UNIVERSITY OF CAPE TOWN

A historical perspective on wind data:

Time, space and vector relationships between ship log data and Cape Royal Astronomical Observatory wind data between 1834 and 1854.

Alexa Brown (BRWALE006)

Supervisor: Chris Lennard (UCT) | Co-supervisor: Stefan Grab (Wits)

MSc minor dissertation

ACDI Climate Change and Sustainable Development

Department of Environmental and Geographical Science, University of Cape Town

August 2016

Artist's impression of an East Indiaman sailing vessel being wrecked by a "hurricane" off Cape Point in April 1848.

Image taken from http://www.19thcenturyshipportraitsinprints.com/. Artist: T. G. Dutton, coloured lithograph (circa 1848), Royal Museums Greenwich, London (0601).

Key words: historical climatology, wind data, ship log data, Cape Town Royal Astronomical Observatory

Abstract

This dissertation assesses the extent to which data from the Climatological Database for the World's Oceans (CLIWOC) reflect newly digitized historical wind data captured at the Royal Astronomical Observatory (RAO) in Cape Town, South Africa from 1834-1854. This follows the historical precipitation reconstructions for Southern Africa by Hannaford *et al*. (2015), using wind data from the CLIWOC database. This project also forms part of a bigger project that is recovering and digitising historical instrumental meteorological data for Southern Africa that have never been analysed before. For Southern Africa, the opportunity to compare historical instrumental data seldom arises due to the paucity of reliable data. However, there is an opportunity to analyse and compare two different wind data sources for a twenty-one year cross over period for south western Africa. Wind, as an indicator of atmospheric conditions, has not been assessed extensively in South African, therefore this project fills an academic gap in historical climatology for the region, and provides newly digitised historical data.

Digitisation and pre-processing steps ensure that the RAO dataset is comparable to the CLIWOC dataset. This is done by replicating wind direction and speed measurement conversions and formatting (Garcia-Herrera *et al.*, 2005), and by mirroring the available time steps of data in each dataset (eliminating data were the other dataset has erroneous or missing data). Spatially scattered data recorded over the sea compared to data recorded at a fixed position introduces inherent limitations, error and noise into the data comparison. Therefore, to eliminate as many uncertainties as possible and minimise the noise in the data, the CLIWOC data are refined further by a) a single observation per day, b) separating three regions of differing seasonal synoptic air flow regimes (west coast, south west peninsula and south coast) and c) all analyses based on seasonally grouped data.

Temporal, spatial and vector relationships are established for each season using scatter plot graphs and Pearson correlations. The different relationships between the data are derived from corresponding wind data (i.e. data of the same day and time), in each dataset for wind speed and wind direction separately. No significant correlation (all p values>0.05) or signal is evident over time, or as the difference in distance changes. However, seasonality is represented consistently in the wind vector distribution heat maps. Significant findings include the observations of anomalous north westerly winds in summer at the RAO, where the CLIWOC data did not pick up similar data for the corresponding region on the west coast.

Historical wind data used herein prove to be reliable by the expected seasonal synoptic flow patterns and characteristics seen in each study region. There is no correlation between the datasets over time and space and the data do not present any clear signals or return events over time. Although corresponding data do not show any correlations, there are typical synoptic flow regimes in each study region which prove that wind data was recorded correctly. Therefore, the datasets are mutually exclusive, but accurate in their intrinsic value. It is only the anomalous summer north westerlies at the RAO which question the reliability of the data, as the same wind regimes were not identifiable in the corresponding CLIWOC data. This anomaly was noted but not studied further. This project highlights the major inconsistencies and limitations in the CLIWOC data. Researchers in the future should use CLIWOC data appropriately to suit the research question and be aware of the inconsistencies that may introduce noise.

Acknowledgements

This work would not be possible without funding through a grant holder bursary from the National Research Fund and the Climate System Analysis Group at UCT.

I thank my friend Michael Hayton for helping me get to Cape Town. Without his humble support, I would not have been able to complete this MSc degree at the University of Cape Town.

Tammy and Wynand Smythe, without your endless support and guidance I would not have been able to complete this MSc degree.

I thank Stefan Grab who has supervised my work for three years, supported me through two international conferences and been so motivating as an academic supervisor. Chris Lennard, thank you for working patiently with me over two years to finish this short paper.

Thank you to the few who have read over my words and seen the mistakes my eyes were blind to.

Acronyms and abbreviations

amsl – above mean sea level

CLIWOC - Climatological Database for the World's Oceans

COP - Conference of the Parties

DJF – December, January and February

IPCC – Intergovernmental Panel on Climate Change

GFS – Global Forecasting System

ITCZ – Inter-Tropical Convergence Zone

JJA – June, July and August

Lat – Latitude

Lon - Longitude

m/s or ms^{-2} – metres per second

MAM – March, April and May

NW'lies – north westerlies

SAHP – South Atlantic High Pressure cell

SAWS – South African Weather Service

S'ly - southerly

SE'lies – south easterlies

SON – September, October and November

SL – Ship log(s)

RAO/ RO – Royal Astronomical Observatory/ Royal Observatory

UNFCCC – United Nations Framework Convention on Climate Change

W'ly – westerly

WASA – Wind Atlas for South Africa

WS – wind speed

WD – wind direction

Table of contents

Contents

List of figures

List of tables

List of equations

Introduction

Researching historical climatology

Historical climatology is a multidisciplinary field of quantitative and qualitative analyses that provide a well-rounded understanding of the state of the past climate and its transformation over time. Knowledge of an historical climate provides a reference base from which to understand how climates have changed over periods in the order of 100s of years. This subsequently allows long term cyclic events and signals to be identified. This knowledge can be reflected onto what we know about contemporary climate phenomena, events and cycles today. Our understanding of the complex and dynamic atmospheric and climate systems, that influence human and biological development, relies on an understanding of what has happened in the past. International pressure to produce reliable and robust information on climate variability and 'climate change' is rising due to the influence of the United Nations Framework Convention on Climate Change (UNFCCC) meetings (Conference of the Parties), and supplementary scientific documents such as the Intergovernmental Panel on Climate Change Fifth Assessment Report (IPCC AR5) (see https://unfccc.int/2860.php for more information about the UNFCCC and https://ipcc.ch/report/ar5/ for access to the working group contributions of the IPCC). Regional decision makers also require accurate climate change information to ensure the physical environment and expanding populations are not vulnerable to surmounting risk and harmful impacts associated with climate changes. Historical climatology supports climate sciences and research by allowing for data recovery and climate reconstructions based on observations that uncover climate changes over long periods of time and over wide regions. Historical climatology is like a telescope that sees climates far in the past; and zooming out allows for trend analysis over variable temporal and spatial scales until the present. It is the prequel to what we know now about contemporary climates around the world today.

The IPCC AR5 concludes that Southern Africa is highly vulnerable to risks of drought (a 10% or 1 – 4,5ºC increasing trend) and decreasing rainfall (30% decreasing trends) by 2100. This will exacerbate the imminent harmful impacts such as terrestrial and marine ecosystem changes and increase the challenges to economic growth and sustainable development because of harmful impacts on health, food security and fresh water availability (Niang *et al.*, 2014; Climate and Development Knowledge Network, 2014; Climate System Analysis Group [CSAG], n.d.). Downscaled and regional climate models predict that Southern Africa is facing temperature increases of up to 4ºC in some regions by 2100 (CSAG, n.d.). The knock on effects of this temperature increase will be disastrous for fragile ecosystems and vulnerable communities. There is scope then, to do as much research as possible to understand the climate system and feedbacks so that potential threats and impacts can be avoided or alleviated appropriately. All fields of research that supplement atmospheric or climate sciences (such as climate modelling, meteorology, cloud physics, paleoclimatology and historical climatology) increase the adaptive capacity of a focus region. For Southern Africa, a region of developing nations this is integral for achieving sustainable development goals (United Nations Sustainable Development Goals, 2015).

Long term atmospheric observations in Southern Africa

Southern Africa is characterised by a general paucity of long term systematically recorded weather data. Very little was written on meteorology during the early years of European settlement (Vogel, 1989 and Forbes, 1977). Without a dense network of meteorological observations and sites, it is difficult to reconstruct a climate and analyse the data signals, correlations and variability over time. Critical understanding of long-term and low-amplitude controls of Southern African climate is then impeded. Therefore, in conjunction with various other international efforts there is a need to recover long term historical data from all available sources in Southern Africa and develop databases compatible with other international databases (Mitchell and Jones, 2005).

Many accounts of the weather exist in mariners' log books, notebooks, diaries, letters and newspapers which offer insight into the past weather around Southern Africa dating back to the 17th century (Hannaford *et al.*, 2015; Nicholson, 2001; Thom and Balkema, 1952). These documents are seldom standardised or homogenous, the descriptive language is subjective and often document a few months or a couple of years at a time per resource. Data are thus difficult to analyse because of the inherent heterogeneities and short time scales. These inherent heterogeneities are caused by instrumental failure, human inaccuracy and observation error. In contrast, the Royal Astronomical Observatory (RAO) of the Cape of Good Hope offers one of very few sources of systematically and meticulously recorded instrumental weather data for South Africa, dating from 1834 until 1976. Digitising these data will provide the longest time series of instrumentally recorded weather data for South Africa. There is a collection of studies on historically reconstructed and modelled climates for temperature and rainfall in Southern Africa, as well as the associations with ocean systems and ENSO events (See for example Hannaford *et al.*, 2015; Neukom *et al.*, 2014; Mackellar *et al.*, 2014; Rouault *et al.*, 2010; Kelso and Vogel, 2007; Nicholson, 2001; Vogel, 1989; Taunton-Clark and Shannon, 1988 and Schumann *et al.*, 1982). Yet wind as an atmospheric variable has not been analysed exclusively or extensively from this data source, or any others.

Analysing wind data

Wind as an important atmospheric circulation indicator, has been less extensively studied compared with temperature and precipitation, especially at long temporal scales. This is because of the paucity of data, the undervaluing of wind data in historical meteorological records and because temperature and rainfall have had more priority and interest in the history of Southern African climatology and social development (Hannaford *et al.*, 2015; Matulla *et al.*, 2008; Vogel, 1989). Interestingly, despite the socio-cultural and economic influence that wind has in the south western regions, for example, the reliance of fishermen on seasonal west coast upwelling for fishing; or the life threatening sea conditions around the rocky coast line, there has been very little uncovered about the historical wind variability for the South West Cape Peninsula region. Additionally, the plethora of research on sea surface temperatures and coastal upwelling is largely unaided by long term wind data (see Rouault, *et al.*, 2010, Taunton-Clark and Shannon, 1988 and Schumann *et al.*, 1982). This exposes a

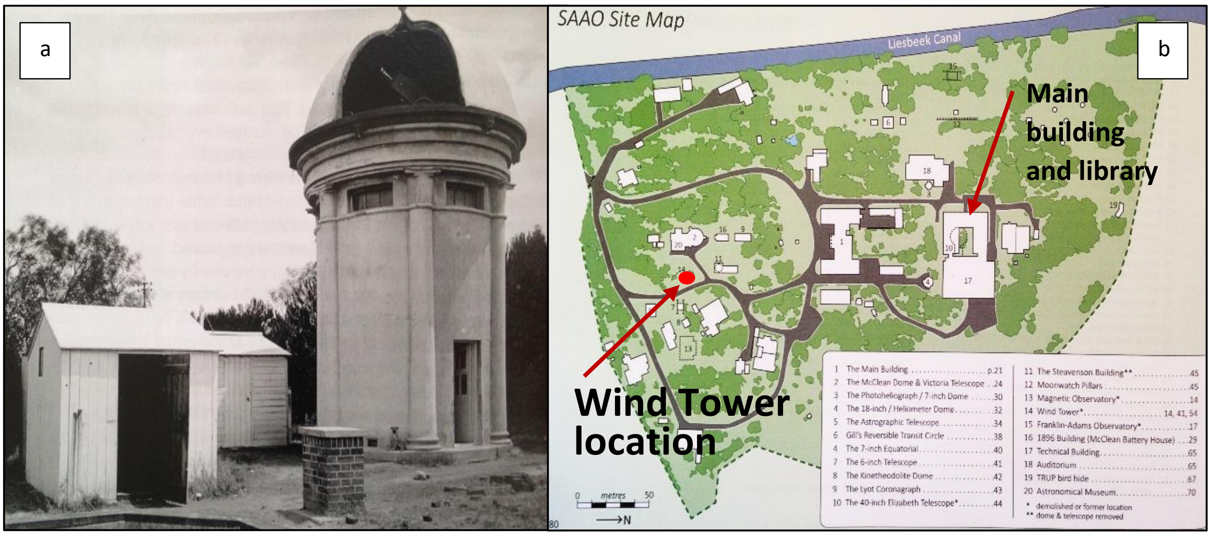


Figure 1 a) Photograph of the original Wind Tower built to house an Osler's Anemometer on the grounds of the RAO. Wind direction observations were made from this tower. The tower was built in 1841 and destroyed in the 1960s. Weather observations continued until 1976; the location of the relocated wind instruments is unknown. b) Map layout of the RAO grounds, including the original site of the Wind Tower where the earliest wind direction recordings were taken from. Photograph and site map taken from I. Glass (2015).

void in our understanding of historical atmospheric circulation controls over Southern Africa, and if there have been any changes over time synonymous with changes in other weather variables. One such source that can provide this long term data are the meteorological notebooks kept at the Cape Royal Astronomical Observatory (RAO) situated in Cape Town, South Africa. Systematic meteorological observations were recorded since 1834 here with the implementation of a meteorological station on the grounds of the RAO (Jackson, 1977) (see Figure 1).

Historical wind data in Southern Africa

The RAO meteorological records provide data recorded four times daily since 1834. The RAO presiding observer and scientist was Irishman Sir Thomas Maclear who took pride in his planetary observations under the auspices of London's Royal Greenwich Observatory and Sir John Herschel (son of Sir William Herschel) (Glass, 2015). Meticulous and systematic accounts of the weather were recorded at the RAO and today provide a daily account of temperature, precipitation, barometric pressure and wind for about 130 years (records phased out with the implementation of modern equipment at the Cape Town International Airport from the 1970s).

Hannaford *et al.* (2015) have statistically reconstructed historical precipitation patterns over Southern Africa. The Climatological Database for the World's Oceans (CLIWOC) meteorological database was created using ship log records that provide wind and barometric pressure data from the English East India Company trading ships (For information on the CLIWOC Project see http://pendientedemigracion.ucm.es/info/cliwoc/ and https://www.knmi.nl/kennis-en-datacentrum/achtergrond/cliwoc. Hannaford *et al.* (2015) were able to statistically superimpose the relationship between wind and precipitation based on contemporary records, onto the wind records of the past. CLIWOC ship log wind data was used for this. Hannaford *et al.* (2015) found wet and dry decadal periods that correspond with documentary reconstructions of El Niño events from the 1810s to the 1830s for Mthatha, Royal National Park and Cape Town. Despite identifying previously unknown precipitation patterns, the integrity of the wind data over such variable temporal and spatial scales was assumed to be reliable enough to infer land based trends. No research or analysis has yet been done to verify the ship log data against land based data. This is because there is no existing digitised wind data to do this. Therefore, a very pertinent opportunity to analyse the ship log records against the RAO records exists.

There are major inconsistencies and limitations in the data sources that impede the ability to rigorously cross check the data for the same time period. For example, a time series cannot be constructed for the ship log data because of the gaps in time between observations. Another major limitation is that the RAO data exists from a fixed point in space, whereas the ship log data are not fixed and are sporadically spaced within a large domain. However, it is still beneficial to compare the data sources to crudely analyse the similarities, differences and independent wind patterns over time. Comparing the data sources will help to (i) understand the quality of data, (ii) variability over time and space and (iii) the relationship between the different datasets. This research paper then seeks to find out the extent to which the ship log records reflect the RAO records. Are the data saying the same thing? This should allow researchers, like Hannaford *et al.* (2015), to better understand the integrity of the historical wind data from Southern Africa in the CLIWOC database and at significant historical data

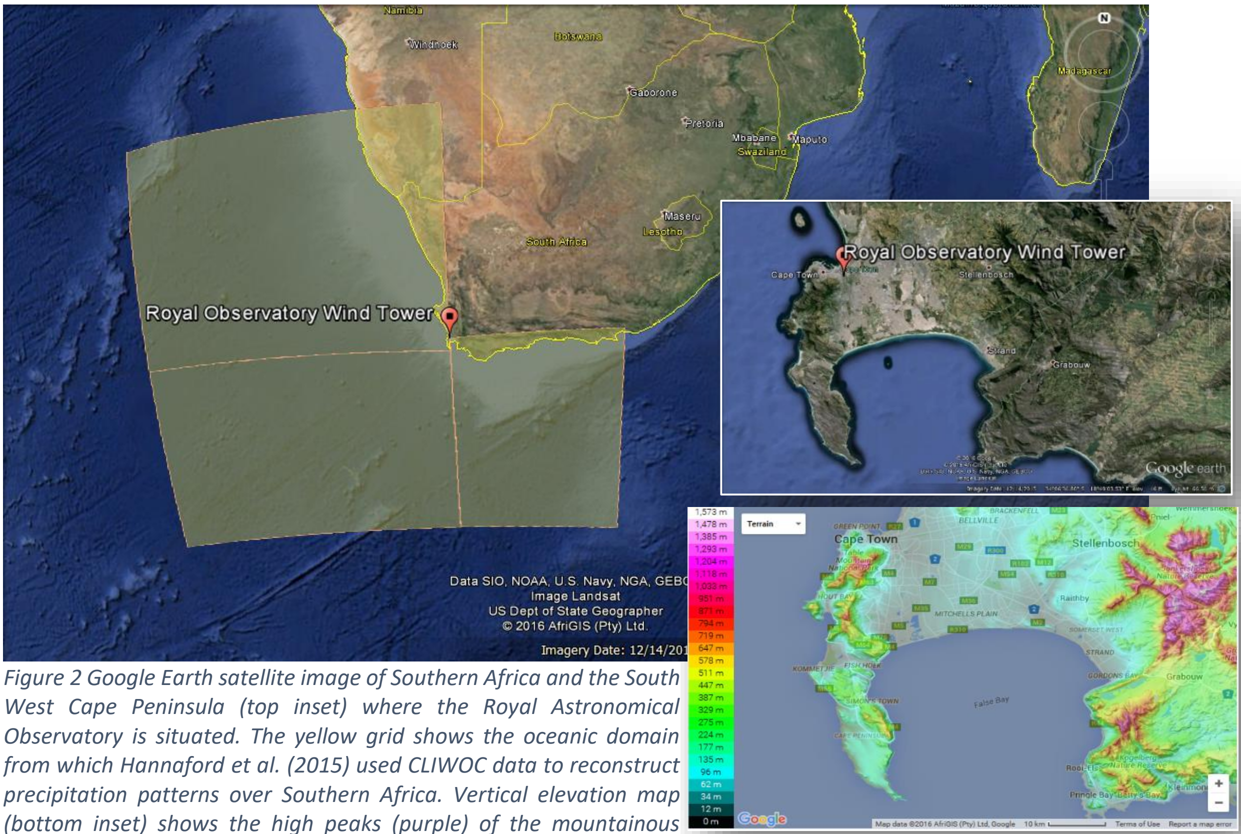


Figure 2 Google Earth satellite image of Southern Africa and the South West Cape Peninsula (top inset) where the Royal Astronomical Observatory is situated. The yellow grid shows the oceanic domain from which Hannaford et al. (2015) used CLIWOC data to reconstruct precipitation patterns over Southern Africa. Vertical elevation map (bottom inset) shows the high peaks (purple) of the mountainous Peninsula and Cape flat areas (light blue). The RAO is at an altitude less than 20meters amsl, but is sheltered from winter westerly winds by the high peaks west of the RAO grounds. Topographic map taken from http://en-gb.topographic-map.com/places/Cape-Town-1310815/ [accessed June 2016].

sources on land. It will also allow a preliminary assessment of the historical wind climatology for Cape Town in South Africa for the period being analysed. This project does not seek to find trends, causation for the signals or comparisons. It is an inquest into the oldest known wind data history for Southern Africa.

Region of study

Geography

The area of study is situated at the south western tip of South Africa, within a domain bound by the coordinates: 26°00'S - 38°00'S, 11°00'E - 26°00'E (Figure 2). The area includes a topographically complex coastal land mass and ocean surface. The mountainous Cape Peninsula is the most extreme South Western Cape of Southern Africa (also known as the Cape of Good Hope). The terrain at the extreme South Western Cape is flat coastal plains bordered by a mountainous peninsula on the west (Cape Peninsula) and a high escarpment boundary to the east, up to 1600 amsl (Lennard, 2014) (Fig. 2). The cold Benguela Current and Benguela Upwelling result in northward flowing cold ocean waters on the western side of the Cape Peninsula and in Table Bay. This cold current along the shore is wind driven. In comparison, the Agulhas Current flows along the southern shoreline of South Africa and brings warmer waters to False Bay on the eastern side of the Cape Peninsula. The intersection of the Benguela Current, Agulhas Current and the Antarctic Circumpolar Current at a retroflection region south of the Cape Peninsula, result in highly variable sea surface temperatures, moisture fluxes and atmospheric conditions (Bonnardot, 2005; Tyson and Preston-Whyte, 2004). The southern hemispheric mid-latitude region of study incorporates the meeting point of different ocean currents and the interface of the Southern Ocean and a topographically complex land mass. The total study region is characteristic of highly dynamic atmospheric systems.

The climate and seasonality at the South Western Cape of South Africa

Lying in the mid latitudes (at 34° south) and on the border of a cold ocean the Cape Peninsula is characterised by a Mediterranean climate. Seasonal weather is controlled by synoptic scale atmospheric dynamics. The oscillation of pressure cells and belts within the sub polar convergence zone drive seasonal climatology. Summer is characterised by hot, dry weather and sustained South Easterly winds (SE'lies) as the stable South Atlantic High Pressure (SAHP) cell shifts southwards over the Atlantic Ocean with the southward shifting of the Inter-Tropical Convergence Zone (ITCZ). Summer SE'lies often reach gale force (+20m/s) for sustained periods (up to 5 days) caused by stable SAHP SE'ly wind deflection over the south west Cape (Figure 3). Winter is characterised by cold, wet weather and infamous sea storm surges due to mid-latitude cyclones that have moved further north with the northwards shifting of the ITCZ (see synoptic chart and wind visualisation in Figure 3). Mid-latitude cyclones that travel from west to east sweep over the Cape bringing cold, wet weather and north easterly wind (NE'lies). Cut off low systems can form on the west coast and travel southwards towards the Cape Peninsula. These cut off low systems are associated with strong winds and stormy weather. Occasionally a section of the SAHP is pushed south of the continental mass to form a ridging high

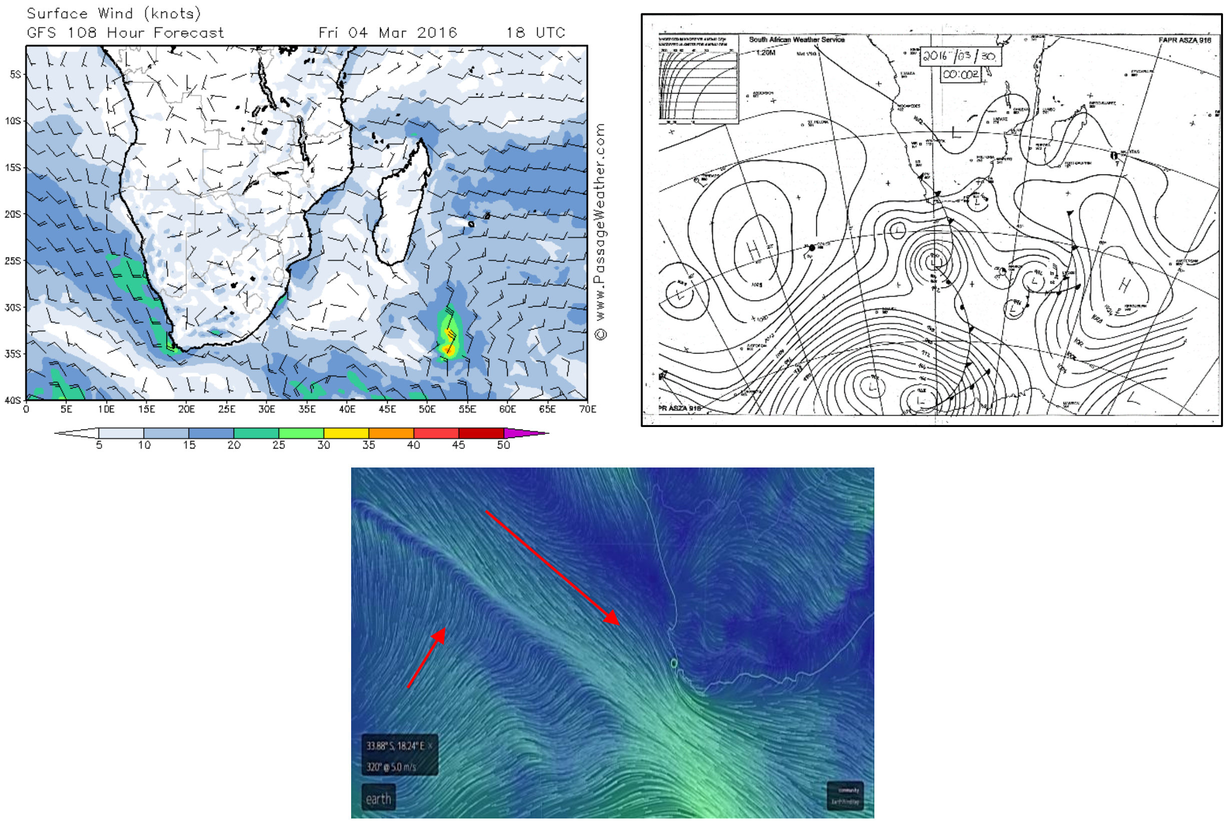


Figure 3 The image of GFS modelled air flow (top left) shows typical southerly summer wind conditions caused by the stable SAHP. The synoptic chart (top right) generated by SAWS shows how mid-latitude cyclones sweep across the SW Cape in a westerly wave during winter months bringing cold, wet weather and NW'ly winds. Opposing winds ahead and behind a mid-latitude frontal system are clear in the (bottom) visualisation of earth's winds which complicate the correlation between winds over time and space (Available at https://earth.nullschool.net/ [accessed 13 August 2016]).

pressure that stimulates easterly flow along the southern coast line (Tyson and Preston-Whyte, 2004). The Cape of Good Hope is well known for the uninterrupted winter sea storm surges that batter the Western Cape coast line, and have caused hundreds of ship wreckages around the peninsula and in Table Bay (Turner, 1988; Jackson 1951). Atmospheric conditions in the study field are therefore highly dynamic and can change hourly.

Hypothesis

The CLIWOC ship log data are reflective of the Royal Astronomical Observatory wind data from 1834 to 1854.

Research questions

1. What is the general wind climatology for the RAO from 1834 to 1854?
2. Are RAO and ship log data observations equivalent for the same day and time?
3. Do ship log and RAO data show a negative correlation with distance between data sources?
4. Do the different data sources capture the same directional vectors and associated wind speeds?

Project aims and objectives

Aim

The aim of this project is to ascertain if the CLIWOC ship log data are reflective of the data captured at the Royal Astronomical Observatory in Cape Town. This aim follows the use of ship log data to infer historical continental climates in a study done by Hannaford *et al.* (2015). This assessment offers an opportunity to measure the extent to which the records are similar or dissimilar, giving insight into the differences between land and sea based wind data.

Objectives

1. Develop a general historical reference climatology for the RAO for 1834 to 1854

    a) Digitise all available RAO wind speed and directional data in the same format as the CLIWOC ship log dataset.

    b) Homogenise the wind speed and direction data (convert to standard unit of measurement).

    c) Calculate the percentage of missing data.

    d) Create seasonal wind roses using all available data.

2. Determine the relationship between data for the same date (data differences as a function of time).

    a) Analyse signals and correlations through the time period being analysed by plotting absolute data values on a time line for both ship logs and RAO data.

    b) Use Pearson correlation to examine if the relationship is statistically significant.

3. Determine the relationship between the data points based on distance away from the RAO (data differences as a function of space).

    a) Calculate the distance between each ship entry and the RAO.
    b) Calculate the difference between the observed recordings for the same time.
    c) Plot the value difference (for wind speed and direction separately) against the difference in distance to assess the data relationship between data for the same date.
    d) Use Pearson correlation to examine if the relationship statistically significant.

4. Flag dominant wind vectors and associated wind speed distributions for the ship log data.

    a) Identify dominant wind directions and associated prevailing wind speed classes using a heat map/ distribution table.
    b) Are these the same as the reference climatology created for the RAO?

The aims and objectives are specifically for assessing if the differences between the data are a function of a) elapsed time or b) distance apart. The wind vector analysis eliminates time and space functions from the differences in the data and allows for signals in wind patterns to be identified for the ship logs and RAO. Thus, despite the imminent time-space issues of simultaneously analysing data, it is possible to identify the time-space and vector relationships between ship log and RAO wind data.

Research impact and rationale

This research forms part of a broader effort to analyse long term historical climate records that have not been digitised or analysed before in Southern Africa. Little has been written on the topic of data historiography, value quantification and/or the development of scientific endeavour in Southern Africa. The nineteenth century played a role in the development of modern day measurement systems and the establishment of scientific and learning centres in Southern Africa. The likes of Brown (1977) and Warner (1978) have made inquests into the historiography of meteorological observation at the Cape by documenting methods of meteorological observation and stating the importance of weather recording to facilitate improved nautical navigation, as well as improving scientific discovery in the fields of land surveying, magnetism, geodesy and astronomy (Brown, 1977). Meteorological observation has played a minor role in the scientific and social development of Cape Town as a scientific centre in the southern hemisphere (Brown, 1977).

Understanding the absolute value of historical wind data is beneficial for quantifying long term climate variability and coupled natural feedback systems. Knowledge of the data source location, instrumental/ observational methods of recording and observers' codes improves the quality of the data analysis. The historiography improves the understanding of the extent of error in the data, and subsequently improves the intrinsic value (quality) of the data when it is analysed quantitatively. Along-side a complete time-series analysis of wind data from the RAO (see Picas, 2015), this study shall expose further the value of the historical data and its connectedness to the only other known historical data from the same time and region, CLIWOC ship log data.

Literature review

Using historical weather records to reconstruct and analyse past climates is a valuable exercise in determining the past climatology and understanding the atmospheric dynamics in a localised context. Added to which, there is a demand for observational data given the rising pressure to produce accurate information on climate change at fine scale resolutions. Historical systematically recorded climate data facilitates our understanding of climate physics for a given area or period and can be used to improve the quality of model representations of observed data or predicted data. All of which is necessary for decision making processes, in all sectors of society that are influenced by weather and climate (Trenberth *et al.*, 2002). This makes long time scale datasets invaluable and motivates the recovery of old data from all possible sources. Climate science is to an extent, dependent on historical records that extend our climate observation networks back in time and thus give us an improved understanding of climate variability over space and time (Brohan *et al.*, 2009).

Using historical observations has enabled climate scientists to analyse a multitude of phenomena/events that relate to climate and weather variability, and vice versa. Historical climate analyses are relevant for long term trend analyses (See for example Luterbacher *et al.*, 2002 and Zhang and Crowley, 1989), as well as analysing significant extreme events (See for example Lennard, 2014; Matulla *et al.*, 2008 and Jonsson and Holmquist, 1995). Historical data can assist in multi-variate climate regime classification, such as Jimenez *et al.* (2009) demonstrates by classifying regional wind regimes within the parameters of large scale atmospheric processes. Jones and Salmon (2005) have used historical observational data to reconstruct large scale atmospheric processes, namely the North Atlantic Oscillation and the Southern Oscillation Index, which improves our knowledge of large scale, low-amplitude processes (Jones and Salmon, 2005). There is interest in the role of historical observational data in datasets such as the Twentieth Century Reanalysis project. The assimilation of wide spread observational data that simulate global tropospheric variability through time and space, act as a reference dataset and as model validation (Compo *et al.*, 2011). Historical climate data uses are diverse and useful in building our knowledge on systems based approaches to climate sciences, particularly when multiple weather variables are included (Trenberth *et al.*, 2002).

In South Africa, there is a serious paucity of data on temporal and spatial scales. Prior to organised nationwide systematic instrumental records in the early 1900s, accounts of weather exist in documentary data that span a few weeks to a few years (Jackson, 1977). Sources such as newspapers, missionaries' journals and letters have thus been the most useful in creating climate chronologies reaching as far back as the 1800s for South Africa (Vogel, 1989; Nicholson, 1981). But alas, these are inconsistent time series datasets that create chronologies of snapshot time periods at large spatial scales, and use predominantly only rainfall and temperature data. Therefore, in order to improve the climatological basis in South Africa, and in the wake of climate change science, there is a need to develop and improve historical climate reconstructions by filling in and extending time periods by using all available sources, especially the longest consistent meteorological records that originate from the Cape Colony in the mid-1800s.

The importance of the "Cape of Storms" as an historical data site

At a time when the English East Indian Company trade routes controlled much seafaring and exploration in the 1700s to the mid-1850s from the North Atlantic to the Indian Ocean, an abundance of logbooks were kept. Information of routes, weather, crew, food and vessel maintenance provides a rich source of data and information on the socio-cultural context of human movement; but also of the weather at the time. Cape Town at the Cape of Good Hope as a strategic port along the route, became very important as a terrestrial control point and centre of knowledge in the early Victorian age (Warner, 1978). The Royal Observatory was established in 1821, which led to major scientific work in astronomy, magnetism and meteorology, recognised and visited by notable scientists such as Sir John Herschel and Sir David Gill (Stoy, 1977). Cape Town's maritime history is an extension of the history of Greenwich Maritime Observatory and British Board of Trade. By extension, the Cape Observatory was a link between science, maritime culture and earth observations for the advancement for marine, terrestrial and astronomical exploration, all of which influenced the other between Cape Town and England since the early 1800s (Brown, 1977).

Cape Town also boasts the longest known continuous record of meteorological observations for the Southern African region. Daily records from the Royal Observatory extend back to 1834. Besides the rainfall and temperature variables, the documents have recorded barometric pressure and wind variables in great detail (often supplemented by written comments on the state of the atmosphere and cloud cover). Barometric pressure is still not digitized and the information about climate variability that can be extracted from the data are largely unknown. The full time series of wind data from the RAO have recently been digitised by Picas (2015).

Today, climate variability in Cape Town is closely monitored for sustainable development, tourism and economic forecasting by using contemporary records (see Klopper *et al.*, 2006, Reason *et al.*, 2011; Schumann *et al.*, 1982). Investigation into seasonal forecasting and periods of drought and flooding still dominate the literature (For example, Neukom, *et al.* 2014; Mackellar *et al.*, 2013; Tadross and Johnston, 2012; Rouault *et al.*, 2010; Thomas *et al.*, 2007 and Reason *et al.*, 2006). Thus, considering the long historical documented connection between Cape Town and climate variability, and the increasing importance of weather observational networks, there is scope for digitizing and analysing the historical wind data and developing a better understanding of climatology for the area.

Focusing on wind data

Cape Town's windy history goes back as far as recalled by historical documents (see image on the cover page as an example) and oral stories. Violent wind driven sea storm surges have caused havoc on the shorelines and in the harbours where ships docked before rounding the Cape Peninsula (Theron, 2011; Turner, 1988 and Forbes, 1977). Seasonal dominant wind patterns characterise the South West Cape windy climate with a hot summer South Easterly and a cold, rainy winter North Westerly. The prevailing winds are driven by synoptic scale air circulation and perturbations inset in the circumpolar westerlies. These dominant wind patterns give Cape Town its unique Mediterranean Climate (Tyson and Preston-Whyte, 2004). Added to which, several oceanic and atmospheric systems control climate modes that act on varying temporal and spatial scales, such as the Southern Annular Mode and ENSO (Hannaford *et al.*, 2015). By inference, wind is an indicator of atmospheric circulation and is useful in understanding the relationship between atmospheric modes and conditions, and the physical interplay between pressure fields, sea surface temperatures and air temperature and precipitation (Hannaford *et al.*, 2015).

There has been serious focus on global temperature data because of anthropogenic global warming (Brazdil *et al.*, 2005). In South Africa there has been a 'preoccupation' in determining return events in extreme events, specifically drought and flood episodes. Temperature and rainfall data are used for this predominantly because of the lack of available and reliable data (Vogel, 1989). However, with the importance of modelling and regionally down scaling climate systems, as well as analysing coupled ocean-climate interactions and long term climate variability, there is benefit in recovering data of all recorded weather variables across multi-variate scales (Tadross and Johnston, 2012 and Reason and Jagadheesha, 2005). Rummukainnen (2013) explains that with international policymaking pressure following the impacts of global warming and extreme events, there is a need to analyse multiple weather variables that will minimise uncertainty in determining changes in extreme events and return periods. This study will focus on wind data specifically, from the RAO meteorological records as well as ship log records from the CLIWOC database.

Ship log data versus Royal Observatory wind data

In a recent paper by Hannaford *et al.* (2015), seasonal precipitation is reconstructed for four stations in Southern Africa using statistical reconstructions based on the significant correlation between wind and precipitation. The complex methods of inference use reanalysis data and regression analyses to reconstruct early 19th century seasonal precipitation. This complicated method is a result of there being no historical observational or instrumental station data on land. However, the earliest instrumental data from the Royal Astronomical Observatory in Cape Town offers an opportunity to correlate the data within the overlap time period, from 1834 to 1854.

Many studies following the CLIWOC project that digitised thousands of ship log records, have agreed that the records are reliable and viable sources of climate information (Garcia-Herrera *et al.*, 2005; Wheeler, 2005; Wheeler, 2001). The RAO data requires stringent quality control, assessment and conversions. However, likenesses and differences in the data and variability may highlight some interesting information about the climate variability for the cross validation time period. This cross comparison will allow us to assess the extent to which the RAO records reflect the ships' logbooks. It is assumed that the data will show some correlation between data that have smaller spatial ranges, and that the datasets will show the same variability over time. However, caution must be paid to the complexities of topography, instrumental error and observer subjectivity that introduce noise and limitations into the data.

Climate science is developing faster with each year that the global temperature rises. This project is a small part of developing the analytical skills required for analysing historical climate data. Although this project does not indicate a major break-through scientific discovery in climate change sciences, it contributes to initiating a broader scale historical data recovery and analysis. As Trenberth *et al.* (2002) points out, the dynamic and interconnectedness of the climate system requires that all weather variables are analysed for a full scope understanding of any trends and variability over time and space. This project seeks to uncover some unknowns about the longest wind record available in South Africa.

Methodology

The methodology used in this dissertation is largely guided by the limitations in the data and the nature of the different datasets. The RAO data has not been digitised before this dissertation, nor pre-processed or analysed before this study. Therefore, it is important to rigorously inspect the data and standardise it in order to understand the data and modify it appropriately. Only once this has been achieved can a reliable comparison with the CLIWOC data be achieved.

Additionally, because the datasets are not homogenous, a method is developed to compare the data over time and space separately. This allows for the relationship between the data to be compared as (i) a function of time, and (ii) a function of distance. Since synoptic atmospheric conditions over time will inevitably create noise in the data, this method best extracts the extent to which the ship log data are reflective of the CLIWOC ship log data in time and space.

Furthermore, a sound understanding of the distribution of wind vector frequencies in each season will determine if the data are capturing typical weather conditions, based on our understanding of contemporary atmospheric circulation detailed by Tyson and Preston-Whyte (2004). That is to say, despite time and space, if the datasets are capturing ideal seasonal patterns, then the observers' perception of the weather is reliable and the data values can be trustworthy.

The first part in this methodology will explore the two wind datasets and explain how each was pre-processed, and then standardised in order to compare them. The next methodology section will define the steps followed in order to achieve the aims set out for this dissertation.

Exploring the datasets

This project will compare the extent to which the ship log data are reflective of the RAO data over time and space. As the wind data from the RAO has not been exploited in this way before, a preliminary data assessment is necessary to identify data characteristics and quality. This will make it possible to homogenise the exclusive dataset so that a reliable comparison can be made with the least possible error. Therefore, attention is paid to the data in this section. The exploration of the data will be followed by a description of the methods used to carry out the comparison of the datasets.

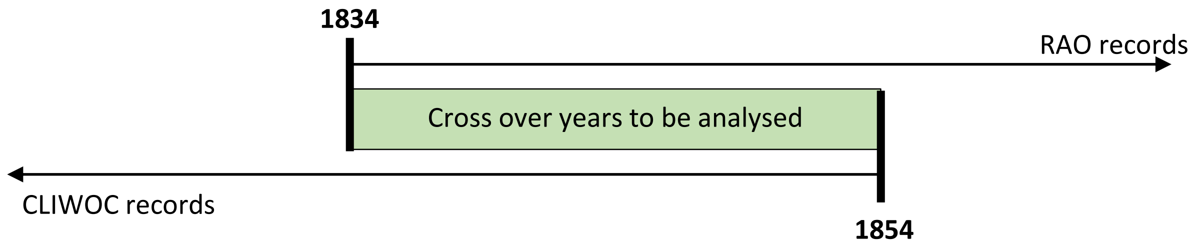


Figure 4 represents the overlapping years in the CLIWOC data and the RAO data, 1834 - 1854.

Two datasets exist in a large geographic area (see figure of grids) that are the only known historical instrumental data sources for the time period being analysed. The CLIWOC data originate from various ships' log books that were recorded in the Atlantic and south Agulhas Oceans surrounding the Cape of Good Hope (moving points). This data was acquired directly from Mathew Hannaford (University of Utrecht), lead author of the report "Early-Nineteenth-Century Southern African Precipitation Reconstructions from Ships' Logbooks" published in *The Holocene*, 2015 (Data are freely available from the CLIWOC home page at http://pendientedemigracion.ucm.es/info/cliwoc/). The other dataset originates from the Royal Astronomical Observatory meteorological registers, situated in Cape Town (a fixed point). Both datasets overlap for a twenty-one year period from 1834 – 1854 (Figure 4). This is the oldest, continuous instrumental data network for the 19th century for Southern Africa covering land and ocean area. The methods used in this research project aim to (i) create a comparable database of two data sources, (ii) perform a preliminary analysis of the RAO data and (iii) assess the extent to which the ship log records reflect the RAO records. The methods described are governed by the limitations and differences in the temporal and spatial characteristics of the data.

Data from the Royal Astronomical Observatory, Cape Town

A description of RAO data

Meteorological observations were recorded at the Cape of Good Hope Royal Astronomical Observatory four times daily in Meteorological Notebooks from 1834 until 1976. Thereafter, meteorological observations were recorded by the South African Weather Service at Cape Town International Airport. Temperature, precipitation, barometric pressure, cloud cover and wind direction and speed were recorded (See Figure 5). For 1834, 1835 and 1836 the readings were taken at "sunrise", "noon", "sunset" and "midnight" (Picas, 2015). 1837 showed no systematic recordings. From 1838 to 1842 the readings were taken at 00h00, 03h20, 09h30, 20h40. From 1843 onwards, recordings were taken at 00h00, 03h30, 09h30, 20h00. Additionally, four "*Term Days*" are included in every year, where the 21st of March, June, September and December had hourly readings, as well as 5 extra hourly readings on either side of the Term Day.

Preceding 1841, visual markers such as trees and chimney smoke plumes were used to indicate wind direction, thereafter an Osler's anemometer instrument was used (Figure 1). The anemometer was mounted atop a 'Wind Tower' situated on the grounds at coordinates

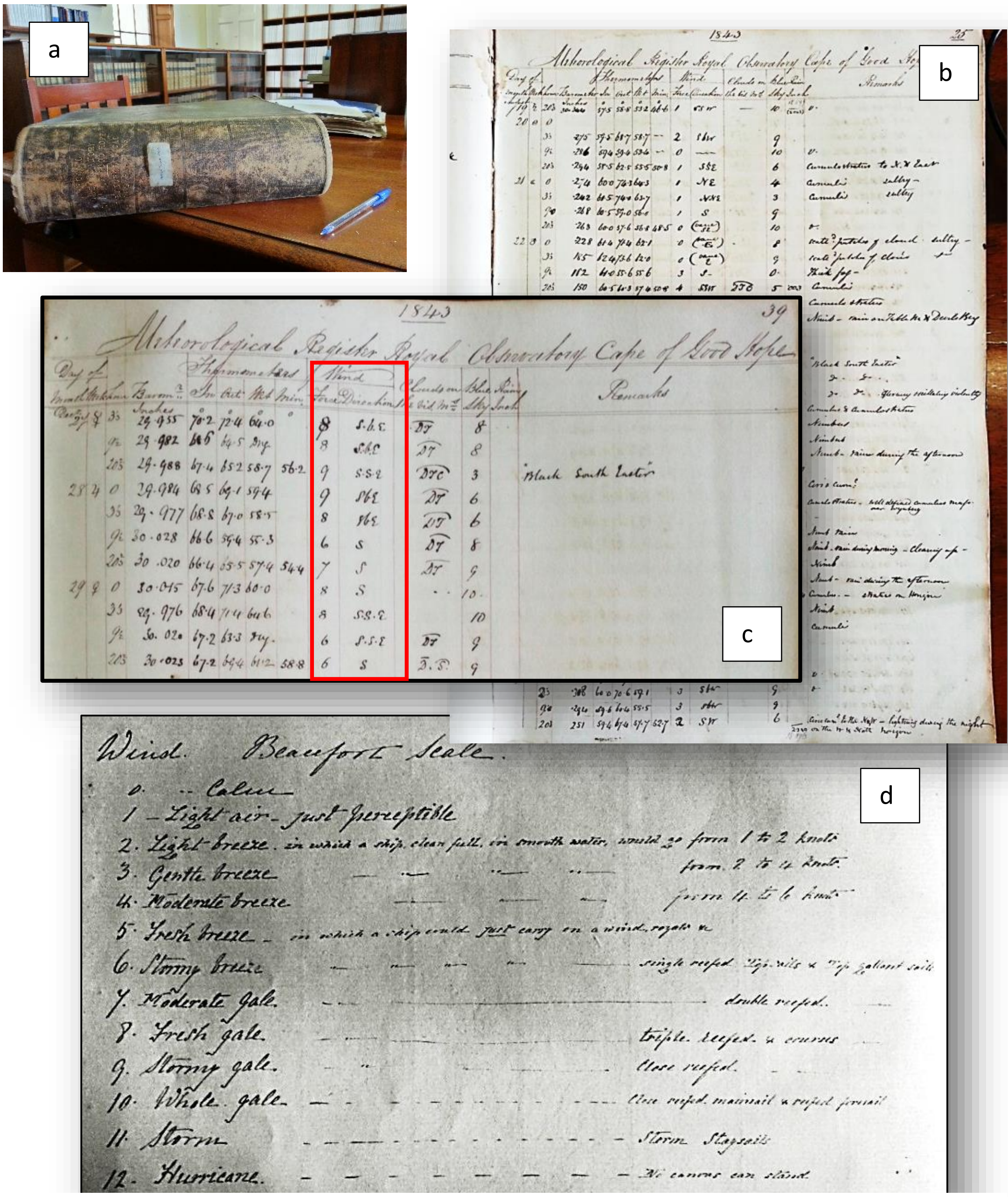
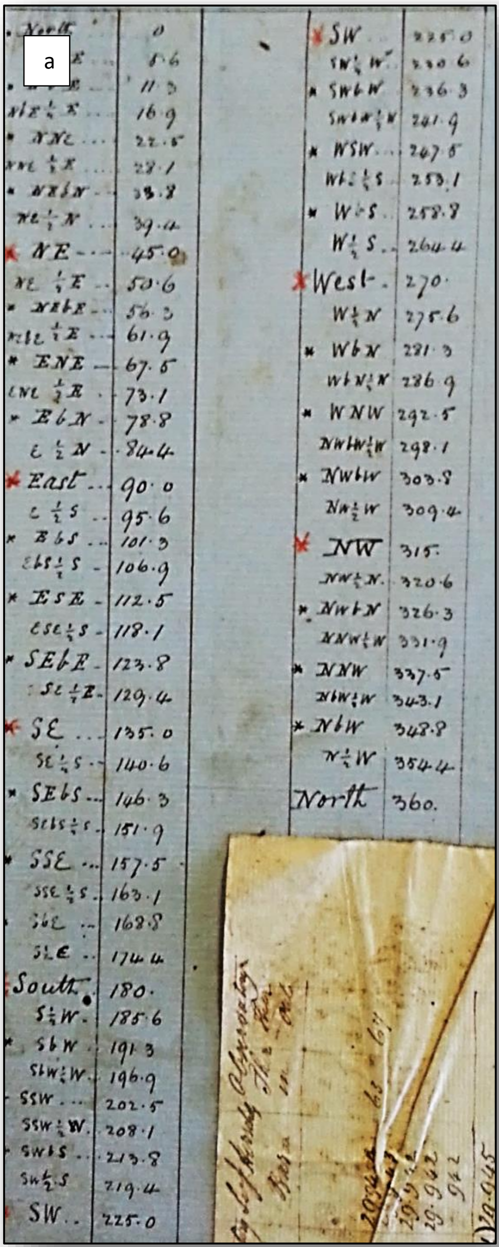
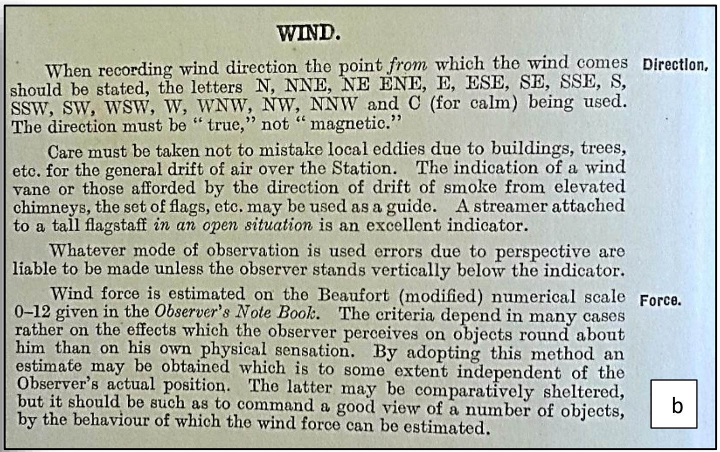


Figure 5a shows a photograph of the original RA0 meteorological register which contains data for 1843-1846. 5b is an example page from inside the register. 5c highlights the "wind" observation column with direction and force data. 5d depicts a hand written note on how to assign the Beaufort Scale codes to the visual interpretations of the wind. This was equivalent to the mariners' instructions. Source: 1854-1858 notebook.

33°56'7.30"S 18°28'39.34"E (see Figure 1) (Sabine, 1851). Wind speed was recorded in the Beaufort Scale where wind strength was assigned a numeric value between 0 (calm) and 12 (Hurricane) according to the observers' visual perception (Table 1 and Figure 5). A standardised code for describing the wind guided the meteorological observer's perception,




WIND.

Direction. When recording wind direction the point *from* which the wind comes should be stated, the letters N, NNE, NE ENE, E, ESE, SE, SSE, S, SSW, SW, WSW, W, WNW, NW, NNW and C (for calm) being used. The direction must be "true," not "magnetic."

Care must be taken not to mistake local eddies due to buildings, trees, etc. for the general drift of air over the Station. The indication of a wind vane or those afforded by the direction of drift of smoke from elevated chimneys, the set of flags, etc. may be used as a guide. A streamer attached to a tall flagstaff *in an open situation* is an excellent indicator.

Whatever mode of observation is used errors due to perspective are liable to be made unless the observer stands vertically below the indicator.

Force. Wind force is estimated on the Beaufort (modified) numerical scale 0–12 given in the *Observer's Note Book*. The criteria depend in many cases rather on the effects which the observer perceives on objects round about him than on his own physical sensation. By adopting this method an estimate may be obtained which is to some extent independent of the Observer's actual position. The latter may be comparatively sheltered, but it should be such as to command a good view of a number of objects, by the behaviour of which the wind force can be estimated.

b

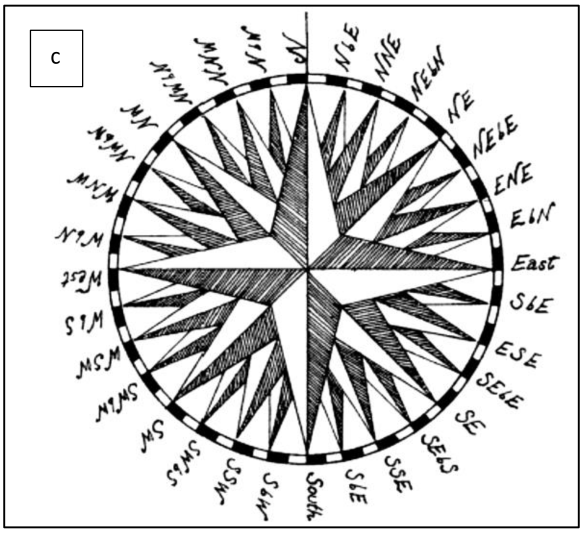


Figure 6a Conversion table used to convert written cardinal letters to decimal degrees (date unknown). 6b A more recent printed note (1933 meteorological notebook) on wind observations. 6c displays a 32-point compass showing directions such as NEbE (north east by east). Figure taken from Wheeler, D. (2005).

as the observer's note on wind in Figure 6 shows. This method and code was identical to the mariners' method of recording wind. The Wind Tower was destroyed in the 1960s, however wind observations continued at the RAO until 1976 (exact location of instruments after the Wind Tower was destroyed is unknown). It is assumed by the comments in the notebooks that all scientific meteorological operations, recordings and instruments were relocated Cape Town International Airport after 1976.

Data recovery and digitisation

The process of capturing the data from the RAO involved photographing every page of the meteorological registers at the South African Astronomical Observatory library in Cape Town from 1834 to the 1960s (the library exists in the original RAO building). Digitization of the data involved manually recording every observation as digital metadata. A date, time and only the wind direction and speed were digitised. This is the first and longest historical continuous chronological record of wind data for Southern Africa.

Data pre-processing and standardisation

In order to standardise the data and make it comparable to the CLIWOC ship log data, RAO digital data was (i) converted to standardised wind measurement conventions and (ii) missing data were accounted for. Wind speed was converted from Beaufort Scale to a middle point number within a range/class of wind speeds in meters per second (Table 1). Wind direction was converted from letters representing compass directions from a 32-point compass to decimal degrees (Figure 6). These conversions and methods of conversions are synonymous with the methods used by the CLIWOC project (Garcia-Herrera *et al*., 2005).

Table 1 Wind velocity classification and conversions from the Beaufort Scale to a metric measurement

Table: Wind velocity classification and conversion table

Beaufort scale	Descriptive term given by Met Office UK	Velocity range: ms^{-2}	**Velocity median: ms^{-2}**
0	calm	0.0 to 0.2	**0.0**
1	light air	0.3 to 1.5	**1.0**
2	light breeze	1.6 to 3.3	**2.6**
3	gentle breeze	3.4 to 5.4	**4.6**
4	moderate breeze	5.5 to 7.9	**6.7**
5	fresh breeze	8.0 to 10.7	**9.3**
6	strong breeze	10.8 to 13.8	**12.3**
7	near gale	13.9 to 17.1	**15.4**
8	gale	17.2 to 20.7	**19.0**
9	severe gale	20.8 to 24.4	**22.6**
10	storm/ whole gale	24.5 to 28.4	**26.8**
11	violent storm	28.5 to 32.6	**30.9**
12	hurricane	32.7+	**35**

Table adapted using the CLIWOC final report available at http://pendientedemigracion.ucm.es/info/cliwoc/ and the Met Office UK Beaufort Wind Force Scale available at http://www.metoffice.gov.uk/weather/marine/guide/beaufortscale.html#moContent.

RAO data reliability and missing data

For the years of 1834 to 1854, 11,14% of the total data are missing from the RAO registers. 66 136 data values are still recorded over the total time period for all observation times of the day (note that occasionally only one wind vector is recorded which is not suitable for high resolution vector analysis and has been ignored in the comparative analysis of this project). When the RAO data are pre-processed to match the CLIWOC ship logs exactly (I.e. Noon readings and matched dates of observation), 7,4% of the RAO data are missing (Table 1). Thus the RAO registers provide adequate data in order to carry out a comparative analysis.

Table 2 Missing data from the RAO dataset

Total missing data from the RAO registers 1834 – 1854	
Entries	33068
Total possible observations	66136
Blank	8307
% missing (= "Blank/Total possible observation*100")	**12,6%**
Total missing data from the RAO registers for matched CLIWOC observation days at Noon	
Entries	2505
Total possible observations	5010
Blank	370
% missing (= "Blank/Total possible observation*100")	**7,4%**

Within the RAO registers, there are variable periods of time where data are missing. Occasionally the reason is unknown/unrecorded. For example, January 1834 has no recorded data for the entire month without explanation. Some of the missing data is attributed to mechanical or human error. Mechanical error refers to problems with the instrument or maintenance repairs. In some cases, as recorded in the registers, issues of vines growing around the anemometer which would have disrupted observations for a time leading up to the point when the observer had noticed the vine. Other cases of mechanical error may be instrumental upgrades or changes (although none occurred during the period analysed in this project). Human error refers to the observers' subjective perceptions of the wind force or direction, and recording values incorrectly. Additionally, the registers refer to periods where the observers took sick leave or attended church on religious holidays, thus neglecting to record observations. As best as is known, during digitization there were no obvious mechanical errors in the period 1834 – 1854.

Data from the Climatological Database for the World's Oceans (CLIWOC)

The CLIWOC project dataset

The first climatological database for the world's oceans has transcribed data from ship log documentary sources found on board ships of various origins (British, Dutch, French and Spanish) that travelled along the English East Indian Company trade route from 1750 to the 1854. The European CLIWOC project aimed to make daily oceanic weather data available for the period 1750 to 1854 for climatological and atmospheric teleconnection analysis (data and reports available at http://pendientedemigracion.ucm.es/info/cliwoc/ and https://www.knmi.nl/kennis-en-datacentrum/achtergrond/cliwoc). Data have had corrections for position, time and measurement and provided in International Maritime Meteorological Archive standard format (Woodruff, 2003). The format provides a date, latitude-longitude position and in situ observed weather variables such as wind direction, wind force, precipitation, fog, ice cover and state of sea and sky. These data are extremely valuable as they cover a very remote area during a period before established meteorological networks existed (Wheeler, 2005). To simplify the issue of recordings at multiple times of the day, only noon recordings were included in the database.

Ship log data at the Cape of Good Hope

The data are meticulously recorded as the ships were reliant on the weather for navigation and survival at sea as well as insurance (Wheeler, 2005 and Hannaford *et al.*, 2015). However, the captured data were recorded sporadically through time and space (further details included in "Limitations"). No complete time series of data exists, but a very rich and dense coverage of the oceans exists with a total of 1 624 log books digitized comprising 273 269 observations (Garcia-Herrera *et al.*, n.d.). Observations in the South Atlantic around the Cape of Good Hope region provide a dense network of 21 533 wind observations after 1800, the period of focus in this study (Figure 10). Hannaford *et al.* (2015) show that around the Cape of Good Hope, there are between 100 to 600 complete wind vector information/observations per year from 1840 to 1854 (that is, a direction and force given for the same time) (see bar graph in figure 10). The Cape of Good Hope, thus has a good source of oceanic historical wind data for the mid-19th century, at the same time that the RAO was recording daily instrumental wind data.

CLIWOC data reliability

Lamb (1982) suggests that despite the inaccuracies inherent in historical and documentary climate data, it is still valuable to correct the error as accurately as possible to analyse the data. The intrinsic value of the data is important enough not to be neglected when investigating historical climatic conditions. Pre-processing and data testing shows

(i) subjective terms used for recorded wind directions around a compass,

(ii) good consistency of 'dual-voyage' observations and

(iii) a high accuracy of Beaufort Scale wind force observations

For more detail consult the CLIWOC Final Report at http://pendientedemigracion.ucm.es/info/cliwoc/Cliwoc_final_report.pdf.

A chi-square statistical test shows that even though customarily a 32-point wind compass was to be used to record wind direction with directions such as *south east by east* (Figure 6c), generally observers used an 8- or 4-point compass for ease of description. This will aggregate directional data into larger sectors, instead of capturing accurate decimal degrees. When 'dual voyages' of two ships sailing the same course at the same time were compared, the wind direction and speeds were similar (with some acceptable deviation) which shows that data was accurately recorded. Lastly the wind forces correlated with a difference of less than one force of the Beaufort Scale which shows that the observers were precise and consistent in their perception of wind force. Thus, CLIWOC data can be reliably used to analyse the climatic conditions because it is consistent throughout time and from ship to ship.

Table 3 A summary of the two historical datasets used for the comparative analysis. The similarities and differences in the data are presented here.

Summary of historical observational data		
Source	**Royal Astronomical Observatory meteorological registers**	**CLIWOC ship log data**
Data acquisition	Digitisation by A Brown and J Picas (Wits University)	CLIWOC dataset available at http://projects.knmi.nl/cliwoc/cliwocdata.htm
Cross over time period	1834 – 1854	
Location	RAO grounds, Cape Town, South Africa (33°56'7,30"S 18°28'39,34"E)	On moving ships within a grid defined in the 'refined region of study' in the oceans surrounding Southern Africa
Periodicity of recording data	4 times daily (00h00, 03h30,09h30,20h00)	Variable; no less than 3 recordings per day per ship
Data time stamp	For the purpose of analysis, noon recordings were used for both datasets (or the closest to noon, within two hours of noon)	
Wind speed measurement units	Beaufort scale (0 – 12)	Beaufort scale (0 - 12)
Wind direction measurement	Compass direction (32-point compass)	Compass direction (32-point compass)
Wind force correction	Converted to meters per second	Converted to meters per second
Wind direction correction	Already recorded as true north (photograph of wind instruction)	Corrected from magnetic north to true north readings by the CLIWOC project

Limitations

Considering inconsistencies in time and space

The most notable limitation in this project is that each data source recorded observations differently allowing for major spatial and temporal inconsistencies. The RAO captured data from a fixed position over time. The location is susceptible to effects of topographic wind shadows and terrain drag (Figure 7). Predominantly, the RAO is sheltered from westerly winds because of the Cape Peninsula mountain range (Lennard, 2014). Inconsistent blocks of missing data create temporal inconsistencies in the RAO dataset. However, the ship log data show different limitations.

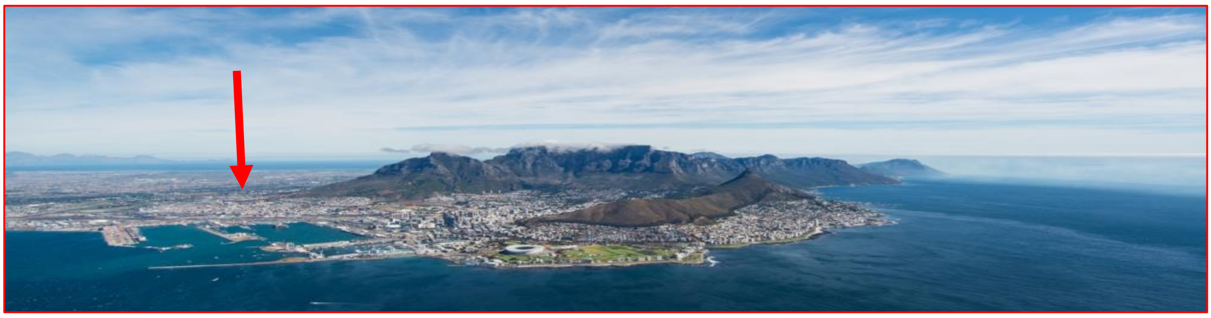


Figure 7 The red arrow points to the location of the RAO and shows the complex terrain around the observatory. This photo faces south. Photo taken from http://www.volvooceanrace.com/en/photo-view/30812.html [accessed June 2015].

Ship log observations come from multiple moving ships. The ships are sporadically spread over a large ocean domain. There is no consistent time step between recordings from the ships. The dataset consists of multiple or single observations for one day, at varying coordinate positions (Figure 8) and with variable gaps of time between available data (Table 4). Therefore, the ship log data are highly incongruent and inconsistent over space and time.

The independent inconsistencies in each dataset make it a challenge to perform a comparison. There are major spatial inconsistencies, temporal inconsistencies and the actual data integrity differs between data sources. As a result, the RAO and ship log observations may show very little correlation to each other in a direct comparative analysis. This weak correlation is not representative of atmospheric dynamics because certain conditions may have weak correlations, but are still plausible. Therefore, the comparison between data must be made a function of time, and of space separately in order to pull out correlations that exists as a function of different atmospheric processes. The limitations will be explained in more detail in the following section as well as a method of working within the constraints of the limitations.

year	month	day	time	lat	long	ID	country	WD	WS
1843	6	7	1100	-35.25	21.85	1421	NL	264	26.8
1843	6	7	1100	-35.31	21.89	1422	NL	297	4.6
1843	6	7	1100	-35.32	16.7	396	NL	197	22.6
1843	6	7	1100	-34.95	22.38	11	NL	354	2.6

Figure 8 The inconsistency in multiple readings for one day is evident in this map. A difference of 22,2m/s between relatively close ships weakens the correlation significantly. To curb this error, a single random observation per day was included in the study. The supplementary table shows the digitised CLIWOC data and the associated information for each ship plotted on the map.

Considering limitations in data integrity

Wind speed is converted from Beaufort Scale to m/s which creates a discrete dataset. The median value of the absolute wind force range in which the Beaufort Scale number lies is used. Therefore, the results may vary by >4m/s around the median value used. Please see

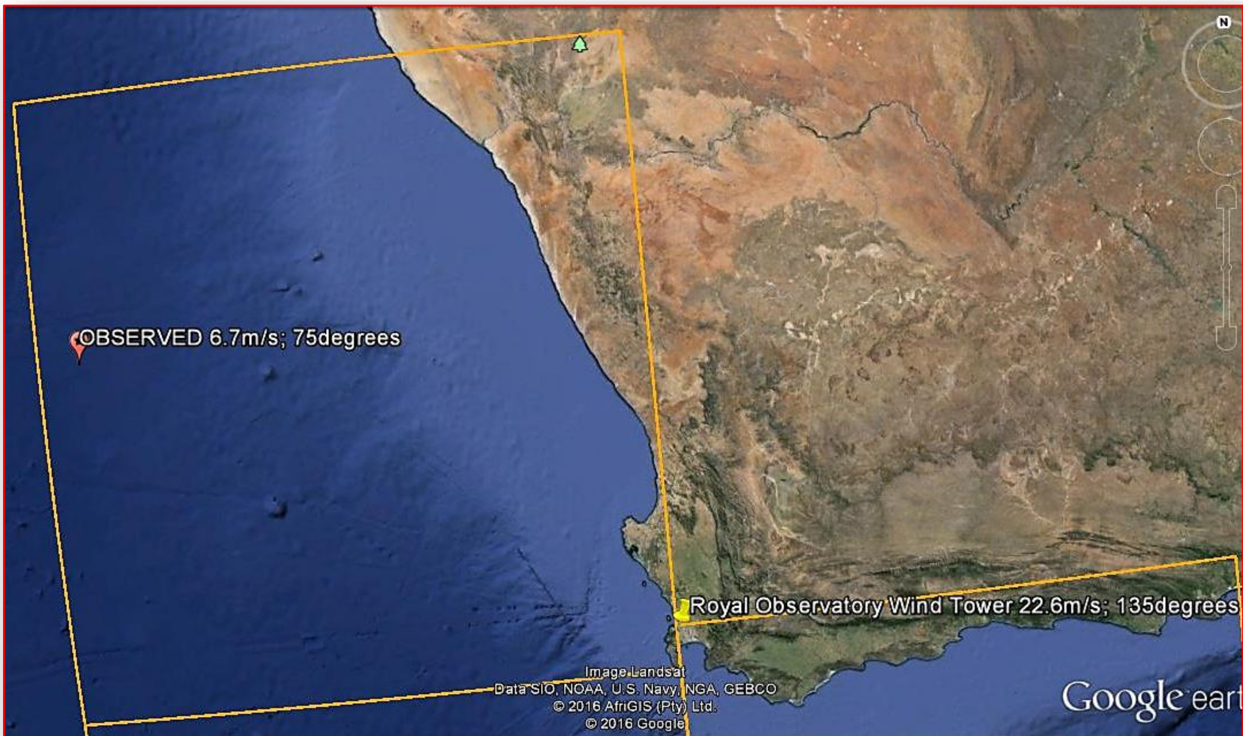


Figure 9 shows that observations far from the observatory (22.6m/s; 135degrees) are vastly different to the observations for a ship far offshore (6.7m/s; 75degrees). However, the data are not implausible. This is an example of how the variance in the data may introduce unnecessary noise and weaken the overall correlation.

Table 1 to determine the proportional variance between wind speed ranges for converting the Beaufort Scale to meters per second. However, as this is the best practice used by Hannaford *et al.* (2015) and documented by Wheeler (2005) then the same convention of analysing wind speed as an indicator of wind force will be used.

Seasonality and meso-scale atmospheric conditions

The atmospheric circulation and resultant wind conditions experienced at varying temporal and spatial scales will introduce additional noise to the data and weaken the correlation between data sources. Seasonality causes travelling meso-scale systems to affect areas smaller than the domain extent. Seasonal atmospheric conditions will influence the correlation and relationship between data during the same day because the data points are variably located in a large domain. Seasonality/ mesoscale phenomena through time will weaken the correlation.

Time series analysis limitations

The data from the RAO is missing 7,4% of the noon observations from 1834 to 1854 (Table 2). Some of this missing data occurs in blocks of time (days/weeks at a time). The ship log data are not continuous over time and only occur at most up to 9 consecutive days before a gap in time in the data occurs. Therefore, the inconsistency in the data does not allow for a reliable time series trend analysis to be done.

Table 4 an example of the inconsistency in the ship log observations. The green blocks indicate the same ship with 8 consecutive days worth of observation. The orange and yellow rows show two different ships that recorded on the same day, with varying observational values. Between the green rows and the blue row, there are 4 days of missing data. WS is given in m/s. "999" shows no observed data for that day.

Y	M	D	TIME	LAT	LONG	ID	COUNTRY	WD	WS m/s
1842	05	09	1100	-34.75	16	160	NL	287	9.3
1842	05	10	1100	-34.97	21.63	395	NL	241	4.6
1842	05	11	1100	-35.22	21.6	395	NL	27	9.3
1842	05	12	1100	-35.33	20.25	395	NL	174	2.6
1842	05	13	1100	-35.4	19.37	395	NL	5	2.6
1842	05	14	1100	-35.15	17.93	395	NL	331	4.6
1842	05	15	1100	-34.92	16.15	395	NL	208	12.3
1842	05	16	1100	-33.43	14.9	395	NL	175	9.3
1842	05	17	1100	-31.88	12.73	395	NL	85	12.3
1842	05	22	1800	-34.68	24	820	NL	84	999
1842	05	23	1900	-34.85	22.42	820	NL	16	999
1842	05	24	1100	-34.25	17.05	159	NL	208	9.3
1842	05	24	1100	-34.3	17.08	157	NL	220	6.7

Creating comparable datasets: a refined region of study

Selecting the domain

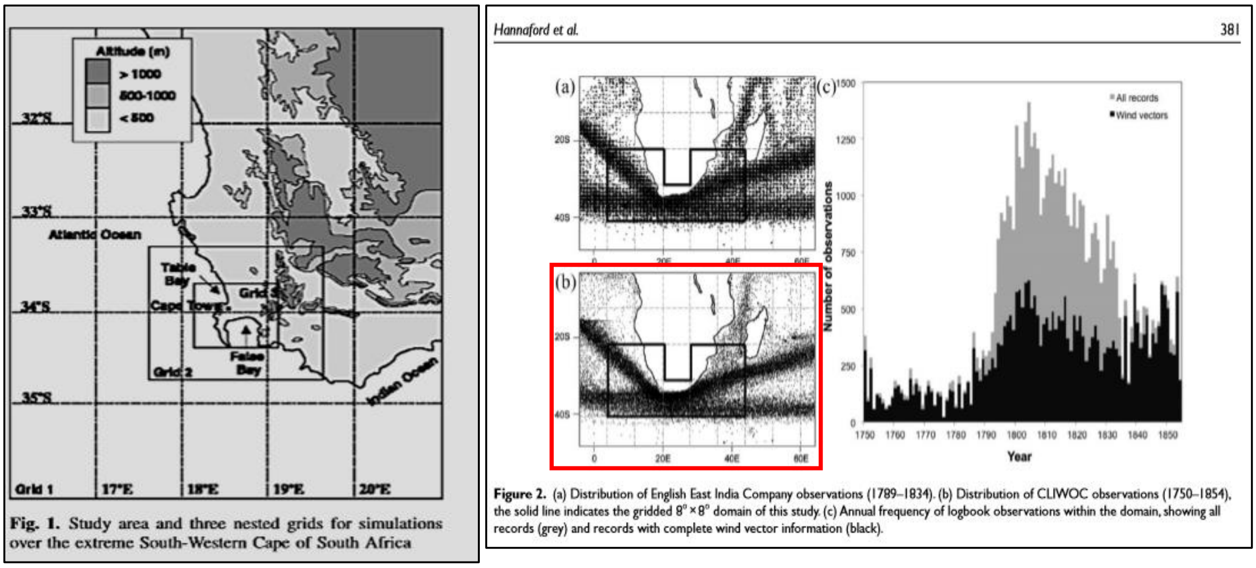


Figure 10 (left) Bonnardot et al. (2005) suggests at the most a 550km x475km domain (largest box) to test meso-scale weather phenomena like sea-land breeze generation. (right) Hannaford et al. (2015) shows the density of ships that travelled along the EEIC trade routes. The black boxes also show the domain Hannaford et al. (2015) used to reconstruct historical precipitation patterns.

The extent of the original 8° x 8° domain used by Hannaford *et al.* (2015) is too large for the purpose of this study (see Figure 2 and Figure 10 for Hannaford's *et al.* (2015) domains). Data within a large domain will introduce significant spatial and temporal error into the data comparison. Therefore, the domain is down-scaled to an acceptable range that will provide adequate ship log data density while minimising the amount of error introduced by sub-grid atmospheric dynamics and phenomena. A sea-breeze simulation study by Bonnardot *et al.* (2005) was originally used to downscale the spatial extent of the domain (Figure 10). However, only 473 ship log data values fall within the biggest grid box of 550kmx470km used by Bonnardot *et al.* (2005) and over 14 000 values at the RAO for the same time period. This difference is not adequate to perform a reliable comparison. Therefore, the extent of the grid is increased further than that of Bonnardot's *et al.* (2005) grids in order to gain enough data for the most reliable comparison.

Three refined study zones

Caution must be paid to the seemingly weak relationships that may be plausible given the complexities of dynamic atmospheric systems over time and space (see "Considering inconsistencies in time and space" for an example). Within the downscaled region of study as seen in the figure of all grids, three boxes are defined by dissimilar and distinct geographical properties such as ocean currents, sea surface temperature and dominant synoptic air flow regimes over time and space (Summary in Table 5). This separation minimises the error caused by varying geographical and atmospheric properties throughout the domain. By performing comparative tests between the RAO and ship logs in each box separately, a meaningful and reasonable understanding of the extent to which the ship log data reflects the RAO data can be quantified.

Table 5 presents the distinctive properties that determine the three study regions for the CLIWOC data.

Distinctive properties in each box within the study region			
	Area 1 (A1)	**Area 2 (A2)**	**Area 3 (A3)**
Ocean current	Cold Benguela current	Turbulent Cape Retroflection	Warmer Agulhas current and retroflection current
Sea surface temperature	Cold	Mixing of warm and cold ocean currents	Warm
Characteristic weather events	Mid-latitude cyclones, coastal low pressures	Mid-latitude cyclones	Ridging high pressures
Dominant summer synoptic airflow	SE'lies	Variable	W'lies and E'lies (ridging high pressure)
Dominant winter synoptic airflow	NW'lies	Variable	Variable

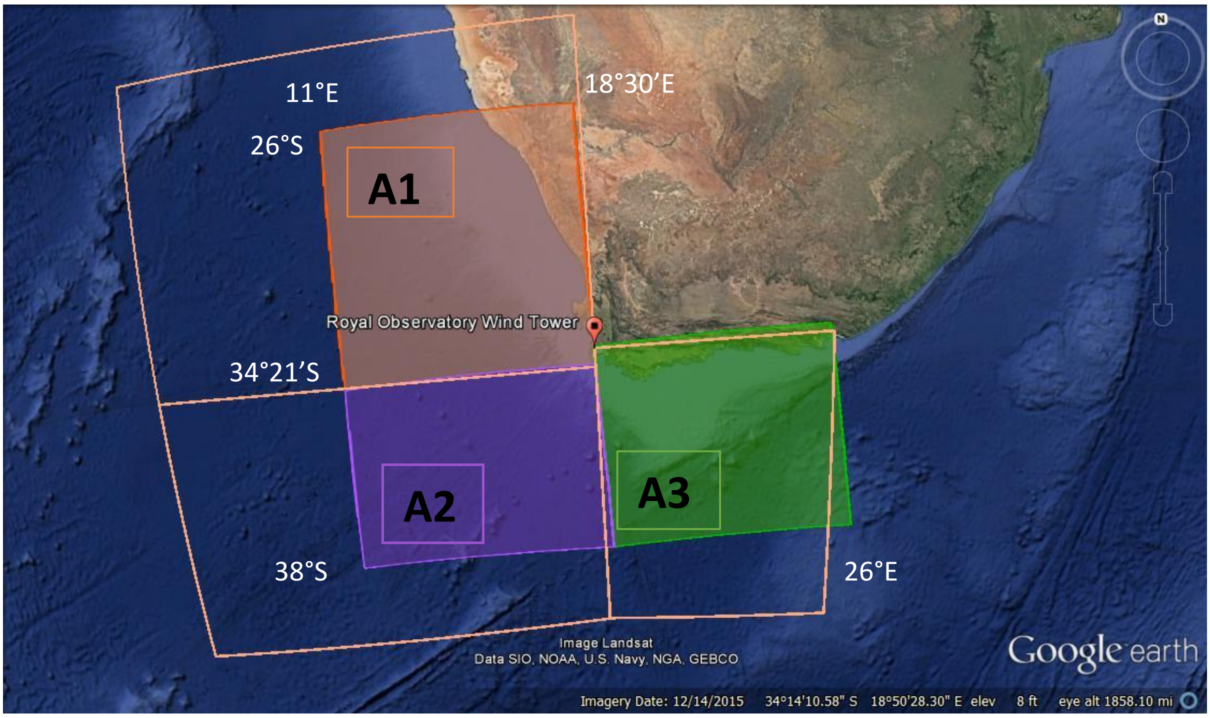


Figure 11 The three selected study areas for this study are shaded in colour. These are substantially smaller than the domain used by Hannaford et al. (2015). The three areas are characterised by differing geographical properties. The RAO is shown with a red pin point marker on the South Western tip of Southern Africa.

Selecting data

The data selected from each dataset was based on data availability in the cross over period, the time of day that the data was recorded and making the datasets as homogeneous as possible. Therefore, data recorded at noon (also recorded as "0h00" and 12:00) was used between 1834 and 1854. Only data with complete information was used (wind speed and direction). The data with matching date stamps were extracted so that the RAO dataset replicated the chronology of the ship log dataset, and *vice versa*. The unmatched dates were discarded from the comparative analysis, but not for establishing a general climatology for the RAO region. In some cases, where more than one observation was made for the same day in the ship log data, only the first observation was included. This random selection of data eliminates the issue of duplicate data for one date stamp.

A new database is built from the pre-selected data from both sources. The database consists of a date stamp, a wind speed (in m/s), a wind direction (in decimal degrees) and a latitude and longitude position for the ship log data. From this database the appropriate comparative analysis can be executed in each defined area and for alternative seasons (or cumulatively). Seasonal and region limited data will satisfy the time-space relationship aims and objectives of this project.

Comparing the data

No established methodology is used for this project, but rather a customised and crude way of testing the data as a function a (i) space (distance between data sources) and (ii) time (equal date stamps) in order to minimise the limitation error. Wind vector frequency analyses allow for an improved understanding of independent wind climate variabilities based on the individual datasets. As this is an inquiry into the newly created RAO wind dataset, the RAO wind data works as the reference data for this project. Seasonal variability and the area in which the data lie complicates the relationship data (see "Limitations" section), therefore each gridded area (Figure 11) will be compared separately and for each season.

For the purpose of comparing the RAO data and the ship log data, only date matched noon readings that have complete vector information (I.e. a wind speed and direction) will be used. All wind directions are true north. All wind speeds have been converted from the Beaufort Scale to metres per second using a proportional wind class range defined by the CLIWOC report (more details on the CLIWOC conversions are available at http://pendientedemigracion.ucm.es/info/cliwoc/Cliwoc_final_report.pdf.)

Establishing a reference climatology

This project aims to develop a general climatology for the RAO by analysing all recorded wind speeds and direction at the RAO station from 1834 to 1854. The RAO dataset provides the reference dataset for this project. Seasonal wind roses will present seasonal patterns in the wind patterns experienced at the RAO. Frequency analysis is a common method of establishing an understanding of the wind characteristics for an area as shown in papers by Wheeler and Saurez-Dominguez (2006) and Brazdil *et al.* (2005). Analysing the frequency of the wind speed classes for 45° directional segments (8 compass directions) for each season is indicative of the wind climatology over time. Therefore, a heat map (frequency distribution table) is used to supplement the observed wind climate at the RAO, and over the ocean.

Determining data relationships on the same date

What do the data show for the same date?

A scatter plot graph will show the difference between observed wind speed and direction recorded on the same day between the RAO and CLIWOC data. Time is made the independent variable in the relationship, showing how the data changes in relation to each other over time. An additional statistical analysis will test the extent of the relationship using a linear Pearson correlation (r value), significance test (p value) and a variability check using the R^2 coefficient (a measure of how much data variability can be explained by the correlation in the data).

Determining data relationships based on distance away from the RAO

Does the data relationship deteriorate as you move further away from the RAO?

The ship log data are given with a coordinate position (latitude; longitude) which allows the great circle distance between the RAO and the ship in question to be calculated. The *law of cosines* formula in Equation 1 calculates the distance between two points on a sphere using trigonometry. Each sources' latitude and longitude coordinates ("lat 1" or "lay2") are used in the trigonometric calculation multiplied by 6371, which equals the radius of the earth in kilometres (Moveable Type Scripts, n.d. [accessed 13 August 2016]). Equation 2 shows the formula used in Microsoft Excel to calculate the great circle distances between the individual ships and the RAO in the database using the coordinate positions (Moveable Type Scripts, n.d. [accessed 13 August 2016]). The calculated distance is given in kilometres.

$$\Delta distance = acos(\sin\varphi 1 * \sin\varphi 2 + \cos\varphi 1 * \cos\varphi 2 * \cos(\Delta\lambda\,) * 6371$$

(1)

= ACOS (SIN(lat1) *SIN(lat2) + COS(lat1) *COS(lat2) *COS(lon2-lon1)) * 6371

(2)

The difference in observations between RAO and CLIWOC data, based on the distance away from the RAO tests the relationship as a function of space. The physical distance between RAO and ship is plotted as the independent variable of a scatter plot graph with corresponding data value differences for the same date on the y axis. An additional statistical correlation analysis will test the strength of the relationship using a linear Pearson correlation (r value), significance test (p value) and a variability check using the R^2 coefficient.

Flagging dominant wind vectors in Area 1, Area 2 and Area 3

Each limited area in the study region will be investigated for dominant wind vectors per season. The same frequency distribution tables (heat maps) as used for the RAO reference climatology section will be replicated here. These distribution tables will be compared against the RAO frequency distributions. This method of analysis will validate (or invalidate) the data by exposing typical (or atypical) variability, and satisfy a previously unknown wind climate variability for the oceans surrounding Southern Africa.

Infographic of how the data is binned and analysed

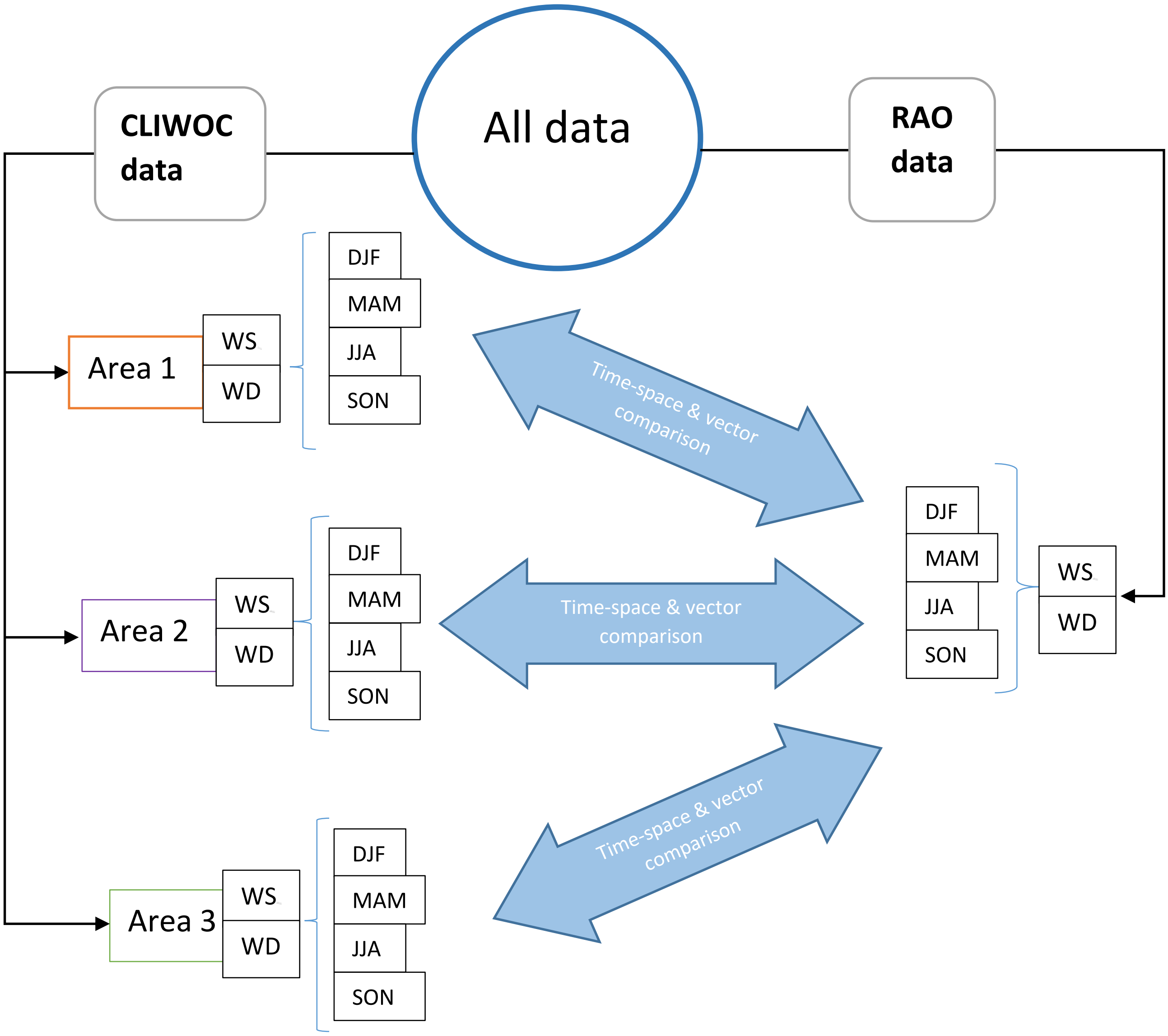


Figure 12 An infographic (flow diagram) or how the data is preselected and then analysed against each other. DJF, MAM, JJA and SON are the seasons. WD is wind direction. WS is wind speed.

Results

Introduction

In this section, the RAO data will establish the reference climatology for 1834 to 1854. Prior to digitisation, RAO data were never before analysed. Therefore, RAO data will be treated as the reference dataset against which the CLIWOC ship log data will be compared. All analyses defined in the methodology are executed in relation to this RAO reference dataset. The spatial and temporal comparisons between the RAO and the CLIWOC ship log data will be presented in two parts that first address wind direction and then wind speed, and then supplemented with a table of statistical correlations. Area specific vector analyses will complete the determined extent to which the ship log records reflect the RAO records. The results are analysed per season where:

Summer is defined by December, January and February (DJF),
Autumn is defined by March, April and May (MAM),
Winter is defined by June, July and August (JJA) and,
Spring is defined by September, October and November (SON).

General climatology for the RAO

Newly digitised data from the RAO for the mid-19th century shows that atmospheric dynamics for the Cape Peninsula are very seldom calm, 0 m/s, accounting for only 7,96% of all recorded noon observations (see Table 6). Instead the frequency distribution in Table 6 and wind rose in Figure 13 show an expected wind climatology based on a contemporary understanding of the wind patterns in the South Western Cape of South Africa (Tyson and Preston-Whyte, 2004). The dominant wind directions are NW'ly or S'ly, with a lower frequency of W'ly and SW'ly winds. Table 6 shows that ten of the twelve wind speed classes had been observed and recorded at the dominant directions, showing a high variability of wind strengths (0 to 28.4m/s), but the dominant direction of sustained wind patterns.

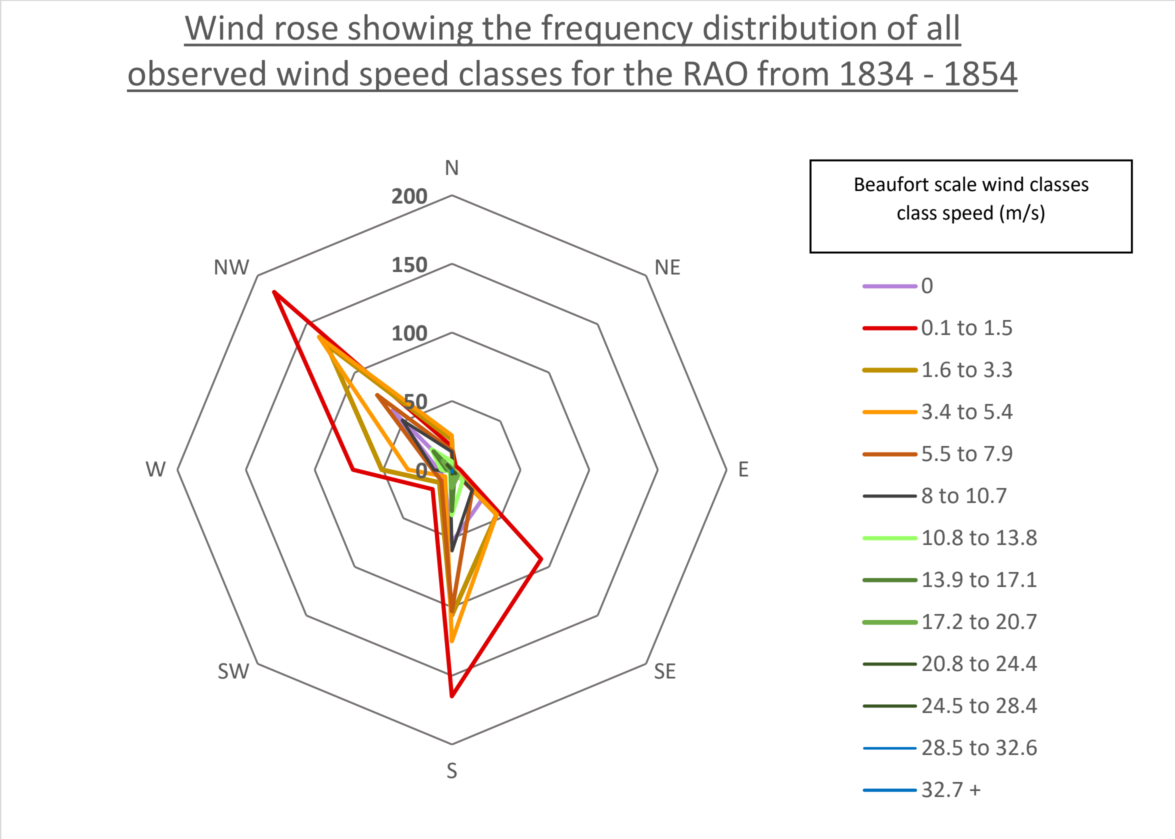


Figure 13 A wind rose shows the frequency of observed wind speeds at a particular direction. The general climatology for the RAO is evidently in the southerly and north westerly direction, with the most occurring wind speeds at 0.1 – 1.5m/s. wind speeds of 1.6 to 5.4 are have a high occurrence.

The frequency distribution Table 6 shows that 33,62% of all captured winds (of all strengths) are from a southerly direction and 13,78% from a SE'ly direction. The stronger winds (>13,9m/s) captured in the SE sector are often referred to as the "Black South Easter" in the RAO meteorological registers (Figure 5). This infamous wind driven by the stable South Atlantic High Pressure cell still adversely affects Cape Town city and which the City of Cape Town government issues warnings for. See the City of Cape Town website for government issued disaster risk management information with regard to severe wind at https://www.capetown.gov.za/en/DRM/Pages/StormsandHighorGaleforceWind.aspx.

33,82% of wind classes are from a NW'ly direction, which can be attributed to the passing winter cold fronts. There is a low frequency of very strong winds that drive intense sea storm surges that Turner (1988) documents. Note that the RAO records only capture 5,67% of the wind observations from a northerly to easterly direction (337,5° to 112,5°). 89,58% of all the wind recorded at the RAO is below 10,7m/s and a substantial 10,42% of all wind experienced at the RAO was gale force strength. Therefore, we can see that the RAO captured dominant wind patterns that are documented today.

Table 6 A detailed quantification of the distribution of wind speeds and associated directions for the RAO from 1834-1854. A southerly and north westerly wind account for 67,44% of all wind data observed at the RAO. With dominant wind speeds between 1,6m/s to 7,9m.s.

		wind direction in 45° decimal degree increments									
	Frequency distribution for RAO	N	NE	E	SE	S	SW	W	NW	No. of events	% of events
	0	0	0	1	31	53	8	10	61	164	7,96
Wind speed class in m/s based on the Beaufort Scale conversion	0.1 to 1.5	17	4	6	92	165	20	72	183	559	27,12
	1.6 to 3.3	22	1	1	46	106	13	51	129	369	17,90
	3.4 to 5.4	25	0	1	46	125	7	32	137	373	18,10
	5.5 to 7.9	14	1	0	21	103	11	14	77	241	11,69
	8 to 10.7	13	1	1	21	59	2	13	51	161	7,81
	10.8 to 13.8	6	0	1	11	33	2	8	21	82	3,98
	13.9 to 17.1	0	0	0	3	30	1	4	19	57	2,77
	17.2 to 20.7	1	1	0	6	13	0	2	9	32	1,55
	20.8 to 24.4	0	0	0	3	3	0	0	5	11	0,53
	24.5 to 28.4	0	0	0	4	1	0	0	5	10	0,49
	28.5 to 32.6	0	0	0	0	2	0	0	0	2	0,10
	32.7 +	0	0	0	0	0	0	0	0	0	0,00
	No. of events	98	8	11	284	693	64	206	697	**2061**	
	% of events	4,75	0,39	0,53	13,78	33,62	3,11	10,00	33,82		

A seasonal breakdown of wind roses in Figure 14 provides insight into the seasonal timing of when the dominant wind patterns are experienced. The seasonal wind roses show ideal seasonal wind patterns for the region. The winter wind rose shows a high frequency of NW'ly winds that are reflected in the distribution table (Table 6). Hence, the NW'ly winds were predominantly winter winds.

The shoulder seasons (autumn and spring) don't show dominant wind directions as strongly winter or summer do. This is for two reasons. Firstly, the frequency of observations is lower, so a clear signal is less evident. Secondly, instead of predominant wind regimes like summer SE'lies and winter NW'lies, the shoulder seasons experience high variability in wind directions due to the shifting atmospheric conditions and changing seasons. For this reason, the average wind force is lower too. Autumn does show a tendency for increased NW'ly winds as mid-latitude cyclones would continue to affect the area. Spring shows a tendency towards S'ly winds, as the SAHP would begin to stabilise. This is representative of an atmosphere in flux with seasonal ITCZ shifting.

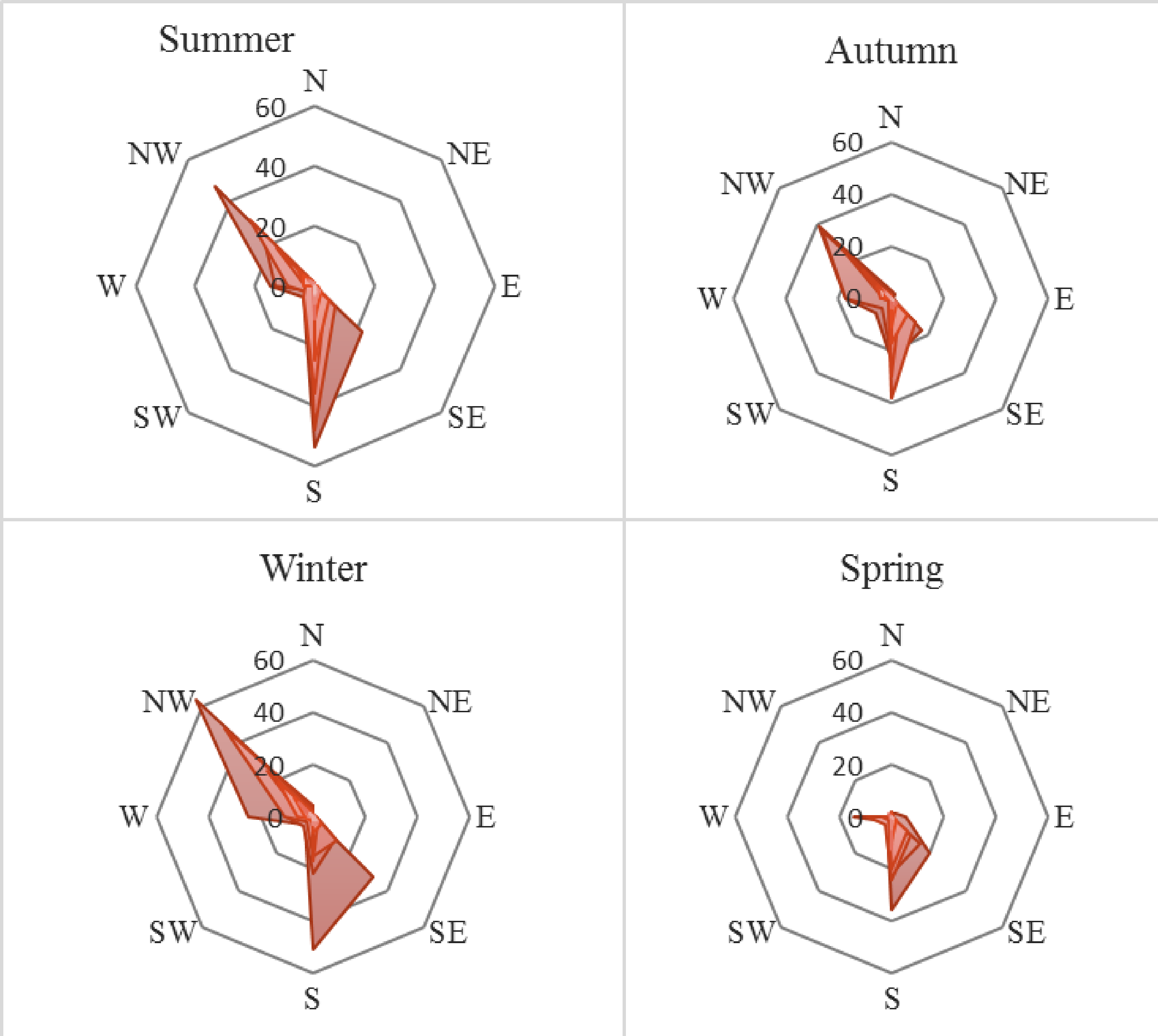


Figure 14 Seasonal wind roses at the RAO reveal dominant wind directions that agree with the seasonal synoptic scale atmospheric conditions in Tyson and Preston-Whyte (2004). Summer and winter show a higher frequency in recordings and produce accurate predominant seasonal wind regimes. The summer time anomaly of north westerly winds is noticeable in the summer wind rose, and is a possible result of sea breeze development in Table Bay as observed by Bonnardot et al. (2015). Spring shows very few observations in comparison to the other seasons, but a predominant southerly wind is noticeable.

The summer season in Figure 14 shows two interesting signals. Firstly, an expected high frequency S'ly and SE'ly wind component, a typical dominant summer wind. The wind rose also shows a high frequency occurrence of NW'ly winds which is not usual for this season. Investigating the monthly wind direction frequency for November, December, January and February might suggest that the unusual NW'ly winds were captured at the very beginning or end of the summer season. However, the monthly frequency analysis in Table 7 shows that the NW'ly winds were captured consistently in the summer months and not at the summer boundary periods. Conversely, the summer (DJF) frequency distribution table for Area 1 (Table 10), the CLIWOC region, most likely to capture the same wind patterns for summer, only captures two occurrences of a NW'ly wind. This phenomenon (noise) in the RAO data can be attributed to the development of late morning sea breezes in the Table Bay region as observed and modelled by Bonnardot *et al.* (2015). There is evidence in the journal kept by Jan van Riebeeck in 1651 – 1655 that show some reports of NW'ly winds in the summer periods (Thom and Balkema, 1952). Therefore, the NW'ly anomaly in the RAO general climatology can be explained by the evidence in historical documents and small scale modelled/observational analyses (Bonnardot *et al.*, 2015).

Table 7 investigates the possibility that the anomalous north westerly winds in summer were captured in the shoulder seasons, which has a higher possibility of occurrence. The shoulder months and summer months consistently recorded NW'lies at the RAO (17 to 20 times per month).

Shoulder season WD frequency at the RAO: NW recorded consistently at the Royal Astronomical Observatory in the summer months. Reason is unknown.				
	Nov	Dec	Jan	Feb
NE	0	0	0	0
SE	27	29	32	0
SW	20	17	22	1
NW	**17**	**20**	**18**	**16**

In conclusion, the RAO experiences seasonal wind patterns that are controlled by seasonal atmospheric fluctuations, as they are today. Additionally, small scale wind patterns like sea breezes, affect daily winds. The wind strength diminishes in the spring and autumn season due to variable atmospheric conditions and fluctuating synoptic conditions. Winter experiences strong winds from the NW direction in winter. Summer experiences strong winds from the S/ SE direction predominantly in summer. The outlying NW'ly winds recorded in summer at the RAO are attributed to sea breeze development as observed and modelled by Bonnardot *et al.* (2015), and seen to occur in the past by Van Riebeeck's journals (Thom and Balkema, 1952).

Determining data relationships for the same date

To begin with, wind direction (WD) data will be addressed first, and then wind speed (WS) data will be addressed.

Wind direction over time in Area 1

Area 1 refers to the region on the west coast of Southern Africa (refer to Figure 11 for a map). Wind directions per observation day, for each season, are plotted over time on scatter plot graphs shown in Figures 15 to 17. All data are very widely spread out, with a higher variance in the ship log (SL) data. The ship log data are very scattered between 90° and 315°, except for summer and spring where the data are more concentrated around 180°. This indicates a dominant summer southerly wind flow regime is captured by the CLIWOC ship logs. The RAO shows consistency in winds at 180° (S) and 315° (NW), with some variability and outliers. The RAO seasonal patterns over time are in agreement with the general climatology discussed in the previous section (see frequency distribution Table 6).

The statistical correlation between the seasonal data is weak; where r<0,131437 and p>0,05 for all seasons in Area 1 (refer to Table 8 for statistical results). Added to which, the spread of data is completely random; where R^2%<1% in all areas during all seasons. Therefore, the data in Area 1 do not show any significant correlation and are completely independent of each other in their absolute values and in their variability.

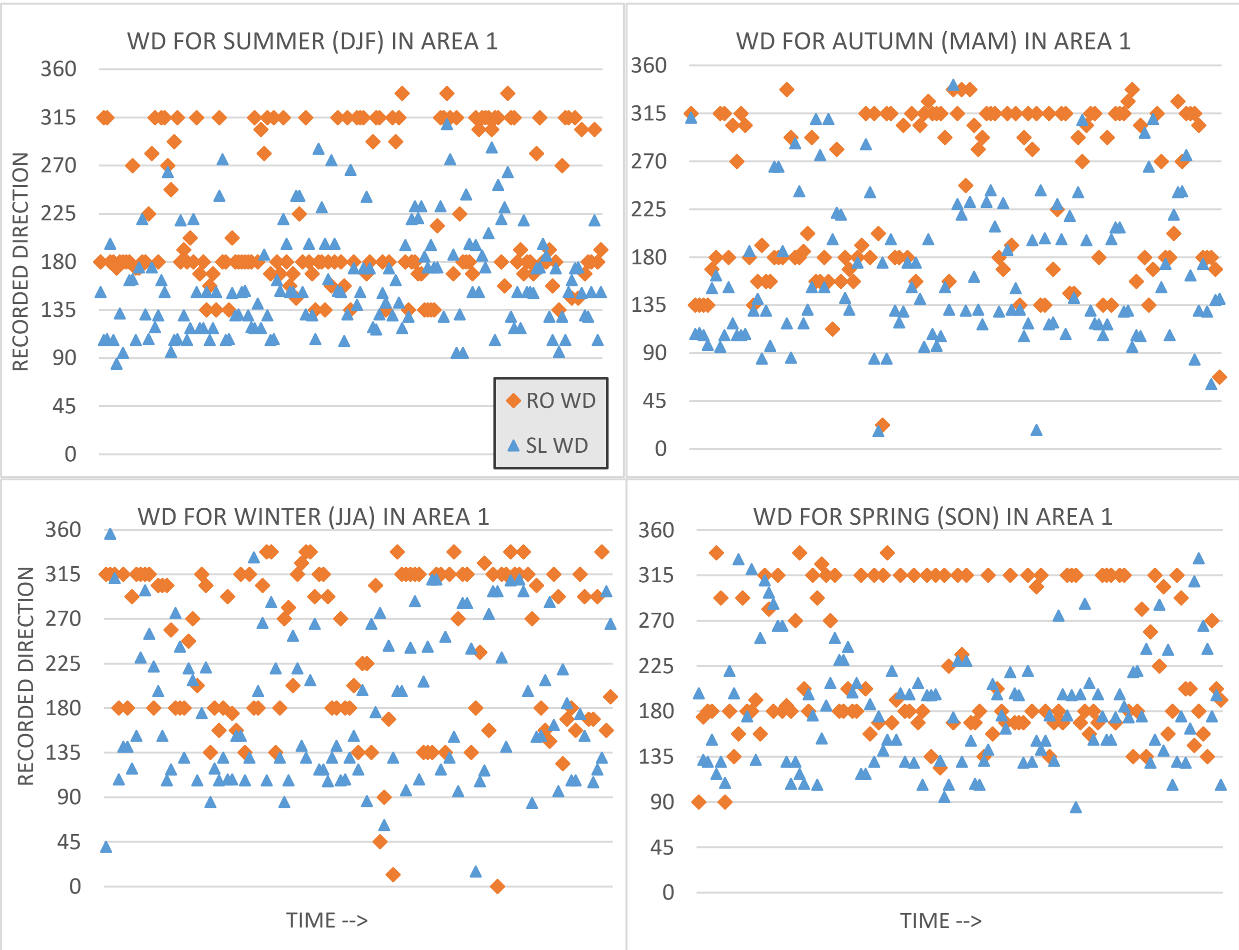


Figure 15 Seasonal scatter plot graphs for WD in Area 1 show no significant signal in the relationship between the data from the RAO and the ship logs from 1834-1854. The data from the RAO (RO WD) does show an aggregation around 180 and 315 degrees. The CLIWOC data has a much higher variance in comparison.

Wind direction over time in Area 2

Area 2 refers to the region south west of the Cape Peninsula (Figure 11). The data for the RAO shows different values to that of Area 1 because of the difference in date matching for Area 2. Nevertheless, the wind patterns evident in the data captured at the RAO has not altered, showing dominant WD at 180° and 315°, and some variance and outlying values. Also, similar to the analysis in Area 1, is the high variability in the CLIWOC ship log data, however the only difference is that more winds between 0° and 90° have been documented.

As with the lack of statistical correlation in Area 1, the data in Area 2 shows no visual similarity in the data distribution over time. Additionally, there are very low correlations for each season where r<1 for each season and the variability shows no similarity either (R^2%<4% for each season) (see Table 6 For statistical correlations). Therefore, the RAO data are independent of CLIWOC data in Area 2 and shows no correlation in absolute values or variability over time.

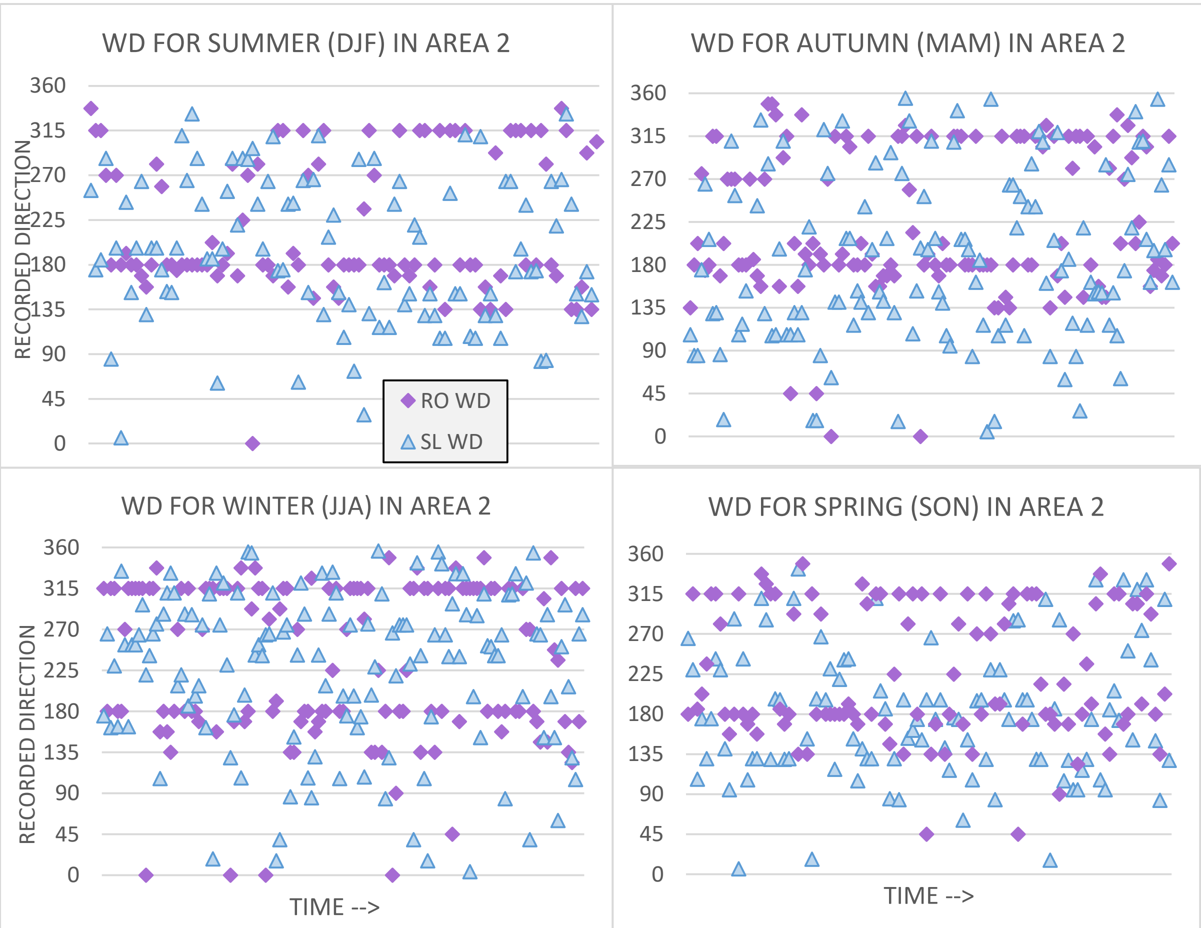


Figure 16 Seasonal scatter plot graphs for WD in Area 2 show no significant signal in the relationship between the data from the RAO and the CLIWOC data from 1834-1854. The RAO data (RO WD) are again aggregated at 180 and 315 degrees. The CLIWOC data shows high variance.

Wind direction over time in Area 3

Area 3 refers to the region to the south of Southern Africa (Figure 11). There is more data within this Area 3 (1194) than Area 1 (522) and Area 2 (475) (Table 8). As a result, the scatter plot graphs in Figure 17 are densely covered with data values. Nevertheless, there are similarities in the visual data. The RAO experienced predominantly S'ly and NW'ly winds while the CLIWOC ship log data are sporadically spread through a 360° range. Summer and winter seasons show the seasonal winds noticeably in both datasets.

Like Area 1 and Area 2, Area 3 also shows no statistical correlation between the datasets in the changes and patterns over time or the variability. In conclusion, there is no correlation between the RAO data and the ship log data in any of the three regions with regard to the recorded wind directions.

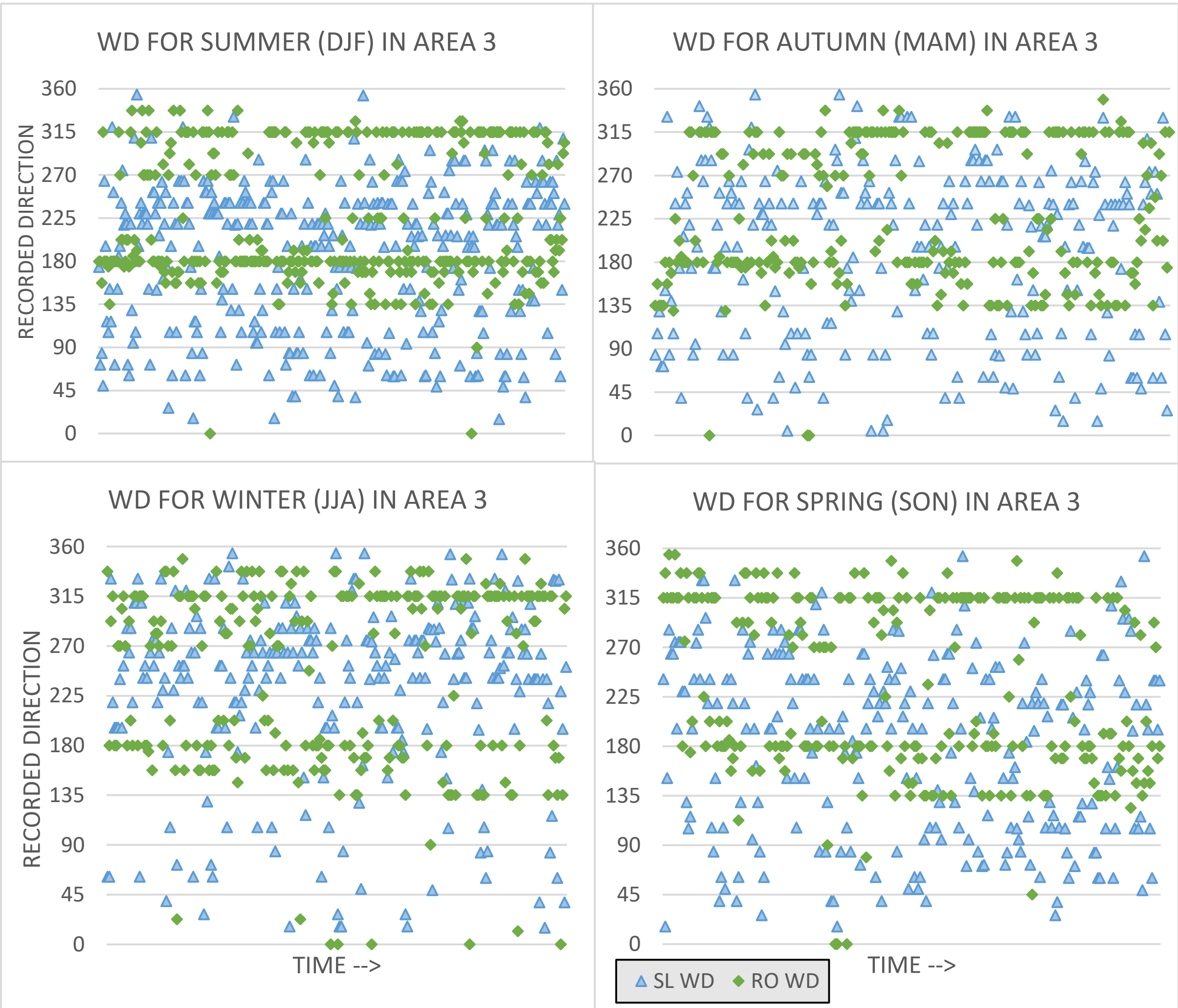


Figure 17 Seasonal scatter plot graphs for WD in Area 3 show no significant signal in the relationship between the data from the RAO and the CLIWOC data from 1834-1854. It is clear that Area 3 contains more observational data in the CLIWOC dataset by the high density of values. Again, the RAO shows aggregation at 180 and 315 degrees, and the CLIWOC shows high variance.

Wind speed over time in Area 1, Area 2 and Area 3

The discrete characteristic of the converted wind speed data (refer to "Considering limitations in data integrity" in the Limitations section) is shown clearly in all scatter plot graphs in Figures 18, 19 and 20 by the definitive wind speed values. Area 1 shows a high variability of wind speeds over time, diminishing any clear signals from being visually observable. Wind speeds did not record above 25m/s. Area 2 shows a very similar dispersion of data, with the exception of a few extreme outlying ship log observations above 25m/s in autumn, winter and spring. These can be attributed to the mid-latitude cyclones. No further investigation will be done to find causation because the purpose of the project is to determine the extent to which the data sources are similar/dissimilar only. Area 3 shows a wider spread of data from calm condition up to 30m/s for all seasons and both data sources. The RAO records more in the lower wind speed classes where as there is higher variation in the wind speeds recorded on the ships. However, there is still no predictable progression or association in the data to show signals and correlations over time.

Statistically there is no significant strong correlation between WS data from the RAO and on board ships in any of the given areas during any of the seasons (Table 6). The coefficient of determination percentages (R^2 %) also suggest that the variability in the data is random and cannot be determined based on any organic standard deviations from a very weak relationship between the data (all R^2 percentages are below 6%).

Conclusion

In conclusion, there is no discernible relationship or signal between the wind direction data and wind speed data for the RAO or the ship log data for any of the areas. The data is randomly scattered throughout time. There is no significant correlation between any of the data. The exceptional seasons that show a significant correlation (Area 1 in winter for wind direction data correlation) only applies to a very weak correlation and is therefore not indicative of data that correlate well. The very low coefficient of determination percentages all suggest that the variability in the data cannot be explained by the correlation of the datasets. To this end, it is evident that the datasets are not well correlated and are completely independent of each other over time.

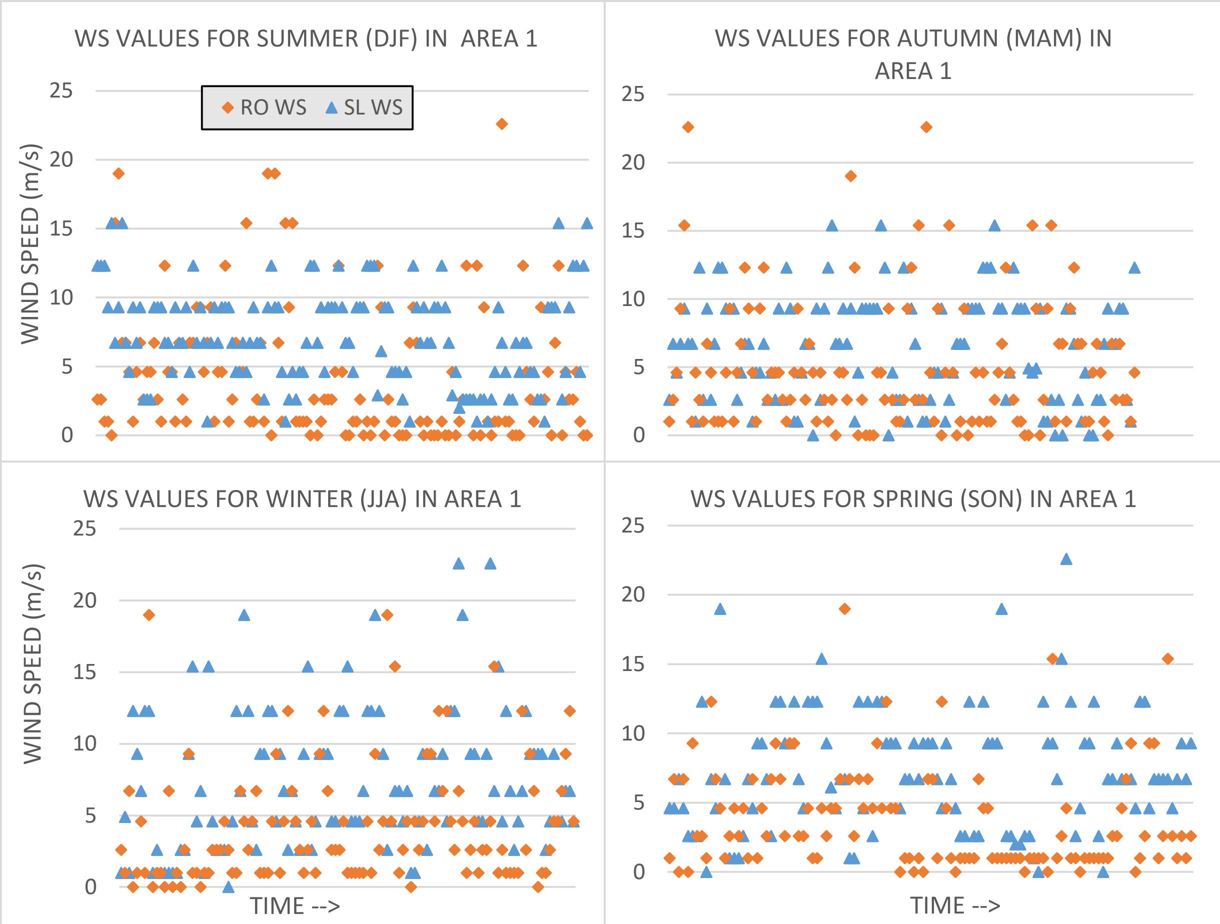


Figure 18 Seasonal scatter plot graphs show plotted WS over time in Area 1 for the RAO and CLIWOC datasets from 1834-1854. Data show a high variance and random scattering, there is no clear signal in the data or relationship between the data.

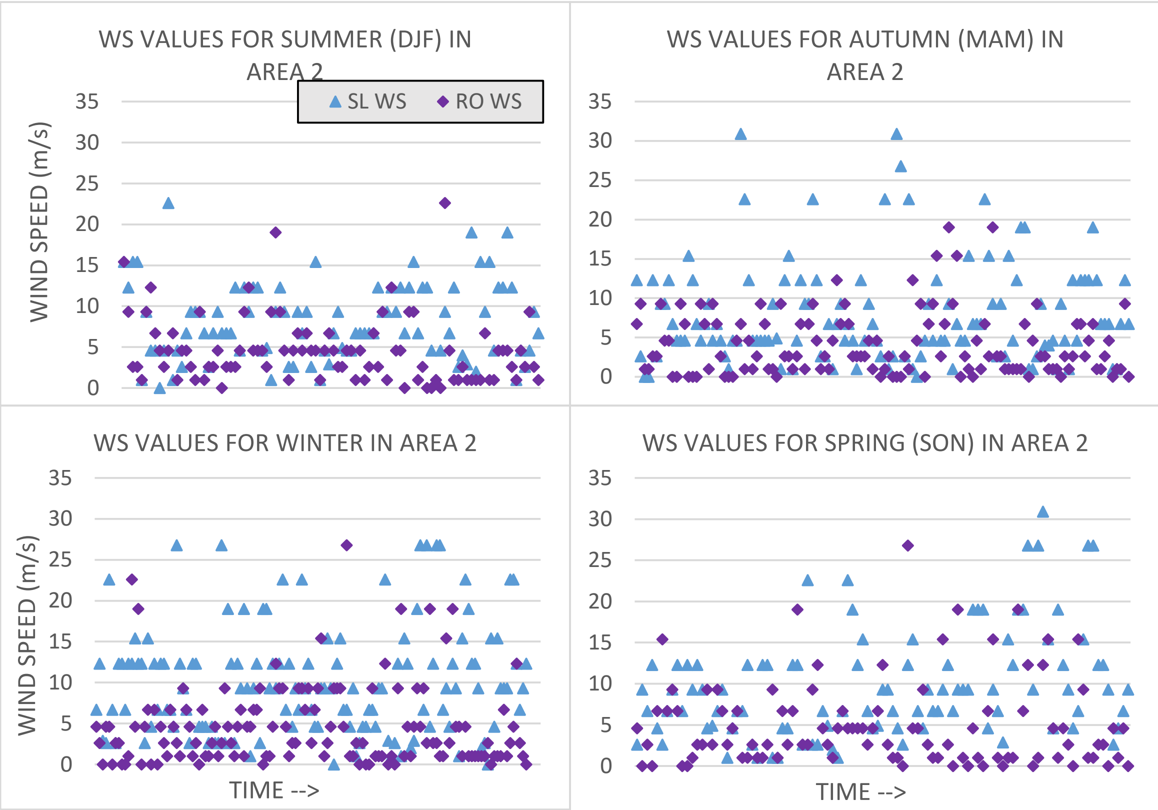


Figure 19 Seasonal scatter plot graphs show plotted WS over time in Area 2 for the RAO and CLIWOC datasets from 1834-1854. There is no significant relationship between the data.

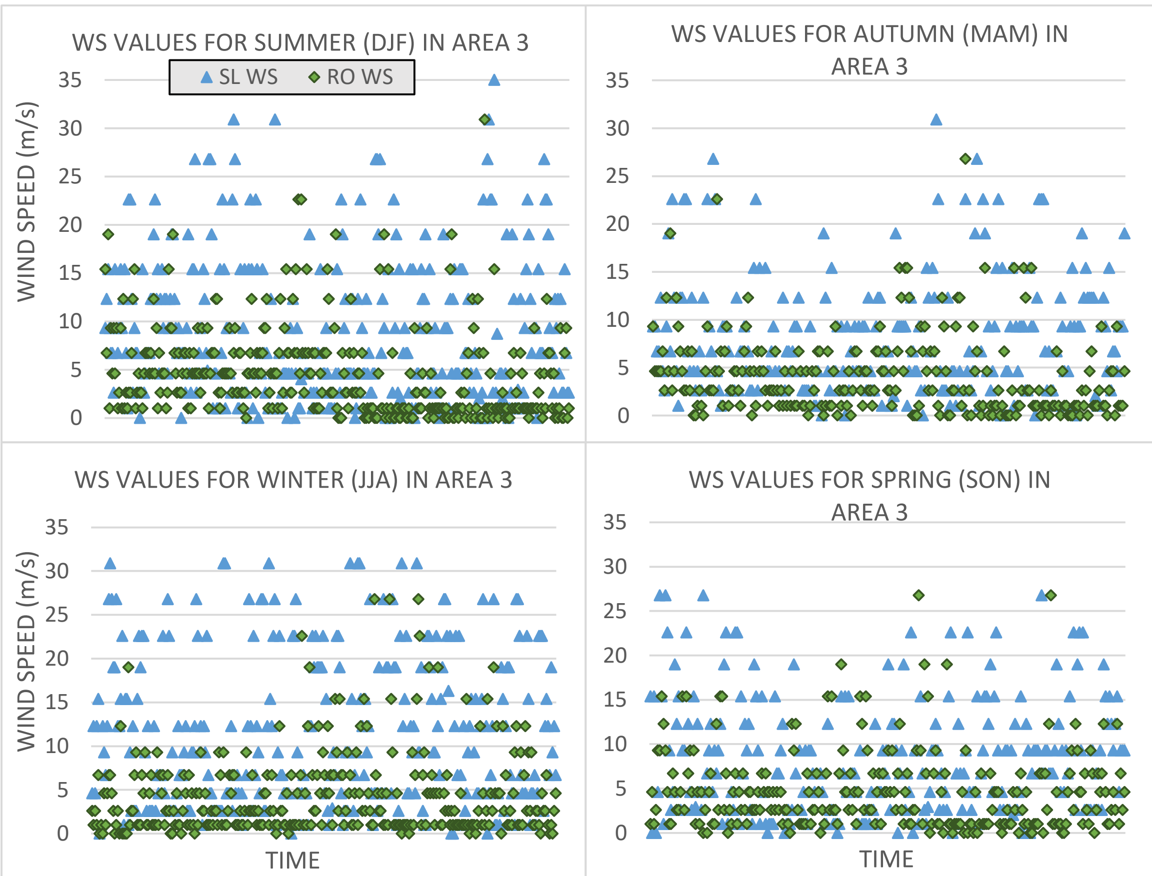


Figure 20 Seasonal scatter plot graphs show plotted WS over time in Area 3 for the RAO and CLIWOC datasets from 1834-1854. Data shows a high variance and random scattering, there is no clear signal in the data or relationship between the data.

Table 8 Statistical correlation results for data values over time. Pearson correlation (r), statistical significance (p value), correlation coefficient and percentage (R^2 and R^2%) and number of observations are given for each dataset comparison between RAO and each area (1-3) for each season. The are no significant statistical correlations. The four cases of statistically significant correlations (green blocks) are random, shown by the corresponding very weak r values and R^2% values. The (top) table comprises wind direction data and the (bottom) table comprises wind speed data.

Pearson correlation between WD data values for the same date

WD	Area 1				Area 2				Area 3			
Time	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
Pearson r	0.029261	0.071098	0.000883072	0.131437	0.083146	0.184337	0.067867	0.090374	0.046174	0.08770265	0.015589	0.017561
P value (stat sig)	0.716016	0.425161	0.000000779815808	0.152441	0.408442	0.035056	0.430698	0.356887	0.384391	0.160146028	0.783205	0.775994
R^2	0.000856	0.005055	0.992460639426583	0.017276	0.006913	0.03398	0.004606	0.008167	0.002132	0.007691755	0.000243	0.000308
R^2 %	0.086	0.505	99.246	1.728	0.691	3.398	0.461	0.817	0.213	0.769	0.024	0.031
Observations	157	128	117	120	101	131	137	106	357	258	314	265

Pearson correlation between WS data values for the same date

WS	Area 1				Area 2				Area 3			
Time	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
Pearson r	0.090592768	0.239649621	0.009491323	0.103868269	0.179551239	0.013594717	0.017003542	0.070362159	0.043544701	0.002547826	0.010196259	0.027614004
P value (stat sig)	0.288866877	0.007345062	0.919809692	0.271434353	0.083340491	0.880876565	0.844813427	0.488898559	0.425577454	0.968416797	0.858549685	0.662646245
R^2	0.00820705	0.057431941	0.000090085205639	0.010788617	0.032238648	0.000184816	0.00028912	0.004950833	0.001896141	0.000006491419185	0.000103964	0.000762533
R^2 %	0.821	5.743	0.009	1.079	3.224	0.018	0.029	0.495	0.190	0.001	0.010	0.076
Observations	139	124	115	114	94	124	135	99	337	244	308	252

Determining the relationship between data based on distance away from the RAO

In order to understand the relationship between the RAO and CLIWOC ship log data as a function of distance, the difference in data value is plotted against the distance the ship is away from the RAO. The results are presented in Figures 21 to 23 for wind direction data, and Figures 24 to 26 for wind speed. The data difference is deduced from RAO and CLIWOC ship log data recorded on the same date (see "Determining data relationships based on distance away from the RAO" for the method used). This will give an indication of whether the weak correlation over time is distorted by the noise introduced by data correlations based on space. It is expected that the correlation in the data will deteriorate as distance increases. This test addresses the issues and limitations introduced by working within a large domain and with sub-scale atmospheric phenomena occurring inside the delimited areas. As with the previous section, this section will first address wind direction and then wind speed, both with associated statistical analyses presented in the tables.

Wind direction correlation in all areas as a function space

There are no significant signals in the data value differences and associated distance away from the RAO, in any of the defined regions of study and for any season. The data are scattered randomly along the x axis (as distance increases). This suggests that despite the distance between the RAO and ships, there is no relationship between the wind direction experienced at the RAO and onboard ships. An extreme difference in the data is +270°, an event that is present in most graphs. This decreases the correlation of data over space and time. The bulk of the differences fall in the positive range, showing that the ships observe wind more to the left (smaller decimal degrees than the recorded RAO data). This bias may be due to topographic effects at the RAO (see "Limitations" for more details; Peterson and Troen, 2012). There will be no further investigation into this bias. It is only important to know that there is a bias, that intensifies the weak correlation. Occasionally, the data value difference is 0°, meaning the ship and RAO recorded exactly the same wind direction for the same date. Due to the fact that this 0° difference occurs randomly over 800km and more, this may be coincidental. Such high variance in the scatter plot graphs in, Figures 21-23, indicates that there is no signal or correlation as a function of distance between directional data.

Area 1

In contrast to Area 2 and 3, Area 1 in Figure 21, shows ships up to 1000km away were recording wind data where other areas captured data up to 600km for Area 1 and 2. This is to do with the nature of shipping routes. Area 1 shows no statistically significant correlation between the data values and the distance away from the RAO (see Table 9). Interestingly, there is a conglomerate of data points within 100km of the RAO that have a directional difference under 90° for summer only. This suggests that in Area 1, the ships within 100km were consistently capturing similar atmospheric conditions to the RAO (within an acceptable range). This is likely because of the dominant SE'ly winds (see Figure 3 for a visualization of the typical conditions).

Area 2

Area 2 (Figure 22) also shows no significant correlation in the data. All seasonal correlations in the data are very weak (all $r<0,12$, all $p>0,05$). Summer shows a strong bias in the positive directional difference range, especially around 100km. Winter shows high variability in the wind directional differences for ships at variable distances away from the RAO. The data in Area 2 therefore does not show any correlation in the data.

Area 3

Area 3 (Figure 23) shows condensed data between 100km to 500km. This data feature is attributable to the ships that navigated around the southern continent, not a data value limitation. There is bias to the left in the ship logs as seen by the majority of positive differences, consistent with Area 1 and 2. Ridging high pressure cells can result in easterly flow in Area 3, adding to the increased left hand shear bias in the ship log data, causing the majority of the data to have a positive difference. The correlation between the data is very weak ($r<0,159$) for all seasons in Area 3 (Table 9).

Seasonal scatter plot graphs of the Δ distance and ΔWD value in Area 1

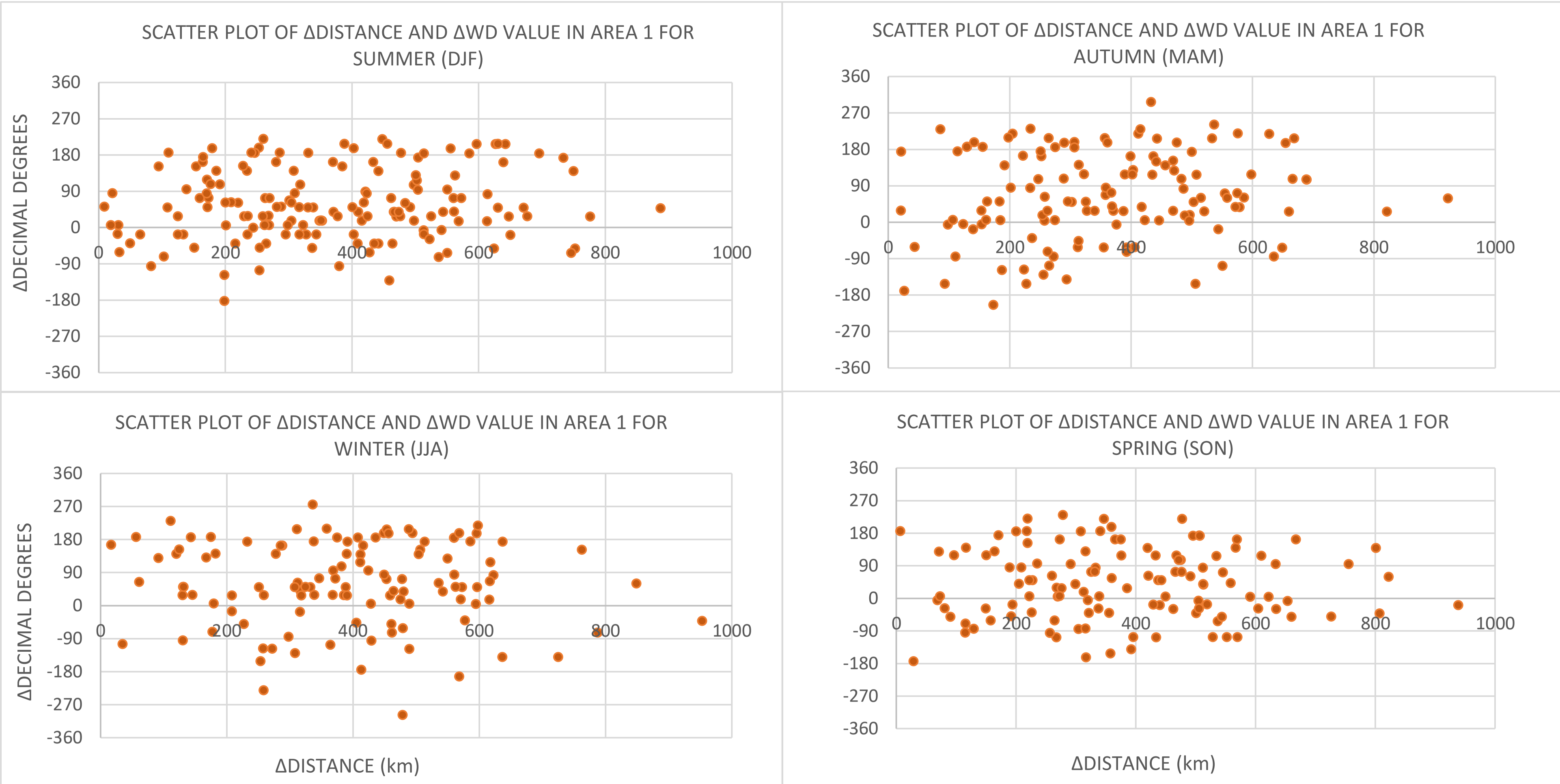


Figure 21 Area 1: Seasonal scatter plot graphs showing the relationship between wind direction value difference and the distance away from the RAO for data recorded on the same day from 1834-1854. There is no clear signal in the data, however it is evident that the CLIWOC data records wind directions more wind shear to the left by the positive average of data differences.

Seasonal scatter plot graphs of the Δ distance and ΔWD value in Area 2

Figure 22 Area 2: Seasonal scatter plot graphs showing the relationship between wind direction value difference and the distance away from the RAO for data recorded on the same day from 1834-1854. There is a noticeable aggregation of data at 100km away from the RAO in summer.

Seasonal scatter plot graphs of the Δ distance and ΔWD value in Area 3

Figure 23 Area 3: Seasonal scatter plot graphs showing the relationship between wind direction value difference and the distance away from the RAO for data recorded on the same day from 1834-1854. The bulk of the data occur within 100km and 500km of the RAO, an indication of the preferred sailing routes. The data show a high variance in value difference throughout the domain, and no clear signal in the relationship between the RAO and CLIWOC data.

Table 9 Statistical relationships for data value difference and distance away from the RAO. Pearson correlation (r), statistical significance (p value), correlation coefficient and percentage (R^2 and R^2%) and number of observations are given for each dataset comparison between RAO and each area (1-3) for each season. The are no significant statistical correlations. The four cases of statistically significant correlations (green blocks) are random, shown by the corresponding very weak r values and R^2% values. The (top) table comprises wind direction data and the (bottom) table comprises wind speed data.

Pearson correlation between Δdistance and ΔWD values of the RAO and ship log data

WD	Area 1				Area 2				Area 3			
Time	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
Pearson r	0.116946	0.141616	0.040117	0.007070805	0.124333	0.060192	0.094159	0.013187	0.123288	0.15583	0.158691	0.068179
P value (stat sig)	0.144668	0.110818	0.667598	0.938904361	0.215424	0.494638	0.273762	0.893274	0.019797	0.012205	0.004822	0.268762
R^2	0.013676	0.020055	0.001609	0.0000499963	0.015459	0.003623	0.008866	0.000174	0.0152	0.024283	0.025183	0.004648
R^2 %	1.37%	2.01%	0.16%	0.00%	1.55%	0.36%	0.89%	0.02%	1.52%	2.43%	2.52%	0.46%
Observations	157	128	117	120	101	131	137	106	357	258	314	265

Pearson correlation between Δdistance and ΔWS values of the RAO and ship log data

WS	Area 1				Area 2				Area 3			
Time	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
Pearson r	0.156359414	0.030211	0.260318	0.009282402	0.162377	0.048492	0.058021	0.029435	0.065143	0.010548	0.000444655	0.004046746
P value (stat sig)	0.066043803	0.739068	0.004959	0.921917368	0.117903	0.592768	0.503854	0.772412	0.23299	0.869796	0.993798951	0.949033142
R^2	0.024448266	0.000913	0.067765	0.0000861630	0.026366	0.002351	0.003366	0.000866	0.004244	0.000111	0.0000001977	0.0000163762
R^2 %	2.445%	0.091%	6.777%	0.009%	2.637%	0.235%	0.337%	0.087%	0.424%	0.011%	0.000%	0.002%
Observations	139	124	115	114	94	124	135	99	337	244	308	252

Wind speed correlation in all areas as a function space

Generally, the ships recorded stronger wind speeds. This is evident by the high number of data found in the negative range of speed differences in the scatter plot graphs in Figures 24, 25 and 26. Added to which, like the wind direction correlations, there is no defined or clear signal in the data correlation with increasing distance. That is to say, the distance away from the RAO does not control the difference in the wind speed values recorded. Area 3 (Figure 26) shows the most extreme variability in the data (albeit there is a higher data density in this region) with wind speed differences exceeding 20m/s frequently. Area 1 (Figure 24) shows a lower variability with wind speed differences mostly contained in the 10m/s difference range. Area 2 (Figure 25), like Area 3, shows more data in the wind speed difference range for winter and spring months (JJA and SON). Ships observed much higher wind speeds, as much as 25m/s, (See Figure 26, winter graph as an example) and more frequently in these regions and seasons. This may be due to intense sea storms caused by travelling mid latitude cyclones and the uninterrupted winds on the ocean surface. Additionally, winter winds may have a westerly component (see wind roses for RAO in Figure 14) and the RAO may be more sheltered by topography from the full force of wind (Figure 7). It is evident that the ships record higher wind speeds in general on the open ocean, thus the correlation between the data is weak in all regions and in all seasons (Table 9).

Area 1

The extended x axis (up to 1000km) for Area 1 in Figure 24, stretches the data differences further away from the RAO, in comparison to Area 2 and 3, but the range of differences is confined to around 10m/s. Of course there are more extreme values but these do not occur as often in comparison to the other areas. Less variability in the data spread and differences suggests that the data in Area 1 are better correlated. However, statistically, there is a very weak correlation, with no significance (for all seasons: $r<0,2$, $p>0,05$ except for winter, but $R^2\%= 6,78\%$) (Refer to statistical correlation summary in Table 9). We cannot assume that Area 1 has the best correlation based on least variance in data spread and difference because the complexities of atmospheric dynamics and data density in each area have not been factored into the calculation. However, the data are not well correlated in this area.

Area 2

There is no significant correlation between the data at the Rao and the ships, in this area. The seasonal recurrences in the spread of data are similar to the wind direction tests. For example, there is a concentration of data with very little value difference at 100km in summer (DJF), much like Figure 22. All statistical correlations show that the data in Area 2 and the RAO are weakly correlated ($r<0,16$, $p>0,05$) (Table 9).

Area 3

There is a wide range of value differences in Area 3 from 20m/s to -30m/s, partly attributed to the high density of data, but also because of the very different geographic location of the data and variable weather conditions. The data in Area 3 is susceptible to different atmospheric conditions than the RAO throughout the year for each season (See Table 5 for characteristics of this region). The scatter plot graphs confirm this. Summer shows the least variance (differences mostly <10m/s), and winter shows the most variance (high frequency of differences of up to 20m/s). Statistically, the correlations between the seasonal data are all weak and not significant ($r<0,06$ and $p>0,05$ for all seasons) (Table 9).

Seasonal scatter plot graphs of the Δ distance and ΔWS value in Area 1

Figure 24 Area 1: Seasonal scatter plot graphs showing the relationship between wind speed value difference and the distance away from the RAO for data recorded on the same day from 1834-1854. The majority of the data in this area record wind speed differences of 10m/s which is a significant increase/decrease. There is no clear tendency for the data differences to deteriorate with distance suggesting that the data is completely uncorrelated over Area 1 spatial domain.

Seasonal scatter plot graphs of the Δ distance and ΔWS value in Area 2

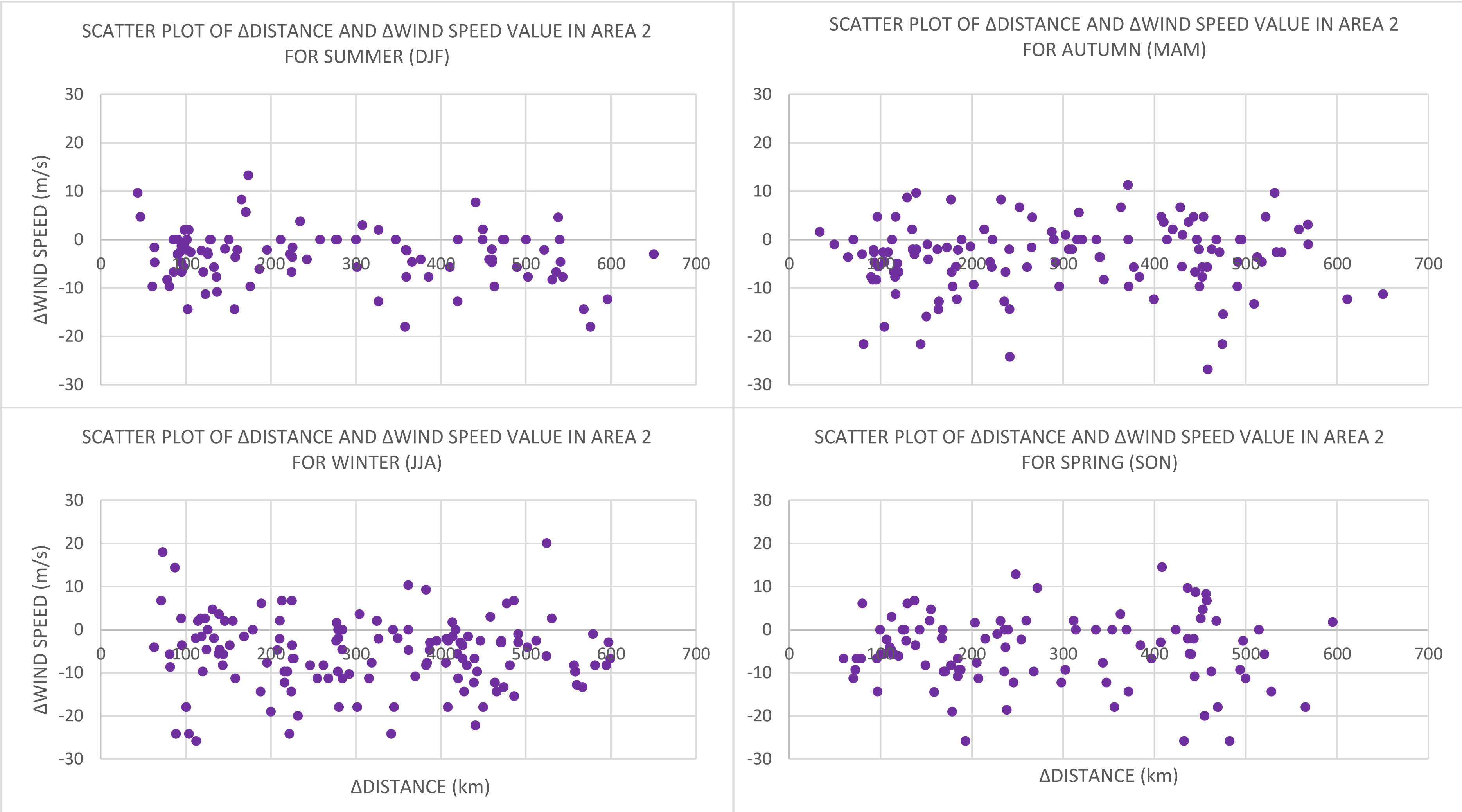


Figure 25 Area 2: Seasonal scatter plot graphs showing the relationship between wind speed value difference and the distance away from the RAO for data recorded on the same day from 1834-1854. The data show major differences in WS, as much as 25m/s and with higher variance compared to Area 1. There is no clear tendency for the data differences to deteriorate with distance suggesting that the data is completely uncorrelated over the Area 2 spatial domain.

Seasonal scatter plot graphs of the Δ distance and ΔWS value in Area 3

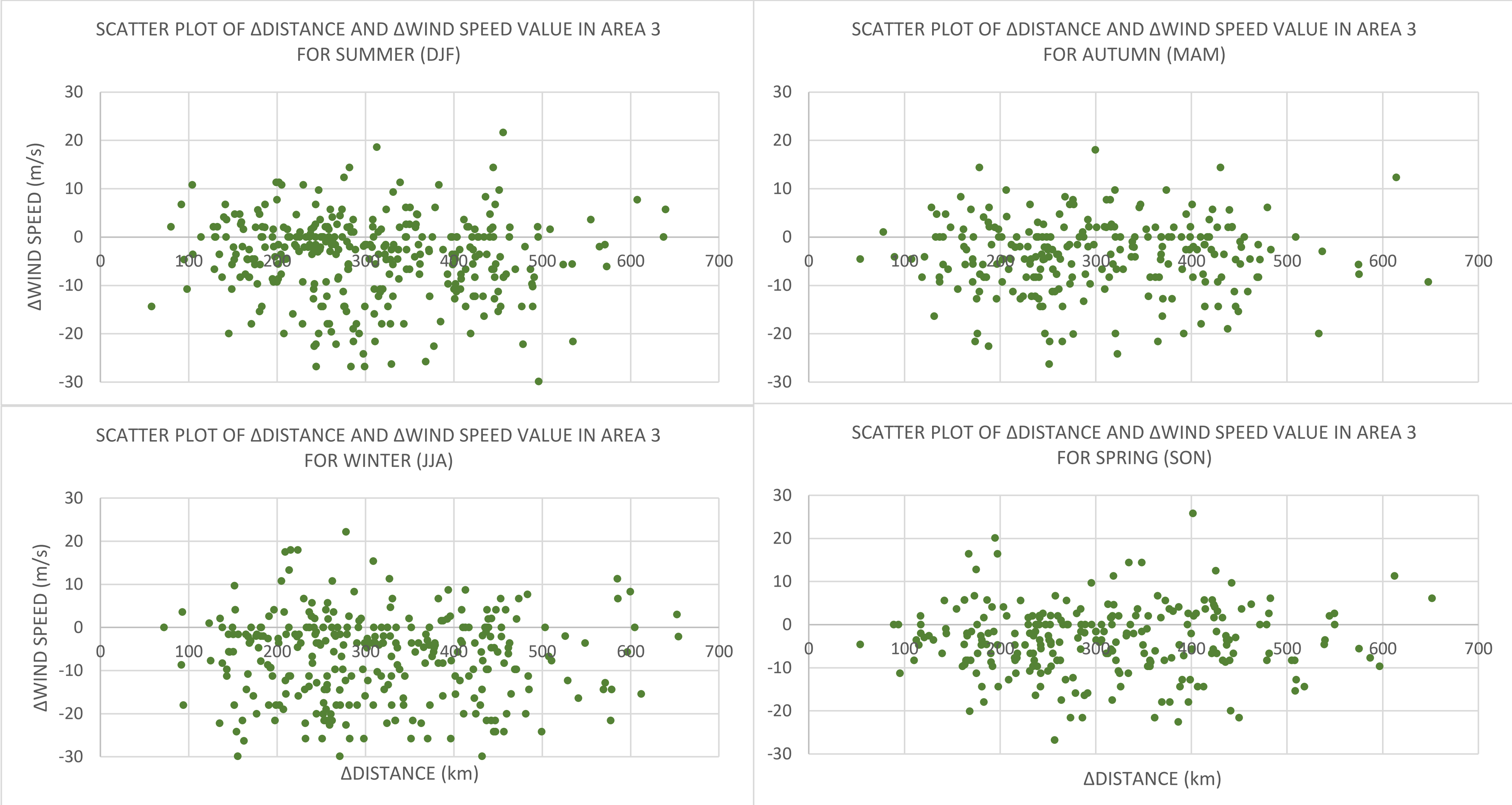


Figure 26 Area 3: Seasonal scatter plot graphs showing the relationship between wind speed value difference and the distance away from the RAO for data recorded on the same day from 1834-1854. There are major differences in the data values (<30m/s) throughout the domain. This suggests that data captured on the same date was vastly different in Area 3 compared to the RAO. Interestingly winter shows the highest average wind speed differences, inferring that ships recorded much higher winds on the south coast, probably during the treacherous gale force storm seas surges (Turner, 1988).

Conclusion

There is no statistically significant correlation between wind data from the RAO and the CLIWOC ship logs. Scatter plot graphs showing the difference in values against the distance between the sources of data show no clear signal or similarities over time that indicate a correlation as a function of distance. The graphs prove that location, season and observer bias all affect the spread and variance of data. These factors complicate and weaken the correlation, but do not invalidate the data. Possible future studies may want to use higher grade statistics such as a multi-variate correlation analysis such as a Principal Component Analysis to separate the noise from absolute data in order to extract proportionally weighted signals (Hannaford *et al.*, 2015). Time did not allow for this during this project.

Dominant wind vector distributions for ship log data

Considering there is no correlation in the data based on time and space, it is interesting then to understand if the three selected regions of study over the ocean show any similarity in the wind vectors. The correlation tests and scatter plot graphs expose the weak correlations over time and space in the data. However, due to the complexities of atmospheric dynamics it is still possible that the data captures similar wind directions and speeds at different times.

This vector distribution analysis downscales the data analysis into the three defined regions and for each season. What is highlighted the most in these heat maps is the similarity in the seasonality. The dominant seasonal wind patterns are exposed in the ship log data with the added unique conditions for each zone (explained later). The synoptic conditions can be seen in the frequency of typical seasonal wind directions. The wind speed classes are generally higher than the RAO recorded. This is in agreement with scatter plot graphs in the previous section on value difference. The frequency distribution of wind classes and vectors is in agreement with the typical synoptic scale atmospheric conditions. This adds more weight to the reliability of the CLIWOC data, and RAO data.

Synoptic characteristics are evident in the wind vector frequency

In summer, Area 1 shows higher frequency of SE'ly wind (66,27%) whereas Area 3 shows a higher frequency of SW'ly winds (50,12%). Area 2 shows equal frequencies for SE and SW (38,51% and 37,27% respectively). These conditions are expected and typical of the summer atmospheric dynamics in the respective regions. In the winter months (JJA), Area 2 and Area 3 very high frequencies of winds with a westerly component, which is linked to winter mid latitude cyclones and synoptic scale atmospheric conditions. Area 2 and 3 have very high frequencies of SW'ly winds of all wind speed classes. The occurrence of strong winds is due to the uninterrupted low pressure cyclones that propagate through these regions but seldom move directly through Area 1. In Area 1 during winter, anomalous 48,80% SE'ly winds are present. This is an example of the type of anomaly one must be cautious of when working with historical data.

Autumn and spring show a spread of wind direction frequencies but lower wind speed classes, in comparison to summer and winter. A dominant SE'ly component exists in all areas for these equinox seasons. Spring shows a higher frequency of NW'lies, compared to autumn that shows a higher frequency of SE'ly and Sw'lies. This indicates that the spring data reflect the less impactful mid-latitude cyclones after winter, where autumn does not. Largely the frequency distribution tables per region for each season are in agreement with the general climatology established at the RAO.

A limitation discovered in this section is the data density per region. Area 1 results are affected by the low number of data values in comparison to Area 3 that is densely populated. The data density affects the frequency distribution tables, so caution must be paid to the relative difference in frequencies per region. For example, Area 1 summer only contains 83 data values. And while the typical atmospheric conditions are present, there is less detail in the data. This is in comparison to Area 3 summer time that has 409 data values, and a more useful distribution table.

Table 10 Each heat map represents the frequency distribution of wind vector information (direction and speed) recorded in Area 1 for each season in the CLIWOC dataset. Summer (top left) shows a lack of data, nevertheless the typical summer SE'er is captured. Winter (bottom left) captures the typical NW'ly winds. Interestingly there are many accounts of a SE'er, but the high frequency 8-10.7m/s winds suggests that these were storm winds possibly associated with travelling low pressure systems (Tyson and Preston-Whyte, 2004).

Frequency of WS for 4 compass quadrants in Area 1 per season

DJF	NE	SE	SW	NW	No. of events	% of events
0	0	0	0	0	0	0,00
0.1 to 1.5	0	4	4	0	8	9,64
1.6 to 3.3	0	5	6	0	11	13,25
3.4 to 5.4	0	9	7	0	16	19,28
5.5 to 7.9	0	11	2	1	14	16,87
8 to 10.7	1	18	4	1	24	28,92
10.8 to 13.8	0	6	2	0	8	9,64
13.9 to 17.1	0	2	0	0	2	2,41
17.2 to 20.7	0	0	0	0	0	0,00
20.8 to 24.4	0	0	0	0	0	0,00
24.5 to 28.4	0	0	0	0	0	0,00
28.5 to 32.6	0	0	0	0	0	0,00
32.7 +	0	0	0	0	0	0,00
No. of events	1	55	25	2	**83**	
% of events	1,20	66,27	30,12	2,41		

MAM	NE	SE	SW	NW	No. of events	% of events
0	0	3	0	0	3	2,52
0.1 to 1.5	0	6	4	1	11	9,24
1.6 to 3.3	2	8	3	3	16	13,45
3.4 to 5.4	1	9	4	2	16	13,45
5.5 to 7.9	3	16	6	4	29	24,37
8 to 10.7	2	18	9	0	29	24,37
10.8 to 13.8	0	5	4	1	10	8,40
13.9 to 17.1	0	2	1	0	3	2,52
17.2 to 20.7	0	0	0	0	0	0,00
20.8 to 24.4	0	0	1	0	1	0,84
24.5 to 28.4	0	1	0	0	1	0,84
28.5 to 32.6	0	0	0	0	0	0,00
32.7 +	0	0	0	0	0	0,00
No. of events	8	68	32	11	**119**	
% of events	6,72	57,14	26,89	9,24		

JJA	NE	SE	SW	NW	No. of events	% of events
0	0	1	0	0	1	0,60
0.1 to 1.5	1	1	5	5	12	7,23
1.6 to 3.3	1	11	7	5	24	14,46
3.4 to 5.4	1	16	8	6	31	18,67
5.5 to 7.9	1	9	10	4	24	14,46
8 to 10.7	3	22	7	5	37	22,29
10.8 to 13.8	3	17	2	1	23	13,86
13.9 to 17.1	1	3	1	1	6	3,61
17.2 to 20.7	0	1	1	2	4	2,41
20.8 to 24.4	0	0	0	2	2	1,20
24.5 to 28.4	0	0	0	0	0	0,00
28.5 to 32.6	0	0	2	0	2	1,20
32.7 +	0	0	0	0	0	0,00
No. of events	11	81	43	31	**166**	
% of events	6,63	48,80	25,90	18,67		

SON	NE	SE	SW	NW	No. of events	% of events
0	0	1	1	1	3	2,46
0.1 to 1.5	1	2	3	1	7	5,74
1.6 to 3.3	2	10	5	0	17	13,93
3.4 to 5.4	0	6	11	2	19	15,57
5.5 to 7.9	0	16	8	2	26	21,31
8 to 10.7	0	17	6	2	25	20,49
10.8 to 13.8	0	14	6	0	20	16,39
13.9 to 17.1	0	2	0	0	2	1,64
17.2 to 20.7	0	1	0	1	2	1,64
20.8 to 24.4	0	1	0	0	1	0,82
24.5 to 28.4	0	0	0	0	0	0,00
28.5 to 32.6	0	0	0	0	0	0,00
32.7 +	0	0	0	0	0	0,00
No. of events	3	70	40	9	**122**	
% of events	2,46	57,38	32,79	7,38		

Table 11 Each heat map represents the frequency distribution of wind vector information (direction and speed) recorded in Area 2 for each season in the CLIWOC dataset. Winter represents the typical winter time times with a westerly component. Summer is predominantly a southerly wind. These heat maps match the typical synoptic scale atmospheric flow regimes and are in agreement with the general climatology of the RAO.

Frequency of WS for 4 compass quadrant for Area 2 per season

DJF	NE	SE	SW	NW	No. of events	% of events
0	0	0	1	0	1	0,62
0.1 to 1.5	2	5	3	3	13	8,07
1.6 to 3.3	1	3	7	3	14	8,70
3.4 to 5.4	6	13	11	4	34	21,12
5.5 to 7.9	2	11	7	6	26	16,15
8 to 10.7	1	17	11	4	33	20,50
10.8 to 13.8	1	8	13	5	27	16,77
13.9 to 17.1	0	4	3	1	8	4,97
17.2 to 20.7	0	1	3	0	4	2,48
20.8 to 24.4	0	0	1	0	1	0,62
24.5 to 28.4	0	0	0	0	0	0,00
28.5 to 32.6	0	0	0	0	0	0,00
32.7 +	0	0	0	0	0	0,00
No. of events	13	62	60	26	**161**	
% of events	8,07	38,51	37,27	16,15		

MAM	NE	SE	SW	NW	No. of events	% of events
0	1	0	0	0	1	0,76
0.1 to 1.5	4	4	1	0	9	6,82
1.6 to 3.3	4	5	7	0	16	12,12
3.4 to 5.4	7	13	9	0	29	21,97
5.5 to 7.9	1	11	9	0	21	15,91
8 to 10.7	2	14	6	0	22	16,67
10.8 to 13.8	2	10	3	0	15	11,36
13.9 to 17.1	2	2	3	0	7	5,30
17.2 to 20.7	0	1	1	0	2	1,52
20.8 to 24.4	0	3	1	2	6	4,55
24.5 to 28.4	0	1	0	0	1	0,76
28.5 to 32.6	0	3	0	0	3	2,27
32.7 +	0	0	0	0	0	0,00
No. of events	23	67	40	2	**132**	
% of events	17,42	50,76	30,30	1,52		

JJA	NE	SE	SW	NW	No. of events	% of events
0	0	0	0	0	0	0,00
0.1 to 1.5	1	4	7	0	12	6,25
1.6 to 3.3	1	4	10	11	26	13,54
3.4 to 5.4	1	5	10	6	22	11,46
5.5 to 7.9	1	7	6	10	24	12,50
8 to 10.7	4	5	15	9	33	17,19
10.8 to 13.8	3	6	14	12	35	18,23
13.9 to 17.1	1	4	3	5	13	6,77
17.2 to 20.7	1	3	2	3	9	4,69
20.8 to 24.4	0	0	6	3	9	4,69
24.5 to 28.4	1	0	2	4	7	3,65
28.5 to 32.6	0	0	1	1	2	1,04
32.7 +	0	0	0	0	0	0,00
No. of events	14	38	76	64	**192**	
% of events	7,29	19,79	39,58	33,33		

SON	NE	SE	SW	NW	No. of events	% of events
0	0	0	0	0	0	0,00
0.1 to 1.5	1	4	0	0	5	4,00
1.6 to 3.3	4	4	4	2	14	11,20
3.4 to 5.4	3	7	7	2	19	15,20
5.5 to 7.9	4	9	8	2	23	18,40
8 to 10.7	0	12	7	4	23	18,40
10.8 to 13.8	0	11	7	3	21	16,80
13.9 to 17.1	0	4	0	0	4	3,20
17.2 to 20.7	0	5	2	1	8	6,40
20.8 to 24.4	0	1	2	0	3	2,40
24.5 to 28.4	0	2	0	3	5	4,00
28.5 to 32.6	0	0	0	0	0	0,00
32.7 +	0	0	0	0	0	0,00
No. of events	12	59	37	17	**125**	
% of events	9,60	47,20	29,60	13,60		

Table 12 Each heat map represents the frequency distribution of wind vector information (direction and speed) recorded in Area 3 for each season in the CLIWOC dataset. The high data density of Area 3 enhances the seasonal differences in wind flow regimes which are slightly different to Area 1 and Area 2. Area 3 shows winds with a more westerly component in all seasons due to the flow of air along the continent as ridging high pressures and travelling disturbances ridge parallel to the continental mass (Tyson and Preston-Whyte, 2004).

Frequency of WS for 4 compass quadrants for Area 3 per season

DJF	NE	SE	SW	NW	No. of events	% of events
0	0	7	4	1	12	2,93
0.1 to 1.5	9	11	26	4	50	12,22
1.6 to 3.3	14	13	19	6	52	12,71
3.4 to 5.4	14	23	28	4	69	16,87
5.5 to 7.9	13	15	23	3	54	13,20
8 to 10.7	9	15	16	5	45	11,00
10.8 to 13.8	7	8	24	4	43	10,51
13.9 to 17.1	4	4	25	3	36	8,80
17.2 to 20.7	0	1	12	3	16	3,91
20.8 to 24.4	0	0	16	2	18	4,40
24.5 to 28.4	1	0	9	0	10	2,44
28.5 to 32.6	0	0	2	1	3	0,73
32.7 +	0	0	1	0	1	0,24
No. of events	71	97	205	36	**409**	
% of events	17,36	23,72	50,12	8,80		

MAM	NE	SE	SW	NW	No. of events	% of events
0	0	3	1	2	6	1,76
0.1 to 1.5	3	4	10	3	20	5,88
1.6 to 3.3	11	12	27	12	62	18,24
3.4 to 5.4	13	14	29	8	64	18,82
5.5 to 7.9	10	8	16	7	41	12,06
8 to 10.7	20	15	21	6	62	18,24
10.8 to 13.8	6	2	15	8	31	9,12
13.9 to 17.1	4	1	10	5	20	5,88
17.2 to 20.7	2	5	4	1	12	3,53
20.8 to 24.4	0	3	10	3	16	4,71
24.5 to 28.4	0	0	1	3	4	1,18
28.5 to 32.6	0	0	2	0	2	0,59
32.7 +	0	0	0	0	0	0,00
No. of events	69	67	146	58	**340**	
% of events	20,29	19,71	42,94	17,06		

JJA	NE	SE	SW	NW	No. of events	% of events
0	1	0	2	2	5	1,15
0.1 to 1.5	14	5	12	12	43	9,93
1.6 to 3.3	5	9	22	14	50	11,55
3.4 to 5.4	15	8	22	27	72	16,63
5.5 to 7.9	11	1	22	11	45	10,39
8 to 10.7	7	5	14	9	35	8,08
10.8 to 13.8	7	7	21	13	48	11,09
13.9 to 17.1	1	2	18	10	31	7,16
17.2 to 20.7	2	2	21	5	30	6,93
20.8 to 24.4	0	1	25	12	38	8,78
24.5 to 28.4	0	1	11	13	25	5,77
28.5 to 32.6	0	0	4	7	11	2,54
32.7 +	0	0	0	0	0	0,00
No. of events	63	41	194	135	**433**	
% of events	14,55	9,47	44,80	31,18		

SON	NE	SE	SW	NW	No. of events	% of events
0	2	1	2	2	7	2,45
0.1 to 1.5	3	4	12	8	27	9,44
1.6 to 3.3	6	15	13	5	39	13,64
3.4 to 5.4	9	11	15	4	39	13,64
5.5 to 7.9	10	9	17	3	39	13,64
8 to 10.7	20	23	10	3	56	19,58
10.8 to 13.8	4	9	10	3	26	9,09
13.9 to 17.1	2	2	18	3	25	8,74
17.2 to 20.7	1	0	9	1	11	3,85
20.8 to 24.4	0	1	8	4	13	4,55
24.5 to 28.4	0	0	4	0	4	1,40
28.5 to 32.6	0	0	0	0	0	0,00
32.7 +	0	0	0	0	0	0,00
No. of events	57	75	118	36	**286**	
% of events	19,93	26,22	41,26	12,59		

Conclusion

The various heat maps/ distribution tables for each region highlight the typical synoptic atmospheric conditions expected within the region and per season. Despite low data density, Area 1 shows a summer SE'ly and NW'ly winds in winter. Autumn and spring largely show a spread of wind vectors in lower range wind speed classes. This is in agreement with the RAO general climatology. Thus, despite the weak correlation between the data, the distribution tables prove that synoptic conditions, relative to the area and season are traceable through data analysis.

Discussion

In this preliminary analysis of the RAO data it has been discovered that the wind data is a rich source of high density, continuous data for South Africa (4 readings per day from 1834 to 1976, including additional Term Days of hourly recordings). This provides 66 136 data values to establish a general understanding of the historical wind climatology for Cape Town, South Africa. For more information about the full length time series analysis see the adjacent study by Picas (2015). Digitisation and pre-processing of the RAO data is a laborious task and requires hours of attention to the detail in the written documents, but this necessary methodological step provides a fundamental understanding of the nature of the data.

Exploring the data

In this research project a qualitative analysis has uncovered that the measurement methods and codes were equivalent to the mariners' ship logbooks (CLIWOC dataset), except the wind direction which was recorded as true north in the RAO registers and magnetic north in the ship logbooks. However, Herrera-Garcia *et al.* (2005) explain that the CLIWOC wind direction data was subsequently corrected to true north. Both data sources contained cardinal compass wind direction around a 32-point compass, which were converted to decimal degrees. Both datasets used the Beaufort Scale to classify wind strengths, according to a prescribed visual interpretation code (Figure 5), and converted to measurements in meters per second. Missing data in the RAO dataset could be attributed to human or instrumental errors, and were accounted for. 11,4% of all the RAO data from 1834 – 1854 was missing, and only 7,4% of all CLIWOC-matched noon observations were missing. Thus, the initial steps to digitize and standardize the RAO dataset depended on a qualitative approach to analyzing the meteorological documents, suggested by Trenberth *et al.* (2002).

For the comparative analysis of the data, it is beneficial that the data share measurement conventions. Although there is inherent error in converting data from a Beaufort scale code to discrete measurements in meters per second, the two datasets carry the same amount of conversion error. This makes them comparable in relative terms. The same occurs for wind direction, where 32 cardinal compass directions are converted to decimal directions. This project exploits the oldest historical wind data from two separate sources that exist in a 20-year period of 1834 to 1854, a fortunate overlap period for the same region. However, despite the advantages in the similarities in the datasets, it is the limitations that guided the methodology of the data comparative analysis.

Each dataset carries its own limitations and errors that cascade into the data analysis (Brohan *et al.*, 2006). This adds additional error and noise to the existing limitations involved in comparing land based and ocean based wind data, that differs greatly over large spatial domains, and with elapsed time. These limitations have guided the methodology in order to minimize cascading error and optimize the value of the inter-comparison (decrease uncertainty). The initial limitations in the CLIWOC data include a) noon only observations, b) insufficient CLIWOC data within a 550km circumference of the RAO, a meso-scale grid prescribed by Bonnardot *et al.* (2008), c) multiple contrasting ship observations on one day and (Figure 8 and 9) d) inconsistent time steps between observation days and variable consecutive days of observation (Table 4). With approximately 1500 yearly data values in the RAO dataset, and only about 500 yearly data values in the CLIWOC dataset, the RAO data could be refined to mirror the CLIWOC dataset by selecting the exact dates for complete wind vector information (wind speed and direction) and including only noon readings. The CLIWOC data is spatially limited to three most suitable regionalized areas around the south western coast. These are Area 1, Area 2 and Area 3 as seen in Figure 11. All analyses are limited to seasonal analyses to enhance synoptic scale conditions. Thus, an inter-comparison between the data is possible by working within the bounds of the limitations in space and time.

Analysis of the data

The results show that overall, there is no correlation between the RAO data and CLIWOC ship log data across all areas and through the whole time period analysed. There is no statistical correlation or significant signal or recurring sequence in the data over time or with distance away from the RAO. It is highly likely that the complex atmospheric dynamics functioning over a range of variable spatial and temporal scales result in an absence of statistical and observed data correlation. Therefore, it is possible that certain scenarios may have a weak, or no correlation but that are plausible. See Figure 4, 8 and 9 as examples of very contrasting observations for the same date. For this reason, the weak correlation is not an indicator of data validity, but rather an indication that the CLIWOC data and RAO data cannot be assumed to reflect each other in future research.

However, the different oceanic regions and the RAO are showing expected seasonal signals of wind regimes for the respective areas in the wind vector analyses. Summer south easterlies are dominant in Area 1 and 2 with summer SE'lies making up 66,27% and 38,51% of observed winds respectively. Area 3 experiences predominantly south westerly summer winds (50,12%). Winter winds are predominantly SE (48,8%) and NW (8,67%) in Area 1, with 22,28% of the winds being recorded above 10,8m/s. Area 2 shows 39,58% SW'lies and 33,33% NW'lies, and 39,07% of winds were recorded above 10,08m/s. Area 3 shows 44,8% SW'lies and 31,18% NW'lies, and 9,61% of the winds were above 10,8m/s. This indicates a high frequency of observed storm weather associated with winds with a westerly component in the CLIWOC data. These are all in agreement with typical seasonal air flow regimes.

In conjunction with the RAO data, the seasonal wind patterns are similar (refer to seasonal wind roses in Figure 14). The RAO shows a high frequency of southerly winds and observed SE'lies which were the majority of the observed summer in Area 1 (66,27%) and 2 (38,51%). Area 3 shows 50,12% SW'ly wind and 23,17% SE'lies. Winter shows a high frequency of NW'lies at the RAO, which agrees with the CLIWOC winter winds (that also show SW'lies). A dissimilarity is the occurrence of the NW'lies that only account for less than 16,15% of the summer NW'lies in the CLIWOC data. Another dissimilarty is that the RAO does not capture SSw'lies in winter months, that are very clearly evident in the oceanic regions. This may be due to wind shadowing effect due to the complex Cape Peninsula topography.

Area 1 and 2 have a predominantly SE'ly wind throughout the year. Each season experiences >48,8% SE'lies in Area 1, and >38,51% for Area 2. Area 3 experiences >42,9% of SW'ly wind in each season. This is explainable by the divergence of south easterly air flow along the southern continent in summer, spring autumn and spring; and the travelling low pressure mid-latitude cyclones in winter (Tyson and Preston-Whyte, 2004). Again for all areas including the RAO, the winter months show very strong winds from the north west and spring shows a lower frequency of strong winds but still from the north west. This indicates trailing and dissipating mid-latitude cyclones as the sub polar low pressure belt recedes further south. Area 3 is an exception again as there is a higher frequency of south westerly winds. Again, this is attributable to the nature of ridging high pressure that follow behind passing mid-latitude cyclones (Tyson and Preston-Whyte, 2004).

Despite the very weak correlations in the data, there is still reason to suggest that the data are not incorrect. The synoptic scale air flow regimes are captured in all areas and for all seasons which suggests that the data are reliable. Notwithstanding, irregularities do still exist in the data, such as the north westerly winds experienced at the RAO in the summer months. It is suggested that these irregularities are the result of micro-scale, diurnal sea breezes as observed and modelled by Bonnardot *et al.* (2015), and that which might only be present at the RAO because of its close proximity to Table Bay and the slopes of Devil's Peak mountain. The journals kept by Jan van Riebeeck in the mid-1600s have provided some evidence that NW'ly winds were captured in summer months in the past (Thom and Balkema, 1952). It is strongly suggested that the anomalies are considered and accounted for in the pre-processing of historical data to understand the extent of uncertainty in contrast to sub-scale phenomena in the results by consulting various mediums of information like documentary data as well as recorded observations (Trenberth *et al.*, 2002). However, despite the irregularities, the synoptic atmospheric conditions are still defined in the wind roses in Figure 14 for the RAO and the wind vector frequency distribution tables (Tables 10, 11 and 12) for CLIWOC data. Therefore, the data are capturing the correct synoptic atmospheric conditions, but do not correlate well over time and space. The RAO are capturing regional winds, that which might add noise to the correlation due to the sub-scale nature of the breeze development.

Supplementary findings in CLIWOC data

It is also interesting that in the data there are indications that the nature of sailing affects what available data there is. For example, the spatial extent of data in Area 1 far exceeds the spatial extent in Area 2 and 3. Area 1 contains data from over 950km away from the RAO, whereas Area 2 and 3 only contain data just under 800km away. This is to do with the routing of sailing around the Cape Peninsula (see Hannaford's et al (2015) map of the CLIWOC data in Figure 10). This kind of spatial extent difference alters the correlation in the data. Hence, it is beneficial to adopt a quantitative and qualitative systems based approach, as Trenberth *et al.* (2002) describes, in order to understand the integrity of the different data. This approach can supplement the intrinsic value of the data.

It is also noticeable that ships within a 100km of the RAO mostly record wind direction within 90° difference and wind speed within 10m/s difference (see summer scatter plot graph in Figure 21). This denotes that the ships closer to Cape Town are capturing similar atmospheric conditions at the same time as the RAO, with an acceptable range of value difference. This adds weight to the argument that the differences and non-correlations over time and space are a function of the complexities of atmospheric dynamics, and not inaccurate observation. The observations both at the RAO and onboard ships are deemed to be accurate given the sophistication and availability of instruments. This is attributed to the training of meteorologists at the RAO and having a necessary code onboard ships disseminated by the meteorological board in London, which is found in the fore notes of the RAO meteorological registers (see "A description of RAO data" and Figures 5 and 6) (Warner, 2012; Sabine, 1851).

Reflecting on the project: what was learnt

This project has shown that recovering available historical data in Southern Africa is important due to the paucity of historical instrumental data and long chronologies of climate data (Vogel, 1989). The RAO provides one such source of long term continuous instrumental data. The CLIWOC database provides an alternative source of climate data that can be exploited like Hannaford *et al.* (2015) showed in reconstructing historical precipitation patterns over the southern continent. The accuracy of the data is acceptable for analysing in situ, and seasonal signals and events. Therefore, the newly digitized wind data from the RAO may be used for alternative analyses and in conjunction with data of another weather variable. Additionally, the data can be used to improve the likes of global reanalysis data like the Atmospheric Circulation Reconstruction over the Earth (ACRE) (for more details on this historical reanalysis project see website at http://www.met-acre.org/).

It was learnt that the data becomes more reliable and more valuable when parametrized by geographical location and seasons. This is due to the range of multi-variable temporal and spatial scales of atmospheric dynamics and events. The parameterization of the data highlights that any future researchers using historical data must be careful of analyzing the data relative to what the research question is. The weak and absent correlations in the datasets mean that signals and relationships between the data will be difficult to extract when analyzing the data cumulatively. Rather, researchers should be aware that the data from the RAO and CLIWOC dataset are subjected to unique complexities and therefore, unique methods of analysis and interpretation must be adopted for a reliable and robust assessment. Future researchers should be aware that the data are reliable for understanding climatology over time, because the data values are accurate. Only the correlation between data over time and space is not significant enough for inter-comparisons.

The limitations in this study have shown that the method of analysis is important and unique to the research question. A qualitative analysis of the data sources allows for an understanding of where to expect inaccuracies in the data. However, fortunately in the time period of this study there have not been any instrumental changes or measurement changes, which simplifies the conversion and modification of the data. Additionally, pre-processing the data in a way that will create two homogenous datasets for inter-comparison without jeopardizing the integrity of the data is important. This is why it is also important to frame the research question well and select data appropriately for the optimal results (least calculation error). The likes of the Southern Annular Mode were not investigated in this project and therefore conclusions about low amplitude temporal scale phenomena cannot be made. A time series analysis could not be performed due to the nature of inconsistent data over time, and sporadically spread data over the oceans. These limitations in the data and results do not hinder our ability to extract some useful information or invalidate the data.

There is no other source of wind data for this time period in the area and so the results drawn out in this project cannot be compared and validated. This project in itself was a way of successfully validating the newly digitized RAO data against CLIWOC data. It was discovered that there is no correlation between the CLIWOC data and the RAO over time and space, but the similar wind vector frequencies show that the data capture typical synoptic scale atmospheric conditions. This is a crude validation of historical wind data for the Southern African continent in itself and an assessment of the poor extent to which the CLIWOC ship log data reflect the RAO data.

Conclusions

This project has successfully created a digital wind dataset for the Royal Astronomical Observatory that is compared to the only other known source of historical wind data for the time and region. Hannaford *et al.* (2015) uses CLIWOC wind data retrieved from ship logbooks surrounding the south western coast of South Africa, to statistically reconstruct precipitation patterns. Hannaford *et al.* (2015) have motivated the research question of the extent to which ocean based CLIWOC data represents land based RAO data. This project is a part of a bigger project that has recovered and analysed the full timeline of instrumentally recorded historical wind data from the RAO. This is in light of the fact that historical data, let alone wind data, is seldom available or reliable enough for comparative analyses in Southern Africa. Also, wind as an atmospheric variable and air flow indicator for South Africa has not been extensively studied in historical climatology. Therefore, the caveat in the knowledge about historical data sources and wind data provides an opportunity to exploit the newly found historical wind datasets from the international CLIWOC dataset and local RAO dataset and perform a comparative analysis.

The limitations in the dataset are what have guided the development of the methodology. A method was developed to work within the constraints of the limitations, but also in order to optimise the chance of finding correlations across temporal and spatial scales. This was a challenge given the high variability in the atmospheric dynamics that occur in the mid-latitudes. The ability to compare data from a stationary point to points within large domains also introduced a limitation to the comparison and correlation of the datasets. Lastly the individual datasets carried their own limitations such as differing data density and variable consistency in recorded data. However, the datasets were refined to mirror each other for the exact dates and complete wind vector information (direction and speed) for noon readings and for each season. The CLIWOC data were regionalised into three geographically significant zones around the south western tip of Southern Africa. Correlations over time and space as well as vector distribution provide a comprehensive understanding of the relationships between the CLIWOC and RAO data.

The results found that the data does not correlate over space and time, and that only three clear signals exist in the data:

i) Ship logs record higher wind speeds,

ii) There is general directional wind shear to the left in the CLIWOC data,

iii) Only CLIWOC data within 100km from the RAO recorded similar wind observations (within an acceptable deviation of 90° for the same date).

Data are not well correlated due to the limitations in the data integrity and the complexity of comparing a highly variable atmospheric variable over a large domain. Even so, the results show that although there is absolutely no significant statistical correlation between the datasets or clear signal in the data, the data are still accurate in their absolute values. This is evident by the typical seasonality shown in wind vector distributions tables for each regionalised area and the RAO. Summer winds were predominantly winds with a southerly component, and winter winds were predominantly a westerly component, with a high frequency of high wind speeds that are linked to severe winter mid-latitude storms. Area 3, on the south coast, shows more westerly winds which represent the typical air flow for the region as ridging high pressures and trailing ends of mid-latitude cyclone systems pass through the region. The evident seasonality shows that the observers were capturing data accurately and reliably.

The anomalous north westerly winds observed at the RAO during summer are not a typical summer season feature. An investigation into monthly wind direction frequency suggests that the NW'lies were a result of passing mid-latitude cyclones that can pass over the continent occasionally. Jan van Riebeeck's journals provide historical evidence of NW'ly winds being recorded in Cape Town in the mid-1600s (Thom and Balkema, 1952). This example of anomalous data highlights that caution must be paid to the complexity of data and that a combined qualitative and quantitative approach to historical climatology is integral to understand the data, any sources of possible error and provide suitable solutions (Trenberth *et al.*, 2002).

This study shows that there are many inconsistencies and limitations that have to be treated with caution when handling historical climate data. Furthermore, future research should use the data appropriately to suit the research question. In conclusion, the CLIWOC data do not correlate with RAO data, but the data are still correct in their absolute values.

References

Websites

19th Century Ship Portraits in Print. n.d. http://www.19thcenturyshipportraitsinprints.com/. Artist: T. G. Dutton, coloured lithograph (circa 1848), Royal Museums Greenwich, London (0601). [14 August 2016].

Atmospheric Circulation Reconstructions over the Earth. n.d. http://www.met-acre.org/ [14 August 2016].

Cameron Beccario. 2016. *Earth: A Visualisation of Global Weather Conditions.* https://earth.nullschool.net/ [14 August 2016].

City of Cape Town, Disaster Risk Management: Storms and Gale Force Wind. Available: https://www.capetown.gov.za/en/DRM/Pages/StormsandHighorGaleforceWind.aspx [28 February 2017].

Climate and Development Knowledge Network. 2014. *The IPCC's Fifth Assessment Report: What's in it for Africa? Executive Summary.* July 2014. Available: http://cdkn.org/wp-content/uploads/2014/04/J1731_CDKN_FifthAssesmentReport_WEB.pdf [28 February 2017].

Climate System Analysis Group. n.d. *Climate Change in Southern Africa and Associated Impacts.* Available: http://media.csag.uct.ac.za/faq/qa_3impacts.html [28 February 2017].

GFS hourly modelled surface Wind. 2016. http://passageweather.com/ [14 August 2016].

Intergovernmental Panel on Climate Change. n.d. Fifth Assessment Reports (AR5). https://ipcc.ch/report/ar5/ [28 February 2017].

Met Office: Beaufort Wind Force Scale. March 2016. http://www.metoffice.gov.uk/weather/marine/guide/beaufortscale.html#moContent [13 August 2016].

Movable Type Scripts, n.d. http://www.movable-type.co.uk/scripts/latlong.html [13 August 2016].

Royal Netherlands Meteorological Institute: Climatological Database for the World's Oceans. n.d. https://www.knmi.nl/kennis-en-datacentrum/achtergrond/cliwoc. [14 August 2016].

South African Weather Service Synoptic Charts. 2016. http://www.weathersa.co.za/observations/synoptic-charts [04 August 2016].

Topographic map: Cape Town. 2016. http://en-gb.topographic-map.com/places/Cape-Town-1310815/ [13 August 2016].

United Nations Framework Convention on Climate Change. 2014. https://unfccc.int/2860.php [28 February 2017].

United Nations Sustainable Development Goals. 2015. http://www.un.org/sustainabledevelopment/sustainable-development-goals/ [28 February 2017].

Volvo Ocean Race: photo galleries. n.d. http://www.volvooceanrace.com/en/photo-view/30812.html [13 August 2016].

Books

Brown, A. C. (1977): *A History of Scientific Endeavour in South Africa.* Cape Town: The Royal Society of South Africa.

Forbes, V. S. (1977): Some Scientific Matters in Early Writings on the Cape. In *A History of Scientific Endeavour in South Africa.* A. C. Brown, Ed. Cape Town: The Royal Society of South Africa. 33–59.

Glass, I. S. (2015): *The Royal Observatory at the Cape of Good Hope: History and Heritage.* Mons Mensa: Cape Town.

Jackson, S. P. (1977): Meteorology and Climatology, In *A History of Scientific Endeavour in South Africa.* A. C. Brown, Ed. Cape Town: The Royal Society of South Africa. 388-408.

Lamb, H. H. (1982*): Climate, History and the Modern World*. Metheun: London.

Niang, I., O.C. Ruppel, M.A. Abdrabo, A. Essel, C. Lennard, J. Padgham, and P. Urquhart, (2014): *Africa*. In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Barros, V.R., C.B. Field, D.J. Dokken, M.D. Mastrandrea, K.J. Mach, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1199-1265. Available: https://ipcc.ch/pdf/assessment-report/ar5/wg2/WGIIAR5-Chap22_FINAL.pdf [28 February 2017].

Reason, C. and Hermes, J. (2011): Climate Change and Variability in Southern Africa and Regional Ocean Influences, In *Observations on Environmental Change in South Africa.* L. Zietsman, Ed. SUN MeDIA, Stellenbosch, 270 – 273.

Sabine, E. (1851). "Observation made at the Magnetic and Meteorological Observatory of the Cape of Good Hope", Vol I., Longman, Brown, Green and Longmans, London.

Stoy, R. H. (1977): Astronomy in South Africa, In *A History of Scientific Endeavour in South Africa.* A. C. Brown, Ed. Cape Town: The Royal Society of South Africa. 409-426.

Tadross, M. and Johnston, P. (2012): Climate Change Projections for Cape Town: Adding Value Through Downscaling, *Sub-Saharan African Cities: A Five-City Network to Pioneer Climate Adaptation through Participatory Research and Local Action*, Local Governments for Sustainability (ICLEI), 1 – 26.

Theron, A. (2011): Climate Change, Sea Level Rise and the Southern African Coastal Zone, In *Observations on Environmental Change in South Africa.* L. Zietsman, Ed. SUN MeDIA, Stellenbosch, 212 – 217.

Thom, H. B. and Balkema, A. A. (1952): *Journal of Jan Van Riebeeck, Vol I, 1651 – 1655.* Cape Town: Amsterdam.

Turner, M. (1988): *Shipwrecks and Salvage: In South Africa - 1505 to the Present*. C. Struik: Cape Town.

Tyson, P. D. and Preston-Whyte, R. A. (2004): The Weather and Climate of Southern Africa, Oxford University Press, South Africa.

Tyson, P. D. and Preston-Whyte, R. A. (2004): Chapter 12: Atmospheric Circulation and Weather over Southern Africa. *In The Weather and Climate of Southern Africa*, Tyson, P. D. and Preston-Whyte, R. A., Eds. Oxford University Press, South Africa. 176–217.

Tyson, P. D. and Preston-Whyte, R. A. (2004): Chapter 13: Ocean-Atmosphere Interactions. *In The Weather and Climate of Southern Africa*, Tyson, P. D. and Preston-Whyte, R. A., Eds. Oxford University Press, South Africa. 218-244.

Warner, B. (2012): *Royal Observatory, Cape of Good Hope 1820–1831: The Founding of a Colonial Observatory Incorporating a biography of Fearon Fallows*. Springer Science & Business Media.

Journals

Bonnardot, V., Planchon, O. and Cautenet, S. (2005): Sea Breeze Development under an Offshore Synoptic Wind in the South-Western Cape and implications for the Stellenbosch Wine-Producing Area. *Theoretical and Applied Climatology*. 81: 203-218.

Brazdil, R., Pfister, C., Wanner, H., von Storch, H. and Luterbacher, J. (2005): Historical Climatology in Europe – The State of the Art. *Climate Change.* 70: 363-430.

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B. and Jones, P. D. (2006): Uncertainty Estimates in Regional and Global Observed Temperature Changes: A New Dataset from 1850. *Journal of Geophysical Research*. 111: D12106(1-21).

Brohan, P., Allen, R., Freeman, J. E., Waple, A. M., Wheeler, D., Wilkinson, C. and Woodruff, S. (2009): Marine Observations of Old Weather. *Bulletin of the American Meteorological Society*. 90: 219-230.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, Jr., B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X.

L., Woodruff, S. D. and Worley, S. J. (2011): Review Article: The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society.* 137: 1 – 28.

Garcia-Herrera. R., Wheeler, D., R., Können, Koek, F., Jones, P and Prieto, M. R. (n.d.): CLIWOC Final Report. Available: http://pendientedemigracion.ucm.es/info/cliwoc/ [28 February 2017]

Garcia-Herrera, R., Können, G. P., Wheeler, D. A., Prieto, M. R., Jones, P. D. and Koek, F. B. (2005): CLIWOC: A Climatological Database for the World's Oceans 1750-1854. *Climatic Change.* 73: 1- 12.

Hannaford, M. J., Jones, J. M. and Bigg, G. R. (2015): Early-Nineteenth-Century Southern African Precipitation Reconstructions from Ships' Logbooks. *The Holocene*. 25(2): 379 – 390.

Jackson, S. P. (1951): Climates of Southern Africa. *South African Geographical Journal*. 33(1): 17-37.

Jimenez, P. A., lez Rouco, J. F. G., Vez, J. P. M., a Bustamante, E. G. and Navarro, J. (2009): Climatology of wind patterns in the north east of the Iberian Peninsula. International. *Journal of Climatology*. 29: 501–525.

Jones, P. D. and Salmon,M. (2005): Preliminary Reconstructions of North Atlantic Oscillation and the Southern Oscillation Index from Measures of Wind Strength and Direction taken during the CLIWOC Period. *Climate Change.* 73: 131-154.

Jonsson, P. and Holmquist, P. (1995): Wind Direction in Southern Sweden 1740 – 1992: Variation and Correlation with Temperature and Zonality. *Theory of Applied Climatology.* 51: 183 -198.

Kelso, C. and Vogel, C. (2007): The Climate of Namaqualand in the Nineteenth Century. *Climate Change*. 83: 357-380.

Klopper, E., Vogel, C. H. and Landman, W. H. (2006): Seasonal Climate Forecasts – Potential Agricultural-Risk Management Tools? *Climate Change.* 76: 73-90.

Lennard, C. (2014): Simulating an Extreme Wind Event in a Topographically Complex Region. *Boundary-Layer Meteorology*. 153: 237-250.

Lennard, C. and Hegerl, G. (2015): Relating Changes in Synoptic Circulation to the Surface Rainfall Response using Self-Organising Maps. *Climate Dynamics.* 44: 861-879.

Luterbacher, J., Xoplaki, E., Dietrich, D., Rickli, R., Jacobeit, J., Beck, C., Gyalistras, D., Schmutz, C. and Wanner, H. (2002): Reconstruction of Sea Level Pressure Fields Over the Eastern North Atlantic and Europe back to 1500. *Climate Dynamics.* 18: 545-561.

MacKellar, N., New, M. and Jack, C. (2014): Observed and Modelled Trends in Rainfall and Temperature for South Africa: 1960-2010. *South African Journal of Science*. 110(7/8): 1-13.

Matulla, C., Schöner, W., Alexandersson, H., von Strorch, H. and Wand, X. L. (2008): European Storminess: late Nineteenth Century to Present. *Climate Dynamics*. 31: 125-130.

Mitchell, T. D. and Jones, P. D. (2005): An Improved Method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology*. 25: 693-712.

Nash, D. J. and Endfield, G. H. (2002): A 19th Century Climate Chronology for the Kalahari Region of Central Southern Africa Derived from Missionary Correspondence. *International Journal of Climatology.* 22: 821-841.

Neukom, R., Nash, D. J., Endfield, G. H., Grab, S. W., Grove, C. A., Kelso, C., Vogel, C. H. and Zinke, J. (2014): Multi- proxy summer and winter precipitation reconstructions for Southern Africa over the last 200 years. *Climate Dynamics*. 42: 2713– 2726.

Nicholson, S. E. (2001): Climatic and Environmental Change in Africa During the Last Two Centuries. *Climate Research*. 17: 123-144.

Peterson E. L. and Troen, I. (2012): Wind Conditions and Resource Assessment. *WIREs Energy Environment*. 1: 206-217.

Picas, J. (2015): *Volcanic Forcing: Possible Impacts on Wind Strength in the Western Cape, 1834-1990?* BSc Geography Honours Project, unpublished paper, University of Witwatersrand, Johannesburg.

Reason, C. J. C. and Jagadheesha, D. (2005): Relationships Between South Atlantic SST Variability and Atmospheric Circulation over the Southern African Region during Austral Winter. *American Meteorological Society.* 3339-3355.

Reason, C. J. C., Landman, W. and Tennant, W. (2006): Seasonal to Decadal Prediction of Southern African Climate and Its Links with Variability of the Atlantic Ocean. *American Meteorological Society*. 941-955.

Rouault, M., Pohl, B. and Penven, P. (2010): Coastal Oceanic Climate Change and Variability from 1982 to 2009 around South Africa. *African Journal of Marine Science*. 32(2): 237-246.

Rummukainnen, M. (2013): Climate Change: Changing Means and Changing Extremes. *Climate Change.* 121: 3-13.

Schumann, E. H., Perrins, L. -A. and Hunter, I. T. (1982): Upwelling along the South Coast of the Cape Province, South Africa. *South Africa Journal of Science*. 78: 238-242.

Taunton-Clark, J. and Shannon, L. V. (1988): Annual and Interannual Variability in the South-East Atlantic during the 20th Century. *South African Journal of Marine Science*. 6: 97-106.

Thomas, D., Twyman, C., Osbhar, H. and Hewitson, B. (2007): Adaptation to Climate Change and Variability: Farmer Responses to Intra-Seasonal Precipitation Trends in South Africa. *Climate Change,* 83: 301 – 322.

Trenberth, K. E., Karl, T. R. and Spence, T. W. (2002): The Need for a Systems Approach to Climate Observations. *Bulletin of the American Meteorological Society*. 83: 1593-1602.

Vogel, C. H. (1989): A documentary derived climatic chronology for South Africa, 1820 – 1900. *Climate Change*. 14: 291–307.

Warner, B. (1978): Early Years of the Magnetic Observatory. *South African Journal of Science*. 74: 82–83.

Wheeler, D. (2001): The Weather of the European Atlantic Seaboard During October 1805: An Exercise in Historical Climatology. *Climatic Change.* 43: 361-385.

Wheeler, D. (2005): An Examination of the Accuracy and Consistency of Ships' Logbook Weather Observations and Records. *Climatic Change.* 73: 97-116.

Wheeler, D. and Suarez-Dominguez, J. (2006): Climatic reconstructions for the northeast Atlantic region A.D. 1685-1700: A new source of evidence from naval logbooks. *The Holocene.* 16: 39–49.

Zhang, J. and Crowley, T. J. (1989): Historical Climate Records in China and Reconstruction of Past Climates. *American Meteorological Society*. 2: 833-849.