



**POLITECNICO**  
**MILANO 1863**

*POLITECNICO DI MILANO*

*School of Civil, Environmental and Land Management Engineering*

*Department of Civil and Environmental Engineering*

*Master of Science in Geoinformatics Engineering*

***Atmospheric water vapor maps generation from  
the stochastic interpolation of  
GNSS Zenith Tropospheric Delays***

**Supervisor:**

*Prof. ssa: GIOVANNA VENUTI*

**Reviewer:**

*Prof.*

**Master Graduation Thesis by:**

*OMAR ABDELGAFAR*

**Student ID.:** *917223*

***Academic Year 2023/2024***



# بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

قَالَ يَاقَوْمِ أَرَأَيْتُمْ إِن كُنتُ عَلَىٰ بَيِّنَةٍ مِّن رَّبِّي وَرَزَقَنِي مِنْهُ رِزْقًا حَسَنًا وَمَا أُرِيدُ أَنْ أُخَالِفَكُمْ إِلَىٰ مَا أَنهَلَكُم عَنْهُ إِن أُرِيدُ إِلَّا الْإِصْلَاحَ مَا اسْتَطَعْتُ وَمَا تَوْفِيقِي إِلَّا بِاللَّهِ عَلَيْهِ تَوَكَّلْتُ وَإِلَيْهِ أُنِيبُ

صدق الله العظيم

*“He said, "O my people, have you considered: if I am upon clear evidence from my Lord and He has provided me with a good provision from Him...? And I do not intend to differ from you in that which I have forbidden you; I only intend reform as much as I am able. And my success is not but through Allah. Upon Him, I have relied, and to Him, I return”*

(The Qur'an, 1965, Surah Hūd: 88)



# ***DEDICATION***

*With heartfelt gratitude to my family, especially the loving souls of my grandparents, I dedicate this work. Their unwavering love and support have shaped my journey profoundly*

.



# *Acknowledgments*

I would like to take this opportunity to express my heartfelt gratitude to all those incredible individuals who have been instrumental in supporting me throughout my journey.

First and foremost, I am immensely grateful to my esteemed academic supervisor, Prof. ssa Giovanna Venuti. Her unwavering support, valuable time, and continuous encouragement have been pivotal in shaping my thesis and providing me with invaluable guidance. Prof. Venuti's expertise and dedication have truly been a beacon of inspiration for me.

A special thanks go to the remarkable team at GReD company, especially Prof. Eugenio Realini and Andrea Gatti. Their generosity in providing us with invaluable case study data has been a significant milestone in my work. Their expertise and willingness to assist have been a true testament to their dedication to fostering academic growth.

I would also like to extend my heartfelt appreciation to my esteemed colleague, Eng. XingYang. Our collaborative efforts and insightful discussions, particularly during the Ordinary Kriging review, have played a vital role in refining my work. XingYang's commitment to excellence and his unwavering support has been invaluable throughout this process.

Furthermore, I am indebted to my beloved parents for their unwavering support and unconditional love. They have been my pillars of strength, and I owe everything I have achieved thus far to their endless encouragement and guidance. This work is dedicated to them as a token of my gratitude and appreciation.

I would also like to express my profound gratitude to my brother, Eng. Ahmed AbdelGafar. His unwavering support and constant presence during the most challenging moments have been a source of immense strength for me. I am truly blessed to have him by my side.

In addition, I would like to extend my heartfelt thanks to my dear friends Eng. Ahmed Elsayed, Eng. Omar El Shwarby, and Eng. Abdullah Bilal. Their unwavering friendship and the incredible moments we have shared along this path have been a constant source of joy and inspiration.

To all the individuals who have played a role, big or small, in supporting and encouraging me, I am eternally grateful. Your belief in me has been a driving force, and I am honored to have had you as part of my journey. Thank you from the bottom of my heart.





# *Abstract*

Weather events monitoring in modern meteorology necessitates the use of advanced techniques, with global navigation satellite systems (GNSS) emerging as a prominent solution. To enhance positioning accuracy, the propagation delay affecting GNSS signals in the troposphere is meticulously modeled and estimated. This delay represents the cumulative effect of signal travel in the zenith direction above the GNSS station and consists of two components: the stratified component, associated with gases in hydrostatic equilibrium, and the turbulence component, attributed to the non-equilibrium behavior of water vapor. Although water vapor also contributes to the stratified part, a precise pointwise estimation of Zenith Total Delays (ZTDs) is crucial for meteorological and forecasting applications.

This thesis focuses on utilizing GNSS-sensed tropospheric delay data to generate high-resolution maps of ZTDs, with a specific emphasis on describing the spatial distribution of water vapor and its temporal evolution. To accomplish this, collaboration was established with the *POLIMI spin-off GReD*, and data from two case studies were acquired for map production. The first case study involved 34 GNSS stations from the SPIN3 network in northern Italy, spanning two years. The second case study focused on a severe weather event in Poland, utilizing data from 278 GNSS reference stations across three networks. Precise Point Positioning (PPP) methodology was employed to derive the ZTD time series with a temporal resolution of 30 seconds.

In the first case study, the topographic variation is significant, ranging from flat areas in the south to the mountainous Albs area in the north. The SPIN3 network exhibits a minimum inner station distance of 19.4 km and a maximum distance of 350 km. Conversely, the second case study covers mostly flat terrain in Poland, except for the southern mountainous region. The three networks in Poland ensure an observation data density of approximately one station per 1120 square kilometers.

Our methodology commenced by disentangling the stratified component from the ZTDs, allowing us to isolate the turbulence component attributed to water vapor. Subsequently, a stochastic prediction approach was adopted, assuming homogeneity and isotropy of the turbulence field. To quantify the spatial correlation among turbulence observations, empirical variograms were computed. Two fitting models, namely the exponential and power models, were considered for variogram fitting. By employing ordinary kriging and utilizing the theoretical variogram model parameters, turbulence maps and estimation error maps were generated for each study area, encompassing specific epochs of interest.

To evaluate the accuracy of our estimations, Leave One Out cross-validation (LOOCV) was performed for both case studies. Additionally, for the second case study, cross-validation was conducted using one of the three networks. The annual average root mean square error (RMSE) obtained for the SPIN3 network in Italy was determined to be 8.8 mm. In the case of Poland, LOOCV yielded an annual average RMSE of 1.2 cm, while cross-validation using another network resulted in an average RMSE of 7.6 mm.

**Key-words:** *GNSS, ZTD, Variogram, Kriging*



# *Abstract in lingua italiana*

Il monitoraggio degli eventi meteorologici nella moderna meteorologia richiede l'uso di tecniche avanzate, con i sistemi globali di navigazione satellitare (GNSS) che stanno emergendo come una soluzione importante. Per migliorare la precisione del posizionamento, il ritardo di propagazione che interessa i segnali GNSS nella troposfera viene meticolosamente modellato e stimato. Questo ritardo rappresenta l'effetto cumulativo del viaggio del segnale in direzione zenitale sopra la stazione GNSS ed è costituito da due componenti: la componente stratificata, associata ai gas in equilibrio idrostatico, e la componente di turbolenza, attribuita al comportamento di non equilibrio del vapore acqueo. Sebbene anche il vapore acqueo contribuisca alla parte stratificata, una precisa stima puntuale degli Zenith Total Delays (ZTDs) è fondamentale per le applicazioni meteorologiche e di previsione.

Questa tesi si concentra sull'utilizzo dei dati di ritardo troposferico rilevati dal GNSS per generare mappe ad alta risoluzione di ZTD, con un'enfasi specifica sulla descrizione della distribuzione spaziale del vapore acqueo e della sua evoluzione temporale. A tal fine è stata avviata la collaborazione con lo spin-off GReD di POLIMI e sono stati acquisiti i dati di due casi studio per la produzione cartografica. Il primo caso di studio ha coinvolto 34 stazioni GNSS della rete SPIN3 nel nord Italia, nell'arco di due anni. Il secondo caso di studio si è concentrato su un grave evento meteorologico in Polonia, utilizzando i dati di 278 stazioni di riferimento GNSS su tre reti. La metodologia Precise Point Positioning (PPP) è stata impiegata per derivare le serie temporali ZTD con una risoluzione temporale di 30 secondi.

Nel primo caso di studio, la variazione topografica è significativa, spaziando dalle aree pianeggianti a sud alla zona montuosa degli Albs a nord. La rete SPIN3 presenta una distanza minima di stazione interna di 19,4 km e una distanza massima di 350 km. Al contrario, il secondo caso di studio copre prevalentemente terreni pianeggianti in Polonia, ad eccezione della regione montuosa meridionale. Le tre reti in Polonia garantiscono una densità di dati di osservazione di circa una stazione ogni 1120 chilometri quadrati.

La nostra metodologia è iniziata separando la componente stratificata dalle ZTD, permettendoci di isolare la componente di turbolenza attribuita al vapore acqueo. Successivamente è stato adottato un approccio di predizione stocastica, assumendo l'omogeneità e l'isotropia del campo di turbolenza. Per quantificare la correlazione spaziale tra le osservazioni di turbolenza, sono stati calcolati variogrammi empirici. Per l'adattamento del variogramma sono stati considerati due modelli di adattamento, vale a dire i modelli esponenziale e di potenza. Utilizzando il kriging ordinario e utilizzando i parametri teorici del modello di variogramma, sono state generate mappe di turbolenza e mappe di errore di stima per ciascuna area di studio, comprendendo epoche specifiche di interesse.

Per valutare l'accuratezza delle nostre stime, è stata eseguita la validazione incrociata Leave One Out (LOOCV) per entrambi i casi di studio. Inoltre, per il secondo studio di caso, è stata condotta una convalida incrociata utilizzando una delle tre reti. L'errore quadratico medio medio annuo (RMSE) ottenuto per la rete SPIN3 in Italia è stato determinato in 8,8 mm. Nel caso della Polonia, LOOCV ha prodotto un RMSE medio annuo di 1,2 cm, mentre la convalida incrociata utilizzando un'altra rete ha prodotto un RMSE medio di 7,6 mm.

**Parole chiave :** *GNSS, ZTD, Variogram, Kriging*



# *Table of contents*

<b>Acknowledgments</b>	<b>vii</b>
<b>Abstract</b>	<b>ix</b>
<b>Abstract in lingua italiana</b>	<b>xi</b>
<b>Table of contents</b>	<b>xiii</b>
<b>List of Figures</b>	<b>xv</b>
<b>List of tables</b>	<b>xix</b>
<b>Acronyms</b>	<b>xxi</b>
<b>1. Introduction</b>	<b>1</b>
<b>2. Background</b>	<b>3</b>
2.1 GNSS signal interaction with the troposphere	3
2.1.1 GPS positioning fundamentals	4
2.1.2 GPS Segments	5
2.1.3 GPS Signal structure	7
2.1.4 GPS Observables	8
2.2 Atmosphere	12
2.2.1 Atmospheric Layers	12
2.2.2 Atmospheric Effects	13
<b>3. Toward Turbulence stochastic modeling</b>	<b>17</b>
3.1 Random variable and Random function	17
3.2 Distribution and Density function	18
3.3 Mean, variance, and covariance for Random variable	19
3.4 Stationarity and Isotropic Assumptions	21
3.5 Turbulence delay as a Random process	23
3.6 Variogram	24
3.6.1 Variogram Cloud	25
3.6.2 Experimental variogram	27
3.6.3 Variogram Model Fitting	28
3.7 Ordinary Kriging	34
	xiii

<b>4. Results and Case Studies</b>	<b>37</b>
4.1 Case Study1	38
4.1.1 Study Area	38
4.1.2 GNSS network	39
4.1.3 Data Processing	43
4.2 Case Study2	69
4.2.1 Study Area and GNSS networks	69
4.2.2 Data Processing	71
<b>5. Conclusion</b>	<b>89</b>
<b>References</b>	<b>91</b>

# List of Figures

Figure 2.1 An Overview of Global and regional satellite navigation .....	3
Figure 2.2 GPS Trilateration Concept .....	4
Figure 2.3 GNSS Segments .....	5
Figure 2.4 GPS Satellites Generations.....	5
Figure 2.5 GPS Control Segment.....	6
Figure 2.6 Navigation message (D Code) frame Structure .....	8
Figure 2.7 The Difference in time between the received and the replica code ( $\Delta TRS$ ).....	8
Figure 2.8 Atmosphere layers classification based on radio waves propagation.....	12
Figure 3.1 ZTD at epoch 00:00:12 on 11/08/2017 in Poland .....	23
Figure 3.2 ZTD at each epoch for station BART1 in Poland.....	23
Figure 3.3 The link vector between point $x_i$ and $x_j$ .....	25
Figure 3.4 Variogram cloud of the estimated turbulence on 11/08/2017 in Poland .....	26
Figure 3.5 Experimental Variogram VS variogram cloud of the estimated turbulence on 11/08/2017 in Poland .....	27
Figure 3.6 Variogram behavior at the origin: (i) Discontinuous behavior, (ii) Continuous but not differentiable behavior, (iii) Continuous and differentiable behavior.....	29
Figure 3.7 Bounded Variogram parameters nugget, sill, and range.....	30
Figure 3.8 Exponential variogram model over the Experimental variogram.....	32
Figure 3.9 Power variogram model over the Experimental variogram where $\alpha = 0.7511$ .....	33
Figure 3.10 Power variogram model with $\alpha = 0.2; 0.5; 1.0; 1.5; \text{ and } 2.0$ .....	33
Figure 4.1 Geographic extent of the study area for Case Study 1 in Italy, showcasing the regions of Lombardy in a blue polygon, Piedmont in a green polygon, and Valle d'Aosta in a red polygon .....	38
Figure 4.2 The full SPIN3 Network permanent stations distribution in Lambordy, Piedmont, and Valle d'Aosta regions.....	39
Figure 4.3 The distribution of the 34 SPIN3 Stations in the study area.....	40
Figure 4.4 SPIN3 Station in Alessandria .....	40
Figure 4.5 Estimated ZTD from SPIN3 Network for Alessandria, Biella, and Bormio stations .....	41
Figure 4.6 Flowchart depicting the data processing methodology for turbulence stochastic prediction using ordinary kriging in case study1 .....	43

Figure 4.7 The Temporal coverage for each station through the period from December 31, 2018, at 23:59:42 to December 31, 2020, at 23:59:42 .....	44
Figure 4.8 Flowchart illustrating the data aggregation process for consolidating all observations from multiple stations into a single matrix while handling empty epochs by assigning NaN values.....	45
Figure 4.9 Flowchart demonstrates how stations are filtered based on the number of NaN values.....	46
Figure 4.10 The Elevation Effect on the Zenith Total Delay.....	47
Figure 4.11 The impact of elevation on the observed Zenit Total Delay (ZTD) by examining three distinct GNSS stations: RUMI with a maximum elevation of 1275.01 m, MANT representing the lowest elevation station of 36.04 m, and MONV station characterized by an average elevation of 587.82 m.....	48
Figure 4.12 ZTD vs Elevation .....	49
Figure 4.13 Flowchart illustrating the steps to how to estimate the turbulence delay starting from ZTD using non-linear least square adjustment.....	51
Figure 4.14 The ZTD, The estimated stratified component, and the turbulence component for station ALSN (on top) and BORM (on bottom) through the period from December 31, 2018, at 23:59:42 to December 31, 2020, at 23:59:42 .....	52
Figure 4.15 Variogram cloud depicting dissimilarities during a half-hour period at 00:29:42 on April 1 <sup>st</sup> , 2019 .....	53
Figure 4.16 Variogram cloud depicting dissimilarities during one day period on April 1 <sup>st</sup> , 2019.....	54
Figure 4.17 Variogram cloud depicting dissimilarities during an hour at 00:59:42 on April 1 <sup>st</sup> , 2019 .....	54
Figure 4.18 Experimental Variogram for April 1st, 2019 at three epochs (00:59:42, 08:59:42, and 18:59:42). The numbers above each point indicate the count of dissimilarities contributing to the average computation within each interval.....	56
Figure 4.19 Theoretical variogram using the Power model for April 1st, 2019 at three epochs (00:59:42, 08:59:42, and 18:59:42).....	57
Figure 4.20 Theoretical variogram using the Exponential model for April 1st, 2019 at three epochs (00:59:42, 08:59:42, and 18:59:42).....	58
Figure 4.21 The Study area grid with a separation distance of 5 km. This grid was projected in (UTM Zone32N) projection coordinate reference system .....	59
Figure 4.22 Trbulence prediction map and the estimation error contour map .....	60
Figure 4.23 Trbulence prediction map and the estimation error contour map .....	61
Figure 4.24 Trbulence prediction map and the estimation error contour map .....	61
Figure 4.25 Flowchart illustrating the steps for ordinary kriging for turbulence delay estimation and the estimation error for each grid point at each epoch.....	62
Figure 4.26 The comparison between the mean, standard deviation (STD), and root-mean-square error (RMSE) of the difference between the estimated turbulence obtained from LOOCV and the observed turbulence for each epoch. Additionally, it displays the mean, STD, and RMSE of the observed turbulence for each epoch. The data presented in the figure spans from January,1 <sup>st</sup> 2019 at 23:59:12 to December,30 <sup>th</sup> 2020 at 23:15:12 .....	64



Figure 4.27 Annual mean bias for SPIN3 Stations .....	65
Figure 4.28 Annual mean RMSE for SPIN3 Stations.....	65
Figure 4.29 The predicted turbulence delay vs the observed turbulence delay and the difference between them for station BORM .....	67
Figure 4.30 The predicted turbulence delay vs the observed turbulence delay and the difference between them for station BUSL.....	67
Figure 4.31 The geographic extent of the study area for Case Study 2 conducted in Poland, along with the positions of the GNSS stations utilized. The ASG-EUPOS network stations are represented by red rectangle markers, the TPI NETpro network stations by green rectangles, and the VRSNet.pl network stations by blue rectangles .....	69
Figure 4.32 Flowchart depicting the data processing methodology for turbulence stochastic prediction using ordinary kriging in case study2.....	71
Figure 4.33 The difference between the orthometric height and the ellipsoidal height. ....	72
Figure 4.34 The Temporal coverage for each station at ASG-EUPOS network through the period on August 11th, 2017, spanning from 00:00:12 to 23:59:12.....	73
Figure 4.35 ZTD vs Elevation .....	74
Figure 4.36 The impact of elevation on the observed Zenit Total Delay (ZTD) by examining three distinct GNSS stations: NWT11 with a maximum elevation of 613.96 m, WLAD1 representing the lowest elevation station of 5.58 m, and ZYWI1 station characterized by an average elevation of 370.91 m .....	74
Figure 4.37 Variogram cloud depicting dissimilarities during a half-hour period at 00:29:42 and a one-hour period at 00:59:42 on August 11 <sup>th</sup> , 2017.....	75
Figure 4.38 Experimental Variogram for August 11 <sup>th</sup> , 2019 at three epochs (00:59:42, 08:59:42, and 18:59:42). The numbers above each point indicate the count of dissimilarities contributing to the average computation within each interval.....	76
Figure 4.39 Theoretical variogram using Power model for August 11 <sup>th</sup> , 2019 at three epochs (00:59:42, 08:59:42, and 18:59:42).....	77
Figure 4.40 Theoretical variogram using Exponential model for August 11 <sup>th</sup> , 2019 at three epochs (00:59:42, 08:59:42, and 18:59:42).....	78
Figure 4.41 The Study area grid that was projected in ETRF2000-PL / CS92 projection coordinate reference system .....	79
Figure 4.42 Trbulence prediction map and the estimation error contour map.....	80
Figure 4.43 Trbulence prediction map and the estimation error contour map.....	81
Figure 4.44 Trbulence prediction map and the estimation error contour map.....	82
Figure 4.45 The comparison between the mean, standard deviation (STD), and root-mean-square error (RMSE) of the difference between the estimated turbulence obtained from LOOCV and the observed turbulence for each epoch. Additionally, it displays the mean, STD, and RMSE of the observed turbulence for each epoch.....	83

Figure 4.46 Annual mean bias for ASG-EUPOS network.....	84
Figure 4.47 Annual mean RMSE for ASG-EUPOS network.....	84
Figure 4.48 The comparison between the mean, standard deviation (STD), and root-mean-square error (RMSE) of the difference between the estimated turbulence obtained from LOOCV and the observed turbulence for each epoch. Additionally, it displays the mean, STD, and RMSE of the observed turbulence for each epoch for VRSNet.pl network.....	85
Figure 4.50 The mean daily RMSE for VRSNet.pl network.....	86
Figure 4.49 The mean daily Bias for VRSNet.pl network.....	86

# *List of tables*

Table 2-1 GPS Binary codes.....	7
Table 4-1 SPIN3 Network -PPP processing setting.....	41
Table 4-2 The 34 SPIN3 stations utilized in our computations, along with the respective province, city, and geodetic coordinates for each station .....	42
Table 4-3 The Bias and the RMSE in cm for 28 SPIN3 station .....	66
Table 4-4 The statistics for SPIN3 GNSS network for the period from January,1st 2019 at 23:59:12 to December,30th 2020 at 23:15:12 .....	68
Table 4-5 the processing strategy employed for the three networks.....	70
Table 4-6 The statistics for ASG-EUPOS GNSS network on August 11th, 2017 from 00:00:12 to 23:59:12 using LOOCV and cross-validation using VRSNet.pl network .....	87



# *Acronyms*

<i>AMCS</i>	Alternate Master Control Station
<i>BLUP</i>	Best Linear Unbiased Prediction
<i>CS</i>	Control Segment
<i>DEM</i>	Digital Elevation Model
<i>DLL</i>	Delay lock loop
<i>ECEF</i>	Earth-center, Earth Fixed coordinate system
<i>GAs</i>	Ground Antennas
<i>GNSS</i>	Global Navigation Satellite system
<i>GPS</i>	Global Positioning System
<i>LOOCV</i>	Leave-One-Out Cross-Validation
<i>MCS</i>	Master Control Segment
<i>MSs</i>	Monitoring Stations
<i>NWP</i>	Numerical Weather Prediction
<i>PDF</i>	Probability Density Function
<i>PMF</i>	Probability Mass Function
<i>PPP</i>	Precise Point Positioning
<i>QZSS</i>	Quasi-Zenith Satellite System
<i>RMSE</i>	Root Mean Square Error
<i>STD</i>	Standard Deviation
<i>SVs</i>	Space Vehicles
<i>ZHD</i>	Zenith Hydrostatic Delay
<i>ZTD</i>	Zenith Total Delay
<i>ZWD</i>	Zenith Wet Delay



# 1. Introduction

The analysis of GNSS signal delays in the neutral atmosphere extends its scope beyond enhancing GNSS positioning through improved tropospheric models. It unlocks a realm of opportunities to measure and comprehend the intricate behavior of atmospheric water vapor, exploring its dynamic temporal and spatial variability. In the realm of meteorology, accurate modeling of water vapor and its dynamics holds the utmost importance for refining numerical weather prediction (NWP) models and effectively forecasting rain events. However, the current limitations of weather models in predicting localized phenomena, such as sudden thunderstorms or torrential rainfall resulting from the rapid formation of clouds on a small scale, necessitate a novel approach. Thus, the establishment of a continuous and dependable water vapor monitoring system at a local level becomes imperative, utilizing a dense GNSS network. Such a system offers valuable insights into water vapor behavior preceding cloud formation, facilitating the integration of this data into NWP models and empowering the development of early warning systems for flood events. This holistic approach to disaster prevention and risk mitigation holds immense promise for safeguarding lives and communities.

The delays affecting GNSS signals as they traverse the troposphere can be attributed to two distinct components: the stratified component and the turbulence component. The turbulence component is directly influenced by the presence of non-equilibrium water vapor along the line of sight between GNSS satellites and receivers. While the stratified component can be reliably modeled using precise ground-level pressure values, addressing the turbulent component poses greater challenges due to its high variability caused by turbulent motions of water vapor, resulting in heterogeneous fluctuations of the refractive index.

This thesis aims to introduce an approach for generating high-resolution water vapor maps in a specific region using a regional GNSS network. Our methodology involves stochastic modeling of the turbulence field as a homogeneous and isotropic random field, employing ordinary kriging interpolation. The thesis comprises several key chapters that contribute to this endeavor.

Chapter 2 provides a comprehensive overview of GNSS positioning concepts, focusing on GPS, along with a discussion on GNSS observations and the Precise Point Positioning (PPP) processing technique. Additionally, it delves into the Earth's atmosphere, exploring its layers and the delays induced by GNSS signal propagation through them.

Chapter 3 presents our methodology for stochastically predicting the turbulence field. It begins with the removal of the stratified component from Zenith Total Delay (ZTD), followed by the introduction of our assumptions applied to the turbulence field. The chapter explores measuring the correlation between observations using variograms and selecting the appropriate variogram model. Furthermore, it introduces the mathematical model for ordinary kriging.

Chapter 4 showcases the application of our methodology through two case studies. The first case study focuses on northern Italy, utilizing the SPIN3 network and spanning a significant time frame of two years. The second case study takes place in Poland, specifically during a severe event, and utilizes three GNSS networks. The results of these case studies are evaluated using Leave-One-Out Cross-Validation (LOOCV) and cross-validation with another network.

Finally, Chapter 5 concludes our work, summarizing the key findings and contributions. It also highlights potential avenues for future research based on our approach.





## 2. Background

### 2.1 GNSS signal interaction with the troposphere

The Global Navigation Satellite System (GNSS) comprises a constellation of satellites that facilitates positioning services for ground receivers by employing a global coordinate reference frame. Ground receivers utilize signals transmitted from satellites with known positions to determine their precise location.

The Transit system, initially developed by the United States Navy during the 1960s, served as the foundational concept for the establishment of GNSS. Subsequently, the United States government advanced this technology by developing the Global Positioning System (GPS) in the 1970s. GPS significantly enhanced the accuracy and coverage capabilities of the transit system. Since achieving full operational capability in 1995, GPS has emerged as the preeminent GNSS system, widely adopted worldwide. It enables users across the globe to attain precise positioning and accurate timing information.

After the United States, other countries developed their navigation satellite systems. In the 1990s, Russia developed the GLONASS system, which provides positioning and timing information to users in Russia and other parts of the world. The European Union started developing the Galileo system in the early 2000s, and it went live in 2016. China has also developed its own satellite navigation system, known as COMPASS-BeiDou, which provides users in China and other parts of the world with positioning and timing information. Japan has created the QZSS (Quasi-Zenith Satellite System) to improve the accuracy and coverage of GPS in Japan. The IRNSS (India's Regional Navigation Satellite System) was created by India. Some of these systems, such as GPS, GLONASS, and Galileo, provide global coverage. On the other hand, some of them, such as QZSS and IRNSS, provide regional coverage [Figure 2.1].







System	GPS	GLONASS	BeiDou	Galileo	QZSS	IRNSS/NavIC
						
Orbit	MEO	MEO	MEO, IGSO, GEO	MEO	IGSO, GEO	IGSO, GEO
Nominal number of satellites	24	24	27, 3, 5	30	3, 1	4, 3
Constellation	6 planes 56° inclination	Walker (24/3/1) 64.8° inclination	Walker (24/3/1) 55° inclination	Walker (24/3/1) 56° inclination	IGSOs with 43° inclination	IGSOs with 29° inclination
Services	SPS, PPS	SPS, PPS	OS, AS, WADS, SMS	OS, CS, PRS	GCS, GAS, PRS, EWS, MCS	SPS, RS
Initial service	Dec 1993	Sep 1993	Dec 2012	2016/2017 (planned)	2018 (planned)	2016 (planned)
Origin	USA	Russia	China	Europe	Japan	India
Coverage	Global	Global	Global	Global	East Asia Oceania region	$-30^\circ < \phi < 50^\circ$ $30^\circ < \lambda < 130^\circ$
Frequency (MHz)	L1 1575:42 L2 1227:60 L5 1176:45	L1 1602:00 L2 1246:00 L3 1202:025	B1 1561:098 B2 1207:14 B3 1268:52	E1 1575:42 E5a 1176:45 E5b 1207:14 E6 1278:75	L1 1575:42 L2 1227:60 L5 1176:45 E6 1278:75	L5 1176:45 S 2492:028
<small>SPS: Standard Positioning Service; PPS: Precise Positioning Service; OS: Open Service; AS: Authorized Service; WADS: Wide Area Differential Service; SMS: Short Message Service; CS: Commercial Service; PRS: Public Regulated Service; GCS: GPS Complementary Service; GAS: GPS Augmentation Service; EWS: Early Warning Service; MCS: Message Communications Service; PS: Precision Service; RS: Restricted Service</small>						

Figure 2.1 An Overview of Global and regional satellite navigation

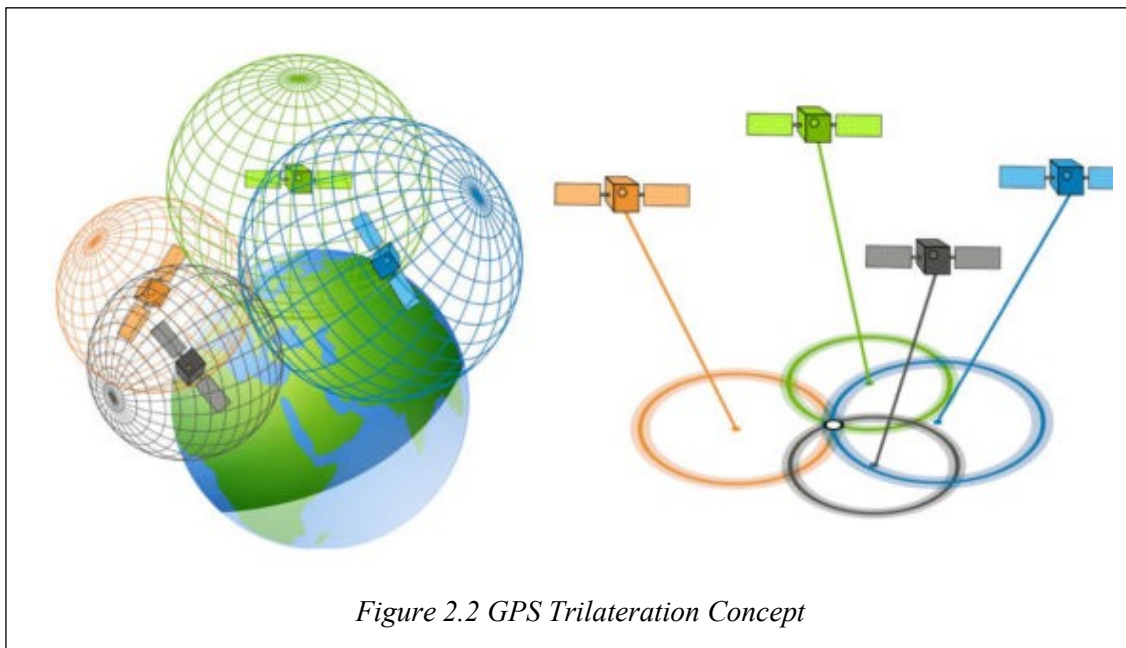
In the next sections, we will focus on GPS as an example of GNSS by understanding its segments, how it works, and its observations that guarantee the positioning.

### 2.1.1 GPS positioning fundamentals

The GPS positioning methodology operates on the fundamental principle of calculating the distances, known as "pseudo ranges," between a receiver situated on Earth and three GPS satellites possessing precisely determined positions. The process of ascertaining the receiver's location relies on the trilateration concept (El-Rabbany, 2002).

GPS signals emanate from satellites positioned in orbit at an altitude of approximately 20,200 km above the Earth's surface. The transmission time for the signal to traverse from a satellite to the receiver, after the generation of a signal replica by the receiver, amounts to approximately 0.07 seconds. By multiplying this temporal discrepancy, referred to as time delay, with the velocity of the signal, corresponding to the speed of light in a vacuum ( $c$ ), the range between the user receiver and the satellite can be reasonably approximated (Wolf & Ghilani, 2006).

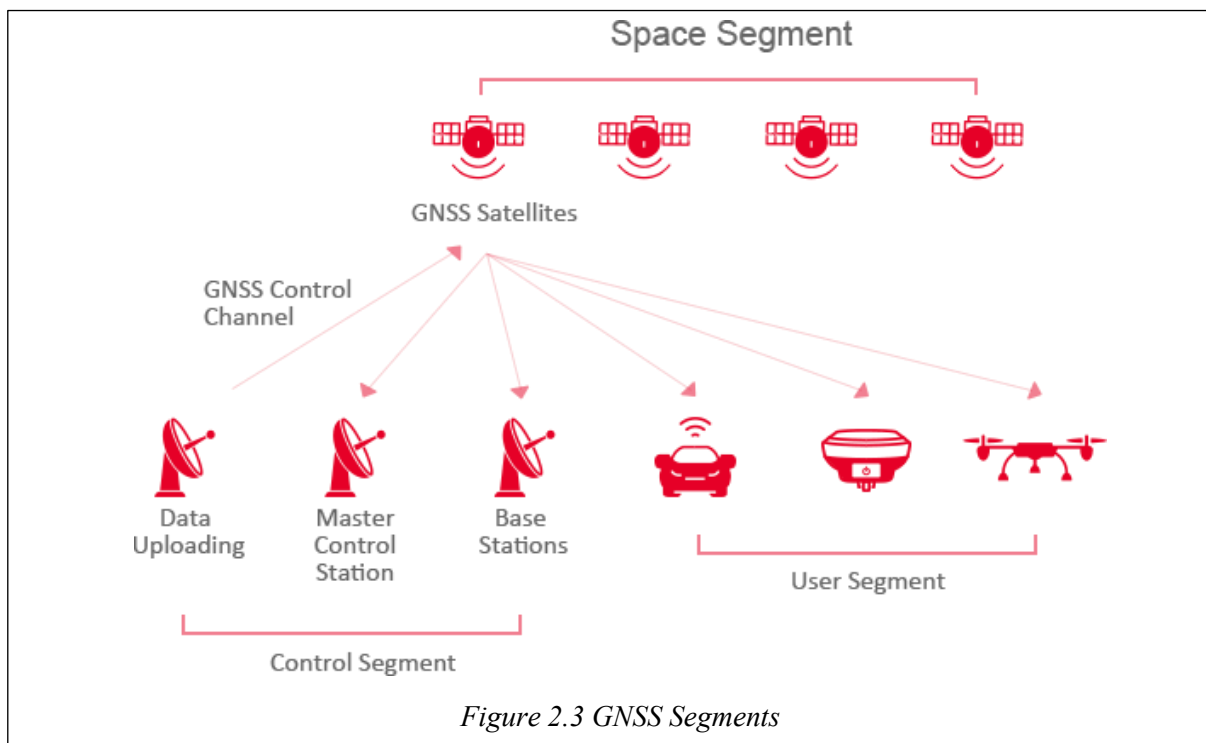
In principle, the localization of the receiver necessitates the knowledge of only three satellite ranges. The receiver's position is determined at the point of intersection among three spheres, wherein each sphere encompasses a respective satellite and possesses a radius equal to the receiver-satellite distance. However, practical considerations require the inclusion of a fourth satellite to account for the receiver's clock offset (El-Rabbany, 2002) Figure [2.2].



*Figure 2.2 GPS Trilateration Concept*

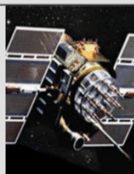
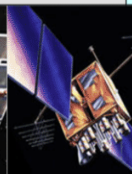
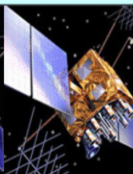
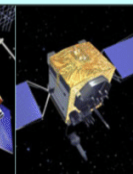
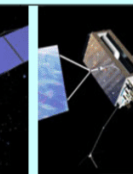
## 2.1.2 GPS Segments

The GPS system can be categorized into three distinct segments: (a) the space segment, (b) the control segment, and (c) the user segment as shown in Figure [2.3].



### 2.1.2.1 Space segment

The GPS Space segment encompasses a satellite constellation that originated with the deployment of a succession of space vehicles (SVs) starting from Block I and advancing to Block III Figure [2.4]. Presently, the constellation consists of 31 satellites orbiting in six distinct orbital planes, positioned at regular intervals of 60° around the equatorial region. These satellites maintain an inclination of 55° relative to the equator and sustain an average orbital altitude of approximately 20,200 km (Biagi, 2022).

Legacy Satellites		Modernized Satellites		
				
<b>Block IIA</b>	<b>Block IIR</b>	<b>Block IIR (M)</b>	<b>Block IIF</b>	<b>GPS III</b>
<b>6 Operational</b>	<b>12 Operational</b>	<b>7 Operational</b>	<b>6 Operational</b>	<b>Now in Production</b>
<ul style="list-style-type: none"> <li>Coarse Acquisition (C/A) code on L1 frequency for civil users</li> <li>Precise P(Y) code on L1 &amp; L2 frequencies for military users</li> <li>7.5-year design lifespan</li> <li>Launched in 1990-1997</li> </ul>	<ul style="list-style-type: none"> <li>C/A code on L1</li> <li>P(Y) code on L1 &amp; L2</li> <li>On-board clock monitoring</li> <li>7.5-year design lifespan</li> <li>Launched in 1997-2004</li> </ul>	<ul style="list-style-type: none"> <li>All legacy signals</li> <li>2nd civil signal on L2 (L2C)</li> <li>New military M code signals for enhanced jam resistance</li> <li>Flexible power levels for military signals</li> <li>7.5-year design lifespan</li> <li>Launched in 2005-2009.</li> </ul>	<ul style="list-style-type: none"> <li>All Block IIR (M) signals</li> <li>3rd civil signal on L5 frequency (L5)</li> <li>Advanced atomic clocks</li> <li>Improved accuracy, signal strength, and quality</li> <li>12-year design lifespan</li> <li>Launched since 2010.</li> </ul>	<ul style="list-style-type: none"> <li>All Block IIF signals</li> <li>4th civil signal on L1 (L1C)</li> <li>Enhanced signal reliability, accuracy, and integrity</li> <li>No Selective Availability</li> <li>Satellites 9+: laser reflectors; search &amp; rescue payload</li> <li>15-year design lifespan</li> <li>Begins launching in 2016.</li> </ul>

*Figure 2.4 GPS Satellites Generations*

### 2.1.2.2 Control segment

The Control Segment (CS) consists of four subsystems: (a) a Master Control Station (MCS), (b) an Alternate Master Control Station (AMCS), (c) a network of four ground antennas (GAs), together with (d) a network of globally-distributed monitor stations (MSs) Figure [2.5] (GPS SPS Performance Standard, 2020).

The primary responsibilities carried out by the ground antennas (GAs) encompass various tasks, including:

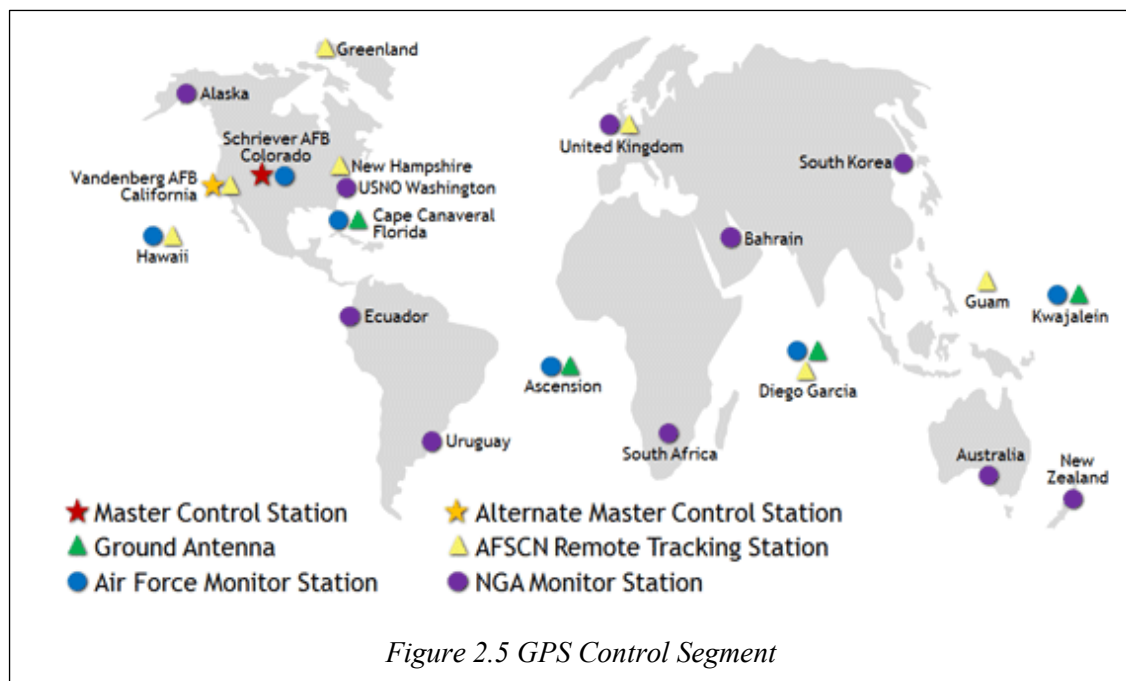
- Tracking and monitoring GPS satellites.
- Collection of atmospheric data and navigation signals.
- Acquiring range/ carrier measurements.
- Transmitting observed data to the Master Control Station (MCS).

For the ground antennas network:

- Updating navigation data.
- Sending commands and loading processor programs to the satellites.
- Collecting telemetry data.
- Providing anomaly resolution and early orbit support.

The MCS oversees all constellation command and control functions, such as:

- Monitoring the health and status of the constellation.
- Resolving satellite maintenance and subsystem anomalies.
- Managing the performance of GPS signals in space to meet performance standards.
- Utilizing data from global monitoring stations to precisely estimate the position of the satellites.
- Updating navigation message parameters, including ephemeris, almanac, and clock corrections.



### 2.1.2.3 User segment

The user segment includes GPS receivers for all military and civilian users, which receive GPS signals afterward solving the navigation equations to estimate their coordinates.

### 2.1.3 GPS Signal structure

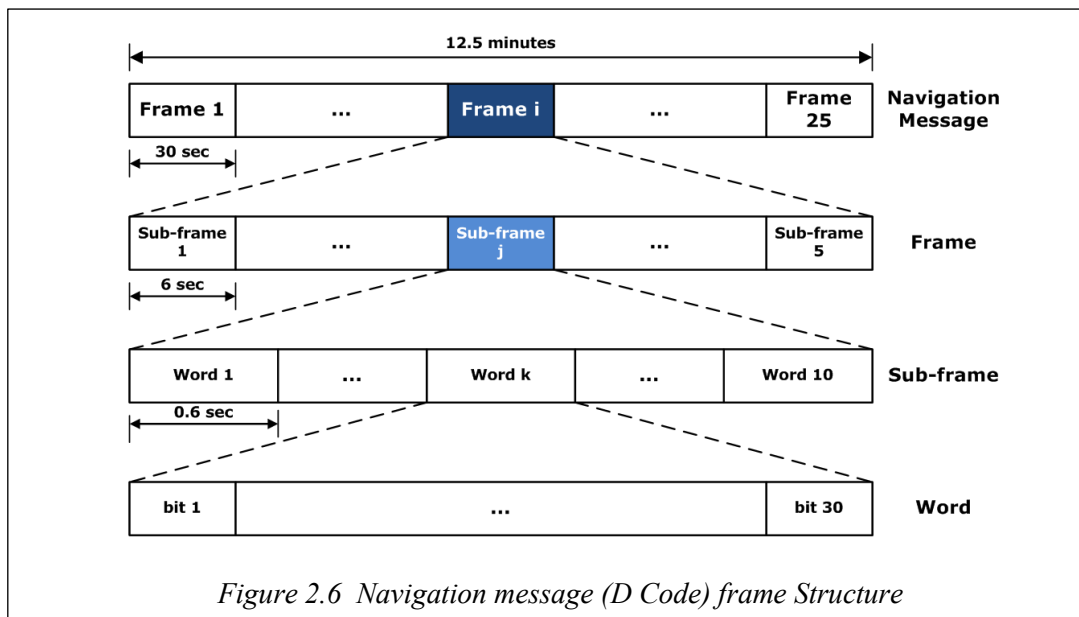
Each satellite in the GPS constellation is onboarded with oscillators that produce a signal with a frequency of 10.23 MHz called the fundamental frequency ( $f_0$ ). From the fundamental frequency, three sinusoidal carrier phases can be generated (L1, L2, and L5 from (Block II-F)) (Biagi, 2022).

Multiplying  $f_0$  by 154, it will give us an L1 carrier(L1) whose frequency is equal to 1575.42 MHz. and the wavelength is 19.03 cm. For L2 carrier(L2), it can be obtained by multiplying  $f_0$  by 120 is equal to 1227.60 MHz with a wavelength of 24.42 cm. The frequency of the L5 carrier is 1176.45 MHz (GPS SPS Performance Standard, 2020).

Accurate determination of receiver position relies on precise measurement of signal travel time from satellites to the receiver. This is achieved through the utilization of pseudorandom noise (PRN) codes, which are generated from a base frequency, denoted as  $f_0$  (Wolf & Ghilani, 2006). At present, the generation of six distinct pseudorandom noise codes is feasible. Among these codes, the C/A code and P-code assume fundamental roles as they are transmitted by each satellite. The C/A code, commonly known as the "coarse acquisition" code, is modulated on the L1 carrier frequency, operating at a frequency of  $0.1f_0$  MHz, corresponding to a wavelength of 293.0 meters. It is important to note that each satellite possesses its unique C/A code. On the other hand, the P-code initially referred to as the *Precise Code*, was transformed in 1994, resulting in an encrypted version designated as P(Y). The adoption of the P(Y) code guarantees improved pseudo-range precision, comparable to that of the C/A code. Recent advancements in GPS satellite technology, particularly the transition from Block II-R to III, have introduced new PRN codes such as L2C and M-code, expanding the available options for satellite-based positioning systems (Biagi, 2022).

One of the codes transmitted by the satellite is the Navigation message also known as the Data code (D-code). It consists of 25 frames of duration 30 seconds each containing 1500bits. Each frame is divided into 5 subframes as shown in Figure [2.6] (Leick, 1995) that contain satellite ephemerides and clock offsets, ionospheric model, and cyclic information about the state of the other orbiting satellites(almanacs) (Biagi,2022).

Name	$f$ (MHZ)	$\lambda$ (m)	Pulse number	T
C/A	$0.1f_0 = 1.023$ MHZ	293.0	1023	1 ms
L2C	$0.1f_0 = 1.023$ MHZ	293.0	1023	10 ms
P(Y)	$f_0 = 10.23$ MHZ	29.3	$3.2703264 \times 1016$	37 weeks
M	$f_0 = 10.23$ MHZ	29.3		
D	50 Hz	$6 \times 10^6$	37500	12.5 min

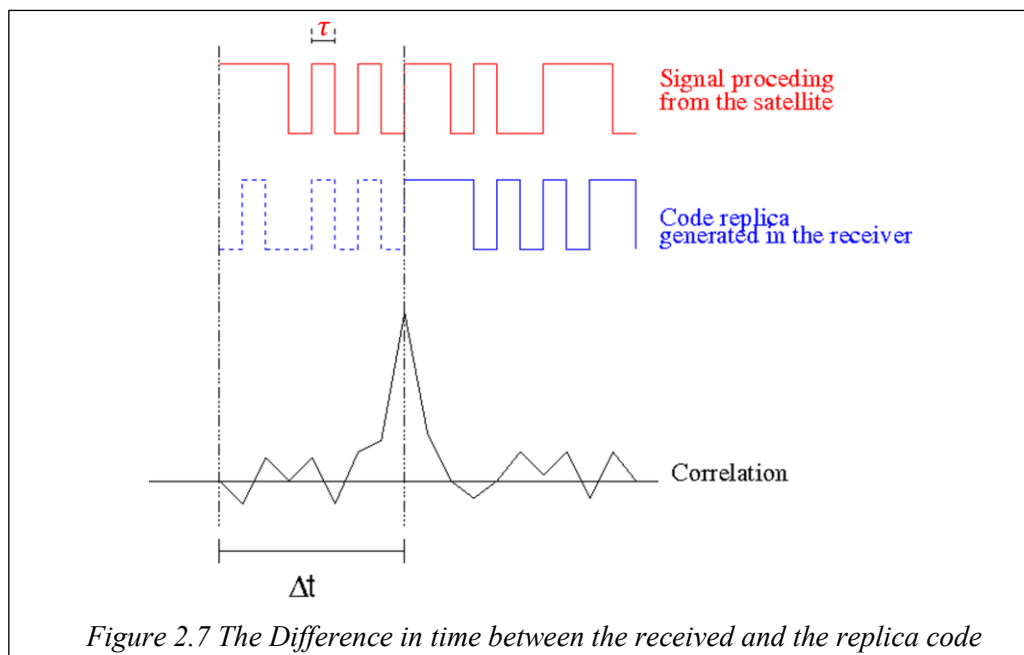


## 2.1.4 GPS Observables

Two primary techniques, code or (Pseudo-range) observation and carrier phase shift measurements, based on either the binary codes or the carrier signals, could be used to calculate the distance between the satellite and the user receiver (Wolf & Ghilani, 2006).

### 2.1.4.1 Pseudo Range observation

Let's consider one epoch ( $t$ ), at this epoch we have several satellites in view. Once the GPS receiver receives a signal which includes the navigation message and the C/A code for each satellite, the receiver automatically separates each of it into different channels (we call these receivers multichannel receivers). The C/A code identifies the satellites in the order letting the receiver know which satellites are on the horizon. For each channel, the C/A code stored will be an input to an electronic device called a correlator also known as a receiver delay lock loop (DLL) (Teunissen & Montenbruck, 2017). Once The DLL received the code, it generates a replica of that code. By comparing the received code with its replica, DLL will be able to determine the shift in time between the two codes ( $\Delta T_R^S$ ). Figure [2.7] (Biagi, 2022).



$$\Delta T_R^S(t) = \tau_R^S + dt_R(t) - dt^S(t) + I_R^S(t) + T_R^S(t) \quad 2-1$$

$\Delta T_R^S$  can be modeled by equation (2-1) where  $\tau_R^S$  is the signal travel time from the satellite signal generator to the DLL in the receiver.  $dt_R(t)$  &  $dt^S(t)$  are the clock offset for the receiver clock and the satellite clock, respectively. The pseudo-range  $P_R^S(t)$  is derived by multiplying  $\Delta T_R^S$  by the signal velocity. It is important to note that the term "pseudo" is employed to signify that the measured value does not represent the actual distance between the satellite and the receiver. This discrepancy arises primarily from the inherent desynchronization between the satellite clock and the receiver clock. Additionally, the pseudo-range is influenced by errors induced during signal propagation through the atmosphere, with particular emphasis on the Ionosphere  $Iono_R^S(t)$  and troposphere layers  $Tropo_R^S(t)$ , thus the pseudo-code observation equation will be:

$$P_R^S(t) = C \times \tau_R^S + C \times [dt_R(t) - dt^S(t) + Iono_R^S(t) + Tropo_R^S(t)] \quad 2-2$$

where:

$$\rho_R^S(t) = C \times \tau_R^S = \sqrt{(X^S - X_R)^2 + (Y^S - Y_R)^2 + (Z^S - Z_R)^2} \quad 2-3$$

where:

$P_R^S(t)$ : is the Pseudo Range at epoch (t) from the receiver (R) to satellite (S).

$(X^S, Y^S, Z^S)$ : The satellite orbital coordinates.

$(X_R, Y_R, Z_R)$ : The receiver ground coordinates.

(C): The speed of light in a vacuum.

$dt_R(t)$ : The receiver clock offset

$dt^S(t)$ : The satellite clock offset.

$Iono_R^S(t)$ : The ionospheric effect.

$Tropo_R^S(t)$ : The tropospheric effect.

From the above equation, we have four unknowns, the receiver coordinates, and its clock offset. To solve this system, we need at minimum 4 equations per epoch so that the need for at least 4 satellites. From this, we can understand the design of the system that guarantees the availability of at least 4 satellites on the horizon at any position on the Earth's surface.

#### 2.1.4.2 Carrier phase observation

The concept of pseudo-range observation can be extended to carrier signals, such as  $L1$  and  $L2$ , where the receiver generates a signal replica upon receiving the satellite signal. In this case, rather than estimating the time shift ( $\Delta T_R^S$ ), the phase difference ( $\varphi_R^S(t)$ ) will be measured:

$$\varphi_R^S(t) = \varphi_R(t) - \varphi^S(t - \tau_R^S) \quad 2-4$$

where:

$\varphi_R^S(t)$ : the phase difference between the satellite received signal and its replica generated by the receiver at epoch  $t$ .

$\varphi_R(t)$ : the receiver oscillator phase at epoch  $t$ .

$\varphi^S(t - \tau_R^S)$ : the satellite oscillator phase at epoch  $t - \tau_R^S$ .

$\tau_R^S$ : the carrier travel time from the satellite to the receiver.

By multiplying the phase difference by the wavelength of the carrier ( $\lambda$ ) we can obtain the carrier phase observation equation  $L_R^S(t)$  given by:

$$\begin{aligned}
 L_R^S(t) &= \lambda\varphi_R^S(t) \\
 &= C\tau_R^S + C[dt_R(t) - dt^S(t)] + Tropo_R^S(t) - Iono_R^S(t) \\
 &\quad + \lambda[\varphi_{0R} - \varphi_0^S + N_R^S(t)]
 \end{aligned}
 \tag{2-5}$$

**where:**

$\lambda$ : the carrier wavelength.

$C$ : the velocity of light in a vacuum.

$dt_R(t)$ : the receiver clock offset.

$dt^S(t)$ : satellite clock offset

$Tropo_R^S(t)$ : tropospheric effect

$Iono_R^S(t)$ : ionospheric effect

$\varphi_{0R}$ : initial phase of the receiver

$\varphi_0^S$ : initial phase of the satellite

$N_R^S(t)$ : the integer number of cycles between the satellite and the receiver "The integer ambiguity".

$\lambda[\varphi_{0R} - \varphi_0^S + N_R^S(t)]$  is known as the ambiguous term  $\lambda\eta_R^S(t)$

In equation (2.5), the term  $N_R^S$  represents the ambiguity in determining the integer number of carrier cycles, which the receiver is unable to directly determine and can only estimate the phase difference. This integer number is subject to change over time due to variations in satellite position and elevation.

Carrier phase observations offer two significant advantages over pseudo-range observations. Firstly, the electronic noise associated with carrier phase measurements is lower, achieving precision levels below 1mm, in contrast to the pseudo-range noise, which typically operates at around 20 cm. Secondly, with the availability of multi-carrier signals such as  $L1$  and  $L2$ , combining these observations enables the creation of a composite measurement, which allows for the mitigation of certain error sources, such as the ionospheric effect (Biagi, 2022).

### 2.1.4.3 Precise Point Positioning Observation

Precise Point Positioning (PPP) emerges as a compelling alternative to differential Global Positioning System (GPS) techniques, which depend on "differential observations" comprising carrier phase or code observations of two stations for processing and resulting in the positioning of one station relative to the other (reference station) (Kaplan & Hegarty, 2017). In contrast, PPP obviates the necessity of reference stations by utilizing undifferentiated observations, encompassing pseudo-range and carrier phase, alongside accurate satellite orbit and clock information.

PPP offers the capability to achieve centimeter-level accuracy in both static and kinematic geodetic point positioning. By leveraging undifferentiated observations and precise orbit and clock products, PPP provides not only precise position estimates but also valuable ancillary information such as station clock offsets and tropospheric zenith total delay (ZTD) (Montenbruck & Teunissen, 2017).

The observation equations for Precise Point Positioning (PPP) typically involve ionosphere-free combinations of dual-frequency GPS pseudo-range ( $P_{r,IF}^S$ ) and carrier phase observations ( $L_{r,IF}^S$ ), as represented by equation (2-6). By employing these combinations, the ionospheric delay effect is mitigated, allowing for more precise positioning solutions. However, in the case of single-frequency observations used in PPP, accurate modeling of the ionospheric delay is necessary to compensate for its influence on the measurements.



$$\begin{aligned}
 P_{r,IF}^S &= \rho_r^S + c(dt_r - dt^S) + Tropo_R^S(t) + e_{IF}, \\
 L_{r,IF}^S &= \rho_r^S + c(dt_r - dt^S) + Tropo_R^S(t) + \lambda_{IF} \Pi_{R,IF}^S + \epsilon_{IF},
 \end{aligned}
 \tag{2-6}$$

**where:**

$P_{r,IF}^S$ : is the ionosphere-free combination  $(f_A^2 P_A - f_B^2 P_B)/(f_A^2 - f_B^2)$  of pseudoranges  $P_A$  and  $P_B$  observed at two distinct signal frequencies  $f_A$  and  $f_B$ .

$L_{r,IF}^S$ : is the ionosphere-free combination  $(f_A^2 L_A - f_B^2 L_B)/(f_A^2 - f_B^2)$  of the corresponding carrier-phases  $L_A$  and  $L_B$ .

$dt^S$ : is the satellite clock offset  $C$  is the vacuum speed of light.

$Tropo_R^S(t)$ : is the Tropospheric effect

$\Pi_{R,IF}^S$ : is the IF combination of the initial phase of the receiver and satellite  $\varphi_{0R}$  and  $\varphi_0^S$  with the integer ambiguity.

$\lambda_{IF}$ : is the IF combination of the carrier-phase wavelengths  $\lambda_A$  and  $\lambda_B$  of signals A and B (e.g., 10.7 cm for GPS L1 and L2).

$e_{IF}, \epsilon_{IF}$ : are the relevant measurement noise components, including multipath of the IF pseudo-range and carrier-phase combinations.

Satellite orbit, as well as clock parameters, are derived from precise ephemerides and considered known. Consequently, the primary unknown parameters in PPP include the receiver position, the clock offset of the receiver, the zenith total delay (ZTD), and the carrier phase ambiguity  $\Pi_{R,IF}^S$  (Montenbruck & Teunissen, 2017).

## 2.2 Atmosphere

The propagation of a GNSS satellite signal from the satellite to the user receiver involves traversing multiple layers of the Earth's atmosphere, which influences the signal characteristics. During its journey through the atmosphere, the signal encounters delays, damping, and refraction as a result of the atmospheric properties, including air density, temperature, and humidity. These effects arise from the interaction between electromagnetic waves and the molecules and particles present in the atmosphere. The aforementioned delays, damping, and refraction have the potential to introduce errors in the determination of the GNSS position solution (Teunissen & Montenbruck, 2017).

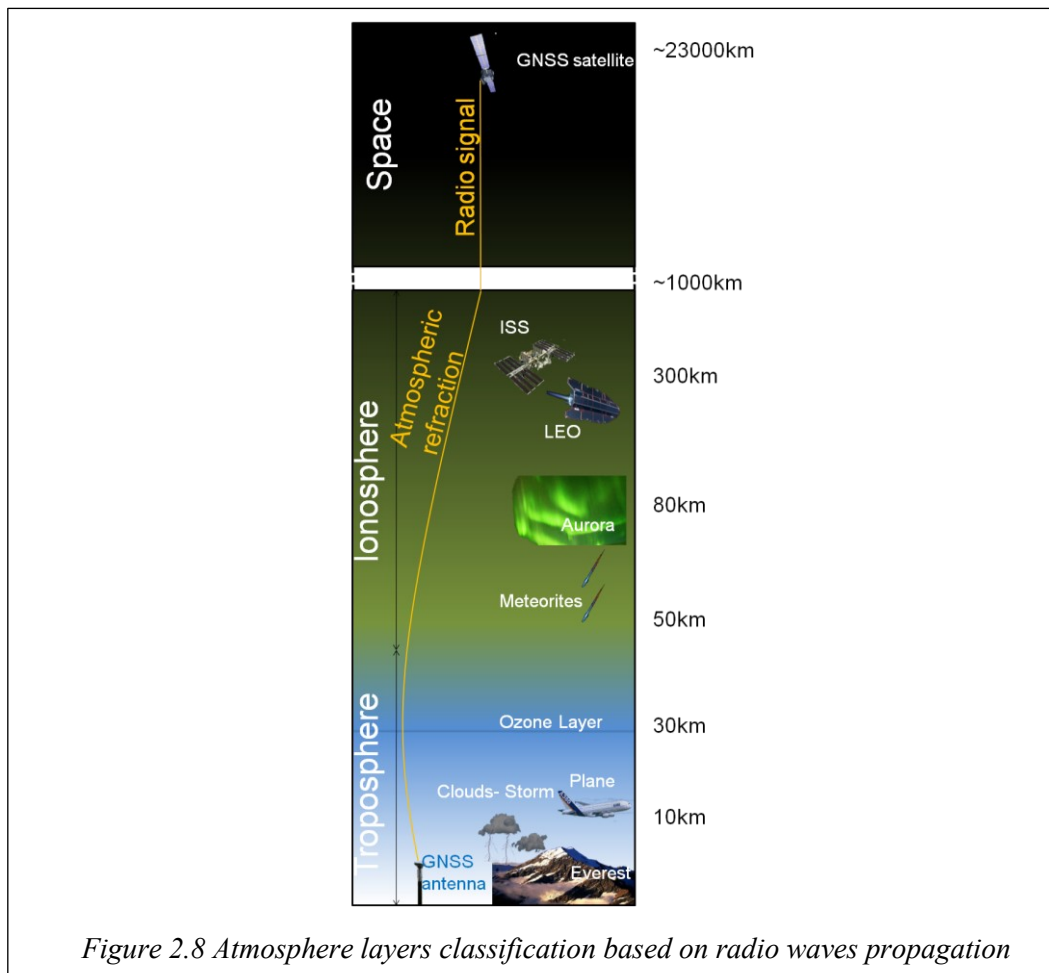


Figure 2.8 Atmosphere layers classification based on radio waves propagation

### 2.2.1 Atmospheric Layers

The Earth's atmosphere can be divided into different layers based on various characteristics, such as composition, temperature, and ionization. For radio wave propagation, the two main layers of interest are the troposphere and the ionosphere (Seeber, 2008) Figure [2.8].

The ionosphere constitutes the uppermost region of the Earth's atmosphere, spanning an altitude range of approximately 50 to 2000 km above the Earth's surface. Within this layer, solar radiation impinges upon the atmosphere at a power density of 1370 W/m<sup>2</sup>, known as the solar constant. During solar flares or similar events, the incident radiation can engage in interactions with atmospheric gas atoms or molecules, resulting in ionization by liberating free electrons and positively charged ions. Concurrently, a process of recombination can take place, where a free electron encounters a positively charged ion and becomes captured by it. This recombination mechanism becomes increasingly prominent at lower altitudes as the density of gas molecules and ions rises, causing them to become more densely packed. The degree of ionization in the ionosphere is governed by the equilibrium between the processes of ionization and recombination, which can vary temporally (Böhm & Schuh, 2013).

The troposphere refers to the lowest layer of the Earth's atmosphere, encompassing approximately 80% of the total atmospheric mass and containing about 99% of the water vapor and aerosols present. Understanding the behavior of these components is crucial for comprehending the effects on GNSS signal propagation, including delays, attenuation, and scintillation. Although the troposphere primarily contributes to signal delay, certain gases situated above it, particularly within the tropopause and stratosphere, also contribute to this delay. The troposphere typically extends up to 20 km in height above sea level, but in most regions, its reach is limited to around 10 km (Teunissen & Montenbruck, 2017).

The influence of water vapor and other atmospheric constituents on the propagation of GNSS signals is subject to variability. Water vapor pressure profiles exhibit substantial disparities from dry pressure profiles and demonstrate a spatial and temporal distribution that is highly erratic. Consequently, the complex behavior of water vapor poses challenges for climatological models in effectively elucidating phenomena associated with this constituent (Teunissen & Montenbruck, 2017).

### 2.2.2 Atmospheric Effects

Based on the atmospheric conditions, the signal propagation experienced a fluctuation in speed. Ionospheric and tropospheric layers will be the primary ones that we consider. The travel time of the signal from the satellite to the receiver ( $\tau_R^S$ ) can be expressed as:

$$\tau_R^S = \int_{\rho_R^S} \frac{dx}{v(x)} \quad 2-7$$

where:

dx: infinitesimal step along the path from the satellite to the user receiver.

v(x): the actual speed along the path.

By substituting in equation 2-2, the pseudo-range will therefore be:

$$\rho_R^S = \int_{\rho_R^S} \frac{cdx}{v(x)} = \int_{\rho_R^S} n(x)dx = \int_{\rho_R^S} dx + \int_{\rho_R^S} (n(x) - 1)dx \quad 2-8$$

$$\int_{\rho_R^S} dx + \int_{\rho_R^S} (n(x) - 1)dx = \rho_R^S + \Delta_R^S \quad 2-9$$

where:

n(x): the Refraction index

$\Delta_R^S$ : the atmospheric delay in unit length

### 2.2.2.1 Ionospheric Effect

The ionosphere is a dispersive medium at which the refraction index is a function of signal frequency (Biagi, 2022). We have two refraction indexes: one for the binary codes ( $\Delta_{gr}^{Iono}$ ) and one for the carrier signal ( $\Delta_{ph}^{Iono}$ ) (Hofmann-Wellenhof et al., 2008).

$$\Delta_{gr}^{Iono} = \frac{40.3 \text{ TEC}}{f^2}, \quad \Delta_{ph}^{Iono} = \frac{-40.3 \text{ TEC}}{f^2} \quad 2-10$$

where:

$\Delta_{gr}^{Iono}$ : the refraction index for PRN codes.

$\Delta_{ph}^{Iono}$ : the refraction index for carriers.

TEC: the total electron content.

f: the frequency of the signal.

Because the ionospheric effect is frequency dependent, utilizing dual frequency receivers for range measurements can remove this effect by properly combining observations made on the two carriers (Kaplan & Hegarty, 2006).

### 2.2.2.2 Tropospheric Effect

Most of the mass of the atmosphere is located in the tropospheric effective height, which extends from the earth's surface up to 40 km. Above 40 km, the density of the higher regions is too small to affect the measurements (Leick, 2004). The troposphere, unlike the ionosphere, is a non-dispersive medium. This means that both PRN codes and GPS carriers are delayed in terms of free-space propagation, not on their frequencies. The tropospheric effect is determined by the tropospheric refractive index ( $N^{tropo}$ ). This index is a function of humidity, pressure, and temperature.

This index can be separated into a dry and a wet component ( $N_d^{Tropo}, N_w^{Tropo}$ ). The dry component or the hydrostatic component is due to the dry gasses such as N<sub>2</sub>, O<sub>2</sub>, CO<sub>2</sub>, Ar, Ne, He, Kr, Xe, CH<sub>4</sub>, H<sub>2</sub>, and N<sub>2</sub>O causing a delay of up to 90% of the tropospheric effect. On the other hand, the wet component is due to the water vapor and it is more difficult to be predicted (Kaplan & Hegarty, 2006; Leick, 2004).

$$N^{Tropo} = N_d^{Tropo} + N_w^{Tropo} \quad 2-11$$

$$N_d^{Tropo} = \overline{K}_1 \frac{Pd}{T}, \quad N_w^{Tropo} = \overline{K}_2 \frac{Pwv}{T} + \overline{K}_3 \frac{Pwv}{T^2} \quad 2-12$$

where:

$\overline{K}_1$ : is a constant equal to  $77.64 \text{ K mb}^{-1}$ .

$\overline{K}_2$ : is a constant equal to  $-12.96 \text{ K mb}^{-1}$ .

$\overline{K}_3$ : is a constant equal to  $3.718 \cdot 10^5 \text{ K}^2 \text{ mb}^{-1}$ .

$Pd$ : is the partial pressure of dry air in mb.

$Pwv$ : is the partial pressure of water vapor in mb.

$T$ : is temprature in K.

From the above equations, we can derive an expression for the tropospheric effect  $\Delta^{Tropo}$

$\begin{aligned}\Delta^{Tropo} &= \Delta_d^{Tropo} + \Delta_w^{Tropo} \\ &= 10^{-6} \int N_d^{Tropo} dx + 10^{-6} \int N_w^{Tropo} dx\end{aligned}$	2-13
---	------

The integration along the zenith direction, as described by equation (2-16), yields the *zenith total delay (ZTD)*, which is the combined effect of the *zenith hydrostatic delay (ZHD)* and the *zenith wet delay (ZWD)*:

$ZHD = 10^{-6} \int N_d^{Tropo}(h) dh$	2-14
$ZWD = 10^{-6} \int N_w^{Tropo}(h) dh$	2-15

Many models had been developed to simulate this effect. The "Saastamoninen model" is one of these. By providing approximations for receiver temperature ( $T$ ), pressure ( $Pd$ ), and humidity ( $Pwv$ ) that can be calculated using Bery formulas. Furthermore, by providing the model with the satellite's position, the approximate position of the receiver, and the satellite elevation  $\eta_R^S$ , the model can estimate the tropospheric effect with an accuracy of about 5:10 % (Biagi, 2022).



### 3. Toward Turbulence stochastic modeling

In this chapter, our objective is to outline the methodology employed to generate turbulence maps based on zenith total delay (ZTD) measurements. To initiate this process, we will introduce fundamental concepts that characterize turbulence as a random phenomenon. Central to this understanding is the definitions of random variables and random functions, which enable the description and analysis of the phenomena of interest. We will also establish key parameters such as mean, variance, and covariance, which provide crucial statistical measures for one and two-dimensional random variables.

Subsequently, we will delve into important concepts and assumptions, such as stationarity and isotropy, which contribute to the foundation of our analysis. Furthermore, we will explore the variogram as a valuable tool for quantifying the variability exhibited by observation stations within our study area. By navigating through the variogram cloud, we will eventually arrive at the variogram function, which encapsulates the essential characteristics of spatial dependence. Finally, we will introduce the kriging interpolation method as a powerful technique employed in the production of turbulence maps. This method allows for the spatial prediction and estimation of turbulence based on the available data. Through the comprehensive examination of these concepts and methods, we aim to provide a robust framework for the generation of turbulence maps in our research endeavor.

#### 3.1 Random variable and Random function

The data samples exhibit temporal and/or spatial dependence and are characterized by multiple variables (multivariate). These samples are obtained at specific points or intervals, span across various epochs, and capture observations of a particular property under study. If we put the time aside, we will have  $n$  observations at each location  $x_\alpha$  that can be noted as follows:

$z(x_\alpha) , where \alpha = 1, \dots, n$	3-1
--	-----

The observation points within the domain  $\mathcal{D}$  constitute a subset of a larger point set. However, due to limitations in resources and cost constraints, it is not feasible to collect an infinite number of observations from this larger set. Nevertheless, by considering the hypothetical scenario of unlimited observations at each point, we define the regional variable  $R_E V$  as we can see in equation(3-2). At each observation location within the domain, a recorded value, known as the *regionalized value*  $z(x_\alpha)$ , represents the outcome of a random mechanism or phenomenon that lies beyond our control. This random mechanism is formally represented as a *random variable*  $Z(x_\alpha)$  (Wackernagel, 2003).

$Z(x) , for all x \in \mathcal{D}$	3-2
------------------------------------	-----

The values observed at all points within the domain originate from an infinite collection of distinct random variables. Each point is associated with a specific random variable, collectively forming a *random function* known as  $R_A F$ . This concept acknowledges the potential variability and uniqueness of each point's observations within the larger set. However, due to practical constraints, we can only observe a limited number of values at the subset of points within the domain (Wackernagel, 2003).

$\{z(x), x \in \mathcal{D}\}$	3-3
-------------------------------	-----

### 3.2 Distribution and Density function

The (cumulative) distribution function (*CDF*), denoted by  $F(x)$ , provides a measure of the probability  $P(\mathbf{x} \leq x)$  that a random variable  $x$  takes on a value less than or equal to a specific value  $x$  (Papoulis, 1991). This function can be defined as

$F_x(x) = P\{\mathbf{x} \leq x\}, \forall x \in \mathbb{R}$	3-4
---	-----

By making the derivative of the distribution function  $F(x)$ , we can get the Density function  $f(x)$ .

$f(x) = \frac{dF(x)}{dx} \geq 0$	3-5
----------------------------------	-----

The form of the distribution function is contingent upon whether the random variable is discrete or continuous in nature. For discrete random variables, it is feasible to assign a probability within the range of 0 to 1 to each possible value as it is shown in (3-6). In this context, the probability of this discrete random variable is commonly referred to as the probability mass function (*PMF*). The PMF, denoted as  $f(x)$ , represents a collection of positive discrete masses associated with each possible value of the random variable (Papoulis, 1991).

$f(x) = \sum_i [p_i * \delta(x - x_i)] = 1$	3-6
---	-----

On the contrary, when dealing with continuous random variables, assigning probabilities to individual values is not feasible. Instead, we employ a probability density function (*PDF*), denoted as  $f_x(x)$ , which enables us to assign probabilities to intervals rather than specific values. The PDF represents a function that characterizes the distribution of the continuous random variable. To determine the probability associated with a particular interval, we calculate the area under the curve of the PDF within that interval. For instance, if the interval of interest lies between  $x_1$  and  $x_2$ , the probability can be obtained by evaluating the integral of the PDF over that interval. This approach allows us to quantify probabilities in the context of continuous random variables (Venuti, 2023).

$P(x_1 < \mathbf{x} < x_2) = \int_{x_1}^{x_2} f_x(x) dx = 1$	3-7
--	-----

When dealing with two-dimensional random variables  $\mathbf{x}$  and  $\mathbf{y}$ , it is essential to determine the joint probability between them for all possible combinations of points  $(x_i, y_j)$  within the domain  $\mathcal{D}$  in the  $x$ - $y$  plane. Simply relying on the individual distribution functions  $F_x(x)$  and  $F_y(y)$  of each random variable is insufficient to fully ascertain the joint statistics of an event (3-8) (Papoulis, 1991; Venuti, 2023).



$\{\mathbf{x} \leq x\} \cap \{\mathbf{y} \leq y\} = \{\mathbf{x} \leq x, \mathbf{y} \leq y\}$	3-8
---	-----

So we can say that the joint distribution  $F(x,y)$  of the two random variables  $\mathbf{x}$  and  $\mathbf{y}$  is the probability of the event

$\mathbf{x} \leq x, \mathbf{y} \leq y = (x, y) \in D, \forall x \text{ and } y \in \mathbb{R}$	3-9
--	-----

is

$F(x, y) = P\{\mathbf{x} \leq x, \mathbf{y} \leq y\}$	3-10
---	------

As previously stated, the joint density function  $f(x,y)$  can be obtained by taking the second-order derivative of the distribution function (3-10) (Papoulis, 1991).

$f(x, y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} \geq 0$	3-11
--	------

Based on the aforementioned information, the probability of the event occurring within area  $\mathcal{A}$ , considering all potential outcomes from random variables  $\mathbf{x}$  and  $\mathbf{y}$   $\{(x, y) \in D\}$ , can be expressed as follows:

$P\{(x, y) \in D\} = \iint_{\mathcal{A}} f(x, y) dx dy = 1$	3-12
---	------

### 3.3 Mean, variance, and covariance for Random variable

To adequately describe a stochastic process, a thorough knowledge of its probability density function (*PDF*) is generally required. Unfortunately, PDFs may not always be available, necessitating the utilization of alternative indicators for the random variable. These indicators commonly include the mean, variance, and covariance.

The mean of a random process, also known as the expected value, corresponds to the first moment of the random variable. It can be computed by integrating the product of the realizations  $x$  and the probability density function  $f(x)$  (Wackernagel, 2003). For continuous random variables, the mean is expressed as follows:

$E\{x\} = \int_{-\infty}^{\infty} x f(x) dx = \mu_x$	3-13
--	------

On the other hand, for discrete random variables, the mean is computed using:

$E\{x\} = \sum_i x_i p_i = m$	3-14
-------------------------------	------

**where:**  $p_i = P(x=x_i)$  denotes the probability mass function of the discrete random variable.

The second indicator is the variance, denoted by the symbol  $\sigma^2$ , which represents the theoretical variance. The variance is a measure of the dispersion or spread of a random variable's distribution. It quantifies the average squared deviation of the random variable  $x$  from its mean  $\mu_x$ , providing insight into the variability of the data (Papoulis, 1991; Venuti, 2023). It can be defined as

$\sigma^2 = E[(x - \mu_x)^2]$	3-15
-------------------------------	------

For continuous random variables, the variance is expressed as follows:

$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x) dx$	3-16
--	------

For discrete random variables, the variance is expressed as follows:

$\sigma^2 = \sum_i p_i (x_i - m)^2, \text{ where } p_i = (x = x_i)$	3-17
---	------

By calculating the square root of the variance of the random variable  $x$ , we obtain the standard deviation, denoted as  $\sigma_x$ . The standard deviation represents a measure of the dispersion or variability of the data, and it is obtained by taking the square root of the variance (Papoulis, 1991). By having the mean  $\mu_x$  and the standard deviation  $\sigma_x$  we can construct a random variable interval by:

$(\mu_x - \alpha\sigma_x < x < \mu_x + \alpha\sigma_x), \text{ where } \alpha = 1, \dots, n$	3-18
--	------

also, we can say that the probability value can be calculated by (Venuti, 2023):

$\text{The probability values} > 1 - \frac{1}{\alpha^2}$	3-19
--	------

In addition to the mean and variance indicators, which provide insights into individual random variables, there is a need for a measure that characterizes the relationship between two random variables,  $x$ , and  $y$ . This led to the introduction of the *covariance*, denoted as  $C_{xy}$ . The covariance quantifies the extent to which changes in one random variable are associated with changes in another (DeGroot, 1986). It is defined as follows:

$C_{xy} = E[(x - \mu_x)(y - \mu_y)]$	3-20
--------------------------------------	------

**where:**  $\mu_x$  and  $\mu_y$  are the mean for random variables  $x$  and  $y$  respectively.

$C_{xy} = E_{xy} \begin{bmatrix} (x - \mu_x)^2 & (x - \mu_x)(y - \mu_y) \\ (y - \mu_y)(x - \mu_x) & (y - \mu_y)^2 \end{bmatrix}$	3-21
--	------

From Equation (3-21), it can be observed that the diagonal elements correspond to the covariance between a random variable and itself, which is equivalent to the variance in Equation (3-22). The covariance between two random variables can be denoted by  $\sigma_{xy}$ .

$C_{xy} = E_{xy} \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix}$	3-22
--	------

The correlation between two random variables, denoted as  $\rho_{xy}$ , can be expressed by dividing the covariance of the variables by their respective standard deviations, as indicated in Equation (3-23). This mathematical relationship provides a measure of the linear association between the two random variables.

$\rho_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y}$	3-23
--	------

**where:**  $\sigma_x$  and  $\sigma_y$  are the standard deviation for the two random variables.

### 3.4 Stationarity and Isotropic Assumptions

If we consider a domain with  $n$  distinct points  $(x_1, \dots, x_n)$ , we can associate  $n$  random variables at each point  $(Z(x_1), \dots, Z(x_n))$ . In this case, the distribution function can be mathematically defined as follows:

$F_{x_1, \dots, x_n}(z_1, \dots, z_n) = P\{\mathbf{Z}(x_1) \leq z_1, \dots, \mathbf{Z}(x_n) \leq z_n\}, \forall z \in \mathbb{R}$	3-24
---	------

In practical applications, the acquisition of comprehensive data on the mono and multivariate distribution functions for all points within a domain is often limited. As a result, it becomes unfeasible to determine the complete set of distribution functions for any given set of points based on the available data from one or several realizations of the  $R_A F$ . Thus, simplification becomes necessary. This simplification is through the introduction of the concept of stationarity (Wackernagel, 2003).

Stationarity in a random process implies that the  $R_A F$  properties remain unchanged regardless of shifting the  $n$  observed points to different locations within the region. This property, also known as *translation invariance*, signifies that the statistical characteristics of the process, such as mean, variance, and covariance, exhibit consistency and do not depend on the specific position of the observed points (Wackernagel, 2003).

From a physical perspective, the observed phenomenon is postulated to exhibit spatial homogeneity, displaying consistent characteristics throughout the entire space and repetitively repeating its behavior across the entire spatial domain (Chiles & Delfiner, 2012).

A random process  $R_A F$ , denoted as  $Z(x)$ , is said to be strictly stationary if, for any  $n$ -set of points  $(x_1, x_2, \dots, x_n)$  and any translation vector  $h$ , the joint distribution function  $F_{x_1, \dots, x_n}(z_1, \dots, z_n)$  is equal to the joint distribution function  $F_{x_1+h, \dots, x_n+h}(z_1, \dots, z_n)$ . Mathematically, this can be represented as:

$F_{x_1, \dots, x_n}(z_1, \dots, z_n) = F_{x_1+h, \dots, x_n+h}(z_1, \dots, z_n), \text{ where } h=(x+h)-x$	3-25
---	------

Strict Stationarity requires specifying a multipoint distribution, which can be complex. To simplify the analysis, an alternative approach considers only pairs of points in the domain, focusing on characterizing the first two moments instead of the entire distribution. Two possibilities arise from this approach. The first is *second-order stationarity*, assuming that the first two moments of the variable remain stationary over the domain. The second possibility is *intrinsic stationarity*, assuming the stationarity of the two moments of the difference between pairs of values at two points. Intrinsic stationarity leads to the concept of a variogram (Wackernagel, 2003).

According to (Chiles & Delfiner, 2012; Cressie, 1993), we can mathematically define second-order stationarity as

$E[Z(x)] = \mu, \quad \forall x \in D$	3-26
--	------

$E[(Z(x) - \mu)(Z(x+h) - \mu)] = C(Z(x) - Z(x+h)) = C(h)$	3-27
---	------

From the given equations, we can conclude that a random function exhibits second-order stationarity when the mean  $\mu$  remains constant for all points and the covariance depends solely on the separation distance  $h$ .

According to the intrinsic stationarity hypothesis, the expected mean of the difference between two random variables  $Z(x)$  and  $Z(x+h)$  known as *increments* can be denoted as

$E[Z(x) - Z(x+h)] = \mu(h) = 0, \quad \forall x, x+h \in D$	3-28
---	------

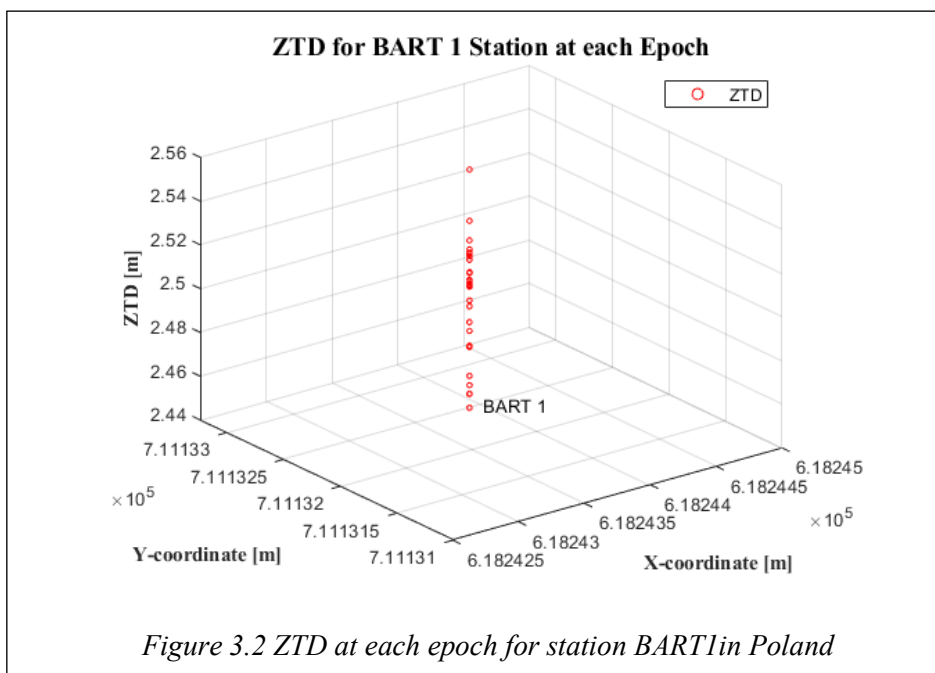
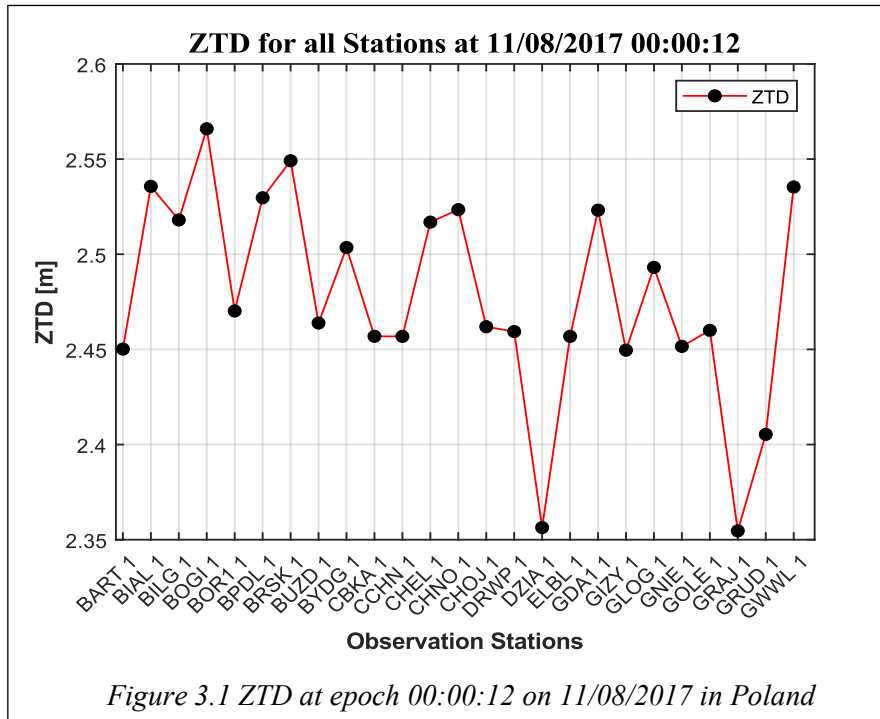
Furthermore, instead of utilizing the covariance, the variogram, which represents the variance of the difference between the two random variables, is employed (Webster & Oliver, 2007). This variogram is defined as

$var[Z(x) - Z(x+h)] = E[\{Z(x) - Z(x+h)\}^2] = 2\gamma(h)$	3-29
--	------

The second assumption we need to consider is *isotropy*, which indicates that a random process maintains its statistical properties irrespective of the direction. In the case of covariance, isotropy implies that the covariance between pairs of random variables depends solely on the distance separating them  $C(\|Z(x) - Z(x+h)\|)$ , while being independent of the particular direction or orientation between the points (Chiles & Delfiner, 2012).

### 3.5 Turbulence delay as a Random process

In our research, we analyze a dataset consisting of observations collected at multiple epochs (denoted as  $t$ ) from a network of  $n$  observation stations. Each station provides a single value at each epoch, as in Figure [3.1]. These observations pertain to the total zenith delay ( $ZTD$ ), which represents the delay experienced by GNSS signals as they propagate through the atmosphere in the zenith direction above each station. The  $ZTD$  is primarily influenced by the presence and distribution of water vapor in the atmosphere above the station. Consequently, different stations exhibit distinct  $ZTD$  values due to the spatial heterogeneity of the water vapor field. Furthermore, even within the same station, the  $ZTD$  values can vary across different epochs, reflecting the temporal variability of water vapor content, as shown in Figure [3.2]. This observation suggests that the  $ZTD$  function is influenced by both spatial (position) and temporal (time) factors, where  $ZTD((E, N), t)$  represents a two-dimensional random variable with  $E$  and  $N$  denoting the station coordinates in our geographical domain.



In the context of our analysis, we treat the ZTD model as a random variable, where the observed ZTD(t) value at a given epoch represents one potential outcome of a stochastic process occurring at a particular station. Considering the network of stations, we encounter distinct random variables for each station, resulting in  $N$ -dimensional random variables due to the  $N$  observed values. However, to simplify the analysis, we address this challenge by assuming stationarity.

This assumption stems from the notion that, when examining closely located stations, the water vapor field tends to exhibit homogeneity, with variations occurring only under specific circumstances. By assuming stationarity, we reduce the  $N$ -dimensional random variables to pairs of points, yielding *two-dimensional random variables*. Under the *stationarity assumption*, pairs of points with equal distances possess the same joint distribution function. Additionally, we assume *isotropy*, implying that not only the relative positions but also the orientations of the pairs remain fixed.

To estimate the spatial characteristics of the ZTD field, we utilize samples obtained from the two-dimensional random variables, representing pairs of points. These samples serve as empirical observations for estimating the variogram, which captures the spatial variability of ZTD values as a function of distance. It is noteworthy that our estimation of the variogram incorporates not only the assumption of stationarity but also *intrinsic stationarity*, accounting for the fixed orientations of the pairs.

By employing these statistical assumptions and leveraging variogram estimation, we aim to gain insights into the spatial dependence and variability of the ZTD field. Ultimately, this approach facilitates the production of turbulence maps based on the observed data, enhancing our understanding of the water vapor distribution and its impact on atmospheric phenomena.

### 3.6 Variogram

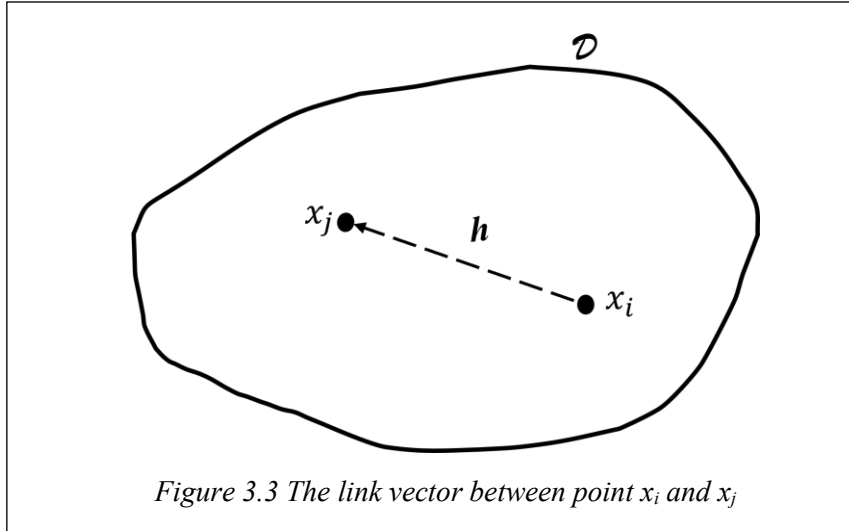
The variogram holds significant importance as a geostatistical tool within the realm of spatial data analysis. It serves as a fundamental component for constructing the groundwork necessary for interpolation through the Kriging method, which will be further explored in subsequent discussions. The fundamental principle underlying the variogram stems from the assumption of intrinsic stationarity, wherein the spatial relationship between pairs of samples is contingent upon their relative positions. This assumption holds within the spatial domain  $\mathcal{D}$ , wherein observations manifest as realizations of random variables  $Z(x_i)$ , generated from a random function  $\mathbf{Z}=Z(x)$ ,  $x \in D$ . Consequently, we refer to this measure as the *variogram function* (Lichtenstern, 2013; Wackernagel, 2003).

To derive the variogram function within the framework of the Kriging method, we will follow the following steps utilizing the observed data:

- 1- Variogram Cloud Estimation: Initially, we estimate what is referred to as a *variogram cloud*, which captures the relationship between the spatial separation of pairs of observed points and the squared difference in their corresponding values. This cloud represents the *empirical variogram*, exhibiting the variability in the data as a function of distance.
- 2- Cloud Segmentation: The variogram cloud is subsequently divided into classes based on a predefined spatial separation criterion. Each class represents a distinct range of spatial distances. Within each class, we calculate the average of the squared differences in values for the corresponding sample pairs. This average value is referred to as the *experimental variogram* for that class.
- 3- Variogram Model Fitting: Finally, we aim to fit the experimental variogram with an appropriate variogram model function. This process is also referred to as *theoretical variogram* fitting. This entails adjusting the parameters of the model to minimize the differences between the experimental variogram and the model-predicted values. The fitting process typically employs a least squares adjustment approach, optimizing the model parameters to achieve the best possible fit.

### 3.6.1 Variogram Cloud

The measurement of variability for a regionalized variable,  $z(x)$ , involves the estimation of dissimilarity  $\gamma^*$  between the values  $z_i$  and  $z_j$  at two pairs of points  $x_i$  and  $x_j$ , situated within our spatial domain. These points are connected by the vector  $h$ , also referred to as the lag, as depicted in Figure [3.3]. The dissimilarity is characterized by its symmetric nature with respect to the lag distance  $h$ , as it is a square quantity. Consequently, the order in which the points are considered does not affect the calculation of dissimilarity [Equation 3-30] (Wackernagel, 2003).



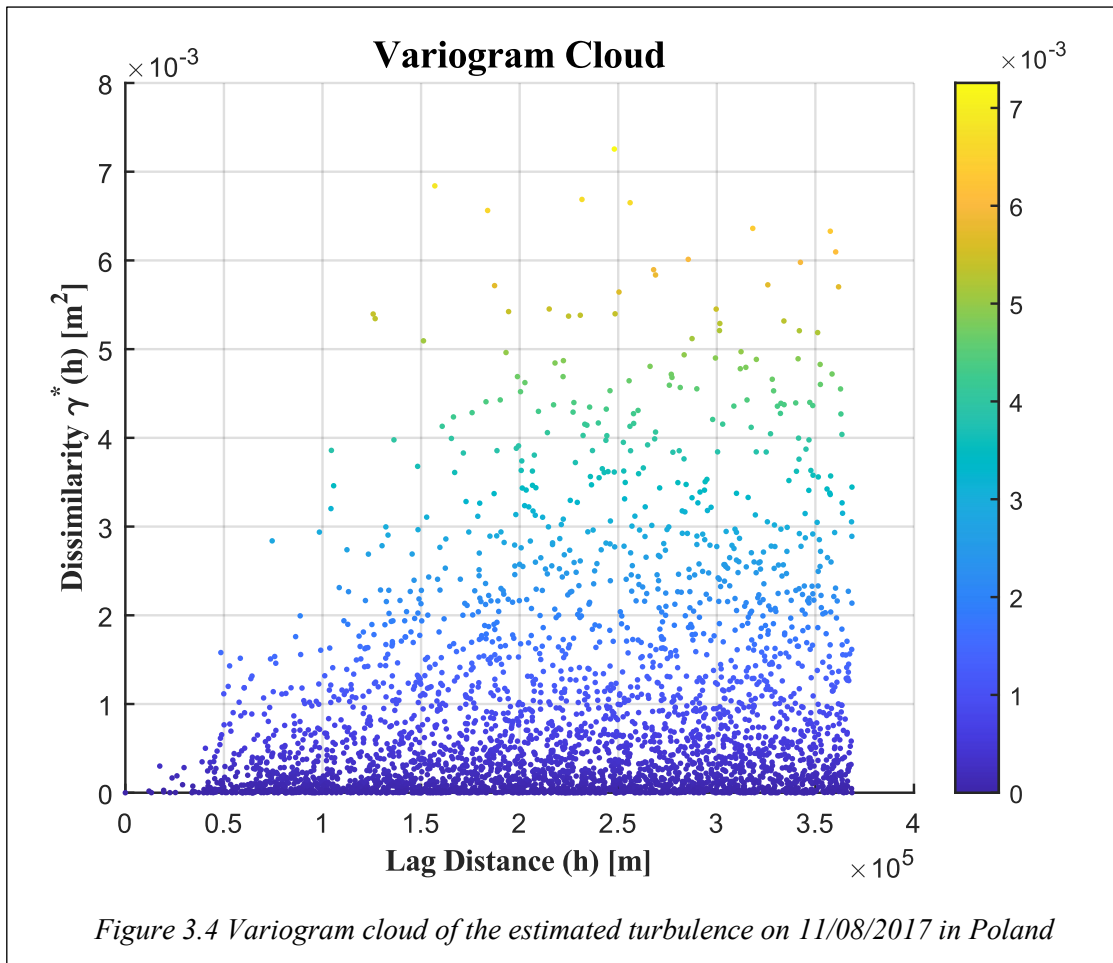
$\gamma^*(-h) = \gamma^*(+h)$	3-30
-------------------------------	------

The dissimilarity between the two values can be measured by

$\gamma_{ij}^* = \frac{1}{2}(z_i - z_j)^2$	3-31
--	------

To express the dissimilarity function in terms of the lag distance  $h$ , the following transformation can be employed:

$\gamma^*(h) = \frac{1}{2}(z_{i+h} - z_i)^2$	3-32
--	------

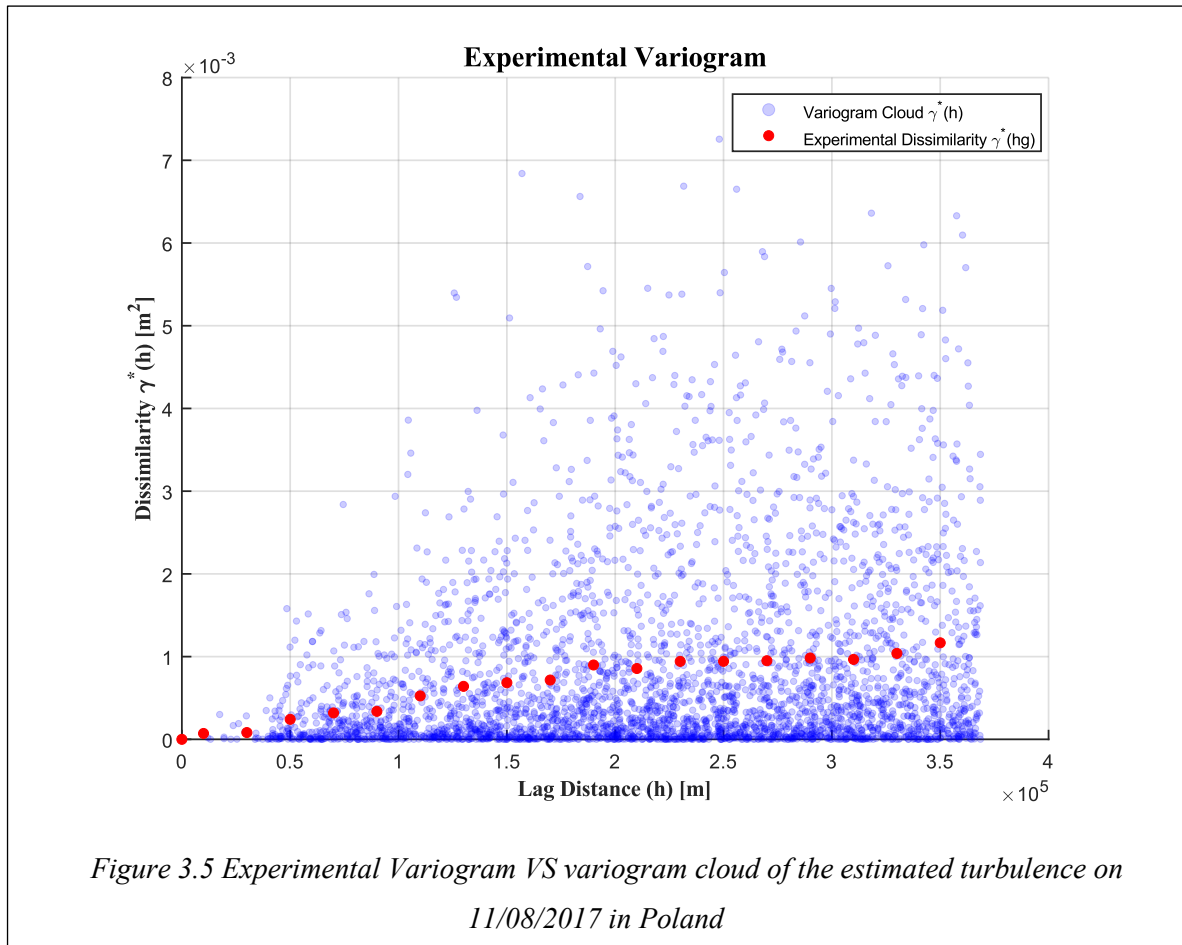


By considering all pairs in our dataset and calculating their dissimilarities with respect to the corresponding lag distances,  $h$ , we can construct a graphical representation known as a *variogram cloud* Figure [3.4]. This cloud consists of  $(n(n-1))/2$  points, each representing the dissimilarity between pairs of values (Kitanidis, 1997). The range of lag distances spans up to half the diameter of the region under study (Wackernagel, 2003).

We can notice from the variogram cloud that the dissimilarity increases by increasing the lag distance. As we mentioned that a variogram cloud is a powerful tool, it allows us to detect Anomalies, and inhomogeneities by looking for high dissimilarity at short distances. Also in all lag distances, we can find many pairs with low dissimilarities (Wackernagel, 2003).



### 3.6.2 Experimental variogram



Upon generating the variogram cloud, the separation vectors  $h$  can be categorized into distinct classes of vectors, denoted as  $\beta$ , based on their lengths falling within specific intervals. The selection of the interval for forming the class of vectors  $\beta$  is contingent upon the specific application at hand (Cressie, 1993). Each vector  $h$  belonging to  $\beta$  is associated with a corresponding group of point pairs, referred to as  $n_c$ . Within each class  $\beta$ , the average dissimilarity,  $\gamma^*(\beta_k)$ , also known as *the classic estimator*, or *the method of moment estimator* (Wackernagel, 2003), can be computed as

$\gamma^*(\beta_k) = \frac{1}{2n_c} \sum_{i=1}^{n_c} (z_{i+h} - z_i)^2, h \in \beta_k \ \& \ k \in \mathbb{N}$	3-33
--	------

The combined implementation of the classic estimator and vector classes  $\beta$  successfully tackles two significant challenges. Firstly, the estimation of the average dissimilarity, represented by  $\gamma^*(\beta_k)$  effectively resolves the problem of encountering multiple values for distance lags,  $h$ . This estimator ensures the preservation of the symmetry property, as stated in Equation 3-30. It is worth noting that while the estimator remains unbiased for estimating the mean  $\mu$  of  $Z(x)$ , it does exhibit sensitivity to outliers due to the adoption of the square difference (Cressie, 1993; Webster & Oliver, 2007).

Secondly, the integration of vector classes  $\beta$  effectively addresses the issue of incomplete data for most lags. In scenarios where observations are unavailable, leading to a lack of dissimilarity values,  $\gamma^*(h)$ , this approach assigns the average dissimilarity of a specific class,  $\beta_k$ , to all lags  $h$  within that class (Lichtenstern, 2013).

Figure [3.5] presents a graphical representation of the experimental variogram across the variogram cloud, where the class width for each category is fixed at 20 km. It is observed that the average dissimilarity increases as the spacing between pairs of points increases. The experimental variogram eventually reaches a plateau, known as the *sill*, which corresponds to the variance of the dataset for large spacings. The behavior of dissimilarity at the origin is of significant importance as it reflects the continuity characteristics of the regionalized variable, indicating whether it is differentiable, not differentiable, or discontinuous. A discontinuity in the variogram at the origin signifies the presence of a *nugget effect*, meaning that at very small scales, the variable values exhibit abrupt changes (Wackernagel, 2003).

### 3.6.3 Variogram Model Fitting

Our objective is to achieve a continuous representation of regional variation, ensuring that the variogram function possesses physical significance. Additionally, we aim to describe the spatial variation in a manner that allows for estimating values in unsampled locations using kriging. Upon examining the experimental variogram, we observed limitations in terms of continuity and the ability to estimate variation across all lag distances. Furthermore, using the experimental variogram values directly in the Kriging system would result in negative variance. Consequently, we have chosen to fit the experimental variogram with a function referred to as *the theoretical variogram*, which addresses these issues (Wackernagel, 2003; Webster & Oliver, 2007).

The theoretical variogram, denoted as  $\gamma(h)$ , is determined by the intrinsic hypothesis, and its conditions can be applied using Equation [3-28] and Equation [3-29]. The formulation for  $\gamma(h)$  is given by:

$\gamma(h) = \frac{1}{2} (Z(x+h) - Z(x))^2$	3-34
---	------

The theoretical variogram exhibits several key properties that are crucial for its understanding and interpretation in geostatistics. First and foremost, the variogram is defined such that its value at the origin ( $h = 0$ ) is always zero Equation [3-35]. This property ensures that there is no spatial variability at a zero lag distance since it compares a location with itself.

$\gamma(h = 0) = 0$	3-35
---------------------	------

Another important property of the variogram is that its values are always positive for all non-zero lag distances [3-36]. This property reflects the fact that the variogram measures the magnitude of spatial variability, which is always a positive quantity.

$\gamma(h) \geq 0$	3-36
--------------------	------

The variogram is also an even function Equation [3-37]. This property implies that the variogram values for positive lag distances are the same as the values for negative lag distances. In other words, the spatial relationship and variability between two locations are the same regardless of the direction of displacement between them.

$\gamma(-h) = \gamma(h)$	3-37
--------------------------	------

Furthermore, the variogram grows slower than the function  $|h|^2$  as  $h$  increases Equation [3-38]. This means that the increase in the variogram with increasing lag distances is less steep than the  $|h|^2$ . This condition is essential for ensuring that the drift term  $\mu(h)$ , which represents any systematic spatial trend or variation in the variable, can be assumed to be zero (Wackernagel, 2003).

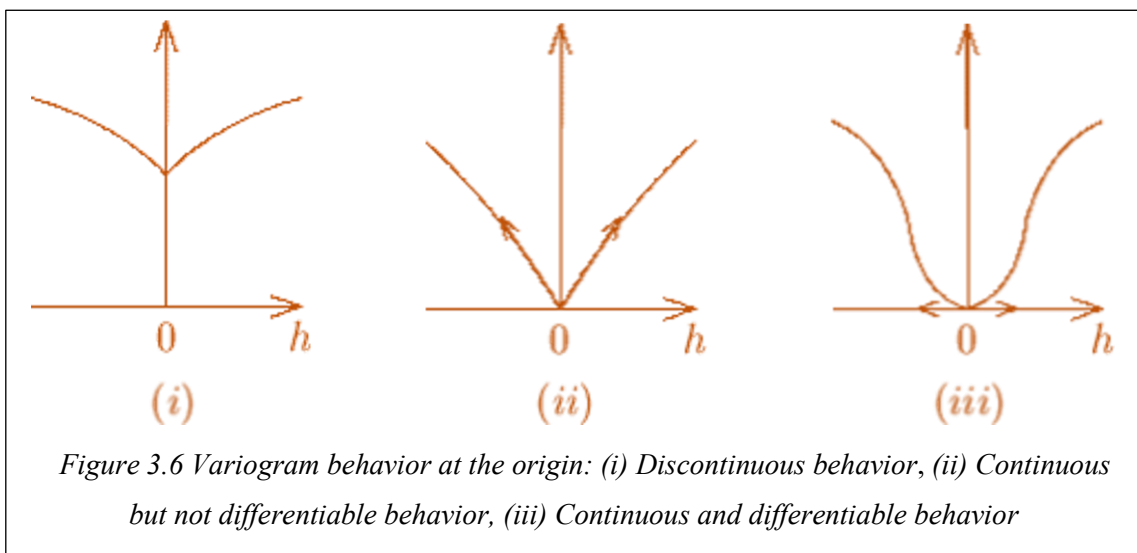
$\lim_{h \rightarrow \infty} \frac{\gamma(h)}{ h ^2} = 0$	3-38
---	------

Lastly, the variogram is a *conditionally negative semidefinite* function (Wackernagel, 2003). This property ensures that the variance of any linear combination of  $n+1$  random variables, which forms a subset of an intrinsic random function, remains positive when the sum of the  $n+1$  weights  $\lambda_\alpha$  is equal to zero. This can be expressed using the following equation:

$\sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_\alpha \lambda_\beta \gamma(x_\alpha - x_\beta) \leq 0, \text{ if } \sum_{\alpha=1}^n \lambda_\alpha = 0$	3-39
---	------

The behavior of the variogram at the origin and large distances holds significant importance when choosing an appropriate variogram model. At the origin, as illustrated in Figure [3.6], the variogram can exhibit distinct characteristics:

1. *Discontinuous behavior*: In this scenario, the variogram does not approach zero as  $h$  tends to zero. This discontinuity, known as the *nugget effect*, adversely affects kriging estimates, resulting in discontinuities in the interpolated surface.
2. *Continuous but not differentiable behavior*: Near the origin, the variogram may demonstrate linear behavior. Although the variogram remains continuous, it lacks differentiability at the origin. As a consequence, the kriging estimator also becomes non-differentiable, and the resulting kriged surface appears linear within the vicinity of the data points.
3. *Continuous and differentiable behavior*: Alternatively, the variogram may display quadratic behavior at the origin, indicating a smooth and continuous function. In this case, the variogram is twice differentiable at the origin, leading to smoother kriging estimates. The resulting kriged surface possesses enhanced smoothness and continuity.

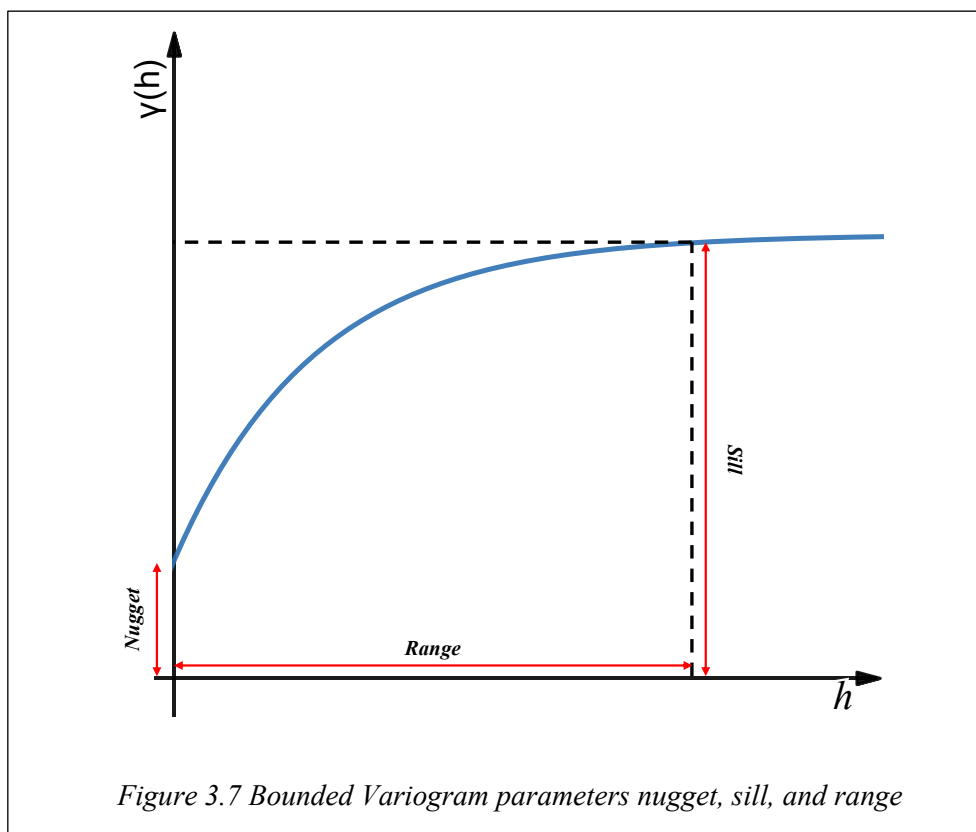


The fitting of the variogram function can be accomplished either visually or numerically using a black box approach. The exact way the variogram function fits the average dissimilarity points is not of utmost importance. What holds significance is the type of continuity assumed for the regional variable and the stationarity hypothesis associated with the random function (Wackernagel, 2003). However, the fitting process poses challenges due to various reasons, as outlined in (Webster & Oliver, 2007):

1. The accuracy of observing the semivariogram is not consistent, which can introduce uncertainties in the fitting process.
2. Instead of assuming isotropic variation, the variation in the regional variable may exhibit anisotropy, posing additional complexities in selecting an appropriate variogram model.
3. The experimental variogram may contain point-to-point fluctuations or noise, making it challenging to identify the underlying spatial structure accurately.
4. Most of the available fitting models involve nonlinearity in one or more parameters, making the fitting process computationally demanding and requiring optimization techniques.

There are two families of functions commonly used to represent spatial variation: unbounded and bounded. In the context of isotropic analysis, the distance between pairs of observations, denoted as  $|h|$ , is used as a scalar measure for lag distance,  $h$  (Webster & Oliver, 2007).

A bounded variogram model refers to a variogram model that has a maximum value beyond which the dissimilarity does not increase. This implies that there is an upper limit to the spatial correlation or dependence. In other words, as the distance between pairs of observations increases, the dissimilarity reaches a plateau or saturation point. This is often the case when there are physical or natural limits to the spatial dependence of a variable. There are a group of bounded models such as *the Spherical Model*, *Gaussian Model*, *Matérn Model*, and *Exponential Model*. In this thesis, our primary focus will be on the exponential variogram model as an example of a bounded model To explore alternative variogram models, valuable insights and detailed explanations can be found in sources such as (Kitanidis, 1997; Webster & Oliver, 2007). These references provide comprehensive analyses and discussions on various variogram models, offering a broader understanding of different modeling approaches.



Bounded variogram models involve three key parameters: *nugget*, *sill*, and *range* Figure [3.7]. The *nugget* ( $c_0$ ) parameter characterizes the discontinuity or variability observed at very short distances Equation [3-40], reflecting fine-scale variation and measurement error that lack spatial correlation. By capturing the randomness and noise present in the data, the nugget effect allows us to account for unexplained fluctuations (Lichtenstern, 2013).

$\gamma(h) \rightarrow c_0 > 0, \text{ as }  h  \rightarrow 0$	3-40
--	------

The *sill* ( $\sigma_Z^2$ ) parameter represents the maximum dissimilarity or variance exhibited by the variable of interest Equation [3-41]. It provides insights into the level of spatial correlation or dependence when the lag distance is sufficiently large. Beyond the range, the sill indicates an upper limit where the dissimilarity ceases to increase, reaching a saturation point in the spatial correlation (Cressie, 1993).

$\gamma(\infty) = \lim_{ h  \rightarrow \infty} \gamma(h)$	3-41
--	------

The *range* ( $a$ ) parameter signifies the lag distance at which the spatial correlation or dependence reaches its maximum value, as defined by the sill. Observations beyond this range exhibit constant or leveled-off dissimilarity. Furthermore, the range provides valuable information about the spatial extent over which observations exhibit significant correlation and offers an indication of the scale of spatial dependence within the dataset.

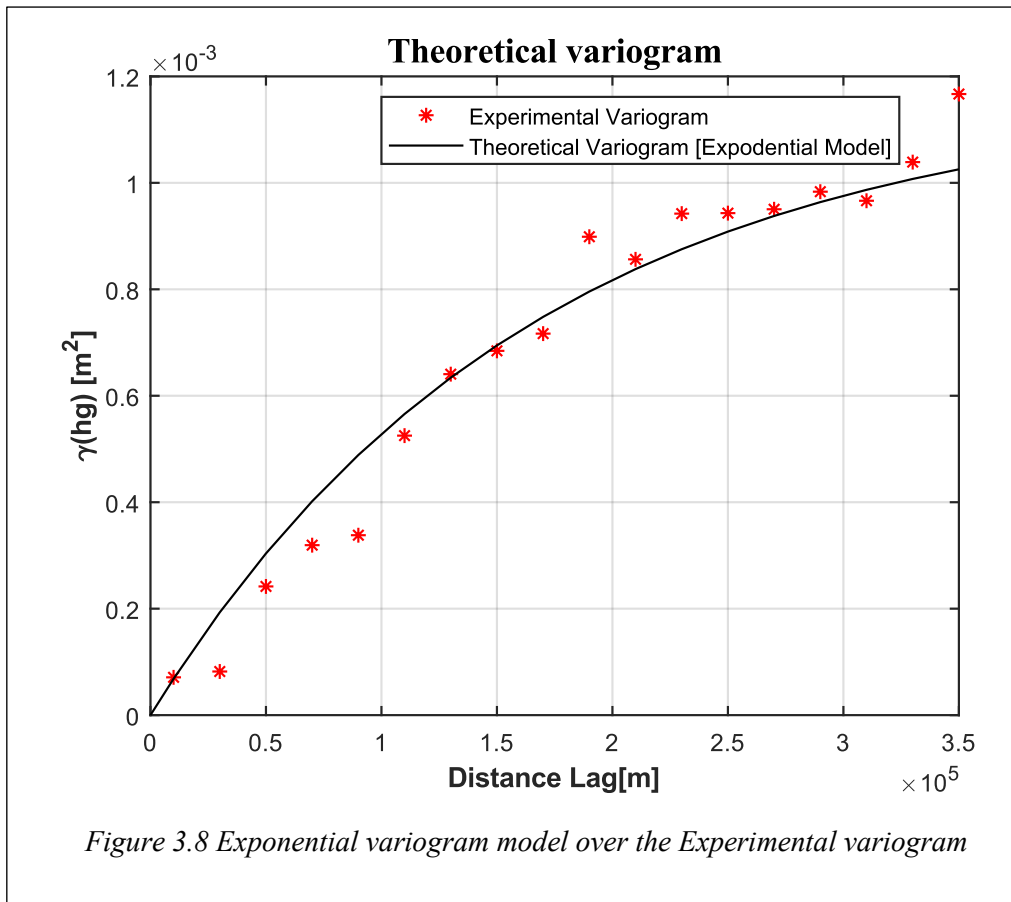
As mentioned earlier, we will now delve into the Exponential model, which can be represented by the equation:

$\gamma_{c,a}^{exp} = c \left( 1 - e^{\left( -\frac{ h }{a} \right)} \right), \quad  h  \geq 0$	3-42
---	------

**where:**

- $c$ : The sill.
- $a$ : The range.

In this model, the range signifies the distance at which the semivariogram reaches approximately 95% of its sill value when  $h = 3a$ . This specific distance is commonly referred to as *the practical or effective range*. In Figure [3.8], we observe that the Exponential model exhibits a continuous behavior but is not differentiable at the origin (Wackernagel, 2003).

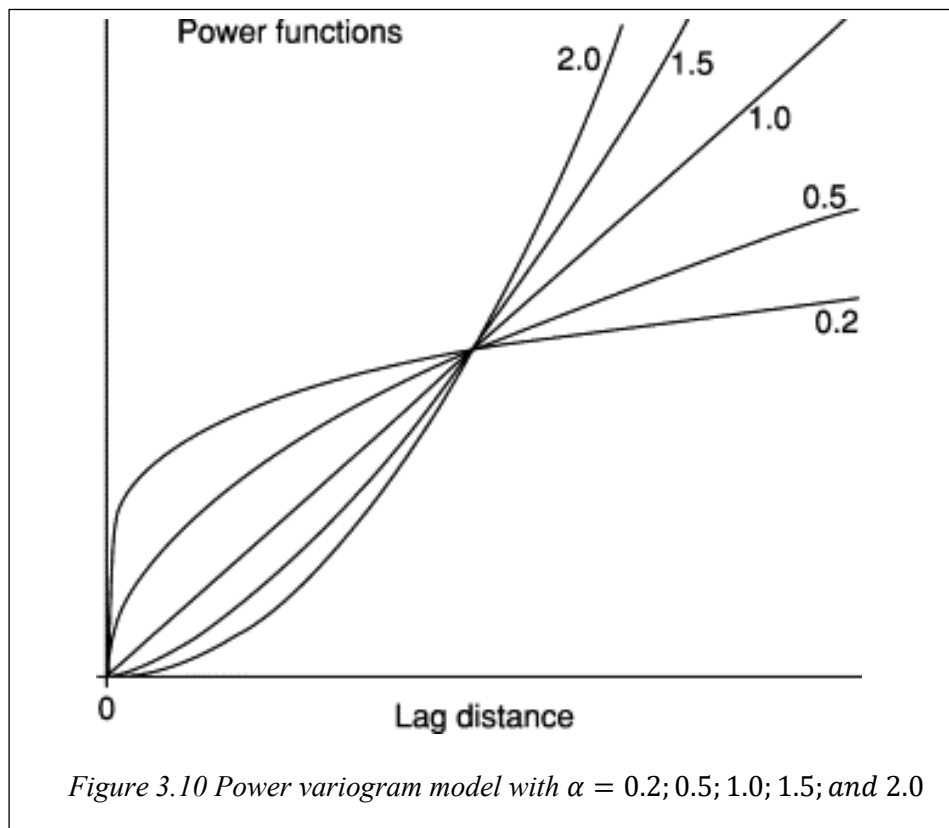
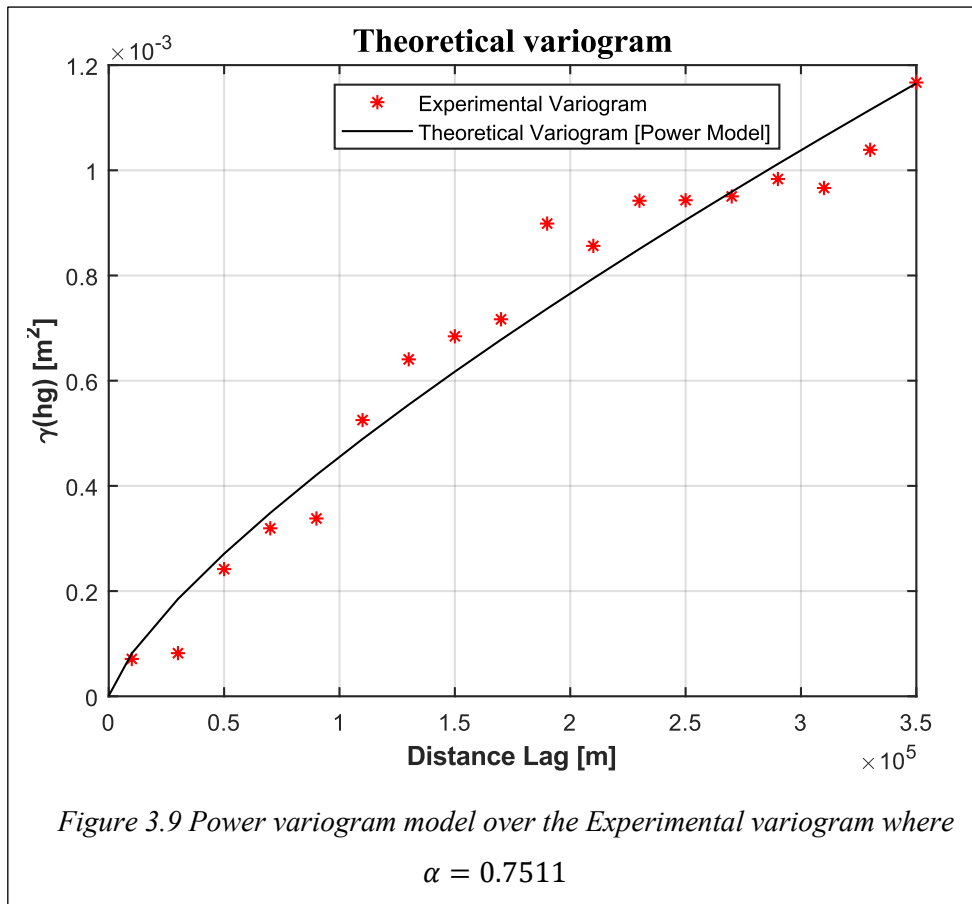


An unbounded variogram model, on the other hand, does not have a maximum value and allows the dissimilarity to increase indefinitely with increasing lag distance. This suggests that there is no inherent upper limit to the spatial correlation or dependence. Unbounded variogram models are often used when the study area is large enough that the spatial correlation does not reach a plateau within the observed range of distances. In this thesis, our specific emphasis will be on the power model, serving as an exemplar of an unbounded variogram model Figure [3.8].

The power model can be mathematically expressed as :

$\gamma(h) = \eta h^\alpha$ for $0 < \alpha < 2$	3-43
--	------

In this model, we have two parameters,  $\eta$  and  $\alpha$ .  $\eta$  describes the intensity of variation and  $\alpha$  describes the curvature. Figure [3.9] illustrates the effect of changing the value of  $\alpha$  on the behavior of the curve. When  $\alpha < 1$ , the variogram curve is convex upward, and when  $\alpha = 1$  then the variogram is linear and  $\eta$  is simply the gradient. When  $\alpha > 1$  then the variogram is concave upwards (Webster & Oliver, 2007).



### 3.7 Ordinary Kriging

Ordinary kriging (OK) is a geostatistical method employed for estimating the value of a random variable,  $Z$ , at a particular location,  $x_0$ . The estimation is based on the utilization of a known variogram function, which characterizes the spatial correlation of the underlying process, as well as a set of neighboring sample points,  $x_i, i=1, \dots, n$ . In ordinary kriging, two fundamental assumptions are made (Wackernagel, 2003).

The first assumption, known as the *model assumption*, posits that the expected value or mean,  $\mu$ , where  $\mu \in \mathbb{R}$  of the second-order stationary random function,  $Z(x)$ , is unknown (Cressie, 1993).

The second assumption, referred to as the *predictor assumption* (Equation [3-44]), establishes that the estimation of  $Z$  at  $x_0, Z_{OK}^*(x_0)$ , is a linear combination of the observed values at the neighboring sample points. The weights,  $\lambda_i$ , assigned to each observation are determined through a set of linear equations, ensuring an unbiased estimate and yielding the best linear unbiased prediction (*BLUP*) of the unknown value at  $x_0$  (Lichtenstern, 2013; Wackernagel, 2003).

$Z_{OK}^*(\mathbf{x}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{x}_i)$	3-44
---	------

To ensure uniform unbiasedness in ordinary kriging, it is necessary for the sum of the weights assigned to the neighboring sample points to equal one Equation [3-45]. This requirement is rooted in the principle that when all the data values are equal to a constant, the estimated value should also be equal to that constant (Wackernagel, 2003).

$Z_{OK}^*(\mathbf{x}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{x}_i), \quad \sum_{i=1}^n \lambda_i = 1$	3-45
---	------

As a consequence of the weights summing up to one in ordinary kriging, the expected error tends to disappear Equation[3-46] (Wackernagel, 2003). In other words, the average difference between the estimated value and the true value diminishes as the number of sample points and the reliability of the estimation procedure increase.

$\begin{aligned} E[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] &= E \left[ \sum_{i=1}^n \lambda_i Z(\mathbf{x}_i) - Z(\mathbf{x}_0) \cdot \underbrace{\sum_{i=1}^n \lambda_i}_1 \right] \\ &= \sum_{i=1}^n \lambda_i E[Z(\mathbf{x}_i) - Z(\mathbf{x}_0)] = 0 \end{aligned}$	3-46
---	------

To quantify the accuracy of our estimation in ordinary kriging, our goal is to minimize the estimation variance. The estimation variance represents the mean square difference between the estimated value  $Z^*(\mathbf{x}_0)$  and the true value  $Z(\mathbf{x}_0)$  at the prediction point, as shown in Equation [3-47].

$\sigma_E^2 := \text{var}(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)) = E \left[ (Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0))^2 \right]$	3-47
--	------

By imposing the unbiased condition as stated in Equation [3-45], where the sum of the weights from 1 to n is equal to one, we can utilize the variogram to estimate the variance of the estimation error (Wackernagel, 2003) as:



$\sigma_E^2 = -\gamma(\mathbf{x}_0 - \mathbf{x}_0) - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) + 2 \sum_{i=1}^n \lambda_i \gamma(\mathbf{x}_i - \mathbf{x}_0)$	3-48
--	------

where:

$\gamma(\mathbf{x}_i - \mathbf{x}_j)$ : The dissimilarity of the random function  $Z$  between observation points  $\mathbf{x}_i$  and  $\mathbf{x}_j$ .  
 $\gamma(\mathbf{x}_i - \mathbf{x}_0)$ : The dissimilarity of the random function  $Z$  between observation points  $\mathbf{x}_i$  and the estimation point  $\mathbf{x}_0$ .

By initiating the process of estimating the unknown weights that minimize the variances in ordinary kriging, we employ the Lagrange multipliers method. Our goal is to minimize the variance of the estimation error while adhering to the unbiasedness constraint, as defined in Equation[3-45]. To achieve this, we introduce the auxiliary function  $f(\lambda_i, \xi_{OK})$ , which combines the variance of the estimation error with the term involving the Lagrange multiplier,  $\xi_{OK}$  (Webster & Oliver, 2007). By optimizing the auxiliary function  $f(\lambda_i, \xi_{OK})$ , while considering the weight constraint, we arrive at the formulation of the ordinary kriging system:

$\begin{cases} \sum_{i=1}^n \lambda_i^{OK} \gamma(\mathbf{x}_i - \mathbf{x}_j) + \xi_{OK} = \gamma(\mathbf{x}_i - \mathbf{x}_0), \text{ for } j = 1, \dots, n \\ \sum_{j=1}^n \lambda_j^{OK} = 1 \end{cases}$	3-49
---	------

The estimation variance of ordinary kriging can be expressed as:

$\sigma_E^2 = \xi_{OK} + \sum_{j=1}^n \lambda_j^{OK} \gamma(\mathbf{x}_j - \mathbf{x}_0) - \gamma(\mathbf{x}_0 - \mathbf{x}_0)$	3-50
---	------

The Kriging system can be represented in matrix form:

$A\lambda = L$	
<p>where:</p> $A = \begin{bmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_1) & \dots & \gamma(\mathbf{x}_1 - \mathbf{x}_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(\mathbf{x}_n - \mathbf{x}_1) & \dots & \gamma(\mathbf{x}_n - \mathbf{x}_n) & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix},$ $\lambda = \begin{bmatrix} \lambda_1^{OK} \\ \vdots \\ \lambda_n^{OK} \\ \xi_{OK} \end{bmatrix} \text{ and } L = \begin{bmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_0) \\ \vdots \\ \gamma(\mathbf{x}_n - \mathbf{x}_0) \\ 1 \end{bmatrix}$	3-51

According to (Wackernagel, 2003), it is mentioned that ordinary kriging is an exact estimator. Specifically, when the estimation point,  $\mathbf{x}_0$ , coincides with one of the data locations, the estimated value is guaranteed to be equal to the observed value at that particular point.

$Z_{OK}^*(\mathbf{x}_0) = Z(\mathbf{x}_i), \text{ if } \mathbf{x}_0 = \mathbf{x}_i$	3-52
---	------

Upon examination of the ordinary kriging solution, it becomes evident that the weights  $\lambda_i^{OK}$  assigned to nearby points in the vicinity of the estimated point are relatively high. Specifically, the first nearest group of points contributes approximately 80% of the total weight. The determination of these weights is influenced by both the spatial position of the neighboring points and the underlying variogram function. In cases where a significant nugget effect is present, the weights assigned to nearby points close to the estimation point tend to be smaller. Additionally, it is worth noting that the weights are contingent upon the specific configuration of the sampling points (Webster & Oliver, 2007).

## 4. Results and Case Studies

This chapter presents the findings of two case studies conducted as part of this thesis, with one study conducted in Italy and the other in Poland. In each study area, a network of GNSS stations is distributed. By processing the observations from these stations, we can determine the precise positions of each station and estimate the *zenith total delay* (ZTD) values.

Following the management of the data sets, we apply the stochastic prediction approaches outlined in the previous chapter. These approaches enable us to generate turbulence maps for each study area, with different time resolutions considered. By utilizing these prediction methods, we aim to capture and visualize the spatial and temporal distribution of turbulence in the respective regions.

To validate the effectiveness of our approach, we employ two different techniques. Firstly, we compare our turbulence predictions with measurements from another network's station at the same location, enabling us to assess the consistency and accuracy of our predictions through external validation. Additionally, we utilize the *Leave-One-Out technique*, systematically excluding individual data points from our analysis, to evaluate the performance of our approach based on the remaining data.

The subsequent sections of this chapter provide a detailed analysis of the results obtained from the case studies conducted in Italy and Poland. Through the integration of GNSS observations, stochastic prediction approaches, and validation procedures, we aim to improve our capability to effectively observe and understand turbulence patterns in these regions.

## 4.1 Case Study1

### 4.1.1 Study Area

The first case study was carried out within the Lombardy, Piemont, and Valle d'Aosta regions of northern Italy, as illustrated in Figure [4.1]. Geographically, the study area spans latitudes ranging from 44°57'N to 45°49'N and longitudes from 6°59'E to 10°13'E. Notably, this region exhibits considerable variability in terrain characteristics. Beginning from relatively flat areas in the southern portion, the topography gradually transitions into undulating hills, ultimately culminating in the prominent presence of the Alps in the northern reaches.

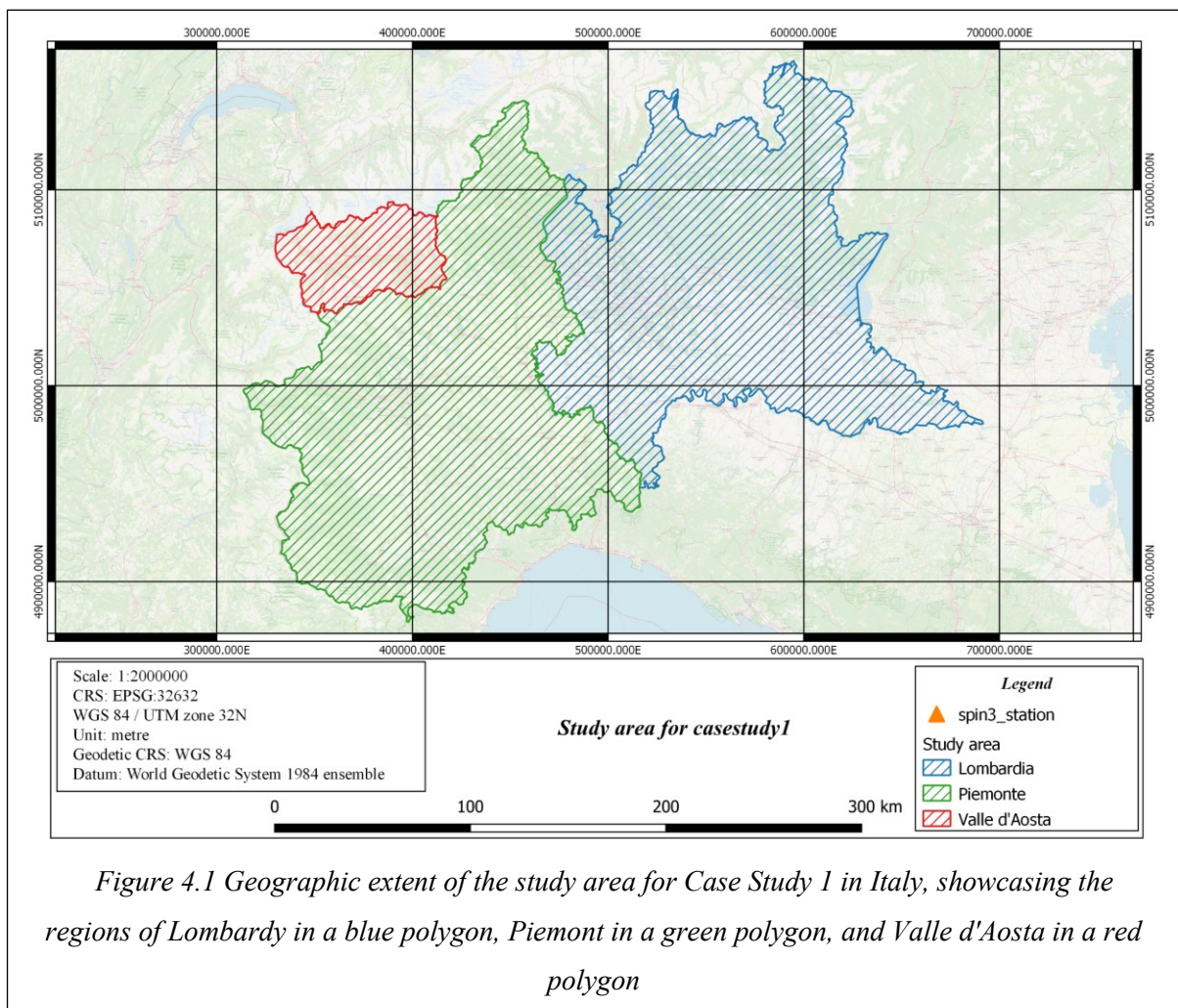


Figure 4.1 Geographic extent of the study area for Case Study 1 in Italy, showcasing the regions of Lombardia in a blue polygon, Piemont in a green polygon, and Valle d'Aosta in a red polygon

## 4.1.2 GNSS network

In this case study, we utilized observations from the GPS, GLONASS, and GALILEO constellations obtained from 34 out of the 39 permanent stations within the Interregional GNSS Positioning Service of Lombardy, Piedmont, and Valle d'Aosta regions (SPIN3 GNSS). Figure [4.2] provides an overview of all 39 stations, while Figure [4.3] specifically depicts the 34 stations used in this thesis. The SPIN3 network is equipped with geodetic receivers capable of receiving signals from multiple satellite constellations, with each receiver connected to a calibrated GNSS antenna. The period of interest for the observations spanned from *December 31, 2018, at 23:59:42 to December 31, 2020, at 23:59:42*. The data was collected at a temporal resolution of 30 seconds. The minimum inner Euclidean distance between stations was found to be *19.4 km*, while the maximum distance reached *340 km*. This network serves various applications, including geodetic, geophysical, geological, precision agriculture, cartography, GIS, engineering, and atmospheric applications (*SPIN3 GNSS – Servizio Di Posizionamento Interregionale GNSS, n.d.*).

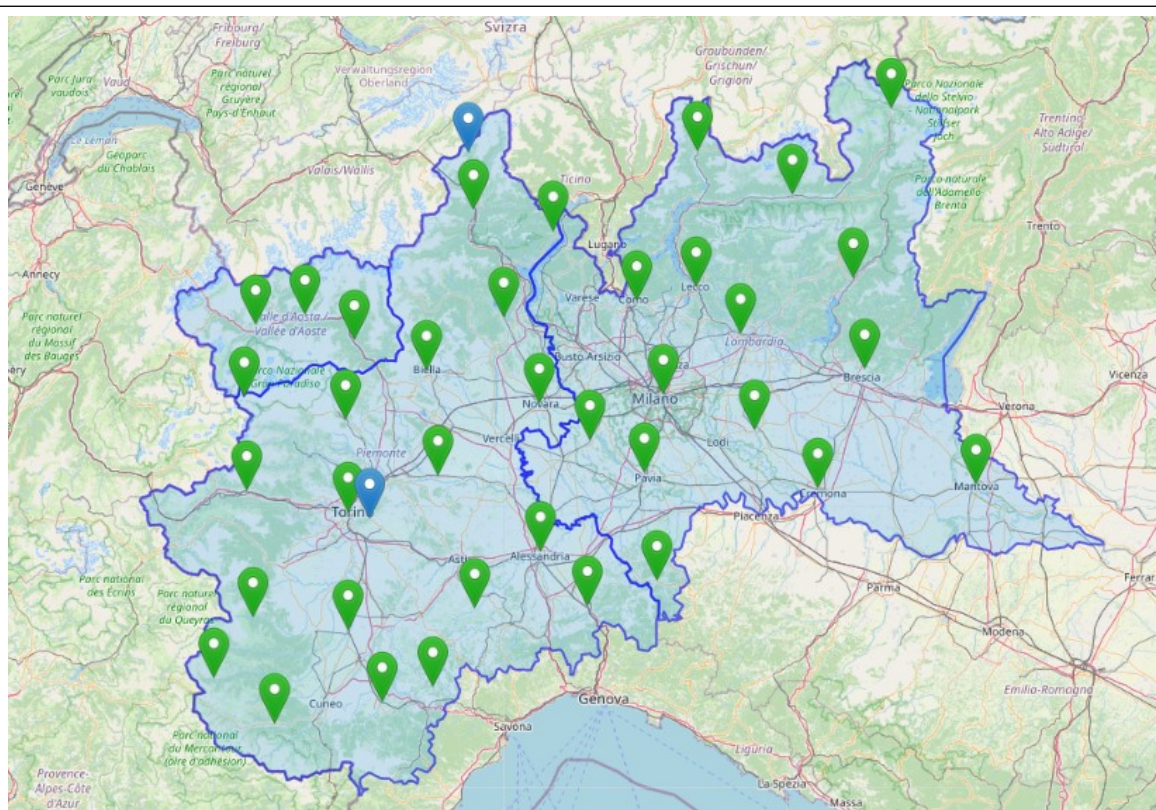


Figure 4.2 The full SPIN3 Network permanent stations distribution in Lombardy, Piedmont, and Valle d'Aosta regions

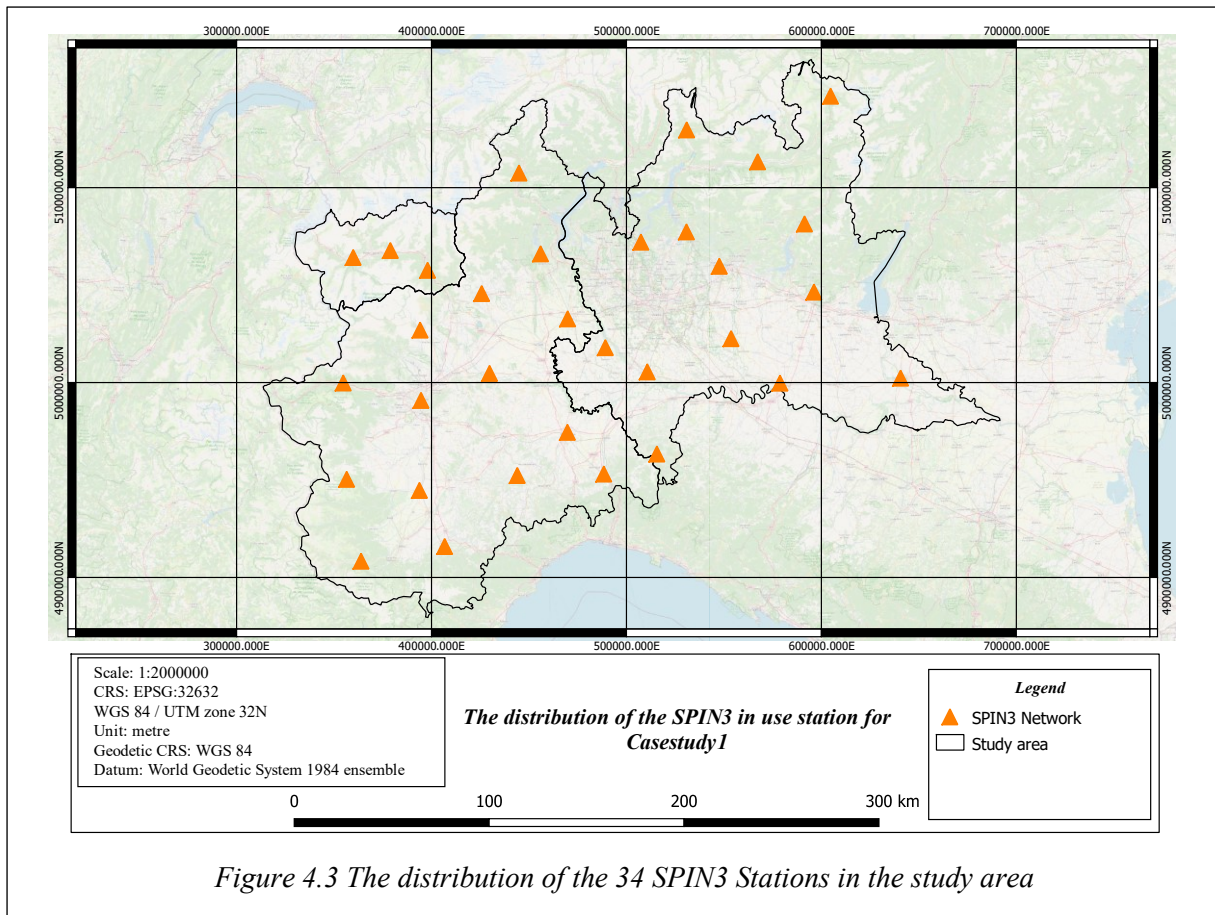


Figure 4.3 The distribution of the 34 SPIN3 Stations in the study area

The observation data were processed using the BERNESE software, employing the Precise Point Positioning (PPP) mode with the processing settings outlined in Table [4.1] as described by (Fermi, 2018). The processing outputs included the precise positions of each network station in the Geodetic coordinate reference system ( $\varphi, \lambda, H$ ) and the Earth-Centered, Earth-Fixed coordinate system ( $ECEF$ ) ( $X, Y, Z$ ). Additionally, the estimation of Zenith Total Delay ( $ZTD$ ) was obtained for each station during the specified period of interest for example the  $ZTD$  for Alessandria, Biella, and Bormio stations in Figure [4.5].

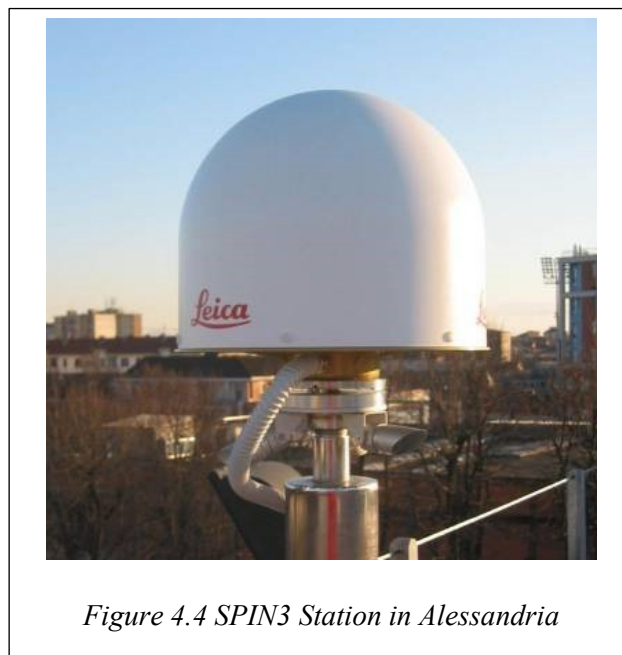
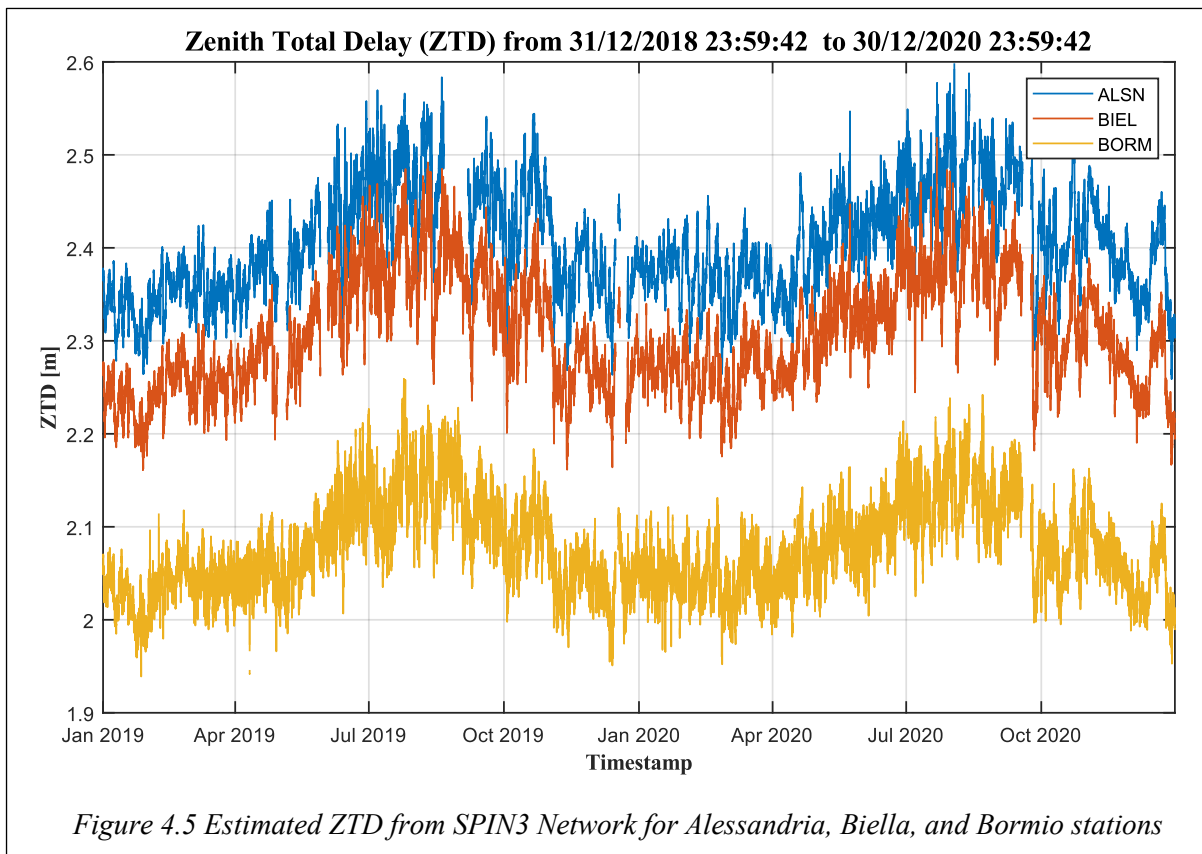


Figure 4.4 SPIN3 Station in Alessandria

<b>Network design</b>	30 dual-frequency GNSS receivers
<b>Received signals</b>	GPS, GLONASS
<b>Processing SW</b>	BERNESE v5.2
<b>Processing rate</b>	30 s
<b>Elevation cutoff</b>	10 degrees
<b>Processed signals</b>	GPS
<b>Processing method</b>	PPP
<b>Ephemeris</b>	CODE (final)
<b>Satellite clocks</b>	CODE 30 s (final)
<b>Ocean loading model</b>	FES2004
<b>Observables</b>	L3 (iono-free)
<b>Mapping functions</b>	GMF dry/wet
<b>Dry tropospheric model</b>	Saastamoinen + GPT
<b>Gradient estimation model</b>	none
<b>ZWD estimation rate</b>	10 min
<b>A priori ZWD difference std</b>	0.02 m



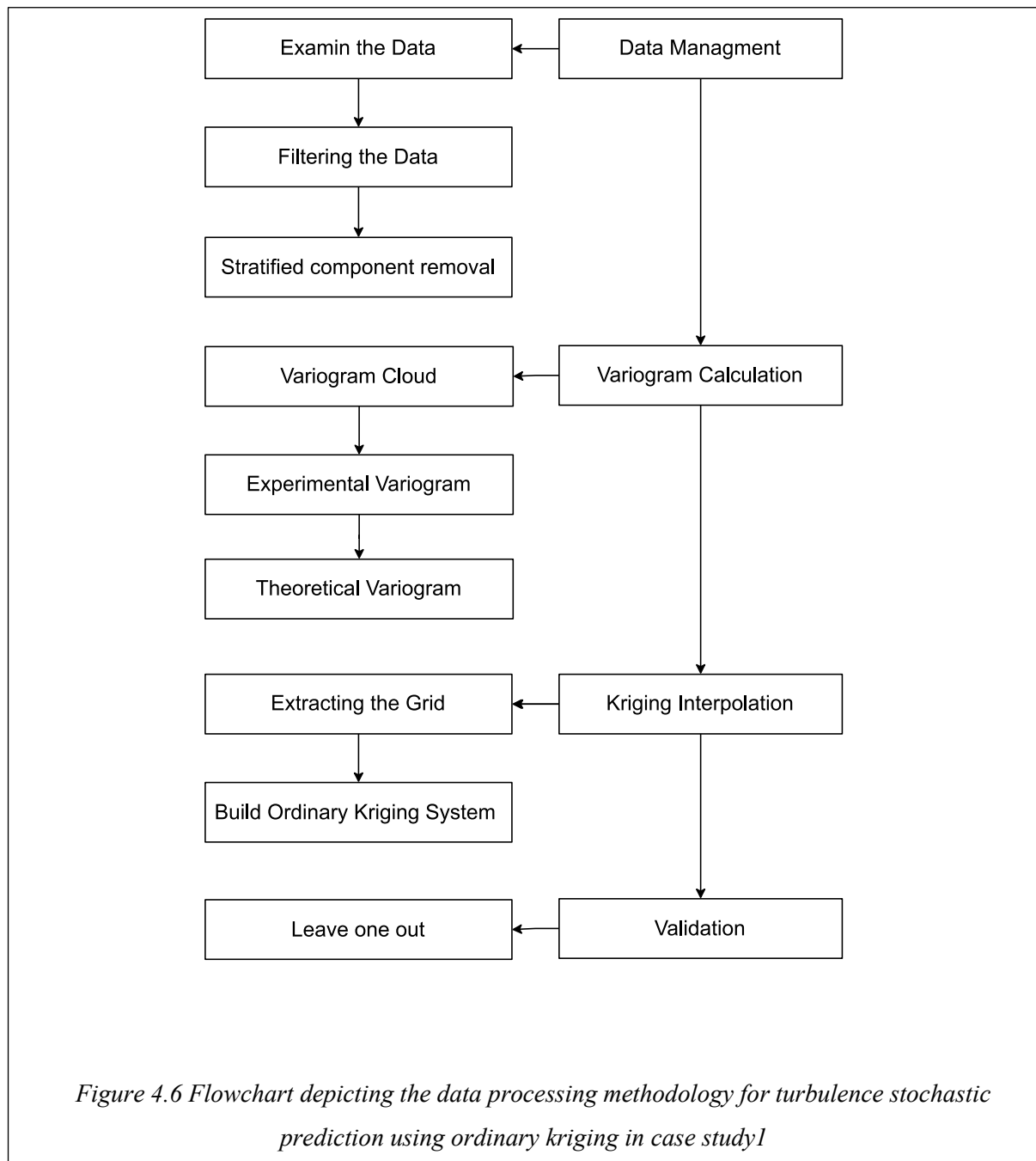
*Table 4-2 The 34 SPIN3 stations utilized in our computations, along with the respective province, city, and geodetic coordinates for each station*

<b>Station</b>	<b>Region</b>	<b>Province</b>	<b>City</b>	<b>Latitude <math>\varphi</math> (°)</b>	<b>Longitude <math>\lambda</math> (°)</b>	<b>Elep.height H(m)</b>
<b>ALSN</b>	PIE	AL	Alessandria	44.92317	8.616345	103.34
<b>BIEL</b>	PIE	BI	Biella	45.56074	8.048051	429.81
<b>BORM</b>	LOM	SO	Bormio	46.46818	10.36397	1211.13
<b>BREU</b>	LOM	BS	Brescia	45.56494	10.23259	179.21
<b>BUSL</b>	PIE	TO	Bussoleno	45.1368	7.152153	442.14
<b>CANL</b>	PIE	AT	Canelli	44.72226	8.292822	159.68
<b>CHIA</b>	LOM	SO	Cliavenna	46.31985	9.401271	340.86
<b>COMO</b>	LOM	CO	Como	45.80216	9.095619	244.89
<b>CREA</b>	LOM	CR	Crema	45.35433	9.685305	86.85
<b>CREM</b>	LOM	CR	Cremona	45.14665	10.00195	61.65
<b>CRSN</b>	PIE	VC	Crescentino	45.19244	8.105767	164.66
<b>CUOR</b>	PIE	TO	Cuorné	45.38791	7.647698	430.57
<b>CURN</b>	LOM	BG	Curno	45.6923	9.612006	251.5
<b>DARF</b>	LOM	BS	Dario B. T.	45.88029	10.17717	233.31
<b>DEMN</b>	PIE	CN	Demonte	44.31572	7.292628	809.74
<b>DOMS</b>	PIE	VB	Domodossola	46.11904	8.286326	313
<b>GOZZ</b>	PIE	NO	Gozzano	45.74674	8.433397	367.24
<b>GREM</b>	LOM	CO	Lomazzo	45.69718	9.035299	311.16
<b>LECO</b>	LOM	LC	Lecco	45.84886	9.396556	226.43
<b>MANT</b>	LOM	MN	Mantova	45.1601	10.7894	103.34
<b>MIL2</b>	LOM	MI	Milano	45.47837	9.229213	429.81
<b>MONV</b>	PIE	CN	Mondovi	44.39036	7.828955	1211.13
<b>NOVR</b>	PIE	NO	Novara	45.44721	8.613964	179.21
<b>OSTA</b>	PIE	CN	Ostana	44.69234	7.18834	442.14
<b>PAVI</b>	LOM	PA	Pavia	45.20298	9.13614	159.68
<b>RUMI</b>	AO	AO	Rumiod	45.71712	7.198678	340.86
<b>SAVI</b>	PIE	CN	Savigliano	44.64764	7.660664	244.89
<b>SERR</b>	PIE	AL	Serravalle S.	44.73106	8.853093	86.85
<b>SONP</b>	LOM	SO	Sondrio	46.16982	9.871937	61.65
<b>TNUS</b>	VDA	VDA	Quart	45.75289	7.442627	164.66
<b>TORI</b>	PIE	TO	Torino	45.06337	7.661278	430.57
<b>VARZ</b>	LOM	PV	Varzi	44.82334	9.197405	251.5
<b>VERR</b>	AO	AO	Verres	45.66467	7.690182	233.31
<b>VIGE</b>	LOM	PV	Vigevano	45.31479	8.861951	809.74



### 4.1.3 Data Processing

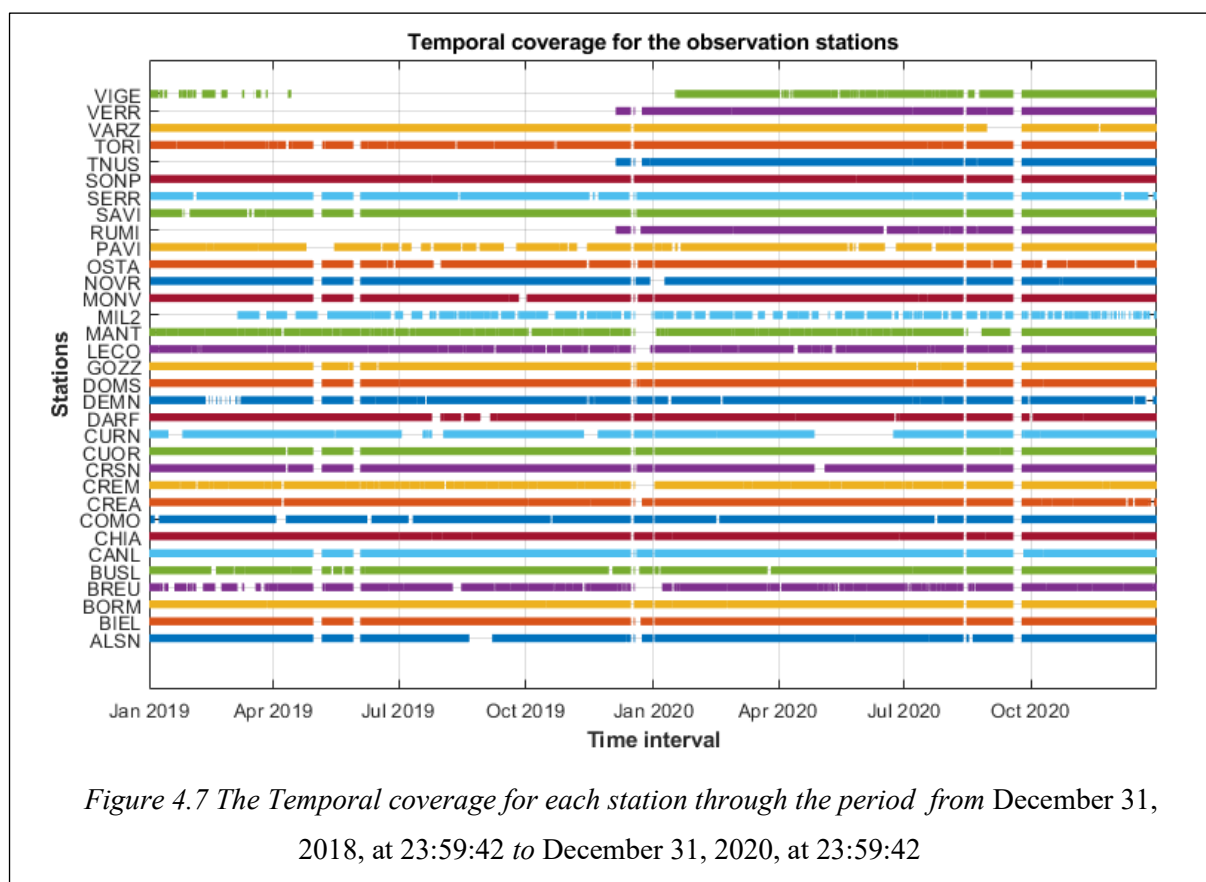
This section provides a comprehensive overview of the various stages involved in our analysis, starting from the initial exploration and management of the observation data. We then proceed to the stochastic prediction procedures, which encompass estimating the variogram cloud, constructing the experimental variogram, and finally determining the theoretical variogram. The theoretical variogram is subsequently utilized as input for the ordinary kriging method. To validate our results, we employ the Leave-One-Out procedure. The entire workflow is depicted in Figure [4.6].

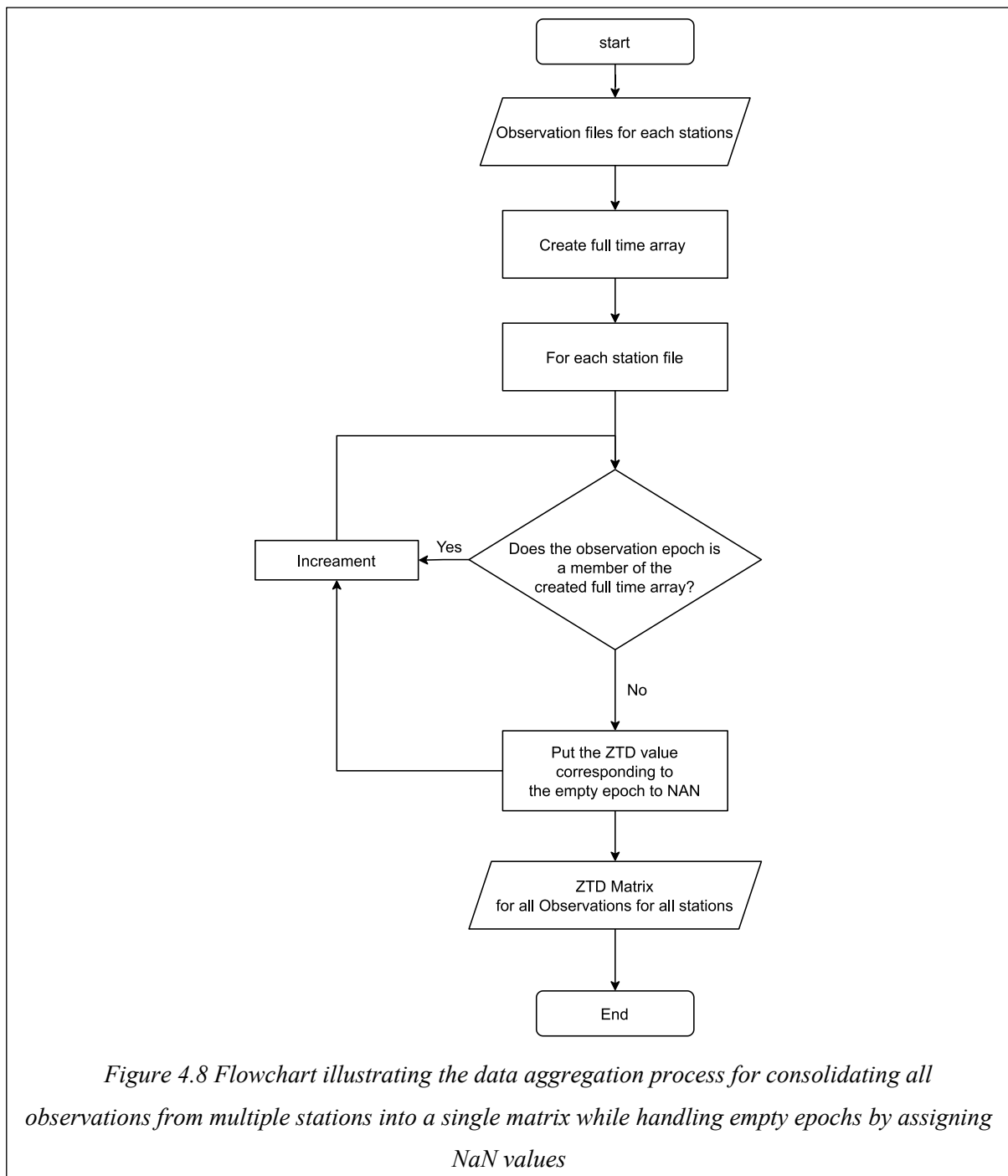


### 4.1.3.1 Data Management

The data management section comprises three primary stages, namely exploration and examination of the observation data file, data filtering, and estimation of the turbulence delay.

During the examination of each observation file for a specific station, the epochs of observation and their corresponding Zenith Total Delay (*ZTD*) estimates are observed. Analysis of the observation files reveals the presence of missing epochs, as depicted in Figure [4.7]. The integration of data from multiple stations into a consolidated matrix poses challenges due to these missing epochs. To address this issue, a time array is created, spanning from *December 31, 2018, at 23:59:42* to *December 31, 2020, at 23:59:42*, with 30-second intervals. A comparison is then made between the observation epochs of each station and the created time array to identify the missing epochs, which are subsequently assigned a "NaN" value. The flowchart presented in Figure [4.8] provides a graphical representation of these sequential steps.





After obtaining the ZTD matrix, it is possible to select a specific interval of interest within the study period. The user can specify the starting and ending epochs for this interval. By doing so, we can extract the corresponding observations from the ZTD matrix, resulting in a specific ZTD matrix for the chosen interval.

Figure [4.7] shows that certain stations had missing or insufficient observations during certain intervals. Therefore, it is necessary to filter the stations based on the number of NaN values compared to the total number of observations for each station in a particular column, which represents the station. If the number of NaN values exceeds 50% of the total number of observations, the station will not be included in our computations. The filtering process algorithm is illustrated in Figure [4.9].

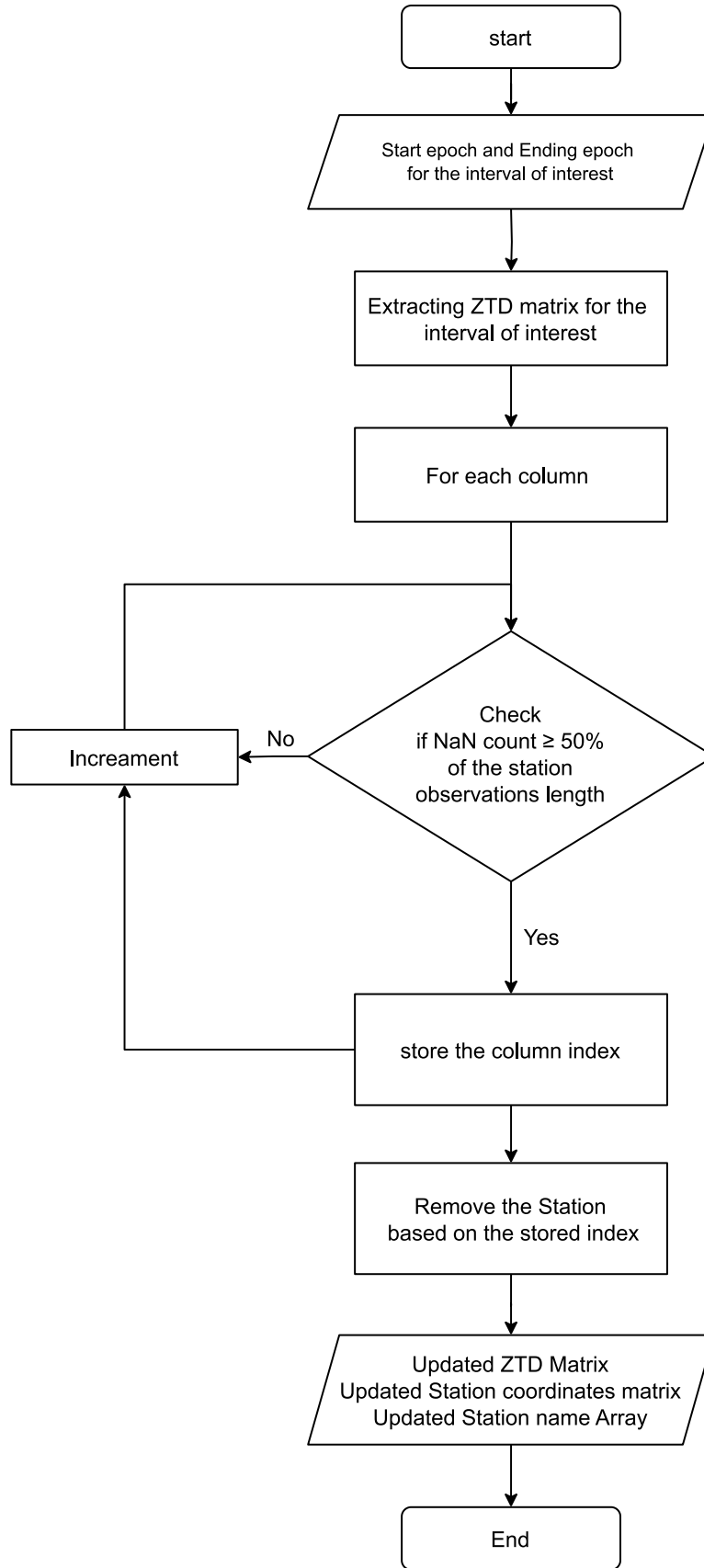


Figure 4.9 Flowchart demonstrates how stations are filtered based on the number of NaN values

In the process of estimating the turbulence delay from the Zenith Total Delay (*ZTD*) observations, it is necessary to delve deeper into the composition of *ZTD*. *ZTD* comprises two distinct components: the stratified part ( $S(h)_x^t$ ) and the turbulence part ( $T(E, N)_x^t$ ). The stratified part primarily originates from the presence of gases in the troposphere that are in a state of hydrostatic equilibrium. Conversely, the turbulence delay arises from the behavior of water vapor, which is not in hydrostatic equilibrium. It is important to note that even though water vapor also contributes to the stratified part, our focus lies on the portion of water vapor that deviates from hydrostatic equilibrium.

$ZTD_x^t = S(h)_x^t + T(E, N)_x^t$	4-1
------------------------------------	-----

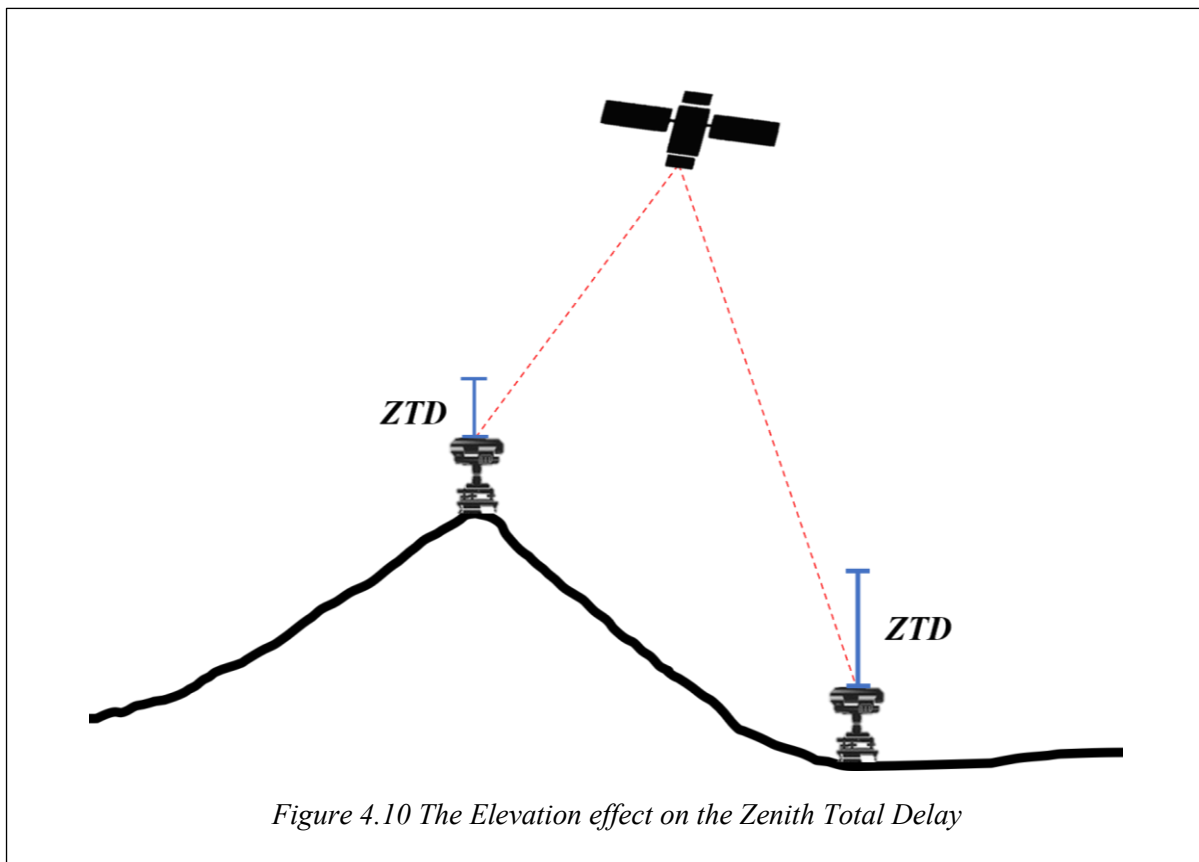
Where:

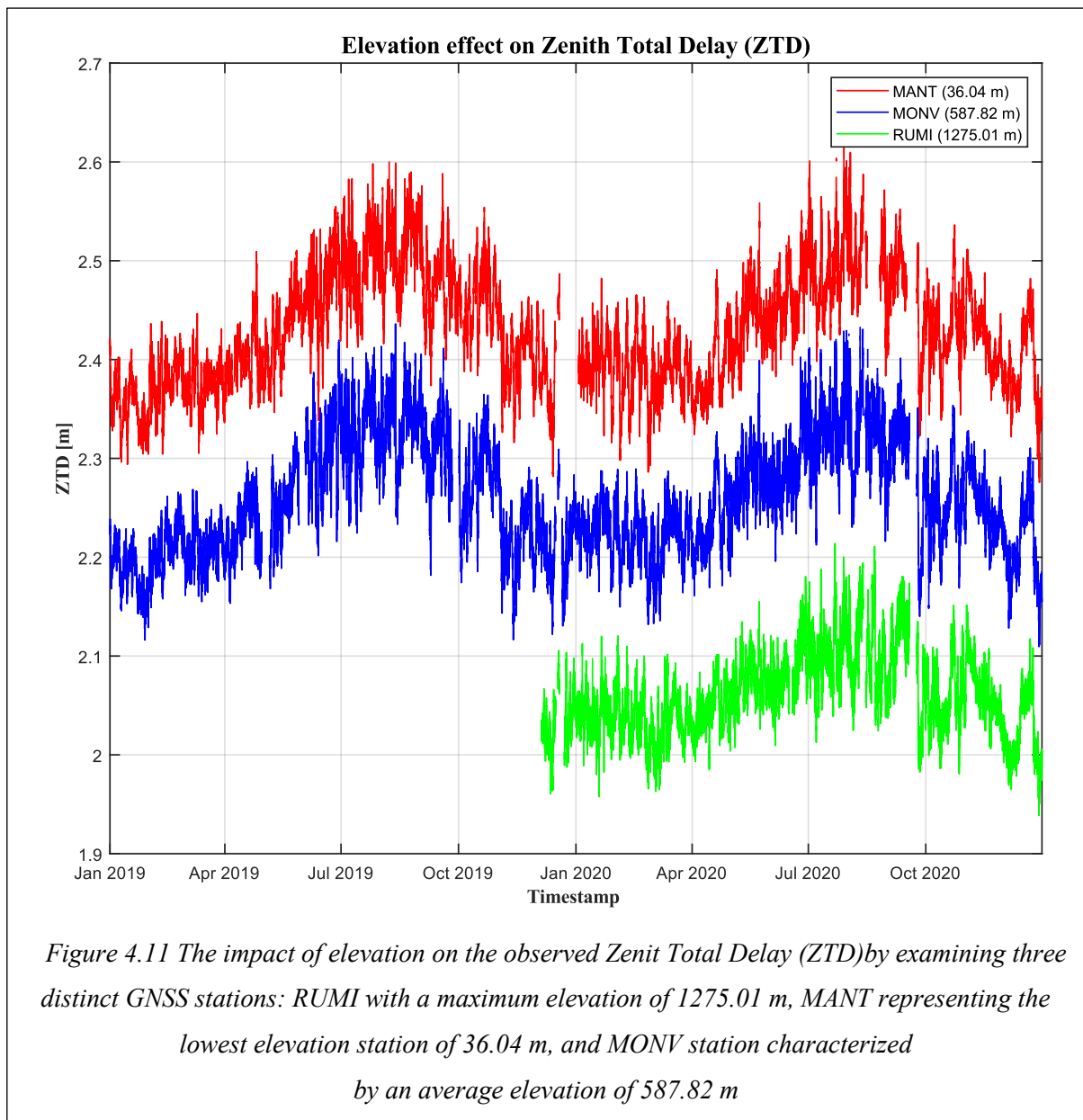
$ZTD_x^t$ : The Zenith total delay at epoch  $t$  for point  $x$ .

$S(h)_x^t$ : The stratified component at epoch  $t$  for point  $x$  that is a function of the point elevation.

$T(E, N)_x^t$ : The Turbulence component at epoch  $t$  for point  $x$  is a function of the point position.

It is crucial to recognize that these two components exhibit different characteristics. The stratified part predominantly depends on the observation station as we can notice when the station is on a high elevation. For instance, when a GNSS station is located in a mountainous region, the GNSS signal traverses multiple layers of the atmosphere compared to a station situated at lower elevations as shown in Figure[4.10]. We can see this effect in our *ZTD* observations as shown in Figure[4.11]. In contrast, the turbulence component is a two-dimensional field.





The estimation of the turbulence component, which characterizes the non-hydrostatic equilibrium portion of water vapor, necessitates the subtraction of the height-dependent contribution from the overall Zenith Total Delay (ZTD) observations. This subtraction process effectively isolates the residual turbulence component, which is of particular interest in our analysis.

Furthermore, to effectively model the stratified component of the Zenith Total Delay (ZTD), it is imperative to investigate the relationship between the elevation of network stations and the corresponding total delay values. Through a comprehensive analysis of the ZTD observations from the SPIN3 network stations, a robust association becomes evident, exhibiting a significant correlation between station elevation and the observed total delay. This strong correlation is visually represented in Figure [4.12].

Notably, we observe a decrease in ZTD with increasing station elevation, aligning with our expectations. For instance, the ZTD at the MANT station, located in a high-elevation mountainous area, is higher compared to the RUMI station. This observation aligns with the notion that higher station elevations correspond to a greater wet air column above, contributing to increased total delay.

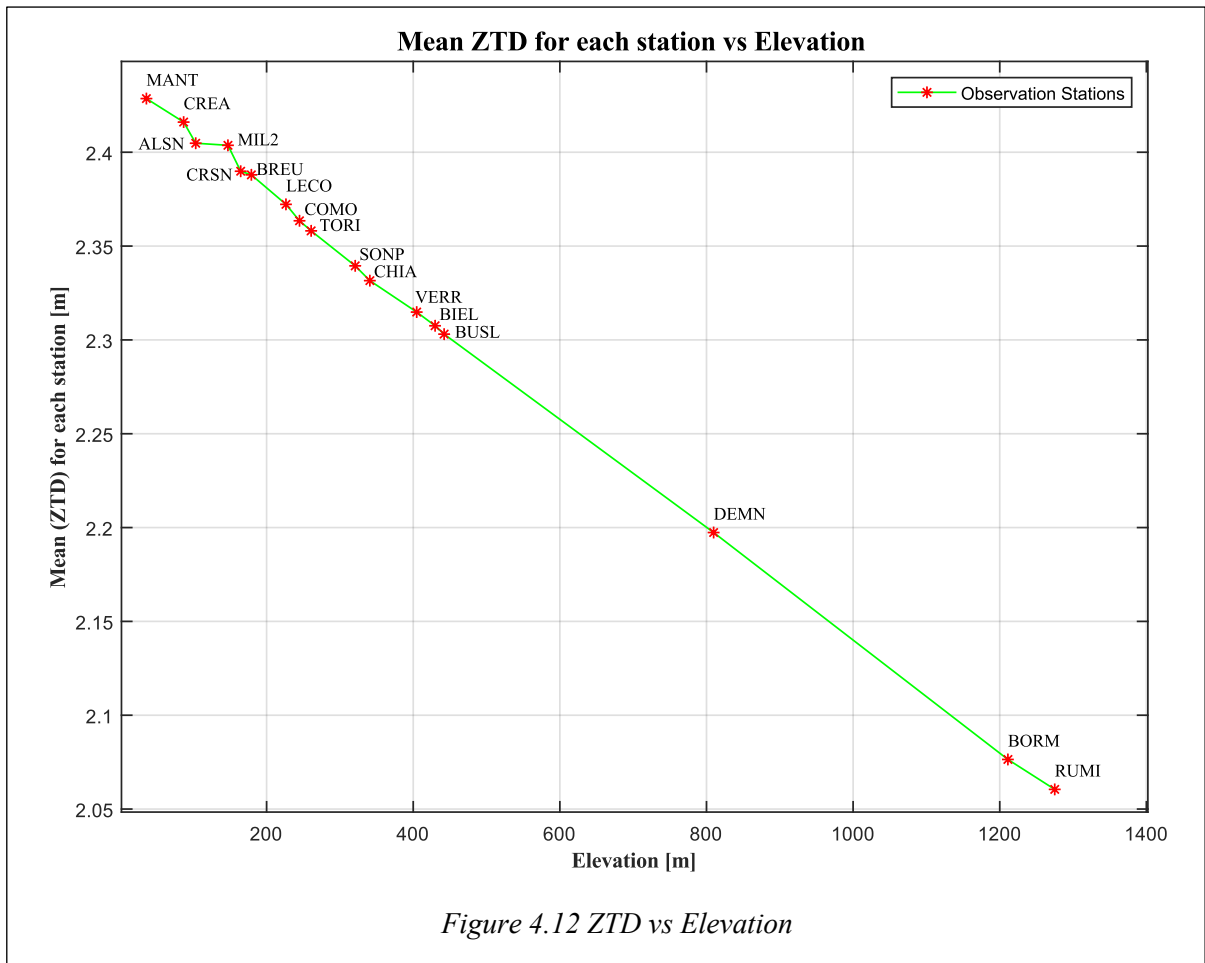


Figure 4.12 ZTD vs Elevation

The relationship between the height and the stratified component can be modeled as

$S(h)_x^t = \beta_t e^{-\alpha_t h_x}$	4-2
--	-----

where:

$S(h)_x^t$ : The stratified component at epoch  $t$  for station  $x$ .

$h_x$ : The elevation of point  $x$ .

$\alpha_t, \beta_t$ : The model coefficients that have to be estimated.

As the model (4.2) is nonlinear, a linearization approach is necessary to obtain a least squares solution. The flowchart presented in Figure [4.13] outlines the steps involved in the solution process. This includes the construction of the *Design matrix A*, the *residual vector L*, and subsequent computation of the *normal matrix N* and the *absolute term* of the system. By employing the least squares method described in Equation [4-3], the values of the model parameters  $\alpha_t$  and  $\beta_t$  can be determined for each epoch and observation station. Subsequently, the turbulence  $T(E, N)_x^t$  can be obtained by subtracting the modeled stratified component  $S(h)_x^t$  from the  $ZTD_x^t$ .

$$ZTD_x^t = \beta_t e^{-\alpha_t h_x}$$

Let :

$$A = [e^{-\tilde{\alpha}_t h_x} \quad -h_x \tilde{\beta}_t e^{-\tilde{\alpha}_t h_x}],$$

$$L = [ZTD_x^t - \tilde{\beta}_t e^{-\tilde{\alpha}_t h_x}],$$

$$X = \begin{bmatrix} \delta\beta = \beta - \tilde{\beta}_t \\ \delta\alpha = \alpha - \tilde{\alpha}_t \end{bmatrix},$$

Then the solution will be:

$$\therefore \begin{bmatrix} \delta\beta \\ \delta\alpha \end{bmatrix} = \underbrace{(A^T A)^{-1}}_N A^T L$$

$$\therefore \begin{bmatrix} \tilde{\beta}_t \\ \tilde{\alpha}_t \end{bmatrix} = \begin{bmatrix} \beta \\ \alpha \end{bmatrix} + \begin{bmatrix} \delta\beta \\ \delta\alpha \end{bmatrix}$$

Thus

$$T(\widehat{E, N})_x^t = ZTD_x^t - \tilde{\beta}_t e^{-\tilde{\alpha}_t h_x}$$

4-3



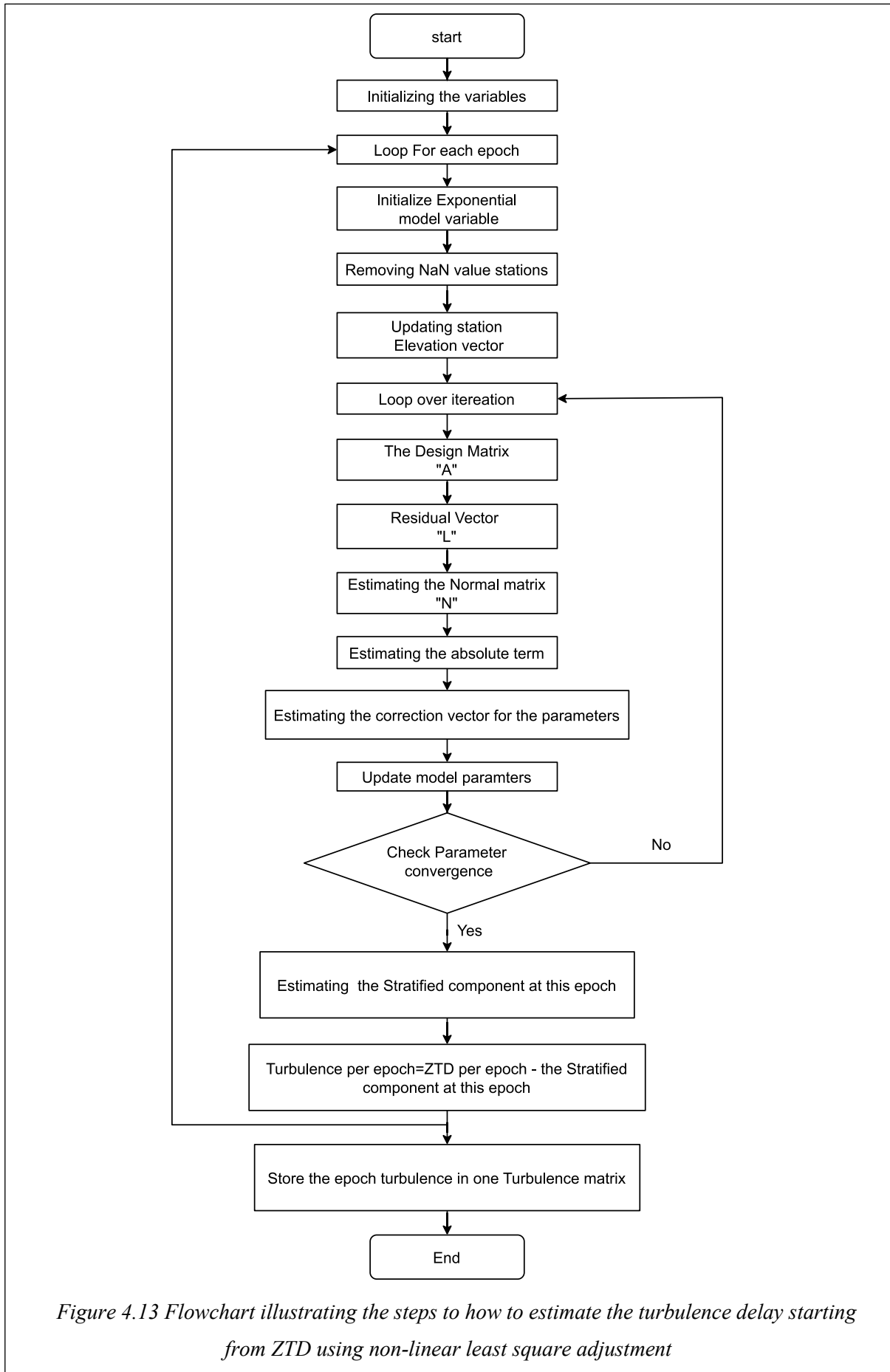


Figure 4.13 Flowchart illustrating the steps to how to estimate the turbulence delay starting from ZTD using non-linear least square adjustment

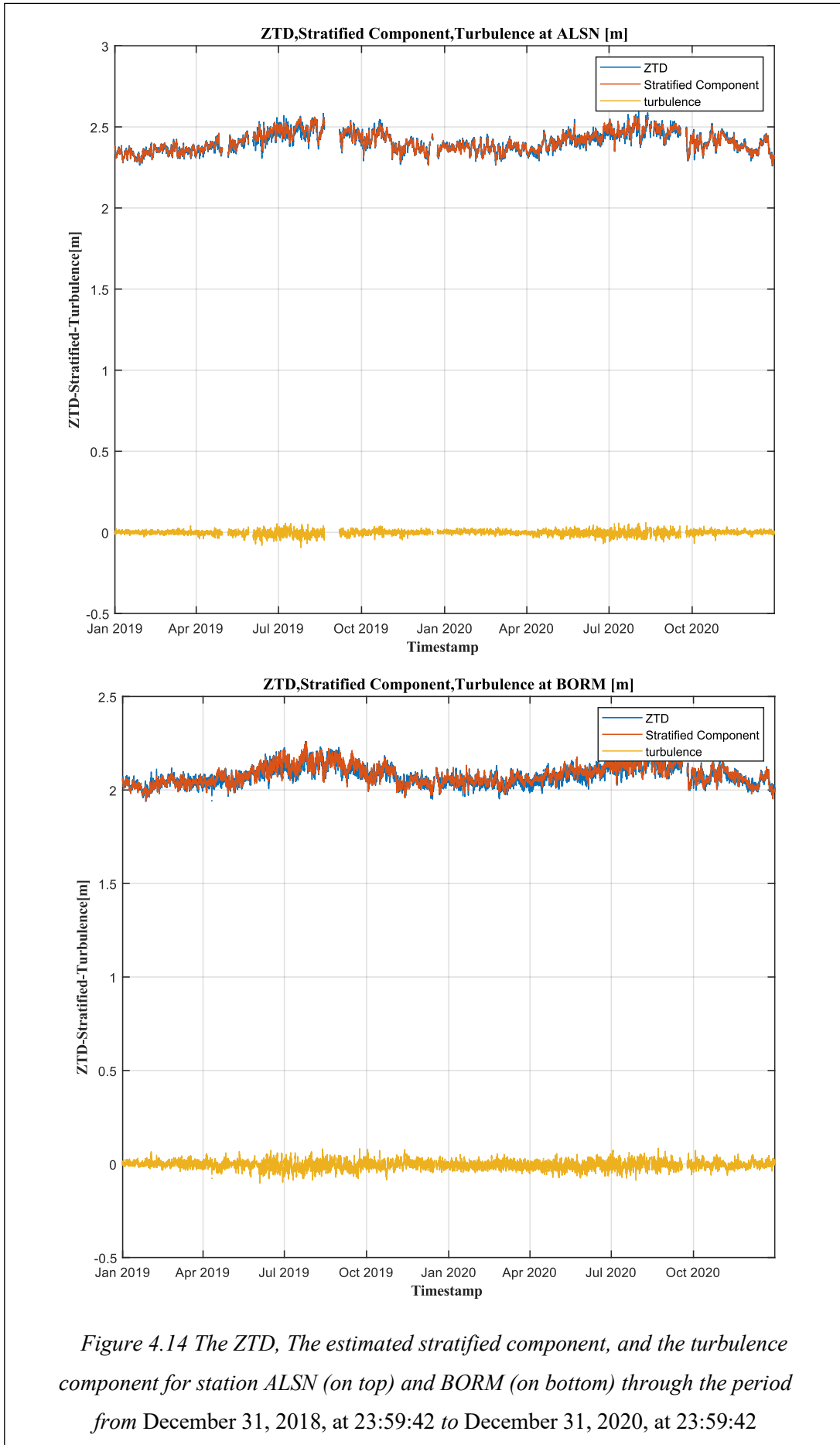


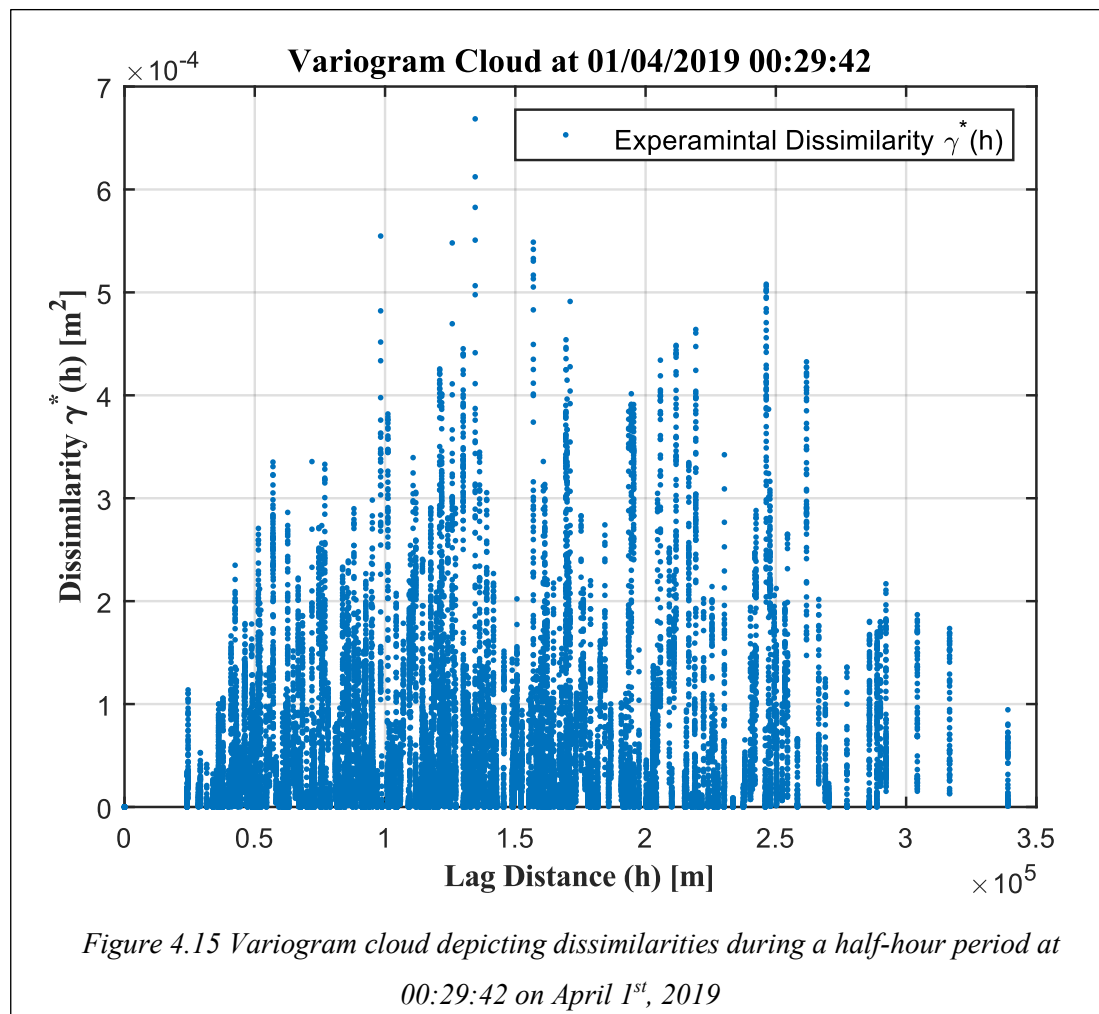
Figure 4.14 The ZTD, The estimated stratified component, and the turbulence component for station ALSN (on top) and BORM (on bottom) through the period from December 31, 2018, at 23:59:42 to December 31, 2020, at 23:59:42

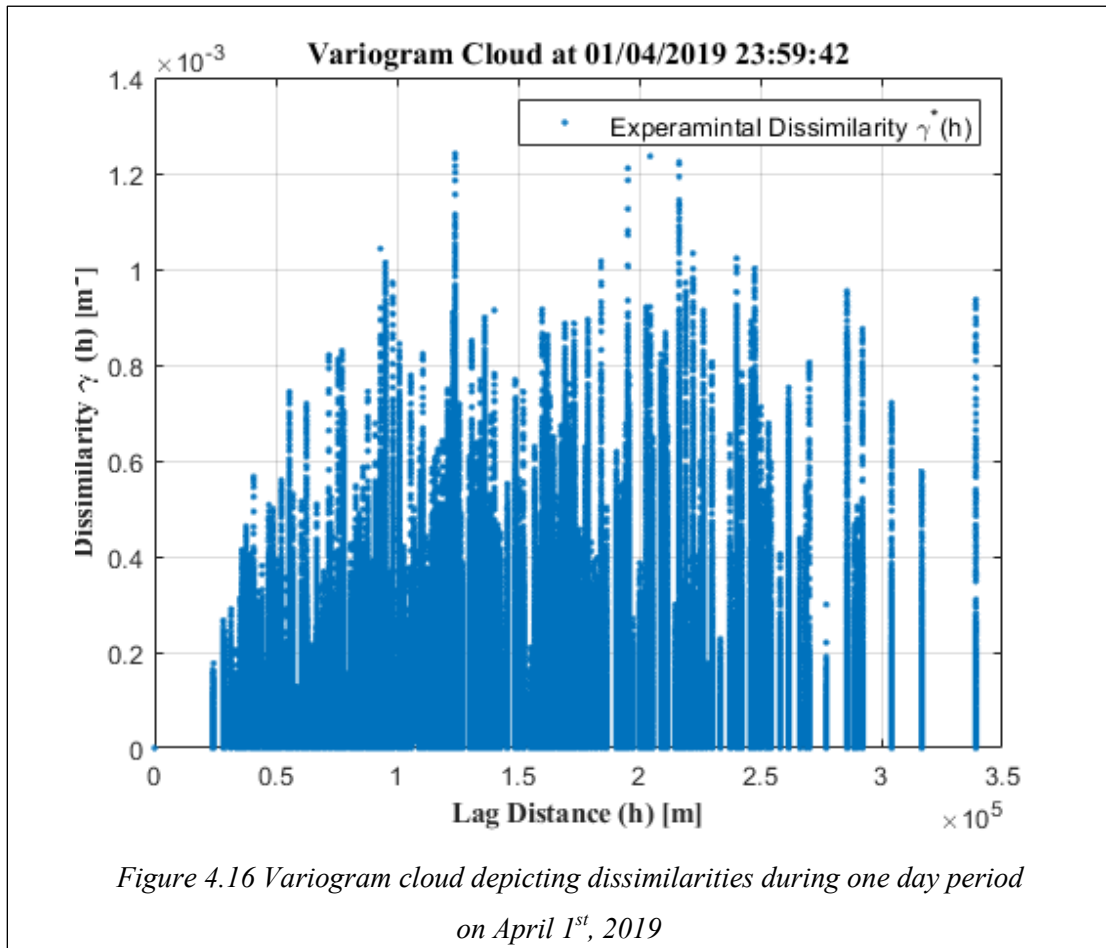
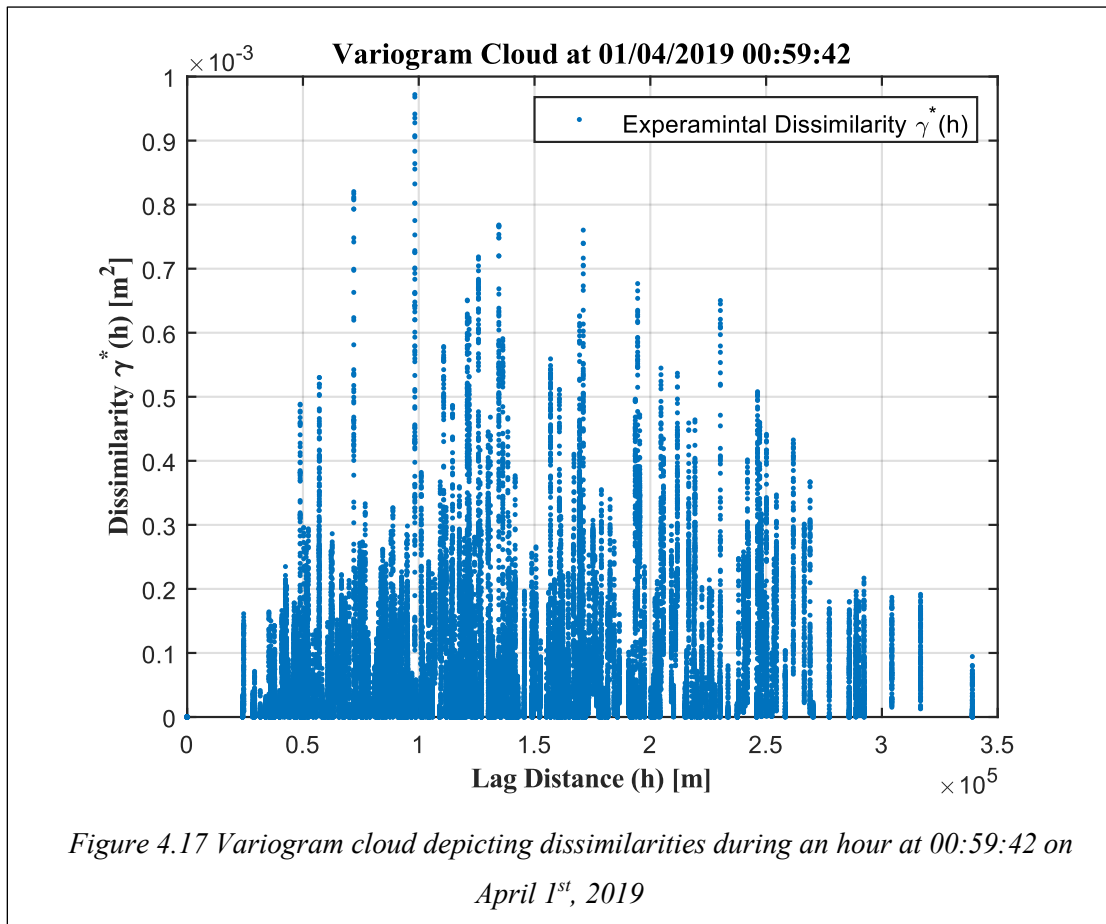
### 4.1.3.2 Variogram Calculation

This section aims to investigate the spatiotemporal variability of turbulence within the geographical coverage of the SPIN3 network. Building upon the comprehensive explanations and concepts outlined in Chapter 3, our objective is to analyze the fluctuations of the turbulence component at both spatial and temporal scales.

To initiate stochastic prediction, the determination of the required temporal resolution for generating turbulence maps is paramount. Considering that the SPIN3 stations observe at a rate of 30 seconds, the chosen temporal resolution can be specified as daily, hourly, or any other suitable interval. Subsequently, the turbulence matrix is segmented at the designated period, resulting in a three-dimensional matrix. Each dimension of this matrix represents the turbulence delay observed for all stations during that particular epoch. By applying intrinsic homogeneity and isotropic assumptions, as elaborated in Section 3.4, we endeavor to estimate epoch-experimental variogram functions. These functions provide insights into the dissimilarity, denoted as  $\gamma_{ij}^*$ , of turbulence values between station pairs at each epoch, as per equation (3-32).

The first step is to compute the variogram cloud. We start by computing the Euclidean distances between stations within the SPIN3 network using the geographical coordinates of each station. Thus, we need to transform the geodetic coordinates ( $\varphi, \lambda, H$ ) or the ECEF coordinates ( $X, Y, Z$ ) to a projection coordinate system in this case study we used (the universal transverse Mercator UTM, zone 32N).





After generating the variogram cloud for each epoch, the subsequent step involves establishing the experimental variogram following Section 3.6.2. The lag distance is initially defined, which serves as the basis for dividing the total distances between stations into intervals. Within each interval, the classic estimator Equation [3-33] is applied to the midpoint. Consequently, it becomes necessary to determine a suitable lag distance that aligns with the distances between stations during a specific epoch.

For instance, let's consider April 1<sup>st</sup>, 2019, where a total of 29 stations are involved in the computations following the filtering process. Among these stations, the minimum distance between any pair is found to be 23.9 km. Choosing a lag distance smaller than this minimum value would result in an absence of data points. Thus, the lag distance must exceed the minimum distance between stations.

Additionally, considering the interest in the behavior near the origin, the lag distance should allow us to capture the variability exhibited by nearby stations. In the given example, a lag distance of 30 km is selected based on the available data to effectively represent the behavior of variability between closely situated stations as shown in Figure [4.18].

In the final step of this section, we proceed with constructing the theoretical variogram, as discussed in Section 3.6.3. Two fitting models, namely the exponential model and the power model, are considered. The power model offers the advantage of preserving the physical interpretation associated with atmospheric turbulence, following the principles of Kolmogorov's theory.

By examining Equation [3-43], which represents the power model, we apply a nonlinear least square adjustment to estimate the two parameters  $(\eta, \alpha)$ . This adjustment allows us to obtain parameter estimations for each epoch  $(\hat{\eta}_t, \hat{\alpha}_t)$ . These estimated parameters serve as key inputs for modeling the theoretical variogram corresponding to each epoch.

To estimate the parameters of the exponential model, as expressed in Equation [3-42], we employ a nonlinear least square approach. Two parameters need to be estimated: the sill "C" and the range "a". To initiate the estimation process, we require initial approximations for these parameters.

For the sill, we can set its initial value to the maximum value of dissimilarities observed at the epoch under consideration. This provides an initial approximation that captures the upper limit of variability in the dataset.

Regarding the range parameter, as mentioned in Section 3.6.3, it can be determined using the formula  $a = h/3$  where "h" represents the distance at which we observe 95% of the sill. This value of "h" corresponds to the distance within which the spatial dependence is strong. By incorporating this value into the range parameter estimation, we establish an initial approximation that aligns with the extent of spatial dependence in the dataset.

It is important to note that the theoretical variogram representation is limited to the half diameter of the study area, as depicted in Figure [4.20]. In our specific case, this means the variogram will be modeled up to a distance of 165 km.

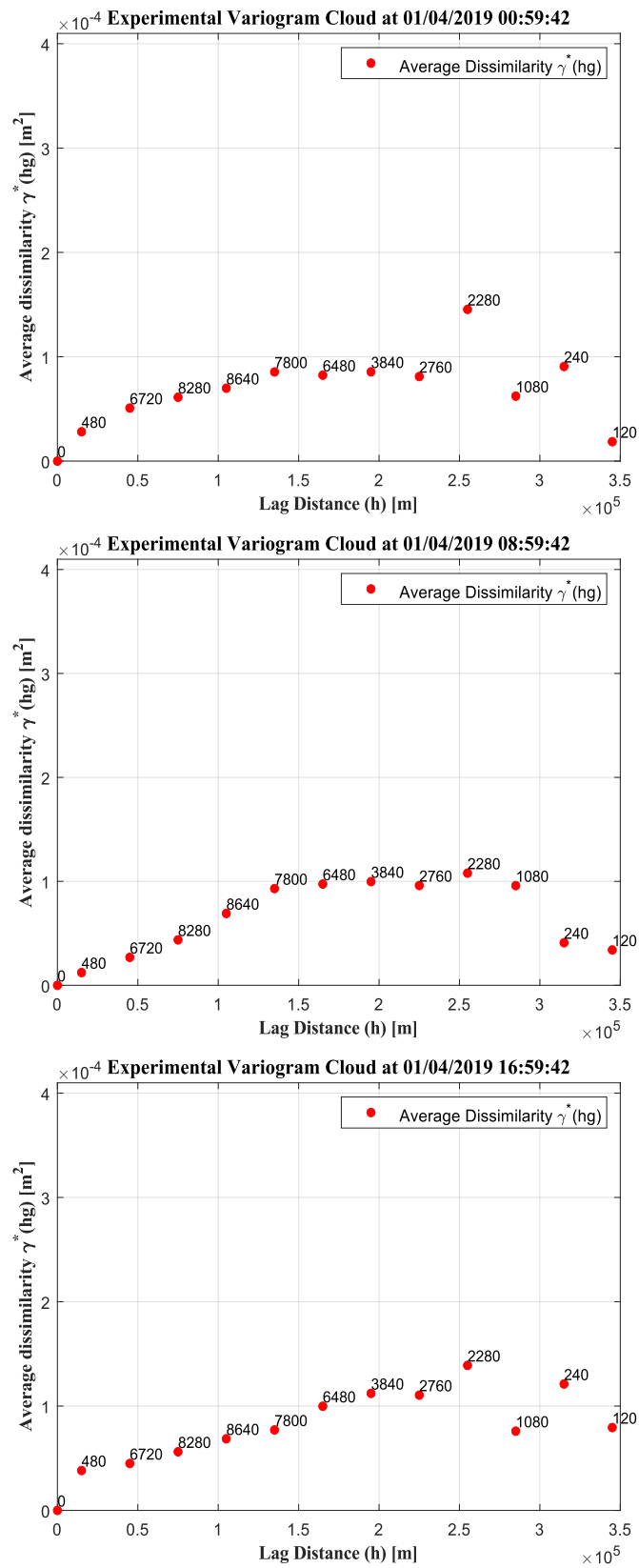


Figure 4.18 Experimental Variogram for April 1st, 2019 at three epochs (00:59:42, 08:59:42, and 18:59:42). The numbers above each point indicate the count of dissimilarities contributing to the average computation within each interval

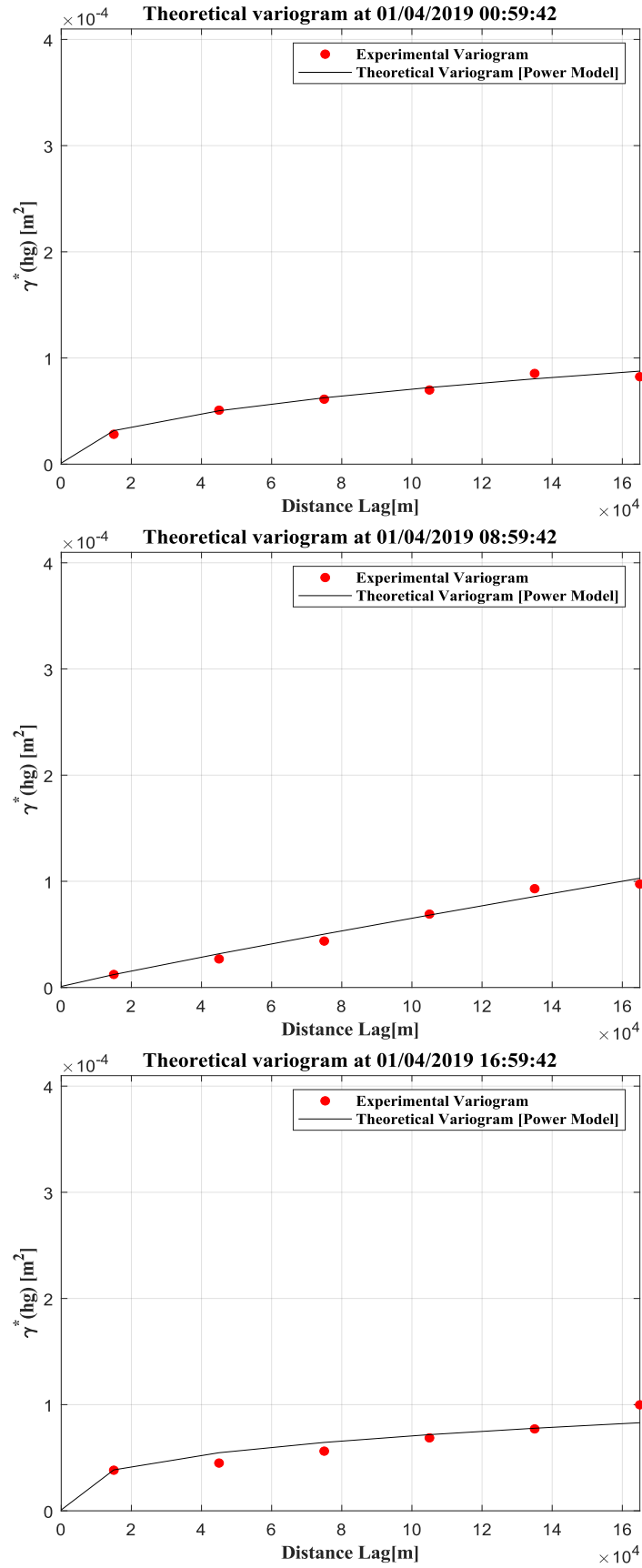


Figure 4.19 Theoretical variogram using the Power model for April 1st, 2019 at three epochs (00:59:42, 08:59:42, and 18:59:42).

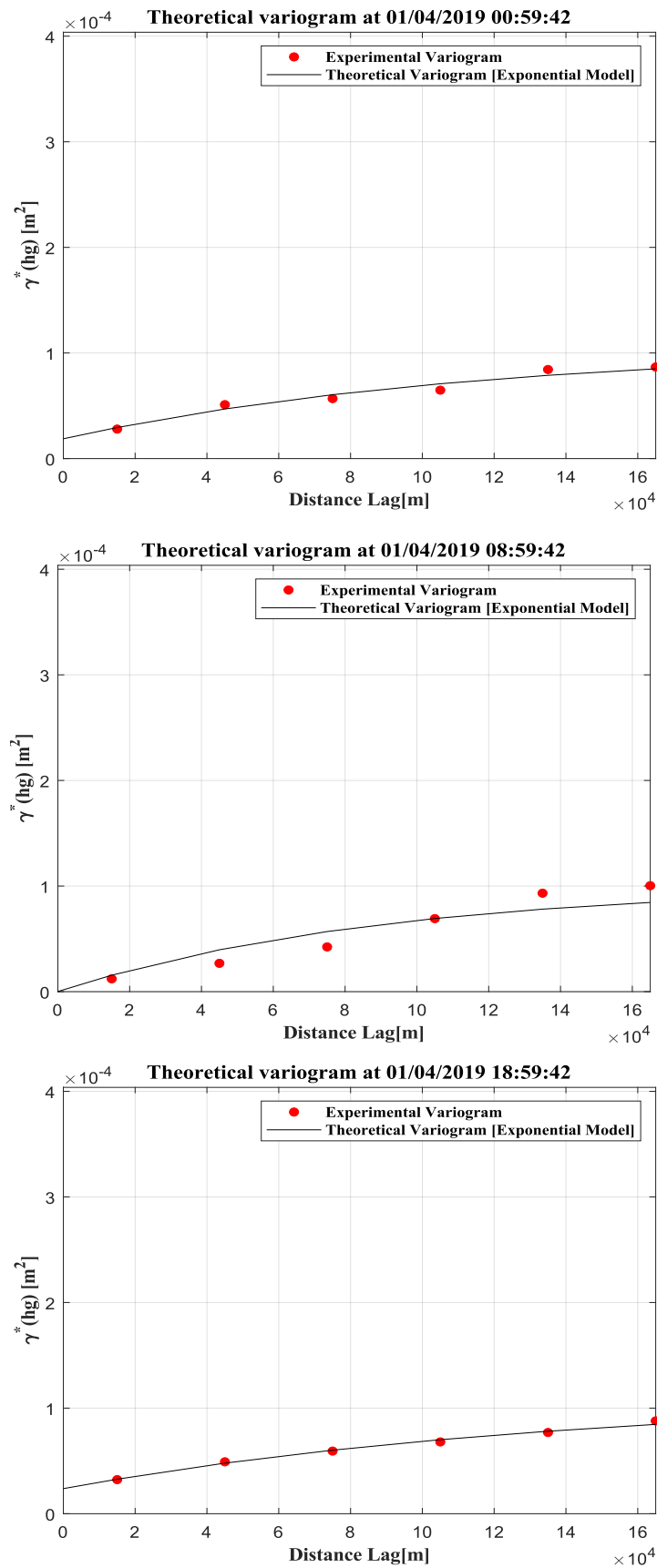


Figure 4.20 Theoretical variogram using the Exponential model for April 1st, 2019 at three epochs (00:59:42, 08:59:42, and 18:59:42)



### 4.1.3.3 Kriging Interpolation

In the pursuit of generating high-resolution turbulence maps, we propose utilizing SPIN3 stations and theoretical variogram functions to employ epoch-wise ordinary kriging. This approach allows us to estimate turbulence values for a grid covering the study area. The grid itself has been derived from a Digital Elevation Model (DEM), with a uniform grid point spacing of 5km, as depicted in Figure [4.21].

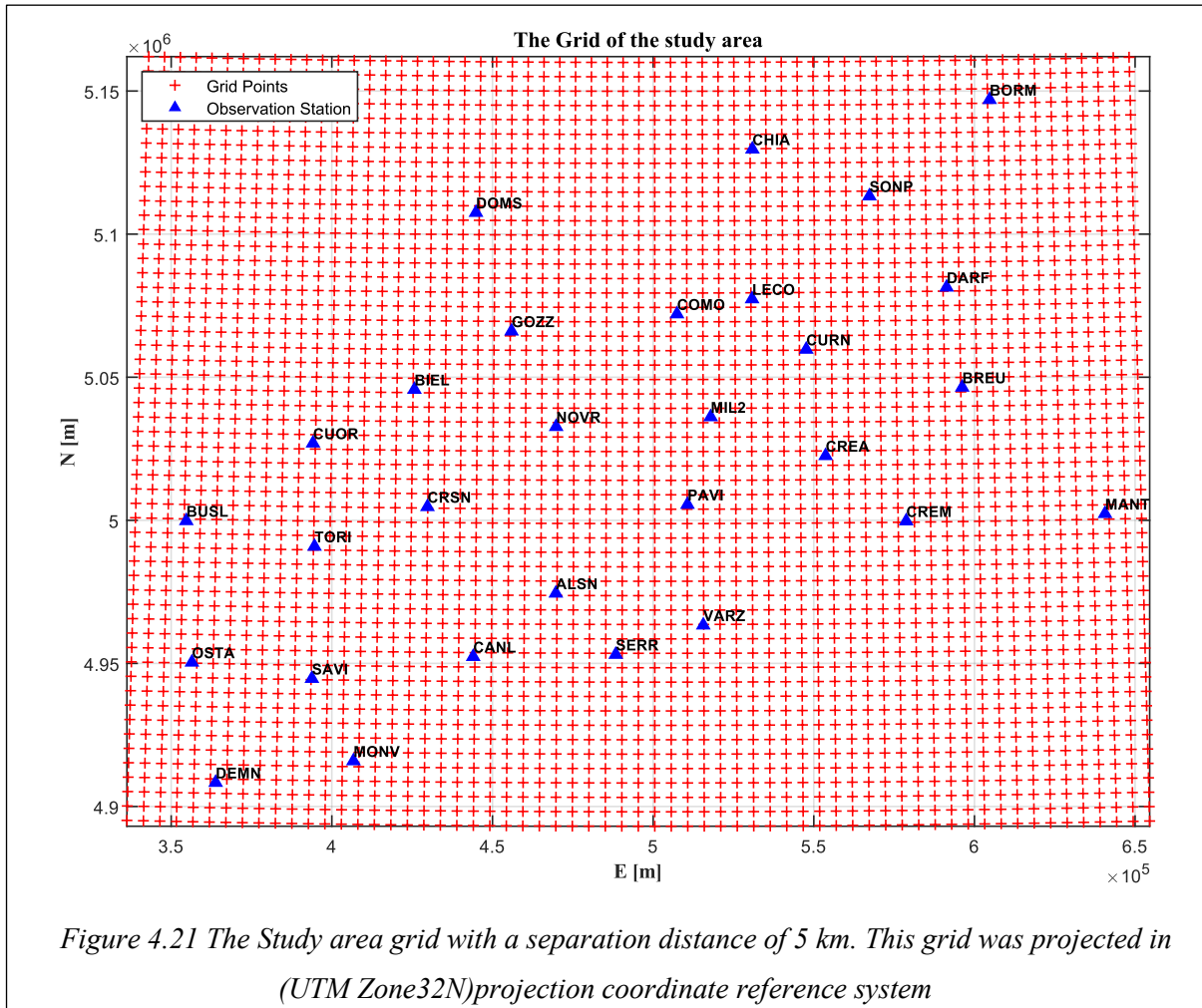
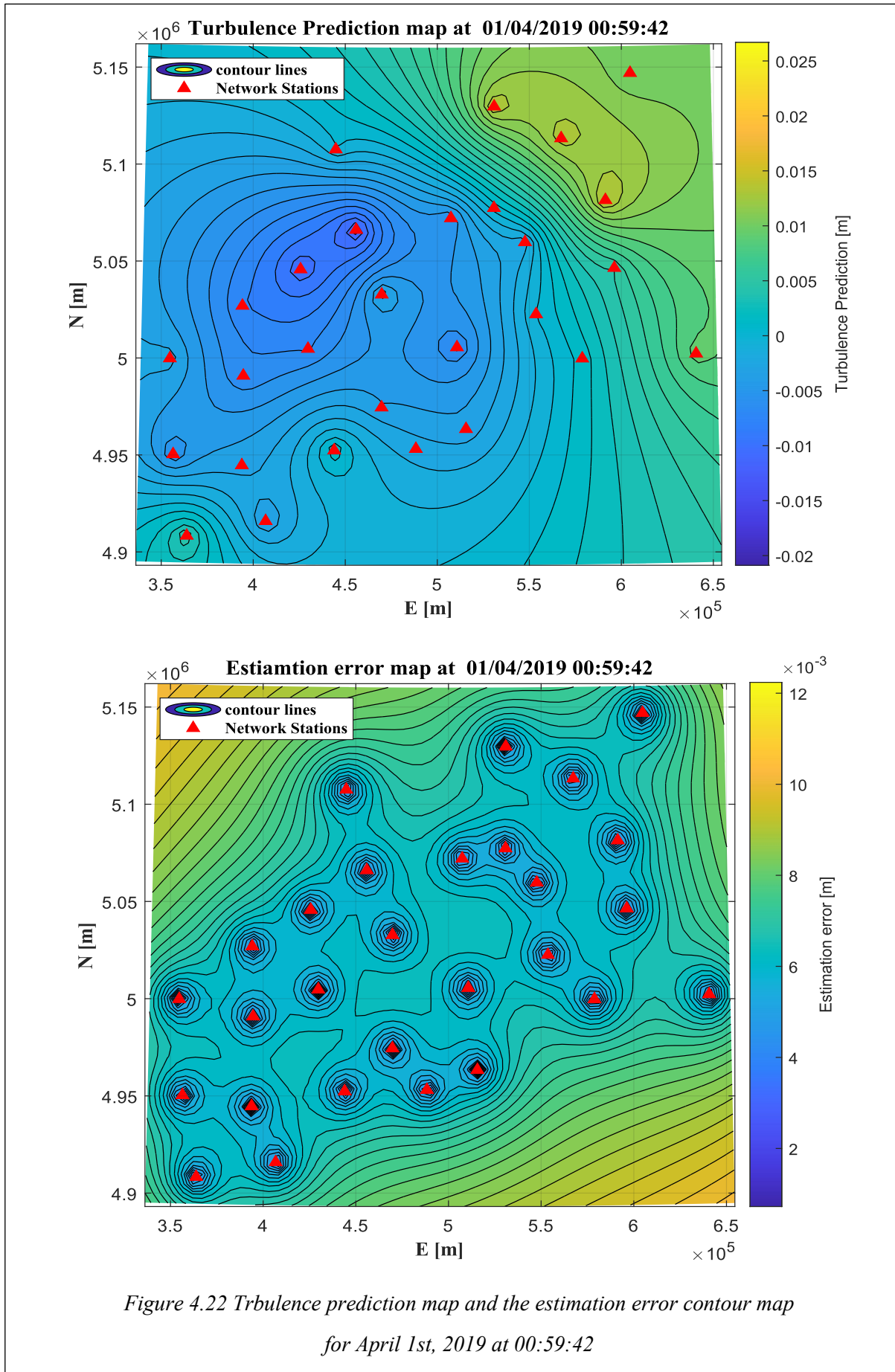


Figure 4.21 The Study area grid with a separation distance of 5 km. This grid was projected in (UTM Zone32N)projection coordinate reference system

Solving the kriging system represents in Equation [3.51], we can estimate not only the turbulence value at each grid point but also the estimation error at that point. The flow chart in Figure[4.25] illustrates the ordinary kriging system we have implemented.



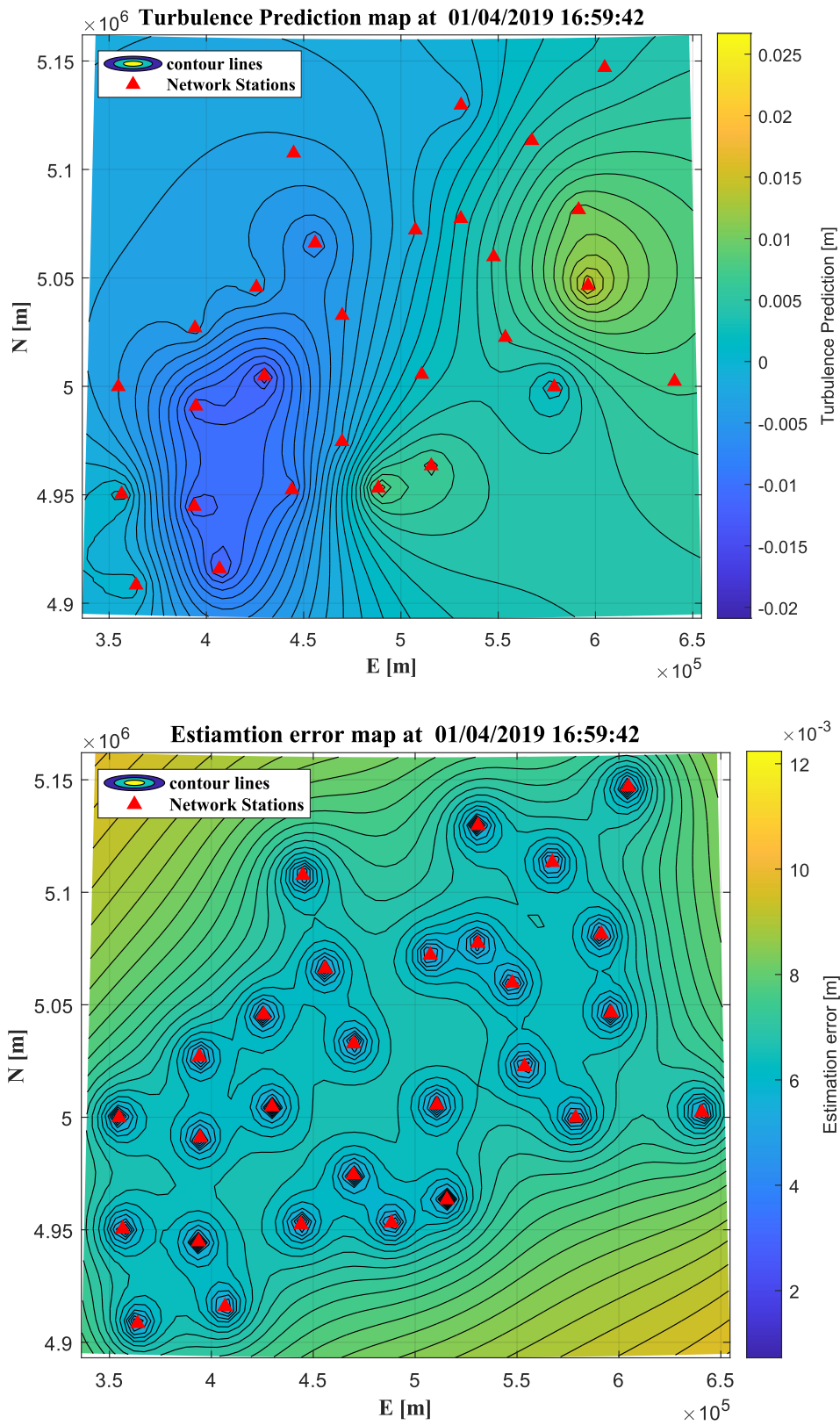


Figure 4.24 Trbulence prediction map and the estimation error contour map for April 1st, 2019 at 18:59:42

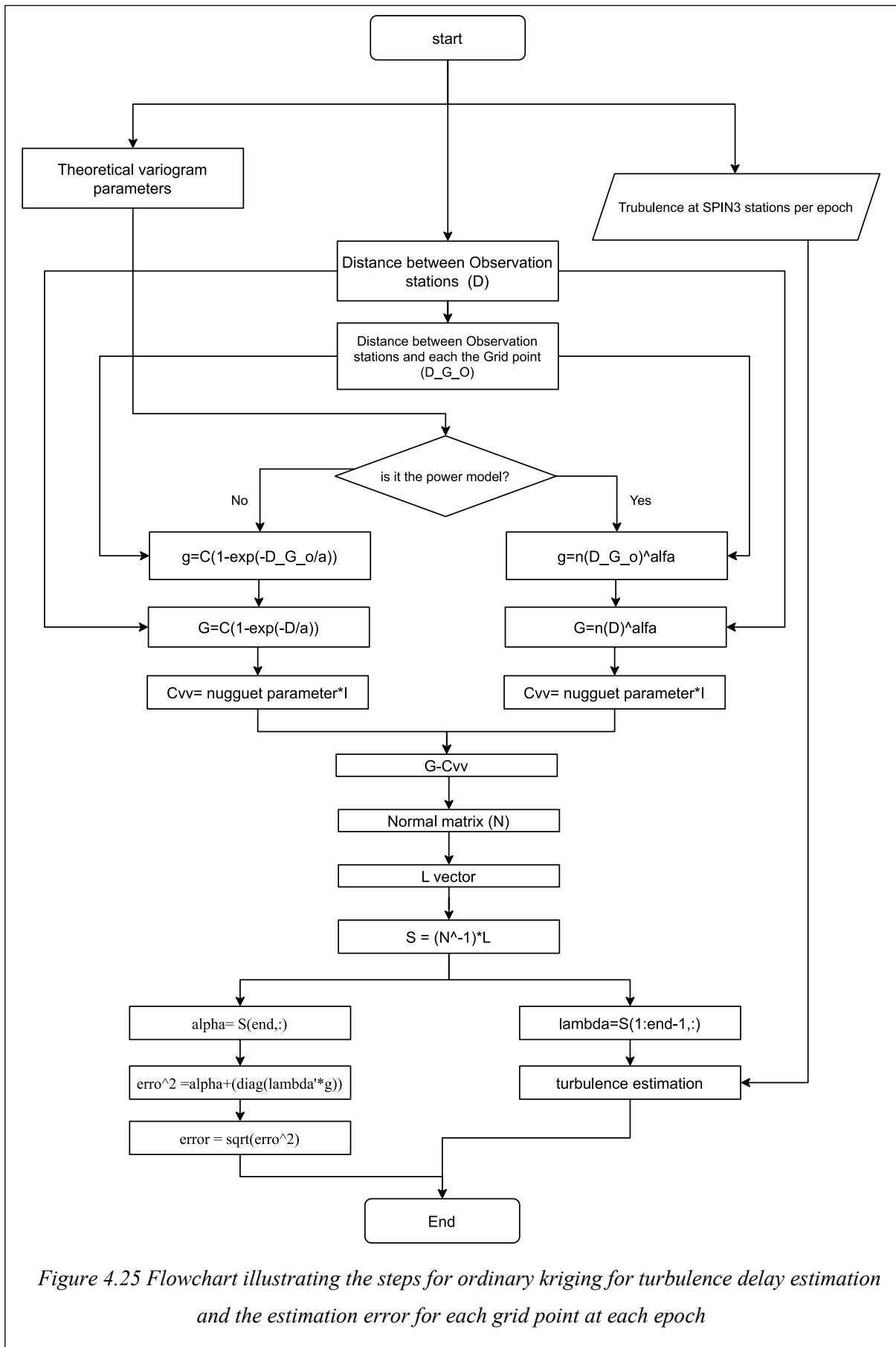


Figure 4.25 Flowchart illustrating the steps for ordinary kriging for turbulence delay estimation and the estimation error for each grid point at each epoch

#### 4.1.3.4 Data Validation

After obtaining turbulence predictions, it is essential to validate the estimation approach employed. In this study, we utilize the *Leave-One-Out Cross-Validation* technique (*LOOCV*), a well-established method known for its nearly unbiased estimation (Efron, 1983). The fundamental concept behind LOOCV involves excluding one measuring point from the dataset and reestimating it using the observations from the remaining points (Longman et al., 2019).

Within the context of this case study, LOOCV was conducted from 01/01/2019 at 23:59:12 to 30/12/2020 at 23:59:12. During this period, a total of 28 SPIN3 Stations was available, and the temporal resolution was *1 day*. For each epoch within this period, one station will be temporarily removed, and stochastic prediction procedures and ordinary kriging will be employed to estimate the turbulence at the removed station. The estimated turbulence value will then be compared to the observed turbulence value at the same epoch for validation. This procedure will be repeated for all stations at each epoch and all epochs.

In order to evaluate the performance of our procedure, we compute the discrepancy between the predicted turbulence delay and the observed turbulence values at each station. Subsequently, we apply Equation [4-4] to determine the bias, Equation [4-5] to calculate the Root Mean Square Error (RMSE) of this difference also Equation [4-6] to determine the standard deviation (STD) of the difference. These evaluations are conducted individually for each station as well as for each epoch.

$Bias = \frac{1}{N} \sum_{i=1}^N (T_{LOOCV_i} - T_{Observed_i})$	4-4
$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (T_{LOOCV_i} - T_{Observed_i})^2}$	4-5
$STD = \sqrt{\frac{1}{N} \sum_{i=1}^N ((T_{LOOCV_i} - T_{Observed_i}) - \mu_{(T_{LOOCV} - T_{Observed})})^2}$	4-6

**where:**

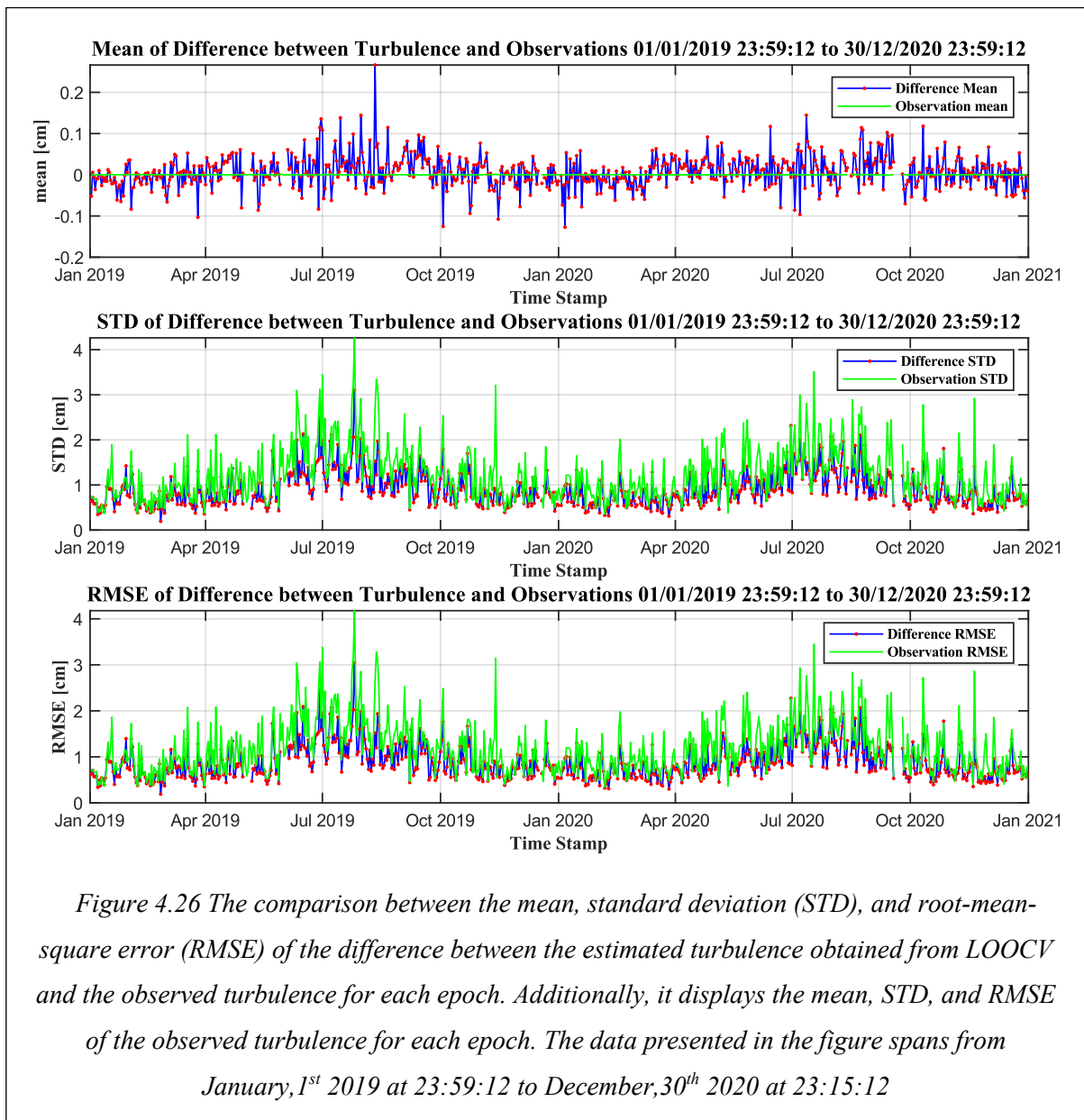
$T_{LOOCV_i}$ : is the turbulence delay estimation from LOOCV.

$T_{Observed_i}$ : is the turbulence delay estimation after removing the stratified part from ZTD.

N: When estimating the bias, RMSE, and STD for each epoch, the value of N corresponds to the number of epochs. On the other hand, when estimating the bias, RMSE, and STD for each station, N represents the number of stations. In the case of estimating the bias, RMSE, and STD for each epoch, N denotes the total number of epochs

$\mu_{(T_{LOOCV} - T_{Observed})}$ : In the case of estimating the standard deviation (STD) for each epoch, this mean is computed by averaging the differences at a particular epoch across all stations. On the other hand, when estimating the STD for each station, the mean is obtained by averaging the differences for all epochs at that specific station.

Upon evaluating the mean, standard deviation (STD), and root-mean-square error (RMSE) for the difference between estimated and observed turbulence for each epoch Figure [4.26], distinct variations in the standard deviation (STD) and root-mean-square error (RMSE) during the spring and summer seasons. As anticipated, these observations align with our expectations, as these two seasons are characterized by elevated levels of atmospheric water vapor, which subsequently intensify turbulence. The pronounced infatuation of STD and RMSE in the depicted seasons attests to the heightened turbulence levels caused by the increased water vapor content.



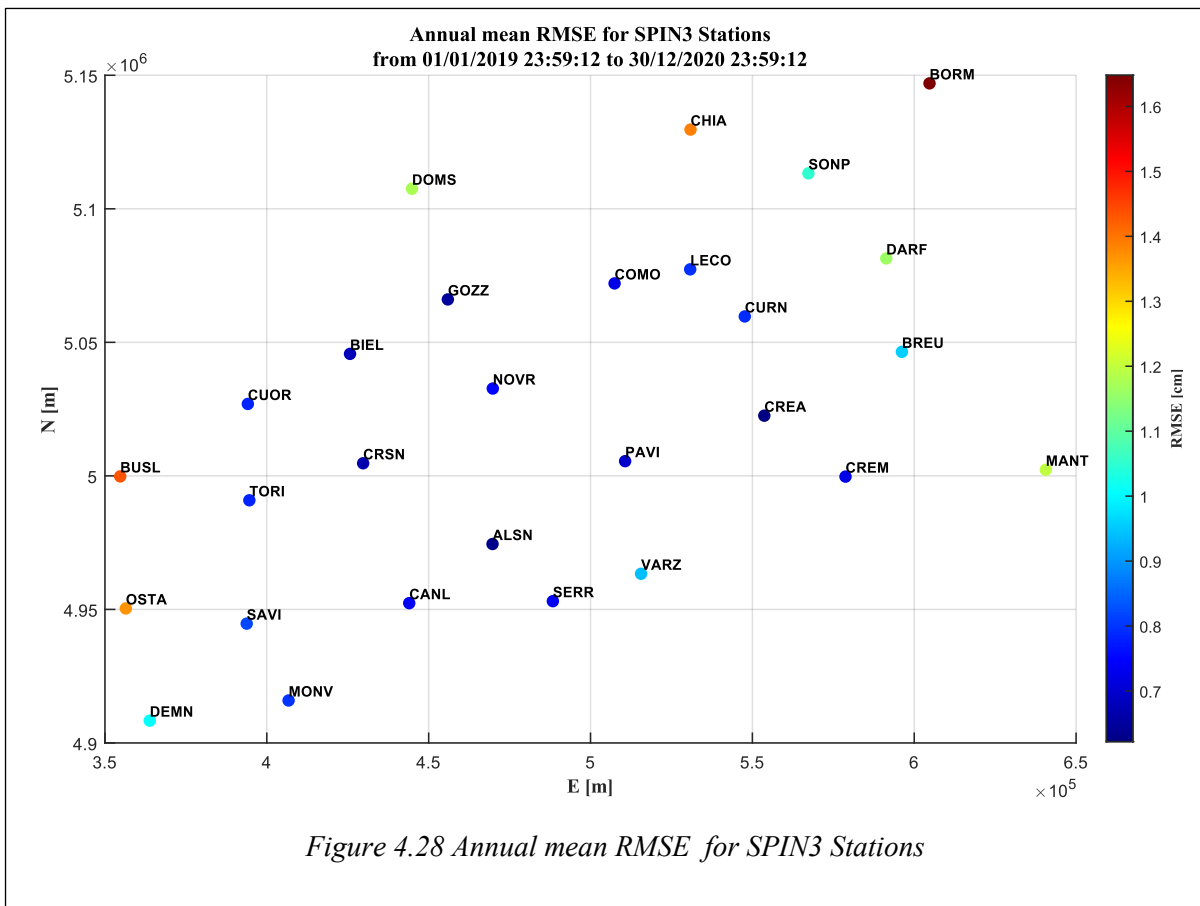
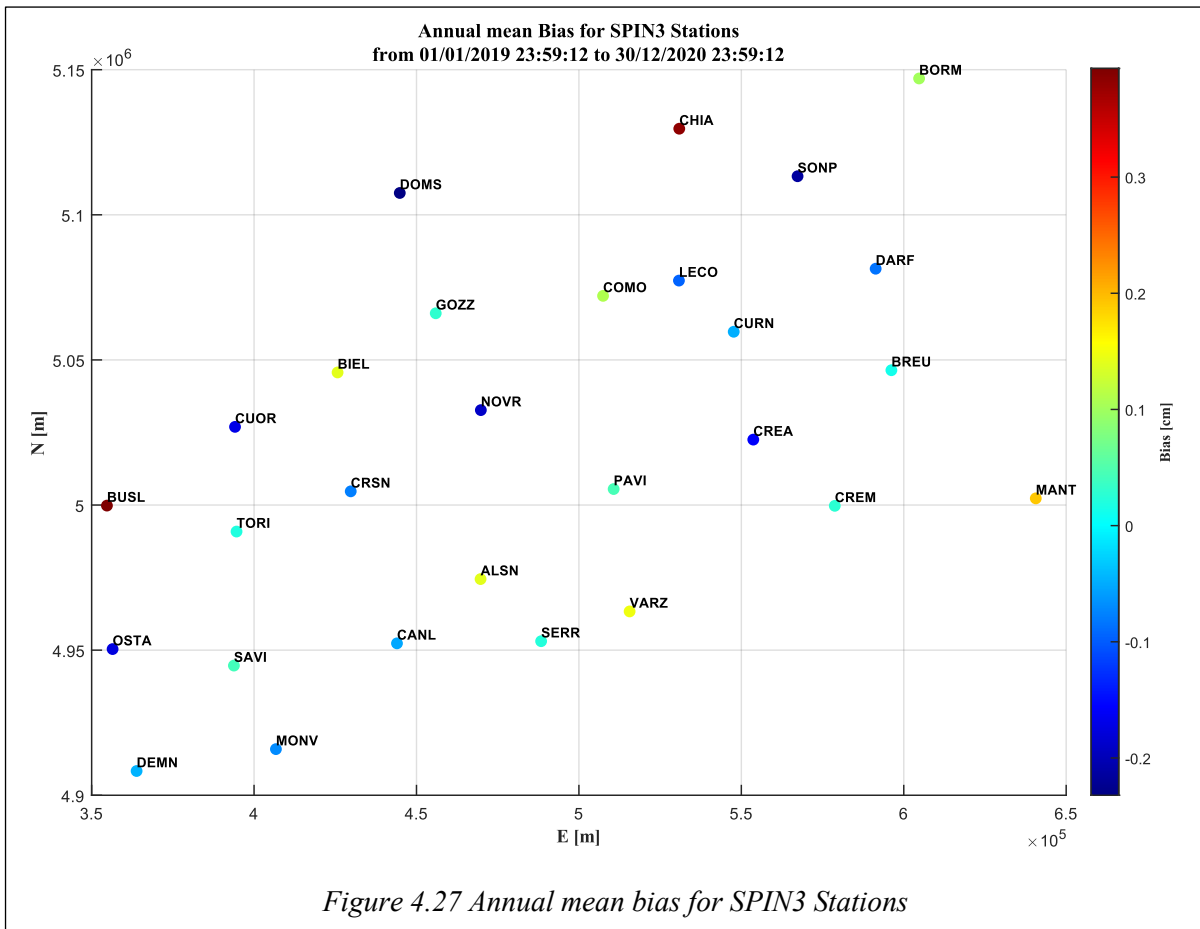
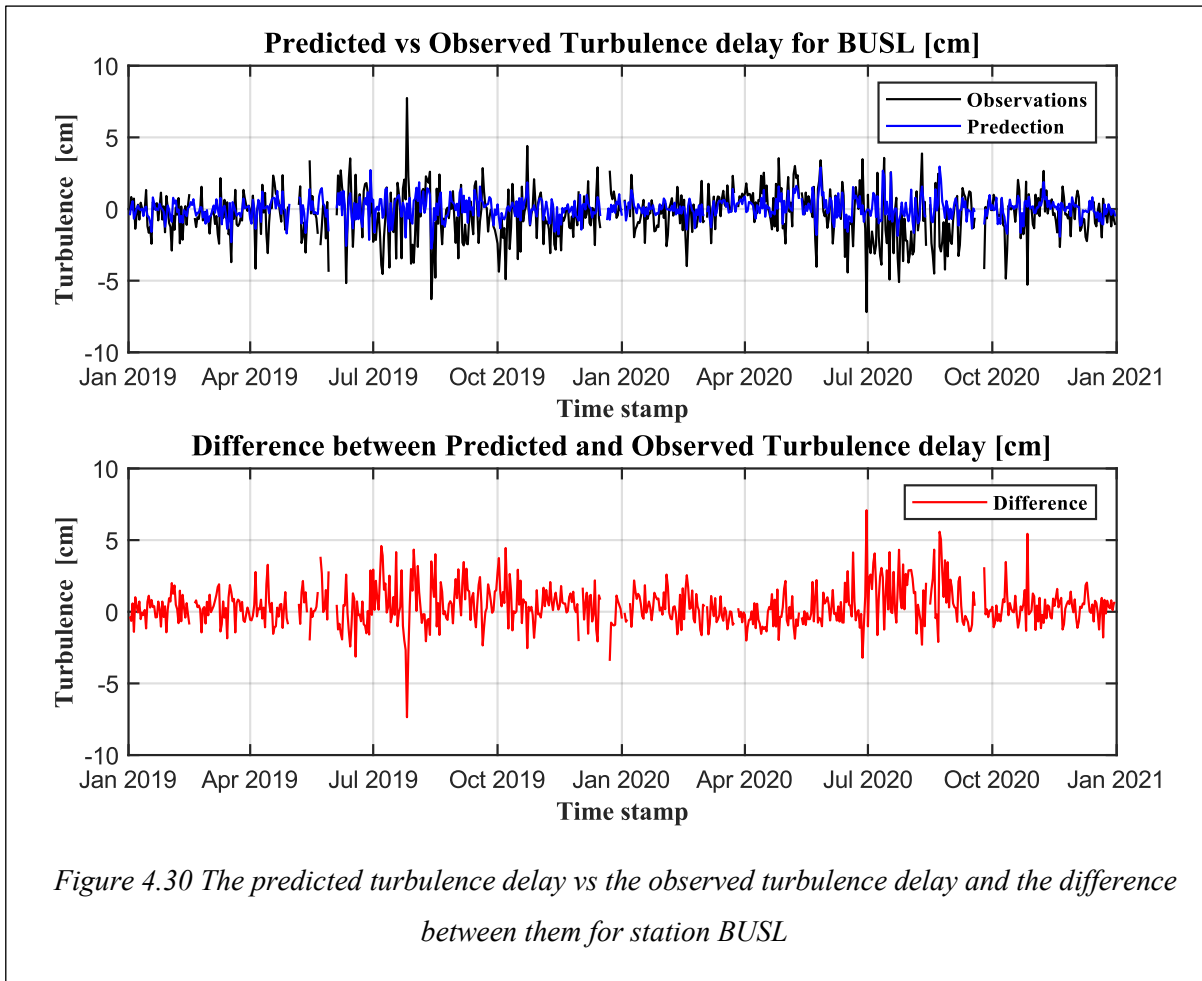
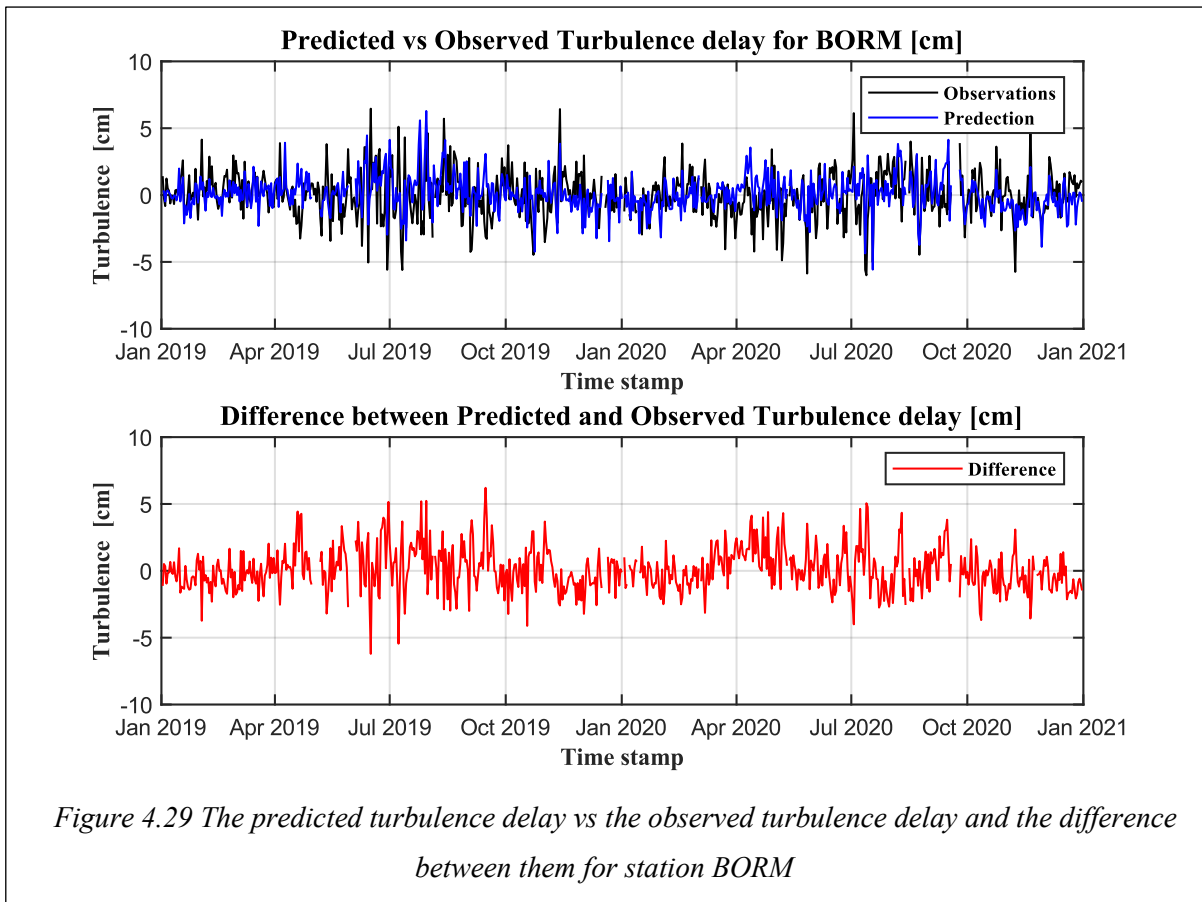


Table 4-3 The Bias and the RMSE in cm for 28 SPIN3 station

<b>Station</b>	<b>Bias(cm)</b>	<b>RMSE(cm)</b>
<i>ALSN</i>	0.1412	0.6333
<i>BIEL</i>	0.1412	0.6786
<i>BORM</i>	0.1017	1.6480
<i>BREU</i>	0.0124	0.9567
<i>BUSL</i>	0.3935	1.4317
<i>CANL</i>	-0.0521	0.7359
<i>CHIA</i>	0.3817	1.3909
<i>COMO</i>	0.1084	0.7262
<i>CREA</i>	-0.1569	0.6220
<i>CREM</i>	0.0292	0.7249
<i>CRSN</i>	-0.0773	0.6662
<i>CUOR</i>	-0.1700	0.7829
<i>CURN</i>	-0.0466	0.7934
<i>DARF</i>	-0.0849	1.1608
<i>DEMN</i>	-0.0442	1.0081
<i>DOMS</i>	-0.2321	1.1771
<i>GOZZ</i>	0.0299	0.6462
<i>LECO</i>	-0.0929	0.7973
<i>MANT</i>	0.1920	1.1960
<i>MONV</i>	-0.0699	0.7991
<i>NOVR</i>	-0.1883	0.7530
<i>OSTA</i>	-0.1743	1.3710
<i>PAVI</i>	0.0448	0.6943
<i>SAVI</i>	0.0423	0.8219
<i>SERR</i>	0.0224	0.7354
<i>SONP</i>	-0.2136	1.0510
<i>TORI</i>	0.0217	0.7859
<i>VARZ</i>	0.1475	0.9406

From the analysis presented in Figure [4.27] and Figure [4.28], as well as Table [4-3], it is evident that certain stations, such as *BORM* and *CHIA*, exhibit high RMSE values. This can be attributed to two main factors. Firstly, these stations are situated in the Albes area, which is known for its challenging atmospheric conditions and increased turbulence. Secondly, the lack of nearby stations limits the effectiveness of interpolation methods in accurately estimating turbulence at these locations. Additionally, stations like *BUSL* and *OSTA*, located in mountainous regions, also demonstrate higher RMSE values due to the complex terrain and associated turbulence patterns.





The comprehensive statistical evaluation of the SPIN3 network, covering the period from January 1, 2019, at 23:59:12 to December 30, 2020, at 23:15:12, is summarized as follows:

*Table 4-4 The statistics for SPIN3 GNSS network for the period from January,1st 2019 at 23:59:12 to December,30th 2020 at 23:15:12*

	<i>Annual mean (cm)</i>	<i>Minimum value (cm)</i>	<i>Maximum value (cm)</i>
<b><i>STD</i></b>	0.9011		3.1055
<b><i>RMSE</i></b>	0.885	0.1898	3.0428

## 4.2 Case Study2

### 4.2.1 Study Area and GNSS networks

The second case study of our thesis was conducted in Poland, which exhibits distinctive topography characterized by lowlands in the central and northern regions, predominantly occupied by the expansive Polish Plain. The southern part of Poland displays a more diverse topography, with the presence of the Sudeten and Carpathian Mountains. The study was conducted on August 11<sup>th</sup>, 2017, in collaboration with (*GReD - Geomatics Research & Development s.r.l.*, n.d.), who provided us with GNSS data from three networks distributed in the area with 278 GNSS stations. These networks include the *ASG-EUPOS*, a state network, as well as the commercial networks *TPI NET pro* and *VRSNet.pl*. The distribution of these networks can be observed in Figure [4-31].

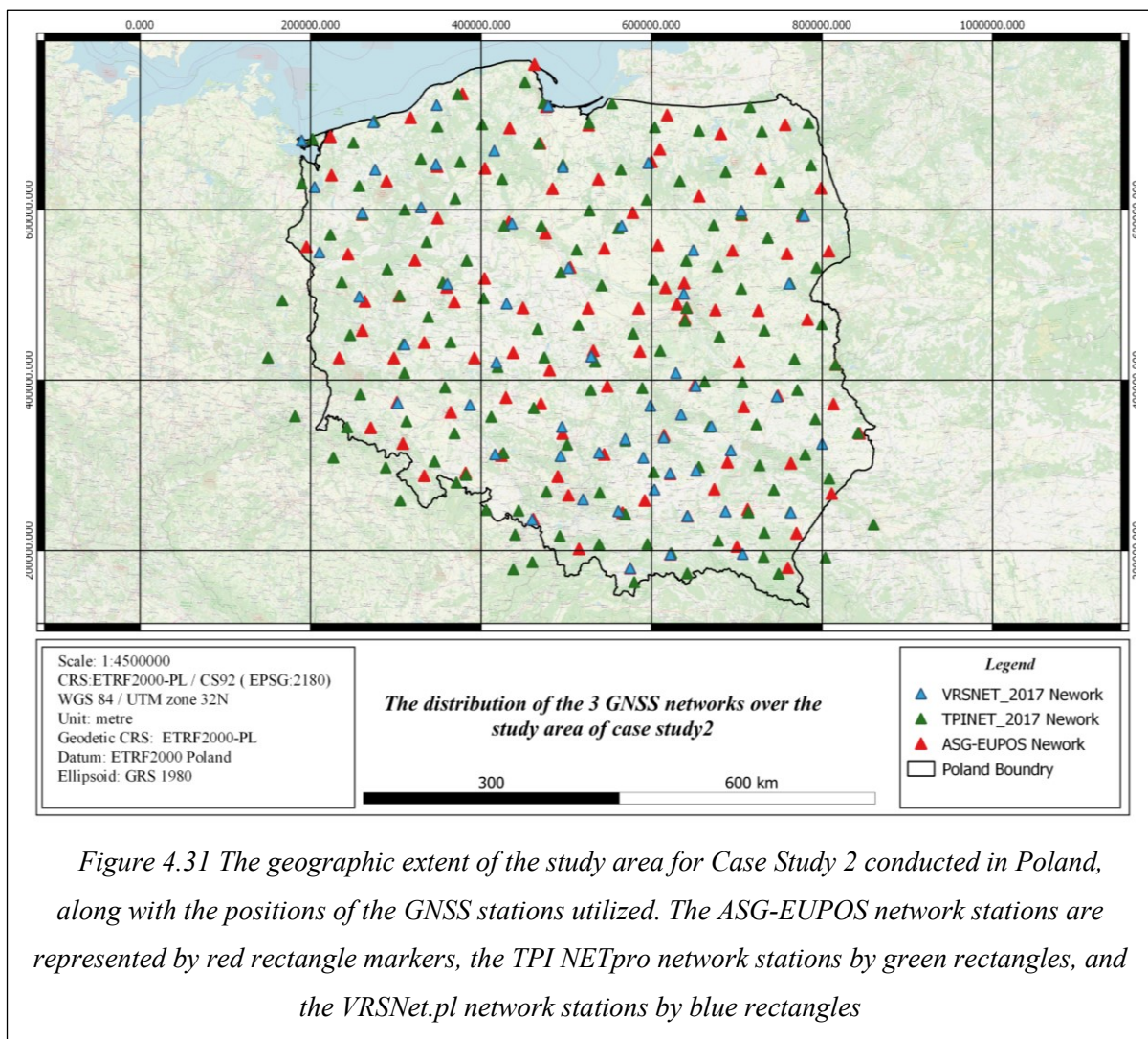


Figure 4.31 The geographic extent of the study area for Case Study 2 conducted in Poland, along with the positions of the GNSS stations utilized. The ASG-EUPOS network stations are represented by red rectangle markers, the TPI NETpro network stations by green rectangles, and the VRSNet.pl network stations by blue rectangles

The ASG-EUPOS network, initiated under the auspices of the Head Office of Geodesy and Cartography in 2008, encompasses a total of 104 stations within the borders of Poland and an additional 24 stations located across international borders. The TPI NETpro network, a privately operated system, has been in operation since 2012 and comprises 121 stations within Poland, with an additional 15 stations positioned outside the country. Similarly, the VRSNet.pl network, established in 2012, consists of 81 stations situated along the border and 6 stations positioned beyond the national borders. These networks collectively provide an average spatial density of approximately 1120 square meters per station (Nykiel et al., 2019).

The processing of the three networks (ASG-EUPOS, TPI NETpro, and VRSNet.pl) was performed using the Precise Point Positioning (PPP) method in the Bernese GNSS software version 5.2. The processing strategy involved utilizing the Code final products and considering only GPS and GLONASS observations. A total of 278 stations with a 30-second observation rate were included in the analysis. Table [4-5] provides a summary of the processing strategy employed for these networks (Nykiel et al., 2019).

*Table 4-5 the processing strategy employed for the three networks  
(ASG-EUPOS, TPI NETpro, and VRSNet.pl)*

<b>Data</b>	30-second daily RINEX from 278 GNSS stations
<b>Satellite systems</b>	GPS + GLONASS
<b>Method</b>	PPP
<b>Software</b>	Bernese GNSS Software ver. 5.2
<b>Position estimation interval</b>	30-second
<b>A priori ZHD</b>	VMF1
<b>Wet tropospheric delays</b>	Estimated with 5-minute interval using VMF1 wet
<b>mapping function</b>	Relative constraint: 2 mm
<b>Tropospheric gradients</b>	Estimated with 15-minute interval using Chen & Herring mapping function. Relative constraint: 0.2 mm
<b>Cut-off angle</b>	5°
<b>Ionospheric delay</b>	1st order delay eliminated using ionospheric free linear combination; High order (2nd and 3rd) effects also included.
<b>Orbits and clocks</b>	CODE Final 5-min. ephemerides and 30-sec. clocks
<b>Reference frame</b>	GS14
<b>Antenna models</b>	Type mean (IGS14)

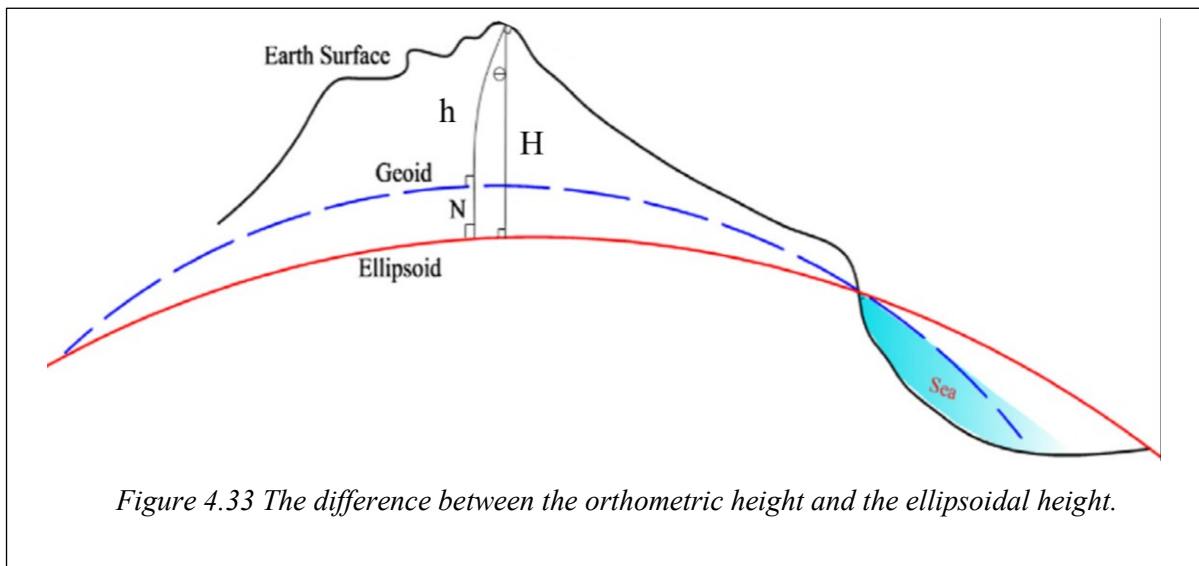


### 4.2.2.1 Data Management

For each network, we have ZTD (Zenith Total Delay) observations for each station during the specified period. Additionally, we possess the positional information for each station in terms of Earth-Centered, Earth-Fixed (ECEF) coordinates (X, Y, Z). In order to align these coordinates with the local projected coordinate system applicable to Poland, namely ETRF2000-PL / CS92, a series of transformations are required.

**Note:** The computations in this case study are exclusively based on data from the ASG-EUPOS network.

To achieve this transformation, we initially convert the ECEF coordinates into geodetic coordinates ( $\varphi, \lambda, H$ ). Subsequently, to obtain the desired Orthometric height ( $h$ ) from the ellipsoidal height ( $H$ ), we employ the official Earth Gravitational Model (EGM2008). The conversion is facilitated by employing Equation [4-7].



$$h = H - N$$

4-7

**where:**

**h:** is the orthometric height.

**H:** is the ellipsoidal height.

**N:** the Geoid undulation.

Once we have obtained the coordinates in the local projected coordinate system (E, N, h), our procedure commences by examining the missing epochs on August 11<sup>th</sup>, 2017, spanning from 00:00:12 to 23:59:12 and assigning “NaN” to them following the methodology depicted in Figure [4.8]. Subsequently, we assess the temporal coverage for all stations, as depicted in Figure [4.34].

The next step involves examining the relationship between the elevation of each station and the mean Zenith Total Delay (ZTD) at that station, as depicted in Figure [4.35]. This analysis helps us understand the influence of elevation on ZTD values. Subsequently, we construct the least square system and estimate the turbulence delay at each epoch for all stations in the ASG-EUPOS network.

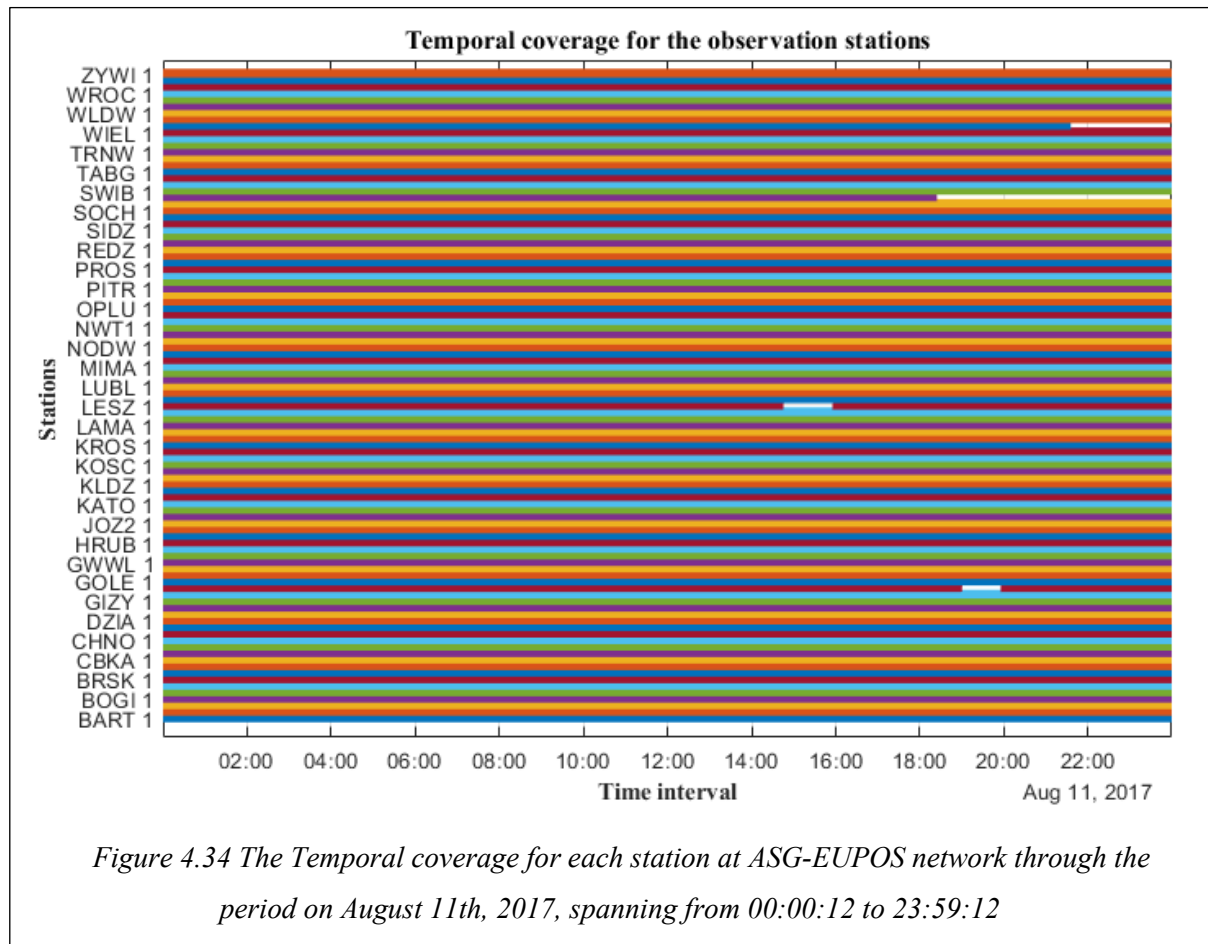
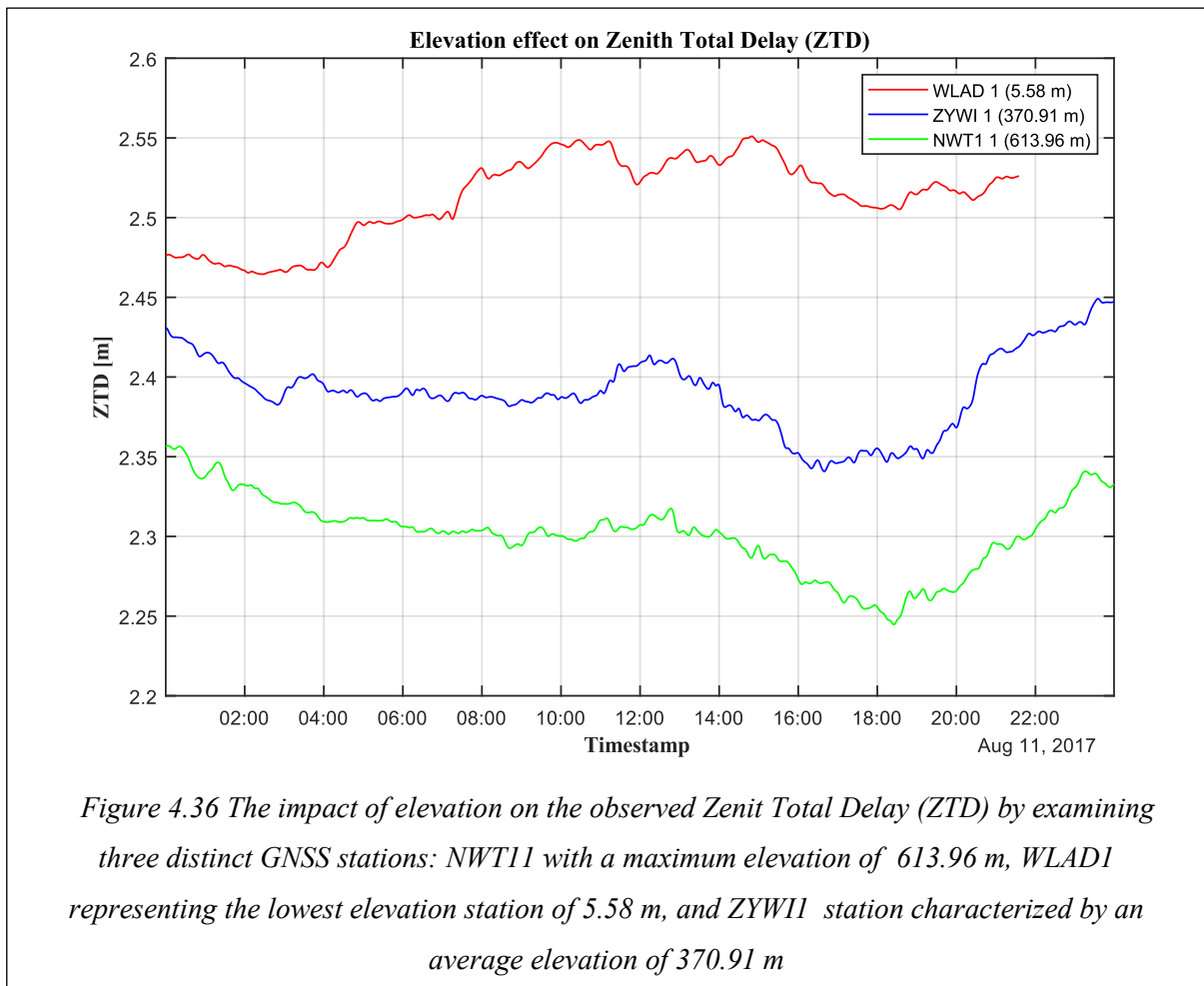
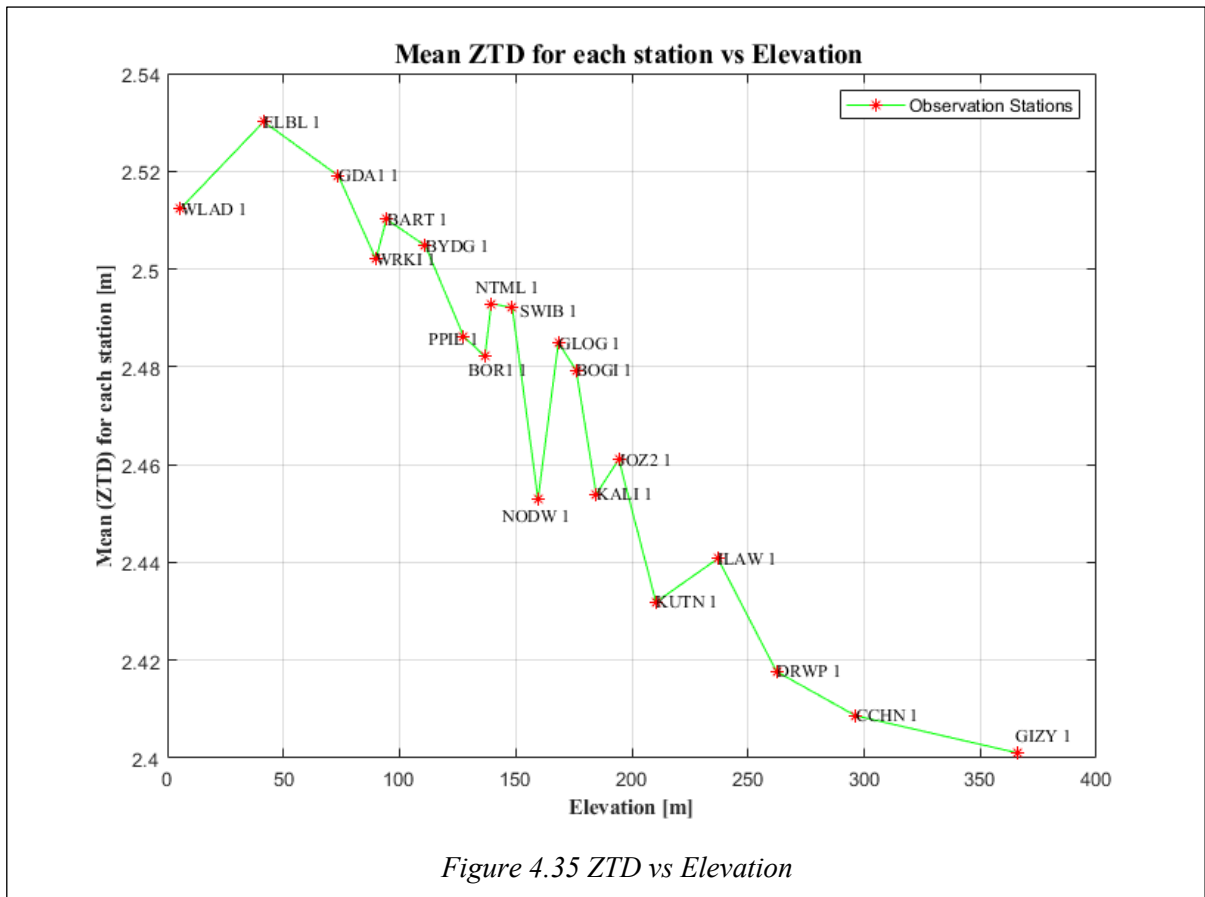


Figure 4.34 The Temporal coverage for each station at ASG-EUPOS network through the period on August 11th, 2017, spanning from 00:00:12 to 23:59:12





### 4.2.2.2 Variogram Calculation

In this section, we initiate the stochastic prediction process by constructing variogram clouds with a temporal resolution of 1 hour, following Equation [3-32] as illustrated in Figure [4.37]. The next step involves calculating the experimental variogram, where we determine the lag distance for applying the classic estimator defined in Equation [3-33]. Based on the minimum distance of 12 km between any pair, we specify the lag distance to be 20 km, as depicted in Figure [4.38].

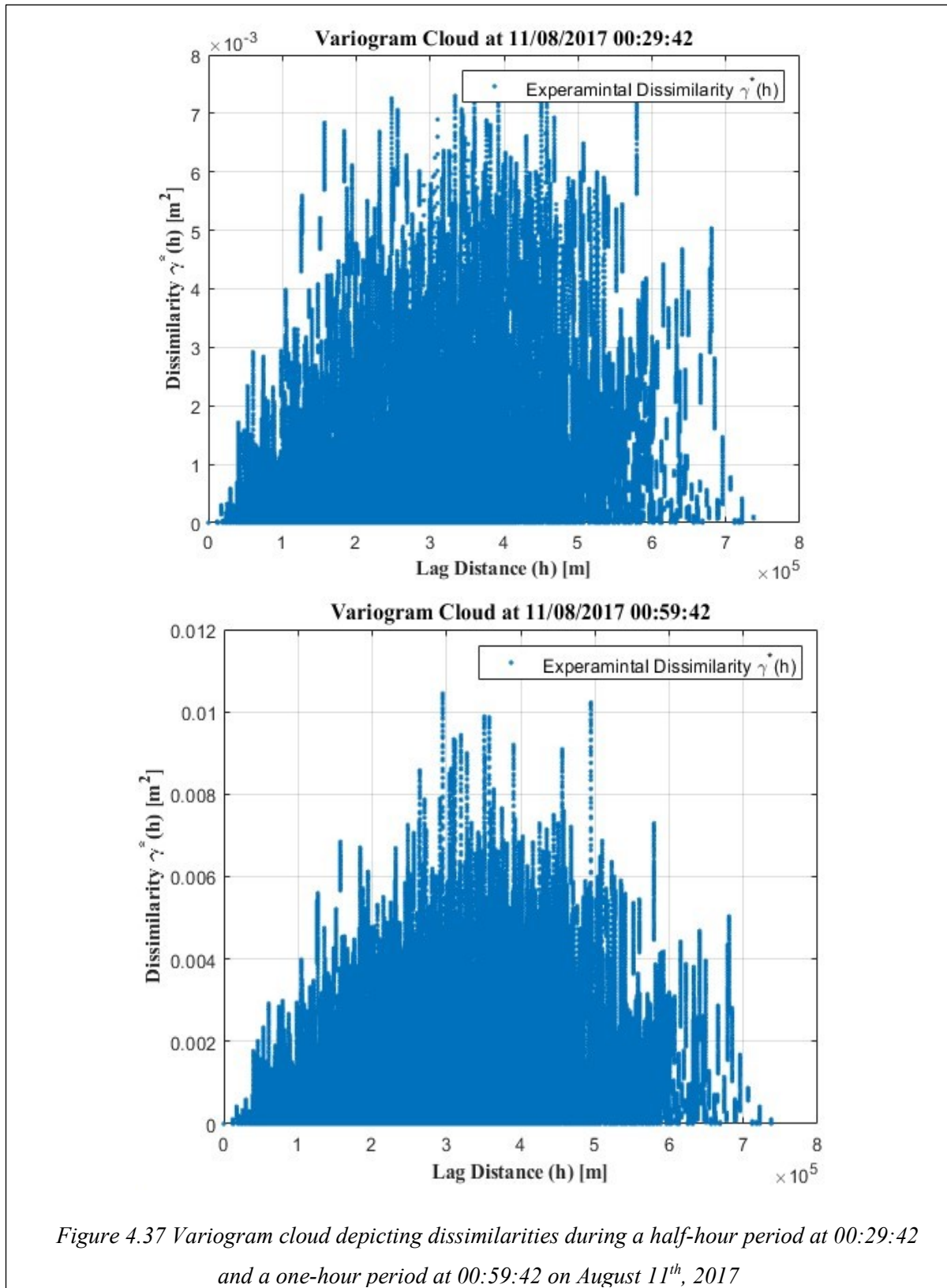


Figure 4.37 Variogram cloud depicting dissimilarities during a half-hour period at 00:29:42 and a one-hour period at 00:59:42 on August 11<sup>th</sup>, 2017

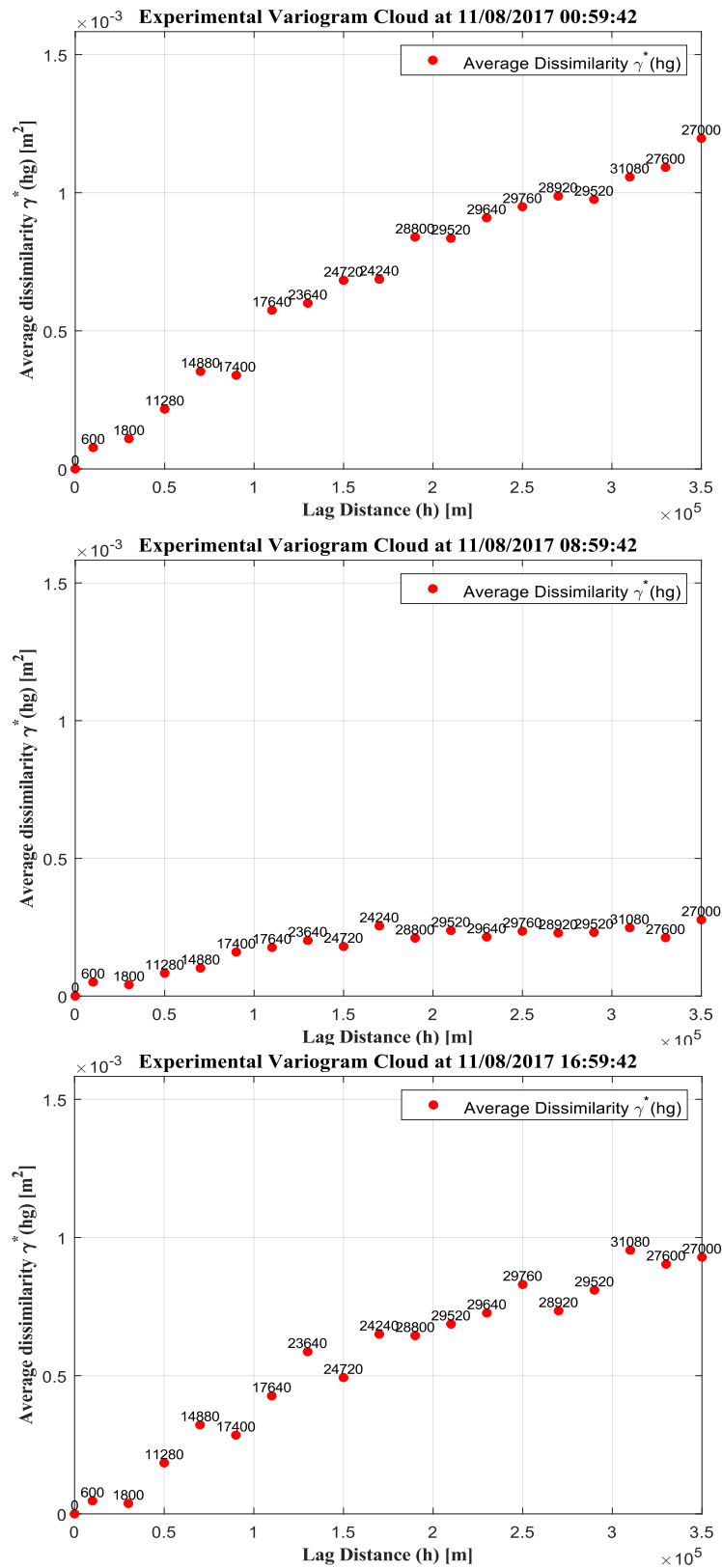


Figure 4.38 Experimental Variogram for August 11<sup>th</sup>, 2019 at three epochs (00:59:42, 08:59:42, and 18:59:42). The numbers above each point indicate the count of dissimilarities contributing to the average computation within each interval

The final step in the stochastic prediction procedure entails fitting either the exponential model or the power model to introduce the theoretical variogram. This is demonstrated in Figure [3.39]. and [3.40].

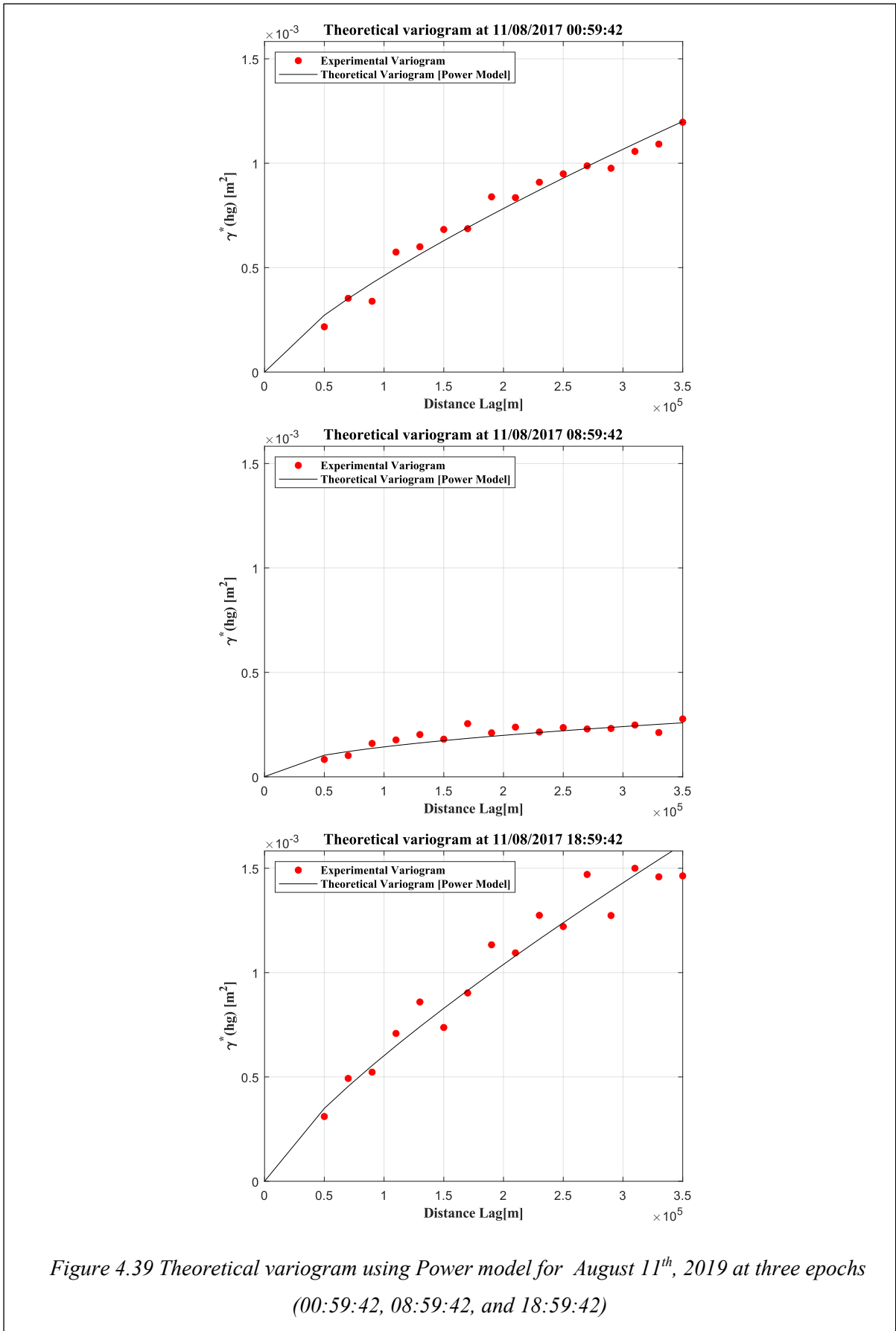


Figure 4.39 Theoretical variogram using Power model for August 11<sup>th</sup>, 2019 at three epochs (00:59:42, 08:59:42, and 18:59:42)

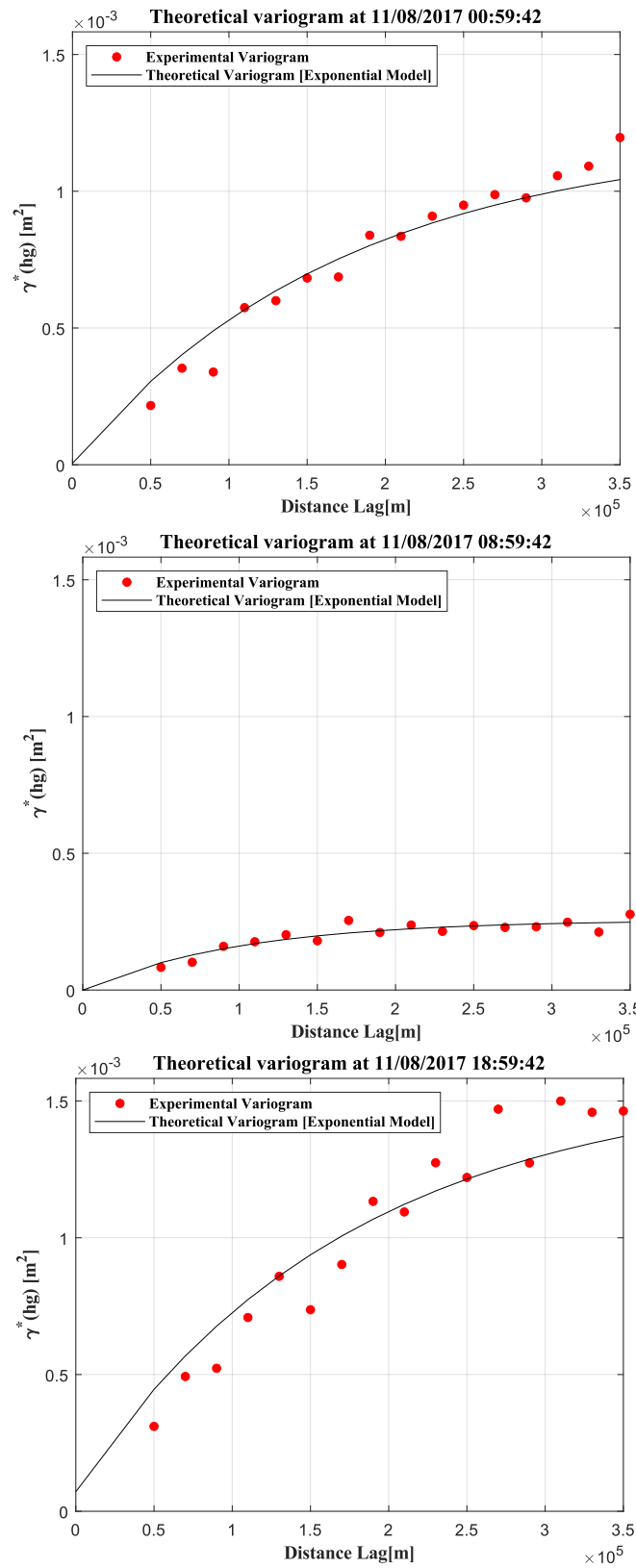
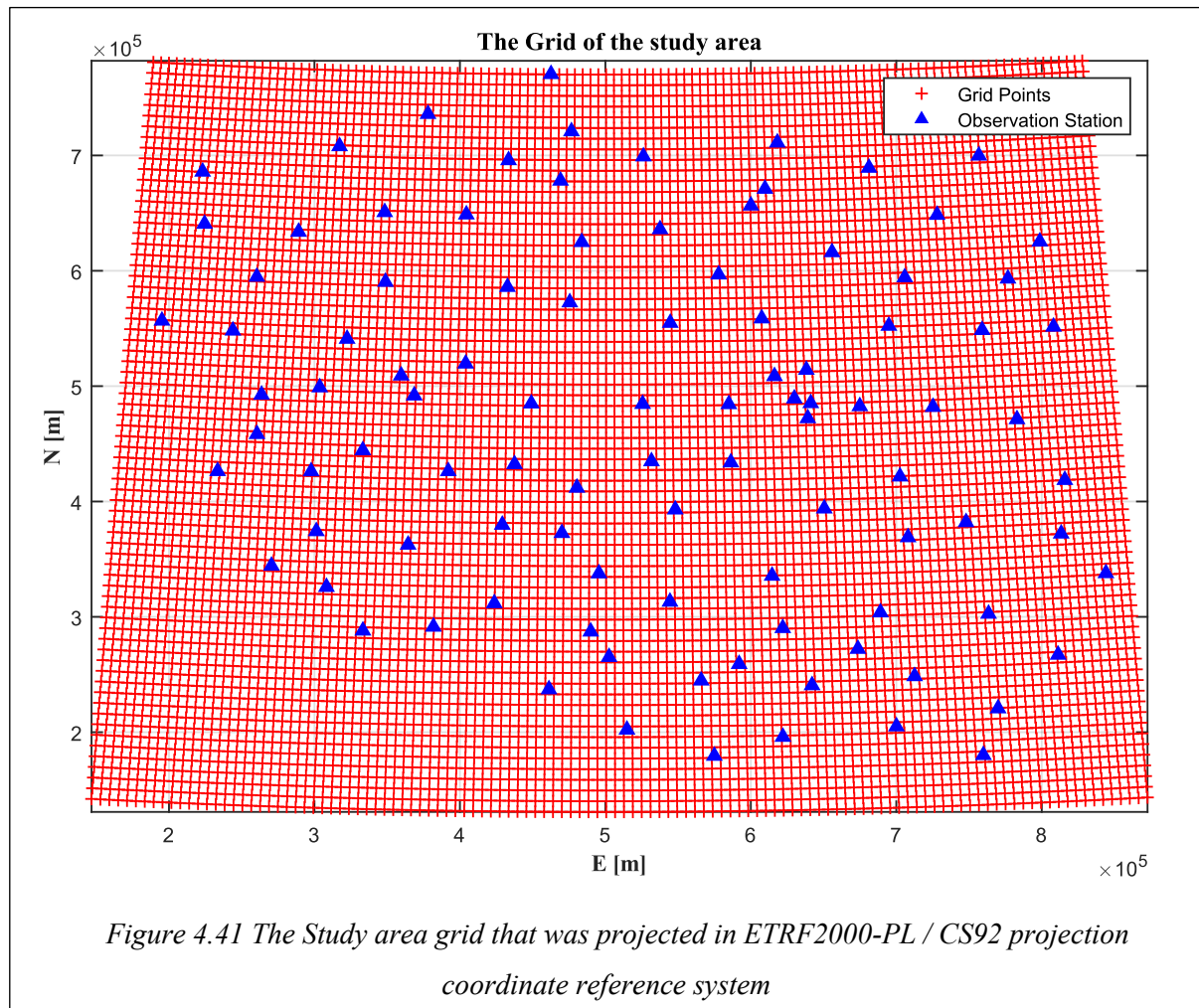


Figure 4.40 Theoretical variogram using Exponential model for August 11<sup>th</sup>, 2019 at three epochs (00:59:42, 08:59:42, and 18:59:42)

### 4.2.2.3 Kriging Interpolation

By adhering to the procedures outlined in section (4.1.3.3), the first step is to extract a uniform grid of the study area. This was achieved by extracting the grid from the Digital Elevation Model (DEM), as depicted in Figure [4.41].

Subsequently, by following the flowchart presented in Figure [4.25] and implementing the ordinary kriging system described in Equation [3.51], a series of turbulence prediction maps and their corresponding estimation error maps can be generated for any desired epoch. These maps are illustrated in Figures [4.42], [4.43], and [4.44].



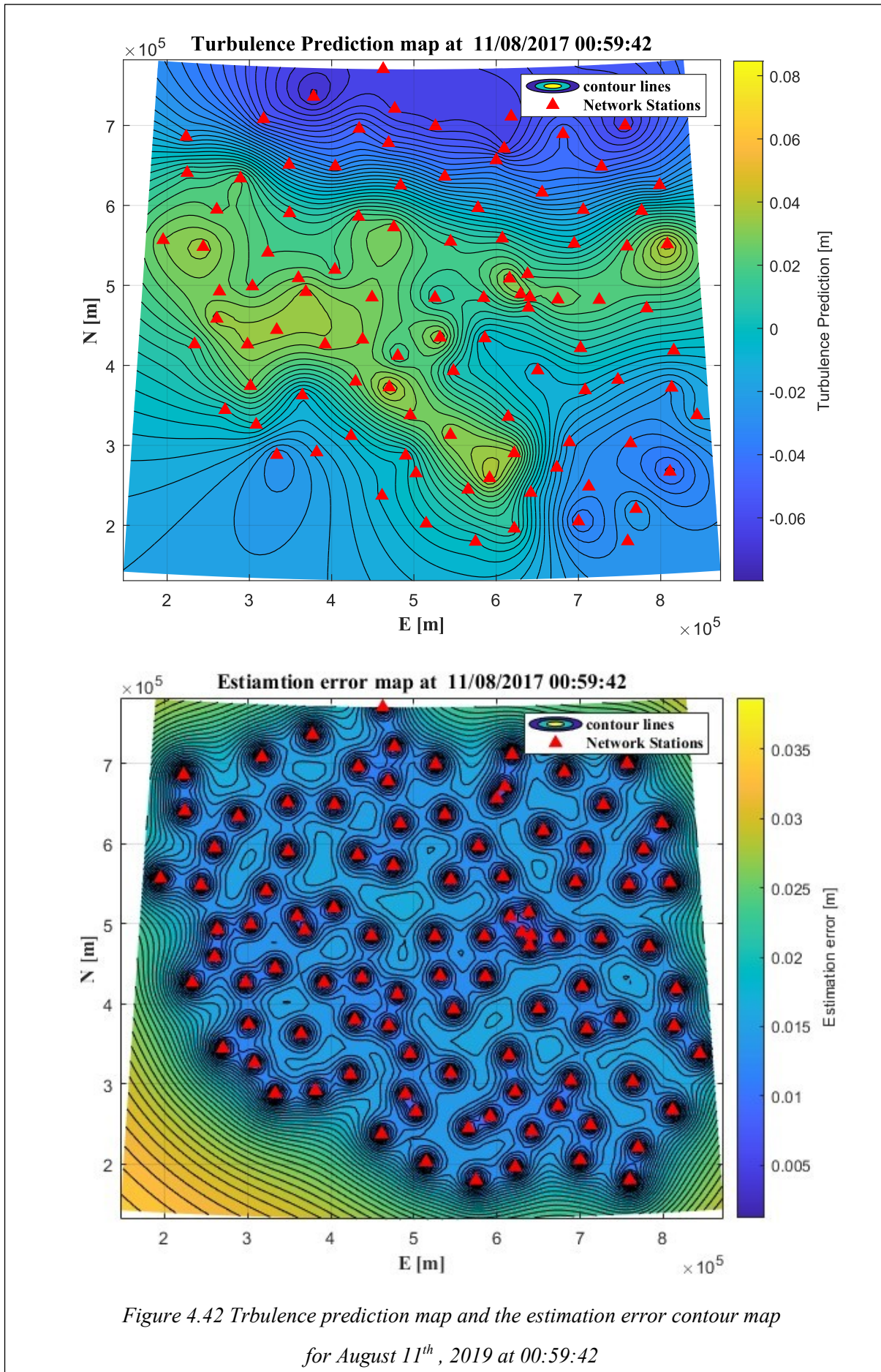


Figure 4.42 Trbulence prediction map and the estimation error contour map for August 11<sup>th</sup> , 2019 at 00:59:42

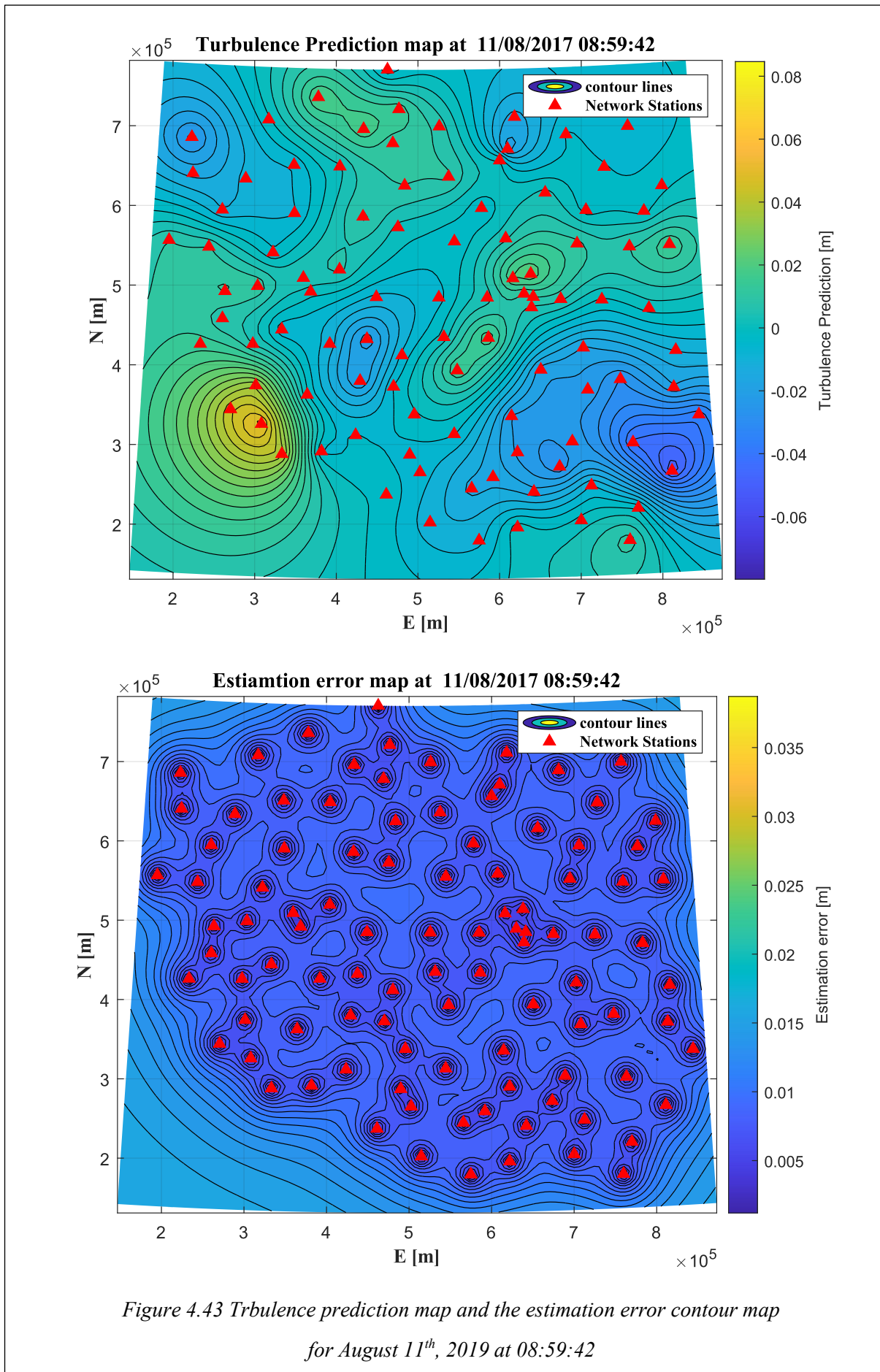


Figure 4.43 Trbulence prediction map and the estimation error contour map for August 11<sup>th</sup>, 2019 at 08:59:42

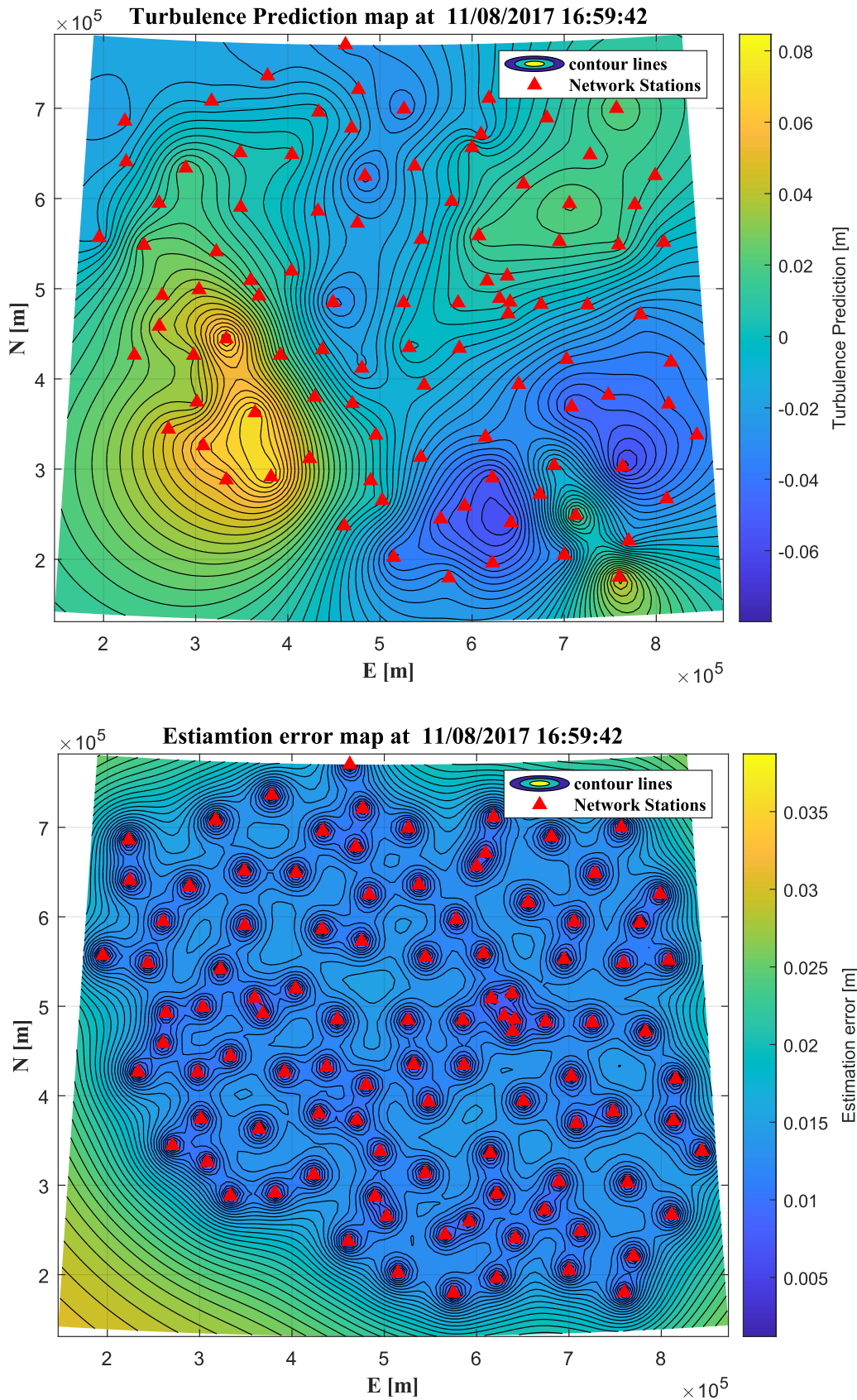


Figure 4.44 Trbulence prediction map and the estimation error contour map for August 11<sup>th</sup>, 2019 at 16:59:42



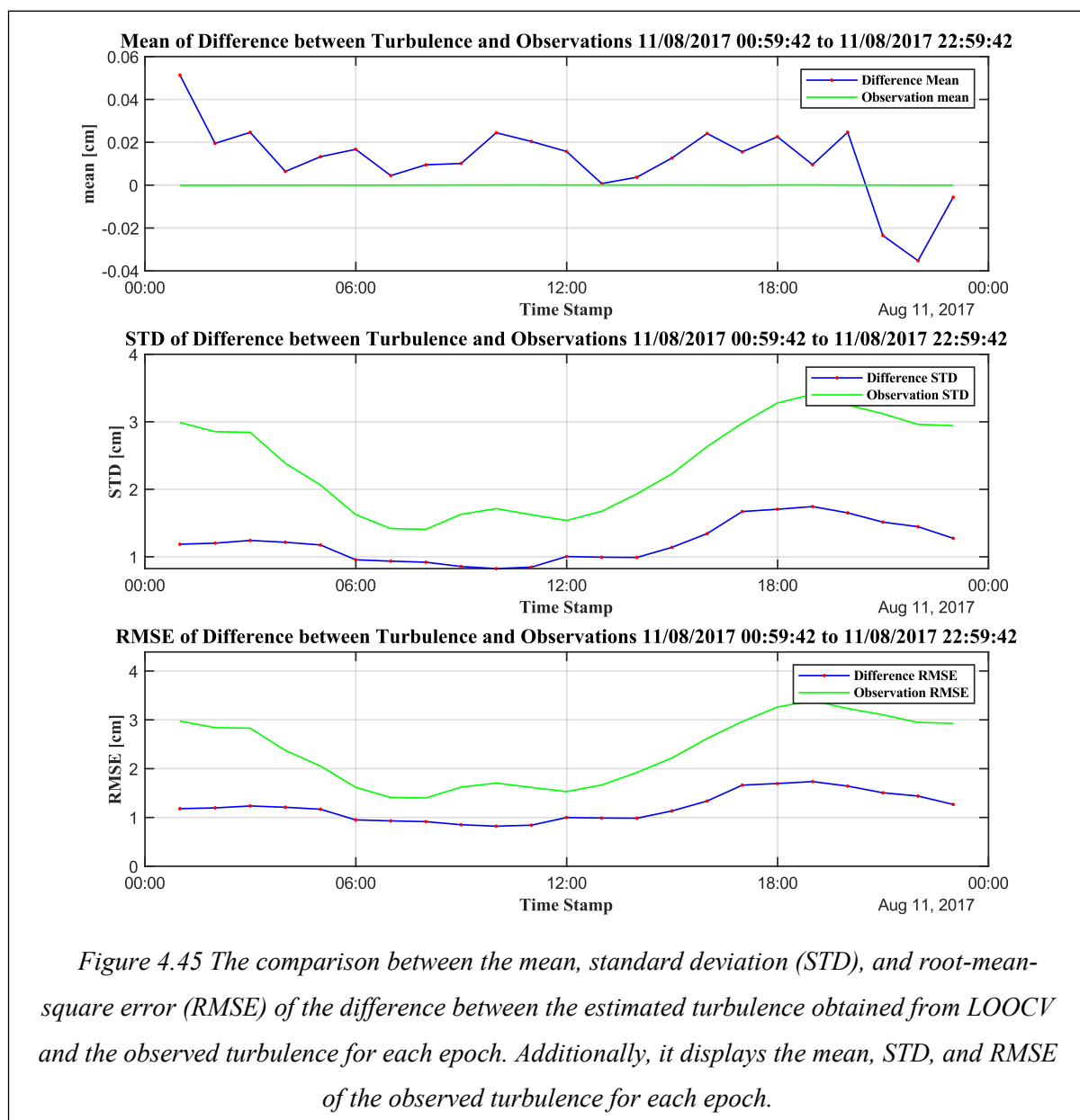
#### 4.2.2.4 Data Validation

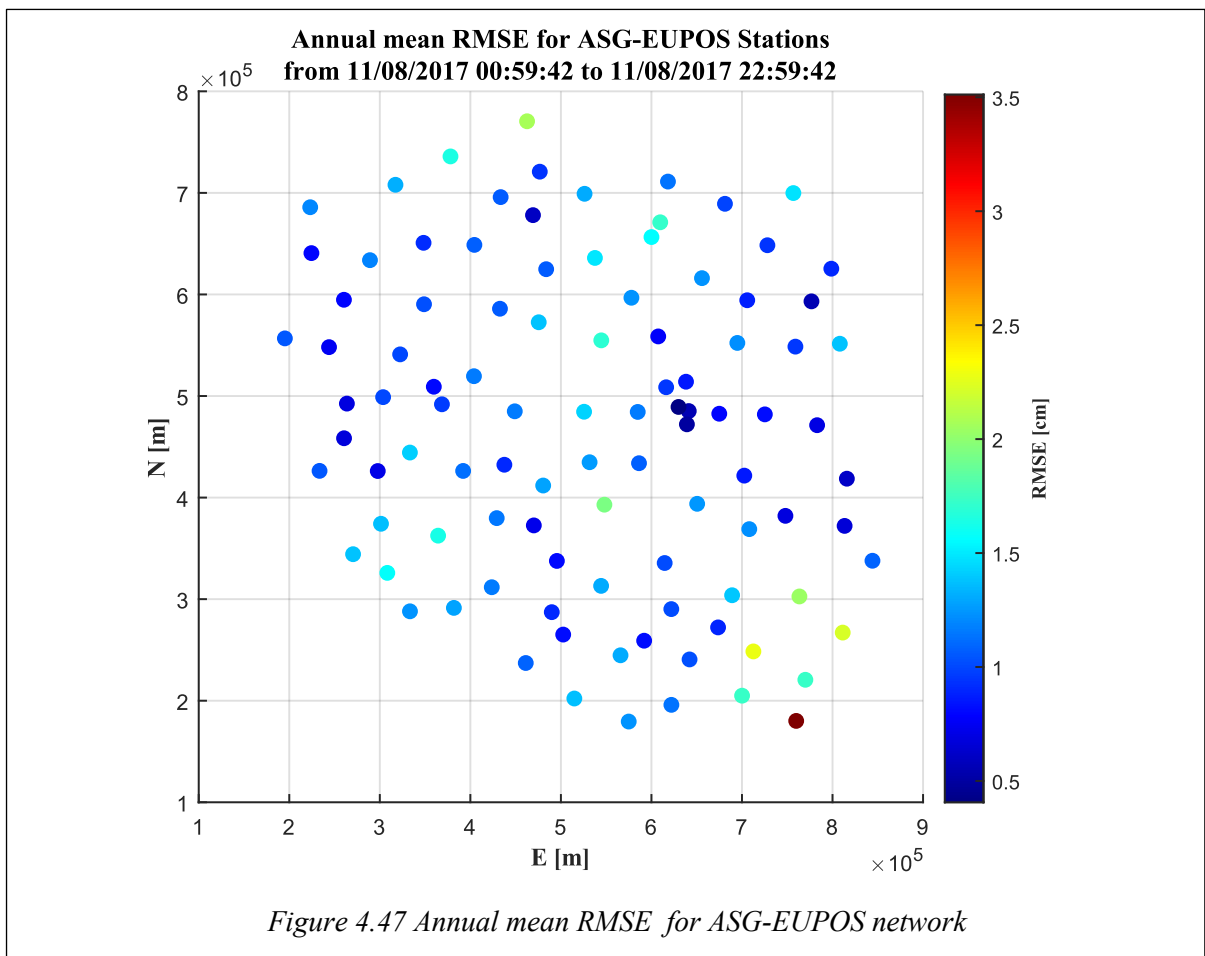
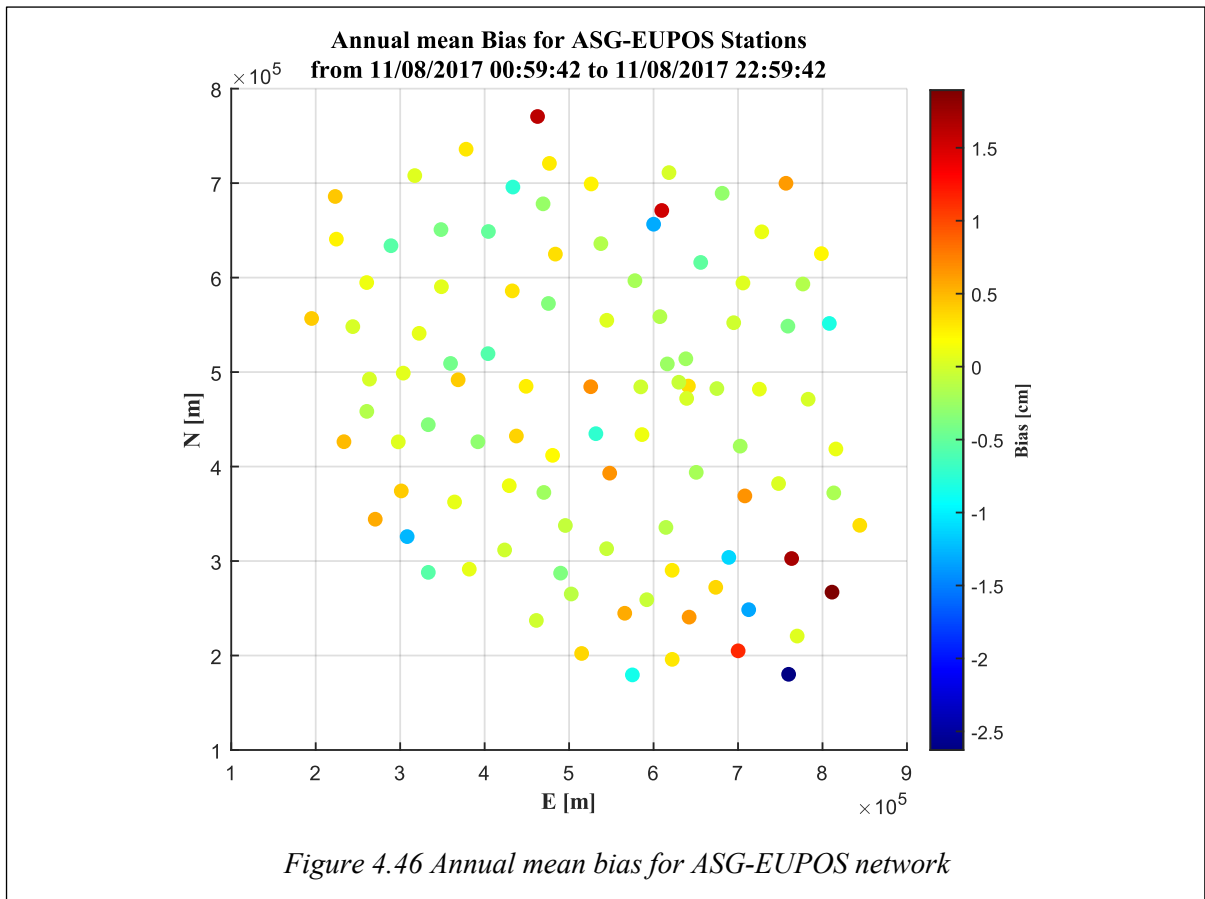
In this case study, we employed two approaches for cross-validation. The first approach, similar to case study 1, involved leaving one station out of the analysis and applying the Leave-One-Out Cross-Validation (LOOCV) technique using the network stations themselves. In the second approach, we utilized one network for the primary prediction and one of the other two networks for cross-validation purposes.

##### A. Leave-One-Out Cross Validation

In this study, LOOCV was employed to assess the prediction performance for all epochs on August 11th, 2017, ranging from 00:00:12 to 23:59:12, with a temporal resolution of 1 hour. The analysis utilized a total of 100 stations associated with the ASG-EUPOS network.

Subsequently, the discrepancy was calculated between the turbulence predicted using LOOCV and the turbulence delay obtained by removing the stratified component from the Zenith Total Delay (ZTD) observations. To evaluate the accuracy of the predictions, statistical analysis was conducted, encompassing the estimation of bias, standard deviation (STD), and root mean square error (RMSE) for this discrepancy. These statistical measures were computed for all epochs and at the individual station level, providing a comprehensive understanding of spatial and temporal prediction accuracy.

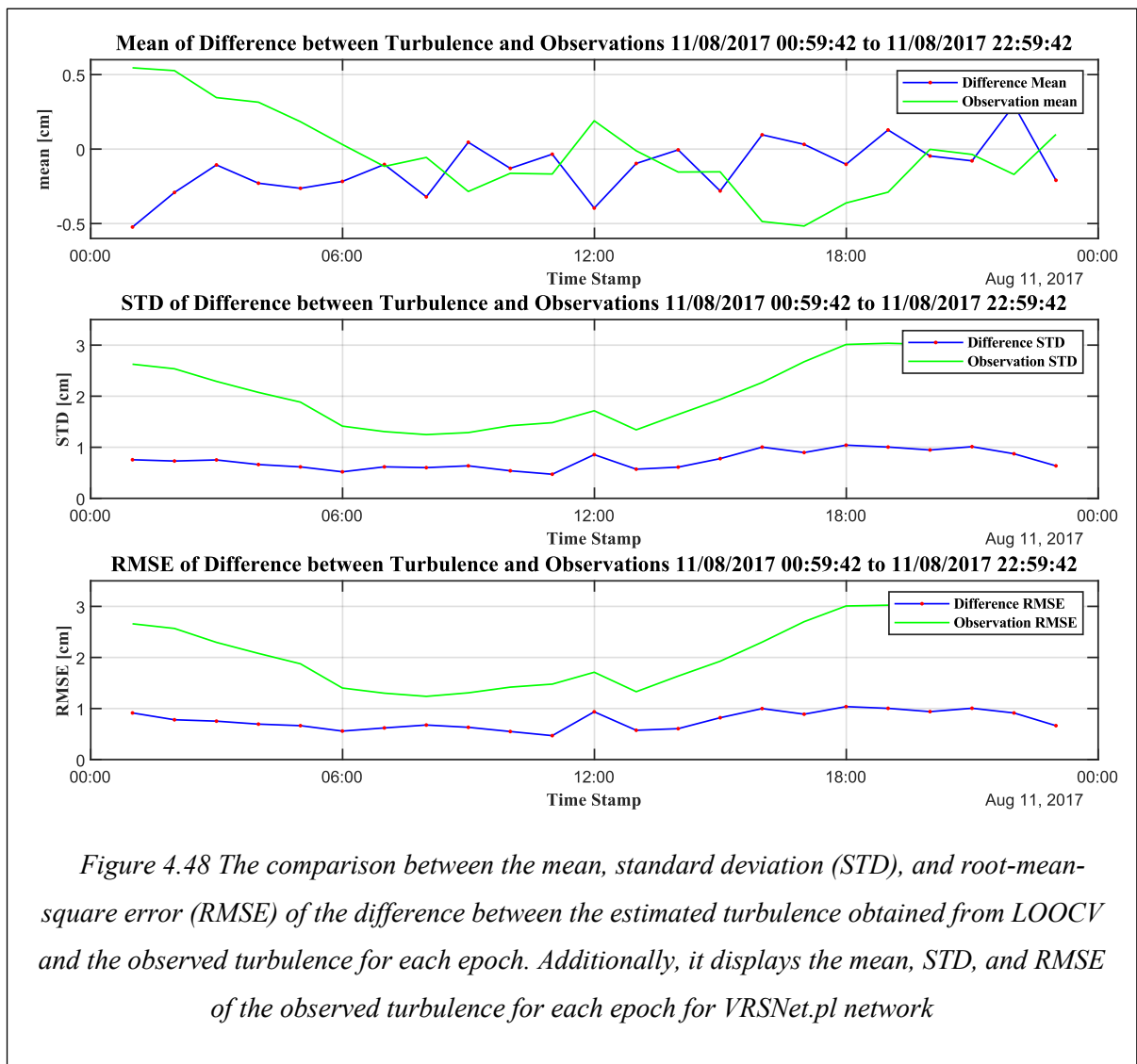




*B. Cross-validation using another network*

The validation of the ASG-EUPOS network was conducted using the VRSNet.pl network. The validation process involved removing the stratified component from the Zenith Total Delay (ZTD) observations at each epoch for all stations in both networks. Subsequently, the stochastic prediction steps described in section (4.2.2.2) were performed, with the difference being the use of the ordinary kriging system to predict at each station in the VRSNet.pl network instead of a uniform grid.

After obtaining the predicted turbulence values at each station in the VRSNet.pl network, a statistical analysis was conducted to assess the prediction accuracy. This analysis involved estimating the bias, standard deviation (STD), and root mean square error (RMSE) of the difference between the predicted turbulence values and the turbulence generated from removing the stratified component from the ZTD observations.



*Figure 4.48 The comparison between the mean, standard deviation (STD), and root-mean-square error (RMSE) of the difference between the estimated turbulence obtained from LOOCV and the observed turbulence for each epoch. Additionally, it displays the mean, STD, and RMSE of the observed turbulence for each epoch for VRSNet.pl network*

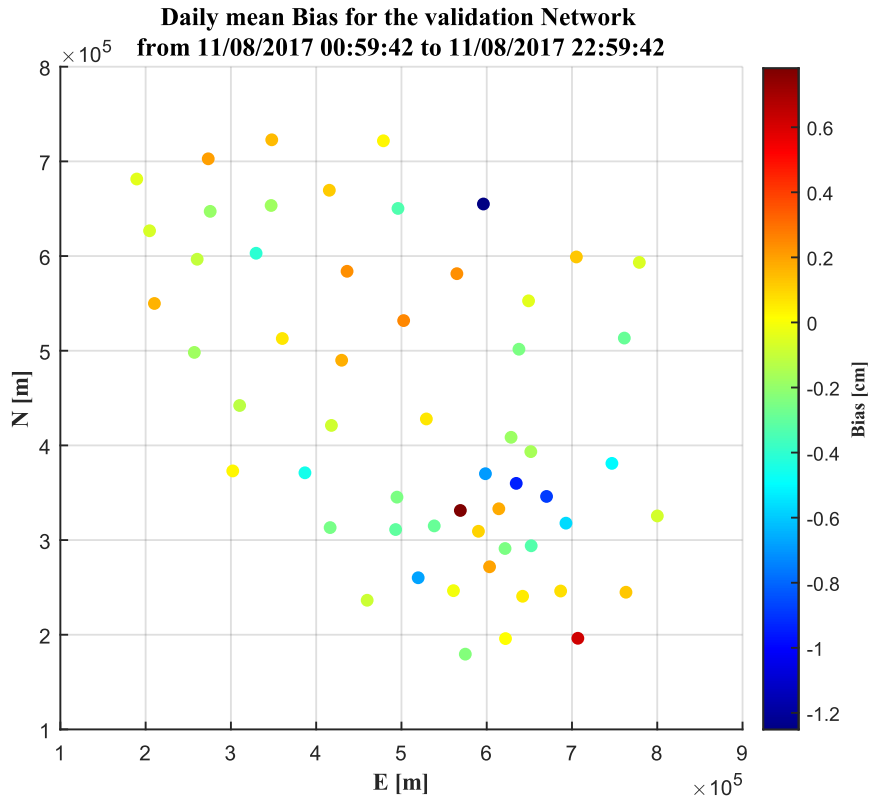


Figure 4.50 The mean daily Bias for VRSNet.pl network

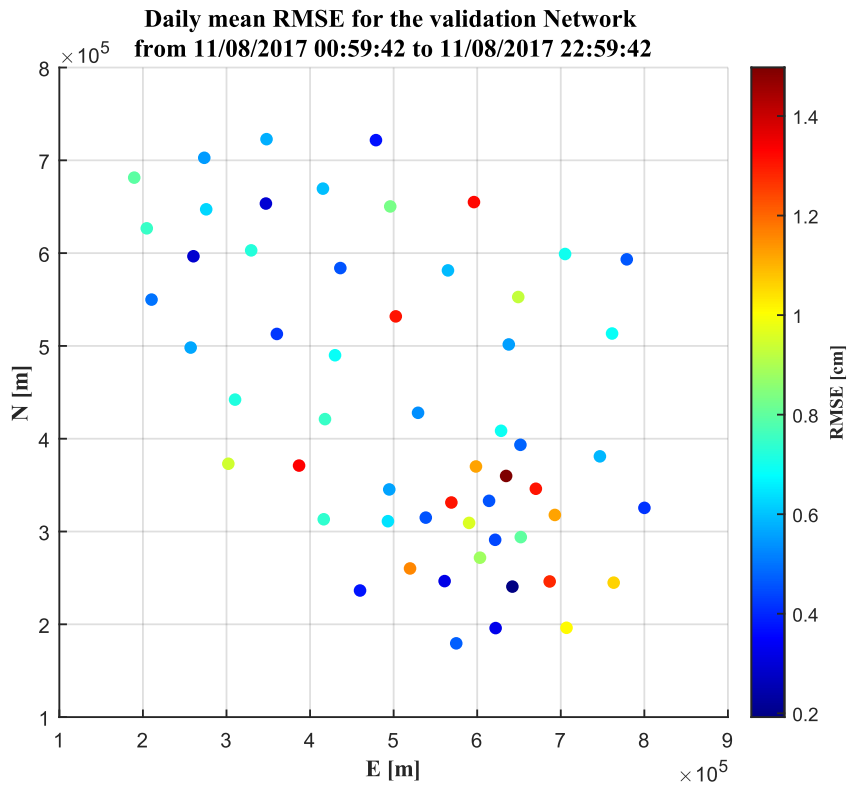


Figure 4.49 The mean daily RMSE for VRSNet.pl network

The comprehensive statistical evaluation of the ASG-EUPOS network after performing LOOCV and cross-validation using VRSNet.pl network, covering the period on August 11<sup>th</sup>, 2017 from 00:00:12 to 23:59:12, is summarized as follows:

*Table 4-6 The statistics for ASG-EUPOS GNSS network on August 11th, 2017 from 00:00:12 to 23:59:12 using LOOCV and cross-validation using VRSNet.pl network*

	<i>LOOCR</i>	<i>Cross-validation VRSNet.pl</i>
<i>Annual mean of Bias(cm)</i>	<i>0.0116</i>	<i>-0.1233</i>
<i>Annual mean of STD(cm)</i>	<i>1.2102</i>	<i>0.7447</i>
<i>Annual mean of RMSE(cm)</i>	<i>1.2043</i>	<i>0.7659</i>

From the validation results obtained through LOOCV and cross-validation using VRSNet.pl, a notable increase in RMSE and STD values can be observed starting from 14:00 UTC, particularly evident in Figure [4.45] and [4.48]. This significant change can be attributed to a severe storm event that occurred on August 11th, 2017. Subsequent to 14:00 UTC, convective cells re-emerged, with additional cells developing and merging overtime under the findings of (Nykiel et al., 2019). These results demonstrate the capability of our estimation method to detect and track the evolution of such events.

Analyzing the mean bias and RMSE for each station, as depicted in Figures [4.46], [4.47], [4.49], and [4.50], reveals that the southeast station exhibited higher RMSE and bias values. This can be attributed to the mountainous terrain in the southern region of Poland where the station is located, which introduces atmospheric instability affecting both the observed zenith total delay (ZTD) and our prediction accuracy.

Moreover, a spatial trend is observed as values progressively increase towards the north. This trend is a consequence of the storm's movement, initially heading northeast and subsequently shifting northward around 18:00 UTC (Nykiel et al., 2019).

These findings highlight the ability of our estimation method to effectively capture and track the development of severe weather events while accounting for regional atmospheric variations.



# Chapter 5

## 5. Conclusion

In this thesis, we have presented a novel approach for generating high-resolution water vapor maps using GNSS ZTDs. By applying the Stochastic prediction approach to the residuals obtained after removing the stratified component from ZTD observations, we were able to capture the correlation among our observations through variograms, assuming homogeneity and isotropy of the water vapor field. However, the turbulent nature of water vapor posed significant challenges in determining the appropriate lag distance for variograms and selecting the fitting model, highlighting the limitations of automatic implementation.

To generate our maps, we employed ordinary kriging interpolation and validated our predictions using LOOCV and cross-validation with the assistance of another network. Two case studies were conducted under different climatic conditions: one in northern Italy using the SPIN3 network for long-period observations, and the other in Poland during a severe storm. The estimation quality, as measured by LOOCV, was approximately 1 cm in both cases, demonstrating the effectiveness of our approach.

From our analyses, we observed that the accuracy of our predictions is influenced by the density and placement of GNSS stations. In mountainous regions, the accuracy tends to be lower compared to flat areas, likely due to fewer stations in proximity. Similarly, stations located near the study area boundaries exhibit reduced accuracy due to a limited number of nearby stations. Additionally, we identified the increased instability of water vapor during the summer season and the capability of our model to accurately capture severe weather events. These findings underscore the relevance of GNSS in meteorology, as it is the only technique capable of providing tropospheric data irrespective of weather conditions.

For future work, we propose the exploration of anisotropic variogram estimation, incorporating not only the separation distance between station pairs but also their orientation. Additionally, given that the effective weights in ordinary kriging are concentrated in the closest 10 stations to the estimation point, local kriging using only nearby stations could be considered. Furthermore, it would be valuable to conduct further validation by comparing our maps with Interferometric Synthetic Aperture Radar (InSAR) and exploring other techniques.

In conclusion, our thesis has introduced a robust methodology for generating high-resolution water vapor maps from GNSS ZTDs. The results obtained from the case studies have demonstrated the effectiveness of our approach, while also revealing the influence of station density and placement, seasonal variations, and severe weather conditions on the accuracy of predictions. The proposed future directions aim to enhance our methodology and expand the validation process, paving the way for advancements in water vapor mapping and its applications in various fields.

---



---

# References

- Biagi, L. (2022). *Master's Lecture: Positioning and Location-Based Service*.
- Böhm, J., & Schuh, H. (Eds.). (2013). *Atmospheric effects in space geodesy*. Springer.
- Chiles, J.-P., & Delfiner, P. (2012). *Geostatistics: Modeling spatial uncertainty* (2nd ed). Wiley.
- Cressie, N. A. C. (1993). *Statistics for spatial data* (Rev. ed). Wiley.
- DeGroot, M. H. (1986). *Probability and statistics* (2nd ed). Addison-Wesley Pub. Co.
- Efron, B. (1983). Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation. *Journal of the American Statistical Association*, 78(382), 316–331.  
<https://doi.org/10.2307/2288636>
- El-Rabbany, A. (2002). *Introduction to GPS: The Global Positioning System*. Artech House.
- Fermi, A. (2018). *Stochastic analysis of Troposphere's non-hydrostatic refractivity field for small- and mid-scale GNSS networks*. <https://www.politesi.polimi.it/handle/10589/139369>
- GPS SPS Performance Standard. (2020). *GLOBAL POSITIONING SYSTEM STANDARD POSITIONING SERVICE PERFORMANCE STANDARD*. 196.
- GReD - Geomatics Research & Development s.r.l. (n.d.). Retrieved 18 June 2023, from <https://www.g-red.eu/>
- Hofmann-Wellenhof, B., Lichtenegger, H., & Wasle, E. (2008). *GNSS--global navigation satellite systems: GPS, GLONASS, Galileo, and more*. Springer.
- Kaplan, E. D., & Hegarty, C. (Eds.). (2006). *Understanding GPS: Principles and applications* (2nd ed). Artech House.
- Kaplan, E. D., & Hegarty, C. (Eds.). (2017). *Understanding GPS/GNSS: Principles and applications* (Third edition). Artech House.
- Kitanidis, P. K. (1997). *Introduction to geostatistics: Applications to hydrogeology*. Cambridge University Press.
- Leick, A. (1995). *GPS satellite surveying* (2nd ed). J. Wiley and Sons.
- Leick, A. (2004). *GPS satellite surveying* (3rd ed). John Wiley.
- Lichtenstern, A. (2013). *Kriging methods in spatial statistics* [Technische Universität München]. <https://mediatum.ub.tum.de/doc/1173364/document.pdf>
- Longman, R. J., Frazier, A. G., Newman, A. J., Giambelluca, T. W., Schanzenbach, D., Kagawa-Viviani, A., Needham, H., Arnold, J., & Clark, M. P. (2019). High-Resolution Gridded Daily Rainfall and Temperature for the Hawaiian Islands (1990–2014). *Journal of Hydrometeorology*, 20(3), 489–508.
- Montenbruck, O., & Teunissen, P. (Eds.). (2017). *Springer Handbook of Global Navigation Satellite Systems* (1st ed. 2017). Springer International Publishing: Imprint: Springer.  
<https://doi.org/10.1007/978-3-319-42928-1>

- Nykiel, G., Figurski, M., & Baldysz, Z. (2019). Analysis of GNSS sensed precipitable water vapour and tropospheric gradients during the derecho event in Poland of 11th August 2017. *Journal of Atmospheric and Solar-Terrestrial Physics*, 193, 105082. <https://doi.org/10.1016/j.jastp.2019.105082>
- Papoulis, A. (1991). *Probability, random variables, and stochastic processes* (3rd ed). McGraw-Hill.
- Seeber, G. (2008). *Satellite geodesy: Foundations, methods, and applications*. Walter de Gruyter.
- SPIN3 GNSS – Servizio di Posizionamento Interregionale GNSS. (n.d.). Retrieved 9 June 2023, from <https://www.spingnss.it/>
- Teunissen, P. J., & Montenbruck, O. (2017). *Springer handbook of global navigation satellite systems* (Vol. 10). Springer.
- Venuti, G. (2023). *Master's Lecture: Geospatial Data Analysis*. Stochastic prediction, Politecnico di Milano.
- Wackernagel, H. (2003). *Multivariate geostatistics: An introduction with applications* (3rd completely rev. ed). Springer.
- Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists* (2nd ed). John Wiley & sons.
- Wolf, P. R., & Ghilani, C. D. (2006). *Elementary surveying: An introduction to geomatics* (11th ed). Pearson Prentice Hall.