An exploration of South Africa's wind climate using station records and reanalysis data

Brendan Argent

Thesis submitted to the University of Cape Town for the degree of Doctor of Philosophy

Supervisor: Bruce Hewiston

Co-supervisors: Chris Lennard and Andrea Hahmann

Department of Environmental and Geographical Sciences

May 2015

Contents

Abstract

Sparse information about the wind climate of South Africa behooves an exploration of the drivers of surface wind speed, especially in the context of wind resource assessment. This work quantifies the coupling between the synoptic circulation states and station-scale flows to develop a process-based regionalisation of wind regimes over the country.

A thorough inspection of available South African Weather Service (SAWS) wind records is conducted and a quality control procedure is applied. The procedure reveals a large proportion of the data are missing and existing data contain numerous errors such that only 107 of the original 960 stations passed the quality control criteria. However, data from these 107 stations only overlap temporally 2% of the time, which makes the data inappropriate for a regionalisation procedure. To ameliorate this, a method for incorporating bias-corrected time series data from a reanalysis data set is developed.

Data from the 0.3° resolution hourly Climate Forecast System Reanalysis (CFSR) between 1989-2010 is selected to improve the temporal coverage of the station data. The raw CFSR data overestimates wind speeds and underestimates the temporal variability and long-term trends. A bias correction method based on the wind speed and direction, time of day and month of the year is developed which successfully removes the mean error on wind speed and direction and improves the correlation with station records. This is achieved without disrupting spatial correlation patterns. Corrected and extended wind time series from each station site are used for the regionalisation.

The regionalisation uses a self-organising map (SOM) to define the archetypal synoptic circulation patterns in the reanalysis data set and the influence of these on the local wind climate is quantified. 12 representative atmospheric states are defined by the SOM that are consistent with the existing literature and capture the major synoptic circulation states. A hierarchical clustering is then used to define wind climate regions based on the coupling

between these circulation states and the extended station data. Six relatively cohesive spatial wind-climate groupings are identified that are physically consistent with the driving synoptic environment and are characteristic in terms of terrain and response to synoptic drivers. This process-based regionalisation facilitates a future assessment of potential changes in the wind climate of South Africa as a result of a warming world.

Acknowledgements

This work has been made possible by a number of people and organisations and I'd like to acknowledge the wonderful support I have received from a few in particular.

Firstly, to the groups that have supported this work financially: To the Harry Crossely Foundation, Woolworths, the Applied Centre for Climate and Earth System Science , the National Research Foundation, the Climate Systems Analysis Group (CSAG) and the South African Wind Atlas project, my sincere thanks. Without these contributions I would not have been to dedicate my time to this project. My thanks especially to the researchers at CSAG with whom I have shared office space over the last few years. Izidine, Hussen, Steve, Mokoena and Michael, you have enriched the many hours I have spent in the office and it has been a pleasure working alongside you.

The data used in this project was provided free of charge by two groups: The South African Weather service and the National Center for Atmospheric Research. An enormous amount of work goes into providing these data sets and I am indebted to the researchers that spend thankless hours putting these data sets together and making them available for work such as this.

In the process of writing this thesis I have read papers by numerous authors, across a range of fields, and have been acutely aware that the work that I do builds on the work done before me. I have been inspired by the passion that shines past the formal language and I am grateful to be a apart of this community; the proverbial shoulders on which I stand.

Special thanks must go to my supervisors at CSAG: Chris Lennard and Bruce Hewitson (as well as the informal advisers, Babatunde Abiodun and Chris Jack) and co-supervisor Andrea Hahmann from the Technical University of Denmark. Their input, guidance and encouragement has greatly enhanced the quality of this thesis. I have learnt an enormous

amount from their expertise and greatly appreciate their patience and support.

To Ros Skelton, Andrew and Muriel Argent who generously read through drafts of my thesis and who's suggestions have improved the clarity and cohesiveness of my writing, my thanks!

My family: Lucy, Jon, Sally and Andrew, my deepest thanks for the bottomless well of support that you have provided over the last few years. You are each inspirational to me and I am incredibly fortunate to have such a solid base.

And finally to my wife, Muriel. Your encouragement in the lowest points, and endless faith in me has made all the difference. Thank you for your love and patience and for walking this path alongside me.

1. Introduction

"No one can tell me,
Nobody knows,
Where the wind comes from,
Where the wind goes."
-AA Milne

The wind has fascinated people for millenia with its seemingly random behaviour and fickle nature. In recent decades, with the increasing use of wind for electricity generation, this fascination has become an economic necessity. As a result, the above excerpt from an A.A. Milne poem, *Wind on the Hill*, is probably not as true as it once was. However, in the context of a growing demand for electricity, in tension with an increasing awareness of the environmental impacts of its generation, questions about the nature of wind, and its potential as an energy source, take on growing importance.

1.1. Energy and Climate Change

When considering large-scale energy production systems, it is no longer possible to avoid the topic of climate change. The International Panel on Climate Change's (IPCC) Fourth Assessment Report (Solomon et al., 2007) concluded that 'most of the observed increase in global average temperature since the mid-20th century is very likely due to the increase in anthropogenic greenhouse gas concentrations', the majority of which are attributed to the consumption of fossil fuels (Edenhofer et al., 2011). Within this context, South Africa is ranked 12th-highest CO_2-equivalent emitting country per capita (Bhattacharyya and Ghoshal, 2010) in the world, largely due to the fact that over 90% of its electricity is

generated from coal power stations (Karekezi, 2002). It is also a country that is particularly vulnerable to the effects of climate change, both socially, for rural farmers (Thomas et al., 2006) and urban communities (Muller, 2007), and environmentally, with recent studies highlighting the potential effects on fauna (Erasmus et al., 2002) and flora (Midgley et al., 2002). Renewable energy technologies, which have a lower carbon footprint (Edenhofer et al., 2011), are therefore a primary means of climate change mitigation, and it is at least partly due to the increased governmental awareness of the potential impacts of climate change that there has been a rapid increase in renewable energy deployment in recent years (Edenhofer et al., 2011).

The 2003 White Paper on Renewable Energy (DoME, 2003), published by the Department of Minerals and Energy, set a target of 10 000GWh installed renewable energy capacity. This is in line with South Africa's carbon footprint reduction goals for 2030 (Rennkamp and Boyd, 2013) and represents roughly 4% of total electricity demand, however, according to recent estimates (Hagemann, 2008), this is just a small proportion of power that could be generated in South Africa by wind energy alone. By comparison, Denmark (currently the world leaders in terms of reliance on wind power) aims to each 50% wind powered electricity by 2020, and to be totally free of fossil fuels by 2050 (Energistyrelsen).

The total available wind power potential in South Africa is a matter of current research. Equation 1 shows that the upper limit in the power contained in the wind depends on the third power of the wind speed.

$$P = 1/2\rho V^3 \tag{1}$$

Here P is the power density of the wind (W/m^2), ρ is the air density (kg/m^3) and V is the wind speed (m/s). Estimates of the available wind power are therefore highly depen-

dent on measurements (or estimates) of the mean wind speeds at potential sites around the country.

The question, 'Where does the wind come from?', can be answered at a number of different levels. On a global scale, the differential heating of the earth's surface produces localised areas of high and low pressure. These pressure differences, combined with the effect of a revolving earth, result in large-scale wind systems. The disruption of these flows by the continents results in large circulation systems like the semi-permanent anticyclones. Smaller systems like fronts and troughs form at the synoptic scale. Locally these systems are further influenced by topography, surface roughness and smaller scale thermal gradients. Finally, the wind at a particular site is the combination of the local-scale circulation features and the micro-scale features like nearby trees or roughness changes.

The fine scale of local and micro-scale variations in wind speed and turbulence mean that there are significant gains to be had in the micro-siting and layout of wind farms. It also means that accurate estimates of the available power require high resolution estimates of wind speeds. Small errors in these estimates may result in large errors in predicted power production because of the cubic dependence on wind speed. This sensitivity, combined with limited availability of measurements, makes the task of locating the best wind sites and quantifying the available power a difficult task, especially in a country as large as South Africa.

1.2. Wind power potential in South Africa

Several studies have attempted to map and quantify the South African wind resources, beginning with the country's first Wind Atlas, which was published by Diab (1995) (Shown in Figure 1). The figure divides the country into regions of 3 types: 'Good', 'Moderate' and 'Low' depending on the mean wind speed from each weather station. Based on an analysis of the available weather station records, this study suggested that wind could supply 7.9TWh/year (or 2.4% of South Africa's electricity needs). While ground-breaking at the time, the atlas was limited by the available data (79 station records, outdated Dines Anemometers mounted close to the ground, and manually recorded records), and the methodology used to create the main atlas in Figure 1 is not elaborated on in the publication. An update of this atlas was produced in 2001 by ESKOM, the Department of Minerals and Energy, the Council for Scientific and Industrial Research (CSIR) and the DANish International Development Agency (DANIDA). This was unfortunately not made public, but some of the results can be accessed through the South African Renewable Energy Resource Data Base [i] and the main Figure from this Atlas is shown in Figure 2). The figure shows the mean speeds from each station as a circle centred on the station position. According to a third-party report (Loy, 2004), this atlas suggested a total wind energy potential of 26TWh/year. A more recent study by DoME and an update based on the same data by Banks and Schaffler (Banks and Schaffler, 2008), estimate the country's wind resource at 60TWh/year and 106TWh/year respectively. This order of magnitude difference in the range of predictions is indicative of the sensitivity of the estimates to the assumptions made.

In *The Mesoscale Wind Atlas of South Africa*, Hagemann (2008), instead of using measurements to estimate the available wind energy potential, used a mesoscale model [ii] to

[i] http://www.sabregen.co.za/

[ii] The Karlsruhe Atmospheric Mesoscale Model (KAMM)

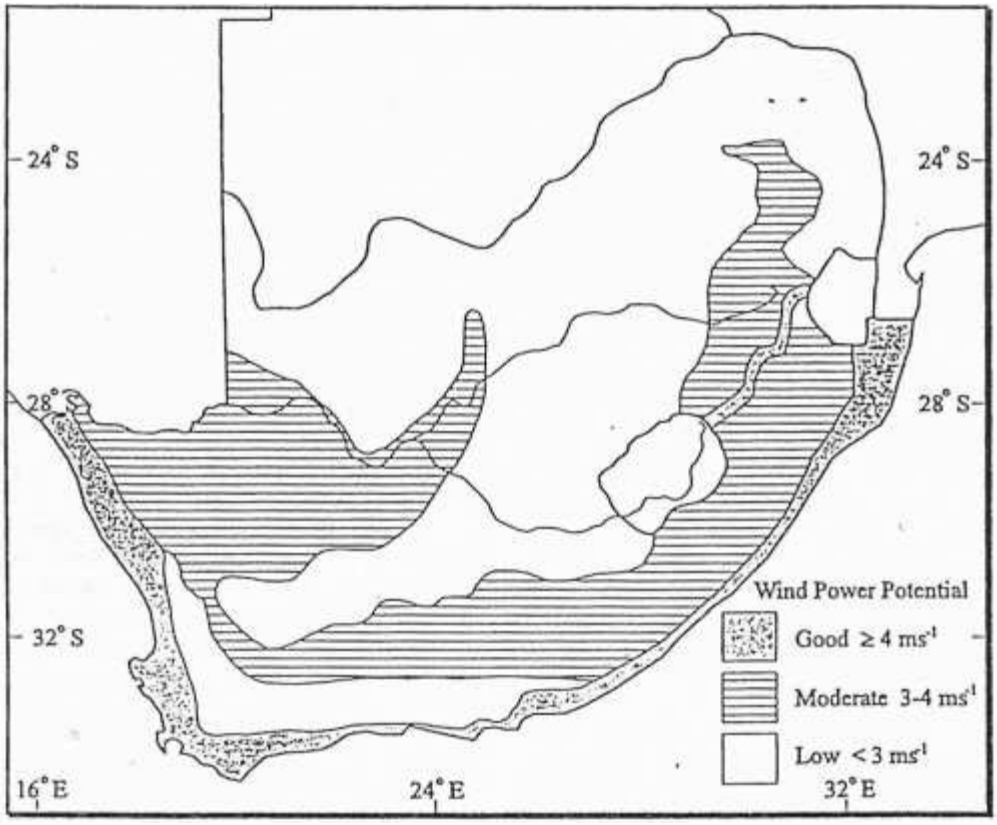

Fig. 1.— South Africa's First Wind Atlas showing general wind speed zones (Diab, 1995).

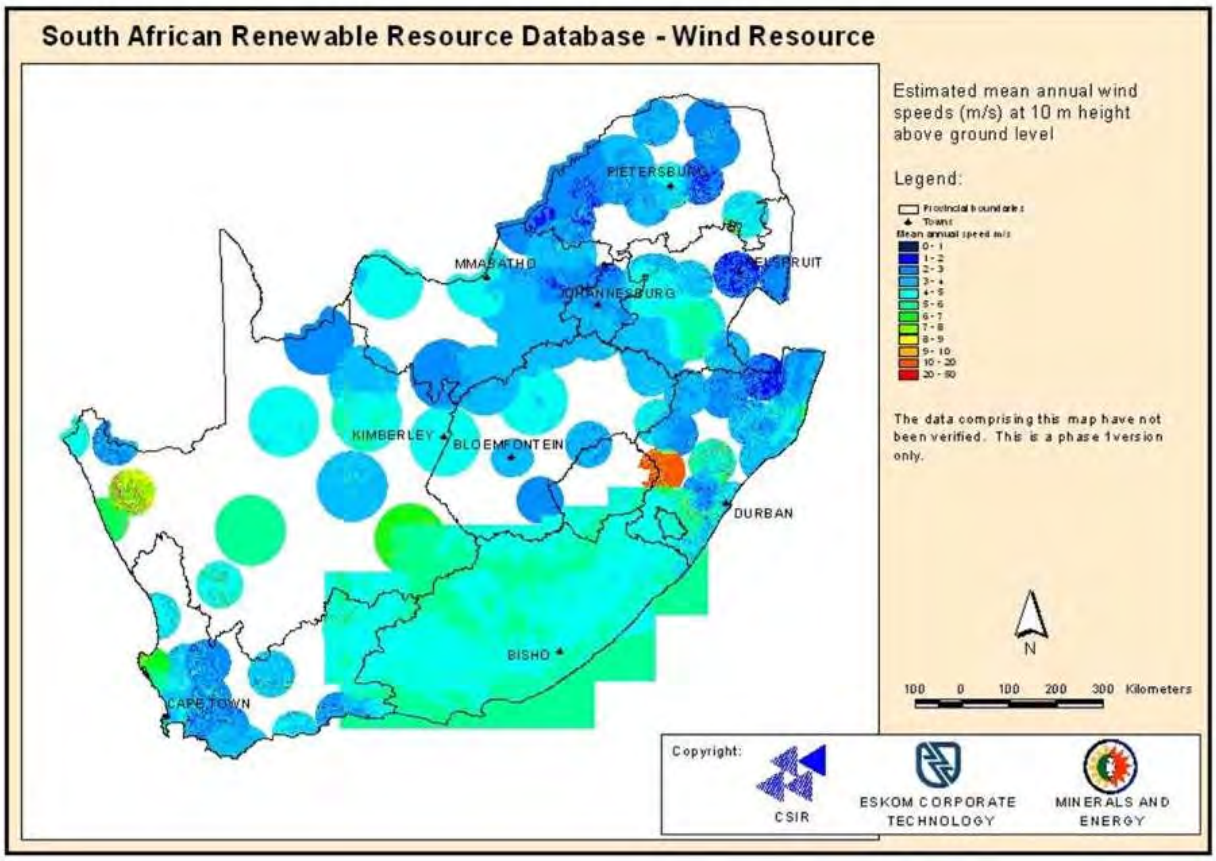

Fig. 2.— ESKOM's Revised South African Wind Atlas, showing mean wind speeds in a circular area centred on each station site

simulate the wind climate and based his estimations on the resulting data. This 'top-down' approach has been used in many other regions [iii] but was a first for this country. In order to limit the computer resources needed to run a full, multi-decadal climate simulation, Hagamann used a novel approach of selecting a 12-month period that 'best represented' the South African climatology, and ran the integration for this period alone. This was a step forward in the efforts to quantify the available wind energy in South Africa. The concluding estimate was that, 'realistically', 80.54TWh of energy could be produced from wind per year [iv].

While this was a good step towards a more robust resource assessment, there are several aspects of this work that could be improved upon. Being limited by available computational capacity, the simulation was only run for a single (albeit, relatively representative) year. This means that the full extent of the inter-annual variability of the wind speed was not included. Additionally, it was only validated against 17, 10m weather station records, none of which were in the regions later identified as potentially good wind energy sites.

The most comprehensive wind atlas to date is the Wind Atlas for South Africa (WASA), of which the first phase was released in 2013. The WASA project, (with which this work is aligned) is a collaboration between the Danish Technical University (DTU), the South African Weather Service (SAWS), the Council for Scientific and Industrial Research (CSIR) and the University of Cape Town (UCT), and aims to facilitate the early phases of wind energy development in the country [v]. This atlas was produced according to the same methodology used for the European Wind Atlas (Troen and Petersen, 1989) and several others

---

[iii] A comprehensive list is available at http://www.windatlas.dk/World/Atlases.html

[iv] This is based on wind speeds at 80m at all sites that had a maximum distance of 4km from roads and 4km from power-lines as well as minimum capacity factors of 30%.

[v] More information about the WASA project can be found at *http* : *//www.wasaproject.info/*.

including the Egyptian (Mortensen et al.) and Russian (Starkof et al., 2000) atlases [vi]. In brief, the first phase makes use of reanalysis data [vii], downscaled by the Karlsruhe Atmospheric Mesoscale Model (KAMM) (Badger et al., 2014) to determine the typical meso-scale conditions. A resource map is generated from each of these with the WAsP micro-scale model (Mortensen et al., 1993), and an average of the resource maps, weighted by the observed frequency of each circulation type, gives the final estimate shown in Figure 3.

The second phase of the project, released in March 2014, used the Weather Research and Forecasting (WRF) model instead of KAMM for the downscaling. The WRF model is computationally more demanding than KAMM, as it simulates the entire period 1980 - 2010, but has advantage that it allows for a more detailed exploration of the circulation dynamics of the region.

The results of the first stage of the WASA atlas were validated against the records from ten, 60m masts erected by the CSIR for this purpose (Hahmann et al., 2015). The time series from the nearest grid-cell in the model to each of the masts was compared with the observed time series at the actual masts. It should be noted that this comparison is not strictly fair to the model. [viii] The mean absolute error in wind speed was found to be 4.2%

[vi] The full details of the methodology are available at http://www.wasaproject.info/.

[vii] Reanalysis: Data from a global climate model, that incorporates observations.

[viii] Even a 'perfect' model, one that exactly reproduced the physics of reality, would not produce the same values as observed at the mast because they are fundamentally different statistics. The mast measures the speed of the air passing a single point in space, while the model determines the mean speed at that height in the grid-cell. Even in relatively flat, uniform terrain, there will be places with relatively higher or lower wind speeds. So at best we would expect the model value to be highly correlated to the measured speeds, and perhaps to have a small, relatively consistent bias for a given wind direction. This topic will

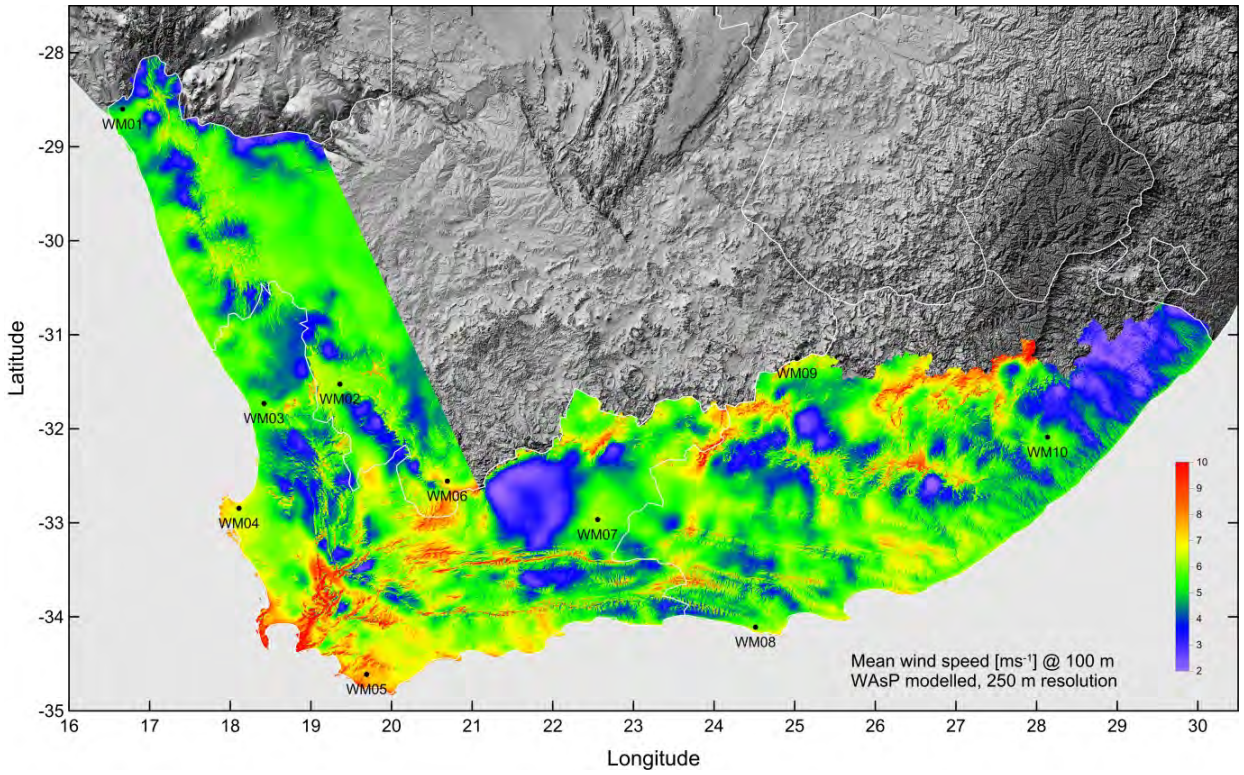

Fig. 3.— Microscale map from the WASA project showing the mean wind speeds, downscaled to 250 meters from the generalised WRF output for the Western Cape and parts of the Northern Cape and Eastern Cape.

across the 10 masts for the 3 year period (Hahmann et al., 2015). While this is a favourable result in the context of wind mapping there remains space for improvement.

Firstly, the validation period could be longer. In Europe, surface wind speeds have decreased over the last three decades (Vautard et al., 2010), and previous estimations of power, based on short validation periods, over predicted the available wind resource. If the measurements are taken in an anomalously windy or calm year, this can skew the estimates. Second, there are only ten validation points for the country and there are likely wind regions that are not represented. Unfortunately these ten stations are the only measurement masts of this kind for which data records are publicly available in South Africa. Due to the competition between wind farm developers, wind records (and research conducted with them) are considered a competitive advantage and are therefore not publicly available. As a result, high quality wind observations are scarce and most of the wind related research that has been performed with meso- and micro-scale models could be improved with a comprehensive observational data set for validation. Ideally this would make use of the changes in the wind measurement instrumentation and the automation of data collection and logging, as well an increase in the number of reporting weather stations.

There are therefore, several factors that motivate for an observation based exploration of the wind climate of South Africa.

1) There have been a limited number of studies of wind in South Africa;

2) there is a wide range of estimates of the available wind resources;

3) there is a lack of reliable observations with which to compare the model estimates;

4) there is a need for a process-based understanding of the factors that affect wind speeds around the country.

---

be addressed in Chapter 3.

It makes sense, therefore, that the available data should be brought together in a methodologically consistent manner, and used to advance the understanding of the processes that drive the various wind climates around the country.

There are two gaps in the literature that this thesis aims to address: a) the lack of a reliable observation-based wind climatology for South Africa, and b) the link between the large-scale circulation and the surface wind observations.

1.3. Motivation

There is a real need for a solid observational base for wind energy research in South Africa. The country is moving towards more renewable power technologies, and wind energy is set to become a significant contributor to the energy mix. Assessment of the available wind resources is primarily conducted with meso- and micro-scale models that can provide high-resolution maps of the wind climate in a region. However, without a good coverage of measurement masts recording the wind characteristics observed in reality, these maps cannot be adequately validated. There are also gaps in the literature regarding the link between the large-scale circulation and regionally observed surface winds, as well as concerning long-term variation in surface wind speeds. A fairly large body of work regarding these issues has been accumulated for areas in Europe and the United States of America [ix], but little has been done in the South African context.

Wind energy is an important source of energy in South Africa, especially in light of a) the recent energy shortages and b) the country's current contribution and vulnerability to climate change. In order to make the best use of this resource, a solid understanding of the

[ix] The literature relating to these areas is addressed in detail in later chapters.

country's wind climates is required at a range of spatial and temporal scales. While the tools to model these variables are available, the data required to validate them is limited and the quality of the data that is available is questionable.

As an additional motivation there are a range of industries in South Africa that are dependent on wind conditions, and enhanced information regarding both regional and local wind conditions over time will be of benefit to these industries. Examples include: fisheries(Jury, 1988; Cury and Roy, 1989), fire (Richardson et al., 1994) and disaster management services and regional planners, pollution modeling (Engelbrecht et al., 2001; Garstang et al., 1996) as well as airports and recreational activities.

It is hoped that this work will convert the millions of individual data points, recorded over decades or generated by models, into information that is useful for these industries.

1.4. Aims

1.4.1. Aim 1

The primary aim of this thesis is to explore the relationship between the observed wind climate of South Africa and its large-scale synoptic forcing. As the literature specific to wind climate in South Africa is relatively sparse, this research seeks to mitigate this by using the larger, international body of work regarding wind climate and energy as a guide to the methods and general approach.

1.4.2. Aim 2

The secondary aim is a response to the lack of publicly available wind data of sufficient quality for a rigorous assessment of wind climate in South Africa. The problem has two

aspects; firstly, there are not enough reporting stations for such a large country, particularly given the complexity of its terrain and the spatial variability of its climate. Secondly, the largest data set of wind records that is available (from the South African Weather Service) is poorly-organised, non-uniformly formatted, lacking in meta-data and in need of rigorous and systematic quality control. The former is being addressed by SAWS and will not be considered in this thesis. The latter forms the motivation for the secondary aim of this thesis, which is to develop a reliable data set of surface wind observations and explore the relevant differences between these data and the available model data.

1.5. Objectives

There are a number of specific objectives to be met in order to achieve the aims of this research.

Objective 1:

1.1) To review the available, observed wind data and compile a single reliable data set for the purpose of validating the results of meso-scale wind modeling over South Africa. This data set would ideally maximise spatial and temporal resolution and continuity, and be based on observations;

1.2) The compilation, screening and quality control of all available wind observations for South Africa;

1.3) The development of a method of combining observational wind data (which is discontinuous in both space and time) with modeled surface wind data (with known and unknown biases);

1.4) The optimisation and automation of this method in preparation for wind climate classification.

Objective 2:

To develop a classification scheme to define regionally consistent, coherent wind climate zones in South Africa and subsequently describe the identified zones.

Objective 3:

To initiate research into the links between the synoptic-scale circulation and the regional surface wind climates in South Africa.

1.6. Thesis outline

The structure of this thesis mirrors the objectives described in the previous section and the results are separated into three Chapters, each dealing with a self-contained, but sequential part of this project. They have been written to be relatively self standing, each with separate introduction, literature review, method and results sections.

The collection, selection, cleaning and reformatting of the observational data from SAWS is described in Chapter 2. An important step in this process is the selection of the inclusion criteria, and the development of methods to identify, and, where possible, resolve inconsistencies in the data. Visualisation techniques are also explored in order to make the large quantities of multi-dimensional data more accessible. A set of stations is selected from

around the country with sufficient quality and quantity of data available in the period 1994 to 2010.

Chapter 3 details the merging of SAWS station observations with a gridded reanalysis product to form a single, reference data set. Three novel bias-correction methods are tested and a combinatory method selected. The validation of the method is described, followed by the implementation at two example stations. The details of the automation of the method for application to the entire domain are also shown, the final result of which is a complete, bias-corrected, 30-year hourly time series for each station point.

In Chapter 4, a self-organising map is used to dis-aggregate the hourly synoptic circulation patterns into 12 representative archetypes. The bias-corrected time series' are split according to the circulation types associated with each hour and the correlations between the surface and 850hPa wind speeds are used to group the stations into 6 regions by means of a hierarchical clustering. The characteristics of each region are explored with respect to the seasonal and diurnal cycles, the shape of the wind speed-direction histograms and the relative correlation with the other regions. Boundaries between regions are subjectively determined by consideration of the terrain around each station and the obvious physical barriers. Finally, the regions are compared with existing climate classifications.

Chapter 5 presents the conclusions of the study, explores some of the caveats to the results and describes possibilities for future research.

2. South African wind records

2.1. Introduction

2.1.1. Introduction

This chapter sets out to compile a comprehensive data set of wind observations from around South Africa, in line with objective 1.2. In order to do this, a review of the available observational wind data in South Africa is given and a quality control procedure developed to ensure that the records are consistent and reliable. More than simply improving the quality of this data set, the purpose of this chapter is to prepare the data for further analysis in the following chapters.

2.1.2. Weather station records

The South African Weather Service (SAWS) is the largest meteorological institution in South Africa and maintains several thousand weather stations across the country. Of these, 960 have records of wind speed and direction. The records from these stations, some of which no longer in operation, are freely available for any non-commercial research and can be accessed through SAWS. The wide distribution and availability of this data makes it the obvious choice for the exploration of the surface wind climate of South Africa.

2.1.3. Data description

The data set provided by SAWS for this research consists of 1083 ASCII files, each containing the record for a single station. Besides the station name, station identity number and coordinates, these files consist of columns for each time-stamp, wind speed and wind

direction. Although some of the stations date back to the beginning of the last century, most of the older records have not yet been digitised and the earliest records in this data set begin in 1960. SAWS maintains several different types of weather stations, ranging from fully manual to automated. Since 1994 however, the stations have been gradually converted to automated weather stations, introducing a greater level of standardisation, and five-minutely wind readings, averaged to hourly values. The data set was initially provided with no meta-data, but over the course of the project, meta-data files for 70 of the stations were made available. From these it is clear that the automated stations include a 10m wind vane and an RM Young cup anemometer recording hourly wind speeds and direction.[1]

Before the station data could be used for any analysis, it required quality assurance processing. The following section describes the quality assurance procedures in the literature and discusses the types of errors that are commonly found in wind observation data. This is followed by a description of the methods used to identify errors in the SAWS data set, the selection of suitable stations as well as a discussion of the results.

2.2. Quality assurance literature

As temperature and precipitation has been extensively studied, most of the literature concerning quality assurance of weather station data is focused on these meteorological variables (Brunet et al., 2006; Wijngaard et al., 2003; Vincent and Zhang, 2002; Karl and Williams, 1987). Gandin (1988) has described the errors found in meteorological data as one of three types: random, systematic and rough errors. Random errors are those associated

[1]Limited meta-data provided by SAWS does not specify the make of the wind vane or the model of the anemometer beyond 'RM Young'. The boom mountings are also not described. Caution should be taken in using the data where high levels of accuracy are required.

with the measurement process itself, are intrinsic to the measurement instrumentation or due to the effects of other variables on the measurement process. As the mean of these errors is usually close to zero it is often considered to be 'noise' in the data, and in most cases these errors are unavoidable and impossible to remove. Systematic errors are those that persist in time and are not centred on zero. The means of these errors are usually referred to as biases. Rough errors are those that are introduced as a result of instrument failure or errors in the data transfer and processing. These are generally the least common but are potentially large in magnitude. It is the identification and removal or correction of the latter two that will be the focus of this section.

In order to identify these errors, several tests are commonly used, both in real-time quality control, and quality assurance of existing data sets (Jimenez et al., 2010b). These tests can be roughly categorised into three classes. The first examines each data point in isolation. Plausibility and contradiction tests are common examples of these. Plausibility tests screen for observations that are physically impossible, like negative wind speeds, or highly unlikely, such as wind speeds far beyond the normal range. These are usually identified by comparison with the mean and variance statistics. Contradiction tests identify conflicting information across multiple variables. Rainfall in conjunction with clear skies would be an example of this, as would dramatic changes in wind speed with no corresponding change in air pressure.

The second class of tests consists of spatial and temporal consistency checks. By comparing observations that are close in either space or time, differences that are outside normal ranges can be flagged as suspect. Several spatial interpolation techniques are commonly used in these tests and have been compared by Eischeid et al. (1995). The third class involve more sophisticated methods, often based on the diagnostic equations, to predict the expected value of each datum. The hydrostatic and geostrophic tests are examples described by Gandin

(1988) in some detail. Gandin, however, recommends the use of multiple simple tests as opposed to single, complex tests, as this approach tends to provide greater information for correcting observations flagged as suspect.

There have, however, been a number of studies focused on quality control of wind observations in particular. Using a set of 41 weather stations in the USA, DeGaetano (1996) developed a quality control method consisting of a set of plausibility and contradiction tests as well as temporal consistency tests, to identify periods of excessive variability, anomalously constant speeds and directions. These were referenced against temperature records to evaluate the possibility of instrumentation failure due to freezing, and wind speed observations from nearby stations. Graybeal (2006) explored the relationships between daily mean and maximum wind speeds in order to develop tests for quality control applications. More recently, Jimenez et al. (2010b) developed a series of plausibility and temporal consistency checks in addition to checks for manipulation of the data during storage, but found that stations too far apart did not allow for effective application of spatial consistency checks. These authors also mention the usefulness of meta-data that describe the process by which weather stations are installed and maintained, as well as how the data is collected and stored. Gandin (1988) in particular, stresses the importance of minimising the processing of observations prior to quality control as this can spread or conceal errors.

Recently, an atlas of extreme winds was published by Kruger et al. (2010), based on the SAWS Weather station wind observations, as a guide to an update of the South Africa building design codes. The selection of stations for inclusion in this analysis was guided by the exposure of the stations and a minimum of 10 years of records was required. Data was screened manually by inspection of the time series plots and years where anomalies were evident were excluded. In an update of this work in 2011, Kruger et al. (2011) applied an exposure correction to the data from stations deemed to be sheltered, or affected by nearby

obstacles or roughness changes.

The methods presented in these studies guided the selection and application of the quality assurance tests applied to the SAWS weather station records in the current research. However, a more thorough quality assurance was deemed necessary, as this study is concerned with the co-evolution of time series across the country, the diurnal and seasonal variability and preparing the data set for further research into the wind climate of South Africa.

2.3. Methodolgy

There are two characteristics of this data set that made it difficult to work with and which were taken into account in the selection of the stations and the data periods. An initial, cursory review of the station files revealed a lack of uniformity in the formatting. Different column dividers, switched columns, different levels of precision and temporal resolution and skipped dates are some of the inconsistencies that are numerous both within, and between, files. The second element is the lack of meta-data. Details of the instrumentation, station siting and relocation, maintenance, calibration and data-processing are not available for the majority of the stations to guide the selection.

In order to deal with these challenges, a four-stage quality assurance procedure was developed and implemented. The first stage read in the data, removed contradictory records and gross errors and assessed which stations contained sufficient data to justify further consideration. The second involved locating the station sites on Google Earth and removing stations that are poorly sited, an extension of the work done by Kruger et al. (2011). In the third stage, a series of quality control measures were applied to remove any anomalous periods in the remaining records, after which a final selection of stations was made. A description of each of these stages is given in the following sections.

2.3.1. Reading the records

In order to read the 1083 station files, a processing script was written in Python [i]. The following is a stepwise description of the checks conducted by this script to identify the data fields and remove the contradictory values.

For each file, the station identity number and coordinates were stored from the file headers. Any duplicates were flagged for a manual check. The columns were identified by the number of characters in each row and the range of values, with any unrecognised formats flagged for manual check. According to the criteria used by Jimenez et al. (2010b) and in line with the limits of the instrumentation, wind speeds below $1m/s$ or above $50m/s$ were removed as well as direction values greater than 360^o or less then 0^o. For any cases where the wind speed was $0m/s$, indicating calm, the direction was set to 0^o (360^o indicating northerly winds). Similarly, where the wind direction is 0^o, the contradictory wind speeds (those greater than $1m/s$) were set to $0m/s$.

The time series for each station was rewritten to a new file with standardised formatting and missing-value flags. During this process, several files were found to be duplicates, reducing the total number of unique stations to 960.

Before continuing with the quality assurance, a coarse filtering of the stations was applied. This was necessary as some of the stations files contained no data records and others had very short records, unsuitable for this work. Following the precedent set by Kruger et al. (2010), the records prior to 1994 were removed as these stations used Dines anemometers, whereas the automated weather stations all used the more modern RM Young cup anemometers. There is no meta-data available from SAWS specifying the type or model of Dines anemometer used and for many stations, prior to 1994, wind direction is only specified

[i]For more information see www.python.org

to the nearest 30^o and wind speeds are only given to the nearest m/s or in some cases, miles per hour. The data is also far more spare with many stations only recording data twice a day. Due to the difficulty in comparing these different types of measurements, and the lack of adequate meta data to adjust the older measurements to match the automated ones, this data was excluded from the study. Only stations with a minimum of three years of data with at least 90% available data between 1994 and 2010 were included. This reduced the number of stations to 243.

2.3.2. Station siting

Since the siting of the station can significantly impact the representivity of the data, stations that were sheltered by nearby objects, or poorly exposed, were removed from the selection. Many of the stations are in close proximity to buildings for convenience and the speeds and directions recorded at these stations are therefore not representative of the larger area. In these cases, the station records are not suitable for this study. An example of a poorly sited station is shown in Figure 4. This photograph comes from the meta-data report (SAWS, 2012b) for Joubertina, one of the few stations for which meta-data is available [ii].

In order to identify sheltered stations, in the absence of the appropriate meta-data, a Google-Earth based survey was conducted. Some work has already been done in this regard both by SAWS, who have begun compiling meta-data reports for each station, and by Kruger et al. (2010). However for the 160 stations not covered by these reports, Google Earth was used to locate the stations and classify their exposure.

The World Meteorological Organisation (WMO) provides a guideline of a minimum distance from sheltering obstacle, of ten times the height of the obstacle(WMO, 2010). This

[ii]At the time of writing these reports were only available for 70 of the SAWS stations.

Fig. 4.— Image of the poorly sited Joubertina weather station from the SAWS meta-data report

criteria was applied by estimating the height of each obstacle close to each of the stations. If the distance to the station was less than ten times this value, it was deemed to be significantly impacting the wind measured by the mast, and the station was excluded. Stations that were deemed too close to steep topographical features were similarly excluded. Due to the fact that coordinates of many of the stations are only specified to 2 decimal places, it was not possible to find some of the stations and these stations were also excluded.

2.3.3. Quality assurance

Having removed the gross errors and the sheltered stations from the data set, a series of temporal continuity tests were applied. These are designed to identify periods of abnormally high or low variance, outliers, and other anomalous data periods that may result from instrumentation errors or that were introduced in the transfer and storage of the data. Three tests were adapted from those described by Jimenez et al. (2010b) to identify periods of excessive or reduced variance. For the first of these, a set of figures was produced for each station consisting of a histogram of the wind speeds, the seasonal and diurnal cycles with mean values and 10th and 90th percentiles, a windrose, and a monthly mean time series. An example of one of these figures is shown in Figure 5 for the Ventersdorp station.

The red bars in the seasonal and diurnal panels show the 10th to 90th percentile range for these variables. The gaps in the time series indicate months in which less than 90% of the hours were recorded. The cut-in speed of the anemometers is $1m/s$, so any speeds below this threshold are registered as calm. This is the reason for the gap in the histogram for the 0.5-$1m/s$ bin. An anomalous period beginning in early 2006 is clearly evident in the time series. In other cases (not shown here), the errors were evident in the other plots, for example, as irregularities in the histogram. In order to speed up these manual checks, the figures were linked to the station coordinates in a Google Earth interface. This allowed

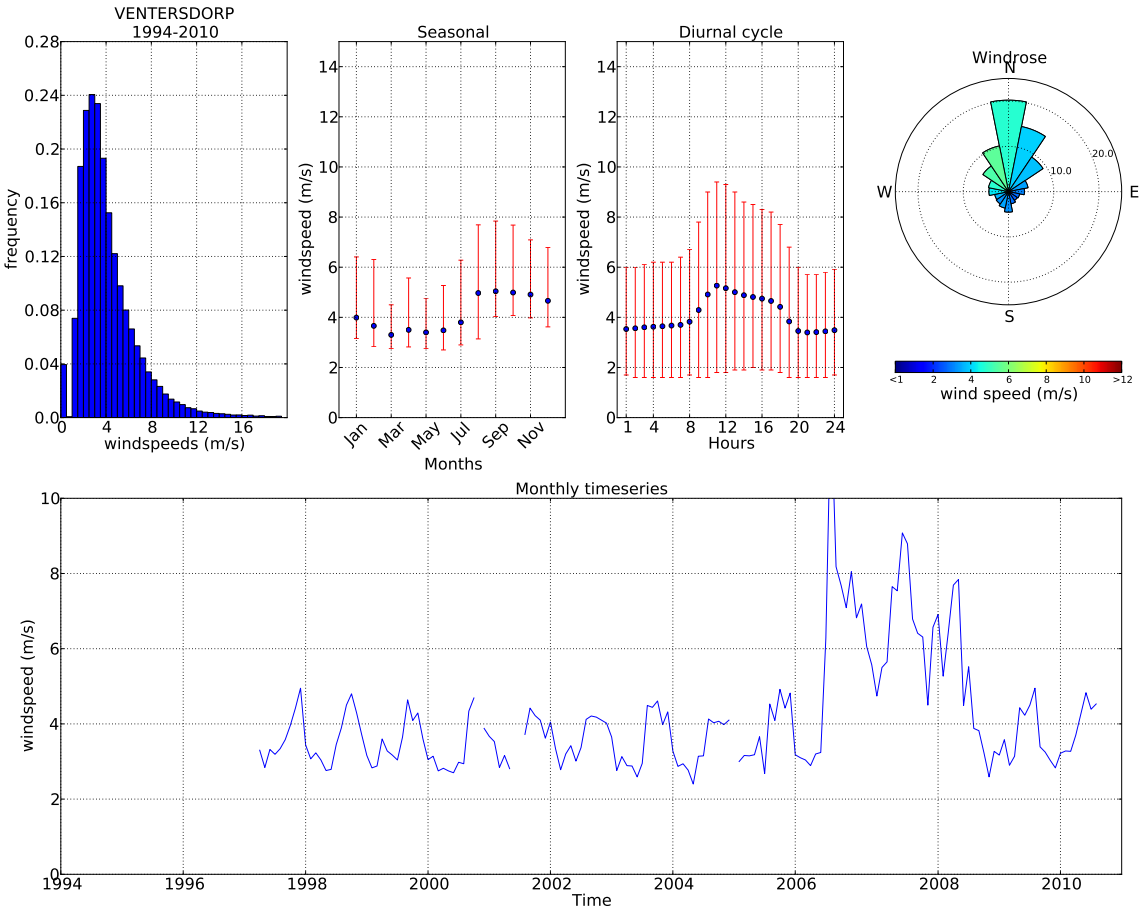

Fig. 5.— The upper panel shows the histogram, seasonal and diurnal cycles for the Ventersdorp station as well as the wind rose. The lower panel is a times series of the mean monthly wind speeds. Gaps in the time series indicate missing data and the empty bin in the histogram is a result of the anemometer cut-in speed of $1m/s$. An anomalous period beginning in early 2006 is clearly evident in the time series.

for easy comparison between nearby stations and facilitated physical interpretation of the results. A screen-shot of this interface can be seen in Figure 6.

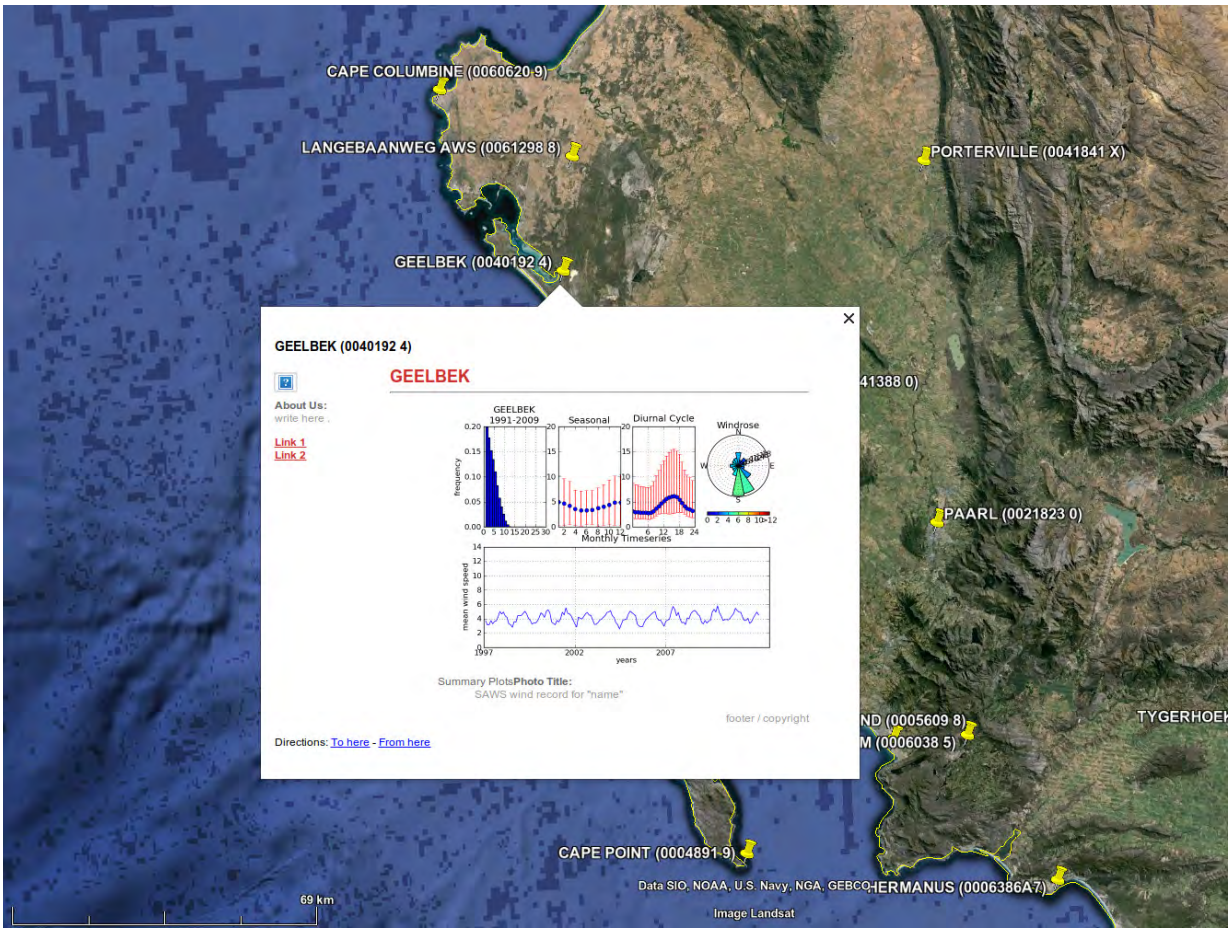

Fig. 6.— A screenshot of the Google earth station interface showing the summary figure for the Geelbek station

The second test identifies suspect steps in wind speed from one hour to the next. Different thresholds have been suggested for stations in other regions, ranging from $5m/s$ (Bailey and McDonald, 1997) to $10m/s$ (Jimenez et al., 2010b). However, due to the wide range of wind climates represented in this data set, a threshold was determined for each station based on the range from the 5th to the 95th percentile of wind speeds. These varied from $3.4m/s$ to $11.3m/s$. The final test identifies repeated values that can occur when either the wind

vane or anemometer becomes stuck in a position. These were divided into repeated calm and non-calm values as zeros are occasionally used to replace missing values. The threshold for erroneous calm periods was determined from inspection of the histogram of the lengths of calm periods across the station records shown in Figure 7. The distribution tapers to zero at

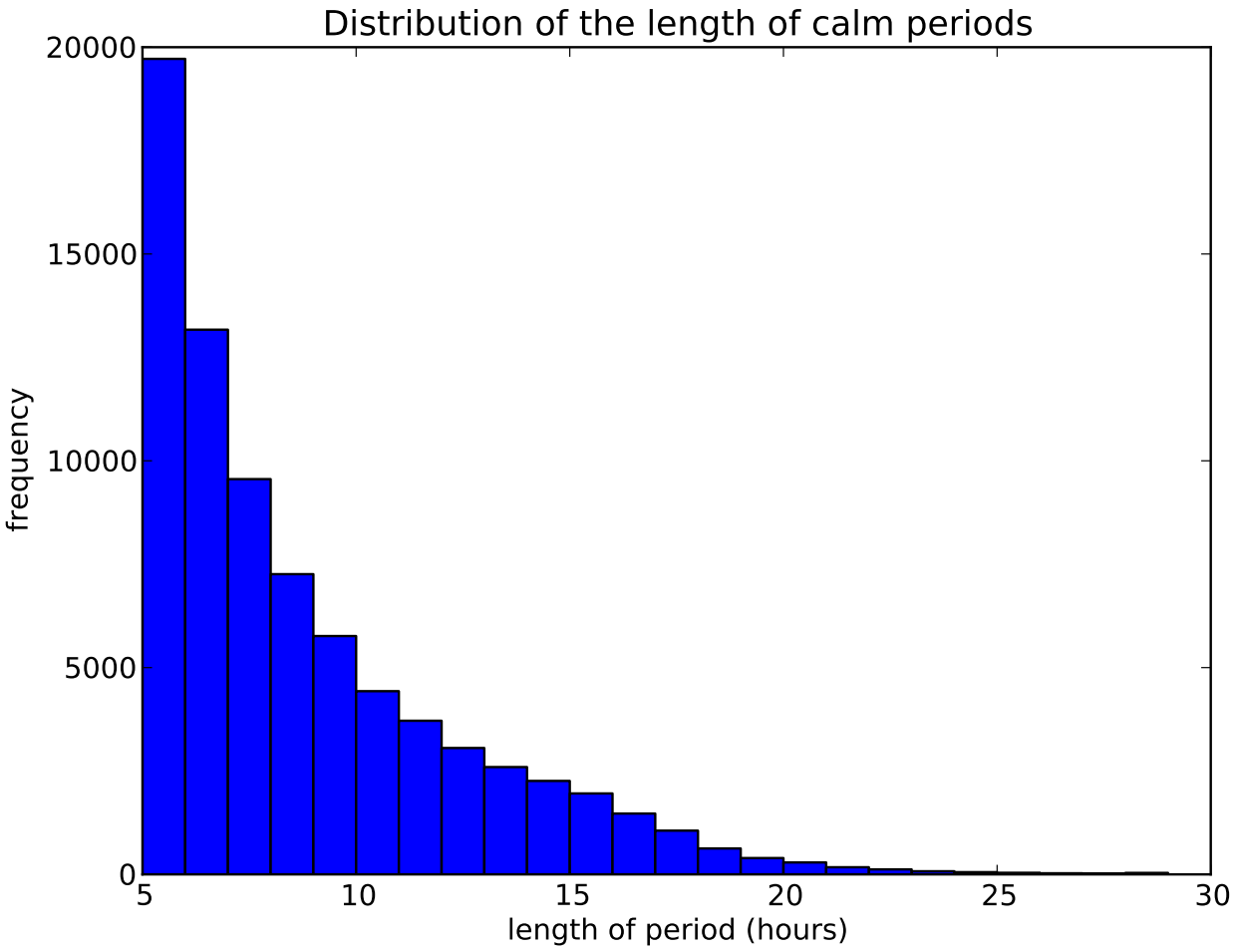

Fig. 7.— The distribution of lengths of calm periods greater than five hours in length in the station time series.

28 consecutive hours (marginally higher than the value found by Jimenez et al. (2010b)). Not shown in Figure 7 are the calm periods shorter than five hours as these were so common that

if displayed, the rest of the figure would be barely discernible. Similarly, the outlying calm periods, up to a maximum of 1479 hours, are not shown. Although infrequent, calm periods longer than 28 hours were deemed to be erroneous. Due to the relatively large distances between the stations it was decided that a spatial continuity test to confirm these calms as plausible would not be suitable. For the non-calm, repeated wind speeds, a threshold of ten repeated values was chosen in a similar manner. Repeated non-calm wind directions were not evident in the records for more than three consecutive values.

While more sophisticated tests such as the geostrophic test or the use of a numerical model to predict the hourly time series may locate further anomalies in this data set, it was decided that these relatively simple tests would provide sufficient assurance of the quality of this data for the current research.

2.4. Discussion of the results

The proportion of errors in this data set is an order of magnitude greater than those typically seen in the literature (Jimenez et al., 2010b), and the majority of the errors are a result of the lack of uniformity and standardisation in the collection and storage of the data. The number of erroneous values found by the different tests are summarised in Table 1. The removal of the false calm periods significantly affected the mean speeds at several stations, as did the removal of the anomalously high wind speeds identified manually from the summary plots. Several of these were found to correspond to a unit change from m/s to mph. Many of the errors removed and corrected in the quality control procedures are relatively small in magnitude and are unlikely to have a significant impact on the outcomes of this study. Similarly, there are likely to be some errors that were missed, however, some of the station records were affected by the corrections.

The following are brief, reference descriptions of the errors described above, and summarised in Table 1:

Read Errors - Data points with anomalous formatting, preventing automation and standardisation.

Gross Errors - Negative speeds or directions, speeds greater than $50m/s$ and directions greater than 360^o and contradictory values ($0m/s$ speed and non calm direction reading or calm direction reading and speeds greater than the calm threshold).

Low Variability(non-calms) - Repeated values (beyond likely repeat threshold) during non calm periods

Low Variability (calms) - Repeated values (beyond likely maximum calm duration) during calm periods indicating high likelyhood of malfunctioning or frozen instrumentation.

High Variability (steps) - Highly unlikely jumps in wind speeds.

High Variability (mean) - Anomalous periods, manually identified via summary plots.

Table 1. Summary of the station errors

Test	Total	% of record
Read Errors	1003000	7.70
Gross Errors	2012000	15.43
Low Variability (non-calm)	303	0.002
Low Variability (calms)	626900	4.80
High Variability (steps)	5790	0.04
High Variability (mean)	8021	0.06

The importance of the quality control for this study is in terms of the confidence that it gives to any conclusions drawn further down the line especially given the variable quality of the raw records.

2.4.1. Station selection

Subsequent to the quality assurance, a final selection of stations was made. In order to be included in the analysis conducted for this thesis, stations were required to:
1) be automated and use RM Young cup anemometers at 10m;
2) contain at least three distinct, 12 month periods of hourly data in the interval 1994-2010 (inclusive) with a data recovery of greater than 90%;
3) be a minimum distance from nearby obstacles of ten times the obstacle's height.

These criteria are in accordance with the standards described in the World Meteorological Organisation's guidelines(WMO, 2010). Although there are a total of 960 stations on records, only 243 of these are automated and just 107 met all of the above criteria. Figure 8 shows the distribution of SAWS stations coloured according to their agreement with these criteria and Figure 9 shows the stations that met all of the data quality criteria. The reasons for the exclusion of each station is detailed in Appendix A. Although a large number of stations have been excluded, the majority are from areas which had a high concentration of stations to begin with, so a fairly good coverage of the country remains.

At this point, it is worth asking whether these criteria are in fact excluding useful data. However, an analysis of the excluded stations shows that the majority have extremely poor records with large sections of missing data, multiple anomalous readings and inconsistencies. There are some stations that are excluded simply for being sheltered, and in these cases,

the weather stations were very poorly sited, and the sheltering can be clearly seen in the wind roses (not shown here) as well as in the reduced variability of wind speeds. Despite the careful attention paid to this process it is possible that a small number of stations have been excluded that have some useful data, but this is a risk that is outweighed by the importance of maintaining the integrity of the data set.

2.5. Conclusions

This data set forms the basis of the exploration of the wind climate of South Africa in the following chapters. With 107 stations, fairly evenly distributed across the country, and with time series that collaboratively span the period 1994 to 2010, this is the most extensive and high-quality observational data set of wind produced for South Africa to date. Although there may still be anomalies within these records, the quality control procedure has removed those errors identified as common in the literature as well as a few that may be unique to this data set. This gives at least a basic level of confidence that this data set can be used to examine the wind climate of South Africa and as a base of comparison for other data sets. Although the spatial coverage is fairly good, and the utmost care was taken in the data selection, there are some remaining limitations.

A notable limitation of this data set is the height of the masts. At 10m, the wind is sensitive to changes in ground cover, and may be influenced by the surrounding surface roughness (Vautard et al., 2010). The less the measurements are affected by the local surface effects, the more representative they are likely to be of the region. This, and the fact that most turbines are located over 50m above the ground, are the main reasons that taller masts are generally used for wind resource assessment purposes. However, since this is the only long term wind archive available, it was used with these limitations in mind and the greatest care was taken to include only the data that passed the quality control tests.

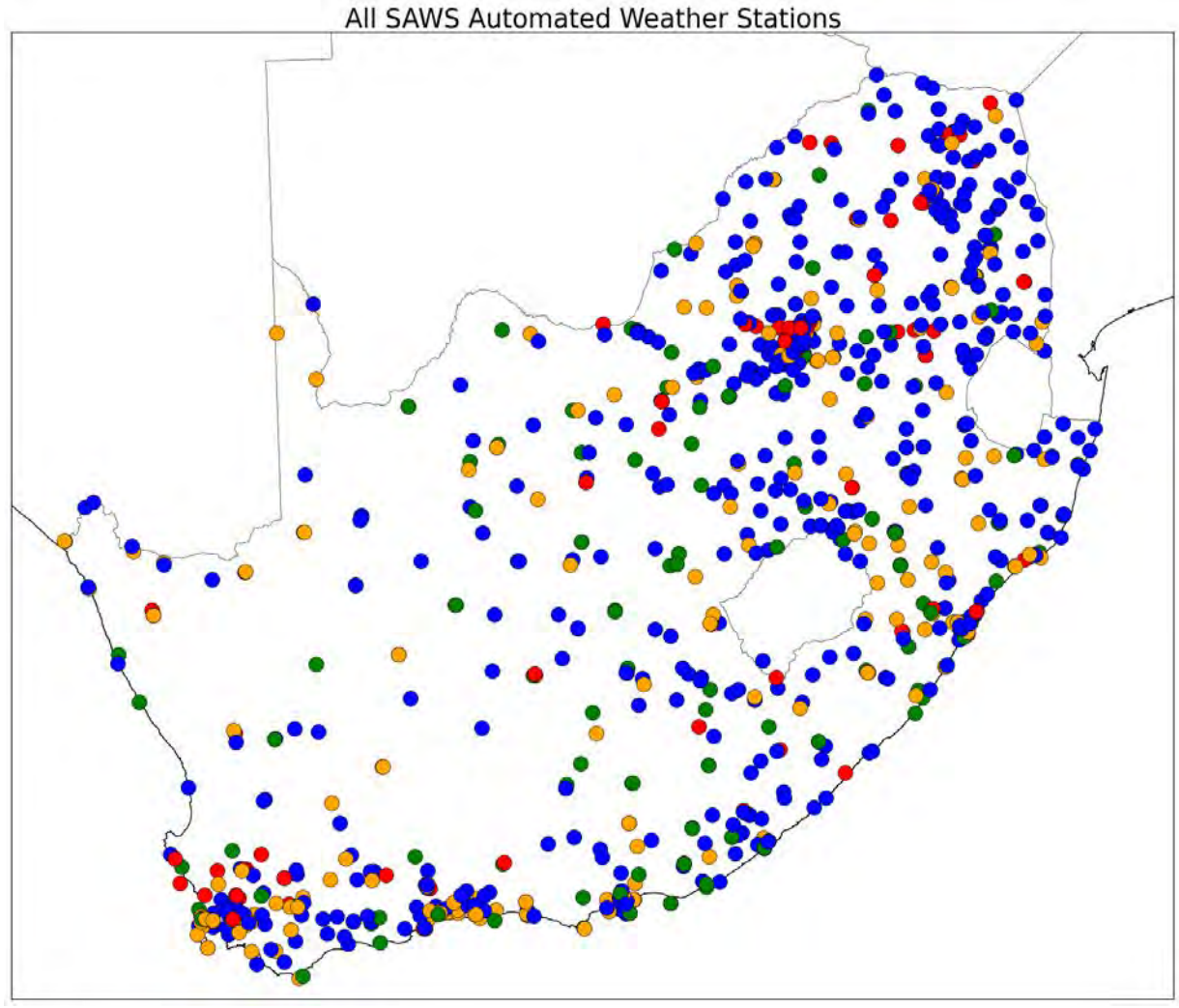

Fig. 8.— The distribution of the SAWS weather stations with wind records. The stations that met all the data quality criteria are in green. In orange are those stations that met the first two criteria but failed the third. The stations in red met only the first data criterion and those in blue failed the first criterion (mainly due to old instrumentation).

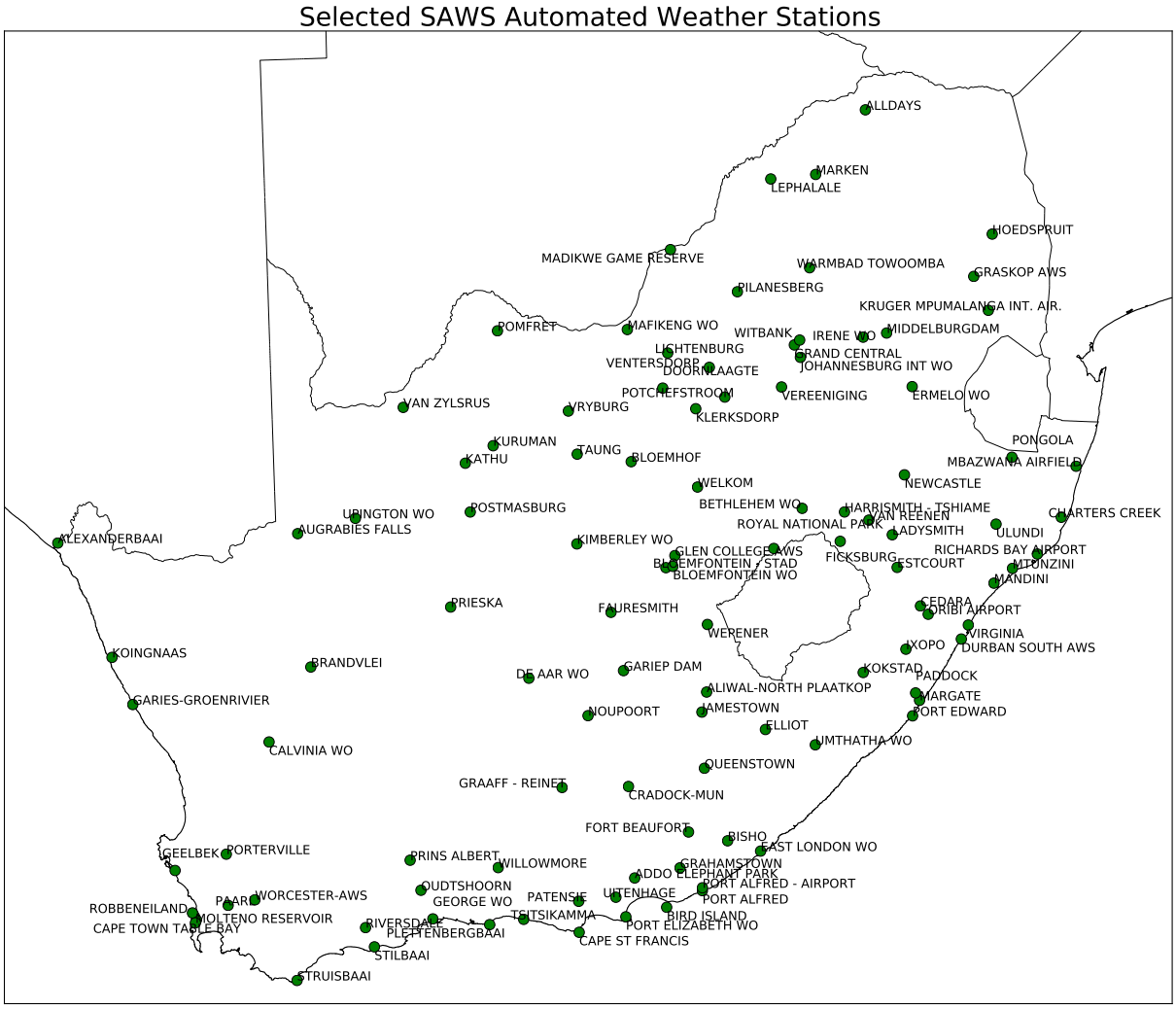

Fig. 9.— The distribution of the selected SAWS weather stations.

Despite the correction of numerous errors, a large proportion of missing data remain. Due to problems ranging from power outages, theft, electronic faults, battery failure and data transfer issues, some of the records are far from complete. Lack of available meta-data limited the success of the data recovery process significantly and increase the uncertainty as to the accuracy of the records as the details of the instrumentation are limited. The large gaps in the records limit the applicability of analyses that require concurrent measurements. In the period 1994-2010 the proportion of concurrent measurements (hours when all the selected stations had records) is only 2%. This is insufficient for comparative time series analyses and the non-uniform distribution of stations limits the extent to which spatial generalisations can be made. The following chapters seeks to address these two limitations.

3. Correction and extension of historical wind records

3.1. Introduction

In the previous chapter, the available observational wind records from the South African Weather Service (SAWS) were standardised with respect to their formatting, and filtered according to the quality and length of the station records and the exposure of the weather station sites. A quality-control (QC) procedure was developed and applied, and 107 stations were selected that met the QC criteria. Despite the fact that this is the most extensive and high-quality observational wind data set available in South Africa, the short length of many of these records, and fairly high proportion of missing data, limit the applicability of comparative time series analysis.

This chapter aims to address this shortcoming by generating complete, 30-year hourly wind speed and direction time series for each of these stations, that are as similar to the actual observations as possible. Following a review of the relevant literature, it is argued that a new method is required to achieve this, combining elements from several established techniques. A method is developed which bias corrects and downscales a high resolution reanalysis data set, in order to extend the station records back in time. A thorough validation is presented and the results are compared to the quality-controlled station records. The resulting data set provides a more suitable basis for the regionalisation of the South African wind climate (which will be covered in Chapter 4) and the exploration of the regional wind response to the large scale circulation patterns.

This chapter begins with a discussion of the differences between weather station records and model time series to contextualise the need to combine the station records with a model data set. Three major differences are highlighted between gridded model output and weather station time series, and each of these is considered in the light of the relevant literature.

A reanalysis data set is selected and the spatial and temporal variability is explored and compared with the station records. A method is then presented that adjusts the model data towards the station records at the station sites. This method is validated and the chapter concludes with a summary of the results for all of the stations.

3.1.1. Model vs observations

Chapter 4 explores the links between the surface wind measured at the stations sites and the synoptic scale circulation. The SAWS station data set compiled in Chapter 2, contains much of this information, especially regarding the seasonal and diurnal variability of wind at the station locations. However, the limited overlap and short length of the records and the high proportion of missing data limit the extent to which these data can be categorised spatially and explored in detail. As regional climate models have been shown to capture long term variability and large scale processes (Liléo et al., 2013), an appropriate product, with high spatial and temporal resolution, could provide a means to generate a longer, complete time series for each station. Before considering the possible options in this regard, there are some important differences between model and observational data that should be considered.

By their nature, weather station observations are not evenly distributed spatially and many of the records do not overlap in time. Climate model data on the other hand, though spatially and temporally complete, are limited by the available computational capacity and our ability to model small-scale processes (Giorgi and Mearns, 1999), to scales on the order of a few kilometers. All models also contain biases (Jun et al., 2008) and because they are inherently of reduced complexity, we do not expect them to be exact matches of reality.

There are also inherent differences between wind speeds from a model grid point, and those measured by a weather station near that point in reality. With advanced instrumenta-

tion technology, and LiDAR, a laser based 'radar' in particular, it is possible to measure the mean wind speed in a large cube of air with a high temporal resolution. Once these instruments are mainstreamed, more robust comparisons with the wind speed estimates produced by atmospheric models will be possible. However, the only available long records of winds in South Africa are from traditional, mast-mounted anemometers and wind vanes, and these measure (almost) instantaneous wind speed and direction at a single point. The representativity of these statistics for the immediate vicinity are relatively high in flat, uniform terrain, but as the terrain complexity rises and distance increases, this relationship becomes more tenuous (Goliger et al., 2013). So there is a scale discrepancy between the model grid point, representing perhaps a 50km area, and a station record from a single point in that area. The more complex the terrain, the greater this may be. If the point is on the top of a hill, the speed measured there is likely to be greater than the mean for the region that includes the hill due to the speed up effect of the hill. The opposite is likely true for a point in a valley.

For an imaginary cube of air near the surface of the earth, the wind inside that cube is determined by the combination of air movements at the boundaries of the cube, the interaction with topography and roughness changes within the cube and potentially local pressure effects resulting from temperature gradients in the cubes. In a climate model, the same applies, except the cube is represented by a finite number of points and the processes are captured by a set of equations. The smaller the distance between these points, the more the model landscape reflects the real landscape, but the representation of the processes may limit the extent to which increasing the resolution can improve the accuracy of a model.

In the hypothetical case where the real and model cubes are identical at the boundaries, any differences between the mean speeds and directions would be due to how the physical features of the landscape are resolved by the model and how the physics are simulated. Being of reduced complexity, the surface features in the model will be smoothed with respect to

reality, regardless of resolution. Since it is the sharp changes in roughness or surface height that will result in the highest and lowest speeds, one would then expect a reduced range of speeds in the model cube. Depending on the nature of the topography in question, this might result in a bias towards a greater or lesser mean speed or simply a narrowing of the range. The degree with which the wind direction is altered by the physical features in the cube will determine the scope of differences in directions between the model and real cube. Once again, the steeper the terrain, the greater the potential for discrepancies. With respect to the thermodynamics, the resolution will similarly have a smoothing effect, and discrepancies may be introduced by inaccuracies in how the physics are modeled.

This case is hypothetical because there are many reasons why the wind at the edges of our hypothetical cube might be different to reality at some instant in time. For example, the timing of a frontal movement, the extent of a convective cell or the strength of a large scale pressure gradient, may be different in the model world when compared to the real one. Although these processes are to some extent captured by climate models Rummukainen (2010), the differences that exist on the large scale are likely to be magnified on the grid cell scale, where a few kilometers can mean the difference between clear skies and rain in the case of a frontal system.

There are also likely to be differences between the model and real cubes as a result of how the model represents the physical processes in the cube, especially in the boundary layer (Alapaty et al., 1997). One of the ways in which models are made to look more like reality, is by incorporating observations of multiple variables by weather station, radiosondes and satellites to 'tie' the model to reality (Saha et al., 2010a). These climate 'reanalyses' are the best approximations of the real world climate available and are generally able to capture the variability over large spatial and large temporal scales (Trenberth et al., 2008; Compo et al., 2011) although they are not free of biases (Christensen et al., 2008). So while we do not

expect the hourly time series from a climate model or reanalysis at a grid level to match the hourly time series at a station exactly, we do expect the inter-annual variability and spatial relationships to be well captured. This assumption will be explored further in the following section and tested in the methods section.

In summary, there are three sources of discrepancy between climate model data and weather station records:
1) The climate model atmosphere is discretised into cubes which are greater in scale than the area generally represented by any particular station.
2) Due to the simplification and approximations made in the physics of climate models, we do not expect climate model data to match reality exactly.
3) Climate models, and even reanalysis products, have inherent biases which reduce the similarity to observations.

With these limitations in mind, the aim of this chapter is to combine the strengths of the climate models in capturing the large scale circulation, long term trends and spatial relationships, with the detailed local information contained in the station records, to produce a long time series of wind data for each station. The following section explores the existing techniques that are used to bridge the scale gap between the regional models and station data, remove systematic biases and include long term variation missing from short station records.

3.2. Literature review

There are three bodies of literature that are of particular relevance to the task of producing long, representative wind records. The first relates to bias correction of climate model data, the second to downscaling of climate model data and the third to incorporating long

term variability. The relative merits of each of these will be considered in the light of what each can offer to the specific aim of this chapter.

3.2.1. Bias correction

In order to use climate model data to study phenomena at a finer scale than the resolution of the global, or even meso-scale models, adjustments to the data need to be made to account for the difference in scale and the bias of the models (Christensen et al., 2008). There are many bias correction techniques used particularly by groups modeling the impacts of climate change. Hydrology and crop modeling are two areas in which bias correction is widely used for precipitation and temperature. Some examples of the bias correction methods commonly used for these variables include: Multiple linear regression (Hay et al., 2002), monthly mean correction (Fowler and Kilsby, 2007), local intensity scaling (Schmidli et al., 2006) and quantile-mapping (Wood et al., 2004; Sun et al., 2006). For a review of these methods, see Ehret et al. (2012) and Themessl et al. (2011). Although generally developed for temperature and precipitation, some of these methods are applicable to the bias correction of wind as well (Wilcke et al., 2013).

Many studies have argued for the need to bias correct model output. Christensen et al. (2008), describe the need for bias correction of climate models in order for the data to be of use at a local scale. Piani et al. (2010) show that spatial distributions of precipitation can be significantly improved by bias correction and Haddeland et al. (2012) found that bias correction of wind, among other variables, improved simulated water fluxes. Infilling and extension of historical records is widely accepted used as a means of reducing biases in climate data sets and the agMERRA (Modern-Era Restrospective Analysis for Research and Applications)and agCFSR (Climate Forecast System Reanalysis) data sets from the AGMIP (Agricultural Model Intercomparison and Improvement Project project) (Ruane et al., 2015)

are examples of these in the context of agricultural modeling.

Although these, and many other studies, have argued for the need to bias correct model output, other authors argue that bias correction is the wrong approach to solving the problems with regional model data. Ehret et al. (2012), for example, questions the physical basis for bias correction of model data and argue that many of these methods disrupt the spatial and temporal relationships conserved by these models. Although this is a concern in the application of a bias correction for this research, both the necessity of bias correction, and the possibility of avoiding some of the pitfalls have been demonstrated. Wilcke et al. (2013) have shown that quantile mapping of rainfall, for example, does not effect the correlation with other variables and retains the temporal quality of the record. These authors also demonstrate the applicability of quantile mapping to variables other than precipitation, including wind.

From the perspective of this study, where the maintainance of spatial relationships is particulary important, the explicit separation of space and time in most of the methods poses a potential problem. However, the ability of some of the bias correction techniques to generate time series that closely match the real temporal evolution and variance is an attractive attribute. Haerter et al. (2011) argue that since different processes operate at different time scales, bias correction should be applied at different timescales as well. A quantile-mapping procedure which incorporates these elements has potential to significantly reduce the bias in the model time series while avoiding some of the pitfalls associated with other bias correction methods.

3.2.2. Downscaling

Downscaling is another approach to dealing with the inherent differences between climate model data and station observations and refers to the processing of model data to make it applicable at a finer scale. Statistical downscaling achieves this by mapping model data to observation points by means of particular transfer functions. These are widely used as simpler, faster alternatives to dynamical downscaling, the nesting of higher resolution models within the region of interest, (Wilby et al., 1998) and for exploring the local impacts of projected climate change (Christensen et al., 2008). Statistical-dynamical approaches lie between these methods.

There are many statistical downscaling techniques in the literature. These range from those that use neural networks (Sailor et al., 2000) to more probabilistic approaches (Kirchmeier et al., 2014), to physical -statistical approaches (Rooy and Kok, 2004) and those that downscale distributions as apposed to atmospheric variables (Devis et al., 2013). These are predominately applied to temperature and precipitation, however, they have been used for other variables, including wind forecasts (Howard and Clark, 2007). The key feature of these methods is the development of a relationship between the grid-scale circulation and the local variability as measured at a specific site.

As an example of a statistical dynamical downscaling, Bergström (2001) used the Meteorological Institute of the University of Uppsala (MIUU) model to characterise forcings by the season, wind speed and direction. Statistical relationships were then developed between the classes defined by these attributes and the wind observed at specific sites. More recently, Badger et al. (2014) used a statistical-dynamical downscaling technique to estimate the wind climate at a potential wind farm site. Wind classes based on speed, direction and Froude number were defined as part of the KAMM-WaSP method. In this method, the results from the Karlsruhe Atmospheric Mesoscale Model (KAMM) are post-processed with the

micro-scale model. Hahmann et al. (2015) applied a similar method driven by the Weather Research and Forecasting (WRF) model as opposed to KAMM for the South African Wind Atlas. The key stage in both of these methods is the generalization of the wind climate by the removal of the speed-up effects of topography and roughness changes as seen by the mesoscale model, and a replacement with the micro-scale effects modeled by the Wind Atlas Analysis and Application Program (WAsP) (Mortensen et al., 1993). In this way, the scale difference between the two is bridged to allow for a fair comparison between the mesoscale grid point time series and observational wind records.

Although downscaling has been successful in many cases, as Pryor et al. (2005) show, the coupling between the regional flow varies between stations and may limit the success of downscaling from regional models. In the follow chapter, this coupling is explored in some detail.

3.2.3. Long term variability

The third set of techniques that is relevant to the objectives of this chapter relates to the long term variability of wind. In the wind energy industry, evaluation of potential wind farm sites is based on on-site wind observations. The short length of the observational records typically available from potential sites poses a problem as these do not capture the variability that the site may experience during the lifetime of a wind farm. For this reason, long term corrections are routinely applied to the records (Liléo et al., 2013). There are many long-term correction (LTC) methods described in the literature, however their common aim is to estimate the wind resource at a site based on a short record (generally 1-3 years) from the site in question, and a reference record (15-20 years). A long period time series is usually obtained, for this purpose, from a nearby measurement mast or from a grid point from a high-resolution model or reanalysis product (Probst and Cárdenas, 2010; Liléo et al., 2013).

The reference record is assumed to be representative of the long term variation of the wind conditions at the site in question. A correction method is then used to determine the long term wind characteristics of the site with the assumption that the variability observed in the past is a good predictor of the variability in the future. This is a source of uncertainty in these methods, as is the choice of reference record, reference period and correction method. Although these long term correction methods have slightly different aims to this study, there is considerable overlap, especially as they are specifically developed for wind, as opposed to temperature or precipitation, and some of the records in this study are similarly short.

Liléo et al. (2013) describes the most common long term wind correction methods as either regressive or non regressive. There are several different regression techniques used in order to define the relationship between the wind climate at the reference and measurement sites respectively. Least squares and principle component regression are examples of these (Thøgersen et al., 2007). Non-regressive methods include the matrix method, which involves applying correction functions that depend on the wind speed and direction, and the quantile-quantile-method (Thøgersen et al., 2010; Liléo et al., 2013) which has similarities to the quantile-mapping methods used for bias correction.

Specifically, these techniques are aimed at estimating the long term statistics of wind that are relevant for wind power generation, and are not as concerned with the daily or hourly time evolution of the wind over the period. However, the aim of using a reanalysis data set to estimate the long term variability in the wind at a site, given a short observation record, is similarity to the task addressed in this chapter.

Another aspect of long term variability is trend analysis. Pryor et al. (2009) for example, explores the trends in wind speeds in the United States as seen in observations, reanalysis products and climate models, finding downward trends across the continent and discrepancies between mean values as well as wind speed trends across the different data types. No

clear relationship could be established between inter-annual variability and long term trends in wind speeds. In another example of a study of long term variability in wind speeds Vautard et al. (2010) explore downward trends in wind speeds in Europe, finding a 5-15% decrease, which is attributed in part to changes in circulation and in part to increased surface roughness. These studies provide a point of comparison for the results of this analysis.

3.2.4. Discussion

There are some challenges that are unique to the task of producing long, representative time series of wind for the South African stations. Apart from the limited station data and relatively low station density, most of the existing methods have been developed in the northern hemisphere and often for variables other than wind. These factors contribute to making this task particularly difficult considering the varied South African terrain.

In terms of developing a method to achieve long term, complete, representative station records, a high resolution reanalysis product would offer a means to extend the station records back in time and a spatial covariance structure would lend itself to extending the correction to the rest of the region. The challenge is then to develop a method that is able to remove the bias in the reanalysis, while maintaining the long term variability and also bridge the scale gap between the model resolution and the station scale.

A new method, that combines elements of the techniques described above, is therefore suggested, to deal with this particular collection of challenges. In the Methodology section, three methods are presented that map from a reanalysis data set to the station records, incorporating elements of some of the correction techniques described above to remove the bias from the data while maintaining the station variance and spatial covariance structure. By combining these methods in a novel manner, some of their respective strengths can be

leveraged and weaknesses avoided. The success of these methods depends largely on the characteristics of the reanalysis data set and so, before presenting the method details, the following section explores the suitability of the selected data set with respect to the spatial and temporal variability of wind speed.

3.3. Reanalysis data

Although the SAWS station records are the best source of surface wind observations available, the deficiencies in the lengths of the records and large proportion of missing data need to be overcome. In line with the recommendations of Liléo et al. (2013) this study uses a reanalysis data set to improve the temporal completeness and continuity of the records. Specifically, the aim of this chapter is to produce a complete, hourly time series for each station, covering the period 1979-2010, that is as similar as possible to existing time series for the period where data is available, while maintaining spatial relationships between stations as well as their time evolution. For the periods where data are not available, these time series would provide an estimate of the likely observations, based on the relationships developed to the reanalysis data.

There are several reanalysis products available, however, a subset of the National Centers for Environmental Prediction's (NCEP) Climate Forecast Systems Reanalysis (CFSR) data set was chosen for this task as this data provides the highest spatial and temporal resolution. The 'Selected Hourly time series Product' consists of U and V components of wind derived at 10m from January 1979 to December 2010 on a 0.313 ° x 0.312° grid in a box bounded by lines at -20° and -40° latitude and 10° and 35° longitude (CFSR, 2010). These data are in fact a collection of hourly forecasts based on the six hourly reanalysis data set (initialised hourly) and can be freely downloaded from the Computational and Information Systems

Laboratory (CISL) Data Archive [i]. The grid domain can be seen in Figure 10, which shows the model terrain height and the positions of the selected SAWS stations.

3.3.1. CFSR: The big picture

In order to explore how the South African wind climate is simulated by the CFSR, the following section describes the CFSR spatial and temporal variability of wind before comparing the results at the 107 SAWS station sites. The specific differences between the stations and the CFSR data will guide the development of the correction developed in the methods section. There are four questions that this section seeks to answer in assessing the representativity of the CFSR wind fields.

How do the CFSR winds compares to those measured at the SAWS stations with respect to:

1) Mean wind speed

2) Inter-annual variability

3) Long terms trends

4) Spatial variability

In order to address these questions, the mean wind speed, standard deviation, and long term trends of the CFSR mean annual wind speed were calculated and are shown in Figure 11.

From the left panel it is clear that there is a large range of mean wind speeds from just over 2m/s to over 5m/s. Speeds are higher in the coastal areas, mountain ranges (excluding the Drakensberg) and in general, the south western half of the country. However, there is a remarkably low inter-annual variability, with standard deviations of mean annual wind speeds at around 1% for most of the country. There are a few areas where the variability

[i]Available at *http* : *//rda.ucar.edu/*. (Accessed June 2013)

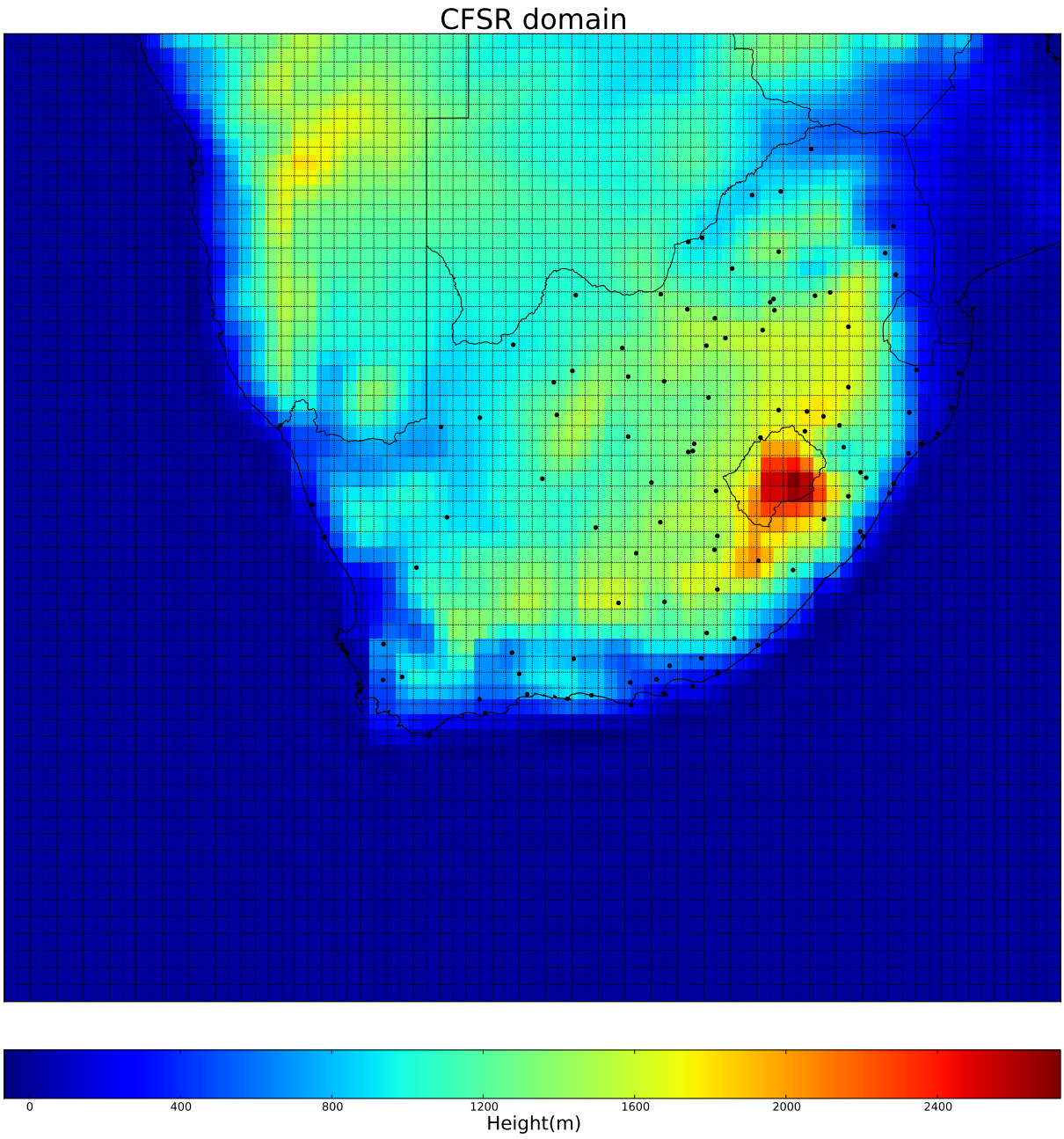

Fig. 10.— CFSR domain subset with terrain height and SAWS station positions indicated. The grid resolution is 0.313 ° x 0.312°, in a box bounded by lines at -20° and -40° latitude and 10° and 35° longitude

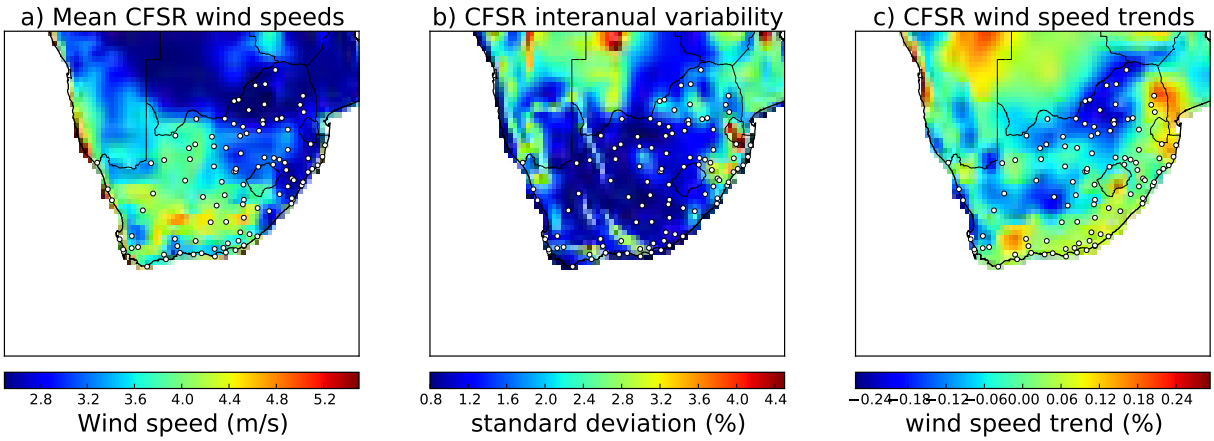

Fig. 11.— a) CFSR Mean annual wind speeds, b) standard deviation of inter-annual wind speeds as a percentage of the mean and c) trends in wind speed as a percentage of the mean, from the hourly 10m CFSR winds in the period 1979 -2010. White markers indicate the SAWS station sites.

is higher, and although it is not immediately clear why this is the case, it may be due to these regions being on the border of zones dominated by different synoptic forcings. Despite high statistical significance (not shown here), the long term trends in wind speed, are all small, with the strongest trends reaching a quarter of a percent per year. This amounts to just over a 7% change in over 30 years, however, most of the trends are nearly an order of magnitude smaller and considerably smaller than the inter-annual variability. In order to explore the results shown in Figure 11, the nearest grid point to each of the selected SAWS stations was extracted, and, using just the hours for which observations were available, the mean and standard deviation of wind speed were calculated. A linear regression was applied to determine the trends in wind speed, the results of which are shown in Figure 12.

From Figure 12a, it is clear that for most stations sites, the CFSR speeds are considerably higher than those observed.[ii] This is an important result as it could significantly

[ii]It should be noted that model winds were set to zero when they drop below 1m/s for

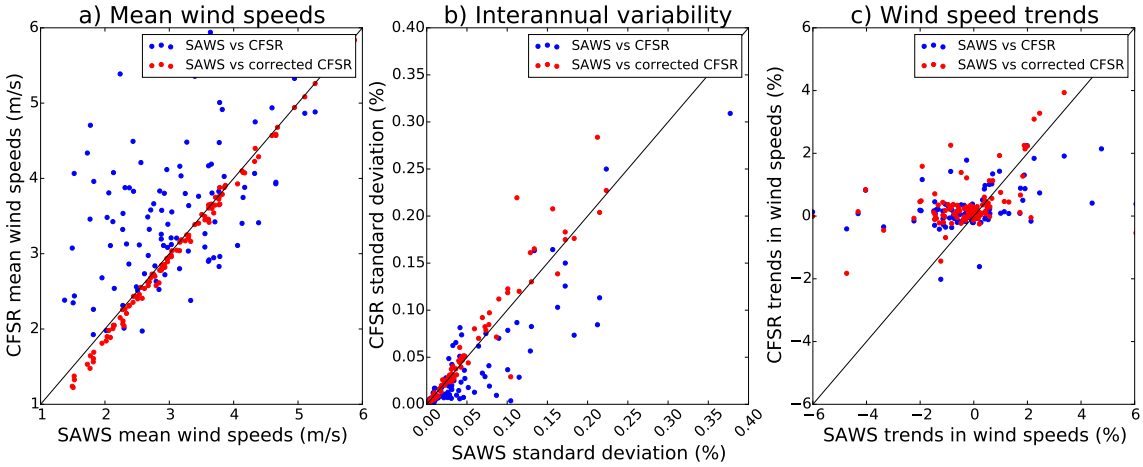

Fig. 12.— A comparison of the mean (a), standard deviation (b) and trends (c) in wind speed from the SAWS weather stations and the corresponding CFSR grid cells. The standard deviations and trends are shown as a percentage of the mean speeds. The SAWS records cover a minimum of three years between (1994 and 2010) and the 10m hourly CFSR time series from the closest grid point to each station, for the same periods, are used for comparison. Each point represents a single station.

impact the results of wind resource assessments that make use of the CFSR winds. It also strengthens the case for a bias correction of the time series. The inter-annual variation in wind speeds, as measured by the standard deviation of mean annual speed, shown in Figure 12b, is a small fraction of a percent for most of the sites. While the CFSR values are generally lower than those observed, the absolute differences are negligible. Figure 12c shows the trends in wind speed. For most of the SAWS stations, these trends are between -2% and 2% per year, however there are some outlier trends of magnitudes of around 5% per year. The CFSR records show a far smaller range of trends, with most corresponding grid cells showing no discernible change over the 30-year period. This is another concerning result in the context of wind resource assessment, and calls into question the low level of trends seen in across the country in Figure 11c.

In order to address the fourth question of this section, regarding the relationships between different regions, the Pearson correlation of wind speed was calculated between each CFSR grid cell corresponding to a SAWS station, and each other station grid point, a total of 11449 values. The same was done for the station time series. Since not all of the station records overlap in time, only those where there at least a year's worth of overlapping hours were used, and the same hours were included for the CFSR correlations to ensure a fair comparison. Figure 13 shows the difference between the cross correlation of each SAWS station record with each other, and the corresponding cross correlations from each CFSR grid point time series, coloured by the distance between stations. The significance of the correlations varies between stations but is generally high, with p-values in the range 0.001 - 0.05.

As one would expect, the correlations reduce with distance, however there are some differences between the spatial relationships as seen by the CFSR records and those that are observed at the stations. In general, the CFSR grid-points that are close together show

---

the sake of comparison with the observations.

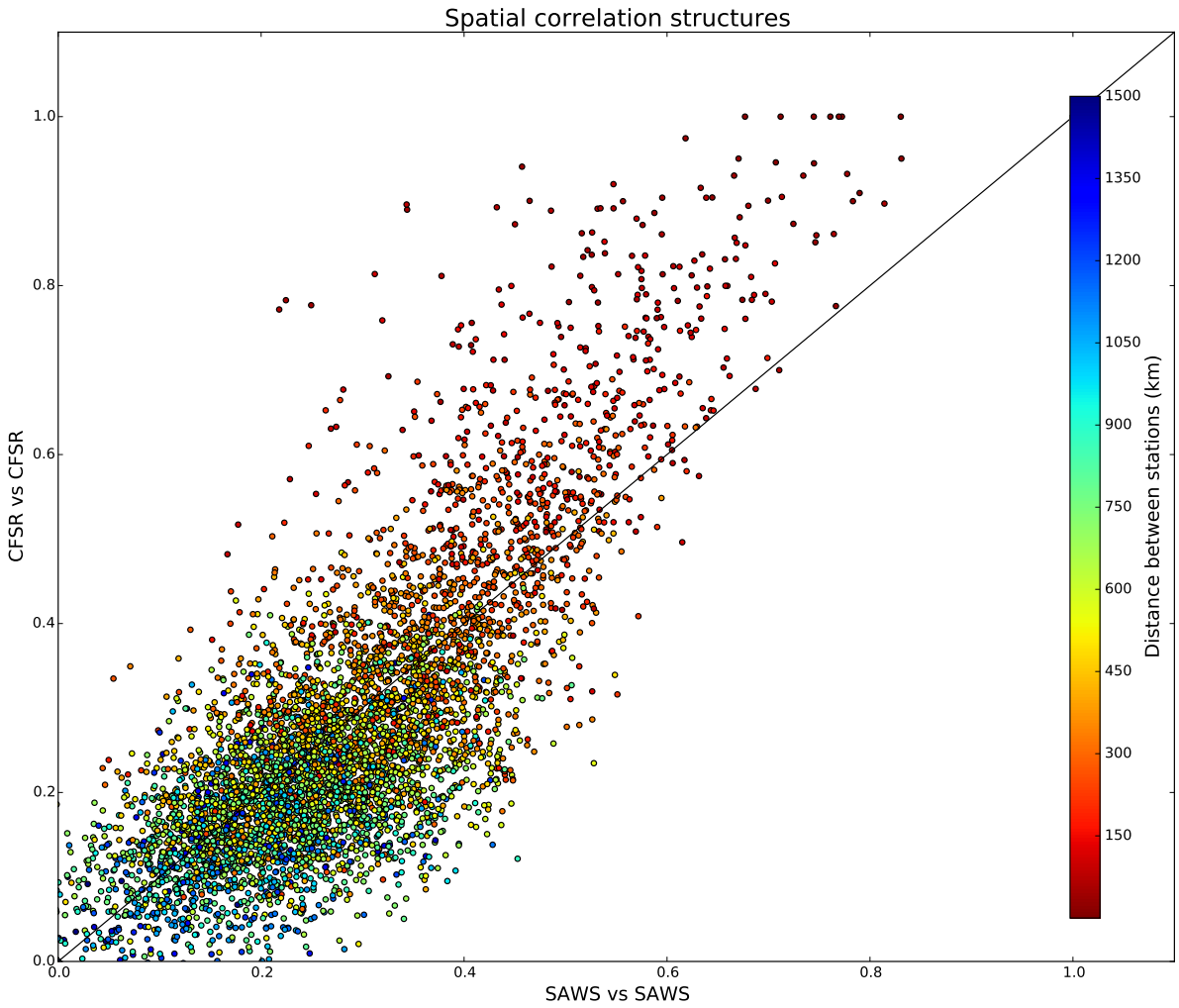

Fig. 13.— A comparison of the observed and CFSR cross correlation patterns between 107 SAWS station sites. Each of the markers represents the correlation between two station points (for the period when their records overlap), with the colour indicating the distance between them. The station records have a minimum of 3 years of data in the period 1994-2010 and the matching years were taken from the 10m CFSR hourly time series for the closest grid point to each station. The grid resolutions is 0.313 ° x 0.312°.

higher correlations than are observed, and those that are far apart show lower correlations than are observed. In other words, the CFSR over estimates the distance dependence of the correlations. There are also many grid points where the CFSR records vastly over estimate the correlation, irrespective of distance. In general, the agreement is poor. In order to get a sense of how these discrepancies translate at individual stations, the following section considers two stations, Koingnaas and De Aar, and their corresponding CFSR grid points, in some detail.

3.3.2. Comparison with stations records

In order to assess how well the CFSR data compares to the station data, two station were selected in different areas and compared to the nearest CFSR grid cells. These two stations represent two distinct terrains, with one on the edge of west coast and the other in the interior of the country.

3.3.3. Koingnaas

Koingnaas is a tiny mining town in the Namakwa region of the west coast of South Africa. The weather station is located at the airport, about 4km from the coast, in very flat terrain and can be seen in Figure 14.

Figure 15 [iii]contains a series of plots which summarise the station record and the corresponding CFSR time series. The period shown is the extent of the station record (in blue) and the breaks in the line are periods where less than 90% of the hours in that month are available. The four plots within the figure show different elements of the station's climatol-

[iii]All times shown in figures and described in the text are in the local time, UTC +2.

Fig. 14.— Koingnaas weather station (SAWS, 2012c)

ogy, with the equivalent CFSR statistics from the nearest model grid point in green. The seasonal and particularly strong diurnal cycles are clear and combine to a fairly typical distribution of wind speeds in the histogram. The CFSR record slightly underestimates the wind speeds throughout the year, more considerably during the winter months, and captures the diurnal cycle fairly well. The CFSR histogram is almost identical to the observed, missing only the frequency of the highest wind speeds. It should be noted that not all of the stations are as well simulated as this example. In order to explore this station further, consider Figures 16 and 17 showing the seasonal and diurnal cycles as well as the histogram, in more detail.

Figure 16 is a combination of the seasonal and diurnal cycles, with hours in the day along the x-axis, and months of the year on the y-axis. The top two panels show the wind speeds from the station and CFSR records respectively, and the bottom two, the corresponding mean directions. Care should be taken in the interpretation of the direction plots as the mean of a set of directions which contain opposing winds may not be meaningful. Differences between the station and CFSR records are therefore more significant than the individual distributions. In this figure, the subtle differences seen in the observed and CFSR diurnal and seasonal cycles are evident. The observed diurnal cycle shifts seasonally from predominately southerly winds throughout the day in summer with a peak in speed at around 17h00, to more northerly winds in winter, swinging towards the west during the day. The winter months are captured fairly well by the CFSR records, slightly underestimating the wind speeds in the mornings and showing a shortened winter season.

The panels in Figure 17 shows the relative frequency of different wind classes. The colour of each square gives the relative frequency with which that wind speed-wind direction

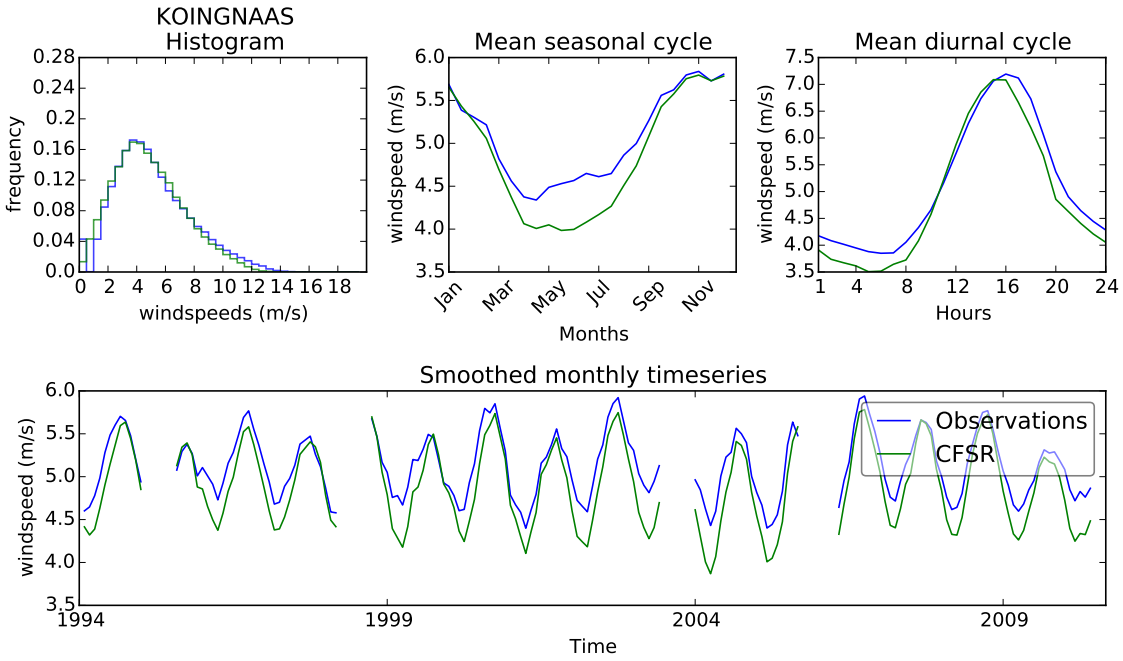

Fig. 15.— Summary figure showing (clockwise from top) the wind speed histogram, seasonal and diurnal cycles and mean monthly wind speed for Koingnaas. In each case the station records are show in blue and the CFSR data in green. The gaps in the time series indicate missing data. The CFSR time series consistently underestimates the wind speed by about 0.1m/s, slightly more between April and July.

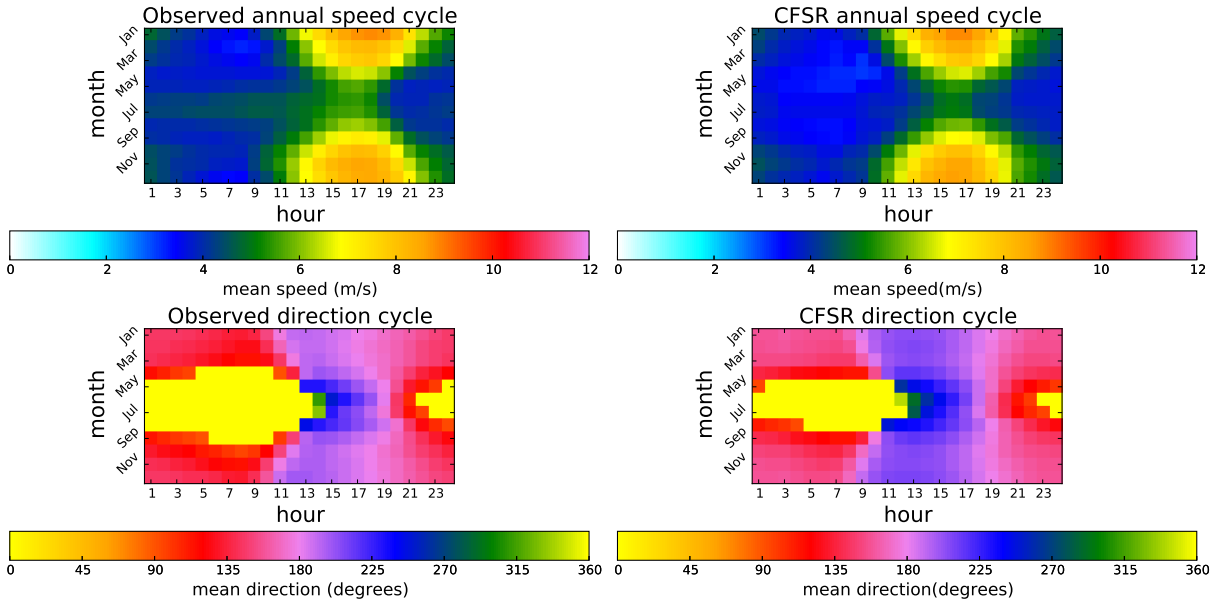

Fig. 16.— Mean observed and CFSR wind speeds and direction for each hour in each month from 1994-2010 for Koingnaas. The timing and intensity of the seasonal and diurnal patterns is well captured with a small directional bias in the shoulder seasons.

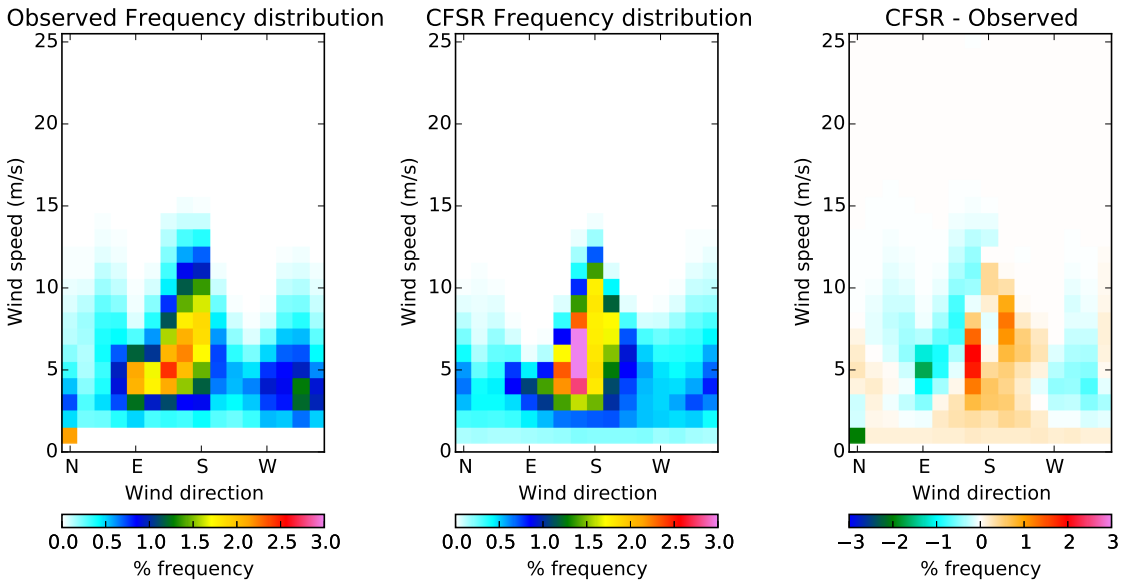

Fig. 17.— 2D histograms of the observed and CFSR wind speeds and directions at Koingnaas from 1994-2010. Although the general shape of the distribution is well captured, the CFSR over estimates the frequency of 4-7 m/s southerly winds and under estimates the frequency of westerlies and easterlies. The bottom left corner bin indicates the frequency of calms.

combination occurred during the observation period. An alternative to the traditional wind rose, these plots highlight the different wind modes and are better suited to displaying the differences between two wind records. The two dominant wind modes observed at Koingnaas are the 4-10m/s South to South-Easterlies and the 3-8m/s North Westerlies. These modes are not as clearly defined in the CFSR time series and while the distribution of southerly classes is fairly well represented, the north westerly classes are more frequent in the observed time series. The bottom left hand bin indicatives the frequency of calms as that the anemometer does not register wind speeds under 1m/s and directions are set to zero in these instances. The same criteria is applied to the CFSR records in the next stage of analysis.

This station is an example of a region where the differences between the CFSR and the station records are minimal. Stations in more complex terrain showed more dramatic disparities but for many of the stations in the flatter areas, the CFSR data produced similar statistics.

3.3.4. De Aar

The De Aar weather station, shown in Figure 18, is situated on the outskirts of the tiny Karoo town of De Aar, far from any significant topography , near the centre of the country.

Figure 19 shows the summary plots comparing the CFSR results with the De Aar station record. With a similar seasonal and diurnal pattern to Koingnaas and a comparable range of wind speeds the only major difference evident here is the lower inter-annual variability. The landscape is similarly homogeneous and CFSR time series matches the observations fairly well up to a consistent 1m/s bias. The error seems to be due to the timing and range of the diurnal cycle with the CFSR peak wind speeds a couple of hours early and a little over 1m/s

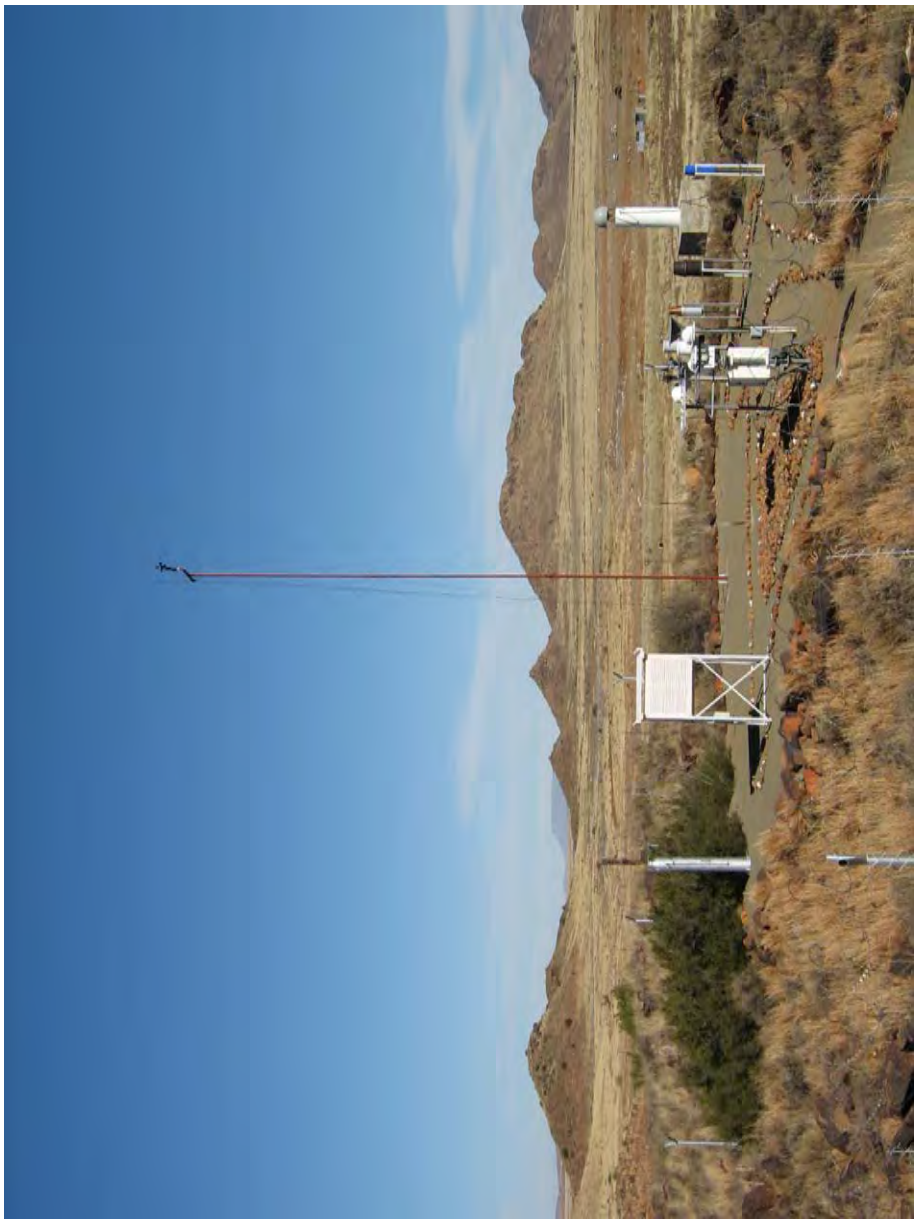

Fig. 18.— De Aar weather station (SAWS, 2012a)

lower than the observed.

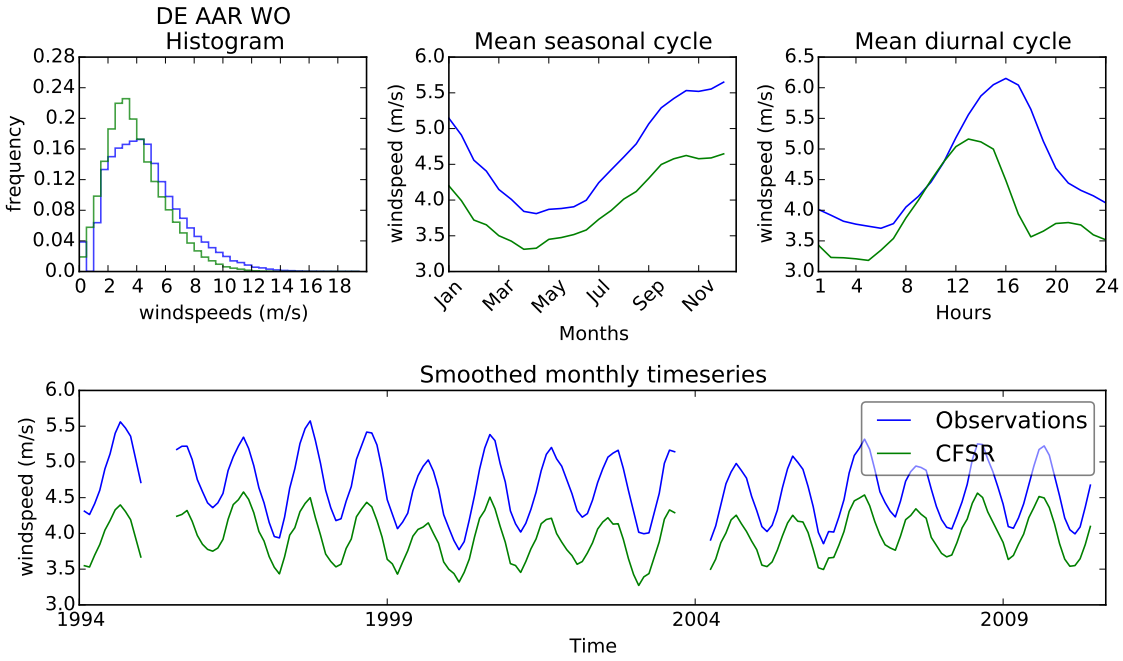

Fig. 19.— Summary figure showing (clockwise from top) the wind speed histogram, seasonal and diurnal cycles and mean monthly wind speed for De Aar. In each case the station records are show in blue and the CFSR data in green. The gaps in the time series indicate missing data. Although the CFSR matches the shape of the diurnal and seasonal cycles, the time series consistently underestimates the wind speed by almost 1m/s.

Figures 20 and 21 show more details on these aspects of the De Aar wind climate.

Figure 20 shows the clearly defined wind modes, although there is more 'smear' than for the Koingnaas station. The two dominant modes for this station are a 2-9m/s South Easterly mode and a 3-10m/s Northerly mode with higher speeds contributed by the Westerlies. Again, the higher end of the distribution is not captured by the CFSR. The seasonal and diurnal pattern of these modes is depicted in Figure 21.

A sharpening of the diurnal cycle occurs in the winter months with 3-4m/s mean wind

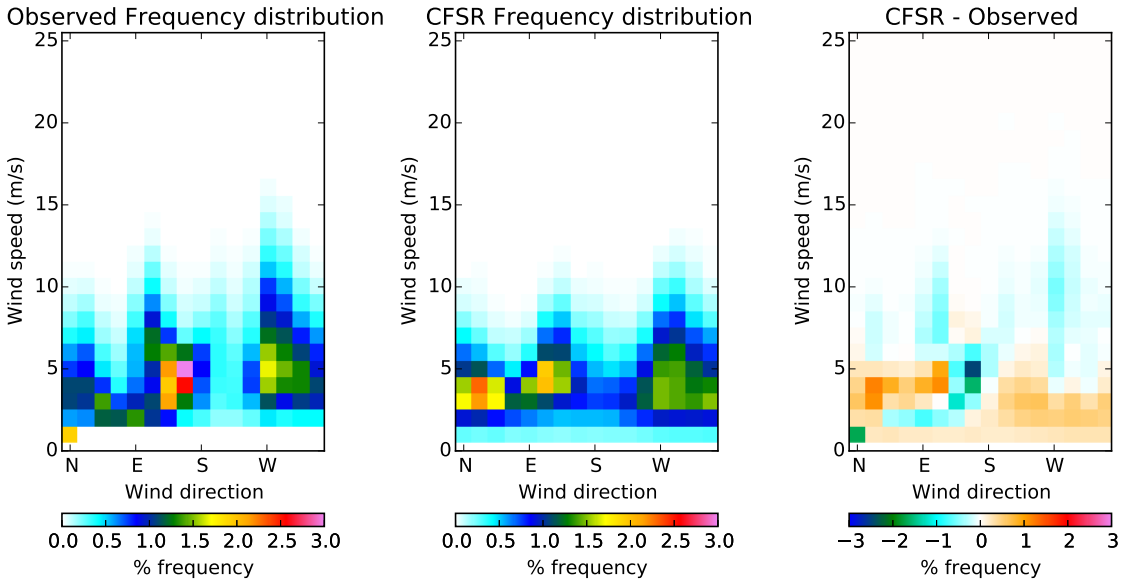

Fig. 20.— 2D histograms of the hourly observed and CFSR wind speeds and directions at De Aar from 1994-2010. The left panl highlights the under estimation of the wind speeds about 6m/s, in favour of lower speeds.

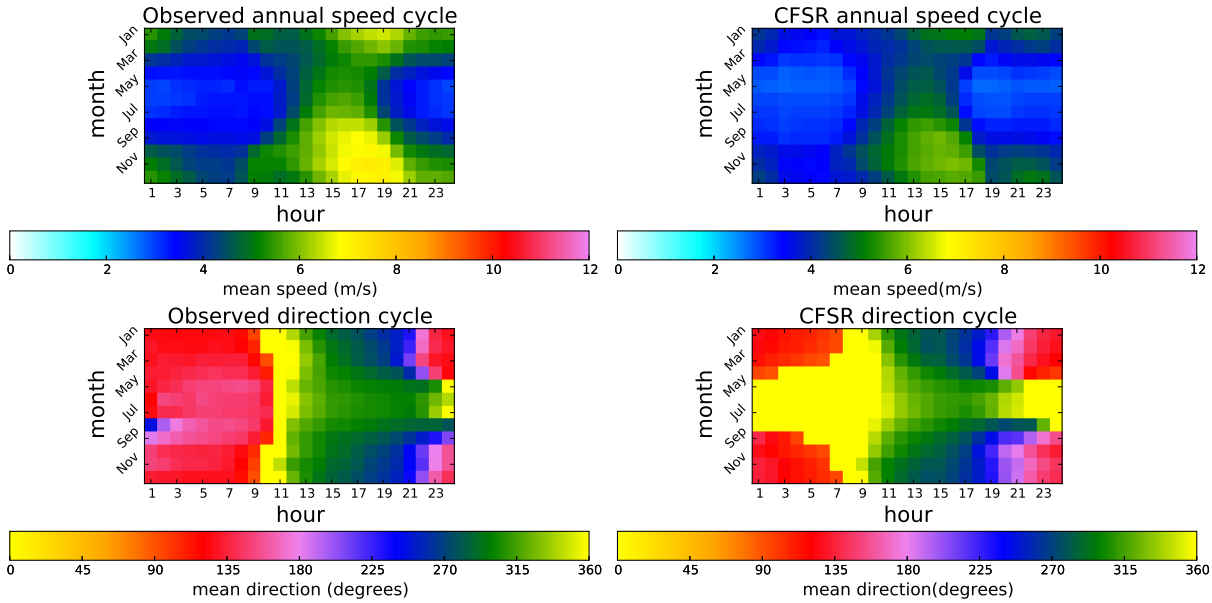

Fig. 21.— Mean observed and CFSR wind speeds and direction for each hour in each month at De Aar from 1994-2010. The under estimate of wind speeds is evident, as is the wind direction bias in the mornings from May to November.

speeds for most of the day jumping to 6m/s mean speeds between 1pm and 7pm. This shift is evident in the CFSR plot whereas the higher speeds observed in the summer mornings and evenings are not. The directional plots show a steady clockwise rotation of the winds through the diurnal cycle with South Easterly morning winds shifting through Easterlies to Northerlies around 10am and then on through North Westerlies in the afternoons into the evening and round through South and back to the South Easterly before midnight. The same pattern is clearly evident in the CFSR records during summer, but remain Northerly during winter, missing the Easterly morning winds.

3.3.5. CFSR validation

This manual comparison is not well suited to high numbers of stations and is relatively subjective. In order to compare the gridded CFSR data with all the station observations, a set of metrics was selected and developed to compare the CFSR wind time series calculated from the U (zonal) and V (meridional) components of the wind at 10m, from the closest non-ocean grid point to each station. The wind speed and direction are given by equations 2 and 3.

$$v = \sqrt{(U^2 + V^2)} \tag{2}$$

$$\theta = arctan(V/U) \tag{3}$$

In order to be able to make quantitative comparisons between wind time series, whether from different stations, or between station records and gridded data sets, it is necessary to select the characteristics to compare, and to define metrics that measure the relevant differences. As this study is concerned with wind climatology, and the implications for wind

energy, the metrics were selected to reflect the needs of these fields and to quantify the differences highlighted in the figures. The metrics chosen and the attributes measured are summarised in Table 2.

For two hourly wind time series A and B, of length n, the root mean square error of wind speeds (RMSE), the mean (mulitplicative) error of wind speeds (MEWS) and mean error of wind direction (MEWD), which are defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(v_{Ai} - v_{Bi})^2} \tag{4}$$

$$MEWS = \frac{1}{n}\sum_{i=1}^{n}(v_{Ai}/v_{Bi}) \tag{5}$$

$$MEWD = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}(\theta_{Ai} - \theta_{Bi}) & for\ |\theta_{Ai} - \theta_{Bi}| < 180 \\ 360 - \frac{1}{n}\sum_{i=1}^{n}(\theta_{Ai} - \theta_{Bi}) & for\ |\theta_{Ai} - \theta_{Bi}| > 180 \end{cases} \tag{6}$$

Table 2. Selected error metrics

Metric	Attribute
Root mean square error	The hourly differences in wind speeds
Mean Error	The magnitude of the bias in wind speed and direction
Distribution Error	The difference in distribution of wind speed-direction "classes"
Cycle error	The difference in seasonality of diurnal cycles
Vector Correlation	The temporal similarity of the time-series

Here n is the number of hours , v and θ are the wind speed and direction respectively the subscripts A and B refer to the two time series. These metrics give a baseline measure of the similarity of two time series of wind. A limitation of the the MEWS and MEWD are that they capture no information about how two time series differ in their time-evolution. The RMSE gives some information on this aspect however further metrics are required to understand the nature of the differences highlighted with these simple measures.

Since the wind speed and direction are not independent variables, the Distribution Error (DE), Cycle Speed and Direction Errors (CSE and CDE) and Vector Correlation (VC) are used to measure the difference between the two-dimensional attributes of the time series. The DE, as defined in equation 7, is essentially the mean absolute difference between the histograms of wind speed-wind direction pairs for the two time series. The method used to calculate this is as follows: Each of the n observations is binned into one of 25 speed bins (each spanning 1m/s) and further into one of 16 direction bins. The total count in each bin is converted to a percentage of the total n. This is repeated for the second time series and the absolute value of the difference in resulting distributions is calculated bin-by-bin. For distributions with no overlap at all, the resulting difference would be 200 and for perfectly matching distributions, 0. This value was therefore divided by 200 to normalise the range to between 0 and 1.

$$DE = \frac{1}{200} \sum_{s=1}^{s=25} \sum_{d=1}^{d=16} |freq_{Asd} - freq_{Bsd}| \tag{7}$$

The s and d subscripts indicate the speed and direction bins respectively.

The calculation of the cycle error uses a similar principle to measure the difference between the seasonal and diurnal cycles of two time series. For the speed error, the wind

speeds are binned according to the month of the year and further by the hour of the day. The mean speed for each bin is calculated to depict the seasonality of the diurnal cycle. Having calculated these values for two time series, a simple differencing results in the mean difference in wind speed per hour per month. The average across all 288 bins gives the cycle speed error in m/s.

$$CSE = \frac{1}{288} \sum_{m=1}^{m=12} \sum_{h=1}^{h=24} (\overline{v}_{Amh} - \overline{v}_{Bmh}) \tag{8}$$

The equivalent process for wind directions, with circular averages, results in the cycle direction error in degrees.

$$CDE = \frac{1}{288} \sum_{m=1}^{m=12} \sum_{h=1}^{h=24} (\overline{\theta}_{Amh} - \overline{\theta}_{Bmh}) \tag{9}$$

The subscripts m and h indicate the month and the hour respectively.

The definition of the vector correlation is that suggested by Crosby et al. (1993).

$$VE = \sqrt{Tr[(\Sigma_{11})^{-1}\Sigma_{12}(\Sigma_{22})^{-1}\Sigma_{21}]}/\sqrt{(2)} \tag{10}$$

Where Σ_{11} is the covariance matrix of A and Σ_{22} is the covariance matrix of B, Σ_{12} is the cross-covariance matrix of A and B and Σ_{21} is the cross-covariance matrix of B and A. Tr denotes the trace and the value of VE is normalised to lie between between 0 and 1.

The calculation of these metrics was done with code written in the python scripting language. [iv] The combination of these metrics gives a clear picture of the differences and similarities between two time series.

[iv]For more information on python see *www.python.org*.

The results of the application of these metrics to each of the SAWS station time series are shown in Table 3.

Table 3. Mean Errors for CFSR station timeseries.

ME (m/s)	RMSE (m/s)	DE [0,1]	CSE (m/s)	CDE (^{o})	VC [0,1]
0.61	2.0	0.36	0.89	-7.2	0.70

From these statistics it is clear that the CFSR grid point time series are able to capture some of the main features of the station time series. The mean wind speeds are on average over estimated by 0.61m/s and the correlation of the wind speeds is fair at 0.70. The spread of the errors is fairly high though, at 2.0m/s, and the Distribution Error of 0.36 means that the 2D histograms of wind speed and direction are often poorly matched. The following section details the development and validation of a bias correction methodology to improve on the biases shown here.

3.4. Correction Methodology

The primary goal of this chapter is to produce complete, 30-year time series of wind speed and direction which can be used for analyses that require concurrent measurements and for the validation of climatological runs like those from the WASA project.

For these purposes the bias correction aimed to achieve the following:
1) Remove the bias in the mean wind speeds and directions from the CFSR time series
2) Correct the mean diurnal and seasonal cycles in the CFSR time series
3) Improve the correlation between the observed and the CFSR time series
4) Maintain the spatial relationships between stations points

A relatively common method for filling gaps in station records is to simply use the mean wind speed and direction (Hagemann, 2008). However, this study is concerned with the co-evolution of wind speeds and directions at different locations and not just with averages. Plugging gaps with an average, especially when there are several consecutive hours missing, would result in unrealistic breaks in the records. A similar problem applies to the use of modelled winds for 'hole-plugging'. For this reason, a method was sought that would give the most realistic and accurate time series as possible. The four methods described below are all based on the concept of taking high resolution reanalysis time series as a starting point and adapting these towards the observations. The CFSR hourly forecast product at 0.3° resolution model data was chosen as the basis for this method due to its high spatial and temporal resolution. However, as with any model, there are biases inherent in the data as we have seen in the previous section. Two methods are presented here, each take somewhat different approaches to removing, or at least, reducing the impact of these biases. Two different combinations of these approaches are also considered. The following section describes the four methods.

3.4.1. Quantile method

The quantile method assumes that the major biases and scale related discrepancies are predictably related to the wind speed and direction in line with method described by Badger et al. (2014). In other words, that for a given wind speed and direction, the difference between the model output and the station record would be fairly consistent. Given this assumption, it is then possible to calculate a bias correction for the CFSR time series for each wind speed-wind direction class. Removing these biases would also achieve a smoother time series than merely plugging the gaps in the record with the mean wind speeds and directions

Stepwise, the application of this method is as follows:

1) The hourly wind direction records for each station in the period 1994-2010 are divided into 20 equal sized bins corresponding to 5% of the record. These bins are then subdivided into 20 speed segments each comprising 5% of those records. This gives a total of 400 wind classes with equal probability of occurrence.

2) For each of the wind classes, the mean error of the wind speed (MEWS) and direction (MEWD) was calculated according to equations 5 and 6.

3) For each hour in the CFSR record, the CFSR wind speed and direction was corrected by multiplication of the MEWS and subtraction of the MEWD corresponding to the class assigned to that hour.

3.4.2. Wind Cycle method

The second method is based on the assumption that the biases in the CFSR time series are dependent on the time of day and the month of the year. For this method, each hour in

CFSR time series is classified according to the time of the day and the month of the year. With 24 hours in a day and 12 months of the year, this results in 288 time classes. For each of these classes, the bias in wind speed and direction is calculated and subtracted from the corresponding CFSR record. The two parts to this application are:

1) The MEWS and MEWD were determined for each calendar month, subdivided into hours of the day.

2) For each hour in the period 1994-2010, the MEWS and MEWD corresponding to that hour, in that month were subtracted from the CFSR wind speed and direction respectively, removing the seasonal and diurnal bias in the CFSR data for each station.

3.4.3. Integrated method

As will become evident in the discussion of these methods, each of these methods has strengths and weakness. In order to get the best of them both, a combination of the two underlying assumptions produces a third method. In this case the classes were defined both according to the time of day and season in the year, but also by the speed and direction. The major limitation with this third method is the ratio of number of classes to the number of observations. Having a full 288 time classes, each subdivided into 400 wind classes would have resulted in most of the classes having just one or two entries. In order to optimise the number of classes, several different resolutions were attempted. The optimal configuration was found to have four seasons, eight times during the day, three speed classes and eight direction classes, giving a total of 768 classes. For each of these classes, the biases were calculated and subtracted from the CFSR records.

3.4.4. Successive method

The second combined method simply applies the wind class method and then the wind cycle method with the second correction calculated after the application of the first. As the diurnal and seasonal cycles are not independent of the wind speed and direction, there is some overlap between these two methods. However, as the second correction is only calculated on the CFSR data once the wind class correction has been applied, only the residual bias is removed.

3.5. Results

Each of the methods described above were applied to each of the stations and the average results are presented in Table 4.

All of the methods, on average, improve all of the error statistics with the exception of the vector correlation. As expected, the wind cycle method dramatically improves the seasonal and diurnal cycles, however, the quantile method has greater success with respect to the mean wind speeds and the vector correlations. Of the combinatory methods, the successive method out performs the integrated approach with respect to the RMSE and CDE, while the integrated approach does better with respect to the ME, DE and VC. In Figure 22, the metrics for each station (before and after correction) are scattered to show the relative success of the correction methods. In the upper left panel, the improvements in VC can be seen for all stations above the unity line with the integrated method showing the best results. The wind cycle method and successive methods seem to reduce the vector correlation for most of the station points. The strength of the wind cycle method can be seen in the top middle panels, where the red markers are obscured beneath the light blue

Table 4. Mean Error Metrics for the CFSR bias correction

Correction	ME (m/s)	RMSE (m/s)	DE [0,1]	CSE (m/s)	CDE(^{o})	VC [0,1]
None	0.61	2.0	0.36	0.89	-7.2	0.70
Quantile	-0.21	1.4	0.31	1.0	-2.1	0.71
Wind cycle	0.10	1.5	0.25	0.98	0.37	0.60
Integrated	-0.070	1.4	0.19	1.0	3.8	0.73
Successive	-0.11	1.3	0.22	1.0	-0.53	0.65

markers of the combined method. Both of these approaches remove, almost completely, the bias in speed in the seasonal and diurnal cycles.

The lower left panel shows the effect of the corrections on the shape of the 2D histogram of speed and direction. In this plot values below the unity line indicate improvements (as opposed to the panel above for the correlations). The quantile approach is the only one that shows an increase in this error for most stations. The combinatory approaches are more effective in this regard with the integrated method showing the most marked improvement. In the lower middle and lower right panels we see the scatter of the more traditional RMSE and speed bias metrics. Here the wind cycle out-performed by the quantile approach, and the combinatory methods also achieve dramatic improvements. For stations where the bias in wind speed was very small to begin with, the corrections increase this bias to a small degree.

Although the successive method achieves similar results to the integrated approach, the integrated method was selected as it shows improvements in all the metrics, including the vector correlation, where the successive method fails (Table 4). It should be noted that this method requires the most classes and therefore has slightly higher data requirements than the other method. This will be discussed further when the results are presented.

3.5.1. Method Validation

Having selected the integrated method as the most viable correction to apply, this section details further validation of the method. Specifically, this section tests the assumption that the correction calculated on a given period is valid for another period, and to what degree the method affects the spatial relationships between station points. These two tests are important because the method will then be used to produce records extending further back

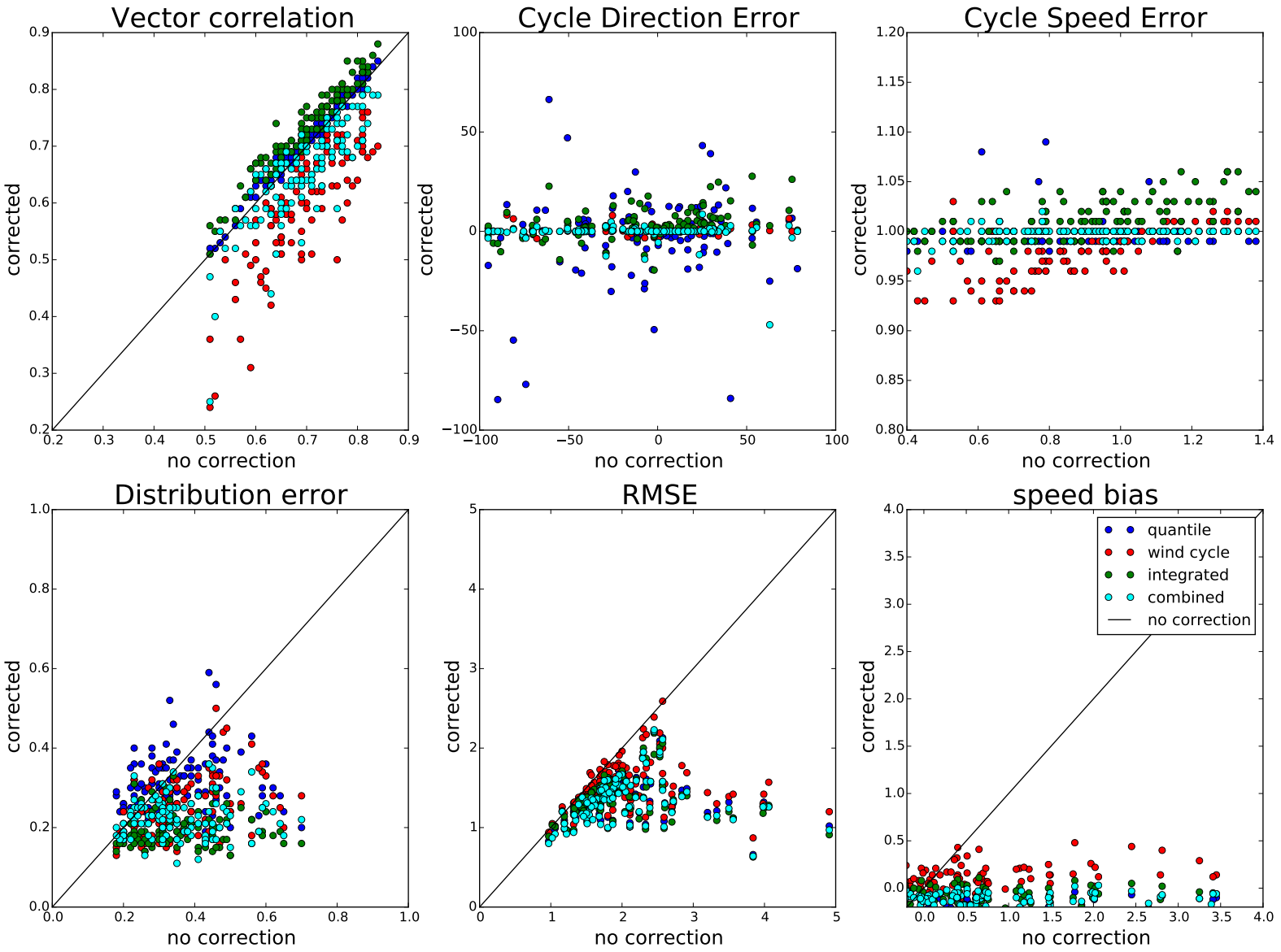

Fig. 22.— Comparison of error statistics with different correction methods. In each case the metrics from before and after the correction are scattered to show the effect of the correction.

than the observational periods and because the following chapter will consider the spatial relationships between different regions.

Themessl et al. (2011) highlights the need for validation of the bias correction on periods independent from the calibration period. The following is an example, in Figure 23 of a station (Alexander Baai) where the method was trained on the period 2004-2010 and then used to extend the record back to 1994. The results are then compared with the observations from this period.

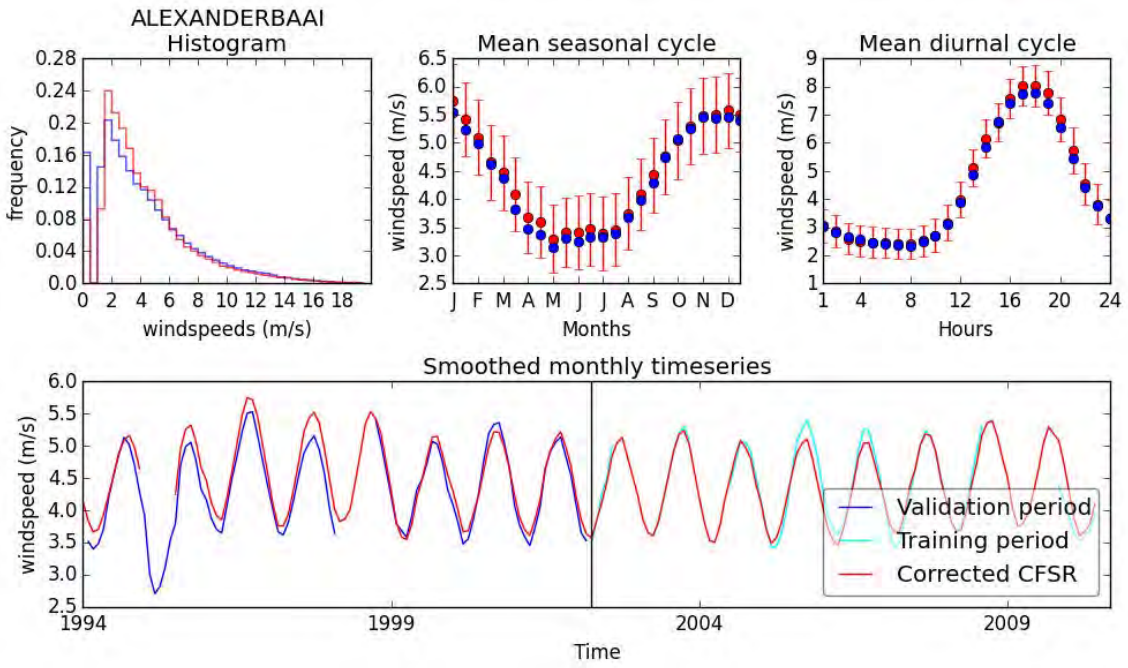

Fig. 23.— The plots in the top row show the histogram, seasonal and diurnal cycles and the lower plot the time series. In each case the blue indicates the observations, and the red the corrected CFSR data. The training period is shown in light blue in the time series. The gap in the corrected CFSR time series is due to a month of missing data in the CFSR data set.

Unfortunately, the calmest year in the record, 1995 is missing from the model data set, but there is very little separating the time series for the rest of the period. Due to the low

inter-annual variability and small trends at most of the stations, the correction appears to be robust with respect to the training period. Varying period length (not shown here) also had minimal impact on the method, at least up to the minimum of 3 years required for inclusion.

Figure 24, similar to Figure 24, shows the effect of the correction to the mean speeds, variability and trends at all the stations.

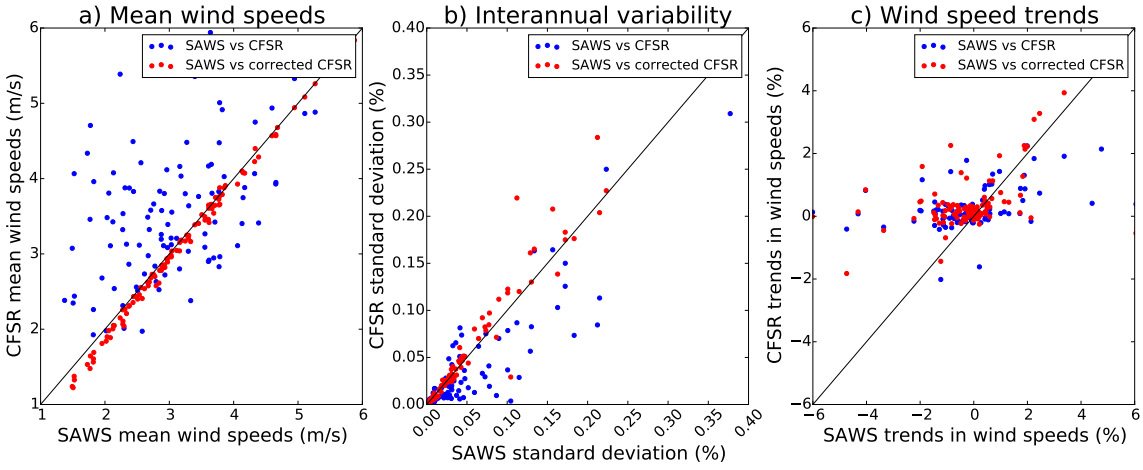

Fig. 24.— A comparison of the mean (a), standard deviation (b) and trends (c) in wind speed from the SAWS weather stations and the corresponding, corrected CFSR grid cells. The SAWS records cover a minimum of three years and the CFSR records for the same periods are used for comparison. Each point represents a single station.

There is a clear improvement, especially with respect to the mean speeds, although the correction results in a slight underestimate of the lower speed stations. The middle panel shows a similar improvement in standard deviation of wind speeds, however slightly over estimating these. There is no major change in the trends, which is not surprising, as the method only takes the average biases into account.

The final test for the method is whether it disrupts the spatial correlation relationships between the stations. In order to test this, the cross-correlation matrices of the stations,

corrected and uncorrected CFSR time series, and their combinations, are shown in Figure 25.

The two right hand panels show the corrected CFSR correlations and the reduction in spread is clearly evident. As the lower right panel shows, there is a bias towards higher-than observed inter-station correlations after the correction.

As an example of the final corrected time series, the following section shows the two stations examined earlier in this chapter (Koingnaas and De Aar), after the application of the correction.

3.5.2. Koingnaas

The following figures (26, 27, 28) show the dramatically improved agreement with the observations from the Koingnaas station. Both the mean seasonal and diurnal cycles (seen in the top middle and left panels of Figure 26) are barely distinguishable from the observed equivalents with respect to both speed and direction. The error bars on the seasonal and diurnal cycles show one standard deviation on the mean difference between the corrected CFSR data and the observed time series. Where the red dots cannot be seen, they lie completely under the blue markers. The corrected histogram (top left in Figure 26), while considerably improved, does not quite capture the shape of the observed distribution, over-estimating the frequency of speeds in the 4-6m/s range and underestimating the frequency of the 2-4m/s range as well as the highest wind range.

Figure 27 compares the corrected 2D histogram of wind speeds and directions with the observed distribution. Although there is a good fit to the general shape, the frequency of the highest winds speeds is still under estimated, particularly from the north easterly direction. Since the correction method starts from the CFSR record, if there are particular observed

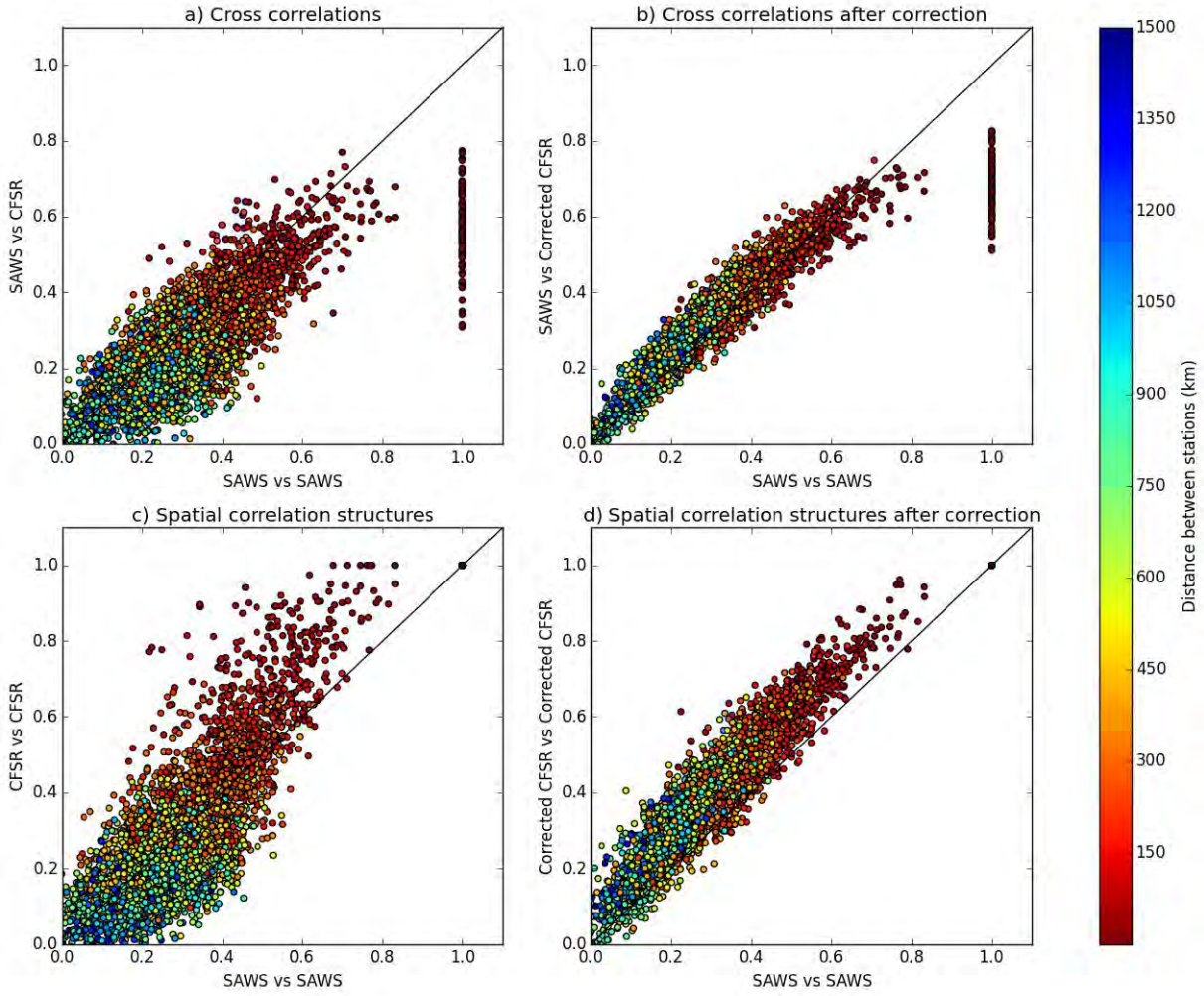

Fig. 25.— A comparison of the observed and corrected CFSR cross correlation patterns between 107 SAWS station sites. Each of the markers represents the correlation between two stations points with the colour indicating the distance between them. a) The correlation between the CFSR time series from the grid points corresponding to the station sites and the observed time series. b) As for a), but after the application of the correction to CFSR time series. c) A comparison of the CFSR spatial correlation structure with the observed spatial correlation structure. d) As for c, but after the application of the correction to the CFSR time series.

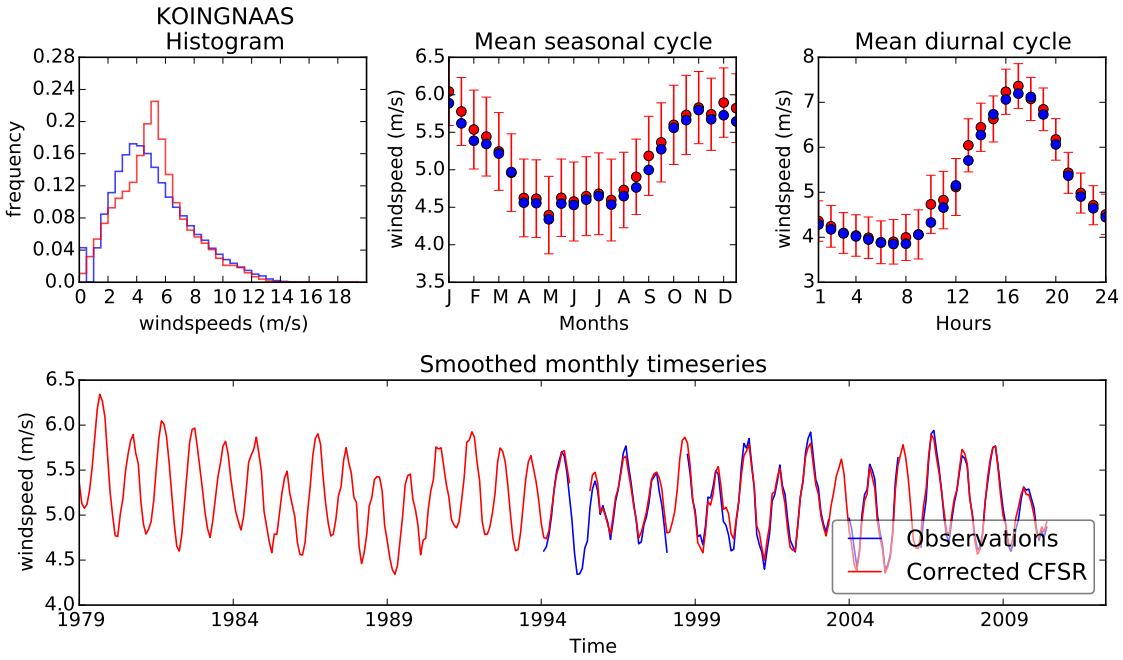

Fig. 26.— Summary of the corrected CFSR data for the Koingnaas station. In the top panel the histogram of winds speeds, seasonal cycle and diurnal cycle are show, and in the lower panel, the mean monthly time series of wind speeds. In each case, the red shows the bias-corrected data from the closest CFSR grid cell.

conditions which are not present in the model data set, the correction method is unable to change the model time series sufficiently to reproduce these especially if there is not a consistent bias. This is one limitation of the correction method that should be kept in mind.

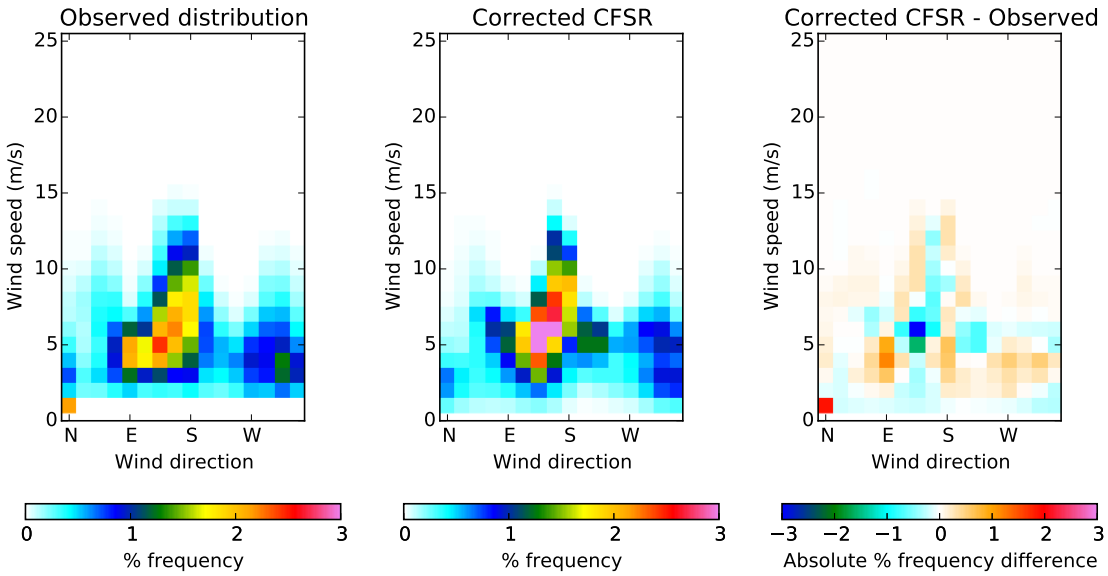

Fig. 27.— 2D histograms of the observed and corrected CFSR wind speeds and directions for Koingnaas from 1994-2010. The bottom left bin represents the frequency of calm hours. Following the correction, the shape of the distribution is improved and the magnitude of the errors are reduced.

The success of the correction is perhaps best seen in Figure 28 where the very close match of the timing of seasonal and diurnal shifts in wind patterns is evident.

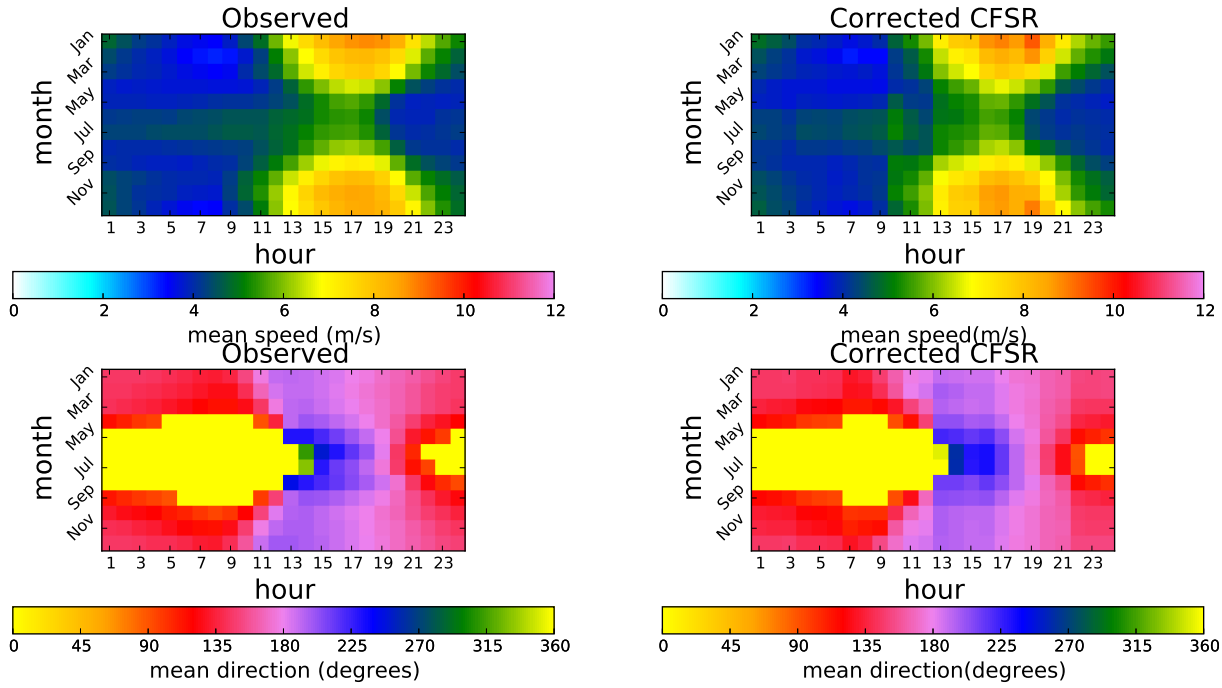

Fig. 28.— Mean observed and corrected CFSR mean wind speeds and direction broken down by hour of the day and month of the year for Koingnaas.

3.5.3. De Aar

At the De Aar station, the corrected time series shows a similar story in the Figures 29-31. Although the two time series are not identical, the mean seasonal and diurnal cycles are almost inseparable and the range of errors is on the order of 1m/s. From the histogram we can see that there is still some difference in the shape of the distributions but the distribution error (as defined in Equation 7) is down from 0.24 to 0.18.

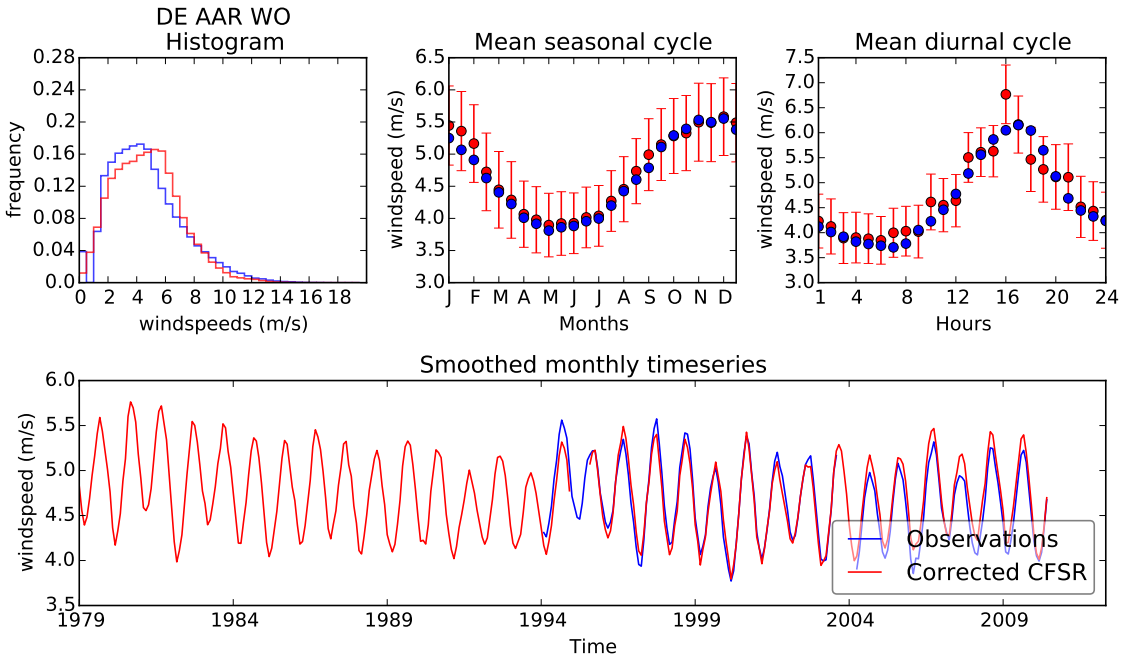

Fig. 29.— Summary of the corrected CFSR data for the De Aar station. In the top panel the histogram of winds speeds, seasonal cycle and diurnal cycle are show, and in the lower panel, the mean monthly time series of wind speeds. In each case, the red shows the bias-corrected data from the closest CFSR grid cell.

Figure 30 shows the improvements in the shape of the 2D histogram, with the average error in bin frequency of less than 1%. The biggest improvement in the shape of the distribution is in the frequency of the high wind speeds which were initially underestimated by

the CFSR grid cell. This is an improvement seen at most of the stations.

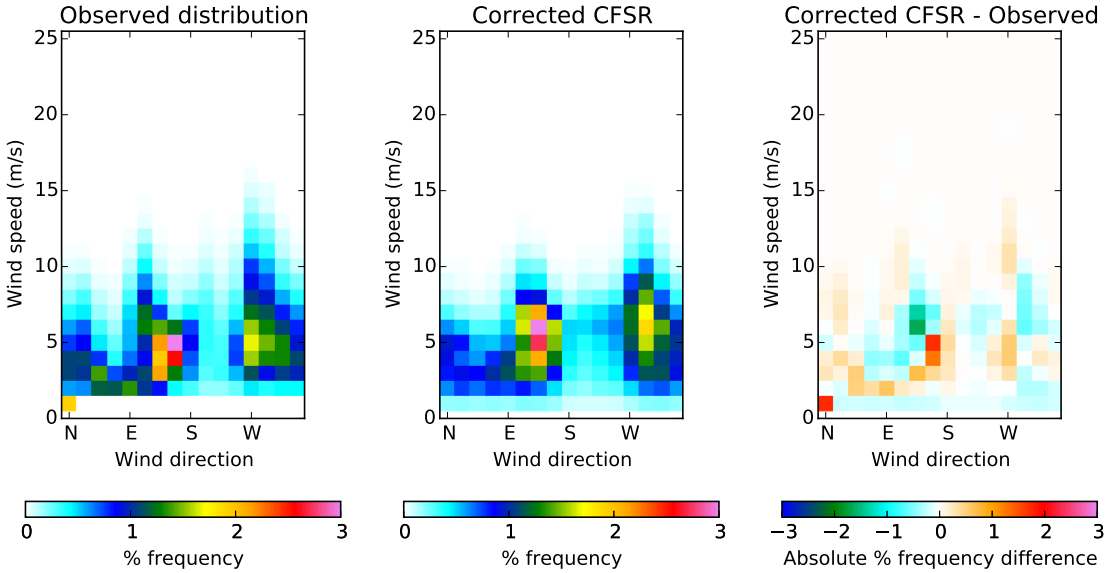

Fig. 30.— 2D histograms of the observed and corrected CFSR wind speeds and directions for De Aar.

In terms of the timing of seasonal shifts, there is a fairly major correction in the wind directions during the winter. In Figure 21, earlier in the chapter, we can see that the CFSR grid cell has predominantly northerly winds in the mornings as opposed to the south easterly winds seen in the station record. In Figure 31 however, the corrected directions match the observations throughout the year. The cycle error (defined in equation 9) is reduced from a mean of -46° to less than 1°.

3.6. Conclusions

In order to create a 30-year time continuous record at each of the SAWS stations, a set of techniques were tested that aim to remove consistent biases from the CFSR grid point records corresponding to the weather station sites. An integration of two approaches,

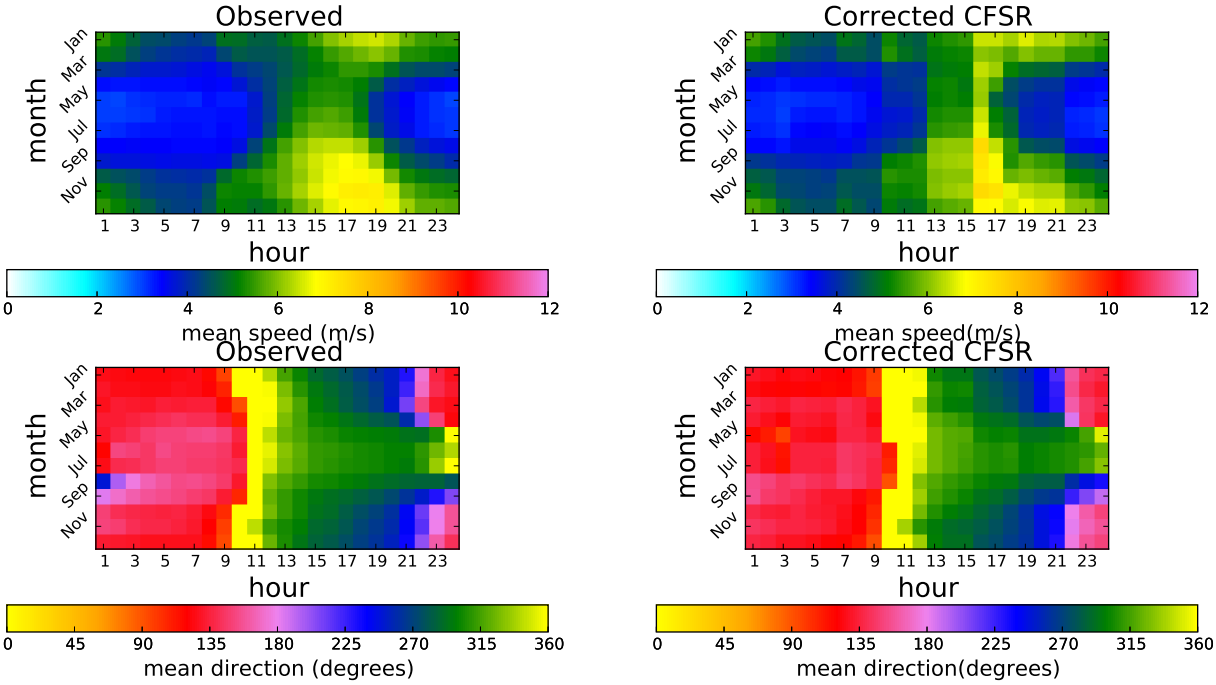

Fig. 31.— Mean observed and corrected CFSR wind speeds and direction for each hour in each month for De Aar

one quantile-based, and the other based on the seasonal and diurnal biases, was shown to consistently reduce both the absolute bias, and spread of errors in wind speed and direction as well as the seasonal and diurnal cycles, however was unable to improve on the correlation between station and CFSR time series. An approach that removes the bias as a function of hour of the day, month as well as wind speed and direction was found to be more effective and was further validated according to the recommendation of Themessl et al. (2011) on periods independent from the calibration period, at two stations with suitably long records. Where existing methods have failed to improve on correlation, and have been shown to drastically alter the spatial relationships between different regions (Themessl et al., 2011), this method not only improves the correlation with observations, but leave the cross-correlations between station points intact.

A caveat for the use of this of the method is the fact that the wind classes are based on the model winds, which may not map cleanly to observed winds. So if there is little overlap in classes, or a single model wind class corresponds to multiple observed wind classes (in the case of long periods of erroneous calm winds) or a weak relationship between the model and station winds to begin with, the correction will have limited success in producing meaningful improvements to the model data. This is because the method relies on the assumption that there is a meaningful relationship between the model and observed winds.

The application of this method to the CFSR time series corresponding to the station records selected in the previous chapter, produces 107, 30-year hourly time series, which are representations of the observations to allow for comparative time series analysis. This data is the basis for the wind regionalisation in the following chapter.

4. South African wind regimes and regions

4.1. Introduction

The previous three chapters prepared the data required for the analysis of the wind climate of South Africa from 1979-2010. Chapter 2 documented the selection of representative weather station records and the application of a stringent quality control of these data. In Chapter 3, a method was developed and applied that corrects the CFSR hourly wind records from the grid points nearest to each of the SAWS stations, towards the observations, and extended them back in time. This chapter makes use of the resulting data set to define anemological wind regions based on how the station wind speeds are controlled by the synoptic scale circulation patterns. The chapter begins with a review of the literature on climate regionalisation and the different clustering methods used, with a focus on the South African context. This is followed by an outline of, and motivation for, the methods in this chapter. A self-organising map is used to dis-aggregate the synoptic circulation patterns into archetypal states and the station records are then correlated with corresponding 850hPa winds for each circulation pattern. Thereafter, an hierarchical clustering algorithm is applied to the correlations to separate the stations into groups representative of the local regions.

4.2. Climate classification and regionalisation in South Africa

Classification of climatic regions is not a new concept, and since Thornthwaite (1948) published 'Towards a rational climate classification', there have been numerous climate classifications for different applications in areas all over the world. Some of the first classifications of South African climate were done by Jackson in 1951 (Jackson, 1951) with the country's first climate atlas dividing the country into 8 climatic regions. This was followed in 1974, by Preston-Whyte's 'Classifying the climate of South Africa: a multi-variate ap-

proach' (Preston-Whyte, 1974) in which 24 variables, from 73 weather stations, were used to define 8 (and 19) regions and this was taken further by Johnston in 1980 with 'Multi-variate climate classification of Southern Africa' (Johnston, 1980).

Since then, climate regionalisations have been by made by Kruger (2002), who described the climatic characteristics of 24 vegetation regions across the country (Figure 32) , and Conradie (2012), who used the Köppen-Geiger classification (Figure 33). Both of these methods use vegetation to define the climate regions as opposed to the actual climate variables.

The majority of climate studies in South Africa have focused, primarily, on rainfall and temperature (Bunkers and Miller, 1996; Harrison, 1984; Landman and Simon, 1999; Taljaard and Phil, 1996; Mason, 1998; Engelbrecht and Engelbrecht, 2015). Mimmack et al. (2000) for example, as part of an examination of various metrics for climate classification, clustered rainfall stations in South Africa by principle component scores. In a similar study from Switzerland, Weber and Kaufmann (1995) developed an 'automated classification scheme for wind fields' based on the spatial similarity of normalised wind fields, clustered with the complete linkage hierarchical method. They were able to identify 12 patterns in the Basel area in the context of high channelling.

Other observational wind studies from around the world demonstrate the utility of regionalising wind observations. For example, a regionalisation of surface winds was conducted by Jimenez et al. (2009) in an area of Spain, using weather station records to characterise the typical wind regimes in each region and link these to the large-scale processes. This study was used as the basis for the validation of the WRF model simulation of the same region (Jimenez et al., 2010a). In another study, Cassola et al. (2008) divided Corsica into anemological regions in order to determine the optimal distribution of wind farms so as to minimise the variability of power into the electrical grid (Burlando, 2009). Other examples include wind regime identification in the Grand Canyon Region (Kaufmann and Whiteman,

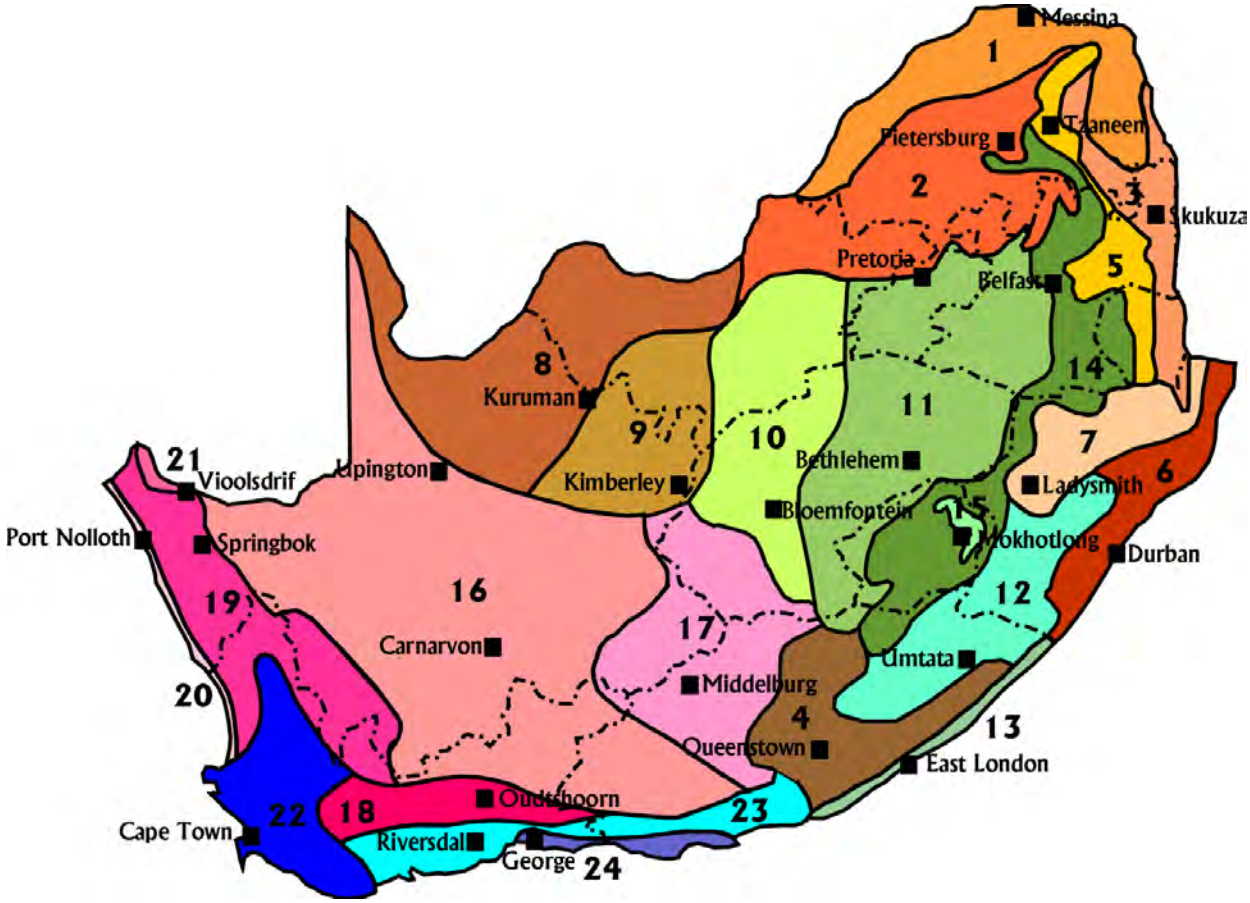

Fig. 32.— Vegetation based climatic regions (Kruger, 2002). Key: 1. Northern Arid Bushveld 2. Central Bushveld 3. Lowveld Bushveld 4. South-Eastern Thornveld 5. Lowveld Mountain Bushveld 6. Eastern Coastal Bushveld 7. KwaZulu-Natal Central Bushveld 8. Kalahari Bushveld 9. Kalaharu Hardveld Bushveld 10. Dry Highveld Grassland 11. Moist Highveld Grassland 12. Eastern Grassland 13. South-Eastern Coast Grassland 14. Eastern Mountain Grassland 15. Alpine Heathland 16. Great and Upper Karoo 17. Eastern Karoo 18. Little Karoo 19. Western Karoo 20. West Coast 21. North-Western Desert 22. South-Western Cape 23. Southern Cape 24. Southern Cape Forest.

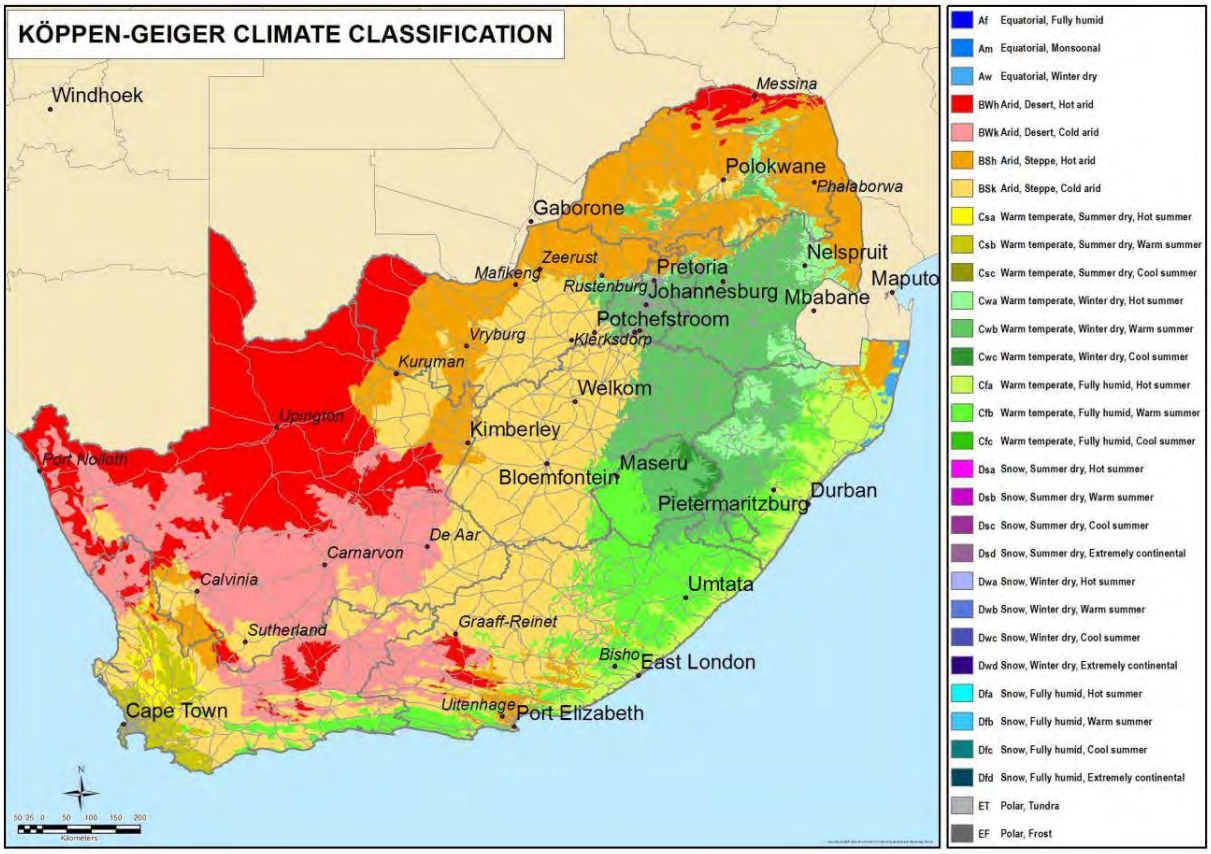

Fig. 33.— CSIR Köppen-Geiger map based on 1985 to 2005 South African Weather Services data on a 1 km x 1 km grid. (Conradie, 2012)

1999) and regionalization in complex terrain (Jimenez et al., 2008). In South Africa, however, there has been very little work done in terms of climate classification with respect to wind since Diab's atlas in 1995 (Diab, 1995) (see Chapter 1, Figure 1). The only rigorous wind regionalisation has been focused on extreme winds within the Wind Atlas for South Africa (WASA) project [v]. High wind events are important from a building-safety perspective (Kruger et al., 2010) but also from a wind energy perspective, and the map shown in Figure 34, produced by Kruger et al. (2011), shows the extreme wind zones identified. Although the processes that lead to extreme winds are different to those that determine vegetation regions and climatic wind zones, these examples provide an interesting comparison to the regionalisation developed in this study.

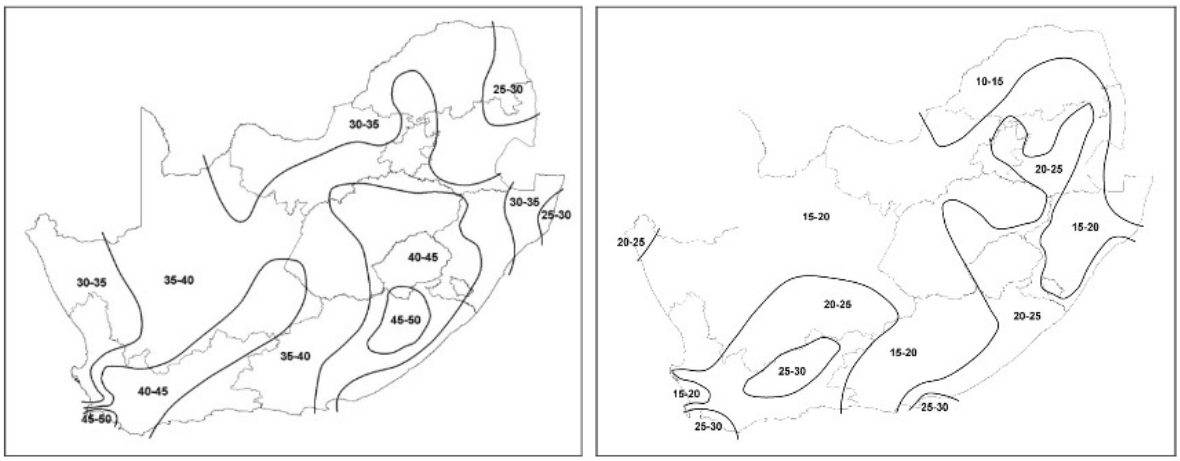

Fig. 34.— Extreme wind regionalisation based on 50 year quantiles for Gust (left, in m/s) and mean hourly wind speed (m/s) (right).

One factor that is stressed by both Bunkers and Miller (1996) and Jimenez et al. (2008), is the importance of the coupling between the synoptic and station scales for wind climate

[v]More information about the WASA project can be found at *http* : *//www.wasaproject.info/*.

classification. In South Africa, the coupling between the wind station data to upper level flow has not been quantified. For example, its it not know to what extent a ridging high-pressure system to the south-west of the country would dominate the wind flow along the west and south-west coast line of the Western Cape or whether a stagnating high-pressure system over the interior in winter would facilitate smaller-scale processes. The method developed here relates the large-scale synoptic state, to the surface wind response, to define wind regions in South Africa. In the following section, the techniques used to achieve this are described.

4.3. Method

The regionalisation method presented here can be broken into three stages. In the first stage, a self-organising map is trained to define the archetypal synoptic circulation states. The second involves correlating the station wind speeds with the 850hPa wind speeds from the closest CFSR grid cell under each of these states. In the final stage, hierarchical clustering is applied to these correlations to produce groups of stations whose wind speed records are similarly dependent on the synoptic states.

4.3.1. Self-organising map

Developed by Kohonen (1990), self-organising maps (SOMs) have been used extensively in climatology, and other fields, to find patterns in complex data sets Hewitson and Crane (2002, 2006); Hope et al. (2006); Lennard and Hegerl (2014). In brief, the method iteratively adjusts the initially random position of a given number of node points in the data space in order to maximise the fit to the data. These node points can be considered to be archetypal representations of the data.

The key parameters in the process are α, which determines how much the nodes are

moved in each iteration (and which decays over time) r, the number of surrounding nodes that are affected, n, the number of nodes, and i, the number of iterations. In the first stage of the method, the node points are distributed (usually at random) in the data space. In the second phase the data points are iteratively compared with the node points. In each case the best matching node point is then moved towards the data point in question, 'dragging' the web of node points with it. It is common practice to apply this training twice, initially with a 'flexible' web of nodes, that is moved quickly and coarsely towards the data, and then a finer training with smaller steps refines the map. For this analysis, the SOMPAK package [vi] was used, as it allows for multi-stage training and full customisability.

The training data used for the SOM consists of hourly 850hPa geopotential heights from the CFSR domain (CFSR, 2010) defined in Chapter 3 for the period 1979-2010. The grid resolution is 0.5 degrees and there are 51 longitude points and 41 latitude points giving a total of just under 550 000 000 data points. Geopotential height was chosen, as it effectively describes the synoptic environment of the surface wind response.

As the purpose of this analysis is to reduce the data set to a small number of representative states, having a large SOM is counter-productive. A fairly small SOM, with 12 nodes (in a 3x4 grid) was selected for this reason and because it falls within the range of climate states found by Preston-Whyte (1974), Johnston (1980), Kruger (2002) and Conradie (2012). Several other sizes (3x5, 4x5 and 5x6) were found to add little extra information and so the smaller size was selected for simplicity. In order to optimise the SOM configuration for this data set, a series of tests was applied to one year of the data. The average distance from

[vi]SOMPAK is freely available at

$$http://www.cis.hut.fi/research/som\_lvq\_pak.shtml.$$

the each data point to the corresponding node point was used as a metric to determine the best combination of settings. The results show that after an initial, coarse training, with a large training radius, the SOM is relatively insensitive to α, r and ns. In order to cope with the large size of the input data, the SOM was trained sequentially with data from each of the years between 1979 and 2010. For the first year, an α of 0.1 and an r of 3 were used for 50000 iterations. For all the following years, α was reduced to 0.05 and r to 1.

In the process of testing different parameter combinations it was noted that when trained on just a year of data, the distribution of data points with respect to each node was in some cases bi-modal. This suggests that in some cases, two distinct states are being combined into a single node and it would make sense to either increase the number of nodes or, as Huva et al. (2014) suggest, exclude the cluster of data points furthest from the centroids in the bimodal nodes as they reduce the representativity of the nodes. However, when run with the full 30 years of data, the data points become more evenly distributed around the centroids with no clear cut-off point. As a further check, the geopotential height fields corresponding to the furthest points from each centroid were compared to the mean fields for each node (not shown). The major difference in each of the cases was the magnitudes of the gradients, with the same general pattern clear. With the vast number of data points, removing these 'outliers' would likely have a negligible effect on the centroid positions and so all data points were retained.

4.3.2. Clustering

In line with the primary aim of this thesis, which is to develop a wind regionalisation based on the synoptic environment, the coupling between the surface wind speed and 850hPa wind speed was chosen as the basis for the clustering. There are many other possible ways to cluster the stations, and basing the regions based on the mean wind speed, diurnal or

seasonal cycles, or mean response to the circulation types, while potentially interesting, are beyond the scope of this thesis and are topics for future work.

In order to quantify this coupling, the wind components at the 850hPa level (where the synoptic states were defined) were extracted from the CFSR data set, for the closest grid point to each of the stations. The hourly wind speeds from each station time series (and its closest CFSR 850hPa gridpoint) were assigned to the corresponding synoptic states produced by the SOM. The Pearson's correlation coefficient was then calculated for each station point, under each synoptic condition. A high correlation indicates that the 850hPa winds exert a strong influence on the surface flow, whereas a low value indicates that the surface wind speed is being determined largely by local processes and is relatively independent of the synoptic scale circulation.

Ward's method of hierarchical clustering was selected for its tendency to produce relatively cohesive clusters (Burlando et al., 2008) and the correlations, as a function of synoptic state, were used as the input for the clustering. The final clusters were then determined by examination of the dendrogram and inferences based on the topographical features and knowledge of the physical processes. In the following section, the results of each of the stages in the process of producing these regions are described.

4.4. Results

4.4.1. Synoptic states

The centroids of each of the clusters produced by the SOM represent the mean atmospheric state for the hours assigned to the node. The mean geopotential height (GPH) and mean sea level pressure (MSLP) fields associated with each of these nodes can be seen in Figures 35 and 36 and the seasonal distribution of the hours assigned to the nodes in Figure

37. Before considering how the measured wind speeds differ under the synoptic conditions shown here, a short description of the states is in order.

The nodes on the left side of the SOM (1, 2, 5, 6, 9 and 10) show different stages in the progression of frontal low pressure systems moving across the south of the country towards the east and the South Atlantic High moving in across the West Coast region. While these synoptic features are more common and more intense during the winter months, they do occur all year round. Nodes 9 and 10 are predominantly seen in the summer months as the continental low connects through to the low pressure system to the south of the country. Nodes 3, 4, 7 and 8, in the top right hand corner of the SOM are dominated by a thermal trough on the west coast and the Indian Ocean High over the eastern parts of the country, and are almost exclusively winter patterns. Nodes 11 and 12 are most frequent during the shoulder seasons with both the continental low pressure (seen prominently in the MSLP fields) in the northern parts of the country, and ridging high pressures to the south.

These distributions are consistent with those found by Hewitson and Crane (2002); Lennard and Hegerl (2014); Pinto (2015). One major differences between these, and SOM patterns seen in the literature is the majority make use of daily, as opposed to the hourly data used for this project. The plots in Figure 38 show how the hours of the day were mapped to each of the SOM nodes. There is clearly a strong diurnal signal in the synoptic patterns. Each of the synoptic states identified by the SOM has a particular signature of hours with which it is associated, and a period, in most cases about 6 hours, before the atmosphere transitions to another state. A similar result was found for the wind field classification over Europe by Kaufmann and Weber (1996). This may have implications for the optimal temporal resolution to be used for SOMs to avoid aliasing and missing short-lived mesoscale patterns.

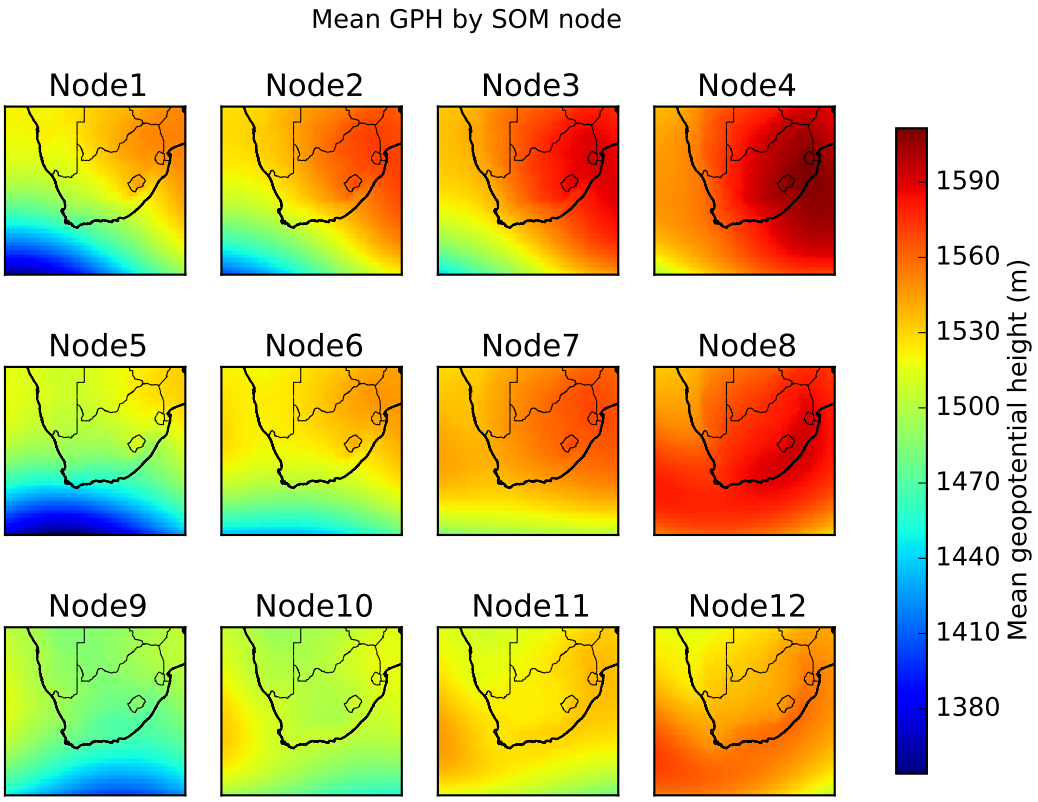

Fig. 35.— Mean 850hPa geopotential height fields for each of the SOM nodes from the hourly CFSR time series at 0.5 ° resolution. Each node represenents an archetypal atmospheric state.

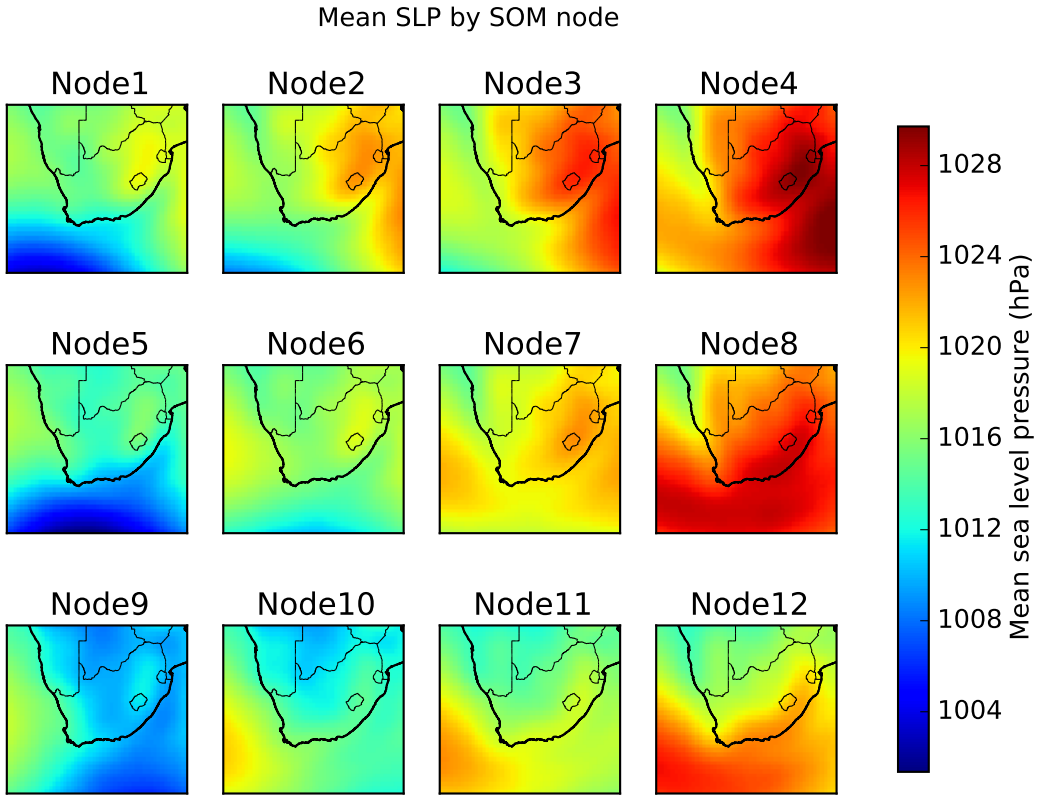

Fig. 36.— Mean sea level pressure fields for each of the SOM nodes from the hourly CFSR time series at 0.5 ° resolution. While the patterns are similar to those in the GPH fields in Figure 35, some shallow features can be seen, such as the trough on the west coast in nodes 4 and 8.

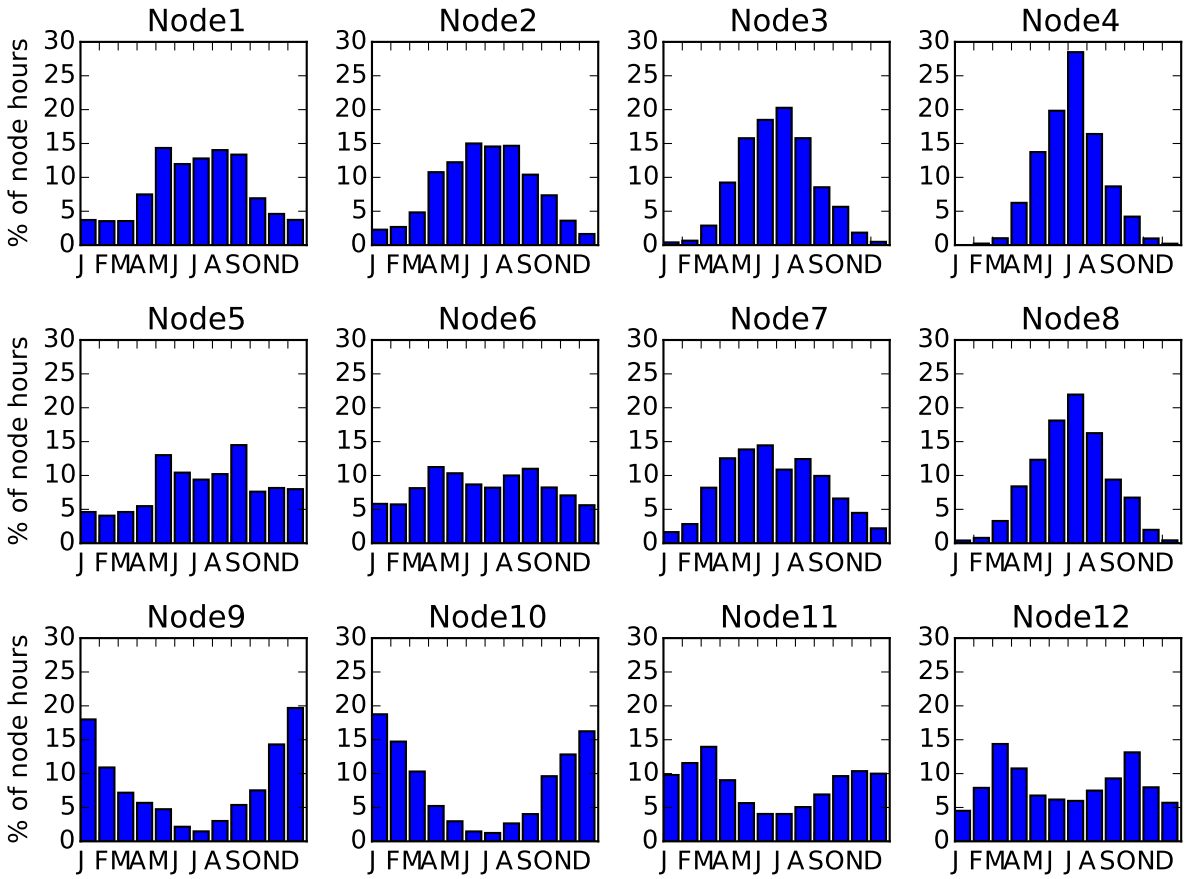

Fig. 37.— Seasonal distribution of SOM node frequencies. Winter frequencies increase towards the top right of the SOM and summer frequencies to the bottom left.

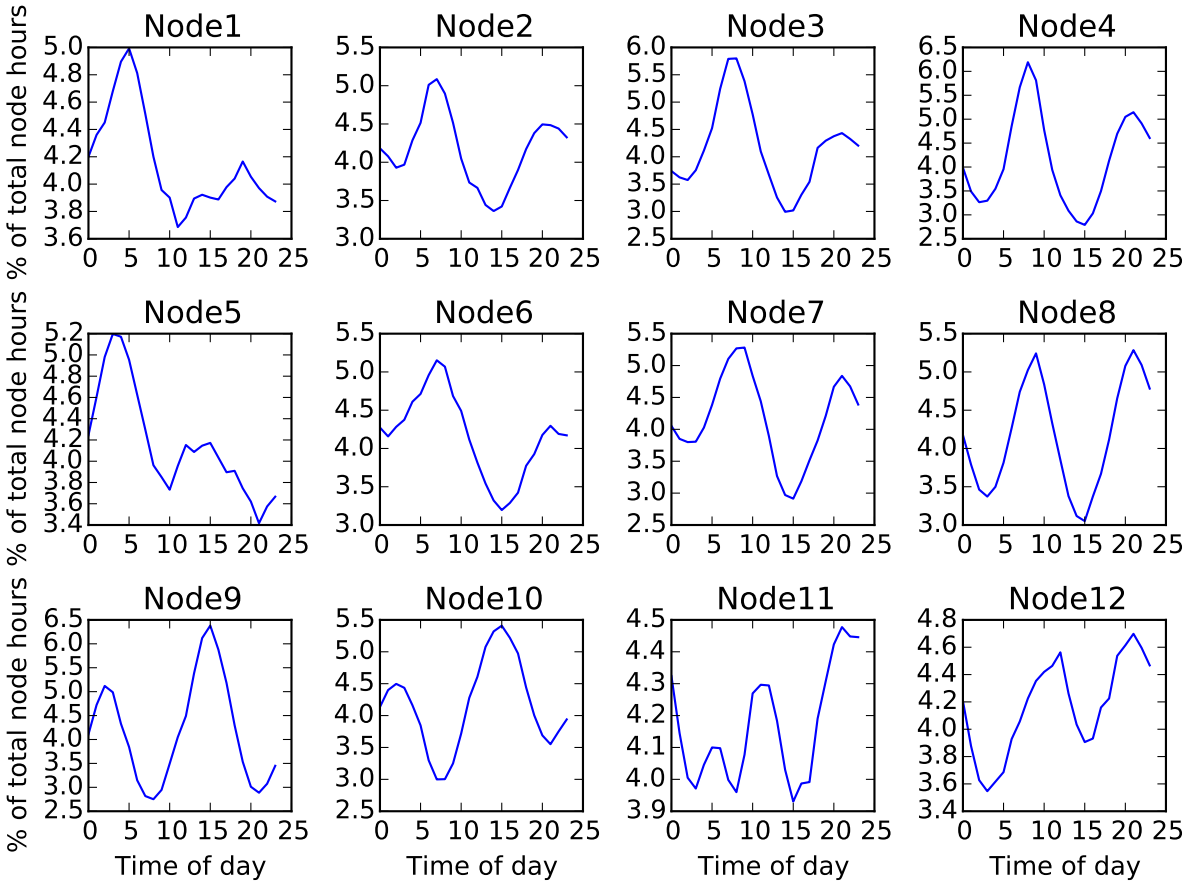

Fig. 38.— Diurnal break down of SOM node frequencies. The proportion of morning hours increases towards the top left of the SOM, with afternoon hours more frequently mapping to nodes in the bottom row, and night time hours to the nodes on the right./note[BA]Normalise

4.4.2. Clusters

The dendrogram for the Ward's Method clustering can be seen in Figure 39. The method initially splits the stations into two groups, with the majority of the inland stations in the first and the coastal stations in the second. The inland group is further divided into a strongly clustered (and spatially cohesive) group in the north-east of the country, and a weakly clustered group with two smaller subdivisions. The coastal group is significantly larger, and four distinct groups stand out at the same level as the two-way split in the inland group. These six clusters, found by 'cutting' the dendrogram tree at 2, will be discussed individually and in some cases divided further.

Cluster 1 consists of 12 stations (denoted in pink in Figure 40) 8 of which are tightly grouped in an area centred on Gauteng, and the remaining 4 in the Freestate Province. The stations in this group are all well correlated with upper level winds across all of the described synoptic states, with mean coefficients in the range of 0.74-0.8. The stations in the Freestate show similar correlations to the those further north, with the exception of the first 3 synoptic states, for which they appear to be less coupled to the upper level flow. The strong coupling to the 850hPa winds in this region are most likely due to the flat terrain, and the lack of strong local processes that would influence the surface wind.

The 27 stations in cluster 2 (in red) are spread across most of the country with the exception of the coastal areas. The cluster is made up of two smaller groups with a roughly north-south separation. These stations have moderate correlations with most of the synoptic states and the mean values are in the range 0.59-0.68.

Of the 14 stations in cluster 3, 8 (in dark blue) are situated right on the edge of the south coast, clearly separated from those just a bit further inland. From west to east, the correlations with the upper level winds under the conditions associated with first 2 synoptic states reduce from 0.5 at Struisbaai, to 0.03 at Port Elizabeth. Higher correlations with

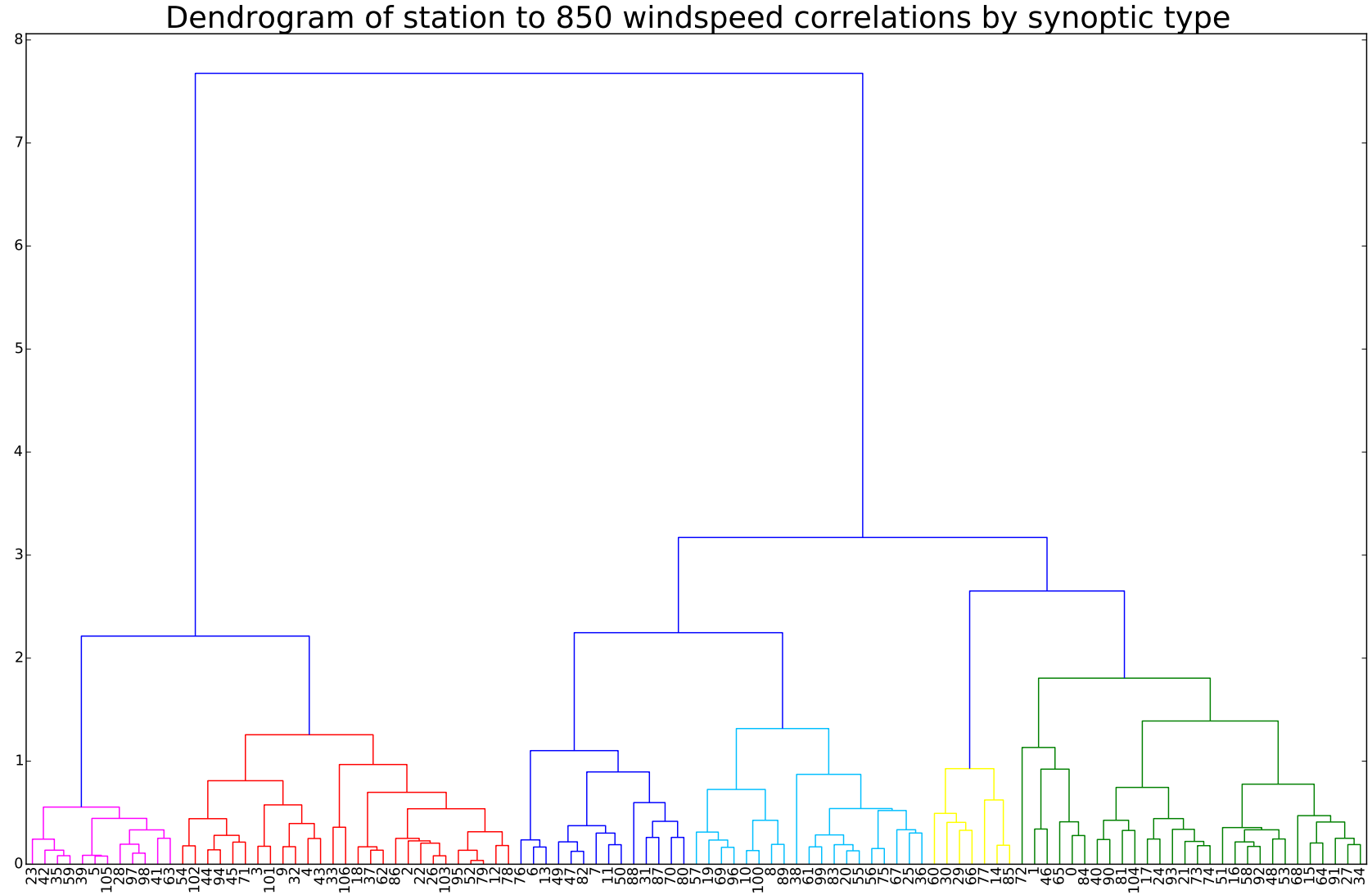

Fig. 39.— Dendrogram showing the clusters produced by Ward's Method. The colours indicate the six clusters that result from 'cutting' the tree at 2, are numbered from left to right, and correspond (with minor changes), to those in Figure 40. The number refer to the alphabetically-ordered station names.

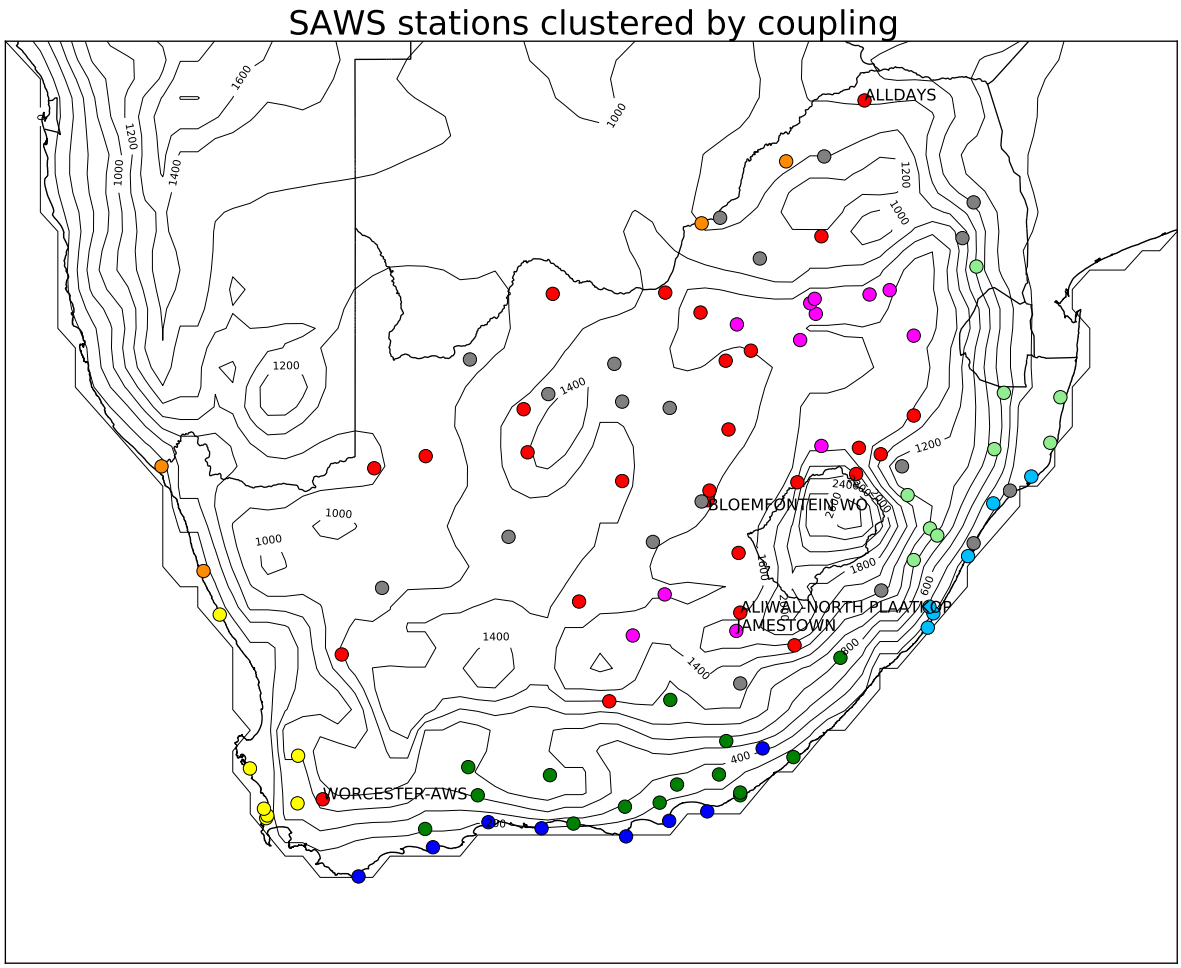

Fig. 40.— Stations clustered by their correlation with 850hPa winds. Pink denotes cluster 1, red denotes cluster 2, dark blue denotes cluster 3, light blue denotes cluster 4, yellow denotes cluster 5, dark green denotes cluster 6a, light green denotes cluster 6b. Grey denotes general outliers, orange denotes the outliers from cluster 6. The contours indicate surface height and a few stations are labeled for reference.

upper level winds under node 9 are also a feature of these stations. The remaining 6 stations (in grey, with the other outliers) are spread out across the country and do not appear to have any physical link to this region, with the possible exception of Queenstown which is just further inland.

The 19 stations that make up cluster 4 are all in the north-eastern part of the country, and are split into coastal and inland groups. The coastal group (light blue) runs from Port Edward up to Richards Bay and as with cluster 3, and are neatly separated from those slightly further inland. These stations show low correlations (0.3-0.51) with all of the synoptic states and are clearly governed by local dynamics. The inland group however is very dispersed, with no immediately obvious reason for their separation from the surrounding stations. It seems likely that these two groups are only clustered together because of a shared disconnect from the upper levels. The inland group is coloured grey as an 'outlier' group.

The cluster 5 is the smallest and most clearly defined (in yellow), encompassing a small region of the Western Cape running from Cape Town, slightly inland, and up the west coast to Garies Groen-Rivier. Topographical features neatly separate this region from the inland station clusters and are perhaps responsible for the low correlation with most of the synoptic states. The exception is node 5, with a 0.72 coefficent, that suggest the strong synoptic forcing of a mid-latitude cyclone.

Cluster 6 is the largest grouping, and contains within it two clear subgroups and several stations that are weakly grouped (indicated by the long connections on the dendrogram). The majority of the stations are along the south and east coast of the country, slightly inland of the very coastal stations. These can be neatly divided north-south into those stations inland of the cluster 3 stations (dark green), and those inland of the cluster 4 coastal stations (light green). The 'outlier' stations include a pair of stations on the northern extreme of the west coast, and a pair on the northern border of the country (in orange). This cluster has the

lowest correlations with the 850hPa winds (0.2-0.4) and are strongly influenced by the local topography inland of the coast.

From the mean correlations in Table 5, we can see the the different influence that each synoptic state has in the different regions. Clusters 3 and 4, the stations stations along the south-east coast, for example (in dark and light blue) respond quite differently to the various synoptic states. Stations in cluster 3 are fairly well correlated with 850hPa winds during hours associated with node 9 (0.66), while cluster 4 stations are not (0.31). During periods characterised by node 7, with a shallow trough effecting the west coast, cluster 5 stations (in yellow) are completely decoupled from the 850hPa winds (0.02) while stations in cluster 1, more effected by the Indian Ocean high, have mean correlations of 0.75. In the following section some of the less intuitive results are explored in more detail.

4.5. Discussion

Synoptic scale circulation patterns are clearly a driving force on the surface flows, and their influence can be seen in the regions they affect. However, as these results reflect, in certain regions, and in certain conditions, surface observations are poorly correlated with the 850hPa winds and as a result, some of the station clusters are not spatially cohesive. This section explores some of the factors that affect the coupling between the surface and upper level flows and how these impact the clustering.

The stations in cluster 1 are the most highly correlated with the 850hPa wind speeds. This is most likely due to the fact that these stations (in pink) are some of the highest above sea level, and therefore the closest to the 850hPa level. This region also lacks local-scale forcings like topography that would give rise to a decoupling. By contrast, the coastal stations have much lower correlations as they are further from the 850hPa level and affected

Table 5. Mean correlations between corrected CFSR 10m wind speeds and 850hPa CFSR windspeeds for 12 nodes found from self-organsing maps and 6 clusters of meteorological stations in South Africa for the years 1979-2010

Synoptic state	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Node 1	0.76	0.62	0.32	0.33	0.65	0.20
Node 2	0.74	0.59	0.25	0.39	0.43	0.17
Node 3	0.75	0.59	0.26	0.42	0.28	0.21
Node 4	0.75	0.57	0.33	0.45	0.11	0.28
Node 5	0.80	0.68	0.53	0.30	0.72	0.29
Node 6	0.76	0.63	0.48	0.34	0.36	0.21
Node 7	0.75	0.59	0.40	0.42	0.02	0.22
Node 8	0.79	0.60	0.51	0.55	0.25	0.40
Node 9	0.80	0.67	0.66	0.31	0.31	0.28
Node 10	0.76	0.62	0.59	0.38	0.17	0.31
Node 11	0.76	0.59	0.54	0.43	0.17	0.31
Node 12	0.78	0.59	0.55	0.51	0.26	0.38

by thermal gradients across the coastline.

The second cluster (in red) is perhaps the least intuitive as it is so dispersed. Many of the stations in this cluster are at a similar altitude to the cluster 1 stations, but have lower correlations, due to other local scale forcings. Consider the two stations in Bloemfontein for example. These stations are just 10km apart, and at roughly the same altitude, and yet the airport station is clustered into group 1 and the station in the town centre is in group 2. There is no major topography in the area so the lower correlations at the city centre are likely due to the sheltering effects of the urban environment.

Conversely, consider two stations, both in the in the same cluster, but in different regions. Worcester (the south-western-most station in red) and Alldays (the northern-most station in red). These two stations are over 1000km apart and are close to stations from other regions and yet are both in cluster 2. Both of these stations are affected by topography. Alldays is in a small valley, and the wind rose (in Figure 41) shows the channelling affect this has on the wind measured at the station. Similarly, the wind at Worcester is channelled by the mountain range just to the north, modulating the effect of the synoptic flows.

Another factor that affects the coupling between the surface and upper level winds is thermally driven surface flows. The heating and cooling of the land surface during the day drives flows that are likely the cause of the diurnal signal we see in the synoptic states. Kaufmann and Weber (1996) found that these flows were partly responsible for the decoupling of the upper level winds from surface winds in Europe, particularly in low lying areas. These thermally driven winds may be responsible for some of the spatial patterns in the clustering of the stations. The stations in cluster 6 (light and dark green) for example, are spatially distinct from both the coastal band (in blue), and the interior stations. Although the distances between the regions is small, there are different thermal flows affecting the coupling in each case. The coastal stations (in blue) experience sea breezes, which typically penetrate

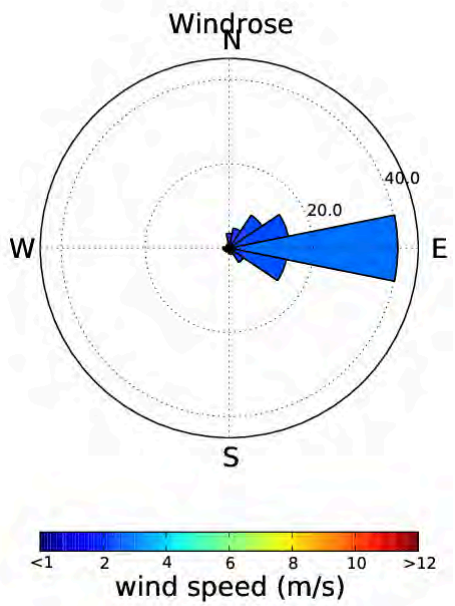

Fig. 41.— Wind rose for the Alldays station from the period 1994-2010. The colour indicates the mean speed in each of the sectors and the length of the bars indicate the frequency as a percentage of the total period. Strong channelling at the site from the local valley system results in the dominance of easterly winds and decoupling from the synoptic environment.

only a few tens of kilometers inland and to as high as 2000m (Estoque, 1962). The cluster 3 stations are likely dominated by drainage flows from the escarpment which may override synoptic flows, especially when these are weak (Kaufmann and Weber, 1996).

Cluster 5 is a good example of a region that has a variable coupling with the 850hPa winds. Figure 42 shows 2D histograms of wind speed and direction for the Cape Town Table Bay station under each of the synoptic states defined by the SOM. Annotated on the plots are the correlation coefficents for the station to 850hPa wind speeds. These range from 0.81 (for node 5) to 0.04 (for node 4), and the nodes with higher correlations (1 and 5) have more tightly defined distributions of wind patterns than those with lower correlations. A similar result was found by Argent et al. (2012) using a different model [vii] and observation set [viii]. In this case, the strong forcing of the mid-latitude cyclones, represented by nodes 1 and 5, override the local processes, like sea breezes, that influence the local wind climate. Nodes 4 and 7 by comparison, represent weaker, winter synoptics, with a shallow low pressure along the west coast (visible in the MSLP fields, and not the GPHs). In these conditions the surface flows are completely decoupled from the larger circulation, which is not the case for any of the other clusters.

4.6. Conclusion

In this chapter, regionalised wind clusters have been developed based on sypoptic forcing strength. The method has produced intuitive regions (for example, coastal versus inland, high altitude versus low) and the synoptic disagregation allows for process-based interpretations. Physically consistent reasons have been offered for the clustering and although the

[vii] The Weather Research and Forcasting model

[viii] A year's worth of observations from the Wind Atlas for South Africa project.

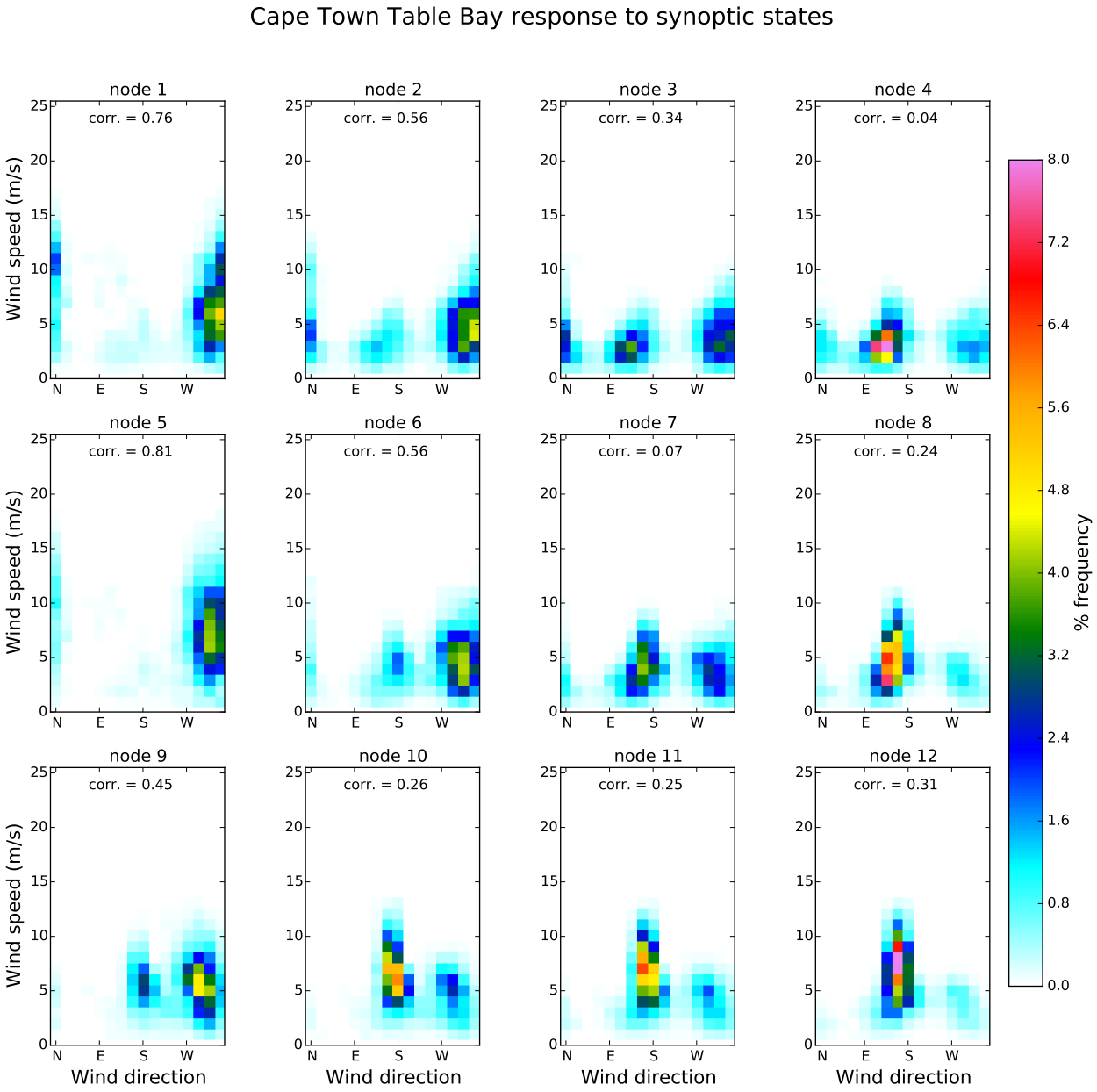

Fig. 42.— 2D histograms of hourly wind speed and direction for Cape Town Table Bay station from 1989-2010, for each synoptic state defined by the SOM. The correlation values indicate the correlation between the 850hPa CFSR wind speeds and those recorded at the station in each case. The signature distribution of winds can be seen for the different synoptic conditions.

relationship between the synoptic circulation and station observations is complex, some regions are found to be strongly controlled by the synoptic environment while others reflect stronger local controls. The degree of coupling varies under the different synoptic conditions.

Specifically, the regional surface to upper level coupling is affected by station altitude and placement, topographic blocking and thermal flows, as well as the large scale circulation systems that they are exposed to. Low level features like the shallow thermal low on the west coast enhance the potential for decoupling, and the deep Indian Ocean high pressure system tends to favour coupling in the north-east of the country. A few outliers are to be expected, and not all of the stations fall neatly into any particular cluster.

An application of this method is in the context of the placement of wind measurement masts for observation campaigns and wind resource assessment. An understanding of the wind regions facilitated by the clustering described in this chapter, allows for the optimal distribution of masts so as to ensure that all the wind climate areas are sampled. The relatively low station density does not support an explicit comparison between the cluster distributions and the climate regions defined by Kruger (2002) and Conradie (2012)(see Figures 32 and 33), and as these regions are related to different drivers we would not expect much overlap, however, there are some similarities, especially along the south and east coast and in the Southern Cape. There is no clear similarity with the extreme wind atlas (Kruger et al., 2011) which is not surprising as the small SOM is unlikely to resolve the unusual synoptic conditions that can lead to extreme wind speeds. The bias correction used to produce this data set is not ideally suited to preserving the extreme winds, as the number of wind classes was designed to maximise coverage of the whole histogram and not focused on the higher wind speeds. However a possible extension of this work would be to explore the extreme wind zones with this long period data set.

In the following chapter, the overall conclusions of the study are discussed as well as

the limitations, and suggestions are made for future work in this area.

5. Conclusions and recommendations

The aim of this study was two-fold: to compile a robust data set of surface wind observations for South Africa, and to use this to explore the relationships between the synoptic circulation and the wind climate. The next section gives an overview of the previous chapters in achieving these aims. This is followed by a discussion of the major conclusions of the study, its limitations and suggestions for future work.

5.1. Overview

In the first chapter, the motivation for this work was described in the context of a coal-reliant and climate-polluting country, with ambitions to shift the energy sector towards more environmentally-friendly renewable sources of energy and wind power in particular. The sparse literature on South African wind climate and the processes that drive it, as well as the lack of reliable wind observations were additional motivating factors.

Wind records from the 960 available South African Weather Service (SAWS) weather stations were obtained for analysis in Chapter 2. These records were found to be irregularly processed, lacking in meta-data and included numerous errors and periods of missing data. A quality control procedure was developed and applied to select the reliable, unsheltered stations, remove any errors and retain the records with a high proportion of available data. This process identified 107 reliable station records with sufficient data remaining after the quality control, in line with objectives 1.1 and 1.2. However, the short length of the records and quantity of missing data meant that only a small proportion of the period 1994-2010 had concurrent observations for all the stations.

Chapter 3 begins with an examination of the literature on the bias correction of model data, long term adjustment based on variability, and downscaling, for a way to fill the gaps

in the station records, and extend them back in time. A subset of the National Centers for Environmental Prediction's (NCEP) Climate Forecast Systems Reanalysis (CFSR) data set was selected for this purpose and was compared with the station records with respect to mean speeds, variability and long term trends. The CFSR data was found to match the general shapes of the wind speed distributions, but to generally over-estimate the mean wind speeds at the SAWS station sites and underestimate the trends and inter-annual variability.

In order to address the primary aim of the thesis, a series of novel methods was developed and tested, to combine the continuity and length of the CFSR records with information about the local climates in the SAWS station records. The most successful was found to be a correction method that adjusts the hourly CFSR reanalysis time series of wind speeds and direction, from the nearest model grid points to each station, towards the observed records as a function of wind speed, direction, time of day and month of the year. This method was found to be robust with respect to the length of training period and able to remove the mean bias in the seasonal and diurnal cycles, reduce the spread of errors, and improve the correlation between the CFSR and station time series. It was demonstrated that the method does not disrupt the spatial relationships between stations, a known deficiency for existing methods. The high number of classes used in this method is possibly limiting for its use with short time series, however was appropriate for the records available. The method was used to generate 30-year hourly time series of wind speed and direction for each of the SAWS station sites, satisfying objectives 1.3 and 1.4.

These time series were used in Chapter 4 to explore the coupling between the large scale circulation and the surface flows measured at the stations. In order to characterise the synoptic climate regimes (objective 2), a self-organising map (SOM) was used to divide the hourly 850hPa CFSR wind fields from the period 1989-2010 into 12 groups. The correlations between the station time series and the nearest grid cell 850hPa wind speeds, under each of

the synoptic conditions defined by the SOM, were used to cluster the stations. A hierarchical clustering produced 6 groups of stations, each driven to a greater or lesser extent by the different synoptic conditions.

The coupling was found to be greater in the interior, at higher altitude stations and away from topographic features. Where there is weak coupling, local processes establish the characteristic wind flow and in both cases, relatively cohesive spatial groupings emerged. The characterisation of the relationship between the surface winds in these regions and the synoptic circulation satisfies the final objective of the thesis by initiating research into the drivers of South African surface wind climate.

5.2. Conclusions

5.2.1. Station records

In terms of the SAWS weather station records, the first major conclusion is that these station records should be used with caution, especially for applications like wind energy resource assessment, which are sensitive to wind speed estimates. The proportion of missing and erroneous data is an order of magnitude higher than was found for other comparable data sets (Jimenez et al., 2010b). Poor siting of stations with respect to nearby buildings and obstacles, and lack of meta data, further reduce the utility of this data set. As a result, there is a relatively low density of reliable stations and understanding of the South Africa wind climate could be improved by siting stations in regions that are data-sparse. Increased availability of meta data may also allow for correction of records with respect to the sheltering of stations, as has been done in some cases by Kruger et al. (2010). A stricter quality control, as a part of the data collection, is needed to improve the data collected in future.

5.2.2. CFSR bias and correction

The CFSR data set used for this study was found to capture many of the features of the surface circulation, but contains large biases with respect to wind speed. In particular, it tends to overestimate surface wind speeds and underestimate interannual and spatial variability. Diurnal and seasonal cycles are variably captured, well in some regions and poorly in others, but in general the magnitude of the diurnal variations is less than is observed. The long term trends observed at some of the stations are almost completely absent from the CFSR time series possibly due to local factors not resolved at the grid resolution of the CFSR data.

The biases in the CFSR time series were found to be largely a function of wind speed, direction, time of day and month of the year and the bias correction was most successful when the interdependence of these biases were taken into account. This implies a processed-based bias in the CFSR as the synoptic circulation have clear seasonal and diurnal characterises and particular wind speed - wind direction distributions. Further work to explore these biases and explain the causes could potentially improve the reanalysis.

The method, which uses elements from a wide range of methodologies and yet is relatively simple to apply, was found to improve the CFSR seasonal and diurnal cycles and remove mean bias, and reduce (to some extent) the root mean square error of wind speed. As opposed to some other techniques (as described by Ehret et al. (2012)), this method was able to preserve, and in fact improve, the representation of spatial relationships between stations as measured by their cross-correlations.

Despite this relative success, some of the time series remain poor representations of the station records due to biases that are independent of those addressed. Mean correlation coefficients between the CFSR and station time series were only marginally improved, from 0.70 to 0.73. As the 10m CFSR winds are derived from the upper level winds, which, as

discussed in Chapter 4, are not always the major driver of surface winds, it is not surprising that these correlations are low to begin with, especially for the regions that are relatively independent of the upper level circulation. The CFSR trends were relatively unaffected by the correction and for the minority of stations with higher inter-annual variability and appreciable long term trends, the extension of the records is less robust than for the rest. A more thorough examination of trends in different reanalysis products is advised for future work especially in the light conflicting trends found in other parts of the world between different data sets (Pryor et al., 2009).

An analysis of the CFSR biases as a function of the synoptic state, though beyond the scope this study, would perhaps give some insight into how well different processes are captured by this, and other reanalyses, with respect to the surface winds. A synoptic circulation based correction may, however, be of limited value in the areas that are not strongly driven from this level. The bias corrected time series and may be useful for a more thorough validation of the Wind Atlas of South Africa (WASA) and as a guide for wind resource assessments. A spatial extension of the bias corrections may be possible within the vicinity of the stations but is not recommended in the areas of complex terrain as the representativity of observations in these regions is limited.

Further analysis of the trends and variability of wind speeds could be done by comparison to other data available including alternative reanalysis products, climate model simulations and satellite winds.

5.2.3. Synoptic regimes and surface wind coupling

The different synoptic states found in Chapter 4 are similar to those described by Hewitson and Crane (2002, 2006), Lennard and Hegerl (2014) and Pinto (2015) and include the

major synoptic states described by Preston-Whyte (1974). These states capture the large scale circulation drivers of wind in South Africa and result in clearly defined wind distributions at the station level, especially those for which the surface to upper level coupling is strong. As particular states are responsible for the highest wind speeds in different regions, trends in the frequency of these states could have major consequences for the future wind resources of the country and are a topic for future work.

The clustering based on the station-to-upper-level coupling produced relatively cohesive regions, and facilitates a process-based interpretation. Some of the factors responsible for the regional differences have been explored and the local processes that sometimes override the synoptic forcings were described. While the regions bear some similarity to previous climate classifications, there is no clear match with the extreme wind regions defined by (Kruger et al., 2011). While unsurprising given the methodological differences and specific drivers of extreme winds, this result highlights the need for further investigation into the processes that drive surface flows in South Africa. The regions described in Chapter 4 may also be used to guide the placement of wind observation masts so as to ensure the sampling of all wind climatic areas.

Some reasons have been suggested for the strong decoupling found especially in the coastal areas. These include topographic influences, thermal flows and shallow atmospheric features. This has implications for downscaling and bias correction methods, as these often assumed that there is strong coupling.

5.3. Concluding remarks

The results of this thesis speak to the growing demand for information about the wind climate of South Africa, in the context of wind resource assessment. The cubic relationship

between wind speed at a given location, and the available wind power, enhances the importance of the reliability of wind measurements and accuracy of modeled wind speeds. The major contributions of this thesis to that goal are, a) the compilation of a robust data set of wind observations, b) the development of a novel method for correcting reanalysis data to fill gaps and extend historical records, c) the dis-aggregation of the synoptic circulation patterns with a self-organising map, and d) the quantification of the degree to which the archetypal synoptic scale circulation patterns govern the surface wind speeds in regions around the country. Additionally, this study lays a platform for studies of future wind climate in a warming world.

REFERENCES

K. Alapaty, S. Raman, and D. S. Niyogi. Uncertainty in the specification of the surface characteristics: A study of prediction errors in the boundary layer. *Boundary-Layer Meterology*, 82:473–500, 1997.

B. Argent, A. Hahmann, C. Lennard, and B. Hewitson. Towards an uncertainty atlas for wind in South Africa. *Proceedings of European Wind Energy Association conference*, 2012.

J. Badger, H. Frank, A. N. Hahmann, and G. Giebel. Wind climate estimation based on mesoscale and microscale modeling: statistical dynamical downscaling for wind energy applications. *American Meteorological Society*, 53:1901–1919, August 2014.

B. H. Bailey and S. L. McDonald. Wind resource assessment handbook, National Renewable Energy Laboratory, New York. *AWS Scientific*, 1997.

D. Banks and J. Schaffler. The potential contribution of renewable energy in South Africa. Technical report, Sustainable Energy & Climate Change Project, Earthlife Africa, Johannesburg, August 2008.

H. Bergström. Boundary-layer Modelling for wind climate estimates. *Wind Engineering*, 25 (5):289–299, 2001.

R. Bhattacharyya and T. Ghoshal. Economic growth and CO_2 emissions. *Environment, Develpment and Sustainability*, 12(2):159–177, April 2010.

M. Brunet, O. Saladié, P. Jones, J. Sigró, E. Aguilar, A. Moberg, D. Lister, A. Walther, D. Lopez, and C. Almarza. The development of a new dataset of Spanish daily adjusted temperature series. *International Journal of Climatology*, (26):1777–1802, 2006.

M. J. Bunkers and J. R. Miller. Definition of Climate regions in the northern plains using a an objective cluster modification technique. *Journal of Climate*, 9:130–146, January 1996.

M. Burlando. The synoptic-scale surface wind climate regimes of the Mediterranean Sea according to the cluster analysis of ERA-40 wind fields. *Theoretical and Applied Climatology*, 96(1-2):69–83, 2009.

M. Burlando, M. Antonelli, and C. F. Ratto. Mesoscale wind climate analysis: identification of anemological regions and wind regimes. *International Journal of Climatology*, 28: 629–641, June 2008.

F. Cassola, M. Burlando, M. Antonelli, and C. F. Ratto. Optimization of the Regional Spatial Distribution of Wind Power Plants to Minimize the Variability of Wind Energy Input into Power Supply Systems. *Journal of Applied Meteorology and Climatology*, 47(12): 3099–3116, 2008.

J. H. Christensen, F. Boberg, O. B. Christensen, P. Lucas-Picher, et al. On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophysical Research Letters*, 35, October 2008.

G. P. Compo, J. S. Whitaker, P. D. Sardeshmukh, N. Matsui, R. J. Allan, X. Yin, B. E. Gleason, R. S. Vose, G. Rutledge, P. Bessemoulin, S. Brönnimann, M. Brunet, R. I. Crouthamel, A. N. Grant, P. Y. Groisman, P. D. Jones, M. C. Kruk, A. C. Kruger, G. J. Marshall, M. Maugeri, H. Y. Mok, Ø. Nordli, T. F. Ross, R. M. Trigo, X. L. Wang, S. D. Woodruff, and S. J. Worley. The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society*, 137(654):1–28, 2011.

D. C. U. Conradie. South Africa's climatic zones: today, tomorrow. July 2012.

D. S. Crosby, L. C. Breaker, and W. H. Gemmill. A proposed definition for vector correlation in geophysics: theory and application. *Journal of Atmospheric and Oceanic Technology*, 10:355–367, June 1993.

P. Cury and C. Roy. Optimal Environmental Window and Pelagic Fish Recruitment Success in Upwelling Areas. *Canadian Journal of Fisheries and Aquatic Sciences*, (46(4)): 670–680, 1989.

A. T. DeGaetano. A quality-control routine for hourly wind observations. *Journal of Atmospheric and Oceanic Technology*, 14(308), August 1996.

A. Devis, N. P. M. van Lipzig, and M. Demuzere. A new statistical approach to downscale qind speed distributions at a site in northern Europe. *Journal of Geophysical Research: Atmospheres*, 118:2272–2283, March 2013.

R. Diab. Wind Atlas of South Africa, Technical report. *Department of Mineral and Energy Affairs, Pretoria*, 1995.

O. Edenhofer, R. Pichs-Mandruga, Y. Sokona, K. Seyboth, P. Matschoss, S. Kadner, T. Zwickel, P. Eickemeier, G. Hansen, S. Schlömer, and C. v. S. (eds). *IPCC, 2011: Summary for Policymakers*. Cambridge University Press, 2011.

U. Ehret, E. Zehe, V. Wulfmeyer, K. Warrach-Sagi, and J. Liebert. Should we apply bias correction to global and regional climate model data. *Hydrology and Earth System Sciences Discussions*, 9:5355–5387, April 2012.

J. K. Eischeid, C. B. Baker, T. R. Karl, and H. F. Diaz. Quality control of long term climatological data using objective analysis. *Journal of Applied Meteorology*, (2787), December 1995.

Energistyrelsen. Energiscenarier frem mod 2020, 2035 og 2050. Online. (Accessed 17.06.2015).

D. o. M. Energy(DoME). White Paper on Renewable Energy, 2003.

C. J. Engelbrecht and F. A. Engelbrecht. Shifts in Köppen-Geieger climate zones over Southern Africa in relation to key global temperture goals. *Theoretical and Applied Climatology*, 123:247–261, 2015.

J. P. Engelbrecht, L. Swapepoel, J. C. Chow, J. G. Watson, and R. T. Egami. PM2. 5 and PM10 concentrations from the Qalabotjha low-smoke fuels macro-scale experiment in South Africa. *Environmental Monitoring and Assessment*, 69:1–15, 2001.

B. F. N. Erasmus, A. S. V. Jaarsveld, S. L. Chown, M. Kshatriya, and K. J. Wessels. Vulnerability of South African animal taxa to climate change. *Global Change Biology*, 8:679–693, 2002.

M. A. Estoque. The sea breeze as a function of the prevailing synoptic situation. *Journal of the Atmospheric Sciences*, 19(3):244–250, 1962.

H. J. Fowler and C. G. Kilsby. Using regional climate model data to simulate historical and future river flows in northwest England. *Climate Change*, 80:337–367, 2007.

L. S. Gandin. Complex Quality Control of Meteorological Observations. *Monthly Weather Review*, (1137), May 1988.

M. Garstang, P. D. Tyson, R. Swap, M. Edwards, and P. Kallber. Horizontal and vertical transport of air over southern Africa. *Journal of Geophysical Research*, 101(19): 23,721–723, October 1996.

F. Giorgi and L. O. Mearns. Introduction to special section: Regional climate modeling revisted. *Journal of Geophysical Research*, 104(6):6335–3352, March 1999.

A. Goliger, A. Kruger, and J. Retief. Representivity of wind measurements for design wind speed estimations. *Proceedings of the 6th European-African Conference on Wind Engineering, Cambridge,United Kingdom*, July 2013.

D. Y. Graybeal. Relationships among daily mean and maximum wind speeds with applications to data quailty assurance. *International journal of climatology*, (26):29–43, 2006.

I. Haddeland, J. Heinke, F. Voss, S. Eisner, C. Chen, S. Hagemann, F. Ludwig, et al. Effects of climate model radiation, humidity and wind estimates on hydrological simulations. *Hydrology and Earth System Sciences*, 16:305–318, February 2012.

J. O. Haerter, S. Hagemann, C. Moseley, and C. Piani. Climate model bias correction and the role of timescales. *Hydrology and Earth System Sciences*, 15:1065–1079, March 2011.

K. Hagemann. *Mesolscale WInd Atlas of South Africa*. PhD thesis, University of Cape Town, 2008.

A. N. Hahmann, C. Lennard, J. Badger, C. L. Vincent, M. C. Kelly, P. J. H. Volker, B. Argent, and J. Refslund. Mesoscale modeling for the wind atlas for South Africa (WASA) Project. Technical report, DTU Wind Energy, February 2015.

M. S. J. Harrison. A generalized classification of South African summer rain-bearing synoptic systems. *Journal of Climatology*, 4(5):547–560, 1984.

L. E. Hay, M. P. Clark, R. L. Wilby, W. J. Gutowski, G. H. Leavesley, Z. Pan, R. W. Arritt, and E. S. Takle. Use of Regional Climate Model Output for Hydrologic Simulations. *Journal of Hydrometeorology*, 3(5):571–590, 2002.

B. C. Hewitson and R. C. Crane. Consensus between GCM climate change projections with empirical downscaling. *International Journal of Climatology*, 26(10):1315–1338, 2006.

B. C. Hewitson and R. G. Crane. Self-organising maps: Applications to synoptic climatology . *Climate Research*, 22:13–26, 2002.

P. K. Hope, W. Drosdowsky, and N. Nicholls. Shifts in the synpotic systems influencing southwest Australia. *Climate Dynamics*, 26:751–764, 2006.

T. Howard and P. Clark. Correction and downscaling of NWP wind speed forecasts. *Applied Metoerology*, 14:105–116, 2007.

R. Huva, R. Dargaville, and P. Rayner. The impact of filtering self-organizing maps: a case study with Australian pressure and rainfall. *International Journal of Climatology*, 35: 624–633, March 2014.

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

P. A. Jimenez, J. F. G. lez Rouco, J. P. M. vez, J. Navarro, E. G. a Bustamante, and F. Valero. Surface wind regionalization in complex terrain. *Journal of Applied Meteorology and Climatology*, 47:308–325, January 2008.

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

P. A. Jimenez, J. F. G. lez Rouco, E. G. a Bustamante, J. Navarro, J. P. M. vez, J. V.-G. de Arellano, J. Dudhia, and A. Munoz-Roldan. Surface wind regionalization over complex terrain: evaluation and analysis of high-resolution WRF simluation. *Journal of Applied Meteorology and Climatology*, 49:268–287, February 2010a.

P. A. Jimenez, J. F. G. Rouco, J. Navarro, J. P. Monta´vez, and E. G. Bustamante. Quality Assurance of Surface Wind Observations from Automated Weather Stations. *Journal of Atmospheric and Ocean Technology*, 27, July 2010b.

P. A. Johnston. Multi-variate climate classification of Southern Africa. June 1980.

M. Jun, R. Knutti, and D. W. Nychka. Spatial analysis to quantify numerical model bias and dependence. *Journal of the American Statistical Association*, 103:934–947, January 2008.

M. R. Jury. A climatological mechanism for wind-driven upwelling near Walker Bay and Danger Point, South Africa. *South African Journal of Marine Science*, 6(1):175–181, 1988.

S. Karekezi. Poverty and energy in Africa - A brief review. *Energy Policy*, 30(11):915–919, 2002.

T. R. Karl and C. N. Williams. An approach to adjusting climatological time series for discontinuous inhomogeneities. *Journal of Climate and Applied Meteorology*, 20(1744), December 1987.

P. Kaufmann and R. O. Weber. Classification of Mesoscale Wind Fields in the MISTRAL Field Experiment. *Journal of Applied Meteorology*, 35(11):1963–1979, 1996.

P. Kaufmann and C. D. Whiteman. Cluster-Analysis Classification of Wintertime Wind Patterns in the Grand Canyon Region. *Journal of Applied Meteorology*, 38(8):1131–1147, 1999.

M. C. Kirchmeier, D. J. Lorenz, and D. J. Vimont. Statistical downscaling of daily wind speed variations. *Journal of Applied Meteorology and Climatology*, 53:660–675, March 2014.

T. Kohonen. The self-organing map. *Proceedings of the IEEE*, 78(99):1464–1480, 1990.

A. C. Kruger. *Climate of South Africa: surface winds*. South African Weather Service, 2002.

A. C. Kruger, A. M. Goliger, J. V. Retief, and S. Sekele. Strong wind climatic zones in South Africa. *Wind and Structures*, 13(1), 2010.

A. C. Kruger, J. V. Retief, and A. M. Goliger. An Updated Description of the Strong-Wind Climate of South Africa. July 2011.

W. A. Landman and J. Simon. Change in the association between Indian Ocean sea-surface temperatures and summer rainfall over South Africa and Namibia. *International Journal of Climatology*, 19(13):1477–1492, 1999.

C. Lennard and G. Hegerl. Relating changes in synoptic circulation to surface rainfall response using self-organising maps. *Climate Dynamics*, 44:861–879, 2014.

S. Liléo, E. Berge, O. Undheim, R. Klinkert, and R. E. Bredesen. Long-term correction of wind measurements: State-of-the-art, guidlines and future work. *Elforsk Technical Report*, January 2013.

D. Loy. Energy-policy Framework Conditions for Electricity Markets and Renewable Energies -21 Country Anaylses. Technical report, June 2004.

S. J. Mason. Seasonal forecasting of South African rainfall using a non–linear discriminant analysis model. *International Journal of Climatology*, 18(2):147–164, 1998.

G. F. Midgley, L. Hannah, D. Millar, M. C. Rutherford, and L. W. Powrie. Assessing the vulnerability of species richness to anthropogenic climate change in a biodiversity hotspot. *Global Ecology & Biogeography*, 11:445–451, 2002.

G. M. Mimmack, S. J. Mason, and J. S. Galpin. Choice of distance matrices in cluster analysis: Defining Regions. *Journal of Climate*, 14:2790–2797, December 2000.

N. G. Mortensen, J. C. Hansen, J. Badger, B. H. Jørgensen, C. B. Hasager, U. S. Paulsen, O. F. Hansen, K. Enevoldsen, L. G. Youssef, U. Said, et al.

N. G. Mortensen, L. Landberg, I. Troen, and E. L. Petersen. Wind atlas analysis and application program (WAsP): Vol. 1: Getting started. *Riso-I-666 (EN), Riso Nat. Lab., Roskilde*, 1993.

M. Muller. Adapting to climate change: water management for urban resilience. *Environment and Urbanization*, 19(99), 2007.

C. Piani, J. O. Haerter, and E. Coppola. Statistical bias correction for daily precipitation in regional climate models over Europe. *Journal of Theoretical Applied Climatology*, 99:187–192, April 2010.

I. Pinto. *Future change in extreme rainfall events and circulation patterns over southern africa*. PhD thesis, University of Cape Town, 2015.

R. A. Preston-Whyte. Climatic Classification of South Africa: A Multivariate Approach. *South African Geographical Journal*, 56(1):79–86, 1974.

O. Probst and D. Cárdenas. State of the art and trends in wind resource assessment. *Energies*, 3:1087–1141, June 2010.

S. C. Pryor, J. T. Schoof, and R. J. Barthelmie. Empirical downscaling of wind speed probability distributions. *Journal of Geophysical Research*, 110, October 2005.

S. C. Pryor, R. J. Barthelmie, D. T. Young, E. S. Takle, R. W. Arritt, D. Flory, W. J. G. Jr., A. Nunes, and J. Roads. Wind speed trends over the contiguous United States. *Journal of Geophysical Research*, 114, 2009.

B. Rennkamp and A. Boyd. Technological capability and transfer for achieving South Africa's development goals. *Climate Policy*, September 2013.

D. M. Richardson, B. W. van Wilgen, D. C. L. Maitre, K. B. Higgins, and G. G. Forsynth. A Computer-Based System for Fire Management in the Mountains of the Cape Province, South-Africa. *Internationsl Journal of Wildland Fire*, 4(1):17–32, 1994.

W. C. D. Rooy and K. Kok. A Combine phyiscal-statistical approach for the downscaling of model wind speed. *Weather and Forecasting*, 19:485–495, June 2004.

A. C. Ruane, R. Goldberg, and J. Chryssanthacopoulos. Climate forcing data sets for agriculture modeling: Merging products for gap-filling and historical climate series estimation. *Agriculture and Forest Meteorology*, 200:233–248, 2015.

M. Rummukainen. State-of-the-art with regional climate models. *WIley Interdisciplinary Reviews: Climate Change*, 1:82–96, January 2010.

S. Saha, S. Moorthi, H. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer, H. Liu, D. Stokes, R. Grumbine, G. Gayno, J. Wang, Y. Hou, H. Chuang, H. H. Juang, J. Sela, M. Iredell, R. Treadon, D. Kleist, P. V. Delst, D. Keyser, J. Derber, M. Ek, J. Meng, H. Wei, R. Yang, S. Lord, H. V. D. Dool, A. Kumar, W. Wang, C. Long, M. Chelliah, Y. Xue, B. Huang, J. Schemm, W. Ebisuzaki, R. Lin, P. Xie, M. Chen, S. Zhou, W. Higgins, C. Zou, Q. Liu, Y. Chen, Y. Han, L. Cucurull, R. W. Reynolds, G. Rutledge, and M. Goldberg. The NCEP Climate Forecast System Reanalysis. *Bulletin of the American Meteorological Society*, 91(8): 1015–1057, 2010a.

S. Saha et al. NCEP Climate Forecast System Reanalysis (CFSR) Selected Hourly Time-Series Products, January 1979 to December 2010. *NCEP Climate Forecast System Reanalysis (CFSR) Selected Hourly Time-Series Products, January 1979 to December 2010, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, CO.*, 2010b.

D. J. Sailor, T. Hu, X. Li, and J. N. Rosen. A neural network approach to local downscaling of {GCM} output for assessing wind power implications of climate change. *Renewable Energy*, 19(3):359–378, 2000.

SAWS. De Aar AWS metadata report. Technical report, 2012a.

SAWS. Joubertina AWS metadata report. Technical report, 2012b.

SAWS. Koingnaas AWS metadata report. Technical report, 2012c.

J. Schmidli, C. Frei, and P. L. Vidale. Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods. *International Journal of Climatology*, 26(5):679–689, April 2006.

S. Solomon, D. Qin, M. Manning, M. Marquis, K. Averyt, M.Tignor, H. M. (eds.), and Z. C. (eds.). *IPCC, 2007: Summary for Policymakers. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, 2007.

A. N. Starkof, P. P. Bezroukikh, M. Borisenko, L. Landberg, et al. *Russian Wind Atlas.* Russian-Danish Institute for Energy Efficiency, 2000.

Y. Sun, S. Solomon, A. Dai, and R. W. Portman. How often does it rain. *Journal of Climate*, 19:916–934, March 2006.

J. J. Taljaard and D. Phil. *Atmospheric circulation systems, synoptic climatology and weather phenomena of South Africa.* Government Printer, South Africa, 1996.

M. J. Themessl, A. Gobiet, and A. Leuprecht. Empirical-statistical down-scaling and error correction of daily precipitation from regional climate models. *International Journal of Climatology*, 31:1530–1544, 2011.

M. L. Thøgersen, M. Motta, T. Sørensen, and P. Nielsen. Measure-correlate-predict methods: Case studies and software implementation. *European Wind Energy Conference & Exhibition, WEA.,*, May 2007.

M. L. Thøgersen, P. Nielsen, T. Sørensen, and L. U. Svenningsen. MCP -Measure-Correlate-Predict, An introduction to the MCP facilities in WindPRO, 2010.

D. S. G. Thomas, C. Twyman, H. Osbahr, and B. Hewitson. Adaptation to climate change and variability: farmer responses to intra-seasonal precipitation trends in South Africa. *Climate Change*, 83:301–322, March 2006.

C. W. Thornthwaite. An approach towards a rational classification of climate. *Geographical Review*, 38(1):55–94, January 1948.

K. E. Trenberth, T. Koike, and K. Onogi. Progress and Prospects for Reanalysis for Weather and Climate. *Eos Transactions American Geophysical Union*, 89(26):234–235, June 2008.

I. Troen and E. Petersen. *European wind atlas*. Risoe National Laboratory,Roskilde, DK, 1989.

R. Vautard, J. Cattiaux, P. Yiou, J. Thepaut, and P. Ciais. Northern Hemisphere atmospheric stilling partly attributed to an increase in surface roughness. *Nature Geoscience*, 3(11):756–761, November 2010.

L. A. Vincent and X. Zhang. Homogenization of daily temperature over Canada. *Journal of Climate*, 15(1322), June 2002.

R. O. Weber and P. Kaufmann. Automated classification scheme for wind fields. *Journal of Applied Meteorology*, 34:1133–1141, May 1995.

J. B. Wijngaard, A. M. G. K. Tank, G. P. Können, et al. Homogeneity of 20th century European daily temperature and precipitation series. *International Journal of Climatology*, 23:679–692, 2003.

R. L. Wilby, T. M. L. Wigley, D. Conway, P. D. Jones, B. C. Hewitson, J. Main, and D. S. Wilks. Statistical downscaling of general circulation model output: A comparison of methods . *Water Resource Research*, 34(11):2995–3008, November 1998.

R. A. I. Wilcke, T. Mendlik, A. Gobiet, et al. Multi-variable error correction of regional climate models. *Climate Change*, 120:871–887, August 2013.

WMO. *WMO guide to meteorological instruments and methods of observations*. Secretariat of the World Meteorological Organization, 2010.

A. W. Wood, L. R. Leung, V. Sridhar, and D. P. Lettenmaier. Hydrological implications of dynamical and statistical approacjes to downscaling climate model outputs. *Climate Change*, 62:189–216, 2004.

6. Appendix A

The following table shows the classification of the South African Weather Service station records according to the quality control criteria described in Chapter 2. A positive 'record quality' indicates that at least 3 years of data were available in the period 1994-2010 that contained at least 90% of the hourly wind speed and direction records. The exposure criteria excludes station that are excessively sheltered by obstacles or buildings in the immediate vicinity in accordance with the World Meteorological Organisations guidelines (WMO, 2010).

Table 6. South African Weather Service weather station details and selection criteria

Station	id	lat	long	record quality	Exposure	included
ADDO ELEPHAN..	0055447A7	-33.4421	25.7485	check	check	yes
ALEXANDERBAA..	0274034A4	-28.5700	16.5287	check	check	yes
ALIWAL-NORTH..	0175678A0	-30.8029	26.8834	check	check	yes
ALLDAYS	0764161B1	-22.6830	29.1111	check	check	yes
AUGRABIES FA..	0281606A5	-28.5950	20.3401	check	check	yes
BETHLEHEM WO	0331585 9	-28.2496	28.3343	check	check	yes
BIRD ISLAND	0036500 7	-33.8389	26.2861	check	check	yes
BISHO	0079504 X	-32.9000	27.2800	check	check	yes
BLOEMFONTEIN..	0261307A4	-29.1204	26.1874	check	check	yes
BLOEMFONTEIN..	0261516B0	-29.1039	26.2981	check	check	yes
BLOEMHOF	0362189 7	-27.6511	25.6219	check	check	yes
BRANDVLEI	0190868 1	-30.4647	20.4785	check	check	yes
CALVINIA WO	0134479A3	-31.4819	19.7617	check	check	yes
CAPE ST FRAN..	0017582A1	-34.2123	24.8352	check	check	yes
CAPE TOWN TA..	0020774 4	-33.9000	18.4300	check	check	yes
CEDARA	0239482 0	-29.5419	30.2650	check	check	yes
CHARTERS CRE..	0339732A9	-28.2000	32.4170	check	check	yes
CRADOCK-MUN	0098190B6	-32.1673	25.6253	check	check	yes
DE AAR WO	0169880 1	-30.6650	23.9928	check	check	yes
DOORNLAAGTE	0586006A4	-24.6000	26.5170	check	check	yes
DURBAN SOUTH..	0240808A2	-29.9650	30.9467	check	check	yes

Table 6—Continued

Station	id	lat	long	record quality	Exposure	included
EAST LONDON ..	0059572B8	-33.0342	27.8294	check	check	yes
ELLIOT	0150620AX	-31.3362	27.8403	check	check	yes
ERMELO WO	0479870 X	-26.4978	29.9839	check	check	yes
ESTCOURT	0300630 8	-29.0069	29.8656	check	check	yes
FAURESMITH	0291570 1	-29.7535	25.3230	check	check	yes
FICKSBURG	0296709AX	-28.8272	27.9043	check	check	yes
FORT BEAUFOR..	0078227A3	-32.7881	26.6294	check	check	yes
GARIEP DAM	0173032 4	-30.5621	25.5278	check	check	yes
GARIES-GROEN..	0157111 3	-30.8650	17.5792	check	check	yes
GEELBEK	0040192 4	-33.1961	18.1242	check	check	yes
GEORGE WO	0012661 7	-34.0019	22.3833	check	check	yes
GLEN COLLEGE..	0293597A6	-28.9420	26.3258	check	check	yes
GRAAFF - REI..	0096072 5	-32.1933	24.5431	check	check	yes
GRAHAMSTOWN	0056917 8	-33.2906	26.5028	check	check	yes
GRAND CENTRA..	0513239 0	-25.9926	28.1401	check	check	yes
GRASKOP AWS	0594626B9	-24.9330	30.8500	check	check	yes
HARRISMITH -..	0332887 3	-28.2800	29.0000	check	check	yes
HOEDSPRUIT	0638081 1	-24.3500	31.0500	check	check	yes
IRENE WO	0513385A2	-25.9100	28.2111	check	check	yes
IXOPO	0210099A7	-30.1525	30.0747	check	check	yes
JAMESTOWN	0148517A9	-31.1211	26.8097	check	check	yes

Table 6—Continued

Station	id	lat	long	record quality	Exposure	included
JOHANNESBURG..	0476399 0	-26.1500	28.2300	check	check	yes
KATHU	0356880 4	-27.6703	23.0063	check	check	yes
KIMBERLEY WO	0290468A9	-28.8057	24.7696	check	check	yes
KLERKSDORP	0436204 1	-26.8683	26.7149	check	check	yes
KOINGNAAS	0184491 4	-30.1955	17.2903	check	check	yes
KOKSTAD	0269894 1	-30.5014	29.3972	check	check	yes
KRUGER MPUMA..	0556173 6	-25.3914	31.0992	check	check	yes
KURUMAN	0393806 4	-27.4332	23.4474	check	check	yes
LADYSMITH	0300454 3	-28.5756	29.7503	check	check	yes
LEPHALALE	0674341 8	-23.6800	27.7000	check	check	yes
LICHTENBURG	0472278 0	-26.1330	26.1670	check	check	yes
MADIKWE GAME..	0585341 8	-24.6903	26.1944	check	check	yes
MAFIKENG WO	0508047 0	-25.8037	25.5428	check	check	yes
MANDINI	0271699 2	-29.1500	31.4000	check	check	yes
MARGATE	0182591A4	-30.8500	30.3330	check	check	yes
MARKEN	0675666 2	-23.6000	28.3830	check	check	yes
MBAZWANA AIR..	0412148 6	-27.4772	32.5981	check	check	yes
MIDDELBURGDA..	0516076 7	-25.7670	29.5500	check	check	yes
MOLTENO RESE..	0020746 6	-33.9375	18.4108	check	check	yes
MTUNZINI	0304357 6	-28.9500	31.7000	check	check	yes
NEWCASTLE	0370856 3	-27.7689	29.9767	check	check	yes

Table 6—Continued

Station	id	lat	long	record quality	Exposure	included
NOUPOORT	0144791 2	-31.1863	24.9604	check	check	yes
ORIBI AIRPOR..	0239699 7	-29.6500	30.4000	check	check	yes
OUDTSHOORN	0028306 5	-33.6000	22.1878	check	check	yes
PAARL	0021823 0	-33.7217	18.9719	check	check	yes
PADDOCK	0182465 7	-30.7544	30.2578	check	check	yes
PATENSIE	0033556 5	-33.7653	24.8233	check	check	yes
PILANESBERG	0548375A4	-25.2573	27.2238	check	check	yes
PLETTENBERGB..	0014545 4	-34.0896	23.3257	check	check	yes
POMFRET	0504050 X	-25.8298	23.5300	check	check	yes
PONGOLA	0410175 X	-27.4139	31.5917	check	check	yes
PORT ALFRED	0037666 6	-33.5594	26.8813	check	check	yes
PORT ALFRED ..	0037574 1	-33.5594	26.8811	check	check	yes
PORT EDWARD	0155394A5	-31.0670	30.2330	check	check	yes
PORT ELIZABE..	0035209B1	-33.9844	25.6108	check	check	yes
PORTERVILLE	0041841 X	-33.0128	18.9772	check	check	yes
POSTMASBURG	0321110 7	-28.3460	23.0786	check	check	yes
POTCHEFSTROO..	0437104A4	-26.7359	27.0755	check	check	yes
PRIESKA	0224400 8	-29.6701	22.7401	check	check	yes
PRINS ALBERT	0048011 7	-33.1830	22.0170	check	check	yes
QUEENSTOWN	0123685 X	-31.9178	26.8779	check	check	yes
RICHARDS BAY..	0305134 6	-28.7378	32.0934	check	check	yes

Table 6—Continued

Station	id	lat	long	record quality	Exposure	included
RIVERSDALE	0010426 8	-34.0961	21.2519	check	check	yes
ROBBENEILAND	0020618 X	-33.7989	18.3744	check	check	yes
ROYAL NATION..	0298791 9	-28.6858	28.9542	check	check	yes
STILBAAI	0010682 0	-34.3689	21.3914	check	check	yes
STRUISBAAI	0003108A7	-34.8008	20.0589	check	check	yes
TAUNG	0360453A0	-27.5452	24.7687	check	check	yes
TSITSIKAMMA	0015692A4	-34.0229	23.8986	check	check	yes
UITENHAGE	0034763 X	-33.7142	25.4350	check	check	yes
ULUNDI	0337738 2	-28.3000	31.4200	check	check	yes
UMTHATHA WO	0127272A4	-31.5300	28.6700	check	check	yes
UPINGTON WO	0317475A8	-28.4111	21.2642	check	check	yes
VAN REENEN	0333682A9	-28.3789	29.3853	check	check	yes
VAN ZYLSRUS	0427083B8	-26.8774	22.0499	check	check	yes
VENTERSDORP	0473559A3	-26.3170	26.8170	check	check	yes
VEREENIGING	0438784 3	-26.5670	27.9500	check	check	yes
VIRGINIA	0241076 6	-29.7670	31.0500	check	check	yes
VRYBURG	0432237 3	-26.9547	24.6527	check	check	yes
WARMBAD TOWO..	0589594 1	-24.9000	28.3300	check	check	yes
WELKOM	0364300 1	-27.9947	26.6658	check	check	yes
WEPENER	0232654 4	-29.9156	26.8469	check	check	yes
WILLOWMORE	0050887 2	-33.3003	23.4839	check	check	yes

Table 6—Continued

Station	id	lat	long	record quality	Exposure	included
WITBANK	0515320 8	-25.8369	29.1887	check	check	yes
WORCESTER-AW..	0022729 X	-33.6639	19.4183	check	check	yes
BABANANGO	0337382 5	-28.3700	31.2200	check	fail	no
BARKLY-OOS (..	0177176 4	-30.9300	27.6000	check	fail	no
BELFAST	0517041 2	-25.6918	30.0344	check	fail	no
BOLEPI HOUSE	0513439A1	-25.8094	28.2564	check	fail	no
CAPE AGULHAS	0003020 4	-34.8264	20.0131	check	fail	no
CAPE POINT	0004891 9	-34.3536	18.4897	check	fail	no
CAPE POINT A..	0004921 6	-34.3531	18.4899	check	fail	no
CAPE TOWN - ..	0020805 3	-33.9204	18.4430	check	fail	no
CAPE TOWN SL..	0004549 2	-34.1483	18.3192	check	fail	no
CAPE TOWN WO	0021178A3	-33.9789	18.6000	check	fail	no
CT-AWS	0021178B8	-33.9789	18.6000	check	fail	no
ELGIN EXP FA..	0006038 5	-34.1394	19.0228	check	fail	no
FRASERBURG	0113025A2	-31.9160	21.5078	check	fail	no
GEORGE WITFO..	0028776 9	-33.9353	22.4269	check	fail	no
GIANTS CASTL..	0268016AX	-29.2653	29.5225	check	fail	no
GREYTOWN	0270155 9	-29.0830	30.6000	check	fail	no
HERMANUS	0006386A7	-34.4322	19.2247	check	fail	no
JHB BOT TUIN..	0475879 0	-26.1500	28.0000	check	fail	no
JOUBERTINA A..	0031650A4	-33.8283	23.8586	check	fail	no

Table 6—Continued

Station	id	lat	long	record quality	Exposure	included
KNYSNA	0014123 3	-34.0490	23.0815	check	fail	no
KOMATIDRAAI	0520691 2	-25.5145	31.9103	check	fail	no
KROONSTAD	0365398 8	-27.6664	27.3136	check	fail	no
LADISMITH	0046480 6	-33.5007	21.2654	check	fail	no
LAINGSBURG	0045642 0	-33.1908	20.8626	check	fail	no
LYDENBURG	0554816A7	-25.1119	30.4766	check	fail	no
MAKATINI RES..	0411323 2	-27.3941	32.1769	check	fail	no
MALMESBURY	0041388 0	-33.4725	18.7183	check	fail	no
MOKOPANE	0633882 7	-24.2056	29.0117	check	fail	no
MOOI RIVER	0268883 6	-29.2175	30.0025	check	fail	no
MOUNT EDGECO..	0241072 9	-29.7000	31.0500	check	fail	no
NELSPRUIT	0555750 9	-25.5000	30.9166	check	fail	no
NGQURA (COEG..	0035288 9	-33.7967	25.6753	check	fail	no
OUDESTAD	0552581 7	-25.1800	29.3300	check	fail	no
PENNINGTON S..	0211324 3	-30.3997	30.6858	check	fail	no
PIETERMARITZ..	0239698 5	-29.6330	30.4000	check	fail	no
POFADDER	0247668A4	-29.1236	19.3890	check	fail	no
POLOKWANE WO	0677802BX	-23.8572	29.4517	check	fail	no
PORT NOLLOTH	0242644 6	-29.2500	16.8684	check	fail	no
PRETORIA UNI..	0513346 0	-25.7700	28.2000	check	fail	no
RIVERVIEW	0339327 3	-28.4413	32.1777	check	fail	no

Table 6—Continued

Station	id	lat	long	record quality	Exposure	included
RUSTENBURG	0511399 X	-25.6500	27.2330	check	fail	no
SEZELA	0211294 6	-30.3939	30.6658	check	fail	no
SHALEBURN	0237618A9	-29.7869	29.3525	check	fail	no
SOMERSET EAS..	0055363 1	-33.0497	25.7192	check	fail	no
SPRINGBOK WO	0214700B2	-29.6694	17.8789	check	fail	no
SPRINGS	0476762A3	-26.2058	28.4386	check	fail	no
STRAND	0005609 8	-34.1411	18.8489	check	fail	no
SUTHERLAND	0088293A6	-32.3996	20.6629	check	fail	no
THABAZIMBI	0587725CX	-24.5829	27.4170	check	fail	no
THOHOYANDOU ..	0723664 6	-23.0872	30.3844	check	fail	no
TOSCA	0504833 6	-25.8781	23.9662	check	fail	no
TWEE RIVIERE..	0461208 4	-26.4717	20.6104	check	fail	no
TYGERHOEK	0007699A0	-34.1497	19.9031	check	fail	no
TZANEEN-WEST..	0679194 5	-23.7389	30.1211	check	fail	no
VANWYKSVLEI	0193561A8	-30.3498	21.8244	check	fail	no
VENTERSTAD	0173527 7	-30.7770	25.7974	check	fail	no
VIOOLSDRIF -..	0276196 8	-28.7694	17.6234	check	fail	no
VRYHEID	0372527 1	-27.7772	30.7961	check	fail	no
ALL SAINTS A..	0126039 0	-31.6575	28.0431	fail	–	no
ATLANTIS	0020846 4	-33.6078	18.4831	fail	–	no
BEAUFORT-WES	0092081 5	-32.3476	22.5731	fail	–	no

Table 6—Continued

Station	id	lat	long	record quality	Exposure	included
BEN MACDHUI	0177789 6	-30.6634	27.9520	fail	–	no
BOTHAVILLE -..	0399894 4	-27.3661	26.6283	fail	–	no
BOURKE'S LUC..	0594520 0	-24.6724	30.8092	fail	–	no
BRONKHORSTSP..	0514408AX	-25.8086	28.7389	fail	–	no
BUFFELSFONTE..	0148352 8	-31.3694	26.6969	fail	–	no
BUFFELSPOORT..	0511855A9	-25.7500	27.4830	fail	–	no
CAPE COLUMBI..	0060620 9	-32.8278	17.8558	fail	–	no
CAROLINA	0480184E1	-26.0681	30.1186	fail	–	no
CERES AWS	0042563 X	-33.3848	19.3212	fail	–	no
CLANWILLIAM	0084671 0	-32.1768	18.8881	fail	–	no
COFFEE BAY	0128237A0	-31.9653	29.1347	fail	–	no
DASSEN ISLAN..	0040146 5	-33.4325	18.0914	fail	–	no
DOHNE - AGR	0079811A0	-2.5350	27.4675	fail	–	no
EAST LONDON ..	0059541 5	-33.0225	27.8200	fail	–	no
EMERALD DALE..	0238806 6	-29.9431	29.9606	fail	–	no
EXCELSIOR CE..	0063807 2	-32.9636	19.4308	fail	–	no
FRANKFORT - ..	0403886A0	-27.2672	28.4946	fail	–	no
GARIES AWS	0157843 6	-30.5611	17.9874	fail	–	no
GEORGE CLIMA..	0028660B1	-34.0080	22.3817	fail	–	no
GIYANI	0724318 9	-23.3142	30.6856	fail	–	no
GOUGH ISLAND..	0000681 X	-40.3500	-9.8800	fail	–	no

Table 6—Continued

Station	id	lat	long	record quality	Exposure	included
HARTEBEESPOO..	0512554 4	-25.7489	27.8303	fail	–	no
HAWEQUAS	0022071 4	-33.6800	19.0500	fail	–	no
JONKERSHOEK	0021748 6	-33.9678	18.9222	fail	–	no
JOUBERTINA S..	0031619A1	-33.8247	23.8606	fail	–	no
KING SHAKA A..	0241186A6	-29.6017	31.1300	fail	–	no
KING SHAKA I..	0241186 1	-29.6108	31.1239	fail	–	no
KIRSTENBOSCH	0020780 4	-33.9919	18.4328	fail	–	no
KNELLPOORTDA..	0232707 1	-29.7861	26.9031	fail	–	no
LADY GREY - ..	0176413 3	-30.8864	27.2344	fail	–	no
LAMBERTSBAAI..	0083572 8	-32.0351	18.3326	fail	–	no
LANGEBAAN	0040005 6	-32.9725	18.1578	fail	–	no
LANGEBAANWEG..	0061298 8	-32.9725	18.1578	fail	–	no
LANGGEWENS	0041347 X	-33.2764	18.7061	fail	–	no
LANSERIA WO	0512746 7	-25.9397	27.9256	fail	–	no
LEVUBU	0723485A0	-23.0798	30.2800	fail	–	no
LINDLEYSPOOR..	0547359 6	-25.5022	26.6961	fail	–	no
MACHADODORP ..	0517372 0	-25.7161	30.2303	fail	–	no
MARA	0722099 1	-23.1500	29.5700	fail	–	no
MARICO	0546630 3	-25.5000	26.3500	fail	–	no
MARION ISLAN..	0000653 1	-46.8830	37.8670	fail	–	no
MOSSEL BAY	0012221 5	-34.1894	22.1317	fail	–	no

Table 6—Continued

Station	id	lat	long	record quality	Exposure	included
MUKUMBANI TE..	0766715AX	-22.9167	30.4075	fail	–	no
NIEUWOUDTVIL..	0133202A2	-31.3777	19.1217	fail	–	no
ORANIA AWS	0227709A8	-29.8185	24.4112	fail	–	no
OTTOSDAL	0435019AX	-26.8200	26.0200	fail	–	no
PHALABORWA A..	0681266D1	-23.9361	31.1547	fail	–	no
PORT ELIZABE..	0035051 8	-33.8524	25.5357	fail	–	no
PORT ELIZABE..	0035176 X	-33.9425	25.6023	fail	–	no
PORT ELIZABE..	0035178A3	-33.9826	25.6138	fail	–	no
POTGIETERSRU..	0633852 5	-24.2000	28.9830	fail	–	no
PRETORIA - P..	0513284 8	-25.7350	28.1764	fail	–	no
PRETORIA UNI..	0513435A4	-25.7519	28.2583	fail	–	no
PRINS ALBERT..	0048010 5	-33.1734	22.0293	fail	–	no
PUNDA MARIA	0768011A8	-22.6800	31.0200	fail	–	no
REDELINGSHUY..	0084058 X	-32.4814	18.5370	fail	–	no
ROBERTSON	0023708A4	-33.7947	19.9011	fail	–	no
S A ASTRONOM..	0020866 9	-33.9350	18.4775	fail	–	no
SECUNDA	0478330 3	-26.4989	29.1878	fail	–	no
SKUKUZA	0596179 3	-24.9927	31.5880	fail	–	no
SWELLENDAM	0008751A1	-34.0217	20.4372	fail	–	no
TSHANOWA PRI..	0766656 8	-22.9353	30.3667	fail	–	no
TSHIKONDENI	0811780 5	-22.5000	30.9300	fail	–	no

Table 6—Continued

Station	id	lat	long	record quality	Exposure	included
TSHIVHASIE T..	0766628 X	-22.9700	30.3500	fail	–	no
UNIVERSITY O..	0304621 8	-28.8517	31.8531	fail	–	no
VAALHARTS AW..	0360597B0	-27.9575	24.8406	fail	–	no
VENETIA MINE	0808567 2	-22.4517	29.3314	fail	–	no
VESLESKARVET	0000002 4	-71.6800	-2.8500	fail	–	no
VREDE	0405326A9	-27.4228	29.1694	fail	–	no
VREDENDAL	0106880A2	-31.6730	18.4961	fail	–	no
WARDEN - HER..	0369118A3	-27.9611	29.0594	fail	–	no
WATERKLOOF A..	0513379 8	-25.8278	28.2236	fail	–	no
WELLINGTON	0021879 8	-33.6511	19.0067	fail	–	no
WOLMARANSSTA..	0398822 3	-27.2139	25.9783	fail	–	no
WONDERBOOM A..	0513369A5	-25.6561	28.2169	fail	–	no