(z_i + \Delta z/2)\right] \tag{2.12}$$

$$g(z_i) = g_o(1 - 2(z_i/R_e)) \tag{2.13}$$

where $T(z_i)$ is the temperature retrieved at particular layer, $\rho(z_i)$ is the atmospheric density at the i^{th} layer and Δz is the thickness of the atmospheric layer, g_0 is the acceleration due to gravity, R_e is the radius of the earth, R is the ideal gas constant. Using HSE, ideal gas law equation and equation 2.10, we obtain the temperature value at each spatial resolution using equation 2.12. The following drawbacks are observed in the Downward Integration method:

1. Mesosphere temperature and density measurements have large uncertainties, which propagate throughout the profile. Improper selection of seed pressure results in the discarding of temperature values for the first 10km to 15km from the temperature profiles to have a validated temperature profile.

2. The temperature within the each spatial resolution (Δz) is assumed to be constant. This assumption results in temperature uncertainty near altitudes having sharp temperature gradients near at regions of Stratopause and the Mesopause.

2.5.2 Non-linear Inversion Technique

The DI method of calculating atmospheric temperature profiles from Rayleigh Lidar system measurements has the limitation of discarding the temperatures at 10km below the seeding altitude, due to the uncertainty in the selected seed pressure value. This problem has been overcome by using non-linear inversion techniques where there exists a reliable and acceptable Signal to Noise Ratio (SNR). The methodology for the retrieval of temperature profiles using inversion techniques is reported by [4]. They explained the potential of the inversion technique using Monte-Carlo simulations in terms of systematic and random uncertainties in temperature profiles in using the inversion techniques. By using the inversion technique, we are able to provide seed pressure values at lower altitudes in the atmosphere where the uncertainty is lower. Thus, the inversion technique virtually extends the altitude range of existing Rayleigh-scatter Lidar from 10km to 15km, which is equivalent to four times the power-aperture product.

2.5.3 Optimal Estimation Method

The shortcomings in the Downward Integration and Grid search methods are overcome by the Optimal Estimation Method. In this technique the forward models are used to completely characterize and retrieve the Atmospheric temperature profiles. The detailed procedure for the formation of the mathematical model, applying the inversion technique, and solving the optimal representation of the model obtained as Bayes Theorem using the Levenberg-Marquardt Algorithm (LMA) is given in [5, 6]. Multiple channels covering different height ranges and vertical resolutions can be used to retrieve a single temperature profile. As a result of the Optimal Estimation Method, robust estimates of temperature can be obtained that are consistent with previous methods while requiring minimal computing resources.

2.6 Standard Error and Uncertainity

$$T(z) = \frac{M_{air} g(z) dz}{R \log(1+X)}; X = \frac{\rho(z) g(z) dz}{P(z+\Delta z/2)} \tag{2.14}$$

$$SE = \frac{dX}{(1+X)\log(1+X)} \tag{2.15}$$

The expressions for temperature (2.14) and SE (2.15) is given in [2]. A seed pressure value from an atmospheric model is used for the initialization of the method for temperature retrieval.

As a result of systematic uncertainty, the mean of observations deviates by a certain amount from the attribute's true value every time an experiment is conducted, whereas random uncertainty refers to fluctuations in the mean of observations caused by fluctuations in the mean of observations caused by a natural fluctuation in observations which is usually unreplicable. In lidar remote sensing, both types of uncertainties are present. As a result, we will need to include variables with both types of uncertainties in our forward model when modeling synthetic lidar counts for testing or error analysis. A list of input variables and their uncertainty characteristics is provided in Table 2.3. Due to their negligible uncertainties of 0.01, the uncertainties in mean molecular mass M, Universal gas constant R, and gravitational acceleration g(z) will not be analyzed.

Table 2.3: Sources of uncertainties

Variable	**Uncertainity**	**Type**	**Value**	**Source**
N(z)	shot noise	Random	$\sqrt{N(z)}$	Photon noise
N(z)	geophysical variability	Systematic	Depends on data	Atoms. variability
P_0	Uncert. in seed pressure	Systematic	$\pm 10\%$	Model
$T_{guess}(z)$	Uncert. in Temperature	Systematic	$\pm 5K$	Model

2.6.0.1 Generation of Synthetic Data

The preliminary tests for the proposed algorithms are conducted with synthetic photon count profiles that are generated by changing the parameters, and are free from noise, unlike the real-time Lidar photon count profiles. In order to incorporate the noise behavior, we used backscattered photons following a Poisson distribution. As a result, the synthetic profiles were initially subjected to Poisson noise. The inclusion of noise alone destabilizes the convergence problem. The synthetic counts also included altitude-dependent geophysical noise at each altitude that had a temporal mean of zero. With this, the synthetically generated photon count profiles are similar to real time photon-count profiles. Also adding the noise term doesn't have an effect on the retrieval of the temperature. Considering the variability of seed pressure at the seeding altitude, seed pressure is affected by -10%, -5%, $+5\%$, and $+1\%$ variation to assess the seed presure uncertainties. These new seed pressure values were used to implement the algorithm.

2.6.0.2 Estimation of uncertainties

The uncertainties are classified as Systematic and Random uncertainties. Whenever an experiment is conducted, systematic uncertainties cause the mean of observed data to deviate over a certain period of time from its true value. Random uncertainties are fluctuations in the mean of the observations. As these values contribute less to uncertainty, uncertainties arising from mean molecular mass (**M**), universal gas constant (**R**), and acceleration due to gravity (g_z) will not be considered.

Below an altitude of 35km, the back-scattered visible light will be absorbed by the Ozone layer. As a result, temperatures are higher than expected. Ozone corrections are therefore applied below 35km altitude. This study conducts all experiments for Upper Atmosphere, no Ozone correction factor is applied in the present analysis. It is not possible to obtain an analytical expression for the uncertainty analysis. As a result, the uncertainty analysis is numerically performed using the Monte-Carlo method. In the present work, uncertainties are studied using Monte Carlo methods.

- **Modelling Systematic uncertainties:** Including the effect of systematic uncertainties is not straight forward. The effect of geophysical variability is compensated in the synthetic profiles so that its temporal mean is zero. The parameter values related to systematic uncertainties are obtained from models or from an experimental source. When all counts are averaged over time at each altitude, geophysical variability is calculated. If we subtract the mean average from each count, we are left with only the mean geophysical variability of zero with random noise superimposed on it. Using a guessed temperature profile, the original synthetic counts (without noise) are added to the average of the counts over time to fit a smooth curve. Temperatures retrieved by the algorithms are stable and convergent despite geophysical variability.

- **Modelling Random uncertainties:** The Shot noise is not only due to the nature of photon counts but also due to the detector electronic used for data acquisition. A detailed discussion on evaluating the random noise is discussed in [7]. An estimate of the shot noise of the Lidar system can be derived from a Poisson distribution around the mean number of counts collected per unit time. For creating synthetic counts, Poisson distributions about the mean are used. Taking a temporal average cancels out geophysical variability at each altitude while taking the mean. Temperature profiles derived from synthetically generated counts are regarded as being within the range of random uncertainty due to shot noise due to temperature standard deviations.

- **Modelling Error in Seed Pressure:** The algorithms used for temperature retrieval need to be initialized with a seed value at top altitude, obtained either from a model or from experimental studies. The seasonal variations in the seed pressure can be accounted for 10% variability of the mean as reported in [8, 9]. The uncertainty due to seed pressure is explained in two categories: A set of synthetic temperature profiles were generated with a seed pressure value defined as $P_{0,True}$ that incorporates Poisson noise as well as geophysical variability. To analyze the effect of uncertainty due to guessed seed pressure value($P_{0,guess}$), we run the algorithm for $P_{0,guess} = P_{0,true} + 10\% P_{0,true}$. Whenever there is only shot noise and no uncertainty in the estimate of seed pressure, i.e. $P_{0,guess} = P_{0,true}$, there is an error in the seed pressure estimate, adding uncertainty to the temperature estimate.

2.6.0.3 Error Propagation: Contribution due to all sources

A total uncertainty in retrieved temperatures is calculated by combining the effects of uncertainty from all sources. A Monte-Carlo analysis is used to estimate the uncertainties. An estimate of the true synthetic count N_{true} (geophysical variability) is enhanced by the following sources of uncertainty:

1. A Poisson random number generator is used to add noise by counting the numbers with a mean of N_{true}.
2. Seed pressure values are selected using a Gaussian random number generator with mean P_{true} and standard deviation $\frac{10}{3}\% P_{true}$.
3. The temperature profiles are predicted using a Gaussian random number generator with mean as T_{true} and standard deviation $\frac{5}{3}$.

For obtaining standard errors of temperature data retrieved from real-world data, an error estimation method of the above type will be employed. A table of the uncertainty budget for the NARL Rayleigh Lidar System is given in Table 2.4.

Table 2.4: NARL Rayleigh Lidar System Uncertainty Budget. $(.)/(.)$ indicates the uncertainty in high and low intensity channels.

Source	**80km**	**70km**	**50km**	**30km**
Detection Noise(N_D)	6K/(.)	5K/(.)	1K/4K	(.)/0.08K
Background Noise Count(N_B)	0.5K/(.)	0.3K/(.)	0.3K/(0.8)	(.)/0.02K
Seed temperature	4K	2K	0.5K	-
Gravity(g)	0.1K	0.1K	0.1K	0.1K
M_{dryair}	1K	0.05K	0.05K	0.05K
Method	15K	8K	3K	0.08K
Total Uncertainty	27K/40K	16K/(.)	4.8K/8.45 K	0.08K/0.33K

2.7 Gravity Waves In The Atmosphere

Atmospheric Gravity Waves (AGW) have been observed and characterized using Rayleigh Lidar ground observations [10, 11, 12, 13, 14, 15, 16]. A few meters to kilometers above ground, the middle atmosphere is characterized by fluctuations in temperature due to Gravity Waves (GW). Currently, only Lidar can provide sufficient vertical and temporal resolution for studying Gravity Waves in the Stratosphere and Mesosphere. As a basis for estimating Gravity Wave induced temperature perturbations T', Lidar measurements are broken down into background temperature (T_0) and perturbations T around the background. A vertical height of the atmosphere and a time instant t are dependencies of parameters T, T_0, and T'. The sources of instabilities include background wind, temperature, and other wave sources like planetary waves. The removal of background sources is therefore essential before analyzing Atmospheric Gravity Waves. To remove background wind from Gravity Wave analyses, a Multi-Resolution Dictionary Learning technique is used.

$$T'(z,t) = T(z,t) - T_0(z,t) \tag{2.16}$$

Observed temperature minus background temperature is then defined as the Gravity Wave signature, as represented in (2.16), where ‘z’ is altitude, $T_0(z,t)$ represents the background temperature profile. $T_0(z,t)$ consists of a series of temperature profiles, $T(z,t)$, at a time, t. The data used for the present analysis is collected using the Rayleigh Lidar system facility available at National Atmospheric Research Laboratory (NARL), Gadanki (13.5^oN , 79.2^oE), India.

The present work on the application of Multi-Resolution Dictionary Learning with Sparsity for the removal of Doppler Shift in the observed spectra of temporal variations, due to the presence of winds in the background in the analysis of Gravity Wave temperature perturbations, constitute a part of the present thesis. In the Wavelet domain, sparsity can reduce the effect of background wind.In the Wavelet domain, sparsity can reduce the effect of background wind. Doppler shifting results in an improvement in horizontal energy density at high frequencies, a reduction in vertical energy density at high frequencies, and a significant transfer of vertical energy density to frequencies above buoyancy frequency. Thus it is important to reduce any effects, that limit the analysis of Gravity Waves need to be addressed with paramount importance. Various waves and wave-like structures shape the structure of temperature and wind in the middle atmosphere. Most analysis approaches for the study of gravity wave activity are dependent on spatial scale separation for removing the initial background, while neglecting the effect of Doppler shifting. Thus the present work is of high importance to the Lidar community.

2.7.1 Characteristics of Gravity Waves

In the present work, it is considered that the wave is propagating along the y-axis. Vertically, the z-axis represents the atmospheric height, while the x-axis shows the observation timeframe. The vertical displacement field $\zeta(x, z, t)$ of a traveling wave, induced by some irregular perturbation, is given by

$$\zeta(x, z, t) = A_\zeta \cos\left(k\bar{\mathbf{a_x}} + m\bar{\mathbf{a_z}} - \hat{\omega}\right) \tag{2.17}$$

where k and m represent the real-valued horizontal and vertical wavenumbers, respectively. $k = \frac{2\pi}{\lambda_x}$ and $m = \frac{2\pi}{\lambda_z}$ with the respective wavelengths λ_x and λ_z and the wave is propagating along the $y - axis$. The subscripts x, y, and z represent the x-axis, y-axis , and z-axis respectively of a 3 dimensional Cartesian coordinate system and $\mathbf{a_x}$, $\mathbf{a_z}$ represent the unit vectors along the x-axis and z-axis respectively. The wave frequency $\hat{\omega} = \frac{2\pi}{T}$($T$ is wave period) can be calculated using the dispersion relation for waves in the xz-plane and expressed as

$$\hat{\omega} = N^2 \left[\frac{k^2}{k^2 + m^2}\right] \tag{2.18}$$

where N defines the buoyancy frequency. By combing the sign components of the wavenumber vector $\bar{K}$=$k\mathbf{a_x}$+$m\mathbf{a_z}$ determine the direction of propagation of the Gravity Wave i.e., the direction of phase progression of the wave. A convenient way to express $\hat{\omega}$ is

$$\hat{\omega} = N\cos(\phi) \tag{2.19}$$

with

$$\phi = \arctan\left(\frac{m}{k}\right) \tag{2.20}$$

where ϕ represents the angle between $\bar{\mathbf{K}}$ and $K{-}axis$. When wavenumber components and the sign of ϕ are combined, it is possible to determine if waves are propagating upwards, downwards, leftwards, or rightwards. In order to determine whether the waves seen in the ground-based Lidar measurements are propagating upward or downward, it is necessary to determine the horizontal phase propagation and the ambient wind. Vertical phase lines indicate downward wave propagation while horizontal phase lines indicate upward wave propagation. A quasi-Doppler shift is caused by the background wind, which gives an observational frequency shifted by ($\bar{\mathbf{K}}.\bar{\mathbf{U}}$) from intrinsic frequency as given as

$$f_{obs} = f_{intrinsic} + \frac{|\bar{\mathbf{K}}||\bar{\mathbf{U}}|\cos(\theta_r)}{2\pi} \tag{2.21}$$

where $\bar{\mathbf{K}}$=$k\mathbf{a_x}$+$m\mathbf{a_z}$, $\bar{\mathbf{U}}$=$u\mathbf{a_x}$+$v\mathbf{a_z}$, θ_r measures the angle between wind (u) and wavenumber (k).

2.7.2 Role of Gravity Waves In Global Circulation

As atmospheric gravity waves break, a dynamic process occurs that contributes significantly to determining the general circulation of middle atmospheres. At the height where the horizontal phase speed equals the horizontal background wind speed in the direction of propagation, Gravity Waves get trapped. This altitude prevents the wave from exceeding it and causes it to be attenuated. A wave gets trapped at a certain altitude called a critical layer, at which critical absorption occurs. As reported in [17], gravity waves are significantly attenuated in the mesosphere (50-90 km). From lidar observations, it is possible to detect Gravity Wave breaking, as shown in [18]. As discussed in detail in the following sections, Deep Kernel Dictionary Learning (DKDL) currently has several advantages over conventional methods. Gravity Waves in the Mesosphere are sufficiently strong to cause significant wind and temperature disturbances in the atmosphere, leading to low atmospheric stability. An overview of the behavior of Gravity Waves in the atmosphere is given in [19, 17, 20, 21, 22].

The coupling of layers in the atmosphere is greatly influenced by atmospheric turbulence. Turbulence occurs when gravity waves break under certain conditions, as detailed in [23]. An analysis of the impact of breaking atmospheric gravity waves on turbulence is presented in [24]. An in-depth investigation of gravity wave breaking and its impact on the middle atmosphere is reported in [25, 26, 27].

2.8 Wind's effect on Atmospheric Gravity Waves

The Gravity Wave characteristics can be interpreted from the Atmospheric Temperature, Density, and Velocity perturbations. In this work, we use temperature perturbation profiles obtained from Rayleigh Lidar system located at Gadanki, Andhra Pradesh, India. The effects background wind on Gravity Wave spectra, and frequency spectra are discussed in detail in the below sections.

2.8.1 Effects of Background Wind On Vertical Wavenumber Spectra

It is mentioned that in the stratosphere a significant departures from true spectral densities at large vertical wavenumbers and is due to background wind changes that increase zonal horizontal phase speeds [28]. The wind shifting theory outlined in this work is considered as initial support for incorporating the effect of winds in the spectral parameterize of Gravity Wave variability within the middle atmosphere.

2.8.2 Effects of Doppler Shifting on the frequency spectra

The articles [29, 19, 30] discusses Doppler Shift resulting from mean background winds. [30] discusses the effects of a nonzero mean flow on observed frequency spectrums of gravity waves with prescribed intrinsic frequencies and wavenumbers. Those two cases establish the limits of possible Doppler-shifting effects when gravity waves propagate in the direction of, or oppose, the mean flow. As a result of Doppler Shifting, horizontal energy density can be significantly increased at higher observed frequencies in the lower and middle atmosphere, vertical energy density can be reduced at higher frequencies, and vertical energy can be transferred significantly to frequencies above buoyancy. It is found that some of the observed frequency spectra support some of the predicted effects.

2.8.3 Propagation of Gravity Waves Through Inhomogeneous Background Winds

A 2D time-dependent simulation of the ray theory is used in [31] to study how small-scale gravity wave packets propagate, refract, and reflect under a time-dependent, vertically, and horizontally inhomogeneous background horizontal wind field. Depending on the parameters of both waves, a static medium-scale wave wind field of sufficient amplitude can channel or critically filter a small-scale wave, or cause significant reflection. A small-scale wave filtering can also be greatly affected by the relative propagation direction between the small-scale and medium-scale waves. Small-scale waves are more likely to become trapped if the phases progress in the same horizontal direction unless reflection occurs first if the phases propagate in opposite horizontal directions. This leads to critical level filtering. For assessing the propagation and dispersion of small-scale waves over large horizontal distances, it is important to account for the time-dependent winds associated with medium-scale-propagating waves.

2.8.4 On the Doppler effect in Gravity Wave Spectra

A statistical wave field affecting gravity waves has been argued as the cause of the invariant shape in the spectrum, according to [32]. Waves with lower wave numbers contribute to tail formation, while those with higher wave numbers are eliminated by dissipative processes at high wave numbers. This layer of waves is Doppler shifted to $m = \infty$, the critical wavelength. It is expected that waves contained in the source spectrum with wavenumbers higher than critical wavenumber (m_c) will be eliminated through absorption in the corresponding critical layer if the mean wind profile is monotonic with altitude from $U(0) = 0$ before going up to $U(z_1) > 0$ under observation. In other words, Doppler shifting is a process of moving from a high to a small scale. Whenever modes are eliminated, the Doppler-shifted power spectrum does not change in shape, but only imposes a wavenumber with high cuts. A detailed explanation of gravity wave spectra and background winds can be found in [33, 34]. As a result, the resulting vertical wavenumber spectrum is profoundly different when monochromatic or continuous source spectrum is used. The continuous case has a spectral tail of 3 due to interference between components, while the monochromatic case has a tail of 1 due to the absence of interference. Although there are previous works that examine the Doppler effect [27, 35], they are based on the vertical wave-action flux conservation. Only permanent gravity-wave sources can benefit from the spectral conservation law in those works.

2.9 Interpreting Atmospheric Gravity Wave Breaking

Detailed explanations of gravity wave breaking from atmospheric density, temperature, and wind velocity perturbations are found in [36, 37]. The detailed procedure for the calculation of Wave breaking characteristics is explained in [38, 39, 40, 41].

2.9.1 Indicators For Wave Breaking

In the presence of gravity wave breaking, the wave phase velocity (C_p) is greater than the background horizontal wind speed ($\bar{\mathrm{U}}$). Gravity waves propagate vertically in the troposphere with increasing amplitudes in temperature perturbations and decreasing densities, leading to convective instability Gravity waves propagate vertically in the troposphere with increasing amplitudes in temperature perturbations and decreasing densities, leading to convective instability ($N^2 < 0$, or $R_i < 0$) or shear instability ($R_i < 0.25$) which occurs at a certain height called breaking level. The Brunt Vaisala Frequency (N) and Richardson

Number (R_i) provide an indication of atmospheric instability, and can be calculated as follows. A Gravity Wave Breaking occurs when the phase velocity (C_p) of the waves exceeds the background horizontal wind ($\bar{\mathbf{U}}$) beyond the gradient. To determine how energy and momentum are transferred between the lower and upper atmospheres, perturbed temperature profiles obtained from Rayleigh lidar observations are examined. By using equation 2.16, the temperature perturbation $T'(z,t)$ is obtained.

$$N^2(z) = \frac{g}{T_0(z,t)}\left[\overline{\frac{dT_0(z,t)}{dz}} + \frac{g}{C_p}\right] \tag{2.22}$$

$$E_p(z) = \frac{1}{2}\left[\frac{g}{N(z)}\right]^2 \left[\overline{\frac{T'_{break}(z,t)}{T_0(z,t)}}\right]^2 \tag{2.23}$$

$$R_i = \frac{N^2}{(\frac{\partial u}{\partial z})^2} \tag{2.24}$$

$$\frac{dT}{dz} = \frac{g}{T_0(z,t)}\left[\overline{\frac{dT'(z,t)}{dz}}\right] \tag{2.25}$$

where $N(z)$ is the Brunt-Vaisala frequency, $T'_{break}(z,t)$ is temperature perturbation at breaking altitude, g is the acceleration due to gravity ($9.8m/s^2$), C_p is the dry air specific heat at constant pressure ($1005JKg^{-1}K^{-1}$). The $\frac{dT'\bar{(}z,t)}{dz}$ value represents the change in rate over the 10km height range of T_0 with respect to altitude z. A list of permutations for N^2 and R_i are given below:

- If $N^2 > 0$, and $R_i < 0.25$, energy is extracted from the mean flow by refracting the internal gravity wave.

- If $N^2 > 0$ and $0.25 < R_i < 2$, the mean flow suffers an energy loss as the internal gravity wave is refracted. As R_i increases, the mean flow loses more energy. As R_i increases, the mean flow loses more energy.

- If $R_i > 2$, ducting and refraction of internal gravity waves are unlikely; the wave quickly dissipates energy.

In order to interpret Wave Breaking, one must identify a large negative temperature gradient ($13km/K$) and strong wind shear ($-40m/s/km$), indicating reduced dynamic and convective stability. Convective stability is reduced when the contours of potential temperature become steep (i.e., become vertical).

2.10 Summary

1. The occurrence of scattering phenomena by molecules and particles is discussed. Rayleigh scattering and Mie scattering process are defined.
2. Relation between back-scattered photons and atmospheric density is explained. The uncertainties in back-scattered photon counts is discussed.
3. The background noise count is defined, and noise due to other sources in the Rayleigh Lidar system are discussed. The Signal to Noise Ratio (SNR) is defined, and statistical techniques to improve the SNR at higher altitudes are discussed.
4. The upper altitude limit of Rayleigh Lidar system, and the characteristics of vertical temperature in the atmosphere has been explained.
5. The uncertainty budget in the atmospheric temperature profiles due to various sources of uncertainties are estimated using synthetically generated photon counts and Monte-Carlo simulations.
6. The effect of background wind on the Gravity Wave vertical wavenumber and criteria to detect wavebreaking are discussed.

Chapter 3

Learning Dictionaries and Deep Learning

3.1 Duality and Optimality

An optimization problem is conceptualized as a dual problem that has a lower bound on its solution compared to a primal (minimization) problem. A duality gap is the result of this difference. For generating the Lagrangian dual problem, the Lagrangian multipliers are applied to the minimization problem in order to add constraints to the objective function. It is necessary to maximize the objective function under the derived constraints in order to find the solution. A Lagrange multiplier produces a resultant, which are primal variables. In the case of two locally convex spaces (X, X^*) and (Y, Y^*) and the function $f : X \to \mathbb{R} \cup \{+\infty\}$, the primal problem can be defined as finding $\hat{x}$ such that $f(\hat{x}) = \inf_{x \in X} f(x)$. It is possible to represent constraints on the objective function (f) by letting $\tilde{f} = f + I_{\text{constraints}}$ where $I_{\text{constraints}}$ is a suitable function on X with a minimum of zero. There is a duality gap when the right-hand side of the inequality is greater than the left-hand side as expressed below

$$\sup_{y^* \in Y^*} -F^*(0, y^*) \leq \inf_{x \in X} F(x, 0), \tag{3.1}$$

where F^* represents the convex conjugate in both variables, and $\sup$ represents the least upper bound. The duality gap has the following properties.

1. The value of duality gap is always greater than or equal to zero.

2. For holding the strong duality case, the duality gap is equal to zero.

3. For the case where the duality gap is strictly positive and greater than zero, the weak duality holds a good fit.

Generally, the objective function is defined as the linear combination of n variables with m constraints, and each constraint places an upper limit on n variables. In a dual problem, the objective function is a linear combination of m values that represent the limits in the m constraints from the primal problem to maximize its value subject to the constraints. An objective function is maximized by a set of n values. A linear combination of m dual variables is constrained by n dual constraints.

The Karush–Kuhn–Tucker (KKT) conditions can be used to find the global maximum of a non-linear problem. The constraints may not necessarily be linear for non-linear problems. Using KKT conditions, it is possible to identify local optima in non-linear problems by.

$$\text{minimize } f_0(x)\text{subject to} \quad f_i(x) \leq 0,\ i \in \{1,\ldots,m\} h_i(x) = 0,\ i \in \{1,\ldots,p\} \tag{3.2}$$

with the domain $\mathcal{D} \subset \mathbb{R}^n$ with non-empty interior, the Lagrangian function

$$\Lambda : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R} \tag{3.3}$$

is defined as

$$\Lambda(x,\lambda,\nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x). \tag{3.4}$$

The Lagrange dual function $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ is defined as 3.5, where the vectors ν and λ correspond to the dual variables or Lagrange multipliers in the equation 3.5.

$$g(\lambda,\nu) = \inf_{x\in\mathcal{D}} \Lambda(x,\lambda,\nu) = \inf_{x\in\mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right). \tag{3.5}$$

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, \quad i = 1,\ldots,m \end{aligned} \tag{3.6}$$

$$\begin{aligned} &\underset{u}{\text{maximize}} \quad \inf_{x} \left(f(x) + \sum_{j=1}^{m} u_j g_j(x) \right) \\ &\text{subject to} \quad u_i \geq 0, \quad i = 1, \ldots, m \end{aligned} \tag{3.7}$$

For any value of $\lambda \geq 0$ and any ν we have $g(\lambda, \nu) \leq p^*$, where p^* is the optimal value of the initial problem. With inequality constraints, a convex minimization problem (3.6) leads to a Lagrangian dual problem given by equation 3.7. The infimum of the Lagrangian dual(3.7) occurs when its gradient equals zero, provided the functions f and $g_1, \ldots, g_m$ are continuously differentiable and is expressed below as

$$\begin{aligned} &\underset{x,u}{\text{maximize}} \quad f(x) + \sum_{j=1}^{m} u_j g_j(x) \\ &\text{subject to} \quad \nabla f(x) + \sum_{j=1}^{m} u_j \, \nabla g_j(x) = 0 \\ &\qquad\qquad\quad u_i \geq 0, \quad i = 1, \ldots, m \end{aligned} \tag{3.8}$$

The equality constraint $\nabla f(x) + \sum_{j=1}^{m} u_j \, \nabla g_j(x)$ is typically non-convex when it comes to dual optimization problems.

3.1.1 Lagranges theorem

Using Lagrangian Multipliers, one can locate the maximum and minimum local maxima of a constraint-based optimization problem. To get the maximum or the minimum of a function $f(x)$ subjected to the equality constraint $g(x) = 0$:

1. Find the stationary points of the Lagrangian function($\mathcal{L}(x, \lambda) = f(x) - \lambda g(x)$) and the Lagrange multiplier λ.
2. The solution to the constrained optimization is always a saddle point of the Lagrangian function.
3. The saddle points can be recognized from definiteness of the Hessian Marix.

A benefit of Lagrangian Multipliers are that they can be used to solve optimization problems without explicitly parameterizing constraints. In this way, Lagrange multipliers can be used to solve constrained optimization problems that are challenging. The KKT conditions generalize the Lagrange multipliers.

3.1.1.1 Statement

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be the objective function, $g\colon \mathbb{R}^n \to \mathbb{R}^c$ be the constraints function, such that their first derivative exists. The optimal solution to (x^*) is given by solving the below equation, such that rank $g(x^*) = c < n$:

$$\text{maximize } f(x)\text{subject to: } g(x) = 0 \tag{3.9}$$

(Here $Dg(x^*)$ denotes the matrix of partial derivatives, $\left[\dfrac{\partial g_j}{\partial x_k}\right]$. Then there exist unique Lagrange multipliers $\lambda^* \in \mathbb{R}^c$ such that $Df(x^*) = \lambda^{*T} Dg(x^*)$.

3.1.2 Single Constraint

It is possible to express the gradient of a function under constraints by combining the gradients of the constraints with the Lagrange multipliers as coefficients, if the function has local maxima or local minima. Think about the optimization problem in the case of only one constraint and two choices. If $f(x_0, y_0)$ is a maximum of $f(x, y)$ for the original constrained problem and $\nabla g(x_0, y_0) \neq 0$, then there exists λ_0 such that (x_0, y_0, λ_0) is a stationary point for the Lagrange function. It is not possible to solve the original problem with all stationary points, as Lagrange multipliers yield only necessary, but not sufficient, conditions for optimality. A particular solution can occur if sufficient conditions are met, but if a particular solution exists, it is a maximum or minimum.

$$\nabla_{x,y} f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right), \qquad \nabla_{x,y} g = \left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}\right) \tag{3.10}$$

Thus we want solution points (x, y) where $g(x, y) = c$ and $\nabla_{x,y} f = \lambda \, \nabla_{x,y} g$, for some λ where the gradients are defined in equation 3.10. Auxiliary functions are introduced into the equation to make these conditions explicit. Refer to equation 3.11.

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda g(x, y), \tag{3.11}$$

and solve $\nabla_{x,y,\lambda} \mathcal{L}(x, y, \lambda) = 0$. To summarize

$$\nabla_{x,y,\lambda} \mathcal{L}(x, y, \lambda) = 0 \iff \begin{cases} \nabla_{x,y} f(x, y) = \lambda \, \nabla_{x,y} g(x, y) \\ g(x, y) = 0 \end{cases} \tag{3.12}$$

It is easily generalized to functions on n variables$\nabla_{x_1,\ldots,x_n,\lambda}\mathcal{L}(x_1,\ldots,x_n,\lambda)=0$ which amounts to solving n + 1 equations in n + 1 unknowns. The constrained extrema of f are critical points of the Lagrangian $\mathcal{L}$, but they are not necessarily local extrema of $\mathcal{L}$.

A numerical optimization problem is also posed by the fact that Lagrangian solutions are not necessarily extreme. As illustrated in the numerical optimization example, the zeros of the gradient are necessarily local minima, so this can be addressed by computing the magnitude of the gradient.

3.1.3 Multiple Constraint

By using a similar argument, Lagrange multipliers can be applied to problems with multiple constraints. We follow the same procedure as a mentioned for single constraint, extending to multiple constraints.

$$\nabla f(\mathbf{x})=\sum_{k=1}^{M}\lambda_k\,\nabla g_k(\mathbf{x}) \quad\Longleftrightarrow\quad \nabla f(\mathbf{x})-\sum_{k=1}^{M}\lambda_k\nabla g_k(\mathbf{x})=0 \tag{3.13}$$

$$\mathcal{L}\left(x_1,\ldots,x_n,\lambda_1,\ldots,\lambda_M\right)=f\left(x_1,\ldots,x_n\right)-\sum_{k=1}^{M}\lambda_k g_k\left(x_1,\ldots,x_n\right) \tag{3.14}$$

and solve

$$\nabla_{x_1,\ldots,x_n,\lambda_1,\ldots,\lambda_M}\mathcal{L}(x_1,\ldots,x_n,\lambda_1,\ldots,\lambda_M)=0 \iff \begin{cases}\nabla f(\mathbf{x})-\sum_{k=1}^{M}\lambda_k\,\nabla g_k(\mathbf{x})=0\\ g_1(\mathbf{x})=\cdots=g_M(\mathbf{x})=0\end{cases} \tag{3.15}$$

We try to find out the individual scalars $\lambda_1,\lambda_2,....\lambda_M$ such that equation 3.13 will get satisfied. Lagrange multipliers are these scalars. As a result, we now have **M** multiple constraints, each having its own set of constraints.

The auxiliary function and solution can be defined as expressed in equation 3.14 and 3.15 respectively. In other words, it involves solving $n+M$ equations in $n+M$ unknowns. When there are multiple constraints, the constraint gradients at the relevant point must be linearly independent.

3.2 Sparse Representations

The Sparse representations represent the signal with few samples without loss of significant information. By switching from l_2 regularization to l_1 regularization, the Sparser solution is promoted. consider the following optimization problem below

$$\min_{\mathbf{x}} ||\mathbf{x}||_{\mathbf{q}}^{\mathbf{q}} \text{ subject to } ||\mathbf{x}||_{\mathbf{p}}^{\mathbf{p}} = \mathbf{1}; \forall \mathbf{q} < \mathbf{p} \tag{3.16}$$

As a result of the optimization problem above (3.16), $x_k^{p-q} = constant$ for all k indicating the same value for all entries other than zero. The constraint $||x||_p^p = 1$ leads to $x = a^{\frac{-1}{p}}$ and the l_p norm is determined by $||x||_q^q = a^{\frac{1-q}{p}}, q < p$. The interpretation of the above result is that for any l_p and l_q norms, the length of l_p norm is shortest in the l_q norm indicating the sparsest solution possible. We find the locus of 'q-ball' in the same space and compare it with the 'p-ball and all entries except one pass through the intersection along the axes. This is demonstrated in Fig. 3.1

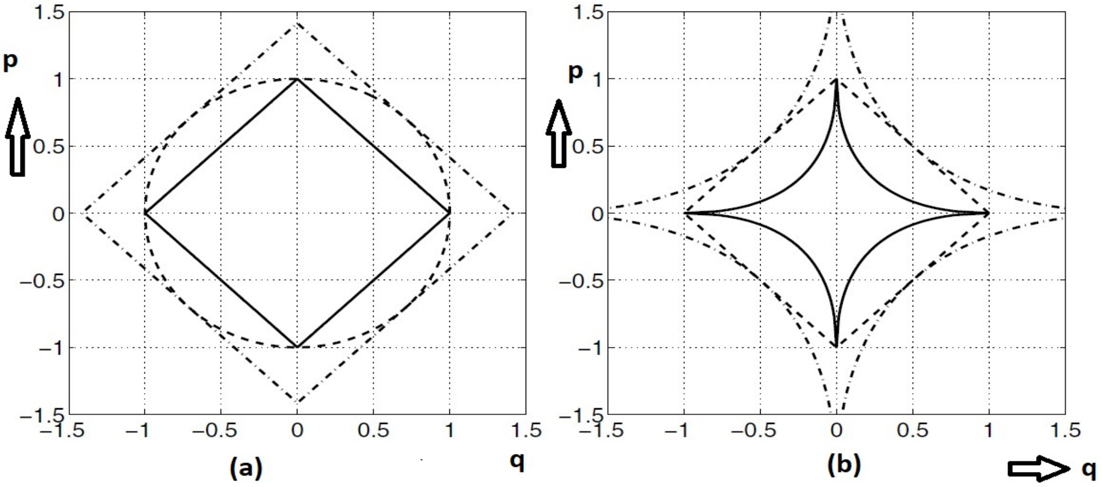


Figure 3.1: Illustrating that a unit-length l_p norm vector(dashed) becomes shortest in $l_q; p > q$ (solid) at its sparest possible.(a) demonstrated for p = 2 and q = 1, and (b) for p = 1 and q = 0.5. Those dashed-dotted lines indicate a non-sparse outcome.

Returning to our basic problem (P_0) in the below equation

$$P_0 : \min_{\mathbf{X}} ||\mathbf{X}||_{\mathbf{0}} \text{ subject to } \mathbf{Y} = \mathbf{DX} \tag{3.17}$$

For the undetermined system of linear equations, $Y = DX$, where $D \in \mathbb{R}^{n \times m}$, $n < m$, we have the following problems in front of us.

1. How to claim the sparsest solution is unique?
2. Is it possible to test the optimality of a candidate solution?

3.2.1 Uncertainty

The uncertainty is quantified in terms of mutual coherence($\mu(A)$).

$$\mu(A) = \max_{1 \le i \ne j \le m} \left| a_i^H a_j \right| \tag{3.18}$$

A lower bound is

$$\mu(A) \ge \sqrt{\frac{m-d}{d(m-1)}} \tag{3.19}$$

Cross-correlations between the columns of a matrix 'A' determine its Mutual Coherence. Formally, let $a_1, \ldots, a_m \in \mathbb{C}^d$represent the columns of the matrix 'A', whose columns are normalized such that $a_i^H a_i = 1 . a_i^H a_i = 1$. As a result, 'A's mutual coherence is defined as equation 3.18. The concept of the mutual coherence [42] is reintroduced to the context of sparse representations by [43, 44, 45, 46]. An algorithm that accurately identifies the sparse signal representation is measured in this way as a measure of the effectiveness of a sub-optimal algorithm, such as matching pursuit or basis pursuit.

3.2.2 Uniqueness

To check the uniqueness of the sparse solution, a metric called '*Spark*' is defined by Donoho and Elad in [45, 47]. A spark identifies a matrix's null space ($\mathbf{A}$) using the l_0 norm. The Spark of a matrix($\mathbf{A}$) indicates how many columns in $\mathbf{A}$ that are linearly dependent. For the defined linear system of equations$\mathbf{Y} = \mathbf{AX}$, has a solution $\mathbf{X}$ satisfying $||\mathbf{x}||_\mathbf{0} < \mathbf{spark}(\mathbf{A})/\mathbf{2}$, which is necessarily the sparsest solution.

$$Spark(\mathbf{A}) \ge \frac{\mathbf{1}}{\mu(\mathbf{A})} \tag{3.20}$$

$$||\mathbf{x}||_\mathbf{0} < \frac{\mathbf{1}}{\mathbf{2}}(\mathbf{1} + \frac{\mathbf{1}}{\mu(\mathbf{A})}) \tag{3.21}$$

The relation between the uniqueness and uncertainty is given by (3.20) and also if the system of linear equations $\mathbf{Y} = \mathbf{A}\mathbf{x}$ has a solution $\mathbf{x}$ obeying (3.21) is the necessarily the sparsest possible.

3.2.3 Sparse Coding Techniques

Essentially, sparse coding involves a linear combination of basic elements that form a dictionary and represent input data as a linear combination of the elements themselves. The elements are called atoms.

3.2.3.1 Efficient Convolutional Sparse Coding

One of the most popular sparse coding technique is the Basis Pursuit Denoising (BPDN) and the optimization problem is given by

$$\arg\min_{\mathbf{x}} \frac{1}{2}\|D\mathbf{x} - \mathbf{s}\|_2^2 + \lambda\|\mathbf{x}\|_1 \tag{3.22}$$

The optimization problem [48, 49] for the convolution sparse coding is given by (3.23).

$$\arg\min_{\{\mathbf{x}_m\}} \frac{1}{2}\left\|\sum_m \mathbf{d}_m * \mathbf{x}_m - \mathbf{s}\right\|_2^2 + \lambda\sum_m \|\mathbf{x}_m\|_1 \tag{3.23}$$

where D represent a dictionary matrix, x represent the sparse representation, λ represent regularization parameter,d_m is a set of M dictionary filters, $*$ denotes convolution, and x_m denotes coefficient maps, whose size is equal to as s.

$$\arg\min_{\{\mathbf{x}_m\},\{\mathbf{y}_m\}} \frac{1}{2}\left\|\sum_m \mathbf{d}_m * \mathbf{x}_m - \mathbf{s}\right\|_2^2 + \lambda\sum_m \|\mathbf{y}_m\|_1 \qquad \textbf{such that } \mathbf{x}_m - \mathbf{y}_m = 0\ \forall m \tag{3.24}$$

The optimization problem (3.23) can also be expressed as equation 3.24 and is solved by Alternating Direction Method of Multipliers (ADMM).

3.2.3.2 Laplacian Sparse Coding

Due to the over-complete code and the individual coding process, the locality and similarity information are lost. In order to preserve the locality and similarity characteristics, the Laplacian Sparse Coding [50] technique is used. Sparse coding encodes the features independently. A mathematical representation of Laplacian Sparse Coding is given by

$$\underset{v_1,\ldots,v_n}{min} \sum_i \|x_i - Uv_i\|_F^2 + \lambda \sum_i \|v_i\|_1 + \frac{\beta}{2} \sum_{ij} \|v_i - v_j\|^2 W_{ij} \tag{3.25}$$

The Laplacian matrix($\mathbf{L}$) is defined as $\mathbf{D} - \mathbf{W}$, where $\mathbf{W}$ represent the similarity features, and $\mathbf{V} = [v_1, v_2, ..., v_n]$.

$$\underset{U,V}{min} \|X - UV\|_F^2 + \lambda \sum_i \|v_i\|_1 + \beta tr(VLV^T) \tag{3.26}$$

subject to $||u_m||^2 = 1$ and avoiding scaling is done by using this constraint.

3.2.4 Least Absolute Shrinkage and Selection Operator (LASSO)

This method formulates the optimization problem as follows:

$$\min_{r \in \mathbb{R}^n} \{ \ \|r\|_1 \} \text{ subject to } \|X - \mathbf{D}R\|_F^2 < \epsilon \tag{3.27}$$

where ϵ represents the LASSO reconstruction error [51].

Assuming a constraint on the L1-norm of the solution vector, it minimizes the least square error to determine r_i:

$$\min_{r \in \mathbb{R}^n} \ \frac{1}{2} \ \|X - \mathbf{D}r\|_F^2 + \lambda \ \|r\|_1 \tag{3.28}$$

where $\lambda > 0$ determines the accuracy of sparse the reconstruction . As a result, the global optimal solution is obtained.

3.3 Dictionary Learning

Given the input dataset $X = [x_1, ..., x_K], x_i \in \mathbb{R}^d$ we wish to find a dictionary

$$\mathbf{D} \in \mathbb{R}^{d \times n} : D = [d_1, ..., d_n] \tag{3.29}$$

and a representation $X = [x_1, ..., x_K], r_i \in \mathbb{R}^n$ such that both $\|Y - \mathbf{D}X\|_F^2$ are minimized and the representations x_i are sparse enough. The optimization problem can be formulated as expressed as below as

$$\underset{\mathbf{D}\in\mathcal{C},x_i\in\mathbb{R}^n}{\operatorname{argmin}} \sum_{i=1}^{K} \|y_i - \mathbf{D}x_i\|_2^2 + \lambda\|r_i\|_0 \tag{3.30}$$

where

$$\mathcal{C} \equiv \{\mathbf{D} \in \mathbb{R}^{d\times n} : \|d_i\|_2 \leq 1 \ \forall i = 1, ..., n\}\lambda > 0 \tag{3.31}$$

$\mathcal{C}$ is required to control $\mathbf{D}$ in order to prevent dictionary atoms from reaching arbitrary values, yet allow them to reach arbitrarily low values at x_i, and x_i is the photon count value at an altitude. The value of λ determines how sparsity is traded off against minimization errors.

3.3.1 Properties of the Dictionary

An undercomplete dictionary represents inputs on a lower-dimensional space, similar to Principal Component Analysis. Data analysis, classification, and pattern recognition can be addressed using Dictionary Learning for dimensionality reduction. A richer representation of data is possible with overcomplete dictionaries. In the presence of sparse representation conditions, an overcomplete dictionary can be used to represent a signal sparsely, and can also be used as a transform matrix (wavelets, Fourier). Compared to fixed predefined transform matrices, dictionaries learned from data can provide sparser solutions.

The optimization problem described in equation 3.30 is convex in one variable if the other is fixed. Finding the optimal sparse coding X for a dictionary $\mathbf{D}$ is known as a sparse approximation (or sometimes just a sparse coding problem). Several algorithms have been proposed to solve the optimization problem defined in equation 3.30.

3.3.2 Method of optimal directions (MOD)

The idea of method of optimal directions (or MOD) [52, 53] is to solve the minimization problem under the condition that the representation vector is only composed of 'T' non-zero components:

$$\min_{\mathbf{D},X}\{\|Y - \mathbf{D}X\|_F^2\} \text{ s.t. } \forall i \ \|x_i\|_0 \leq T \tag{3.32}$$

where, F denotes the Frobenius norm. By alternating between sparse coding through a method such as matching pursuit and updating the dictionary by calculating a Moore-Penrose pseudo-inverse, the Dictionary update problem is solved by $\mathbf{D} = YX^{+}$ where X^{+} is a Moore-Penrose pseudo-inverse. Following this update, $\mathbf{D}$ is re-normalized according to the constraints, and a new sparse code is obtained. These steps are repeated until a satisfactory amount of convergence is reached (or until there is a satisfactory amount of residue). It took only a few iterations for MOD to converge on low-dimensional input data X. A drawback of the MOD is the difficulty in computing the pseudo-inverse when the data is high dimensional. As a result of this shortcoming, other algorithms were developed.

3.3.3 K-Singular Value Decomposition

The K-Singular Value Decomposition (KSVD) [54, 55, 56, 57] performs Singular Value Decomposition (SVD) during dictionary update, and is basically a generalization of K-means. This ensures that each element in y_i is encoded by a linear combination of no more than T_0 elements as do MOD elements as expressed below

$$\min_{\mathbf{D},X}\{\|Y-\mathbf{D}X\|_F^2\} \text{ s.t. } \forall i \ \|x_i\|_0 \leq T_0 \tag{3.33}$$

As a first step in the algorithm, we fix the Dictionary (**D**), then find the best X under the above constraint (using Orthogonal Matching Pursuit [58]). We iteratively update the atoms in Dictionary $\mathbf{D}$ and represent the entire process mathematically as

$$\|Y-\mathbf{D}X\|_F^2 = \|Y-\sum_{i=1}^{K} d_i x_T^i\|_F^2 = \|E_k - d_k x_T^k\|_F^2 \tag{3.34}$$

As part of the algorithm, the residual matrix must be approximated by rank-1, updated with d_k. This method also has same drawback of being efficient for signals with relatively low dimensional and has the problem of being stuck at local minima.

3.3.4 Stochastic Gradient Descent

In Stochastic Gradient Descent (SGD) [59, 60] method, the iterative projection with first order Stochastic Gradient Descent is used to update the dictionary and to project it on the constraint set $\mathcal{C}$. The mathematical representation of method at i^{th} iteration is expressed by using

$$\mathbf{D}_i = \text{proj}_{\mathcal{C}} \left\{ \mathbf{D}_{i-1} - \delta_i \nabla_{\mathbf{D}} \sum_{i \in S} \|y_i - \mathbf{D}x_i\|_2^2 + \lambda \|x_i\|_1 \right\} \tag{3.35}$$

where S is a random subset of $\{1...K\}$ and δ_i is a gradient step.

3.3.5 Lagrange Dual Method

As an alternative, dictionary learning can also be viewed in terms of Lagrangian multipliers problem and expressed below as

$$\mathcal{L}(\mathbf{D}, \Lambda) = \text{tr}\left((Y - \mathbf{D}X)^T (Y - \mathbf{D}X)\right) + \sum_{j=1}^{n} \lambda_j \left(\sum_{i=1}^{d} \mathbf{D}_{ij}^2 - c \right) \tag{3.36}$$

where c constrains the norm of the atoms and λ_i comprise the diagonal matrix Λ, **R** represent the sparse representation. After minimization, the Lagrange dual is expressed as

$$\mathbf{D} : \mathcal{D}(\Lambda) = \min_{\mathbf{D}} \mathcal{L}(\mathbf{D}, \Lambda) = \text{tr}(X^T X - XR^T (RR^T + \Lambda)^{-1} (XR^T)^T - c\Lambda) \tag{3.37}$$

$$\mathbf{D}\mathbf{D}^T = (RR^T + \Lambda)^{-1} (XR^T)^T \tag{3.38}$$

By solving the equation 3.37 with conjugate gradients, we get **D** as expressed in equation 3.38. Due to the fact that the number of dual variables n in this problem is far less than the number of variables in the primary problem, it is less computationally challenging to solve this problem (3.37).

3.3.6 Regularization and Dictionary Learning

There is a common problem with dictionary learning for sparse representations. Regularized versions of K-SVD and other algorithms were used to regularize[61] the error criterion. This study used either representation norms or atom coherence measures. The norm of the representations or the atom coherence measure were used here. In order to cure the problem, the objective is to change the below expression

$$f_\mu(D,\ X) = \|\mathrm{Y} - DX\|_F^2 + \mu \|X\|_F^2 \tag{3.39}$$

when $\mu > 0$ reduced the possibility of atoms with similar directions by encouraging small magnitudes of representation coefficients. In cases of rank deficiency, this regularization is typically applied to least-squares problems.

$$f_\gamma(D,\ X) = \|\mathrm{Y} - DX\|_F^2 + \gamma\|D^T D - I\|_F^2 \tag{3.40}$$

In contrast, an alternative approach (3.39) decreases the mutual coherence among atoms, thus preventing the appearance of groups of atoms almost linearly dependent, and then the optimization problem takes the form of equation 3.40.

3.4 Dictionary Learning: Local feature Identifiability

Given a sparse linear combination of the atoms of a true Dictionary$\mathbf{D_0}$, we explain how to recover $\mathbf{D_0}$ through Dictionary Learning. We use the property that if D_0 is locally identifiable then the Dictionary Learning objective function is local minimum. We use l_1-minimization Dictionary Learning that retrieves the Dictionary through minimization of the average l_1-norm of the linear coefficients.

With the μ-coherent reference Dictionary, i.e., a dictionary with a maximum pairwise column inner-product of $\mu \in [0; 1)$, there is still local identification even for random linear coefficient vectors with $O(\mu^{-2})$ non-zeros on average. Furthermore, if the sparsity level exceeds $O(\mu^{-2})$, there is no local identification in the reference Dictionary.

However, we have discussed Dictionary Learning and two of the solutions namely MOD, and KSVD for optimizing Dictionary Learning. This methodology even has restrictions like speed and memory used, restriction to low dimensions, operating on a Single-Scale, and lack of Invariances. The Dictionary Learning methodology. Dictionary Learning has dealt with Speed and Memory used to some extent.

3.4.1 l_0, l_1, and l_p minimization

Underdetermined linear equations generally have more unknowns than equations, and there are an infinite number of possible solutions. For the equation system $\mathbf{Y} = D\mathbf{X}$, we want to find a solution for $\mathbf{X}$. It is necessary to impose additional constraints (such as smoothness) to find a unique solution to such a system. The constraint of sparsity limits compressed sensing solutions to those with a small number of nonzero coefficients. Under-determined linear equations do not always have sparse solutions. A compressed sensing framework, however, allows for the recovery of a unique sparse solution to an under-determined system.

3.4.1.1 Efficiency of l_1-minimization over l_0-minimization problem

The l_0 minimization problem has the following form:

$$\min ||\mathbf{X}||_0 : \mathbf{Y=DX} \tag{3.41}$$

The optimization problem is solved using continuous approximation, Orthogonal Matching Pursuit (OMP) [58, 62, 63, 64], thresholding methods [65, 66, 67, 68], l_1 method and weighted l_1 method [69].

The l_1-norm is the convex envelope of the $||\mathbf{X}||_0$ over the region $X : ||X||_\infty \leq 1$ and the minimization problem is given by

$$\min ||\mathbf{X}||_1 : \mathbf{Y=DX} \tag{3.42}$$

The l_0 and l_1 minimization problems are equivalent if l_0 corresponds to a unique solution to the l_1-problem. The l_0 and l_1 minimization [70] problems exhibit strong similarities when a unique answer to l_0 corresponds to the l_1-problem. At the moment, in the present work, strong equivalence between l_0 and l_1 is considered. The strong equivalence is quantified through Mutual Coherence (MC) Condition, Restricted Isometric Property (RIP), Null Space Property (NSP), and Range Space Property (RSP). In many situations, the strong equivalence conditions can only partially explain about the l_1-method. Based on probabilistic analysis, it is found that l_1 minimization is more suitable for linear systems with sparse solutions than indicated by equivalence criteria. The l_0 and l_1 problems are said to be equivalent if and only if the RSP is satisfied with the linear system's sparsest solution.

3.5 Multi-Resolution Analysis

By adjusting the resolution, a particular task's information can be retrieved. The application of Multi-Resolution Analysis to the Atmospheric Gravity Waves is first reported in [71]. MRA breaks a signal up into individual components, that produce the original signal by adding the individual components together. MRAs are associated with Wavelets, which decompose data variability into physically meaningful and interpretable elements.

An approximation of function ‘f’ with a resolution of 2^{-j} is represented as a discrete grid of samples with local averages of ‘f’ proportional to 2^j. Approximations at multiple resolutions are therefore composed of embedded grids. The MRA of the Lebesgue space $L^2(\mathbb{R}))$ is composed of discrete subspaces satisfying certain self-similarity relations in time and space, as well as completeness and regularity.

Multi-Resolution Approximation Errors We use Wavelets as the basis of orthogonal complements of multi-resolution approximation spaces. Errors in projection on multi-resolution approximation bases are dependent on Wavelet coefficient decay rates. To reduce the approximation error, we use the folded Wavelets. The vanishing moment of folded wavelets is sufficient for obtaining low approximation errors. Wavelet bases provide a better representation of uniform regular signals. There exists an association between differentiability in the Wavelet linear approximation error decay and Soblev regularity ie the Wavelet function $\psi \in \mathbf{L^2[0,1]}$ is in Soblev space $\mathbf{W^S[0,1]}$, where s represent the Soblev exponent. If the Wavelet function satisfies the Soblev regularity, the decay of linear approximation is fast and the approximation error is less.

3.6 Criteria For Selecting Neural Network Architectures

When it comes to nonlinear activation functions, even a small change in the parameter increases the computational cost. Convolution Neural Networks address this problem effectively by exploiting the linear dependency of linear features, which is not possible with Neural Networks used for regression. Following are guidelines for selecting the number of layers and hidden neurons in neural networks.

1. **Degenerate Problems and Degenerate Solutions:** Degenerate problems are those in which $y = f(x)$. The problems require a degenerate solution that copies the input to the output as unmodified.

2. **Neural Networks For Linear Separation:** For solving linear classification and regression problems, perceptron can be used. When linearly separable problems exist, the dimension of the Neural Network is of $|x|$ input nodes and $|y|$ output nodes. If $|x|$ comprises non-linearly independent features, dimensionality reduction techniques can be applied to transform the input into a vector with linearly independent components, which is called the Kernel function. As a result, we will need a size of $|e|$ for the input, where e represents the eigenvectors of x. Thus, if a problem can be linearly separated, then the correct number and size of hidden layers is zero.

3. **Neural Networks For Non-linear Separation:** Non-linear separable problems are problems for those whose solution is not a hyper-plane. Using one hidden layer, a neural network can approximate a continuously differentiable function to arbitrary

precision if it consists of a continuously differentiable function. According to this, there is one hidden layer if the problem is continuously differentiable.

4. **Neural Networks For Arbitrary Boundaries:**This is suitable for cases where the decision boundary is multi-discontinuous. We need to increment the number of hidden layers by 1 to account for the extra complexity of the problem. One continuous component of the decision boundary is learned by each neuron in the second hidden layer. Functions that combine the output layer's weight matrix into a single boundary are found in the interaction with the output layer's weight matrix.

5. **Neural Networks For Abstraction:** An abstraction level, higher than problems is also applicable to problems. A higher level of abstraction is required when selecting nodes within a layer rather than patterns in the input. Therefore, more than two hidden layers are required.

3.7 Heuristics

The Heuristics act as guidelines to correct the dimensionality of the Neural Network. The guidelines will help to choose the number of hidden layers and their sizes

1. **Building incrementally complex systems:** We need to develop complex network configurations only if the simpler configurations on a network are insufficient. First, we have to go with the linear approach, if it fails we have to move to the non-linear approach.

2. **Prefer Increasing Sizes over Layers:** Prefer increasing the size of the layer, rather than increasing the number of hidden layers. An inefficiency in learning a decision function is the result of a neural network with a given number of hidden layers. It is better to avoid increasing the number of hidden layers if we have reason to believe that the complexity of the problem is appropriate for the number of layers we added.

3. **Processing the data better:** The failure of training represents that the data we are using requires additional processing steps. Consequently, when we train, the problem is not the number of hidden layers, but rather optimizing their parameters. Preprocessing steps may include dimensionality reduction techniques to extract independent features, and batch normalization to ease the training difficulty, or we could add a dropout layer to reduce overfitting.

3.7.1 Choosing the Hyperparameters

Grid search is used to optimize Hyperparameters. The best suitable activation function for hidden neurons is Rectified Linear Unit (ReLU) which reduces the vanishing gradient problem. A Neural Network's depth and width are determined by the number of hidden layers and neurons per layer. Hyperparameters are the configurations related to weights, biases, drop rate, and learning rate.

Overfitting is the most commonly observed problem while learning. The model is overfit because it produces low errors when responding to seen data during the training phase, creating large errors with test data or validation data sets.

$$BN(u) = \frac{\gamma(u-\mu)}{\sigma} + \beta \tag{3.43}$$

where γ and β are the learned paameters. Dropout and Batch Normalizations(BN) 3.43 can be used to suppress the problem. By using equation 3.43, we standardize the hidden layers with a mean(μ) and standard deviation(σ). BN is applied before the activation function and speeds up convergence. Dropout removes neurons from the network by multiplying their activation by zero (called Drop rate) at probability(r). By combining multiple independent models, Dropout works similarly to cascaded combinations of networks. In order to compensate for the decrease in model capacity, the depth or width of the network is increased.

The network parameters are always selected in a way along the steepest descent direction in order to minimize errors. The selection of parameters is governed by the learning rate. The selection of a low learning rate results in time-consuming and poor training phases and choosing a high learning rate lead to divergence. Mini-batch methods are employed to avoid this trade-off, since this reduces training time by updating neual network parameters. A larger kernel size in the Convolutional Neural Network can capture steeper wave motions and a small kernel size is used for shallow variations.

3.8 Guidelines for Training Neural Networks

The following below are the rules of thumb used to train the Neural Network.

1. **Number of Layers:** Begin with two hidden layers (excluding the last layer).
2. **Number of nodes (size) of intermediate layers:** The number of nodes should follow the geometric progression. About half of the input data features should be in the first

layer. In the next layer, the size is half what it was in the previous one.

3. **Number of nodes (size) of output layer for Classification:** The output layer is one if the classification is binary. It is the class size that determines how big a multi-class classifier.
4. **Size of output layer for regression:** One if there is only one response. It is the number of responses determining the size of the multi-response regression.
5. **Activation for intermediate layers:** ReLU is taken into account as activation.
6. **Activation for output layer:** For binary classification, use sigmoid, for multiclass classification, use softmax and for regression, use linear. It is recommended that the last layer of an autoencoder is linear if the input data is continuous, otherwise, it should be sigmoid or softmax if the input data is binary or multi-level categorical.
7. **Dropout Layers:** Do not include the Input layer in Dropout Layers (unless defining it separately). Dropout rate should be set to 0.5. To reduce the dropout rate to less than 0.5, increase the layer size instead of reducing the dropout rate.
8. **Data preprocessing:** Prior to using the data for training the model, perform data scaling. Use Minmax scaler. The standard scaling in the same library will work if this does not work well. Regression requires scaling.
9. **Split data to train, valid, test:**Split the data into train, test, and validation data.
10. **Class weights:** Include class weights in the model to balance loss in unbalanced data. Binary classifiers should have weights of: $0 \leftarrow \frac{\text{number of 1's}}{\text{data size}}$, $1 \leftarrow \frac{\text{number of 0's}}{\text{data size}}$. If the data is extremely unbalanced (rare events), class weight may not be effective.
11. **Optimizer:** Use ‘Adam’ algorithm, with default learning rate.
12. **Loss in classification:** Entropy may be used to measure loss in binary classification, entropy may be used to measure loss in multi-class classification. If the labels are hot-encoded, then the entropy may be sparse for integer labels.
13. **Loss in regression:** Use Mean Square Error (MSE).
14. **Metrics for Classification:** Use accuracy that shows the percent of correct classifications.

15. **Metric for Regression:** Use Root Mean Square Error (RMSE).
16. **Epochs:** Generally consider 20 epochs and check the accuracy and loss of the model after training. If the improvement in accuracy is not observed within 20 epochs, make the epoch 100.
17. **Batch size:** Select the batch size from the geometric progression of 2. Always choose a higher value for unbalanced data.
18. **Oscillating loss:** There may be a convergence problem if there is oscillating loss during training. Changing the batch size or reducing the learning rate may help.
19. **Custom Metric:** False Positive Rate is an important metric for unbalanced binary classification.

3.9 Performance Metrics

The Minimum-Mean Square Error (MMSE), Peak Error (PE), and the Reconstruction Error (RE) serve as performance indicators for Wavelets. The MMSE is the measure of the estimator quality. The MMSE can also be given as the posterior mean of the parameters to be estimated. The Squared Prediction Error (SPE) metric and T^2-metric are used for detection of gravity wave breaking events.

1. **Minimum Mean Sqaure Error (MMSE)** The term MMSE refers to the estimation of the quadratic loss function. Ideally, for a good estimator, the MMSE approches to zero.

   $$MMSE = \frac{1}{N}\sum_{n=1}^{N}\left[\mathbf{y} - \hat{\mathbf{y}}\right]^2 \tag{3.44}$$

2. **Peak Error (PE)** The Peak Error (PE) represents the maximum distance between the original and reconstructed signals. For the best method, the PE should be as small as possible.

   $$PE = max_{i=1}^{N}\left[\mathbf{y} - \hat{\mathbf{y}}\right] \tag{3.45}$$

   where $\mathbf{y}$ is the actual value, $\hat{\mathbf{y}}$ is the measured value.

3. **Reconstruction Error (RE)**

   $$RE = ||\mathbf{y} - \mathbf{D_T x_N}||_2^2 \tag{3.46}$$

where $\mathbf{y}$ is the actual value, $\mathbf{x}_N$ indicates the percentage of N coefficients in the transform domain. Sparse representation gets better with faster decay of reconstruction error. For an ideal method, the reconstruction error should fall at a faster rate when compared to the other methods.

4. **Squared Prediction Error**

   SPE is defined as

   $$SPE = ||Y - \hat{Y}||_2^2 \tag{3.47}$$

   in which Y is the input sequence and $\hat{Y}$ is the predicted output. A feature is detected when the SPE statistic is higher than the threshold.

5. **T^2-Metric**

   The T^2 Statistic at k^{th} bin is defined as

   $$T^2 = h_k^T \Theta^{-1} h_k; \Theta = \frac{H^T H}{(n-1)} \tag{3.48}$$

   where $H \in R^{m \times n}$ encoded features, and Φ is their covariance matrix.

3.9.1 Kullback-Leibler Divergence

In the case of discrete probability distributions P and Q in the same probability space, $\mathcal{X}$, the relative entropy (D_{KL}) also called Kullback-Leibler divergence from Q to P is defined as

$$D_{\text{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right) \tag{3.49}$$

which is equivalent to

$$D_{\text{KL}}(P \parallel Q) = -\sum_{x \in \mathcal{X}} P(x) \log\left(\frac{Q(x)}{P(x)}\right) \tag{3.50}$$

As a result, its probability difference is expected to be the logarithm of the difference between probabilities P and Q, with the expectation derived from probabilty P. In order to define relative entropy, for all x, $Q(x) = 0$ implies $P(x) = 0$ (absolute continuity). Whenever $P(x)$ is zero, the corresponding term's contribution will also be zero.

$\lim_{x \to 0^+} x\log(x) = 0$. For distributions P and Q involving continuous random variables, relative entropy is defined as

$$D_{\text{KL}}(P \parallel Q) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx \tag{3.51}$$

where p and q denote the probability densities of P and Q.

The divergence of P from Q or the divergence of Q to P, which describes the asymmetry in Bayesian inference, which involves a prior updating Q followed by a posterior updating P as expressed in equation 3.50.

3.9.1.1 Interpretations

Entropy relative to Q to P is often denoted $D_{\text{KL}}(P \parallel Q)$. As an analogy, KL divergence can be equated to Relative Entropy of P relative to Q in the information theory. It can be expressed in Bayesian inference terms, $D_{\text{KL}}(P \parallel Q)$, that is, how much information is lost when Q is used to approximate P. To find the distribution Q nearest to P, minimize KL divergence and calculate information projections.

3.9.1.2 Properties

1. The relative entropy is always positive, $D_{\text{KL}}(P \parallel Q) \geq 0$, referred to as Gibbs' inequality, with $D_{\text{KL}}(P \parallel Q)$ being zero if and only if $P = Q$ almost everywhere. The entropy $\mathrm{H}(P)$ sets a minimum value for the cross-entropy $\mathrm{H}(P, Q)$, and Kullback-Leibler divergence is the number of extra bits needed to identify value x drawn from X, corresponding to Q, rather than true distribution P.

2. Independent distributions have additive relative entropy.

$$D_{\text{KL}}(P \parallel Q) = D_{\text{KL}}(P_1 \parallel Q_1) + D_{\text{KL}}(P_2 \parallel Q_2). \tag{3.52}$$

$$D_{\text{KL}}(\lambda p_1 + (1-\lambda)p_2 \parallel \lambda q_1 + (1-\lambda)q_2) \leq \lambda D_{\text{KL}}(p_1 \parallel q_1) + (1-\lambda) D_{\text{KL}}(p_2 \parallel q_2) \tag{3.53}$$

If P_1, P_2 specify two independent distributions, then the joint distribution is $P(x, y) = P_1(x)P_2(y)$, and Q, Q_1, Q_2 can therefore be expressed as (3.52) is convex pair of probability mass functions (p, q), i.e. if (p_1, q_1) and (p_2, q_2) are a pair of probability mass function as (3.53) with $\lambda = [0, 1]$

3.10 Methods For Feature Extraction

The feature extraction and classification can be done using the Discriminant functions. Fisher Discrimination is the most widely accepted method for classification tasks.

1. **Fisher discrimination** Sparse representation-based dictionary learning is a method of classifying features based on sparse representations. The reconstruction error can be effectively utilized for pattern classification by constraining the coefficients of coding to have a small intraclass scatter and a large interclass scatter at the end of sparse coding. A more detailed analysis of Fisher Discriminant Dictionary Learning is reported in [72]. For handling non-linear classification problems, the introduction of Kernels in the Fisher Discriminant Function is reported by [73]. In this work, the Kernel trick was used for making Kernel Approximations. The Kernel trick transforms the non-linear classification problem into the linear classification problem in the feature space.

2. **Ada-Boost SVM classification** In this work, the Dictionary Learning technique is used for classification by combing the Dictionary Learning techniques with Ada-Boost algorithms [74]. In the Ada-Boost algorithm, sparse coefficients are used as weights.

3. **Dicriminative Deviation** The discrimination-based Dictionary learning is inherently unstable. The deviation is difference of the samples x_i and it is mean $\bar{x}$. The discriminative deviation-based dictionary learning[75] requires only one tuning parameter and handles discrimination in a controlled manner, thus being more stable than previous work.

4. **Sparse Classification Frame work**A joint framework for Dictionary Learning with sparse linear classifier for the principle feature extraction is reported in [76]. In this work, at the learning stage, to promote discriminability of Dictionary Learning, we consider discriminative sparse code errors along with reconstruction errors and classification errors. The performance of the Deep Dictionary Learning framework with sparse coding is reported in [76].

5. **Hierarchial Dicrimination** In the Hierarchial Discriminative Deep Dictionary Learning [77] technique, a discriminative layer was embedded within a Deep Dictionary Learning framework. In addition, local sparse representation goals are approximated in forward pass, through introduction of local regularization. In the work reported, the advantages of both the Discriminative approach and Deep Learning representations are exploited.

6. **Locality Constrained Sparse Coding** In this work, locality-constrained sparse coding, and discriminative constraints are added to the Dictionary Learning and the Sparse coding optimization function. The expression for the locality-constrained sparse and dictionary learning and the algorithm for its implementation given in detail in [78] .

7. **Discrimination Based on Soft thresholding** In this work a new sparse coding scheme for classification tasks for working with large data sets[79] is discussed.

3.11 Kernel Density Estimation: Selection Of Optimal Bandwidth

The T^2-Statistic, and Squared Prediction Error (SPE) defined in [80, 81] by equations 3.48 and 3.47 are used for feature detection. The T^2 metric is used to measure minute differences in feature spaces. Based on the threshold limit, the SPE metric can detect irregularities in residual space and identify their features. By selecting the optimal bandwidth, threshold limits can be calculated using Kernel Density Estimation (KDE) [80]. A Kernel Density Estimation (KDE) [80] uses density to estimate unknown numbers. Let 'Q' be a random variable for the variable event at point q, and $P_Q(q)$ be the probability density function for the event.

$$P_Q(q) = \frac{1}{n\mathbf{W}} \sum_{i=1}^{n} \big(K(q - q_i)\big) \tag{3.54}$$

where q_i represents a sample of the dataset, $\mathbf{W}$ specifies the bandwidth, and $K(.)$ satisfies the following criteria:

$$K(q) > 0; \sum_{0}^{\infty} K(q)dq = 1; \tag{3.55}$$

The optimal bandwidth [82] in the KDE is critical to obtaining an accurate Probability Density Function (PDF), and the criteria defined in [83] are used for the calculation of the optimal bandwidth ($\mathbf{W_{opt}}$) and expressed as .

$$W_{opt} = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{\frac{1}{5}} \approx 1.06\,\hat{\sigma}\,n^{-1/5} \tag{3.56}$$

where $\hat{\sigma}$ and n represent data samples (n) with standard deviations. Based on KDE, thresholds are calculated at a confidence level of 99%, with a false alarm rate of 1%. As long as the threshold limit exceeds the SPE statistic and T^2-Statistic, a feature has been detected. To make the W_{opt} value more robust for both the long-tailed and skew distributions, it would be best to replace $\hat{\sigma}$ with another parameter A', which is given by

$$A' = min(\hat{\sigma}, IQR/1.34) \tag{3.57}$$

In addition, reducing the factor from 1.06 to 0.9 will improve the model. Finally, we would express the final formula as

$$h = 0.9\ \min\left(\hat{\sigma}, \frac{IQR}{1.34}\right) n^{-\frac{1}{5}} \tag{3.58}$$

where $\hat{\sigma}$ is the standard deviation of the samples, n is the sample size, and IQR is the Inter-Quartile Range.

3.12 Summary

1. The generalized expressions and formulation of optimization problem with constraints are discussed using Lagranges theorem.
2. Formulation of optimization problem for Dictionary Learning, and techniques for solving the Dictionary Learning optimization problem are discussed.
3. Performance Metrics for estimating the performance of Dictionary Learning algorithms are defined.
4. A brief discussion is provided on the selection criteria of optimal bandwidth for Kernel Density Estimation.

Chapter 4

Dictionary Learning and Penalized Maximum Likelihood

Dictionary Learning combined with Penalized Maximum-Likelihood is used to extend Rayleigh Lidar's measurement range. We use l_1 and l_2 regularizations for achieving a higher measurement range than the present existing situation. The upper altitude limit of Rayleigh Lidar is discussed in detail in [3].

Extending the ranging height, while having acceptable level of Signal to Noise Ratio (SNR) at higher altitudes is a necessity for Rayleigh Lidar System. Increasing Power Aperture product and updating the hardware will achieve the objective. The Dictionary Learning technique is another means of achieving high SNR. After denoising the backscattered signal using Dictionary Learning, the Penalized Maximum Likelihood method has improved the accuracy of atmospheric temperatures as compared to the existing method. A measure of the method's performance is assessed in terms of improvement in ranging height and in terms of Standard Error in the temperature parameter retrieved.

Improvement of Signal to Noise Ratio (SNR) alone is not sufficient for achieving higher altitudes, an acceptable value of SE is also essential at the altitudes of the extension in the retrieved parameter through the introduction of prior knowledge of temperature parameter profiles.

The atmospheric temperature retrieval method based on fundamental laws of physics is reported in [2, 84] gives the reliable temperature profiles. It is mentioned that the uncertainty decreases in proportion to density from the top altitude down to the next lower altitude. Temperature profiles in the top 10 km to 15 km need to be ignored because of uncertainties in seed pressure at the seed altitude while using the method reported in [2] temperature retrieval. To solve the disadvantage, an inversion method is proposed by Sica and Jaya Khanna called Grid Search (GS) [4] method using mathematical inversion. The drawback has been overcome by providing seed value at lower altitude where the uncertainty in seed pressure value is less in mesosphere. The GS method has performance improvement with the equivalence of four times the power aperture product. The GS method partially explains the contribution of different sources to the uncertainty of temperature. The Optimal Estimation Method (OEM) [6] for temperature retrievel gives a complete description of uncertainty. The forward models presented in [6] completely characterize the measurements and provide feasibility for retrieval of the temperature profile, the dead time, and the background counts. Another technique for the retrieval of atmospheric temperature by taking seed values from O_2 Airglow measurements is given by Taori in [85]. Though this method reduces the SE by removing uncertainty in seed parameter value, it requires colocated seed values that put a limitation on the method. A relative comparison of atmospheric temperature retrieval from the Microwave Limb Sounder (MLS) and SABER-TIMED instruments with ground-based Lidars is given in [86].

4.1 Existing Signal Denoising Techniques

The method of Moving Averages (MA) is the simplest among all denoising techniques. Many denoising techniques based on stationary and non-stationary signal processing schemes including Kalman Filtering (KF) [87], Wavelet Transform(WT) [88, 89, 90], Empirical Mode Decomposition (EMD) [91, 92] are used. A relative comparison of SNR at different heights with the moving average method, wavelet thresholding, and EMD for the Rayleigh Lidar is given in [91]. The EMD method has good adaptability, thereby improving SNR by a factor of 2.3. The EMD has the disadvantage of a mode mixing that has been overcome by the Variational Mode Decomposition (VMD) [93, 94]. In the VMD technique, optimization algorithms are used to choose mode parameters. Combined with a quadratic penalty term, the mode number parameter provides a better SNR. Inclusion of prior information for improving SNR makes us to choose Machine Learning and Neural Networks for denoising. Least Squares Support Vector Machine (LS-SVM) [95], and Wavelet neural networks [88, 96] are some of the Machine Learning based SNR improvement techniques.

4.2 Dictionary Learning For Lidar Signal Denoising

The measured data ($\mathbf{Y} \in \mathbb{R}^{N\times m}$) is expressed as the sum of the linear combination of the product of the Dictionary matrix ($\mathbf{D} \in \mathbb{R}^{N\times K'}$) with Sparse coefficient matrix ($\mathbf{X} \in$

Algorithm 4.1: Basic methodolgy of Dictionary Learning

```
procedure DICTIONARY LEARNING
Initialization: Choose the photon count data with $SNR \geq 3$.
Sparse Coding:
    Implement (4.2)
    Decompose $\mathbf{DX}$ to the sum of $K'$ matrices with rank 1.
Dictionary Update:
For each column $k'$ in $\mathbf{D}$ update using $\omega_{k'}$, and $\mathbf{x_{k'}}^{T'}$. $\omega_{k'}$ define a group of indices pointing $\mathbf{y}_i$. $\omega_{k'} = i|1 < i < k', X_{k'}^{T'}(i) \neq 0$
Calculate the matrix of representation errors $\mathbf{E_{K'}}$. The expression for the representation error is given by $\mathbf{E_{k'}} = \mathbf{Y} - \sum_{j \neq k'} \mathbf{d_j X^j_{T'}}$. Restrict $\mathbf{E_{k'}}$ to columns matching $\omega_{k'}$, and $\mathbf{E^R_{k'}}$.. Get $\mathbf{E^R_{k'}}$ by choosing a column in $\mathbf{E'_K}$(nearest neighbour).
Calculate SVD for $\mathbf{E^R_{k'}}$. Arrange singular values of $\mathbf{E^R_{k'}}$ in descending order, and replace singular vectors corresponding to the smallest value of singular values by zeros in $\mathbf{E^R_{k'}} = \mathbf{U \Delta W^{T}}'$.
Update $\mathbf{d'_k}$ to include $\mathbf{U}$, and $\mathbf{x^R_{k'}}$ as the first column of $\mathbf{W}$. Eliminate atoms that are close to each other.
Eliminate atoms that are used by less than b(arbitrary value) training examples.
```

$\mathbb{R}^{K' \times m}$), and noise matrix ($\mathbf{V} \in \mathbb{R}^{N \times m}$) and expressed as

$$[\mathbf{Y}]_{\mathbf{N \times m}} = [\mathbf{D}]_{\mathbf{N \times K'}} [\mathbf{X}]_{\mathbf{K' \times m}} + [\mathbf{V'}]_{\mathbf{N \times m}} \tag{4.1}$$

where N refers to the number of altitude rannge bins, m represents the number of profile integrations, and K' represents any size. The learned Dictionary for the sparse representation of a Lidar data set, leads to a global approach and the same Dictionary is used regardless of the time or the offset of the data. The denoising of photon count profiles is done through thresholding and the denoised photon count profiles, obtained are inverted to retrieve temperature profiles using Algorithm 4.1. In this work, the Sparse Coding is achieved through Batch Orthogonal Matching Pursuit (BOMP), and the Dictionary update is achieved by K-Singular Value Decomposition (K-SVD). K-SVD [97, 55, 98] uses the sparse coding method combined with dictionary updates to generalize the k-means clustering method.

4.2.1 Sparse coding

Coding sparsely is implemented using Batch-Orthogonal Matching Pursuit (BOMP). The sparse coding steps minimize the representation error $||\mathbf{Y} - \mathbf{DX}||^{\mathbf{2}}_{\mathbf{2}}$. The Dictionary($\mathbf{D}$) is fixed and the sparse matrix($\mathbf{X}$) is variable. Specifically, the BOMP algorithm is optimized for sparse-coding large sets of signals and selecting the most highly correlated atom. The procedure for implementation of the algorithm is given in [99].

4.2.2 Dictionary update

At each iteration the k-means clustering has two steps: given $d_k'^K$, assign training examples to their nearest neighbors, and update d_k'. Singular Value Decomposition is the process of representing a given square matrix into the product of upper triangular matrix($\mathbf{U}$), Singular value matrix(Δ), and lower triangular matrix ($\mathbf{W}$).

$$\min_{\mathbf{D},\mathbf{X}} \left\{ \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_2^2 \right\} \quad \text{subject to} \quad \forall i, > \|\mathbf{x}_i\|_0 \leq \delta \tag{4.2}$$

The minimization of the optimization problem (4.2) corresponds to searching for possible the Dictionary and solution of equation 4.2 gives sparse representations of observations. A detailed description of the algorithm for sparse coding and dictionary updating can be found in Algorithm 4.1. Implementing the Algorithm 4.1 for once, forms a single iteration. The choice of selecting a number of iterations is dependent on the residual matrix $\mathbf{E_{k'}}$ and the tolerance value ϵ.

4.2.3 Thresholding and Denoising

The important sources of noise in the Rayleigh Lidar system are random noise, and background noise. The attenuation of random noise ($v' \sim \mathcal{N}(0, \sigma'^2)$) is achieved through the sparse a Dictionary domain. The noise cannot be represented by the sparse approximation, as it is attenuated during the sparse coding phase. By subtracting the average of all the photon counts between the altitudes 125km to 140km from the background noise can be removed.

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{v}' \tag{4.3}$$

where $\mathbf{y}$ represent the observed data, $\hat{\mathbf{y}}$ is the true data.

The denoising is done in the Dictionary domain using the thresholding technique. The observed data is the sum of true data and noise (4.3). The σ' value is selected by choosing trails that give good SNR at higher altitudes. The true data has a sparse representation in Dictionary bases. So

$$\overline{\mathbf{y}} = \mathbf{D^T}\mathbf{y} = \mathbf{D^T}\hat{\mathbf{y}} + \mathbf{D^T}\mathbf{v}' = p' + z' \tag{4.4}$$

where $\overline{\mathbf{y}}$, p', and z' are arbitrary variables. The majority of elements in p' are 0 or close to 0, and $z' \sim \mathcal{N}(0, \sigma'^2 I)$. Choose the prior, as

$$p'' \sim a\mathcal{N}(0, \sigma_1'^2) + (1-a)\mathcal{N}(0, \sigma_2'^2) \tag{4.5}$$

where a denote an arbitrary value, $\sigma_1'^2$ represent the variance of “significant" coefficients, and $\sigma_2'^2$represent the variance of “insignificant" coefficients. Then $\tilde{p'}$ = $E'(p''/y)$ = $\tau(y)y$, where $E'(.)$ is the expectation of function, $\tau(y)$ denote shrinkage factor, depending on prior variances σ_1^2 and $\sigma_2'^2$. Coefficients small in magnitude represent the noise, and coefficients large in magnitude represent the actual signal. Apply the inverse transform in the Dictionary domain to obtain the denoised data as $\hat{\mathbf{y}}$= $\mathbf{D}\tilde{\mathbf{p}}'$.

4.3 Penalized Maximum Likelihood: Retrieval of Atmospheric Temperature Profile

The Lidar back scattered signal is of the form as (4.6) :

$$P^*(z) = d'(z)\exp\left(-\int_0^z \alpha(x)dx\right) \tag{4.6}$$

where $P^*(z)$ represents the back-scattered signal power, $d'(z)$ is system dependent function, $\alpha(x)$ is a parameter related to the temperature, x is an arbitrary variable, and the exponential term gives the transmission coefficient.

$$Y = F(X) + V' \tag{4.7}$$

$$\rho(z_i) = C(n(z_i) - B)(z_0 - z_i)^2 \tag{4.8}$$

$$n(z_i) = \frac{P(z_i)M_{air}}{CRz_i^2T(z_i)}\exp\left\{\int_{z_0}^{z_i}\frac{-M_{air}g(z)}{RT(z)}dz\right\} \tag{4.9}$$

where N denote number of range bins, $\rho(z_i)$ represent atmospheric density at i^{th} layer and $F : \mathbb{R} \to \mathbb{R}$ is a non-linear operator. The forward model (4.7) is used for inverting the temperature profile. The atmospheric density profile is retrieved using (4.8) and (4.9). Equation 4.7 is the forward model used that describes the relationship between photon counts ($n(z)$) and the temperature parameter, where C is Lidar System Constant, B is background photon count, z_0 is the altitude of the top layer, $g(z)$ denote acceleration due to gravity, R denote universal gas constant, M_{air} represent molar mass of dry air, z is the height above the earth surface. The value of constant C is affected by the two-way transmission coefficient, the laser output power, and the optics and quantum efficiency.

4.3.1 Maximum Likelihood for Poisson Data

It is convenient to express (4.9) in the discrete form as

$$P^* = d' \odot \exp\{-L\alpha^*\} \tag{4.10}$$

where $\odot$ is element wise multiplication, also called Hadamard product, P^* is the noise free observed vector such that $P_i^*, i = 1, 2, 3, ...N$,$d'(z_i)$ is known system function vector $i = 1, 2, 3, ...N$, $\alpha_j^* = \alpha(z_j), j = 1, 2, ...m$, m is number of profiles, L is operator discretizing the integral. From here in the rest of the work, we assume that the noise is Poisson distributed. The noise at each range bin satisfies identical and independent(iid) processes. We retrieve temperature from the back-scattered signal $P^*(z)$. Here α is a variable that depends on temperature, $\exp(-L\alpha)$ constitute the atmospheric transmission coefficient.

The Maximum Likelihood Estimate for the photon count profile is represented as

$$\mathbb{P}(P^*|\alpha) = \prod_{i=1}^{N}\{e^{-d_i e^{-L\alpha_i}}(d_i e^{-L\alpha_i})^{P_i^*}\}/P_i^*! \tag{4.11}$$

$$J(\alpha) = \sum_{i=1}^{N}\{P_i^* \log(d_i) - (L\alpha)_i P_i^* - d_i e^{L\alpha} - \log(P_i^*!)\} \tag{4.12}$$

4.3.2 Iterative Algorithm: Solution to Maximum Likelihood Function

First of all, maximise the log-likelihood equation given by equation 4.12 for solving α. There exists a maximum log likelihood if $N \geq M$. The L^T sums the ranges from the highest bin to the lowest acceptable range bin. The first-order KKT (Krush-Khun-Tucker) conditions are necessary conditions for ensuring optimality under constraints.

$$(\nabla J)_j(\alpha) = \sum_{i=1}^{N}(L_{i,j}(d_i e^{(L\alpha)_i})) \tag{4.13}$$

$$H(\alpha) = -\{\sum_{i=1}^{N}(L_{i,j}L_{i,k}(d_i e^{(L\alpha)_i}))\}\forall j,k = 1,2,3,...M \tag{4.14}$$

The concavity of $J(\alpha)$ can be proved using gradient and Hessian matrix(H). Let $v \in \mathbb{R}^N$, the expression for gradient and the hessian matrix is given in (4.13) and (4.14).

4.3.2.1 Checking Uniqueness of Maximum and Convergence

Let $v \neq 0$, $v^T H(\alpha)v < 0$ indicate that $J(\alpha)$ has a unique maximum and is concave. The assumption is that P_N^* indicates the last part of P^* and it is not equal to zero or otherwise it is good to consider data up to where P_N is not equal to zero.

4.3.2.2 Maximum Likelihood ($KKTl_1$)

Detailed solution to (4.12) is obtained from the method of successive approximations. Final solution to equation 4.12 can be given as

$$\alpha^{(n+1)} := \frac{L^T(d \odot e^{-L\alpha^{(n)}})}{(L^T P)} \odot \alpha^{(n)} \tag{4.15}$$

Measured values of temperature are obtained by applying the line search method to (4.15). By comparing equations (4.9) and (4.10), $P^*(z_i) = N(z_i)$,L is an operator discretizing the integral, $d(z_i) = \frac{P(z_i)M}{CRz_i^2 T(z_i)}$,$\alpha(z) = \frac{Mg(z)}{RT(z)}$ atmospheric temperature profile can be retrieved iteratively using equation 4.15. To assess the sensitivity of the above algorithm to noise, the value of SNR is varied by with C in (4.8), which depends upon Rayleigh Lidar System specifications.

The optimization problem is an unconstrained optimization and an initial value of α is required for iteratively deriving the value of temperature at all range bins. The initial procedure is as follows: Identify the value of N such that $P^*(z_N) \neq 0$.

4.3.2.3 Penalised Maximum Likelihood ($KKTl_2$)

A penalty term accounting for regularization can improve the log-likelihood, and one option is to include a l^2 weighted-regularization(γ) in (4.12). So (4.12) turns to

$$S\hat{(\alpha)} = J(\alpha) - \gamma||\alpha||_2^2 \tag{4.16}$$

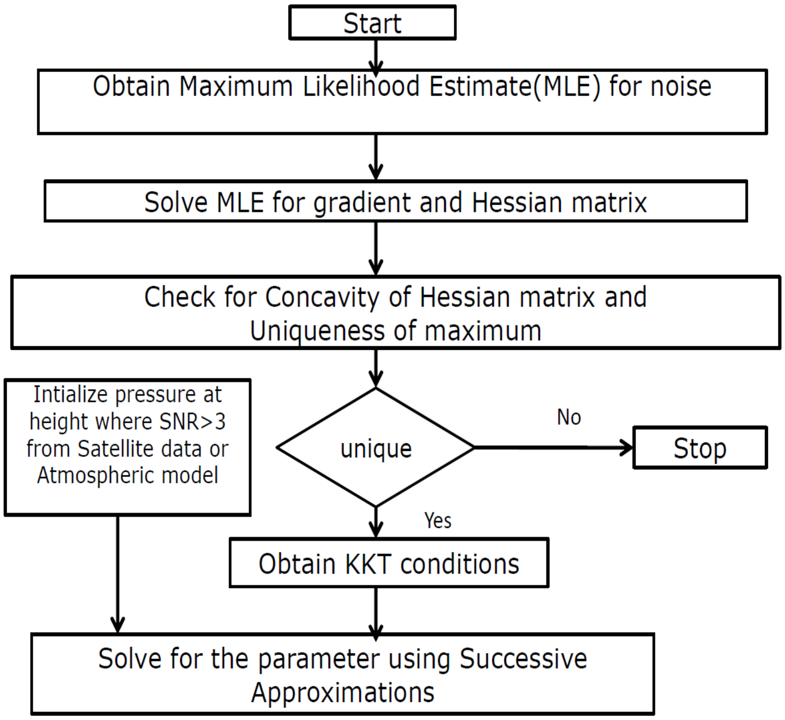


Figure 4.1: $KKTl_2$ Algorithm.

The above equation is solved using the same procedure followed for KKT-Maximum Likelihood.

$$\alpha^{(n+1)} := \frac{L^T(d \odot e^{-L\alpha^{(n)}})}{(L^T P + 2\gamma\alpha^{(n)})} \odot \alpha^{(n)} \tag{4.17}$$

Solution to (4.16) is given as (4.17). Generally, the regularisation parameter is considered as a ratio between the variance of likelihood and the variance prior to it. Since the distribution considered here is Poisson, it is assumed that inverse variance is the regularization parameter, so that the regularization parameter dont not depend on SNR. The term $2\gamma\alpha^{(n)}$ is the weighting term that reduces the error at higher altitudes. The prior profile for initialization of the algorithm is obtained from atmospheric model MSISE-00, or Rayleigh, Sodium, and Potassium resonance Lidars, Fe-Boltzmann Lidar, and satellite instruments like MLS, and SABER. SNR has no effect on the regularization parameter. A flowchart showing the implementation steps of $KKTl_2$ method is shown in Fig. 4.1.

4.4 Performance Analysis

The SNR in the case of Poisson error in the simplest form is defined as the ratio of received photon counts(S) due to scatters, the square root of the sum of the source counts(S), and background count(B) and represented mathematically as (4.18).

$$SNR = \frac{S}{\sqrt{S+B}} \approx \sqrt{S}, S >> B \tag{4.18}$$

The RLS has given optimal performance up to 90 km in the initial years, and has gone down to 78 km in course of time. The common practice for increasing SNR is by co-adding the photon count profiles in space and time or increasing the power aperture product. Adding data in time series improves SNR by a factor of the sample size squared. However, loss of minute information occurs as the integration time increases. Data from 1999 to 2011 has high SNR profiles, as the system was operated under optimal performance conditions and hence it is used for training the Dictionary. Each photon count profile is collected with a range resolution of 0.3km and with a time integration of 4 minutes. The Dictionary learned with the data set of size $\mathbf{200} \times \mathbf{450}$, where 450 is the number of profiles used for training and 200 is the number of range bins.

During this study, this Dictionary is trained using data with high SNR acquired between 1998 and 2011. The patch size is of dimension 16×16, and K-SVD is iterated 80 times. Here we have chosen 80 iterations, where the representation error tolerance reaches a tolerance (δ) of 0.001 to get the optimal set of the Dictionary atoms. A noise variance of $\sigma' = 3$,$\sigma'_1 = 4$, and $\sigma'_2 = 8$ are used for training the Dictionary. SKL metric and uncertainty analysis is performed for data collected on 5^{th}, January 2011.

Table 4.1: Comparison of SNR.

$Date$	$Height(km)$	30	45	60	75	85
28^{th}Dec, 1998	SNR	198	45	12	4	2
Energy=530mJ, PRR=20Hz	SNR_{DL}	225	53	14	5	4
11^{th}Jan,1999	SNR	135	34	8	4	1
Energy=290mJ, PRR=20Hz	SNR_{DL}	166	37	11	5	2
16^{th}Jan 2008	SNR	183	48	15	4	1
Energy=310mJ, PRR=50Hz	SNR_{DL}	229	53	16	6	3
5^{th} Jan 2011	SNR	102	22	5	3	1
Energy=580mJ, PRR=50Hz	SNR_{DL}	105	29	6	4	2

Table 4.2: Comparison of heights where SNR is 3.

Date	28Dec,1998	11Jan, 1999	16Jan, 2008	5Jan, 2011
height(km)	81	78	78	75
height$_{DL}(km)$	87	82	84	78

$$SNR_{DL} = \sqrt{\left(\frac{\hat{\mathbf{y}}}{\mathbf{y} - \hat{\mathbf{y}}}\right)} \tag{4.19}$$

SNR improvement using Dictionary Learning (4.19) and conventional techniques (4.18) are shown in Fig. 4.2 and Fig. 4.3. The mentioned specific days are selected when there is a decrease in the measurement range of the Rayleigh Lidar System. For a given threshold ($SNR \geq 3$), the Dictionary Learning technique has shown the same value of SNR at higher heights than the conventional technique. A minimum height of 4 km and a maximum height of 6 km as mentioned in Table-4.1 and Table-4.2.

To show the improvement in SNR, the SKL metric is used for the performance comparison between the DL technique, and conventional technique shown in Fig. 4.4. Figure 4.4, it is observed that the SNR of the Dictionary Learning technique falls at a slower rate than the conventional technique. SKL metric score for both techniques is zero after 92 km, where the signal and noise counts are indistinguishable in clear sky conditions.

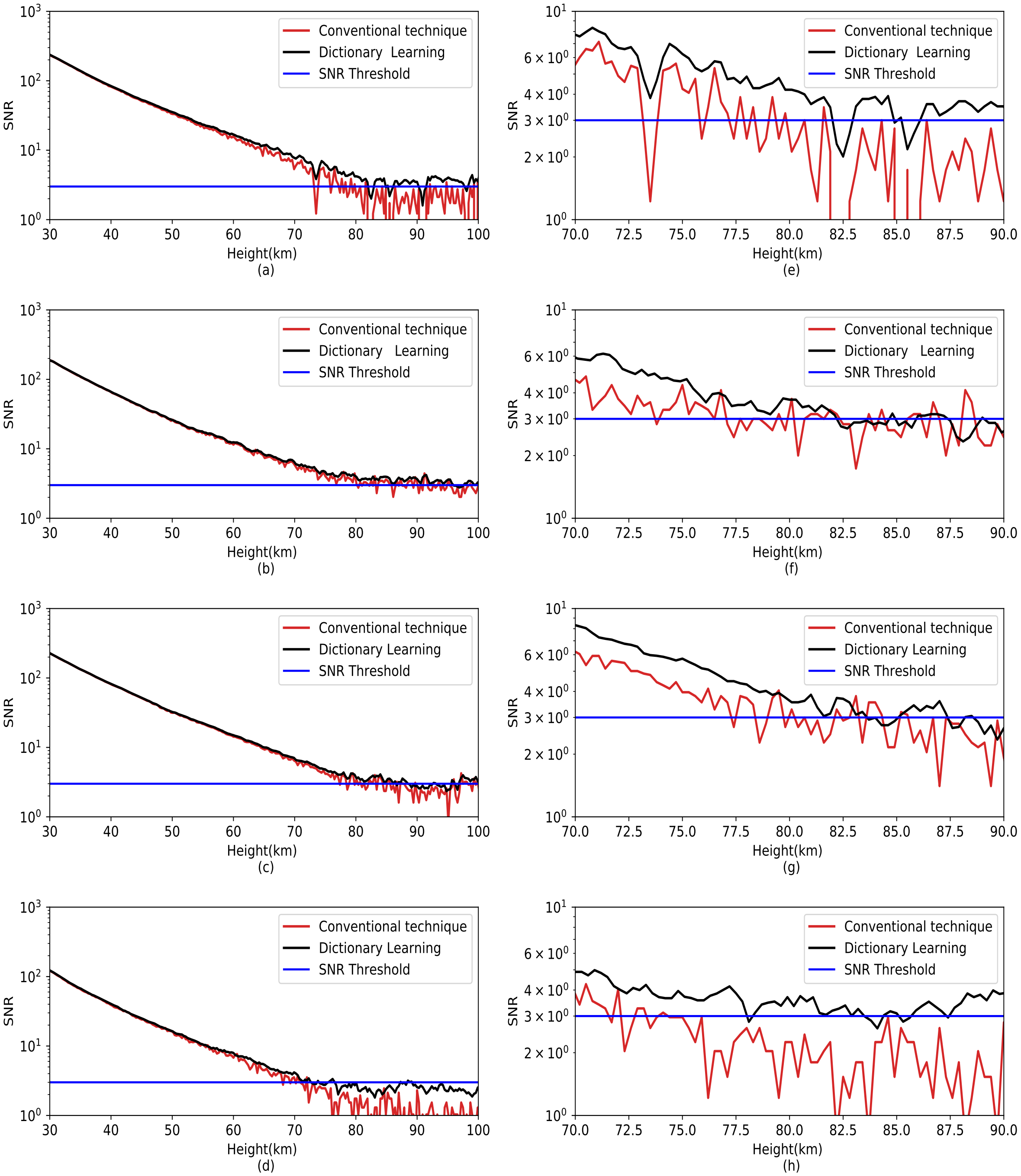
4.4.1 Evaluating the benefits of adding regularization ($2\gamma\alpha$) term in extending the measuring altitude

There is no analytical expression to assess the performance of the incorporated prior term($2\gamma\alpha$) in (4.15). The contribution of the prior is analyzed using simulation.

1. To the observed photon count profile(true profile), a structure of the photon count profile is added in higher range bins, and the resultant profile is called the estimated profile. Compare temperatures obtained using the observed photon profile and the estimated photon count profile using the $KKTl_2$ are exact.

2. The parameters for the initialization of the algorithm are chosen to have less error between the exact, and the estimated temperature profiles.

3. The sensitivity is estimated using Monte-Carlo simulation. To assess the sensitivity, we have generated 1000 Poisson realizations with the same mean of the estimated



(a) (e) (b) (f) (c) (g) (d) (h)

Figure 4.2: Comparison of Signal to Noise ratio using a conventional technique, and the Dictionary Learning for 1 hour integration time with a range resolution of 300m on Figure (a), (e):28^{th} December 1998 (b),(f):11^{th} January 1999, (c),(g):16^{th} January 2008, and (d),(h):5^{th} January 2011.

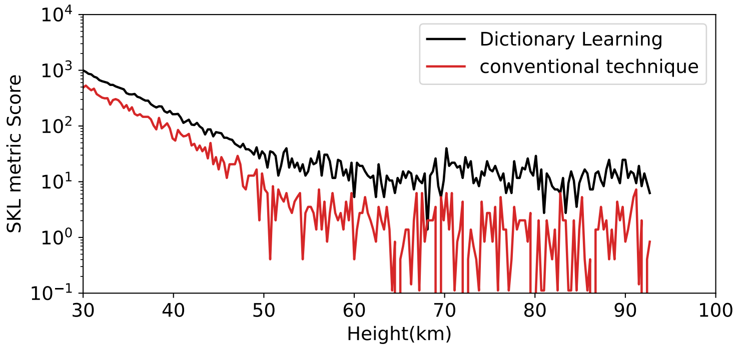


Figure 4.3: Comparison of SKL metric between the Dictionary Learning, and the Conventional technique for 1 hour integration time with a range resolution of 300m on 5^{th} January 2011.

photon profile, and we compute the estimated temperature profiles for all 1000 realizations using the $KKTl_2$ algorithm and observed that their variance is minimum in comparison with the exact shape. This indicates that the $KKTl_2$ manages to recover the presence and intensity of the temperature curve at higher altitudes.

4.4.2 Error estimation in $KKTl_2$ method for temperature retrieval

The $KKTl_2$ method is a numerical method, and there is no analytical expression for computing SE. The error in the temperature retrieved is analyzed using Monte Carlo Method. Proper selection of seed pressure is important as it has more contribution to SE than all other sources of errors like instruments, electronics used for counting, and geophysical variabilities. The detailed procedure for the analysis of SE for the $KKTl_2$ method is reported in [100] and Table 4.3. The uncertainties due to seed Pressure in the temperature profiles are shown in Fig. 4.4 for the observed data on 5th January 2011. The uncertainty

Table 4.3: NARL Rayleigh Lidar System Uncertainty Budget. $(.)/(.)$ shows the uncertainty in channels with high and low intensity for the $KKTl_2$ method.

Source	80km	70km	50km	30km
Detection Noise(N$_D$)	6K/(.)	5K/(.)	1K/4K	(.)/0.08K
Background Noise(N$_B$)	0.5K/(.)	0.3K/(.)	0.3K/(0.8)	(.)/0.02K
Seed temperature	4K	2K	0.5K	-
Gravity(g)	0.1K	0.1K	0.1K	0.1K
Molecular Mass(M$_{dryair}$)	1K	0.05K	0.05K	0.05K
Method	15K	8K	3K	0.08K
Total Uncertainty	27K/40K	16K/(.)	4.8K/8.45 K	0.08K/0.33K

variations are valid as the temperatures derived with $KKTl_2$ using actual data and synthetically generated data are identical.

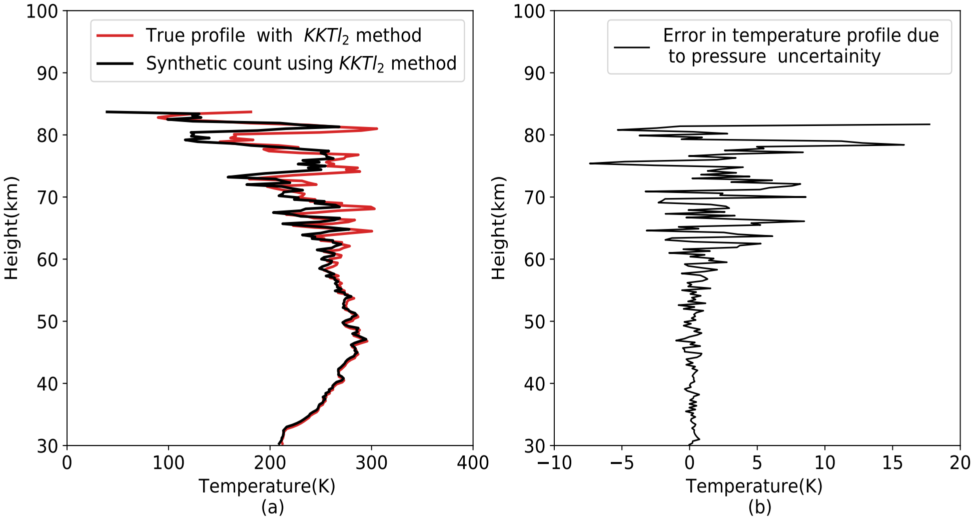


Figure 4.4: Temperature samples are obtained by randomly selecting seed pressure (P_0) that has a mean of $P_{0_{true}}$, and standard deviation of $\sigma = 3.33\% P_{0_{true}}$, where $P_0 = 2.093 \times 10^{-3} hPa$ at 95 km, (a) Synthetic count temperature profile, (b) Uncertainty in temperature due to seed pressure uncertainty for 5^{th} January 2011.

A relative comparison of temperature and SE retrieved using the HC method, SABER, and $KKTl_2$ method is given in Fig. 4.5. The average of all the cases is taken to arrive at the overall increase in the measurement range of the instrument. The algorithm is tested with varying seed pressure values to check its robustness. Improvement in the measurement range is accounted for with respect to SE. The difference in SE for the two methods is less than 6K up to 48 km as observed from Fig. 4.5. We noticed an improvement in the measurement range by 6 km at higher altitudes where SNR is less. A SE of 8K is observed at 78 km, and 84 km with the HC method, and the $KKTl_2$ method respectively. The improvement in measuring altitudes from a minimum of 4 km, and to a maximum of 7km, and having an average height improvement of about 6 km. It has been observed from Fig. 4.5(e) and Fig. 4.5(g), that the temperature retrieved by the proposed method is warmer than the temperature measured with SABER in the stratopause region by 5K.

An ensemble of temperature profiles retrieved using the $KKTl_2$ method, for 1 hour integration time of observed data in clear sky conditions from March 1998 to March 2020 in two days each month with a spatial resolution of 0.3 km, and temporal resolution of 4 minutes is given in Fig. 4.6. From Fig. 4.6(b) it is reported in [100] that “ the temperatures

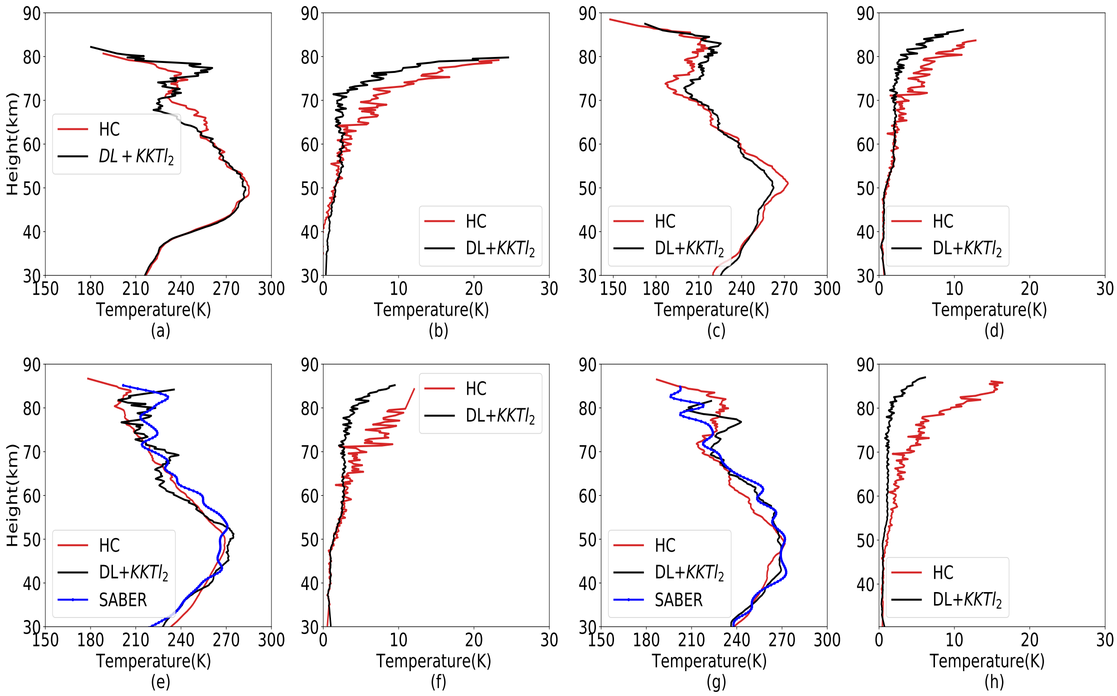
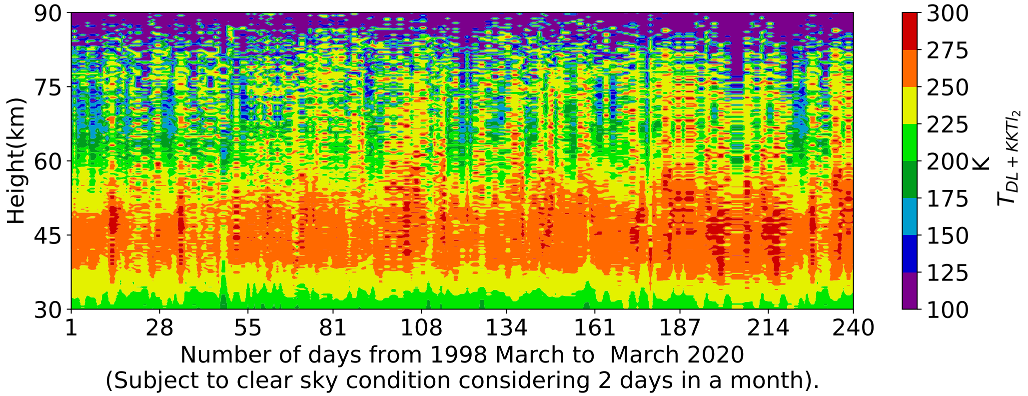
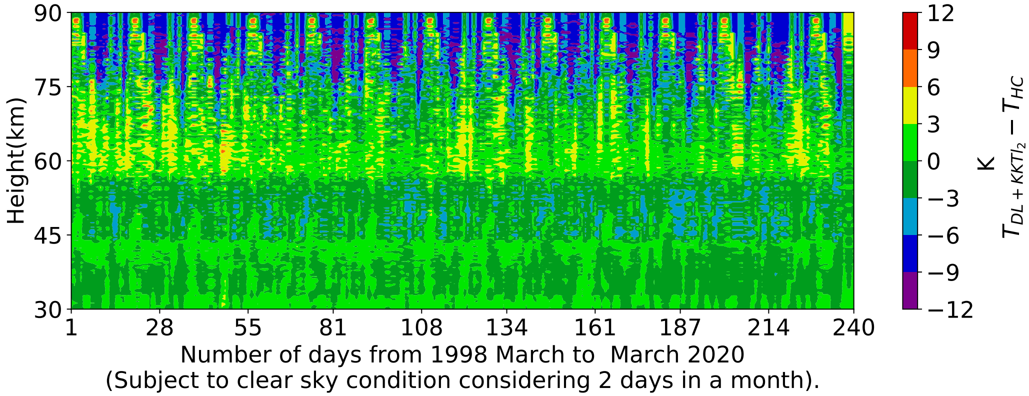


Figure 4.5: Comparison of temperature ((a), (c), (e), (g)) and SE ((b), (d),(f),(h)) using HC, the Dictionary Learning technique for denoising, and the $KKTl_2$ method for 28^{th} December 1998,11^{th} January 1999, 16^{th} January 2008, and 5^{th} January 2011 initialized with seed pressure of $2.093 \times 10^{-3} hPa$ at 95 km.

retrieved with $KKTl_2$ are lower than the temperatures retrieved with the HC method, and the difference varies from 0K to 3K for most of the profiles and a maximum of 12K in the height range of 70km to 85km. In the height range of 55km to 70km, the temperatures retrieved with the $KKTl_2$ method tend to be warmer than the temperatures retrieved with the HC method. The temperatures retrieved with $KKTl_2$ are lower than those retrieved with the HC method at the stratopause." The difference in temperatures retrieved with $KKTl_2$ and the HC method is zero. However the difference in temperature($T_{KKTl_2} - T_{HC}$) is $< 1K$ in the altitude range of 30km to 40km and $< 3K$ upto 70km suggests that there exists bias throughout the profile. The following possible reasons for the discrepancy between the two methods are reported in [100]. Further investigations are required to reduce the SE to $0.5K$ in the stratosphere and $2K$ in the mesosphere.



(a)



(b)

Figure 4.6: Ensemble of temperature profiles and their difference with $KKTl_2$ and HC methods.

4.5 Summary

1. The problem of having low SNR at higher altitudes is addressed using Dictionary Learning and Penalized Maximum Likelihood estimation.

2. The Dictionary Learning using thresholding is applied to improve SNR of lidar system at higher altitudes.

3. The relation between atmospheric density and back-scattered photon counts is formulated in the setting of Maximum Likelihood Estimation. The Maximum Likelihood Estimation is penalized to further improve SNR at higher altitude.

4. The uncertainty in the temperature profiles in using the proposed algorithm is estimated using Monte-Carlo simulations. The uncertainty is found to be 27K at an

altitude of 80km.

5. The measurement range of Rayleigh Lidar system has improved by 6km in using the proposed method for lidar photon count denoising.

Chapter 5

Multi-Resolution Dictionary Learning

We propose a method for discriminating true data from the effects of unknown sources that occur during the acquisition of data by combining Multi-Resolution Dictionary Learning (MRDL) with Instantaneous Frequency Estimation (IFE). By decomposing the signals into multi-resolutions, Wavelet Multi-Resolution Analysis (MRA) is used, all of which are performed in the Wavelet Analysis domain (W_A). The learned Dictionaries are then applied to the measured data to extract the features. As a result of the promotion of sparsity and IFE, we aim to extract hidden features in both magnitude and frequency domains of the measured data. A use case is presented in which the proposed technique is implemented and its importance is discussed in terms of identifying and extracting Doppler shifts that occur as a result of horizontal wind in Gravity Wave frequency spectra.

5.1 Temporal and Spatial Feature Extraction

The temporal information can be extracted using Short Time Fourier Transform and Spatial information can be obtained from Multi-Resolution Analysis. Data consists of hidden information in it. Wavelets are used for understanding the behavior of a wave parameter simultaneously in two domains (like temporal and spatial domains). To transform discrete-time data to segmented into frames, we Fourier transform each frame, and we store the magnitude and phase information for each frame in a matrix. The expression for Discrete STFT is given in

$$\mathbf{STFT}\{x[n]\}(m,\omega) \equiv X(m,\omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega n} \tag{5.1}$$

where x[n] is data, w[n] is the window, m is discrete, and ω is continuous, and during the implementation, the STFT both the variable m and ω are discretized and quantized. When selecting a window size, it is important to keep in mind that a narrow window has good time resolution, at cost of poor frequency resolution, whereas a wide window has good frequency resolution, but with poor time resolution. Window functions are selected based on the application. If the frequency components of the original signal are well distinguished, then we can sacrifice some frequency resolution for having a better time resolution as the spectral components are well separated from each other.

Multi-Resolution Modeling or Multi-Resolution Analysis (MRA) mathematics have applications in problems that have important features at multiple scales of spatial information in the present case. MRA refers to the splitting of the original data into individual components, that produce the original signal back when all the components are added together. The MRA analysis is related to wavelets or wavelet packets, and there also exist non-wavelet techniques that decompose any signal into multiple components. Several other multi-resolution representation techniques are available, such as filtering, Steerable pyramid [101], and Multi-Resolution Analysis (MRA) [10, 11, 102](Wavelet Packet Decomposition (WPD) in general). By choosing a range of parameters (space or time) in bandpass filtering, the desired parameter of the required scale can be obtained. In multiscale representations, steerable pyramids are more suitable for higher-order differential equations. An orthogonal wavelet filter bank can be used for the MRA. Each iteration decomposes the measured parameter data into details and approximations. The extraction of time-frequency features [103, 104] together with each resolution enhances the separation of the contributions from the unknown sources. The MRA can also detect sudden transient events like Wave Breaking in atmospheric remote sensing, and reductions or increases in transfer of potential energy to surrounding environments as a result of Atmospheric Gravity Waves. The changes are localized to certain spatial scales or frequency bands. These changes are investigated by applying MRA to raw data that are not visualized from raw data directly. We take the advantage of concept and apply this to suppress Doppler shift in the spatial domain due to horizontal background wind along with the Dictionary Learning technique. The MRA representation is done using Discrete Wavelet Transform. The Maximum Overlap DWT (MODWT) distributes the signal's energy to all scaling coefficients and detail coefficients.

An example of applying the Wavelet Maximum Overlap Multi-Resolution Analysis (MODWTMRA) for decomposing the temperature perturbation profiles into details and analysis coefficients is shown in Fig. 5.1, four minutes of time is given to each profile, and

Spatial resolution is 300m. The Fig.5.1(a) represent the vertical temperature perturbation profile, Fig. 5.1(b-g) represents the temperature perturbations with spatial resolutions of 0.6km, 1.2km, 2.4km, 4.8km, 9.6km, 19.2km, and 38.4km respectively. Also, MRA can be used to identify the different frequencies of waves in temperature perturbations based on variations in time.

5.2 Dictionary Learning In Wavelet Analysis Domain

The Wavelet Transform has advantage with respect to Dictionary Learning for its good time-frequency characteristics. The mathematical expression for Dictionary Learning using wavelets is given by

$$\arg\,\min \|\mathbf{Y} - \mathbf{W}_{\mathrm{S}}\mathbf{D}\mathbf{X}\|_F^2 \,\mathrm{subject\ to} \|\mathbf{x}_i\|_0 \leq \delta \tag{5.2}$$

where $\mathbf{D}$ represents the learned dictionary, $\mathbf{X}$ represents the coefficient vector, Y is the training data set, $\mathbf{W_s}$ represents wavelet transformation domain. δ is the number of nonzero coefficients for each atom. The Dictionary Learning in the synthesis domain can also be represented in the Wavelet Analysis($\mathbf{W_A}$) domain as

$$\arg\min_{\mathbf{D},\mathbf{X}} \|\mathbf{W}_A\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \quad \text{subject to} \quad \|\mathbf{x}_i\|_0^0 \leq \delta \quad \forall i \tag{5.3}$$

where $\mathbf{W}_A$ represent the Wavelet analysis operator.

Representation of $\mathbf{y}$ in terms of a dictionary $\mathbf{D}$ solves the problem

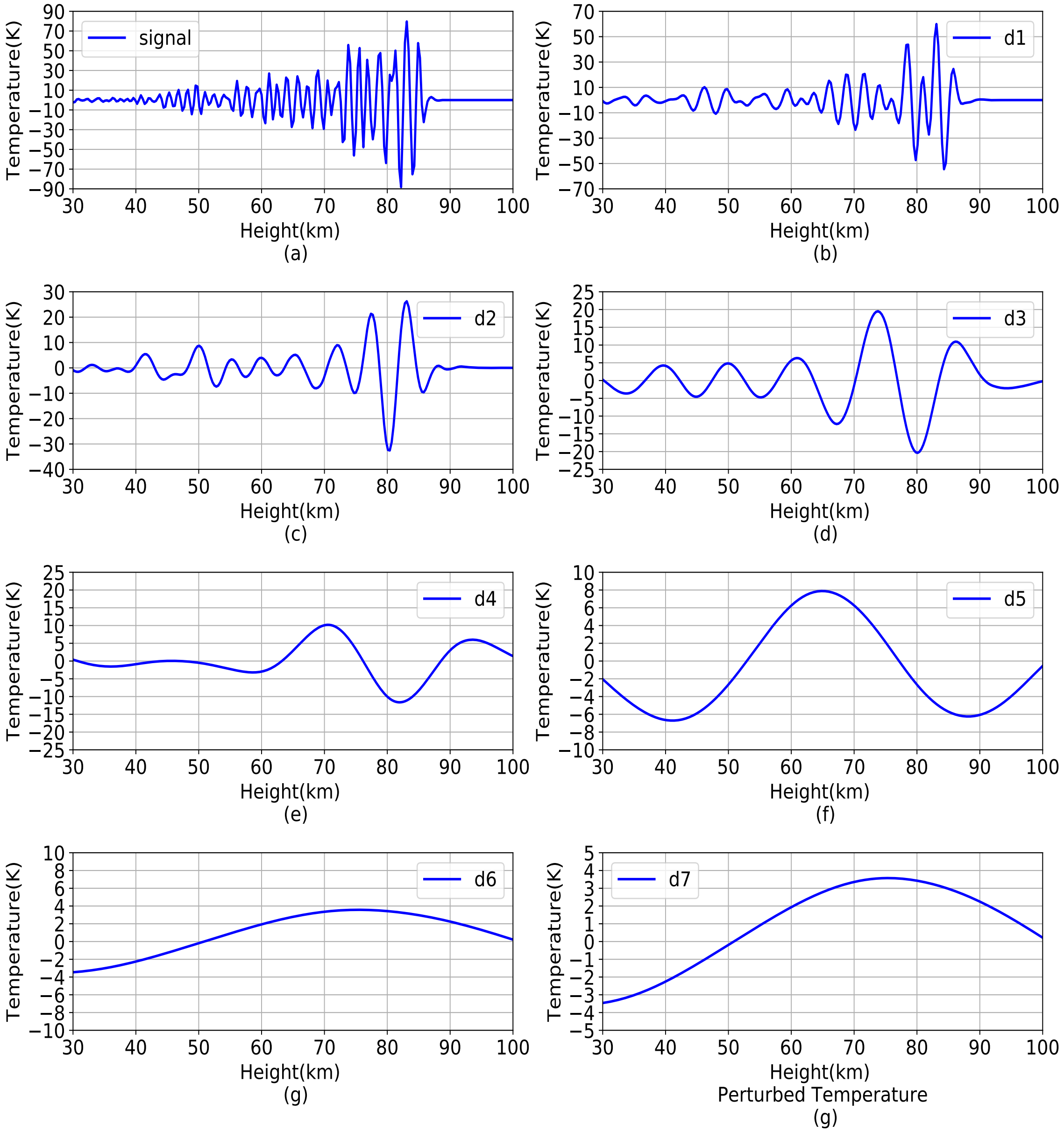
$$(P_{0,\epsilon}) \quad \min_{\mathbf{x}} \|\mathbf{x}\|_0^0 \text{ subject to } \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \leq \epsilon \tag{5.4}$$

$$\left(P_{0,\epsilon}^{W}\right)_{\text{analysis}} \min_{\mathbf{x}} \|\mathbf{x}\|_0^0 \text{ subject to } \|\mathbf{W}_A\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \leq \epsilon \tag{5.5}$$

where the goal here is to find the sparsest representation that can explain $\mathbf{y}$ as $\mathbf{DX}$ with no more error than ϵ. The same problem is represented by (5.5) in the analysis domain as equation 5.5.

5.3 Multi-Resolution Dictionary Learning

According to the methodology reported in [105], MRA combined with WPD can be applied to discriminate when the measured data is influenced by another source. Block diagrams



(a) (b)

(c) (d)

(e) (f)

(g) Perturbed Temperature (g)

Figure 5.1: Decomposition of a temperature perturbation profile into individual spatial scale profiles with details coefficients with 7 levels of decomposition.

and flowcharts are provided in Fig. 5.2 and Fig. 5.3 for reducing the effects of other sources in measured data. Combining WPD and Dictionary Learning results in the lowest Reconstruction Error (RE) among multi-resolution representations.

```
Algorithm 5.1: Multi-Resolution Dictionary Learning(MScDL)
Input:tolerence($\epsilon$), regularization parameter($\alpha$), patch size, training data, test data,
    no. of. iterations, type of wavelet, no. of. decomposition levels, no. of. atoms,
    size of dictionary
Output:The $\mathbf{D_b}$ combined with DWT, give the effective multi-resolution
    representation.
if ($SNR > 3$) && (observation time $\geq$ 4 hours) then
        • set $D_b = 0\forall\ b\epsilon\mathbb{R}^K$.
        • Learn and update the $\mathbf{D_b}$ separately for all
          decomposition using KSVD.
        • Solve the optimization problem (5.5) for sparse
          representation of each band using the iterative
          shrinkage Algorithm given in.
        • Feed the learned $\mathbf{D_b}$ and $\mathbf{x_b}$ to IFE Algorithm
          for the removal of Doppler Shift.
    else
        Look for other data.
    end
end
```

The Dictionary is trained and updated using the procedure described in [56]. Dictionary Learning in the wavelet domain uses the same training data as in the analysis domain (W_A) of Multi-Resolution decomposition. Figure 5.4 illustrates the structure of MRA and Dictionary Learning of Fig. 5.2. In the wavelet analysis domain, sparse matrices and dictionaries are created to represent temporal variations with specific resolutions as

$$\forall b \ \underset{\mathbf{D_b},\mathbf{X_b}}{\arg\min} \|(\mathbf{W}_A\mathbf{Y})_b - \mathbf{D}_b\mathbf{X}_b\|_F^2 \text{ s.t. } \|\mathbf{x}_{i,b}\|_0^0 \leq \epsilon \tag{5.6}$$

where ϵ denote error tolerence(=10^{-2}). Instructions for implementing MScDL can be found in Algorithm 5.1. The Dictionary($\mathbf{D_b}$) is produced by combining the fixed basis(B) with Dictionary($\mathbf{D}$):

$$(\mathbf{D_b})_{\mathbf{n'}\times\mathbf{K'}} = \mathbf{B}_{\mathbf{n'}\times\mathbf{K'}}(\mathbf{D})_{\mathbf{n'}\times\mathbf{K'}} \tag{5.7}$$

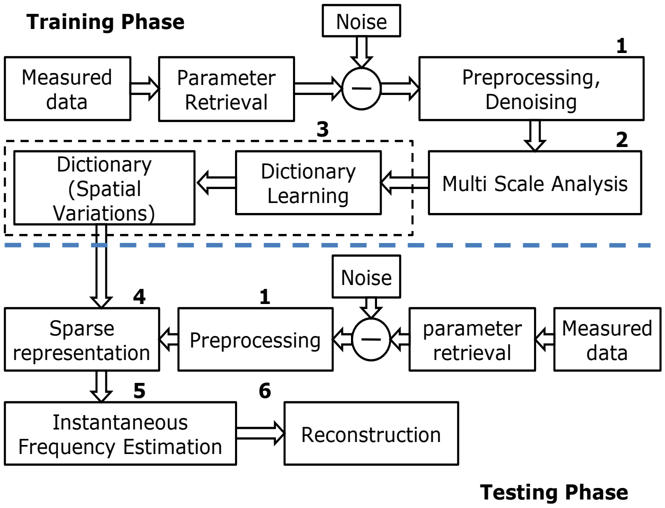


Figure 5.2: Internal representation of Multi-Resolution Dictionary Learning. Block numbers are displayed on each block so that the technique can be understood better.

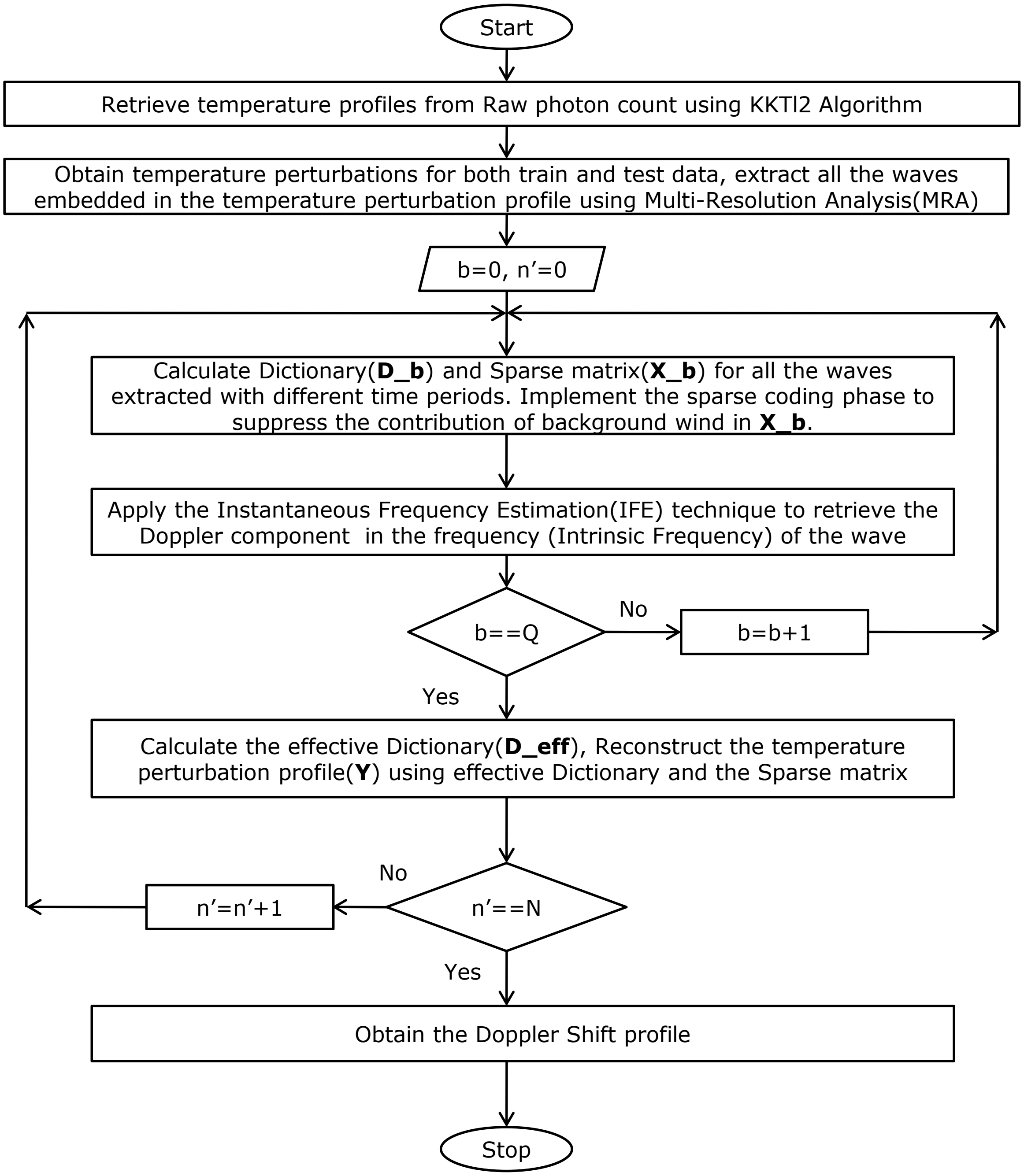
where $1 \leq b \leq Q$, and Q is count of wave decompositions that can be obtained from the temperature perturbations after applying MRA. The kernel function is a fixed basis (Fourier Transform, Wavelet Transform). .

$$(P_0^w)^{\text{global}} \; \min_{\mathbf{x}} \sum_b \sum_k \left\| [\mathbf{W}_a \mathbf{y}]_b^k - \mathbf{D}\mathbf{x}_b^k \right\|_2^2 \; subject\ to \; \sum_b \sum_k \left\| \mathbf{x}_b^k \right\|_0^0 \leq \ell \tag{5.8}$$

$$||\mathbf{x}||_{\mathbf{2,0}} = ||\beta||_{\mathbf{0}} \tag{5.9}$$

where $\mathbf{D}_b$ gives temporal information about AGW and $\mathbf{x}_b^k$ gives information about spatial analysis of AGW, $\beta = [||b_1||_2, ||b_2||_2, ..., ||b_n'||_2]^T$, $b_1, b_2...b_n'$ are columns of $\mathbf{x}$. The problem (5.8) can be solved with algorithm described in [69]. Sparse code recovery may be expressed as equation 5.9 where ϵ coverges to l_2-norm that denote noise. $||\mathbf{x}||_{\mathbf{2,0}}$ stands for l_2/l_0 norm of $\mathbf{x}$,

In the analysis domain, sparse coding is mathematically represented as (5.9). In the analysis domain, patches are coded by using the BOMP algorithm, based on the appropriate sub-dictionaries. As with single-scale dictionary learning, all operations are performed on sub-dictionaries. The process can be viewed as (5.8), which takes pursuits on patches. The sparse matrix is defined as constraints as $||\mathbf{x}||_{\mathbf{2,0}}$ in this work.



Q= Number of Levels of Decomposition

Figure 5.3: Retrieval procedure for hidden features.

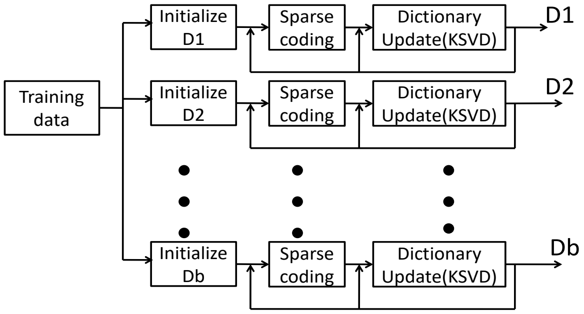


Figure 5.4: A method for learning multi-resolution dictionaries. D_1, $D_2...D_b$ indicate the Dictionaries corresponding to each wave, decomposed by the measured parameter profile (**Y**) using MRA.

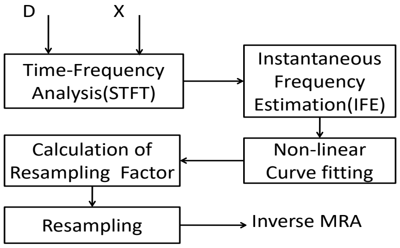


Figure 5.5: Resampling factor calculation methodology.

5.4 Instantaneous Frequency Estimation

A Doppler-shifted version of the intrinsic frequency is observed. The rows of the sparse matrix for each scale, give the temporal information of waves. Doppler Shift can be removed using Instantaneous Frequency Estimation (IFE), [106, 107, 108] and a series of steps in the methodology is shown in Fig. 5.5. Figure 5.5 is the internal representation of reducing background wind effect by Instantaneous Frequency Estimation of Fig. 5.2. The IFE utilizes Short-Time Fourier Transform (STFT), to get Instantaneous Frequency (IF) vector. The peak of STFT distribution for each sampling interval is the IF. Projection of time-frequency curve on the frequency axis is the Instantaneous Frequency. The resampling sequence of IF's is obtained by fitting data to time-frequency distribution curve. When Doppler shift is present in a profile, each subsequent signal period will contain n of time samples using equation 5.10.

$$n = \frac{f_{s1}}{f_1} = \frac{f_{s2}}{f_2} = \frac{f_{s3}}{f_3} = ... = \frac{f_{sN}}{f_N} = \frac{f_s}{f_0} = constant \tag{5.10}$$

Table 5.1: Anology between Atmospheric Gravity Waves and Multi-scale Dictionary Learning

Block Name (Block number)	Input	Output
Parameter data (1)	Temporal Variations	Preprocessed data
MRA (2)	Temporally varying data	Temporal profiles decomposed to 5 levels
Dictionary Learning, Dictionary (3)	Each individual level of all temporal profiles	Dictionary Matrix(D_b), sparse matrix(X_b)
Sparse Representation (4)	$(D_b)_i$, $(X_b)_i$ $\forall 1\leq b \leq Q$; $1\leq i \leq n'$	X_b after applying $l_2/l_0 - norm$
IFE(5)	X_b	IF vector, resampling factor,Hidden Feature
Reconstruction (6)	$(D_b)_i$, $(X_b)_i$ $\forall 1\leq b \leq Q$; $1 \leq i \leq n'$	D_{eff}, X_{eff}, and $\hat{Y}$

where $f_1, f_2, .., f_N$ denote the IFs of the signal, $f_{s1}, f_{s2}, ..., f_{sN}$ indicate instantaneous sampling rates of the signal with Doppler shift; f_S, f_0 denote the sampling rate and the IF of the original signal and n denotes the sampling numbers per period of the signal. Consequently, the resampling interval of the signal acquired is determined by

$$constant = f_i \times dt_i = f_0 \times dt \tag{5.11}$$

where dt_i and dt represent the resampling interval of the signal with Doppler shift and the given original signal without resampling, respectively.

5.4.1 IFE based on Time-Frequency Distribution (TFD)

For the present case, obtaining the IF's through Time-Frequency distribution is more beneficial.

$$f_{IF_i} = \frac{\sum_{i=-\infty}^{\infty} f.P_i(\bar{t}, f)df}{\sum_{i=-\infty}^{\infty} P_i(\bar{t}, f)df} \tag{5.12}$$

In equation 5.12, the IF is defined as the first moment of the Time-Frequency distribution. There are two approaches to obtaining the IF. One is to calculate the IF from equation 5.10 directly, and the maximum of TFD is IF obtained using equation 5.12. The first one is chosen in our study as it is the most suitable and valid one, and the mathematical representation is given in equations 5.10 and 5.11. The first one is the most useful methods for

Time-Frequency distribution analysis and has the advantages like easy interpretation, and fast calculation. Discrete STFT distribution STFT(K', m') is a matrix with K' rows and m' columns, where K' denote number of coherent process. There should be a maximum value corresponding when there are more than one component in each column.

5.5 Reconstruction

The perturbation profile in the Wavelet analysis domain for $\mathbf{y}$ is expressed as

$$\mathbf{D}_{\text{eff}}\mathbf{x} = \left(\bigcup_{b}\bigcup_{k}\mathbf{D}_{b}\mathbf{x}_{b}^{k}\right) \tag{5.13}$$

where $\bigcup_{b}\mathbf{D}_{b}$ gives temporal information about AGW and $\bigcup_{k}\mathbf{x}_{b}^{k}$ gives information about spatial analysis of AGW.

5.6 A Case Study: The Discrimination of Horizontal Background Winds on Gravity Wave Signatures

For the purpose of analyzing atmospheric gravity waves, we have to suppress the contribution of other sources. The perturbations must be isolated from background sources that develop instabilities. Using MRDL with sparsity promotion, with IFE, this work detects and removes the effect of background wind [28] leading to Doppler-shift. GPS Radiosonde (in situ) [109] and Meteor Radar [110] are used to validate the Doppler shift profiles.

There are discussions on effects of horizontal wind on Gravity Wave characteristics in [28, 30] as: “A rise in horizontal energy density is observed in temperature perturbations owing to background winds, and a decrease in vertical energy density at lower and higher frequencies, respectively. However, the vertical energy density is observed to be high above buoyancy frequency". The procedure for the extraction of Gravity Wave parameters from Rayleigh Lidar data is explained in [111, 36]. The time, frequency and vertical height, vertical wave number form the Fourier transform pairs.

5.6.1 Selection of Data Set and Processing

A temperature perturbation is obtained by retrieving temperature profiles using the algorithm described in [100]. Vertical heights of GWs range from a few meters to kilometers, causing fluctuations in the middle atmosphere temperature. We collected test data on 21 April 2014 at 20:00 Indian Standard Time (IST) for the removal of background wind. For training and testing, data with large temporal variations observed for more than four hours and $SNR > 3$ is selected. These data selections were made to maintain acceptable SNRs while maintaining a high resolution. Altitude and system noise limit the maximum and minimum vertical wavelengths, respectively.An observation period of GWs is limited by the integration time associated with individual profiles.

Data for training is a set of photon count profiles collected in 2014 with a $SNR > 3$. The training data is 184×3046. The rows represent the vertical altitudes of the atmosphere between 25.1 km and 80 km, whereas the columns represent temporal profiles in train and test data. Test data is of shape 184×60. For each wave, the learned dictionary and sparse matrix are 184×3046 and 3046×60, respectively. Dictionary learning is performed with a regularization parameter of 1, having error tolerance (ϵ) of 0.01, for 80 iterations.

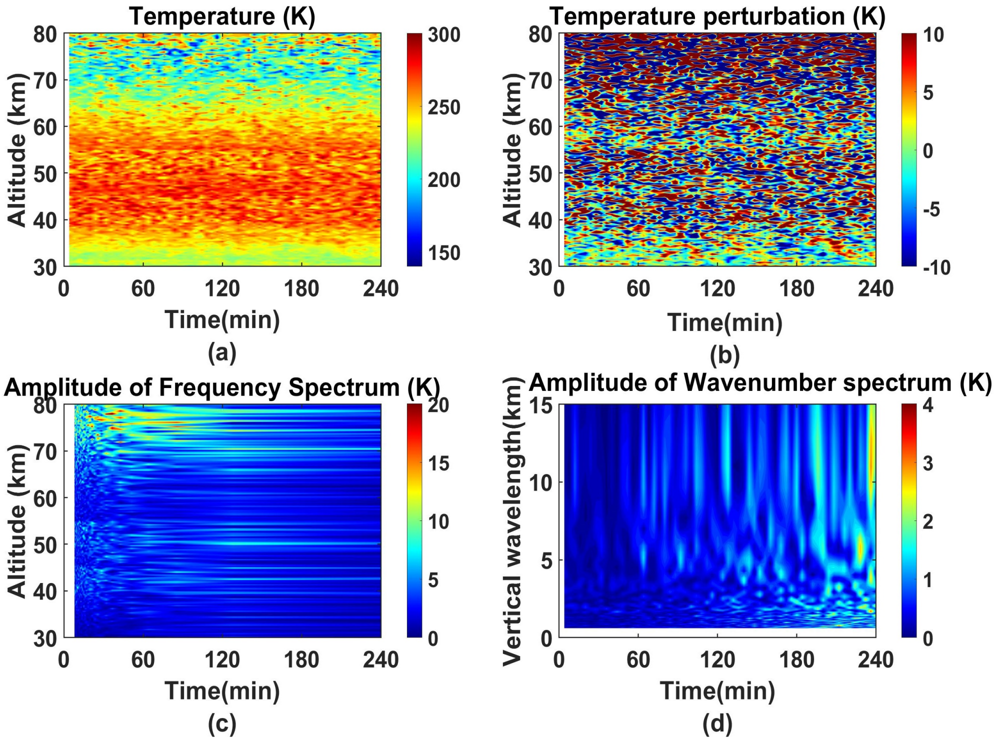


Figure 5.6: Test data collected on 21^{st}, April 2014. (a). Temperature profiles (b) Temperature perturbations, (c) Frequency domain representation of temporal profiles (d) Identification of the dominant vertical wavenumber

Figure 5.6 illustrates how to identify the dominant vertical wavelengths. Figure. 5.6(a) provides temperatures at a four-minute resolution. An altitude-dependent temperature perturbation and frequency spectrum can be seen in Figures 5.6(b) and (c). Fig.5.6(c) shows that waves are with a dominant time period between 15 and 80 minutes. The present gravity wave data is dominated by waves with vertical wavelengths between 4.2km and 7km and 10km to 15km. The vertical number dominance is observed between $2km^{-1}$ and $4km^{-1}$, and vertical wavelength dominance is observed around 5km to 7km and 9km to 13km from Fig. 5.6(d). Horizontal wavelength and intrinsic frequency are calculated using methodology reported in [112] after compensating for background wind.

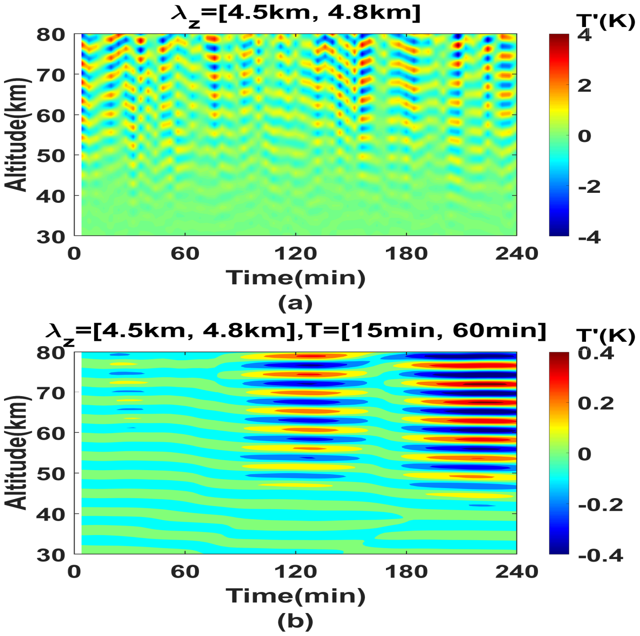


Figure 5.7: Temperature perturbations for photon profiles collected on 21 April 2014. (a) $\lambda_z = [4.5km, 4.8km]$, (b) $\lambda_z = [4.5km, 4.8km]$ with dominant period of 15min to 60min.

The vertical wavelength band-pass filtering with lower-limit(λ_{z_L}) and higher limit(λ_{z_H}) of 4.5km and 4.8km respectively is implemented on Fig. 5.6 and shown in Fig. 5.7(a). Fig. 5.7(b) shows spatially and temporally filtered temperature perturbations lasting between 15 minutes and 60 minutes. A contour plot of vertical wavelength-filtered temperature perturbations plotted to both spatial and temporal scales identifies the monochromatic waves. As can be seen from Fig. 5.7(b), one vertical wavelength cycle corresponds to bright and low-intensity stripes along altitude. Monochromatic waves have a stationary pattern of either propagating upwards or downwards and their amplitudes are time-independent.

After identifying the dominant vertical wavenumbers, the MRDL process is applied to the wavelet analysis domain. We learn Dictionary and Sparse matrices separately from the training set for each resolution based on the temperature perturbations. Using sparse coding phase, background wind contributions to the energy enhancement are corrected by applying the l_2/l_0 norm for each wave, and background wind contributions to frequency are corrected using the IFE method.

The proposed method is evaluated using performance indicators such as Minimum Mean Square Errors (MMSE), Peak Errors (PE), and Reconstruction Errors (RE). Robustness is measured using wavelet test signals as shown in Table. 5.2. An analysis of the proposed method's robustness against changes in both amplitude and frequency is conducted using the Doppler test signal, testing sampling and quantization errors with block signals, testing robustness against transient changes with bumps, and testing the method's phase variation and frequency variation with the heavysine test signal. From Table 5.2, it is evident that MMSE is lower for chirp signals. The Peak Error between the chirp signal and the reconstructed profile is the lowest among other test signals. Better performance with the Chirp signal reveals that the proposed method is more capable of handling frequency variations by preserving them, which is useful for us in calculating the Doppler shift due to background winds in the temporal perturbation profiles.

Table 5.2: Evaluating the performance of the algorithm with wavelet test signals

Test Signal	Doppler	Blocks	Bumps	Heavysine	Chirp
MMSE	1.49e-2	8.32e-5	3.4e-6	1.06e-2	8.744e-12
PE	3.9	0.00981	8.501e-3	1.17	9.18e-5

5.6.2 Selection of Parameters for Dictionary Learning and Frequency Estimation

According to performance metrics and the time resolution required for separation, an optimal combination of Wavelet parameters is chosen, including how many decompositions are made, what type of Wavelet is used , and selection of parameters in DL and IFE in the algorithm. Wavelet types with the lowest RE value are chosen to determine how many levels to decompose. From Table5.3, MMSE and PE are lower for the 'Haar' Wavelet with three levels of decomposition than with four levels.

5.6.2.1 Selection of type of Wavelet

The Fig. 5.8 illustrates a temporal profile decomposed into five levels with time periods of 8 minutes, 16 minutes, 32 minutes, 64 minutes, and 128 minutes. From Table 5.3 and Fig. 5.8, the combination of Haar' wavelets with 4 levels of decomposition is used in this work.

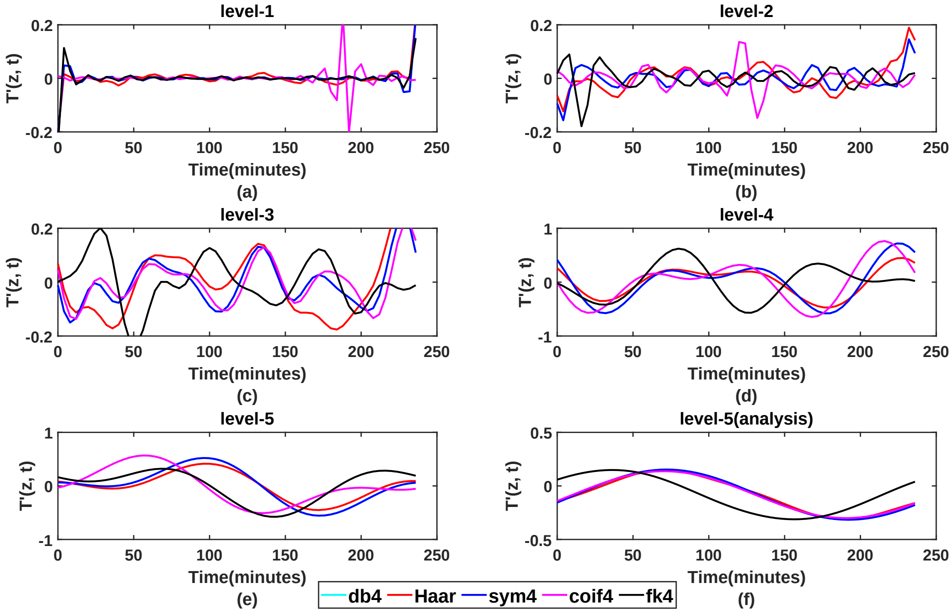


Figure 5.8: Wavelet Decompositions of a temporal profile with DB4, Haar, Sym4, Coif4, and Fejer-Korovkin (fk4) Wavelets with time intervals of (a) 8 min, (b)16 min, (c) 32 min, (d) 64 min, (e) 128 min and (f) analysis coefficients of level 5. Decompositions obtained with db4' and Haar wavelets exhibit no difference in order.

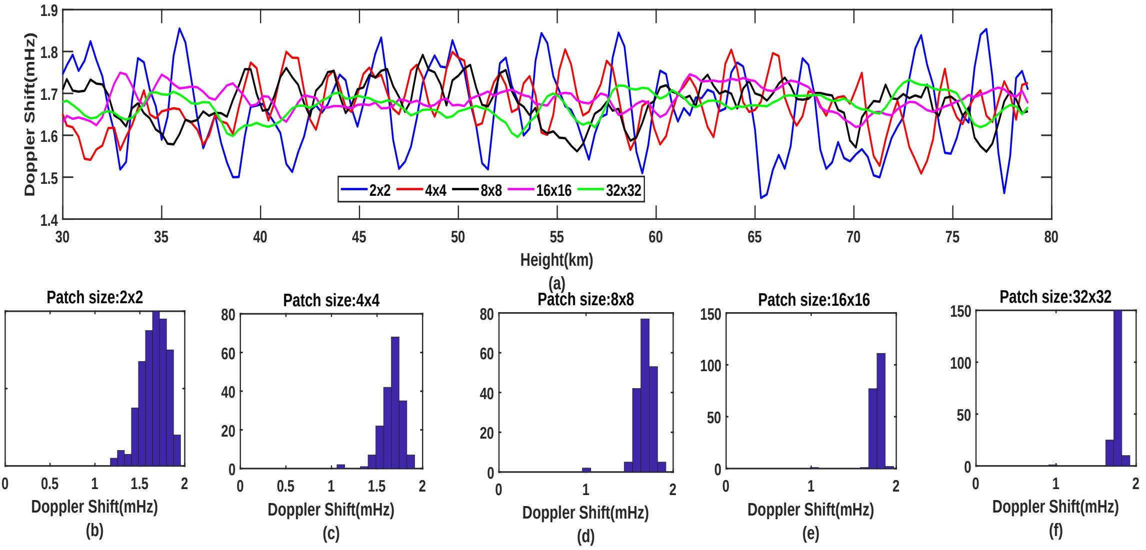


Figure 5.9: Identification of patch size. (a) Variations between patch sizes in Doppler shift profiles. (b), (d), (e), (f) illustrate the efficiency of patch size in preserving small changes in Doppler Shift.

Table 5.3: Comparison of Performance Indicators according to Wavelet type and level of decomposition (Q).

Q	Wavelet	db4	Haar	sym4	coif4	fk4
1	RE	2.487e-12	2.487e-12	4.381e-13	4.377	9.912
	PE	8.421e-13	5.511e-16	1.984e-13	1.124	0.9183
	MMSE	1.862e-26	2.051e-32	6.878e-28	0.021	0.016
2	RE	8.115e-12	8.115e-12	1.049e-12	9.013	8.5270
	PE	2.268e-12	7.771e-16	4.344e-13	1.9356	2.129
	MMSE	8.578e-26	2.054e-32	3.539e-27	0.086	0.0756
3	RE	1.961e-11	4.425e-16	3.270e-12	7.793	7.973
	PE	3.329e-12	1.221e-15	6.109e-13	2.1504	1.4071
	MMSE	1.848e-25	2.958e-32	2.207e-26	0.077	0.090
4	RE	1.961e-11	**8.881e-16**	1.442e-12	1.877	5.846
	PE	3.455e-12	**8.881e-16**	1.442e-12	0.721	1.612
	MMSE	4.016e-25	**2.054e-32**	6.031e-26	0.008	0.043
5	RE	9.9128	4.035e-15	9.023e-12	3.902	6.342
	PE	3.41e-12	9.92e-16	2.224e-12	0.790	1.769
	MMSE	4.142e-25	2.052e-32	8.265e-26	0.018	0.052

In order to remove background wind from GWs, the number of levels of decomposition is crucial.

5.6.2.2 Decomposition Levels

The number of levels of decomposition is unaffected by the amplitude of the background wind-induced Doppler component. Decompose the profile into more levels, and the representation will be more accurate under RE. A maximum of five levels of decomposition are used for the present temporal profiles.

5.6.2.3 Patch Size

The performance of dictionary learning was examined using Doppler Shift profiles containing different patch sizes are shown in Fig. 5.9(a). According to Fig.5.9(a), the patch size of 2×2 yields more insight into the Doppler shift variation than the other patch sizes. Doppler Shift profiles become more biased towards maximum values as the patch size increases. Doppler Shift information can be lost when patch sizes are large, and can be observed from Fig.5.9(b-f).

5.6.2.4 Window size

In this study, the window size is $20.6mHz$. Depending on the required time resolution, the window size is selected to separate background wind.

5.6.3 Applying MScDL Algorithm

The detailed procedure for applying the proposed method to reduce the effect of Doppler shift due to horizontal background wind is given below.

1. Use training data to implement MRA.
2. Coefficients at each range bin for all the temporal profiles are divided into non-overlapping individual blocks.
3. Determine the sparsity based on the dictionary learning algorithm on each scale using (5.8).
4. Dictionary and sparse matrices are the results after representing each wave.
5. Reduce the effect of background wind by using equations 5.9 and 5.11.
6. The effective data representation is the union of all the data at different waves at each altitude.

The procedure for the reconstruction of temperature perturbations is as follows.

1. For each wave, multiply the Sparse matrix with the appropriate Dictionary matrix.
2. Reconstruct the perturbation profiles by applying the IMRA to the sum of individual resolution representations using (5.13).

For the present learned Dictionary, mutual coherence is 0.0485, that indicates the learned Dictionary is efficient. By obtaining the hyper-parameters by grid search, and tuning them with Wind Velocity data derived from Meteor Radar (MR) measurements, this work achieves the best possible results in Dictionary Learning.

5.6.4 Retrieving Doppler Shift

The procedure for removing background wind is explained at the altitude of 45km as an examples. The same procedure is repeated at all the altitudes.

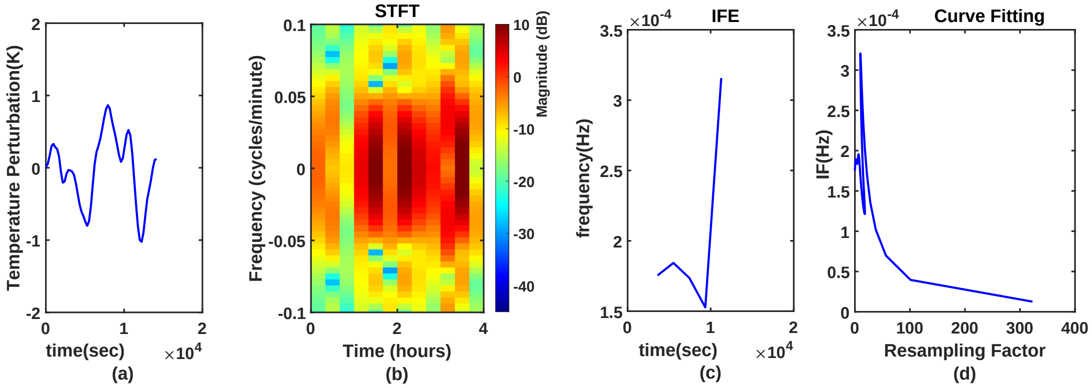


Figure 5.10: In order to better understand the IFE method, intermediate stages are explained for a single time profile at 45km height. (a) Temporal perturbation profile, (b) time-frequency variation (c) determining IF, (d) estimating resampling factor.

Figure 5.10(a), illustrates the selected temporal profile, and shows it in both the frequency and time domains with STFT, as shown in Fig. 5.10(b). The sampling frequency(f_0) for the STFT representation is 4.16mHz, window width is 20.64mHz. A graph of Instantaneous Frequency (IF) can be found in Fig. 5.10(c), which is calculated with (5.10). A frequency range of 0.169mHz to 0.363mHz is observed for the IF. In order to reduce the Doppler shift, the resampling factor needs to be applied to the time intervals. The sampling frequency vector and the IF vector are fitted with non-linear least-squares with 95% accuracy as shown in, Fig. 5.10(d). The fitting is done using the below equation

$$\hat{f_{IF}} = a * (f_{IF} - b)^{\hat{n}}; a \in [-9.578e - 08, 4.272e - 07], b \in [1938, 1.102e + 04] \quad (5.14)$$

where a =1.657e-07, b =6478 and $\hat{n} = 1$. The resampling factor is determined by dividing $\hat{f_{IF}}$ by f_s. In order to determine the quality of the fit, we use SSE, R-Square, Adjusted R-Square, and RMSE. Doppler shift and IFE-derived Doppler shift are the actual and predicted responses, respectively. As a result of the current fitting technique used in this work, SSE, R-square, Adjustable R-Square value, and RMSE are 7.45e-7, 0.833, 0.732 and 4.98e-4 repectively.

5.6.5 Validating with Insitu and Remote Sensing Instruments

Using the Rayleigh Lidar, IFE is able to obtain Doppler shift profiles in comparison to GPS Radiosonde (in situ) and Meteor Radar (MR). We are able to obtain Doppler shift profiles for height domains of 25km to 35km and 72km to 80km by substituting zonal wind data, wind direction, and horizontal wavelength data.

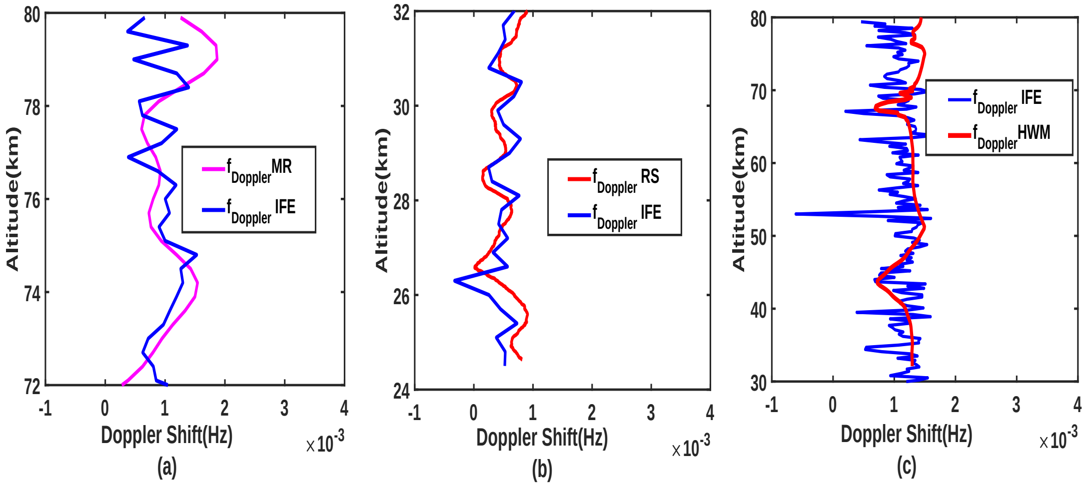


Figure 5.11: Comparison of Doppler Shift profile, (a)IFE and MR, (b)IFE and GPS-Radiosonde , compare Doppler Shift profiles in the height range of 25km to 35km, (c) HWM and IFE.

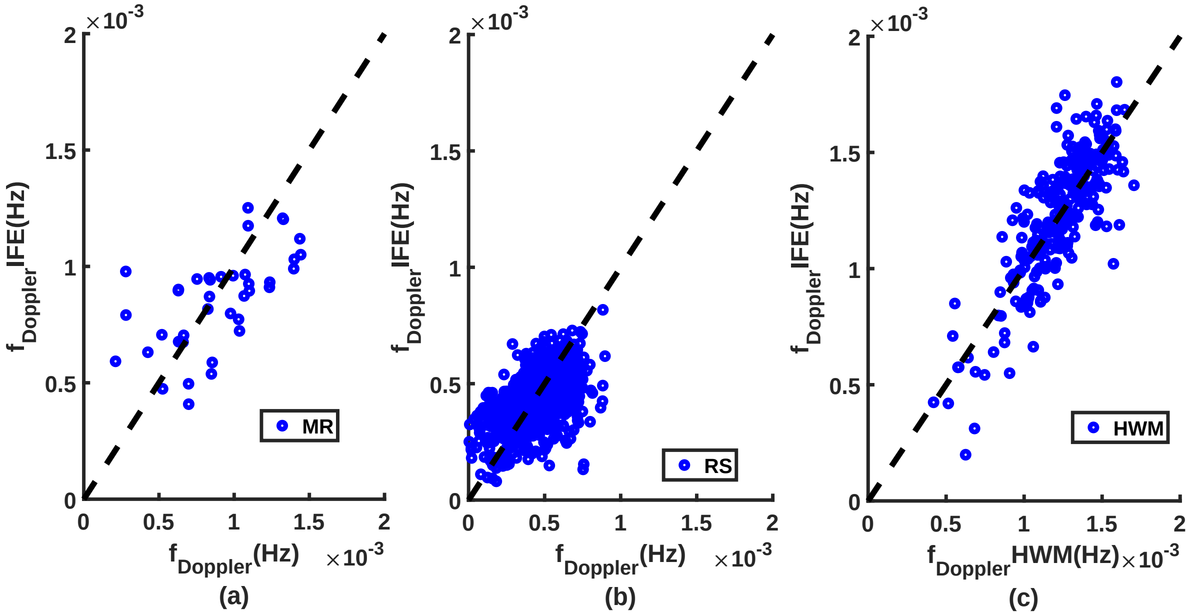


Figure 5.12: Retrieved Doppler shift comparison for data acquired on 7 April 2014, 21 April 2014, 9 January 2015, and 11 August 2016. with Doppler shift obtained from (a). Meteor Radar (MR) having spatial resolution of 2km, (b). Radiosonde (RS), with 0.3km spatial resolution, (c). Horizontal Wind Model (HWM), with 1km spatial resolution

The dot product of the wavenumber and velocity vectors is the Doppler shift. Comparison of Doppler shifts between IFE and MR are shown in Fig. 5.11(a). At an altitude of 70km to 76km, a percentage difference in the Doppler shift is observed of 0 to 3. When retrieved from 24km to 32km heights, GPS Radiosonde Doppler profiles are in good agreement with IFE Doppler profiles, as shown in Fig. 5.11(b). There is a 0 to 0.2 percentage difference between the Doppler shift estimated by Radiosonde and the Instantaneous Frequency Estimation. At altitudes between 30km and 80km, the Doppler shift can vary between 0mHz and 1.7mHz. For a comparison of the Doppler shift between HWM [113] and IFE is given in Fig. 5.11(c). IFE is cross-correlated with MR and RS by 0.72 and 0.85, and can be concluded from Fig. 5.12(a), and Fig. 5.12(b) respectively. Doppler shift profiles derieved from wind velocities are compared with Doppler shifts retrieved from HWM is shown in Fig. 5.12(c). In terms of cross-correlation, HWM and IFE are 0.7.

5.6.6 Addressing of background sources

When checking against background removal, the relative amplitude threshold (a_e) [114, 112] defined as the wave amplitude required for thresholding against instabilities, assuming a minimum Richardson number (R_i) greater than 0.25. The R_i greater than 0.25 ensures the absence of instabilities. For the case where $R_i \in [0, 0.25]$ corresponds to wave-induced fluctuations, then a single vertical temperature (or density) profile provides the intrinsic gravity wave frequency. The expression for amplitude threshold is given by

$$a_e = \frac{g|m|}{N^2} . \frac{|T'(z,t)|}{|T_0(z,t)|}; a_e \in [0,1] \tag{5.15}$$

In order to consider a wind removal as successful, the remnants of wind ($\langle T'(z,t) \rangle > 0$) are smaller than the errors introduced by the instrument and the algorithm used for the temperature retrieval. Precision and accuracy describe the error budget of an instrument, where precision describes noise-like errors that affect measured parameters randomly.Precision and accuracy describe the error budget of an instrument, where precision describes noise-like errors that affect measured parameters randomly. As the background is removed along with the errors characterized by accuracy, the uncertainty is more closely related to precision than to accuracy. The method used for temperature retrieval for the present work is $KKTl_2$ algorithm [100]. The detailed procedure for obtaining an uncertainty budget is given in [115]. If the visible spectrum wave lengths are being used, uncertainty sources include detection of the signals, correction of saturation, extraction of background noise, and tying on the temperature to the top of the profiles, and ozone absorption. In addition to molecular extinction, gravity acceleration, and dry air molecular mass, which are all determined by the instrument and data processing algorithm, the magnitudes of these factors depend on instrument and algorithm [115], and the altitude of interest. Equation 5.16 ensures that background wind is removed, and criteria are defined in [116].

$$< T'(z,t) >_{remnant} << \Delta T_S + \Delta T_R + \Delta T_M \tag{5.16}$$

where ΔT_S represents systematic uncertainty, ΔT_R represents random uncertainty, and ΔT_M represents atmospheric temperature retrieval algorithm uncertainty.

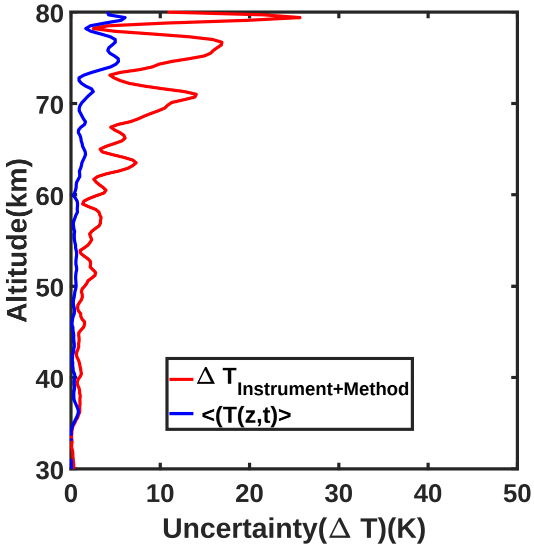


Figure 5.13: Temperature uncertainty budget for data of 21 April 2014.

Table 5.4: NARL Rayleigh Lidar System Uncertainty Budget. $(.)/(.)$ denotes the degree of uncertainty in high and low channels of Rayleigh Lidar.

Source	80km	70km	50km	30km
Detection Noise(N_D)	6K/(.)	5K/(.)	1K/4K	(.)/0.08K
Background Noise(N_B)	0.5K/(.)	0.3K/(.)	0.3K/(0.8)	(.)/0.02K
Seed Value	4K	2K	0.5K	-
Gravity(g)	0.1K	0.1K	0.1K	0.1K
Molecular Mass(M_{dryair})	1K	0.05K	0.05K	0.05K
$KKTl_2$ method	15K	8K	3K	0.08K
Total Uncertainty	27K/40K	16K/(.)	4.8K/8.45 K	0.08K/0.33K

Temperature perturbations free of wind effect and the uncertainty associated with the instrument and algorithm for temperature retrieval [100] can be seen in Fig. 5.13 and Table 5.4. The methodology for the calculation of uncertainty due to different sources in the temperature profile is estimated using the procedure mentioned in [115]. There are two major sources of uncertainty: detector noise and the algorithm used for temperature retrieval. At 80km, uncertainty reaches a maximum of 20K to 40K. Dry air molecular mass and gravity are considered negligible sources of uncertainty in temperature.

Table 5.5: Seasonal variations of the Doppler shift distribution function in 2014.

Season	Winter	Summer	Southwest Monsoon
No. of. Days	11	7	9
Vertical Wavelength(km)	2, 3, 5 to 7, 8, 10 to 15	8 to 12, 13 to 15	7 to 12
f_{dmax}(mHz)	1.7	1.7	1.7
f_{dmin}(mHz)	0.376	0.383	0.322
μ(mHz)	1	1.04	1.01
σ(mHz)	0.24	0.226	0.23
Skewness	0.186	-0.102	-0.08
Kurtosis	-0.39	-0.14	-0.289

μ and σ represent average and standard deviation respectively.

5.7 Effect on Vertical Wavenumber in Reducing Doppler Shift

The wind free profiles were reconstructed using equation 5.13. To examine the effect of the reduction of Doppler shift on the vertical wavenumber, a relative comparison of reference vertical wavenumber power spectrum called source spectrum has to be made, with the resultant vertical wavenumber spectrum calculated which is free from background wind effect. The reference altitude can be chosen between the altitude range of 30km to 80km. For the present work, the reference altitude is considered 45km. The source spectrum is the spectrum at a reference altitude of 45km. Using Fourier transforms, we calculate it at $z = 45km$ at all sampling intervals. In the same way, the resultant spectrum consists of the Fourier transformation of the vertical perturbation profile at all intervals obtained before and after IFE at altitude $z = z1$, where the background wind is $U = U(z1)$, $25km \leq z1 \leq 80km$. Fig. 5.14 represents the relation between a wavenumber (m_0) with background wind U_0 at the reference level and the wavenumber (m) obtained after

Doppler Shift compensation, where m_c represents the critical wavenumber. For the present data, the critical wavenumber is $0.5km^{-1}$. Here the notation m_0 indicates the parameter at the reference altitude.

The refraction of wavenumbers is one of the dynamic processes that govern the vertical wavenumber power spectrum. The reduction in the vertical wavenumber is studied by

$$m = \frac{N}{|c_h - \bar{U}\cos(\phi)|} = \frac{N}{|c_h - \frac{2\pi f_d}{k}|} \tag{5.17}$$

where c_h represent horizontal phase speed and f_d is Doppler shift obtained through IFE. Refraction is responsible for this reduction in wavenumber, refraction is responsible for this reduction in wavenumber, which increases ω. The strongest effect of refraction is observed near the neighborhood of critical wavenumber(m_c) and the marked spectral tail is produced as shown in Fig. 5.14. For the resultant spectrum, in the region between $[m/m_c, m_M/m_c]$ and $m_M > m$ conservative Doppler shift dominates the spectrum. If the wavenumber(m) is greater than a wavenumber(m_M), the effects of instabilities start prevailing. From the figure, in the resultant spectrum, the value of m/m_c is between 10 to 25, the region where the Doppler shifting is dominant. For the Doppler wind compensated wavenumber, the critical wavenumber shifts left towards the region where the turbulence and dissipation of energy are dominant in the source wavenumber spectrum.

The seasonal variations of Doppler Shift and the intrinsic frequency are shown in Fig. 5.15(a), 5.15(b), 5.15(c) and 5.15(d), 5.15(e), 5.15(f) respectively. The Doppler shift distribution is inferred from the histogram using a normal distribution. Skewness and Kurtosis with 99.9 confidence intervals are given in Table 5.5. Winter is observed at the observational site from December to February, summer from March to May, and the North-East monsoon from June to August. The South-West monsoon occurs from September to November). The North East Monsoon season is not provided with data due to the non-availability of data. Seasonally, the intrinsic frequency of the summer season is perfectly correlated with altitude. The statistics in Table 5.5 indicate a clear deviation from normal distributions for Doppler shift. The following inferences are made from Table 5.5.

1. During the South West Monsoon season, the minimum Doppler shift occurs, and during all seasons, the maximum is 1.7mHz.

2. During the winter season, the Doppler shift's standard deviation is higher, while during the summer, it is lower.

3. Doppler shift distribution curves are moderately skewed in all seasons. The Doppler

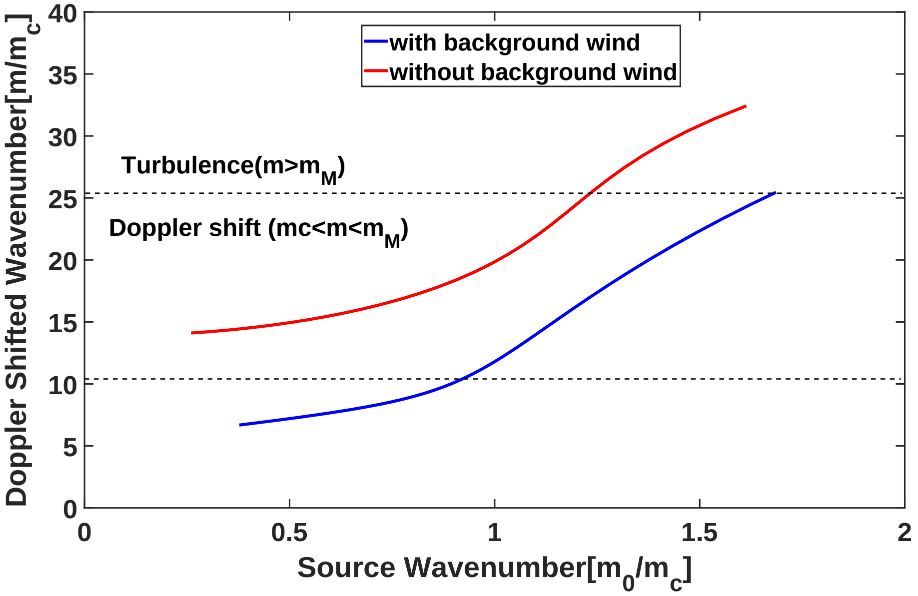


Figure 5.14: The refracted wavenumber(m) as a function of the source wavenumber(m_0) for data on 21 April 2014. The turbulent and Doppler-shifted regions are marked based on the vertical wavenumber without background wind compensation. The value of m/m_M is 25 and 35 with and without background wind reduction respectively.

Shift profiles in the South West Monsoon season follow a normal distribution function when the skewness is minimum. A large variation in the horizontal wind causes the distribution function of the Doppler shift profiles to deviate from normal in the Winter season.

A detailed study is conducted on reduction of background wind on the seasonal variations of vertical wavenumber. In the Winter Season, due to the background wind, the instabilities start prevailing at a lower value of m_M than the waves without background wind. From Fig. 5.14 it is observed that because of the presence of background wind, the waves will become absorbed ($m_0 > m_M$) at lower vertical wavenumbers resulting in loss of information about actual characteristics of Gravity Waves. The results are consistent with the numerical studies and Gravity wave studies at the Gadanki location. The difference in vertical wavenumber between seasons of Summer and South West Monsoon is smaller because the distribution of Doppler shift follows a normal pattern in these seasons.

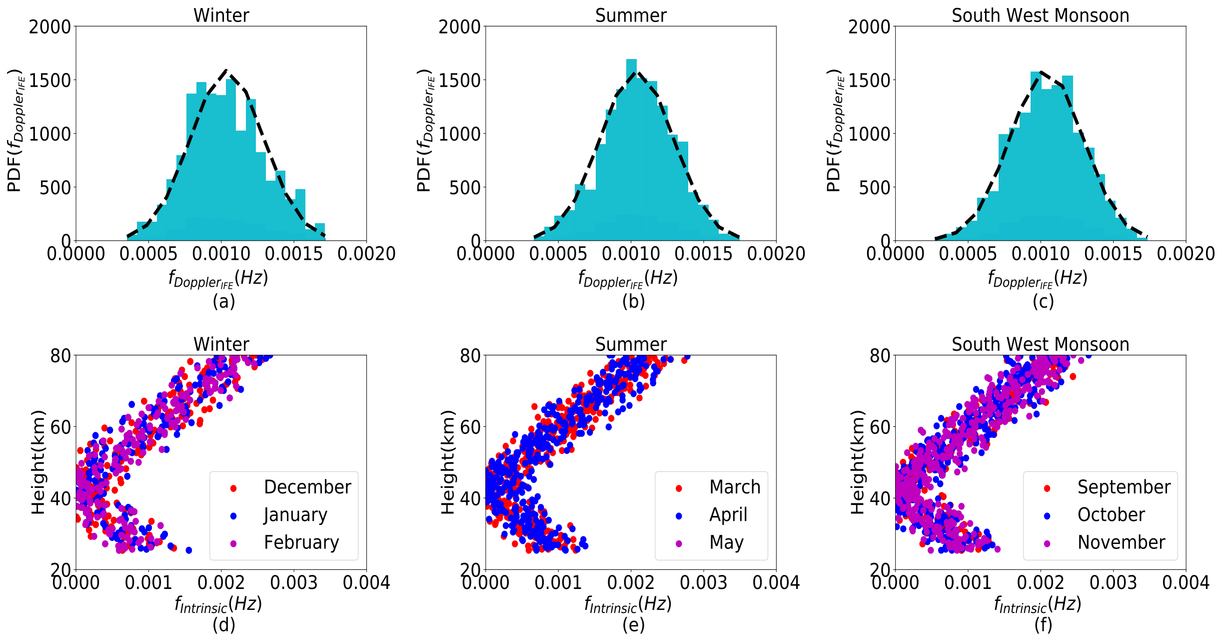


Figure 5.15: Seasonal distribution of Doppler shift for data on 2014. (a) Winter, (b) Summer, and (c)South West Monsoon. The relationship between intrinsic frequency and altitude during the (d) Winter, (e) Summer, and (f) South West Monsoon seasons.

5.8 Summary

1. Using MRDL and IFE, we distinguish the effect of background wind on temperature perturbations.
2. Wavelet transforms are combined with learned dictionaries to produce a Multi-Resolution Dictionary Learning algorithm.
3. The Dictionary is trained in discrete wavelet bands. As a result, learning is simple and efficient. In the sparse coding scheme, the MRDL is incorporated.
4. Temperature perturbations are checked after reconstruction for remnants of background wind.
5. When compared to other seasons, the Doppler shift values in the Winter season differs significantly from the Normal distribution.
6. A study is conducted on seasonal variations of wind free temperature perturbations to assess vertical wavenumber patterns.

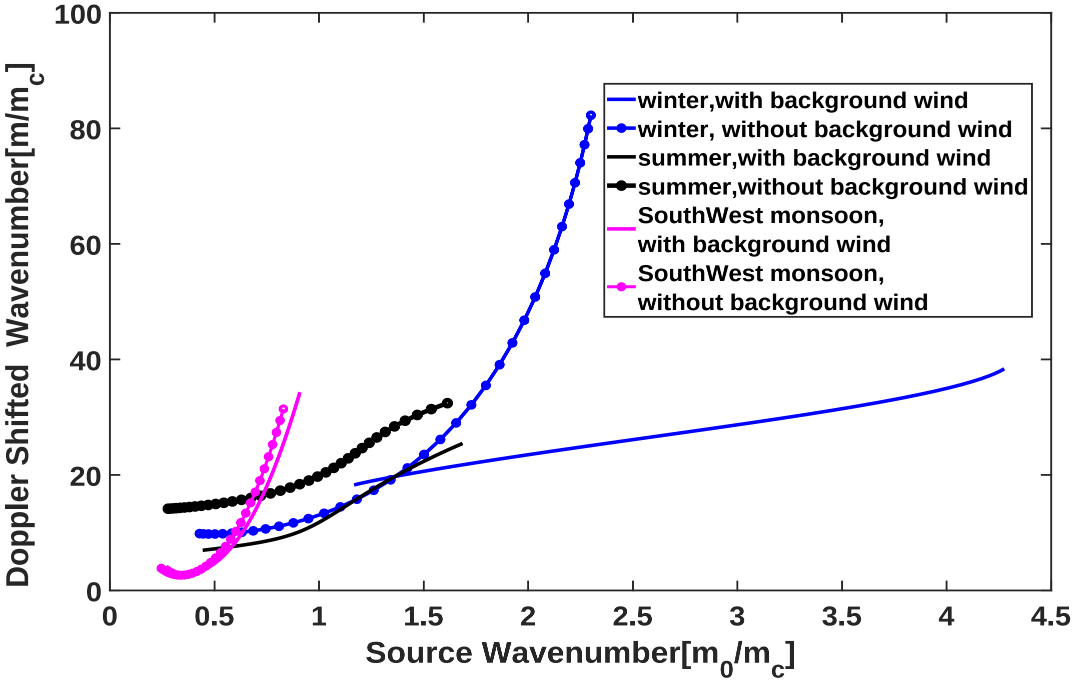


Figure 5.16: Seasonal variations of Vertical Wavenumber(m) as a function of the source wavenumber(m_0) for the data colleceted in 2014. Waves with $m_0 > m_c$ should be interpreted as the prevailing absorption of waves through critical layers.

Chapter 6

Deep Kernel Dictionary Learning

Introduction of Kernel functions in the KSVD technique and conventional Dictionary Learning techniques improve the classification accuracy. The Kernel Dictionary Learning technique is used for solving non-linear problems having a high computational cost. It has been proven in [117] that non-linear Dictionary Learning techniques are superior to their linear counterparts.

6.1 Kernel Mapping

The Kernel mapping can be seen as a transformation to a higher dimension where the problems become linear. The function $\Psi : X \to Y$ represent the information of its domain(X) and its co-domain(Y). The widely used definition of function $\Psi : X \to Y$ Ψ is a subset of $X \times Y$ consisting of all the pairs $(x, \Psi(x))$ for $x \in X$. A linear classifier is used to solve non-linear problems with the Kernel, with help of the Kernel trick. The kernel functions to transform the inseparable data to the linearly separable data and transform them into higher dimensions. The mathematical representation of the Kernel trick is given by

$$K(X, Y) =< \Psi(X), \Psi(Y) > \tag{6.1}$$

where ‘K’ is kernel function, and X, Y are of n-dimensional. ‘Ψ’ represent the feature map from n-dimension to m-dimension space ($m > n$). Any learning algorithm that can be represented as the inner product of data can run kernel function and defined as kernel method.

6.2 Kernel Approximations

The benefit of using the feature maps obtained explicitly in comparison with Kernel trick is that they reduce the computational cost of learning when working with large data sets.

6.2.1 Mercer's Theorem

The Mercers theorem can characterize which functions can be used as Kernel function. It defines that any symmetric positive-definite functiom can be represented as “square of sum of a convergent sequence of product functions". First, the Mercers theorem is explained with a for a continuous Kernel as given by

$$K : [a, b] \times [a, b] \rightarrow \mathbb{R} \tag{6.2}$$

where symmetric means that $K(x, y) = K(y, x)$ for all $x, y \in [a, b]$.

The Kernel semi-definiteness is guaranteed if the kernel satisfies the below equation.

$$\sum_{i=1}^{n} \sum_{j=1}^{n} K(x_i, x_j) c_i c_j \geq 0, \forall x_1, x_2 ..., x_n \in [a, b] \tag{6.3}$$

where $c_1, ..., c_n$ are arbitrary real numbers. Associated with the Kernel (K), there exists a linear operator on functions defined by equation 6.4.

$$[T_K \psi](x) = \int_a^b K(x, s) \psi(s) \, ds, \psi \in L_2[a, b] \tag{6.4}$$

$$(g, Kg) = g^T \cdot Kg = \sum_{i=1}^{N} \sum_{j=1}^{N} g_i \, K_{ij} \, g_j \geq 0 \tag{6.5}$$

where g_i represent a positive semi-definite. **Discrete analogy of Mercer's Theorem** The definition of Mercers theorem in the discrete domain (6.5) is equivalent to a positive semi-definite matrix. The matrix$K$ of dimension N, that satisfies equation 6.5

6.2.2 Representer Theorem

$$f \mapsto E\left((x_1, y_1, f(x_1)), ..., (x_n, y_n, f(x_n))\right) + g\left(\|f\|\right). \tag{6.6}$$

$$f^* = \text{argmin}_{f \in H_k} \left\{ E\left((x_1, y_1, f(x_1)), ..., (x_n, y_n, f(x_n))\right) + g\left(\|f\|\right) \right\} \tag{6.7}$$

$$f^*(\cdot) = \sum_{i=1}^{n} \alpha_i k(\cdot, x_i), \alpha_i \in \mathbb{R} \forall 1 \leq i \leq n \tag{6.8}$$

The Representer theorem states that a minimizer f^* of a regularized empirical risk functional defined over an RKHS(H_k) can be expressed as finite linear combination of kernel products evaluated on the input points in the training set data($(x_1, y_1), \ldots, (x_n, y_n) \in \mathcal{X} \times \mathbb{R}$) that as a strictly increasing function and an arbitrary error function $E\colon\ (\mathcal{X} \times \mathbb{R}^2)^n \to \mathbb{R} \cup \{\infty\}$ as represented in equations 6.7 and 6.8. Representer theorems are practically useful as they dramatically reduce the computational complexity of empirical risk minimization problems.

6.3 Kernel Dictionary Learning

Using Kernel theory, Dictionary Learning can be expressed as follows.

$$\kappa\left(\mathbf{x}, \mathbf{x}'\right) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle = \Phi(\mathbf{x})^T \Phi(\mathbf{x}') \tag{6.9}$$

$$\mathbf{K}_{i,j} = \kappa(\mathbf{x}_i, \mathbf{x}_j) = \left\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \right\rangle, \quad \forall i, j = 1..N. \tag{6.10}$$

$$\underset{\mathbf{A},\mathbf{\Gamma}}{\text{argmin}} \|\Phi(\mathbf{X}) - \Phi(\mathbf{X})\mathbf{A}\mathbf{\Gamma}\|_F^2 \quad s.t. \quad \forall i = 1..N \quad \|\boldsymbol{\gamma}_i\|_0 \leq q. \tag{6.11}$$

As a result of using the feature space, Dictionary Learning optimization can be represented by equation (6.11).

6.4 Reproducing Kernel Hilbert Spaces

RKHS is the space relating to a kernel that generates all functions in the space(X). In the RKHS, all evaluation functions are continuous.

$$L_x : f \mapsto f(x) \, \forall f \in H. \tag{6.12}$$

$$f(x) = L_x(f) = \langle f,\ K_x \rangle_H \quad \forall f \in H. \tag{6.13}$$

Assume X be an arbitrary set and H is a Hilbert space of real valued functions on X. At a point x, this function is evaluated over the Hilbert space of functions as represented in equation 6.12. Another definition of the RKHS is through representing the evaluation function as the inner product of f with a function K_x in H. Reproducing kernels are characteristic functions for Hilbert spaces H from which the RKHS expressed in equation 6.13. Since K_x is a function defined on $X \in \mathbb{R}$ and as K_x is in H we have the transformed equation 6.14.

$$K_x(y) = L_y(K_x) = \langle K_x,\ K_y \rangle_H \tag{6.14}$$

where $K_y \in H$ is the element in H associated to L_y. The reproducing kernel can be defined by H as a function $K : X \times X \to \mathbb{R}$ by $K(x, y) = \langle K_x,\ K_y \rangle_H$. Kernel function($K : X \times X \to \mathbb{R}$) in terms of inner product is represented by

$$\begin{aligned} \sum_{i,j=1}^{n} c_i c_j K(x_i, x_j) &= \sum_{i=1}^{n} c_i \left\langle K_{x_i}, \sum_{j=1}^{n} c_j K_{x_j} \right\rangle_H \\ &= \left\langle \sum_{i=1}^{n} c_i K_{x_i}, \sum_{j=1}^{n} c_j K_{x_j} \right\rangle_H \\ &= \left\| \sum_{i=1}^{n} c_i K_{x_i} \right\|_H^2 \geq 0 \end{aligned} \tag{6.15}$$

where $n \in \mathbb{N}, x_1, \ldots, x_n \in X,$ and $c_1, \ldots, c_n \in \mathbb{R}$.

6.4.1 Sampling Techniques

The sampling techniques are classified into two types: data-independent and data-dependent techniques. The sampling techniques that depend on data are called Explicit Kernel Approximations and those that are independent of data are Implicit Kernel Approximations.

$$k(\mathbf{x}, \mathbf{x}') = \langle \psi(\mathbf{x}), \psi(\mathbf{x}') \rangle_{\mathcal{V}} \tag{6.16}$$

Here the constraint is that the $\langle \cdot, \cdot \rangle_{\mathcal{V}}$ must be a proper inner product. The kernel trick avoids the usage of explicit mapping techniques required for working with linear learning algorithms to learn a nonlinear function. Using the Kernel trick, the kernel is expressed in the form of a "feature map" $\psi \colon \mathcal{X} \to \mathcal{V}$ which satisfies equation 6.16

6.4.1.1 Implicit Kernel Approximations

1. **Random Fourier Feature Generation:** The Cosine and Sine functions are used as basis functions in Random Fourier Feature generation. The sequence of steps involved in the Random Fourier Feature generation [118] is given in Algorithm 1.

Algorithm 6.1: Random Fourier Feature Generation

```
(a) Input: Consider a shift-invariant Kernel $K(x, y) = K(x - y)$.
(b) Compute: Fourier Transform $p$ of the Kernel $K$:
        $p(\omega) = \frac{1}{2\pi} \int_{R^d} \exp -j\omega^T K(\delta d\Delta)$
(c) Draw 'L' identical and independent samples ($\omega_1, \omega_2 \dots \omega_L \in R^d$) from 'p'.
(d) Return: $z(x) = \frac{1}{\sqrt{L}}[\cos \omega_i^T, \sin \omega_i^T]$ for all $1 \leq i \leq L$, $z \in R^{2L}$.
```

2. **Quasi-Random Fourier Feature Generation:** The Quasi-Random Fourier feature [119] is a variant of Random Fourier Feature generation through Monte-Carlo simulations. In this technique, we draw samples in step(c) of Algorithm 1 using Monte Carlo simulations.

3. **Random Maclaurin Feature Maps** The Maclaurin Feature maps are suitable for polynomial kernels, and the detailed algorithm for the implementation is given in [120].

4. **Compact Random Feature Maps** The data is encoded to a very high dimensional space using like Random Maclaurin Feature Map and Johnson-Lindenstrauss random projections. The detailed algorithm for implementation of Compact Random Feature Maps is given in [121].

```
Algorithm 6.2: Quasi Random Fourier Feature Generation
    (a) Input: Consider a shift-invariant Kernel $K(x, y) = K(x - y)$.
    (b) Compute: Fourier Transform $p$ of the Kernel $K$:
            $p(\omega) = \frac{1}{2\pi} \int_{R^d} \exp -j\omega^T K(\delta d\Delta)$
    (c) Generate: low discrepancy sequence $t_1, t_2, ..t_L$
    (d) Draw 'L' identical and independent samples ($\omega_1, \omega_2 \dots \omega_L \in R^d$) using $\omega = \psi^{-1}(t)$.
    (e) Return: $z(x) = \frac{1}{\sqrt{L}}[\cos \omega_i^T, \sin \omega_i^T]$ for all $1 \leq i \leq L$, $z \in R^{2L}$.
```

6.4.1.2 Explicit Kernel Approximations

These types of Kernel approximations are data-dependent. This kind of technique can be used for any kernel functions while the earlier mentioned techniques for feature generation are suitable for shift-invariant kernel and polynomial Kernels respectively. The well-known technique of the Explicit Kernel Approximation is the Nystroem Method [122, 123, 124]. The basis function in the case of Nystroem method is obtained by sampling the training example, unlike data-independent sine and cosine basis functions. Relative comparison between the Nystrom method and Random Fourier Feature generation is reported in [123].

6.4.2 Properties of Kernel Functions

Consider any Space X and $K_1(.,.)$, and $K_2(.,.)$ over the space X. Then $K(.,.)$ is a kernel with the following properties being satisfied.

1. $k(x, y) = K_1(x, y) + K_2(x, y)$
2. $k(x, y) = aK_1(x, y)$ where $a > 0$
3. $k(x, y) = f(x).f(y)$ for any function f on X
4. $k(x, y) = K_1(x, y).K_2(x, y)$
5. $k(x, y) = \frac{K_1(x,y)}{\sqrt{K_1(x,x)\sqrt{K}_1(y,y)}}$

6.5 Inclusion of Kernels in Dictionary Learning

A discussion of introducing kernels in dictionaries [117, 125, 126] and suggests a method to learn deep kernel dictionaries. Let $\Psi : \mathbb{R}^n \rightarrow \mathbb{F}$ translate $\mathbb{R}^n$ through $\mathbb{F}$. To carry out the feature map explicitly in higher dimensional spaces, Mercer Kernels are commonly used. The Mercer Kernel is a mathematical expression $\kappa(\mathbf{x}, \mathbf{y})$ mapped onto $\mathbf{y_i}$ to produce a semi-definite matrix $\kappa(\mathbf{y_i}, \mathbf{y_j})$. The dictionary model used in this study is $\mathbf{D} = \mathbf{BA}$, where $\mathbf{B}$ represents the predefined basis, and $\mathbf{A}$ is an adaptive dictionary. Data $\mathbf{Y}$ in the feature space is represented as $\boldsymbol{\Psi}(\mathbf{Y})$ and a learned Dictionary is represented as $\boldsymbol{\Psi}(\mathbf{D})$. Feature space learned dictionary can be found by

$$\boldsymbol{\Psi}(\mathbf{D}) = \boldsymbol{\Psi}(\mathbf{Y})\mathbf{A} \tag{6.17}$$

The mathematical expression for KDL is represented below. In [117] reported that non-linear Dictionary Learning is better at discriminating features than linear Dictionary Learning.

$$\underset{\mathbf{A},\mathbf{X}}{\text{argmin}} \|\Psi(\mathbf{Y}) - \Psi(\mathbf{Y})\mathbf{AX}\|_F^2 \quad s.t. \quad \forall i = 1..n \quad \|\boldsymbol{X}_i\|_0 \leq T_0. \tag{6.18}$$

where T_0 represents the sparsity.

$$\|\Psi(\mathbf{Y}) - \Psi(\mathbf{Y})\mathbf{AX}\|_F^2 = \mathbf{tr}((\mathbf{I} - \mathbf{AX})^T \mathbb{K}(\mathbf{Y}, \mathbf{Y})\ (\mathbf{I} - \mathbf{AX})) \tag{6.19}$$

$$[\mathbb{K}(\mathbf{Y}, \mathbf{Y})]_{ij} = [\langle \Psi(\mathbf{Y}), \Psi(\mathbf{Y}) \rangle]_{ij} = \kappa(y_i, y_j) \tag{6.20}$$

KDL's classical optimization problem (6.18) is solved using MOD or KSVD, and the objective function is expressed in equations(6.19) and (6.20).

6.6 Deep Dictionary Learning

Atmospheric Remote Sensing has seen a significant increase in application of Deep Learning techniques, as these technologies provide significant solutions for both feature extraction and classification tasks. It is possible to use Deep Learning in a wide variety of atmospheric remote sensing applications, including detecting turbulence [127], estimating atmospheric gravity wave parameters [40], retrieving parameter profiles [128], detecting the height of atmospheric boundary layers [129], improving air quality prediction [130], and forecasting [131].

Through the introduction of kernels, DL is applied to nonlinear problems, resulting in more straightforward classification and regression tasks if the data are projected to higher dimensional feature space. In [117] it is reported that kernels in sparse representations have been introduced into the KSVD Dictionary Learning, called Kernel KSVD (KKSVD) and this is the foundation for the proposed method. As part of KKSVD, a "Base Dictionary" is learned in which all signals are represented, and a "Coefficient Dictionary" in which the Base Dictionary is updated.

Complementing Deep Learning with Dictionary Learning (DDL) [132] , uses a multi-layer feature extraction. The Sparse matrix ($\mathbf{X}$) and Dictionary ($\mathbf{D}$) are learned from the output of each Neural Network. The Sparse matrix obtained at the previous level serves as input data to the Neural Network. When the Dictionary is learned from the individual layers, it performs differently from the Dictionary learnt from the overall neural network model. An increase in layers increases the likelihood of overfitting. BN is used at the end of each layer to overcome the over-fitting problem.

The Deep Dictionary Learning algorithm [132], a recent algorithm introduction, enhanced the capability of LSTM Autoencoders, Convolution Neural Networks, and K-SVD Dictionary Learning. It is more efficient to detect sudden and short-duration abnormal events by including Kernels into Deep Dictionary Learning. A case study using satellite data is provided to demonstrate the significance of the proposed method in detecting Wave Breaking events (AURA-Microwave Limb Sounder). Data from ground-based RL and MR measurements are used to validate the proposed methodology for Wave Breaking event detection.

6.6.1 Greedy Deep Dictionary Learning

In the work reported in [133], the representation of Dictionary Learning methodology in terms of Deep Learning and using of basic blocks for Dictionary Learning for creating Deep Neural Networks are explained. In this technique, the Deep architecture is not learned in a single step, is learned greedily [134]. The learning of Dictionaries in a greedy fashion has the following advantages.

1. Learning with basic blocks is easy.

2. Selection of learning parameters for learning the network is easy. It also avoids the problem of overfitting the parameters.

Ideally, we solve the optimization problem below

$$\min_{D_1,\ldots D_N,Z} \|Y - D_1\Phi(D_2\Phi(\ldots\Phi(D_N Z)))\|_F^2 + \|X\|_1 \tag{6.21}$$

We solve the optimization problem, layer by layer. For the 1^{st} layer the problem is

$$Y = D_1 X_1; \;\; X_1 = \Phi(D_2(\Phi(D_3 \ldots \Phi_{N'} X))) \tag{6.22}$$

After learning the coefficients for the 1^{st} layer, the second layer can be learned as a single layer of Dictionary Learning as expressed by

$$\Phi^{-1}(X_1) = D_2 X_2 \tag{6.23}$$

Repeating the same process upto the penultimate layer, we get,

$$X_{N-1} = \Phi(D_N X) \Rightarrow \Phi^{-1}(X_{N-1}) = D_N X \tag{6.24}$$

At the last layer of Neural Network, the optimization problem is given below and is solved using alternating minimization.

$$\min_{D_N,X} \|\Phi^{-1}(X_{N-1}) - D_N X\|_F^2 + \mu\|X\|_1 \tag{6.25}$$

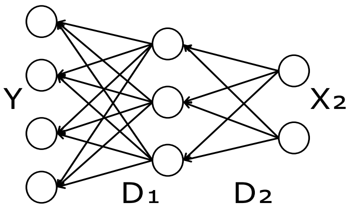


Figure 6.1: Deep Dictionary Learning Schematic Diagram

6.7 Deep Kernel Dictionary Learning Methodology

Sub-section 6.6 discussed Deep Learning's relationship with Dictionary Learning, the introduction of Kernels, and Deep Dictionary Learning's techniques. In Algorithm 6.3, the DKDL procedure is given for any sort of activation or kernel function. Even when there is a linear activation function, there is no single dictionary that can represent dictionaries at

Algorithm 6.3: 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.
        $X_1 \leftarrow \min_{X} ||Y - \Psi_1(\Phi_1(D_1X))||_F^2;\ D_1 \leftarrow \min_{\Psi_1(D_1)} \left\|Y - \Psi_1(\Phi_1(D_1X))\right\|_F^2$
    2. For all the levels (ie., the effective Deep Kernel Dictionary Learning problem), repeat until convergence.
        $Y = \Psi_{N'}\Bigg(Y_{N'}..\bigg(\Psi_2\Big(Y_2(\Psi_1(Y_1))\Big)\bigg)\Bigg)$
        $X'_N \leftarrow \min_{X'_N} \left\|Y - \Psi_1(D_1)\Phi_1\left(\Psi_2(D_2)\Phi_2\Big(\dots\Psi_{N'}\Phi_{N'}\Big(D_{N'}X\Big)\Big)\right)\right\|_F^2 + \lambda\left\|X_{N'}\right\|_1$
        $D'_N \leftarrow \min_{\Psi_{N'}(D_{N'})} \left\|Y - \Psi_1(D_1)\Phi_1\left(\Psi_2(D_2)\Phi_2\Big(\dots\Psi_{N'}\Phi_{N'}\Big(D_{N'}X\Big)\Big)\right)\right\|_F^2 + \lambda\left\|X_{N'}\right\|_1$
```

Algorithm 6.4: Linear Activation Function and Kernel Function Testing

```
    1. Getting an effective Dictionary from individual Dictionaries.
        $D = \Psi_1(D_1)\Psi_2(D_2)...\Psi_{N'}(D_{N'})$
    2. Implement the Sparse coding for the test sample data($Y_{test}$)
        $X_{test} = \min_{X_{test}} ||Y_{test} - DX_{test}||_F^2 + \lambda|||X_{test}||_1$
```

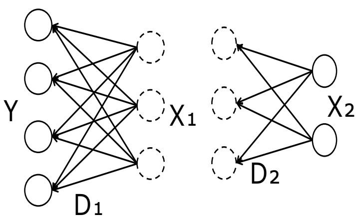


Figure 6.2: Layer-wise schematic diagram of Deep Dictionary Learning

Algorithm 6.5: Testing Algorithm for non-linear activation

```
1. Generate features for the first level
       $X_1 \leftarrow \min_{X} ||Y_{test} - \Psi_1(\Phi_1(D_1X))||_F^2;$
2. Generate features for the final level
       $X_{test} = \min_{X_{test}} \left\| Y_{test} - \Psi_1(D_1)\Phi_1 \left( \Psi_2(D_2)\Phi_2 \left( \dots \Psi_{N'}\Phi_{N'} \left( D_{N'}X_{test} \right) \right) \right)_F^2 + \lambda \|X_{test}\|_1$
```

multiple levels since the Dictionary Learning is bi-linear. As an example, if the sample has a dimensionality of m, if the first dictionary has a size of $m \times n_1$ and the second dictionary has a size of $n_2 \times n_3$, then we can get an effective dictionary of size $m \times n_3$.

Kernel functions can be included in two ways: either using the Kernel Trick or explicitly using techniques reported in [135, 123, 122] that use feature maps. The Kernel Trick is not as efficient when working with large data sets as explicit feature maps. Instead of using the Nystroem method for Kernel approximation, we use the Kernel Trick.

Implementation of (6.17) and (6.20) in (6.22) and (6.21) leads to the DKDL Algorithm 6.3 and KSVD is used to solve the optimization problem. A single effective Dictionary can be obtained by combining the kernel transformed dictionaries. Procedure to get an effective dictionary is given in Algorithm 6.4 and l_1-minimization is used to obtain the features. The testing phase cannot be implemented using a single dictionary for non-linear activation functions. Tests are conducted according to the procedure described in Algorithm 6.4.

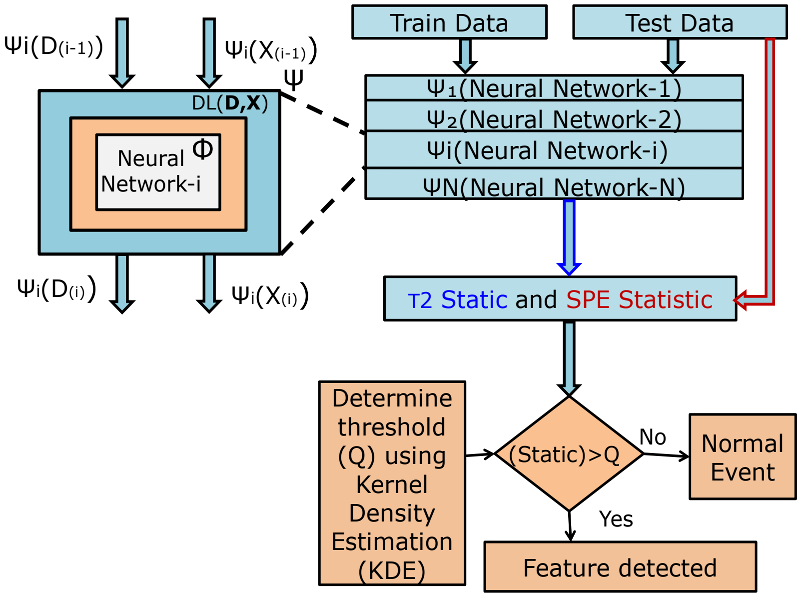


Figure 6.3: Feature detection with Deep Kernel Dictionary Learning. T^2 and SPE metrics exceeding the threshold limit denote a feature has been detected. In Deep Dictionary Learning, the highlighted block represents how kernels are incorporated.

6.7.1 Relevance To Existing Algorithms

The proposed DKDL is compared here with CLSTM, KCLSTM, and a generalized Auto-Encoder. The configuration of neural networks used in each differs. CLSTM, KCLSTM, and DKDL use different Kernel functions. The DKDL algorithm 6.6 is an variant of Auto-Encoder. Alternatively, we can find sparse representations through learning a dictionary that is similar to denoising AE as reported in [132]and represents a linearly activated hidden layer. Using Algorithm 6.3, we get the following mathematical expression for Dictionary Learning as represented by

$$\hat{Y} = \Psi_1(D_1)\left(\Psi_2(D_2)\left(\ldots\Psi_{N'}\left(D_{N'}X\right)\right)\right) = D_S X \tag{6.27}$$

$$\left\|Y - \hat{Y}\right\|_F^2 + \lambda \left\|X\right\|_1 \tag{6.28}$$

where D_S and D_A denote the effective dictionary for linear activation functions in the synthesis domain and analysis domain, respectively. By ignoring the sparsity term of (6.28), the optimization problem is given by below equation that contains a denoising auto-encoder [136].

$$||Y - D_S D_A \hat{Y}||_F^2 \tag{6.29}$$

6.7.2 Kernel Maps of Data Independence

This groundbreaking idea originated from the work described in [137]. Our work uses Radial Basis Functions (RBFs) that are data independent. The Fourier Transform of the RBF kernel $\kappa(\Delta) = \exp(\frac{-\Delta}{2\sigma^2})$ represents the normal probability distribution $p_k(\omega) \sim \mathcal{N}(0,\, 1/\sigma^2)$, where $\Delta = ||x_i - x_j||_F^2$. As a result, the features correspond to 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. The relative advantages and disadvantages of data-independent and data-dependent Kernel maps is given in [138].

6.7.3 Limitations And Improvements

As the number of layers in a NN increases, solving optimization problem becomes difficult in Algorithm 6.3 increases and kernels must be explicitly approximated. This method is known as the Nystroem method.

6.7.4 Comparison to Existing Methods

The DKDL method is validated with CLSTM, KCLSTM, and DDL. Each method aims to determine the weights of the network, represent the data, and extract features. They differ from each other in linear, non-linear, and Dictionary Learning based feature extraction applications. There is no relation between the proposed methodology and the hierarchical or structured dictionary learning techniques. Generally, all DL techniques are single level learning techniques and relationships are between the atoms within the dictionary.

6.8 Detection of Atmospheric Gravity Wave Breaking

Through the MLS-AURA satellite data, the altitude of the atmosphere is calculated based on vertical temperature profiles ranging from 316hPa to 0.001hPa. From 2014 to 2019, all global gravity WB events were mapped to temperature perturbations using MLS. Preprocessing steps are carried on data to achieve a spatial resolution of 0.3km, the training data

```
Algorithm 6.6: Feature detection using Deep Kernel Dictionary Learning
Input: tolerence($\epsilon$), regularization parameter($\lambda$), 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 6.3,
       6.4, and 6.5.
    2. Calculate the $T^2$-metric and SPE metrics.
    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
```

has a size of 128, where rows represent altitude varying between 25.1km and 80km, and columns represent latitude variations (-82^oS to $+82^o$N) of Wave Breaking events.

A wave breaking event was observed on 21 April 2014, 7 April 2014, and 09 January 2015 in temperature perturbation and velocity perturbation obtained with RLS and MRS, respectively, and other test data were used to validate the proposed method. The test data is choosen on satellite overpass ($[13^oN, 14^oN]$, $[79^oE, 80^oE]$) over the Gadanki region. The

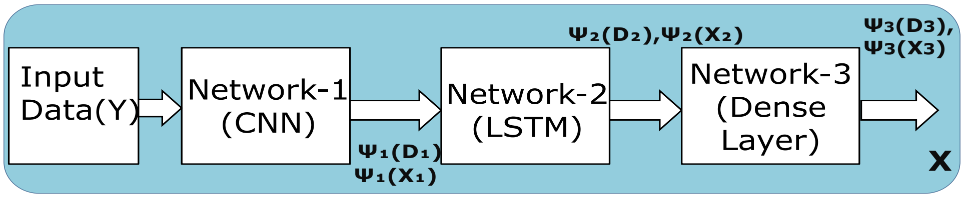


Figure 6.4: Architecture of DKDL to detect wave breaking events.

DKDL algorithm is applied on above mentioned data. Every day, the test data is 128×256, with rows indicating spatial resolution and columns indicating latitude as $13^o N$ to $14^o N$.

Using the previous literature pertaining to this study, we selected the grid search space for tuning hyper-parameters. In this case, the learning rate is set between $[10^{-3}, 10^{-4}]$, and the batch size and dropout parameters are determined accordingly. In order to learn convolution neural networks, linear activation functions α of 0.011. Details about the layers in the Neural Networks are given in Table 6.1 and a specification of the hyper-parameters used in NN training is given in Table 6.2. Radial Basis Kernel is used in the current analysis, and the regularization parameter λ is 0.01.

6.8.1 Application and Evaluation Of The Proposed Method

Ten epochs of training result in a stable loss, as shown from Fig. 6.6. Therefore, in the present experiment, the network will stop learning after 20 epochs. Each iteration uses 10 samples, which means 10 samples are used per batch. As performance indicators, the Kernel function is used to calculate DR and FDR, comparison is done with CLSTM and KCLSTM. CLSTM is rigorously used for feature detection. KCLSTM and CLSTM detect features well, but cannot capture significant ones. It is impossible to separate abnormal points from the original data. Using CLSTM and Kernel DL, all features can be detected with a DR of 100%.

The Fig. 6.8 shows the SPE metric and T^2 metric calculated using data from 21 April 2014. Wave-breaking events are detected when both metrics exceed the threshold limit. At 75km, the WB altitude is detected. For the purpose of explaining the method, the above-mentioned data is provided.

The Table 6.3 and 6.4 evaluate the proposed method under varying kernel approximation components. By increasing the number of kernel approximation components, feature detection should become more efficient. A Kernel approximation with size of 128, which takes 12.57 seconds to train, has the highest DR of 0.95. For the present study, 128 Kernel

Table 6.1: Specifications of input, and output to DKDL algorithm

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

Table 6.2: Least validation loss corresponding to regression models

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

approximation components are used, and both DR and FDR are assessed the same way.

Table 6.5 and 6.6 show the DR and Threshold limits for T^2 and SPE metric statistics. As a result of the experiments, DKDL appears to perform better than any other method in finding features, and inclusion of Kernels in DKDL improve the DR. Table 6.6 shows the thresholds for T^2 and SPE metrics calculated using KDE for each data set to detect wave breaking event. Ideally for a having a good DR, threshold limit must be high for T^2 and a

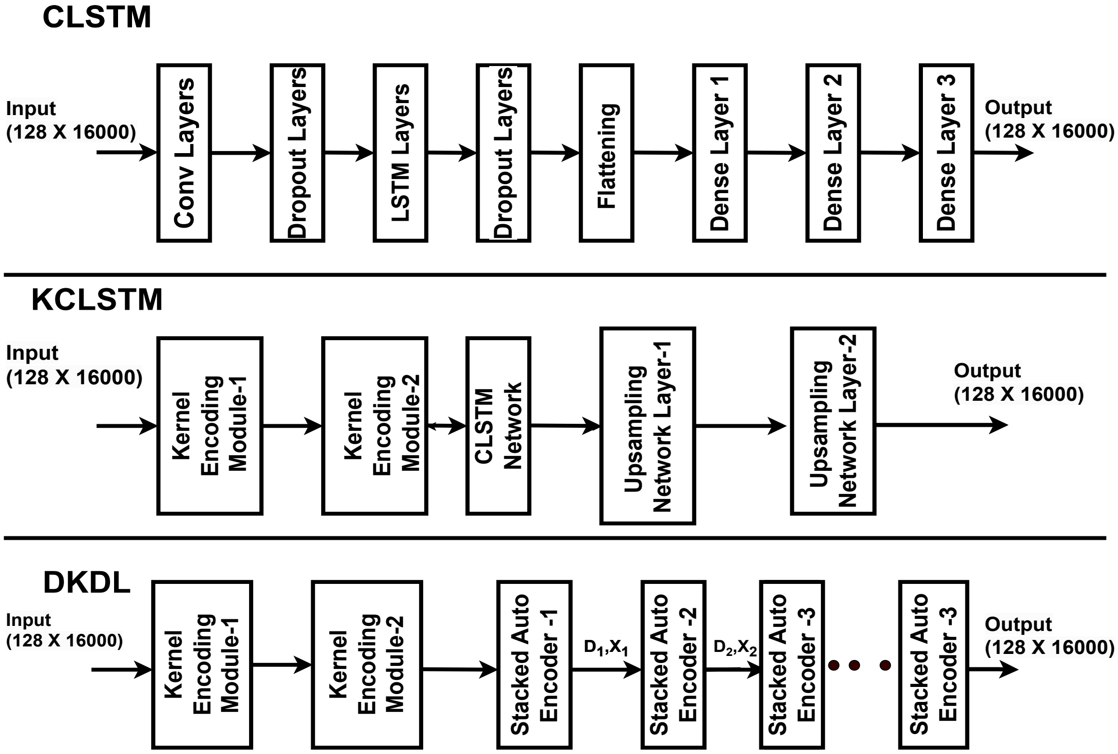


Figure 6.5: CLSM, KCLSTM, and DKDL architectures.

Table 6.3: Characterization of DR against Kernel 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 6.4: Characterization of Time(seconds) against Kernel 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

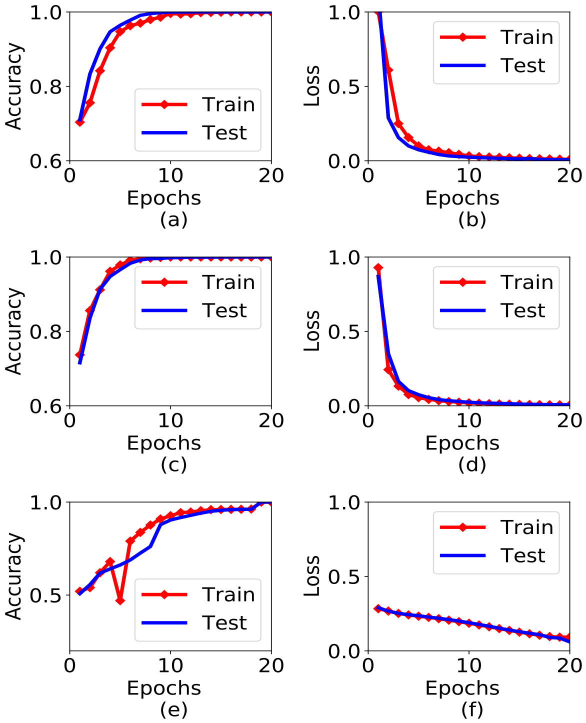


Figure 6.6: Accuracy and loss curves for data on 21 April 2014. (a) and (b) DKDL, (c) and (d) KCLSTM, (e) and (f) CLSTM .

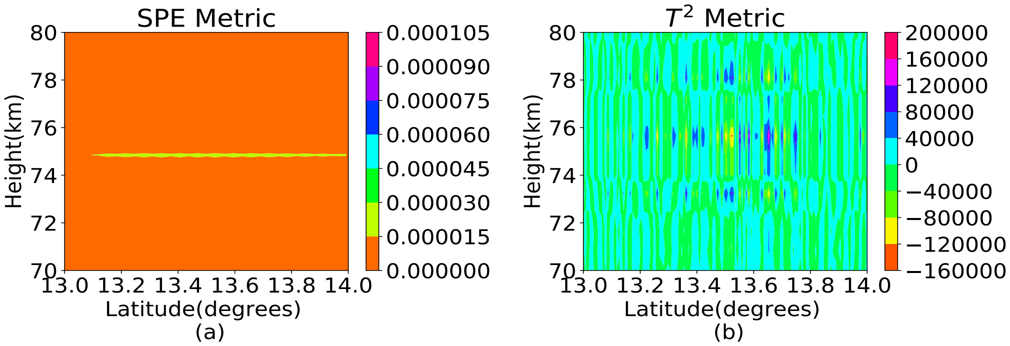


Figure 6.7: Wave Breaking event detection on 21 April 2014 . (a) SPE static and (b) T^2-static.

low for SPE.

Table 6.7 compares the three methods. For all three methods, the network parameters used for training are same and FDR is lower than 6%. Comparing CLSTM, KCLSTM, and DKDL, CLSTM has the largest training error and lowest DR based on SPE. CLSTM and KCLSTM produce poor results compared to DKDL.

Table 6.5: Gravity Wave breaking Detection Rate (DR) 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 6.6: Threshold limitsT^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)

Overall, the proposal performs better than 6% against false detection as shown in Table 6.8. From Tables 6.5 and 6.8, the DRs and FDRs of T^2 and SPE static are better than existing methods in terms of feature extraction.

Table 6.7: Characterization of DR with respect to training loss

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

Table 6.8: Characterization of False Detection Rate(FDR)

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)

6.8.2 Validation of the proposed method

The detected wave breaking events are validated by RL and MRS measurements as shown in Table 6.9 and Fig. 6.8. On the reported days, the Wave Breaking patterns are also seen in the temperature perturbations and the velocity perturbations obtained with the MRS. Identifying Wave Breaking [37] from Atmospheric density, and Wind profiles is through identification of large negative gradients($10Kkm^{-1}$) (2.25) of temperature, exchange of potential energy(2.23) to environment, and a strong wind shear of $-40ms^{-1}km^{-1}$ (indicate the presence of reduced convective and dynamic stability (2.25) and (2.23)). Reduction in Convective stability is observed when the potential temperature contours get steepen as observed from Fig.6.8(e).

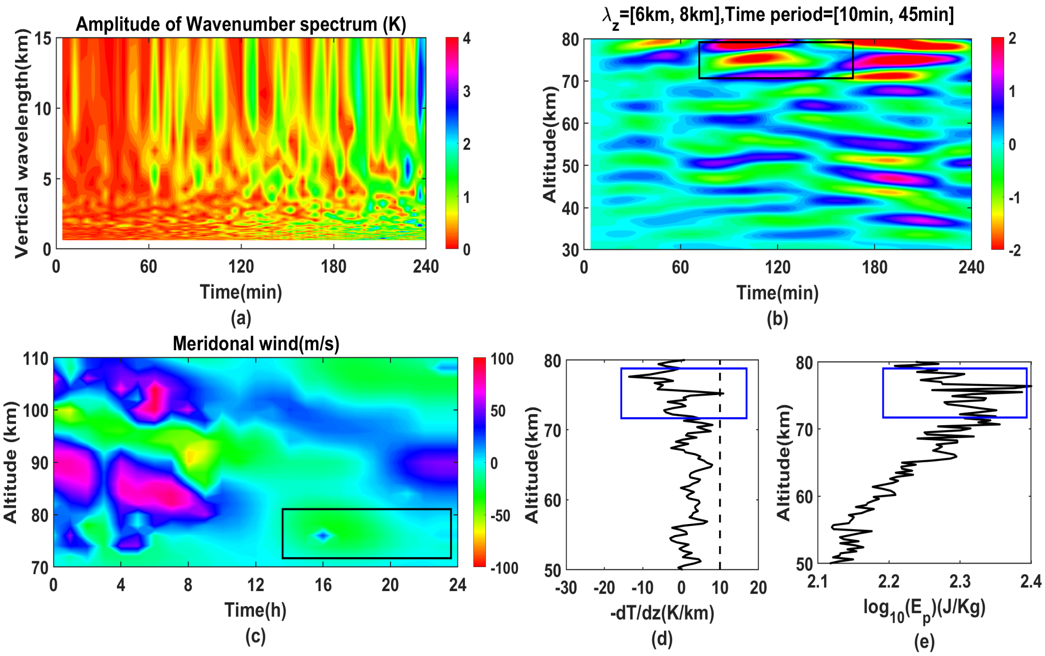


Figure 6.8: Ensuring occurrence of a wave breaking event. (a) vertical wavenumber dominance , (b) Wave progression, (c) Meridional Wind velocities, (d) temperature gradient, and (e) Potential Energy(E_p) obtained on 21 April 2014. The rectangle markers denote the breaking altitudes

From Fig. 6.8(a), a dominant vertical wavenumber was found to be $1.046Kkm^{-1}$. It is observed that there is a large negative temperature gradient of $13Kkm^{-1}$ and strong wind shear of $-40ms^{-1}km^{-1}$ is observed from Fig. 6.8(b), Fig. 6.8(d) and Fig. 6.8(c) respectively. Potential energy transfer(E_p) is shown at 75km altitude in Fig. 6.8(e) and shows that Wave Breaking has taken place. It is possible to verify the instability of the

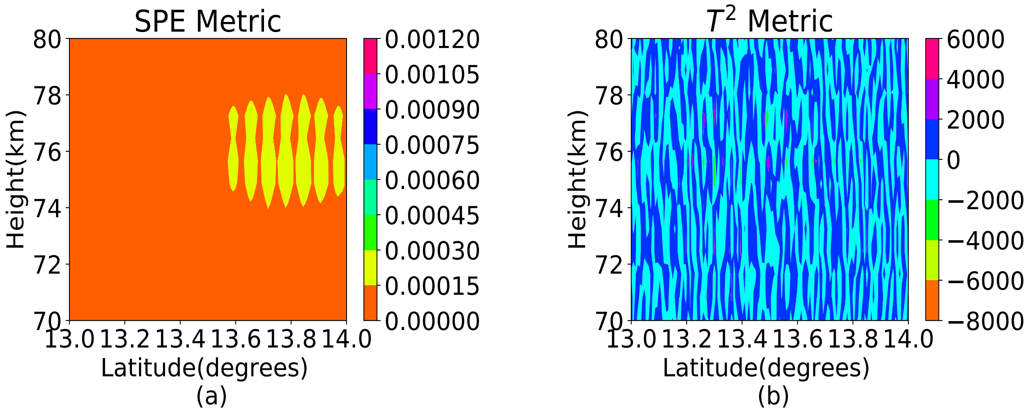
Table 6.9: Threshold limits for DR and FDR based on T^2(SPE) metrics using Deep Kernel Dictionary Learning

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 6.10: Characteristics of Gravity Wave Breaking Events.

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

atmosphere created by Wave Breaking by using equations 2.23, and 2.24. Wave Breaking events for the data listed in Table 6.10 are shown in Figures 6.9,6.10, 6.11, and 6.12.



(a)
(b)

Figure 6.9: (a) Detection of Wave Breaking event based on SPE metric, (b) Wave Breaking event detection using T^2 metric based on temperature perturbations for data collected on 22 November 2017.

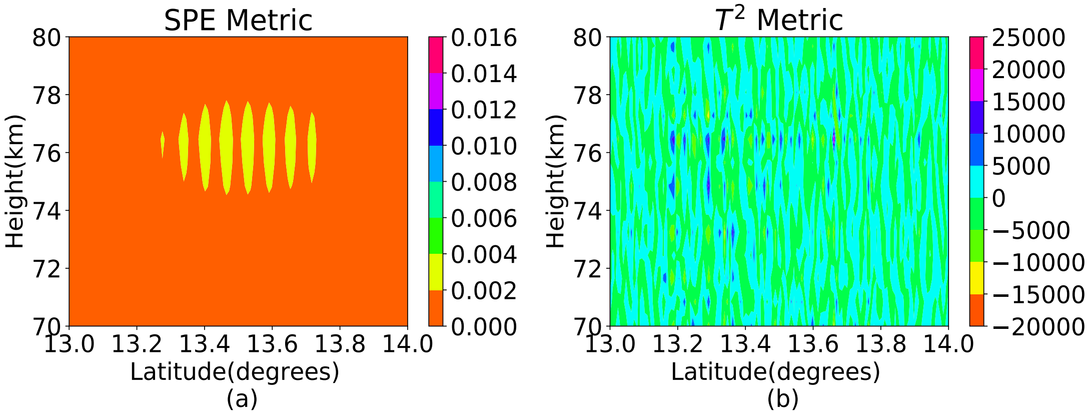


Figure 6.10: (a) Wave Breaking event detection using SPE metric, (b) Wave Breaking event detection using T^2 metric based on temperature perturbations measured on 2 February 2007.

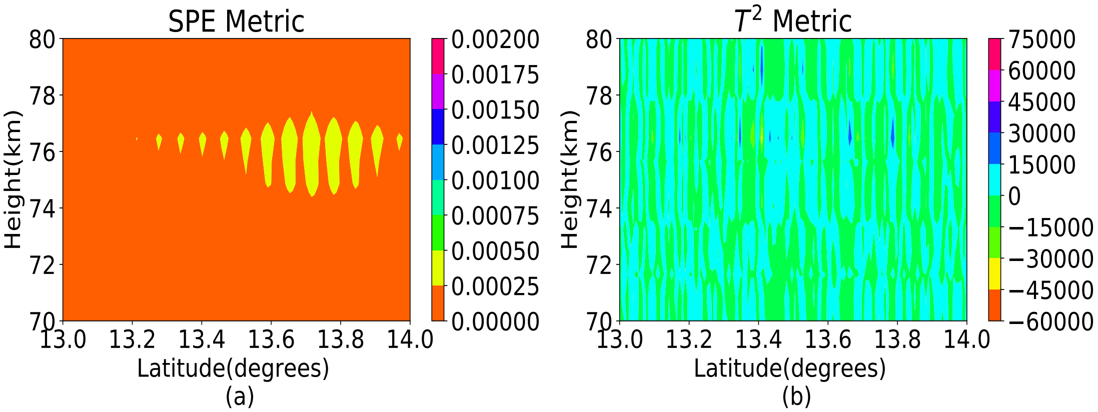


Figure 6.11: (a) Wave Breaking detection using SPE metric, (b) Wave Breaking detection using T^2 metric based on temperature perturbations measured on 21 April 2014.

6.8.3 Applicability Of the proposed method

Non-linear optimization problems have a higher classification accuracy than linear optimization problems. The formed effective Dictionary is capable of detecting features in a very effective manner due to the multi-level dictionary learning capabilities. The proposed method is more effective if:

1. There is a non-linear learning problem (as most optimization problems are non-linear), and achieving high classification accuracy. Using DDL, optimization problems can be solved more accurately.

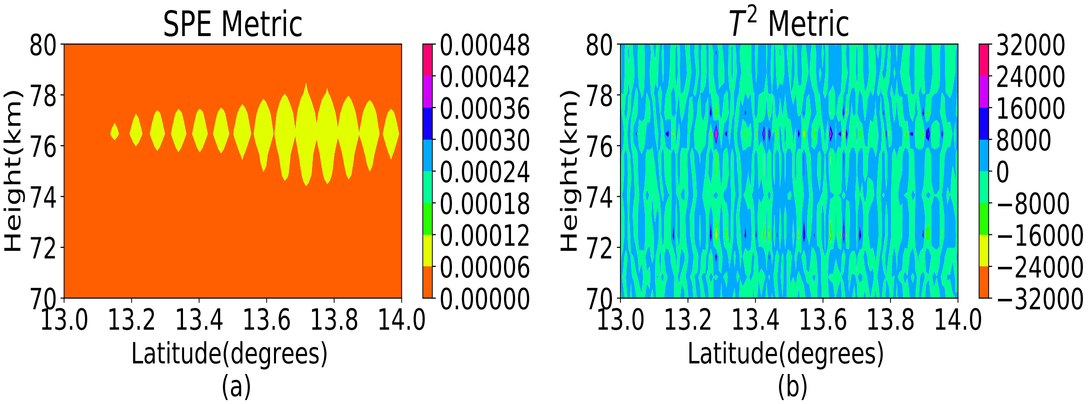


Figure 6.12: (a) Wave Breaking event detection using SPE metric, (b) Wave Breaking event detection based on T^2 metric from temperature perturbations collected on 5 January 2018.

2. With few test data, it requires fewer neural network layers for feature detection. Kernel approximation has a lower computational complexity.

3. Compared to other kernel functions, the RBF kernel in DKDL can identify transient events more effectively. Atmospheric Gravity Wave Breaking is the term used here to describe sudden transient events. Thresholding reveals breaking events in the probability distribution of RBF kernel.

6.9 Summary

1. A Deep Kernel Dictionary Learning algorithm is tested to detect Gravity Wave breaking events with a case study.

2. The proposed method can serve as an effective method for detecting transient events such as Wave Breaking, Turbulence, and Lightnings.

3. Detecting patterns of aerodynamic parameter or process can be accomplished by applying the proposed method to any data.

4. In the long run, we must solve the issues associated with this new methodology.

5. In the future, research might blend other AI solutions into Gravity Waves, thus moving this technology forward.

6. Kernel functions enhance the detection of transient events in Deep Dictionary Learning.
7. It was shown that the proposed methodology provides a detection rate of 95% in cases where Gravity Waves break.

Chapter 7

Conclusion

The reduction in the measurement range of the Rayleigh lidar system over time is addressed using signal processing techniques, and useful results have been obtained. The main highlights of this work are:

1. The dictionary learning technique is used for Rayleigh lidar signal denoising.
2. An iterative method for atmospheric temperature retrieval is presented by solving a penalized maximum likelihood method to reduce the SE.
3. Monte Carlo simulations are used to calculate uncertainties in retrieved temperatures.
4. Both methods in combination improved the measurement range of the Rayleigh lidar system by 6 km.
5. There is a reduction in the SE when photon counts data are denoised using the dictionary learning technique in combination with the $KKTl_2$ method.

MScDL and IF are used to reduce the effect of other background sources. Wavelet decomposition is used to decompose the data to different temporal resolutions. In the Wavelet analysis domain of the Wavelet transform, dictionaries are trained for all decomposition levels. DWTs, DCTs, and single-resolution learned dictionaries are outperformed by learned effective dictionaries. In the frequency and amplitude domains, the algorithm is better suited to measured data, affected by background. We apply the proposed Multi-scale Dictionary Learning algorithm to Gravity Wave signatures to reduce presence of wind in GW's. Doppler shift is examined in relation to Wavelet type, decomposition levels, patch size, and Window size. Using the MScDL, Harr Wavelets with four decomposition's and patches of 2×2 are selected chosen. Based on data from RS, MR, and HWM, the frequency shifts at various altitudes are compared with the Doppler shift profiles. Reconstruction of the profiles is necessary to determine whether the temperature perturbations contain any background wind remnants. During Winter, the Doppler shift distribution departs from normal distribution compared to other seasons, as shown in the study on seasonal variations.

Based on the Dictionary Learning method and kernel adaptations, we propose a method for detecting Wave Breaking events. The method is called “DKDL”. Fourier features are used to implement Kernel approximations. As a result of cascading features from one layer to another, better results are achieved compared with cascading features from the first layer. With inputs that are kernel transformed, the proposed method is identical to the denoising Auto-Encoder. Atmospheric gravity wave breaking events, the DR and FDR of the proposed algorithm DKDL are demonstrated using different data sets. A good detection rate can be achieved through the selection of hyper-parameters using Grid-Search. Compared to CLSTM and KCLSTM, DKDL improved the DR by 33% and 26%. A deep kernel dictionary learning algorithm that incorporates discriminative approaches such as Fisher discrimination and locality-constrained sparse coding, and a DKDL methodology that incorporates supervised learning could further enhance Detection Rates and reduce false alarm rates in order to analyze Gravity Wave Breaking events more accurately.

7.1 Summary

1. We improve the measurement range of Rayleigh Lidar system by 6km using Dictionary Learning in conjunction with Penalized Maximum Likelihood.
2. With Multi-Resolution Dictionary Learning, horizontal background wind effects on atmospheric gravity waves are removed. Insitu instruments like GPS Radiosondes and Meteor Radar verify the results obtained from Multi-Resolution Dictionary Learning.
3. Kernels introduced in Deep Dictionary Learning technique provide 33 higher accuracy in detecting Gravity Wave breaking events than existing methodologies.

Bibliography

[1] V. A. Kovalev and W. E. Eichinger, *Elastic lidar: theory, practice, and analysis methods*. John Wiley & Sons, 2004.

[2] A. Hauchecorne and M.-L. Chanin, "Density and temperature profiles obtained by lidar between 35 and 70 km," *Geophysical Research Letters*, vol. 7, no. 8, pp. 565–568, 1980.

[3] P. Argall, "Upper altitude limit for rayleigh lidar," in *Annales Geophysicae*, vol. 25, no. 1. Copernicus GmbH, 2007, pp. 19–25.

[4] J. Khanna, J. Bandoro, R. Sica, and C. T. McElroy, "New technique for retrieval of atmospheric temperature profiles from rayleigh-scatter lidar measurements using nonlinear inversion," *Applied Optics*, vol. 51, no. 33, pp. 7945–7952, 2012.

[5] A. Povey, R. Grainger, D. Peters, and J. Agnew, "Retrieval of aerosol backscatter, extinction, and lidar ratio from raman lidar with optimal estimation," *Atmospheric Measurement Techniques*, vol. 7, no. 3, pp. 757–776, 2014.

[6] R. J. Sica and A. Haefele, "Retrieval of temperature from a multiple-channel rayleigh-scatter lidar using an optimal estimation method," *Applied optics*, vol. 54, no. 8, pp. 1872–1889, 2015.

[7] Z. Liu, W. Hunt, M. Vaughan, C. Hostetler, M. McGill, K. Powell, D. Winker, and Y. Hu, "Estimating random errors due to shot noise in backscatter lidar observations," *Applied optics*, vol. 45, no. 18, pp. 4437–4447, 2006.

[8] I. Bipm, I. IFCC, and I. Iupac, "Evaluation of measurement data-supplement 1 to the'guide to the expression of uncertainty in measurement," *Propagation of distributions using a Monte Carlo method*, 2008.

[9] J. C. for Guides in Metrology, "Evaluation of measurement data—guide to the expression of uncertainty in measurement," *JCGM*, vol. 100, no. 2008, pp. 1–116, 2008.

[10] F. Chane-Ming, F. Molinaro, and J. Leveau, "Wavelet techniques applied to lidar temperature profiles in the middle atmosphere to study gravity waves," in *IGARSS'97. 1997 IEEE International Geoscience and Remote Sensing Symposium Proceedings. Remote Sensing-A Scientific Vision for Sustainable Development*, vol. 1. IEEE, 1997, pp. 581–583.

[11] F. Chane-Ming, F. Molinaro, J. Leveau, P. Keckhut, and A. Hauchecorne, "Analysis of gravity waves in the tropical middle atmosphere over la reunion island (21 s, 55 e) with lidar using wavelet techniques," in *Annales Geophysicae*, vol. 18, no. 4. Copernicus GmbH, 2000, pp. 485–498.

[12] R. Wilson, M. Chanin, and A. Hauchecorne, "Gravity waves in the middle atmosphere observed by rayleigh lidar: 2. climatology," *Journal of Geophysical Research: Atmospheres*, vol. 96, no. D3, pp. 5169–5183, 1991.

[13] M. Rauthe, M. Gerding, J. Höffner, and F.-J. Lübken, "Lidar temperature measurements of gravity waves over kühlungsborn (54° n) from 1 to 105 km: A winter-summer comparison," *Journal of Geophysical Research: Atmospheres*, vol. 111, no. D24, 2006.

[14] C. J. Nappo, *An introduction to atmospheric gravity waves.* Academic press, 2013.

[15] C. Lu, S. Koch, and N. Wang, "Determination of temporal and spatial characteristics of atmospheric gravity waves combining cross-spectral analysis and wavelet transformation," *Journal of Geophysical Research: Atmospheres*, vol. 110, no. D1, 2005.

[16] K. Sato and M. Yamada, "Vertical structure of atmospheric gravity waves revealed by the wavelet analysis," *Journal of Geophysical Research: Atmospheres*, vol. 99, no. D10, pp. 20 623–20 631, 1994.

[17] T. Tsuda, "Characteristics of atmospheric gravity waves observed using the mu (middle and upper atmosphere) radar and gps (global positioning system) radio occultation," *Proceedings of the Japan Academy, Series B*, vol. 90, no. 1, pp. 12–27, 2014.

[18] P. M. Franke and R. L. Collins, “Evidence of gravity wave breaking in lidar data from the mesopause region,” *Geophysical research letters*, vol. 30, no. 4, 2003.

[19] D. C. Fritts, “Gravity wave saturation in the middle atmosphere: A review of theory and observations,” *Reviews of Geophysics*, vol. 22, no. 3, pp. 275–308, 1984.

[20] J. R. Booker and F. P. Bretherton, “The critical layer for internal gravity waves in a shear flow,” *Journal of fluid mechanics*, vol. 27, no. 3, pp. 513–539, 1967.

[21] R. Lindzen, “Gravity waves in the mesosphere,” *Dynamics of the middle atmosphere*, pp. 3–18, 1984.

[22] R. S. Lindzen, “Turbulence and stress owing to gravity wave and tidal breakdown,” *Journal of Geophysical Research: Oceans*, vol. 86, no. C10, pp. 9707–9714, 1981.

[23] C. O. Hines, “Generation of turbulence by atmospheric gravity waves,” *Journal of Atmospheric Sciences*, vol. 45, no. 7, pp. 1269–1278, 1988.

[24] H.-L. Liu, P. B. Hays, and R. G. Roble, “A numerical study of gravity wave breaking and impacts on turbulence and mean state,” *Journal of the Atmospheric Sciences*, vol. 56, no. 13, pp. 2152 – 2177, 1999.

[25] D. C. Fritts and M. J. Alexander, “Gravity wave dynamics and effects in the middle atmosphere,” *Reviews of geophysics*, vol. 41, no. 1, 2003.

[26] D. C. Fritts, C. Bizon, J. A. Werne, and C. K. Meyer, “Layering accompanying turbulence generation due to shear instability and gravity-wave breaking,” *Journal of Geophysical Research: Atmospheres*, vol. 108, no. D8, 2003.

[27] C. Hines, “Generalizations of the richardson criterion for the onset of atmospheric turbulence,” *Quarterly Journal of the Royal Meteorological Society*, vol. 97, no. 414, pp. 429–439, 1971.

[28] S. D. Eckermann, “Effect of background winds on vertical wavenumber spectra of atmospheric gravity waves,” *Journal of Geophysical Research: Atmospheres*, vol. 100, no. D7, pp. 14 097–14 112, 1995.

[29] S. D. Eckermann, “Influence of wave propagation on the doppler spreading of atmospheric gravity waves,” *Journal of the atmospheric sciences*, vol. 54, no. 21, pp. 2554–2573, 1997.

[30] D. C. Fritts and T. VanZandt, "Effects of doppler shifting on the frequency spectra of atmospheric gravity waves," *Journal of Geophysical Research: Atmospheres*, vol. 92, no. D8, pp. 9723–9732, 1987.

[31] C. Heale and J. Snively, "Gravity wave propagation through a vertically and horizontally inhomogeneous background wind," *Journal of Geophysical Research: Atmospheres*, vol. 120, no. 12, pp. 5931–5950, 2015.

[32] C. O. Hines, "Doppler-spread parameterization of gravity-wave momentum deposition in the middle atmosphere. part 1: Basic formulation," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 59, no. 4, pp. 371–386, 1997.

[33] M. Pulido and G. Caranti, "Power spectrum of a gravity wave propagating in a shearing background wind," *Geophysical research letters*, vol. 27, no. 1, pp. 101–104, 2000.

[34] M. Pullman, I. Gurung, M. Maskey, R. Ramachandran, and S. A. Christopher, "Applying deep learning to hail detection: A case study," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 10 218–10 225, 2019.

[35] C. Warner and M. McIntyre, "On the propagation and dissipation of gravity wave spectra through a realistic middle atmosphere," *Journal of Atmospheric Sciences*, vol. 53, no. 22, pp. 3213–3235, 1996.

[36] B. Ehard, B. Kaifler, N. Kaifler, and M. Rapp, "Evaluation of methods for gravity wave extraction from middle-atmospheric lidar temperature measurements," *Atmospheric Measurement Techniques*, vol. 8, no. 11, pp. 4645–4655, 2015.

[37] J. Li, R. Collins, X. Lu, and B. Williams, "Lidar observations of instability and estimates of vertical eddy diffusivity induced by gravity wave breaking in the arctic mesosphere," *Journal of Geophysical Research: Atmospheres*, vol. 126, no. 4, p. e2020JD033450, 2021.

[38] X. Lu, A. Z. Liu, G. R. Swenson, T. Li, T. Leblanc, and I. S. McDermid, "Gravity wave propagation and dissipation from the stratosphere to the lower thermosphere," *Journal of Geophysical Research: Atmospheres*, vol. 114, no. D11, 2009.

[39] H. Sit and C. J. Earls, "Deep learning for classifying and characterizing atmospheric ducting within the maritime setting," *Computers & Geosciences*, vol. 157, p. 104919, 2021.

[40] D. Matsuoka, S. Watanabe, K. Sato, S. Kawazoe, W. Yu, and S. Easterbrook, "Application of deep learning to estimate atmospheric gravity wave parameters in reanalysis data sets," *Geophysical Research Letters*, vol. 47, no. 19, p. e2020GL089436, 2020.

[41] C. E. Stringari, M. Prevosto, J.-F. Filipot, F. Leckler, and P. V. Guimarães, "A new probabilistic wave breaking model for dominant wind-sea waves based on the gaussian field theory," *Journal of Geophysical Research: Oceans*, vol. 126, no. 4, p. e2020JC016943, 2021.

[42] D. L. Donoho, X. Huo *et al.*, "Uncertainty principles and ideal atomic decomposition," *IEEE transactions on information theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[43] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE transactions on Information Theory*, vol. 58, no. 2, pp. 1094–1121, 2012.

[44] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ^{1} minimization," in *PROC. NATL ACAD. SCI. USA 100 2197–202*. Citeseer, 2002.

[45] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[46] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2558–2567, 2002.

[47] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM Journal on Applied Mathematics*, vol. 49, no. 3, pp. 906–931, 1989.

[48] B. Wohlberg, "Efficient convolutional sparse coding," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 7173–7177.

[49] J. Qiu and M. Yang, "Adaptive convolutional sparse coding with weighted l_1-norm," in *2019 Chinese Control Conference (CCC)*, 2019, pp. 7746–7750.

[50] S. Gao, I. W.-H. Tsang, and L.-T. Chia, “Laplacian sparse coding, hypergraph laplacian sparse coding, and applications,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 92–104, 2013.

[51] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[52] K. Engan, S. Aase, and J. Hakon Husoy, “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, 1999, pp. 2443–2446 vol.5.

[53] K. Engan, S. O. Aase, and J. H. Husoy, “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. IEEE, 1999, pp. 2443–2446.

[54] W. Gui and X. Shao, “Onset detection using leared dictionary by k-svd,” in *2014 IEEE Workshop on Advanced Research and Technology in Industry Applications (WARTIA)*. IEEE, 2014, pp. 406–409.

[55] R. Rubinstein, T. Peleg, and M. Elad, “Analysis k-svd: A dictionary-learning algorithm for the analysis sparse model,” *IEEE Transactions on Signal Processing*, vol. 61, no. 3, pp. 661–677, 2012.

[56] M. Aharon, M. Elad, and A. Bruckstein, “K-svd: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Transactions on signal processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[57] Z. Jiang, Z. Lin, and L. S. Davis, “Label consistent k-svd: Learning a discriminative dictionary for recognition,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 11, pp. 2651–2664, 2013.

[58] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in *Proceedings of 27th Asilomar conference on signals, systems and computers*. IEEE, 1993, pp. 40–44.

[59] J. B. Crockett, H. Chernoff *et al.*, "Gradient methods of maximization." *Pacific Journal of Mathematics*, vol. 5, no. 1, pp. 33–50, 1955.

[60] M. R. Hestenes, "Multiplier and gradient methods," *Journal of optimization theory and applications*, vol. 4, no. 5, pp. 303–320, 1969.

[61] P. Irofti and B. Dumitrescu, "Regularized algorithms for dictionary learning," in *2016 International Conference on Communications (COMM)*, 2016, pp. 439–442.

[62] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of computational mathematics*, vol. 9, no. 3, pp. 317–334, 2009.

[63] D. L. Donoho, Y. Tsaig, I. Drori, and J.-l. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit, submitted to," in *IEEE Trans. on Information theory*. Citeseer, 2006.

[64] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[65] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 57, no. 11, pp. 1413–1457, 2004.

[66] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[67] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *Journal of Fourier analysis and Applications*, vol. 14, no. 5-6, pp. 629–654, 2008.

[68] A. Maleki, "Coherence analysis of iterative thresholding algorithms," in *2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2009, pp. 236–243.

[69] E. Treister and I. Yavneh, "A multilevel iterated-shrinkage approach to $l\_\{1\}$ penalized least-squares minimization," *IEEE transactions on signal processing*, vol. 60, no. 12, pp. 6319–6329, 2012.

[70] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted 1 minimization," *Journal of Fourier analysis and applications*, vol. 14, no. 5, pp. 877–905, 2008.

[71] F. Chane-Ming, F. Molinaro, and J. Leveau, "Wavelet techniques applied to lidar temperature profiles in the middle atmosphere to study gravity waves," in *IGARSS'97. 1997 IEEE International Geoscience and Remote Sensing Symposium Proceedings. Remote Sensing - A Scientific Vision for Sustainable Development*, vol. 1, 1997, pp. 581–583 vol.1.

[72] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 543–550.

[73] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Mullers, "Fisher discriminant analysis with kernels," in *Neural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop (cat. no. 98th8468)*. Ieee, 1999, pp. 41–48.

[74] M. Barstuğan and R. Ceylan, "A discriminative dictionary learning-adaboost-svm classification method on imbalanced datasets," in *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, 2017, pp. 1–4.

[75] N. Khan and M. Tappen, "Stable discriminative dictionary learning via discriminative deviation," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, 2012, pp. 3224–3227.

[76] Y. Kong, T. Wang, F. Chu, Z. Feng, and I. Selesnick, "Discriminative dictionary learning-based sparse classification framework for data-driven machinery fault diagnosis," *IEEE Sensors Journal*, vol. 21, no. 6, pp. 8117–8129, 2021.

[77] U. Rodríguez-Domínguez and O. Dalmau, "Hierarchical discriminative deep dictionary learning," *IEEE Access*, vol. 8, pp. 142 680–142 690, 2020.

[78] J. Bin, Z. Jing, and Y. Zhiyong, "Learning a discriminative dictionary for locality constrained coding and sparse representation," in *Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, 2014, pp. 1–6.

[79] A. Fawzi, M. Davies, and P. Frossard, "Dictionary learning for fast classification based on soft-thresholding," *International Journal of Computer Vision*, vol. 114, no. 2, pp. 306–321, 2015.

[80] Q. Chen, R. Wynne, P. Goulding, and D. Sandoz, "The application of principal component analysis and kernel density estimation to enhance process monitoring," *Control Engineering Practice*, vol. 8, no. 5, pp. 531–543, 2000.

[81] Y. Cui, J. Yang, and X. Zhang, "New cfar target detector for sar images based on kernel density estimation and mean square error distance," *Journal of Systems Engineering and Electronics*, vol. 23, no. 1, pp. 40–46, 2012.

[82] B. A. Turlach, "Bandwidth selection in kernel density estimation: A review," in *CORE and Institut de Statistique*. Citeseer, 1993.

[83] D. M. Bashtannyk and R. J. Hyndman, "Bandwidth selection for kernel conditional density estimation," *Computational Statistics & Data Analysis*, vol. 36, no. 3, pp. 279–298, 2001.

[84] M. Chanin and A. Hauchecorne, "Lidar studies of temperature and density using rayleigh scattering," *In International Council of Scientific Unions Middle Atmosphere Handbook*, vol. 13, pp. 87–98, 1984.

[85] A. Taori, A. Jayaraman, K. Raghunath, and V. Kamalakar, "A new method to derive middle atmospheric temperature profiles using a combination of rayleigh lidar and o 2 airglow temperatures measurements," in *Annales Geophysicae*, vol. 30, no. 1. Copernicus GmbH, 2012, pp. 27–32.

[86] R. Wing, A. Hauchecorne, P. Keckhut, S. Godin-Beekmann, S. Khaykin, M. Martic, W. Steinbrecht, T. J. McGee, J. Sullivan, and E. McCullough, "Intercomparisons between lidar and satellite instruments in the middle atmosphere," in *EGU General Assembly Conference Abstracts*, 2020, p. 11725.

[87] F. Rocadenbosch, C. Soriano, A. Comerón, and J.-M. Baldasano, "Lidar inversion of atmospheric backscatter and extinction-to-backscatter ratios by use of a kalman filter," *Applied optics*, vol. 38, no. 15, pp. 3175–3189, 1999.

[88] H.-T. Fang and D.-S. Huang, "Noise reduction in lidar signal based on discrete wavelet transform," *Optics Communications*, vol. 233, no. 1-3, pp. 67–76, 2004.

[89] J. Mao, "Noise reduction for lidar returns using local threshold wavelet analysis," *Optical and Quantum Electronics*, vol. 43, no. 1, pp. 59–68, 2012.

[90] Z. Zhou, D. Hua, Y. Wang, Q. Yan, S. Li, Y. Li, and H. Wang, "Improvement of the signal to noise ratio of lidar echo signal based on wavelet de-noising technique," *Optics and Lasers in Engineering*, vol. 51, no. 8, pp. 961–966, 2013.

[91] P. Tian, X. Cao, J. Liang, L. Zhang, N. Yi, L. Wang, and X. Cheng, "Improved empirical mode decomposition based denoising method for lidar signals," *Optics Communications*, vol. 325, pp. 54–59, 2014.

[92] M. Sarvani, K. Raghunath, and S. V. B. Rao, "Lidar signal denoising methods-application to narl rayleigh lidar," *Journal of Optics*, vol. 44, no. 2, pp. 164–171, 2015.

[93] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE transactions on signal processing*, vol. 62, no. 3, pp. 531–544, 2013.

[94] H. Li, J. Chang, F. Xu, Z. Liu, Z. Yang, L. Zhang, S. Zhang, R. Mao, X. Dou, and B. Liu, "Efficient lidar signal denoising algorithm using variational mode decomposition combined with a whale optimization algorithm," *Remote Sensing*, vol. 11, no. 2, p. 126, 2019.

[95] B.-Y. Sun, D.-S. Huang, and H.-T. Fang, "Lidar signal denoising using least-squares support vector machine," *IEEE Signal processing letters*, vol. 12, no. 2, pp. 101–104, 2005.

[96] H.-T. Fang, D.-S. Huang, and Y.-H. Wu, "Antinoise approximation of the lidar signal with wavelet neural networks," *Applied optics*, vol. 44, no. 6, pp. 1077–1083, 2005.

[97] Z. Gao, Q. Li, R. Zhai, and F. Lin, "Laser range data denoising via adaptive and robust dictionary learning," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 8, pp. 1750–1754, 2015.

[98] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.

[99] H. Wang, J. Vieira, P. Ferreira, B. Jesus, and I. Duarte, "Batch algorithms of matching pursuit and orthogonal matching pursuit with applications to compressed sensing," in *2009 International Conference on Information and Automation*. IEEE, 2009, pp. 824–829.

[100] V. S. Sreekanth, K. Raghunath, and D. Mishra, “Dictionary learning technique and penalized maximum likelihood for extending measurement range of a rayleigh lidar,” *Journal of Applied Remote Sensing*, vol. 14, no. 3, p. 034529, 2020.

[101] P. J. Burt and E. H. Adelson, “The laplacian pyramid as a compact image code,” in *Readings in computer vision.* Elsevier, 1987, pp. 671–679.

[102] A. Yadav, A. Aryasomayajula, and R. AhmedAnsari, “Multiresolution analysis based sparse dictionary learning for remotely sensed image retrieval,” in *2019 Women Institute of Technology Conference on Electrical and Computer Engineering (WITCON ECE).* IEEE, 2019, pp. 76–80.

[103] S. G. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Transactions on signal processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[104] J. Medel, A. Savakis, and B. Ghoraani, “A novel time-frequency feature extraction algorithm based on dictionary learning,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE, 2016, pp. 4895–4899.

[105] B. Ophir, M. Lustig, and M. Elad, “Multi-scale dictionary learning using wavelets,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 1014–1024, 2011.

[106] N. A. Khan and B. Boashash, “Instantaneous frequency estimation of multicomponent nonstationary signals using multiview time-frequency distributions based on the adaptive fractional spectrogram,” *IEEE Signal Processing Letters*, vol. 20, no. 2, pp. 157–160, 2012.

[107] A. Zhang, F. Hu, Q. He, C. Shen, F. Liu, and F. Kong, “Doppler shift removal based on instantaneous frequency estimation for wayside fault diagnosis of train bearings,” *Journal of Vibration and Acoustics*, vol. 136, no. 2, 2014.

[108] Y. Abdoush, G. Pojani, and G. E. Corazza, “Adaptive instantaneous frequency estimation of multicomponent signals based on linear time–frequency transforms,” *IEEE Transactions on Signal Processing*, vol. 67, no. 12, pp. 3100–3112, 2019.

[109] M. V. Ratnam, N. Pravallika, S. R. Babu, G. Basha, M. Pramitha, and B. K. Murthy, “Assessment of gps radiosonde descent data,” *Atmos. Meas. Tech.*, vol. 7, pp. 1011–1025, 2014.

[110] S. V. B. Rao, S. Eswaraiah, M. Venkat Ratnam, E. Kosalendra, K. Kishore Kumar, S. Sathish Kumar, P. Patil, and S. Gurubaran, "Advanced meteor radar installed at tirupati: System details and comparison with different radars," *Journal of Geophysical Research: Atmospheres*, vol. 119, no. 21, pp. 11–893, 2014.

[111] A. Dörnbrack, S. Gisinger, and B. Kaifler, "On the interpretation of gravity wave measurements by ground-based lidars," *Atmosphere*, vol. 8, no. 3, p. 49, 2017.

[112] V. N. Gubenko, A. G. Pavelyev, and V. E. Andreev, "Determination of the intrinsic frequency and other wave parameters from a single vertical temperature or density profile measurement," *Journal of Geophysical Research: Atmospheres*, vol. 113, no. D8, 2008.

[113] D. Drob, J. Emmert, G. Crowley, J. Picone, G. Shepherd, W. Skinner, P. Hays, R. Niciejewski, M. Larsen, C. She *et al.*, "An empirical model of the earth's horizontal wind fields: Hwm07," *Journal of Geophysical Research: Space Physics*, vol. 113, no. A12, 2008.

[114] V. N. Gubenko, A. G. Pavelyev, and V. E. Andreev, "Determination of the intrinsic frequency and other wave parameters from a single vertical temperature or density profile measurement," *Journal of Geophysical Research: Atmospheres*, vol. 113, no. D8, 2008.

[115] T. Leblanc, R. J. Sica, J. A. Van Gijsel, S. Godin-Beekmann, A. Haefele, T. Trickl, G. Payen, and G. Liberti, "Proposed standardized definitions for vertical resolution and uncertainty in the ndacc lidar ozone and temperature algorithms–part 2: Ozone dial uncertainty budget," *Atmospheric Measurement Techniques*, vol. 9, no. 8, pp. 4051–4078, 2016.

[116] C. Strube, M. Ern, P. Preusse, and M. Riese, "Removing spurious inertial instability signals from gravity wave temperature perturbations using spectral filtering methods," *Atmospheric measurement techniques*, vol. 13, no. 9, pp. 4927–4945, 2020.

[117] H. V. Nguyen, V. M. Patel, N. M. Nasrabadi, and R. Chellappa, "Kernel dictionary learning," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 2021–2024.

[118] A. Rahimi, B. Recht *et al.*, "Random features for large-scale kernel machines." in *NIPS*, vol. 3, no. 4. Citeseer, 2007, p. 5.

[119] J. Yang, V. Sindhwani, H. Avron, and M. Mahoney, “Quasi-monte carlo feature maps for shift-invariant kernels,” in *International Conference on Machine Learning*. PMLR, 2014, pp. 485–493.

[120] P. Kar and H. Karnick, “Random feature maps for dot product kernels,” in *Artificial intelligence and statistics*. PMLR, 2012, pp. 583–591.

[121] R. Hamid, Y. Xiao, A. Gittens, and D. DeCoste, “Compact random feature maps,” in *International conference on machine learning*. PMLR, 2014, pp. 19–27.

[122] P. Drineas, M. W. Mahoney, and N. Cristianini, “On the nyström method for approximating a gram matrix for improved kernel-based learning.” *journal of machine learning research*, vol. 6, no. 12, 2005.

[123] T. Yang, Y.-F. Li, M. Mahdavi, R. Jin, and Z.-H. Zhou, “Nyström method vs random fourier features: A theoretical and empirical comparison,” *Advances in neural information processing systems*, vol. 25, pp. 476–484, 2012.

[124] S. Kumar, M. Mohri, and A. Talwalkar, “Sampling techniques for the nystrom method,” in *Artificial intelligence and statistics*. PMLR, 2009, pp. 304–311.

[125] A. Golts and M. Elad, “Linearized kernel dictionary learning,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 726–739, 2016.

[126] Q. Li, X. Wu, L. Xu, L. Yao, and K. Chen, “Multi-feature kernel discriminant dictionary learning for classification in alzheimer’s disease,” in *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 2017, pp. 1–6.

[127] A. M. Vorontsov, M. A. Vorontsov, G. A. Filimonov, and E. Polnau, “Atmospheric turbulence study with deep machine learning of intensity scintillation patterns,” *Applied Sciences*, vol. 10, no. 22, p. 8136, 2020.

[128] D. Malmgren-Hansen, V. Laparra, A. A. Nielsen, and G. Camps-Valls, “Statistical retrieval of atmospheric profiles with deep convolutional neural networks,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 158, pp. 231–240, 2019.

[129] N. Kumar, K. Soni, and R. Agarwal, “Prediction of temporal atmospheric boundary layer height using long short-term memory network,” *Tellus A: Dynamic Meteorology and Oceanography*, vol. 73, no. 1, pp. 1–14, 2021.

[130] J. Ma, J. C. Cheng, C. Lin, Y. Tan, and J. Zhang, “Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques,” *Atmospheric Environment*, vol. 214, p. 116885, 2019.

[131] W. Qiao, W. Tian, Y. Tian, Q. Yang, Y. Wang, and J. Zhang, “The forecasting of pm2. 5 using a hybrid model based on wavelet transform and an improved deep learning algorithm. ieee access 7: 142814–142825,” 2019.

[132] S. Tariyal, A. Majumdar, R. Singh, and M. Vatsa, “Deep dictionary learning,” *IEEE Access*, vol. 4, pp. 10 096–10 109, 2016.

[133] S. Tariyal, H. Aggarwal, and A. Majumdar, “Greedy deep dictionary learning for hyperspectral image classification,” in *2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2016, pp. 1–4.

[134] Y. Bengio, *Learning deep architectures for AI*. Now Publishers Inc, 2009.

[135] C. Williams and M. Seeger, “Using the nyström method to speed up kernel machines,” in *Proceedings of the 14th annual conference on neural information processing systems*, no. CONF, 2001, pp. 682–688.

[136] R. Salakhutdinov and G. Hinton, “Deep boltzmann machines,” in *Artificial intelligence and statistics*. PMLR, 2009, pp. 448–455.

[137] A. Rahimi and B. Recht, “Random features for large-scale kernel machines,” in *Advances in Neural Information Processing Systems*, J. Platt, D. Koller, Y. Singer, and S. Roweis, Eds., vol. 20. Curran Associates, Inc., 2008.

[138] T. Yang, Y.-f. Li, M. Mahdavi, R. Jin, and Z.-H. Zhou, “Nyström method vs random fourier features: A theoretical and empirical comparison,” in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012.

[139] M. Schwartz, N. Livesey, and W. Read, “Mls/aura level 2 temperature v004, greenbelt, md, usa, goddard earth sciences data and information services center (ges disc),” 2015.

[140] A. E. Hedin, E. Fleming, A. Manson, F. Schmidlin, S. Avery, R. Clark, S. Franke, G. Fraser, T. Tsuda, F. Vial *et al.*, “Empirical wind model for the upper, middle and

lower atmosphere," *Journal of atmospheric and terrestrial physics*, vol. 58, no. 13, pp. 1421–1447, 1996.

List of Publications

Referred Journals

1. Varanasi Satya Sreekanth, Karnam Raghunath, Deepak Mishra, “Deep Kernel Dictionary Learning for detection of wave breaking features in Atmospheric Gravity Waves’, Computers Geosciences, Volume 176, 2023, 105361.**(IF:5.176).**
2. Varanasi Satya Sreekanth, Karnam Raghunath, Deepak Mishra, **"Multi-resolution dictionary learning for discrimination of hidden features: A case study of atmospheric gravity waves"**, Elsever-Signal Processing, Volume 204, 2023, 108831, ISSN 0165-1684, https://doi.org/10.1016/j.sigpro.2022.108831.**(IF:4.729)**
3. Varanasi Satya Sreekanth, Karnam Raghunath, Deepak Mishra,**“Dictionary learning technique and penalized maximum likelihood for extending measurement range of a Rayleigh lidar,**" SPIE-J. Appl. Rem. Sens. 14(3) 034529 (30 September 2020) https://doi.org/10.1117/1.JRS.14.034529.**(IF:1.57)**

Presentations

1. Varanasi Satya Sreekanth, Karnam Raghunath, Deepak Mishra “Discrimination Of Doppler Shift In Atmospheric Gravity Waves Due To Horizontal Background Wind Using Dictionary Learning ", held between 31^{st} January to 4^{th} February, 2022, at 21^{st} National Space Science Symposium Organized by ISRO and IISER Kolkata. (Role: Presenter)
2. K Raghunath, M Venkatratnam, V Satya Sreekanth, T Rajendra Prasad, G Sandeep “Initial observations of atmospheric ozone with NARL DIAL system ", held between 31^{st} January to 4^{th} February, 2022, at 21^{st} National Space Science Symposium Organized by ISRO and IISER Kolkata. (Role: Co-Presenter)

3. V S Sreekanth , K Raghunath and Deepak Mishra, “Iterative Algorithm for Atmospheric Temperature Retrieval with Application to Lidar",29^{th} International Laser Radar Conference,,held at Hefei,Anhui,China between $24^{th} - 28^{th}$June 2019 with complete travel grant from Jet Propulsion Laboratory(JPL)-National Aeronautics Space Administration(NASA). (Role: Presenter)

Workshops Attended

1. Varanasi Satya Sreekanth,“Artificial Intelligence In Remote Sensing", organized by IEEE-Geosciences and Remote Sensing Society between $21^{st} - 23^{rd}$ October, 2021. (Role: Attendee)

2. Sreekanth Varanasi,“OPTICA Laser Congress", organized by OPTICA(Optical Society of America), between $3^{rd} - 7^{th}$ October, 2021. (Role: Attendee).

3. Varanasi Satya Sreekanth,“Deep Learning for Sensor Signal Analytics", organised by IEEE Signal Processing Society and Academy of Scientific and Innovative Research-Central Scientific Instruments Organisation(AcSIR-CSIO) between $2^{nd} - 6^{th}$ August, 2021. (**Appreciation certificate from IEEE Signal Processing Society**). (Role: Attendee)

4. Sreekanth Varanasi,“OSA Laser Congress", organized by Optical Society of America, between $13^{th} - 16^{th}$ October, 2020. (Role: Attendee)

Appendix A

Data Availability and Instruments Used

A.1 Data and Computer Code Availability

1. Hardware Specifications: Code was tested on a modern PC with 8 GB RAM, Intel i5 processor.
2. Software required: Matlab, Python 3.6+ with the numpy, scikitlearn, matplotlib, and scipy packages; Tensorflow, Tensorflow-gpu, Keras, and Pytorch
3. Program language: Python, and MATLAB.
4. Details on how to access the source code: https://github.com/VARANASISATYASREEKANTH
5. The data used in the present work is available at $https://www.narl.gov.in$ and can be accessed by readers upon request to data center.
6. The data of the Horizontal Wind Model is available at $https://www.alpendac.eu$
7. The MLS-Aura data is available at:$https://mls.jpl.nasa.gov/eos-aura-mls/data-access$
8. The SABER instrument data is available at $https://saber.gats-inc.com$

A.2 GPS Radiosonde

NARL(13.5^0N,79.2^0E) launches GPS Radiosonde (RD-11G, Meisei make) every day around 17 hours 30 minutes Local Time (12 UTC). The GPS Radiosonde provides profiles of Pressure, Temperature, Relative humidity, Wind Speed, and Wind Direction up to the balloon burst altitude typically 30km to 35km. The maximum height reached by Radiosonde is 33km on 21 April 2014.

A.3 Meteor Radar

The Meteor Radar(MR) operates at 35.25 MHz, was installed at Sri Venkate-swara University (SVU)(13.63^0N,79.4^0E), Tirupati, India, and is used to measure horizontal winds in the Mesosphere and Lower Thermosphere (MLT). It is located 30 km away from the Gadanki, which provides information about winds in the region between 70 km and 110 km with a spatial resolution of 2km.

A.4 AURA:Microwave Limb Sounder

The Earth Observation Satellite(EOS) MLS-Aura[139] makes measurements of atmospheric composition, temperature, humidity and cloud ice that are mandatory to (1) track the stratospheric ozone layer stability, (2) enhancement of predictions related to climate change and variability, and (3) maintaining the global air quality. MLS observes thermal microwave emission from Earth's 'limb' (the edge of the atmosphere) viewing forward along the Aura spacecraft flight direction, scanning its view from the ground to 90 km every 25 seconds. Aura is placed at 705 km altitude of near polar orbit. AURA provides a global coverage with 15 orbits per day. Vertical profiles are retrieved every 165 km along the suborbital track, covering -82^oS to 82^oN latitudes on each orbit.

A.5 Horizontal Wind Model

The Horizontal Wind Model(HWM)[113, 140] is an empirical horizontal neutral wind model. Taking into account the earth's rotation, tilt, and orbit around the sun, the model represents changes in the middle and upper atmosphere winds. With a vertical altitude resolution of 1km, the HWM provides zonal and meridional winds for specified latitude, longitude, and time. Model and data root mean square differences are $15m/s$ in the mesosphere and $10m/s$ in the stratosphere for zonal winds, and $10m/s$ and $5m/s$ respectively for meridional winds.

A.6 SABER-TIMED Instrument

The Sounding of the Atmosphere using Broadband Emission Radiometry (SABER) instrument is one of four instruments on NASA's TIMED (Thermosphere Ionosphere Mesosphere Energetics Dynamics) satellite. SABER provides the global measurements of the atmosphere with a 10-channel broadband limb-scanning infrared radiometer having spectral range from 1.27 µm to 17 µm. The measurements provide the kinetic temperature, pressure, geopotential height, volume mixing ratios for the trace species, the cooling and heating rates of CO2, O3, and O2 bands, and the chemical heating rates for the atmospheric region extending from 60 km to 80 km covering the mesosphere and lower thermosphere.