
Masters Theses

Student Theses and Dissertations

Spring 2023

Tropospheric Correction for INSAR using Machine Learning

Ngo Hi Kenny Yue

Missouri University of Science and Technology

Follow this and additional works at: https://scholarsmine.mst.edu/masters_theses



Part of the [Geological Engineering Commons](#)

Department:

Recommended Citation

Yue, Ngo Hi Kenny, "Tropospheric Correction for INSAR using Machine Learning" (2023). *Masters Theses*. 8155.

https://scholarsmine.mst.edu/masters_theses/8155

This thesis is brought to you by Scholars' Mine, a service of the Missouri S&T Library and Learning Resources. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact scholarsmine@mst.edu.

TROPOSPHERIC CORRECTION FOR INSAR USING MACHINE LEARNING

by

NGO HI KENNY YUE

A THESIS

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE IN GEOLOGICAL ENGINEERING

2023

Approved by:

Dr. Jeremy Maurer, Advisor

Dr. Ryan Smith

Dr. Ardhendu. Tripathy

© 2023

Ngo Hi Kenny Yue

All Rights Reserved

ABSTRACT

Interferometric Synthetic Aperture Radar (InSAR) is a popular technique for studying Earth's surface deformation caused by phenomena like earthquakes and subsidence. However, its accuracy is limited by tropospheric delays caused by water vapor in the atmosphere. This limitation can be overcome by using methods that correct for tropospheric noise, such as statistical, empirical, and predictive approaches. This study explores the potential of using machine learning algorithms to predict the zenith total delay caused by tropospheric effects in InSAR measurements. The study employs two different machine learning algorithms, random forest, and neural networks, to learn the relationship between numerical weather prediction model data and InSAR parameters in Continental USA and the globe. The neural network model outperforms both the random forest model and the traditional approach, reducing the RMSE by approximately 30%. The study demonstrates that machine learning algorithms can effectively correct tropospheric noise in most interferograms, resulting in a 30-60% improvement in Pennsylvania and Hawaii. However, the neural network model faces challenges in making predictions in areas with high variability in local climate and weather patterns. Overall, this research presents a promising approach for improving InSAR accuracy by using machine learning algorithms to correct for tropospheric noise.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to the following individuals.

Dr. Jeremy Maurer, for providing me with the opportunity to work on this thesis and for his guidance and mentorship throughout the process.

Dr. Ryan Smith and Dr. Ardhendu. Tripathy giving me valuable comments and guidance as my committee.

Joyce Hsieh, for her unwavering support and belief in me.

Everyone in GEM-lab who have provided me with valuable feedback and support.

TABLE OF CONTENTS

	Page
ABSTRACT.....	iii
ACKNOWLEDGMENTS	iv
LIST OF ILLUSTRATIONS.....	viii
LIST OF TABLES.....	x
NOMENCLATURE	xi
 SECTION	
1. INTRODUCTION	1
1.1. PROBLEM.....	1
1.2. TROPOSPHERIC CORRECTION METHODS.....	2
1.3. ZENITH DELAY.....	4
1.4. MOTIVATION	5
1.5. PROPOSED APPROACH.....	6
2. DATASET	8
2.1. NWP MODELS	8
2.2. GNSS DATA	10
2.3. TIMEFRAME	12
2.4. DATA VALIDATION.....	13
2.5. ACCURACY OF THE DATA	16
2.6. DATA EXTRACTION AND PREPROCESSING	17
3. METHOD.....	21
3.1. RANDOM FOREST MODEL.....	21

3.1.1. The Concept of Random Forest	21
3.1.2. Implementation of Model	22
3.2. ARTIFICIAL NEURAL NETWORK.....	23
3.2.1. Concept of Neural Network	24
3.2.2. Gradient Descent.....	28
3.2.3. Learning Rate.....	29
3.2.4. Loss Function.....	30
3.2.5. Implementation	31
3.3. COMPUTATION OF TROPOSPHERIC DELAY WITH TRADITIONAL METHOD	33
4. RESULT.....	35
4.1. COMPARISON RESULTS BETWEEN MODELS	35
4.1.1. CONUS Model	35
4.1.2. Global Model	38
4.2. CORRELATION BETWEEN PARAMETERS AND ERROR.....	44
4.2.1. Elevation	44
4.2.2. Latitude.	45
4.3. RANDOM FOREST FEATURE IMPORTANCES.....	46
4.4. COMPARE PREDICTION WITH TRADITIONAL METHOD.....	47
4.5. COMPARE PREDICTION BY STATIONS	50
4.5.1. IABL Station.....	50
4.5.2. LINC Station.....	53
4.6. APPLICATION TO INSAR.....	56
4.6.1. Background of Test Site.....	56

4.6.2. Implementation	58
5. DISCUSSION	65
5.1. GNSS BIAS	65
5.2. WEATHER MODEL TIME DEPENDENCY	66
5.3. FUTURE SCOPE	69
5.3.1. A Different Machine Learning Approach.....	69
5.3.2. Using Weather Model with Higher Spatial Resolution	70
5.3.3. Implement of GOES Data.....	70
6. CONCLUSION	72
APPENDIX.....	74
REFERENCE.....	75
VITA.....	80

LIST OF ILLUSTRATIONS

	Page
Figure 2.1. Illustration of weather model parameters in 3D cube.	9
Figure 2.2. ZTD estimation from a single GPS station (Benevides. 2013)	11
Figure 2.3. The distribution of GNSS station around the globe with ZTD and ZTD with elevation and latitude on 01-01-2017 at 00:00:00 UTC.	12
Figure 2.4. Plot of a subset of semivariogram in pressure at 3 different elevations.	14
Figure 2.5. Variogram of weather model parameters, hydrostatic, and wet delay	15
Figure 2.6. ZTD (m) versus elevation (m).	18
Figure 3.1. Illustration of a single neuron (source from https://www.kaggle.com/code/ryanholbrook/a-single-neuron)	25
Figure 3.2. Graphical representation of the ReLU activation function above	27
Figure 3.3. Illustration of gradient descent from O'Reilly	28
Figure 3.4. Comparison of different learning rate.	30
Figure 3.5. Mean square error during training.	32
Figure 4.1. Residual for CONUS RF model prediction.	35
Figure 4.2. Residual for CONUS NN model prediction.	36
Figure 4.3. CONUS NN model prediction RMSE.	37
Figure 4.4. CONUS model prediction average error.	37
Figure 4.5. Residual plot of Random Forest result.	38
Figure 4.6. Residual plot of Neural Network result.	39
Figure 4.7. Histogram of RMSE by GNSS station from RF and NN model.	39
Figure 4.8. Globe NN model prediction average error.	41

Figure 4.9. Globe NN model prediction RMSE	41
Figure 4.10. Global RF model prediction average error.....	42
Figure 4.11. Global RF model prediction RMSE.....	42
Figure 4.12. Global WM prediction average error.....	43
Figure 4.13 Global WM prediction RMSE.....	44
Figure 4.14. Scatter plot between RMSE (m) and elevation of stations with comparison between NN model and RF model.....	45
Figure 4.15. Scatter plot between RMSE and latitude of stations with comparison between NN model and RF model.....	46
Figure 4.16. Compare RMSE from between model prediction and GNSS true ZTD of each station and plot by month.....	50
Figure 4.17. Histogram compare of RMSE by month for station IABL between each model prediction.....	51
Figure 4.18. Time series of predictions for station IABL.....	52
Figure 4.19. Residuals plot of station IABL compare with three model.....	53
Figure 4.20. Monthly RMSE for station LINC.....	54
Figure 4.21. Time series of prediction for station LINC.....	54
Figure 4.22. Residuals plot of station LINC compare with three model.....	55
Figure 4.23. Topography of study area Pennsylvania and Hawaii.....	57
Figure 4.24. Tropo-correction for interferogram in Pennsylvania with NN model and RAiDER.....	59
Figure 4.25. Tropospheric correction for interferogram in Hawaii with NN model and RAiDER.....	62
Figure 5.1. Distribution of GNSS station by latitude.....	66
Figure 5.2. Tropospheric delay changes in the increment of 10 mins.....	67
Figure 5.3. GOES-17 band 10 imager showing water vapor content.....	71

LIST OF TABLES

	Page
Table 1.1. Comparison of state-of-the-art atmospheric correction method	4
Table 2.1. Training and testing data distribution.	19
Table 3.1. Architecture of model	31
Table 3.2. Hyperparameters	32
Table 4.1. Importance of feature and weightings	47
Table 4.2. Comparison between random forest, neural network, and traditional method	48
Table 4.3. Compare of RMSE with different methods by date (Pennsylvania).	60
Table 4.4. Percentage improvement from original std (Pennsylvania).....	61
Table 4.5. Compare of RMSE with different methods by date (Hawaii).	63
Table 4.6. Percentage improvement from original std (Hawaii)	64
Table 5.1. Compare correction without interpolation in time in Hawaii.....	68

NOMENCLATURE

Symbol	Description
P	Pressure
T	Temperature
e	Partial pressure of water vapor
STD	Slant Total Delay
ZTD	Zenith Total Delay
NWP	Numerical Weather Prediction
θ	incidence angle from radar satellite
LOS	Line Of Sight

1. INTRODUCTION

1.1. PROBLEM

Interferometric Synthetic Aperture Radar (InSAR) is a remote sensing technique that can accurately measure surface deformation on the Earth's crust. InSAR has been extensively used to study various geophysical phenomena such as earthquakes (Barnhart & Lohman, 2013, Funning et al., 2005), subsidence (Buzzanga et al., 2020), volcanoes (Albino et al., 2020), and others. However, there are limitation into wide-ranging applicability of InSAR due to the accuracy of its measurements which is affected by atmospheric delays. Atmospheric delays consist of both ionospheric and tropospheric. Ionospheric delays are triggered by change of free electrons along the signal's path, leading to a phase advancement. However, ionosphere is comparatively easier to model and have been resolved with empirical method (Liang et al. 2019) hence, this thesis will only focus on the tropospheric component. Tropospheric delay arises from variations in the refractive index of the atmosphere (troposphere) and can cause significant errors in Synthetic Aperture Radar (SAR) satellite measurements, especially for long-wavelength deformation signals. Therefore, it is crucial to correct for the tropospheric delay in InSAR measurements to obtain accurate and reliable results. Tropospheric delays are caused by spatio-temporal changes in atmospheric refractivity related to pressure, temperature, and water vapor. It can cause up to 15–20 cm of interferogram signals (Zebker et al., 1986, Rosen et al., 1999, Massonnet et al., 1998, Hanssen, 2001, Hooper and Bekaert 2014). Some 20 % changes in the water vapor content in the atmosphere could result with delay of 10 cm (Zebker et al.,

1997) and can greatly obscure the other signal of interest such as crustal deformation and other ground movement. Tropospheric delay remains challenging to address.

1.2. TROPOSPHERIC CORRECTION METHODS

There are three main categories of tropospheric correction methods used in InSAR: empirical, statistical, and predictive corrections. Empirical corrections rely on analyzing a significant amount of InSAR data and are usually applied on a per-case basis. Statistical corrections use the information from the interferometric phase and other data sources, such as ground-based GPS measurements, to estimate the tropospheric delay. Predictive corrections utilize atmospheric models, such as Numerical Weather Prediction (NWP) models, to predict the tropospheric delay based on atmospheric parameters, such as temperature, pressure, and water vapor.

The process of empirical correction eliminates the correlation between the phase values in interferograms and changes in topography (Lin et al. 2010; Bekaert et al. 2015; C. W. Wicks et al. 2002; Zebker, 2021). This technique is advantageous because it does not depend on external data and can be easily computed. However, it requires consistently correlated high-quality pixels (Zebker, 2021), and it may not account for fluctuations in atmospheric water vapor due to turbulence (Massonnet and Feigl, 1995; Hanssen, 2001).

Another class of methods, statistical correction, uses time averaging to reduce the tropospheric phase lag from time series InSAR (Massonnet and Feigl, 1995; Zebker et al., 1997) and stacking (Rosen et al., 2018; Sandwell & Sichoix, 2000) are two examples of this class of methods. Recently, methods based on least squares have been proposed to estimate and remove tropospheric noise using shared data between multiple interferograms

(Kang et al., 2021; Kirui et al., 2022; Tymofyeyeva & Fialko, 2018). These methods are very effective at removing tropospheric noise if the underlying deformation style (e.g., linear deformation rate) is known, such as slow slip events.

Auxiliary data for prediction is the last major method for correcting phase delays in the troposphere. Examples of ancillary data include GNSS measurements of zenith total delay (Wadge et al., 2002; Onn and Zebker, 2006; Yu et al., 2018; Shamshiri et al., 2020), multispectral satellite observations of water vapor such as the Moderate Resolution Imaging Spectrometer (MODIS) onboard the Terra and Aqua satellites (Liang, 2014), and NWP such as the European Center for Medium-Term Forecasts (ECMWF) ERA-5 global NWP model. There are multiple open-source software packages and online services make use of NWP to provide tropospheric delays globally such as Python Based Atmospheric Phase Screen estimation (PyAPS) (Jolivet, Agram, et al., 2014), Direct integration method along the LOS direction (D-LOS) (Hu & Mallorquí, 2019), and Generic Atmospheric Correction Online Service for InSAR (GACOS) (Yu et al., 2018). However, the accuracy of such corrections depends on the spatial and temporal resolutions of auxiliary data or NWP model outputs (Cao et al., 2021). Bekaert et al (2015) reviewed of different state-of-the-art atmospheric correction method and it's briefly summarized in Table 1.1.

The development of NWP models has significantly improved our understanding of atmospheric behaviors. NWP models such as the North American Regional Reanalysis (NARR; Mesinger et al., 2006), the European Reanalysis – Interim (ERA-I; Dee et al., 2011), the Modern-Era Retrospective analysis for Research Applications (MERRA-2; Gelaro et al., 2017), and ECMWF-HRES (Haiden et al., 2018) could allow tropospheric estimation in wider range of area.

Table 1.1. Comparison of state-of-the-art atmospheric correction method

Method	Inputs	Limitation
Weather model	GAMS	High temporal resolution, but low spatial resolution of 30 -80 km
GPS measurement	GNSS ZTD	Not constant everywhere in Spatial and temporal resolution
Multi spectral observation	MODIS/ MERIS	Re-visit time of the instrument are not simultaneously acquired with SAR data and require interpolation in time
GPS + Spectrometer data	GNSS ZTD + Spectrometer data	Spectrometer only available under cloud-free and daylight
Empirical method	Topography (DEM) + Interferograms phase	Only in non-deforming region or in spatial band insensitive to deformation
Power-law model	Topography (DEM) + Interferograms phase	An estimate of spatially varying tropospheric signal that require defined boundary of stagnation

1.3. ZENITH DELAY

The zenith total delay in the signal path between a GNSS receiver and satellite, also known as the path delay, is primarily dependent on the satellite's position. To simplify calculations and create a more generalizable model, empirical models often use the zenith

total delay (ZTD) and reproject to the slant total delay (STD). To convert the ZTD to STD, a mapping function is applied. The mapping function is a simple equation as equation (1):

$$ZTD = \frac{STD}{\cos(\theta)} \quad (1)$$

Where θ refers to the incidence angle to the satellite. Zenith delay is only discussed in model prediction for this thesis. Slant delay is used when discuss of the InSAR application.

1.4. MOTIVATION

Machine learning (ML) has been extensively adopted in various fields in the past two decades due to the advancements in cheaper and more powerful computing power and sophisticated algorithms. ML algorithms enable computers to extract complex relationships from sizeable volumes of data, and the growing of geospatial data being generated in recent years has created new opportunities to understand and deliver insights from large, high-dimensional datasets. ML has the potential to develop new methods for processing and analyzing large, complex datasets where traditional methods may fail, and it improves accuracy with short computation times compared to empirical approaches.

In estimating of tropospheric delay, previous studies have shown successful application with machine learning. For example, Shamsiri et al.(2020) used machine learning Gaussian process (GP) regression and GNSS derived zenith total delay (ZTD) to predict phase delay trigger by troposphere. By applying the ML technique, the root mean square error (RMSE) of zenith total delay prediction reduces on average by 83%, compared to a 50% reduction obtained by using traditional approaches. In another study, Selbesoglu (2020) developed an artificial neural network model with weather station and GNSS data

to estimate tropospheric wet delay. Their model was able to predict the zenith wet delay for up to six hours with a root mean square error (RMSE) of 1.5 cm.

Instead of estimating the ZTD with a fixed equation, machine learning approach may be an alternative in leveraging product derived from NWP models which are globally and hourly available. NWP models is not limited spatially as the GNSS ZTD and could possibly be used in areas where no GNSS station available. The trained model will be beneficial to provide an alternative correction method for InSAR processes which result with getting accurate subsurface deformation measurement.

1.5. PROPOSED APPROACH

The thesis aims to investigate the capability of machine learning techniques in predicting the zenith total delay (ZTD) at GNSS stations based on NWP model parameters and geographic coordinate information. A traditional approach that utilizes the same NWP model parameters to estimate ZTD is set as a baseline model for this study. This traditional approach involves raytracing of refractivity, which will be discussed further in the method section.

The necessary data is divided into two separate datasets based on time: a training and testing set. Two machine learning algorithms are employed to learn the relationship between NWP parameters and ZTD from the training set. The prediction which evaluate the performance of the model uses the testing set only.

The first algorithm used is random forest, which is commonly used for classification and regression problems. It is utilized as a reference for the development of the machine learning model due to its ease of implementation and interpretability. The

second algorithm used is a fully connected neural network, which can solve more complex problems when larger datasets are available.

The models' performance is evaluated by comparing the predicted delays with the actual delays at GNSS stations to determine if there was any improvement over the baseline model. The desired target accuracy of the prediction is set to be within 2 cm in root mean square error (RMSE) from all models. The models are trained on two datasets based on the geographic boundary: one that includes data from the Continental US and the other that includes data from the globe.

The accuracy of the prediction is further compared with any dependency variables such as elevation, latitude, and time of year. Two examples of individual GNSS stations are chosen for in-depth analysis.

After evaluating the model's accuracy, in preparation for tropospheric correction for InSAR, the optimized model will perform tropospheric estimates to make corrections at two locations in the US: Pennsylvania and Hawaii. These two locations represent different circumstances and environments, and the test provides a better understanding of the model's capabilities. Several date pairs of interferograms at these locations are used to showcase the correction power by calculating the change in standard deviation.

2. DATASET

Data of NWP models and GNSS ZTD are download through RAiDER package. GitHub repository is linked as follow (<https://github.com/dbekaert/RAiDER>)

2.1. NWP MODELS

ERA-5 is a weather model developed by the European Centre for Medium-Range Weather Forecasts (ECMWF). It is the fifth-generation reanalysis dataset that provides global atmospheric conditions from 1979 to present. The ERA-5 model uses a state-of-the-art numerical weather prediction (NWP) system to assimilate a wide range of observations, including surface weather stations, radiosondes, and satellite measurements, into a comprehensive and consistent atmospheric model (Hersbach, H et al. 2020).

The ERA-5 model provides a range of atmospheric variables at various spatial and temporal resolutions. These variables include temperature, wind speed, humidity, precipitation, cloud cover, and other parameters that describe the state of the atmosphere. The model has a horizontal resolution of 31 km and 137 vertical levels, providing a detailed picture of atmospheric conditions from the surface up to 80 km altitude.

The ERA-5 model also includes a variety of auxiliary data products, such as surface solar radiation and soil moisture, which can be used for various applications, including weather forecasting, climate monitoring, and research. The outputs that are of interest are Pressure (in Kelvin), Pressure (in Pa), and partial pressure of water vapor (in Pa).

The high spatial and temporal resolution of the ERA-5 model makes it an essential tool for various weather-related applications, such as predicting extreme weather events, studying the impact of climate change, and supporting renewable energy production. The

ECMWF continues to improve the ERA-5 model, and updates are released regularly to provide the most accurate and reliable atmospheric data possible.

We used the output of weather model as the input of our deep learning model. These data are uniformly gridded as a 3D cube (netCDF) and referenced with the associate latitude, longitude, and elevation. (Figure 2.1)

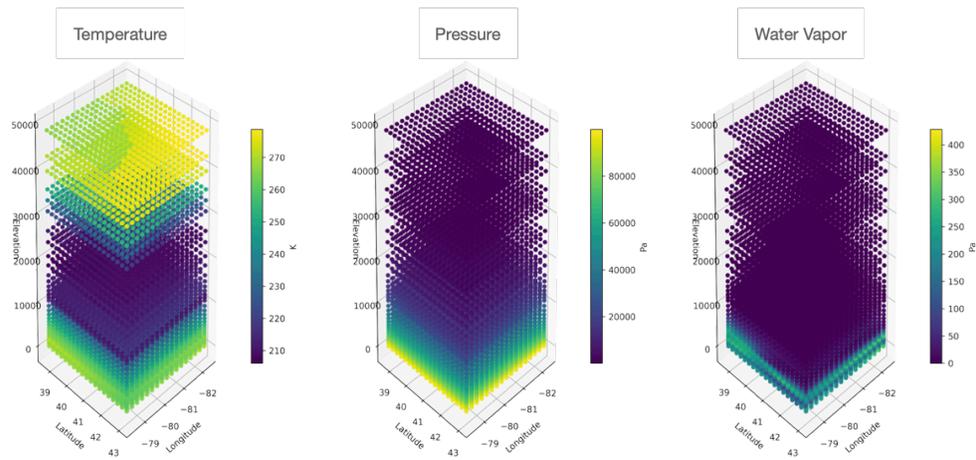


Figure 2.1. Illustration of weather model parameters in 3D cube.

The ERA5 model is created using over 240 input parameters derived from a wide range of datasets, including satellite, in-situ, and surface elevation data. These datasets include information from sources such as stratospheric sounding units, infrared and microwave sounders, all-sky microwave radiances, GPS radio occultation, geostationary radiances, atmospheric motion vectors, scatterometers, WAVE integrated parameters, ground-based radar, and conventional methods such as meteorological stations, aircraft, ships, buoys, and radiosondes. These input parameters contribute to generating global analysis models of atmospheric variables, including pressure, temperature, partial pressure of water vapor, humidity profile, cloud liquid water, precipitation, ocean surface wind

speed, surface and cloud top temperature, wind vectors, ozone, backscatter, soil moisture, rain rates, snow depth, and snow cover. (ERA5, 2023)

The model is freely available to the public and can be accessed through the ECMWF data archive via C3S Climate Data Store.

2.2. GNSS DATA

GNSS stations are ground-based receivers that collect signals from Global Navigation Satellite Systems (GNSS) such as GPS, GLONASS, Galileo, and BeiDou. These stations collect precise positioning and timing data, which can be used for a variety of applications, including navigation, surveying, and geophysics.

One of the most important applications of GNSS stations is in the estimation of Zenith Total Delay (ZTD) products. Delay occurs when radar signals pass through the atmosphere before reaching the receiver and is caused by atmospheric variables such as temperature, pressure, and partial pressure of water vapor. (Yu et al. 2021).

To generate ZTD products, GNSS stations collect data on the signals received from the GNSS satellites and the station's position. At a GNSS station, the receiver collects signals from multiple satellites at various slant angles. As shown in Figure 2.2, the illustration depicts an upside-down cone on top of a GPS station. To calculate ZTD, the slant delays for multiple GPS satellites are measured and then combined using a process called precise point positioning (PPP) or a similar approach. PPP allows for precise determination of the position of a GPS receiver by using GPS measurements and precise satellite orbit and clock information (Benevides. 2013).

By combining the slant delays from multiple GPS satellites, the ZTD can be determined with a high degree of accuracy. The resulting ZTD value represents the total delay in the GPS signal as it passes through the Earth's atmosphere from the GPS satellite to the GPS receiver at the zenith.

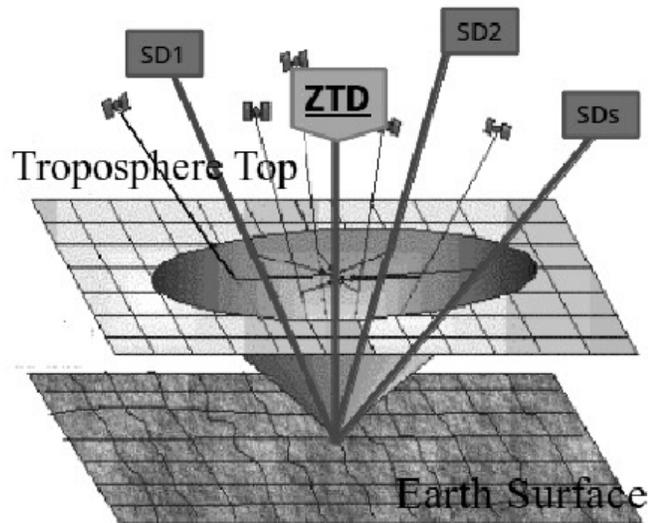


Figure 2.2. ZTD estimation from a single GPS station (Benevides. 2013)

GNSS ZTD data is collected from the Raytracing Atmospheric Delay Estimation for RADAR (RAiDER) package, it is created by Caltech as to calculate tropospheric correction for Radar. In supplement it also provides access to the Nevada Geodetic Lab to retrieve data by region or location with a 5-minute temporal resolution at locations around the globe (Blewitt, 2018). In this thesis, we focus on the area of the US continent and the globe between 50N 68.5W 30N 125W with a total of ~8,000 GNSS stations selected at 11:00:00 UTC and 89N 180W 89S 180E with a total of ~17,000 stations selected at 00:00:00 UTC, respectively. The data are taken from the years 2017 to 2022. Not all stations have continuous data available for analysis; therefore, some stations consist of discontinued data.

The distribution of GNSS stations around the globe is not evenly spread across every continent. Referring to Figure 2.3, most GNSS stations are in North America, Europe, and Japan. As shown in Figure 2.3, zenith total delay (ZTD) correlates with changes in elevation and latitude. ZTD increases as latitude approaches the equator (0°) and at locations closer to sea-level ($< 1000\text{m}$). As the ray propagates through a thinner troposphere, the effect on the delay is smaller.

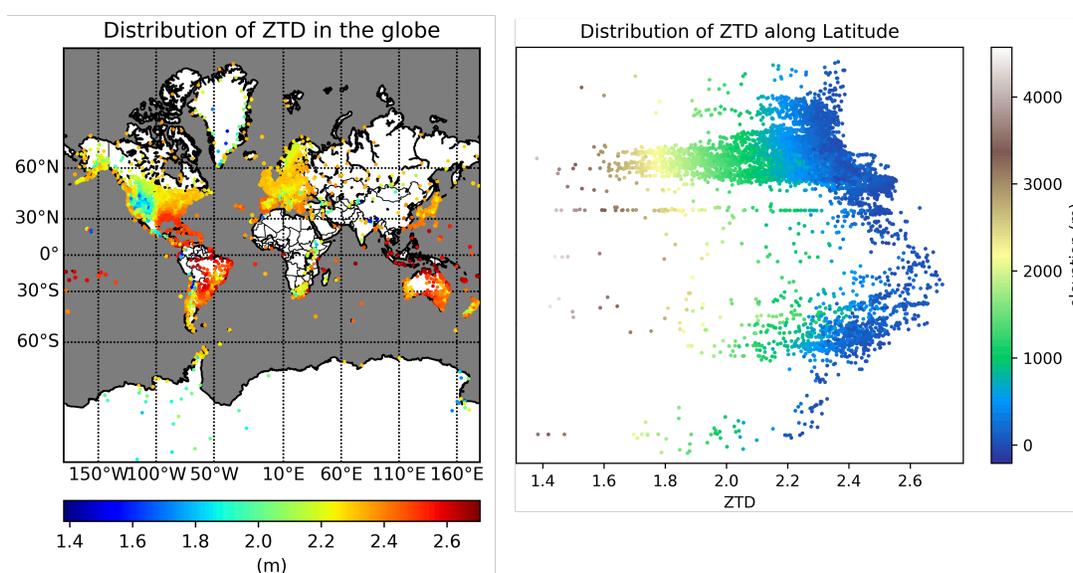


Figure 2.3. The distribution of GNSS station around the globe with ZTD and ZTD with elevation and latitude on 01-01-2017 at 00:00:00 UTC.

2.3. TIMEFRAME

The timeframe for the data ranges over six years from the beginning of 2017 to the end of 2022. Although ZTD data from GNSS stations have a higher temporal resolution, it was decided to match the timeframe with the NWP model's temporal resolution to keep consistency. However, not all GNSS stations essentially have continuous data for long periods of time, which may result in some data discontinuity due to no available data on some date or time. For NWP model, due to the nature of a model it is generally not an

issue, as each datetime and location can be treated as a singular observation and dismissed if needed. For this thesis, the US continent model takes daily measurements, and for the global model, it takes measurements at 15-day intervals for training with the respective ML model.

2.4. DATA VALIDATION

To determine if we can improve spatial resolution of ERA-5 output for the machine learning algorithms, we checked if it was relevant, representative, and comprehensive. The data underwent a geostatistical analysis to determine the spatial correlation between ERA-5 outputs and ZTD.

We first sampled the NWP model between 51N 120W 48N 123W at pressure levels of 400m, 1000m, and 4000m. Then, we calculated the distance between points and grouped them by distances to measure the variance between the response variables. This generates semi-variograms that provide information on the variation change with increasing lag distance. In general, points that are closer will have higher covariance/more similarity than points that are farther away (see Figure 2.4).

Variograms consist of measurements of the variance and covariance of weather parameters, such as pressure, temperature, and partial pressure of water vapor content, in comparison to hydrostatic delay and wet delay derived from RAiDER. The variance of pressure, temperature, and partial pressure of water vapor tends to have high variation in the shorter range at lower elevations (~400m), but the variance tends to decrease as altitude increases.

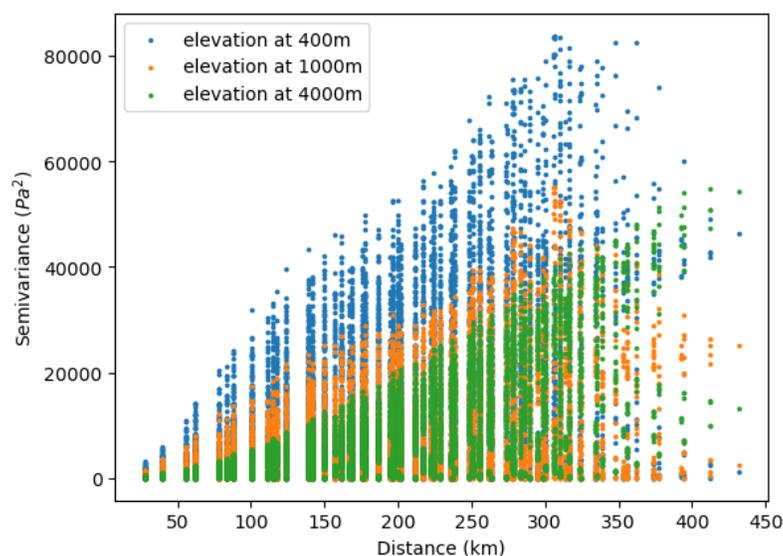


Figure 2.4. Plot of a subset of semivariogram in pressure at 3 different elevations.

Variograms are an essential tool in geostatistics, as they help us understand the spatial correlation between variables. Specifically, they provide information about how the variance of a particular variable changes as a function of distance between locations. In the context of this study, the variograms help us to evaluate the spatial correlation between the ERA-5 outputs. Figure 2.5 illustrates the variability of NWP model products. Figure 2.5a reflects of the correlation between elevation change and wet delay, which is similar to the partial pressure of water vapor (Figure 2.5b). This indicates that wet delay is reliant on the partial pressure of water vapor. Partial pressure of water vapor, temperature, and pressure exhibit distinct spatial variations (Figure 2.5c-e). Although partial pressure of water vapor and pressure display similar variability at different altitudes, their correlation distance differs, with partial pressure of water vapor being 200 km and pressure being over 300 km. At a point of 49°N and 122°W, the changing in partial pressure of water vapor versus elevation (Figure 2.5b) indicates most of the variation in water vapor is in the area close to

the surface of the earth. (Lower troposphere <10km). This is consistent with the estimated spatial variation of water vapor partial pressure at distinctive heights (Figure 2.5c), which demonstrate a significant variability in partial pressure of water vapor in the portion of the atmosphere that is closer to the Earth's surface with a variance of around 12000 Pa² at 400 m, compared to 4000 Pa² at 4000 m.

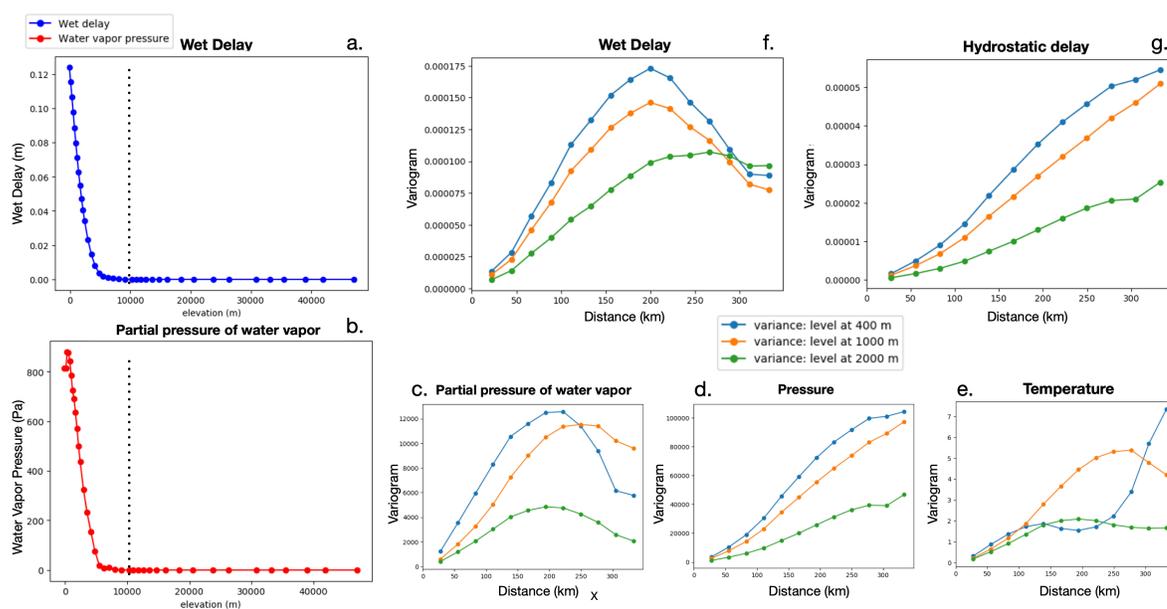


Figure 2.5. Variogram of weather model parameters, hydrostatic, and wet delay

The results of geostatistical analysis demonstrate a limitation of using interpolation for NWP model output. Variograms indicate that values cannot be interpolated from 0 km due to the regular gridded spatial resolution of ERA-5 (~30 km). Interpolation performed between grid points of the NWP model results in smoothing out of data due to the absence of additional information. Previous research by Cao et al. (2021) attempted to address this issue by applying a 3D interpolation scheme for NWP model (ERA5) parameters. Their approach increased the resolution of the NWP model parameters from 30 km to 2.5 km,

followed by raytracing along the LOS. The study resulted in a noise reduction of 87% in interferograms after applying correction. However, the success of this approach may still be limited by the geostatistical analysis mentioned earlier. Thus, the effectiveness of this approach could be attributed to the calculation of raytracing along the line-of-sight (LOS) instead of increasing the resolution of the weather model output.

2.5. ACCURACY OF THE DATA

In GNSS data analysis, outliers can cause significant problems in data processing and affect the accuracy of the results. To address this issue, the Nevada Geodetic Laboratory provides an uncertainty factor called sigZTD for the ZTD data. However, it does not automatically remove the outliers from the data. Therefore, it is crucial to filter out the GNSS station acquisition with high uncertainty to ensure data integrity for further analysis. In this thesis, the threshold used for filtering is a sigZTD of 0.01 or higher. Removing these high-uncertainty data points can improve the overall accuracy of the results.

The ERA5 Numerical Weather Prediction (NWP) model's accuracy and uncertainty rely on various factors, such as the spatial and temporal resolution, physics parameterization schemes, and input data quality. With a spatial resolution of approximately 31 km and a temporal resolution of one hour, ERA5 is one of the most advanced NWP models available. Nevertheless, the accuracy of the model remains uncertain due to the inherent complexities of atmospheric processes. The accuracy may also fluctuate depending on the location and period studied, with regions with complex terrain or extreme weather conditions displaying lower accuracy than others. Moreover,

the quality of input data from satellites, radiosondes, and ground-based stations can impact the model's precision.

2.6. DATA EXTRACTION AND PREPROCESSING

The extraction of NWP model products is a crucial step in this study, as it forms the basis for subsequent analysis. To ensure that the data are representative of the atmospheric conditions at each station, a two-step interpolation process is employed. The first step involves vertical interpolation, which is performed by interpolating the NWP model output into 51 fixed pressure levels. The method used for interpolation is similar to that used by Cao et al. (2021), who found that the atmospheric refractivity beyond 40 km can be ignored, while it increases significantly towards the lowermost 5 km of the atmosphere. To account for this change in refractivity, different sampling steps are used for different altitude ranges (Figure 2.6), with finer sampling steps employed for the lower altitudes. The second step involves horizontal interpolation, which is performed by interpolating the NWP model output to the GNSS stations based on their geographic coordinates. This generates a comprehensive database comprising 155 parameters, including pressure, temperature, water vapor, latitude, and elevation. Each row in the database represents an individual observation for a specific station at a given date and time. This approach ensures that the data are spatially and temporally consistent, enabling accurate evaluation of the spatial correlation between the NWP model outputs. It is worth noting that the accuracy of the extracted NWP model products is dependent on the spatial and temporal resolution of the NWP model output and the accuracy of the interpolation

techniques employed. Therefore, careful consideration must be given to these factors when interpreting the results of this study.

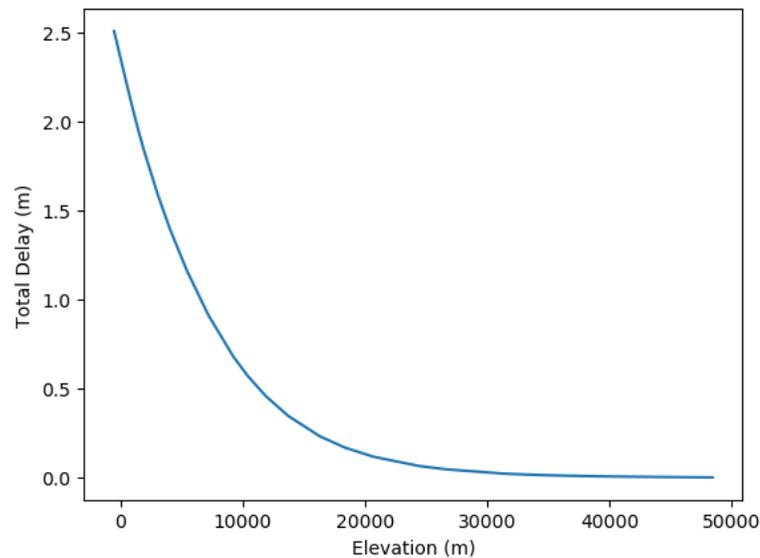


Figure 2.6. ZTD (m) versus elevation (m).

After extracting data from the NWP model, it needs to be processed into at least two different datasets (training and testing) before the training phase. The training dataset is the most crucial aspect of machine learning as it teaches the machine learning model about the relationship between input and labelled data. It trains the algorithms iteratively in batches along with the labelled data.

The validation dataset is used during the training process to monitor the progress of the training phase. The model sees this data throughout the training, but it doesn't learn from it. Instead, it's used as an indicator of overfitting and underfitting of the model. Overfitting occurs when the model is too complex and learns the noise within the data, resulting in a model that is not generalizable to new data. Underfitting occurs when the model is too simple and does not learn the underlying patterns in the data.

The testing dataset is used to evaluate the performance and accuracy of the model. It's a sample of data that is not included in the training of the model and gives an unbiased evaluation of the final model's performance. Testing and validation dataset are usually shared because they are excluded from training.

To ensure fair training without data leakage, the data is split in time. The training dataset consists of the years in 2017-2020, while the validation / testing datasets are from the year in 2021-2022. This helps ensure that the model is generalizable and can make accurate predictions on new data.

It's important to note that the size of each dataset may vary depending on the amount of data available and the complexity of the problem. Additionally, it's crucial to ensure that the datasets are balanced and representative of the overall population.

Table 2.1. Training and testing data distribution.

	US model	Global model
Training	5,644,952	1,082,502
Testing	2,086,682	457,480
Total	7,731,634	1,539,982

Feature scaling is a pre-processing technique used in machine learning to normalize the data features to a similar scale. Machine learning models rely on numerical features to make predictions and they may have different units or scales. By scaling the features to a similar range, we can ensure that no feature dominates over the others in influencing model predictions. Feature scaling can also improve the convergence of machine learning algorithms, such as gradient descent, by allowing them to reach the optimal solution more

quickly. Common scaling methods include standardization (scaling to zero mean and unit variance) and normalization (scaling to a range between 0 and 1). In this thesis, we employ normalization as our preprocessing step instead of standardization. The reason for this is that the weather model parameters at each altitude level do not follow a normal distribution. Standardization may not accurately represent the data after scaling.

Normalization is also known as min-max scaling, it is done by having each feature subtracting the minimum value from each column of data and divided by its range (max – min). The value in a column is bounded between a fixed range of 0 and 1.

$$feature_{norm} = \frac{(feature - feature_{min})}{(feature_{max} - feature_{min})} \quad (2)$$

3. METHOD

3.1. RANDOM FOREST MODEL

Random Forest is a type of ensemble learning algorithm that combines multiple decision trees to make predictions. In this approach, each tree is built using a random subset of the training data and features, resulting in a diverse set of trees that collectively make more accurate and stable predictions than a single decision tree. During prediction, the algorithm aggregates the output of all trees to generate the final prediction.

3.1.1. The Concept of Random Forest. A decision tree is a simple machine learning algorithm that can be used for both regression and classification. It has been applied in various fields such as finance, medicine, marketing, and engineering (Zheng et al., 2022, Liu et al., 2022). The algorithm builds a tree-like model in which each internal node represents a test on a particular feature, each branch represents the outcome of that test, and each leaf node represents a class label or a regression value. Decision tree uses splitting criteria to divide the dataset into subsets based on the value of an attribute. Each split taking the model closer to the final decision.

To evaluate the quality of the split, metrics such as mean square error (MSE) are used. decision trees are prone to overfitting and can be sensitive to noise and outliers, fitting strongly with the given dataset and failing to generalize when extracting complex relations in datasets. Random forest, on the other hand, predicts results with greater precision when multiple decision trees form a committee, particularly when the individual trees have no correlation with one another. The random forest algorithm is an extension of the bagging

method and uses feature randomness in creating an uncorrelated forest of decision trees.

The key concepts are as follows:

- **Bagging (Bootstrap Aggregation):** Since decision trees are sensitive to the data they are trained on; bootstrapping is done by randomly sampling the whole training data with replacement. This results in tree structures that can vary significantly even with relatively insignificant changes to the training set.
- **Feature Randomness:** This forces variation within the decision trees in the model and generates low-correlation and high-diversity models to minimize possible error from a single decision tree model.
- **Aggregating:** Each decision tree model is a prediction of the forest. To make a prediction with the forest, the mean of all tree's predictions is taken as the model output.

3.1.2. Implementation of Model. This thesis employed random forest consisting of 100 trees for the dataset described earlier. It was discovered that adding more estimators (trees) did not increase the accuracy of the model, but it did lead to an increase in computational time. Inversely, using fewer trees had an adverse impact on prediction quality. The mean squared error (MSE) was used to assess the suitability of features or thresholds for splitting samples. While having more data is typically advantageous for machine learning algorithms, the random forest algorithm does not require as many samples as neural networks. It was also observed that enhancing the size of training samples lead to an increase in computational time and did not improve prediction accuracy.

3.2. ARTIFICIAL NEURAL NETWORK

Artificial Neural Network (ANN) is a type of machine learning algorithm that is inspired by the structure and function of biological neural networks. It has become increasingly popular in the past decade, especially in geological scientific research, due to its ability to analyze and learn from complex data sets. The basic building block of an ANN is a neuron, which receives inputs, performs a computation, and produces an output. In an ANN, multiple neurons are connected to form layers, and layers are stacked together to form a network. ANNs or neural networks are fundamental processes using statistical methods for training computational algorithms that split, sort, and transform a set of data to maximize the ability to classify, predict clusters, or discover patterns in a target dataset (Reichstein et al. 2019).

ANNs can be used for both classification and regression problems. During the training process, the weights of the connections between neurons are regulated based on the difference between the predicted output and the actual output, using an optimization algorithm such as backpropagation. The objective is to minimize the error between the predicted output and the actual output, so that the network can generalize to new data.

To use ANN effectively, it is important to consider some key points, such as the network architecture. The number of layers and neurons in each layer can have a significant effect on the performance of the network. Choosing the right architecture for a given problem is important and can involve experimentation and tuning. Additionally, ANNs require a large amount of training data to learn the underlying patterns in the data. The quality and quantity of the data used for training can impact the accuracy of the model.

Another important consideration is regularization techniques. ANNs can easily overfit the training data, which can lead to poor generalization performance. Regularization techniques such as L1/L2 regularization, dropout, and early stopping can be used to prevent overfitting. Moreover, the choice of activation functions in the neurons can also impact the performance of the network. Common activation functions include sigmoid, rectified linear activation unit (ReLU), and tanh.

Interpretability is another important factor to consider. One potential drawback of ANN is that they can be difficult to interpret, especially for complex architectures. Techniques such as visualization and sensitivity analysis can be used to help understand the model's behavior and how it makes predictions.

ANN can solve a wide variety of problems with the help of a backpropagation learning algorithm. There are various types of ANN, and the most common type consists of three layers containing an input layer, a hidden layer, and an output layer. Nodes in each layer connect with another layer with weights and thresholds. A neural network with more than three layers, including input and output, is considered a deep learning algorithm or deep neural network. Deep learning can take in unstructured data in its raw form, and it can distinguish categories of data from one another. Early uses of machine learning involved the classification of land cover and clouds with the use of high-resolution satellite data (Pritt & Chern, 2018, Bai et al., 2016).

3.2.1. Concept of Neural Network. ANN (or multi-layer perceptron (MLP)) become popular in the field of scientific research especially when handling large datasets. ANNs consist of neurons that are interconnected nodes in the hidden layers, that process information and make predictions based on that information. The input layer receives

data, which is then processed by the neurons in the hidden layers using a set of weighted connections. The output layer produces the final prediction or output (Figure 3.1).

The performance of a neural network depends on several factors, including the architecture of the network, the training algorithm used, and the size and quality of the training dataset. The architecture of the network, which determines the number of layers and the number of neurons in each layer, can have a significant impact on the performance of the network. For example, a deeper network with more layers can learn more complex representations but may also be more difficult to train.

In addition, the choice of training algorithm can also have a significant impact on the performance of the network. The backpropagation algorithm, which is commonly used to train neural networks, adjusts the weights between neurons based on the difference between the predicted output and the actual output. Other algorithms, such as stochastic gradient descent and adaptive learning rate methods, can also be used to train neural networks. The objective is to minimize the error between the predicted output and the actual output, so that the network can generalize to new data.

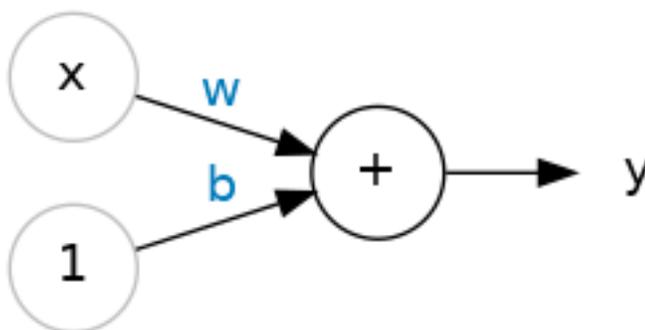


Figure 3.1. Illustration of a single neuron (source from <https://www.kaggle.com/code/ryanholbrook/a-single-neuron>)

Deep learning and neural networks are closely related concepts, but there are some key differences between them. Neural networks are a type of machine learning algorithm that are designed to mimic the structure and function of the human brain. They are composed of interconnected nodes that process information and generate output based on a set of inputs. Deep learning, on the other hand, is a subset of neural networks that involves training neural networks with many layers (Goodfellow et al., 2016). These deep neural networks are capable of learning more complex features and patterns from data than shallow neural networks, and they have been particularly successful in tasks such as image and speech recognition (LeCun et al., 2015).

In other words, deep learning is a more advanced form of neural networks that involves training neural networks with many layers, enabling them to learn more complex features and patterns from data. While neural networks can be shallow or deep, deep learning specifically refers to neural networks with many layers.

The idea of neural network model can be formulated by equation (3):

$$y = \phi(z) = \phi(\mathbf{w}^T \mathbf{x} + b) \quad (3)$$

Where \mathbf{x} is the input, along with a bias, b and \mathbf{w} is the weight. The neuron's output is determined by combining the input vector \mathbf{x} , weight vector \mathbf{w} , and scalar bias value b , which are added together and passed as the argument of the activation function ϕ . (Emmert-Streib, F et al., 2020).

In a neural network, the activation function is used to determine the output of each neuron in the network based on the weighted sum of the inputs. The choice of activation function can have a significant impact on the performance of the network, as it determines

the nonlinearity of the model and can affect the ability of the network to learn complex relationships in the data.

Activation functions are typically chosen to be nonlinear, as this allows the network to model complex nonlinear relationships in the data. There are several commonly used activation functions, including sigmoid, tanh, and ReLU. In this thesis we used ReLU as our activation function and it is used in the hidden layers and allows for faster training and better performance in deep neural networks.

The choice of activation function and its parameters can have a significant impact on the performance of the network. Researchers have shown that choosing the right activation function can lead to better performance and faster training times in deep neural networks (Glorot and Bengio, 2010). It can be defined by the following equation (4):

$$f(x) = \max(0, x) \quad (4)$$

It describes the situation as value form x is negative, the function will output as zero. On the other hand, when x is positive, the function will output as x . Figure 3.2. shows the graphical representation of the equation described above.

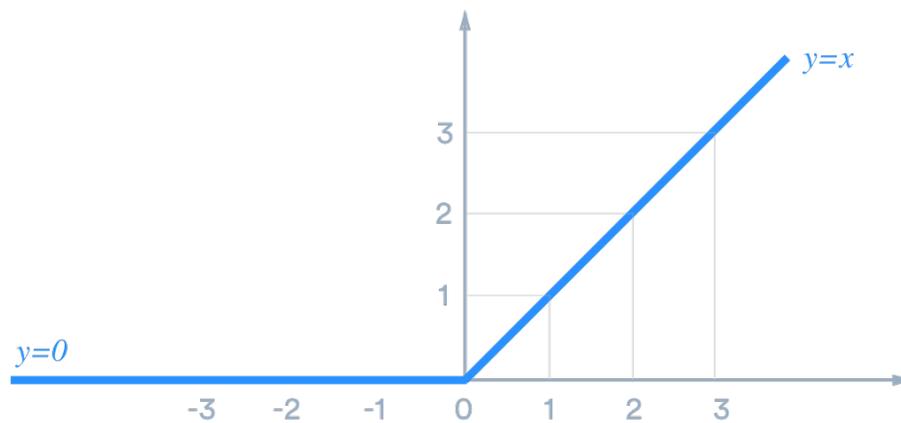


Figure 3.2. Graphical representation of the ReLU activation function above

3.2.2. Gradient Descent. Gradient descent (GD) is an optimization algorithm used in machine learning to minimize the cost function. It works by calculating the gradients of the cost function with respect to the model parameters and updating them in the opposite direction of the gradient to reach the minimum cost. The algorithm iteratively takes small steps towards the minimum cost, and step size is determined by the learning rate (Figure 3.3).

The GD algorithm can be computationally expensive when the training data is large, and the model parameters are high dimensional. There are several variants of GD, including batch gradient descent, mini-batch gradient descent, and stochastic gradient descent. The choice of variant depends on the size of the training data, computational resources available, and the convergence rate required. GD is a widely used optimization algorithm in machine learning due to its simplicity and effectiveness (Goodfellow et al., 2016).

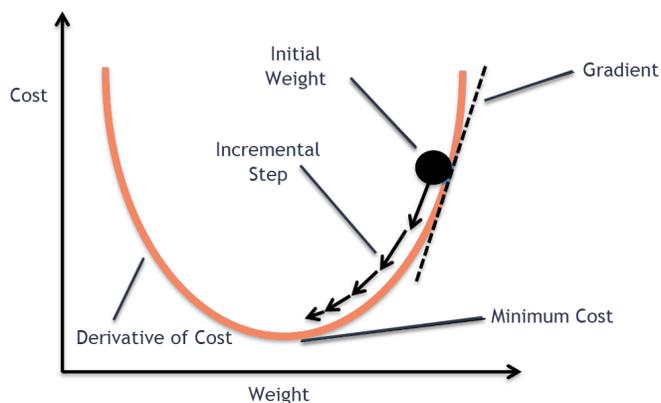


Figure 3.3. Illustration of gradient descent from O'Reilly

Stochastic Gradient Descent (SGD) is an optimization algorithm that improves upon the traditional Gradient Descent (GD) algorithm. Unlike GD, which computes the

gradient of the cost function over the entire training set, SGD updates the model's parameters by computing the gradient over a single randomly chosen training instance (batch size) at each iteration (epoch). This results in a faster convergence rate and better generalization performance, especially in large-scale datasets. (Bottou 2012)

3.2.3. Learning Rate. The learning rate is a crucial hyperparameter that governs the magnitude of the steps taken while optimizing a model. It plays a significant role in determining the extent of the changes made to the model's parameters during the training process. If the learning rate is set too high, the model can converge quickly, but it may overshoot the optimal solution. Conversely, a low learning rate can lead to slow convergence and may get trapped in a local minimum. Hence, selecting an appropriate learning rate is essential for optimizing the model's overall performance.

For example, in the popular optimization algorithm SGD, the learning rate determines the step size taken in the direction of the negative gradient of the loss function. A small learning rate means smaller steps and slower convergence, while a large learning rate means larger steps and faster convergence (Goodfellow et al. 2016).

An appropriate learning rate is crucial for achieving the best performance of a machine learning model. If the learning rate is set too high, the model may diverge and fail to converge to a good solution. If it is set too low, the model may converge too slowly or get stuck in a local minimum. Therefore, it is often necessary to experiment with different learning rates to find the optimal value for a given problem (Figure 3.4).

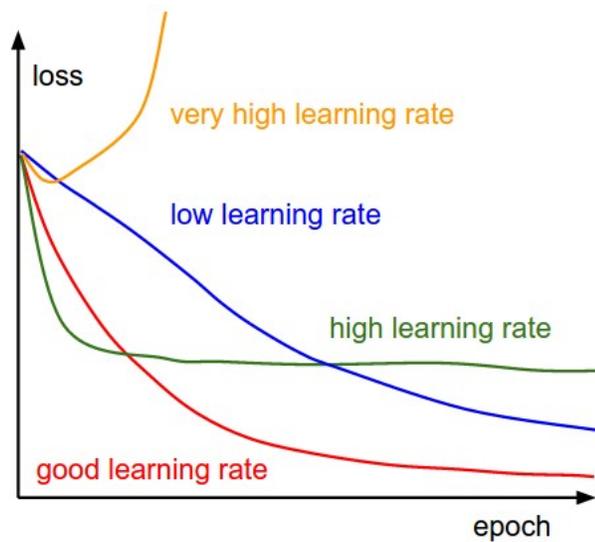


Figure 3.4. Comparison of different learning rate.

3.2.4. Loss Function. Mean Squared Error (MSE) is a commonly used loss function in machine learning for regression problems. It measures the average of the squared differences between the predicted values and the target values and can reference with the following equation (5).

$$MSE = \frac{1}{n} \sum_{i=1}^n (pred_i - label_i)^2 \quad (5)$$

where n is the number of training examples.

The goal of training a machine learning model is to minimize the loss function. In the case of MSE, this means finding the set of model parameters that result in the smallest average squared error between the predicted and actual target values.

MSE has several desirable properties, including being a differentiable, convex function that is easy to optimize. However, it can be sensitive to outliers in the data, which can lead to overfitting if not addressed properly. MSE is a useful and widely used loss function in regression problems due to its simplicity and ease of optimization.

3.2.5. Implementation. The architecture of the neural network model is compiled with 11 hidden layers, and the details can be found in Table 3.1. The network has a funnel structure, which is a commonly used design. The model begins with 155 neurons, the same number as the input parameters, to account for all possible features. In each consecutive hidden layer, the number of neurons remains the same as the previous layer. However, from the next hidden layer onwards, the number of neurons is halved from the previous layer. This process is repeated until reaching the final layer, which outputs one neuron to ensure one prediction value is obtained.

Network described is constructed with python package of Tensorflow – Keras.

Table 3.1. Architecture of model

Layer	Input layer	Hidden layer 1	Hidden layer 2	Hidden layer 3	Hidden layer 4	Hidden layer 5	Hidden layer 6	Hidden layer 7	Hidden layer 8	Hidden layer 9	Hidden layer 10	Hidden layer 11	Output layer	Total
Number of Neurons	155	155	155	80	80	40	40	20	20	10	10	5	1	-
Parameters	-	24335	24335	12560	6560	3280	1680	840	440	220	120	60	6	74436

In machine learning, hyperparameters such as learning rate, batch size, and epochs need to be carefully tuned for optimal performance. However, determining the best values for these hyperparameters can be a tedious and time-consuming process involving trial and error. To streamline this process, the randomized search cross-validation method from the Scikit-Learn model selection module is used for this thesis. This method allows for setting out a range of marginal values for the hyperparameters and then randomly searches for the optimal combination of hyperparameters automatically during the training runs. This

approach can save considerable time and effort in finding the best hyperparameters for a given model. The final set of optimal hyperparameters can be summarized in Table 3.2, for easy reference and reproducibility.

In Figure 3.5, the progress during training is shown. The training curve appears to have a fairly high learning rate, but the validation curve closely follows the training curve until it reaches its minimum. Additionally, increasing the number of epochs does not appear to aid in convergence or result to a lower MSE. Therefore, it only adds more computational time.

Table 3.2. Hyperparameters

Hyperparameters	Values
Learning Rate	0.0001
Batch Size	128
Epochs	50

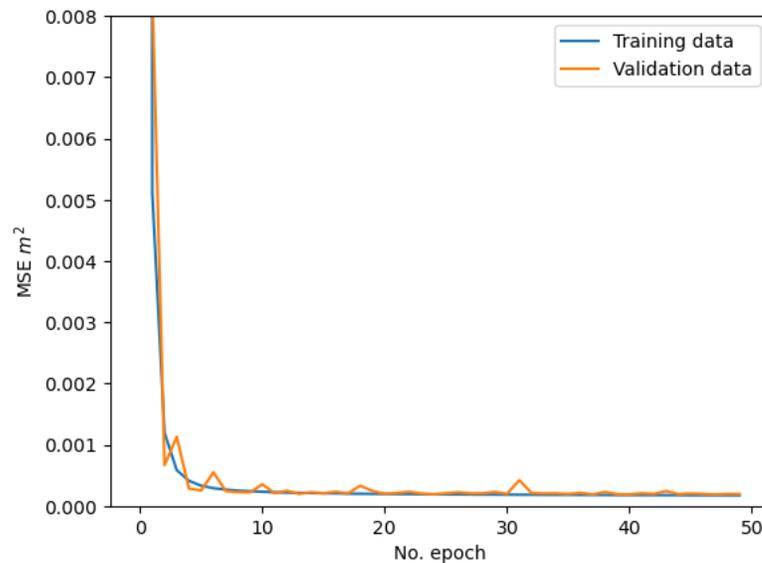


Figure 3.5. Mean square error during training.

3.3. COMPUTATION OF TROPOSPHERIC DELAY WITH TRADITIONAL METHOD

In this thesis, the theoretical model is used as the baseline for predicting zenith total delay (ZTD) at the same station and timeframe as comparisons. The tropospheric phase delay, which is mainly caused by refractivity (N), is composed of hydrostatic, wet, liquid, and ionospheric components (Smith and Weintraub, 1953; Davis et al., 1985). These delays result from variations in temperature, pressure, and relative humidity that are spatially correlated with topography. This study concentrates solely on the hydrostatic and wet components of refractivity. Refractivity, N, can be expressed in the following manner as described in equation (6) (Hanssen, 2021):

$$N = (k_1 \frac{P}{T})_{hydro} + (k'_2 \frac{e}{T} + k_3 \frac{e}{T^2})_{wet} = N_{hydro} + N_{wet} \quad (6)$$

where P represents the total atmospheric pressure, T is the temperature, and e is the partial pressure of water vapor at a specific height. The coefficients k_1 , k'_2 , and k_3 are empirical constants estimated in literature as $k_1 = 77.6 \text{ K hPa}^{-1}$, $k'_2 = 23.3 \text{ K hPa}^{-1}$, and $k_3 = 3.75 \cdot 10^5 \text{ K}^2 \text{ hPa}^{-1}$ (Smith and Weintraub, 1953).

The phase delay caused by troposphere, ϕ_{tropo} , at a given height, $h = h_0$, represents the accumulation of refractivity between h_0 and the top of the troposphere h_{top} along the radar's line of sight as explained by Hanssen (2001). The tropospheric phase delay can be calculated by integrating the refractivity value at grid points using the following equation (7):

$$\phi_{tropo} = \frac{-4\pi \cdot 10^{-6}}{\lambda \cos\theta} \int_{h_0}^{h_{top}} (N_{hydro} + N_{wet}) dh \quad (7)$$

Where θ is the incidence angle, λ is the radar wavelength, and $-4\pi/\lambda$ is a conversion factor used to convert from pseudo range increase to phase delay.

A package RAiDER is used for deriving this pointwise delay. The python script will download the NWP model and create a 3D map where it calculates all the refractivity at each grid nodes. The final product (hydrostatic total and wet total) is derived using a 3D grid interpolator by considering latitude, longitude, and elevation of interest. The zenith total delay is the summation of the hydrostatic total and wet total. All the NWP model ZTD discuss later in the result section are referring to the product derived from RAiDER package.

4. RESULT

4.1. COMPARISON RESULTS BETWEEN MODELS

4.1.1. CONUS Model. The preliminary version of the model architecture described in Section 3.2.2 was trained on GNSS stations situated in the US continent. The testing dataset consisted of data collected from the period of 2021 to 2022. To evaluate the accuracy of the model's predictions, a histogram was generated to display the distribution of residuals for each GNSS station (refer to Figure 4.1 and 4.2). The histogram showed a normal distribution with a mean close to zero 0.00039 m for RF model and -0.00144m for NN model. The RMSE shows for both models are well below 1 cm which is 0.972 cm and 0.874 cm for RF model and NN model respectively. NN model with a lower RMSE suggests that the models' predictions were highly accurate and exhibited consistent performance.

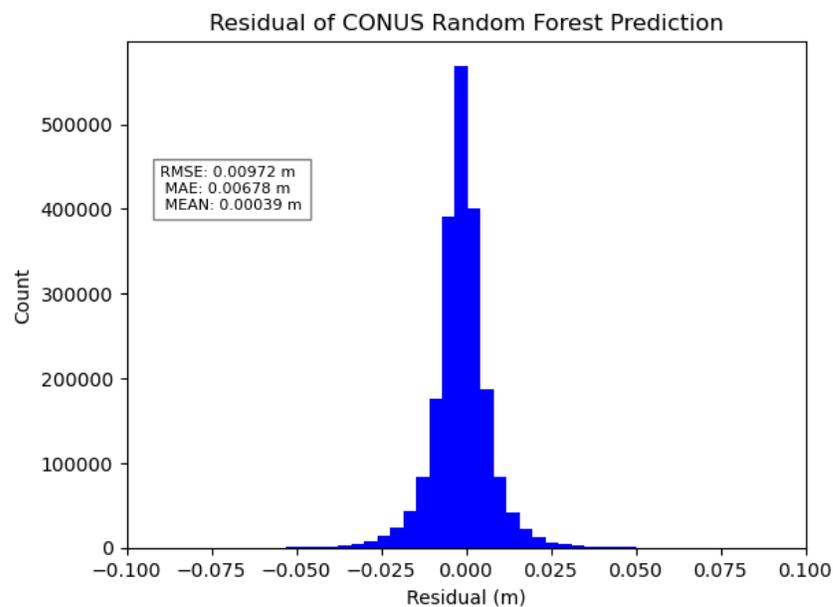


Figure 4.1. Residual for CONUS RF model prediction

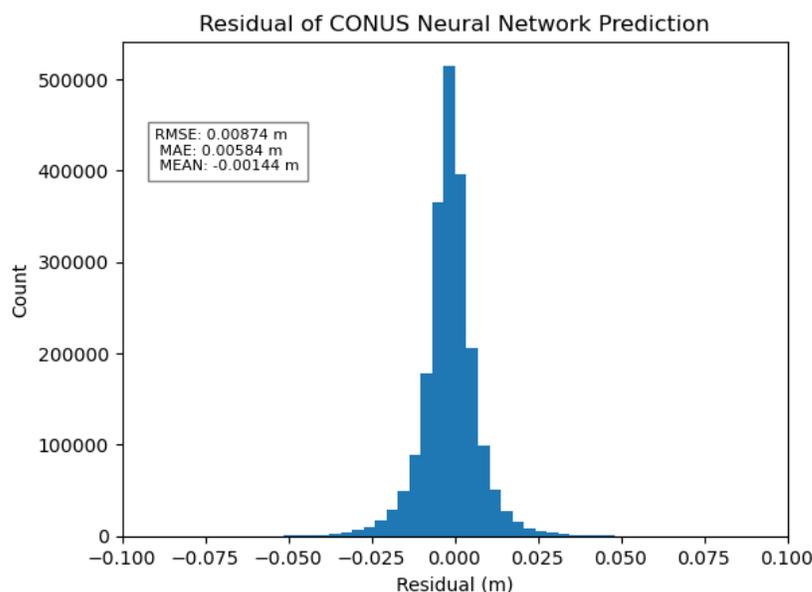


Figure 4.2. Residual for CONUS NN model prediction.

To further evaluate the model's performance, Figures 4.3 and 4.4 were utilized to show the root mean square error and average error for individual GNSS stations, respectively. These figures demonstrate that the model's RMSE predictions in the continental US were less than 2 cm, signifying that the model performed exceptionally well in forecasting the target values. The accuracy of the model's predictions in this region is a positive indication of its ability to forecast complex weather patterns at a local level, which could possibly be extended to a global scale. Overall, these results provide strong support for the capability of machine learning to accurately predict zenith total delay from NWP model features.

Accuracy of the prediction seems to decrease as GNSS that are closer to the equator (Figure 4.3). From the northern part up in Canada with a RMSE range from 0.0025 m – 0.0050 m to the southern part down in Texas, Louisiana, and Florida states, we see an increase of RMSE up to 0.0125 m. This is possibly related to the increase of humidity in

the atmosphere from the ocean. The same kind of effect also appear to some inland area such as the city of Toronto where huge lake could introduce large variation in humidity in the atmosphere and could be difficult for the model to predict.

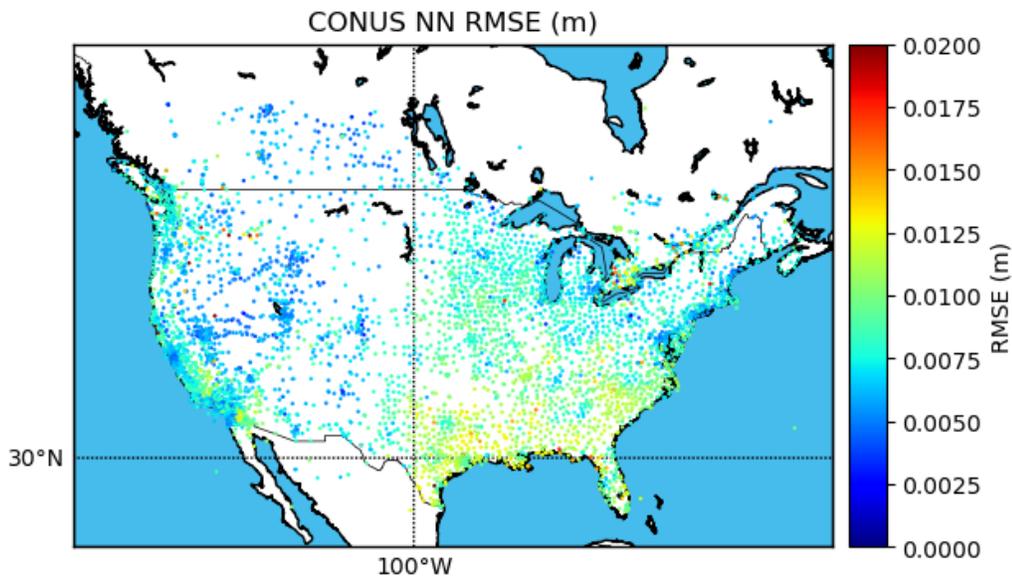


Figure 4.3. CONUS NN model prediction RMSE.

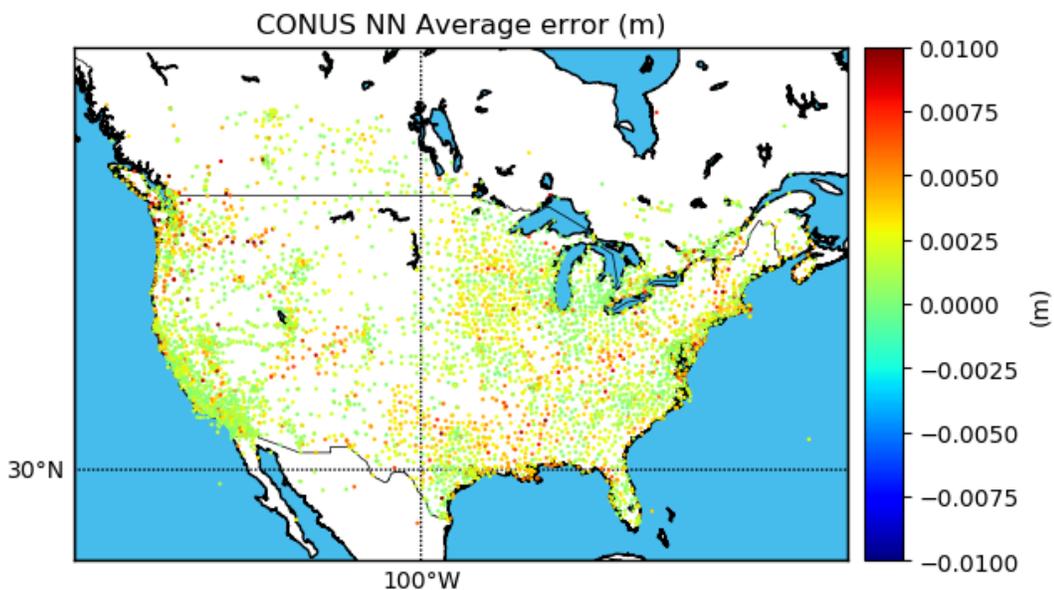


Figure 4.4. CONUS model prediction average error.

4.1.2. Global Model. This section is to compare the performance of the global Neural Network (NN) model and the global Random Forest (RF) model, which were trained using data from a 15-day interval with fewer input features than the previous model. While the previous model focused on daily data at a specific time, this new model considers the temporal and spatial variability of the data, covering a larger geographical area with more diverse latitude and elevation values. Figures 4.5 and 4.6 demonstrate the performance of the two models, where both models exhibit a normal distribution with a mean close to zero. The normal distribution indicates that the predictions made by both models are close to the actual values, which is an essential factor in evaluating the accuracy of the models. The mean value from the residual plot shows that random forest model does not seem to have a bias with value very close to zero, however, neural network model may seemingly have slightly bias in overpredicting with the mean value of 0.224 cm.

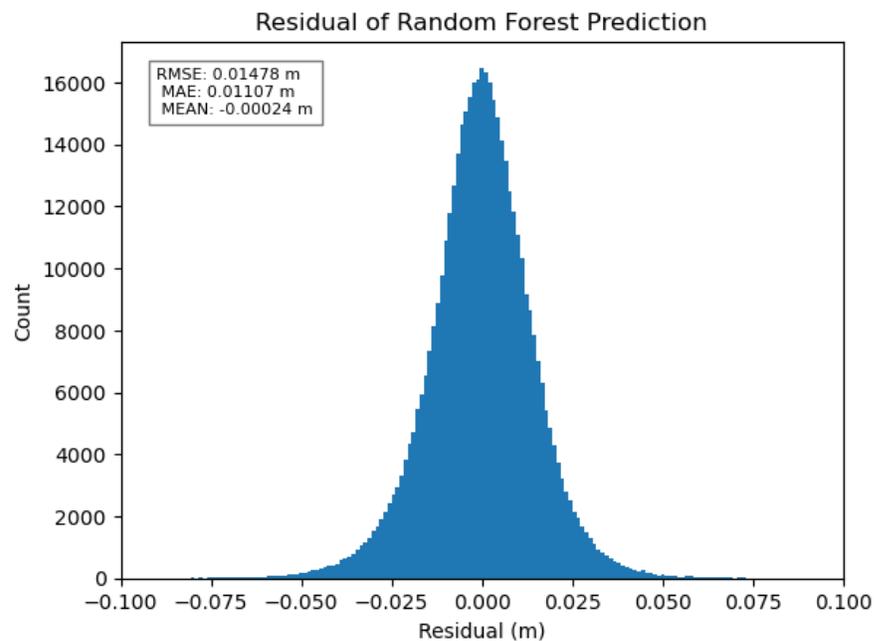


Figure 4.5. Residual plot of Random Forest result.

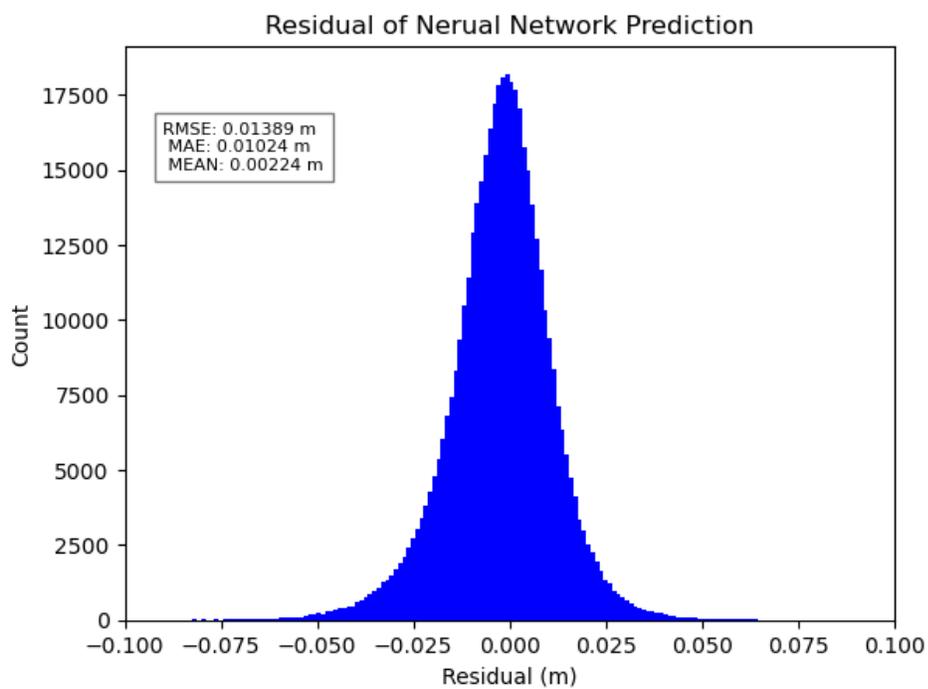


Figure 4.6. Residual plot of Neural Network result.

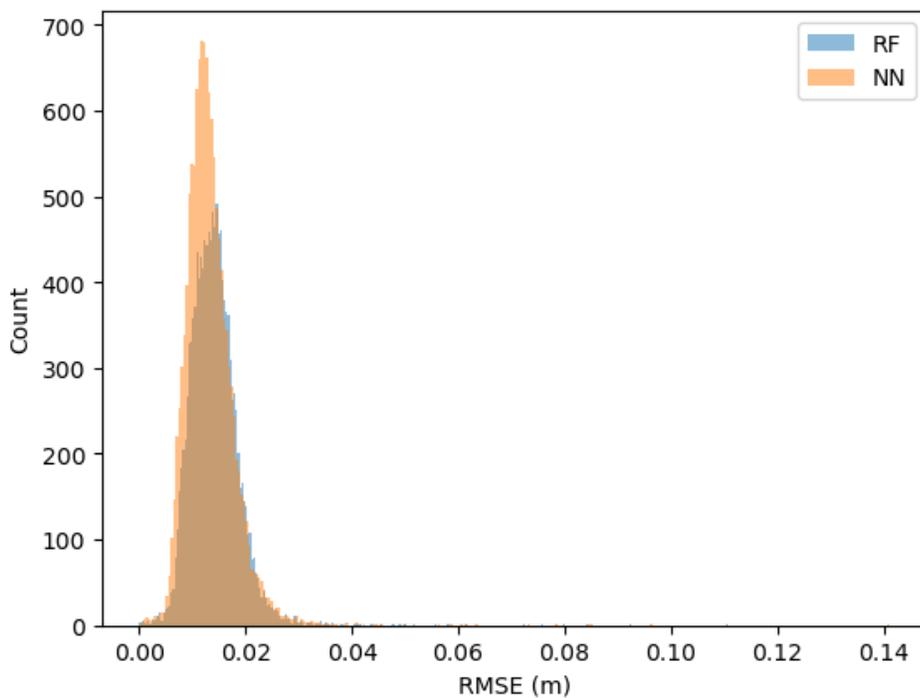


Figure 4.7. Histogram of RMSE by GNSS station from RF and NN model.

In Figure 4.7, a direct comparison of the RMSE distribution from all stations between the NN and RF models is presented. While the NN model has a slightly lower mean RMSE than the RF model, the standard deviation distributions are similar. Additionally, the histogram reveals that over 80% of stations have an RMSE lower than 0.02 m.

When comparing the two models, it is evident that the neural network model is more accurate in predicting zenith total delay, with an overall RMSE of 1.39 cm. This result is below the target of 2 cm set for this study, indicating that the neural network model is well-suited for this task. On the other hand, the RF model shows a slightly higher overall RMSE of 1.48 cm, still within the margin but suggesting that it may not be as accurate as the neural network model.

Taken together, these metrics and plots suggest that the neural network model is the more suitable option for predicting zenith total delay due to its higher accuracy compared to the RF model. However, it is worth noting that both models performed exceptionally well in predicting zenith total delay with high accuracy.

The residual plots highlight the potential of machine learning techniques in improving the accuracy of atmospheric correction in remote sensing applications. Overall, these findings demonstrate the promise of machine learning in advancing our understanding of complex atmospheric processes and improving the accuracy of zenith total delay predictions.

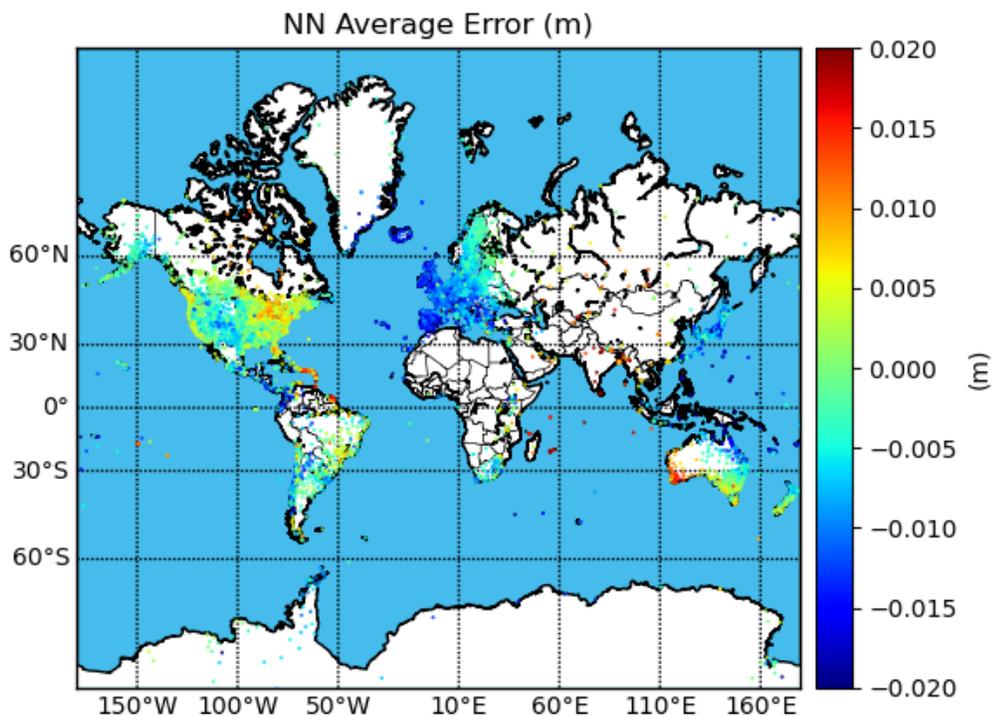


Figure 4.8. Globe NN model prediction average error.

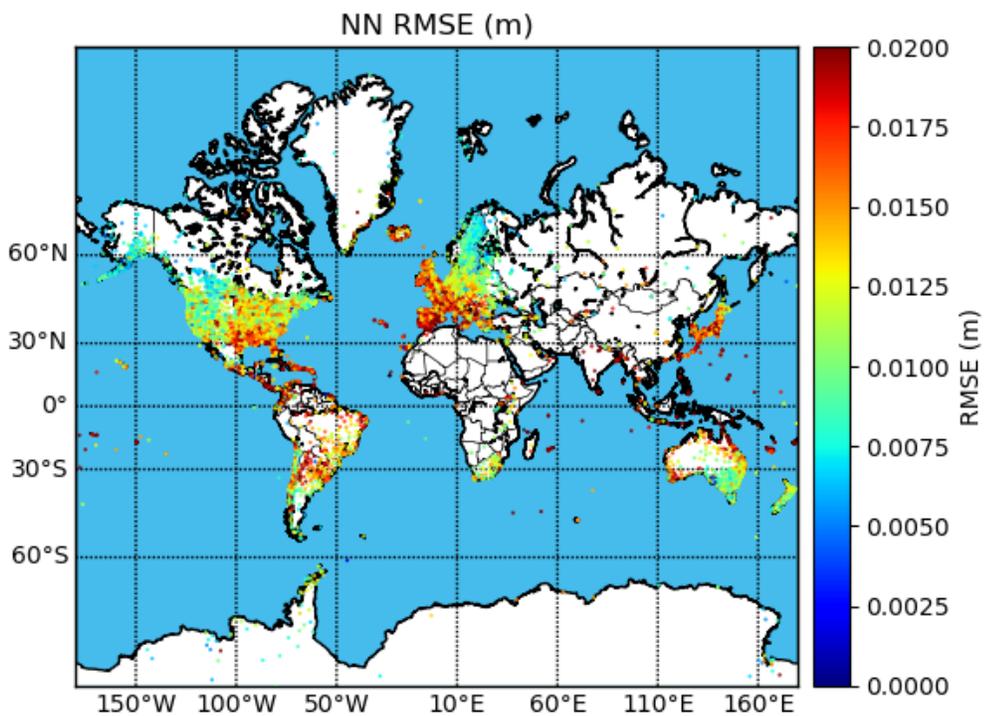


Figure 4.9. Globe NN model prediction RMSE

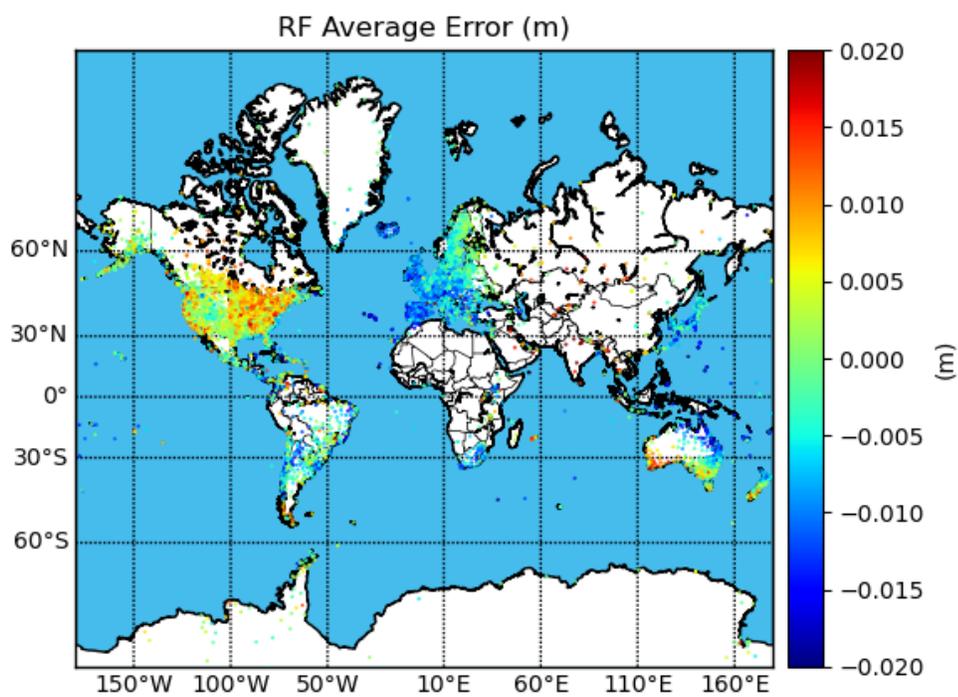


Figure 4.10. Global RF model prediction average error.

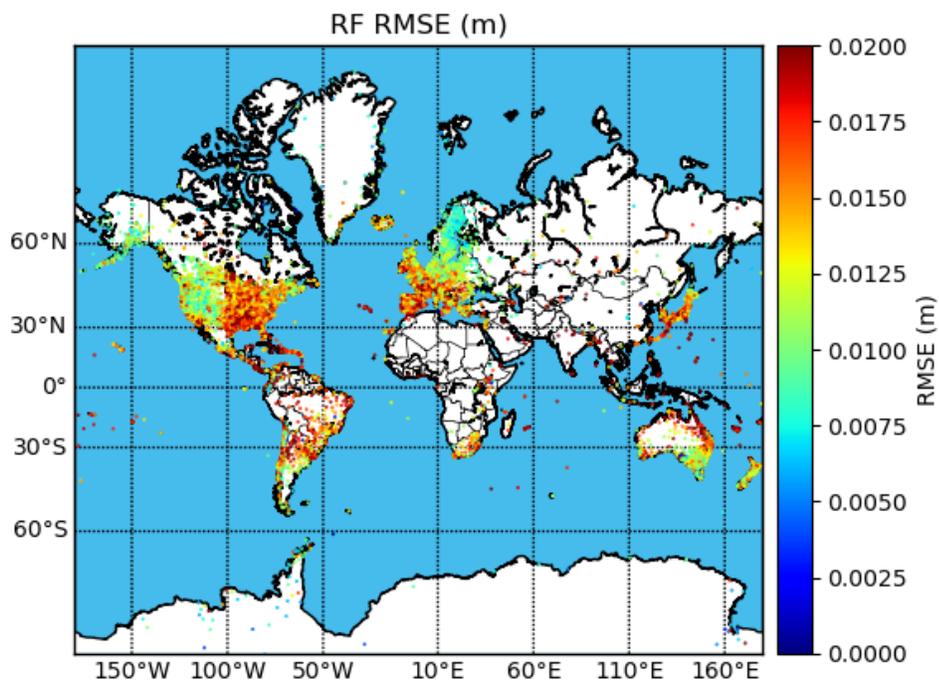


Figure 4.11. Global RF model prediction RMSE.

Compared to the machine learning model, the NWP model derived ZTD was calculated at the same GNSS stations as presented earlier. Figures 4.12 and 4.13 depict the average error and RMSE at GNSS stations in the globe. These plots display a significant range of error that appears to follow a systematic pattern, which correlates with the continent or longitude. Interestingly, Europe and Japan seem to have better predictions, while North America shows the lowest RMSE and average error. These results may suggest that the NWP model performs better in certain geographical areas than others. Additionally, it is worth noting that the systematic pattern observed in the error range may indicate the presence of unaccounted-for factors that affect the NWP model's predictions. Further research is necessary to investigate and identify these factors and how they influence the accuracy of the NWP model's predictions.

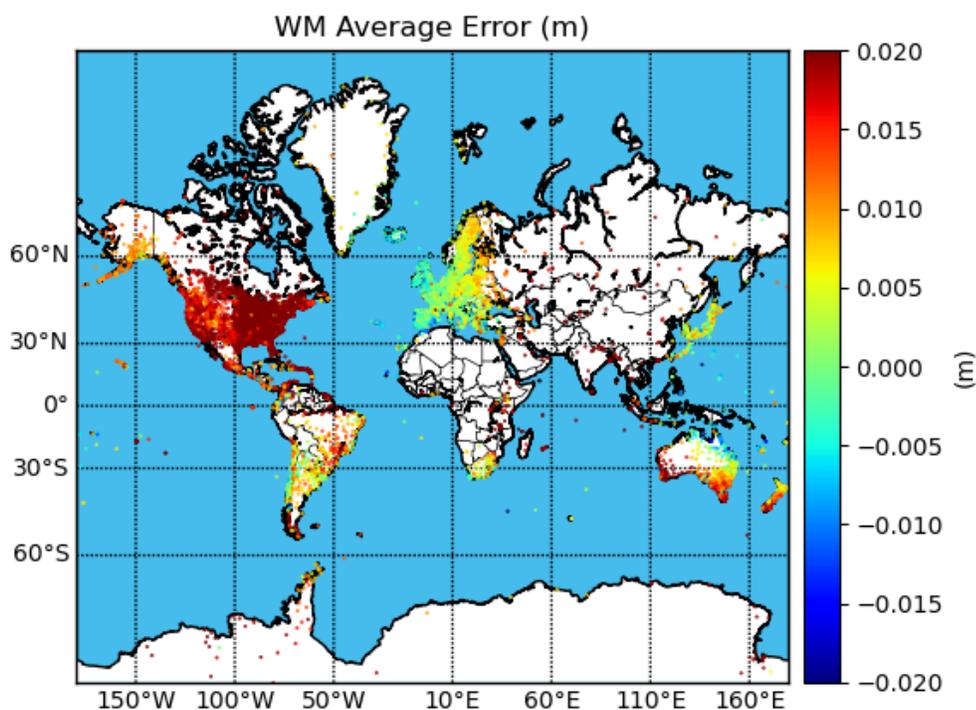


Figure 4.12. Global WM prediction average error.

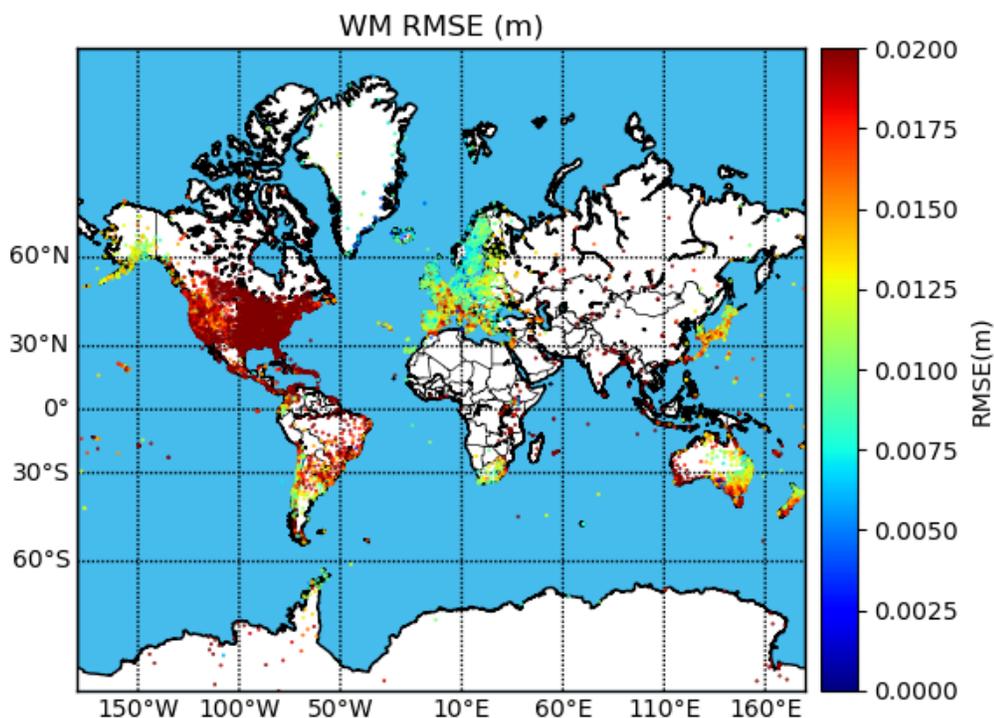


Figure 4.13 Global WM prediction RMSE.

4.2. CORRELATION BETWEEN PARAMETERS AND ERROR

4.2.1. Elevation. The variability in the accuracy of Global Navigation Satellite System (GNSS) stations is to be expected due to the individual circumstances surrounding each station. Factors such as station height and latitude can differ greatly from one station to another, leading to a wide range of potential errors in the data collected. This variability is demonstrated in Figure 4.14, where the relationship between the root mean square error (RMSE) and station elevation shows a wide range of potential errors. The dependency is not linear and can vary depending on the region. The plot suggests that the prediction of station RMSE is dependent on elevation, with higher RMSE in area of low elevations and lower RMSE in area at high elevations. This trend observed for both neural network (NN) and random forest (RF) models. As anticipated trend is primarily because at lower

elevations, there is typically more water vapor in the atmosphere, leading to increased atmospheric delay. Conversely, at higher elevations, the atmosphere tends to be drier, resulting in less atmospheric delay. However, the impact of elevation on prediction error seems to decrease for stations above an altitude of 1000 m. This is the case in general but at some cases we see in higher elevation it can have a large RMSE as well. The impact of elevation on ZTD can also be influenced by local factors such as topography, vegetation, and meteorological conditions, which can differ from one region to another.

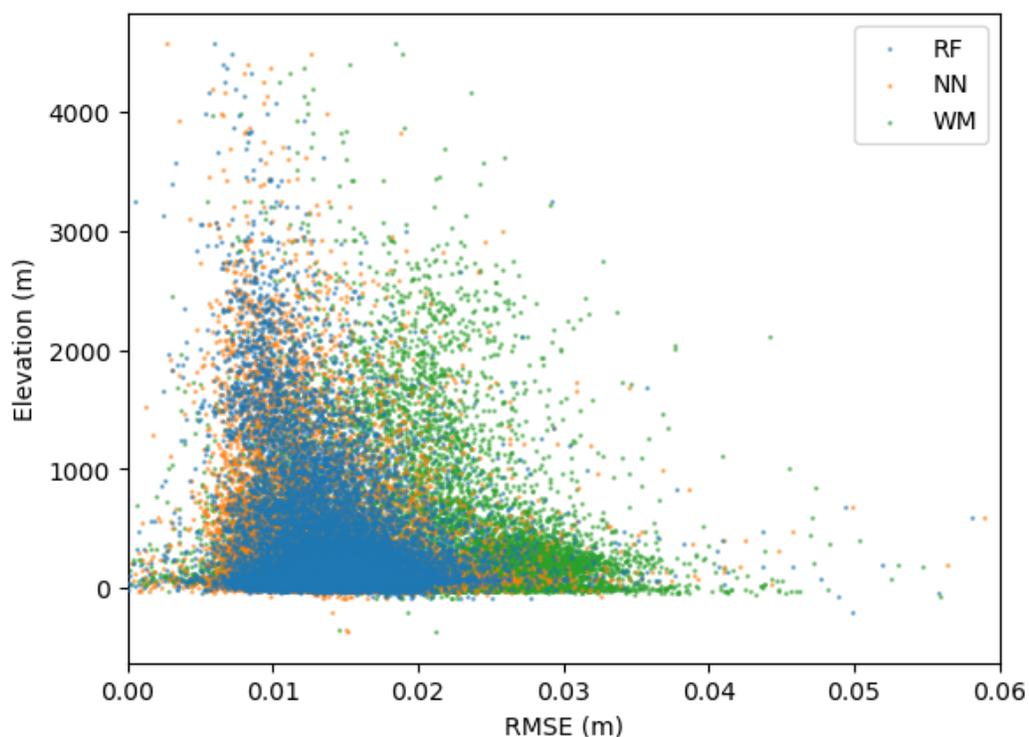


Figure 4.14. Scatter plot between RMSE (m) and elevation of stations with comparison between NN model and RF model.

4.2.2. Latitude. Figure 4.15 have shown that ZTD is influenced by latitude, and that the dependency is not linear. In general, ZTD is higher at lower latitudes and lower at higher latitudes. This trend is since at lower latitudes, there is typically more water

vapor in the atmosphere, leading to increased atmospheric delay. Conversely, at higher latitudes, the atmosphere tends to be drier, resulting in less atmospheric delay.

Furthermore, the dependency of ZTD on latitude is not constant across different regions of the world as seen in Figure 4.8 - 4.11. In some areas, the impact of latitude on ZTD may be more significant than in others due to local atmospheric conditions.

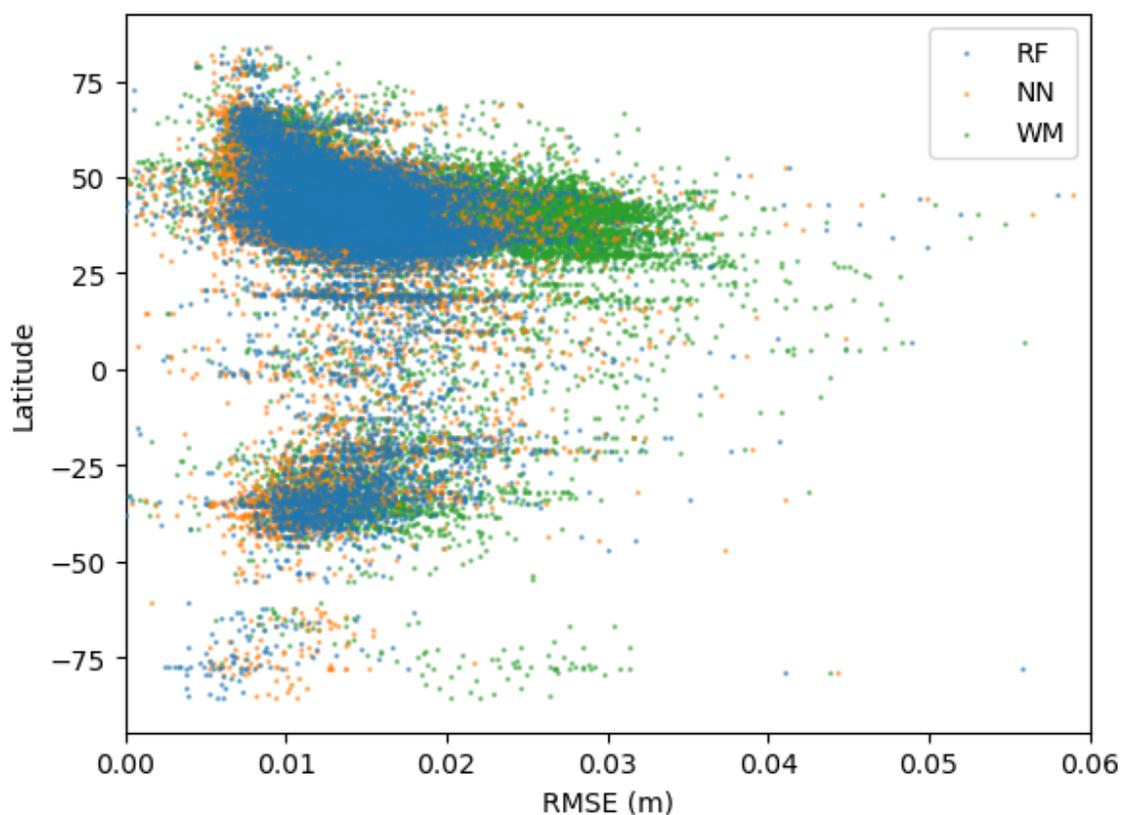


Figure 4.15. Scatter plot between RMSE and latitude of stations with comparison between NN model and RF model.

4.3. RANDOM FOREST FEATURE IMPORTANCES

One advantage of the random forest model is its interpretability, as compared to other machine learning models. This is achieved by reviewing each feature importance

during the splitting of node and branch in decision trees and the creation of forest. The weighting of each feature ranges from 0 to 1 and sums up to 1. These weights indicate the importance of the input features when determining a split in a decision tree.

Table 4.1 presents the eight most important features with their corresponding weights. It can be observed that the elevation of the GNSS station is the most important feature on the list. This means that the elevation of the GNSS station is the most influential factor in predicting the target values. Other important features include the water vapor profile specifically below 5 km.

This information is valuable in helping us understand which features are most useful in predicting the target values. By knowing which features have the most impact, we can focus our efforts on improving data quality and ensuring that these features are accurately measured and recorded. This knowledge can also guide future model development and refinement, leading to more accurate and reliable predictions.

Table 4.1. Importance of feature and weightings

Hgt_m	e_14	e_8	e_15	e_16	e_10	e_5	e_20
0.794	0.1246	0.0181	0.0147	0.00536	0.00399	0.00387	0.00240

4.4. COMPARE PREDICTION WITH TRADITIONAL METHOD

To thoroughly evaluate the accuracy of the machine learning prediction, it is compared with the prediction from traditional Empirical model as described earlier. This comparison is conducted over the exact same timeframe, ensuring a fair and direct

comparison between the two approaches. Table 4.2 presents several properties of the residuals obtained from both the machine learning and Empirical models, along with a third comparison method. Analyzing the properties of the residuals can provide additional insight into the performance of each model. For example, the mean and standard deviation of the residuals can indicate the overall bias and precision of each model, respectively. The comparison of these properties can help in determining the strengths and weaknesses of each model and guide further development of GNSS prediction models.

Table 4.2. Comparison between random forest, neural network, and traditional method

	Random Forest	Neural Network	Traditional Method
Root Mean Square Error [m]	0.01478	0.01389	0.02017
Mean residual [m]	-0.00024	0.00224	0.01301
Amount of residual > 2cm [%]	7.77	6.51	42.5
Mean Absolute Error [m]	0.01107	0.01024	0.01614

After analyzing the results, both the random forest and neural network predictions demonstrate significantly lower errors compared to the Traditional method approach. The RMSE and MAE errors are lower by approximately 5-6 mm, leading to a 27% reduction in error. Additionally, the number of residuals exceeding the 2 cm threshold is significantly

lower with model utilized machine learning, particularly neural network predictions. It shows that Traditional method have more than 40 % of estimation with residual larger than the threshold while machine learning approach have more 90 % of estimation within the 2 cm threshold. It indicates machine learning approach have higher accuracy than Traditional method.

A bias in the ZTD prediction is identified with the Traditional approach. While the Traditional method produces a mean of over 1.3 cm, the mean residuals for both the random forest and neural network models is very close to zero. This reflects that the Traditional method have the tendency to overpredict the ZTD. Comparatively machine learning algorithms are close to no bias, except for a slight overprediction by the neural network model.

To further analyze the prediction corresponding to error fluctuate with times of year, Figure 4.16 illustrates a month-by-month comparison of RMSE for all three models. In general, the prediction error is lower in the months of winter and spring, while the months of June, July, August, and September being the highest errors within a year. This is likely related to the increased of moisture in the atmosphere which increase more fluctuation in wet delay during the summer months, which is much more difficult to determine precisely compared to the dry part, thus increasing the error. However, it appears that both machine learning predictions handle these seasonal patterns better than the Traditional method. Comparing between machine learning approaches, neural network model appears have lower RMSE than random forest model in most of the months except in the month of July and August.

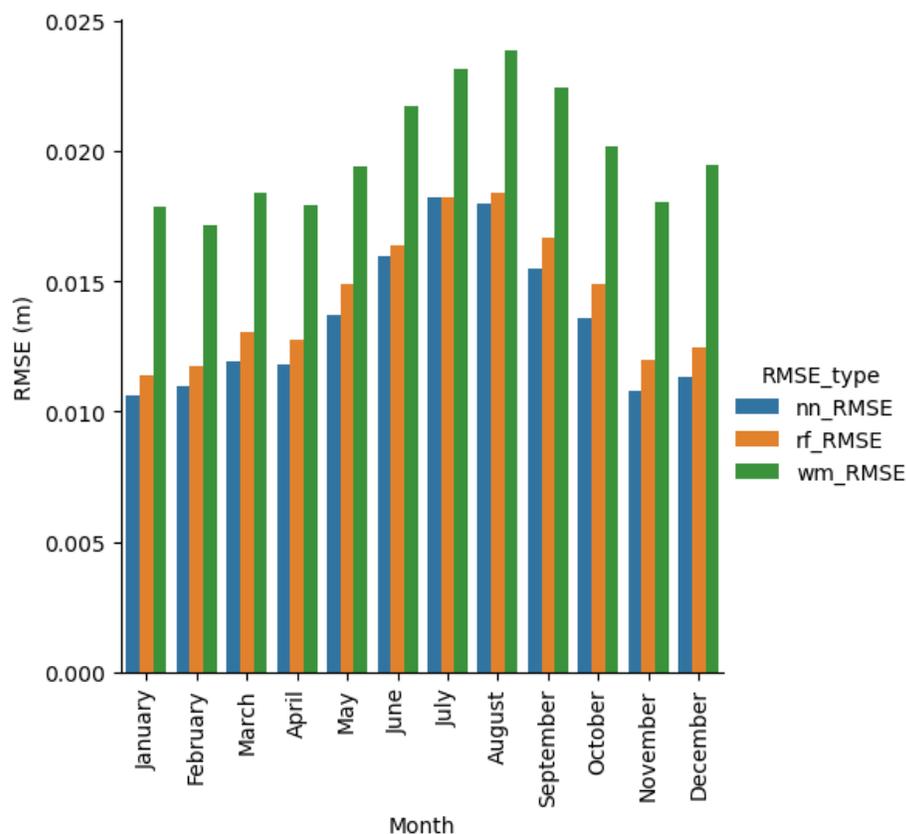


Figure 4.16. Compare RMSE from between model prediction and GNSS true ZTD of each station and plot by month.

4.5. COMPARE PREDICTION BY STATIONS

To further analyze the properties of the prediction from the neural network, random forest, and traditional method, two stations are selected to further investigate.

4.5.1. IABL Station. The station is situated at (-92.4309 W, 40.7408 N) and has an elevation of 239.93 m. It is in the center of the United States, in the south of Iowa, near the border of Missouri. The area is characterized by relatively flat topography without any mountains. Figure 4.17 depicts the monthly variation of RMSE for this station. It is observed that the monthly distribution of RMSE exhibits similar characteristics to the error observed for all the stations. The error is commonly lower during winter months. The

difference in summer months is not as significant, where the RMSE of neural network predictions in typical month are less than half compared to Traditional approach but it's not the case during summertime. It indicates at sites that are in lower elevation will have a difficult time in predicting high accuracy ZTD in summertime due to the increase volume of moisture in the atmosphere.

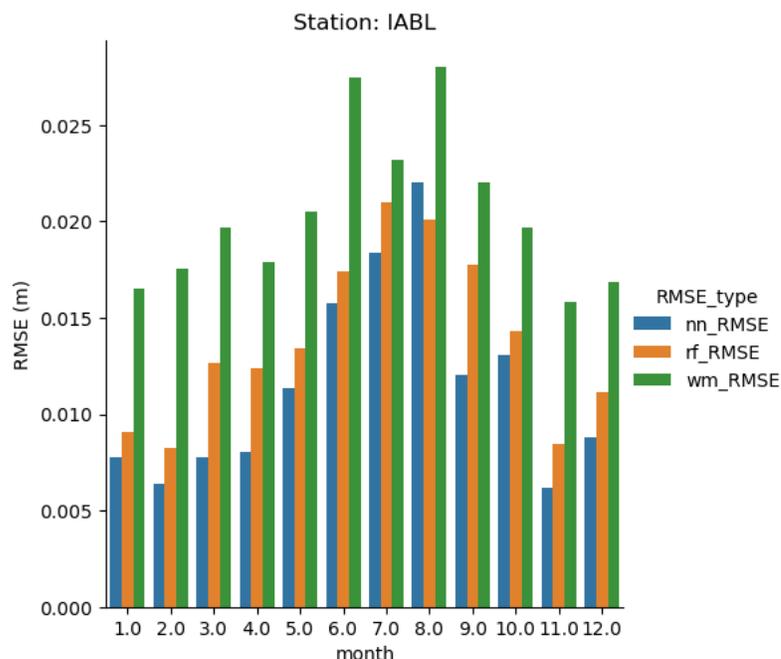


Figure 4.17. Histogram compare of RMSE by month for station IABL between each model prediction.

Upon closer examination of the residuals of the time series predictions, as depicted in Figures 4.18 and 4.19, it becomes apparent that the initial observation holds true. Specifically, the Traditional model tends to overpredict the ZTD, a trend that is especially prominent during the summer months. Conversely, the neural network proves capable of handling these extreme events, though it does have a tendency to underpredict the delay. However, the line of neural network prediction fits closely with the actual delay, as

evidenced by both the time series and residual plots of the predictions. Interestingly, while it is indistinguishable from the random forest model, the neural network stands out as being markedly different from the traditional method.

Indeed, this divergence is reflected in the metric of RMSE, with the neural network achieving a value of 1.15 cm and the traditional method lagging behind at 1.36 cm - almost double that of the former. This difference is particularly noteworthy given that the maximum error recorded from this station can exceed ± 6 cm. Figure 4.19 underscores the robustness of both machine learning methods, as neither displays a strong bias and the residual plot indicates that the error is consistently close to zero throughout the year.

However, the traditional method presents a different picture altogether, with a bias of 0.02 m evident throughout the year. The root cause of this bias remains elusive, and it is possible that it may stem from some constant in the equation or other factors. Thus, further investigation is necessary to address this issue and determine a path forward.

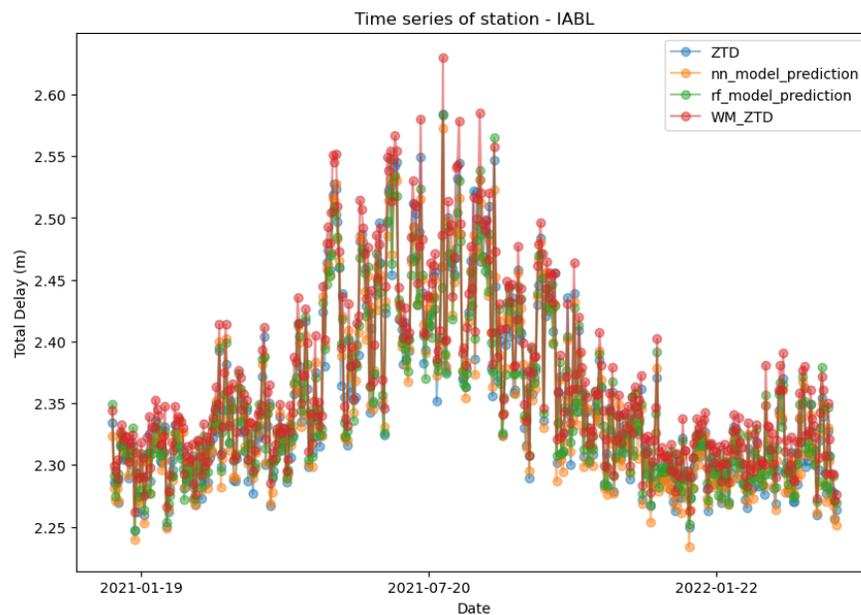


Figure 4.18. Time series of predictions for station IABL

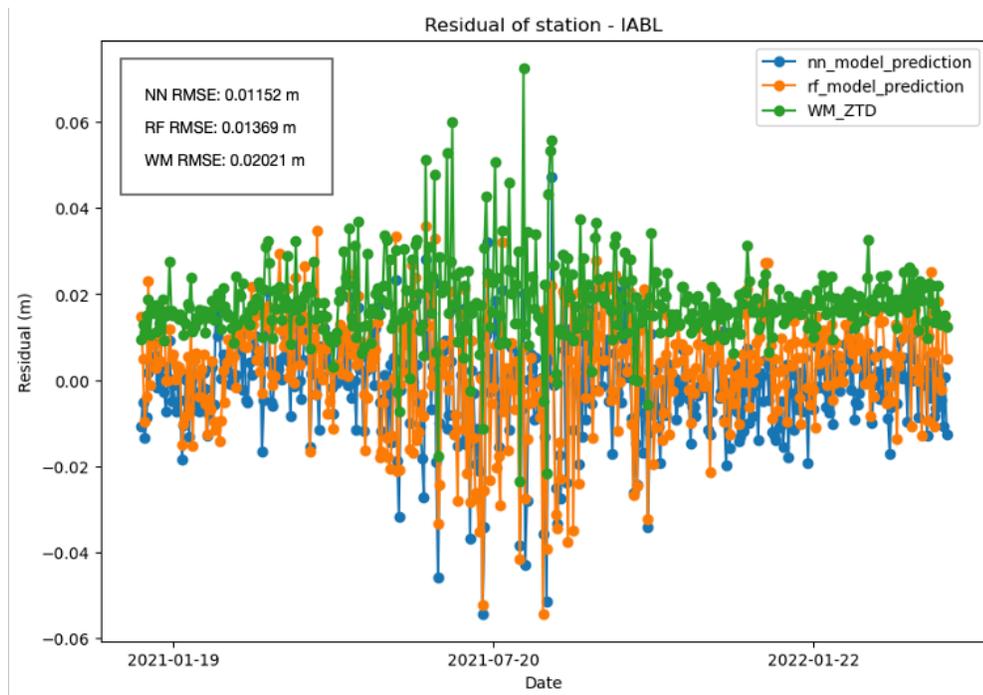


Figure 4.19. Residuals plot of station IABL compare with three model.

4.5.2. LINC Station. The station is located at (-119.01 W, 37.63 N) in the eastern state of California near by a ski resort at Mammoth Mountain, with an elevation of 3056.94 m. This station is selected for comparison to evaluate the correlation between elevation and tropospheric delay. It is known that higher elevations have lower water vapor content, resulting in lower tropospheric delay. However, a similar seasonal pattern of higher tropospheric delay in the summer months and lower effect during the winter months can still be observed at higher elevations (Figure 4.20). In contrast, Figure 4.21 shows that the residual plot of the station still exhibits a positive bias or overprediction for the Traditional method. It is surprising to note that the Global neural network model behaves as an under-predictor at this station, while the Global random forest model remains close to zero throughout the year. The overall RMSE for all models is much lower than 2 cm, ranging between 1 – 1.4 cm.

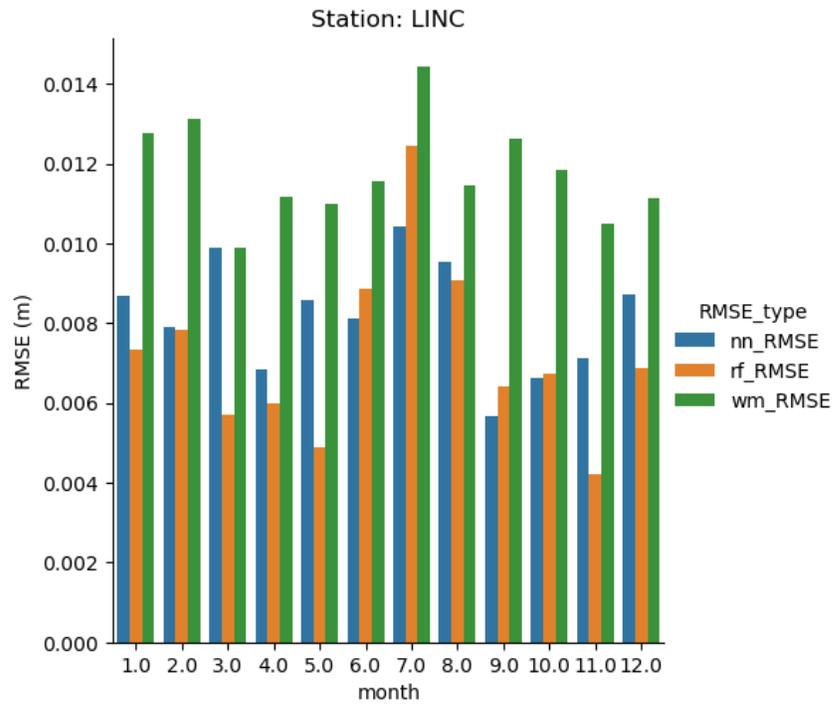


Figure 4.20. Monthly RMSE for station LINC.

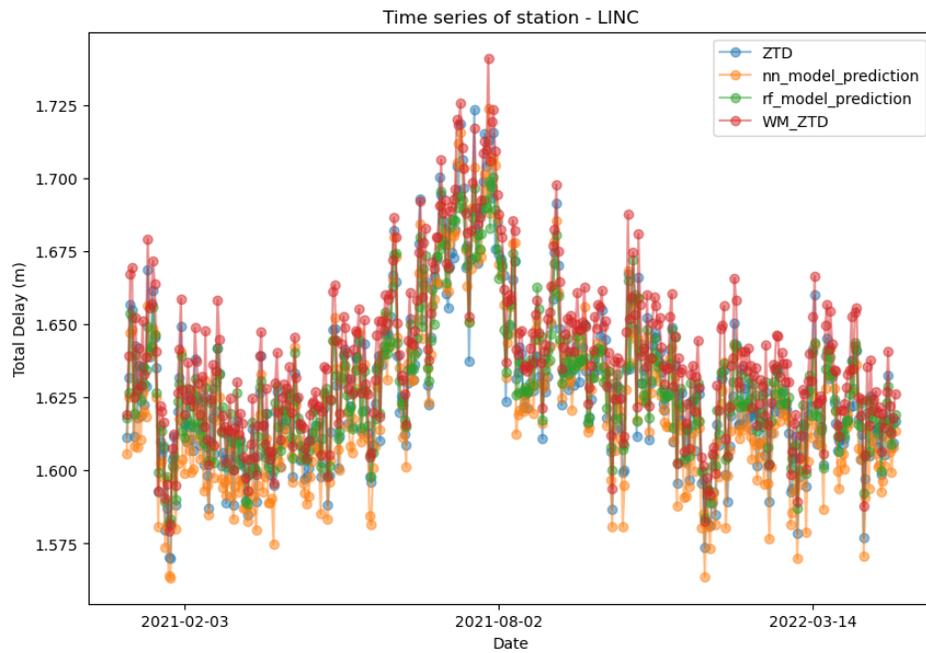


Figure 4.21. Time series of prediction for station LINC.

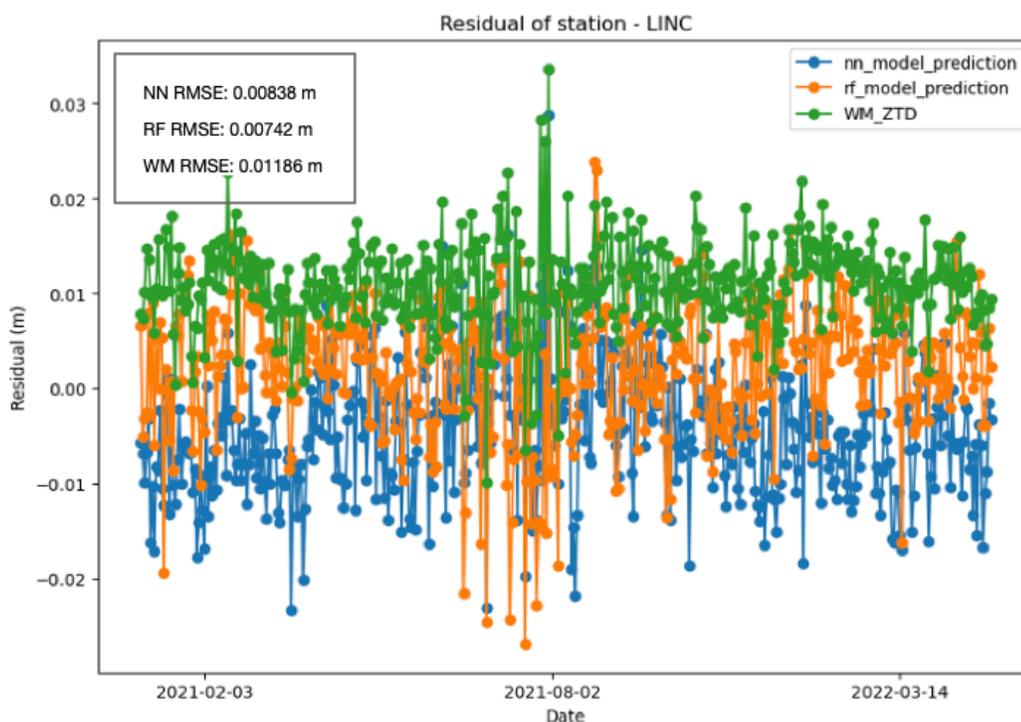


Figure 4.22. Residuals plot of station LINC compare with three model.

The elevation of the station appears to have an impact on the improvement of random forest prediction compared to the neural network model. The RMSE values suggest that the random forest model is marginally more precise by approximately 1 cm. However, these findings are preliminary in nature and further tests are necessary to assess the correlation more accurately between weather model parameters and the target variable.

This thesis is of interest of the ability to predict the magnitude of the ZTD. During the process of tropospheric correction, this bias will be canceled out from each other when calculating the interferometric total delay from two date. As long as the bias are consistent within each method, it does not raise the flag for getting the correct interferometric delay between two dates.

4.6. APPLICATION TO INSAR

4.6.1. Background of Test Site. An investigation was conducted to determine the effectiveness of different tropospheric InSAR correction methods in two regions, Pennsylvania, and Hawaii. The geocoded unwrapped interferogram (GUNW) and its related products were downloaded from ARIA-tools, an open-source Python package that contains tools to manipulate ARIA standard InSAR products (Buzzanga et al., 2020).

In Pennsylvania, there were eight ascending interferograms from February 2020 to November 2020 at around 23:00:00 UTC (Track 004, 033, 077, and 106). The study area covered an area of approximately 440 km by 500 km (Figure 4.23a), extending from the south of Pennsylvania to the north of Virginia across the Appalachian Mountain range.

There was no significant crustal deformation other than landslides and possible subsidence. The topography in this area increases from sea-level up to 900. The northern part of the interferogram had mountain ranges with a maximum elevation of 900 m, which can act as a sink for air flow, introducing different atmospheric conditions in between them. The testing area in this study exhibited large tropospheric phase delays (>5 cm). The weather model grid node has a coverage of 18 x 16 grid points.

In Hawaii, seven ascending interferograms were collected over the largest island in Hawaii between July and November 2019 at around 4:30:00 UTC (Track 124) with a short temporal baseline (6-12 days). Signals from these interferograms should be clear from deformation and the major measurement are tropospheric signals. The landmass, spanning around 130 km in width and 160 km in length, is in the midst of an oceanic expanse and harbors two shield volcanoes - Mauna Loa and Kilauea - that are currently active and positioned towards the southern region of the island. Additionally, the island is also marked

by two mountain summits - Mauna Kea and Mauna Loa - which have the highest altitudes on the island, reaching up to 4,170 m and 3,980 m correspondingly. The topography of the study area ranges from sea level to 4,500 m across the island, presenting a challenging environment for InSAR measurements.

The interferograms obtained were impacted by the unpredictable weather conditions prevalent on the island. These weather patterns are characterized by a complicated blend of turbulent mixing and local climate, because of the diverse distribution of humidity in troposphere in Hawaii. Additionally, the wind patterns around the towering volcanoes on the island are intense and fluctuate rapidly, further contributing to the variability of the delay. These factors pose challenges due to the high variability in water vapor content in the air, making it difficult to accurately correct for tropospheric delays. Despite these challenges, no eruptions were recorded during the study period, allowing the team to focus solely on the effects of the island's atmospheric conditions on InSAR measurements. For weather model coverage it has 5 x 8 grid points with about 25 points overlapping with our study area (Figure 4.23b).

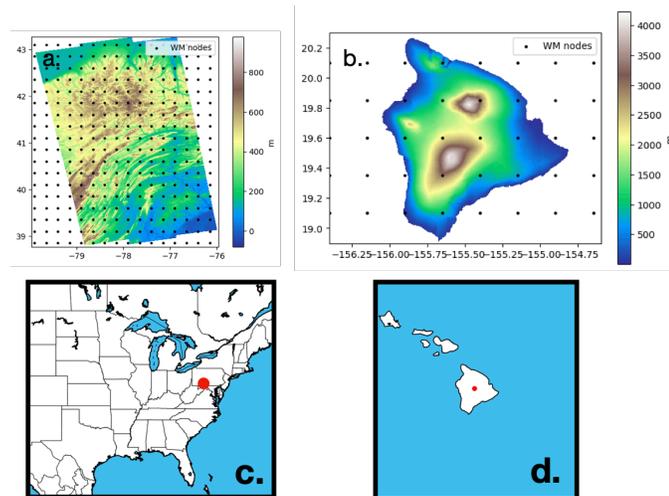


Figure 4.23. Topography of study area Pennsylvania and Hawaii.

4.6.2. Implementation. Figures 4.24 and 4.25 provide a comparison between eight examples in Pennsylvania and seven examples in Hawaii of the estimated tropospheric delays from the Global Neural Network model prediction and the RAIiDER (Traditional Method) prediction. The original interferograms were converted from phase delay (π) to meters (m). Since an interferogram is a measurement of relative change, a pixel was selected as a reference point to reflect the surrounding pixel change with respect to that pixel. To ensure more representative results, the selected interferograms for the application had an average coherence value of 0.8 or above. For each example, the date is shown from left to right, with the first row displaying the raw interferogram where the tropospheric noise are partly correlated with the change in topography. The second and third rows show the predicted tropospheric delay from the Global Neural Network model and RAIiDER (weather model), respectively. The last two rows display the corrected interferogram from the two methods that are dereferenced from the common pixel and converted to the line of sight with the equation mentioned in Section 1.4.

Both the Global Neural Network and RAIiDER predictions (Traditional Method) were able to estimate the delays and have similar magnitude as reflect in the original interferogram, although not always in the correct location in Pennsylvania. While similar magnitudes and spatial patterns of the delay were achieved, there are different results observed when compared.

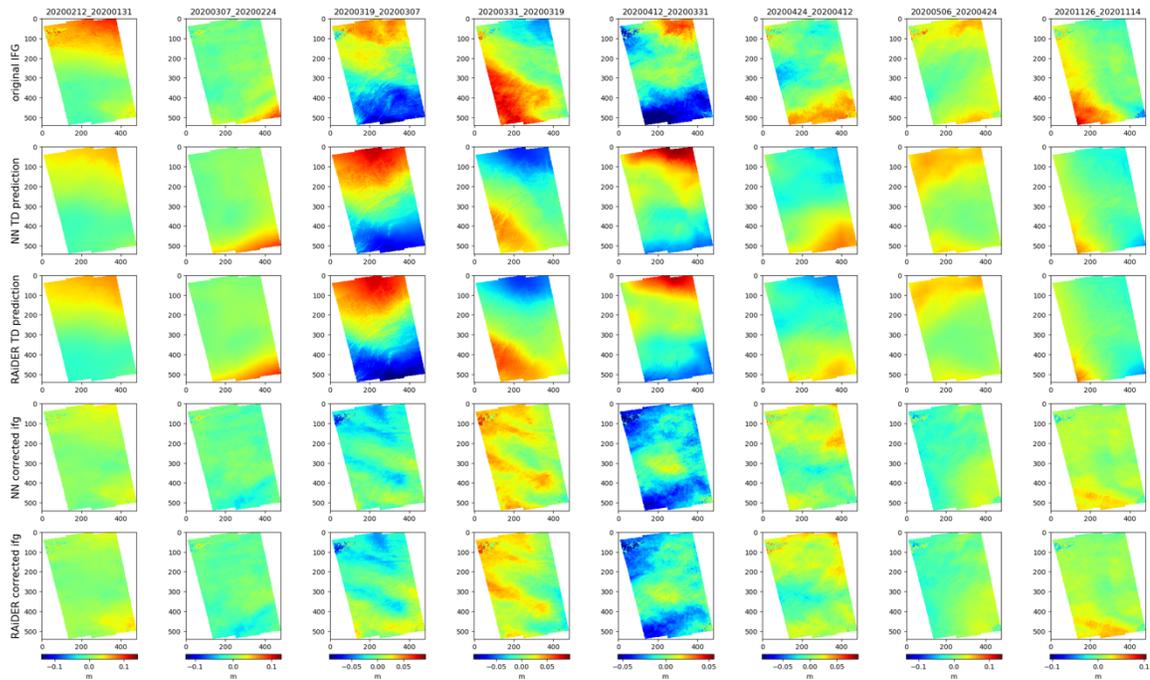


Figure 4.24. Tropo-correction for interferogram in Pennsylvania with NN model and RAiDER.

While both the neural network model and RAiDER were effective in correcting most of the noise from the troposphere, there appears to be a limitation to the extent of correction that can be achieved. By examining each interferogram in Figure 4.24, we can observe the correction made by the proposed method. Although some data may suggest that the Global neural network model outperformed the RAiDER prediction, it is still challenging to compare the two methods visually. To measure the improvement of the interferogram after correction, a metric was utilized that calculated the change in standard deviation from the original interferogram to the corrected ones. Both RAiDER and the neural network were able to minimize the standard deviation by 30-60%.

Table 4.3. Compare of RMSE with different methods by date (Pennsylvania).

Date pair	Original IFG std (m)	NN corrected IFG std (m)	WM corrected IFG std (m)
20200212_20200131	0.03148	0.01173	0.01133
20200307_20200224	0.01450	0.00958	0.00979
20200319_20200307	0.03778	0.01378	0.01773
20200331_20200319	0.03168	0.01334	0.01662
20200412_20200331	0.02324	0.01362	0.01258
20200424_20200412	0.01976	0.01312	0.01362
20200506_20200424	0.01865	0.01777	0.01553
20201126_20201114	0.02263	0.01017	0.01161

From the date pairs of 20203019_20200307, 20200331_20200319, 20200412_20200331 in Figure 4.24, we can observe that these interferograms were strongly affected by the tropospheric delay signal with added noise between $\pm 5 - 10$ cm, especially at the northeastern and southwestern part of the interferogram. However, two out of three days show that the Neural network model outperformed the traditional method in the reduction of noise by $\sim 20\%$. In five out of eight ($\sim 60\%$) interferograms, Neural network model was able to achieve a lower reduction in noise by a minimum of $\sim 10\%$.

Detail metrics of each interferogram std and percentage change are in Table 4.3 and Table 4.4.

Table 4.4. Percentage improvement from original std (Pennsylvania)

Date pair	NN corrected	WM corrected
Slave_Master	[%]	[%]
20200212_20200131	-62.73	-64.02
20200307_20200224	-33.91	-32.49
20200319_20200307	-63.51	-53.08
20200331_20200319	-57.90	-47.54
20200412_20200331	-41.38	-45.87
20200424_20200412	-33.61	-31.06
20200506_20200424	-4.70	-16.76
20201126_20201114	-55.04	-48.69

As far as in the testing region in Pennsylvania, we did not observe any reverse effect from the correction method, which means it did not add more noise to the original interferogram. The absence of any reverse effect from the correction method in the testing region is an important observation. The ability to correct for tropospheric delays is crucial for obtaining accurate measurements in interferometry. The presence of noise in interferograms can significantly affect the accuracy of measurements, leading to incorrect results. The use of correction methods such as the Global Neural Network model and RAiDER is therefore essential in ensuring the accuracy of the measurements. The absence of any added noise from the correction method in the testing region suggests that the methods used in this study were effective in minimizing noise and correcting for tropospheric delays.

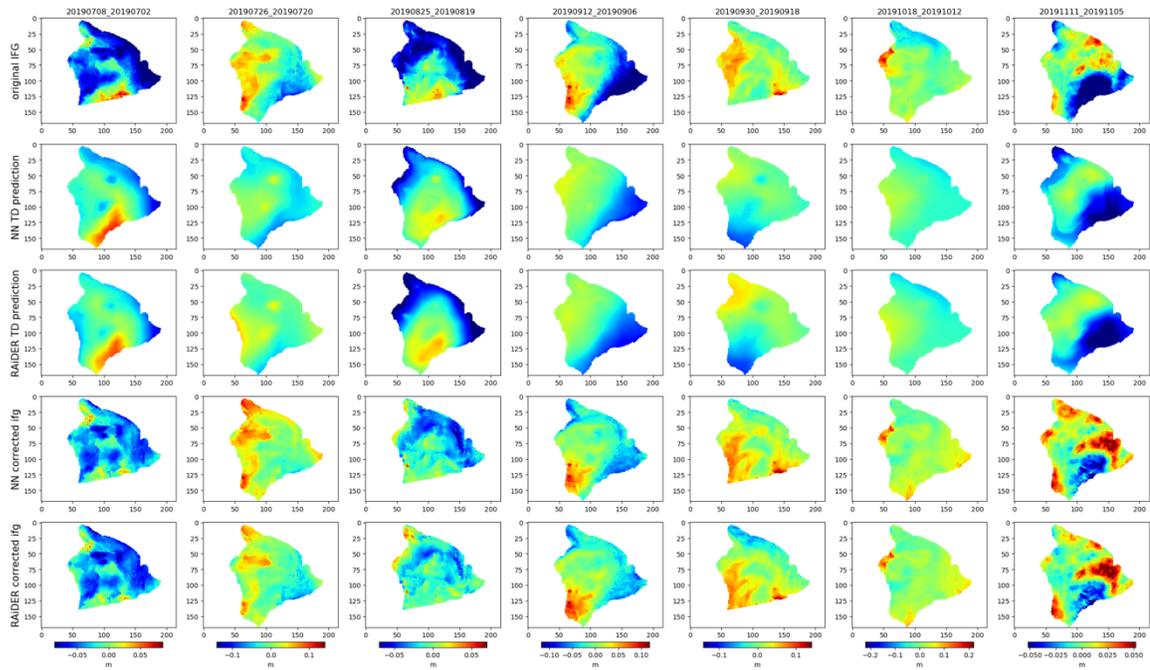


Figure 4.25. Tropospheric correction for interferogram in Hawaii with NN model and RAiDER.

The same correction methods were applied to Hawaii, although it was a challenging task due to its unique characteristics as a tropical island and its unpredictable weather patterns, especially from the nearby volcanoes. Despite these challenges, both the Global Neural Network and RAiDER methods were able to achieve approximately 30% noise reduction. However, there were only two out of the seven dates where the Global Neural Network method outperformed the RAiDER correction. It is important to note that on one date pair (20190930_20190918), both methods were unable to reduce noise but instead introduced more noise to the original interferogram. Nevertheless, the Global Neural Network model introduced only about 1% of noise compared to the approximately 20% increase from the RAiDER method.

The observations from Figure 4.25 showed that most of the tropospheric delay occurs around the coastal regions and nearby mountains where rain shadow effects might

occur. The example of Hawaii demonstrated that turbulent mixing and local weather patterns are difficult to predict and are greatly limited by the coverage of weather model grid points. There is so much variation and alteration within each node in Hawaii, and the spatial resolution of the weather model is unable to provide enough detail of change between the nodes. Therefore, the noise reduction in Hawaii was not able to achieve a similar correction as in Pennsylvania. In other words, the prediction is limited by the spatial resolution of the input from weather model parameters. These challenges highlight the need for further research and development of correction methods for areas with unique weather patterns and topographies such as Hawaii.

Detail metrics of each interferogram std and percentage change are in Table 4.5 and Table 4.6. In summary, these finding provides confidence in the use of these correction methods in future studies.

Table 4.5. Compare of RMSE with different methods by date (Hawaii).

Date pair Slave_Master	Original IFG std (m)	NN corrected IFG std (m)	WM corrected IFG std (m)
20190708_20190702	0.03538	0.02218	0.02328
20190726_20190720	0.04080	0.03094	0.03094
20190825_20190819	0.03355	0.01803	0.01497
20190912_20190906	0.05401	0.03243	0.03331
20190930_20190918	0.02866	0.02891	0.03431
20191018_20191012	0.04085	0.03096	0.02875
20191111_20191105	0.03065	0.02290	0.02170

Table 4.6. Percentage improvement from original std (Hawaii)

Date pair	NN corrected [%]	WM corrected [%]
20190708_20190702	-37.32	-34.20
20190726_20190720	-24.17	-24.17
20190825_20190819	-46.28	-55.37
20190912_20190906	-39.95	-38.32
20190930_20190918	0.86	19.70
20191018_20191012	-24.23	-29.62
20191111_20191105	-25.27	-29.18

5. DISCUSSION

5.1. GNSS BIAS

The ZTD measurement provided by GNSS stations is highly accurate and not constrained by areas with crustal deformation. However, the distribution of GNSS stations is not uniform, especially in Asia and Africa continents. This non-uniformity could lead to bias in the prediction of the machine learning model, as it relies on the ZTD data of GNSS stations from all around the world for training. Figure 5.1 depicts the count of GNSS stations by latitude, which shows a significant difference in the number of GNSS stations in the northern hemisphere compared to the southern hemisphere. This variation is mainly due to the concentration of GNSS stations in three primary regions, i.e., the USA, Europe, and Japan. Two-thirds of the global GNSS stations are located in the northern hemisphere, which may result in the model predicting better in this region. Moreover, the distribution of GNSS stations is heavily skewed towards these regions, which can cause bias in the machine learning model, making it perform better in these areas. Therefore, the lack of GNSS stations in some regions, especially Asia and Africa, can potentially impact the accuracy of the machine learning model.

The non-uniform distribution of GNSS stations can pose challenges for machine learning models, particularly in regions with limited data. In some areas, the lack of GNSS data can result in larger errors in ZTD prediction. For instance, in tropical regions where convective processes significantly affect precipitation formation, the scarcity of GNSS data can limit our understanding of GNSS ZTD and NWP parameters, leading to incorrect forecasts of rainfall intensity and distribution. Similarly, coastal areas with complex

atmospheric conditions require more precise ZTD measurements, increasing the need for GNSS data. Although increasing the density of GNSS stations in underrepresented regions may not be feasible, weighted sampling of GNSS stations by region can minimize location bias and improve the model's accuracy.

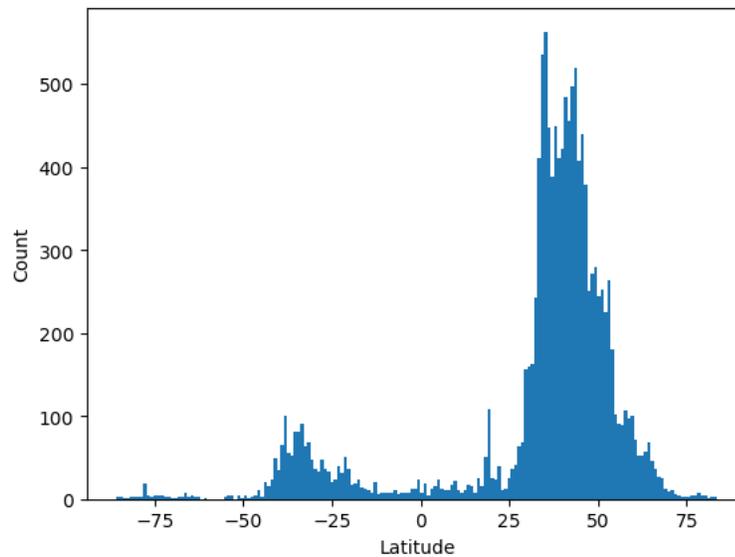


Figure 5.1. Distribution of GNSS station by latitude.

5.2. WEATHER MODEL TIME DEPENDENCY

The temporal resolution of ERA-5 is hourly, which can limit the accuracy of estimation when InSAR acquisitions occur between hours. The application of InSAR for the Hawaii region is an example of such cases. As weather patterns can shift and clouds can form or dissipate within an hour, obtaining an accurate tropospheric delay for correction can be challenging. Moreover, turbulent components are time-sensitive, and while their prediction may be correct, their location can be inaccurate. As a result, the accuracy of correction can be affected.

To demonstrate this issue, the Global neural network model was used to predict tropospheric delay over Hawaii from 4:00:00 UTC to 5:00:00 UTC, with a 10-minute increment, as shown in the Figure 5.2 below. The increments were obtained through linear interpolation of parameters from two weather models since they were not available in the original weather model. The figure shows a red phase change on the western side of the island starting at 0 minutes, and with each increment, the red phase change disappears to the west, along with other noticeable features. Although this may not be a significant issue in predicting the tropospheric effect in most cases, the actual weather pattern change in the area remains highly uncertain because there is not a way to validate such weather pattern change. An animated version of this plot is available on the GitHub repository to observe the transition of weather patterns within an hour.

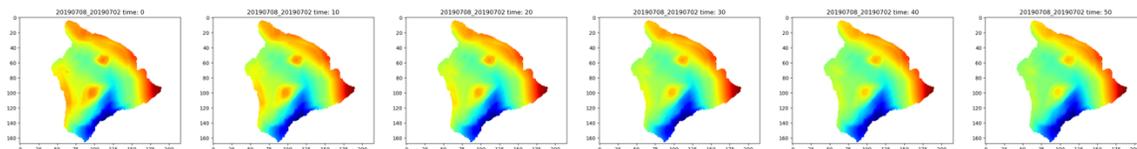


Figure 5.2. Tropospheric delay changes in the increment of 10 mins.

A test was conducted to compare the results obtained by getting the weather model for the closest hour (04:00:00 UTC) and getting the exact minute that matches the SAR acquisition time through linear interpolation (04:30:00 UTC). From Table 5.1, close to 90 % of the interferograms tested show an increase in standard deviation by at least 0.42% and up to 6.31%. Interestingly, for a date pair (20190912_20190906), the standard deviation decreased by 2% from the original result. Although these changes are not significant, it is important to note that weather models have limitations in obtaining the most accurate parameters to match with the SAR acquisition time. Hence, the resolution of

the weather model will be key to providing high-accuracy tropospheric prediction for InSAR correction.

ERA-5 provides high-resolution atmospheric data for InSAR correction; however, its hourly temporal resolution can limit its accuracy for both machine learning and traditional methods. Weather patterns and turbulent components can change rapidly within an hour, particularly in tropical areas such as Hawaii. Hence, getting an accurate tropospheric delay for correction can be challenging. The figure and table presented in this section demonstrate the impact of obtaining the correct interpolate weather model through time, which can result in increased standard deviation. Therefore, it is crucial to consider the limitations of the weather model in obtaining the most accurate parameters to match the SAR acquisition time, and the resolution of the weather model should be considered to ensure high-accuracy tropospheric prediction for InSAR correction.

Table 5.1. Compare correction without interpolation in time in Hawaii.

Date pair	Original IFG std (m)	NN corrected by interp hour std (m)	NN corrected by closest hour std (m)	Percentage change [%]
20190708_20190702	0.03538	0.02218	0.02358	+6.31
20190726_20190720	0.04080	0.03094	0.03107	+0.42
20190825_20190819	0.03355	0.01803	0.01850	+2.61
20190912_20190906	0.05401	0.03243	0.03177	-2.04
20190930_20190918	0.02866	0.02896	0.03058	+5.60
20191018_20191012	0.04085	0.03096	0.03169	+2.36
20191111_20191105	0.03065	0.02290	0.02304	+0.61

5.3. FUTURE SCOPE

There are so much needed to be done and to add on to this research in tropospheric correction. Machine learning have shown some promising result from the usage of weather model outputs (Pressure, Temperature, and water vapor) to predict tropospheric correction and comparable to traditional methods. Here are some ideas would be great for future work.

5.3.1. A Different Machine Learning Approach. This study has shown that machine learning can effectively utilize weather model output to predict tropospheric effect. However, it should be noted that fully connected neural networks may not be able to account for surrounding pixels. This is because, in this study, each pixel is treated as an individual setting, and there is no mechanism to consider the neighboring pixels. To address this issue, future work could focus on using a convolutional neural network (CNN) style of filter-based machine learning architecture. This approach can account for the three-dimensional data structure from the weather model and directly train with the interferograms' target. However, one of the main challenges of using a CNN is the resolution mismatch between the weather model and interferogram since the weather model is coarser. Nevertheless, this approach can offer better results by capturing spatial dependencies and providing a more accurate prediction of the tropospheric effect. Furthermore, the use of transfer learning could also be explored in future studies to further improve the performance of the CNN by leveraging pre-trained models on large datasets.

Another possible architecture that is similar to transfer learning involves utilizing a stacking model. This includes obtaining multiple individual machine learning outputs and using them to create another hierarchical neural network model to learn the relationship

with the actual ZTD. This approach can take advantage of all the different models and consolidate them into a larger, stacked machine learning model.

5.3.2. Using Weather Model with Higher Spatial Resolution. The current ERA5 weather model has limitations due to its spatial resolution of 25-30 km. This can be problematic in areas with high variability in local weather patterns, such as Hawaii, as it can limit the model's ability to capture changes within its grid points. To address this, it is suggested to test a higher resolution weather model for its performance and application in InSAR tropospheric correction. For instance, the HRES climate analysis model from ECMWF provides atmospheric and climate variables every 6 hours with a spatial resolution of approximately 10 km. By using a higher resolution weather model, it may be possible to improve the accuracy ZTD prediction hence improve InSAR tropospheric correction in areas with complex local climate conditions.

5.3.3. Implement of GOES Data. Yu et al. (2021) and Bekaert et al. (2015) have demonstrated that spectrometry can provide highly accurate predictions of wet delay, which in turn can improve predictions of interferometric delay. The Geostationary Operational Environmental Satellite (GOES) system is a valuable resource for weather forecasting, meteorological research, and environmental analysis by the National Oceanic and Atmospheric Administration (NOAA). GOES-16 and 17 provide coverage of the Western Hemisphere with a high temporal resolution of 5-10 minutes depending on the mode of data collection, and spatial resolution ranging from 0.5 km to 2 km depending on the wavelength. The satellite uses an Advanced Baseline Imager (ABI) with 16 different spectral bands, including two visible channels, four near-infrared channels, and ten infrared channels. The satellite is particularly useful in bands 8, 9, and 10, which monitor water

vapor/moisture in the troposphere at low, mid, and upper levels, respectively. The data are collected using an infrared band, and the product is a Brightness Temperature that indicates higher water vapor content with lower temperatures due to absorption of energy by water vapor. While the data are not available globally, they can be used as supplementary data to overcome the low spatial resolution in weather models. Figure 5.3 is an example of the band 10 (low troposphere moisture) from GOES-17 imager through the NOAA GOES image viewer, which captures changes in moisture content in Hawaii at 02:50 UTC, 03:20 UTC, and 03:50 UTC. The figure shows the transition of moisture across the island with high spatial and temporal resolution. A gif image for better visualization is available on the GitHub repository.

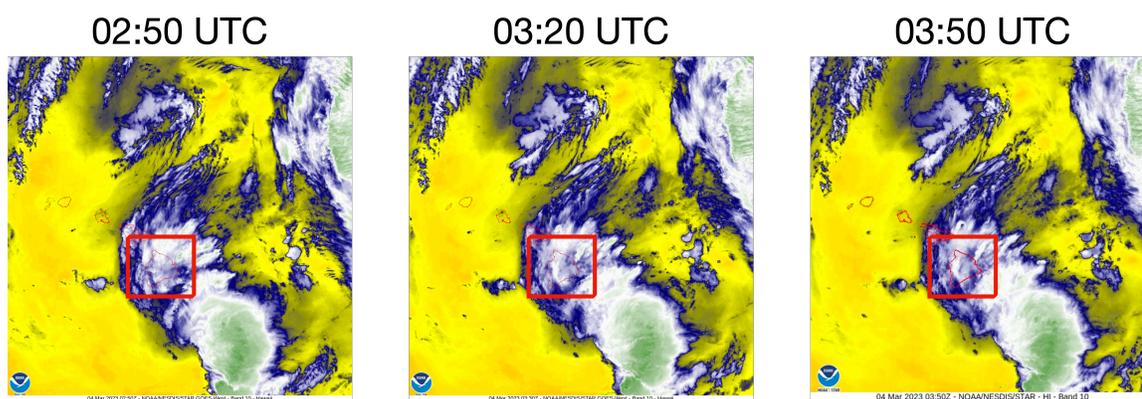


Figure 5.3. GOES-17 band 10 imager showing water vapor content.

6. CONCLUSION

The study employs machine learning models based on numerical weather prediction (NWP) parameters to forecast zenith total delay. The first model, trained on US GNSS stations, yields accurate predictions with a low root mean square error (RMSE) of 0.008734 m, demonstrating its potential for global applications. The second model, a global model trained on a more extensive area with diverse latitudes and elevations, shows normal distribution with precise predictions. However, the neural network model outperforms both models with an overall RMSE of 1.39 cm, which is less than the target of 2 cm. The prediction accuracy of the neural network model correlates with height and latitude, which may be due to the distribution of moisture at different altitudes and latitudes. Additionally, there is a dependence of prediction error on the time of year, which could be attributed to the amount of water vapor in the troposphere during the season.

Notably, both random forest and neural network predictions exhibit significantly lower errors than the traditional method approach. The RMSE, originally 2 cm, was reduced to 1.39 cm, indicating a reduction of approximately 30%. The traditional method approach has a bias in the delay estimation, with an approximate overprediction of delay by 0.5 to 1 cm. The cause of such bias remains unclear. Monthly comparison of RMSE demonstrates that both machine learning predictions outperform the traditional method, suggesting that machine learning may handle wet delay more effectively. While random forest could perform similarly to the neural network approach, building many individual trees from training data leads to large files when storing trained weights for future use.

Regarding the InSAR application, the neural network model effectively corrects tropospheric noise (reduction of ~30-60%) in most interferograms. However, the model

faces challenges in making predictions in areas with high variability in local climate and weather patterns. This is likely due to the limitation of the spatial resolution of weather model output, as the model depends on the input data resolution.

In conclusion, the results indicate that the machine learning models accurately predict GNSS station data and may have practical applications. However, it is important to note that the model's performance may vary depending on the context and dataset used. Further testing and optimization may be necessary before deploying the model in real-world scenarios. Additionally, the bias in the traditional method approach requires further investigation to improve its performance. Finally, the limitations of the model in areas with high variability in local climate and weather patterns suggest the need for future research to overcome these challenges.

APPENDIX

Feature Importance for model

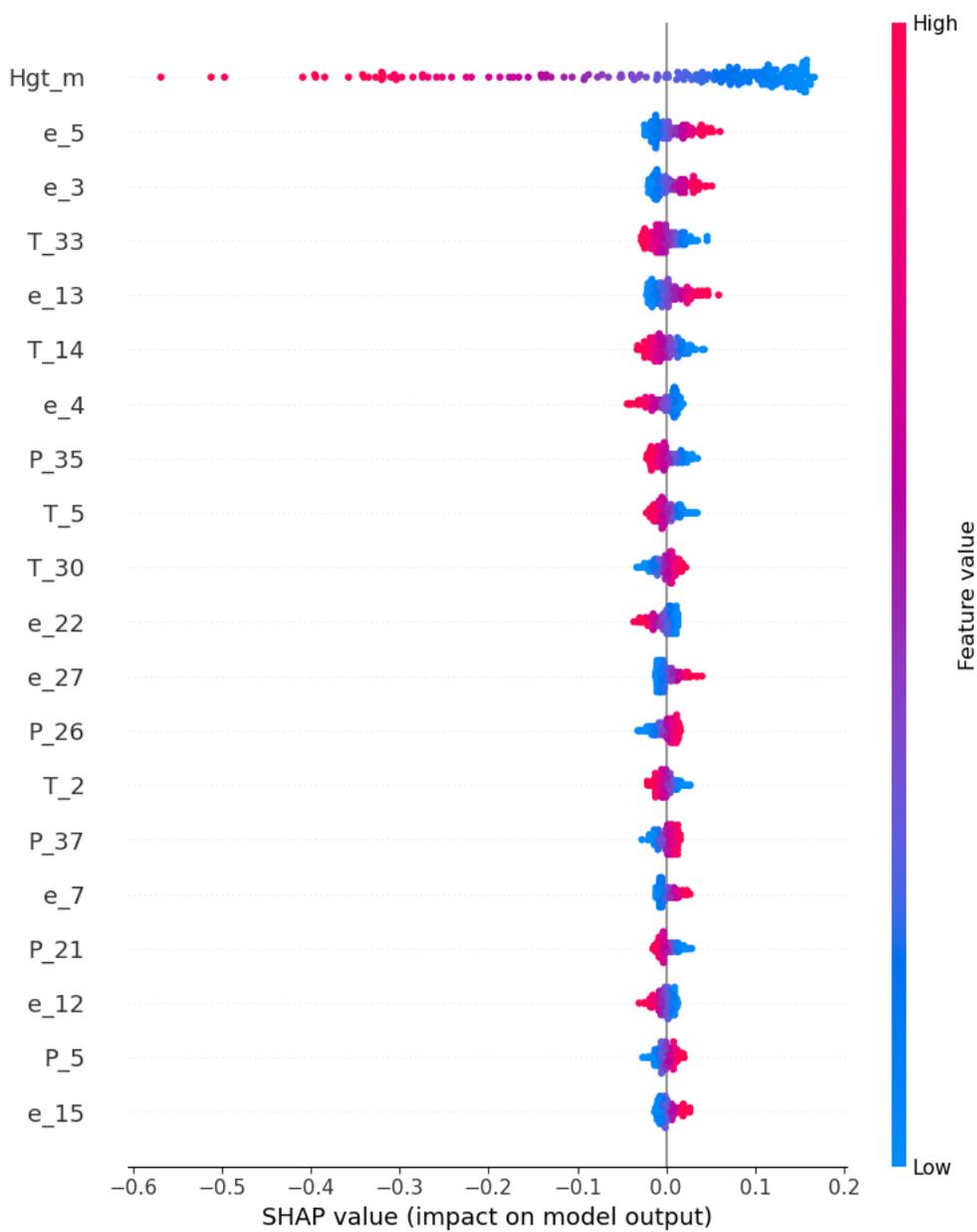


Figure A.1. SHAP Feature importance for model.

REFERENCE

- Albino, F., Biggs, J., Yu, C., & Li, Z. (2020). Automated Methods for Detecting Volcanic Deformation Using Sentinel-1 InSAR Time Series Illustrated by the 2017–2018 Unrest at Agung, Indonesia. *Journal of Geophysical Research: Solid Earth*, 125(2), e2019JB017908.
- Bai, T., Li, D., Sun, K., Chen, Y., & Li, W. (2016). Cloud Detection for High-Resolution Satellite Imagery Using Machine Learning and Multi-Feature Fusion. *Remote Sensing* 2016, Vol. 8, Page 715, 8(9), 715
- Barnhart, W. D., & Lohman, R. B. (2013). Phantom earthquakes and triggered aseismic creep: Vertical partitioning of strain during earthquake sequences in Iran. *Geophysical Research Letters*, 40(5), 819–823
- Bekaert D., M. Karim, L. Justin, H. Hua, P. Agram, S. Owen, G. Manipon, N. Malarout, M. Lucas, G. Sacco, L. Pan, S. Sangha, and ARIA team (2019), Development and Dissemination of Standardized Geodetic Products by the Advanced Rapid Imaging and Analysis (ARIA) Center for Natural Hazards, The International Union of Geodesy and Geophysics (IUGG), Montreal
- Bekaert, D. P. S., Hooper, A., & Wright, T. J. (2015). A spatially variable power law tropospheric correction technique for InSAR data. *Journal of Geophysical Research: Solid Earth*, 120(2), 1345–1356
- Bekaert, D. P. S., Walters, R. J., Wright, T. J., Hooper, A. J., & Parker, D. J. (2015). Statistical comparison of InSAR tropospheric correction techniques. *Remote Sensing of Environment*, 170, 40–47
- Benevides, Pedro & Catalao, Joao & Miranda, Pedro & Chinita, Maria. (2013). Analysis of the relation between GPS tropospheric delay and intense precipitation. *Proceedings of SPIE - The International Society for Optical Engineering*.
- Blewitt, G., W. C. Hammond, and C. Kreemer (2018), Harnessing the GPS data explosion for interdisciplinary science, *Eos*, 99
- Bottou, L. (2012). Stochastic Gradient Descent Tricks. In *Neural Networks: Tricks of the Trade* (pp. 421-436). Springer Berlin Heidelberg.
- Buzzanga, B., Bekaert, D. P. S., Hamlington, B. D., & Sangha, S. S. (2020). Towards Sustained Monitoring of Subsidence at the Coast using InSAR and GPS: An Application in Hampton Roads, Virginia. *Geophysical Research Letters*, 47,

- Cao, Y., Jónsson, S., & Li, Z. (2021). Advanced InSAR Tropospheric Corrections From Global Atmospheric Models that Incorporate Spatial Stochastic Properties of the Troposphere. *Journal of Geophysical Research: Solid Earth*, 126(5),
- Davis, J., Herring, T., Shapiro, I., Rogers, A., & Elgered, G. (1985). Geodesy by radio interferometry: effects of atmospheric modeling errors on estimates of baseline length. *RadioScience*, 20
- Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., & Dehmer, M. (2020). An Introductory Review of Deep Learning for Prediction Models With Big Data. *Frontiers in Artificial Intelligence*, 3, 4.
- Funning, G. J., Parsons, B., Wright, T. J., Jackson, J. A., & Fielding, E. J. (2005). Surface displacements and source parameters of the 2003 Bam (Iran) earthquake from Envisat advanced synthetic aperture radar imagery. *Journal of Geophysical Research: Solid Earth*, 110(B9).
- Glorot, X. and Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249-256).
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning* (Vol. 1). MIT Press.
- Hanssen, R. F. (2001). *Radar interferometry: data interpretation and error analysis* (Vol. 2). Springer Science & Business Media.
- Hanssen, R. F. (2001). *Radar Interferometry*. 2
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Holm, E., Janiskova, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., & Thepaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 2020, 1– 51.
- Hooper, A., & Bekaert, D. P. (2014). Correction of atmospheric and topographic effects for InSAR in volcanic regions: a review and new methods based on permanent scatterer monitoring. *ISPRS Journal of Photogrammetry and Remote Sensing*, 97, 167-186.
- Hu, Z. B., & Mallorquí, J. J. (2019). An Accurate Method to Correct Atmospheric Phase Delay for InSAR with the ERA5 Global Atmospheric Model. *Remote Sensing*, 11(17).

- Jolivet, R., Agram, P. S., Lin, N. Y., Simons, M., Doin, M.-P., Peltzer, G., & Li, Z. (2014). Improving InSAR geodesy using Global Atmospheric Models. *Journal of Geophysical Research: Solid Earth*, 119(3), 2324–2341
- Kang, Y., Lu, Z., Zhao, C., Xu, Y., Kim, J. woo, & Gallegos, A. J. (2021). InSAR monitoring of creeping landslides in mountainous regions: A case study in Eldorado National Forest, California. *Remote Sensing of Environment*, 258, 112400
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Liang, S. (2014). Narrowband to broadband conversions of land surface albedo II: Validation using MODIS data. *Remote Sensing of Environment*, 140, 205-217.
- Liu, Y., & Jade Morton, Y. (2022). Improved Automatic Detection of GPS Satellite Oscillator Anomaly using a Machine Learning Algorithm. *NAVIGATION: Journal of the Institute of Navigation*, 69(1), 69
- Massonnet, D., & Feigl, K. L. (1998). Radar interferometry and its application to changes in the Earth's surface. *Reviews of Geophysics*, 36(4), 441-500.
- Massonnet, D., and Feigl, K. L. (1998), Radar interferometry and its application to changes in the Earth's surface, *Rev. Geophys.*, 36(4), 441– 500
- Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E. H., Ek, M. B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., & Shi, W. (2006). North American Regional Reanalysis. *Bulletin of the American Meteorological Society*, 87(3), 343–360
- Onn, F., & Zebker, H. A. (2006). Correction for interferometric synthetic aperture radar atmospheric phase artifacts using time series of zenith wet delay observations from a GPS network. *Journal of Geophysical Research: Solid Earth*, 111(B9).
- Pritt, M., & Chern, G. (2018). Satellite image classification with deep learning. *Proceedings - Applied Imagery Pattern Recognition Workshop, 2017-October*. <https://doi.org/10.1109/AIPR.2017.8457969>
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195-204.

- Rosen, P. A, Gurrola, Eric M. and Agram, Piyush and Cohen, Joshua and Lavalley, Marco and Riel, Bryan V. and Fattahi, Heresh and Aivazis, Michael A.G. and Simons, Mark and Buckley, Sean M. "The InSAR Scientific Computing Environment 3.0: A Flexible Framework for NISAR Operational and User-Led Science Processing," IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 2018, pp. 4897-4900
- Rosen, P. A., Hensley, S., & Joughin, I. R. (1999). Synthetic aperture radar interferometry. *Proceedings of the IEEE*, 87(3), 288-307.
- Rousset, B., Agram, P. S., Jolivet, R., & Simons, M. (2018). A statistical correction for InSAR tropospheric phase delay using time series of GPS data: Application to Sentinel-1 in California. *Geophysical Research Letters*, 45(14), 6988-6996
- Shamshiri, R., Motagh, M., Nahavandchi, H., Haghshenas Haghghi, M., & Hoseini, M. (2020). Improving tropospheric corrections on large-scale Sentinel-1 interferograms using a machine learning approach for integration with GNSS-derived zenith total delay (ZTD). *Remote Sensing of Environment*, 239, 111608
- Selbesoglu, M. O. (2020). Prediction of tropospheric wet delay by an artificial neural network model based on meteorological and GNSS data. *Engineering Science and Technology, an International Journal*, 23(5), 967–972
- Smith, E., & Weintraub, S. (1953). The constants in the equation for atmospheric refractiveindex at radio frequencies. *Proceedings of the IRE*, 41, 1035–1037.
- Tymofyeyeva, E., & Fialko, Y. (2018). Geodetic evidence for a blind fault segment at the southern end of the San Jacinto Fault Zone. *Journal of Geophysical Research-Solid Earth*, 123(1), 878–891.
- Wadge, G., Webley, P. W., James, I. N., Bingley, R., Dodson, A., Waugh, S., Veneboer, T., Puglisi, G., Mattia, M., Baker, D., Edwards, S. C., Edwards, S. J., & Clarke, P. J. (2002). Atmospheric models, GPS and InSAR measurements of the tropospheric water vapour field over Mount Etna. *Geophysical Research Letters*.
- Wang, C., Moreira, A., Datcu, M., Wu, F., Xiao, R., Chen, C., Dai, K., Tang, X., Cheng, J., Pirasteh, S., Wu, M., Shi, X., Zhou, H., & Li, Z. (2022). Removing InSAR Topography-Dependent Atmospheric Effect Based on Deep Learning. *Remote Sensing 2022*, Vol. 14, Page 4171, 14(17), 4171. <https://doi.org/10.3390/RS14174171>
- Yu, C., Li, Z. H., Penna, N. T., & Crippa, P. (2018). Generic Atmospheric Correction Model for Interferometric Synthetic Aperture Radar Observations. *Journal of Geophysical Research: Solid Earth*, 123, 9202–9222.

- Yu, C., Li, Z., & Penna, N. T. (2018). Interferometric synthetic aperture radar atmospheric correction using a GPS-based iterative tropospheric decomposition model. *Remote Sensing of Environment*, 204, 109–121.
- Yu, C., Li, Z., & Blewitt, G. (2021). Global Comparisons of ERA5 and the Operational HRES Tropospheric Delay and Water Vapor Products With GPS and MODIS. *Earth and Space Science*, 8(5), e2020EA001417.
- Zebker, H. (2021). Accuracy of a Model-Free Algorithm for Temporal InSAR Tropospheric Correction. In *Remote Sensing* (Vol. 13, Issue 3).
- Zebker, H. A., & Goldstein, R. M. (1986). Topographic mapping from interferometric synthetic aperture radar observations. *Journal of Geophysical Research: Solid Earth*, 91(B5), 4993-4999.
- Zebker, H. A., Rosen, P. A., & Hensley, S. (1997). Atmospheric effects in interferometric synthetic aperture radar surface deformation and topographic maps. *Journal of Geophysical Research: Solid Earth*, 102(B4), 7547–7563.
- Zheng, Y., Lu, C., Wu, Z., Liao, J., Zhang, Y., & Wang, Q. (2022). Machine Learning-Based Model for Real-Time GNSS Precipitable Water Vapor Sensing. *Geophysical Research Letters*, 49(3).
- <https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation#heading-Observations>, Observation of ERA-5, Feb 2023
- https://github.com/ngohikennyue/Tropo_corrections , GitHub repository for this thesis - Tropospheric correction

VITA

Ngo Hi Kenny Yue received his Bachelor of Science in Geospatial Computing and Minor in Geology from Brigham Young University – Idaho in 2019. In May 2023 he received his master's degree in Geological Engineering from Missouri University of Science and Technology