Multiscale Characterization of the Probability Density Functions of Velocity and Temperature Increment Fields

by
Adam Ward DeMarco

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Marine, Earth, and Atmospheric Sciences

Raleigh, North Carolina

2017

APPROVED BY:

Dr. Sukanta Basu
Co-chair of Advisory Committee

Dr. Russell Philbrick
Co-chair of Advisory Committee

Dr. Anantha Aiyyer

Dr. Steven Fiorino

Dr. Yang Zhang

ABSTRACT

DEMARCO, ADAM WARD. Multiscale Characterization of the Probability Density Functions of Velocity and Temperature Increment Fields. (Under the direction of Dr. Sukanta Basu and Dr. Russell Philbrick.)

The turbulent motions with the atmospheric boundary layer exist over a wide range of spatial and temporal scales and are very difficult to characterize. Thus, to explore the behavior of such complex flow enviroments, it is customary to examine their properties from a statistical perspective. Utilizing the probability density functions of velocity and temperature increments, δu and δT, respectively, this work investigates their multiscale behavior to uncover the unique traits that have yet to be thoroughly studied. Utilizing diverse datasets, including idealized, wind tunnel experiments, atmospheric turbulence field measurements, multi-year ABL tower observations, and mesoscale models simulations, this study reveals remarkable similiarities (and some differences) between the small and larger scale components of the probability density functions increments fields.

This comprehensive analysis also utilizes a set of statistical distributions to showcase their ability to capture features of increments probability density functions (pdfs) across multiscale atmospheric motions. Also, an approach is proposed for estimating their pdfs utilizing the maximum likelihood estimation (MLE) technique, which has never been conducted utilizing atmospheric data. Using this approach, we reveal the ability to estimate higher order moments accurately with a limited sample size, which has been a persistent concern for atmospheric turbulence research. With the use robust Goodness of Fit (GoF) metrics, we quantitatively reveal the accuracy of the distributions to the diverse dataset. Through this analysis, it is shown that the normal inverse Gaussian (NIG) distribution is a prime candidate to be used as an estimate of the increment pdfs fields. Therefore, using the NIG model and its parameters, we display the variations in the increments over a range of scales revealing some unique scale-dependent qualities under various stability and flow conditions. This novel approach can provide a method of characterizing increment fields with the sole use of only four pdf parameters. Also,

we investigate the capability of the current state-of-the-art mesoscale atmospheric models to predict the features and highlight the potential for use for future model development. With the knowledge gained in this study, a number of applications can benefit by using our approach, including the wind energy and optical wave propagation fields.

BIOGRAPHY

Major Adam DeMarco was born and raised in the surburbs of Pittsburgh, PA. He graduated from Penn State University in 2003 with a Batchelor of Science degree in Meteorology. Following his undergraduate program, he joined the U.S. Air Force and was commissioned as a Second Lieutenant to serve as a Meteorologist. Since then, the author has been stationed around the world providing detailed weather forecasts in support of numerous air and ground operations. In 2012, he went back to school to receive a Master of Science degree in Meteorology from the Naval Postgraduate School in Monterey, CA. Then in 2014, he was selected by the U.S Air Force to pursue a PhD in Meteorology at North Carolina State University. Upon graduation, the author will return to service and continue his career as a U.S. Air Force weather officer.

ACKNOWLEDGEMENTS

This experience has been quite an exciting adventure! I'm very grateful to the U.S. Air Force for providing me with the resources for this wonderful opportunity, but the views expressed in this dissertation are those of the author and do not reflect the official policy or position of the U.S. Air Force, Department of Defense, or the U.S. Government. Without the help from some very influential people, my success during my PhD studies would not have been possible. First, I would like to thank my co-advisor, Dr. Sukanta Basu, for his guidance and involvement in my work. He led me down a path with a clear and coherent vision which was instrumental for my success! In addition, I would like to thank my other co-advisor, Dr. Russell Philbrick. He has had a wealth of experience both in academia and in government and was willing to pass along to me the knowledge he gained, which will continue to guide me as I continue my career as an Air Force officer and scientist.

Also, I would like to thank my other committee members, Dr. Anantha Aiiyer, Dr. Steven Fiorino, and Dr. Yang Zhang for their time and assistance during my research. Many thanks also go out to my fellow group members, Dr. Ping He, Dr. Yao Wang, Dr. Chris Nunalee, and Pat Hawbecker, for being such a great sounding board during this time. They were always willing and able to assist me when I was in bind. Their advice and support was greatly appreciated and needed. Similarly, I am very thankful for the computational resources provided to me by the Air Force Research Laboratory, Department of Defense Supercomputing Resource Center and Dr. Kevin Keefer from the Air Force Institute of Technology. During my studies, I never had to worry if I was going to run out of resources thanks to their support. Additionally, I would be remiss if I did not thank my Air Force leadership, specifically, Lt Col Brian Kabat, for the potential they saw in me and the support they have given me throughout this endeavor. Also, many thanks to my parents for instilling in me the strong work ethic that has carried me throughout my life.

Lastly, but certainly not least, I am eternally gratefully to my loving and encouraging wife, Meg, and two boys, Lucas and Alex, for their constant support and understanding throughout this long process. Without them, I would certainly not be as dedicated or motivated. So to put it simply, they are why I do what I do. I love you and appreciate all you have done for me. Thank you!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations

ABL	Atmospheric Boundary Layer
A-D	Anderson Darling
DoD	Department of Defense
cdf	Cumulative Distribution Function
edf	Empirical Distribution Function
ESS	Extended Self-Similarity
EV	Extreme value
GHSST	Generalized Hyperbolic Skew Student-t's
GoF	Goodness of Fit
IR	Inertial Range
i.i.d.	Independent and identically distributed
KH	Kelvin-Helmholtz
K-S	Kolmogorov-Smirnov
LNSS	Log Normal Superstatistic
MLE	Maximum Likelihood Estimation
MME	Method of Moments Estimation
NIG	Normal Inverse Gaussian
NWP	Numerical Weather Prediction
PBL	Planetary Boundary Layer

pdf	Probability Distribution Function
Q-Q	Quantile-Quantile
RE	Reynolds number
SF	Structure Function
TKE	Turbulent kinetic energy
USAF	United States Air Force
VG	Variance Gamma
WRF	Weather Research and Forecasting

Chapter 1

Introduction

> Could I just look at something which everybody had been looking at for a long time and find something dramatically new?
>
> Benoit Mandelbrot (2010)

1.1 Motivation

The Earth's atmosphere is highly complex with motions spanning across a wide range of spatio-temporal scales, from the large globally driven circulations to the small-scale turbulent eddies. The size of these features range from thousands of kilometers to on the order of millimeters and can occur over a range of weeks to seconds [246]. This extensive range leads to an exorbitant number of degrees of freedom, making it nearly impossible to accurately predict the changes in the flow over these scales.

Throughout the years, many scientists have examined the characteristics of turbulence in order to better understand its behavior and how it evolves. The progress made has brought forth a vast number of ideas and new approaches to examining turbulent media. However, due to the randomness and non-linear nature of these flows, to this day, it is still extremely difficult to fully encompass multi-scale and multi-dimensional motions with the sole use of first principles (i.e., the Navier-Stokes equations) due to the inability to have a closed-form solution. In fact, this issue has been coined, by many, as one of the most important unsolved problems in modern physics [253]. Thus, in the field of fluid dynamics, it is quite customary to describe the turbulent characteristics with the use of statistics in order to identify similarities and differences between various flow types (e.g, Reynolds number, boundary conditions, large-scale forcing, *etc.*). Another area of turbulence research focuses on identifying the dynamical structures of turbulence (e.g., vortex elements [148, 195]). Though, for this work, the primary method of evaluation is geared towards expanding on the former point of view. In fact, in a highly regarded fundamental turbulence book, Monin and Yaglom (1971, [180]) stated:

> "The theory of turbulence by its very nature cannot be other than statistical, i.e. an individual description of the fields of velocity, pressure, temperature, and other characteristics of turbulent flow is in principle impossible. Moreover, such description would not be useful even if possible, since the extremely complicated and irregular nature of all the fields eliminates the possibility of using exact values of them in any practical problems..."

Utilizing the techniques and theories addressed in this book, countless efforts, some of which will be addressed in detail in Chapter 2, were made to identify statistical properties of various turbulent flows. From an experimental perspective, a number of works strive to uncover statistical scaling properties and universal behavior. At the heart of this work is the 1-D velocity (or temperature) increment series, δu (or δT), see Eq. 1.1,

$$\delta u(r) = u(x+r) - u(x), \tag{1.1}$$

where r is the spatial separation (similarly τ represents temporal separation). It has been shown that the probability density functions (pdfs) of both velocity and temperature increments, across a wide spectrum of scales, have very unique and self-similar properties [185, 251, 257], which are quite different from classical turbulence theory [138].

To illustrate the pdf features, Figure 1.1 shows an example of a normalized velocity increment pdf from atmospheric measurements (grey symbols) with a Gaussian distribution overlaid.

In many applications, it is customary to assume a Gaussian (or normal) distribution for random continuous processes. However, from the pdf perspective it becomes clear that there are distinct differences between what is observed and the normal distribution. For instance, the tails of the pdf are much heavier (i.e., more extended) for the observed data, indicating that extreme wind events (i.e., strong wind gusts) have a much higher probability of occurrence compared to the assumed Gaussian distribution (see black arrow). Predicting this deviation is of great importance to a number of application, in particular the wind energy industry. Therefore, it is important to gain a comprehensive understanding into these features over a range of various time and space scales given the extreme difference observed in Figure 1.1.

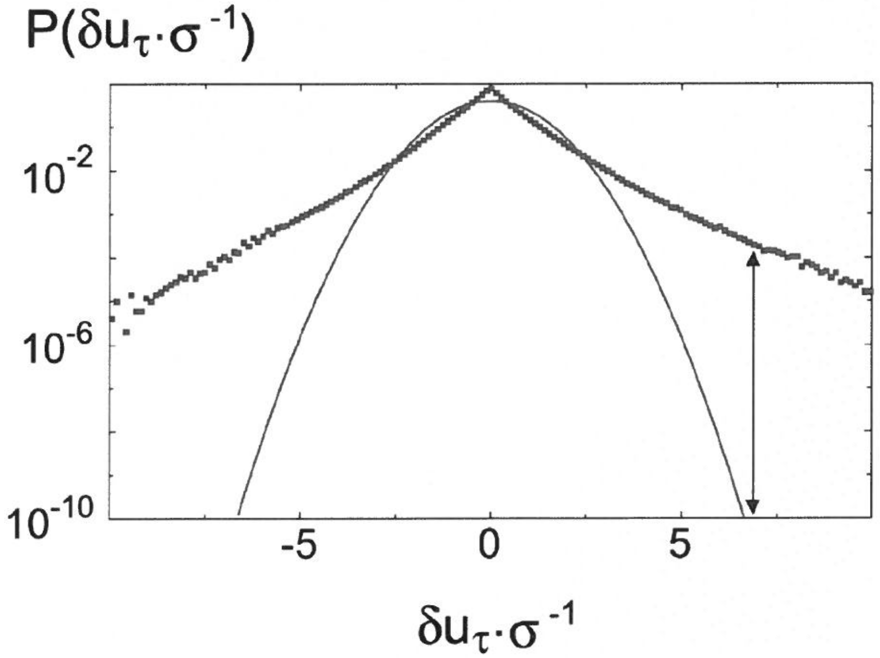


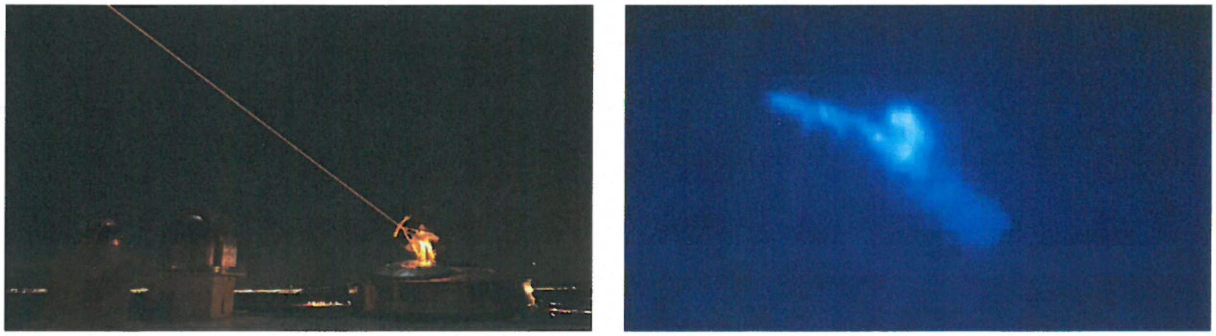
Figure 1.1: Atmospheric normalized velocity increment pdf at ($\tau = 4$ s) overlaid with a Gaussian distribution as indicated with the solid line (adapted from [48]).

Even though a number of works have been dedicated to pdfs of these fields, there is still no consensus into whether or not these features can be accurately captured by a particular pdf model. A few studies have shown the ability of pdf models to capture the traits of velocity and temperature increments (e.g., log-normal [56] and stretched-exponential [64]) from a visual perspective. However, no study has attempted to provide a quantitative analysis into the quality of the fits, especially across a wide range of scales. Therefore, this work is devoted to filling that void. Furthermore, examination of multi-year wind and temperature datasets provides a unique opportunity to explore the long-term statstical behavior of these fields. Unfortunately, these extensive datasets have not been utilized in the past for this type of evaluation, thus this work will be the first to reveal the atmospheric boundary layer (ABL) characteristics of the wind speed and temperature increment pdfs using such long-term data. Having the ability to properly characterize and estimate these increments can be of great benefit to a number of practical applications.

From a military perspective, for example, the Department of Defense (DoD) has become increasingly interested in understanding and predicting the effects small-scale atmospheric turbulence has on optical wave propagation in order to minimize the impacts on various DoD space and airborne surveillance applications (see Figure 1.2). The influences of turbulence can have direct effects on the propagation of light and sound waves in the atmosphere. For instance, a phenomenon called scintillation, which is the fluctuation of the beam intensity, is due almost exclusively to atmospheric temperature variations [152, 208]. Additionally, outside of the small scale range, the larger scale influences are still not well-defined [10], thus work is still being conducted to futher enhance our understanding of the turbulent nature of our very complex atmosphere. Further, the larger scales often describe the sources and character of small scales, because of the way they are linked. The energy of the large scales provides the driving force for small scales, and the rates of turbulent dissipation determines the lifetime of large scales. Therefore, with a proper characterization of velocity and temperature increments, improvements can be made in this arena. As eloquently stated by Lawrence and Strohben (1970, [152]):

> "A serious problem is associated with the use of the Gaussian model for optical propagation; we believe its use should be discouraged. The Gaussian model contains only a single scale size, although at least two independent scales are important

for optical effects. We shall see that the small-scale fluctuations are primarily responsible for intensity scintillation effects, and the large-scale fluctuations produce optical phase effects. A Gaussian model having too few adjustable parameters ties these together in an arbitrary and unrealistic way and engenders misleading and erroneous predictions."



(a) (b)

Figure 1.2: a) United States Air Force (USAF) laser-guided telescope in Albuquerque, New Mexico [169], b) Blurred image of a satellite obtained from the telescope indictive of turbulence impacts [169].

Furthermore, proper characterization of the velocity and temperature increments and identifying a suitable pdf model can aid in the testing and validation of mesoscale models. For instance, conventional validation methods generally utilize traditional statistical methods, such as root mean square error and correlation coefficient to compare observations against model results [124]. However, these techniques do not provide a rigorous evaluation. From this standpoint, the characterization of the observed atmospheric pdf increments can be used as a surrogate for providing a benchmark for future model developement.

1.2 Objectives and Science Questions

The objective of this research is to investigate the behavior of multi-scale wind and temperature increment fields. Also, an approach will be proposed for estimating their pdfs by utilizing the maximum likelihood estimation techinque and a set of pdf models. In this work, a variety of observational and modeling datasets are used to explore the boundary layer characteristics of wind and temperature increments. With the use Goodness of Fit (GoF) techniques, an examination into the model pdfs is conducted to identify a suitable model for practical use. Therefore, the primary science questions raised by this research are:

1) Do velocity and temperature have a universal behavior across a range of spatio-temporal scales? This science question will be addressed in Chapters 5, 6, 7, 8, 9, and 10.

2) If so, can we identify a universal probability density function (pdf) model which can be used for both velocity and temperature increments? This science question will be addressed in Chapters 5, 6, 7, 8, 9, and 10.

3) Utilizing the maximum likelihood estimation method, can we accurately estimate the pdfs assuming an underlying distribution? This science question will be addressed in Chapter 4.

4) Do various stability conditions influence the behavior of the pdf fields? If so, do the pdf models still capture these traits? This science question will be addressed in Chapters 6, 8, 9, and 10.

5) How well do the state-of-the-art mesoscale models capture the observed pdfs of wind and temperature increments? This science question will be addressed in Chapter 10.

1.3 Dissertation Outline

The overarching theme of this research is to use a myraid of long-term, research grade datasets to identify the similarities and differences in the wind and temperature pdfs from different scales of motion. Therefore, the chapters of this dissertation are organized as follows:

- Chapter 2 provides a brief background into the pdf concepts and brief literature review of the current state and understanding of the increment pdfs.

- In Chapter 3, an introduction into the family of pdf models which have the capability to represent the unique behavior of multiscale turbulent phenomena. Also, there is a discussion on the maximum likelihood estimation technique which was employed throughout this work.

- Chapter 4 discusses some of the current inherent shortfalls that exist in terms of the limited sample size that are pervalent in typical geophysical datasets. This study showcases the performance of a model pdf, Normal Inverse Gaussian (NIG), and the ability to estimate it using the maximum likelihood estimation technique. This type of analysis was never considered for turbulence data. The material presented in this chapter is accepted and in press as the following publication: DeMarco, A.W. and Basu, S. (2017) Estimating Higher-Order Structure Functions from Geophysical Turbulence Time-Series: Confronting the Curse of the Limited Sample Size, Physical Review E.

- Chapter 5 examines high-Reynolds number flows from laboratory and atmospheric boundary layer settings to explore the model pdfs fits to determine how well each model can represent the small-scale turbulence data.

- In Chapter 6, using a state-of-the-art, stably stratified wind tunnel dataset, we will explore the variability in pdf parameters as a function of stability.

- Chapter 7 focuses specifically on the tails of the mesoscale range wind distributions with emphasis on the extreme ramp-up and ramp-down characteristics. This work is specifically beneficial to the wind energy industry for predictive purposes. The material presented in this chapter is in peer-review as the following publication: DeMarco, A.W. and Basu, S. (2017) On the Tails of the Wind Ramp Distributions, Wind Energy.

- Utilizing the dataset from the previous chapter, Chapter 8 tests the different pdf models against each other using GoF techniques (i.e., Kolmogorov-Smirnov and Anderson-Darling tests) to determine which model is able to fit the data the best. We also show the impacts of day versus night to reveal how the pdf parameters are affected by the different atmospheric diurnal conditions.

- Chapter 9 evaluates the mesoscale temperature increments similar to Chapter 8's wind speed examination.
- The state-of-the-art mesoscale model, Weather Research and Forecasting (WRF) will be evaluated in Chapter 10 to see how well a handful of planetary boundary layer schemes are able to capture the pdf behavior in both wind and temperature.
- In Chapter 11, a conclusion of the work along with future directions is summarized.
- Appendix A provides the computational code utilized in this work.

Chapter 2

Background

The study of turbulence is rooted in the understanding and characterization of the statistical properties of the flow. From this point of view, the probability density functions (pdfs) of two-point difference (or increments, see Eq. 1.1) of these stochastic variables can shed light into unique features. The reasons for increased interest and an ability to accurate estimate the pdfs of these flows is that it can uncover all the statistical moments (i.e., mean, variance, *etc.*) of the turbulence field. Moreover, the tails of the distribution can provide great insight into predicting extreme events [251]. Thus, evaluating and determining an estimation of turbulence can be beneficial to a number of fields, as addressed in Chapter 1.

By definition, the p^{th} order moment of a variable x about the origin ($x = 0$) can be computed as an integral over the pdf of the velocity or temperature increment:

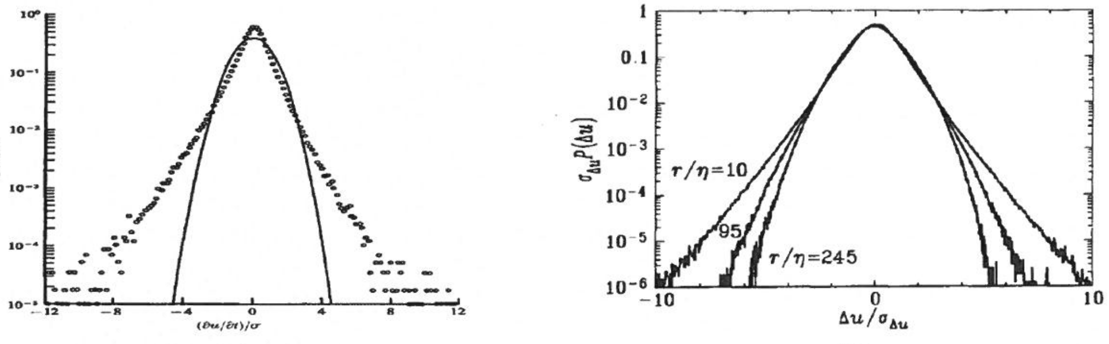
$$\langle x^p \rangle = \int_{-\infty}^{\infty} x^p f(x) dx, \tag{2.1}$$

where x is δu or δT for velocity and temperature increments, respectively, $f(x)$ is the pdf of x and the angular bracket denotes spatial averaging. The absolute value of x is simply the structure function (SF) formulation and this relation describes the linear dependence of structure. Traditionally, the study of turbulence deals extensively with the SF of a given variable to determine unique scaling features, however, the pdfs of these increments possess important information which can easily be lost using the SF approach, as it only computes the mean quanity of the increments. Therefore, increment pdfs are a necessary tool to evaluate the distinct properties of stochastic variables, such as intermittency.

Sreenivasan (1999, [237]) defines turbulence as intermittent and having a non-Gaussianity appearance: "Roughly speaking, intermittency means that extreme events are far more probable than can be expected from Gaussian statistics and that the probability density functions (pdfs) of increasingly smaller scales are increasingly non-Gaussian . . ." Thus, the study of intermittency for fine-scale turbulent behavior is most conveniently conducted utilizing the pdfs. Throughout the remainder of this chapter, a review of various turbulence pdf features, such as those due to intermittency, will be discussed starting with velocity increments.

Before delving into the different features of the pdfs of velocity and temperature increments, it is important to understand how to interpret a pdf plot. From a visual perspective these distributions are depicted by either a pdf or cumulative distribution function (cdf) in linear-linear, log-linear, and log-log coordinates. This work is primarily focus on last two vantage points. However, we caution that portraying pdf results via log-linear can mask some of the important features near the peak of the distribution, which will be highlighted in Chapter 8. Similarly, depicting results in a linear-linear fashion can hide the behavior of the tails. The visual comparison between these three plots provides a fast and intuitive view of the nature of the data. First, a power law distribution will appear as a convex curve in the linear-linear and log-linear plots and as a straight line in the log-log plot. "Next, a Gaussian distribution will appear as a bell-shaped curve in the linear-linear plot, appear as an inverted parabola in the log-linear plot and as strongly concave sharply falling curve in the log-log plot. Last, an exponential distribution will appear as a convex curve in the linear-linear plot, as a straight line in the log-linear plot and as a concave curve in the log-log plot [233]."

2.1 Velocity Increment PDFs



(a) (b)

Figure 2.1: a) An early example of the pdf of small-scale velocity gradient ($\delta u/\delta t$) from experimental data over the ocean indicating a clear deviation from Gaussian. (adapted from [257], Figure 8). b) The transition of normalize velocity increments ($\Delta u/\sigma_{\Delta u}$) from small ($r/\eta = 10$) to large ($r/\eta = 245$) scales. Here, the separation distance r is being normalized by η, the Kolmogorov length scale, $\eta = (\nu^3/\epsilon)^{1/4}$, where ν is the viscosity of the fluid (adapted from [109], Figure 5.1).

In 1956, Albert Einstein remarked that the most meaningful quantity to study in stochastic processes is the increment [84]. Over the last several decades, the characterization of turbulent flows utilizing the pdfs of the two-point velocity increments has gained momentum [24, 31, 43, 56, 100, 128, 156, 212, 248, 257, the list goes on]. The evolution of pdfs provides an interesting perspective in the description of velocity increments over a range of spatial and/or temporal scales. Many geophysical turbulence studies have shown that the pdfs of velocity increments are scale-dependent and change steadily within the inertial range, for example, see Figure 2.1. When viewed on a log-linear plot it can become clear that these distributions exhibit strong departure from Gaussianity (i.e. heavy tail and convex peaks near the core) at small increments, then become more Gaussian as separation increases towards the energy input scale [95]. Batchelor and Townsend (1949, [30]) found evidence that deviation from Gaussianity is stronger as the *RE* number is increased and with decreasing separation, which was later confirmed by Kuo and

Corrsin (1971, [143]) and others [253]. This behavior is considered to give further affirmation of the intermittency in the flow, which can also be depicted in the deviation from the structure function scaling exponent (ζ_p), discussed further in Section 2.4.

One specific feature of pdfs that varies with scale is the peakedness of the distribution (i.e., increased peakedness relative to a Gaussian distribution with decreasing separation). In fact, an intriguing finding was revealed by Huang *et al*, (2011) indicating that the maximum value of the velocity increment (i.e., mode or peak of the distribution) scales consistent with inverse of the first order structure function with respect to separation distance [118]. From a physical perspective the variations in the peak are often thought to be "...caused by the random and gentile fluid motion in the center of the ramps leading up to the sharp velocity gradient [43, 140]."

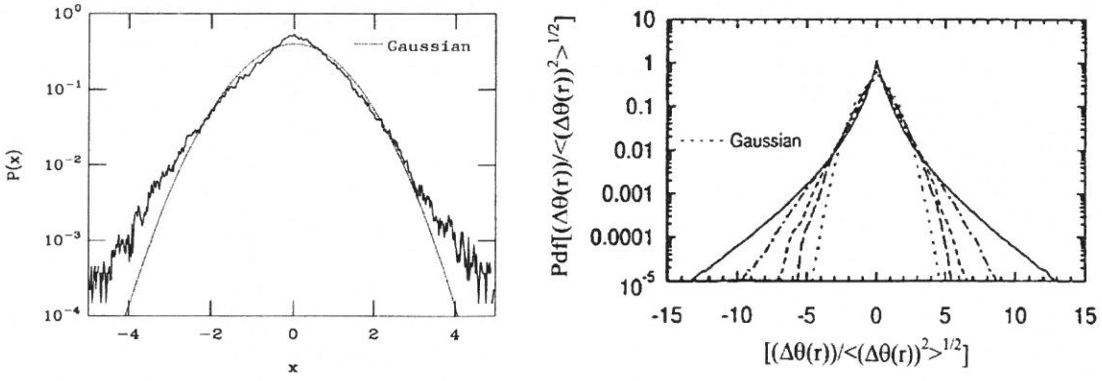
Another characteristic of pdfs is the skewness factor, believed to be related to the vortex folding and stretching process as the energy is transferred between the large and small turbulent eddies [63]. For instance, pdf have been shown to display increased negative skewness as thermal stability increases [95].

Furthermore, early experimental evidence revealed that the tails of the pdfs, contributing to higher-order moment estimation, have an exponential behavior, semi-heavy tail features (i.e., leptokurtic) for both the velocity increment and derivative fields [12, 94, 99, 128, 248, 251, 257]. By many [12, 110], this exponential tail was subsequently represented with a stretched exponential in order to capture the extended tail behavior and aid in the determination of the higher-order moment of velocity increments. Physically speaking, the exponential tails of the pdf are thought to be caused by occasional sharp velocity gradients (rounded-off shocks,[140], slender vortex filaments [195]). Thus it is believed that the degree of heavy tailedness is proportional to the RE number, that is higher RE leads to heavier tails [25].

Significant efforts have been made to determine if a particular pdf model is capable of capturing these ever-changing pdf features. One of the most well-accepted ideas is the suggestion that the turbulence behavior follows a stretched exponential pdf, $\left(e^{-c|x|^{\beta}}\right)$, where x represents the increment and β is the "stretching" parameter from 0 to 1 and represents the tail behavior [37, 128, 248]. This particular model has shown reasonable qualitative fit to turbulence data near the dissipation range (smallest-scales) where the tails are heavier, [24]. However, there are a few concerns with this model: (i) this is only a one-parameter (β) pdf and cannot estimate all the features of empirical data; (ii) this model assumes symmetry in the pdf, which has been shown not to be the case for turbulence data; (iii) it focus on the tails of the distribution and cannot capture the full features [95].

In the early nineties, Castaing *et al.* (1990, [56]), utilizing the Kolmogorov's 1962 (K62, [139, 199]) log-normal model, a fitted distribution that is "weakly curved" near the dissipative range, which deviates from the earlier exponential finding. A year later, a similar result was also confirmed using direct numerical simulation (DNS) [264]. However, as remarked in [56] there are some issues with this model in terms of properly capturing the center and skewness of the distribution. Similarly, in practice the log-normal model is considered to provide a modest fit for experimental data [244, 268]. Interestingly, studies within other geophysical disciplines (e.g., solar wind fluctuations [204, 234]) have experimented with the log-normal model, confirming that it provides representative fits to other geophysical turbulent flows. Unfortunately, these attempts to accurately characterize the pdfs of velocity increments were focused on fitting the tails, thus a prediction of the core or "shape parameter" of the distribution was lacking. Obtaining precise characterization is important for discerning between different theories.

Additional pdf models were later proposed in the early 2000s, that were believed to accurately capture the full behavior of the velocity increment distribution. Beck and Cohen (2003, [32]) introduced the concept of a superstatistic (aka "statistics of a statistics") incorporating the log-normal concepts of Castaing *et al.* (1990, [56]) and having the ability to capture unique pdf features. Also, Barndorff-Nielsen (2004, [24]) illustrated, in a parsimonious manner, that a member of the generalized hyperbolic distribution family, normal inverse Gaussian, could perform equally as well for turbulent quantities. The NIG model has been proposed and tested against the empirical data also revealing a remarkable fit to the data, most recently [43, 44, 166].



(a) (b)

Figure 2.2: Similiar examples as Figure 2.1, but for small scale temperature a) adapted from Ching (1991,[64]). b) adapted from Warhaft (2000, [269]).

These two models will be discussed in more detail in the next chapter. Now, despite these developments, we are still lacking a general consensus into whether or not a given pdf model can provide acceptable estimation of the turbulence behavior, especially in a quantitative manner. Therefore, this research is geared to explaining and expanding on these ideas.

2.2 Temperature Increment PDFs

In the study of fluid flows, the temperature field behaves in the passive sense in which the velocity acts as a vehicle advecting this scalar fields around where it has no significant effects on the turbulent dynamics. However, in larger scale motions and in stratified environments, such as the atmopshere, temperature fluctuations can directly effect the velocity field (via buoyancy forces). Thus, it is important to investigate difference between both quantities. The study of the fluctuations of the local temperature contain considerable information about both dynamics and transport processes [104]. Some other examples of scalar fields include pollutants, humidity, and pressure, which will not be studied here, but have implications in dispersion modeling and optical turbulence research.

The pdf of temperature increments, or similarly, passive scalars have received considerable attention but not nearly as much as the velocity counterpart [14, 64, 141, 178, 214, 226, 226, 236, 251, 269]. Within these studies, it is well discussed that the temperature increments pdfs in small-scale settings behave somewhat differently than velocity, such that the deviation from Gaussian is much more pronounced in temperature, for example see Figure 2.2. The temperature increments "...is affected both by the intermittency of the energy transfer and the intermittency of the temperature variance; this can explain the large deviation from the Gaussian curve, Castaing *et al.* (1990 [56])."

Furthermore, it has been shown that the temperature pdfs, within laboratory flows, show a much more heavier tail compared to the velocity counterpart, which indicates a more anomolous behavior in the temperature fluctuation field [56]. Even though, there are differences between the two variables, pdf of temperature increments have also shown to fit with stretched exponentials well [64]. However, a similar concern arises with regards to the issues of the stretched exponential fit addressed in Section 2.1. Warhart (2000, [269]) remarked that the exponential tails evident in the scalar fields was due to anomalous mixing as the fluid moves beyond the integral length scale. This behavior was also shown in [126] to describe the small-scale temperature pdfs that have a Gaussian appearance near the core and exponential tails beyond 1σ. In terms of the skewness features of the temperature pdfs, Ould-Rouis *et al.* (1995, [203]) found that "...the experimental skewness factor for the distribution of $\Delta\theta$ (temperature difference) grows when separation is decreased, in total contradiction with the local isotropy assumption which stipulates that it should be zero." Though, the velocity increments tend towards isotropy with decrease separation, the temperature field does not exhibit this feature and is a major difference between the two.

Another example of the temperature pdfs come from LES results, computed by Métais and Lesieur (1992, [177]) and experimental evidence shown by Sreenivasan *et al.* (1979, [240]). These studies revealed that the tails of the temperature fluctuations are exponential, whereas the longitudinal velocity fluctuations were Gaussian. This behavior demonstrates that the temperature variations are strained by its velocity counterpart. Eventhough there are good examples in numerical and laboratory experiments, the features of temperature increments within the ABL still remains largely undocumented.

In terms of the scaling characteristic the temperature field within the inertial range is often assumed to scale similar to velocity (i.e., $\zeta_2 = 2/3$, based on Kolmogorov-Obukhov-Corrsin (KOC), [71, 138, 198]). However, similar to K41, this theory is also not consistent with the experimental findings of the temperature pdfs and leads to anomolous scaling characteristics.

Figure 2.2 showcases two examples of pdfs of temperature increments within the inertial range. We will probe further into the characteristics of temperature in the atmosphere as well as under various stability conditions.

2.3 Mesoscale Increment PDFs

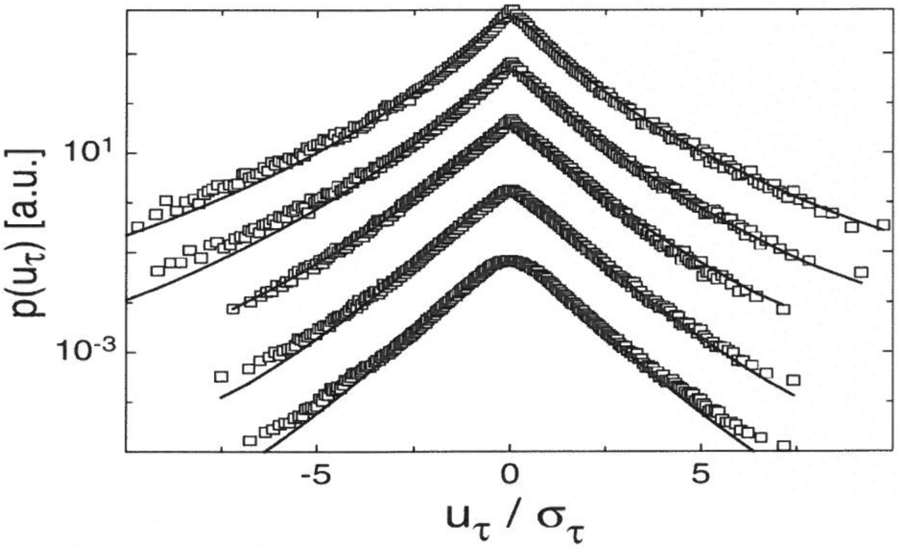


Figure 2.3: This example shows the evolution of the velocity increment pdf for larger scale motions from 2.5, 25, 250, and 4,000 sec, from top-bottom respectively. The solid lines represent the log-normal model (adapted from [49]).

In contrast to the evaluation of the pdfs within microscale turbulent fields, wind speed scaling properties within the mesoscale range (approximately sub-hourly to sub-daily time scales) still lacks an extensive examination. Due to a number of external forces and influences, such as diurnal variations, atmospheric stability, and topographical interaction (i.e., inhomogeneities), it is not expected that a universal scaling description can be made within this regime [151]. Moreover, in the mesoscale range, the scaling properties have shown to vary with topography and the atmospheric conditions.

Thus, only a handful of papers exist [21, 49, 53, 105, 133, 135, 136, 137, 150, 151, 155, 185, 219, 249] attempting to uncover a range of scaling characteristics. They analyzed data from around the world, however, several of these papers focused on wind gusts (τ being on the order of seconds), and thus, their findings are not directly the associated with the mesoscale wind. For instance, Böttcher *et al.* (2007, [49]) found that the pdfs of velocity increments within the ABL have a similar to isotropic turbulence in the small-scale range, but are quiet different in the larger scales (up to ~1 hr), see Figure 2.3. While, Liu *et al.* (2010, [156]) finds that between sub-second to sub-hour that the tails of distribution fit remarkably well to truncated stable distributions. They conclude that the pdfs of the velocity increments of the atmospheric turbulence are not similar to what is observed in the laboratory turbulence. Later, Liu *et al.* (2013, [155]) showed both log-normal and truncated Lévy flights, but neither fit the data well.

For mesoscale regimes, Muzy *et al.* (2010, [185]) and Baïle and Muzy (2010, [21]) provided an evaluation of statistical description of the mesoscale wind increments. They concluded that the large-scale (hrs-days) motions do in fact resemble those intermittency-cascade properties found in small-scale laboratory turbulence. Similar results were discovered by Liu and Hu (2013, [155]). In fact, some empirical evidence suggest that the wind speed is more intermittent in the mesoscale range than the inertial range [53, 151]. Additionally, Telesca and Lovallo (2011, [249]) found a strong height dependence on the scaling characteristic, but they did not provide a physical explanation. It was speculated in Kiliyanpilakkil and Basu (2015, [136]) that the buoyancy effects are at the root of this height-dependency trait as buoyancy force dominates at higher heights above the surface [246].

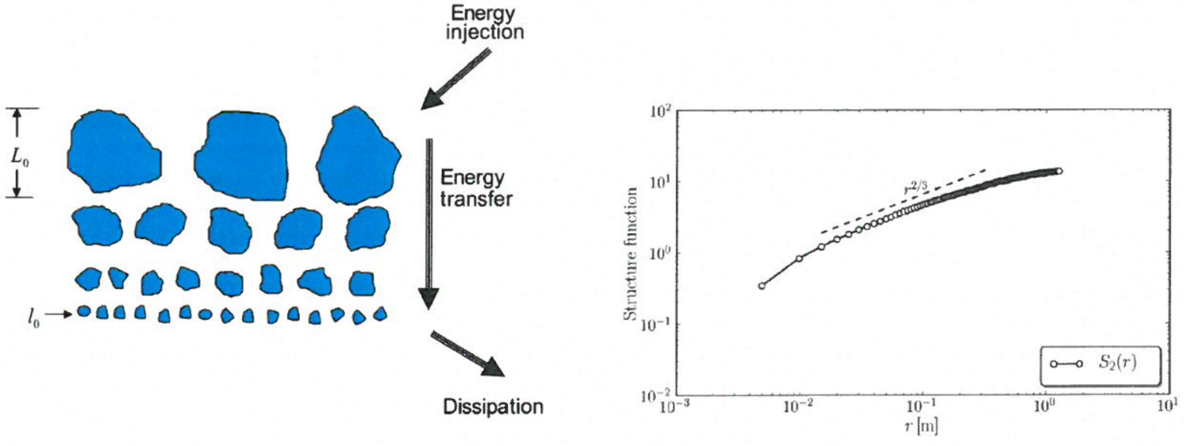
A significant issue with most of these studies was the use of very limited amount of observational data (often with durations of a few hours to merely a few days). Despite, quasi-universal features being revealed, many of these studies concluded that further experimental evidence is required to confirm these finding. The scaling properties of the temperature field within the mesoscale range has not received any attention in terms of the pdfs. This work will address the characteristics of the temperature pdfs and will attempt to uncover features that may shed light other unique features.

In conclusion, the past research in this arena has proved to be very insightful for characterizing turbulence phenomena. However, researchers are still struggling with providing a theoretical framework for explaining the dynamics behind known features of turbulence (i.e., intermittency and energy cascade). As mentioned by many, more work is required to gain further knowledge regarding a practical approach for estimating and representing the pdf models. In the next chapter, we will discuss the pdf models which exhibit strong characteristics of non-Gaussianity as well as our estimation techinques which we will employ throughout this research.

2.4 A Brief Review of Other Statistical Descriptions of Turbulence

One of the earilest, and still highly regarded, theories was first introduced by Lewis F. Richardson in 1922 [216]. In his work, he proposed the concept of the kinetic energy cascade in which energy from high Reynolds (RE) number flows is injected into the flow by large-scale instabilities and transferred through the intermediate scales (aka the inertial range), until the energy is dissipated, by viscosity, at small scales and converted to molecular heat, see Figure 2.4a.

Taking this idea a step further, in the 1941, Andrey Kolmogorov (K41) [138] along with his student Alexander Obukhov [198] derived a very powerful formulation based on dimensional analysis, from the Navier-Stokes equation to describe this process. They postulated that under



(a) (b)

Figure 2.4: a) Depiction of small-scale turbulent energy cascade (adapted from Andrews and Philips (2005, [10])). Based on K41 theory, the inertial range in fully-developed turbulence is bounded by a region of eddies sizes between the outer (L_0) and inner length scale (l_0). b) Example of "2/3" scaling law within a turbulent random field (adapted from Wittwer (2013, [272])).

homogeneity, isotropic, and a constant dissipation rate, energy cascades from large to small scales through the inertial sub-range (IR) with a scale-invariant relationship of ($\zeta_2 = 2/3$) until the energy is dissipated and transferred to molecular heat. The energy cascade process is dependent on only the characteristic size of the energy input, (integral length scale L) and the mean energy dissipation rate, $\langle\epsilon\rangle$. Under the assumption of homogeneity and isotropy, turbulence can be universally described by employing the K41 theory, which has become the foundation of turbulence research, as described by Eq. 2.2,

$$S_p = \langle\big[u(x+r) - u(r)\big]^p\rangle = \langle|\delta u|^p\rangle \sim (\langle\epsilon\rangle r)^{\zeta p}, \tag{2.2}$$

where S_p is the p^{th} order structure function with respect to the spatial separation, r (equivalently the temporal separation, τ) and mean energy dissipation rate, $\langle\epsilon\rangle$. The angular brackets here indicate spatial averaging, $|\delta u|^p$ is the p^{th} order absolute moment of the velocity increments, and ζ_p is the scaling exponent. The scaling exponent is believed to have universal traits such that $\zeta_p = p/3$.

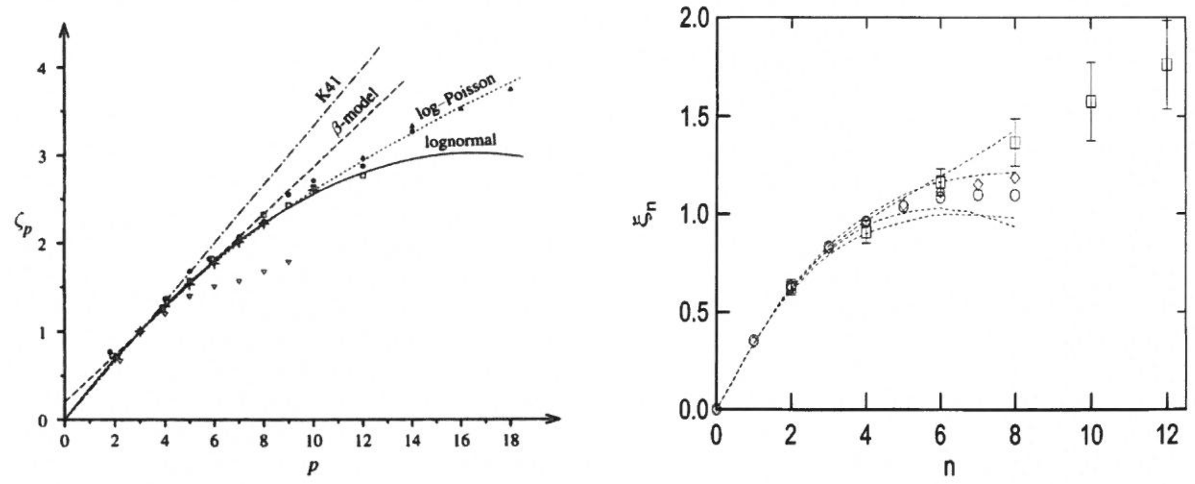
This statistical approach has paved the way for turbulence research over the last 70 years leading to a countless number of theoretical, experimental, and numerical modeling studies. However, shortly after K41's landmark theory, objections to the existence of a universal scaling behavior were made by Landau and Lifshitz (1944, [147]). They suggested that the large scale fluctuations played a role in the the energy dissipation rate and the scaling exponent was not universal. Thus, as a result, Kolmogorov and Obukhov in 1962 (KO62) proposed a refined similiarity hypothesis (RSH) remarking that $\langle \epsilon \rangle$, is not an average quantity throughout the energy cascade process, but rather has a varying, "local" dissipation rate, ϵ_r, over a eddy size, r. This modification assumes a log-normal probability density function for ϵ_r and that that the logarithm of the energy distribution in the inertial range, was Gaussian [139, 199]. This leads to the scaling exponents within the inertial range behaving in a non-linear fashion.

Following this change, experimental evidence confirmed this deviation in the scaling exponent $\zeta_p \neq p/3$, giving further credence to the existence of an intermittency characteristic in small-scale turbulence. Figure 2.5a along with Eq. 2.3 illustrates this non-linear feature in the scaling exponents as a function of p from various experiments along with a few cascade models (overlaid, e.g., log-normal [139, 139], β [97, 196], log-Poisson [225]) which have been developed to predict the scaling behavior.

Figure 2.5 also illustrates the scaling exponent for scalar fields revealing smaller exponents as the order of the moments increase. These models seems to give a good description of the turbulence behavior from experimental evidence. However, issues arise in the empirical identification of high-order structure functions resulting in a statistical inaccuracy which causes major problems using this SF approach. The existance and properties of small-scale intermittency still remains an open question. In Chapter 4, the higher-order moment estimation issue will be addressed in more details. For a comprehensive summary of turbulence cascade models, please refer to Frisch (1995, [95]) and Sreenivasn and Antonia (1997, [239]). The scaling exponent is described as,

$$\zeta p = \frac{p}{3} + \frac{\mu}{18}(3p - p^2), \tag{2.3}$$

where μ is referred to as the intermittency correction and believed to be ~ 0.25, [18], which corresponds to $\zeta_2 \sim 0.70$.



(a) (b)

Figure 2.5: a) Exponent ζ_p vs. p. Inverted white triangles: data from Van Atta and Park (1972) [258]; black circles, white squares and black triangles: data from Anselment *et al.* (1984) with increasing Re: + signs: wind tunnel data using ESS; straight dash-dot line from $\zeta_p = p/3$ (K41). The model fit lines are from the various cascade models. (adapted from [95]), b) The scaling exponents (ξ_n) for the scalar increments ($\Delta\theta_r$) with the separation distance in the inertial range. Squares are experimental results from Antonia *et al.* (1984, [14]), with vertical bars showing the uncertainty in the data. Circles and diamonds are from Meneveau *et al.* (1990, [174]. The dashed lines are estimates of the uncertainty (adapted from [239]).

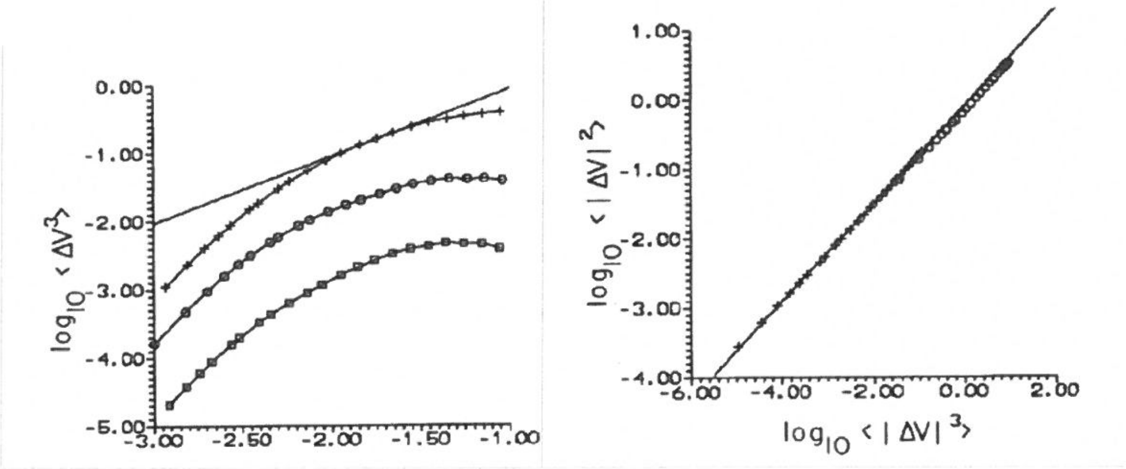
To expand these efforts, alternative, and perhaps more informative, statistical techniques can be used. For instance, examination into the normalized third and fourth order moments has become a method for capturing intermittent phenomena [239]. For instance, Eqs. 2.4 and 2.5 are referred to as the skewness (S_r) and flatness (F_r), respectively. The S_r factor is considered highly sensitive to the deviation from a normal distribution, which according to K41 theory this value should saturate at zero (S=0, [175, 253]). Whereas, the F_r is also a good metric for determining the degree of intermittency and is quite simply equal to three for pure Gaussian distributions. As the flatness values increase the distributions exhibit a greater departure from Gaussian. These quantities can prove valuable as statistical metrics regarding the behavior of the flow as a function of separation distance, (r). The skewness, S_r, and the flatness, F_r, are defined as,

$$S_r = \frac{\langle |\delta u|^3 \rangle}{\langle |\delta u|^2 \rangle^{3/2}}, \tag{2.4}$$

$$F_r = \frac{\langle |\delta u|^4 \rangle}{\langle |\delta u|^2 \rangle^{2}}. \tag{2.5}$$

Outside of the pdfs, spectra, and structure function approaches, one can use other types of scaling features to uncover various characteristics of turbulence phenomena. For instance, in the mid-1990s, Roberto Benzi and his colleagues introduced a novel concept called extended self-similarity (ESS) in which they identified a consistent scaling behavior of velocity/temperature increments that extend well beyond the inertial sub-range, into the dissipation range (see Figure 2.6 as an example, [38]). They approached the problem by plotting 3^{rd}-order structure functions against other ordered structure functions, which represented a relative scaling relationship against 3^{rd}-order, Eq. 2.6. This approach has since been examined experimentally through numerous laboratory, and some atmospheric, studies for a variety of Reynolds number flows [8, 35, 36, 38, 40, 78, 245, 276, just to name a few]. As for the temperature field, numerous other studies [5, 14, 16, 41, 61, 68, 154] have been conducted examining the scaling behavior of this field, through both SF and ESS analysis resulting in some variation between the velocity and temperature fields within the inertial sub-range, which is believed to be a result of stronger intermittency within the temperature field [178, 226, 269].

$$\zeta^*_{p,q} = \frac{\zeta_p}{\zeta_q} \tag{2.6}$$



(a) (b)

Figure 2.6: a) Third order SF of velocity increments versus spatial separation (not labeled in x-axis) for various Reynolds number flows, the solid line represents a slope of 2/3 in the inertial sub-range. b) The corresponding ESS (2^{nd} vs. 3^{rd} order) indicating larger range of 2/3 scaling (adapted from [38]).

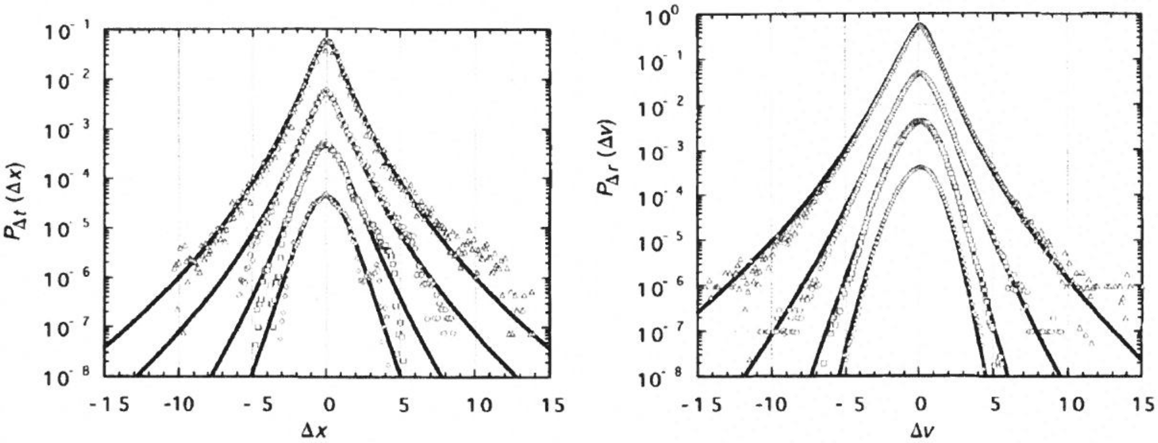
More recently, it was shown by Kiliyanpilakkil, et.al., [135, 136] that the wind speed characteristics from various locations within the atmospheric boundary layer indicates an ESS-based scaling relationship within the mesoscale range (scales larger than inertial sub-range, between sub-hourly to sub-daily). Remarkably, these scaling results were shown to be consistent with the ESS literature within the inertial range of the velocity field.

Chapter 3

Statistical Probability Density Functions and Parameter Estimation Techniques

Unlike examining turbulence series via scaling exponent of S_p over a range of scales, utilizing the pdfs of velocity increments can shed additional light into some of the unique statistical characteristics of the flow. Unfortunately, at this time, there is no universal underlying pdf capable of describing the statistical properties of turbulence. In order to determine possible pdf candidates for the sample data, one needs to use estimation techniques for comparison. In this work, we will show that the maximum likelihood estimation (MLE) approach is ideal in the sense that it seeks to find values of the pdf parameters from model distributions that maximize the likelihood that the observation data came from a given distribution [270]. Thus, using this estimation method we will evaluate a handful of known distributions to determine if the a number of observed data is accurately captured by a practicular model pdf.

The pdfs models used in this research are discussed below along with the techniques utilized to estimate the parameters of the pdfs. Some of these models extend beyond the field of fluid dynamics and have never been considered in this field as potential fits to the velocity and/or temperature increments from a wide range of scales. The benefit of identifying a pdf to represent turbulence is that the distribution will be able to characterize the flow with few parameters, even across distinct experiments. Quite interestingly, there are other research arenas, (e.g., financial, stock market variations [23, 42, 102, 205, 262, 278]) where observed stochastic quantities, like turbulence, deviate from a Gaussian distribution. In fact, some studies have attempted to draw analogies between the scaling properties of the two fields, see Figure 3.1. As a result of these works, studies in other disciplines, beyond hydrodynamic turbulence, have designated a mixture of various Gaussian distribution types, which capture the scale-dependent behavior of the increments.



(a) (b)

Figure 3.1: a) The stock market price changes (Δx) for different τ=640 sec, 5,120 sec, 40,960 sec, and 163,804 sec (from top to bottom) . b) turbulence flow (Δv) for different separations (δr = 3.3η, 18.5η, 138η, and 325η) from small to large scale, situated from top to bottom, respectively, (adapted from [102]).

A popular family of pdfs that is widely considered in the field of statistics and finance is the generalized hyperbolic distributions, [205]. In this work, we focus on a couple of these models as potential fits to the empirical turbulence data. The beauty of these distributions are that they are capable of representing not only the range of intermittent features of the pdfs (e.g., heavy tails), but they also can be estimated easily using a well-known estimating tool, known as the MLE. Throughout this work, we will be comparing these models against the observational data and between one another to determine the accuracy of the fits. As its been shown in mostly the finance community, see Figure 3.2, these models possess the traits which also exist in turbulence. In the following sections, the pdf models will be discusses along with some of their key features.

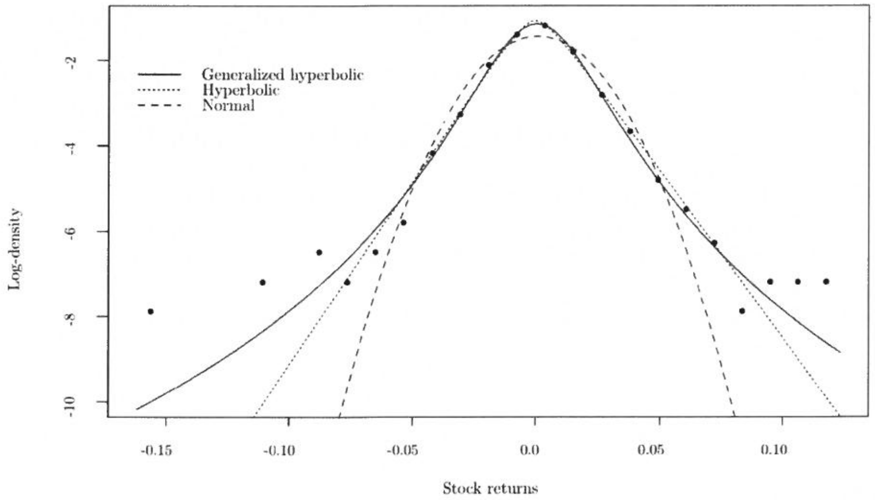


Figure 3.2: An illustration of a daily IBM-stock returns fitted with generalized hyperbolic, hyperbolic and Gaussian distributions (adapted from [42]).

3.1 Normal Inverse Gaussian (NIG)

The one-dimensional normal inverse Gaussian (NIG) pdf contains four-parameters (α, β, δ, and μ) and is defined as follows [22, 205, 221]:

$$
\begin{aligned}
f_{NIG}(x;\alpha,\beta,\mu,\delta) &= \frac{\alpha}{\pi}\exp\left(\delta\sqrt{\alpha^2-\beta^2}-\beta\mu\right)\frac{1}{\sqrt{\phi(x)}} \\
&\quad \times \exp(\beta x), K_1\left(\delta\alpha\sqrt{\phi(x)}\right). \qquad (3.1)
\end{aligned}
$$

where

$$
\phi(x) = 1 + \left[\frac{x-\mu}{\delta}\right]^2 \qquad (3.2)
$$

The parameter α controls the steepness of the pdf, where a small α signifies a heavy tail. β is the skewness parameter, such that a negative β indicates that the pdf is skewed to the left. μ is the centrality or translation parameter and is slightly different from the arithmetic mean of the distribution. δ is a scale or peakedness parameter and it controls the shape of the pdf near its mode. α and δ are always positive and $0 < |\beta| < \alpha$. This distribution is a sub-class of the generalized hyperbolic family. For NIG, λ, which is also a scale parameter, is equal to -$\frac{1}{2}$, which leads to a K_1 modified Bessel function of third kind of order one [3]. For an illustrative example of the behavior of this distribution, Figure 3.3 shows idealized cases for which two of the parameters vary. This representation highlights the inherent ability of this distribution to capture the unique traits of the increment pdfs.

The parameters of the NIG distribution are explicitly related to the first four central moments [24, 130]. The benefit of this pdf model is that all moments are finite and do not suffer from undefined or infinite moment estimation. Let us denote the sample mean, standard deviation, skewness, and kurtosis by m_1, m_2, m_3, and m_4, respectively. Further define: $\gamma = \frac{3}{m_2\sqrt{\left(3m_4-5m_3^2\right)}}$. Subsequently, the pdf parameters can be estimated, via the method of moments (MME) approach, as follows [130]:

$$
\alpha = \sqrt{\beta^2+\gamma^2}; \qquad (3.3a)
$$

$$
\beta = \frac{\left(m_2 m_3 \gamma^2\right)}{3}; \qquad (3.3b)
$$

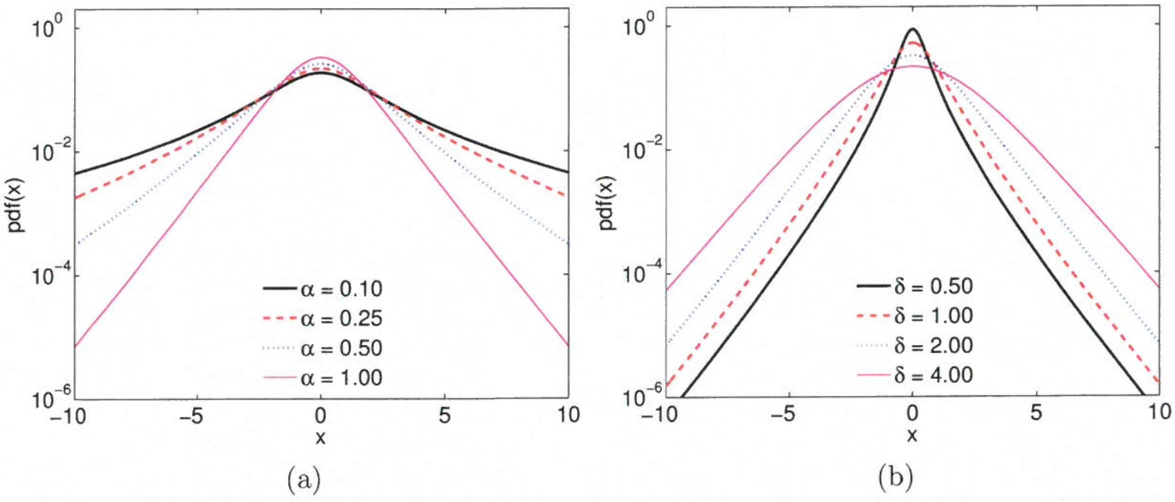


Figure 3.3: NIG pdfs for different values of α (left panel) and δ (right panel). For all the pdfs, μ and β are assumed to be equal to zero. For the left figure, δ is kept constant at 2. In the right figure, α is taken as 4.

$$\delta = \frac{m_2^2\gamma^3}{\beta^2+\gamma^2}; \tag{3.3c}$$

$$\mu = m_1 - \frac{\beta\delta}{\gamma}. \tag{3.3d}$$

The tails of the distribution exhibit the following traits:

$$f_x(x) \sim C|x|^{-3/2}\exp(-\alpha|x|+\beta x), \qquad \text{as } x \to \pm\infty. \tag{3.4a}$$

Thus, the heaviest tail decays as:

$$f_x(x) \sim C|x|^{-3/2}\exp(-\alpha|x|+|\beta x|) \qquad \text{when} \begin{cases} \beta < 0 \text{ and } x \to -\infty, \\ \beta > 0 \text{ and } x \to +\infty, \end{cases} \tag{3.4b}$$

and the light tail decaying as:

$$f_x(x) \sim C|x|^{-3/2}\exp(-\alpha|x| - |\beta x|), \qquad \text{when} \begin{cases} \beta < 0 \text{ and } x \to +\infty, \\ \beta > 0 \text{ and } x \to -\infty. \end{cases} \tag{3.4c}$$

Following the approach by Rydberg [217], numerically generated NIG distributed variates can be constructed using this approach:

- Sample a Inverse Gaussian distribution, $Z = \mathrm{IG}(\delta, \gamma)$.
- Sample a Gaussian distribution, $Y = \mathrm{Normal}(0,1)$.
- Return the NIG random variates, $X = \mu + \beta Z + \sqrt{Z}Y$.

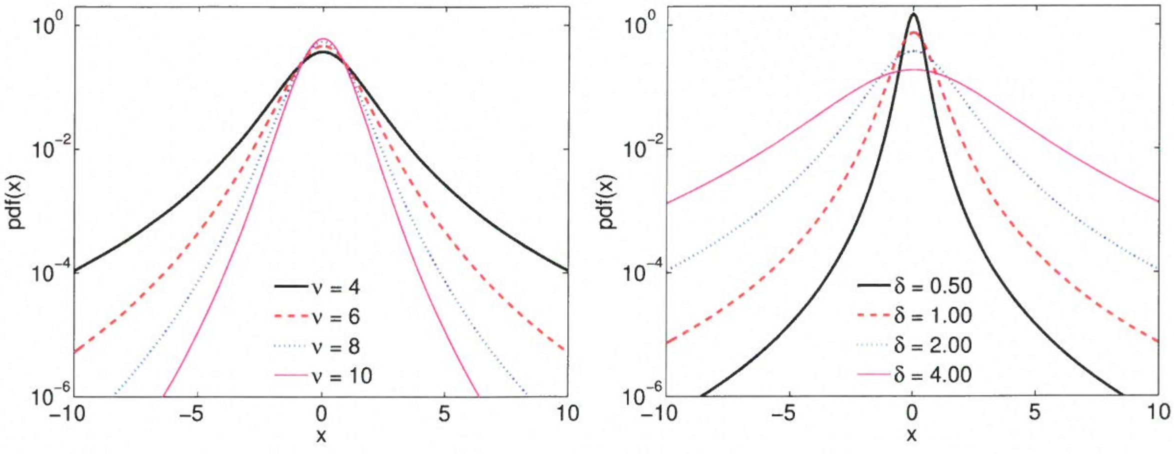
where $\gamma = \sqrt{\alpha^2 - \beta^2}$

3.2 Generalized Hyperbolic Skew Student's t (GHSST)

Another subclass of the generalized hyperbolic family of distributions is GHSST. It has been shown to fit the distributions of financial data very accurately and contains the following parameters (ν, β, μ, and δ) [2, 42, 89, 153, 205].

The skewness and the heavy tailedness of the GHSST are determined jointly by the combination of the parameter values of β and ν. With ν fixed, a lower value of β implies a more negative skewness or left-skewness as well as heavier tails. On the other hand, with β fixed, as ν becomes larger the density becomes less skewed and has lighter tails. Figure 3.4 shows idealized cases for which two of the parameters vary.

The GHSST distribution is a subclass of the generalized hyperbolic (GH) family, which exhibits both heavy tailed and skewed properties of a distribution. These features are uniquely characterized by the following four parameters (ν, β, μ, and δ) [1, 2, 205] given by this probability distribution function (pdf):



(a) (b)

Figure 3.4: GHSST pdfs for different values of ν (left panel) and δ (right panel). For all the pdfs, μ assumed to be equal to zero and $\beta \sim 0$. For the left figure, δ is kept constant at 2. In the right figure, ν is taken as 4.

$$f_x(x;\nu,\beta,\mu,\delta) = \frac{2^{\frac{1-\nu}{2}}\delta^{\nu}}{\sqrt{\pi}\Gamma(\nu/2)}\left(\frac{y_x}{|\beta|}\right)^{-\frac{\nu+1}{2}} K_{\frac{\nu+1}{2}}\left(|\beta|y_x\right)e^{\beta(x-\mu)}, \tag{3.5}$$

where $y_x = \sqrt{\delta^2 + (x-\mu)^2}$ and Γ is the gamma function. The parameter ν controls the tails of the distribution and represents the number of degrees of freedom. The heaviness of the tails are inversely proportional to the value of ν. Like NIG, β is considered the skewness parameter, such that a negative β indicates that the pdf is skewed to the left, while μ and δ are similar paramters as in NIG. $K_\nu(x)$ is the modified Bessel function defined as [3]:

The pdf parameters have the following form and the moments can easily be estimated for the sample statistics [2]:

$$m_1 = \mu + \frac{\beta\delta^2}{\nu-2}; \tag{3.6a}$$

$$m_2 = \left(\frac{2\beta^2\delta^4}{(\nu-2)^2(\nu-4)} + \frac{\delta^2}{\nu-2}\right)^2; \tag{3.6b}$$

$$m_3 = \frac{2\sqrt{\nu - 3}\beta\delta}{[2\beta^2\delta^2 + (\nu-2)(\nu-4)]^{3/2}}\left[3(\nu-2) + \frac{8\beta^2\delta^2}{\nu-6}\right]; \tag{3.6c}$$

$$m_4 = 3 + \frac{6}{[2\beta^2\delta^2 + (\nu-2)(\nu-4)]^2}$$
$$\left[(\nu-2)^2(\nu-4) + \frac{16\beta^4\delta^2(\nu-2)(\nu-4)}{\nu-6} + \frac{8\beta^4\delta^4(5\nu-22)}{(\nu-6)(\nu-8)}\right]. \tag{3.6d}$$

The skewness and kurtosis, higher-order moments, do not exist when $\nu \leq 6$, and $\nu \leq 8$, respectively. As it turns out, this limitation is a hindrance for use of GHSST for turbulence data. Their tails decay slower than the observations and other models. It can be proved that a much slower decay of tail distributions will lead to the divergence of variance. More details regarding this will be discussed later in Chapter 7.

The tails of the distribution exhibit the following traits:

$$f_x(x) \sim C|x|^{-\nu/2-1}\exp(-|\beta x| + \beta x) \qquad \text{as } x \to \pm\infty. \tag{3.7a}$$

Thus, the heaviest tail decays as:

$$f_x(x) \sim C|x|^{-\nu/2-1} \qquad \text{when} \begin{cases} \beta < 0 \text{ and } x \to -\infty, \\ \beta > 0 \text{ and } x \to +\infty, \end{cases} \tag{3.7b}$$

and the light tail decays as:

$$f_x(x) \sim C|x|^{-\nu/2-1}\exp(-2|\beta x|) \qquad \text{when} \begin{cases} \beta < 0 \text{ and } x \to +\infty, \\ \beta > 0 \text{ and } x \to -\infty. \end{cases} \tag{3.7c}$$

Such that, when $\beta > 0$ the left hand tail decreases as $|x|^{-(\nu/2+1)}\exp(2\beta x)$, while the right hand tail decreases as $|x|^{-(\nu/2+1)}$. When $\beta < 0$, the behaviour of the two tails is the opposite [42].

Following Aas and Hobæk Haff 2005 [1], numerically generated GHSST distributed variates were constructed using this approach:

- Sample a Gamma distribution, $G = \text{Gamma}(\nu/2, \delta^2/2)$.
- Obtain the Inverse Gamma by $Z = G^{-1}$.
- Sample a Gaussian distribution, $Y = \text{Normal}(0,1)$.
- Return the GHSST random variates, $X = \mu + \beta Z + \sqrt{Z}Y$.

3.3 Variance Gamma (VG)

The one-dimensional VG distribution has four parameters (α, β, μ, and λ, [161, 205, 211, 223], Eq. 3.3). Similar to the NIG, this distribution is also a sub-class of the generalized hyperbolic distribution familiy. Thus, the α, β, and μ parameters represent the same features as NIG. However, unlike NIG, where λ = -1/2, within the VG distribution λ varies and $\delta = 0$. This distribution type has never been thought to represent turbulence data and is mostly used for financial data analysis (e.g., share market returns) [161]. As λ increases towards ∞, the skewness and kurtosis of a random variable tend towards the skewness and kurtosis of a normal random variable is calculated from:

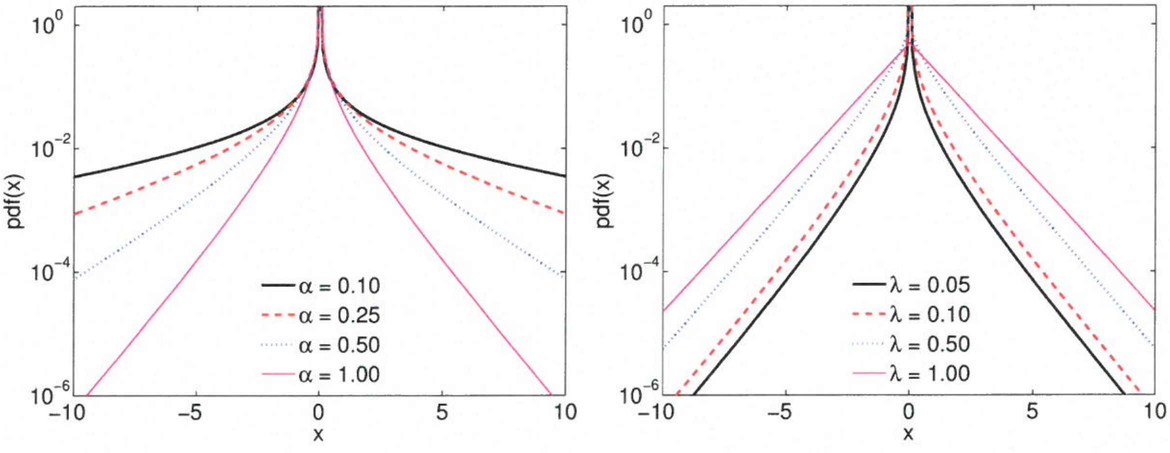
$$f_{VG}(x;\lambda,\alpha,\beta,\mu) \quad = \quad \frac{2\left(\frac{\alpha^2-\beta^2}{2}\right)^{\lambda}}{\sqrt{2\pi}\Gamma(\lambda)}\left(\frac{|x-\mu|}{\alpha}\right)^{\lambda-\frac{1}{2}} K_{\lambda-\frac{1}{2}}\left(\alpha|x-\mu|\right)e^{\beta(x-\mu)}, \tag{3.8}$$

where $\lambda = 3/(m_4\text{-}3)$ and Γ is the gamma function. Figure 3.5 shows idealized cases for which two of the parameters vary.

Subsequently, the pdf parameters can be estimated, via the method of moments approach, as follows [205]:

$$\psi = \frac{2\lambda}{m_2^2}; \tag{3.9a}$$

$$\alpha = \sqrt{\psi + \beta^2}; \tag{3.9b}$$



(a) (b)

Figure 3.5: VG pdfs for different values of α (left panel) and λ (right panel). For all the pdfs, μ and β assumed to be ~ 0. For the left figure, λ is kept constant at 1. In the right figure, α is taken as 1.

$$\beta = \frac{m_3 \lambda}{3\sqrt{m_2^2}}; \tag{3.9c}$$

$$\mu = m_1 - \frac{m_2 m_3 \lambda}{3}. \tag{3.9d}$$

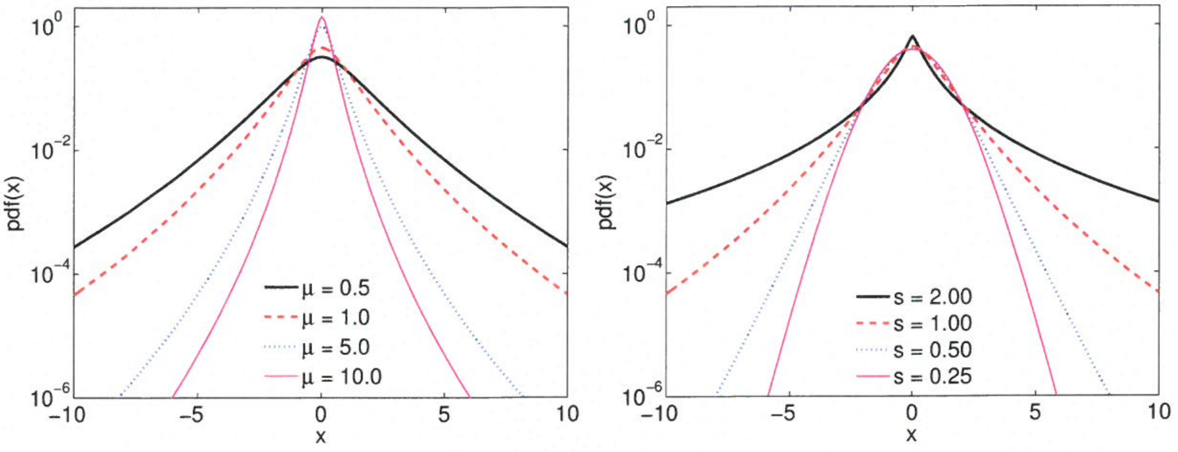
3.4 Log-Normal Superstatistics (LNSS)

The concept of superstatistics (aka "statistics of a statistics") was proposed by Beck and Cohen 2003 [32]. Here, the log-normal superstatistics (LNSS) pdf can be written as:

$$f_{LNSS}\left(x;\mu,s\right) = \frac{1}{2\pi s}\int_0^\infty b^{-1/2}\exp\left\{\frac{-\left(\ln\left(b/\mu\right)\right)^2}{2s^2}\right\}e^{-(1/2)bx^2}db. \tag{3.10}$$

Figure 3.6 will show idealized cases for which two of the parameters vary.

The non-central second-order moment and the flatness of this distribution can be written as follows: [31]:



(a) (b)

Figure 3.6: LNSS pdfs for different values of μ (left panel) and s (right panel). For the left figure, s is kept constant at 1. In the right figure, μ is taken as 1.

$$\langle x^2 \rangle = \sqrt{w}/\mu, \tag{3.11a}$$

$$F_2 = \frac{\langle x^4 \rangle}{\langle x^2 \rangle^2} = 3w, \tag{3.11b}$$

where $w = e^{s^2}$. These equations are easily solved to estimate the unknown parameters μ and s. Customarily, LNSS pdfs are estimated using MME techniques, thus for this work will be primarily folllowing this method. Also, as a cautionary note, the integral of LNSS cannot be evaluated in closed form and due to extremely high computational costs associated with numerical integration of the pdf, the MLE technique is problematic and will require further evaluation to determine a suitable alternative method. In the meantime, the MME-based estimation technique was utilized. Also, the use of MME will later reveal further evidence that MLE is a much more powerful and robust estimating tool.

In the present work, these method of moments-based estimates (MME) shown for all four models are used as initial conditions for the Maximum likelihood estimation (MLE) computations. A commercial MLE function from MATLAB™, employing the Nelder-Mead method, is utilized, however, an expectation-maximization algorithm is presented below as an alternative appraoch for computing MLE.

3.5 Maximum Likelihood Estimation

Maximum likelihood estimation (MLE) is a well-established statistical method to estimate the parameters of a given model (say NIG) which make the sample data, $\mathtt{x}_i = \mathtt{x}_1;\mathtt{x}_2::\mathtt{x}_N$, the most probable outcome. Let $\mathtt{f}(\mathtt{x}|\ )$ denote the underlying NIG pdf of the sample data with parameters $(\ _1 = \ ,\ _2 = \ ,\ _3 = \ $, and $\ _4 = \ )$. In the case of independent, identically distributed NIG variates, $\mathtt{f}(\mathtt{x}|\ )$ can then be expressed as a joint probability function:

$$\mathtt{f}(\mathtt{x}|\ ) = \mathtt{f}_1(\mathtt{x}_1|\ )\mathtt{f}_2(\mathtt{x}_2|\ ):::\mathtt{f}_N(\mathtt{x}_N|\ ): \tag{3.12}$$

However, in order to determine the model parameters based on observed data, we have to solve an inverse problem. The first step is to define a likelihood function, where the parameters are a function of the fixed data, as follows:

$$\mathtt{L}(\ |\mathtt{x}) = \mathtt{f}(\mathtt{x}|\ ): \tag{3.13}$$

In the case of the NIG pdf, this likelihood function is simply a four dimensional surface sitting above a four-dimensional hyperplane covering the NIG parameters. For computational ease, it is customary to maximize the log-likelihood (instead of $\mathtt{L}$) to obtain the MLE estimates. The log-likelihood function for the NIG pdf can be written as:

$$\begin{aligned} \ln(\mathtt{L}(\ |\mathtt{x})) &= -\mathtt{N}\ \ln(\ ) + \mathtt{N}\ \ln(\ ) + \mathtt{N}\ (\ \ -\ \ )- \\ \frac{1}{2}\sum_{i=1}^{N}\ln(\ (\mathtt{x}_i)) &+ \ \sum_{i=1}^{N}\mathtt{x}_i + \sum_{i=1}^{N}\ln\left[\mathtt{K}_{\ 1}(\ \ \sqrt{\ (\mathtt{x}_i)})\right]; \end{aligned} \tag{3.14}$$

where N is the sample size. Next, to find the parametric values which "maximizes" the log-likelihood function, we impose:

$$\frac{\partial \ln\left(L\left(\theta|x\right)\right)}{\partial \theta_k} = 0, \tag{3.15}$$

$$\frac{\partial^2 \ln\left(L\left(\theta|x\right)\right)}{\partial \theta_k^2} < 0, \tag{3.16}$$

where k varies from 1 to 4.

Karlis [130] developed an expectation-maximization algorithm 3.1 to easily maximize Eq. (3.14). He leveraged on the fact that the partial derivative of $\ln(L)$ with respect to β leads to a simple relationship connecting the sample mean ($\overline{x}$) and some of the NIG parameters:

$$\overline{x} = \mu + \frac{\delta\beta}{\gamma}. \tag{3.17}$$

He also derived the relationships between the conditional expectations of the inverse Gaussian (IG) distribution and NIG parameters (see [130] for details). A pseudo-code for his approach is provided below:

3.6 Extreme Value Theory - Hill Estimator

Another aspect of this research involves the behavior of the pdfs tails, which is dictated by extreme and rare events which also has a strong influence on the higher-order moment statistics (used to determine intermittency in small scale turbulence). Extreme-value (EV) theory concepts primarily developed for use in finance and statistics will be explored. Over the years, several estimators for γ^* (the tail-index) have been proposed in the literature, including (but not limited to): Pickand's estimator [209], Hill estimator [111], and the Dekkers-Einmahl-de Haan estimator [77]. For this work, the Hill estimator is used and is determined from the following

ALGORITHM 3.1: Expectation-Maximization for NIG pdf

```
Initialize:
    NIG parameters ($\alpha^{(1)}, \beta^{(1)}, \mu^{(1)}, \delta^{(1)}$) using the MME approach
Compute:
    $L^{(1)}$ using Eq. (3.14) and initialized parameters
for $k = 1$ to $k_{max}$ do
    Compute:
    E-Step - The conditional expectations, $s_i$ and $w_i$
        $s_i \leftarrow \frac{\delta^{(k)}\sqrt{\phi^{(k)}(x_i)}}{\alpha^{(k)}} \frac{K_0(\delta^{(k)}\alpha^{(k)}\sqrt{\phi^{(k)}(x_i)})}{K_1(\delta^{(k)}\alpha^{(k)}\sqrt{\phi^{(k)}(x_i)})},$
        $w_i \leftarrow \frac{\alpha^{(k)}}{\delta^{(k)}\sqrt{\phi^{(k)}(x_i)}} \frac{K_{-2}(\delta^{(k)}\alpha^{(k)}\sqrt{\phi^{(k)}(x_i)})}{K_{-1}(\delta^{(k)}\alpha^{(k)}\sqrt{\phi^{(k)}(x_i)})},$
$i = 1, ..., N$
    Compute:
    M-Step - Update the parameters using $s_i$ and $w_i$ from previous step
        $\overline{s} \leftarrow \sum_{i=1}^{N} \frac{s_i}{N},$
        $\widehat{M} \leftarrow \overline{s},$
        $\widehat{\Lambda} \leftarrow \frac{N}{\sum_{i=1}^{N} \left(w_i - \frac{1}{\widehat{M}}\right)},$
        $\delta^{(k+1)} \leftarrow \sqrt{\widehat{\Lambda}},$
        $\gamma^{(k+1)} \leftarrow \frac{\delta^{(k+1)}}{\widehat{M}},$
        $\beta^{(k+1)} \leftarrow \frac{\sum_{i=1}^{N} x_i w_i - \overline{x} \sum_{i=1}^{N} w_i}{N - \overline{s} \sum_{i=1}^{N} w_i},$
        $\mu^{(k+1)} \leftarrow \overline{x} - \beta^{(k+1)} \overline{s},$
        $\alpha^{(k+1)} \leftarrow \sqrt{\left(\gamma^{(k+1)}\right)^2 + \left(\beta^{(k+1)}\right)^2}.$
    Compute:
    $L^{(k+1)}$ using Eq. (3.14) with updated parameters
    if $\left|\frac{L^{(k+1)} - L^{(k)}}{L^{(k)}}\right| < 10^{-10}$ then
        break
    end if
end for
```

equations,

$$\gamma^* = \left(\frac{1}{k} \sum_{i=1}^{k} \log \left(\frac{\Phi_i}{\Phi_{k+1}} \right) \right)^{-1}, \tag{3.18}$$

where $k = 1, ..., N-1$. It is customary to estimate γ^* value through a graphical representation called a Hill plot. This type of depiction is formed by plotting γ^* versus k [82]. For EV distributions (e.g., Pareto), estimated γ^* are expected to stabilize (i.e., exhibit non-fluctuating behavior) with increasing values of k. Illustrated in Figure 3.7, for a dataset depicting internet traffic, it can be seen that the tail behavior exhibits a "quasi" power-law behavior as the estimated γ^* value plateaus. The magnitude of the tail-index value provide insight into the heaviness of the tails, in which small values signify heavy tails while larger γ^* values indicate a decrease in the slope of the pdf tails as shown on a log-linear plot. Additional usage will also provide a quantitative method for estimating the maximum order statistic achievable given a particular PDF of a given increment, as discussed more in Chapter 4 and Chapter 7.

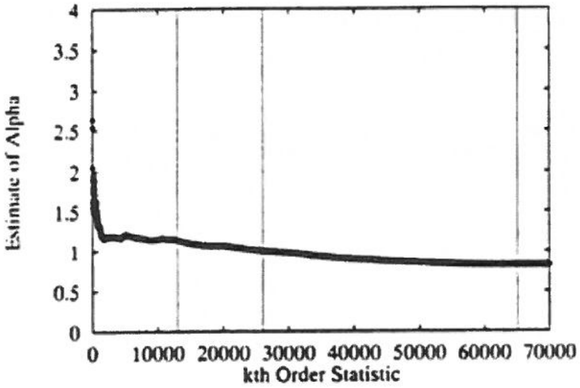


Figure 3.7: Example of a Hill plot for the transmittion times of web files. The asymptotic value occurs at $\alpha \sim 1$, signifying a Pareto-type distribution (adapted from [72]).

Chapter 4

Estimating Higher-Order Structure Functions from Geophysical Turbulence Time-Series: Confronting the Curse of the Limited Sample Size

4.1 Introduction

[1]Characterization of small-scale turbulence by higher-order statistical moments has a long and rich history in both laboratory and geophysical settings. Dating back nearly 50 years, Van Atta and Chen [257] conducted an extensive atmospheric boundary layer (ABL) experiment to measure up to fourth-order moments of atmospheric velocity fluctuations over the ocean. Their study revealed some intriguing similarities, and differences, between what was observed

---

[1] The material presented in this chapter is accepted and in press as the following publication: DeMarco, A.W. and Basu, S. (2017) Estimating Higher-Order Structure Functions from Geophysical Turbulence Time-Series: Confronting the Curse of the Limited Sample Size, Physical Review E

and the existing turbulence hypotheses [e.g., Kolmogorov and Obukhov 1941 (KO41), [138, 198]]. These findings led to increased efforts (see [15, 103, 241, 244, 250, 258, 273, just to name a few]) to acquire extensive measurements of ABL flows in order to further explore the statistical and dynamical features of turbulence (e.g., power-law scaling, nonlinearity in energy cascade). At the same time, researchers in a myriad of other geophysical disciplines, from oceanography (e.g., phytoplankton biomass distribution in turbulent coast waters [159, 224]) to magneto-hydrodynamics (e.g., intermittency in solar wind fluctuations [54, 115, 168, 210, 218]), initiated their own investigations in this intriguing research arena. Due to the lack of modern instrumentation, the majority of the initial studies were conducted solely within the Eulerian framework (by invoking the so-called Taylor's hypothesis). Fortunately, several recent contributions have attempted to fill the void in the Lagrangian framework (see [19, 55, 78, 252, and references therein]).

Over the years, the utilization of higher-order statistics not only became common practice for (in)validating various hypotheses against experimental findings, but it also enabled the research community to gain a better understanding of different types of turbulent flows. For example, many studies have been conducted which strive to make a distinction between active and passive scalars [59, 70, 141] as well as contrasting atmospheric convection [5, 172] from Rayleigh-Bénard convection [41, 57, 157, 230, 275]. For other illustrative examples, please refer to the outstanding books by Frisch [95] and Tsinober [253].

In parallel with the analysis and characterization studies, a handful of pioneering works focused on the development of cascade models with the inherent ability to capture the observed higher-order scaling behaviors. These models ranged from a simplistic β-model which seeks to resemble the intermittent behavior of turbulent cascade from a geometric point-of-view [97, 196] to a more complex probabilistic model, which describes the energy cascade (multiplicative) process within the inertial-range in a multifractal framework [39, 96, 165]. Likewise, a number of other simplified dynamical models (e.g., shell models, stochastic Burgers equation, [27, 65, 122, 158, 200]) were developed to mimic a number of intrinsic traits of three-dimensional turbulence.

From our perspective, this line of research is still far from being matured.

It is important to acknowledge that the aforementioned higher-order characterization and related modeling activities are not only of importance from a pedagogical point of view, but they are also beginning to make impacts in diverse practical applications, including (but not limited to): combustion [51, 144]; wind energy [184], and atmospheric modeling [260]. We strongly believe that the gamut of applications can be significantly broadened with further analyses of various geophysical data sets. Unfortunately, the typical sample size of such data sets is orders of magnitude fewer than their laboratory counterparts. This disparity in sample size poses a serious challenge for higher-order statistical analyses using the traditional sample moments. In this paper, we advocate an alternative approach to confront this challenge.

4.2 Limited Sample Size Problem

It is common knowledge in the turbulence research community that capturing higher-order moments of turbulent variables can be rather difficult without a substantial sample size of experimental data [12, 34, 94, 95, 219, 250, 259, 264, 267, 283]. In particular, the ability to accurately estimate the moments can be directly related to the tails of the underlying probability density function (pdf), which signify rare events [24].

Thus, in laboratory settings, it is customary to measure turbulence with upwards of 10^7 samples using hot-wire anemometry (sampling rate on the order of several kHz). However, acquisition of hot-wire data in a natural geophysical setting is quite challenging. For example, in the case of ABL field experiments, one needs to perform meticulous hot-wire calibration at short regular intervals in order to account for the ever-changing, diurnally varying ABL flow parameters [142, 170]. As a viable alternative, the ABL community widely uses sonic anemometers (sampling rate of $\sim$ 20 Hz) for measuring turbulent velocity fields. In contrast to hot-wires, these sensors require much less periodic calibration and maintenance. However, these instruments can only collect $O(20\text{–}40)$ thousand samples during a measurement time window of 15–30 min. Publications, from as early as the 1970s, pointed out that such short time series are not adequate for estimating moments beyond fourth-order [250, 257]. Unfortunately, one cannot circumvent this problem by simply using longer time-records owing to the frequent interference of non-stationary and non-turbulent motions (a.k.a mesoscale motions). An earlier work attempted to tackle this problem using cumulants [28]. It was demonstrated that a cumulant-based approach can reliably estimate the so-called intermittency exponent [95] from short ABL time series. However, the cumulants involve logarithmic functions of the velocity increments; as such, they are more influenced by the peak of the pdf rather than its tails. Thus, their usage in the estimation of higher-order moments is questionable. As an alternative, in this paper, we illustrate a maximum likelihood-based moment estimation technique which can provide statistically accurate higher-order moments from relatively short geophysical series.

4.3 Quantifying Uncertainty in Structure Function Estimates

It is customary to quantify the behavior of fine-scale fully developed turbulence, using structure function (SF) analysis, see Eq. 2.2.

Many laboratory and geophysical turbulence studies have shown that the pdfs of velocity increments, pdf$[\delta u]$, are scale-dependent and change steadily within the inertial sub-range. Specifically, these distributions have been shown to exhibit strong non-Gaussian behavior at small increments, then become more Gaussian as separation increases [24, 31, 52, 167, 168, 234, 239]. A few years ago, Barndorff-Nielsen *et al.* [24] demonstrated that the normal inverse Gaussian (NIG) distribution has the inherent ability to capture such scale-dependent traits in a parsimonious manner. An illustrative example is shown in Figure 4.1. Here, following the approach by Rydberg [217], we have numerically generated three NIG distributed variates with different parameter settings, see Section 3.1 for details. This figure clearly attests to the fact that the NIG distribution can indeed capture heavy-tailed (●), moderate-tailed (▲), and near-Gaussian (■) distributions with appropriate choice of parameters. In this work, without loss of generality, we generated NIG distributed variates with $\mu = 0$ and $\beta = 0$. This choice was made

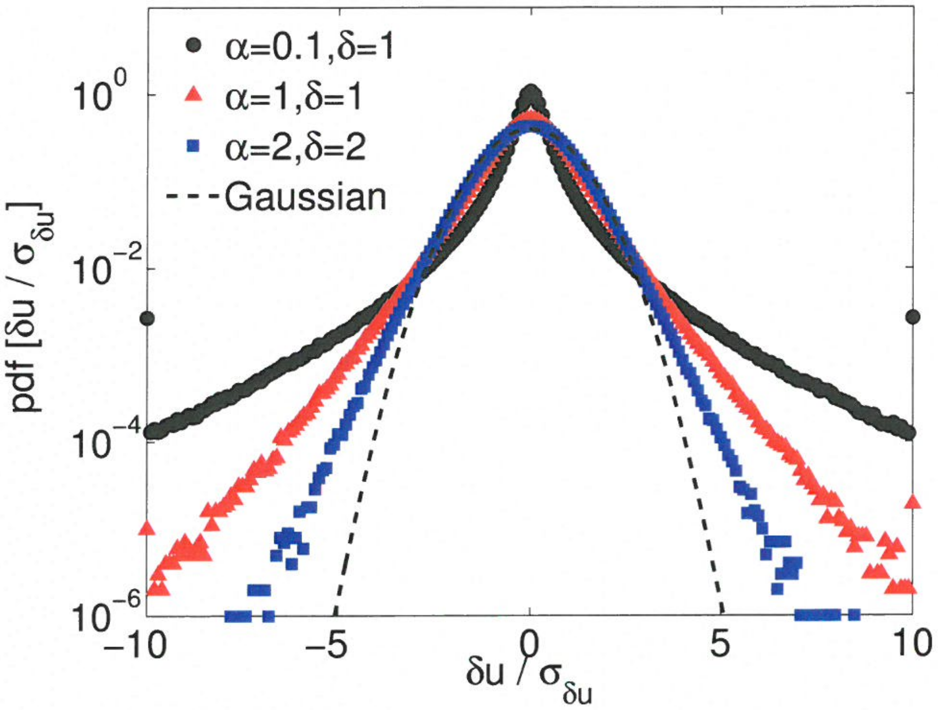


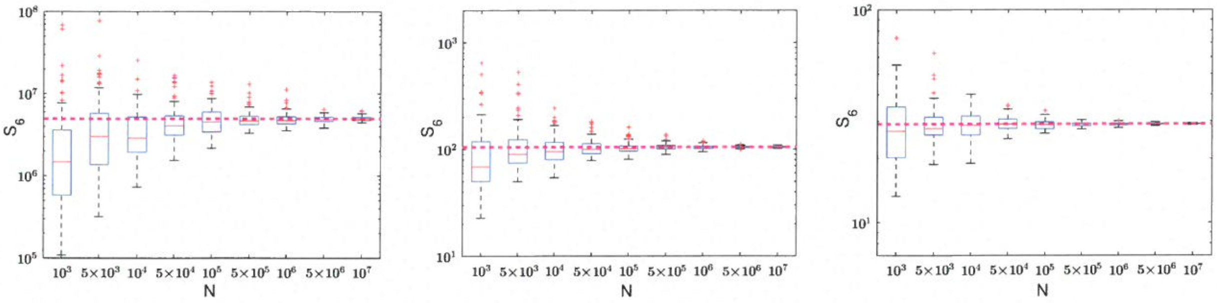
Figure 4.1: NIG distributed variates with three different parameter combinations: (a) $\alpha = 0.1, \delta = 1$, (b) $\alpha = 1, \delta = 1$, and (c) $\alpha = 2, \delta = 2$. For all these cases, the parameters μ and β are assumed to be equal to zero. For each case, 10^7 samples were generated using Rydberg's algorithm [217]. The distributions were normalized by the standard deviation $\sigma_{\delta u}$. A Gaussian pdf is overlaid (dashed line) as a reference.

so that Eq. 4.4 becomes analytically tractable [26] and reduces to:

$$S_p = E\left(|x|^p\right) = \frac{2^{\frac{p+1}{2}}\delta^p}{\pi\left(\alpha\delta\right)^{\frac{p-1}{2}}}\exp\left(\alpha\delta\right)\Gamma\left(\frac{p+1}{2}\right)K_{\frac{p-1}{2}}\left(\alpha\delta\right). \tag{4.1}$$

Next, using the NIG distributed variates, we quantify the impacts of pdf shapes and sample sizes on the uncertainty of the SF estimates. We consider a wide range of sample sizes (N) from 10^3 to 10^7 for each of the three pdfs shown in Figure 4.1. In order to obtain reliable statistics, we generate 100 realizations for each case. Based on the numerically generated variates, we compute SF using Eq. 2.2. Without loss of generality, we focus on the sixth-order SF (S_6).

The decision to consider S_6 as a test statistic was not arbitrary. In turbulence literature,



(a) (b) (c)

Figure 4.2: "Empirical" S_6 box plots for three different NIG distributions with parameter combinations: (a) $\alpha = 0.1, \delta = 1$, (b) $\alpha = 1, \delta = 1$, and (c) $\alpha = 2, \delta = 2$. The parameters μ and β are assumed to be equal to zero. The sample sizes (N) are varied from 10^3 to 10^7. One hundred realizations are used for the construction of the box plots. The dashed magenta lines represent the true S_6 values based on Eq. 4.1.

there is considerable interest in the accurate estimation of S_6, since its scaling exponent is directly used to determine the intermittency exponent, $\mu^* \equiv 2 - \zeta_6$, which relates to the behavior of the underlying non-Gaussian distributions.

The estimated (henceforth "empirical") S_6 values are shown in Figure 4.2 utilizing a standard box plot notation where 50% of the data lie within the blue "box" and the red line segment within the "box" is the median value of the data. The "whiskers" (i.e., the vertical dashed line segments) correspond to $\pm 2.7\sigma$, while the $+$ are the outliers. The following observations can be made based on this figure:

- For all the cases, with increasing sample size the S_6 estimates converge towards the true values as would be expected.

- For comparable sample size, the uncertainty of the S_6 estimate is much higher for the heavy-tailed case than the near-Gaussian case; the moderate-tailed case falls in between.

- For small sizes, the S_6 estimates are strongly biased for the heavy-tailed case. The bias decreases for the moderately tailed case, while for the near-Gaussian case, the estimates are close to unbiased.

These findings have significant implication on geophysical data analyses and require scrutiny in to whether a collection of samples are robust enough to provide an accurate estimation of a specific higher-order moment. In a recent paper, Dudok Dudok de Wit [83] made an interesting contribution in this arena by borrowing ideas from the Extreme Value (EV) theory. He proposed a simple approach to compute the maximum moment order (p_{max}) for which reliable estimation of structure function ($S_{p_{max}}$) is feasible for a given spatial or time series. In the following section, using the aforementioned NIG distributed variates, we elaborate on Dudok de Wit's approach followed by a recommendation for certain improvement.

4.4 Estimating p_{max} from Limited Data

Let us denote the rank-ordered (in decreasing order) absolute value of the velocity increments as: $_k = |\delta u|_k$, where $k = 1, ..., N$. Dudok de Wit [83] showed that for small values of k, $_k$ versus k follows the well-known Zipf power-law behavior:

$$_k \propto \left(\frac{k}{N}\right)^{-\frac{1}{\gamma^*}}, \tag{4.2}$$

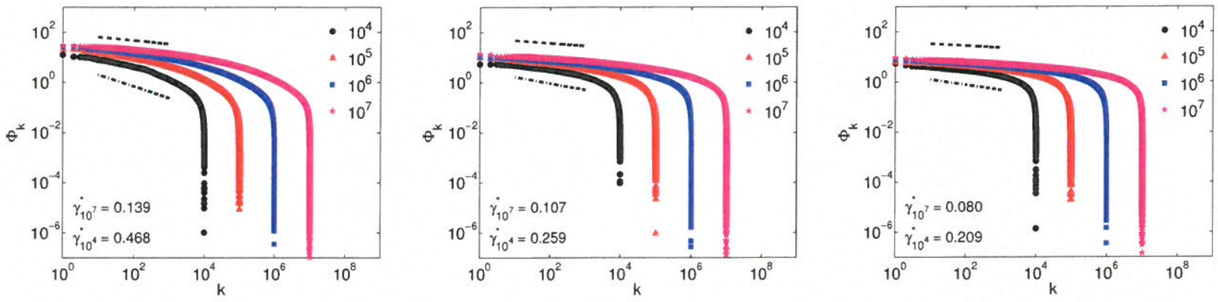
where, γ^* is known as the tail index (aka shape parameter). Via analytical derivations (with one minor approximation), Dudok de Wit [83] related γ^* to p_{max} as follows:

$$p_{max} = \left\lfloor \frac{1}{\gamma^*} \right\rfloor - 1, \tag{4.3}$$

where the floor bracket denotes the integer part.

In Figure 4.3, the rank-ordered plots for three NIG distributed variates are shown. In these log-log plots, the power-law behavior is clearly discernible for small values of k. Following Dudok de Wit's approach, γ^* values are estimated using ordinary linear regression over the range of $10 \leq k \leq 1000$. These values along with the estimated p_{max} values are reported in Table 4.1.

As shown in [83], and illustrated here in Table 4.1, there is a clear dependence on p_{max} values



(a) (b) (c)

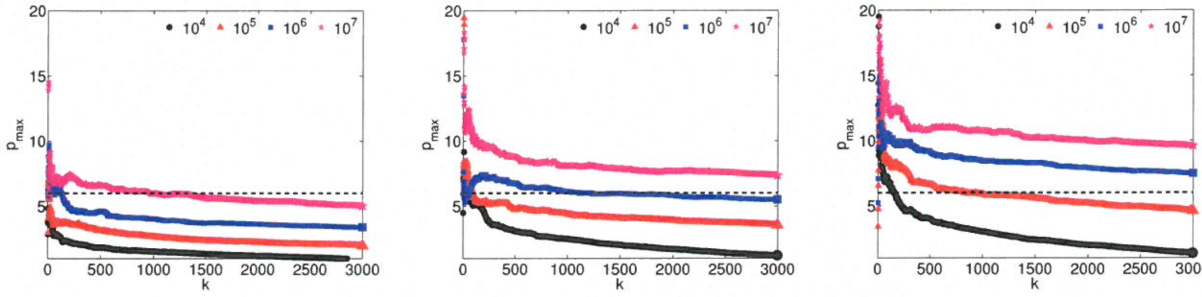
Figure 4.3: Rank-order (a.k.a. Zipf) plots for NIG distributed variates with three different parameter combinations: (a) $\alpha = 0.1, \delta = 1$, (b) $\alpha = 1, \delta = 1$, and (c) $\alpha = 2, \delta = 2$. The parameters μ and β are assumed to be equal to zero. The sample sizes (N) are varied from 10^4 to 10^7. The tail-indices (γ^*) are estimated for $N = 10^4$ (dot-dashed) and $N = 10^7$ (dashed) and reported in the bottom-left corner of the plots.

Table 4.1: Tail-index (γ^*) and maximum moment order (p_{max}) for NIG distributed variates of varying sample sizes and with three different parameter combinations.

	$\alpha = 0.1, \delta = 1$		$\alpha = 1, \delta = 1$		$\alpha = 2, \delta = 2$	
N	γ^*	p_{max}	γ^*	p_{max}	γ^*	p_{max}
5×10^3	0.583	0	0.346	1	0.265	2
1×10^4	0.468	1	0.259	2	0.209	3
5×10^4	0.344	1	0.206	3	0.157	5
1×10^5	0.294	2	0.185	4	0.140	6
5×10^5	0.213	3	0.146	5	0.117	7
1×10^6	0.204	3	0.130	6	0.109	8
5×10^6	0.155	5	0.109	8	0.088	10
1×10^7	0.139	6	0.107	8	0.080	11

for the various sample sizes. Additionally, we also notice that the shape of the pdfs directly influences p_{max}. For instance, in the case of the heavy-tailed distribution ($\alpha = 0.1$ and $\delta = 1$), a minimum sample size of 10^7 is required in order to provide a reliable estimate of S_6. However, for a near-Gaussian distribution ($\alpha = 2$ and $\delta = 2$), only 10^5 samples are needed. We would like to note that these results are in complete agreement with the ones reported in Figure 4.2.

Even though the overall approach of Dudok de Wit [83] is quite elegant and powerful, the tail index estimation component is rather subjective and not statistically robust. It can be significantly improved by employing one of the well-tested estimators from the field of EV theory [85] instead of using linear regression. In this work, we use the popular Hill estimator as described in Section 3.6. For EV distributions (e.g., Pareto), estimated γ^* (or, related p_{max} via Eq. 4.3) are expected to stabilize (i.e., exhibit non-fluctuating behavior) with increasing values of k.



(a) (b) (c)

Figure 4.4: Hill plots for NIG distributed variates with three different parameter combinations: (a) $\alpha = 0.1, \delta = 1$, (b) $\alpha = 1, \delta = 1$, and (c) $\alpha = 2, \delta = 2$. The parameters μ and β are assumed to be equal to zero. The sample sizes (N) are varied from 10^4 to 10^7. The dashed black lines represent $p_{max} = 6$.

To illustrate this technique, we present, in Figure 4.4, Hill plots using the same three NIG distributed variates and range of sample sizes as in Figure 4.3. Even though it captures heavy tails [2], a NIG distribution is not formally an EV distribution. Thus, perfect stabilization of the Hill plots is not anticipated. Nonetheless, Figure 4.4(a), shows that all four curves smoothly and slowly decrease below the dashed line, representing $p_{max} = 6$. This behavior indicates that, for this particular heavy-tailed distribution, even with 10^7 samples we cannot provide an accurate estimation of S_6. However, as the pdf shapes are modified and tend to become more Gaussian, the Hill plots [Figures 4.4(b) and 4.4(c)] reveal that higher order moments are achievable for smaller sample sizes. Another important feature that Figure 4.4 portrays is that sixth-order moment estimations pose issues with sample size less than 10^5 irrespective of the shape of the distribution. In fact, this characteristic can also be seen through the "empirical" sixth-order SF

---

[2] It is mathematically more accurate to say that the NIG distributions capture semi-heavy tails.

calculations displayed in Figure 4.2.

In summary, the approach proposed by Dudok de Wit [83] and our suggested modification provide guidelines into the p_{max} values attainable by the conventional structure function analysis. In the next section, we advocate an alternative approach, based on maximum likelihood estimation (MLE), which has the ability to reliably compute structure functions well beyond p_{max}. However, to employ such an approach, one needs to make an assumption about the underlying pdf of the velocity increments. In the present work, we assume it to be following the NIG distribution [24].

4.5 Maximum Likelihood-based Structure Function Estimation

If the velocity increments follow the NIG distribution, the structure functions can be re-written as:

$$S_p = E\left(|x^p|\right) = \int_{-\infty}^{\infty} |x|^p f_{NIG}(x; \ ; \ ; \ ; \ )\, dx: \tag{4.4}$$

In this case, x equals u If the parameters of the NIG distribution are known (or estimated), one can simply use a numerical integration approach to compute the integral.

However, in a typical laboratory or geophysical setting, these NIG parameters are not known *a priori*. One can then use the method of moments approach (see Section 3.1) to estimate them, but this technique has been shown to be unstable [131]. A more robust approach would be to use the maximum likelihood estimator (MLE; [130, 186]), for details see Section 3.5. Since, MLE is a computationally very expensive technique and occasionally requires numerous iterations for convergence, in this work, we make use of the MME-based estimates as the initial conditions for the MLE computations.

In order to illustrate the superiority of MLE over MME, we make use of the Kolmogorov-Smirnov (K-S) statistic based on the empirical distribution function (edf; [73]):

$$D = \sup_{x} |F_{EMP}(x) - F_{NIG}(x)| \ ; \tag{4.5}$$

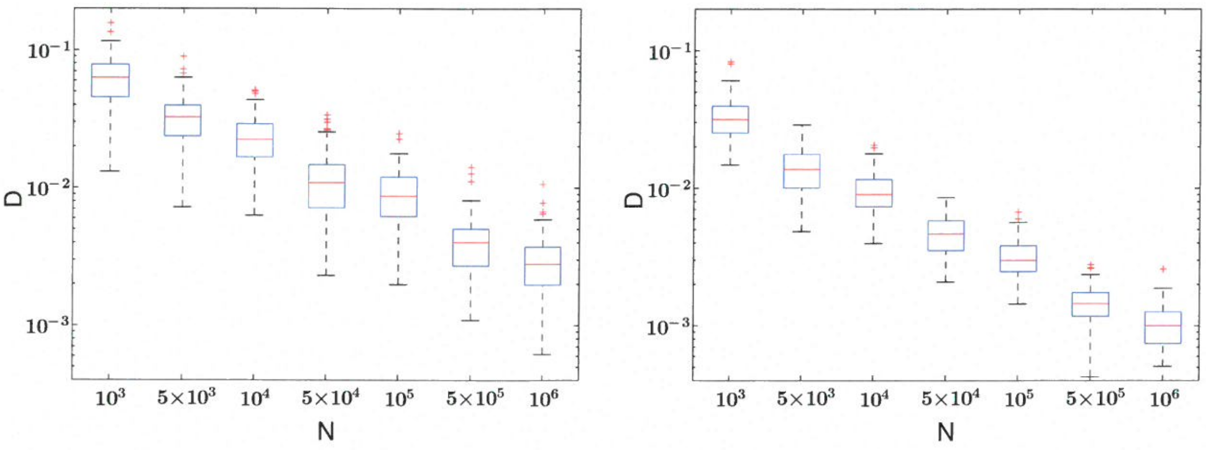
where F_{EMP} and F_{NIG} are the empirical and estimated NIG distribution functions, respectively. The results associated with a heavy-tail distribution ($\alpha = 0.1, \delta = 1$) are shown in Figure 4.5. Without any doubt, for all sample sizes, MLE outperforms MME by having (i) lower (median) values of D, and (ii) reduced uncertainty bounds of D. Similar comparative results were obtained for moderate-tail and near-Gaussian pdfs as well (not shown here for brevity).

Within the MLE computation, we explored two types of numerical algorithms: Expectation-Maximization (E-M) [80] and Nelder-Mead method [190]. Both these approaches produced near-identical results (up to the fourth decimal points). Our implementation of the E-M algorithm follows [130]. We chose this specific approach due to its simplicity and robustness for the NIG distribution. A pseudo-code is included in Section 3.1. For the Nedler-Mead method, we used a commercial MLE function from MATLAB™. Further details regarding this approach can be found in Lagarias *et al.* (1998, [145]).

Using the idealized NIG distributed variates generated in Figure 4.2, we now apply the MLE approach to find out if an improvement over the conventional approach is feasible. The MLE-based S_6 estimates are displayed in Figure 4.6 using the standard box plot notation as before. Of note, the sample size range is reduced from 10^7 in Figure 4.2 to 10^6 in Figure 4.6 since MLE provides converged statistics with far fewer data samples.

There are distinct differences in results between Figure 4.2 and Figure 4.6. First of all, for all sample sizes and pdf shapes, the estimated S_6 values are nearly unbiased in Figure 4.6. Second, the uncertainty bounds are significantly reduced in Figure 4.6 in comparison to the corresponding box plots in Figure 4.2. In other words, the MLE-based structure function approach is much more robust and more reliable in comparison to the conventional structure function approach, especially in a small sample setting.

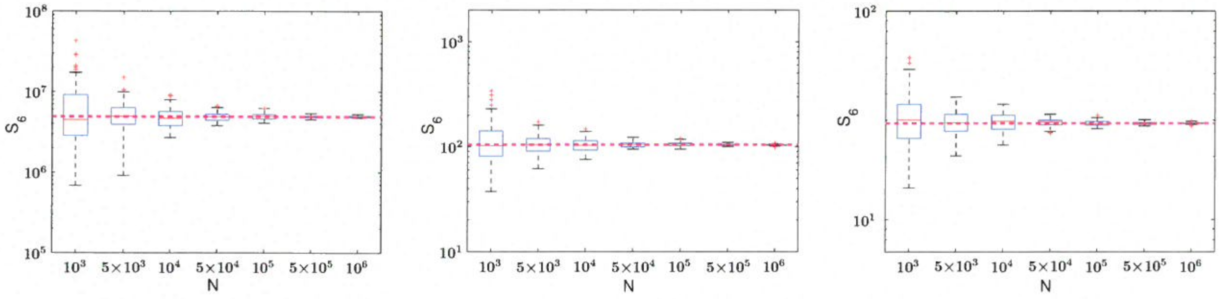
In the following section, we will investigate if this conclusion also holds in the case of correlated samples.



(a)

(b)

Figure 4.5: Box plots of K-S test statistic (D) comparing MME (left panel) versus MLE (right panel) results. The following parameters are utilized to generate the random variates: $\alpha = 0.1$, $\beta = 0$, $\mu = 0$, and $\delta = 1$ (heavy-tailed). The sample sizes (N) are varied from 10^3 to 10^6. As before, 100 realizations are used for the construction of the box plots.



(a)

(b)

(c)

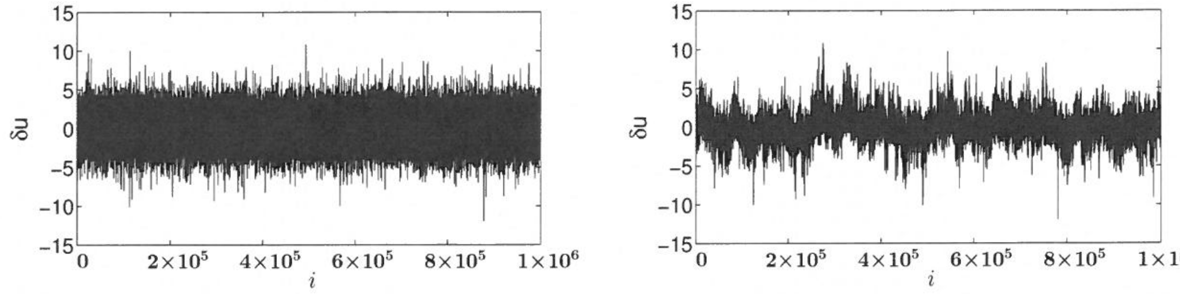
Figure 4.6: MLE-based S_6 box plots for three NIG distributions with parameter combinations: (a) $\alpha = 0.1, \delta = 1$, (b) $\alpha = 1, \delta = 1$, and (c) $\alpha = 2, \delta = 2$. The parameters μ and β are assumed to be equal to zero. The sample sizes (N) are varied from 10^3 to 10^6. One hundred realizations are used for the construction of the box plots. The dashed magenta lines represent the true S_6 values based on Eq. 4.1.

4.6 Effects of Correlation

To this point, our analysis has been focused on independent and identically distributed (i.i.d) random NIG variates. However, in practice, the samples may be correlated. For example, in the case of velocity increments, Anselmet *et al.* [12] and Huang *et al.* [117] reported scale-dependent temporal correlations. In order to quantify the impact of correlation on higher-order moment estimation, we synthetically generated 100 realizations of correlated NIG variates. Each realization is created via the following steps:

1. As before, generate i.i.d NIG variates (sample size of one million) using the approach by Rydberg [217]. We assumed a moderate-tailed distribution (=1, = 0, = 0, and =1).
2. Generate a red-noise series (sample size of one million) with a spectral slope between zero and minus 1. The magnitude of this slope is randomly selected from a uniform distribution.
3. Perform simple histogram matching between the i.i.d NIG series and the red-noise series. The resultant series preserves the NIG pdf accurately, and it is correlated. However, it does not follow the exact correlation structure of the red noise.

Illustrative examples of i.i.d and correlated NIG variates are shown in Figure 4.6. Even though they look distinctly different (due to correlation structure), both realizations follow identical pdf.

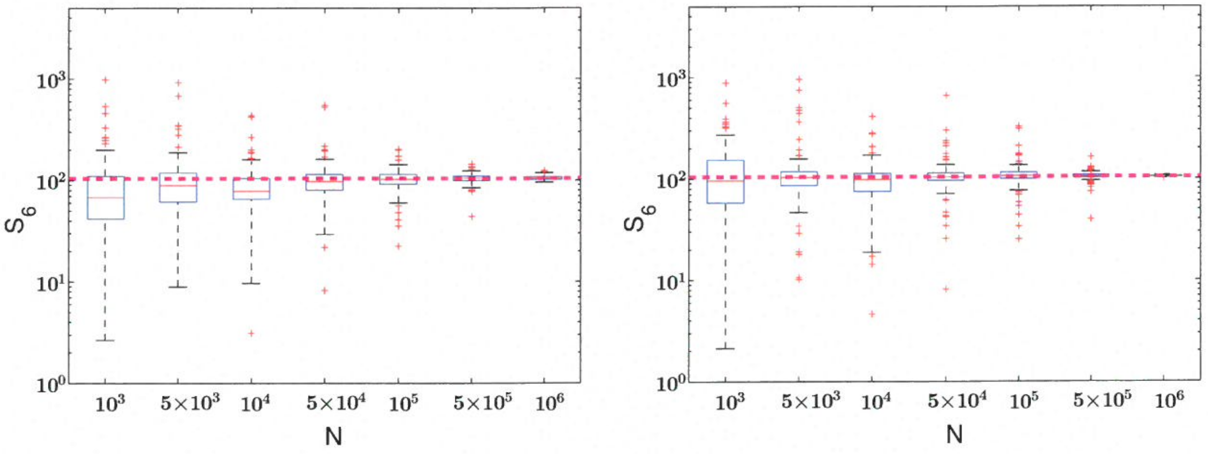


(a) (b)

Figure 4.7: Realizations of i.i.d (top panel) and correlated (bottom panel) NIG variates. Both realizations follow the same NIG distribution with prescribed parameters: =1, = 0, = 0, and =1.

Let us denote each of the 100 correlated NIG realizations as: $u_1^{(k)}$,..., $u_M^{(k)}$; where M $= 10^6$ and k varies from 1 to 100. From each realization, we extract a contiguous subset $u_i^{(k)}$,..., $u_{i+N-1}^{(k)}$;

where $i \geq 1$ and $i + N - 1 \leq M$. We vary the subsample size N from 10^3 to 10^6. Since we have a total of 100 realizations, for a specific value of N, we also get 100 subsamples. These subsamples are used to estimate S_6 using the aforementioned \empirical" and MLE approaches. These results are shown in Figure 4.8.



(a) (b)

Figure 4.8: \Empirical" (top panel) and MLE-based (bottom panel) S_6 box plots for correlated NIG variates. The following parameters are utilized to generate the random variates: $\alpha = 1$, $\beta = 0$, $\mu = 0$, and $\delta = 1$. The sample sizes (N) are varied from 10^3 to 10^6. One hundred realizations are used for the construction of the box plots. The dashed magenta lines represent the true S_6 value based on Eq. 4.1.

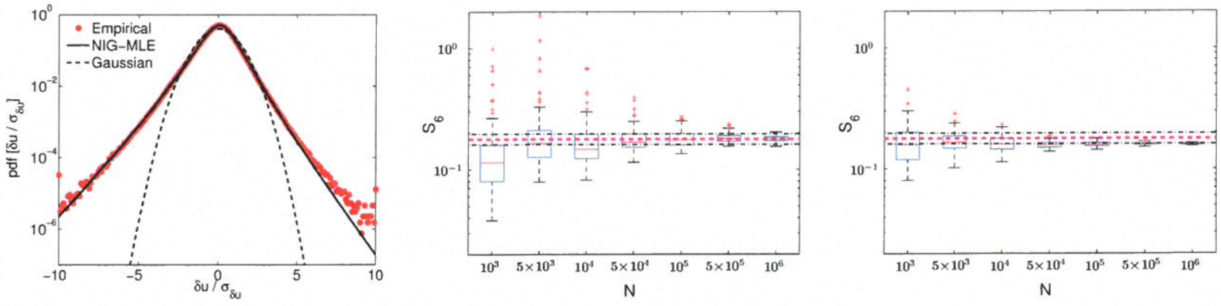
Comparing the results from Figure 4.8 to the corresponding box plots in Figure 4.2, it is clear that the uncertainty bounds have significantly increased for the correlated samples. Also, the number of outliers have increased substantially. Still, for this difficult scenario, the MLE-based estimates outperforms the \empirical" ones in terms of bias and uncertainty bounds. It is quite remarkable that for all sample sizes, the MLE-based estimates are unbiased. It is needless to say that of both the estimates do converge to the true S_6 value for large sample sizes; however, the convergence is slower than the i.i.d scenario.

4.7 Wind Tunnel Data

To further showcase the strength of the MLE approach, we utilize laboratory data from the Office National d'Etudes et de Recherches Aérospatiales (ONERA) S1 wind tunnel. These velocity measurements were obtained with a DISA 55M01 constant temperature system with Wollaston wire (3 mm diameter and 0.35 mm length for all the air flows). The sample size for this data is approximately ($\sim 13 \times 10^6$) long with a sampling rate of 25 kHz. Following the conventional turbulence analysis techniques, we have invoked Taylor's hypothesis to convert this time series to a spatial series and normalized the original series to have zero mean and unit standard deviation. For further details on this data set and past ONERA S1 wind tunnel experiments, please refer to [98, 127, 164].

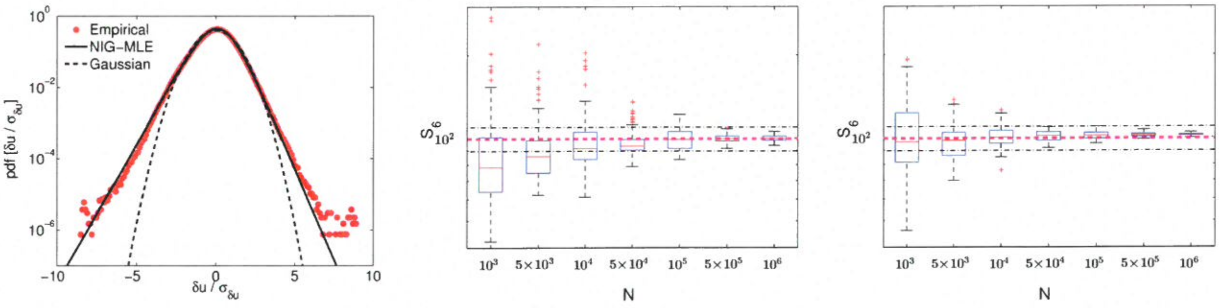
Figures 4.9(a) and 4.10(a) depict pdfs of velocity increments for two specific temporal separations (τ): 4×10^{-4} sec and 1.35×10^{-2} sec, respectively. Given their distinctly different tail behavior, we utilize them to evaluate the strengths (weaknesses) of the proposed MLE-based structure function estimation approach. As before, we randomly selected 100 subsets of varying sample sizes for both empirical and MLE-based moment estimations. We consider the S_6 values based on Eq. 2.2 using the entire data set as the "truth."

For small sample sizes, the conventional estimates are quite uncertain (middle panels of Figures 4.9 and 4.10). In addition, for such cases, the median S_6 estimates are significantly outside the ± 10% uncertainty bound. In contrast, the MLE-based estimates offer low uncertainty for all sample sizes (right panels of Figures 4.9 and 4.10). For the large τ case, the MLE-based median estimates are almost identical to the "true" value for all sample sizes. The performance is somewhat poorer for small τ. In this case, the MLE-based median values are approximately ± 10% lower than the "truth" for all sample sizes. For small sample sizes ($N \leq 10^4$), the MLE-based approach unequivocally outperforms the conventional approach. Based on these results, we strongly recommend the geophysical community to utilize the proposed MLE-based approach for analyzing data sets with limited sample size.



(a) (b) (c)

Figure 4.9: Analyses of ONERA S1 wind tunnel data. Selected increment: $r = 8.2$ mm or $\tau = 4 \times 10^{-4}$ sec, (a) corresponding PDF with NIG-MLE fit, (b) "Empirical" S_6 box plots, (c) NIG-MLE S_6 box plots. One hundred realizations are used for the construction of the box plots. The dashed magenta line represent the S_6 values based on Eq. 2.2 using the entire data set. The dot-dashed lines represent an uncertainty of ± 10% around the magenta line.



(a) (b) (c)

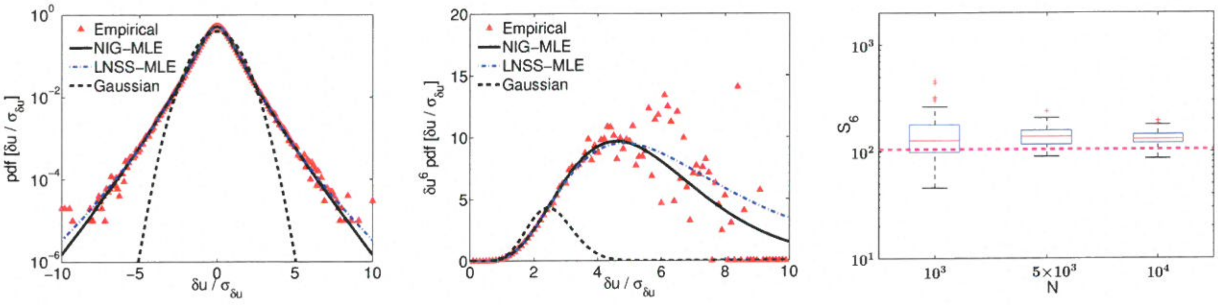
Figure 4.10: Same a Figure 4.9, except for $r = 277.8$ mm or $\tau = 1.35 \times 10^{-2}$ sec.

4.8 Alternative Probability Density Functions

Throughout this paper, we have demonstrated the strengths of the MLE-based approach to estimate higher-order moments. For this purpose, we needed to make an assumption about the underlying pdf of increments. Given its versatility in capturing pdfs of different shapes and forms [86], we opted to use the NIG pdf.

We acknowledge that the NIG pdf has not been fully vetted by the turbulence research community at large. Thus, its appropriateness to represent velocity and/or scalar increments for diverse types of flows remains to be confirmed. In the recent past, other types of pdfs have been proposed (e.g., [32, 56, 128]), and some of them have already gained popularity. In this section, we would like to find out if our MLE-based moment estimation approach can be utilized in conjunction with another type of pdf.

In this proof-of-concept exercise, we choose to use the log-normal superstatistics (LNSS) pdf [31, 32, 33] given its strikingly different form compared to the NIG pdf. First, we generate i.i.d. NIG variates ($\alpha = 1, \beta = 0, \mu = 0, \delta = 1$) of different sample sizes. We then estimate S_6 by utilizing the MLE approach. However, unlike before, we make an (incorrect) assumption that the underlying distribution is governed by the LNSS pdf. The results are shown in Figure 4.11. The top-panel plot shows that the difference between the NIG and LNSS pdfs are remarkably small (except near the tails). Visually (without using any metrics), both the pdfs appear to be equally representative of the generated random variates. More surprisingly, the LNSS-MLE approach overestimates the true S_6 value only by a small margin (see bottom panel of Figure 4.11). Undoubtedly, the most important finding is that the estimation bias remains independent of sample size. This result re-confirms the strength of the proposed MLE approach for small sample setting.



(a) (b) (c)

Figure 4.11: Top panel: comparison of fitted NIG and LNSS pdfs for i.i.d NIG variates. The sample size is 10^6. Bottom panel: MLE-based S_6 box plot for i.i.d. NIG variates with the (incorrect) assumption of LNSS as the underlying pdf. Due to high computational costs associated with numerical integration of the LNSS pdf, the sample sizes (N) are only limited to 10^4 in this example. The following parameters are utilized to generate the random variates: $\alpha = 1$, $\beta = 0$, $\mu = 0$, and $\delta = 1$. One hundred realizations are used for the construction of the box plot. The dashed magenta line represents the true S_6 value based on Eq. 4.1.

4.9 Concluding Remarks

Maximum likelihood estimation (MLE) is an age-old technique and is utilized in countless disciplines. For reasons unknown, the turbulence community has not leveraged on this statistically robust approach for the estimation of higher-order structure functions. In this work, we demonstrate that unbiased estimation of sixth-order structure function with relatively small samples (less than 10^5) is feasible with MLE, whereas, the conventional approach fails under these circumstances. Moreover, the unbiasedness of the MLE approach is not affected by correlation within the samples. Last, this approach performs remarkably well for two entirely different types of pdfs: normal inverse Gaussian and log-normal superstatistics. In our future work, we will continue to investigate the prowess of the MLE-based higher-order moment estimation approach by systematically using other observed data sets from turbulence to larger-scales motions. Additionally, we will inter-compare various potential pdf candidates using rigorous metrics (e.g., Kolmogorov-Smirnov and Anderson-Darling).

Chapter 5

Intercomparison of Probability Density Functions of Small-Scale Velocity Increments

5.1 Introduction

The understanding and characterization of turbulence begins from a fine scale perspective. It is believed that turbulence within the so-called inertial range (IR) behaves in a universal manner and can possibly be predicted from a statistical perspective. However, as discussed in Chapter 2, we are still lacking a general consensus into whether or not turbulence can be represented by a universal underlying pdf. The implications of having this knowledge extends beyond the research and continues to be of great importance to many application including wind energy, mesoscale modeling and optical wave propagation. In this work, an examination into the small-scale velocity increment pdfs will be conducted utilizing high Reynolds number (RE) data obtained from controlled laboratory (wind-tunnel) and near-surface atmospheric environments.

Within this study, we showcase four unique model pdfs, as introduced in Chapter 3, and their ability to fit the experimental data. With the use of maximum likelihood estimation (MLE) techniques and two robust Goodness of Fit (GoF) metrics, we quantitatively reveal the accuracy of the fits. To delve further into the pdf model, we also display the variations of the parametric values (i.e., α, δ, etc.) over a range of spatial scales to highlight some of the similarities and differences between the turbulence datasets. In the following sections, we will briefly introduce the datasets for this analysis. Then, the need to utilize a GoF metrics to quantify the accuracy of the model will be discussed. In the final section, we will test our estimation against empirical pdfs from two turbulence datasets and reveal the GoF results as well as explore the scale-dependent variation in the pdf parameters.

5.2 Description of Data

For this study, two different velocity series are analyzed. Using both laboratory and atmospheric data provides additional information regarding impacts of external forces and universal qualities. Therefore, it is important to distinguish between the two sets of data. First, we utilize a high-quality 30 min long ($\sim 18 \times 10^6$ data points) high RE turbulence series collected at the Surface Layer Turbulence and Environmental Science Test (SLTEST) facility located in the western Utah Great Salt Lake desert under near-neutral atmospheric condition [142, 170]. For typical small-scale atmospheric turbulence measurements a maximum of 30 minutes data is used which is believed to ensure stationarity of the flow and minimizes the larger scale influences. A fast-response (10 kHz) hot-wire sensor was used at a height of 2.1 m above ground level. Various characteristics of this series include integral time scale ($T_i = 5.71$ s) and length scale ($L_i = 34.22$ m) as defined in Eq. 5.1 as:

$$T_i = \int_0^\infty R(\tau) d\tau, \tag{5.1a}$$

$$L_i = U \cdot T_i, \tag{5.1b}$$

where $R(\tau)$ is the autocorrelation function. Next, we used a velocity data set from the Office National d'Etudes et de Recherches Aérospatiales (ONERA) S1 wind tunnel as described in Chapter 4. This data set was utilized in the previous chapter. For further details, please refer to [98, 127, 164]. We assume stationarity in both instances, therefore we also invoked Taylor's hypothesis to convert this time series to a spatial series and normalized the original series by zero mean and standard deviation $\left(\frac{u-\langle u\rangle}{\sigma_u}\right)$, which is typically performed in turbulence research. Table 5.1 shows a summary of the mean flow characteristics from the two experiments.

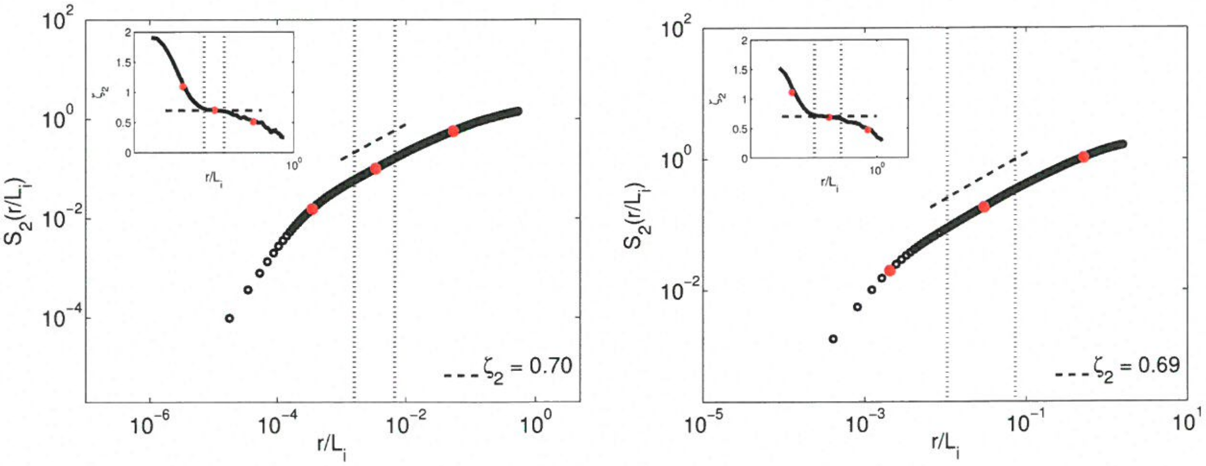
Table 5.1: Mean Flow Characteristics of the Measurements. U is the mean velocity, σ_u is the standard deviation, T_i and L_i represent the chacteristic time and length scale, respectively and f is the data sampling frequency.

Experiment	U (ms^{-1})	σ_u (ms^{-1})	T_i (s)	L_i (m)	f (kHz)
SLTEST	5.99	0.74	5.71	34.22	10
ONERA wind tunnel	20.55	1.70	0.10	2.02	25

5.3 Structure Function Analysis

Prior to discussing the pdfs of the velocity increments, it is important to observe the typical second-order structure function (S_2) representation of the turbulence data in order to gain insight into the traditional scaling behavior. Figure 5.2 shows S_2 for the two datasets and reveals that in fact there is a discernible IR for both datasets. The clear ($\zeta_2 \sim 0.70$) scaling regime can be reliably estimated between the two vertical dotted lines and it is confirmed by the local slope figures (inset). This scaling exponent also depicts the deviation from K41, $\zeta_2 = 2/3$, which is an indication of small-scale intermittent behavior within both datasets. Additionally, a more pronounced dissipation range for the high RE atmospheric data and a wider scaling regime within the IR for the wind tunnel experimental data can be observed. Finally, the atmospheric data, SLTEST, has a narrower scaling range compared to the ONERA wind tunnel data which we believe is due to the uncontrolled, natural environment, affected from large-scale motions and anisotropic forcings (i.e., buoyancy), while the wind-tunnel experiment does not possess external factors.

Within in the plots there are three red dots (●). The middle dot is within the IR while the other two represent the smaller (far left dot) and larger (far right dot) scales. Each of those selected scales represent a single velocity increment pdf and will be used to further examine how they differ over different spatial separations, r (or equivalently temporal separation, τ). The differences for the r values between the sites is due to the sampling frequency, but they approximately correspond to the same turbulent regime.



(a) (b)

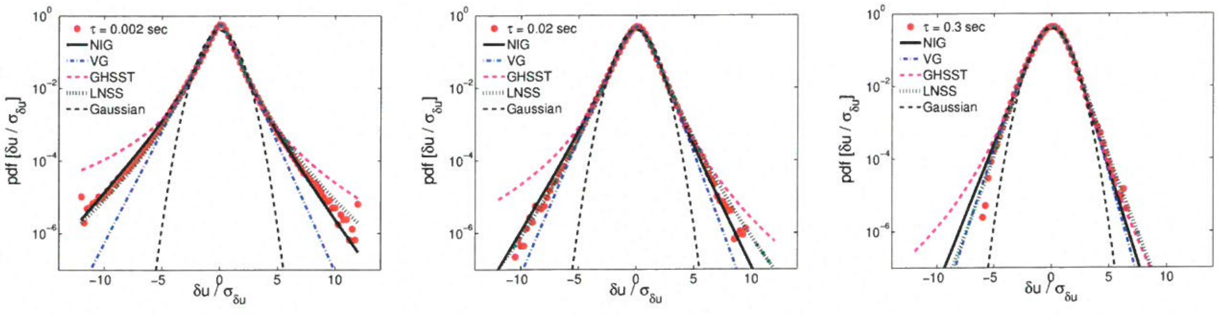
Figure 5.1: Second-order structure function (S_2) as a function of normalized spatial separation (r/L_i) from turbulence series a) SLTEST [142] and b) ONERA-Modane S1 wind tunnel [98]. The insets depict the corresponding local slopes ($\zeta_2 = d(log(S_2)/d(log(r/L_i))$ from S_2 with the dashed line representing the expected IR scaling exponent, $\zeta_2 \sim 0.70$. The scaling exponent can be reliably estimated within the vertical dotted lines. The red dots were selected to represent three distinctly different regimes within the turbulence signal.

5.4 Probability Density Function Results

Next, the following figures depict the pdfs of the normalized velocity increments for the two different datasets over three ranges of normalized spatial separation (r/L_i), as indicated by the red dots in Figure 5.1. Also, overlaid in these figures are the four pdf models that were estimated based on the observed data using the approaches discussed in Chapter 3. The first observation as shown in Figures 5.2 and 5.3 via a log-linear representation is that there is a clear deviation from a Gaussian distribution for small separations (Figures 5.2a and 5.3a). However, as r increases the pdfs begin to approach Gaussian distribution. Additionally, there is also a marginal negative skewness that is discernible in these plots, as indicated by the slight leftward shifting of the tails. As discussed previously, these features are not surprising and are well documented. Visually, all four models fit the observed pdfs to some degree of accuracy; however, there are subtle differences that should be noted. For instance, Figures 5.2a and 5.3a, representing small separation, the GHSST model shows a drastic over-prediction near the tails of the distributions, which is also observed at larger separations (Figures 5.2c and 5.3c). Furthermore, the VG model in Figures 5.2a and 5.3a, depicts a significant under-prediction near the tails resulting in a linear and rapid decay from high to low probability. The remaining fits (i.e., NIG and LNSS) are quite accurate and may be considered as reliable estimations.

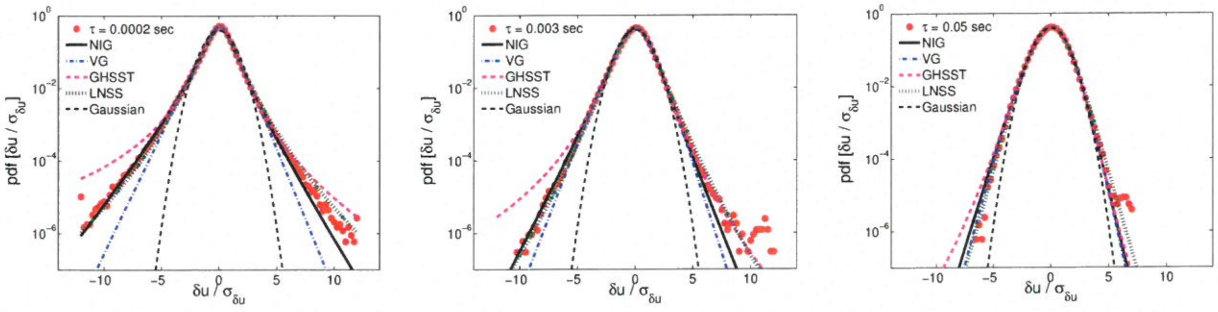
5.5 Goodness of Fit Techniques

Though it is typical to visualize pdfs of increments in this fashion, certain aspects of the pdf models, do not capture the full empirical distribution as accurate as they appear. A prevalent and widely-employed method to explore the quantitative differences between empirical distribution function (edf) from a sample and a hypothesized cumulative distribution function (cdf) are metrics referred to as Goodness of Fit (GoF) test. Interestingly, within the turbulence literature this form of statistical evalution is severely lacking. To perform these tests, one needs to convert the pdf into an edf. Let $x_1 \leq x_2 \leq ... \leq x_n$ be the order statistic of a random sample



(a) (b) (c)

Figure 5.2: The pdfs of normalized velocity increments ($\delta u/\sigma_{\delta u}$) corresponding to $r/L_i = 3.5 \times 10^{-4}$, $r/L_i = 3.5 \times 10^{-3}$, and $r/L_i = 5.4 \times 10^{-2}$ are shown in the left, middle, and right panels, respectively for the SLTEST atmospheric data. The equivalent τ values are shown in the legend. The empirical pdfs are then fitted using MLE with the four distributions described in Chapter 3. The dashed line in all figures designate the Gaussian distribution for reference.



(a) (b) (c)

Figure 5.3: The pdfs of normalized velocity increments ($\delta u/\sigma_{\delta u}$) corresponding to $r/L_i = 2.0 \times 10^{-3}$, $r/L_i = 3.0 \times 10^{-2}$, and $r/L_i = 5.1 \times 10^{-1}$ are shown in the left, middle, and right panels, respectively for the ONERA wind tunnel data. The equivalent τ values are shown in the legend. The empirical pdfs are then fitted using MLE with the four distributions described in Chapter 3. The dashed line in all figures designate the Gaussian distribution for reference.

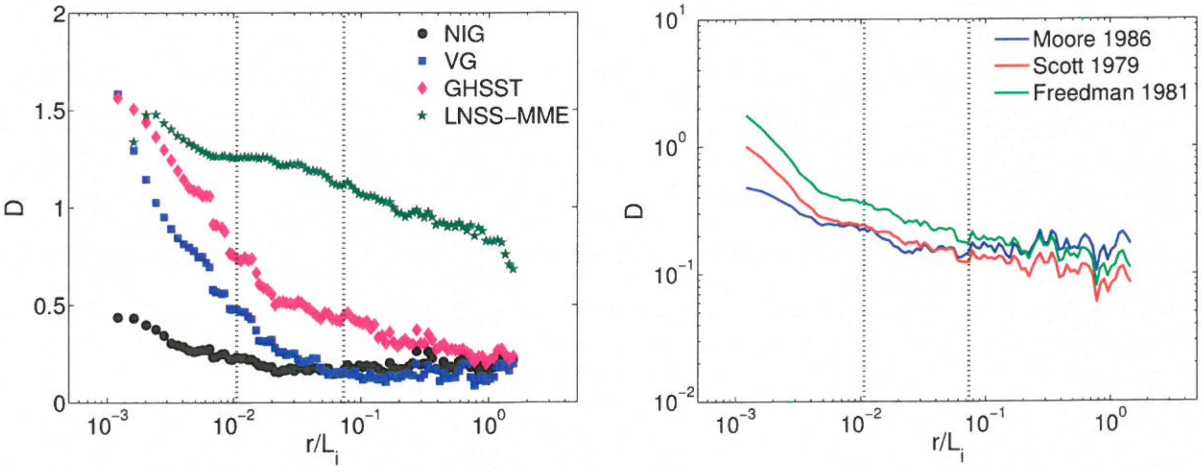
from a given pdf (e.g., pdf$[\delta u/\sigma_{\delta_u}]$). The edf is a step function calculated from a given sample of data, of size n, that provides an approximation of the underlying data:

$$F_{EMP}(x) = \frac{\text{number of samples observations} \leq x}{n}, \qquad -\infty < x < \infty. \tag{5.2}$$

As briefly introduced in Chapter 4, the Kolmogorov-Smirnov (K-S, see Eq. 4.5) statistic is one of many quantitative tests that is performed, based on the edf. Though, this edf technique is more robust since it uses the full set of data, compared to the alterative GoF tests that are commonly utilized using histogram binning (e.g., Pearson's χ^2 test). However, the choice of bin width and total number of bins has been shown to produce varying results and requires knowledge of the underlying distribution [3, 66, 222]. However, over the years, there have been a handful of proposed guidelines to be used, yet these approaches still have their shortcomings. Below, we will show an example to illustrate the impacts histogram binning has our the GoF estimation using three known binning techniques. The following itemized list depicts three formulations of binning options, which will be tested using the ONERA dataset. For each of the methods, the dynamic range (i.e., the minimum and maximum bin value) was set to the minimum and maximum of the velocity increments for the given sample data.

1. Moore (1986, [183]): $B = 2n^{2/5}$, where n is the sample size and B is the total number of bins . The bin width h is determined by $\lceil \frac{\max x - \min x}{B} \rceil$, where the ceiling bracket denotes the integer part [261].
2. Scott (1979, [222]): $h = 3.49 s n^{-1/3}$, where s is the standard deviation and h is the proposed bin width.
3. Freedman and Diaconis (1981, [90]): $h = 2 * IQR(x) n^{-1/3}$, where IQR is the interquartile range of the random variable, x.

Figure 5.4 shows two examples using histogram binning for determining the K-S statistic. In addition to the binning concerns addressed above, the use of MME as a estimation technique can also be shown as a limitation, also addressed previously in Chapter 4. For instance, Figure 5.4a shows the K-S test results (D) for the ONERA data plots as a function of r/L_i using the Moore (1986) method, item (1). Each curve represents a different model which was shown in the pdf plots in Figures 5.2 and 5.3. From a test statistic perspective, the models are all different and more or less in agreement with the visual inspection as shown in the pdf figures (i.e., NIG is better than VG and GHSST for smaller scales). However, notice that LNSS-MME is the worst from a statistical point of view, despite being very accurate from a visual perspective. This



(a) (b)

Figure 5.4: a) K-S statistic histogram binning of ONERA data compared against the various pdf models as a function of r/L_i on a log-linear plot. b) K-S evaluation using various binning techniques (described in itemized list above) for the ONERA turbulence dataset comparing the NIG distribution as a function of separation r/L_i on a log-log plot. The black symbols in (a) correspond to the blue curve in (b).

is due to the inaccuracy that exists using the MME techinque as compared to MLE, which was previously observed in Section 4.5[1]. Therefore, the use of MLE is the optimum choice for model estimation. Also, as addressed in Section 3.4, LNSS model is extremely inefficient using MLE and therefore, will not be utilized to the same extent as the other models. Now, in terms of binning issues specifically, Figure 5.4b showcases the K-S test using three different binning methods for ONERA data comparing against the NIG pdf model fit. Clearly, using different binning methods leads to different statistical results. So, to avoid the ad-hoc assumptions with histogram binning and the poor accuracy in the result, the remaining portions of this work will be conducted using edf-based statistics.

The following techniques are widely utilized in the field of statistics to compare the fit of a model to observations. For the purpose of this study we have elected to choose two, the K-S and Anderson-Darling (A-D) test statistics. The K-S test is a supremum test statistic which

---

[1] In all the previous studies (e.g., [31]) which have used this model, they have only explored the MME-based estimation technique.

is based on the maximum distance between the empirical and population distributions (e.g., NIG). This test only provides information regarding a single maximum value, instead of testing for the entire fit of the data. Thus, the best fit is obtained as $D \to 0$. More details were discussed in Chapter 4.

An alternative test called Anderson-Darling (A^2) test is a quadratic statistic used to determine whether or not a given sample of data, of size n, is drawn from a particular distribution [9, 73, 160], see Eq. 5.3:

$$A^2 = n \int_{-\infty}^{\infty} \{F_{EMP}(x) - F(x)\}^2 \psi(x) dF(x), \tag{5.3}$$

where $\psi(x) = \left[F(x)(1 - F(x))\right]^{-1}$ and $F(x)$ is the estimated distribution function of the model.

We elected to apply this statistic because it puts more emphasis and weight on the tails of the distributions, which is important for the higher order moment estimations [119, 213]. Using these metrics, we evaluate the quality of the model fit to the empirical data. Tables 5.2 and 5.3 depict the corresponding statistics for Figures 5.2 and 5.3, respectively. From the K-S statistic (D), Table 5.2 shows that NIG outperforms VG and GHSST for all the r/L_i value except 5.4×10^{-2}. While, the A-D (A^2) test reveals a very similar conclusion, though with larger magnitude results. These higher values are expected as A^2 depends on the sample size and puts more weight on the tails. Also, as r/L_i increases the statistics begin to improve, suggesting that the model fits are more in line with the observations. However, for GHSST the A^2 diminishes slightly as the tails of this distiribution never fully line up with the observations. In Table 5.3, not surprisingly, the results reveal almost identical trends. This finding demonstrates that the two datasets have similar traits and the models are able to capture this feature. From the comparison of this subset of pdfs, the NIG model reveals the most reliable fit of the turbulence pdfs. Next, we will show how the models compare over the range of scales.

Figure 5.5 shows the K-S statistics for the atmospheric dataset SLTEST in (a) and the wind tunnel data in (b). Again, the trends and magnitudes are nearly the same between the datasets

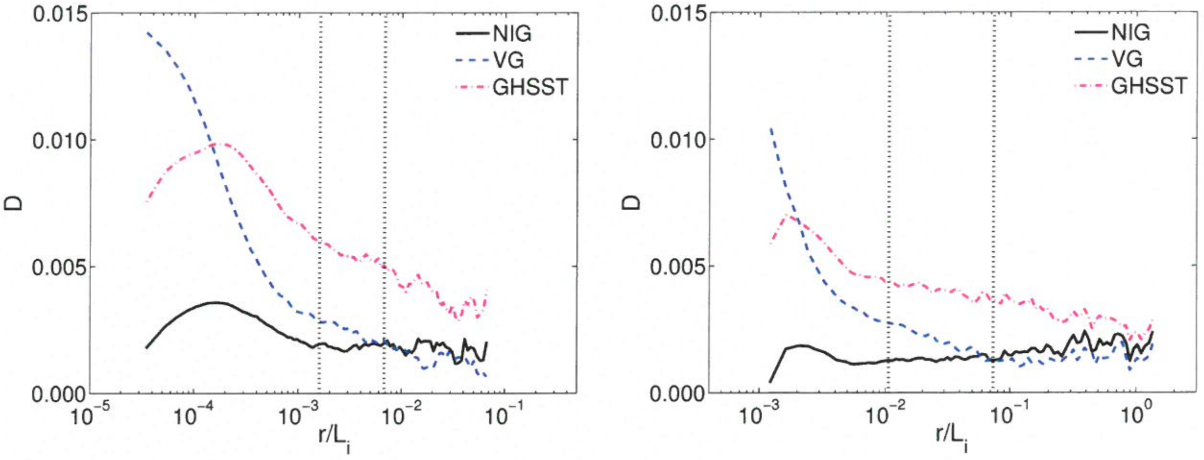
Table 5.2: Goodness of Fit Test for SLTEST data, corresponding to Figure 5.2 for three select separations

pdf Model	r/L_i	SLTEST	
		D	A^2
NIG	3.5×10^{-4}	3.11×10^{-3}	2.11×10^{2}
	3.5×10^{-3}	1.87×10^{-3}	4.35×10^{2}
	5.4×10^{-2}	1.46×10^{-3}	2.42×10^{3}
VG	3.5×10^{-4}	5.67×10^{-3}	1.23×10^{4}
	3.5×10^{-3}	2.41×10^{-3}	2.35×10^{3}
	5.4×10^{-2}	8.34×10^{-4}	1.67×10^{3}
GHSST	3.5×10^{-4}	8.93×10^{-3}	4.60×10^{3}
	3.5×10^{-3}	5.38×10^{-3}	4.55×10^{3}
	5.4×10^{-2}	3.00×10^{-3}	5.38×10^{3}

Table 5.3: Goodness of Fit Test for ONERA data, corresponding to Figure 5.3 for three select separations

pdf Model	r/L_i	ONERA	
		D	A^2
NIG	2.0×10^{-3}	1.70×10^{-3}	1.17×10^{2}
	3.0×10^{-2}	1.28×10^{-3}	5.52×10^{3}
	5.1×10^{-1}	1.96×10^{-3}	3.95×10^{3}
VG	2.0×10^{-3}	4.92×10^{-3}	3.66×10^{4}
	3.0×10^{-2}	1.99×10^{-3}	8.52×10^{4}
	5.1×10^{-1}	1.37×10^{-3}	4.95×10^{3}
GHSST	2.0×10^{-3}	6.20×10^{-3}	1.96×10^{3}
	3.0×10^{-2}	3.98×10^{-3}	2.32×10^{3}
	5.1×10^{-1}	2.93×10^{-3}	4.00×10^{3}

and the models. At the smallest scales the NIG model can capture the velocity increments more accurately than VG and GHSST. As separation increase and through the IR the model fits gradually improve, though NIG is outperforming. Then at larger r/L_i as the pdf tends towards a Gaussian distribution, the models are more in line with one another. This re-confirms the results found in the tables above and shows the full trends through the range of turbulent scales. Now, that we have identified NIG as the model of choice for these datasets, we will look closer at the parameters of the model (α, β, μ, and δ, see Section 3.1 for details) to see how they vary as a function of r/L_i.



(a) (b)

Figure 5.5: KS plots for the different distribution as a function of normalized separation distance (r/L_i) for SLTEST (a) and ONERA (b). For these computations, the velocity increments are not normalized. The inertial-range scaling exponents can be reliably estimated within the vertical dotted lines (SLTEST) and vertical dashed lines (ONERA).

Figure 5.6 shows the dependence of the four estimated NIG parameters for the two datasets SLTEST and ONERA on r/L_i. Figure 5.6a depicts α parameter which controls the steepness of the pdf. Clearly, there are difference between the two datasets at small scales indicating heavier tails near the dissipation range for the higher RE atmospheric data as compared to the laboratory experiment. However, as the scales increase near IR, α becomes more consistent between the datasets suggesting universal features within the IR. The β parameter (Figure 5.6b) which depcits skewness reveals that the atmospheric measurements (SLTEST) has a more negatively skewed behaviour in the increments, but as r/L_i increases the both approach zero and are nearly identical within their respective IR as demarcated by the vertical dashed and dotted lines. Figure 5.6c) shows the μ parameters as a function of r/L_i is an indication of the centrality of the peak. Both datasets are nearly negligible at small scales, but increase through the IR and into the large scales. The atmospheric data remains closer to zero with increasing scales compared to the laborartory. For all practical purpose, the μ parameter, which is a measure of centrality, is more or less neglected as it is close to zero thoughout the IR and small scale ranges. Finally, δ, which is the peakedness or scale parameter, has fairly consistent trends between the two datasets. These results suggest a quasi-universal behavior of the small-scale velocity increments between two different sets of experiments.

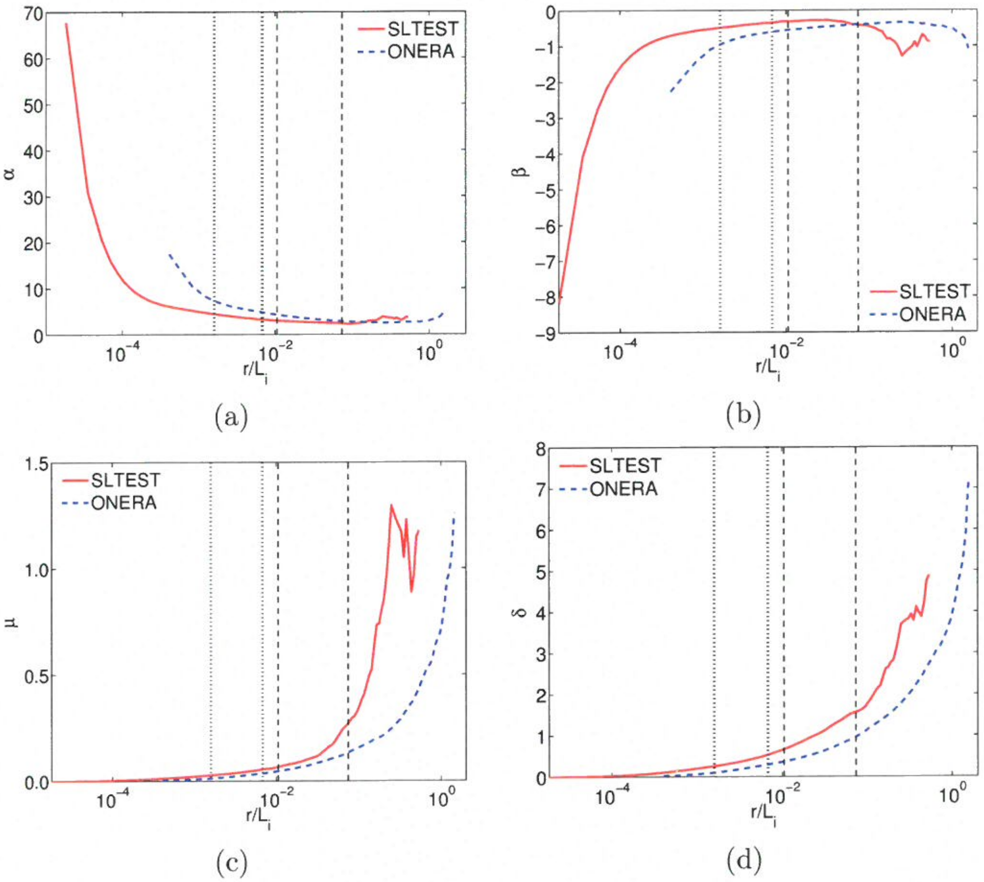


Figure 5.6: Dependence of the NIG pdf parameters on the normalized separating distance (r/L_i) for SLTEST (red-solid line) and ONERA (blue-dashed line). For these computations, the velocity increments are not normalized. The inertial-range scaling exponents can be reliably estimated within the vertical dotted lines (SLTEST) and vertical dashed lines (ONERA).

5.6 Conclusion

The turbulence data presented and analyzed here shows remarkable similiarities between the two distinctly different experimental settings. The pdfs of each dataset revealed highly non-Gaussian distributions at small scales outside of the IR with moderately heavy tails within the IR. As the scales increased near the energy input scale the distributions become near-Gaussian which is confirmation of the results from previous studies. Similarly, the pdf models showcased in this study provided evidence that these unique pdf features can indeed capture and represent turbulence datasets with great accuracy. However, the difference between the models were explored with the use of robust GoF statistics. Our results indicated that a member of the generalized hyperbolic distribution family, normal inverse Gaussian (NIG), provides the most reliable estimation to pdfs of velocity increments over a range of turbulent scales. As a result, we were able to present the parameters of NIG as a function of separation distance, which revealed similar features between the datasets and near universal characteristics with the IR. However, as flow characteristics change, a question may arise as to how do these pdf features vary. Therefore, in the next chapter, an examination into the impacts of stability will be discussed.

Chapter 6

Characterization of Small-Scale Velocity and Temperature Increments in Stably Stratified Boundary Layer Flows

6.1 Introduction

In this chapter, we showcase the impact stability has on boundary layer flows by investigating the small-scale velocity and temperature increments from a high-quality wind tunnel experiment with modifying stable stratification [201]. The atmosphere is fundamentally a stratified environment, therefore it is important to gain insight into possible implications of variations in stability conditions. From a physical perspective, the presence of stable stratification implies that when a fluid parcel is displaced from its equilibrium position, a restoring force will act on it due buoyancy. Two important conditions exist in these type of flow scenarios. First, under stratified environments the vertical motions are suppressed and the turbulent flow tends to be more anisotropic or restricted in the vertical direction, which suggests possible variations in the scaling characteristic. In 1959 and 1962, Bolgiano [45, 46] hypothesized that a secondary scaling regime exists at larger scale outside the IR due to stable stratification referred to the "buoyancy sub-range". Unlikely K41, which assumes isotropic and homogeneous conditions and energy cascade depending on energy dissipation rate, Bolgiano [45, 46] considered potential temperature variance flux ($\partial\theta'^2/\partial t$), as a key variable. Based on dimensional analysis he proposed that scaling within the buoyancy sub-range would scale as $\sim 3p/5$ and $\sim p/5$ for the velocity and temperature, respectively. Since this theory, many have experimentally tested and have found similarities in the scaling results (e.g, [4, 5]). Furthermore, when stability conditions are enhanced internal gravity wave formation and Kelvin-Helmholtz (KH) instability are supported [253] and give rise to further non-linear interactions. For instance, internal wave breaking occurs in the presence of stable stratification and is an important generation source for turbulence within geophysical flows. To date, our understanding of stably stratified environments remains incomplete which leads to poor prediction of these conditions in our atmospheric models [163, 271], thus further examination is required.

The impacts stable stratification has on turbulence has been widely studied over the years [6, 29, 45, 46, 58, 114, 162, 231, 251, 265, and references therein]. However, from a pdf perspective only a handful of studies have been conducted within atmospheric flows. For example, Thoroddsen and Van Atta (1992, [251]) found that "...the temperature gradient pdfs exhibit very distinct exponential tails over a wide range of turbulent RE number, while the corresponding pdfs of the temperature fluctuations show close to Gaussian behaviour." Likewise, Chu *et al.* (1996, [67]) examined various surface-layer velocity and temperature fluctuation pdfs and revealed similiar results. Furthermore, Alisse and Sidi (2000, [6]) examined the effects stability, in the free-troposphere, has on the scaling characteristics for horizontal wind and temperature within vertical atmospheric profiles. They showed two distinct regimes for velocity fluctuation pdfs depending on whether or not the flow was calm or turbulent, while the temperature increments pdfs remained close to each other regardless of the flow conditions. Also, their results showed that the temperature fluctuation pdfs had heavier tails compared to wind fluctuations.

Additionally, Vindel *et al.* (2011, [265]) found that as stratification becomes more stable for near-surface observations, intermittency decreases or in other words the pdf tails of small-scale velocity increments are less heavy, which is a different results from Alisse and Sidi. Investigations under more controlled environments, such as wind tunnel and large-eddy simulation experiments, revealed that the scalar (i.e., temperature) and the velocity gradient pdfs depart from Gaussian and exhibit strong exponential tails [123, 177]. Despite these studies, very few have fully examined the increment pdfs under various stable conditions, thus this work will attempt to uncover features of the temperature and velocity increment pdfs in a systemic manner.

Building on our previous results, this chapter will estimate the increment pdfs with the normal inverse Gaussian (NIG) and generalized hyperbolic skewed student's t (GHSST) distributions. From this perspective, we are able to estimate the four-parametric distribution to reveal the variations in the velocity and temperature increment series as a function of stability. Therefore, a major objective of this study is to determine, through analysis of pdfs, as to whether or not there is evidence of variations in the turbulent characteristics as stability condition vary. First, we will introduce the dataset used for this study.

6.2 Description of Datasets and Methodology

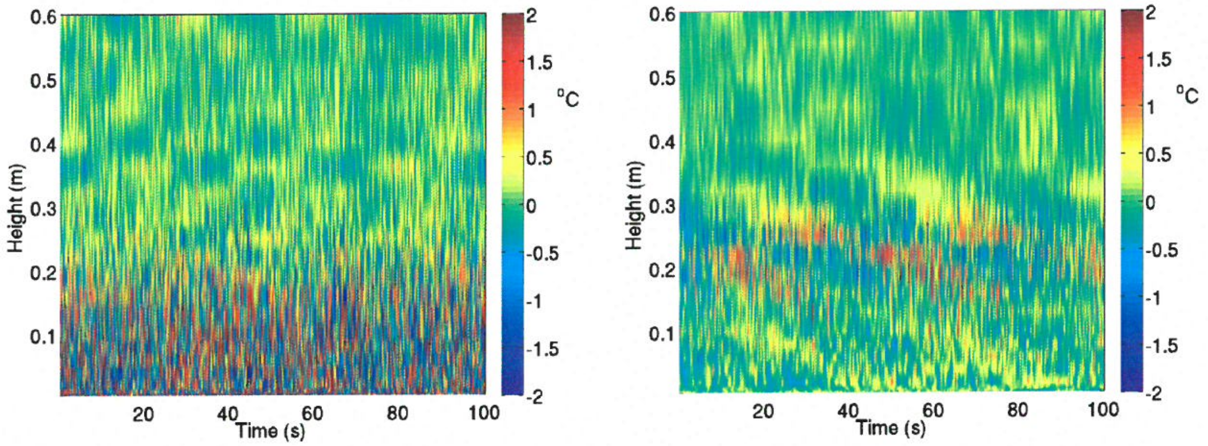
Generally, wind tunnel experiments are configured for neutrally stratified flows [272]. For this analysis, however, we examine the data from a high-quality wind-tunnel experiment conducted by Ohya (2001, [201]), for which he explored the turbulent flow characteristics (i.e., turbulent intensities and fluxes) for a variety of stability conditions, from near-neutral to strongly stable. The increased stability was imposed by modifying the temperature structure, such that the floor was gradually cooled to around 10°C, while the flow aloft was heated to 40-50°C resulting in eight different stability classes. The range of stability in terms of bulk Richardson number, Eq. 6.1 is between 0.12 and 1.31 from near-neutral to strongly stable, using:

$$Ri_h = \frac{gh\Delta\Theta}{\Theta_0 U_\infty^2}, \tag{6.1}$$

where $\Delta\Theta$ is the temperature difference between surface and ambient air, Θ_0 is the average absolute temperature over the whole boundary-layer depth (h), U_∞ is the ambient velocity and g is the gravity constant. Also, from a RE perspective, the flow configuration changes from a large RE_h of $\sim$ 50,000 to a small value of $\sim$ 25,000. The itemized list below provides some of the details pertaining to the experimental set-up, more details regarding this wind tunnel experiment can be found here [201, 202]:

- Measurements of temperature and velocity were taking using X-type hot-wire and I-type cold-wire.
- Sampling rate of 500 Hz with a low-pass filter of 200 Hz applied with a sampling period of approximately 100-150 sec leading to $\sim$ 50,000 data samples per level.
- The turbulent quantities were measured 9 m downstream from the in-flow source.
- There were 20 measurement levels from 0.005 m to 0.6 m with an average boundary layer height estimated around 0.45 m.
- For this analysis, we only considered data from within the logarithmic sub-layer, which we determined using similiar techniques as [62] which resulted in data from 0.04 m to 0.36 m.

This extensive dataset provides a myriad of flow situations conducive to examining the behavior of velocity and temperature increments of a range of separations. To illustrate the differences between various stability regimes, Figure 6.1 depicts two time versus height plots of temperature fluctuations ($T' = T - \langle T \rangle$). Under near-neutral conditions (left panel), the temperature fluctuations remain random in appearance throughout the lower portion of the boundary layer for the entire measurement period. However, as stability increases (right panel) evidence of a wave-like pattern is present in the middle of the domain. This provides visual confirmation of different flow characteristics between the experiments.



(a)

(b)

Figure 6.1: Time versus height plot of the temperature fluctuations (T') for two different flow configurations a) near-neutral and b) moderately stable.

However, in lieu of utilizing Ohya's eight stability classes for our analysis, we elected to apply four alternative stability classifications based on the local Obukhov length and the height above ground. According to Nieuwstadt (1984-1985, [192, 193, 194]) the use of Monin-Obukhov's local similarity theory (Eq. 6.2) is a more fundamental length scale under stable stratification. From this we can then use a "universal" function of a single, dimensionless stability parameter ($\zeta = z/\Lambda$) as a revised stability class structure, see Eq. 6.2:

$$\Lambda = \frac{-\overline{\Theta} u_*^3}{\kappa g (\overline{w'\theta'})}, \tag{6.2}$$

where Λ is the local Obukhov length scale, $u_* = (\overline{u'w'}^2 + \overline{v'w'}^2)^{1/4}$ is the friction velocity, k is the Von-Kármán (assumed as 0.4), and Θ is the mean temperature. A similar approach was used in Basu *et al.* (2006, [29]). For this work, the breakdown of the four new stability regimes along with the number of time series measurements within those classes are shown in Table 6.1, where S1 through S4 range from weakly stable to highly stable conditions, respectively.

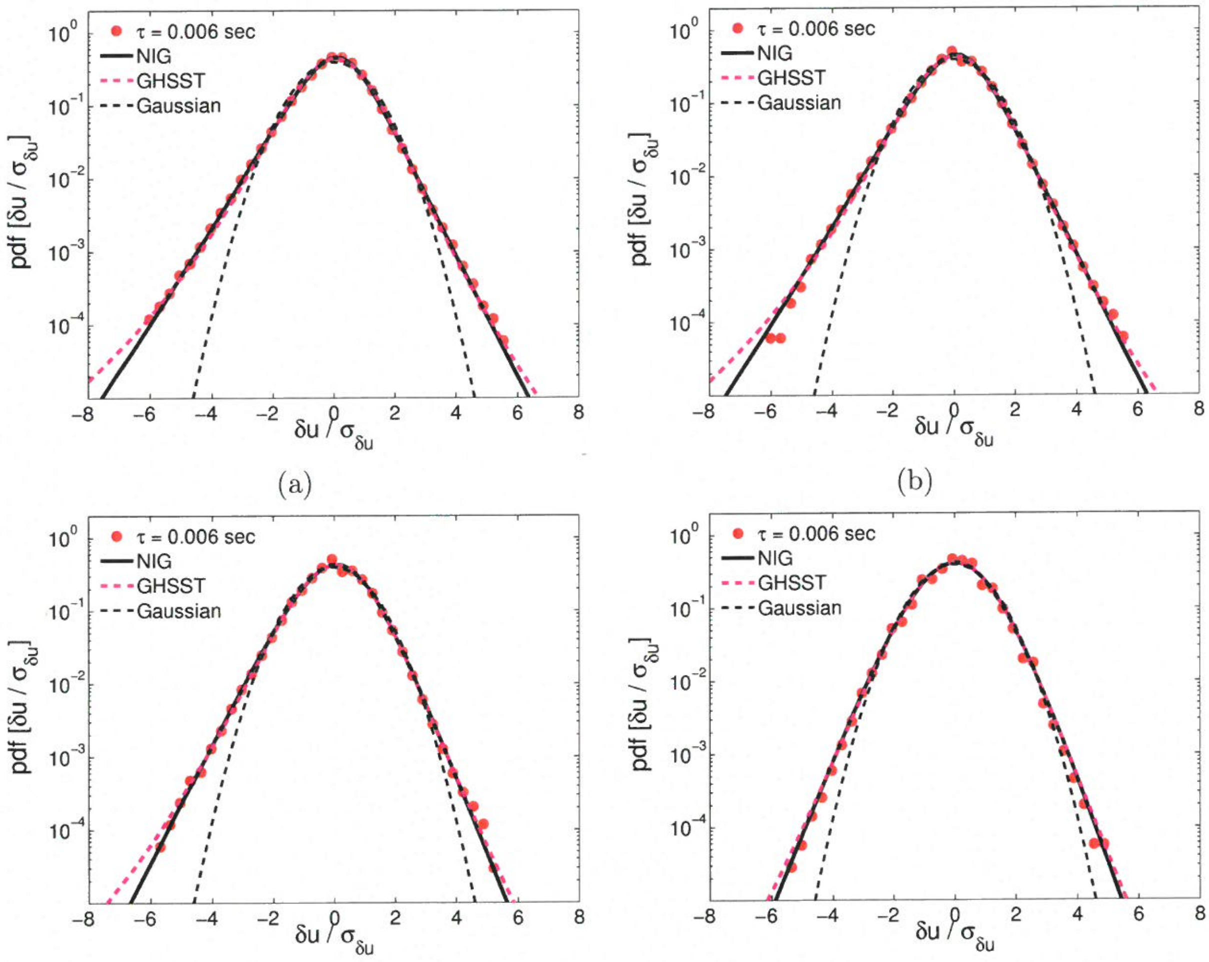
Table 6.1: Number of samples in each stability class

Class	Stability (ζ)	Wind Tunnel Measurements
S1	0.00-0.25	16
S2	0.25-0.50	17
S3	0.50-1.00	16
S4	> 1.00	17

6.3 Probability Density Function Model Fit

As a first illustrative example, Figure 6.2 shows the median velocity increment pdfs at small scales for the four different stability conditions. To compute the median pdf, we took the median of all the data at $\tau = 0.006$ secs for the 16 time-series representing S1 (weakly stable) and plotted the corresponding pdfs (e.g., Figure 6.2a). Overlaid on each figure are NIG and GHSST estimates along with a Gaussian distribution. An initial observation is that the two models provide an accurate representation of the observed data regardless of the stability class. Also, all of the observed data appear to moderately deviate from Gaussian, however, not nearly as significant as what was indicated in the previous chapter. This suggest that the amount of intermittency in this experiment is not as significant and the *RE* values are less. However, there does appear to be marginal differences between each of the stability regimes ranging from moderately heavy tail for the weakly stable environment (Figure 6.2a) to light tailed distribution in the strong stability case (Figure 6.2d). Given the change in stability one would expect a change in the velocity increment distribution at small scales. For instance, with increased stability, temporal and/or spatial concentration of the small scale turbulent structure(s) (i.e., internal intermittency) become surpressed, which can be seen here.

Furthermore, we also observe a slightly more negatively skewed pdfs as stability decreases. Typically, experimental evidence suggest that the skewness factor (see Eq. 2.2) is proportional to *RE* of the flow, ranging from -0.3 to -0.7 [181, 243, 264, 273]. This quantity is also considered to be a direct measure of vortex stretching and related to the energy transfer from large to small

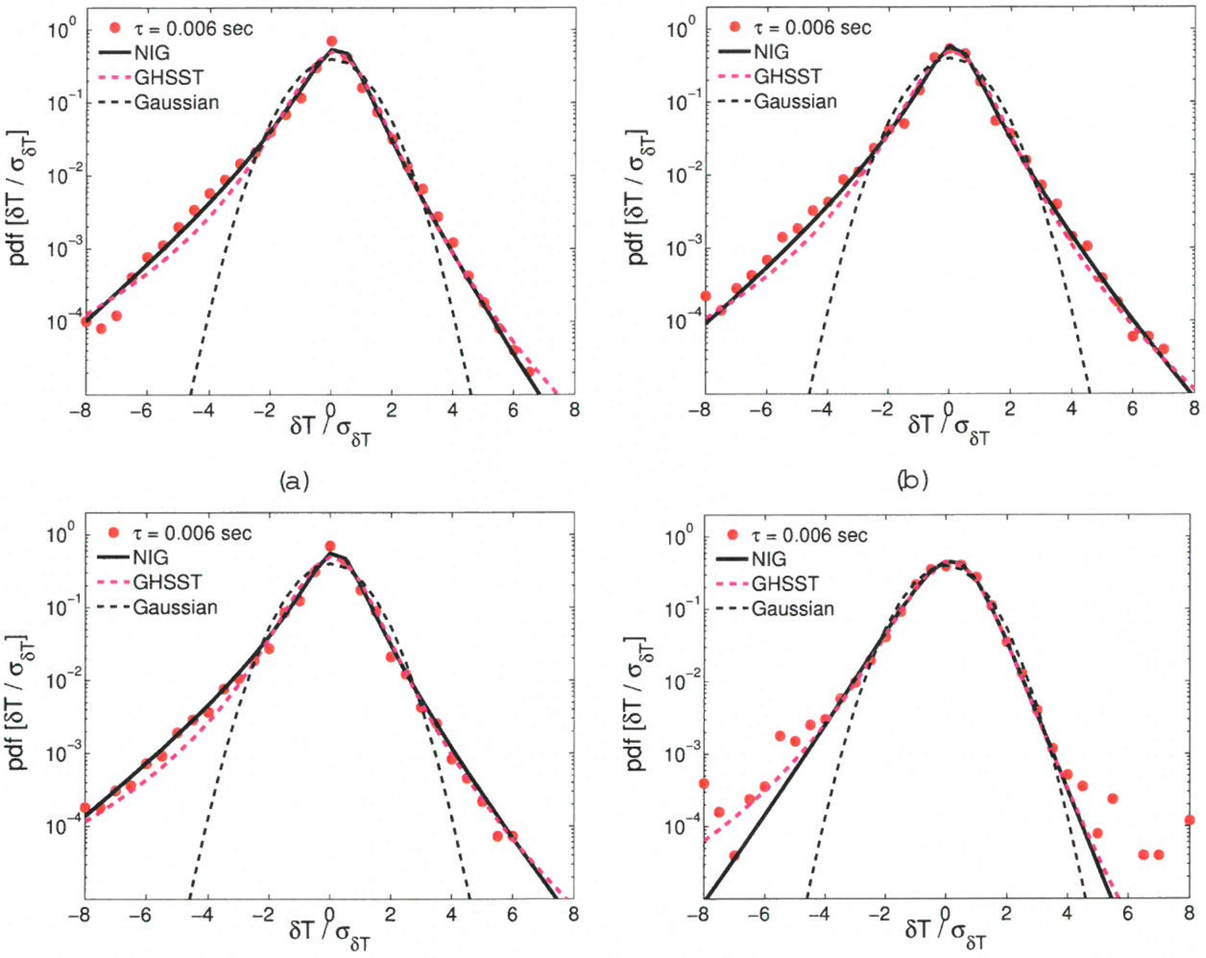


(a) (b) (c) (d)

Figure 6.2: Median velocity increment pdfs for four different stability classes using Ohya wind-tunnel data: a) S1, $0 < \zeta \leq 0.25$; b) S2, $0.25 < \zeta \leq 0.50$; c) S3, $0.50 < \zeta \leq 1.0$; and d) S4, $\zeta > 1.0$.

scales. Therefore, with increased skewness (in the negative sense) the 3-D turbulence features are thought to exhibit increased stretching of the vortex elements, whereas when skewness decreases (when stability increases) the opposite effect occurs. Since, in this experiment the RE decreases with increased stability our results are consistent with previous findings and further suggest a dependence on the velocity increments relative to the stability conditions.

In regards to the temperature increments under varying stable stratification, in Figure 6.3



(a)

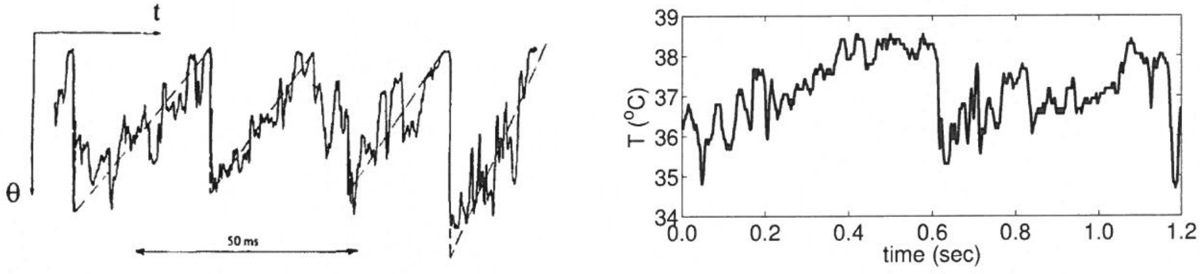
(b)

(c)

(d)

Figure 6.3: Same as Figure 6.2 except the median temperature increment pdf.

we see the pdf models (NIG and GHSST) again are capable of estimating the various features of the temperature increment pdfs even though heavier tailed and skewed features of the distributions are present. This skewed behavior also continues to be present as stability increases with only a minor decrease as strong stability conditions are imposed. It was observed by Stewart (1969, [243]) under a stratified atmosphere that the skewness factor for scalar derivatives ($\partial\theta/\partial t$) was on the order of unity (i.e., -1), which is far from a value of zero required for local isotropy. These asymmetries in the pdfs are thought to be generated by the mean temperature gradients associated with the flow boundary conditions. [17, 112, 176, 235, 238]. Moreover, when examining the scalar series a consistent ramp-cliff feature is shown, see Figure 6.4 for a comparison between a past experiment and our work. These sudden negative jumps in the temperature field are believed to contribute to the skewness we observe here. Clearly, there are similar temperature qualities that are present in boundary layer flows that contribute to these skewed pdfs. Furthermore, from a physical perspective, these features are attributed to a temperature front indicated by cool and warm fluids entrained by the two counterflowing structures [13, 269]. The ramp-cliff behavior is a large-scale feature, which are influenced by boundary conditions [132], on the order of the integral scale, while the sharp leading edge is evident in the small-scales. This omnipresent characteristic of scalar fields is the cause for the anisotropic nature and the heavy tailed features in the increment pdfs.



(a) (b)

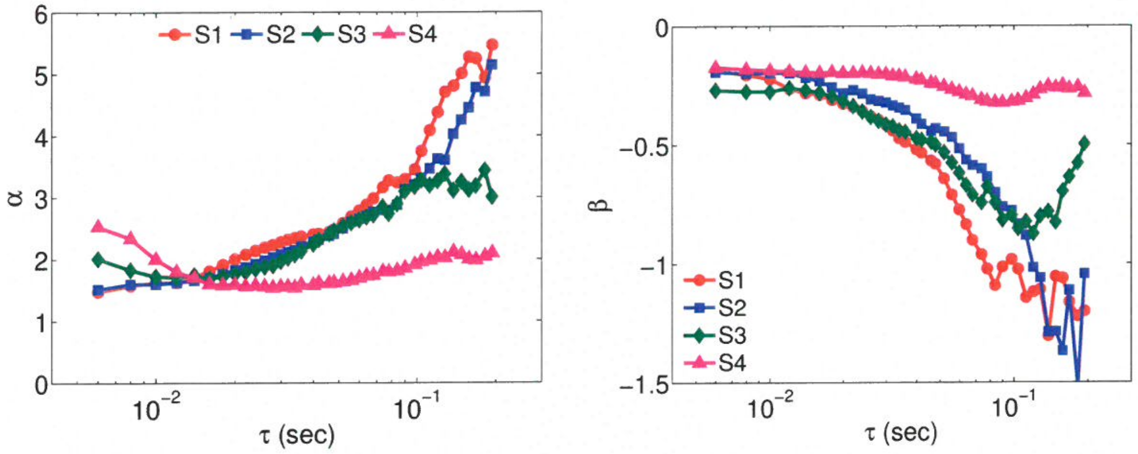
Figure 6.4: Temperature ramp-cliff features from a) Sreenivasan (1979, [240]) b) Series segment from our analysis, Ohya 2001.

6.4 Stability Impacts on the Probability Density Functions

Now that we have shown the models are able to fit the distributions well, we can then evaluate the velocity and temperature increments over a range of separation and stability conditions using the various parameters of the models. For brevity, the NIG model will be used since visually this distribution reveals the most accurate fit to the data.

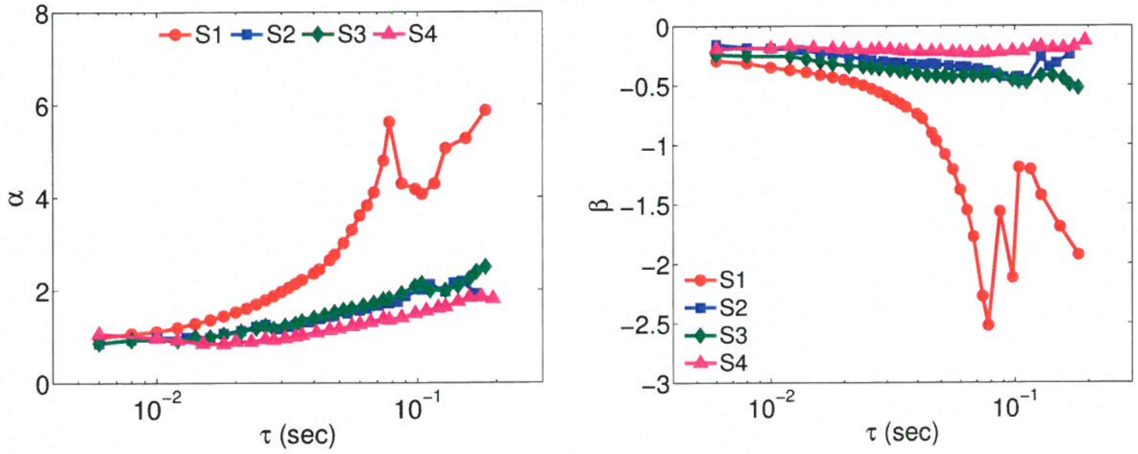
For each stability class, the estimated NIG model parameters were isolated and binned for the range of separations (τ) from 0.006 sec to 0.2 sec. From the binned data, we then took the median value (50th percentile) of the parameters for each separation and presented the results in Figures 6.5 and 6.6 for the velocity and temperature increments, respectively. For these computations, the increments are normalized by their respective standard deviations (i.e., σ_{δ_u}). Also, only the α and β parameters are shown here since they describe the tail and skewed features of the distribution, as discussed in Section 3.1. Figure 6.5a shows α parameter for the four stability classes. At small scales, the tails become lighter with increased stability since α is inversely proportional to the heaviness of the tails (i.e., small α signifies a heavy tail). However, as separation increases the reverse occurs and the increased stability contributes to a heavier tail. In other words, the intermittency of the flow is stronger (resulting in heavier tails) at larger scales when stability conditions are enhance as compared to less stable environments. It is known that under high `RE` and near-neutral flows that distributions become more Gaussian (i.e. lighter tails) with increased separation, but the stability impacts for larger scales is not as well-known. Furthermore, the β parameter remains nearly constant for stronger stability, but changes for the other conditions with increased separation. This indicates that the tails and skewness of the strongly stable regime remain nearly consistent throughout the range of separation.

The variations in α and β for the temperature increments are depicted in Figure 6.6. Unlike in the velocity increment, α at small separations are nearly identical for all stability classes, suggesting that tails of the distribution are equally heavy. Also, the magnitude of α is less suggesting even heavier tails for temperature as opposed to velocity increments, which can be visually seen in two pdf figures. However, as separation increases, the three higher stability regimes remain heavier tailed while the near-neutral conditions behave smiliar to their velocity counterpart and become lighter tailed while approaching Gaussian. This result indicates that the temperature (scalar) field is strongly influenced by the stratification and leads to heavier tailed distributions with increased separations. Furthermore, as seen with the velocity increments, β increases for S1 as separation increases, but the other three regimes remain constant.



(a) (b)

Figure 6.5: Median values of NIG parameters a) α and b) β as a function of τ from the velocity datasets from Ohya (2001, [201]) based on the four different stability classes: S1, $(0 < \zeta \leq 0.25)$; S2, $(0.25 < \zeta \leq 0.50)$; S3, $(0.50 < \zeta \leq 1.0)$; and S4, $(\zeta > 1.0)$.



(a) (b)

Figure 6.6: Same as Figure 6.5, except for the temperature dataset.

Figures 6.5 and 6.6 showcase the feature of the velocity and temperature increment fields in a parametric manner. Utilizing this method to evaluate the evolution of the statistical characteristics of the field reveals the variability of the parameters as a function of stability and separation.

6.5 Conclusion

Using a controlled wind tunnel experiment with different stability conditions, in this work we showed that the velocity and temperature increment pdfs display some uniquely different features. With the use of the NIG pdf model, we highlighted the specific changes in the parameters of the pdf and showed the variation as stability and separation increased. From these results, a few important conclusions can be made: (i) The velocity and temperature increment pdfs are heavy tailed at small separation, however, the temperature distribution is slightly heavier indicating larger intermittency with in the field. (ii) Under strongly stable conditions the α and β parameter for velocity and temperature vary less compared to the other stability regimes, as the scales of the flow and stability increases. (iii) The temperature increment field behaves similarly for moderately to strongly stable conditions, whereas in the velocity increment field, the strongly stable regime is separated by the other three classes. This indicates that the temperature structure is strongly influenced with moderately less stability as opposed to its velocity counterpart. However, to gain further insight into the influences stability has on small-scale turbulence within the atmospheric boundary layer, field measurements are desired to detect if similar trends exist. In the following chapters, an examination into the behavior of larger scale (i.e., mesoscale) pdfs of velocity and temperature increments will be conducted.

Chapter 7

Quasi-Universality of Mesoscale Wind Ramp Distributions

7.1 Introduction

[1] In this study, we examine the probability density functions (pdfs) of long-term observed wind speed increments within the mesoscale range (from 10 minutes to 6 hours) of the atmospheric boundary layer from diverse geographic locations. Throughout the majority of the work thus far, the discussion has been primarily focused on the small-scale turbulent fields. However, in the atmosphere a number of large scale motions are present, such as frontal boundaries, low pressure systems, buoyancy driven flows, meandering motions, and other mesoscale disturbances. The variability of mesoscale phenomena is not only present throughout the year, but also across various locations throughout the world creating further reasons to believe differences exist, such as the tails of the distribution. The knowledge of the tails of the wind speed distribution (i.e. wind ramps) can be advantageous for predicting extreme wind events, which impacts a range of industries (e.g., wind energy) which will be the focus for this study.

Wind energy is the fastest growing renewable energy source in the world. The total installed

---

[1] The material presented in this chapter is in peer-review as the following publication: DeMarco, A.W. and Basu, S. (2017) On the Tails of the Wind Ramp Distributions, Wind Energy

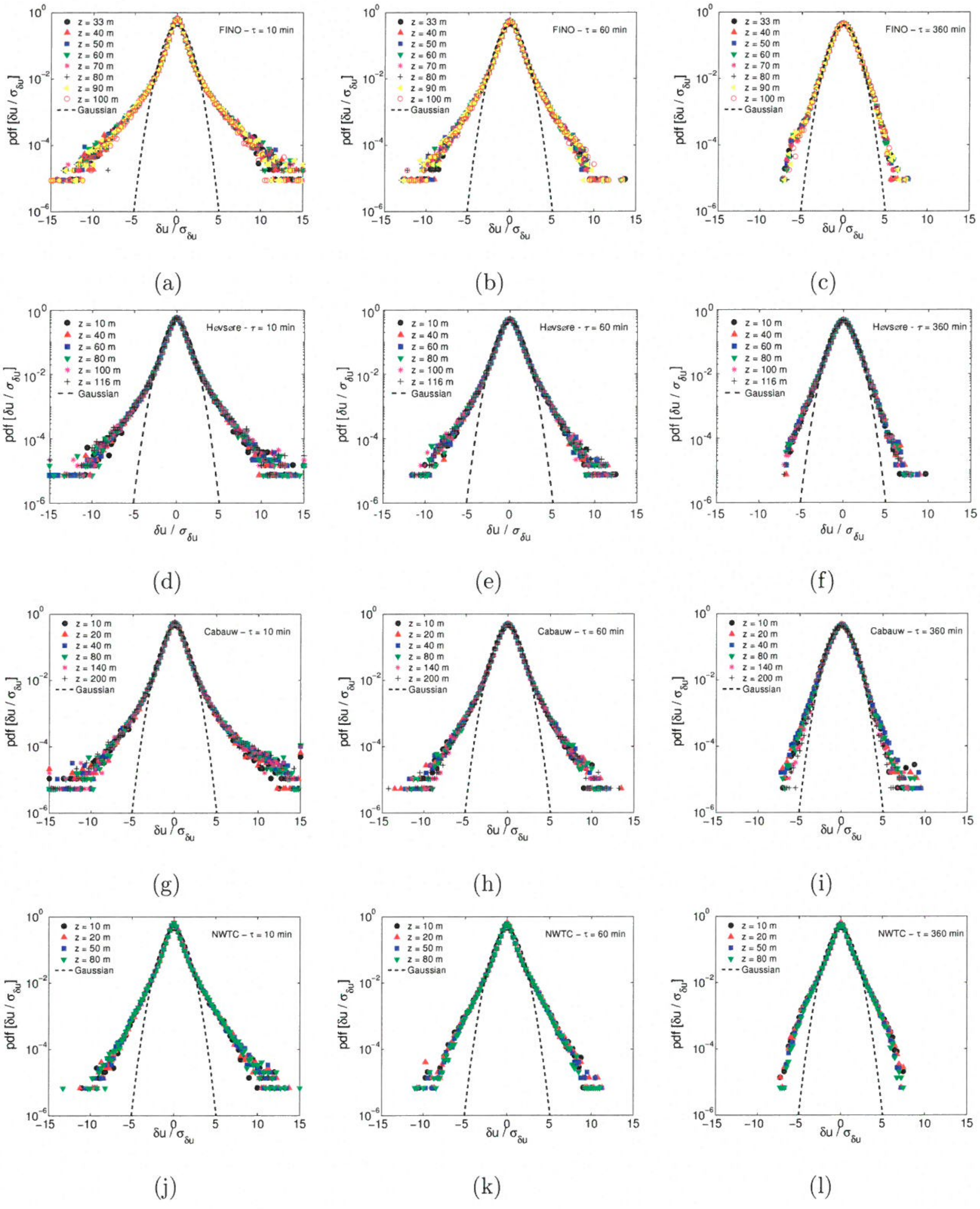
capacity has increased impressively during the last decade, passing from 10 GW in 1998 to 158 GW in 2009.* This is an encouraging development if we consider the issue of managing power systems with high penetration of wind power. Indeed, the variable nature of wind generation makes it difficult to reach the basic requirement of balancing the electricity demand by an equal production. Short-term forecasts of wind power production up to 2 or 3 days ahead can facilitate the management of power systems by operators. Wind power forecasts are useful for performing various power system management functions, such as the dynamic quantification of reserves or the optimization of combined wind-hydro power plant scheduling.1,2 They also prove valuable when incorporated into bidding strategies for participating in electricity markets.

One of the major challenges facing the wind energy industry is the accurate prediction of sudden and sharp fluctuations in the wind field (a.k.a. wind ramps) near the lower part of the atmospheric boundary layer (ABL; [88, 101, 106, 277]). These not-so-rare and inauspicious events can drastically modulate deficiencies (ramp down) and/or surpluses (ramp-up) in wind power production causing disruptions in operations and energy supply balance. More specifically, the sporadic behavior of wind power generation, due to the variable nature of the wind, makes it challenging to manage the balance of energy demand. Thus, having the ability to provide a reliable short-term forecast for these events can greatly facilitate the control and management of this high-demand power generating systems. Therefore, as the demand for more reliable wind power increases there is an ever-present need for further advancement in the understanding of how to properly characterize and quantify the ramp events [129]. It is well-known in the literature that various meteorological factors can contribute to ramp events including (but not limited to): thunderstorm outflows, low-level jets, dry lines, cold fronts [91]. Since most of these phenomena are location-specific (for example, dry lines occur predominantly over the southern Great Plains of the US), one would naïvely expect the statistical properties of the ramp events to also be site-dependent. In this paper, we confront this expectation with reality. By making use of long-term observational data from several field sites, we demonstrate that a key trait of wind ramp statistics behaves in a quasi-universal manner.

Before delving into the detailed quantitative aspects, we provide qualitative support for our claim via Figure 7.1. Here, we have plotted the probability density functions (pdfs) of wind speed increments (δu) normalized by standard deviation ($\sigma_{\delta u}$) displayed in a log-linear fashinon, from the four locations for all available tower height at 10 min, 60 min and 360 min temporal separation (τ), situated from left to right, respectively. For a specific time increment (τ), the wind speed increments (or ramps) are defined in Eq. 1.1. From wind energy perspective, τ values on the order of a few minutes to a few hours (the so-called mesoscale regime) are of utmost importance [47, 129, 255, 280]. They portray strong peakedness near the mode of the distribution, and more importantly, they all possess tails which are much heavier than a Gaussian pdf. These tails seem to depend on τ in a subtle yet systematic manner. Also, the pdfs from all locations display remarkable similiarities for a specific τ and the appearance of the distributions change steadily over the range of separation. This finding is quite remarkable given the differences in the locations of the datasets and variations in the sample time-periods (i.e., different years for each site). Furthermore, comparing these pdfs to the small scale data shown in the previous chapters, it is quite remarkable that the distributions are fairly similiar in appearance, such that the non-Gaussian behavior is present for smaller τ value and the distributions become increasingly more Gaussian with increased separation. However, one striking difference is that normalized mesoscale wind speed increments possess heavier tails than turbulent increments, suggesting that time-scales on the order of minutes result in higher probability for rare events.

Additionally, there is also clear similiarities between the various heights overlaid on each plot which reveals that there are no obvious height-dependences present. Though, this is evident across all locations, there are varying degrees of intermittency present between the different locations. For instance, the tails of the distribution for NWTC do not exhibit the same features as the other locations. Figure 7.1j shows NWTC 10 min wind speed increment pdf which does not have heavy tails as significant as the other 10 min pdfs. In fact, the data at NWTC suggests that probability of a extreme event with a normalized wind speed increment $\delta u/\sigma_{\delta u}$ value of -15 has virtualy a zero percent possibility of occurrence as indicated over the sample 10 years data, while at the other locations this extreme event is much more likely to occur. These unexpected findings inspired us to probe further into this problem by addressing a suite of science questions in this paper:

- Do the pdf tails corresponding to the ramp-up and ramp down events behave differently?
- How do the tails depend on the height (above ground level)?
- What is the impact of aggregation (filtering) on the tails?
- Can the dependence of tail properties on the sample size be quantified?



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure 7.1: Probability density functions of wind speed increments for FINO1 (a-c), Høvsøre (d-f), Cabauw (g-i), and NWTC (j-l) for the three sets of τ values from 10 min, 60 min and 360 mins.

To the best of our knowledge, none of these questions have been answered in the literature in a comprehensive manner. We do point out that a handful of studies [48, 49, 155, 185, 215, 219]. provided important building blocks for our research. Unfortunately, several of these papers focused on wind gusts (τ being on the order of seconds), and thus, their findings cannot be very relevant for mesoscale wind ramps. More critically, most of these studies utilized very limited amount of observational data (often with durations of a few hours to merely a few days). Consequently, in lieu of converged statistics, the results from these works cannot be faithfully generalized. In this respect, the present study is unique. It utilizes multi-year wind datasets from four tall-tower sites: FINO 1 (North Sea), Høvsøre (Denmark), Cabauw (the Netherlands), and NWTC (USA). Since these sites are quite diverse in nature, we have more confidence in generalizing the outcomes.

7.2 Description of Observational Data

In order to capture a sufficient amount of mesoscale phenomena and have a data sample size that provides robust statistics, it is necessary to obtain datasets that have a large number of samples and cover a multi-year time period. Furthermore, datasets which comprise of various geographical and climatological conditions are also ideal to test variability in the data. Therefore, we examined four different datasets for our analysis made up of homogeneous and complex terrain. For each location different types of research-grade instruments are utilized. Naturally, with such long-term data, all of the data sets contain variable amount of missing data. However, the average amount of missing data for the entire collection is less than $\sim 4\%$ and in some cases, well below 1%. Thus, we believe the available data allows a robust statistical analysis. The descriptions and results are ordered by the complexity of the terrain, from uniform, ocean surface to mountainous topography. Table 7.1 and Figure 7.2, along with the following subsections, provide more details into these locations. This data will be utilized throughout the remaining chapters.

Table 7.1: Description of measurement sites

Site	Elevation (m; mean sea level)	Location	No. of Months
FINO 1	0	54.01° N, 6.59° E	91
Høvsøre	0	58.44° N, 8.15° E	132
Cabauw	- 0.7	51.97° N, 4.93° E	170
NWTC	1855	39.91° N, 105.23° W	132

7.2.1 FINO 1

It is an offshore platform in the North Sea [87, 191, 254]. It consists of a 100 m tall meteorological tower equipped with wind speed measurement sensors (cup anemometers) at heights of 33 m, 40 m, 50 m, 60 m, 70 m, 80 m, 90 m, and 100 m. A total of 91 months of wind speed data collected over a period of nine years (2004–2012) are utilized in the present study. Each time-series (output rate: 10 min) contains ~478 thousand samples.

7.2.2 Høvsøre

This meteorological tower is situated in a rural area close to the west coast of Jutland, Denmark and played pivotal role in numerous wind energy studies [125, 206]. We analyze 10 min-average wind data from six levels: 10 m, 40 m, 60 m, 80 m, 100 m, and 116 m collected during the years 2005–2015. In this case, each time-series consists of ~567 thousand samples.

7.2.3 Cabauw

The Cabauw Experimental Site for Atmospheric Research (CESAR) tower is located in the western part of the Netherlands [182, 256, 263]. We use 170 months of 10 min-average wind speed data from the years 2001-2015 (~736 thousand samples) measured by propeller wind vanes at heights of 10 m, 20 m, 40 m, 80 m, 140 m, and 200 m. We would like to point out that even though the landscape at Cabauw is quite flat and open (grassland), the existence of wind breaks and scattered villages cause significant disturbances in the near-surface region [263].

7.2.4 NWTC

We analyze multi-year (2004–2014) wind data from a 80 m tall tower (called M2) located at the foothills of the Colorado Rocky near Boulder, Colorado and maintained by the National Renewable Energy Laboratory's (NREL) National Wind Technology Center (NWTC). This location represents complex terrain and is prone to various wind flows and disturbances [134]. The NWTC dataset includes 1-min averaged, cup anemometer-based, wind speed time series from four heights: 10 m, 20 m, 50 m, and 80 m. Each time-series is made up of ~5.78 million points with virtually no data gaps. In order to compare similar sampling rates, we created a 10 min-average series (sample size: ~578 thousand) by simple moving averaging (followed by downsampling). This adjusted dataset will be utilized throughout this work.

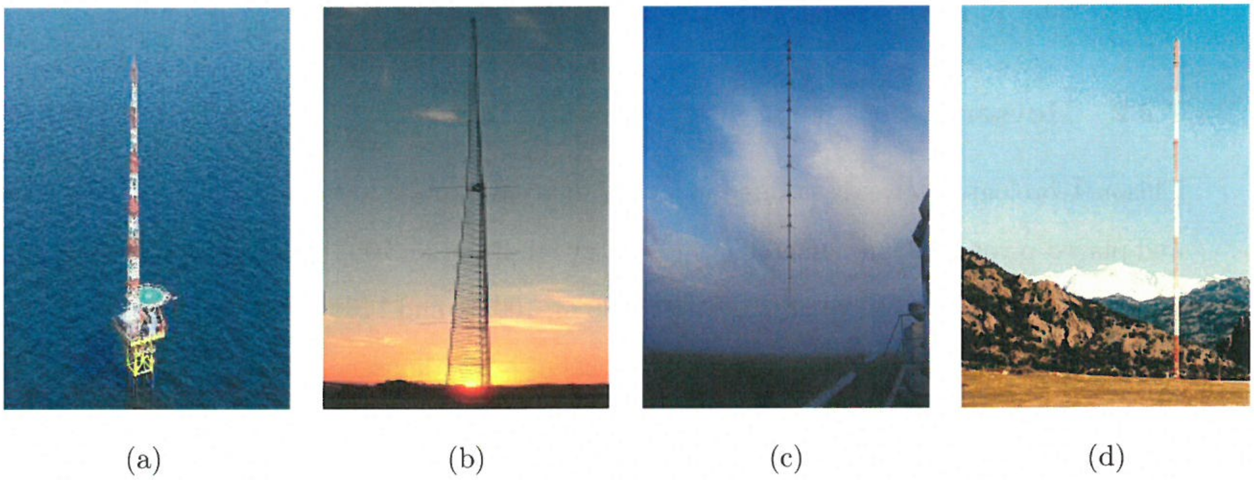

(a) (b) (c) (d)

Figure 7.2: a) FINO-1 (100 m) 2004–2012, b) Høvsøre (116.5 m) 2005–2015, c) Cabauw (200 m) 2001–2015, d) NWTC (M2; 80 m) 2004–2014

7.3 Methodology

In order to investigate the tail features of the wind ramp events, we have borrowed a well-established methodology, called the Hill plot [82], from the extreme value (EV) theory [75, 85, 107], also see Section 3.6 and Chapter 4. In this section, we explain this approach in detail by using synthetically generated random variates from two heavy-tailed distributions.

By definition, a heavy-tailed distribution (F) satisfies [72]:

$$\overline{F}(x) = 1 - F(x) \sim \frac{\lambda}{x^{\gamma}}; \qquad x \to \infty, \gamma > 0, \tag{7.1}$$

where $\overline{F}$ is the so-called complementary cumulative distribution function (ccdf). λ is a positive constant and γ is known as the tail-index (or shape parameter). In principle, γ can be estimated from the slope, $-\frac{d \log \overline{F}(x)}{d \log x}$ [72]. However, in practice, the most common approach is to invoke the concept of order statistics [232].

The rank-ordered values (in decreasing order) of x can be written as: $\Phi_k = x_k$, where $k = 1, ..., N$. If the variates x follow Eq. 7.1, it is expected to exhibit the following power-law behavior (a.k.a. Zipf law; [282]) as shown in Eq. 4.2. Over the years, several estimators for γ have been proposed in the literature, including (but not limited to) Pickand's estimator [209], Hill estimator [111], and the Dekkers-Einmahl-de Haan estimator [77]. In this work, we use the popular Hill estimator (γ_H), as described in Section 3.6. For EV distributions (e.g., Pareto), estimated γ_H is supposed to stabilize with increasing values of k. In Figure 7.3, we show an illustrative example utilizing generalized Pareto (GP) distributed variates.

The pdf of the GP distribution can be written as:

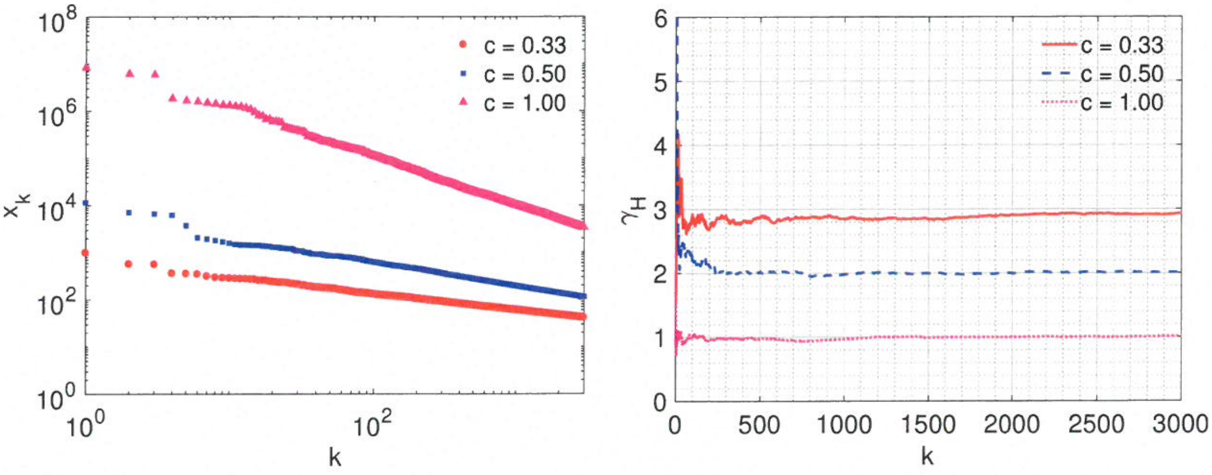
$$f(x) = \left(\frac{1}{a}\right)\left(1 + \frac{c(x-b)}{a}\right)^{-1-1/c}; \qquad x > b, c > 0, \tag{7.2}$$

where a, b, c are the parameters of GP. By integrating this equation, one can derive the ccdf of GP as:

$$\overline{F}(x) = \left(1 + \frac{c(x-b)}{a}\right)^{-1/c}. \tag{7.3}$$

Thus, the ccdf of GP is expected to decline as a power-law with tail index $\gamma = 1/c$.

In Figure 7.3 (left panel), the rank-order plots for the GP distribution are shown for three



(a) (b)

Figure 7.3: a) Rank-order (a.k.a. Zipf) plots for generalized Pareto distributed variates with three different c values. The parameters a and b are assumed to be equal to 1 and 0, respectively. b) The estimated γ_H values for these cases. It is clear that $\gamma_H \approx 1/c$ for $k > 1000$.

values of c. For each case, the sample size is 10^7 and the parameters a and b are assumed to be equal to 1 and 0, respectively. By construction, only positive random variates are generated in this case. The tail indices are determined via the Hill plot in the right panel of Figure 7.3. Clearly, the γ_H values rapidly stabilize towards $\frac{1}{c}$ for all the three cases, as would be desired. This example attests to the prowess of the Hill plot in estimating the tail indices from a rather simple EV distribution. Next, we investigate the usefulness of the Hill plot using a far more complicated distribution with two distinct tail behaviors.

The generalized hyperbolic skew student's t (GHSST) distribution is often used in financial modeling and risk management [2, 42]. It has the innate ability to fit pdfs with heavy tails and significant asymmetry. A brief overview of this distribution is provided in Section 3.2. A realization of the GHSST variates is shown in the top-panel of Figure 7.4. Large positive values, signifying a heavy right tail, can be readily observed in this plot.

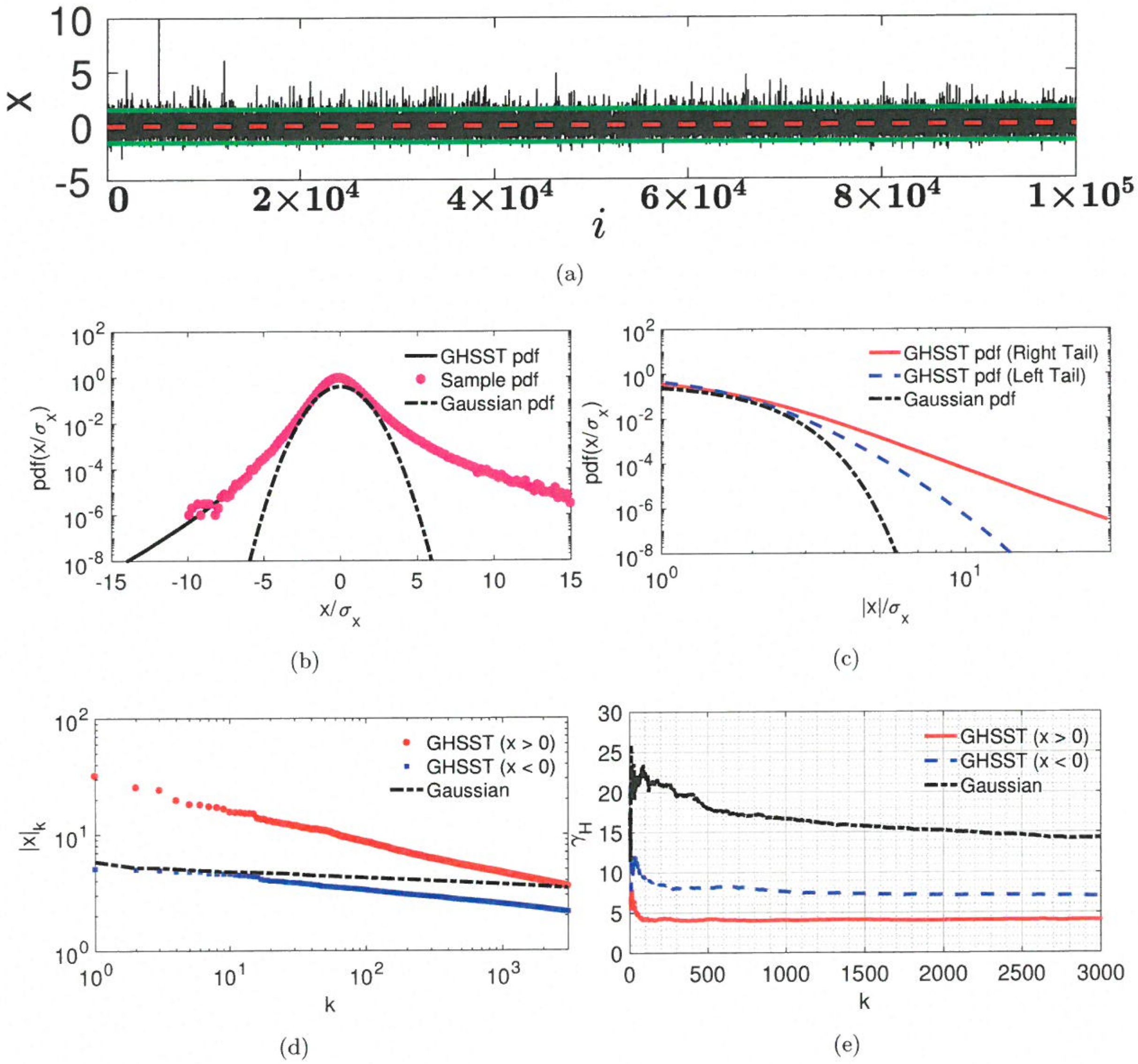


Figure 7.4: A realization of the GHSST variates (sample size = 10^7) is generated using the following parameters: $\nu = 6$, $\beta = 0.5$, $\mu = -0.125$, and $\delta = 1$. A subset of these variates is shown in (a) as an illustrative example. The mean of the variates is depicted by the dashed red line. The green lines denote three times the standard deviation around the mean. The analytical (black line; [2]) and sample (magenta circles) pdfs are shown in (b). For comparison, a Gaussian pdf (dot-dashed line) with zero mean and unit variance is overlaid on this panel. The tails of the GHSST and Gaussian pdfs are shown (c). Since the right and the left tails of the GHSST pdf behave differently, they are shown separately. Clearly, the right tail exhibits a linear behavior in this log-log representation. Rank-order (a.k.a. Zipf) plots for the GHSST and Gaussian distributed variates are shown in (d). Estimated tail indices (γ_H) utilizing the Hill plot are documented in (e).

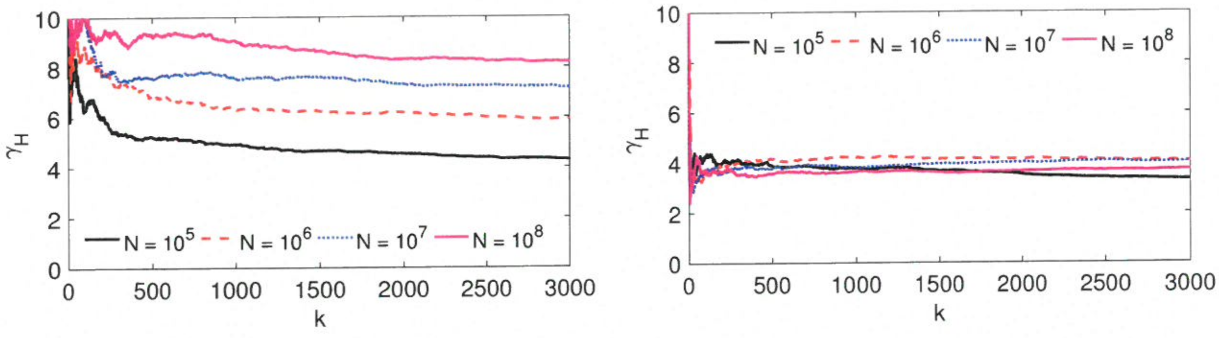
The pdf of the generated GHSST variates is shown in the middle-left panel of Figure 7.4. For comparison, a Gaussian pdf is overlaid on this plot. Clearly, both the left and right tails of the GHSST variates are much heavier than the Gaussian pdf. They also show different decaying behaviors as evident in the middle-right panel. For large values of x, the right tail portrays a quasi-linear appearance in this log-log representation. In other words, the right tail is characterized by a power-law distribution which is in-line with the asymptotic limit discussed in Section 3.2. In contrast, the left tail strongly departs from linearity highlighting its mixed-exponential-power-law behavior. The rank-order plot, shown in the bottom-left panel of Figure 7.4, provides further supporting evidence.

At this point, we would like to point out that Eq. 7.1 has limited applicability in real-world scenarios. For such cases, this equation should be generalized as:

$$\overline{F}(x) = 1 - F(x) \sim \frac{L(x)}{x}; \qquad x \to \infty, \gamma > 0, \tag{7.4}$$

where, L is a slowly varying function (e.g., exponential). For large x, $L(x)$ may be approximated as a constant λ.

The tail indices from the GHSST variates are estimated via the Hill plot and shown in the bottom-right panel of Figure 7.4. These results should be interpreted carefully by taking into consideration Eq. 7.4. In the case of right tail, the γ_{H} values stabilize rapidly as in the case of GP distributed variates. For the left tail, the γ_{H} values are significantly higher; also, the stabilization is slightly slower (difficult to detect in this figure without zooming in). In this case, the exponential term modulates the power-law tail. Please note that the estimated γ_{H} values are very high for the random Gaussian variates and they never stabilize as the tails simply follow exponential behavior. In Figure 7.5, we document the influence of sample size on γ_{H} estimation using the GHSST variates. In the case of the right tail (exhibiting power-law behavior), γ_{H} values are more-or-less insensitive to sample size (N). In contrast, for the left tail, γ_{H} keeps increasing as N increases. This non-convergence essentially corroborates the fact that the left tail of the GHSST pdf does not exhibit a purely power-law behavior; rather, it follows a mixed-exponential-power-law.[2]



(a)

(b)

Figure 7.5: Sensitivity of estimated γ_H values with respect to sample size. All the GHSST variates are generated using the same parameters as in Figure 7.5. (a) and (b) correspond to the left and right tails, respectively.

In summary, with the aid of randomly generated variates, we have demonstrated that the Hill plots can be very effective in characterizing different types of tail behaviors. It is also computationally very efficient. For these reasons, in the following section, we will invoke this methodology to address the science questions posed in the Introduction. The idealized examples shown in Figures 7.3, 7.4, and 7.5 will provide guidance for interpreting the wind ramp characteristics observed within various observational datasets.

7.4 Evaluation of the Tail-Index

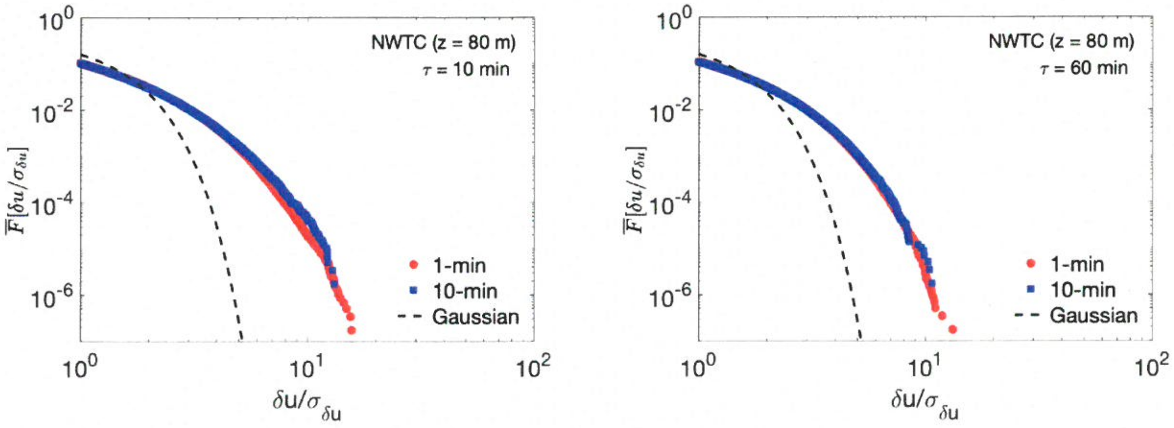
Given that the NWTC dataset offers the largest sample size, we select it first for comprehensive analysis. The original granularity of the wind time-series is 1-min. Henceforth, we refer to this series as $\mathrm{NWTC}_{1\mathrm{min}}$. In order to study the effects of aggregation on the tail characteristics,

---

[2] In a recent paper [79], we reported a similar trend for the normal inverse Gaussian (NIG) distribution, which also possesses mixed-exponential-power-law tails.

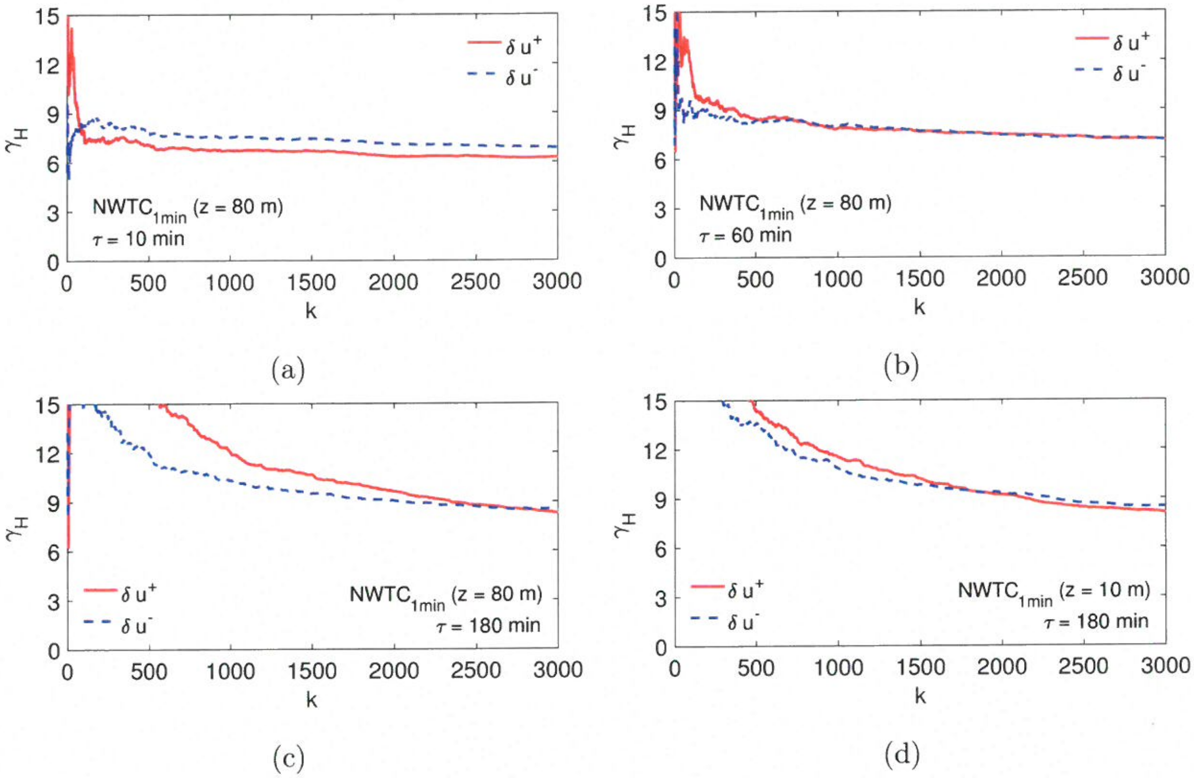
we created a 10-min-average series (sample size: ~578 thousand) by simple moving averaging (followed by downsampling) of the NWTC$_{1\text{min}}$ series. This new series will be identified as NWTC$_{10\text{min}}$.

The ccdf ($\overline{F}$) for both the NWTC$_{1\text{min}}$ and NWTC$_{10\text{min}}$ time-series are shown in Figure 7.6. The left and right panels represent $\tau = 10$ min, and $\tau = 60$ min, respectively. On these plots, $\overline{F}$ for a Gaussian distribution is also shown for comparison. In this log-log representation, we only focus on the right tail (ramp-up) of the pdf. Several remarks can be made from this figure. First of all, both the NWTC$_{1\text{min}}$ and NWTC$_{10\text{min}}$ cases clearly portray non-Gaussian tails. The implication of this heavy-tail behavior is rather crucial for the wind energy community. For example, the exceedance probability of a strong ramp-up event of magnitude $5\sigma_{\delta u}$ is very small (much less than 10^{-6}) if one assumes Gaussianity. According to observations, however, the exceedance probability is almost 10^{-2}. In other words, the assumption of Gaussianity leads to severe underestimation of extreme wind ramp events.



(a) (b)

Figure 7.6: Complementary cumulative distribution function ($\overline{F}$) from the NWTC$_{1\text{min}}$ and NWTC$_{10\text{min}}$ datasets at $z = 80$ m are shown. (a) and (b) represent time increments (τ) of 10 min and 60 min, respectively. The wind increment values are normalized by the corresponding standard deviations ($\sigma_{\delta u}$). A Gaussian pdf is overlaid (dashed line) as a reference.



(a) (b) (c) (d)

Figure 7.7: The Hill plots for wind ramp distributions based on the NWTC$_{1\text{min}}$ time-series. The red (solid) and blue (dashed) lines represent ramp-up (δu^+) and ramp-down (δu^-) cases, respectively. (a) - (d) correspond to the following scenarios, respectively: (i) τ= 10 min, z = 80 m; (ii) τ= 60 min, z = 80 m; (iii) τ= 180 min, z = 80 m; and (iv) τ= 180 min, z = 10 m.

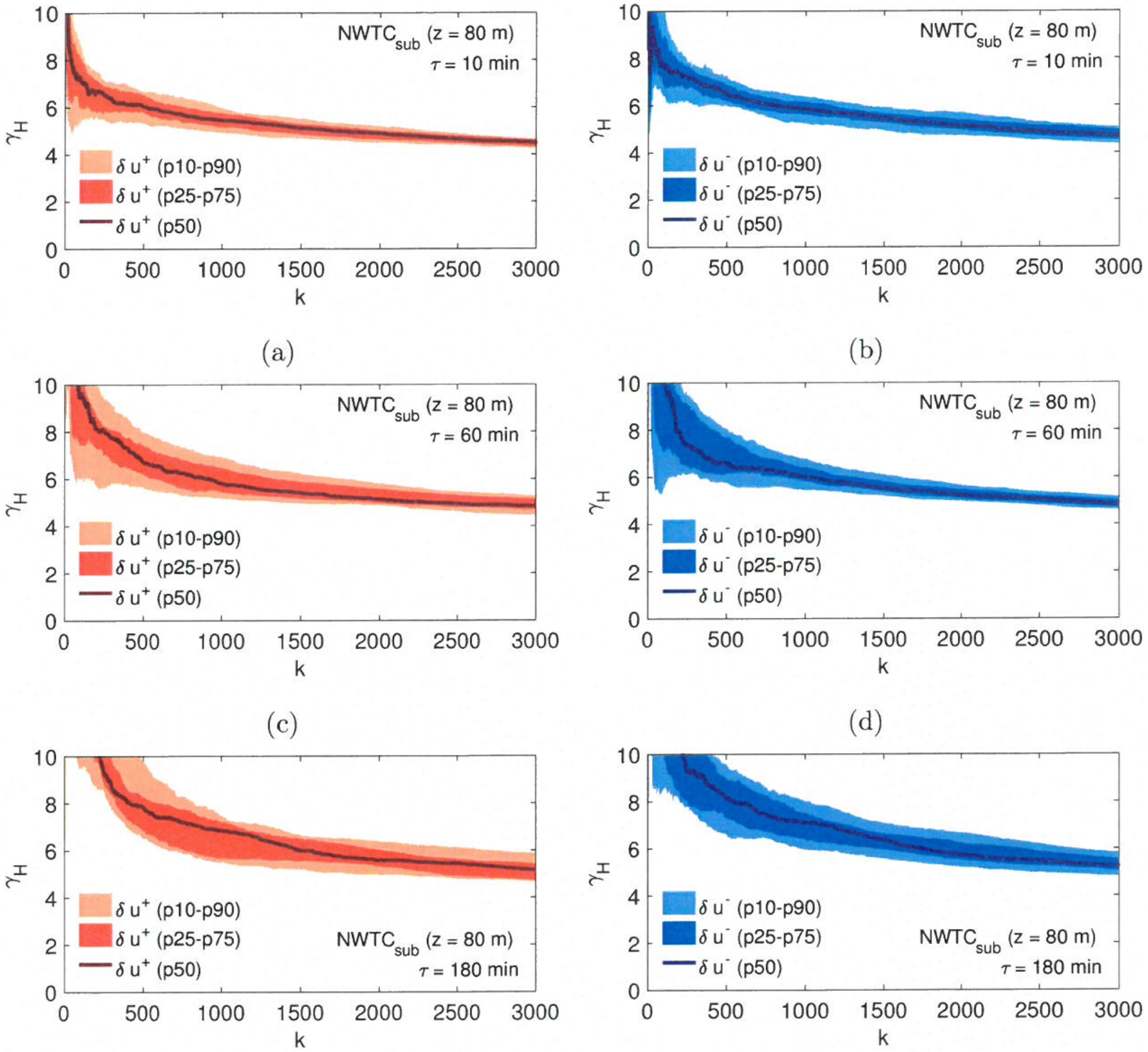
In Figure 7.1, we reported that the tails of the wind ramp pdfs systematically depend on τ. Thus, it is not surprising that the same dependence is also evident from the ccdfs. From Figure 7.6, one can discern that, in comparison with $\tau = 10$ min, the right tail decays faster in the case of $\tau = 60$ min. Later on, we will establish that this trend is actually monotonic in the range of $\tau = 10 - 360$ min.

According to Figure 7.6, the agreement between NWTC$_{1\text{min}}$- and NWTC$_{10\text{min}}$-based $\overline{F}$ curves are excellent up to $\delta u \approx 5\sigma_{\delta_u}$. Beyond that point, the NWTC$_{1\text{min}}$-based $\overline{F}$ curve starts to decline faster. With the aid of the Hill plots, we will further investigate if this discrepancy is due to the disparity in sample sizes or it is an artifact of aggregation.

The Hill plots for NWTC$_{1\text{min}}$ are shown in Figure 7.7. The values of γ_H are found to be noticeably higher for larger τ values. In other words, the wind ramp pdf tails decay faster for larger τ values, which is in-line with our earlier finding. Both the ramp-up (noted as δu^+) and ramp-down (δu^-) cases follow similar trends. However, the γ_H curves never fully stabilize for either case in the considered range ($1 < k \leq 3000$). Thus, we can deduce that the wind ramp pdf tails do not obey power-laws. Later on, we will explore further if these NWTC data-based results also hold for other datasets.

In order to quantify the effect of sample size on the γ_H values, we adopted a Monte-Carlo-type strategy. From the NWTC$_{1\text{min}}$ time-series (sample size 5.78 million), we extract one hundred contiguous subsets from random locations. Each subset is called NWTC$^i_{\text{sub}}$ and contains 578 thousand samples. The index i varies from 1 to 100 to demarcate each subset. We then perform Hill plot analysis on each subset separately and compute ensemble statistics. These ensemble Hill plots are shown in Figure 7.8. The ramp-up and ramp-down cases are shown separately in the left and right panels, respectively.

The overall trends of the γ_H values reported in Figure 7.7 and Figure 7.8 are qualitatively very similar. However, the magnitudes of γ_H for the NWTC$^i_{\text{sub}}$ cases are significantly lower than the NWTC$_{1\text{min}}$ series. We would like to remind the readers that similar sample size dependency were reported earlier in the case of the left tail (depicting a mixed-exponential-power-law behavior) of the GHSST pdf (refer to bottom-left panel of Figure 7.5). Thus, on the basis of the Hill plot analyses and ccdf plots, we can confidently claim that the wind ramp distributions do not exhibit power-law tails.



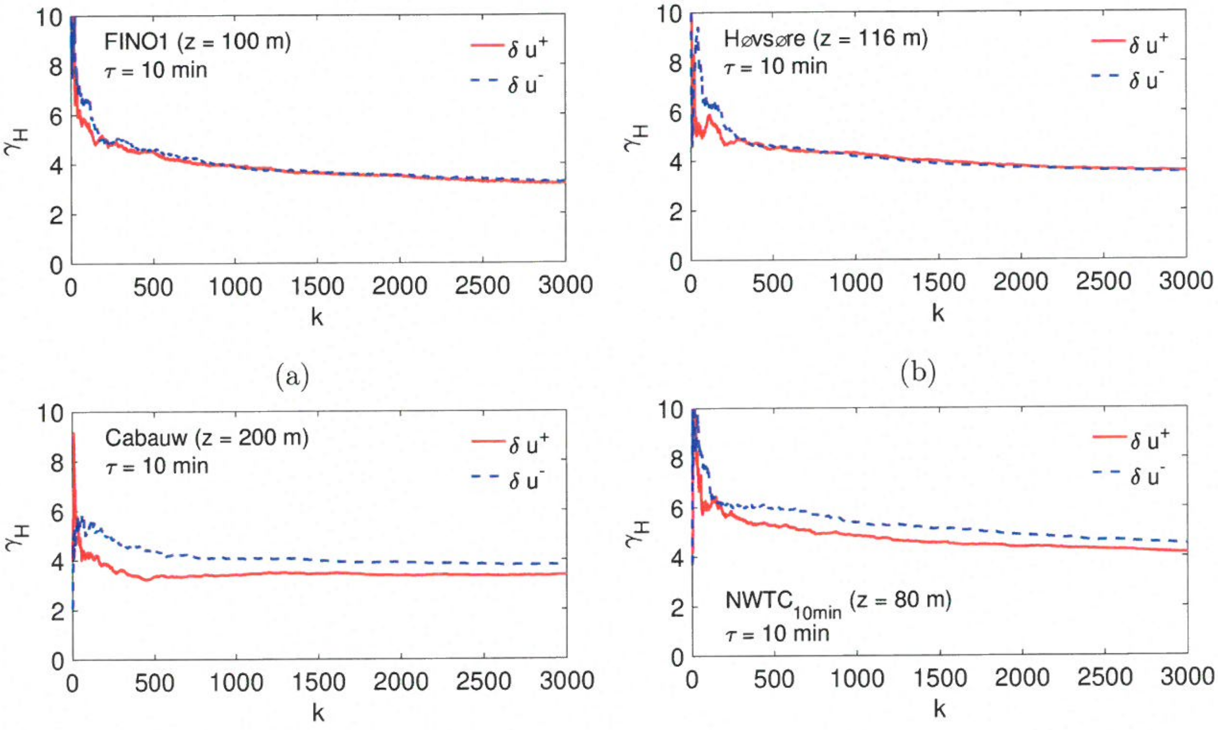
(a) (b) (c) (d) (e) (f)

Figure 7.8: The Hill plots for wind ramp distributions based on the subsets of 1-min-average NWTC wind speed time-series. Each subset contains contiguous 578 thousand samples. A total of one hundred randomly selected subsets are utilized for these plots. The top, middle, and bottom panels represent τ = 10 min, 60 min, and 180 min, respectively. Ramp-up and ramp-down results are shown in left and right panels, respectively. The solid lines, dark shaded areas, and the light shaded areas correspond to the medians, 25th-75th percentile ranges, and 10th-90th percentile ranges, respectively.

In order to further bolster our claim, we have computed γ_{H} for three other locations: FINO 1, Høvsøre, and Cabauw. Wind data from the topmost sensor levels are utilized. Figures 7.9 – 7.11 show the Hill plots for three different values of τ, 10 min, 60 min and 180 min, respectively. In each of these figures, we also included the Hill plot for the $\mathrm{NWTC}_{10\mathrm{min}}$ series. Based on these figures, several assertions can be made. First of all, the Hill plots from all the locations look remarkably similar. For (almost) all the cases, the differences between the ramp-up and ramp-down events are marginal. At the same time, for all plots, the values of γ_{H} do not stabilize and continue to decrease with increasing k. Thus, we can safely rule out power-law being a viable candidate for wind ramp distributions. In-line with our earlier finding, the γ_{H} values exhibit dependence on τ. Thus, stable distributions [232] should not be used to characterize wind ramp distributions.

Earlier, we have concluded that the γ_{H} values strongly depend on sample size. Now, in order to probe the impact of aggregation, we compare the γ_{H} values for the $\mathrm{NWTC}_{10\mathrm{min}}$ series (bottom right panels of Figures 7.9 – 7.11) against the corresponding values from the $\mathrm{NWTC}^{\mathrm{i}}_{\mathrm{sub}}$ series (reported in Figure 7.8). Please note that these series have identical sample size; albeit, they have different granularity. From the plots, it is quite evident that the γ_{H} values from the $\mathrm{NWTC}_{10\mathrm{min}}$ and $\mathrm{NWTC}^{\mathrm{i}}_{\mathrm{sub}}$ series are very much comparable. Thus, within the limited filtering range of 1-min to 10-min, the aggregation effect is negligible. However, one should not downplay the effects of sample size.

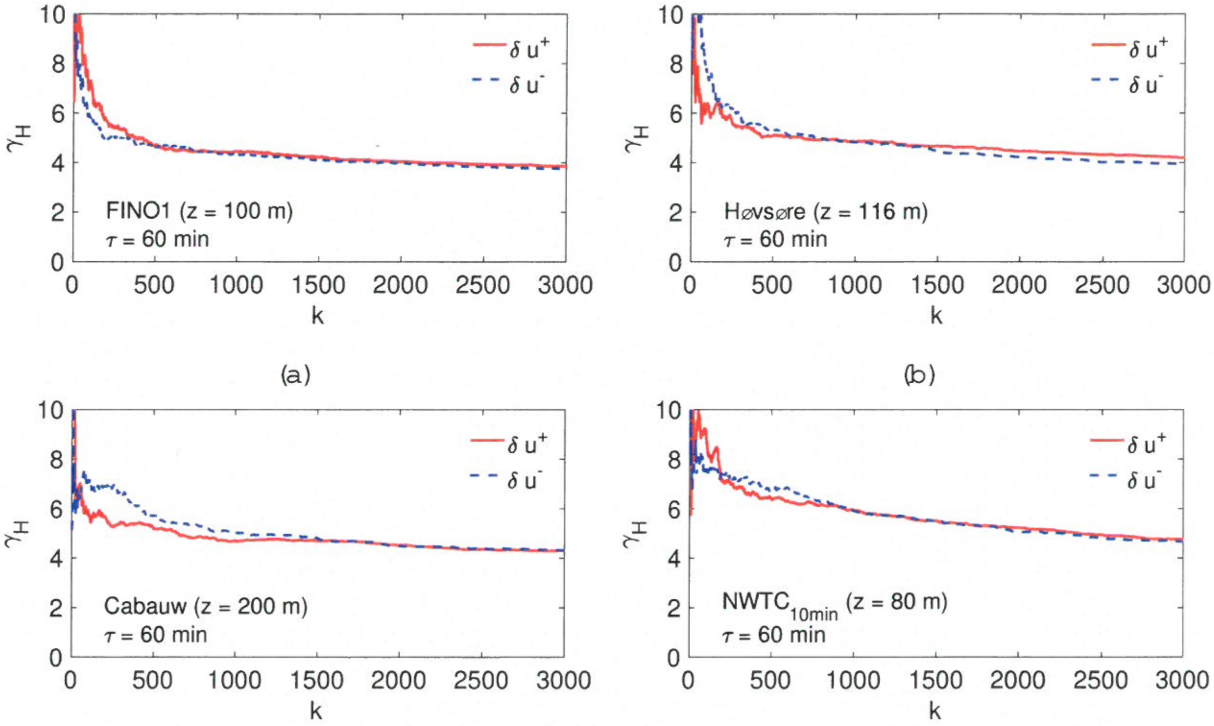
Thus far, we have only focused on wind data from the topmost sensors of all the four meteorological towers. It would be interesting to find out if/how the γ_{H} values depend on sensor height. Instead of plotting several individual Hill plots, we opt for plotting averaged γ_{H} values so as to report all the results succinctly in Figures 7.12 and 7.13. The values of $\langle\gamma_{\mathrm{H}}\rangle$ are computed for $2000 \leq k \leq 3000$. We intentionally (and incorrectly) assume that over this range the values of γ_{H} have fully stabilized. Despite this ad-hoc assumption, the results are quite revealing. For both the ramp-up and ramp-down cases, $\langle\gamma_{\mathrm{H}}\rangle$ increase monotonically from ~ 4 (at $\tau = 10$ min) to $\sim 6-7$ (at $\tau = 360$ min). The diversity in $\langle\gamma_{\mathrm{H}}\rangle$ values across various heights is rather small (especially for $\tau < 180$ min). Among all the locations, the spread of $\langle \gamma_H \rangle$ is the most significant at Cabauw; the $\langle \gamma_H \rangle$ curves from the top two sensors, located at heights of 140 m and 200 m, seem to branch out from others for $\tau > 180$ min. We speculate that certain meteorological processes (e.g., low-level jets) influence the wind ramp statistics at higher altitudes. However, we need more observational datasets from higher altitudes (possibly collected by lidars and/or sodars) to shed further lights on this intriguing finding. To evaluate the quasi-universal behavior between these datasets, we will examine the probability of exceedance over a range of standard deviations and two temporal scales.



(a) (b)

(c) (d)

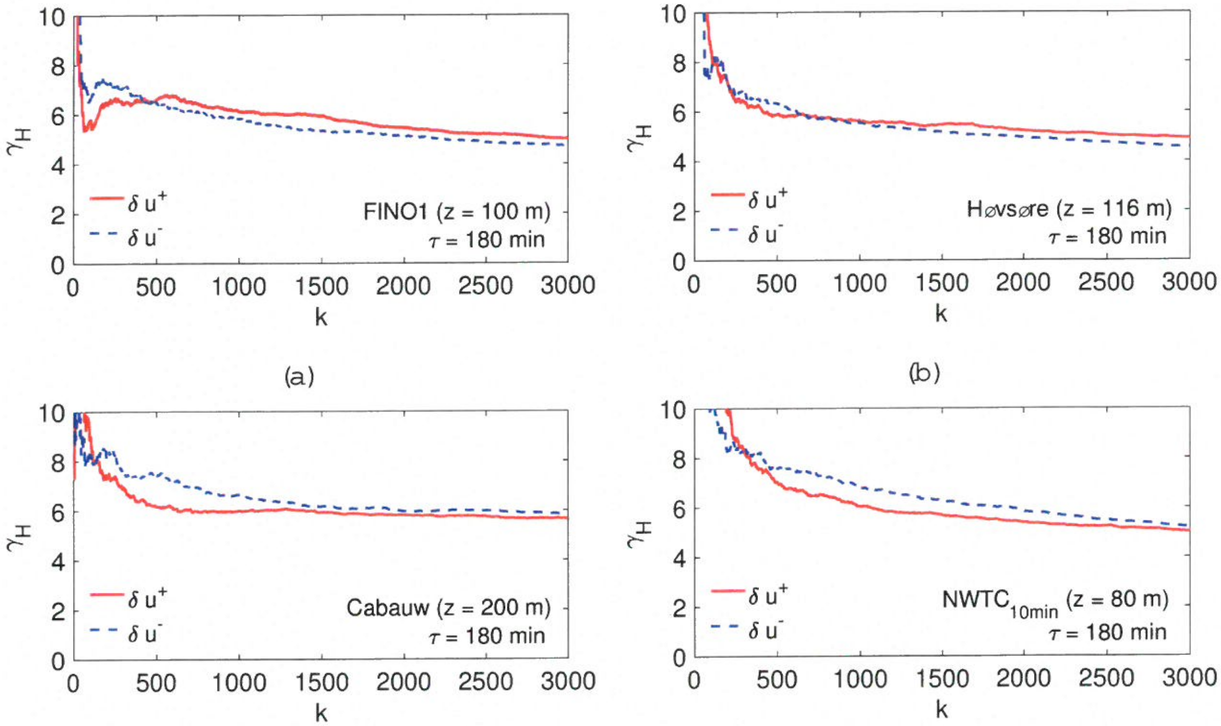
Figure 7.9: The Hill plots of wind ramp distributions from four field sites: FINO1 ((a); sensor height: 100 m), Høvsøre ((b); sensor height: 116 m), Cabauw ((c); sensor height: 200 m), NWTC ((d); sensor height: 80 m). For all the cases, 10-min-averaged wind speed are utilized. The time increment (τ) is 10 min. The ramp-up (δu^+) and ramp-down (δu^-) statistics are denoted by red (solid) and blue (dashed) lines, respectively.



(a) (b)

(c) (d)

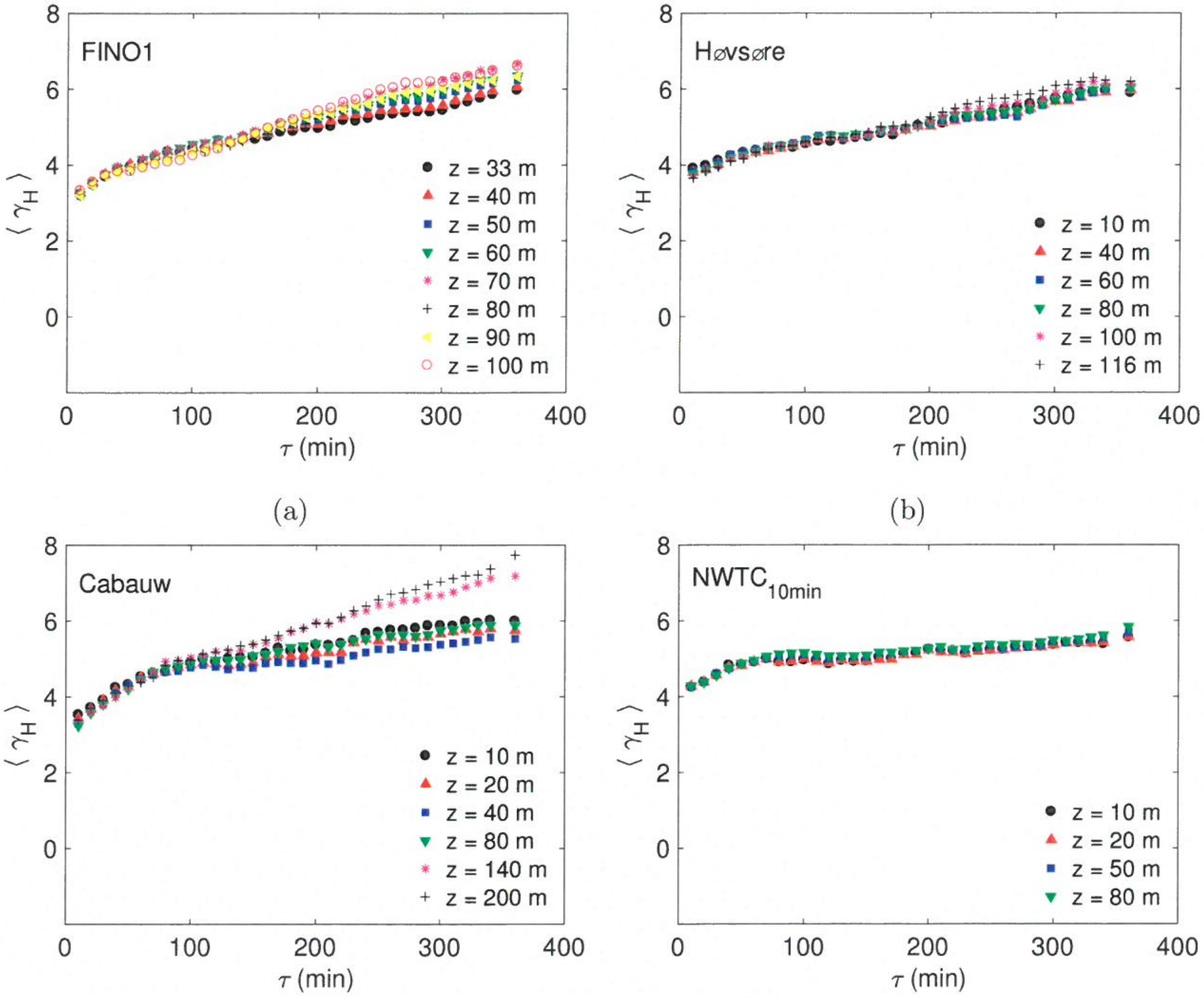
Figure 7.10: Same as Figure 7.9, except for time increment (τ) of 60 min.



(a) (b)

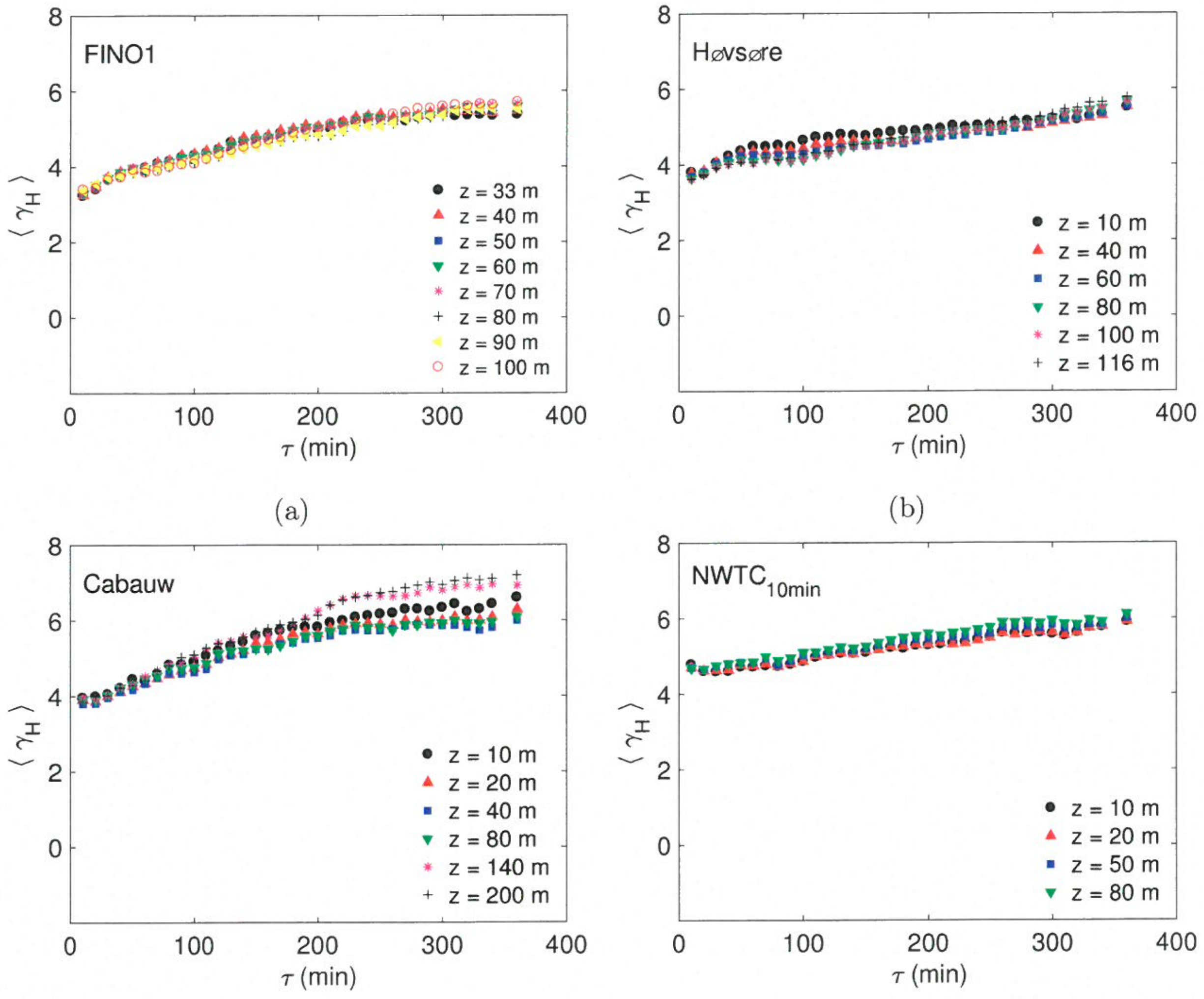
(c) (d)

Figure 7.11: Same as Figure 7.9, except for time increment (τ) of 180 min.



(a) (b) (c) (d)

Figure 7.12: The averaged values of tail indices ($\langle \gamma_H \rangle$) as a function time-increment (τ). Wind ramp-up data from all the available sensors from all the four sites are utilized here.



(a) (b)

(c) (d)

Figure 7.13: Same as Figure 7.12, except for ramp-down events.

7.5 Probability of Exceedance

Tables 7.2 and 7.3 showcase the probability of exceedance (%) for the right side of the distribution for a range of standard deviations from $3\sigma_u$ (near the core) up to $10\sigma_u$ (near the tails) for the 10 min and 60 min increment data, respectively. To compute these values, the approach along with the code is shown in Section A.6. Since the increments are normalized by their standard deviation, the coefficient corresponds to the $\delta u/\sigma_{\text{u}}$ value on Figure 7.1. Of note, for this analysis, the pdfs are converted to edf to avoid the use of histogram binning of the pdf, which has previously been shown to be sensitive to bin width selection.

From these tables the following conclusions can be made: (i) As σ_{u} increases the probability of exceeding a particular value decreases, which is clearly depicted from the monotonically decreasing probability with increasing δu, shown in Figure 7.1. (ii) The exceedance values for each site regardless of the tower levels are nearly identical. Further proving that the pdfs of wind speeds increments in the mesoscale range are not dependent on height. iii) The differences between the locations for a given σ_{u} value are minimal, but do indicate differences which are similiar to the visual inspection from the pdf plots. For instance, the NWTC reveals a lower probability of exceedance of the $10\sigma_{\text{u}}$ values (i.e. extreme events), however the other sites are not far behind.

With regard to the 60 min exceedance probability (Table 7.3), the similiarities between the values are more consistent indicating that from a 60 min increment perspective the data behaves in a quasi-universal manner, more so than at 10 min. Another interesting observation can be made from this table. For instance, the FINO-1 tower shows viritually identical exceedance values for the middle 4 to 5 tower levels. This suggests that over the ocean a very uniformed boundary layer is formed resulting in a consistent wind speed increment at a separation of 60 mins. In other words, this result indicates that the intensity of turbulence does not change much over a given time period, as ocean surfaces do not experience large diurnal cycles [246]. The influence on diurnal variability on the wind speed increments will be addressed in more details in the following chapter. Furthermore, for all locations the probability to exceed $10\sigma_{\text{u}}$ is lower than with the smaller τ increment, which indicates a reduction in the tails features as τ increases. This type of analysis provides a quantative metric into the similiarities and differences between the datasets and can shed light into potential physical explanatiions as to the features causing these contrasting features.

Table 7.2: Exceedance Probability (%) for right tail (τ = 10 min)

site	z (m)	F >$3\sigma_{\delta u}$	F >$5\sigma_{\delta u}$	F >$7\sigma_{\delta u}$	F >$10\sigma_{\delta u}$
FINO-1	33	9.58e-1	2.26e-1	6.76e-2	1.10e-2
	40	9.57e-1	2.24e-1	7.05e-2	1.08e-2
	50	9.57e-1	2.37e-1	7.36e-2	1.37e-2
	60	9.82e-1	2.39e-1	7.33e-2	1.37e-2
	70	9.67e-1	2.38e-1	7.13e-2	1.15e-2
	80	9.56e-1	2.32e-1	6.85e-2	1.33e-2
	90	9.67e-1	2.34e-1	7.12e-2	1.14e-2
	100	9.16e-1	2.01e-1	5.65e-2	1.03e-2
Høvsøre	10	8.09e-1	1.25e-1	2.85e-2	5.00e-3
	20	8.33e-1	1.40e-1	3.22e-2	5.55e-3
	40	8.54e-1	1.46e-1	3.20e-2	6.10e-3
	80	8.62e-1	1.54e-1	3.33e-2	5.18e-3
	140	8.56e-1	1.57e-1	3.53e-2	5.73e-3
	200	8.36e-1	1.52e-1	3.53e-2	5.58e-3
Cabauw	10	7.55e-1	1.34e-1	4.12e-2	1.13e-2
	20	7.73e-1	1.36e-1	4.31e-2	1.27e-2
	40	7.81e-1	1.47e-1	5.06e-2	1.63e-2
	80	7.83e-1	1.55e-1	5.16e-2	1.61e-2
	140	7.72e-1	1.47e-1	4.58e-2	1.55e-2
	200	7.67e-1	1.38e-1	4.52e-2	1.41e-2
NWTC	10	1.01e0	1.56e-1	2.71e-2	3.11e-3
	20	1.04e0	1.62e-1	2.87e-2	3.11e-3
	50	1.06e0	1.70e-1	3.30e-2	4.15e-3
	80	1.07e0	1.71e-1	3.37e-2	4.49e-3

7.6 Conclusions

In this study, we analyzed several long-term wind speed datasets comprised of four different geographical locations, from offshore to complex terrain. We showed that the wind ramp pdfs from all the sites reveal amazingly similar shape characteristics. Most interestingly, the tails of the wind ramp pdfs are much heavier than Gaussian and decay faster as time increments increase. With the aid of the Hill plots, we showed that the extreme ramp-up and ramp-down events behave similarly from a statistical point-of-view. Moreover, the tail-index statistics exhibited minimal dependence with respect to height above the ground.

Another important aspect of these results showed that the tails of the wind ramp distribu-

Table 7.3: Exceedance Probability (%) for right tail ($\tau = 60$ min)

site	z (m)	F $>3\sigma_{\delta u}$	F $>5\sigma_{\delta u}$	F $>7\sigma_{\delta u}$	F $>10\sigma_{\delta u}$
FINO-1	33	8.43e-1	1.15e-1	2.04e-2	1.18e-3
	40	8.43e-1	1.13e-1	2.07e-2	6.92e-4
	50	8.36e-1	1.15e-1	2.08e-2	4.57e-4
	60	8.37e-1	1.16e-1	2.06e-2	4.42e-4
	70	8.42e-1	1.14e-1	2.05e-2	4.42e-4
	80	8.04e-1	1.16e-1	2.10e-2	4.42e-4
	90	8.03e-1	1.18e-1	2.06e-2	4.38e-4
	100	7.57e-1	1.08e-1	2.44e-2	1.32e-3
Høvsøre	10	7.38e-1	8.81e-2	1.47e-2	1.67e-3
	40	7.42e-1	8.92e-2	1.54e-2	1.85e-3
	60	7.63e-1	9.03e-2	1.39e-2	1.67e-3
	80	7.65e-1	9.34e-2	1.36e-2	1.30e-3
	100	7.46e-1	9.00e-2	1.41e-2	1.48e-3
	116	7.12e-1	8.40e-2	1.47e-2	1.49e-3
Cabauw	10	6.33e-1	7.12e-2	1.28e-2	6.88e-4
	20	6.59e-1	7.51e-2	1.41e-2	8.22e-4
	40	6.85e-1	8.29e-2	1.51e-2	1.09e-3
	80	6.72e-1	8.02e-2	1.53e-2	8.21e-4
	140	6.38e-1	7.58e-2	1.43e-2	5.48e-4
	200	6.20e-1	7.56e-2	1.34e-2	1.78e-3
NWTC	10	1.02e0	1.23e-1	1.69e-2	5.19e-4
	20	1.02e0	1.20e-1	1.64e-2	5.19e-4
	50	1.00e0	1.08e-1	1.40e-2	3.46e-4
	80	9.91e-1	1.04e-1	1.21e-2	5.19e-4

tions do not follow a power-law distribution, rather modulated by a slowly varying function. We speculate this function to be an exponential. As shown in earlier chapters, several types of pdfs from the generalized hyperbolic distribution family (e.g., GHSST, NIG) can be considered as ideal candidates for capturing the tail characteristics of wind ramp distributions. We also highlighted the pros and cons associated with the different models, which should be taken into account when representing these quasi-universal distributions.

As the wind energy industry continues to flourish and wind ramp prediction becomes increasingly important, the results from this study should be utilized for model validations and improvement. It would be critical to find out if the state-of-the-art numerical weather prediction models and time-series forecasting tools are able to capture the extreme ramp behaviors accurately. It is also envisaged that the contemporary synthetic wind speed generators (e.g., [74, 189]), heavily relying on statistical information, will tremendously benefit from our findings.

Chapter 8

The Probability Density Functions of Wind Speed Increments in the Mesoscale Range

8.1 Introduction

Extending the work from the previous chapter, the focus will now shift to determining a pdf model which is capable of capturing the features depicted in Chapter 7 over a range of separation. As discussed in detail in Section 2.3, only a few studies have addressed the statistical characteristics of wind speed within the mesoscale range. In these studies, generally speaking, the pdfs of wind speed increments are shown to portray similar feature and fit typical small-scale model, yet a general consensus into which model provides the most reliable fit is still lacking and there has not been a quantitative examination.

An additional factor that has not been addressed is the variability on the mesoscale wind increments pdfs between day and night conditions. The Earth's atmosphere undergoes drastic changes within the boundary layer between the day and night cycle. For instance, buoyancy and shear generation contributes differently to the total turbulent kinetic energy (TKE) production depending the time of day. Recently, Kiliyanpilakkil and Basu (2015, [136]) studied the impacts night and day environments have on mesoscale wind scaling characteristics and found a larger exponent (ζ_2) for second order structure function during night time over land as compared to a more convective, day time regime. This finding suggests more intermittency occurs within the night time boundary layer with respect to height. Also, they revealed that over the ocean there is a negligible difference in ζ_2 ensued between day and night. Therefore, a closer examination into the impacts these atmospheric condition has on the pdf features will be conducted.

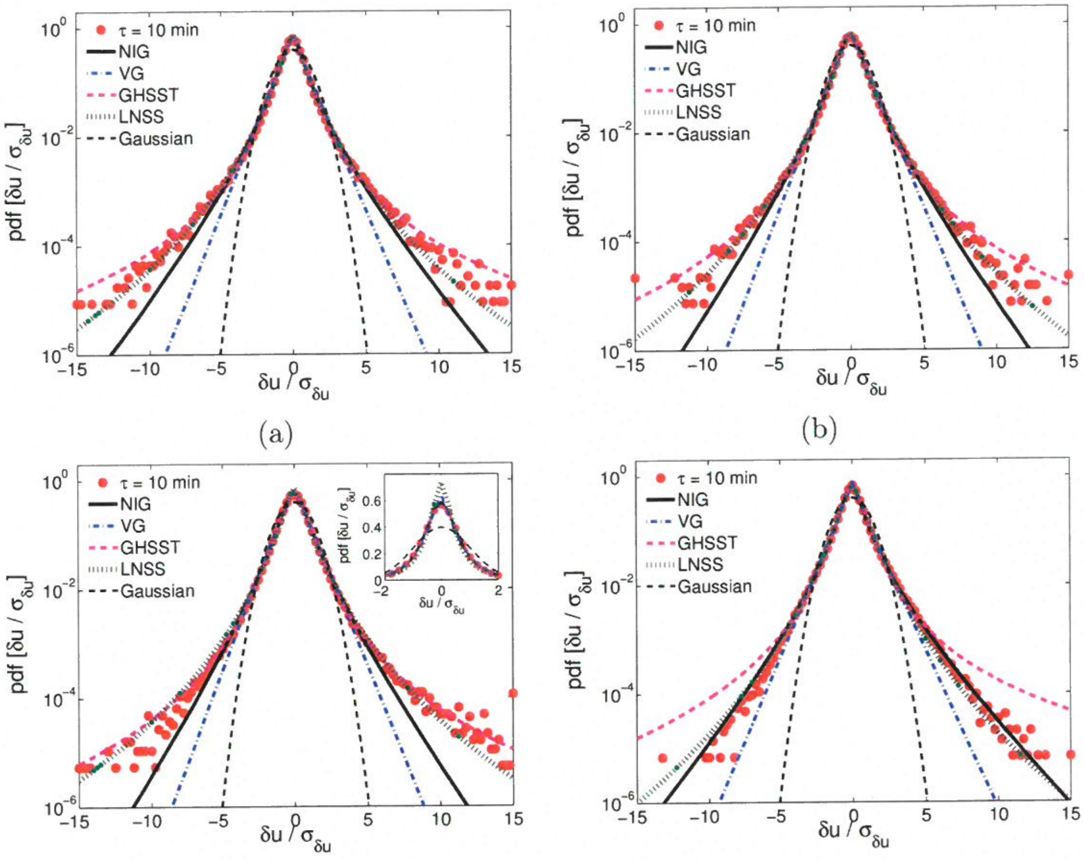
A main objective for this work is to determine if a particulr model is able to capture the behavior of the mesoscale wind speed increments, similiar to the small-scale counterparts. We estimate the parameters of these models using the maximum likelihood estimation technique, as previously shown, and examine the quality of the fit using two goodness of fit (GoF) metrics, Kolmogorov-Smirnov (K-S) and the two-sided Anderson-Darling (A-D) tests for our evaluation. We will discuss the pros and cons of each model and with the use of the model parameters, we will address the impacts day and night have on pdf characteristics.

8.2 Mesoscale Probability Density Functions with Model Fit

The observed wind speed increment pdfs from the four locations and tower heights (FINO1 (100m), Høvsøre (100m), Cabauw (80m) and NWTC (80m)) are presented in Figures 8.1 and 8.2 for 10-min and 360-min temporal separation (τ), respectively. Overlaid on each figure are the four different pdf models discussed in Chapter 3. The parameters of these individual models were estimated using the MLE-based approached. However, for the LNSS model the method of moments approach is utilized, see Eq. 3.11. Due to extreme computational expense for the MLE calculation for LNSS, we are limited to showcasing the results in this manner. Thus, our intentions are to highlight the limitations with the proposed LNSS model and estimation method. In Figure 8.1c the inset shows the linear-linear representation of the 10 min wind speed increment pdf for the 80 m height at Cabauw. This plot is highlighting the over estimation of LNSS-MME near the peak of the distribution, which is not as evident in the parent plot. Thus, we caution that portraying pdf results via log-linear can mask some of the important features near the peak of the distribution. Similarly, depicting results in a linear-linear fashion can hide the behavior of the tails, as mentioned in Chapter 2. Therefore, the use of GoF testing leads to a more robust examination into the quality of a given model's fit, a similar discussion was made in Section 5.5.

The pdf shapes in Fig 8.1 clearly display features that are synonymous with the small-scale counterpart, in which the tails extended outward towards the larger extreme values and the peaks are narrower than Gaussian. Though, this is evident across all locations, there are varying degrees of intermittency present depending on the location. For instance, the tails of the distribution for NWTC do not exhibit the same features as the other locations. This artifact is likely due to the variations in the geography/topography, as NWTC is located at the foothills of the Colorado Rockies. With regard to the models fit, all models visually provide a very reliable fit up to about 10^{-3} probability, however near the tails there are clear difference between all models and locations. From a qualitative perspective, GHSST and LNSS models provide the best fit near the tails for FINO-1, Høvsøre and Cabauw, while at NWTC, NIG and LNSS appear to capture the shape most accurately.

As the separation increases, Figure 8.2 provides evidence that the shapes of the pdfs tend towards a Gaussian shape, similiar to the observed trends in small-scale turbulence. However, at NWTC there is still a quasi-heavy tail feature, which suggests a higher likelihood of stronger wind gusts as opposed to the other locations at 6 hr increments. Also, at NWTC there appears to be a moderate peakedness, relative to the Gaussian distribution, near the mode, which is a signature of increased intermittency still contributing to the flow characteristics at 6 hour time scales. Overall, the pdf models' fits at τ=360 min are all generally in line except for the marginal over-estimation of GHSST across the board. While VG distribution appears be the most consistent fit for all locations at this separation. These results signify three points: (i) The wind speed increments within the mesoscale range behave similar to the small-scale turbulent regime, as remarked by Muzy *et al.* (2010, [185]); (ii) Though, features appear to



(a)

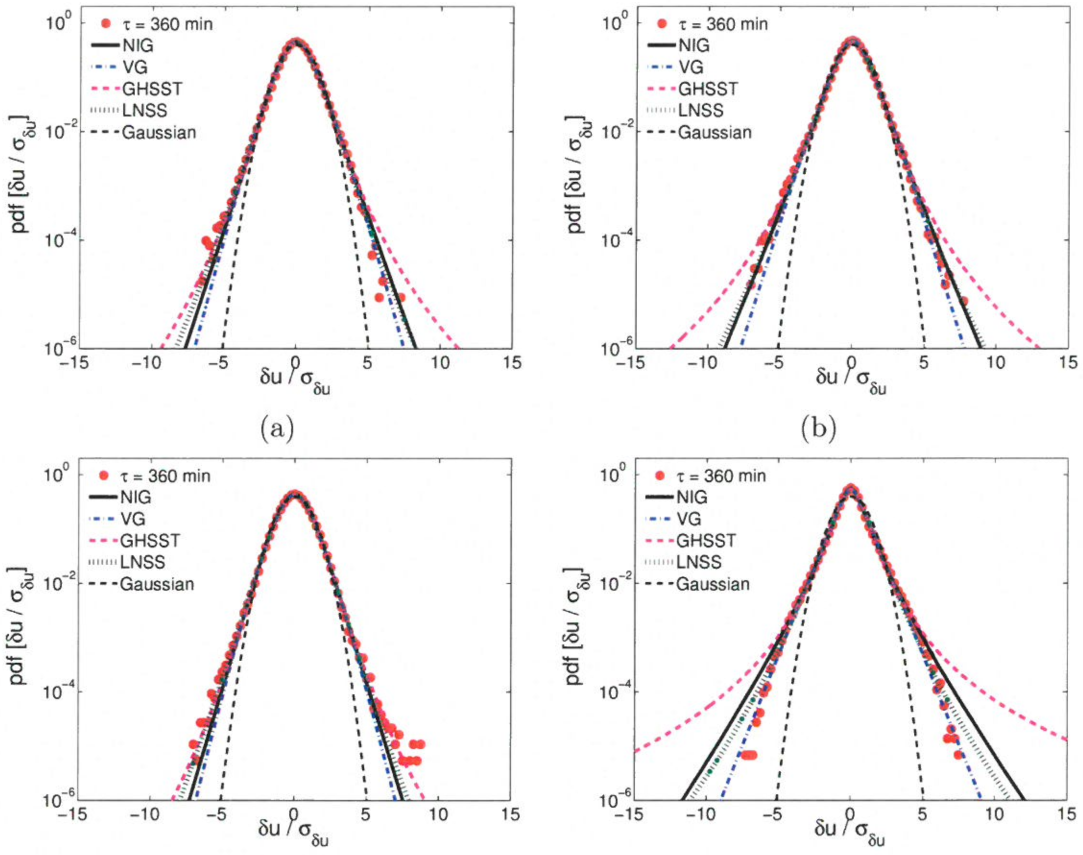
(b)

(c)

(d)

Figure 8.1: 10-min wind speed increment pdf for the following locations and tower heights in order of complexity from smooth ocean surface to complex terrain. a) FINO1-100 m; b) Høvsøre-100 m; c) Cabauw-80 m; d) NWTC-80 m. Overlaid on each figure are the estimation model pdfs using MLE for NIG, VG, and GHSST and MME for LNSS. The dashed black line represents the Gaussian distribution.

be quasi-universal, differences in the shapes are present across diverse geographical locations, as discussed in Chapter 7; (iii) Each of the pdf models is capable of revealing the unique features of the underlying data, yet further quantitative evaluation is need to determine the appropriateness of a given model of choice.



(a) (b) (c) (d)

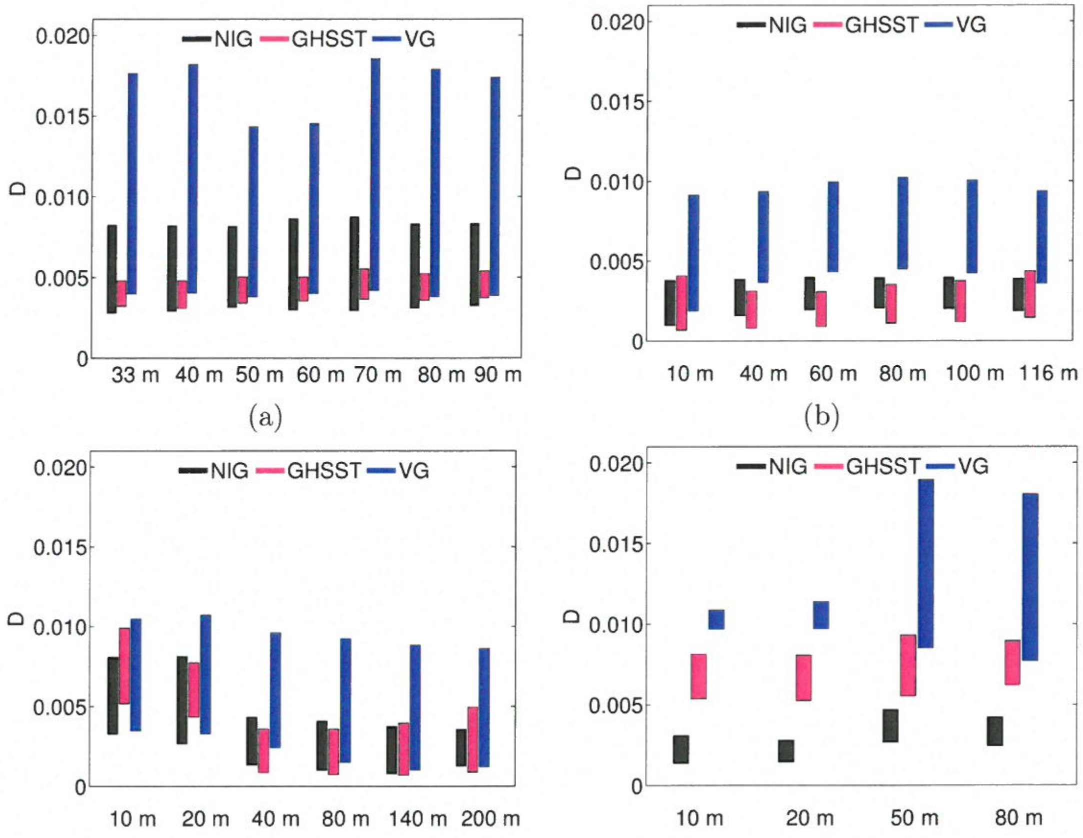
Figure 8.2: Same as Figure 8.1 except for the 360-min wind speed increment pdf.

8.3 Quantile-Quantile Plots

Given the accuracy of NIG and GHSST throughout this work, we will now utilize a Quantile-Quantile (Q-Q) plot representation to show an alternative quality of the fit. Q-Q plots are commonly used in statistics to compare and determine if two distributions are the same. In a non-parametric way, we are comparing the observed wind speed increments across various temporal separations (10, 60, and 360 min) against the corresponding estimated pdf models (NIG in Figure 8.4 and GHSST in Figure 8.5). The predicted NIG values are constructed by an inverse cdf function, as shown in Appendix A.1. A similiar computation is done for GHSST (not shown). If the sample and theoretical distributions are identical then the data points will fall on the dashed red line.

In Figure 8.4, it is clear to see that NIG, using the MLE technique, can estimate well the normalized wind speed increment sample for all locations and increments up to ~ ±5. In other words, this confirms the accuracy of the up to $5\sigma_{\mathrm{u}}$. Moreover, as the separation increases, the model fits are significantly improved and more or less fall in line with the observations. However, as indicated in the pdf plots, the fit of the model is problematic such that NIG underestimates the tails of the distribution for small τ values, especially for the three North Sea locations (top three rows). But for NWTC, the NIG fit is virtually perfect and can be confidently used for estimating the mesoscale wind speed increment data for this complex terrain environment. From a quantative perspective, it is clear to see, from the legends, that the K-S and A-D results in-line up with the visual comparisons. However, since A-D puts more emphasis on the tails, the deviation between locations is more amplified due to the test statistic being more sensitivity to the accuracy of the prediction of the extreme and rare events.

In a similar manner, Figure 8.5 shows the comparison of the observed wind speed increments against the GHSST model. Interestingly, despite the visual accuracy of GHSST, the Q-Q representation reveals a not-so-accurate result. Unlikely NIG, GHSST equally overestimates, from each side of the distribution, the observed data beyond ±5. This remains fairly consistent throughout the sites and for the various τ values. However, with NWTC the GHSST model is a very poor estimator, likely due to the steeper tail characteristics evident from the pdf plots. Again, this "3-D" like turbulent feature is an artifact of the complex terrain and is therefore a factor that needs to be considered in terms of providing a suitable pdf model for predicting the wind ramp conditions. Given the fact, that these two model lead to relatively poor estimation near the tails of the distribution, we will now evaluate the GoF of these models utilizing the K-S and two-tailed A-D test statistics.



(a) (b)

(c) (d)

Figure 8.3: K-S (D) box plots for the wind speed time series increments from a) FINO-1, b) Høvsøre, c) Cabauw, and d) NWTC towers.

8.4 Goodness of Fit Evaluation

Figure 8.3 shows the K-S edf statistic (D) comparing NIG, GHSST, and VG against the wind speed increment for the particular sites, where the lower the value represents the better fit. Each box plot contains the D values for a range of separations from 10 to 360 mins. From this perspective, the results are further indicating that VG is a model that is unable to represent the mesoscale wind speed increments well enough to be considered. This result is in agreement with the visual inspection of the pdfs in Figure 8.1 as VG is shown to drastically under estimate the observed data for small increments, leading to a poor statistical test. However, NIG and

GHSST reveal more adequate fit to the data with D values remaining below 0.01, with GHSST minimally providing a more accurate fit. Also, the differences between the fit as a function of height is minimal within each location, suggesting the there is no height dependence. The clear outlier is the NWTC location in Figure 8.3d, in this instance NIG well outperforms the other two models. Again, this location represents the complex terrain and distributions do not portray similar heavy-tailed features from the other near-homogeneous terrain location. This finding provides an indication that the mesoscale range behavior is quasi-universal and perhaps two models pdfs can be capable of approximating the variations of all ABL flows.

As previous shown in Chapter 5, the Anderson-Daring (A-D) test statistics provides a usefuly metric for determining the GoF giving more weight to the tails of the distribution. As we have presented the tail behavior is a unique feature for the mesoscale range wind speed increments (see Section 5.5 for detail). A-D statistic can also reveal the difference between the right and left tails of the distribution separately. Equation 8.1 shows the modified statistics based on the right and left tail, respectively [160] and the results from mesoscale wind speed increments are displayed in Table 8.1,

$$A_R^2 = n \int_{-\infty}^{\infty} \frac{\{F(x) - F_{EMP}(x)\}^2}{1 - F(x)} dF(x), \tag{8.1a}$$

$$A_L^2 = n \int_{-\infty}^{\infty} \frac{\{F(x) - F_{EMP}(x)\}^2}{F(x)} dF(x), \tag{8.1b}$$

where $F_{EMP}(x)$ is the empirical distribution function, $F(x)$ is the estimated distribution function of the model, and n is the sample size of the evaluated points of the corresponding distributions. These results show the 5%, 95%, and median values from all levels for the given observational tower and respective τ values. Generally, speaking the GHSST model outperforms NIG for the first three sites, however NIG is much more accurate model compared to GHSST for NWTC, due to its lighter tails. Also, as separation increases, the NIG model, for the left and right tails, indicates an improvement in the estimation for first three sites, but the opposite is true for NWTC. However, GHSST gets worse as separation increases for all sites. Another key finding is that the spread of values are largest at Cabauw providing evidence that the pdfs of wind speed increments vary with height as well as the NIG prediction. Finally, for the locations which portary heavier tail features (North Sea sites), the general conclusion suggests that the model estimation is less accurate for the right tail behavior compared to the left, for small τ.

These statistically evaluations not only reveal a quantative metric for the accuracy of the model fits, but can also divulge information regarding specific differences in particular pdf feature (i.e., skewed, heavy tailedness). Thus, using the NIG model and the steepness parameter (α), we will examine the variations in the tail features as a function of τ for day and night time period, which possess distinctly different atmospheric boundary characteristics.

8.5 Day and Night Impacts on Wind Speed Increments

The diurnal variations are one of the key signatures within the ABL. As incoming solar radiation warms up the earth's land surface changes in the near surface characteristics occur leading to turbulence transport processes modifying the structure of the ABL [246]. While, during the night hours, the surface begins to cool due to long wave radiative cooling creating a more stable ABL which restrict turbulence due to buoyancy. From a wind speed increment perspective, few have studied the differences between day and night characteristics. As an example, it was shown in Vindel et al. (2008, [266]) that for a wide range of atmospheric stability throughout the diurnal cycle, the flatness (see Eq. 2.5) changes considerably due to intermittency variations. For instance, for their study they evaluated atmospheric boundary layer wind increments for a single field campaign and found that during night time hours the degree of intermittency was reduced, resulting in a near-Gaussian distribution form during the night and a more non-Gaussian feature during the day. Unfortunately, this analysis was only conducted during a very short time period and examined only for small scale (sub-minute) increments. Therefore, using the long-term time series data, we will investigate the variations in the pdf characteristic

Table 8.1: $A_L^2$and A_R^2 statistics for the mesoscale wind speed increments the four sites at $\tau =$ 10 min, 60 min, and 360 min. The results are the median, 5%, and 95% height values for the NIG and GHSST models. See Eq. 8.1.

			PDF Model					
			NIG			**GHSST**		
Sites	τ (min)	tail	5%	50%	95%	5%	50%	95%
FINO 1	10	L	4.22	5.50	5.74	1.57	3.35	3.83
		R	6.44	6.83	7.63	3.05	5.25	5.99
	60	L	6.72	7.82	19.93	0.73	4.46	5.46
		R	2.02	2.93	7.64	3.10	7.78	9.90
	360	L	3.57	4.87	6.27	0.96	1.31	2.29
		R	1.04	1.54	2.95	5.07	6.22	8.66
Høvsøre	10	L	67.39	86.04	124.19	7.58	17.54	23.19
		R	88.13	107.93	178.58	21.47	35.61	44.70
	60	L	75.80	183.97	217.21	13.89	15.20	17.07
		R	89.75	99.40	140.44	36.02	44.53	57.21
	360	L	37.96	51.06	70.96	19.21	54.64	61.30
		R	17.79	32.02	49.36	133.73	181.57	190.37
Cabauw	10	L	209.06	2167.52	7744.77	21.83	92.36	113.89
		R	2023.15	3511.50	8824.67	10.94	45.71	57.23
	60	L	92.92	621.94	849.50	24.15	32.25	39.66
		R	156.30	706.29	833.47	12.18	21.92	61.05
	360	L	11.20	62.67	180.56	27.60	116.27	405.58
		R	43.74	168.87	268.83	42.27	78.08	241.78
NWTC	10	L	36.87	69.10	81.88	534.22	697.23	750.15
		R	7.68	9.85	12.13	897.94	1215.57	1510.12
	60	L	76.63	119.70	136.73	1393.07	1442.79	1634.45
		R	93.86	130.72	180.15	2142.66	2444.92	2587.43
	360	L	403.69	509.09	647.96	1877.15	2342.44	2874.33
		R	417.24	503.04	657.06	2398.66	2835.15	3445.95

between day and night time measurement.

Now that the model evaluation has been conducting and the NIG-based estimation has shown to provide an acceptable representation of the mesoscale wind speed increments, we will utilize the α parameter as an illustrative example[1]. Before we break down the day-night

---

[1] GHSST also provides a quality estimation of the mesoscale wind pdfs, however, since the moments of NIG are all finite, this model is the most desirable to use.

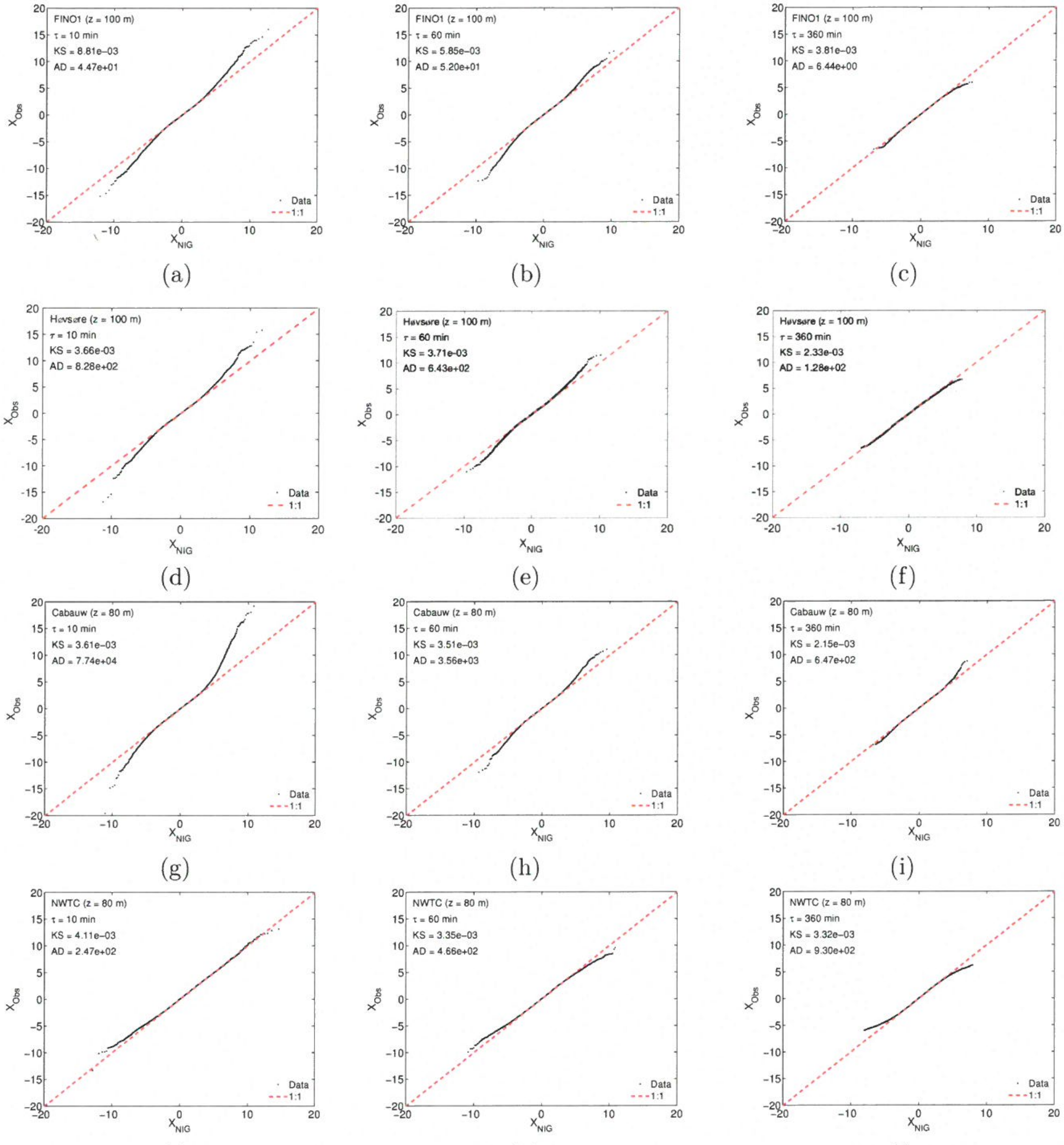
series, Figure 8.6 shows α (estimated from the normalized wind speed increment) as a function of τ for the four locations. Overall the location there doesn't appear to be much of a height dependence, except for at larger scales (i.e., $\tau > 120$ mins) at Cabauw. Also, the α values are monotonically increase as a function of τ, which suggests that all the distributions are tending towards Gaussian as separation increases. This can also be seen with pdf plots shown in Figures 8.1 and 8.2. To investigate the impacts night and day has on the wind speed increments, the data is separated by two, six hour periods, 11 to 17 local time for day and 23 to 05 local time for night.

Figure 8.7 shows α as a function of τ for the four tower locations. The left and right columns depict the day and night time results, respectively. From this perspective, there are clear and interesting similiarities and differences between α during day and night time conditions. From an initial observation, α increasing with respect to increasing τ for both day and night time. This indicates a change in the steepness (or heaviness of the tail) of the pdfs from semi-heavy tail at small increments to near-Gaussain as τ approaches 360 min in both conditions. However, comparing day and night at small scales, there is a minor indication that the tails of the distribution are slightly heavier during night time conditions as opposed to day time. This suggests that at 10 min increments, the night time ABL exhibits slight more intermittent features. Furthermore, comparing FINO-1 day versus night there is not a marked difference between the behavior of the pdfs. This interesting finding can also be supported by the fact that over the ocean, where the ABL does not possess a strong diurnal cycle, one would not expect to see significant difference in the turbulent features even within the mesoscale.

In terms of the height dependence, there are strikingly different features that exist between day and night. For instance, at all locations there does not appear to be significant different in α with respect to height. However, for the night time condtions, especially over land, the steepness of the distribution clearly is changing with height. For example, Cabauw shows that α increases more rapid within increase height as separation increases suggesting that the tails of the distribution remain heavier tailed (more intermittency) beyond 60 min at lower heights above the ground. Similiar results ensue over NWTC, though not as drastic. For Høvsøre the changes in the α occur at larger scales, though it is still a marine environment and would be less influenced by diurnal fluctuations. Using NIG and the steepness parameter α provided a useful method to diagnosis the features of the wind speed increment pdf during the day and night conditions.

8.6 Conclusion

In this work, we analyzed the wind speed increments from long term time series wind measurement over diverse geographical locations. A quasi-universal multiscaling behavior was shown for these distributions over a range of separation from 10 min to 6 hr. Utilizing the MLE technique, we illustrated the ability of a number of pdf models to capture the inherent features of the empirical pdfs. Though these models visually revealed a degree of percision in the fits, two goodness of fit metrics were explored to determine the model with the best fit. The reliability of both the NIG and GHSST distributions were demonstrated through the visual inspection of Q-Q plots along with K-S and two-tailed A-D test statistics. From these statistics it was shown that NIG is more accurate over complex terrain while GHSST can capture the wind speed increment characteristics over homogeneous terrain. Using the NIG-based estimation, specifically α, we showcased the changes in the steepness of the pdf between day and night time conditions revealing interesting features for various terrain types and temporal separations. In future work, we will explore the numerical weather prediction models capability of capturing these mesoscale features to use as a benchmark. However, first we will explore the mesoscale temperature increments and determine if a suitable model is capable of representing this variable.



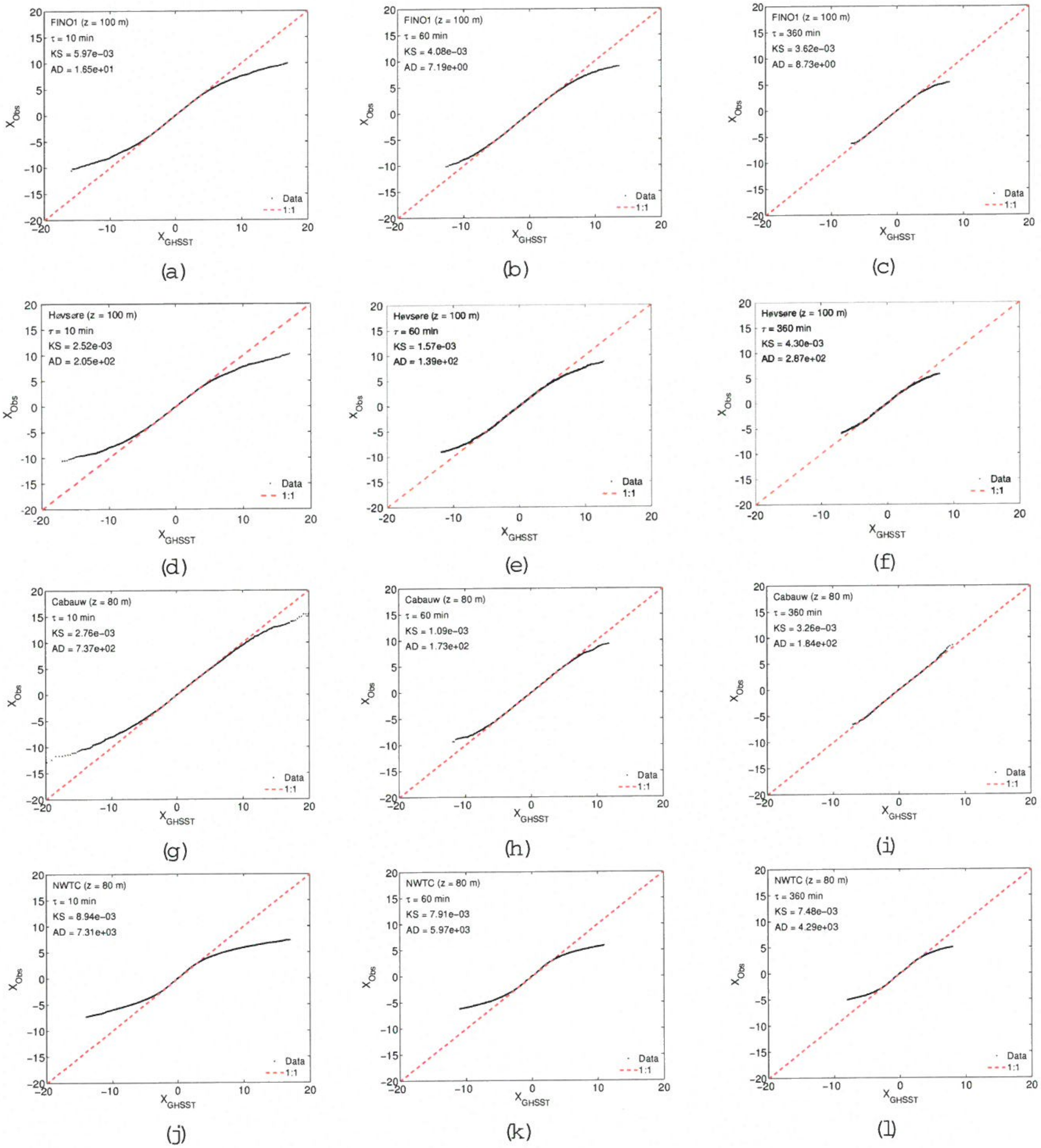
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure 8.4: Q-Q plots comparing wind speed increments against the estimated NIG model distributions for three different τ (10 min, 60 min, and 360 min) and four locations (FINO1, Høvsøre, Cabauw, and NWTC). The corresponding K-S and A-D test statistics are depicted in the legend.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure 8.5: Q-Q plots comparing wind speed increments against the estimated GHSST model distributions for three different τ (10 min, 60 min, and 360 min) and four locations (FINO1, Høvsøre, Cabauw, and NWTC).

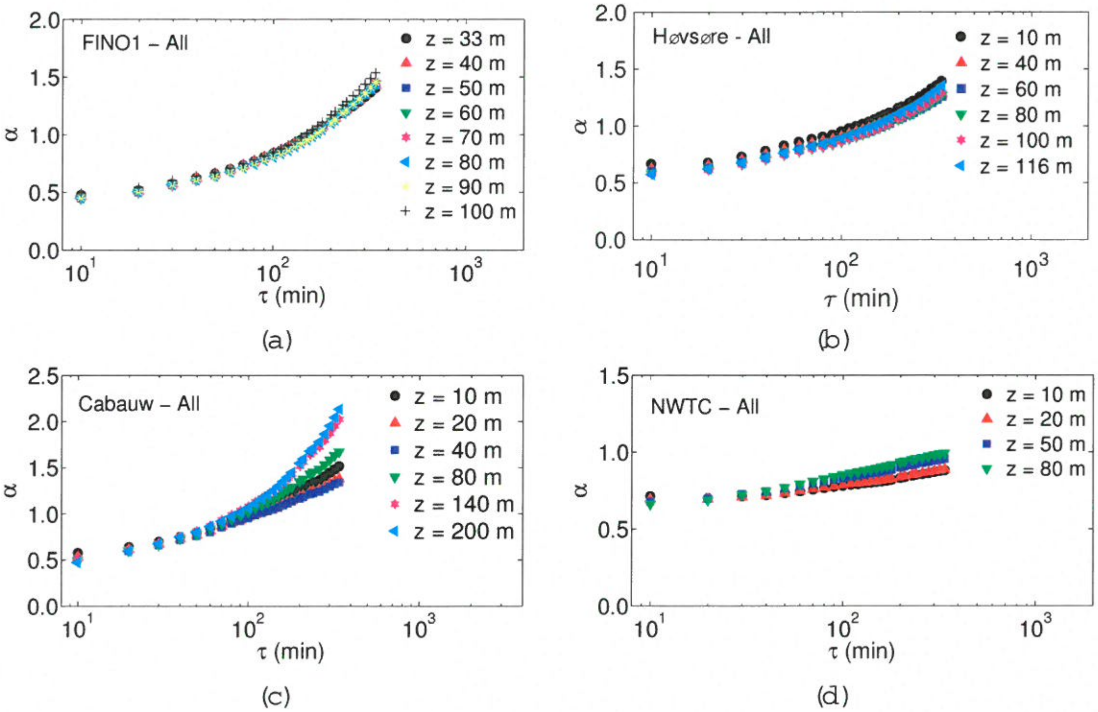
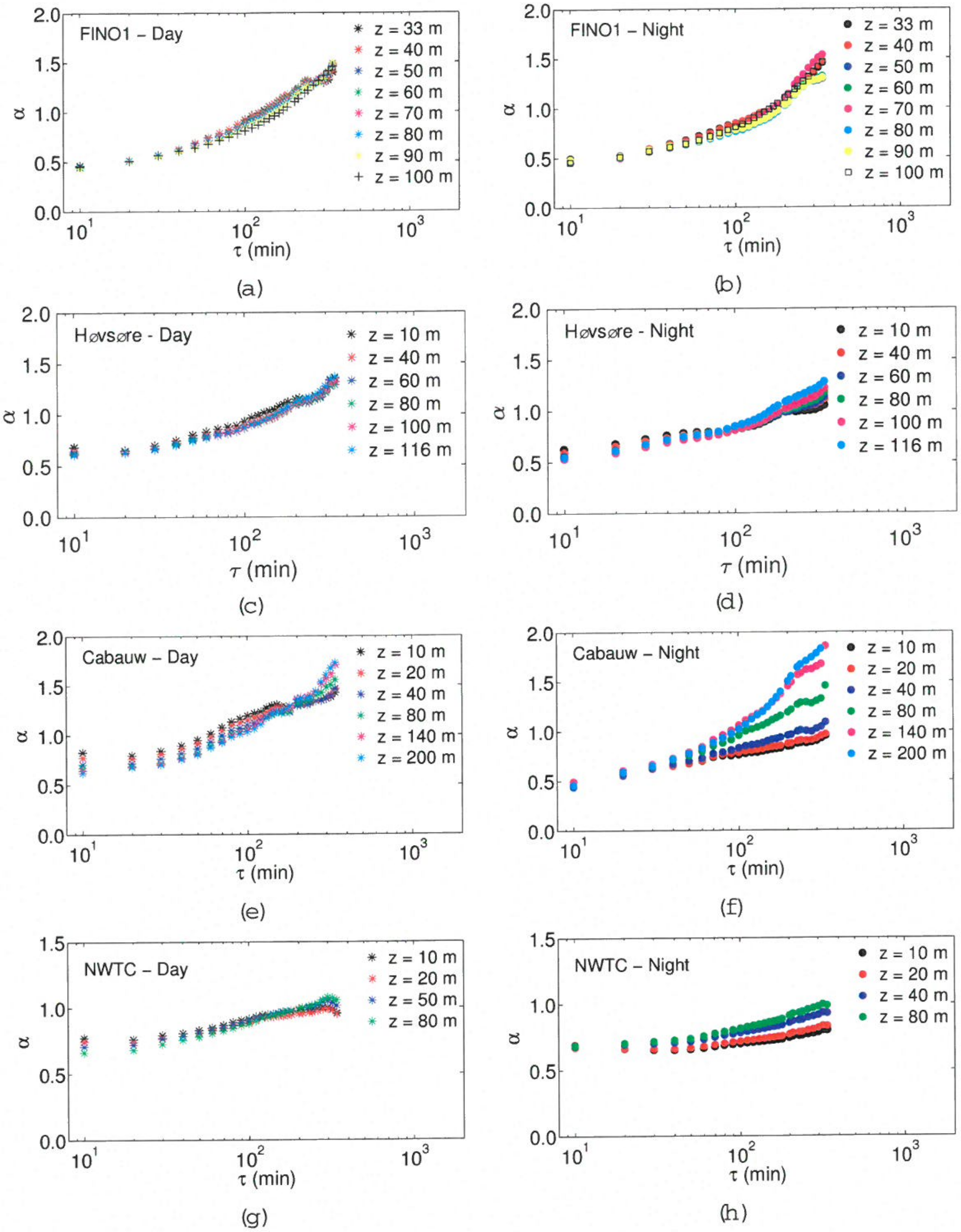


Figure 8.6: NIG parameter (α) for full series of wind speed increments as a function of τ for a) FINO1, b) Høvsøre, c) Cabauw, and d) NWTC for different heights.



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

Figure 8.7: NIG parameter (α) for day (left column) and night (right column) series of wind speed increments as a function of τ for a)-b) FINO1, c)-d) Høvsøre, e)-f) Cabauw, and g)-h) NWTC for different heights.

Chapter 9

Characterizing the Probability Density Functions of Temperature Increments in the Mesoscale Range

9.1 Introduction

In this chapter, we investigate the properties of the mesoscale temperature increments within the ABL utilizing a set of near-surface tower data as used in the pervious two chapters. Details of the observational data are provided in Section 7.2. Thus far, we have demonstrated that the normalized mesoscale wind speed increments behave in a quasi-unverisal manner over various geographical locations and can reasonably be estimated with NIG and/or GHSST via the MLE-based approach. Therefore, these results led us to investigate whether or not a similiar behavior and capability exist within the temperature increment field in the ABL.

Within the small scale temperature field, experimental evidence suggests statistical differences between the velocity and temperature increments (e.g., [14, 174, 239]). These difference reveal stronger intermitteny (i.e., larger fluctuations) within the temperature field, leading to smaller scaling exponents (ζ_p) for higher order moments ($p > 3$), see Section 2.4. These studies determined ζ_p values under which temperature was treated essentially as a passive scalar (i.e., the velocity field was unaffected by the transport of the scalar). However, in larger scale motions it is believed that temperature characteristics, are considered as an active scalar, in which temperature fluctuations directly affect the velocity field, via buoyancy forces (e.g., convection and stably stratified flows). Therefore, the temperature field, within large-scale atmospheric flows, is an important aspect to the dynamics of turbulent structures due to the inherent buoyancy forces [45, 181]. Buoyancy manifests itself in a number of ways, e.g., in the premature truncation of the upper end of the inertial range, and in a direct influence on large scales. However, there are no studies examining the traits of temperature increments within the mesoscale range, especially from a pdf perspective. Unlike wind speed in the mesoscale range, the temperature field exhibits strong variations as a result of the diurnal fluctuations that naturally occur due to the change in the short and longwave solar radiation flux over the course of a diurnal cycle. As such, one may expect that temperature increments in the mesoscale range also behave quite differently compared to the wind speed.

A major motivation of this work is related to the impacts temperature fluctuations (or equivalently density fluctuations) have on optical wave propagation. For instance, it is known that fluctuations of scintillation (variations in optical properties through a medium) is directly proportional to the changes in temperature fluctuations [92]. This relationship is still widely studied within the small-scale range; however, influences on the larger scale fluctuations are still a mystery. "Eddies of scale sizes smaller than L_0 (i.e., outer length scale of turbulence) are assumed statistically homogeneous and isotropic, whereas those equal to or larger than L_0 (i.e., eddy sizes on the order of 10s of meters) are generally non-isotropic (or anisotropic) and their structure is not well understood, Andrews and Phillips (2005, [10])." In practice, one generally assumes that the temperature fluctuations (δT) behave under K41 law, in which the small-scale features are universal and possess a scaling exponent of $\zeta_2 = 2/3$ through the inertial range. However, as we have shown this assumption does not necessarily hold for observed data for small scale temperature and/or wind increments. Therefore, the intent of this work is to expand this idea well beyond the inertial range to identify the pdf characteristics of temperature increments within the mesoscale range to determine the behavior and showcase the capability of pdf models to capture these traits. Given the stark diurnal variations, we will also explore the similiarities and differences in the pdfs between day and night time conditions.

9.2 Probability Density Functions of Mesoscale Temperature Increments

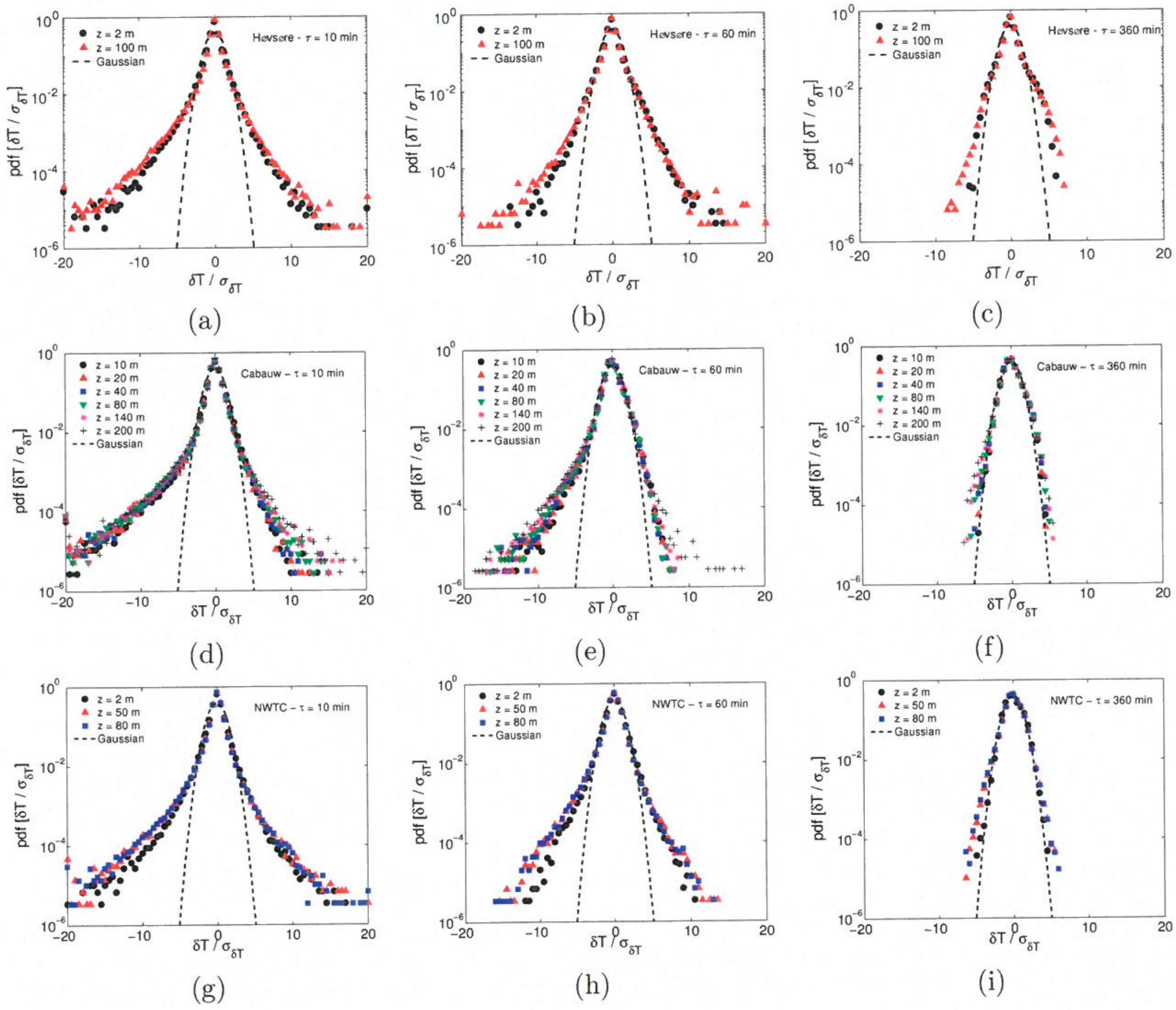
Figure 9.1 shows the pdfs of normalized temperature increments of three locations (Høvsøre, Cabauw, and NWTC) for three separation temportal increments (τ), no temperature data is available for FINO1. Overlaid on each plot are the corresponding heights above the ground. From this perspective, it is clear to see that the series of increments across the locations all exhibit non-Gaussian behavior. Also, as separation increases the distributions begin to revert back to a more Gaussian shape. However, unlike the wind speed increments (e.g., Figure 8.1) for the same locations and time periods, the tails of the temperature distribution are clearly heavier [1]. Additionally, the cores of these distribution reveal more peakedness indicating that a large amount of the fluctuations are present around the center. This corroborates the ideas presented in the small-scale temperature increment field.

Furthermore, and perhaps more intriguing, is the skewness evident in the left tails of each of the locations (Cabauw shows a much more skewed appearance through the 60 min increment). As shown in Chapter 6 with the wind-tunnel experiment, the temperature field also revealed a strong left skewed distribution, which was speculated to be due to the ramp-cliff features in the temperature field. However, in the mesoscale range this negative tail feature is more pronounced and is likely due to the larger scale atmospheric features which produce abrupt negative changes in temperature (i.e., cold frontal passage and convective downburst). This artifact is not observed in positive tail since drastic positive changes in temperature do not

---

[1] Of note, the scale on the x-axis for the temperature increment extends to $\delta T/\sigma_{\delta T} \pm 20$, whereas the wind speed range was between $\delta u/\sigma_{\delta u} \pm 15$, further amplifying the degree of heavy tailedness.

generally exist. Finally, a subtle difference with respect to height is noticable, such that slightly less intermittency is shown at smaller scales suggesting that fluctuations closer to the surface are less prevalent as oppose to temperature fluctuations higher in the ABL. These results provide an interesting perspective in the characteristics of the mesoscale temperature increments revealing that 10 min to 6 hr scales behave similar to small scale scalar increments. Next, we will evaluate fits of the pdf models.



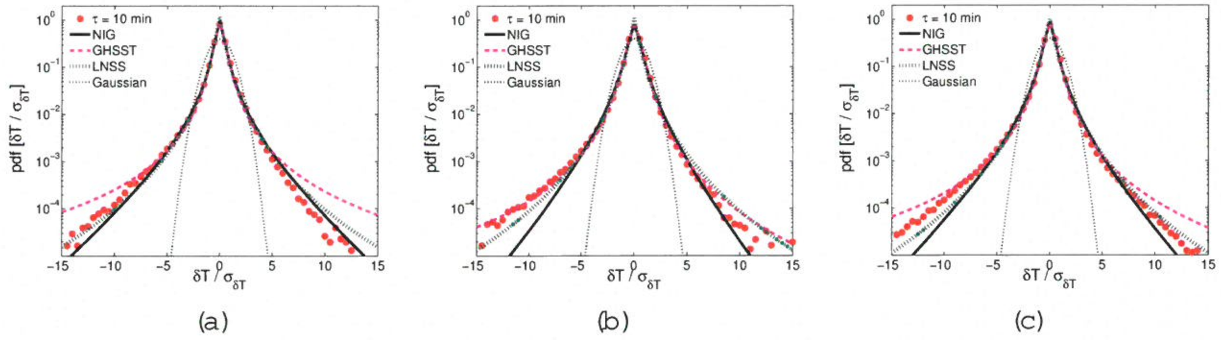
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 9.1: Probability density functions for mesoscale temperature increments across various geographical locations and a range of temporal separation τ. (a-c) Høvsøre, (d-f) Cabauw, (g-i) NWTC. Overlaid on each figure are the heights of the temperature measurements

Figure 9.2 is an illustrative example to highlight the fit potential for the mesoscale temperature increments for $\tau = 10$ min. Utilizing the MLE-based estimation for NIG and GHSST and MME-based for LNSS, clearly the models perform well with a few exceptions. For instance, NIG slightly underestimates the left tails of the distribution for Cabauw and NWTC, while GHSST overestimates the tails in all cases except for the heavy left tail for Cabauw. Moveover, LNSS shows deviation in both tails for all sites and it also over predicts the core of the distribution similar to what was revealed in Figure 8.1. The underestimation and overestimation in the tails for NIG and GHSST, respectively, are similar features seen in previous results of small and mesoscale increments. Therefore, to gain a full appreciation of the fit, we will evaluation the accuracy of NIG and GHSST from a quantative perspective.



(a) (b) (c)

Figure 9.2: τ = 180 min. a) Høvsøre-100 m; b) Cabauw-80 m; c) NWTC-80 m

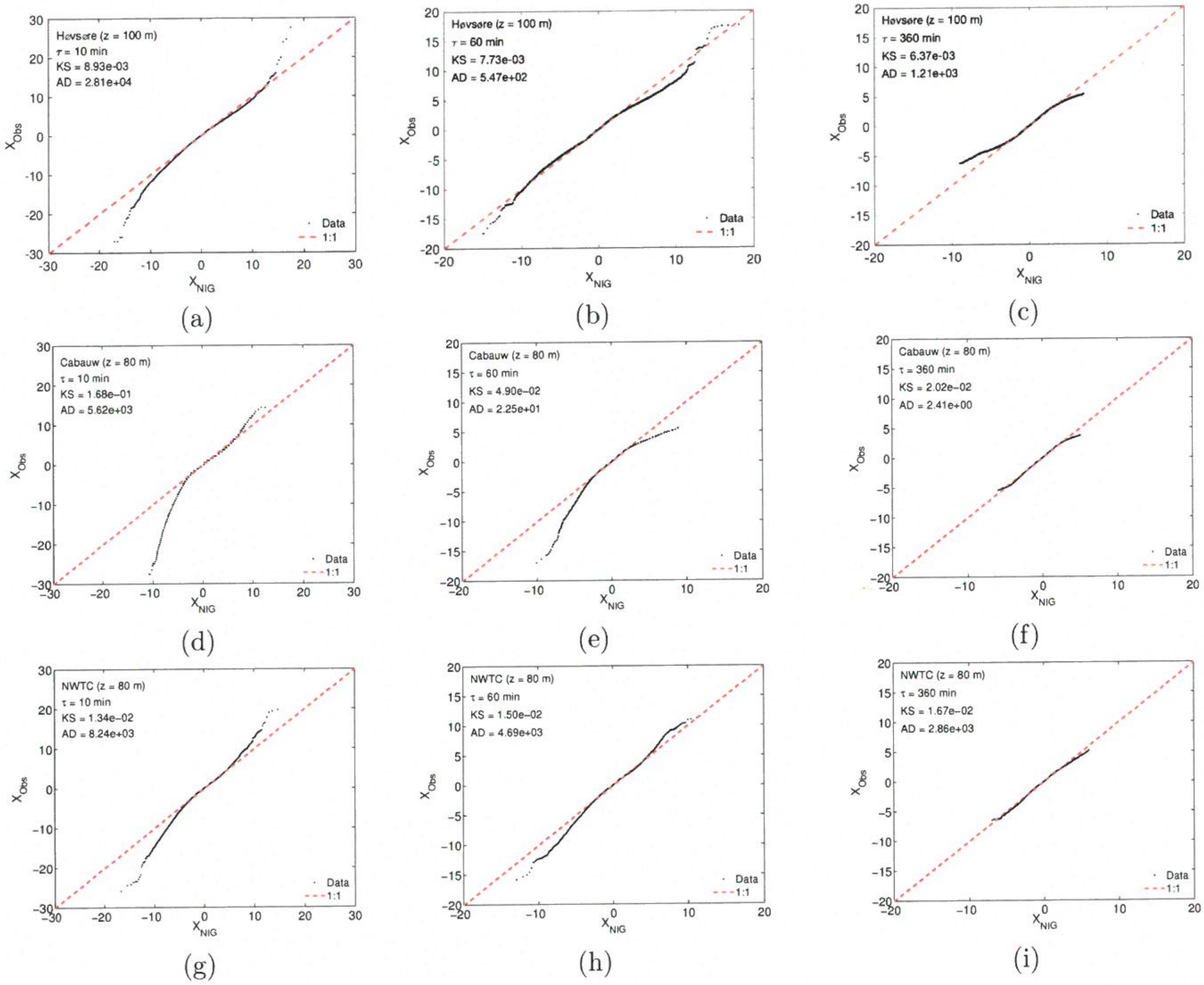
9.3 Quantile-Quantile Plots and Goodness of Fit Test

The Q-Q plots shown in Figures 9.3 and 9.4 reveal the quality of the fits for NIG and GHSST, respectively for three different τ values at 10, 60, and 360 mins. Similar methods are used in Section 8.3. At first glance, comparing the two different models, NIG appears to have an overall more accurate estimation of the temperature increment distributions across all sites, except for Cabauw. At this location, the distribution showed a heavier and more skewed tailed compared to Høvsøre and NWTC and therefore, GHSST outperforms. Furthermore, the NIG model is consistently underestimating the left tail of the distributions across all locations for τ less than 60 min, due to the heavy nature of the left tail. This feature is a result of more extreme negative fluctuations occuring in the temperature increment fields. However, the core of the distribution is well captured by NIG. Thus, due to the very poor prediction with GHSST over NWTC and Høvsøre, one could easily argue that NIG is far more superior that GHSST and could confidently be used for estimating mesoscale temperature increments. To further illustrate the accuracies of the models, as before, we will compute the two-tailed Anderson-Darling test statistic comparing these two models against observations.

Table 9.1 showcases the two-tailed A-D results for the mesoscale temperature increments comparing observations with NIG and GHSST. A few conclusions can be drawn from this table. First, at Cabauw, the A-D results show a much lower statistic compared to the other locations, which can be misleading, given the results obtained from the Q-Q plots. The percision of the temperature measurements at this location are only to two decimal places (e.g., 273.15), which creates many non-unique values when calculating the increments. Since A-D computes its statistic based the size of the edf/cdf (i.e., n in Eq. 8.1) and the edf/cdf is determined from only unique values, the statistic is therefore small in comparison to other locations. With that said, only an inter-site or relative comparison can reasonably be conducted. This issue is not relevant for wind speed given that the percision of the wind speed data and the number of non-unique values are nearly the same for all locations. Next, the left tail for $\tau = 10$ min for all locations show a worse prediction as the values are generally higher for the left tail. Also, the GHSST results are generally higher (least accurate estimation) for NWTC and Høsvøre and lower for Cabauw. Finally, these statistics collaborate, the previous findings, that NIG is a more appropriate model for estimating the temperature increments within the mesoscale range. Therefore, we will use the α parameter for NIG to illustrate the variations in the pdfs as a function of τ and then explore how the diurnal variations impacts these features.

Table 9.1: $A_L^2$and A_R^2 statistics for the mesoscale temperature increments for the four sites at τ = 10 min, 60 min, and 180 min. The results are the median, 5% and 95% height averaged values for the NIG and GHSST models. See Eq. 8.1.

			PDF Model					
			NIG			**GHSST**		
Sites	τ (min)	tail	5%	50%	95%	5%	50%	95%
Høvsøre	10	L	62.18	69.83	77.47	1092.15	1823.04	2553.93
		R	48.43	50.26	52.09	195.56	279.87	364.19
	60	L	21.23	34.80	48.38	235.48	249.35	263.22
		R	43.09	47.52	51.95	402.22	410.21	418.20
	360	L	226.53	381.81	537.09	596.08	783.24	970.41
		R	207.22	417.11	626.99	859.78	1226.31	1592.84
Cabauw	10	L	10.38	719.34	5288.94	1.09	1.29	1.50
		R	1.00	1.28	1.76	1.08	1.33	1.51
	60	L	3.70	6.77	17.21	0.77	0.90	1.30
		R	0.96	1.14	1.27	1.50	1.63	1.71
	360	L	0.53	0.70	0.98	0.86	1.11	1.45
		R	2.57	3.25	3.61	3.85	4.52	5.00
NWTC	10	L	106.15	168.24	170.99	115.17	474.89	563.84
		R	26.87	37.10	108.35	35.94	90.43	114.30
	60	L	65.23	146.97	196.33	137.10	502.72	516.38
		R	60.64	89.14	136.01	70.33	130.53	221.47
	360	L	558.31	637.80	717.30	425.60	496.68	567.76
		R	585.73	613.08	640.43	568.28	596.88	625.47



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 9.3: Q-Q plots comparing temperature increments against the estimated NIG model distributions for three different τ (10 min, 60 min, and 360 min) and three locations (Høvsøre (100 m), Cabauw (80 m), and NWTC (80 m)).

9.4 Day and Night Impacts on Temperature Increments

Before discussing the day and night comparison, Figure 9.5 shows α as a function of τ for the three observational towers from the entire dataset with each height overlaid on the plots. From this perspective, one can clearly see that the height dependence is more evident than what was shown in the wind speed increments (see Figure 8.6). For instance, the α values suggest

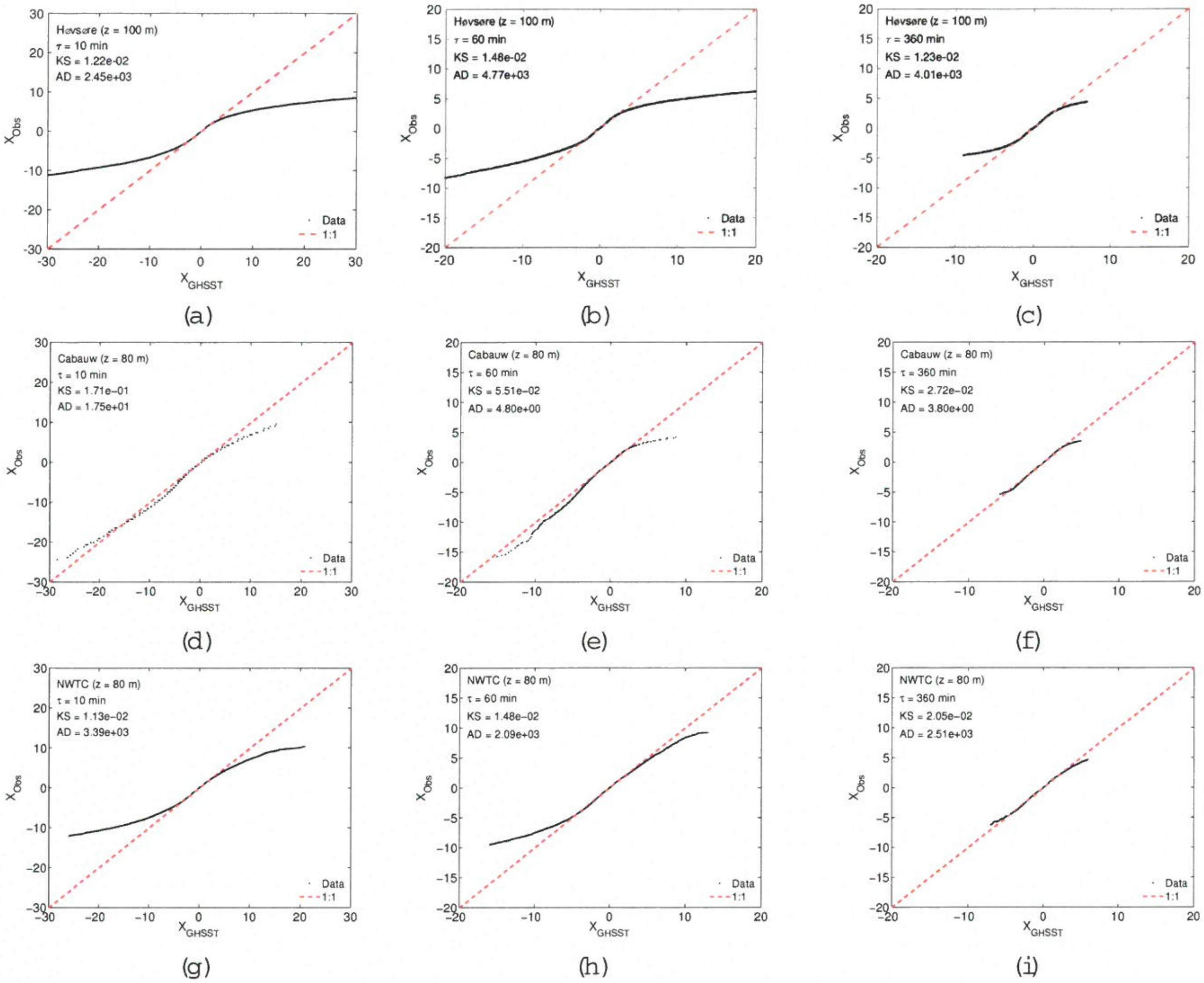
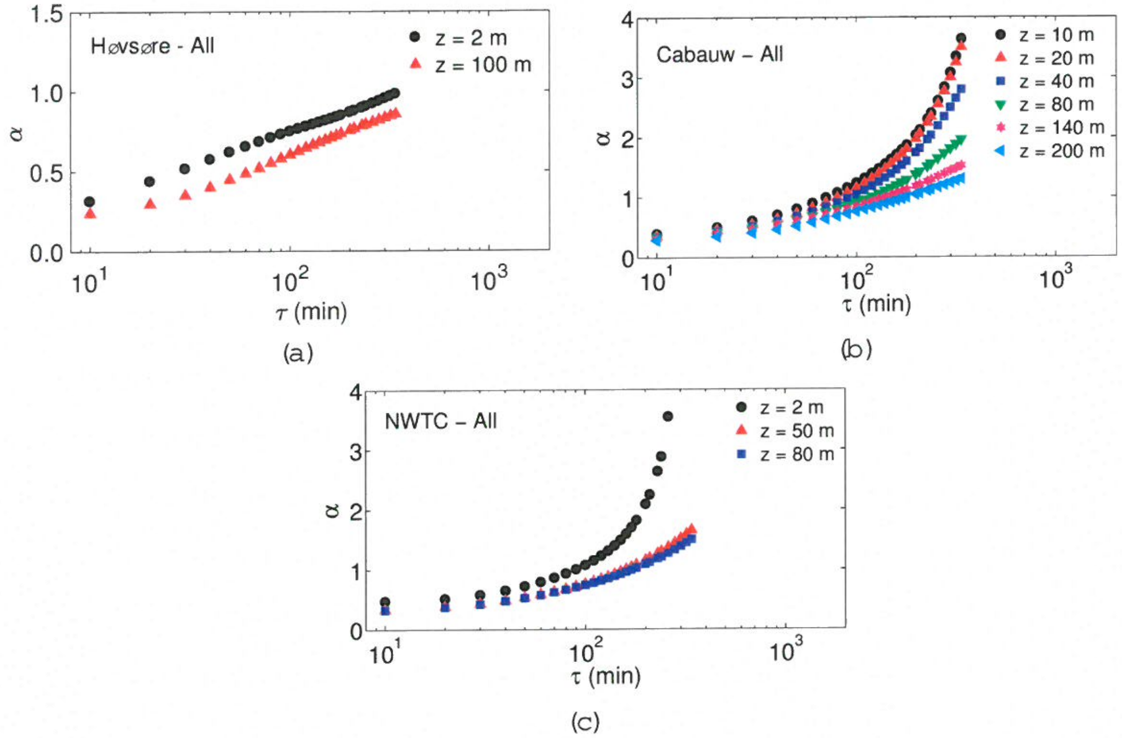


Figure 9.4: Same as Figure 9.3 except the comparison between observed temperature increment and GHSST.

that for higher altitude the temperature increments are slight more intermittent (lower α) for small τ and the intermittency difference becomes more enhanced as separation increases, especially for Cabauw and NWTC (more complex terrain). The intermittency is more pronounced at higher altitudes and larger τ, suggesting that variations in the larger scale increments is greater at higher heights as opposed to near-surface where the distribution approach Gaussian. For the near coastal site, Høvsøre, shows a very consistent, monotonically increasing trend with increasing separation demonstrating a quasi-universal near-surface boundary layer for marine environments. This behavior is not surprising since the diurnal variations are minuscule near ocean surface, however, this parameter-based representation is unique given that we have proposed an underlying distribution to describe these features.



(a)

(b)

(c)

Figure 9.5: NIG parameter (α) for full series of temperature increments as a function of τ for a) Høvsøre, b) Cabauw, and c) NWTC for different heights.

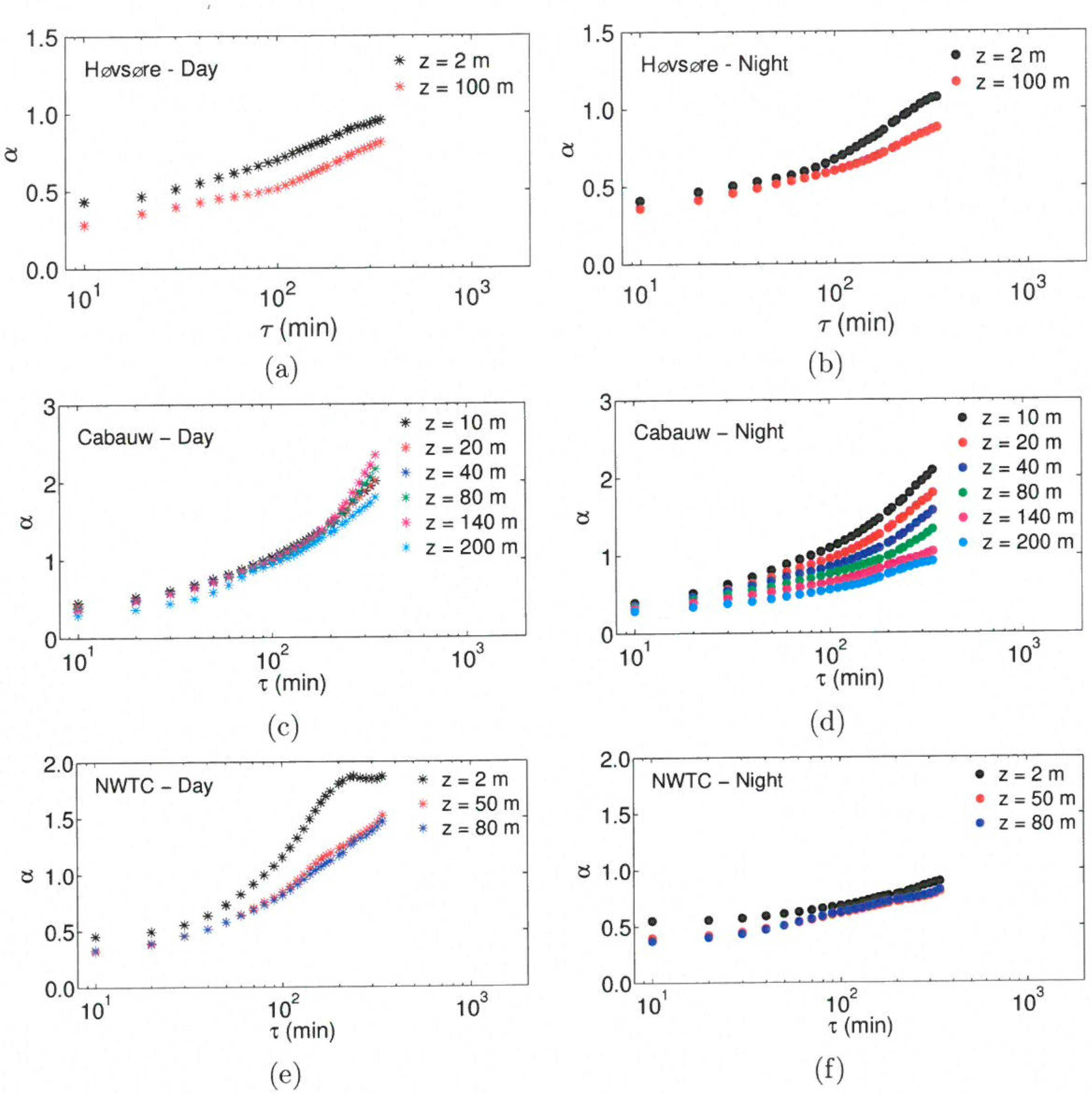
Next, the temperature data was separated into day and night series in accordance with the same procedures used in Section 8.5. Figure 9.6 reveals interesting differences between day and night. In particular, Figures 9.6c and 9.6d (Cabauw) show a distinct difference between night and day, such that the α values, as a function of τ, change uniformly with respect to height suggesting a well mixed environment up to 200 m. However, during the night time period there are clear differences between the heights, where the lower levels show a more Gaussian behavior (larger α) at larger scales while more intermittency (lower α) is evident at higher heights. This intriguing finding indicates that as stability increases near the surface, during the night, the effects are felt beyond 30 m in temporal scale and more intermittency (lower α) is observed at higher altitudes at larger scales. Interestingly, yet not surprisingly, the full series shown in Figure 9.5b depicts a hybrid of the day and night plots.

As for Figures 9.6a and 9.6b, at Høvsøre, the change in α with increased separation (τ) is not as evident. However, there appears to be a minor change with respect to height at larger scales, but not as pronounced as Cabauw. The day time results is nearly identical to the full series result further indicating the marine influence on the temperature increment series.

Finally, the complex terrain (i.e., NWTC) figures show the α values during night time increase slowly with increasing τ with no obvious height dependence as shown over Cabauw. However, this slow increase in α suggests that the temperature fluctuations remain relatively intermittent compared to the day time in which the distribution quickly progress to a more Gaussian distribution with increase separation. The 2 m curve during the day shows a strong difference perhaps indicating that the very near surface is characterized differently than 50 and 80 m heights. These results are remarkable given that physical evidence, and differences between day and night conditions can be observed and explained through the examination of a single NIG model parameter. More work is needed to further explore other physical processes, however, we can have confidence that the NIG model is a suitable tool to estimate the distribution of the mesoscale temperature increments.

9.5 Conclusion

In this work, we examined the mesoscale temperature increments and found that there are indeed quasi-universal features that exist over a long term period of measurements. The tails of the corresponding temperature increment pdfs do portary remarkable similiarities to the small scale countparts. In fact, the temperature increment in the mesoscale range exhibits heavier tailed and more skewed distributions compared to small-scale temperature increments.



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

Figure 9.6: NIG parameter (α) for day (left column) and night (right column) series of temperature increments as a function of τ for a)-b) Høvsøre, c)-d) Cabauw, and e)-f) NWTC for different heights.

A similiar conclusion was found for the comparison between the mesoscale wind speed and small-scale velocity suggesting that in general mesoscale motions, ~10 min or greater, have a larger degree of intermittency than the small scale. Then we showcased the ability of each model pdf and found that, like wind speed, NIG and GHSST were the most reliably through statistical tests and evaluation. Finally, using NIG α parameter as a proxy, we highlighted the variability in the shape (specifically steepness) of the distribution between night and day over a range of separation. The results suggest that the distribution of the temperature increments are in fact modulated by the diurnal variations and are strongly dependent on height over land, but are uniform over ocean. Future work will be required to thoroughly evaluate the larger scale impacts on optical wave propagation. However, providing a model capable of predicting the features of the temperature increment field in the mesoscale range is an sufficient tool for further analysis. In the next chapter, we will explore the Weather Research and Forecasting (WRF) model to determine how well it is able to predict the wind speed and temperature increments over the course of a one year model simulated period.

Chapter 10

Weather Research and Forecasting Model Comparison Study of Wind Speed and Temperature Probability Density Functions

10.1 Introduction

Numerical Weather Prediction (NWP) models are used for many applications, for instance forecasters rely on models to provide them with insight on how the weather will evolve over time. Also, researchers use NWP models to explore past events to understand the physical nature of complex atmospheric environments. For instance, the Air Force has become interested in leveraging mesoscale models to develop forecast tools for optical turbulence (C_n^2) forecasts for various ground and airborne based laser communication systems [7]. These forecast rely on accurate estimation of small-scale turbulent processes. However, due to the complexity of the atmosphere and the inability for these models to resolve motions (e.g., turbulence) which occur on the smallest scales, a general misrepresentation of these processes occur [93].

To illustrate variations in the model results for different model configurations, Figure 10.1a shows the wind speed increments for a range of τ (temporal separation) from the New Zealand Meteorological Service synethically generated wind series [255] over various wind farm locations throughout the country. This data was created from the Fifth-Generation Penn State/National Center for Atmospheric Research (NCAR) Mesoscale Model (MM5) at 12-km horizontal resolution and 10 min sampling over a five year period to aid in the prediction in wind fluctuations (i.e., extreme winds) for wind energy purposes. Figure 10.1b shows the results from a simulated wind dataset generated by the Weather Research and Forecasting (WRF) model run using Yonsei University (YSU) PBL scheme on a 2-km grid over the continental United States at a five min resolution over a seven year period (2007-2013). This dataset was also created for wind energy applications [81].

Clearly, these models are quiet different even though they are both attempting to represent the boundary layer wind speed characteristics. The NIWA dataset does not capture the large fluctuations as evident by the quasi-Gaussian distribution shape over a range of temporal separation in Figure 10.1a, extremely underestimating the probability of occurrence of rare/large fluctuating events. In contrast, in Figure 10.1b the WTK results appear to provide a better representation, of the mesoscale wind speed increment characteristics as shown in the previous chapters. However, the smaller scale (τ= 10 min) fluctuations (e.g., the tails of the distribution) are much more pronounced, indicating an overestimation in the extreme events [1]. These drastic differences are alarming. Thus, it is essential to ensure the we understand the current state-of-the-art models' ability to properly characterize scale-dependent traits of the bounday layer processes.

The objective of this study is to evaluate the performance of seven different PBL parameterizations within the WRF model (V3.6.1) and validate their performance based on wind speed and temperature observational data from a one-year period over three simulated tower sites. These tower sites have already been discussed at length in the previous chapters, they are

---

[1] The x-axis scale extends to ±20, which is well beyond our previous wind speed increments results.

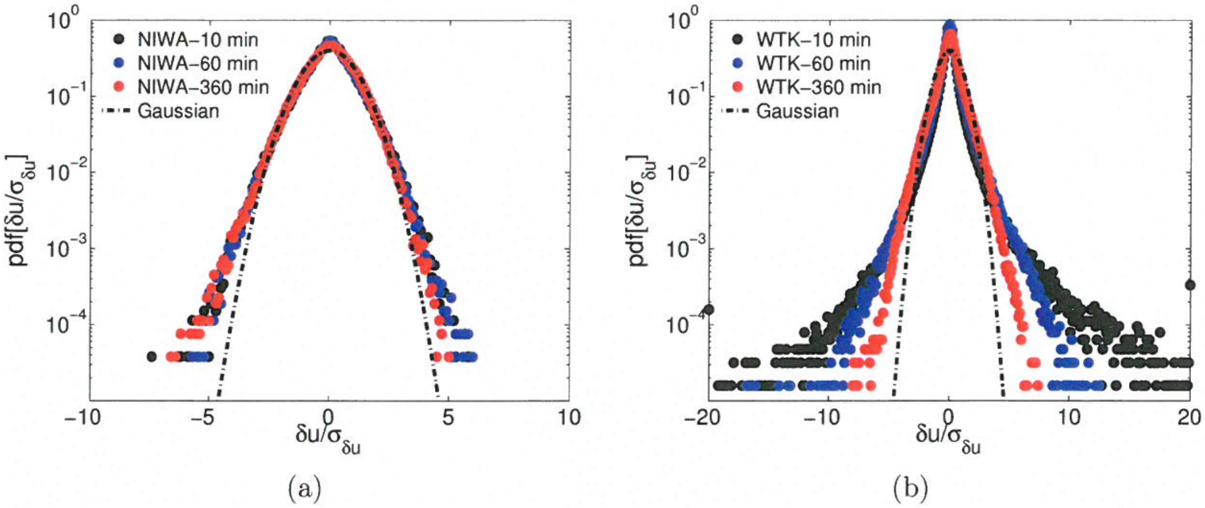


Figure 10.1: Examples of mesoscale model-based synthetic wind speed increments from a) The National Institute of Water and Atmospheric Research (NIWA) in New Zealand [255] and b) The National Renewable Energy Laboratory's (NREL) Wind Integration National Dataset (WIND) Toolkit (WTK) over the continental United States [81].

FINO-1, Høvsøre and Cabauw towers. Using the increment pdf method for both wind speed and temperature, an evaluation of the state-of-the-art models will be conducted. First, an introduction and description of the model set-up/configuration and the specifics of the PBL schemes will be discussed. Then, we will show how the WRF model simulations perform against observation using the pdfs, for a visual representation, as well as goodness of fit metrics, such as two sample Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D). Next, we will highlight the NIG-based estimate and corresponding fit for day and night conditions. Finally, we will briefly discuss the potential for using this approach for developing a benchmark for model development and validation.

10.2 Planetary Boundary Layer Parameterizations

Within the atmospheric planetary boundary layer (PBL) processes occur on a wide range of spatio-temporal scales. For modeling purposes, parameterizations are required to account for the sub-grid scale (SGS) motions (e.g., turbulence) for which the model is unable to resolve.

Over the years research has been conducted to determine suitable approximations for these motions by using similarity theories that have been hypothesized by [20, 76, 138, 179, 281]. Over time, these theories have been tested based on observational and laboratory studies and have been applied to NWP models leading to the generation of numerous turbulence schemes. There are two main catagories of parameterization which are imposed in the model, local and non-local. Local schemes, as the name suggests, only consider neighboring vertical levels in the model for their calculations, whereas non-local schemes communicate with mulitple levels to represent the impacts of vertical mixing within the PBL [274]. The local schemes are believed to restrict the depth of vertical mixing associate with larger turbulent eddies, while non-local schemes have been shown to depict a deeper PBL during convection and strong nighttime winds [69].

Another factor that is addressed in the development of PBL schemes is the turbulent closure problem (i.e., more unknown variables than available equations), which uses Reynolds–averaged Navier–Stokes equations. To account for this limitation, some level of approximation is required either first- or second-order which parameterizes either the covariance or triple-covariance term, respectively. For first-order, the flux-gradient (K) theory approach is used, (see Eq. 10.1).

$$\overline{u'w'} = -K_m \; \frac{\partial \overline{u}}{\partial z}^{\,!} \; ; \qquad \overline{v'w'} = -K_m \; \frac{\partial \overline{v}}{\partial z}^{\,!} \; ; \qquad \overline{\theta' w'} = -K_h \; \frac{\partial \overline{\theta}}{\partial z}^{\,!} \; . \tag{10.1}$$

Higher order schemes have been developed invoking explicted calculations of the prognostic turbulent kinetic energy equation (TKE, see Eq 10.2). In some instances, there is a non-integer order scheme. For example, 1.5-order closure schemes estimate second-order TKE by computing second-order moments of some variables (e.g. potential temperature covariance) and first-order for others [69, 246].

$$\frac{\partial e}{\partial t} = -\frac{1}{\rho}\frac{\partial \overline{\rho w' e'}}{\partial Z} - \overline{u' w'}\frac{\partial U}{\partial Z} - \overline{v' w'}\frac{\partial V}{\partial Z} + \beta(\overline{w'\theta'}) - \epsilon. \tag{10.2}$$

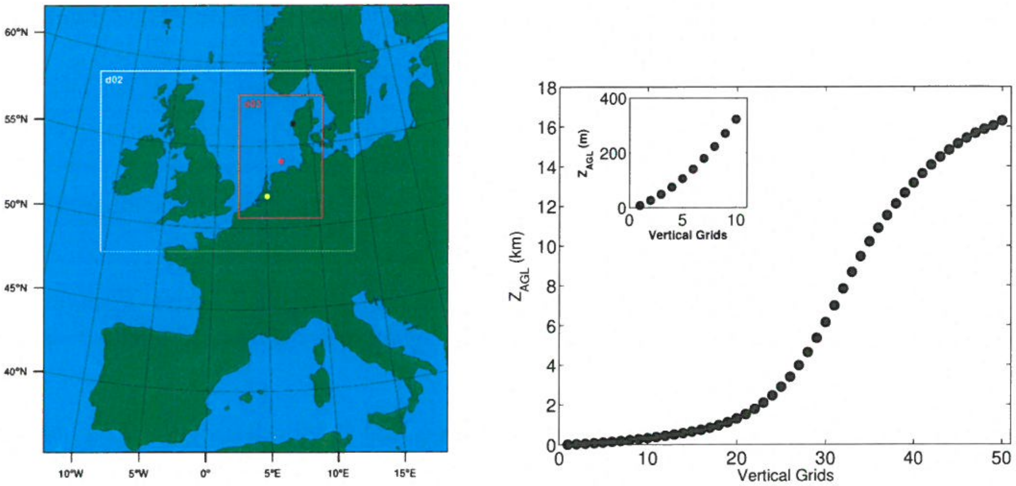
It has been shown that the PBL parameterizations have the greatest impact on the modeling of surface variables. Thus, for this work, we have chosen seven different PBL schemes to evaluate which are briefly described here:

- Mellor-Yamada-Janjić (MYJ) [120, 121, 173] - 1.5 order TKE closure scheme w/local mixing, invokes prognostic equations for TKE.
- Mellor-Yamada-Nakanishi-Niino level 2 (MYNN2) [187]- 1.5 order TKE closure scheme (similar to MYJ) introduces three mixing lengths for buoyancy, surface and turbulence layers.
- Mellor-Yamada-Nakanishi-Niino level 3 (MYNN3) [188] - 2.0 order TKE closure scheme (similar to MYNN2), with higher order moment computation.
- Quasi norma scale eliminationl (QNSE) [247] - 1.5 order local TKE closure and provides improved potential temperature profiles.
- Total energy-mass flux (TEMF) [11, 171, 227] - 1.5 order TKE closure with both a local and non-local component activated for stable and unstable conditions, respectively.
- University of Washington (UW) moist turbulence [50] - 1.5 order local TKE closure and performs well in night-time stable boundary layers.
- Yonsei University (YSU) [113] - Non-local K closure (First-order) scheme. Determines height of the PBL better by explicit treatment of the entrainment processes.

Associated with each of these schemes are surface layer schemes, based on Monin-Obukhov similiarity theory, (see Eq. 6.2, as an example). Therefore, the following surface layer-based scheme were incorporated with the various PBL parameterizations: Eta similiarity with MYJ and UW; MYNN surface layer with MYNN2 and MYNN3; MM5 similarity with YSU; QNSE similiarity with QNSE; and TEMF similiarity with TEMF. In addition to these PBL-based schemes, the following section will discuss the other key aspects of the WRF model set-up.

10.3 Model Configuration

The WRF model is a fully compressible non-hydrostatic mesoscale NWP model that is used for both operational and research purposes. The models infrastructure is extensive and incorporates numerous phases, which is described in more detail in the WRF technical notes by Skamarock



(a) (b)

Figure 10.2: a) WRF domain configuration. The circles represent the approximate locations of the simulated towers at Høvsøre (●), FINO-1 (●), and Cabauw (●). b) Model's vertical grid resolution versus height in AGL (km). The inset depicts the 10 lowest vertical grid points which are considered in this study.

et al. (2004 and 2008, [228, 229]). To evaluate the statistical properties of PBL wind speed and temperature increments, a one-year period (2012) simulation for all parameterizations were generated over the North Sea. For all the model simulations, the WRF pre-processing system (WPS) was used to construct 3-one way nested domains (d01-d03). The innermost domain (d03) consists of a 3 x 3 km horizontal resolution with a domain size of 552 x 813 km. The middle domain (d02) had a grid spacing of 9 km, while the outer domain had 27 x 27 km horizontal resolution. The vertical resolution for the three domains contained 51 eta levels from the surface up to 16 km AGL with at least 23 levels below 2 km to better resolve the PBL features (see Figure 10.2). The center of the domain was strategically placed to encompass the three tower locations with enough separation and influence from the sides of the domain. The simulated tower data was generated by utilizing the *tslist* option on WRF and was placed at the nearest grid point associated with the observed tower latitude and longitude dimensions. For the two intermost domains, U and V components of the wind and temperature were output every time step (45 and 15 secs) for d02 and d03, respectively for each tower level up to 16 km AGL. A

10-min moving average and downsampling was performed for this increment analysis in order to directly compare with the 10 min sampling rate of the observational data, similar approach was used for NWTC in Chapter 8.

The U.S. Geological Survey (USGS) data was utilized for the boundary conditions (i.e., land-use and topography) at 2 resolution for the middle domain (d02). However, for the innermost domain (d03) the newly implement GMTED2010 30-sec high-resolution data was incorporated [197] to test if any effects are seen by changing the topography. Furthermore, the European Centre for Medium-Range Weather Forecasts ERA-Interim archive data was used for the initial and boundary conditions and was nudged in the simulations every six hours during the one year period using the WRF Four-Dimensional Data Assimilation (FDDA) method. Analysis nudging of the temperature, wind, and water vapor mixing ratio are applied in both domains with no nudging in the PBL. No feedback from inner domain to outer domain is applied to avoid information exchange from nests and parent domains, since our intent is to test the specific behavior of the domains and not have the averaged fields from the child domain fed back into the parent. For this study, we were interested in understanding the impacts between the domains and did not want the influences from higher to lower resolution. The physics packages for the simulations consisted of the following parameterizations: Microphysics-WRF single-moment 5-class scheme (WSM-5); shortwave and longwave radiation scheme, Rapid Radiative Transfer Model for Global Climate Models (RRTMG); Cumulus-Kain Fritsch cumulus scheme; and land surface-Noah scheme. The cumulus scheme was only applied for d01-d02 where the grid resolution was greater than 5 km. For the finer domain (d03 at 3 km resolution), explicit microphysics accounts for all the cloud processes [242]. Moreover, the land surface model (LSM) is comprised of soil temperature and moisture in four layers, which attempts to account for the sub-grid scale interactions between the land-atmosphere interactions. The LSM plays an important role with more complex PBL scheme, so further work is necessary to test the sensitivities between various LSM schemes.

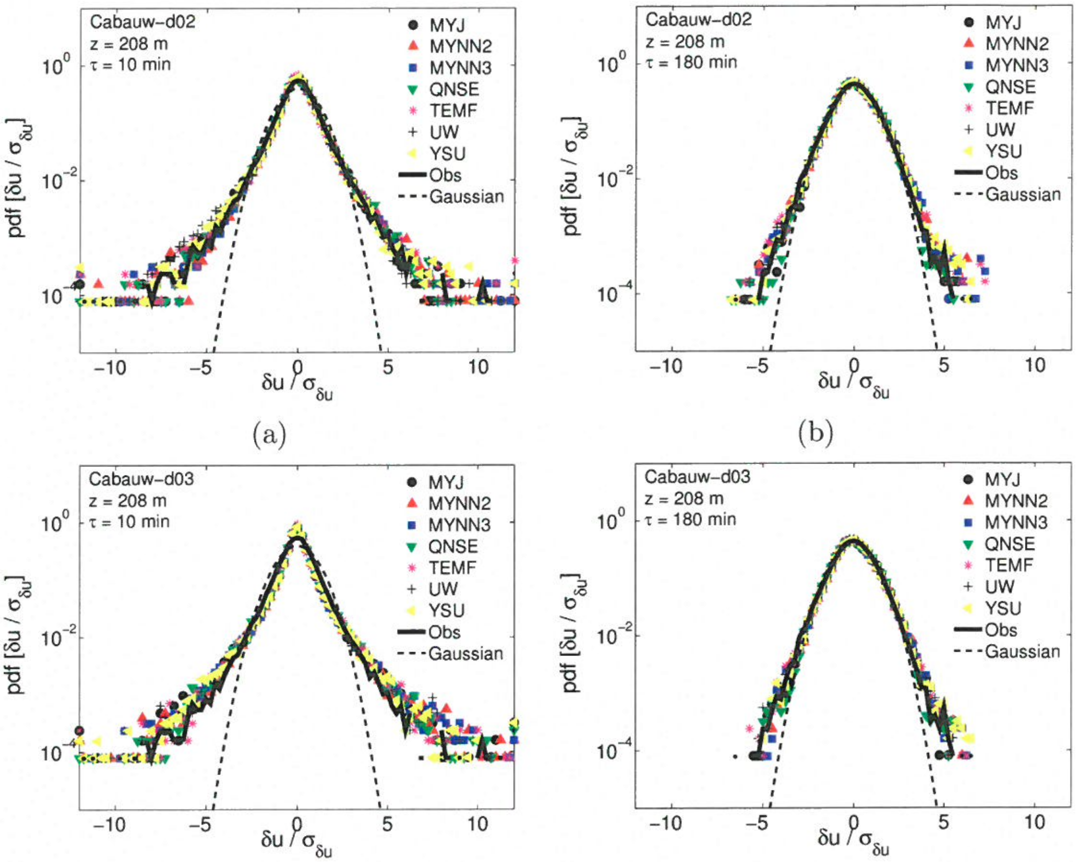
10.4 Wind Speed and Temperature Probability Density Functions from WRF

For an illustrative example, Figures 10.3 and 10.4 depict the wind speed and temperature increments for τ at 10 min (left column) and 180 min (right column) for the Cabauw tower location at ~200 m AGL. Also, included in these figures are the differences between the 9 km resolution domain (d02, top row) and the 3 km resolution domain (d03, bottom row). Overlaid on each figure are the seven simulations with the different PBL schemes (colored symbols) and the corresponding observed wind speed and temperature increments at Cabauw ($z = 200$ m). From a visual perspective, all simulations are lined up with the observations remarkably well. However, there are distinct differences which require further evaluation. For instance, in Figure 10.3, the tails of the distributions for the model simulations, at $\tau =$ 10 min, are slightly heavier than the observations, while each model shows a slight variation between one another. However, as separation increases to 180 min the model and observation appear to be more in line.

Furthermore, the higher resolution simulations (d03) are indicating an even larger difference in the tail features as well as near the core of the distribution. Figures 10.11c and d gives an indictation how these features vary between times of day suggesting that the deviations between observations and models are more pronounced during the day time. Even after the moving averaging operation, this distinct feature over a higher degree of intermittency is present. It was recently shown that increased resolution does not necessarily improve results [108, 279]. Moreover, Frehlich and Sharman (2008, [93]) discussed the effective model resolution based on structure function (SF) and spectra analysis. They remarked, "As NWP models move to higher resolutions, the smaller scale features, such as mountain waves and convectively induced gravity waves, will introduce modifications to the model-derived average spatial statistics.." Thus, increasing the resolution may not necessarily result in improved wind speed increment statistics. Additionally, it was shown by Larsén *et al.* (2012, [149]), using model simulations from a range of resolutions (from 50 to 15 km), that the simulated winds from mesoscale models remove the variations because of spatial and temporal averaging effect by revealing a steeper spectal slope (-3) compared to observations (-5/3, equivalently 2/3 in structure function scaling) at higher wavenumbers (> 2 day^{-1} or equivalent temporal scale, $\tau < 12$ hr). Given our results, suggesting larger fluctuations with higher resolutions, and the previous studies, it is clear that further improvement on sub-grid scale processes are required. To address the subtle difference between the PBL schemes, in the next section, we will provide a statistical evaluation of pdf fits, but first we will discuss the variations in the temperature increment field.

Figure 10.4 reveals the temperature increment pdfs from the one-year model simulations and observations from the same tower location/height. Similar to the wind speed increment pdfs, these pdfs closely resemble the corresponding observed temperature increments over Cabauw at a tower height of 200 m. However, unlike with wind speed, the higher resolution model results are more closely in line with the observations, except for TEMF. Also, with increasing temporal separation the models continue to closely mirror the observations. This suggests that regardless of the PBL scheme the mesoscale models can accurately predict the temperature increments well. However, as shown previously in Chapters 8 and 9, the visual inspection is only a part of the story, to fully grasp the fit of the pdfs and a statistical evaluation is required to provide discriminatory power.

Additionally, in these chapters, a much longer period (10+ years) of time was evaluated as opposed to the 1-year period utilized in this modeling studying. Therefore, to illustrate that 1-year WRF data is sufficient for this type of increment analysis, Figure 10.5 shows the comparison between the 1-year Cabauw observed wind speed increment data (left panel) and the 14-year Cabauw observed data (right panel). The 1-year simulated model data is overlaid same as the previous two figures. From this visual perspective, it is clear to see that the majority of the distributions have a similar appearance, however, the tails of the distribution for the 14-year period extends beyond the model data merely due to the fact that more data is available. Also, from a statistical perspective, these two produce nearly the same results. For instance,



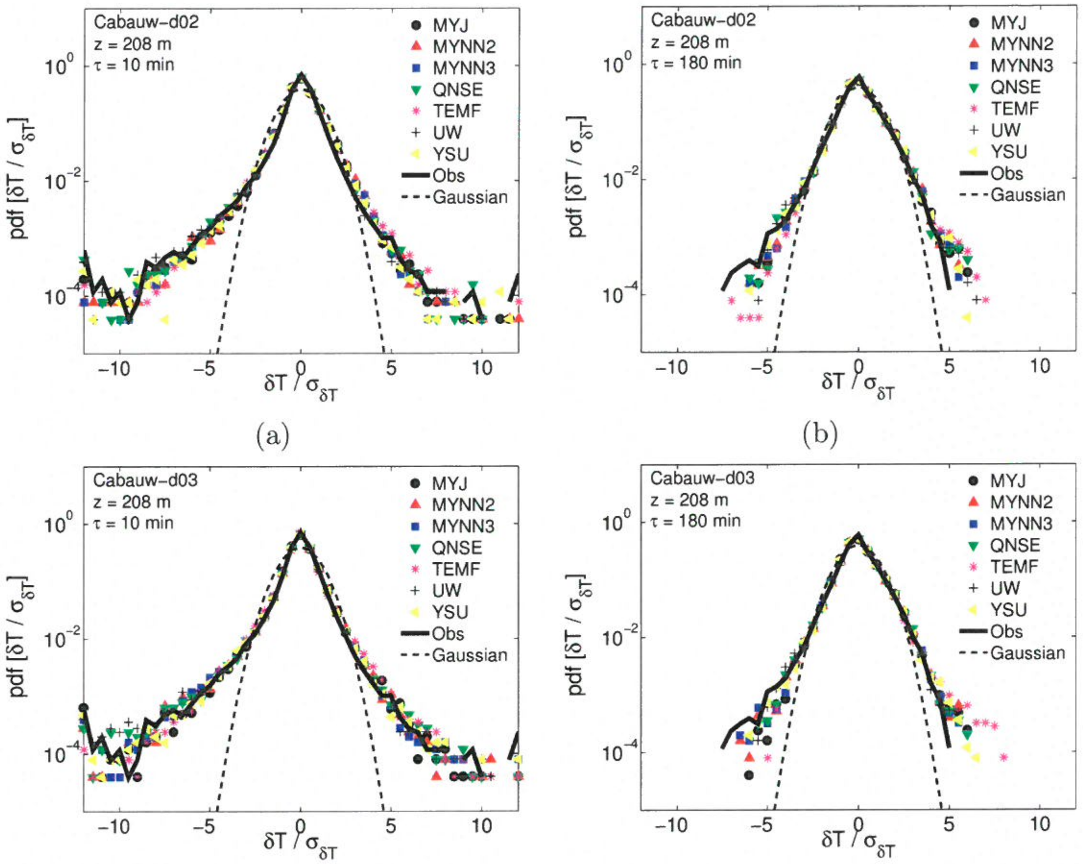
(a) (b)

(c) (d)

Figure 10.3: Pdf of wind speed increments comparing observations and WRF results for seven PBL schemes. (a) and (b) are wind speed increments for $\tau = 10$ and 180 min from d02 domain and (c) and (d) are same except for d03. The series are normalized by their respective standard deviation of their increments (e.g., $\sigma_{\delta u}$).

the K-S test statistic for the Cabauw 1-year data comparing the TEMF scheme (left panel) is 0.0776. Similiarly, for the 14-year data the comparison between 1-year model simulation using the TEMF scheme and 14-year observational data produces a 0.0776 K-S test statistic. This example provides evidence that generating a 1-year simulation is sufficient for evaluating mesoscale features within NWP.

Another factor with the evaluation of pdfs is the choice of bin selection. This issue was briefly addressed in Section 5.5 using small-scale turbulence data to test the sensitivity of the

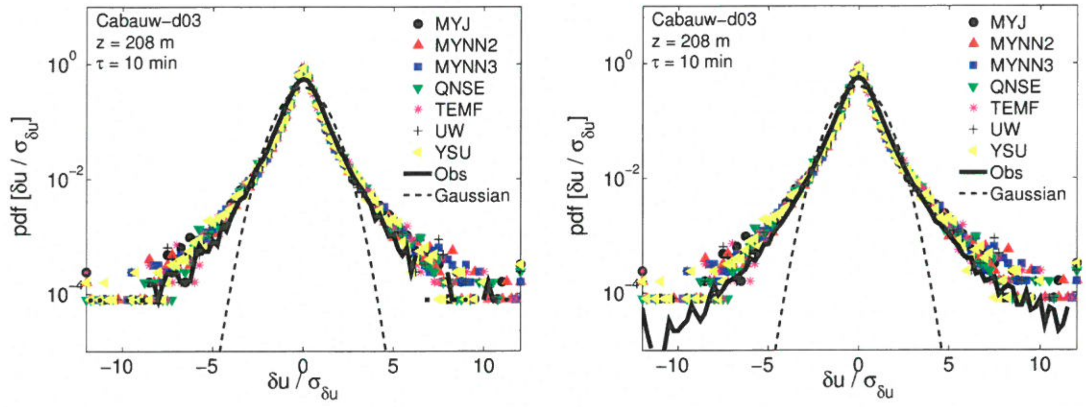


(a) (b)

(c) (d)

Figure 10.4: Same as Figure 10.3 except for pdf of temperature increments

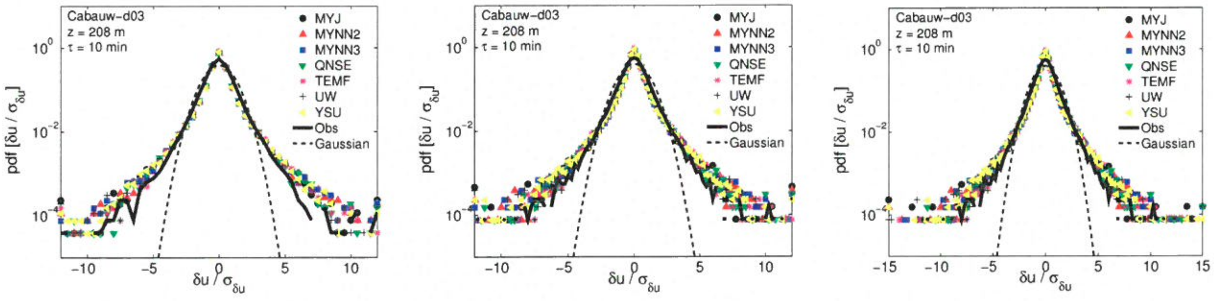
GoF test statistics. In the case with the pdfs displayed above (specifically on Figure 10.3c) with MYJ), it is observed that the tails of the pdfs are not monotonically decreasing, such that the probability of extreme events predicted by the models is higher than the smaller $\delta u/\sigma_{u}$ values. To test ths cause of this artifact, Figure 10.6 shows three different bin selection based on the bin width and maximum/minimum range. When increasing the bin with from 0.25 in Figure 10.3 to 0.5 in Figure 10.6a) the upward trend near the tails, especially for MYJ, is still evident. Similarily, as you decrease the bin width to 0.1 in Figure 10.6b) the trend continues. However, when the maximum/minimum range increases the upward trend begins to erase. Therefore, when displaying and comparing pdfs, it is important to choose a proper bin set-up otherwise



(a) (b)

Figure 10.5: Wind speed increment pdf at τ = 10 mins, a) Cabauw 1-year observations and 1-year d03 modeling results for the simulated tower identical to Figure 10.3c), b) Cabauw 14-year observations and 1-year d03 modeling results for the simulated tower. These two figures produce nearly identical K-S results between the two comparing the PBL schemes and the observations.

results can vary leading to improper characterization and evaluation. The next section will also showcase the variations in results for the other two simulated towers, FINO1 and Høvsøre.



(a) (b) (c)

Figure 10.6: Different bin-widths to showcase the sensitivity to the choice of pdf binning. The following is the dynamic range (maximum and minimum) and bin-widths for the figures above: a) ±12 and 0.5 width, b) ±12 and 0.1, c) ±15 and 0.25.

10.5 Statistical Evaluation of the Probability Density Functions

To examine the variations between the different PBL schemes, two 2-sample goodness of fit tests will be conducted using Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D). Similar to the previous K-S and A-D tests, the 2-sample version utilizes the empirical distribution functions (edf) of the sample. In this case, the comparisons are between the observed increment distributions and the distributions from the seven PBL simulations for the one-year period over the three tower locations. Eq. 10.3 shows the K-S (D_{nm}) test statistic calculation,

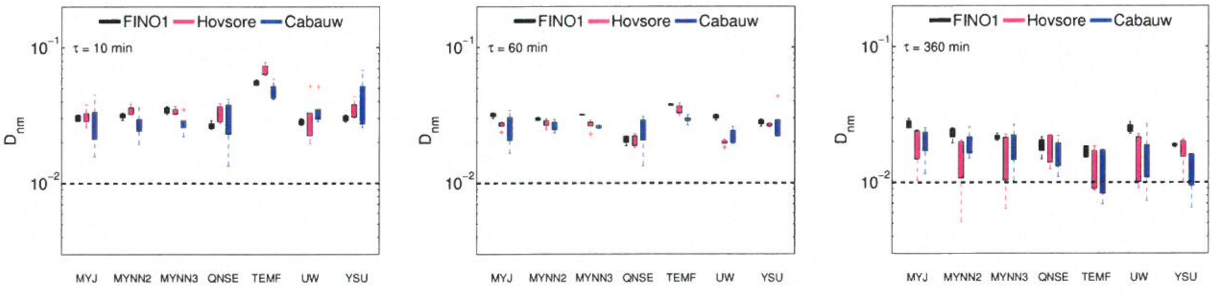
$$D_{nm} = \sup_{x} |F_n(x) - F_m(x)|, \tag{10.3}$$

where F_m represents the different model simulations and F_n is the observed data. In addition to the K-S statistics, a critical value can be determined, for given a confidence interval (α), for which one can either accept or reject the hypothesis that the two distributions are the same, see Eq. 10.4.

$$D_\alpha = c_\alpha \sqrt{\frac{n+m}{nm}} \tag{10.4}$$

where D_α is the Kolmogorov-Smirnov critical value, n and m are the sample sizes for the observed and model data, respectively, and c_α is 1.63 based on an α value of 0.01. For this set of data, D_α is ~0.01. Therefore, the hypothesis is accepted if the calculated D_{nm} value is less than 0.01 and vice-versa. Figure 10.7 through Figure 10.10 show box plots of the various PBL schemes for each location (FINO1, Høvsøre, Cabauw), temporal separations τ (10, 60, and 360 mins) horizontal resolution (d02 and d03) and variable increment (wind speed and temperature). Within in "box" are the D_{nm} values for each height at the different locations. Also, on each figure is a horizontal dashed line indicating the D_α value.

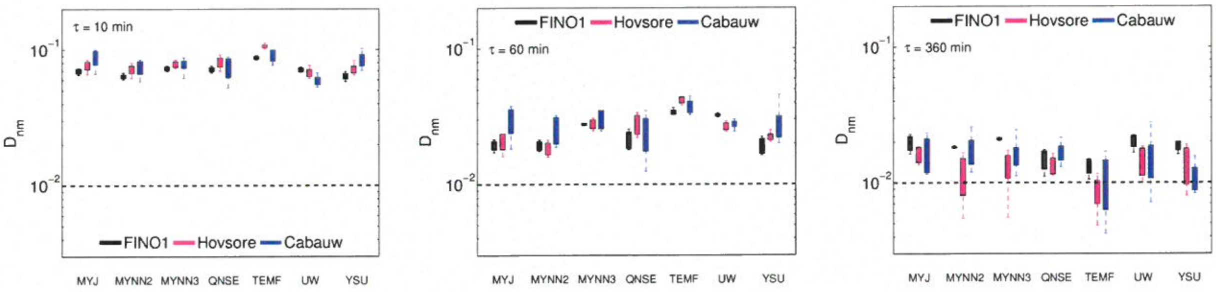
Figure 10.7 is the wind speed increment comparison for the 9 km horizontal resolution domain for 10, 60, and 360 min separations. First, the D_{nm} results for nearly all three plots reveal that the null hypothesis can be rejected, such that the model and observational data are not the same distribution. However, as temporal separation increases the quality of the fit improves. In terms of the various model, there are little differences for each location. However, TEMF scheme performs the worst at small separations, which may be due to the non-local/local aspects of the scheme leading to inconsistency in the long-term statistics. Despite previous studies (e.g., [116]) revealing TEMF's ability to capture the thermodynamic vertical structure of the PBL well, our results indicate that the wind speed increments within this scheme are problematic from a long-term statistical perspective. However, as separation increases TEMF indicates an improvement and slightly outperforms the other models at 360 min increments. The models overall show more variability (i.e., more spread in the box plots) at Cabauw suggesting that in more complex terrain greater differences ensue for the wind speed increments at small scale. Likewise, as separation increases, the variability in the overall results are more diverse suggesting further uncertainty with respect to height at larger increments. Although the variability at Cabauw is greater, the test statistic at small scales across all the models (except UW) reveal the lowest value is achieved at Cabauw. In terms of the best fit between all the models, it is difficult to make a concrete conclusion, however, for 10 to 60 min increments QNSE and MYJ appear to perform slightly better with the two MYNN schemes indicating the least amount of variability (i.e., more consistent prediction throughout lower BL).



(a) (b) (c)

Figure 10.7: Boxplots of D_{nm} for d02 for a) 10 min, b) 60 min, and c) 360 min increments. The set of box plots represent different locations for each of the seven PBL schemes (x-axis).

Figure 10.8 shows the D_{nm} results for wind speed increments as the resolution of the model increases. As it was shown in Figure 10.3c, the model pdfs for small scale separation deviate the most and shows a greater degree of intermittency within the increment series. This finding is confirmed by the Figure 10.8a as the D_{nm} values are nearly half a decade (or a factor of 5) higher than the d02 results for $\tau = 10$ mins. Similiar to d02, the differences in the models are difficult to distinguish, though TEMF is also inferior to the other models. However, as separation increases to 360 min TEMF improves to at or below the 0.01 critical value with the other models closely behind[2]. In fact, the higher resolution results begin to slightly outperform the lower resolution at larger scales suggesting that the impact on the resolution is lost at large scale motions. This is finding is different than what was previously indicated when examining the smaller mesoscale range and the fact that higher resolutions performed worse.



(a) (b) (c)

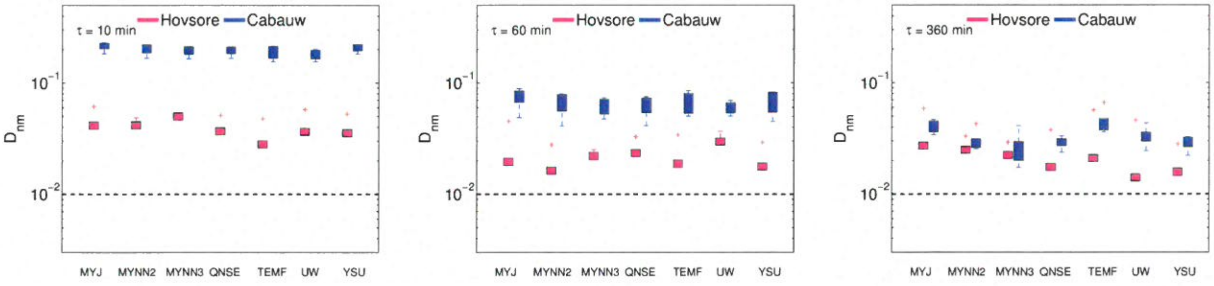
Figure 10.8: Same as Figure 10.7 except D_{nm} results for d03.

Next, we will examine the temperature increment comparison between the two tower locations (Høvsøre and Cabauw). Figures 10.9 and 10.10 are in the same format of the previous two figures, however, there are greater discrepancies between the sites in terms of the temperature. For instance, at small scales the (D_{nm}) results are Cabauw are nearly a factor of 5 higher than Hovsore. There are two possible causes for these differences: (i) the terrain influences are more

---

[2] The model results are largely above the D_α value of 0.01 indicating the models do not fit the observations according to K-S test, however, this statistc has shown to be overly sentitive, due to slight shifts in the distribution [146]. Further evaluation is required to provide an alternative approach.

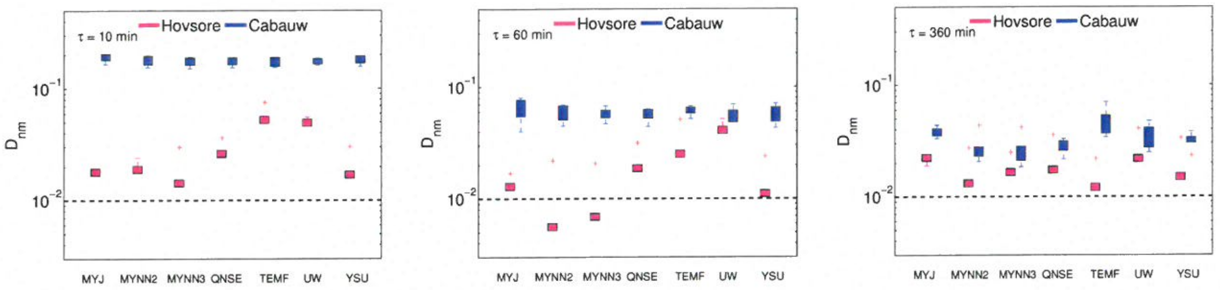
pronounced at Cabauw compared to the Høvsøre resulting in a difficulty for the models to capture the diurnal variations with great accuracy. (ii) As addressed in Section 9.3, the poor decimal percision of the observed temperature measurements at Cabauw leads to problematic statistical results, even though Figure 10.4 suggests otherwise. Though as separation increases the these difference are minimized. In terms of the model to model comparison, again, there are only subtle differences to report. However, for Figure 10.9a, the small scale Temperature resuls show TEMF as the more superior model opposed to the wind speed increments. This results in more in line with [116] in their evaluation of the local and non-local impacts on the thermodynamic structure. Even though, from a visual perspective, the TEMF model shows more deviations from observations in the tails of distribution (see Figure 10.4), the statistics suggest that this difference is not as large of a factor for determining the quality of the fit. Furthermore, for Figure 10.10, there is clear reversal in TEMF with the two MYNN schemes outperforming all the models for all nearly the entire range of separation. This suggests that the higher-order TKE schemes, at higher resolutions, are more capable of predicting the observational temperature increments. To explore further, we will examine the 2-sample A-D statistic A^2_{nm} to determine a more approximation test for the tails of the distributions.



(a) (b) (c)

Figure 10.9: Same as Figure 10.7 except D_{nm} results for temperature increments for only Høvsøre and Cabauw towers.

Similar to the 2-sample K-S test, A^2_{nm} compares two-sample distributions against one an-



(a) (b) (c)

Figure 10.10: Same as Figure 10.9 except the D_{nm} results for d03

other (e.g., observed versus WRF model data). In this computation, we are taking into account the number of "ties" that exist in the data. In other words, the repeated samples within the data are being omitted since this affects the results in a negative way, see Eq. 10.5:

$$A_{nm}^{2} = \sum_{i=1}^{k} \frac{1}{n_i} \sum_{j=1}^{L-1} \frac{l_j (N M_{ij} - n_i B_j)^2}{B_j (N - B_j)}; \tag{10.5}$$

where M_{ij} is the cumulative distribution function (cdf) from each individual sample observation (n) and model (m) with k equaling two for the number for series and n_i being the individual sample sizes. B_j is the combined cdf and N is the total sample size of B_j evaluated for length of the unique (non-repeated) sample, $L - 1$. Finally, l_j represents the pdf of the combined model and observed data series. The results for the A_{nm}^{2} statistic are presented in Table 10.1 for the wind speed increment comparison for the three locations and four different values at d02 domain (this domain will only be shown here). In this Table, the results show the median values from all levels for the given observational tower and respective values. A few conclusions can be drawn from this table.

First, the overall trend of the test statistc suggests that as separation increases the result generally improves. Second, in terms of the specific schemes, it becomes more clear that the TEMF parameterization for small scale wind speed increments is by far underperforming compared to the other models for all locations. As mentioned in Angevine *et al.* (2010, [11]),

Table 10.1: The median two sample Anderson-Darling test (A^2_{nm}) results comparing the various WRF model wind speed increment distributions against the observed pdfs for 2012 for d02 (9 km).

		SITES											
		FINO-1				**Høvsøre**				**Cabauw**			
	τ (min)	10	60	180	360	10	60	180	360	10	60	180	360
PBL Schemes	**MYJ**	21.91	29.72	29.04	20.47	27.22	21.72	10.02	13.63	33.82	24.83	18.83	17.23
	MYNN2	28.32	28.98	24.07	18.41	42.36	23.29	7.09	10.24	30.36	23.56	13.49	11.79
	MYNN3	34.17	32.37	23.17	14.07	48.23	25.14	7.13	10.37	25.06	23.37	11.87	11.92
	QNSE	16.62	10.31	12.23	9.98	28.65	11.57	7.09	12.01	54.06	27.13	10.03	17.35
	TEMF	146.08	51.20	20.17	7.68	226.90	49.86	8.40	4.93	119.87	30.74	10.65	5.90
	UW	17.00	24.46	27.24	17.59	17.30	11.38	5.67	9.45	43.92	19.45	6.55	5.82
	YSU	24.53	22.79	22.58	11.49	30.74	21.00	10.23	12.44	50.06	25.98	5.86	3.90

this hybrid local/non-local scheme tends to generate an excessive amount of moisture flux in the lower cloud bounary layer, which perhaps modifies to the wind speed fluctuation over the North Sea locations, considering shallow stratocumulus clouds are present in this area. But, as separation increases this scheme begins to dominate indicating that this influence is lost within large scale motions. Furthermore, the QNSE and UW appear to be more accurate for the two locations which are near or on an ocean surface (FINO-1 and Høvsøre). It was discussed in Cohen *et al.* (2015, [69]) that UW and QNSE schemes provide a more accurate depiction of stable boundary layers, which is a characteristic of maritime environments. The two MYNN schemes perform worse over the water but are the most accurate, especially at small scales, over land. From these statistcs, a few conclusions can be made based on the performances of the different PBL parameterizations. Next, we will examine the temperature increments using the A-D test statistic.

In Table 10.2 depicts the median A^2_{nm} results for the mesoscale temperature increments series over two tower locations comparing PBL schemes for the d02 domain against observations. Similar to the results shown in Figure 10.9, the accuracy at Cabauw is problematic, but it greatly improves with separation. Again, the model to model comparison does not provide a significant demarcation, however, at τ=10 min for Høvsøre it is clear that TEMF is more accurate on average over all the levels with a value of 58.28. While, throughout all separations,

Table 10.2: Same as Table 10.1 except A^2_{nm} for temperature increments

		SITES							
		Høvsøre				**Cabauw**			
	τ (min)	10	60	180	360	10	60	180	360
PBL Schemes	**MYJ**	114.82	37.68	44.74	97.34	758.68	81.94	29.35	61.43
	MYNN2	71.15	13.05	15.34	43.16	726.76	57.25	13.79	15.29
	MYNN3	82.10	13.37	13.80	35.46	742.80	50.89	12.62	8.63
	QNSE	76.98	21.36	18.62	43.55	746.55	55.26	14.29	23.71
	TEMF	58.28	22.20	48.37	98.76	782.13	70.87	38.57	67.42
	UW	84.64	30.87	31.01	59.76	791.91	49.07	11.83	33.43
	YSU	77.34	15.15	9.97	29.32	720.78	50.71	12.37	14.62

YSU performs slightly better than the two MYNN schemes, while MYJ performs the worse. Over Cabauw, the best predictions are from YSU. YSU is a non-local scheme which has pros and cons relative to the time of day or year, however over a long-term period it is shown to handle the temperature increment data over land the best.

The K-S and A-D 2-sample statistical evaluations of the long-term wind speed and temperature increments between observations and various PBL schemes were able to provide some level of quantative assessment. Even though from a visual perspective, the various schemes appeared to be very similar, it was shown that there were difference and could be related to previous findings from the various PBL schemes. TEMF showed to be inferior at small scales for wind speed increments, while QNSE and MYJ outperforms the other models overall. However, from a statistcal perspective, for temperature increments TEMF is better at small scales and YSU is slightly outperforms over land. In the next section, we will use the NIG distribution estimation to further showcase model and observed data comparison between day and night time conditions.

10.6 Diurnal Variations in Wind Speed and Temperature Increments in WRF

As discussed in Chapters 8 and 9, diurnal variations occur within the wind speed and temperature increment fields. Also, it is well-known that different PBL schemes have diverse solutions under various stability conditions. Therefore, to illustrate the difference between the models and observations over day and night time conditions, we estimated the model's increment distributions using NIG, separating the data following the same procedures used in Section 8.5. We then show how the α parameters compare with the estimated α parameters of the observations from the three tower locations[3]. Figures 10.11 and 10.12 show a series of plots with the α value for τ values of 10, 60, 180, and 360 min for wind speed and temperature increments, respectively from d02 and d03 domains. We performed bootstrapping on the observations and calculated α 100 times for each randomized sample of $\delta u/\sigma_{\delta u}$ to obtain a range of α values as indicated by the grey shaded regions. The left column of each figure represents the day time increment series, while the right column is the night. First, Figure 10.11 shows two tower locations (Cabauw (a-d) and FINO1 (e-h)) at the 10 m and 40 m height, respectively. From this perspective, it is clear to see that the estimated α values for all models are more or less in line with the observations with some distinct differences:

- There is more deviation at Cabauw than FINO1, for night and day, especially at larger scales indicating that the more homogeneous BL conditions provide a more accurate prediction.
- The two Cabauw day time plots (Figures 10.11a and 10.11c) show the largest deviations at larger scales MYJ performing the worst.
- Figure 10.11a provides the worst overall estimation in α where the model underestimates and overestimates it at small and large scales, respectively. This result was also shown in Figure 10.3c as the higher resolution results indicated heavier tails at small scales. This indicates that the day time series contributes the most to the high level of intermittency.
- The variation between night and day is clear, showing lower α values (more intermittency/steeper pdf) throughout the range of scales and the models all more in line with

---

[3] Given the comparison of NIG and the mesoscale observations shown in Chapters 8 and 9, we are confident that NIG-based estimations are capable of capturing the scale-dependent traits for the WRF simulations.

observations.

- The night time prediction is the best over the water and is nearly perfect at FINO1 at z = 40 m.

Finally, we will examine the temperature increments in a similar fashion. Figure 10.12 shows the range of α values comparing the model and observations for Cabauw (z = 10 m) and Høvsøre (z = 2 m). Similar to the previous figure, the models within a reasonable range of α values compared to observations. However, there are more significant deviations which need to be addressed:

- Like with the wind speed increments, the temperature field reveals a greater deviation at Cabauw than Høvsøre, for night and day, especially at larger scales indicating that the more homogeneous BL conditions provide a more accurate prediction.
- At small scales within the Cabauw plots, the α values for all the model simulations mischaracterize the steepness of the temperature pdfs, resulting in a less heavy tail within the models overall.
- In constrast to the statistical evaluation of temperature evaluation for the full increment series, the TEMF model produces the largest deviations for the day time conditions over Cabauw.
- The night time conditions at the two locations reveal a poor (underestimation) prediction of the α values for the models, at intermediate scales, suggesting that more temperature fluctuations exist in the model.
- At Høvsøre, the estimation for the day time, d02 results are the best, while the other plots indicate that there is a drastic underestimation in α for the large scale temperature increments for near surface temperature. However, all the models are well in line with one another.

Given the ability of NIG, this type of parameter-based comparison can be a key method of evaluating the perform of various models for long-term wind speed increments over diverse geographical locations. Although, this method cannot provide a direct physical explanation, it may be a valuable evaluation tool for the development of future PBL schemes.

To determine how well the α values from the models compare with the observations, a root mean squared error (RMSE) calculation was performed for each τ value. Tables 10.3 and 10.4 show the summed RMSE from the range of separations indictated on Figures 10.11

Table 10.3: RMSE summed over the range of separation as depicted in Figure 10.11 comparing the α values obtained from wind speed observations against the seven different PBL schemes.

	PBL schemes						
	MYJ	**MYNN2**	**MYNN3**	**QNSE**	**TEMF**	**UW**	**YSU**
Cabauw - Day, d02	1.96	1.14	1.15	2.21	0.93	1.31	1.49
Cabauw - Night, d02	1.10	1.38	1.42	0.73	1.00	1.91	1.09
FINO1 - Day, d02	0.36	0.33	0.42	0.24	0.25	0.36	0.46
FINO1 - Night, d02	0.35	0.27	0.37	0.25	0.24	0.42	0.34
Cabauw - Day, d03	1.91	1.93	1.70	2.24	1.30	1.39	1.55
Cabauw - Night, d03	0.67	0.59	0.43	0.89	0.89	0.24	0.66
FINO1 - Day, d03	0.30	0.27	0.33	0.51	0.70	0.56	0.46
FINO1 - Night, d03	0.41	0.28	0.27	0.50	0.51	0.49	0.28

and 10.12 for the wind speed and temperature increments, respectively. From this perspective, some conclusions can be drawn regarding the capabilities of the given PBL schemes to accurately estimate the α values. For instance, clearly from Table 10.3 overall the TEMF model appears to be the most accurate despite having issues at small scales ($\tau = 10$ min) from the GoF tests over the entire range of data as shown in Figure 10.7. Also, over Cabauw during the day QNSE appears to perform the worst, but does the best at predicting the night time boundary layer. Moreover, the differences between the d02 and d03 results suggest only marginal differences over the range of separations despite d03 depicting more deviations in α over Cabauw in Figure 10.11. Finally, over the ocean all the models produce the lowest RMSE suggesting that the models are more capable of handling smooth surfaces as compared to land surfaces.

In Table 10.4 the temperature increment data from observations are compared against the model results from Cabauw and Høvsøre. Overall, the RMSE for the temperature results are not as good compared to the wind speeds. However, the best model for the d02 domain is MYNN2 for temperature for day time conditions, while YSU is a close second. However, the largest RMSE during the day is TEMF which is different from the α values estimated from wind speed increments. Also, similar to the wind speed results over the ocean all the models for the temperature increments perform better over near coastal environments (i.e., Høvsøre) as opposed to the land surface at Cabauw. Using RMSE to determine the quality of the α

Table 10.4: RMSE summed over the range of separation as depicted in Figure 10.12 comparing the α values obtained from temperature observations against the seven different PBL schemes.

	PBL schemes						
	MYJ	**MYNN2**	**MYNN3**	**QNSE**	**TEMF**	**UW**	**YSU**
Cabauw - Day, d02	1.98	1.07	0.66	2.05	3.24	1.01	0.90
Cabauw - Night, d02	1.51	1.91	1.97	1.86	1.16	1.74	1.30
Høvsøre - Day, d02	0.38	0.59	0.61	0.59	0.31	0.66	0.86
Høvsøre - Night, d02	0.92	1.04	1.11	1.07	0.86	1.08	1.16
Cabauw - Day, d03	1.24	0.79	0.83	0.70	3.24	0.74	0.63
Cabauw - Night, d03	2.14	1.31	1.31	2.51	1.69	2.07	1.93
Høvsøre - Day, d03	0.94	1.06	1.10	1.28	1.19	1.46	1.29
Høvsøre - Night, d03	1.08	1.21	1.28	1.49	1.66	1.75	1.37

estimates from the model provide a quantative metric to evaluate their performance against the observations.

10.7 Conclusion

In this chapter, a detailed and quantative comparison was made for the wind speed and temperature increments using seven different WRF PBL schemes. Using a one year simulation and observations from three different tower locations near the North Sea, we showcased the ability of the state-of-the-art model to capture the scale-dependent features of the increments. From a visual perspective, the models all were able to represent the observed data well. However, when higher resolution simulations were used the models showed heavier tailed features which weren't evident in the observed data. Using the 2-sample GoF metrics, K-S and A-D, we further identified various flaws with the set of PBL schemes. Finally, using the NIG-based estimation, we showed the differences in the α parameter between models and observations for day and night conditions. This approach provides a potential benefit for evaluating the model's performance and testing future model development.

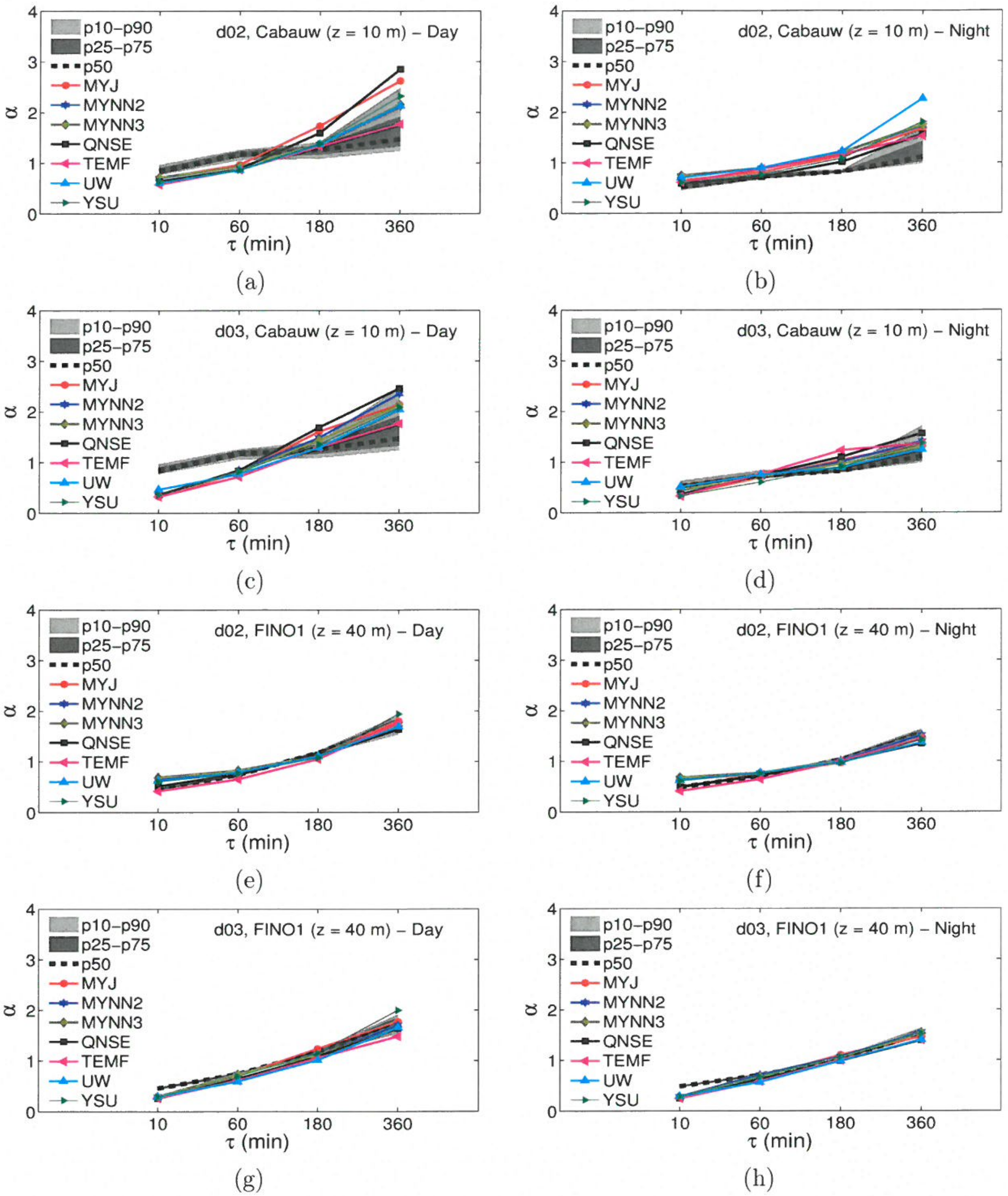
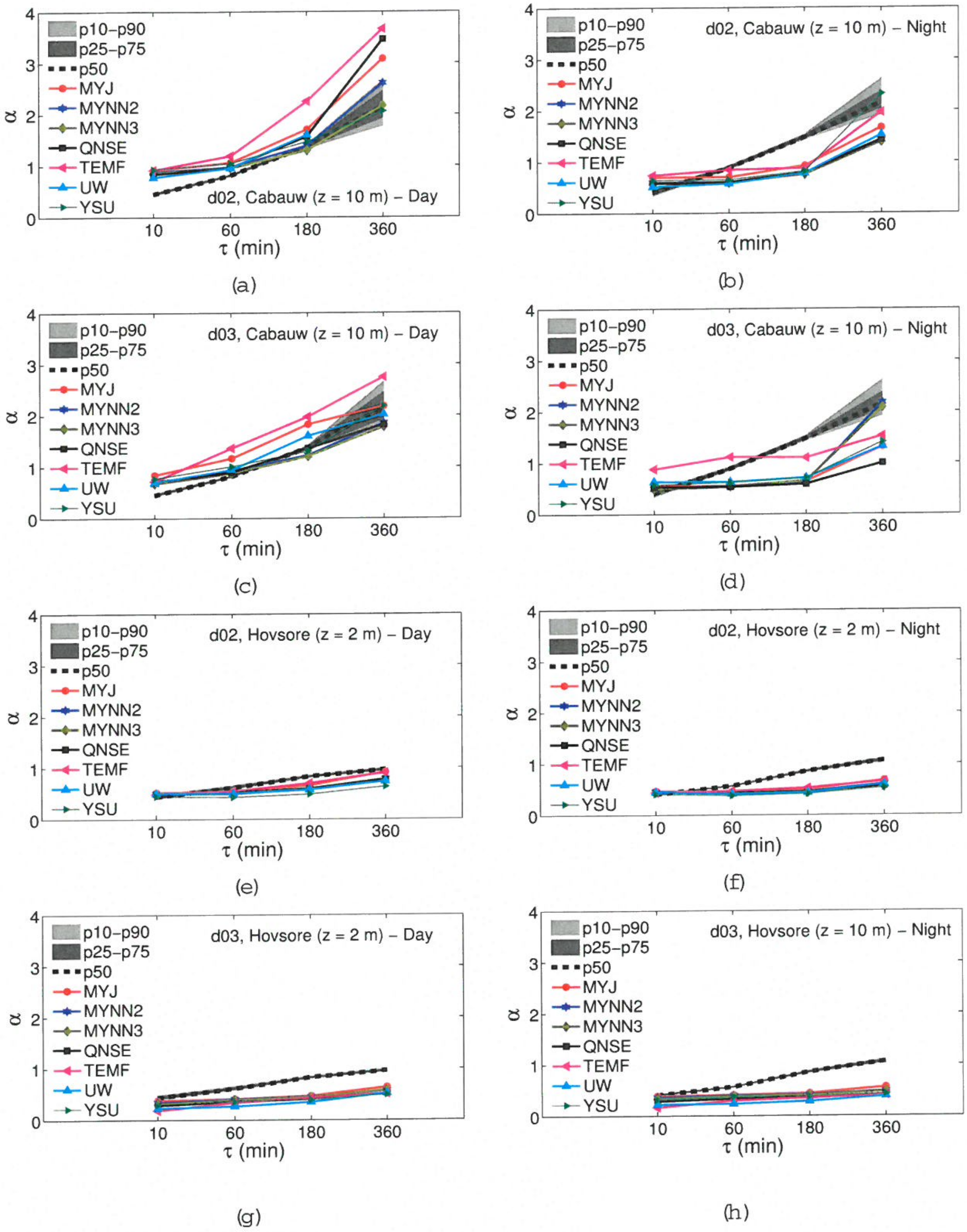


Figure 10.11: NIG parameter (α) for day (left column) and night (right column) series of wind speed increments as a function of τ comparing seven PBL schemes against model estimated α values for (a-d) Cabauw and (e-f) FINO1 for different heights for the d02 (a,b,e,f) and d03 (c,d,g,h) domains. The observational data is computed using bootstrap method with 100 randomized samples. The dashed black lines, dark grey shaded areas, and the light grey areas correspond to the medians, 25^{th}-75^{th} percentile ranges, and 10^{th}-90^{th} percentile ranges, respectively.



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

Figure 10.12: Same as Figure 10.11 except for Temperature increments for (a-d) Cabauw and (e-f) Høvsøre for different heights the d02 (a,b,e,f) and d03 (c,d,g,h) domains.

Chapter 11

Conclusions and Future Directions

11.1 Summary of Work

The purpose of this research was geared towards uncovering the statistical behavior of velocity and temperature increments within the atmospheric boundary layer across a range of scales. Utilizing a set of statistical distributions and a robust statistical estimation technique (i.e. maximum likelihood estimation (MLE)), this study revealed a capability of accurately capturing features of both the velocity and temperature increment fields. With the use of a variety of goodness of fit (GoF) metrics, it was shown that the normal inverse Gaussian (NIG) distribution is a prime candidate to be used as an estimate of the increment pdfs for a wide range of atmospheric motions. This finding answered my second and third questions posed in Chapter 1 as to whether or not a given statistical distribution model is able to capture the behavior of velocity and temperature fields across a range of spatio-temporal scales. This study is the first to perform this type of comprehensive examination into the behavior of the increments fields.

Utilizing diverse datasets, including idealized, wind tunnel experiments, atmospheric turbulence field measurements, multi-year ABL tower observations, and mesoscale models simulations, this study revealed remarkable similiarities (and some differences) between the small and larger scale components of the probability density functions increments fields. The features of these pdfs indicated a strong scale-dependent trait by which the tails of the distribution were highly non-Gaussian at small separation and steadily transitioned to a more Gaussian shape as separation increased. However, differences between the temperature and wind fields were evident, such that the tails of the temperature increments were heavier and slightly negatively skewed compared to the velocity counterparts. This finding provided an answer to my first science questions, addressed in Chapter 1, as to the universal behavior of velocity and temperature increments. Additionally, under various stability conditions, both in the large scale atmospheric flows and controlled wind tunnel enviroments, the features of the pdfs marked differences which revealed a larger degree of intermittency as separation increased. Even under these variations, the NIG model was still accurate in capturing these unique traits. This finding helped answer my fourth science question posed in Chapter 1.

In Chapter 4, we first demonstrated the capabilities of the NIG model and maximum likelihood estimation (MLE) method to deliver an unbiased estimation of higher-order structure function with relatively small samples (less than 10^5). With the use of extreme value theory (e.g., Hill plots), the behavior of the tails of the distribution revealed the limitations in the higher order moment estimation. The MLE approach was thoroughly tested using an idealized and observed dataset giving further credence to its ability for practical use for other increment fields. This work was then extended in Chapter 5 to explore the ability of other pdf models to match experiemental small scale turbulence data, both in the atmospheric boundary layer and in controlled experiments. Remarkably, both of these datasets portrayed very similar traits and we accurately estimated with the NIG model. The other models showed potential, but according to the GoF tests NIG outperformed generalized hyperbolic skew student's t (GHSST), variance gamma (VG), and lognormal superstatistics (LNSS). Also, the limitations of the method of moments estimation (MME) and histogram binning techniques were shown through the use of the LNSS model and the GoF evaluations. This confirmed that MLE-based estimation is superior to MME. While, the empirical distribution function approach was shown to be best for evaluting pdf models. Finally, the NIG parameters (α, β, μ, and δ) were presented revealing the capability of using parameter-based identification of the unique features of the velocity increments. In Chapter 6, the impact stability has on the small scale velocity and temperature increments were explored. From this analysis, it was clear to see, with the use of the NIG model parameters, that various stability conditions modulate the flow across a range of scales. This influence suggests that a greater degree of intermittency exist for larger scale flows under higher stability, with the temperature field exhibiting a stronger influence to changes in stability.

In Chapter 7, the features of the mesoscale characteristics of long-term wind speed increments were explored over diverse geographical locations. Such an extensive dataset has never been used to evaluate long-term statistics. It was discovered that, from a visual perspective, these fields are quasi-universal. They also appear to behave in a similiar manner to the small scale motions, except for exhibiting slightly heavier tails. We evaluated the wind ramp, up and down, conditions observed in this extensive dataset to uncover the similiarities. Using the probability of exceedence as a measuring stick, it became clear that the differences between each location and height were minor, but some deviations existed especially near the tails at ±10 standard deviations. The most striking contrast between the locations occurred at NWTC suggesting that complex terrain induces differences within the wind fields which are not observed near quasi-homogeneous environments. The finding requires further investigation by studying other complex terrain settings.

In Chapter 8, we examined the same mescoscale wind dataset to reveal the quality of the pdf models' fit to the normalized wind speed increments. Unlike with the small-scale turbulence field, GHSST showed the best match for the mesoscale range at small scales ($\tau = 10$ min) due to the model's ability to capture the heavier tailed features. The GoF metrics revealed similar results. However, from a Q-Q plot perspective, the NIG model showed some promise. Therefore, as before, the NIG-based parameter (α) was used to explore the differences between day and night increment fields. Quite interestingly, α for day time over all locations showed virtually no height dependence along with portraying similiar features between locations. However, during the night, over land, the influences of the stable boundary layer appeared to play a role in the variations, such that more intermittency was observed at lower heights and at larger scales. This finding was very similar to the wind tunnel study further corroborating the influences of stability regardless of the scale of motion.

The mesoscale temperature field across the diverse sites was examined next. It became clear that the pdfs of temperature behave differently than the wind speed. For instance, overall the tails of the distributions were heavier and slightly more left skewed than the wind. However, this behavior was similar to the small scale temperature field. Also, the pdf models were shown to fit the temperature increments with some accuracy as indicated by the GoF statistics and the Q-Q plots. From this perspective, again, the NIG model was chosen and we showed the variations between the sites during day and night conditions. Interestingly, during the night over land an opposite result ensued as more intermittency was evident at higher heights with increased separation. However, over the ocean uniform features in the pdf were observed. This feature confirms the fact that diurnal variations are minimal over the ocean providing further evidence that this unique approach for characterizing the ABL flows is a valid technique.

Finally, to address my fifth science question the WRF mesoscale model over a one-year simulated period was examined to identify the capability of seven different PBL parameterizations to capture these unique features of wind speed and temperature increments. We produced sub-minute time-series output data using the *tslist* option over three tower locations around the North Sea. Remarkably, from a visual perspective, minimal difference were observed between the mesoscale wind speed observations and all the different simulations. However, after examining the GoF statistics, the TEMF model underperformed compared to the other schemes at 10 min increments. Also, the higher resolution run (3 km) was shown to produce an excessive amount of intermittency, especially in the wind field, indicating that higher resolution runs do not necessarily improve results. Within the temperature field, the results for Cabauw were shown to be problematic as they deviated significantly from Høvsøre from a quantitative perspective. However, more work is required to investigate the root cause. Additionaly, the α parameter from NIG was used to showcase the diurnal variations between the WRF simulations and the observations. Overall, the model appeared to line up well with the observations, with the most significant differences coming from the higher resolution simulations. This approach provides a potential benefit for evaluating the model's performance and testing future model development.

11.2 Future Directions

Although this work was thoroughly conducted, we have just stratched the surface on the capability of utilizing our approach for understanding the characterisitics of the velocity and scalar increment fields. In turbulence research, which utilizes pdfs to characterize the increment fields, it is customary to simply visualize the quality of a model to represent a set of empirical turbulence data. However, we have shown that visual evaluation of the quality of pdf fits can lead to a mischaracterization. Thus, the need for statistical comparisons are essential for obtaining an accurate representation of the fits. In future turbulence research it is highly recommended to perform this type of GoF evaluation before claiming that a given pdf model provides a reliable estimation of empirical data.

One benefit of utilizing the NIG pdf as an estimate for velocity and temperature increment series is that one can easily compute all the moments of the distribution. With this capability, a direct connection can be made between NIG and other statistical characterization methods, such as structure function and extended self-similiarity (ESS), which utilizes the absolute moments of the increments. For instance, recent work introduced by Kiliyanpilakkil and Basu (2015-2016, [135, 136]) discussed the statistical similiarities of mesoscale wind speed from a scaling (or more specifically, ESS) perspective. However, the connection between ESS (or SF) and pdfs are not thoroughly addressed in this work. It would be an intriguing to further investigate the relationship between these methods. Since ~~the~~ inception of ESS, a number of turbulence researchers ~~has~~ have used it as an alternative for evaluating scaling characteristics with the inertial and mesoscale ranges. However, to this day, no physically explanation into why such an "extended" scaling range exists in turbulence under ESS.

> "In spite of several attempts to explain the success of ESS the latter is still not fully understood and we do not know how much we can trust scaling exponents derived by ESS. It would be nice to have at least one instance for which ESS not only works, but does so for reasons we can rationally understand (Chakraborty et al., 2010 [60])."

Therefore, we strongly believed that in conjunction with pdfs, the ESS behavior and other scaling descriptions, can be understood more extensively from statistical and physical perspectives. In fact, now that a suitable model pdf has been identified, a possible relationship between increment pdfs and ESS can be proposed using NIG. Efforts in this direction have recently been conducted by Basu 2016 (unpublished work). However, now that there has been a more rigorous evaluation of NIG across multiscales, more efforts can be made in this arena.

In terms of understanding the stability influences on the increment fields, our work uncovered interesting features within small scale wind tunnel experiments and the mesoscale range in the ABL. However, the small-scale atmospheric fields under various stability conditions were not addressed and need further evaluation. Additionally, due to the constraints of our datasets, we were limited to the lowest 200 m of the ABL. Therefore, in future field experiments it is recommended to extended the observational footprint farther into the ABL to understand the turbulent features that are present throughout the lower atmosphere, specifically near the top of the boundary layer.

Furthermore, we showcased the ability of the WRF mesoscale models, with seven different PBL parameterizations, to capture the wind speed and temperature increment field with some level of accuracy. However, there are uncertainties and limitations which exist when using the modeling and statistical approach proposed in the work. For instance, the mesoscale model has many tunable parameters such as initial and boundary conditions and other physical parameterization scheme, such as with land surface schemes, which could have potential impact on the prediction of these increment fields. Therefore, further sensitivity tests on these various model set-up conditions will increase the understanding of predictability of the models using our statistical approach. In fact, an objective way to estimate the uncertainty of a model is to perform an ensemble prediction, which could then be applied to our statistical methods for the increment fields. Limitations in this technique also require further work. In particular, our approach for characterizing the increment fields was examined from a 1-D perspective and was only conducted using the two atmospheric variables. To gain greater knowledge of the full picture, a future study should use of the 3-D model output space to not only evaluate the time-series of a given variable, but to uncover the multi-dimensional characterization. Additionally, the higher resolution results showed the greatest deviation from the observed data. Thus, a closer look into different horizontal grid resolutions, such as 1 km, could shed light into the effective model resolution, similar to the structure function and spectra approach conducted by Frehlich and Sharman [93].

From an application perspective, we gained a greater understanding of the wind ramp characteristics benefiting the wind energy industry (see Chapter 7). This work showed that the tail features of the long-term mesoscale wind increments revealed amazingly similar shape characteristics from several well-known field sites from around the world. We strongly believe that our results will tremendously benefit the future development of ramp forecasting methodologies. However, differences at NTWC (i.e., over complex terrain) were evident, thus further studies at other diverse geographical locations, utilizing our approach, would provide additional insight into the quasi-universal features shown in this study.

In terms of optical wave propagation, additional work is needed to make a connection between the eddy structure and the pdfs. For instance, the 1-D statistical description provided in this work, can only provide a glimsp into the full picture of turbulence. Therefore, using observational data from Light Detection and Ranging (LIDAR) systems or other state-of-the-art measurement tools, can allow one to visualize the turbulent eddies in the ABL and employ our statistical description simulanteously to gain insight from both angles and provide a 3-D picture of the behavior of velocity and scalar increment fields. As an illustrative example, Figure 11.1 shows a time versus height plot of a scalar field (water vapor mixing ratio acting as a great tracer) which depicts the development of the daytime ABL near Philadelphia, PA. From this perspective, one can observe the building of the thermal convective plumes from

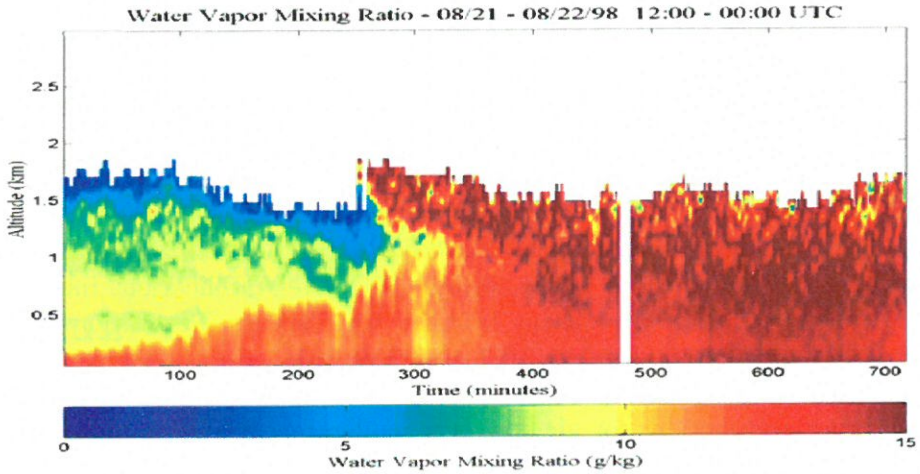


Figure 11.1: 12-hr time-height plot of water vapor mixing ratio (g/kg) from LIDAR measurements near Philadelphia, PA in August 1998 (adapted from Philbrick and Hallen (2015, [207])).

12:00 UTC (8:00 am local time) to 17:00 UTC (1:00 pm local time) throughout the lower portion of the ABL. Then as the ABL interacts and mixes with the residual layer, evidence of turbulent eddies can be observed. Moreover, turbulent eddies within the atmosphere may not necessarily fall under the category of statistically homogeneous and isotropic and as stated by Andrews and Philips (2005, [10]) the eddies of larger size are generally nonisotropic and their structure is not well defined. There, using this type of measurement in conjunction with our 1-D statistical approach for characterization turbulent features, a clearer picture can be presented to understand the behavior of various atmospheric flows for a wide range of scales.

Finally, given that optical turbulence (C_n^2) is directly related to the temperature increment field, δT, a future study can leverage on our approach using NIG as a tool for estimating temperature fluctuations, as well as humidity, in both the vertical and horizontal direction for applications in the field of optical wave propagation.

REFERENCES

[1] K. Aas and I. Hobæk Haff. NIG and skew Student's t: Two special cases of the generalised hyperbolic distribution, 2005. SAMBA/01/05, Norwegian Computing Center, Oslo.

[2] K. Aas and I. Hobæk Haff. The generalized hyperbolic skew students t-distribution (GHSST). *Journal of Financial Econometrics*, 4(2):275–309, 2006.

[3] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*, volume 55. Dover Publications Inc., 1972.

[4] S. I. Adelfang. On the relations between wind shears over various altitude intervals. *Journal of Applied Meteorology*, 10(1):156–159, 1971.

[5] K. G. Aivalis, K. R. Sreenivasan, Y. Tsuji, J. C. Klewicki, and C. A. Biltoft. Temperature structure functions for air flow over moderately heated ground. *Physics of Fluids*, 14(7):2439–2446, 2002.

[6] J. R. Alisse and C. Sidi. Experimental probability density functions of small-scale fluctuations in the stably stratified atmosphere. *Journal of Fluid Mechanics*, 402:137–162, 2000.

[7] R. J. Alliss and B. D. Felton. Validation of optical turbulence simulations from a numerical weather prediction model in support of adaptive optics design. In *Advanced Maui Optical and Space Surveillance Technologies Conference*, volume 1, 2009.

[8] G. Amati, R. Benzi, and S. Succi. Extended self-similarity in boundary layer turbulence. *Physical Review E*, 55(6):6985, 1997.

[9] T. W. Anderson and D. A. Darling. A test of goodness of fit. *Journal of the American Statistical Association*, 49(268):765–769, 1954.

[10] L. C. Andrews and R. L. Phillips. *Laser beam propagation through random media*, volume 1. SPIE press Bellingham, WA, 2005.

[11] W. M. Angevine, H. Jiang, and T. Mauritsen. Performance of an eddy diffusivity–mass flux scheme for shallow cumulus boundary layers. *Monthly Weather Review*, 138(7):2895–2912, 2010.

[12] F. Anselmet, Y. Gagne, E. J. Hopfinger, and R. A. Antonia. High-order velocity structure functions in turbulent shear flows. *Journal of Fluid Mechanics*, 140:63–89, 1984.

[13] R. A. Antonia, A. J. Chambers, D. Britz, and L. W. B. Browne. Organized structures in a turbulent plane jet: Topology and contribution to momentum and heat transport. *Journal of Fluid Mechanics*, 172:211–229, 1986.

[14] R. A. Antonia, E. J. Hopfinger, Y. Gagne, and F. Anselmet. Temperature structure functions in turbulent shear flows. *Physical Review A*, 30(5):2704, 1984.

[15] R. A. Antonia, B. R. Satyaprakash, and A. J. Chambers. Reynolds number dependence of velocity structure functions in turbulent shear flows. *Physics of Fluids*, 25(1):29–37, 1982.

[16] R. A. Antonia and R. J. Smalley. Anomalous scaling of velocity and temperature structure functions. *Physical Review E*, 63(2):025301, 2001.

[17] R. A. Antonia and C. W. Van Atta. Accuracy of structure functions of temperature in turbulent shear flows. *Physics of Fluids*, 21(7):1096–1099, 1978.

[18] A. Arnéodo, C. Baudet, F. Belin, R. Benzi, B. Castaing, B. Chabaud, R. G. Chavarria,

et al. Structure functions in turbulence, in various flow configurations, at Reynolds number between 30 and 5000, using extended self-similarity. *Europhysics Letters*, 34(6):411, 1996.

[19] A. Arnéodo, R. Benzi, J. Berg, L. Biferale, E. Bodenschatz, A. Busse, E. Calzavarini, et al. Universal intermittent properties of particle trajectories in highly turbulent flows. *Physical Review Letters*, 100(25):254504, 2008.

[20] S. P. S. Arya. Parameterizing the height of the stable atmospheric boundary layer. *Journal of Applied Meteorology*, 20(10):1192–1202, 1981.

[21] R. Baïle and J-F. Muzy. Spatial intermittency of surface layer wind fluctuations at mesoscale range. *Phys. Rev. Lett.*, 105(25):254501, 2010.

[22] O. E. Barndorff-Nielsen. Exponentially decreasing distributions for the logarithm of particle size. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 353–1674, pages 401–419. The Royal Society, 1977.

[23] O. E. Barndorff-Nielsen. Probability and statistics: Self-decomposability, finance and turbulence. In *Probability towards 2000*, pages 47–57. Springer, New York, 1996.

[24] O. E. Barndorff-Nielsen, P. Blæsild, and J. Schmiegel. A parsimonious and universal description of turbulent velocity increments. *European Physical Journal B*, 41:345–363, 2004.

[25] O. E. Barndorff-Nielsen and J. Schmiegel. Incremental similarity and turbulence. *arXiv preprint arXiv:1509.06515*, 2015.

[26] O. E. Barndorff-Nielsen and R. Stelzer. Absolute moments of generalized hyperbolic distributions and approximate scaling of normal inverse Gaussian lévy processes. *Scandinavian Journal of Statistics*, 32:617–637, 2005.

[27] S. Basu. Can the dynamic eddy-viscosity class of subgrid-scale models capture inertial-range properties of Burgers turbulence? *Journal of Turbulence*, 10(12):1–16, 2009.

[28] S. Basu, E. Foufoula-Georgiou, B. Lashermes, and A. Arnéodo. Estimating intermittency exponent in neutrally stratified atmospheric surface layer flows: a Robust framework based on magnitude cumulant and surrogate analyses. *Physics of Fluids*, 19(11):115102, 2007.

[29] S. Basu, F. Porté-Agel, E. Foufoula-Georgiou, J-F. Vinuesa, and M. Pahlow. Revisiting the local scaling hypothesis in stably stratified atmospheric boundary-layer turbulence: an Integration of field and laboratory measurements with large-eddy simulations. *Boundary-Layer Meteorology*, 119(3):473–500, 2006.

[30] G. K. Batchelor and A. A. Townsend. The nature of turbulent motion at large wavenumbers. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 199, pages 238–255. The Royal Society, 1949.

[31] C. Beck. Superstatistics in hydrodynamic turbulence. *Physica D: Nonlinear Phenomena*, 193(1):195–207, 2004.

[32] C. Beck and E. G. D. Cohen. Superstatistics. *Physica A: Statistical Mechanics and its Applications*, 322:267–275, 2003.

[33] C. Beck, E. G. D. Cohen, and S. Rizzo. Atmospheric turbulence and superstatistics. *Europhysics News*, 36(6):189–191, 2005.

[34] F. Belin, P. Tabeling, and H. Willaime. Exponents of the structure functions in a low temperature helium experiment. *Physica D*, 93(1):52–63, 1996.

[35] R. Benzi, L. Biferale, S. Ciliberto, M. V. Struglia, and R. Tripiccione. On the intermittent energy transfer at viscous scales in turbulent flows. *Europhysics Letters*, 32(9):709, 1995.

[36] R. Benzi, L. Biferale, A. Crisanti, G. Paladin, M. Vergassola, and A. Vulpiani. A random process for the construction of multiaffine fields. *Physica D: Nonlinear Phenomena*, 65(4):352–358, 1993.

[37] R. Benzi, L. Biferale, G. Paladin, A. Vulpiani, and M. Vergassola. Multifractality in the statistics of the velocity gradients in turbulence. *Physical Review Letters*, 67(17):2299, 1991.

[38] R. Benzi, S. Ciliberto, R. Tripiccione, C. Baudet, F. Massaioli, and S. Succi. Extended self-similarity in turbulent flows. *Physical Review E*, 48(1):R29, 1993.

[39] R. Benzi, G. Paladin, G. Parisi, and A. Vulpiani. On the multifractal nature of fully developed turbulence and chaotic systems. *Journal of Physics A: Mathematical and General*, 17(18):3521, 1984.

[40] R. Benzi, M. V. Struglia, and R. Tripiccione. Extended self-similarity in numerical simulations of three-dimensional anisotropic turbulence. *Physical Review E*, 53(6):R5565, 1996.

[41] R. Benzi, R. Tripiccione, F. Massaioli, S. Succi, and S. Ciliberto. On the scaling of the velocity and temperature structure functions in Rayleigh-Bénard convection. *Europhysics Letters*, 25(5):341, 1994.

[42] B. M. Bibby and M. Sørensen. Hyperbolic processes in finance. *Handbook of Heavy Tailed Distributions in Finance*, pages 211–248, 2003.

[43] B. Birnir. The Kolmogorov–Obukhov statistical theory of turbulence. *Journal of Nonlinear Science*, 23(4):657–688, 2013.

[44] B. Birnir, X. Chen, and L. Can. Moments and probability density functions in turbulent boundary layers. 2015.

[45] R. Bolgiano. Turbulent spectra in a stably stratified atmosphere. *Journal of Geophysical Research*, 64(12):2226–2229, 1959.

[46] R. Bolgiano. Structure of turbulence in stratified media. *Journal of Geophysical Research*, 67(8):3015–3023, 1962.

[47] A. Bossavy, R. Girard, and G. Kariniotakis. Forecasting ramps of wind power production with numerical weather prediction ensembles. *Wind Energy*, 16(1):51–63, 2013.

[48] F. Böttcher, C. H. Renner, H. P. Waldl, and J. Peinke. On the statistics of wind gusts. *Boundary-Layer Meteorology*, 108(1):163–173, 2003.

[49] F. Böttcher, St., Barth, and J. Peinke. Small and large scale fluctuations in atmospheric wind speeds. *Stochastic Environment Research and Risk Assessment*, 21:299–308, 2007.

[50] C. S. Bretherton and S. Park. A new moist turbulence parameterization in the Community Atmosphere Model. *Journal of Climate*, 22(12):3422–3448, 2009.

[51] G. C. Burton and W. J. A. Dahm. Multifractal subgrid-scale modeling for large-eddy simulation. II. backscatter limiting and a posteriori evaluation. *Physics of Fluids*, 17(7):075112, 2005.

[52] R. Calif and F. G. Schmitt. Modeling of atmospheric wind speed sequence using a lognormal continuous stochastic equation. *Journal of Wind Engineering and Industrial Aerodynamics*, 109:1–8, 2012.

[53] R. Calif and F. G. Schmitt. Multiscaling and joint multiscaling description of the atmospheric wind speed and the aggregate power output from a wind farm. *Nonlinear Processes in Geophysics*, 21(2):379–392, 2014.

[54] V. Carbone. Time intermittency in a shell model for fully developed hydromagnetic turbulence. *Europhysics Letters*, 27(8):581–586, 1994.

[55] B. Castaing. Lagrangian and eulerian velocity intermittency. *European Physcial Journal B*, 29(3):357–358, 2002.

[56] B. Castaing, Y. Gagne, and E. J. Hopfinger. Velocity probability density functions of high Reynolds number turbulence. *Physica D: Nonlinear Phenomena*, 46(2):177–200, 1990.

[57] B. Castaing, G. Gunaratne, L. Kadanoff, A. Libchaber, and F. Heslot. Scaling of hard thermal turbulence in Rayleigh-Bénard convection. *Journal of Fluid Mechanics*, 204(1):1–30, 1989.

[58] D. Cava and G. G. Katul. The effects of thermal stratification on clustering properties of canopy turbulence. *Boundary-Layer Meteorology*, 130(3):307–325, 2009.

[59] A. Celani, M. Cencini, A. Mazzino, and M. Vergassola. Active and passive fields face to face. *New Journal of Physics*, 6(1):72, 2004.

[60] S. Chakraborty, U. Frisch, and S. S. Ray. Extended self-similarity works for the Burgers equation and why. *Journal of Fluid Mechanics*, 649:275–285, 2010.

[61] G. R. Chavarria, C. Baudet, and S. Ciliberto. Extended self-similarity of passive scalars in fully developed turbulence. *Europhysics Letters*, 32(4):319, 1995.

[62] R. G. Chavarria, S. Ciliberto, C. Baudet, and E. Lévêque. Scaling properties of the streamwise component of velocity in a turbulent boundary layer. *Physica D: Nonlinear Phenomena*, 141(3):183–198, 2000.

[63] L. Chevillard, B. Castaing, E. Lévêque, and A. Arnéodo. Unified multifractal description of velocity increments statistics in turbulence: Intermittency and skewness. *Physica D: Nonlinear Phenomena*, 218(1):77–82, 2006.

[64] E. S. C. Ching. Probabilities for temperature differences in Rayleigh-Bénard convection. *Physical Review A*, 44(6):3622, 1991.

[65] E. S. C. Ching and W. C. Cheng. Anomalous scaling and refined similarity of an active scalar in a shell model of homogeneous turbulent convection. *Physical Review E*, 77(1):015303, 2008.

[66] H. K. Cho, K. P. Bowman, and G. R. North. A comparison of gamma and lognormal distributions for characterizing satellite rain rates from the tropical rainfall measuring mission. *Journal of Applied Meteorology*, 43(11):1586–1597, 2004.

[67] C. R. Chu, M. B. Parlange, G. G. Katul, and J. D. Albertson. Probability density functions of turbulent velocity and temperature in the atmospheric surface layer. *Water Resources Research*, 32(6):1681–1688, 1996.

[68] S. Cioni, S. Ciliberto, and J. Sommeria. Temperature structure functions in turbulent convection at low prandtl number. *Europhysics Letters*, 32(5):413, 1995.

[69] A. E. Cohen, S. M. Cavallo, M. C. Coniglio, and H. E. Brooks. A review of planetary boundary layer parameterization schemes and their sensitivity in simulating southeastern us cold season severe weather environments. *Weather and Forecasting*, 30(3):591–612, 2015.

[70] P. Constantin. Scaling exponents for active scalars. *Journal of Statistical Physics*, 90(3-4):571–595, 1998.

[71] S. Corrsin. On the spectrum of isotropic temperature fluctuations in an isotropic turbulence. *Journal of Applied Physics*, 22(4):469–473, 1951.

[72] M. E. Crovella, M. S. Taqqu, and A. Bestavros. Heavy-tailed probability distributions in the World Wide Web. In R. J. Adler, R. E. Feldman, and M. S. Taqqu, editors, *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, pages 3–26. Birkhäuser, Boston, 1998.

[73] R. B. d'Agostino and M. A. Stephens, editors. *Goodness-of-Fit Techniques.* Marcel Dekker, Inc., 1986.

[74] G. D'Amico, F. Petroni, and F. Prattico. First and second order semi-Markov chains for wind speed modeling. *Physica A*, 392:1194–1201, 2013.

[75] L. De Haan and A. Ferreira. *Extreme Value Theory: An Introduction.* Springer Science+Business Media, LLC, New York, 2007. 417 pp.

[76] J. W. Deardorff. Stratocumulus-capped mixed layers derived from a three-dimensional model. *Boundary-Layer Meteorology*, 18(4):495–527, 1980.

[77] A. L. M. Dekkers, J. H. J. Einmahl, and L. De Haan. A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, 17(4):1833–1855, 1989.

[78] J. Delour, J. F. Muzy, and A. Arneodo. Intermittency of 1D velocity spatial profiles in turbulence: a Magnitude cumulant analysis. *European Physcial Journal B*, 23(2):243–248, 2001.

[79] A. W. DeMarco and S. Basu. Estimating higher-order structure functions from geophysical turbulence time series: Confronting the curse of the limited sample size. *Physical Review E*, 95:052114, 2017.

[80] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B-Methodological*, pages 1–38, 1977.

[81] C. Draxl, A. Clifton, B-M. Hodge, and J. McCaa. The Wind Integration National Dataset (WIND) toolkit. *Applied Energy*, 151:355–366, 2015.

[82] H. Drees, L. De Haan, and S. Resnick. How to make a Hill plot. *Annals of Statistics*, 28:254–274, 2000.

[83] T. Dudok De Wit. Can high-order moments be meaningfully estimated from experimental turbulence measurements? *Physical Review E*, 70(5):055302, 2004.

[84] A. Einstein. *Investigations on the Theory of the Brownian Movement.* Courier Corporation, 1956.

[85] P. Embrechts, C. Klüppelberg, and T. Mokosh. Modeling extreme events. *Journal of Business*, 38(1):34–105, 1997.

[86] A. Eriksson, E. Ghysels, and L. Forsberg. *Approximating the probability distribution of functions of random variables: A new approach.* Cirano, 2004.

[87] B. Ernst and J. R. Seume. Investigation of site-specific wind field parameters and their effect on loads of offshore wind turbines. *Energies*, 5:3835–3855, 2012.

[88] C. Ferreira, J. Gama, L. Matias, A. Botterud, and J. Wang. A survey on wind power ramp forecasting. Technical report, Argonne National Laboratory (ANL), 2011.

[89] R. Finlay. *The variance gamma (VG) model with long range dependence.* PhD thesis, University of Sydney, 2009.

[90] D. Freedman and P. Diaconis. On the histogram as a density estimator: L 2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(4):453–476, 1981.

[91] J. Freedman, M. Markus, and R. Penc. Analysis of West Texas wind plant ramp-up and

ramp-down events. *AWS Truewind, LLC*, 2008.

[92] R. Frehlich. Laser scintillation measurements of the temperature spectrum in the atmospheric surface layer. *Journal of the Atmospheric Sciences*, 49(16):1494–1509, 1992.

[93] R. Frehlich and R. Sharman. The use of structure functions and spectra from numerical model output to determine effective model resolution. *Monthly Weather Review*, 136(4):1537–1553, 2008.

[94] F. N. Frenkiel and P. S. Klebanoff. On the lognormality of the small-scale structure of turbulence. *Boundary-Layer Meteorology*, 8(2):173–200, 1975.

[95] U. Frisch. *Turbulence.* Cambridge University Press, 1995. p.296.

[96] U. Frisch and G. Parisi. Fully developed turbulence and intermittency. *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, 88:71–88, 1985.

[97] U. Frisch, P-L. Sulem, and M. Nelkin. A simple dynamical model of intermittent fully developed turbulence. *Journal of Fluid Mechanics*, 87(04):719–736, 1978.

[98] Y. Gagne, B. Castaing, C. Baudet, and Y. Malécot. Reynolds dependence of third-order velocity structure functions. *Physics of Fluids*, 16(2):482–485, 2004.

[99] Y. Gagne, E. J. Hopfinger, and U. Frisch. A new universal scaling for fully developed turbulence: the Distribution of velocity increments. In *New Trends in Nonlinear Dynamics and Pattern-Forming Phenomena*, pages 315–319. Springer, 1990.

[100] Y. Gagne, M. Marchand, and B. Castaing. Conditional velocity pdf in 3-D turbulence. *Journal de Physique II*, 4(1):1–8, 1994.

[101] C. Gallego-Castillo, A. Cuerva-Tejero, and O. Lopez-Garcia. A review on the recent history of wind power ramp forecasting. *Renewable and Sustainable Energy Reviews*, 52:1148–1157, 2015.

[102] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge. Turbulent cascades in foreign exchange markets. *Nature*, 381(6585):767, 1996.

[103] C. H. Gibson, G. R. Stegen, and S. McConnell. Measurements of the universal constant in Kolmogorov's third hypothesis for high Reynolds number turbulence. *Physics of Fluids*, 13(10):2448–2451, 1970.

[104] J. P. Gollub, J. Clarke, M. Gharib, B. Lane, and O. N. Mesquita. Fluctuations and transport in a stirred fluid with a mean gradient. *Physical Review Letters*, 67(25):3507, 1991.

[105] R. B. Govindan and H. Kantz. Long-term correlations and multifractality in surface wind speed. *Europhysics Letters*, 68(2):184, 2004.

[106] B. Greaves, J. Collins, J. Parkes, and A. Tindal. Temporal forecast uncertainty for ramp events. *Wind Engineering*, 33(4):309–319, 2009.

[107] E. J. Gumbel. *Statistics of Extremes.* Columbia Univ. Press, New York, 1958.

[108] S. R. Hanna and R. Yang. Evaluations of mesoscale models' simulations of near-surface winds, temperature gradients, and mixing depths. *Journal of Applied Meteorology*, 40(6):1095–1104, 2001.

[109] J. A. Herweijer. *The small-scale structure of turbulence.* PhD thesis, Technische Universiteit Eindhoven, 1995.

[110] J. A. Herweijer and W. van de Water. Universal shape of scaling functions in turbulence. *Physical Review Letters*, 74(23):4651, 1995.

[111] B. M. Hill. A simple general approach to inference about the tail of a distribution. *Annals*

of Statistics, 3(5):1163–1174, 1975.

[112] M. Holzer and E. D. Siggia. Turbulent mixing of a passive scalar. *Physics of Fluids*, 6(5):1820–1837, 1994.

[113] S-Y. Hong, Y. Noh, and J. Dudhia. A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, 134(9):2318–2341, 2006.

[114] E. J. Hopfinger. Turbulence in stratified fluids: A review. *Journal of Geophysical Research: Oceans*, 92(C5):5287–5303, 1987.

[115] T. S. Horbury and A. Balogh. Structure function measurements of the intermittent MHD turbulent cascade. *Nonlinear Processes in Geophysics*, 4(3):185–199, 1997.

[116] H-Y. Huang, A. Hall, and J. Teixeira. Evaluation of the WRF PBL parameterizations for marine boundary layer clouds: Cumulus and stratocumulus. *Monthly Weather Review*, 141(7):2265–2271, 2013.

[117] Y. X. Huang, F. G. Schmitt, Z. M. Lu, and Y. L. Liu. Autocorrelation function of velocity increments time series in fully developed turbulence. *Europhysics Letters*, 86(4):40010, 2009.

[118] Y. X. Huang, F. G. Schmitt, Q. Zhou, X. Qiu, X. D. Shang, Z. M. Lu, et al. Scaling of maximum probability density functions of velocity and temperature increments in turbulent systems. *Physics of Fluids*, 23(12):125101, 2011.

[119] S. R. Hurst, E. Platen, and S. T. Rachev. Option pricing for asset returns driven by subordinated processes. Technical report, The Australian National University, 1995.

[120] Z. I. Janjić. The step-mountain coordinate: physical package. *Monthly Weather Review*, 118(7):1429–1443, 1990.

[121] Z. I. Janjić. The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Monthly Weather Review*, 122(5):927–945, 1994.

[122] M. H. Jensen, G. Paladin, and A. Vulpiani. Intermittency in a cascade model for three-dimensional turbulence. *Physical Review A*, 43(2):798, 1991.

[123] M. A. Jiménez and J. Cuxart. Study of the probability density functions from a large-eddy simulation for a stably stratified boundary layer. *Boundary-Layer Meteorology*, 118(2):401–420, 2006.

[124] I. T. Jolliffe and D. B. Stephenson. *Forecast verification: a practitioner's guide in atmospheric science*. John Wiley & Sons, 2003.

[125] H. E. Jørgensen, T. Mikkelsen, S. E. Gryning, S. Larsen, P. Astrup, and P. E. Sørensen. Measurements from Høvsøre met mast. Technical report, Technical Report Risø, 2008.

[126] M-C. Jullien, P. Castiglione, J Paret, and P. Tabeling. Vorticity statistics in the 2D enstrophy cascade and tracer dispersion in the Batchelor regime. page 168. Cambridge University Press, 2001.

[127] H. Kahalerras, Y. Malecot, Y. Gagne, and B. Castaing. Intermittency and Reynolds number. *Physics of Fluids*, 10(4):910–921, 1998.

[128] P. Kailasnath, K. R. Sreenivasan, and G. Stolovitzky. Probability density of velocity increments in turbulent flows. *Physical Review Letters*, 68:2766–2769, 1992.

[129] C. Kamath. Understanding wind ramp events through analysis of historical data. In *Transmission and Distribution Conference and Exposition, 2010 IEEE PES*, pages 1–6. IEEE, 2010.

[130] D. Karlis. An EM type algorithm for maximum likelihood estimation of the normal–inverse Gaussian distribution. *Statistics and Probability Letters*, 57:43–52, 2002.

[131] D. Karlis and J. Lillestöl. Bayesian estimation of NIG models via Markov chain Monte Carlo methods. *Applied Stochastic Models in Business and Industry*, 20(4):323–338, 2004.

[132] G. G. Katul, C. Angelini, D. De Canditiis, U. Amato, B. Vidakovic, and J. D. Albertson. Are the effects of large scale flow conditions really lost through the turbulent cascade? *Geophysical research letters*, 30(4), 2003.

[133] R. G. Kavasseri and R. Nagarajan. A multifractal description of wind speed records. *Chaos, Solitons & Fractals*, 24(1):165–173, 2005.

[134] N. D. Kelley, G. N. Scott, and B. J. Jonkman. The great plains turbulence environment: its origins, impact and simulation. Number NREL/CP-500-40176, Golden, CO, 2006.

[135] V. P. Kiliyanpilakkil and S. Basu. Extended self-similarity of atmospheric boundary layer wind fields in mesoscale regime: Is it real? *Europhysics Letters*, 112(6):64003, 2016.

[136] V. P. Kiliyanpilakkil, S. Basu, A. Ruiz-Columbié, G. Araya, L. Castillo, B. Hirth, and W. Burgett. Buoyancy effects on the scaling characteristics of atmospheric boundary layer wind fields in the mesoscale range. *Physical Review E*, 92(3):033005, 2015.

[137] K. Koçak. Examination of persistence properties of wind speed records using detrended fluctuation analysis. *Energy*, 34(11):1980–1985, 2009.

[138] A. N. Kolmogorov. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. In *Doklady Akademii Nauk SSSR*, volume 30–4, pages 299–303, 1941.

[139] A. N. Kolmogorov. A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. *Journal of Fluid Mechanics*, 13(01):82–85, 1962.

[140] R. H. Kraichnan. Turbulent cascade and intermittency growth. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 434–1890, pages 65–78. The Royal Society, 1991.

[141] R. H. Kraichnan. Anomalous scaling of a randomly advected passive scalar. *Physical Review Letter*, 72(7):1016–1019, 1994.

[142] G. J. Kunkel and I. Marusic. Study of the near-wall-turbulent region of the high-Reynolds-number boundary layer using an atmospheric flow. *Journal of Fluid Mechanics*, 548:375–402, 2006.

[143] A. Y. S. Kuo and S. Corrsin. Experiments on internal intermittency and fine-structure distribution functions in fully turbulent fluid. *Journal of Fluid Mechanics*, 50(02):285–319, 1971.

[144] V. R. V. Kuznetsov and V. A. Sabel'nikov. *Turbulence and Combustion.* New York, NY; Hemisphere Publishing, 1990.

[145] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright. Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9(1):112–147, 1998.

[146] F. Lampariello. On the use of the Kolmogorov-Smirnov statistical test for immunofluorescence histogram comparison. *Cytometry Part A*, 39(3):179–188, 2000.

[147] L. D. Landau and E. M. Lifshitz. *Fluid Mechanics.* Russian edition, 1944.

[148] A. S. Lanotte, S. K. Malapaka, and L. Biferale. On the vortex dynamics in fractal fourier

turbulence. *The European Physical Journal E*, 39(4):1–8, 2016.

[149] X. G. Larsén, S. Ott, J. Badger, A. N. Hahmann, and J. Mann. Recipes for correcting the impact of effective mesoscale resolution on the estimation of extreme winds. *Journal of Applied Meteorology and Climatology*, 51(3):521–533, 2012.

[150] M. K. Lauren, M. Menabde, and G. L. Austin. Analysis and simulation of surface-layer winds using multiplicative cascade models with self-similar probability densities. *Boundary-Layer Meteorology*, 100(2):263–286, 2001.

[151] M. K. Lauren, M. Menabde, A. W. Seed, and G. L. Austin. Characterisation and simulation of the multiscaling properties of the energy-containing scales of horizontal surface-layer winds. *Boundary-Layer Meteorology*, 90(1):21–46, 1999.

[152] R. S. Lawrence and J. W. Strohbehn. A survey of clear-air propagation effects relevant to optical communications. *Proceedings of the IEEE*, 58(10):1523–1545, 1970.

[153] P. Lengua Lafosse. An empirical application of stochastic volatility models to Latin-American stock returns using generalized hyperbolic skew student's t-distribution (GHSST). *Pontificia Universidad Católica del Perú*, 2015.

[154] E. Lévêque, R. G. Chavarria, C. Baudet, and S. Ciliberto. Scaling laws for the turbulent mixing of a passive scalar in the wake of a cylinder. *Physics of Fluids*, 11(7):1869–1879, 1999.

[155] L. Liu and F. Hu. Cascade-like and scaling behavior of wind velocity increments in the atmospheric surface layer. *Physica A*, 392(23):5808–5816, 2013.

[156] L. Liu, F. Hu, X. L. Cheng, and L. L. Song. Probability density functions of velocity increments in the atmospheric boundary layer. *Boundary-Layer Meteorology*, 134:243–255, 2010.

[157] D. Lohse and K. Q. Xia. Small-scale properties of turbulent Rayleigh-Bénard convection. *Annual Review of Fluid Mechanics*, 42:335–364, 2010.

[158] E. N. Lorenz. Low order models representing realizations of turbulence. *Journal of Fluid Mechanics*, 55(03):545–563, 1972.

[159] S. Lovejoy, W. J. S. Currie, Y. Tessier, M. R. Claereboudt, E. Bourget, J. C. Roff, and D. Schertzer. Universal multifractals and ocean patchiness: phytoplankton, physical fields and coastal heterogeneity. *Journal of Plankton Research*, 23(2):117–141, 2001.

[160] A. Luceño. Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators. *Computational Statistics & Data Analysis*, 51(2):904–917, 2006.

[161] D. B. Madan and E. Seneta. The variance gamma (VG) model for share market returns. *Journal of Business*, pages 511–524, 1990.

[162] L. Mahrt. Stratified atmospheric boundary layers. *Boundary-Layer Meteorology*, 90(3):375–396, 1999.

[163] L. Mahrt. Stably stratified atmospheric boundary layers. *Annual Review of Fluid Mechanics*, 46:23–45, 2014.

[164] Y. Malécot, C. Auriault, H. Kahalerras, Y. Gagne, O. Chanal, B. Chabaud, and B. Castaing. A statistical estimator of turbulence intermittency in physical and numerical experiments. *The European Physical Journal B-Condensed Matter and Complex Systems*, 16(3):549–561, 2000.

[165] B. B. Mandelbrot. Possible refinement of the lognormal hypothesis concerning the distribution of energy dissipation in intermittent turbulence. In *Statistical Models and Turbu-*

lence, pages 333–351. Springer, 1972.

[166] J. U. Márquez and J. Schmiegel. Modelling turbulent time series by BSS-processes. In *The Fascination of Probability, Statistics and their Applications*, pages 29–52. Springer, 2016.

[167] E. Marsch and C-Y. Tu. Non-Gaussian probability distributions of solar wind fluctuations. *Annales Geophysicae*, 12–12:1127–1138, 1994.

[168] E. Marsch and C-Y. Tu. Intermittency, non-Gaussian statistics and fractal scaling of MHD fluctuations in the solar wind. *Nonlinear Processes in Geophysics*, 4(2):101–124, 1997.

[169] D. Martin. The battle above. CBS 60 Minutes, April 2015.

[170] I. Marusic and G. J. Kunkel. Streamwise turbulence intensity formulation for flat-plate boundary layers. *Physics of Fluids*, 15(8):2461–2464, 2003.

[171] T. Mauritsen, G. Svensson, S. S. Zilitinkevich, I. Esau, L. Enger, and B. Grisogono. A total turbulent energy closure model for neutrally and stably stratified atmospheric boundary layers. *Journal of the Atmospheric Sciences*, 64(11):4113–4126, 2007.

[172] I. Mazzitelli and A. S. Lanotte. Active and passive scalar intermittent statistics in turbulent atmospheric convection. *Physica D*, 241(3):251–259, 2012.

[173] G. L. Mellor and T. Yamada. Development of a turbulence closure model for geophysical fluid problems. *Reviews of Geophysics*, 20(4):851–875, 1982.

[174] C. Meneveau, K. R. Sreenivasan, P. Kailasnath, and M. S. Fan. Joint multifractal measures: Theory and applications to turbulence. *Physical Review A*, 41(2):894, 1990.

[175] P. Mestayer. Local isotropy and anisotropy in a high-Reynolds-number turbulent boundary layer. *Journal of Fluid Mechanics*, 125:475–503, 1982.

[176] P. G. Mestayer, C. H. Gibson, M. F. Coantic, and A. S. Patel. Local anisotropy in heated and cooled turbulent boundary layers. *Physics of Fluids*, 19(9):1279–1287, 1976.

[177] O. Métais and M. Lesieur. Spectral large-eddy simulation of isotropic and stably stratified turbulence. *Journal of Fluid Mechanics*, 239:157–194, 1992.

[178] F. Moisy, H. Willaime, J. S. Andersen, and P. Tabeling. Passive scalar intermittency in low temperature helium flows. *Physical Review Letters*, 86(21):4827, 2001.

[179] A. S. Monin and A. M. F. Obukhov. Basic laws of turbulent mixing in the surface layer of the atmosphere. *Contrib. Geophys. Inst. Acad. Sci. USSR*, 151(163):e187, 1954.

[180] A. S. Monin and A. M. Yaglom. *Statistical Fluid Mechanics: Mechanics of Turbulence*, volume 1. MIT Press, Cambridge, MA, 1971. 769 pp.

[181] A. S. Monin and A. M. Yaglom. *Statistical fluid mechanics: Mechanics of Turbulence*, volume 2. MIT Press, 1975.

[182] W. A. A. Monna and J. G. Van der Vliet. *Facilities for research and weather observations on the 213 m tower at Cabauw and at remote locations.* KNMI De Bilt, The Netherlands, 1987.

[183] D. S. Moore. Tests of chi-squared type. *Goodness-of-Fit Techniques*, 634(5), 1986.

[184] A. Morales, M. Wächter, and J. Peinke. Characterization of wind turbulence by higher-order statistics. *Wind Energy*, 15:391–406, 2012.

[185] J. F. Muzy, R. Baïle, and P. Poggi. Intermittency of surface-layer wind velocity series in the mesoscale range. *Physical Review E*, 81(5):056308, 2010.

[186] I. J. Myung. Tutorial on maximum likelihood estimation. *Journal of Mathematical Psy-*

chology, 47(1):90–100, 2003.

[187] M. Nakanishi and H. Niino. An improved Mellor–Yamada level-3 model with condensation physics: Its design and verification. *Boundary-Layer Meteorology*, 112(1):1–31, 2004.

[188] M. Nakanishi and H. Niino. An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Boundary-Layer Meteorology*, 119(2):397–407, 2006.

[189] N. B. Negra, O. Holmstrøm, B. Bak-Jensen, and P. Sørensen. Model of a synthetic wind speed time series generator. *Wind Energy*, 11:193–209, 2008.

[190] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.

[191] T. Neumann, K. Nolopp, M. Strack, H. Mellinghoff, H. Söker, E. Mittelstaedt, W. J. Gerasch, and G. Fischer. Erection of German offshore measuring platform in the North Sea. *DEWI Magazin*, 23:32–46, 2003.

[192] F. T. M. Nieuwstadt. Some aspects of the turbulent stable boundary layer. In *Boundary-Layer Meteorology*, pages 31–55. Springer, 1984.

[193] F. T. M. Nieuwstadt. The turbulent structure of the stable, nocturnal boundary layer. *Journal of the Atmospheric Sciences*, 41(14):2202–2216, 1984.

[194] F. T. M. Nieuwstadt. A model for the stationary, stable boundary layer. In J. C. R. Hunt, editor, *Turbulence and Diffusion in Stable Environments*, pages 149–179. Clarendon Press, Oxford, UK, 1985.

[195] A. Noullez, G. Wallace, W. Lempert, R. B. Miles, and U. Frisch. Transverse velocity increments in turbulent flow using the RELIEF technique. *Journal of Fluid Mechanics*, 339:287–307, 1997.

[196] E. A. Novikov and R. W. Stewart. The intermittency of turbulence and the spectrum of energy dissipation fluctuations. *Izv. Geophys. Ser.*, 3:408–413, 1964.

[197] C. G. Nunalee, Á. Horváth, and S. Basu. High-resolution numerical modeling of mesoscale island wakes and sensitivity to static topographic RELIEF data. *Geoscientific Model Development*, 8(8):2645–2653, 2015.

[198] A. M. Obukhov. Spectral energy distribution in a turbulent flow. In *Dokl. Akad. Nauk SSSR*, volume 32–1, pages 22–24, 1941.

[199] A. M. Obukhov. Some specific features of atmospheric turbulence. *Journal of Geophysical Research*, 67(8):3011–3014, 1962.

[200] K. Ohkitani and M. Yamada. Temporal intermittency in the energy cascade process and local Lyapunov analysis in fully-developed model turbulence. *Progress Theoretical Physics*, 81(2):329–341, 1989.

[201] Y. Ohya. Wind-tunnel study of atmospheric stable boundary layers over a rough surface. *Boundary-Layer Meteorology*, 98(1):57–82, 2001.

[202] Y. Ohya, D. E. Neff, and R. N. Meroney. Turbulence structure in a stratified boundary layer under stable conditions. *Boundary-Layer Meteorology*, 83(1):139–162, 1997.

[203] M. Ould-Rouis, F. Anselmet, P. Le Gal, and S. Vaienti. Statistics of temperature increments in fully developed turbulence part ii. experiments. *Physica D: Nonlinear Phenomena*, 85(3):405–424, 1995.

[204] N. S. Padhye, C. W. Smith, and W. H. Matthaeus. Distribution of magnetic field components. *Journal of Geophysical Research*, 106(A9):18–635, 2001.

[205] M. Paolella. *Intermediate Probability: A Computational Approach.* John Wiley & Sons, 2007.

[206] A. Peña, R. Floors, A. Sathe, S. E. Gryning, R. Wagner, M. S. Courtney, X. G Larsén, et al. Ten years of boundary-layer and wind-power meteorology at Høvsøre, Denmark. *Boundary-Layer Meteorology*, 158(1):1–26, 2016.

[207] C. R. Philbrick and H. D. Hallen. Lidar investigations of atmospheric dynamics. In *SPIE Optical Engineering+ Applications*, pages 96120C–96120C. International Society for Optics and Photonics, 2015.

[208] R. L. Phillips and L. C. Andrews. Universal statistical model for irradiance fluctuations in a turbulent medium. *Journal of the Optical Society of America*, 72(7):864–870, 1982.

[209] J. Pickands III. Statistical inference using extreme order statistics. *Annals of Statistics*, 3(1):119–131, 1975.

[210] J. J. Podesta, M. A. Forman, C. W. Smith, D. C. Elton, Y. Malécot, and Y. Gagne. Accurate estimation of third-order moments from turbulence measurements. *Nonlinear Processes in Geophysics*, 16(1):99–110, 2009.

[211] P. D. Praetz. The distribution of share price changes. *Journal of Business*, pages 49–55, 1972.

[212] A. Praskovsky and S. Oncley. Probability density distribution of velocity differences at very high Reynolds numbers. *Physical Review Letters*, 73(25):3399, 1994.

[213] K. Prause. *The generalized hyperbolic model: Estimation, financial derivatives, and risk measures.* PhD thesis, University of Freiburg, 1999.

[214] A. Pumir. Small-scale properties of scalar and velocity differences in three-dimensional turbulence. *Physics of Fluids*, 6(12):3974–3984, 1994.

[215] M. Ragwitz and H. Kantz. Indispensable finite time corrections for Fokker-Planck equations from time series data. *Physical Review Letters*, 87(25):254501, 2001.

[216] L. F. Richardson. *Weather prediction by numerical process.* Cambridge University Press, 1922.

[217] T. H. Rydberg. The normal inverse Gaussian Lévy process: simulation and approximation. *Communication in Statistical Stochastic Models*, 13(4):887–910, 1997.

[218] C. Salem, A. Mangeney, S. D. Bale, and P. Veltri. Solar wind magnetohydrodynamics turbulence: anomalous scaling and role of intermittency. *Astrophyscial Journal*, 702(1):537–553, 2009.

[219] F. Schmitt, D. Schertzer, S. Lovejoy, and Y. Brunet. Empirical study of multifractal phase transitions in atmospheric turbulence. *Nonlinear Processes in Geophysics*, 1(2/3):95–104, 1994.

[220] F. W. Scholz and M. A. Stephens. K-sample Anderson–Darling tests. *Journal of the American Statistical Association*, 82(399):918–924, 1987.

[221] D. J. Scott, D. Würtz, C. Dong, and T. T. Tran. Moments of the generalized hyperbolic distribution. *Computational Statistics*, 26:459–476, 2011.

[222] D. W. Scott. On optimal and data-based histograms. *Biometrika*, pages 605–610, 1979.

[223] E. Seneta. Fitting the variance-gamma model to financial data. *Journal of Applied Probability*, 41(A):177–187, 2004.

[224] L. Seuront, F. Schmitt, Y. Lagadeuc, D. Schertzer, and S. Lovejoy. Universal multifractal analysis as a tool to characterize multiscale intermittent patterns: Example of phytoplank-

ton distribution in turbulent coastal waters. *Journal of Plankton Research*, 21(5):877–922, 1999.

[225] Z. S. She and E. Leveque. Universal scaling laws in fully developed turbulence. *Physical Review Letters*, 72(3):336, 1994.

[226] B. I. Shraiman and E. D. Siggia. Scalar turbulence. *Nature*, 405(6787):639–646, 2000.

[227] A. P. Siebesma, P. M. M. Soares, and J. Teixeira. A combined eddy-diffusivity mass-flux approach for the convective boundary layer. *Journal of the atmospheric sciences*, 64(4):1230–1248, 2007.

[228] W. C. Skamarock. Evaluating mesoscale NWP models using kinetic energy spectra. *Monthly Weather Review*, 132(12):3019–3032, 2004.

[229] W. C. Skamarock, J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X. Y. Huang, W. Wang, and J. G. Powers. A description of the advanced research WRF version 3. In *Technical Report NCAR/TN-457+STR*, 2008.

[230] L. Skrbek, J. J. Niemela, K. R. Sreenivasan, and R. J. Donnelly. Temperature structure functions in the Bolgiano regime of thermal convection. *Physical Review E*, 66(3):036303, 2002.

[231] W. D. Smyth and J. N. Moum. Length scales of turbulence in stably stratified mixing layers. *Physics of Fluids*, 12(6):1327–1342, 2000.

[232] D. Sornette. *Critical Phenomena in Natural Sciences: Chaos, Fractals, Self-organization and Disorder: Concepts and Tools*. Springer-Verlag, Heidelberg, 2006.

[233] D. Sornette. Probability distributions in complex systems. In *Encyclopedia of Complexity and Systems Science*, pages 7009–7024. Springer, 2009.

[234] L. Sorriso-Valvo, V. Carbone, P. Veltri, G. Consolini, and R. Bruno. Intermittency in the solar wind turbulence through probability distribution functions of fluctuations. *Geophysical Research Letters*, 26(13):1801–1804, 1999.

[235] K. R. Sreenivasan. On local isotropy of passive scalars in turbulent shear flows. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 434–1890, pages 165–182. The Royal Society, 1991.

[236] K. R. Sreenivasan. The passive scalar spectrum and the Obukhov–Corrsin constant. *Physics of Fluids*, 8(1):189–196, 1996.

[237] K. R. Sreenivasan. Fluid turbulence. *Reviews of Modern Physics*, 71(2):S383, 1999.

[238] K. R. Sreenivasan and R. A. Antonia. Skewness of temperature derivatives in turbulent shear flows. *The Physics of Fluids*, 20(12):1986–1988, 1977.

[239] K. R. Sreenivasan and R. A. Antonia. The phenomenology of small-scale turbulence. *Annual Review of Fluid Mechanics*, 29(1):435–472, 1997.

[240] K. R. Sreenivasan, R. A. Antonia, and D. Britz. Local isotropy and large structures in a heated turbulent jet. *Journal of Fluid Mechanics*, 94(4):745–775, 1979.

[241] K. R. Sreenivasan and B. Dhruva. Is there scaling in high-Reynolds-number turbulence? *Progress of Theoretical Physics Supplement*, 130:103–120, 1998.

[242] D. J. Stensrud. *Parameterization schemes: keys to understanding numerical weather prediction models*. Cambridge University Press, 2009.

[243] R. W. Stewart. Turbulence and waves in a stratified atmosphere. *Radio Science*, 4(12):1269–1278, 1969.

[244] R. W. Stewart, J. R. Wilson, and R. W. Burling. Some statistical properties of small scale

turbulence in an atmospheric boundary layer. *Journal of Fluid Mechanics*, 41(01):141–152, 1970.

[245] G. Stolovitzky and K. R. Sreenivasan. Scaling of structure functions. *Physical Review E*, 48(1):R33, 1993.

[246] R. B. Stull. *An Introduction to Boundary Layer Meteorology*, volume 13. Springer, 1988.

[247] S. Sukoriansky, B. Galperin, and V. Perov. Application of a new spectral theory of stably stratified turbulence to the atmospheric boundary layer over sea ice. *Boundary-Layer Meteorology*, 117(2):231–257, 2005.

[248] P. Tabeling, G. Zocchi, F. Belin, J. Maurer, and H. Willaime. Probability density functions, skewness, and flatness in large Reynolds number turbulence. *Physical Review E*, 53(2):1613, 1996.

[249] L. Telesca and M. Lovallo. Analysis of the time dynamics in wind records by means of multifractal detrended fluctuation analysis and the Fisher–Shannon information plane. *Journal Statistical Mechanics: Theory and Experiment*, 2011(07):P07001, 2011.

[250] H. Tennekes and J. C. Wyngaard. The intermittent small-scale structure of turbulence: Data-processing hazards. *Journal of Fluid Mechanics*, 55(01):93–103, 1972.

[251] S. T. Thoroddsen and C. W. Van Atta. Exponential tails and skewness of density-gradient probability density functions in stably stratified turbulence. *Journal of Fluid Mechanics*, 244:547–566, 1992.

[252] F. Toschi and E. Bodenschatz. Lagrangian properties of particles in turbulence. *Annual Review of Fluid Mechanics*, 41:375–404, 2009.

[253] A. Tsinober. *An Informal Introduction to Turbulence*, volume 63. Kluwer Academic Publishers, Dordrecht, 2001.

[254] M. Türk, K. Grigutsch, and S. Emeis. The wind profile above the sea – investigations basing on four years of FINO 1 data. *DEWI Magazin*, 33:12–16, 2008.

[255] R. Turner, X. Zheng, N. Gordon, M. Uddstrom, G. Pearson, R. de Vos, and S. Moore. Creating synthetic wind speed time series for 15 New Zealand wind farms. *Journal of Applied Meteorology and Climatology*, 50(12):2394–2409, 2011.

[256] A. P. van Ulden and J. Wieringa. Atmospheric boundary layer research at Cabauw. *Boundary-Layer Meteorology*, 78(1):39–69, 1996.

[257] C. W. Van Atta and W. Y. Chen. Structure functions of turbulence in the atmospheric boundary layer over the ocean. *Journal of Fluid Mechanics*, 44(01):145–159, 1970.

[258] C. W. Van Atta and J. Park. Statistical self-similarity and inertial subrange turbulence. In *Statistical Models and Turbulence*, pages 402–426. Springer, 1972.

[259] W. van de Water and J. A. Herweijer. High-order structure functions of turbulence. *Journal of Fluid Mechanics*, 387:3–37, 1999.

[260] H. F. C. Velho, R. R. Rosa, F. M. Ramos, R. A. Pielke, G. A. Degrazia, C. R. Neto, and A. Zanandrea. Multifractal model for eddy diffusivity and counter-gradient term in atmospheric turbulence. *Physica A*, 295(1):219–223, 2001.

[261] W. N. Venables and B. D. Ripley. *Modern applied statistics.* Springer Science & Business Media, 2013.

[262] J. H. Venter and P. J. de Jongh. Risk estimation using the normal inverse Gaussian distribution. *Journal of Risk*, 4:1–24, 2002.

[263] J. W. Verkaik and A. A. M. Holtslag. Wind profiles, momentum fluxes and roughness

lengths at Cabauw revisited. *Boundary-Layer Meteorology*, 122:701–719, 2007.

[264] A. Vincent and M. Meneguzzi. The spatial structure and statistical properties of homogeneous turbulence. *Journal of Fluid Mechanics*, 225:1–20, 1991.

[265] J. M. Vindel and C. Yagüe. Intermittency of turbulence in the atmospheric boundary layer: Scaling exponents and stratification influence. *Boundary-Layer Meteorology*, 140(1):73–85, 2011.

[266] J. M. Vindel, C. Yagüe, and J. M. Redondo. Relationship between intermittency and stratification. *Nuovo Cimento C. Geophysical Space Physics*, 31:669–678, 2008.

[267] J. M. Vindel, C. Yagüe, and J. M. Redondo. Structure function analysis and intermittency in the atmospheric boundary layer. *Nonlinear Processes in Geophysics*, 15(6):915–929, 2008.

[268] L-P. Wang, S. Chen, J. G. Brasseur, and J. C. Wyngaard. Examination of hypotheses in the Kolmogorov refined turbulence theory through high-resolution simulations. part 1. velocity field. *Journal of Fluid Mechanics*, 309:113–156, 1996.

[269] Z. Warhaft. Passive scalars in turbulent flows. *Annual Review of Fluid Mechanics*, 32(1):203–240, 2000.

[270] D. S. Wilks. *Statistical methods in the atmospheric sciences*, volume 100. Academic press, 2011.

[271] O. Williams, T. Hohman, T. Van Buren, E. Bou-Zeid, and A. J. Smits. The effect of stable thermal stratification on turbulent boundary layer statistics. *Journal of Fluid Mechanics*, 812:1039–1075, 2017.

[272] A. R. Wittwer, G. S. Welter, and A. M. Loredo-Souza. Statistical analysis of wind tunnel and atmospheric boundary layer turbulent flows. In *Wind Tunnel Designs and Their Diverse Engineering Applications*. INTECH, 2013.

[273] J. C. Wyngaard and H. Tennekes. Measurements of the small-scale structure of turbulence at moderate Reynolds numbers. *Physics of Fluids*, 13(8):1962–1969, 1970.

[274] B. Xie, J. C . H. Fung, A. Chan, and A. Lau. Evaluation of nonlocal and local planetary boundary layer schemes in the WRF model. *Journal of Geophysical Research: Atmospheres*, 117(D12), 2012.

[275] V. Yakhot. Kolmogorov law for statistically stationary turbulence: Application to high-Rayleigh-number Bénard convection. *Physical Review Letters*, 69(5):769, 1992.

[276] V. Yakhot. Mean-field approximation and a small parameter in turbulence theory. *Physical Review E*, 63(2):026307, 2001.

[277] Q. Yang, L. K. Berg, M. Pekour, J. D. Fast, R. K. Newsom, M. Stoelinga, and C. Finley. Evaluation of WRF-predicted near-hub-height winds and ramp events over a Pacific Northwest site with complex terrain. *Journal of Applied Meteorology and Climatology*, 52(8):1753–1763, 2013.

[278] J. Yoon and S. Song. A numerical study of adjusted parameter estimation in normal inverse Gaussian distribution. *Korean Journal of Applied Statistics*, 29(4):741–752, 2016.

[279] H. Zhang, Z. Pu, and X. Zhang. Examination of errors in near-surface temperature and wind from wrf numerical simulations in regions of complex terrain. *Weather and Forecasting*, 28(3):893–914, 2013.

[280] H. Zheng and A. Kusiak. Prediction of wind farm power ramp rates: A data-mining approach. *Journal of Solar Energy Engineering*, 131(3):031011, 2009.

[281] S. S. Zilitinkevich and J. W. Deardorff. Similarity theory for the planetary boundary layer of time-dependent height. *Journal of the Atmospheric Sciences*, 31(5):1449–1452, 1974.

[282] G. K. Zipf. *Human Behavior and the Principle of Least Effort: an Introduction to Human Ecology.* Addison-Wesley Press, Cambridge, 1949.

[283] L. Zubair. *Studies in Turbulence using Wavelet Transforms for Data Compression and Scale Separation.* PhD thesis, Yale University, 1993.

APPENDIX

Appendix A

Computational Codes

A.1 Normal Inverse Gaussian Distribution (NIG)

Estimating NIG parameters using method of moments estimation (MME) from sample moments, see Eq. 3.3:

```
%Method of Moments Estimation of Parameters for NIG Distribution
%Written by Adam DeMarco: 4/15/2016

function [alpha, beta, mu, delta, zeta, chi] = nigpar(m1,m2,m3,m4)

%Input:
%m1 = sample mean
%m2 = sample stdev
%m3 = sample skewness
%m4 = sample kurtosis

%Output:
%Method of Moment estimation of alpha, beta, mu, delta, zeta, and chi
%parameters for NIG

%Reference: Karis (2002; Statistics & Probability Letters; see page 46)
%Similar notation appears in Aas and Haff (2006) NIG and Generalized Hyperbolic
%Skew Student's t document (page 8). I am using their notation

m4x   = m4 - 3;  %Flatness (a.k.a. excess kurtosis)

if m4x > (5/3)*m3^2
    gam   = 3/(m2*sqrt(3*m4x - 5*m3^2));
    beta  = (m3*m2*(gam^2))/3;
```

```
    delta = (m2^2)*(gam^3)/(beta^2+gam^2);
    mu    = m1 - beta*delta/gam;
    alpha = sqrt(gam^2 + beta^2);
    zeta  = (1 + delta*sqrt(alpha^2 - beta^2))^(-1/2);
    chi   = beta*zeta/alpha;

else
    beta  = NaN;
    delta = NaN;
    mu    = NaN;
    alpha = NaN;
    zeta  = NaN;
    chi   = NaN;
end
```

Estimating NIG parameters using maximum likelihood estimation (MLE) from MME:

```
%Maximum Likelihood Estimation (MLE) of NIG parameters
%Written by Adam DeMarco: 4/15/2016

function [alpha, beta, mu, delta, zeta, chi] = nigpar_mle(x,alpha,beta,mu,delta)

% Inputs: the increment series (x), alpha, beta, mu, and delta obtained
% Method of Moments.
% Outputs: alpha, beta, mu, delta, zeta, and chi estimated using MLE

lb = [0 -Inf -Inf 0];
ub = [Inf Inf Inf Inf];

if isnan(alpha) == 0 & isnan(beta) == 0 & isnan(mu) == 0 & isnan(delta) == 0
    %TolFun: Termination tolerance for the objective function value.
    %TolX: Termination tolerance for the parameters.
    options = statset('MaxIter',1000, 'MaxFunEvals',1000,...
        'TolFun',1e-20,'TolX',1e-20,'Display','final');

    start = [alpha,beta,mu,delta];
    phat  = mle(x, 'pdf',@nigpdf, 'start',start,...
        'lower',lb, 'upper',ub, 'options',options);

    alpha = phat(1);
    beta  = phat(2);
    mu    = phat(3);
    delta = phat(4);

    zeta  = (1 + delta*sqrt(alpha^2 - beta^2))^(-1/2);
    chi   = beta*zeta/alpha;
else
    alpha = NaN;
    beta  = NaN;
```

```
    mu    = NaN;
    delta = NaN;
    zeta  = NaN;
    chi   = NaN;
end
```

Calculating probability density function (pdf) of NIG using model parameters:

```
%Probability Distribution Function (PDF) of NIG
%Written by Adam DeMarco: 4/15/2016

function pdf = nigpdf(x,alpha,beta,mu,delta)

% Inputs: set of values from increments to be evaluated (x)
% alpha, beta, mu, and delta from Maximum Likelihood estimations
% Output: pdf of NIG
% Reference: Scott et al. (2011) Moments of the generalized hyperbolic distribution
% Computational Statistics

if isnan(alpha) == 0 & isnan(beta) == 0 & isnan(mu) == 0 & isnan(delta) == 0

    T1  = (alpha/pi)*exp(delta*sqrt(alpha^2-beta^2) - beta*mu);
    phi = 1 + ((x-mu)/delta).^2;
    T2  = phi.^(-1/2);
    T3  = besselk(1, (delta*alpha*(phi.^(1/2)) ));
    T4  = exp(beta*x);
    pdf = T1.*T2.*T3.*T4;

else
    pdf = NaN*x;
end
```

Calculating cumulative distribution function (cdf) of NIG using model parameters:

```
%Cumulative Distribution Function generation of NIG
%Written by Adam DeMarco: 12/20/2016
%This code is based on a code from FileExchange

function F = nigcdf(x,alpha,beta,mu,delta)

% Inputs: the increment series (x), alpha, beta, mu, and delta obtained MLE
% Output: CDF of NIG using integration (F)

N    = length(x);
F    = zeros(N,1);

pdfHandle = @(x) nigpdf(x,alpha,beta,mu,delta);

if beta>=0
```

```
F(1) = quadgk(pdfHandle,-inf,x(1));
else
F(1) = quadgk(pdfHandle,-100,x(1));
end

if isnan(F(1))==1

    return;
end
for k = 2 : N
    F(k) = F(k-1) + quadgk(pdfHandle,x(k-1),x(k));
    if mod(k,1000) == 0
        fprintf('%d\t%d\n',k,N);
    end
end
```

Calculating Inverse cdf of NIG using model parameters, used for generating Q-Q plots:

```
%Inverse CDF of NIG
%Written by Adam DeMarco: 06/14/2017
function Xi = invnigcdf(Fi,alpha,beta,mu,delta)

% INPUT: Fi - EDF from obtained from observation data
% The NIG parameters - alpha, beta, mu and delta
% OUTPUT: Xi - The x values evaluating the NIG CDF estimation
% determined from spline interpolation

x = -25:0.001:25;
N = length(x);

pdfHandle = @(x) nigpdf(x,alpha,beta,mu,delta);
F(1) = quadgk(pdfHandle,-inf,x(1));

for k = 2 : N
    F(k) = F(k-1) + quadgk(pdfHandle,x(k-1),x(k));
    if mod(k,1000) == 0
        fprintf('%d\t%d\n',k,N);
    end
end

[F,indx] = unique(F);
x        = x(indx);
clear indx;

indx = find(isnan(F)==1); F(indx) = []; x(indx) = [];
clear indx;
indx = find(isinf(F)==1); F(indx) = []; x(indx) = [];
clear indx;
```

```
Xi = interp1(F,x,Fi,'spline',NaN);
```

Calculating NIG parameters using Expectation-Maximization code as described in ALGORITHM 3.1:

```
%NIG Parameter Estimation by Expectation-Maximization
%Following Karlis (2002)
%Written by Adam DeMarco: 1/15/2017

function [alpha, beta, mu, delta, zeta, chi] = nigpar_EM(x,alpha,beta,mu,delta)

% INPUT: alpha, beta, mu, delta obtained from MME, x is increment series
% OUTPUT: NIG parameters estimated using Expectation-Maximization

n    = length(x);
Lold = -Inf;

for k = 1:10^6

    %Expectation Step-----------------------------------------------------
    phi = 1 + ((x-mu)/delta).^2;
    T1  = (delta/alpha)*sqrt(phi);
    T2  = besselk(0, (delta*alpha*(phi.^(1/2)) ))./besselk(1,
        (delta*alpha*(phi.^(1/2)) ));
    S   = T1.*T2;

    T3  = (alpha/delta)*(1./sqrt(phi));
    T4  = besselk(-2, (delta*alpha*(phi.^(1/2)) ))./besselk(-1,
        (delta*alpha*(phi.^(1/2)) ));
    W   = T3.*T4;
    %---------------------------------------------------------------------

    %Maximization Step----------------------------------------------------
    M   = mean(S);
    Lam = 1/mean(W - 1/M);

    delta = sqrt(Lam);
    gamma = delta/M;
    Num   = sum(x.*W) - mean(x)*sum(W);
    Den   = n - mean(S)*sum(W);
    beta  = Num/Den;
    mu    = mean(x) - beta*mean(S);
    alpha = sqrt(gamma^2 + beta^2);
    zeta  = (1 + delta*sqrt(alpha^2 - beta^2))^(-1/2);
    chi   = beta*zeta/alpha;
    %---------------------------------------------------------------------

    %Compute Log-Likelihood-----------------------------------------------
    T1    = -n*log(pi);
```

```
    T2    = n*log(alpha);
    T3    = n*(delta*gamma - beta*mu);
    T4    = -0.5*sum(log(phi));
    T5    = beta*sum(x);
    T6    = sum(besselk(1, (delta*alpha*(phi.^(1/2)) )));
    L     = (T1+T2+T3+T4+T5+T6); %Log-Likelihood function
    %-----------------------------------------------------------------------

    if abs((L-Lold)*100/Lold) < 1e-5
       fprintf('%s%d\n','Number of Iterations:',k);
       break;
    else
       Lold = L;
    fprintf('%d\t%f\t%f\t%f\t%f\t%f\n',k,alpha,beta,mu,delta,L);
    end
end
```

A.2 Generalized Hyperbolic Skewed Student t's Distribution (GHSST)

Estimating GHSST parameters using MLE from Eq. 3.6 assuming $\beta = 0$:

```
%Maximum Likelihood Estimation (MLE) of NIG parameters
%Written by Adam DeMarco: 1/30/2017

function [nu, beta, mu, delta] = ghsstpar_mle(x,nu,beta,mu,delta)

% Inputs: the increment series (x), nu, beta, mu, and delta obtained
% sample moments.
% Outputs: nu, beta, mu, and delta estimated using MLE

lb = [0,-Inf,-Inf,0];
ub = [Inf,Inf,Inf,Inf];

if isnan(nu) == 0 & isnan(beta) == 0 & isnan(mu) == 0 & isnan(delta) == 0
    %TolFun: Termination tolerance for the objective function value.
    %TolX: Termination tolerance for the parameters.
    options = statset('MaxIter',1000, 'MaxFunEvals',1000,...
       'TolFun',1e-20,'TolX',1e-20,'Display','final');

    start = [nu,beta,mu,delta];

    [phat,pci] = mle(x,'pdf',@ghsstpdf,'start',start,...
       'lower',lb, 'upper',ub,'options',options,'optimFun','fminsearch');
```

```
    nu    = phat(1);
    beta  = phat(2);
    mu    = phat(3);
    delta = phat(4);

else
    nu    = NaN;
    beta  = NaN;
    mu    = NaN;
    delta = NaN;
end
```

Calculating pdf of GHSST using model parameters:

```
%PDF of Generalized Hyperbolic Skew Student's t-Distribution (GHSST)
%Written by Adam DeMarco: 1/30/2017

function pdf = ghsstpdf(x,nu,beta,mu,delta)

% Inputs: set of values from increments to be evaluated (x)
% nu, beta, mu, and delta from Maximum Likelihood estimations
% Output: pdf of GHSST

%Reference: Aas and Haff (2006; Journal of Financial Econometrics)
%NOTE: Aas and Haff has a typo in Bessel function (negative sign was
%missing); I can verify it i 2 ways: (a) compare with Eq. (3) of their
%paper and (b) Eq. (9.46) of Paolella's paper.

if isnan(nu) == 0 & isnan(beta) == 0 & isnan(mu) == 0 & isnan(delta) == 0

    T1  = 2^((1-nu)/2) * (delta^nu) * ((abs(beta))^((nu+1)/2));
    T2  = besselk(-(nu+1)/2,sqrt(beta^2*(delta^2 + (x-mu).^2)));
    T3  = exp(beta*(x-mu));
    T4  = gamma(nu/2)*sqrt(pi);
    T5  = (sqrt(delta^2 + (x-mu).^2)).^((nu+1)/2);
    pdf = (T1.*T2.*T3)./(T4.*T5);
else
    pdf = NaN*x;
end
```

Calculating cdf of GHSST using model parameters:

```
%Cumulative Distribution Function generation of GHSST
%Written by Adam DeMarco: 12/20/2016
%This code is based on a code from FileExchange

function F = ghsstcdf(x,nu,beta,mu,delta)
```

```
% Inputs: the increment series (x), nu, beta, mu, and delta obtained MLE
% Output: CDF of GHSST using integration (F)

N    = length(x);
F    = zeros(N,1);

pdfHandle = @(x) ghsstpdf(x,nu,beta,mu,delta);

if beta>0
F(1) = quadgk(pdfHandle,-inf,x(1));
elseif beta<=0
F(1) = quadgk(pdfHandle,-100,x(1));
end

for k = 2 : N
    F(k) = F(k-1) + quadgk(pdfHandle,x(k-1),x(k));
    if mod(k,1000) == 0
        fprintf('%d\t%d\n',k,N);
    end
end
```

A.3 Variance Gamma (VG)

Estimating VG parameters using MME from sample moments, see Eq. 3.9:

```
%Method of Moments Estimation of Parameters for VG Distribution
%Written by Adam DeMarco: 1/24/2017

function [alpha, beta, mu, lambda] = vgpar(m1,m2,m3,m4)

%Inputs:
%m1 = sample mean
%m2 = sample stdev
%m3 = sample skewness
%m4 = sample kurtosis;

%NOTE: These are "approximate" estimates. They are useful as a
%starting condition for MLE.

%Output:
%Method of Moment estimation of alpha, beta, mu, lambda parameters for VG

%I used Paolella's chapter and Seneta (2004). NOTE: They use different
%notations.

m2s    = m2^2;     %Stdev -> Variance
```

```
m4x    = m4 - 3; %Flatness
lambda = 3/(m4x);

psi    = 2*lambda/m2s;

theta = sqrt(m2s)*m3*lambda/3;
beta  = m3*lambda/(3*sqrt(m2s));

alpha = sqrt(psi + beta^2);

mu     = m1 - theta;
```

Estimating VG parameters using MLE from MME:

```
%Maximum Likelihood Estimation (MLE) of VG parameters
%Written by Adam DeMarco: 1/27/2017

function [alpha, beta, mu, lambda] = vgpar_mle(x,alpha,beta,mu,lambda)

% Inputs: the increment series (x), alpha, beta, mu, and lambda obtained
% Method of Moments.
% Outputs: alpha, beta, mu, and lambda estimated using MLE

lb = [0,-Inf,-Inf,0];
ub = [Inf,Inf,Inf,Inf];

if isnan(alpha) == 0 & isnan(beta) == 0 & isnan(mu) == 0 & isnan(lambda) == 0
   %TolFun: Termination tolerance for the objective function value.
   %TolX: Termination tolerance for the parameters.

   options = statset('MaxIter',1000, 'MaxFunEvals',1000,...
       'TolFun',1e-20,'TolX',1e-20,'Display','final');

   start = [alpha,beta,mu,lambda];

   [phat,pci] = mle(x,'pdf',@vgpdf,'start',start,...
       'lower',lb, 'upper',ub,'options',options,'optimFun','fminsearch');

   alpha = phat(1);
   beta  = phat(2);
   mu    = phat(3);
   lambda= phat(4);

else
   alpha = NaN;
   beta  = NaN;
   mu    = NaN;
   lambda= NaN;
```

```
end
```

Calculating pdf of VG using model parameters:

```
%Probability Distribution Function (PDF) of VG
%Written by Adam DeMarco: 1/27/2017

function pdf = vgpdf(x,alpha,beta,mu,lambda)

% Inputs: set of values from increments to be evaluated (x)
% alpha, beta, mu, and delta from Maximum Likelihood estimations
% Output: pdf of NIG
% Reference: Scott et al. (2011) Moments of the generalized hyperbolic distribution
% Computational Statistics

if isnan(alpha) == 0 & isnan(beta) == 0 & isnan(mu) == 0 & isnan(lambda) == 0

    T1 = 2*((alpha^2 - beta^2)/2)^lambda;
    T2 = sqrt(2*pi)*gamma(lambda);
    T3 = (abs(x-mu)/alpha).^(lambda-0.5);
    T4 = besselk(lambda-0.5, (alpha*abs(x-mu)));
    T5 = exp(beta*(x-mu));
    pdf = (T1./T2).*T3.*T4.*T5;

else
    pdf = NaN*x;
end
```

Calculating cdf of VG using model parameters:

```
%Cumulative Distribution Function generation of VG
%Written by Adam DeMarco: 1/24/2017
%This code is based on a code from FileExchange

function F = vgcdf(x,alpha,beta,mu,lambda)

% Inputs: the increment series (x), alpha, beta, mu, and lambda obtained MLE
% Output: CDF of VG using integration (F)

N    = length(x);
F    = zeros(N,1);

pdfHandle = @(x) vgpdf(x,alpha,beta,mu,lambda);

if beta>0
F(1) = quadgk(pdfHandle,-inf,x(1));
elseif beta<=0
F(1) = quadgk(pdfHandle,-100,x(1));
end
```

```
for k = 2 : N
    F(k) = F(k-1) + quadgk(pdfHandle,x(k-1),x(k));
    if mod(k,1000) == 0
        fprintf('%d\t%d\n',k,N);
    end
end
```

A.4 Lognormal Superstatistics (LNSS)

Calculating pdf of LNSS using MME obtained from Eq. 3.11:

```
%This code generates LN superstatistics pdf using Eq. (15) of Beck (2004)
%Written by Adam DeMarco: 12/20/2016

function p = LN_Superstatistics(du_range,s,mu)

% Inputs:
% du_range=the histogram binned range of the increments
% mu and s are the two parameters describing the distribution
% Output: PDF of LNSS

N=length(du_range);

for k = 1:length(du_range)
    B   = [1e-3:1e-3:1e3];
    Y   = (1/(2*pi*s))*(B.^(-1/2)).*exp(- (log(B/mu)).^2 / (2*s^2))
         .*exp(-1/2*B*du_range(k)^2);
    p(k) = trapz(B,Y);
    if mod(k,100)==0
    fprintf('%d\t%d\n',k,N);
    end
end
```

A.5 Hill Estimator

Calculating Hill plots which depict the tail-index for a given distribution, Eq. 3.18:

```
%Hill Estimator
%Written by Adam DeMarco: 12/23/2016

function [muH] = Hill_Estimator(X,Kmax)
```

```
%Input: X = Increment series, Kmax = sample fraction typically 3000
%Note: Kmax >= 2

Indx = find(isnan(X) == 1); X(Indx) = [];
Xs = sort(X,'descend');

%------------------------------------------------------------------------

den1 = cumsum(log(Xs(1:Kmax-1)));
den2 = ([2:Kmax]' - 1).*log(Xs(2:Kmax));
den  = den1-den2;
num  = [1:Kmax-1].';
muH  = num./den;
%-------------------------------------------------------------------------
```

A.6 Probability of Exceedance

Calculating exceedance probability for a range of σ (standard deviations) from 1-10 for the left and right tails for a given increment series:

```
function [Fx,Fy] = ComputeExceedance(x,F)
%Written by Adam DeMarco: 6/13/2017
% INPUT: x: the values for which the empirical cdf (F) was evaluated.
%        F: the vector of values of the empirical cdf.
% OUPUT: Fx: the exceedance probability of ramp-up or right tail in %
%        Fy: the same as Fx, except for ramp-down or left tail.

y1 = [1:10]; y2 = [-1:-1:-10];

Fx = zeros(size(y1));
for i = 1:numel(y1)
    Indx = find(x>y1(i));
    if numel(Indx) >= 1
        Fx(i) = (1-F(Indx(1)))*100;
    else
        Fx(i) = NaN;
    end
end

Fy = zeros(size(y2));
for i = 1:numel(y2)
    Indx = find(x<y2(i));
    if numel(Indx) >= 1
        Fy(i) = F(Indx(end))*100;
```

```
    else
        Fy(i) = NaN;
    end
end
```

A.7 Anderson-Darling Test Statistic

Integral form of the Anderson-Darling statistic as shown in Eq. 5.3:

```
function [AD,N] = AndersonDarlingIntegral(x,Fn,F)
% Written by: Adam DeMarco: 4/15/2017
% INPUT
% Fn = empirical distribution function (EDF)
% F = theoretical cumulative distribution function (CDF)
% x = the value which the EDF was evaluated
% OUTPUT
% AD = Anderson Darling statistics
% N = length of x
% Luceno 2006 - Fitting the GP distribution to data using maximum GOF estimators

T1 = (Fn - F).^2;
T2 = F.*(1-F);
T  = T1./T2;

N  = length(x);

AD = N*trapz(x,T);
clear T1 T2
```

Integral form of the Anderson-Darling statistic, separately evaluating the left and right tails as shown in Eq. 8.1:

```
function [ADL,ADR,N] = AndersonDarlingIntegral_tails(x,Fn,F)

% Written by: Adam DeMarco: 5/28/17
% Fn is the EDF
% F is the true CDF
% ADL - Anderson-Darling Left tail
% ADR - Anderson-Darling Right tail
% Luceno 2006 - Fitting the GP distribution to data using maximum GOF estimators

T1 = (Fn - F).^2;
T2 = (1-F);
T  = T1./T2;

N  = length(x);
```

```
ADR = N*trapz(x,T);

clear T1 T2 T

T1 = (Fn - F).^2;
T2 = F;
T  = T1./T2;

ADL = N*trapz(x,T);
clear T1 T2 T
```

Two-sample Anderson-Darling statistic, taking the “ties” into consideration, which are a direct result of discrete parent populations from ([220], Eq. 6), see Eq. 10.5:

```
function [AD] = adtest2_ties(X,Y)
%2-sample Anderson Darling test statistic
%Takes into account the issue of "ties" (i.e., duplicate CDF values)
%Written by: Adam DeMarco: 6/30/2017
%Based on Eq. (6) of Scholz and Stephens (1987)
% INPUT:
% X and Y are the two increment series for comparsion
% OUTPUT:
% AD is the 2-sample Anderson Darling test statistic

[NR,NC] = size(X); if NC > 1; X = X'; end;
[NR,NC] = size(Y); if NC > 1; Y = Y'; end;

n1 = length(X); n2 = length(Y); N = n1 + n2;

Z  = [X;Y];
Zs = unique(Z); %Automatically sorts in ascending order

f1 = hist(X,Zs); M1 = cumsum(f1);
f2 = hist(Y,Zs); M2 = cumsum(f2);
l  = hist(Z,Zs); B = cumsum(l);

%These terms are computed for j = 1:L-1
den = B.*(N-B); den(end) = [];
num1 = l.*(N*M1 - n1*B).^2; num1(end) = [];
num2 = l.*(N*M2 - n2*B).^2; num2(end) = [];

T1   = 1/(n1*N)*sum(num1./den);
T2   = 1/(n2*N)*sum(num2./den);

AD   = T1+T2;
```