Modelling axial and circular data for vegetation stripes at Marion Island

by

Francois Victor von Holtzhausen

Supervisors: Dr. Priyanka Nagar, Prof. Andriette Bekker

Submitted in partial fulfillment of the requirements for the degree

MSc. Advanced Data Analytics

in the Department of Statistics

University of Pretoria

February 7, 2025

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Abstract

In this study, a bivariate model is proposed to analyse the joint distribution of Marion Island's vegetation stripe and wind direction data. The objective is to investigate whether wind contributes to the formation of these irregular vegetation stripe patterns. Using a copula-based approach, the joint density function is modelled with a bivariate wrapped Cauchy circular component combined with various circular and axial distributions. Due to multimodality in the data, a finite mixture model is proposed to accurately model the overall density. This finite mixture model incorporates the slope angle and cone aspect as concomitant variables. The results indicate that a three latent component finite mixture model with von Mises and axial normal distributions as marginals provides the best fit. Using the proposed model it was determined that wind influences vegetation stripe orientation on the southern sides of cones, while no clear relationship is observed on the northern sides, likely due to harsher wind and sunlight exposure. These findings highlight the role of wind and other environmental factors, such as cone aspect and slope, in shaping vegetation patterns.

Keywords: axial data, circular data, copula, expectation-maximisation algorithm, mixture model, vegetation stripes

Declaration of Authorship

I, Francois Victor von Holtzhausen, declare that the dissertation, which I hereby submit for the degree *MSc. Advanced Data Analytics* at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signed: Francois Victor von Holtzhausen

Date: 2024/11/26

Acknowledgements

I would like to first thank my supervisors Dr. Nagar and Prof. Bekker, for their invaluable guidance and mentorship throughout the various phases of this project. Their constant support, insightful feedback, and encouragement were instrumental in shaping the direction and success of this study.

Secondly, I would like to thank Prof. le Roux and Ms. Schoombie, not only for providing the data required for this study but also for their willingness to address any questions regarding Marion Island and its vegetation. Prof. le Roux's contributions were especially appreciated in understanding and interpreting the results of the proposed model.

Thirdly, I would like to thank Prof. Lagona and Dr. Mingione for their significant input in the development and derivation of the proposed model, which enhanced the study's theoretical foundation significantly.

Finally, I would like to thank my mom for her never-ending support, encouragement, motivation and love.

Regarding the data used in this paper, the following ethics number is valid: 2024-PLR1

Contents

List of Figures

List of Tables

List of Abbreviations

AXN	Axial Normal distribution
AXWC	Axial wrapped Cauchy distribution
BIC	Bayesian information criterion
cdf	Cumulative distribution function
CDL	Complete data log-likelihood
CFD	Computational fluid dynamics
CWC	Conditional wrapped Cauchy distribution
EM	Expectation-maximisation
ECDL	Expected complete data log-likelihood
E	East
glm	Generalised linear model
km	Kilometer
km^2	Square kilometer
m	Meter
m.sec^{-1}	Meters per second
MLE	Maximum likelihood estimates
mm	Millimeter
N	North
NE	Northeast
NW	Northwest
S	South
SE	Southeast
SW	Southwest
T	Transpose
VM	von Mises distribution
W	West
WC	Wrapped Cauchy distribution
°C	Degree Celsius

Chapter 1

Introduction

This chapter provides an overview of the problem statement, the aim, and the objectives of the study. Additionally, it examines relevant previous research pertaining to the specific problem under investigation and highlights other studies conducted on Marion Island over the years.

1.1 Problem statement

Linear vegetation patterns are frequently observed in climatically extreme environments such as arid, polar, alpine, and wetland environments [1–3]. The general idea behind these vegetation stripes is that the vegetation or substrate found on hills or slopes would grow directly downhill. This is because the plants found upslope would protect the neighbouring plants found downslope from burial caused by moving substrate. Furthermore, neighbouring plants also provide wind shelter for one another in the immediate environment [4]. For this reason, it is expected that vegetation stripes will always be orientated vertically (aligned with the slope of the surrounding topography) as this will provide maximum protection from burial caused by gravity-related processes. However, this is not the case, just by looking at satellite images of these various environments it can be observed that numerous vegetation stripes deviate away from the expected vertical orientation and that this deviation varies considerably [4].

Various different hypotheses were formulated and studies were conducted in order to understand the orientation of these irregular stripe and band-like vegetation patterns. In these studies, a recurring theme occurred being that wind and slope angles are seen as key contributors to the non-vertical orientation of vegetation stripes. However, there is still an incomplete understanding regarding which environmental factors influence the occurrence and nature of vegetation patterning. Since most of these studies were only conducted on a small scale or on single landforms [5, 6] or through simulation studies [7] and have not yet been quantitatively tested.

Fortunately, vegetation stripes are also present on Marion Island (specifically on its various scoria cones found all over the island) for which there are various data variables associated with both the specific vegetation stripes found all over the island as well as the various environmental factors to which the stripes may be exposed to. Furthermore, due to the strong

dominance of westerly winds in the sub-Antarctic region, Marion Island offers an excellent study system when examining the impacts of wind directions. For this reason, we can conduct a large-scale analysis by developing a statistical model to explain the deviation from the vertical orientation and finally determine whether or not wind direction (as well as various other factors) truly has an effect on the orientation of the vegetation stripes found on Marion Island.

It should however be noted that the statistical analysis of this specific data combination is quite different from the traditional analysis of bivariate continuous data. This is due to the fact that wind direction is observed as circular data, and the orientation of the vegetation stripes is axial data. The general idea behind axial data is that it is random lines that do not have a natural orientation or specific start and end points associated with them [8]. There exist numerous methods and distributions in directional statistics to model axial and circular data separately. However, to the best of our knowledge, approaches to analyse the joint distribution of mixed axial and circular data have not yet been investigated.

1.2 Aim and objective

The aim of this study is to determine whether or not wind direction (as well as various other factors) truly affects the orientation of the vegetation stripes found on Marion Island. This will be achieved by using a proposed bivariate distribution for handling mixed circular and axial data. The model will be constructed using a copula-based approach to introduce a circular copula that relies on the toroidal density created by S. Kato and A. Pewsey (2015) [9]. To deal with data heterogeneity, which results from the fact that not all vegetation stripes experience the same environmental conditions, a finite mixture of circular-axial bivariate distributions will be fitted to the data. The mixing weights for this finite mixture model will be obtained based on the available environmental covariates (concomitant variables) [10]. Using this approach will allow for the grouping of vegetation stripes that feature a similar dependence on wind direction into the same clusters while still taking into consideration the environmental factors that it experienced.

1.3 Background

Marion Island is situated just North of the Antarctic Polar Convergence in the South Indian Ocean [11] and is one of the six island archipelagos in the subantarctic [12]. The climate is maritime with temperatures being cold but constant throughout the year ranging from between 5.2°C (degree Celsius) and 10.9°C at sea level in the summer months and between 1.6°C and 6.6°C at sea level in the winter months [13], with the top of the island being 7°C colder. In terms of precipitation, it remains quite high year-round with 2000mm (millimeter) per annum [14]. Similarly, the island also experiences windy conditions throughout the entire year with an average of less than 10 windstill days and more than 100 days of extreme winds (more than 15 m.sec^{-1} (meters per second)) between 1992 and 2003 [14]. The island has a predominantly Northwesterly wind which averages at 10 m.sec^{-1} and which makes up 60% of the captured

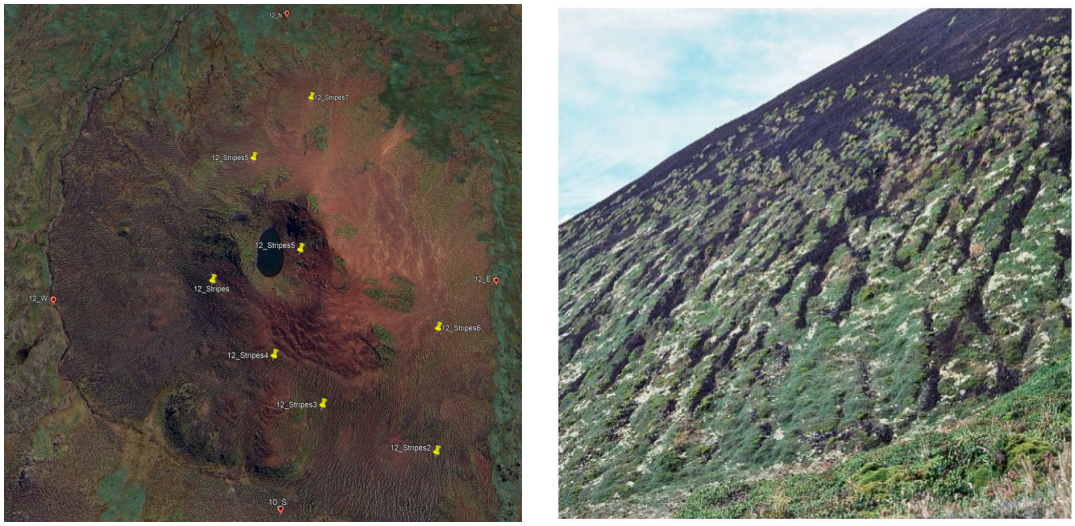
wind readings. Whereas, less than 10% of the wind reading indicates a far weaker Easterly wind which averages at less than 6 m.sec^{-1} in wind speed [14]. It should however be noted that Marion Island has experienced a rapid change in its climate, with temperatures rising at a rate double the mean global warming rate with an average annual increase of 0.04°C since 1969 [12], as well as a significant decline in captured rainfall readings since the 1960's [15]. Changes in wind conditions over the same period are less clear. However, the annual mean wind speed has increased and the mean wind direction has seen a shift to a more Northerly direction especially during the late summer and autumn seasons [15].

In terms of topography, Marion Island is 290km^2 (square kilometer) in area and rises to a height of 1230m (meter). The island consists of one big mountain (volcano) in its center which slopes down to a coastal plain. On the Northern and Southern sides of the island are a 4 to 5 km (kilometer) wide area that gently rises from sea level to the foot of the mountain at around 300m in altitude [12]. The island has 130 pyroclastic scoria cones scattered all over the island which originated from previous volcanic eruptions [4]. Scoria cones mostly comprise of loose, rough volcanic rocks. They are typically disproportionately exposed to winds due to the cones being around 50 to 200m taller than the surrounding landforms [12]. Furthermore, they are predominantly located on these wider and more wind-exposed areas on the island away from cliffs and ridges that could otherwise offer some wind shelter [12] as can be seen in Figure 1.1.

FIGURE 1.1: Aerial view of Marion Island and its scoria cones.

Marion Island houses a variety of plant species including 22 native species, 18 introduced species, and three species of unknown status [12]. The vegetation on these scoria cones is quite diverse as it ranges from sparse cryptogram cover to nearly closed vascular plant cover. However, the cone's vegetation is mostly dominated by the following four plant species: the cushion-forming Azorella selago Hook, the grass Argostis magellanica Lam, the creeping fern Blechnum penna-marina Kuhn, and finally the low-growing woody shrub Acaena magellanica [4]. The growth pattern of these plants found specifically on the various scoria cones exhibits an atypical characteristic, wherein they appear to align themselves in linear or stripe formations that correspond to their respective positions on the cones, see Figure 1.2b. As mentioned in Section 1.1, it was hypothesised that these plants grow in these specific formations due to neighbouring plants benefiting from each other by providing not only protection from burial

by scoria in their immediate environment but also protection from these extreme wind conditions [4]. Thus, it is expected that these vegetation stripes to be vertically orientated and aligned with the immediate environment's slope. However, by specifically looking at Figures 1.2a and 1.2b it can be seen that the orientation of these various stripes varies significantly from the expected vertical orientation. Furthermore, it can be seen that the orientations also vary across the different segments or aspects of the specific cone, which comprises of the northwest (NW), northeast (NE), southwest (SW) and southeast (SE) aspects. This is also the case with the other 130 scoria cones having their own unique variations in their various aspects.



(a) (b)

FIGURE 1.2: (a) Aerial view of scoria cone with visible vegetation stripe patterns and (b) frontal view of a scoria cone with visible vegetation stripe patterns. Note. Reprinted from "Vegetation of subantarctic Marion and Prince Edward Islands," by Smith and Mucina, 2006, Strelitzia, 19, 708. Copyright 2006.

Marion Island was annexed by the Union of South Africa in late 1947 from the British government which was in control of the island since 1908 [16]. This was quite a strategically important event as Marion Island is one of the most significant sites for the collection and climatological data and marine and terrestrial research in the vast oceanic region from as early as the 1940's and 1960's respectively [17]. Furthermore, Marion Island has one of the most active soil and perma frost environments in the world and plays a key role in understanding periglacial environments [17]. Shortly after gaining control of the island the South African government installed the first meteorological station on the Transvaal Cove on the North-East coast of Marion Island which was administered by the Weather Bureau division of the Department of Transport [16]. The station was manned by small parties of volunteers from various departments and was stationed six months at a time until the next relief party arrived [16]. This weather station remained in commission until the early 2010's when a new research base was built replacing all of the older decaying buildings on the site.

During the 1950's Marion Island had a permanent snowline at 600m in altitude and ice at the summit [17]. However, this has drastically receded as Marion Island experienced an average yearly temperature increase of 0.04°C between 1969 and 1999 [18]. Due to the snowline receding, researchers could finally take a detailed look at the surface of higher altitude areas on the Island. For this reason, the sorted stripes widely found on Marion Island were first described by B. Van Zinderen (1975) [19]. Sorted stripes were originally described as patterned ground consisting of alternating lines of coarse and fine stones orientated down the nearest steepest slope [20]. Although vegetation stripes were observed and mentioned during these initial studies the research conducted on Marion Island regarding the variation of the orientation of linear features in the sub-Antarctic was mostly focused on the sorted stripes.

As research developed through the years specifically regarding patterned ground it was observed that sorted stripes as well as vegetation stripes surrounding it were not necessarily orientated down the steepest slope but rather across. This occurred at numerous study sites around the Southern Hemisphere such as Marion Island [21], Maquarie Island [22], South Georgia Island [23], South Shetland Islands [24] and Kerguelen Island [25] just to name a few. Multiple hypotheses were developed regarding this strange phenomenon. Some speculated that erosion due to slopewash [26] or snow cover [24] was a core reason behind these sorted stripes [26]. While others believed that sorted stripes and sediment movement occurred due to frost creep associated with needle ice activity resulting from the seasonal freezing and thawing occurring in the landscapes [11, 27, 28]. Another hypothesis was that the sorted stripes originated due to aeolian processes in the various areas [13, 21, 29, 30].

The hypothesis that wind was one of the key reasons for the existence of the sorted stripes and vegetation stripes found in the areas was first proposed by K. Hall (1979) [21]. Hall measured the orientation of sorted stripes at 12 specific locations and determined that there was a clear correlation between the orientation of the stripes and wind direction as the sorted stripes aligned predominantly parallel to the wind direction [13, 21]. A study by S.D. Holness (2001) [13] further reinforced this hypothesis of wind being the primary factor for the specific orientation of these sorted stripes. In this paper, Holness conducted a more thorough investigation by systematically searching for all sorted stripes at the specific 12 locations and determining their orientation [13]. Holness also incorporated the varying orientations of the slopes where these sorted stripes were found as part of the study. It was concluded that sorted stripes consistently aligned with the maximum local gradient and were predominantly found on slopes with a windward aspect, although they also occasionally appear on low-angle leeward slopes [13]. Furthermore, it was found that in the presence of cross-winds, there was quite a significant absence of sorted stripes as they were quite infrequent [13]. Finally, he concluded that when sorted stripes were found on nearly horizontal surfaces they would align parallel to the dominant wind direction.

The literature consists of quite a few contradicting hypotheses regarding the origins of these sorted stripes. Even though the paper by S.D. Holness (2001) [13] is quite conclusive there are

just as strong arguments and evidence for the other hypotheses. Furthermore, one drawback of the study conducted by Holness which he noted himself was that no theoretical statistical models could be developed due to the nonexistence of specific surface-level wind data at the locations of the sorted stripes. This was because all of the available wind data at the time was just from the weather station which was located on the Northeast coast of the island and for this reason they had to assume that this was representative of the entire island's wind behaviour.

Fortunately, in a recent study conducted by K.A Goddard, K.J Craig, et al. (2022) [31], a computational fluid dynamics (CFD) model was developed to generate reliable wind pattern data for Marion Island at surface level (30m resolution). With this CFD model, local variation in wind speed and direction can be accounted for due to the island's heterogeneous topography.

Instead of sorted stripes, the focus of this study will be on the vegetation stripes found on Marion Island and specifically on how the wind as well as other factors such as cone aspect and slope angle influences the orientation of these stripes. It is important to understand the growth patterns of these stripes. Not only do these vegetation stripes affect ecosystem functionality in climatically extreme environments (altering sediment and water movement) [6, 32], but they are quite sensitive to changing environmental conditions and can possibly be used as early warning mechanisms for environmental tipping points and exosystem collapse [33]. However, the influence of wind conditions on vegetation stripe patterns in the sub-Antarctic has not been explored prior to this research.

In Table A.1, A.2 and A.3 found in Appendix A, a brief review of recent relevant literature is presented specifically on the wind of Marion Island as well as the effect of wind on plants in some other areas of the world. The table contains a short summary of the goal of each of the studies, what methodologies were followed during the specific studies, and the conclusion of each of the studies. The papers are listed in chronological order and start with P.C. Le Roux and M.A. McGeoch (2008) [4] and end with M. Momberg et al. (2023) [34].

1.4 Outline of study

This study is organised as follows: Chapter 2 provides a detailed description of how the data for vegetation stripes and wind direction were collected. It also includes an exploration of the data to gain a comprehensive understanding of the various variables. Chapter 3 presents background information on directional statistics, including the distinction between circular and axial variables and the marginal distributions applicable to each. This chapter also introduces copula functions and their application in modelling a circular-axial joint distribution. Chapter 4 outlines the methodology for constructing the proposed finite mixture model, which applies the circular-axial joint density to the multimodal dataset. It details the parameter estimation process using the expectation-maximisation (EM) algorithm and discusses the goodness of fit measures used to identify the best-fitting model. Chapter 5 discusses the simulation model employed to evaluate the robustness of the proposed model under different scenarios. The chapter

also presents the results obtained from applying the model to the Marion Island dataset, followed by an interpretation of these findings. Finally, Chapter 6 concludes the study with a discussion of the key conclusions and suggestions for future research.

Chapter 2

Data Description and Exploration

This chapter provides a detailed account of the processes used to obtain the wind direction data and the vegetation stripe data. Additionally, it includes a thorough exploration of the dataset to ensure a clear understanding of the variables that will be used in this study.

2.1 Data description

Throughout this study, two datasets from Marion Island are utilised. The first is a wind dataset derived from a CFD model, and the second is a vegetation stripe dataset, with observations obtained from a 1m resolution digital elevation model. In conjunction with the vegetation stripe dataset, this study also incorporates a Google Earth document containing numerous markers (refer to Figure 1.2a) that indicate the locations of relevant scoria cones and 150 specific vegetation stripes of interest. This Google Earth document, along with the elevation model readings, was previously obtained and utilised in a study by M. Momberg, D.W. Hedding, et al. (2021) [35], which investigated wind stress as a driver of fine-scale variation in plant communities on Marion Island.

The wind dataset contains simulated data originating from a CFD model [31]. The CFD model uses the full digital surface model of Marion Island and simulates airflow over the topography by iteratively solving partial differential equations [34]. The model was built using wind data captured by sonic anemometers at 17 different locations spread out over the island, as can be seen in Figure 2.1, from April 2018 until March 2020. Note however that this model predicts the wind speed and direction at numerous systematic grid points on the island. Resulting in each cone having numerous observations. However, only data points closest to a specific vegetation stripe's coordinate were extracted since extracting all of the data points in an aspect will introduce a lot of real-world heterogeneity that the vegetation stripe most likely does not experience. In Figure 2.2 it can be observed that the actual location of the vegetation stripe is marked with a red dot, whereas the closest data point from the simulated dataset is marked with a blue dot.

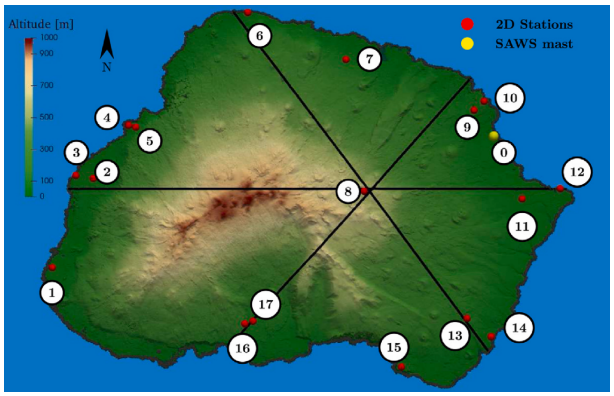


FIGURE 2.1: Sonic anemometers locations over Marion Island. Note: Reprinted from "Investigation of ecologically relevant wind patterns on Marion Island using Computational Fluid Dynamics and measured data," by Goddard *et al.*, 2022, Ecological Modelling, 464, 109827. Copyright 2022.

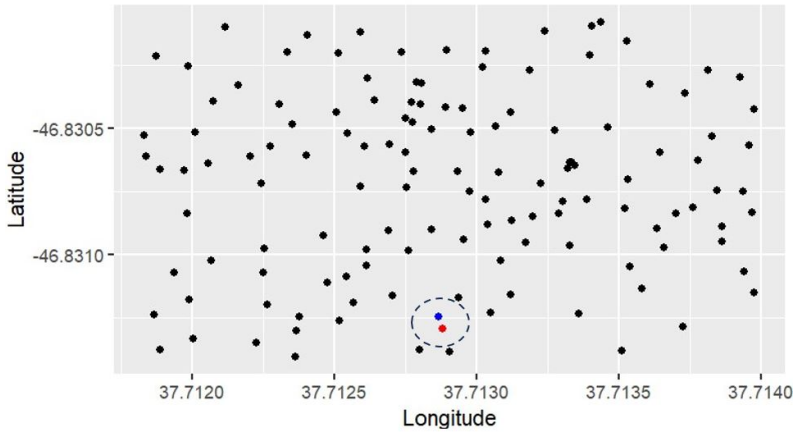


FIGURE 2.2: Plot of simulated computational fluid dynamics model data points with the red dot representing the actual location of the vegetation stripe and the blue dot representing the closest simulated wind observation point.

In Figure 2.2, it is evident that the distance between the actual location of the vegetation stripe and the nearest simulated data point is remarkably close. This suggests that there should be no difficulty in accurately determining the direction of the wind affecting the particular vegetation stripe. However, there are a few instances where the distances between the actual and simulated points are not as close and where the topography between the two data points drastically changes. This results in the closest simulated wind direction not accurately representing the winds that the vegetation stripe might have encountered. In cases like these, the vegetation stripes as observations were excluded.

For this study, vegetation stripe observations were specifically extracted from 35 of the 130 scoria cones identified on Marion Island. These 35 cones were split into four aspects (NW, NE, SE, SW) as indicated in Figure 2.3. This approach was followed such that it could be determined if the vegetation stripes' orientation found in different areas of the cones was influenced differently by the various external factors. This specific subset of the island's scoria cones was selected based on the following three rules:

1. They exhibited vegetation stripes in at least three out of the four aspects of the cone (NE, NW, SE, SW).
2. Had a somewhat regular cone shape and no drastic changes in terms of topography.
3. Had a relatively consistent slope within each aspect.

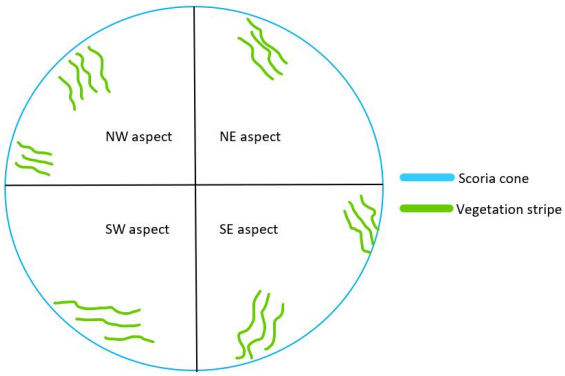


FIGURE 2.3: Illustration of how cone aspects were obtained.

The pre-specified vegetation stripes' orientation was extracted using Google Earth imagery (which was last updated 2024/02/26) and its bag of various tools in order to calculate the circular average of five distinct vegetation stripes (measured from South to North) surrounding a specific stripe marker as indicated by Figure 2.3 (where fewer stripes were present, the circular mean orientation of all the stripes was used). However, with some of the vegetation stripes, it was not possible to accurately obtain the orientation of the vegetation stripes as the stripes were not distinct enough. Again, the observations in cases like these were excluded .

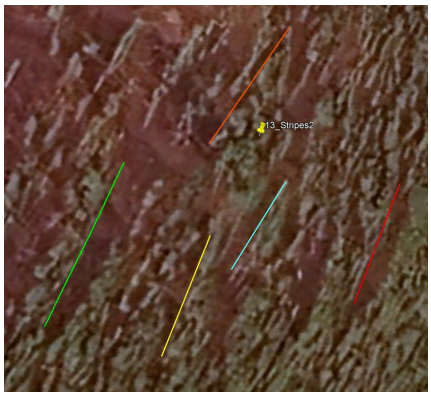


FIGURE 2.4: Illustration of how vegetation stripe direction was obtained from Google Earth Imagery.

Due to both the above-mentioned issues, as well as our three-rule system in terms of the selection of applicable scoria cones, only 133 of the total 150 vegetation stripe observations in the dataset were utilised for this study. Furthermore, from the two datasets, the following additional environmental information was available.

TABLE 2.1: Wind data (obtained from computational fluid dynamics model)

Variable	**Units**	**Description**
Longitude	decimal degrees	
Latitude	decimal degrees	
u_x	m.sec^{-1}	Component of the wind velocity in the x direction (E-W, x>0 means wind blowing from West to East).
u_y	m.sec^{-1}	Component of the wind velocity in the y direction (i.e. N-S, y>0 means wind component flowing from South to North).
u_z	m.sec^{-1}	Component of the wind velocity in the z direction, w (i.e. up or down).
U-2D	m.sec^{-1}	Two dimensional horizontal wind speed U-2D=$\sqrt{u_x^2 + {u_y}^2}$.
U-3D	m.sec^{-1}	Three dimensional horizontal wind speed U-3D=$\sqrt{u_x^2 + u_y^2 + u_z^2}$.
θ	degrees	Two-dimensional wind direction (obtained by using the Equation: 2.1).

The output of the CFD model was only a three-dimensional wind velocity vector. Therefore, Equation 2.1 was used to determine the two-dimensional horizontal wind direction (θ):

$$\theta = (270^\circ - atan2(u_y, u_x) \text{ x } \frac{180}{\pi}) \bmod 360^\circ, \tag{2.1}$$

where

$$atan2(u_y, u_x) = \begin{cases} tan^{-1}(\frac{u_y}{u_x}) & \text{, if } u_x > 0 \\ tan^{-1}(\frac{u_y}{u_x}) + \pi & \text{, if } u_x < 0 \text{ and } u_y \geq 0 \\ tan^{-1}(\frac{u_y}{u_x}) - \pi & \text{, if } u_x < 0 \text{ and } u_y < 0 \\ \frac{\pi}{2} & \text{, if } u_x = 0 \text{ and } u_y > 0 \\ -\frac{\pi}{2} & \text{, if } u_x = 0 \text{ and } u_y < 0 \\ undefined & \text{, if } u_x = 0 \text{ and } u_y = 0. \end{cases}$$

TABLE 2.2: Stripe data (obtained from digital elevation model and Google Earth document).

Variable	**Units**	**Description**
Name		Sample name, with north (N), east (E), west (W) and south (S) indicating the scoria cone aspect.
Latitude of marker	Decimal degrees	
Longitude of marker	Decimal degrees	
Slope angle	Degrees	Extracted from Marion Island digital elevation model.
Aspect		Slope aspect; extracted from Google Earth document.
Altitude	meters above sea level	Extracted from Marion Island digital elevation model.
Stripe direction	Degrees	Obtained by calculating the circular mean of 5 distinct stripes surrounding specific stripe marker.

2.2 Exploratory analysis

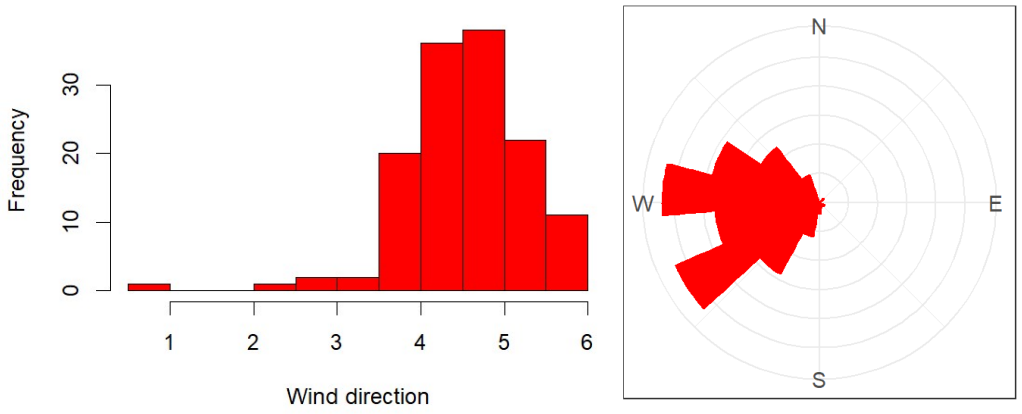
The next section examines the descriptive statistics for both the wind and stripe directions. The shape features of these variables will be explored using histograms and rose diagrams. Additionally, the correlation between the two directions will be analysed through a scatter plot and the calculation of a correlation coefficient.

2.2.1 Wind direction

The wind direction variable is a circular variable with a domain between 0 and 2π, and it has the following descriptive statistics and distribution.

TABLE 2.3: Descriptive statistics of wind direction variable.

Mean direction	Median direction	Mean resultant length	Circular variance	Circular standard deviation
1.413483	1.432109	0.7955687	0.204431	0.67631



(a) Histogram of wind direction data. (b) Rose diagram of wind direction data.

FIGURE 2.5: Distribution of wind direction variable.

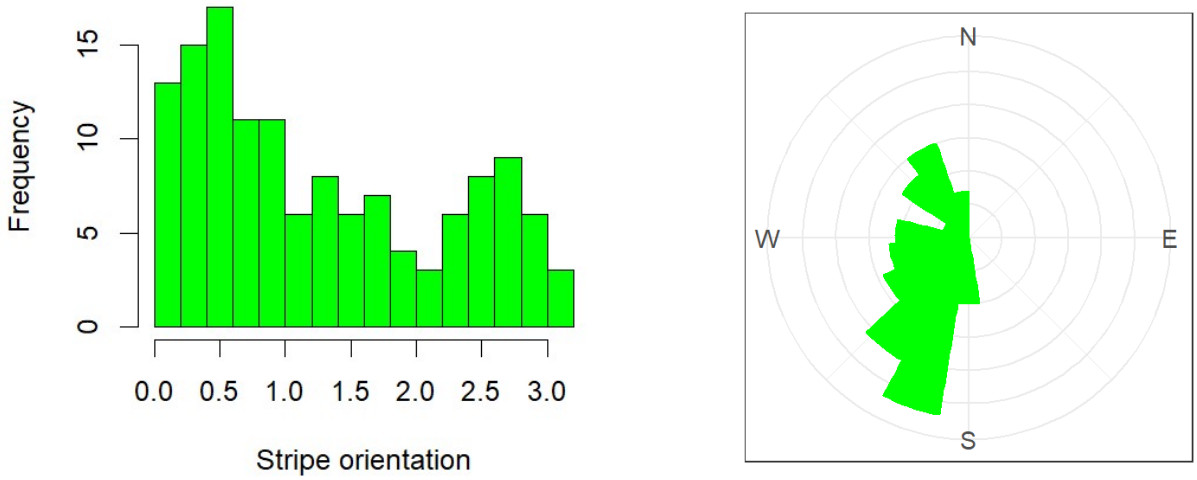
From Figure 2.5, it is observed that the wind mostly comes from the westerly direction (West to East). This is in line with what is expected since from historical weather data it is known that the three dominant wind directions on Marion Island are Westerly, North-Westerly, and South-Westerly [31]. Furthermore, it can also be seen that the wind direction data is quite concentrated. This is further confirmed in Table 2.3 from the mean resultant length, (which is measured between 0 and 1 with values closer to 1 indicating higher concentration and values closer to zero indicating higher dispersion) and circular variance (which is also measured between 0 and 1 with values closer to 1 indicating higher dispersion and values closer to zero indicating higher concentration).

2.2.2 Stripe orientation

The stripe orientation variable is an axial variable with a domain between 0 and π, and has the following descriptive statistics and shape features.

TABLE 2.4: Descriptive statistics of vegetation stripe variable.

Mean direction	Median direction	Mean resultant length	Circular variance	Circular standard deviation
1.182635	0.9965248	0.6260779	0.3739221	0.9677609

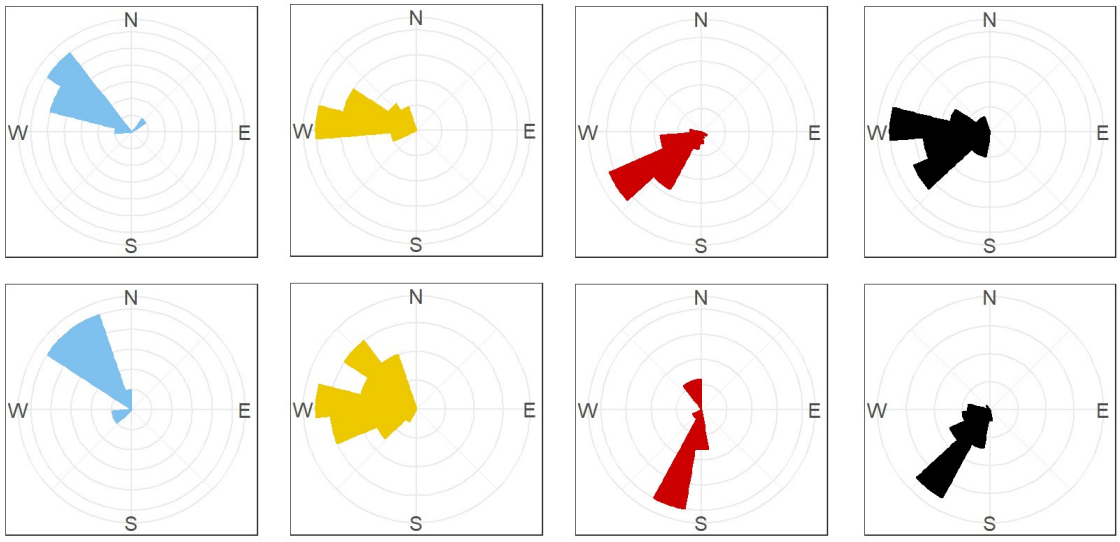


(a) Histogram of stripe orientation data.

(b) Rose diagram of stripe orientation data.

FIGURE 2.6: Distribution of vegetation stripe variable.

From Table 2.4 and Figure 2.6, it is observed that the stripe orientation has a multimodal distribution and that the vegetation mostly grows in the South-Westerly / North-Easterly direction. Furthermore, by looking at the somewhat high mean resultant length and low circular variance it can be seen that the observations for stripe directions are quite concentrated and not dispersed.



(a) NE (b) NW (c) SE (d) SW

FIGURE 2.7: Rose diagrams for wind direction (top) and stripe orientation (bottom) by aspect (NW, NE, SW, SE).

2.2.3 Joint distribution of stripe orientation and wind direction

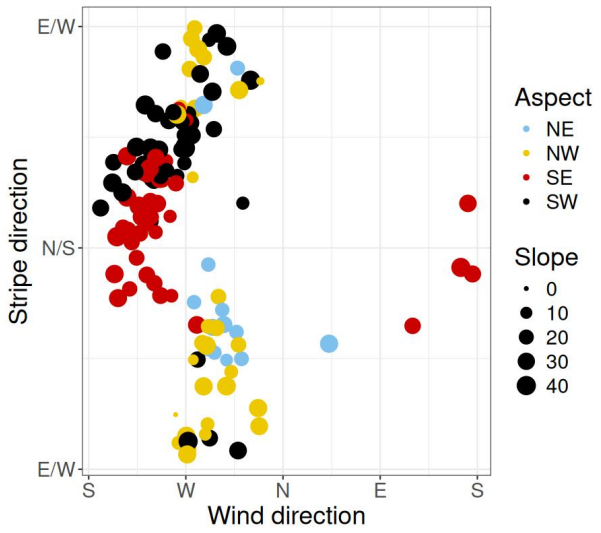


FIGURE 2.8: Scatter plot of the joint density of the vegetation stripe orientation and wind direction variables. Data points are coloured according to the cone's aspect, while their size is proportional to the associated cone's slope.

The joint distribution of wind and stripe directions is displayed in Figure 2.8. It is noted that the wind direction and stripe orientation variables do not seem to be correlated. By calculating the circular version [36] of Pearson's product-moment correlation we get $\rho = -0.1452$. This indicates that the variables are slightly negatively correlated. However, by specifically colouring the data points according to the cone's aspect, it is observed that there is clear evidence of separated groups and clusters. Figures 2.7 and 2.8 indicate that locations with South-Westerly winds and stripes pointing toward the North/South or North-East/South-West directions are mostly found in the Southern part of the cone. Where on the other hand, locations in the Northern part of the cone usually experience North-Westerly winds, with stripe directions more aligned horizontally to the cone shape. By using a mixture of axial and circular distributions that also consider and incorporate additional concomitant variables, we can not only group the data based on different patterns of correlation but also evaluate how the aspect and slope affect cluster membership.

Chapter 3

Statistical Background

The aim of this study is to develop a mixture of bivariate distributions, which will be obtained by binding a univariate circular distribution and a univariate axial distribution using a copula density. This chapter explores potential distributions that can be considered for axial and circular marginals. Additionally, it examines the specific copula chosen to construct the joint density.

3.1 Marginal distributions for the wind direction variable

Directional random variables are discussed extensively in [36–38]. Assume a direction observed in a plane, such as the wind direction variable, is represented as a circular random variable X with values on the unit circle $[0, 2\pi)$. The density function is given by $f_{\text{circ}}(x)$. For $f_{\text{circ}}(x)$ to be a valid circular density it needs to meet the following three requirements [36]:

- $f_{\text{circ}}(x) \geq 0$ almost everywhere on $(-\infty, \infty)$,
- $f_{\text{circ}}(x + 2\pi) = f_{\text{circ}}(x)$ almost everywhere on $(-\infty, \infty)$,
- $\int_0^{2\pi} f_{\text{circ}}(x)dx = 1$.

From the above requirements, it can be observed that $f_{\text{circ}}(\text{x})$ is a nonnegative function that integrates to 1 over the support $[0, 2\pi)$ and that it is also 2π periodic. Under these specific requirements, the cumulative distribution function (cdf) of the circular random variable which is expressed as [36]:

$$F_{\text{circ}}(x) = \int_0^x f_{\text{circ}}(t)dt, \quad \text{where } x \in [0, 2\pi). \tag{3.1}$$

Some interesting properties relating to $F_{\text{circ}}(x)$ includes [36]:

- Similar to $f_{\text{circ}}(\text{x})$, $F_{\text{circ}}(x)$ retains periodicity ie. $F_{\text{circ}}(x + 2k\pi) = F_{\text{circ}}(x)$,
- $F_{\text{circ}}(x)$ is discontinuous at $2k\pi$ ie. $\lim_{x \to 2k\pi} F_{\text{circ}}(x) = 1 \neq F_{\text{circ}}(2k\pi) = F_{\text{circ}}(0) = 0$, for any integer k,
- $F^{-1}_{\text{circ}}(u)$ is a continuous mapping from the unit interval $u \in [0, 1)$ to the unit circle $[0, 2\pi)$.

To model the circular wind direction variable, this study considers the two most commonly used circular density functions: the von Mises (VM) distribution and the wrapped Cauchy (WC) distribution. Both of these distributions are unimodal distributions which are symmetric at the location parameter $\mu \in [0, 2\pi)$ and shaped by their concentration parameter κ which controls the concentration of the distribution around μ [36].

The von Mises distribution, $VM(\mu, \kappa)$, has the following density [36]:

$$f_{\text{VM}}(x; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(x-\mu)}, \tag{3.2}$$

where $0 \leq \mu < 2\pi$, $\kappa > 0$, I_0 is the modified Bessel function of the first kind and order 0, and is given by the following:

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} e^{\kappa \cos\theta} d\theta. \tag{3.3}$$

The von Mises distribution can be approximated by wrapping the normal distribution around the circumference of a unit circle. This is achieved by using the following [36]:

$$x = l (\text{mod} 2\pi), \tag{3.4}$$

where l represents a linear distribution (ie. the normal distribution) with a random linear observation d. Furthermore, if l has a density f_l then the corresponding density of the wrapped function x_w is [36]:

$$f_{X_w}(x_w) = \sum_{k=-\infty}^{\infty} f_l(x_w + 2\pi k). \tag{3.5}$$

Now in order to derive the density of the wrapped Cauchy distribution, $WC(\mu, \kappa)$, Equation 3.4 is used to wrap the linear Cauchy distribution's density onto the unit circle. The Cauchy distribution's density is given by the following [36]:

$$f(d; \mu, \kappa) = \frac{1}{\pi} \frac{\kappa}{\kappa^2 + (d-\mu)^2}, \quad \text{where} \quad -\infty < \mu < \infty, \quad \kappa > 0. \tag{3.6}$$

The wrapped Cauchy distribution, $WC(\mu, \kappa)$, has the following density [36]:

$$f_{\text{WC}}(x; \mu, \xi) = \frac{1}{2\pi} \frac{1-\xi^2}{1+\xi^2-2\xi\cos(x-\mu)} \quad \text{where} \quad \xi = e^{-\kappa}, \quad 0 \leq \mu < 2\pi, \quad \kappa > 0. \tag{3.7}$$

Figure 3.1 illustrates the densities of both the von Mises (black) and wrapped Cauchy (red) distributions. The plot also highlights the effects of the concentration parameter κ on both circular distributions. From this density plot, it can be observed that the wrapped Cauchy distribution is more concentrated compared to the von Mises distribution.

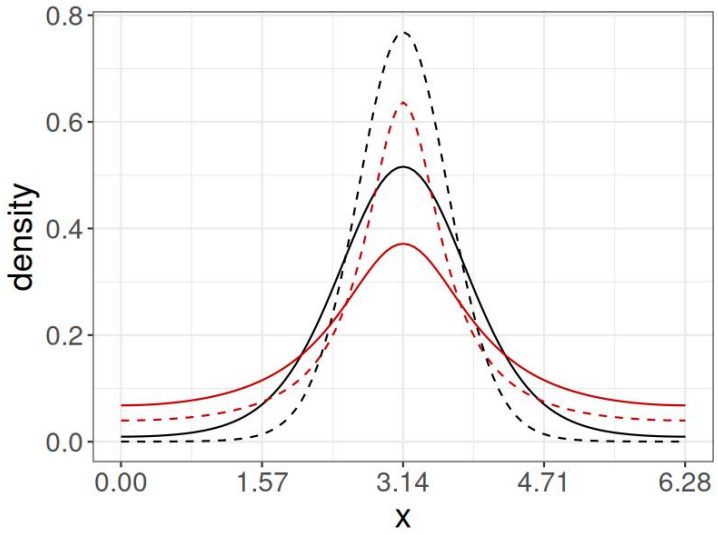


FIGURE 3.1: von Mises (black) and wrapped Cauchy (red) circular distributions, centered at $\mu = \pi$ and varying concentrations $\kappa = 2, 0.4$ (continuous) and $\kappa = 4, 0.6$ (dashed).

3.2 Marginal distributions for the vegetation stripe orientation variable

Axial data refers to observations of axes and are closely connected to circular data with a few key differences. Axial variables represent axes or the orientation of a segment in a plane that has neither a start nor an endpoint. Whereas, circular data represents a specific direction in the plane. Furthermore, axial data otherwise known as half-circle data only lie within the domain of $[0, \pi)$. Now assume that a random axial observation, like the orientation of a vegetation stripe, is indicated by variable Y. Up until recently numerous studies fitted axial data by changing the domain to $[0, 2\pi)$ (by multiplying the angle with 2) and then used the various circular distributions to model it [8]. However, in a study conducted by B. Arnold and A. SenGupta (2006) [8] they proposed a new method to obtain axial densities $f_{\text{axial}}(y)$, this method entails the wrapping of circular density onto the semi-circle by using:

$$Y = X(\text{mod})\pi. \tag{3.8}$$

Similar to the circular densities, in order for $f_{\text{axial}}(y)$ to be a valid axial density it needs to meet the following requirements [8]:

- $f_{\text{axial}}(y) \geq 0$ almost everywhere on $(-\infty, \infty)$,
- $f_{\text{axial}}(y + \pi) = f_{\text{axial}}(y)$ almost everywhere on $(-\infty, \infty)$,
- $\int_0^{\pi} f_{\text{axial}}(y)dy = 1$.

From the above requirements, it can be observed that $f_{axial}(y)$ is a nonnegative function that integrates to 1 over the support $[0, \pi)$ and that it is also π periodic. Under these specific requirements, the cumulative distribution function of a continuous axial random variable is expressed as:

$$F_{\text{axial}}(y) = \int_0^y f_{\text{axial}}(t)dt, \quad \text{where } y \in [0, \pi). \tag{3.9}$$

Some interesting properties of $F_{\text{axial}}(y)$ includes [8]:

- Similar to f_{axial}(y), $F_{\text{axial}}(y)$ retains periodicity
  ie. $F_{\text{axial}}(y + k\pi) = F_{\text{axial}}(y)$ so $F_{\text{axial}}(y + (k+1)\pi) - F_{\text{axial}}(y + k\pi) = 1$,
- $F_{\text{axial}}(y)$ is discontinuous at π ie. $\lim_{y \to k\pi} F_{\text{axial}}(y) = 1 \neq F_{\text{circ}}(k\pi) = F_{\text{axial}}(0) = 0$, for any integer k,
- $F^{-1}_{\text{axial}}(u)$ is a continuous mapping from the unit interval $u \in [0, 1)$ to the half-circle $[0, \pi)$.

To model the vegetation stripe orientation axial variable, Equation 3.4 will be used to wrap both the von Mises and wrapped Cauchy distributions on the half-circle. Using this approach ensures that both of the distributions are still unimodal and symmetric at the location parameter μ. However, the domain for both of the distributions is $[0, \pi)$.

The axial von Mises distribution or axial normal distribution, $AXN(\mu, \kappa)$, has the following density [8]:

$$f_{AXN}(y; \mu, \kappa) = [\pi I_0(\kappa)]^{-1} \cosh\{\kappa \cos(y - \mu)\}, \tag{3.10}$$

where $0 \leq y < \pi$, $0 \leq \mu < \pi$, $\kappa > 0$, I_0 is the modified Bessel function of the first kind and order 0 as defined in Equation 3.3.

The axial wrapped Cauchy distribution, $AXWC(\mu, \kappa)$, has the following density:

$$f_{AXWC}(y; \mu, \xi) = \frac{1}{\pi} \frac{1 - \xi^4}{1 + \xi^4 - 2\xi^2 \cos(2(y - \mu))} \tag{3.11}$$

where $\xi = e^{-\kappa}$, $0 \leq y < \pi$, $0 \leq \mu < \pi$, $\kappa > 0$.

Figure 3.2, illustrates the density of both the axial normal (black) and axial wrapped Cauchy (red) density. The plot also highlights the effects of the concentration parameter κ on both axial distributions. From this distribution plot, it can be observed that the axial wrapped Cauchy distribution is more concentrated compared to the axial normal distribution.

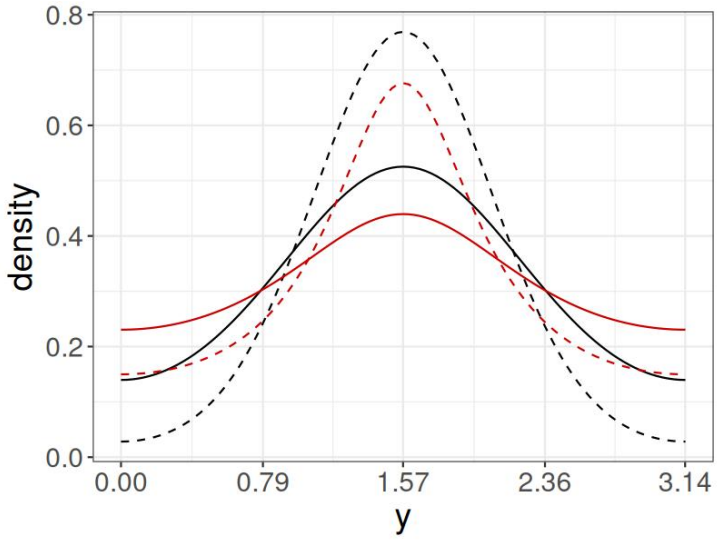


FIGURE 3.2: Axial normal (black) and axial wrapped Cauchy (red) distributions, centered at $\mu = \frac{\pi}{2}$ and varying concentrations $\kappa = 2, 0.4$ (continuous) and $\kappa = 4, 0.6$ (dashed).

3.3 Circular-axial bivariate distribution

A copula-based approach was considered in order to obtain a circular-axial bivariate distribution. A copula function is a function that expresses a multivariate distribution or a joint distribution of various random variables in terms of individual marginal distributions [39]. Copulas are used to express the interdependencies between variables explicitly as well as to model the correlation between these variables [39].

In order for a copula function to capture and portray the correct properties of a joint distribution function it must meet the following three requirements [39]:

- The copula function is an increasing function of its marginal distribution inputs.
- The copula function always returns a valid probability.
- If all the marginal distribution inputs are equal to 1 except for one, then the copula is equal to the value of the marginal distribution which was not equal to 1.

An appealing property of the copula approach is that it allows us to consider the marginal distributions separately from the dependency structure between the variables. Using Sklar's theorem [39], the bivariate copula function can be defined as:

$$F_{X_1,X_2}(x_1, x_2) = C(F_{X_1}(x_1), F_{X_2}(x_2)), \tag{3.12}$$

where C represents the copula function, $F_{X_i}(x_i)$ represents the distribution function of X_i for $i = 1, 2$ and $F_{X_1,X_2}(x_1, x_2)$ represents the joint distribution between variables X_1 and X_2. Furthermore, if $F_{X_1,X_2}(x_1, x_2)$ is continuous everywhere on $\mathbb{R}$, it can be differentiated with respect

to x_1 and x_2 to obtain the density which is expressed as:

$$f_{X_1,X_2}(x_1,x_2) = c(F_{X_1}(x_1), F_{X_2}(x_2)) f_{X_1}(x_1) f_{X_2}(x_2), \tag{3.13}$$

where $f_{X_i}(x_i)$ represents the density of X_i for $i = 1,2$ that is obtained by differentiating each $F_{X_i}(x_i)$ with its respective X_i, and c is the copula density function which is obtained by differentiating the copula function C.

In order to obtain the bivariate circular-axial density, a few different tools are needed.

One of the most well-known copula functions is the Gaussian copula which is defined as [40]:

$$C_\rho(x_1,x_2) = \Phi_\rho(\Phi^{-1}(x_1), \Phi^{-1}(x_2)) \qquad \text{where } 0 \leq x_1, x_2 \leq 1, \tag{3.14}$$

where Φ_ρ is the cdf of the bivariate standard normal distribution with correlation parameter $\rho \in [-1,1]$, and Φ^{-1} is the quantile function of the standard normal distribution.

Furthermore, the copula density function of the Gaussian copula is expressed as [40]:

$$c_\rho(F(x_1), F(x_2)) = \frac{\varphi_\rho(\Phi^{-1}(F(x_1)), \Phi^{-1}(F(x_2)))}{\varphi(\Phi^{-1}(F(x_2)))\varphi(\Phi^{-1}(F(x_2)))}, \tag{3.15}$$

where φ_ρ is the density of a bivariate standard normal with marginal densities φ. Equation 3.15 shows that Gaussian copula's density can be exploited to specify new bivariate distributions on $\mathbb{R}^2$ with the desired marginals.

Considering the above result, we can exploit the Quantile Function Theorem which is given by the following theorem [41]:

Theorem 1 (Quantile Function Theorem) *Let F be a cdf. If $F^{-1} : (0,1) \to (-\infty, \infty)$ is defined by*

$$F^{-1}(y) = \inf\{x : F(x) \geq y\}, \quad 0 < y < 1,$$

and U has the distribution of $U(0,1)$, then $X = F^{-1}(U)$ has cdf $F()$.

With Theorem 1 it is possible to transform any circular random variable to another circular random variable with a specific desired distribution. For example, let X be a random variable from the von Mises distribution with cdf F_{VM} and let F^{-1}_{WC} be the inverse cdf or quantile function of a wrapped Cauchy random variable. By using Theorem 1, the random variable $X' = F^{-1}_{\text{WC}}(F_{\text{VM}}(X))$ is distributed as a wrapped Cauchy with cdf F_{WC}. Similarly, Theorem 1 can be used not only to switch between axial random variables but also to transform an axial random variable into a circular random variable. For example, let Y be distributed as an axial normal distribution with cdf F_{AXN} and let F^{-1}_{WC} be the inverse cdf or quantile function of a wrapped Cauchy circular random variable. Under this setting, the random variable $Y' = F^{-1}_{\text{WC}}(F_{\text{AXN}}(Y))$ is distributed as a circular wrapped Cauchy with cdf F_{WC}.

Unfortunately, it is not possible to utilise standard copula functions such as the Gaussian, Clayton and Student's t copula since they lack the periodicity requirement which is required to specify a joint distribution with circular marginals [42] i.e.

$$\begin{aligned} c(F_{circ}(x_1), 0) &= c(F_{circ}(x_1), 1) \quad \text{where } 0 \leq F_{circ}(x_1) \leq 1, \\ c(0, F_{circ}(x_2)) &= c(1, F_{circ}(x_2)) \quad \text{where } 0 \leq F_{circ}(x_2) \leq 1. \end{aligned}$$

The group of copulas that meet the above requirements and are able to specify a joint distribution with circular marginals are called circulas [43]. To obtain the circular joint density it is custom to use the following [43]:

$$f_{circX_1,X_2}(x_1, x_2) = 4\pi^2 c_{circ}(2\pi F_{circX_1}(x_1), 2\pi F_{circX_2}(x_2)) f_{circX_1}(x_1) f_{circX_2}(x_2), \tag{3.16}$$

where $f_{circX_i}(x_i)$ represents the circular density of X_i which is obtained by differentiating each $F_{circX_i}(x_i)$ with its respective X_i, and c_{circ} is the circula density function.

However, in this study, a circula is obtained by replacing the bivariate normal density in Equation 3.15 with the five-parameter bivariate wrapped Cauchy density which was introduced by S. Kato and A. Pewsey (2015) [9]. The five-parameter bivariate wrapped Cauchy density is given by the following:

$$\begin{aligned} f_{KP}(x_1, x_2) = c(c_0 - c_1 \cos(x_1 - \mu_1) - c_2 \cos(x_2 - \mu_2) - c_3 \cos(x_1 - \mu_1)\cos(x_2 - \mu_2) \\ - c_4 \sin(x_1 - \mu_1)\sin(x_2 - \mu_2))^{-1}, \quad \text{for } 0 < x_1, x_2 \leq 2\pi, \end{aligned} \tag{3.17}$$

where x_i is a random wrapped Cauchy variable with distribution $WC(\mu_i, \kappa_i)$ such that $0 < \mu_i \leq 2\pi$, $0 \leq \xi_i < 1$, $\xi_i = e^{-\kappa_i}$.
Furthermore:

- $c = (1-\rho^2)(1-\xi_1^2)(1-\xi_2^2)/(4\pi^2)$,
- $c_0 = (1+\rho^2)(1+\xi_1^2)(1+\xi_2^2) - 8|\rho|\xi_1\xi_2$,
- $c_1 = 2(1+\rho^2)\xi_1(1+\xi_2^2) - 4|\rho|(1+\xi_2^2)\xi_2$,
- $c_2 = 2(1+\rho^2)(1+\xi_1^2)\xi_2 - 4|\rho|\xi_1(1+\xi_1^2)$,
- $c_3 = -4(1+\rho^2)\xi_1\xi_2 + 2|\rho|(1+\xi_1^2)(1+\xi_2^2)$,
- $c_4 = 2\rho(1-\xi_1^2)(1-\xi_2^2)$,
- $-1 < \rho < 1$.

The various circular and axial distributions, along with the circula function discussed in this chapter, will serve as foundational components in constructing the proposed modeling framework for this study.

Chapter 4

Methodology

While the concept of a circular-axial distribution is quite uncommon, Chapter 2 highlights concerns regarding latent heterogeneity and multimodality, especially for the stripe direction variable. This multimodality in the observations most likely originated from the influence of biotic and abiotic factors, such as soil moisture, slope angle, surrounding topography, and specific aspects of the cone where the stripe is located, which collectively have an effect on the orientation of the vegetation stripes. Unfortunately, capturing this complex behaviour accurately using a single model presents significant challenges. To address this issue, the proposed density will be incorporated into a finite mixture model that also accounts for the various concomitant variables. Furthermore, the EM algorithm will be used to maximise the likelihood and estimate the various parameters.

4.1 The proposed joint model

For this study, a special case of Equation 3.17 is used where it is assumed $\xi_i = 0$ for $i = 1, 2$. By making this assumption the following bivariate wrapped Cauchy density is obtained:

$$f_{KP}(x_1, x_2) = \frac{1 - \rho^2}{4\pi^2 \left(1 + \rho^2 - 2 \mid \rho \mid \cos(x_1)\cos(x_2) - 2\rho \sin(x_1)\sin(x_2)\right)}, \tag{4.1}$$

where $0 \leq x_1, x_2 \leq 2\pi, \rho \in [-1, 1]$.

The univariate marginal distributions of this joint density are circular uniform distributions which is expressed as:

$$g(x_1) = g(x_2) = (2\pi)^{-1},\ 0 \leq x_1, x_2 < 2\pi, \tag{4.2}$$

with the following cdfs:

$$G(x_1) = \frac{x_1}{2\pi},\ G(x_2) = \frac{x_2}{2\pi}, \tag{4.3}$$

and inverses:

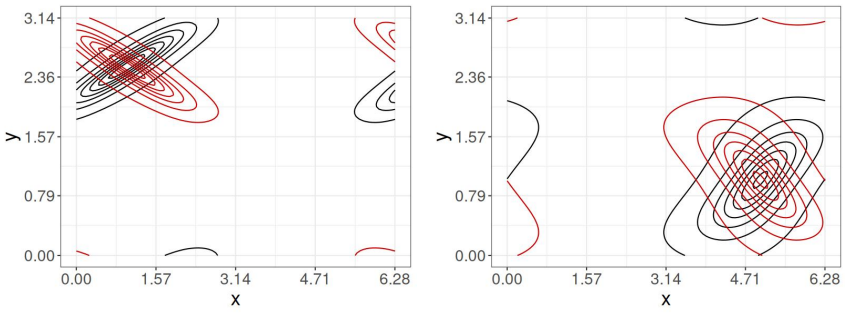
$$G^{-1}(a) = 2\pi a,\ G^{-1}(b) = 2\pi b. \tag{4.4}$$

Now, using the result from Chapter 3 the bivariate circular-axial distribution of the circular random variable X and the axial random variable Y can be obtained. Using Equation 3.15 and Theorem 1 the density function of the bivariate wrapped Cauchy circula $g_{KP}(F_{\text{circ}}(x), F_{\text{axial}}(y))$ can be calculated and is given by the following:

$$\begin{aligned} g_{KP}(F_{\text{circ}}(x), F_{\text{axial}}(y)) &= \frac{f_{KP}(G^{-1}(F_{\text{circ}}(x)), G^{-1}(F_{\text{axial}}(y)))}{g(G^{-1}(F_{\text{circ}}(x)))g(G^{-1}(F_{\text{axial}}(y)))} \\ &= \frac{1-\rho^2}{4\pi^2\,(1+\rho^2-2\mid\rho\mid\cos(2\pi F_{\text{circ}}(x))\cos(2\pi F_{\text{axial}}(y))-2\rho\sin(2\pi F_{\text{circ}}(x))\sin(2\pi F_{\text{axial}}(y)))\,\frac{1}{4\pi^2}} \\ &= \frac{1-\rho^2}{(1+\rho^2-2\mid\rho\mid\cos(2\pi F_{\text{circ}}(x))\cos(2\pi F_{\text{axial}}(y))-2\rho\sin(2\pi F_{\text{circ}}(x))\sin(2\pi F_{\text{axial}}(y)))}. \end{aligned} \tag{4.5}$$

Finally, using $g_{KP}(F_{\text{circ}}(x), F_{\text{axial}}(y))$ and Equation 3.13 the following bivariate circular-axial density can be obtained:

$$f_{circ,axial}(x,y) = g_{KP}(F_{\text{circ}}(x), F_{\text{axial}}(y))f_{circ}(x)f_{axial}(y). \tag{4.6}$$



(a) (b)

FIGURE 4.1: Example of the bivariate densities for: (a) circular von Mises and axial normal marginals with $\mu_{circ} = 1, \kappa_{circ} = 1$ and $\mu_{ax} = 2.5, \kappa_{ax} = 4$; (b) circular wrapped Cauchy and axial normal marginals with $\mu_{circ} = 5, \kappa_{circ} = 0.4$ and $\mu_{ax} = 1, \kappa_{ax} = 2$.

As can be seen in Figure 4.1, some properties of the joint density function include, a unimodal density whose shape is determined by ρ, if the value of ρ is positive (negative) then the correlation will be positive (negative) respectively. Furthermore, if $\rho = 0$ then it indicates that the two variables are independent. For large values of $|\rho|$ it indicates that x_1 and x_2 are close together in probability. Finally, in terms of the shape of the density contours, it is approximately elliptical-shaped near each of the modes and will distort as it gets further away from the modes [9].

4.2 The proposed mixture model

A finite mixture model can be utilised to determine the overall joint density from a weighted sum of K different underlying densities. Each underlying density is associated with its own α_k, which represents the probability of belonging to the kth class. The general finite mixture model is given by the following [44]:

$$f(x,y) = \sum_{k=1}^{K} \alpha_k f_k(x,y), \tag{4.7}$$

where $\boldsymbol{\alpha}$ is a vector of mixture weights α_k $(k = 1, 2, ..., K)$ such that each $\alpha_k > 0$ and the sum of all α_k is equal to 1. And $f_k(x,y)$ is the joint density function of the underlying component K.

A finite mixture model will be used to account for the data's multimodality. Assume that the underlying K components follow the bivariate circular-axial density:

$$f_{circ,axial}(x,y;\boldsymbol{\theta}_k) = g_{KP_k}(F_{circ}(x), F_{axial}(y)) f_{circ}(x;\mu_{circ_k},\kappa_{circ_k}) f_{axial}(y;\mu_{axial_k},\kappa_{axial_k}), \tag{4.8}$$

where $\boldsymbol{\theta}_k = (\rho_k, \mu_{circ_k}, \kappa_{circ_k}, \mu_{axial_k}, \kappa_{axial_k})$.

However, for the inclusion of the various environmental variables, a concomitant variable mixture model will be used [45]. Using this model it is possible to account for the heterogeneity caused by these specific covariates. The difference between a normal finite mixture model and a concomitant variable mixture model is that the mixing proportions α_k's can be modelled as depending on a set of covariates $z = (z_1, \ldots, z_p)$. Where the relationship between these probabilities and the covariates can be specified in terms of $(K-1) \times p$ regression coefficients through a generalised linear model (glm) with link function h, say $h(\alpha_k(z)) = z^\top \boldsymbol{\beta}_k$ [46, 47]. For this study, a logit link function will be used. Therefore, the mixing proportions α_k can be calculated as follows:

Case 1: For only two classes (K=2)
To model the log-odds of the mixing proportions for the K=2 case the following is used [48]:

$$\log\left(\frac{\alpha_2(z)}{\alpha_1(z)}\right) = z^\top \boldsymbol{\beta}, \tag{4.9}$$

where $\alpha_1(z)$ is the probability that observation i belongs to class 1 and $\alpha_2(z)$ that it belongs to class 2. Furthermore, z is the vector of covariates for each observation and $\boldsymbol{\beta}$ is the regression coefficients for class 2. Now, by using the fact that the sum of α_k should equal to 1, both of the α's can be solved simultaneously and express both of the mixing proportions as:

$$\alpha_2(z) = \frac{\exp(z^\top \boldsymbol{\beta})}{1 + \exp(z^\top \boldsymbol{\beta})}, \tag{4.10}$$

$$\alpha_1(z) = \frac{1}{1 + \exp(z^\top \boldsymbol{\beta})}. \tag{4.11}$$

Case 2: For more than two classes (K>2)
To model the log-odds of the mixing proportions for the K>2 case the following is used [48]:

$$\log\left(\frac{\alpha_k(z)}{\alpha_1(z)}\right) = z^\top \beta_k \quad \text{for} \quad k = 2, 3, \ldots, K \tag{4.12}$$

where $\alpha_k(z)$ is the probability that observation i belongs to the kth class and β_k is the regression coefficients for class k. Now, using the fact that the sum of α_k should equal to 1, all of the α's can be solved simultaneously and express all of the mixing proportions as:

$$\alpha_1(z) = \frac{1}{1 + \sum_{k=2}^{K} \exp(z^\top \beta_k)}, \tag{4.13}$$

$$\alpha_k(z) = \frac{\exp(z^\top \beta_k)}{1 + \sum_{k=2}^{K} \exp(z^\top \beta_k)}. \tag{4.14}$$

So by combining Equation 4.8 and the defined glm approach the following concomitant variable mixture model can be obtained:

$$f_{circ,axial}(x, y; \boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{k=1}^{K} \alpha_k(z; \boldsymbol{\beta}) f_{circ,axial}(x, y; \boldsymbol{\theta_k}). \tag{4.15}$$

4.3 Maximum likelihood estimation

In this study, the well-known EM algorithm [49] was used to find the maximum likelihood estimates (MLE) of the finite mixture model's parameters $\boldsymbol{\theta}$. In the finite mixture model framework, it is usually assumed that the observations $\{x_i, y_i\}_{i=1}^{n}$ are conditionally independent given the K latent classes. For this reason, the likelihood of the mixture model can be expressed as:

$$L(\boldsymbol{\theta}, \boldsymbol{\beta}; \boldsymbol{x}, \boldsymbol{y}) = \prod_{i=1}^{n} \sum_{k=1}^{K} f_{circ,axial}(x_i, y_i, \theta_k) \alpha_k(z_i; \boldsymbol{\beta}). \tag{4.16}$$

The EM algorithm is an iterative algorithm that starts from a set of initial values of $\boldsymbol{\theta}$ which get updated at each iteration until convergence is met. The algorithm combines the observed data with their latent posterior probabilities which are given by:

$$u_{ik} = \begin{cases} 0 & \text{if observation } i \text{ is not from component } k, \\ 1 & \text{if observation } i \text{ is from component } k, \end{cases}$$

to maximise the complete data log-likelihood (CDL) and obtain the MLE's of the parameters. The CDL function is given by:

$$\log L(\boldsymbol{\theta}, \boldsymbol{\beta}; \boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} \sum_{k=1}^{K} u_{ik} [\log(f_{circ,axial}(x_i, y_i | u_{ik}; \boldsymbol{\theta}_k)) + \log(\alpha_k(z_i; \boldsymbol{\beta}))]. \tag{4.17}$$

With the EM algorithm, each iteration consist out of an expectation step (E step) and a maximisation step (M step). In the **E step**, the posterior probabilities are updated based on the current iteration t's set of parameters $\alpha_k^{(t)}$ and $\boldsymbol{\theta}^{(t)} = (\boldsymbol{\theta}_1^{(t)}, \ldots, \boldsymbol{\theta}_K^{(t)})$, by using the following:

$$\begin{aligned}
\gamma_{ik} &= \mathbb{E}(u_{ik} \mid x_i, y_i, \boldsymbol{\theta}_k^{(t)}) \\
&= \mathbb{P}(u_{ik} = 1 \mid x_i, y_i, \boldsymbol{\theta}_k^{(t)}) \\
&= \frac{P(u_{ik} = 1, x_i, y_i \mid \boldsymbol{\theta}_k^{(t)})}{P(x_i, y_i \mid \boldsymbol{\theta}_k^{(t)})} \\
&= \frac{P(x_i, y_i \mid u_{ik}, \boldsymbol{\theta}_k^{(t)}) P(u_{ik} = 1 \mid \boldsymbol{\theta}_k^{(t)})}{P(x_i, y_i \mid \boldsymbol{\theta}_k^{(t)})} \\
&= \frac{\alpha_k^{(t)} f_{circ,axial}(x_i, y_i \mid u_{ik}, \boldsymbol{\theta}_k^{(t)})}{\sum_{k=1}^{K} \alpha_k^{(t)} f_{circ,axial}(x_i, y_i \mid u_{ik}, \boldsymbol{\theta}_k^{(t)})}.
\end{aligned} \tag{4.18}$$

Now the expected complete data log-likelihood (ECDL) can be obtained by replacing u_{ik} in Equation 4.17 with $\mathbb{E}(u_{ik} = 1 \mid x_i, y_i, \theta_k^{(t)}) = \gamma_{ik}$, the ECDL is given by:

$$Q(\boldsymbol{\theta}) + Q(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{k=1}^{K} \gamma_{ik} [\log(f_{circ,axial}(x_i, y_i | u_{ik}; \boldsymbol{\theta}_k)) + \log(\alpha_k(\boldsymbol{z}_i; \boldsymbol{\beta}))]. \tag{4.19}$$

In the **M step** the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are updated by maximising the ECDL. Since the ECDL is the sum of two functions that depend on independent sets of paramaters, $Q(\boldsymbol{\theta})$ and $Q(\boldsymbol{\beta})$ can be maximised separately. For the maximisation of $Q(\boldsymbol{\theta})$'s parameters the weighted inference of margins approach is used, this allows us to obtain the MLE's for each of the marginal distributions independently using γ_{ik} as the weights. Now that both the circular and axial marginals' MLEs are obtained, it can be used to numerically maximise the ρ parameter using the log of Equation 4.8. To maximise $Q(\boldsymbol{\beta})$, a fitted weighted multinomial regression with a logit link function was used.

4.4 Goodness of fit measure

To determine which of the four different possible combinations for the marginal distributions provides the best fit for the density $f_{circ,axial}(x_i, y_i; \boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{k=1}^{K} \alpha_k(\boldsymbol{u}_i; \boldsymbol{\beta}) f_{circ,axial}(x_i, y_i; \boldsymbol{\theta}_k)$, the Bayesian Information Criterion (BIC) was considered. Additionally, the BIC score will be used to identify the optimal value of K. The model with the lowest BIC score is considered to be the best fit. By using the BIC score it is possible to determine which model provides the best balance between goodness of fit and model complexity [50] .

The BIC score can be calculated by using the following equation:

$$BIC = -2ln(L(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}; \boldsymbol{x}, \boldsymbol{y})) + 2ln(n)r\,, \tag{4.20}$$

where $L(\hat{\theta}, \hat{\beta}; x, y)$ being the maximum likelihood function of the specific model, r being the amount of estimated parameters in the model and n being the total number of observations. From Equation 4.20, it can be seen that the BIC score rewards a model that fits the data well (higher maximum likelihood), but penalises a model that is more complex (more parameters) or less parsimonious.

Chapter 5

Application

The following Chapter presents the simulation study, designed to evaluate the robustness of the proposed model across a wide range of variables and varying numbers of latent classes. Additionally, it includes the results obtained from applying the proposed model to the Marion Island dataset, followed by an in-depth interpretation of these findings in the discussion section.

5.1 Simulation study

The simulation study can be used to test the model under different scenarios and ensure that the proposed model estimates the various parameters correctly for the specified number of latent classes. The simulation study is conducted by generating 200 replicas, each with $n = 600$ observations, from the model described in Equation 4.15 by using Algorithm 1.

Algorithm 1 Simulation from a mixture of copula-based axial-circular distributions

Require: a vector of covariates $\boldsymbol{z}$ and parameters $(\boldsymbol{\beta}, \boldsymbol{\mu_{circ}}, \boldsymbol{\kappa_{circ}}, \boldsymbol{\mu_{axial}}, \boldsymbol{\kappa_{axial}})$
Ensure: $n \geq 2$
for $i = 1, \ldots, n$ **do**
1: Generate mixing proportions:
a: Compute $\eta_{ij} = \boldsymbol{z}_i^\top \boldsymbol{\beta}_k,\ k = 2, \ldots, K$
b: Compute $\alpha_{ik} = \frac{\exp(\eta_{ik})}{1+\sum_{k=2}^{K}\exp(\eta_{ik})},\ k = 2, \ldots, K$
c: Compute $\alpha_{i1} = \frac{1}{1+\sum_{k=2}^{K}\exp(\eta_{ik})}$
2: Simulate latent class membership:
By using a multinomial distribution with the generated mixing proportions as probabilities such that $u_{ik} \sim Multinom(\alpha_{i1}, ..., \alpha_{iK})$
3: Simulate axial-circular observations:
a: Draw $v \sim Unif(0, 1)$
b: Compute $y_i = F^{-1}_{\text{axial}}(v; \boldsymbol{\mu_{axial}}, \boldsymbol{\kappa_{axial}})$
c: Compute $\delta_1 = 2\pi(F_{\text{axial}}(y_i; \boldsymbol{\mu_{axial}}, \boldsymbol{\kappa_{axial}}))$
d: Draw $\delta_2 \sim f_{CWC}(\delta \mid \delta_1)$,
where $f_{CWC}(\delta \mid \delta_1)$ is the conditional wrapped Cauchy obtained by S. Kato and A. Pewsey (2015) [9], Th. 1, Sec. 2.4, with $\kappa_1 = \kappa_2 = 0$
e: Compute $x_i = F^{-1}_{\text{circ}}(2\pi^{-1}\delta_2; \boldsymbol{\mu_{circ}}, \boldsymbol{\kappa_{circ}})$
end for

Using the Algorithm 1, the proposed model can be tested on scenarios where we have $K \in \{2,3\}$ latent classes. And use specific parameter values to test the model on a whole range of the data domain $((x,y) \mid 0 \leq x \leq 2\pi, 0 \leq y \leq \pi)$. Two covariates are also included in the simulation study where the values of the covariates have been randomly generated from $N(0,4)$ and $Ber(0.5)$ and the corresponding $\boldsymbol{\beta}$ regression coefficients from $N(0,4)$. To estimate the β's and the mixing probabilities the glm() function was used for K=2 and the multinom() function from the nnet package was used for K>2 in R software. All of the parameters MLE's were obtained using the optimize() function in R software. Furthermore, the EM algorithm was set to run for either a maximum of 100 iterations or until the convergence threshold of 0.0001 was met ($|ECDL_{new} - ECDL_{old}| > 0.0001$)).

With this specific decision on covariate parameters, the model can be tested on both continuous and categorical variables. Even though these few scenarios are not exhaustive, we are still able to explore potential real-world scenarios where latent components may or may not overlap and show positive or negative correlations, while still being able to account for different effects of covariates on the class membership probabilities. Using this simulation study, the equal-tail 95% confidence intervals can be obtained for each parameter for each of the four different possible combinations of the circular-axial marginal distributions. The simulation study results for both the $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ parameters can be found in Table 5.1 and 5.2. From the results, it can be seen that for all the cases the parameter estimates are accurate and that the true values are included in the 95% confidence intervals. Thus, highlighting the fact that the proposed model is capable of correctly and accurately recovering the parameters for both the observed and latent components.

TABLE 5.1: Results of the simulation study: estimation of the regression parameters.

Family	K	β_0		β_1		β_2	
		true	est	true	est	true	est
VM-AXN	2	-2.41	-2.43 (-2.94, -2.00)	0.55	0.56 (0.43, 0.71)	2.17	2.18 (1.72, 2.77)
	2	-0.09	-0.06 (-0.51, 0.38)	0.64	0.66 (0.50, 0.83)	0.12	0.10 (-0.43, 0.60)
	3	1.32	1.36 (1.01, 1.75)	1.17	1.19 (1.01, 1.46)	-2.93	-2.96 (-3.67, -2.37)
VM-AXWC	2	-0.86	-0.86 (-1.11, -0.55)	0.23	0.23 (0.14, 0.33)	0.37	0.37 (0.01, 0.71)
	2	-0.86	-0.88 (-1.31, -0.55)	0.37	0.37 (0.26, 0.49)	0.54	0.55 (-0.02, 1.13)
	3	0.23	0.22 (-0.10, 0.45)	-0.07	-0.08 (-0.17, 0.01)	-0.19	-0.20 (-0.60, 0.17)
WC-AXN	2	-1.26	-1.32 (-2.10, -0.67)	3.67	3.87 (2.82, 5.17)	0.69	0.75 (-0.10, 1.72)
	2	-0.09	-0.07 (-0.66, 0.59)	0.64	0.65 (0.45, 0.91)	0.12	0.11 (-0.62, 0.79)
	3	1.32	1.36 (0.87, 1.91)	1.17	1.21 (0.95, 1.56)	-2.93	-3.02 (-3.93, -2.13)
WC-AXWC	2	-0.86	-0.87 (-1.40, -0.38)	0.23	0.25 (-0.14, 0.77)	0.37	0.38 (-0.06, 0.79)
	2	-0.86	-0.87 (-1.18, -0.552)	0.37	0.38 (0.21, 0.55)	0.54	0.55 (-0.03, 1.21)
	3	0.23	0.25 (-0.14, 0.77)	-0.07	-0.08 (-0.22, 0.06)	-0.19	-0.23 (-0.77, 0.32)

TABLE 5.2: Results of the simulation: estimation of the parameters of the density.

Family	K	μ_{circ}		κ_{circ}		μ_{axial}		κ_{axial}		ρ	
		true	est	true	est	true	est	true	est	true	est
VM-AXN	1	1	1.00 (0.92, 1.09)	2	2.03 (1.74, 2.31)	0.5	0.50 (0.40, 0.61)	2	2.03 (1.70, 2.34)	-0.45	-0.45 (-0.50, -0.39)
	2	5	5.00 (4.94, 5.07)	6	6.19 (4.87, 8.16)	2	2.00 (1.93, 2.06)	5	5.12 (4.06, 6.31)	0.60	0.58 (0.49, 0.65)
	1	1	1.00 (0.90, 1.10)	3	3.09 (2.48, 3.87)	0.5	0.49 (0.35, 0.62)	2	2.08 (1.54, 2.62)	-0.45	-0.39 (-0.50, -0.11)
	2	5	5.00 (4.92, 5.07)	5	5.05 (3.69, 6.71)	2	2.00 (1.92, 2.07)	5	5.14 (3.96, 6.79)	0.60	0.57 (0.46, 0.64)
	3	3	3.00 (2.95, 3.04)	10	10.10 (8.31, 12.70)	1.5	1.50 (1.45, 1.54)	9	9.13 (7.63, 11.10)	0.10	0.09 (-0.06, 0.17)
VM-AXWC	1	1	0.99 (0.86, 1.11)	3	3.03 (1.99, 3.92)	0.5	0.61 (0.13, 2.86)	0.3	0.31 (0.15, 0.43)	-0.45	-0.43 (-0.52, -0.36)
	2	5	4.98 (4.91, 5.08)	5	5.24 (3.60, 6.83)	2	2.00 (1.83, 2.15)	0.55	0.56 (0.45, 0.65)	0.60	0.57 (0.45, 0.66)
	1	1	1.00 (0.92, 1.10)	2	2.04 (1.68, 2.40)	0.5	0.51 (0.20, 0.83)	0.3	0.31 (0.13, 0.41)	-0.45	-0.41 (-0.49, -0.31)
	2	5	4.98 (4.93, 5.07)	6	5.99 (4.60, 7.37)	2	2.00 (1.92, 2.08)	0.7	0.70 (0.64, 0.75)	0.60	0.58 (0.50, 0.64)
	3	3	3.00 (2.95, 3.04)	10	10.02 (8.31, 12.30)	1.5	1.50 (1.48, 1.52)	0.9	0.90 (0.88, 0.92)	0.10	0.09 (-0.08, 0.18)
WC-AXN	1	1	1.00 (0.76, 1.24)	0.3	0.30 (0.25, 0.37)	0.5	0.50 (0.37, 0.61)	2	2.02 (1.69, 2.42)	-0.45	-0.44 (-0.50, -0.38)
	2	5	5.00 (4.98, 5.02)	0.9	0.90 (0.88, 0.92)	2	2.00 (1.95, 2.05)	5	5.07 (4.37, 6.00)	0.60	0.59 (0.53, 0.64)
	1	1	1.00 (0.71, 1.29)	0.3	0.31 (0.21, 0.40)	0.5	0.50 (0.33, 0.67)	2	2.11 (1.55, 2.81)	-0.45	-0.43 (-0.52, -0.34)
	2	5	5.00 (4.98, 5.03)	0.9	0.90 (0.87, 0.92)	2	2.00 (1.92, 2.07)	5	5.21 (3.65, 7.18)	0.60	0.58 (0.46, 0.66)
	3	3	3.00 (2.86, 3.16)	0.5	0.50 (0.41, 0.58)	1.5	1.50 (1.45, 1.55)	9	9.19 (7.44, 11.70)	0.10	0.09 (-0.06, 0.20)
WC-AXWC	1	1	1.01 (0.82, 1.23)	0.3	0.31 (0.23, 0.37)	0.5	0.49 (0.20, 0.80)	0.3	0.31 (0.16, 0.41)	-0.45	-0.42 (-0.50, -0.29)
	2	5	5.00 (4.98, 5.02)	0.9	0.90 (0.88, 0.92)	2	2.00 (1.92, 2.06)	0.7	0.71 (0.64, 0.76)	0.60	0.58 (0.51, 0.65)
	1	1	1.05 (0.70, 1.43)	0.3	0.31 (0.18, 0.42)	0.5	0.55 (0.13, 1.21)	0.3	0.33 (0.14, 0.51)	-0.45	-0.40 (-0.51, -0.22)
	2	5	5.00 (4.98, 5.03)	0.9	0.90 (0.87, 0.92)	2	2.00 (1.81, 2.13)	0.55	0.57 (0.46, 0.66)	0.60	0.57 (0.46, 0.66)
	3	3	3.01 (2.87, 3.14)	0.5	0.50 (0.39, 0.59)	1.5	1.49 (1.48, 1.52)	0.9	0.90 (0.87, 0.92)	0.10	0.09 (-0.05, 0.19)

5.2 Real data application

In this section, the proposed model will be applied to the Marion Island dataset discussed in Chapter 2, where the axial variable represents the vegetation stripe orientation, and the circular variable corresponds to the wind direction measured at $n = 133$ locations across the island. To model the latent class membership probabilities, the cone aspect (NE, NW, SE, SW) and slope angle will be included as covariates. The cone aspect identifies the quadrant of the cone where each vegetation stripe observation is situated, while the slope angle quantifies the inclination of the terrain surrounding each vegetation stripe orientation observation. To determine the best-fitting model for the Marion Island data, four different combinations of circular-axial marginal densities, (VM-AXN, VM-AWC, WC-AXN, and WC-AWC) will be evaluated, along with three potential values for the number of underlying latent classes ($K = 2, 3, 4$). The models will be assessed using the ECDL and BIC scores, with results summarised in Table 5.3.

From Table 5.3, it can be concluded that based on the lowest BIC score the best-performing model overall is a model with K=3 underlying components and VM-AXN marginals. Furthermore, the final parameter estimates are summarised in Tables 5.4 and 5.5, as well as each parameter's 90% equal-tail confidence intervals which were obtained using a 10 000 interval bootstrap procedure. From these results, it is evident that the confidence interval bands are quite spread out and wide. This is most likely due to the limited number of observations, allowing the model to estimate a range of possible different results. If we had more observations, as was the case in the simulation study, the confidence intervals would be narrower and would have reflected greater certainty in the parameter estimates.

TABLE 5.3: Expected complete-data log-likelihood and Bayesian Information Criterion for each estimated model.

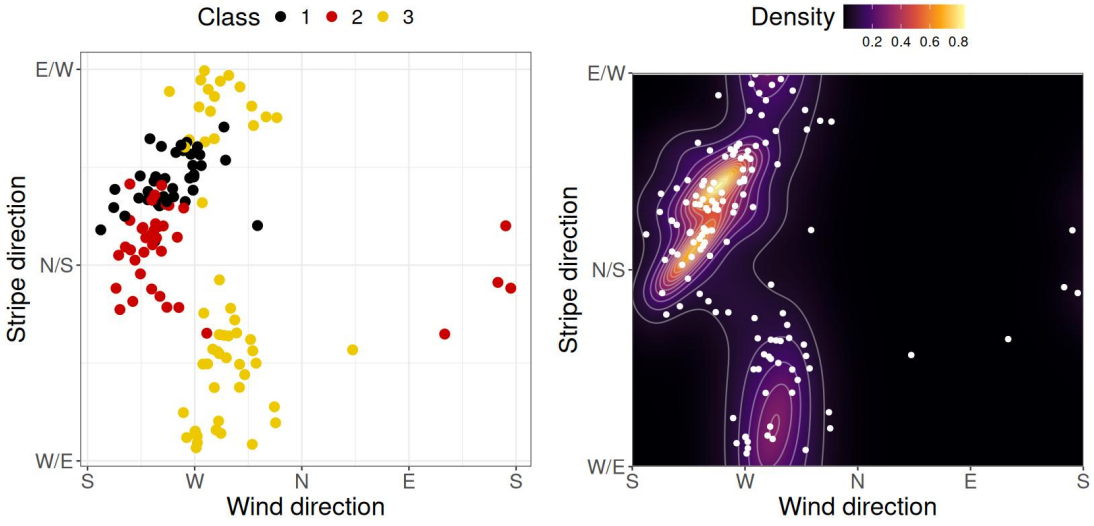
Model / K	**logL**			**BIC**		
	2	3	4	2	3	4
VM-AXN	-182.14	**-156.02**	-147.32	427.86	**419.62**	446.23
WC-AXN	-185.50	-162.28	-139.66	434.58	432.15	430.93
VM-AXWC	-192.88	-167.05	-149.59	460.53	454.43	468.56
WC-AXWC	-192.87	-167.83	-148.57	460.52	456.00	470.65

TABLE 5.4: Point estimates (95% equal-tail confidence intervals) for the parameters of the density for K=3 latent classes.

$Class$	μ_{circ}	κ_{circ}	μ_{axial}	κ_{axial}	ρ
1	-1.92 (-2.04, -1.81)	5.60 (4.66, 10.07)	0.68 (0.62, 0.75)	18.96 (15.0, 36.7)	0.25 (0.09, 0.43)
2	-2.32 (-2.45, -2.20)	5.00 (4.26, 9.21)	0.14 (0.05, 0.23)	10.01 (8.42, 19.7)	0.35 (0.12, 0.49)
3	-1.16 (-1.26, -1.05)	5.99 (5.42, 10.03)	-1.25 (-1.39, -1.10)	3.13 (2.63, 5.41)	0.07 (-0.17, 0.23)

TABLE 5.5: Estimated regression coefficients for K=3 latent classes. Reference class is $Class = 1$.

Class	***NE***	***NW***	***SE***	***SW***	*Slope*
2	-15.30 (-207, -5.93)	-6.81 (-37.5, 6.89)	11.40 (4.88, 149)	-11.01 (-159, -5.17)	0.30 (0.05, 8.83)
3	7.35 (1.34, 63.6)	-1.50 (-9.71, 35.2)	-17.12 (-99.9, -13.2)	-10.16 (-69, -4.67)	0.05 (-0.04, 0.23)



(a) (b)

FIGURE 5.1: (a) Estimated classification results of the model with $K = 3$. (Each observation has been assigned to a latent component where class 1=black, class 2=red, class 3=gold, (b) Overall joint density of the Marion Island data.

FIGURE 5.2: Conditional distribution in the form of rose diagrams of wind direction (top) and vegetation stripe orientations (bottom), given the allocation within three latent classes, where class 1=black, class 2=red, class 3=gold.

In Figure 5.1, the estimated classification results for the proposed finite mixture model are visualised, alongside the overall joint density fitted to the Marion Island dataset. Figure 5.1 provides a detailed representation of how the model classifies the observations based on the fitted density. Additionally, Figure 5.2 offers an alternative perspective through the use of rose

diagrams, illustrating the conditional distributions of both the wind direction (circular) and vegetation stripe orientation (axial) for the observations within each class. These diagrams highlight the distinct distributional patterns captured by the model for each class. By using both figures, it can be seen that classes 1 and 2 demonstrate a positive correlation between the wind direction and vegetation stripe orientation.

By specifically focusing on class 1, which consists of vegetation stripe observations that are mostly located on the Southwestern aspects of the scoria cones. It can be seen that it has a positive correlation (≈ 0.25) between the variables, with vegetation stripes growing towards the northeast (southwest) and wind blowing mostly from the southwest.

Class 2, consists of observations that are predominantly located on the Southeastern quadrant of the scoria cones. It can also be observed that it has a positive correlation between the variables (≈ 0.35), with vegetation stripes having a north (south) orientation and winds that have a more southern direction compared to the winds from class 1. Apart from this, the observations in both class 1 and class 2 behaves quite similarly. Furthermore, by specifically looking at the *Slope* regression coefficient (5.5) we can observe that the scoria cone surfaces on which class 2's vegetation stripe observations are located have a significantly steeper slope compared to class 1's southeastern aspects.

Finally, the third class mainly consists of vegetation stripes which are located on the Northern quadrants of the scoria cones. It can be seen that there is no significant correlation between the variables (≈ 0.07). Furthermore, the vegetation stripe on the Northeastern quadrant was orientated along both the southwestern (northeastern) and southeastern (northwestern) axis despite the wind blowing consistently from the northwestern direction. A similar pattern, although with more variation, was observed with the northwesterly located vegetation stripes, with orientation spanned across almost all of the axis (with only north (south) orientations being rare) while experiencing the same wind from the northwestern direction. Thus, while wind direction was relatively consistent on these aspects of scoria cones (westerly to north-westerly winds), stripe direction was significantly variable. Furthermore, for this component, the effect of slope was negligible.

5.3 Discussion

Our findings, based on data collected from scoria cones under various environmental conditions on Marion Island, suggest that wind direction does contribute to the specific orientation of vegetation stripes. However, this is only the case on the southern sides of these cones. The dominant winds likely influence stripe orientation by allowing neighboring plants to shelter those located downwind. As we know, strong winds can negatively impact plants by causing leaf damage, increasing water loss, slowing photosynthesis, and even breaking stems [51]. Therefore, wind shelter can positively affect plant growth and survival, leading to a pattern

where plants grow behind existing vegetation or any structure that blocks the dominant wind [52].

On the other hand, there was no clear link between wind direction and vegetation stripe orientation on the northern sides of scoria cones. This was quite unexpected and might be a clear indication of important environmental differences between the northern and southern aspects of the scoria cones. One possible explanation is that northern aspects experience stronger winds, which overwhelm the protective effects of neighboring plants. In such conditions, plants may not benefit from the shelter of others, and different factors could drive stripe formation. However, from the CFD simulations, we know that wind speeds vary more between the east and west sides of the cones than between the north and south sides.

Another possible explanation for the difference could be varying levels of sunlight and surface temperatures [53]. Equator-facing slopes (north) receive more direct sunlight and tend to be warmer, which generally provides better-growing conditions in cold environments. For example, on Macquarie Island, plant cover has been observed to vary based on aspect, possibly due to differences in temperature (although this can also be due to differences in wind exposure). This suggests that on the warmer northern slopes, plants might not need as much shelter from neighboring vegetation, reducing the link between wind direction and stripe orientation.

In contrast to wind direction and cone aspect, from the results, it can be seen that slope angle has a weak influence on the orientation of vegetation stripes. This aligns with similar findings from other studies in semi-arid areas [1]. So even though slope angle could affect how water or substrate moves, it does not seem to have a strong enough impact on the formation of vegetation stripes.

However, the influence of slope steepness may also potentially be related to the relative orientation of wind direction and slope direction. For example, if the dominant wind is blowing directly up a slope (parallel with the slope aspect), we would expect vegetation stripes to be orientated vertically (a strong positive correlation), irrespective of slope steepness. However, in the opposite scenario where the wind is blowing across the slope (perpendicular to the slope), we would anticipate vegetation stripes that follow the same orientation as the wind only on flatter surfaces. Instead, on steeper slopes, the orientation of the vegetation stripes should reflect the influence of both wind direction and the movement of the substrate downslope. Therefore, while the slope is clearly not a strong driver of vegetation stripe orientation in this system, there are still potential mechanisms of influence that could be examined.

Chapter 6

Conclusion

In this study, a bivariate density for handling circular and axial data was developed. This was done to model the joint density of Marion Islands vegetation stripe orientation data and wind direction data to finally determine if wind is the reason for the occurrence of these irregular stripes. This was achieved by using a copula-based approach, specifically using a bivariate wrapped Cauchy circula with a variety of different circular and axial distributions. For the marginal distributions, the von Mises and wrapped Cauchy distributions were considered as the circular distributions and the axial normal and axial wrapped Cauchy distributions for the axial variables were considered. However, the joint density of the wind direction variable and vegetation stripe orientation variable display multimodality. To deal with this issue a finite mixture model was used to identify various underlying latent components in the data. Additionally, the finite mixture model incorporated two concomitant variables: the slope angle and the cone aspect, such that the model could account for the heterogeneity caused by these covariates in the model. This was achieved by using a generalised linear model with a logit link function to determine the various class membership probabilities. For the maximum likelihood estimation of the proposed concomitant variable mixture model, the EM algorithm was used. We implemented a simulation study in order to verify that the proposed model returns the correct estimated parameters. Finally, twelve different combinations of circular-axial models with $K = 1, 2, 3$ latent components were fitted to the Marion Island data set. With the BIC goodness of fit measure, it was determined that the best-fitting model was a model with three underlying components with von Mises and axial normal distributions as marginal distributions.

With this fitted model it can be concluded that wind direction influences vegetation stripe orientation, but only on the southern sides of cones. This is likely due to dominant winds allowing plants to shelter others downwind, enhancing growth. On the northern sides of the cones, no clear link between wind and stripe orientation was observed, possibly due to stronger winds or different sunlight exposure creating warmer conditions that reduce the need for shelter. Slope angle has a weak influence, though interactions between wind direction and slope steepness may still affect stripe orientation. Overall, it can be concluded that wind as well as other environmental factors such as the cone aspect influences the orientation of vegetation stripes found on Marion Island.

For future work, the scope could be expanded by increasing the number of combinations of marginal distributions considered. Additionally, incorporating other potentially significant covariates, such as soil moisture, solar surface irradiance, and ground temperature, may enhance the model's applicability. Developing a regression model to predict axial variables based on circular variables and adapting the estimation algorithm to accommodate spatially correlated data are also promising directions for further research.

Bibliography

[1] V. Deblauwe *et al.*, "The global biogeography of semi-arid periodic vegetation patterns," *Global Ecology and Biogeography*, vol. 17, no. 6, pp. 715–723, 2008.

[2] M. F. Bekker and G. P. Malanson, "Linear forest patterns in subalpine environments," *Progress in Physical Geography*, vol. 32, no. 6, pp. 635–653, 2008.

[3] M. Rietkerk, S. C. Dekker, P. C. De Ruiter, and J. van de Koppel, "Self-organized patchiness and catastrophic shifts in ecosystems," *Science*, vol. 305, no. 5692, pp. 1926–1929, 2004.

[4] P. C. Le Roux and M. A. McGeoch, "Spatial variation in plant interactions across a severity gradient in the sub-antarctic," *Oecologia*, vol. 155, pp. 831–844, 2008.

[5] J. E. Baartman, A. J. Temme, and P. M. Saco, "The effect of landform variation on vegetation patterning and related sediment dynamics," *Earth Surface Processes and Landforms*, vol. 43, no. 10, pp. 2121–2135, 2018.

[6] S. W. Morgan, J. B. Kirkpatrick, and M.-B. Di Folco, "Wind-controlled linear patterning and cyclic succession in tasmanian sphagnum mires," *Journal of Ecology*, vol. 98, no. 3, pp. 583–591, 2010.

[7] F. Zhang, H. Zhang, M. R. Evans, and T. Huang, "Vegetation patterns generated by a wind driven sand-vegetation system in arid and semi-arid areas," *Ecological Complexity*, vol. 31, pp. 21–33, 2017.

[8] B. C. Arnold and A. SenGupta, "Probability distributions and statistical inference for axial data," *Environmental and Ecological Statistics*, vol. 13, pp. 271–285, 2006. DOI: 10.1007/s10651-004-0011-8.

[9] S. Kato and A. Pewsey, "A Möbius transformation-induced distribution on the torus," *Biometrika*, vol. 102, no. 2, pp. 359–370, Mar. 2015, ISSN: 0006-3444. DOI: 10.1093/biomet/asv003. [Online]. Available: https://doi.org/10.1093/biomet/asv003.

[10] C. M. Dayton and G. B. Macready, "Concomitant-variable latent-class models," *Journal of the American Statistical Association*, vol. 83, no. 401, pp. 173–178, 1988. DOI: 10.1080/01621459.1988.10478584. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/01621459.1988.10478584.

[11] S. D. Holness, "Sediment movement rates and processes on cinder cones in the maritime subantarctic (marion island)," *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group*, vol. 29, no. 1, pp. 91–103, 2004.

[12] V. R. Smith and L. Mucina, "Vegetation of subantarctic marion and prince edward islands," *The Vegetation of South Africa, Lesotho and Swaziland. Strelitzia*, vol. 19, pp. 698–723, 2006.

[13] S. D. Holness, "The orientation of sorted stripes in the maritime subantarctic, marion island," *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group*, vol. 26, no. 1, pp. 77–89, 2001.

[14] S. Chown and P. W. Froneman, *The Prince Edward Islands: land-sea interactions in a changing ecosystem*. African Sun Media, 2008.

[15] P. C. Le Roux and M. A. McGeoch, "Changes in climate extremes, variability and signature on sub-antarctic marion island," *Climatic Change*, vol. 86, no. 3, pp. 309–329, 2008.

[16] J. La Grange, "The south african station on marion island, 1948–53," *Polar Record*, vol. 7, no. 48, pp. 155–158, 1954.

[17] W. Nel *et al.*, "Earth science research on marion island (1996–2020): A synthesis and new findings," *South African Geographical Journal*, vol. 103, no. 1, pp. 22–42, 2021.

[18] V. R. Smith, "Climate change in the sub-antarctic: An illustration from marion island," *Climatic Change*, vol. 52, no. 3, pp. 345–357, 2002.

[19] E. M. Van Zinderen Bakker, "Geoecology of the marion and prince edward islands (sub-antarctic)," *Erdwissenschaftliche Forschung*, 1975.

[20] A. L. Washburn, "Classification of patterned ground and review of suggested origins," *Geological Society of America Bulletin*, vol. 67, no. 7, pp. 823–866, 1956.

[21] K. Hall, "Sorted stripes orientated by wind action: Some observations from sub-antarctic marion island," *Earth Surface Processes*, vol. 4, no. 3, pp. 281–289, 1979.

[22] E. Löffler, "Macquarie island: A wind-molded natural landscape in the subantarctic," *Polar Geography*, vol. 8, no. 4, pp. 267–286, 1984.

[23] D. W. Walton and T. Heilbronn, "Periglacial activity on the subantarctic island of south georgia," *National Academy Press*, 1983.

[24] K. Hall, "Some observations regarding sorted stripes, livingston island, south shetlands," *Permafrost and Periglacial Processes*, vol. 5, no. 2, pp. 119–126, 1994.

[25] K. Hall, "Sorted stripes on sub-antarctic kerguelen island," *Earth Surface Processes and Landforms*, vol. 8, no. 2, pp. 115–124, 1983.

[26] J. M. Selkirk, "Active vegetation-banked terraces on macquarie island," *Zeitschrift für Geomorphologie*, vol. 42, no. 4, pp. 483–496, 1998.

[27] J. Boelhouwers, S. Holness, and P. Sumner, "Geomorphological characteristics of small debris flows on junior's kop, marion island, maritime sub-antarctic," *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group*, vol. 25, no. 4, pp. 341–352, 2000.

[28] J. Peterson, J. Scott, and E. Derbyshire, "Sorted stripes of periglacial origin," *Australian Geographer*, vol. 15, no. 5, pp. 325–328, 1983.

[29] D. W. Hedding, "Spatial inventory of landforms in the recently exposed central highland of sub-antarctic marion island," *South African Geographical Journal*, vol. 90, no. 1, pp. 11–21, 2008.

[30] D. W. Hedding, W. Nel, and R. L. Anderson, "Aeolian processes and landforms in the sub-antarctic: Preliminary observations from marion island," *Polar Research*, vol. 34, no. 1, p. 26 365, 2015.

[31] K. Goddard, K. Craig, J. Schoombie, and P. Le Roux, "Investigation of ecologically relevant wind patterns on marion island using computational fluid dynamics and measured data," *Ecological Modelling*, vol. 464, p. 109 827, 2022.

[32] J. A. Ludwig, D. J. Tongway, and S. G. Marsden, "Stripes, strands or stipples: Modelling the influence of three landscape banding patterns on resource capture and productivity in semi-arid woodlands, australia," *Catena*, vol. 37, no. 1-2, pp. 257–273, 1999.

[33] M. Rietkerk *et al.*, "Evasion of tipping in complex systems through spatial pattern formation," *Science*, vol. 374, no. 6564, eabj0359, 2021.

[34] M. Momberg *et al.*, "Factors determining nest-site selection of surface-nesting seabirds: A case study on the world's largest pelagic bird, the wandering albatross (diomedea exulans)," *Ibis*, vol. 165, no. 1, pp. 190–203, 2023.

[35] M. Momberg, D. W. Hedding, M. Luoto, and P. C. le Roux, "Exposing wind stress as a driver of fine-scale variation in plant communities," *Journal of Ecology*, vol. 109, no. 5, pp. 2121–2136, 2021.

[36] K. V. Mardia and P. E. Jupp, *Directional statistics*. John Wiley & Sons, 2009.

[37] S. R. Jammalamadaka and A. SenGupta, *Topics in circular statistics*. World Scientific, 2001.

[38] C. Ley and T. Verdebout, *Modern Directional Statistics*. Chapman and Hall, 2020.

[39] R. B. Nelsen, *An introduction to copulas*. Springer Science & Business Media, 2007.

[40] H. Joe, *Dependence modeling with copulas*. CRC press, 2014.

[41] L. J. Bain and M. Engelhardt, *Introduction to probability and mathematical statistics*. Duxbury Press Belmont, CA, 1992, vol. 4.

[42] E. García-Portugués, R. M. Crujeiras, and W. González-Manteiga, "Exploring wind direction and so2 concentration by circular- linear density estimation," *Stochastic Environmental Research and Risk Assessment*, pp. 1055–1067, 27 2013.

[43] M. Jones, A. Pewsey, and S. Kato, "On a class of circulas: Copulas for circular distributions," *Annals of the Institute of Statistical Mathematics*, vol. 67, pp. 843–862, 2015.

[44] G. J. McLachlan, S. X. Lee, and S. I. Rathnayake, "Finite mixture models," *Annual Review of Statistics and its Application*, vol. 6, no. 1, pp. 355–378, 2019.

[45] M. Wedel, "Concomitant variables in finite mixture models," *Statistica Neerlandica*, vol. 56, no. 3, pp. 362–375, 2002.

[46] C. M. Dayton and G. B. Macready, "Concomitant-variable latent-class models," *Journal of the American Statistical Association*, vol. 83, no. 401, pp. 173–178, 1988.

[47] A. K. Formann, "Linear logistic latent class analysis for polytomous data," *Journal of the American Statistical Association*, vol. 87, no. 418, pp. 476–486, 1992.

[48] A. J. Dobson and A. G. Barnett, *An introduction to generalized linear models*. Chapman and Hall/CRC, 2018.

[49] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[50] E. A. Mohammed, C. Naugler, and B. H. Far, "Emerging business intelligence framework for a clinical laboratory through big data analytics," *Emerging Trends in Computational Biology, Bioinformatics, and Systems Biology: Algorithms and Software Tools. New York: Elsevier/Morgan Kaufmann*, pp. 577–602, 2015.

[51] J. Grace, "Wind damage to vegetation," *Commentaries in Plant Science*, vol. 17, pp. 209–20, 1976.

[52] M. Momberg, D. W. Hedding, M. Luoto, and P. C. le Roux, "Species differ in their responses to wind: The underexplored link between species fine-scale occurrences and variation in wind stress," *Journal of Vegetation Science*, vol. 32, no. 6, e13093, 2021.

[53] N. B. Fitzgerald and J. B. Kirkpatrick, "Wind distortion in alpine and subantarctic plants is constant among life forms but does not necessarily reflect prevailing wind direction," *Arctic, Antarctic, and Alpine Research*, vol. 49, no. 4, pp. 521–535, 2017.

Appendix A

Code and additional Marion Island studies

All code and data used during this study can be found in the following GitHub repository: https://github.com/Chip0712/Marion-Island-Project

TABLE A.1: Literature review of recent studies part 1.

Title of article	**Goal of research**	**Methodology**	**Conclusion/Results**
Spatial variation in plant interactions across a severity gradient in the sub-Antartic [4]	To test the stress gradient hypothesis in a climatically extreme sub-Antarctic location with a low variety in plant species, ie. Marion Island.	Generalised linear models were used to analyse environmental and vegetation data, with error distributions and link functions selected based on response variable properties. Species richness data were examined with a Poisson error distribution and logit link function, while microclimate and species cover data were analysed using a normal distribution. Spatial Analysis by Distance Indices (SADIE) and L(t) function indices were modelled accordingly. Log-likelihood tests assessed relationships between SADIE indices and explanatory variables. Ordinal logistic regression was used for L(t) function indices, with corrections for over and underdispersion. Stepwise model building and false discovery rate adjustment were used to control errors in the analyses.	The study on Marion Island's scoria cone revealed a clear altitudinal severity gradient, with stronger winds and lower temperatures at higher altitudes. Stronger interspecific spatial associations at higher altitudes support the stress gradient hypothesis, indicating greater facilitation relative to competition. Directional spatial patterns suggest protection from winds and substrate stability as primary mechanisms, with intense positive interactions at higher altitudes. This supports the stress gradient hypothesis in the sub-Antarctic, highlighting spatial patterns among plant species within severity gradients.
Wind controlled linear patterning and cyclic succession in Tasmanian Sphagnum mires [6]	To investigate the characteristics and causes of striped patterning in minerotrophic Sphagnum cristatum mires on the western Central Plateau in Tasmania.	The study used GIS software, specifically ArcMap 9.2, for spatial analysis, utilising scanned and aerial photographs to map striped mires and measure stripe orientations and lengths. Fieldwork involved detailed surveys within selected mire areas, including transects for plant species cover estimation, height measurements, and peat depth analysis. Laboratory analyses included core extraction for peat layer examination and macrofossil identification. Anemometers were deployed to capture wind speed and direction, and statistical analyses, such as ANOVA tests, were performed using Minitab Statistical Software to test hypotheses regarding stripe orientation, ridge dimensions, peat depth, plant species distribution, and wind speed variations.	The study identified consistent stripe orientations independent of slope aspect, with Richea scoparia dominant on north-west-facing slopes and Sphagnum cristatum on south-east-facing slopes, while Baloskion australe prevailed in swales. The research noted increased exposure of striped mires to both north-westerly and south-westerly winds. Strong north-westerly winds formed rotors behind ridges, affecting R. scoparia growth inversely, while south-westerly winds influenced asymmetry in R. scoparia on ridges. The study inferred cyclic ridge migration due to differential growth rates of R. scoparia in response to wind speed variation, highlighting the unique wind-controlled patterning and cyclic succession in these mires.

TABLE A.2: Literature review of recent studies part 2.

Title of article	Goal of research	Methodology	Conclusion/Results
Aeolian processes and landforms in the sub-Antarctic: preliminary observations from Marion Island [30]	To investigate aeolian processes and their geomorphic impacts on Marion Island, Additionally, it evaluates the potential impact of climate change on aeolian processes and the dispersal of vegetation on the island.	The study focused on three sites on Marion Island displaying evidence of aeolian processes, conducting thorough ground climate and sediment flux measurements in April/May 2013. It utilised sensors and data loggers to monitor meteorological conditions, wind speed, wind direction, and rainfall, while employing Big Spring Number Eight sediment traps and surface sediment samplers to assess aeolian sediment flux. Additionally, texture analysis, orientation measurement of ripples, and transect measurements were employed to delve into landform morphology and sediment characteristics.	The study identifies aeolian landforms like mega-ripples and wind-modified terraces on Marion Island, noting a relatively low annual aeolian sediment flux due to consistently wet surfaces and limited sediment availability. It highlights the wind's role in dispersing plant material, emphasising the need for further research on synoptic climate-wind relationships and long-term monitoring of aeolian processes amidst climate change.
Wind distortion in alpine and subantarctic plants is consistent among life forms but not necessarily reflect prevailing wind direction [53]	Investigate if non-woody plants display consistent prevailing wind deformation similar to woody plants in windy environments.	The study investigated wind distortion effects on vegetation in alpine Tasmania and sub-antarctic Macquarie Island. Vertical photography was conducted at four locations, with Mount Wellington monitored monthly for over a year. Various plant forms were observed for wind-induced distortion in quadrants of varying sizes. Statistical analysis included circular statistics, ANOVA, paired t-tests, and chi-squared tests with Monte Carlo simulation. Binary logistic regression was used to assess the relationship between quadrant aspects and plant distortion presence.	The study conducted in maritime alpine and subantarctic environments shows that wind distortion in vegetation can be assessed using various plant types. Contrary to expectations, the dominant wind direction inferred from plant distortion doesn't always align with the strongest winds. Instead, vegetation responds to winds causing the most damage to foliage, varying by environmental context. This highlights the complex relationship between wind patterns and vegetation responses.
Species differ in their responses to wind the under-explored link between species fine-scale occurrences and variation in wind stress [52]	To evaluate the impact of wind stress on species distribution models of vascular plants on Marion Island. Furthermore, the study seeks to determine whether including wind stress improves predictions of species occurrence and cover compared to traditional variables. Additionally, the study aims to understand how different plant species respond to wind stress and highlight the importance of incorporating wind metrics into distribution models, particularly given the changing wind patterns.	Using data obtained from 1440 heterogeneous quadrants (in terms of topology, geology and biology), generalised linear models/generalised additive models/boosted regression trees were used to model the occurrence and cover of species. Variable importance scores were calculated by taking the mean Pearson correlation coefficient which was obtained by randomly ordering the predictor variables and calculating the Pearson correlation between the predictor variables 10 times.	The research findings demonstrate the significant impact of wind stress on species occurrence and cover for vascular plants on Marion Island. Wind stress emerges as a key predictor alongside other factors, remaining influential even after considering other variables. Notably, wind stress outperforms temperature and precipitation in predicting occurrence and cover for certain species, particularly those adapted to open, wet environments and pteridophytes. This highlights the importance of incorporating wind metrics into distribution models, enriching our understanding of Marion Island's ecological dynamics amidst changing wind patterns.

TABLE A.3: Literature review of recent studies part 3.

Title of article	**Goal of research**	**Methodology**	**Conclusion/Results**
Investigation of ecologically relevant wind patterns on Marion Island using Computational Fluid Dynamics and measured data [31]	Estimate high-resolution wind speed, direction, and turbulence map for Marion Island (MI) using CFD for the study of terrestrial ecology.	Utilising CFD, wind properties were estimated using a Reynolds Averaged Navier–Stokes approach with a κ-ϵ turbulence closure model in ANSYS FLUENT 2019R3 software. Validation involved comparing experimental measurements with numerical simulation results for full-scale atmospheric flow around Marion Island (MI) and Prince Edward Island (PEI), validated against two years of wind measurements.	The study established a high-resolution wind map for Marion Island using CFD, providing reliable estimates of wind speed, direction, and turbulence. Despite moderate errors, the CFD model demonstrated mean absolute errors ranging from 14.8% to 32.6% in predicting wind properties. This research facilitates future studies on Marion Island to assess wind-related factors crucial for understanding terrestrial ecology without invasive surveillance methods. Additionally, it underscores the potential of CFD modelling in estimating spatial variation in wind parameters at biologically-relevant scales on other islands, advancing island ecology research methodologies.
Factors determining nest-site selection of surface-nesting seabirds: a case study on the world's largest pelagic bird, the Wandering Albatross (Diomedea exulans) [34]	Investigate environmental variables (such as topographical variables, vegetation, wind characteristics, and geological characteristics) affecting nest-site selection by Wandering Albatrosses on Marion Island	A cross-check of results from a generalised additive model and a generalised linear model was used to determine which variables are the most important factors regarding the location of the presence/absence of albatross nests. Highly correlated variables were determined and removed based on the Variance Inflation Factor. Wind velocity, turbulence, vegetation type, elevation, geology, terrain roughness, and distance from the coast were investigated as predictors. (All of the wind data was obtained using the same CFD that we will be using in this paper)	The study revealed elevation as the primary factor influencing nest-site selection for Wandering Albatrosses on Marion Island, with nests predominantly located at lower elevations. Proximity to the coast and terrain roughness also played significant roles, while wind velocity showed a subtle relationship with nest occurrence. Coastal vegetation types were favored for nesting, contrasting with the absence of nests in polar desert vegetation areas above a certain elevation. These findings highlight the vulnerability of nesting habitats to climate change, particularly regarding vegetation type and wind characteristics. Overall, the study sheds light on the intricate interactions of environmental factors, such as wind, in shaping the nest-site selection of surface-nesting seabirds, providing crucial insights for conservation efforts.