Machine Learning Techniques for Prediction of Storms using GNSS Receiver measurements

A thesis submitted

in partial fulfillment for the degree of

Doctor of Philosophy

by

Deepak Singh Bisht

Under the joint supervision of

Dr. Rama Rao Nidamanuri

Professor
Department of Earth and Space Sciences
IIST, Thiruvananthapuram,
Department of Space, Government of India

Dr. Thota Narayan Rao

Sci. /Eng.–'SG' & Head
Clouds and Convective Systems Group
NARL, Gadanki,
Department of Space, Government of India

Department of Earth and Space Sciences
Indian Institute of Space Science and Technology
Thiruvananthapuram-695547

June 2025

CERTIFICATE

This is to certify that the thesis entitled ***'Machine Learning Techniques for Prediction of Storms using GNSS Receiver measurements'*** *submitted by* ***Mr. Deepak Singh Bisht*** *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 research work carried out by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institution or University for the award of any degree or diploma.*

Dr. Thota Narayan Rao
Sci./Eng.– 'SG' & Head
Clouds and Convective Systems Group
National Atmospheric Research Laboratory,
Gadanki, Dept. of Space, GoI.

Dr. Rama Rao Nidamanuri
Professor
Dept. of Earth and Space Sciences
Indian Institute of Space Science and Technology,
Thiruvananthapuram, Dept. of Space, GoI.

Dr. Anand Narayanan
Professor & Head
Dept. of Earth and Space Sciences
Indian Institute of Space Science and
Technology, Thiruvananthapuram, Dept. of
Space, GoI.

DECLARATION

I declare that this thesis entitled ***'Machine Learning Techniques for Prediction of Storms using GNSS Receiver measurements'*** *submitted in partial fulfillment of the degree of Doctor of Philosophy is a record of original work carried out by me under the joint supervision of* ***Dr. T. Narayana Rao,*** *and,* ***Dr. Rama Rao Nidamanuri****, and has not formed the basis for the award of any other degree or diploma, in this or any other Institution or University. 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: 05 June 2025

Deepak Singh Bisht

(SC18D016)

This thesis is dedicated to

Bhagwaan Krishna, Maa Durga, Golu Devta, &

My beloved Family

ACKNOWLEDGEMENTS

Writing a thesis is harder than I thought and more rewarding than I could have ever imagined but what made the entire process look like a journey with many challenges and enriching experiences worth cherishing is the constant efforts of numerous people in supporting me throughout the same. It gives me abundant pleasure in acknowledging, what remains due to all of them, a sense of gratitude and gratefulness.

To start with, I acknowledge with utmost gratitude, a sense of indebtedness towards my supervisors Prof. Rama Rao Nidamanuri, Professor, IIST and Dr. T. Narayan Rao, Scientist, NARL for their patient guidance, enthusiastic encouragement, and constructive criticism throughout the course of my Ph.D. journey. Their vision, guidance, and unwavering support allowed me the freedom to explore and transformed me from a doctoral student into an independent researcher.

I would like to extend my sincere appreciation to the members of my doctoral committee Dr. Govindan Kutty M., Dr. Basudeb Ghosh, and Dr. Amit P. Kesarkar for their thoughtful suggestions, continuous monitoring, and encouragement.

I would like to express the gratitude to National Atmospheric Research Laboratory for providing me with the resources and opportunities to pursue this research and for the financial support. I also convey my gratitude to the technical staff, Padma Rao, Vijayan, and Mohan whose contributions made this work possible. I am also thankful to the Clouds and Convective Systems Group scientists at NARL, Dr. B. Radhakrishna, Dr. Anish Nair, Dr. Satheesh, Mr. Chandrakanth, Ms. Gayatri, for their invaluable support, and encouragement during the course of my tenure.

I owe a special debt of gratitude to Dr. Jayachandran V. and Akshit Sharma for their camaraderie, intellectual discussions, and support throughout this journey. Akshit introduced me to machine learning which ultimately laid the foundation of my research work. His selfless assistance in research as well as in personal matters are priceless. The brief but inspiring discussions with Dr. Jayachandran always motivated me to work harder. With them, I felt like a family.

I have been overwhelmed by the kind support of my friends and lab mates, Dr. Shridhar, Abhijeet, Donali, and Nidhi Sharma who made this journey memorable. The cricket matches with

Abhijeet, Jitesh, and Sritam provided much-needed breaks, helping me relieve stress and making the challenging PhD journey more manageable. A special mention goes to Kasturi Di, and Harshit Upreti who supported me through both the highs and lows, not only during my Ph.D. journey but in my personal life as well.

It is my privilege to express my regards and sincere thanks to Dr. Praveen Kumar, NCMRWF for providing substantial guidance and support during my final years of Ph.D. I am also thankful to my colleagues Dr. Vivekananda Hazra, Dr. Gopinath, Dr. Ashutosh, Dr. Durgesh, Mr. Navin Kumar, Dr. Bibhuti, and Dr. Upal Saha at NCMRWF, India.

It would be an incomplete task and unjust if I do not mention my family. My parents Mrs. Radha Bisht and Mr. Narayan Singh Bisht whose sacrifices, love, and endless support have been my motivation throughout this journey. They backed me at every moment and provided me nothing less than I deserved. A hearty warm thanks to my younger sister Neelam Bisht for always standing by me no matter what came my way. To my partner, Nidhi, I am so glad that our paths crossed. You are always there for me, whether it's talking about science or life, and I am excited for the next chapters of our journey together.

Last but certainly not least, I offer my heartfelt praise to Kanha Ji, Golu Devta, and Maa Durga for their love, guidance, and blessings throughout my life. To anyone I may have unintentionally overlooked, I extend my sincere gratitude.

June 2025 (Deepak Singh Bisht)

ABSTRACT

Integrated Water Vapor (IWV) is a crucial parameter in atmospheric studies, influencing weather patterns, storm development, and climate dynamics. This thesis investigates the diurnal variability of IWV, its prediction using machine learning (ML) techniques, and its application in storm nowcasting using data from a GNSS receiver network in Gadanki, India, over a three-year period (2018–2021). The study provides insights into IWV dynamics, demonstrating the potential of GNSS-based IWV data and advanced ML approaches in enhancing storm prediction and disaster preparedness.

The diurnal and seasonal variations of IWV were analyzed using harmonic analysis on high-resolution GNSS-derived IWV data. The study revealed that the diurnal (24-hour) harmonic dominates IWV variations, being significant for 93% of the time, while the semi-diurnal (12-hour) harmonic is significant for 36% of the time. Seasonal analysis showed that diurnal amplitudes are highest during monsoon, while semi-diurnal amplitudes peak in winter and post-monsoon. Vertically Integrated Moisture Flux Convergence (VIMFC) emerged as the primary driver of IWV variability, dictating the timing and magnitude of IWV peaks across all seasons. This detailed understanding of IWV variability establishes a foundation for its application in storm prediction.

A Light Gradient Boosting Machine (LightGBM) model was developed to predict IWV with lead times ranging from 30 to 120 minutes. The model was trained using GNSS data, meteorological parameters, and derived IWV features, achieving a correlation coefficient of > 0.99 with observed IWV and a root mean square error (RMSE) of < 1 mm for shorter lead times. Even for longer lead times, the RMSE remained < 3 mm, demonstrating the robustness of the model. Sensitivity analysis revealed that IWV features are critical for short-term predictions, while additional predictors gain importance for longer lead times and stormy days.

Building on this, a hybrid nowcasting model was developed that combines predicted IWV, brightness temperature, and threshold-based parameters. This model achieved an accuracy of 97.5% with a false alarm rate of just 5%, significantly improving upon traditional methods. By utilizing predicted IWV, the model provides up to two hours of lead time for storm events, enabling better disaster preparedness. The study also validated the predictive potential of IWV gradients, demonstrating that moisture buildup occurs 1–4 hours prior to storm onset.

Additionally, multiple ML models, including Random Forest (RF), Decision Trees (DT), XGBoost, Logistic Regression, CatBoost and Artificial Neural Networks (ANN), were evaluated for storm nowcasting using GNSS-derived IWV and meteorological data. Random Forest emerged as the best-performing model, achieving high accuracy and robustness with AUC-ROC and Cohen's Kappa scores exceeding those of other models. The combination of IWV and brightness temperature was identified as the most significant predictor set, highlighting the utility of GNSS-based IWV data in real-time storm prediction systems.

This thesis underscores the importance of GNSS-based IWV data in understanding atmospheric processes and advancing storm nowcasting techniques. By integrating data-driven methods and physical insights, the research provides a robust framework for early storm warning systems, aiding in disaster management, and reducing the socio-economic impacts of extreme weather events.

Table of Contents

Table of Figures

List of Tables

Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
AUC-ROC	Area Under the Receiver Operating Characteristic Curve
AWS	Automatic Weather Stations
CAPE	Convective Available Potential Energy
DT	Decision Tree
ECMWF	European Centre for Medium-Range Weather Forecasts
EPS	Ensemble Prediction System
ERA5	Reanalysis Fifth Generation
FAR	False Alarm Rate
FEMA	Federal Emergency Management Agency
FTIR	Fourier Transform Infrared Spectroscopy
GBDT	Gradient Boosting Decision Trees
GEFS	Global Ensemble Forecast System
GEO	Geostationary Earth Orbit
GFS	Global Forecast System
GNSS	Global Navigational Satellite Systems
GOSS	gradient-based one-side sampling
GPS	Global Positioning System
HPC	High-performance computing

IGS	International GNSS Service
INSAT	Indian National Satellite System
IWV	Integrated Water Vapor
JD	Julian Day
LEO	Low Earth Orbit
LIDAR	Light Detection and Ranging
LightGBM	Light Gradient Boosting Machine
MAE	Mean Absolute Error
MCC	Mesoscale Convective Complexes
MCS	Mesoscale Convective Systems
MJO	Madden-Julian Oscillation
ML	Machine Learning
NARL	National Atmospheric Research Laboratory
NWP	Numerical Weather Prediction
POD	Probability of Detection
PPP	Precise Point Positioning
PWV	Precipitable Water Vapor
RF	Random Forest
RMSE	Root Mean Square Error
SH	Specific Humidity
SLHF	Surface latent heat flux
SM	Soil Moisture

TEC	Total Electron Count
UTC	Coordinated Universal Time
VIMF	Vertically Integrated Moisture Flux
VIMFC	Vertically Integrated Moisture Flux Convergence
VMF	Vienna Mapping Function
WRF	Weather Research and Forecasting
WVR	Water Vapor Radiometer
ZHD	Zenith Hydrostatic delay
ZTD	Zenith Total Delay
ZWD	Zenith Wet Delay

List of Symbols

Δt_{iono}	Ionospheric Delay
f	Frequency
H_z	Hertz
P_s	Surface pressure
T_s	Surface Temperature
ϕ	Latitude
H	Height of GNSS Station above sea level
Π	Conversion factor
ρ_w	Density of liquid water
R_d	Specific gas constant
R_w	Specific gas constant for water vapor
T_m	Mean temperature of the atmosphere
k_3	Empirical Constant
k_2'	Empirical Constant
a_n	Coefficient representing the sine component for the n-th harmonic
b_n	Coefficient representing the cosine component for the n-th harmonic
t	Time in hours
ϵ	Higher-order harmonics and other non-periodic components
σ_n	Phase shift of the n-th harmonic
A_n	Amplitude of the n-th harmonic
P_n	Time of day when the IWV reaches its peak

$\boldsymbol{\sigma^2}$	Variance of the residuals
$\boldsymbol{\alpha}$	Significance level
u, v	Zonal and meridional wind components
$\mathbf{p_t}$	Pressure at the top of the atmospheric layer
g	Acceleration due to gravity
$\frac{\partial(\mathbf{uq})}{\partial \mathbf{x}}, \frac{\partial(\mathbf{vq})}{\partial \mathbf{y}}$	Horizontal gradients of moisture flux
$\boldsymbol{f(x)}$	The final predicted value for input feature vector x
$\boldsymbol{M}$	The total number of decision trees
$\boldsymbol{T_m(x)}$	The prediction from the m-th tree
$\boldsymbol{\Omega(T_m)}$	regularization term
$\boldsymbol{p(x)}$	The probability of storm occurrence
$\boldsymbol{w_0}$	Intercept term
$\boldsymbol{w_i}$	Coefficients for the features x_i
$\boldsymbol{z^{(l)}}$	The linear combination of weights, inputs, and biases for the l -th layer
$\boldsymbol{A^{(l)}}$	Activation output for the l -th layer
$\boldsymbol{W^{(l)}}$	Weight matrix for layer l
$\boldsymbol{b^{(l)}}$	Bias term for layer l
$\boldsymbol{r}$	Correlation coefficient
$\boldsymbol{R^2}$	Coefficient of Determination
$\boldsymbol{p_o}$	Observed agreement

p_e	Expected agreement
κ	Kappa score
T_b	Brightness Temperature

Chapter 1

Introduction

1.1 Introduction to storms

Storms are severe weather phenomena characterized by violent winds, heavy precipitation (rain, snow, hail), lightning, and thunder and often cause severe damage to property and life (Figure 1.1) (Williams, 2017; Halverson, 2024; Atlas *et al.* 1963). These disturbances occur when the atmosphere becomes highly unstable due to various factors. like surface heating, topographic lifting of air and convergence of air masses, etc., while the mid- and upper-tropospheric moisture aids the rapid development of storms. The storms are on the rise in recent years causing flash floods and lightning and the increase is partly attributed to climate change.

(a) Flash floods: These occur when excessive rainfall leads to rapid flooding, particularly in urban areas with poor drainage and in mountainous terrain. Flash floods are one of the deadliest consequences of severe thunderstorms, especially in regions with limited capacity to absorb heavy rains (Doswell *et al.* 1996).

(b) Hailstorms: These are caused by strong updrafts within thunderstorms that lift raindrops into freezing levels, resulting in the formation of hailstones that can cause damage to crops, vehicles, and buildings (Foote *et al.* 2016). In India, hailstorms are commonly observed in the West Bengal during the premonsoon season due to the convergence of different air masses.

(c) Lightning strikes: Thunderstorms are the primary drivers of lightning activity, which poses risks to human safety, infrastructure, and wildfires (Rakov & Uman, 2003). Recent World Meteorological Organization (WMO) reports show that lightning now become a major threat to human life.

Figure 1.1: *Examples of destruction caused by storms include: (a) the 2013 Kedarnath disaster, and (b) flooding of human settlements.*

The intensity and scale of storms vary, and they can be classified based on their size, duration, and the processes driving them (Newton, 1967; Dimri *et al.* 2015).

1.1.1 Small-Scale Storms

Small-scale storms, often referred to as local thunderstorms, are typically short-lived and affect relatively confined areas, usually ranging from a few square kilometers to tens of kilometers (Moller, 2001; Brooks and Weiss, 1999, Newton, 1967). These storms develop due to localized convective processes where warm, moist air rises rapidly, leading to cloud formation and precipitation. Thunderstorms are a common example of small-scale storms and can bring hazards such as lightning, heavy rainfall, hail, and even tornadoes (Harper *et al.* 2000, Atlas *et al.* 1963).

1.1.2 Mesoscale Storms

Mesoscale refers to weather systems that occur on a scale ranging from 20 to several hundred kilometers, persisting for hours or even days (Wang and Johnson, 2006). Examples include squall lines, mesoscale convective systems (MCS), and mesoscale convective complexes (MCC). These systems are driven by organized convection and often involve clusters of thunderstorms that can extend across large areas, leading to significant precipitation and strong winds (Houze Jr, 2004).

Squall lines are one of the example. These are long, narrow bands of thunderstorms that can stretch across hundreds of kilometers. They typically form ahead of cold fronts and produce severe weather, including strong winds and heavy rain (Houze, 2014). Apart from that MCC and MCS systems are there. These systems represent clusters of thunderstorms that behave as a single organized system, often producing widespread rainfall, flash flooding, and strong winds. MCCs can cover an area larger than 100,000 square kilometers, persisting for 6-12 hours or more (Maddox, 1980; Tyagi *et al.* 2012). Mesoscale systems are important because they often lead to prolonged and widespread severe weather events. For example, MCSs are responsible for much of the heavy rainfall and flash flooding observed in many regions, particularly in the mid-latitudes and tropical regions (Schumacher & Johnson, 2005).

1.1.3 Synoptic-Scale Storms

At the largest scale are tropical cyclones, hurricanes, and typhoons, which are essentially the same type of storm but named differently depending on their location (Emanuel, 2003). These massive

storm systems form over warm ocean waters and can grow to span thousands of kilometers. They are driven by the energy released from the condensation of water vapor, making them highly destructive in terms of both wind speed and rainfall.

For example, low-pressure systems form over warm tropical waters and can intensify into hurricanes (Atlantic) or typhoons (Pacific) or cyclones (North Indian Ocean). Their structure includes a well-defined eye surrounded by an eyewall of intense thunderstorms, which generate strong winds and heavy rains (Emanuel, 2003). Hurricanes are classified based on their wind speed according to the Saffir-Simpson scale, with Category 5 hurricanes having sustained winds greater than 157 mph (252 km/h) (NOAA, 2021). These storms can cause catastrophic damage due to wind, storm surges, and flooding.

These large-scale systems have far-reaching impacts, often affecting entire regions and causing significant disruptions to infrastructure, ecosystems, and human settlements. For instance, Hurricane Katrina (2005) and Typhoon Haiyan (2013) caused widespread devastation and loss of life due to storm surges, flooding, and extreme winds (Knabb *et al.* 2006; Lagmay *et al.* 2015).

Figure 1.2: *Example of (a) small (b) meso (c)synoptic scale storms (Source: Google images).*

1.2 Why Is Storm Prediction Important?

Storms are among the most destructive natural phenomena, often resulting in significant human and economic losses. Accurate storm prediction is essential for mitigating these impacts, as it provides the basis for early warnings and preparedness strategies. The ability to forecast storms, especially high-impact events like cyclones, thunderstorms, and hurricanes, directly affects a

community's ability to respond effectively, thus saving lives and reducing damage to property and infrastructure (Morss *et al.* 2022; Frame *et al.* 2017; Esteban *et al.* 2015).

1.2.1 Human Lives and Safety

Storms are a leading cause of mortality during natural disasters, with fatalities often arising from drowning, windborne debris, and structural collapses. Accurate storm prediction allows authorities to issue early warnings (Hermans *et al.* 2022) giving individuals and communities the opportunity to take protective measures, such as evacuation or sheltering in place. For example, Hurricane Katrina in 2005 caused over 1,800 deaths, with many lives lost due to the failure to evacuate in time (Knabb *et al.* 2006). However, effective forecasting and timely evacuation orders have been credited with reducing death tolls in other major storm events (Blake *et al.* 2013). Storm warnings allow people to avoid traveling or working in dangerous areas, such as coastal zones during storm surges or flood-prone areas during intense rain. This is particularly critical for vulnerable populations, such as those living in poorly constructed buildings or coastal communities exposed to high storm surge risks.

1.2.2 Economic Damage and Infrastructure Protection

Storms cause billions of dollars in economic losses each year due to the destruction of homes, businesses, infrastructure, and agriculture. The ability to predict these events allows governments and businesses to prepare for the impending disaster by securing property, deploying emergency services, and protecting critical infrastructure. The World Bank estimates that extreme weather events, including storms, cost the global economy more than $300 billion annually (Hallegatte *et al.* 2016; Vogt-Schilb, & Bangalore, 2017). Economic impacts are most profound in sectors like construction, transportation, and energy, where disruptions can lead to long-term economic damage. For instance, Hurricane Harvey (2017) caused $125 billion in damages in Texas, much of which was attributed to flooding that could have been mitigated by better urban planning and preparedness (Blake & Zelinsky, 2018). Accurate forecasts allow cities and regions to implement mitigation strategies such as reinforcing buildings, shoring up levees, or shutting down critical infrastructure like airports and power plants ahead of a storm. For example, Japan has significantly reduced storm-related damage by implementing stringent building codes and flood defenses, supported by accurate storm predictions and risk management strategies (Mimura *et al.* 2011).

1.2.3 Agriculture and Food Security

Agriculture is highly vulnerable to storm events, with impacts including crop destruction, soil erosion, and livestock loss. Predictive capabilities help farmers and agricultural businesses protect crops and livestock ahead of storms, improving food security and reducing economic losses. For example, early warnings about the 2019 Typhoon Hagibis in Japan helped farmers mitigate losses by harvesting crops early and securing livestock before the storm hit (Shimozono *et al.* 2020). In regions heavily reliant on agriculture, such as South Asia or sub-Saharan Africa, storm prediction is vital for minimizing the economic and social impacts of extreme weather events. Storms can devastate crops, leading to food shortages and rising food prices, which disproportionately affect lower-income communities (FAO, 2015).

1.2.4 Water Management and Flood Control

Storms often lead to extreme rainfall and flooding, which can overwhelm drainage systems, cause rivers to overflow, and result in widespread water damage. Predictive models provide critical information to water management authorities, allowing them to implement preemptive measures like reservoir drawdowns, the construction of temporary flood barriers, and the preparation of emergency response teams (Medema *et al.* 2008). In many cases, storm predictions can help mitigate the risks associated with urban flooding, which has become a growing problem due to increased urbanization and insufficient drainage systems (Bakker & Buisman, 2020; Zhang *et al.* 2016).

1.2.5 Risk Assessment and Disaster Preparedness

Accurate storm prediction is essential for risk assessment and disaster preparedness. Governments, businesses, and communities rely on forecast data to plan and execute disaster response strategies. Risk assessments informed by storm forecasts help prioritize which regions are most likely to be affected and ensure that resources are deployed efficiently. This reduces the financial and logistical burden on emergency services and ensures that help reaches vulnerable areas on time. For example, predictive models have improved the deployment of emergency services during hurricanes in the United States. Agencies such as the Federal Emergency Management Agency (FEMA, 2020) use storm predictions to pre-position personnel, supplies, and equipment in areas likely to be affected by a storm, enhancing response times and overall effectiveness (FEMA, 2020)

1.3 Conventional Techniques for Predicting Storms

Storm prediction has significantly improved over the years through advances in computational capabilities and observational technologies (Bauer *et al.* 2015, Liu *et al.* 2012). However, conventional techniques such as Numerical Weather Prediction (NWP) models and Ensemble Forecasting remain the foundation of modern meteorological forecasting. Both approaches have distinct strengths and limitations, which must be considered when utilizing them for storm prediction.

1.3.1 Numerical Weather Prediction (NWP) Models

Numerical Weather Prediction (NWP) models are essential for simulating atmospheric processes and forecasting storms (Steppeler *et al.* 2003). These models operate by solving a set of mathematical equations known as the Navier-Stokes equations, which describe the flow of fluid in the atmosphere (Temam 1995). NWP models consider a wide array of meteorological variables, such as temperature, humidity, wind speed, and pressure, to simulate the behavior of the atmosphere over time. They are utilized on both global and regional scales, providing detailed forecasts ranging from a few hours to several days ahead. NWP models divide the atmosphere into a grid system where each grid point represents the state of the atmosphere. The model use initial conditions observational data from satellites, radars, radiosondes, and surface stations to predict how the atmosphere will evolve.

Prominent examples of NWP models include: Global models like the European Centre for Medium-Range Weather Forecasts (ECMWF) model and the Global Forecast System (GFS) used by the United States (Bauer *et al.* 2015). Regional models, such as the Weather Research and Forecasting (WRF) model, which is applied to provide higher-resolution forecasts over smaller geographic regions (Skamarock *et al.* 2019).

There are some disadvantages associated with NWP models as they are Computationally Intensive. Running NWP models requires significant computational power due to the vast amount of data processing and the complexity of the atmospheric equations involved. NWP models must solve millions of calculations across thousands of grid points for each time step of the forecast. The finer the resolution (i.e., the smaller the grid points), the more accurate the prediction, but this also

increases computational demands. High-performance computing (HPC) resources are required to run these models efficiently (Bauer *et al.* 2015).

Apart from that it suffers with Initial Condition Errors and Model Bias. The accuracy of NWP models depends heavily on the quality of the initial conditions—the data fed into the model at the start of the simulation. Errors in observational data, incomplete data coverage (especially over oceans and remote regions), and delays in real-time data assimilation can lead to inaccuracies in the initial conditions. Even small errors in the initial state can propagate and amplify over time, a phenomenon known as sensitivity to initial conditions (Lorenz, 1969). Furthermore, NWP models can introduce biases due to simplifications and assumptions made in representing atmospheric processes, such as cloud microphysics, convection, and turbulence (Rodwell & Palmer, 2007).

1.3.2 Ensemble Forecasting

Ensemble forecasting is an extension of traditional NWP that seeks to account for the inherent uncertainties in weather prediction. Instead of running a single deterministic forecast, ensemble forecasting involves running multiple model simulations, each with slightly perturbed initial conditions or different model configurations. These simulations generate a range of possible outcomes, providing a probabilistic forecast rather than a single prediction (Anderson 1996). The concept of ensemble forecasting is based on the chaotic nature of the atmosphere (Leutbecher, M. and Palmer 2008). Small uncertainties in initial conditions can lead to vastly different outcomes over time. By running multiple forecasts with small variations in these conditions, meteorologists can assess the range of potential outcomes and the likelihood of different storm scenarios.

Prominent examples of ensemble systems include ECMWF Ensemble Prediction System (EPS), which produces medium-range probabilistic forecasts for up to 15 days (Buizza *et al.* 2003) and **NCEP Global Ensemble Forecast System (GEFS)** used by the National Weather Service in the United States, which provides ensemble forecasts to improve the accuracy and reliability of medium- to long-range predictions (Toth & Kalnay, 1997). Ensemble forecasts provide a probability distribution of possible outcomes, allowing forecasters to express the likelihood of various storm scenarios (e.g., storm intensity, track, and timing). They help mitigate the limitations of deterministic forecasts, which are sensitive to initial condition errors and cannot fully capture the uncertainty of storm evolution.

There are few disadvantages associated with this technique including it is highly Resource-Intensive. Ensemble forecasting requires significant computational resources because multiple simulations must be run simultaneously. Running 20 to 50 ensemble members with slight variations in initial conditions can be demanding, especially at higher resolutions (Buizza *et al.* 2003). This computational burden is one of the key challenges in expanding the use of ensemble forecasting, particularly for real-time applications where fast results are needed. Another challenges of ensemble forecasting is balancing model complexity with resolution. High-resolution models can provide more detailed forecasts, but they also increase computational costs. As a result, ensemble forecasts often sacrifice spatial or temporal resolution to maintain a feasible number of ensemble members. This trade-off can lead to less detailed predictions, especially in regions requiring high-resolution data to accurately predict small-scale phenomena like thunderstorms or local flooding (Leutbecher & Palmer, 2008).

Despite these limitations, ensemble forecasting has become a critical tool for storm prediction (Wu *et al.* 2020; Flowerdew *et al.* 2010). It allows forecasters to better understand the range of possible outcomes and communicate forecast uncertainty to decision-makers and the public. For example, ensemble forecasts have been crucial for predicting the potential tracks of storms, which are notoriously difficult to predict due to their sensitivity to initial conditions and small changes in environmental factors (Zhu *et al.* 2002).

1.3.3 Radar-Based Nowcasting

Radar-based nowcasting plays a critical role in short-term storm prediction. By using real-time radar data, this technique can provide high-resolution forecasts for the next few hours, making it particularly useful for predicting thunderstorms and severe weather (Wilson *et al.* 1998; Prudden *et al.* 2020).

One of the primary advantages of radar-based nowcasting is its high temporal and spatial resolution, which allows for detailed observation of storm structure, intensity, and movement. This level of detail enables timely and accurate short-term forecasts, crucial for predicting severe weather events such as thunderstorms and squall lines. For example, studies have demonstrated that radar data provides critical insights into the organization and life cycle of convective systems, enabling forecasters to issue timely warnings (Wilson *et al.* 1998). Additionally, radar systems

provide continuous real-time data, which is essential for tracking the rapid development of localized weather phenomena (Pfister *et al.* 1999). This makes radar-based nowcasting particularly effective for localized predictions, where detailed and area-specific forecasts are required to mitigate potential impacts on communities and infrastructure (Kober and Tafferner, 2009).

Despite its advantages, radar-based nowcasting has notable limitations. It is most effective for short-term forecasts, typically up to a few hours, as it relies on the extrapolation of current storm movement and intensity. This approach becomes less effective for longer-term predictions because radar data alone cannot capture broader atmospheric trends or the dynamic evolution of storm systems. For instance, convective storms can intensify or dissipate rapidly due to changes in atmospheric conditions, which radar-based methods often fail to predict accurately (Wilson *et al.* 1998, Davis 2001). However, the accuracy of radar-based nowcasting is highly dependent on radar coverage. In regions with sparse or low-quality radar infrastructure, the reliability of forecasts may be significantly reduced. Ground clutter, signal attenuation, and beam blockages caused by terrain or tall structures can further degrade the quality of radar observations (Fabry *et al.* 1994). Another challenge is the complex interpretation of radar data, which requires expertise to account for artifacts and noise, such as anomalous propagation, that can introduce errors in the forecast (Collier, 1996).

1.3.4 Radiosonde and Satellite Observations for nowcasting application

Radiosonde and satellite observations contribute significantly to the larger picture of storm prediction by providing valuable atmospheric data. Radiosondes are balloon-borne instruments that provide vertical profiles of temperature, humidity, and pressure, which are crucial for understanding the atmospheric structure and identifying potential instability that could lead to storm development (Vaisala, 2013). Satellites, on the other hand, offer broad coverage of atmospheric conditions, allowing for continuous monitoring of large areas, which is particularly useful for tracking storm systems over oceans and remote regions.

Radiosondes, which are balloon-borne instruments, provide detailed vertical profiles of temperature, humidity, and pressure. These data are invaluable for understanding the thermodynamic structure of the atmosphere and calculating instability indices such as Convective Available Potential Energy (CAPE) and the K Index, which are commonly used to assess the

potential for convective storms (Vaisala, 2013). The high vertical resolution of radiosonde data allows meteorologists to identify atmospheric layers prone to instability, making it an essential tool for localized storm forecasting. Satellites, on the other hand, provide broad spatial coverage, enabling continuous monitoring of large geographic areas, including remote and oceanic regions where ground-based observations are sparse. Real-time data from satellites, such as those from the Geostationary Operational Environmental Satellites (GOES), offer critical information about cloud formation, water vapor distribution, and storm movement, facilitating timely warnings and decision-making during rapidly evolving weather events (Schmit *et al.* 2017).

Despite their advantages, radiosonde and satellite observations have notable limitations. Radiosondes are typically launched only twice daily, resulting in limited temporal resolution that can make it challenging to capture rapid atmospheric changes associated with storm development. Furthermore, their spatial coverage is restricted to specific launch locations, leaving large data gaps over oceans, mountainous regions, and sparsely populated areas (Seidel *et al.* 2009). While satellites excel in providing broad coverage, their spatial resolution is often insufficient to detect small-scale atmospheric features, such as localized convection, which are critical for accurate storm predictions (Fabry *et al.* 1994). Additionally, interpreting satellite data is complex, as the information retrieved is influenced by various atmospheric and surface conditions, requiring sophisticated algorithms and expertise to ensure accuracy (Collier, 1996). These limitations highlight the need for integrating radiosonde and satellite data with complementary techniques, such as GNSS and radar observations, to improve the accuracy and lead time of storm forecasts.

1.4 Evolving Methods of Storm Prediction: Water Vapor Variation-Based Studies

Nowcasting refers to the prediction of weather a few minutes to 6 hours in advance. As traditional methods of storm prediction, such as numerical weather prediction (NWP) models and radar-based nowcasting, face limitations in accuracy and lead time, newer approaches have emerged to complement these methods. One such evolving method is the study of water vapor variations in the atmosphere. Water vapor plays a crucial role in the formation and intensification of storms, and its continuous monitoring offers an innovative way to predict severe weather events like thunderstorms, hurricanes, and tropical cyclones.

1.4.1 Importance of Water Vapor in Storm Development

Water vapor is a fundamental component of the Earth's atmosphere and serves as the primary source of moisture for cloud formation and precipitation (Jacob, 2001; Schneider *et al.* 2010). Latent heat release, which occurs when water vapor condenses into liquid, is a key energy source that fuels storm development and intensification (Williams, 2017). The higher the concentration of water vapor in a given region, the greater the potential for storm formation, as the available moisture can lead to strong convective processes (Ralph *et al.* 2004; Sherwood *et al.* 2010; Del Genio and Kovari, 2002). In particular, water vapor is vital for the development of mesoscale convective systems and tropical cyclones. For example, in tropical regions, where the atmosphere is typically rich in moisture, water vapor variations can indicate the onset of cyclonic activity. As water vapor accumulates in the lower layers of the atmosphere, it creates conditions conducive to deep convection, leading to the development of storm systems (Emanuel, 1991).

1.4.2 Water Vapor Variation as a tool for Nowcasting Storms

Water vapor buildup has been increasingly recognized as a precursor to storm formation. By monitoring changes in atmospheric water vapor content, meteorologists can identify areas of potential convective activity well before traditional forecasting tools detect the early signs of storms. For instance, rapid increases in water vapor in the lower troposphere can indicate the likelihood of severe thunderstorm development (Schmit *et al.* 2017).

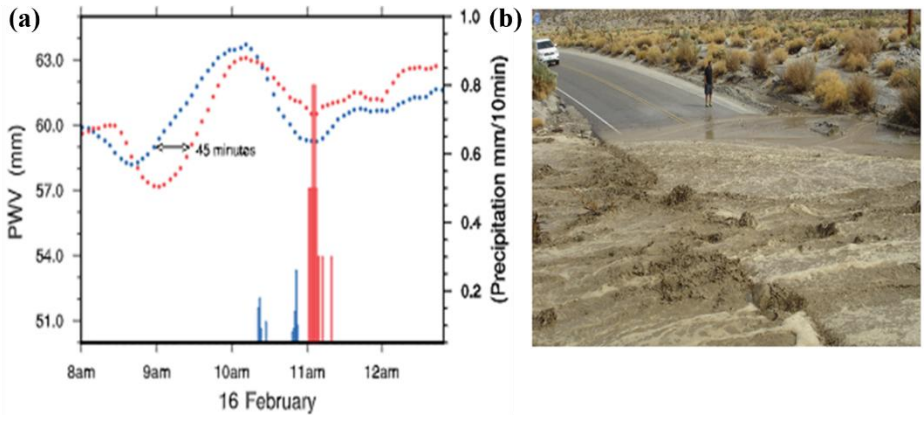

Figure 1.3: *(a) Moisture built up before the storm's occurrence (Adams et al., 2011) (b) Flash flood caused by storm.*

Recent research has shown that Integrated Water Vapor (IWV) measurements—an estimate of the total column of water vapor in the atmosphere can provide valuable insights into storm predictability. IWV measurements have been linked to the occurrence of atmospheric rivers—narrow corridors of concentrated moisture that can lead to heavy precipitation and flooding when they make landfall (Zhu & Newell, 1998). Monitoring these water vapor-rich phenomena allows meteorologists to predict potential storm impacts with greater accuracy.

Figure 1.3 (a) is representing a case study (Adams *et al.* 2011) where it can be seen abnormal enhancement in IWV before storm occurrence at 2 different locations. Identifying such storms can be useful for nowcasting of storm events.

1.5 Importance of Water Vapor in the Atmosphere

Water vapor plays a fundamental role in atmospheric processes and weather patterns, acting as a critical regulator of both weather dynamics and climate systems. Its unique properties and behaviors are essential in various meteorological phenomena, influencing everything from local weather events to global climate patterns. Understanding the distribution, movement, and impact of water vapor in the atmosphere is crucial for accurate weather prediction and climate modeling.

1.5.1 Water Vapor as a Greenhouse Gas

Water vapor is the most abundant greenhouse gas in the atmosphere, contributing significantly to the natural greenhouse effect, which keeps the Earth's surface warm enough to sustain life (Mitchell, 1989; Held and Soden ,2000; El Zein and Chehayeb *et al.* 2005). It absorbs and re-emits infrared radiation, trapping heat within the atmosphere. According to Kiehl and Trenberth (1997), water vapor accounts for approximately 50% of the Earth's greenhouse effect, while clouds contribute an additional 25%. This makes water vapor a key player in regulating the Earth's temperature. Unlike other greenhouse gases like carbon dioxide (CO_2) and methane (CH_4), which are directly emitted by human activities, the concentration of water vapor is controlled indirectly through feedback processes. As the atmosphere warms due to increased CO_2, for example, the capacity of the atmosphere to hold water vapor increases, leading to more evaporation and higher concentrations of atmospheric water vapor. This feedback amplifies warming, creating what is known as a positive feedback loop (Held & Soden, 2000).

1.5.2. Role in the Hydrological Cycle

Water vapor is the gaseous phase of water within the hydrological cycle, the continuous movement of water between the Earth's surface and the atmosphere (Mitchell, 1989; Held and Soden ,2000; El Zein and Chehayeb *et al.* 2005). It forms when water evaporates from oceans, rivers, lakes, and other water bodies or through the process of transpiration in plants. Once in the atmosphere, water vapor can travel long distances before condensing into clouds and eventually returning to the surface as precipitation (rain, snow, sleet, or hail). The cycling of water vapor is crucial for maintaining the Earth's climate equilibrium (Schneider *et al.* 2010; Galewsky *et al.* 2016; Allan *et al.* 2020). It facilitates the transfer of heat between the surface and the atmosphere. When water evaporates, it absorbs heat from the environment, cooling the surface. This absorbed energy, known as latent heat, is released when water vapor condenses back into liquid during cloud formation, playing a vital role in driving atmospheric circulation and storm development (Rogers & Yau, 1989).

1.5.3. Influence on Weather Systems and Storm Development

Water vapor is a critical factor in the development of clouds, precipitation, and storms (Sherwood *et al.* 2010). The presence of sufficient water vapor in the lower atmosphere can lead to the formation of cumulus clouds, which are precursors to thunderstorms and other severe weather phenomena. As water vapor rises, it cools and condenses, releasing latent heat. This process drives convection, which is responsible for the development of towering thunderclouds and, in extreme cases, the formation of severe storms such as hurricanes, typhoons, and tornadoes (Emanuel, 2003).

Tropical cyclones, for example, are powered primarily by the condensation of water vapor in the atmosphere. As water evaporates from warm ocean surfaces, it increases the moisture content of the air. When this moist air rises and condenses, it releases large amounts of latent heat, which intensifies the storm system. Hurricanes, in particular, are fueled by large quantities of water vapor, and their strength is often linked to the amount of moisture in the atmosphere (Emanuel, 1991). Water vapor also plays a critical role in precipitation formation. As water vapor condenses into cloud droplets, these droplets coalesce and grow in size, eventually becoming heavy enough to fall

to the Earth as precipitation. This process is key in determining rainfall patterns and the distribution of freshwater resources across the planet (Wallace & Hobbs, 2006).

1.5.4 Water Vapor and Atmospheric Stability

The amount of water vapor in the atmosphere is also a critical factor in determining atmospheric stability (Galewsky *et al.* 2010; Sherwood *et al.* 2010). A moist atmosphere is more likely to become unstable, leading to convection and the development of thunderstorms. Conversely, a dry atmosphere tends to be more stable, with fewer chances for cloud development and precipitation. The measure of atmospheric moisture, known as precipitable water vapor (PWV), is often used by meteorologists to assess the potential for severe weather events (Bevis *et al.* 1992). Convective Available Potential Energy (CAPE), a common metric for measuring atmospheric instability, is highly dependent on the moisture content of the atmosphere. High values of CAPE typically indicate a greater likelihood of strong convection and severe weather, as they signify the presence of significant amounts of water vapor that can fuel storm development (Markowski & Richardson, 2011).

1.5.5 Water Vapor as a Climate Feedback Mechanism

Water vapor acts as both a short-term weather factor and a long-term climate feedback mechanism. As the atmosphere warms due to increased concentrations of CO_2 and other greenhouse gases, its ability to hold water vapor increases—a process governed by the Clausius-Clapeyron equation. This means that for every degree Celsius increase in atmospheric temperature, the air can hold about 7% more water vapor (Held & Soden, 2000).

This increased water vapor enhances the greenhouse effect, leading to further warming—a process known as water vapor feedback. This feedback is one of the most significant contributors to the sensitivity of the Earth's climate system to changes in greenhouse gas concentrations. Studies show that water vapor feedback amplifies the warming effect of CO_2 by about two to three times (Dessler, 2010).

1.6 Existing Techniques to Measure Water Vapor

Due to its importance, water vapor is continuously monitored using a variety of methods. Radiosondes, satellite-based sensors, and ground-based GNSS (Global Navigation Satellite

System) networks are all employed to measure the distribution of water vapor in the atmosphere. GNSS, in particular, has proven to be an effective tool for real-time monitoring of precipitable water vapor, providing valuable data for both weather forecasting and climate studies (Champollin *et al.* 2005; Bock *et al.* 2005). Satellite missions like the Aqua satellite, part of NASA's Earth Observing System, have provided critical data on water vapor content and distribution, improving our understanding of its role in climate and weather systems (Schneider *et al.* 2010). Each method has its own strengths and limitations in terms of spatial and temporal resolution, accuracy, and coverage

1.6.1 Satellite-Based Remote Sensing

Satellites provide global coverage and are capable of measuring water vapor in both the upper and lower atmosphere. They are typically divided into two types based on their orbits: Geostationary (GEO) and Low Earth Orbit (LEO) satellites.

GEO satellites orbit the Earth at the same rotational speed as the planet, allowing them to remain fixed over a particular region. Instruments onboard, such as infrared sensors, monitor water vapor continuously, providing valuable data for weather forecasting. GEO satellites are particularly effective for tracking large-scale atmospheric phenomena like tropical cyclones and moisture patterns. However, their coverage is limited to specific regions, and their resolution decreases with increasing distance from the equator (Schmit *et al.* 2017).

LEO satellites orbit at lower altitudes (approximately 200–800 km) and provide more detailed measurements, with higher spatial resolution than GEO satellites. Instruments such as microwave sensors can penetrate clouds and provide accurate water vapor measurements even in regions covered by dense clouds or precipitation. LEO satellites provide near-global coverage as they pass over different areas of the Earth during each orbit (Wentz *et al.* 2007). However, their temporal resolution is limited since they only pass over the same location at specific intervals.

1.6.2 In Situ Measurements

In-situ measurements refer to direct observations of atmospheric parameters, typically through the use of radiosondes and dropsondes. Radiosonde are balloon-borne instruments measure temperature, pressure, humidity, and wind as they ascend through the atmosphere. Radiosondes

provide highly accurate vertical profiles of water vapor but are limited in spatial coverage, as they are typically launched from fixed ground stations twice daily. This limited temporal resolution poses challenges for continuous monitoring (Seidel *et al.* 2009). Similar to radiosondes, dropsondes are released from aircraft and measure the atmospheric parameters as they descend. They are particularly useful during severe weather events like hurricanes, where vertical profiles of water vapor and other atmospheric parameters are needed for storm prediction (Franklin *et al.* 2003). However, the deployment of dropsondes is costly and cannot be done routinely, making it a tool primarily for research or extreme weather events.

1.6.3 Ground-Based Remote sensing Observation

Ground-based systems provide continuous monitoring of water vapor with high temporal resolution but are limited in spatial coverage compared to satellite observations. Key ground-based methods include WVR, LIDAR, FTIR, and GNSS meteorology.

WVR measures microwave radiation emitted by water vapor in the atmosphere to estimate the total column of water vapor. It is widely used in radio astronomy and satellite communication but is susceptible to interference from clouds and precipitation. Whereas LIDAR Uses laser pulses to measure atmospheric properties, including water vapor. LIDAR provides high-resolution vertical profiles of water vapor but is limited to clear-sky conditions as clouds and precipitation can attenuate the laser signals (Eichinger *et al.* 1999). FTIR instruments measure the absorption of infrared radiation by atmospheric gases, including water vapor. This technique provides accurate column measurements of water vapor but requires clear skies and is usually limited to research settings (Schneider *et al.* 2010).

1.6.4 GNSS Meteorology

This technique has gained significant attention in recent years due to its high accuracy, continuous operation, and global availability. Global Navigation Satellite System (GNSS) meteorology is a relatively new and highly effective method for monitoring water vapor. GNSS receivers, which are widely used for positioning and navigation, can also be employed to measure water vapor in the atmosphere. This technique exploits the fact that water vapor in the atmosphere causes a delay in the propagation of GNSS signals as they travel from satellites to ground-based receivers. By

analyzing these delays, scientists can estimate the amount of precipitable water vapor (PWV)—the total amount of water vapor in a column of atmosphere above the receiver (Jade *et al.* 2005.

The concept of using GNSS for meteorological purposes was first proposed by Bevis *et al.* (1992), and since then, the method has been widely adopted for weather monitoring and forecasting. GNSS meteorology is particularly valuable because it provides high temporal resolution, continuous measurements, and operates under all weather conditions, unlike optical or infrared methods.

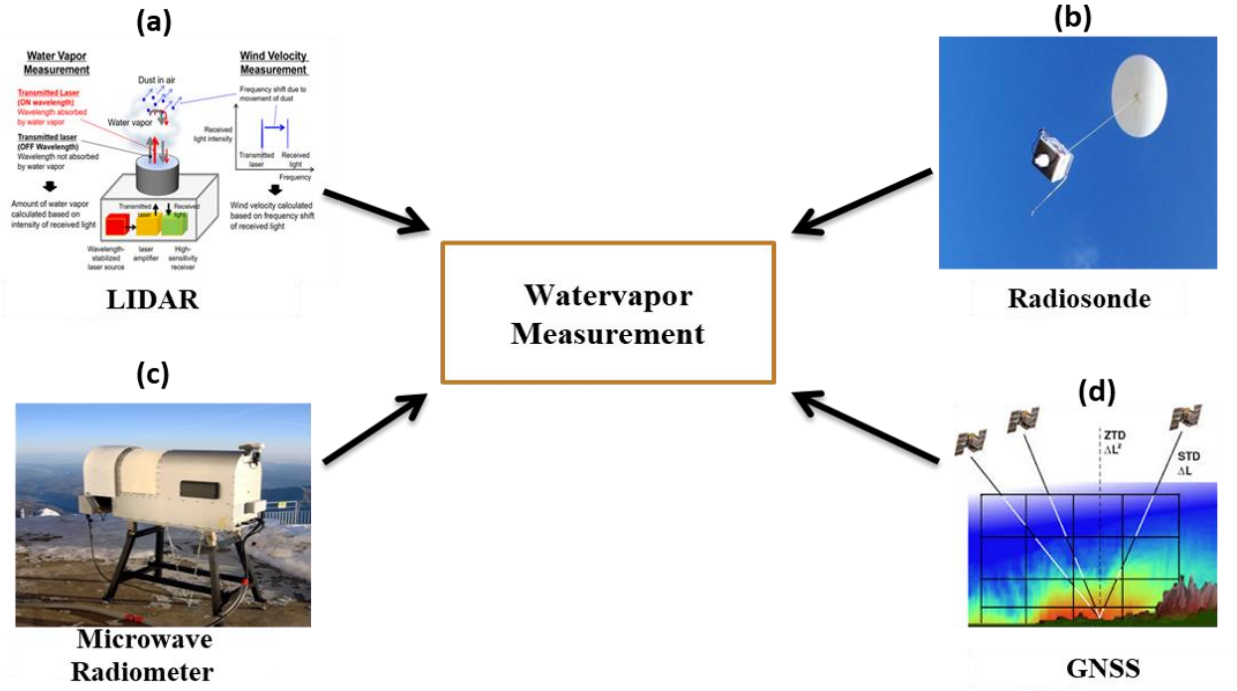

Figure 1.4: *Pictorial representation of different techniques of measure atmospheric water vapor including (a) LIDAR, (b) Radiosonde, (c) Microwave Radiometer, (d) GNSS.*

1.7 How GNSS Measures Water Vapor

As GNSS signals pass through the atmosphere, they are delayed due to the refractivity of the atmosphere, which depends on pressure, temperature, and the amount of water vapor. This delay is known as the zenith total delay (ZTD), which consists of two components:

ZHD Caused by the dry components of the atmosphere (primarily nitrogen and oxygen). ZWD Caused by water vapor in the atmosphere. Once the ZHD is determined (based on surface pressure observations), the ZWD can be estimated. The ZWD is directly related to the amount of PWV, allowing meteorologists to derive real-time water vapor measurements (Bevis *et al.* 1992).

1.8 Why GNSS for Water Vapor Measurement?

Global Navigation Satellite System (GNSS) meteorology has emerged as a powerful tool for measuring atmospheric water vapor due to its many distinct advantages over traditional techniques. From its continuous monitoring capability to high temporal resolution, GNSS offers a unique and valuable solution for both weather forecasting and climate research. Below, we explore in detail why GNSS has become a favored method for water vapor measurement, supported by real-world examples and references.

1.8.1 Continuous Monitoring and High Temporal Resolution

GNSS receivers provide continuous, near-real-time measurements of atmospheric water vapor, which is essential for weather forecasting and storm prediction. Unlike radiosondes or satellite-based sensors that take measurements at fixed intervals (e.g., twice daily for radiosondes or once every few hours for some satellites), GNSS data is collected continuously, providing updates at intervals as short as every 5 to 30 minutes. This high temporal resolution enables meteorologists to detect rapid changes in water vapor content, making GNSS a crucial tool for tracking fast-evolving weather phenomena such as thunderstorms, flash floods, and tropical cyclones (Bevis *et al.* 1992; Van Baelen *et al.* 2011).

In contrast, satellite-based sensors, while excellent for global coverage, often have limited revisit times, especially in the case of Low Earth Orbit (LEO) satellites. Geostationary satellites provide more frequent data but are limited by their spatial resolution and coverage of high-latitude regions. The continuous nature of GNSS measurements makes it superior for monitoring localized or short-term weather events that evolve rapidly over time.

1.8.2 All-Weather Capability

A major advantage of GNSS is its ability to operate in all weather conditions. Unlike optical and infrared sensors, which can be significantly affected by cloud cover, rain, or snow, GNSS signals are relatively unaffected by these adverse weather conditions. The propagation of GNSS signals is primarily impacted by atmospheric water vapor, which is precisely what this technique measures. This makes GNSS particularly useful for monitoring storm systems, where other measurement methods might fail due to poor visibility or interference from clouds (Shoji *et al.* 2009; Bock *et al.*

2005; Vaquero-Martinez *et al.* 2022). In extreme weather situations, such as during the onset of a tropical cyclone, GNSS receivers can continue to provide accurate water vapor measurements when traditional satellite and optical instruments may be limited. This capability is vital for maintaining continuous data flow during critical periods, enhancing the ability to issue timely weather warnings.

1.8.3 Global Coverage and Dense Networks

One of the unique advantages of GNSS is its potential for global coverage. GNSS networks, such as GPS, GLONASS, Galileo, and BeiDou, already span the globe for navigation purposes, and many of these stations are now being used for meteorological applications. The global infrastructure of GNSS stations, combined with the increasing deployment of new stations, makes it possible to monitor water vapor levels in almost any region on Earth. This includes remote and otherwise under-observed regions, such as the oceans, deserts, and polar areas, where traditional meteorological networks are sparse (Gutman *et al.* 2004).

For instance, Japan's GEONET system consists of over 1,200 GNSS stations that provide continuous data on water vapor. This dense network has proven invaluable for tracking water vapor variations associated with convective storms, improving local weather forecasts, and advancing storm prediction capabilities (Shoji *et al.* 2009).

The development of multi-GNSS systems (combining data from GPS, Galileo, GLONASS, and BeiDou) increases both the accuracy and coverage of water vapor measurements. As more GNSS satellites are launched and more ground stations are established, the spatial and temporal resolution of water vapor data will continue to improve, providing even better global coverage for weather monitoring and forecasting.

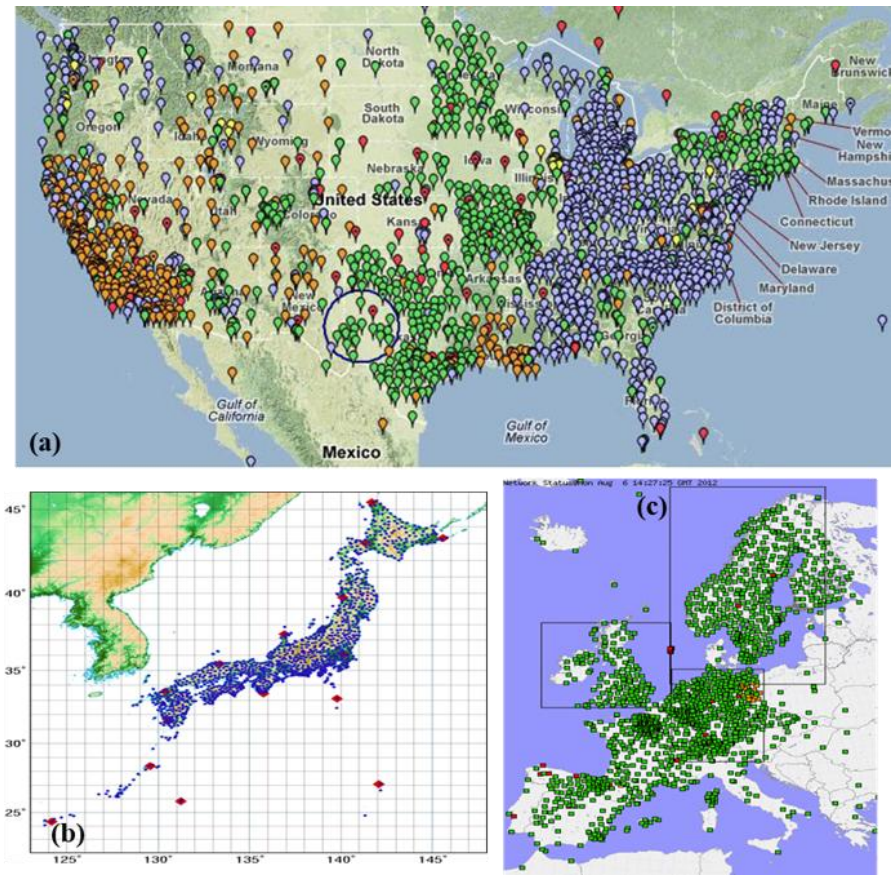

Figure 1.5: *A dense GNSS network in (a) USA (b) Japan (c) Europe.*

1.8.4 No Mechanical Movement Required

Unlike some meteorological instruments, GNSS receivers do not require any mechanical parts or moving components, which reduces the risk of mechanical failure and the need for regular maintenance. This increases the reliability of GNSS as a long-term monitoring tool. Mechanical failures can be a significant issue for instruments like radiosondes, where the entire system is subject to environmental wear and tear, particularly in extreme weather conditions (He *et al.* 2021). The static nature of GNSS receivers also makes them cost-effective and easy to maintain compared to other ground-based water vapor measurement systems, such as weather radars or LIDAR systems, which require more complex infrastructure and regular upkeep.

1.8.5 Highly Accurate Water Vapor Measurement

The GNSS-based approach to water vapor measurement is built upon the accurate determination of the Zenith Total Delay (ZTD) experienced by GNSS signals as they pass through the

atmosphere. This delay is caused by both the dry (hydrostatic) and wet (water vapor) components of the atmosphere. By separating the Zenith Wet Delay (ZWD) from the ZTD, meteorologists can derive the Precipitable Water Vapor (PWV)—the total column of water vapor in the atmosphere above the GNSS station (Bevis *et al.* 1992).

Studies have shown that GNSS-based measurements of PWV are highly accurate and compare favorably with traditional methods such as radiosondes, satellite-based measurements, and water vapor radiometers (WVRs). Radiosondes measure PWV from 0 to over 70 mm with an accuracy of approximately 2–5 mm RMSE, though performance varies in dry or cold conditions (Bevis *et al.* 1992). Satellite-based systems, like MODIS or COSMIC, cover 0–50+ mm with accuracies of 4–10 mm RMSE for MODIS (King *et al.* 2003) and 0.2–7 mm for COSMIC (Kursinski *et al.* 2008). WVRs measure PWV from ~2.5–50+ mm with 1–3 mm RMSE accuracy, but are sensitive to precipitation (Buehler *et al.* 2012). In fact, GNSS measurements have been shown to have an uncertainty of less than 2 mm of PWV, which is sufficiently accurate for weather forecasting and climate studies (Bock *et al.* 2007). This high accuracy, combined with the continuous nature of GNSS data, makes it an indispensable tool for storm prediction, particularly in rapidly changing weather conditions.

1.8.6 Cost-Effective

Using GNSS for meteorology is highly cost-effective because the GNSS infrastructure is already in place for navigation and positioning. Repurposing GNSS stations for atmospheric water vapor measurement requires minimal additional investment, primarily in the form of software and data processing capabilities. This makes GNSS a financially viable option for enhancing weather prediction systems, particularly in developing regions where the cost of deploying specialized meteorological equipment might be prohibitive (Gutman *et al.* 2004).

In contrast, deploying radiosondes, LIDAR, or other specialized instruments involves significant costs for hardware, deployment, and maintenance, making GNSS a more attractive option for large-scale meteorological networks.

1.8.7 Integration with Multi-GNSS and Other Systems

The use of multi-GNSS systems significantly improves the spatial and temporal resolution of water vapor measurements. By combining data from multiple satellite constellations (such as GPS, Galileo, GLONASS, and BeiDou), meteorologists can achieve better coverage and redundancy in their measurements (Casallas-García *et al.* 2023). This is particularly useful in regions where individual GNSS signals may be blocked or weakened by buildings, mountains, or other obstacles (Shoji *et al.* 2009).

1.9. Literature Review

Understanding the distribution and variability of atmospheric water vapor is crucial due to its significant role in weather phenomena, climate dynamics, and the Earth's energy balance. Integrated Water Vapor (IWV), or precipitable water, represents the total amount of water vapor present in a vertical column of the atmosphere. Accurate measurement and analysis of IWV are essential for weather forecasting, climate modeling, and studying atmospheric processes such as radiation balance, the hydrological cycle, energy transport, and global warming (Held & Soden, 2000; Trenberth *et al.* 2005). Traditional methods of measuring atmospheric water vapor, such as radiosondes and microwave radiometers, have limitations in spatial and temporal resolution. The advent of Global Navigation Satellite Systems (GNSS) has revolutionized atmospheric sensing by providing high-resolution, continuous IWV data, opening new avenues for studying atmospheric processes, including the diurnal variations of IWV.

The utilization of GNSS for atmospheric sensing began with the groundbreaking work of Bevis *et al.* (1992), who demonstrated that delays in GNSS signals caused by atmospheric water vapor could be quantified to estimate IWV accurately. This method involved measuring the Zenith Wet Delay (ZWD) experienced by GNSS signals as they pass through the atmosphere, which is directly related to the amount of water vapor along the signal path. Bevis *et al.*'s study laid the foundation for GNSS meteorology, providing a cost-effective and reliable alternative to traditional atmospheric water vapor measurement techniques. The high temporal resolution of GNSS observations allowed for continuous monitoring of IWV, capturing rapid changes in atmospheric moisture content that were previously undetectable.

Following this pioneering work, Rocken *et al.* (1993) and Duan *et al.* (1996) further validated the use of GNSS for IWV retrieval. They demonstrated that GNSS-derived IWV measurements were in good agreement with those obtained from radiosondes and microwave radiometers. The ability of GNSS to provide all-weather observations made it an invaluable tool for atmospheric studies, particularly in regions with sparse observational networks. Solheim *et al.* (1997) developed a multicomponent integrated water vapor sensing system that combined GNSS data with surface meteorological measurements, improving the accuracy of IWV estimations and enhancing the capability to monitor atmospheric water vapor in real-time.

Throughout the late 1990s and early 2000s, significant advancements were made in refining GNSS-based IWV retrieval techniques. Rocken *et al.* (1993) highlighted the potential of GNSS in capturing diurnal variations of IWV, emphasizing its utility in studying atmospheric processes occurring over short timescales. By providing continuous data, GNSS enabled the detection of subtle changes in atmospheric moisture content associated with diurnal heating and cooling cycles. Liou *et al.* (2001) compared GNSS-derived IWV with microwave radiometer data, confirming the reliability of GNSS measurements. Wang and Zhang (2008) conducted comprehensive evaluations of GNSS IWV measurements over various climatic regions, demonstrating high correlations with radiosonde data. These studies established GNSS as a robust tool for atmospheric water vapor sensing.

The accurate retrieval of IWV using GNSS facilitated a wide range of atmospheric applications. Flores *et al.* (2000) and Bender *et al.* (2011) utilized GNSS data to develop four-dimensional tomographic models of atmospheric water vapor (Champollion *et al.* 2005, Champollion *et al.* 2009, Chen *et al.* 2017). This approach provided detailed insights into the spatial and temporal distribution of moisture in the atmosphere, enhancing the understanding of weather systems and climate processes. In monsoonal studies, Radhakrishna *et al.* (2019) employed GNSS-derived IWV to investigate the spatial coherence of water vapor and its relationship with rainfall over the Indian subcontinent, contributing to improved modeling of monsoon dynamics and precipitation patterns. Singh *et al.* (2019) integrated GNSS signal delay data into numerical weather prediction models, enhancing short-range weather forecasts over India by incorporating high-resolution IWV data to better represent atmospheric moisture fields.

1.9.1. Diurnal variation of IWV

Continuous IWV probing capability by GNSS systems allowed several researchers to understand the diurnal variation of IWV and underlying mechanisms causing such variation. Dai *et al.* (2002) analyzed IWV over North America using data from 54 stations over four years, finding significant diurnal variations that varied spatially, with higher amplitudes in the western United States and subtropics. The phase of the diurnal cycle peaked around noon in winter and shifted from mid-afternoon to midnight in summer. Their findings emphasized the importance of accounting for diurnal IWV variations in weather models to improve humidity measurements and forecasting accuracy. In Europe, Champollion *et al.* 2005) observed clear diurnal patterns influenced by atmospheric tides and regional meteorological conditions, attributing the variations to the superposition of semidiurnal and diurnal atmospheric tides and local convective processes. Similarly, Bock *et al.* (2007) studied the diurnal cycle over Africa, revealing strong variations associated with atmospheric waves and convection, particularly over the Sahel region during the monsoon season. Li *et al.* (2015) analyzed diurnal IWV variations over the Tibetan Plateau, finding significant diurnal cycles with peaks in the late afternoon influenced by thermal convection and large-scale circulation. The elevated terrain enhances thermal contrasts, leading to pronounced diurnal variations.

Subedi *et al.* (2023) examined the diurnal and seasonal variation of IWV over Nepal, observing late afternoon peaks and early morning minima influenced by monsoonal activity and local topography. In Russia's Volga-Ural region, Kalinnikov and Khutorova (2017) found that the diurnal harmonic phase occurred between 14:00 and 17:00 local time, strongly influenced by surface air temperature in summer. Ortiz de Galisteo *et al.* (2011) studied diurnal cycles in Spain, finding that nighttime IWV variations were more similar across stations compared to afternoon variations, suggesting significant local effects during daytime. Wu *et al.* (2003) investigated the diurnal variation of IWV over a mountainous area on Sumatra Island, observing a distinct pattern with IWV increasing during the day and peaking in the late afternoon, attributed to thermally induced local circulations and convective processes. Torri *et al.* (2019) extended this research by analyzing the impact of the Madden-Julian Oscillation (MJO) on diurnal cycles, finding significant effects on amplitude and peak IWV times. Li *et al.* (2008) examined diurnal variations in central Japan, finding that moisture advection and sea breeze circulations were key factors influencing the

variations. Meza *et al.* (2020) analyzed diurnal IWV variations over Central and South America, finding stronger variations over land due to local topography and land-sea interactions. Radhakrishna *et al.* (2015) studied the diurnal variation over the continental United States, highlighting the influence of topography and large-scale circulation patterns.

Despite global interest, studies focusing on the diurnal variation of IWV in India, particularly in semi-arid regions, are limited. Emmanuel *et al.* (2018) analyzed the diurnal variation over Gadanki and Trivandrum using radiosonde observations, noting limitations due to poor temporal resolution. Renju *et al.* (2015) investigated seasonal and diurnal variations using GPS data, observing significant diurnal cycles influenced by monsoon activity and surface heating. Jadala *et al.* 2020) focused on Hyderabad, finding strong diurnal harmonics and a positive correlation between IWV and surface temperature. Kannemadugu *et al.* (2022) investigated diurnal variations across 18 stations in India, noting that the variations were not uniform and depended on seasons and locations. For semi-arid locations like Hyderabad and Bangalore, noticeable phase lags between seasons were observed, emphasizing the influence of local topography and monsoonal winds. Researchers studied diurnal and seasonal variations over the Indian subcontinent, finding more pronounced diurnal variations during the monsoon season due to increased moisture influx and convective activities (Ratna *et al.* 2016; Diri *et al.* 2016). Barman *et al.* (2017) analyzed diurnal variations over Northeast India, finding distinct variations for each site due to complex terrain and varying monsoonal influences.

The diurnal variation of IWV is influenced by several factors, including surface heating, local topography, land-sea interactions, moisture advection, atmospheric dynamics, convective processes, surface temperature, atmospheric tides, and monsoonal influences. Surface heating leads to increased evaporation and atmospheric moisture during the day (Meza *et al.* 2020). Local topography affects atmospheric circulation patterns, influencing moisture distribution (Wu *et al.* 2003; Kannemadugu *et al.* 2022). Land-sea interactions introduce breeze circulations impacting IWV (Li *et al.* 2008). Moisture advection and atmospheric dynamics like the MJO (Torri *et al.* 2019) also contribute to variations. Convective processes redistribute moisture vertically (Dai *et al.* 2002), and positive correlations between IWV and surface temperature indicate temperature's role (Jadala *et al.* 2020). Understanding these factors is essential for accurate modeling and prediction of IWV variations, particularly for weather forecasting and climate studies.

1.9.2 Nowcasting of storms:

Recognizing water vapor's critical role in convective storm development, researchers explored the use of GNSS-derived IWV data for storm nowcasting. Sapucci *et al.* (2019) observed that sudden increases, or "jumps," in GNSS-derived IWV occurred before intense rain events in Brazil. They proposed using specific IWV thresholds as indicators of imminent heavy rainfall, offering potential improvements in early warning systems. Benevides *et al.* (2015) incorporated GNSS-IWV trends into rainfall nowcasting models, demonstrating enhanced forecast accuracy by utilizing real-time IWV observations to capture rapid atmospheric changes preceding precipitation events. Yao *et al.* (2017) developed a short-term rainfall forecasting method based on GNSS-derived IWV, achieving improved detection rates by analyzing parameters such as the magnitude, variation, and rate of change of IWV. In tropical regions, Manandhar *et al.* (2018) found that rapid increases in IWV were associated with imminent rainfall events, highlighting the potential for real-time applications in storm prediction.

While GNSS-derived IWV data showed promise in storm nowcasting, challenges remained. High false alarm rates were observed when using IWV parameters alone (Yao *et al.* 2017; Manandhar *et al.* 2018), necessitating the integration of additional meteorological predictors to improve reliability. To address these challenges, hybrid models combining GNSS IWV data with other atmospheric parameters were developed. Guerova *et al.* (2022) created a "convective storm demonstrator" that integrated IWV measurements with instability indices derived from atmospheric profiles, achieving an 83% predictability and reduced false alarms. This model demonstrated the benefits of a multi-parameter approach. Benevides *et al.* (2019) employed neural networks to merge satellite-derived cloud top temperatures, IWV, and meteorological sensor data, enhancing rainfall forecasting accuracy. Zhao *et al.* (2020) incorporated GNSS-derived precipitable water vapor and meteorological data using support vector machines, improving detection rates and reducing false alarms compared to traditional threshold-based methods.

1.9.3 Usage of AI/ML techniques for nowcasting

Advancements in artificial intelligence (AI) and machine learning (ML) have further opened new avenues for analyzing complex atmospheric data and improving storm prediction. ML models can handle large datasets with multiple variables, identifying patterns that may not be apparent through traditional statistical methods (Wu *et al.* 2023). Suparta and Alhasa (2016) applied adaptive neuro-

fuzzy inference systems to model precipitable water vapor in the absence of a GPS network, demonstrating the potential of ML in regions with limited observational infrastructure. Łoś *et al.* (2020) developed a GNSS-based machine learning storm nowcasting system that integrated GNSS IWV data with meteorological parameters to predict convective storms. The system demonstrated the potential of AI in meteorological applications, reducing false alarms and improving forecast reliability. Mostajabi *et al.* (2019) used ML techniques to nowcast lightning occurrence from commonly available meteorological parameters, enhancing prediction capabilities for severe weather warnings. Feng *et al.* (2021) applied deep learning methods for precipitation nowcasting, showing significant improvements over numerical weather prediction models by capturing spatial-temporal patterns in radar and satellite data.

The integration of GNSS-derived IWV data with machine learning techniques holds significant potential for advancing storm prediction and atmospheric research. However, challenges remain, including data quality and availability, model generalization across different climatic regions, and the computational resources required for advanced ML models (Chen *et al.* 2020; Feng *et al.* 2021). Addressing these challenges is essential for leveraging the full potential of GNSS and ML technologies in atmospheric sciences.

1.10 Research Gaps

Despite significant advancements, several research gaps remain:

- **a.** Most studies focus on real-time monitoring or nowcasting with minimal lead time. There is a lack of research on utilising GNSS-derived IWV to predict storms before they occur, providing sufficient lead time for early warning systems and disaster preparedness.
- **b.** While AI/ML techniques have been applied using traditional meteorological data, their integration with GNSS-derived IWV data remains underexplored. Utilizing AI/ML models to analyze IWV variations could enhance the accuracy and lead time of storm predictions.
- **c.** Traditional nowcasting techniques predominantly rely on instability indices estimated from radiosonde data, which have limited temporal resolution. Incorporating high-resolution IWV data could improve nowcasting models.

d. Radar-based nowcasting can result in high false alarm rates due to the temporal evolution of storms. Satellite observations may lead to false positives as not all growing clouds develop into storms.

e. There is a paucity of research focused on predicting storms before they occur, especially with lead times sufficient for proactive measures and public safety interventions.

f. Integrating multiple data sources—such as GNSS-derived IWV, radar, satellite imagery, and surface observations—could enhance storm prediction accuracy. Developing models that effectively combine these datasets, particularly using AI/ML techniques, is a key research area.

g. Most research has been conducted in regions with dense observational networks. Studies focused on regions like India and other semi-arid areas are needed to improve understanding and prediction of storm events in diverse climatic conditions.

1.11 Objectives of the Thesis

Based on the research gaps identified—specifically, the limited utilization of GNSS-derived Integrated Water Vapor (IWV) for early storm prediction, the underexplored integration of Artificial Intelligence (AI) and Machine Learning (ML) techniques with IWV data, and the reliance on traditional nowcasting methods with inherent limitations—this thesis aims to address these challenges through the following objectives

1. Understanding the Diurnal Cycle of IWV
2. Prediction of IWV Using AI/ML Techniques
3. Nowcasting of Storms Using Predicted IWV with ML Techniques
4. Evaluation of ML-Based Storm Nowcasting Techniques

1.12 Thesis Overview

Building upon the identified research gaps and objectives, this thesis is structured to systematically explore and address the challenges in predicting storms in advance using GNSS-derived Integrated Water Vapor (IWV) data and machine learning techniques. The following is a summary of each chapter, incorporating the specific focus and methodologies as per your instructions.

Chapter 2 provides a comprehensive description of the data sources and methodologies employed throughout the research. It details the collection and processing of GNSS-derived IWV data from

a network of seven receivers in Gadanki, India, over a three-year period (September 2018 to August 2021). The chapter outlines the acquisition of satellite brightness temperature data and other relevant meteorological observations necessary for the study. It discusses the preprocessing steps, including quality control measures and handling of missing or outlier data, ensuring the reliability of the datasets used. The methodologies for harmonic analysis of IWV, development of machine learning models and threshold-based hybrid modeling are introduced. This chapter sets the foundation for the analyses conducted in subsequent chapters, providing the necessary background on the tools and techniques employed.

In chapter 3, the focus is on investigating the factors responsible for the diurnal variability of IWV in the Gadanki. Utilizing the GNSS-derived IWV data collected over three years, a detailed harmonic analysis is performed to determine the significant diurnal and semi-diurnal components. The study finds that the diurnal (24-hour) harmonic is significant 93% of the time, while the semi-diurnal (12-hour) harmonic is significant 36% of the time. Seasonal and monthly variations in the amplitudes and phases of these harmonics are examined, revealing that diurnal amplitudes of 2–3 mm and semi-diurnal amplitudes of 1.5–2 mm are most common across all seasons. The analysis identifies two distinct semi-diurnal cycles with different peak times. Importantly, the research discovers that the vertically integrated moisture flux convergence (VIMFC), representing the transport term of IWV, is the primary factor determining the diurnal cycle of IWV at both seasonal and diurnal scales. This chapter provides critical insights into the atmospheric processes influencing IWV variations, which are essential for improving weather prediction models and understanding the moisture dynamics in semi-arid regions.

Building upon the understanding of IWV variability, chapter 4 focuses on developing a machine learning technique to predict IWV, aiming to enhance the nowcasting of storm events. Specifically, the Light Gradient Boosting Machine (LightGBM) algorithm is employed due to its efficiency and accuracy with large datasets and its ability to handle complex non-linear relationships. The model is trained using the IWV data from Gadanki, along with relevant meteorological parameters such as temperature, humidity, and wind speed. The predictive capability of the model is thoroughly evaluated, demonstrating its effectiveness in forecasting IWV values. By accurately predicting IWV, the model provides crucial information for anticipating atmospheric changes that precede storm development, thereby contributing to improved early warning systems.

In chapter 5, the predicted IWV values obtained from the LightGBM model are utilized in conjunction with satellite-derived brightness temperature data to develop a hybrid model for storm prediction. This two-step process involves integrating the nowcasted IWV with satellite observations through threshold analysis to predict storms at different time horizons. The hybrid model uses the strengths of both data sources: the temporal precision of IWV predictions and the satellite-based brightness temperature. Thresholds are determined based on historical storm events to identify critical values indicative of storm formation. The chapter details the methodology for combining these datasets, the development of the threshold-based predictive model, and the evaluation of its performance. Results indicate that the hybrid model enhances the accuracy and lead time of storm forecasts compared to traditional nowcasting methods, offering a practical approach for operational meteorology.

Expanding on the hybrid approach, chapter 6 explores the use of a standalone machine learning-based model to nowcast storms in advance. The model incorporates GNSS-derived IWV, satellite brightness temperature, and additional meteorological data to predict storm events without relying solely on threshold analysis. Advanced machine learning algorithms, such as ensemble methods and deep learning techniques, are applied to capture the complex interactions among these variables. The model is trained and validated using historical data, and its performance is compared against both the hybrid model from Chapter 5 and traditional nowcasting techniques. The standalone ML model demonstrates improved prediction accuracy and reduced false alarm rates, highlighting its potential for practical implementation in early warning systems. The chapter discusses the advantages of this approach, including its adaptability to different regions and conditions, and addresses potential limitations related to data availability and computational requirements.

The final chapter 7 summarizes the key findings and contributions of the thesis. It reflects on how the research objectives were achieved, emphasizing the advancements made in understanding the diurnal cycle of IWV, the development of machine learning techniques for IWV prediction, and the improvement of storm nowcasting methods using both hybrid and standalone models. The limitations of the study are acknowledged, such as the reliance on data from a single region and potential challenges in generalizing the models to other areas. Recommendations for future research are proposed, including exploring additional data sources like radar observations, refining

the models for different climatic regions, and investigating the integration of real-time data streams for operational use. The chapter concludes with final remarks on the impact of the research and its potential to enhance early warning systems and disaster preparedness in semi-arid regions and beyond.

Chapter 2

Instrumentation and data analysis

This study primarily uses a network of GNSS receiver measurements augmented by INSAT-3D and automatic weather station (AWS) measurements at Gadanki. In this chapter, a brief description on the above data sets and study region is provided. The GNSS-based retrieval method provides high temporal resolution and accuracy, enabling continuous monitoring of atmospheric water vapor. The data collection period, preprocessing steps, and quality control measures are explained to ensure the reliability of the datasets used in the subsequent analyses. Different analysis methods are also included in this chapter, like harmonic analysis for determining main modes of IWV variability, machine learning techniques for IWV prediction, and hybrid modeling approaches for storm nowcasting.

2.1 Datasets

2.1.1 NARL's Integrated GNSS Receiver and AWS Network

The National Atmospheric Research Laboratory (NARL) established an integrated network of GNSS receivers (and AWS) within an 6-8 km radius around Gadanki, India, in 2018, specifically to study atmospheric water vapor, which is crucial for understanding weather and climate dynamics (Bevis *et al.* 1992, Anthes *et al.* 2008, Anthes *et al.* 2000).

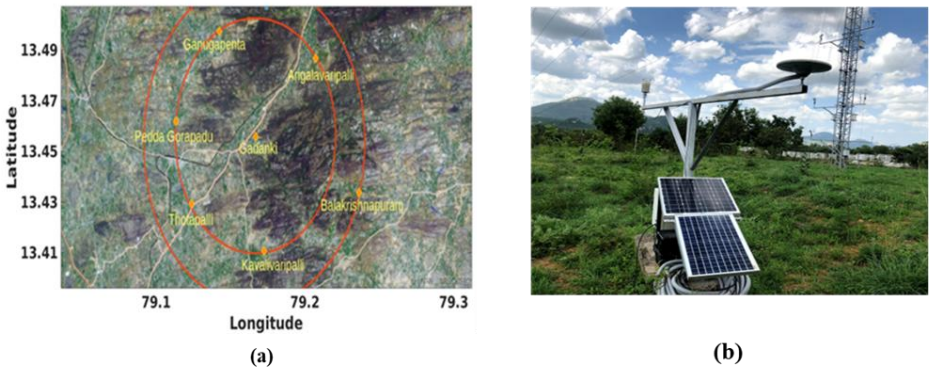

Figure 2.1: *(a) GNSS receiver network surrounding Gadanki, (b) A standalone GNSS receiver (resembling a flat plane), with a collocated AWS receiver AWS receiver on the left side.*

The seven stations are strategically located at Gadanki, Kavalavari Palli, Balakrishnapuram, Arigalavari Palli, Pedha Gorapadu, Thotapalli, and Ganugapenta, as shown in Figure 2.1(a), with detailed station information provided in Table 2.1.

Table 2.1: *Position Information of GNSS receiver stations of GNSS Network around Gadanki.*

Station Name	Latitude	Longitude	Altitude (in m)
Gadanki	13.460777	79.173229	246.239
Kavalavari Palli	13.410932	79.179656	300.026
Balakrishnapuram	13.435176	79.245098	290.638
Arigalavari Palli	13.493126	79.214549	199.744
Pedha Gorapadu	13.467129	79.11684	293.89
Thotapalli	13.429995	79.128733	270.947
Ganugapenta	13.504319	79.147499	345.99

This network plays a crucial role in providing high-resolution, continuous monitoring of atmospheric water vapor in the region. The study utilized GNSS data from the NARL network, with a temporal resolution of 15 min, 30 min, collected over a three-year period from September 2018 to August 2021. The raw GNSS data, recorded at 1-second intervals, were analyzed using the GAMIT software package to estimate Integrated Water Vapor (IWV), following established GNSS meteorology procedures (Ware *et al.* 1997). For processing, precise final ephemeris data from the International GNSS Service (IGS) and the Vienna Mapping Function (VMF) were used to estimate the Zenith Total Delay (ZTD), which was then converted to IWV (Boehm *et al.* 2006). The conversion required meteorological parameters, such as surface pressure and temperature, which were collected using Automatic Weather Stations (AWS) integrated with each GNSS receiver. The AWS data, available at a 30-second resolution, were averaged over 30-minute intervals before being used in GAMIT processing (Rocken *et al.* 1993).

The AWS used in this study is all-in-one type manufactured by Vaisala (WXT536), which provides basic meteorological parameters: temperature, humidity, wind speed, wind direction, pressure and rainfall.

2.1.2 Other Data Sources

Along with GNSS-derived Integrated Water Vapor (IWV) data, this research also utilizes additional data sources to provide a comprehensive understanding of the atmospheric variables that influence IWV. These data sources contribute to a holistic analysis of moisture dynamics, atmospheric processes, and storm prediction.

2.1.2.1 ERA5 for Vertically Integrated Moisture Flux (VIMF) and Divergence

To analyze the transport mechanisms and sources influencing atmospheric moisture, we utilized ERA5 reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA5 provides high-resolution global data, which includes vertically integrated moisture flux (VIMF) and its divergence (Hersbach *et al.* 2020). The VIMF represents the net movement of water vapor through the atmosphere, which is crucial for understanding changes in IWV. By integrating ERA5 data with GNSS-derived IWV estimates, we are able to determine how moisture advection and convergence influence the diurnal variations of IWV over Gadanki.

2.1.2.2 Fifty-Metre Flux Tower for Soil Moisture Retrieval

To study the influence of soil moisture on atmospheric water vapor, we used data from a 50-metre flux tower located at Gadanki. This tower provides crucial surface measurements, including soil moisture, temperature, latent and sensible heat fluxes, and surface winds (Sandeep *et al.* 2014). Soil moisture data are particularly important in understanding evapotranspiration and how it affects atmospheric water vapor content.

2.1.2.3 INSAT 3DR Satellite-Based Brightness Temperature Data

We also utilized satellite-based brightness temperature (T_b) data from the INSAT 3DR satellite, an advanced meteorological satellite that provides infrared and thermal data crucial for monitoring atmospheric conditions (Murugavel *et al.* 2019). T_b data serve as a proxy for cloud top temperatures, which are critical for identifying convective activities that may lead to storm development.

2.1.2.4 X-Band Dual-Polarization Radar Data

We also used X-band dual-polarization radar data, which is located at Gadanki, to identify and track storm events in real-time. The radar provides high temporal resolution observations and dual-polarization capabilities are useful for detailed analysis of precipitation, storm structure, and

hydrometeor classification (Ryzhkov *et al.* 2016). The dual-polarization feature helps distinguish different types of precipitation and track the evolution of storm systems.

2.2 Methodology

2.2.1 Retrieval of IWV from GNSS

GNSS consists of multiple satellites in medium Earth orbits, ground control stations, and user receivers that together enable precise positioning and timing measurements. GNSS signals are transmitted from satellites and received by ground-based GNSS receivers. As these signals travel through the atmosphere, they are subject to delays due to interactions with the medium in different atmospheric layers, primarily the ionosphere and troposphere. The GNSS signal delay caused by the troposphere is used in this study to estimate IWV, which is an important measure of the total atmospheric moisture in a vertical column (Bevis *et al.* 1992; Ware *et al.* 1997).

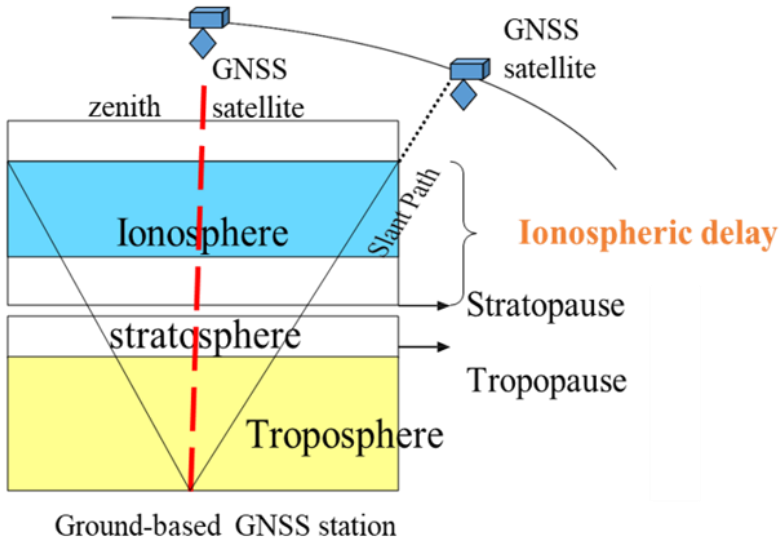

Figure 2.2: *Diagram illustrating the movement of GNSS signals through the atmosphere.*

2.2.1.1 GNSS Signal Delays and the Role of the Atmosphere

The use of GNSS to retrieve atmospheric water vapor involves taking advantage of the propagation delays that occur as the signal passes through the atmosphere. The primary atmospheric delays include ionospheric delay and tropospheric delay. The tropospheric delay, in turn, is composed of two components: the hydrostatic (or dry) delay and the wet delay. The wet delay is directly related to the atmospheric water vapor content, which is what we aim to estimate.

The total delay experienced by the GNSS signal as it travels from the satellite to the receiver can be expressed as:

$$Total\ Delay = Ionospheric\ Delay + Tropospheric\ Delay \quad (2.1)$$

The ionosphere is a region of the atmosphere containing free electrons, and it causes signal refraction and delays that are frequency-dependent. The ionospheric delay (Δt_{iono}) is inversely proportional to the square of the signal frequency (f), (as shown in the equation below), and hence it can be mitigated using dual-frequency GNSS observations.

$$\Delta t_{iono} = \frac{40.3 \times TEC}{f^2} \quad (2.2)$$

Where the constant 40.3 in the equation represents a scaling factor used to calculate the ionospheric delay. TEC is the Total Electron Content along the signal path, measured in electrons per square meter and f is the frequency of the GNSS signal, measured in Hz.

GNSS satellites transmit signals on two frequencies (L1 and L2), and by combining these dual-frequency signals, we can effectively eliminate the ionospheric delay. This approach significantly improves the precision of GNSS-derived atmospheric products, including ZTD (Kouba & Héroux, 2001).

The tropospheric delay, which is independent of signal frequency, contains valuable information about atmospheric moisture. It can be represented as:

$$Troposhperic\ Delay = Hydrostatic\ Delay + Wet\ Delay \quad (2.3)$$

(a) Hydrostatic Delay (ZHD): This component is mainly due to dry gases in the atmosphere and accounts for approximately 90% of the total tropospheric delay.

(b) Wet Delay (ZWD): This component arises from atmospheric water vapor and is the most variable. It is directly proportional to the amount of water vapor in the atmosphere and is used for IWV estimation.

2.2.1.2 Estimation of Zenith Total Delay (ZTD)

Once the ionospheric delay is eliminated, the next step is to estimate the Zenith Total Delay (ZTD) from the GNSS signal. The ZTD represents the delay experienced by the signal in the zenith direction and can be estimated using precise point positioning (PPP) or network processing techniques. For this study, the raw GNSS data recorded at 1-second intervals were processed using the GAMIT software package, which is widely used for GNSS data analysis and atmospheric parameter estimation (Herring *et al.* 2010). GAMIT uses precise satellite orbits and clock corrections, final ephemeris data from the International GNSS Service (IGS), and the Vienna Mapping Function (VMF) to estimate ZTD. ZHD can be accurately estimated using models such as the Saastamoinen model, which takes into account surface pressure (P_s) and latitude (ϕ):

$$ZHD = 0.0022768 \times \frac{P_s}{1 - 0.00266\cos(2\phi) - 0.00028\mathrm{H}} \tag{2.4}$$

Where H represents the height of the GNSS station above sea level in kilometers. Apart from this we have Hopfield model also to estimate ZHD.

$$ZHD = \frac{10^{-6}}{5} * [40136 + 148.72 * (T_s - 273.16)] * 77.64 * \frac{P_s}{T_s} \tag{2.5}$$

Where T_s is surface temperature, and P_s is surface Pressure (Bevis *et.al.* 1992).

2.2.1.3 Sensitivity Analysis of Hopfield and Saastamoinen Models for Estimating ZHD

To check the sensitivity of both the model we have done the sensitivity analysis. Figure 2.3 compares the Zenith Hydrostatic Delay (ZHD) estimated using the Hopfield and Saastamoinen models over Julian days 281 to 287. The Hopfield model (depicted in blue) and the Saastamoinen model (depicted in red) show close agreement overall, with both models capturing the variability in ZHD across the analyzed period. However, small deviations between the two models are observed, particularly during rapid changes in ZHD values. These differences may arise because of different assumptions for the two models.

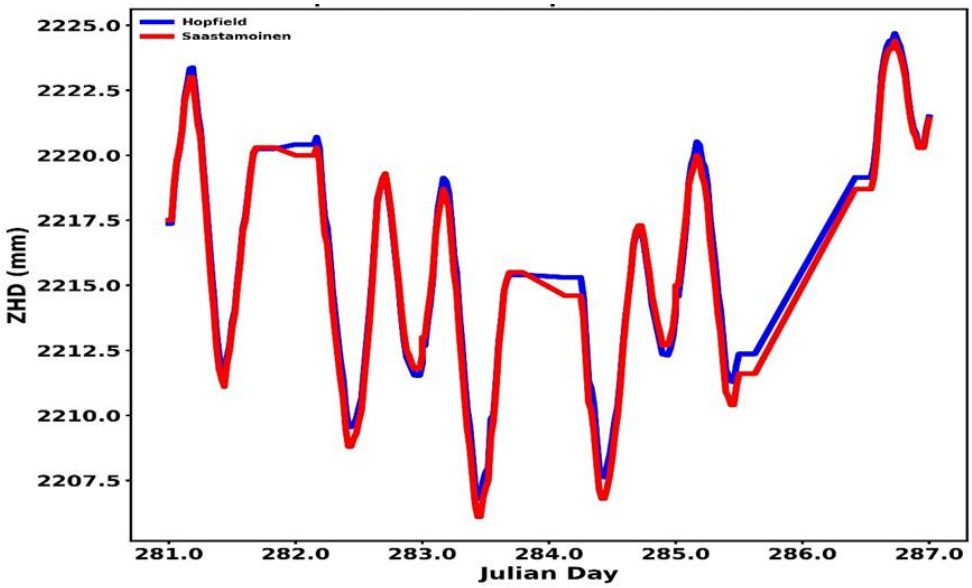

Figure 2.3: *A time series plot between Hopfield and Saastamoinen model for estimation of ZHD.*

2.2.1.4 Conversion of ZTD to IWV

To estimate IWV, the ZWD must be converted into a measure of the total atmospheric water vapor. This conversion is achieved using the following equation:

$$IWV = \Pi \cdot ZWD \tag{2.6}$$

Where Π is the conversion factor, which depends on the temperature, pressure, and specific properties of the atmosphere. It can be expressed as:

$$\Pi = \frac{\rho_w \cdot R_d \cdot (\frac{k_3}{T_m} + k_2')}{R_w . (10^6)} \tag{2.7}$$

Where ρ_w is the density of liquid water (typically 1000 kg/m³), R_d is the specific gas constant for dry air (287.05 J/(kgK), R_w is the specific gas constant for water vapor (461.5 J/(kgK)), k_3 and k_2' are empirically determined constants, T_m is the mean temperature of the atmosphere, which can be estimated using surface temperature (T_s) obtained from AWS data.

The surface temperature and pressure data needed for conversion are collected from Automatic Weather Stations (AWS) integrated with the GNSS receivers. These meteorological data are available at a 30-second resolution and are averaged over 30-minute intervals before being used in the GAMIT processing.

2.2.1.5 Sensitivity Analysis of IWV Retrieval Using Vienna and Neill Mapping Functions

To evaluate the sensitivity of Vienna and Neill mapping functions, we conducted a comparative analysis of their performance. The figure 2.4 shows a sensitivity analysis comparing two mapping functions—Vienna Mapping Function (VMF) and Neill Mapping Function (NMF) for IWV retrieval. The analysis spans Julian days 174 to 179, with IWV values measured in millimeters (mm). The IWV values derived using the VMF (shown as a red dashed line) closely align with those obtained using the NMF (depicted by a solid blue line). Both mapping functions exhibit similar temporal variability and capture the peaks and troughs of IWV with high consistency. However, slight deviations are observed at specific peaks, indicating subtle differences in sensitivity to the atmospheric conditions. The overall agreement suggests that both mapping functions are robust for IWV estimation, though the choice of mapping function may lead to minor variations in extreme IWV values.

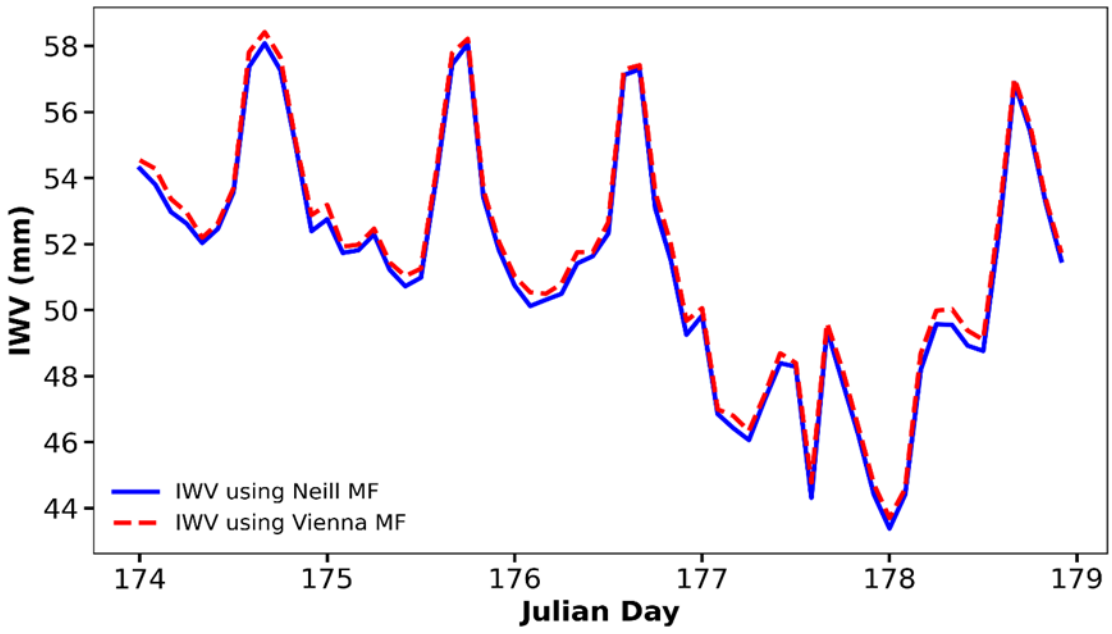

Figure 2.4: *Sensitivity Analysis of IWV Retrieval Using Vienna and Neill Mapping Functions.*

2.2.1.6 Sensitivity of IWV Retrieval to Temporal Resolution Using GAMIT

The sensitivity analysis of IWV retrieval to different temporal resolutions was conducted to understand the impact of varying time intervals on the accuracy and variability of IWV measurements. This analysis is crucial because the temporal resolution directly influences the

ability to capture rapid changes in atmospheric water vapor, which is essential for studying dynamic weather phenomena such as storms and heavy rainfall events. By evaluating IWV retrievals at different time horizons (15, 30, 60 and 120 minutes), we aimed to determine the optimal temporal resolution that balances capturing short-term variability and computational efficiency. Figure 2.5 shows IWV estimated with different averaging periods. At finer resolutions (15 and 30 minutes), the blue and orange lines exhibit higher variability and capture rapid fluctuations in IWV, particularly during dynamic weather conditions. In contrast, coarser temporal resolutions (60 and 120 minutes, represented by the green and red lines) appear smoother, as short-term variations are averaged out. This smoothing effect is especially evident in the later stages (around Julian days 278–280), where finer resolutions detect sharper declines in IWV compared to the coarser ones. This analysis underscores the importance of temporal resolution in IWV retrievals, particularly for capturing short-term variations and rapid changes in atmospheric water vapor.

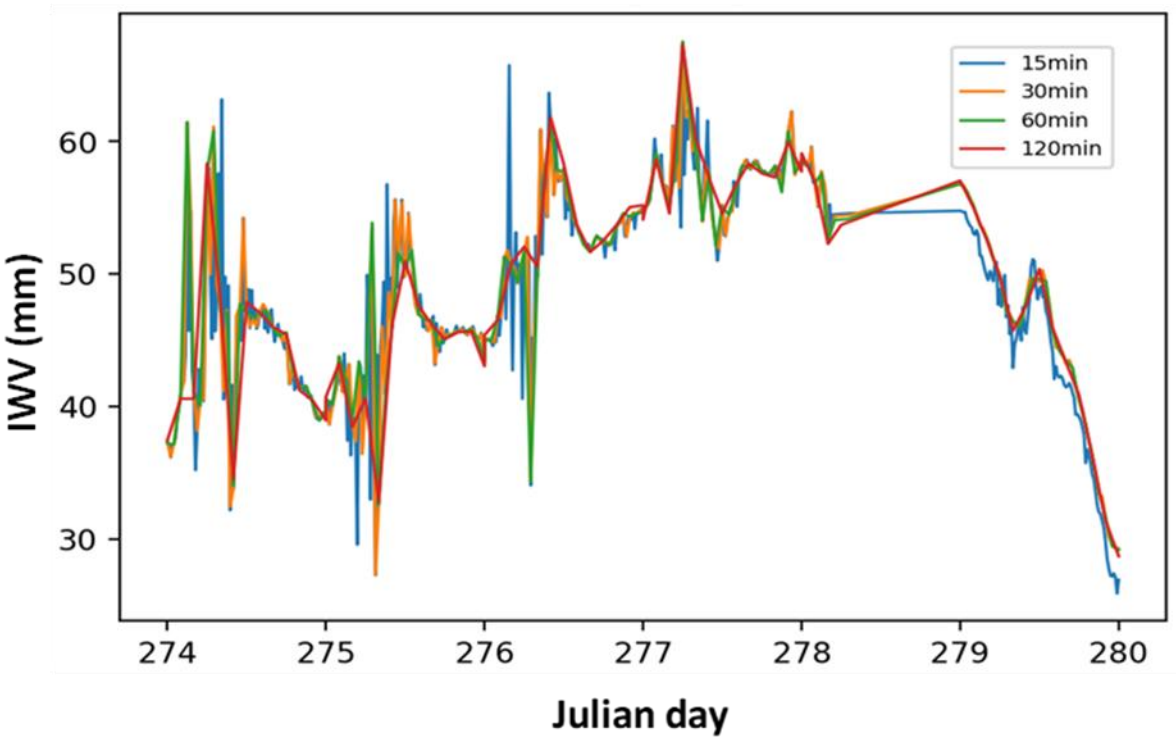

Figure 2.5: *Sensitivity analysis of IWV Retrieval to Temporal Resolution.*

2.2.1.7 Validation of GNSS-Derived IWV

The accuracy of GNSS-derived IWV is assessed by comparing it against independent radiosonde measurements made at Gadanki (Figure 2.6) . Radiosondes provide vertical profiles of atmospheric parameters, including temperature, pressure, and humidity, which can be used to estimate IWV. In this study, a high correlation coefficient (0.97) was observed between GNSS-derived IWV and Radiosonde-derived IWV, with a bias of 3.24 mm and a root mean square error (RMSE) of 2.86 mm. This high level of agreement demonstrates the reliability of GNSS- derived IWV for further analyses of atmospheric water vapor variability and storm prediction.

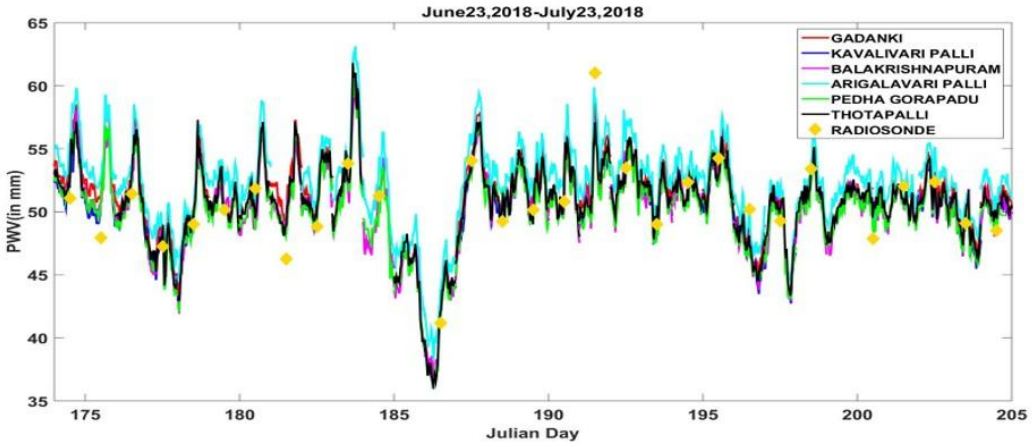

Figure 2.6: *Temporal variation of IWV retrieved from GNSS receiver measurements and their comparison with those obtained with radiosonde.*

2.2.2 Harmonic Analysis

To analyze the diurnal variability of Integrated Water Vapor (IWV), harmonic analysis is applied to capture the periodic components of IWV fluctuations over the study region. This analysis allows us to identify which harmonic components, such as the diurnal (24-hour) or semi-diurnal (12-hour) cycles, are dominant in the IWV data. The approach focuses on understanding the key patterns and factors that drive the diurnal changes in IWV.

2.2.2.1 Data Averaging and Diurnal Anomaly Calculation

Initially, the IWV data from all seven GNSS stations (located at Gadanki, Kavalavari Palli, Balakrishnapuram, Arigalavari Palli, Pedha Gorapadu, Thotapalli, and Ganugapenta) are averaged to represent the regional variation in IWV, rather than focusing on local variations at a single site.

Averaging the IWV from all stations helps in eliminating small-scale local effects and enhances the signal representing the broader regional atmospheric moisture.

Once the regional IWV values are obtained, diurnal anomalies are calculated by subtracting the daily mean IWV from each hourly IWV value for each day. This subtraction removes the daily average variability, allowing us to focus specifically on the diurnal patterns and deviations from the daily average. The resulting diurnal anomalies provide insights into the intraday changes in IWV, which can be crucial for understanding the processes driving moisture dynamics. The diurnal anomalies are further averaged for each of the four seasons observed in India (Attri & Tyagi ,2010) , defined as:

- **Winter**: January and February
- **Pre-Monsoon**: March, April, and May
- **Monsoon**: June, July, August, and September
- **Post-Monsoon**: October, November, and December

The seasonal averaging helps in highlighting how the diurnal cycle of IWV varies across different times of the year, providing insights into the influence of seasonal changes, such as monsoonal circulation and thermal convection, on atmospheric moisture.

2.2.2.2 Harmonic Analysis Method

To quantify the periodic components of the IWV diurnal variability, harmonic analysis is conducted on the diurnal and seasonal diurnal anomalies. Harmonic analysis allows us to represent the time series of IWV anomalies in terms of sinusoidal functions, capturing the most significant periodic variations. The goal is to determine the amplitude and phase of the dominant harmonics, specifically focusing on the diurnal (24-hour) and semi-diurnal (12-hour) components. The diurnal anomalies are modeled using the following equation:

$$IWV(t) = \sum_{n=1}^{2} \left(a_n \sin \frac{2n\pi t}{24} + b_n \cos \frac{2n\pi t}{24} \right) + \epsilon \tag{2.8}$$

Where:

- n=1,2: Represents the harmonic components with periods of 24 hours (diurnal) and 12 hours (semi-diurnal), respectively.
- a_n: Coefficient representing the sine component for the n-th harmonic.
- b_n: Coefficient representing the cosine component for the n-th harmonic.
- t: Time in hours, expressed in Coordinated Universal Time (UTC).
- ϵ: Represents higher-order harmonics and other non-periodic components.

The above equation can also be expressed in a simplified form:

$$IWV(t) = A_n \sin\left(\frac{2\pi nt}{24} + \sigma_n\right) \tag{2.9}$$

Where $A_n = sqrt\{a_n^2 + b_n^2\}$: Represents the amplitude of the n-th harmonic and $\boldsymbol{\sigma_n}$ Represents the phase shift of the n-th harmonic.

The amplitude A_n indicates the strength of the harmonic component, while the phase σ_n determines the timing of the maximum IWV for the given harmonic. The phase can also be translated into the time of day when the IWV reaches its peak, denoted as P_n:

$$P_n = \frac{\left(\frac{\pi}{2} - \sigma_n\right) \cdot 12}{\pi n} \tag{2.10}$$

The value 12 in the equation converts the phase σ_n from radians to hours within a 24-hour diurnal cycle for the n^{th} harmonic, where $12/\pi$ maps 2π radians to 24 hours, scaled by n. This equation allows us to determine the time of maximum IWV for each harmonic, giving a more intuitive understanding of when the peak diurnal and semi-diurnal variations occur throughout the day. The time is always expressed in Coordinated Universal Time (UTC) to maintain consistency and facilitate comparison with other studies and global observations.

2.2.3 Fisher's Significance Test

Fisher's significance test is a widely used statistical method in harmonic analysis to determine whether or not the amplitudes and phases of the Fourier series coefficients are statistically

significant. In the context of this study, Fisher's test is employed to verify whether the observed harmonics (i.e., the diurnal and semi-diurnal components) are significant or not.

Fisher's significance test is applied to the time series after computing the harmonic components. The test statistic is calculated as follows:

$$F = \frac{\frac{1}{2}(a_n^2 + b_n^2)}{\frac{\sigma^2}{N}} \tag{2.11}$$

Where:

- F: Fisher's test statistic.
- a_n and b_n: Fourier series coefficients for the n-th harmonic.
- σ^2: Variance of the residuals, which represents the part of the time series not explained by the harmonic model.
- N: Number of data points in the time series.

The calculated F value is compared to a critical value from the Fisher distribution table, which depends on the degrees of freedom and the significance level (α), typically set at 0.05. If the calculated F value is greater than the critical value, the null hypothesis that the harmonic has no real contribution to the signal can be rejected. This indicates that the harmonic component is statistically significant and contributes to the observed variability in IWV.

2.2.4 Vertically Integrated Moisture Flux Convergence

The diurnal cycle of Integrated Water Vapor (IWV) is influenced by various factors, including evaporation, condensation, land/sea breeze circulation, precipitation, the region's orography, and moisture advection. To examine how moisture advection affects the diurnal variation of IWV, the Vertically Integrated Moisture Flux Convergence (VIMFC) has been analyzed using wind and specific humidity data from the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis Fifth Generation (ERA5) dataset over a four-year period (2018-2021). The VIMFC represents the net convergence or divergence of atmospheric moisture over a given

location, providing insights into the contribution of horizontal moisture transport to IWV variability.

2.2.4.1 Estimation of VIMFC

VIMFC is calculated from wind components (u and v) and specific humidity (q) using the following formula:

$$VIMFC = \frac{1}{g}\int_{p_s}^{p_t}\left(\frac{\partial(\mathrm{uq})}{\partial \mathrm{x}} + \frac{\partial(\mathrm{vq})}{\partial \mathrm{y}}\right)dp \quad (2.12)$$

Where:

q: Specific humidity (kg/kg).

- u, v: Zonal and meridional wind components (m/s), respectively.
- p_s: Surface pressure (hPa).
- p_t: Pressure at the top of the atmospheric layer (typically set to 300 hPa in this study).
- g: Acceleration due to gravity (9.81 m/s).
- $\left(\frac{\partial(\mathrm{uq})}{\partial \mathrm{x}}, \frac{\partial(\mathrm{vq})}{\partial \mathrm{y}}\right)$: Horizontal gradients of moisture flux, representing moisture advection.

The VIMFC provides a measure of the amount of moisture convergence or divergence per unit area over a given period.

The resulting VIMFC values represent the net effect of moisture convergence or divergence over the column extending from the surface to 300 hPa.

2.2.4.1.1 Positive VIMFC: Indicates moisture convergence, meaning that water vapor is accumulating in the column. This often leads to an increase in IWV and is indicative of conditions favorable for cloud formation and possibly precipitation. High positive VIMFC values represent strong convergence, which can be associated with convective activity and storm development.

2.2.4.1.2 Negative VIMFC: Indicates moisture divergence, meaning that water vapor is moving away from the column. This results in a decrease in IWV and can indicate subsidence or drying

conditions. High negative values represent strong divergence, often linked to clear skies and stable atmospheric conditions.

2.2.5 Used Machine Learning Methods

2.2.5.1 Light GBM

LightGBM (Ke *et al.* 2017) is a powerful gradient boosting framework that builds multiple decision trees sequentially. Each tree is trained to minimize the residual errors from the previous trees, ultimately forming a robust model for regression tasks.

Mathematical Description: The prediction from LightGBM is based on an ensemble of decision trees, which is represented as:

$$f(x) = \sum_{\{m=1\}}^{M} T_m(x) \tag{2.13}$$

Where:

- f(x): The final predicted value for input feature vector x.
- M: The total number of decision trees.
- $T_m(x)$: The prediction from the m-th tree.

The objective of LightGBM is to minimize the loss function, which consists of a fitting error term $l(y_i,f(x_i))$ and a regularization term $\Omega(T_m)$ to control overfitting:

$$L(y, \hat{y}) = \sum_{\{i=1\}}^{n} l\big(y_i, f(x_i)\big) + \sum_{\{m=1\}}^{M} \Omega(T_m) \tag{2.14}$$

Where:

- $l(y_i,f(x_i))$: The loss for each training instance, typically the Mean Squared Error (MSE) for regression tasks.

$$l(y_i, f(x_i)) = (y_i - f(x_i))^2 \quad (2.15)$$

- $\Omega(T_m)$: The regularization term, which controls the complexity of each tree and helps prevent overfitting.

LightGBM uses gradient boosting to optimize this objective, building trees that iteratively minimize the error.

2.2.5.2 Storm Nowcasting Using Classification Models

In Chapter 6, several classification models including LightGBM, Logistic Regression, Random Forest, XGBoost, CatBoost, and Artificial Neural Networks (ANNs)—are used for storm classification. Each model has its unique way of addressing classification tasks. Here's a detailed look at each:

2.2.5.2.1 Logistic Regression

Logistic regression is a linear model that predicts the probability of a binary outcome using the logistic function (Hosmer *et al.* 2013):

$$p = \frac{1}{1 + e^{\{-z\}}} \quad (2.16)$$

Where:

$$z = w_0 + \sum_{\{i=1\}}^{n} w_i x_i \quad (2.17)$$

- w_0: Intercept term.
- w_i Coefficients for the features x_i.

The model is trained using the same binary cross-entropy loss as LightGBM, with the objective of maximizing the likelihood of correct predictions.

2.2.5.2.2 Random Forest

Random Forest is an ensemble learning method that builds multiple decision trees using random subsets of data and features, and then aggregates their outputs (Breiman, 2001).

For classification, the final prediction is determined by majority voting from all decision trees:

$$f(x) = mode\{T_1(x), T_2(x), \dots, T_m(x)\} \tag{2.18}$$

Where:

- $T_m(x)$: Prediction from the mmm-th decision tree.
- $f(x)$: The final classification output.

The advantage of Random Forest lies in its ability to reduce overfitting and improve generalization by averaging the results of multiple trees.

2.2.5.2.3 Extreme Gradient Boosting (XGBoost)

XGBoost is an optimized implementation of gradient boosting that uses a regularization term to enhance model generalizability. The model is built sequentially by adding decision trees that minimize the objective function (Chen & Guestrin, 2016):

$$L(\theta) = \sum_{\{i=1\}}^{n} l(y_i, \hat{y_i}) + \sum_{\{m=1\}}^{M} \Omega(f_m) \tag{2.19}$$

Where:

- $l(y_i, \hat{y_i})$: The loss function for each observation, typically binary cross-entropy for classification.

- $\Omega(f_m)$: Regularization term, which includes terms to penalize the number of leaves and leaf weights, reducing overfitting.

Each tree improves on the previous iteration by learning from the residual errors.

2.2.5.2.4 CatBoost

CatBoost builds gradient-boosted decision trees, designed specifically to handle categorical features efficiently. The model uses ordered boosting, which prevents overfitting by ensuring that each tree is built without data leakage from the current training instance (Dorogush *et al.* 2018).

The objective function for CatBoost is similar to that of XGBoost:

$$L(\theta) = \sum_{\{i=1\}}^{n} l(y_i, \hat{y}_i) + \sum_{\{m=1\}}^{M} \Omega(f_m) \tag{2.20}$$

Where:

$L(\theta)$ is Total loss function to minimize, with model parameters θ, y_i is True target value for sample i , $\hat{y}_i$ is predicted value for sample i *and* $\Omega(f_m)$ is Penalty for tree complexity.

The key advantage of CatBoost is its ability to handle categorical features without requiring extensive preprocessing, and the use of ordered boosting for stable training.

2.2.5.2.5 Artificial Neural Networks (ANNs)

ANNs are made up of interconnected layers of neurons that help to model complex relationships in the data. The feedforward calculation in an ANN is given by (Goodfellow *et al.* 2016):

$$z^{(l)} = W^{(l)} A^{(l-1)} + b^{(l)} \tag{2.21}$$

$$A^{(l)} = f(z^{(l)}) \tag{2.22}$$

Where:

- $z^{(l)}$: The linear combination of weights, inputs, and biases for the l -th layer.
- $A^{(l)}$: Activation output for the l -th layer.

- $W^{(l)}$: Weight matrix for layer l.
- $b^{(l)}$:Bias term for layer l.
- f: Activation function (e.g., ReLU, sigmoid).

The model is trained using backpropagation to minimize a loss function (typically binary cross-entropy in classification tasks).

2.2.5.3 Evaluation Metrics for Classification Models

2.2.5.3.1 Area Under the Receiver Operating Characteristic Curve (AUC-ROC)

The AUC-ROC measures the model's ability to distinguish between the positive and negative classes (storm vs. no storm). It plots the True Positive Rate (TPR) against the False Positive Rate (FPR) (Fawcett, 2006):

$$TPR = \frac{TP}{TP + FN} \tag{2.23}$$

$$FPR = \frac{FP}{FP + TN} \tag{2.24}$$

The AUC value ranges between 0 and 1, where a higher value indicates better model discrimination. A value close to 1 indicates that the model is good at distinguishing between storm and non-storm events.

2.2.5.3.2 Cohen's Kappa Score

Cohen's Kappa measures the agreement between the predicted classifications and the actual labels, adjusted for chance agreement (Cohen, 1960):

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{2.25}$$

Where: p_o is Observed agreement (i.e., the proportion of times the model and actual classifications agree). p_e is Expected agreement by chance.

A Kappa score of 1 indicates perfect agreement, while a score of 0 suggests that agreement is no better than random.

2.2.5.4 Evaluation Metrics for Prediction Outcomes

The following metrics are also used to evaluate the prediction of storm events:

2.2.5.4.1 Probability of Detection (POD)

Measures the proportion of correctly detected storm events:

$$POD = \frac{d}{d + c} \times 100 \tag{2.26}$$

Where: d is True positives (correctly predicted storm events) and c: False negatives (missed storm events).

2.2.5.4.2 False Alarm Rate (FAR)

Measures the proportion of false alarms:

$$FAR = \frac{b}{b + d} \times 100 \tag{2.27}$$

Where: b is False positives (incorrectly predicted storm events) and d is True positives.

2.2.5.4.3 Miss Rate

Measures the proportion of missed storm events:

$$Miss\ Rate = \frac{c}{c + d} \times 100 \tag{2.28}$$

2.2.5.4.4 Model Accuracy

Measures the proportion of correctly predicted events (storm or no storm):

$$Model\ Accuracy = \frac{a + d}{a + b + c + d} \times 100 \tag{2.29}$$

Where a is True negatives (correctly predicted non-storm events), b is False positives, c is False negatives, d is True positives.

Chapter 3

Diurnal variability of IWV and underlying mechanisms

3.1 Introduction

Numerous authors have analyzed the diurnal variations of IWV over different regions of the world using GNSS observations and attributed the variations to . For instance, Dai *et al.* (2002) studied the diurnal variation of IWV over North America using data from 54 stations over 4 years (1996-2000). The study shows that the diurnal variation is significant and varies spatially, with higher amplitude in the western United States and the subtropics. The phase of the Diurnal cycle peaks around noon (100 -1400 LST) in winter, from midafternoon to midnight in summer, and for autumn a couple of hours earlier than in summer.

Kalinnikov and Khutorva (2017) studied the diurnal variations of IWV over the Volga –Ural region of Russia from 16 GPS stations during 2013-2015.The diurnal harmonic phase for winter and summer falls within the period from 14:00 to 17:00. Ortiz de Galisto *et al.* (2011) studied annual and seasonal diurnal cycles at 10 stations over Spain for 7 years (2002-2008). They have found that both the annual and seasonal diurnal cycles at the different stations are similar to each other during the night than during the afternoon and the mechanisms changing the IWV are influenced more by local effects during the daytime than during the night. Their results also showed that the diurnal cycle in winter is quite similar at all locations, whereas in summer it is quite different.

Wu *et al.* (2003) studied the diurnal variation of IWV at Koto Tabang, a mountainous area of Sumatra Island, in the dry season from June to August 2001. They have observed a distinct diurnal variation, where IWV increases during the daytime reaching its maximum in the late afternoon at about 1700 LST. They have suggested that the diurnal variation of IWV is caused by the transport of water vapor by thermally induced local circulations. Torri *et al.* (2019) also studied the diurnal variation of IWV using data from stations of the Sumatran GPS Array (SuGAr) during 2008-2013. Their results show that Madden-Julian Oscillation has an impact on the diurnal cycles in both daily mean and diurnal amplitudes. Large variations of amplitudes are seen during the active phase of MJO and small variations during the suppressed phase of MJO. They have also seen a shift in the IWV peak, the peak moves from 19 LST in the active phase to 22 LST in the suppressed phase.

Li *et al.* (2008) found that the diurnal maximum appears during 1800 - 2000 LST in Japan, whereas the diurnal minimum appears near the noon for mountainous areas and basins but in the early morning for plains. Their results suggested that the differences in phase of the IWV diurnal cycle between different locations are due to moisture advection. Meza *et al.* (2020) analyzed annual and

seasonal diurnal IWV variations over central and South America for the period 2007-2013 using principal component analysis (PCA). They have classified the stations into four categories inland, coastal (stations located less than 5km from a water body), valley (stations surrounded by mountains), and undefined (stations that do not fall in any of the previous three categories). The analysis shows that PWV exhibits a clear diurnal cycle with a maximum in the afternoon and a minimum in the early morning. The diurnal amplitude of PWV is higher over Central America than over South America, which is attributed to the stronger land-sea breeze circulation over the former. The diurnal cycle is also affected by local topography, with higher PWV values observed over mountainous stations. The authors suggest that surface heating is the main driver of the diurnal cycle, with the daily cycle of insolation playing a key role.

Radhakrishna *et al.* (2014) studied the diurnal variation of water vapor over the continental United States obtained from GPS and North American Regional Reanalysis. The study showed that the mean seasonal diurnal cycle of Water vapor has both spatial and temporal variation across the continental United States. The diurnal cycle also shows significant distinct variation across seasons. The phase of the diurnal cycle occurs at 6 UTC (middle of the night) for the Rockies and the western United States, whereas for the Eastern side of the Great Plains to the Midwest United States phase occurs at 0 UTC (late afternoon).

Kannemadugu *et al.* (2022) analyzed the diurnal variation of 18 stations in India and observed that diurnal variation is not uniform across all the stations and is dependent on seasons and location. At Hyderabad, Bangalore, Madurai, and Bhopal (semi-arid locations), it has been observed there is a phase lag between the seasons and the phase lag is different for different locations. For Hyderabad, the diurnal peak of winter and post-monsoon is lagging by 3-4 hrs to pre-monsoon and monsoon diurnal peaks. For Bangalore, the diurnal peak of monsoon is lagging by 1-2 hrs to pre-monsoon, winter, and post-monsoon diurnal peaks. For Bhopal and Madurai, the diurnal peak of pre-monsoon and post-monsoon is lagging by 1-2 hrs to winter and monsoon diurnal peaks. Barman *et al.* (2017) analyzed the diurnal variation of 5 stations in northeast India and observed that within the region of the study itself, diurnal variation is distinct for each site. Stations in Assam Valley do not have a distinct diurnal cycle.

Nirmala Bai *et al.* (2020) studied the diurnal variation of IWV over a tropical station (Hyderabad) in India and observed that the annual diurnal cycle IWV is minimum in the morning and maximum in the evening. They also observed a shift in the phase of the IWV cycle with seasons. Diurnal

harmonic is strongest and more dominant than 12hr and 8hr harmonics and a positive correlation of 0.44 is observed between IWV and surface temperature for the whole year. The correlation between IWV and surface temperature is not uniform throughout the seasons.

Even though several studies, as elucidated above, have examined the diurnal cycle of IWV and suggested the factors responsible for this variation for different parts of the world, there have been very less studies about the diurnal variation of water vapor in India and in semi-arid regions. The studies over India talk provide a general picture of the diurnal cycle of IWV, do not study about the factors influencing the variation in detail, and clearly indicate that there is spatial and topographic variability in the diurnal cycle of IWV. All these studies show that the diurnal variation of IWV is dependent on so many factors. The area of study in this paper Gadanki, India (13.46° N, 79.1733° E) is in a tropical semi-arid region in Southern India, surrounded by mountains. This chapter is structured to cover multiple aspects of IWV variability, including identifying significant harmonics in the diurnal cycle, understanding the underlying mechanisms contributing to these variations, and assessing how these patterns change across different seasons. The analysis includes the quantification of diurnal, semidiurnal, and terdiurnal components and their respective amplitudes and phases. We also examine the contribution of meteorological factors, such as evaporation, moisture advection, and atmospheric dynamics, which influence the diurnal and seasonal variability of IWV. By analyzing these harmonics and understanding the physical mechanisms behind IWV changes, this chapter aims to shed light on the intricate atmospheric processes that govern water vapor behavior in semi-arid tropical regions like Gadanki. Such insights are crucial for enhancing the accuracy of weather forecasting models, particularly in predicting rainfall and storm events, which are highly sensitive to variations in atmospheric moisture.

3.2. Data and Methodology

The data used in this study have been described in detail in Chapter 2. Here, we provide a brief description of the dataset used in this chapter.

This study utilized the GNSS data from the NARL GNSS network with a time resolution of 15 minutes over a period of 3 years, from September 2018 to August 2021. The raw GNSS data at 1-second intervals for all stations were analyzed using GAMIT software to estimate Integrated Water Vapor (IWV). The precise final ephemeris from the International GNSS Service (IGS) and the

Vienna Mapping Function (VMF) were used for the processing of raw GNSS data and estimation of Zenith Total Delay (ZTD). The meteorological data (surface pressure and temperature) needed for converting ZTD to IWV were obtained from the Automatic Weather Station (AWS). The AWS data were available at a 30-second resolution and were averaged over 30-minute intervals before being used in GAMIT processing.

3.3 Results

3.3.1. Quantification of Significant Harmonics

Previous studies have indicated that the diurnal harmonic (S1) is generally stronger than the semidiurnal harmonic (S2) in the mean diurnal variation of IWV, with a few exceptions during certain seasons. In this study, both the mean diurnal variation and day-to-day variations in IWV are analyzed using harmonic analysis and Fischer's significance test, as described in Chapter 2, to identify which harmonics are significant. Harmonic analysis involves decomposing the IWV time series into different sinusoidal components, each representing variations with specific frequencies. This method helps in understanding the periodic behavior of IWV and identifying the key contributors to its variability. Fischer's significance test is used to determine whether the identified harmonics are statistically significant, thus allowing us to categorize the variation patterns.

In this study, harmonics are classified into five modes based on their contribution to the diurnal variation:

Mode 1: Only the diurnal harmonic is significant, indicating that the primary variation occurs over a 24-hour period, typically driven by solar radiation and evaporation of soil moisture/water.

Mode 2: Only the semidiurnal harmonic is significant, suggesting that the variation follows a 12-hour cycle, which could be influenced by atmospheric tides or other periodic forcings.

Mode 3: Both diurnal and semidiurnal harmonics are significant, indicating a complex interaction between daily solar heating and other sub-daily atmospheric processes.

Mode 4: Neither diurnal nor semidiurnal harmonic is significant, implying that the variation is either irregular or dominated by non-periodic factors, such as transient weather systems or localized convection.

Mode 5: Both diurnal and terdiurnal (8-hour period) harmonics are significant, highlighting additional complexity in the IWV variation, possibly due to rapid changes in local meteorological conditions or multiple overlapping forcing mechanisms.

Figure 3.1 shows the percentage distribution of different modes, based on above classification, that are significant. For 93% of the time, the diurnal harmonic is significant, indicating the strong influence of the daily solar cycle on IWV variability. The semidiurnal harmonic is significant for 36% of the time, suggesting that atmospheric tides also play a role in modulating IWV. Approximately 31% of the time, both diurnal and semidiurnal harmonics contribute to the variation, demonstrating the interplay between different periodic processes. Additionally, there are 53 days where both diurnal and terdiurnal harmonics are significant, pointing to more complex sub-daily variations that could be linked to rapid shifts in atmospheric dynamics or localized convection events.

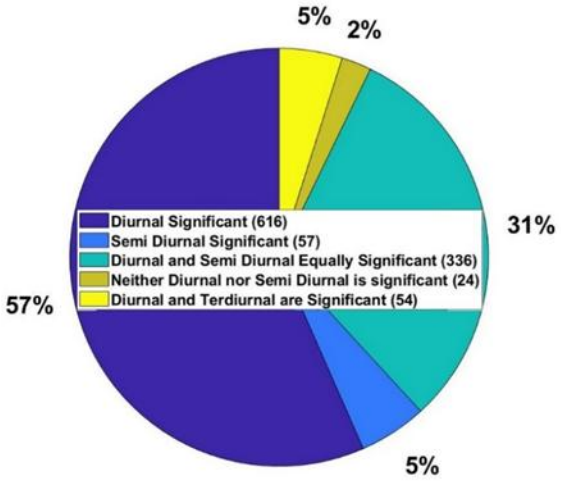

Figure 3.1: *Percentage distribution of different significant harmonics for the whole data.*

The classification of these modes helps in understanding the dominant factors influencing IWV variability in Gadanki. For example, the dominance of the diurnal harmonic across most of the dataset suggests that solar radiation and associated surface processes are the primary drivers of water vapor variability. However, the presence of semidiurnal and terdiurnal harmonics indicates that other processes, such as atmospheric tides, mesoscale circulations, and localized meteorological phenomena, also contribute significantly to the observed variations. Understanding these modes and their significance provides valuable insights into the behavior of IWV, which is

critical for improving the representation of water vapor in weather models, particularly in regions with complex topography and varying climatic conditions.

3.3.2. Seasonal and Monthly Variation of Amplitudes and Phases

The seasonal distribution of different harmonics, as depicted in Figure 3.2, reveals that the diurnal harmonic (S1) is significant for more than 90% of the data in all seasons, indicating that the diurnal cycle driven by solar radiation is the dominant factor influencing IWV variability. Semidiurnal harmonics (S2) are more significant during winter and pre-monsoon seasons, suggesting that atmospheric tides and other sub-daily forcings play a more prominent role during these periods.

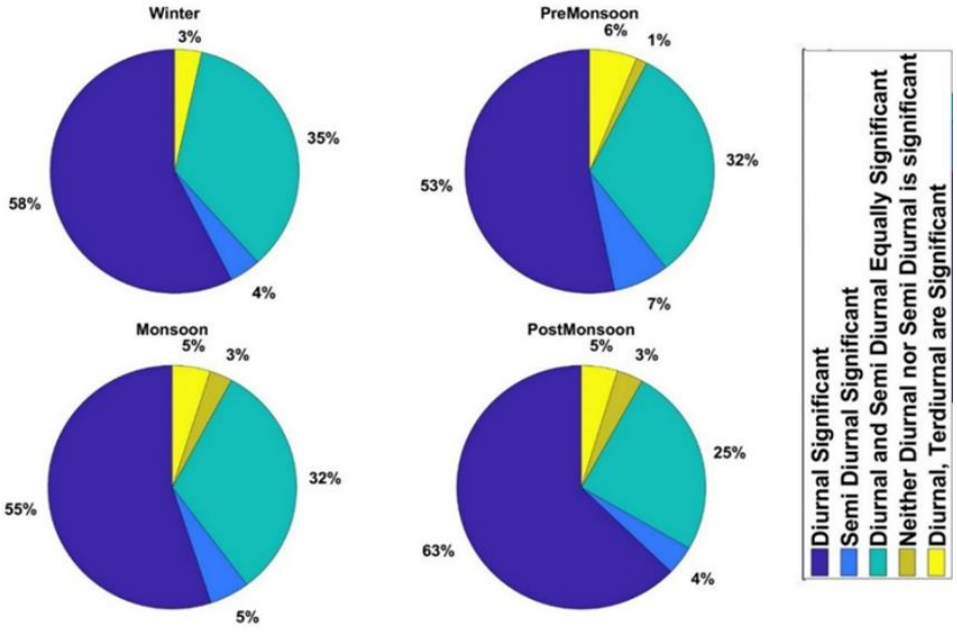

Figure 3.2: *Percentage distribution of different significant harmonics for the whole data in different seasons.*

The seasonal distribution also shows that diurnal amplitudes greater than 5 mm occur more frequently in pre-monsoon and post-monsoon seasons, which could be attributed to stronger solar heating and enhanced evaporation during these times. In contrast, semidiurnal variations of approximately 1.5 mm are more common during winter, monsoon, and post-monsoon seasons, which may be linked to atmospheric tides and the influence of mesoscale processes.

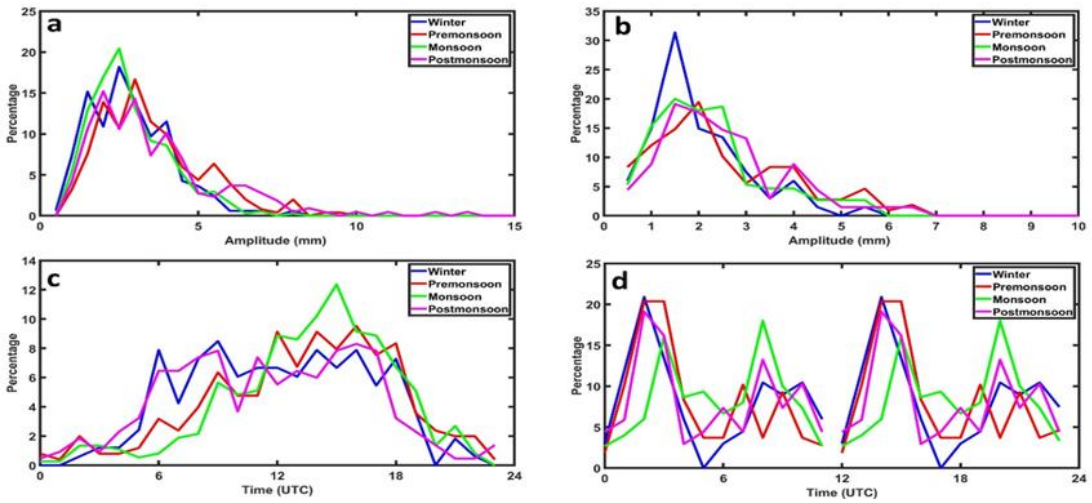

Figure 2.3: *Percentage occurrence of a) diurnal amplitudes (from Mode-1, Mode-3 and Mode-5 Cases, Diurnal significant days) b) Semi-Diurnal Amplitudes (from Mode-2 and Mode-3 Cases, Semi Diurnal Significant days) c) Diurnal Phase (from Mode-1, Mode-3 and Mode-5 Cases, Diurnal significant days) and d) Semi-Diurnal Phase (from Mode-2 and Mode-3 Cases, Semi Diurnal Significant days) across different seasons.*

Figure 3.3 provides a detailed depiction of the percentage occurrence of amplitudes and phases for both diurnal and semidiurnal harmonics across different seasons. The occurrence of IWV maxima is more frequent from afternoon to night, consistent with the timing of peak solar heating and subsequent moisture convergence. Distinct semidiurnal cycles are observed in all seasons, but their peak occurrences vary, reflecting the influence of different atmospheric processes, such as thermal tides and convection.

Figure 3.4 shows the monthly variation in amplitudes and phases, indicating that both diurnal and semidiurnal amplitudes are highest in November. This suggests that the transition period between monsoon and winter is characterized by significant moisture variability, with average diurnal amplitudes ranging from 2.3 to 3.8 mm and semidiurnal amplitudes ranging from 1.4 to 2 mm. The higher amplitudes in November could be attributed to the combined effects of residual moisture from the monsoon and the onset of wintertime atmospheric stability, which enhances the modulation of IWV.

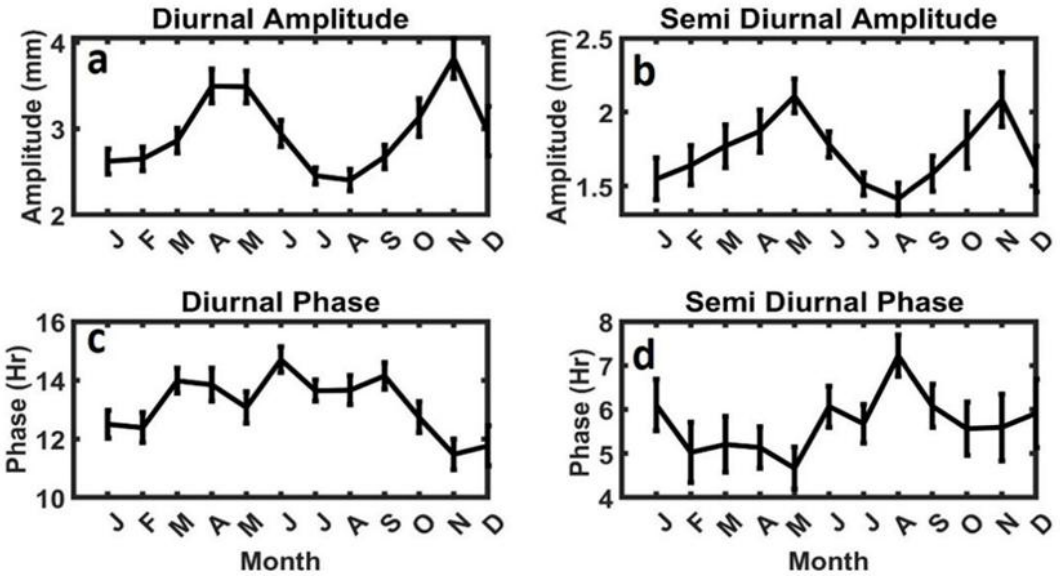

Figure 3.4: *Monthly variation of a) Diurnal amplitude b) Semi-Diurnal Amplitude c) Diurnal Phase and d) Semi Diurnal phase for the whole data. The error bar in Fig 4a & 4b represents the standard error of amplitudes in mm, whereas the error bar in Fig 4c & 4d represents standard error of phases in h.*

3.3.3. Mean Seasonal Diurnal Variation

Figure 3.5 shows the mean diurnal variation for each season, highlighting the distinct seasonal patterns in the diurnal cycle of IWV. The diurnal cycle of IWV is characterized by a low in the early morning, typically around sunrise, followed by an increase throughout the day until evening or night, and then a subsequent decrease during the late night to early morning hours. This diurnal pattern is influenced by the daily cycle of solar heating, which drives evaporation and atmospheric turbulence. The timing and amplitude of these variations are season dependent, reflecting changes in solar radiation, atmospheric stability, and moisture availability across different seasons.

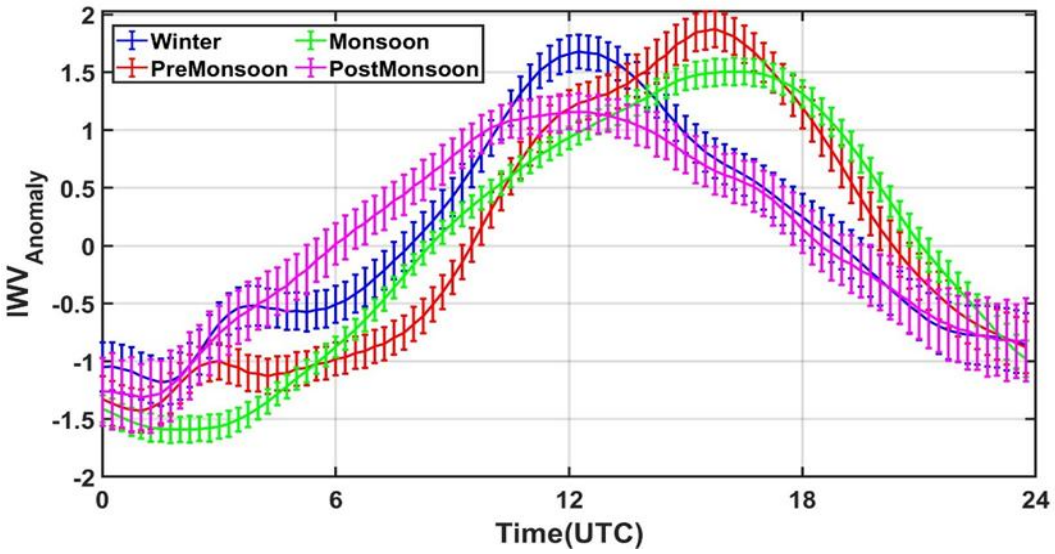

Figure 3.5: *Seasonal mean diurnal cycle of IWV at Gadanki. The unit of Y-axis is mm. The error bar represents the standard error of the anomalies in mm.*

During winter and post-monsoon seasons, the diurnal cycle of IWV peaks at 12:15 UTC, whereas in the pre-monsoon and monsoon seasons, the peak occurs later at 15:00 UTC. This shift in peak timing can be attributed to differences in surface heating and atmospheric dynamics. In the pre-monsoon and monsoon seasons, increased solar radiation and convective activity lead to a delayed and more pronounced peak in IWV, whereas in winter and post-monsoon seasons, the relatively stable atmosphere results in an earlier peak.

The mechanisms contributing to these variations include evaporation, moisture advection, and topographical influences. Surface latent heat flux (SLHF), which represents the rate of evaporation from the surface, and vertically integrated moisture flux convergence (VIMFC), which indicates the net horizontal transport of moisture, are key indicators used to understand these processes. Figure 3.6 illustrates the relationship between IWV, VIMFC, and SLHF, showing that the diurnal variation in IWV is closely related to VIMFC. Moisture convergence in the lower atmosphere leads to an increase in IWV, while divergence results in a decrease. The phase lag between seasonal IWV variations can be explained by the combined effects of SLHF and VIMFC, where the timing of maximum evaporation and moisture convergence varies across seasons, leading to differences in the diurnal cycle of IWV.

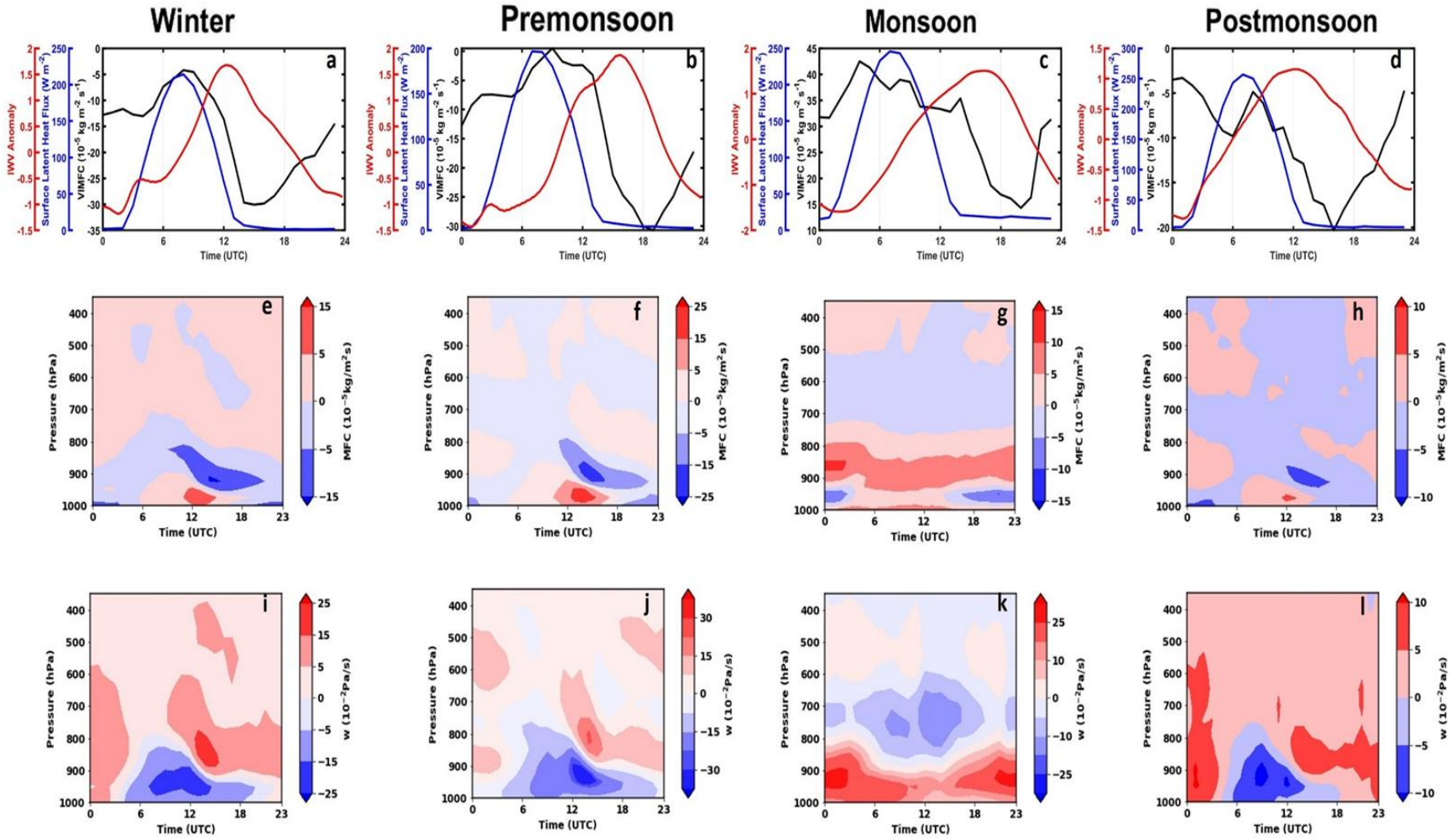

Figure 3.6: *Intercomparison of Diurnal cycle of IWV, Vertically Integrated Moisture Flux convergence (+ve is convergence and -ve is divergence) and Surface Latent Heat Flux from ERA5 in different seasons (a-d). Diurnal cycle of Moisture Flux convergence (+ve is convergence and -ve is divergence) at different pressure levels for different seasons (e-h). Diurnal cycle of Vertical wind (+ve is downdrafts and -ve is updrafts) at different pressure levels for different seasons (i-l).*

3.3.4. Diurnal Harmonic Significant Days

Figure 3.7 shows the diurnal variation of individual diurnal harmonic significant days. It is evident that the diurnal variation in IWV is strongly influenced by the vertically integrated moisture flux convergence (VIMFC), which dictates both the timing and magnitude of IWV peaks. The peak in IWV generally occurs when there is a transition from moisture divergence to convergence, indicating that the accumulation of moisture in the atmospheric column is a key factor driving the diurnal maxima. The strong correlation between IWV peaks and VIMFC peaks highlights the significance of moisture transport processes, such as advection and convergence, in regulating IWV variability.

In addition to VIMFC, the day-to-day variation in IWV is influenced by other atmospheric factors. Changes in VIMFC patterns, driven by synoptic-scale weather systems, can significantly impact the moisture content in the atmosphere. For instance, the passage of frontal systems or the influence of low-pressure areas can alter the convergence patterns, thereby affecting IWV levels. Precipitation events, particularly those occurring on preceding days, can lead to a temporary depletion of atmospheric moisture, followed by a recovery phase where evaporation and moisture advection contribute to replenishing IWV. This recovery phase can further modulate the diurnal cycle, creating variability in the timing and magnitude of subsequent peaks.

The surface latent heat flux (SLHF), which represents the rate of evaporation from the surface, plays a crucial role in the early morning buildup of IWV. During the daytime, increased solar radiation enhances SLHF, leading to higher evaporation rates and contributing to the gradual increase in IWV. The boundary layer processes, including turbulence and vertical mixing, also influence the distribution of water vapor, especially during the transition from night to day. Turbulence within the boundary layer facilitates the upward transport of moisture, supporting the increase in IWV during the morning hours. The interplay between VIMFC, SLHF, synoptic-scale weather systems, and boundary layer dynamics governs the diurnal variability of IWV. The complexity of these interactions highlights the need for detailed observational and modeling studies to improve our understanding of moisture processes and their role in atmospheric dynamics over tropical regions like Gadanki.

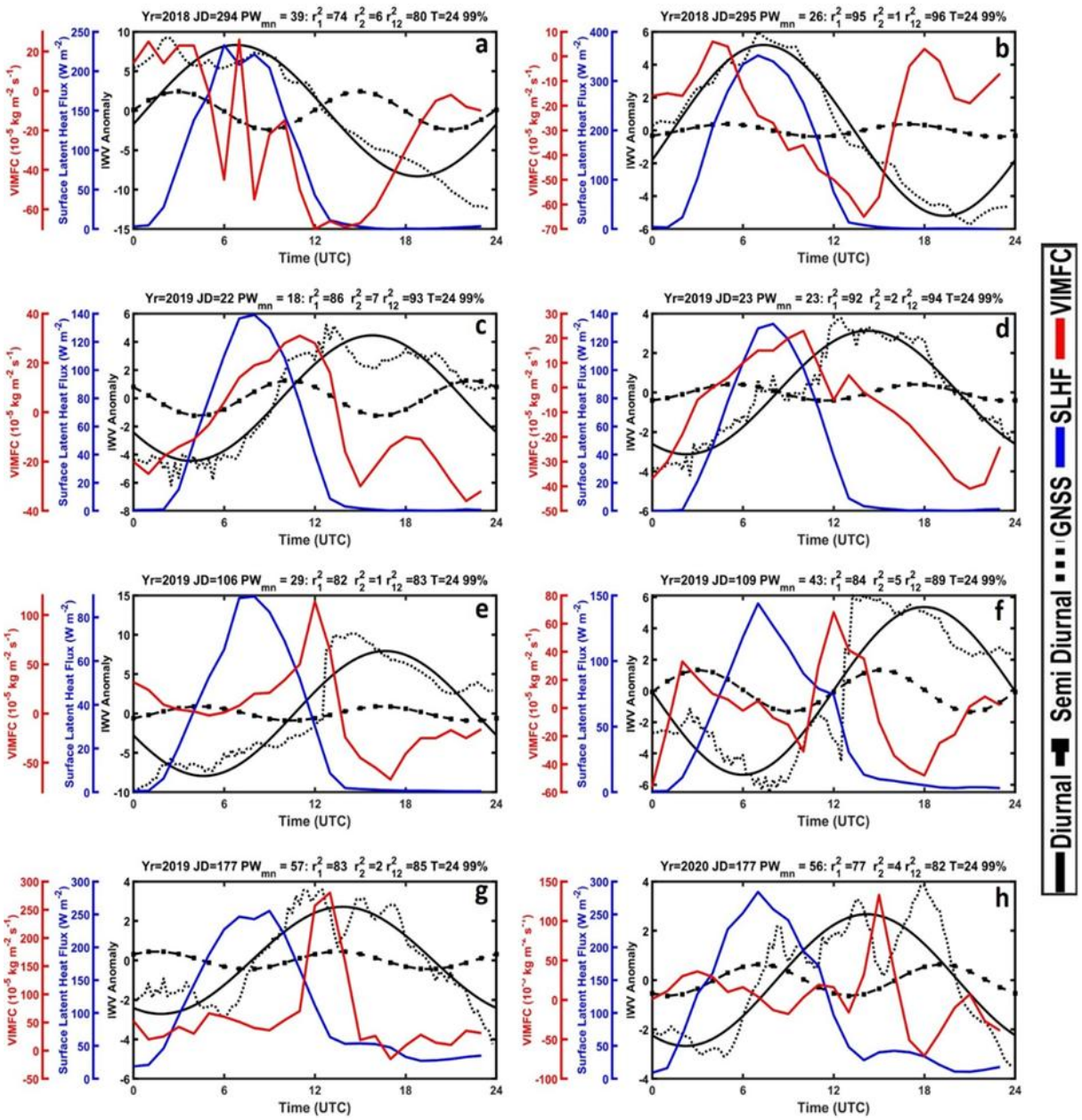

Figure 3.7: *IWV diurnal cycle of individual days for only diurnal significant days. Details on the top of each subplot are Yr is Year, JD is Julian Day, PWmn is Mean IWV of that day, r12 is % contribution of Diurnal harmonic to the diurnal variation of IWV, r22 is % contribution of Semi-Diurnal harmonic to the diurnal variation of IWV, r122 is % contribution of both Diurnal and semi-diurnal harmonics to the diurnal variation of IWV, T is the Time period of significant harmonics and the percentage number at the last is the confidence level at which particular harmonic is significant.*

3.3.5 Both Diurnal and Semidiurnal Harmonic Significant Days

Figure 3.8 shows the days when both diurnal and semidiurnal harmonics are significant. Typically, the diurnal harmonic is stronger than the semidiurnal harmonic, indicating that solar-driven processes dominate the moisture variation. The presence of both harmonics suggests a complex interaction between daily solar heating and atmospheric tides, leading to variability in the moisture content. The variation pattern is heavily influenced by changes in the vertically integrated moisture flux convergence (VIMFC), which affects the timing and magnitude of IWV peaks.

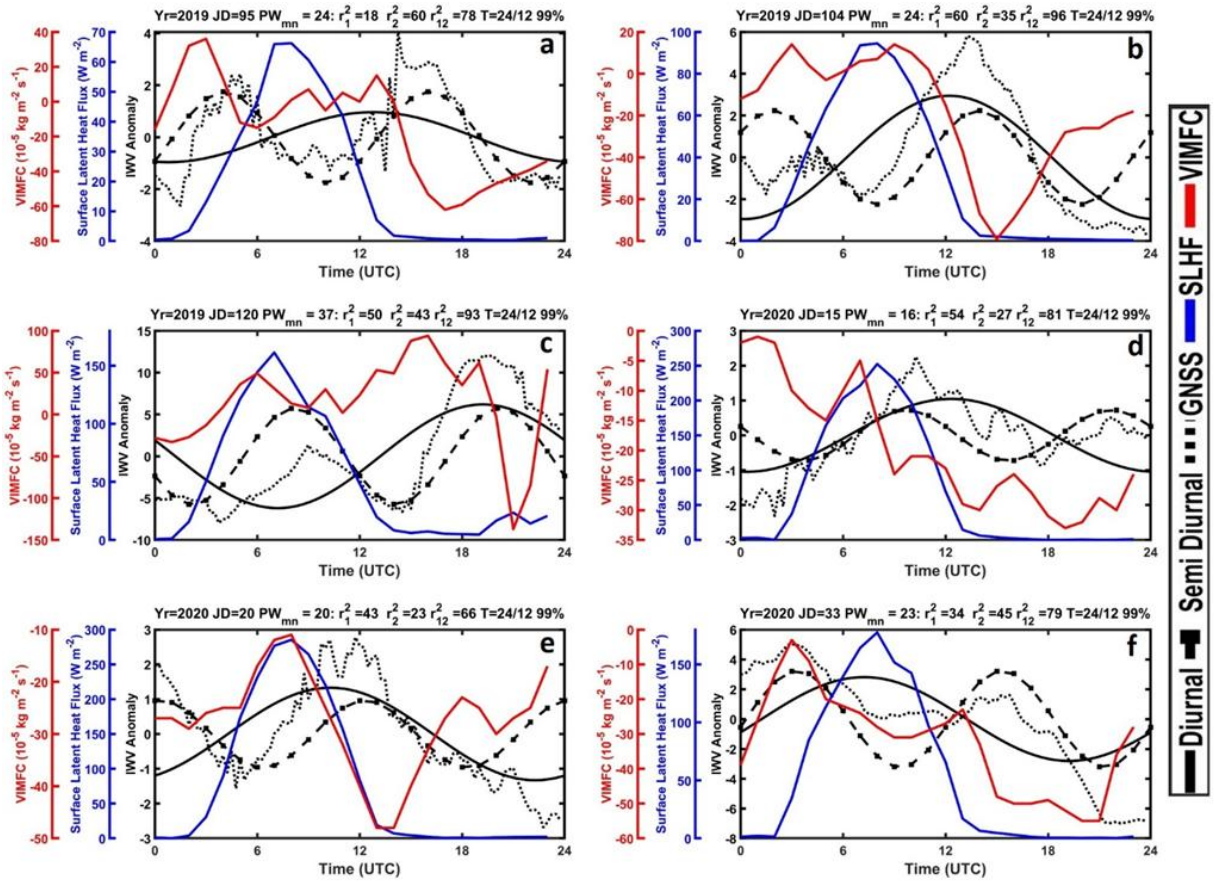

Figure 3.8: *IWV diurnal cycle of individual days for both diurnal and semi-diurnal significant days. Details on the top of each subplot are Yr is Year, JD is Julian Day, PWmn is Mean IWV of that day, r12 is % contribution of Diurnal harmonic to the diurnal variation of IWV, r22 is % contribution of Semi-Diurnal harmonic to the diurnal variation of IWV, r122 is % contribution of both Diurnal and semi-diurnal harmonics to the diurnal variation of IWV, T is the Time period of significant harmonics and the percentage number at the last is the confidence level at which particular harmonic is significant.*

Sudden shifts in VIMFC, such as changes from convergence to divergence or vice versa, result in abrupt peaks or dips in IWV. These dynamics highlight the role of synoptic weather conditions,

mesoscale processes, and atmospheric stability in modulating the contributions of both harmonics to the observed variability in IWV. This interaction is particularly pronounced during transitional weather periods, when multiple atmospheric processes overlap, leading to increased complexity in the IWV diurnal cycle.

3.3.6 Neither Diurnal nor Semidiurnal Harmonic is Significant

Figure 3.9 shows the days on which neither diurnal nor semidiurnal harmonic is significant. On these days, the diurnal variation of IWV appears to be irregular or highly influenced by non-periodic factors. The diurnal variation can be random or perturbed due to advection of moisture into the atmosphere by transient synoptic systems or local convection events. For example, on Julian Day (JD) 260 of the years 2018 and 2021 (Figure 3.13(a, f)), the diurnal variation is characterized by low IWV levels at the beginning of the day, followed by a sudden increase and subsequent decrease. This pattern is attributed to moisture advection and precipitation events that occurred after significant moisture buildup at 16 UTC and 13 UTC for JD 260 of 2018 and 2021, respectively. The advection of moisture into the atmosphere at 14 UTC for JD 260 of 2018 and at 13 UTC for JD 260 of 2021 led to substantial variability in IWV, disrupting any potential harmonic pattern.

In the absence of such moisture advection or precipitation events, the IWV variations on these days might have exhibited a clearer diurnal or semidiurnal harmonic. However, the presence of these atmospheric disturbances leads to a lack of significant periodicity, resulting in random or complex diurnal patterns. Figure 3.13(b-e) shows instances where either 8-hour or 6-hour harmonics were significant, or where no significant harmonic patterns were observed at all. It is evident that on these days, the diurnal variation is not driven by typical forcings like vertically integrated moisture flux convergence (VIMFC) or surface latent heat flux (SLHF), but rather by stochastic processes and local atmospheric conditions.

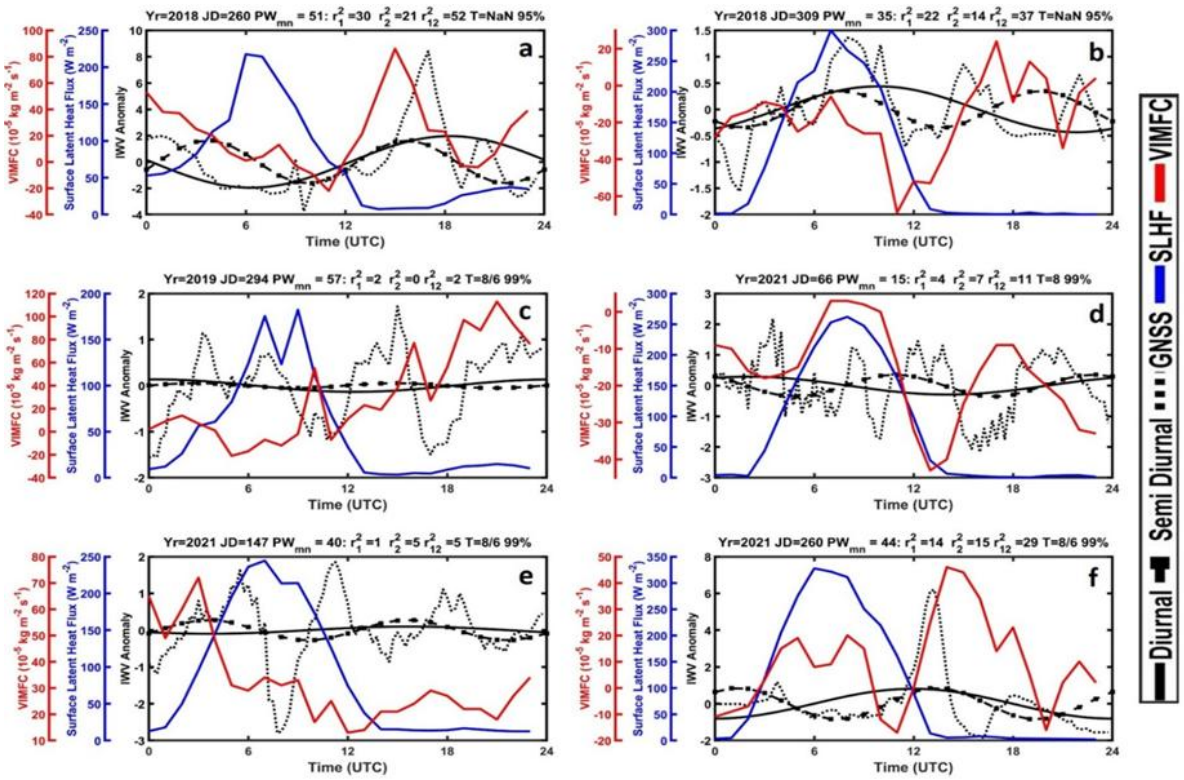

Figure 3.9: *IWV diurnal cycle of individual days for Neither Diurnal nor Semi-diurnal Harmonic is significant. Details on the top of each subplot are Yr is Year, JD is Julian Day, PWmn is Mean IWV of that day, r12 is % contribution of Diurnal harmonic to the diurnal variation of IWV, r22 is % contribution of Semi-Diurnal harmonic to the diurnal variation of IWV, r122 is % contribution of both Diurnal and semi-diurnal harmonics to the diurnal variation of IWV, T is the Time period of significant harmonics and the percentage number at the last is the confidence level at which particular harmonic is significant.*

3.4 Discussion

The diurnal variation in IWV is a complex function influenced by a combination of atmospheric processes, including evaporation, vertical air motion, precipitation, convection, and turbulence. These factors can be categorized into three components: sources, sinks, and transport mechanisms of IWV. Solar radiation is the primary source that drives evaporation from the surface, while the atmospheric wind regime plays a crucial role in transporting water vapor horizontally and vertically within the atmosphere. The contribution of moisture convergence (VIMFC) is particularly important in determining the mean seasonal diurnal variation of IWV. Significant diurnal harmonics are observed throughout all seasons, indicating that the diurnal cycle is primarily driven by solar radiation and related surface processes. However, the impact of synoptic-scale weather systems, including low-pressure areas and frontal systems, can alter moisture

convergence patterns, adding complexity to the diurnal variability. For both diurnal and semidiurnal significant days, the occurrence of IWV maxima is closely linked to VIMFC and precipitation events. Precipitation acts as a sink for atmospheric moisture, leading to a reduction in IWV, while subsequent evaporation and moisture advection can cause a recovery in IWV levels.

The study also underscores the role of mesoscale circulations, such as land-sea breeze interactions and local convection, in modulating IWV variation. During the monsoon and post-monsoon seasons, the enhanced convective activity and frequent occurrence of precipitation events significantly influence IWV. Moisture advection from surrounding regions and the vertical transport of water vapor due to convection contribute to the observed diurnal peaks. These processes are particularly important in tropical semi-arid regions like Gadanki, where the interaction between large-scale and local atmospheric dynamics leads to significant variability in water vapor content. Understanding the interplay between VIMFC, surface latent heat flux (SLHF), and atmospheric dynamics provides valuable insights into the diurnal and seasonal variability of IWV. The detailed analysis presented in this study highlights the importance of considering both large-scale synoptic conditions and local meteorological factors when assessing IWV variability. Such insights are crucial for improving the representation of water vapor processes in weather and climate models, ultimately enhancing the accuracy of weather forecasting, particularly for rainfall and storm events that are highly sensitive to variations in atmospheric moisture.

3.5 Conclusion

This study analyzed the diurnal variation of IWV in the Gadanki region using three years of IWV data. The mean seasonal and day-to-day diurnal variations were subjected to Fischer's significance test to determine the significance of various harmonics. Results indicated that the diurnal harmonic (S1) is significant for 93% of the time, demonstrating the dominant role of solar radiation in driving IWV variability. The occurrence of diurnal cycles with amplitudes between 2-3 mm was found to be the most frequent, with peaks typically observed during the evening to nighttime hours. The semidiurnal harmonic (S2), while generally weaker than the diurnal harmonic, was found to be significant for 36% of the time. The amplitude of semidiurnal cycles varied across seasons, with a maximum of 1.5 mm occurring in winter, monsoon, and post-monsoon, while higher amplitudes of 2 mm were more common in pre-monsoon. Additionally, two distinct types of semidiurnal

cycles were identified, differing in the timing of their peaks, which varied across seasons. The monthly analysis revealed that both diurnal and semidiurnal amplitudes were highest in November, reflecting the combined effects of residual moisture from the monsoon and the onset of winter atmospheric stability.

The study also highlighted the importance of VIMFC in dictating the mean seasonal diurnal variation of IWV. VIMFC plays a crucial role in the timing and magnitude of IWV peaks, with significant correlations observed between IWV maxima and moisture convergence. Precipitation, acting as both a moisture sinks and a modulator of atmospheric moisture content, was also found to influence the phase of IWV cycles, particularly during the recovery phase following rainfall events. The presence of distinct diurnal, semidiurnal, and terdiurnal harmonics underscores the complex interplay between solar-driven processes, atmospheric tides, mesoscale circulations, and synoptic weather systems in regulating IWV variability. Understanding these interactions is essential for improving the representation of water vapor processes in weather and climate models. Enhanced modeling of IWV is critical for improving weather forecasts, especially for rainfall and storm events that are highly sensitive to variations in atmospheric moisture.

Future studies should aim to extend these findings to other regions of India, accounting for diverse climatic conditions and topographical influences. Such research will help refine our understanding of the factors affecting IWV variability and improve predictive capabilities for weather and climate phenomena.

Chapter 4

Prediction of IWV using a machine learning technique

4.1 Introduction

The atmospheric water vapor is one of the fundamental component required for virtually all-weather phenomena occur ring in the troposphere and, therefore, plays a central role in the hydrological cycle, energy transfer through latent heating, and radiation budget (Bevis *et al.* 1992; Holloway and Neelin, 2009; Schiro and Neelin, 2019). The energy released during water vapor condensation also affects the vertical stability of the atmosphere and influences the weather system and associated precipitation patterns (Sherwood *et al.* 2010; Adams and Souza, 2009). Water vapor is a potent greenhouse gas with a strong absorption potential of outgoing longwave radiation and controls the Earth's radiation budget. Recent studies used IWV and its convergence/divergence as a prognostic variable to understand large-scale atmospheric dynamics, mon soon/storm onset, and also in numerical weather prediction (NWP) models (Bretherron *et al.* 2004; Peters and Neelin, 2006; Takiguchi *et al.* 2000; Puviarasan *et al.* 2015).

AI/ML techniques are becoming popular for prediction purposes. Several studies have made successful attempts for the prediction of IWV using different AI-based techniques like ANN, adaptive neuro-fuzzy inference system (ANFIS), support vector machine (SVM), and multilayer perceptron (MLP) considering ground parameters such as temperature, pressure, humidity to train the model (Suparta and Alhasa, 2013; Yue and Ye, 2019). Studies have shown that the ANFIS model shows better performance than MLP models for PWV forecast (Suparta and Alhasa, 2013). Though these techniques offer better predictions, the predicted values are, presently, being used to fill IWV data gaps rather than used to predict IWV in advance in a nowcasting mode. A few attempts have been made to predict IWV using ANN and genetic algorithm 6 and 12 h in advance using IWV and intrinsic mode functions of IWV (Yue and Ye, 2019). The maximum relative error is found to be ∼20% for the 6 h forecast, whereas it is much worse for the 12 h forecast. Even though the studies report reasonably good forecasts, training the ANN with huge data is computationally expensive and time-consuming. Recently, several studies have focused on the nowcasting of storms and rainfall based on IWV characteristics [using data obtained from Global Navigation Satellite System (GNSS) receivers], using preset thresholds or AI techniques (Yue and Ye, 2019; Manandhar *et al.* 2018; Sapucchi *et al.* 2019; Li *et al.* 2020; Benevides *et al.* 2015; Yao *et al.* 2017). These methods are highly successful in predicting storms/rainfall accurately, but all of them suffer from false alarms. Nevertheless, these techniques warrant the knowledge of accurate

IWV with the shortest possible latency, if preferable, in advance (predicted values) for improved nowcasting of storms. Accurate prediction of IWV allows the prediction of a storm in advance and provides more lead time for disaster preparedness. Also, the predicted data can be used to fill the existing data gaps in the IWV time series. Keeping the above aspects in mind, an attempt has been made in this work to develop a fast and reliable model to predict IWV with different lead times (30 min to 2 h) adapting a robust ML technique, gradient boosting machine (GBM). The model is trained using GNSS receiver data and evaluated its output over different synoptic conditions. The sensitivity of the predicted IWV on predictors is also examined to know their importance.

4.2 Methodology

GNSS technique has, now, become a promising means to obtain IWV, as this technique offers IWV with high temporal resolution virtually in all weather conditions (Bevis *et al.* 1992). For this study, two and half years (June 2018–December 2020) of GNSS receiver measurements made at Gadanki (13.4593 N, 79.1684 E) have been used. The detailed methodology, including data preprocessing, feature selection, and model development, has been discussed in Chapter 2. Here, we provide a summary specific to the machine learning approach and the implementation details used for IWV prediction.

4.2.1 LightGBM Approach

In this study, we utilize LightGBM, a Gradient Boosting Machine (GBM) model, to predict IWV. Compared to conventional Gradient Boosting Decision Trees (GBDT), LightGBM employs a gradient-based one-side sampling (GOSS) approach to speed up the training process while maintaining the accuracy. This makes it highly suitable for fast, memory-efficient IWV prediction with high temporal resolution. Figure 4.1 presents a systematic diagram illustrating the data preprocessing and feature engineering steps, followed by splitting the data into training, testing, and validation sets, conducting hyperparameter tuning, and ultimately predicting the desired parameter.

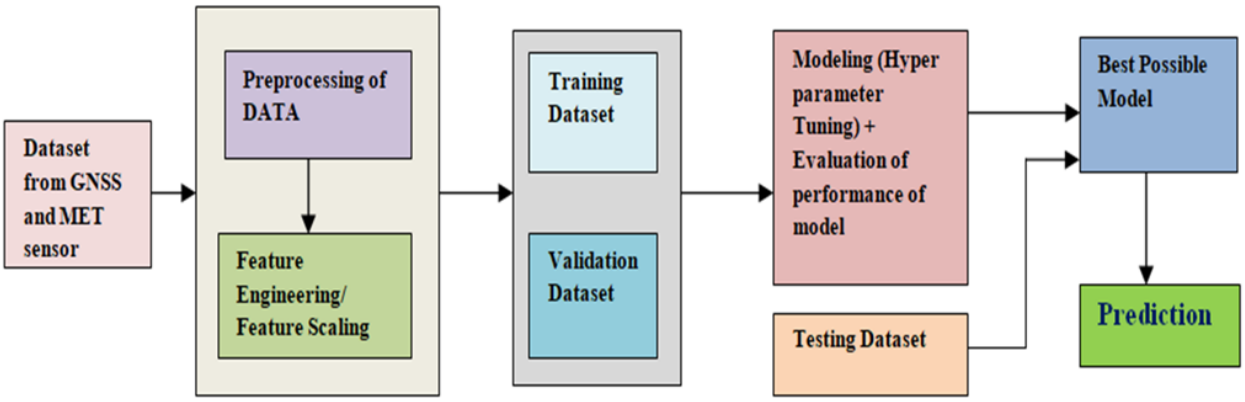

Figure 4.1: *Systematic Diagram of prediction using LightGBM technique.*

4.2.2 Model Setup and Training

Selection of parameters (predictors) is very crucial, which helps the model to predict the predictand accurately. Variation in any parameter, including IWV, depends on source, sink and transport terms. For IWV prediction, air temperature (indicator for evaporation), specific humidity (SH), soil moisture (SM) (for latent heat flux), rainfall (removes water vapor from the atmosphere and acts as a sink), and horizontal vertically integrated moisture flux (hereafter referred to as moisture flux, which represents the horizontal transport of moisture and is estimated from ERA5 reanalysis data, which was later interpolated to 30 min. resolution) are the ideal parameters. In addition, radiation (short and longwave) and soil temperature measurements are also related to moisture through evaporation of water (from water bodies) and soil moisture.

Ideally, all the parameters should be independent of each other and should have a link with IWV. The selected parameters are, therefore, correlated with each other to know their interdependency (Figure 4.2). As expected, the soil temperature and shortwave radiation show strong correlation with air temperature and, therefore, they are excluded from the analysis. Also, the longwave radiation is also excluded as it does not show a strong correlation with IWV. Moreover, the study tries to develop the prediction model with easily available data sets, like GNSS receiver and AWS measurements. Finally, these independent parameters (air temperature, specific humidity, soil moisture, moisture flux and rain rate) have been given to the model to train it for prediction.

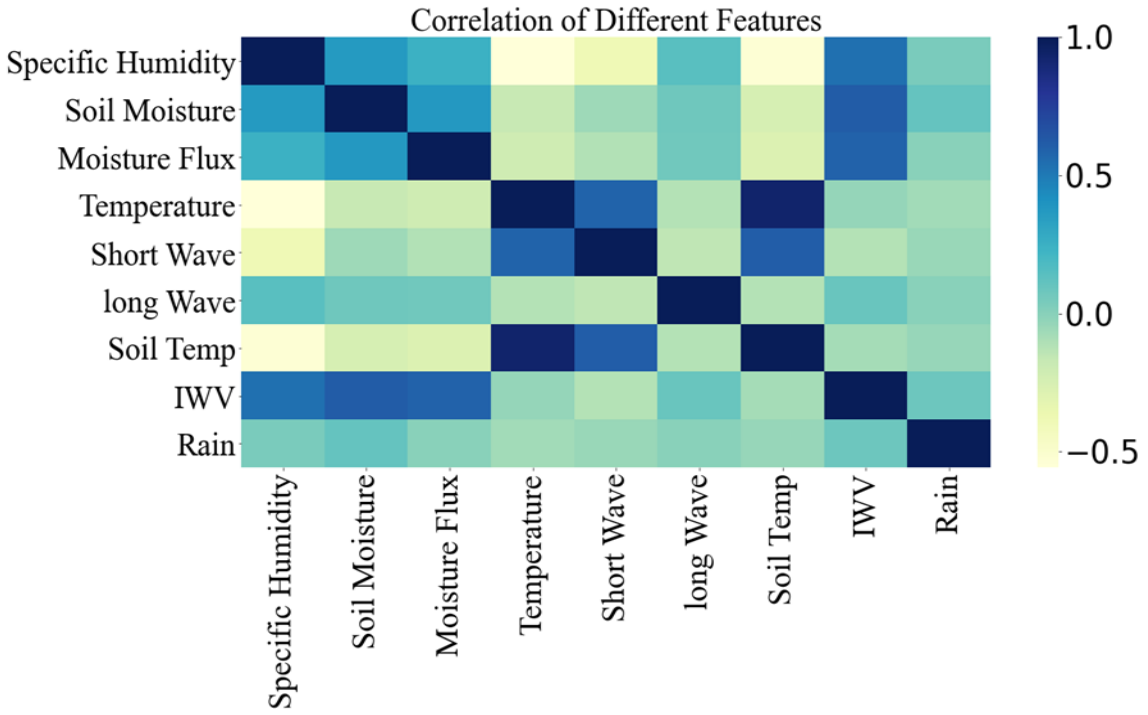

Figure 4.2: *Correlation coefficient between different parameters considered for the analysis.*

The LightGBM model was trained using two and a half years of GNSS-derived IWV data, spanning from June 2018 to December 2020. The dataset was divided into two parts: June 2018 to December 2019 for model training and 2020 for validation. Training the model involved selecting optimal hyperparameters through iterative tuning, using a combination of early stopping and grid search to avoid overfitting and ensure robust performance.

Besides physical parameters, model also requires several features in the predictors for the prediction of IWV. For instance, IWV follows a diurnal pattern and understanding this pattern makes the prediction easy in normal conditions. However, the diurnal pattern is often disturbed by the occurrence of storms, during which the IWV increases dramatically. Similarly, gradients and double gradients in selected parameters are useful for the short-term prediction of IWV. It is also imperative to assess the importance of each of these parameters/features in the prediction of IWV. The number of splits of features in lightGBM processing (information gain) indicates the importance of the parameter/feature. Figure 4.3 presents the feature importance (in terms of %, estimated from the number of times the feature is used for prediction compared to a total number of splits) in the prediction of IWV with different lead times. For instance, Sinusoidal IWV has the

highest splits for all time horizons and therefore is the most important feature for the prediction of IWV. Also, the feature importance varies with predictions with different lead times. The feature importance of meteorological parameters, like temperature, specific humidity, moisture flux and soil moisture, increases with prediction with increased lead times. On the other hand, the gradients of parameters and sinusoidal patterns are more important for short term prediction (30 min). Among all parameters/features, rainfall shows the lowest value of feature importance.

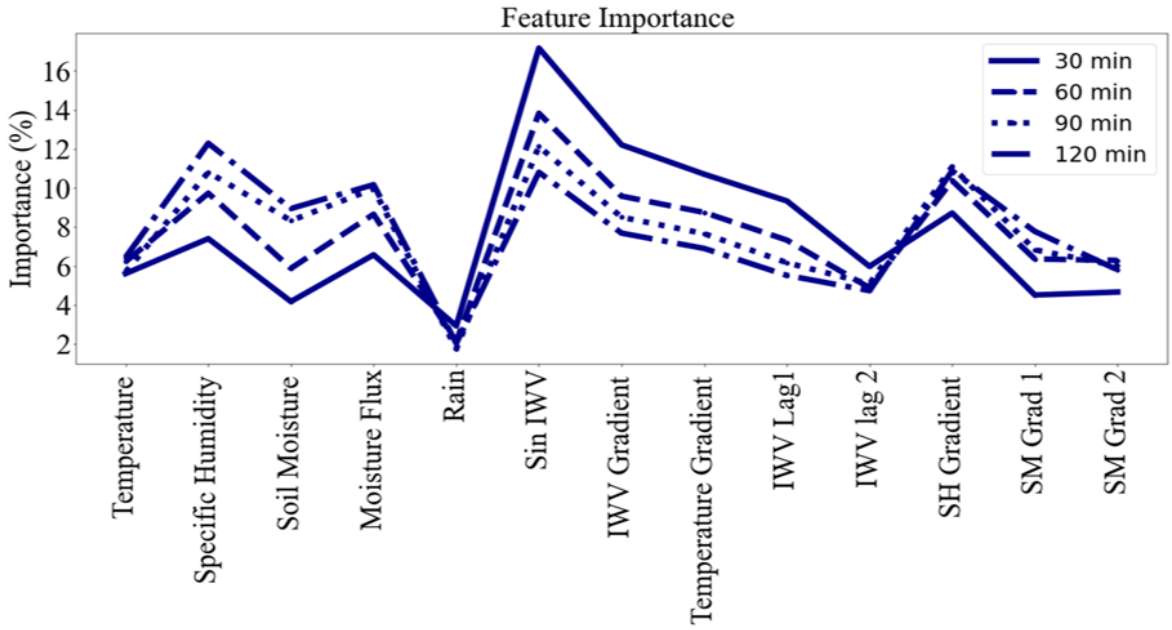

Figure 4.3: *Feature importance of selected meteorological parameters and their patterns (as obtained by the model) for predictions with different lead times.*

4.3 Results and Discussion

4.3.1 Model Performance

In the present study, a machine learning-based LightGBM model was developed to predict IWV for different lead times. To evaluate the model's performance, data from April 2019 and June 2019 were used for testing, representing typical dry and wet months for the Gadanki region. The model was implemented with different lead times ranging from 30 minutes to 2 hours, with 30-minute intervals. The uncertainty in the model is presented in terms of daily RMSE values, as shown in Figure 4.4. The RMSE values for different lead times are referred to as $RMSE_{30}$, $RMSE_{60}$, $RMSE_{90}$, and $RMSE_{120}$, with the subscript representing the prediction lead time.

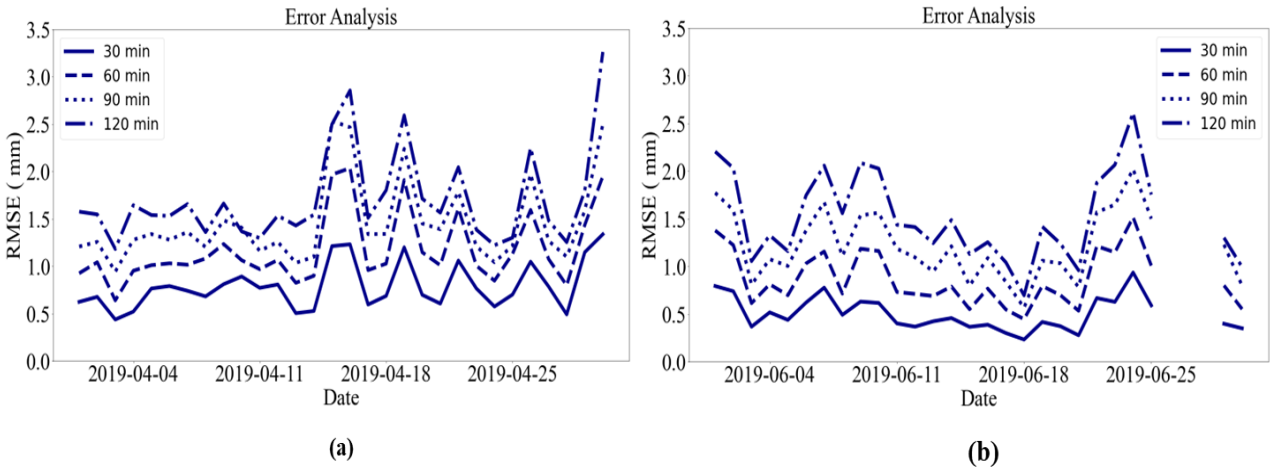

Figure 4.4: *RMSE values for predicted IWV with different lead times, indicating the testing of the model performance during an (a) dry season (April 2019) and (b) rainy season (June 2019). The gap in (b) is due to the data gap.*

The daily $RMSE_{30}$ values for the selected two months were consistently less than 1 mm, indicating that the prediction was highly accurate for a 30-minute lead time. As the lead time increased, the RMSE values also increased, with maximum values reaching up to 3 mm for a 120-minute lead time prediction. However, about 80% of the time, the RMSE was less than 2 mm, even for 120-minute lead times. This demonstrates the robustness of the LightGBM model in accurately capturing the variations in IWV with minimal error across different lead times. Upon a detailed examination of the days with larger RMSE values, it was observed that these large values were not always associated with rainy periods. This finding highlights the influence of other factors, such as atmospheric turbulence and convective activities, in causing deviations in IWV predictions.

The model-predicted IWV values are validated using the data from 2020. The data are first segregated into different seasons: winter (January and February), pre-monsoon (March–May), monsoon (June–September), and post-monsoon (October–December), following the India Meteorological Department. It facilitates the assessment of the model performance in different seasons. Figure 4.5 presents a comparison between the predicted IWV with a 30-minute lead time and the reference GNSS-derived IWV values for different seasons. The color indicates the percentage occurrence of IWV at that value. Linear regression fit is performed, and the derived statistics are also included in Fig. 4.5. The correlation coefficient was found to be 0.99 across all seasons, indicating excellent agreement between predicted and observed IWV values. The majority of the data points lie along the diagonal in all seasons, except during the monsoon season, where

some scatter can be observed, likely due to the presence of long-lasting storms with multiple convective cells. The bias during the monsoon season was found to be 2.39 mm, whereas the bias in other seasons was much lower, ranging between 0.5 mm and 0.64 mm. A slightly larger bias and the scatter in the data during the monsoon could be due to the higher occurrence of storms, particularly the long-lived storms with multiple convective cells. A slightly larger bias and scatter during the monsoon season could be due to the higher occurrence of storms, particularly the long-lived storms with multiple convective cells.

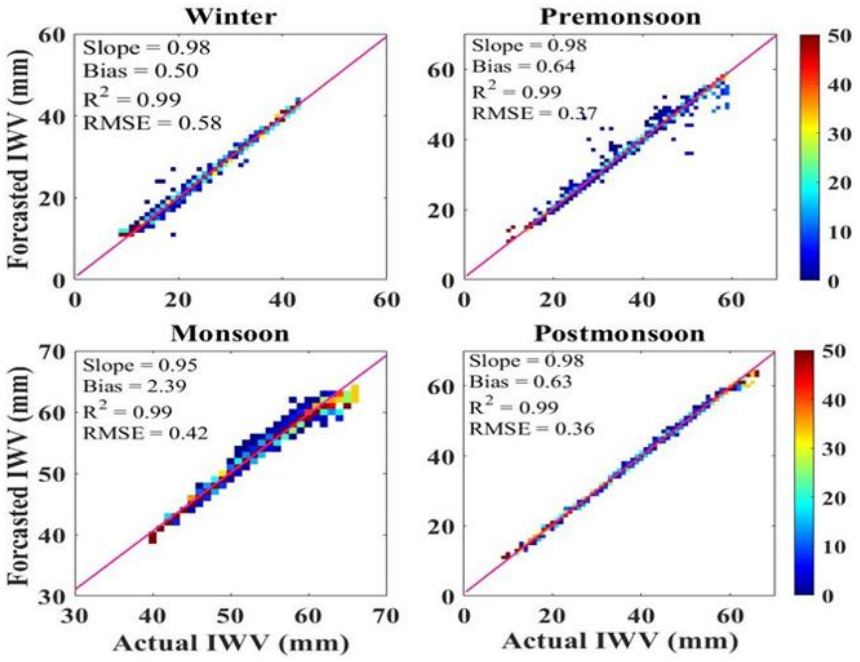

Figure 4.5: *Comparison of model predicted IWV with that measured by GNSS receiver in different seasons. Color bar indicates the percentage of occurrence of data points. Linear regression fit (red line) and parameters of the fit (slope, intercept (bias), RMSE and R2) are also shown in the figure.*

The validation exercise has been repeated for different time scales, like 60, 90, and 120 min, to check the capability of the model. Table 4.1 shows the RMSE values with different lead times during different seasons. Clearly, the model performance is outstanding with RMSE values ≤ 2 mm (except for monsoon with 120 min lead time). The RMSE values are less than 1 mm for predictions up to 60 min in advance. As the lead time increases, the error (RMSE) in the predictability of the model also increases but remains within the tolerable limit (less than the uncertainty obtained in the testing data).

Table 4.1: *RMSE values obtained from predicted IWV with different lead times during different seasons.*

Season	$RMSE_{30}$	$RMSE_{60}$	$RMSE_{90}$	$RMSE_{120}$
Winter	0.58	0.64	1.23	1.85
Premonsoon	0.37	0.75	1.11	1.75
Monsoon	0.42	0.81	1.36	2.02
Postmonsoon	0.36	0.51	1.03	1.63

4.3.2 Sensitivity Analysis

Two types of sensitivity analyses have been performed to examine the dependency on predictors. Presently, moisture flux is estimated from reanalysis data, which typically will be available to users with some latency. However, for real time applications, one should depend on numerical forecasting model outputs, which may have some errors. So, to examine the impact of moisture flux on the prediction, the IWV prediction was made by varying reanalysis moisture flux values by ±10% and ±20%, and each time RMSE values are estimated with predicted values. The RMSE value with a 20% variation of moisture flux from reanalysis data is 0.69 mm (0.05 mm variation from the original value). It indicates that one can use NWP models out of moisture flux (as long as the error is within 20%) to predict IWV.

Table 4.2: *Sensitivity analysis on reanalysis moisture flux data.*

Type	Reanalysis (mm)	+10% (mm)	+20% (mm)	-10% (mm)	-20% (mm)
RMSE	0.64	0.69	0.69	0.69	0.69

The sensitivity of IWV prediction on other predictors is also tested by excluding them (temperature and humidity, SM, moisture flux, and rain). Clearly, the exclusion of any of the parameters (apart from IWV) does not change the RMSE considerably. It also indicates that IWV is the most

important variable (predictor). Note that the present analysis includes all the data, including storm days. Some of these predictors are important for storm days as they contain crucial information (temperature drop, IWV reduction after rain, etc.) for IWV prediction on storm days.

Table 2.3: *RMSE values obtained from predicted IWV with different predictors, highlighting the sensitivity of various predictors.*

Time/Type	$RMSE_{30}$	$RMSE_{60}$	$RMSE_{90}$	$RMSE_{120}$
All	0.63	0.93	1.16	1.37
Except Temperature and Humidity	0.54	0.89	1.07	1.27
Except Soil moisture	0.66	0.93	1.17	1.40
Except moisture flux	0.65	0.89	1.12	1.36
Except rain	0.65	0.93	1.14	1.38

4.3.3 Storm Event Analysis

Past studies have shown that the variation of IWV will be quite different from that of simple diurnal variation. The storm-associated convective motions, large-scale convergence, and related processes increase IWV dramatically before the occurrence of precipitation. This increase in IWV is used as one of the predictors for storm/rain occurrence in several studies (Yue and Ye, 2019; Manandhar *et al.* 2018; Sapucchi *et al.* 2019; Li *et al.* 2020; Benevides *et al.* 2015; Yao *et al.* 2017). The applicability of the LightGBM for storm days has been tested by examining whether or not the predicted IWV shows such an increase. Fig. 4.6 shows predicted IWV with different lead times along with rainfall (30 min accumulated) and GNSS-derived IWV (for reference), highlighting the chosen ML model's capability on a storm day (July 15, 2019). Fig. 4.6 clearly shows some interesting features. The IWV variation is found to be complex during the storm period (14:30–19:30 UT), with a large initial rise, subsequent reduction, and then an increase, corresponding well with rainfall variation. The signature of IWV jump or buildup starts 2 h before the rain begins and the peak IWV is observed 30 min before the rainfall peak. The model-predicted IWV is able to capture this variability in a well-defined way for all lead times. Although the model is able to capture the IWV jump before the storm, small deviations from the GNSS derived IWV

are seen during subsequent variations. These variations are, probably, responsible for the scatter (albeit small) during the monsoon period in Fig. 4.5.

The model's ability to accurately capture these complex variations, even with longer lead times, highlights its potential for operational use in nowcasting and weather prediction applications. This capability is particularly beneficial for mitigating the impacts of extreme weather events, as it provides critical lead time for emergency preparedness and response efforts, ultimately contributing to reducing the risk posed by sudden and severe weather conditions.

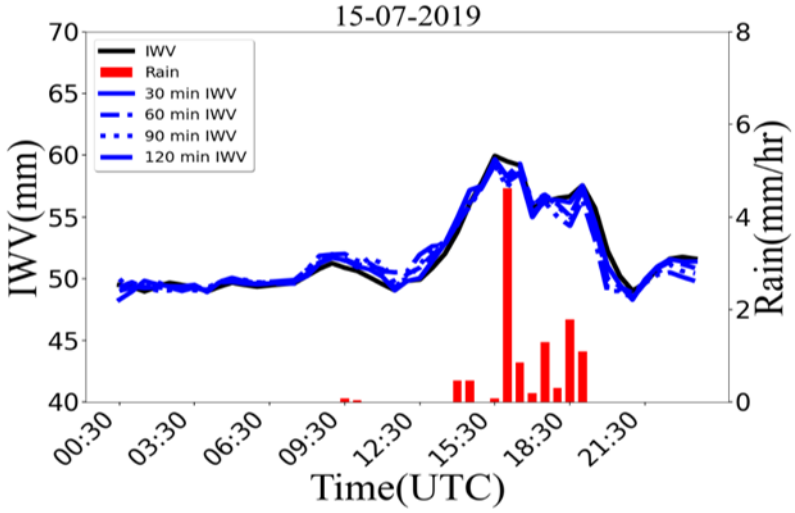

Figure 4.6: *Temporal variation of predicted (different lead times) and retrieved IWV on a storm day (15 July 2019). Figure also shows the AWS-measured rainfall.*

Although the algorithm was developed with post-processed data, its performance has also been tested with real-time data to check its applicability for real-time applications (not shown here). The performance is found to be nearly equal to that of post-processed data (R^2 = 0.90, RMSE 2.17 mm for 30 min prediction).

4.4 Conclusion

The application of machine learning techniques, particularly LightGBM, for predicting IWV has demonstrated highly promising results, providing accurate predictions across a variety of weather conditions and lead times. The model's effectiveness in predicting IWV during both regular and stormy conditions showcases its potential for operational use in nowcasting, which is critical for early warnings, disaster preparedness, and emergency response. The ability to predict sudden

changes in atmospheric moisture content, as seen during storm events, makes this approach particularly valuable for mitigating the impacts of extreme weather.

The feature importance and sensitivity analyses offer valuable insights into the dynamics of the different predictors. Sinusoidal IWV patterns emerged as the most influential feature for short-term predictions, emphasizing the importance of accurately capturing the diurnal cycle of IWV. Short-term meteorological gradients were also crucial, especially for immediate forecasts, while variables related to moisture flux and transport processes gained importance for medium-term predictions, particularly during storm conditions.

Future work will focus on further refining the model for real-time applications, addressing computational efficiency, and expanding its applicability to different geographic regions with varying climatic conditions. The aim is to enhance the overall accuracy and reliability of nowcasting, ultimately contributing to more effective early warning systems and disaster risk reduction.

Chapter 5

Nowcasting of storms using predicted IWV and brightness temperature

5.1 Introduction

Storms are life-threatening weather events with huge damage potential to property as they are accompanied by intense rain, flash flooding, hail, lightning, violent winds, and even tornadoes. Recent studies have shown that these storms can be nowcasted (1–6 h in advance) with reasonable accuracy with the help of rapidly updated high temporal resolution datasets obtained with AWS (Mostajabi *et al.* 2019), Doppler weather radars, satellites (cloud maps with infrared channels), lightning networks, and multiwave length radiometers (Madhulatha *et al.* 2013), GNSS networks (Benevides *et al.* 2015; Yao *et al.* 2017; Manandhar *et al.* 2018). In recent years, blending techniques employing different datasets are also being developed for improved nowcasting (Chkeir *et al.* 2023). Nowcasting significantly reduces fatalities and property damage and also plays a key role in crisis management.

The GNSS-based nowcasting techniques are increasingly becoming popular due to the availability of high resolution, low latency, and round-the-clock data of IWV (Bevis *et al.* 1992; Benevides *et al.* 2015; Yao *et al.* 2017; Manandhar *et al.* 2018; Martinez *et al.* 2021). Furthermore, the GNSS receiver networks are being established at a rapid phase in all the countries, providing a unique opportunity to map water vapor and its variability on a global scale. It is known that water vapor is the basic ingredient for convective storms/rain (Neelin *et al.* 2009), but their relation is complex and varies from land to ocean and also between different oceanic regions (Bretherton *et al.* 2004; Ahmed and Schumacher, 2017). At a local scale, moisture builds up rapidly before the occurrence of rain fall/storm and this signature has been used in several studies to nowcast rainfall/storm (Benevides *et al.* 2015; Yao *et al.* 2017; Manandhar *et al.* 2018; Sapucci *et al.* 2019; Zhao *et al.* 2020; Guerova *et al.* 2022; Chkeir *et al.* 2023).

The GNSS-based nowcasting models typically use a thresh old magnitude and/or gradient of IWV for identifying storm/rain occurrence (Benevides *et al.* 2015; Yao *et al.* 2017; Manandhar *et al.* 2018). Some of the studies used hybrid models, wherein GNSS data are coupled with meteorological data or radar observations or lightning data to predict the storm (Benevides *et al.* 2015; Yao *et al.* 2017; Chkeir *et al.* 2023) noted a positive correlation between the large gradients in IWV and rainfall occurrence and therefore devised a method to nowcast rainfall based on IWV increment. Although the model's predictability is good (75%), false alarm is also high at 40%–65%, depending on the threshold IWV gradient. Yao *et al.* (2017) used three parameters of IWV:

magnitude, variation, and rate, as indicators for nowcasting rainfall. The inclusion of other IWV parameters has increased the detection rate by 7%; however, the false alarm rate is nearly remained the same (66%). Subsequent studies also used IWV increment and IWV slope (Zhao et al., 2019) or magnitude of IWV and rate of IWV (Manandhar *et al.* 2018) or magnitude of IWV, IWV increment, and rate of IWV (Li *et al.* 2020) with different thresholds to improve the predictability of storms/rainfall and also to reduce the false alarm rate. Zhao *et al.* (2020) improved the rainfall forecasting model with the help of five predictors, monthly IWV, seasonal IWV, ZTD fluctuations, and their first derivatives. The model, indeed, increased the detection rate to 95% and reduced the false alarm rate to less than 30%. To reduce the false alarm rate, researchers focused on blending techniques by employing data from a variety of instruments, such as AWS, GNSS receivers, infrared channels of satellites, and DWRs. The model developed by Benevides *et al.* (2019) employing satellite-derived cloud top temperature and pressure along with IWV, air pressure, relative humidity, surface air pressure, and temperature to predict rainfall events had a success rate of 64% and a false alarm rate of ∼22%. Guerova *et al.* (2022) developed a "convective storm demonstrator" using IWV and instability indices. The implementation of this algorithm on the data collected at 12 locations in Bulgaria reveals that the predictability (false alarm rate) of the model is 83% (38%). It can be inferred from the above studies that IWV-derived parameters can be used as primary predictors of storm events, while a well-chosen secondary parameter reduces the false alarms.

Recently, AI/ML techniques have become popular in geosciences, including GNSS-based prediction of IWV (Suparta and Alhasa, 2016) and nowcasting applications (Sangiorgio *et al.* 2020; Los *et al.* 2020; Liu *et al.* 2022; Chkeir *et al.* 2023). Some of these studies used zenith tropospheric delay (ZTD) to predict extreme events (Sangiorgio *et al.* 2020). The usage of predicted IWV allows nowcasting of storms/rainfall in advance, providing more time for disaster managers.

As seen above, some of the studies are qualitative and others are based on simple gradient methods quantifying the changes in water vapor. However, the nowcasting of storms by the above approaches provides a very short lead time for disaster preparedness, 5–30 min. Furthermore, false alarm is found to be large with all models/algorithms. In the present study, a hybrid model is developed to nowcast storms with different lead times (up to 2 h) by blending predicted IWV by

employing an AI/ML technique with GNSS receiver and AWS data as inputs with satellite measurements. This model combines the strengths of earlier techniques and provides a robust model to improve the detection capability of storms with reduced false alarms. Furthermore, early prediction of storms will provide more lead time for disaster preparedness.

5.2 Data and Methodology

5.2.1 Data

The GNSS receiver measurements obtained during September 2018–December 2020 have been used for the present study. In addition, other meteorological parameters, such as surface temperature, surface pressure, wind speed and direction, and accumulated rainfall obtained from AWS, are used. Near real-time brightness temperature (T_b) retrieved from Indian National Satellite (INSAT-3DR) at 30-min resolution is also used to monitor cloud growth and decay. An X-band dual-polarization Doppler weather radar has been used to detect the storms and their movement in and around Gadanki. The radar is capable of providing such information up to a range of 80 km.

The detailed methodology of IWV retrieval from GNSS measurements has been discussed in Chapter 2. Therefore, an overview of the approach specific to storm nowcasting is only presented below.

5.2.2 Methodology

The nowcasting technique proposed in the present study is a two-step process: 1) prediction of IWV using an ML technique (light GBM) and 2) fixing thresholds to chosen predictors for nowcasting storms.

5.2.2.1 IWV Prediction

The first step involves predicting IWV using a machine learning approach (LightGBM), which has been discussed in detail in Chapter 4. IWV prediction plays a vital role in identifying the moisture buildup associated with storm development. Using GNSS receiver data, the LightGBM model predicts IWV values with different lead times ranging from 30 to 120 minutes. It is shown in Chapter 4 that the model was able to predict IWV with good accuracy up to 2 h with seasonal RMSE $\leq$ 2 mm.

5.2.2.2 Threshold Calibration for Nowcasting

The second step involves setting thresholds for chosen predictors to identify storms. Three predictors are considered in addition to the magnitude of IWV: changes in IWV magnitude, temporal gradients in IWV, and changes in T_b. The thresholds are calibrated using historical storm data, ensuring that the model can effectively distinguish between storm and non-storm events. The chosen thresholds are optimized to minimize false alarms while retaining high sensitivity to storm occurrences.

5.2.3 Evaluation Metrics

To evaluate the performance of the hybrid storm nowcasting model, the following metrics have been used: Probability of Detection (POD), False Alarm Rate (FAR), Miss Rate (MR), Model Accuracy (MA). All these metrics are defined and discussed in Chapter 2 in detailed way.

5.3 Results and Discussion

5.3.1 Evaluation of predicted IWV for storm cases

As seen in Chapter 4, the LightGBM performed well in predicting IWV with RMSE less than 2 mm in all seasons. The same technique is validated here on data collected during several storm events occurred in Gadanki region. The evaluation included a diverse set of storm cases, such as isolated thunderstorm events, mesoscale convective systems, and long-duration convective storms, to thoroughly test the robustness of the hybrid model. The first stage of the evaluation involved predicting IWV up to a lead time of 120 minutes. The predicted IWV values were compared with observed data for various storm events. The validity of the algorithm has been tested here for various storm cases [two isolated thunderstorm events (July 15 and August 7, 2020) occurring at different times of the day and one long event (July 8, 2020) with multiple cells of convection] at different lead times (30–120 min) in Fig. 5.1. The increase in IWV is seen in almost all events, while it is concomitant with the rainfall in some cases and observed much before the event in other cases. The predicted IWV could accurately reproduce the observed temporal variability of IWV for all three events and for all lead times. The model can predict even the storm-associated increase in IWV accurately. Therefore, the predictive model will be quite effective in nowcasting the storms, as it is possible to obtain IWV and monitor its variation before the storm occurrence. The RMSE in forecast IWV for all lead times is < 2 mm. The model prediction of IWV is compared

with the observed IWV for all the storm cases, and the statistics of correlation analysis for different lead times are shown in Table 5.1. A very good correlation is found with a correlation coefficient ≥ 0.85 and RMSE < 1.2 mm up to 120-min prediction.

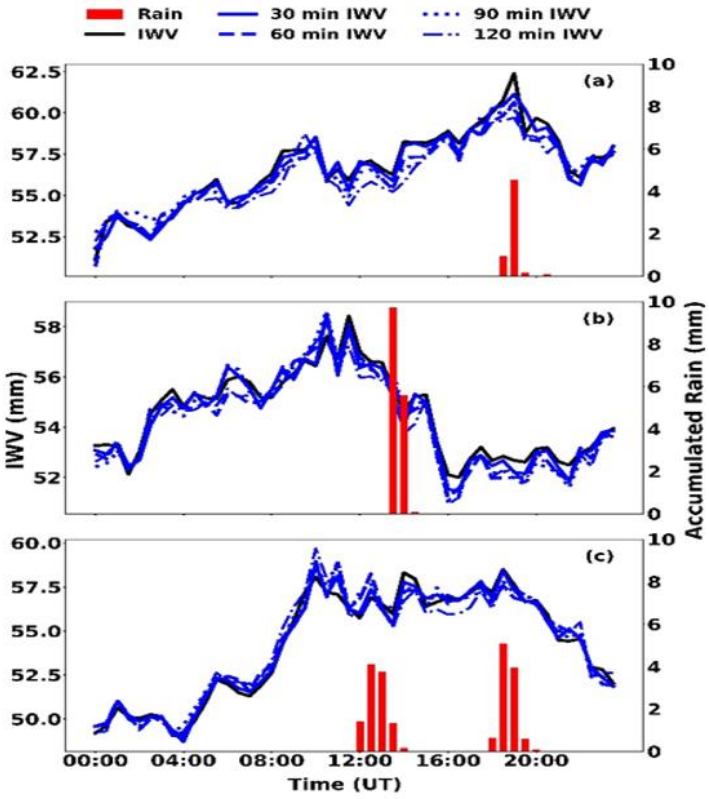

Figure 5.1: *Predicted IWV with light GBM technique along with GNSS-derived IWV and rainfall on 3 storm days, (a) 15 July 2020, (b) 07 August 2020 and (c) 8 July 2020.*

Table 5.1: *Statistics of correlation analysis for storm days, depicting the accuracy of model's prediction of IWV for different lead times. R^2 and MAE indicate correlation coefficient and mean absolute error, respectively.*

Statistical Metric	30 min	60 min	90 min	120 min
R^2	0.96	0.93	0.90	0.85
RMSE	0.42	0.72	0.91	1.17
MAE	0.36	0.50	0.66	0.84

5.3.2 Determining Optimal Thresholds

To identify storms unambiguously with an automated algorithm using moisture buildup as a criterion, several storms need to be examined to arrive at thresholds for predictor(s). For this purpose, storms passing close to Gadanki location (within 1 km) have been detected using measurements from the X-band dual-polarization radar at Gadanki.

Figure 5.2 shows a typical example of storm occurrence (on 13 July 2019) as seen by the X-band radar, automatic weather station (AWS), INSAT-derived Tb and GNSS receiver. The spatial variation of reflectivity factor at horizontal polarization (Z_H) at 8° elevation around Gadanki (10 km X 10 km), as observed by X-band polarimetric radar (Figure 5.2a), clearly shows convective cells ($Z_H \geq 40$ dBZ) embedded in a large precipitating system. The cells on the southwest side passed over the study site and produced a good amount of rainfall at Gadanki. The AWS measurements clearly show copious rainfall (about 45 mm in 5 hours) starting from 15:30 UTC at Gadanki. The AWS measurements logged at 1 sec. resolution are averaged over 30 min. (accumulated for rainfall) to match with GNSS receiver measurements. Sudden drop in temperature and rapid increase in wind speed can also be seen at the time of rainfall. The GNSS-derived IWV started increasing at 12 UTC, reached a maximum at the time of peak rainfall, and decreased rapidly after the rainfall. The dramatic increase in IWV is much more than the climatological daily diurnal variations (dashed line), indicating that the increase is associated with the storm. The increase in IWV started ~4 hours before the rain (storm) occurrence at Gadanki and the increase is about 8 mm. It is clear from the Figure 5.2b that the IWV increase is an excellent predictor for the occurrence of storm, as the increase is seen well ahead of the storm.

In addition to the above parameters, Benevides *et al.* (2019) suggested the use of gradient in T_b to reduce the false alarm rate. Dramatic change in T_b can be clearly seen (Figure 5.2b) prior to and during the occurrence of storm/rainfall. The T_b value has reduced by ~ 90 K in 3 hours. Even prior to the storm, the reduction in T_b was significant, indicating that it can also act as good predictor. Figure 2b clearly shows the signature of storm in various surface meteorological parameters and T_b, i.e., reduction in temperature and T_b, increase in wind speed, high IWV and temporal changes in IWV. To identify and predict the storms, one should use proper thresholds to the above parameters. Larger thresholds will only identify intense storms, while smaller thresholds will increase false alarm rates. In other words, different thresholds provide different prediction

outcomes; therefore, determining an ideal threshold value for high prediction capacity and least false alarm rate is essential. Data from September 2018 to December 2019 are used to identify storms and the data corresponding to chosen predictors during the passage of these storms are used to generate the threshold values for predictors. It includes all types of convective storms/cells, including 33 mild ($40 \leq Z_H < 45$ dBZ), 17 intense ($45 \leq Z_H < 50$ dBZ), and 15 extreme ($Z_H > 50$ dBZ) storms. The number of storms is large during the monsoon season (65), allowing to obtain robust thresholds for predictors. Winter and premonsoon seasons have been not considered as there were not enough storm occurrences to meaningfully derive thresholds for predictors.

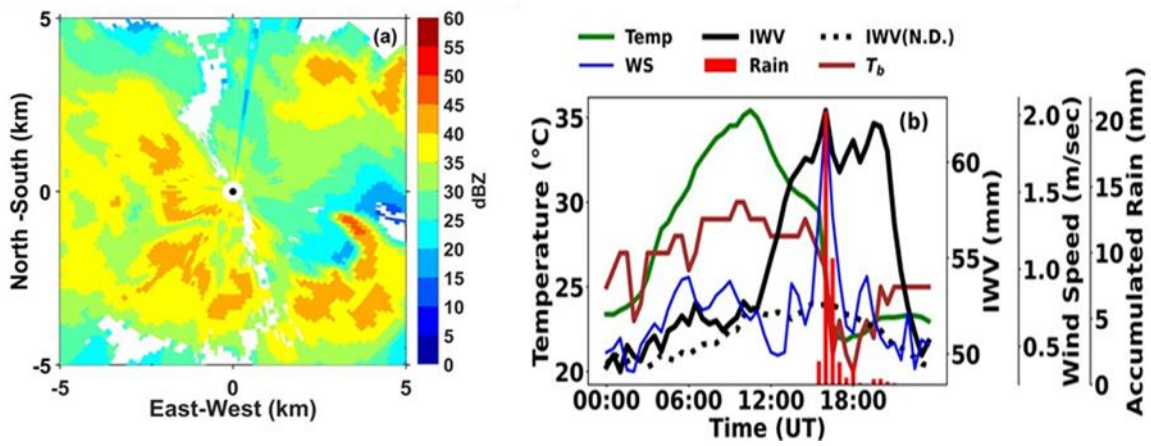

Figure 5.2: *(a) Plane Position Indicator (PPI) at 8° elevation obtained with an X-band dual polarization radar at Gadanki, showing embedded convective cells in the reflectivity field around Gadanki. (b) Temporal variation of meteorological parameters, temperature and wind speed obtained with an AWS, GNSS-based IWV and INSAT-derived T_b.*

Histograms of the maximum value of IWV, temporal gradients of IWV, temperature and wind speed and Tb, and the time at which the IWV starts to increase (additional IWV) relative to the time of storm (rainfall) (Δt) occurrence are plotted using the data from storm days, identified as per the procedure outlined above. The magnitude of gradients in temperature are found to be different for day and night, smaller during the night (peak is at 3 K) than in the day (peak is at 5 K). However, it is found from many case studies that the rapid change in temperature and wind speed occurred at the time of event, while the gradients are found to be weaker prior to the storm/rainfall occurrence. They may be good to do nowcasting with a lead time of 30 min, but may not be suitable for longer lead times. Histograms of remaining parameters, magnitude of IWV, temporal gradients in IWV and T_b and Δt are shown in Figure 5.3.

The IWV is generally high during the monsoon season and additional IWV associated with storms increases the IWV further. The histogram of maximum IWV on storm days shows values larger than 50 mm with a peak in the range of 58-62 mm (Figure 5.3(a)). Not only IWV values, but the rate of increase of IWV is also high on the storm days with all the cases showing a rate ≥ 1.54 mm hr^{-1}. Rate of IWV values as large as 5-9 mm hr^{-1} were also observed, nevertheless, their occurrence is found to be less (Figure 5.3(b)). However, the histogram shows that on 50% of the days, the rate of IWV values is nearly equal to 2 mm hr-1 (1-5-2.5 mm hr^{-1} bin). The storms can also be recognized by the rapid growth of clouds. The histogram of temporal change in T_b shows values ≥ 5 K hr^{-1} (Figure 5.3(c)). However, the cloud growth is much larger in many of the cases (~70% occurrence). From Figure 5.3(d) it is evident that in most of the cases, IWV enhancement can typically be seen two hours before a storm, indicating that the majority of the cases can be identified two hours before the storm. However, some storms show an increase in IWV 1 hr before the occurrence of storm/rainfall. These details, particularly the lowest values in the histograms (representatives of weakest storms), are used to arrive at thresholds to identify the storms. Therefore, the thresholds of IWV, rate of IWV, and rate of T_b are finalized as 50 mm, 1.56 mm hr^{-1}, and 5 K km^{-1}, respectively.

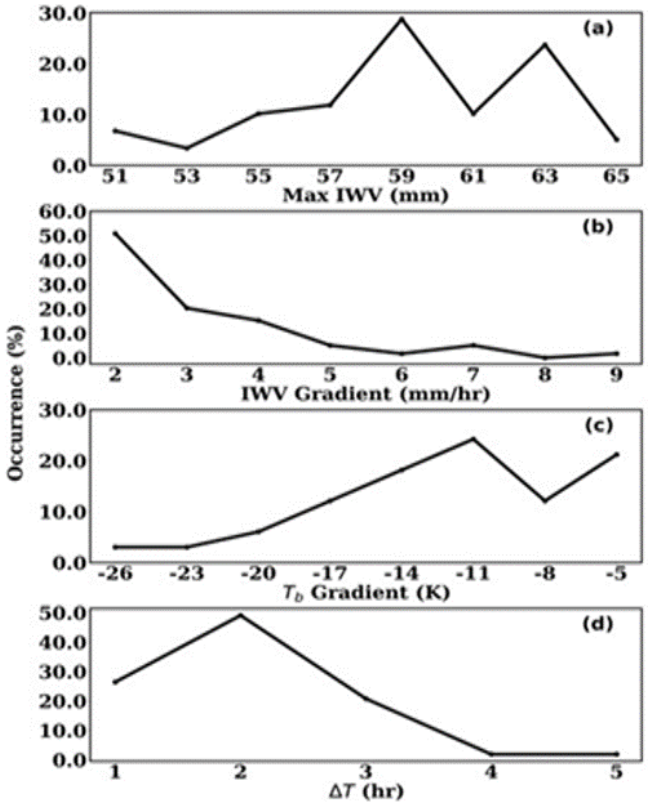

Figure 5.3: *Histograms of (a) maximum IWV, (b) IWV gradient, (c) Tb gradient, and (d) generated from observations during 65 storm days.*

The above thresholds are compared and contrasted with those available in the literature at different geographical locations. Except for T_b, which was used only by Benevides *et al.* (2019), other parameters are extensively used by several researchers for identifying/predicting storms/rainfall (Benevides *et al.* 2015; Yao *et al.* 2017; Sapucci *et al.* 2019; Zhao *et al.* 2020). Yao *et al.* (2017) examined several rain events in Zhejiang province in the subtropical monsoon-affected zone in China and noted maximum IWV and rate of IWV are in the range of 51.1 mm - 73.9 mm and 0.93-7 mm hr^{-1}, respectively. The threshold IWV value used in the present study, i.e., 50 mm of maximum IWV, is very close to the lowest range noted by Yao *et al*. (2017). Also, the rate of IWV in the present study is within the range observed by Yao *et al.* (2017). Sapucci *et al.* (2019) also reported a rate of 2 mm hr^{-1} IWV before intense rainfall in CHUVA vale campaign in Brazil. Further, they noted that the IWV peaks 16-96 min. before the maximum rainfall and used the average rate of IWV to nowcast storms. Benevides *et al.* (2015) also noted the change in IWV around 1.4 mm hr^{-1}, which is very close to the value used in the present study, for intense rain events in Portugal. However, some of the earlier studies have used much smaller thresholds for

IWV gradient (for ex., 0.6 by Yao *et al.* 2017) for identifying light rain. Zhao *et al.* (2020) also used a smaller threshold for IWV gradient for their rainfall detection algorithm in China. All these studies considered the increase in IWV from smallest to largest values of IWV before the storm and then normalized by the time to obtain hourly values. As shown in Figure 5.2b, the minimum to maximum IWV includes the diurnal variability, which is a slowly varying component and increases the time between the minimum and maximum IWV, thereby reducing the IWV gradient. However, IWV above the climatological diurnal variation is associated with the storm in the present study (13 July 2019). This increase in IWV is quite rapid and the gradient is also large as seen in the case study. Also, given the large diurnal amplitudes of IWV (of the order of >2 mm), smaller gradients may increase the false alarm. Therefore, a larger gradient in IWV is used in the present study.

Considering the large seasonal variability in IWV, earlier studies suggested the use of different IWV thresholds for different seasons for identifying storms (Yao *et al.* 2017; Zhao *et al.* 2020). Given such constraints, the thresholds derived in the present study are valid only for the monsoon season, and for the other seasons, the thresholds need to be fine-tuned.

5.3.3 Validation and Sensitivity Analysis

The validation of storm prediction using the above two-step process has been carried out using observations from the year 2020. A total of 54 storms have been identified during the monsoon season of 2020 (validation period), i.e., from X-band radar reflectivity, decrease in temperature, increase in wind speed, IWV and change in T_b=5 K hr^{-1} and considered them as reference dataset. The procedure outlined in Section 5.3.2, i.e., first predicting the IWV using light GBM and then using the thresholds for predicted IWV magnitude and variation and INSAT-derived T_b, have been followed to nowcast storms. The nowcasted storms are tested against the reference dataset to validate the algorithm and thresholds.

The algorithm's sensitivity on the use of one predictor or multiple predictors has been tested by comparing with reference dataset and the evaluation statistics are shown in Table 5.2. Since, the smallest values of histograms were taken as thresholds for the respective parameters, the model is able to detect all the storms/rain events with single or multiple parameters (except for stability indices, like CAPE and K index), as evidenced by 100% probability of detection. However, the false alarm rate is high (47%-55%), when the model uses only single parameter (IWV/IWV

gradient/ T_b gradient) or two parameters (IWV and IWV rate) for nowcasting. For instance, the IWV exceeds 50 mm on many days as we are taking data from monsoon season, during which the IWV is generally high. This increases false alarm rate alarmingly to 53%. Similarly, the usage of smallest gradient for Tb gradient increases the false alarm rate to 55%. One can notice high false alarm rate of 48% even with a larger threshold of IWV gradient than those available in the literature. Usage of smaller threshold will drastically increase the false alarm rate and reduces the model accuracy.

Traditionally meteorologists employ stability indices, like convective available potential energy (CAPE), K-index, etc., estimated from vertical soundings of temperature and humidity, for predicting storms. The above parameters are estimated using fifth-generation ECMWF atmospheric reanalysis of the global climate (ERA 5) data 30, 60, 90, 120 m before the storm/rainfall occurrence to capture the instability during prestorm environment. Similar to Figure 5.3, histograms have been constructed and the smallest values are considered as thresholds for nowcasting. Based on histograms, the thresholds for CAPE and K index are finalized as 500 J Kg^{-1} and 30 K, respectively. The 30 K threshold agrees with many earlier studies in India, the 500 J Kg^{-1} threshold for CAPE is smaller than the reported values, which typically are found to be larger than 1000 J Kg^{-1}.

Table 5.2: *Evaluation statistics of single, double and 3-parameter schemes in predicting storms along with threshold values used for evaluation.*

Threshold Values	**Probability of Detection (%)**	**Miss Rate (%)**	**False Alarm (%)**	**Model Accuracy (%)**
IWV (mm) > 50	100	0	53.4	49.18
IWV Rate (mm/hr) > 1.56	100	0	48.0	58.19
T_b Rate (K/hr) < -5	100	0	55.00	45.90
IWV > 50, IWV Rate >1.56	100	0	47.50	59.80
IWV > 50, IWV Rate > 1.56, T_b Rate < -5	100	0	5.20	97.50

CAPE (J/kg) > 500, K-Index (K) > 30	87.04	12.96	33.80	74.59
CAPE (J/kg) > 500, K-Index(K) > 30, PW > 50, PW Rate > 1.56	87.04	12.96	11.32	89.34

However, these storms are mostly intense in nature. The inclusion of all storms in the present study, including weak storms, reduced the threshold. It is also possible that ERA5 may not be able to properly represent the prestorm environment for all storms. Nevertheless, the usage of these two parameters certainly reduced the false alarms compared to single and double-parameter models based on IWV or Tb. Still, the model missed some of the storms.

Multiparameter models are found to be working well in nowcasting storms. For instance, introduction of the rate of change in brightness temperature as a third parameter in addition to the IWV and IWV gradient parameters has resulted in a noticeable improvement in the prediction algorithm. The three-parameter model lowered the false alarm rate drastically to 5.2% and increased the model accuracy to 97.50%. Similarly, inclusion of IWV and IWV parameters in CAPE and K-index based model, reduces the false alarm rate to 11% and increases model accuracy to 89%. It is clear from Table 5.2 that the three-parameter model using IWV, IWV gradient and T_b gradient works well with 100% detection capability of storms with a smaller false alarm rate.

The above analysis is carried out with 30 min. predicted IWV. To examine the sensitivity of predicted IWV with different lead times on storm nowcasting, the IWV predicted at different time intervals (30, 60, 90 and 120 min) has been (like in Figure 5.1) used in the model. Three parameter model with IWV, IWV gradient and T_b gradient as inputs are used for sensitivity analysis. It can be seen from Table 5.3 that, no storms were missed with 30 minutes lead time (detection rate=100%), but about 12% of total storms were not detected with 120 min lead time. The algorithm's performance degrades gradually with the increase in lead time. It can be seen that the probability of detection decreases gradually from 100% to 88%, missing rate and false alarms increase to 0 and 5% to 12%, respectively, as lead time increases from 30 min. to 120 min. Nevertheless, the model's accuracy remains above 90% even with 120 min. lead time, which is comparable or better than the existing methods/models in the literature.

Table 5.3: *Sensitivity of 3-parameter scheme in the prediction of storms with different lead times.*

Lead Time	Probability of Detection (%)	Miss Rate (%)	False Alarm Rate (%)	Model Accuracy (%)
30 min	100	0	6.89	96.72
60 min	96.29	3.70	8.77	94.26
90 min	94.44	7.40	10.71	91.80
120 min	87.50	12.50	12.72	89.34

5.4 Conclusion

The GNSS-based storm prediction techniques, which are increasingly popular in recent years, essentially follow one of the following approaches: AI/ML techniques or threshold-based schemes. In the present study, a new hybrid scheme is introduced based on the strengths of both approaches. The algorithm first predicts the IWV using a machine learning technique (light GBM) with GNSS and surface measurements as inputs and then nowcast storms based on multiparameter threshold algorithm. The accuracy of predicted IWV with 2 hours lead time is ~ 2 mm and the accuracy is found to be much better with shorter lead times (30 to 90 minutes). The main advantage of the present scheme is that it uses predicted IWV to nowcast storms, which provides more time for disaster preparedness and rescue operations. The thresholds employed in the present algorithm are purely based on 18 months of GNSS-based IWV measurements made during the stormy days, which are identified based on radar and AWS observations. The threshold for magnitude of IWV agrees quite well with those available in the literature at different geographical locations, indicating the robustness of this threshold. The study also clearly shows the increase in IWV starts few hours (1-4 hours) before the storm/rainfall occurrence, indicating the predictive potential of IWV gradient. However, the IWV gradient used in the present study (1.56 mm hr^{-1}) is larger than those used in some of the earlier studies. The smaller thresholds for IWV gradient in earlier studies are mainly due to the inclusion of routine diurnal variation in the estimation of gradient. In the current study climatological monthly diurnal variation of IWV has been used to identify storm associated IWV gradient and the threshold for gradient IWV is chosen from the storm associated IWV

gradient histogram. Smaller thresholds of IWV gradient will increase false alarm rates dramatically as reported in earlier studies.

Evaluation of the model has been carried out by comparing its outcome with a reference dataset of 54 stormy days in the year 2020 that were identified with an X-band radar and AWS. The sensitivity of the individual predictors and combination of predictors in nowcasting storms has also been studied. In addition, the efficacy of traditionally used stability indices for storm prediction has tested. Because of the smaller thresholds used in the present study, the probability of detection of storms is found to be 100% with IWV magnitude, IWV gradient, T_b gradient and any combination of these parameters. However, the false alarm rate is startlingly high with the usage of single or two predictors for nowcasting purpose. The inclusion of T_b gradient to IWV magnitude and gradients drastically reduced the false alarm rates to ~5% and increased the success ratio to 97.5%. The stability parameters, like CAPE and K index failed to predict storms accurately with 87% probability of detection and ~34% false alarm rate. The addition of IWV parameters improved the accuracy of the model by ~15%. The study unequivocally demonstrates the importance of IWV in nowcasting storms as it improves the model's success ratio with both traditional stability indices and T_b gradient.

The ability of the three-parameter model (IWV magnitude and gradient and Tb gradient) in nowcasting with different lead times has also been tested. It is noted, in general, that the model performance degrades with increase in lead time with increase in the percentage of missing storms and false alarm rates. However, the model's accuracy still remains close to 90% with 120 min. lead time, which is much better than the existing methods.

Chapter 6

Evaluation of AI/ML techniques for nowcasting storms

6.1 Introduction

The accurate prediction of storms is of paramount importance for effective disaster management and saving human life. Traditional storm forecasting methods, those based on numerical weather prediction products, have limitations in accurately predicting the variations in Integrated Water Vapor (IWV), a critical parameter influencing storm development and movement. Artificial Intelligence and Machine Learning (AI/ML) techniques are very powerful tools and are becoming increasingly popular in geoscience and earth sciences. Recently, several AI/ML-based models have been developed to nowcast storms using radar, lightning sensor, satellite and radiometer measurements (Roberts and Rutledge 2003). To overcome the limitations in traditional extrapolation methods of cell and area tracking of weather radar echoes, several studies focused on employing AI/ML techniques for nowcasting storms. Convolutional long short-term memory (LSTM) models with radar observations as input have been successful in nowcasting storms (Kim and Chandrasekar, 2021; Kim and Ushio 2022). Yang and Yuan (2023) argued that these techniques have a good potential to nowcast low-intensity precipitation events, but yield poor results in predicting high-intensity storms. They proposed customized multi-scale deep learning framework for nowcasting storms using weather radar observations, which ensures multi-scale spatiotemporal consistency. Recent attempts follow multi-sensor data fusion approaches to improve nowcasting of storms, including observations from weather radar, meteorological satellite and lightning observations along with numerical weather prediction model outputs (Leinonen *et al.* 2022; Ortland *et al.* 2023; Li *et al.* 2024).

In recent years, Global Navigation Satellite System (GNSS) technology has emerged as a promising solution for IWV measurement and analysis, offering improved precision and accuracy in storm nowcasting (Benevides *et al.* 2015; Yao *et al.* 2017; Manandhar *et al.* 2018; Martinez *et al.* 2022). Numerous methods have been developed to improve the predictability of storm occurrence by leveraging the concept of unusual enhancement in IWV before the storm. These techniques primarily depend on any or all of the factors related to IWV, like threshold magnitude of IWV, temporal change in IWV, rate of change of IWV (Benevides *et al.* 2015; Yao *et al.* 2017; Manandhar *et al.* 2018; Martinez *et al.* 2022). Some of the earlier methods based on single or two parameters were successful in predicting the storm/rainfall occurrence with 70%-90% accuracy but suffer with large false alarm rates (as large 70%). Benevides *et al.* (2019) developed an

algorithm integrating cloud top temperature and pressure with IWV, air pressure, relative humidity, surface air pressure, and temperature to predict rainfall events to reduce the false alarm rate. Los *et al.* (2020) proposed a methodology that combines GNSS-based IWV with wet refractivity profiles to predict storms in Poland with an accuracy of over 87%. Manandhar *et al.* (2019) analyze different meteorological parameters alongside IWV to create a data-driven machine learning algorithm for rainfall prediction, achieving a true detection rate of 80.4%. Khaniani *et al.* (2021) employed artificial neural network (ANN) models with IWV and other parameters as inputs, which outperformed standard meteorological forecasts. As discussed in chapter 4 we have developed a machine learning approach for the prediction of IWV using measurements from GNSS receiver and automatic weather station (AWS) with varying lead times. Their approach yielded good results not only for fair weather conditions but also during the storm conditions. Using predicted IWV and INSAT-derived T_b, in Chapter 5, devised a multi-parameter scheme to nowcast storms (Bisht *et al.* 2022; Bisht *et al.* 2024).

As noted above, all the studies either used thresholds of IWV and rate of IWV or some machine learning models to predict the storm without dwelling much on the efficacy of the model. Also, these studies do not dwell much on the comparison with other machine learning models. The present study, therefore, aims to develop and evaluate various machine- and deep-learning-based models (Logistic Regression, XGBoost, Random Forest, Catboost, LightGBM, and ANN) with GNSS IWV and INSAT-derived T_b as inputs for nowcasting storm events with various lead times ranging from 30 minutes to 120 minutes. The findings of this research hold implications for real-time storm warning systems, providing valuable insights into the potential of GNSS-based IWV data and machine learning models in storm nowcasting.

6.2 Data and Methodology

The detailed methodology of each Machine learning model has been described in chapter 2. Therefore, only summary of those models is given below.

6.2.1 Data Collection and Preprocessing

The data from multiple instruments, including GNSS receiver (Trimble NetR9), AWS (Vaisala WXT536) and Indian National Satellite (INSAT) – 3DR, have been used for nowcasting the storm. GNSS Integrated Water Vapor (IWV) data were obtained from a Global Navigation Satellite System (GNSS) receiver located at Gandaki, NARL, India. The GNSS receiver-sampled 1 Hz data

are post processed with GAMIT/GAMIK software by considering the data within the cone of ±75° from zenith. The pressure from the AWS is used to correct for Zenith Hydrostatic Delay from Zenith Tropospheric Delay to obtain Zenith Wet delay, which is then converted into IWV using the relations obtained for this region (Bisht *et al.* 2022). A detailed comparison with radiosonde derived IWV shows that the error in GNSS-derived IWV is less than 2 mm. The AWS measured pressure, temperature, wind speed and direction, humidity, and rainfall are also used for identification and prediction of the storm. The cloud growth information is obtained from Infrared brightness temperature (BT) from INSAT 3DR.

The data collection period spanned from July 2018 to December 2020, focusing specifically on the monsoon season. The dataset was divided into training, validation, and test subsets. The training dataset comprised data from 2018 and 2019, while the validation and test datasets contained data from 2020. Prior to conducting the analysis, quality control checks were implemented to ensure the reliability and consistency of the data (i.e., removal of erroneous or corrupted data points). Small gaps (less than 2 min) in the time-series of RINEX data are filled with interpolation techniques. If the data gap is longer than 20 min., that half an hour data is not considered for further analysis. We also synchronized the data from different sensors to align the timestamps accurately.

6.2.2 Model Selection and Training

Several machine learning algorithms have been evaluated to determine the most suitable model for storm prediction. To determine the best-performing model, a comparative analysis has been conducted using performance metrics such as Area Under the Receiver Operating Characteristic curve (AUC-ROC) and Cohen's kappa score. These metrics provide valuable insights into the model's classification accuracy and agreement with the ground truth data.

Based on the evaluation from figure 6.1, it is evident that the random forest algorithm is best among all as both ROC AUC and Kohen Kappa scores are higher than that of other models. Although, logistic Regression ROC AUC score is high but corresponding Cohen Kappa Score is way less indicating illogical prediction of storm events. Based on the Analysis, RF is selected as the primary model due to its superior performance in terms of AUC-ROC and kappa scores. RF also offers

advantages in handling missing values and outliers, making it well-suited for our dataset (Bisht *et al.* 2024).

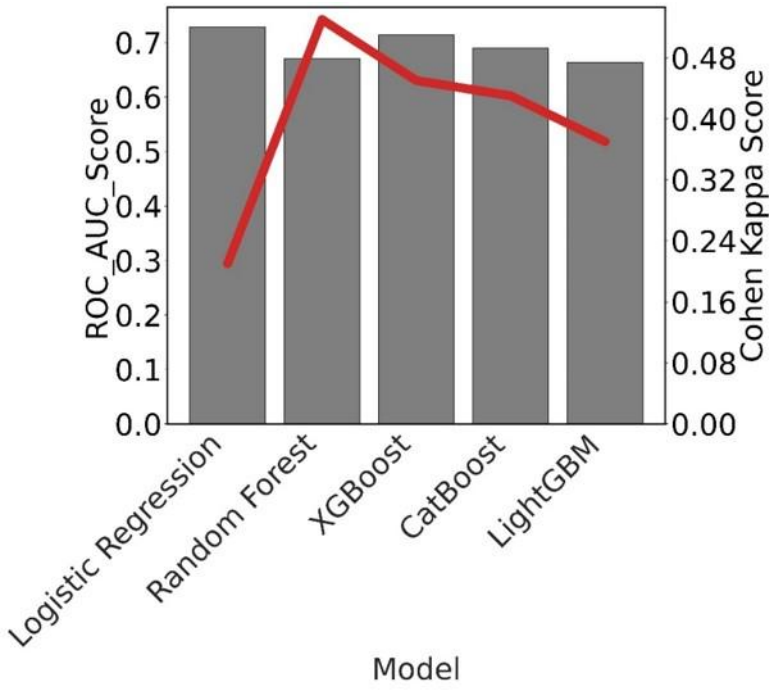

Figure 6.1: *Determining AUC ROC and Kohen Kappa Score for selecting Best ML model.*

In addition to the RF model, the potential of deep learning model is also explored by employing ANN model for the present purpose. Deep neural networks are known for their ability to capture complex relationships within data, and the ANN model was evaluated alongside RF to assess its effectiveness in storm prediction. Both models were trained on the training dataset and validated using the validation dataset, while the test dataset was used for final model evaluation.

The performance of the models was evaluated for different lead times (30, 60, 90, and 120 minutes) to assess their ability to predict storm events in advance. The evaluation focused on metrics such as true detection rate, false alarm rate, miss rate, and overall accuracy.

6.3 Results and Discussion

6.3.1 Model Performance and Feature Importance

The results of our analysis showed that the random forest model outperformed other machine learning algorithms, including Logistic Regression, XGBoost, Catboost, LightGBM and ANN, in

predicting storm events. Based on the evaluation metrics, including AUC-ROC and Cohen's kappa score, RF emerged as the most effective model, achieving an accuracy rate of 93.07% in nowcasting storms (Table 6.1). The true detection rate was 93.42%, while the false alarm rate was acceptable at 5.33%. These results indicate that the RF model is well-suited for real-time storm prediction, providing reliable and accurate forecasts.

Table 6.1: *Statistics for predictability RF model for different time horizons.*

Time Horizon	True Detection (%)	Miss Rate (%)	False Alarm (%)	Accuracy (%)
Real Time	93.42	6.58	5.33	93.07
30 mins	91.37	8.63	7.92	91.67
60 mins	88.58	11.42	10.45	89.12
90 mins	85.64	14.36	13.79	84.99
120 mins	81.33	18.67	17.73	82.40

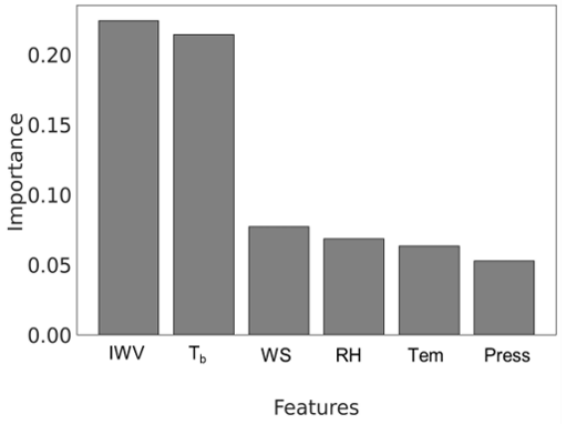

Figure 6.2: *Feature importance of various parameters employed in RF model.*

Feature importance analysis for RF model (Figure 6.2) revealed that IWV was the most critical parameter for storm prediction, followed by T_b, followed by wind speed, RH, T and P. This analysis highlights the importance of GNSS-derived IWV and satellite-based T_b as key indicators of storm potential, as also seen in Chapter 4. Though other meteorological parameters show large variation at the time of storm (like reduction in T, increase in WS and RH, etc.), but do not show such signature much before the storm, thereby, lacking the predictive potential.

6.3.2 Performance Across Different Time Horizons

The performance of the predictive models was assessed for different time horizons, ranging from 30 to 120 minutes. As seen in Table 6.1, the random forest model achieved accuracy rates of 91.67%, 89.12%, 84.99%, and 82.40% for lead times of 30, 60, 90, and 120 minutes, respectively. The corresponding true detection rates were 91.37%, 88.58%, 85.64%, and 81.33%. While the accuracy and detection rates decreased slightly with longer lead times, the model still maintained a reasonable level of accuracy, indicating its potential for operational use in nowcasting. Not only the true detection rate but for miss rate and false alarm rate, the RF model showed good statistics. False alarm was only 7.92% for 30 min which is quite acceptable in real time applications.

The predicting capability of deep neural-based ANN model has been also been examined and it is found that ANN model also performed well, achieving an accuracy rate of 91.45% (Table 6.2) in real-time scenario analysis. The true detection rates and false alarm rates of the ANN model were comparable to those of the RF model, suggesting that ANN is a viable option for storm prediction. Usually Deep learning models need large amount of data to train itself. Based on the availability of the dataset ANN model also performed well in this case.

Table 6.2: *Statistics for predictability ANN model for different time horizons.*

Time Horizon	True Detection (%)	False Alarm (%)	Miss Rate (%)	Accuracy (%)
Real Time	91.78	8.22	6.98	91.45
30 mins	89.91	10.09	9.27	90.15
60 mins	87.02	12.98	11.81	87.47
90 mins	83.89	16.11	14.79	84.25
120 mins	79.67	20.33	18.14	80.10

6.3.3 Comparative Analysis of RF and ANN Models

To establish the statistical significance of the performance differences between the RF and ANN models, hypothesis testing was conducted. The results indicated that the differences in performance between the two models were statistically insignificant ($p > 0.05$), implying that both models have similar capabilities in predicting storm events. However, the ANN model's computational efficiency provides a distinct advantage, making it a more efficient choice for

operational storm prediction, especially with larger datasets. The results of this study indicate that both RF and ANN have strengths that can be used for storm prediction. While RF demonstrated superior accuracy, the ANN model showed promise in terms of efficiency and computational performance. For future research, exploring ensemble methods or hybrid models that combine the strengths of RF and ANN could further enhance storm prediction accuracy and reliability, particularly when dealing with limited data. This approach holds significant promise for advancing storm nowcasting techniques and strengthening our ability to predict and prepare for storm events.

6.4 Conclusion

A comprehensive analysis has been presented in this chapter, highlighting the potential of machine learning-based approaches, particularly RF and ANN models, in nowcasting storm events. By leveraging GNSS-based IWV data, satellite-derived T_b, and other meteorological parameters, a robust framework for storm prediction has been developed for providing accurate and timely nowcasts of storms. The RF model exhibited high accuracy and robustness in predicting storm events, achieving a true detection rate of 93.42% and an accuracy rate of 93.07% in real-time scenarios. Feature importance analysis emphasized the critical role of IWV and T_b in storm prediction, highlighting the value of integrating GNSS and satellite data for improving storm nowcasting.

The performance of the models across different time horizons demonstrated their potential for providing early warnings, with accuracy rates remaining above 80% even for lead times of up to 120 minutes. The comparative analysis of RF and ANN models indicated that both models are suitable for storm prediction, with ANN offering additional advantages in terms of computational efficiency. Future work will focus on developing ensemble models that combine the strengths of RF and ANN, aiming to further improve prediction accuracy and reliability. Additionally, expanding the application of these models to different geographic regions and integrating additional atmospheric predictors will be key areas of future research. The insights gained from this study have significant implications for weather forecasting, disaster management, and infrastructure planning, ultimately contributing to enhanced preparedness and mitigation of storm impacts.

Chapter 7

Summary and Future scope

7.1 Summary

Storms are one of the deadliest weather hazards with devastating damage potential due to the accompanying severe winds, intense rainfall, and lightning. There is a tremendous improvement in the forecast of these storms by numerical weather prediction models in recent years due to better data availability for initialization of model and higher computational power. NWP models are proven their capability in forecasting large-scale storms, but often fail in predicting small-scale thunderstorms. Recently, GNSS-derived IWV-based nowcasting techniques are becoming popular by monitoring the moisture buildup. Moreover, GNSS-derived IWV's all weather and continuous capability makes it more attractive parameter for nowcasting of storms. The present thesis provides various nowcasting approaches for storm prediction (threshold-based techniques and AI/ML techniques) using 3 years of GNSS receiver measurements made at Gadanki.

The moisture builds up before the storm occurrence as moisture is the basic ingredient for cloud formation (cloud drop). Earlier researchers, therefore, mainly used the IWV variation (along with its magnitude) as a parameter for nowcasting of storms. However, the IWV exhibits a prominent diurnal cycle governed by a variety of processes, surface evaporation, moisture flux, and sea-breeze circulation. To delineate the moisture-build up associated with storm, one should understand the diurnal variation of IWV in different seasons. The first work, therefore, focuses on quantifying the diurnal variability of IWV in different seasons and understanding the underlying processes responsible for the diurnal variability. GNSS data from a network of seven receivers collected over a three-year period (September 2018 to August 2021) were used to investigate the diurnal and semi-diurnal variations in IWV. Harmonic analysis revealed that the diurnal harmonic (24-hour cycle) is more prominent and significantly stronger than the semi-diurnal harmonic, with notable seasonal and monthly variations in amplitude and phase. Both diurnal and semi-diurnal oscillations in IWV amplitude show two peaks in their annual cycle, November and April-May. The average diurnal amplitude range varies between 2.3 to 3.8 mm, whereas the average semi-diurnal amplitude range varies between 1.4 to 2 mm. It is also observed that the diurnal variations with amplitudes greater than 5 mm are occurring predominantly during the pre-monsoon and post-monsoon. The percentage occurrence of diurnal phase shows a broad peak during noon – midnight. The average phase of diurnal harmonic varies between 12-13 UTC in winter, 13.5-14.5 UTC in pre-monsoon, 14-15 UTC in monsoon, and 11.5 to 13 UTC in post-monsoon. This

indicates that IWV peaks late in hotter seasons of monsoon and pre-monsoon than the colder seasons. VIMFC combined with vertical wind are found to be the major factors contributing to the observed IWV diurnal variation over Gadanki and also for occurrence of seasonal differences in the phase. Diurnal phase of IWV is coinciding with the time at which maximum convergence is occurring at lower pressure levels for all seasons except monsoon. For monsoon, the diurnal peak is coinciding with the time when the atmosphere at lower pressure levels is changing from convergence to divergence

Long-term measurements of the GNSS receiver at Gadanki, India, have been used to develop a ML technique – light gradient boosting machine (lightGBM) for the prediction of IWV with different lead times. A variety of data sets related to IWV (representing source, sink, and transport) and short-scale features of IWV (gradients, sinusoidal pattern) have been used to train the model. Model performance is validated in different seasons. The predicted IWV at different lead times (30–120 min) perfectly captures the temporal variability of measured IWV with a correlation coefficient >0.99. The root mean square error (RMSE) of predicted IWV with 30 min lead time is less than 1 mm in all seasons. Nevertheless, the RMSE for predicted IWV with longer lead times increases with lead time but always remains <3 mm. The bias is slightly larger during the monsoon, mainly due to the higher occurrence of longer-duration rainy events. Even in those days, the model is able to accurately (correlation coefficient >0.99) predict the enhanced IWV before the rain occurrence. Sensitivity analysis and feature importance analysis on different predictors used in the model reveal that the IWV features are more important for short-scale prediction, like 30 min, whereas the importance of other predictors is high for longer lead time prediction (1–2 h) and on storm days.

A hybrid model for nowcasting of storms has been developed by employing predicted IWV with LightGBM and estimated thresholds for three storm predictors. The utilization of predicted IWV allows more lead time for disaster preparedness. The efficacy of light GBM technique in predicting IWV has been tested on 54 stormy days (from the year 2020), identified with a collocated polarimetric weather radar observations. The predicted IWV agrees very well with observed IWV with rms errors <1.2 mm (correlation coefficient >0.85) for predictions with a lead time up to 2 h. Among several predictors considered for nowcasting, IWV is found to have a great predictive potential, as the moisture buildup is seen few hours (1–4 h) prior to the occurrence of

storm/rainfall. Thresholds for chosen predictors [magnitude of IWV, change in IWV, and change in brightness temperature (Tb)] are finalized using the data from known stormy days (65 days from the years 2018 and 2019). The sensitivity analysis of the predictors independently and in combination in predicting storms reveals that 1 and 2 parameter-based predictions detect storms accurately means there is no miss rate but produce large false alarm rates. The three-parameter scheme reduced the false alarm rate drastically to 5% and improved the model accuracy to 97%, which is much better than the existing methods.

In addition to above hybrid model, which is based on thresholds, various AI/ML approaches have been followed, including Random Forest, Support Vector Machine, Decision Trees, and Artificial Neural Networks, to develop a stand-alone ML model for storm nowcasting using GNSS-derived IWV and additional meteorological parameters. Among the models evaluated, Random Forest demonstrated the highest accuracy and robustness. The contribution of different predictors, including IWV time series, brightness temperature, surface temperature, pressure, and wind speed, was also analyzed, with IWV and brightness temperature emerging as the most significant predictors. Evaluation metrics, such as AUC-ROC score and Cohen's Kappa, were used to assess model performance. The deep learning-based ANN model also demonstrated the potential of using GNSS-based IWV variations for real-time storm warning systems. Based on Cohen's Kappa and AUC-ROC score, Random Forest emerged as the best method for storm prediction, achieving a true detection rate greater than 81% for a 2-hour lead time, while ANN also showed similar capability with a true detection rate of nearly 80% for the same lead time, demonstrating the applicability of advanced data-driven approaches in improving storm prediction accuracy and aiding disaster preparedness and response in semi-arid regions.

7.2 Future Scope

The findings of this thesis offer promising directions for further research and development in the field of atmospheric science and storm prediction. The following are the key areas for future exploration:

7.2.1 Expansion to Broader Geographical Regions

Future studies could expand the methodology used in this research to different geographical regions, such as coastal, mountainous, and urban areas, to evaluate how IWV variability and storm

prediction models perform under varying climatic and environmental conditions. This expansion would enhance the generalizability of the prediction models.

7.2.2 Integration with Additional Data Sources

Incorporating additional atmospheric data, such as soil moisture, NWP outputs and other satellite-derived parameters, can enhance the predictive capabilities of the models. Using these diverse data sources would provide a better understanding of the interactions between atmospheric components, potentially improving IWV prediction and storm forecasting accuracy.

7.2.3 Real-Time Implementation and Early Warning Systems

A major future goal is the real-time application of these models in operational weather forecasting. Integrating the prediction and nowcasting models into existing early warning systems can help meteorological agencies provide timely alerts for extreme weather events, thereby improving disaster preparedness and emergency response efforts.

7.2.4 Refinement of Regional and Seasonal Thresholds

Developing region-specific and season-specific thresholds for storm prediction can help improve model accuracy and reduce false alarms. Future work could focus on adapting the storm prediction thresholds to account for regional climatic characteristics, making the models more versatile for different weather conditions.

7.2.5 Hybrid Modeling Approaches for Better Accuracy

This thesis introduced a hybrid approach combining machine learning and threshold-based techniques for storm nowcasting. Future research could further enhance hybrid models by incorporating probabilistic methods and uncertainty quantification, providing better predictions under varying data conditions, and ensuring improved reliability in early warning systems.

References

Adams, D.K. and Souza, E.P., 2009. CAPE and convective events in the Southwest during the North American monsoon. Monthly Weather Review, 137(1), pp.83-98.

Adams, D.K., McGillicuddy Jr, D.J., Zamudio, L., Thurnherr, A.M., Liang, X., Rouxel, O., German, C.R. and Mullineaux, L.S., 2011. Surface-generated mesoscale eddies transport deep-sea products from hydrothermal vents. science, 332(6029), pp.580-583.

Ahmed, F. and Schumacher, C., 2017. Geographical differences in the tropical precipitation-moisture relationship and rain intensity onset. Geophysical Research Letters, 44(2), pp.1114-1122.

Allan, R.P., Barlow, M., Byrne, M.P., Cherchi, A., Douville, H., Fowler, H.J., Gan, T.Y., Pendergrass, A.G., Rosenfeld, D., Swann, A.L. and Wilcox, L.J., 2020. Advances in understanding large-scale responses of the water cycle to climate change. Annals of the New York Academy of Sciences, 1472(1), pp.49-75.

Anderson, J.L., 1996. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. Journal of climate, 9(7), 1518.

Anthes, R.A., Bernhardt, P.A., Chen, Y., Cucurull, L., Dymond, K.F., Ector, D., Healy, S.B., Ho, S.P., Hunt, D.C., Kuo, Y.H. and Liu, H., 2008. The COSMIC/FORMOSAT-3 mission: Early results. Bulletin of the American Meteorological Society, 89(3), 313.

Anthes, R.A., Rocken, C. and Ying-Hwa, K., 2000. Applications of COSMIC to meteorology and climate. Terrestrial Atmospheric and Oceanic Sciences, 11(1), 115.

Atlas, D., Booker, D.R., Byers, H., Douglas, R.H., Fujita, T., House, D.C., Ludlum, F.H., Malkus, J.S., Newton, C.W., Ogura, Y. and Schleusener, R.A., 1963. Severe local storms: A review. Severe Local Storms, 1.

Attri, S. D., & Tyagi, A. (2010). Climate Profile of India. India Meteorological Department, Ministry of Earth Sciences, Government of India. Retrieved from http://www.imd.gov.in/

Baba shaeb Kannemadugu, H., Ranganathan, K., Gharai, B. and Seshasai, M.V.R., 2022. GNSS-GPS derived integrated water vapor and performance assessment of ERA-5 data over India. Journal of Atmospheric and Solar-Terrestrial Physics, 227, 105807.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R. and Fuentes, J., 2001. FLUXNET: A new tool to study the temporal and spatial

variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bulletin of the American Meteorological Society, 82(11), 2415.

Barman, P., Jade, S., Kumar, A. and Jamir, W., 2017. Inter annual, spatial, seasonal, and diurnal variability of precipitable water vapour over northeast India using GPS time series. International journal of remote sensing, 38(2), 391.

Bauer, P., Thorpe, A. and Brunet, G., 2015. The quiet revolution of numerical weather prediction. Nature, 525(7567), 47.

Bender, M., Dick, G., Ge, M., Deng, Z., Wickert, J., Kahle, H.G., Raabe, A. and Tetzlaff, G., 2011. Development of a GNSS water vapour tomography system using algebraic reconstruction techniques. Advances in Space Research, 47(10), 1704.

Benevides, P., Catalao, J. and Miranda, P.M.A., 2015. On the inclusion of GPS precipitable water vapour in the nowcasting of rainfall. Natural Hazards and Earth System Sciences, 15(12), 2605.

Benevides, P., Catalao, J. and Nico, G., 2019. Neural network approach to forecast hourly intense rainfall using GNSS precipitable water vapor and meteorological sensors. Remote Sensing, 11(8), 966.

Bengtsson, L., Hodges, K.I. and Hagemann, S., 2004. Sensitivity of the ERA40 reanalysis to the observing system: determination of the global atmospheric circulation from reduced observations. Tellus A: Dynamic Meteorology and Oceanography, 56(5), 456.

Bevis, M., Businger, S., Herring, T.A., Rocken, C., Anthes, R.A. and Ware, R.H., 1992. GPS meteorology: Remote sensing of atmospheric water vapor using the global positioning system. Journal of Geophysical Research: Atmospheres, 97(D14), 15787.

Bisht, D.S., Kumar, D.P., Amarjyothi, K. and Saha, U., 2024. Bias correction of satellite precipitation estimates using Mumbai-MESONET observations: A random forest approach. Atmospheric Research, 107858.

Bisht, D.S., Rao, T.N., Rao, N.R. and Chandrakanth, S.V., 2024. Nowcasting of Storms Using Predicted Integrated Water Vapor with a Machine Learning Technique and Satellite Brightness Temperature. IEEE Transactions on Geoscience and Remote Sensing.

Bisht, D.S., Rao, T.N., Rao, N.R., Chandrakanth, S.V. and Sharma, A., 2022. Prediction of integrated water vapor using a machine learning technique. IEEE Geoscience and Remote Sensing Letters, 19, 1.

Blake, E. S., and D. A. Zelinsky, 2018: National Hurricane Center tropical cyclone report: Hurricane Harvey (17 August– 1 September 2017). NHC Tech. Rep. AL092017, 77 pp. https://www.nhc.noaa.gov/data/tcr/AL092017_Harvey.

Blake, E., Kimberlain, T., Berg, R., Cangialosi, J. and Bevin, I.I., 2013. J (2013) Tropical cyclone report hurricane sandy. National hurricane center [online]

Bock, O., Bouin, M.N., Walpersdorf, A., Lafore, J.P., Janicot, S., Guichard, F. and Agusti-Panareda, A., 2007. Comparison of ground-based GPS precipitable water vapour to independent observations and NWP model reanalyses over Africa. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 133(629), pp.2011-2027.

Bock, O., Guichard, F., Janicot, S., Lafore, J.P., Bouin, M.N. and Sultan, B., 2007. Multiscale analysis of precipitable water vapor over Africa from GPS data and ECMWF analyses. Geophysical research letters, 34(9).

Bock, O., Keil, C., Richard, E., Flamant, C. and Bouin, M.N., 2005. Validation of precipitable water from ECMWF model analyses with GPS and radiosonde data during the MAP SOP. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 131(612), pp.3013-3036.

Boehm, J., Werl, B., & Schuh, H. (2006). Troposphere Mapping Functions for GPS and Very Long Baseline Interferometry from European Centre for Medium-Range Weather Forecasts Operational Analysis Data. Journal of Geophysical Research: Solid Earth, 111(B2).

Breiman, L., 2001. Random forests. Machine learning, 45, 5.

Bretherton, C.S., Peters, M.E. and Back, L.E., 2004. Relationships between water vapor path and precipitation over the tropical oceans. Journal of climate, 17(7), pp.1517-1528.

Brooks, H.E. and Weiss, S.J., 1999. Severe local storms. In Proceedings of the WMO/UNESCO Sub-forum on Science and Technology in Support of Natural Disaster Reduction, 12.

Buehler, S.A., Östman, S., Melsheimer, C., Holl, G., Eliasson, S., John, V.O., Blumenstock, T., Hase, F., Elgered, G., Raffalski, U. and Nasuno, T., 2012. A multi-instrument comparison of integrated water vapour measurements at a high latitude site. Atmospheric Chemistry and Physics, 12(22), pp.10925-10943.

Buizza, R., Richardson, D.S. and Palmer, T.N., 2003. Benefits of increased resolution in the ECMWF ensemble system and comparison with poor-man's ensembles. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 129(589), 1269.

Casallas-García, A., Hernández-Deckers, D. and Mora-Páez, H., 2023. Understanding convective storms in a tropical, high-altitude location with in-situ meteorological observations and GPS-derived water vapor. Atmósfera, 36(2), 225-238.

Champollion, C., Flamant, C., Bock, O., Masson, F., Turner, D.D. and Weckwerth, T., 2009. Mesoscale GPS tomography applied to the 12 June 2002 convective initiation event of IHOP_2002. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 135(640), 645.

Champollion, C., Masson, F., Bouin, M.N., Walpersdorf, A., Doerflinger, E., Bock, O. and Van Baelen, J., 2005. GPS water vapour tomography: preliminary results from the ESCOMPTE field experiment. Atmospheric research, 74(1-4), 253.

Chen, B., Liu, Z., Wong, W.K. and Woo, W.C., 2017. Detecting water vapor variability during heavy precipitation events in Hong Kong using the GPS tomographic technique. Journal of Atmospheric and Oceanic Technology, 34(5), 1001.

Chen, L., Cao, Y., Ma, L. and Zhang, J., 2020. A deep learning-based methodology for precipitation nowcasting with radar. Earth and Space Science, 7(2), p.2019EA000812.

Chen, T. and Guestrin, C., 2016, August. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

Chkeir, S., Anesiadou, A., Mascitelli, A. and Biondi, R., 2023. Nowcasting extreme rain and extreme wind speed with machine learning techniques applied to different input datasets. Atmospheric Research, 282, 106548.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 20(1), 37.

Collier, C.G., 1996. Applications of weather radar systems: a guide to uses of radar data in meteorology and hydrology.

Dai, A., Wang, J., Ware, R.H. and Van Hove, T., 2002. Diurnal variation in water vapor over North America and its implications for sampling errors in radiosonde humidity. Journal of Geophysical Research: Atmospheres, 107(D10), ACL-11.

Davis, R.S., 2001. Flash flood forecast and detection methods. In Severe convective storms. Boston, MA: American Meteorological Society, 481.

Del Genio, A.D. and Kovari, W., 2002. Climatic properties of tropical precipitating convection under varying environmental conditions. Journal of Climate, 15(18), pp.2597-2615.

Dessler, A.E., 2010. A determination of the cloud feedback from climate variations over the past decade. Science, 330(6010), 1523.

Dimri, A.P., Niyogi, D., Barros, A.P., Ridley, J., Mohanty, U.C., Yasunari, T. and Sikka, D.R., 2015. Western disturbances: a review. Reviews of Geophysics, 53(2), 225.

Dimri, A.P., Yasunari, T., Kotlia, B.S., Mohanty, U.C. and Sikka, D.R., 2016. Indian winter monsoon: Present and past. Earth-science reviews, 163, pp.297-322.

Dorogush, A.V., Ershov, V. and Gulin, A., 2018. CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363.

Doswell, C.A., Brooks, H.E. and Maddox, R.A., 1996. Flash flood forecasting: An ingredients-based methodology. Weather and forecasting, 11(4), 560.

Duan, B., Hugentobler, U., Chen, J., Selmke, I. and Wang, J., 2019. Prediction versus real-time orbit determination for GNSS satellites. GPS Solutions, 23, pp.1-10.

ECMWF (2017). ERA5 Reanalysis Data. Retrieved from https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5.

El Zein, A.L. and Chehayeb, N.A., 2015. The effect of greenhouse gases on earth's temperature. International Journal of Environmental Monitoring and Analysis, 3(2), p.74.

Emanuel, K., 2003. Tropical cyclones. Annual review of earth and planetary sciences, 31(1), 75.

Emanuel, K.E.R.R.Y.A., 1991. The theory of hurricanes. Annual Review of Fluid Mechanics, 23(1), 179.

Emmanuel, M., Sunilkumar, S.V., Ratnam, M.V., Muhsin, M., Parameswaran, K. and Murthy, B.K., 2018. Diurnal variation of the tropospheric water vapour over a coastal and an inland station in Southern Indian Peninsula. Journal of Atmospheric and Solar-Terrestrial Physics, 179, 11.

Esteban, M., Valenzuela, V.P., Yun, N.Y., Mikami, T., Shibayama, T., Matsumaru, R., Takagi, H., Thao, N.D., De Leon, M., Oyama, T. and Nakamura, R., 2015. Typhoon Haiyan 2013 evacuation preparations and awareness. International Journal of Sustainable Future for Human Security, 3(1), 37.

Fabry, F., Bellon, A., Duncan, M.R. and Austin, G.L., 1994. High resolution rainfall measurements by radar for very small basins: the sampling problem reexamined. Journal of Hydrology, 161(1-4), 415.

FAO. (2015). The impact of disasters on agriculture and food security. Food and Agriculture Organization of the United Nations. https://www.fao.org/3/a-i5128e.pdf

Fawcett, T., 2006. An introduction to ROC analysis. Pattern recognition letters, 27(8), 861.

FEMA. (2020). Hurricane preparedness. Federal Emergency Management Agency. https://www.fema.gov/emergency-managers/risk-management/hurricanes

Feng, J., Zhou, X., Zhang, Y., & Sun, J. (2021). A deep learning approach for precipitation nowcasting by radar. Atmospheric Research, 249, 105292.

Flores, A., Ruffini, G. and Rius, A., 2000, February. 4D tropospheric tomography using GPS slant wet delays. In Annales geophysicae (Vol. 18, No. 2, 223). Berlin/Heidelberg: Springer-Verlag.

Flowerdew, J., Horsburgh, K., Wilson, C. and Mylne, K., 2010. Development and evaluation of an ensemble forecasting system for coastal storm surges. Quarterly Journal of the Royal Meteorological Society, 136(651), pp.1444-1456.

Foote, G.B. and Knight, C. eds., 2016. Hail: a review of hail science and hail suppression.

Frame, T., Harrison, G., Hewson, T. and Roberts, N., 2017. Meteorological risk: extra-tropical cyclones, tropical cyclones and convective storms. Science for disaster risk management, 246.

Franklin, J.L., Black, M.L. and Valde, K., 2003. GPS dropwindsonde wind profiles in hurricanes and their operational implications. Weather and forecasting, 18(1), .32.

Galewsky, J., Steen-Larsen, H.C., Field, R.D., Worden, J., Risi, C. and Schneider, M., 2016. Stable isotopes in atmospheric water vapor and applications to the hydrologic cycle. Reviews of Geophysics, 54(4), pp.809-865.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Gopikrishnan, G.S., Kuttippurath, J., Thapliyal, P.K. and Shukla, M.V., 2023. Validation of INSAT-3D and INSAT-3DR Temperature Profile Retrievals Using Ground-Based, Satellite, and Reanalysis Data. Journal of Geophysical Research: Atmospheres, 128(22), p.e2023JD038912.

Guerova, G., Douša, J., Dimitrova, T., Stoycheva, A., Václavovic, P. and Penov, N., 2022. GNSS storm nowcasting demonstrator for Bulgaria. Remote Sensing, 14(15),.3746.

Gutman, S.I., Sahm, S.R., Benjamin, S.G., Schwartz, B.E., Holub, K.L., Stewart, J.Q. and Smith, T.L., 2004. Rapid retrieval and assimilation of ground based GPS precipitable water observations at the NOAA Forecast Systems Laboratory: Impact on weather forecasts. Journal of the Meteorological Society of Japan. Ser. II, 82(1B), 351.

Hallegatte, S., Vogt-Schilb, A., Bangalore, M. and Rozenberg, J., 2016. Unbreakable: building the resilience of the poor in the face of natural disasters. World Bank Publications.

Halverson, J.B., 2024. An Introduction to Severe Storms and Hazardous Weather. Taylor & Francis.

Harper, B., Granger, K. and Hall, S., 2000. Severe thunderstorm risks. Natural Hazards and the Risks they Pose to South-East Queensland, 11.

He, Z., Wang, D., Qiu, X., Jiang, Y., Li, H. and Shu, A., 2021. Application of radar data assimilation on convective precipitation forecasts based on water vapor retrieval. Meteorology and Atmospheric Physics, 133(3), 611.

Held, I.M. and Soden, B.J., 2000. Water vapor feedback and global warming. Annual review of energy and the environment, 25(1), pp.441-475.

Hermans, T.D., Šakić Trogrlić, R., van den Homberg, M.J., Bailon, H., Sarku, R. and Mosurska, A., 2022. Exploring the integration of local and scientific knowledge in early warning systems for disaster risk reduction: a review. Natural Hazards, 114(2), pp.1125-1152.

Herring, T. A., King, R. W., & McClusky, S. C. (2010). Introduction to GAMIT/GLOBK, Release 10.4. Massachusetts Institute of Technology.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D. and Simmons, A., 2020. The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), pp.1999-2049.

Holloway, C.E. and Neelin, J.D., 2009. Moisture vertical structure, column water vapor, and tropical deep convection. Journal of the atmospheric sciences, 66(6), pp.1665-1683.

Holton, J. R. (2004). An Introduction to Dynamic Meteorology (4th ed.). Academic Press.

Hosmer Jr, D.W., Lemeshow, S. and Sturdivant, R.X., 2013. Applied logistic regression. John Wiley & Sons.

Houze Jr, R.A., 2004. Mesoscale convective systems. Reviews of Geophysics, 42(4).

Houze Jr, R.A., 2014. Cloud dynamics. Academic press.

Jacob, D., 2001. The role of water vapour in the atmosphere. A short overview from a climate modeller's point of view. Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy, 26(6-8), pp.523-527.

Jadala, N.B., Sridhar, M., Dashora, N. and Dutta, G., 2020. Annual, seasonal and diurnal variations of integrated water vapor using GPS observations over Hyderabad, a tropical station. Advances in Space Research, 65(1), pp.529-540.

Jadala, N.B., Sridhar, M., Dashora, N. and Dutta, G., 2020. Annual, seasonal and diurnal variations of integrated water vapor using GPS observations over Hyderabad, a tropical station. Advances in Space Research, 65(1), pp.529-540.

Jade, S., Vijayan, M.S.M., Gaur, V.K., Prabhu, T.P. and Sahu, S.C., 2005. Estimates of precipitable water vapour from GPS data over the Indian subcontinent. Journal of atmospheric and solar-terrestrial physics, 67(6), pp.623-635.

Kalinnikov, V.V. and Khutorova, O.G., 2017, March. Diurnal variations in integrated water vapor derived from a GPS ground network in the Volga–Ural region of Russia. In Annales Geophysicae (Vol. 35, No. 3, pp. 453-464). Göttingen, Germany: Copernicus Publications.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.

Kiehl, J.T. and Trenberth, K.E., 1997. Earth's annual global mean energy budget. Bulletin of the American meteorological society, 78(2), pp.197-208.

Kim, D-K and Ushio, T. 2022. Nowcasting Meso-γ-Scale Convective Storms Using Convolutional LSTM Models and High-Resolution Radar Observations. Tellus A: Dynamic Meteorology and Oceanography, 74(2022), 17–32. DOI: https://doi. org/10.16993/tellusa.37

Kim, E. and Chandrasekar, V., 2021, July. Evaluation of Machine Learning Based Nowcasting Between Storms Over Different Geographical Regions. In 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS (pp. 7172-7175). IEEE.

King, M.D., Menzel, W.P., Kaufman, Y.J., Tanré, D., Gao, B.C., Platnick, S., Ackerman, S.A., Remer, L.A., Pincus, R. and Hubanks, P.A., 2003. Cloud and aerosol properties, precipitable water, and profiles of temperature and water vapor from MODIS. IEEE Transactions on Geoscience and Remote Sensing, 41(2), pp.442-458.

Knabb, R.D., Rhome, J.R. and Brown, D.P., 2006. Tropical cyclone report: Hurricane katrina, august 23-30, 2005. Fire Engineering, 159(5), pp.32-40.

Kober, K. and Tafferner, A., 2009. Tracking and nowcasting of convective cells using remote sensing data from radar and satellite. Meteorologische Zeitschrift, 1, pp.75-84.

Kouba, J. and Héroux, P., 2001. Precise point positioning using IGS orbit and clock products. GPS solutions, 5, pp.12-28.

Kursinski, E.R., Hajj, G.A., Schofield, J.T., Linfield, R.P. and Hardy, K.R., 1997. Observing Earth's atmosphere with radio occultation measurements using the Global Positioning System. Journal of Geophysical Research: Atmospheres, 102(D19), pp.23429-23465.

Leinonen, J., Hamann, U., Germann, U. and Mecikalski, J.R., 2022. Nowcasting thunderstorm hazards using machine learning: The impact of data sources on performance. Natural Hazards and Earth System Sciences, 22(2), pp.577-597.

Leinonen, J., Hamann, U., Sideris,I. V., & Germann, U. (2023).Thunderstorm nowcasting with deeplearning: A multi-hazard data fusionmodel. Geophysical Research Letters,50, e2022GL101626. https://doi.org/10.1029/2022GL101626

Leutbecher, M. and Palmer, T.N., 2008. Ensemble forecasting. Journal of computational physics, 227(7), pp.3515-3539.

Li, G., Kimura, F., Sato, T. and Huang, D., 2008. A composite analysis of diurnal cycle of GPS precipitable water vapor in central Japan during Calm Summer Days. Theoretical and applied climatology, 92, pp.15-29.

Li, H., Wang, X., Wu, S., Zhang, K., Chen, X., Qiu, C., Zhang, S., Zhang, J., Xie, M. and Li, L., 2020. Development of an improved model for prediction of short-term heavy precipitation based on GNSS-derived PWV. Remote Sensing, 12(24), p.4101.

Li, J., J. Zheng, B. Li, *et al.*, 2024: Quantitative applications of weather satellite data for nowcasting: Pro gress and challenges. J. Meteor. Res., 38(3), 399–413, doi: 10.1007/s13351-024-3138-6.

Li, X., Dick, G., Ge, M., Heise, S., Wickert, J. and Bender, M., 2014. Real-time GPS sensing of atmospheric water vapor: Precise point positioning with orbit, clock, and phase delay corrections. Geophysical Research Letters, 41(10), pp.3615-3621.

Liou, Y.A., Teng, Y.T., Van Hove, T. and Liljegren, J.C., 2001. Comparison of precipitable water observations in the near tropics by GPS, microwave radiometer, and radiosondes. Journal of Applied Meteorology and Climatology, 40(1), pp.5-15.

Liu, Y., Weerts, A.H., Clark, M., Hendricks Franssen, H.J., Kumar, S., Moradkhani, H., Seo, D.J., Schwanenberg, D., Smith, P., Van Dijk, A.I.J.M. and Van Velzen, N., 2012. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. Hydrology and earth system sciences, 16(10), pp.3863-3887.

Lorenz, E.N., 1969. The predictability of a flow which possesses many scales of motion. Tellus, 21(3), pp.289-307.

Łoś, M., Smolak, K., Guerova, G. and Rohm, W., 2020. GNSS-based machine learning storm nowcasting. Remote Sensing, 12(16), p.2536.

Maddox, R.A., 1980. Mesoscale convective complexes. Bulletin of the American Meteorological Society, pp.1374-1387.

Madhulatha, A., Rajeevan, M., Venkat Ratnam, M., Bhate, J. and Naidu, C.V., 2013. Nowcasting severe convective activity over southeast India using ground-based microwave radiometer observations. Journal of Geophysical Research: Atmospheres, 118(1), pp.1-13.

Manandhar, S., Lee, Y.H., Meng, Y.S., Yuan, F. and Ong, J.T., 2018. GPS-derived PWV for rainfall nowcasting in tropical region. IEEE transactions on geoscience and remote sensing, 56(8), pp.4835-4844.

Markowski, P. and Richardson, Y., 2011. Mesoscale meteorology in midlatitudes. John Wiley & Sons.

Medema, W., McIntosh, B.S. and Jeffrey, P.J., 2008. From premise to practice: a critical assessment of integrated water resources management and adaptive management approaches in the water sector. Ecology and Society, 13(2).

Meza, A., Mendoza, L., Natali, M.P., Bianchi, C. and Fernández, L., 2020. Diurnal variation of precipitable water vapor over Central and South America. Geodesy and Geodynamics, 11(6), pp.426-441.

Mimura, N., Yasuhara, K., Kawagoe, S., Yokoki, H. and Kazama, S., 2011. Damage from the Great East Japan Earthquake and Tsunami-a quick report. Mitigation and adaptation strategies for global change, 16, pp.803-818.

Mitchell, J.F., 1989. The "greenhouse" effect and climate change. Reviews of Geophysics, 27(1), pp.115-139.

Moller, A.R., 2001. Severe local storms forecasting. In Severe convective storms (pp. 433-480). Boston, MA: American Meteorological Society.

Morss, R.E., Vickery, J., Lazrus, H., Demuth, J. and Bostrom, A., 2022. Improving tropical cyclone forecast communication by understanding NWS partners' decision timelines and forecast information needs. Weather, Climate, and Society, 14(3), pp.783-800.

Mostajabi, A., Finney, D.L., Rubinstein, M. and Rachidi, F., 2019. Nowcasting lightning occurrence from commonly available meteorological parameters using machine learning techniques. Npj Climate and Atmospheric Science, 2(1), p.41.

Neelin, J.D., Peters, O. and Hales, K., 2009. The transition to strong convection. Journal of the Atmospheric Sciences, 66(8), pp.2367-2384.

Newton, C.W., 1967. Severe convective storms. In Advances in geophysics (Vol. 12, pp. 257-308). Elsevier.

Ortiz de Galisteo, J.P., Cachorro, V., Toledano, C., Torres, B., Laulainen, N., Bennouna, Y. and De Frutos, A., 2011. Diurnal cycle of precipitable water vapor over Spain. Quarterly Journal of the Royal Meteorological Society, 137(657), pp.948-958.

Ortland, S.M., Pavolonis, M.J. and Cintineo, J.L., 2023. The Development and Initial Capabilities of ThunderCast, a Deep Learning Model for Thunderstorm Nowcasting in the United States. Artificial Intelligence for the Earth Systems, 2(4), p.e230044.

Peters, O. and Neelin, J.D., 2006. Critical phenomena in atmospheric precipitation. Nature physics, 2(6), pp.393-396.

Pfister, A. and Cassar, A., 1999. Use and benefit of radar rainfall data in an urban real time control project. Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere, 24(8), pp.903-908.

Prudden, R., Adams, S., Kangin, D., Robinson, N., Ravuri, S., Mohamed, S. and Arribas, A., 2020. A review of radar-based nowcasting of precipitation and applicable machine learning techniques. arXiv preprint arXiv:2005.04988.

Puviarasan, N., Sharma, A.K., Ranalkar, M. and Giri, R.K., 2015. Onset, advance and withdrawal of southwest monsoon over Indian subcontinent: A study from precipitable water measurement using ground based GPS receivers. Journal of Atmospheric and Solar-Terrestrial Physics, 122, pp.45-57.

Radhakrishna, B., Fabry, F., Braun, J.J. and Van Hove, T., 2015. Precipitable water from GPS over the continental United States: Diurnal cycle, intercomparisons with NARR, and link with convective initiation. Journal of Climate, 28(7), pp.2584-2599.

Radhakrishna, B., Rao, T.N. and Saikranthi, K., 2019. Spatial coherence of water vapor and rainfall over the Indian subcontinent during different monsoon seasons. Journal of Hydrometeorology, 20(1), pp.45-58.

Rakov, V.A. and Uman, M.A., 2003. Lightning: physics and effects. Cambridge university press.

Ralph, F.M., Neiman, P.J. and Wick, G.A., 2004. Satellite and CALJET aircraft observations of atmospheric rivers over the eastern North Pacific Ocean during the winter of 1997/98. Monthly weather review, 132(7), pp.1721-1745.

Ratna, S.B., Cherchi, A., Joseph, P.V., Behera, S.K., Abish, B. and Masina, S., 2016. Moisture variability over the Indo-Pacific region and its influence on the Indian summer monsoon rainfall. Climate Dynamics, 46, pp.949-965.

Renju, R., Suresh Raju, C., Mathew, N., Antony, T. and Krishna Moorthy, K., 2015. Microwave radiometer observations of interannual water vapor variability and vertical structure over a tropical station. Journal of Geophysical Research: Atmospheres, 120(10), pp.4585-4599.

Roberts, R.D. and Rutledge, S., 2003. Nowcasting storm initiation and growth using GOES-8 and WSR-88D data. Weather and Forecasting, 18(4), pp.562-584. Refrerences

Rocken, C., Ware, R., Van Hove, T., Solheim, F., Alber, C., Johnson, J., Bevis, M. and Businger, S., 1993. Sensing atmospheric water vapor with the Global Positioning System. Geophysical research letters, 20(23), pp.2631-2634.

Rodwell, M.J. and Palmer, T.N., 2007. Using numerical weather prediction to assess climate models. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 133(622), pp.129-146.

Rogers, R.R. and Yau, M.K., 1989. A short course in cloud physics. Bull. Amer. Meteor. Soc, 45, p.619.

Ryzhkov, A., Zhang, P., Reeves, H., Kumjian, M., Tschallener, T., Trömel, S. and Simmer, C., 2016. Quasi-vertical profiles—A new way to look at polarimetric radar data. Journal of Atmospheric and Oceanic Technology, 33(3), pp.551-562.

Sandeep, A., Rao, T.N., Ramkiran, C.N. and Rao, S.V.B., 2014. Differences in atmospheric boundary-layer characteristics between wet and dry episodes of the Indian summer monsoon. Boundary-layer meteorology, 153, pp.217-236.

Sangiorgio, M., Barindelli, S., Guglieri, V., Biondi, R., Solazzo, E., Realini, E., Venuti, G. and Guariso, G., 2020. A comparative study on machine learning techniques for intense convective rainfall events forecasting. In Theory and Applications of Time Series Analysis: Selected Contributions from ITISE 2019 6 (pp. 305-317). Springer International Publishing.

Sapucci, L.F., Machado, L.A., de Souza, E.M. and Campos, T.B., 2019. Global P ositioning S ystem precipitable water vapour (GPS-PWV) jumps before intense rain events: A potential application to nowcasting. Meteorological Applications, 26(1), pp.49-63.

Schiro, K.A. and Neelin, J.D., 2018. Tropical continental downdraft characteristics: mesoscale systems versus unorganized convection. Atmospheric Chemistry and Physics, 18(3), pp.1997-2010.

Schmit, T.J., Griffith, P., Gunshor, M.M., Daniels, J.M., Goodman, S.J. and Lebair, W.J., 2017. A closer look at the ABI on the GOES-R series. Bulletin of the American Meteorological Society, 98(4), pp.681-698.

Schneider, M., Hase, F. and Blumenstock, T., 2006. Ground-based remote sensing of HDO/H 2 O ratio profiles: introduction and validation of an innovative retrieval approach. Atmospheric Chemistry and Physics, 6(12), pp.4705-4722.

Schneider, T., O'Gorman, P.A. and Levine, X.J., 2010. Water vapor and the dynamics of climate changes. Reviews of Geophysics, 48(3).

Schneider, T., Teixeira, J., Bretherton, C.S., Brient, F., Pressel, K.G., Schär, C. and Siebesma, A.P., 2017. Climate goals and computing the future of clouds. Nature Climate Change, 7(1), pp.3-5.

Schumacher, R.S. and Johnson, R.H., 2005. Organization and environmental properties of extreme-rain-producing mesoscale convective systems. Monthly weather review, 133(4), pp.961-976.

Schumacher, R.S., 2017. Heavy rainfall and flash flooding. In Oxford research encyclopedia of natural hazard science.

Seidel, D.J., Berger, F.H., Diamond, H.J., Dykema, J., Goodrich, D., Immler, F., Murray, W., Peterson, T., Sisterson, D., Sommer, M. and Thorne, P., 2009. Reference upper-air observations for climate: Rationale, progress, and plans. Bulletin of the American Meteorological Society, 90(3), pp.361-369.

Sherwood, S.C., Roca, R., Weckwerth, T.M. and Andronova, N.G., 2010. Tropospheric water vapor, convection, and climate. Reviews of Geophysics, 48(2).

Shimozono, T., Tajima, Y., Kumagai, K., Arikawa, T., Oda, Y., Shigihara, Y., Mori, N. and Suzuki, T., 2020. Coastal impacts of super typhoon Hagibis on Greater Tokyo and Shizuoka areas, Japan. Coastal Engineering Journal, 62(2), pp.129-145.

Shoji, Y., Kunii, M. and Saito, K., 2009. Assimilation of nationwide and global GPS PWV data for a heavy rain event on 28 July 2008 in Hokuriku and Kinki, Japan. Sola, 5, pp.45-48.

Singh, R., Ojha, S.P., Puviarasan, N. and Singh, V., 2019. Impact of GNSS signal delay assimilation on short range weather forecasts over the Indian region. Journal of geophysical research: atmospheres, 124(17-18), pp.9855-9873.

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Liu, Z., Berner, J., Wang, W., Powers, J.G., Duda, M.G., Barker, D.M. and Huang, X.Y., 2019. A description of the advanced research WRF version 4. NCAR tech. note ncar/tn-556 str, 145.

Solheim, F.S. and Ware, R.H., University Corp for Atmospheric Research UCAR, 1997. Atmospheric water vapor sensing system using global positioning satellites. U.S. Patent 5,675,081.

Steppeler, J., Hess, R., Schättler, U. and Bonaventura, L., 2003. Review of numerical methods for nonhydrostatic weather prediction models. Meteorology and Atmospheric Physics, 82, pp.287-301.

Subedi, A., Ghimire, B.D., Nagarkoti, S. and Khadka, B., 2023. On the Dynamic Variability of Global Positioning System (GPS) derived Precipitable Water Vapour (PWV) over three regions of Nepal.

Suparta, W. and Alhasa, K.M., 2016. Modeling of tropospheric delays using ANFIS.

Takiguchi, H., Kato, T., Kobayashi, H. and Nakaegawa, T., 2000. GPS observations in Thailand for hydrological applications. Earth, planets and space, 52(11), pp.913-919.

Temam, R., 1995. Navier–Stokes equations and nonlinear functional analysis. Society for industrial and applied mathematics.

Torcasio, R.C., Mascitelli, A., Realini, E., Barindelli, S., Tagliaferro, G., Puca, S., Dietrich, S. and Federico, S., 2023. The impact of GNSS Zenith Total Delay data assimilation on the short-term precipitable water vapor and precipitation forecast over Italy using the WRF model. Natural Hazards and Earth System Sciences Discussions, 2023, pp.1-29.

Torri, G., Adams, D.K., Wang, H. and Kuang, Z., 2019. On the diurnal cycle of GPS-derived precipitable water vapor over Sumatra. Journal of the Atmospheric Sciences, 76(11), pp.3529-3552.

Toth, Z. and Kalnay, E., 1997. Ensemble forecasting at NCEP and the breeding method. Monthly Weather Review, 125(12), pp.3297-3319.

Trenberth, K.E., Fasullo, J. and Smith, L., 2005. Trends and variability in column-integrated atmospheric water vapor. Climate dynamics, 24, pp.741-758.

Tyagi, A., Sikka, D.R., Goyal, S. and Bhowmick, M., 2012. A satellite based study of pre-monsoon thunderstorms (Nor'westers) over eastern India and their organization into mesoscale convective complexes. Mausam, 63(1), pp.29-54.

Vaisala. (2013). Radiosonde RS92-SGP: User's Guide. Vaisala Inc. Retrieved from https://www.vaisala.com

Van Baelen, J., Reverdy, M., Tridon, F., Labbouz, L., Dick, G., Bender, M. and Hagen, M., 2011. On the relationship between water vapour field evolution and the life cycle of precipitation systems. Quarterly Journal of the Royal Meteorological Society, 137(S1), pp.204-223.

Van Den Broeke, M.S., 2017. Polarimetric radar metrics related to tornado life cycles and intensity in supercell storms. Monthly Weather Review, 145(9), pp.3671-3686.

Vaquero-Martínez, J. and Antón, M., 2021. Review on the role of GNSS meteorology in monitoring water vapor for atmospheric physics. Remote Sensing, 13(12), p.2287.

Vaquero-Martinez, J., Anton, M., Chan, K.L. and Loyola, D., 2022. Evaluation of Water Vapor Product from TROPOMI and GOME-2 Satellites against Ground-Based GNSS Data over Europe. Atmosphere, 13(7), p.1079.

Wallace, J.M. and Hobbs, P.V., 2006. Atmospheric science: an introductory survey (Vol. 92). Elsevier.

Wang, B. and Johnson, R.H., 2006. Mesoscale processes. The Asian Monsoon, pp.331-356.

Wang, J. and Zhang, L., 2008. Systematic errors in global radiosonde precipitable water data from comparisons with ground-based GPS measurements. Journal of Climate, 21(10), pp.2218-2238.

Ware, R., Alber, C., Rocken, C. and Solheim, F., 1997. Sensing integrated water vapor along GPS ray paths. Geophysical Research Letters, 24(4), pp.417-420.

Wentz, F.J., Ricciardulli, L., Hilburn, K. and Mears, C., 2007. How much more rain will global warming bring? Science, 317(5835), pp.233-235.

Williams, E.R., 2017. Meteorological aspects of thunderstorms. In Handbook of Atmospheric Electrodynamics, Volume I (pp. 27-60). CRC Press.

Wilson, J.W., Crook, N.A., Mueller, C.K., Sun, J. and Dixon, M., 1998. Nowcasting thunderstorms: A status report. Bulletin of the American Meteorological Society, 79(10), pp.2079-2100.

Wu, P., Hamada, J.I., Mori, S., Tauhid, Y.I., Yamanaka, M.D. and Kimura, F., 2003. Diurnal variation of precipitable water over a mountainous area of Sumatra Island. Journal of Applied Meteorology and Climatology, 42(8), pp.1107-1115.

Wu, W., Emerton, R., Duan, Q., Wood, A.W., Wetterhall, F. and Robertson, D.E., 2020. Ensemble flood forecasting: Current status and future opportunities. Wiley Interdisciplinary Reviews: Water, 7(3), p.e1432.

Wu, Y., Jiang, N., Xu, Y., Yeh, T.K., Xu, T., Wang, Y. and Su, W., 2023. Improving the capability of water vapor retrieval from Landsat 8 using ensemble machine learning. International Journal of Applied Earth Observation and Geoinformation, 122, p.103407.

Yang, S., & Yuan, H. (2023). Acustomized multi-scale deep learningframework for storm nowcasting. Geophysical Research Letters,50, e2023GL103979. https://doi.org/10.1029/2023GL103979

Yao, Y., Shan, L. and Zhao, Q., 2017. Establishing a method of short-term rainfall forecasting based on GNSS-derived PWV and its application. *Scientific reports*, *7*(1), p.12465.

Yue, Y. and Ye, T., 2019. Predicting precipitable water vapor by using ANN from GPS ZTD data at Antarctic Zhongshan Station. Journal of Atmospheric and Solar-Terrestrial Physics, 191, p.105059.

Zhang, F., Minamide, M. and Clothiaux, E.E., 2016. Potential impacts of assimilating all-sky infrared satellite radiances from GOES-R on convection-permitting analysis and prediction of tropical cyclones. Geophysical Research Letters, 43(6), pp.2954-2963.

Zhao, Q., Liu, Y., Ma, X., Yao, W., Yao, Y. and Li, X., 2020. An improved rainfall forecasting model based on GNSS observations. IEEE Transactions on Geoscience and Remote Sensing, 58(7), pp.4891-4900.

Zhao, Q., Yang, P., Yao, W. and Yao, Y., 2019. Hourly PWV dataset derived from GNSS observations in China. Sensors, 20(1), p.231.

Zhu, Y. and Newell, R.E., 1998. A proposed algorithm for moisture fluxes from atmospheric rivers. Monthly weather review, 126(3), pp.725-735.

Zhu, Y., Toth, Z., Wobus, R., Richardson, D. and Mylne, K., 2002. The economic value of ensemble-based weather forecasts. Bulletin of the American Meteorological Society, 83(1), pp.73-84.

List of Publications

1. Bisht, D.S., Rao, T.N., Rao, N.R., Chandrakanth, S.V. and Sharma, A., "*Prediction of integrated water vapor using a machine learning technique"*, IEEE Geoscience and Remote Sensing Letters, 19, 1, 2022. (DOI: 10.1109/LGRS.2022.3217094)

2. Bisht, D.S., Rao, T.N., Rao, N.R. and Chandrakanth, S.V., "*Nowcasting of Storms Using Predicted Integrated Water Vapor with a Machine Learning Technique and Satellite Brightness Temperature"*, IEEE Transactions on Geoscience and Remote Sensing, 62, 5801608, 2024. (DOI: 10.1109/TGRS.2024.3429525)

3. Bisht, D.S., Rao, T.N. and Rao, N.R., "*Machine learning-based nowcasting of storm events: utilizing GNSS-based IWV, satellite and meteorological data for storm Nowcasting*". (preparing for submission to IEEE Transactions on Geoscience and Remote Sensing)

4. Chandrakanth, S.V., Rao, T.N, Gayatri V, **Bisht, D.S.**, "*Understanding the diurnal cycle of integrated water vapour over Gadanki region".* (manuscript ready to submit)

Reviewer Experience

1. Peer-reviewed a manuscript for publication in npj Climate and Atmospheric Science – Nature, 2024

Conferences Attended

1. IEEE India Geoscience and Remote Sensing Young Researchers Conclave 2020 IEEE Geoscience and Remote Sensing Society Kerala Chapter

Title: Now-casting of storms from GNSS-based precipitable water vapor data using machine learning approach

2. International Symposium on Tropical Meteorology on Changing Climate: Consequences and Challenges Indian Meteorological Society, Cochin Chapter

Title: A comprehensive study on machine learning approaches to storm nowcasting.

3. Presented **"Overview of AI/ML Work at NCMRWF"** at Momentum User Workshop.

4. Participated in **NVIDIA AI Summit 2024.**

Training/Schools Attended

1. **Workshop on Advances In GNSS Technologies & Applications (GNSS-18)**
   Research and Training Unit For Navigational Electronics Osmania University, Hyderabad, India.
2. **Workshop on Advanced GNSS Signal Processing (GNSS-18)**
   Research And Training Unit For Navigational Electronics Osmania University, Hyderabad, India.
3. **Workshop on Emerging Trends In GNSS Meteorology**
   Space Applications Centre (ISRO), Ahmedabad, India
4. **Workshop on NAVIC/GNSS: Technique and Applications**
   National Atmospheric Research Laboratory, Gadanki, India
5. Completed Certified course **"MOOC Machine Learning in Weather and Climate"** organized by ECMWF (online).

Talks and Lectures Delivered

1. Delivered a talk on **"AI/ML applications in Meteorology"** at National Centre for Medium Range Weather Forecasting.

2. Delivered two lectures on **“Machine Learning and Weather Forecasting”** to Scientific Staff at National Centre for Medium Range Weather Forecasting.

3. Delivered a lecture on **“Artificial Intelligence and its application in atmospheric sciences”** to Kolkata University Students.

4. Conducted a workshop on **"Python and Machine Learning"** for students of Kolkata University, organized at NCMRWF.

5. Volunteered and participated in **“Workshop on commemorating 30 years of operational NWP and IMD’s 150 Years”**