Evaluating Human-Processed Forecasts Provided to the Aviation Industry in South Africa

by

Quinton Jacobs

Submitted in partial fulfilment of the requirements

for the degree of

MASTER OF SCIENCE

in the

Faculty of Natural and Agricultural Sciences

University of Pretoria

February 2015

Evaluating Human-Processed Forecasts Provided to the Aviation Industry in South Africa

Quinton Jacobs

Supervisor:	Prof. Willem A. Landman
Department:	Department of Geography, Geoinformatics and Meteorology
Faculty:	Faculty of Natural and Agricultural Sciences
University:	University of Pretoria
Degree:	Master of Science

ABSTRACT

The main objective of this study is to determine whether human-forecaster at the main airports of South Africa add value to the raw numerical weather prediction model output when they provide forecast services to the aviation industry. Data set pairs of weather observations made at these airports and terminal aerodrome forecasts were constructed for the three forecast systems: the human-forecaster, persistence forecast and the raw output from the 12 km resolution the Unified Model administered by the South African Weather Service. These three data set pairs were independently evaluated by the in house developed forecast verification system of the South African Weather Service. A Monte Carlo method was used to determine the significance of the verification results obtained from calculating the proportion correct, hit rate, false alarm ratio, critical success index, Heidke skill score and the Pierce skill score of the various forecasts. In general, it was found that the forecaster-adapted forecasts are superior to the raw model output, thus providing the evidence that the aviation industry may benefit most from forecasts routinely issued by South African Weather Service forecasters.

I declare that the thesis that I hereby submit for the degree Master of Science at the University of Pretoria is my own work and has not previously been submitted by me for degree purposes at any other university or institution.

__________________ ____________

SIGNATURE DATE

University of Pretoria, 2015

PREFACE

The aviation industry is an integral part of modern society since it provides a means of transport for people and goods that is both efficient, quick and safe when compared to other means of transport. Moreover, this industry has vastly reduced the time needed for people and goods to be transported over large distances across the globe. There are, however, various factors that influence the safety of air travel when ensuring that goods and passengers reach their destination safely. For example, weather in its various forms contains numerous hazards for the aviation industry. It is the function of aviation weather authorities to provide the aviation industry with quality weather forecast products in order to mitigate possible negative weather-related delays or disasters. These weather forecast products are used in the planning and execution phases of air travel in order to ensure the safety of the users of the aviation industry, be it passengers or cargo, as well as the flight crews on-board.

This research investigates the quality of these weather forecast products provided to the aviation industry in South Africa. The focus of this research study will be the terminal aerodrome forecast, issued by forecasters at the five main airports in South Africa. The aim of the research is to determine whether the forecasts issued by the human-forecaster at the five main airports do, in fact, improve the raw forecast from a numerical weather prediction model and subsequently add value to the forecasts. This investigation will be done by comparing the forecasts of human-forecaster to the raw numerical weather prediction model data for the same airports.

The hypothesis to be tested is:

Terminal aerodrome forecasts issued by South African Weather Service forecasters for major airports in South Africa improve on raw numerical weather prediction model forecasts

To test the hypothesis the following steps needs to be taken:

1. The compilation of data sets comprising observations, terminal aerodrome forecasts produced by human-forecaster, numerical weather prediction model output and persistence forecast as a control.

2. The various forecast sets (forecaster, model and persistence) will be verified by using the operational aviation forecast verification system developed for use in South Africa.

3. The Monte Carlo method will be applied to the data sets to determine the significance of the results obtained by the forecast verification system

4. Case studies extracted from the data sets should be evaluated in order to assess forecasts produced for cases of significant weather events.

This dissertation contains six chapters. Chapter 1 gives a broad discussion of the general weather at the five chosen airports. It also discusses the importance of the aviation industry and its significance for the South African Weather Service, the meteorological authority for aviation in South Africa. Related work done on aviation verification abroad is also discussed in Chapter 1. The threat that numerical weather models hold over forecasters is also briefly discussed. The hypothesis is stated at the end of Chapter 1. Chapter 2 discuss the data used for this study in detail. The construction and interpretation of aviation meteorological observation and forecast data is also discussed in detail. Chapter 3 discusses the methodologies used to construct the data sets. The utility of the forecast verification system and how the Monte Carlo method was applied to the data is also discussed in Chapter 3. All the verification indexes calculated by the verification system are discussed in detail in Chapter 3. The verification indexes used are discussed as well as the motivation presented for using the specific indexes. A skill score for determining whether the human-forecaster out-performs the numerical weather forecast model of the Weather Service and how it is to be implemented is also given in Chapter 3. Chapter 4 focuses on the results resulting from the calculations obtained through the methods discussed in Chapter 4. Chapter 5 will focus on case studies from the data to verify whether the findings of Chapter 4 hold for significant events. The final chapter deals with conclusions and recommendations.

ACKNOWLEDGEMENTS

Many thanks and my appreciation to:

- Firstly, God Almighty, for giving me the strength, ability and perseverance to complete this study.
- My wife, Karien, for making this journey with me; for giving me the support and love I needed to conduct this study and for keeping me grounded.
- Prof. Willem A. Landman (Chief Researcher, Meteorologist, Council for Scientific and Industrial Research, Pretoria) for the guidance given to get this study off the ground; and for his advice and enthusiasm during this study. Without his guidance this study would never have reached fruition.
- Stefanie Landman (Scientist: Prediction Research, South African Weather Service) for providing the numerical weather prediction model data and advice on how to extract it.
- Karin Oxley and Anastasia Demertzis (Librarians, South African Weather Service), who were always happy to help sourcing articles and books.
- Britta Bolzmann (Principal Librarian of National Meteorological Library, German Meteorological Service) for going the extra mile in obtaining an old work instruction for use in my literature study.
- Garth Samson (Climate Liaison Officer, Eastern Cape, Port Elizabeth, South African Weather Service) for help with accessing data and for spurring me on to further my studies.
- Hugh van Niekerk (Regional Manager: Eastern Cape, South African Weather Service) for allowing me to do this research and for his active support.
- Mike Edwards (Retired, South African Weather Service) for getting the ball rolling and enabling me to develop the systems that led to this study.
- South African Weather Service for paying for this study.
- Ricky Woods (English Teacher, Alexander Road High School, Port Elizabeth) for ensuring that the use of the English language is correct.
- All my friends and family, for the support, encouragement and prayers during this study.

LIST OF ABBREVIATIONS

ACSA	Airports Company of South Africa
AMD	Amended Terminal Aerodrome Forecast
ATM	Air Traffic Management
AUTO	Indicator for an automatically generated Aviation Observation
AvianEval	The Aviation Forecast Verification System used in the South African Weather Service
BCFG	Fog patches
BECMG	Becoming, Gradual Change Group
BKN	Broken Cloud – 5 to 7 oktas of cloud
BR	Mist
ß	Frequency Bias
CAVOK	Ceiling and Visibility Okay
CB	Cumulonimbus Cloud
CNL	Canceled Terminal Aerodrome Message
COR	Correction of Aviation Meteorological Message
CSI	Critical Success Index
CSS	Clayton Skill Score
DSS	Doolittle Skill Score
DWD	German Weather Service (*Deutscher Wetterdienst*)
EDS	Extreme Dependency Score
EFS	Ensemble Forecasting System
F	False Alarm Rate

FABL	Bloemfontein Bram Fischer Airport
FACT	Cape Town International Airport
FALE	Durban King Shaka International Airport
FAOR	Johannesburg O R Tambo International Airport
FAPE	Port Elizabeth Airport
FAR	False Alarm Ratio
FEW	Few Cloud – less than 2 oktas of cloud
FG	Fog
FM	From, Change Group which restarts the Meteorological Conditions at given time
FTP	File Transfer Protocol
FvUMSS	Human Forecaster versus Unified Model Skill Score
FZ	Freezing Precipitation
GMT	Greenwich Medial Time GSS Gilbert Skill Score
H	Hit Rate
HSS	Heidke Skill Score
HZ	Haze
ICAO	International Civil Aviation Organization
KT	Knot
LOR	Log of Odd Ratio
MA	Meteorological Authority
METAR	Meteorological Aerodrome Report
METSERV	The data server of the South African Weather Service in Pretoria

NCEP	National Centers of Environment Prediction
NIL	Empty Meteorological Message indicator
NSC	No Significant Cloud
NWP	Numerical Weather Prediction Model
OVC	Overcast Cloud – 8 oktas of cloud
PC	Proportion Correct
PROB	Probability, used with change groups
PSS	Pierce Skill Score
Q	Yule's Q
QNH	Station pressure reduced to sea level
r	Forecast Rate
RA	Rain
RMK	Remark in Aviation Observation
s	Base Rate
SA	South Africa
SAST	South African Standard Time
SAWS	South African Weather Service
SCT	Scattered Cloud – 2 to 4 oktas of cloud
SEDI	Symmetric Extremal Dependence Index
SEDS	Symmetric Extreme Dependency Score
SN	Snow
SHRA	Showers of rain
SHSN	Showers of snow

SKC	Sky Clear
SPECI	Special Meteorological Report
TAF	Terminal Aerodrome Forecast
TCU	Towering Cumulus Cloud
TEMPO	Temporal Change Group
TSRA	Thundershower of rain
TSSN	Thundershower of snow
UM	Unified Model
UK Met Office	United Kingdom Meteorological Office
Δv	Change in wind speed between the actual wind speed and forecasted wind speed
v_{act}	Actual Wind Speed
v_{fc}	Forecasted Wind Speed
VV	Vertical Visibility
WMO	World Meteorological Organization
XAANT	One of the versions of the Unified model as run locally in South Africa by the South African Weather Service

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 TOPOGRAPHY

South Africa (SA) is a large country. It is ranked 25th in the world according to land area (CIA, 2013). Comparatively speaking, France will fit into SA twice and the United States of America is only about eight times bigger than SA (SouthAfrica.info, 2012).

SA is predominantly situated in the subtropics of the Southern Hemisphere, with the northern regions protruding into the tropics. Topographically, SA varies in height from sea-level to 3408 meters above sea-level. The country is dominated by a high plateau, which is surrounded by ragged mountain ranges dropping down to the coastal regions in the south and east, with a smoother drop in the west (CIA, 2013).

1.2 RAINFALL

The central and eastern parts of the country are primarily summer rainfall regions. Precipitation is dominated by summer thunderstorms and they are produced in the warm moist air masses coming down from the tropical belt (Tyson & Preston-Whyte, 2000).

The Western Cape, which comprises the south-western parts of SA, is dominated by winter precipitation. The passing of mid-latitude frontal systems are the primary cause of rainfall in the Western Cape (Tyson & Preston-Whyte, 2000).

Within the borders of SA there exist in the west, arid and semi-arid regions where the average rainfall drops below 200 mm per annum. These arid and semi-arid areas are in the zone between the tropical and mid-latitude rainfall regions (Tyson & Preston-Whyte, 2000).

The region along the south coast of SA receives rainfall from both tropical and mid-latitude origin. Therefore, along the South coast, precipitation occurs throughout the year (Tyson & Preston-Whyte, 2000).

1.3 SIGNIFICANT AND SEVERE WEATHER

Severe weather over SA is widespread and occurs frequently. The main severe weather phenomenon in SA is severe thunderstorms with or without hail. Even without hail, these storms can cause local flooding and lightning. Gale force winds occur frequently along the south and south east coasts and heavy rains occur from tropical or mid-latitude origin.

Severe weather causes various hazards to the aviation industry. Thunderstorms can lead to various problems such as low cloud bases and poor visibility making landing and take-off hazardous. Thunderstorms encountered *en-route*, can lead to icing and turbulence. Operations at the airport can also be affected. For example, thunderstorms cause interruptions in the refuelling of aircraft. In fact, thunderstorms are the primary meteorological reason for delays and diversions at airports.

Heavy rain can lead to low cloud bases and poor visibility. Flooding of the airport can also result. Heavy rain can cause the prevailing conditions to deteriorate to such a degree that minimum instrumentation landing levels are exceeded. These poor visibility and low cloud base conditions result in delays and diversions.

Gale force winds can cause problems during landing and take-off. Wind can become strong enough to exceed the flight specifications of small aircraft. Gale force winds can also create dangerous cross-wind components, creating problems for pilots to land the aircraft.

Another major aviation hazard is turbulence and, more specifically, mountain wave turbulence. There are various cases of light aircraft fatalities owing to mountain wave turbulence (van der Mescht & Eloff, 2013).

1.4 MAJOR AIRPORTS IN SOUTH AFRICA

South African airports are highly regarded in the world, receiving impressive results during assessments. In the World Airport Awards 2013 the three main airports of SA made the top 100 list (Douglas, 2013; Skytrax, 2013). Cape Town International is ranked 22nd; Durban King Shaka International is ranked 26th and Johannesburg O R Tambo International is ranked 28th. The highest ranking American airport was Cincinnati in 30th place. These three airports are also ranked in the top 10 busiest airports on the African Continent, with O R Tambo in first place; Cape Town third and King Shaka in the 10th position.

Figure 1.1 shows the number of passengers (arrivals and departures) for March 2012 to February 2013 at the Airports Company of South Africa (ACSA) owned airports (Airports Company South Afirca, 2013). Figure 1.1 shows that millions of people make use of the airports of ACSA. Johannesburg O R Tambo International Airport (FAOR) handles over 19 million passengers, whereas the smaller regional airports handle fewer than 700 000 people. The coastal airports handle more passengers than the interior airports, with O R Tambo being the exception.

Figure 1.2 displays the movement (the number of aircraft taking-off and landing) for the same airports. Numerous flight training schools use Port Elizabeth Airport (FAPE) as their base of operations. Therefore, the general aviation activity in Port Elizabeth catapults Port Elizabeth to the third spot in movement in SA. King Shaka International only serves commercial aviation. Flight schools and smaller privately operated aircraft, commonly known as general aviation, use the other airports in the Durban area, resulting in fewer aircraft taking off and landing at King Shaka International Airport (Airports Company South Afirca, 2013).

SA has 567 airports, placing it eleventh in the World on the list of most airports in a country (Aneki.com, 2013). Most of these airports are small local airports serving the general aviation industry. Therefore, whilst commercial aviation in SA is significant, general aviation makes up a vast amount of aircraft movement in the South African Airspace.

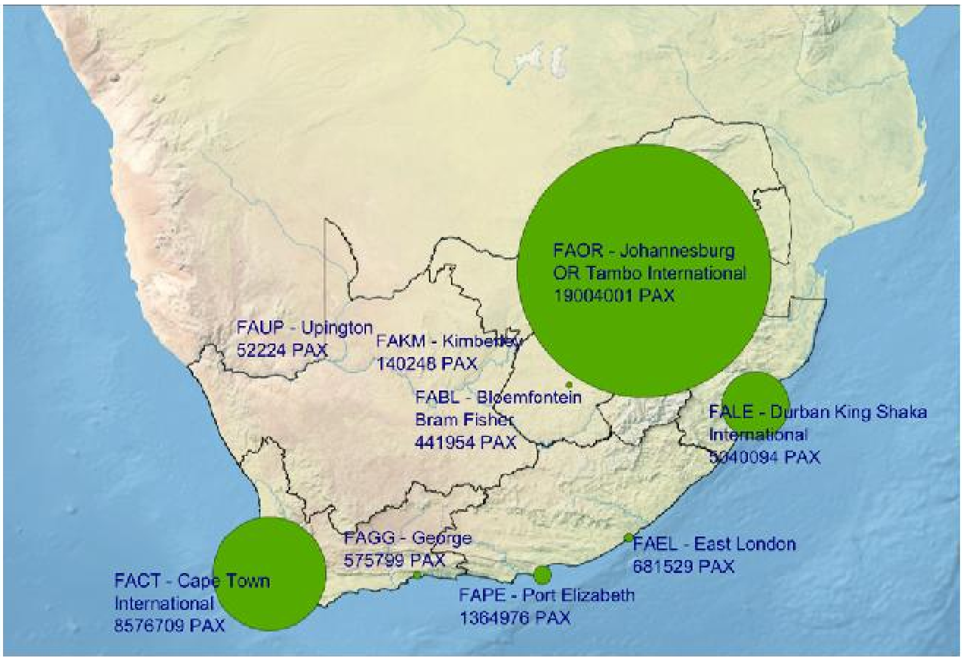

Figure 1.1: ACSA passengers for March 2012 to February 2013

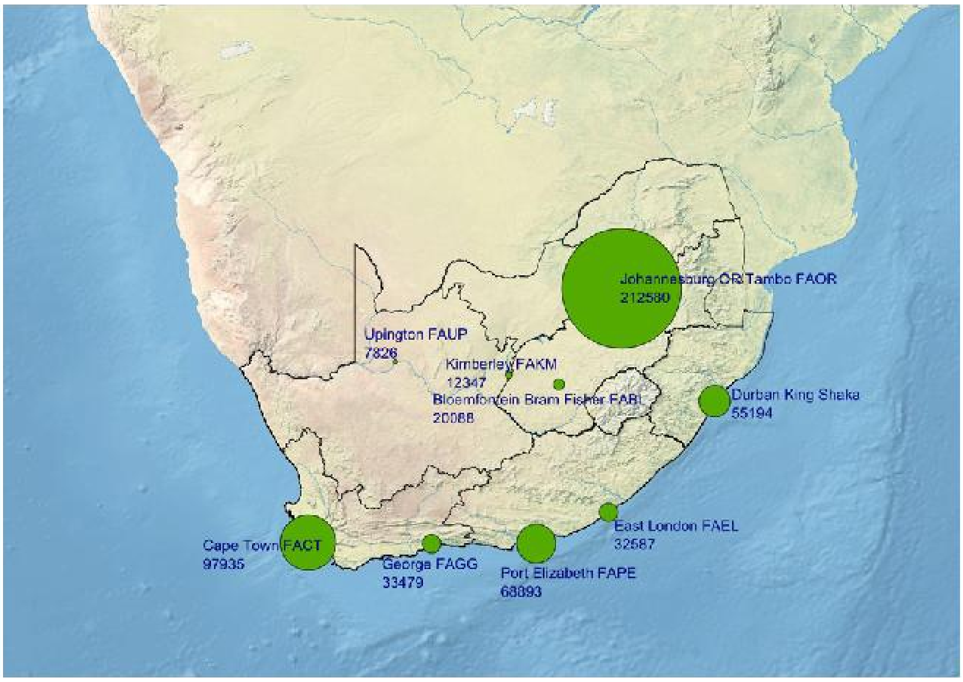

Figure 1.2: ACSA movements March 2011 to February 2012

1.5 AERONAUTICAL WEATHER SUMMARIES FOR THE FIVE MAIN AIRPORTS COMPANY SOUTH AFRICA (ACSA) AIRPORTS

The weather experienced over SA is diverse. Therefore, it is necessary to explore the aeronautical summaries of the aerodromes used in this study to understand what is expected for each region. In this study the five main ACSA airports will be used. The airports are situated in different climatic zones of SA (Figure 1.3). Cape Town International Airport (FACT) is situated in the South Western Cape rainfall region; Port Elizabeth Airport (FAPE) is situated in the South Coast region; Durban King Shaka International Airport (FALE) is located in the KwaZulu-Natal Coast rainfall region; Bloemfontein Bram Fisher Airport (FABL) is located in the Central Interior rainfall region; and Johannesburg O R Tambo International Airport (FAOR) is situated in the North Eastern Interior rainfall regions (Landman, *et al*., 2001).

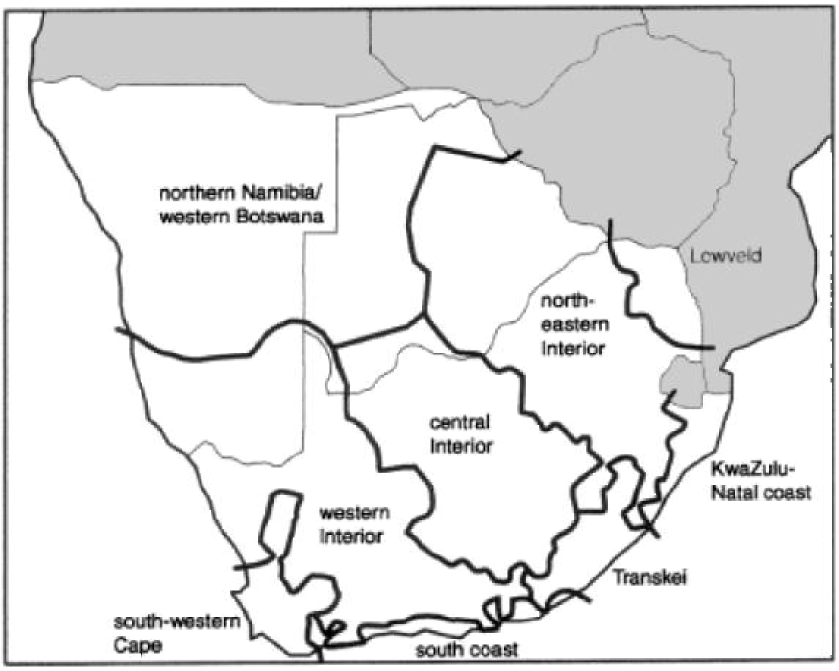

Figure 1.3: The nine rainfall regions in SA (ADAPTED from Landman, *et al*., 2001)

1.5.1 BLOEMFONTEIN BRAM FISHER INTERNATIONAL AIRPORT (FABL)

1.5.1.1 Location

Bloemfontein Bram Fisher International Airport is located 10 km north-east of the city of Bloemfontein. The runway is at a height of 1359 m above mean sea-level. The area around the airport is flat with industry to the west. Recorded observations date back to 5 December 1961 (SAWS, 2012).

1.5.1.2 *Wind speed and direction*

Bloemfontein Bram Fisher Airport experiences highly variable wind in terms of strength and direction as depicted in Figure 1.4. However, north-easterly winds tend to be slightly dominant. Wind speed averages range between 3.7 m/s in autumn to 4.9 m/s in spring. Winds tend to die down in the evening, especially in winter. In summer, the wind tends not to drop, as thunderstorm activity generates gust fronts (SAWS, 2012).

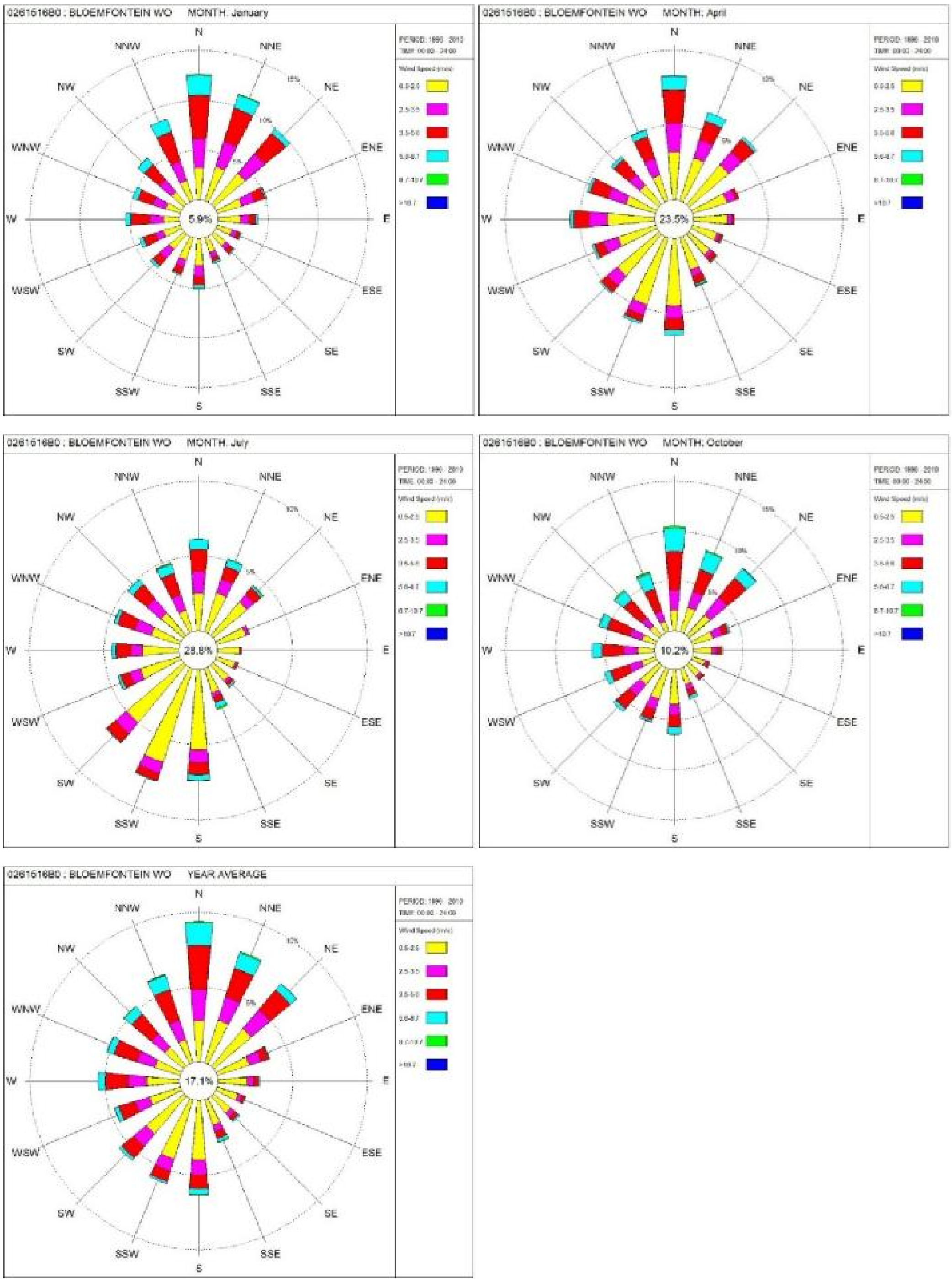

Figure 1.4: Wind Roses Bram Fischer International Airport- January, April, July, October and Average 1996-2010 (SAWS, 2012)

1.5.1.3 Fog, smog and dust storms

Fog is rare – only occurring on about five days per year. The fog tends to occur more in the winter and autumn on clear nights after a rare spell of rain. The type of fog that does occur is normally radiation fog. Winter nights can lead to smog due to the formation of a temperature inversion. This usually happens when the winds are light overnight. Prolonged dry spells can lead to dust storms. These dry spells happen early in the thunderstorm season. Dust can be picked up by the strong gust fronts caused by thunderstorm downdraughts (SAWS, 2012).

1.5.1.4 Low cloud

Low cloud (below 1500 ft) is rare at Bloemfontein Bram Fisher International Airport and is normally only associated with heavy thunderstorms (SAWS, 2012).

1.5.1.5 Precipitation

Precipitation at Bloemfontein Bram Fisher International Airport is predominantly caused by air mass thunderstorms approaching the airport from the north-west. Trough and frontal systems associated thunderstorms can approach from the west or south west. Hail at the airport is quite rare with only about 3% of the thunderstorms reaching the airport contain hail (SAWS, 2012).

1.5.2 CAPE TOWN INTERNATIONAL AIRPORT (FACT)

1.5.2.1 Location

FACT is located in the Cape Flats at a height of 44 m above sea-level, 17 km east south east of the City of Cape Town. The Cape Flats consists mainly of settlements, formal and informal with some industry. FACT is surrounded by various mountain ranges Table Mountain (1085 m) to the west; the Hottentots Holland Mountains (1590 m) to the east and Tygerberg (1843 m) to the north. Observation data at this point date back to 1957 (SAWS, 2012).

1.5.2.2 Wind speed and direction

During summer months southerly winds dominate, while north-westerly and northerly winds are most common in winter as depicted in Figure 1.5. Winds originating from the southern sector are prevalent in autumn and spring, but north-westerly winds are also present. Strong winds are common reaching gale force at times. From April to

May the average wind speed is 4.7 m/s and from November to January 6.5 m/s. November is the month with the highest average wind speed. Winds tend to calm at night both during winter and summer, averaging only 4.8 m/s in the early morning hours. Average wind speed peaks in the late afternoon (8.5 m/s) in Summer, but in early afternoon in winter (5.8 m/s) (SAWS, 2012).

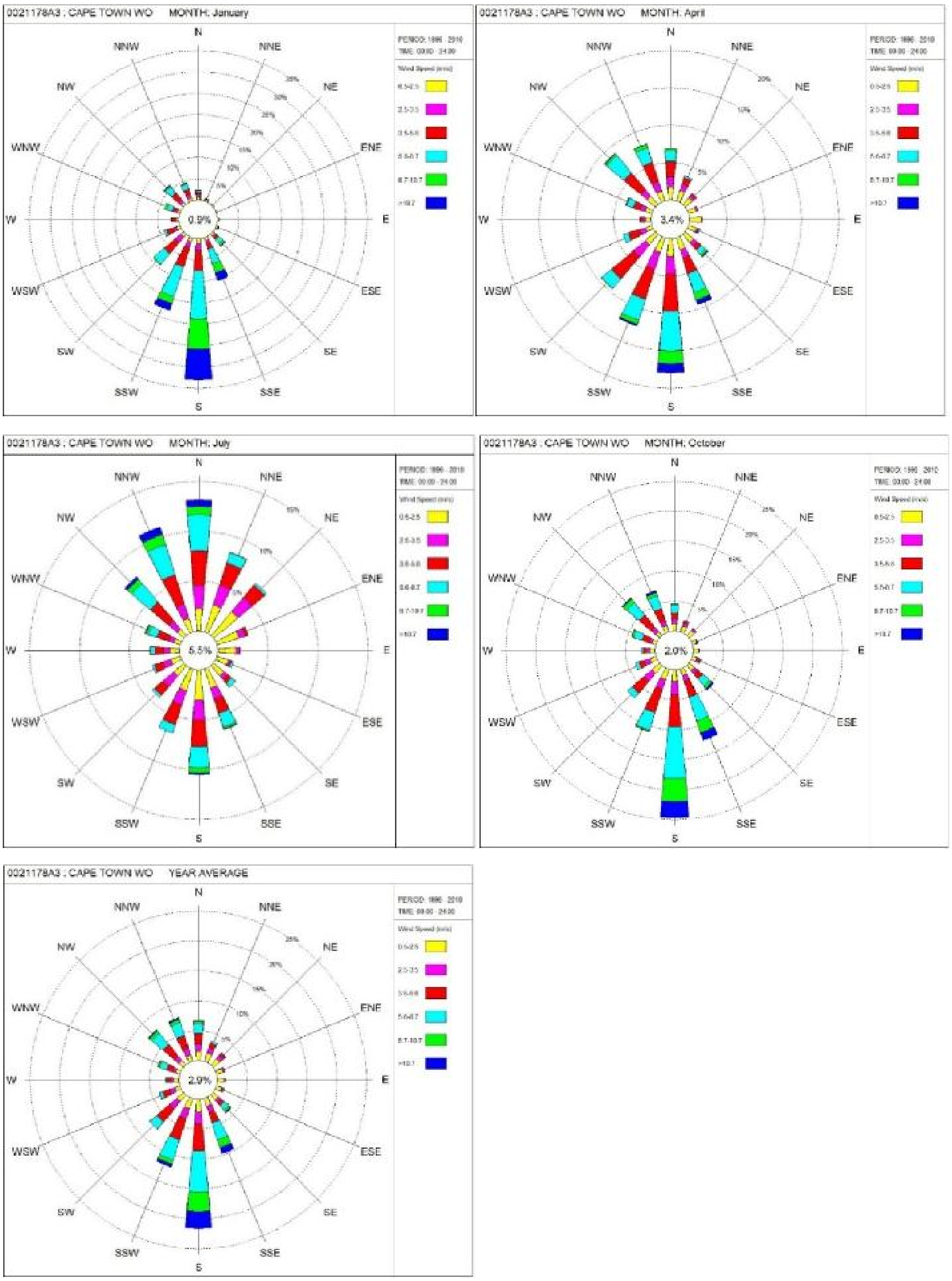

Figure 1.5: Wind Roses Cape Town International Airport- January, April, July, October and Average 1996-2010 (SAWS, 2012)

1.5.2.3 Fog, smog and dust storms

Owing to the Benguela Ocean current along the West coast of SA, advection fog is quite common in the region. FACT averages 45 days of fog per year, with the period of February to July being the most prone. The synoptic conditions mainly responsible for the advection fog are the coastal low, formed owing to the prefrontal flow off the escarpment. These lows, once they move through the region, tend to produce an influx of moist cool sea air over the Cape Flats, resulting in advection fog. Smog is rarely observed owing to the high wind speeds observed throughout the year (SAWS, 2012).

1.5.2.4 Low cloud

Low cloud (below 1500 ft) is mostly associated with frontal passages and fog events. Overcast conditions with cloud bases dropping to below 300 m occur frequently in the north westerly flow preceding the frontal passage. Once the frontal system passes and the wind turns to the south west cloud tends to break up and bases lift. Low cloud bases can also occur when fog starts to lift to become stratus cloud. The cloudiest time of the year is during the rainy season between April and August when coastal lows and frontal systems are more frequent (SAWS, 2012).

1.5.2.5 Precipitation

Precipitation is mainly in the form of drizzle or widespread rain, owing to the passing coastal lows or frontal passage. After the frontal passage showers can occur from the cold air cumulus cloud in the cold air behind the frontal system. Warm fronts from the frontal systems do not affect the airport as they normally pass south of southern Africa. The upper atmospheric trough of the associated frontal system occasionally intensifies and becomes a closed-off low, which is called a cut-off low. These systems generate heavy rains which severely impact operations at the airport. Thunderstorms are rare at FACT. Thunderstorms usually result from frontal passage, but only for about seven days per annum (SAWS, 2012).

1.5.3 OR TAMBO INTERNATIONAL AIRPORT (FAOR)

1.5.3.1 Location

O R Tambo International Airport is located 20 km to the east-north-east of the heart of Johannesburg and 40 km south of Pretoria. Both runways at OR Tambo International Airport are more than 1676 m above mean sea level. The airport is surrounded by industry and skyscrapers with the highest terrain to the south of the airport.

Observations at O R Tambo International Airport have been conducted since its opening on 1 September 1953 (SAWS, 2012).

1.5.3.2 Wind speed and direction

North-westerly winds occur all year round, but different seasons result in other wind directions becoming more dominant. The ridging of high pressure systems in summer results in easterly and south-easterly wind becoming more frequent. In winter the passage of frontal system and troughs force the south westerly winds to be more frequent. The two transitional seasons are not affected by the summer or winter systems as frequently as in the two main seasons. Therefore, the north westerly tends to dominate. Wind usually starts as a north-easterly wind in the early morning. During the course of the day wind backs through the north to north-westerly. This backing of the wind is more pronounced in summer than in winter (SAWS, 2012).

The wind speed has a clear diurnal effect as depicted in Figure 1.6. Mostly calm at night, increasing in strength as the day progresses and dying down in the evening. Summer thundershowers can cause gust fronts in the late afternoon and evening breaking this pattern. Average wind speed varies from 3.3 m/s in March and April to 4.6 m/s in October. Midday winds reach 4.2 m/s on average, but are somewhat stronger in winter (SAWS, 2012).

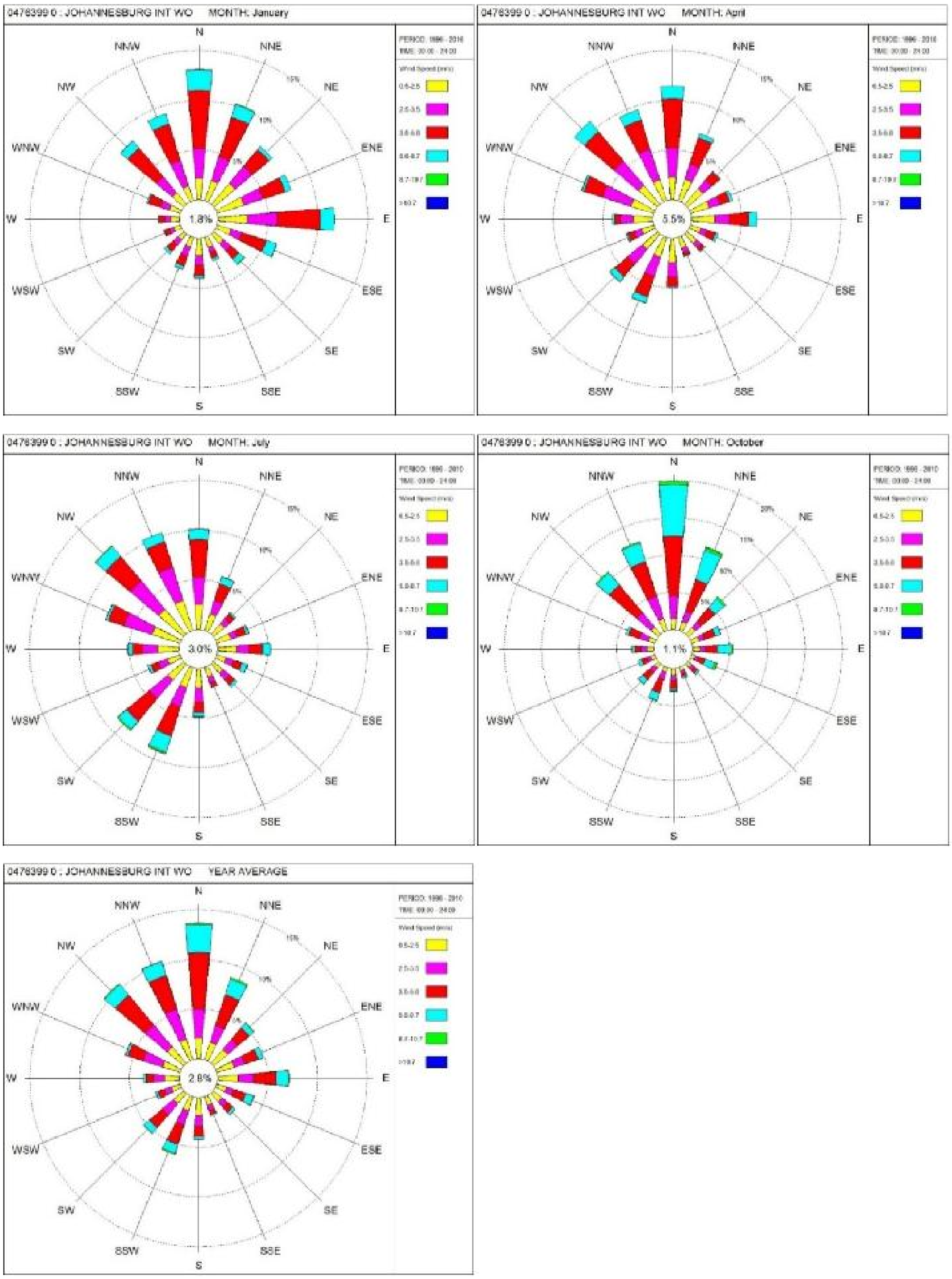

Figure 1.6: Wind Roses Johannesburg OR Tambo International Airport- January, April, July, October and Average 1996-2010 (SAWS, 2012)

1.5.3.3 Fog, smog and dust storms

Visibility at the airport is generally good. Fog does occur around two to three days per month; March and April having four to five days of fog on average. The fog occurring at O R Tambo is both radiation and advection fog. Radiation fog generally occurs after thunderstorm activity and the sky clears with calm winds. Advection fog is usually due to the ridging high pressures bringing moist air through the Lowveld from the east or north east of the airport. Due to the high altitude of OR Tambo International Airport, the resulting cloud from the subtropical air mass usually reaches the airport as fog (SAWS, 2012).

Smog is quite common in winter months owing to the vast amounts of industry around and temperature inversions developing from clear nights. Visibility usually drops below 5000 m as a result of pollutants trapped under these inversions. For the most part, such inversions are removed through surface heating. Therefore, smog generally occurs in the evening until midmorning (SAWS, 2012).

Dust and sand storms can occur at the airport, but are rare. Strong winds in spring after a dry winter or gust fronts emanating from a squall line of thunderstorm after a dry spell, can lead to visibility dropping below 5000 m (SAWS, 2012).

1.5.3.4 Low cloud

Low cloud (below 1500 ft) usually occurs in the morning and improves as the day progresses. The occurrence of low cloud is associated with winds ranging from south-east all the way to the north-west. Frontal systems can lead to low cloud associated with a south westerly wind. Low cloud mostly occurs in the summer months, while winter months are dominated by clear conditions (SAWS, 2012).

1.5.3.5 Precipitation

O R Tambo Airport experiences about 80 thunderstorms per year and 90% of these storms occur in spring, summer and early autumn. Thunderstorms experienced at the airport are usually associated with the convergence zone of an eastward moving through-line or cut-off low in the upper atmosphere. Thunderstorms are generally experienced in the afternoon into the evenings. 6% of thunderstorms reaching O R Tambo produce hail, mainly between the months of October to December (SAWS, 2012).

O R Tambo is located in a summer rainfall region, receiving the bulk of its rain in January and the least in August. Spring and autumn also receive a fair amount of precipitation, with somewhat more in spring. December is the month when the rainfall intensity is the highest (SAWS, 2012).

1.5.4 DURBAN INTERNATION AIRPORT (FADN)

King Shaka International Airport became operational in May 2010 (King Shaka International, 2013). This airport has not been operational for long enough to establish an aeronautical summary. Although Durban International Airport is situated about 45 km to the south-west of King Shaka International, historical data from the former are used in the aeronautical summary presented here. Whilst the weather experienced at the two airports will differ somewhat, they are located close enough to one another so that the weather patterns affecting the airports should be similar even though location may cause slight differences.

1.5.4.1 Location

Durban International is located 15 km from the city business district of Durban in the southern industrial basin. The airport, city and harbour are connected by the N2 highway. The airport is only 10 m above mean sea-level. The recorded observations are from 1 September 1953 until May 2010 when the observations moved to King Shaka International. Automatic weather information is still recorded at the airport (SAWS, 2012).

1.5.4.2 Wind speed and location

The wind generally blows parallel to the coast, with south-westerly and north-easterly winds being dominant and occurring around the same frequency throughout the year as depicted in Figure 1.7. Wind variation is also very small averaging 4.3 m/s in June up to 5.5 m/s in October. A diurnal effect of the wind is also quite noticeable. Wind tends to be calmer at night and early morning averaging 3.9 m/s, but increasing during the day reaching its maximum speed in the late afternoon with an average of about 6.3 m/s (SAWS, 2012).

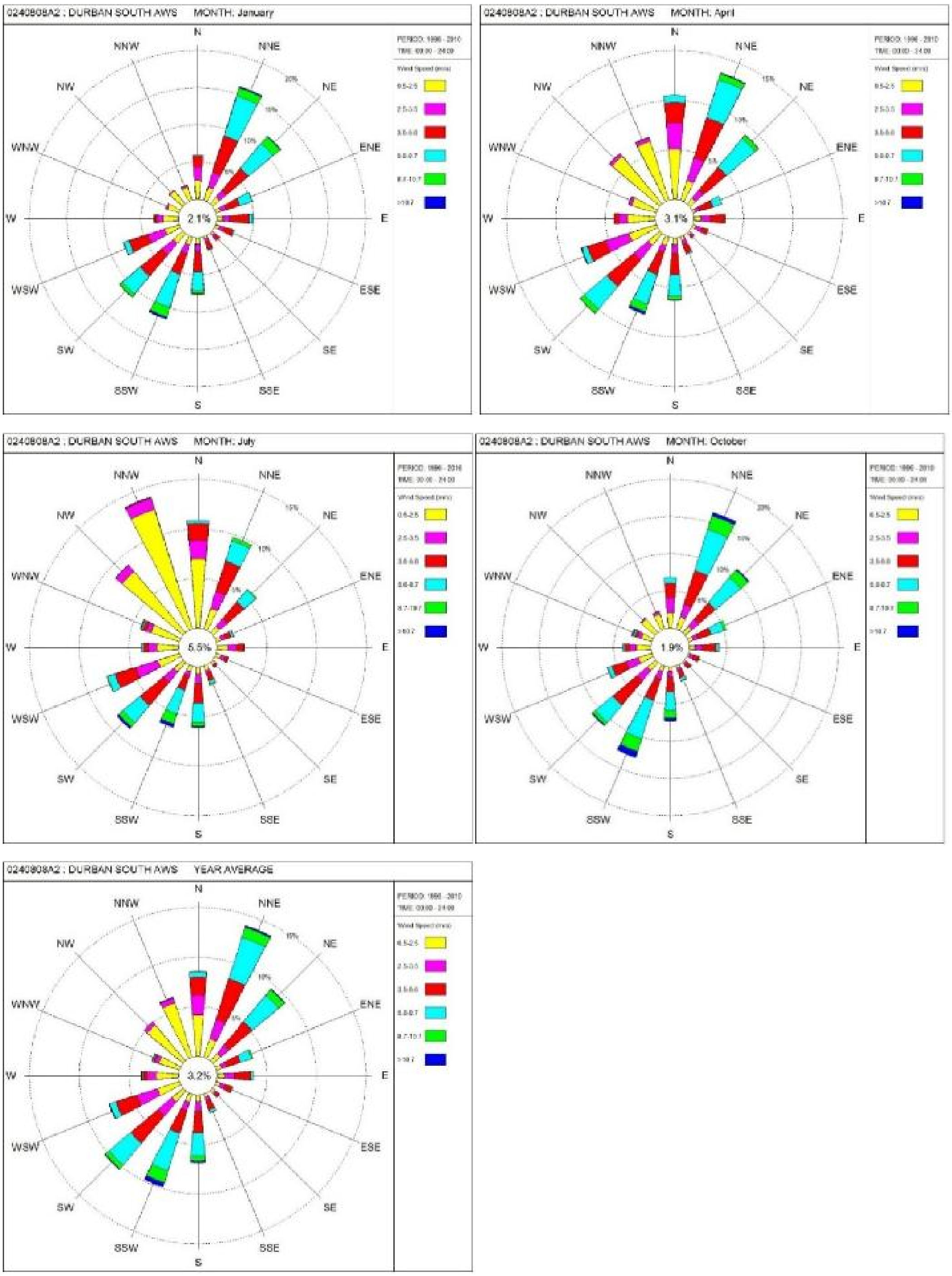

Figure 1.7: Wind Roses Durban International Airport- January, April, July, October and Average 1996-2010 (SAWS, 2012)

1.5.4.3 Fog, smog and dust storms

Fog occurs about two to three day per year and is usually shot lived and patchy. Shallow ground mist can occur overnight and in the early hours of the morning. Smog occurs in winter when temperature inversion traps the pollutants from the surrounding industries. Visibility is generally good when the wind is blowing from the south to south-west (SAWS, 2012).

1.5.4.4 Low cloud

Durban and the surrounding area is wet and humid with annual precipitation of about 1400mm. Owing to the high humidity brought on by the warm Agulhas current, only about 45% of the daylight hours of summer experience sunshine.

The occurrence of low cloud (below 1500 ft) is associated with certain types of rain-bearing synoptic systems. Summer is the season when Durban International has the highest frequency of low cloud. Coastal low and frontal passages also lead to low clouds rolling in off the sea. The ridging high pressure systems behind these frontal systems also enhance onshore flow, resulting in low cloud. Low cloud is associated with late afternoons, evenings and early mornings (SAWS, 2012).

1.5.4.5 Precipitation

Summertime precipitation is primarily from thunderstorms drifting out to sea from the interior. Tropical depressions and tropical cyclones can also reach the kwaZulu Natal coastline. Coastal low and frontal passages can bring in some maritime moisture, which leads to occasional showers. Winter is mostly associated with fair weather conditions, but the passage of coastal lows and the associated frontal systems can result in rain showers. Heavy rains can occur when a well-developed upper trough or cut-off lows move into the region. Thunderstorms tend to develop over the warm Agulhas current in winter, which can drift onto the coast and adjacent interior (SAWS, 2012).

Thunderstorms are quite common in Durban. Durban International averages about 38 thunderstorms in a year, with March and November averaging six thunderstorms a month. Thunderstorms occur mainly in summer, but one day with thunderstorms per winter month is common. Hail is a rare occurrence, as the storm would normally have hailed out by the time it reaches the surface. When hail does occur it is usually of small size. On average, one day per year can contain hail. (SAWS, 2012).

1.5.5 PORT ELIZABETH AIRPORT (FAPE)

1.5.5.1 Location

FAPE is situated in the Walmer suburb of Port Elizabeth, which is in the southern area of the city. The airport is west of Algoa Bay on the Cape Recife Peninsula 7.5 km from the coastline. The runway is 59 m above mean sea-level and the surrounding area is quite flat. Within 50 km of the airport the highest ground reaches 610 m. Skyscrapers are located to the west of the airport and are mostly residential flats and hotels, before moving into the central business area of Port Elizabeth. FAPE has the longest observational record of all the airports in this study, dating back to 1937 (SAWS, 2012).

1.5.5.2 Wind speed and direction

Wind near the airport tends to blow parallel to the coastline as depicted in Figure 1.8. From autumn to spring west to south westerly winds dominate. These south-westerly winds are due to coastal low and frontal system passages. In summer north-east to easterly winds are dominant. These easterly winds are due to the easterly troughs developing over the interior in summer (SAWS, 2012).

Gale force winds are a frequent occurrence in Port Elizabeth. Spring to early summer tends to have the strongest winds averaging 6.2 m/s from October to December. May to July wind speeds drop to 4.5 to 5.0 m/s. Winds at night are lighter than in daytime averaging 4.0 to 5.0 m/s. Maximum wind speeds are generally reached by mid-afternoon averaging 8.2 m/s in summer and 6.0 m/s in winter (SAWS, 2012).

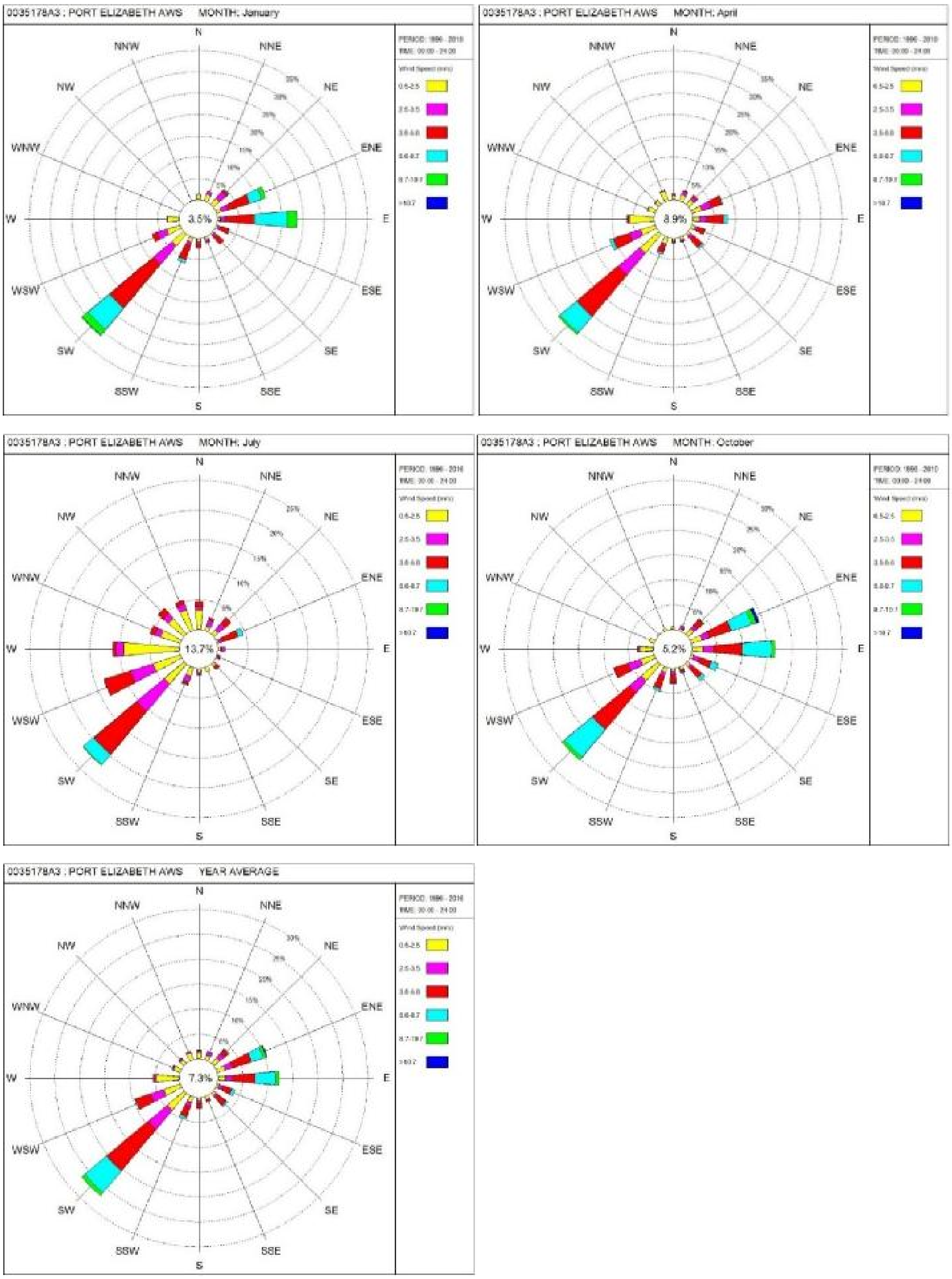

Figure 1.8: Wind Roses Port Elizabeth Airport- January, April, July, October and Average 1996-2010 (SAWS, 2012)

1.5.5.3 Fog, smog and dust storms

The reduction of visibility is mostly related to haze and ground fog. These events occur owing to the north-easterly flow in summer; when the dew point temperatures are quite high. Fog events usually clear by mid-morning. Smog can occur in winter. The harbour and industry do contribute to the pollutants, but the Walmer Township with its wood fires is usually the main contributor. In summer, light rain or drizzle can also reduce visibility significantly, due to the easterly flow resulting from ridging high pressure systems (SAWS, 2012).

1.5.5.4 Low cloud

Low cloud (below 1500 ft) occurs throughout the year. Winter and early spring yields the highest number of clear days. Low cloud usually occurs overnight and in early morning, lifting during the day. The passage of coastal lows and frontal systems tends to yield low cloud coming in when the wind veers to the south-west (SAWS, 2012).

1.5.5.5 Precipitation

Port Elizabeth is part of the Southern Cape climate region. This region receives rain all year round, although summer is the driest season (SAWS, 2012). Thunderstorms do occur at the airport at times in summer for around two days per month. The Groot Winterhoek mountain ridge to the north of the airport usually diverts the thunderstorms to the east. Thus, whilst the region does receive thunderstorms, FAPE usually misses out on these storms. Since the airport is close to the coast, thundershowers that do occasionally reach the airport have usually hailed out and therefore hail is quite rare (SAWS, 2012).

Precipitation not originating from thunderstorms is also quite common since Port Elizabeth is also affected by mid-latitude systems. When a frontal system approaches, the preceding coastal low can cause enough moisture to be fed onto the coastline to result in rain and showers. Port Elizabeth rarely receives rainfall from the passage of the frontal system itself, as most of the rain will already have fallen. After the passage of the frontal system the ridging high brings cold moist air resulting in rain and showers. Cold air cumulus developing in the unstable cold air behind the frontal system is also fed onto the coastal region resulting in showers at FAPE. Well-developed upper troughs and cut-off lows can cause heavy rains, resulting in poor visibility, low cloud ceiling and flooding of the runway (SAWS, 2012).

1.6 THE IMPORTANCE OF AVIATION TO THE SOUTH AFRICAN WEATHER SERVICE

As mentioned previously, SA has many airports with high volumes of aircraft movement. International Aviation allows a Meteorological Authority to charge for services rendered to aviation as it is considered a commercial service and not a public good service (WMO, 2007). In SA this is regulated income for the South African Weather Service (SAWS). Figure 1.9 shows the aviation income as a percentage of total operational income for the years 2008, 2010 and 2013 (SAWS, 2008; 2010 and 2013).

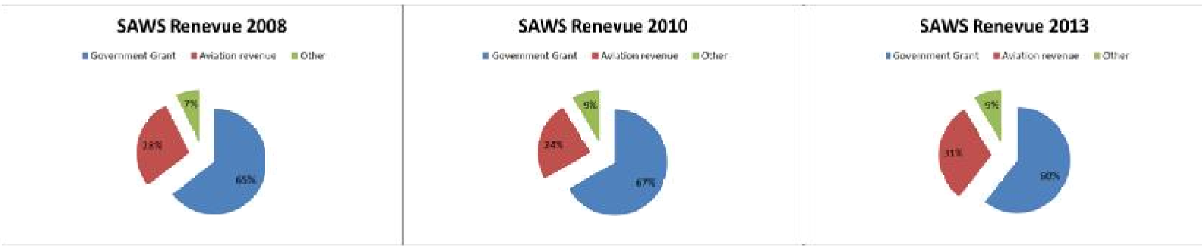

Figure 1.9: SAWS Revenue 2008, 2010 and 2013

The Government Grant constitutes 60 – 67% of the total income for SAWS. This income is to be used for the public good services SAWS provides to the citizens of SA. In 2001, when the South African Weather Bureau became SAWS, this government agency became a parastatal agency. Becoming a parastatal agency enabled SAWS to generate its own income in order to subsidise the shortfall of the Government Grant. From Figure 1.9 it is quite clear that aviation income is the primary commercial income of SAWS. Since the aviation industry pays for the service, SAWS has to determine that the quality of services it provide meet international standards and that these standards are maintained. Verification methods need to be applied on these products to determine its quality and areas of improvement. For the 2012/3 operational year, aviation forecasts exceeded the targets set 90% for Terminal Aerodrome Forecasts (TAF) and 92% for Trend (landing) forecasts (SAWS, 2013).

1.7 TERMINAL AERODROME FORECAST VERIFICATION METHODS USED ABROAD

TAF verification has increased in importance over the last twenty years. Meteorological Authorities (MA) are required to provide evidence that their commercial forecasts for the aviation industry are reliable and can be used with confidence by pilots of commercial airlines. As development in the field of aviation forecast verification progressed, MA finally have the tools at their disposal to verify TAF messages effectively. The verification of TAF messages ultimately results in safer air traffic around the globe.

The TAF verification method for South African airports to be presented here was developed, using only the International Civil Aviation Organization (ICAO) criteria for its basis. This verification method will be briefly introduced in the next section and is discussed detail in the next chapter.

Implementing an automated TAF verification system has been an active aim of MA across the globe. Only a few articles have been published on this topic (Gordon, 1993; Balzer, 1995; Mahringer, 2008). Verification methodologies applied to TAF messages found around the world will be investigated in chronological order.

1.7.1 NEW ZEELAND

The methodology employed in New Zealand is to obtain hourly values for the following elements:

- Wind Speed
- Wind Direction
- Visibility
- Ceiling (lowest cloud layer with 4 oktas more coverage, but limited to a height of 2500 ft (760 m) (feet are used as cloud ceiling is reported in feet rather than metres)

The actual observational data are used for the verification purposes. These observational messages are called Metereological Aerodrome Report (METAR), which are routine aviation observational reports. Special meteorological report (SPECI) messages are the same as METAR messages, but are issued when significant changes have occurred since the previous METAR of SPECI message issued at the aerodrome.

Two sets of values are obtained. The first is the prevailing values of the TAF message at the given time period. The second is any alternative value obtained form change groups valid for the time period (Gordon, 1993). There are two types of change groups:

1. Temporal change (TEMPO)

2. Becoming, gradual change (BECMG)

These hourly values are used to determine minimum values of ceiling and visibility for three-hour blocks or periods. TEMPO values are only considered if the probability is greater than 50% and they are valid for at least one hour of the block. The minimum values obtained for the blocks are then compared to all the METAR and SPECI (actual observations in the three-hour period) and a contingency table is constructed. Persistence is used for the period of the TAF between the issue time and the start of the TAF (Gordon, 1993). In this context persistence implies that the weather conditions between the issue time and start time is assumed to stay unchanged.

For wind speed a root mean square error is calculated using the TAF wind speed and actual wind speed. Wind direction is handled similarly, with the comparison range taken as -180° to 180°. Again a root mean square error is calculated (Gordon, 1993).

The values obtained from the process discussed above are compared to the criteria cited in Annex 3 of ICAO 11th Edition, as it was the edition in use at the time. The results aimed for were the operationally desirable accuracy of aerodrome forecasts (TAFs) as given in the mentioned edition of Annex 3 of ICAO (Gordon, 1993).

Whilst the methodology proposed by Gordon (1993) was proposed more than twenty years ago, it does form the basis for the TAF verification systems various other countries, namely Denmark (Hilden, *et al.*, 1998), Hong Kong (Chan, 2000) and the Netherlands (Jacobs & Maat, 2004), to name some examples. No further published works could be found to determine if this methodology is still being used in New Zealand.

1.7.2 GERMANY

In 1994 the German Weather Service (DWD) issued its work instruction regarding TAF verification. This document would in later years be superseded by the methodology of Austro Control discussed in the next section. The German approach was more complex than that of the Meteorological Service of New Zealand Limited. The DWD used a system of weighted values and handled scalar and vector components differently (Balzer, 1995).

For scalar elements such as cloud base, cloud amount and visibility, the forecast hourly values are weighted according to the following:

- Initial values are weighted at 100%
- Values following the becoming (BECMG) and from (FM) change groups are weighted at 100%

- Temporal (TEMPO) groups are weighted at 40% if no probability (PROB) group precede it. Otherwise, TEMPO groups are weighted at the given PROB percentage.

Wind direction values are calculated by determining the vectorial difference between the forecasted and observed winds. Two special cases are considered. If the wind speed of both the forecast and actual are less or equal to 5 kt the error is set to zero. If an error needs to be calculated from variable wind and a directional wind, the errors are calculated using the directional wind for the variable wind and the wind direction opposite the directional wind. An average of the two errors are then used (Balzer, 1995).

1.7.3 AUSTRO CONTROL

The Austro Control method was the successor of the German (DWD) methodology discussed in Section 1.7.2. This methodology sets out to create a system using all the available aerodrome data.

Austro Control argued that various methods exist to verify TAF forecasts ranging from time interval blocks to continuous value verification. The main area of contention is the fact that ICAO Annex 3 (in this case the 15th Edition of 2004) states two different sets of requirements for TAF verification – the one being the criteria which govern the composition of TAF messages; the other is given in Appendix B of the ICAO Annex 3, stating tolerances for the TAF components, which can contradict the criteria used to write TAF messages (Mahringer, 2008).

Austro Control identified three applications for TAF verification (Mahringer, 2008):

1. Feedback to management – identifying the quality of the forecast and its trend over time.
2. Feedback to forecasters – identifying the areas where the forecasts need to improve and where the forecasters need to improve or for individual evaluation.
3. Feedback to customers – how TAF forecasts can be used most effectively by the customer.

Austro Control see a TAF as a collection of forecast periods rather than forecasts for any given point in time. The use of the change groups, *i.e*. TEMPO, BECMG, PROB and PROB TEMPO has the shortest period of one hour. TEMPO implies a temporal change in conditions valid for half of the period; BECMG stands for becoming, suggesting a gradual change and PROB is probability. The only exception is the FM (from) group which is time-specific to the minute. This fact is evident on the criteria in ICAO Annex 3, which state ranges considered to have similar effects on flight operations. Thus, any given hour in a TAF message, contains ranges of values within the criteria stipulated in ICAO Annex 3 (Mahringer, 2008).

Verification is done on the maximum and minimum values valid for the given hour in the TAF message. These minimum and maximum values are used to eliminate the need to quantify the PROB groups, as the range of values is already defined. Customer-based verification PROB and PROB TEMPO groups can be excluded as this is operationally applied by many users of the TAF message. TAF messages are also checked for syntax errors (Mahringer, 2008).

As was the case in Germany, which preceded this methodology, values are calculated for wind, visibility, significant weather and ceiling. These variables in the TAF message are verified using ICAO Annex 3 amendment criteria and regulations that may apply locally. ICAO Annex 3 Appendix B is not used. Operationally, TAF messages are used for planning and therefore the thresholds are more meaningful to the clients. All variables except wind direction are evaluated and entered into two separate contingency tables (Mahringer, 2008; Jolliffe & Stephenson, 2012). The highest and lowest values are evaluated against the two corresponding contingency tables (Mahringer, 2008).

Significant weather is grouped into similar weather phenomena to ensure proper statistical results. Ceiling and visibility are conducted in a similar manner using the thresholds given in ICAO Annex 3 as the ranges. Wind speed is evaluated using the continuous sensor data as wind is quite variable and the wind speed recorded is the 10 minute average wind. Austro Control only evaluates significant variations of wind direction when the wind speed exceeds 7 kt. A 30° threshold is used for directional variance. Thus, all winds within the 30° threshold are considered to be correct. A VRB (variable) wind is considered to be the expected wind and the wind direction 180° of the prevailing wind. The greatest error of the two is then used (Mahringer, 2008).

Austro Control MET and Air Traffic Management (ATM) have set operationally significant wind speed values in accordance to ICAO Annex 3 Appendix 5. The values agreed upon were 7, 15, 25, 35, 45 and 55 kt. Wind speed is verified the same as the others. For the verification of wind gusts the hourly wind maximums are used from the continuous wind data. Austro Control only verifies wind gusts above 30 kt (Mahringer, 2008).

The problem arising from the change groups is solved using the minimum and maximum value approach. Since these values will cover the range of change expected within the time period, change groups are covered. This methodology therefore does not distinguish between different types of change groups resulting in all values being considered (Mahringer, 2008).

Comparing the resulting calculations from the pairs of contingency tables, conclusions can be made regarding the quality of the forecasts for given events. The timing issues of these events are not always clear, but trends in forecasting are easily identified (Mahringer, 2008).

1.8 TAF VERIFICATION IN SOUTH AFRICA

In SA, there is no documented evidence of TAF verification. However, an aviation product verification system has since been conceived and is now being tested in the operational forecast environment of SAWS. The methodology of this system will be discussed in the methodology chapter, as this system will be used to conduct the research for this dissertation.

1.9 NUMERICAL WEATHER MODEL PREDICTION IN SOUTH AFRICA

Numerical weather prediction models (NWP) are invaluable tools for operational weather forecasters. NWP models are used in forecasting to provide the forecasters with possible future states of the atmosphere based on numerical solutions of atmospheric motion. These future states of the atmosphere provide the forecaster with a first guess for forecasting in the short and long term, when just relying on the actual data no longer applies.

SAWS have been using the National Centers of Environmental Prediction (NCEP) Ensemble Forecasting System (EFS) operationally since 2000. The NCEP model is used in short to midrange forecasts. Being an ensemble of model forecasts it is considered to be an efficient way of determining the probability of the forecast. The ability to determine probability is due to the nature of ensemble models – a combination of NWP models all giving their outputs for the given variable. The more models agree, the higher the probability and confidence of the total ensemble. The resolution of these models are 2˚ and 1˚ (Tennant, *et al*., 2007).

A different model is required for short range and now-casting use. The model that SAWS adapted for South African conditions and use was the United Kingdom Meteorological Office (UK Met Office) developed Unified Model (UM). A global version of the model is run at the UK Met Office four times a day with a horizontal resolution of 40 km. These model runs are used as the boundary conditions for any regional run versions of the UM, in order to produce 12 km horizontal resolution NMP data over Southern Africa. In SA, SAWS is running the UM for the Southern African Region with various parameters since 2006 (Landman, *et al*., 2012).

SAWS runs the UM in three basic configurations (Landman, *et al*., 2012):

- 12 km horizontal resolution model run without data assimilation. It has 38 vertical levels and is run off the 18Z UM global model run for initialization. It includes the Southern African subcontinent with a large chunk of the surrounding oceans. This model produces output for every hour up to 48 hours.
- 12 km horizontal resolution with data assimilation. Same as above, but it uses a 3D variation data assimilation. Data assimilation is a statistical method of combining observational data with the corresponding first guess field of the UM. This is done every 6 hours. This method ensures that the first guess field is corrected according to the observations. This UM run is very process intensive, but is expected to result in improved forecasts.
- 15 km horizontal resolution model run with no data assimilation. This run is computed on a smaller area, covering only SA between 22°S to 35°S and 15°E to 34°E. With the lower resolution and smaller area, this UM run is less taxing than the others and is therefore available to the forecasters earlier in the day.

These models have been updated quite frequently with the latest update released in January 2011. The UM is ideally suited for aviation forecasts as it has hourly values and is valid for 48 hours, which is sufficient for the compilation of TAF messages (Landman, *et al*., 2012). The iteration of the UM used in this dissertation is the XAANT run.

1.10 THE THREAT TO HUMAN-FORECASTER BEING REPLACED BY RAW NUMERICAL WEATHER PREDICTION MODELS

Over the last decade, with the increase of computation power of modern computers and supercomputers, NWP modelling has improved by leaps and bounds. With the faster and more powerful computers of today, the resolution of NWP has increased dramatically. In SA the UM is running at a 12 km horizontal resolution (Landman, *et al.*, 2012). The advancement of model resolution has contributed, along with much research and development in NWP, to increase the accuracy of the forecasts issued by NWP models.

With the increased reliability of NWP and its increased accuracy, the internet has spouted numerous websites using NWP data to present weather forecasts to the masses. Most of these websites contain raw model data products with no human intervention, but since the models are becoming more reliable and accurate with time these forecasts are sufficient for general use. These NWP raw model forecasts add to the stress of the forecasting profession as the weather forecaster is being threatened by redundancy and may in future only have a supervisory role in the field of forecasting.

1.11 THE HYPOTHESIS

As previously stated in this chapter, aviation forecast products are important both to the aviation industry and to SAWS. The aviation industry requires reliable forecasts to operate and transfer goods and passengers safely and it is dependent on aviation forecasts to plan the flights – fuel requirements, time to destination, alternatives, etc. SAWS as the Meteorological Authority of SA, is allowed to charge the aviation industry for services rendered. These services account for almost all the commercial income for SAWS and over a quarter of the total operational income of SAWS, second only to the Government Grant for Public Good (SAWS, 2008; SAWS, 2010; SAWS, 2013).

The questions that arise are the following:

- Are the forecast products provided by SAWS of good quality?
- Do the forecast products of SAWS meet International Standards?
- How do alternative forecast strategies compare with aviation forecasting?
- How do the human-forecaster compare with alternative forecast strategies?

According to the 2012/3 SAWS Annual Report, the Aviation Forecasts have performed better than the targets set out as stated in section 1.6 (SAWS, 2013). These targets are higher than the targets expected by ICAO (ICAO, 2010). Thus, the first two questions seem to be answered by the verification of the aviation products of human-forecaster. The other questions need to be answered by verification of alternative forecast strategies.

The hypothesis for this study is therefore as follows:

Terminal aerodrome forecasts issued by South African Weather Service forecasters for major airports in South Africa improve on raw numerical weather prediction model forecasts.

The data and the methodologies used to test the hypothesis will be discussed in detail in the next two chapters. The data sets will be used to compare the performance of the different forecast strategies and are not intended to be used to compare different forecasting teams against each other. The five airports are all located in different climatic zones and therefore are not objectively comparable against one another.

1.12 CHAPTER SUMMARY

SA is a country with a variety of climatic zones. The aviation industry in SA is highly regarded, with the main South African airports doing well in international awards for airports (Skytrax, 2013). The aviation industry is heavily dependent on Meteorological Services, as weather conditions can provide various hazards to this industry. MA can charge for services rendered to the Aviation industry, as the ICAO and the World Meteorological Organization (WMO) granted the MA the right to do so (WMO, 2007). Aviation is currently the most significant commercial entity for SAWS, accounting for about 24 – 31% of SAWS's income (SAWS, 2008; SAWS, 2010; SAWS, 2013).

Various methodologies for TAF verification have been adapted for operational use across the globe. The model suggested by Gordon (1993) for use in New Zealand has been adapted by various other MA including Denmark (Hilden, *et al.*, 1998), Hong Kong (Chan, 2000) and the Netherlands (Jacobs & Maat, 2004). The methodology proposed that TAF messages should be divided into three-hourly blocks and the minimum and maximum values to be obtained for the variables during these blocks. These minimum and maximum values for the variables are then compared to the minimum and maximum values for the variables according to the observational data (METAR and SPECI). Two contingency tables are drawn up and used to derive the verification indexes (Gordon, 1993).

In Germany a similar methodology was used, but the variables were weighted instead. These weighted values were also compared to the observational data (METAR and SPECI) and a contingency table was drawn up. Again the contingency table is used to derive the verification indexes (Balzer, 1995). This methodology was superseded by the methodology proposed by Austro Control (Mahringer, 2008). The verification system adopted is similar to the method of Gordon (Gordon, 1993), but hourly periods are used instead of three-hourly periods. Continuous observational data sources are also used, when available, to supplement the aviation observational messages (METAR and SPECI) (Mahringer, 2008). No documented evidence for TAF verification exists for SA. The verification system developed and tested in SA will be discussed in the methodology chapter as this verification system was used for conducting the research for this dissertation.

NWP models are invaluable tools to forecasters as they provide forecasters with possible future outcomes of weather. In SAWS, the NCAR EPS has been in use since 2000 (Tennant, *et al*., 2007). However, for aviation use, which is more now-casting and short range forecasting, an additional NWP model is required. The UK Met Office is providing SAWS with a modified version of their UM adapted for the Southern African region, running at 12km resolution (Landman, 2012). This model is ideally suited for the task as it has 48 hours of forecasted data and is provided hourly.

The research question of this dissertation is whether the human-forecaster in employ of SAWS, working in Aviation, can improve on raw NWP model-generated TAF messages. To test the hypothesis, the NWP TAF messages as well as persistence TAF messages need to be constructed and then evaluated against the observational data in order to determine which of these data sets performed best. A Monte Carlo method will also be applied to the data sets to determine the significance of the results (Wilks, 2011). The data required to build the three data sets for the three different forecast systems will be discussed in Chapter 2, whereas the methodologies will be covered in Chapter 3.

CHAPTER 2

DATA

In order to test the hypothesis, data sets need to be constructed. Three pairs of data sets will be constructed for this study. The first data set pair is the operational actual data in the form of Aviation Observations (METAR and SPECI) and the operational TAF messages. The second data set pair is the same operational actual data and the persistence TAF messages. The persistence TAF messages are constructed using the operational actual data of the previous day. This will be discussed in Section 3.2. The last data set pair is the operational actual data and the raw UM derived data for the five aerodromes. The derivation of these TAF messages from the UM will be discussed in Section 2.2.

The data used to construct the three dataset pairs come from three sources:

1. The TAF messages for the selected stations, issued by the operational forecasters in the forecasting offices of SAWS.
2. The XAANT version of the South African-run UM. The model data is used to construct TAF messages from an alternative source to the human generated TAF messages. This is the main data source against which the human-generated TAF messages will be measured and compared for the testing of the hypothesis.
3. The operational actual data (METAR and SPECI) for the stations used in the study. The operational actual data is the data set used to evaluate the forecasts. The values in the actual data are the values the forecasts intend to emulate. A control data set is constructed out of the operational actual data, based on the principal of persistence –the actual data of today are used to forecast for tomorrow.

The following three sections describe these three data sources in detail. The construction and origin of these data sources will also be investigated.

2.1 TERMINAL AERODROME FORECAST (TAF) MESSAGES GENERATED BY THE OPERATIONAL FORECASTERS

A TAF is a concise, alphanumerically coded forecast of the expected prevailing conditions at an aerodrome during the period for which the TAF is valid. As with all aviation weather-related messages and products, TAF messages are governed by the ICAO Annex 3 (ICAO, 2010) and WMO Technical Report No. 49 (WMO, 2007). These two documents are the regulations of two governing bodies for Meteorological Service in International Air Navigation and basically contain the same information.

According to the WMO (WMO, 2007), a TAF is to be interpreted as the prevailing conditions for the time indicated in the forecast. Thus, the values given in the TAF are assumed to be the values of highest probability to occur during the time period of the TAF. The occurrence of change groups and temporal weather values is to be interpreted as being the most likely time of these changes or events. Upon the issuance of a new TAF, be it a routine, correction, amended TAF, nil or cancelled TAF, all previous TAF messages will no longer be valid.

Any TAF, with the exception of cancelled or nil TAFs, should contain at least the following components: (WMO, 2007; ICAO, 2010):

1. Forecast type identifier – TAF, AMD TAF or COR TAF, for normal TAF, Amended TAF and Correction TAF
2. ICAO location indicator – a four-digit alphabetical code identifying the aerodrome. South African location indicators start with FA, as Southern Africa is the F-block, for example, Cape Town International is FACT.
3. Time and day of issue – This is given in Greenwich Median Time (GMT) in the following format *ddHHmmZ,* where *dd* is the day of issue and *HH* the hour of issue and *mm* the minute of issue. The *Z* indicates GMT.
4. When applicable, indicate whether the forecast have been issued as a NIL TAF, owing to missing TAF. This can be due to no actual data being recorded on which to base a TAF.
5. Period of validity – Also given in GMT as the TAF is written in GMT. The period group is given as $d_sd_sH_sH_s/d_ed_eH_eH_e$, where d_sd_s is the day and H_sH_s the hour in which the TAF periods starts. d_ed_e is the day and H_eH_e the hour that the TAF period ends.

6. When applicable, indicate whether the forecast has been cancelled (CNL). Cancelled TAF are issued when an aerodrome can no longer be kept under constant review.
7. Surface wind – In SA knots are used to indicate wind speed. Therefore, the wind group is encoded as $W_dW_dW_dW_sW_sGW_gW_gKT$. The $W_dW_dW_d$ is the wind direction rounded to the nearest 10°. If the wind direction is given rounded to a 5°, the speed exceeds 100 kt. W_sW_s is the 10-minute average speed given in knots. If applicable, when a wind gust exceeds the 10-minute wind by 10kt, the GW_gW_g is added to indicate the gust strength to the forecast. The wind group is terminated by the wind scale indicator, which is KT for knots.
8. Visibility – The most likely worst visibility is given in metres rounded to the closest 100 m. The visibility is always expressed as a four digit number. If visibility is below 1000 m, a leading 0 is added. Visibility greater than 10000 m is expressed as 9999. Under certain conditions the CAVOK code can be used to express the visibility and cloud. CAVOK stands for Ceiling And Visibility OKay and is used only when:
    a. Visibility exceeds 10000 m
    b. No cloud is below the minimum threshold of the aerodrome. This is normally 5000 ft, but Cape Town International uses 6500 ft, owing to the mountainous terrain surrounding the aerodrome
    c. No cumulonimbus cloud expected
9. If visibility is given as ≤ 5000 m, a present weather indicator has to be given to indicate why the visibility is reduced.
10. Cloud - one or more layers of cloud, if applicable. When no cloud is expected no significant cloud (NSC) or CAVOK can be used. Cloud are reported as $C_aC_aC_aC_bC_bC_b$, where $C_aC_aC_a$ is the cloud cover expressed as FEW (few), SCT (scattered), BKN (broken) or OVC (overcast) and $C_bC_bC_b$ is the cloud base in 100 ft expressed as a 3 digit numerical value with leading zeroes. Expected cumulonimbus cloud is indicated with a CB at the end. When vertical visibility is expressed, the cloud cover is replaced by VV (vertical visibility).
11. Expected changes to any of the forecast variables, if applicable.

The three major international airports, Johannesburg O R Tambo, Cape Town International and Durban King Shaka airports have TAF messages valid for 30 hours. Bloemfontein Bram Fisher and Port Elizabeth airports have 24 hour TAF messages. These messages are issued routinely every 6 hours at 04Z, 10Z, 16Z and 20Z hours, where the Z indicates time in GMT. In South African Standard Time (SAST) these TAF messages are issued at 00, 06, 12 and 18 hours. The validity period starts two hours after the issue time for routine messages. The validity time of correction and Amendment TAF messages is the same as that of the TAF message it replaces.

Changes in TAF messages are governed by the Section 1.3 of Appendix 5 of WMO Technical Report No. 49 (WMO, 2007). The information is the same as the SPECI criteria in Section 2 of Appendix 5 of WMO Technical Report No. 49 (WMO, 2007) and Section 2 of Appendix 5 of ICAO Annex 3 (ICAO, 2010). The change thresholds are as follows:

- A wind direction change > 60° if the wind before or/and after was ≥ 10 kt
- Mean wind speed change ≥ 10 kt
- Wind gust changes ≥ 10 kt, with the mean wind speed ≥ 15 kt
- Wind directional changes though values of operational significance as determined and agreed on, for the aerodrome. This includes change of runways and significant cross wind components
- Visibility deteriorating or improving through the following levels – 150, 350, 600, 800, 1500 and 3000 m. If the area has significant amount of general aviation, 5000 m is also included
- The onset, cessation or change of intensity of the following:
    - Freezing precipitation - FZ
    - Precipitation – moderate or heavy, including showers, rain and snow
    - Dust storms
    - Sand storms
    - Other weather phenomena reducing the visibility
- The onset and cessation of the following weather phenomena in isolation or combination:

  - Ice crystals
  - Freezing fog
  - Drifting snow, sand or dust
  - Blowing snow, sand or dust
  - Thundershowers precipitating or not
  - Squalls
  - Tornado or waterspout funnel clouds
- When the ceiling, lowest BKN or OVC cloud layer, passes through one of the following levels – 100, 200, 500 and 1000 ft. When the aerodrome is often used by general aviation, 1500 ft is also used
- When a layer of cloud below 1500 ft changes from OVC or BKN to SCT, FEW or NSC, and vice versa
- Vertical visibility changes through the following values – 100, 200, 500 and 1000 ft
- Any other criteria locally agreed on for the aerodrome.

In SA, TAF messages are terminated with the expected minimum and maximum temperatures (in °C) for the aerodrome during the forecast validity period. The format is $TXt_xt_x/d_xd_xH_xH_xZTNt_nt_n/d_nd_nH_nH_nZ$, where t_xt_x is the maximum temperature and $d_xd_xH_xH_x$ is the day and hour of the maximum temperature expected. t_nt_n is the minimum temperature and $d_nd_nH_nH_n$ is the day and hour the minimum temperature is expected. Temperatures below zero degrees Celsius are written with the minus sign replaced by an M before the numerical number. Temperatures are expressed as whole numbers.

2.1.1 TERMINAL AERODROME FORECAST LOCATIONS CHOSEN FOR THIS STUDY

Only the five most important airports in SA are chosen for this study. These airports are chosen as the greatest movement of aircraft and number of passengers pass through these airports. These five airports are also nicely spread throughout SA to provide an indication of different climates and therefore different challenges to the forecasters. Different forecast offices are located at these airports resulting in the forecasts being written by different sets of forecasters. These airports are:

1. Johannesburg O R Tambo International Airport

2. Cape Town International Airport
3. Durban King Shaka International Airport
4. Bloemfontein Bram Fisher International Airport
5. Port Elizabeth Airport

There is one major airport that was not included in the study, namely Nelspruit Kruger National International Airport. This airport was excluded as five airports will suffice. The data period for this study was chosen to be two years, 1 February 2011 to 31 January 2013. This equates to 731 days, as 2012 was a leap year. Four routine forecasts are issued daily and three data sets need to be constructed and evaluated. This gives:

$$(365 + 366)_{days} * 4_{forecast/day} * 5_{stations} * 3_{datasets} = 43860_{total\,forecasts} \quad (2.1)$$

Thus, according to formula 2.1, using only five stations, the maximum number of forecasts to be evaluated is 43,860. If six stations were used the total would have been 52,632. This would have resulted in the evaluation of 8,772 more forecasts. If the Monte Carlo method, discussed in 2.2.5, is taken into consideration the figure becomes 219,343,860 forecasts for five stations and 263,212,632 for six stations. This results in 43,868,772 more forecasts to evaluate. Given these figures, five stations is more than enough to test the hypothesis.

2.2 THE UNIFIED MODEL

The NWP model used to construct the computer-generated TAF messages is the UK Met Office UM administrated by SAWS, particularly the XAANT model run. The history of the model was discussed in Section 1.9. The XAANT model run has a 12 km horizontal resolution and 38 levels (Landman, 2012).

A TAF is considered to cover the prevailing conditions within an 8 km radius of the reference point of the aerodrome (WMO, 2007; ICAO, 2010). With the XAANT model having a 12 km horizontal resolution, there are always grid point values within this radius of the reference point. Using one of the UM grid points minimizes errors resulting from interpolation between grid points to the exact location.

To construct the TAF messages from the UM the appropriate data need to be extracted from the model data files. The research section of SAWS, provided the model data as small regions around the five aerodromes. The grid point to be used for the aerodrome was determined by finding the grid point closest to the aerodrome. The reference grid point was recorded to be used in order to extract the data.

For the construction of the UM TAF messages the following fields were extracted from the reference grid points:

- Relative humidity at 500 hPa – used in the calculation of the total totals index to determine if cumulonimbus cloud is present
- Relative humidity at 850 hPa – used in the calculation of the total totals index to determine if cumulonimbus cloud is present
- Temperature at 500 hPa – used in the calculation of the total totals index to determine if cumulonimbus cloud is present
- Temperature at 850 hPa – used in the calculation of the total totals index to determine if cumulonimbus cloud is present
- Low cloud cover – used to determine the low cloud cover of the cloud group in the TAF
- Convective Rain – used to determine whether convective precipitation is present
- Fog Index – used to determine whether there will be fog
- Surface dew point temperature – used in the calculation of cloud base
- Surface temperature – used in the calculation of cloud base and to determine the minimum and maximum temperatures for the temperature group in the TAF
- Surface Pressure – used in the calculation of the cloud base
- Categorical Snow – used to determine if snowfalls are expected to occur
- Total rain – used to determine precipitation
- 10m U-wind component – used to construct the wind vector
- 10m V-wind component – used to construct the wind vector
- Visibility – used to determine the prevailing visibility

Wind, cloud cover and visibility values can be used directly from the model output as the fields are available. Cloud bases, however, needed to be calculated. Present weather also needs to be determined.

2.2.1 WIND GROUP

Wind within the UM is stored as a U-wind component and a V-wind component. These components need to be combined to calculate the wind vector. The wind vector ultimately gives the direction and speed.

To obtain the wind direction the algorithm in equation 2.2 was used (Revering, n.d.):

$$Wind_{Direction} = \begin{cases} if\ V > 0 \rightarrow \theta = 180 \\ if\ U < 0\ and\ V < 0 \rightarrow \theta = 0 \\ if\ U > 0\ and\ V < 0 \rightarrow \theta = 360 \\ \left(\frac{180}{\pi}\right)\tan^{-1}\frac{U}{V} + \theta \end{cases} \qquad (2.2)$$

Wind speed is determined using the Pythagoras Theorem and converting the result to knots as given in equation 2.3 (Revering, n.d.):

$$Wind_{speed} = \frac{\sqrt[2]{U^2+V^2}}{0.5148} \qquad (2.3)$$

The UM does not have a field for wind gusts. For the construction of the UM TAF messages, a gust of 10 kt greater than the wind speed is added, if the wind speed is ≥15 kt.

2.2.2 VISIBILITY

The visibility value used in the UM terminal aerodrome is the value from the visibility field of the UM. This model value is rounded off to an integer. If the visibility is ≥10000 m the value is set to 9999. If the value is < 1000 m, leading zeros are added to accommodate the four digits.

2.2.3 PRESENT WEATHER

There are many weather descriptors that can be used in a TAF leading to numerous combinations (WMO, 2007; ICAO, 2010). Conditions such as sand and dust storms cannot be determined using model output data. For the purpose of generating TAF messages using model data, the options of present weather was restricted to: fog, mist, rain, showers, thundershowers, snow and haze.

Snow is determined using the categorical snow field of the UM. The snow categorical field is a binary field. This means the value is either 1 or 0 – true or false. If the value is 1, snow is expected and rainfall is changed to snow.

Rainfall is determined using the categorical rainfall field of the UM. As with the snow categorical field, it is also a binary field. Rainfall is further investigated to determine if convective precipitation is present. The convective precipitation field and the total precipitation field of the UM are considered using the following logic given in equation 2.4:

$$Rain_{Convective} \rightarrow Ra_{Conv} \geq \frac{Ra_{Total}}{2} \tag{2.4}$$

Therefore rainfall is upgraded to convective precipitation (showers) if the convective precipitation value is at least half of the total precipitation. To determine the possibility of thundershowers the Total-Totals Index (TT) is calculated using equation 2.5 (Revering, n.d.):

$$TT = T_{850} + Td_{850} - 2T_{500} \tag{2.5}$$

The temperatures in equation 2.5 are expressed in degrees Celsius. The 850 hPa dew point temperature does not exist in the UM model output data and needs to be calculated using the equation 2.6 (Schlatter & Baker, 1991):

$$Td_{850} = \left(\frac{RH_{850}}{100}\right)^{1/8}(112 + 0.9T_{850}) + 0.1T_{850} - 112 \tag{2.6}$$

The Total-Totals Index is used to differentiate between showers and thundershowers. A TT ≥ 45 is used to upgrade the showers to thunderstorms. Thus, if the model is giving precipitation, the following combinations are possible:

- RA – Rain
- SN – Snow
- SHRA – Showers of rain
- SHSN – Showers of snow
- TSRA – Thundershowers of rain
- TSSN – Thundershowers of snow

Reduction of visibility other than precipitation is considered to be one of the following:

- FG – Fog if fog index ≥ 50 and visibility is < 1000 m
- BCFG – Fog patches if fog index ≥ 50 and 1000 m < visibility ≤ 3000 m
- BR – Mist if fog index ≥ 50 and 3000 < visibility ≤ 5000 m

- HZ – Haze if fog index < 50 and visibility ≤ 5000 m.

The fog index is taken from the fog field of the UM model output data. Haze is used as the default present weather indicator if the visibility ≤ 5000 m and fog and precipitation is not expected.

2.2.4 CLOUDS

The cloud cover is taken directly from the raw low cloud cover field (LC_{Cover}) of the UM, as only the lowest level of cloud cover is required. Medium and high level cloud is of no consequence in aviation. The values are assigned as given in equation 2.7:

$$Cloud_{Amount} = \begin{cases} NSC \rightarrow LC_{Cover} < 10\% \\ FEW \rightarrow 10\% \leq LC_{Cover} < 25\% \\ SCT \rightarrow 25\% \leq LC_{Cover} < 50\% \\ BKN \rightarrow 50\% \leq LC_{Cover} < 90\% \\ OVC \rightarrow LC_{Cover} \geq 90\% \end{cases} \quad (2.7)$$

For CAVOK conditions the UM terminal aerodrome will read 9999 NSC, which evaluates the same as a CAVOK code would do. The UM TAF messages will not be used operationally and are only evaluated resulting in the evaluation of these forecasts being correct.

Cloud base is a different problem all together. No easy way exists to determine the cloud base of low cloud that is not convective. With the fields available for use, the approach used was to calculate the lifting condensation level using iteration (Schlatter & Baker, 1991). The first step is to determine the lifting condensation level pressure. To obtain the lifting condensation pressure, the mixing ratio line value for the surface dew point temperature and surface pressure is calculated. The dry adiabatic line for the surface temperature and surface pressure is also determined. The pressure is then decreased incrementally and new temperatures are calculated, using the new pressure, mixing ratio line values and the dry adiabatic line values. The temperature from the mixing ratio line is subtracted from the temperature from the dry adiabatic line. If the absolute difference < 0.00001 the pressure level is considered to be the pressure at the lifting condensation level.

To calculate the mixing ratio line the saturated vapour pressure (e_{Sat}) needs to be determined first apply Equations 2.8, 2.9 2.10 and 2.11. (a_0, a_1 and a_2 are used to simplify equation 2.11) (Schlatter & Baker, 1991):

$$a_0 = 23.832241 - 5.02808\log_{10}(T) \tag{2.8}$$

$$a_1 = 1.3816(10^{-7})\left(10^{(11.344-0.0303998T)}\right) \tag{2.9}$$

$$a_2 = 0.0081328\left(10^{\left(3.49149-\frac{1302.8844}{T}\right)}\right) \tag{2.10}$$

$$e_{Sat} = 10^{\left(a_0-a_1+a_2-\frac{2949.076}{T}\right)} \tag{2.11}$$

T temperature in Kelvin. Once the saturated vapour pressure is calculated the mixing ratio line W, through T and p are calculated, apply equation 2.12, where p is the surface pressure in mb (Schlatter & Baker, 1991):

$$W = \frac{622e_{Sat}}{p-e_{Sat}} \tag{2.12}$$

To calculate the dry adiabatic line O, equation 2.13 is applied (Schlatter & Baker, 1991), with T is in Kelvin and p in mb (Schlatter & Baker, 1991):

$$O = T\left(\frac{1000}{p}\right)^{0.288} \tag{2.13}$$

To calculate the temperature on the mixing ratio line (T_{MR}) at a given pressure, equations 2.14 and 2.15 are applied, where x is used to simplify equation 2.15 (Schlatter & Baker, 1991):

$$x = \log_{10}\left(\frac{Wp}{622+W}\right) \tag{2.14}$$

$$T_{MR} = 10^{(0.0498646455x+2.4082965)} - 280.23475 + 38.9114\left(10^{(0.0915x-1.2035)}\right)^2 \tag{2.15}$$

To calculate the temperature on the dry adiabatic line (T_{DA}) at a given pressure equation 2.16 is used (Schlatter & Baker, 1991):

$$T_{DA} = O\left(\frac{p}{1000}\right)^{0.288} - 273.16 \tag{2.16}$$

The lifting condensation level pressure (p_{lcl}) is then calculated using equation 2.17 (Schlatter & Baker, 1991):

$$(T,p) \rightarrow |T_{DA} - T_{MR}| < 0.00001, p_{lcl} = p \tag{2.17}$$

where p is the pressure value used to calculate both T_{DA} and T_{MR}. This process calculates a pressure value, but a height value is required. To obtain the height value the hypsometric equation 2.18 (Holton, 1992) is used:

$$Z_t \equiv Z_2 - Z_1 = \frac{R}{g_0}\int_{p_2}^{p_1} T d \ln p \tag{2.18}$$

Z_1 and Z_2 are the geopotential heights at pressure levels p_1 and p_2 respectively. Z_T is the thickness of the geopotential. If one calculates a mean temperature $\langle T \rangle$ equation 2.18 can be simplified to equation 2.19, where H the height in metres, R the universal gas constant and g_0 the average gravitational acceleration (Holton, 1992):

$$H = \frac{R\langle T \rangle}{g_0} \tag{2.19}$$

$\langle T \rangle$ is calculated using the following calculation, equation 2.20:

$$\langle T \rangle = \frac{T_{sfc} + T_{lcl}}{2} \tag{2.20}$$

Then 2.18 becomes equation 2.21(Holton, 1992):

$$Z_T = H \ln\left(\frac{p_1}{p_2}\right) \tag{2.21}$$

If one considers the surface pressure to be p and the calculated lifting condensation pressure p_{lcl}, 2.21 can be rewritten as:

$$Z_{Cloud\ Base} = -H \ln\left(\frac{p_{lcl}}{p}\right) \tag{2.22}$$

The result from equation 2.22 is rounded to the lowest 100 ft and converted to the correct format for cloud base. This method works well for the interior stations, but is less accurate along the coast. Given the variables available within the UM this method performs as well as can be expected. It does out-perform any of the imperial methods for estimating cloud base (Revering, n.d.).

The UM has data hourly and therefore the constructed TAF messages are constructed using *FMddHHmm* groups for each hour data set. This creates TAF messages that do not comply with all the regulations, but that are coded correctly (WMO, 2007; ICAO, 2010). These TAF messages will only be used in the evaluation process, where the TAF restrictions do not matter.

In the next section the observational data will be discussed. This is the data against which the TAF messages will be evaluated.

2.3 OBSERVATIONAL DATA

The observational data captured for use in Aviation is written in an alphanumerical code called a Meteorological Aerodrome Report (METAR). In some cases, changes to the prevailing weather need to be recorded in Special Aerodrome Reports (SPECI). A SPECI is issued when certain thresholds are breached which have operational importance to Aviation. METAR and SPECI specifications and criteria are governed by Chapter 4 and Appendix 5 of both the ICAO Annex 3 (ICAO, 2010) and the WMO WMO technical Report No. 49 (WMO, 2007).

An Aviation observation contains the following information (WMO, 2007; ICAO, 2010):

1. Report type identifier – METAR for a routine observation and SPECI for a special report. If SPECI criteria have been reached since the last observation during a routine observation, a METAR is issued followed by an identical SPECI to indicate the significant change.
2. Location indicator – Four digit alphanumeric code used by ICAO to identify the aerodrome for which the observation is done.
3. Observation time – Day and time of observation written as *ddHHmmZ*, where *dd* is the day, *HH* the hour and *mm* the minute of the observation. The *Z* indicates time as GMT.
4. If applicable, an indicator is used when the METAR is automatically generated using only instrument data (AUTO) or when the METAR is missing (NIL).
5. Surface wind – In SA knots are used to indicate wind speed. Therefore, the wind group is encoded as $W_dW_dW_dW_sW_sGW_gW_gKT$. The $W_dW_dW_d$ is the wind direction rounded to the nearest 10°. If the wind direction is given rounded to a 5°, the speed exceeds 100 kt. W_sW_s is the 10-minute average speed given in knots. If applicable, when a wind gust exceeds the 10-minute wind by 10 kt, the GW_gW_g is added to indicate the gust strength to the forecast. The wind group is terminated by the wind scale indicator, which is KT for knots.
6. When the wind direction varies > 60°, but < 180°, the wind variation range will be included. The wind variation group is coded as $d_1d_1d_1Vd_2d_2d_2$, where

$d_1d_1d_1$ and $d_2d_2d_2$ are the limits of the variation. The limits are ordered in such a way that moving clockwise between the limits, the variation is contained.

7. Visibility – The most likely visibility is given in meters rounded to the closest 100 m. The visibility is always expressed as a four digit number. If visibility is below 1000 m, a leading 0 is added. Visibility greater than 10000 m is expressed as 9999. Under certain conditions the CAVOK code can be used to express the visibility and cloud. CAVOK stands for Ceiling And Visibility OKay and is used only when:
   a. Visibility exceeds 10000 m
   b. No cloud is below the minimum threshold of the aerodrome. This is normally 5000 ft, but Cape Town International uses 6500 ft, owing to the mountainous terrain surrounding the aerodrome
   c. No cumulonimbus cloud observed
8. In conditions where visibility is low, runway visual range information can be included in the observation.
9. If visibility is ≤ 5000 m, a present weather indicator has to be given to indicate why the visibility is reduced.
10. Cloud - one or more layers of cloud, if applicable. When no cloud is expected no significant cloud (NSC) or CAVOK can be used. Clouds are reported as $C_aC_aC_aC_bC_bC_b$, where $C_aC_aC_a$ is the cloud cover expressed as FEW (few), SCT (scattered), BKN (broken) or OVC (overcast) and $C_bC_bC_b$ the cloud base in 100ft expressed as a three digit numerical value with leading zeroes. Cloud type is expressed for towering cumulus (TCU) and cumulonimbus clouds (CB) and added to the end of the cloud group. When vertical visibility is expressed the cloud cover is replaced by VV (vertical visibility).
11. Current air temperature and dew point temperature – two digit numerical rounded values of the two temperatures in Celsius, encoded as T_aT_a/T_dT_d. T_aT_a is the rounded air temperature and T_dT_d is the rounded dew point temperature. Negative numbers are expressed with a leading M instead of the – sign.
12. QNH – the QNH is the station pressure reduced to sea level. It is reported as a four digit numerical code of the floored QNH pressure value preceded by a Q to indicate the QNH group. If the pressure is below 1000 hPa, a leading 0 is added after the Q.

13. If trend forecasts are issued for the station, the trend forecast follows the QNH. Trend forecasts are not investigated in this study and will therefore not be elaborated on further.
14. Any remarks can be added to the end following a RMK code.

In SA, hourly observations are done at the observation station during the operational hours of the observation office. During daytime half hourly observations are done in some offices. All the stations chosen for this study have 24-hour operational observations.

The aviation observational data will be used (i) to evaluate the forecasts and (ii) construction of the persistence TAF messages. The observational data is the meteorologically observed data that occurred at the time of observation and is therefore the test data set for the forecasts valid at the time of the observation. The persistence TAF generation will be discussed in 3.2.

2.4 SUMMARY

To test the hypothesis three data sets were required. This first data set consisted of the human generated TAF messages and the operational observation messages. The second data set consisted of TAF messages extracted from the NWP model, in this case the UK Met Office UM, administrated by SAWS; and the operational observational messages. The third data set consisted of the persistence TAF messages and the operational observational data. These three data sets were constructed using only three data sources:

1. Operational observational data – This data was used in all three data sets as the values against which the TAF messages were evaluated and to construct the persistence TAF messages
2. The operational human-generated TAF messages
3. The UM

Section 2.1 discussed in detail the construction of terminal forecasts as done in SA. TAF messages were constructed according to Annex 3 (ICAO, 2010) and WMO Technical Report No-49 (WMO, 2007). This section also discussed the coding of TAF messages and the order of the code elements in the message. This order was used in the constructed messages for the NWP and persistence TAF messages as well.

Section 2.2 discussed the extraction and computation from the UM required to construct the TAF messages from the raw NWP model. Section 2.3 discussed in detail the construction of observational aviation messages also according to Annex 3 (ICAO, 2010) and WMO Technical Report No-49 (WMO, 2007).

Chapter 3 discuss how these three data sources are used for this study, including the construction of the persistence TAF messages, verification considerations and significance calculations.

CHAPTER 3

METHODS

The data sources as described in Chapter 2 are used to construct the three data set pairs. The TAF messages in the three data sets should be constructed so that data set pairs are evaluated against data set pairs.

For the airports used in this study, human-generated TAF messages are valid for 24/30 hour periods and are operationally issued every six hours and updated as needed. Therefore, TAF messages need to be generated from the UM and actual data using exactly the same time periods as the human-generated data. This will ensure that all the data sets contain TAF messages of matching lengths and valid for the same periods.

Amended TAF messages will be excluded from this comparison to eliminate any bias that could result. Therefore, the terminal aerodrome forecast verification will be conducted from the first issued forecasts during the routine dissemination of the TAF messages.

The evaluation will be conducted using the AvianEval verification system discussed in Section 3.4. The results from the AvianEval verification system will be used to determine which of the three data sets is superior. Knowing which data set is the best is not enough to satisfy the hypothesis, therefore significance has to be determined for each data set. This is done to validate whether or not the results from the AvianEval calculations are significant. The significance of the data set results will be determined using the Monte Carlo or Bootstrap method which will be discussed in detail in Section 3.6 (Wilks, 2011).

3.1 TERMINAL AERODROME FORECAST VERIFICATION BASED ON CONTINGENCY TABLES

A TAF is considered to be an indication of the most likely prevailing conditions expected at the aerodrome, during the period of validity. It is also expected to be concise. ICAO recommends limiting the use of change groups to five or fewer within the forecast period (ICAO, 2010).

A TAF is a planning tool, used by pilots to plan their fights. It was never intended to emulate the complex and high variability of the atmosphere. A set of criteria was developed to define operationally significant values for the weather elements given in the TAF (WMO, 2007; ICAO, 2010). These criteria were discussed in section 2.1.1. The values for the components of the TAF message represents all possible values within the TAF criteria where the value falls. Therefore, the value given in the TAF is resistant to variations within the value range of the given criteria category.

Since there exist strict TAF criteria, any change to any of the TAF components not resulting in a change according to the TAF criteria, do not jeopardise the validity and correctness of the TAF. Therefore the problem of TAF verification can be simplified and conducted as a binary event (Gordon, 1993; Balzer, 1995; Mahringer, 2008). The decision to approach TAF verification as binary simplifies the verification process. The binary approach reduces the problem to a yes/no question. Is the forecast value in the same category as the observation? This leads to the construction of a 2x2 contingency table (Jolliffe & Stephenson, 2012):

Table 3.1: 2x2 Contingency Table

		Event Observed		
		Yes	No	Total
Event Forecasted	Yes	*a*	*b*	*a* + *b*
	No	*c*	*d*	*c* + *d*
	Total	*a* + *c*	*b* + *d*	*n* = *a* + *b* + *c* + *d*

a is defined as a hit or Correct Forecast

b is defined as a False Alarm

c is defined as a miss, or Incorrect Forecast

d is defined as an Insignificant Event or correct rejection.

The verification system developed in-house by SAWS uses this binary approach. In Section 3.4.4 the process of constructing the contingency tables and the resulting calculations are discussed.

SAWS uses only five verification indexes. The main index used to determine the quality of the forecast is Accuracy, also known as Proportion correct (Jolliffe & Stephenson, 2012). Accuracy is the percentage of correct forecasts out of the total forecasts issued. This is not the best index to use as it is not specific on any level of skill. Probability of detection or hit rate is also used. The hit rate determines the correct forecasts out of the total of observed events. This determines the ability to detect significant events. The False alarm ratio is also used to determine the number of false alarms issued by the forecasters. The false alarm ratio is the percentage of false alarms forecast within all the forecast events. To determine skill, the Heidke skill score is used. The formulae for the verification indexes used by SAWS and others calculated for use in this study are given in sections 3.4.5 and 3.5.

3.2 PERSISTENCE TERMINAL AERODROME FORECASTS

The concept of persistence used in this study is quite simple. Persistence makes one assumption – that the weather tomorrow will be the same as that of today. This assumption makes the construction of a persistence data set easy. Since the observational data is already available, the persistence TAF messages are already determined. The persistence terminal forecasts only need to be constructed.

The construction of the persistence terminal forecast is simple. Take each observation and add a day to the time of the observation. Then construct a TAF valid for the same time period as the human-forecast generated TAF messages. The persistence TAF messages are constructed as a collection of “FM” (FM = from) change groups. The “FM” change group is used, because according to TAF coding rules, all elements must be given in the “FM” group. In essence, this implies that the TAF prevailing conditions restart when an “FM” group is used. The process of constructing persistence TAF is shown in Figure 3.1.

Since the persistence TAF will only be used in verification of the forecast, it does not adhere to the criteria given in the ICAO Annex 3 (ICAO, 2010) and WMO Technical Report No.49 (WMO, 2007) as it has more than the prescribed number of change groups. Thus, the persistence TAF does not adhere to the definition of the operational TAF – it is not concise and has too many change groups. For the intended use this is good since it provides a unique evaluation set for each evaluation to be conducted.

The persistence data set is to be used as a control in the testing of the hypothesis. The other two data sets are obtained from both human and raw model data generated forecasts. The persistence data set is therefore the only data set constructed using assumptions. The persistence data set will also be subjected to all the verification methods and its significance will also be determined.

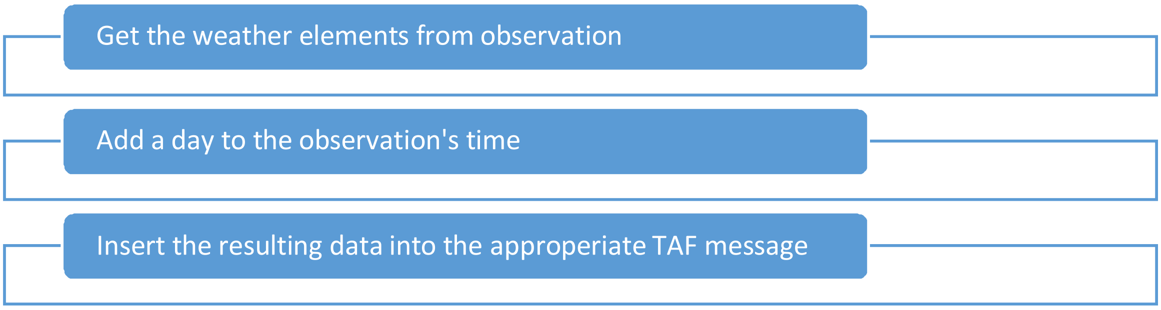

Figure 3.1: Procedure to construct the Persistence TAF messages

3.3 METHOLOGY USED IN THE SOUTH AFRICAN WEATHER SERVICE FOR WRITING TERMINAL AERODROME FORECASTS

The writing of TAF messages are done in strict adherence to ICAO Annex 3 (ICAO, 2010) and WMO Technical Report No. 49 (WMO, 2007). These two manuals are used by SAWS Training section to train forecasting students in Aviation Forecasting.

Figure 3.2 shows the standard operational procedure for Aviation Forecasting. Previous forecasts are evaluated and compared to the actual data. The actual data is analysed and a conceptual model is constructed. Once an idea of the weather is formed from the actual data the guidance forecast and NWP models are consulted to further build the conceptual model for the weather in the geographical area of responsibility of the forecaster. Once the forecaster has completed building a conceptual model of the weather in his/her region the TAF messages can be constructed at the appropriate time for dissemination. The TAF message is constructed using the TAF Criteria as stipulated in section 1.3 of Appendix 5 of WMO Technical Report No. 49 (WMO, 2007).

Once a TAF message has been issued, it is kept under constant review. This is done to ensure that the TAF stays on track. Once the forecast starts to veer off course the forecaster has to amend the TAF. The forecaster needs to keep his/her conceptual model up to date to ensure that the forecast stays on track.

These TAF messages are disseminated globally. An archive of meteorological messages is stored by SAWS, which includes all TAF messages sent by SAWS Forecasters. This archive is used to determine the quality of the TAF messages using the AvianEval verification system, which will be discussed next.

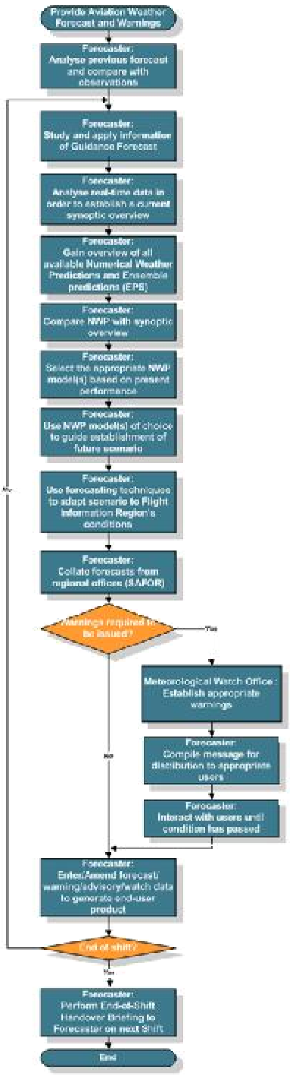

Figure 3.2: Standard Operational Procedure for Aviation Forecasting of the South African Weather Service (SAWS, 2007)

3.4 AVIANEVAL VERIFICATION SYSTEM

The AvianEval verification system was developed in-house at SAWS by the author. This system was the result of a request to implement a system for the evaluation of METAR trend forecasts in 2007. The verification system was developed and expanded to include the evaluation of METAR/SPECI trend forecasts; TAF messages and take-off data. The AvianEval verification system was developed independently of aviation forecast systems in other countries. It was built on the interpretation by the author of the criteria given in the ICAO Annex 3 (ICAO, 2010).

Compared to the systems described in 1.4, this system is in line with international practice. In this section the operational functionality and methodology used in the AvianEval verification system will be discussed. The only focus will be on the TAF evaluation function of the system, as this is the only area that applies to the study at hand. The other functions of the verification system have similar implementations.

3.4.1 OPERATIONAL FUNCTION OF THE AVIANEVAL VERIFICATION SYSTEM

The AvianEval verification system has two operational modes – the monthly report mode and the ad hoc mode. The monthly report mode runs the AvianEval verification system for the chosen month and generates a report for the stations and options chosen during the run. The stations cannot be set up in this mode. The ad hoc mode is more flexible as it allows the user to choose the time range of the evaluation to be executed. It also allows access to the configuration of the stations in the various evaluations and the set-up of the operation directories. Figure 3.3 shows the operational flow of the AvianEval verification system.

A research mode of the AvianEval verification system is applied instead of the operational monthly report mode for use in this study. The research mode evaluates only TAF messages for the five chosen stations over the two-year period of the data. The research mode also calculates more verification indexes than the other methods, as discussed in Section 3.1.

Figure 3.3 shows the logical flow of the AvianEval verification system as it would apply to the ad hoc mode. Figure 3.4 displays the configuration logic used in the AvianEval verification system. In the step of loading the configuration files, the verification system checks if any configuration is present. If none is found, it prompts the configuration dialog to set up a configuration for use. If an existing configuration is found it is loaded for use. This is done to enable the monthly report option to run as one has no way of configuring the system when entering the monthly report mode. In the ad hoc mode one is presented with the full operational functions of the verification system and alterations to the configuration can be made there.

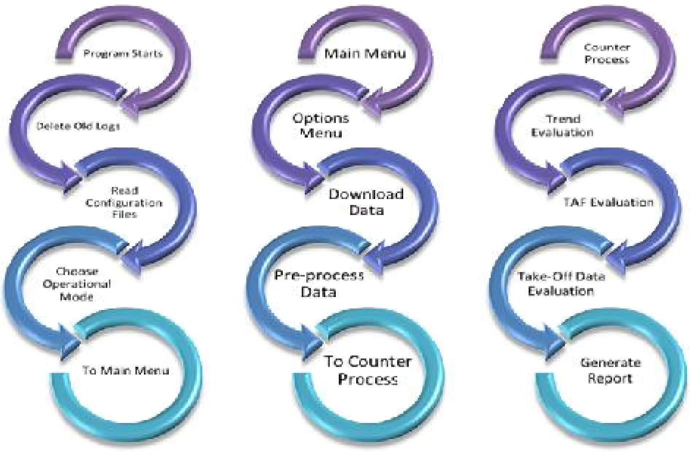

Figure 3.3: Flow Diagram of the Operational Flow of the AvianEval Verification System

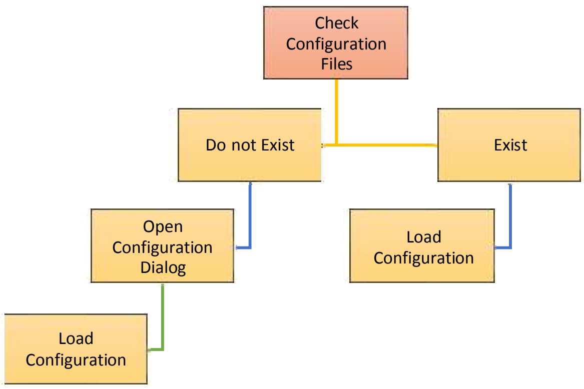

Figure 3.4: Flow Diagram of the Loading of Configuration Files into the AvianEval Verification System

The first action a user gets when launching the AvianEval verification system is to choose between the two operating options - monthly report mode or ad hoc mode. Figure 3.5 shows the flow diagram for this operation. One can generate reports on a daily, weekly, monthly, seasonally, yearly or any combination from one date to another using the ad hoc mode.

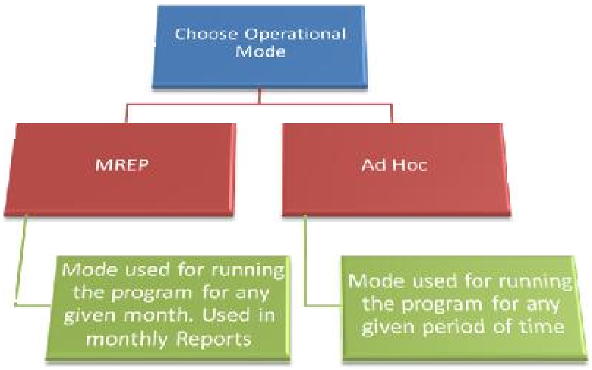

Figure 3.5: The Flow Diagram for Choosing Operational mode of the AvainEval Verification System

Operationally, there is no difference in the way the programme works in either mode. The monthly report mode just uses a predefined configuration and evaluated for the given month only. The ad hoc mode is more customizable. The only difference is the report output name. The monthly report will result in reports named *ICAO-AVI-EVAL-yyyy-MM.xls*, where ICAO stands for the office where the AvianEval verification system was run; yyyy-MM the year and month for which the report is valid. In the case of ad hoc mode the file name changes to *ICAO-AVI-EVAL-AD-HOC*-$y_sy_sy_sy_s$-M_sM_s-d_sd_s-*TO*-$y_ey_ey_ey_e$-M_eM_s-d_ed_e*.xls,* where ICAO is again the office where the report was run. $y_sy_sy_sy_s$-M_sM_s-d_sd_s is the starting date of the report and $y_ey_ey_ey_e$-M_eM_s-d_ed_e is the date on which the report ends.

Figure 3.6: AvianEval Main Screen in Ad Hoc Mode

In the ad hoc mode the user is presented with the main screen (Figure 3.6) where the programme can be set up for use and the period for evaluation can be chosen. There are two calendar drop-down boxes to choose the starting and ending dates of the verification period. Three buttons are found on the bottom of the option screen - Start, Config and Exit buttons. The Start button executes the evaluation process. The Exit button exits the program.

The Config button opens the configuration dialog (Figure 3.7). In this dialog one can choose the working directory and specify the office to use in the filenames of the resultant reports. There are four more tabs for the configuration of the four evaluation processes: Counter Section, Trend Evaluation, TAF Evaluation and Take-Off Data Evaluation. These four tabs are exactly the same. The left hand column contains all possible ICAO indicators for SA from FAAA to FAZZ. On the right hand side appear the stations that will be evaluated for the specific evaluation. On the bottom there are three buttons – Clear to clear all the data in the dialogs; Save and Exit to save all the changes to the configuration files and a Cancel button to close the dialog without saving the changes.

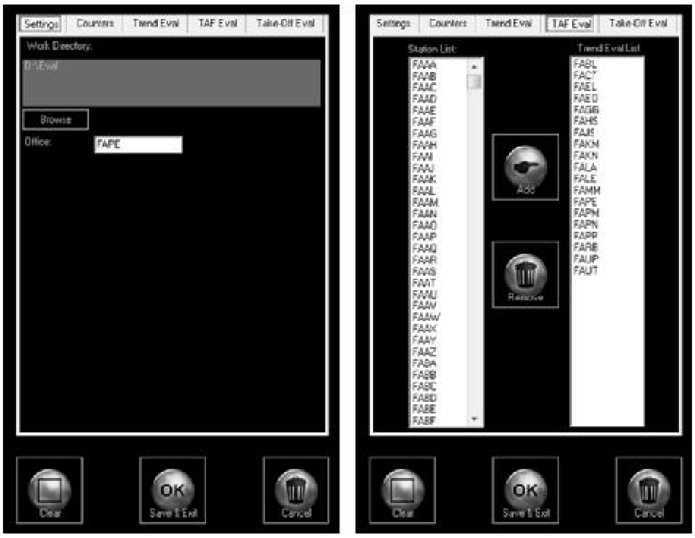

Figure 3.7: Two Views of the Configuration Dialog

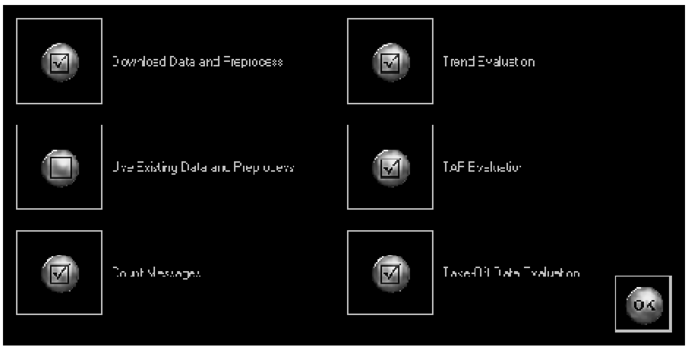

Figure 3.8: Option Dialog of the AvianEval Verification System

Once the evaluation of the user is set up the process can be started by pressing the Start button. This will present the options dialog to the user in both monthly report and ad hoc modes. Figure 3.8 shows the Options Dialog. The Options Dialog contains six selectable options: Download Data and Pre-process; Use Existing Data and Pre-process; Count Messages; Trend Evaluation; TAF Evaluation and Take-Off Data Evaluation.

The user can choose either the download option or to use existing data. This enables the user to obtain data from SAWS Archive or to use data that was acquired previously. For this study the second option will be used. The other four options are only checkboxes to select or unselect the evaluation processes. The report generated at the end of the programme run will create tabs in the Excel Sheet for all processes, but the processes that were not run will be empty.

3.4.2 DOWNLOADING DATA FOR THE AVIANEVAL VERIFICATION SYSTEM

The SAWS archive contains dumps of Meteorological Messages received by the *MetServ* servers in Pretoria. Selected messages are saved to a text dump of meteorological messages. Amongst these saved messages are all the data needed by the AvianEval verification system. At 0000Z the *MetServ* servers in Pretoria save the daily dump into an Archive directory and compress it to conserve space. To acquire the archive the AvianEval programme connects to the *MetServ* server via File Transfer Protocol (FTP) and downloads the file dated the same as the date it is currently processing. After the file has been downloaded to the working directory specified in the configuration, it proceeds to extract the archive and obtains the text file it contains.

Since the text file contains all messages dumped, the AvianEval verification system searches for the bulletins containing METAR, SPECI, TAF and Take-Off Data messages. The dump file is read into memory as a string for processing. In the dump file all messages start with an OK and mostly end with the = sign. Therefore, the AvianEval system replaces all "OK " with "=OK ". Since the abbreviation "CAVOK" is used in METAR, SPECI and TAF messages, "CAV=OK" is replaced by "CAVOK". Double "==" is replaced by "=". "METAR FA", "SPECI FA" and "TAF FA" is replaced by "~METAR FA", "~SPECI FA" and "~TAF FA" respectively. The "FA" is required to ignore messages from outside South African borders. For Take-Off Data "FBZA" is replaced with "~FBZA". "=" is the replaced by "=~". This ensures that all messages within the dump file are separated by a ~-sign.

Once the data from the dump file are processed the "~" is used to divide the dump file string into an array of strings containing the separated messages. The ~-sign is used because it is not a text character that is used in the compilation of meteorological messages. Therefore, the ~-sign should be unique and can be used as a message separator. The programme loops through all these resulting strings. It checks whether the message starts with "METAR FA", "SPECI FA", "TAF FA" and "FBZA". Once one of these search strings is found, the string is then further investigated. The string is checked for "OK". If an "OK" is found, it is checked for "CAVOK". If it does contain "OK", but not "CAVOK" it is a heading and not a message and is therefore ignored. Successful messages are stored in an *ArrayList*, but the message is also checked if the *ArrayList* contains the message already. This is done to eliminate duplicates in the data used for evaluation. Once all the messages for the day have been processed the array-list is saved to a file *yyyyMMDD.aed*, where *yyyyMMdd* is the date of the data it contains. For this study the data sets generated from persistence and model data, need to be in files of the same format. The downloaded data is now ready for the Pre-processing process.

3.4.3 THE PRE-PROCESSING PROCESS

The pre-processing of the data is an essential step in the operation of the AvianEval system. The pre-processing is responsible for detecting coding errors in the messages and automatically fixing common mistakes. Mistakes that require user intervention or are not defined as common mistakes are thrown out as an error and the user is provided with a popup-dialog requiring the user to correct the message or ignore the message manually. These changes in the messages are logged and included in the report.

The pre-processing starts by adding any message in the data for the selected days into the *ArrayList* object. Before the messages are added to the *ArrayList*, the messages are sent to the Build method in the corresponding class of the message. If the message is an Actual (METAR/SPECI), the Build method in the Actual class is invoked. In the Build method various changes are made to the original message to ensure that the message will pass the build action.

For example a variable wind is denoted as “VRB” in the code. The wind direction in all the classes is stored as an integer and VRB is not a string that can be converted to an integer. Therefore, the “VRB” is replaced by “999” in the message. Direction in degrees only goes to 360; therefore a value of 999 is not a definite wind direction and is used as a place-holder for the variable wind. In the case where the winds are not there and are given as "/////KT" it is replaced with "88888KT". The "999" and "888" are used in the system to ignore wind calculations as these are either unavailable (888) or not significant (999).

Other noteworthy automatic changes to the messages include common typological errors and code that used to be included in the bulletins, but are removed from the code in the last update of Annex 3 (ICAO, 2010). This includes "SKC", sky clear, which has been removed. The system replaces it with "NSC", no significant cloud, which now includes the function of "SKC".

3.4.4 EVALUATION OF TERMINAL AERODROME FORECASTS

Trend forecasts and TAF forecasts are evaluated using the same methodology. The only difference between the two methods is the time frequency – Trends are valid for 2 hours and are therefore evaluated in minute intervals. TAF messages are valid for between 9 to 30 hours and are therefore evaluated in hour intervals. The Trend or TAF class object stored in memory is converted to a Forecast class object. Thus the comparison to the actual is done as an Actual object being compared to a Forecast object.

All the actuals are stored in an *ArrayList* of Actual objects and the Forecast Objects in another *ArrayList*. The forecast is then stepped according to type (Trend/TAF) in time intervals. The Actual *ArrayList* is searched to find the corresponding Actual object for the given time. The Actual *ArrayList* is comprised by Actual class objects valid for all the time steps. If no Actual object exists the actual object of previous time step is repeated. This ensures that there are Actual objects available for every comparison time step. An Actual object gets repeated only for an hour. After an hour, if no data is available, no Actual object exists. This ensures that periods of no observational data are not evaluated.

The evaluation uses a 2x2 Contingency Table as given in Table 3.1 (Jolliffe & Stephenson, 2012).

3.4.4.1 Wind direction

Wind direction is evaluated as shown in Figure 3.9. The wind direction is determined. If both wind direction values are 999 it is an insignificant event and the evaluation is aborted. If a value of 888 is found the evaluation for wind direction is aborted. If the evaluation has not been aborted the wind speeds are compared. This is done owing to the SPECI criteria given in section 2.2.2 a) of Annex 3 (ICAO, 2010). From 2.2.2 a) it follows that at least one of the two winds must be ≥ 10kt. If both wind speeds are < 10kt the evaluation is aborted as an insignificant event.

If by this stage the evaluations have not been aborted the difference in wind direction is calculated by the equation 3.1:

$$\Delta\theta = \left|\sin\theta_{act} - \sin\theta_{fc}\right| \quad (3.1)$$

If Δθ > 0.5, the change in wind direction is > 60°, and is therefore considered a miss.

If Δθ ≤ 0.5, the change of wind direction is still within the criteria of 2.2.2 a) (ICAO, 2010) and is therefore considered a correct forecast. The findings of the evaluation comparison are then added to the tally of the total and wind direction contingency tables.

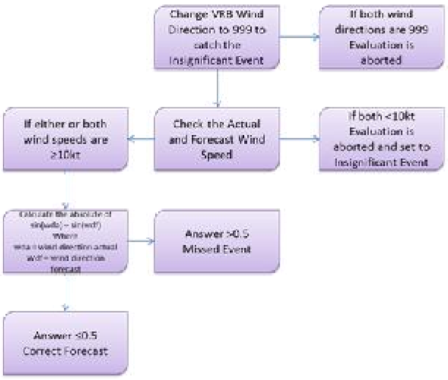

Figure 3.9: Flow Diagram for Wind Direction Evaluation

3.4.4.2 Wind speed and gusts

Figure 3.10 shows the flow diagram of the wind speed evaluation. Wind gusts are evaluated the same as wind speed. SPECI criteria 2.2.2 b) (ICAO, 2010) stipulates that a change of ≥ 10 kt is significant. Therefore, if both the actual wind speed and the forecasted wind speed is < 10 kt, the change in wind speed cannot exceed 10 kt and the evaluation is aborted and assigned as an insignificant event. A value of 88 for either the actual or forecasted wind speeds will abort the evaluation as well, as it indicates a missing wind group – the evaluation is thus ignored.

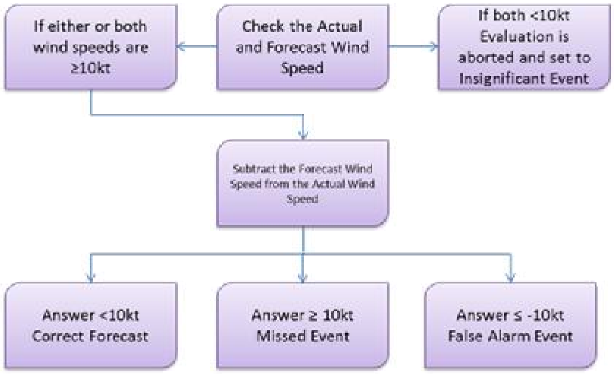

Figure 3.10: Flow Diagram for Wind Speed and Gusts

If at least one of the wind speeds exceeds 10 kt the difference in wind speed is calculated using equation 3.2:

$$\Delta v = v_{act} - v_{fc} \tag{3.2}$$

If $\Delta v < |10|kt$ then the wind speed amounts to a correct forecast, owing to the fact that the change does not exceed the SPECI criteria. If Δv > 10 kt the actual wind speed was stronger than the forecasts and it is considered a miss. If Δv < -10 kt the forecast wind speed was stronger than the actual and it is considered a false alarm.

3.4.4.3 Visibility

The SPECI criteria governing visibility as discussed in section 2.2.3 (ICAO, 2010), gives the visibility thresholds as 150, 350, 600, 800, 1500, 3000 and 5000 m. The visibility values are categorized according to equation 3.3:

$$C_{vis} = \begin{cases} vis > 5000m \xrightarrow{yields} 9 \\ 5000m \geq vis > 3000m \xrightarrow{yields} 7 \\ 3000m \geq vis > 1500m \xrightarrow{yields} 6 \\ 1500m \geq vis > 800m \xrightarrow{yields} 5 \\ 800 \geq vis > 600m \xrightarrow{yields} 4 \\ 600 \geq vis > 350m \xrightarrow{yields} 3 \\ 350 \geq vis > 150m \xrightarrow{yields} 2 \\ vis \leq 150m \xrightarrow{yields} 1 \end{cases} \tag{3.3}$$

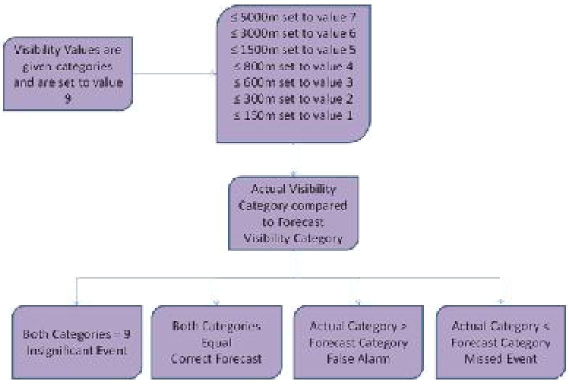

Figure 3.11: Flow diagram for the visibility evaluation process

Equation 3.3 is calculated for both the actual and forecast visibility. If both actual and forecasted visibility categories are 9, then the evaluation is aborted and assigned as an insignificant event. If both categories are the same, both actual and forecast visibilities are in the same SPECI criteria range and it is considered a correct forecast. If the actual visibility category > forecast visibility category, the visibility was better than forecasted and it is considered a false alarm. If the forecast visibility category > actual forecast category, then the visibility was worse than forecasted and it is considered a miss. This process is shown in Figure 3.11.

3.4.4.4 The present weather

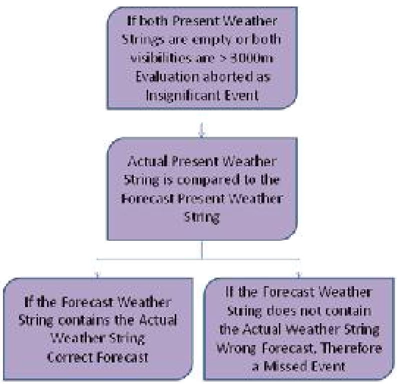

Figure 3.12: Flow diagram for present weather evaluation

SPECI criteria for the present weather was discussed in section 2.2.4 (ICAO, 2010). A decision was made to limit the present weather evaluation to 3000 m visibility or lower as the present weather can be reported regardless of other SPECI criteria. The evaluating process is shown in Figure 3.12.

If both the actual and forecasted the present weather strings are empty the evaluation is aborted and is regarded as an insignificant event. If the visibility of both the actual and forecast is > 3000m the evaluation is also aborted and is regarded as an insignificant event.

The forecasted present weather string is checked to see if it contains the actual present weather string. If it does, it is considered a correct forecast. If not, it is a miss. No false alarms are calculated for the present weather.

3.4.4.5 Cloud cover

Cloud cover and cloud base are interdependent. SPECI criteria 2.2.5 (ICAO, 2010) cover cloud-related changes. Cloud SPECI conditions only come into play once the cloud cover is either broken (BKN) or overcast (OVC) and the cloud base is below 1500ft. Therefore, to consider a cloud cover SPECI at least one of the actual or forecast cloud bases has to be below 1500ft. If both cloud bases are > 1500ft the evaluation is aborted and is assigned to be an insignificant event.

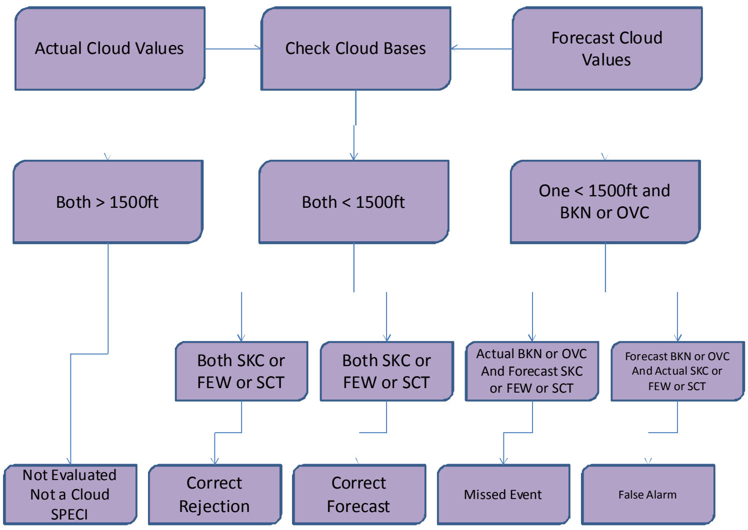

Figure 3.13: Flow diagram for the cloud base evaluation

The evaluation of the cloud cover is computed according to equation 3.4:

$$\langle NIL|FEW|SCT\rangle \overset{changes}{\Longleftrightarrow} \langle BKN|OVC\rangle \tag{3.4}$$

The actual cloud cover and forecast cloud cover are compared. If both actual cloud cover and forecast cloud cover is BKN or OVC, it is considered a correct forecast. If both actual cloud cover and forecast cloud cover is no cloud (NIL); few clouds (FEW) or scattered cloud (SCT), the cloud cover is insignificant and is evaluated as such. If the actual cloud cover is BKN or OVC and the forecast cloud cover is not, then it is a miss. If the forecast cloud cover is BKN or OVC and the actual not, then it is a false alarm. Figure 3.13 shows the flow diagram for the cloud cover evaluation.

3.4.4.6 Cloud base

Cloud base is only significant if the cloud cover is BKN or OVC. If both the actual cloud cover and forecast cloud cover are not BKN or OVC, the evaluation is aborted as an insignificant event. As with visibility, the criteria for cloud bases are given as significant levels, 100, 200, 500, 1000 and 1500 ft. Cloud base values are categorized using equation 3.5:

$$C_{cb} = \begin{cases} cb > 1500ft \xrightarrow{yields} 9 \\ 1500ft \leq cb > 1000ft \xrightarrow{yields} 4 \\ 1000ft \leq cb > 500ft \xrightarrow{yields} 3 \\ 500ft \leq cb > 200ft \xrightarrow{yields} 2 \\ 200ft \leq cb > 100ft \xrightarrow{yields} 1 \\ cb \leq 100ft \xrightarrow{yields} 0 \end{cases} \tag{3.5}$$

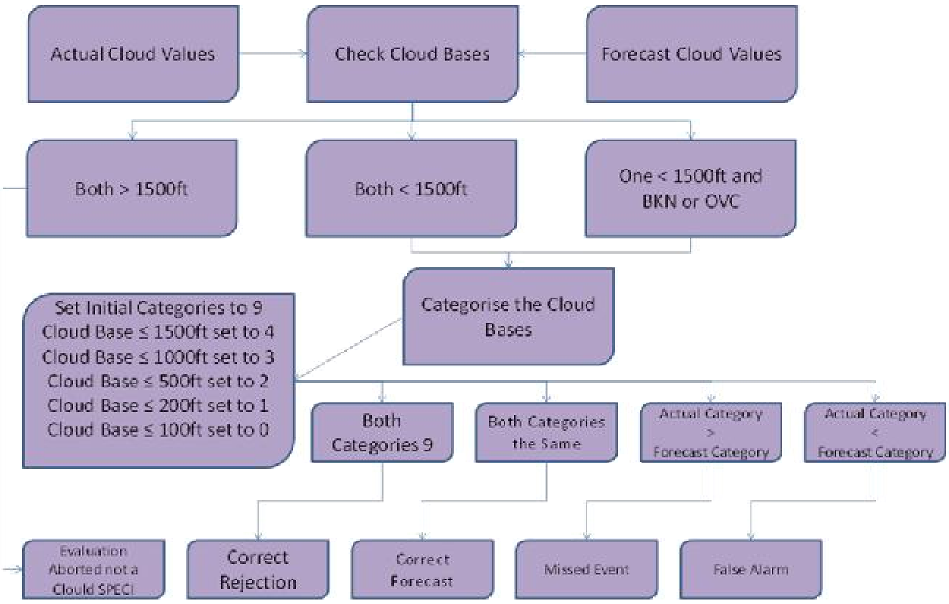

Figure 3.14: Flow diagram for the cloud base evaluation

Figure 3.14 shows the flow diagram for the cloud base evaluation. If both the actual cloud base category and forecast category are 9, then the evaluation is aborted as an insignificant event. If both the actual and forecast categories are the same, then it is a correct forecast. If the actual category > forecast category, then the forecast was worse than the actual cloud base and is assigned to be a false alarm. If the forecast category > actual category, then the actual conditions are worse than expected and it is a miss.

3.4.5 THE GENERATED REPORT

Once the requested evaluations have been completed a report is then generated. For use in this study a research option was added to accommodate additional forecaster indexes in order to evaluate the results. The programme calculates these indexes from the contingency table values that it calculated. The indexes calculated from the contingency given in Table 3.1 (Jolliffe & Stephenson, 2012):

Base Rate (s): $$s = \frac{a+c}{n} \qquad (3.6)$$

Forecast Rate (r): $$r = \frac{a+b}{n} \qquad (3.7)$$

Frequency Bias (ß): $$ß = \frac{a+b}{a+c} \qquad (3.8)$$

Hit Rate (H): $$H = \frac{a}{a+c} \qquad (3.9)$$

False Alarm Rate (F): $$F = \frac{b}{b+d} \qquad (3.10)$$

False Alarm Ratio (FAR): $$FAR = \frac{b}{a+b} \qquad (3.11)$$

Proportion Correct (PC): $$PC = \frac{a+d}{n} \qquad (3.12)$$

Critical Success Index (CSI): $$CSI = \frac{a}{a+b+c} \qquad (3.13)$$

Gilbert Skill Score (GSS): $$GSS = \frac{H-F}{\frac{(1-sH)}{(1-s)}+\frac{F(1-s)}{s}} \qquad (3.14)$$

Heidke Skill Score (HSS): $$HSS = \frac{2s(1-s)(H-F)}{s+s(1-2s)H+(1-s)(1-2s)F} \qquad (3.15)$$

Peirce Skill Score (PSS): $$PSS = \frac{ad-bc}{(b+d)(a+c)} = H - F \qquad (3.16)$$

Clayton Skill Score (CSS): $$CSS = \frac{a}{a+b} - \frac{c}{c+d} \qquad (3.17)$$

Doolittle Skill Score (DSS): $$DSS = \frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \qquad (3.18)$$

Log of Odds Ratio (LOR): $$\theta = \frac{ad}{bc} \quad (3.19)$$

$$LOR = \ln \theta \quad (3.20)$$

Yule's Q (Q): $$Q = \frac{ad-bc}{ad+bc} = \frac{\theta-1}{\theta+1} \quad (3.21)$$

Extreme Dependency Score (EDS):

$$EDS = \frac{2\ln\left[\frac{a+c}{n}\right]}{\ln\left[\frac{a}{n}\right]} - 1 \quad (3.22)$$

Symmetric Extreme Dependency Score (SEDS):

$$SEDS = \frac{\ln\left[\frac{(1-s)F}{H}+s\right]}{\ln[Hs]} \quad (3.23)$$

Symmetric Extremal Dependence Index (SEDI):

$$SEDI = \frac{\ln F - \ln H + \ln(1-H) - \ln(1-F)}{\ln F + \ln H + \ln(1-H) + \ln(1-F)} \quad (3.24)$$

The report presents the calculated data in contingency tables consisting of the 7 elements: totals, wind direction, wind speed, visibility, present weather, cloud cover and cloud base. The totals row of calculated indexes is calculated from the totals of all the components to give values for the forecast as a whole. In the next section the verification indexes will be discussed in more detail.

3.5 DISCUSSION ON VERIFICATION INDEXES

In Section 3.4.6 verification indexes calculated by the AvianEval verification system were given. In this section these verification indexes (Jolliffe & Stephenson, 2012) will be examined more thoroughly.

3.5.1 BASE RATE (S)

From equation 3.6 the base rate is the number of events observed out of the total verification periods. This implies that the base rate is an observational characteristic. The base rate is also known as the climatological probability or sample climate of the weather event.

The occurrence of weather events are out of the control of the forecaster. Therefore, the base rate is not directly relevant to forecasting skill, but skill scores can be dependent on the base rate. This makes verification indexes reliant on the base rate sensitive to variation in climate and observed events.

3.5.2 FORECAST RATE (R)

Where the base rate is the total of events observed out of the total of the verifications, the forecast rate is the forecasted events out of the total verifications. The forecast rate is just an indication of the probability of events being forecasted, out of the given sample of data.

3.5.3 FREQUENCY BIAS (ß)

The frequency bias is the ratio of forecasted events over the observed events. Bias is not an indication of skill. It is an indication of where the forecast tends to be over-forecasted or under-forecasted. A bias value of "+1" is neutral, resulting in a forecast without bias. Values of greater than "+1" are over-forecasted. This could be correct in certain conditions where some false alarms can be tolerated, but missed forecasts should be avoided. A bias can also be used to fine-tune forecast systems based on the bias of the forecast system. Changes can be made to improve the bias closer to the value of "+1".

3.5.4 HIT RATE (H)

The hit rate or probability of detection is exactly as the two names suggest. It is a performance measurement that calculates the hits out of the total number of events. This measurement is particularly useful as it forms part of the alternative value set, reducing the variables from four to three. The contingency table can be rewritten as base rate, hit rate and false alarm rate. These three performance measures can be used to write the performance measures discussed in this chapter. Using this alternative set, some of the formulae are simplified rather than using the contingency table values.

What makes the hit rate valuable in particular is that it indicates the ability of the forecaster to pick up the commencement of an event. This is of great importance in aviation as the onset of an event is critical for flight planning and for the diverting of aircraft already in flight.

3.5.5 FALSE ALARM RATE (F)

The false alarm rate is the opposite of the hit rate. The F is the number of false alarms out of all the non-events. For wind direction the false alarm is not calculated as the wind direction is only considered to be a hit or a miss. There can, however, be conditions that result in correct rejections, which can lead to the false alarm rate becoming “0” instead of “∞”.

3.5.6 FALSE ALARM RATIO (FAR)

The false alarm ratio is the number of false alarms out of the total forecasted events. F is dependent on the observational data set, whereas the FAR is based on the forecast data set. This makes the FAR the preferred measurement of false alarms. The forecaster or forecast system controls only the forecasted data and has no control over the observations. Since the forecast data set is the data set being verified, the FAR is more relevant. SAWS uses the FAR for its evaluation purposes instead of F (SAWS, 2008; 2010 and 2013).

3.5.7 PROPORTION CORRECT (PC)

The proportion correct or accuracy, is the total correct forecasts out of the total data set. It is therefore an indication of the correctness of the forecast. It is, however, a flawed measurement to consider in isolation. For very rare events, not forecasting the events can yield a higher PC than trying to forecast these events. Therefore, PC should not be considered in isolation as it does not provide any information on the abilities of the forecaster. SAWS uses the PC score for evaluating their forecasters (SAWS, 2008; 2010 and 2013).

3.5.8 CRITICAL SUCCESS INDEX (CSI)

The critical success index is also known as the Threat or Gilbert score. The CSI is the correct forecasted events, divided by the correct forecasts, misses and false alarms. Correct rejections do not affect the CSI. Therefore, the CSI is base rate dependent. SAWS also considers the CSI, as it is an indication of the effect of both misses and false alarms on the forecast. If the misses or false alarms are huge, the CSI score plummets.

3.5.9 GILBERT SKILL SCORE (GSS)

One of the various skill scores, the Gilbert skill score was developed by Gilbert in 1884. The GSS is a further development on the thread score or critical success index. Gilbert wanted to determine the skill required to forecast rare events, by finding a way to eliminate correct forecast by chance. The GSS has a heavy dependence on the correct rejections in the contingency table, unlike the CSI that does not.

3.5.10 HEIDKE SKILL SCORE (HSS)

The Heidke skill score is, in essence, the proportion of correct measurement, which has been calibrated to become linear. The HSS is related to the GSS. The GSS can be expressed as equation 3.25:

$$GSS = \frac{HSS}{2-HSS} \qquad (3.25)$$

The HSS is generous to an under-predicting system and harsh to an over-predicting system, when the system has positive skill. The HSS is a more vigorous skill measurement than the previously mentioned scores.

3.5.11 PIERCE SKILL SCORE (PSS)

The Pierce Skill Score is expressed as the difference between the hit rate and the false alarm rate. It therefore implies that, if the hit rate exceeds the false alarm rate, positive skill is displayed. The PSS handles forecasting systems in an opposite manner to the HSS. Therefore, the PSS is harsh to under-predicting systems and generous to over-predicting systems. The PSS is not dependent on the base rate.

3.5.12 CLAYTON SKILL SCORE (CSS)

The Clayton skill score is similar to the PSS. The PSS subtracts the false alarm rate from the hit rate. The CSS subtracts the Detection of Failure (DOF) (equation 2.27) from the Frequency of Hits (FOH) (equation 3.27):

$$FOH = \frac{a}{a+b} \qquad (3.26)$$

$$DOF = \frac{c}{c+d} \qquad (3.27)$$

The CSS is mainly used to determine the economic value of the forecast. Therefore, a positive result has economic value, whereas a negative value is not economically viable.

3.5.13 DOOLITTLE SKILL SCORE (DSS)

The Doolittle skill score is different to other skill scores. It is likened to a chi-squared measure. A chi-squared measure evaluates how expectations relate to the actual results. The data needs to be raw, random, independent variables, mutually exclusive and a large enough set of data (Investopedia, n.d.).

Equation 3.18 can be expressed as a chi-squared measure as:

$$DSS = \sqrt{\frac{\chi^2}{n}} \qquad (3.28)$$

The Gilbert, Heidke, Clayton and Doolittle skill scores are all equitable, non-regular skill scores. These skill scores all tend to punish biased forecasting systems.

3.5.14 LOGS OF ODDS RATIO (LOR)

The odds ratio (θ) is simply the ratio of the product of the correct forecasts and correct rejections to the product of the misses and hits. This, in effect, is the ratio between the odds of a hit and the odds of false alarms. These are calculated as using equations 3.29, 3.30 and 3.31:

$$\omega_H = \frac{H}{1-H} \qquad (3.29)$$

$$\omega_F = \frac{F}{1-F} \qquad (3.30)$$

$$\theta = \frac{\omega_H}{\omega_F} \qquad (3.31)$$

Thus, the LOR is the natural logarithmic of the odds ratio. When the LOR is negative, the odds of the false alarms ≥ 50% of the forecasts. If the LOR is > 0 then the odds of hits > 50% of the forecasts (IDRE, UCLA, 2013).

3.5.15 YULE'S Q (Q)

The Yule's Q was developed by Yule in 1900. It is named after A. Quetelet, a Belgian statistician. Yule's Q is developed for use with the 2x2 contingency table (Table 3.1). No correlations need to be made to calculate Q. Q determines the reduction of error, proportional to the two categories of the dichotomized data (Adeyemi, 2011). The Yule's Q tends to award higher scores for biased systems than for unbiased systems.

3.5.16 EXTREME DEPENDENCY SCORE (EDS)

Extremely rare events are not handled properly by the verification measurements discussed up to this point and the calculated results become meaningless. The extreme dependency score was developed to address this problem. EDS will not degenerate as the other skill scores do. It is however, necessary to remove bias from the forecast system, prior to calculating EDS. EDS is not suitable for general verification as it is easily manipulated.

3.5.17 SYMMETRICAL EXTREME DEPENDENCY SCORE (SEDS)

To address the shortcomings of EDS, the symmetrical extreme dependency score was developed. SEDS is irregular as the values do not converge where H = F = 0 and H = F = 1. SEDS has various properties that are useful in verification. SEDS is independent of the base rate; it shows complement symmetry and is a regular verification measure. SEDS has the added benefit that it does not degenerate either for extremely rare or for common events.

3.5.18 SYMMETRICAL EXTREMAL DEPENDENCE INDEX (SEDI)

One of the primary problems with EDS and SEDS is their dependence on the base rate. To remove base rate dependency, the symmetrical extremal dependence index (SEDI) was developed. SEDI can be viewed as a normalization of the odds ratio, as its numerator is a transformed version of the odds ratio. The advantage of the SEDI is that the values range is now [-1, 1] (Ferro & Stephenson, 2011).

Table 3.2 indicate the properties of EDS SEDS and SETI (Ferro & Stephenson, 2011):

Table 3.2: Properties of three verification measures, ADAPTED from Ferro, *et al*. (2011)

Property	EDS	SEDS	SETI
Non degenerate limit	✓	✓	✓
Base rate independent	✗	✗	✓
Nontrivial to hedging	✗	✓	✓
Regular	✗	✗	✓
Fixed Range [-1, 1]	✗	✗	✓
Asymptotically equitable	✗	✓	✓
Meaningful origin	✗	✓	✓
Complement symmetric	✗	✗	✓
Transpose symmetric	✗	✓	✗

As can be seen from Table 3.2, SEDI has addressed most of the shortcomings of EDS, it is important to remember that these verification methods work best on recalibrated forecasting data (Ferro & Stephenson, 2011).

3.6 MONTE CARLO METHOD

Verification of the forecast data is not enough, since their statistical significance also has to be determined. Significant results enable us to test the hypothesis. The statistical method employed to determine the significance of the forecast data in this study, is the Monte Carlo method. The Monte Carlo method is also known as bootstrapping (Wilks, 2011).

The forecast data sets are one-sample data. Therefore, permutations do not exist. The best option available to resample the data for the purpose of significance, is to resample the data set using elements of the data set. One then creates a data set of equal length. One can randomize the original data set, by using replacement, to replace each forecast in the data set with a randomly selected forecast out of the original data set. The bootstrap or Monte Carlo method uses resampling with replacement (Wilks, 2011).

Conceptually speaking, the Monte Carlo method can be likened to placing the forecasts on separate slips of paper. These slips of paper are then thrown into a hat. The contents of the hat are then shaken and a random slip is drawn from the hat. The forecast data of the drawn slip are then used to record the forecast for the day for which it was drawn. Once the forecast have been recorded for the day, the slip is returned to the hat and the contents of the hat are shaken again. For the next day a slip is again drawn from the hat. The drawn slip may or may not be the same as a slip that was previously drawn. Thus, resampling with replacement ensures that the entire data set is available for all n days within the data set period (Wilks, 2011).

The randomized data sets tend to smooth out in significantly large quantities. Since the data set consists of 365 + 366 days (2 years), 5000 randomized sets will be constructed to smooth the curve. The original and randomized data sets will be verified using the verification measures described in Section 3.5. Once all the datasets have been verified, the scores for each verification measure are sorted according to value. For most of the calculated values from the verification scores, are sorted high to low, as for most of the verification scores 100% is the goal. False alarm rate and ratio, however, are sorted low to high, as false alarms are considered to be best avoided and a lower value is better. The datasets will be ranked from top to bottom and the ranking of the original data set will be used to determine the significance of the data. Since there are 5001 data sets, 1 original and 5000 randomized sets, equation 3.32 will be used to determine the significance:

$$Sig_{Org} = 100\left(1 - \frac{Rank_{Org}}{5001}\right) \tag{3.32}$$

For Sig_{Org} to be significant, the resultant value needs to be > 95. This implies that the original forecast data set should not be ranked lower than 250.

3.6.1 IMPLEMENTATION OF THE MONTE CARLO METHOD

The Monte Carlo method was built into the AvianEval verification system. The final evaluation of the data set and the randomized data sets are performed as described in Section 3.4, but the reports were saved according to the verification measures instead of a combined output file.

The data set had to be processed to remove any coding errors from the original data set. This is done to eliminate unexpected crashes when the program is running. The original data set is read to memory into a list. The original data set is then processed through the TAF Verification process as discussed in Section 3.4.5. The results are stored into files corresponding to the verification measures.

Once the original run is completed, 5000 randomized data sets are generated. These data sets are created using the list of forecasts created from the original data set. The list has n number of forecasts. For each day of the period a random integer value ranging from 0 to n is generated. This random number is then used to choose the forecast to be used as the forecast of the day. The forecast is then transformed by replacing the time of the original forecast value with the corresponding time for the date that the forecast was drawn. No forecasts are ever removed from the list; therefore the list acts as the conceptual hat with the forecast list members as the strips of paper. Once the new data set has been created, it is verified using the exact methods used on the original data. The values are saved to the corresponding files, by adding the results to the file. Once all of the 5000 randomized data sets are completed and verified the programme exits.

There are now files for each verification measure corresponding to the station and description of the data that was evaluated. These files are saved as comma separated files, which can be read into Microsoft Excel. Once in Excel, the values are sorted. The cell number is the ranking. The ranking and values of the variables for the stations will be compiled into Excel spreadsheets by hand for the study of the results. The results will be examined in Chapter 4.

3.7 THE CHOICE AND JUSTIFICATION OF RESULTS TO DISCUSS

The forecast verification at SAWS is primarily concerned with the following verification indexes:

1. Proportion Correct
2. Hit Rate
3. False Alarm Ratio

4. Critical Success Index
5. Heidke Skill Score

These five performance indexes cover a number of areas relevant to forecasting. PC, while not the most correct index to use in isolation, is generally the performance index used in the targets set to SAWS forecasters and is also the index by which ICAO expresses its desired levels of accuracy on the different aviation forecasts (ICAO, 2010). H presents the ability of the forecasting system to detect significant events, hence the alternative name – probability of detection. The FAR is used instead of the F. The FAR is forecaster-orientated, whereas F is observation-orientated. The false alarms resulting from forecasts is desired, not the false alarms in observation.

The CSI investigates both misses and false alarms in a single verification index. Thus, if the misses and false alarms become large, the CSI drops in value. The HSS is based on the PC, and therefore is kinder to systems under forecasting significant events, than to systems that tend to over forecast the events (Jolliffe & Stephenson, 2003; Wilks, 2011 and Jolliffe & Stephenson, 2012).

Of the remaining verification measures various skill scores exist. In order for a skill score to determine whether one forecast is better that another, the skill score should be equitable, implying that the skill score should treat both random and constant forecasts the same (Wilks, 2011). Of the remaining forecast measures only the PSS is equitable and will therefore be used in examining the results. The HSS is also equitable. Therefore, for examining the results in order to come to a conclusion, the following six performance measurements will be evaluated:

1. Proportion Correct
2. Hit Rate
3. False Alarm Ratio
4. Critical Success Index
5. Heidke Skill Score
6. Pierce Skill Score

3.8 FORECASTER VERSUS UNIFIED MODEL (UM) SKILL SCORE

In order to test the hypothesis conclusively the performance indexes of the forecaster will be compared to the performance indexes of the UM. In order to do this, the generic skill score formula will be used to construct the Forecaster versus UM Skill Score, FvsUMSS.

A skill score is defined as the percentage of improvement of a forecast system (A), in comparison to a reference forecast (A_{Ref}) (Wilks, 2011). The generic formula for a skill score is given in equation 3.33 (Wilks, 2011):

$$SS_{Ref} = \frac{A - A_{Ref}}{A_{Perf} - A_{Ref}} \times 100\% \qquad (3.33)$$

Thus, using the UM performance indexes as the reference forecast equation, 3.33 becomes:

$$FvsUMSS = \frac{A_F - A_{UM}}{A_{Perf} - A_{UM}} \times 100\% \qquad (3.34)$$

Where:

- A_F is the forecaster performance index value
- A_{UM} is the UM performance index value
- A_{Perf} is the value for a perfect forecast. A_{Perf} = 1 for all the performance indexes, except for F and FAR, where A_{Perf} = 0

If the human-forecaster is out-performing the UM forecast, the value will be positive. If the value is negative, the UM is performing better than the human-forecaster. Both forecasts will be considered equal if the result of equation 3.34 is zero (Wilks, 2011). The Forecaster versus UM Skill Score will therefore indicate whether or not the forecasters are performing better than the UM, which is the basis of this hypothesis.

3.9 SUMMARY

The verification methodology of TAF messages was briefly discussed in section 3.1. The methodology for constructing the persistence TAF messages was dealt with next in section 3.2. Persistence forecasting makes the assumption that the weather tomorrow will be the same as the weather today. The methodology of the compilation of TAF messages by human-forecaster at SAWS was discussed in section 3.3, citing its conformance to the prescriptions given in Annex 3 (ICAO, 2010) and WMO Technical Report No.49 (WMO, 2007).

The workings of the AvianEval verification system were discussed in detail for terminal aerodrome forecast verification in section 3.4. The different operational modes of the AvianEval system were discussed. The method and source of its data were explained and how the relevant operational messages were extracted from the archived dump of meteorological messages as they are stored on the servers of SAWS. The importance and functionality of the pre-processing process were discussed in section 3.4.3.

The verification process, based on the SPECI criteria in Annex 3 (ICAO, 2010) was discussed in detail for the six meteorological variables being verified in section 3.4.5. The verification indexes that the AvianEval system can calculate were briefly discussed in 3.5. These verification indexes are:

- Base Rate (s)
- Forecast Rate (r)
- Frequency Bais (ß)
- Hit Rate (H)
- False Alarm Rate (F)
- False Alarm Ratio (FAR)
- Proportion Correct (PC)
- Critical Skill Index (CSI)
- Gilbert Skill Score (GSS)
- Heidke Skill Score (HSS)
- Pierce Skill Score (PSS)
- Clayton Skill Score (CSS)
- Doolittle Skill Score (DSS)
- Logs of Odds Ratio (LOR)
- Yule's Q (Q)
- Extreme Dependency Score (EDS)
- Symmetrical Extreme Dependency Score (SEDS)
- Symmetrical Extremal Dependence Index (SEDI)

The implementation of the Monte Carlo method to determine the significance of the results was discussed in section 3.6. Only six of the calculated verification indexes are used for the study of results. Proportion correct, hit rate, false alarm ratio, critical success index are Heidke skill score are used at SAWS and are therefore included in the results. The Pierce skill score is also considered as it is the only remaining score which is equitable. A new skill score, the forecaster versus Unified Model skill score, is also calculated to compare the human-generated forecast results with the corresponding results of the UM as the base forecast to determine the forecast system resulting in the best results. Now that all the data have been collected and constructed; and the verification systems (AvianEval and the Monte Carlo method) applied to the data sets, the results for the two-year period will be discussed in Chapter 4.

CHAPTER 4

DISCUSSION OF RESULTS

In order to study the results and answer the hypothesis, the results need to be evaluated and filtered, to determine which results should be used in the investigation of the hypothesis. Calculated verification results might be irrelevant or unnecessary to test the hypothesis and can therefore be ignored. It can also be argued that the properties of verification indexes do not contribute to answering the hypothesis.

In this chapter the results generated will be scrutinized and evaluated to determine which results to use to test the hypothesis. In the remainder of this chapter, the performance indexes will be discussed. This will be conducted for each of the stations separately as this study is comparing the different forecast systems to one another. The aerodromes for this study are located in vastly different climatic zones. Thus, the aerodromes present the forecasters working there with different challenges and can therefore not be compared to one another. For the data in the results, data up to the fourth decimal is considered significant digits.

4.1 PROPORTION CORRECT (PC)

In this section figures 4.1, 4.2, 4.3 and 4.4 are discussed. Figure 4.1 display values for the 22Z TAF messages for all five aerodrome and all three forecast systems, figure 4.2 the 04Z TAF messages, figure 4.3 the 10Z TAF messages and 16Z TAF messages. The significance values calculated using the Monte Carlo method are also displayed on the graphs as values. ICAO specifies that the accuracy of TAF components should be above 80% and cloud above 70%. Therefore, any PC value above 80% is considered to be meeting the requirements set by ICAO (ICAO, 2010).

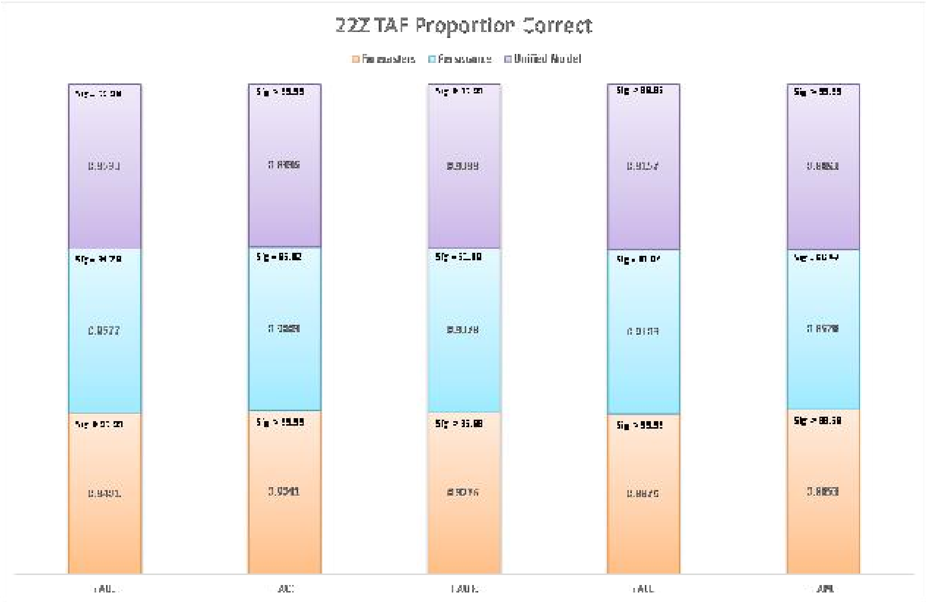

Figure 4.1: 22Z TAF Proportion correct

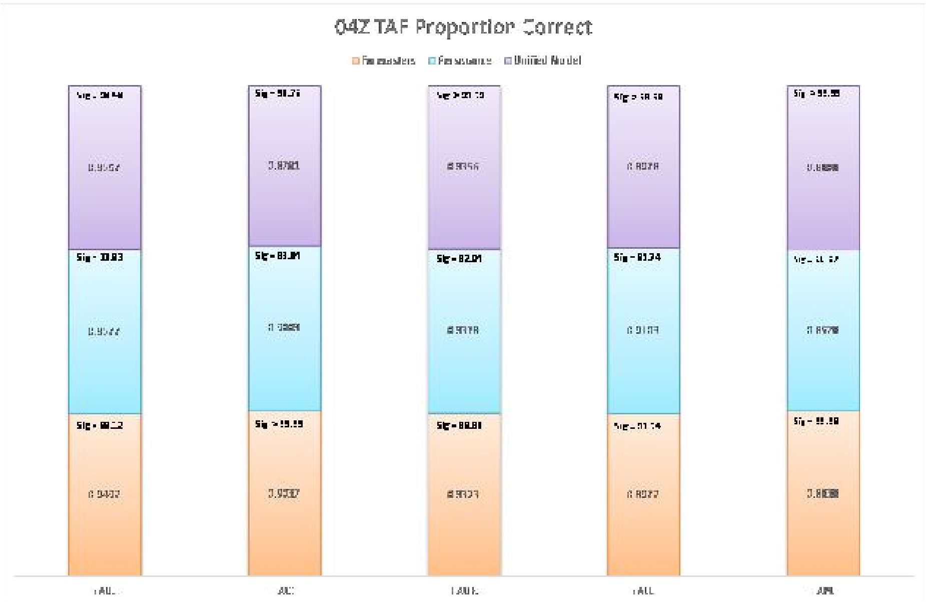

Figure 4.2: 04Z TAF Proportion correct

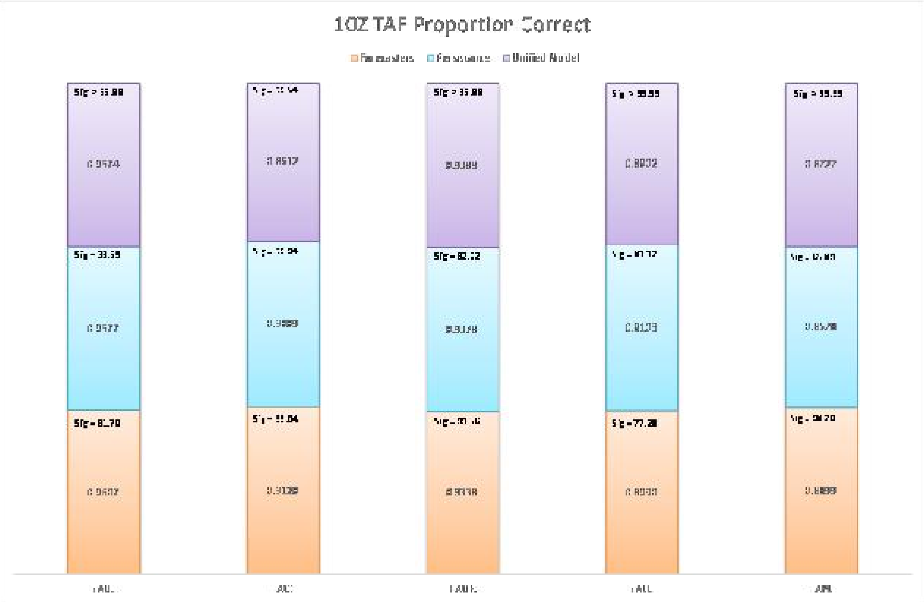

Figure 4.3: 10Z TAF Proportion correct

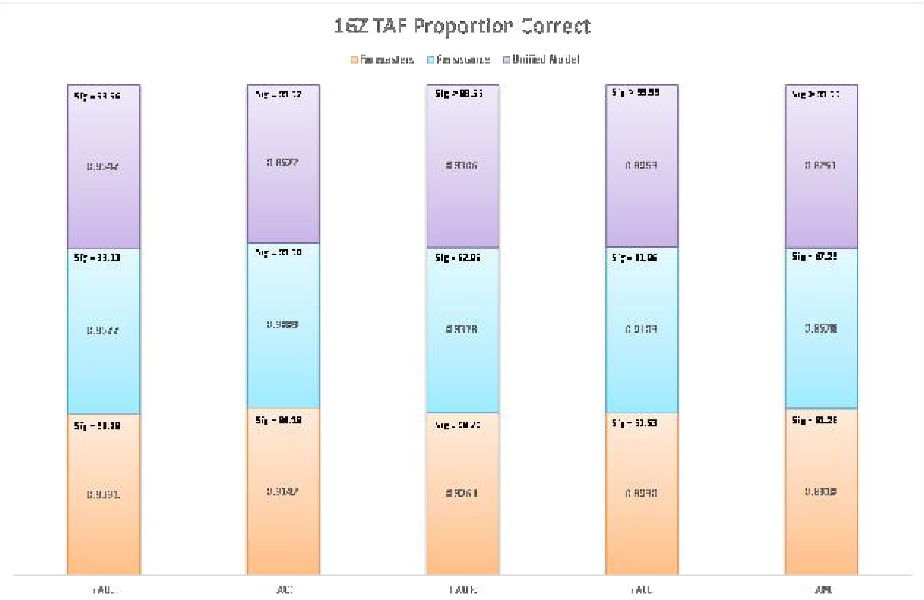

Figure 4.4: 16Z TAF Proportion correct

4.1.1 BLOEMFONTEIN BRAM FISCHER AIRPORT (FABL)

PC values are all in the mid-90s for all three forecast systems. The values of the human-forecaster show the highest variance with a variance value of 2.2%. The UM are more consistent, with a variance of 0.4%. Persistence shows a value of 95.8%, which is maintained for all the issue time groups (on all four figures), as the way persistence TAF messages are constructed. All the persistence performance indexes were the same for all the issue times. Therefore, for persistence, only one value will be considered for each station and performance messure.

When considering the significance of the results, using the Monte Carlo method in Chapter 3, the UM and the human-forecaster results were significant. Only the 10Z human TAF had significance below 95%, but still above 90%. Thus, the UM and the human-forecaster results are reliable, as they out-score random forecasts significantly. The persistence PC significance is only 33 – 35%. Thus, whilst persistence yields good results, it has a poor significance.

PC values for UM and human-forecaster are almost indistinguishable. This indicated that both forecasting systems are relatively similar, when considering PC. Based on the PC performance index, both systems perform the same and the results are significant for both systems.

4.1.2 CAPE TOWN INTERNATIONAL AIRPORT (FACT)

PC values show greater variability than the PC values for FABL. The values for human-forecaster PC are 90 – 92%. The variance of the human-forecaster PC values is only 1.1%. The persistence PC value was 90.1%. The UM PC values were between 86 and 90%. The variance for the UM PC values is 3.9%.

Using the Monte Carlo method on the PC values, significance was >99% for all the time periods and all three data sets, except the 16Z human-forecaster, which dropped to 86.18%.

Thus, based on the PC index, there are no real differences between the human-forecaster and the persistence forecasts. All three forecast systems are significant. The UM forecast did lack behind the human-forecaster, which gives the human-forecaster the edge.

4.1.3 JOHANNESBURG OR TAMBO INTERNATIONAL AIRPORT (FAOR)

The PC values are very consistent throughout. The PC values vary between 92 and 94% for all three data sets. The variance for the human-forecaster PC forecasts is 0.768%. The variance for the UM PC forecasts is 0.933%. The persistence forecast yielded a PC value of 93.8%.

Results from the Monte Carlo method on the PC values indicates that both the human-forecaster and UM values are >99% significant. The persistence forecast data set is >90% significant. With the PC variances being so small and the values obtained for the human-forecaster and UM nearly identical, this indicates that both forecasting systems are equal in performance and exceeds the expected requirements.

Both these forecasting systems are found to be highly significant. Therefore, no call can be made on the forecast systems based on PC.

4.1.4 DURBAN KING SHAKA INTERNATIONAL AIRPORT (FALE)

PC values are slightly lower than for Bloemfontein, Cape Town and Johannesburg. These PC values vary between 88 and 92% for the human-forecaster and UM forecasts. The interesting result is the persistence forecast PC value, which was 91.0%. The variance of the human-forecaster PC values is 1.15%. The variance for the UM PC is 6.72%, which is almost six times greater than that of the human-forecaster.

The significance values obtained for PC values are quite interesting. The UM PC values are all at least 99.9% significant. The 04Z and 22Z TAF messages issued by the human-forecaster are >99% significant for PC. The TAF messages issued at 10Z and 16Z are not significant at all for the PC. The significant value for the 10Z TAF for PC is 77.28%. The significance value of PC for the 16Z TAF messages dropped to 53.53%. The PC values of the persistence forecast are all around 80-82%.

The PC values for the human-forecaster do not vary as much as those of the UM and the values are in the same range. It is, however, noteworthy that the UM was always significant and the forecaster was not. The UM wind this round, as it is more significant.

4.1.5 PORT ELIZABETH AIRPORT (FAPE)

PC values are all between 88 and 90% for both the human-forecaster and UM. The persistence forecast PC is 85.8%. The variance of the human-forecaster PC is 0.72%. The variance for the UM PC is 1.36%.

The results from the Monte Carlo method are similar to those significance values of Cape Town International. The human-forecaster and UM significance values were >99% for present correct, except the 16Z TAF messages of the human-forecaster, which dropped to 81.26%. The persistence forecasts showed significance values of 66 - 68%.

The PC variances for both the human-forecaster and UM are very small and both data sets are highly significant except the evening TAF messages, which had a drop in significance. Therefore, no real difference between the two forecast systems is evident.

4.2 HIT RATE

Figures 4.5, 4.6, 4.7 and 4.8 show the H values for the four sets of TAF messages. The significance values calculated using the Monte Carlo method are also displayed on the graphs as values.

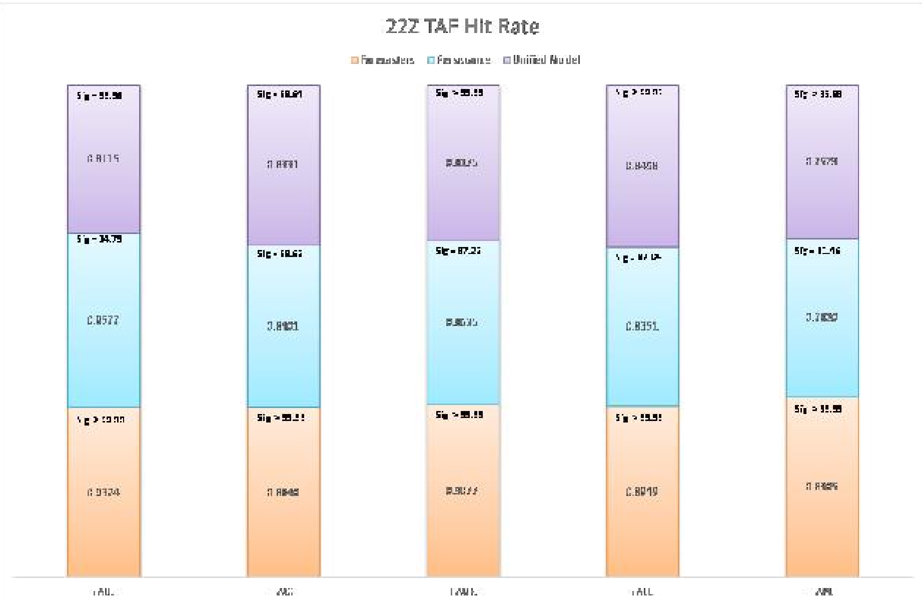

Figure 4.5: 22Z TAF Hite Rate

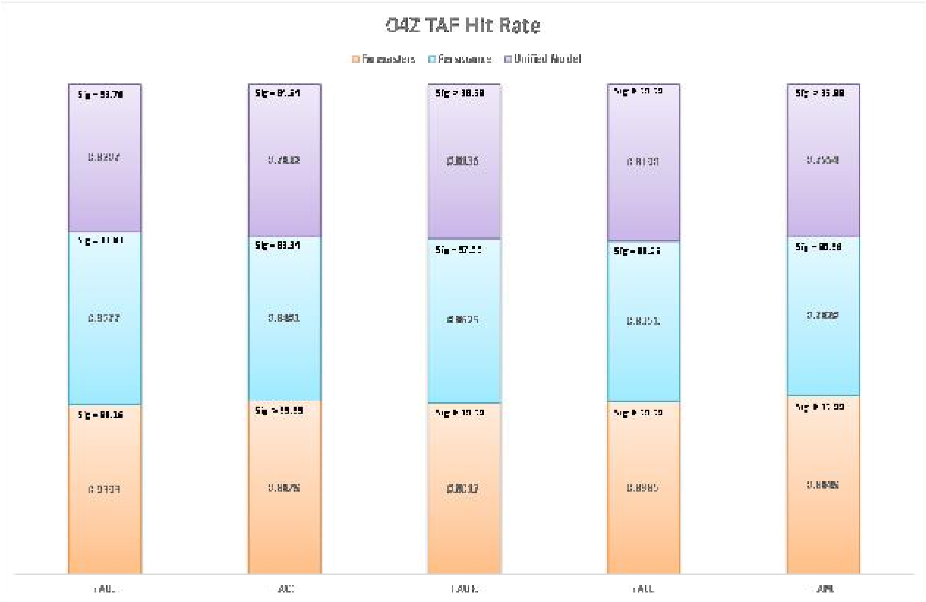

Figure 4.6: 04Z TAF Hit Rate

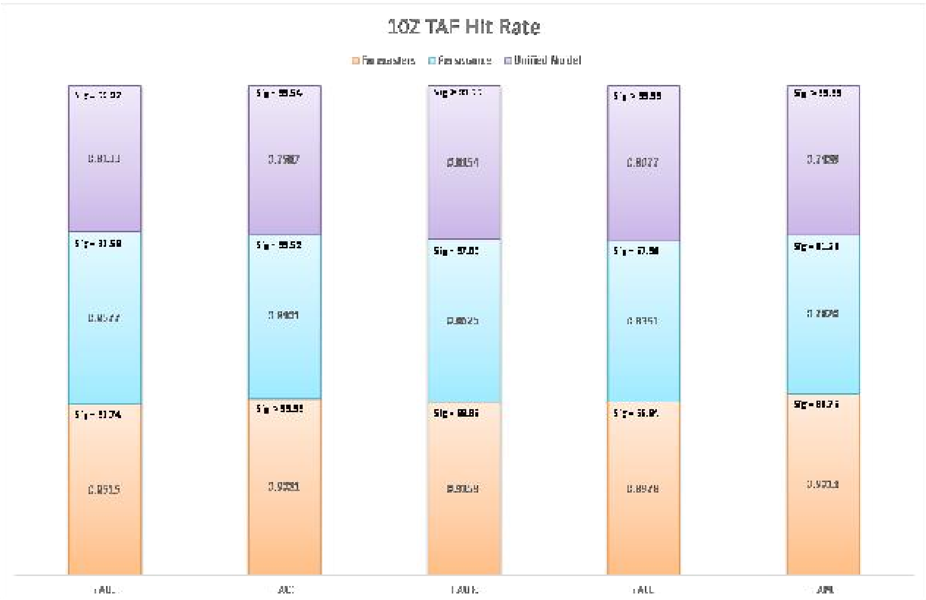

Figure 4.7: 10Z TAF Hit Rate

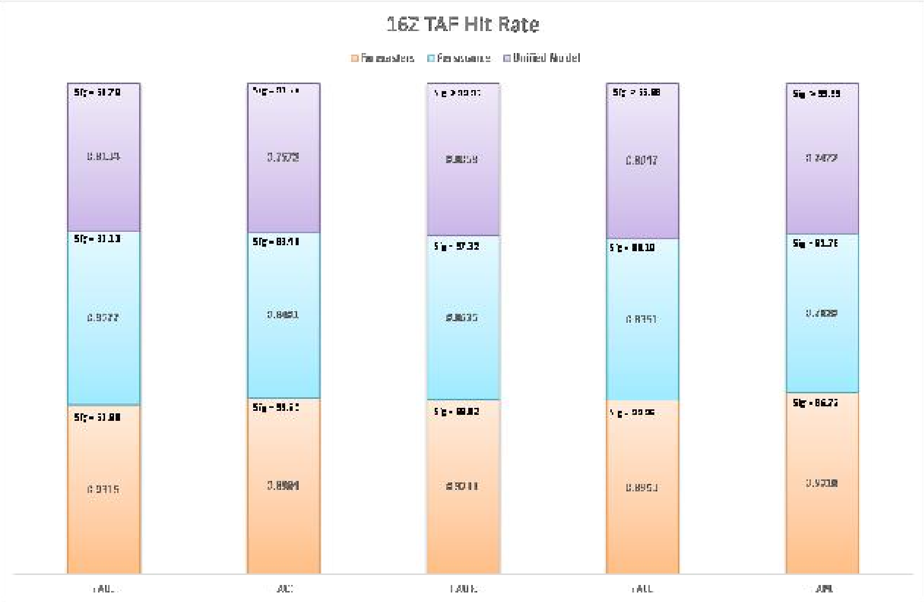

Figure 4.8: 16Z TAF Hit Rate

4.2.1 BLOEMFONTEIN BRAM FISCHER AIRPORT (FABL)

The H scores of the human-forecaster are between 93 and 96%. The variance for the human-forecaster H is 1.993%. The H scores for the UM drop to 81.83%. The variance for the UM H is 0.919%. The persistence forecast H is 95.77%.

The H significance from the Monte Carlo method, is >98% for most of the UM and human-forecaster data sets, but the UM 04Z TAF messages and the human-forecaster 10Z and 16Z TAF messages dropped to just over 93%. The significance of the persistence data sets was only 33.35%.

Whilst persistence did yield the highest H, the significance of the persistence data sets were very low and therefore not significant. The human-forecaster and UM H significance did drop slightly below the 95% threshold, but was highly significant when this was not the case. Since the human-forecaster scores where more than 10% higher than those of the UM, it must be concluded that the ability of the human-forecaster to detect the onset of significant weather is superior to that of the UM.

4.2.2 CAPE TOWN INTERNATIONAL AIRPORT (FACT)

The H values for human-forecaster are 88 - 91%. The variance for the H of the human-forecaster is 1.811%. The UM H values are 75 - 84%. The variance for the UM is 7.586%. The H value for the persistence forecast is 84%.

The Monte Carlo method showed significance for all the three forecasts data sets, >99%. Only the significance of the UM 04Z TAF messages' was at 94.64%. This significance result indicates that all three forecast systems were viable with regard to H.

With all three forecast systems being significant, the spoils should go to the human-forecaster for achieving the highest H values. The lowest H value for the human-forecaster data sets is greater than the highest result from the UM and the persistence forecasts.

4.2.3 JOHANNESBURG OR TAMBO INTERNATIONAL AIRPORT (FAOR)

H values for human-forecaster are 90 - 93%. The H variance for the human-forecaster is 1.94%. The UM H values are 80 - 82%. The H variance for the UM is 1.155%. The H for persistence forecasting is 86.2%.

The significant values of the Monte Carlo method for H are >99% for both human-forecaster and UM data sets. The persistence forecast was also >97% significant. This indicates that all three forecast systems are identifying significant weather changes significantly better than just by chance.

Since all three forecast systems are significant, none of the systems can be ignored. The H values for Johannesburg O R Tambo International indicate that the human-forecaster does have the edge here as the H values are higher than those of the UM and the persistence forecasting systems.

4.2.4 DURBAN KING SHAKA INTERNATIONAL AIRPORT (FALE)

The human-forecaster H values are 89 - 90%. The human-forecaster H variance is only 0.367%. The UM H is 80 - 85%. The UM H variance is 4.213%. The persistence forecast H is 83.5%.

H significance obtained from the Monte Carlo method, for both human-forecaster and UM data sets is >99%. The persistence significance is 87 - 89%.

In terms of H, the human-forecaster does have the edge over the UM. Both the human-forecaster and UM are highly significant, but the human-forecaster data sets obtained higher values.

4.2.5 PORT ELIZABETH AIRPORT (FAPE)

Human-forecaster H values are 88 - 91%. The variance of human-forecaster H values is 1.718%. The UM H values were the lowest of all the airports, around 74 - 76%. The persistence forecast H value is 78.2%, which is also higher than the UM H values.

The significance values obtained from the Monte Carlo method, show that all the data sets from the human-forecaster and UM are >95%, which makes them all significant. The persistence forecast has significant values of 80 - 82%.

With both the UM and human-forecaster systems determined as significant, the human-forecaster has a clear lead with regard to H. This is due to almost 15% higher values for H by human-forecaster. Persistence also out-scored the UM, but is not as significant.

4.3 FALSE ALARM RATIO

Figures 4.9, 4.10, 4.11 and 4.12 show the FAR values for the four sets of TAF messages. The significance values calculated using the Monte Carlo method are also displayed on the graphs as values. The FAR has a negative orientation. A FAR score of “0” indicates a perfect forecast. Thus, values closer to “0” indicate better forecasts.

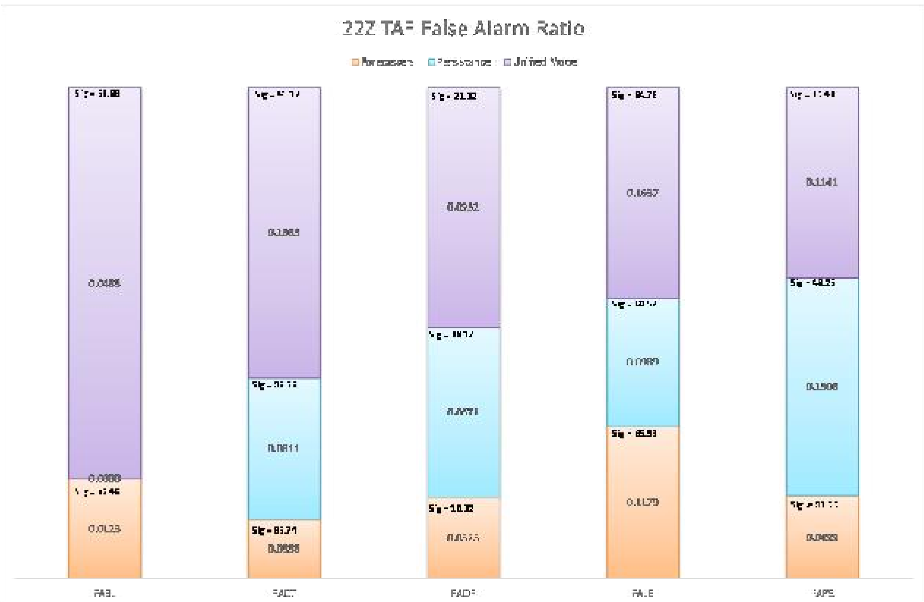

Figure 4.9: 22Z TAF False Alarm Ratio

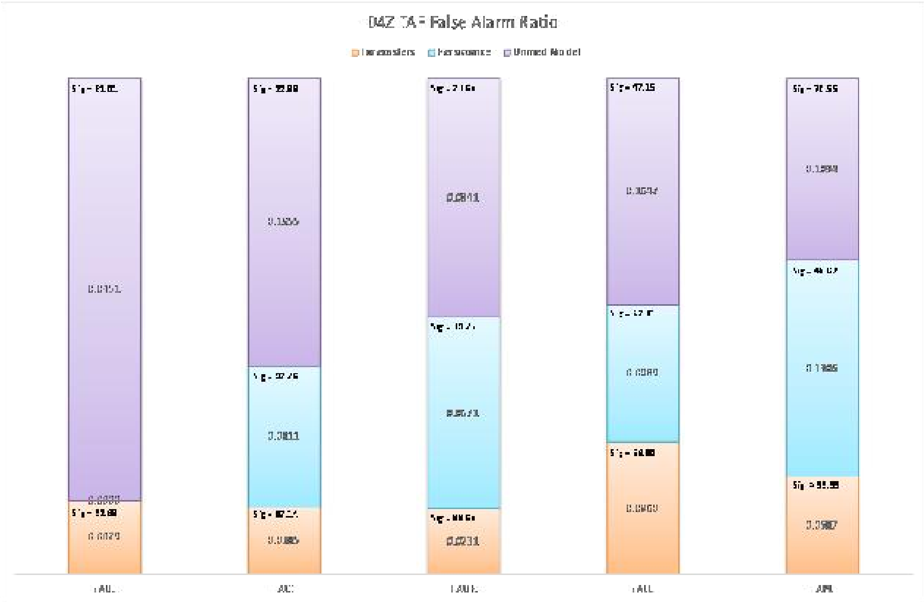

Figure 4.10: 04Z TAF False Alarm Ratio

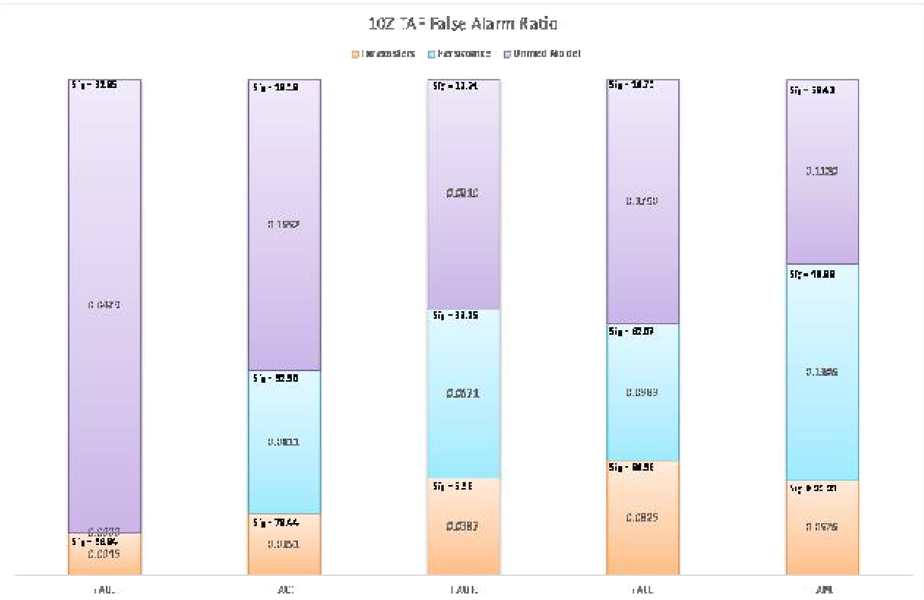

Figure 4.11: 10Z TAF False Alarm Ratio

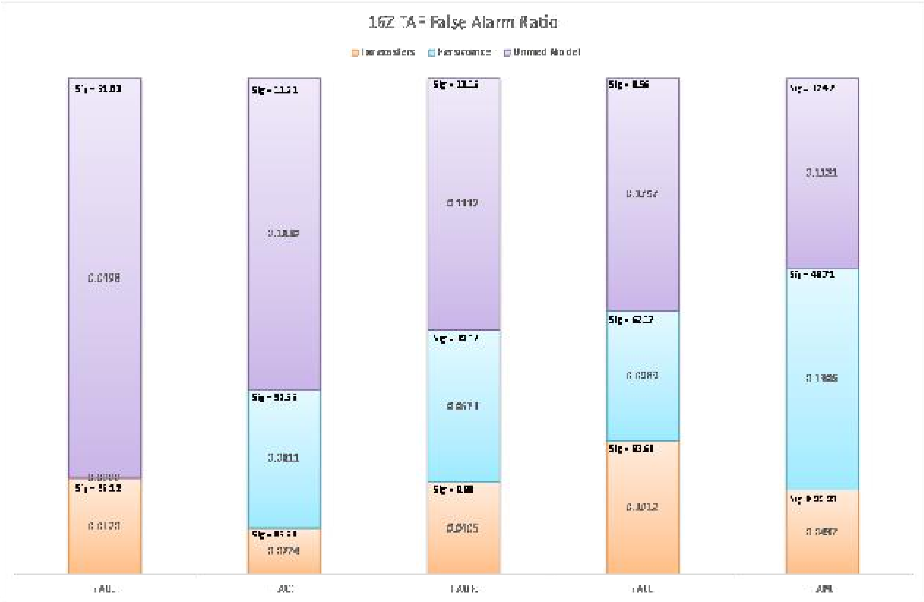

Figure 4.12: 16Z TAF False Alarm Ratio

4.3.1 BLOEMFONTEIN BRAM FISCHER AIRPORT (FABL)

The contingency tables for the persistence forecasts of FABL only had hits and misses. Therefore, no false alarm values were recorded. This means that no FAR values could be calculated. The FAR for the human-forecaster data sets are below 1.5%. The variance for human-forecaster false alarm rate is 0.776%. The FAR values for the UM data were 4.5%, which is also very low. The variance for UM FAR is 0.434%.

Significant values for FAR obtained from the Monte Carlo method are very volatile. The human-forecaster managed to reach significant values for the 10Z and 16Z TAF messages, with significant values of >95%. However, the significance values drop to 92.68% for the 04Z TAF messages and 82.46% of the 22Z TAF. The main reason that these significance values tend to be higher with regard to the afternoon and evening TAF messages is that significant weather is more common in the afternoon and evenings when the forecaster can get the timing better using now-casting tools. The UM FAR significance values are low, generally between 51 and 62%, and the 10Z TAF messages drop down to only 32.85%.

The FAR is not really significant, as chance can yield better results. The human-forecaster's forecast did manage to obtain significance at times and is lower than the UM. Therefore, the human-forecaster takes the lead.

4.3.2 CAPE TOWN INTERNATIONAL AIRPORT (FACT)

FAR values for human-forecaster are 2.4%. The variance for human-forecaster FAR is 1.112%. The UM FAR values are 16 - 19%. This is close to the upper limit of 20% enforced by SAWS on forecasts. The variance for the UM FAR is 1.781%. The persistence forecast achieved a FAR value of 6.7%.

The Monte Carlo method significance values for FAR are again volatile. The human-forecaster did achieve a significant value of 99.74% for their 22Z TAF messages, but the other human-forecaster data sets were not significant (04Z was 87.14%; 10Z was 78.44% and 16Z was 85.68%). The UM achieved very low significance values. All the significance for the UM were <42%. The significance of the UM FAR dropped as low as 11.21% for the 16Z TAF messages. The persistence data sets had significant values of 92 - 93%.

If one considers that only one data set was significant out of the twelve data sets and the relative high FAR values from the UM, the human-forecaster should get the nod for FAR.

4.3.3 JOHANNESBURG OR TAMBO INTERNATIONAL AIRPORT (FAOR)

FAR values are <5% for the human-forecaster data sets. The variance of the human-forecaster FAR is 1.742%. FAR values for the UM are 8 - 12%. The variance for FAR values with regard to the UM is 2.71%.The significance of FAR values obtained using the Monte Carlo method is low. The only forecast data set that was anywhere near significant was the 04Z TAF message of the human-forecaster with a value of 88.68%. The 16Z TAF messages, also by the human-forecaster, have a significant value of 0.98%! Except for the 04Z TAF messages of the human-forecaster, all the significant values were <40%.

From the above discussion, it is clear that better FAR can be obtained by chance. With the significance being so low, no real winner can be chosen, although the 04Z human-forecaster data set came closest to being significant.

4.3.4 DURBAN KING SHAKA INTERNATIONAL AIRPORT (FALE)

The FAR values of the human-forecaster are 8-12%. The variance for the FAR is 3.542%. The FAR values for the UM are 16-18%. The variance for the UM FAR values is 1.298%. The FAR value for persistence is 9.6%.

Significance values calculated using the Monte Carlo method were all below the 95% level of significance. The data set with the highest significance was the 16Z TAF messages by the human-forecaster, with a value of 93.68%. The UM FAR significance reached 8.56% for the 16Z TAF messages. Therefore, all three forecast systems are not better than chance in terms of FAR.

Owing to all data found to be not significant, with regard to FAR, no call can be made on which system is superior. It does seem that FAR values can be improved by chance.

4.3.5 PORT ELIZABETH AIRPORT (FAPE)

Human-forecaster FAR values are 4.6%. The FAR variance for human-forecaster is only 0.902%. The UM FAR values are 10 - 12%. The FAR variance for the UM is only 0.478%. The persistence FAR value is 13.06%.

The resulting significance values of the Monte Carlo method are quite interesting. The human-forecaster data sets were all >99.99% significant, while the UM and persistence data set were not significant at all. The highest value achieved by the UM was 70.55% for the 04Z TAF messages and the persistence data sets were all just above 48%.

The FAR values of the human-forecaster were around 5% less than those of the UM FAR. Also, the human-forecast significance with regard to FAR was significant at >99.99%, while the UM only managed a significance of 70.55% for the 04Z TAF data set. Therefore, the human-forecaster win this round.

4.4 CRITICAL SUCCESS INDEX

Figures 4.13, 4.14, 4.15 and 4.16 show the CSI values for the four sets of TAF messages. The significance values calculated using the Monte Carlo method are also displayed on the graphs as values.

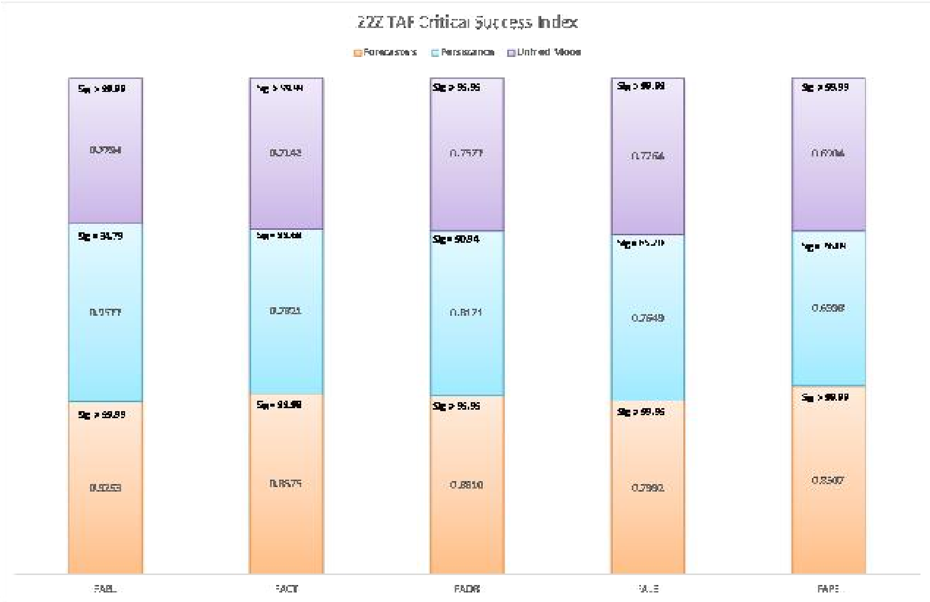

Figure 4.13: 22Z TAF Critical Success Index

Figure 4.14: 04Z TAF Critical Success Index

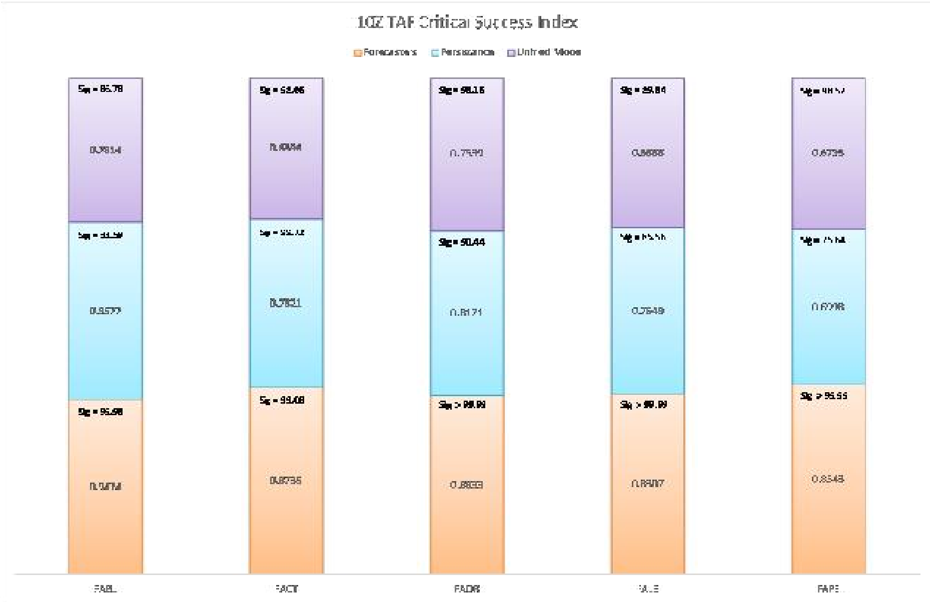

Figure 4.15: 10Z TAF Critical Success Index

Figure 4.16: 16Z TAF Critical Success Index

4.4.1 BLOEMFONTEIN BRAM FISCHER AIRPORT (FABL)

The CSI investigates both hits and misses. Therefore, forecast systems that fared well in both the H and FAR, should do well. The CSI values for human-forecaster are 92.95%. The variance of CSI values is 2.63%. The UM CSI values are 77 - 79%, but the 04Z TAF messages increased to 95.77%. This increase has given the variance for CSI a high value of 17.828%. The persistence CSI value is 95.77%.

Both the human-forecaster and UM scored statistical significant values in the Monte Carlo method for CSI. The UM did drop below the 95% cut-off for the 10Z TAF messages (86.78%) and 16Z TAF messages (91.92%). Persistence was only at 33.35% for CSI.

Persistence scored the highest value, but was not significant for the CSI. Whilst the UM did out-score the human-forecaster in one data set, it fell short in the other three data sets, for the CSI. The UM was also not always significant, whereas the human-forecaster was always significant for CSI values. Therefore, the human-forecaster win this round.

4.4.2 CAPE TOWN INTERNATIONAL AIRPORT (FACT)

The CSI values for human-forecaster are 85 - 88%. The variance of these CSI values for the human-forecaster is 1.726%. The CSI values of the UM are 64 - 79%. This, as was the case in FABL, leads to a high variance for the UM CSI, which is 13.471%. The persistence forecast CSI is 78.21%.

Statistical significance values for the CSI obtained with the Monte Carlo method for human-forecaster are all >96%, and therefore all significant. The CSI values for the UM are >98%, but the 16Z TAF data sets were only 84.80% significant. The significance for persistence was all >99%.

All three forecast systems were found to be significant, except the 16Z TAF data set of the UM. The human-forecaster scored higher values than the UM and therefore win this round.

4.4.3 JOHANNESBURG OR TAMBO INTERNATIONAL AIRPORT (FAOR)

CSI values for human-forecaster were 88 - 89%. The variance of the CSI score for human-forecaster was only 0.564%. The CSI values for the UM were 73.82%. The variance of the CSI values of the UM was 8%. The persistence forecast value for the CSI was 81.2%.

The results for significance from the Monte Carlo method for the CSI for human-forecaster were all >99% significant. The UM significance for the CSI was also significant with three of the data sets being >99%, but the value of the 16Z TAF messages dropped to 92.38%. The persistence forecasts also yielded a high value of just over 90% for all data sets for CSI.

With both the human-forecaster and UM being significant for the CSI, the winner is the human-forecaster for achieving higher values throughout.

4.4.4 DURBAN KING SHAKA INTERNATIONAL AIRPORT (FALE)

The CSI values for human-forecaster are 79 - 84%. The variance of these CSI values for human-forecaster is 3.152%. The CSI values for the UM are 68 - 77%. The variance of the CSI values for the UM is 7.874%. The persistence forecast CSI value was 76.49%.

The Monte Carlo method obtained significance values for the CSI were again >99% significant for the human-forecaster data sets. The significance values of the UM for the CSI are a mixed bag of results. The 04Z and 22Z TAF messages achieved >99% significance, but the 10Z TAF messages only had a significance value of 89.84%. The 16Z TAF messages fared even worse, with a significance score of 73.43%. The significance of the persistence forecast data sets for the CSI are all just over 85%.

All the human-forecaster data sets were significant, but only half of the UM data sets were significant. If you consider that the human-forecaster achieved higher values as well, the human-forecaster takes this round as well.

4.4.5 PORT ELIZABETH AIRPORT (FAPE)

The CSI values for human-forecaster are 83 - 87%. The variance of the CSI values for human-forecaster is 2.283%. The CSI values of the UM are 67 - 70%. The variance of the CSI values for the UM is 2.058%. The CSI value for the persistence forecast is 69.98%.

The significant values obtained using the Monte Carlo method for the human-forecaster data sets are all >99% significant. The significant values for the UM are, as was the case with Durban King Shaka Airport, only half >99% significant (22Z and 04Z TAF messages). The 10Z TAF messages are also significant with a score of 98.52%, but the significance of the 16Z TAF messages dropped to 91.5%. The significant values of the persistence forecast data sets are all just above 75% for the CSI.

The human-forecaster data sets are all significant; the UM 16Z TAF messages are not. The CSI scores of the human-forecaster are also consistently higher than the UM. Therefore, the human-forecaster take this round.

4.5 HEIDKE SKILL SCORE

Figures 4.17, 4.18, 4.19 and 4.20 show the HSS values for the four sets of TAF messages. The significance values calculated using the Monte Carlo method are also displayed on the graphs as values.

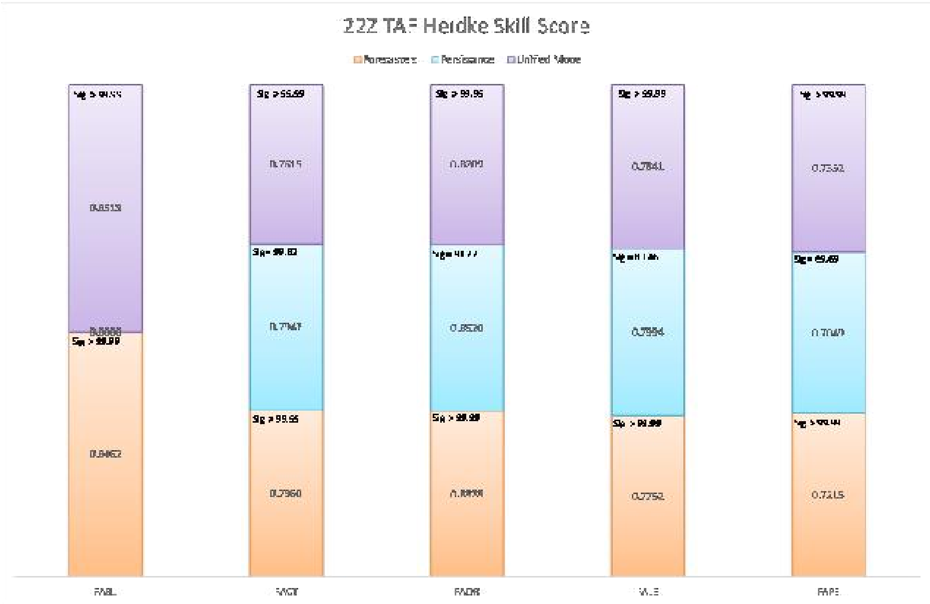

Figure 4.17: 22Z TAF Heidke Skill Score

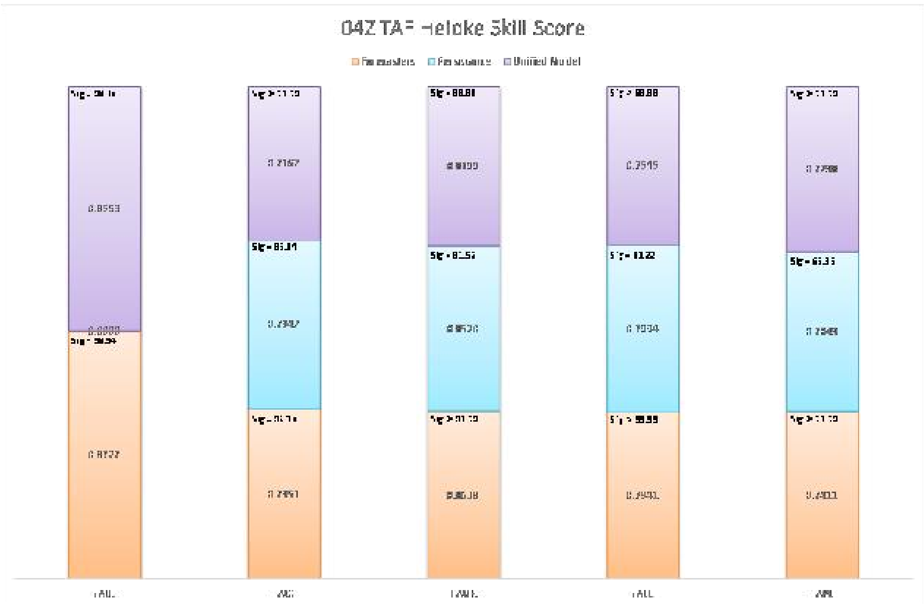

Figure 4.18: 04Z TAF Heidke Skill Score

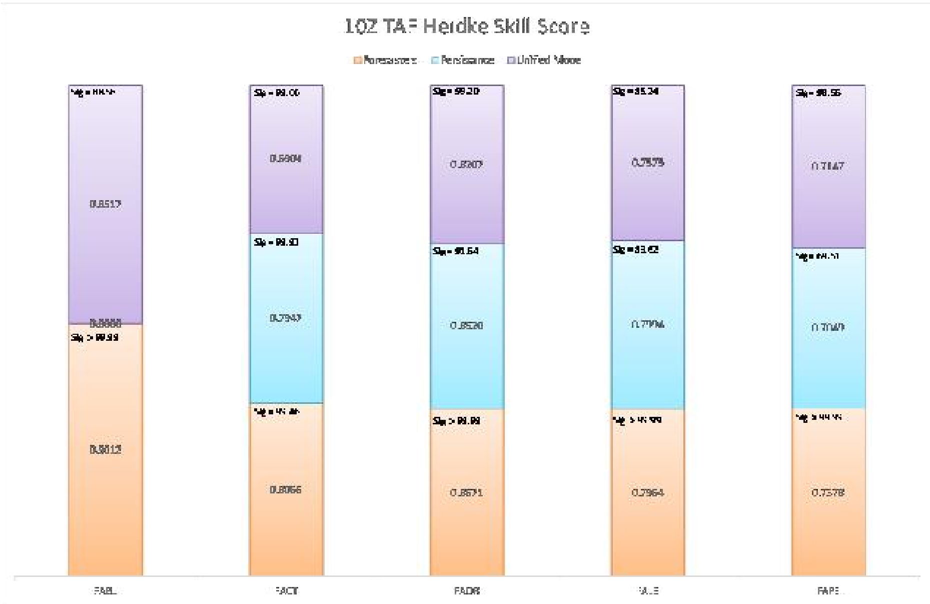

Figure 4.19: 10Z TAF Heidke Skill Score

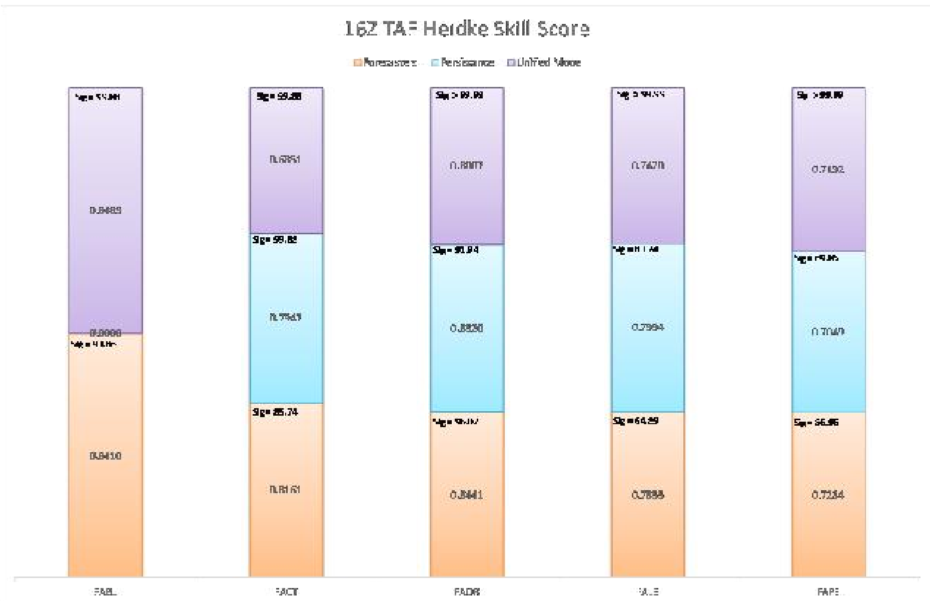

Figure 4.20: 16Z Heidke Skill Score

4.5.1 BLOEMFONTEIN BRAM FISCHER AIRPORT (FABL)

The HSS values for human-forecaster are 84 - 91%. The variance for the HSS for human-forecaster is 6.019%. The HSS values for the UM are 85 - 86%. The variance of the HSS values for the UM is 0.744%. Since the persistence data sets had no false alarm values at FABL, the HSS could not be calculated.

The Monte Carlo method obtained significance values of the HSS for the human-forecaster data sets are significant >98%, except for the 16Z TAF messages, where significance is 93.86%. The significant values for the UM HSS show the same pattern as for the human-forecaster. All the data sets are significant >98%, but the 10Z TAF messages have a significance value of 88.58%.

HSS values have similar values and significance. Therefore no clear decision can be made. Both forecast systems are performing at the same level with regard to the HSS.

4.5.2 CAPE TOWN INTERNATIONAL AIRPORT (FACT)

The HSS values for the human-forecaster are 79 - 82%. The variance for the HSS values for human-forecaster is 2.106%. The HSS values for the UM are 68 - 77%. The variance of the HSS for the UM is 7.642%. The persistence forecast HSS value is 79.47%.

The Monte Carlo method obtained significant values for the HSS values of the human-forecaster are all significant >98%, except the 16Z TAF messages with a significance score of 93.86%. The significance of the UM Heidke skill were all significant except for the 10Z TAF messages, which has a significance value of 88.58%. This result is similar to the result found at FABL.

The significance of the HSS values is similar for both the human-forecaster and the UM. The human-forecaster out-performed the UM with higher scores and therefore wins this round.

4.5.3 JOHANNESBURG OR TAMBO INTERNATIONAL AIRPORT (FAOR)

The HSS values for human-forecaster are 84 - 87%. The variance of the HSS values of the human-forecaster is 2.296%. The HSS values of the UM are 80 - 83%. The variance of the HSS values for the UM is 2.017%. The HSS of the persistence forecast is 85.2%.

From the Monte Carlo method, the HSS significance values calculated for the human-forecaster are all significant >96%. The HSS significance values for the UM are all significant >99%. The HSS significance for the persistence forecast data sets are 91.92%.

Both the human-forecaster and UM data sets are significant. The HSS values do not differ by a large degree. The human-forecaster have a slight edge, but the resulting values are close. Both forecast systems are basically on par, as the slightly higher values of the human-forecaster are weighed against the higher significance of the UM.

4.5.4 DURBAN KING SHAKA INTERNATIONAL AIRPORT (FALE)

The HSS values for the human-forecaster are 77 - 80%. The variance of the HSS values for the human-forecaster is 2.125%. The HSS values for the UM are 73 - 79%. The variance of the HSS for the UM is 4.663%. The HSS for the persistence forecast is 79.94%.

The Monte Carlo method obtained HSS significance for the human-forecaster are significant for three of the data sets. The16Z TAF messages dropped to 64.89%. The significance values of the HSS for the UM are all significant. The HSS significant scores for the persistence forecast is just above 83%.

The values of the human-forecaster and UM for the HSS are similar. The UM was totally significant, whereas the human-forecaster was totally significant for three of its data sets. There is therefore no real difference in the two forecast systems with regard to the HSS.

4.5.5 PORT ELIZABETH AIRPORT (FAPE)

The HSS values for human-forecaster are 72 - 75%. The variance of the HSS for human-forecaster is 1.958%. The HSS values for the UM are 71 - 74%. The variance of the HSS values for the UM is 2.05%. The HSS for the persistence forecast is 70.49%.

The Monte Carlo method obtained significance values of the HSS for human-forecaster are all significant, except the 16Z TAF messages with a value of 86.96%. The significant scores for the UM based on the HSS are all significant >98%. The significance of the HSS for the persistence forecast is around 69%.

There is no real difference between the values of the HSS for the human-forecaster and UM. Both the human-forecaster and the UM are significant, although the 16Z TAF messages for the human-forecaster are not. Therefore, there is no difference between the two forecast systems with regard to the HSS.

4.6 PIERCE SKILL SCORE

Figures 4.21, 4.22, 4.23 and 4.24 show the PSS values for the four sets of TAF messages. The significance values calculated using the Monte Carlo method are also displayed on the graphs as values.

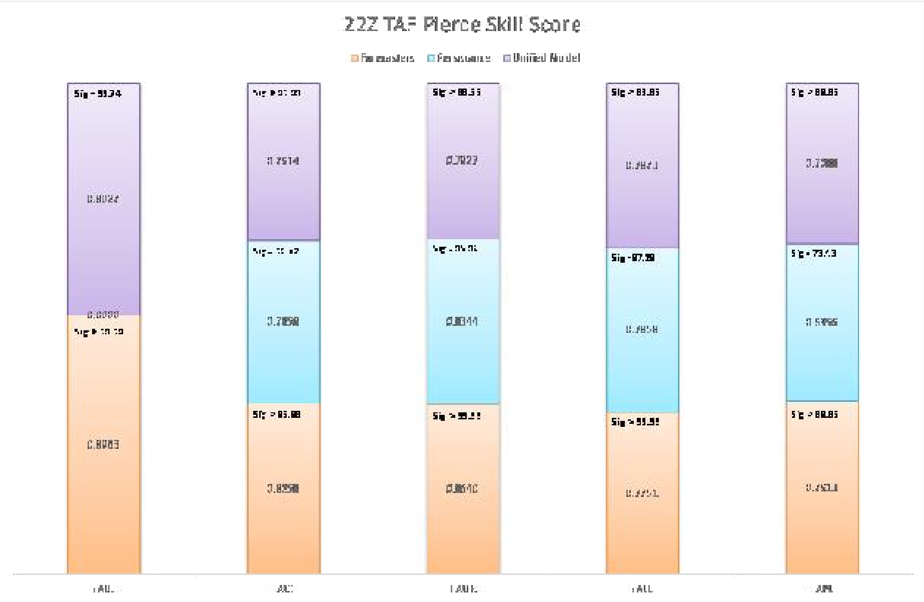

Figure 4.21: 22Z TAF Pierce Skill Score

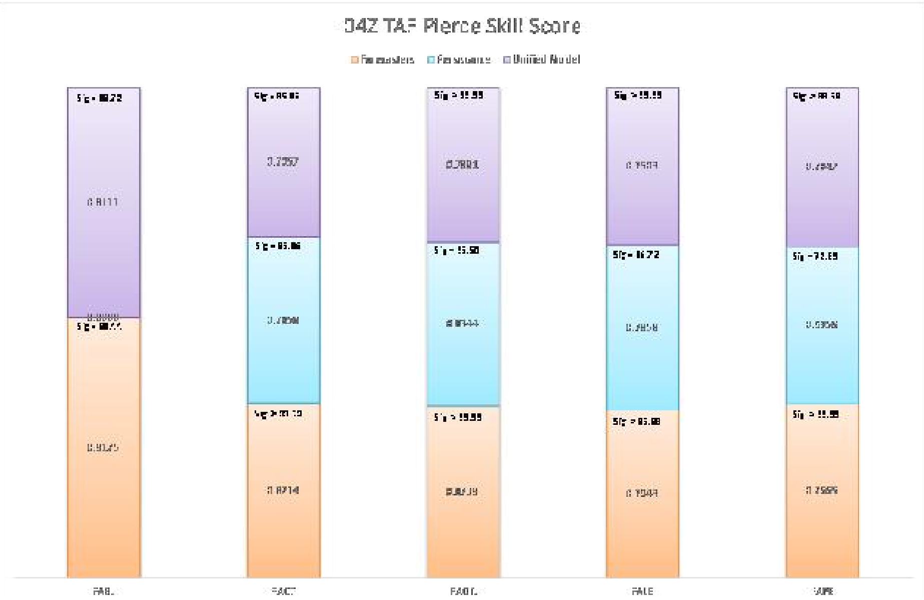

Figure 4.22: 04Z TAF Pierce Skill Score

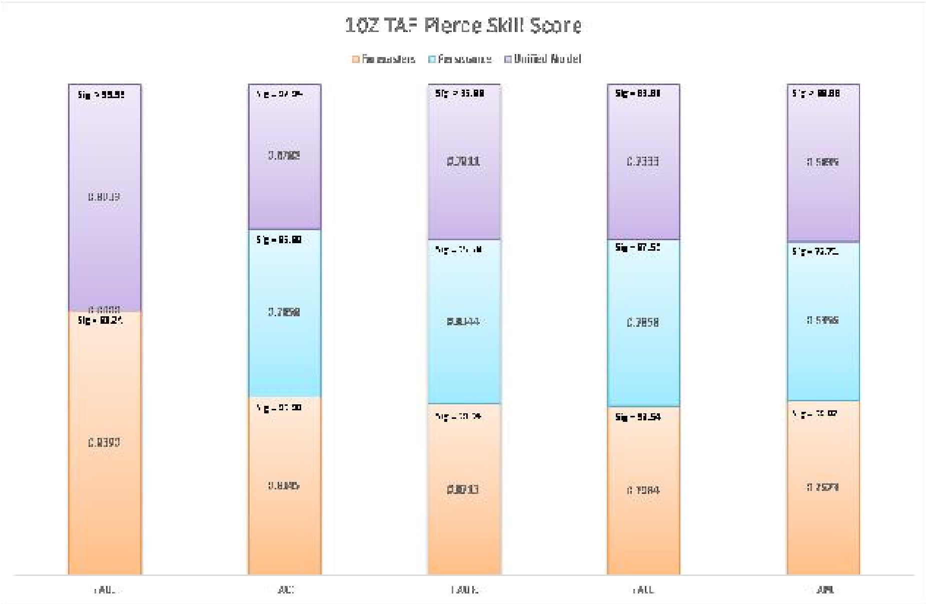

Figure 4.23: 10Z TAF Pierce Skill Score

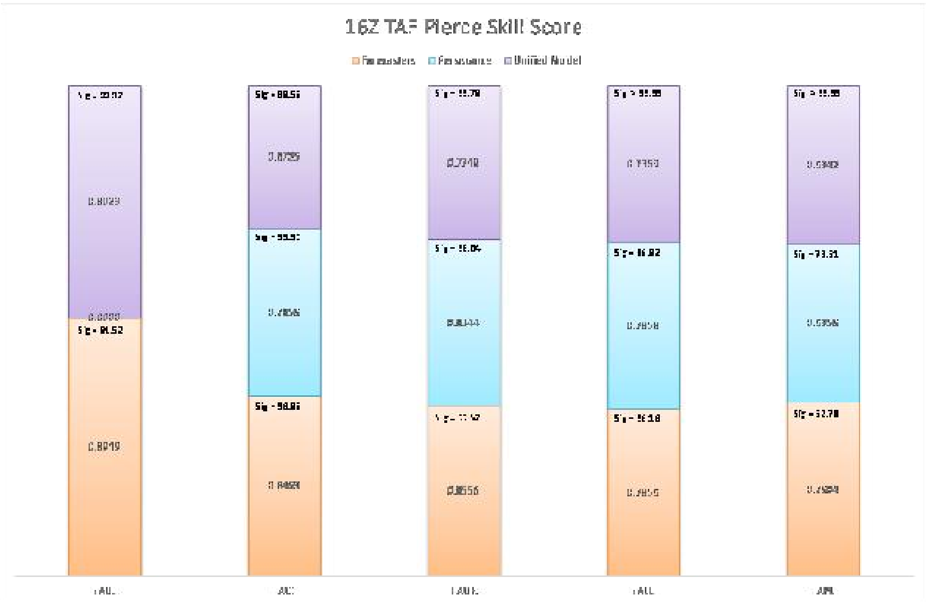

Figure 4.24: 16Z TAF Pierce Skill Score

4.6.1 BLOEMFONTEIN BRAM FISCHER AIRPORT (FABL)

The PSS values for human-forecaster are 89 - 94%. The variance of the PSS values of human-forecaster is 4.409%. The PSS values of the UM are 80 - 82%. The variance of the PSS for the UM is 0.843%. Once again, there are no false alarms for the persistence data sets at FABL, which resulted in the PSS not being calculated.

The Monte Carlo method was used to determine the significance values for the PSS. The significance values for the PSS for the human-forecaster are significant >98% for the 04Z and 22Z TAF messages, but the 10Z and 16Z TAF messages have significant values of 93 - 95%,, thus just missing the cut-off. The significance values of the PSS for the UM are all significant >98%.

The human-forecaster have higher PSS, but were not always significant. The PSS for the UM are all found to be significant, but the values are around 10% less than those of the human-forecaster. Therefore, the human-forecaster win this round.

4.6.2 CAPE TOWN INTERNATIONAL AIRPORT (FACT)

The PSS value for human-forecaster is 82 - 85%. The variance of the PSS for human-forecaster is 2.55%. The PSS values of the UM are 67 - 77%. The variance of the PSS values is 8.924%. The PSS value for the persistence forecast is 78.56%.

The significant values, obtained from the Monte Carlo method, for the PSS of the human-forecaster are all significant >96%. The significance values for the PSS for the UM are also all significant >95%. Even the persistence forecasts are significant >99% for the PSS.

All the forecasts for all three forecast systems are found to be significant. The human-forecaster have the highest values for the PSS and are therefore the winners in this round. The human-forecaster were also more consistent than the UM.

4.6.3 JOHANNESBURG O R TAMBO INTERNATIONAL AIRPORT (FAOR)

The PSS value for human-forecaster is 85 - 88%. The variance of the PSS for human-forecaster is 1.829%. The PSS of the UM are 77 - 80%. The variance of the PSS for the UM is 1.78%. The PSS value for the persistence forecast is 83.44%.

The Monte Carlo derived PSS significance values for human-forecaster are significant >99%. The PSS significance values for the UM are also significant >99%. Even the persistence forecast PSS significance is significant >95%.

All forecast systems are significant for the PSS. The human-forecaster has the highest values for the PSS and therefore win this round.

4.6.4 DURBAN KING SHAKA INTERNATIONAL AIRPORT (FALE)

The PSS values for human-forecaster are 77 - 80%. The variance of the Peirce skill score values for human-forecaster is 2.322%. The PSS values for the UM are 73 - 79%. The variance of the PSS values for the UM is 5.395%. The persistence forecast has a PSS of 78.58%.

The PSS significance values, obtained by the Monte Carlo method for the human-forecaster, are significant >96%. The PSS significant values for the UM are also significant >99%. The significance of the persistence model for the PSS are 86 - 88%.

Both the forecasts of the UM and the human-forecaster are significant. The differences in the values are not that great. The human-forecaster is slightly ahead, but really too close to call.

4.6.5 PORT ELIZABETH AIRPORT (FAPE)

The PSS values for the human-forecaster are 76 - 77%. The variance of the PSS for human-forecaster is 0.619%. The PSS values for the UM are 68 - 71%. The variance of the PSS values for the UM is 1.924%. The PSS for the persistence forecast is 69.56%.

The Monte Carlo method obtained significance values for the PSS for human-forecaster are >99% significant, with only the 16Z TAF messages dropping to 92.78%. The significance of the PSS for the UM are all >99% significant. The significance of the PSS for the persistence forecasts are around 72.74%.

All the data sets for the human-forecaster and UM are >99% significant, except the 16Z TAF messages of the human-forecaster. The human-forecaster do have a clear lead with regard to the higher PSS. This gives the human-forecaster a slight lead. In the next section the human-forecaster will be evaluated directly against the UM to determine the magnitude of the difference between the two forecast systems.

4.7 FORECASTERS VERSUS UM SKILL SCORE

To quantify the results for the forecaster versus UM skill score equation (3.34) will be used. The results will be categorized according to equation 4.1:

$$if\ FvUMSS = \begin{cases} |FvUMSS| \leq 10\% \rightarrow Neutral\ Result \\ 10\% < |FvUMSS| \leq 25\% \rightarrow Minor\ Victory\ Result \\ 25\% < |FvUMSS| \leq 50\% \rightarrow Major\ Victory\ Result \\ |FvUMSS| > 50\% \rightarrow Comprehensive\ Vitory\ Result \end{cases} \tag{4.1}$$

The sign of the FvUMSS will determine which one of the forecast systems achieved the winning result given equation 4.1. A positive value will hand the winning result to the human-forecaster as the forecast is better than the reference forecast. A negative value indicates that the reference value was better than the forecast; hence the winning result goes to the UM, which is the reference forecast. We will once again investigate the five airports as separate entities because the offices cannot be compared with one another owing to different climatic factors, resulting in different forecasting challenges.

4.7.1 BLOEMFONTEIN BRAM FISCHER AIRPORT (FABL)

Figure 4.25 shows the values of the human-forecaster versus the UM skill score for the six forecast indexes and the four data sets for FABL. Positive values are in favour of the human-forecaster and negative values favour the UM.

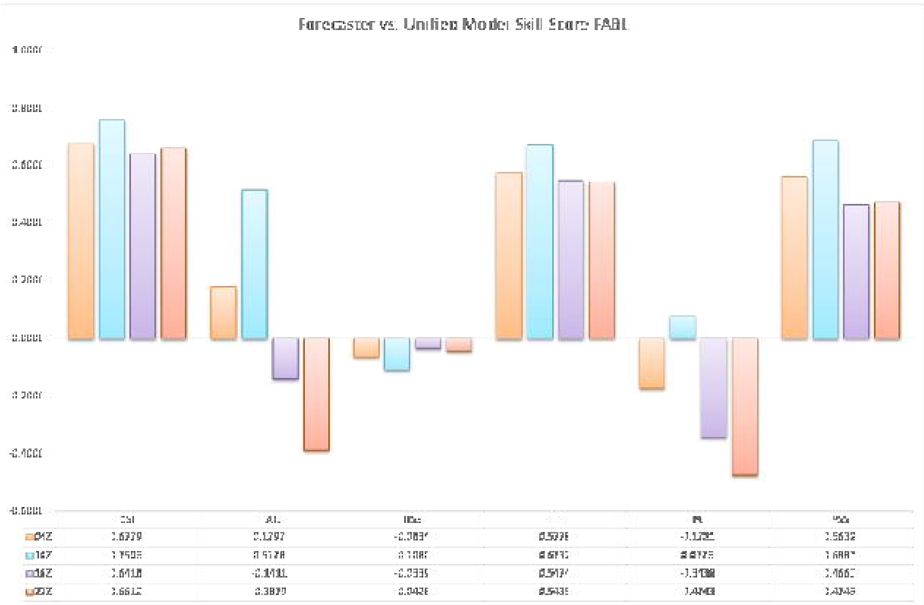

Figure 4.25: Forecaster vs. UM Skill Score FABL

The FvUMSS values for the PC are:

- 04Z – Minor victory for the UM.
- 10Z – Neutral result.
- 16Z – Major victory for the UM.
- 22Z – Major victory for the UM.

Thus, in terms of the PC, the UM scored one minor victory and two major victories. There is one neutral result. This gives the PC at FABL to the UM.

FvUMSS H scores show a resounding victory for the human-forecaster scoring comprehensive victories in all four time periods. This indicates the superior ability of human-forecaster to predict the onset of significant events.

The FvUMSS FAR values are:

- 04Z – Minor victory for the human-forecaster.
- 10Z – Comprehensive victory for the human-forecaster.
- 16Z – Minor victory for the UM.
- 22Z – Major victory for the UM.

From these values for the FAR, one can see that overall the scores of the two forecast systems are similar. Both forecasting systems scored a minor victory for one of the time periods, but the human-forecaster managed a comprehensive victory while UM scored a major victory. Thus, the human-forecaster takes the FAR for FABL.

The FvUMSS scores for CSI also yield a resounding victory for the human-forecaster, with all four datasets scoring comprehensive victories for the human-forecaster. These victories are mostly due to the comprehensive victories achieved in the H, since the CSI is dependent on both H and F.

The FvUMSS values for the HSS are all in favour of the UM. There is only one of the time periods (10Z) that recorded a minor victory, whilst the other time periods are neutral. The values show that both these systems are really on equal footing with regard to the HSS, but the UM is in the lead by a narrow margin.

The FvUMSS values for the PSS are heavily in favour of the human-forecaster. Two of the time data sets are major victories, while the other two are comprehensive victories. Thus, with the PSS being the difference between the H and the F, these results are to be expected. The human-forecaster beat the UM comprehensively regarding the H and this is reflected in the results for the PSS.

4.7.2 CAPE TOWN INTERNATIONAL AIRPORT (FACT)

Figure 4.26 shows the values of the human-forecaster versus the UM skill score for the six forecast indexes and the four time data. Positive values are in favour of the human-forecaster and negative values favour the UM.

The FvUMSS values are unique and resoundingly in favour of the human-forecaster. Only one of the FvUMSS scores was negative – the HSS for the 22Z TAF messages. In this instance, the UM scored a minor victory. Of the other FvUMSS scores, the human-forecaster scored three minor victories, four major victories and thirteen comprehensive victories. There were two neutral results as well. The FvUMSS indicates that the human-forecaster forecasting system is superior to the UM in all facets at FACT.

Figure 4.26: Forecaster vs. UM Skill Score FACT

4.7.3 JOHANNESBURG O R TAMBO INTERNATIONAL AIRPORT (FAOR)

Figure 4.27 shows the values of the human-forecaster versus the UM skill score for the six forecast indexes and the four time data sets. Positive values are in favour of the human-forecaster and negative values favour the UM.

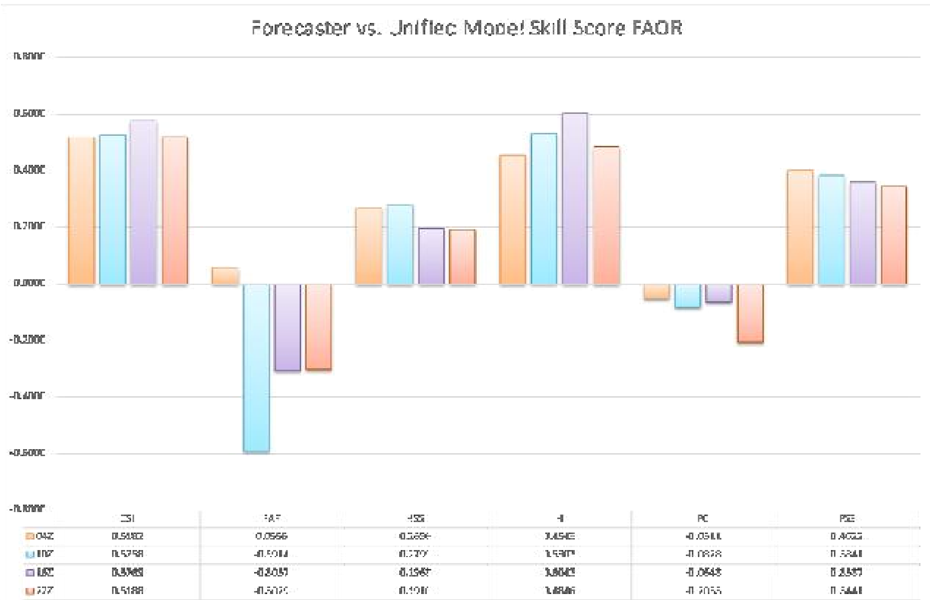

Figure 4.27: Forecaster vs. UM Skill Score FAOR

The FvUMSS values for PC are all for the UM. However, three of the time periods only recorded neutral results, while the 22Z time period barely managed a minor victory. Therefore, for the PC index, the UM has the slight edge.

The FvUMSS values for the H are resoundingly in favour of the human-forecaster. Two major victories and two comprehensive victories for the human-forecaster are recorded. This again shows the superior capability of the human-forecaster over the UM to anticipate the onset of significant weather.

The FvUMSS values for the FAR are in favour of the UM. The UM scored two major victories and one comprehensive victory for false alarm. One neutral result is also recorded. This result for the FAR shows that the UM is superior at determining the cessation of significant events.

Owing to the dominance shown by the human-forecaster, with the H FvUMSS scores, the FvUMSS scores for the CSI were also a resounding success. The human-forecaster scored comprehensive victories for the CSI.

The FvUMSS values for the HSS are also heavily in favour of the human-forecaster. Major victories have been achieved by the human-forecaster. The FvUMSS for the PSS are largely in favour of the human-forecaster, resulting in major victories for all four of the time periods.

4.7.4 DURBAN KING SHAKA INTERNATIONAL AIRPORT (FALE)

Figure 4.28 shows the results of the human-forecaster versus the UM skill score for the six forecast indexes and the four datasets. Positive values are in favour of the human-forecaster and negative values favour the UM.

The FvUMSS values for the PC are mostly neutral results, but in favour of the human-forecaster. The 22Z TAF has scored a major victory for the UM. Therefore, the UM takes PC.

The FvUMSS values for the H again show overwhelming results in favour of the human-forecaster. The human-forecaster scored four comprehensive victories for the H.

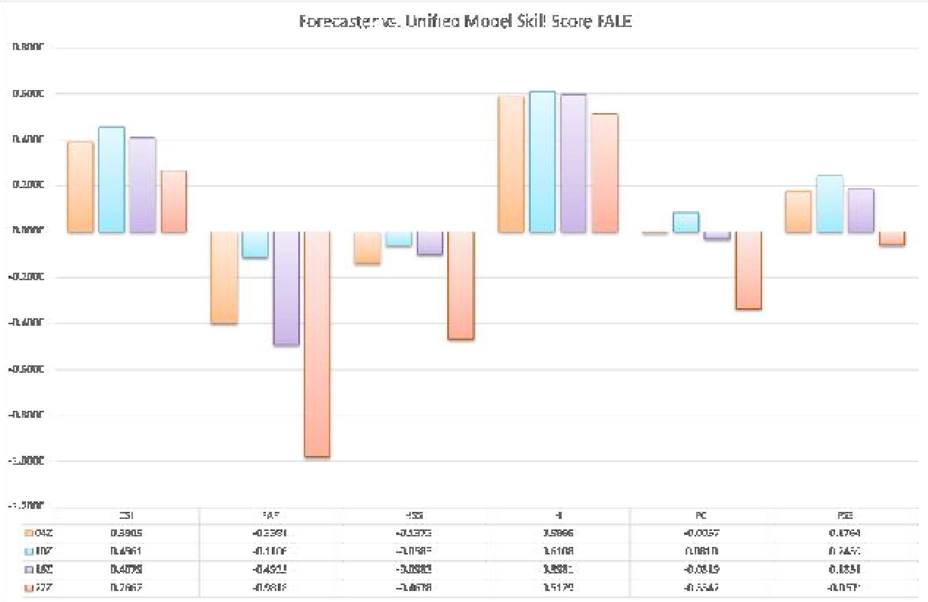

Figure 4.28: Forecaster vs. UM Skill Score FALE

The FvUMSS values for FAR are in favour of the UM. One minor victory, two major victories and one comprehensive victory were achieved for the FAR of the UM. The UM wins this one hands down.

Four major victories were achieved by the human-forecaster for the FvUMSS values of the CSI. These major victories are again as a result of the dominance over the UM with regard to H. The main reason that only major victories were achieved is the dominance of the UM with regard to the FAR. The H dominance is greater and resulted in the human-forecaster taking the CSI.

The FvUMSS values for the HSS are in favour of the UM. One minor and one major victory were achieved by the UM. There were also two neutral results, which were also in favour of the UM.

The FvUMSS values for the PSS favour the human-forecaster. Three minor victories were achieved by the human-forecaster. There was one neutral result slightly favouring the UM.

FALE is the airport in the study showing the strengths of the UM best. The UM managed to beat the human-forecaster at several instances. The UM achieved one comprehensive victory, four major victories and two minor victories – more than at any of the other airports. The human-forecaster, owing to their superiority in H, still managed four comprehensive victories, four major victories and three minor victories over the UM.

4.7.5 PORT ELIZABETH AIRPORT (FAPE)

Figure 4.29 shows the results of the human-forecaster versus the UM skill score for the six forecast indexes and the four datasets for FAPE. Positive values are in favour of the human-forecaster and negative values favour the UM.

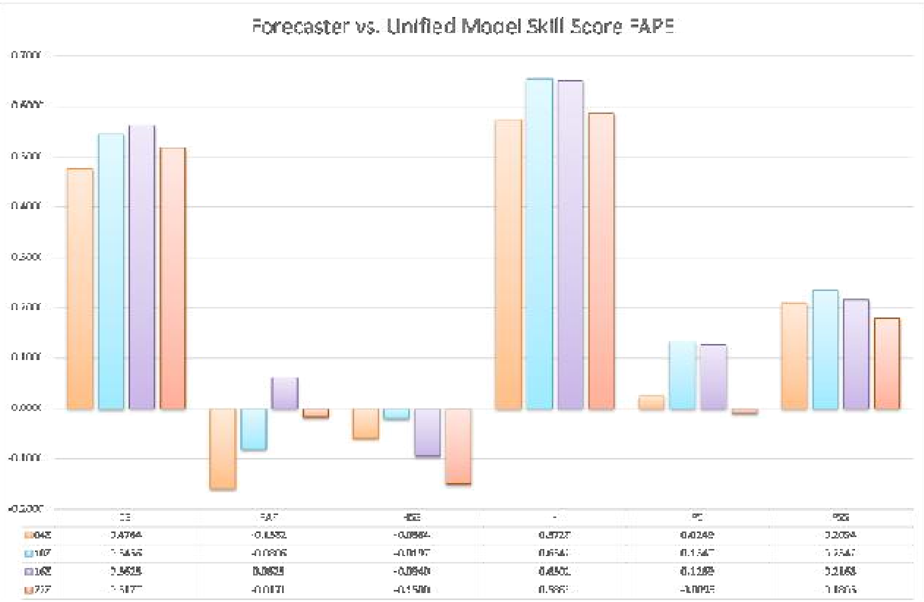

Figure 4.29: Forecasters vs. UM FAPE

The FvUMSS values for PC favour the human-forecaster. Two minor victories are recorded for the human-forecaster. There were also two neutral results, one for the human-forecaster and one for the UM. Therefore, the human-forecaster takes PC, but only just.

The FvUMSS values for the H are resoundingly in favour of the human-forecaster. The human-forecaster scored comprehensive victories for all four of the time periods. Again human-forecaster show utter dominance when it comes to the ability to detect the onset of significant weather.

The FvUMSS values for FAR are slightly in favour of the UM. Three of the time periods recorded neutral results, but the UM managed a minor victory with the 04Z TAF data set. Therefore, for the FAR, the UM just takes it.

The FvUMSS values for the CSI are again resoundingly in favour of the human-forecaster. Three comprehensive victories and one major victory are recorded for the human-forecaster. The main reason for the 04Z TAF data set missing out on the comprehensive victory is the minor victory that the UM scored in the FAR.

The FvUMSS values for the HSS are in favour of the UM. Three of the data sets are neutral, but the UM did manage to record a minor victory for the HSS. The UM takes the HSS, but with a small margin.

The FvUMSS values for the PSS favour the human-forecaster. Four minor victories were achieved for the human-forecaster. Again the dominance of the human-forecaster in the H led to this result – victory to the human-forecaster.

4.8 SUMMARY OF RESULTS

The results as discussed in the previous sections of this chapter show some clear indications with regard to the three forecast systems. Without comparing the airports with one another, the following can be concluded from these results:

1. PC (accuracy) is not an ideal indicator of the forecasting strengths of the three forecast systems. The PC values for all the forecasting systems are high. Both the human-forecaster and UM are significant. Results obtained through the FvUMSS for PC were very volatile.
2. The UM suffers severely with a low H. The UM H values were constantly >8% lower than the human-forecaster for the same time data set. These low H values resulted in the human-forecaster taking major and comprehensive victories for H, CSI and in some cases, PSS. When comparing the human-forecaster to the UM, the human-forecaster are superior in terms of H at all of the airports in this study. This is due to the ability of the human-forecaster to adapt to the situation, where the model can't change without being run again with new data. Even without this advantage (22Z and 04Z TAF data sets) the human-forecaster is still superior.
3. False alarms are a problem to qualify. The human-forecaster does out-perform the UM with regard to FAR, but all three forecast system lose out to chance when significance is taken in consideration. FAPE is a notable exception, where the human-forecaster are always significant.

The onset and cessation of significant weather conditions at an airport are the two most important factors for pilots. These significant weather conditions affect the planning and execution of the flight plans. The superiority shown by the human-forecaster with regard to H, does make the human-forecaster-generated forecasts more valuable to the aviation end user.

The results discussed in this Chapter cover the entire validation period from 1 February 2011 to 31 January 2013. These results indicate how the forecast systems fare on average. The question that still needs to be answered is how these systems fare during actual significant weather situations to see if the same results still hold for days where significant weather did occur. To examine this question various case studies will be conducted on days within this period, when significant weather occurred at one or more of the airports of this study. This will further enhance the hypothesis or raise interesting questions. These case studies are the topic of the next chapter.

CHAPTER 5

CASE STUDIES

In chapter 4, the results were discussed for the entire period from February 2011 to January 2013. In the discussion it was found that both the human-forecaster and the UM performed quite well, although it was found that the ability of the human-forecaster to detect the onset of significant or severe weather outperformed the UM in most cases. The question that still remains is whether the findings in chapter 4 are still valid when looking at significant or severe weather events.

In this chapter all events that caused SPECI criteria to come into effect are investigated. These events range from heavy rain owing to cut-off lows and thunderstorms, to fog events and strong winds. There were numerous days when significant weather events occurred in the regions where the airports are located. The days when these events occurred were isolated and the evaluation system was used on the event days only. This resulted in small data sets for the events. The results were examined and only the events that had usable data were retained. This process resulted in seven case studies to be discussed. One of these case studies, 7 April 2012, contained two of the airports (FACT and FAPE) being affected by the prevailing system.

The results for all eight airports examined in the case studies only yielded usable results for PC, CSI and H forecaster versus UM skill score. Therefore the discussion will be limited to these three verification indexes. The reason that the other values did not compute is the shortness of the period of the verification. The verification system works best on longer periods as short periods can result in no misses of false alarms which leads to division by zero. The case studies will be presented in chronological order.

5.1 26 MAY 2011

A cold front moved out of the north-eastern parts of southern Africa with the Atlantic Ocean high ridging in behind it, as indicated in figure 5.1. The ridging of the Atlantic Ocean high resulted in rain and showers in Kwa-Zulu Natal. This day yielded heavy rains in the area around FALE.

From the results given in figure 5.2, both the human-forecaster and the UM handled the event evenly. The FvsUMSS scores resulted in neutral values for CSI and H and the PC is a minor victory for the human-forecaster. From the results of this case study, the human-forecaster barely improved on the UM forecast.

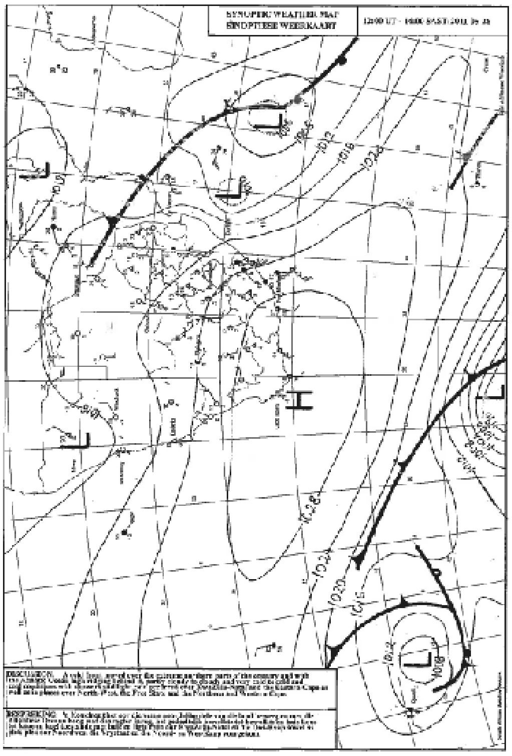

Figure 5.1: Synoptic Weather Map 26 May 2011, SAWS Weather Bulletin May 2011

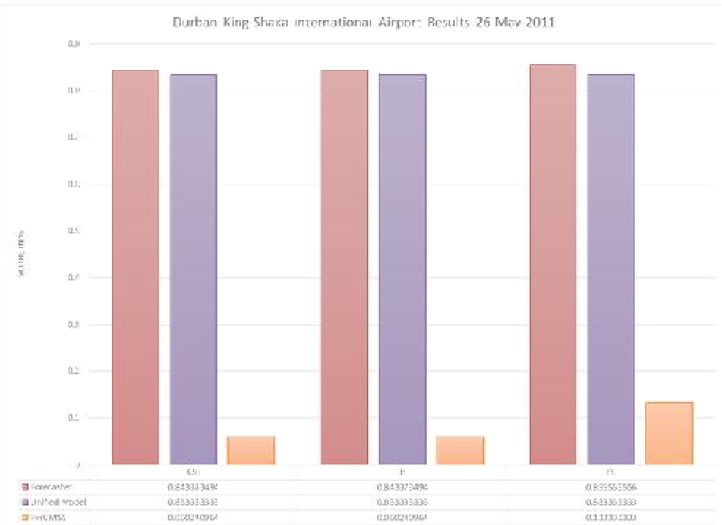

Figure 5.2: Durban King Shaka International Airport 26 May 2011

5.2 8 JUNE 2011

A low pressure system over the country, with an upper air cut-off low occurred on 8 June 2011, as indicated in figure 5.3. This synoptic configuration led to rain, showers and thundershowers over large areas of SA. The affected area included FABL.

From the results shown in Figure 5.4, the human-forecaster scored between 80% and 90% for all three verification indexes. The UM did not fare as well, scoring 68.75% for all three verification indexes. This result equates to two major victories and one comprehensive victory for the human-forecaster. Once again the superior H of the human-forecaster is coming to the fore. It is also notable that the PC score for the human-forecaster was almost 20% higher than that of the UM, which resulted in the comprehensive victory for the human-forecaster. Therefore, the human forecasters improved on the guidance given by the UM.

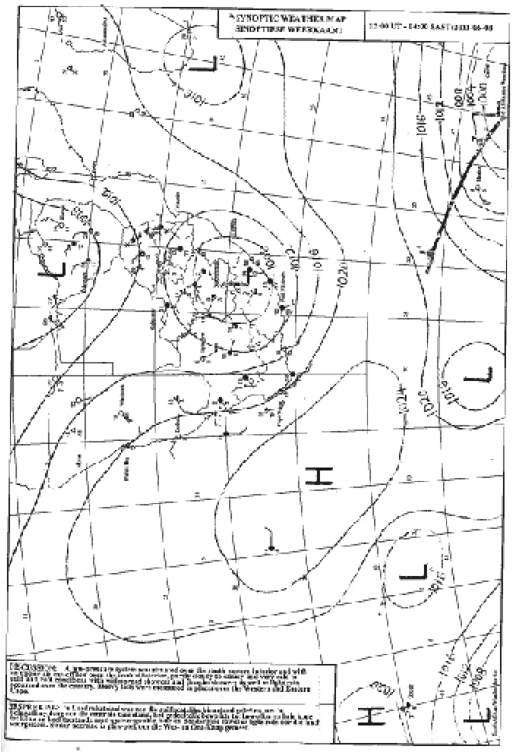

Figure 5.3: Synoptic Weather Map 8 June 2011, SAWS Weather Bulletin June 2011

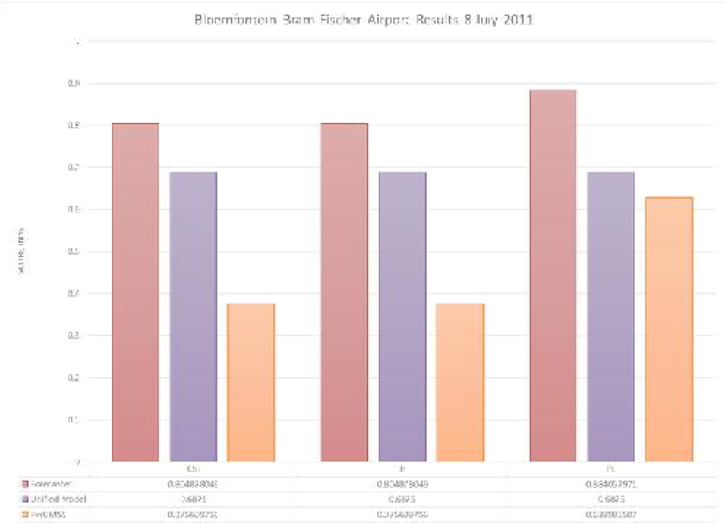

Figure 5.4: Bloemfontein Bram Fischer Airport Results 8 July 2011

5.3 11 NOVEMBER 2011

A tropical trough developed over the central interior of southern Africa extending over the Eastern Cape as indicated in figure 5.5. This configuration induced an easterly flow in Port Elizabeth resulting in heavy fog conditions over the airport. The heavy fog conditions reached the minimum operational criteria for FAPE (cloud base 200ft or lower). Planes were forced to divert to their respective alternate airports, awaiting the conditions to improve in order to land at FAPE. This case study is not a severe weather event, but a significant event – Air traffic to and from FAPE was adversely affected, costing the air travel companies millions of rands to accommodate stranded passengers.

What is interesting for this case is that all three values of the H, CSI and PC were exactly the same as indicated by figure 5.6. This outcome implies that for both the human and unified forecasts no false alarms were reported and no non-events. In this case study the UM scores poorly. The human-forecaster has scores of 88.89% where the UM only scored 56.52%. This low performance scores for the UM handed the human-forecaster three conclusive victories.

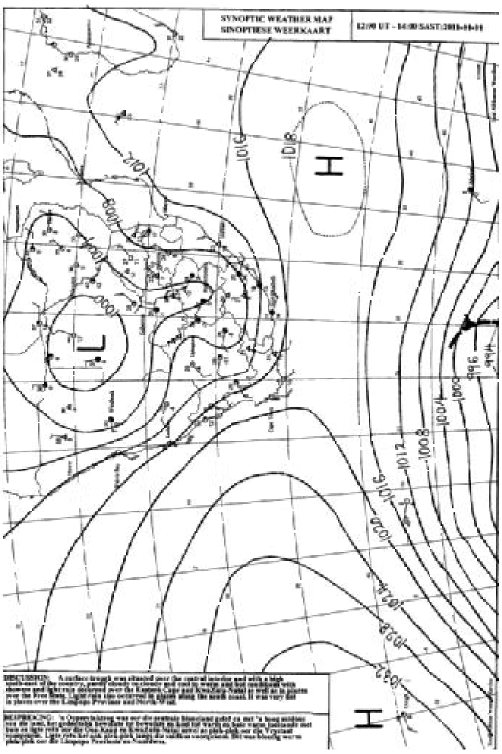

Figure 5.5: Synoptic Weather Map 11 November 2011, SAWS Weather Bulletin May 2011 (SAWS, 2014)

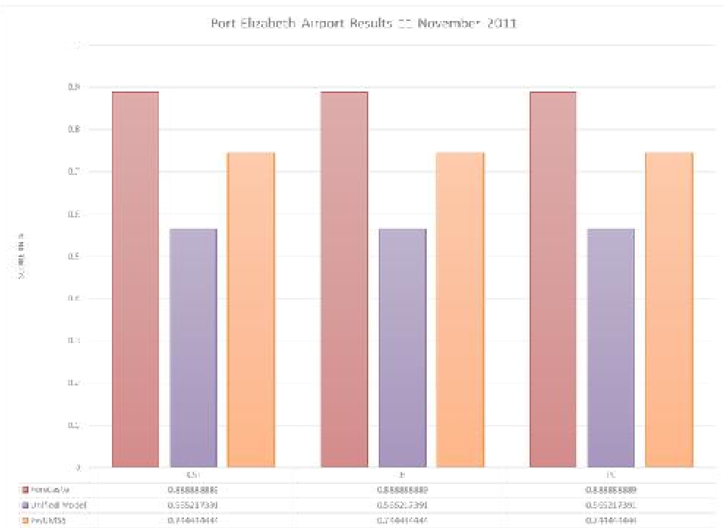

Figure 5.6: Port Elizabeth Airport Results, 11 November 2011

5.4 7 APRIL 2012

A frontal system and a tropical trough merged to yield widespread rain and showers as indicated on figure 5.7. On 7 April 2012 both FACT and FAPE recorded significant weather and both cases will be discussed. Of all the case studies 7 April 2012 presents the most interesting results.

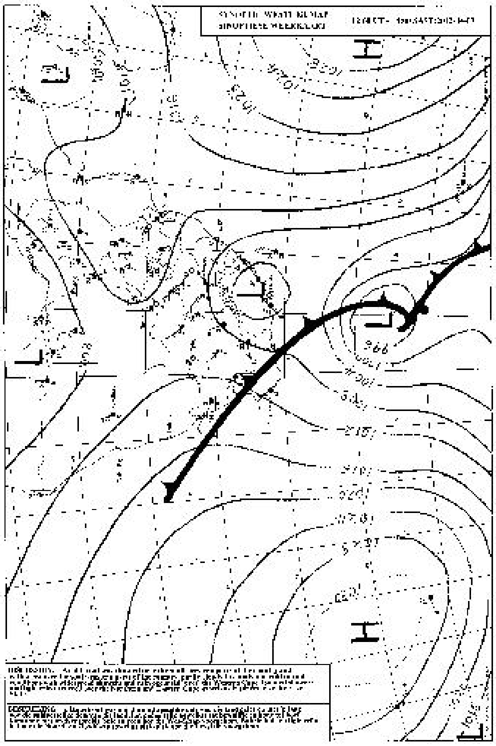

Figure 5.7: Synoptic Weather Map 7 April 2012, SAWS Weather Bulletin April 2012 (SAWS, 2014)

5.4.1 CAPE TOWN INTERNATIONAL AIRPORT

Here again there is an example of only hits and misses in the data for 7 April 2012 at FACT for both data sets given in figure 5.8. The human-forecaster scored 80% and the UM only managed 56.36%. This poor showing by the UM handed the human-forecaster three comprehensive victories.

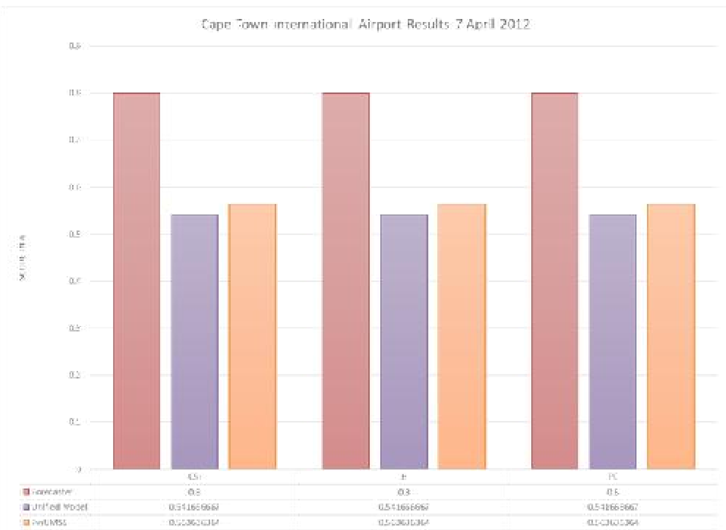

Figure 5.8: Cape Town International Airport Results 7 April 2012

5.4.2 PORT ELIZABETH AIRPORT

The results for FAPE for 7 April 2012 are astounding. This is an example of where the UM had provided a close to perfect forecast as indicated by figure 5.9. The UM had the same values for all three verification indexes, meaning it only had hits and misses. The UM managed to score 95.65%! The forecasters managed to score 93.33% for the H. The CSI was a respectable 80.77% and the PC was 81.48%. Whilst these scores by the human-forecaster were all above 80%, which is considered better than required, the UM was near perfect. This resulted in three comprehensive victories for the UM. Even the H, for which the human-forecaster achieved 93%, was comprehensively out-scored by the UM. This case study provided evidence that the NWP model can occasionally provide raw forecasts that are hard to beat.

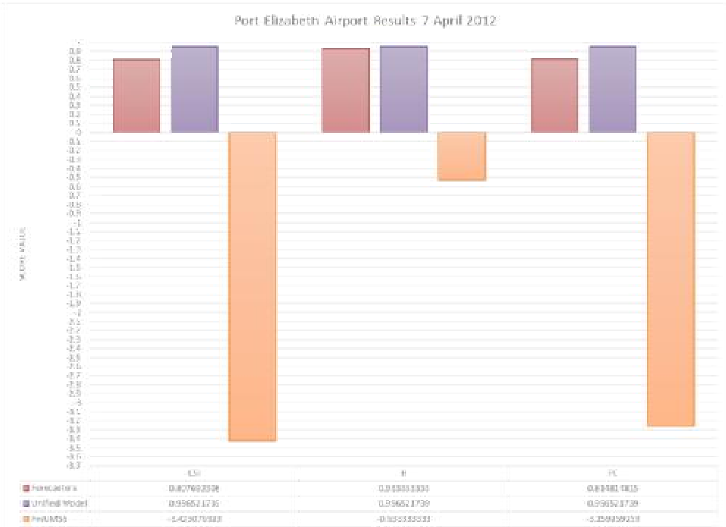

Figure 5.9: Port Elizabeth Airport Results 7 April 2012

5.5 6 – 8 JULY 2012

During the period 6 to 8 July 2012 a series of frontal systems and a well-developed upper trough moved in over the country yielding heavy rains and cold conditions as shown in figure 5.10. Bloemfontein, Cape Town and Port Elizabeth were affected, but only FACT yielded usable verification data, as the results calculated for Port Elizabeth and Bloemfontein resulted in a division of zeroes.

The human-forecaster managed scores of above 80% for the PC and H and close to 80% for the CSI as indicated on figure 5.11. The UM scored 65.5% for all three. This low score for the UM gives the human-forecaster one major victory and two comprehensive victories.

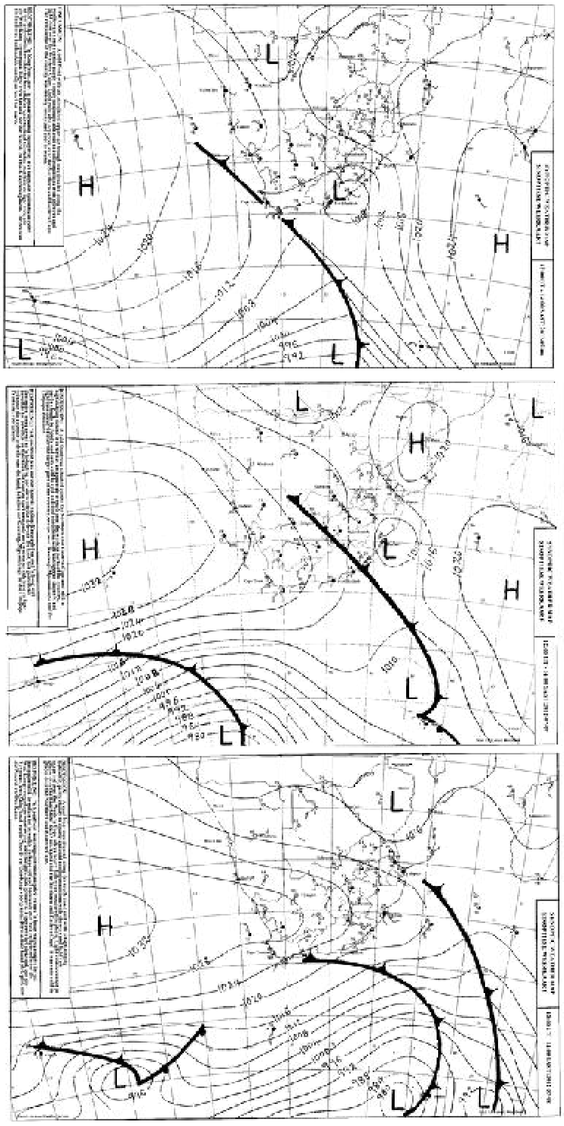

Figure 5.10: Synoptic Weather Maps 6 to 8 July 2012, SAWS Weather Bulletin July 2012 (SAWS, 2014)

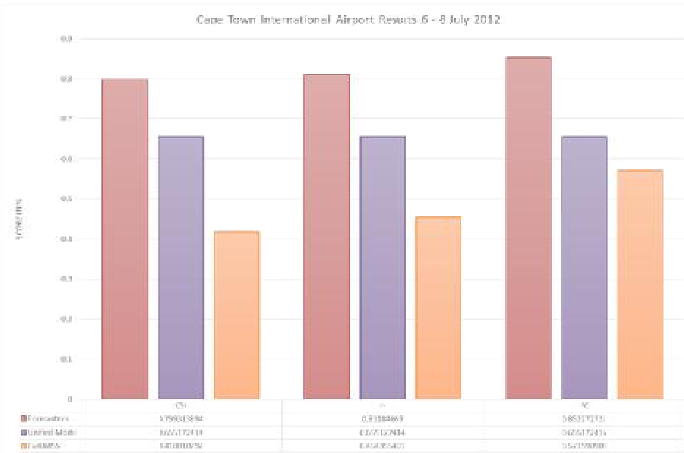

Figure 5.11: Cape Town International Results 6 - 8 July 2012

5.6 15 SEPTEMBER 2012

A tropical easterly trough developed over the western interior of Southern Africa as indicated in figure 5.12. This resulted in thunder showers over the eastern interior. Thunder showers with hail were reported in Johannesburg. Therefore, FAOR was investigated. This is the only thundershower related case study.

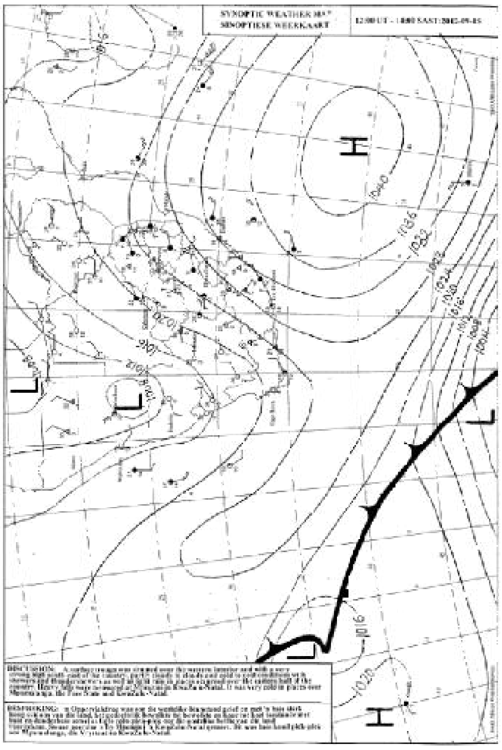

Figure 5.12: Synoptic Weather Map 15 September 2012, SAWS Weather Bulletin September 2012 (SAWS, 2014)

The human-forecaster scored 78% for the PC and 76.6% for the CSI as indicated in figure 5.13. The H was better at 84.5%. The UM scored 72% throughout for all three verification indexes. With the scores for PC and CSI being close, the human-forecaster scored two minor victories. The improvement in the score for the H gives the human-forecaster a major victory. This case study provides more evidence of the enhanced ability of the human-forecaster to predict the onset of significant weather.

Figure 5.13: Johannesburg OR Tambo International Airport Results 15 September 2012

5.7 10 – 11 DECEMBER 2012

A surface low with an associated cut-off low developed over Botswana and extended over the central interior as indicated in figure 5.14. This cut-off low resulted in heavy rains in Kwa-Zulu Natal. FALE was also affected and used for this case study.

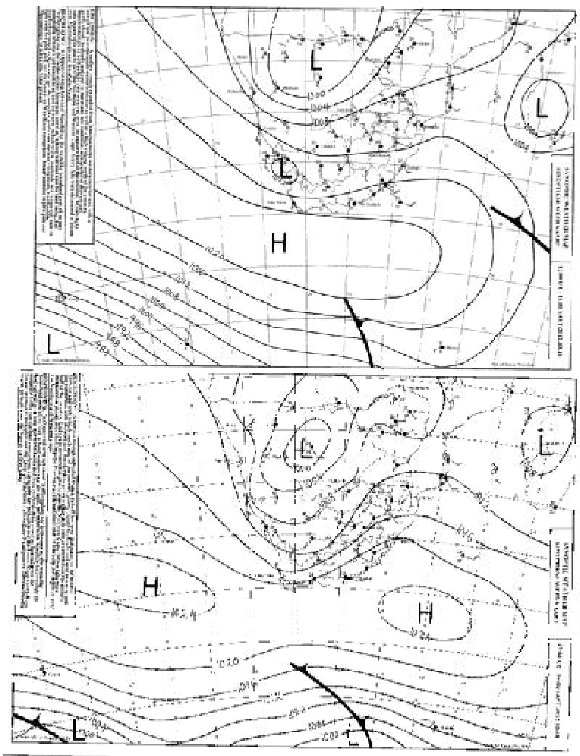

Figure 5.14: Synoptic Weather Maps for 10 - 11 December 2012, SAWS Weather Bulletin December 2012 (SAWS, 2014)

The human-forecaster scored 81.3% for the PC, 87.8% of H and 73.2% for the CSI as indicated on figure 5.15. The UM only managed 56.4% for all three indexes. Owing to the difference between the values of the CSI for the human-forecaster and UM being less than the other two verification indexes, the human-forecaster scored a major victory. The FvsUMSS scores for the H and PC were conclusive victories.

Figure 5.15: Durban King Shaka International Airport 10 - 11 December 2012

5.8 CONCLUSIONS DERIVED FROM CASE STUDIES

The above case studies support the results found for considering the entire study period. Moreover, these case studies support the notion that the human-forecaster have an edge in the detection of the onset of significant weather events. These findings are mostly due to the fact that the human-forecaster can adapt to changes and modify their forecasts accordingly. However, the case study for 7 April 2012 at Port Elizabeth provides evidence that the UM is a reliable model.

For Port Elizabeth on 7 April 2012 the UM timing of the significant event was spot-on, resulting in extremely high verification numbers. This result left the human-forecaster in its wake. Therefore, whilst the human-forecaster are more flexible to adapt to significant weather than the UM, the 7 April 2012 case study does provide evidence that the model can capture the timing and location of significant events. Unfortunately for the UM, this is more the exception to the rule, than the norm.

From the findings in Chapters 4 and 5, the point has been reached where the conclusion can be drawn and recommendations can be made. Has the hypothesis been proven? The answer will be provided in Chapter 6.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

In chapter 1 the background for this study was discussed and the need for this study was explained. Chapter 2 discussed the data used in this study. Chapter 3 discussed the way that aviation products are constructed. The methodologies used in this study to generate the data sets and conducting the tests were also discussed in chapter 3. Next, the verification system was applied to the data sets to derive the results of the study were also discussed in chapter 3.

These results were then discussed in chapter 4. Case studies were drawn up, in chapter 5, out of the data sets and the verification system was again applied to these case study data sets to determine whether the results from the period in totality are also reflected in the periods where significant weather events did occur at the airports. Now, we reached the point where the hypothesis can be evaluated.

6.1 CONCLUSION ON THE HYPOTHESIS

From Section 1.11 the hypothesis was defined as follows:

Terminal aerodrome forecasts issued by South African Weather Service forecasters for major airports in South Africa improve on raw numerical weather prediction model forecasts.

In Section 4.9, the summary of the results, three items of note were found:

1. PC is not an ideal verification measure to differentiate between forecast systems since all the forecasting systems score high PC values. The forecaster versus UM skill score values were also volatile
2. The H of the human-forecaster is superior to that of the UM
3. False alarm rates are problematic to quantify as better results seem to be yielded by chance.

From the case studies in Chapter 5, it was found that the case studies also agree with the findings for the total period. However, one case study, 7 April 2012 at FAPE, provided evidence of the value of the UM as an excellent weather forecast model.

Unfortunately, in general, the UM suffers a low H compared to that of the human-forecaster, which swings the pendulum soundly in favour of the human-forecaster.

Therefore, we look again at the question: do the human-forecaster employed by SAWS add value to the raw numeric model product forecasts? It was shown in Chapters 4 and 5 that human-forecaster indeed improve on the NWP models. There are, however, cases where the UM has out-performed the human-forecaster (Port Elizabeth 7 April 2012), but the human-forecaster, in general, out-performed the UM significantly over the two year sample period. In that fact lies the trap for human-forecaster to be cautious of – the numerical weather model is an excellent guide to the expected weather conditions, but it does have challenges when predicting the timing of extreme events. When the timing of the numerical weather model is spot-on, the results are accurate.

6.2 FUTURE OF NUMERICAL WEATHER PREDICTION MODELS

The research field of NWP is constantly improving the quality of the NWP models and the resolution, both horizontally and vertically, and is expected to continue to improve in future. Based on current trends global NWP model should improve its horizontal resolution to only few kilometres and the days the model can be used for can extend by two more days by 2030 (Thorpe, 2012).

It therefore implies that regional NWP models will have even greater resolution than their global counterparts. Research is also developing methods of improving the NWP output, by applying filters and bias to improve the guidance forecast offered to the forecasters. Thus, since the human forecasters improve on the guidance forecasts, the further improvement of NMP models will result in even more reliable aviation forecasts being issued by the aviation forecasting centres around the globe.

6.3 RECOMMENDATIONS TO AVIATION WEATHER SERVICE PROVIDERS

Human-forecasters are invaluable to the aviation forecast industry. The strongest advantage that human-forecasters have over NWP model output products is their adaptability. A human-forecaster can react if the weather conditions change to unexpected conditions and can issue amendments. A human-forecaster can, owing to this adaptability, update timing of significant events. This intervention leads to higher

H, which saves the aviation industry millions of rands annually. Knowing when significant weather is expected enables the aviation industry to plan more efficiently and reduces additional operational costs accrued when events occur earlier than expected.

There is, however, a major challenge to be investigated – the volatility of the cessation of significant events. The timing when significant events are to end is seemingly random. In the Monte Carlo test the FAR values scored low significance, suggesting that chance is better. Thus, even though the human-forecaster working for SAWS, scored well within the requirements set forth by ICAO, room for improvement still exists.

6.4 RECOMMENDATIONS TO THE AVIATION INDUSTRY

Human-forecasters are an invaluable resource available to the aviation industry. A pilot can get his or her weather information for free from web-based sites using raw model data only. These raw model forecasts are of a high quality and can seem to be accurate enough for use. But using raw model forecasts has two distinct pitfalls – the first being the inability of the model to react if the forecast does not go according to plan; and the second is the low H of the raw model output forecasts.

As was proven by the case study of FAPE on 7 July 2012, the numerical weather products can get the forecasts spot-on at times, but this is more the exception to the rule, than the norm. The NWP model is a tool to use to generate a forecast. Human-forecaster are trained to interpret these NWP model products and to apply scientific decision making techniques to arrive at the most likely conditions. Thus, human-forecasters use the numerical forecasting products with real-time observational data such as satellite, RADAR and personal observations to validate the NWP model product and subsequently, to produce a forecast. Human-forecasters are also able to adapt to changes in the weather and to amend the forecast accordingly.

Therefore, users of weather products in the aviation industry can use the NWP model products freely available to get an idea of what is expected. Most of these web-based weather forecasting sites are geared for general use. It is helpful for general weather, for surfers and even for skippers of ships out at sea. But aviation is a very specialized forecast and is best served by human-forecaster.

The vastly superior H obtained by the human-forecaster is also an important factor for the aviation industry. Higher H implies fewer diversions as the onset of significant weather events are detected more frequently. Thus, pilots know when significant weather is expected and can make their flight plans accordingly, with a smaller chance of disruption. In SA, as is the case in many countries around the world, the aviation industry pays for the weather products received. Why not then use the resources that are at the disposal of the aviation industry, and which in the long run delivers better results?

REFERENCES

Adeyemi, O., 2011. Measures of Association for Research in Educational Planning and Administration. *Research Journal of Mathematics and Statistics,* 25 09, 3(3), pp. 82-90.

Airports Company South Afirca, 2013. *ACSA Airport Passengers and Movents.* [Online]
Available at: http://www.acsa.co.za/home.asp?pid=117

Aneki.com, 2013. *Countries with the most airports.* [Online] Available at: http://www.aneki.com/airports.html?t=Countries+with+the+Most+Airports&table=fb193&places==*&order=desc&orderby=fb193.value&decimals=--&dependency=independent&number=25&cntdn=asc&r=seeRelated&c=&measures=Country--Airports&units=--&file=airports&map=on

Balzer, K., 1995. *TAF Verifikation. Eine Dokumentation der Methodik.* Offerbach: Deutscher Wetterdienst.

Chan, S. T., 2000. *TAF Improvement Developing of Aviation Forecast Verification System In Hong Kong, China.* Hong Kong, China, World Meteorological Orginazation.

CIA, 2013. The World Factbook 2013-2014. In: *The World Factbook 2013-2014.* Washington DC: Central Intelligence Agency.

Douglas, K., 2013. *How we made it in Africa.* [Online] Available at: http://www.howwemadeitinafrica.com/africas-top-three-airports-rank-higher-than-any-airports-in-the-usa/26124/
[Accessed 03 09 2013].

Ferro, C. A. T. & Stephenson, D. B., 2011. Extremal Dependence Indices: Improved Verification Measures for Deterministic Forecasts of Rare Binary Events. *Weather and Forecasting,* Volume 26, pp. 699-713.

Gordon, N., 1993. *Verification of Terminal Forecasts.* Vienna, VA, s.n., pp. 254-257.

Hilden, A., Kjær, O., Feldberg, R. & Hansen, J., 1998. *Forecast Verification Report September 1998,* Copenhagen: DMI.

Holton, J. R., 1992. An Introduction to Dynamic Meteorology. In: *An Introduction to Dynamic Meteorology.* Third Edition ed. London: Academic Press Limited.

ICAO, 2010. *International Standards and Recommended Practices: Annex 3 to the Convention on International Civil Aviation Meteorological Service for International Air Navigation.* Seventeenth ed. Montréal(Quebec): International Civil Aviation Organization.

IDRE, UCLA, 2013. *FAQ: How do I interpret odds ratios in logistic regression?.* [Online]
Available at: http://www.ats.ucla.edu/stat/mult_pkg/faq/general/odds_ratio.htm
[Accessed 31 10 2013].

Investopedia, n.d. *Chi Square Statistic.* [Online]
Available at: http://www.investopedia.com/terms/c/chi-square-statistic.asp
[Accessed 31 10 2013].

Jacobs, A. J. M. & Maat, N., 2004. Numerical Guidence Methods for Decision Support in Aviation Meteorological Forecasting. *Weather and Forecasting,* Volume 20, pp. 82 - 100.

Jolliffe, I. T. & Stephenson, D. B. eds., 2003. *Forecast Verification: A Practitioner's Guide in Atmospheric Science.* s.l.:John Wiley and Sons: Chichester.

Jolliffe, I. T. & Stephenson, D. B. eds., 2012. *Forecast Verification A Practitioner's Guide in Atmospheric Science.* Second ed. Chichester(West Sussex): Wiley-Blackwell.

King Shaka International, 2013. *King Shaka International Home Page.* [Online]
Available at: http://kingshakainternational.co.za/

Landman, S., 2012. *A multi-model ensemble system for short-range weather prediction in South Africa.* Pretoria: University of Pretoria.

Landman, S., Engelbrecht, F.A., Engelbrecht, C.J., Dyson, L.L & Landman, W.A,. 2012. A short-range weather prediction system for South Africa based on a multi-model approach. *Water SA,* 38(5), pp. 765-774.

Landman, W. A., Mason, S. J., Tyson, P. D. & Tennant, W. J., 2001. Retro-active Skill of Multi-tiered Forecasts of Summer Rainfall over Southern Africa. *International Journal of Climatology,* Volume 21, pp. 1 - 19.

Mahringer, G., 2008. Terminal aerodrome forecast verification in Austro Control using time windows and ranges of forecast conditions. *Meteorological Aplications*, 1(15), pp. 113-123.

Rae, K. J., Van Niekerk, H. B. & Poolman, E., 2013. *Work Instructions Severe Weather Warning System,* Pretoria: s.n.

Revering, A., n.d.. *Andrew Revering's List of Meteorological Formulas.* [Online] Available at: http://www.aprweather.com/pages/calc.htm [Accessed 2 June 2014].

SAWS, 2007. *Provide Aviation Weather Forecast and Warnings.* [Online] Available at: http://tqm.saws.local/qms/EM_Forecast_Warn/EM2__Provide_Avi_Weath_Forecast_Warnings.htm
[Accessed 8 10 2013].

SAWS, 2008. *Annual Report 2007/2008.* Pretoria(Gauteng): South African Weather Service.

SAWS, 2010. *Annual Report 2009/2010.* Pretoria(Gauteng): South African Weather Service.

SAWS, 2012. *Aeronautical Summary Bram Fischer International Airport 1996 - 2010.* Pretoria(Gauteng): South African Weather Service.

SAWS, 2012. *Aeronautical Summary Cape Town Internation Airport 1996-2010.* Pretoria(Gauteng): South African Weather Service.

SAWS, 2012. *Aeronautical Summary Durban International Airport 1996 - 2010.* Pretoria(Gauteng): South African Weather Service.

SAWS, 2012. *Aeronautical Summary OR Tambo International Airport 1996 - 2010.* Pretoria(Gauteng): South African Weather Service.

SAWS, 2012. *Aeronautical Summary Port Elizabeth Airport 1996 - 2010.* Pretoria(Gauteng): South African Weather Service.

SAWS, 2013. *Annual Report 2012/2013.* Pretoria(Gauteng): South African Weather Service.

SAWS, 2014. *South African Weather Service - Climate Publications - 2011 - November.* [Online] Available at: http://www.weathersa.co.za/web/index.php?option=com_content&view=article&id=285&Itemid=516&jsmallfib=1&dir=JSROOT/2011/November [Accessed 11 May 2014].

SAWS, 2014. *South African Weather Service - Climate Publications - 2012 - April.* [Online] Available at: http://www.weathersa.co.za/web/index.php?option=com_content&view=article&id=285&Itemid=516&jsmallfib=1&dir=JSROOT/2012/April [Accessed 2014 May 2014].

SAWS, 2014. *South African Weather Service Climate Publications 2012 July.* [Online] Available at: http://www.weathersa.co.za/web/index.php?option=com_content&view=article&id=285&Itemid=516&jsmallfib=1&dir=JSROOT/2012/July [Accessed 14 May 2014].

SAWS, 2014. *Weathersa Climate Publications 2012 December.* [Online] Available at: http://www.weathersa.co.za/web/index.php?option=com_content&view=article&id=285&Itemid=516&jsmallfib=1&dir=JSROOT/2012/December [Accessed 15 May 2014].

SAWS, 2014. *Weathersa Climate Publications 2012 September.* [Online] Available at:

http://www.weathersa.co.za/web/index.php?option=com_content&view=article&id=285&Itemid=516&jsmallfib=1&dir=JSROOT/2012/September [Accessed 2014 May 2014].

Schlatter, T. W. & Baker, D. V., 1991. *Algorithms, Comparisons and Source References by Schlatter and Baker.* [Online] Available at: http://wahiduddin.net/calc/density_algorithms.htm

Skytrax, 2013. *World Airport Awards.* [Online] Available at: http://www.worldairportawards.com/Awards_2013/top100.htm [Accessed 03 09 2013].

SouthAfrica.info, 2012. *South Africa's geography.* [Online] Available at: http://www.southafrica.info/about/geography/geography.htm#.UiYPGDanqy8 [Accessed 03 09 2013].

Tennant, W. J., Toth, Z. & Rae, K. J., 2007. Application of the NCEP Ensemble Prediction System to Medium-Range Forecasting in South Africa: New Products, Benefits, and Challenges. *Weather and Forecasting,* Volume 22, pp. 18-35.

Thorpe, A., 2012. Weather forecasts in 2030. *ECMWF Newsletter No. 131 Spring 2012*, p. 1.

Tyson, P. D. & Preston-Whyte, R. A., 2000. *The Weather and Climate of Southern Africa.* Second ed. Cape Town(Western Cape): Oxford University Press South Africa.

van der Mescht, D. & Eloff, P. J., 2013. Mountain wave-induced rotors in the lee of the Hex River Mountains. *South African Geographical Journal,* pp. 117-1331.

Wilks, D. S., 2011. *Statistical Methods in the Atmospheric Sciences.* 3rd ed. Oxford, United Kingdom: Academic Press.

WMO, 2007. *Guide to Aeronautical Meteorological Services and Cost Recovery WMO-No. 904 Second Edition.* 2nd ed. Geneva: World Meteorological Organization.

WMO, 2007. *Technical Regulations Basic Documents No. 2 Volume II - Meteorological Service for International Air Navigation WMO-No. 49.* 2007 ed. Geneva: World Meteorological Organization.