Hourly Scale Model of Wind Magnitude and Direction Based on Statistical Analysis

Maayan Shimoni
Advisor- Anna Clarke
Faculty of Aerospace Engineering
Technion - Israel Institute of Technology

November 17, 2023

Abstract

Wind modelling is relevant to many fields, such as aircraft performance, wind energy, fatigue loads and others. Wind is a stochastic process that is effected by many factors such as season, geographic and meteorologic conditions, topography and the list goes on and on. Hence, wind has a wide range of behaviors and can be analyzed and modelled in many ways. The main goal of our research is to model wind in different scales, starting with a long-scale model that will provide one value per hour, and proceeding to a short-scale model that will use this value as a mean. The latter will model such phenomena as gusts, turbulence and dust devils. The shorter scale models will be more useful for aircraft in a context of landings, takeoffs and maneuvers.

In this project we will focus on the long scale model, which will eventually produce hourly values of wind speed and direction. This model can be useful for estimating wind energy production from wind turbines, or a wise design of runways for aircraft, since the performances of aircraft in takeoffs and landings is highly effected by the relative wind speed. For example, takeoffs are more effective when there is headwind, since this increases the airspeed and, hence, the lift. The typical wind behavior in the airport area is a crucial factor.

The wind magnitude and direction will be treated separately. The wind magnitude will be modelled using a method based on stochastic differential equations (SDEs). The resulting model will be able to produce stochastic processes with a given probability distribution, for both magnitude and direction. The model of the wind magnitude will also show an exponentially decaying autocorrelation (and autocorrelation coefficient).

Wind speed is usually modelled by non-Gaussian distributions. For instance, The Weibull distribution is a popular probability distribution used to model wind speeds, since it has shown a good fit to observed wind speed data in many locations around the world in a time scale of hours. Therefore, this distribution particularly will be used.

The direction of the wind will be modelled using a mixture of Von Mises distributions, which is the circular analogue of the normal distribution. This was chosen due to the flexibility of the model that can represent many different wind regimes. Real world data from a meteorological station in New Zealand Wellington, that is taken from the national climate database in [1], will be analyzed, and the model will be examined according to this data. A comparison between the statistical properties of the data and the stochastic process generated by the model is also provided.

1 Introduction

In this work we take from literature two known wind models - one for the wind magnitude and one for wind direction, and we will examine the two models separately so that each variable has its own known pdf.

For the magnitude of the wind: We will use a systematic method to build a wind speed model based on stochastic differential equations. This model is taken from [3] and produces a stochastic process with a desired probability distribution (such as Weibull, Beta, Gamma, etc), and an exponentially decaying autocorrelation function. Exponentially decaying autocorrelations are very common in context of wind modelling, both for hourly wind speed measurements in the time frame of hours [5], and for wind speed measurements on a 1-s basis in the time frame of minutes [7]. In many wind regimes, as time goes by, the effect of former states is weakening. Recent values have a stronger influence on the future values compared to previous values. This statement is valid usually in a time frame of few hours, and outside this range, non-stationary phenomena related to seasonal effects can be observed, as was explained in [5]. The model strives to reproduce an exponential autocorrelation, however does not guarantee this due to reasons reviewed in Section 2. We must note that the autocorrelation behavior of wind can vary depending on the location, time frame, and atmospheric conditions. Therefore, the validity of the proposed model is limited to cases for which the autocorrelation is approximately decaying exponentially.
From a statistical point of view, the wind speed is characterized by its probability distribution and autocorrelation, hence the wind speed model should be able to reproduce these defined characteristics. There are probability distributions that are most frequently used in the literature in context of wind modelling, and the most common one - Weibull, will be used in this research. In [3] the approach is to transform Ornstein–Uhlenbeck (OU) process to stochastic processes with various distributions using a memoryless transformation. In [21] only the Weibull distribution is used due to its commonness in wind statistical modelling, and a decent fit to the data is shown. The model exhibited in these two articles will be used here. It is important to note that transformation of the OU process is not the only way to model wind in an hourly scale described in the literature. In [22] an SDE based wind model is constructed from different probability distributions by customizing the drift and the diffusion terms of the SDE, using regression theory. This model also assures an exponentially decaying autocorrelation.

For the wind direction: We will use a directional model given in [8], comprised of a finite mixture of von Mises (vM-pdf) distributions. The von Mises distribution, also known as the circular normal distribution, is a continuous probability distribution on the circle, and is commonly used for directional statistics. In [8] the model is used for directional data of wind, from two different stations in the Canarian Archipelago. As the directional wind speed in this location has a seasonal behaviour, the proposed model was applied for two monthly intervals: winter and autumn months (January–March and October–December), and spring and summer months (April–September). The same model was applied in [9], as four different stations were selected, which are representative of the most complex wind speed and direction distributions in the Canary Islands. In this article there was no separation to seasons. Both articles showed valid estimations for different wind regimes, proving the flexibility of the proposed model. Important to note that in [9], a joint probability density function of wind speed and direction was used, as the wind direction and speed can be modelled as either independent or dependent variables. However, the model produced slightly better results when the variables were treated as dependent. In [9] and [8] the parameters of the mVM (multiple Von Mises) pdf were estimated using the Least Squares and Maximum Likelihood methods. In this project the Maximum Likelihood method will be adopted. This algorithm is the key to adjust a distribution of VM functions to a variety of real wind distributions.
In [17] the isotropic model is suggested. The isotropic Gaussian model is derived from the following assumptions: (1) the wind speed component along the prevailing wind direction is normally distributed with non-zero mean and a given variance; (2) the orthogonal wind speed component is independent and normally distributed with zero mean and the same variance. A joint density function of wind speed and direction can be obtained from polar coordinate transformation, and a distribution of the direction alone can be computed by integrating over the speed.
In [20] the anisotropic Gaussian model is proposed, it uses the same hypotheses as the isotropic model,

but the two variances do not have to be the same.
In [9] the suggested model, based on von Mises distributions, is compared to the isotropic and anisotropic models. The proposed VM model provided better fits to the real world data in all analysed cases. Hence the model given in [9] will be used.

Project Organization: One section is dedicated to each variable - the wind speed magnitude, and the wind speed direction. Next, a combination of the two is considered, in order to create a wind field in space.

2 Wind Speed

2.1 The Ornstein–Uhlenbeck process

The governing SDE is:

$$\begin{gathered} dX(t) = a(X(t),t)\cdot dt + b(X(t),t)\cdot dW(t) \\ X(0) = X_0 \end{gathered} \tag{1}$$

where $X(t)$ is the variable at time t; X_0 is a deterministic value or a random value; $W(t)$ is the Wiener process at time t and it is used to represent the integral of a white Gaussian noise; $dW(t) \sim N(0,dt)$ is the random increment of the Wiener process; $a(x(t),t)$ is the drift term; $b(x(t),t)$ is the diffusion term. The drift term is the deterministic part of the system and directly determines the expectation. The diffusion term characterizes the variance of $X(t)$, and determines the noise intensity in the system. This kind of equation can be viewed as an ordinary differential equation where an additional term is included to model the stochastic dynamical behavior related to variable $X(t)$.
Despite the fact that the differential formulation is widely used in the literature, since Wiener process is nowhere differentiable, in a strictly mathematical sense, Eq.1 is not fully correct. The truly correct form of the equation is the integral form:

$$X(t) - X_0 = \int_0^t a(X(t),t)\cdot dt + \int_0^t b(X(t),t)\cdot dW(t), \quad t\in[0,T] \tag{2}$$

The first integral is an ordinary Riemann-Stieltjes integral, and the second one is a stochastic Ito's integral. Nevertheless, in this project we will be using the heuristic differential form of the SDE.
For stationary processes:

$$\begin{gathered} a(X(t),t) = a(X(t)) \\ b(X(t),t) = b(X(t)) \end{gathered} \tag{3}$$

We will use a specific form of the SDE where $a(X(t),t) = -\alpha\cdot(X(t)-\mu)$ and $b(X(t),t) = b$:

$$dX(t) = -\alpha\cdot(X(t)-\mu)dt + b\cdot dW(t), \quad t\in[0,T] \tag{4}$$

The initial condition of X is:

$$X(0) \sim N(\mu, \frac{b^2}{2\alpha}) \tag{5}$$

The resulting process $X(t)$ is the Ornstein–Uhlenbeck process and it has the following properties:

$$E[X(t)] = E[X_0]e^{-\alpha t} + \mu\cdot(1-e^{-\alpha t}) \tag{6}$$

$$Var[X(t)] = \frac{b^2}{2\alpha}(1-e^{-2\alpha t}) + Var[X_0]\cdot e^{-2\alpha t} \tag{7}$$

$$R_{XX}[t_1,t_2] = \frac{b^2}{2\alpha}\left(e^{-\alpha\cdot|t_1-t_2|} - e^{-\alpha(t_1+t_2)}\right) + Var[X_0]\cdot e^{-\alpha(t_1+t_2)}, \quad \forall\ t_1,t_2\in[0,T] \tag{8}$$

For simplicity, the Ornstein-Uhlenbeck process is adapted to a standard Normal distribution, so $\mu = 0$, and $b = \sqrt{2\alpha}$. When assuming $X_0 \sim N(0,1)$:

$$E[X(t)] = \mu = 0 \tag{9}$$

$$Var[X(t)] = \frac{b^2}{2\alpha} = 1 \tag{10}$$

$$R_{XX}[\tau] = \frac{b^2}{2\alpha}e^{-\alpha\cdot\tau} = e^{-\alpha t} \tag{11}$$

In this case, the first moment does not vary with respect to time and the autocorrelation only depends on τ, so the process is wide sense stationary. Moreover, since $X(t)$ is a Gaussian process, the process becomes strict sense stationary. It is important to note that the process is also Markov, meaning that the future value is independent of its past history, and depends only on the current value.
For obtaining a process that is distributed differently it is required to apply a memoryless transformation on the OU process.

2.2 Memoryless transformation

As mentioned previously, the two-parameter Weibull distribution has shown a good fit to the wind empirical distributions in many locations around the world for long time scales, and is the most frequent choice for representation of wind speed data for wind energy calculation purposes. However, there are other non-Gaussian distributions that are commonly used for wind models, such as Rayleigh, Beta, Gamma, Nakagami and so on. These distributions can be utilized by applying a memoryless transformation.
For a Gaussian process $X(t)$, the transformation is:

$$Y(t) = g[X(t)] = F^{-1}[\phi[X(t)]] \tag{12}$$

where ϕ is the standard Gaussian cumulative distribution function, and F is the non-Gaussian continuous cumulative distribution function. The excepted stochastic process $Y(t)$ will have the desired distribution F. This transformation is said to be memoryless since the value of the new process $Y(t)$ at an arbitrary t depends only on the value of $X(t)$ at t. It is known that since $X(t)$ is a Gaussian process, $Y(t)$ will be Gaussian only if the transformation is linear. So we will receive a non-Gaussian distribution only if the transformation is non-linear. Additionally, $Y(t)$ is stationary and Markov, since $X(t)$ is stationary and Markov and since the transformation is memoryless.
If the OU process is not standard normal, it is necessary to normalize it to obtain the standard process:

$$Y(t) = g[X(t)] = F^{-1}\left[\phi\left[\frac{X(t)-\mu}{\frac{b}{\sqrt{2\alpha}}}\right]\right] \tag{13}$$

where ϕ, the standard Gaussian cumulative distribution function, is:

$$\phi\left[\frac{X-E[X]}{\sqrt{Var[X]}}\right] = \frac{1}{2}\left(1 + erf\left[\frac{X-E[X]}{\sqrt{2Var[X]}}\right]\right) \tag{14}$$

If X is standard normal, then:

$$\phi[X] = \frac{1}{2}\left(1 + erf\left[\frac{X}{\sqrt{2}}\right]\right) \tag{15}$$

where erf, the error function, is defined as:

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt \tag{16}$$

In this project we will use the Weibull distribution as the distribution of the wind model, while keeping in mind that the model is versatile and enables the use of other distributions. For the Weibull distribution F is computed as follows:

$$F_{Weibull}(u) = 1 - exp\left[\left(\frac{u}{\lambda}\right)^k\right], \quad \forall u > 0 \tag{17}$$

where $\lambda > 0$ and $k > 0$ are the scale and shape parameters, respectively. The resulting process Y, obtained by applying the transformation, is a Weibull distributed stochastic process. The statistical properties of $Y(t)$ are:

$$E[Y(t)] = \mu_w = \lambda \cdot \Gamma\left[1 + \frac{1}{k}\right] \tag{18}$$

$$Var[Y(t)] = \sigma_w^2 = \lambda^2 \cdot \Gamma\left[1 + \frac{2}{k}\right] - \mu_w^2 \tag{19}$$

$$\rho_{YY}(t_i,t_j) \approx \rho_{XX}(t_i,t_j) = R_{XX}(t_i,t_j) = e^{-\alpha|t_j-t_i|} \tag{20}$$

where Γ is the Gamma function and is defined as follows:

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt \tag{21}$$

Eq.20 states that the autocorrelation coefficient of $Y(t)$ is approximately same as the autocorrelation coefficient of the Ornstein-Uhlenbeck process. This result has been determined empirically by analyzing a number of realizations. The goodness of this appoximation depends on the parameters of the Weibull distribution, as discussed in Section 2.3. Important to note that this is not true for the autocorrelation itself because of the difference in the definition. For a stationary process X, the definition of autocorrelation function between times t and $t+\tau$ is

$$R_{XX} = E[X_t X_{t+\tau}] \tag{22}$$

And the autocorrelation coefficient is:

$$\rho_{XX} = \frac{E[(X_t - \mu)(X_{t+\tau} - \mu)]}{\sigma^2} \tag{23}$$

For OU process the two terms are identical because $\mu = 0$ and $\sigma^2 = 1$. For the Weibull process it is not the case ($\mu \neq 0$ and $\sigma^2 \neq 1$), so there is a significant difference between the autocorrelation and autocorrelation coefficient.

It is worth noting that in [3], five other distributions were tested in addition to Weibull: Three Parameter Beta, Four-parameter generalized Gamma, Log-Pearson 3, Nakagami, and One-parameter Rayleigh. The autocorrelation coefficients of all six generated processes were exponentially decaying, as in Eq.20. Therefore, this approximation is also correct for other kinds of PDFs. Parameters λ and k are directly taken from the Weibull fit of the wind speed data, while α can be easily computed from the exponential fit to the autocorrelation coefficient of the wind speed data.

2.3 Numerical Results

Firstly, in order to produce the OU process, a numerical integration must be applied. This will be done using Milstein integration. The Milstein approximation to the true solution X is the Markov chain defined as follows:

$$X_{i+1} = X_i + a[X_i] \cdot \Delta t + b[X_i]\Delta W_i + \frac{1}{2} b[X_i] \frac{\partial b}{\partial x}[X_i]((\Delta W_i)^2 - \Delta t) \tag{24}$$

However, in our case $b = \sqrt{2\alpha}$ and does not depend on X, hence:

$$X_{i+1} = X_i - \alpha \cdot X_i \cdot \Delta t + \sqrt{2\alpha} \cdot \Delta W_i \tag{25}$$

where Δt is the integration step, and $\Delta W_i \sim N(0, \Delta t)$

Secondly, the transformation is applied for X, where the inverse cumulative distribution is known from Eq.17 and equals to:

$$F_w^{-1} = \lambda \cdot [-ln(1-u)]^{\frac{1}{k}} \tag{26}$$

In order to extract specific constants we will deal with wind data that has been used in [21]. This data was measured in Baring head, New Zealand (latitude 41°25' south, longitude 174°52' east), during the whole year of 2014. Each element of the data is an hourly mean wind speed: 8760 elements in total. The decision to use such a long time period stems from the assumption that the process is ergodic, meaning that the statistical properties of the process observed over a long time period (time average) will be similar to the statistical properties of the process observed across many different samples of the process (ensemble average).

In Fig. 1 a histogram of the data PDF is exhibited with the Weibull fit. The Weibull distribution describes the data set well, mainly in the higher speeds range. There are deviations due to the randomness of the data, and the quantity of values may not be big enough to observe a better fit. However, due to the general similarities to the Weibull distribution (the right tail that characterizes the Weibull distribution), and due to its simplicity, we will continue with this distribution.

The autocorrelation coefficient of the data with an Exponential fit is shown in Fig. 2. The autocor-

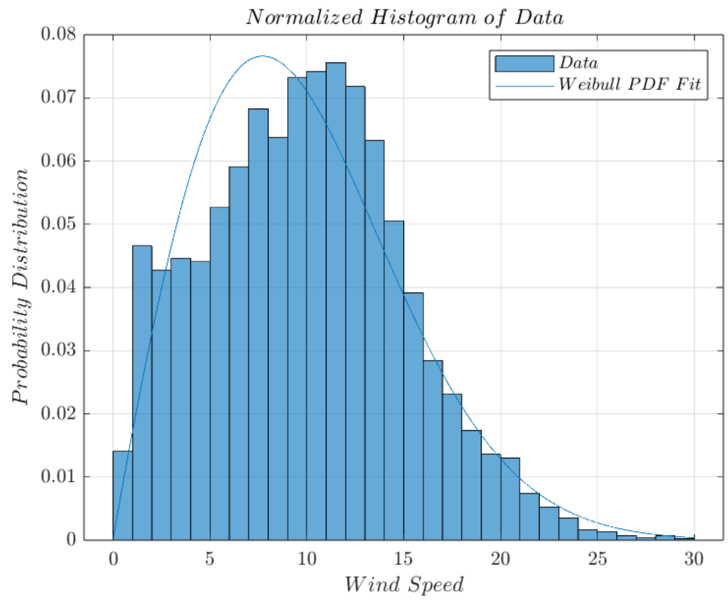

Figure 1: Data histogram and the Weibull PDF fit

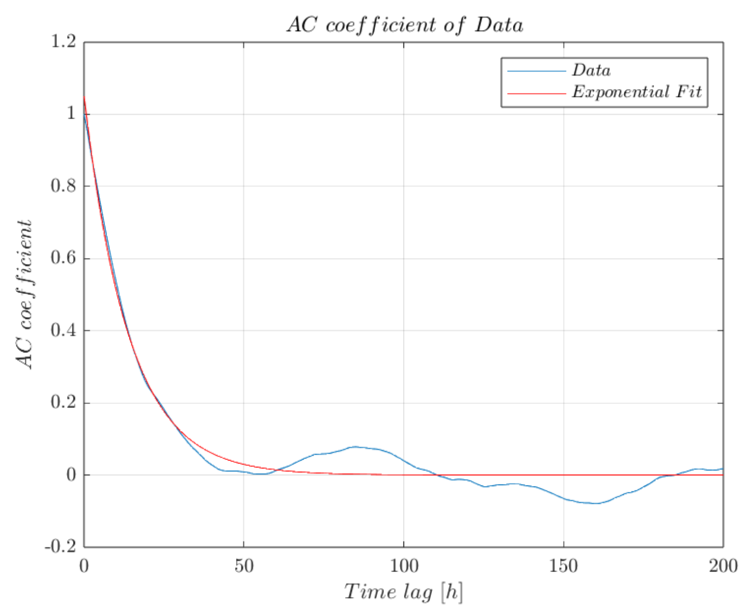

Figure 2: Data-driven autocorrelation coefficient and the exponential fit

relation coefficient was taken for future comparison to the generated process Y (as was introduced in Eq.20). The autocorrelation coefficient of the given data is decaying, however fluctuates around zero after 50 hours lag. The periodicity after 50 hours is a usual phenomena, caused by prevailing regional wind patterns and systematic recurrence of local wind regimes. For expanded explanations in the matter see [5]. Therefore, the exponentially decaying approximation is acceptable for the considered location only for time frame of approximately 50 hours. The exponent coefficient of the fit is -0.07142. We will difine this coefficient to be the parameter α of the SDE.
Overall, the parameters of the fits are:

$$\begin{aligned} \alpha &= 0.07142 \\ \lambda &= 11.0792 \\ k &= 1.96744 \end{aligned} \tag{27}$$

λ and k are from the Weibull fit, and α is from the exponential fit. These parameters will be used for generating the wind trajectories based on the suggested model. A few simulated wind trajectories based on the extracted parameters can be observed in Fig. 3. The statistical properties of the process

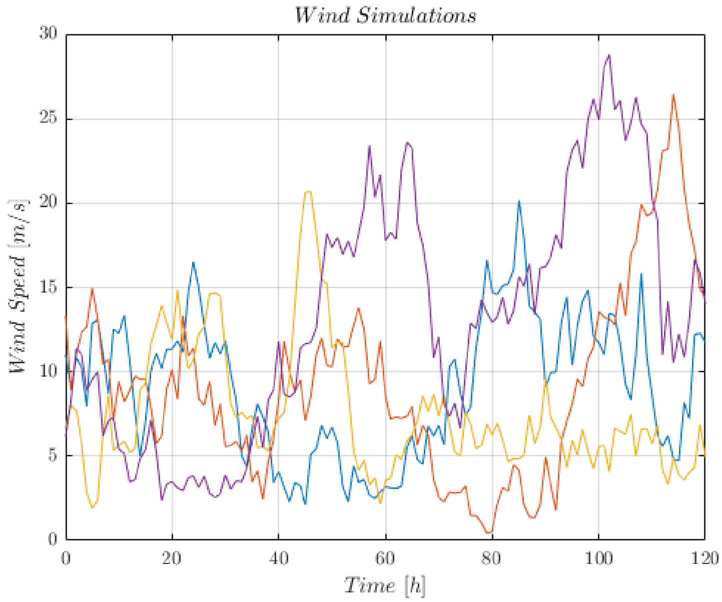

Figure 3: Four wind speed trajectories generated by the model

generated by the model will be obtained on the basis of 1000 realizations. The mean and standard deviation of the fitted Weibull distribution are:

$$\begin{aligned} \mu &= 9.822 \\ \sigma &= 5.2106 \end{aligned} \tag{28}$$

In comparison to the data:

$$\begin{aligned} \mu_{data} &= 9.8623 \\ \sigma_{data} &= 5.1267 \end{aligned} \tag{29}$$

By applying ensemble average on 1000 realizations, we can find the mean and the standard deviation of the generated stochastic process, as shown in Fig. 4. As expected, there is a match between the mean and standard deviation of the model to the properties of the Weibull fit. The statistical properties match from the very beginning of the time axis, because the process is stationary from $t = 0$ due to our choice of the initial condition as a standard normal variable. Important no note that the process is ergodic, meaning that we can also find the mean and standard deviation from time average if we simulate one realization for long enough time, as shown in Fig. 5. The mean and standard deviation converge with time to the statistical mean and standard deviation, as expected.

As was done for the mean and standard deviation, we will find the autocorrelation coefficient using ensemble average and time average. It will be computed from $t = 0$ since the process is stationary. If the initial value X_0 were deterministic, the process would not be stationary from initial time (as can be seen in Eq. 6,7,8). In this case, we would have to find from which time t the process may be considered stationary. The autocorrelation coefficient shown in Fig. 6 was computed from the ensemble average of over 1000 realizations, each one was simulated for 120 steps ($dt = 1[hr]$, $t_{final} = 120[hr]$). The autocorrelation coefficient of Y is almost identical to the theoretical expression $\rho_{yy}(\tau) = e^{-0.07142\tau}$. This expression represents the exponential fit of the data (recall Fig.2), and it is both the autocorrelation $R_{XX}(\tau)$ and the autocorrelation coefficient $\rho_{XX}(\tau)$ of process X. The plot clearly shows that the approximation in Eq.20 is valid, and can be used for this example, for the time scale of 50 hours (approximately two days). In order to see how many realizations are needed for obtaining an accurate estimate of ρ_{YY}, we will compute the RMSE as a function of a number of sample functions ($t_{final} = 120[hr]$ remains as above). The RMSE is decaying significantly until 1000 realizations, and

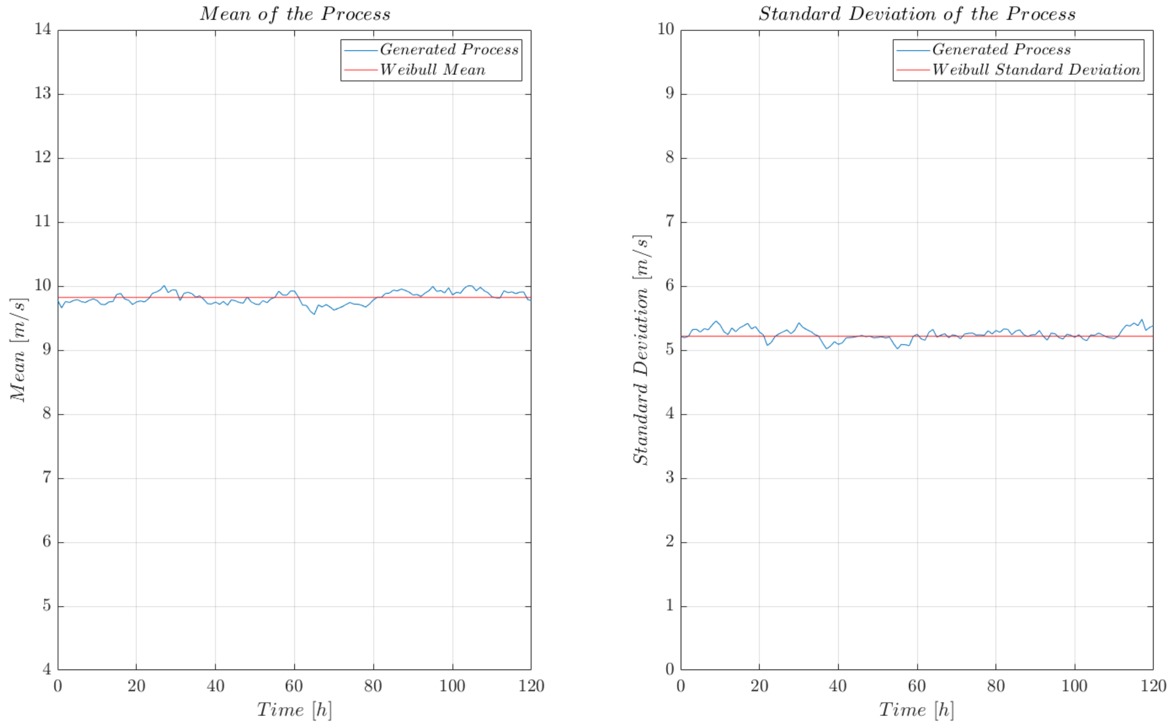

Figure 4: Statistical properties of the Weibull distribution vs. the generated process using ensemble average

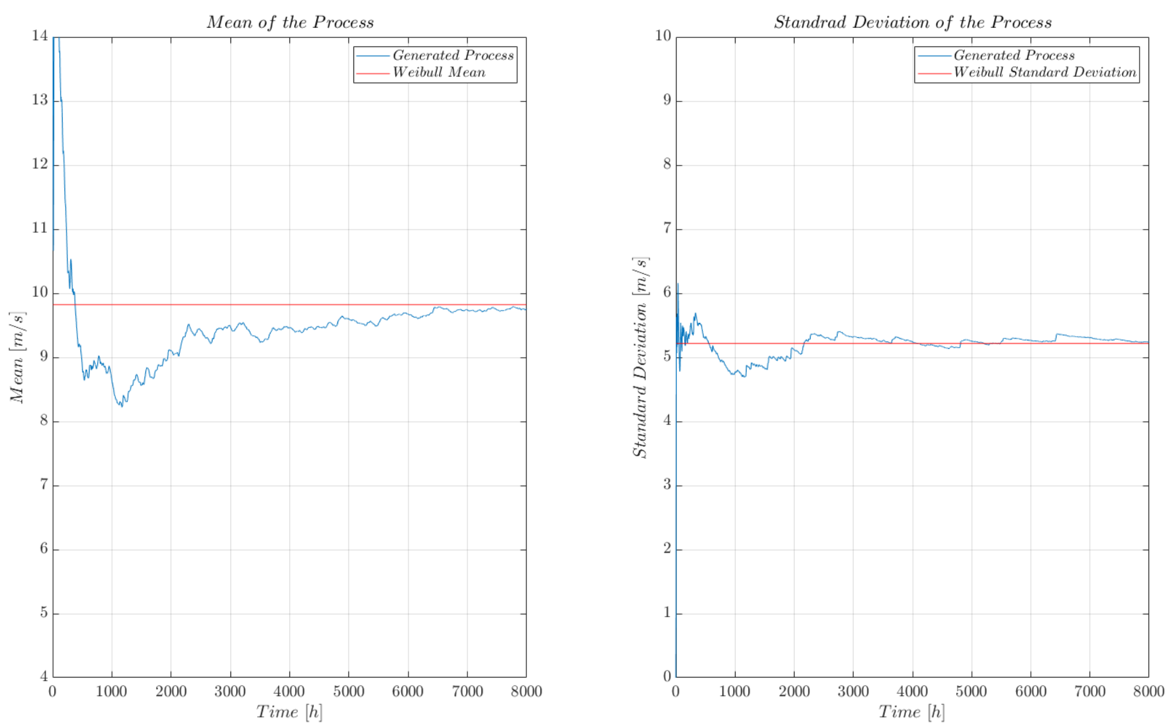

Figure 5: Statistical properties of the Weibull distribution vs. the generated process using time average

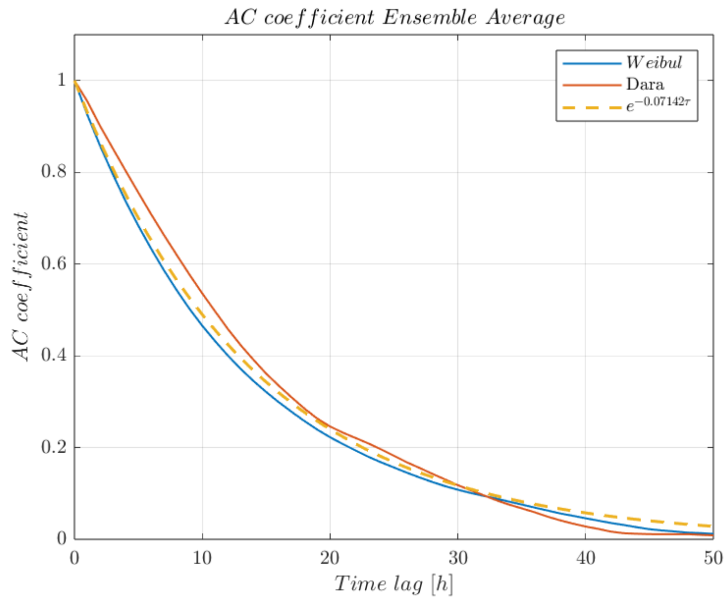

Figure 6: The autocorrelation coefficient of the process computed by ensemble average vs. theoretical and data

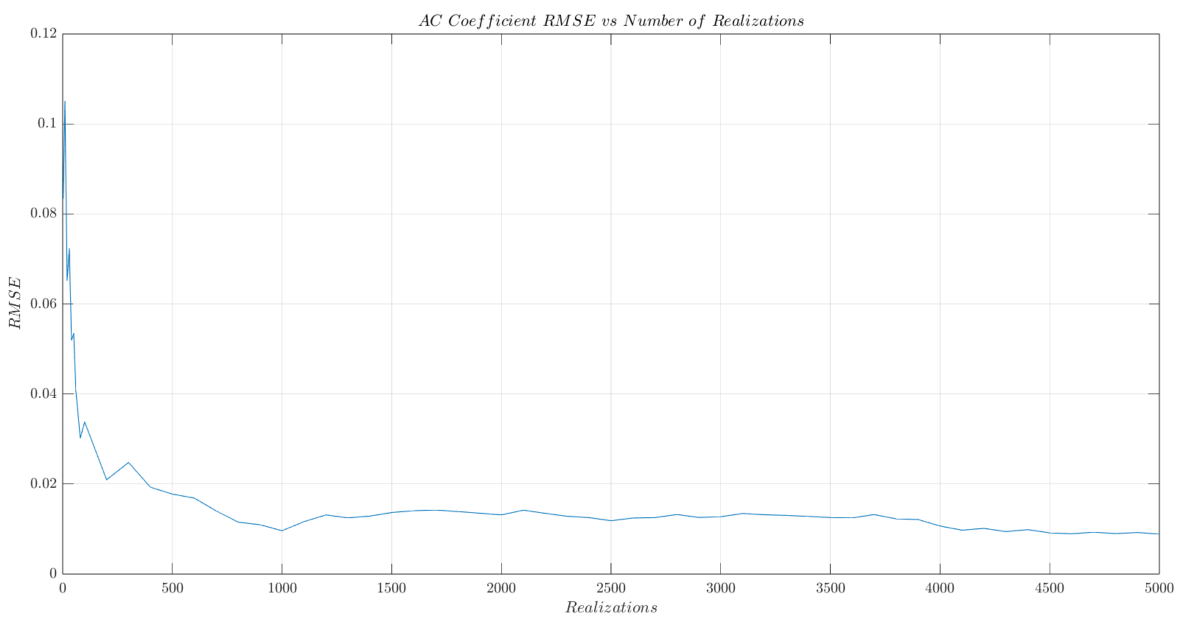

Figure 7: RMSE between ρ_{YY} computed by ensemble average and the analytical approximation $e^{-0.07142\tau}$

from this value the RMSE stabilizes around 0.01. There is no significant improvement from around this value, so there is no advantage in running more realizations.
Since the process is ergodic, it is expected that the estimated correlation computed from one realization will approach the analytical curve as t becomes larger. This is demonstrated in Fig. 8 for $t_{final} = 10,000[H]$. The autocorrelation coefficient computed from this one realization is almost identical to the theoretical expression, same as the ensemble average.
We can also understand the ergodicy by the correlation, by a graph showing the RMSE between the calculated autocorrelation coefficient of one realization and the analytical expression, as a function of the final time of the generated process: As we can see the RMSE is getting smaller as the final time is

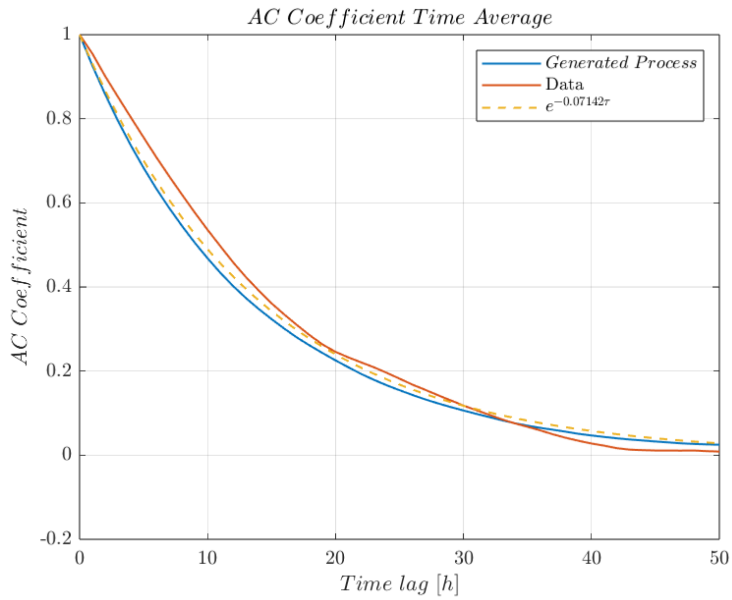

Figure 8: Autocorrelation coefficient of the process computed by time average vs. theoretical and data

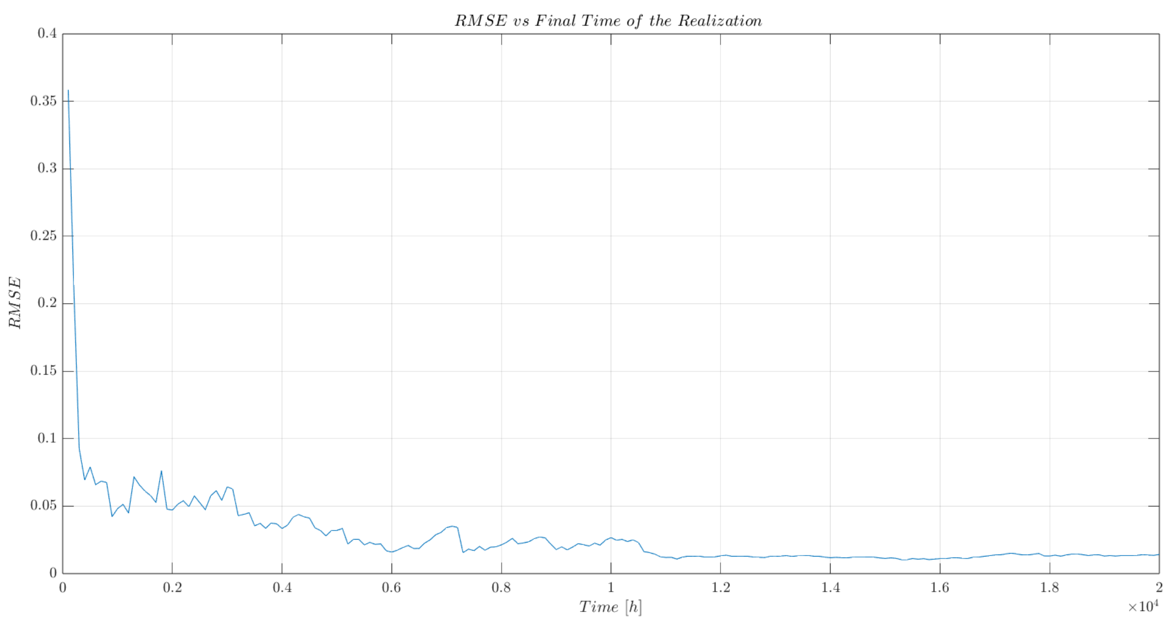

Figure 9: RMSE between ρ_{YY} computed by time average and the analytical approximation $e^{-0.07142\tau}$

getting bigger. From $t_{final} \approx 12,000$ $[hr]$ the RMSE stabilizes at 0.01. From around this value there is no significant improvement in accuracy of the estimation of ρ_{YY}.
As it was mentioned in subsection 2.2, the goodness of the approximation in Eq.20 depends on the Weibull parameters. We will test this statement by calculating the RMSE between the computed autocorrelation coefficient of the Weibull process (from time average of 10000 $[hr]$, although ensemble average can be used as well) and the theoretical ρ_{YY} as a function of k and λ, as shown in Fig. 10. We can conclude that the value of the shape parameter k has a significant effect on the RMSE. For values above $k = 2.5$ the RMSE converges to 0.01. The scale parameter λ, on the other hand, barely effects the error. This graph is useful for future work of modelling wind with the Weibull distribution, since it indicates when the approximation of Eq.20 is correct.

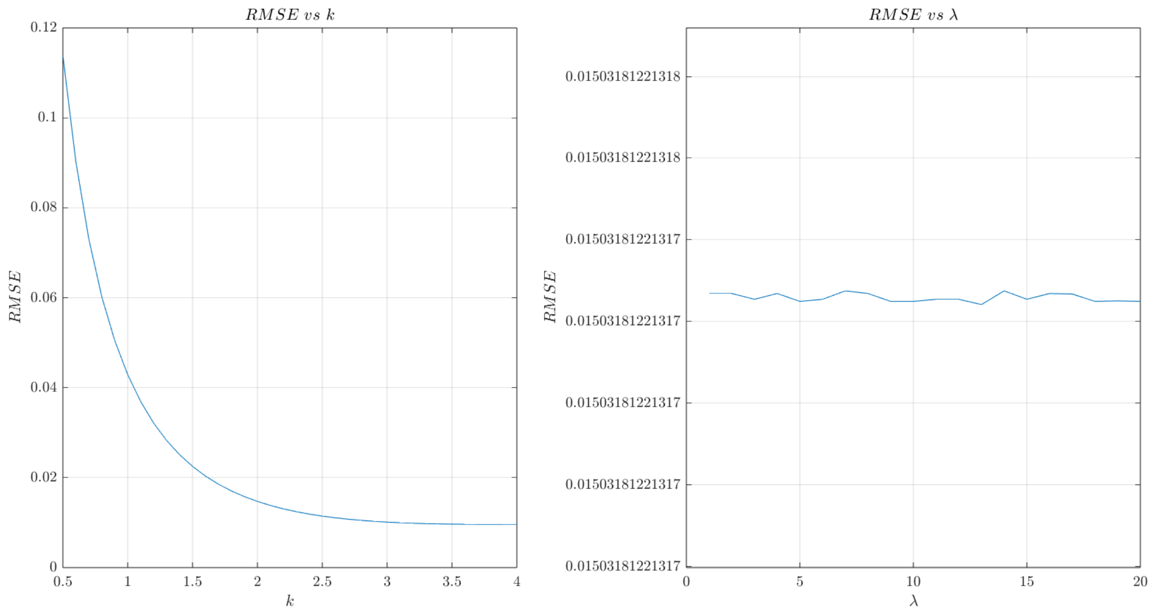

Figure 10: RMSE between the autocorrelation coefficient of the generated process computed from time average and the theoretical autocorrelation coefficient as a function of k and λ

We are also interested in investigating the PSD of the process, and comparing it to the PSD of the real world data. Power spectral density is the distribution of the power of a signal in the frequency domain. Analyzing the PSD of the wind can provide insights on the behavior of the wind, if it is turbulent or laminar, and it also may by relevant when we refer to aircraft stability and natural frequencies.
The definition of periodogram for a continuous time process is as follows:

$$p_x(f) = \frac{1}{T}\tilde{x}(f)\tilde{x}^*(f) = \frac{1}{T}\int_0^T x(\zeta)e^{-j2\pi f\zeta}d\zeta \int_0^T x(\eta)e^{j2\pi f\eta}d\eta \tag{30}$$

However, in the simulation the expression will be used numerically. The numerical form of $\tilde{x}(f)$ is:

$$\tilde{x}(f) = \int_0^T x(\zeta)e^{-j2\pi f\zeta}d\zeta = \sum_{k=1}^{\frac{T}{\Delta t}} x_k e^{-j2\pi fk\Delta t}\Delta t \tag{31}$$

With this implementation, we can find the PSD of the data. For comparison purposes, we will simulate the process we modelled for the same time period - one year, with a time step of 1 hour (8760 values), and compute its PSD, as shown in Fig. 11. It is clear that the lower frequencies are dominant, meaning that the wind is characterized by constant prime modes, while higher frequencies modes (that can be referred to as turbulence or circulation) are relatively minor. These results are expected since the data and the generated values are hourly mean values, hence phenomena of high frequencies such as turbulence and gusts, can not be observed in this long time scale results. Thus, the wind speed in hourly scale can be considered as a simple low-pass filter corresponding to the daily, seasonal, and annual effects (long term effects).

3 Wind Direction

3.1 Directional Data

The previously investigated model dealt only with the wind magnitude, but hasn't addressed wind direction. Wind direction is an essential information for aircraft performances in landings and takeoffs, wind energy, and many other research fields. Airports plan and design the use of runways on the

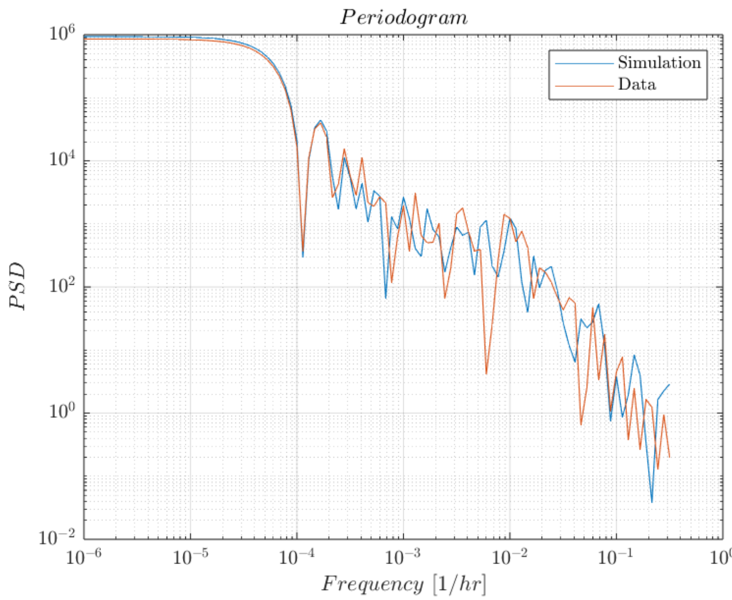

Figure 11: Power spectral density of data and the simulated process

basis of the wind direction. For the wind energy research, the direction of the wind is important for determining the location and the orientation that turbines should be imposed. Wind direction can change dramatically with the season, hence from looking on the direction distribution through one year, a meaningful data may not be found. We suggest taking data from a certain season over several consecutive years and fitting a distribution.
The seasons in Wellington fall under the following months:

- Summer – December, January and February
- Autumn – March, April and May
- Winter – June, July and August
- Spring – September, October and November

We will use data from four consecutive years- March 2014 until March 2018. A large amount of samples will raise the chance to find reliable statistical distribution and properties. Obviously the separation of the seasons is not deterministic and may change, however since we examine several years, perhaps special patterns of each season will be apparent. One way to exhibit the directional data is a wind rose, which is a graphic tool, usually used by meteorologists to give a succinct view of how wind speed and direction are typically distributed at a particular location. The code for the generation of the wind roses is taken from [12], and the results are shown in Fig. 12. As we can see this area is characterized by primarily northern and southern winds in all four seasons. There are slight differences between the winter/ autumn and spring/ summer seasons: in the spring and summer the south winds also tend east. However, there are no significant differences between the seasons, hence we will investigate the wind without referring to specific seasons (as was done in [9]). This does not necessarily happen in other locations, which may have clear seasonal differences, and in which case the directional modelling should consider the season. Furthermore, the model that will be used for wind direction is very flexible, and it was already proved in the literature that it is capable of describing many kinds of directional behaviors. Indeed, the histograms in Fig. 13 are not significantly different from one another and roughly represent the same wind behavior. The wind rose for the whole year of 2014, which corresponds to the wind speed data that was used in section 2 is shown in Fig. 14. This rose diagram is not dramatically different from any of the other four diagrams describing different seasons for the years 2014-2018, hence we will refer to this diagram as a reliable representation of the typical wind in the tested location. The model we will develop in this project will be examined according to this data, and the model will be two dimensional in order to match the collected data.

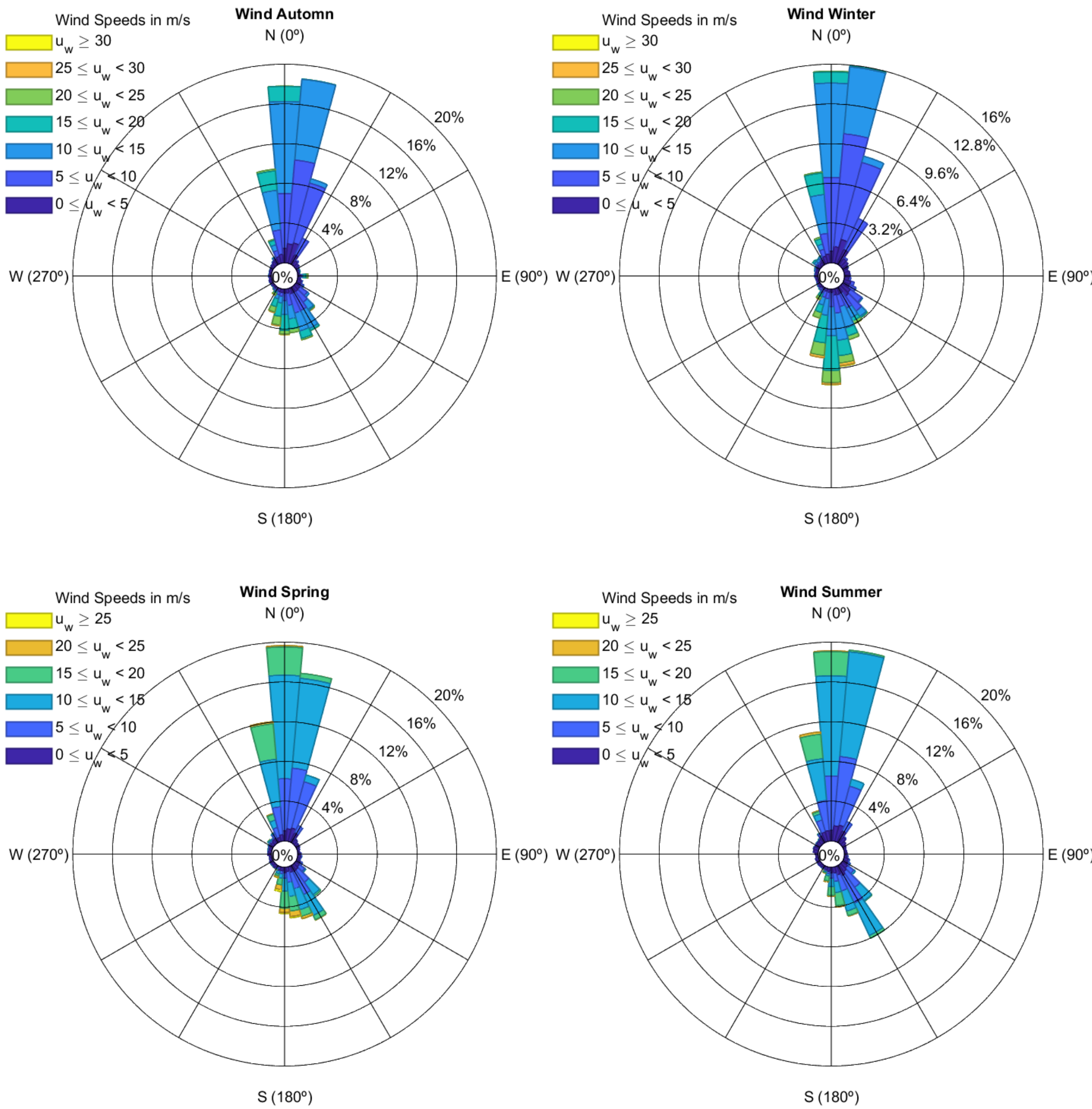

Figure 12: Wind roses for the four seasons for a duration of four years

3.2 Mixture of Von Mises Distributions as a Wind Direction Model

Von Mises distribution is a continuous probability distribution on the circle. It is often used to model wind direction data because it is well-suited to circular or directional data that exhibit a unimodal (single-peaked) distribution around a central direction. However, it may not be appropriate for data with complex multimodal (multi-peaked) distributions. In such cases, it is possible to use a mixture of Von Mises distributions.

The proposed continuous probability distribution $mVM(\theta)$ is comprised of a sum of N different Von Mises probability densities:

$$mVM(\theta) = \sum_{i=1}^{N} \omega_i VM_j(\theta) \tag{32}$$

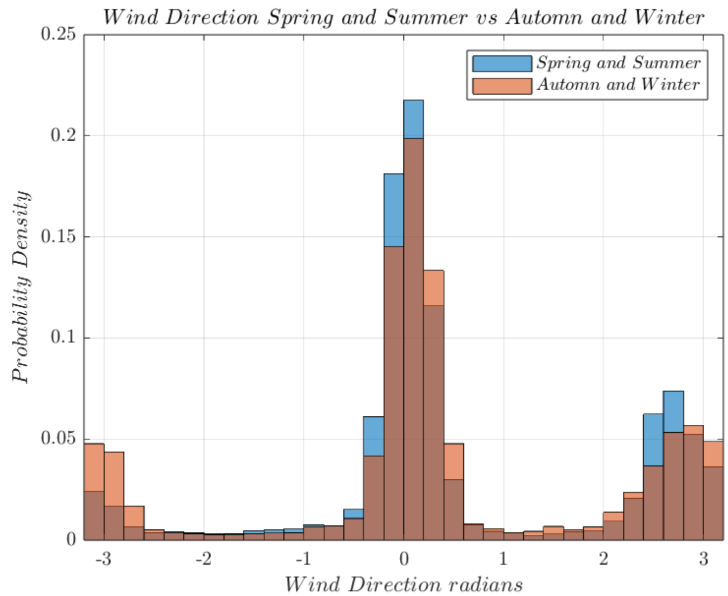

Figure 13: Comparison between the wind direction histograms of spring and summer, and autumn and winter

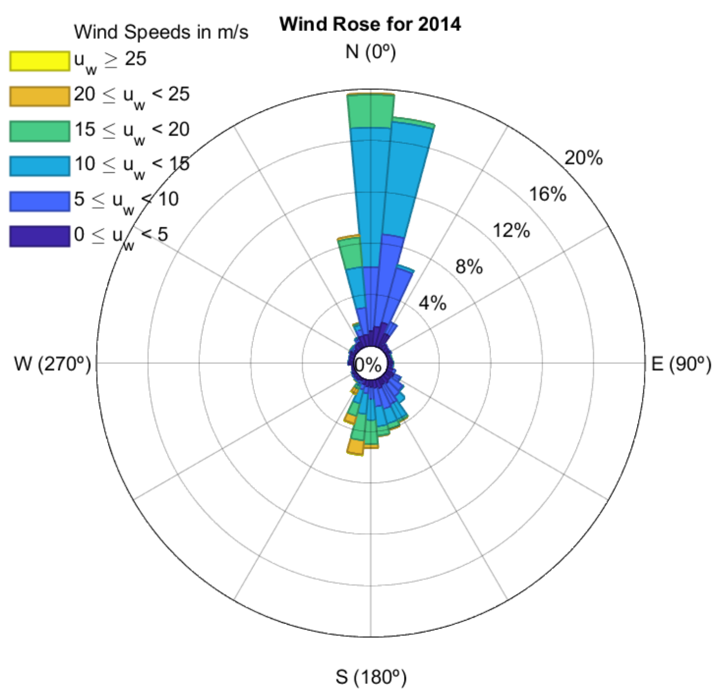

Figure 14: Wind Rose for the year 2014

where ω_i are non-negative quantities that satisfy the following conditions:

$$0 \leq \omega_i \leq 1 \quad (i = 1, ...N), \quad \sum_{i=1}^{N} \omega_i = 1 \tag{33}$$

A Von Mises probability density function is:

$$f_{VM}(\theta) = \frac{1}{2\pi I_0(k_i)} exp\left[k_i cos(\theta - \mu_i)\right] \\ -\pi \leq \theta \leq \pi \tag{34}$$

where μ_i is the mean direction, and the parameter k_i is known as the concentration parameter. The distribution is unimodal and is symmetrical about $\theta = \mu_i$. For large values of k, the distribution is concentrated around the mean direction, and when $k = 0$, the pdf is a uniform distribution on $[0, 2\pi]$. $I_0(k_i)$ is the modified Bessel function of the first kind and order zero:

$$I_0(k_i) = \sum_{j=0}^{\infty} \frac{1}{(j!)^2} \left(\frac{k_j}{2}\right)^{2j} \tag{35}$$

Therefore, in total, the wind direction pdf is:

$$f_{mVM}(\theta) = \sum_{i=1}^{N} \frac{\omega_i}{2\pi I_0(k_i)} exp\left[k_i cos(\theta - \mu_i)\right] \\ -\pi \leq \theta \leq \pi \tag{36}$$

Important to note that zero degrees matches the north direction and $\pi, -\pi$ is the south (positive is clockwise). The parameters of the mixture of von Mises distributions that fit the data set, are approximated based on the Expectation-Maximization algorithm. The code is taken from [10], and explanations on the algorithm itself can be found in [4].

The suggested model is very flexible and can match many wind behaviors, since the superposition of the density functions can produce a variety of distributions with one or several modes. We want to examine the fit of the model to the data set (2014, Wellington), while looking at different number of components of VM distributions, as shown in Fig. 15. It is worth noting, that each distribution

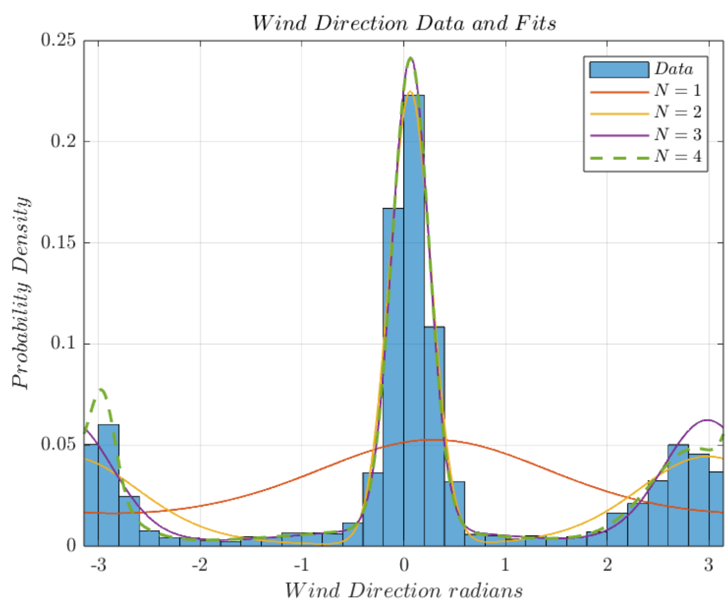

Figure 15: Frequency histogram of wind directions with different mVM fits

function has the same value in the edges $[-\pi, \pi]$. It is the same angle so it has to have the same value of probability. The parameters of the distribution functions used in Fig. 15 are given in Table 1.
The fit for $N = 1$ is a single von Mises pdf, and it exhibits the characteristics of a von Mises distribution - it has a mean value and a concentration factor that effects the dispersion around the mean. However, it is obvious that this fit hardly represents the data, while the alternative fits, (comprised from multiple von Mises distributions) represent the data decently. The main difference between the fits for higher N, is the level of accuracy of the fit in the edges $-\pi$ and π. It appears that the mixtures that include more distributions in the superposition (bigger N), are more accurate.
In order to verify this assumption, we will use the coefficient of determination R^2, defined as:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{37}$$

Numerical values of the parameters of the mVM-pdfs			
PDF 1	PDF 2	PDF 3	PDF 4
$\mu_1 = 0.294$ $k_1 = 0.586$ $\omega_1 = 1$	$\mu_1 = 0.064$ $k_1 = 23.797$ $\omega_1 = 0.578$	$\mu_1 = 2.984$ $k_1 = 5.224$ $\omega_1 = 0.3442$	$\mu_1 = 3.007$ $k_1 = 5.475$ $\omega_1 = 0.3375$
	$\mu_2 = 2.989$ $k_2 = 23.797$ $\omega_2 = 0.422$	$\mu_2 = 0.067$ $k_2 = 30.595$ $\omega_2 = 0.5311$	$\mu_2 = 0.0686$ $k_2 = 28.647$ $\omega_2 = 0.5516$
		$\mu_3 = -0.028$ $k_3 = 0.956$ $\omega_3 = 0.1247$	$\mu_3 = 1.2853$ $k_3 = 0.993$ $\omega_3 = 0.0699$
			$\mu_4 = -0.861$ $k_4 = 3.427$ $\omega_4 = 0.041$

Table 1: mVM parameters

where SS_{res} is the residual sum of squares:

$$SS_{res} = \sum_i (y_i - f_i)^2 \tag{38}$$

y_i are the values of the existing data set; f_i are the fitted values; SS_{tot} is the total sum of squares (proportional to the variance of the data):

$$SS_{tot} = \sum_i (y_i - \bar{y})^2 \tag{39}$$

and $\bar{y}$ is the mean of the observed data.
In the best case, the modelled values exactly match the observed values, resulting in $SS_{res} = 0$ and $R^2 = 1$. A baseline model, which always predicts $\bar{y}$, will have $R^2 = 0$. Models that have worse predictions than this baseline will have a negative R^2.

We want to explore the relationship between the number N of VM distributions that produce the mixture pdf, to the goodness of fit. Hence, for each number N we calculate R^2, as shown in Fig. 16. There is an immediate improvement when the used pdf comprises a mixture of multiple VM distributions, and there is also an apparent relationship of a bigger R^2 as the number of components N increases. The possibility to use a mixture of VM distributions is the key to the high accuracy of this model, enabling it to represent complex wind direction behaviors that may differ from one another in season, location, and more. However, in this particular example, R^2 does no change dramatically for $N > 2$, since there were mostly north and south winds - two prime modes. Therefore, two components are sufficient for a very high accuracy of modelling: for N=2 $R^2 = 0.987$. Additionally, we can have a shorter running time of the fit algorithm, in comparison to models with a larger number of components.

4 Modelling Wind Field in Space Using Radial Basis Functions

So far, the model was built according to data gathered in a constant point in space. The model can describe wind distribution in a specific location for an hourly time scale, however it does not give us any information about the wind behavior in space.
Obviously, wind changes not only with time but also in space. Wind can change in different geographic areas due to local topography, the presence of mountains, valleys, coastlines and landscapes. The weather also causes local differences. For example, temperature gradient causes pressure gradients since warm air is less dense, while cold air is denser, and air flows from areas of high pressure to areas of low pressure. Hence, the wind vector may change dramatically over the area. For instance, in an airport, wind can be southern at a certain point, but northern a hundred meters away from it. We can simulate wind in space using radial basis functions and their interpolation.

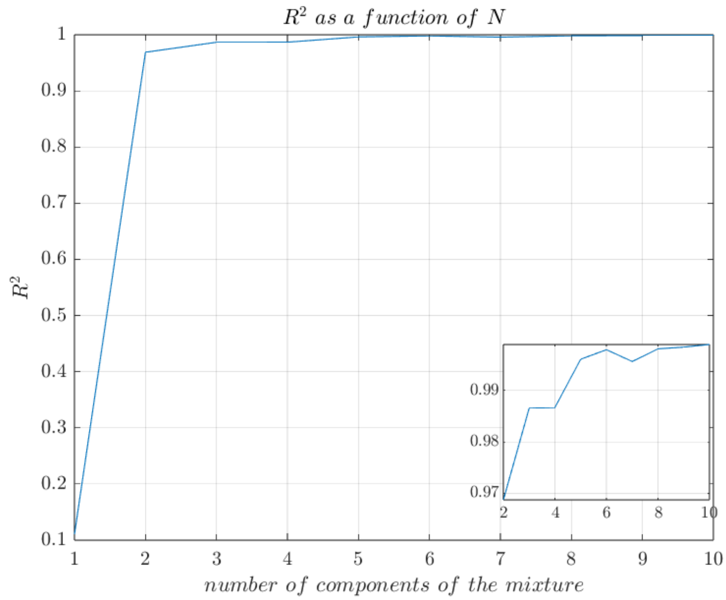

Figure 16: Coefficient of determination as a function of the number of components of the mixture distribution

Radial basis function is a real-valued function whose value depends only on the distance between the input and some fixed point c:

$$\phi(x) = \phi(|x - c|) \tag{40}$$

We will use the radial function to produce 'weights' for all squares in a grid of a size that can be determined by the user. Suppose we generate wind speed and direction for every square in the grid, but we are interested in creating a smooth wind field. To achieve this, we can compute the wind speed and direction as a superposition of known values that will be defined in specific squares, according to the weights computed by the radial functions.
The radial function we will use to set the weights is:

$$\phi(r) = e^{-\tau \cdot \left[(x - x_{fixed})^2 + (y - y_{fixed})^2\right]} \tag{41}$$

where x_{fixed}, y_{fixed} are the values defining the square where the weight equals 1 (where a specific wind speed and direction are inputted by the user); τ is the constant of the decaying exponent and also can be chosen by the user. The bigger it is, the less the assigned value effects the other squares. It is possible to assign few different models - each for one square in the grid, and the rest of the squares will have a superposition of the neighboring models according to the radial weights assigned for the squares. The number of models can be a parameter of the user's choice as well. An example is presented in Fig. 17: this wind field, is defined by three wind vectors - speed and direction (V, θ):

- square 1,1: $V = 20[\frac{m}{}s]$, $\theta = 0°$
- square 1,20: $V = 10[\frac{m}{}s]$, $\theta = -45°$
- square 20,20: $V = 15[\frac{m}{}s]$, $\theta = -120°$

and the radial function for each of them is:

- square 1,1: $e^{-0.1 \cdot \left[(x-1)^2 + (y-1)^2\right]}$
- square 1,20: $e^{-0.03 \cdot \left[(x-1)^2 + (y-20)^2\right]}$
- square 20,20: $e^{-0.05 \cdot \left[(x-20)^2 + (y-20)^2\right]}$

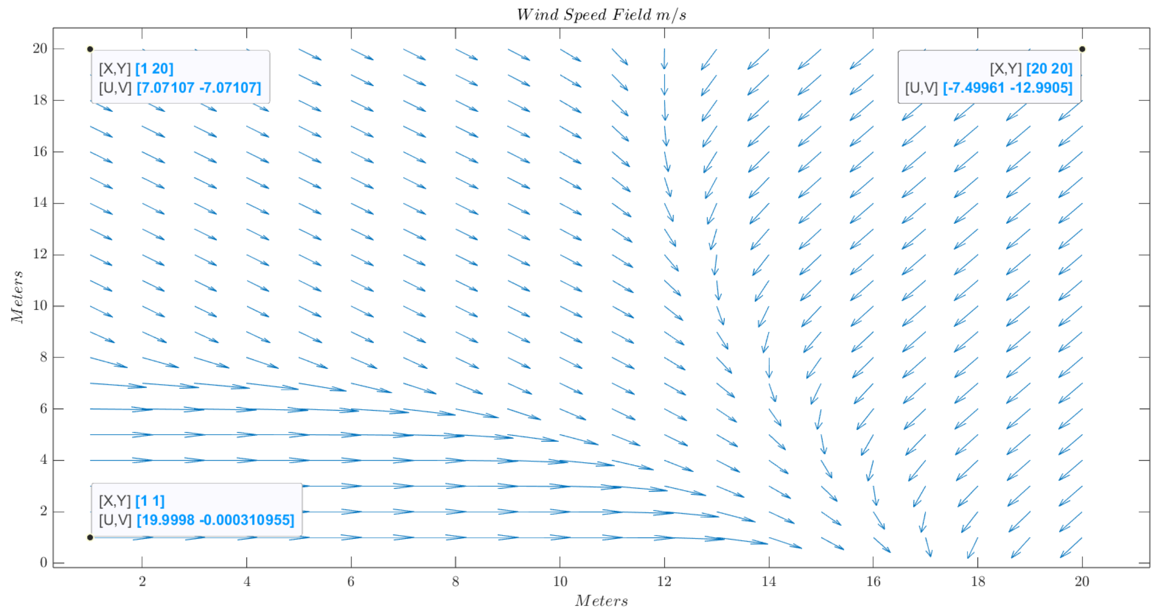

Figure 17: Wind field based on interpolation between three different predefined wind vectors

As we can see in the plot, the values computed in each of these three squares are not perfectly what was defined, due to minor effects from the other two radial functions. This effect comes from the normalization that was applied on the weights:

$$\begin{aligned} V_{square_{xy}} &= \frac{w_1}{w_1+w_2+w_3} \cdot V_{model1} + \frac{w_2}{w_1+w_2+w_3} \cdot V_{model2} + \frac{w_3}{w_1+w_2+w_3} \cdot V_{model3} \\ \theta_{square_{xy}} &= \frac{w_1}{w_1+w_2+w_3} \cdot \theta_{model1} + \frac{w_2}{w_1+w_2+w_3} \cdot \theta_{model2} + \frac{w_3}{w_1+w_2+w_3} \cdot \theta_{model3} \end{aligned} \tag{42}$$

where w_i is the weight assigned for wind model i. Important to note that for the interpolation of the angles to be correct, it is required to convert the angle to radians, between $[-\pi, \pi]$. We can increase the τ parameter and that will decrease even more the small effect that exists between the 3 squares. For instance, we can set the weights in square $[1, 1]$ to be: $w_1 = 1,\ w_2, w_3 << w_1$, and by that receive: $V_{[1,1]} \approx V_{model1},\quad \theta_{[1,1]} \approx \theta_{model1}$.
Additionally, the assigned values, may be generated from stochastic models, while controlling the parameters of their distributions. For example, the wind in square [1,20] can be generated from a speed model with a Weibull distributed speed and a direction model of northern-western winds (which mVM distribution can generate). The wind in square [1,1] can be generated from a model of a Rayleigh distribution and western winds, and the wind in square [20,20] can be generated from another Weibull distribution, with different parameters, and a direction model with east as the prime mode. This method, allows us to create a variety of wind pictures with just a few wind models, using a simple interpolation between them. Important to note that wind fields are dynamic in time, and it can be expressed in our method by the fact the the assigned values are originated from statistical models that each time will produce different values.

5 Conclusions

In this project two aspects of wind modelling were investigated: wind speed magnitude and wind direction. We analyzed a speed model based on the OU process, and then performed a memoryless transformation to a Weibull distribution in order to receive a stochastic process that represents the data well. This approach reproduces both the probability distribution and the exponentially decaying autocorrelation coefficient of the wind speed data. However, for $\tau > 50$ hours the exponentially decaying autocorrelation coefficient did not represent the real data due to diurnal and seasonal effects, that

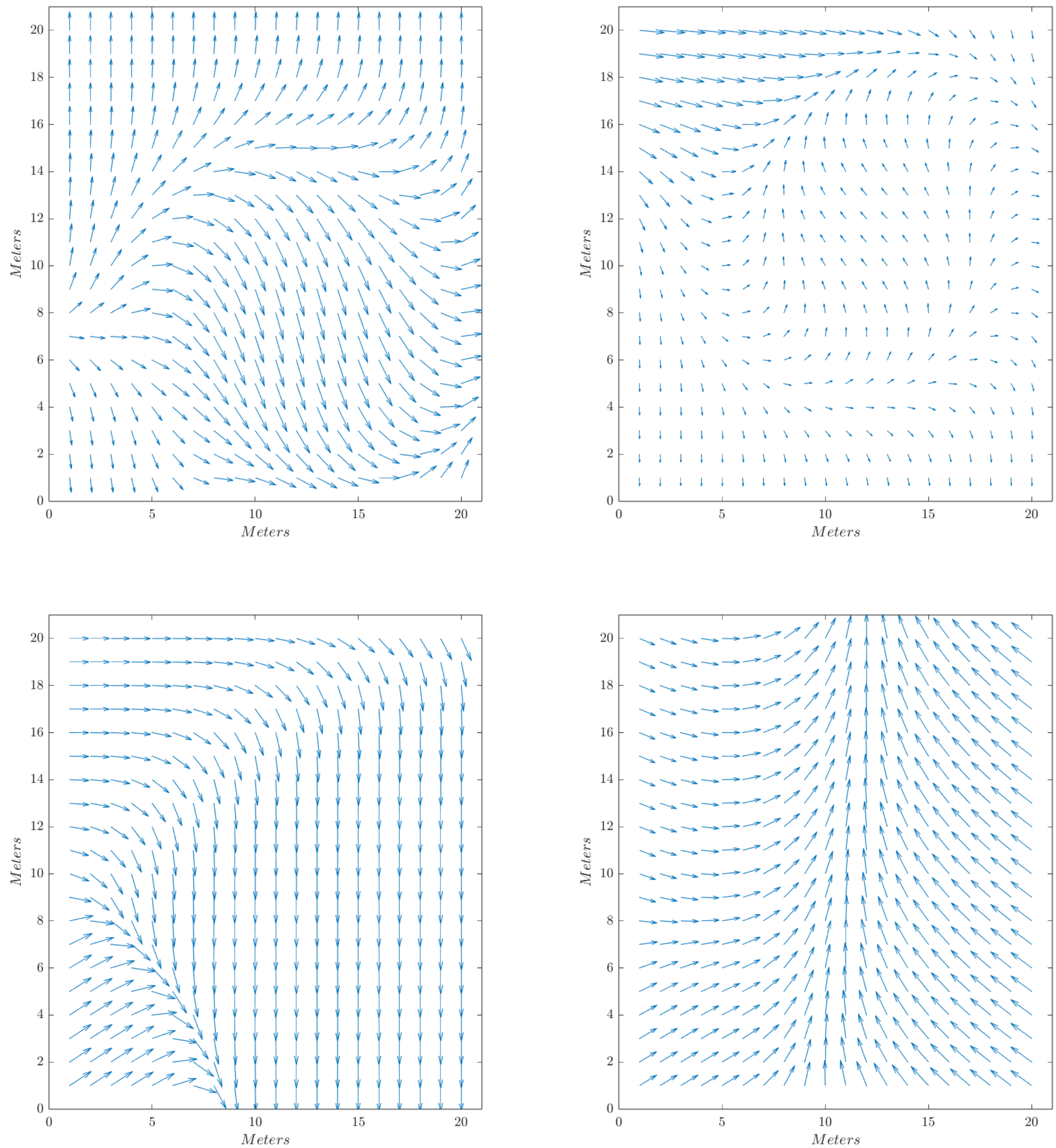

Figure 18: Four different wind behaviors in space based on radial basis functions

are not considered in the statistical model. Thus, the goodness of this model is limited to this time range.
The directional model had showed that a Von Mises mixture distribution with two components is adequate for describing the directional data. This conclusion was also supported by R^2 coefficient analysis, as it was specifically concluded that increasing the number of components N of the mixture, causes the value of the coefficient of determination R^2 to increase, meaning the fit is improved. However, the variations in R^2 are not pronounced for values of N greater than two. Next, we suggested a method to simulate wind in space using an interpolation based on radial basis functions. The method

considers known wind models (for both magnitude and direction) for specific points in space, and by interpolation between them we can receive a solution of the wind in other points in space. This method relies on the assumption that wind is continuous in space.

6 Future Work

Future work will focus on expanding the wind model and explore innovative models to complex phenomena. Overall the model will have a superposition of a few layers:

1. Long time scales model for wind speed and direction (one value per hour). The time frame we are interested in can be 24 hours: a day of operation of the airfield. For realistic results, a vertical component of the wind must be modelled as well (change of the wind with altitude).

2. Short time scale models for wind speed and direction (one value per second). This will be important during final landing stage of an aircraft/parachute, which may last 10-20 sec. May be modeled by linear SDEs.

3. Wind gusts (speed and direction), modeled by non-linear SDEs (one or maybe even more values per second). Can be modelled by one of the langevin-type model of turbulence - the pitchfork bifurcation normal form. This form of SDE can represent gusts that randomly jump between two solutions.

4. More complex wind effects, such as dust devils, which can be modeled by non-linear SDEs and a specific structure in space, (which can be predefined, such as a cone for dust devils, but have stochastic parameters of cone height and radius).

The analysis of wind will be deepened as well, through investigation of Navier Stokes equations. These can be helpful for thoroughly understanding the air flow in space and time.

The model investigated here can serve as a building block for the future extended model. It has a User Interface, which is coded in Matlab, App Designer, allows the user to tune the parameters and observe resulting wind behaviors, and can be developed further in the future.
Matlab code and Matlab app in GUI are availave in:
https://figshare.com/articles/software/research_project_code_zip/24559900

References

[1] New Zealand's National Climate Database https://cliflo.niwa.co.nz/.

[2] J. Pablo Arenas-López and Mohamed Badaoui. A Fokker–Planck equation based approach for modelling wind speed and its power output. *Energy Conversion and Management*, 222, 2020.

[3] J. Pablo Arenas-López and Mohamed Badaoui. The Ornstein-Uhlenbeck process for estimating wind power under a memoryless transformation. *Energy*, 213, 2020.

[4] D. J. Best and N. I. Fisher. The bias of the maximum likelihood estimators of the von mises-fisher concentration parameters. *Communications in Statistics - Simulation and Computation*, 10(5), 1981.

[5] A. C. Brett and S. E. Tuller. The autocorrelation of hourly wind speed observations. *Journal of Applied Meteorology*, 30(6), 1991.

[6] Tony Burton, Nick Jenkins, David Sharpe, and Ervin Bossanyi. *Wind Energy Handbook, Second Edition*. 2011.

[7] Rudy Calif. PDF models and synthetic model for the wind speed fluctuations based on the resolution of Langevin equation. *Applied Energy*, 99, 2012.

[8] 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), 2008.

[9] José A. Carta, Penélope Ramírez, and Celia Bueno. A joint probability density function of wind speed and direction for wind energy analysis. *Energy Conversion and Management*, 49(6), 2008.

[10] Christopher Schymura. Von Mises Mixture Matlab Package https://github.com/chrschy/mvmdist, 2016.

[11] Crispin Gardiner. *Stochastic Methods, A Handbook for the Natural and Social Sciences*. Dunedin, New Zealand, fourth edition edition, 2008.

[12] Daniel Pereira. Wind Rose Matlab Function https://dpereira.asempyme.com/windrose/, 2023.

[13] Wen Liang Hung, Shou Jen Chang-Chien, and Miin Shen Yang. Self-updating clustering algorithm for estimating the parameters in mixtures of von Mises distributions. *Journal of Applied Statistics*, 39(10), 2012.

[14] Richard A. Johnson and Thomas E. Wehrly. Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, 73(363), 1978.

[15] K. V. Mardia. Linear circular correlation coefficients and rhythmometry. *Biometrika*, 63(2), 1976.

[16] N. Masseran, A. M. Razali, K. Ibrahim, and M. T. Latif. Fitting a mixture of von Mises distributions in order to model data on wind direction in Peninsular Malaysia. *Energy Conversion and Management*, 72, 2013.

[17] B. McWilliams, M. M. Newman, and D. Sprevak. The probability distribution of wind velocity and direction. *WIND ENGNG.*, 3(4 , 1979), 1979.

[18] B. McWilliams and D. Sprevak. ESTIMATION OF THE PARAMETERS OF THE DISTRIBUTION OF WIND SPEED AND DIRECTION. *Wind Engineering*, 4(4), 1980.

[19] A Sarkar, S Singh, and D Mitra. Wind climate modeling using Weibull and extreme value distribution. *International Journal of Engineering, Science and Technology*, 3(5), 2011.

[20] R. Weber. Estimator for the standard deviation of wind direction based on moments of the Cartesian components. *Journal of Applied Meteorology*, 30(9), 1991.

[21] Rafael Zárate-Miñano, Marian Anghel, and Federico Milano. Continuous wind speed models based on stochastic differential equations. *Applied Energy*, 104, 2013.

[22] Rafael Zárate-Miñano and Federico Milano. Construction of SDE-based wind speed models with exponentially decaying autocorrelation. *Renewable Energy*, 94, 2016.