

ÖREBRO UNIVERSITY

MASTER'S THESIS

Mobile Robot Wind Mapping

Author:

Aidin HASSANZADEH

Supervisor:

Dr. Erik SCHAFFERNICHT

Mobile Robotics and Olfaction Lab
School of Science and Technology

April 2014

Declaration of Authorship

I, Aidin HASSANZADEH, declare that this thesis titled, 'Mobile Robot Wind Mapping' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

“To my parents and Sahar”

Aidin Hassanzadeh

ÖREBRO UNIVERSITY

Abstract

Faculty Name

School of Science and Technology

Master's of Science

Mobile Robot Wind Mapping

by Aidin HASSANZADEH

Statistical gas distribution mapping has recently become a prominent research area in the robotics community. Gas distribution mapping using mobile robots aims for building map of gas dispersion in an unknown environment using the sampled gas concentrations accompanied by the corresponding atmospheric variables. In this context, wind is considered as one of the main driving forces and recently exploited as an environmental bias in the the modelling process. However, the existing approaches utilizing the wind data are based on very simple averaging window methods which do not take the specific spatio-temporal wind variations into account appropriately.

In the current thesis work, under the heading of statistical wind modelling, the various aspects of the existing approaches to model both temporal and spatial wind variations are studied. Accordingly, in the undertaking of *Mobile Robot Wind Mapping (MRWM)* task, three individual methods for statistically *wind speed modelling*, *wind direction modelling* and *spatial wind mapping* are proposed and implemented.

Particularly, wind speed is modelled in form of a Gaussian distribution where the valid averaging scale is defined using an online adaptive approach, namely *Time-Dependent Memory Method (TDMM)*. The wind direction is modelled by means of the mixture-model of Von-Mises distribution and for the spatial mapping of modelled wind data, a recursive approach based on Linear Kalman filter is utilized. The proposed approaches for statistically wind speed and direction modelling are applied to and evaluated by real wind data, collected specifically for this project. The wind mapping algorithm is implemented and tested using simulated data.

Acknowledgements

I would like to thank my supervisor Erik Schaffernicht for the useful comments, remarks and engagement through the learning process of this master thesis.

Contents

Declaration of Authorship	i
Abstract	iii
Acknowledgements	iv
Contents	v
List of Figures	vii
List of Tables	viii
1 Introduction	1
1.1 Problem Formulation	1
1.2 Task Description	3
1.3 Methodology	3
1.4 Wind Data	4
1.5 Thesis Outline	4
2 The Wind Phenomenon	6
2.1 The Wind Theory	6
2.1.1 Origins of Wind	6
2.1.2 Wind Temporal Variations	6
2.1.3 Wind Spatial Variations	8
2.2 Wind Descriptors	10
2.2.1 Wind Speed	11
2.2.1.1 Speed Mean	11
2.2.1.2 Wind Speed Frequency-Distributions	11
2.2.1.3 Turbulence Fluctuations	14
2.2.1.4 Turbulence Intensity	16
2.2.1.5 Wind Speed Increments	16
2.2.2 Wind Direction	16
2.2.2.1 Direction Mean	17
2.2.2.2 Trigonometric Moments	17
2.2.2.3 Trigonometric Moments about Mean Direction	18

2.2.2.4	Wind Direction Frequency-Distributions	18
3	Statistical Wind Speed Modelling	20
3.1	Background	20
3.2	Requirements	21
3.3	Methodology	22
3.3.1	Theory	22
3.3.2	Practical Considerations	24
3.4	Experimental Results	26
4	Statistical Wind Direction Modelling	33
4.1	Background	33
4.2	Requirements	34
4.3	Methodology	36
4.4	Experimental Results	39
4.4.1	Wind Direction Analysis	39
4.4.2	Wind Direction Modelling	42
5	Wind Mapping	49
5.1	Overview	49
5.1.1	Background	49
5.1.2	Requirements	51
5.2	Methodology	52
5.2.1	Problem Formulation	52
5.2.2	Kalman Filter	55
5.2.3	Practical Considerations	56
5.3	Experimental Results	57
6	Conclusions	64
6.1	Summary	64
6.2	Future Work	65
A	ORU.WIND Data Specification	67
B	Wind Speed Modelling Results	69
C	Wind Direction Modelling Results	74
	Bibliography	79

List of Figures

2.1	Vad der Hoven wind spectrum	8
2.2	Spatial vs temporal wind variability	10
2.3	Weibull distribution	12
2.4	Rayleigh distribution	13
2.5	Inverse-Gaussian distribution	14
2.6	Generalized Extreme Value distribution	15
2.7	Von-Mises distribution	19
3.1	Empirical density vs Estimated distribution - ORU.WIND.OS subjected to TDMM	28
3.2	Empirical density vs Estimated distribution - ORU.WIND.OC subjected to TDMM	29
3.3	TDMM failure to detect averaging scale	30
3.4	Box-whisker plots of the NRMSE and R-squared values for estimated Gaussian models	31
4.1	Wind rose	34
4.2	Histograms of ORU.WIND outdoor wind direction time series	40
4.3	Histograms of ORU.WIND indoor wind direction time series	41
4.4	Concentration parameter κ vs averaging interval length ΔT	42
4.5	Concentration parameter κ vs 40-seconds wind speed averages.	43
4.6	Estimated Von-Mises mixture distribution model vs empirical frequency distribution at ORU.WIND.OC.100 Segment 54	45
4.7	Estimated Von-Mises mixture distribution model vs empirical frequency distribution at ORU.WIND.OC.100 Segment 13	46
4.8	Box-whisker plots of the NRMSE and R-squared values for estimated Von-Mises mixture models	47
5.1	Graphical model for stochastic mapping process	53
5.2	The plots showing the mapping results with observation step 100cm for Gaussian data	60
5.3	The plots showing the mapping results with observation step 200cm for Gaussian data	61
5.4	The plots showing the mapping results with observation step 100cm for turbulent data	62
5.5	The plots showing the mapping results with observation step 200cm for turbulent data	63

List of Tables

1.1	ORU.WIND data statistics	4
2.1	Atmospheric model classifications	9
2.2	Structure of Atmospheric Boundary Layer	10
3.1	The sequence of wind speed and corresponding instantaneous statistics . .	25
3.2	Statistical moments and errors for ORU.WIND.OS/OC subject to wind speed density estimation	28
3.3	The overall performance of estimated Gaussian models - NRMSE	31
3.4	The overall performance of estimated Gaussian models - R^2	32
4.1	Statistical moments and errors for ORU.WIND.OC.100 segment 13 at different numbers of sectors	45
4.2	The overall performance of estimated Von-Mises mixture models - NRMSE	47
4.3	The overall performance of estimated Von-Mises mixture models - R^2 . .	48
5.1	The performance of the mapping process with respect to different initial- ization - GRF with Gaussian correlations	58
5.2	The performance of the mapping process with respect to different initial- ization - GRF with turbulent correlations	58
A.1	ORU.WIND Data Measurement Specification	68
B.1	Statistical moments and errors for ORU.WIND.OS.100 subject to wind speed density estimation	70
B.2	Statistical moments and errors for ORU.WIND.OS.150 subject to wind speed density estimation	71
B.3	Statistical moments and errors for ORU.WIND.OC.100 subject to wind speed density estimation	72
B.4	Statistical moments and errors for ORU.WIND.OC.150 subject to wind speed density estimation	73
C.1	Performance statistics for ORU.WIND.OS.100 subject to wind direction density estimation	75
C.2	Performance statistics for ORU.WIND.OS.150 subject to wind direction density estimation	76
C.3	Performance statistics for ORU.WIND.OC.100 subject to wind direction density estimation	77
C.4	Performance statistics for ORU.WIND.OC.150 subject to wind direction density estimation	78

Chapter 1

Introduction

The aim of the current work presented in this thesis is to create a statistical framework that could be utilized in mobile robots for wind mapping purposes. To a great extent, such model not only is required to encapsulate the intrinsic uncertainties of wind in time and space domains, but also should cope with the specific requirements of a typical mobile robot mapping task. The purpose of this chapter is to define the problem formulation and the research objectives the subject MRWM task, and briefly describes the relevant methodology and the thesis structure.

1.1 Problem Formulation

Gas Distribution Modelling (GDM) refers to the task of mapping gas concentrations in the environment while aiming to interpret the spatio-temporal distributed measurements, by exploiting gas concentrations and the related atmospheric variables, as accurate as possible. Thus far, the analytical methods, importantly Computational Fluid Dynamics (CFD), have been exploited to resolve the problem of modelling gas distribution. The analytical gas distribution methods utilize mathematical equations governing atmosphere, dispersion, chemical and physical process. Although analytical models in GDM provide valuable results, they are computationally intense and strongly controlled by underlying environmental assumptions. That is, the traditional numerical mapping methods become inapplicable when employed to build high resolution gas distribution maps in real time and within unknown environments, which is the case in real world problems.

The more recent method for building map of gas distributions is Statistical Gas Distribution Modelling (SGDM). SGDM takes gas concentrations as random variables and

approaches gas distribution mapping as statistical density estimation problem. To build gas distribution maps statistically, several methods have been proposed. Based on the type of statistics involved in the estimation problem, the SGDM may be categorized into two groups. I) The methods that only use the first moment statistics or extreme values of gas concentration data (e.g. [1] and [2]) II) The methods which incorporate both first and the second moment statistics, mean and variance such as efforts in [3], [4] and [5].

Apart from the above discussed SGDM approaches which basically only rely on gas concentration measurements, there are a few methods which besides gas concentrations also incorporate local wind data. To be specific, in DM+V/W method [6] 30-seconds wind averaging scheme is used by which the shape of the kernel function is specified. Similarly, in [7] the average wind vector along DM+V/W was used, though, a method of wind measurement using microdrone was introduced and experimented. Hernandez et al [8] uses the 30-second wind average vector in its arguments against biometric gas source localization methods. In general, the methods previously utilized the wind information in SGDM were only based on a fixed length simple averaging window, but other crucial parameter, such as the valid averaging length and the higher statistical moments have been neglected so far.

Considering the chaotic nature of wind, it is apparent that plugging a simple wind average vector into SGDM could not lead to factual results. In fact, SGDM methods utilizing wind information suffer from three major downsides. First, the wind is averaged as per a fixed length averaging wind, in which the validity averaging scale has not been considered. Second, they have only made use of wind average vector, but not considering any higher statistical moments. Third, the spatial variability of wind data has not been taken into account while applying to gas distribution modelling.

In order to have a viable SGDM, a truthful description of wind data is vital. Uncertainties in wind speed and direction degrades the performance of gas distribution mapping. The effect of wind and its specific characteristics, wind turbulence, on gas dispersion has been studied with intense consideration. The wind and wind turbulence are taken as key parameters in the existing atmospheric dispersion models by which gas plume dilution and diffusion are affected. The emission concentration is inversely proportional to the wind speed. This fact can easily be seen in one of simplest form of dispersion modelling at ground level, the Gaussian dispersion model:

$$\chi(x, y, 0; H) = \frac{Q}{\sigma_y \sigma_z \mu} \exp\left[-\frac{1}{2}\left(\frac{y}{\sigma_y}\right)^2\right] \exp\left[-\frac{1}{2}\left(\frac{H}{\sigma_z}\right)^2\right] \quad (1.1)$$

Here, without going to the every paramter invovled, the concentration is given by χ and the wind speed by μ . Indeed, Gas plume diffusion is affected by wind turbulence.

The eddy diffusion coefficient is proportional to the product of wind velocity and the function turbulence which was encapsulated by the standard deviations of the emission distribution, namely σ_y and σ_z .

That is, wind is one of the primary environmental variables influencing gas dispersion and accordingly gas distribution modelling as well. As mentioned above, the wind information incorporated into the existing gas distribution methods, is limited to first moment statistics only, and its spatial variability has not been considered either. It is vital to collect and provide wind information in more trustful manner wind innate characteristics that could lead a reliable distribution model. This thesis project aims at investigating statistical properties and methods for wind modelling, owing to provide a trustworthy wind information to plug into existing statistical gas distribution mapping techniques.

1.2 Task Description

The main goal of this thesis work is to develop a framework for Mobile Robot Wind Mapping (MRWM), which first is capable to model wind and its natural fluctuations in space and time at fairly high resolution, and second does suit the specific requirements of the mobile robot gas distribution task.

To accomplish this, three ordered subtasks are addressed. Initially, the state-of-art techniques, specifically in meteorology and wind power, for modelling wind and its temporal variability are explored and studied. The method appropriate to a typical statistical gas distribution modelling task is selected and implemented using the real near-surface wind data as per the specific condition of wind measurements with mobile robots . Next, as per the wind formulation, the framework for building wind grid maps in space is formulated and implemented. Specific attention is paid at existing approaches in wind modelling and mapping available in the fields of mobile robot olfaction and meteorology. Finally, the constructed wind map is put under evaluation using simulated data.

1.3 Methodology

The main objective the current thesis work is to construct a wind grid mapping method that's reliable and robust to address the temporal and spatial uncertainties which are typical to the statistical gas distribution framework. The modelling was implemented on an offline wind dataset measured by a portable anemometer at Orebro university

WIND.ORU.	OS.100	OS.150	OC.100	OC.150	IL.100	IL.150	IH.100	IH.150
N	371087	352826	362680	351936	349757	347985	352936	355683
\bar{u} [ms^{-1}]	1.13	1.40	1.04	0.98	0.037	0.03	0.10	0.15
σ_u [ms^{-1}]	0.61	0.68	0.51	0.53	0.026	0.02	0.05	0.09
min_u [ms^{-1}]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
max_u [ms^{-1}]	4.31	4.66	3.92	3.74	0.23	0.37	0.96	1.08
$\bar{\theta}$ [$^{\circ}$]	152.97	213.79	222.01	144.37	67.04	30.52	175.06	180.57
κ	0.75	2.29	3.080	4.974	0.214	1.047	3.493	3.242
min_{θ} [$^{\circ}$]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
max_{θ} [$^{\circ}$]	360.00	360.00	360.00	360.00	357.30	357.30	359.10	360.00

TABLE 1.1: ORU.WIND data statistics. In total, the *ORU.WIND* data set consists of 8 indoor and outdoor data sets. The name of data sets start with “ORU.WIND”, following with two letter word indicating the type of data. OS and OC are for “OUTDOOR SIMPLE” and “OUTDOOR COMPLEX”, while IS and IO stand for “INDOOR SIMPLE” and “INDOOR COMPLEX”, respectively. Data set names end with three digit number referring to the elevation at which the measurement was conducted.

campus. The proposed mapping is applied to artificial data simulated as *Gaussian Random Field (GRF)*.

1.4 Wind Data

The wind data applied for this thesis work is from the particular source *ORU.WIND*. *ORU.WIND* is a wind data set recorded specifically for this thesis work. The measurement was done by a YOUNG 81000 3D ultrasonic anemometer, set up as a portable station. The detailed characteristics of measurement are summarized in Table A.1 in Appendix A. In total, the *ORU.WIND* data set consists of 8 indoor and outdoor measurements starting from three-hour wind measurement. The wind measurements are performed at stationary positions with fixed portable masts at 100 and 150 centimetres height. In order to reproduce the complexities involved with a typical mobile robot mapping task, wind measurements are performed with two degrees of complexity. Outdoor recordings are done at two specific locations involving relatively simple and complex terrains, open space grass field and a field occupied by trees and nearby cars. For indoor recordings the complexity was applied as the low and high occupancy times. The complete list of *ORU.WIND* data set is presented in Table 1.1.

1.5 Thesis Outline

This thesis report is organised as follows.

In the Chapter 2, the introductory objectives describing the wind origin, wind spatio-temporal variations and wind descriptors are presented. In Chapter 3, the focus is on

wind speed statistical modelling. Several methods are discussed and the approach to model wind speed data is proposed, implemented and evaluated. In Chapter 4, the research efforts is to model wind direction. In Chapter 5, the algorithm for building spatial wind map is outlined and described. Finally, in Chapter 6, the work is wrapped up by pointing out future research directions and presenting concluding remarks.

Chapter 2

The Wind Phenomenon

This chapter provides basic ideas and theories related to the nature of wind and its respective implications. In Section 2.1, the wind origin and its unique inherent uncertainties is studied. Specifically, the spatio-temporal variation of wind is reviewed and explained, individually. Proceeding to the end of the chapter, Section 2.2 present a short summary on the available descriptors to formulate wind speed and direction.

This chapter is mainly guided by [9], unless otherwise stated.

2.1 The Wind Theory

2.1.1 Origins of Wind

Wind is generally defined as the motion of gas molecules over the surface in a particular direction. The main origin of wind is the difference in pressure imposed by uneven solar radiation and thermal differences, although other artificial sources could cause air movements [9]. The air flow, specifically surface winds, does not follow a uniform pattern and they embody with several sources of uncertainty. In fact, the most remarkable characteristic of wind in meteorology, wind power and indeed in GDM, is its uncertainty and variability in time and space.

2.1.2 Wind Temporal Variations

Wind is subject to temporal variation at multiple scales. In general, the temporal variations of wind could be classified into 4 main categories: (I) Long-term variations, (II)

annual and seasonal variations, (III) synoptic and diurnal variations and (IV) turbulence. These broad scales of wind fluctuations have made wind a very complex and challenging problem in various fields.

The long term wind variations is observed at very large scales from years to decades. The perennial temperature changes may have the greatest effect on long-term wind variations, although there are other several natural and human causes effecting wind variability in long term, such as volcanic eruptions, changes in solar radiations and *El Niño*¹. The smaller scale of wind variabilities occur annually and seasonally. Theses sorts of variabilities are closely coupled with the seasonal temperature changes. Compared to intra-year variations of wind which due to difficulties of sampling process are complex to characterize, It was observed seasonal variations characteristics are more convenient to model. Synoptic² and Diurnal³ variations are found in shorter temporal scales. These scale of variations quantitatively and qualitatively are rooted into the large-scale weather patterns such as low and high pressure regions and different weather fronts. The fastest fluctuation in wind is associated with time scales ranging from minutes to seconds or even to time scales below than a second, known as turbulence.

The main cause of turbulence are forces affected by topography, but the source of these forces could be different. Based on how topography influence turbulence, it could be classified into two groups: Thermal and Mechanical.

Thermal turbulence is referred to the kind of irregularities of air movement that induced from the effect of heating surfaces. It does occur as the result of clash between rising thermals and prevailing wind in local atmosphere. Thermal turbulence usually observed at higher altitudes in comparison with Mechanical turbulence.

Mechanical turbulence is caused by physical characteristics and shape of the topography. *Wind shears*⁴ caused by frictional drag forces effecting wind, and the lift force and the deflection of flow caused by large obstructions on the ground are the examples that could lead to Mechanical turbulence [11]. However, there might be some circumstances where thermal or mechanical effects are both contributed in generation of turbulence in the same time.

There have been several efforts to investigate the temporal wind variability over various time scales, but the first generic study of temporal wind variations was reported by

¹El Niño is referred to the class of events related to the presence of a warm ocean current periodically developing off western coast of South America by which the Pacific Ocean climate can be effected. [10]

²A particular meteorological scale in which atmospheric motion occurs in a typical range of several hundreds of kilometers [10]

³Atmospheric cycles completed within 24 hours.[10]

⁴Wind shear is referred to 'the local variation of the wind vector or any of its components in a given direction'. [10]

Issac Van Der Hoven in 1957 [12] where the horizontal wind speed fluctuations were analysed through the power density spectrum over a broad range of frequencies. Figure 2.1 presents the Van Der Hoven wind spectrum. The distinct characteristic of the wind spectrum is the *spectral gap* by which the turbulent motions and other temporal scales of wind trends are separated. The two peaks at the left side of wind spectrum approximately centred at ‘4 days’ and ‘10 hours’ resemble synoptic and diurnal patterns, respectively. On the other hand, The the turbulent flows peaks at 1 minute and range from less than a minute to 1 hour.

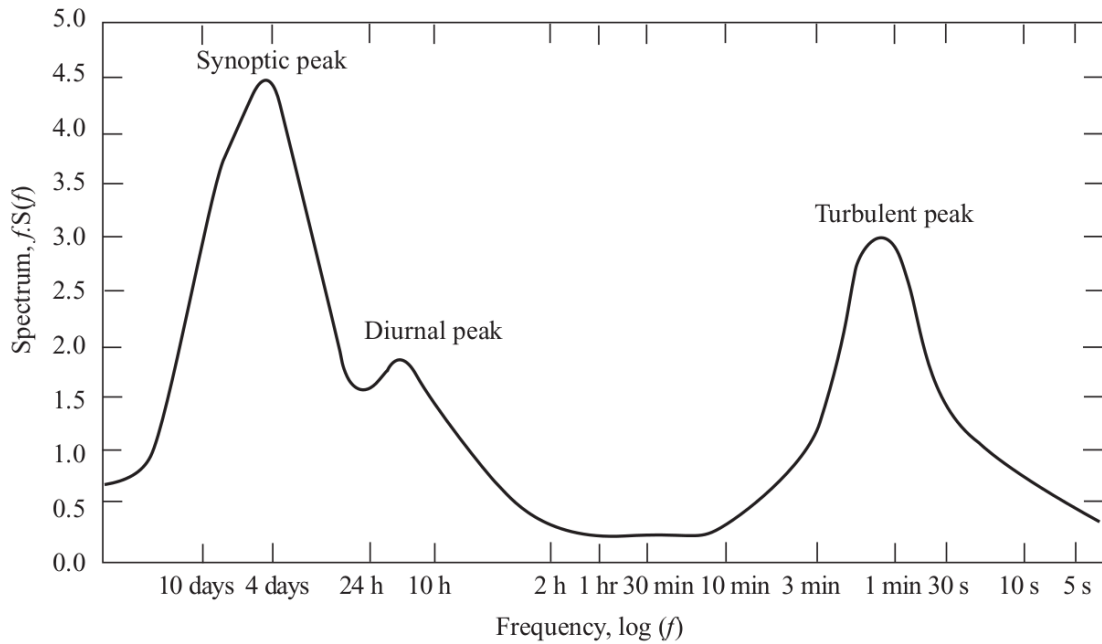


FIGURE 2.1: Vad der Hoven wind spectrum - Cycles per time (1957) [Adapted from [9]]

2.1.3 Wind Spatial Variations

The wind spatial variability is persistent over a variety of scales. On the very coarse scales, the variability stems from the insolation rate interconnected with latitudes. In other words, the sparsity of latitudes is the phenomenon that influences surface heating trend and eventually results to different climatic regions with various windiness behaviour. At the smaller scales, the spatial variance in wind is attached to the types of geo-systems, i.e. vegetations and terrain types. The type of geo-system impacts on the amount of absorption of solar energy form ground by which wind trend affected. Over the very small scales, the friction with earth surface is adjoined with spatial variation of wind. Topography and man-made structures have a notable effect on the wind behaviour.

Scale	Distance Scale	Examples
Macroscale		
Planetary	1000–4000km	Westerlies, Trade Winds,...
Synoptic	1000–5000km	Midlatitude Cyclones, Hurricanes,...
Mesoscale	1–2000km	Thunderstorms, Tornadoes
Microscale	< 2km	turbulence, Dust Devils, Wind Gusts,...

TABLE 2.1: Atmospheric models are classified as per variability in space

In meteorological sense, the wind trend fluctuations in space are classified under the headings of the three of atmospheric models, namely macroscale, mesoscale and microscale. The rough picture of these three scales is depicted in Table 2.1.

Formally, as per *Glossary of Meteorology (American Meteorological society)* [10] the macroscale systems range through thousands of kilometres and affected by synoptic events, e.g. anticyclones, cyclones, fronts and jet streams. Mesoscale models fall in between micro- and meso-scales, and mantle areas in range of few to several hundred kilometres, particularly from 2 to 2000 kilometres. Thunderstorms, tornado and extratropical cyclones, and wind trends that generated by topographical effects such as mountain waves and sea and land breezes are from this group. The microscale atmospheric models are located at the lowest side of scale spectrum and work on small spatial scales less than 2 kilometres.

Figure 2.2 draws parallel between wind spatial and temporal variability. It can be seen that the scale of variability in space and time are directly proportional to each other. That is, the shorter wind variation in time correspond to smaller spatial scales.

The wind phenomena related to microscale atmospheric systems are limited to the ones which basically stem from atmospheric boundary layer (ABL). Atmospheric boundary layer (ABL) or planetary boundary layer (PBL) is the bottom layer of troposphere which is under the influence of surface frictional effects. ABL approximately occupies 10% of atmospheric layer ranging from 0.5–2 km above ground. More specifically, ABL is divided into further two sublayers: *Ekman* and *Surface*, as shown in Table 2.2. Ekman layer is referred to the part of ABL in which wind direction changes are observed. Surface layer comprises the lower region of ABL whose height varies from few meters in stable condition to 20–50 meters in stable condition. The main attribute of Surface layer is that the vertical flux usually is assumed to be steady. In ABL, except a tiny part of Surface which is under the effect of molecular forces, the process of exchange flux is turbulent [13].

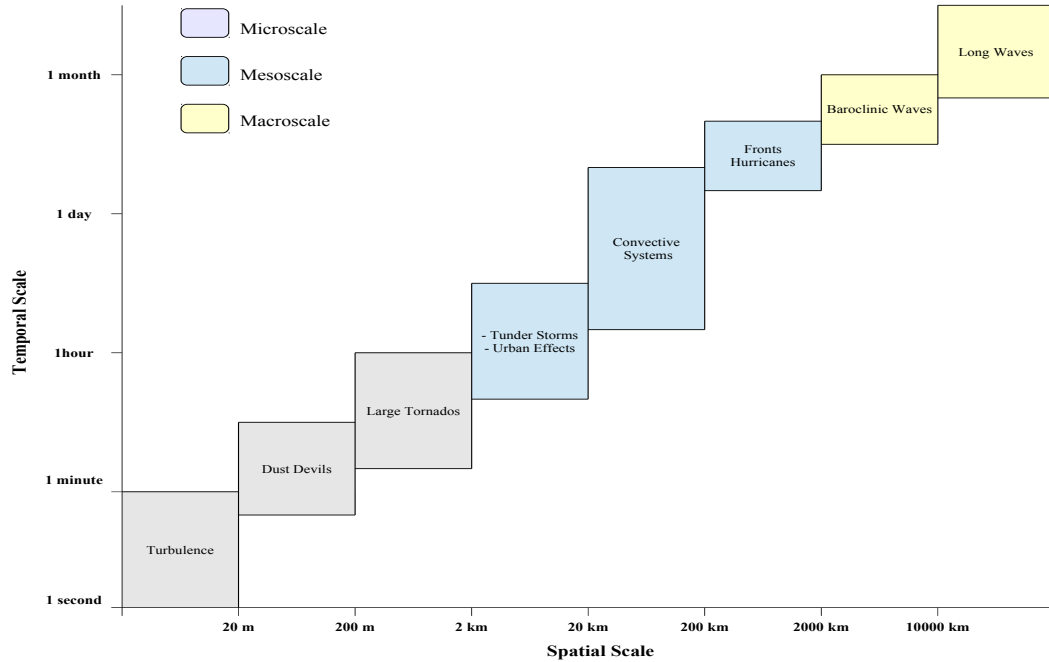


FIGURE 2.2: Spatial vs temporal wind variability

Layer Name	Height	Exchange	
Ekman	1000m	turbulent	Non-constant Flux
Surface	20m	Turbulent	Constant Flux
	1m	Turbulent/Molecular	
	0.01m	Molecular	
	0.001m	Molecular	

TABLE 2.2: Structure of Atmospheric Boundary Layer [adapted from [13]]

2.2 Wind Descriptors

Similar to any moving object in space, wind could be characterized by its magnitude (wind speed) and direction, thereby one may represent wind as a vector quantity in a three dimensional space. The wind vector is decomposed into three components in the reference frame, where the x-axis is a long wind-mean direction. The y-axis is in the horizontal plane and the z-axis is oriented upwards. In literature, x-, y- and z-axis respectively are referred to longitudinal, lateral and vertical directions. However, while dealing with ABL wind patterns close to surface, the vertical component of surface wind is fairly small compared to its horizontal component and assumed to equal zero. As a consequence, the representation of wind generally is constrained to two-dimensional horizontal planar space with longitudinal and lateral directions.

Apart from vector representation of wind, wind could also be described as two separate scalar variables: wind speed and wind direction. In other words, wind is represented in

polar coordinate system as the vector magnitude and orientation are defined as wind speed and wind direction.

Traditionally, wind speed and direction qualities are represented by primary deterministic variables. But, due to the broad scale of uncertainties in wind it more common to consider wind as a stochastic process and analyse statistically. that is, owing to have a more in-depth analysis wind speed and taking advantage of statistical tools, wind speed and direction essentially are described as random variables.

From statistical point of view, having considered the application and the scale of variation involved with wind representation, several statistical descriptors have been introduced to model and analyse air movement characteristics. The following highlights some of important wind speed and direction descriptors.

2.2.1 Wind Speed

2.2.1.1 Speed Mean

Wind speed traditionally is described by its first moment statistics, speed mean. Wind speed mean is referred to the average quantity of instantaneous horizontal wind speed collected at fixed point over a specific period of time. The mean of N samples of scalar horizontal wind speed u_i is

$$\bar{u} = \frac{1}{N} \sum_{j=1}^N u_j \quad (2.1)$$

Officially, according to the the World Meteorological Organization (WMO) standard, the recommended averaging period may range from 10 to 60 minutes; 10 minute as the most common practice. This benefits from incorporating much as fluctuations to estimate the wind speed mean. The 10-minute wind speed mean usually is applied for short- and long-term wind modelling and prediction in wind energy and mesoscale meteorology.

2.2.1.2 Wind Speed Frequency-Distributions

Apart from wind speed moment statistics, it is also customary to describe wind speed statistics with the probability frequency-distribution functions. The eminence of probability frequency-distribution basically lies in its distinct strength in variation analysis. Wind speed Probability frequency-distribution has been widely used to address the wind inherent spatial and temporal variabilities in climatic studies and wind power estimation.

Indeed, in the literature, there could be found several wind speed frequency-distribution models aiming to comply with either the characteristics of different wind speed regimes or the the requirement of applications.

The majority of wind speed frequency-distribution models are dedicated to mesoscale processes in which wind is facing with extensive variations in space and time. Generally, due to inhomogeneity of dynamic properties of mesoscale systems, wind speed distributions are non-Gaussian and tend to be skewed. By way of illustration, the followings may be listed as the key probability frequency-distribution in mesoscale models.

1. *Weibull* In literature, particularly in the area of wind energy, the two parameter Weibull distribution is most commonly distribution model describing the skewed wind speed statistics. It is given by

$$f_X(x; \alpha, \lambda) = \begin{cases} \left(\frac{\alpha}{\lambda}\right) \left(\frac{x}{\lambda}\right)^{\alpha-1} \exp\left(-\frac{x}{\lambda}\right)^\alpha, & x \geq 0. \\ 0, & x < 0. \end{cases} \quad (2.2)$$

where $\lambda > 0$ and $\alpha > 0$ are respectively, scale and shape parameters (Figure 2.3).

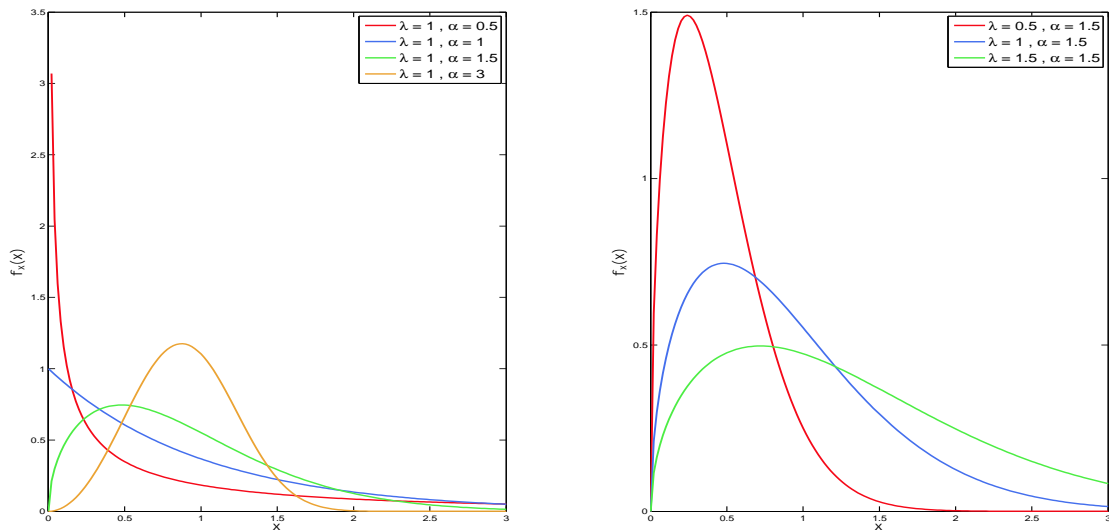


FIGURE 2.3: Left: 2-parameter Weibull distribution given different shape parameters while scale parameter being constant; Right: 2-parameter Weibull distribution given different scale parameters while shape parameter being constant

The prevalence of Weibull distribution has been known by its simplicity and unique characteristic in describing wind power estimation. Wind speed Weibull distribution with shape parameter α leads to cubed wind speed Weibull distribution with shape $\frac{\alpha}{3}$ parameter. This simplifies the computation wind power which is proportional to the cubed wind speed.

Nevertheless, the validation of Weibull distribution basically relies on experimental observations rather than solid theoretical grounds. The Weibull distribution does not hold in modelling of wind regimes dominated by null wind speeds and bimodal distribution [14].

2. *Rayleigh* distribution is the special case of Weibull distribution with shape parameter equals 2 (Figure 2.4). The Rayleigh function with scale parameter λ is defined as

$$f_X(x; \lambda) = \begin{cases} \frac{x}{\lambda^2} \exp\left(-\frac{x^2}{2\lambda^2}\right), & x \geq 0. \\ 0, & x < 0. \end{cases} \quad (2.3)$$

Rayleigh distribution is observed as the bivariate distribution of uncorrelated

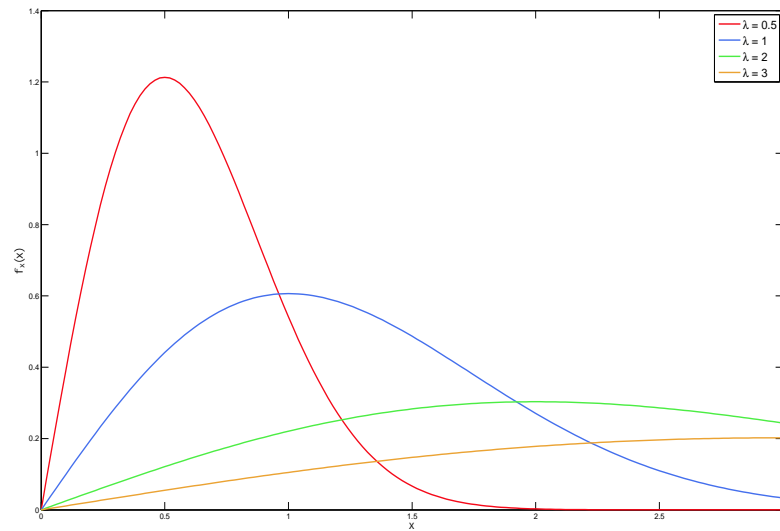


FIGURE 2.4: Rayleigh distribution given different scale parameters

Gaussian orthogonal wind components with zero variances. Thus by, it could be a good approximation for wind regimes whose components are uncorrelated and tend to be normally distributed.

3. *Weibull-3* is the generalized form the 2-parameter Weibull distribution with the additional position parameter β locating the distribution along the abscissa.

$$f_X(x; \alpha, \lambda, \beta) = \begin{cases} \left(\frac{\alpha}{\lambda}\right) \left(\frac{x-\beta}{\lambda}\right)^{\alpha-1} \exp - \left(\frac{x-\beta}{\lambda}\right)^{\alpha}, & x \geq 0. \\ 0, & x < 0. \end{cases} \quad (2.4)$$

4. *Inverse-Gaussian* distribution is a 2-parameter distribution model which is suitable for wind data with low frequencies of low wind speed (High wind speeds are more probable). As the result, It is an alternative method to Weibull-3, which, in

terms of complexity, involves with simpler parameter estimation methods. Inverse-Gaussian with mean parameter μ and shape parameter α is (Figure 2.5):

$$f_X(x; \mu, \alpha) = \begin{cases} \left(\frac{\alpha}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left(\frac{-\alpha(x-\mu)^2}{2\mu^2 x}\right), & x \geq 0. \\ 0, & x < 0. \end{cases} \quad (2.5)$$

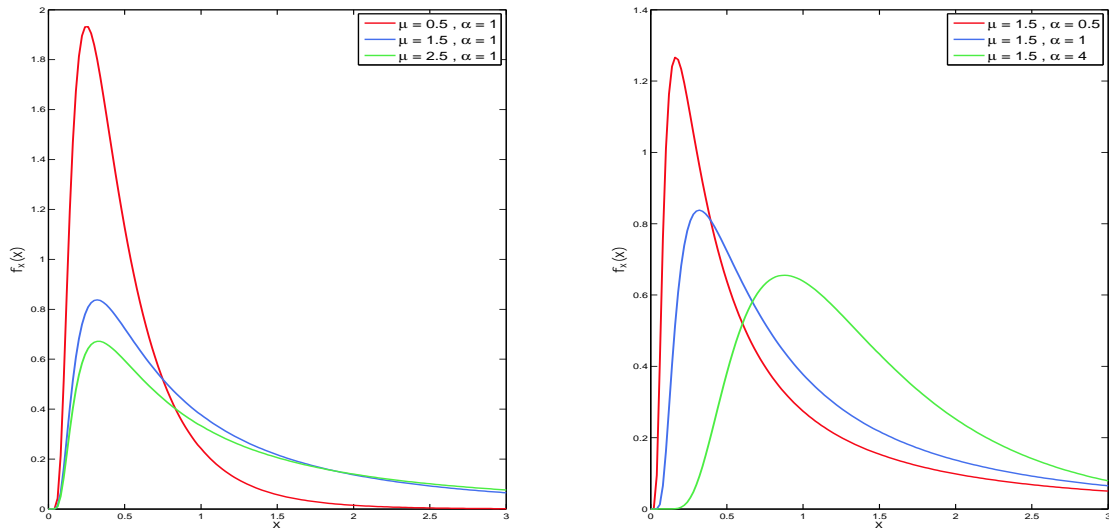


FIGURE 2.5: Left: Inverse-Gaussian distribution given different mean parameters while shape parameter being constant; Right: Inverse-Gaussian distribution given different shape parameters while mean parameter being constant

5. *Generalized Extreme Value (GEV)* GEV is assumed as an alternative method to Weibull-3 distribution to address the negative skewness of tropical wind speed data collected. It is limited to use in negatively skewed distributed wind data. Accordingly, GEV is defined by shape, scale and position parameters, α , λ , and β respectively (Figure 2.6).

$$f_X(x; \alpha, \lambda, \beta) = \begin{cases} \exp\left\{-\left[1 + \alpha\left(\frac{x-\beta}{\lambda}\right)\right]^{-\frac{1}{\alpha}}\right\}, & x \geq 0. \\ 0, & x < 0. \end{cases} \quad (2.6)$$

2.2.1.3 Turbulence Fluctuations

Following *Reynolds* decomposition, it is convenient to assume the time dependent wind speed, in short time scales, as the combination of a constant mean part, resembling the seasonal, synoptic and diurnal effects, and fluctuating parts with zero mean [15]. That is,

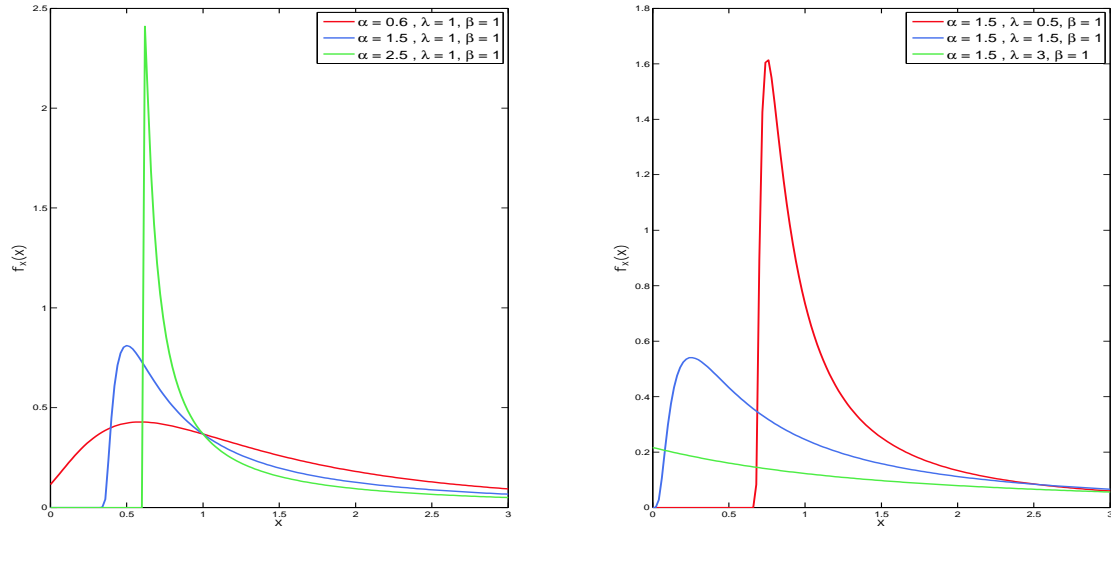


FIGURE 2.6: Left: Generalized Extreme Value distribution given different shape parameters while scale and position parameters being constant; Right: Generalized Extreme Value distribution given different scale parameters while shape and position parameters being constant

the wind vector $\mathbf{u}(t)$ is constructed as the sum of wind speed mean \bar{u} and random parts along longitudinal, lateral and vertical directions, $\chi_u(t)$, $\chi_v(t)$ and $\chi_w(t)$, respectively.

$$\begin{aligned} \mathbf{u}(t) &= \bar{\mathbf{u}} + \chi_u(t) \\ \mathbf{v}(t) &= \chi_v(t) \\ \mathbf{w}(t) &= \chi_w(t) \end{aligned} \quad (2.7)$$

Due to complexity of ABL turbulence, It usually involves with simplifications. ABL turbulence is usually considered isotropic. That is, at given point, the rate of variations in longitudinal, lateral and vertical fluctuations are equal.

$$\sigma_u = \sigma_v = \sigma_w \quad (2.8)$$

Turbulent fluctuations is roughly Gaussian. Granted the averaging interval is long enough to incorporate all the fluctuation, wind speed fluctuations could be approximated with Gaussian distribution about wind speed mean and the standard deviation of σ_u . In particular, approximation of fluctuations with Gaussian distribution could be justified such that the turbulence level with respect to mean is small. [16].

$$\begin{aligned} f(u) &= \mathcal{N}(u; \bar{u}, \sigma_u) \\ &= \frac{1}{\sigma_u \sqrt{2\pi}} e^{-(u-\bar{u})^2 / 2\sigma_u^2} \end{aligned} \quad (2.9)$$

For high wind speeds, lateral and vertical compared to longitudinal fluctuations is usually insignificant and may be neglected.

2.2.1.4 Turbulence Intensity

The turbulence intensity is non-dimensional measure of overall turbulence which is defined as the relative variability of fluctuations, standard deviation σ to the wind speed mean \bar{u} .

$$I = \frac{\sigma}{\bar{u}} \quad (2.10)$$

2.2.1.5 Wind Speed Increments

Along with the moment statistics of fluctuations itself, turbulence may be described and analysed by using the statistics of fluctuations differences, known as wind speed increments. Simply, at a given time t , wind increment $\mathbf{u}_{t;\tau}$ is defined as difference between

$$\mathbf{u}_{t;\tau} = \mathbf{u}_{t+\tau} - \mathbf{u}_t \quad (2.11)$$

Wind speed increments with very short time intervals accounts for wind gusts, and essentially could be applied to extreme wind speed estimation and risk analyses. It is well known that these PDFs are highly intermittent and does not follow gaussian distributions, therefore higher statistical moments of such increments are needed for a proper description and estimation of possible extreme events.

2.2.2 Wind Direction

The wind direction is a circular variable which ranges over a period of $0-2\pi$ radian. Circular variables have the specific characteristic that the lowest and highest value of sample space are connected to each other, imposing discontinuity over data samples. As a result of this discontinuity, the circular data could not be averaged by conventional linear methods and particular consideration should be taken while computing the directional mean.

2.2.2.1 Direction Mean

For N samples of circular variable θ , the circular mean is computed by transforming the angular variables into Cartesian coordinates and taking averages of Cartesian components individually:

$$\bar{S} = \frac{1}{N} \sum_{j=1}^N \sin \theta_j, \quad \bar{C} = \frac{1}{N} \sum_{j=1}^N \cos \theta_j \quad (2.12)$$

$$\bar{\Theta} = \begin{cases} \tan\left(\frac{\bar{S}}{\bar{C}}\right), & \text{if } \bar{C} \geq 0. \\ \tan\left(\frac{\bar{S}}{\bar{C}}\right) + \pi, & \text{if } \bar{C} < 0. \end{cases} \quad (2.13)$$

2.2.2.2 Trigonometric Moments

In directional statistics, it is common to estimate the p th trigonometric moment m'_p in polar coordinates

$$m'_p = a_p + ib_p = \bar{R}_p \exp(i\bar{\theta}_p) \quad (2.14)$$

Where

$$a_p = \frac{1}{N} \sum_{j=1}^N \cos(p\theta_j), \quad b_p = \frac{1}{N} \sum_{j=1}^N \sin(p\theta_j) \quad (2.15)$$

Here, \bar{R} is denoted as *p*th mean resultant length [17]. Accordingly the first trigonometric moment is referred to

$$m'_1 = \bar{C} + i\bar{S} = \bar{R} \exp(i\bar{\theta}) \quad (2.16)$$

It is notable that trigonometric moments in directional data is at greater advantage compared to linear distribution. As, directional distribution of data could be described by their trigonometric moments. That is

$$\mathbb{E}[e^{ip\theta}] = \alpha_p + i\beta_p = \rho_p \exp(i\mu_p) \quad (2.17)$$

given

$$\alpha_p = \mathbb{E}[\cos(p\theta)], \quad \beta_p = \mathbb{E}[\sin(p\theta)] \quad (2.18)$$

2.2.2.3 Trigonometric Moments about Mean Direction

The p th trigonometric moment about direction mean $\bar{\theta}$ is denoted as

$$m_p = \bar{a}_p + i\bar{b}_p \quad (2.19)$$

where

$$\bar{a}_p = \frac{1}{N} \sum_{j=1}^N \cos p\theta_j - \bar{\theta} \quad \bar{b}_p = \frac{1}{N} \sum_{j=1}^N \sin p\theta_j - \bar{\theta}, \quad (2.20)$$

2.2.2.4 Wind Direction Frequency-Distributions

The wind direction could also be represented by frequency-distribution models. The wind direction frequency distribution model is a convenient statistical instrumental tool in harvesting wind direction and wind-induced fatigue analysis. In literature, there have been several studies conducted on the subject frequency-distribution of wind direction, such as uniform and various wrapped distributions [17].

However, in the domain of wind energy, the most important and commonly used distribution is Von-Mises distribution, also known as circular normal. Von-Mises is unimodal symmetric distribution describing directional data with the two specific parameters: mean and concentration. That is, for wind direction, $0 \leq \theta < 2\pi$ with mean parameter μ and concentration parameter κ Von-Mises distribution is defined as (Figure 2.7)

$$f_{\Theta}(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\theta - \mu)) \quad (2.21)$$

Here $\kappa \geq 0$ and $I_0(i)$ is the modified Bessel function of the first kind and order 0.

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\kappa \cos \theta) d\theta \quad (2.22)$$

Von-Mises distribution notably benefits from statistical similarities to the linear normal distribution. In particular, for sufficiently large value of concentration parameter, it could be approximated to normal distribution. That is, for very large values of κ :

$$\begin{aligned} \therefore \theta - \mu &\approx 0, \quad (\text{for large } \kappa) \\ \therefore \cos(\theta - \mu) &\approx 1 - \frac{(\theta - \mu)^2}{2} \end{aligned} \quad (2.23)$$

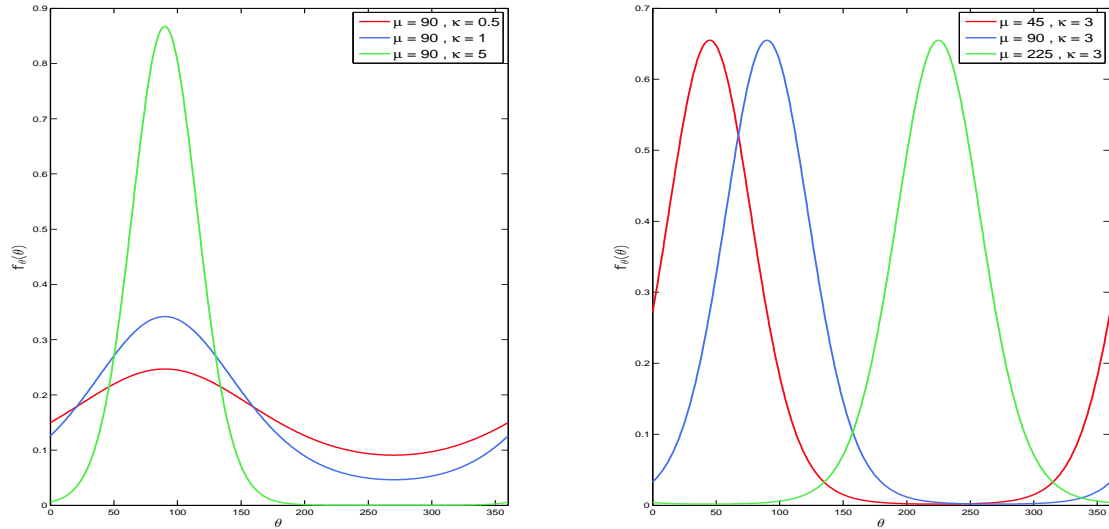


FIGURE 2.7: Left: Von-Mises distribution given different concentration parameters while mean parameter being constant; Right: Von-Mises distribution given different mean parameters while concentration parameter being constant

Thus by, the Von-Mises distribution function could be approximated to

$$\begin{aligned}
 f_{\Theta}(\theta; \mu, \kappa) &\approx \frac{1}{2\pi I_0(\kappa)} \exp\left(\kappa - \frac{\kappa(\theta - \mu)^2}{2}\right) \\
 &\propto \exp\left(\frac{-\kappa(\theta - \mu)^2}{2}\right)
 \end{aligned}
 \tag{2.24}$$

which corresponds to PDF of normal distribution with mean μ and variance κ^{-1} : $\mathcal{N}(\theta; \mu, \kappa^{-1})$.

Chapter 3

Statistical Wind Speed Modelling

In this chapter, the approach adapted for modelling short-scale wind speed, turbulence, is introduced and elaborated. The wind speed is modelled as pure Gaussian distribution, and particularly the main course of work is about to determine its governing statistical moments, mean and variance, in a real time manner. The methodology for estimation of the averaging interval, the associated mathematical details and experimental results are presented.

3.1 Background

In the statistical sense, it is customary to represent wind speed fluctuations by means of probability distribution function (PDF). Several statistical distribution models have been proposed to approximate wind speed variations. However, most of these efforts are involved with modelling meso-scale wind speed regimes and only a limited number of distribution methods are available for short term wind speed fluctuations, turbulence [18]. Recalling the mentioned classification of wind regimes in Chapter 2, any statistical models under the heading mesoscale systems, although valuable, are not consistent with the specific details of Mobile Robot Wind Mapping (MRWM), where wind or more truly wind fluctuations should be viewed as microscale turbulence.

It is convenient to assume that the idealized turbulence is nearly Gaussian distributed. Atmospheric microscale wind speed fluctuations have the very similar structure and, provided that the wind speed field consists of a series of segments with different average values, follow Gaussian distribution [16]. That is, the atmospheric turbulence over the periods of constant wind speed average is considered to be Gaussian distributed. Although, this assumption of Gaussianity is not strictly valid for every circumstances,

particularly for close surface wind flows as the case for the mobile robot wind data measurements, it still could be used as the PDF approximation in many practical situations. In general, Gaussian distribution is the most convenient way to model short-term wind speed fluctuations.

The central issue in estimation of the Gaussian statistics of a short-term wind speed field is averaging. Difficulty in averaging arises when the scale of variation is not fixed and changes occur over a broad range of time scales, from few seconds to several months [19]. This is in parallel with the arguments in [20] criticizing the fixed 10 minutes averaging interval. Indeed, the main goal in the modelling of wind speed fluctuations is to identify a valid interval in which wind speed mean and its variations could truly be defined. This happens provided that the wind speed fluctuation is reasonably stationary ergodic process.

3.2 Requirements

The first step approaching to modelling wind speed fluctuations is to adapt a procedure by which the statistical properties of collected wind speed data are characterized and modelled. Here, as mentioned in previous section, the key issue is to determine the expected scale of variations that could reasonably postulate wind speed fluctuation. In literature, there are several available methods that could be used in this purpose. Gupta [21] employed the variable interval time averaging method (VITA) by which a fixed short scale averaging interval is employed to detect the burst frequency of ABL turbulence. In [22], a unified method of a collection of statistical procedures is proposed to determine “acceptable averaging distance”.

However, the favourable approach within the framework of mobile robot SGDM is the one that suits the inherent and implicit requirements of microscale wind speed modelling. The existing mobile robot SGDM methods particularly aims for mapping distribution of gases locally while limited to the very short time scales. In most of the scenarios, the robot only spends less than a minute for measurement and could not wait for a long course of time, for instance several minutes or hours. It is apparent that in the MRWM task, an approach is of great interest which could model wind speed in very short time spans. In other words, wind speed modelling incorporating MRWM, basically lies under the heading microscale wind speed systems and should be able to model wind in very short scales without need to any user intervention.

To accomplish this, an online averaging method introduced by [23] called *Time-Dependent Memory Method (TDMM)* is implemented. TDMM particularly is chosen since, in

essence it is an online framework that could extract the averaging length in real time. Moreover, it is fully automated approach that eliminates any user interactions. The details and practical issues of the adapted TDMM method is presented in the next section.

3.3 Methodology

3.3.1 Theory

The wind speed time series recorded over the period of $[t - T \ t]$ is described by $\mathbf{u}(t)$. In a sense, T is considered as the sampling period and t referred to the *running time*. The instantaneous wind speed mean of ΔT of previous recordings is estimated as

$$\bar{\mathbf{u}}(\gamma, \Delta T) = \frac{1}{\Delta T} \int_{t-\Delta T}^t \mathbf{u}(\gamma) d\gamma, \quad \Delta T < T \quad (3.1)$$

This formulation turns out to be a *real time* local wind averaging. It is notable that the upper bound of the integral is limited to the present time t and the averaging only incorporates the present and the past wind speed recordings. Being independent of future observations does make the averaging scheme in 3.1 suitable for online applications where the instantaneous mean is required to be computed in real time.

In general, the main question in computation of instantaneous wind speed mean, $\bar{\mathbf{u}}(t, \Delta T)$ is to define the extent of the previously recorded data that is required to consider in averaging process. In particular, it is to estimate the *requisite interval* ΔT such that it is a valid scale of $\bar{\mathbf{u}}(t, \Delta T)$ while standing at its lowest limit. The *requisite interval* ΔT varies randomly with time and ABL wind average is a realistic measure as much as it matches to actual scale of fluctuations which is not dependent on t .

From the statistical point of view, corollary of idealized turbulence, a valid estimate of the requisite ΔT corresponds to the interval $[t - \Delta T \ t]$ where the mean of fluctuations is zero and instantaneous average steadies. That is, for the stationary wind speed time series, the *requisite interval* is truthful provided that $\bar{\mathbf{u}}_n^{(N\Delta T)}$ is almost same for the every sample of interval of $[t - \Delta T \ t]$, $\bar{\mathbf{u}}_n^{(N\Delta T)}$. Thereby, in order to estimate the valid averaging interval this is the constancy of instantaneous means which is required to be monitored. By this means, The first detected interval with approximately of constant instantaneous mean is referred to the shortest valid *requisite interval*, ΔT .

However, one may evaluate fluctuations to determine the shortest valid averaging interval. That is, in lieu of instantaneous means, the mean of fluctuations, not only is

investigated for steadiness, but also for the constancy of zero. Given the requisite interval ΔT , wind speed fluctuations $\chi(t, \Delta T)$ is estimated by subtracting ABL mean $\bar{u}(t, \Delta T)$ from instantaneous wind speed $u(t)$.

$$\chi(t, \Delta T) = u(t) - \bar{u}(t, \Delta T) \quad (3.2)$$

The sequence of fluctuations is estimated over the entire set of $[t - \Delta T, t]$, by which the time series of $\{\chi(\gamma) : \gamma \in [t - \Delta T, t]\}$ is formed. Here $\chi(\gamma)$ is identical to $\chi(\gamma, \Delta T)$, and for the sake of simplicity, in further references $\chi(\gamma)$ will be used. Then the time series of fluctuations $\chi(\gamma)$, by means of *mean squared error (MSE)*, is undergone test for optimality. The optimal requisite interval corresponds to the ΔT whose MSE is equal to zero.

Formally, MSE,

$$\text{MSE} = \left\{ \frac{1}{\Delta T} \int_{t-\Delta T}^t \chi(\gamma) d\gamma \right\}^2 \quad (3.3)$$

$$\begin{aligned} \text{MSE} &= \left\{ \left(\frac{1}{\Delta T} \right)^2 \int_{t-\Delta T}^t \int_{t-\Delta T}^t \chi(\gamma_1) \chi(\gamma_2) d\gamma_1 d\gamma_2 \right\} \\ &= \bar{\chi}^2(t, \Delta T) \end{aligned} \quad (3.4)$$

Motivated by the limited number of wind speed recordings available, by substitution of $\gamma_1 = \gamma - \frac{\tau}{2}$ and $\gamma_2 = \gamma + \frac{\tau}{2}$, MSE further interpreted as

$$\text{MSE} = \frac{1}{\Delta T} \int_{-\Delta T}^{\Delta T} \left(\frac{1}{\Delta T - |\tau|} \int_{t-\Delta T + \frac{|\tau|}{2}}^{t - \frac{|\tau|}{2}} \chi(\gamma - \frac{\tau}{2}) \chi(\gamma + \frac{\tau}{2}) d\gamma \right) d\tau \quad (3.5)$$

which is interesting, in essence that the inner integral (integration inside brackets) is autocorrelation. The expression inside the brackets () can be taken as a function of the three variables: t , τ and ΔT . However, for the sufficiently large value of ΔT autocorrelation loses its dependency on ΔT . More specifically, when ΔT is remarkably greater than any value of τ in which autocorrelation is not zero, the scale factor of the inner integral is approximated with $\frac{1}{\Delta T}$ by which the autocorrelation is made independent of ΔT . That is to say, for the reasonably large value of ΔT , the inner integral in Equation 3.5 keeps unchanged at a value that is not function of ΔT any more, say $C(t, \tau)$:

$$\text{MSE} = \frac{1}{\Delta T} \int_{-\Delta T}^{\Delta T} C(t, \tau) d\tau \quad (3.6)$$

The autocorrelation is steady at $C(t, \tau)$. This implies that the whole integral in () is

also steady with more increasing ΔT . Now, here is a definite integral that maintains constant while changing the integral bound and this is not the case unless the whole integral in Equation 3.6 is approximately equal to zero. Consequently, the valid estimate of *requisite interval* is such that MSE of random part of ABL wind speed corresponds to zero. Indeed, the wind speed mean estimated within this interval could account for the valid estimate of instantaneous mean and the left residuals for the fluctuations.

$$\text{MSE} = \frac{1}{\Delta T} \int_{-\Delta T}^{\Delta T} C(t, \tau) d\tau \approx 0 \quad (3.7)$$

3.3.2 Practical Considerations

In this section TDMM method, Algorithm 1 is described with further details, particularly by focusing on the empirical implications. The entire process is exemplified by a short sequence of wind speed recordings adjusted from WIND-ORU dataset.

Input: t : Current time, T_s : Sampling rate

Output: ΔT_{valid}

```

begin
  % initialization;
   $\Delta T = 0$ ;
   $n = 2$ ;
  while ( $t > 2\Delta T$ ) do
     $\Delta T = \Delta T + nT_s$ ;
    foreach  $\gamma \in [t - \Delta T \quad t]$  do
      Detrend  $\{u(\gamma) : \gamma \in [\gamma - \Delta T \quad \gamma]\}$ ;
      Estimate  $\bar{u}(\gamma, \Delta T)$ ;
      Estimate  $\chi(\gamma, \Delta T)$ ;
    end
    Estimate MSE for  $\{\chi(\gamma) : \gamma \in [t - \Delta T \quad t]\}$ ;
    if  $MSE \approx 0$  then
       $\Delta T_{valid} := \Delta T$ ;
      return  $\Delta T_{valid}$ ;
    end
  end
end

```

end

Algorithm 1: TDMM Alogrithm

The actual sampling times, starting from 00:00:000, and the related sample number of wind speed recordings are listed in the first two columns of the Table. The third column is dedicated to the instantaneous wind speed values, such that the wind speed $u(\gamma)$ sampled with frequency f_{sr} , here $20Hz$, corresponds to the sample $n = \gamma \cdot f_{sr}$ is given by

Sample No.	Real Time	Wind Speed [ms^{-1}]	Inst. Mean [ms^{-1}]	Random Part [ms^{-1}]
1	00:00.000	0.190		
2	00:00.050	0.220		
3	00:00.100	0.180		
4	00:00.150	0.180		
5	00:00.200	0.180		
6	00:00.250	0.190		
7	00:00.300	0.180		
8	00:00.350	0.180		
9	00:00.400	0.180		
10	00:00.450	0.160		
11	00:00.500	0.160		
12	00:00.550	0.130		
13	00:00.600	0.150		
14	00:00.650	0.130		
15	00:00.700	0.130		
16	00:00.750	0.130	0.153	-0.023
17	00:00.800	0.130	0.148	-0.018
18	00:00.850	0.130	0.143	-0.013
19	00:00.900	0.130	0.138	-0.008
20	00:00.950	0.130	0.135	-0.005
21	00:01.000	0.130	0.132	-0.002
22	00:01.050	0.170	0.136	0.034
23	00:01.100	0.120	0.133	-0.013
24	00:01.150	0.120	0.132	-0.012
25	00:01.200	0.120	0.131	-0.011

TABLE 3.1: The sequence of wind speed with corresponding instantaneous statistics computed as per requisite interval $\Delta T = 0.45$ second.

\mathbf{u}_n

$$\mathbf{u}_n = \mathbf{u}(\gamma \cdot f_{sr}) \quad (3.8)$$

Instantaneous wind speed means and fluctuations are respectively in columns 4 and 5.

The main task of TDMM is to determine the shortest valid averaging interval ΔT . To accomplish this, MSE is estimated and tested for validity over a range of requisite intervals sorted in ascending order, say $\frac{1}{f_{sr}}, \frac{2}{f_{sr}}, \frac{4}{f_{sr}}, \dots$. Thereby, the interval estimated in this manner is the shortest time scale such that, it may be adapted for the valid instantaneous mean estimation and the corresponding fluctuations.

Provided the requisite interval ΔT , equally $N_{\Delta T}$ data samples, wind speed mean in the present time t is computed by backward averaging, i.e. averaging over the sequence of $\{\mathbf{u}(\gamma) : \gamma \in [t - \Delta T \ t]\}$.

Averaging over previously recorded data brings up a practical consideration, which is that at least $2N_{\Delta T}$ of data recordings are required to compute instantaneous means, providing the entire set of wind speed fluctuations. In other words, the complete set of fluctuations $\{\chi(\gamma) : \gamma \in [t - \Delta T \ t]\}$ could not be build, unless $2N_{\Delta T}$ of past data is available. For instance, standing at present time 00:01.200, the wind speed means are

computed from 00:00.750 to 00:01.200. Notice that, in time 00:00.750. $(t - \Delta T)$, the wind speed mean and accordingly the fluctuation is estimated by looking 0.500 (ΔT) seconds back of wind speed samples.

Another practical consideration is raised while the requisite intervals are applied to the validity test and are compared with zero. Precisely, the validity of requisite interval, from numerical computational perspective, is about the condition that MSE is sufficiently close to zero. Here sufficiently is referred to the measure of closeness, which is specified by the measurement uncertainties. In other words, this is the accuracy (ACC) of anemometer in wind speed measurement that determines the closeness of MSE to zero. In this particular work where ACC is “ $\pm 1\% + 0.05 \frac{m}{s}$ ”, the zero condition is defined by

$$\begin{aligned} \text{MSE} &\leq (\text{ACC} \cdot \bar{u}(t, \Delta T))^2 \\ &\leq (0.01 \cdot \bar{u}(t, \Delta T) + 0.05)^2 \end{aligned} \quad (3.9)$$

3.4 Experimental Results

The proposed method to model wind speed data is evaluated using *ORU.WIND* outdoor dataset. The *ORU.WIND* dataset is modified while wind speed time series are divided into equal segments of 5 minutes length, each consisting of 6001 samples. Accordingly, each segment is treated as a single stage of wind speed measurements and the TDMM algorithm is applied to detect the shortest valid averaging interval. For every segment, the probability distribution of model output is estimated over the detected intervals. The constructed distributions along the corresponding MSE trends are visualized and studied.

The performance of the approximation is evaluated by comparing the experimental density of wind speed fluctuations with predictive Gaussian distribution model. The experimental frequency distribution of wind speed fluctuations are estimated using a kernel density function, where the density of N samples wind speed fluctuations \mathbf{u}_i , $1 < i < N$ is given by

$$\hat{f}_{\mathbf{u}}(\mathbf{u}) = \frac{1}{Nh} \sum_{i=1}^N K \left[\frac{x - \mathbf{u}_i}{h} \right] \quad (3.10)$$

Here h is bandwidth and K is kernel density function. In the current work, the Gaussian kernel is used.

$$K \left[\frac{x - \mathbf{u}_i}{h} \right] = \left[\frac{1}{\sigma\sqrt{\pi}} \right] \exp \left[- \frac{(x - \mathbf{u}_i)^2}{2h^2} \right] \quad (3.11)$$

Normalized Root mean square error (NRMSE) and coefficient determination (R^2) are used to measure the deviation between the empirical density and approximated density function. The NRMSE accumulates the magnitudes of the residuals in predictions, normalized by the range of observed values of a variable being predicted. The lower NRMSE is, the better performance is achieved. NRMSE for the experimental kernel density $\hat{f}_{\mathbf{u}}(\mathbf{u})$ with estimated Gaussian distribution model $f_{\mathbf{u}}(\mathbf{u})$ is given by

$$NRMSE = \frac{\left[\frac{1}{N} \sum_{i=1}^N (\hat{f}_{\mathbf{u}}(\mathbf{u}_i) - f_{\mathbf{u}}(\mathbf{u}_i))^2 \right]^{\frac{1}{2}}}{\max(\hat{f}_{\mathbf{u}}(\mathbf{u}_i)) - \min(\hat{f}_{\mathbf{u}}(\mathbf{u}_i))} \quad (3.12)$$

The R^2 value indicates the explanatory power of the model by constructing the ratio of the sum of residuals to the sum of squares; the larger R^2 is, the more powerful approximation model is to predict the distribution.

$$R^2 = 1 - \frac{\sum_{i=1}^N (\hat{f}_{\mathbf{u}}(\mathbf{u}_i) - f_{\mathbf{u}}(\mathbf{u}_i))^2}{\sum_{i=1}^N (\hat{f}_{\mathbf{u}}(\mathbf{u}_i) - \overline{\hat{f}_{\mathbf{u}}(\mathbf{u})})^2} \quad (3.13)$$

Figures 3.1 and 3.2 demonstrate 4 specific examples of experiments where TDMM algorithm was applied to ORU.WIND. outdoor data and could achieved to detect the valid interval. In the figures, row (a) and (b) correspond to wind data measured at 100cm and 150cm height, respectively. Having considered the height of measurement, no significant difference has been observed.

The process of validating the requisite interval is presented by the line graphs where the amount of minimum square error MSE of fluctuations is plotted against to interval length ΔT . To be descriptive, at a given time the TDMM algorithm does seek for the valid interval by evaluation the MSE of residuals while iterating from the shortest to longest requisite averaging interval ΔT . As expected, according to the Equation 3.7, the shortest valid averaging interval is found when MSE approaches to zero, that is while increasing the scale of ΔT , MSE slowly decreases towards zero and maintains constant.

Notably, It could be seen that it is not always the case and the MSE value may fluctuate over the value of zero. Recall that, here an idealized model of turbulence is used and in real world scenarios, particularly the near surface wind speed fluctuations could be far from ideal turbulence model and show non-ergodic behaviour.

In the meantime, the wind speed histogram of wind fluctuations and the corresponding empirical kernel density are compared to the estimated Gaussian distribution function,

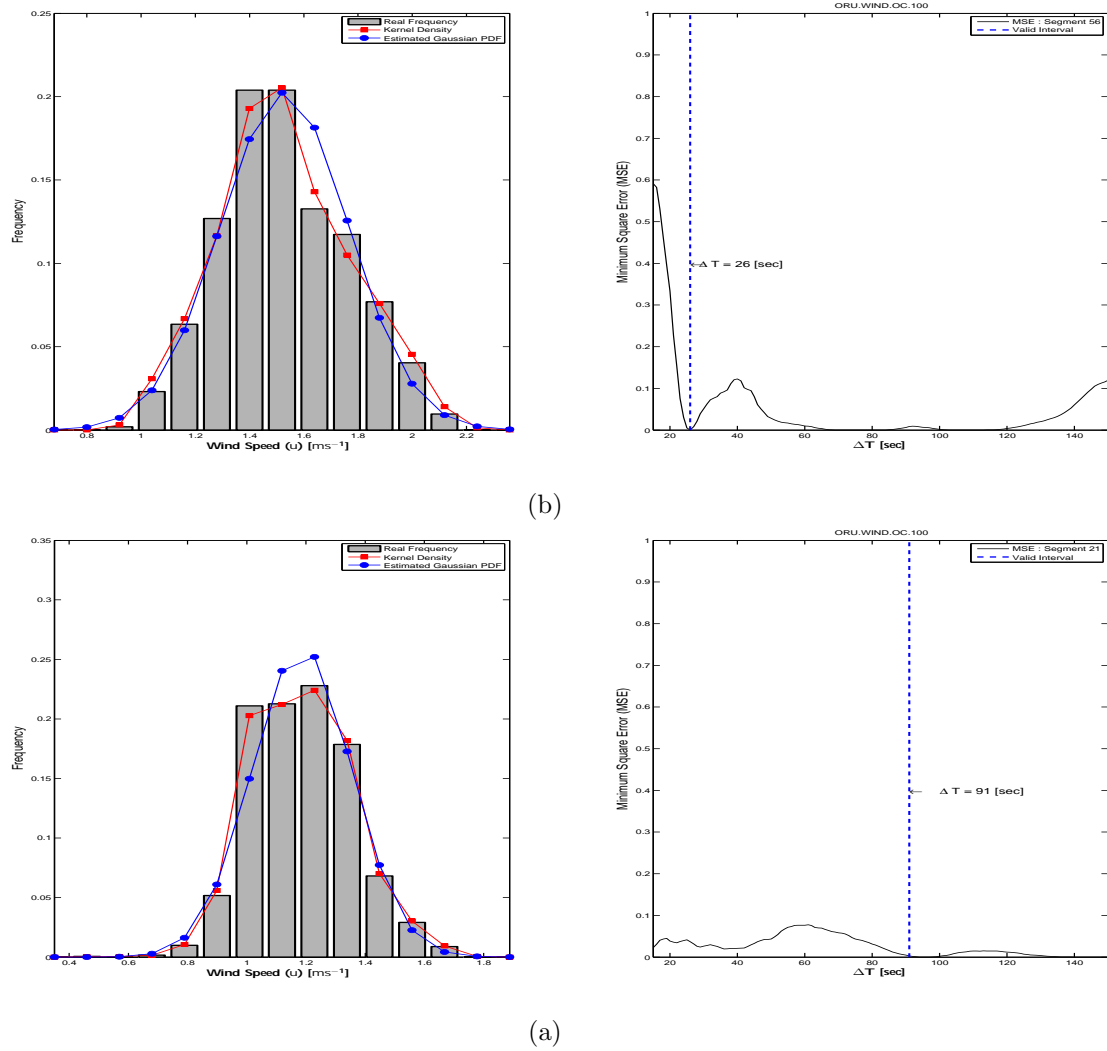


FIGURE 3.1: The empirical density is compared to the estimated distribution using the averaging interval detected by TDMM algorithm. The rows (a) and (b) correspond to ORU.WIND.OS.100 segment 21. and ORU.WIND.OS.150 segment 56. respectively. The line graphs at the right side of figure mimics the whole process of detection while relating the requisite interval ΔT to MSE . The detected interval is marked by the dashed line in blue.

Dataset.	Segment No.	Kurtosis	Skewness	NRMSE	R^2
ORU.WIND.OS.100	21	2.93	0.22	0.800	0.960
ORU.WIND.OS.150	56	2.60	0.23	0.789	0.956
ORU.WIND.OS.100	35	2.79	0.23	0.884	0.987
ORU.WIND.OS.150	37	2.43	0.25	0.775	0.949

TABLE 3.2: Statistical moments and errors for Probability Density Functions compared to the empirical distribution estimated from the kernel density, for ORU.WIND.OS/OS dataset.

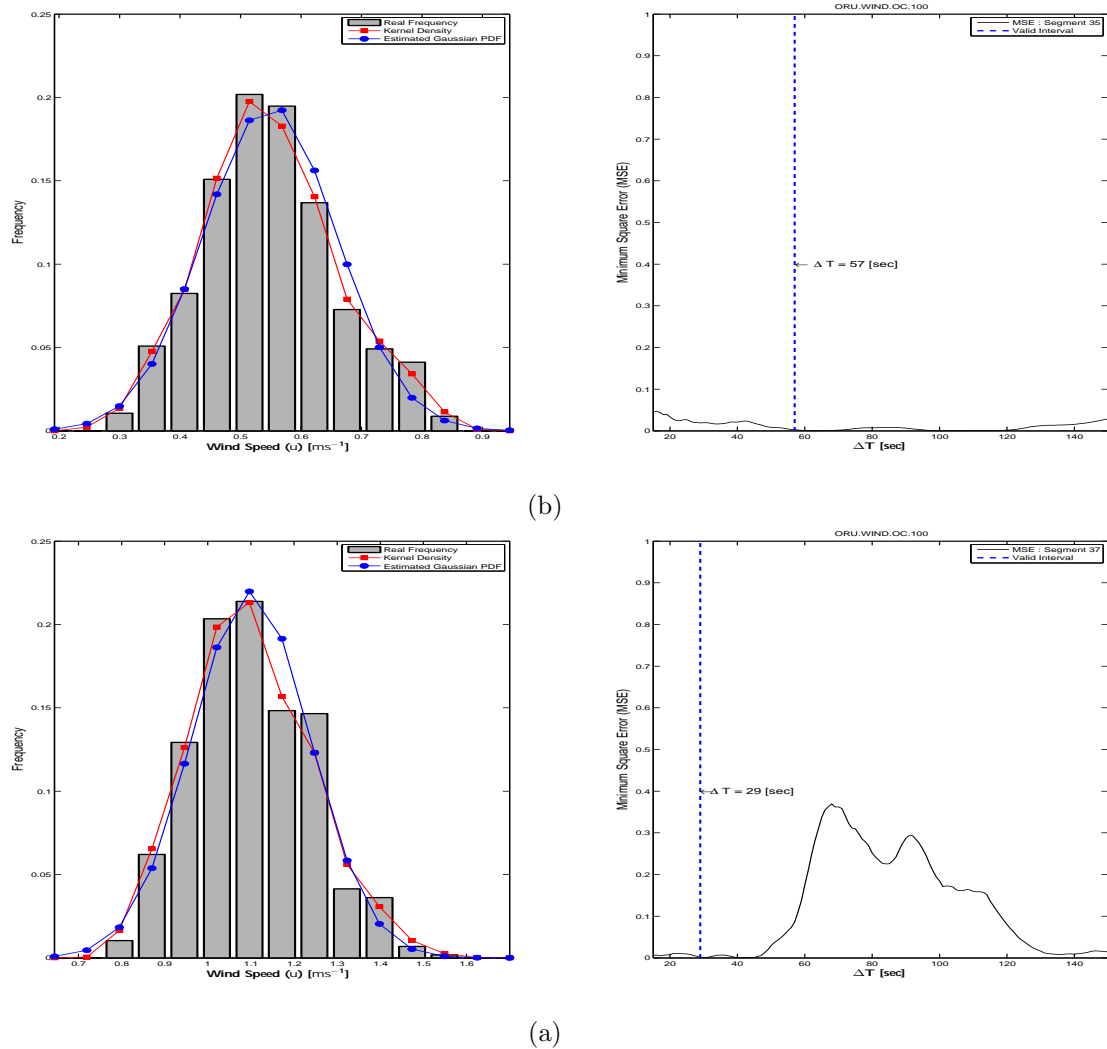


FIGURE 3.2: The empirical density is compared to the estimated distribution using the averaging interval detected by TDMM algorithm. The rows (a) and (b) correspond to ORU.WIND.OC.100 segment 37. and ORU.WIND.OC.150 segment 35. respectively. The line graphs at the right side of figure mimics the whole process of detection while relating the requisite interval ΔT to MSE . The detected interval is marked by the dashed line in blue.

depicted at left sides of the figures. Recall here the values of the skewness and the kurtosis of the subject wind speed segments, shown in Table 3.2 are close to those of Gaussian distribution, that is 0 for the skewness and 3 for the kurtosis. Indeed, according to NRMSE and R^2 goodness of fit values presented in Table 3.2, the Gaussian distribution is a good approximation to these wind speed fluctuations.

However, in chaotic situations TDMM may not succeed to detect a valid interval in a limited amount of time. It is also likely that, the shape of empirical distribution of wind speed fluctuations deviates from Gaussian. Figure 3.3 is exemplified one of the circumstances where TDMM algorithm fails to detect a valid averaging scale. As can

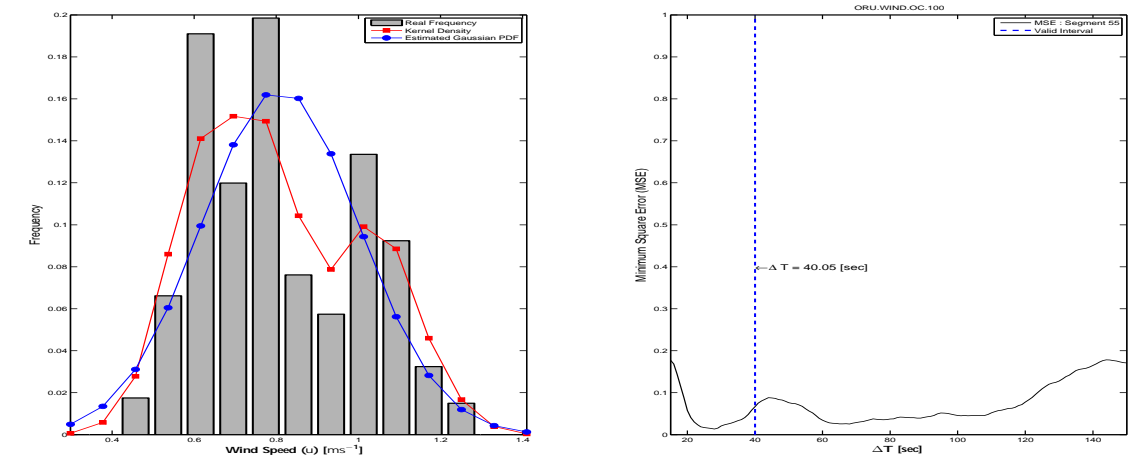


FIGURE 3.3: An example when TDMM failed to detect a valid averaging scale. The averaging scale is set to a predefined value. Kurtosis = 2.06, Skewness=0.35

be seen, The kurtosis and skewness values of wind speed fluctuations are 2.05 and 0.35, respectively, which are apart from Gaussianity. This would be the effect of sudden wind changes (wind gusts) which are pronounced within low wind speed field over short times scales.

The overall performance of the estimated Gaussian distribution models while incorporating the TDMM method for every 5 minutes segments of wind speed data is summarized in Figure 4.8 where the NRMSE (left-column) and R-Square (right-column) values of the estimated Gaussian models are shown in form of Box-whisker plots. Each row entry in the plots corresponds to one of particular ORU.WIND outdoor and indoor data sets. The supporting data comprising of the goodness of fit measures of every single wind speed segment is given in Appendix B through tables B.1 to B.4.

Across the entire data set, the NRMSE for ORU.WIND.OS.100 varies from 2.80–20.30% of the observed range, NRMSE for the ORU.WIND.OS.150 varies from 2.60–21.50% of the observed range, NRMSE for the ORU.WIND.OS.100 varies from 2.90–23.00% of the observed range, and the NRMSE for ORU.WIND.OS.150 varies from 2.40–20.60% of the observed range. Having considered the NRMSE values of all ORU.WIND data sets, the Gaussian distribution model based on the scaling provided by TDMM achieved the highest accuracy of 2.40% and not worth than 23.00% at NRMSE. Meanwhile, in terms of percentage of variance explained by the model, R^2 for ORU.WIND.OS.100 ranges from 71.20–99.30% , R^2 for ORU.WIND.OS.150 ranges from 54.40–99.80%, R^2 for ORU.WIND.OS.100 ranges from 52.40–99.30 % , and R^2 for ORU.WIND.OS.150 ranges from 76.80–99.50%.

The quantitative summary statistics of NRMSE and R^2 of ORU.WIND wind speed data sets are presented in Table 3.3 and Table 3.4, respectively. In Table 3.3, the NRMSE

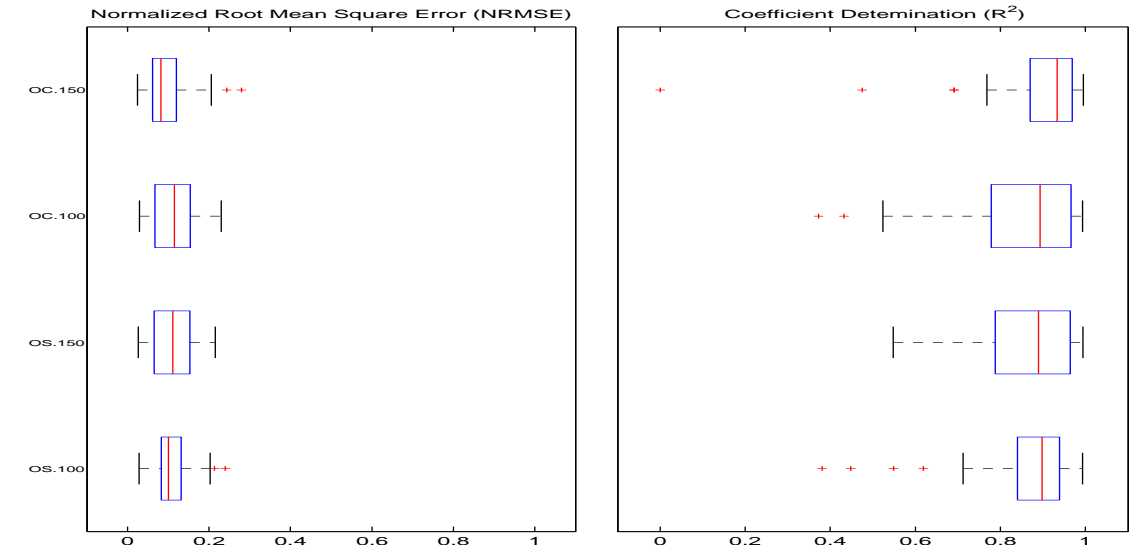


FIGURE 3.4: The overall performance of the estimated Gaussian distribution models while incorporating TDMM for every 5 minutes segments of wind speed data are shown. Each row entry in the plots corresponds to a particular ORU.WIND outdoor data sets where the spread of performance metrics, NRMSE (left-column) and R-Square (right-column), are illustrated in form of Box-whisker plots

Dataset ID	NRMSE < 0.1 [%]	NRMSE < 0.2 [%]	NRMSE \geq 0.2 [%]
ORU.WIND.OS.100	50.82	95.08	4.92
ORU.WIND.OS.150	44.83	94.83	5.17
ORU.WIND.OC.100	38.33	86.67	13.33
ORU.WIND.OC.150	60.34	94.83	5.17
Overall	48.58	92.85	7.15

TABLE 3.3: The overall performance of the estimated Gaussian distribution models for ORU.WIND outdoor data sets are summarized while the NRMSE value of each wind direction segment is compared to the band limits: 0.1 and 0.2. The overall results are shown in percentage.

values of the estimated Gaussian models corresponding to the every wind speed segments of each data set are compared to the band limits: 0.1 and 0.2. On average, 92.85% and 48.58% of the estimated Gaussian models have NRMSE values less than 0.2 and 0.1, respectively. Only 7.15% of approximated Gaussian models have NRMSE values greater than 0.2. Likewise, in Table 3.4, the R^2 values of the estimated Gaussian models are compared to the band limits: 0.6 and 0.7. Overall, 94.93% and 89.45% of the estimated Gaussian models have R^2 values greater than 0.6 and 0.7, respectively. Only 5.07% of approximated Gaussian models have NRMSE values less than 0.6.

Dataset ID	$R^2 > 0.7$ [%]	$R^2 > 0.6$ [%]	$R^2 \leq 0.6$ [%]
ORU.WIND.OS.100	91.80	95.08	4.92
ORU.WIND.OS.150	86.21	93.10	6.90
ORU.WIND.OC.100	86.67	95.00	5.00
ORU.WIND.OC.150	93.10	96.55	3.45
Overall	89.45	94.93	5.07

TABLE 3.4: The overall performance of the estimated Gaussian distribution models for ORU.WIND outdoor data sets are summarized while the R^2 value of each wind direction segment is compared to the band limits: 0.1 and 0.2. The overall results are shown in percentage.

Chapter 4

Statistical Wind Direction Modelling

This chapter is dedicated to the method employed to describe and characterize the temporal variations of wind direction. In order to cope with the the uncertainties in wind direction induced from near surface wind measurements and unpredictability in environment, specific to the task of mobile robot wind mapping, a mixture of Von-Mises distribution is utilized. The background and overview are given in Sections [4.1](#) and [4.2](#). The methodology and theory of the utilized model is presented in section [4.3](#) which followed by the experimental results and the discussions in the next section, [4.4](#).

4.1 Background

Another aspect of wind uncertainties is encapsulated by wind direction. Pursuant to the Raynold decomposition of separating wind average from its fluctuating parts, presented in Section [2.2.2](#), the mean wind vector in 2D horizontal plane is aligned along the longitudinal direction. At a fixed point, the longitudinal wind direction is determined as the predominant observed direction within a specific interval of time. However, wind may experience various scales of variability and generally is described by frequency-distribution models.

Traditionally, wind direction is represented by discrete frequency-distributions, particularly as wind roses (Figure [4.1](#)). Although wind roses are valuable tools to describe and visualize wind direction and speed concurrently, they are prone to the very coarse bins that make them less useful for detailed analysis. Owing to eliminate this short coming, continuous frequency-distribution satisfying the specific requirement of circular

data are utilized to model wind direction statistics. In the wind energy and meteorology

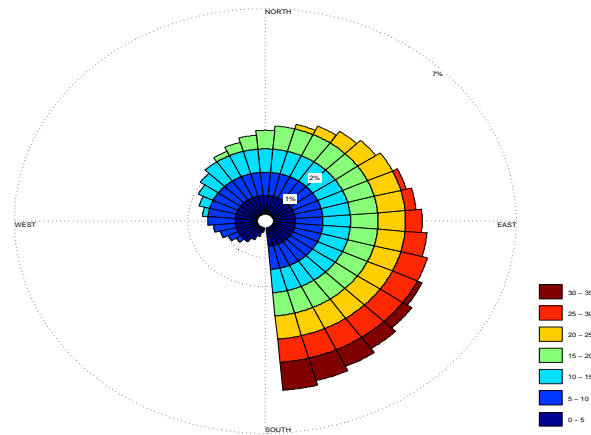


FIGURE 4.1: An example of wind rose describing the directional data

literature, only a limited number of continuous frequency-distributions are available by which wind direction is modelled. Most importantly, Smith [24] decomposes horizontal wind speed into x and y components and obtains wind direction distribution model by transforming their bivariate normal distribution into polar coordinates. In McWilliams et al. [25, 26] wind direction is invoked from the assumption of isotropic Gaussian model for horizontal wind speed. That is, horizontal wind speed is assumed to be uncorrelated and to follow Gaussian distribution along the longitudinal and lateral directions, respectively with non-zero and zero means. In essence, McWilliams model is isotropic since the variance of longitudinal and lateral components are considered to be equal. Wind direction distribution is estimated by transforming wind vector from Cartesian to polar coordinates. Weber generalizes the McWilliams isotropic model and presented anisotropic model in which the variance of longitudinal and lateral components are not presumed to be equal. Carta et al. [27] presents a distribution model based on the mixture of Von-Mises distribution. The Carta et al. undoubtedly is supposed to be the most convenient approach to model wind direction uncertainties as benefiting from the capability of modelling wind regimes with asymmetric and multimodal distributions patterns.

4.2 Requirements

The overall performance of a mobile robot wind mapping task is tightly linked to the precision of the utilized wind direction descriptors so that accurately modelling wind direction is crucial for the fulfilment of the wind mapping process.

Over the moderate temporal scales, it is generally acceptable to consider wind direction as a highly concentrated quantity, but it is not particularly true for MRWM task, where wind direction recorded by mobile robots may involve with considerable variations. In general, there are two major sources for these variations. The first and foremost one does originate from the specific setup of wind measurement using robots where the wind data is recorded at the proximity to the ground surface. As discussed in Chapter 2, under the effect of frictional drag forces exerted from earth, the near surface wind is highly influenced by turbulent fluxes. Near ground turbulence is the phenomenon that most likely causes changes in wind direction. The second cause of uncertainty corresponds to the inherent characteristics of mapping wind with mobile robots operating in complex environments. Mapping wind in real world using robots is to involve with dynamic environment with moving obstacles. This process basically is subject to unpredictable events by which wind direction could be affected. Consequently, it is not uncommon to see changes in wind direction behaviour while mapping wind by mobile robot.

In general, the main course of problem in wind direction modelling within the framework of MRWM is the range and the quality of uncertainties that to be dealt with. Due to the facts mentioned, from statistical perspective, wind direction may possess a multiplicity of modes and it is not possible to well characterize the wind direction variations with only the first two statistical moments in form of unimodal distributions. In particular, under several circumstances the temporal dispersion of wind direction is prone to be multimodal such that the traditional symmetric unimodal distribution functions such as [24] and [25] fail to describe wind direction uncertainties.

In this theThe objective function is formulated as the sum of squared error between the cumulative relative frequencies and the theoretical cumulative distribution function. As a modification to the Carta et al. method, the initialization of parameters is done using the maximum likelihood estimation. This model is with great advantage since it adaptively estimates the number of mixtures and accordingly models the distribution mixture-model parameters.

4.3 Methodology

This Section is mainly guided by [27], unless otherwise stated.

The Von-Mises mixture-model distribution $f_{\Theta}(\theta; \mu, \kappa, \omega)$ is the weighted sum of N components of Von-Mises distributions which is defined by the equation,

$$f_{\Theta}(\theta; \mu, \kappa, \omega) = \sum_{j=1}^N \omega_j f_{\Theta}(\theta; \mu_j, \kappa_j) \quad (4.1)$$

where θ is wind direction measurement, ω_j , $j = 1, 2, \dots, N$ are mixing weights and, $g_{\Theta}(\theta; \mu_j, \kappa_j)$, $j = 1, 2, \dots, N$ are the component Von-Mises densities. The mixing weights ω_j satisfy $\sum_{j=1}^N \omega_j = 1$ and $0 \leq \omega_j \leq 1$. Formally, the complete Von-Mises mixture-model is parametrized by the three mean, concentration and mixing-weight vectors (\mathcal{R}^N) respectively, μ , κ and ω each comprising of N quantities.

The Von-Mises 1-dimensional frequency distribution with mean direction $0 \leq \mu_j \leq 2\pi$ and concentration parameter $\kappa_j \geq 0$ is given by

$$g_{\Theta}(\theta; \mu_j, \kappa_j) = \frac{1}{2\pi I_0(\kappa_j)} \exp(\cos(\theta - \mu_j)) \quad (4.2)$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind and order zero:

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\kappa \cos \theta) d\theta \quad (4.3)$$

The cumulative distribution function of mixture-model of N Von-Mises distribution is given by

$$F_{\Theta}(\theta; \mu, \kappa, \omega) = \sum_{j=1}^N \frac{\omega_j}{2\pi I_0(\kappa_j)} \int_0^{\theta} \exp[\kappa_j \cos(\theta - \mu_j)] d\theta \quad (4.4)$$

Non-linear programming is used to estimate the $3 \times N$ parameters of mixture-model, namely μ, κ and ω . In particular, the mixture-model parameters are estimated in form least square curve fitting problem resolved by Levenberg-Marquardt Method. In particular, Levenberg-Marquardt algorithm, shown in Algorithm 2, was applied to cumulative distribution function (CDF), formulated as the sum of squared differences between cumulative relative frequencies (empirical CDF) and the mixture-model cumulative distribution function given in Equation 4.4 are minimized to estimate the unknown parameters. The relative frequencies of wind direction observations are estimated as the frequencies normalized with the total number of wind recordings. That is, the whole range of 0 to 2π is divided into T direction sectors and the relative frequency is estimated for each of

Input: $f(\cdot)$: Objective function; $y \in \mathcal{R}^n$: Observed data points, $x \in \mathcal{R}^n$: Independent variable; $p \in \mathcal{R}^m$: Initial parameters; $W \in \mathcal{R}^{n \times n}$: Residuals weights as diagonal matrix

Output: $p \in \mathcal{R}^m$: fit parameters

```

begin
  % initialization;
   $\lambda := \tau \max(a_{ii})$ 
   $\nu := 2$ ;
   $A := J^T W J$ ;  $B := J^T W [y - f(x; p)]$ ;
   $\chi^2(x; p) = \frac{1}{2} [y - f(x; p)]^T W [y - f(x; p)]$ ;
  while ( $Iter < Iter_{max}$  and not found) do
    Iter := Iter + 1;
    Solve  $(A + \lambda I)h_{lm} = B$ ;
    if ( $\|h_{lm}\| < \epsilon$  or  $\|\chi^2(x; p)\| < \epsilon$  or  $\|B\| < \epsilon$ ) then
      | found := true;
    else
      |  $A := J^T W J$ ;  $B := J^T W [y - f(x; p)]$ ;
      |  $p_{new} = p + h_{lm}$ ;
      |  $\chi^2(x; p) = \frac{1}{2} [y - f(x; p)]^T W [y - f(x; p)]$ ;
      |  $\chi^2(x; p_{new}) = \frac{1}{2} [y - f(x; p_{new})]^T W [y - f(x; p_{new})]$ ;
      |  $\rho = \frac{\chi^2(x; p) - \chi^2(x; p_{new})}{\frac{1}{2} h_{lm}^T (\lambda h_{lm} - B)}$ ;
      | if  $\rho > 0$  then
      | |  $p := p_{new}$ ;
      | |  $\lambda := \lambda \times \max\{1/3, 1 - (2\rho - 1)^3\}$ ;
      | |  $\nu := 2$ ;
      | else
      | |  $\lambda := \lambda \times \nu$ ;
      | |  $\nu := 2 \times \nu$ ;
      | end
    end
  end
end

```

Algorithm 2: Levenberg-Marquardt Alogrithm (modified from [28] and [29])

theses sector. Here, the number of sectors should necessarily be greater or equal than number of mixtures.

For the direction sector “ $\theta_{m-1} - \theta_m$ ” the relative frequencies is obtained by

$$p_m = \frac{n_m}{\sum_{i=1}^T n_i} \quad (4.5)$$

where n_i is number of observations falling in sector i . Empirical CDF of relative frequencies “ $\theta_0 - \theta_1, \theta_1 - \theta_2, \dots, \theta_{T-1} - \theta_T$ ” is then estimated by summing up of relative frequencies.

$$P_1 = p_1, \quad P_2 = P_1 + p_2, \quad \dots, \quad P_T = P_{T-1} + p_T \quad (4.6)$$

Accordingly, the least square objective function is formulated by putting mixture-model CDF Equation 4.4 along with empirical CDF.

$$S = \sum_{K=1}^T \left\{ P_k - \sum_{j=1}^N \frac{\omega_j}{2\pi I_0(\kappa_j)} \int_0^\theta \exp[\kappa_j \cos(\theta - \mu_j)] d\theta \right\}^2 \quad (4.7)$$

The least square problem is to minimize S subject to the inequality constraints

$$\begin{aligned} 0 \leq \mu \leq 2\pi, \quad \kappa \geq 0 \\ \sum_{j=1}^N \omega_j = 1 \quad \omega \geq 0 \end{aligned} \quad (4.8)$$

Likewise to any convergence based problem, the initial estimate of the parameters in Von-Mises mixture-model is a key point to consider. In order to set up the starting parameters, first the observed wind directional data is distributed over N consecutive sectors, say " $\theta_0 - \theta_1, \theta_1 - \theta_2, \dots, \theta_{N-1} - \theta_N$ ", and then the starting parameters are estimated for every of these sectors using Maximum Likelihood Estimate (MLE). Formally, the initial parameter values for the m th component of mixture model corresponding to the sector m , i.e $\theta_{m-1} - \theta_m$, are obtained as follow:

1. The starting value of direction mean (μ_m) is estimated as the angular average of the recorded directions grouped into the sector m

$$\mu_m = \begin{cases} \tan\left(\frac{\bar{S}_m}{\bar{C}_m}\right), & \bar{C}_m > 0, \bar{S}_m \geq 0. \\ \frac{\pi}{2}, & \bar{C}_m = 0. \\ \arctan\left(\frac{\bar{S}_m}{\bar{C}_m}\right) + \pi, & \bar{C}_m < 0. \\ \pi, & \bar{C}_m = -1, \bar{S}_m = 0. \\ \arctan\left(\frac{\bar{S}_m}{\bar{C}_m}\right) + 2\pi, & \bar{C}_m > 0, \bar{S}_m < 0. \\ \frac{3\pi}{2}, & \bar{C}_m = 0, \bar{S}_m < 0. \end{cases} \quad \begin{aligned} \bar{S}_m &= \frac{1}{n_m} \sum_{j=1}^{n_m} \sin \theta_j, \\ \bar{C}_m &= \frac{1}{n_m} \sum_{j=1}^{n_m} \cos \theta_j \end{aligned} \quad (4.9)$$

2. The starting value of concentration parameter (κ_m) is obtained as

$$\kappa_m = \{23.2904 - 16.8617(\bar{C}_m^2 + \bar{S}_m^2)^{0.25} - 17.4749 \exp[-(\bar{C}_m^2 + \bar{S}_m^2)]\}^{-1} \quad (4.10)$$

3. Mixture weight coefficient (ω_m) is initialized uniformly with respect to other mixing weights.

$$\omega_m = \frac{n_m}{N} \quad (4.11)$$

4.4 Experimental Results

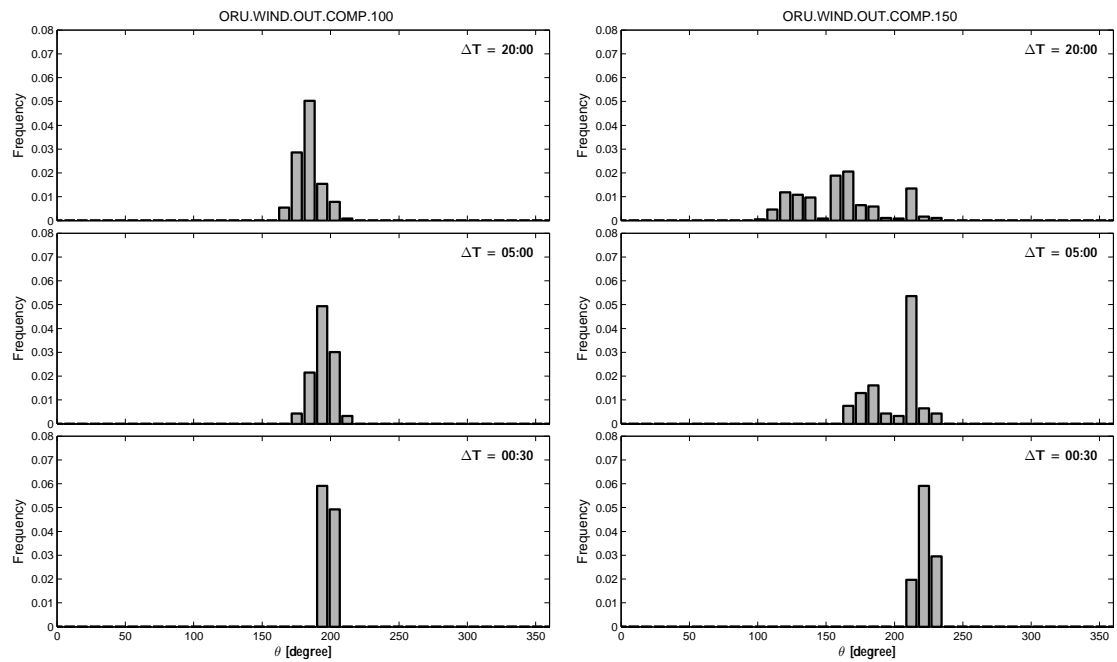
4.4.1 Wind Direction Analysis

Near surface wind direction data collected from mobile robotic platform is analysed to investigate its overall temporal variability. The better understanding of wind direction behaviour, particularly its variability, will definitely help to interpret and model wind direction under various experimental settings. Here, the emphasis of wind direction analysis is on the *ORU.WIND* dataset as it consists of wind data recorded at different measurement scenarios within relatively long periods. In particular, wind data collected in outdoor and indoor setting at 100 and 150 setting are studied.

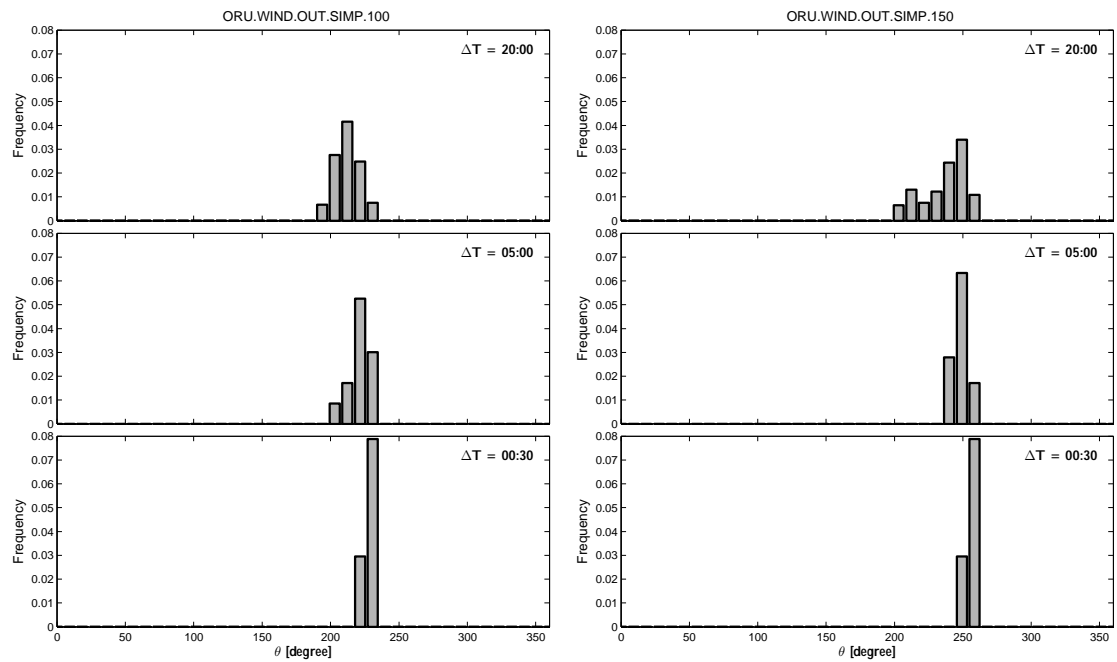
Due to the specific characteristic of wind direction data of being periodic at the interval of $0 - 360^\circ$, the use of linear second moment statistics is convoluted with the inherent discontinuity. In order to avoid this difficulty, in this thesis work, the wind direction variability is quantified and studied using the the metric of concentration parameter κ .

Primarily, wind direction variability is investigated by temporal scales in which wind direction data are averaged. In particular, the dispersion of wind direction time series are analysed using histogram plots at three specific time intervals: 30 seconds, 5 minutes and 20 minutes. The histograms are shown in Figures 4.2 and 4.3, respectively for *ORU.WIND* outdoor and indoor datasets. Overall, its apparent that the spread of wind direction increases with the length of time intervals. However, compared to *ORU.WIND* outdoor, the wind direction of *ORU.WIND* indoor has too much dispersion which basically is justified with its very low speeds. What is interesting in the frequency of distribution of wind direction time series is that the prevailing directions within the examined intervals are almost similar, even with indoor case which has too much dispersion.

Figure 4.4 draws parallel with the results obtained from dispersion plots as it relates the concentration parameter κ to the averaging scale ΔT . Expectedly, the concentration of wind direction stands at its highest values within shorter intervals. That is, in shorter time scales, wind direction tends to be unimodal and more concentrated at a prevailing direction, whereas over the longer times scales, wind direction shows greater variability, prone, the multiplicity of modes. The more insights into the wind direction variability analyses is produced by Figure 4.5. That is, wind direction concentration parameter, as the measure of magnitude of variability, is quantified against with different wind speed levels. Further to the observations in the previous chapter of the most commonly detected valid averaging intervals, 40-seconds intervals of wind data is used to generate wind direction concentrations. The Figure 4.5 is quite revealing in a sense that there

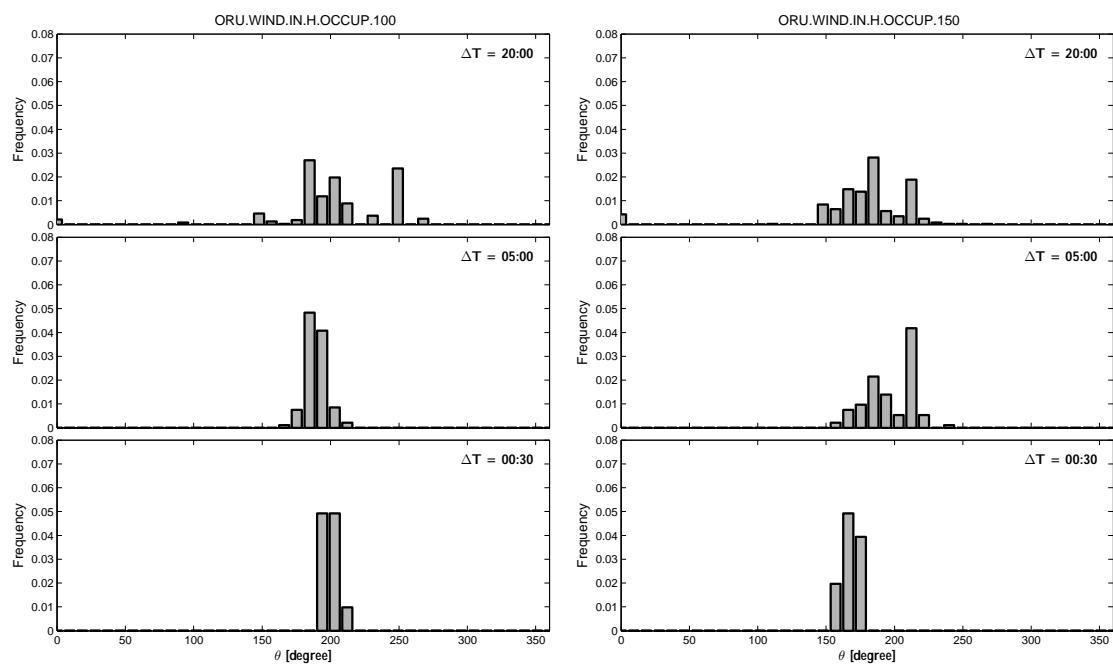


(b)

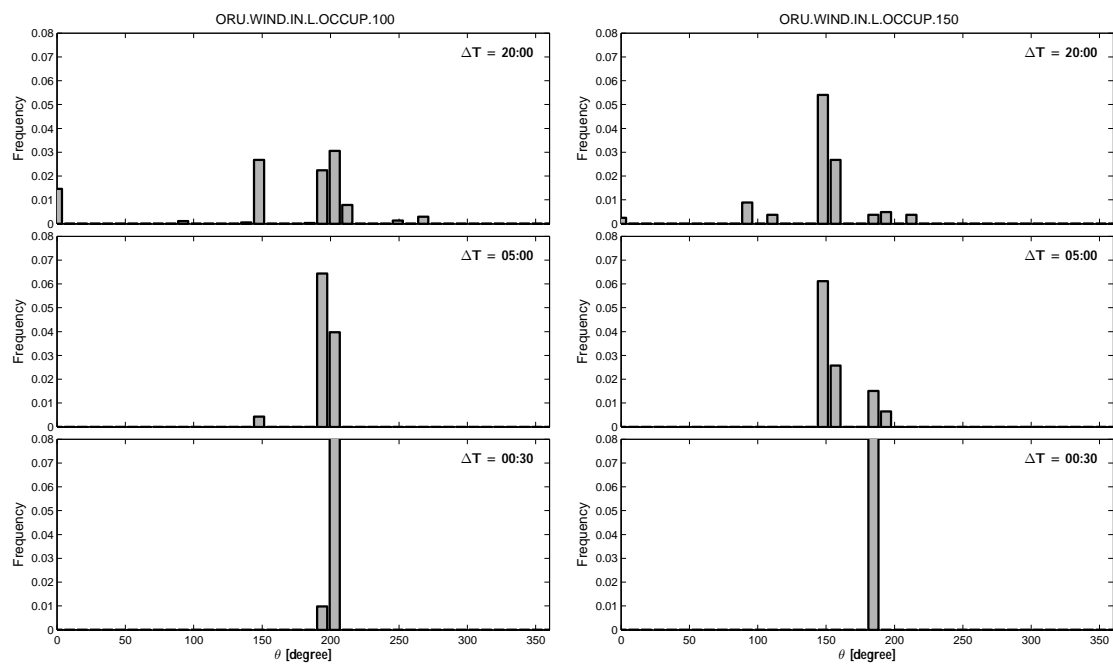


(a)

FIGURE 4.2: Histograms of ORU.WIND outdoor wind direction time series at 30 seconds, 5 minute and 20 minutes intervals; (a): Simple terrain (b): Complex terrain



(b)



(a)

FIGURE 4.3: Histograms of ORU.WIND indoor wind direction time series at 30 seconds, 5 minute and 20 minutes intervals; (a): Low occupancy (b): High occupancy

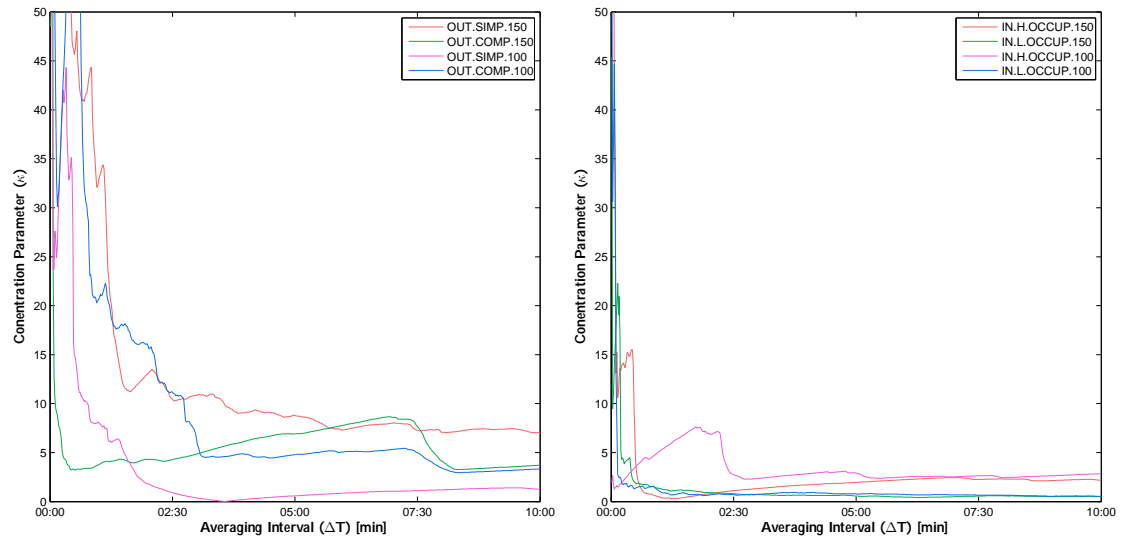


FIGURE 4.4: The dependence of wind direction variability (represented as concentration parameter κ) on the averaging interval length for *ORU.WIND* outdoor (Left) and indoor (right) datasets.

is direct relationship between wind direction variability and wind speed. Wind direction concentration increases with increasing wind speeds. Indeed, taking all datasets together, the highest rate of wind direction variability is associated with indoor wind dataset bearing the lowest wind speed range. This observation is in agreement with *ORU.WIND* indoor histogram plots in Figure 4.3.

4.4.2 Wind Direction Modelling

The *ORU.WIND* outdoor dataset is used to assess the reliability and performance of Von-Mises mixture model for statistical modelling of wind direction distributions. Following the procedure adapted for wind speed fluctuations modelling, wind direction time series are split into equally 5 minutes segments. But, the mixture model is only applied to the parts of segments which are identified as being a valid averaging intervals by TDMM algorithm.

The preference of the number of sectors and the number of mixture components are the key issues in configuration of Von-Mises mixture-model. Apart from the computation complexity involved, the more accurate empirical distribution models are obtained from sufficiently large number of sectors. However, the large number of sectors does not necessarily result the optimum solution and even could fail to find any solution at all. This mainly occurs because too many number sectors leave several number of sectors empty leading optimization to singularity matrix problems. Accordingly, in order to

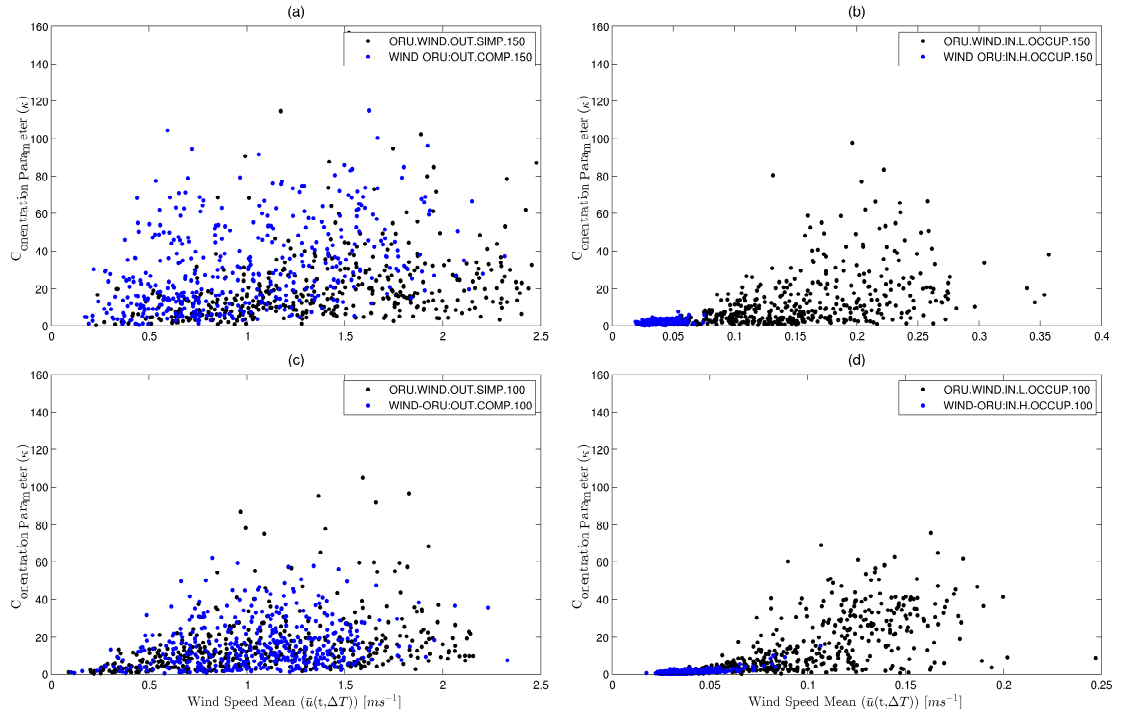


FIGURE 4.5: The dependence of wind direction variability (represented as concentration parameter κ) on 40-second wind averages for ORU.WIND outdoor (left column) and indoor (right column) datasets.

avoid such difficulty and any user intervention, in the current thesis work, the number of mixtures is specified by number of detected peaks of experimental frequency distribution.

To evaluate the performance of estimated distribution model compared to empirical frequency distribution of wind direction data, three kinds of statistical test of goodness of fit are applied:

1. Normalized Root Mean Square Error (NRMSE)

The Normalized Root Mean Square Error (NRMSE) accumulates the magnitudes of the residuals in predictions normalized by the range of observed values of a variable being predicted, given by:

$$NRMSE = \frac{\left[\sum_{k=1}^T \frac{1}{T} (p_k - f_{\Theta}(\theta_k))^2 \right]^{\frac{1}{2}}}{\max(p_k) - \min(p_k)} \quad (4.12)$$

where T is the number of sectors, p_k is empirical frequency distribution value at sector k and $f_{\Theta}(\theta_k)$ is the estimated probability distribution function at the center of the sector.

2. R^2

R squared test or coefficient determination is applied to the empirical and estimated cumulative densities of wind directions as given by

$$R^2 = 1 - \frac{\sum_{k=1}^T (P_k - F_{\Theta}(\theta_k))^2}{\sum_{k=1}^T (P_k - \overline{F_{\Theta}(\theta)})^2} \quad (4.13)$$

where T is the number of sectors, P_k is empirical cumulative distribution value at sector k and $F_{\Theta}(\theta_k)$ is the estimated cumulative distribution value at the center of the sector. $\overline{F_{\Theta}(\theta)}$ is referred to the mean of estimated cumulative distribution function. In R squared test, the larger value of R^2 values ($R^2 \in [0, 1]$) corresponds to better fit between empirical and estimated data.

3. Kolmogorov-Smirnov (KS)

The Kolmogorov-Smirnov test is defined as the maximum error between the empirical and estimated cumulative distribution values given by:

$$KS = \max |P_k - F_{\Theta}(\theta_k)| \quad (4.14)$$

where T is the number of sectors, P_k is empirical cumulative distribution value at sector k and $F_{\Theta}(\theta_k)$ is the estimated cumulative distribution value at the center of the sector. In KS test of goodness of fit, the smaller errors corresponds to better fit between empirical and estimated data.

In Figure 4.6, an example of single peak unimodal wind direction distribution is shown and compared to the estimated Von-Mises mixture distribution model. Here, the estimated model comprises of a single component with 37 sectors. The goodness-of-fit statistics obtained for RMSE, R^2 and K-S metrics are 0.064, 0.0999 and 0.024, respectively. Obviously, the estimated models is a good fit to the empirical frequency distribution.

Figure 4.7 illustrates an example of wind direction ORU.WIND.OC.100 distribution for a bimodal case, particularly segment 13, compared with a Von-Mises mixture distribution model of 2 components. The number of mixtures is adaptively identified by number of peakiness in empirical frequency distribution. It is obvious that the the mixture model is a good approximate of empirical distribution. In meanwhile, the performance of the Von-Mises mixture model is studied in function of different number of sectors different number sectors at table 4.1. In particular, sector factors 17, 27, 37 and 47 are utilized to construct the Von-Mises mixture model and the corresponding obtained values of

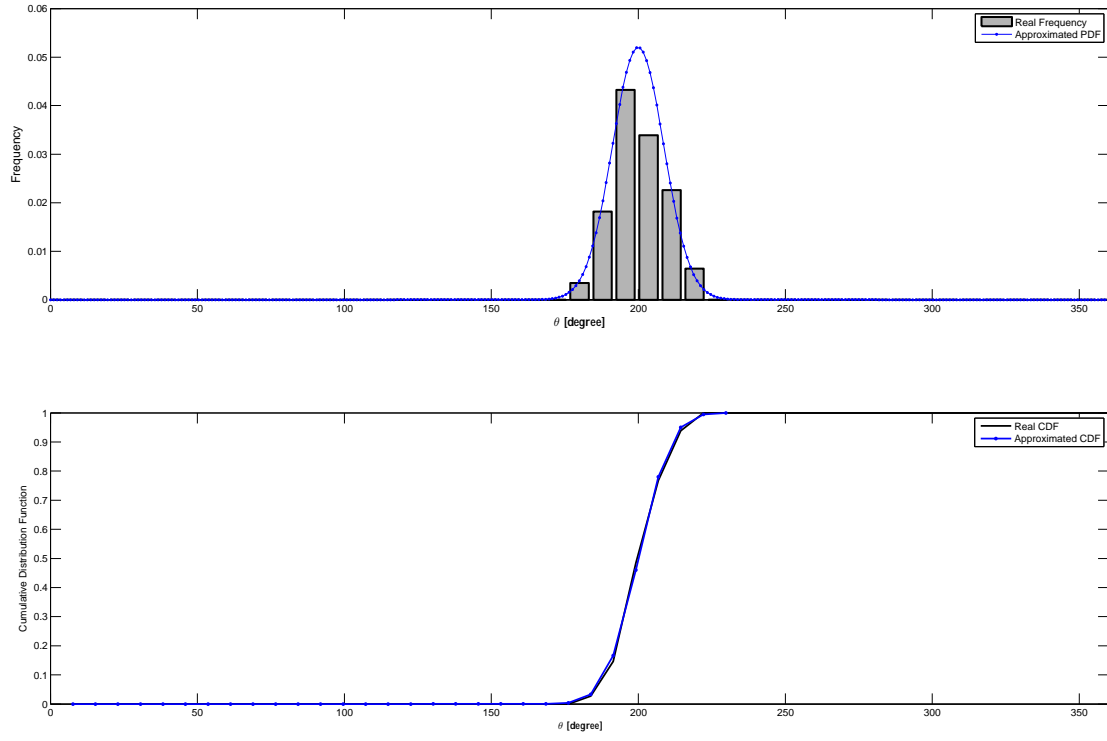


FIGURE 4.6: Top: Estimated Von-Mises mixture distribution plotted against empirical frequency distribution at ORU.WIND.OC.100 segment 54 ($T=47$). Bottom: Cumulative distribution function of estimated mixture model against the empirical cumulative frequency distribution. (Number of mixtures specified by the number of detected peaks)

T	N	ΔT [sec]	μ [rad]	κ	ω	RMSE	R^2	K-S
17	1	22	2.769	3.37	1.00	0.104	0.993	0.080
27	2	22	2.291	31.63	0.48	0.076	0.999	0.033
			3.194	23.72	0.52			
37	2	22	2.297	35.64	0.49	0.054	0.999	0.049
			3.211	22.08	0.51			
47	2	22	2.306	26.47	0.50	0.039	0.999	0.024
			3.233	23.25	0.50			

TABLE 4.1: The statistical moments and the errors for the approximated Von-Mises mixture model with different numbers of sectors compared to the experimental frequency distribution for ORU.WIND.OC.100 segment 13. R^2 and K-S test were applied to Cumulative distribution function, while RMSE was applied to probability distribution function.

goodness-of-fit statistics, R^2 and K-S, are presented in Table 4.1. It can be observed that performance estimation is directly related to the number of sectors, such that by increasing the number of sectors the performance of fitting is improved, decreasing RMSE and KS while increasing R^2 . Nonetheless, the higher number of sectors is computationally involved and improvements are not noticeable sectors greater than 27.

The overall performance of the estimated Von-Mises mixture distribution models while incorporating the TDMM method for every 5 minutes segments of wind speed data is

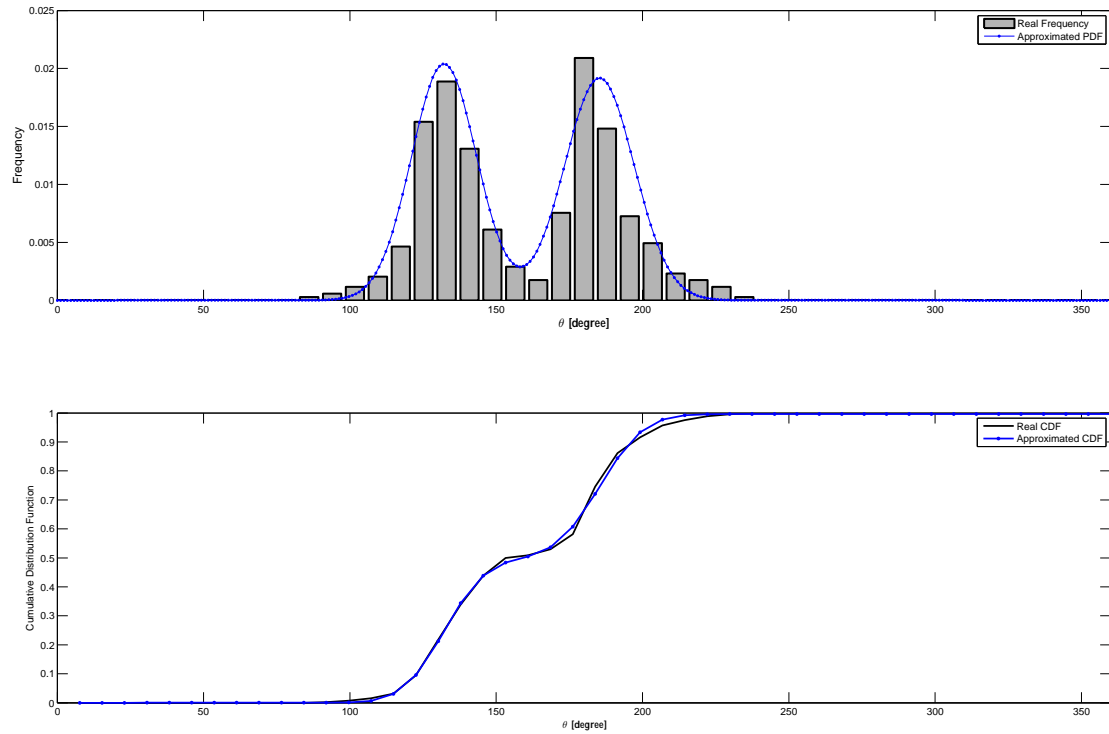


FIGURE 4.7: Top: Estimated Von-Mises mixture distribution plotted against empirical frequency distribution at ORU.WIND.OC.100 segment 13 ($T=47$). Bottom: Cumulative distribution function of estimated mixture model against the empirical cumulative frequency distribution. (Number of mixtures specified by the number of detected peaks.)

summarized in Figure 4.8 where the NRMSE (left-column) and R-Square (right-column) values of the estimated Gaussian models are shown in form of Box-whisker plots. Each row entry in the plots corresponds to one of particular ORU.WIND outdoor and indoor data sets. The supporting data comprising of the goodness of fit measures of every single wind speed segment is given in Appendix C through tables C.1 to C.4.

Across the entire data set, the NRMSE for ORU.WIND.OS.100 varies from 1.10–12.50% of the observed range, NRMSE for the ORU.WIND.OS.150 varies from 1.20–13.90% of the observed range, NRMSE for the ORU.WIND.OC.100 varies from 0.80–12.80% of the observed range, and the NRMSE for ORU.WIND.OC.150 varies from 1.00–12.70% of the observed range. Having considered the NRMSE values of all ORU.WIND data sets, the Von-Mises mixture distribution model achieved the highest accuracy of 0.80% and not worth than 13.90% at NRMSE. Meanwhile, in terms of percentage of variance explained by the model, R^2 for ORU.WIND.OS.100 ranges from 72.00–99.70%, R^2 for ORU.WIND.OS.150 ranges from 73.30–99.70%, R^2 for ORU.WIND.OC.100 ranges from 78.20–99.80%, and R^2 for ORU.WIND.OC.150 ranges from 81.80–99.80%.

The quantitative summary statistics of NRMSE and R^2 of ORU.WIND wind speed data

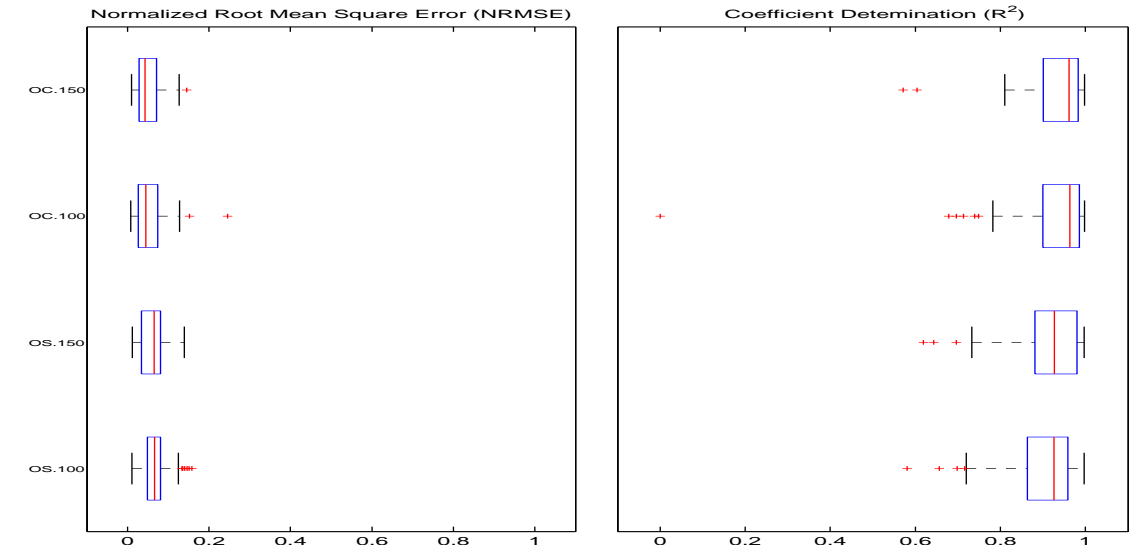


FIGURE 4.8: The overall performance of the estimated Von-mises mixture distribution models for ORU.WIND outdoor wind direction data are illustrated. Each row entry in the plots corresponds to a particular ORU.WIND outdoor data set where the spread of performance metrics, NRMSE (left-column) and R-Square (right-column), are illustrated in form of Box-whisker plots. For each data set, the mixture models were estimated per every 5-minutes segments while only considering the averaging scale supplied by TDMM.

Dataset ID	NRMSE < 0.1 [%]	NRMSE < 0.2 [%]	NRMSE \geq 0.2 [%]
ORU.WIND.OS.100	83.61	100.00	0.00
ORU.WIND.OS.150	86.21	100.00	0.00
ORU.WIND.OC.100	88.33	98.33	1.67
ORU.WIND.OC.150	93.10	100.00	0.00
Overall	87.81	99.58	0.42

TABLE 4.2: The overall performance of the estimated Von-Mises mixture models for ORU.WIND outdoor data sets are summarized while the NRMSE value of each wind direction segment is compared to the band limits: 0.1 and 0.2. The overall results are shown in percentage.

sets are presented in Table 4.2 and Table 4.3, respectively. In Table 4.2, the overall performance of the estimated Von-Mises mixture models for ORU.WIND outdoor data sets are summarized while the NRMSE value of each wind direction segment is compared to the band limits: 0.1 and 0.2. On average, 99.58% and 87.81% of the estimated Von-Mises mixture models have NRMSE values less than 0.2 and 0.1, respectively. Only 0.42% of approximated mixture models have NRMSE values greater than 0.2. Likewise, in Table 4.3, the R^2 values of the estimated mixture models are compared to the band limits: 0.8 and 0.9. Overall, 89.51% and 70.49% of the estimated Von-Mises mixture models have R^2 values greater than 0.8 and 0.9, respectively. Only 10.49% of approximated mixture models have NRMSE values less than 0.8.

Dataset ID	$R^2 > 0.9$ [%]	$R^2 > 0.8$ [%]	$R^2 \leq 0.8$ [%]
ORU.WIND.OS.100	65.57	86.89	13.11
ORU.WIND.OS.150	65.52	87.93	12.07
ORU.WIND.OC.100	75.00	86.67	13.33
ORU.WIND.OC.150	75.86	96.55	3.45
Overall	70.49	89.51	10.49

TABLE 4.3: The overall performance of the estimated Von-Mises mixture models for ORU.WIND outdoor data sets are summarized while the R^2 value of each wind direction segment is compared to the band limits: 0.8 and 0.9. The overall results are shown in percentage.

Chapter 5

Wind Mapping

The main purpose of this chapter is to introduce and describe the proposed framework for the wind mapping task. It starts with a general overview of the existing spatial prediction methods in Section 5.1. Different types of spatial interpolation methods and their attributes briefly are explained. Section 5.2 is dedicated to the theory and the methodology involved with of the proposed wind mapping framework. And finally in Section 5.3, the experimental results are presented.

5.1 Overview

5.1.1 Background

Besides the temporal uncertainties observed in near-surface wind signal, it is subject to spatial variability. In general, within the short-scale spatial domains, wind is considered as a highly correlated phenomenon, featuring statistical dependency with local points in close neighbourhood. The closer the points are, the stronger the statistical dependency is.

The spatial variations of wind resources is an important factor in many applications. In particular, within the task of SGDM, the accurate construction of gas distribution model closely depends upon a realistic interpretation of wind spatial variability. Typically, the wind spatial variations are mapped through girded wind information, so-called wind maps. Wind maps have been utilized to describe the wind patterns of areas not covering by the wind measurement.

In meteorology and the field of wind energy, the spatial mapping of unsampled wind resources generally is constructed by spatial interpolation. Spatial interpolation is referred to a series of techniques by which the unsampled data is estimated using a limited amount of measurements at nearby locations. The spatial interpolation methods applied to environmental data is generally classified into two main groups: deterministic and statistical methods, commonly know as geostatistical methods.

Deterministic interpolation methods create spatial maps from measured points using pure mathematical functions. Based on the heuristic of the function employed, there are several methods available. Trend Surface Mapping (TSM), Inverse Distance Weighting (IDW) and Radial Based Function (RBF) may be identified as the three commonly used deterministic interpolation techniques.

1. *Trend Surface Mapping (TSM)*, also known as Trend Surface Analysis(TSA), is an interpolation method in which mapped data is estimated by expansion of low-order polynomial of spatial coordinates. Based on the governing scale of the map grid nodes involved in interpolation process, TSM is classified into Global Polynomial Interpolation (GPI) and Local Polynomial Interpolation (LPI). As the name suggests, GPI procedures incorporate the entire set of scattered grid nodes, whereas LPI methods only use limited number of nodes in proximity to estimate the polynomial coefficients.
2. *Inverse Distance Weighting (IDW)* is assumed as one of most commonly used deterministic interpolation techniques in spatial environmental mapping. The main idea behind IDW is that the similarity of a pair of observations is bounded to their distance, as the scatter points in nearby are more similar compared to the ones in far distance. IDW performs the interpolation of the unsampled scattered points by a linear weighted combination of sampled data points, where the weights are inversely related to the distance to sampling locations.
3. *Radial Based Function (RBF)* is referred to a series of exact deterministic interpolation techniques which aims to fit a surface through all the sampled scatter points. Based on the type of function employed, there are several numbers of RBF models available in literature. The most commonly RBF models in wind data interpolation may include: Completely Regularized Spline (RBF-CRS), Spline With Tension (RBF-SWT), Thin Plate Spline (RBF-TPS), Multiquadric (RBF-M) and Inverse Multiquadratic (RBF-IM)).

Geostatistics is a subdivision in statistical science which basically is involved with spatial or spatiotemporal stochastic processes, originally rooted in prediction of probability

distributions of ore grades for mining operations. Geostatistics can be viewed as an extension of deterministic spatial interpolation techniques which employs the concept randomness in sampling of scatter points.

Kriging techniques, also known Gaussian process regression, are the mostly used geostatistical interpolation methods. In a sense, Kriging is very similar to IDW where it utilizes a linear weighted average of sampled scatter points to interpolate spatial maps. Nevertheless, it expands upon IDW as it combines the idea of inverse distance heuristic with the spatial correlation of sampled points, so-called semivariogram. In fact, semivariogram lies at the heart of Kriging by relating semivariance to separation distances.

On the basis of the methods of estimating the averaging weights, several Kriging techniques have been proposed, most importantly Simple Kriging, Ordinary Kriging and Kriging with trends, also known as Universal Kriging. In particular, the main difference among the proposed Kriging methods so far, is rooted in their stochastic and stationarity assumptions. In Simple Kriging, all the sampled points are assumed stationary at first moment with a known value for entire scatter points, while the Ordinary Kriging is only first moment stationary at local neighbourhood. Kriging with trend, the first moment of nearby observations is fitted by a general polynomial trend model, such as linear trend model.

5.1.2 Requirements

Geostatistical interpolation methods or, without loss of generality, Kriging predictors are supposed to be the optimal statistical predictor which have been widely used in spatial interpolation mapping tasks. Nonetheless, the efficiency of Kriging methods is coupled with the quality and density of available observations. The main issue in Kriging is rooted into fitting variogram models which is computationally demanding and time consuming. In addition, fitting variogram is inefficient in the cases with too few number of sample points.

In general, the classical interpolation methods, so far have been employed for spatial wind mapping, are basically inapplicable to the tasks when continuously updating of the spatial map is required. Particularly, in a typical MRWM task supposedly to build spatial maps within the time constraints imposed by specific requirements of SGD, Kriging is deemed to be inefficient.

Thus by, having all things considered, in the current thesis work we propose an online spatial interpolation approach applicable to mobile robot SGDM. The proposed method is powered by Linear Kalman filter, which has the potential advantages of producing

optimal estimates maps in real time. The Kalman filter algorithm estimates the optimal map recursively which is tailored to a class of state space model estimation. The spatial domain subject to the wind mapping process is taken as a high-dimensional space model which recursively is updated by the streaming input observations.

Kalman filter particularly is chosen since, first Kalman filter is a linear discrete-time, finite dimensional model that is well suited to digital computer or digital signal processor. Second, the mathematical formulation is in terms of state-space model. Forth, it is a recursive algorithm where the output is updated using previous state estimate and new observations.

5.2 Methodology

5.2.1 Problem Formulation

The mapping problem intended to be solved involves with recursively estimation of the optimal map of a local wind field in a real time manner. That is, provided a sequence (history) of uncertain observations over a range of discrete time intervals, the main goal is to predict map estimates stochastically. Notice that within the framework of MRWM, the observation may be referred to any components of measured wind data, i.e. wind speed and wind mean.

Assume that the observations $z_{x_1;t_1}, \dots, z_{x_n;t_n}$ are collected by the robot within an observable and spatially stochastic process, where $x_i \in m$ are some specific location in the spatial domain m and $t \in \{1, 2, \dots\}$ is a discrete index of times. As convenient to probabilistic robotics, here the spatial domain m at time stamp t is represented in form of 2-dimensional Euclidean evenly grid spaced map. Put simply, m is a location based map of wind field which could be viewed as a n-dimensional dynamic state space vector. That is the mapping process considered has a finite number of measurable objects that are collected in a particular state vector m_t .

$$m_t = \begin{bmatrix} m_{x_1;t} \\ m_{x_2;t} \\ \cdot \\ \cdot \\ m_{x_n;t} \end{bmatrix} \quad (5.1)$$

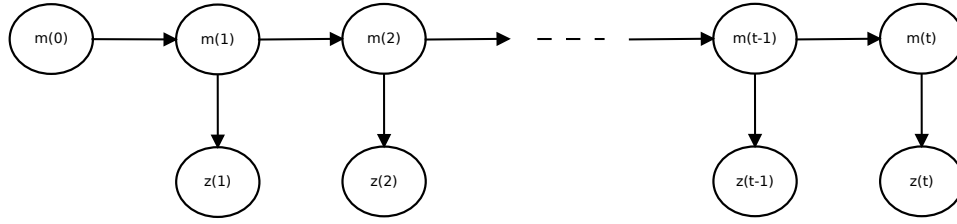


FIGURE 5.1: The stochastic mapping process as a graphical model. Each vertical slice represents a time step. The top row node in each slice represents the map state $m(t)$ and the bottom row node in each slice represents the observable variable $z(t)$

Analogously to linear dynamic systems, the problem of estimation of wind map m over a range of discrete time intervals could be referred to a simple form of linear discrete space model in which no control action is present. Put simply, the values of map state vector m_t at time stamp t does follow the Markov stochastic dynamics as only they are linearly dependent on previous subsequent map state.

To be illustrative, Figure 5.1 shows the generic graphical model of the stochastic mapping process based on the assumption of discrete space model. Typically, there is a sequence of observations $\{z_1, \dots, z_{t-1}, z_t\}$ and the main objective is to determine $p(m_t|z_t)$, called the map posterior belief at time stamp t . Obviously, the values for $p(z_t|m_t)$ are known and it is assumed at given time t , the observation z_t is independent of the rest of observation history. Here the Markov property (first order) is to hold such that the belief of current state is only depends on the previous state. In general, at discrete time stamp t the true estimate of m_t is observed by the measurement vector z_t and eventually transmitted to the next state map model.

Mathematically, the linear state space model is given by dynamic and observation models.

1. *Dynamic Model*

Dynamic model describes the stochastic dynamics of the system and how it transmits over time intervals. Here, it is assumed that the dynamics model of map does not change and is persistent over time. That is, dynamics models is identically linear with no control inputs involved, and in a sense the map is static with respect to state transition.

$$p(m_t|m_{t-1}) = p(m_{t-1}|m_{t-2}) \quad (5.2)$$

It is customary to represent the stochastic properties of dynamics model by probability density. In case of occupancy grid maps with discrete variables, Bernoulli distributions is the common choice. But, in wind mapping process where the

subject data, wind speed and direction are continuous variable and basically represented by their first and second moment statistics, Bernoulli density could not be a valid choice. Then by, here the dynamics model is represented by the multi-dimensional Gaussian distribution

$$p(m_t|m_{t-1}) \sim \mathcal{N}(\mu_t, \Sigma_t) \quad (5.3)$$

where $\mu_t \in \mathcal{R}^n$ is the state vector storing the stochastic properties of every grid cell in the map, and $\Sigma_t \in \mathcal{R}^{n \times n}$ is the full covariance matrix.

2. *Observation Model*

Observation model describes the likelihood of an observations at given time step, formally it is the conditional probability of observation $z(t)$ at the map $m(t)$.

$$p(z_t|m_t) \quad (5.4)$$

Likewise to dynamics model, the inaccuracies in observations usually are represented probabilistically using parametric density functions. Here due to specific characteristic of sampling wind data within the MRWM framework, single measurement per each cell grid, the linear model with additive Gaussian noise would suffice for the observation model. It is assumed that at a given time spot t , the observable process z_t has a component of measurement Gaussian error given by the measurement equation

$$z_t = C_t m_t + Q_t \quad (5.5)$$

where $C_t \in \mathcal{R}^{m \times n}$ is the measurement matrix and $\sigma_n \in \mathcal{R}^n$ is the Gaussian measurement noise covariance matrix. Notice that, based on the assumption of single cell measurement at a time, the measurement matrix could be simplified to a $(1 \times n)$ vector whose all elements are zero except the cell corresponding the current cell under measurement.

$$C_t = [0 \quad \dots \quad 1 \quad \dots \quad 0] \quad (5.6)$$

$$C_t[k] = \begin{cases} 1, & k = c. \\ 0, & k \neq c. \end{cases}$$

which eventually leads to the observation model

$$z_t = C_t m_t + \sigma_n \quad (5.7)$$

where $C_t \in \mathcal{R}^{1 \times n}$ is the measurement model vector and $\sigma_n \in \mathcal{R}$ is the Gaussian measurement noise matrix with zero mean.

5.2.2 Kalman Filter

Provided a linear Gaussian stochastic space model while streaming of input observations, Kalman filter, also known as linear quadratic estimation (LQE), is an optimal choice to estimate map posterior beliefs recursively. In a sense, Kalman filter is optimal, since it optimizes the Kalman gain as a function of the relative confidence of the observation and the current state estimate.

The input to the Kalman filter algorithm is the map posterior belief at $t - 1$, modelled by the first and second moment statistics respectively μ_{t-1} and Σ_{t-1} . The output of the Kalman filter algorithm is the updated map posterior using the current observation z_t . From a theoretical standpoint, implementation of the Kalman filter for the real-time spatial mapping of wind data is summarized with the following steps:

1. *Prediction Step*

In the prediction step, the current estimates of the posterior map are obtained. Following the assumption of the static state space model for state transition, Equation 5.2, the current state space model is identical to the previous state.

$$\hat{\mu}_t = \mu_{t-1}, \quad \hat{\Sigma}_t = \Sigma_{t-1} \quad (5.8)$$

2. *Update Step*

The update step involves optimizing the current estimate of the map belief using the state observations. Basically, the optimization is performed by means of the relative confidence of the observation and the current state estimate, so-called Kalman gain. Particularly, at the time stamp t , the current map belief is updated by the Kalman gain and the difference between the current estimate $\hat{\mu}_t$ and the actual observation z_t .

$$\mu_t = \hat{\mu}_t + K_t(z_t - C_t \hat{\mu}_t) \quad (5.9)$$

where, Kalman gain K_t is estimated from the estimate covariance matrix $\hat{\Sigma}_t$ and the observation matrix C_t

$$K_t = \hat{\Sigma}_t C_t^T (C_t \hat{\Sigma}_t C_t^T + \sigma_n^2)^{-1} \quad (5.10)$$

Likewise the covariance matrix $\hat{\Sigma}_t$ is updated as

$$\Sigma_t = (I - K_t C_t) \hat{\Sigma}_t \quad (5.11)$$

5.2.3 Practical Considerations

The Kalman Filter is supposed to be the best optimal estimator for linear systems disturbed by additive white Gaussian noise, so-called linear quadratic systems, which as mentioned in Section 5.1.2 has its own strength in its simplicity and ability to estimate optimal state space model recursively. However, the basic linear Kalman filter, particularly within the framework of MRWM, may fail to be an optimal approach when utilized for building large-scale maps or maps with very fine grids.

In a sense, the ordinary linear Kalman filter is notably fragile in circumstances in which the state space model is very large. In fact this mainly occurs because the Kalman filter applied to large-scale state space model deals with high-dimensional matrix computations and as a result it is deemed to be computationally and even data intensive. That is, for a N -dimensional map space model, the computation complexity for mean and covariance matrix stands at $\mathcal{O}(N)$ and $\mathcal{O}(N^2)$, respectively.

Owing to address this complexity, several computationally efficient implementations of Kalman filter have been proposed for large-scale state space models, such as Ensemble Kalman filter (EKF) [30] and Kalman filtering using Broyden–Fletcher–Goldfarb–Shanno (LBFGS) [31]. However, due to linear observation model and the simplification imposed by wind data measurements one may simply use the compressed covariance matrix structure. In particular, by employing the idea that closer cells in neighbourhood bear higher correlation compared to the ones in far distance, the high dimensionality of covariance matrix could be truncated using a windowing strategy as proposed by [5].

Another complexity in ordinary Kalman filtering arises when, the assumptions of linear state transitions and linear measurements with additive unimodal Gaussian error are violated. In particular, as shown in Chapter 4, the wind direction dispersion may follow the multiplicity of modes and generally could not be described by single mode distribution models. This observation under the Gaussian linear assumption of map posterior beliefs, make the ordinary Kalman filter intractable. However, there is a range of nonlinear Kalman filter variations that could be utilized in this purpose, for example non-Gaussian Kalman filter (NGKF) [32]. A procedure that proposes MMSE estimator for linear, non-Gaussian problems, where the Gaussian mixture model is exploited to approximate non-Gaussian distributions.

5.3 Experimental Results

Due to the shortcoming of the existing wind data as well as the limitation imposed by the timing plan of the thesis project, the proposed wind mapping algorithm is implemented and evaluated by means of simulated data. The main aim of the simulation is to demonstrate the mapping process while collecting wind observations and gradually updating the wind map. In the mean time, the performance of the algorithm to predict unsampled data is evaluated incorporating different mapping parameters, initial map covariance and the number observations. It is worthwhile to note, the current mapping simulation is a generic experiment that could be extended to every wind subject attributes, wind speed mean, wind direction mean, wind speed variance and wind direction concentration parameter. Simulation of mapping process basically involves with two distinct subtasks: the robot wandering and the wind data simulation.

The suitable robot exploration strategy for SGDM, due to sophisticated flow dispersion models is non-trivial task and is not the concern of the current thesis work, e.g. [33]. Here a very basic wandering scheme is employed as the robot zizags the map with a particular step length, called observation step. That is, the robot wanders around the mapping environment till completely covers the area.

The method for data simulation is *Gaussian Random Field (GRF)*, where a 2-dimensional random field is generated following the Gaussian distribution. GRF is supposed to be the most commonly used model describing spatial stochastic processes. As GRF modelled by the assumption of Gaussianity for each position in the spatial domain, it is fully determined by the corresponding mean and covariance functions.

For a 2-dimensional $n_1 \times n_2$ grid map $m = \{m_1, m_2, \dots, m_{n_1 n_2}\}$, consisting of $n_1 n_2$ jointly random variables, with covariance function $f_c(\cdot)$ defining covariance matrix $S \in \mathcal{R}^{n_1 n_2 \times n_1 n_2}$, the Gaussian random field is generated as a sample realization of $n_1 n_2$ -dimensional Gaussian distribution $\mathcal{N}(0, S)$. Here a conditional simulation method based on lower-upper triangular decomposition (LU) from [34] is used to simulate the random field. Considering LU decomposition of covariance matrix $S = LU$ and given some sampled data z_d , the sample S is generated by

$$m = A_{21}L_{11}^{-1}z_d + L_{22}\epsilon \quad (5.12)$$

as partitioned lower-triangular matrix L is defined as

$$L = \begin{bmatrix} L_{11} & 0 \\ A_{21} & L_{22} \end{bmatrix} \quad (5.13)$$

Number of Observations	σ	Step [cm]	RMSE
237	2	100	0.00300
237	5	100	0.00058
237	10	100	0.00023
58	2	200	0.03190
58	5	200	0.00620
58	10	200	0.00330

TABLE 5.1: The mapping algorithm is applied to the Gaussian data. The performance of the mapping process, quantified by means of RMSE, is evaluated with respect to different observation steps and initial variance values.

Number of Observations	σ	Step [cm]	RMSE
237	2	100	0.0130
237	5	100	0.0029
237	10	100	0.0016
58	2	200	0.1238
58	5	200	0.0424
58	10	200	0.0268

TABLE 5.2: The mapping algorithm is applied to the turbulent data. The performance of the mapping process, quantified by means of RMSE, is evaluated with respect to different observation steps and initial variance values.

In order to reproduce the low and the high degrees of variability in the spatially distributed data, two types of 30-by-30 Gaussian random fields, corresponding to the Gaussian and the turbulent correlations, are generated.¹ The grided spatial data are assumed to be of the fixed resolution with the cell size 50×50 cm. The generated data is served as the true spatial observations to the mapping algorithm.

The following procedure is used for the initialization of the mapping algorithm:

- The initial estimate of the spatial map uniformly is set to a value slightly larger than the first collected observation.
- Inspired by [5], the initial covariance matrix is defined in form 2-dimensional isotropic Gaussian distribution. In this connection, the starting variance of each grid cell is the key parameter affecting the overall performance of the mapping process. The initial value of covariance matrix is closely connected to physical attributes of environment.

Figures 5.2 and 5.3 demonstrate the simulation of the mapping process of the Gaussian data with two different observations steps, particularly 100cm and 200cm. Obviously, the shorter observation step corresponds to the fewer number of observations. For each observation step, three different initial variance values, which are 2 (row a), 5 (row b) and 10 (row c), are experimented. The performance of the mapping algorithm with respect to different initial variance values and the observation step lengths is quantified in Table 5.1.

¹The Matlab code (Paul Constantine, 26 Apr 2012) was used to generate GRFs. <http://www.mathworks.com/matlabcentral/fileexchange/27613-random-field-simulation>.

Overall, as could be expected, mapping with the shorter observation steps and the larger variance values, as utilizing more information about the environment, leads to more accurate map estimates. On the other hand, too few samples along with small values for initial variance, would deteriorate the performance of mapping process.

The results of simulation of the mapping process for turbulent data are presented in Figures 5.4 and 5.5. The corresponding performance measures related to different observation steps and the initial variance values are listed in Table 5.2. Similar to the simulations in the Gaussian case, the more accurate mapping results do come from higher number observations and large variance values, and indeed no significant difference was observed.

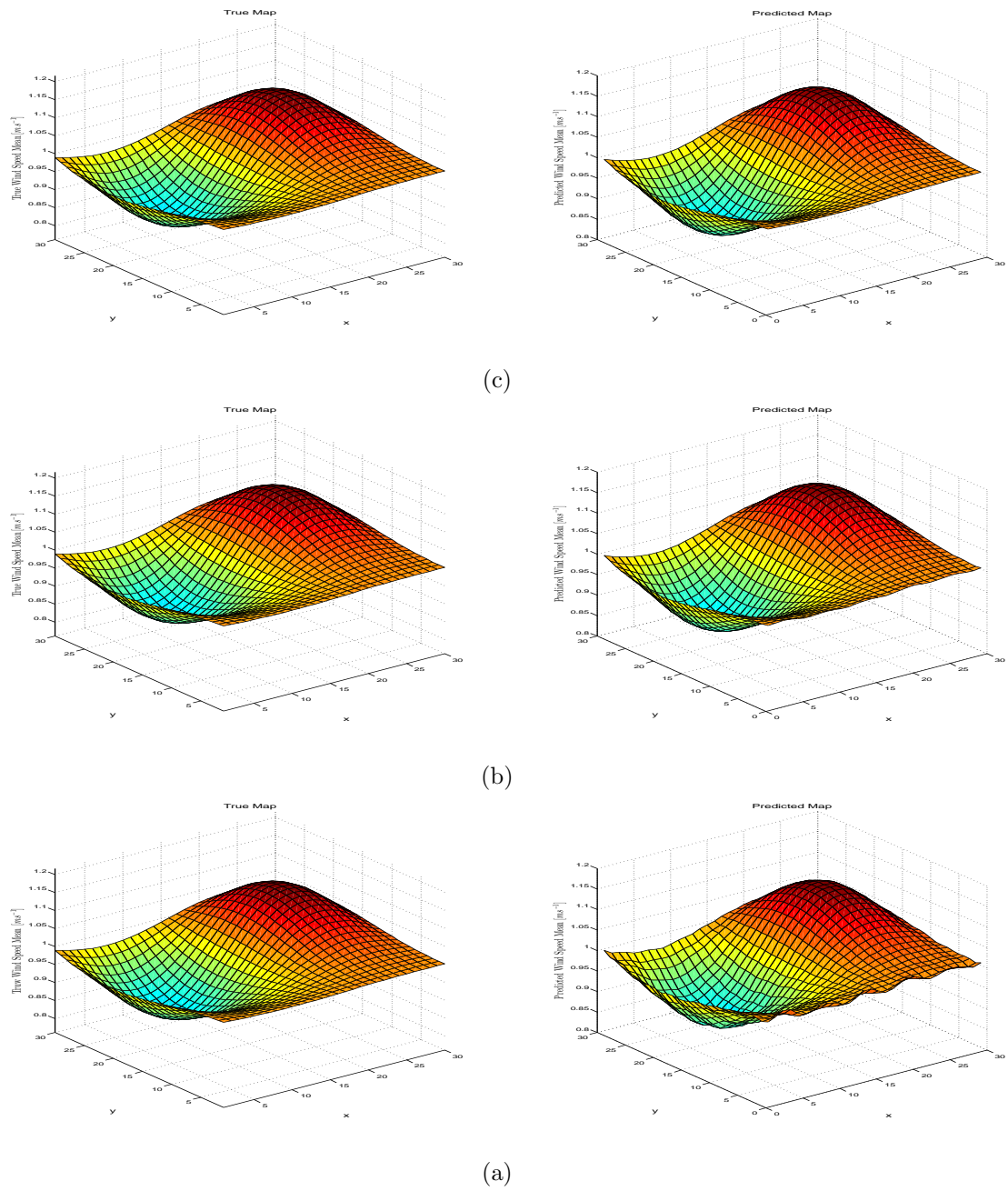


FIGURE 5.2: The plots show the mapping results with observation step 100cm as incorporating three different initial variance value (Gaussian data). (a): 2 (b): 5 (c):

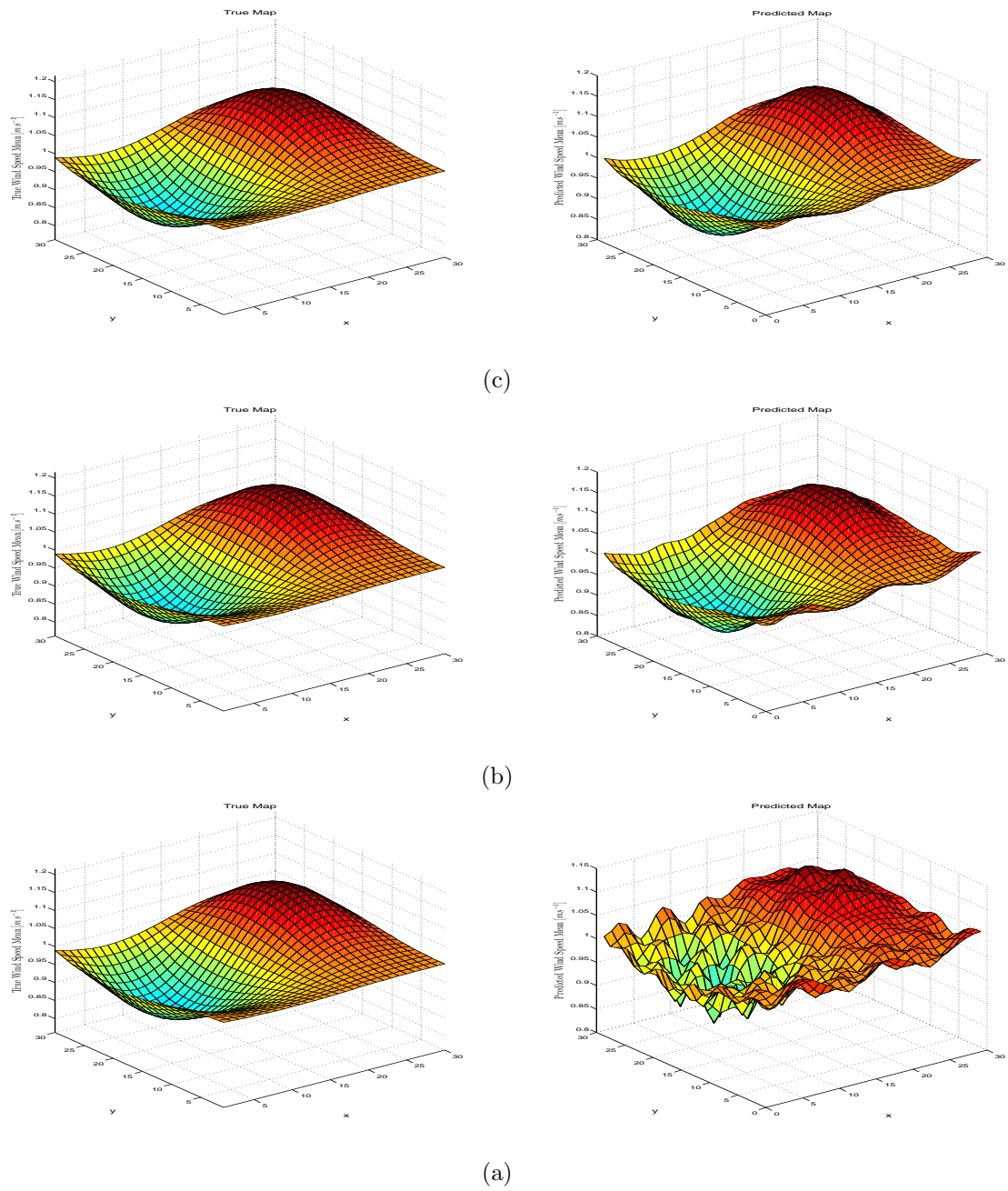


FIGURE 5.3: The plots show the mapping results with observation step 200cm as incorporating three different initial variance value (Gaussian data). (a): 2 (b): 5 (c):

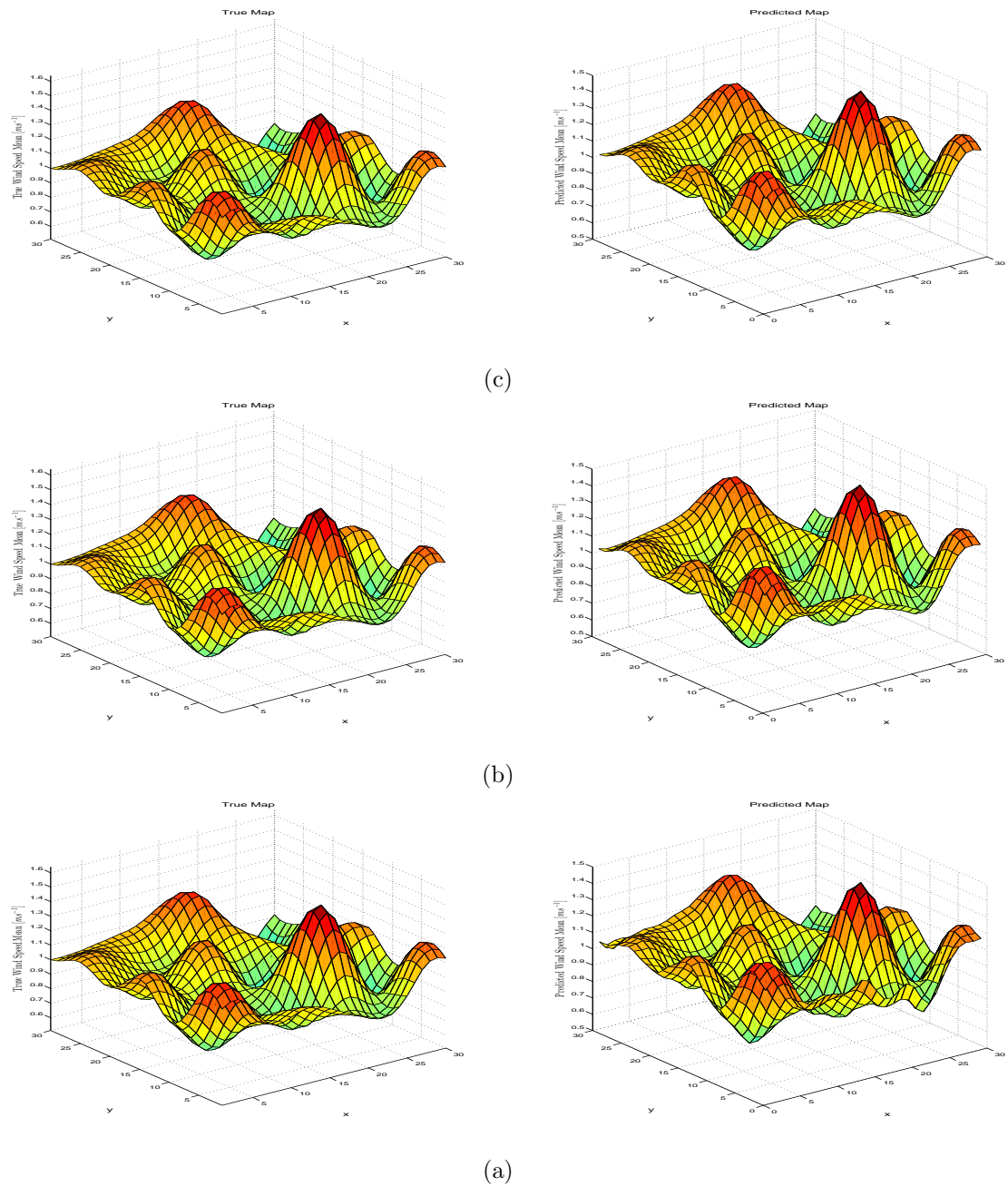


FIGURE 5.4: The plots show the mapping results with observation step 100cm as incorporating three different initial variance value (turbulent data). (a): 2 (b): 5 (c):

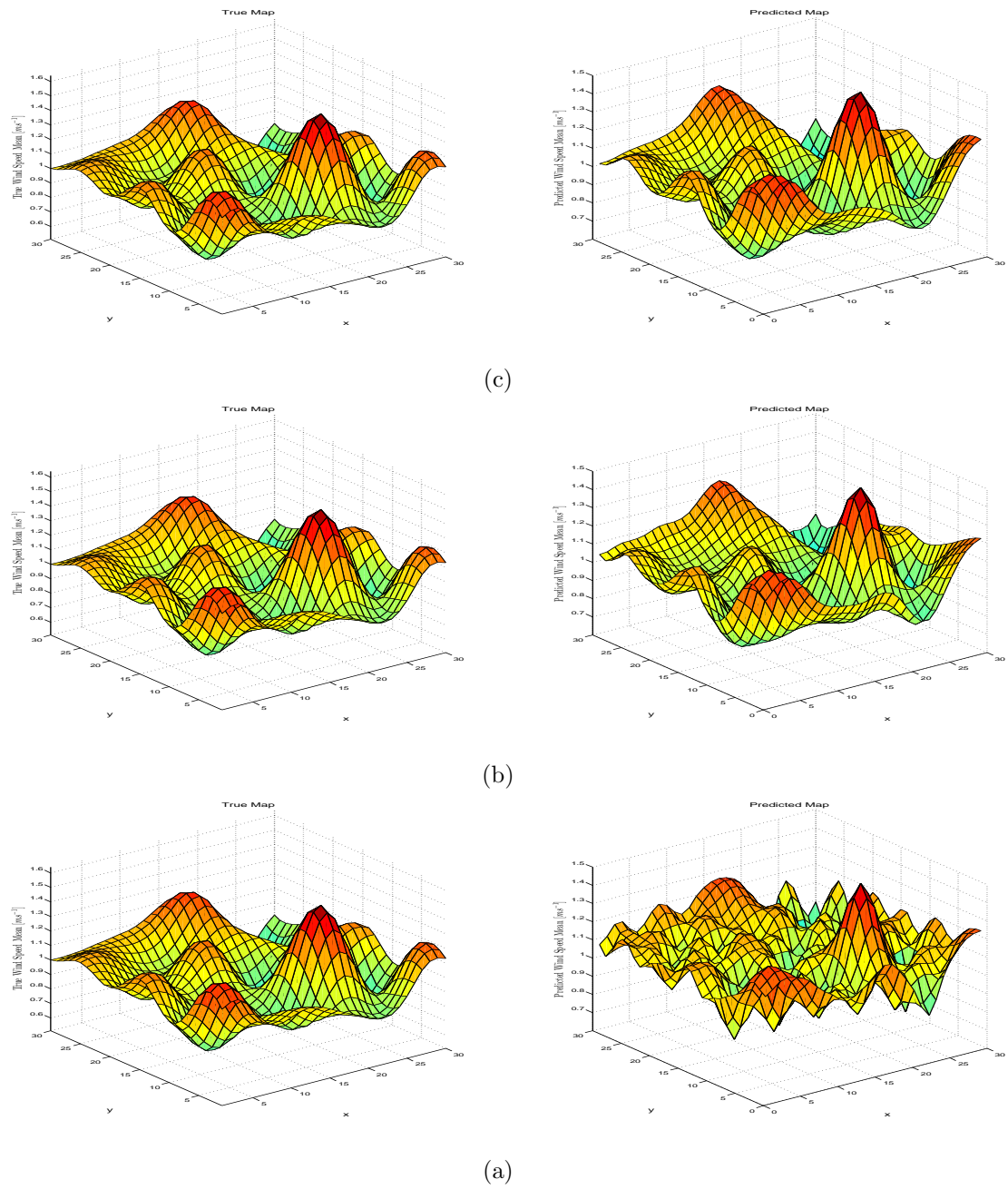


FIGURE 5.5: The plots show the mapping results with observation step 200cm as incorporating three different initial variance value (turbulent data). (a): 2 (b): 5 (c):

Chapter 6

Conclusions

The final chapter begins with an overview of the framework proposed for the MRWM task in Section 6.1 and then the future works addressing the improvements and the unsolved problems are reviewed in Section 6.2.

6.1 Summary

In this thesis, a framework for constructing wind maps utilizing in mobile robots was proposed. The overriding purpose of the framework was to address the spatio-temporal variabilities inherent in near-surface wind measurements, while complying with the specific requirements of a typical mobile robotic SGDM task. To accomplish this, three distinct subtasks were resolved, namely “statistical wind speed modelling”, “statistical wind direction modelling” and “wind mapping”.

1. *Statistical wind speed modelling*

The temporal uncertainties observed in short-scale wind speed measurements were modelled using a single peak Gaussian distribution. Under the assumption of the ideal turbulence, TDMM method was utilized to extract the shortest valid averaging interval incorporating into the wind speed distribution modelling. For evaluation purpose, ORU.WIND outdoor and indoor data sets were used and the estimated Gaussian distribution models were compared to empirical densities using the goodness-of-fit measures. Although, the TDMM method proved to be acceptable in the majority of cases, it failed in the presence of abrupt wind speed changes (wind gusts) and at very low wind speed patterns, commonly observed in indoor environments.

2. *Statistical wind direction modelling*

For construction of the statistical wind direction model, due to the observed multimodality in wind directional data, a mixture-model of von-mises distribution was used. The mixture model parameters were estimated under the formulation of a non-linear least-square problem where the error between empirical and theoretical cumulative distribution was minimized. The mixture model was applied to “ORU.WIND” outdoor data set and the performance of the system was evaluated with respect to the parameters of the model.

3. *Wind Mapping*

Inspired by the conventional occupancy grid maps commonly used in the area of mobile robotics, the wind spatial mapping was addressed by considering n -dimensional linear state space model. The mapping algorithm basically utilized the linear Kalman filter to recursively update the spatial map of data with streaming of input observations. For the validation purpose, the proposed mapping model was applied to artificial data simulated using Gaussian random field.

6.2 Future Work

The future works to improve the proposed framework of MRWM could be summarized to the followings:

- Due to the shortcoming of the available wind data, the mapping algorithm only was applied to simulated data. The primary future work will be to explore the effectiveness of the proposed framework at mapping of real world scenarios.
- The proposed spatial mapping framework is based the ordinary linear Kalman filtering technique, presented in Chapter 5, does not take into account the possibility of multiplicity of modes, as particularly observed in wind direction measurements. Clearly, this kind of formulation could not be a comprehensive solution for a real world wind mapping task, and as an extension to the present model, a non-linear version of Kalman filtering techniques, such as [32], may be utilized.
- The short-scale wind speed fluctuations were addressed using a pure single peak Gaussian distribution model. However, it was observed that in presence of wind gustiness, empirical wind speed density might deviate from the bell shape distribution model and indeed the Gaussian distribution could not be a rational choice within these particular cases. Alternatively, one may adopt a collection of distribution classes applicable to short-scale wind speed fluctuations, as proposed by [35], where the best fitted model adaptively chosen by a trained classifier model.

-
- In the current work, the wind speed and the wind direction attributes were represented by means of the two distinct distribution models. However, it would be valuable, to have both wind speed and wind direction in a single statistical model. In order to follow, the representation of wind data could be extended to the form of a joint probability statistical model incorporating both wind speed and wind attributes as the marginal distribution functions, e.g. [36].

Appendix A

ORU.WIND Data Specification

Project	Mobile robot wind mapping
Instrument	YOUNG 8100 3D ultrasonic anemometer, Range: 0-40 m/s, Resolution: 0.01 m/s, Accuracy: $\pm 1\%$ rms + ± 0.05 m/s (0-30 m/s), $\pm 3\%$ rms (30-40 m/s) Range: 0-40 m/s Resolution: 0.01 m/s Accuracy: $\pm 1\%$ rms + ± 0.05 m/s (0-30 m/s), $\pm 3\%$ rms (30-40 m/s)
Format	ASCII
Sampling rate	20 Hz
Measurement elements	U, V, W, 2D Wind Speed Magnitude, 3D Wind Speed Magnitude, Azimuth, Elevation, Speed of Sound, Sonic Temperature, Error Code
Units	Wind Speed: m/s, Direction: degrees, Temperature: $^{\circ}\text{C}$

TABLE A.1: ORU.WIND Data Set Specification

Appendix B

Wind Speed Modelling Results

SegId	$N_{\Delta T}$	ΔT	Mean [ms^{-1}]	Std [ms^{-1}]	Kurtosis	Skewness	NRMSE	R^2
1	1260	01:03.000	1.337	0.189	2.333	0.138	0.095	0.926
2	340.000	00:17.000	0.264	0.191	2.440	-0.532	0.171	0.714
3	300.000	00:15.000	0.173	0.077	2.568	0.351	0.096	0.922
4	320.000	00:16.000	0.744	0.208	2.234	-0.005	0.096	0.913
5	300.000	00:15.000	0.865	0.171	2.943	0.066	0.096	0.893
6	640.000	00:32.000	1.561	0.238	2.382	-0.293	0.093	0.938
7	300.000	00:15.000	0.385	0.115	2.763	0.350	0.080	0.955
8	880.000	00:44.000	1.049	0.411	2.097	0.577	0.240	0.381
9	780.000	00:39.000	1.086	0.166	2.817	0.344	0.058	0.971
10	300.000	00:15.000	1.491	0.201	3.159	-0.546	0.110	0.856
11	1160.000	00:58.000	0.318	0.247	2.308	0.341	0.125	0.840
12	300.000	00:15.000	0.455	0.134	3.299	0.935	0.159	0.717
13	680.000	00:34.000	0.039	0.075	3.517	0.018	0.042	0.985
14	360.000	00:18.000	0.461	0.192	4.122	0.967	0.112	0.891
15	300.000	00:15.000	0.893	0.313	3.861	0.849	0.089	0.937
16	801.000	00:30.000	1.131	0.220	3.062	0.243	0.072	0.947
17	1200.000	01:00.000	1.457	0.226	2.766	0.453	0.132	0.879
18	660.000	00:33.000	1.111	0.173	2.756	-0.271	0.087	0.931
19	1280.000	01:04.000	1.093	0.166	2.747	-0.381	0.078	0.951
20	2540.000	02:07.000	1.048	0.295	3.111	0.605	0.121	0.848
21	1820.000	01:31.000	1.186	0.168	2.929	0.218	0.078	0.960
22	740.000	00:37.000	0.256	0.224	2.613	-0.549	0.138	0.818
23	320.000	00:16.000	1.264	0.141	2.833	-0.288	0.050	0.982
24	1240.000	01:02.000	1.082	0.180	2.693	0.598	0.123	0.863
25	1920.000	01:36.000	0.283	0.276	3.264	0.457	0.073	0.960
26	700.000	00:35.000	0.241	0.108	2.625	0.167	0.066	0.967
27	320.000	00:16.000	1.988	0.114	2.551	-0.286	0.093	0.892
28	440.000	00:22.000	1.085	0.135	2.930	-0.400	0.063	0.966
29	540.000	00:27.000	1.488	0.271	4.347	1.238	0.145	0.811
30	300.000	00:15.000	1.062	0.183	2.527	0.533	0.125	0.867
31	460.000	00:23.000	0.371	0.211	1.976	0.264	0.193	0.619
32	1420.000	01:11.000	0.965	0.230	2.632	-0.350	0.093	0.922
33	1220.000	01:01.000	1.317	0.271	2.912	-0.139	0.084	0.930
34	480.000	00:24.000	0.521	0.276	2.525	0.222	0.119	0.870
35	440.000	00:22.000	1.631	0.211	4.238	-0.272	0.082	0.930
36	300.000	00:15.000	1.476	0.111	2.495	-0.412	0.099	0.910
37	300.000	00:15.000	0.200	0.151	2.839	0.608	0.121	0.845
38	1620.000	01:21.000	1.381	0.408	5.025	1.427	0.157	0.712
39	801.000	00:30.000	0.785	0.300	2.218	0.376	0.165	0.726
40	520.000	00:26.000	1.947	0.228	2.682	-0.624	0.129	0.881
41	620.000	00:31.000	1.674	0.483	2.646	0.086	0.075	0.953
42	620.000	00:31.000	1.318	0.326	4.222	1.114	0.139	0.806
43	940.000	00:47.000	2.482	0.667	2.558	0.115	0.096	0.918
44	620.000	00:31.000	0.924	0.316	3.649	1.004	0.149	0.754
45	420.000	00:21.000	2.150	0.223	2.153	-0.299	0.127	0.876
46	1180.000	00:59.000	0.701	0.421	6.502	0.998	0.083	0.933
47	1500.000	01:15.000	0.848	0.414	2.417	0.145	0.102	0.911
48	760.000	00:38.000	1.220	0.504	1.950	0.022	0.174	0.737
49	300.000	00:15.000	0.166	0.313	1.952	0.375	0.203	0.549
50	560.000	00:28.000	2.253	0.459	2.865	0.262	0.057	0.977
51	460.000	00:23.000	2.310	0.379	2.355	-0.073	0.087	0.939
52	580.000	00:29.000	2.637	0.390	2.193	0.339	0.147	0.796
53	340.000	00:17.000	0.591	0.358	2.267	-0.139	0.104	0.903
54	440.000	00:22.000	1.956	0.472	2.354	0.000	0.113	0.885
55	340.000	00:17.000	1.339	0.264	2.485	-0.220	0.083	0.946
56	540.000	00:27.000	1.680	0.542	1.982	-0.411	0.213	0.448
57	640.000	00:32.000	1.954	0.665	2.197	0.173	0.128	0.855
58	300.000	00:15.000	2.033	0.186	2.942	0.213	0.028	0.993
59	1200.000	01:00.000	1.866	0.424	2.676	-0.074	0.082	0.955
60	580.000	00:29.000	1.086	0.380	1.972	0.312	0.186	0.661
61	801.000	00:30.000	1.009	0.235	2.864	0.248	0.054	0.974

TABLE B.1: Statistical moments and errors for Probability Density Functions compared to the empirical distribution, for ORU.WIND.OS.100.

SegId	$N_{\Delta T}$	ΔT	Mean [ms^{-1}]	Std [ms^{-1}]	Kurtosis	Skewness	NRMSE	R^2
1	700	00:35.000	1.373	0.367	2.061	0.378	0.177	0.669
2	1260	01:03.000	1.373	0.354	2.321	-0.039	0.103	0.917
3	460	00:23.000	0.950	0.256	2.598	-0.057	0.064	0.964
4	420	00:21.000	2.908	0.340	1.997	0.110	0.142	0.852
5	600	00:30.000	1.963	0.301	2.559	0.163	0.061	0.967
6	360	00:18.000	3.445	0.346	3.198	-0.534	0.061	0.968
7	500	00:25.000	2.478	0.383	2.224	0.200	0.121	0.899
8	300	00:15.000	1.647	0.122	2.672	-0.048	0.065	0.966
9	460	00:23.000	2.025	0.181	2.959	-0.360	0.063	0.965
10	640	00:32.000	2.806	0.454	2.946	0.862	0.155	0.734
11	520	00:26.000	1.758	0.347	2.825	-0.311	0.069	0.957
12	1100	00:55.000	2.122	0.623	2.241	0.457	0.160	0.761
13	1360	01:08.000	1.721	0.418	3.387	0.491	0.082	0.942
14	520	00:26.000	1.378	0.315	3.028	0.757	0.137	0.812
15	1040	00:52.000	1.259	0.375	2.287	0.510	0.168	0.672
16	300	00:15.000	1.079	0.230	2.231	0.516	0.187	0.597
17	1640	01:22.000	1.101	0.476	2.369	0.349	0.123	0.841
18	960	00:48.000	3.232	0.321	3.481	-0.350	0.038	0.989
19	801	00:30.000	2.463	0.253	3.671	0.827	0.091	0.932
20	340	00:17.000	1.309	0.605	1.892	0.097	0.215	0.591
21	300	00:15.000	2.132	0.150	2.507	0.073	0.059	0.972
22	801	00:30.000	2.574	0.576	2.716	0.191	0.059	0.972
23	460	00:23.000	0.968	0.167	2.736	0.604	0.126	0.850
24	360	00:18.000	1.903	0.207	2.674	0.406	0.087	0.932
25	780	00:39.000	2.846	0.355	2.798	0.702	0.128	0.832
26	820	00:41.000	0.902	0.417	2.089	0.317	0.175	0.707
27	720	00:36.000	1.376	0.602	1.827	-0.135	0.206	0.548
28	1800	01:30.000	2.588	0.472	2.685	0.258	0.054	0.977
29	801	00:30.000	1.337	0.281	2.321	-0.086	0.078	0.940
30	760	00:38.000	1.073	0.200	2.588	-0.192	0.096	0.916
31	2540	02:07.000	-0.130	0.310	3.202	0.347	0.148	0.788
32	560	00:28.000	1.394	0.241	3.131	0.045	0.066	0.959
33	2160	01:48.000	3.014	0.354	3.212	-0.084	0.045	0.980
34	1380	01:09.000	2.869	0.504	2.433	-0.472	0.170	0.599
35	500	00:25.000	1.211	0.188	2.238	0.104	0.104	0.923
36	320	00:16.000	1.615	0.234	2.018	-0.026	0.151	0.857
37	1180	00:59.000	0.382	0.500	2.286	-0.254	0.125	0.865
38	300	00:15.000	2.349	0.275	3.241	0.236	0.060	0.966
39	320	00:16.000	0.924	0.124	2.453	-0.355	0.121	0.866
40	300	00:15.000	1.475	0.144	2.229	-0.337	0.136	0.825
41	920	00:46.000	0.297	0.103	3.327	-0.217	0.036	0.990
42	820	00:41.000	1.337	0.220	2.419	0.213	0.075	0.961
43	360	00:18.000	0.876	0.133	2.288	0.064	0.088	0.943
44	560	00:28.000	0.514	0.145	1.848	0.128	0.209	0.643
45	300	00:15.000	0.439	0.059	3.364	0.077	0.026	0.994
46	600	00:30.000	1.687	0.115	3.491	0.049	0.031	0.991
47	760	00:38.000	0.650	0.201	2.074	0.122	0.153	0.825
48	440	00:22.000	1.249	0.170	2.395	0.408	0.125	0.880
49	380	00:19.000	0.884	0.223	2.267	-0.497	0.174	0.718
50	420	00:21.000	0.567	0.207	2.857	-0.530	0.099	0.918
51	780	00:39.000	1.019	0.246	2.310	-0.518	0.168	0.709
52	460	00:23.000	1.492	0.371	2.170	0.475	0.176	0.680
53	380	00:19.000	0.810	0.138	2.982	-0.332	0.057	0.973
54	480	00:24.000	0.465	0.238	2.646	-0.042	0.105	0.872
55	801	00:30.000	0.809	0.193	2.056	0.346	0.182	0.727
56	520	00:26.000	1.529	0.236	2.596	0.228	0.069	0.956
57	400	00:20.000	0.983	0.306	2.777	0.416	0.118	0.832
58	460	00:23.000	2.663	0.254	1.990	-0.092	0.142	0.860

TABLE B.2: Statistical moments and errors for Probability Density Functions compared to the empirical distribution, for ORU.WIND.OS.150.

SegId	$N_{\Delta T}$	ΔT	Mean [ms^{-1}]	Std [ms^{-1}]	Kurtosis	Skewness	NRMSE	R^2
1	320	00:16.000	1.236	0.214	1.894	0.036	0.176	0.792
2	800	00:40.000	0.736	0.150	3.477	0.832	0.108	0.886
3	300	00:15.000	0.145	0.030	3.283	0.902	0.144	0.779
4	360	00:18.000	0.856	0.111	2.646	-0.206	0.060	0.971
5	700	00:35.000	1.522	0.254	3.076	0.929	0.163	0.620
6	540	00:27.000	1.375	0.166	2.217	-0.003	0.109	0.917
7	680	00:34.000	1.029	0.169	2.759	-0.243	0.056	0.979
8	600	00:30.000	0.443	0.113	3.015	-0.369	0.056	0.971
9	740	00:37.000	0.298	0.247	2.218	0.334	0.167	0.760
10	300	00:15.000	0.692	0.113	2.323	-0.217	0.093	0.935
11	360	00:18.000	1.474	0.222	2.662	-0.231	0.064	0.966
12	1300	01:05.000	1.023	0.261	2.485	0.595	0.155	0.746
13	440	00:22.000	1.540	0.595	1.854	0.318	0.230	0.432
14	440	00:22.000	1.806	0.210	2.433	0.179	0.070	0.960
15	640	00:32.000	0.264	0.199	4.386	-1.173	0.133	0.803
16	460	00:23.000	1.986	0.191	3.061	-0.287	0.029	0.993
17	340	00:17.000	0.746	0.108	3.907	-0.873	0.104	0.897
18	360	00:18.000	1.340	0.123	2.558	0.182	0.062	0.972
19	380	00:19.000	1.268	0.326	2.892	0.629	0.126	0.844
20	1320	01:06.000	1.474	0.449	2.301	0.393	0.140	0.847
21	300	00:15.000	1.528	0.367	2.343	0.339	0.154	0.830
22	1060	00:53.000	0.397	0.358	2.521	0.340	0.089	0.925
23	340	00:17.000	0.879	0.152	2.472	-0.316	0.081	0.947
24	1140	00:57.000	0.512	0.194	2.510	-0.006	0.108	0.915
25	420	00:21.000	0.667	0.223	2.030	0.099	0.154	0.724
26	540	00:27.000	1.712	0.485	2.391	0.509	0.142	0.812
27	300	00:15.000	1.085	0.106	2.556	0.200	0.067	0.969
28	340	00:17.000	0.747	0.198	3.232	-0.100	0.064	0.970
29	1060	00:53.000	0.410	0.142	2.986	0.301	0.045	0.983
30	820	00:41.000	0.791	0.357	2.307	0.242	0.101	0.901
31	1380	01:09.000	1.196	0.369	2.033	0.365	0.187	0.651
32	560	00:28.000	0.242	0.301	1.969	0.191	0.212	0.641
33	300	00:15.000	3.103	0.335	2.046	0.041	0.132	0.889
34	600	00:30.000	0.599	0.437	1.815	0.232	0.228	0.373
35	620	00:31.000	0.782	0.188	2.792	0.226	0.041	0.987
36	320	00:16.000	0.825	0.308	2.242	-0.119	0.159	0.708
37	580	00:29.000	1.100	0.137	2.830	0.328	0.051	0.978
38	1100	00:55.000	1.034	0.758	3.158	1.039	0.200	0.524
39	620	00:31.000	0.793	0.415	1.942	-0.127	0.182	0.648
40	660	00:33.000	1.330	0.226	2.846	-0.177	0.033	0.990
41	460	00:23.000	0.233	0.208	3.176	0.282	0.032	0.992
42	560	00:28.000	1.489	0.254	2.150	-0.230	0.134	0.862
43	320	00:16.000	1.479	0.214	2.441	-0.370	0.098	0.920
44	840	00:42.000	1.743	0.383	2.647	0.618	0.143	0.805
45	300	00:15.000	1.438	0.146	3.170	0.554	0.068	0.962
46	540	00:27.000	0.862	0.463	2.323	-0.380	0.154	0.715
47	440	00:22.000	1.492	0.577	2.222	0.109	0.135	0.811
48	360	00:18.000	1.511	0.182	3.273	0.627	0.099	0.914
49	800	00:40.000	1.672	0.309	2.653	0.053	0.059	0.970
50	940	00:47.000	1.499	0.245	2.447	0.246	0.090	0.945
51	440	00:22.000	0.450	0.224	2.844	0.628	0.127	0.842
52	380	00:19.000	1.177	0.374	2.086	0.014	0.171	0.749
53	300	00:15.000	0.780	0.120	2.174	-0.042	0.128	0.881
54	520	00:26.000	1.033	0.181	2.302	0.190	0.102	0.928
55	460	00:23.000	0.332	0.077	3.152	-0.218	0.047	0.980
56	460	00:23.000	1.286	0.274	1.762	0.153	0.209	0.624
57	440	00:22.000	0.671	0.175	2.401	-0.118	0.120	0.908
58	680	00:34.000	0.934	0.210	2.228	-0.002	0.122	0.884
59	400	00:20.000	0.758	0.179	2.035	-0.170	0.155	0.798
60	580	00:29.000	1.149	0.253	3.332	0.832	0.134	0.810

TABLE B.3: Statistical moments and errors for Probability Density Functions compared to the empirical distribution, for ORU.WIND.OC.100.

SegId	$N_{\Delta T}$	ΔT	Mean [ms^{-1}]	Std [ms^{-1}]	Kurtosis	Skewness	NRMSE	R^2
1	1300	01:05.000	1.016	0.171	2.861	-0.059	0.052	0.978
2	620	00:31.000	1.818	0.221	2.861	-0.010	0.047	0.982
3	320	00:16.000	2.313	0.283	2.725	0.341	0.085	0.933
4	960	00:48.000	1.874	0.374	2.309	-0.112	0.100	0.914
5	1060	00:53.000	2.016	0.217	2.646	0.387	0.066	0.965
6	801	00:30.000	1.124	0.100	2.642	-0.246	0.060	0.969
7	360	00:18.000	0.714	0.129	2.214	0.008	0.103	0.921
8	560	00:28.000	1.574	0.280	2.109	0.328	0.158	0.779
9	720	00:36.000	1.255	0.197	2.937	-0.018	0.075	0.932
10	720	00:36.000	1.135	0.144	2.518	0.245	0.066	0.968
11	1580	01:19.000	0.639	0.371	2.900	0.411	0.128	0.822
12	400	00:20.000	0.163	0.181	2.683	0.152	0.073	0.958
13	900	00:45.000	0.561	0.275	3.052	0.822	0.155	0.768
14	420	00:21.000	1.262	0.101	2.465	0.228	0.083	0.953
15	300	00:15.000	0.747	0.124	2.542	0.079	0.058	0.974
16	860	00:43.000	1.912	0.231	3.343	0.708	0.106	0.905
17	420	00:21.000	2.112	0.291	3.848	0.504	0.048	0.979
18	720	00:36.000	0.673	0.142	2.667	-0.076	0.076	0.956
19	540	00:27.000	0.217	0.119	2.299	-0.273	0.133	0.827
20	380	00:19.000	1.293	0.182	2.084	0.065	0.162	0.805
21	320	00:16.000	1.367	0.168	3.189	0.122	0.024	0.995
22	460	00:23.000	0.755	0.109	2.441	-0.054	0.061	0.970
23	340	00:17.000	1.146	0.183	2.578	-0.050	0.057	0.972
24	340	00:17.000	1.867	0.137	3.216	0.032	0.028	0.992
25	520	00:26.000	0.733	0.202	3.156	-0.134	0.062	0.965
26	801	00:30.000	2.331	0.342	2.071	0.055	0.141	0.861
27	1240	01:02.000	0.022	0.326	3.196	0.829	0.148	0.812
28	2400	02:00.000	1.584	0.294	2.794	-0.075	0.039	0.987
29	420	00:21.000	0.121	0.168	2.083	0.085	0.137	0.834
30	300	00:15.000	1.577	0.165	3.065	0.244	0.028	0.994
31	340	00:17.000	0.823	0.156	2.668	0.517	0.115	0.903
32	2220	01:51.000	1.469	0.427	2.835	0.837	0.206	0.691
33	1320	01:06.000	0.824	0.164	2.845	0.341	0.074	0.963
34	760	00:38.000	0.960	0.140	1.865	-0.021	0.179	0.776
35	1140	00:57.000	0.549	0.110	2.817	0.240	0.047	0.981
36	580	00:29.000	0.770	0.098	2.925	0.467	0.104	0.870
37	340	00:17.000	0.676	0.069	2.434	0.251	0.078	0.949
38	340	00:17.000	0.114	0.192	1.627	-0.245	0.280	0.000
39	1060	00:53.000	1.821	0.153	2.993	-0.514	0.069	0.960
40	820	00:41.000	1.527	0.371	1.695	-0.042	0.244	0.475
41	360	00:18.000	2.709	0.394	2.748	0.250	0.076	0.938
42	460	00:23.000	0.975	0.115	2.347	-0.302	0.110	0.911
43	300	00:15.000	1.054	0.072	3.387	0.163	0.032	0.990
44	1480	01:14.000	1.070	0.181	2.587	0.088	0.081	0.944
45	300	00:15.000	0.753	0.110	2.799	0.444	0.076	0.957
46	300	00:15.000	0.587	0.084	2.064	0.129	0.143	0.870
47	300	00:15.000	0.210	0.023	4.219	-0.187	0.137	0.692
48	780	00:39.000	0.356	0.085	2.866	-0.222	0.095	0.912
49	400	00:20.000	0.390	0.069	2.407	-0.276	0.092	0.929
50	300	00:15.000	0.462	0.068	2.319	-0.179	0.120	0.891
51	1080	00:54.000	0.128	0.064	3.167	0.203	0.100	0.917
52	300	00:15.000	0.107	0.050	2.768	-0.365	0.081	0.932
53	320	00:16.000	0.949	0.100	2.319	-0.046	0.104	0.925
54	300	00:15.000	0.599	0.070	3.100	0.596	0.094	0.920
55	300	00:15.000	0.650	0.061	2.832	-0.333	0.083	0.933
56	1300	01:05.000	0.911	0.093	3.456	0.461	0.058	0.971
57	340	00:17.000	0.513	0.040	3.977	-0.028	0.061	0.969
58	660	00:33.000	0.677	0.119	2.358	0.092	0.079	0.947

TABLE B.4: Statistical moments and errors for Probability Density Functions compared to the empirical distribution, for ORU.WIND.OC.150.

Appendix C

Wind Direction Modelling Results

SegId	$N_{\Delta T}$	ΔT	N (# Mixtures)	$NRMSE$ (PDF)	R^2 (PDF)	R^2 (CDF)	$K - S$ (CDF)
1	1260	01:03.000	1	0.058	0.947	0.999	0.018
2	340	00:17.000	1	0.068	0.903	0.990	0.050
3	300	00:15.000	5	0.087	0.858	0.990	0.051
4	320	00:16.000	1	0.085	0.847	0.998	0.074
5	300	00:15.000	1	0.021	0.986	1.000	0.018
6	640	00:32.000	1	0.090	0.848	0.992	0.027
7	300	00:15.000	1	0.064	0.959	0.784	0.115
8	880	00:44.000	4	0.158	0.698	0.897	0.104
9	780	00:39.000	1	0.133	0.734	0.832	0.178
10	300	00:15.000	1	0.065	0.917	0.994	0.078
11	1160	00:58.000	1	0.071	0.924	0.997	0.033
12	300	00:15.000	1	0.073	0.927	0.984	0.117
13	680	00:34.000	6	0.135	0.716	0.999	0.027
14	360	00:18.000	1	0.061	0.954	1.000	0.017
15	300	00:15.000	2	0.076	0.920	0.922	0.072
16	801	00:30.000	1	0.043	0.966	0.999	0.011
17	1200	01:00.000	2	0.151	0.720	0.977	0.058
18	660	00:33.000	1	0.140	0.656	0.946	0.155
19	1280	01:04.000	1	0.072	0.932	0.998	0.056
20	2540	02:07.000	1	0.112	0.738	0.944	0.136
21	1820	01:31.000	1	0.075	0.837	0.980	0.105
22	740	00:37.000	4	0.081	0.868	0.998	0.015
23	320	00:16.000	1	0.030	0.971	1.000	0.003
24	1240	01:02.000	1	0.052	0.951	0.997	0.060
25	1920	01:36.000	2	0.054	0.920	0.999	0.050
26	700	00:35.000	1	0.080	0.879	0.999	0.044
27	320	00:16.000	1	0.125	0.581	1.000	0.001
28	440	00:22.000	1	0.021	0.987	1.000	0.019
29	540	00:27.000	1	0.081	0.840	0.999	0.070
30	300	00:15.000	1	0.046	0.962	0.992	0.042
31	460	00:23.000	1	0.146	0.797	0.937	0.068
32	1420	01:11.000	6	0.065	0.945	1.000	0.009
33	1220	01:01.000	1	0.059	0.960	0.999	0.034
34	480	00:24.000	1	0.102	0.871	0.999	0.031
35	440	00:22.000	1	0.059	0.954	1.000	0.029
36	300	00:15.000	1	0.043	0.963	0.999	0.052
37	300	00:15.000	3	0.049	0.952	0.999	0.031
38	1620	01:21.000	1	0.064	0.941	1.000	0.040
39	801	00:30.000	1	0.073	0.891	0.999	0.053
40	520	00:26.000	1	0.054	0.934	0.999	0.063
41	620	00:31.000	1	0.048	0.963	1.000	0.029
42	620	00:31.000	1	0.069	0.911	0.999	0.039
43	940	00:47.000	3	0.027	0.988	1.000	0.024
44	620	00:31.000	1	0.053	0.946	0.999	0.040
45	420	00:21.000	1	0.049	0.938	0.999	0.038
46	1180	00:59.000	3	0.032	0.982	1.000	0.008
47	1500	01:15.000	1	0.069	0.925	0.999	0.023
48	760	00:38.000	3	0.035	0.973	1.000	0.014
49	300	00:15.000	5	0.108	0.808	0.999	0.052
50	560	00:28.000	1	0.069	0.917	0.999	0.064
51	460	00:23.000	1	0.051	0.944	1.000	0.036
52	580	00:29.000	1	0.035	0.962	0.999	0.040
53	340	00:17.000	1	0.095	0.863	0.999	0.042
54	440	00:22.000	1	0.026	0.987	1.000	0.032
55	340	00:17.000	2	0.065	0.935	1.000	0.020
56	540	00:27.000	1	0.011	0.997	1.000	0.015
57	640	00:32.000	1	0.076	0.873	0.999	0.048
58	300	00:15.000	1	0.077	0.884	0.997	0.043
59	1200	01:00.000	1	0.070	0.938	0.999	0.032
60	580	00:29.000	2	0.069	0.937	1.000	0.023
61	801	00:30.000	1	0.048	0.954	0.999	0.043

TABLE C.1: Performance statistics for the estimated distribution models compared to the empirical distribution, for wind direction data ORU.WIND.OS.100. (N: number of mixtures)

SegId	$N_{\Delta T}$	ΔT	N (# Mixtures)	$NRMSE$ (PDF)	R^2 (PDF)	R^2 (CDF)	$K - S$ (CDF)
1	700	00:35.000	1	0.065	0.935	0.999	0.023
2	1260	01:03.000	1	0.029	0.986	1.000	0.013
3	460	00:23.000	1	0.029	0.987	1.000	0.024
4	420	00:21.000	1	0.042	0.957	0.999	0.052
5	600	00:30.000	1	0.061	0.925	0.999	0.055
6	360	00:18.000	2	0.038	0.965	1.000	0.015
7	500	00:25.000	1	0.107	0.752	0.998	0.082
8	300	00:15.000	1	0.012	0.997	1.000	0.006
9	460	00:23.000	1	0.023	0.986	1.000	0.023
10	640	00:32.000	1	0.067	0.927	0.997	0.072
11	520	00:26.000	1	0.081	0.882	0.998	0.063
12	1100	00:55.000	2	0.062	0.932	1.000	0.016
13	1360	01:08.000	1	0.028	0.989	1.000	0.029
14	520	00:26.000	1	0.020	0.993	1.000	0.017
15	1040	00:52.000	1	0.042	0.969	1.000	0.039
16	300	00:15.000	1	0.081	0.775	0.996	0.119
17	1640	01:22.000	1	0.105	0.853	0.998	0.067
18	960	00:48.000	1	0.017	0.994	1.000	0.007
19	801	00:30.000	1	0.108	0.619	0.995	0.109
20	340	00:17.000	1	0.033	0.982	1.000	0.014
21	300	00:15.000	1	0.066	0.881	1.000	0.002
22	801	00:30.000	1	0.037	0.980	1.000	0.014
23	460	00:23.000	2	0.119	0.643	0.996	0.059
24	360	00:18.000	1	0.021	0.988	1.000	0.027
25	780	00:39.000	1	0.085	0.880	0.999	0.059
26	820	00:41.000	1	0.067	0.927	1.000	0.030
27	720	00:36.000	1	0.085	0.894	0.999	0.062
28	1800	01:30.000	1	0.034	0.969	1.000	0.026
29	801	00:30.000	1	0.073	0.913	0.999	0.047
30	760	00:38.000	6	0.060	0.917	0.999	0.012
31	2540	02:07.000	1	0.086	0.892	0.999	0.051
32	560	00:28.000	1	0.097	0.811	0.999	0.052
33	2160	01:48.000	1	0.070	0.933	0.999	0.029
34	1380	01:09.000	1	0.072	0.928	0.999	0.041
35	500	00:25.000	3	0.119	0.761	0.994	0.076
36	320	00:16.000	2	0.139	0.733	0.995	0.096
37	1180	00:59.000	1	0.071	0.905	0.999	0.034
38	300	00:15.000	1	0.071	0.913	0.999	0.023
39	320	00:16.000	1	0.013	0.995	1.000	0.002
40	300	00:15.000	1	0.026	0.986	1.000	0.003
41	920	00:46.000	2	0.076	0.905	0.999	0.047
42	820	00:41.000	1	0.062	0.953	0.999	0.034
43	360	00:18.000	2	0.126	0.696	0.996	0.069
44	560	00:28.000	1	0.084	0.870	0.999	0.046
45	300	00:15.000	1	0.070	0.934	0.000	0.110
46	600	00:30.000	1	0.023	0.983	1.000	0.041
47	760	00:38.000	1	0.064	0.899	0.999	0.040
48	440	00:22.000	1	0.036	0.974	0.999	0.014
49	380	00:19.000	1	0.040	0.963	1.000	0.026
50	420	00:21.000	3	0.120	0.809	0.999	0.042
51	780	00:39.000	4	0.096	0.847	0.964	0.042
52	460	00:23.000	1	0.017	0.993	1.000	0.007
53	380	00:19.000	1	0.027	0.984	1.000	0.023
54	480	00:24.000	3	0.060	0.893	0.998	0.080
55	801	00:30.000	1	0.072	0.917	0.999	0.050
56	520	00:26.000	1	0.077	0.878	0.999	0.042
57	400	00:20.000	1	0.067	0.905	0.999	0.061
58	460	00:23.000	1	0.051	0.928	0.999	0.059

TABLE C.2: Performance statistics for the estimated distribution models compared to the empirical distribution, for wind direction data ORU.WIND.OS.150. (N: number of mixtures)

SegId	$N_{\Delta T}$	ΔT	N (# Mixtures)	$NRMSE$ (PDF)	R^2 (PDF)	R^2 (CDF)	$K - S$ (CDF)
1	320	00:16.000	1	0.070	0.900	0.999	0.057
2	800	00:40.000	1	0.040	0.967	1.000	0.016
3	300	00:15.000	3	0.246	0.000	0.864	0.055
4	360	00:18.000	1	0.011	0.997	1.000	0.007
5	700	00:35.000	1	0.026	0.987	1.000	0.025
6	540	00:27.000	2	0.051	0.943	1.000	0.023
7	680	00:34.000	1	0.025	0.989	1.000	0.014
8	600	00:30.000	1	0.054	0.954	0.999	0.051
9	740	00:37.000	2	0.064	0.938	1.000	0.020
10	300	00:15.000	1	0.066	0.924	0.460	0.118
11	360	00:18.000	1	0.064	0.919	0.999	0.045
12	1300	01:05.000	1	0.090	0.853	0.998	0.050
13	440	00:22.000	2	0.086	0.871	0.999	0.049
14	440	00:22.000	1	0.025	0.987	1.000	0.017
15	640	00:32.000	2	0.092	0.859	1.000	0.010
16	460	00:23.000	1	0.056	0.929	0.999	0.036
17	340	00:17.000	1	0.027	0.984	1.000	0.016
18	360	00:18.000	1	0.089	0.813	1.000	0.034
19	380	00:19.000	2	0.115	0.749	0.997	0.084
20	1320	01:06.000	1	0.035	0.979	1.000	0.036
21	300	00:15.000	2	0.121	0.713	0.995	0.079
22	1060	00:53.000	5	0.050	0.961	0.999	0.026
23	340	00:17.000	1	0.028	0.981	1.000	0.040
24	1140	00:57.000	1	0.040	0.969	0.999	0.021
25	420	00:21.000	1	0.026	0.986	1.000	0.023
26	540	00:27.000	2	0.152	0.696	0.989	0.030
27	300	00:15.000	1	0.027	0.981	1.000	0.008
28	340	00:17.000	1	0.089	0.911	0.999	0.030
29	1060	00:53.000	1	0.028	0.984	0.999	0.041
30	820	00:41.000	1	0.109	0.782	0.998	0.051
31	1380	01:09.000	2	0.023	0.992	1.000	0.013
32	560	00:28.000	3	0.067	0.904	0.999	0.034
33	300	00:15.000	1	0.021	0.987	1.000	0.051
34	600	00:30.000	5	0.029	0.977	1.000	0.009
35	620	00:31.000	2	0.128	0.679	0.991	0.098
36	320	00:16.000	1	0.056	0.919	1.000	0.024
37	580	00:29.000	1	0.015	0.996	1.000	0.019
38	1100	00:55.000	4	0.033	0.965	0.999	0.028
39	620	00:31.000	1	0.037	0.973	1.000	0.036
40	660	00:33.000	2	0.042	0.958	1.000	0.016
41	460	00:23.000	4	0.048	0.952	0.999	0.029
42	560	00:28.000	2	0.046	0.969	1.000	0.011
43	320	00:16.000	1	0.008	0.998	1.000	0.014
44	840	00:42.000	1	0.037	0.985	1.000	0.016
45	300	00:15.000	1	0.009	0.997	1.000	0.020
46	540	00:27.000	4	0.066	0.921	0.999	0.074
47	440	00:22.000	1	0.086	0.818	0.997	0.087
48	360	00:18.000	1	0.015	0.994	1.000	0.024
49	800	00:40.000	1	0.019	0.992	1.000	0.014
50	940	00:47.000	1	0.074	0.897	0.999	0.034
51	440	00:22.000	2	0.034	0.980	1.000	0.011
52	380	00:19.000	1	0.017	0.993	1.000	0.013
53	300	00:15.000	1	0.094	0.739	0.995	0.085
54	520	00:26.000	1	0.019	0.990	1.000	0.017
55	460	00:23.000	1	0.043	0.966	1.000	0.026
56	460	00:23.000	1	0.100	0.713	0.997	0.077
57	440	00:22.000	1	0.074	0.941	0.999	0.038
58	680	00:34.000	1	0.020	0.992	1.000	0.034
59	400	00:20.000	1	0.031	0.976	1.000	0.008
60	580	00:29.000	1	0.058	0.956	1.000	0.030

TABLE C.3: Performance statistics for the estimated distribution models compared to the empirical distribution, for wind direction data ORU.WIND.OC.100. (N: number of mixtures)

SegId	$N_{\Delta T}$	ΔT	N (# Mixtures)	$NRMSE$ (PDF)	R^2 (PDF)	R^2 (CDF)	$K - S$ (CDF)
1	1300	01:05.000	1	0.010	0.998	1.000	0.006
2	620	00:31.000	1	0.073	0.889	0.998	0.057
3	320	00:16.000	1	0.026	0.987	1.000	0.010
4	960	00:48.000	1	0.030	0.981	1.000	0.024
5	1060	00:53.000	1	0.017	0.993	1.000	0.019
6	801	00:30.000	1	0.055	0.931	1.000	0.015
7	360	00:18.000	1	0.012	0.996	1.000	0.009
8	560	00:28.000	1	0.049	0.965	0.999	0.038
9	720	00:36.000	1	0.012	0.997	1.000	0.006
10	720	00:36.000	1	0.039	0.975	1.000	0.015
11	1580	01:19.000	3	0.078	0.927	1.000	0.021
12	400	00:20.000	2	0.091	0.876	1.000	0.023
13	900	00:45.000	1	0.079	0.815	0.998	0.051
14	420	00:21.000	1	0.043	0.936	1.000	0.000
15	300	00:15.000	1	0.055	0.949	1.000	0.021
16	860	00:43.000	1	0.016	0.995	1.000	0.004
17	420	00:21.000	1	0.033	0.975	1.000	0.008
18	720	00:36.000	1	0.067	0.864	0.997	0.063
19	540	00:27.000	3	0.079	0.873	0.999	0.038
20	380	00:19.000	1	0.101	0.866	0.998	0.072
21	320	00:16.000	1	0.055	0.907	0.998	0.063
22	460	00:23.000	1	0.056	0.959	1.000	0.019
23	340	00:17.000	1	0.024	0.989	1.000	0.029
24	340	00:17.000	1	0.045	0.964	1.000	0.032
25	520	00:26.000	1	0.025	0.983	1.000	0.032
26	801	00:30.000	1	0.080	0.871	0.998	0.045
27	1240	01:02.000	2	0.063	0.936	0.999	0.022
28	2400	02:00.000	1	0.041	0.958	0.998	0.069
29	420	00:21.000	2	0.101	0.905	1.000	0.018
30	300	00:15.000	1	0.012	0.996	1.000	0.031
31	340	00:17.000	1	0.066	0.904	0.999	0.035
32	2220	01:51.000	5	0.097	0.810	0.998	0.050
33	1320	01:06.000	1	0.028	0.988	1.000	0.013
34	760	00:38.000	2	0.067	0.901	0.999	0.039
35	1140	00:57.000	2	0.039	0.975	1.000	0.005
36	580	00:29.000	1	0.089	0.873	0.998	0.085
37	340	00:17.000	1	0.030	0.975	1.000	0.010
38	340	00:17.000	5	0.071	0.881	0.997	0.046
39	1060	00:53.000	1	0.019	0.991	1.000	0.003
40	820	00:41.000	2	0.127	0.571	0.981	0.197
41	360	00:18.000	1	0.026	0.983	1.000	0.018
42	460	00:23.000	1	0.035	0.973	1.000	0.023
43	300	00:15.000	1	0.042	0.976	1.000	0.021
44	1480	01:14.000	1	0.088	0.910	0.999	0.049
45	300	00:15.000	1	0.026	0.984	1.000	0.017
46	300	00:15.000	1	0.021	0.990	1.000	0.022
47	300	00:15.000	1	0.031	0.981	1.000	0.016
48	780	00:39.000	1	0.032	0.964	1.000	0.038
49	400	00:20.000	1	0.067	0.888	0.999	0.067
50	300	00:15.000	1	0.037	0.972	1.000	0.018
51	1080	00:54.000	1	0.056	0.917	0.999	0.053
52	300	00:15.000	2	0.145	0.604	0.992	0.083
53	320	00:16.000	1	0.028	0.985	0.999	0.031
54	300	00:15.000	1	0.043	0.958	1.000	0.015
55	300	00:15.000	1	0.030	0.970	1.000	0.000
56	1300	01:05.000	1	0.074	0.879	0.999	0.050
57	340	00:17.000	1	0.041	0.951	1.000	0.002
58	660	00:33.000	1	0.028	0.978	1.000	0.029

TABLE C.4: Performance statistics for the estimated distribution models compared to the empirical distribution, for wind direction data ORU.WIND.OC.150. (N: number of mixtures)

Bibliography

- [1] Achim J. Lilienthal and Tom Duckett. Building gas concentration gridmaps with a mobile robot. *Robotics and Autonomous Systems*, 48(1):3–16, 2004.
- [2] Amy Loutfi, Silvia Coradeschi, Achim J. Lilienthal, and Javier Gonzalez. Gas distribution mapping of multiple odour sources using a mobile robot. *Robotica*, 27(2):311–319, June 4 2009. URL <http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=3754296>.
- [3] A.J. Lilienthal, M. Reggente, M. Trincavelli, J.-L. Blanco, and J. Gonzalez. A statistical approach to gas distribution modelling with mobile robots - the kernel dm+v algorithm. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 570–576, Oct 2009. doi: 10.1109/IROS.2009.5354304.
- [4] Cyrill Stachniss, Christian Plagemann, and AchimJ. Lilienthal. Learning gas distribution models using sparse gaussian process mixtures. *Autonomous Robots*, 26(2-3):187–202, 2009. ISSN 0929-5593. doi: 10.1007/s10514-009-9111-5. URL <http://dx.doi.org/10.1007/s10514-009-9111-5>.
- [5] Jose Luis Blanco, Javier G. Monroy, Achim Lilienthal, and Javier Gonzalez-Jimenez. A kalman filter based approach to probabilistic gas distribution mapping. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, pages 217–222, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1656-9. doi: 10.1145/2480362.2480409. URL <http://doi.acm.org/10.1145/2480362.2480409>.
- [6] M. Reggente and A.J. Lilienthal. Using local wind information for gas distribution mapping in outdoor environments with a mobile robot. In *Sensors, 2009 IEEE*, pages 1715–1720, Oct 2009. doi: 10.1109/ICSENS.2009.5398498.
- [7] P.P. Neumann, S. Asadi, A.J. Lilienthal, M. Bartholmai, and J.H. Schiller. Autonomous gas-sensitive microdrone: Wind vector estimation and gas distribution mapping. *Robotics Automation Magazine, IEEE*, 19(1):50–61, March 2012. ISSN 1070-9932. doi: 10.1109/MRA.2012.2184671.

- [8] Victor Hernandez Bennetts, Achim Josef Lilienthal, Patrick Neumann, and Marco Trincavelli. Mobile robots for localizing gas emission sources on landfill sites: Is bio-inspiration the way to go? *Frontiers in Neuroengineering*, 4(20), 2012. ISSN 1662-6443. doi: 10.3389/fneng.2011.00020. URL <http://www.frontiersin.org/neuroengineering/10.3389/fneng.2011.00020/abstract>.
- [9] Tony Burton, Nick Jenkins, David Sharpe, and Ervin Bossanyi. *Wind energy handbook; 2nd ed.* Wiley, Hoboken, NJ, 2011.
- [10] American Meteorological Society. Glossary of meteorology, 2013. URL http://glossary.ametsoc.org/wiki/Main_Page.
- [11] R.B. Stull. *An Introduction to Boundary Layer Meteorology*. Atmospheric sciences library. Kluwer Academic Publishers, 1988. URL <http://books.google.se/books?id=794jYAAACAAJ>.
- [12] Isaac Van der Hoven. Horizontal wind speed in the frequency range from 0.0007 to 900 cycles per hour. volume 14, pages 160–164, 1957.
- [13] T. Foken and C.J. Nappo. *Micrometeorology*. Springer, 2008. ISBN 9783540746669. URL <http://books.google.se/books?id=2y0yHjsevvMC>.
- [14] Penélope Ramírez and José Antonio Carta. The use of wind probability distributions derived from the maximum entropy principle in the analysis of wind energy. a case study. *Energy Conversion and Management*, 47(15–16):2564 – 2577, 2006. ISSN 0196-8904. doi: <http://dx.doi.org/10.1016/j.enconman.2005.10.027>. URL <http://www.sciencedirect.com/science/article/pii/S0196890405002876>.
- [15] John C.. Wyngaard. Cambridge University Press, 2010. URL <http://dx.doi.org/10.1017/CB09780511840524>.
- [16] T. Laubrich. Statistical analysis and stochastic modelling of boundary layer wind speed. *The European Physical Journal Special Topics*, 174(1):197–206, 2009. ISSN 1951-6355. doi: 10.1140/epjst/e2009-01100-1. URL <http://dx.doi.org/10.1140/epjst/e2009-01100-1>.
- [17] Kanti V. Mardia and Peter E. Jupp. *Basic Concepts and Models*, pages 25–56. John Wiley & Sons, Inc., 2008. ISBN 9780470316979. doi: 10.1002/9780470316979.ch3. URL <http://dx.doi.org/10.1002/9780470316979.ch3>.
- [18] Rudy Calif. {PDF} models and synthetic model for the wind speed fluctuations based on the resolution of langevin equation. *Applied Energy*, 99(0):173 – 182, 2012. ISSN 0306-2619. doi: <http://dx.doi.org/10.1016/j.apenergy.2012.05.007>. URL <http://www.sciencedirect.com/science/article/pii/S0306261912003558>.

- [19] F. Böttcher, St. Barth, and J. Peinke. Small and large scale fluctuations in atmospheric wind speeds. *Stochastic Environmental Research and Risk Assessment*, 21(3):299–308, 2007. ISSN 1436-3240. doi: 10.1007/s00477-006-0065-2. URL <http://dx.doi.org/10.1007/s00477-006-0065-2>.
- [20] S. Lovejoy, D. Schertzer, and J. D. Stanway. Direct evidence of multifractal atmospheric cascades from planetary scales down to 1 km. *Phys. Rev. Lett.*, 86: 5200–5203, May 2001. doi: 10.1103/PhysRevLett.86.5200. URL <http://link.aps.org/doi/10.1103/PhysRevLett.86.5200>.
- [21] A. K. Gupta. Short-time averaging of turbulent shear flow signals. In *G. Treviño, J. C. Hardin, B. Douglas, and E. L. Andreas (eds.), Current Topics in Nonstationary Analysis, World Scientific, Singapore*, pages 159–173, 1996.
- [22] Ernest Agee Gluhovsky, Alexander. A definitive approach to turbulence statistical studies in planetary boundary layers. In *Journal of the Atmospheric Sciences*, volume 52, pages 1682–1690, 1994. doi: [http://dx.doi.org/10.1175/1520-0469\(1994\)051\(1682:ADATTS\)2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(1994)051(1682:ADATTS)2.0.CO;2).
- [23] George Treviño and Edgar L. Andreas. Averaging intervals for spectral analysis of nonstationary turbulence. *Boundary-Layer Meteorology*, 95(2):231–247, 2000. ISSN 0006-8314. doi: 10.1023/A:1002632004254. URL <http://dx.doi.org/10.1023/A/3A1002632004254>.
- [24] Smith OE. Direct evidence of multifractal atmospheric cascades from planetary scales down to 1 km. *Proceedings of the international symposium on probability and statistics in the atmospheric sciences, Honolulu, Hawaii, Amer. Meteor. Soc.*, pages 162–168, 1971.
- [25] Newmann MM McWilliams B and Sprevak D. The probability distribution of wind velocity and direction. *Wind Eng*, pages 269–273, 1979.
- [26] McWilliams B and Sprevak D. The estimation of the parameters of the distribution of wind speed and direction. *Wind Eng*, pages 227–238, 1980.
- [27] José A. Carta, Celia Bueno, and Penélope Ramírez. Statistical modelling of directional wind speeds using mixtures of von mises distributions: Case study. *Energy Conversion and Management*, 49(5):897 – 907, 2008. ISSN 0196-8904. doi: <http://dx.doi.org/10.1016/j.enconman.2007.10.017>. URL <http://www.sciencedirect.com/science/article/pii/S0196890407003767>.
- [28] K. Madsen, H. B. Nielsen, and O. Tingleff. Methods for non-linear least squares problems (2nd ed.), 2004. URL <http://f>.

- [29] Henri Gavin. The levenberg-marquardt method for nonlinear least squares curve-fitting problems. *Department of Civil and Environmental Engineering Duke University*, September 2011.
- [30] S. Gillijns, O.B. Mendoza, J. Chandrasekar, B. L R De Moor, D.S. Bernstein, and A. Ridley. What is the ensemble kalman filter and how well does it work? In *American Control Conference, 2006*, pages 6 pp.–, June 2006. doi: 10.1109/ACC.2006.1657419.
- [31] H. Auvinen, J. M. Bardsley, H. Haario, and T. Kauranne. The variational kalman filter and an efficient implementation using limited memory bfgs. *International Journal for Numerical Methods in Fluids*, 64(3):314–335, 2010. ISSN 1097-0363. doi: 10.1002/fld.2153. URL <http://dx.doi.org/10.1002/fld.2153>.
- [32] I. Bilik and J. Tabrikian. Optimal recursive filtering using gaussian mixture model. In *Statistical Signal Processing, 2005 IEEE/SP 13th Workshop on*, pages 399–404, July 2005. doi: 10.1109/SSP.2005.1628628.
- [33] Gideon Kowadlo and R. Andrew Russell. Robot odor localization: A taxonomy and survey. *Int. J. Rob. Res.*, 27(8):869–894, August 2008. ISSN 0278-3649. doi: 10.1177/0278364908095118. URL <http://dx.doi.org/10.1177/0278364908095118>.
- [34] MichaelW. Davis. Production of conditional simulations via the lu triangular decomposition of the covariance matrix. *Mathematical Geology*, 19(2):91–98, 1987. ISSN 0882-8121. doi: 10.1007/BF00898189. URL <http://dx.doi.org/10.1007/BF00898189>.
- [35] Rudy Calif, Richard Emilion, and Ted Soubdhan. Classification of wind speed distributions using a mixture of dirichlet distributions. *Renewable Energy*, 36(11):3091 – 3097, 2011. ISSN 0960-1481. doi: <http://dx.doi.org/10.1016/j.renene.2011.03.024>. URL <http://www.sciencedirect.com/science/article/pii/S0960148111001388>.
- [36] E. Erdem and J. Shi. Comparison of bivariate distribution construction approaches for analysing wind speed and direction data. *Wind Energy*, 14(1):27–41, 2011. ISSN 1099-1824. doi: 10.1002/we.400. URL <http://dx.doi.org/10.1002/we.400>.