Research Article

The Development of a Customization Framework for the WRF Model over the Lake Victoria Basin, Eastern Africa on Seasonal Timescales

R. Argent, X. Sun, F. Semazzi, L. Xie, and B. Liu

*Department of Marine, Earth, and Atmospheric Sciences, 2800 Faucette Drive, Rm. 1125 Jordan Hall,
North Carolina State University, Raleigh, NC 27695-8208, USA*

Correspondence should be addressed to R. Argent; reargent@ncsu.edu

Received 17 May 2014; Revised 11 August 2014; Accepted 13 August 2014

Academic Editor: Richard Anyah

Copyright © 2015 R. Argent et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Lake Victoria, Africa, supports millions of people. To produce reliable climate projections, it is desirable to successfully model the rainfall over the lake accurately. An initial step is taken here with customization of the Weather, Research, and Forecast (WRF) model. Of particular interest is an asymmetrical rainfall pattern across the lake basin, due to a diurnal land-lake breeze. The main aim is to present a customization framework for use over the lake. This framework is developed by conducting several series of model runs to investigate aspects of the customization. The runs are analyzed using Tropical Rainfall Measuring Mission rainfall data and Climatic Research Unit temperature data. The study shows that the choice of parameters and lake surface temperature initialization can significantly alter the results. Also, the optimal physics combinations for the climatology may not necessarily be suitable for all circumstances, such as extreme years. The study concludes that WRF is unable to reproduce the pattern across the lake. The temperature of the lake is too cold and this prevents the diurnal land-lake breeze reversal. Overall, this study highlights the importance of customizing a model to the region of research and presents a framework through which this may be achieved.

1. Introduction

Lake Victoria is a key water resource in Eastern Africa. The lake is surrounded by three countries: Uganda, Kenya, and Tanzania. It is the largest freshwater lake in Africa, covering an area of 69,000 km² [1], but is also one of the shallowest; the bathymetry within the lake is asymmetrical east to west with an average depth of 45 m [2]. Approximately 85% of the inflow comes from rainfall directly into the lake itself, with the remaining inflow from tributaries [3]. The lake also feeds one of the sources of the River Nile, which then runs through Uganda, South Sudan, Sudan, and Egypt.

The population of the Lake Victoria Basin (LVB) is around 35 million [4] and depends on the lake for many resources. The lake provides food, water, income, and electricity to many of the people living in this region [5]. For example, the lake is the industrial and domestic water supply for around 5 million people living in the major cities that directly surround the lake [4]. It is also vitally important for

transportation and is used extensively, yet it is one of the most dangerous waterways in the world due to the severity of the weather [1].

In addition to this, 238 million people live in the River Nile basin, from all the tributary sources to the Mediterranean Sea. This large population is highly dependent on one singular water resource [6]. The countries downstream of Lake Victoria are dependent on the water released into the Nile from the Lake, as it supplies vital water to homes and agriculture as well as hydroelectric power through dams [3]. Upstream developments can affect water use downstream as they may determine when and how much water is reaching people and communities [3].

Seto et al. [7] show that both the northern side of Lake Victoria and the Nile Basin in Egypt are likely to see significant urban growth, therefore, increasing the importance of the lake and its water resources. Consequently, changes in long and short term water availability have the potential for disastrous effects. Future modeling is important

to determine if and how the region will change on multiple timescales.

The climatology over the lake is determined by both local and large scale influences [8, 9]. The dynamics associated with the geography of the Rift Valley Complex, the surrounding climate regimes, and synoptic scale features transient over the region all combine to create a highly complex and interesting climatology over the LVB. Situated in the Rift Valley region, the lake has mountainous regions to both the west and the east. The lake and surrounding land induce a large land-lake breeze circulation system [9].

The most predominant of the large scale features that influence the region is the Intertropical Convergence Zone (ITCZ). As the ITCZ 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, OND) when they are known as the short rains. 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 [9].

In addition to the ITCZ, two large teleconnections that occur in the Pacific and Indian Oceans influence the region: the El Niño—Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD). Both of these have been linked to East African rainfall, although there is ambiguity about how these two teleconnections are linked and which one exerts the most influence [10–13].

On a more local scale, the land-lake breeze is created by the surrounding high topography, the lake itself, the high insolation in the region, and the resulting temperatures gradients [9]. At night, the lake is at its warmest in comparison to the surrounding land, resulting in air flowing onto the lake and subsequent rising motion. This results in large amounts of convection, and therefore cloudiness and rainfall, overnight [14]. Anyah and Semazzi [15] found that the strength of the breeze and consequently the amount of rainfall are dependent on the lake surface temperature (LST). Anyah et al. [8] found that the high topography enhanced the breeze both during the day and at night. The mean easterly flow and asymmetrical lake temperatures (due to differing lake depths) mean that the rainfall caused by the diurnal cycle is not centered on the lake and is actually focused on the north western side [2, 14].

These complex influences interact to create an asymmetrical rainfall pattern seen across the LVB (Figure 1). This pattern needs to be captured accurately by models to build confidence in the accuracy of future projections and forecasts.

Global climate models (GCMs) are used all around the globe for future projections. In order to capture finer and more accurate details of the precipitation pattern, it is necessary to downscale through a regional climate model (RCM). It is necessary to tune any RCM used so it best reflects the dynamics and physics of the region of interest. The majority of model parameterizations are designed with a particular region in mind; therefore, it is important to choose appropriate options. The process of adjusting the tunable parameters and attributes within the regional climate model is known as customization and is important regardless

FIGURE 1: Tropical Rainfall Measuring Mission (TRMM) data showing the total rainfall (mm) over the LVB basin for the 1999 short rains (OND). The asymmetrical pattern over the lake can be clearly seen.

of the model being used. Additionally, unlike most inland regions where RCMs have been used, Lake Victoria and the corresponding land-lake breeze create an additional source of complexity and therefore uncertainty.

Previous customization in this region is very limited. The Climate Modeling Laboratory (Climlab) at North Carolina State University, in particular Sun et al. [16] and Davis et al. [17], performed substantial research in customization for the LVB using the RegCM3 model. Much more recently in 2011, Pohl et al. [18] conducted research into the customization of the Weather, Research, and Forecasting model over Eastern Africa. However, their focus was a large East African domain rather than specifically the lake basin, which is the focus of this study. Recently, Sun et al. [19] showed that the LST is important when modeling the rainfall over the lake and that, by changing the initial LST, the rainfall pattern can be influenced.

The WRF model is selected for this study due to its increasing use and the resulting large community of modelers. However, this is a suggested framework which could be applied to any model and ideally in the future would be part of an ensemble of models focused on the region.

It is important that customization is approached in a comprehensive manner. The following work proposes a framework which encompasses several aspects of customization and could be used for other regions around the globe; this study is focused on the Lake Victoria Basin and consequently will be suitable for other similar regions. Section 2 lays out the process and methods used to determine this framework. Section 3 presents the results and discusses how they can be interpreted. Finally, Section 4 concludes this study.

2. Methods

2.1. Model and Data. This project is primarily based upon the customization of a model for the LVB during the short rains (OND). The model customized is the Weather, Research, and Forecasting (WRF) model, version 3.3 [20]. The model runs are initiated with the NCEP FNL (Final) Operational Global

FIGURE 2: The domains used in WRF. The large domain, one, is at a grid spacing of 50 km and the smaller, two, nested within one has a grid spacing of 10 km.

Analysis data [21] and the Optimum Interpolation SST data set is used [22]. The datasets can be found through the University Corporation for Atmospheric Research (UCAR) [23].

Data from the Tropical Rainfall Measuring Mission (TRMM) is used for validation of the model. The data used in this effort are acquired as part of the activities of NASA's Science Mission Directorate and are archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC). The TRMM rainfall 3B42 dataset [24] is chosen as the most suitable, because it combines different precipitation estimates including both satellite coverage and station data, which resolves the problem of very limited in situ data in the region [14].

The parent domain is extensive enough to reproduce large scale features such as Kelvin waves and the Indian Ocean SST [25]. It is chosen to be similar to the domain used by Paeth and Hense [26], for Kelvin waves, but also takes into account domains used previously within Climlab at NCSU, such as those used by Bowden [27]. The nested domain covers a much smaller area centered over the Lake (Figure 2). The grid spacing of the large domain is 50 km and the nest is 10 km. Additionally, all runs use the US Geological Survey (USGS) land use data, linear lateral boundary conditions with four relaxation points for the boundary, and the alternative lake temperature initialization option within WRF unless otherwise stated in the individual experiments.

The run is initialized on 00z October 1, 1999. This year is chosen as it represents the climatology of the region [18]. However, only November is used for analysis to allow for the model to spin up and to capture the middle month of the short rains. The middle month is chosen as the basin experiences the most consistent rainfall coverage in this month. Figure 3 shows the TRMM data for the three months in the short rains. In October, the rains mostly occur in the north and in December to the south; however, in November the entire lake basin experiences the rainy season.

2.2. Evaluation Metrics. To evaluate the model output with respect to the TRMM observations, the root mean square error (RMSE), mean absolute error (MAE), 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. [28] and Pielke [29]. The skill of the model runs is evaluated using the specifications described by Pielke [29]. 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). Skill score 2 requires equating two different SDs, where within plus/minus 10% of the observational SD is considered to be equal.

The runs are analyzed and ranked based on how well they performed statistically; the best quartile and the worst quartile are identified. Evaluation metrics are calculated for each individual domain size and then consolidated into an overall ranking based upon the results in the different domains. In order to analyze the lake in detail, a subset of domain two is also considered, referred to as the lake domain.

Finally, to analyze how well the model reproduces the rainfall distribution, plots of the total rainfall and difference plots, where TRMM data has been subtracted from the model output, are created for both domains.

2.3. Experimental Design. The proposed framework consists of six experiments in order to comprehensively customize the model. These are as follows: (1) a comparison of the different physics options available in WRF, (2) a comparison of different SST datasets as well as different methods of initializing the Lake Surface Temperature, (3) further analysis on the dynamics found in the optimal runs and how they compare to the expected circulation over the lake, (4) a comparison of results for extreme years with the climatological year used for the initial analysis, (5) a comparison of results with the lake removed from statistical analysis, and (6) a comparison between the optimal physics options with respect to temperature and the optimal options with respect to rainfall.

Experiment 1, Physics Options, aims to test which physics options within WRF give optimal results. Three base combinations are used: the default settings in WRF (run A), the combination used in the East African Community Feasibility Study (run B) [1], and the combination found in preliminary customization tests conducted for just five days (run C). Beyond this, multiple variations of the base combinations are also used (runs D to M) (Table 1).

In addition to the different parameters available for customization, WRF also has many options that can be changed depending on the suitability for the domain in question. In this domain, several of these options are of interest, particularly because the aim of this customization is for a seasonal run. Further runs are conducted in order to test these parameters. The tests conducted are (a) boundary conditions: an investigation into linear versus exponential boundary and the number of relaxation points within the boundary, (b) additional parameters (henceforth referred to as the climate parameters) suggested for use in climate runs such as updating the deep soil temperature and creating bucket values for radiation and rainfall, and (c) the land use data set; WRF has the option of using land use data provided by USGS or MODIS.

FIGURE 3: The total rainfall from TRMM data for each month of the short rains, (a) October, (b) November, and (c) December, shows the transition of the rainy season over the region with the most consistent cover in November.

TABLE 1: The combinations of physics options used for each run. The base combinations are runs A–C. For further information on each option please see references [35, 37, 38, 40–52].

Run	Short wave	Long wave	Cumulus	Microphysics	Planetary boundary layer	Land surface
A	Dudhia	RRTM	Kf	Single 3	Yonsei	Noah
B	CAM	CAM	kf	Single 5	Yonsei	Noah
C	Dudhia	RRTM	Grell 3D	Eta	ACM2	Noah
D	CAM	CAM	BMJ	Single 5	Yonsei	Noah
E	CAM	CAM	Grell 3D	Eta	ACM2	Noah
F	CAM	CAM	BMJ	Eta	ACM2	Noah
G	Dudhia	RRTM	Grell 3D	Eta	ACM2	5-layer
H	Dudhia	CAM	Grell 3D	Eta	ACM2	Noah
I	CAM	CAM	kf	single 6	Yonsei	Noah
J	Dudhia	RRTM	Grell 3D	Eta	ACM2	RUC
K	Dudhia	RRTMG	Grell 3D	Eta	ACM2	Noah
L	Dudhia	RRTM	BMJ	Eta	ACM2	Noah
M	Dudhia	RRTM	Grell 3D	Eta	Yonsei	Noah

While Sun et al. [19] showed the importance of the LST, it is important to determine that these results were not just based upon the data set or the way in which the temperature is initialized. This section aims to satisfy those uncertainties. For experiment 2, Lake Surface Temperature, four different runs are conducted. These take into account two different LST

initialization methods and two different SST data sets that WRF uses. The default initialization option interpolates from the nearest water source and consequently is dependent on the SST data set and changes within it. The alternative option uses the air temperature as a proxy that it then sets as the LST for the entire run [30]. The SST datasets are the previously

mentioned Optimally Interpolated SST and the ECMWF ERA Interim data [31], obtained from the ECMWF Data Server. Each data set is run with each different initialization method. The model setup for all four runs is the same as combination H from part 1. This combination is chosen as it is the best performing combination in preliminary testing (not discussed here).

For experiment 3, Optimal Physics and Circulation Patterns, the physics combinations that gave the best results are chosen from the initial preliminary tests. The runs for this chapter are started earlier than previous runs, in order to ensure that spinup is not a problem. Unfortunately, lack of initialization data meant that the earliest date from which these runs could be started is September 12. November remains the focus of these runs. In addition to this longer time period, there are other set-up differences from the preliminary testing, based on the results of experiment one. These are as follows: exponential Boundary Conditions with a relaxation zone of 9 grid spaces, use of the climate parameters, and using the USGS land use data. The remaining set-up is as previously stated. In addition to the previously mentioned evaluation metrics, plots are created of monthly average temperature, vector wind, and vertical and horizontal motion cross sections over the lake throughout the day. The temperature is compared to data from the Climatic Research Centre at the University of East Anglia [32].

Experiment 4, Extreme Years, uses the same basic set-up, combination E, as experiment 3 and consists of four different years to represent extremes. The results in experiment 3 represent 1999 which is acting as a climatological control run. The differences between the statistics for each extreme year and the control run are calculated. A large amount of the variation within the short rains is due to a combination of ENSO and IOD [33, 34]. Consequently, it was decided that the extreme years would be based on the indices of these two teleconnections. 2002 and 2006 are both positive years with both positive ENSO and IOD indices. The negative years are 2005 and 2010, with opposite signals.

The model runs for both experiments 5, Lake Masking Comparison, and 6, Temperature Comparison, are the same as performed and analyzed for the initial customization in experiment 1. However, in experiment 5, for every run statistics are also calculated with the lake masked and therefore that region of the output removed from the calculations. These results are then compared to those found in experiment 1. For experiment 6, the analysis is conducted with respect to temperature instead of rainfall and validated with Climatic Research Unit data. The results are compared with the results from part one to determine if the same physics options are optimal across multiple variables.

3. Results and Discussion

3.1. Physics Options. When considering both the rainfall distribution and the evaluation metrics, it is apparent that while one physics combination cannot be objectively stated as the best, some combinations show significantly higher skill and therefore should be deemed more appropriate for use. Table 2 shows the statistical ranking from the RMSE, both

of the individual domains and the overall ranking for all three metrics based upon how each combination performs in each domain. Figure 4 shows the rainfall pattern over the lake domain for all of the different physics combinations. Overall, Figure 4 shows that none of the options used are able to capture the rainfall pattern over the lake and that this result extends to other lakes within the lake domain, where the rainfall is also underestimated. Another consistent pattern is that all the runs overestimate the rainfall over the Congo rainforest in the western side of the domain.

Run L is consistently in the upper quartile of all three statistical error scores and shows a greater amount of skill than the majority of the other combinations, particularly in domain 2, when analyzed using the skill scores. Runs D and E also perform relatively well. They both consistently appear in the upper half of all the error statistic rankings in both the overall ranking and those for each domain. However, for the mean absolute error, both runs outperform the other combinations and make up the majority of the upper quartile in addition to run L. The two runs also show similar amounts of skill as run L.

The worst runs are identified as I, G, and B. While there is substantial variation within the lower quartile, these three runs are consistently in the lower statistics in RMSE, MAE, and SD. Additionally, none of these runs show skill in the skill scores. These runs dramatically overestimate the rainfall over large areas of the region and do not capture the rainfall distribution well. B and I both use the Kain Fritsch (KF) cumulus scheme [35], which appears to lead to much higher amounts of rainfall, possibly due to being developed for the continental USA [36].

Figure 5 shows the difference between modeled and observed (TRMM) rainfall. The overall distribution of the rainfall is more accurately reflected in runs L, D, and E where they capture more of the pattern surrounding the lake and throughout the domain. In particular, these runs show the asymmetry on either side of the lake, capturing the low rainfall on the western shore and high rainfall on the eastern shore. The rainfall over the Indian Ocean is also much more accurate in runs L and D than in many of the other runs, which tend to overestimate this region. It is suspected that the use of the Betts-Miller-Janjic (BMJ) [37, 38] cumulus scheme is instrumental in these results as it is a consistent factor between the two runs.

All runs show an excess of rainfall over the Congo rainforest and deficient rainfall over the lake itself. Although there is significant variability between the different runs, the majority of the combinations show too much rainfall over the Indian Ocean and underestimate over the remaining continent.

Run H, whilst performing well in the evaluation metrics, does not capture the correct rainfall distribution and largely underestimates over the majority of the continental region (Figure 4). It is suspected that the good statistical results are due to the fact that many of the other runs significantly overestimate the rainfall. This highlights the importance of not just basing the results on one particular form of analysis, such as one error metric like RMSE.

Finally, the additional runs show that using the climate parameters within WRF gives better results and should be

TABLE 2: (a) The Ranking of all the different runs with their corresponding RMSE (best to worst in descending order) separated by domain. (b) The overall RMSE, SD and MAE ranking based on all the domains.

(a)					
Root Mean Square Error					
Domain 1		Domain 2		Lake Domain	
Ranking	Error	Ranking	Error	Ranking	Error
E	66.5	L	105.7	H	138.9
A	70.7	D	106.4	L	153.4
H	72.2	H	114.9	G	159.5
K	75.2	E	117.9	D	162.4
C	77.4	C	126.5	A	163.3
J	80.3	F	127.2	E	166.3
L	82.7	A	127.3	K	172.4
D	84.3	K	130.0	C	184.4
F	87.1	J	137.8	M	185.1
B	91.3	M	142.5	I	189.4
I	91.4	G	148.2	J	192.3
M	92.1	B	154.0	B	194.7
G	113.9	I	159.6	F	198.8

(b)		
RMSE	SD	MAE
Overall	Overall	Overall
Ranking	Ranking	Ranking
H	H	C
L	L	E
<u>A</u>	A	<u>L</u>
D	C	A
E	<u>E</u>	H
C	K	D
K	D	F
F	<u>M</u>	K
M	F	J
<u>J</u>	G	<u>G</u>
G	J	M
B	B	B
I	I	I

The lines represent the upper and lower quartiles, in the SD ranking runs A, C and E and runs F, G and J are equal hence the displaced lines.

used for future runs. The differences between the statistical results for both the boundary conditions and the land use tests are very small and do not have a large impact on the results. USGS and an exponential boundary with 9 relaxation points are however found to be best in this situation. However, these areas need greater study to help determine the optimal results.

None of the parameter combinations resolve the complex pattern over the lake, regardless of how well they performed otherwise. Many of them capture the rainfall distribution surrounding the lake, although they rarely reproduce the correct rainfall amounts. When considering the rainfall over the lake itself, all of the runs underestimate the rainfall. This underestimation is particularly large over the western half

of the lake, where greater rainfall is shown in the observed data. Some of the runs, such as D and L, reproduce the correct wet and dry areas on either side of the lake but fail to reproduce the rainfall over the lake, resulting in a dry-dry-dry-wet pattern instead of the dry-wet-dry-wet pattern in the observations.

This first experiment highlights the importance of customization for a particular region and the correct selection of parameters. For example, changing the cumulus parameter from the KF to the BMJ scheme changes a model run from being one of the most accurate (D) to one of the least accurate (B). It also confirms that it is difficult to fully determine what the overall optimal physics are, as the number of combinations is too extensive to fully investigate.

FIGURE 4: Total November simulated rainfall (mm) over the lake domain for the different physics options (a-m) and TRMM observations for comparison (n). Note the lack of rainfall over western side of the lake in each figure.

FIGURE 5: The difference between the WRF output and the TRMM data for each combination in domain two. Blue shows where the model underestimates and red where it overestimates.

FIGURE 6: November difference (in mm) between TRMM data and WRF model output over domain 2.

3.2. Lake Surface Temperature. Experiment two compares two methods of initializing the lake surface temperature and two different datasets. Once again, none of these runs are able to capture the asymmetrical rainfall pattern over the lake. One run, using the OI SST and the original initialization (for this experiment only, run O), produces a significantly worse distribution than the other three (Figure 6). Although the remaining three are more comparable, using the OI SST and the alternative initialization method (henceforth run OA) gives the optimal performance.

Statistically, O performs the worst in both the RMSE and SD tests and second worst in the MAE. In most cases, the difference from the other runs is also considerable. For example, in domain 2, run OA has the lowest RMSE of 111.8 mm, comparable to the other two runs (henceforth E and EA, using the ECMWF SST data and the original and alternative initialization, resp.) which have errors of 116.3 mm and 117.5 mm, respectively. Conversely, run O has an error of 136.4 mm. The error in run O also increases as each domain focuses on the lake in greater detail; while the other runs do show this as well, it is to a lesser extent. For instance, between the RMSE for domain 1 and domain 2, run O saw a difference of 61.2 mm whereas run OA only saw a difference of 39.1 mm. This shows that the main region of error in the runs is in the lake region, but in run O this error is worse than in the others. In both the RMSE and MAE, OA performs the best; again, the difference between this run and the other two is far smaller

in comparison to the difference to O. In the SD, E slightly outperforms OA but the differences remain small, with the largest difference being around 13 mm and the majority under 10 mm.

The rainfall distribution and difference from TRMM data over the large domains show that the rainfall and the SST are very similar in all four runs. The key differences are found when the lake is examined in greater detail. Run O shows an overestimation over the lake, whereas the other three show an underestimation (Figure 6). It is also observed in the difference plot that the lake shows an asymmetry of results, which means that the model is not capturing the necessary pattern over the lake.

Both the initialization process and the SST data have the potential to significantly influence the rainfall and how accurately it is represented. However, when using the second method of initialization, the SST data becomes of less importance to the lake region, as the initialization then uses air temperature as a proxy. The alternative initialization does lead to an improvement, but it does not fully resolve the problem as the lake is still the most significant source of error.

It is already recognized by Sun et al. [19] that the LST is very important. This experiment shows that differences in data and initialization options do not significantly increase the ability of the model to capture the LST and consequently the rainfall. Sun et al. [39] look at this problem in greater detail with a coupled lake surface temperature model.

3.3. Optimal Physics and Circulation Patterns. Overall the rerun optimal physics choices from experiment one perform as expected. This result is consistent in both the statistics and the rainfall distribution. All runs show a similar pattern, overestimation of the Atlantic Ocean (in the large domain) and central continent (in particular the Congo rainforest) and underestimation of large continental regions and the lake (difference plots are analyzed but are not shown here). One area of difference is the representation of the Indian Ocean where the run containing the BMJ cumulus scheme outperforms the others. Again, it is highlighted that none of the runs are able to reproduce the asymmetrical pattern over the lake.

The next analysis is of the physics and circulation surrounding the lake. The results in this experiment are given for just one of the optimal combinations tested, with the same combination as run E in experiment 1; similar results are found for the other runs. The first additional variable to be considered is the mean temperature, analyzed on a mean monthly timescale for comparison with CRU. Overall a general negative bias is observed, with WRF underestimating the temperature over the majority of the continent by several degrees. This may be a contributing factor to the underestimation of rainfall over large areas of the domain. It is noteworthy that a bias of even two degrees could have a significant impact when looking at future changes, as the bias is around the same order of magnitude expected in future warming.

Following the temperature analysis, the circulation over the lake is considered. If the land lake breeze was forming, there would be convergence over the lake rising motion and consequently rainfall at night, with the reverse during the day. East-West cross sections of the zonal and vertical winds do appear to show a reversal (Figure 7). The rising motion however is weak and inconsistent when it should be much stronger; similarly, the zonal flow onto the lake is much weaker than the corresponding flow off the lake during the day. Analysis of the vector winds (Figure 8) confirm that there is strong flow on land during the day but also show that there is a large amount of divergence occurring over the lake at night, which would not be expected if the model is accurately reproducing the diurnal circulation.

The relative humidity (Figure 7) cross sections are consistent with the wind patterns, low humidity over the lake during the day and higher humidity at night. However, very little moisture appears to be above 4 km in the atmosphere as expected with large convective storms. In the west of the domain, there are large areas of moisture throughout the atmosphere, occurring over the Congo Rainforest and we would expect the night time moisture over the lake to look similar to this. Consequently, the circulation reversal that occurs at night is not taking place, and thus the land lake breeze system is not being fully resolved by this model. This is most likely due to incorrect representation of the lake surface temperature as discussed by Sun et al. [19].

It is theorized that the models inability to capture the flow is due to the temperature of the lake. The daytime flow is captured by the model, as this is the time period when the lake is colder than the shore. Although the LST being too cold may influence the strength, it does not impact the occurrence

TABLE 3: RMSE (mm) for each extreme year and for both domains. 2010 saw a large increase in error.

	Year	Domain 1	Domain 2
Control	1999	65.8	119.7
Positive	2006	73.4	126.1
	2002	69.9	151.3
Negative	2005	64.9	123.2
	2010	73.8	153.9

of the flow. This flow will still influence divergence over the lake and consequently low rainfall. However the lake, due to having an underestimated LST, is never warmer than the surrounding land and this prevents the night time reversal of the land-lake breeze. This constant one way flow or very weak reversal explains the lack of rainfall over the lake in the model, as the rainfall is primarily due to convergence at night with the reversal of the land-lake breeze.

The lack of rainfall not only prevents the model from reproducing the asymmetrical pattern but also means 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 is able to reproduce the correct rainfall pattern, it is necessary to get the fundamental flow over the lake correct and at present time this experiment has shown that this is not occurring.

3.4. Extreme Years. The extreme years do not show comparable results statistically when compared to the control year of 1999. 2005 is the only one with comparable error results; the remaining three years all have larger errors; in particular, 2002 and 2010 have much larger errors with 2010 being the worst (Table 3). The inaccuracy corresponds to the strength of the ENSO signal as 2005 has the weakest ENSO index and 2010 has the strongest. The inaccuracy does not correspond to the strength of the IOD signal.

Figure 9 shows the rainfall for each year with the corresponding TRMM rainfall and the difference between the two. The positive and negative extreme years are observed in the TRMM data; however, the model varies less significantly. The model appears to overestimate the low rainfall years and underestimate the high rainfall years meaning that WRF is not accounting for the changes in these large scale influences. The model still overestimates large areas over the Congo Rainforest and once again the lake is still not accurately represented. The asymmetrical pattern is reproduced by the TRMM data in all five years; however, this pattern is not reflected in the WRF rainfall in any of the years.

Experiment four reveals that WRF is unable to reproduce the influence of the large scale features on the region. Consequently, the same combination of physics options cannot reproduce the results to the same accuracy for an extreme year as for the original climatology year. The influences on the region are different during years when one of the large scale teleconnections is dominant. Further customization is needed for those years to produce another parameterization combination for years of particularly strong teleconnection

FIGURE 7: (a) Horizontal motion, (b) vertical motion, and (c) relative humidity at 00z and 12z on the November 22. The lake is situated in the center of the image, at approximately grid spaces 100–130.

modes; in the case of the region, the mode of focus would be ENSO.

3.5. Lake Mask and Temperature Comparisons. Overall, with the lake masked and taken out of the statistical analysis, the results are not significantly different. A few of the runs move up or down in the rankings statistically and this is particularly prominent in the lake domain, where, as expected, the lake would have the largest impact. However, the upper quartile is only impacted on one instance and not significantly; therefore, overall it appears that the optimal combinations are unaffected by the lake masking. Consequently, the lake does influence the results, but in this case not significantly enough

to bias the results. These results highlight that one feature does have the potential to bias the results of the entire domain and therefore caution should be used.

This raises the question as to whether customization 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 customize 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.

The comparison to temperature does give different results. When considering the RMSE and the MAE, the

FIGURE 8: Vector winds over the lake at 00z and 12z on the November 22 at the 850 hPa pressure level. Wind speed represented using wind barbs in knots.

optimal physics combinations for temperature are different to those for rainfall with the only exception being run L which remains within the upper quartile for both variables. The worst quartile for temperature contains both H and D across several evaluation metrics, which are in the upper quartile for the rainfall. The SD results are not in agreement with the RMSE and MAE showing that the runs with the least error do not necessarily have the most consistent bias unlike many of the rainfall related results.

Overall, there is a smaller range of error for the temperature results than there is for the rainfall results. The majority of the errors are within a degree of one another and many are much less. There is still a generally negative bias of the domain with some of the better runs showing less of this bias. As previously mentioned, the bias is significant enough that it may be of consequence if looking at a warming trend.

It is shown that it is important to not just check the variable of interest but also analyze 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 represent the entire climate system well. Therefore, it may be necessary to balance optimization across several variables. In this case, run L would be the optimal solution, as it performed well in both temperature and rainfall analyses.

4. Conclusion

Each experiment of this study has highlighted an important area of the customization process. This is necessary for confirming that WRF is producing the most accurate results possible for a particular region, in this case a month of the short rains season over the LVB. The individual experiments all emphasize individual aspects of this crucial process and also agree with and reinforce initial ideas and theories.

Some of the main components of the customization process are the parameters and options available in WRF. Some schemes will always be better suited to particular cases. In this study, several of the schemes are stressed as either being suitable or being 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 show a much better representation of rainfall over the Indian Ocean. The majority of runs produce a large overestimation over the ocean, whereas this scheme does not. Other options appear responsible for inaccurate results such as the KF cumulus scheme, which results in large amounts of precipitation and consequently an overestimation over a large proportion of the domain.

Masking the lake does not have a significant impact on which combinations produce the most accurate results. However, it does produce minor modifications in the ranking of how well the combinations perform, especially within the lake domain. Conversely, the temperature comparison does cause the accuracy of the combinations to change. This emphasizes 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 is that of run L as it produces good results for both rainfall and temperature. It also contains the BMJ scheme. This is the combination that would be recommended for use in future studies.

However, the best combination for the climatology is not necessarily the best for the extreme years. It appears that it is necessary to perform additional customization for these years. For this region, the ENSO mode has the largest impact on the accuracy of the runs and it is important to adapt the customization to support this source of variability.

Finally, the main emphasis of this study was on the asymmetrical pattern that occurs over the LVB and whether or not 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 is sometimes able to capture the asymmetry on each shore; however, it is categorically unable to reproduce the correct rainfall distribution directly over the lake.

It is recognized that the model is unable to initialize the lake surface temperature correctly, causing this incorrect rainfall distribution. While Sun et al. [19] have shown the critical nature of the LST, this study shows that this problem with LST occurs regardless of which initialization or data set is used. WRF offers two options of initializing the lake; one, a direct interpolation of SST, creates 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 initialization causes too little rainfall regardless of the other parameters or variables within the run. It is theorized that this lack of rainfall is due to the cold temperature preventing the reversal of the land-lake breeze.

Overall, it appears that a customization study can improve the results over the general domain, but it does not improve the results directly over the lake. Future work will aim to combine this framework with the coupled model developed in Sun et al. [39].

This framework provides a method of comprehensively customizing a model for a particular region and highlights important aspects of the customization process. It can be

FIGURE 9: Total rainfall in November in Domain two for each extreme year. WRF rainfall is on the right and TRMM rainfall on the left and the difference between them is the lower image.

used for future work, not just around Lake Victoria, but in other situations with prominent geological features which dominate the local circulation.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This project was funded through the NSF Grant no. 553210.

References

- [1] East African Community (EAC), *Enhancing Safety of Navigation and Efficient Exploitation of Natural Resources over Lake Victoria and its Basin by Strengthening Meteorological Services*, East African Community, Nairobi. Kenya, 2011.
- [2] Y. Song, F. H. M. Semazzi, L. Xie, and L. J. Ogallo, "A coupled regional climate model for the Lake Victoria Basin of East Africa," *International Journal of Climatology*, vol. 24, no. 1, pp. 57–75, 2004.
- [3] J. V. Sutcliffe and Y. P. Parks, *The Hydrology of the Nile*, International Association of Hydrological Sciences Wallingford, Oxfordshire, UK, 1999.
- [4] World Bank, "Africa—second phase of lake Victoria environmental management project," Project Information Document AB3561, World Bank, Washington, DC, USA, 2008, <http://documents.worldbank.org/curated/en/2008/09/9853300/africa-second-phase-lake-victoria-environmental-management-project>.
- [5] M. J. Ntiba, W. M. Kudoja, and C. T. Mukasa, "Management issues in the Lake Victoria watershed," *Lakes and Reservoirs: Research and Management*, vol. 6, no. 3, pp. 211–216, 2001.
- [6] Nile Basin Initiative, "State of the River Nile Basin Report," 2012.
- [7] K. C. Seto, B. Güneralp, and L. R. Hutya, "Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 40, pp. 16083–16088, 2012.
- [8] R. O. Anyah, F. H. M. Semazzi, and L. Xie, "Simulated physical mechanisms associated with climate variability over Lake Victoria basin in East Africa," *Monthly Weather Review*, vol. 134, no. 12, pp. 3588–3609, 2006.
- [9] G. C. Asnani, *Tropical Meteorology: 2005 Revision*, G.C. Asnani, 1993.

- [10] S. K. Behera, J.-J. Luo, S. Masson et al., "Paramount impact of the Indian Ocean Dipole on the East African short rains: a CGCM study," *Journal of Climate*, vol. 18, no. 21, pp. 4514–4530, 2005.
- [11] D. Manatsa, C. H. Matarira, and G. Mukwada, "Relative impacts of ENSO and Indian Ocean dipole/zonal mode on east SADC rainfall," *International Journal of Climatology*, vol. 31, no. 4, pp. 558–577, 2011.
- [12] C. O. Clark, P. J. Webster, and J. E. Cole, "Interdecadal variability of the relationship between the Indian Ocean zonal mode and East African coastal rainfall anomalies," *Journal of Climate*, vol. 16, no. 3, pp. 548–554, 2003.
- [13] C. J. Schreck III and F. H. M. Semazzi, "Variability of the recent climate of eastern Africa," *International Journal of Climatology*, vol. 24, no. 6, pp. 681–701, 2004.
- [14] S. E. Nicholson and X. Yin, "Mesoscale patterns of rainfall, cloudiness and evaporation over the Great Lakes of East Africa," in *The East African Great Lakes: Limnology, Palaeolimnology and Biodiversity*, vol. 12, pp. 93–119, 2002.
- [15] R. O. Anyah and F. H. M. Semazzi, "Simulation of the sensitivity of Lake Victoria basin climate to lake surface temperatures," *Theoretical and Applied Climatology*, vol. 79, no. 1–2, pp. 55–69, 2004.
- [16] L. Q. Sun, F. H. M. Semazzi, F. Giorgi, and L. Ogallo, "Application of the NCAR regional climate model to eastern Africa 1. Simulation of the short rains of 1988," *Journal of Geophysical Research D: Atmospheres*, vol. 104, no. 6, pp. 6529–6548, 1999.
- [17] N. Davis, J. Bowden, F. Semazzi, L. Xie, and B. Önl, "Customization of RegCM3 Regional Climate Model for eastern Africa and a tropical Indian Ocean domain," *Journal of Climate*, vol. 22, no. 13, pp. 3595–3616, 2009.
- [18] B. Pohl, J. Crétat, and P. Camberlin, "Testing WRF capability in simulating the atmospheric water cycle over Equatorial East Africa," *Climate Dynamics*, vol. 37, no. 7–8, pp. 1357–1379, 2011.
- [19] X. Sun, L. Xie, F. H. M. Semazzi, and B. Liu, "Effect of lake surface temperature on the spatial distribution and intensity of the precipitation over Lake Victoria Basin," *Monthly Weather Review*. In press.
- [20] W. Wang, C. Bruyère, M. Duda et al., *WRF-ARW Version 3 Modeling System User's Guide*, Mesoscale & Microscale Meteorology Division, National Center for Atmospheric Research, Boulder, Colo, USA, 2011.
- [21] National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, "NCEP FNL Operational Model Global Tropospheric Analyses, Continuing from July 1999," Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, 2000.
- [22] R. W. Reynolds, N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, "An improved in situ and satellite SST analysis for climate," *Journal of Climate*, vol. 15, no. 13, pp. 1609–1625, 2002.
- [23] 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. NCEP Version 2.0 OI Global SST and NCDC Version 3.0 Extended Reconstructed SST Analyses. 1986, updated weekly; Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, 2013, <http://rda.ucar.edu/datasets/ds277.0/>.
- [24] G. J. Huffman, R. F. Adler, D. T. Bolvin et al., "The TRMM Multi-satellite Precipitation Analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales," *Journal of Hydrometeorology*, vol. 8, no. 1, pp. 38–55, 2007.
- [25] S. Nicholson, "A review of climate dynamics and climate variability in eastern Africa," in *Limnology, Climatology and Paleoclimatology of the East African Lakes*, pp. 25–56, 1996.
- [26] H. Paeth and A. Hense, "On the linear response of tropical African climate to SST changes deduced from regional climate model simulations," *Theoretical and Applied Climatology*, vol. 83, no. 1–4, pp. 1–19, 2006.
- [27] J. H. Bowden, *Multi-Scale Climate Change Modeling Study over the Greater Horn of Africa*, North Carolina State University, 2008.
- [28] D. Carvalho, A. Rocha, M. Gómez-Gesteira, and C. Santos, "A sensitivity study of the WRF model in wind simulation for an area of high wind energy," *Environmental Modelling & Software*, vol. 33, pp. 23–34, 2012.
- [29] R. A. Pielke, *Mesoscale Meteorological Modeling*, Academic Press, New York, NY, USA, 2002.
- [30] W. C. Skamarock, J. B. Klemp, J. Dudhia et al., "A description of the advanced research WRF version 3," NCAR Technical Note, 2008.
- [31] D. P. Dee, S. M. Uppala, A. J. Simmons et al., "The ERA-Interim reanalysis: configuration and performance of the data assimilation system," *Quarterly Journal of the Royal Meteorological Society*, vol. 137, no. 656, pp. 553–597, 2011.
- [32] P. Jones and I. Harris, "CRU Time Series (TS) high resolution gridded datasets," University of East Anglia Climatic Research Unit (CRU), NCAS British Atmospheric Data Centre, 2008, http://badc.nerc.ac.uk/view/badc.nerc.ac.uk_ATOM_dataent.1256223773328276.
- [33] NOAA/National Weather Service, "Cold and Warm Episodes by Season," 2013, http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml.
- [34] JAMSTEC, "Indian Ocean Dipole—2012," 2013, http://www.jamstec.go.jp/frsgc/research/d1/iod/e/iod/dipole_mode_index.html.
- [35] J. S. Kain, "The Kain-Fritsch convective parameterization: an update," *Journal of Applied Meteorology*, vol. 43, no. 1, pp. 170–181, 2004.
- [36] J. S. Kain and J. M. Fritsch, "A one-dimensional entraining/detraining plume model and its application in convective parameterization," *Journal of the Atmospheric Sciences*, vol. 47, no. 23, pp. 2784–2802, 1990.
- [37] Z. I. Janjic, "The step-mountain eta coordinate model: further developments of the convection, viscous sublayer, and turbulence closure schemes," *Monthly Weather Review*, vol. 122, no. 5, pp. 927–945, 1994.
- [38] Z. I. Janjic, "Comments on 'Development and evaluation of a convection scheme for use in climate models,'" *Journal of the Atmospheric Sciences*, vol. 57, article 3686, 2000.
- [39] X. Sun, L. Xie, F. H. M. Semazzi, and B. Liu, "A numerical investigation of the precipitation over Lake Victoria basin using a coupled atmosphere-lake limited-area model," *Advances in Meteorology*, vol. 2014, Article ID 960924, 15 pages, 2014.
- [40] F. Chen and J. Dudhia, "Coupling and advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: model implementation and sensitivity," *Monthly Weather Review*, vol. 129, no. 4, pp. 569–585, 2001.
- [41] S. A. Clough, M. W. Shephard, E. J. Mlawer et al., "Atmospheric radiative transfer modeling: a summary of the AER codes," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 91, no. 2, pp. 233–244, 2005.

- [42] W. D. Collins, P. J. Rasch, B. A. Boville et al., "Description of the NCAR community atmosphere model (CAM 3.0)," Tech. Rep., 2004.
- [43] J. Dudhia, "Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model," *Journal of the Atmospheric Sciences*, vol. 46, no. 20, pp. 3077–3107, 1989.
- [44] G. A. Grell and D. Dévényi, "A generalized approach to parameterizing convection combining ensemble and data assimilation techniques," *Geophysical Research Letters*, vol. 29, no. 14, pp. 381–384, 2002.
- [45] S.-Y. Hong, J. Dudhia, and S.-H. Chen, "A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation," *Monthly Weather Review*, vol. 132, no. 1, pp. 103–120, 2004.
- [46] S.-Y. Hong, Y. Noh, and J. Dudhia, "A new vertical diffusion package with an explicit treatment of entrainment processes," *Monthly Weather Review*, vol. 134, no. 9, pp. 2318–2341, 2006.
- [47] S. Hong and J. J. Lim, "The WRF single-moment 6-class microphysics scheme (WSM6)," *Journal of the Korean Meteorological Society*, vol. 42, pp. 129–151, 2006.
- [48] E. J. Mlawer, S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, "Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave," *Journal of Geophysical Research D: Atmospheres*, vol. 102, no. 14, pp. 16663–16682, 1997.
- [49] J. E. Pleim, "A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: model description and testing," *Journal of Applied Meteorology and Climatology*, vol. 46, no. 9, pp. 1383–1395, 2007.
- [50] E. Rogers, T. Black, B. Ferrier, Y. Lin, D. Parrish, and G. DiMego, "Changes to the NCEP meso eta analysis and forecast system: increase in resolution, new cloud microphysics, modified precipitation assimilation, modified 3DVAR analysis," NWS Technical Procedures Bulletin, 2001.
- [51] T. G. Smirnova, J. M. Brown, and S. G. Benjamin, "Performance of different soil model configurations in simulating ground surface temperature and surface fluxes," *Monthly Weather Review*, vol. 125, no. 8, pp. 1870–1884, 1997.
- [52] T. G. Smirnova, J. M. Brown, S. G. Benjamin, and D. Kim, "Parameterization of cold-season processes in the MAPS land-surface scheme," *Journal of Geophysical Research D: Atmospheres*, vol. 105, no. 3, pp. 4077–4086, 2000.