ABSTRACT

ARGENT, ROWAN ELIZABETH. Customisation of the WRF Model Over the Lake Victoria Basin in East Africa. (Under the direction of Dr. Fredrick H. M. Semazzi.)

Lake Victoria lies in Eastern Africa surrounded by Tanzania, Kenya and Uganda, and is one of the major sources of the Nile. Between the countries surrounding the lake and those further downstream it supports millions of people. The lake receives 80% of its inflow from rainfall directly into the lake, so it is necessary to be able to model this rainfall accurately, in order to produce reliable climate projections. The initial step in this process is customisation of the model for this region. Of particular interest is an asymmetrical pattern that forms across the lake basin due to a diurnal land-lake breeze. It is hypothesised that the Weather, Research and Forecasting (WRF) model is unable to accurately reproduce this pattern due to deficient hydrodynamics. The main aim in this customisation study is to determine whether or not this pattern can be reproduced. The customisation was performed by conducting several series of model runs in order to investigate certain aspects of the customisation. These aspects are; parameter and options within WRF, lake surface temperature (LST) initialisation, physics in the best and worst runs, extreme years, optimal customisation for temperature and bias due to the lake. In each section, a set of model runs was conducted and then analysed using Tropical Rainfall Measuring Mission (TRMM) rainfall and Climatic Research Unit (CRU) temperature data. The model runs were compared to these observational datasets using standard statistical tools, such as the Root Mean Square Error (RMSE). The study showed that the parameters chosen and the way in which the LST is initialised have the potential to greatly improve or worsen the results. Also, it cannot be assumed that the optimal combination for rainfall in a year representing the climatology will be optimal for extreme years and/or other variables within the climate system. Despite these

complexities, one model combination (N) did perform consistently better than the others across multiple experiments. A major contributor to this good performance is the use of the BMJ cumulus scheme, which more accurately represents rainfall over the Indian Ocean than the other schemes. Finally, the study concluded that WRF is unable to reproduce the asymmetrical pattern across the lake, although it is capable of capturing the elements of the pattern that occur over land on the eastern and western shores. It is suspected that this is due to the temperature of the lake being too low and preventing the diurnal reversal of the land lake breeze. In reality, the land-lake breeze reverses at night, causes convergence and consequently rainfall, but the lower model temperature prevents this from happening. Overall, this study highlights the importance of customising a model to the region of research.

Customisation of the WRF Model Over the Lake Victoria Basin in East Africa

by
Rowan Elizabeth Argent

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Marine, Earth and Atmospheric Science

Raleigh, North Carolina

2014

APPROVED BY:

Dr. F. H. M. Semazzi
Committee Chair

Dr. L. Xie

Dr. A. Hanna

BIOGRAPHY

Rowan Elizabeth Argent was born in Sheffield, England where she lived until she moved to Peebles, Scotland at the age of 11. Here she developed her interest in maths, physics and hiking in the hills of the Scottish Borders often accompanied by some form of rain or drizzle. This led to a B.Sc. from the University of East Anglia in meteorology and oceanography. During her degree she spent an additional year at the University of Arizona in Tucson where she got to experience a very different climate and culture. Her experience in Tucson along with her desire to continue in education brought her to North Carolina State University where she plans to continue into the PhD program.

ACKNOWLEDGMENTS

Firstly, I would like to thank my advisor, Dr. Semazzi, for his help and encouragement throughout the duration of this project. Additionally I would like to thank my committee members, Dr. Xie and Dr. Hanna, for their feedback and advice.

I would also like to thank my colleagues within Climlab, Kara Smith, Pascal Waniha and Masilin Gudoshava who were endless sources of knowledge and help. Additionally all my friends and colleagues in the MEAS department were always there to bounce ideas off and go for emergency cups of tea.

Finally I would like to thank my parents and sister for their continual (long distance) support and love and Michael Angus, for his unwavering help, love and cake supplying during times of stress.

This project was funded through the NSF grant number 553210.

TABLE OF CONTENTS

LIST OF TABLES vii
LIST OF FIGURES viii
1. Introduction 1
1.1. Background on the Lake Victoria Basin 1
1.2. Current Climatology 2
1.2.1. Atmospheric Influence 2
1.2.2. Local Circulation 7
1.2.3. Lake Influence 9
1.2.4. Resulting Circulation 10
1.3. Future 11
1.4. Socioeconomic Factors 13
1.4.1. Dependency on the lake 13
1.4.2. Future Changes 16
1.5. Difficulties in Projections 17
1.6. Modeling and Customisation 19
1.6.1. Overview 19
1.6.2. Previous customisation 20
1.7. Project Overview and Hypothesis 22
1.7.1. Project Overview 22
1.7.2. Hypothesis 22
2. Initial Customisation 24
2.1. Introduction 24
2.2. Hypothesis 25
2.3. Methods 25
2.3.1. Preliminary Customisation 25
2.3.2. Full customisation 30
2.3.3. Additional Runs 34
2.3.3.1. Boundary Conditions 35

2.3.3.2. Climate parameters 35
2.3.4. Land use..36
2.3.5. Lake Surface Temperature and Sea Surface Temperature 37
2.4. Results 37
2.4.1. Preliminary customisation 37
2.4.2. Full customisation...40
2.4.2.1. Parameters..40
2.4.3. Additional runs 47
2.5. Conclusion..48
3. Lake Surface Temperature...50
3.1. Introduction 50
3.2. Hypothesis...52
3.3. Methods..52
3.4. Results and Discussion..54
3.4.1. Statistical Results..54
3.4.2. Visual Rainfall Results 55
3.4.3. Lake Surface Temperature...59
3.5. Conclusion...61
4. Optimal Customisation 63
4.1. Introduction 63
4.2. Hypothesis...64
4.3. Methods..64
4.4. Results and Discussion..66
4.4.1. Rainfall 66
4.4.2. Statistical Results..71
4.4.3. Other Variables...74
4.4.3.1. Temperature 74
4.4.3.2. Relative Humidity...76
4.4.3.3. Wind Flow 77

4.5. Conclusion 80
5. Extreme Years 82
5.1. Introduction 82
5.2. Hypothesis 83
5.3. Methods 83
5.4. Results 85
5.5. Conclusion 88
6. Lake Mask 90
6.1. Introduction 90
6.2. Hypothesis 91
6.3. Methods 91
6.4. Results 92
6.5. Conclusion 94
7. Temperature Comparison 96
7.1. Introduction 96
7.2. Hypothesis 96
7.3. Methods 97
7.4. Results and Discussion 97
7.5. Conclusion 101
8. Conclusions 103
9. Future work 107
REFERENCES 109
Appendix 1 117
Appendix 2 118
Appendix 3 119
Appendix 4 120
Appendix 5 121
Appendix 6 122
Appendix 7 123

LIST OF TABLES

Table 2.1: Combinations of the parameters used for initial customisation. 28

Table 2.2: Combinations of schemes for longer runs. 33

Table 2.3: The Runs produced for the BC options. 35

Table 2.4: Runs used for the climate parameters. 36

Table 2.5: The runs conducted for the Land Use section. 37

Table 2.6: The RMSE and Standard Deviation for each of the different schemes chosen in each of the domains. 38

Table 2.7: The order of accuracy according to each statistical test (best to worst in descending order). 41

Table 3.1: The runs produced for this section 52

Table 4.1: The combinations used in all the runs for this chapter 64

Table 4.2: The Standard Deviation over all the different domains and time periods (ND = November, December). Green indicates low error fading to red with high error. 72

Table 6.1: The order of accuracy according to each statistical test. Blue and orange represent the combinations that have changed with the lake masked, down and up respectively. 92

Table 6.2: Difference between the runs for each statistic. 94

Table 7.1. The order of accuracy for each statistic 98

Table A1: Preliminary Customisation Results 117

Table A2: Skill Score for Customisation, figure in green represent skill. 118

Table A3: Statistical results for the LST Chapter, Red shows an increase in error and green shows a decrease. 119

Table A4: RMSE and MAE for best chapter, green represents the lowest error, fading toe red for the worst. 121

Table A5: Skill scores for Best runs. 122

Table A6: Difference in Statistics between control and extreme year. Green represents when the extreme year outperforms the control. 123

LIST OF FIGURES

Figure 1.1: a) Location and Topography of Lake Victoria in meters (Anyah et al., 2006), b) Map of the Lake Victoria Basin (East African Community (EAC), 2004) 1

Figure 1.2: Mean circulation over Africa in a) the boreal winter and b) the boreal summer. Dotted line indicates ITCZ (Nicholson, 1996). 4

Figure 1.3: SST and surface wind anomalies during a dipole event, Anomalies of SSTs and winds exceeding 90% significance are indicated by shading and bold arrows, respectively. (Saji et al., 1999) 5

Figure 1.4: The annual rainfall (mm) contoured over the lake, showing the asymmetrical pattern (Asnani, 1993). 10

Figure 1.5: Probability of Urban Growth (Seto et al., 2012) 17

Figure 2.1: The domains used for the initial run. 26

Figure 2.2: The domains for the longer runs 31

Figure 2.3: Difference between the WRF model rainfall output and the TRMM observations (mm). 43

Figure 2.4: November and December rainfall for all runs and TRMM in mm. 44

Figure 2.5: As previous figure except for the lake domain. 46

Figure 3.1: November Difference (in mm) between TRMM data and WRF model output over the domain 2. 57

Figure 3.2: East-West Cross section of November rainfall in mm. 58

Figure 3.3: The SST and LST for each run and each domain 60

Figure 4.1: Rainfall difference between WRF and TRMM over the total time period (OND). 67

Figure 4.2: Total Rainfall (OND) over domain 1 for all runs plus TRMM. 68

Figure 4.3: As 4.2 but for Domain 2 69

Figure 4.4: As 4.3 but for the Lake domain. 70

Figure 4.5: November rainfall for Bbest, TRMM and Ybest 73

Figure 4.6: Difference between mean OND CRU temperature and WRF temperature (°C)... 75

Figure 4.8: As 4.7 but for U wind. 78

Figure 4.9: Abest on the 22nd of November showing the same wind direction during the day. 80

Figure 5.1: Total TRMM rainfall for OND for each extreme year 85
Figure 5.2: Difference between WRF and TRMM rainfall over OND in domain two. 87
Figure 5.3: WRF rainfall on the right and TRMM rainfall on the left for each extreme year. Total rainfall over OND for domain two in mm. 88
Figure 7.1: The difference between CRU and WRF for the mean temperature over November and December for the upper and lower quartile combinations. 100
Figure 7.2: The difference between CRU and WRF for the mean temperature over November and December for run N. 101
Figure A1: Difference in rainfall between WRF and TRMM data over November in mm. ..120

1. Introduction

1.1. Background on the Lake Victoria Basin

Lake Victoria is a freshwater lake in Eastern Africa. The lake itself is surrounded by three countries: Uganda to the north, Kenya to the east and Tanzania to the south. Additionally Rwanda and Burundi are both located in the Lake Victoria Basin (LVB) and run off area of the lake (Fig 1.1) (Sutcliffe and Parks, 1999). It is the largest freshwater lake in Africa covering an area of 69,000 km² (East African Community (EAC), 2011), but it is also one of the shallowest in Eastern Africa with an average depth of 45m (Song et al., 2004).

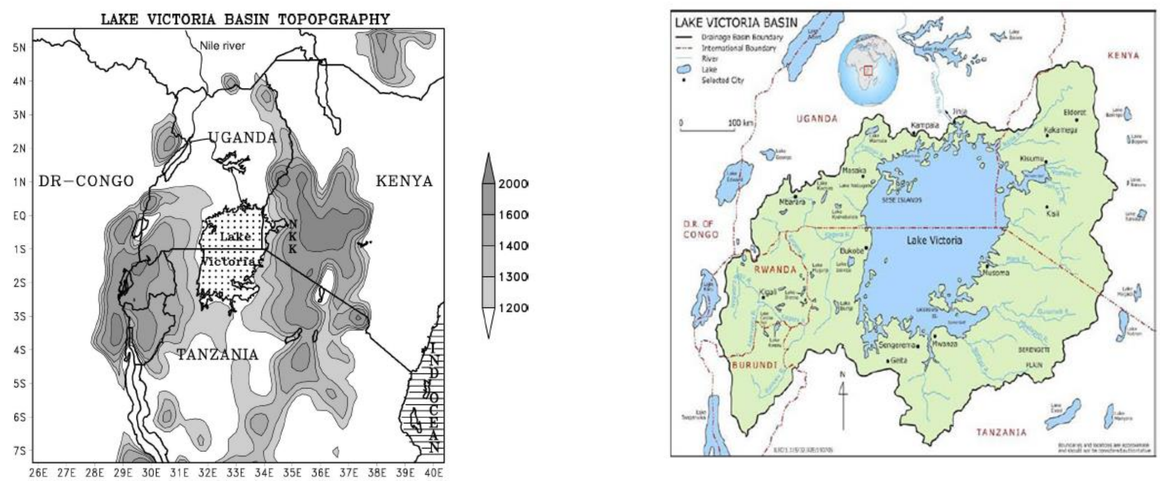


Figure 1.1: a) Location and Topography of Lake Victoria in meters (Anyah et al., 2006), b) Map of the Lake Victoria Basin (East African Community (EAC), 2004)

Approximately 85% of the inflow comes from rainfall directly into the lake itself. The remaining inflow is from tributary inflow, predominantly from the Kagera River (Sutcliffe and Parks, 1999). The lake feeds one of the sources of the River Nile, which then runs through Uganda, South Sudan, Sudan and Egypt. Consequently, the lake is very important for a large area with a population in the millions (World Bank, 2008).

The lake lies in a complex geographical region; it is surrounded by multiple different climate regimes and varying topography (Fig. 1.1a). Situated in the Rift Valley region, the lake has mountainous regions to both the west and the east. These mountains are one of the many influences on the climate in the lake basin. The lake and surrounding land induce a large land-lake breeze circulation system. The strength of this breeze, and consequently the areas that see the most rainfall due to it, are strongly influenced by the topography surrounding the lake (Anyah et al., 2006) and various attributes of the lake, including the lake bathymetry. To the west lies the Congo rainforest, a large moisture source, and to the east, the Kenyan plains and the Indian Ocean. The climate over the lake is influenced substantially by the Indian Ocean with a significant majority of moisture flowing in from the east (Anyah et al., 2006).

1.2. Current Climatology

1.2.1. Atmospheric Influence

The climatology over the lake is determined by both local and large scale influences (Anyah et al., 2006; Asnani, 1993). The dynamics associated with the geography of the Rift Valley Complex, the surrounding climate regimes (section 1.1), and synoptic scale features transient over the region all combine to create a highly complex and interesting climatology over the

LVB.

The most predominant of the large scale features that influence the region is the Inter-Tropical Convergence Zone (ITCZ). The ITCZ follows the overhead position of the sun and migrates across the region twice a year during the annual cycle (Fig. 1.2). As it moves across the region it causes two rainy seasons. These rains occur during the spring (March-April-May) when they are known as the long rains and during the autumn (October-November-December) when they are known as the short rains (Anyah et al., 2006). While the long rains are responsible for the largest amounts of precipitation, it is actually the short rains that determine much of the variability of the annual rainfall (Nicholson, 1996). The movement of the ITCZ is also responsible for the dry periods throughout the year. The wind direction reverses during the rainy seasons with a north easterly flow during the Northern Hemisphere winter and a south easterly flow during the Northern Hemisphere summer (Asnani, 1993). These opposing flows are both thermally stable and result in subsiding air in the region which leads to lower rainfall outwith the ITCZ movement (Nicholson, 1996).

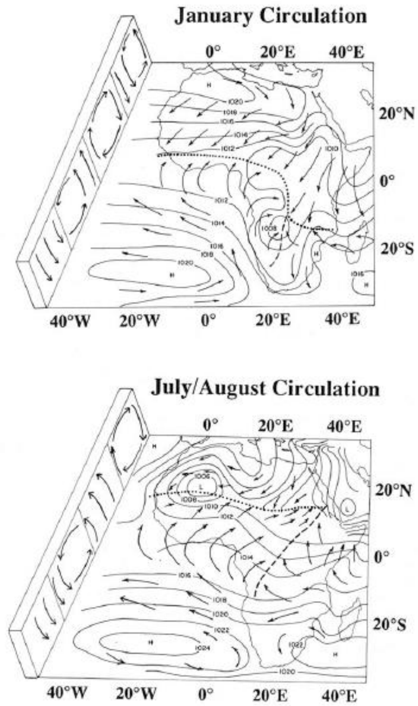


Figure 1.2: Mean circulation over Africa in a) the boreal winter and b) the boreal summer. Dotted line indicates ITCZ (Nicholson, 1996).

As previously mentioned the OND season tends to be responsible for the majority of the variability of annual rainfall and one of the largest sources of variance within the rainfall comes from Sea Surface Temperature (SST) (Nicholson, 1996). In this case both the Indian and the Pacific Oceans play a crucial role. The two oceans are linked physically by the Indonesian throughway, the water transport is controlled by the monsoonal winds, and atmospherically through the Walker Circulation (Black et al., 2003).

There are two large teleconnections that occur in the Pacific and Indian Oceans and appear to influence the region; these are the El Niño – Southern Oscillation (ENSO) and the Indian

Ocean Dipole (IOD), both of which are SST patterns. The IOD is associated with the gradient of the SST in the Indian Ocean (Fig. 1.3). In the Indian Ocean the SSTs are usually warmer in the eastern Indian Ocean and cooler in the west although this temperature gradient reverses periodically. This reversal of temperature anomaly is known at the IOD (Saji et al., 1999). This pattern is most prominent during the boreal autumn months.

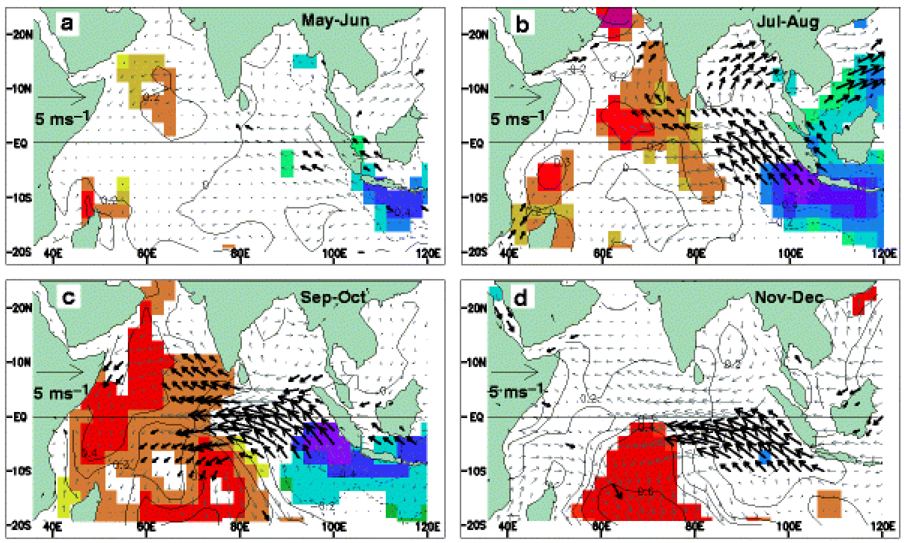


Figure 1.3: SST and surface wind anomalies during a dipole event, Anomalies of SSTs and winds exceeding 90% significance are indicated by shading and bold arrows, respectively. (Saji et al., 1999)

ENSO is a source of interannual variability around the globe. It is an ocean-atmospheric coupled oscillation which is determined by a warming of the SST near South America and the consequent changes to circulation. It modulates rainfall over a large proportion of the African continent (Nicholson and Kim, 1997). However, the connection between the two is not well understood and there is some ambiguity and controversy regarding which mode is dominant in the region.

Both Clark et al., (2003) and Schreck and Semazzi, (2004) found that ENSO is connected to rainfall in the East African region, specifically the OND rains. Additionally the Empirical Orthogonal Functions performed by Schreck and Semazzi,(2004) suggested that ENSO was responsible for 29% of the OND variability, in particular that El Nino years see an increase in rainfall over the Greater Horn of Africa region. However, other studies such as Behera et al., (2005) and Manatsa et al, (2011) found that the IOD is more influential than ENSO. Saji and Yamagata, (2003a) suggested that ENSO only really influences East African rainfall when it coincides with IOD, yet IOD is consistently connected to variable rainfall in the region regardless of ENSO. In agreement with this, the modeling study of Latif et al., (1999) showed that during the enhanced rains of 1997/1998 the Indian Ocean SSTs were more influential than the Pacific. Tierney et al., (2013) also support the idea that ENSO's influence is smaller, particularly over longer timescales. They found that the Pacific's influence waned on timescales longer than 10 years and the low pass patterns were influenced to a greater extent by changes to the Walker Circulation within the Indian Ocean.

There is ambiguity about how these two teleconnections are linked. Clark et al. (2003) and Black et al. (2003) suggest that the IOD is driven by or at least triggered by ENSO. However, Saji and Yamagata (2003b) conclude that IOD is independent of ENSO, although the two teleconnections have the ability to influence one another, either enhancing or reducing the other signal. Both Black et al. (2003) and Manatsa et al. (2011) found that the rainfall response to IOD is non-linear and non-symmetric; for example the response to positive IOD is much greater than the negative mode. Manatsa et al. (2011) additionally found the same to be true of El Nino, but speculated that the role of ENSO has been amplified in past studies due to co-occurrences. Finally, Black et al. (2003) found that the impacts of IOD were only really seen in particularly strong years when the reversal occurred for several months. What is clear is that this is an area of uncertainty and while both teleconnections may have an influence on the region the extent of that influence may well be determined by the other. Additionally, while ENSO and IOD are the most studied measures of variance, there are undoubtedly other large scale features that have the potential to impact the region and are currently less well studied. An example of this can be found in a study by Paeth and Hense (2006) who found that Kelvin and Rossby waves, caused by heating the Tropical Atlantic, had the ability to influence rainfall over large regions of Africa.

1.2.2. Local Circulation

In addition to the large features that influence the region there are local effects to be considered. As previously mentioned the mountains around the lake and the lake itself influence the local climate, in particular creating a land-lake breeze. A land-lake breeze is formed by the same process as a sea breeze. It is caused by the thermal capacity of the lake being much greater than that of the land. During the day the land becomes much warmer, consequently air rises, causing low pressure. A breeze is created as air moves on land towards the low pressure and away from the high pressure created from cooler lake surfaces. At night the opposite occurs as the water becomes warmer than the rapidly cooling land. The land-lake breeze over Lake Victoria is created by the surrounding high topography, the lake itself, the high insolation in the region and the resulting temperatures gradients (Asnani, 1993).

At night the lake is at its warmest in comparison to the surrounding land, resulting in air flowing into the lake and subsequently rising motion. This results in large amounts of convection, and therefore cloudiness and rainfall, overnight (Nicholson and Yin, 2002). Anyah and Semazzi (2004) found that the strength of the breeze and consequently the amount of rainfall is dependent on the lake surface temperature (LST). They found that a warmer lake results in greater rainfall, but also that any changes in LST are not proportional to the corresponding changes in the precipitation. While the high topography surrounding the lake is not necessarily essential to the land-lake breeze, it does have the ability to influence and enhance the system. Anyah et al. (2006) found that the high topography enhanced the breeze both during the day and at night. During the day the high mountains heat faster and create an upslope wind, however, at night this is reversed. They found that at night the cooler air flowing down the slope of the eastern mountains cooled the land further which enhances the land-lake breeze. During the day the upslope wind also takes the land-lake breeze effects further inland.

Additionally, the mountain rain shadows influence the rainfall; the east is in the shadow resulting in lower amounts of precipitation, while in the west the air begins to rise causing greater rainfall amounts (Asnani, 1993). The combination of large scale and local flow means that different areas of the lake see very different diurnal patterns (Asnani, 1993). Both flows are necessary to create the current rainfall pattern (Mukabana and Pielke, 1996). The flow over the lake is easterly throughout the year with a northern/southern component that changes over the year. The sea breeze flows from the east and is weakened and strengthened, depending on the time of day, by the land lake breeze (Asnani, 1993). This easterly flow means that the rainfall caused by the diurnal cycle is not centred on the lake and is actually focused on the north western side of the lake (Nicholson and Yin, 2002).

1.2.3. Lake Influence

As well as the local topography, the lake surface temperature (LST) also influences the atmospheric circulation over the basin. The LST is not consistent over the whole area due to different depths of the lake; the western side of the lake is shallower and consequently has a higher temperature. The heat is then distributed around the lake by the circulation within it (Song et al., 2004). The circulation flows anti-clockwise due to predominantly eastern flow on the surface (Song et al., 2002). This also influences the asymmetrical rainfall pattern across the lake along with the topography and prevailing winds (Song et al., 2004). The LST of the lake can intensify or weaken the flow and the convergence and divergence over the lake. With warmer LSTs there are greater rainfall amounts, however this relationship is not proportional (Anyah and Semazzi, 2004).

1.2.4. Resulting Circulation

The current consensus is that these complex influences interact to create the rainfall pattern seen across the LVB. This pattern is an asymmetry across the lake which is also known as the dry-wet-dry-wet (DWDW) pattern (Fig. 1.4). The amount that each of these features contributes towards this pattern is not fully understood and is the subject of ongoing study (Dr. F. Semazzi – personal communication, 2014).

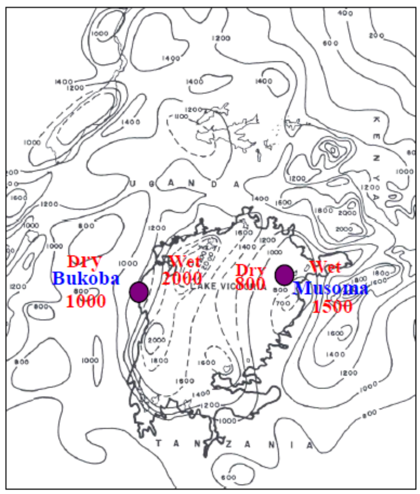


Figure 1.4: The annual rainfall (mm) contoured over the lake, showing the asymmetrical pattern (Asnani, 1993).

It is all these complex influences that make forecasting and future projection for the LVB region difficult and an area in need investigation. These patterns need to be captured accurately by models to build confidence in the accuracy of future projections and forecasts. In particular, it is important to capture the asymmetrical pattern across the lake. A correct representation of this pattern will imply that the correct physics in the regions are being reproduced within the model and constitute a significant element of accurate representation of the lake weather and climate. This will give confidence in both in current studies and in future projections.

1.3. Future

It is not known with confidence how the rainfall in this region is going to change in the future. Many different publications have suggested future projections, but not all of these are in agreement. The most significant of these publications are the Intergovernmental Panel on Climate Change (IPCC) reports. The most recent IPCC report (AR5) confirms with increased confidence that there is warming due to anthropogenic emissions and that the planet will continue to warm (IPCC, 2013a). This warming trend may influence changes in rainfall over the lake basin. Schreck and Semazzi (2004) found that the second most important EOF over the Lake Victoria region was a dipole pattern connected to a climate warming trend so it is apparent that any warming will impact the region. Bowden and Semazzi (2007) did further study into this area, and also concluded that this pattern is potentially linked to a climate warming signal. They also concluded that it could result in the area around Lake Victoria experiencing less rainfall and more severe drought as it falls into the southern dipole region, whilst the northern region is likely to see greater rainfall.

The latest finalised IPCC report is the fourth assessment report (AR4). This supports one of the theories about change in rainfall over Eastern Africa. The general projections are that the sub tropics will see a decrease in rainfall and the tropics will see little change or a slight increase (Christensen et al., 2007). The report suggests that all of Africa will warm and concludes that this change in rainfall is consistent with changes in water vapour content in a warmer atmosphere (Christensen et al., 2007). Additionally, there is a consensus within the models used for the IPCC report that East Africa will experience an increase in rainfall (Christensen et al., 2007). However, this projection theory is not in agreement with Williams and Funk (2011) who suggest that an increase in the gradient of the warm pool in the Indian Ocean could potentially extend the Walker circulation, causing descending air over Eastern Africa. The authors also suggest that moist air would be less likely to be transported inland over the continent; both of these changes would act to decrease rainfall over the region. These changes to the walker circulation were initially suggested by Funk et al. (2008). Due to the IPCC report being based on an ensemble of models it is considered to be the most reliable interpretation, however it is important to be aware of other possible theories.

Finally an additional factor is how the influence of the teleconnections mentioned above will change over time. Tierney et al. (2013) suggest that the Indian Ocean has a greater influence over longer timescales than the Pacific, meaning that when looking at patterns over decades, it will be important to consider the Indian Ocean in greater detail.

The future of the rainfall over the Lake Victoria Basin is unclear from current knowledge. It is apparent that more research will need to be conducted in order to create more conclusive projections. In order to create these better projections it is necessary to resolve some of the complex issues mentioned in the section 1.2.

1.4. Socioeconomic Factors

1.4.1. Dependency on the lake

The population of the region surrounding Lake Victoria has a huge dependency on the lake. The population of the basin is around 35 million (World Bank, 2008) and the lake has the potential to provide food, water, income and electricity to the people living in this region. The basin supports many economic activities and is a source of agriculture, livestock, forestry, hydropower, transport and tourism (Ntiba et al., 2001). The lake is the industrial and domestic water supply for around 5 million people living in the major cities directly surrounding the lake (World Bank, 2008).

It is also vitally important for transportation as it is still one of the cheapest methods of communication between the three countries and the many cities surrounding the lake (World Bank, 2008). This water resource experiences a massive amount of use and is one of the most dangerous waterways in the world; this is due to the complexity of the climate and the severity of the weather over the lake basin (East African Community (EAC), 2011). Many accidents, and consequently deaths, occur on the lake both within the fishing and transport communities, often leaving behind multiple dependents (East African Community (EAC), 2011).

Fishing and agriculture are highly dependent on the Lake. Lake Victoria supports the largest inland fishing industry in the world, producing 500,000 tons per year with a value of around $600 million (World Bank, 2008). This fishery supports around 800,000 people in the industry, plus all of their dependents (World Bank, 2008). The basin agriculture consists of crops such as maize, cotton, tobacco, beans, sugarcane and coffee, as well as livestock and forestry (Ntiba et al., 2001). Irrigation from the lake is used to supplement rainfall for some perennial crops (Sutcliffe and Parks, 1999) and in 1999 Sutcliffe and Parks (1999) stated that the irrigation could use as much as 5% of the mean outflow.

The basin is very important for the production of hydroelectric power, particularly in Uganda at the source of the Victoria Nile (World Bank, 2008). The singular outflow of the Lake goes through Nalubaale Hydroelectric Power Station and the water is currently released according to an agreed curve, decided on by Uganda and Egypt after the dam was completed in 1954 (Sutcliffe and Petersen, 2007; Sutcliffe and Parks, 1999). The artificial nature of the release means that lake levels will fall or rise unnaturally depending on the water release (Sutcliffe and Petersen, 2007). The demand for hydroelectric power has the potential to act as an incentive not to follow the original release rules and may cause lake levels to fluctuate (Sutcliffe and Parks, 1999).

238 million people live in the River Nile basin, from all the tributary sources to the Mediterranean Sea. This is a very large population dependent on one water resource (Nile Basin Initiative, 2012). Within Uganda, South Sudan and Egypt over 90% of their populations live within in the Nile River Basin (Nile Basin Initiative, 2012). The countries downstream of Lake Victoria are also dependent on the water released into the Nile from Lake Victoria, as it supplies vital water to homes and agriculture as well as hydroelectric power through dams (Sutcliffe and Parks, 1999). Farming is one of the main sources of income for communities living along the Nile river basins (Rahman, 2013). Many communities downstream from Lake Victoria are dependent on the ebb and flow of the seasonal flood plain. Upstream developments can affect water use downstream as they may determine when and how much water is reaching people and communities (Sutcliffe and Parks, 1999).

Water resources are of vital importance all over the world. Many of the countries with access to Lake Victoria and further downstream along the River Nile rely on these resources very heavily. It is important that the water is used fairly and is accessible to all. (Rahman, 2013) state that water resources have the potential to be a cause of tension and consequently conflict. Previously Egypt has had the majority of the control over the water use from the Nile, however many of the other countries are beginning to question the treaties and request a greater supply of water, which has the potential to escalate into conflict (Rahman, 2013). Should water supplies lessen in the future due to less rainfall the potential of escalation into conflict may increase.

1.4.2. Future Changes

For all of the reasons given in section 1.4.1, it is apparent that any changes in rainfall, and consequently lake and river levels, could have serious impacts on both the countries surrounding the lake and those further downstream (World Bank, 2008). Seto et al. (2012) shows that the northern side of Lake Victoria is likely to see huge urban growth to the extent that most of the northern coastline is urban in the future (Fig 1.5). The Nile Basin in Egypt is also expected to see large urban growth (Seto et al., 2012). In particular, the economic effects will be felt in the fishing, shipping and hydroelectric industries (World Bank, 2008). Assessments of water resources are currently being conducted within the LVB (Mehta et al., 2013). These are only of any use if we have confidence in the future climate and how that will impact the hydrology of the region. Finally, an improvement in meteorological understanding resulting in better forecasts and information may help to reduce the number of deaths occurring on the lake (East African Community (EAC), 2011). Consequently, changes in long term or short term patterns have the potential to cause disastrous effects and future modeling is important to determine if and how the region will change on multiple timescales.

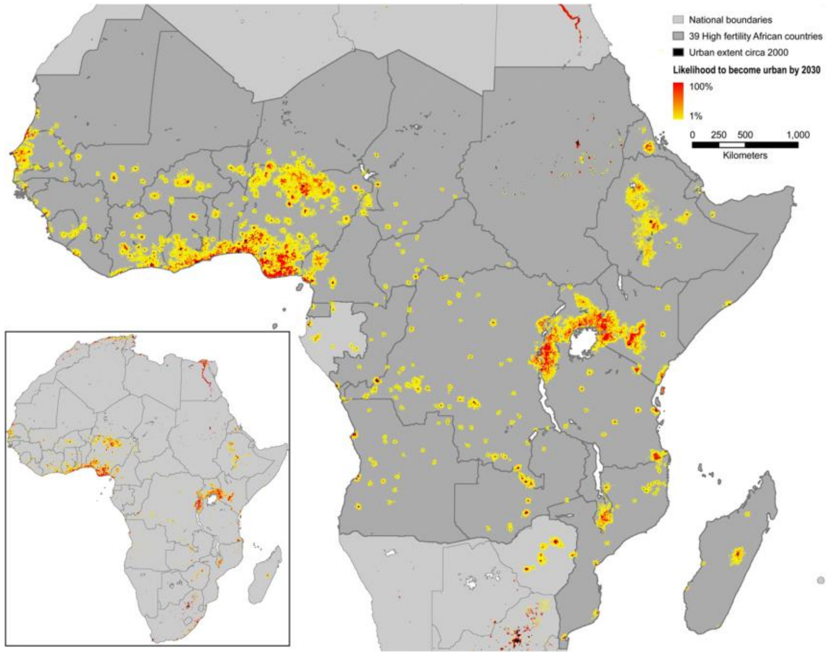


Figure 1.5: Probability of Urban Growth (Seto et al., 2012)

This review reinforces the reasons why it is so important to accurately model and project the rainfall and its asymmetry across the Lake Basin. A resolution of the uncertainties, better understanding and greater confidence in what is likely to happen would lead to a greater ability to adapt and mitigate any future changes, therefore lessening the socioeconomic impacts.

1.5. Difficulties in Projections

This region is a hard area to model accurately. The surrounding topography and geography are complex. The interactions between different scale flows over the lake make the region complicated climatically. These factors combined create a difficult region to model.

Additionally complicating matters are the lack of observations. Observations are vital for validating. This region lacks station data (Nicholson and Yin, 2002; East African Community (EAC), 2011) which limits model comparison both spatially and temporally.

The other key limitation on the climate projections in this region is the lake itself. The lake also lacks both climate and hydrology observations. The circulation and variables such as lake surface temperatures are not fully known or understood, nor is the interaction between the lake and the atmosphere fully understood. Song et al. (2004) found that a model with a coupled lake component gave better results than those that did not include the lake, so it is known that it is important to include the lakes interaction with the atmosphere and that this is an important component of the climate system.

The IPCC does highlight that the models used to create these projections contain some systematic errors, including, among others, the displacement of the ITCZ and an overestimation of rainfall across large areas of the continent. This means that the projections exhibit high levels of uncertainty (Christensen et al., 2007). The number of models covering the African continent is limited and many of those that do exist are focused on Southern Africa, again limiting the confidence in the projections (Christensen et al., 2007). Eighteen out of the twenty-one models used in the IPCC report agreed that Eastern Africa would see an increase in rainfall meaning that this projection, at least, is relatively robust (Christensen et al., 2007). It is this use of ensembles, used by the IPCC and other projects such as the Co-ordinated Regional Downscaling Experiment (CORDEX), that gives the future projects greater confidence than those run by individual models.

This section has discussed the difficulties faced when trying to accurately model and project for this region. It is these problems and issues that need to be improved and overcome to accurately produce results for this region. A resolution of these issues and further modeling studies in the region will give greater confidence when producing the results for the region.

1.6. Modeling and Customisation

1.6.1. Overview

Global Climate Models (GCMs) are used all around the globe for future projections. One of the most important ensembles of GCMs is that of the Coupled Model Intercomparison Project (CMIP). This is the collection of modeling groups that are used for the analysis within the IPCC. It is the projections from these models that the IPCC relies on to provide the long term future projections around the globe. When considering the IPCC's rainfall projections for Eastern Africa, the results are mixed. While overall there appears to be a trend in the region getting wetter the differences between the 25^{th} and 75^{th} percentile are relatively large, including a difference in sign in the near future projection. In addition to this, many of the results are less the one standard deviation of the difference, meaning that the signal cannot be definitively attributed to a warming climate (IPCC, 2013b). Furthermore Eden et al. (2012) suggest that while GCMs can show skill in precipitation around the globe, the continent of Africa is still one region with significant error. Consequently these models are not necessarily the optimum way in which to look to the future, or even current climate, when focused on a particular region. In order to capture finer and more accurate details of the precipitation pattern it is necessary to downscale through a Regional Climate Model (RCM).

The process of model customisation involves adjustment of the tuneable parameters and attributes within the Regional Climate Model. This is in order to optimise the model performance for a particular region. Unlike most inland regions where RCMs have been used, Lake Victoria creates an additional source of complexity and therefore uncertainty.

The Weather, Research and Forecasting (WRF) model has a global community of users, registered in 130 different countries (UCAR, 2013), and has been evaluated for many different regions around the world. A few recent examples of the many studies using WRF around the globe include Osuri et al., (2012) for the northern Indian Ocean, Jimenez et al., (2013) for the Iberian Peninsula, Spain, Zhao et al., (2012) for the Great Lakes in the USA and Rao et al., (2012) for Guyana. The LVB, however, has not been studied in great detail.

1.6.2. Previous customisation

Previous customisation in this region is very limited. The Climate Modeling Laboratory (Climlab) at North Carolina State University, in particular Sun et al. (1999) and Davis et al. (2009) performed substantial research in the customisations for the LVB. However, this was using the RegCM3 model which is not the model of choice for this project.

Much more recently in 2011, Pohl et al. (2011) conducted research into the customisation of the WRF over Eastern Africa. This is the model that is used in this project and consequently this previous research is of use. The main aim of their study was to determine which parameterizations were the most influential and had the largest impact on the current hydrological cycle. Their focus was a large East African domain rather than specifically the lake basin. They concluded that it was important to conduct sensitivity tests on regions due to many parameterisations having significant impacts on the hydrological cycle, however they concluded that there were still additional parameters to test.

As the main area of focus for this project is the land-lake breeze circulation, it is important to take into account the different scales involved in the process. In Crosman and Horel (2010) they analyse the modeling of a land-sea breeze and state that this cannot be modeled with a grid spacing larger than around 1km. This focused on the circulation directly surrounding the coast and therefore it can be considered that the circulation of this study is larger than the normal constituents of a sea-breeze, potentially due to the size and depth of the lake as well as the surrounding mountainous regions. However, this is an important factor to consider. Song et al. (2002) managed to reproduce the large circulation which occurs across the entirety of the basin with a grid spacing of 20km and consequently it is shown that the reversal can be captured at a coarser grid spacing.

1.7. Project Overview and Hypothesis

1.7.1. Project Overview

While the overall aim is to achieve projections with greater confidence, the first step in this process is to customise the model for the region. This is the main focus of study for this project. The following chapters will include an in depth study of how the WRF model should be customised for this region. There is an emphasis on the models ability to reproduce the asymmetrical pattern that is considered to be an important attribute of the regional climate and therefore should be modelled accurately. With an accurate customisation the confidence in the model's ability should increase, which in turn will result in more accurate projects and a better idea of any future changes to rainfall.

1.7.2. Hypothesis

The standard version of the WRF model has a highly deficient formulation of lake hydrodynamics. It is hypothesised that the region directly over Lake Victoria cannot be improved through the traditional model customisation procedure, adopted in many previous RCM investigations. However it is also hypothesised that customisation will have significant positive impacts on the surrounding land region. Consequently, a central objective in this study is to develop a standard procedure that may be adopted in future studies for the customization of RCMs over LVB. Within this hypothesis, and in the following six chapters, six sub hypotheses will be applied and assessed:

- The default settings for WRF are not adequate to represent the climate over the Lake

Victoria Basin. By customising the settings for the region the statistical results and rainfall distribution accuracy will be vastly improved and the asymmetric pattern will be more apparent in the results.

- That the SST data and the way in which the lake surface temperature is initialised will influence the accuracy of the rainfall distribution.
- The physics in the best model runs will show distinct differences to the physics in the worst runs and that the best runs will capture more of the diurnal cycle and asymmetrical pattern over the lake.
- That the best combination will also provide accurate results for years with other influences such as ENSO.
- That the inaccurately represented lake creates a bias in the results for the entire domain. This bias impacts which set of parameters are optimal for predicting rainfall in the region.
- The best customisation based on rainfall will also provide the best customisation for temperature.

2. Initial Customisation

2.1. Introduction

Customisation, as mentioned in section 1.6, is a critical component of modeling. It is necessary to have confidence in the models and to do this the model needs to be suitable for the region of the study. Therefore customisation of the model should be conducted to make sure the results obtained are optimal.

Models such as WRF, the model of choice for this study, have many different options and parameters that may be changed and adapted for particular regions, situations, and length of run (Wang et al., 2011). Among the options are many different parameters for aspects of the model such as; convection (cumulus) schemes, microphysics, the planetary boundary layer and the radiation. Each parameter for every option contains a different way of realising the problem. One example of this is the microphysics schemes, which contain different numbers of processes; one may contain only a few variables and not include ice-processes, whereas another may have up to ten variables, including ones such as hail and graupel (Skamarock et al., 2008). This is just one example; there are many different variations within the parameters and options in WRF. In addition to the parameter options, other specifics need to be considered, such as the Boundary conditions or the land use dataset.

This chapter goes through the Initial process used to customise the model for the Lake Victoria Basin. Section 2.3 provides the method and rationale behind the options chosen while section 2.4 will provide the results and discussion. Section 2.5 concludes the chapter.

2.2. Hypothesis

The default settings for WRF are not adequate to represent the climate over the Lake Victoria Basin, by customising the settings for the region the statistical results and rainfall distribution accuracy will be vastly improved.

2.3. Methods

The methods section in this chapter is for the main customisation, however it will also cover many details of the methods applied in the following chapters and many of the methods and analysis will also be used in the assessment of the other hypotheses. Consequently this section can be used to provide a general overview, and will frequently be referred to in the subsequent chapters.

This project is primarily based upon the customisation of a model for the Lake Victoria Basin (LVB) in Eastern Africa. The model chosen to be customised is the Weather, Research and Forecasting Model (WRF), version 3.3 (Wang et al., 2011). This model is used for the entirety of this project; initially to produce small runs for a preliminary customisation, followed by longer runs, which are conducted for both a final customisation and additional sections.

2.3.1. Preliminary Customisation

The focus of this project is the short rains; these are the rains that fall in October, November and December. These rains are also responsible for the majority of the rainfall variance in the region (Nicholson, 1996). For the preliminary customisation the runs are five days long and are started at 00z on the 26th of November 1999 and finished at 00z on the 1st of December. This period is chosen both because it represents a time slots in the middle of the short rains and because Pohl et al. (2011) stated that this year was representative of the climatology.

Two domains are chosen over the region, one nested within the other. Both encompass the LVB. The domains chosen are shown in figure 2.1. The parent domain has a grid spacing of 12km and covers a large region over Eastern Africa. The nest however is much more focused on the lake basin itself and has a grid spacing of 4km.

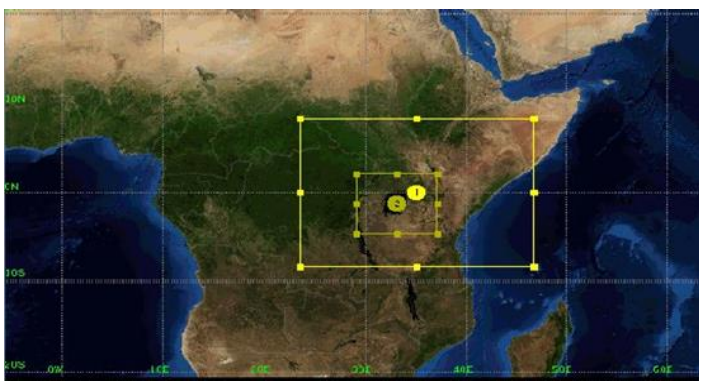


Figure 2.1: The domains used for the initial run.

These runs are initiated with Reanalysis data, in this case the NCEP FNL (Final) Operational Global Analysis data (National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, 2000). This data has a grid spacing of 1°x1°

and is prepared every six hours. Reanalysis means that it is a combination of both observations and model output.

In order to save time, the customisation was divided up between this project and another study that focused on the same region. It was decided that each project should focus on 3 different parameters. This project considered both radiations schemes (Long wave (LW) and short wave (SW)) and the land surface parameters (Lsurf). The other studied the cumulus, microphysics and planetary boundary layer schemes. This study used WRF 3.3 while the other used WRF 3.2. This summary concentrates mainly on the three parameter options it was directly involved in, however the results from the other project will be briefly touched upon.

Each of the three parameterisations has multiple options which are varied over several runs. For each parameter, the different options are changed one at a time from the default run and then compared. The differences between the runs are used to ascertain which parameters are likely to give the most accurate results in the region. The combinations of the three parameters can be seen in table 2.1.

Table 2.1: Combinations of the parameters used for initial customisation.

Run	LW	SW	Lsurf
a	RRTM	Dudhia	Noah
b	GFDL	Dudhia	Noah
c	CAM	Dudhia	Noah
d	RRTMG LW	Dudhia	Noah
e	New Goddard	Dudhia	Noah
f	RRTM	Goddard	Noah
g	RRTM	GFDL	Noah
h	RRTM	CAM	Noah
j	RRTM	RRTMG SW	Noah
i	RRTM	New Goddard	Noah
k	RRTM	Dudhia	5-layer
l	RRTM	Dudhia	RUC
m	RRTM	Dudhia	Pleim-Xiu

In order to determine the accuracy the of the WRF output observations are required. Data from the Tropical Rainfall Measuring Mission (TRMM) (Goddard Space Flight Center, 2013) is used in this project to determine how well the model is performing. TRMM was chosen due to a number of reasons. Firstly, this region has very limited in situ and station data (Nicholson and Yin, 2002) which is particularly problematic when rainfall can vary over very small scales spatially. There are stations surrounding the lake but no data from over the lake itself. Therefore it is decided that the TRMM rainfall 3B42 dataset (NASA, 2013) would be most suitable as the dataset is a combination of different precipitation estimates including both satellite coverage and station data. This means that the full area of interest is covered consistently, including the lake, which may just be approximated in datasets reliant on station data. This data set has a grid spacing of 0.25°x0.25° over the whole domain. Additionally, this data is at a three hourly temporal resolution which means it is one of the few data sets that is at a small enough time scale to enable diurnal results to be analysed.

While TRMM data provides the best and most comprehensive data available, it should be considered that it is not error free, and that the lack of surface observations also prevents the satellite data from being validated. In 2000, Adler et al. described the merging of satellite and rain gauge data for 3B42 dataset. They stated that TRMM had a 10% positive rainfall bias when compared to GPCP data and in some regions this extended to 20%, particularly in areas of high rainfall. Adeyewa and Nakamura (2003) also validated the TRMM data with rainfall gauge data over various regions in Africa and stated that the 3B43 (monthly version of 3B42) data had the closest agreement with rain gauge data. However neither of these studies focused on the East African or Lake Victoria Basin.

The TRMM data also has a random error estimate within the dataset. This is a root mean squared error set based upon the calculations by Huffman (1997). When plotted for November 1999 it shows significant errors over the lake and the Congo Rainforest, this is not unexpected as Huffman (1997) stated that the error increases with greater rainfall and these two regions would experience some of the larger precipitation amounts within the domain. Additionally areas with a greater amount of rain gauge data have lower errors, again lack of observations is known to be a problem in this region. While these errors exist and need to be considered the TRMM data set still provides the most comprehensive data set available for use in this project.

Statistics are used to evaluate and analyse how well the model runs perform with respect to the TRMM observations. For the preliminary runs the Root Mean Square Error (RMSE) and the standard deviation (SD) of the difference between the WRF output and the observations are calculated. These are considered and calculated in the same way as Carvalho et al. (2012) and Pielke (2002). For each run, the RMSE and SD are calculated for the total rainfall over the region in the time period. Additionally two individual days, days one and five, were analysed separately. This gives an indication of how well the model is performing with respect to the observations. Pohl et al. (2011) only considered the RMSE when they conducted their sensitivity study. Carvalho et al. (2012) used both RMSE and SD and suggested that it is important to look at both as the SD indicates the constant error and if it is high then, even with a low RMSE, it means that the error is random and hard to interpret.

Difference plots, where TRMM data has been subtracted from the model output, were created for both domains to determine if particular regions were being over or underestimated by the model and if the model was able to capture the asymmetrical pattern that the hypothesis focusses on. The majority of the results of the preliminary customisation are based on the RMSE and SD.

2.3.2. Full customisation

To avoid any effects of spin up, the time scale should be increased for the full customisation. In this case, seasonal runs of October, November and December were chosen in order to continue looking at the short rains. The runs are now initiated on 00z 1st October and run until 00z 1st of January, still for the year 1999. This larger time scale avoided a problem with spin up as well as giving a better indication of how well WRF is able to capture the seasonal rainfall.

At the same time as the temporal change a spatial domain change is also implemented to capture a larger area. The new domains are to be large enough to catch large scale features such as Kelvin waves along with a significant amount of the Indian Ocean, as the importance of the SST is well known (Nicholson, 1996). The large domain chosen is similar to that used by Paeth and Hense (2006), in which they considered kelvin waves, but also took into account domains used previously within the climate modelling laboratory here at NCSU, such as those used by Bowden (2008). The nested domain covers a much smaller area centred over the Lake. Both can be seen in Figure 2.2. The grid spacing of the large domain is increased to 50 km and the smaller domain to 10km. Additionally, these runs used the alternative lake temperature initialisation, which will be discussed further in chapter 3.

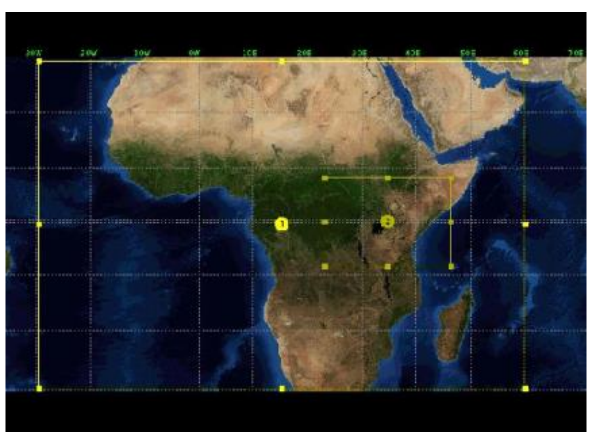


Figure 2.2: The domains for the longer runs

The preliminary customisation runs are short enough so that they do not need a separate SST data set, but for the longer runs it is necessary to have updating SST. The larger runs also use the NCEP FNL for the initialisation and require the additional SST data. In this case the Optimum Interpolation SST data set is used (Reynolds et al., 2002). The data set has a temporal resolution of one week and a spatial grid spacing of 1°x1°. The data set can be found through National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, and National Climatic Data Center/NESDIS/NOAA/U.S. Department of Commerce. (1986, updated weekly.). The addition of this dataset allows the model to capture the changes of the SST over the season more accurately, and consequently any impacts that these changes may have on the accuracy of the rainfall.

A selection of combinations are chosen to be used for the longer runs. There are four base combinations which are: the default settings in WRF (run A), the combination used by Pohl et al. (2011) (run D), the combination used in the East African Community Feasibility Study (run B) (East African Community (EAC), 2011), and the combination found in the preliminary customisation (run E). Beyond this, multiple variations of the base combinations are also used. The different schemes can be seen in the table 2.2 below.

Table 2.2: Combinations of schemes for longer runs.

Run	SW	LW	Cu	Mphys	PBL	Surface
A	Dudhia	RRTM	Kf	Single 3	Yonsei	Noah
B	CAM	CAM	kf	Single 5	Yonsei	Noah
C	CAM	CAM	BMJ	Single 5	Yonsei	Noah
D	Dudhia	RRTMG	Kf	Morrison double	ACM2	Noah
E	Dudhia	RRTM	Grell 3D	Eta	ACM2	Noah
F	CAM	CAM	Grell 3D	Eta	ACM2	Noah
G	CAM	CAM	BMJ	Eta	ACM2	Noah
H	Dudhia	RRTM	Grell 3D	Eta	ACM2	5-layer
I	Dudhia	CAM	Grell 3D	Eta	ACM2	Noah
J	CAM	CAM	kf	single 6	Yonsei	Noah
L	Dudhia	RRTM	Grell 3D	Morrison double	ACM2	Noah
M	Dudhia	RRTMG	Grell 3D	Eta	ACM2	Noah
N	Dudhia	RRTM	BMJ	Eta	ACM2	Noah
O	Dudhia	RRTM	Grell 3D	Eta	Yonsei	Noah
P	Dudhia	RRTM	Grell 3D	Eta	ACM2	RUC

For these longer runs, additional statistical tools are warranted. The Mean Absolute Error (MAE) is added to the statistical analysis for this purpose. Willmott and Matsuura (2005) found that the MAE can give a better indication of the model skill as, unlike RMSE, it is not as significantly biased by the large errors. RMSE, SD and the graphical plots produced were all still used as previously described in section 2.3.1. For the longer runs, the times analysed are the months of November, December and the total over both months. October was avoided in order to account for spin up problems. In addition to this, each domain was interpolated to the grid spacing of the other allowing a direct comparison between domains.

The runs are analysed based on how well they performed statistically; the best quartile and the worst quartile were selected (thirteen runs were analysed so in this study the best and worst three were selected). This is done both for all results as well as being broken down into domain size and grid spacing. In order to analyse the lake in detail, a subset of domain two is also considered, referred to as the lake domain. This is therefore a subset of the results produced in domain two, however it gives a better indication of the error directly surrounding the lake.

Along with the addition of MAE, the skill of the model runs are evaluated using the specifications described by Pielke (2002). It is stated that skill is shown when the RMSE is less than the SD of the observations (referred to as skill score 1) or when the SD of the model approximately equates to the SD of the observations (skill score 2). In order to consider these specifics the SD of the observations and the WRF output are also calculated. Skill score 2 requires equating two different SDs, where within plus/minus 10% of the observational SD is considered to be equal.

2.3.3. Additional Runs

In addition to the different parameters available for customisation, WRF also has many options that can be changed depending on if they are suitable for the domain in question. In this domain, several of these options are of interest, in particular because the aim of this customisation is for a seasonal run. All of the analysis methods used in the evaluation of each of the options below is the same as mentioned in the previous section. None of these runs used the alternative lake initialisation.

2.3.3.1. Boundary Conditions

One of the most important options is the lateral boundary conditions (BCs). The BCs control how the global model data is interpolated down to the WRF model grid spacing. The BCs in WRF are based on the grid spacing; the first grid space is a specific boundary zone with the following specified region for boundary relaxation. The default setting is five spaces and a linear boundary. For climate runs an exponential boundary is expected to be better (Giorgi et al., 1993; Marbaix et al., 2003). A series of runs are conducted in order to determine which BC is better (table 2.3). The runs consist of two different combinations, based on those previously mentioned in section 2.3.2, two different types of boundary, both linear and exponential, and finally a varied number of relaxation boundary points. According to the WRF guide (Wang et al., 2011) the appropriate setting for regional climate runs are exponential with nine spaces for relaxation.

Table 2.3: The Runs produced for the BC options.

Run name	Run Combination	Type	Relaxation Boundary
AA	F	linear	4
AB	F	exponential	9
AC	I	linear	4
AD	I	exponential	9
AE	I	exponential	4
AF	I	exponential	14

2.3.3.2. Climate parameters

WRF contains additional parameters that are advised for regional climate runs, these can be turned on if required. The four recommended options consist of two bucket parameters, for both radiation and rainfall, as well as an option that allows for soil temperature updates and an option that calculates the SST skin temperature. The bucket parameters measure the rainfall in a different way which prevents numerical error in the extremely high numbers that are possible in longer runs (Wang et al., 2011). A series of runs were conducted with and without these additional options. Three different combinations were selected and then run with and without the climate parameters to give an indication as to whether they were beneficial. The combinations used can be seen in table 2.4. The ADD specific in the Run combination is due to the fact that this run was not specified in 2.2.2 due to using the modis land use category that will be covered shortly.

Table 2.4: Runs used for the climate parameters.

Run	Run Combination	Climate Parameters
BA	F	N
BB	F	Y
BC	E	N
BD	E	Y
BE	F+Modis	N
BF	F+Modis	Y

2.3.4. Land use

WRF contains two different land use options for use within the model. These are the Modis option and the USGS options. They contain different land use specifications and a different number of land use categories (Wang et al., 2011). Once again a series of runs are conducted to test whether one land use holds significant advantages over the other. Two different combinations are used and run with both of the land use options. These were then compared to see if either of the land use options performed better than the other (Table 2.5). ADD again shows that the combination wasn't previously listed due this time to containing some of the additional climate options previously mentioned.

Table 2.5: The runs conducted for the Land Use section.

Run	Run Combination	Land Use
CA	F+climate parameters	USGS
CB	F+climate parameters	Modis
CC	I	USGS
CD	I	Modis

2.3.5. Lake Surface Temperature and Sea Surface Temperature

The Lake surface temperature and how it is interpolated by the model is important for determining the rainfall over the lake. Additionally the Sea surface temperature is very important as one of the interpolation methods relies on the SST. This problem is more significant than those previously discussed and will consequently be the focus of chapter 3.

2.4. Results

2.4.1. Preliminary customisation

The results of the preliminary customisation show which schemes appear to give the most accurate results. In the case of the three parameters considered in this study, both long and shortwave radiation and Surface layer, the results are mostly very comparable making it hard to determine exactly which schemes are the best. Table 2.6 shows the RMSE and SD for the total rainfall (full results can be found in Appendix 1). In most cases the results in table 2.6 represent the overall conclusions from the results; however that is not the case for the land surface options.

Table 2.6: The RMSE and Standard Deviation for each of the different schemes chosen in each of the domains.

Domain One

LW	RMSE	Std. Dev.
RRTM	42.30	42.11
GFDL	54.90	53.00
CAM	43.80	43.80
RRTMG LW	44.96	44.49
New Goddard	42.34	42.30

SW	RMSE	Std. Dev.
Dudhia	42.30	42.11
Goddard	55.23	51.20
GFDL	41.84	41.50
CAM	54.30	50.58
RRTMG SW	51.46	48.55
New Goddard	48.36	45.59

Land Surface	RMSE	Std. Dev.
Noah	42.2958	42.1123
5-layer	43.1509	43.1218
RUC	41.7797	41.4436
Pleim-Xiu (usgs no lake)	43.8431	43.6697

Domain Two

LW	RMSE	Std. Dev.
RRTM	51.72	48.74
GFDL	54.49	52.36
CAM	51.29	44.11
RRTMG LW	51.90	49.04
New Goddard	53.10	46.35

SW	RMSE	Std. Dev.
Dudhia	51.72	48.74
Goddard	55.46	55.46
GFDL	48.48	46.58
CAM	54.96	54.94
RRTMG SW	52.71	52.52
New Goddard	50.99	50.90

Land Surface	RMSE	Std. Dev.
Noah	51.72	48.74
5-layer	53.02	48.17
RUC	51.23	49.40
Pleim-Xiu (usgs no lake)	50.11	47.54

For the Long wave radiation, two parameters appeared consistently better than the others. The Rapid Radiative Transfer Model (RRTM - default) and the CAM parameters both appeared to give better results than the others, when considering both domains and the total, day 1 and day 5 rainfall. RRTM showed the best results when considering the RMSE as it always contained one of the best two results in each time frame and domain. CAM showed the second best as it contained one of the two best in four out of six options. These two options were also the best when compared using the SD, although they appeared more equal in the SD. In both RMSE and SD, RRTM gave better results for domain 1 and CAM gave better results for domain two. The remaining options did not perform as well; in particular the GFDL scheme was consistently worse.

For the short wave radiation, the GFDL scheme consistently outperformed the remaining schemes in both domain sizes. Interestingly, this was the scheme that performed the poorest in the long wave parameters. GFDL was consistently better than the others and Dudhia, the default setting, consistently performed second best. These results were found in both the RMSE and the SD. While the other parameter options did not perform as well, many of the results were similar. The Goddard scheme gave the worst results in both the RMSE and SD and the different domains.

Finally, the land use parameter did not produce conclusive results. For the RMSE the RUC parameter produced the best results for both domains, with the Noah option giving the second best results. For the SD this was not true, in this case both the 5-layer option and the RUC option gave reasonable results however it was hard to determine which of the two performed better. 5-layer gave better results for domain 2 whereas RUC appeared better for domain 1. The Pleim-Xiu option consistently produced the worst results.

The parallel study concluded that particular schemes for the Cumulus, Microphysics and Planetary Boundary Layer were optimal. They were Grell 3D, ETA and ACM2 respectively.

These results and those produced by the other study act as a basis for the more in-depth customisation that follows these preliminary runs.

2.4.2. Full customisation

2.4.2.1. Parameters

Table 2.7 summarises the overall performance of each run. The RMSE and the SD had very consistent results with I, N and C being frequently in the upper quartile and J, H and B in the lower quartile. Although runs F and H also feature in the best results for a particular domain or grid spacing, they were not consistently in the upper quartile. The MAE gave differing results, in this case instead of run I appearing in the upper quartile it was replaced by G which was then consistently in the top results. In the MAE ranking, run I was much closer to the middle spectrum of the runs. The worst results for MAE were the same as for RMSE and SD and J was constantly the worst run. The default setting of WRF, run A, often appeared around the centre of the analysis; customisation does therefore lead to improved results. Using these statistical tools, C and N appear best as they give the best results and lowest errors in all of the considered metrics.

Table 2.7: The order of accuracy according to each statistical test (best to worst in descending order).

Root Mean Square Error					
Total	d01	d02	dLake	50km	10km
I	F	N	I	I	N
N	I	C	N	C	C
C	C	I	C	N	I
F	A	G	A	A	F
A	G	F	E	F	A
G	N	A	H	G	E
E	E	E	F	E	G
M	M	M	M	M	M
O	P	O	O	O	O
P	B	P	G	P	P
H	O	H	J	B	H
B	H	B	P	H	B
J	J	J	B	J	J

Standard Deviation					
Total	d01	d02	dLake	50km	10km
I	F	I	I	I	I
N	I	N	N	C	N
C	C	C	H	N	C
F	A	F	C	F	F
A	G	A	F	A	A
M	N	M	A	G	E
E	M	E	M	M	G
G	E	G	E	E	M
O	P	O	O	O	O
P	B	P	G	P	P
H	O	H	J	B	H
B	H	B	B	H	B
J	J	J	P	J	J

Mean Absolute Error					
Total	d01	d02	dLake	50km	10km
C	N	C	C	C	C
N	C	N	N	N	N
G	G	G	F	G	F
F	F	F	M	A	G
M	A	M	E	F	P
E	M	E	I	E	E
A	E	A	A	I	I
I	I	I	G	M	M
P	P	P	H	P	A
O	B	O	O	O	O
H	O	H	P	H	H
B	H	B	B	B	B
J	J	J	J	J	J

When looking at the skill scores (appendix 2) there is a range of results. In the first skill score; I, N, C and G all show skill in the majority of domain 1, C and N then also show some skill in domain 2. Therefore all the runs with the lowest errors show some form of skill. The opposite is also observed, as none of H, J and B show skill in any domain. In the second skill score, the results are much more mixed, particularly in the first domain with many of the runs showing varying skill in the different time periods. For example, B was skilful in November and November/December total but not December alone. C, I and N all show skill in the second domain, and run I shows a small amount in the lake domain. With respect to the worst results, J and B do show a small amount of skill in the first domain whereas H still shows no skill at all.

C and N are consistently the best runs. The common factor between them is the Betts-Millar-Janjic (BMJ) cumulus scheme; all the other parameters are different. Additionally, the MAE shows that run G is also one of the more accurate options, this run is a mixture of the scheme used in C and N, again with BMJ. Both C and G contain both CAM radiation schemes which would be expected to be better as they are designed for a climate model (Wang et al., 2011).

Of the worst runs, two, B and J, contained kain-fritsch (KF) as a cumulus scheme, suggesting that this is not a suitable option for this region. The third run contained a different surface layer which was the 5-layer scheme option instead of the usually recommended Noah scheme, again suggesting that this scheme is unsuitable.

There are definite visual differences between the runs, as well as those observed in the statistics Fig. 2.3 shows the difference between model and observed rainfall (TRMM), for each of the runs in table 2.2. C, N and G all show a very mixed result, as all show a large amount of variation with respect to over or under estimation. One aspect that is very apparent in these three runs is that they appear to represent the Indian Ocean far better than many of the other runs. Run I was also one of the better results, however it shows a very big underestimation over the entire continent. The worst runs overestimate the rainfall, in particular J and B with the KF parameter. In these cases large areas of both the Indian Ocean and the African continent appear to be significantly overestimated.

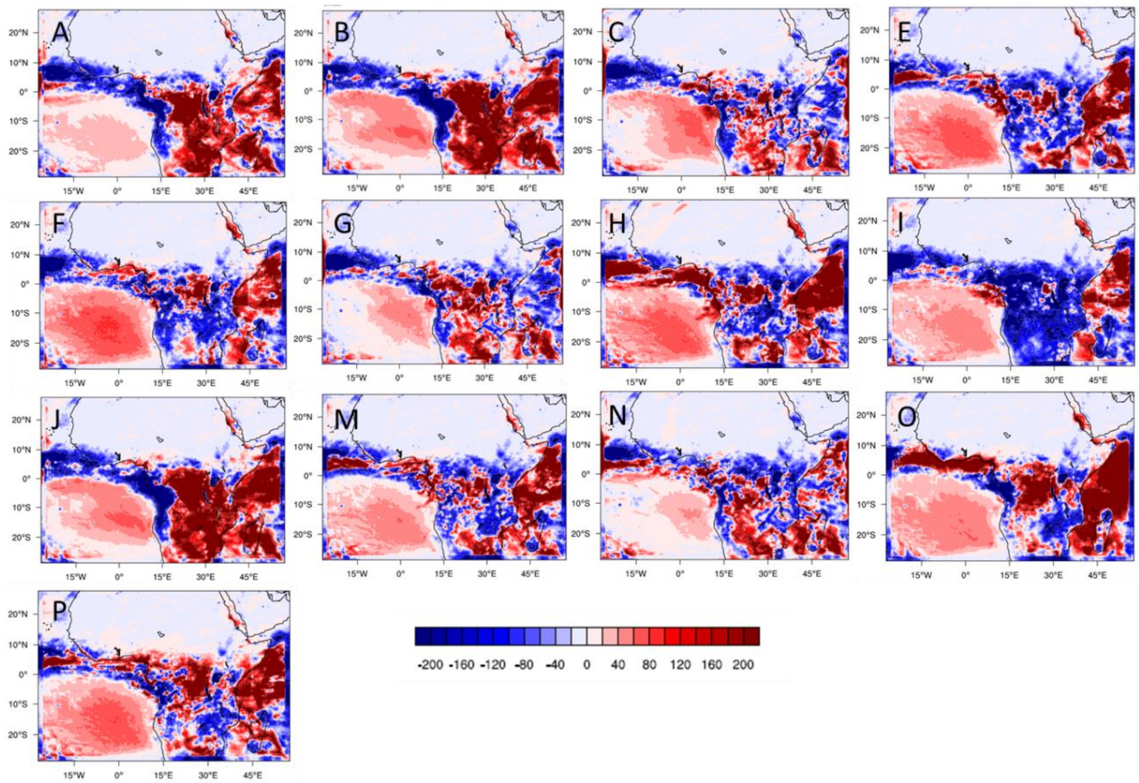


Figure 2.3: Difference between the WRF model rainfall output and the TRMM observations (mm).

When comparing the model rainfall to the TRMM pattern, the majority of the runs overestimate the Congo rainforest and underestimate over the lake (Fig 2.4). While still producing too much precipitation, the runs C, N and G seem to capture the distribution of the rainfall relatively well, in particular over the Indian Ocean when compared to many of the other runs. The remaining runs appear to overestimate to a greater extent. The statistically worse runs produce much more rainfall and additionally don't appear to reproduce the rainfall distribution. Run I performs well statistically but in direct comparison to observations does not reproduce the rainfall distribution well. This highlights the importance of analysing the results in multiple ways, before deciding on the best options.

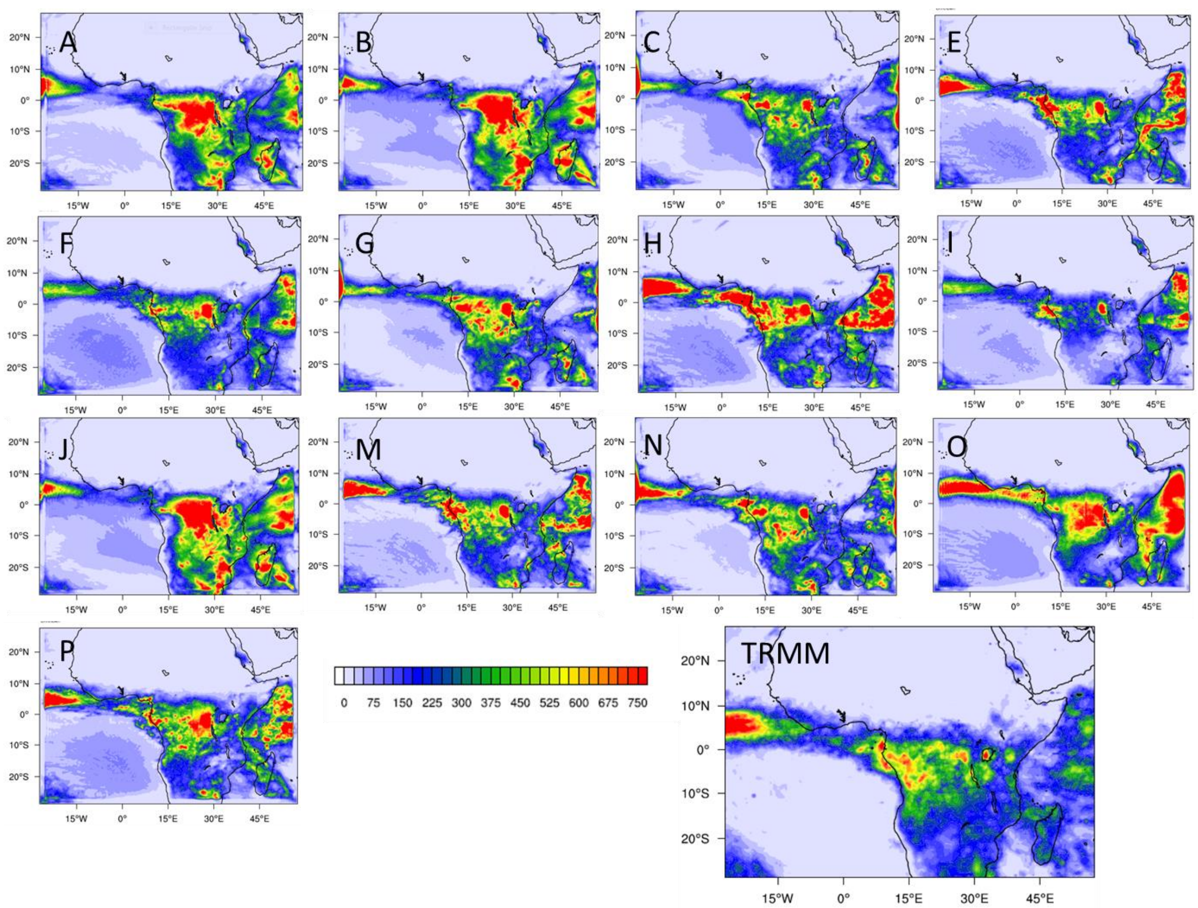


Figure 2.4: November and December rainfall for all runs and TRMM in mm.

When looking at the domain centred on the lake, it is apparent that none of the runs correctly reproduce the rainfall over the lake (Fig 2.5). All of the runs show an underestimation over the lake and none of them show any indication of being able to reproduce the asymmetrical pattern clearly visible in the TRMM data. Some of the runs, in particular those with the smallest errors, seem to correctly reproduce the pattern surrounding the lake. Although many of them seem to overestimate the amount of rainfall, they show a sweeping band of rainfall coming up to the lake from the south west. This adds to the asymmetry in the TRMM data with the eastern shore much wetter than the western shore. The worst quartile runs tend to produce rainfall around the majority of the lake and therefore do not accurately capture the pattern. Run I, again, does not seem to accurately capture any of the rainfall pattern indicating that it may not be one of the best runs, despite good performance in the statistical tests. An additional point is that Lake Victoria isn’t the only lake that the runs reproduce incorrectly. The other lakes on the domain, while much smaller (or only partially included) also don’t appear to be represented accurately.

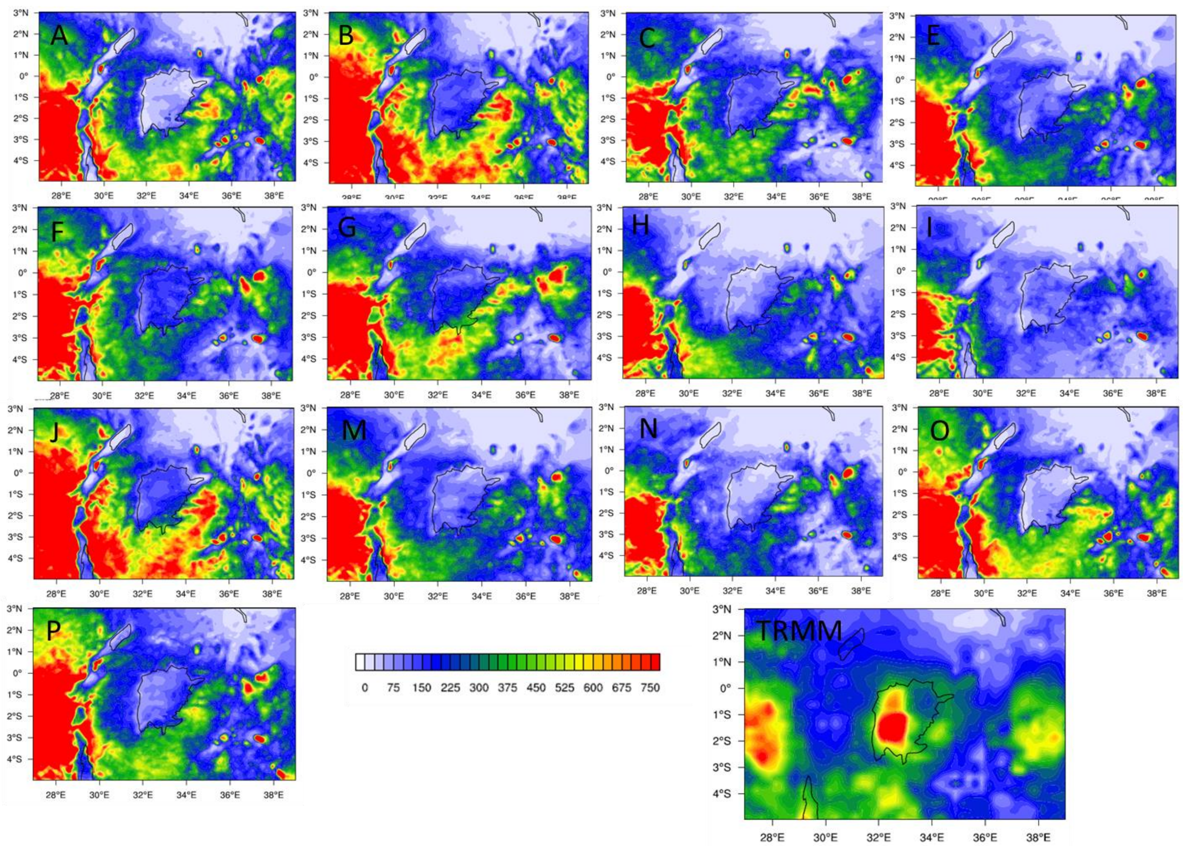


Figure 2.5: As previous figure except for the lake domain.

Finally, two of the runs proposed for the customisation did not complete. Both of them finished with an undetermined error before the full run had been completed. These were the only two runs to contain the microphysics scheme Morrison and it is suspected that this scheme caused some form of instability causing the model to crash. The results that it did produce show a large overestimation over a large proportion of the domain and did not give realistic results. In addition to this, the few statistics that were produced were far worse than the other runs.

Overall, one run alone does not outperform all the rest, however it is relatively easy to summarise which ones consistently result in lower errors and therefore give a better accuracy. In this customisation, two runs seemed to consistently be in the top results with lower errors in all the different domains and grid spacing combinations. These runs were N and C. At the opposite end of the spectrum J, H and B were consistently the worst runs.

2.4.3. Additional runs

For the boundary conditions, in all cases the linear boundary with 4 points of relaxation gave better results. However, the differences between the linear boundary and the exponential boundaries with 4 and 9 relaxation points was on an order of magnitude less than the difference between the upper and lower quartiles shown in table 2.7. The boundary condition is therefore of a lesser importance. The exception to this is the exponential run with 14 relaxation points, in this run, the differences were much larger and this boundary appeared to give significantly worse results. In all cases the RMSE and SD agreed. Finally most of them initially appeared better for domain 1 but were then worse for domain 2.

With the addition of the climate parameters (table 2.4) every run used gave better results. All saw an improvement in RMSE and SD and the improvements increased as the domain got smaller, so it can be concluded that use of these climate parameters do lead to better results.

Finally the land use difference was assessed using two different runs and these came back with opposing results. In one Modis appeared better and in the other USGS. In both cases the RMSE and SD agree with one another. The differences were again small suggesting that in this case it does not have a large influence. Further study of this may be needed to clarify this completely, as the results of this section were inconclusive.

2.5. Conclusion

The study in this chapter has highlighted the importance of customisation for a particular region and the correct selection of parameters. For example, changing the cumulus parameter from the KF to the BMJ can change a model run from being one of the most accurate (C) to one of the least accurate (B). It has also shown that it is always difficult to fully determine what the best results are, as the amount of combinations is too extensive to enable an entirely full investigation and it is also important to look at how each parameter works in each domain and grid spacing.

The hypothesis has been proven correct in that the default settings were not adequate to represent the region of this study and by customising there have been improved results. Over the entire domain customisation has had a big impact, however the lake was represented poorly regardless of the parameter combination.

None of the parameter combinations manage to resolve the complex pattern over the lake, regardless of how well they performed otherwise. Many of them appeared to capture the pattern surrounding the lake, although often not reproducing the correct rainfall amounts. When considering the rainfall over the lake itself, all of the runs underestimated the rainfall. This underestimation was particularly large over the western half of the lake, where greater amounts of rainfall is shown in the observed data. Some of the runs got the correct wet and dry areas on either side of the lake, but failed to reproduce the rainfall over the lake resulting in a dry-dry-dry-wet pattern instead of a dry-wet-dry-wet pattern.

3. Lake Surface Temperature

3.1. Introduction

It has already been established that Lake Victoria is a vital component of the local circulation and therefore is partially responsible for the rainfall pattern in the region. The lake creates a land-lake breeze (Asnani, 1993), which interacts with the large scale flow and produces the asymmetrical pattern seen over the basin. Anyah and Semazzi (2004) found that the LST influences the strength of the breeze, and consequently the amount of rainfall over the lake. Therefore, it is known that the Lake surface temperature (LST) is a critical component in the system.

It is vital that these patterns are captured in model runs. If the model run does not adequately represent the DWDW pattern then it is not reproducing the correct influence from the lake. It has been shown that changing the lake surface temperature of the lake within the WRF model does indeed impact the rainfall over the lake (Xia Sun – Personal communication, 2013). Therefore it appears it is necessary for WRF to have the correct lake temperatures in order to accurately represent the lake.

WRF has two means of interpolating the LST (Wang et al., 2011). The original method, and default setting, takes the LST as a direct interpolation from the nearest resolved SSTs. This method essentially extends the SST field over to the lake area. Due to the different depths, volumes and dynamics this is inevitably inaccurate. This was recognised within the modelling community and a second interpolation method was developed.

The second interpolation method involves air temperature instead of SST. The daily average air temperature over the lake is calculated and this is substituted into the lake as an LST proxy in order to try to avoid unrealistic interpolation of temperatures from the oceans (Wang et al., 2011).

There are many different SST data sets available and it is necessary to choose which one to use within WRF. In the first method of interpolation, the choice of dataset will have an impact on the LST and therefore is likely to be of importance. It should not however have such a large influence on the lake region if using the second method of interpolation.

It is not only Lake Victoria that is likely to be affected by this issue. There are many other large lakes around the world that will impact the surrounding climate yet will not be resolved correctly in the model. This has been a problem in the Great Lakes of North America (Zhao et al., 2012) and can also be seen in this model over some of the other lakes of Eastern Africa (Fig. 2.5). There are undoubtedly other lakes that also fall into this category and are not being represented correctly. This is turn could be resulting in severe errors within the model output.

This chapter discusses the effect that different initialisation and SST datasets can have on the results of a WRF model run. Section 3.3 provides a brief methods summary, 3.4 the results and discussion and 3.5 the conclusion.

3.2. Hypothesis

That the SST data and the way in which the lake surface temperature is initialised will influence the accuracy of the rainfall distribution.

3.3. Methods

In order to investigate this issue, four different runs are conducted. These took into account the two different initialisations and two different SST data sets. Each data set was run with each different initialisation method. Consequently two runs were using the default WRF lake surface temperature initialisation and two were using the more recently introduced alternative initialisation. The specifics of these runs are shown in table 3.1. The model set up for all four runs is the same as combination I from chapter 2. This combination was chosen as it was the best performing combination in preliminary testing.

Table 3.1: The runs produced for this section

Runs	SST data	Initialisation
O	OISST	Default
OA	OISST	Alternative
E	ECMWF	Default
EA	ECMWF	Alternative

Two different SST datasets were used. The first one is OISST which is the optimally interpolated sea surface temperature, details of which can be found in chapter two. The second dataset used is the ECMWF ERA Interim data which is obtained from the ECMWF

Data Server (ECMWF, accessed 2013). This data is at a 0.75°x0.75° and is updated once a day (ECMWF, 2014). The different initialisation methods have been described in the previous section and will be described as the default and alternative initialisation respectively.

It is necessary to look at the differences between the four runs, both to see the impacts of the lake surface temperature on the rainfall and to see the how the model is interpreting the different SST and LST. The difference between the WRF runs and the observations are compared both visually over the whole domain as done previously and also as cross- sections running North-South and East-West across the lake. The cross-sections show the TRMM rainfall, the WRF produced rainfall and the difference in rainfall. These are analysed along with the statistics mentioned in chapter 2 in order to give an indication of which runs produce the more realistic rainfall results. Run OA was used as a control run and all the others were compared. The statistics were considered individually as well as a comparison between them. For example, the difference between the average RMSE of two different runs was analysed. Additionally the SST is plotted from the wrfout files. These are compared and analysed.

3.4. Results and Discussion

3.4.1. Statistical Results

All the statistical results from this chapter can be found in appendix 3. OA outperforms all of the other options in the RMSE score. In the case of both of the E runs, the RMSE of OA is constantly better. The average difference between OA and E or EA is around 5mm and individually the difference ranges from 2 to 9 mm resulting in a small and consistent difference over the three domains. This means that the differences between the two model runs are not biased by a particular region, but are spread over the domains equally. Again this is an order of magnitude lower than those seen in the initialisation chapter. This is not the case for run O; in this run the average difference is 30.9mm and this difference does increase with each domain. The differences in the lake domain are all over 60mm whereas in domain one they are less than 5mm. This makes it clear that the main area of increased error in run O does occur over the lake region, as expected due to the difference in lake initialisation. In all cases the RMSE increases as the domains centre on the Lake, however this is most noticeable in the worst run O.

When considering Standard deviation, the results are slightly different for E and EA. E outperforms OA in the standard deviation, which if following the advice of (Carvalho et al., 2012) would imply that this is the better choice due to a more consistent error. The differences between the two SDs are very small with a maximum of 13mm. EA and OA result in such similar SDs that it can be concluded that there would be no benefit in choosing one over the other. The SD results for O are comparable to the results for the RMSE both in relation to OA and the values produced. Once more all show an increase in SD as the domain size decreases and again it is more noticeable in O. This increase is much less apparent in E and EA suggesting that the lake is a smaller cause of error in these runs.

The Mean absolute error shows very similar results to the RMSE however, in this case the increase with domain size in very obvious in all four runs. The resulting differences between OA and the other runs show a very similar pattern to the RMSE with the same consistency in the differences as exhibited previously. The most apparent dissimilarity is that the average difference between OA and O is much lower, around 8mm. This is probably due to the calculation of the MAE in comparison to the RMSE. As RMSE squares the difference it increases the influence of any large errors within the domain, meaning that the badly represented lake and the resulting errors give a much larger RMSE and therefor a larger difference between the runs.

3.4.2. Visual Rainfall Results

As all four runs contain, the same parameters it is expected that overall they should be relatively similar over land with the exception of lake. This is the case in the large domain (Appendix 4); all runs show a similar pattern in the rainfall differences. This pattern is generally an underestimation of rainfall over the continent and an overestimation over the oceans. Both SST datasets show the oceanic overestimation which suggests that there isn't a large amount of difference between these two datasets. The biggest difference that can be seen, even in domain one, is that the lake is influenced by the changes between runs. In the O run the lake is clearly highly overestimated whereas in the others it is underestimated.

In domain two it is easier to see the differences in the LVB region (Fig.3.1). O with the direct SST interpolation displays the worst results as could already be inferred from the statistics. It grossly overestimates the rainfall by approximately 1000mm and does not capture any of the expected pattern over the lake. It is theorised that this may be due to warmer lake temperatures on the lake causing constant convective systems instead of the usual land-lake breeze. The temperature is too warm to cause the necessary temperature gradient for the land-lake breeze. The remaining options also do not capture the correct pattern even though the differences are smaller. The asymmetry on the lake shows that it does not capture the dry-wet-dry-wet pattern as it is underestimating more over the region that should experience greater rainfall.

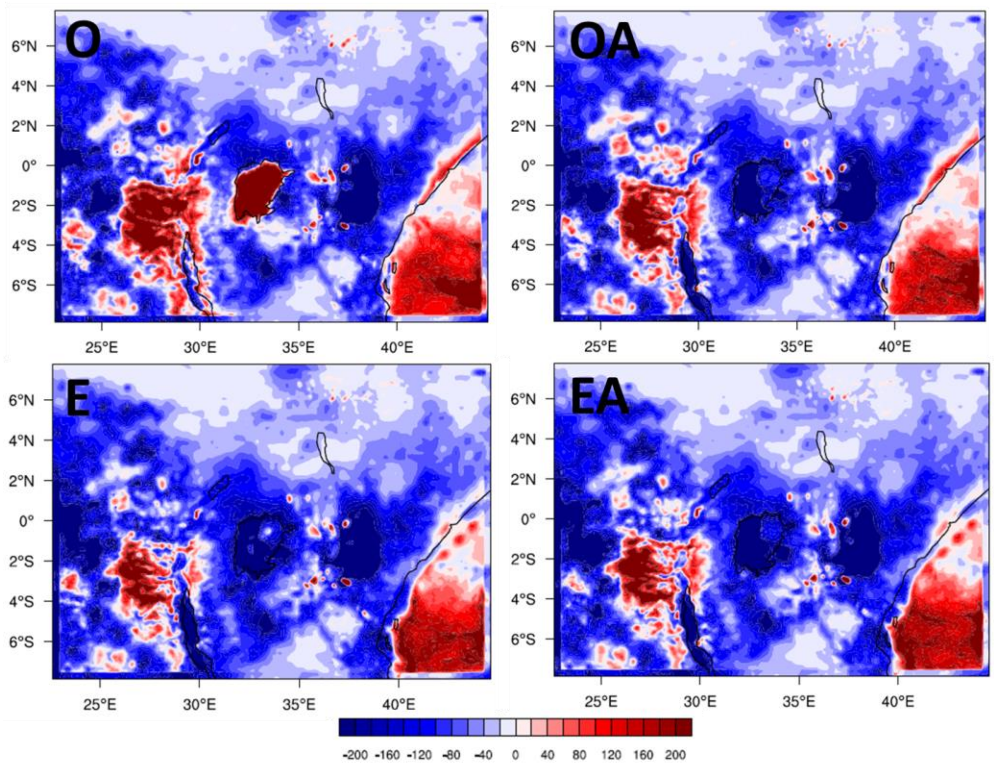


Figure 3.1: November Difference (in mm) between TRMM data and WRF model output over the domain 2.

OA and EA have a very similar pattern over the lake and the surrounding regions. This is expected due to them both being interpolated in the same way and consequently having very similar proxy LST. E also shows a similar pattern, however this was unexpected due to the LST being an SST interpolation.

The inability to capture the pattern can also be seen in fig. 3.2. This shows a cross section of the rainfall in each of the runs as well as the TRMM rainfall. In the TRMM it is evident that the dry-wet-dry-wet pattern exists whereas none of these runs capture that over the lake. On either side of the lake the WRF values are much closer to the correct values but over there is no evidence of the peak in rainfall to the west. E does see a slightly larger peak however this is over the centre of the lake, not towards the western side, and is still producing too little rainfall. It is apparent that O hugely overestimates with the peak rainfall reaching around 1300 mm, in reality the TRMM reaches approximately 350mm over the lake. None of the other runs produce that as all vastly underpredict the rainfall (<200mm). The Cross section patterns in December show similar patterns.

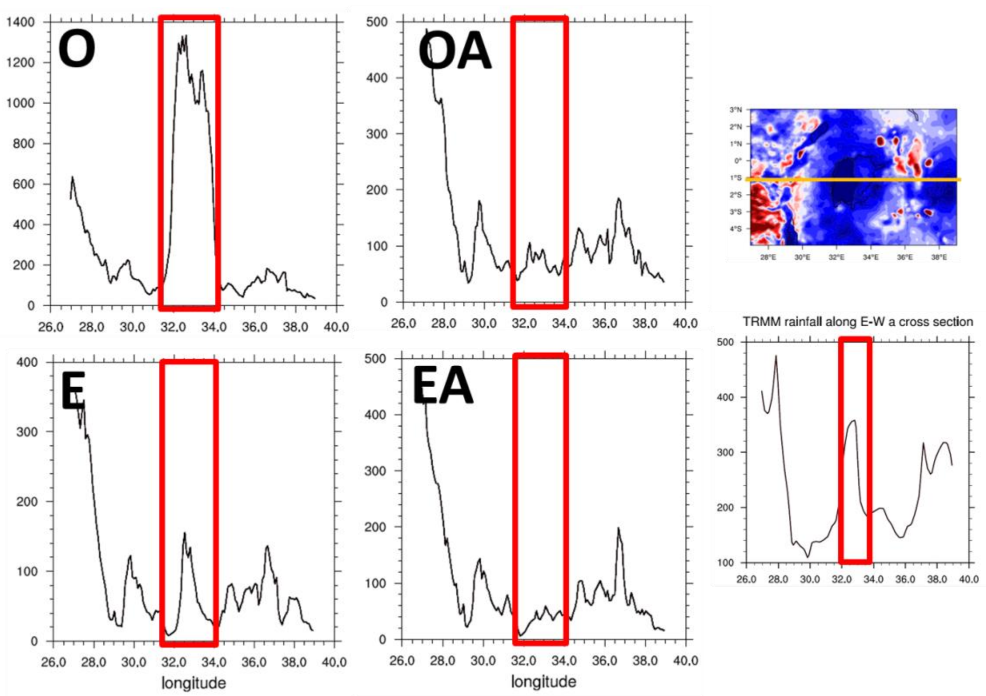


Figure 3.2: East-West Cross section of November rainfall in mm.

3.4.3. Lake Surface Temperature

The main difference between the SST datasets is that the OISST dataset covers the entire globe and has values even where there are landmasses (Fig. 3.3). The ECMWF does not do this and has coastal boundaries. These boundaries are much cooler and it is suspected that the reason the E run is more similar to the alternative initialisation runs is due to the interpolated temperature being cooler by approximately 4°C. The WRF Guide (Wang et al., 2011) specifies that the LST is taken from the nearest resolved water body, which in this case would be the cooler coastal areas, this however has not been confirmed. Other than the land mask and subsequent coastal areas, the SST datasets generally look similar.

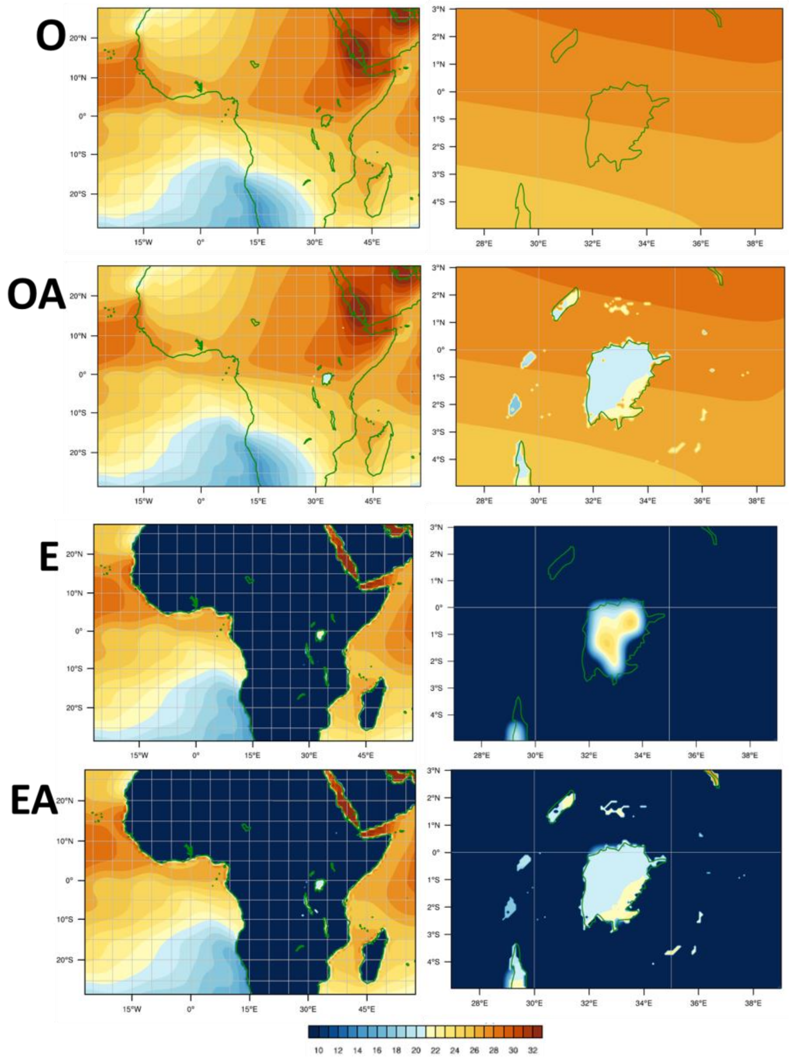


Figure 3.3: The SST and LST for each run and each domain.

An alternative theory for E is that the ECMWF data did include a lake to an extent but not fully as the geographical shape is incorrect. The temperature however is much lower and therefore closer to the proxy values and not O. Additionally the shape may lead to the central rainfall pattern seen in the WRF cross sections. The lake temperature in O appears to be the SST temperatures extended and interpolated over land. EA and OA have the same LST, as expected, and this is lower than both O and E. The alternative initialisation also captures the surrounding lakes, including many of the smaller ones, whereas the default option does not.

The much warmer LST temperatures in the O run support the idea of the convective system and the lack of temperature gradient. The lower temperatures in the other three runs would cause a lower amount of moisture and therefore less rainfall.

3.5. Conclusion

In this case the hypothesis has been confirmed; both the initialisation and the SST data have the potential to influence the rainfall and how accurately it is represented. However when using the second method of initialisation the SST data becomes of less importance to the lake region as the initialisation then uses air temperature as a proxy. The alternative initialisation did lead to an improvement but it is not fully resolved as the lake still appears to be a significant area of error. Overall OA is the best option as it outperformed the others in the RMSE and MAE and performed comparably in the SD. Additionally it resolves the geographical extent of the lake and the other smaller lakes.

The initialisation, and the consequent lake temperature, can have a huge impact on how well the lake is represented. In one case (O), the rainfall is dramatically overestimated and in the others it is underestimated. This means that should it not be correctly initialised the results produced by the model may not accurately represent what is actually happening. Even with the second initialisation this model still isn't producing the correct rainfall pattern over the lake as it does not reproduce uneven rainfall over the lake. This problem is in need of further investigation and potentially a coupled model. This could also be an issue with lakes in other regions and it is important to be aware of the difficulties and problems posed by the lake not being represented correctly.

4. Optimal Customisation

4.1. Introduction

Different parameters and schemes are designed to change the physics and dynamics within the runs produced by WRF. This is in order to account for differences in run specifications, such as location and length of run. As they affect the physics and dynamics, they will undoubtedly impact the circulation over the lake basin and consequently the rainfall in the region. An understanding of how particular schemes or parameters impact the circulation and rainfall will lead to a better understanding into why some schemes work and others do not.

It is hoped that the runs that gave the best results previously will capture the circulation over the lake resulting in better accuracy. In addition to this, the physics analysed may give an indication of the aspects of the circulation that the model is able to reproduce along with the ones that it is unable to capture. Additionally it is expected that the worst ones will show significant differences from the best runs and that this will also indicate which elements of those runs the model is unable to reproduce.

This chapter examines these best and worst runs in more detail by looking at other variables in addition to the rainfall and analysing the differences between the runs. As in previous chapters, section 4.3 examines the methods, 4.4 the results and discussion and 4.5 the conclusion.

4.2. Hypothesis

The physics in the best model runs will show distinct differences to the physics in the worst runs and the best runs will capture more of the diurnal cycle and asymmetrical pattern over the lake.

4.3. Methods

The runs chosen for this chapter were the ones that gave the best and worst results from preliminary testing. These preliminary runs masked out the lake region, as the lake was badly represented, therefore biasing the results. The best three runs and the worst four were chosen. Three runs were chosen as this represented the upper quartile, four were chosen for the worst runs due to very similar results meaning it was difficult to determine which run outperformed the other. These runs were not covered in this document, however they followed the same methods as in chapter 2. The combinations and parameters used can be seen in the table below (table 4.1).

Table 4.1: The combinations used in all the runs for this chapter

Run	LW	SW	Cu	Mphys	PBL	Surface
Abest	CAM	CAM	Grell 3D	Eta	ACM2	Noah
Bbest	CAM	CAM	BMJ	Eta	ACM2	Noah
Cbest	CAM	CAM	Grell 3D	Eta	Yonsei	Noah
Wworst	CAM	CAM	kf	Single 5	Yonsei	Noah
Xworst	CAM	CAM	Kf	rrison dou	ACM2	Noah
Yworst	CAM	CAM	Grell 3D	Eta	ACM2	5-layer
Zworst	CAM	CAM	Grell 3D	rrison dou	ACM2	Noah

The runs for this chapter were started on Sept 12^{th} in order to avoid any spin up complications while allowing analysis of the whole three months (OND). The remaining set up is the same as concluded in the final section of the second chapter; an exponential BC with a relaxation zone of 9, use of the climate parameters and using USGS land use data. The use of the climate parameters specifies that the radiation scheme has to be CAM which does eliminate the ability to determine if the radiation schemes have a large influence.

For this chapter the same statistics as previously mentioned are used, RMSE, SD and MAE. Due to the smaller number of runs, instead of looking at the quartiles, the runs were ranked based on how they performed in each measure, over multiple domains and temporal resolutions. The same plots to be analysed visually have also been made. Additionally, plots have been created of temperature, vector wind and the vertical and horizontal motion over the lake using cross sections, East-West over the lake. Temperature is the monthly average, whereas the plots showing motion and wind are considered at four different times throughout the day, 00z, 06z, 12z and 18z. This should give an indication of how the diurnal features and physics over the lake are being represented.

The temperature was compared to data from the Climatic Research Centre at the University of East Anglia (University of East Anglia Climatic Research Unit (CRU)., 2008). This provides monthly data on a 0.5°x0.5° grid spacing.

4.4. Results and Discussion

4.4.1. Rainfall

When considering the difference between the WRF and TRMM, all the best runs show a similar pattern over domain 1 (Fig. 4.1.) They tend to overestimate the rainfall over the Atlantic Ocean by approximately 100mm, and central continent, around the region of the Congo rainforest by several hundred mm. Simultaneously, the Sahel region is generally underestimated. Two of the runs, Abest and Cbest, both also overestimate the Indian Ocean, whereas Bbest shows more of an underestimation. This is similar to the results in chapter 2. Bbest contains the BMJ scheme which appears to do very well over the ocean. All of them underestimate the lake which can be seen even at this scale and grid spacing.

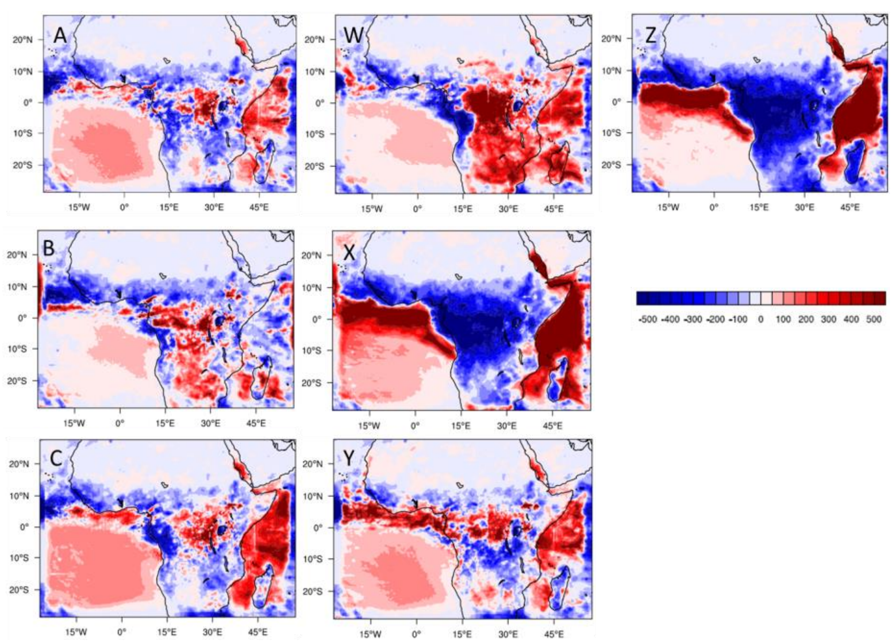


Figure 4.1: Rainfall difference between WRF and TRMM over the total time period (OND).

The worst runs show a wider variation. Runs Xworst and Zworst both severely overestimate over the oceans and underestimate over the continent (>500mm). They both contain the Morrison Microphysics scheme and are the only two runs to do so. It appears that this scheme is unsuitable for these runs as these were also the combinations that did not complete in chapter 2. Wworst and Yworst both look similar to the best runs and seem to capture the pattern of rainfall, however they overestimate more than the best runs, resulting in higher error scores.

When comparing the wrfout rainfall to the TRMM rainfall (Fig. 4.2), all the best runs and W and Y worst show a similar pattern to the TRMM rainfall, the main difference being the amount of over or underestimation that each captures. Xworst and Zworst however show a completely different pattern that has very little rainfall overland and therefore appears unrealistic.

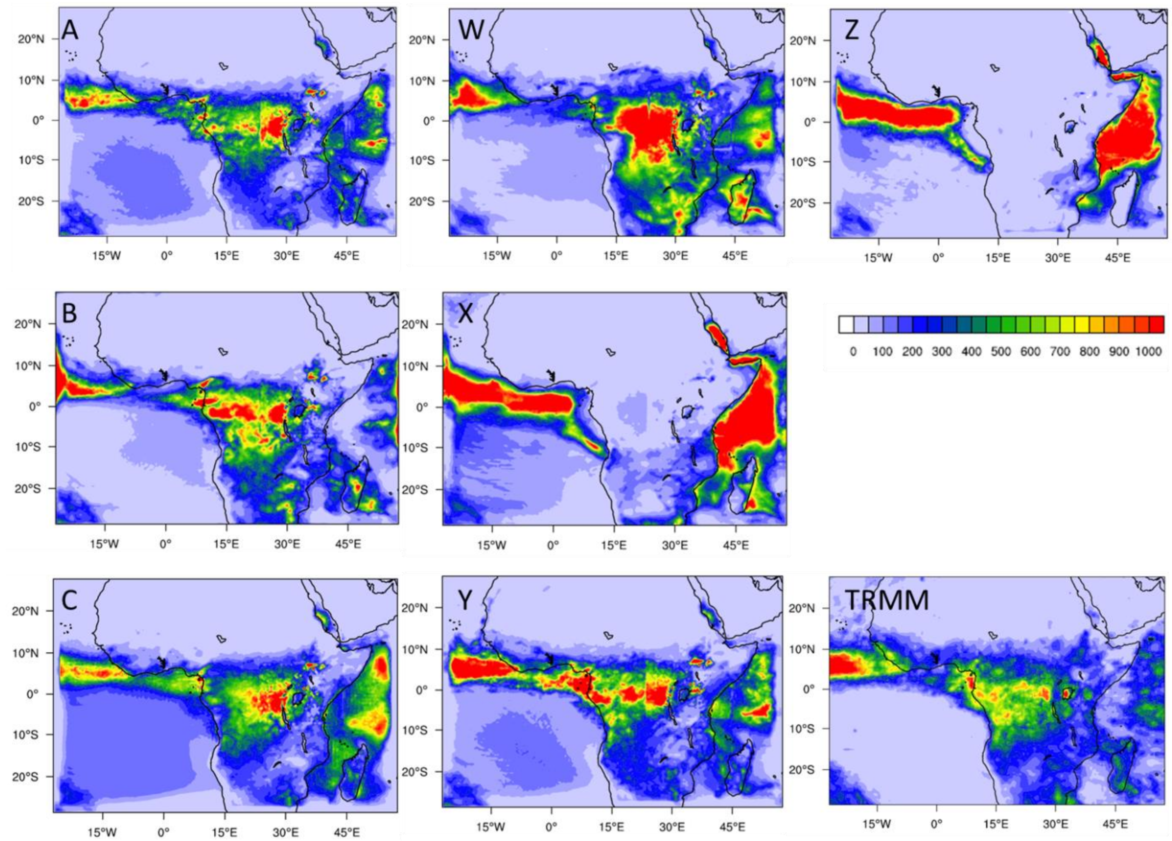


Figure 4.2: Total Rainfall (OND) over domain 1 for all runs plus TRMM.

Figure 4.3 repeats figure 4.1 for the domain with higher grid spacing, all the best runs, Wworst and Yworst show a similar result with large overestimation to the west of the lake and otherwise a general underestimation over land. Bbest again performs best the Indian Ocean, due to the BMJ scheme. Wworst shows a particularly bad overestimation over the Congo rainforest. It can be theorised that this would be due to the combination of the KF cumulus scheme and the Single five microphysics scheme not resolving the moisture and storms that occur over the rainforest accurately. The same inaccurate pattern from domain one can be seen in Xworst and Zworst. These results are confirmed by looking at the rainfall instead of the difference in domain 2 (not shown). Abest and Bbest seem to show the closest results as the rainfall appears less than the others. In addition to this, Bbest also shows the most accuracy over the Indian Ocean.

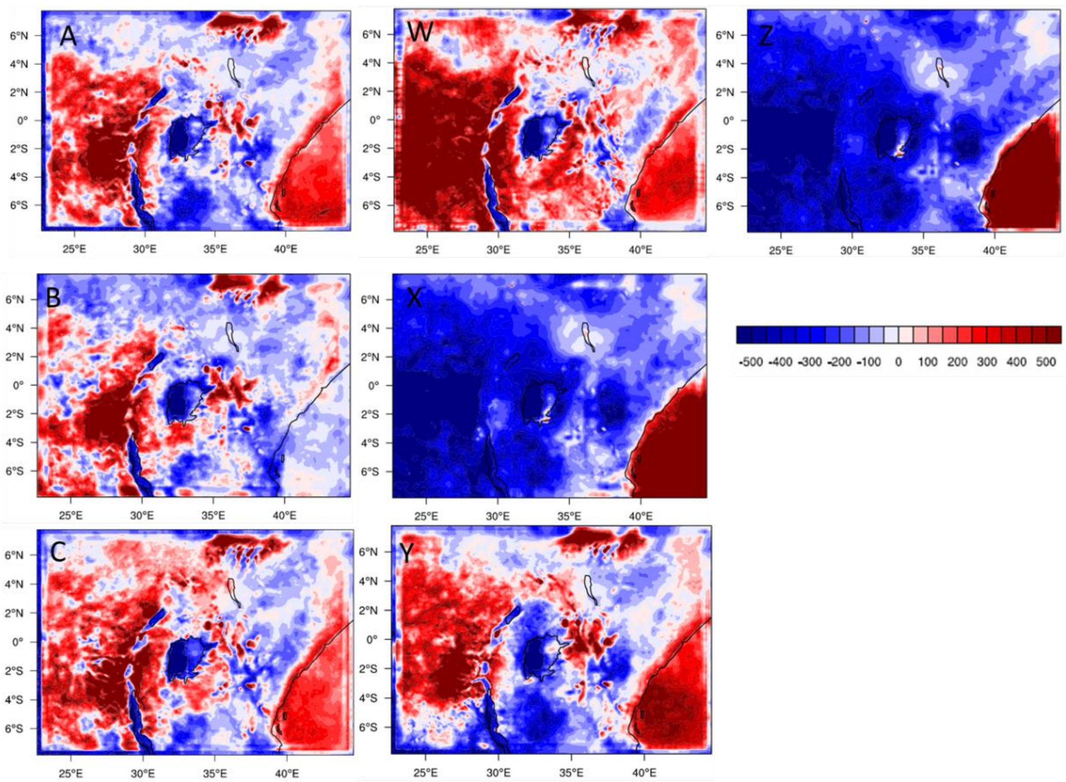


Figure 4.3: As 4.2 but for Domain 2.

The lake domain shows the same patterns as the larger domains, but it also emphasises the fact that none of the runs capture the asymmetrical pattern that is known to occur over the lake. Regardless of how good the RMSE, SD and MAE are and which combinations are used, there seems to be a fundamental flaw in the representation of the lake; the correct pattern does not occur in the models. The asymmetrical pattern shown in the difference confirms that the real asymmetrical pattern is not captured as the area of high underestimation represents the area that should see large amounts of rainfall.

When looking at the TRMM rainfall, the DWDW pattern occurs over the lake but none of the runs produce this (Fig 4.4). Some of the runs show a very small increase over the lake but the amount is far too small, under by approximately 800mm, and the positioning is not correct (the west side that should see greater rainfall sees less). Abest, Best and Yworst seem to capture the dry area to the west of the lake and the wet area to the east, but do not resolve the lake correctly.

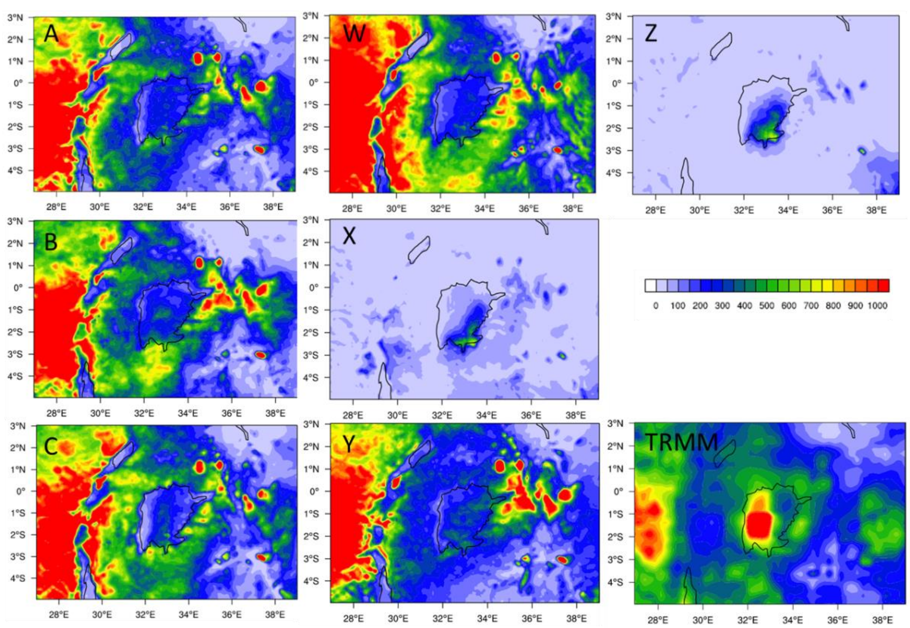


Figure 4.4: As 4.3 but for the Lake domain.

4.4.2. Statistical Results

The RMSE and MAE can be found in Appendix 5, the SD in table 4.2. When considering the RMSE overall, the best three runs outperformed the worst, however when looking at the details it less clear cut. Domain 1 and 2 both show the best three runs doing well, unlike the smaller lake domain. In domain one Abest shows consistently better results and while Xworst and Zworst perform worse than the others, there is greater ambiguity in the other runs. In domain two, this ambiguity increases although overall the results are the same and the best runs still show lower RMSE than the worst. It is when looking at the lake in detail that the results become less defined and in this case run Wworst actually out performs Bbest.

The SD (Table 4.2) shows very similar results in domains one and two with the best runs showing lower SDs than the worst and therefore being classified as better. However, the opposite it true when focusing on the lake. The two runs Xworst and Zworst that had previously been producing the worst error suddenly show the best SD. These runs show an underestimation over the entire domain so it is suspected that this is due to the error being very consistent and represented as a constant bias giving a lower SD than the other runs.

Table 4.2: The Standard Deviation over all the different domains and time periods (ND = November, December). Green indicates low error fading to red with high error.

SD			d01 original					d01 to d02		
	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
Abest	150.031	70.6108	65.5568	69.1231	114.242	143.679	66.6077	63.1994	66.0517	110.111
Cbest	172.441	71.6889	72.1205	78.7199	133.071	166.004	67.7859	70.13	76.1811	129.156
Bbest	183.74	85.5213	88.5096	74.2659	139.474	162.726	79.1349	79.7832	67.0721	124.007
Worst	195.869	77.0793	87.9037	99.6821	163.275	191.423	74.9544	86.0833	97.2073	160.051
Yworst	174.939	86.1737	74.3922	86.0161	135.897	167.861	81.6938	71.5088	82.7311	131.518
Xworst	399.513	142.426	141.827	198.668	307.064	398.828	141.421	141.036	197.671	306.327
Zworst	400.48	156.179	134.821	187.858	293.131	398.724	154.237	133.58	186.246	291.682

SD			d02 original					d02 to d01		
	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
Abest	286.487	131.012	119.202	102.125	202.199	272.546	114.624	117.076	99.6911	198.342
Cbest	268.079	124.904	111.067	100.253	188.184	242.661	104.878	104.538	96.4674	177.709
Bbest	264.101	117.49	130.666	82.8846	190.407	259.699	110.666	131.276	79.8335	188.525
Worst	334.412	122.675	141.243	135.085	252.173	330.804	118.318	140.179	132.86	249.709
Yworst	302.736	148.852	113.02	112.481	205.836	285.601	131.501	109.875	110.698	202.237
Xworst	461.482	162.506	171.348	201.534	344.855	452.197	160.192	170.083	194.706	337.942
Zworst	449.954	196.416	140.245	178.279	299.641	462.746	200.581	142.011	184.705	308.284

SD			Lake d02					Lake d02 to d01		
	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
Abest	361.378	126.871	164.574	120.339	266.531	352.156	111.843	162.918	118.339	264.313
Cbest	357.178	127.006	162.089	123.357	266.763	335.102	109.365	155.142	117.508	255.193
Bbest	384.15	138.177	196.522	113.202	286.885	368.922	126.787	191.99	106.52	276.454
Worst	402.191	120.113	187.533	154.243	307.843	383.519	109.769	179.8	139.699	295.068
Yworst	340.745	155.394	131.45	109.185	224.011	316.445	131.09	127.384	107.718	218.855
Xworst	206.697	103.196	93.4642	83.2585	154.455	205.201	102.402	94.1146	84.0335	154.673
Zworst	202.171	103.964	88.917	83.9296	153.101	200.822	103.243	89.2919	85.0797	153.838

The MAE shows slightly different results for domain one and the lake domain while d02 shows results consistent with the RMSE and SD. In d01 and the lake domain, Wworst performs better than Cbest. The remaining worst runs do not perform as well as Cbest. This means that in two of the error tests Wworst produces results better than some of the best runs. Overall though it does not outperform one run in particular, meaning that the best runs still have the better results.

The skill scores show very little skill outside of domain one (Appendix 6). Conversely, the only case that doesn't show skill in skill score one in domain one is in November in Bbest (Figure 4.5). When comparing November rainfall it does appear that the pattern isn't represented correctly. In the TRMM data the rainfall is south by the coast as well as further east over the Congo Rainforest. Bbest does not capture this coastal rainfall but some of the others, including Yworst do. Outside of domain 1 only 2 sections show skill and they are both in December of domain 2 for run Bbest. This occurs even though the rainfall distribution in the TRMM data is picked up by the best runs. This is due to the incredibly large (>1000mm) overestimation all of the runs exhibit to the west of the lake (exception Xworst and Zworst).Using the second skill score the skill is significantly decreased, with only Abest and Cbest showing majority skill in d01. Xworst and Zworst show a small amount of skill in d02. None of the runs show any skill in the lake domain in either of the skill scores.

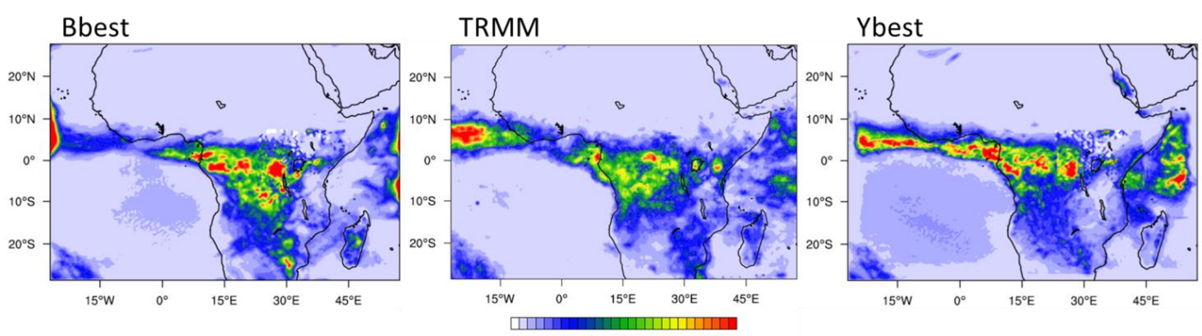


Figure 4.5: November rainfall for Bbest, TRMM and Ybest

4.4.3. Other Variables

4.4.3.1. Temperature

The differences between the WRF output and the CRU temperature appear to be fairly consistent with little variation over the different time periods. Over the majority of the continent, the WRF model underestimates the temperature. The only exception to this appears to be a small region in the south west of the continent (Fig. 4.6.). Therefore in general the model produces a negative temperature bias. This negative bias may influence the rainfall; a large amount of the continent sees an underestimation of rainfall, which may be linked to the fact that the temperature is colder than the observations. This however does not apply to the Congo Rainforest, where often the temperature is lower than the observed data. Despite this, the model overestimates rainfall. The two runs that very much followed that rule were Xworst and Zworst, which were far colder than the observations and also had very little rainfall.

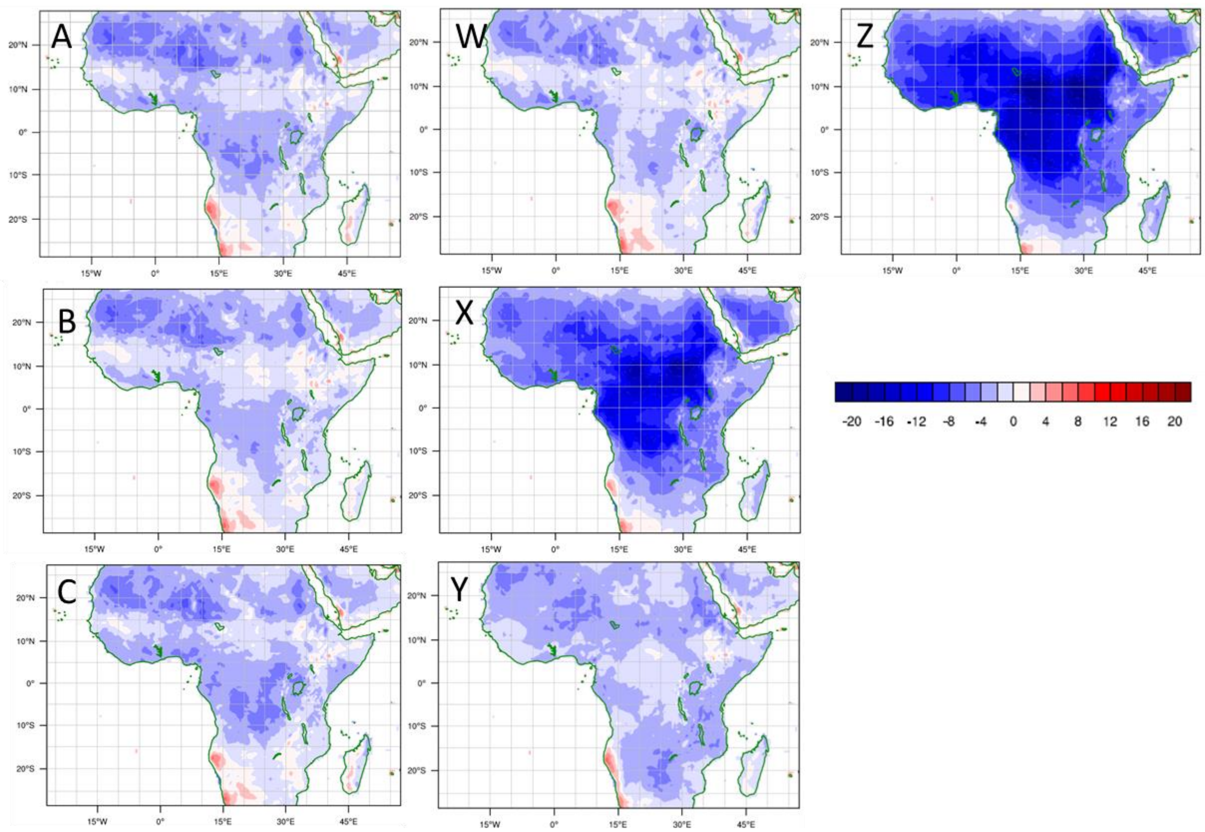


Figure 4.6: Difference between mean OND CRU temperature and WRF temperature (°C).

Domain two follows the same pattern with the exception of some small regions that are either far cooler or far warmer than the CRU observations (not shown). These appear to occur over the high altitude regions, so it may be that the model does not quite resolve these correctly. Over all three domains, Bbest is the closest to the observations. These are just the average temperatures over the entire time period and it may be that looking at the diurnal cycle or the maximum or minimum may give a better indication of how the models are representing the temperature.

4.4.3.2. Relative Humidity

The relative humidity was considered at six hourly intervals in order to determine if any of the runs picked up the diurnal cycle. In some cases there does appear to be a diurnal pattern that would fit with the expected rainfall (Fig 4.7). That is:

- 0z – high moisture over the lake
- 6z – Moisture moves west over the lake
- 12z – low moisture over the lake but high surrounding it
- 18z – moisture appears to be moving west.

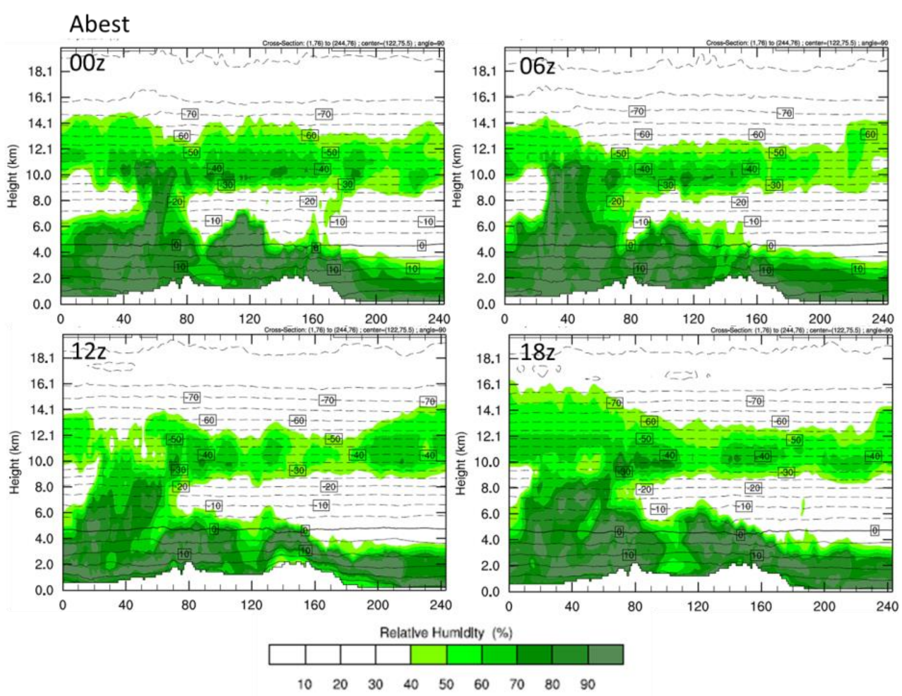


Figure 4.7: Relative Humidity throughout the day for run Abest. Day shown is the 22nd of November and the lake is situated in the centre of the image, at approximately grid spaces 100-130.

As the daily rainfall occurs at night, there should be high humidity, with low humidity during the day when the land-lake breeze is flowing inland. This pattern did not appear consistently but was found over some days in all the best runs. It was found less frequently in the worst runs. Most of the runs showed a much higher humidity over the rainforest which extended further into the upper troposphere. Very little moisture over the lake seemed to extend into the upper troposphere therefore making it appear unlikely that large cumulonimbus clouds were forming over the lake as would be expected during the diurnal cycle. In the most runs the moisture does appear to be moving to the west which would agree with the mean flow over the region.

Wworst appeared to have low humidity over the lake and didn't appear to follow any of the diurnal pattern; Yworst followed the pattern slightly more. Xworst and Zworst display an unrealistic humidity pattern with low humidity at the surface and very high humidity in the upper atmospheres.

4.4.3.3. Wind Flow

An EW cross section of the wind is also used to determine if the diurnal cycle is captured. The wind was considered using the three different components U, V and W, where the U and W winds were more indicative of the diurnal cycle. The U wind did capture what would be expected from the diurnal cycle as well as the mean easterly flow. In the Abest, Cbest, Wworst and Yworst runs there appears to be a reversal of the wind direction around the lake during the day. Fig. 4.8 shows the wind flowing off the lake during the day and towards the lake at night. This is consistent with what would be expected due to the diurnal cycle and the land-lake breeze, however the wind flowing towards the lake at night sometimes appears to be further west than expected. In Bbest, the off lake flow during the day seemed to be there but the on lake flow did not, therefore it was not allowing for the nighttime convergence. The Xworst and Zworst seemed to catch a very weak version of this pattern, although the upper level winds did not appear realistic.

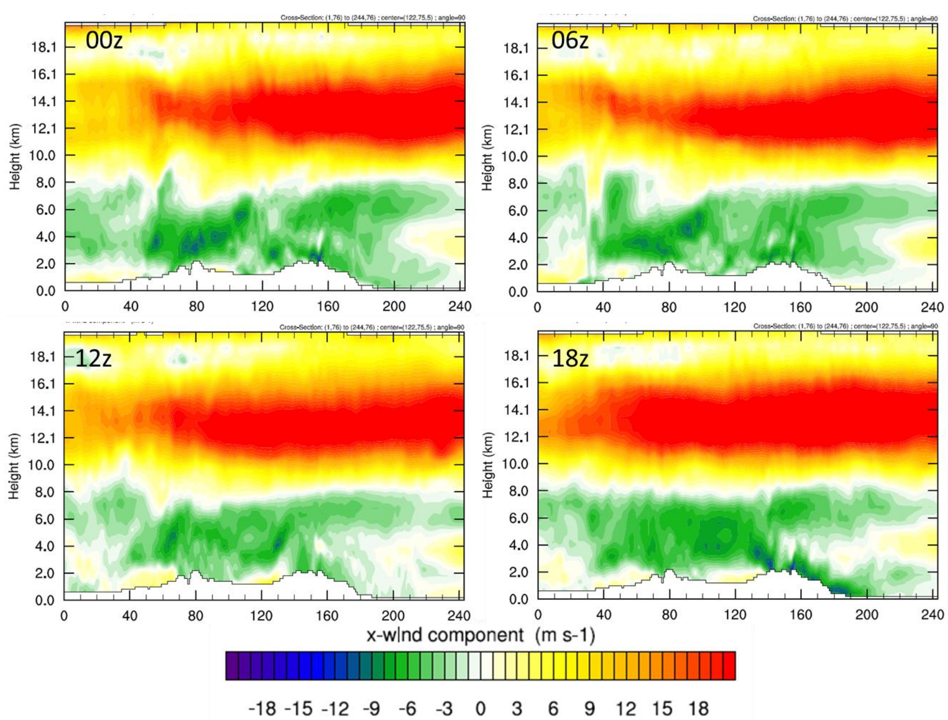


Figure 4.8: As 4.7 but for U wind.

The V wind did not show a large amount of variation between the different runs and did not show any evidence of the flow being represented by the model. The W wind was slightly more informative about the diurnal cycle. Ideally there would have been rising motion overnight and sinking during the day in order to fit in with the land-lake breeze. This did seem to occur within some of the runs in particular Abest, Cbest, Wworst and Yworst however the upward motion in some of these runs appeared to be very weak. The upward motion was also not consistently visible and often the downward motion seems to be much stronger than the upward. Bbest didn't capture the vertical distribution as well as the others; the upward motion did appear to be stronger but occurred at all times so was not consistent with the land-lake breeze pattern. Xworst and Zworst showed very little vertical motion in the troposphere, which did not show any similarity to the pattern expected.

Finally vector winds were also analysed (Fig 4.9). Nearly all the runs seemed to capture the land-lake breeze flowing off the lake but not flowing onto the lake. In many of the runs the flow away from the lake continued at all times. This would explain the lack of rainfall over the lake as the air is descending and therefore there is no convergence, convection and consequently no rainfall. It would also explain why the previous cross sections always showed the daytime flow much stronger than the nighttime flow. This may be due to the colder model temperatures which, instead of a reversal of the land-lake breeze at night, cause the flow to continue in the same direction or only cause a very weak reversal. This would be due to the lake not being warmer than the surrounding land, but this will need further investigation.

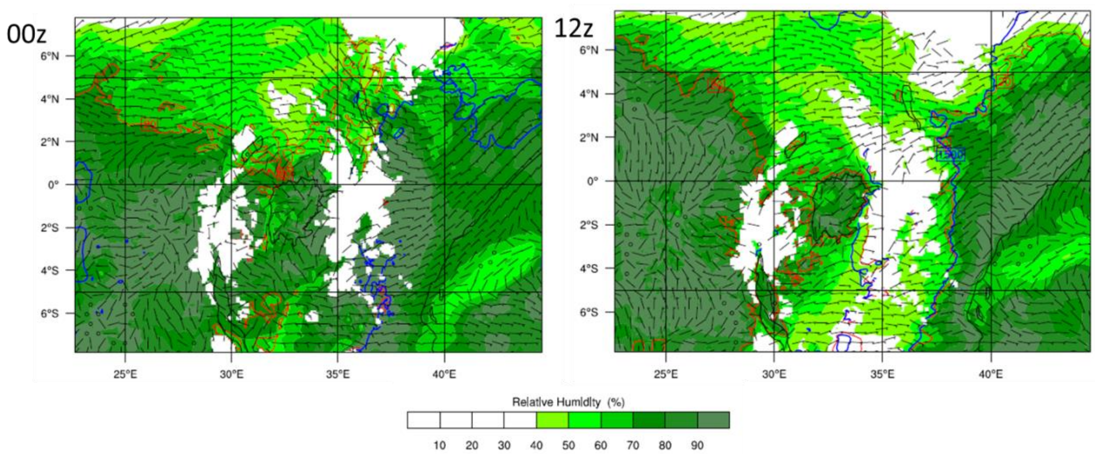


Figure 4.9: Abest on the 22nd of November showing the same wind direction during the day.

4.5. Conclusion

Our hypothesis was both accurate and inaccurate, as the different model runs did show distinct differences between one another yet these differences were not always directly separated by the best and worst run divide. In addition to this, the best runs did not show the diurnal cycle or asymmetrical pattern more than the worst runs as none of the runs managed to capture this pattern correctly.

Some elements of the diurnal cycle are captured; the flow during the daytime is exhibited by the models. The land-lake breeze during the daytime is reproduced, however it does not seem to exist at night. It is theorised that this may be due to the temperature of the lake still remaining too low and the cycle not reversing as the lake is never warmer than the surrounding land. This constant one way flow or very weak reversal would explain the lack of rainfall over the lake, as it is primarily due to convergence at night with the reversal of the land-lake breeze. The lack of rainfall means that it does not pick up the asymmetrical pattern but also that it is impossible to judge if the pattern would be there should the model be producing the correct amount of rainfall. Before it is possible to ascertain for certain whether the model reproduces the pattern, it is necessary to get the fundamental flow over the lake correct and at present time this is not occurring.

Over the rest of the domain, in particular in domain one, the distribution of rainfall is often reproduced, qualitatively if not quantitatively. . The model shows a tendency to underestimate over land and overestimate over the oceans and the Congo Rainforest. This could be linked to the fact that the temperature tends to have a cold bias of a few degrees, which could have an impact when looking at future changes as the bias is around the same order of magnitudes could be expected in future warming.

Finally it is clear from these runs that some parameters do not work in particular situations. In this case, the Morrison microphysics scheme did not work and was not suitable for these runs. It did not produce realistic results and it is important to note what the runs are doing and not just the statistical results as even if a run completes it does not necessarily mean it has produced accurate and reliable results.

5. Extreme Years

5.1. Introduction

Thus far the analysis has focused on a year representing the climatology according to Pohl et al. (2011). As mentioned in chapter one, there are several large scale features that have the potential to influence the region, such as ENSO and the IOD. As this year is the climatology, it does not take into account any of these large scale features. If WRF is to be used for future runs, it also needs to be to reproduce the rainfall from these years to the same extent.

As discussed in chapter one, these large scale features are often linked to extreme rainfall either influencing very heavy rainfall with the potential to cause flooding or creating years of drought. It is necessary that these extreme years can be reproduced, however it is unclear that the results for these extreme runs will be comparable to the results for the climatology year. These schemes may no longer be the optimal combinations due to a change in the physics and the influence upon the region.

This chapter explores whether this is the case and if the optimum combination reproduces the rainfall for the extreme year to the same accuracy as it does for the climatology. The sections will follow the same order as in previous chapters.

5.2. Hypothesis

That the best combination will also provide accurate results for years with other influences such as ENSO.

5.3. Methods

To investigate whether the best combination of parameters is also capable of capturing years with more extreme variance a selection of runs are performed covering different extreme scenarios. The set-up for the runs are exactly the same as that of run Abest as used in chapter 4, the combination of parameters for which can be found in table 4.1. The data used to initialise and validate the model is the same as the previous runs, the exception being that the years have changed. Four different years are chosen to represent extremes.

The analysis used to determine if the model can also replicate the rainfall during extreme years is the same as that used in the previous chapters. RMSE, SD, MAE along with the visual plots are all used. The results for Abest in the previous chapter represent 1999 which is acting as a control run. The differences between the statistics for each extreme year and the control run are calculated.

It has already been discussed that a large amount of the variation within the short rains is due to a combination of ENSO and IOD. Consequently it was decided that the extreme years would be based on the indices of these two teleconnections (NOAA/ National Weather

Service, 2014; JAMSTEC, 2012). Additionally an EOF of TRMM data over the region (Masilin Gudoshava – personal communication, 2013) was used to confirm both that this is the main source of variability and the years which experienced the largest influence. These years may not necessarily have the greatest or smallest rainfall, however they are determined by the different modes of these teleconnections and it is the changes in dynamics that are needed to ensure comparable results. The years and their specific characteristics are listed below:

- The positive years are 2002 and 2006, both of these years had positive ENSO and IOD indices. 2002 was a stronger ENSO year with an index of 1.3 and an IOD index of 0.27 averaged over the three months. 2006 had an ENSO index of 1 but a stronger IOD index of 0.57.
- The negative years are 2005 and 2010. 2010 is much stronger than 2005 in both indices with ENSO at -1.5 and IOD at -0.21. 2005 has indices of -0.16 for IOD and -0.5 for ENSO.

TRMM data shows that this resulted in greater rainfall in the positive years and less rainfall in the negative years (Fig. 5.1).

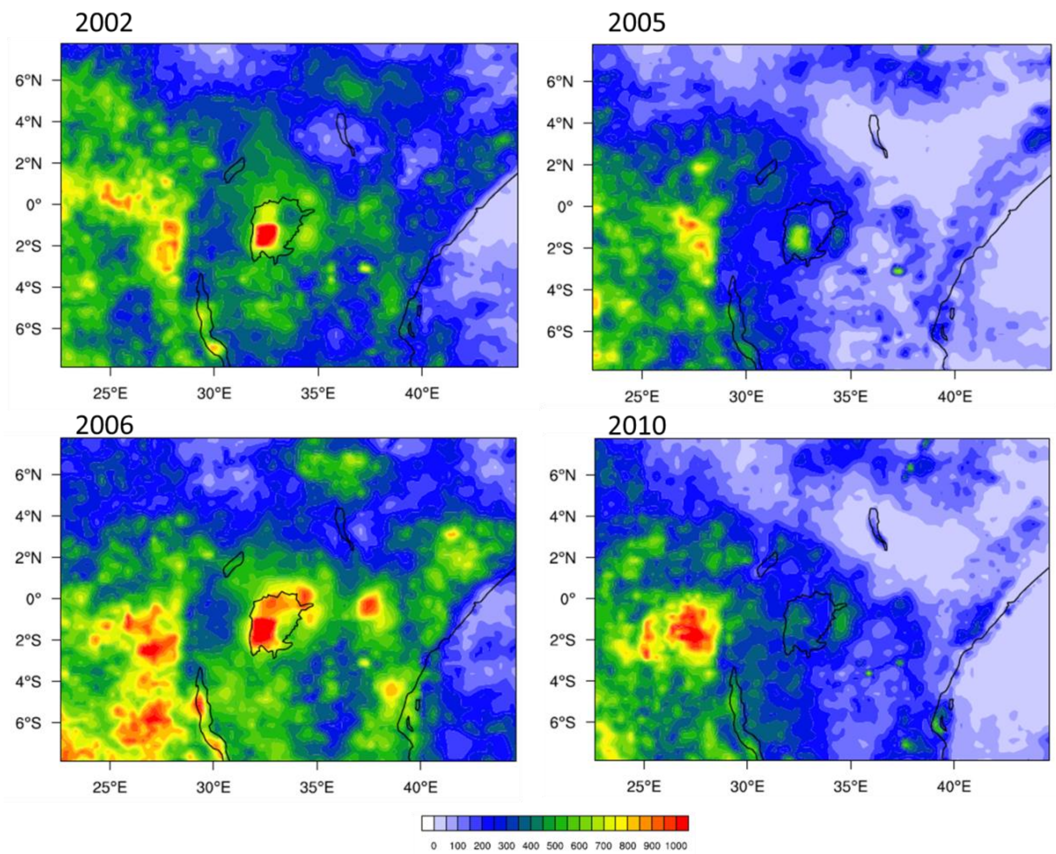


Figure 5.1: Total TRMM rainfall for OND for each extreme year

5.4. Results

Statistically the extreme years showed a range of results (Appendix 7). Out of the four runs conducted only one appeared to perform similarly, if not slightly better, than the control run. This run was for the year 2005; the remaining runs all performed worse to greater or lesser extents. 2005 showed a very small increase in error in the RMSE but then showed a decrease in both the SD and MAE.

2006 was the second closest but did see a decrease in accuracy. This change did not seem to be influenced by domain; no increase or decrease of error was found as the domains focused upon the lake. On average, 2006 saw an average difference over the individual months from the control run of about 9mm in the RMSE.

The other two runs were substantially less accurate. In particular 2010 was badly represented and had the worst results statistically. It had an average difference in RMSE over the individual months of 40mm and over the lake domain of nearly 70mm. Both 2002 and 2010 showed the accuracy decreasing as they focused on the lake, implying that the main problem and source of error was due to the area around the lake. In this case, it may be East Africa in general as it is the East African region that is influenced by changes in ENSO.

The runs increase in error in the same order as the increasing strength of the ENSO index. This order is 2005 with the best accuracy, 2006, 2005 and then 2010 with the worst accuracy. The runs were chosen using both ENSO and IOD but it appears in this case that as the ENSO signal gets stronger (either El Nino or La Nina) the model becomes less accurate. This does not seem to apply the IOD as the order in which the IOD index strengthened is not the same as that of the ENSO index.

Visually, as could be expected, the runs of the low rainfall years tend to overestimate and the runs of high rainfall tend to overestimate (Fig 5.2). This suggests that WRF is not compensating for the change in mode. Years 2005 and 2006, which were the weakest in strength, seemed to reproduce the rainfall distribution accurately but overestimated and underestimated respectively (Fig 5.3). In 2002 and 2010 the model did not appear to accurately capture the rainfall distribution.

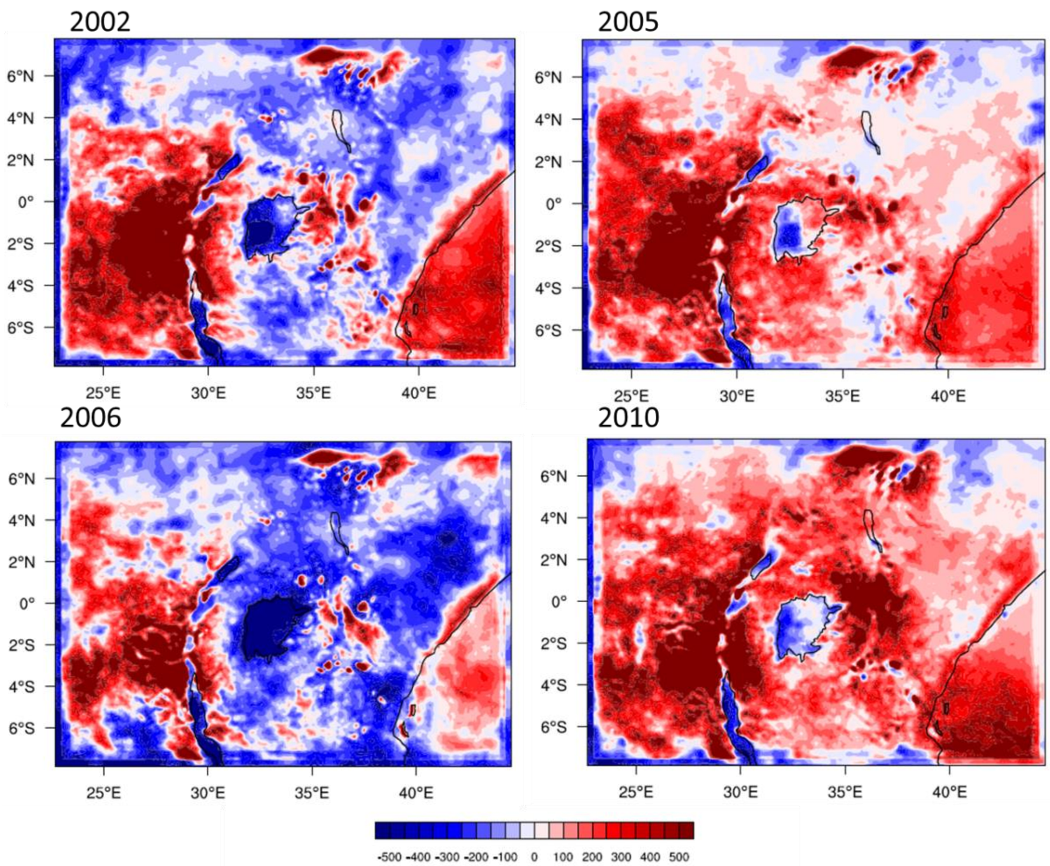


Figure 5.2: Difference between WRF and TRMM rainfall over OND in domain two.

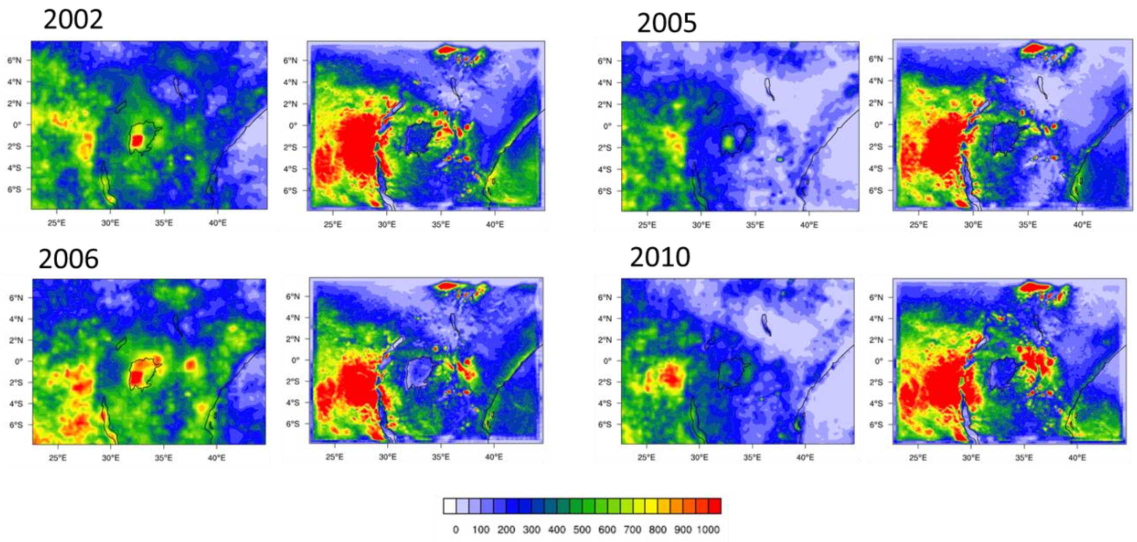


Figure 5.3: WRF rainfall on the right and TRMM rainfall on the left for each extreme year. Total rainfall over OND for domain two in mm.

Looking at the lake domain none of the runs are able to capture the correct asymmetrical rainfall pattern. There is an underestimation of the rainfall over the lake regardless of whether it is a wet or dry year. Even in the years where all of the surrounding land is highly overestimated the lake is still an area of underestimation. In 2005 and 2006 WRF does seem to reproduce the dry area on the western side of the lake and the wetter area on the eastern sides but there is no indication of the asymmetrical pattern over the lake itself.

5.5. Conclusion

In this chapter the hypothesis has been proven incorrect; the same combination cannot be relied upon to reproduce the results for extreme years to the same accuracy as the original climatology year. It appears that when the ENSO signal is strong the model cannot or does not account for this change and the accuracy of the model decreases. In the weaker years, it appears to remain relatively accurate.

It may be that the model needs another parameterisation combination for years of particularly strong ENSO mode that can be swapped in when necessary. This would be because the influences in the region are different when one of the large scale teleconnections is dominant.

The model is still unable to represent the pattern over the lake, and in the two most extreme runs it does not seem to be able to reproduce the rainfall distribution around the lake unlike the better runs.

6. Lake Mask

6.1. Introduction

It is known that the lake is very important to the circulation over the LVB, and that it is partially responsible for the asymmetrical rainfall pattern that occurs over the region (Asnani, 1993). It is also known, and has been explored further in this study, that the lake is badly represented and that this has the potential to incorrectly reproduce the rainfall over the region, in particular a lack of precipitation over the lake itself. The asymmetrical pattern that occurs over the lake is not represented in the models.

There is uncertainty in how much this problem with the lake extends into the region and whether this could produce a bias within the results with respect to the customisation. It is known that the lake should influence the rainfall over the surrounding area however the main area of error does appear to be over the lake itself. If the error over the lake is very high, it could be creating a large RMSE and yet the area surrounding the lake may actually be well represented. The aim of this chapter is to explore whether what appears to be the best and worst combinations of parameters with lake still represent the best and worst runs if the lake, and consequently the error, are taken out of the region. The following sections are ordered as previously.

6.2. Hypothesis

The inaccurately represented lake creates a bias in the results for the entire domain. This bias impacts which set of parameters are optimal for predicting rainfall in the region.

6.3. Methods

The runs for this section are the same as performed and analysed for the initial customisation in chapter 2. The runs and combinations in table 2.2 are used for this analysis. Similarly, the same statistical and visual analysis as has been previously described was also used.

The additional method used is that in every run all the statistics were also calculated with the lake masked and therefore that region of the output removed from the calculations. Each run was considered both with the lake included, as in every other chapter, and with the lake masked in order to remove any bias caused by the badly represented lake. The previous results, found in chapter 2, were used and then compared to the results found without the lake. A direct comparison shows whether or not the lake significantly biases the result and if the combinations have the same order of accuracy. The runs conducted for chapter three (Lake Surface Temperature chapter) were also considered. These are analysed without the lake and compared to the statistical results with the lake.

6.4. Results

Overall, with respect to the runs performed in the initialisation chapter, the hypothesis is disproved. Table 6.1 shows the order of combinations when the lake is masked, blue represents the combinations that have moved down and orange represents the combinations that moved up in the order. The runs that appear best and worst overall for the original rainfall results also appear to be best and worst when the lake is masked. This suggests that the lake does not create enough of a bias to significantly affect the results. Within the individual domains and grid spacing there were a few differences but these changes were minimal with just two or three runs being swapped in order.

Table 6.1: The order of accuracy according to each statistical test. Blue and orange represent the combinations that have changed with the lake masked, down and up respectively.

Root Mean Square Error					
Total	d01	d02	dLake	50km	10km
I	F	N	I	I	n
N	I	C	n	C	I
C	C	I	C	N	C
F	A	G	H	A	F
A	G	F	A	F	A
G	N	A	E	G	E
E	E	E	F	E	G
M	M	M	M	M	M
O	P	O	O	O	O
P	B	P	G	P	P
H	O	H	J	B	H
B	H	B	P	H	B
J	J	J	B	J	J

Standard Deviation					
Total	d01	d02	dLake	50km	10km
I	F	I	I	I	I
N	I	N	N	C	N
C	C	C	C	N	C
F	A	F	H	A	F
A	G	A	A	F	A
M	N	M	F	G	E
G	M	G	M	M	G
E	E	E	O	E	M
O	P	O	E	O	O
P	B	P	G	P	P
H	O	H	B	B	H
B	H	B	J	H	B
J	J	J	P	J	J

Mean Absolute Error					
Total	d01	d02	dLake	50km	10km
C	N	C	C	C	C
N	C	N	N	N	N
G	G	G	M	G	G
F	F	F	E	A	F
M	A	M	I	F	E
E	E	A	F	E	I
A	I	E	A	I	M
I	M	I	H	M	A
P	P	P	G	P	P
O	B	O	O	O	H
H	O	H	P	H	O
B	H	B	B	B	B
J	J	J	J	J	J

In all of the different statistics the lake domain showed the most changes. This would be expected as being focused directly on the basin means that any influence the lake has would impact this domain the most. In the case of the SD and MAE there were a few changes where runs performed better or worse and moved up or down the order in which the runs were ranked. The majority of the upper and lower quartile runs remained the same, but it highlights that the lake does hold an influence over the domain if only small. Therefore it is important to consider whether a particular feature may impact the results especially in a domain that primarily focuses on that feature.

In the lake temperature runs, the same parameterisation was used, so without the lake they should be comparable in error. The difference in RMSE between run O and OA before the lake is masked averages at around 30 mm with the lake mask this reduces to around 10mm (Table 6.2). This shows that the value of the error can be significantly changed by one large area of error, particularly the case in the RMSE. It is apparent however that the lake still causes enough bias to the region to make O significantly worse than OA. Masking out the lake does however give O better results than E and EA, showing that this huge area of error can bias how the region is represented. Finally, it changes the SD results, while in the RMSE and MAE OA still outperforms the other runs in the SD the other runs now show smaller error on average than OA does.

Table 6.2: Difference between the runs for each statistic.

RSME	Original	Masked Lake
OA-O	-31.00	-10.98
OA-E	-4.99	-9.22
OA-EA	-4.05	-3.66
Std Dev		
OA-O	-39.59	1.44
OA-E	2.01	11.78
OA-EA	-0.13	0.13
MAE		
OA-O	-8.51	-7.56
OA-E	-5.99	-10.28
OA-EA	-4.21	-4.03

It has to be considered whether, if knowing that there is a big area of error that cannot be accounted for, it makes more sense to customise without it to avoid any bias over the remainder of the region.

6.5. Conclusion

The results in this chapter do indicate that the lake has the potential to bias the region and affect the results of the overall customisation. While the masking of the lake in these customisation runs did not influence the final result, there was some change within the different domains. In particular in the domains that focused on the lake there were changes to the order and which runs outperform others. Had any changes occurred within the upper or lower quartile, it may have influenced which combinations were considered the overall best and worst.

As observed in the lake temperature chapter, there can be a large change in the regional error caused by an anomalously bad representation of one particular feature. This raises the question as to whether customisation should include areas that are known to be difficult to reproduce, or if the remainder of the domain should be focused on in order to customise as accurately as possible. In the case of this study, the lake is the large area of uncertainty, but there are potentially other regions around the globe that are modelled insufficiently and that continue to influence the regions surrounding them. While in this study the overall result was not impacted by the lake, it may be that for a smaller domain focused on the region of error, like the lake domain, it could have influenced the overall findings.

7. Temperature Comparison

7.1. Introduction

This study is concentrated on representing the regional rainfall and as such, that has been the major focus of all the preceding chapters. However, it is important to look at other variables, to make sure the model is producing reasonable atmospheric circulation. Results are not always correct for the right reason, so it is important to consider what other variables are doing at the same time. Additionally, due to the complexity of the atmosphere, many variables are connected, meaning that a different conclusion may be reached by considering other properties within the run. In this chapter, temperature is being considered in addition to rainfall.

Ideally the best run for temperature will also result in the best run for rainfall as that would confirm that the model is reproducing the correct results. As different parameters have different impacts on the run, this may not necessarily be the case however. This chapter investigates how the best combinations for rainfall compare to those for temperature. The order of sections is the same as previous chapters.

7.2. Hypothesis

The best customisation based on rainfall will also provide the best customisation for temperature.

7.3. Methods

Similar to chapter six the runs used for this chapter are also in table 2.2. Again the analysis was the same, but in this case the analysis was also run for the temperature. The same statistics and visual plots are produced but using the mean temperature over the time period and comparing it to the CRU data specified in chapter four.

This is in order to determine if the parameter combinations are also optimal for the temperature. Again these are analysed over multiple time periods and all domains to determine if the best rainfall combinations are also the best for temperature. The results from the temperature analysis are compared to that of the rainfall analysis in order to determine which runs are best for each and which combinations reproduce both variables with the greatest degree of accuracy. .

7.4. Results and Discussion

The results for temperature do differ from those for rainfall. Within the statistics RMSE and MAE agree, whereas SD gives slightly different results. In the RMSE and MAE the runs that come out with consistently the lowest error scores are P, N and A (Table 7.1). N is the only run that is consistently good for both the rainfall and the temperature (table 2.7 shows the comparable results for rainfall). The ones that consistently come out worse are I, C and F. Two of the runs that produced the best results in rainfall are now amongst the worst for temperature. In the SD the results are different from the other statistics and agree more with the rainfall results. The best combinations appear to be G, H and C while the worst are hard to define as when looking at all the differing domains and grid spacing no runs consistently underperform. The most common runs that are in the lower quartile are M and E.

Table 7.1. The order of accuracy for each statistic.

Root Mean Square Error					
Total	d01	d02	dLake	50km	10km
P	P	P	P	P	P
N	H	N	N	N	N
A	O	A	A	A	A
H	N	H	H	O	H
O	M	O	O	H	M
M	A	M	M	E	O
E	E	E	E	M	E
J	J	B	J	B	B
B	B	G	G	J	J
G	C	J	B	G	G
C	G	C	C	C	C
I	I	I	I	F	F
F	F	F	F	I	I

Standard Deviation					
Total	d01	d02	dLake	50km	10km
H	H	H	G	H	H
C	P	C	F	C	C
G	O	G	I	G	G
P	C	F	H	P	F
F	N	I	P	A	P
N	A	B	C	F	A
A	J	N	N	I	N
O	B	A	A	N	O
I	M	O	M	O	I
J	E	J	E	B	B
B	G	P	O	J	J
M	F	M	J	E	M
E	I	E	B	M	E

Mean Absolute Error					
Total	d01	d02	dLake	50km	10km
P	P	P	P	P	P
A	O	A	A	A	A
N	A	N	N	N	N
O	H	O	O	O	O
H	N	H	H	H	H
M	M	M	M	B	B
E	E	E	E	E	E
B	B	B	B	M	M
J	J	J	g	J	J
G	C	G	J	G	G
C	G	C	C	C	C
I	I	I	I	F	F
F	F	F	F	I	I

The differences between the errors from run to run are smaller than that of the rainfall. For example, the best runs are producing a RMSE of between 1.5° and 2.0°, yet run C which in many cases appeared to be one of the worst seemed to be closer to 2.5°. F and I seemed to be larger outliers as they did have bigger RMSE, F in particular showed errors of up to 11°. However, generally this seems a much smaller difference than that of the rainfall which saw the RMSE go from around 70mm for the best to over 100mm for one month's rainfall.

P used a different surface layer parameter which may account for a better result in temperature. The RUC scheme therefore performed better for temperature than Noah or 5-layer. There does not appear to be any consistency between the good runs in order to determine if certain parameter options are better than others. All of the worst runs contained CAM in the LW radiation option; suggesting that the CAM option may not be optimal for this region under these conditions. This is an unexpected finding, due to CAM being designed for longer runs.

The plots and the statistical bias show that nearly all the runs underestimate to a large degree over the continent; the average bias is around 2° (Fig 7.1). This may be an issue when customising in order to do future runs as 2 degree warming is the same order of magnitude as a potential climate warming, which would make the signal very hard to distinguish.

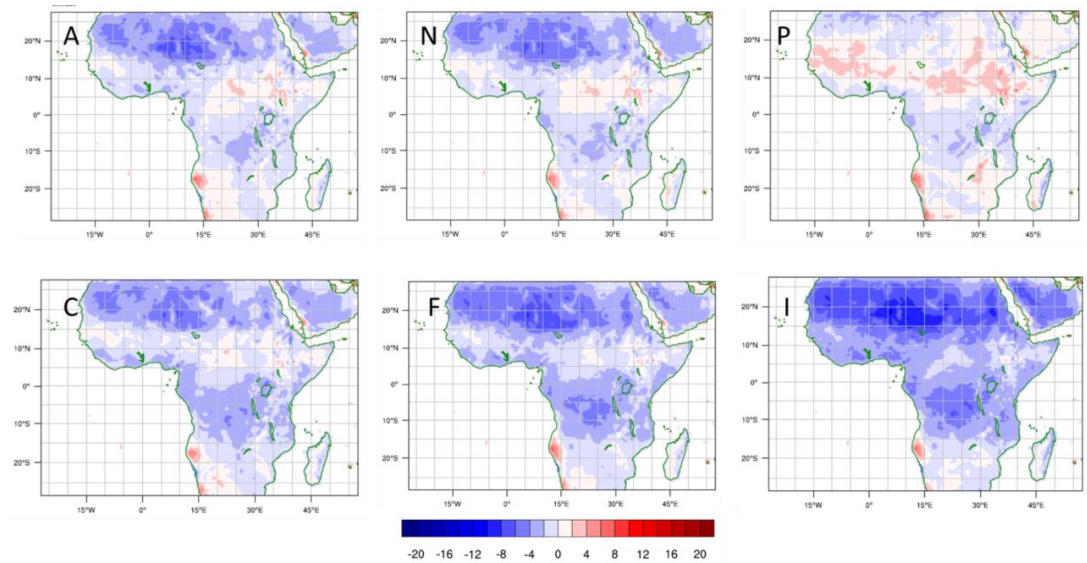


Figure 7.1: The difference between CRU and WRF for the mean temperature over November and December for the upper and lower quartile combinations.

The worst runs do show a substantial cold bias over the continent whereas the best runs tend exhibit regions of positive bias, although the overall pattern remains a negative one. In the lake domain, the high altitude areas seem to vary very quickly between over and underestimation of temperature presumably in areas of steep topography (Fig 7.2). This seems to occur in all the runs not just the worst ones.

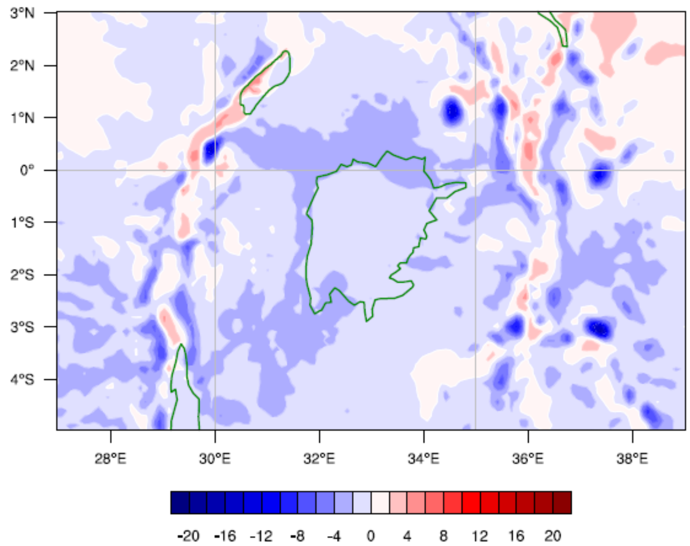


Figure 7.2: The difference between CRU and WRF for the mean temperature over November and December for run N.

7.5. Conclusion

In this case the stated hypothesis has been proven wrong. The best results for rainfall do not correspond to the best results for temperature. Only one of the runs seems to be optimal for both rainfall and temperature and that is run N. The others do not connect and in some cases what appears to be one of the best results for rainfall is one of the worst results for temperature. However it is also important to note that the error differences between the runs are relatively small, meaning that most combinations reproduce temperature to a similar level of accuracy.

It is shown that it is important to check not just the variable of interest but also analyse other variables in order to determine that the model is producing realistic results and that good results in one variable are not being produced at the expense of others. It is important to weigh up which combinations are the best to use as it is necessary to produce the best results for the variable of interest but it also needs to represent the rest of the system well. Therefore it may be necessary to balance the two. In this case run N would be the optimal solution as it seems to be one of the better runs in both cases.

8. Conclusions

Each chapter in this study has highlighted an important area of the customisation process. This is necessary for confirming that WRF is producing the most accurate results possible for a particular region, in this case the LVB. The individual chapters have all emphasised individual aspects of this crucial process but have also agreed and reinforced initial ideas and theories.

Some of the main components of the customisation process are the parameters and options available in WRF. Some schemes will always be better suited to particular runs. In this study, several of the schemes were stressed as either being suitable or unsuitable for this region. The scheme that stood out for producing optimal results is the BMJ cumulus scheme. All the runs produced with this scheme showed a much better representation over the Indian Ocean. The majority of runs produced a large overestimation over the ocean, whereas this scheme did not. The CAM radiation schemes (LW and SW) appeared in many of the better rainfall runs, however thy also appeared in the less accurate temperature runs so more investigation may be needed in this respect.

The schemes that did not appear to produce good results are the KF cumulus scheme and the Morrison Microphysics scheme. The KF scheme resulted in large amounts of precipitation and consequently an overestimation over a large proportion of the domain. The Morrison scheme appeared to cause some form of instability; some of the runs did not complete whereas others showed unrealistic results. This implied that neither of these schemes were suitable for the model run conducted here.

It was concluded that masking the lake did not have significant impact on which combinations produced the most accurate results. However it did produce minor modifications, especially within the lake domain. Conversely, the temperature comparison did cause the accuracy of the combinations to change. This emphasises that it is important to investigate more than one variable and not just the variable of primary focus. It is necessary to confirm that other elements of the circulation are also represented correctly. In the case of this study the best combination was that of run N as it produced good results for both rainfall and temperature. It also contained the BMJ scheme. This is the combination that would be recommended for use in future studies.

Finally, it can be concluded that the best combination for the climatology is not necessarily the best for the extreme years. It appears that it is necessary to perform additional customisation for these years. For this region, the ENSO mode appeared to have the largest impact on the accuracy of the runs and it would be important to adapt the customisation to support this changing modes.

The main emphasis of this study was on the asymmetrical pattern that occurs over the LVB and whether WRF was able to reproduce this pattern. The pattern is dry over the western shore, wet over the western lake, dry over the eastern lake and wet over the eastern shore.

The model was sometimes able to capture the asymmetry on each shore; however it was unable to reproduce the correct rainfall directly over the lake.

It is suspected that the main problem with this is that the model is unable to initialise the lake surface temperature correctly. WRF offers two options of initialising the lake; one, a direct interpolation of SST, create a high LST which produces too much rainfall over the lake. The second one produces a proxy for LST from air temperature and this gives a more accurate but much colder LST, resulting in too little rainfall over the lake surface.

This second initialisation causes too little rainfall regardless of the other parameters or variables within the run. This included positive ENSO years that caused large amounts of rainfall over the remainder of the region and yet the lake still produces very little. It is theorised that this lack of rainfall is due to the cold temperature preventing the reversal of the land-lake breeze. This breeze should reverse at night causing convergence and consequently rainfall over the lake. It does not, as the temperature is too low so the breeze continues to flow inland causing air to sink over the lake instead of rise and therefore producing no rainfall.

Overall it appears that a customisation study can improve the results over the general domain, but it does not improve the results directly over the lake. It will be necessary to correct the LST, potentially with a coupled model, before accurate rainfall can be produced.

Only then will it be possible to see if the model fully resolves the correct asymmetrical pattern over the Lake Victoria Basin.

9. Future work

More work is necessary to confirm some of the theories produced by this study. It would be of interest to do more analysis on the circulations that occur when the lake is misrepresented. Does the high temperature cause constant convection? And does the low temperature prevent the reversal of the land-lake breeze? Not only would this offer a more complete explanation of these two questions, but it may also offer an insight as to what would happen should there be any changes to the natural temperature gradient surrounding the lake.

It is known that the TRMM data contains errors and this needs to be taken into consideration and investigated in more detail. Additionally a greater set of observations would allow for better validation of the model and the rainfall it produces.

As mentioned previously a coupled lake model may enable more realistic rainfall amounts and patterns and therefore would be a vital addition to the model system. This would allow for further investigation and projections over the area.

Within this study it would be of interest to consider some of the schemes in more detail, such as CAM and Morrison. CAM appeared to perform as one of the best options for rainfall but worst for temperature, raising the question of why this occurs. It would also be useful to know why the Morrison scheme does not appear to be compatible with these runs. Additionally it would be of use to do further analysis on some of the aspects only mentioned here very briefly. This would include a more detailed study of the boundary conditions, the climate parameters and the land use. These were considered but all in very preliminary studies and may be of greater importance than anticipated.

It would also be of interest to separate the region further into different regions, such as mountainous regions, rainforest or plains. This could give an indication of which areas the model is able to reproduce most accurately and which other regions may also need more work to improve the results. Additionally, as mentioned in the introduction, finer grid spacing may indicate smaller circulations occurring around the coast that may be of further interest and worthy of future study. An investigation into the different scales of circulation in this region may give a better indication of what size of flow is influencing the lake in greater detail.

Finally it has been shown that extreme years need further customisation in order to incorporate the influence of large scale climate features on the region. The currently discussed feature is ENSO, however other features may represent different modes and may need additional customisation. It is important to identify these large-scale features and adapt the customisation to include their influence.

REFERENCES

Adeyewa, Z. D. and K. Nakamura, (2003). Validation of TRMM radar rainfall data over major climatic regions in Africa. *Journal of Applied Meteorology*, *42*(2).

Adler, R. F., G. J.Huffman, D. T. Bolvin, S. Curtis, and E. J. Nelkin, (2000). Tropical rainfall distributions determined using TRMM combined with other satellite and rain gauge information. *Journal of Applied Meteorology, 39*(12), 2007-2023.

Anyah, R. O., F. H. M. Semazzi, and L. Xie (2006), Simulated physical mechanisms associated with climate variability over Lake Victoria basin in East Africa, *Mon. Weather Rev., 134*(12), 3588-3609, doi: 10.1175/MWR3266.1.

Anyah, R. and F. Semazzi (2004), Simulation of the sensitivity of Lake Victoria basin climate to lake surface temperatures, *Theor. Appl. Climatol., 79*(1-2), 55-69, doi: 10.1007/s00704-004-0057-4.

Asnani, G. (1993), *Tropical Meteorology: 2005 Revision*, GC Asnani.

Behera, S., J. Luo, S. Masson, P. Delecluse, S. Gualdi, A. Navarra, and T. Yamagata (2005), Paramount impact of the Indian Ocean dipole on the East African short rains: A CGCM study, *J. Clim., 18*(21), 4514-4530, doi: 10.1175/JCLI3541.1.

Black, E., J. Slingo, and K. Sperber (2003), An observational study of the relationship between excessively strong short rains in coastal East Africa and Indian Ocean SST, *Mon. Weather Rev., 131*(1), 74-94, doi: 10.1175/1520-0493(2003)131<0074:AOSOTR>2.0.CO;2.

Bowden, J. H. (2008), Multi-scale climate change modeling study over the Greater Horn of Africa, North Carolina State University.

Bowden, J. H. and F. H. M. Semazzi (2007), Empirical analysis of intraseasonal climate variability over the greater horn of Africa, *J. Clim., 20*(23), 5715-5731, doi: 10.1175/2007JCLI1587.1.

Carvalho, D., A. Rocha, M. Gómez-Gesteira, and C. Santos (2012), A sensitivity study of the WRF model in wind simulation for an area of high wind energy, *Environmental Modelling & Software*, *33*, 23-34.

Christensen, J. H. et al. (2007), : Regional Climate Projections. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.

Clark, C., P. Webster, and J. Cole (2003), Interdecadal variability of the relationship between the Indian Ocean zonal mode and East African coastal rainfall anomalies, *J. Clim.*, *16*(3), 548-554, doi: 10.1175/1520-0442(2003)016<0548:IVOTRB>2.0.CO;2.

Crosman, E. T. and J. D. Horel (2010). Sea and lake breezes: a review of numerical studies. *Boundary-layer meteorology*, *137*(1), 1-29.

Davis, N., J. Bowden, F. Semazzi, L. Xie, and B. Onol (2009), Customization of RegCM3 Regional Climate Model for Eastern Africa and a Tropical Indian Ocean Domain, *J. Clim.*, *22*(13), 3595-3616, doi: 10.1175/2009JCLI2388.1.

East African Community (EAC) (2004), The vision and strategy framework for management and development of Lake Victoria basin. http://www.eac.int/index.php/lvbc-home/37-lvbc-content/43-lvbc.html.

East African Community (EAC) (2011), Enhancing Safety of Navigation and Efficient Exploitation of Natural Resources over Lake Victoria and its Basin by Strengthening Meteorological Services.

ECMWF (2014), ECMWF Data, Frequently Asked Questions, *2013*. http://www.ecmwf.int/products/data/archive/data_faq.html#ERAres

ECMWF (accessed 2013), ERA Interim, Daily Fields, *2013*. http://apps.ecmwf.int/datasets/data/interim_full_daily/

Eden, J. M., Widmann, M., D. Graweand S. Rast, (2012). Skill, Correction, and Downscaling of GCM-Simulated Precipitation. *Journal of Climate*, *25*(11).

Funk, C., M. D. Dettinger, J. C. Michaelsen, J. P. Verdin, M. E. Brown, M. Barlow, and A. Hoell (2008), Warming of the Indian Ocean threatens eastern and southern African food security but could be mitigated by agricultural development, *Proceedings of the National Academy of Sciences*, *105*(32), 11081-11086.

Giorgi, F., M. R. Marinucci, G. T. Bates, and G. Decanio (1993), Development of a 2nd-Generation Regional Climate Model (Regcm2) .2. Convective Processes and Assimilation of Lateral Boundary-Conditions, *Mon. Weather Rev.*, *121*(10), 2814-2832, doi: 10.1175/1520-0493(1993)121<2814:DOASGR>2.0.CO;2.

Goddard Space Flight Center (2013), TRMM - Tropical Rainfall Measuring Mission, *2011-2013*. http://pmm.nasa.gov/TRMM

Huffman, G. J. (1997). Estimates of root-mean-square random error for finite samples of estimated precipitation. *Journal of Applied Meteorology*, *36*(9), 1191-1201.

IPCC, (2013a): Summary for Policymakers. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker,T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)].Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

IPCC, (2013b): Annex I: Atlas of Global and Regional Climate Projections [van Oldenborgh, G.J., M. Collins, J. Arblaster, J.H. Christensen, J. Marotzke, S.B. Power, M. Rummukainen and T. Zhou (eds.)]. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

JAMSTEC (2012), Indian Ocean Dipole, *2013*. http://www.jamstec.go.jp/frsgc/research/d1/iod/e/iod/dipole_mode_index.html

Jimenez, P. A., J. Dudhia, J. F. Gonzalez-Rouco, J. P. Montavez, E. Garcia-Bustamante, J. Navarro, J. V. de Arellano, and A. Munoz-Roldan (2013), An evaluation of WRF's ability to reproduce the surface wind over complex terrain based on typical circulation patterns, *J. Geophys. Res. -Atmos.*, *118*(14), 7651-7669, doi: 10.1002/jgrd.50585.

Latif, M., D. Dommenget, M. Dima, and A. Grotzner (1999), The role of Indian Ocean sea surface temperature in forcing east African rainfall anomalies during December-January 1997/98, *J. Clim.*, *12*(12), 3497-3504, doi: 10.1175/1520-0442(1999)012<3497:TROIOS>2.0.CO;2.

Manatsa, D., C. H. Matarira, and G. Mukwada (2011), Relative impacts of ENSO and Indian Ocean dipole/zonal mode on east SADC rainfall, *Int. J. Climatol.*, *31*(4), 558-577, doi: 10.1002/joc.2086.

Marbaix, P., H. Gallee, O. Brasseur, and J. P. van Ypersele (2003), Lateral boundary conditions in regional climate models: A detailed study of the relaxation procedure, *Mon. Weather Rev.*, *131*(3), 461-479, doi: 10.1175/1520-0493(2003)131<0461:LBCIRC>2.0.CO;2.

Mehta, V. K., O. Aslam, L. Dale, N. Miller, and D. R. Purkey (2013), Scenario-based water resources planning for utilities in the Lake Victoria region, *Phys. Chem. Earth*, *61-62*, 22-31, doi: 10.1016/j.pce.2013.02.007.

Mukabana, J. R. and R. A. Pielke (1996), Investigating the influence of synoptic-scale monsoonal winds and mesoscale circulations on diurnal weather patterns over Kenya using a mesoscale numerical model, *Mon. Weather Rev., 124*(2), 224-243, doi: 10.1175/1520-0493(1996)124<0224:ITIOSS>2.0.CO;2.

NASA (2013), Mirador, data acess made simple - Product Summary and Product Description, *2011-2013*. http://mirador.gsfc.nasa.gov/collections/TRMM_3B42__007.shtml

National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce (2000), NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999, *2011-2013*. http://rda.ucar.edu/datasets/ds083.2

National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, and National Climatic Data Center/NESDIS/NOAA/U.S. Department of Commerce. (1986, updated weekly.), NCEP Version 2.0 OI Global SST and NCDC Version 3.0 Extended Reconstructed SST Analyses., *2011-2013*. http://rda.ucar.edu/datasets/ds277.0.

Nicholson, S. (1996), A review of climate dynamics and climate variability in eastern Africa, *The limnology, climatology and paleoclimatology of the East African lakes*, 25-56.

Nicholson, S. E. and X. Yin (2002), Mesoscale patterns of rainfall, cloudiness and evaporation over the Great Lakes of East Africa, *The East African Great Lakes: Limnology, Palaeolimnology and Biodiversity, 12*, 93-119.

Nicholson, S. and E. Kim (1997), The relationship of the El Nino Southern oscillation to African rainfall, *Int. J. Climatol., 17*(2), 117-135, doi: 10.1002/(SICI)1097-0088(199702)17:2<117::AID-JOC84>3.0.CO;2-O.

Nile Basin Initiative (2012), State of the River Nile Basin Report. http://nileis.nilebasin.org/content/state-river-nile-basin-report;

NOAA/ National Weather Service (2014), Cold and Warm Episodes by Season., *2013*. http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml

Ntiba, M., W. Kudoja, and C. Mukasa (2001), Management issues in the Lake Victoria watershed, *Lakes & Reservoirs: Research & Management*, *6*(3), 211-216.

Osuri, K. K., U. C. Mohanty, A. Routray, M. A. Kulkarni, and M. Mohapatra (2012), Customization of WRF-ARW model with physical parameterization schemes for the simulation of tropical cyclones over North Indian Ocean, *Nat. Hazards*, *63*(3), 1337-1359, doi: 10.1007/s11069-011-9862-0.

Paeth, H. and A. Hense (2006), On the linear response of tropical African climate to SST changes deduced from regional climate model simulations, *Theoretical and applied climatology*, *83*(1-4), 1-19.

Pielke, R. A. (2002), *Mesoscale Meteorological Modeling*, Academic press.

Pohl, B., J. Crétat, and P. Camberlin (2011), Testing WRF capability in simulating the atmospheric water cycle over Equatorial East Africa, *Clim. Dyn.*, *37*(7), 1357-1379, doi: 10.1007/s00382-011-1024-2.

Rahman, M. A. (2013), Water Security: Ethiopia–Egypt Transboundary Challenges over the Nile River Basin, *Journal of Asian and African Studies*, *48*(1), 35-46.

Rao, Y. V. R., L. Alves, B. Seulall, Z. Mitchell, K. Samaroo, and G. Cummings (2012), Evaluation of the weather research and forecasting (WRF) model over Guyana, *Nat. Hazards*, *61*(3), 1243-1261, doi: 10.1007/s11069-011-9977-3.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang (2002), An improved in situ and satellite SST analysis for climate, *J. Clim.*, *15*(13), 1609-1625.

Saji, N., B. Goswami, P. Vinayachandran, and T. Yamagata (1999), A dipole mode in the tropical Indian Ocean, *Nature*, *401*(6751), 360-363, doi: 10.1038/43855.

Saji, N. and T. Yamagata (2003a), Possible impacts of Indian Ocean Dipole mode events on global climate, *Clim. Res.*, *25*(2), 151-169, doi: 10.3354/cr025151.

Saji, N. and T. Yamagata (2003b), Structure of SST and surface wind variability during Indian ocean dipole mode events: COADS observations, *J. Clim.*, *16*(16), 2735-2751, doi: 10.1175/1520-0442(2003)016<2735:SOSASW>2.0.CO;2.

Schreck, C. and F. Semazzi (2004), Variability of the recent climate of eastern Africa, *Int. J. Climatol.*, *24*(6), 681-701, doi: 10.1002/joc.1019.

Seto, K. C., B. Güneralp, and L. R. Hutyra (2012), Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools, *Proceedings of the National Academy of Sciences*, *109*(40), 16083-16088.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X. Huang, W. Wang, and J. G. Powers (2008), A Description of the Advanced Research WRF Version 3.

Song, Y., F. H. Semazzi, and L. Xie (2002), Development of a coupled regional climate simulation model for the Lake Victoria basin, *The East African Great Lakes: Limnology, Palaeolimnology and Biodiversity: Dordrecht, Kluwer Academic Publishers*, 189-208.

Song, Y., F. Semazzi, L. Xie, and L. Ogallo (2004), A coupled regional climate model for the Lake Victoria basin of East Africa, *Int. J. Climatol., 24*(1), 57-75, doi: 10.1002/joc.983.

Sun, L. Q., F. H. M. Semazzi, F. Giorgi, and L. Ogallo (1999), Application of the NCAR regional climate model to eastern Africa - 1. Simulation of the short rains of 1988, *Journal of Geophysical Research-Atmospheres, 104*(D6), 6529-6548, doi: 10.1029/1998JD200051.

Sutcliffe, J. V. and Y. P. Parks (1999), *The Hydrology of the Nile*, International Association of Hydrological Sciences Wallingford, Oxfordshire, UK.

Sutcliffe, J. V. and G. Petersen (2007), Lake Victoria: derivation of a corrected natural water level series, *Hydrol. Sci. J. -J. Sci. Hydrol., 52*(6), 1316-1321, doi: 10.1623/hysj.52.6.1316.

Tierney, J. E., J. E. Smerdon, K. J. Anchukaitis, and R. Seager (2013), Multidecadal variability in East African hydroclimate controlled by the Indian Ocean, *Nature, 493*(7432), 389-392.

UCAR (2013), The Weather Research and Forecasting Model. http://wrf-model.org/index.php

University of East Anglia Climatic Research Unit (CRU). (2008), CRU Time Series (TS) high resolution gridded datasets, (Internet). NCAS British Atmospheric Data Centre, *2013*. http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__dataent_1256223773328276

Wang, W., C. Bruyère, M. Duda, J. Dudhia, D. Gill, H. Lin, J. Michaelakes, S. Rizvi, and X. Zhang (2011), WRF-ARW Version 3 Modeling System User's Guide, *Mesoscale & Microscale Meteorology Division, National Center for Atmospheric Research, Boulder, USA*.

Williams, A. P. and C. Funk (2011), A westward extension of the warm pool leads to a westward extension of the Walker circulation, drying eastern Africa, *Clim. Dyn., 37*(11-12), 2417-2435, doi: 10.1007/s00382-010-0984-y.

Willmott, C. J. and K. Matsuura (2005), Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, *Climate Research, 30*(1), 79.

World Bank (2008), *Africa - Second Phase of Lake Victoria Environmental Management Project*. http://documents.worldbank.org/curated/en/2008/09/9853300/africa-second-phase-lake-victoria-environmental-management-project;

Zhao, L., J. Jin, S. Wang, and M. B. Ek (2012), Integration of remote-sensing data with WRF to improve lake-effect precipitation simulations over the Great Lakes region, *J. Geophys. Res. -Atmos.*, *117*, D09102, doi: 10.1029/2011JD016979.

APPENDIX

Appendix 1

Table A1: Preliminary Customisation Results

RMSE	Domain 1			Domain 2		
run	total	Day 1	Day 5	total	Day 1	Day 5
a	42.30	11.08	17.77	51.72	13.01	23.64
b	54.90	13.78	21.61	54.49	16.20	24.35
c	43.80	11.02	17.89	51.29	12.98	24.11
d	44.96	11.49	18.41	51.90	13.17	23.76
e	42.34	11.49	18.27	53.10	13.96	24.06
f	55.23	12.58	20.87	55.46	15.55	23.12
g	41.84	10.16	17.38	48.48	11.62	23.16
h	54.30	12.06	20.50	54.96	14.46	23.19
i	51.46	11.80	19.87	52.71	14.02	23.30
j	48.36	11.45	19.03	50.99	13.97	23.09
k	43.15	11.09	17.75	53.02	12.85	24.08
l	41.78	11.02	17.56	51.23	13.01	23.00
m	43.84	11.46	17.87	50.11	13.54	23.66

SD	Domain 1			Domain 2		
run	total	Day 1	Day 5	total	Day 1	Day 5
a	42.11	10.65	17.61	48.74	12.71	20.52
b	53.00	12.94	21.60	52.36	15.55	21.13
c	43.80	10.69	17.45	44.11	12.82	18.85
d	44.49	11.01	18.34	49.04	12.87	20.59
e	42.30	11.02	17.76	46.35	13.62	18.62
f	51.20	11.74	20.67	55.46	14.68	22.12
g	41.50	9.85	17.32	46.58	11.42	21.46
h	50.58	11.35	20.32	54.94	13.80	22.01
i	48.55	11.13	19.77	52.52	13.43	21.91
j	45.59	10.76	18.97	50.90	13.31	21.99
k	43.12	10.77	17.45	48.17	12.63	19.56
l	41.44	10.58	17.51	49.40	12.66	20.51
m	43.67	10.84	17.67	47.54	13.03	20.61

Appendix 2

Table A2: Skill Score for Customisation, figure in green represent skill.

Skill Score One

	d01 original			d01-d02			d02 original			d02 to d01			Lake d02			Lake d01 to d01		
RSME	Nov	Dec	ND	Nov	Dec	ND	Nov	Dec	ND	Nov	Dec	ND	Nov	Dec	ND	Nov	Dec	ND
A	70.68	78.09	126.90	69.53	76.25	124.41	127.27	110.73	222.72	125.44	109.20	220.16	163.31	128.84	277.29	155.45	121.92	262.03
B	91.25	104.78	170.42	90.05	102.91	167.97	154.02	163.38	296.97	155.42	164.32	300.13	194.70	174.91	346.44	189.92	172.70	340.31
C	84.32	69.62	132.37	76.07	62.99	117.45	106.38	80.48	162.27	106.17	75.83	157.13	162.41	119.18	256.29	156.54	105.96	236.77
E	77.40	96.12	141.25	74.32	92.64	136.17	126.52	99.43	203.13	125.50	99.51	202.93	184.39	110.13	271.52	181.97	107.88	267.80
F	66.53	69.12	114.54	64.60	66.52	110.74	117.94	103.62	201.80	116.06	104.48	201.48	166.30	115.87	262.91	164.02	115.25	260.61
G	87.05	74.83	139.18	78.71	67.95	124.05	127.22	87.86	194.42	127.83	85.15	193.10	198.83	126.62	308.48	193.41	118.06	294.77
H	113.91	122.90	199.52	110.09	118.80	193.88	148.20	154.14	270.48	149.48	159.51	277.51	159.46	116.17	259.57	157.16	111.90	253.59
I	72.22	76.08	130.03	70.66	73.81	127.00	114.93	90.73	187.96	112.27	89.86	185.25	138.91	103.84	223.17	138.73	104.97	224.44
J	91.45	107.93	172.65	90.22	105.84	170.12	159.61	167.78	308.05	160.17	168.63	309.98	189.40	191.87	358.50	182.32	186.63	346.18
L	172.14	0.00	0.00	170.50	0.00	0.00	173.99	0.00	0.00	177.65	0.00	0.00	138.28	0.00	0.00	132.07	0.00	0.00
M	75.18	88.94	133.81	71.96	84.96	128.30	130.04	104.16	212.37	129.72	104.97	213.24	172.39	112.77	259.43	170.83	109.89	255.79
N	82.73	80.66	134.79	73.86	74.00	119.27	105.74	78.54	160.12	105.94	76.70	158.88	153.36	103.61	235.19	149.53	98.83	226.95
O	92.14	105.18	179.51	90.31	103.41	176.36	142.47	113.72	240.05	142.96	115.47	242.70	185.09	125.24	293.98	176.65	119.42	282.62
P	80.31	85.75	135.62	77.41	81.86	130.39	137.83	130.91	250.10	138.06	133.45	253.85	192.27	156.43	331.03	188.70	153.11	324.70

Skill Score Two

	d01 original			d01-d02			d02 original			d02 to d01			Lake d02			Lake d01 to d01		
SD WRF	Nov	Dec	ND	Nov	Dec	ND	Nov	Dec	ND	Nov	Dec	ND	Nov	Dec	ND	Nov	Dec	ND
A	70.50	76.37	125.09	69.34	74.47	122.53	123.19	104.94	212.31	120.24	102.08	207.14	161.53	125.73	272.33	153.44	118.67	256.61
B	90.18	101.49	165.66	88.92	99.49	163.02	142.65	143.42	263.82	142.19	142.12	262.64	186.74	154.73	318.35	180.99	150.48	309.07
C	84.30	69.39	132.31	76.02	62.79	117.43	104.29	80.41	160.44	104.30	75.80	155.52	161.88	118.81	256.27	156.16	105.89	236.90
E	77.40	95.14	140.59	74.32	91.53	135.37	126.42	99.38	203.12	125.53	99.17	202.75	184.27	109.14	270.66	182.07	107.00	267.33
F	66.31	69.01	114.54	64.39	66.37	110.74	117.79	102.33	200.56	115.16	101.80	197.85	165.85	115.63	262.17	163.64	115.16	260.08
G	87.05	74.45	139.04	78.67	67.60	123.97	127.20	87.78	194.42	127.86	85.08	193.14	197.87	124.73	305.70	192.66	116.41	292.42
H	112.76	120.22	195.13	108.83	115.87	189.08	148.19	150.70	268.27	149.27	154.82	273.54	153.41	115.51	253.48	152.17	111.49	248.73
I	70.07	75.50	127.22	68.51	73.26	124.24	102.87	84.12	168.14	102.72	84.65	169.17	117.02	88.07	181.51	117.89	89.23	184.02
J	90.35	104.55	167.74	89.07	102.32	164.99	147.99	145.02	272.23	146.85	143.31	269.66	182.72	165.40	326.92	175.00	157.30	311.14
L	165.76			163.81			149.55			147.95			119.53			112.05		
M	75.18	88.00	133.15	71.95	83.89	127.51	129.97	103.89	212.03	129.25	104.13	211.83	172.23	112.51	258.99	170.91	109.69	255.67
N	82.73	79.63	134.19	73.85	72.98	118.75	98.03	76.84	150.04	98.25	74.89	148.58	143.14	100.53	221.13	139.51	95.15	212.13
O	90.76	101.83	174.46	88.82	99.81	170.94	137.90	109.74	230.91	134.60	108.50	226.30	182.59	123.31	289.31	176.63	117.32	277.17
P	80.11	84.28	133.91	77.17	80.20	128.44	131.06	122.01	233.30	127.43	121.03	229.08	187.88	150.00	319.87	183.99	146.61	313.05

Appendix 3

Table A3: Statistical results for the LST Chapter, Red shows an increase in error and green shows a decrease.

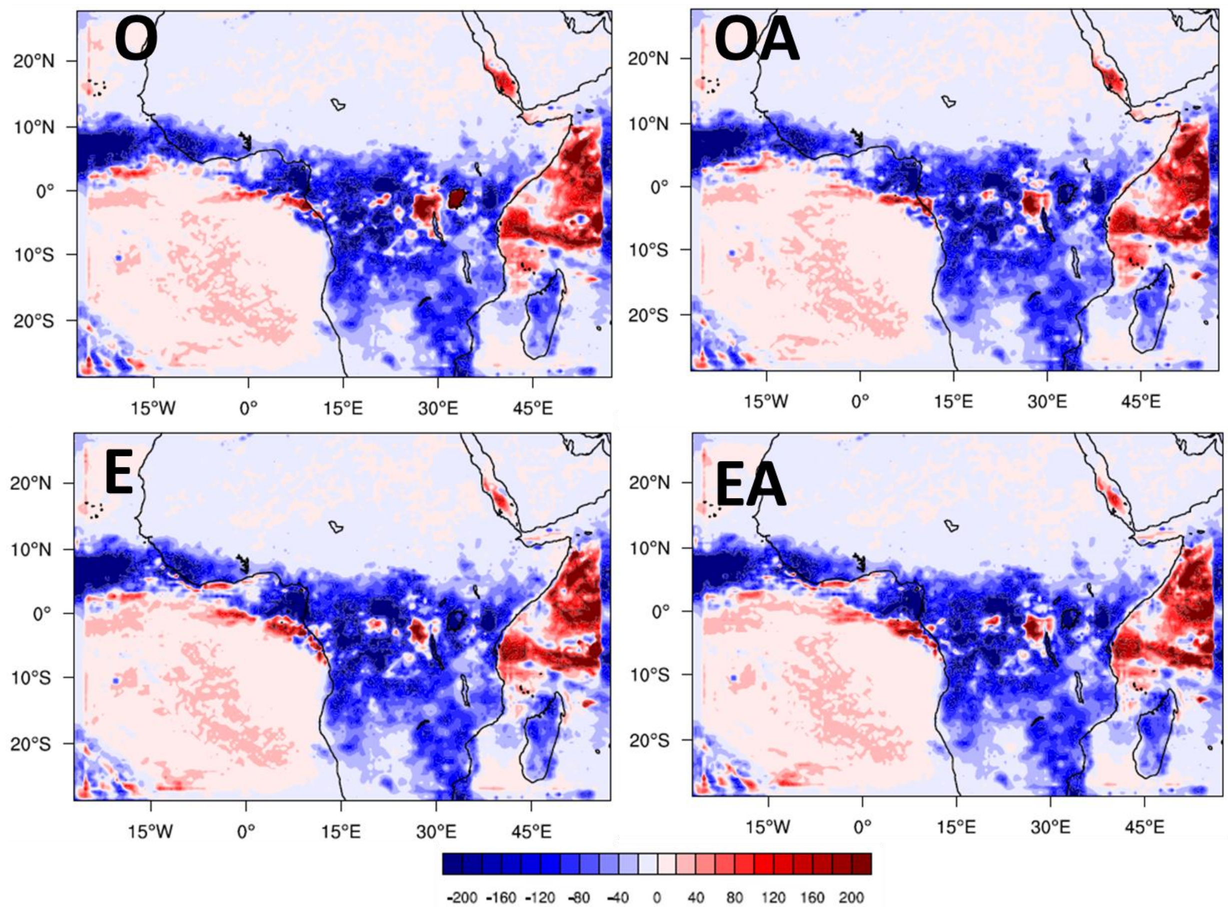
RSME	d01 original		d01 to d02		d02 original		d02 to d01		Lake d02		Lake d01 to d01		Average Difference
	Nov	Dec	Nov	Dec	Nov	Dec	Nov	Dec	Nov	Dec	Nov	Dec	
O	75.174	77.5721	73.2046	75.0006	136.422	118.548	135.223	116.905	204.027	177.143	198.141	168.093	
OA	72.682	76.6107	71.1891	74.1759	111.758	91.4153	108.865	91.2474	137.077	105.539	136.834	106.07	
OA-O	-2.492	-0.9614	-2.0155	-0.8247	-24.664	-27.1327	-26.358	-25.6576	-66.95	-71.604	-61.307	-62.023	**-30.9992**
E	74.9556	85.9265	73.4239	83.3571	116.29	95.0573	114.127	93.9616	141.62	112.029	140.736	111.876	
OA-E	-2.2736	-9.3158	-2.2348	-9.1812	-4.532	-3.642	-5.262	-2.7142	-4.543	-6.49	-3.902	-5.806	**-4.9914**
EA	75.6207	79.6224	73.9148	77.3288	117.359	94.6651	115.299	94.1921	141.168	111.575	139.507	111.861	
OA-EA	-2.9387	-3.0117	-2.7257	-3.1529	-5.601	-3.2498	-6.434	-2.9447	-4.091	-6.036	-2.673	-5.791	**-4.0541**

Std Dev	d01 original		d01 to d02		d02 original		d02 to d01		Lake d02		Lake d01 to d01		Average Difference
	Nov	Dec	Nov	Dec	Nov	Dec	Nov	Dec	Nov	Dec	Nov	Dec	
O	73.3952	77.1518	71.4244	74.6064	132.864	117.634	132.785	116.481	203.438	177.07	197.634	168.248	
OA	70.5924	76.1187	69.1053	73.7117	99.7461	84.9142	98.5386	86.1045	115.562	88.3478	116.249	88.7053	
OA-O	-2.8028	-1.0331	-2.3191	-0.8947	-33.1179	-32.7198	-34.2464	-30.3765	-87.876	-88.7222	-81.385	-79.5427	**-39.5864**
E	72.4925	85.3196	70.958	82.7768	98.472	82.7795	98.093	83.0004	102.625	82.4543	102.854	81.6947	
OA-E	-1.9001	-9.2009	-1.8527	-9.0651	1.2741	2.1347	0.4456	3.1041	12.937	5.8935	13.395	7.0106	**2.0147**
EA	73.3133	78.8102	71.6024	76.5451	102.381	83.841	102.14	84.7998	111.053	87.3645	110.142	87.3168	
OA-EA	-2.7209	-2.6915	-2.4971	-2.8334	-2.6349	1.0732	-3.6014	1.3047	4.509	0.9833	6.107	1.3885	**-0.1345**

MAE	d01 original		d01 to d02		d02 original		d02 to d01		Lake d02		Lake d01 to d01		
	Nov	Dec	Nov	Dec	Nov	Dec	Nov	Dec	Nov	Dec	Nov	Dec	
O	37.3326	39.8043	36.86	38.9747	87.2257	66.6793	86.7603	66.2741	129.825	92.8691	127.215	91.0311	
OA	37.1385	39.6286	36.728	38.8729	81.2773	60.9092	80.07	60.6722	106.383	75.404	105.273	76.3688	
OA-O	-0.1941	-0.1757	-0.132	-0.1018	-5.9484	-5.7701	-6.6903	-5.6019	-23.442	-17.4651	-21.942	-14.6623	**-8.5105**
E	38.7962	43.7663	38.2602	42.96	87.5987	65.6594	87.0742	65.2877	114.949	85.9261	113.482	86.7904	
OA-E	-1.6577	-4.1377	-1.5322	-4.0871	-6.3214	-4.7502	-7.0042	-4.6155	-8.566	-10.5221	-8.209	-10.4216	**-5.9854**
EA	38.7626	42.0476	38.2139	41.205	87.1232	64.21	86.6329	63.845	111.651	82.332	109.936	83.2623	
OA-EA	-1.6241	-2.419	-1.4859	-2.3321	-5.8459	-3.3008	-6.5629	-3.1728	-5.268	-6.928	-4.663	-6.8935	**-4.2080**

Appendix 4

Figure A1: Difference in rainfall between WRF and TRMM data over November in mm.



Appendix 5

Table A4: RMSE and MAE for best chapter, green represents the lowest error, fading toe red for the worst.

RSME	d01 original					d01 to d02				
	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
Abest	150.331	70.8587	65.8446	69.1695	114.297	143.924	66.845	63.4689	66.1229	110.146
Cbest	173.209	71.6942	72.2015	79.9314	134.191	166.944	67.7888	70.2359	77.5452	130.451
Bbest	183.744	85.7091	88.508	74.6061	139.665	162.727	79.391	79.7873	67.3823	124.136
Worst	200.842	77.3151	88.9299	102.743	167.722	196.68	75.2158	87.166	100.423	164.712
Yworst	176.447	86.4489	74.5005	86.8596	136.848	169.65	82.0267	71.6494	83.701	132.637
Xworst	403.711	142.454	142.568	202.772	311.972	403.149	141.452	141.805	201.905	311.383
Zworst	400.56	156.45	134.83	188.858	293.658	398.845	154.497	133.587	187.34	292.29

RSME	d02 original					d02 to d01				
	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
Abest	290.917	134.012	119.653	103.367	203.91	281.498	118.8	118.454	101.88	202.158
Cbest	277.473	127.884	111.919	104.747	193.292	260.622	110.69	106.698	103.749	187.441
Bbest	264.895	117.579	131.02	83.1201	191.069	260.725	110.764	131.757	80.1043	189.381
Worst	392.404	132.094	154.194	164.834	296.691	392.34	128.588	153.86	164.554	296.867
Yworst	308.967	153.228	113.065	114.639	207.394	296.024	138.084	110.219	113.865	205.342
Xworst	489.146	175.693	183.532	203.699	357.802	482.566	173.861	182.859	197.573	352.661
Zworst	483.916	205.333	162.521	181.898	322.124	493.505	208.371	163.545	187.759	329.025

RSME	Lake d02					Lake d02 to d01				
	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
Abest	367.803	132.107	165.694	120.97	268.397	358.407	117.412	164.03	118.83	266.032
Cbest	373.846	134.037	163.96	130.577	275.178	351.58	116.597	157.019	124.521	263.521
Bbest	393.701	140.885	200.285	114.962	292.826	377.365	129.27	195.422	107.889	281.648
Worst	454.787	127.489	194.254	187.699	351.452	438.563	117.198	186.93	184.254	341.314
Yworst	342.215	163.271	133.748	109.37	224.747	316.878	138.82	130.051	107.64	219.997
Xworst	401.69	142.348	168.01	135.39	290.789	400.027	140.357	166.807	138.079	291.7
Zworst	410.756	145.085	174.238	135.607	298.594	409.754	142.835	173.72	138.538	300.708

MAE	d01 original					d01 to d02				
	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
Abest	88.3542	36.6058	33.5118	36.8534	61.4995	86.1999	35.8217	32.5259	35.6294	59.7401
Cbest	106.896	39.9027	39.2623	45.6125	76.8135	104.869	39.3123	38.4249	44.5898	75.2371
Bbest	86.952	38.0067	35.9927	33.4503	59.5321	83.0525	36.8589	34.318	31.5561	56.4403
Worst	113.114	38.0071	44.3609	51.075	88.8907	111.808	37.3957	43.7593	50.3495	87.9558
Yworst	101.612	41.1505	38.017	44.0297	73.4789	99.4846	40.2195	37.0685	42.886	71.9702
Xworst	219.922	73.813	70.0978	91.7379	155.569	220.523	73.952	70.0452	91.8593	155.883
Zworst	215.84	72.2104	66.6155	86.7371	149.594	216.127	72.395	66.5196	86.763	149.669

MAE	d02 original					d02 to d01				
	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
Abest	188.845	71.2952	74.8335	64.3264	128.039	184.265	67.635	73.8472	64.1543	127.238
Cbest	185.393	72.4864	73.991	70.3133	129.808	177.628	66.7754	71.4749	71.4571	127.706
Bbest	155.247	65.9812	74.053	47.6002	103.712	154.449	65.0484	75.0133	46.8699	103.795
Worst	267.209	78.7473	106.377	111.514	201.118	269.977	78.8938	107.749	112.917	203.352
Yworst	208.741	86.6042	74.0726	75.2074	137.317	204.062	83.6862	72.5285	74.7782	136.012
Xworst	366.779	136.899	130.315	115.514	242.302	363.498	135.234	130.452	113.557	240.558
Zworst	374.332	148.957	122.521	110.674	231.772	382.348	151.126	124.87	114.197	237.858

MAE	Lake d02					Lake d02 to d01				
	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
Abest	236.376	76.0467	110.6	73.507	172.487	235.512	72.6771	109.974	75.1194	175.247
Cbest	247.623	78.4872	114.644	89.3759	187.633	237.604	72.9076	110.819	87.5644	182.709
Bbest	244.222	80.3091	124.527	69.7354	175.119	238.053	77.2072	122.369	68.4379	170.832
Worst	326.047	81.7602	143.52	142.99	252.745	317.428	78.7077	139.114	142.423	247.45
Yworst	230.78	92.6783	92.3822	69.5886	152.166	223.792	87.1593	89.8535	69.3037	150.286
Xworst	344.808	101.646	140.357	106.94	246.599	344.052	99.1845	139.209	109.627	247.686
Zworst	358.03	105.171	149.941	106.636	256.417	357.675	102.404	149.26	109.401	258.467

Appendix 6

Table A5: Skill scores for Best runs.

RSME			d01 original					d01 to d02		
	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
best										
Abest	150.331	70.8587	65.8446	69.1695	114.297	143.924	66.845	63.4689	66.1229	110.146
Bbest	183.744	85.7091	88.508	74.6061	139.665	162.727	79.391	79.7873	67.3823	124.136
Cbest	173.209	71.6942	72.2015	79.9314	134.191	166.944	67.7888	70.2359	77.5452	130.451
worst										
Wwosrt	200.842	77.3151	88.9299	102.743	167.722	196.68	75.2158	87.166	100.423	164.712
Xworst	403.711	142.454	142.568	202.772	311.972	403.149	141.452	141.805	201.905	311.383
Yworst	176.447	86.4489	74.5005	86.8596	136.848	169.65	82.0267	71.6494	83.701	132.637
Zworst	400.56	156.45	134.83	188.858	293.658	398.845	154.497	133.587	187.34	292.29

RSME			d02 original					d02 to d01		
	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
best										
Abest	290.917	134.012	119.653	103.367	203.91	281.498	118.8	118.454	101.88	202.158
Bbest	264.895	117.579	131.02	83.1201	191.069	260.725	110.764	131.757	80.1043	189.381
Cbest	277.473	127.884	111.919	104.747	193.292	260.622	110.69	106.698	103.749	187.441
worst										
Wwosrt	392.404	132.094	154.194	164.834	296.691	392.34	128.588	153.86	164.554	296.867
Xworst	489.146	175.693	183.532	203.699	357.802	482.566	173.861	182.859	197.573	352.661
Yworst	308.967	153.228	113.065	114.639	207.394	296.024	138.084	110.219	113.865	205.342
Zworst	483.916	205.333	162.521	181.898	322.124	493.505	208.371	163.545	187.759	329.025

RSME			Lake d02					Lake d01 to d01		
	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
best										
Abest	367.803	132.107	165.694	120.97	268.397	358.407	117.412	164.03	118.83	266.032
Bbest	393.701	140.885	200.285	114.962	292.826	377.365	129.27	195.422	107.889	281.648
Cbest	373.846	134.037	163.96	130.577	275.178	351.58	116.597	157.019	124.521	263.521
worst										
Wwosrt	454.787	127.489	194.254	187.699	351.452	438.563	117.198	186.93	184.254	341.314
Xworst	401.69	142.348	168.01	135.39	290.789	400.027	140.357	166.807	138.079	291.7
Yworst	342.215	163.271	133.748	109.37	224.747	316.878	138.82	130.051	107.64	219.997
Zworst	410.756	145.085	174.238	135.607	298.594	409.754	142.835	173.72	138.538	300.708

SD-wrf			d01 original					d01 to d02		
	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
best										
Abest	204.218	88.7921	75.8073	86.406	149.275	204.218	85.9411	73.9331	84.4383	146.789
Bbest	277.025	111.966	111.833	107.633	201.627	277.025	106.088	103.235	101.526	188.118
Cbest	221.307	84.1146	82.3383	93.0864	165.719	221.307	81.0714	80.7681	91.444	163.311
worst										
Wwosrt	286.497	110.175	115.309	128.844	223.691	286.497	108.854	113.847	126.983	221.369
Xworst	404.982	131.636	139.231	206.72	316.416	404.982	130.192	138.301	205.638	315.404
Yworst	262.997	118.297	97.2884	108.931	191.044	262.997	115.512	95.3808	106.73	188.619
Zworst	389.401	141.911	126.961	188.183	288.815	389.401	139.68	125.838	186.735	287.669

SD-wrf			d02 original					d02 to d01		
	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
best										
Abest	362.623	169.29	144.476	120.562	256.329	357.909	159.146	146.18	120.157	257.942
Bbest	360.298	157.48	164.349	108.48	263.534	357.46	152.293	164.739	106.942	262.875
Cbest	335.353	156.649	125.378	117.017	234.251	321.205	144.638	122.485	115.95	230.402
worst										
Wwosrt	443.121	174.234	182.585	171.214	338.732	438.679	171.5	179.971	169.672	335.359
Xworst	359.581	101.903	127.48	171.984	277.562	350.377	99.4172	127.02	163.739	269.847
Yworst	356.708	184.279	127.242	118.864	239.798	345.179	172.694	126.277	118.035	238.608
Zworst	347.711	143.56	87.9533	145.765	225.353	359.779	147.36	90.6587	152.167	234.411

SD-wrf			Lake d02					Lake d01 to d01		
	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
best										
Abest	435.353	172.52	180.441	141.027	310.954	429.778	161.206	180.443	138.56	309.631
Bbest	461.745	183.276	217.563	138.516	346.417	446.585	173.01	212.783	131.774	335.421
Cbest	411.087	161.556	166.191	140.118	298.076	393.929	147.675	160.095	134.809	288.036
worst										
Wwosrt	459.053	162.287	194.046	168.94	349.956	441.993	154.72	186.824	163.619	337.975
Xworst	51.7576	19.284	29.9667	13.7714	40.501	52.3295	17.5094	32.6161	12.5444	42.5414
Yworst	399.098	193.597	139.049	123.811	257.174	376.681	174.127	134.787	119.91	249.824
Zworst	56.7981	24.1864	24.3057	14.7742	37.1682	56.5195	23.1794	24.9056	14.2162	37.3698

Appendix 7

Table A6: Difference in Statistics between control and extreme year. Green represents when the extreme year outperforms the control.

RMSE

d01	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
1999-2006	-30.735	3.7444	-7.5097	-33.9045	-35.883	-30.989	2.5956	-6.7393	-34.0311	-35.337
1999-2002	-4.71	1.3534	-4.0537	-0.0656	0.409	-4.836	0.5658	-3.5091	0.1667	1
1999-2005	-8.075	-2.1511	0.8884	-0.5859	-0.98	-8.596	-2.614	1.2095	-1.082	-1.218
1999-2010	-35.075	-7.4943	-7.9299	-12.1471	-18.877	-35.462	-8.3309	-7.4736	-12.5826	-18.846

d02	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
1999-2006	-0.516	8.037	-6.456	-16.278	-17.991	6.039	-2.162	-3.607	-8.768	-6.152
1999-2002	-37.561	-1.08	-31.626	-19.663	-33.218	-39.124	-9.385	-32.804	-17.42	-31.444
1999-2005	-23.018	-30.882	-3.497	23.884	22.599	-28.312	-41.776	-4.451	23.4904	22.973
1999-2010	-89.424	-48.879	-34.202	-31.504	-45.466	-92.052	-59.929	-33.188	-32.672	-45.359

d0L	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
1999-2006	-14.745	27.146	-14.73	-41.335	-43.801	-8.671	13.333	-10.108	-32.93	-30.977
1999-2002	-80.897	-14.368	-58.469	-54.425	-76.555	-71.378	-20.325	-52.265	-47.391	-65.769
1999-2005	-40.201	-78.45	4.639	32.996	42.91	-37.107	-84.548	8.269	31.3436	46.359
1999-2010	-181.625	-95.494	-62.86	-68.14	-97.895	-158.697	-88.817	-52.337	-63.495	-86.029

SD

d01	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
1999-2006	-28.255	3.701	-7.2095	-32.5259	-33.608	-28.582	2.5453	-6.4557	-32.7088	-33.129
1999-2002	-4.562	1.2143	-4.342	0.3804	0.626	-4.708	0.428	-3.7729	0.5555	1.183
1999-2005	-8.377	-9.3583	0.6164	-0.4755	-0.993	-8.84	-2.7906	0.9732	-1.013	-1.23
1999-2010	-34.956	-7.4192	-8.0984	-12.1867	-18.826	-35.355	-8.2662	-7.6403	-12.6509	-18.802

d02	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
1999-2006	-0.451	5.198	-5.841	-14.145	-15.176	-1.019	-6.338	-4.574	-8.7429	-7.458
1999-2002	-36.073	-3.758	-21.063	-20.844	-28.989	-36.668	-12.433	-19.213	-19.6259	-25.256
1999-2005	10.159	-17.921	9.636	25.7447	38.627	12.978	-24.924	11.513	25.9809	43.202
1999-2010	-38.004	-32.865	-11.512	-27.839	-17.919	-27.255	-40.058	-5.524	-27.4849	-10.25

d0L	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
1999-2006	-18.239	22.004	-15.702	-38.076	-42.726	-12.311	7.708	-11.093	-29.857	-29.78
1999-2002	-64.126	-16.653	-28.049	-55.056	-59.342	-54.025	-23.205	-21.803	-48.041	-47.639
1999-2005	28.172	-45.68	32.102	35.4759	72.45	32.739	-53.4	36.429	34.8056	78.416
1999-2010	-62.718	-60.163	-8.13	-55.38	-26.629	-32.83	-52.6	3.484	-48.215	-10.026

MAE

d01	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
1999-2006	-13.9028	1.2211	-4.2545	-11.2246	-15.3051	-13.5627	1.2221	-4.0602	-11.301	-15.1496
1999-2002	-2.4735	0.1697	-0.7124	-0.722	-1.3216	-2.47	0.1187	-0.756	-0.6146	-1.2746
1999-2005	0.1724	2.4432	-0.2571	1.7691	-1.1837	-0.2475	2.2038	-0.264	1.4693	-1.5011
1999-2010	-15.0158	-0.7696	-4.0212	-3.58	-10.1805	-15.7481	-1.2088	-4.1581	-3.8746	-10.5474

d02	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
1999-2006	-18.843	-12.4806	-10.6716	-9.0763	-14.298	-12.937	-13.1628	-9.063	-5.1642	-8.121
1999-2002	-31.794	-17.9682	-17.312	-15.8983	-20.81	-30.953	-18.4258	-18.935	-15.1217	59.3816
1999-2005	-5.477	-24.6627	0.413	24.6369	22.071	-9.001	-26.9866	-1.4115	24.5929	20.799
1999-2010	-59.742	-32.3878	-25.6455	-11.4712	-28.739	-65.701	-35.569	-27.4108	-12.835	-31.505

d0L	Total	Oct	Nov	Dec	ND	Total	Oct	Nov	Dec	ND
1999-2006	-34.599	2.9402	-17.24	-35.054	-37.362	-30.624	0.9912	-14.613	-30.8336	-30.317
1999-2002	-45.574	-15.9683	-29.887	-40.073	-37.677	-44.228	-17.3332	-27.914	-39.5526	-35.969
1999-2005	-21.27	-53.0053	8.152	29.414	37.365	-19.83	-52.4029	8.714	29.1334	38.697
1999-2010	-138.339	-62.0013	-47.615	-43.172	-71.347	-134.333	-59.4829	-44.26	-44.2386	-68.768