List of all relevant changes made the manuscript

1. The disagreement among researchers on results of future precipitation has also been quoted in this manuscript (p1, l11). The question is, what is the new knowledge given in this article?

The short comings of the previous climate change studies on the study area of UBNRB and the new knowledge given in this article is included. This is shown in the manuscript with track changes on page 4, l4to l33

2. The objective of the paper, is it an evaluation of downscaling methods? or is it evaluation of climate model results for future precipitation and temperatures? or both?

The objective of the paper is the objective of this study is to construct and analyze detailed climate change scenarios for precipitation, maximum and minimum temperature over Upper Blue Nile River Basin at required resolution which can be used for further hydrological impact study. This can be achieved through the inclusion of multi-model approach and two downscaling methods by incorporating acceptable number of weather stations which has long time series and reliable observed climate data to appreciate the uncertainties coming from GCMs and the process of downscaling methods to overcome the short comings of the previous studies on the study area. Mean while evaluation of both climate models and downscaling techniques were carried out to account the uncertainties of both climate models and downscaling methods. This is shown in the manuscript with track changes on page 4, l34 and on page 5, l1 and 5.

3. The paper gives a lot of details and too many numbers make it very difficult to follow a clear story line that serves the key message of the paper. E.g.,evaluation is made at individual stations, and then for the whole catchment (p11, l23), and large differences were found among the models for the later. What does this mean for the overall uncertainty of the analysis?

This issue has been addressed in AC1 with explanation. Some texts are deleted not to make confusion for the readers and to make the paper more clear. This is shown in the manuscript with track changes on page 5, l6 to l16, page 13, l13, l14, l17 to l20, page14, l15, 16, 28,32, page15, l15, page16, l15-32, page17, l21-23 and on page 18, l22,23,25 and 26 and page19, l2-9.

4. The paper has too many tables and figures, the readability of the paper could have improved if limiting the analysis to precipitation only.

Figure 9 and 11 are removed from the paper as per the comment above to make the paper more readable. However, we intentionally include the projection of maximum and minimum temperature in the paper as they are the most required climatic variables for hydrological models to study impacts of climate change on the hydrology and water resources of the study area and could be the authors future focus.

5. P1, l11, “However, a large uncertainty between different Global Circulation Models (GCM) and downscaling methods exist that makes reliable conclusions for a sustainable water management difficult.” This is known for many years now, please give what is new that the reader is expecting from this paper.

In the paper multi-model approach was applied together with inclusion of two widely used statistical downscaling in order to see and evaluate the uncertainties' of both climate models and downscaling methods by incorporating adequate number and spatially representative weather stations. This is shown in the manuscript with track changes on page 1, l13 to l18 and on page 5, l2 and 5.

6. P1, l14, LARS-WG, SDSM; give full name when appears for first time in manuscript: accepted The Abstract is not easy to follow. Try to mention the key message (objective) of the paper, and key results, without many details. Too many models, and too many numbers makes it difficult to grasp the main findings of the paper.

Accepted and we revised accordingly and it is shown in the marked up manuscript.

7. P2, l15, climate change and climate variability mentioned on the same line. What is the difference between variability and change, please make critical discussion on this issue.

The difference and critical discussion on this issue is included. This is shown in the manuscript with track changes on page 2, l25 to l31 in the context of the study area.

8. p3, l8, there are more studies on Upper Blue Nile climate, e.g., Tesemma et al., 2010; Gebre et al., 2014, Beyene et al., 2010, among others

Accepted and included in the Authors comment#1 and in the revised manuscript as well. This is shown in the manuscript with track changes on page 3, l24 to l28 in the context of the study area.

9. P5, l15, 6 best performed models, how selected? mention few lines about criteria, and selection process.

Detail description of the procedure was included in the Authors comment #1 and here the selection criteria and summarized process is incorporated. This is shown in the manuscript with track changes on page 7, l16 to l23.

10. Limitation of LARS WG, If LARS-WG is applied for individual station, it significantly distorted the spatial correlation between stations.

Accepted and validation was done and presented. This is shown in the manuscript with track changes on page 19, l22 to l34 and on page 34, Figure 13 .

11. Figure 2, font is too small.

Accepted and corrected. This is shown in the manuscript with track changes on page 29, Figure2.

12. Table 2: Showing only percentage of passing tests might be enough.

Accepted and corrected. This is shown in the manuscript with track changes on page 25, Table 2

13) Table3: Please describe how you selected these predictors.

it is described in the paper under section 4.2 page 11, l15-20

1 4) Table 5 & 6: I am sure these table can be exchanged to figures for readers to easily understand the results.

Accepted and corrected. This is shown in the manuscript with track changes on pages 26 and 27, Table 5 and 6 deleted and accordingly on page 30 and 31, Figure 5 and 6 are added (Tables 5 and 6 are changed to Figures 5 and 6 respectively).

Analyzing the future climate change of Upper Blue Nile River Basin (UBNRB) using statistical down scaling techniques

Dagnenet Fenta Mekonnen[1,2], Markus Disse[1]

[1]Chair of Hydrology and River Basin Management, Faculty of Civil, Geo and Environmental Engineering, Technische Universität München, Arcisstrasse 21, 80333, Munich, Germany.
[2]Amhara Regional State Water, Irrigation and Energy Development Bureau, Bahirdar, Ethiopia
Correspondence to: Dagnenet Fenta (dagnfenta@yahoo.com)

Abstract. Climate change is becoming one of the most arguable and threatening issues in terms of global context and their responses to environment and socio/economic drivers. Its direct impact becomes critical for water resource development and indirectly for agricultural production, environmental quality, economic development, social well-being. However, a large uncertainty between different Global Circulation Models (GCMs) and downscaling methods exist that makes reliable conclusions for a sustainable water management difficult. Hence, to value the uncertainty of GCMs and downscaling processes, - a multi-model approach from a systematically selected six CMIP3 GCMs and one CMIP5 GCMs were used to construct climate change scenarios of precipitation, maximum and minimum temperature for the UBNRB ~~In order to understand the future climate change of the Upper Blue Nile River Basin,~~ by applying two widely used statistical down scaling techniques namely Long Ashton Research Station Weather Generator (LARS-WG) and Statistical Down Scaling Model (SDSM) ~~models were applied~~as they are computationally less demanding and efficient. ~~Six CMIP3 GCMs for LARS WG (CSIRO-MK3, ECHAM5-OM, MRI-CGCM2.3.2, HaDCM3, GFDL-CM2.1, CCSM3) model while HadCM3 GCM and canESM2 from CMIP5 GCMs for SDSM were used for climate change analysis.~~

The downscaled precipitation results from ~~the prediction of~~ the six GCMs by LARS WG showed inconsistency and large inter model variability, two GCMs showed decreasing trend while 4 GCMs showed increasing in the range from -7.9 % to +43.7 % while the ensemble mean of the six GCM result showed increasing trend ranged from 1.0 % to 14.4 %. ~~NCCCS GCM predicted maximum increase in mean annual precipitation. However, the projection from HadCM3 GCM is consistent with the multi model average projection, which predicts precipitation increase from 1.7% to 16.6%.~~ Conversely, the result from all GCMs showed a similar continuous increasing trend for maximum temperature (Tmax) and minimum temperature (Tmin) in all three future time periods. ~~-~~The change for mean annual Tmax may increase from 0.4 ^{o}c to 4.3 ^{o}c whereas the change for mean annual Tmin may increase from 0.3 ^{o}c to 4.1^{o}c

~~Meanwhile~~Even though, both HadCM3 and canESM2 GCMs using SDSM agree with respect to the direction of 21st century ~~the result from SDSM showed an increasing trend for all three climate variables (~~precipitation, minimum and

maximum temperature changes, there are considerable variability in magnitude~~) from both HadCM3 and canESM2 GCMs~~. The relative change of mean annual precipitation downscaled from 5 different scenarios of two GCMs range from 2.1 % to 43.8 % while the change for mean annual Tmax and Tmin may increase from 0.4 ^{o}c to 2.9 ^{o}c and from 0.3 ^{o}c to 1.6 ^{o}c respectively. The change in magnitude for precipitation is higher in RCP8.5 scenarios than the others as expected. The ~~present~~ study results illustrate that both down scaling techniques have shown comparable and good ability to simulate the current local climate variables which can be adopted for future climate change study ~~with high confidence~~ for the UBNRB. However, based on the comparative performance evaluation results of this study, SDSM would be more robust and can be applied at higher confidence for downscaling large scale GCMs outputs to finer scales to suit for hydrological models for impact assessment in the study area of UBNRB. ~~In order to see the comparative downscaling results from the two down scaling techniques, HadCM3 GCM of A2 scenario was used in common. The result obtained from the two down scaling models were found reasonably comparable and both approaches showed increasing trend for precipitation, Tmax and Tmin. However, the analysis of the downscaled climate data from the two techniques showed, LARS WG projected a relatively higher increase than SDSM.~~

Key words: Climate Change, GCM, statistical down scaling, LARS WG, SDSM; UBNRB

1. Introduction

The impacts of climate change on the hydrological cycle in general and on water resources in particular are of high significance due to the fact that all natural and socio/economic system critically depend on water. The direct impact of climate change can be variation and changing pattern of water resources availability and hydrological extreme events such as floods and droughts, with many indirect effects on agriculture, food and energy production and overall water infrastructure (Ebrahim *et al.*, 2013). The impact may be worse on trans-boundary Rivers like Upper Blue Nile River where competition for water is becoming high from different economic, political and social interests of the riparian countries and when runoff variability of upstream countries can greatly affect the downstream countries (Kim, 2008; Semenov and Barrow, 1997).

According to IPCC (2007), between 75 and 250 million people are projected to be exposed to increased water stress due to climate change in Africa by 2020. The increasing water demand of upstream countries in the Nile Basin coupled with climate change impacts can affect the availability of water resources for downstream countries and in the basin, that could result in resource conflicts and regional insecurities. Moreover, climate variability, the way climate fluctuates yearly and seasonally above or below a long-term average value, caused by changes in forcing factors such as variation in seasonal extent of the Inter Tropical Convergence Zone (ITCZ) like El Niño and La Niña events, ~~Current climate variability~~ is already imposing a significant challenge to Ethiopia by affecting food security, water and energy supply, poverty reduction and sustainable socio-economic development efforts. To mitigate these challenges, ~~T~~the Ethiopian government is therefore carried out a series of studies on Upper Blue Nile river Basin (UBNRB), which have been identified as an economic "growth

corridor", focused on identifying irrigation, hydropower potential and the use of the extensive water resources of the basin (BCEOM, 1998; USBR, 1964; WAPCOS, 1990).~~, with less attention for climate change and its impact.~~ As the result, large scale irrigation and hydro-power projects including the Grand Ethiopian Renaissance Dam (GERD), the largest hydroelectric power plant in Africa, ~~has~~ have been ~~constructed~~ identified and ~~is~~ being constructed ~~to mitigate the~~as mitigation measure for the ~~-~~impacts of climate change. However, most studies were given less emphasis for climate change and its impact on the hydrology of the basin, ~~H~~hence, identifying local impacts of climate change at ~~a catchment~~ basin level is quite important especially in UBNRB for the sustainability of large scale water resource development projects, for proper water resource management leading to regional security and looking for the possible mitigation measures otherwise the consequences becoming catastrophic.

To this end, several individual researches have been done to study the impacts of climate change on the water resources of Upper Blue Nile River Basin. Taye *et al.* (2011) reviewed some of the research outputs and concluded that clear discrepancies were observed particularly on the projection of precipitation. For instance, the result obtained from (Bewket and Conway, 2007; Conway, 2000; Gebremicael *et al.*, 2013) showed that there is no significant change on the amount of rainfall and there is no consistent patterns or trends observed. Kim (2008) used the outputs of six GCMs ~~to project~~for the projection of future precipitations and temperature, the result suggested that the changes in mean annual precipitation from the six GCMs range from -11 % to 44 % with a change of 11% from the weighted average scenario at 2050s. On the other hand, the changes in mean annual temperature range from 1.4°C to 2.6^{o}C with a change of 2.3^{o}C from the weighted average scenario. Likewise, Yates and Strzepek (1998a) used 3 GCMs and the result revealed that the changes in precipitation range from -5% to 30% and the change in temperature range from 2.2^{o}c to 3.5^{o}c. Yates and Strzepek (1998b) also used 6 GCMs and the result showed in the range from -9% to 55% for precipitation while temperature increased from 2.2^{o}c to 3.7 ^{o}c. Another study done by Elshamy *et al.* (2009), ~~–~~used 17 GCMs and the result showed that Changes in total annual precipitation range between −15 % to +14 % but the ensemble mean of all models showed almost no change in the annual total rainfall. ~~Moreover~~While, all models predict the temperature to increase between 2^{o}C and 5^{o}C. Gebre and Ludwig (2014), used five biased corrected 50km x 50km spatial resolution GCMs for RCP4.5 and RCP8.5 scenarios to down scale the future climate change of 4 watershed (Gilgel Abay, Gumara, Ribb and Megech) located in Tana sub basin of UBNRB for the time period of 2030s and 2050s. The result suggested that the selected five GCMs disagree on the direction of future prediction of precipitation but multimodal average monthly and seasonal precipitation may generally increases over the watersheds.

For the historical context, the discrepancies could be due to the period and length of data analyzed and the failure to consider stations which can represent the spatial variability of the basin and also errors induced from observed data. For the future context, ~~apart from~~beside the above mentioned reasons, discrepancies could be due to the difference of GCMs and scenarios

used for downscaling, the downscaling techniques applied (can be dynamical and statistical), selection of representative predictors, the period of analysis and spatial and temporal resolution of observed and predictor dataset.

To address uncertainty in projected climate changes, the (IPCC, 2014) thus recommends using a large ensemble of climate change scenarios produced from various combinations of Atmospheric Ocean General Circulation Model (AOGCMs) and forcing scenarios. However, it can become prohibitively time consuming to assess the climate change, using simultaneously many climate change scenarios and many Statistical Down scaling models. As a result, researchers typically assess the climate change and its impacts under only one or a few climate change scenarios selected arbitrarily with no justification for instance (Elshamy *et al.*, 2009; ~~Kim and Kaluarachchi, 2009;~~ Kim, 2008) used only A1B and A2 scenarios respectively. Yet, there is no any hard rule to select an appropriate subset of climate change scenarios among the wide range of possibilities (Casajus et al., 2016).

~~Recently, there is great advancement in Global Circulation Models (GCMs) to represent large scale (global and continental) climate fairly well, However, they often fail to simulate less scale climate features which are required by hydrological models to carry out impact studies, particularly for precipitation (Semenov *et al.*, 1997). To overcome this problem two sets of techniques have emerged as a means to bridge this resolution gap (Fowler *et al.*, 2007; Wilby *et al.*, 2002) by employing either dynamic down scaling or statistical down scaling methods. Even though, statistical down scaling is problematic in producing realistic future climate change scenario because of recognized inter-variable biases in host GCMs, it has practical advantage and is the more promising option in situations where low-cost, rapid assessments of localized climate change impacts are required (Wilby *et al.*, 2002).~~

Although climate models are usually responsible for high uncertainty in climate change impact analysis, the processes of downscaling that ensures to narrow down the scale discrepancy between the coarse scale GCMs and the required local scale climate variables for hydrological models should be investigated for their contribution. Many downscaling models have been developed in the past decade, to bridge the resolution gap between the coarse resolution GCMs and the required local scale for hydrological models to carry out impact studies, ~~(Fowler *et al.*, 2007)~~, however, no single model has been found to perform well over all the regions and time scales. Thus, evaluations of different models are critical to understand the applicability of the existing models (Dibike and Coulibaly, 2005; Ebrahim et al., 2013; Fiseha et al., 2012; Goodarzi et al., 2015; Hashmi et al., 2011; Khan et al., 2006; Qian et al., 2004; Wilby et al., 2004; Wilby and Wigley, 1997; Xu, 1999). However, it remains difficult to directly compare the skill of different downscaling models (Goly et al., 2014).

Apart from the GCMs and downscaling techniques, most of the previous studies e.g (Beyene et al., 2010; Elshamy et al., 2009; Kim, 2008), used CRU, NFS and other gridded data sets constructed based on the interpolation of a few stations in Ethiopia, which has relatively less accuracy as compared with the station based data (Worqlul *et al.*, 2014). Therefore, the objective of this study is to construct and analyze detailed climate change scenarios for precipitation, maximum and minimum temperature over Upper Blue Nile River Basin at required resolution which can be used for further hydrological

impact study. This can be achieved through the inclusion of systematically selected multiple GCMs and two downscaling methods by incorporating acceptable number of weather stations which has long time series and reliable observed climate data to appreciate the uncertainties coming from GCMs and the process of downscaling methods to overcome the short comings of the previous studies on the study area .

~~Therefore, the objective of this study is to analyze and to better comprehend the possible future climate trend of Upper Blue Nile River Basin by applying widely used and more plausible statistical down scaling techniques.~~

~~Recently, there is great advancement in Global Circulation Models (GCMs) to represent large scale (global and continental) climate fairly well, However, they often fail to simulate less scale climate features which are required by hydrological models to carry out impact studies, particularly for precipitation (Semenov *et al.*, 1997). To overcome this problem two sets of techniques have emerged as a means to bridge this resolution gap (Fowler *et al.*, 2007; Wilby *et al.*, 2002) by employing either dynamic down scaling or statistical down scaling methods. Even though, statistical down scaling is problematic in producing realistic future climate change scenario because of recognized inter variable biases in host GCMs, it has practical advantage and is the more promising option in situations where low cost, rapid assessments of localized climate change impacts are required (Wilby *et al.*, 2002).~~

Generally, downscaling methods are classified in to dynamic and statistical downscaling (Fowler *et al.*, 2007; Wilby *et al.*, 2002). Dynamic downscaling nests higher resolution Regional Climate Models(RCMs) into coarse resolution GCMs to produce complete set of meteorological variables which are consistent each other. The outputs from this method is still not at required scale to what the hydrological model require. Statistical downscaling overcomes this challenge moreover it is computationally undemanding, simple to apply and provides the possibility of uncertainty analysis (Dibike *et al.*, 2005; Semenov *et al.*, 1997; Wilby *et al.*, 2002). Extensive details on the strength and weakness of the two methods can be found (Wilby *et al.*, 2007; Wilby *et al.*, 1997). ~~Recently~~Among the different possibilities, two well recognized statistical downscaling tools, ~~are made available to the broader climate change impact study community. The first one implements~~ a regression based ~~method and is referred to as~~ Statistical Down-Scaling Model (SDSM) (Wilby *et al.*, 2002) ~~while the second is~~and a stochastic weather generator called Long Ashton Research Station Weather Generator (LARS-WG) (Semenov *et al.*, 1997; Semenov *et al.*, 2002) were chosen for this study. They have been tested in various regions e.g., (Chen *et al.*, 2013; Dibike *et al.*, 2005; Dile *et al.*, 2013; Elshamy *et al.*, 2009; Fiseha *et al.*, 2012; Hashmi *et al.*, 2011; Hassan *et al.*, 2014; Maurer and Hidalgo, 2008; Yimer *et al.*, 2009) under different climatic conditions of the world. ~~They are most widely used for climate change impact studies (Wilby and Dawson, 2013). Therefore, statistical down scaling technique of LARS WG and SDSM were applied for this study.~~

2. Description of Study Area

The Upper Blue Nile River Basin (UBNRB) extends from 7°45' to 13° N and 34°30' and 37°45' E. It is one of the most important major basin of Ethiopia because it contributes to 45% of the countries surface water resources, 20% of the population, 17% of the landmass, 40% of the nation's agricultural product and large portion of the hydropower and irrigation potential of the country (Elshamy *et al.*, 2009). The whole UBNRB has an area coverage of 199,812 km^2 (BCEOM, 1998). For this study Rahad, Gelegu and Dinder sub catchments that do not flow through the main river stem to Sudan is excluded. The basin area coverage is 176,000km^2 which is about 15% of total area of 1.12 million km^2(Awulachew *et al.*, 2007) of Ethiopia . The elevation ranges between 489 m.a.s.l downstream on the western side to 4261m.a.s.l upstream at Mount Ras Dashen in the north-eastern part.

The Upper Blue Nile River itself has an average annual run-off of about 49 BCM. In addition, the Dinder, Galegu and Rahad rivers have a combined annual run-off of about 5 BCM. The rivers of the Upper Blue Nile River Basin contribute on average about 62 per cent of Nile total at Aswan. Together with contributions of the Baro-Akobo and Tekeze rivers, Ethiopia accounts for 86 per cent of run-off at Aswan (BCEOM, 1998). The climate of Ethiopia is mainly controlled by the seasonal migration of the Inter-tropical Convergence Zone (ITCZ) following the position of the sun relative to the earth and the associated atmospheric circulation. It is also highly influenced by the complex topography. The whole UBNRB has long term mean annual rainfall, minimum and maximum temperature of 1452 $mmyr^{-1}$, 11.4°C and 24.7°C respectively as calculated by this study from 15 rainfall and 25 temperature gauging stations from the period 1984-2011. The mean seasonal rainfall based on the above data showed about 238 mm, 1065 mm, and 148 mm occurred in Belg (October-January), Kiremit (July-September), and Bega (February-May) respectively, in which about 74 % of rainfall concentrates between June and September (Kiremit season).

3. Datasets

3.1 Local data sets

The historical precipitation, maximum and minimum temperature data for the study area were obtained from Ethiopian Meteorological Agency (EMA), which were analyzed and checked for further quality control. A considerable length of time series data were missed in almost all available stations and hence 15 rainfall and 25 temperature stations which have long time series and relatively short time missing records were selected. Filling missed or gap records was the first task for further meteorological data analysis. This task was done using the well-known methodology of inverse distance weighing method (IDW). To check the quality of the data, the Double Mass Curve analysis (DMC) were used. DMC is a cross correlation

between the accumulated totals of the gauge in question against the corresponding totals for a representative group of nearby gauges.

3.2 Large scale datasets

High uncertainty is expected in climate change impact studies if the simulation result is relied up on a single GCM due to the fact that each GCM has different spatial and temporal resolution with different assumptions of atmospheric processes (Kim *et al.*, 2009). Hence, a new version of the LARS-WG5.5 ~~-~~was applied for this study that incorporates predictions from 15 GCMs which were used in the IPCC's Fourth Assessment Report (AR4) based on Special Emissions Scenarios SRES B1, A1B and A2 for three time windows as listed in Table 1~~Table 1~~. ~~At the time of this study~~However, the fifth phase of Coupled Model Inter Comparison Project (CMIP5) climate models based on the new radiative forcing scenarios (Representative Concentration Pathway, RCP) which were used for IPCC Fifth Assessment Report (AR5) were not incorporated in to it at the time of the study. ~~the new version of LARS WG5.5.~~

~~However, a~~As it is difficult to process all the incorporated 15 CMIP3 GCMs and as it is expected large differences in predictions of climate variables among the GCMs, the~~ir~~ performance of GCMs in simulating the current climate variables of ~~in t~~the study area (UBNRB) ~~in particular and for Ethiopia in general were~~ should be evaluated and best represented GCMs were selected. The MAGICC/SCEGEN computer program tool was used for the performance evaluation of the embedded 15 GCMs in LARS WG5.5 database, ~~–~~as it is a standard method for selecting models on the basis of their ability to accurately represent current climate, either for a particular region and/or for the globe.

In this study, we used a semi-quantitative skill score that rewards relatively good models and penalizes relatively bad models as suggested by user manual Wigley (2008). The statistics used for model selection are pattern correlation (R2), Root mean square error (RMSE), bias (B), and a bias-corrected RMSE (RMSE-corr). The analysis was done separately for precipitation and temperature and finally an average score value was taken for model selection. ~~and s~~Six best performed GCMs have been selected for this study namely: HadCM3, GFDL-CM2.1, ECHAM5-OM, CCSM3, MRI-CGCM2.3.2, and CSIRO-MK3 in the order of their performance ~~by MAGICC/SCEGEN computer program tools as suggested by Wigley (2008)~~ to construct future precipitation, maximum and minimum temperature in the UBNRB for the time period of 2030s, 2050s and 2080s under A1B, A2 and B1 scenarios.

Moreover, ~~A~~atmospheric large scale predictor variables used for representing the present condition were obtained from the National Centre for Environmental Prediction (NCEP) reanalysis data set. CanESM2, second generation Canadian Earth System Model (ESM) developed by Canadian Centre for Climate Modelling and Analysis (CCCma) of Environment Canada that represents ~~the IPCC Fifth Assessment Report (AR5)~~CMIP5 and HadCM3~~, the third version of Atmosphere Ocean General Circulation Model (AOGCM)~~ outputs from the Hadley Centre, United Kingdom(UK) representing ~~AR4~~ CMIP3

were used in SDSM for the construction of daily local meteorological variables corresponding to their future climate scenario.

The reasons for selecting these two GCMs were due to the fact that they are models that made daily predictor variables freely available to be directly fed into SDSM covering the study area with a better resolution. Additionally, they are the most used GCMs in previous studies such as (Dibike *et al.*, 2005; Dile *et al.*, 2013; Hassan *et al.*, 2014; Yimer *et al.*, 2009), and HadCM3 ranked first in performance evolution done by MAGICC/SCEGEN computer program tools and its downscaled results match with the ensemble mean of the six GCMs used in LARS WG model. ~~Moreover~~Furthermore, they can represent two different scenario generations describing the amount of green house gases(GHGs) in the atmosphere in the future. HadCM3 GCM used emission scenarios of A2 (separated world scenario) in which the co2 concentration projected to be 414ppm, 545ppm and 754ppm and B2 (the world of technological inequalities) where the co2 concentration to be expected 406ppm, 486ppm and 581ppm at the time period of 2020s, 2050s and 2080s respectively(Semenov and Stratonovitch, 2010) that were used in the CMIP3 for the IPCC's AR4 (IPCC, 2007). While canESM2 ~~GCM~~ represents the latest and wide range of plausible radiative forcing scenarios, which include a very low forcing level (RCP2.6), where radiative forcing peaks at approximately 3 Wm^{-2}, approximately 490 ppm co2 equivalent before 2100 and then decline to 2.6Wm^{-2}; two medium stabilization scenarios (RCP6 and RCP 4.5) in which radiative forcing is stabilised at 6Wm^{-2} (approximately 850 ppm co2 equivalent) and 4.5 Wm^{-2} (approximately 650 ppm co2 equivalent) after 2100 respectively, and one very high baseline emission scenario (RCP8.5) for which radiative forcing reaches >8.5 Wm^{-2} (1370 ppm co2 equivalent) by 2100 and continues to rise for some time ~~(RCP8.5)~~ that were used for the IPCC's AR5, (IPCC, 2014).

The NCEP dataset were normalized over the complete 1961-1990 period data, and interpolated to the same grid as HadCM3 (2.5° latitude x 3.75° longitude) and canESM2 (2.8125° latitude x 2.8125° longitude) from its horizontal resolution of (2.5° latitude x 2.5° longitude), to represent the current climate conditions. NCEP reanalysis data were normalized and interpolated as (Hassan *et al.*, 2014):

$$un = \frac{(ut-ua)}{\sigma u} \quad \text{.....} \quad (11)$$

In which *un* is the normalized atmospheric variable at time t, *ut* is the original data at time t, *ua* is the multiyear average during the period, and σu is the standard deviation.

The canESM2 outputs ~~were downloaded~~ for three different climate scenarios namely: RCP 2.6, RCP 4.5 and RCP 8.5 for the period ~~1961~~2006-~~2099~~ 2100 while the outputs of HadCM3 ~~were~~ for A2a (medium-high) and B2a (medium-low) emission scenarios ~~of the IPCC Special Report on Emission Scenarios~~ for the period 1961-~~2100~~2099~~. The outputs of the models~~ were obtained on a grid by grid box basis for the study area from the Environment Canada website http://ccds-

dscc.ec.gc.ca/index.php?page=dst-sdi (the "a" in A2a and B2a refers the ensemble member in the HadCM3 A2 and B2 experiments). The archive of canESM2 and HadCM3 GCM output contains 26 daily predictor variables each as listed in Table 3~~Table 3~~.

4. Methodology

4.1 Description of LARS-WG Model

LARS-WG is a stochastic weather generator which can be used for the simulation of weather data at a single station under both current and future climate conditions. These data are in the form of daily time-series for a group of climate variables, namely, precipitation, maximum and minimum temperature and solar radiation (Chen *et al.*, 2013; Semenov *et al.*, 1997). LARS-WG uses a semi-empirical distribution (SED) that is defined as the cumulative probability distribution function(CDF) to approximate probability distributions of dry and wet series, daily precipitation, minimum and maximum temperatures.

$$\boldsymbol{EPM} = \{\boldsymbol{a_0}, \boldsymbol{a_i}, \boldsymbol{h_i}, \boldsymbol{i} = \boldsymbol{0}, \dots, \boldsymbol{10}\} \quad (2~~2~~)$$

EPM is a histogram of the distribution of 23 different intervals (a_{i-1}, a_i) where $a_{i-1} < a_i$ (Semenov et al., 2002), which offers more accurate representation of the observed distribution compared with the 10 used in the previous version. By perturbing parameters of distributions for a site with the predicted changes of climate derived from global or regional climate models, a daily climate scenario for this site could be generated and used in conjunction with a process-based impact model for assessment of impacts.

In general, the process of generating synthetic weather data can be categorized in three distinct steps: model calibration, model validation and scenario generation as represented in Figure 2~~Figure 2~~ (a), which are briefly described by (Semenov *et al.*, 2002) as follows.

The inputs to the weather generator are the series of daily observed data (precipitation, minimum and maximum temperature) of the base period (1984-2011)and site information (latitude, longitude and altitude) are the inputs to the LARSWG. After the input data preparation and quality control, the observed daily weather data at a given site were used to determine a set of parameters for probability distributions of weather variables. These parameters are used to generate a synthetic weather time series of arbitrary length by randomly selecting values from the appropriate distributions, having the same statistical characteristics as the original observed data but differing on a day-to-day basis . The LARS WG distinguishes wet days from dry days based on whether the precipitation is greater than zero. The occurrence of precipitation is modelled by alternating wet and dry series approximated by semi empirical probability distributions. The statistical characteristics of the observed

and synthetic weather data are analyzed to determine if there are any statistically-significant differences using Chi-square goodness of fit test (KS) and the means and standard deviation using t and F test respectively by changing the parameters of LARS-WG (number of years and seed number).

To generate climate scenarios at a site for a certain future period and an emission scenario, the LARS-WG baseline parameters, which are calculated from observed weather for a baseline period (1984-2011), are adjusted by the Δ-changes for the future period and the emissions predicted by a GCM for each climatic variable for the grid covering the site. In this study, the local-scale climate scenarios based on the SRES A2, A1B and B1 scenario simulated by the selected six GCMs are generated for the time periods of 2011–2030, 2046–2065, and 2080–2099 to predict the future change of precipitation and temperature in UBNRB.

Δ-changes were calculated as relative changes for precipitation and absolute changes for minimum and maximum temperatures (Eq. 3 and 4),respectively. No adjustments for distributions of dry and wet series and temperature variability were made, because this would require daily output from the GCMs which is not readily available from LARS WG data set (Semenov *et al.*, 2010).

$$\Delta \boldsymbol{T_i} = \left(\overline{\boldsymbol{T}}_{\boldsymbol{GCM,FUT,i}} - \overline{\boldsymbol{T}}_{\boldsymbol{synt,Base,i}}\right) \quad (3)$$

$$\Delta \boldsymbol{P_i} = \left(\overline{\boldsymbol{P}}_{\boldsymbol{GCM,FUT,i}} \Big/ \overline{\boldsymbol{P}}_{\boldsymbol{synt,Base,i}} \right) \quad (4)$$

In above equations, ΔT_i and ΔP_i are climate change scenarios of the temperature and precipitation, respectively, for long-term average for each month ($1 \leq i \leq 12$); $\overline{T}_{GCM,FUT,I}$ the long term average temperature simulated by the AOGCM in the future periods per month for three time periods; $\overline{T}_{Synth,Base,i}$ is the long term average temperature simulated by the model in the period similar to observation period (in this study 1984-2011) for each month. The above calculations are true for precipitation as well.

For obtaining time series of future climate scenarios, climate change scenarios are added to the observations values by employing the change factor (CF) method (Eq. 5 and 6) (in this study 1984-2011):

$$\boldsymbol{T} = \boldsymbol{T_{obs}} + \Delta \boldsymbol{T} \quad \textbf{(5)}$$

$$\boldsymbol{P} = \boldsymbol{P_{obs}} + \Delta \boldsymbol{P} \quad \textbf{(6)}$$

T and P; time series of the future climate scenarios of temperature and precipitation (2011-2100) and T_{obs} and P_{obs} ; observed temperature and precipitation. So, in LARS-WG downscaling unlike SDSM, large-scale atmospheric variables are not directly used in the model, rather, based on the relative mean monthly changes between current and future periods predicted by a GCM, local station climate variables are adjusted proportionately to represent climate change (Dibike *et al.*, 2005).

4.2 Description of SDSM

The SDSM is best described as a hybrid of the stochastic weather generator and regression based in the family of transfer function methods~~. D~~due to the fact that a multiple linear regression model is developed between a few selected large-scale predictor variables and local-scale predictands such as temperature and precipitation to condition local scale weather parameters from large scale circulation patterns. The stochastic component of SDSM enables the generation of multiple simulations with slightly different time series attributes, but the same overall statistical properties. (Wilby *et al.*, 2002) . It requires two types of daily data, the first type corresponds to local predictands of interest (e.g. temperature, precipitation) and the second type corresponds to the data of large-scale predictors (NCEP and GCM) of a grid box closest to the station.

The SDSM model categorizes the task of downscaling into a series of discrete processes such as quality control and data transformation, screening of predictor variables, model calibration and weather and scenario generation as shown in Figure 2~~Figure 2~~(b). ~~-~~Detail procedures and steps can be found (Wilby *et al.*, 2002) for further reading. Screening potentially useful predictor-predictand relationships for model calibration is one of the most challenging but very crucial stage in the development of any statistical down scaling model. It is because of the fact that the selection of appropriate predictor variables largely determines the success of SDSM and also the character of the downscaled climate scenario (Wilby *et al.*, 2007). After routine screening procedures, the predictor variables that provide physically sensible meaning in terms of their high explained variance, correlation coefficient (r) and the magnitude of their probability (p value) were selected.

The model calibration process in SDSM was used to construct downscaled data based on multiple regression equations given daily weather data (predictand) and the selected predictor variables. The model was structured as monthly model for both daily precipitation and temperature downscaling. Consequently, twelve regression equations were developed for twelve months. Bias correction and variance inflation factor was adjusted until the model replicate the observed data. The weather generator helps to validate the calibrated model ideally using independent data. This operation generates the ensembles of synthetic daily weather data for the specified period with the help of regression model weights along with parameter file prepared during model calibration. To compare the observed and simulated data, SDSM has provided summary statistics function that summarizes the result of both the observed and simulated data. Time series of station data and large scale predictor variable information (NCEP reanalysis data) were divided into two groups; for the period from 1984-1995/ 1984-2000 and 1996-2001/ 2001-2005 for model calibration and validation of HadCM3/canESM2 GCMs respectively.

The Scenario Generator operation produces ensembles of synthetic daily weather series given observed daily atmospheric predictor variables supplied by a GCM either for current or future climate(Wilby *et al.*, 2002). The scenario generation produced 20 ensemble members of synthetic weather data for 139 years (1961-2099) from HadCM3 A2a and B2a scenarios and for 95 years (2006-2100) from canESM2 for RCP2.6, 4.5 and 8.5 scenarios, and the mean of the ensemble members was calculated and used for further analysis. The generated scenario was divided into three time windows of 30 years of data (2011-2040), (2041-2070) and (2071-2100) henceforth called 2030s, 2050s and 2080s, respectively.

~~5. Results and Analysis~~

~~5.1~~4.3 Model performance evaluation criteria

A simulation of mean daily and monthly rainfall, Tmax and Tmin, during the calibration and validation of the SDSM and LARSWG time series were checked by using graphical representation and statistical performance index. Performance indicators such as mean absolute error (MAE), root mean square error RMSE), Bias (B), coefficient of determination (R^2), NasheSutcliffe Model Efficiency (NSE) were used to evaluate the performance of the GCMs and downscaling techniques reproducing the present climate variables of the study area and are defined as;

$$R^2 = \frac{\left[\sum_{i=1}^{n}(X_i-\mu_x)(Y_i-\mu_x)\right]^2}{\sum_{i=1}^{n}(X_i-\mu_x)^2\sum_{i=1}^{n}(Y_i-\mu_x)^2} \quad \text{..... (7~~7~~)}$$

$$MAE = \frac{\sum_{i=1}^{n}|X_i-Y_i|}{n} \quad \text{..... (8~~8~~)}$$

$$RMSE = \sqrt{\sum_{i=1}^{n}(X_i-Y_i)^2} \quad \text{..... (9~~9~~)}$$

$$NSE = 1 - \frac{\frac{1}{n}\sum_{I=1}^{I=n}(X_i-Y_i)^2}{\frac{1}{n}\sum_{i=1}^{i=n}(X_i-\mu_x)^2} \quad \text{..... (10~~10~~)}$$

$$Bias = \frac{\sum_{i01}^{i=n}X_i}{n} - \frac{\sum_{i=1}^{i=n}Y_i}{n} \quad \text{..... (11~~11~~)}$$

$$Relative\ Bias\ (\%) = \frac{1}{n}\sum_{i=1}^{i=n}\frac{(X_i-Y_i)}{X_i} \quad \text{..... (12~~12~~)}$$

In the above equations X_i and Y_i are i-th observation and simulated data by the model, respectively. μ_x and μ_y are the average of all data of X_i and Y_i in the study population and n is the number of all samples to be tested.

5. Results and Analysis

5. ~~2~~ 1 Calibration and validation of LARS-WG

To verify the performance of LARS-WG, in addition to the graphic comparison, some statistical tests were performed. The Kolmogorov–Smirnov (KS) test is performed to test equality of the seasonal distributions of wet and dry series (WDSeries), distributions of daily rainfall (RainD), and distributions of daily maximum (TmaxD) and minimum (TminD) temperature. The F-test is performed on testing equality of monthly variances of precipitation (RMV) while the t test is performed on verifying equality of monthly mean rainfall (RMM), monthly mean of daily maximum temperature (TmaxM), and monthly mean of daily minimum temperature (TminM). ~~calculated from observed and generated data~~. All of the tests calculate a p-value, which is used to accept or reject the hypotheses that the two sets of data (observed and generated) could have come from the same distribution at the 5% significance level . Therefore, number of tests that results a p value less than 5% out of the total number of 8 dry/wet seasons or 12months were recorded for each stations. The average number of P values less than 5% recorded from 26 stations and percentage failed from the total of 8 seasons or 12 months ~~The test results have~~ has been presented in Table 2~~Table 2,. where the numbers show how many tests gave significant different results at the 5% significance level out of the total number of tests of 8 seasons or 12months. A large number indicates a poor performance.~~ It can be seen from Table 2 that LARS WG performs very well for all parameters except RMM and RMV. On the other hand, LARS WG performs poor (i.e. average of 2.2 % and 17.3 % of the months of a year were recorded a P value < 5 %) for the monthly mean and variance of precipitation respectively. ~~It can be seen from Table 2 that the average number of significant different results for seasonal wet and dry series distributions and for the daily rainfall distributions (RainD) were 0 and 1.67 out of 8 and 12 respectively; for the monthly means (RMM) is 0.3 and for the monthly variance (RMV) is 2.1out of 12. The average numbers of significant results for TminD, TminM, TmaxD, and TmaxM are zero or close to zero.~~ From these numbers, it can be noted that the model is ~~more~~ less capable in simulating ~~the monthly means and the daily rainfall distributions of each month in comparison to~~ the monthly variances than the other parameters.

For illustrative purpose, graphical representation of monthly mean and standard deviation of the simulated and observed precipitation, Tmax and Tmin were constructed in Figure 3~~Figure 3~~ for randomly chosen Gondar station as it has been difficult to present the result of all stations. It can be seen from Figure 3~~Figure 3~~ that observed and simulated monthly mean precipitation, Tmax and Tmin matches very well. However, as it is known for being difficult to simulate the standard deviations in most statistical downscaling studies, the performance of the standard deviation is less accurate as compared to the mean (Figure 3~~Figure 3~~ b) . Generally, according to the obtained statistical performance measure values and from graphical representation, the performance of the model for simulation ~~and prediction~~ of the three climatic variables in all stations across UBNRB is acceptable and reasonably well.

5.3 2 Down scaling with LARS-WG

The result of precipitation prediction by using LARS-WG model from six multi model GCMs under three scenarios (A1B, B1 and A2) for three time periods were presented in Table 5 and plotted in Figure 4Figure 4 for illustrative purpose. In Figure 4Figure 4, each box–whisker plot represents the prediction of precipitation across all stations of UBNRB under a single scenario for each GCM and the result revealed that there are no coherent change trends observed among various GCMs' for predicting precipitation. NCCCSM GCM was found the most unstable GCM in predicting precipitation across UBNRB stations particularly under A2 scenario at the time period of 2080s while MPEH5 was relatively stable across all stations as compared to others.

After downscaling the future climate predictions at all stations from the selected six GCMs, the projected precipitation analysis for the areal UBNRB was calculated from the point rainfall stations using Thiessen polygon method. The result analysis revealed that, GCMs disagree on the direction of precipitation change, two GCMs (MIHR and GFCM21) result decreasing trend whereas a majority or four GCMs (NCCSM, Hadcm3, MPEH5 and MIHR) result increasing trend from the reference period in all three time periods. The results from Figure 5Table 5 showed that NCCCS reported maximum increase while GFCM21 reported highest reduction. From the model average, B1 scenario projected maximum increase of mean annual relative change of precipitation at 2030s and 2050s while A2 scenario for 2080s. For 2030s, the relative change of mean annual precipitation projected between (-2.3 % and + 6.5 %) for A1B, (-2.3 % and +7.8 %) for B1 and (-3.7 % and +6.4 %) for A2 emission scenarios. At 2050s, the relative change in precipitation range from (-8 % and +22.7 %) for A1B, (-2.7 % and +22 %) for B1 and (-7.4 % and +8.7 %) for A2 scenarios. In the time of 2080s, the relative change of precipitation projected may vary from(-7.5 % and +29.9 %) for A1B, (-5.3 % and +13.7 %) for B1 and (-5.9 % and +43.8 %) for A2 emission scenarios. The multi model average relative change mean annual precipitation result showed that in the future precipitation may generally increases over the basin in the range of 1 %-14.4 % which is consistent matches with the result from HadCM3 GCM (0.8 %-16.6 %) as it is shown in Figure 5Table 5 .

In a different way from precipitation, the projection of mean annual Tmax and Tmin have coherent increasing change trends were observed from all the six GCMs under all scenarios in all three future time periods. At 2080s, for A1b, B1 and A2 scenarios, the change in mean annual Tmax and Tmin is more pronounced than 2030s in all GCMs from three scenarios. At 2080s, 2050s and 2030s, the mean annual Tmax may increase up to +4.3^{o}c, +2.1^{o}c and +0.7^{o}c for A2, A1B and A1B scenarios respectively. The result calculated from the ensemble mean showed that mean annual Tmax my increase up to +0.5 ^{o}c, +1.8 ^{o}c and +3.6 ^{o}c by 2030s, 2050s and 2080s respectively under A2 scenario which is consistent in line with the results of from both GFCM21 and HadCM3 GCMs (Figure 5). Likewise, UBNRB may experience an increase mean annual Tmin may increase up to +4.1^{o}c, +1.9^{o}c and +0.7^{o}c at 2080s, 2050s and 2030s respectively whereas the result calculated from the

~~multi model average showed that mean annual Tmin change may reach up to~~ +0.6 °c, +1.8 °c and +3.6 °c by 2030s, 2050s and 2080s respectively from the multi model average.

5. ~~4~~ 3 Screening variable, model calibration and validation of SDSM

Initially, offline correlation analysis was performed using SPSS software between predictands and NCEP re-analysis predictors to identify an optimal lag and physically sensible predictors for climate variables of precipitation, Tmax and Tmin. Analysis of the offline correlation revealed that ~~in most stations and predictors~~ an optimal lag or time shift does not improve the correlation of predictands and predictors for this particular study. Average partial correlation of precipitation of all stations with predictors is shown in the Figure 7~~Figure 5~~ which indicates all stations followed the same correlation pattern ~~with predictors~~ (both in magnitude and direction) that illustrates all stations can have identical physically sensible predictors with a few of exceptions. Furthermore, there exist a number of predictors that have correlation coefficient values in the range of 20 %-45 % for precipitation across all stations as shown in Figure 7~~Figure 5~~. This range is considered to be acceptable when dealing with precipitation downscaling (Wilby *et al.*, 2002) because of its complexity and high spatial and temporal variability to downscale.

~~The predictands of all stations and NCEP reanalysis predictors close to or within the stations are then used in SDSM for the final analysis.~~ The predictor variables identified for each downscaling GCMs and for the corresponding local climate variables conducted in this study are summarized in Figure 8~~Figure 6~~. From the selected predictors, it is observed that different large scale atmospheric variables control different local variables. For instance, set of temp, mslp, s500, s850, p8_v, p500, shum are the most potential or meaningful predictors for temperature and s500, s850, p8_u, p_z, pzh, p500 for precipitation of the study area respectively, which is consistent with the result of offline correlation analysis.

The graphical comparison between the observed and generated rainfall, Tmax and Tmin were run to enhance the confidence of the model performance, as shown in Figure 9~~Figure 7~~ and Figure 10~~Figure 8~~ for Gondar station only. Examination of Figure 9~~Figure 7~~ showed that the calibrated model reproduces the monthly mean precipitation and mean and standard deviation of daily Tmax, Tmin, and mean dry-spell length values quite well. However, the wet-spell length were consistently underestimated and also less accurate in reproducing variance of observed precipitation. As Wilby *et al.* (2004) point out, downscaling models are often regarded as less able to model the variance of the observed precipitation with great accuracy.

Furthermore, the performance of the model was evaluated by statistical performance indicators of (MAE, RMSE, R^2, NSE and BIAS) as summarized in Table 4~~Table 4~~. The result of statistical analysis revealed that the model is much better in simulating Tmax and Tmin than precipitation, because of the high dynamical properties of precipitation makes it difficult to simulate. After accomplishing a satisfactory calibration **(Figure 9)**, the multiple regression model is validated using an

independent set of data outside the period for which the model is calibrated as discussed under section 4, and the results obtained are shown in Figure 9~~Figure 7~~ and Table 4~~Table 4~~. Examination of ~~the~~ Figure 9~~Figure 7~~, Figure 10~~Figure 8~~ and Table 4~~Table 4~~ revealed that the model is successfully validated but at lesser accuracy as compared to calibration for both GCMs. In general, the result analysis of performance measure and graphical representation of observed and simulated both for calibration and validation revealed that the model performs very well in simulating the climate variables with high degree of accuracy.

5.~~5~~ 4 Down scaling with SDSM

Results of down scaling future climate scenario of areal UBNRB using SDSM calculated from all stations using Thiessen polygon methods are summarized from Figure 6~~Table 6~~. The magnitude of future climate change at each station has different pattern and magnitude using different scenarios as can be seen the variation in Figure 11~~Figure 10 and 12~~. The overall analysis of the result of the whole UBNRB from Figure 6 ~~the~~ ~~table~~ indicates, a general increase in mean annual precipitation for three time windows (2030s, 2050s and 2080s) under all scenarios in the range of 2.1 % to 43.8 %~~ under the A2a and RCP8.5 scenarios respectively~~. ~~At 2080s, t~~The maximum/~~-~~minimum relative change of mean annual precipitation is projected to be 43.8 %/2.1 %, 29.5 %/3.5 % and 19 %/2.1% at 2080s, 2050s and 2030s under RCP8.5 scenario of canESM2 ~~GCM while the minimum relative change of mean annual precipitation to be~~ and ~~2.1% under~~ H3B2a scenario of HadCM3 ~~GCM. At 2050s, the maximum and minimum relative mean annual relative change to be 29.5% and 3.5% under RCP8.5 of canESM2 GCM and H3B2a of HadCM3 GCM scenarios respectively. Mean while, at 2030s, the maximum and minimum relative change of mean annual precipitation projected to be 19% and 2.1% under RCP8.5 of canESM2 and H3B2a of HadCM3 scenarios~~ respectively ~~respectively.~~ ~~In~~In general, RCP8.5 scenario of canESM2~~-~~GCM resulted pronounced increase in all three time periods whereas scenario B2a of HadCM3 GCM reported minimum change over the study area.

~~The result analysis at monthly basis as shown from Figure 9 revealed that, canESM2 GCM reported increasing mean monthly precipitation in all months of the year except January, February and March whereas HadCM3 result showed both increasing and decreasing pattern over the whole months of a year. Mean monthly precipitation may increase with a maximum value of 214.5% in the month of October for canESM2 under RCP8.5 by 2080s and may reduce at maximum of 26.6% in the month of February under scenario RCP2.6 by 2080s. The result also indicates, HadCM3 GCM predicted relatively small increase as compared to canESM2, with maximum value of 46.6% in the month of November under scenario A2a by 2020s and maximum reduction in the month of May by 30.4% under scenario A2a by 2080s. Furthermore, the spatial variability analysis of the result showed, the Eastern part may get a more pronounced increase in precipitation compared to the Western part of the study area (Figure 11). Seasonally, both HadCM3 and canESM2 GCM reported precipitation may increase in Summer and Autumn in the range of 1.3% to 27.6% and 8.4% to 89% respectively. However, result from HadCM3 showed reduction of precipitation in spring across the three time windows with a value range from~~

~~10.5% to -21.7%. Also, canESM2 reported precipitation reduction in winter season in the range of -0.9% to -6.1% as summarized in Figure 9.~~

Regarding to temperature, the down scaling result of Tmax and Tmin showed increasing trend consistently in all months, seasons in three time periods under all scenarios with mean annual value ranges from 0.5 ^{o}C to 2.6 ^{o}C and 0.3 ^{o}c to 1.6 ^{o}C under scenario RCP8.5 and H3B2a respectively. RCP 8.5 scenario reported maximum change while H3B2a scenario reported minimum change both for Tmax and Tmin in all three time periods as compared to other scenarios. The analysis of down scaling result illustrates maximum temperature may become much hotter as compared to minimum temperature in all scenarios and time periods in the future across UBNRB.

5.~~6~~ 5 Comparative performance ~~and downscaling results~~evaluation of LARS-WG and SDSM

Chen *et al.* (2013) argued that though major source of uncertainty are linked to GCMs and emission scenarios, uncertainty related to the choice of downscaling methods give less attention on climate change analysis. Therefore, in this study, comparative performance evaluation of the downscaling methods was done and ~~T~~the average partial Pearson's correlation coefficient (R) values of the observed and simulated for all stations ~~as~~ presented in Figure 12~~Figure 12,~~ F~~f~~or precipitation, at daily time series R value was 0.21 in LARS WG while 0.43/042 for HadCM3/canESM2 using SDSM. ~~for HadCM3 and 0.42 for canESM2~~ W~~w~~hereas, R value for daily Tmax were 0.61 using LARS WG and 0.75/~~ and~~ 0.76 ~~using SDSM~~ for HadCM3/~~ and~~ canESM2 using SDSM respectively. ~~–~~The R value for precipitation at monthly basis has improved significantly to 0.79 using LARS WG while 0.84 for both HadCM3 and canESM2 using SDSM. ~~and f~~For Tmax R value was 0.89 using LARS WG and 0.91/~~ and~~ 0.92 for HadCM3/~~ and~~ canESM2 using SDSM respectively. ~~In general, t~~The result from the two downscaling models suggested that both SDSM and LARS-WG approximate the observed climate data corresponding to the current state reasonably well.

~~However, LARS-WG under estimated the standard deviations of the three climatic variables largely for most of the months of the year (poor at modelling inter annual variability), and less performance in simulating daily time series of climate variables as compared to SDSM.~~

For future simulation, HadCM3 GCM under A2 scenario was used in common for two (LARS WG and SDSM) down scaling methods for comparison purpose. The results obtained from the two down scaling models were found reasonably comparable and both approaches showed increasing trend for precipitation, Tmax and Tmin. However, the magnitude ~~analysis~~ of the downscaled climate data from the two techniques as presented in Figure 14~~Figure 13~~ showed different.~~,~~ LARS-WG ~~resulted in a relatively higher increase~~over predicts than SDSM~~,~~. ~~which does not lead to identical conclusions~~ . LARS-WG predict relative change of mean annual precipitation about 16.1 % and an average increase in mean annual Tmax and Tmin about 3.7 ^{o}C and 3.6 ^{o}C respectively at 2080s while SDSM predicts relative change in mean annual precipitation

only about 9.7 % and an average increase in Tmax and Tmin about 2 °c and 1.3 °C respectively in the same period. The difference ~~in the down scaled climate variables~~ could be due to the fact that SDSM uses large scale predictor variables from GCM outputs to predict local scale climate variables using multiple linear regression, while the LARS WG is analysed by applying the change factors from the GCM output of only those variables which directly correspond to the predictands. Moreover, because of the well known fact that GCMs are not very reliable in simulating precipitation, the error induced from the GCM output for precipitation will propagate the error of downscaling that makes the performance of LARS-WG to downscale precipitation more questionable (Dibike et al., 2005). Therefore, based on the above facts SDSM would be more robust and can be applied at higher confidence for downscaling large scale GCMs outputs to finer scales to suit for hydrological models for impact assessment in the UBNRB.

6. Discussions and conclusions

The uncertainty related to climate change analysis can be due to climate models and downscaling methods among many others. In this study, we employed multi model approach to see the uncertainties came from different GCMs. In total, 21 systematically selected future climate scenarios were produced for each time period, which we might think representative to understand fully and to project plausibly the future climate change in the study area and to retain information about the full variability of GCMs. Moreover, ~~W~~we applied two widely used statistical down scaling methods, namely the regression downscaling technique (SDSM) and the stochastic weather generation method (LARS WG) for this particular study.

The down scaling result ~~analysis~~ reported from the six GCMs used in LARS-WG showed large inter model differences, 2 GCMs reported precipitation may decrease while 4 GCMs reported precipitation may increase in the future. The large inter model differences of the GCMs ~~on the direction of future precipitation~~ showed the uncertainties of GCMs associated with their differences of resolution and assumptions of physical atmospheric processes to represent local scale climate variables which are typical characteristics for Africa and because of low convergence in climate model projections in the area of UBNRB (Gebre *et al.*, 2014). ~~In 2030s, the relative change in mean annual precipitation projected may vary from -3.7% to 7.8%. At 2050s and 2080s, the relative change in mean annual precipitation projected between -8% to +22.7% and -7.5% to +43.8% respectively.~~ The multi model average result showed that in the future precipitation may generally increases over the basin in the range of 1 %-14.4 % which agreed with the result from HadCM3 GCM (0.8 %-16.6 %), ~~However, the multi model average of the six GCMs showed increasing pattern for precipitation, Tmax and Tmin and the magnitude of the projection which has better agreement with HadCM3 GCM projection which~~this indicates HadCM3from CMIP3 GCMs has a better representation of local scale climate variables in the study area consistent with the previous study result by Kim *et al.* (2009) and (Dile *et al.*, 2013) in the same study area.

~~Furthermore~~ Further uncertainty analysis of~~,~~ HadCM3 GCM from CMIP3 ~~applied in IPCCAR4~~ and canESM2 GCM from CMIP5 ~~applied in IPCC AR5 were~~ used by SDSM was carried out for precipitation. The downscaled results from the two GCMs modelled by SDSM suggested that mean annual precipitation may generally increase in the range of 2.1 % to 43.8 %. ~~For comparison, the performance indicator result presented in Table 4 has shown, the ratio of MAE relative to mean for daily precipitation was 9.7% and 5.3% while for monthly precipitation was 9.1% and 4.8% respectively for HadCM3 and canESM2 GCMs. The ratio of RMSE relative to mean for daily precipitation was 12.6 % and 8.6 % while for monthly precipitation it was 13% and 7.3% respectively for HadCM3 and canESM2 GCMs during calibration period. During validation period these values were reported 25.2% /22.9% and 26.6% /19.1% for the ratio of MAE to mean for daily and monthly precipitation using HadCM3/canESM2 GCMs respectively whereas 30.7%/ 30% and 38.1%/27.4% for the ratio of RMSE to mean. The ratio of MAE and RMSE to mean is less than 5% both for daily and monthly Tmax and Tmin using the mentioned two GCMs by SDSM techniques. In general, the result suggested that~~ However, canESM2 better performs than HadCM3 in reproducing the current climate variables of UBNRB both in calibration and validation consistently (Table 4).~~.~~ The better performance of canESM2 could be due to the fact that modelling was based on the new set of radiative forcing scenario (RCP) that replaced SRES emission scenarios, constructed for IPCC AR5 where the impacts of land use and land cover change on the environment and climate is explicitly included. Also, it is one of the earth system models which has additional features that incorporates various important biogeochemical processes which is the limitation of CMIP3 GCMs like HadCM3. ~~Further performance statistical analysis revealed that e~~Even though, the simulation of large scale precipitation has improved since IPCC AR4, GCMs still continues to perform less well for precipitation as compared to temperature and therefore downscaling of precipitation becomes more complex and difficult to reproduce the base scenario as compared to downscaling of temperature (Fowler et al., 2007) also confirmed in this study (Table 4). However, a direct comparison between the projection from the two datasets (HadCM3 and canESM2) is not possible as seen from ~~Table 6~~Figure 6, as they used different scenarios describing the amount of Green House Gases (GHGs) in the atmosphere differently.

LARS WG as it is a stochastic simulation tool that is commonly used to produce synthetic climate data of any length with the same characteristics as the input record, it simulate weather separately for single sites; therefore, the resulting weather series for different sites are independent of each other, which can lost a very strong spatial correlation that exists in real weather data during simulation. Although, a few stochastic models have been developed to produce weather series simultaneously at multiple sites preserving the spatial correlation, mainly for daily precipitation, such as space–time models, non-homogeneous hidden Markov model and nonparametric models typically use a K-Nearest Neighbour (K-NN) procedure (King et al., 2015), they are complicated in both calibration and implementation and are unable to adequately reproduce the observed correlations (Khalili et al., 2007).

To test the capability of LARS WG spatial correlation of stations, the simple Pearson's correlation coefficient (R^2) value was calculated for two stations Abaysheleko and Bahirdar and the result from Figure **13** revealed that the spatial correlation of

the stations was distorted /decreased/ from the original to a lesser extent as expected. Even though, LARS WG has limitation to preserve the spatial correlation of climate variables, it can be applied for downscaling climate change scenario for the area of UBNRB satisfactorily with caution to hydrological impact models, as spatial distribution of precipitation may have essential effects on the discharge of a river and the formation of floods.

In conclusion, a multi model average result of this study showed a general increasing trend for all three climatic variables (precipitation, Tmax and Tmin) in all three time periods applying ~~multi model average from~~ LARS-WG and SDSM downscaling techniques. The positive change of precipitation in future can be a good opportunity for the farmers who are engaged in rain fed agriculture to maximize their agricultural production and to change their lively hoods. However, this information cannot be a guarantee for irrigation farming because precipitation is not the only factor contributing to affect the flow of the river, which is the main source ~~of~~ for irrigation. Evapotranspiration, dynamics of land use land cover, proper water resource management and other climatic factors, which are not yet assessed by this study can influence the flow of the river directly and indirectly. Furthermore, the result from this study ~~Figure 9 and~~ (Figure 11~~Figure 10 (only RCP4.5 at 2050s)~~ revealed that, maximum positive precipitation change may occur in Autumn (Sep.-Nov.) when most agricultural crops get matured and start harvesting~~)~~ while minimum precipitation change may occur during summer (June-August), when about 80% of the annual rainfall occurred, this climate variability can be potential threat for the farmers, who have limited ability to cope with the negative impacts of climate variability ~~agricultural production~~ and overall ongoing economic development efforts in the basin ~~as the crops do not require much rain when they are matured~~.

In general, ~~the results of future climate change from multi model GCMs and applying two widely used downscaling techniques for all three climatic variables~~this study ~~have~~ has shown that climate change will occur plausibly that may affect the water resources and hydrology of the UBNRB, and the study proposed ~~so~~ the outputs of canESM2 ~~GCMs~~ESM with new sets of emission scenarios downscaled by SDSM technique can be applied for further impact analysis with high degree of certainty. ~~–~~ Moreover, the paper provides information that the choice of downscaling methods has a contribution in the estimation of climate change and has to be done in the same way as we consider multiple GCMs to acknowledge the uncertainty range.

References

Awulachew, S.B. *et al.*, 2007. Water resources and irrigation development in Ethiopia, 78p. (Working Paper 123). Colombo, Sri Lanka: International Water Management Institute.

BCEOM, 1998. Abbay river basin integrated development master plan, section II, volume V-water resources development. Ministry of Water Resources, Addis Ababa, Ethiopia(part 1- irrigation and drainage).

Bewket, W., Conway, D., 2007. A note on the temporal and spatial variability of rainfall in the drought-prone Amhara region of Ethiopia. International Journal of Climatology, 27(11): 1467-1477.

Beyene, T., Lettenmaier, D.P., Kabat, P., 2010. Hydrologic impacts of climate change on the Nile River Basin: implications of the 2007 IPCC scenarios. Climatic change, 100(3-4): 433-461.

Casajus, N. *et al.*, 2016. An Objective Approach to Select Climate Scenarios when Projecting Species Distribution under Climate Change. PloS one, 11(3): e0152495.

Chen, H., Guo, J., Zhang, Z., Xu, C.-Y., 2013. Prediction of temperature and precipitation in Sudan and South Sudan by using LARS-WG in future. Theoretical and Applied Climatology, 113(3-4): 363-375.

Conway, D., 2000. The climate and hydrology of the Upper Blue Nile River. The Geographical Journal, 166(1): 49-62.

Dibike, Y.B., Coulibaly, P., 2005. Hydrologic impact of climate change in the Saguenay watershed: comparison of downscaling methods and hydrologic models. Journal of hydrology, 307(1): 145-163.

Dile, Y.T., Berndtsson, R., Setegn, S.G., 2013. Hydrological Response to Climate Change for Gilgel Abay River, in the Lake Tana Basin-Upper Blue Nile Basin of Ethiopia. PloS one, 8(10): e79296.

Ebrahim, G.Y., Jonoski, A., van Griensven, A., Di Baldassarre, G., 2013. Downscaling technique uncertainty in assessing hydrological impact of climate change in the Upper Beles River Basin, Ethiopia. Hydrology Research, 44(2): 377-398.

Elshamy, M.E., Seierstad, I.A., Sorteberg, A., 2009. Impacts of climate change on Blue Nile flows using bias-corrected GCM scenarios. Hydrology and Earth System Sciences, 13(5): 551-565.

Fiseha, B., Melesse, A., Romano, E., Volpi, E., Fiori, A., 2012. Statistical downscaling of precipitation and temperature for the Upper Tiber Basin in Central Italy. International Journal of Water Sciences, 1.

Fowler, H., Blenkinsop, S., Tebaldi, C., 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. International journal of climatology, 27(12): 1547-1578.

Gebre, S.L., Ludwig, F., 2014. Hydrological Response to Climate Change of the Upper Blue Nile River Basin: Based on IPCC Fifth Assessment Report (AR5). Journal of Climatology & Weather Forecasting, 2015.

Gebremicael, T., Mohamed, Y., Betrie, G., van der Zaag, P., Teferi, E., 2013. Trend analysis of runoff and sediment fluxes in the Upper Blue Nile basin: A combined analysis of statistical tests, physically-based models and landuse maps. Journal of Hydrology, 482: 57-68.

Goly, A., Teegavarapu, R.S., Mondal, A., 2014. Development and evaluation of statistical downscaling models for monthly precipitation. Earth Interactions, 18(18): 1-28.

Goodarzi, E., Dastorani, M., Talebi, A., 2015. Evaluation of the change-factor and LARS-WG methods of downscaling for simulation of climatic variables in the future (Case study: Herat Azam Watershed, Yazd-Iran). Ecopersia, 3(1): 833-846.

Hashmi, M.Z., Shamseldin, A.Y., Melville, B.W., 2011. Comparison of SDSM and LARS-WG for simulation and downscaling of extreme precipitation events in a watershed. Stochastic Environmental Research and Risk Assessment, 25(4): 475-484.

Hassan, Z., Shamsudin, S., Harun, S., 2014. Application of SDSM and LARS-WG for simulating and downscaling of rainfall and temperature. Theoretical and applied climatology, 116(1-2): 243-257.

IPCC, 2007. Climate Change 2007. Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, Pachauri, R.K and Reisinger, A.(eds.)].

IPCC, 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

Khalili, M., Leconte, R., Brissette, F., 2007. Stochastic multisite generation of daily precipitation data using spatial autocorrelation. Journal of hydrometeorology, 8(3): 396-412.

Khan, M.S., Coulibaly, P., Dibike, Y., 2006. Uncertainty analysis of statistical downscaling methods. Journal of Hydrology, 319(1): 357-382.

Kim, U., Kaluarachchi, J.J., 2009. Climate change impacts on water resources in the Upper Blue Nile River Basin, Ethiopia1. Wiley Online Library.

Kim, U.K., J. J.; Smakhtin, V. U., 2008. Climate change impacts on hydrology and water resources of the Upper Blue Nile River Basin, Ethiopia. Colombo, Sri Lanka: International Water Management Institute, 27p (IWMI Research Report 126).

King, L.M., McLeod, A.I., Simonovic, S.P., 2015. Improved weather generator algorithm for multisite simulation of precipitation and temperature. JAWRA Journal of the American Water Resources Association, 51(5): 1305-1320.

Maurer, E.P., Hidalgo, H.G., 2008. Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods. Hydrology and Earth System Sciences, 12(2): 551-563.

Qian, B., Hayhoe, H., Gameda, S., 2004. Evaluation of the stochastic weather generators LARS-WG and AAFC-WG for climate change impact studies. Climate Research, 29(1): 3.

Semenov, M.A., Barrow, E.M., 1997. Use of a stochastic weather generator in the development of climate change scenarios. Climatic change, 35(4): 397-414.

Semenov, M.A., Barrow, E.M., Lars-Wg, A., 2002. A stochastic weather generator for use in climate impact studies. User Manual: Hertfordshire, UK.

Semenov, M.A., Stratonovitch, P., 2010. Use of multi-model ensembles from global climate models for assessment of climate change impacts. Climate research (Open Access for articles 4 years old and older), 41(1): 1.

Taye, M.T., Ntegeka, V., Ogiramoi, N., Willems, P., 2011. Assessment of climate change impact on hydrological extremes in two source regions of the Nile River Basin. Hydrology and Earth System Sciences, 15(1): 209-222.

USBR, 1964. Land and water resources of the Blue Nile basin, Ethiopia. Main report. United States Bureau of Reclamation, Washington, DC.

WAPCOS, 1990. Preliminary water resources development master plan for Ethiopia. Prepared for EVDSA. Addis Ababa, Final report.

Wigley, T.M., 2008. MAGICC/SCENGEN 5.3: User manual (version 2). NCAR, Boulder, CO, 80.

Wilby, R. *et al.*, 2004. Guidelines for use of climate scenarios developed from statistical downscaling methods.

Wilby, R., Dawson, C., Barrow, E., 2007. SDSM user manual-a decision support tool for the assessment of regional climate change impacts.

Wilby, R.L., Dawson, C.W., 2013. The statistical downscaling model: insights from one decade of application. International Journal of Climatology, 33(7): 1707-1719.

Wilby, R.L., Dawson, C.W., Barrow, E.M., 2002. SDSM—a decision support tool for the assessment of regional climate change impacts. Environmental Modelling & Software, 17(2): 145-157.

Wilby, R.L., Wigley, T., 1997. Downscaling general circulation model output: a review of methods and limitations. Progress in Physical Geography, 21(4): 530-548.

Worqlul, A.W. *et al.*, 2014. Comparison of rainfall estimations by TRMM 3B42, MPEG and CFSR with ground-observed data for the Lake Tana basin in Ethiopia. Hydrology and Earth System Sciences, 18(12): 4871-4881.

Xu, C.-y., 1999. From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. Progress in Physical Geography, 23(2): 229-249.

Yates, D.N., Strzepek, K.M., 1998a. An assessment of integrated climate change impacts on the agricultural economy of Egypt. Climatic Change, 38(3): 261-287.

Yates, D.N., Strzepek, K.M., 1998b. Modeling the Nile Basin under climatic change. Journal of Hydrologic Engineering, 3(2): 98-108.

Yimer, G., Jonoski, A., Van Griensven, A., 2009. Hydrological response of a catchment to climate change in the upper Beles river basin, upper blue Nile, Ethiopia. Nile Basin Water Engineering Scientific Magazine, 2: 49-59.

Table 1~~1~~: Selected Global climate models from IPCC AR4 incorporated into the LARS-WG

Research centre	Country	GCM	Model acronym	Grid Resolution	Emission Scenarios	Time Periods
Common Wealth Scientific and Industrial Research Organization	Australia	CSIRO-MK3	CSMK3	1.9x1.9^o	A1B, B1	B,T1,T2,T3
Max-Plank Institute for Meteorology	Germany	ECHAM5-OM	MPEH5	1,9x1.9^o	A1B,A2,B1	B,T1,T2,T3
National Institute for Environmental Studies	Japan	MRI-CGCM2.3.	MIHR	2.8x2.8^o	A1B,B1	B,T1,T2,T3
UK Meteorological Office	UK	HadCM3	HADCM3	2.5x3.75^o	A1B,A2,B1	B,T1,T2,T3
Geophysical Fluid Dynamics Lab	USA	GFDL-CM2.1	GFCM21	2x2.5^o	A1B,A2,B1	B,T1,T2,T3
National Centre for Atmospheric Research	USA	CCSM3	NCCCS	1.4x1.4^o	A1B,B1	B,T1,T2,T3

B: baseline; T1: 2011–2030; T2: 2046–2065; T3: 2081–2100

Table 2~~2~~**:** Results of the average statistical tests comparing the observed data ~~for~~ from 26 ~~sites~~ stations with synthetic data generated through LARS-WG. The numbers in the table show ~~how many~~the average numbers of tests gave P value less than ~~significant results at the~~ 5 % significance level.

~~No.~~	~~Stations~~	~~KS test~~		~~t test~~	~~F test~~	~~KS test~~	~~t test~~	~~KS test~~	~~t test~~
~~-~~	~~-~~	~~WDseries~~	~~RainD~~	~~RMM~~	~~RMV~~	~~TminD~~	~~TminM~~	~~TmaxD~~	~~TmaxM~~
~~1~~	~~Abaysheleko~~	~~0~~	~~0~~	~~1~~	~~5~~	~~0~~	~~0~~	~~0~~	~~0~~
~~2~~	~~Adet~~	~~0~~	~~0~~	~~0~~	~~1~~	~~0~~	~~0~~	~~0~~	~~0~~
~~3~~	~~Alemketema~~	~~0~~	~~0~~	~~0~~	~~1~~	~~0~~	~~0~~	~~0~~	~~0~~
~~4~~	~~Anger~~	~~0~~	~~0~~	~~0~~	~~5~~	~~0~~	~~0~~	~~0~~	~~1~~
~~5~~	~~Angerguten~~	~~0~~	~~0~~	~~0~~	~~7~~	~~0~~	~~0~~	~~0~~	~~0~~
~~6~~	~~Arjo~~	~~0~~	~~0~~	~~0~~	~~4~~	~~0~~	~~0~~	~~0~~	~~0~~
~~7~~	~~Ayehu~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~
~~8~~	~~Ayira~~	~~0~~	~~0~~	~~0~~	~~2~~	~~0~~	~~1~~	~~0~~	~~1~~
~~9~~	~~Bahirdar~~	~~0~~	~~0~~	~~1~~	~~3~~	~~0~~	~~0~~	~~0~~	~~0~~
~~10~~	~~Bedele~~	~~0~~	~~0~~	~~1~~	~~1~~	~~0~~	~~0~~	~~0~~	~~0~~
~~11~~	~~Dangila~~	~~0~~	~~0~~	~~1~~	~~2~~	~~0~~	~~0~~	~~0~~	~~0~~
~~12~~	~~Dbirhan~~	~~0~~	~~0~~	~~0~~	~~1~~	~~0~~	~~0~~	~~0~~	~~0~~
~~13~~	~~Dedesa~~	~~0~~	~~0~~	~~0~~	~~2~~	~~0~~	~~0~~	~~0~~	~~1~~
~~14~~	~~Dmarkos~~	~~0~~	~~0~~	~~0~~	~~1~~	~~0~~	~~0~~	~~0~~	~~0~~
~~15~~	~~Dtabor~~	~~0~~	~~0~~	~~1~~	~~3~~	~~0~~	~~0~~	~~0~~	~~0~~
~~16~~	~~Fitche~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~
~~17~~	~~Gatira~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~
~~18~~	~~Gidayana~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~
~~19~~	~~Gimijabet~~	~~1~~	~~0~~	~~1~~	~~3~~	~~-~~	~~-~~	~~-~~	~~-~~
~~20~~	~~Gondar~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~
~~21~~	~~Motta~~	~~0~~	~~0~~	~~0~~	~~1~~	~~0~~	~~0~~	~~0~~	~~0~~
~~22~~	~~Mselam~~	~~0~~	~~0~~	~~0~~	~~2~~	~~0~~	~~1~~	~~0~~	~~0~~
~~23~~	~~Nedjo~~	~~0~~	~~0~~	~~0~~	~~3~~	~~0~~	~~0~~	~~0~~	~~0~~
~~24~~	~~Shambu~~	~~0~~	~~0~~	~~1~~	~~2~~	~~0~~	~~0~~	~~0~~	~~0~~
~~25~~	~~Yetnora~~	~~0~~	~~0~~	~~0~~	~~2~~	~~0~~	~~1~~	~~0~~	~~0~~
~~26~~	~~Zege~~	~~0~~	~~0~~	~~0~~	~~3~~	~~0~~	~~0~~	~~0~~	~~0~~
~~-~~	~~Average~~	~~0.04~~	~~0~~	~~0.27~~	~~2.08~~	~~0~~	~~0.12~~	~~0~~	~~0.12~~
~~-~~	~~No. of tests~~	~~8~~	~~12~~	~~12~~	~~12~~	~~12~~	~~12~~	~~12~~	~~12~~

Tests	KS-test		t-test	F-test	KS-test	t-test	KS-test	t-test
Parameters	WDseries	RainD	RMM	RMV	TminD	TminM	TmaxD	TmaxM
Average	0.04	0.00	0.27	2.08	0	0.12	0	0.12
Total	8	12	12	12	12	12	12	12
% failed	0.48	0.00	2.24	17.31	0	1	0	1

Table 3~~3~~: Name and description of all NCEP predictors on HadCM3 & canESM2 grid

Variables	Descriptions	variables	Descriptions
temp	Mean temperature at 2 m	s500 +	Specific humidity at 500 hpa height
mslp	Mean sea level pressure	s850+	Specific humidity at 850 hpa height
p500	500 hpa geopotential height	**_f	Geostrophic air flow velocity
p850	850 hpa geopotentail height	**_z	Vorticity
rhum *	Near surface relative humidity	**_u	Zonal velocity component
r500*	Relative humidity at 500 hpa	**_v	Meridional velocity component
r850*	Relative humidity at 850 hpa	**zh	Divergence
shum	Near surface specific humidity	**thas	Wind direction
Prec+	Total precipitation		

(**) refers to different atmospheric levels: the surface (p_), 850 hpa height (p8), and 500 hpa height (p5)
(*) refers predictors only found from HadCM3, (+) refers predictors only for canESM2

Table 4~~4~~: Average values of statistical performance indicators for all stations using SDSM

	Performance indicators	Climate variables	Mean Daily Precipitation (mm)		Mean monthly Precipitation (mm)		Mean daily Tmax (^{o}C)		Mean dailyTmin (^{o}C)	
		GCMs	HadCM3	canESM2	HadCM3	canESM2	HadCM3	canESM2	HadCM3	canESM2
Cal	Mean	Observed	7.94	7.94	121.7	121.7	24.9	24.8	11.5	11.6
		Simulated	7.97	7.87	127.8	123.7	24.9	24.8	11.5	11.6
	R^2		0.93	0.94	0.98	0.99	0.99	0.99	0.99	1.00

	MAE	value	0.8	0.4	11.1	5.9	0.1	0.1	0.1	0.0
		Ratio relative to mean(%)	9.7	5.3	9.1	4.8	0.5	0.2	1.1	0.2
	RMSE	value	1.0	0.7	15.8	8.9	0.2	0.1	0.2	0.0
		Ratio relative to mean(%)	12.6	8.6	13.0	7.3	0.6	0.3	1.3	0.3
	NSE		0.82	0.92	0.97	0.99	0.96	0.99	1.00	1.00
	Bias		0.51	0.20	4.28	3.65	0.09	0.02	0.09	0.01
Validation*	R^2		0.70	0.71	0.90	0.95	0.97	0.95	0.96	0.95
	MAE	value	2.0	1.8	32.4	23.2	0.4	0.5	0.3	0.4
		Ratio relative to mean(%)	25.2	22.9	26.6	19.1	1.5	1.8	2.4	3.6
	RMSE	value	2.4	2.4	46.4	33.4	0.5	0.6	0.3	0.5
		Ratio relative to mean(%)	30.7	30.0	38.1	27.4	1.9	2.2	3.0	4.3
	NSE		0.37	0.55	0.83	0.92	0.90	0.87	1.00	0.99
	Bias		1.13	0.92	14.77	13.20	0.26	0.33	0.18	0.31

+ calibration period for canESM2 (1984-2000), for HadCM3 (1984-1995) * Validation period for canESM2 (2001-2005), for HadCM3 (1996-2001)

~~**Table 5: Relative change mean annual precipitation and change in Tmax and Tmin modeled from six GCMs for three time periods of UBNRB as compared from the reference period of 1984-2011 by using LARS-WG**~~

	1984-2011	2030s (2011-2035)			2050s (2046-2065)			2080s (2080-2099)		
GCMs/Scenario		A1B	B1	A2	A1B	B1	A2	A1B	B1	A2
Co2 concentration (ppm)		410	418	414	492	541	545	538	674	754
Mean annual rainfall (mm)	1417.5									
CSMK3	Relative Change of precipitation(%)	-2.3	-2.3		-4.2	-2.7		-7.0	-5.3	
GFCM21		-1.4	-0.6	-3.7	-8.0	0.7	-7.4	-7.5	-2.2	-5.9
HADCM3		2.1	0.8	1.7	4.4	2.1	3.5	12.9	4.1	16.7
MIHR		1.9	3.7		5.5	5.5		10.2	6.0	
MPEH5		1.8	3.3	-0.5	2.5	5.8	4.2	6.0	1.4	3.3
NCCCS		6.5	7.8	6.4	22.8	22.0	8.7	29.9	13.7	43.7
Model average		1.4	2.1	1.0	3.8	5.6	2.2	7.4	3.0	14.5
Mean daily Tmax (^{o}C)	24.7									
CSMK3	Change in maximumTemperature (^{o}C)	0.4	0.4		1.3	0.9		2.2	1.5	
GFCM21		0.7	0.6	0.7	2.2	1.4	1.9	3.1	2.0	3.6
HADCM3		0.5	0.5	0.4	1.7	1.4	1.8	3.1	2.0	3.7
MIHR		0.6	0.6		2.0	1.6		3.5	2.6	
MPEH5		0.5	0.4	0.6	1.8	1.4	1.8	4.1	2.7	4.3
NCCCS		0.6	0.5	0.6	1.5	0.9	1.7	2.2	1.4	3.0
Model average		0.6	0.5	0.6	1.8	1.3	1.8	3.0	2.0	3.6
Mean daily Tmin (^{o}C)	11.4									
CSMK3	change in minimum Temperature (^{o}C)	0.3	0.3		1.1	0.8		1.9	1.3	
GFCM21		0.7	0.6	0.7	2.2	1.4	1.9	3.1	2.0	3.6
HADCM3		0.5	0.5	0.4	1.7	1.4	1.8	3.1	2.0	3.7
MIHR		0.7	0.7		2.1	1.8		3.6	2.7	
MPEH5		0.5	0.4	0.6	1.8	1.5	1.8	4.1	2.7	4.1
NCCCS		0.6	0.5	0.6	1.5	0.9	1.7	2.2	1.4	3.0
Model average		0.6	0.5	0.6	1.7	1.3	1.8	3.0	2.0	3.6

~~**Table 6: Relative change of mean annual precipitation, change of mean annual Tmax and Tmin for three time periods as compared to the baseline period of UBNRB using SDSM for HadCM3 and canESM2 GCMs under different scenarios**~~

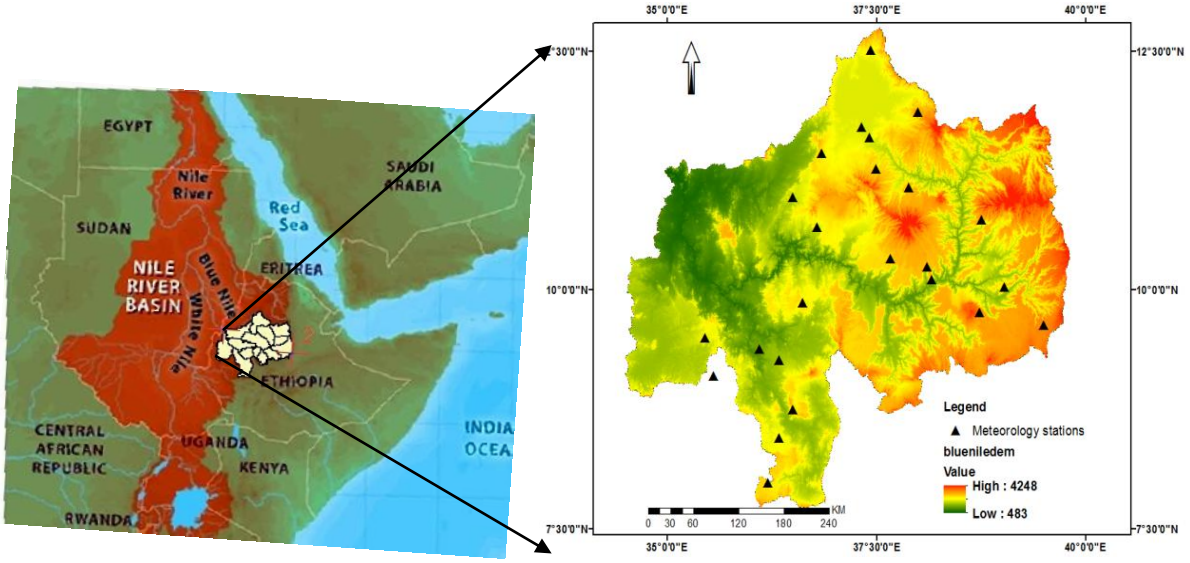


Figure 1~~1~~: Location Map of the study area

a) — b)

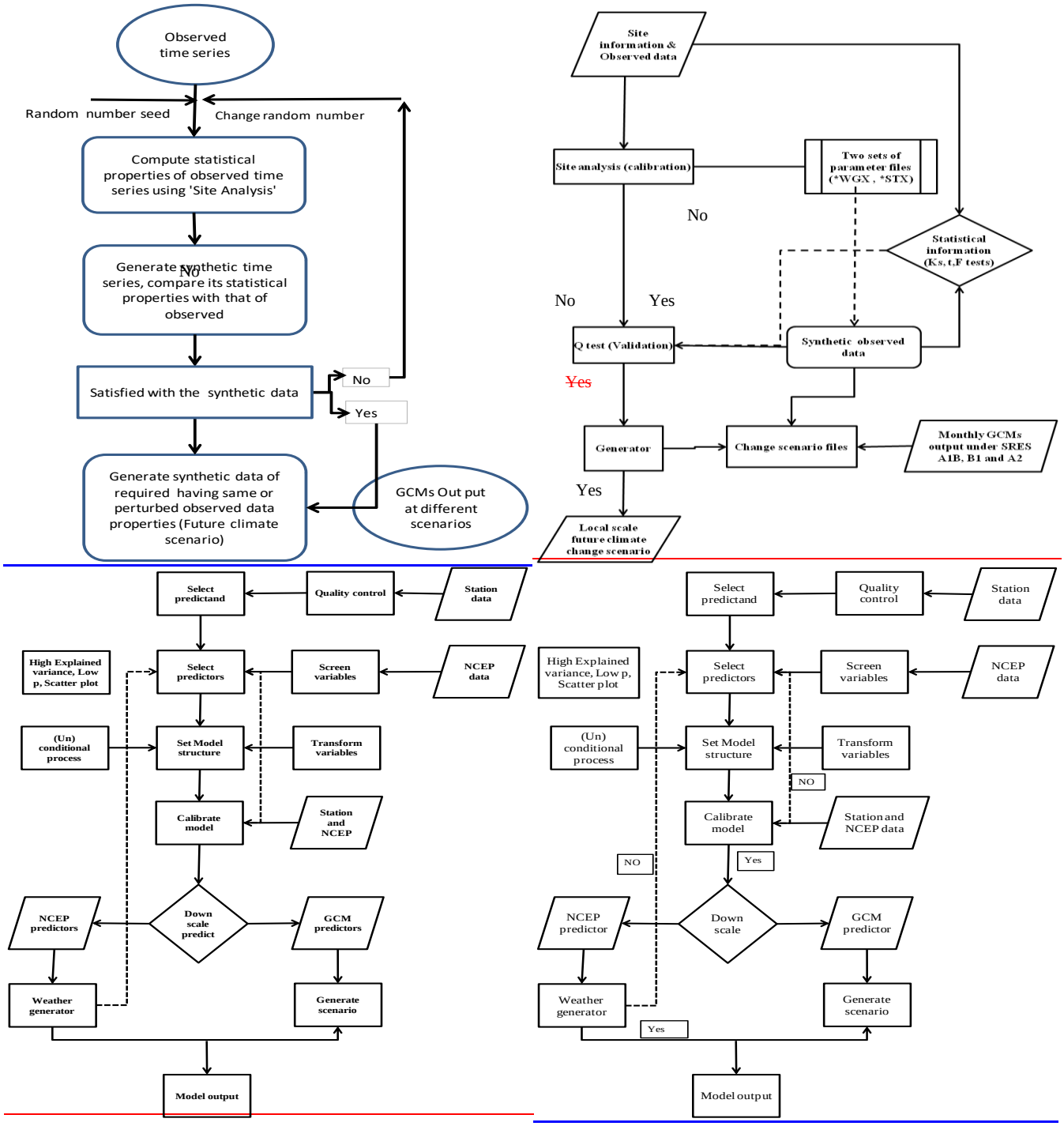


Figure 2~~2~~: Schematic diagram of a) LARS WG analysis b) SDSM analysis source (Wilby *et al.*, 2002)

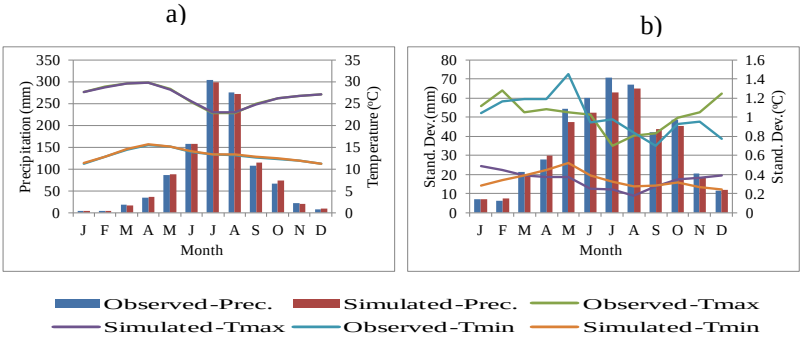


Figure 33: Observed and simulated a) mean monthly precipitation,Tmax and Tmin ; b) standard deviation of precipitation, Tmax and Tmin using LARS-WG

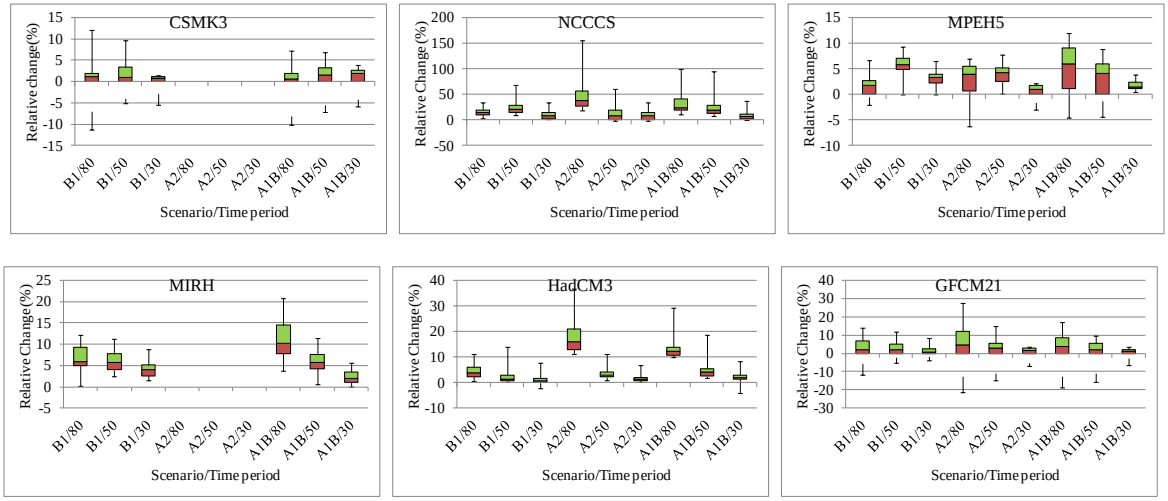
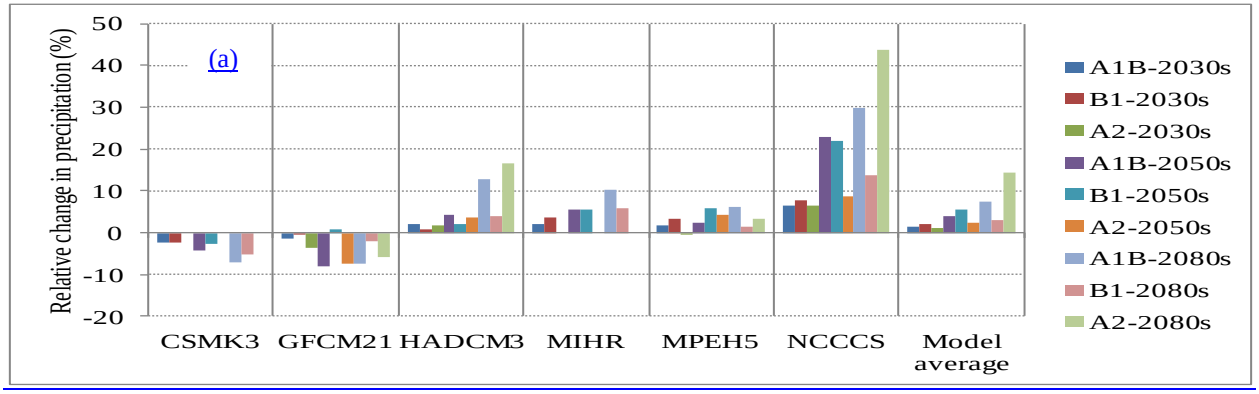


Figure 44: Box plots showing the relative change of precipitation (%) for each six selected GCMs downscaled from 15 stations by using LARS-WG for scenarios (B1, A2 and A1B) during three time periods as compared to the base line. Box boundaries indicate the 25th and 75th percentiles, the line within the box marks the median, whiskers below and above the box indicate the 10th and 90th percentiles,

B1/80: B1 scenario time period 0f 2080s, B1/50: B1 scenario time period 2050s, B1/30: B1 scenario time period 2030s



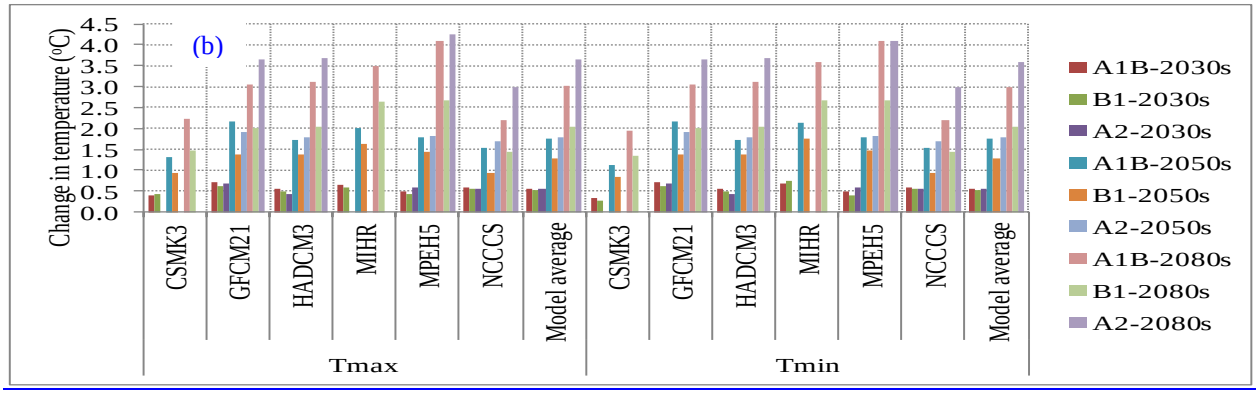


Figure 5: (a) Relative change mean annual precipitation and (b) change in Tmax and Tmin modeled from six GCMs for three time periods of UBNRB as compared from the reference period of 1984-2011 by using LARS-WG

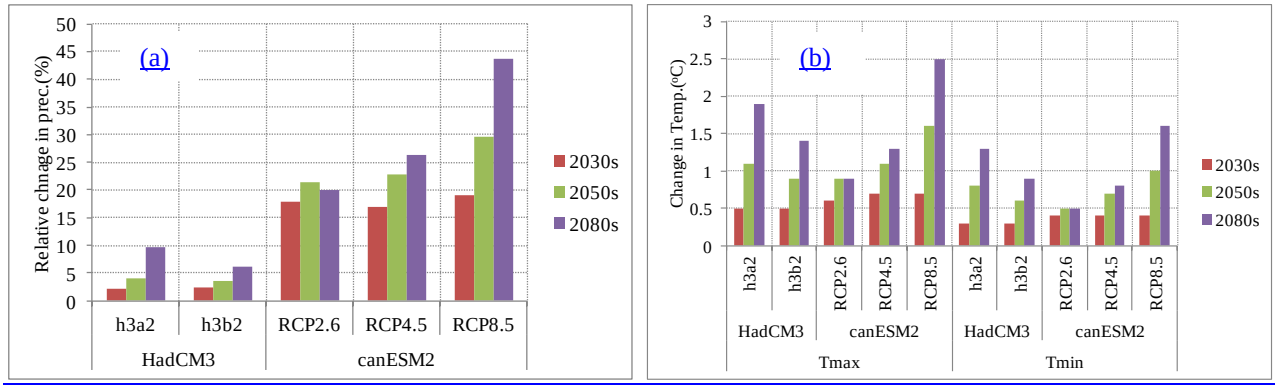


Figure 6: (a) Relative change of mean annual precipitation, and (b) change of mean annual Tmax and Tmin for three time periods as compared to the baseline period of UBNRB using SDSM for HadCM3 and canESM2 GCMs under different scenarios

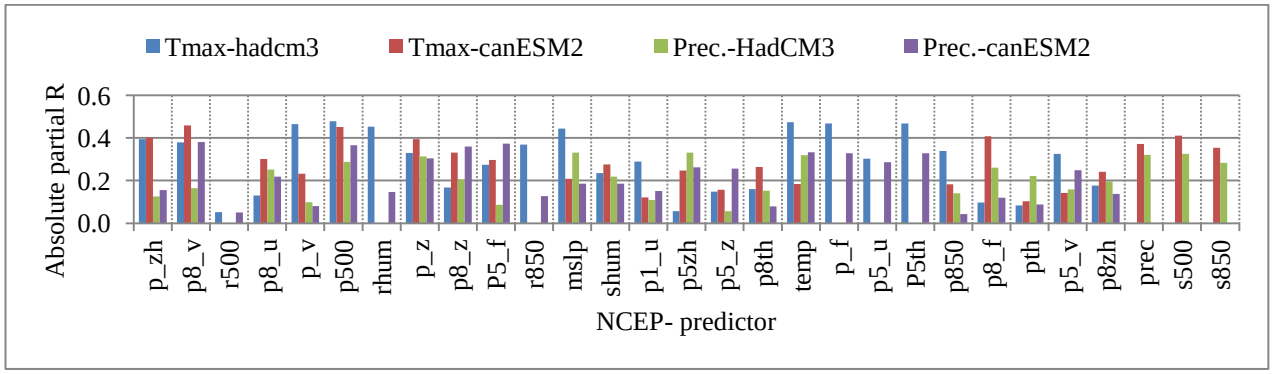


Figure 7~~5~~: Average partial correlation coefficient values of all stations for precipitation and Tmax with NCEP- reanalysis predictors

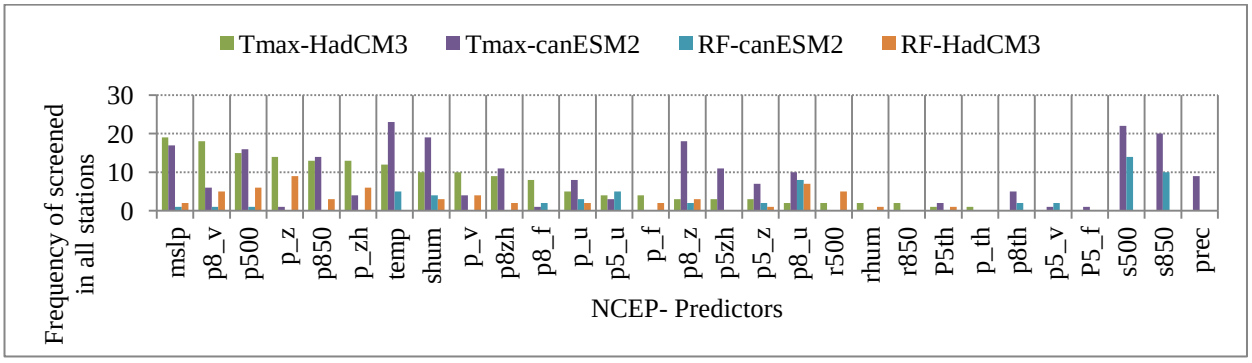


Figure 8~~6~~: Screened NCEP- predictor variables for observed Tmax and precipitation from two GCMs, the maximum frequency is 15 for precipitation and 25 for Tmax

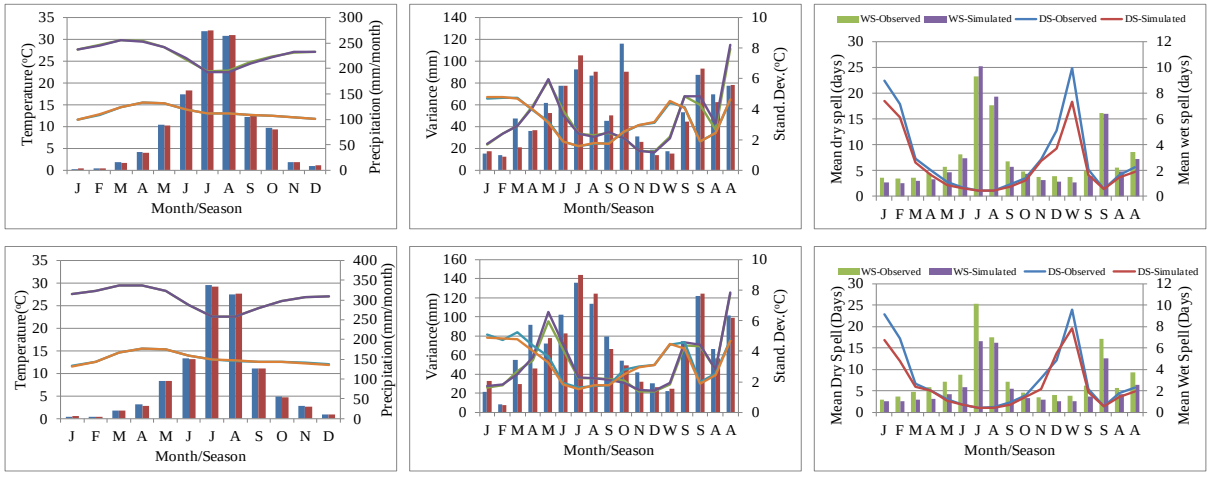
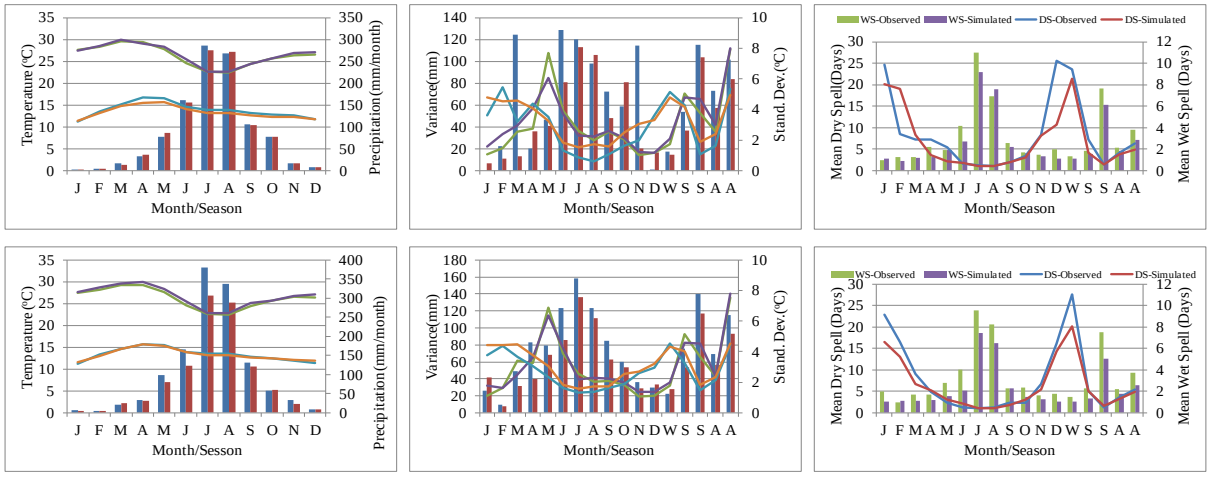


Figure 9~~7~~: calibration of observed and simulated of precipitation, maximum and minimum temperature for the Gondar station using SDSM from canESM2 and HadCM3 from top to bottom



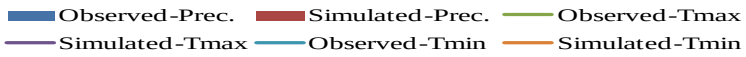


Figure 108: Validation of observed and simulated of precipitation, maximum and minimum temperature for Gondar station using SDSM from canESM2 and HadCM3 from left to right respectively

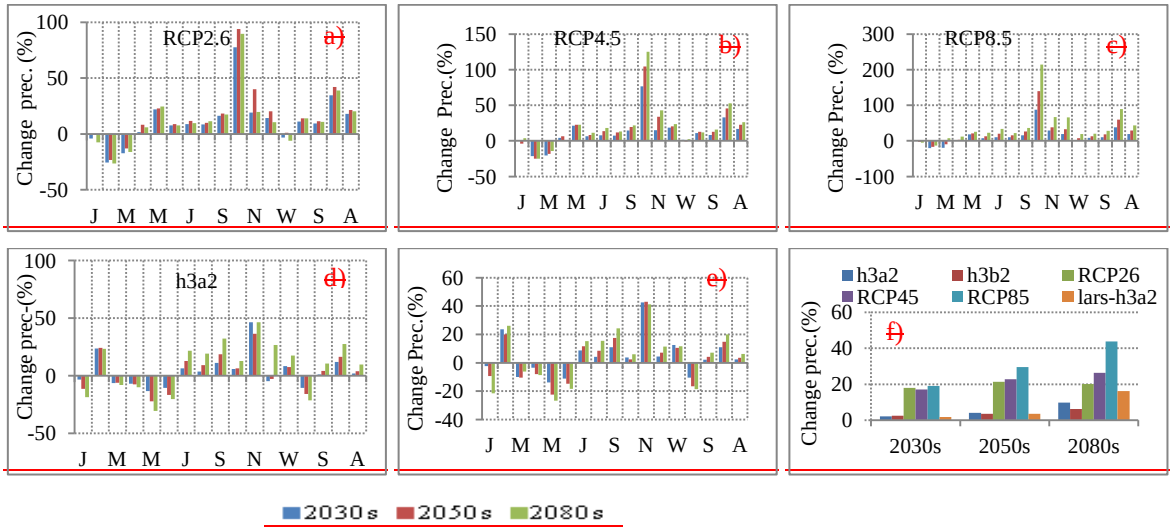
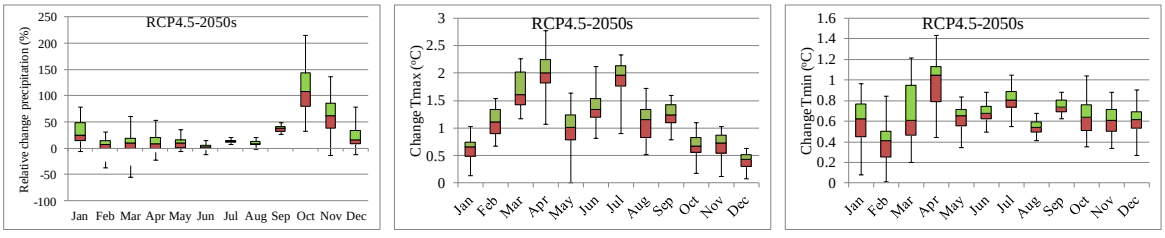


Figure 9: Relative change in monthly mean precipitation for three time periods as compared to the baseline period of UBNRB area a) canESM2/RCP2.6, b) canESM2/RCP4.5, c) canESM2/RCP8.5, d) hadCM3/A2a and e) hadCM3/B2a scenario f) relative change in mean annual precipitation for all scenarios.

(a)



(b)

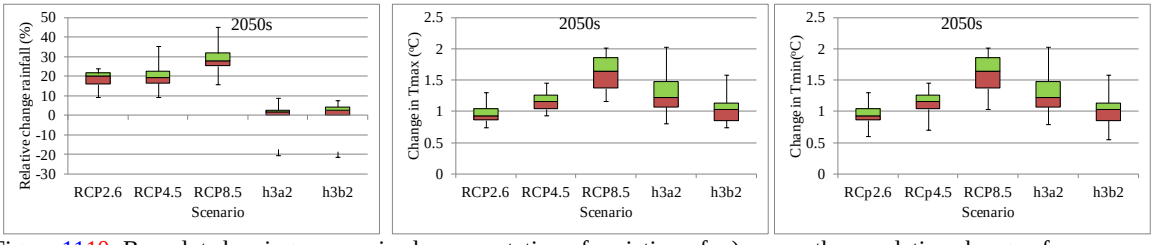
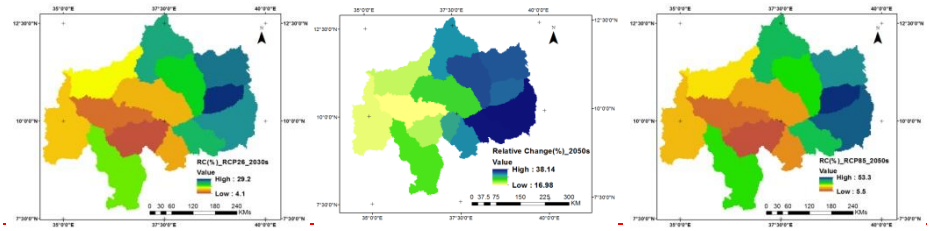


Figure 1110: Box plot showing summarized representation of variation of a) upper three: relative change of mean monthly precipitation, change maximum and minimum temperature from left to right for UBNRB across all stations under RCP4.5 at 2050s Using SDSM b) lower three: Relative change of mean annual precipitation, change in mean annual maximum and minimum temperature from left to right at 2050s for scenarios (RCP2.6, 4. 5 and 8.5 and SRES A2 and B2) for UBNRB using SDSM.



~~**Figure 11: Spatial distribution of the relative change in precipitation under RCP2.6, 4.5 and 8.5 scenario 2050s from left to right by using SDSM**~~

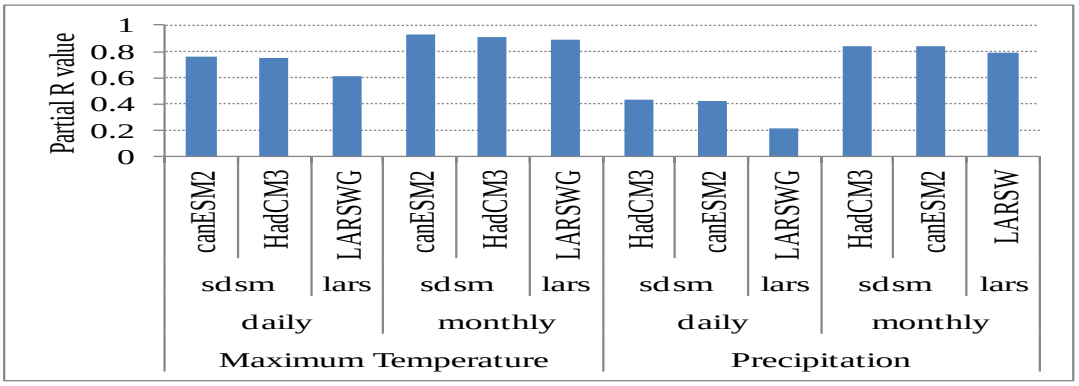


Figure 12~~12~~**:** Performance comparison of LARSWG and SDSM at different time scale

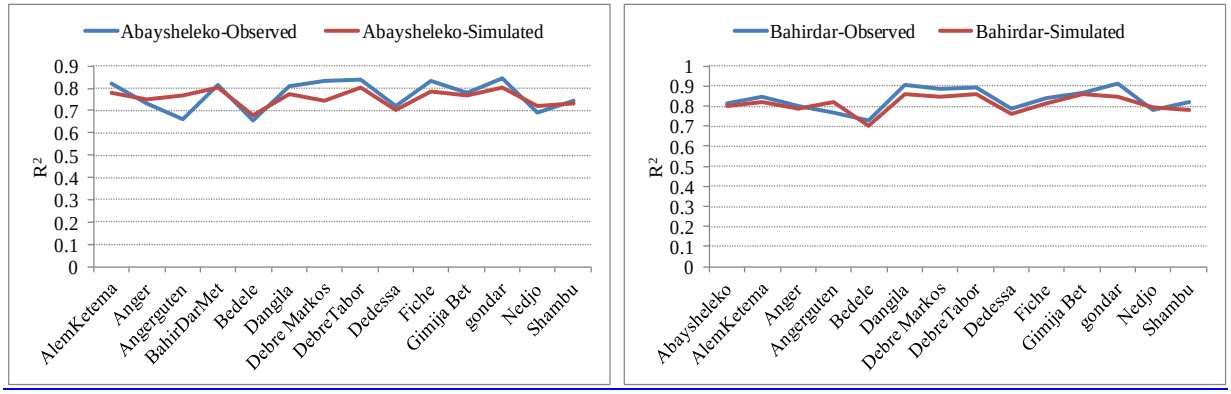
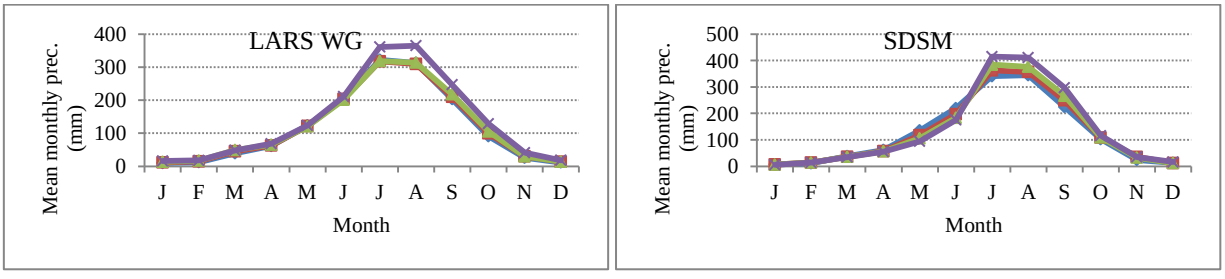


Figure 13: Spatial correlation coefficient (R^2) of Abaysheleko (left) and Bahirdar (right) weather stations with others for monthly precipitation from 1984-2011.



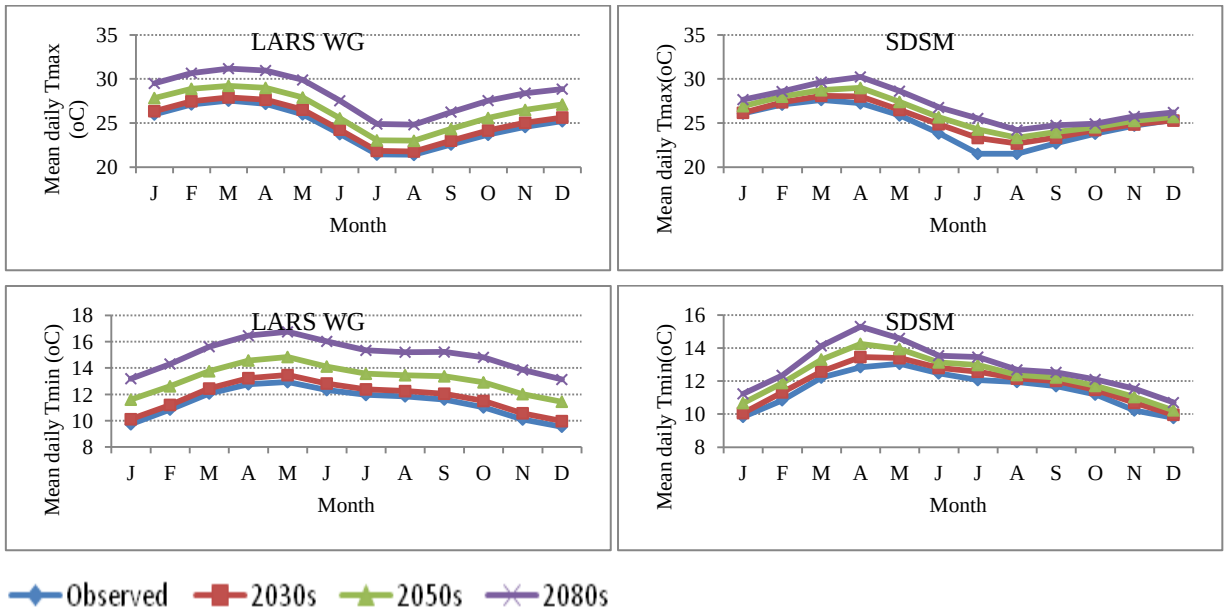


Figure 14~~13~~: General trend in precipitation, Tmax and Tmin at UBNRB corresponding to a climate change scenario downscaled using LARS WG and SDSM from HadCM3 GCM for a2 scenario~~o~~

Analyzing the future climate change of Upper Blue Nile River Basin (UBNRB) using statistical down scaling techniques

Dagnenet Fenta Mekonnen[1,2], Markus Disse[1]

[1]Chair of Hydrology and River Basin Management, Faculty of Civil, Geo and Environmental Engineering, Technische Universität München, Arcisstrasse 21, 80333, Munich, Germany.
[2]Amhara Regional State Water, Irrigation and Energy Development Bureau, Bahirdar, Ethiopia
Correspondence to: Dagnenet Fenta (dagnfenta@yahoo.com)

Abstract. Climate change is becoming one of the most arguable and threatening issues in terms of global context and their responses to environment and socio/economic drivers. Its direct impact becomes critical for water resource development and indirectly for agricultural production, environmental quality, economic development, social well-being. However, a large uncertainty between different Global Circulation Models (GCMs) and downscaling methods exist that makes reliable conclusions for a sustainable water management difficult. Hence, to value the uncertainty of GCMs and downscaling processes, - a multi-model approach from a systematically selected six CMIP3 GCMs and one CMIP5 GCMs were used to construct climate change scenarios of precipitation, maximum and minimum temperature for the UBNRB ~~In order to understand the future climate change of the Upper Blue Nile River Basin,~~ by applying two widely used statistical down scaling techniques namely Long Ashton Research Station Weather Generator (LARS-WG) and Statistical Down Scaling Model (SDSM) ~~models were applied~~as they are computationally less demanding and efficient. ~~Six CMIP3 GCMs for LARS WG (CSIRO-MK3, ECHAM5-OM, MRI-CGCM2.3.2, HaDCM3, GFDL-CM2.1, CCSM3) model while HadCM3 GCM and canESM2 from CMIP5 GCMs for SDSM were used for climate change analysis.~~

The downscaled precipitation results from ~~the prediction of~~ the six GCMs by LARS WG showed inconsistency and large inter model variability, two GCMs showed decreasing trend while 4 GCMs showed increasing in the range from -7.9 % to +43.7 % while the ensemble mean of the six GCM result showed increasing trend ranged from 1.0 % to 14.4 %. ~~NCCCS GCM predicted maximum increase in mean annual precipitation. However, the projection from HadCM3 GCM is consistent with the multi model average projection, which predicts precipitation increase from 1.7% to 16.6%.~~ Conversely, the result from all GCMs showed a similar continuous increasing trend for maximum temperature (Tmax) and minimum temperature (Tmin) in all three future time periods. ~~-~~The change for mean annual Tmax may increase from 0.4 °c to 4.3 °c whereas the change for mean annual Tmin may increase from 0.3 °c to 4.1°c

~~Meanwhile~~Even though, both HadCM3 and canESM2 GCMs using SDSM agree with respect to the direction of 21st century ~~the result from SDSM showed an increasing trend for all three climate variables (~~precipitation, minimum and

maximum temperature changes, there are considerable variability in magnitude~~) from both HadCM3 and canESM2 GCMs~~. The relative change of mean annual precipitation downscaled from 5 different scenarios of two GCMs range from 2.1 % to 43.8 % while the change for mean annual Tmax and Tmin may increase from 0.4 ^{o}c to 2.9 ^{o}c and from 0.3 ^{o}c to 1.6 ^{o}c respectively. The change in magnitude for precipitation is higher in RCP8.5 scenarios than the others as expected. The ~~present~~ study results illustrate that both down scaling techniques have shown comparable and good ability to simulate the current local climate variables which can be adopted for future climate change study ~~with high confidence~~ for the UBNRB. However, based on the comparative performance evaluation results of this study, SDSM would be more robust and can be applied at higher confidence for downscaling large scale GCMs outputs to finer scales to suit for hydrological models for impact assessment in the study area of UBNRB. ~~In order to see the comparative downscaling results from the two down scaling techniques, HadCM3 GCM of A2 scenario was used in common. The result obtained from the two down scaling models were found reasonably comparable and both approaches showed increasing trend for precipitation, Tmax and Tmin. However, the analysis of the downscaled climate data from the two techniques showed, LARS WG projected a relatively higher increase than SDSM.~~

Key words: Climate Change, GCM, statistical down scaling, LARS WG, SDSM; UBNRB

1. Introduction

The impacts of climate change on the hydrological cycle in general and on water resources in particular are of high significance due to the fact that all natural and socio/economic system critically depend on water. The direct impact of climate change can be variation and changing pattern of water resources availability and hydrological extreme events such as floods and droughts, with many indirect effects on agriculture, food and energy production and overall water infrastructure (Ebrahim *et al.*, 2013). The impact may be worse on trans-boundary Rivers like Upper Blue Nile River where competition for water is becoming high from different economic, political and social interests of the riparian countries and when runoff variability of upstream countries can greatly affect the downstream countries (Kim, 2008; Semenov and Barrow, 1997).

According to IPCC (2007), between 75 and 250 million people are projected to be exposed to increased water stress due to climate change in Africa by 2020. The increasing water demand of upstream countries in the Nile Basin coupled with climate change impacts can affect the availability of water resources for downstream countries and in the basin, that could result in resource conflicts and regional insecurities. Moreover, climate variability, the way climate fluctuates yearly and seasonally above or below a long-term average value, caused by changes in forcing factors such as variation in seasonal extent of the Inter Tropical Convergence Zone (ITCZ) like El Niño and La Niña events, ~~Current climate variability~~ is already imposing a significant challenge to Ethiopia by affecting food security, water and energy supply, poverty reduction and sustainable socio-economic development efforts. To mitigate these challenges, ~~T~~the Ethiopian government is therefore carried out a series of studies on Upper Blue Nile river Basin (UBNRB), which have been identified as an economic "growth

corridor", focused on identifying irrigation, hydropower potential and the use of the extensive water resources of the basin (BCEOM, 1998; USBR, 1964; WAPCOS, 1990).~~, with less attention for climate change and its impact.~~ As the result, large scale irrigation and hydro-power projects including the Grand Ethiopian Renaissance Dam (GERD), the largest hydroelectric power plant in Africa, ~~has~~ have been ~~constructed~~ identified and ~~is~~ being constructed ~~to mitigate the~~as mitigation measure for the ~~-~~impacts of climate change. However, most studies were given less emphasis for climate change and its impact on the hydrology of the basin, ~~H~~hence, identifying local impacts of climate change at ~~a catchment~~ basin level is quite important especially in UBNRB for the sustainability of large scale water resource development projects, for proper water resource management leading to regional security and looking for the possible mitigation measures otherwise the consequences becoming catastrophic.

To this end, several individual researches have been done to study the impacts of climate change on the water resources of Upper Blue Nile River Basin. Taye *et al.* (2011) reviewed some of the research outputs and concluded that clear discrepancies were observed particularly on the projection of precipitation. For instance, the result obtained from (Bewket and Conway, 2007; Conway, 2000; Gebremicael *et al.*, 2013) showed that there is no significant change on the amount of rainfall and there is no consistent patterns or trends observed. Kim (2008) used the outputs of six GCMs ~~to project~~for the projection of future precipitations and temperature, the result suggested that the changes in mean annual precipitation from the six GCMs range from -11 % to 44 % with a change of 11% from the weighted average scenario at 2050s. On the other hand, the changes in mean annual temperature range from 1.4°C to 2.6^{o}C with a change of 2.3^{o}C from the weighted average scenario. Likewise, Yates and Strzepek (1998a) used 3 GCMs and the result revealed that the changes in precipitation range from -5% to 30% and the change in temperature range from 2.2^{o}c to 3.5^{o}c. Yates and Strzepek (1998b) also used 6 GCMs and the result showed in the range from -9% to 55% for precipitation while temperature increased from 2.2^{o}c to 3.7 ^{o}c. Another study done by Elshamy *et al.* (2009), ~~–~~used 17 GCMs and the result showed that Changes in total annual precipitation range between −15 % to +14 % but the ensemble mean of all models showed almost no change in the annual total rainfall. ~~Moreover~~While, all models predict the temperature to increase between 2^{o}C and 5^{o}C. Gebre and Ludwig (2014), used five biased corrected 50km x 50km spatial resolution GCMs for RCP4.5 and RCP8.5 scenarios to down scale the future climate change of 4 watershed (Gilgel Abay, Gumara, Ribb and Megech) located in Tana sub basin of UBNRB for the time period of 2030s and 2050s. The result suggested that the selected five GCMs disagree on the direction of future prediction of precipitation but multimodal average monthly and seasonal precipitation may generally increases over the watersheds.

For the historical context, the discrepancies could be due to the period and length of data analyzed and the failure to consider stations which can represent the spatial variability of the basin and also errors induced from observed data. For the future context, ~~apart from~~beside the above mentioned reasons, discrepancies could be due to the difference of GCMs and scenarios

used for downscaling, the downscaling techniques applied (can be dynamical and statistical), selection of representative predictors, the period of analysis and spatial and temporal resolution of observed and predictor dataset.

To address uncertainty in projected climate changes, the (IPCC, 2014) thus recommends using a large ensemble of climate change scenarios produced from various combinations of Atmospheric Ocean General Circulation Model (AOGCMs) and forcing scenarios. However, it can become prohibitively time consuming to assess the climate change, using simultaneously many climate change scenarios and many Statistical Down scaling models. As a result, researchers typically assess the climate change and its impacts under only one or a few climate change scenarios selected arbitrarily with no justification for instance (Elshamy *et al.*, 2009; ~~Kim and Kaluarachchi, 2009;~~ Kim, 2008) used only A1B and A2 scenarios respectively. Yet, there is no any hard rule to select an appropriate subset of climate change scenarios among the wide range of possibilities (Casajus et al., 2016).

~~Recently, there is great advancement in Global Circulation Models (GCMs) to represent large scale (global and continental) climate fairly well, However, they often fail to simulate less scale climate features which are required by hydrological models to carry out impact studies, particularly for precipitation (Semenov *et al.*, 1997). To overcome this problem two sets of techniques have emerged as a means to bridge this resolution gap (Fowler *et al.*, 2007; Wilby *et al.*, 2002) by employing either dynamic down scaling or statistical down scaling methods. Even though, statistical down scaling is problematic in producing realistic future climate change scenario because of recognized inter-variable biases in host GCMs, it has practical advantage and is the more promising option in situations where low-cost, rapid assessments of localized climate change impacts are required (Wilby *et al.*, 2002).~~

Although climate models are usually responsible for high uncertainty in climate change impact analysis, the processes of downscaling that ensures to narrow down the scale discrepancy between the coarse scale GCMs and the required local scale climate variables for hydrological models should be investigated for their contribution. Many downscaling models have been developed in the past decade, to bridge the resolution gap between the coarse resolution GCMs and the required local scale for hydrological models to carry out impact studies, ~~(Fowler *et al.*, 2007)~~, however, no single model has been found to perform well over all the regions and time scales. Thus, evaluations of different models are critical to understand the applicability of the existing models (Dibike and Coulibaly, 2005; Ebrahim et al., 2013; Fiseha et al., 2012; Goodarzi et al., 2015; Hashmi et al., 2011; Khan et al., 2006; Qian et al., 2004; Wilby et al., 2004; Wilby and Wigley, 1997; Xu, 1999). However, it remains difficult to directly compare the skill of different downscaling models (Goly et al., 2014).

Apart from the GCMs and downscaling techniques, most of the previous studies e.g (Beyene et al., 2010; Elshamy et al., 2009; Kim, 2008), used CRU, NFS and other gridded data sets constructed based on the interpolation of a few stations in Ethiopia, which has relatively less accuracy as compared with the station based data (Worqlul *et al.*, 2014). Therefore, the objective of this study is to construct and analyze detailed climate change scenarios for precipitation, maximum and minimum temperature over Upper Blue Nile River Basin at required resolution which can be used for further hydrological

impact study. This can be achieved through the inclusion of systematically selected multiple GCMs and two downscaling methods by incorporating acceptable number of weather stations which has long time series and reliable observed climate data to appreciate the uncertainties coming from GCMs and the process of downscaling methods to overcome the short comings of the previous studies on the study area .

~~Therefore, the objective of this study is to analyze and to better comprehend the possible future climate trend of Upper Blue Nile River Basin by applying widely used and more plausible statistical down scaling techniques.~~

~~Recently, there is great advancement in Global Circulation Models (GCMs) to represent large scale (global and continental) climate fairly well, However, they often fail to simulate less scale climate features which are required by hydrological models to carry out impact studies, particularly for precipitation (Semenov *et al.*, 1997). To overcome this problem two sets of techniques have emerged as a means to bridge this resolution gap (Fowler *et al.*, 2007; Wilby *et al.*, 2002) by employing either dynamic down scaling or statistical down scaling methods. Even though, statistical down scaling is problematic in producing realistic future climate change scenario because of recognized inter variable biases in host GCMs, it has practical advantage and is the more promising option in situations where low cost, rapid assessments of localized climate change impacts are required (Wilby *et al.*, 2002).~~

Generally, downscaling methods are classified in to dynamic and statistical downscaling (Fowler *et al.*, 2007; Wilby *et al.*, 2002). Dynamic downscaling nests higher resolution Regional Climate Models(RCMs) into coarse resolution GCMs to produce complete set of meteorological variables which are consistent each other. The outputs from this method is still not at required scale to what the hydrological model require. Statistical downscaling overcomes this challenge moreover it is computationally undemanding, simple to apply and provides the possibility of uncertainty analysis (Dibike *et al.*, 2005; Semenov *et al.*, 1997; Wilby *et al.*, 2002). Extensive details on the strength and weakness of the two methods can be found (Wilby *et al.*, 2007; Wilby *et al.*, 1997). ~~Recently~~Among the different possibilities, two well recognized statistical downscaling tools, ~~are made available to the broader climate change impact study community. The first one implements~~ a regression based ~~method and is referred to as~~ Statistical Down-Scaling Model (SDSM) (Wilby *et al.*, 2002) ~~while the second is~~and a stochastic weather generator called Long Ashton Research Station Weather Generator (LARS-WG) (Semenov *et al.*, 1997; Semenov *et al.*, 2002) were chosen for this study. They have been tested in various regions e.g., (Chen *et al.*, 2013; Dibike *et al.*, 2005; Dile *et al.*, 2013; Elshamy *et al.*, 2009; Fiseha *et al.*, 2012; Hashmi *et al.*, 2011; Hassan *et al.*, 2014; Maurer and Hidalgo, 2008; Yimer *et al.*, 2009) under different climatic conditions of the world. ~~They are most widely used for climate change impact studies (Wilby and Dawson, 2013). Therefore, statistical down scaling technique of LARS WG and SDSM were applied for this study.~~

2. Description of Study Area

The Upper Blue Nile River Basin (UBNRB) extends from 7°45' to 13° N and 34°30' and 37°45' E. It is one of the most important major basin of Ethiopia because it contributes to 45% of the countries surface water resources, 20% of the population, 17% of the landmass, 40% of the nation's agricultural product and large portion of the hydropower and irrigation potential of the country (Elshamy *et al.*, 2009). The whole UBNRB has an area coverage of 199,812 km^2 (BCEOM, 1998). For this study Rahad, Gelegu and Dinder sub catchments that do not flow through the main river stem to Sudan is excluded. The basin area coverage is 176,000km^2 which is about 15% of total area of 1.12 million km^2(Awulachew *et al.*, 2007) of Ethiopia . The elevation ranges between 489 m.a.s.l downstream on the western side to 4261m.a.s.l upstream at Mount Ras Dashen in the north-eastern part.

The Upper Blue Nile River itself has an average annual run-off of about 49 BCM. In addition, the Dinder, Galegu and Rahad rivers have a combined annual run-off of about 5 BCM. The rivers of the Upper Blue Nile River Basin contribute on average about 62 per cent of Nile total at Aswan. Together with contributions of the Baro-Akobo and Tekeze rivers, Ethiopia accounts for 86 per cent of run-off at Aswan (BCEOM, 1998). The climate of Ethiopia is mainly controlled by the seasonal migration of the Inter-tropical Convergence Zone (ITCZ) following the position of the sun relative to the earth and the associated atmospheric circulation. It is also highly influenced by the complex topography. The whole UBNRB has long term mean annual rainfall, minimum and maximum temperature of 1452 $mmyr^{-1}$, 11.4°C and 24.7°C respectively as calculated by this study from 15 rainfall and 25 temperature gauging stations from the period 1984-2011. The mean seasonal rainfall based on the above data showed about 238 mm, 1065 mm, and 148 mm occurred in Belg (October-January), Kiremit (July-September), and Bega (February-May) respectively, in which about 74 % of rainfall concentrates between June and September (Kiremit season).

3. Datasets

3.1 Local data sets

The historical precipitation, maximum and minimum temperature data for the study area were obtained from Ethiopian Meteorological Agency (EMA), which were analyzed and checked for further quality control. A considerable length of time series data were missed in almost all available stations and hence 15 rainfall and 25 temperature stations which have long time series and relatively short time missing records were selected. Filling missed or gap records was the first task for further meteorological data analysis. This task was done using the well-known methodology of inverse distance weighing method (IDW). To check the quality of the data, the Double Mass Curve analysis (DMC) were used. DMC is a cross correlation

between the accumulated totals of the gauge in question against the corresponding totals for a representative group of nearby gauges.

3.2 Large scale datasets

High uncertainty is expected in climate change impact studies if the simulation result is relied up on a single GCM due to the fact that each GCM has different spatial and temporal resolution with different assumptions of atmospheric processes (Kim *et al.*, 2009). Hence, a new version of the LARS-WG5.5 ~~-~~was applied for this study that incorporates predictions from 15 GCMs which were used in the IPCC's Fourth Assessment Report (AR4) based on Special Emissions Scenarios SRES B1, A1B and A2 for three time windows as listed in Table 1~~Table 1~~. ~~At the time of this study~~However, the fifth phase of Coupled Model Inter Comparison Project (CMIP5) climate models based on the new radiative forcing scenarios (Representative Concentration Pathway, RCP) which were used for IPCC Fifth Assessment Report (AR5) were not incorporated in to it at the time of the study. ~~the new version of LARS WG5.5.~~

~~However, a~~As it is difficult to process all the incorporated 15 CMIP3 GCMs and as it is expected large differences in predictions of climate variables among the GCMs, the~~ir~~ performance of GCMs in simulating the current climate variables of ~~in t~~the study area (UBNRB) ~~in particular and for Ethiopia in general were~~ should be evaluated and best represented GCMs were selected. The MAGICC/SCEGEN computer program tool was used for the performance evaluation of the embedded 15 GCMs in LARS WG5.5 database, ~~–~~as it is a standard method for selecting models on the basis of their ability to accurately represent current climate, either for a particular region and/or for the globe.

In this study, we used a semi-quantitative skill score that rewards relatively good models and penalizes relatively bad models as suggested by user manual Wigley (2008). The statistics used for model selection are pattern correlation (R2), Root mean square error (RMSE), bias (B), and a bias-corrected RMSE (RMSE-corr). The analysis was done separately for precipitation and temperature and finally an average score value was taken for model selection. ~~and s~~Six best performed GCMs have been selected for this study namely: HadCM3, GFDL-CM2.1, ECHAM5-OM, CCSM3, MRI-CGCM2.3.2, and CSIRO-MK3 in the order of their performance ~~by MAGICC/SCEGEN computer program tools as suggested by Wigley (2008)~~ to construct future precipitation, maximum and minimum temperature in the UBNRB for the time period of 2030s, 2050s and 2080s under A1B, A2 and B1 scenarios.

Moreover, ~~A~~atmospheric large scale predictor variables used for representing the present condition were obtained from the National Centre for Environmental Prediction (NCEP) reanalysis data set. CanESM2, second generation Canadian Earth System Model (ESM) developed by Canadian Centre for Climate Modelling and Analysis (CCCma) of Environment Canada that represents ~~the IPCC Fifth Assessment Report (AR5)~~CMIP5 and HadCM3~~, the third version of Atmosphere Ocean General Circulation Model (AOGCM)~~ outputs from the Hadley Centre, United Kingdom(UK) representing ~~AR4~~ CMIP3

were used in SDSM for the construction of daily local meteorological variables corresponding to their future climate scenario.

The reasons for selecting these two GCMs were due to the fact that they are models that made daily predictor variables freely available to be directly fed into SDSM covering the study area with a better resolution. Additionally, they are the most used GCMs in previous studies such as (Dibike *et al.*, 2005; Dile *et al.*, 2013; Hassan *et al.*, 2014; Yimer *et al.*, 2009), and HadCM3 ranked first in performance evolution done by MAGICC/SCEGEN computer program tools and its downscaled results match with the ensemble mean of the six GCMs used in LARS WG model. ~~Moreover~~Furthermore, they can represent two different scenario generations describing the amount of green house gases(GHGs) in the atmosphere in the future. HadCM3 GCM used emission scenarios of A2 (separated world scenario) in which the co2 concentration projected to be 414ppm, 545ppm and 754ppm and B2 (the world of technological inequalities) where the co2 concentration to be expected 406ppm, 486ppm and 581ppm at the time period of 2020s, 2050s and 2080s respectively(Semenov and Stratonovitch, 2010) that were used in the CMIP3 for the IPCC's AR4 (IPCC, 2007). While canESM2 ~~GCM~~ represents the latest and wide range of plausible radiative forcing scenarios, which include a very low forcing level (RCP2.6), where radiative forcing peaks at approximately 3 Wm^{-2}, approximately 490 ppm co2 equivalent before 2100 and then decline to 2.6Wm^{-2}; two medium stabilization scenarios (RCP6 and RCP 4.5) in which radiative forcing is stabilised at 6Wm^{-2} (approximately 850 ppm co2 equivalent) and 4.5 Wm^{-2} (approximately 650 ppm co2 equivalent) after 2100 respectively, and one very high baseline emission scenario (RCP8.5) for which radiative forcing reaches >8.5 Wm^{-2} (1370 ppm co2 equivalent) by 2100 and continues to rise for some time ~~(RCP8.5)~~ that were used for the IPCC's AR5, (IPCC, 2014).

The NCEP dataset were normalized over the complete 1961-1990 period data, and interpolated to the same grid as HadCM3 (2.5° latitude x 3.75° longitude) and canESM2 (2.8125° latitude x 2.8125° longitude) from its horizontal resolution of (2.5° latitude x 2.5° longitude), to represent the current climate conditions. NCEP reanalysis data were normalized and interpolated as (Hassan *et al.*, 2014):

$$un = \frac{(ut-ua)}{\sigma u} \quad \text{.....} \quad (11)$$

In which *un* is the normalized atmospheric variable at time t, *ut* is the original data at time t, *ua* is the multiyear average during the period, and σu is the standard deviation.

The canESM2 outputs ~~were downloaded~~ for three different climate scenarios namely: RCP 2.6, RCP 4.5 and RCP 8.5 for the period ~~1961~~2006-~~2099~~ 2100 while the outputs of HadCM3 ~~were~~ for A2a (medium-high) and B2a (medium-low) emission scenarios ~~of the IPCC Special Report on Emission Scenarios~~ for the period 1961-~~2100~~2099~~. The outputs of the models~~ were obtained on a grid by grid box basis for the study area from the Environment Canada website http://ccds-

dscc.ec.gc.ca/index.php?page=dst-sdi (the "a" in A2a and B2a refers the ensemble member in the HadCM3 A2 and B2 experiments). The archive of canESM2 and HadCM3 GCM output contains 26 daily predictor variables each as listed in Table 3~~Table 3~~.

4. Methodology

4.1 Description of LARS-WG Model

LARS-WG is a stochastic weather generator which can be used for the simulation of weather data at a single station under both current and future climate conditions. These data are in the form of daily time-series for a group of climate variables, namely, precipitation, maximum and minimum temperature and solar radiation (Chen *et al.*, 2013; Semenov *et al.*, 1997). LARS-WG uses a semi-empirical distribution (SED) that is defined as the cumulative probability distribution function(CDF) to approximate probability distributions of dry and wet series, daily precipitation, minimum and maximum temperatures.

$$\boldsymbol{EPM} = \{\boldsymbol{a_0}, \boldsymbol{a_i}, \boldsymbol{h_i}, \boldsymbol{i} = \boldsymbol{0}, \dots, \boldsymbol{10}\} \quad (2~~2~~)$$

EPM is a histogram of the distribution of 23 different intervals (a_{i-1}, a_i) where $a_{i-1} < a_i$ (Semenov et al., 2002), which offers more accurate representation of the observed distribution compared with the 10 used in the previous version. By perturbing parameters of distributions for a site with the predicted changes of climate derived from global or regional climate models, a daily climate scenario for this site could be generated and used in conjunction with a process-based impact model for assessment of impacts.

In general, the process of generating synthetic weather data can be categorized in three distinct steps: model calibration, model validation and scenario generation as represented in Figure 2~~Figure 2~~ (a), which are briefly described by (Semenov *et al.*, 2002) as follows.

The inputs to the weather generator are the series of daily observed data (precipitation, minimum and maximum temperature) of the base period (1984-2011)and site information (latitude, longitude and altitude) are the inputs to the LARSWG. After the input data preparation and quality control, the observed daily weather data at a given site were used to determine a set of parameters for probability distributions of weather variables. These parameters are used to generate a synthetic weather time series of arbitrary length by randomly selecting values from the appropriate distributions, having the same statistical characteristics as the original observed data but differing on a day-to-day basis . The LARS WG distinguishes wet days from dry days based on whether the precipitation is greater than zero. The occurrence of precipitation is modelled by alternating wet and dry series approximated by semi empirical probability distributions. The statistical characteristics of the observed

and synthetic weather data are analyzed to determine if there are any statistically-significant differences using Chi-square goodness of fit test (KS) and the means and standard deviation using t and F test respectively by changing the parameters of LARS-WG (number of years and seed number).

To generate climate scenarios at a site for a certain future period and an emission scenario, the LARS-WG baseline parameters, which are calculated from observed weather for a baseline period (1984-2011), are adjusted by the Δ-changes for the future period and the emissions predicted by a GCM for each climatic variable for the grid covering the site. In this study, the local-scale climate scenarios based on the SRES A2, A1B and B1 scenario simulated by the selected six GCMs are generated for the time periods of 2011–2030, 2046–2065, and 2080–2099 to predict the future change of precipitation and temperature in UBNRB.

Δ-changes were calculated as relative changes for precipitation and absolute changes for minimum and maximum temperatures (Eq. 3 and 4),respectively. No adjustments for distributions of dry and wet series and temperature variability were made, because this would require daily output from the GCMs which is not readily available from LARS WG data set (Semenov *et al.*, 2010).

$$\Delta \boldsymbol{T_i} = \left(\overline{\boldsymbol{T}}_{\boldsymbol{GCM,FUT,i}} - \overline{\boldsymbol{T}}_{\boldsymbol{synt,Base,i}}\right) \quad (3)$$

$$\Delta \boldsymbol{P_i} = \left(\overline{\boldsymbol{P}}_{\boldsymbol{GCM,FUT,i}} \Big/ \overline{\boldsymbol{P}}_{\boldsymbol{synt,Base,i}} \right) \quad (4)$$

In above equations, ΔT_i and ΔP_i are climate change scenarios of the temperature and precipitation, respectively, for long-term average for each month ($1 \leq i \leq 12$); $\overline{T}_{GCM,FUT,I}$ the long term average temperature simulated by the AOGCM in the future periods per month for three time periods; $\overline{T}_{Synth,Base,i}$ is the long term average temperature simulated by the model in the period similar to observation period (in this study 1984-2011) for each month. The above calculations are true for precipitation as well.

For obtaining time series of future climate scenarios, climate change scenarios are added to the observations values by employing the change factor (CF) method (Eq. 5 and 6) (in this study 1984-2011):

$$\boldsymbol{T} = \boldsymbol{T_{obs}} + \Delta \boldsymbol{T} \quad \textbf{(5)}$$

$$\boldsymbol{P} = \boldsymbol{P_{obs}} + \Delta \boldsymbol{P} \quad \textbf{(6)}$$

T and P; time series of the future climate scenarios of temperature and precipitation (2011-2100) and T_{obs} and P_{obs} ; observed temperature and precipitation. So, in LARS-WG downscaling unlike SDSM, large-scale atmospheric variables are not directly used in the model, rather, based on the relative mean monthly changes between current and future periods predicted by a GCM, local station climate variables are adjusted proportionately to represent climate change (Dibike *et al.*, 2005).

4.2 Description of SDSM

The SDSM is best described as a hybrid of the stochastic weather generator and regression based in the family of transfer function methods~~. D~~due to the fact that a multiple linear regression model is developed between a few selected large-scale predictor variables and local-scale predictands such as temperature and precipitation to condition local scale weather parameters from large scale circulation patterns. The stochastic component of SDSM enables the generation of multiple simulations with slightly different time series attributes, but the same overall statistical properties. (Wilby *et al.*, 2002) . It requires two types of daily data, the first type corresponds to local predictands of interest (e.g. temperature, precipitation) and the second type corresponds to the data of large-scale predictors (NCEP and GCM) of a grid box closest to the station.

The SDSM model categorizes the task of downscaling into a series of discrete processes such as quality control and data transformation, screening of predictor variables, model calibration and weather and scenario generation as shown in Figure 2~~Figure 2~~(b). ~~-~~Detail procedures and steps can be found (Wilby *et al.*, 2002) for further reading. Screening potentially useful predictor-predictand relationships for model calibration is one of the most challenging but very crucial stage in the development of any statistical down scaling model. It is because of the fact that the selection of appropriate predictor variables largely determines the success of SDSM and also the character of the downscaled climate scenario (Wilby *et al.*, 2007). After routine screening procedures, the predictor variables that provide physically sensible meaning in terms of their high explained variance, correlation coefficient (r) and the magnitude of their probability (p value) were selected.

The model calibration process in SDSM was used to construct downscaled data based on multiple regression equations given daily weather data (predictand) and the selected predictor variables. The model was structured as monthly model for both daily precipitation and temperature downscaling. Consequently, twelve regression equations were developed for twelve months. Bias correction and variance inflation factor was adjusted until the model replicate the observed data. The weather generator helps to validate the calibrated model ideally using independent data. This operation generates the ensembles of synthetic daily weather data for the specified period with the help of regression model weights along with parameter file prepared during model calibration. To compare the observed and simulated data, SDSM has provided summary statistics function that summarizes the result of both the observed and simulated data. Time series of station data and large scale predictor variable information (NCEP reanalysis data) were divided into two groups; for the period from 1984-1995/ 1984-2000 and 1996-2001/ 2001-2005 for model calibration and validation of HadCM3/canESM2 GCMs respectively.

The Scenario Generator operation produces ensembles of synthetic daily weather series given observed daily atmospheric predictor variables supplied by a GCM either for current or future climate(Wilby *et al.*, 2002). The scenario generation produced 20 ensemble members of synthetic weather data for 139 years (1961-2099) from HadCM3 A2a and B2a scenarios and for 95 years (2006-2100) from canESM2 for RCP2.6, 4.5 and 8.5 scenarios, and the mean of the ensemble members was calculated and used for further analysis. The generated scenario was divided into three time windows of 30 years of data (2011-2040), (2041-2070) and (2071-2100) henceforth called 2030s, 2050s and 2080s, respectively.

~~5. Results and Analysis~~

~~5.1~~4.3 Model performance evaluation criteria

A simulation of mean daily and monthly rainfall, Tmax and Tmin, during the calibration and validation of the SDSM and LARSWG time series were checked by using graphical representation and statistical performance index. Performance indicators such as mean absolute error (MAE), root mean square error RMSE), Bias (B), coefficient of determination (R^2), NasheSutcliffe Model Efficiency (NSE) were used to evaluate the performance of the GCMs and downscaling techniques reproducing the present climate variables of the study area and are defined as;

$$R^2 = \frac{\left[\sum_{i=1}^{n}(X_i-\mu_x)(Y_i-\mu_x)\right]^2}{\sum_{i=1}^{n}(X_i-\mu_x)^2\sum_{i=1}^{n}(Y_i-\mu_x)^2} \quad \text{..... (7~~7~~)}$$

$$MAE = \frac{\sum_{i=1}^{n}|X_i-Y_i|}{n} \quad \text{..... (8~~8~~)}$$

$$RMSE = \sqrt{\sum_{i=1}^{n}(X_i-Y_i)^2} \quad \text{..... (9~~9~~)}$$

$$NSE = 1 - \frac{\frac{1}{n}\sum_{I=1}^{I=n}(X_i-Y_i)^2}{\frac{1}{n}\sum_{i=1}^{i=n}(X_i-\mu_x)^2} \quad \text{..... (10~~10~~)}$$

$$Bias = \frac{\sum_{i01}^{i=n}X_i}{n} - \frac{\sum_{i=1}^{i=n}Y_i}{n} \quad \text{..... (11~~11~~)}$$

$$Relative\ Bias\ (\%) = \frac{1}{n}\sum_{i=1}^{i=n}\frac{(X_i-Y_i)}{X_i} \quad \text{..... (12~~12~~)}$$

In the above equations X_i and Y_i are i-th observation and simulated data by the model, respectively. μ_x and μ_y are the average of all data of X_i and Y_i in the study population and n is the number of all samples to be tested.

5. Results and Analysis

5. ~~2~~ 1 Calibration and validation of LARS-WG

To verify the performance of LARS-WG, in addition to the graphic comparison, some statistical tests were performed. The Kolmogorov–Smirnov (KS) test is performed to test equality of the seasonal distributions of wet and dry series (WDSeries), distributions of daily rainfall (RainD), and distributions of daily maximum (TmaxD) and minimum (TminD) temperature. The F-test is performed on testing equality of monthly variances of precipitation (RMV) while the t test is performed on verifying equality of monthly mean rainfall (RMM), monthly mean of daily maximum temperature (TmaxM), and monthly mean of daily minimum temperature (TminM). ~~calculated from observed and generated data~~. All of the tests calculate a p-value, which is used to accept or reject the hypotheses that the two sets of data (observed and generated) could have come from the same distribution at the 5% significance level . Therefore, number of tests that results a p value less than 5% out of the total number of 8 dry/wet seasons or 12months were recorded for each stations. The average number of P values less than 5% recorded from 26 stations and percentage failed from the total of 8 seasons or 12 months ~~The test results have~~ has been presented in Table 2~~Table 2,. where the numbers show how many tests gave significant different results at the 5% significance level out of the total number of tests of 8 seasons or 12months. A large number indicates a poor performance.~~ It can be seen from Table 2 that LARS WG performs very well for all parameters except RMM and RMV. On the other hand, LARS WG performs poor (i.e. average of 2.2 % and 17.3 % of the months of a year were recorded a P value < 5 %) for the monthly mean and variance of precipitation respectively. ~~It can be seen from Table 2 that the average number of significant different results for seasonal wet and dry series distributions and for the daily rainfall distributions (RainD) were 0 and 1.67 out of 8 and 12 respectively; for the monthly means (RMM) is 0.3 and for the monthly variance (RMV) is 2.1out of 12. The average numbers of significant results for TminD, TminM, TmaxD, and TmaxM are zero or close to zero.~~ From these numbers, it can be noted that the model is ~~more~~ less capable in simulating ~~the monthly means and the daily rainfall distributions of each month in comparison to~~ the monthly variances than the other parameters.

For illustrative purpose, graphical representation of monthly mean and standard deviation of the simulated and observed precipitation, Tmax and Tmin were constructed in Figure 3~~Figure 3~~ for randomly chosen Gondar station as it has been difficult to present the result of all stations. It can be seen from Figure 3~~Figure 3~~ that observed and simulated monthly mean precipitation, Tmax and Tmin matches very well. However, as it is known for being difficult to simulate the standard deviations in most statistical downscaling studies, the performance of the standard deviation is less accurate as compared to the mean (Figure 3~~Figure 3~~ b) . Generally, according to the obtained statistical performance measure values and from graphical representation, the performance of the model for simulation ~~and prediction~~ of the three climatic variables in all stations across UBNRB is acceptable and reasonably well.

5.3 2 Down scaling with LARS-WG

The result of precipitation prediction by using LARS-WG model from six multi model GCMs under three scenarios (A1B, B1 and A2) for three time periods were presented in Table 5 and plotted in Figure 4Figure 4 for illustrative purpose. In Figure 4Figure 4, each box–whisker plot represents the prediction of precipitation across all stations of UBNRB under a single scenario for each GCM and the result revealed that there are no coherent change trends observed among various GCMs' for predicting precipitation. NCCCSM GCM was found the most unstable GCM in predicting precipitation across UBNRB stations particularly under A2 scenario at the time period of 2080s while MPEH5 was relatively stable across all stations as compared to others.

After downscaling the future climate predictions at all stations from the selected six GCMs, the projected precipitation analysis for the areal UBNRB was calculated from the point rainfall stations using Thiessen polygon method. The result analysis revealed that, GCMs disagree on the direction of precipitation change, two GCMs (MIHR and GFCM21) result decreasing trend whereas a majority or four GCMs (NCCSM, Hadcm3, MPEH5 and MIHR) result increasing trend from the reference period in all three time periods. The results from Figure 5Table 5 showed that NCCCS reported maximum increase while GFCM21 reported highest reduction. From the model average, B1 scenario projected maximum increase of mean annual relative change of precipitation at 2030s and 2050s while A2 scenario for 2080s. For 2030s, the relative change of mean annual precipitation projected between (-2.3 % and + 6.5 %) for A1B, (-2.3 % and +7.8 %) for B1 and (-3.7 % and +6.4 %) for A2 emission scenarios. At 2050s, the relative change in precipitation range from (-8 % and +22.7 %) for A1B, (-2.7 % and +22 %) for B1 and (-7.4 % and +8.7 %) for A2 scenarios. In the time of 2080s, the relative change of precipitation projected may vary from(-7.5 % and +29.9 %) for A1B, (-5.3 % and +13.7 %) for B1 and (-5.9 % and +43.8 %) for A2 emission scenarios. The multi model average relative change mean annual precipitation result showed that in the future precipitation may generally increases over the basin in the range of 1 %-14.4 % which is consistent matches with the result from HadCM3 GCM (0.8 %-16.6 %) as it is shown in Figure 5Table 5 .

In a different way from precipitation, the projection of mean annual Tmax and Tmin have coherent increasing change trends were observed from all the six GCMs under all scenarios in all three future time periods. At 2080s, for A1b, B1 and A2 scenarios, the change in mean annual Tmax and Tmin is more pronounced than 2030s in all GCMs from three scenarios. At 2080s, 2050s and 2030s, the mean annual Tmax may increase up to +4.3^{o}c, +2.1^{o}c and +0.7^{o}c for A2, A1B and A1B scenarios respectively. The result calculated from the ensemble mean showed that mean annual Tmax my increase up to +0.5 ^{o}c, +1.8 ^{o}c and +3.6 ^{o}c by 2030s, 2050s and 2080s respectively under A2 scenario which is consistent in line with the results of from both GFCM21 and HadCM3 GCMs (Figure 5). Likewise, UBNRB may experience an increase mean annual Tmin may increase up to +4.1^{o}c, +1.9^{o}c and +0.7^{o}c at 2080s, 2050s and 2030s respectively whereas the result calculated from the

~~multi model average showed that mean annual Tmin change may reach up to~~ +0.6 °c, +1.8 °c and +3.6 °c by 2030s, 2050s and 2080s respectively from the multi model average.

5. ~~4~~ 3 Screening variable, model calibration and validation of SDSM

Initially, offline correlation analysis was performed using SPSS software between predictands and NCEP re-analysis predictors to identify an optimal lag and physically sensible predictors for climate variables of precipitation, Tmax and Tmin. Analysis of the offline correlation revealed that ~~in most stations and predictors~~ an optimal lag or time shift does not improve the correlation of predictands and predictors for this particular study. Average partial correlation of precipitation of all stations with predictors is shown in the Figure 7~~Figure 5~~ which indicates all stations followed the same correlation pattern ~~with predictors~~ (both in magnitude and direction) that illustrates all stations can have identical physically sensible predictors with a few of exceptions. Furthermore, there exist a number of predictors that have correlation coefficient values in the range of 20 %-45 % for precipitation across all stations as shown in Figure 7~~Figure 5~~. This range is considered to be acceptable when dealing with precipitation downscaling (Wilby *et al.*, 2002) because of its complexity and high spatial and temporal variability to downscale.

~~The predictands of all stations and NCEP reanalysis predictors close to or within the stations are then used in SDSM for the final analysis.~~ The predictor variables identified for each downscaling GCMs and for the corresponding local climate variables conducted in this study are summarized in Figure 8~~Figure 6~~. From the selected predictors, it is observed that different large scale atmospheric variables control different local variables. For instance, set of temp, mslp, s500, s850, p8_v, p500, shum are the most potential or meaningful predictors for temperature and s500, s850, p8_u, p_z, pzh, p500 for precipitation of the study area respectively, which is consistent with the result of offline correlation analysis.

The graphical comparison between the observed and generated rainfall, Tmax and Tmin were run to enhance the confidence of the model performance, as shown in Figure 9~~Figure 7~~ and Figure 10~~Figure 8~~ for Gondar station only. Examination of Figure 9~~Figure 7~~ showed that the calibrated model reproduces the monthly mean precipitation and mean and standard deviation of daily Tmax, Tmin, and mean dry-spell length values quite well. However, the wet-spell length were consistently underestimated and also less accurate in reproducing variance of observed precipitation. As Wilby *et al.* (2004) point out, downscaling models are often regarded as less able to model the variance of the observed precipitation with great accuracy.

Furthermore, the performance of the model was evaluated by statistical performance indicators of (MAE, RMSE, R^2, NSE and BIAS) as summarized in Table 4~~Table 4~~. The result of statistical analysis revealed that the model is much better in simulating Tmax and Tmin than precipitation, because of the high dynamical properties of precipitation makes it difficult to simulate. After accomplishing a satisfactory calibration **(Figure 9)**, the multiple regression model is validated using an

independent set of data outside the period for which the model is calibrated as discussed under section 4, and the results obtained are shown in Figure 9~~Figure 7~~ and Table 4~~Table 4~~. Examination of ~~the~~ Figure 9~~Figure 7~~, Figure 10~~Figure 8~~ and Table 4~~Table 4~~ revealed that the model is successfully validated but at lesser accuracy as compared to calibration for both GCMs. In general, the result analysis of performance measure and graphical representation of observed and simulated both for calibration and validation revealed that the model performs very well in simulating the climate variables with high degree of accuracy.

5.~~5~~ 4 Down scaling with SDSM

Results of down scaling future climate scenario of areal UBNRB using SDSM calculated from all stations using Thiessen polygon methods are summarized from Figure 6~~Table 6~~. The magnitude of future climate change at each station has different pattern and magnitude using different scenarios as can be seen the variation in Figure 11~~Figure 10 and 12~~. The overall analysis of the result of the whole UBNRB from Figure 6 ~~the~~ ~~table~~ indicates, a general increase in mean annual precipitation for three time windows (2030s, 2050s and 2080s) under all scenarios in the range of 2.1 % to 43.8 %~~ under the A2a and RCP8.5 scenarios respectively~~. ~~At 2080s, t~~The maximum/~~-~~minimum relative change of mean annual precipitation is projected to be 43.8 %/2.1 %, 29.5 %/3.5 % and 19 %/2.1% at 2080s, 2050s and 2030s under RCP8.5 scenario of canESM2 ~~GCM while the minimum relative change of mean annual precipitation to be~~ and ~~2.1% under~~ H3B2a scenario of HadCM3 ~~GCM. At 2050s, the maximum and minimum relative mean annual relative change to be 29.5% and 3.5% under RCP8.5 of canESM2 GCM and H3B2a of HadCM3 GCM scenarios respectively. Mean while, at 2030s, the maximum and minimum relative change of mean annual precipitation projected to be 19% and 2.1% under RCP8.5 of canESM2 and H3B2a of HadCM3 scenarios~~ respectively ~~respectively.~~ ~~In~~In general, RCP8.5 scenario of canESM2~~-~~GCM resulted pronounced increase in all three time periods whereas scenario B2a of HadCM3 GCM reported minimum change over the study area.

~~The result analysis at monthly basis as shown from Figure 9 revealed that, canESM2 GCM reported increasing mean monthly precipitation in all months of the year except January, February and March whereas HadCM3 result showed both increasing and decreasing pattern over the whole months of a year. Mean monthly precipitation may increase with a maximum value of 214.5% in the month of October for canESM2 under RCP8.5 by 2080s and may reduce at maximum of 26.6% in the month of February under scenario RCP2.6 by 2080s. The result also indicates, HadCM3 GCM predicted relatively small increase as compared to canESM2, with maximum value of 46.6% in the month of November under scenario A2a by 2020s and maximum reduction in the month of May by 30.4% under scenario A2a by 2080s. Furthermore, the spatial variability analysis of the result showed, the Eastern part may get a more pronounced increase in precipitation compared to the Western part of the study area (Figure 11). Seasonally, both HadCM3 and canESM2 GCM reported precipitation may increase in Summer and Autumn in the range of 1.3% to 27.6% and 8.4% to 89% respectively. However, result from HadCM3 showed reduction of precipitation in spring across the three time windows with a value range from~~

~~10.5% to -21.7%. Also, canESM2 reported precipitation reduction in winter season in the range of -0.9% to -6.1% as summarized in Figure 9.~~

Regarding to temperature, the down scaling result of Tmax and Tmin showed increasing trend consistently in all months, seasons in three time periods under all scenarios with mean annual value ranges from 0.5 ^{o}C to 2.6 ^{o}C and 0.3 ^{o}c to 1.6 ^{o}C under scenario RCP8.5 and H3B2a respectively. RCP 8.5 scenario reported maximum change while H3B2a scenario reported minimum change both for Tmax and Tmin in all three time periods as compared to other scenarios. The analysis of down scaling result illustrates maximum temperature may become much hotter as compared to minimum temperature in all scenarios and time periods in the future across UBNRB.

5.~~6~~ 5 Comparative performance ~~and downscaling results~~evaluation of LARS-WG and SDSM

Chen *et al.* (2013) argued that though major source of uncertainty are linked to GCMs and emission scenarios, uncertainty related to the choice of downscaling methods give less attention on climate change analysis. Therefore, in this study, comparative performance evaluation of the downscaling methods was done and ~~T~~the average partial Pearson's correlation coefficient (R) values of the observed and simulated for all stations ~~as~~ presented in Figure 12~~Figure 12,~~ F~~f~~or precipitation, at daily time series R value was 0.21 in LARS WG while 0.43/042 for HadCM3/canESM2 using SDSM. ~~for HadCM3 and 0.42 for canESM2~~ W~~w~~hereas, R value for daily Tmax were 0.61 using LARS WG and 0.75/~~ and~~ 0.76 ~~using SDSM~~ for HadCM3/~~ and~~ canESM2 using SDSM respectively. ~~–~~The R value for precipitation at monthly basis has improved significantly to 0.79 using LARS WG while 0.84 for both HadCM3 and canESM2 using SDSM. ~~and f~~For Tmax R value was 0.89 using LARS WG and 0.91/~~ and~~ 0.92 for HadCM3/~~ and~~ canESM2 using SDSM respectively. ~~In general, t~~The result from the two downscaling models suggested that both SDSM and LARS-WG approximate the observed climate data corresponding to the current state reasonably well.

~~However, LARS-WG under estimated the standard deviations of the three climatic variables largely for most of the months of the year (poor at modelling inter annual variability), and less performance in simulating daily time series of climate variables as compared to SDSM.~~

For future simulation, HadCM3 GCM under A2 scenario was used in common for two (LARS WG and SDSM) down scaling methods for comparison purpose. The results obtained from the two down scaling models were found reasonably comparable and both approaches showed increasing trend for precipitation, Tmax and Tmin. However, the magnitude ~~analysis~~ of the downscaled climate data from the two techniques as presented in Figure 14~~Figure 13~~ showed different.~~,~~ LARS-WG ~~resulted in a relatively higher increase~~over predicts than SDSM~~,~~. ~~which does not lead to identical conclusions~~ . LARS-WG predict relative change of mean annual precipitation about 16.1 % and an average increase in mean annual Tmax and Tmin about 3.7 ^{o}C and 3.6 ^{o}C respectively at 2080s while SDSM predicts relative change in mean annual precipitation

only about 9.7 % and an average increase in Tmax and Tmin about 2 °c and 1.3 °C respectively in the same period. The difference ~~in the down scaled climate variables~~ could be due to the fact that SDSM uses large scale predictor variables from GCM outputs to predict local scale climate variables using multiple linear regression, while the LARS WG is analysed by applying the change factors from the GCM output of only those variables which directly correspond to the predictands. Moreover, because of the well known fact that GCMs are not very reliable in simulating precipitation, the error induced from the GCM output for precipitation will propagate the error of downscaling that makes the performance of LARS-WG to downscale precipitation more questionable (Dibike et al., 2005). Therefore, based on the above facts SDSM would be more robust and can be applied at higher confidence for downscaling large scale GCMs outputs to finer scales to suit for hydrological models for impact assessment in the UBNRB.

6. Discussions and conclusions

The uncertainty related to climate change analysis can be due to climate models and downscaling methods among many others. In this study, we employed multi model approach to see the uncertainties came from different GCMs. In total, 21 systematically selected future climate scenarios were produced for each time period, which we might think representative to understand fully and to project plausibly the future climate change in the study area and to retain information about the full variability of GCMs. Moreover, ~~W~~we applied two widely used statistical down scaling methods, namely the regression downscaling technique (SDSM) and the stochastic weather generation method (LARS WG) for this particular study.

The down scaling result ~~analysis~~ reported from the six GCMs used in LARS-WG showed large inter model differences, 2 GCMs reported precipitation may decrease while 4 GCMs reported precipitation may increase in the future. The large inter model differences of the GCMs ~~on the direction of future precipitation~~ showed the uncertainties of GCMs associated with their differences of resolution and assumptions of physical atmospheric processes to represent local scale climate variables which are typical characteristics for Africa and because of low convergence in climate model projections in the area of UBNRB (Gebre *et al.*, 2014). ~~In 2030s, the relative change in mean annual precipitation projected may vary from -3.7% to 7.8%. At 2050s and 2080s, the relative change in mean annual precipitation projected between -8% to +22.7% and -7.5% to +43.8% respectively.~~ The multi model average result showed that in the future precipitation may generally increases over the basin in the range of 1 %-14.4 % which agreed with the result from HadCM3 GCM (0.8 %-16.6 %), ~~However, the multi model average of the six GCMs showed increasing pattern for precipitation, Tmax and Tmin and the magnitude of the projection which has better agreement with HadCM3 GCM projection which~~this indicates HadCM3from CMIP3 GCMs has a better representation of local scale climate variables in the study area consistent with the previous study result by Kim *et al.* (2009) and (Dile *et al.*, 2013) in the same study area.

~~Furthermore~~ Further uncertainty analysis of~~,~~ HadCM3 GCM from CMIP3 ~~applied in IPCCAR4~~ and canESM2 GCM from CMIP5 ~~applied in IPCC AR5 were~~ used by SDSM was carried out for precipitation. The downscaled results from the two GCMs modelled by SDSM suggested that mean annual precipitation may generally increase in the range of 2.1 % to 43.8 %. ~~For comparison, the performance indicator result presented in Table 4 has shown, the ratio of MAE relative to mean for daily precipitation was 9.7% and 5.3% while for monthly precipitation was 9.1% and 4.8% respectively for HadCM3 and canESM2 GCMs. The ratio of RMSE relative to mean for daily precipitation was 12.6 % and 8.6 % while for monthly precipitation it was 13% and 7.3% respectively for HadCM3 and canESM2 GCMs during calibration period. During validation period these values were reported 25.2% /22.9% and 26.6% /19.1% for the ratio of MAE to mean for daily and monthly precipitation using HadCM3/canESM2 GCMs respectively whereas 30.7%/ 30% and 38.1%/27.4% for the ratio of RMSE to mean. The ratio of MAE and RMSE to mean is less than 5% both for daily and monthly Tmax and Tmin using the mentioned two GCMs by SDSM techniques. In general, the result suggested that~~ However, canESM2 better performs than HadCM3 in reproducing the current climate variables of UBNRB both in calibration and validation consistently (Table 4).~~.~~ The better performance of canESM2 could be due to the fact that modelling was based on the new set of radiative forcing scenario (RCP) that replaced SRES emission scenarios, constructed for IPCC AR5 where the impacts of land use and land cover change on the environment and climate is explicitly included. Also, it is one of the earth system models which has additional features that incorporates various important biogeochemical processes which is the limitation of CMIP3 GCMs like HadCM3. ~~Further performance statistical analysis revealed that e~~Even though, the simulation of large scale precipitation has improved since IPCC AR4, GCMs still continues to perform less well for precipitation as compared to temperature and therefore downscaling of precipitation becomes more complex and difficult to reproduce the base scenario as compared to downscaling of temperature (Fowler et al., 2007) also confirmed in this study (Table 4). However, a direct comparison between the projection from the two datasets (HadCM3 and canESM2) is not possible as seen from ~~Table 6~~Figure 6, as they used different scenarios describing the amount of Green House Gases (GHGs) in the atmosphere differently.

LARS WG as it is a stochastic simulation tool that is commonly used to produce synthetic climate data of any length with the same characteristics as the input record, it simulate weather separately for single sites; therefore, the resulting weather series for different sites are independent of each other, which can lost a very strong spatial correlation that exists in real weather data during simulation. Although, a few stochastic models have been developed to produce weather series simultaneously at multiple sites preserving the spatial correlation, mainly for daily precipitation, such as space–time models, non-homogeneous hidden Markov model and nonparametric models typically use a K-Nearest Neighbour (K-NN) procedure (King et al., 2015), they are complicated in both calibration and implementation and are unable to adequately reproduce the observed correlations (Khalili et al., 2007).

To test the capability of LARS WG spatial correlation of stations, the simple Pearson's correlation coefficient (R^2) value was calculated for two stations Abaysheleko and Bahirdar and the result from Figure **13** revealed that the spatial correlation of

the stations was distorted /decreased/ from the original to a lesser extent as expected. Even though, LARS WG has limitation to preserve the spatial correlation of climate variables, it can be applied for downscaling climate change scenario for the area of UBNRB satisfactorily with caution to hydrological impact models, as spatial distribution of precipitation may have essential effects on the discharge of a river and the formation of floods.

In conclusion, a multi model average result of this study showed a general increasing trend for all three climatic variables (precipitation, Tmax and Tmin) in all three time periods applying ~~multi model average from~~ LARS-WG and SDSM downscaling techniques. The positive change of precipitation in future can be a good opportunity for the farmers who are engaged in rain fed agriculture to maximize their agricultural production and to change their lively hoods. However, this information cannot be a guarantee for irrigation farming because precipitation is not the only factor contributing to affect the flow of the river, which is the main source ~~of~~ for irrigation. Evapotranspiration, dynamics of land use land cover, proper water resource management and other climatic factors, which are not yet assessed by this study can influence the flow of the river directly and indirectly. Furthermore, the result from this study ~~Figure 9 and~~ (Figure 11~~Figure 10 (only RCP4.5 at 2050s)~~ revealed that, maximum positive precipitation change may occur in Autumn (Sep.-Nov.) when most agricultural crops get matured and start harvesting~~)~~ while minimum precipitation change may occur during summer (June-August), when about 80% of the annual rainfall occurred, this climate variability can be potential threat for the farmers, who have limited ability to cope with the negative impacts of climate variability ~~agricultural production~~ and overall ongoing economic development efforts in the basin ~~as the crops do not require much rain when they are matured~~.

In general, ~~the results of future climate change from multi model GCMs and applying two widely used downscaling techniques for all three climatic variables~~this study ~~have~~ has shown that climate change will occur plausibly that may affect the water resources and hydrology of the UBNRB, and the study proposed ~~so~~ the outputs of canESM2 ~~GCMs~~ESM with new sets of emission scenarios downscaled by SDSM technique can be applied for further impact analysis with high degree of certainty. ~~–~~ Moreover, the paper provides information that the choice of downscaling methods has a contribution in the estimation of climate change and has to be done in the same way as we consider multiple GCMs to acknowledge the uncertainty range.

References

Awulachew, S.B. *et al.*, 2007. Water resources and irrigation development in Ethiopia, 78p. (Working Paper 123). Colombo, Sri Lanka: International Water Management Institute.

BCEOM, 1998. Abbay river basin integrated development master plan, section II, volume V-water resources development. Ministry of Water Resources, Addis Ababa, Ethiopia(part 1- irrigation and drainage).

Bewket, W., Conway, D., 2007. A note on the temporal and spatial variability of rainfall in the drought-prone Amhara region of Ethiopia. International Journal of Climatology, 27(11): 1467-1477.

Beyene, T., Lettenmaier, D.P., Kabat, P., 2010. Hydrologic impacts of climate change on the Nile River Basin: implications of the 2007 IPCC scenarios. Climatic change, 100(3-4): 433-461.

Casajus, N. *et al.*, 2016. An Objective Approach to Select Climate Scenarios when Projecting Species Distribution under Climate Change. PloS one, 11(3): e0152495.

Chen, H., Guo, J., Zhang, Z., Xu, C.-Y., 2013. Prediction of temperature and precipitation in Sudan and South Sudan by using LARS-WG in future. Theoretical and Applied Climatology, 113(3-4): 363-375.

Conway, D., 2000. The climate and hydrology of the Upper Blue Nile River. The Geographical Journal, 166(1): 49-62.

Dibike, Y.B., Coulibaly, P., 2005. Hydrologic impact of climate change in the Saguenay watershed: comparison of downscaling methods and hydrologic models. Journal of hydrology, 307(1): 145-163.

Dile, Y.T., Berndtsson, R., Setegn, S.G., 2013. Hydrological Response to Climate Change for Gilgel Abay River, in the Lake Tana Basin-Upper Blue Nile Basin of Ethiopia. PloS one, 8(10): e79296.

Ebrahim, G.Y., Jonoski, A., van Griensven, A., Di Baldassarre, G., 2013. Downscaling technique uncertainty in assessing hydrological impact of climate change in the Upper Beles River Basin, Ethiopia. Hydrology Research, 44(2): 377-398.

Elshamy, M.E., Seierstad, I.A., Sorteberg, A., 2009. Impacts of climate change on Blue Nile flows using bias-corrected GCM scenarios. Hydrology and Earth System Sciences, 13(5): 551-565.

Fiseha, B., Melesse, A., Romano, E., Volpi, E., Fiori, A., 2012. Statistical downscaling of precipitation and temperature for the Upper Tiber Basin in Central Italy. International Journal of Water Sciences, 1.

Fowler, H., Blenkinsop, S., Tebaldi, C., 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. International journal of climatology, 27(12): 1547-1578.

Gebre, S.L., Ludwig, F., 2014. Hydrological Response to Climate Change of the Upper Blue Nile River Basin: Based on IPCC Fifth Assessment Report (AR5). Journal of Climatology & Weather Forecasting, 2015.

Gebremicael, T., Mohamed, Y., Betrie, G., van der Zaag, P., Teferi, E., 2013. Trend analysis of runoff and sediment fluxes in the Upper Blue Nile basin: A combined analysis of statistical tests, physically-based models and landuse maps. Journal of Hydrology, 482: 57-68.

Goly, A., Teegavarapu, R.S., Mondal, A., 2014. Development and evaluation of statistical downscaling models for monthly precipitation. Earth Interactions, 18(18): 1-28.

Goodarzi, E., Dastorani, M., Talebi, A., 2015. Evaluation of the change-factor and LARS-WG methods of downscaling for simulation of climatic variables in the future (Case study: Herat Azam Watershed, Yazd-Iran). Ecopersia, 3(1): 833-846.

Hashmi, M.Z., Shamseldin, A.Y., Melville, B.W., 2011. Comparison of SDSM and LARS-WG for simulation and downscaling of extreme precipitation events in a watershed. Stochastic Environmental Research and Risk Assessment, 25(4): 475-484.

Hassan, Z., Shamsudin, S., Harun, S., 2014. Application of SDSM and LARS-WG for simulating and downscaling of rainfall and temperature. Theoretical and applied climatology, 116(1-2): 243-257.

IPCC, 2007. Climate Change 2007. Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, Pachauri, R.K and Reisinger, A.(eds.)].

IPCC, 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

Khalili, M., Leconte, R., Brissette, F., 2007. Stochastic multisite generation of daily precipitation data using spatial autocorrelation. Journal of hydrometeorology, 8(3): 396-412.

Khan, M.S., Coulibaly, P., Dibike, Y., 2006. Uncertainty analysis of statistical downscaling methods. Journal of Hydrology, 319(1): 357-382.

Kim, U., Kaluarachchi, J.J., 2009. Climate change impacts on water resources in the Upper Blue Nile River Basin, Ethiopia1. Wiley Online Library.

Kim, U.K., J. J.; Smakhtin, V. U., 2008. Climate change impacts on hydrology and water resources of the Upper Blue Nile River Basin, Ethiopia. Colombo, Sri Lanka: International Water Management Institute, 27p (IWMI Research Report 126).

King, L.M., McLeod, A.I., Simonovic, S.P., 2015. Improved weather generator algorithm for multisite simulation of precipitation and temperature. JAWRA Journal of the American Water Resources Association, 51(5): 1305-1320.

Maurer, E.P., Hidalgo, H.G., 2008. Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods. Hydrology and Earth System Sciences, 12(2): 551-563.

Qian, B., Hayhoe, H., Gameda, S., 2004. Evaluation of the stochastic weather generators LARS-WG and AAFC-WG for climate change impact studies. Climate Research, 29(1): 3.

Semenov, M.A., Barrow, E.M., 1997. Use of a stochastic weather generator in the development of climate change scenarios. Climatic change, 35(4): 397-414.

Semenov, M.A., Barrow, E.M., Lars-Wg, A., 2002. A stochastic weather generator for use in climate impact studies. User Manual: Hertfordshire, UK.

Semenov, M.A., Stratonovitch, P., 2010. Use of multi-model ensembles from global climate models for assessment of climate change impacts. Climate research (Open Access for articles 4 years old and older), 41(1): 1.

Taye, M.T., Ntegeka, V., Ogiramoi, N., Willems, P., 2011. Assessment of climate change impact on hydrological extremes in two source regions of the Nile River Basin. Hydrology and Earth System Sciences, 15(1): 209-222.

USBR, 1964. Land and water resources of the Blue Nile basin, Ethiopia. Main report. United States Bureau of Reclamation, Washington, DC.

WAPCOS, 1990. Preliminary water resources development master plan for Ethiopia. Prepared for EVDSA. Addis Ababa, Final report.

Wigley, T.M., 2008. MAGICC/SCENGEN 5.3: User manual (version 2). NCAR, Boulder, CO, 80.

Wilby, R. *et al.*, 2004. Guidelines for use of climate scenarios developed from statistical downscaling methods.

Wilby, R., Dawson, C., Barrow, E., 2007. SDSM user manual-a decision support tool for the assessment of regional climate change impacts.

Wilby, R.L., Dawson, C.W., 2013. The statistical downscaling model: insights from one decade of application. International Journal of Climatology, 33(7): 1707-1719.

Wilby, R.L., Dawson, C.W., Barrow, E.M., 2002. SDSM—a decision support tool for the assessment of regional climate change impacts. Environmental Modelling & Software, 17(2): 145-157.

Wilby, R.L., Wigley, T., 1997. Downscaling general circulation model output: a review of methods and limitations. Progress in Physical Geography, 21(4): 530-548.

Worqlul, A.W. *et al.*, 2014. Comparison of rainfall estimations by TRMM 3B42, MPEG and CFSR with ground-observed data for the Lake Tana basin in Ethiopia. Hydrology and Earth System Sciences, 18(12): 4871-4881.

Xu, C.-y., 1999. From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. Progress in Physical Geography, 23(2): 229-249.

Yates, D.N., Strzepek, K.M., 1998a. An assessment of integrated climate change impacts on the agricultural economy of Egypt. Climatic Change, 38(3): 261-287.

Yates, D.N., Strzepek, K.M., 1998b. Modeling the Nile Basin under climatic change. Journal of Hydrologic Engineering, 3(2): 98-108.

Yimer, G., Jonoski, A., Van Griensven, A., 2009. Hydrological response of a catchment to climate change in the upper Beles river basin, upper blue Nile, Ethiopia. Nile Basin Water Engineering Scientific Magazine, 2: 49-59.

Table 1~~1~~: Selected Global climate models from IPCC AR4 incorporated into the LARS-WG

Research centre	Country	GCM	Model acronym	Grid Resolution	Emission Scenarios	Time Periods
Common Wealth Scientific and Industrial Research Organization	Australia	CSIRO-MK3	CSMK3	1.9x1.9^o	A1B, B1	B,T1,T2,T3
Max-Plank Institute for Meteorology	Germany	ECHAM5-OM	MPEH5	1,9x1.9^o	A1B,A2,B1	B,T1,T2,T3
National Institute for Environmental Studies	Japan	MRI-CGCM2.3.	MIHR	2.8x2.8^o	A1B,B1	B,T1,T2,T3
UK Meteorological Office	UK	HadCM3	HADCM3	2.5x3.75^o	A1B,A2,B1	B,T1,T2,T3
Geophysical Fluid Dynamics Lab	USA	GFDL-CM2.1	GFCM21	2x2.5^o	A1B,A2,B1	B,T1,T2,T3
National Centre for Atmospheric Research	USA	CCSM3	NCCCS	1.4x1.4^o	A1B,B1	B,T1,T2,T3

B: baseline; T1: 2011–2030; T2: 2046–2065; T3: 2081–2100

Table 2~~2~~**:** Results of the average statistical tests comparing the observed data ~~for~~ from 26 ~~sites~~ stations with synthetic data generated through LARS-WG. The numbers in the table show ~~how many~~the average numbers of tests gave P value less than ~~significant results at the~~ 5 % significance level.

~~No.~~	~~Stations~~	~~KS test~~		~~t test~~	~~F test~~	~~KS test~~	~~t test~~	~~KS test~~	~~t test~~
~~-~~	~~-~~	~~WDseries~~	~~RainD~~	~~RMM~~	~~RMV~~	~~TminD~~	~~TminM~~	~~TmaxD~~	~~TmaxM~~
~~1~~	~~Abaysheleko~~	~~0~~	~~0~~	~~1~~	~~5~~	~~0~~	~~0~~	~~0~~	~~0~~
~~2~~	~~Adet~~	~~0~~	~~0~~	~~0~~	~~1~~	~~0~~	~~0~~	~~0~~	~~0~~
~~3~~	~~Alemketema~~	~~0~~	~~0~~	~~0~~	~~1~~	~~0~~	~~0~~	~~0~~	~~0~~
~~4~~	~~Anger~~	~~0~~	~~0~~	~~0~~	~~5~~	~~0~~	~~0~~	~~0~~	~~1~~
~~5~~	~~Angerguten~~	~~0~~	~~0~~	~~0~~	~~7~~	~~0~~	~~0~~	~~0~~	~~0~~
~~6~~	~~Arjo~~	~~0~~	~~0~~	~~0~~	~~4~~	~~0~~	~~0~~	~~0~~	~~0~~
~~7~~	~~Ayehu~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~
~~8~~	~~Ayira~~	~~0~~	~~0~~	~~0~~	~~2~~	~~0~~	~~1~~	~~0~~	~~1~~
~~9~~	~~Bahirdar~~	~~0~~	~~0~~	~~1~~	~~3~~	~~0~~	~~0~~	~~0~~	~~0~~
~~10~~	~~Bedele~~	~~0~~	~~0~~	~~1~~	~~1~~	~~0~~	~~0~~	~~0~~	~~0~~
~~11~~	~~Dangila~~	~~0~~	~~0~~	~~1~~	~~2~~	~~0~~	~~0~~	~~0~~	~~0~~
~~12~~	~~Dbirhan~~	~~0~~	~~0~~	~~0~~	~~1~~	~~0~~	~~0~~	~~0~~	~~0~~
~~13~~	~~Dedesa~~	~~0~~	~~0~~	~~0~~	~~2~~	~~0~~	~~0~~	~~0~~	~~1~~
~~14~~	~~Dmarkos~~	~~0~~	~~0~~	~~0~~	~~1~~	~~0~~	~~0~~	~~0~~	~~0~~
~~15~~	~~Dtabor~~	~~0~~	~~0~~	~~1~~	~~3~~	~~0~~	~~0~~	~~0~~	~~0~~
~~16~~	~~Fitche~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~
~~17~~	~~Gatira~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~
~~18~~	~~Gidayana~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~
~~19~~	~~Gimijabet~~	~~1~~	~~0~~	~~1~~	~~3~~	~~-~~	~~-~~	~~-~~	~~-~~
~~20~~	~~Gondar~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~
~~21~~	~~Motta~~	~~0~~	~~0~~	~~0~~	~~1~~	~~0~~	~~0~~	~~0~~	~~0~~
~~22~~	~~Mselam~~	~~0~~	~~0~~	~~0~~	~~2~~	~~0~~	~~1~~	~~0~~	~~0~~
~~23~~	~~Nedjo~~	~~0~~	~~0~~	~~0~~	~~3~~	~~0~~	~~0~~	~~0~~	~~0~~
~~24~~	~~Shambu~~	~~0~~	~~0~~	~~1~~	~~2~~	~~0~~	~~0~~	~~0~~	~~0~~
~~25~~	~~Yetnora~~	~~0~~	~~0~~	~~0~~	~~2~~	~~0~~	~~1~~	~~0~~	~~0~~
~~26~~	~~Zege~~	~~0~~	~~0~~	~~0~~	~~3~~	~~0~~	~~0~~	~~0~~	~~0~~
~~-~~	~~Average~~	~~0.04~~	~~0~~	~~0.27~~	~~2.08~~	~~0~~	~~0.12~~	~~0~~	~~0.12~~
~~-~~	~~No. of tests~~	~~8~~	~~12~~	~~12~~	~~12~~	~~12~~	~~12~~	~~12~~	~~12~~

Tests	KS-test		t-test	F-test	KS-test	t-test	KS-test	t-test
Parameters	WDseries	RainD	RMM	RMV	TminD	TminM	TmaxD	TmaxM
Average	0.04	0.00	0.27	2.08	0	0.12	0	0.12
Total	8	12	12	12	12	12	12	12
% failed	0.48	0.00	2.24	17.31	0	1	0	1

Table 3~~3~~: Name and description of all NCEP predictors on HadCM3 & canESM2 grid

Variables	Descriptions	variables	Descriptions
temp	Mean temperature at 2 m	s500 +	Specific humidity at 500 hpa height
mslp	Mean sea level pressure	s850+	Specific humidity at 850 hpa height
p500	500 hpa geopotential height	**_f	Geostrophic air flow velocity
p850	850 hpa geopotentail height	**_z	Vorticity
rhum *	Near surface relative humidity	**_u	Zonal velocity component
r500*	Relative humidity at 500 hpa	**_v	Meridional velocity component
r850*	Relative humidity at 850 hpa	**zh	Divergence
shum	Near surface specific humidity	**thas	Wind direction
Prec+	Total precipitation		

(**) refers to different atmospheric levels: the surface (p_), 850 hpa height (p8), and 500 hpa height (p5)
(*) refers predictors only found from HadCM3, (+) refers predictors only for canESM2

Table 4~~4~~: Average values of statistical performance indicators for all stations using SDSM

	Performance indicators	Climate variables	Mean Daily Precipitation (mm)		Mean monthly Precipitation (mm)		Mean daily Tmax (^{o}C)		Mean dailyTmin (^{o}C)	
		GCMs	HadCM3	canESM2	HadCM3	canESM2	HadCM3	canESM2	HadCM3	canESM2
Cal	Mean	Observed	7.94	7.94	121.7	121.7	24.9	24.8	11.5	11.6
		Simulated	7.97	7.87	127.8	123.7	24.9	24.8	11.5	11.6
	R^2		0.93	0.94	0.98	0.99	0.99	0.99	0.99	1.00

	MAE	value	0.8	0.4	11.1	5.9	0.1	0.1	0.1	0.0
		Ratio relative to mean(%)	9.7	5.3	9.1	4.8	0.5	0.2	1.1	0.2
	RMSE	value	1.0	0.7	15.8	8.9	0.2	0.1	0.2	0.0
		Ratio relative to mean(%)	12.6	8.6	13.0	7.3	0.6	0.3	1.3	0.3
	NSE		0.82	0.92	0.97	0.99	0.96	0.99	1.00	1.00
	Bias		0.51	0.20	4.28	3.65	0.09	0.02	0.09	0.01
Validation*	R^2		0.70	0.71	0.90	0.95	0.97	0.95	0.96	0.95
	MAE	value	2.0	1.8	32.4	23.2	0.4	0.5	0.3	0.4
		Ratio relative to mean(%)	25.2	22.9	26.6	19.1	1.5	1.8	2.4	3.6
	RMSE	value	2.4	2.4	46.4	33.4	0.5	0.6	0.3	0.5
		Ratio relative to mean(%)	30.7	30.0	38.1	27.4	1.9	2.2	3.0	4.3
	NSE		0.37	0.55	0.83	0.92	0.90	0.87	1.00	0.99
	Bias		1.13	0.92	14.77	13.20	0.26	0.33	0.18	0.31

+ calibration period for canESM2 (1984-2000), for HadCM3 (1984-1995) * Validation period for canESM2 (2001-2005), for HadCM3 (1996-2001)

~~**Table 5: Relative change mean annual precipitation and change in Tmax and Tmin modeled from six GCMs for three time periods of UBNRB as compared from the reference period of 1984-2011 by using LARS-WG**~~

	1984-2011	2030s (2011-2035)			2050s (2046-2065)			2080s (2080-2099)		
GCMs/Scenario		A1B	B1	A2	A1B	B1	A2	A1B	B1	A2
Co2 concentration (ppm)		410	418	414	492	541	545	538	674	754
Mean annual rainfall (mm)	1417.5									
CSMK3	Relative Change of precipitation(%)	-2.3	-2.3		-4.2	-2.7		-7.0	-5.3	
GFCM21		-1.4	-0.6	-3.7	-8.0	0.7	-7.4	-7.5	-2.2	-5.9
HADCM3		2.1	0.8	1.7	4.4	2.1	3.5	12.9	4.1	16.7
MIHR		1.9	3.7		5.5	5.5		10.2	6.0	
MPEH5		1.8	3.3	-0.5	2.5	5.8	4.2	6.0	1.4	3.3
NCCCS		6.5	7.8	6.4	22.8	22.0	8.7	29.9	13.7	43.7
Model average		1.4	2.1	1.0	3.8	5.6	2.2	7.4	3.0	14.5
Mean daily Tmax (^{o}C)	24.7									
CSMK3	Change in maximumTemperature (^{o}C)	0.4	0.4		1.3	0.9		2.2	1.5	
GFCM21		0.7	0.6	0.7	2.2	1.4	1.9	3.1	2.0	3.6
HADCM3		0.5	0.5	0.4	1.7	1.4	1.8	3.1	2.0	3.7
MIHR		0.6	0.6		2.0	1.6		3.5	2.6	
MPEH5		0.5	0.4	0.6	1.8	1.4	1.8	4.1	2.7	4.3
NCCCS		0.6	0.5	0.6	1.5	0.9	1.7	2.2	1.4	3.0
Model average		0.6	0.5	0.6	1.8	1.3	1.8	3.0	2.0	3.6
Mean daily Tmin (^{o}C)	11.4									
CSMK3	change in minimum Temperature (^{o}C)	0.3	0.3		1.1	0.8		1.9	1.3	
GFCM21		0.7	0.6	0.7	2.2	1.4	1.9	3.1	2.0	3.6
HADCM3		0.5	0.5	0.4	1.7	1.4	1.8	3.1	2.0	3.7
MIHR		0.7	0.7		2.1	1.8		3.6	2.7	
MPEH5		0.5	0.4	0.6	1.8	1.5	1.8	4.1	2.7	4.1
NCCCS		0.6	0.5	0.6	1.5	0.9	1.7	2.2	1.4	3.0
Model average		0.6	0.5	0.6	1.7	1.3	1.8	3.0	2.0	3.6

~~**Table 6: Relative change of mean annual precipitation, change of mean annual Tmax and Tmin for three time periods as compared to the baseline period of UBNRB using SDSM for HadCM3 and canESM2 GCMs under different scenarios**~~

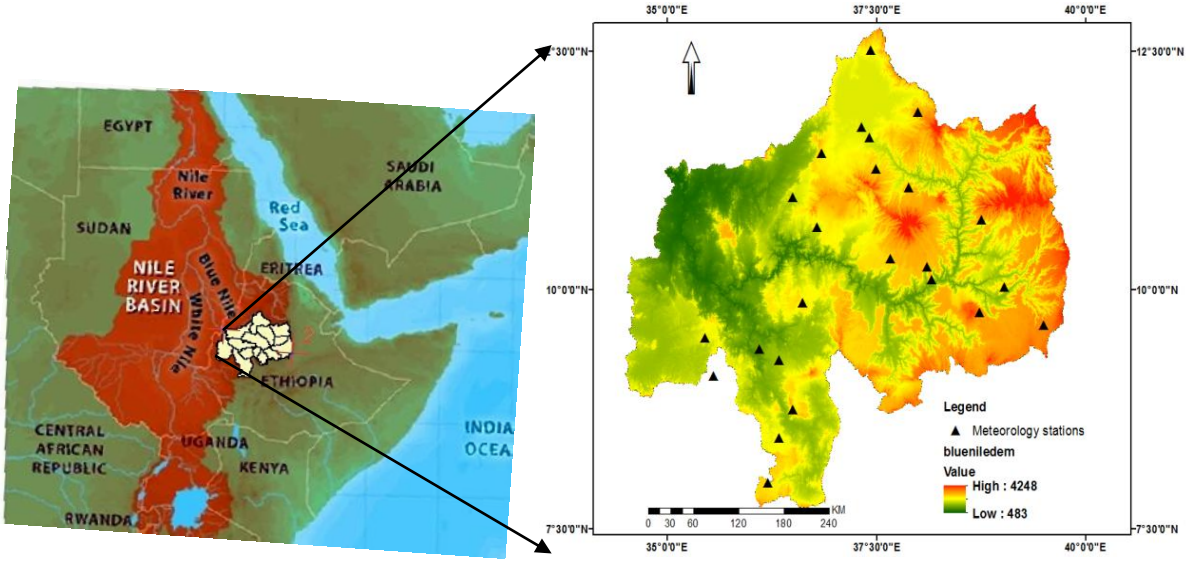


Figure 1~~1~~: Location Map of the study area

a) — b)

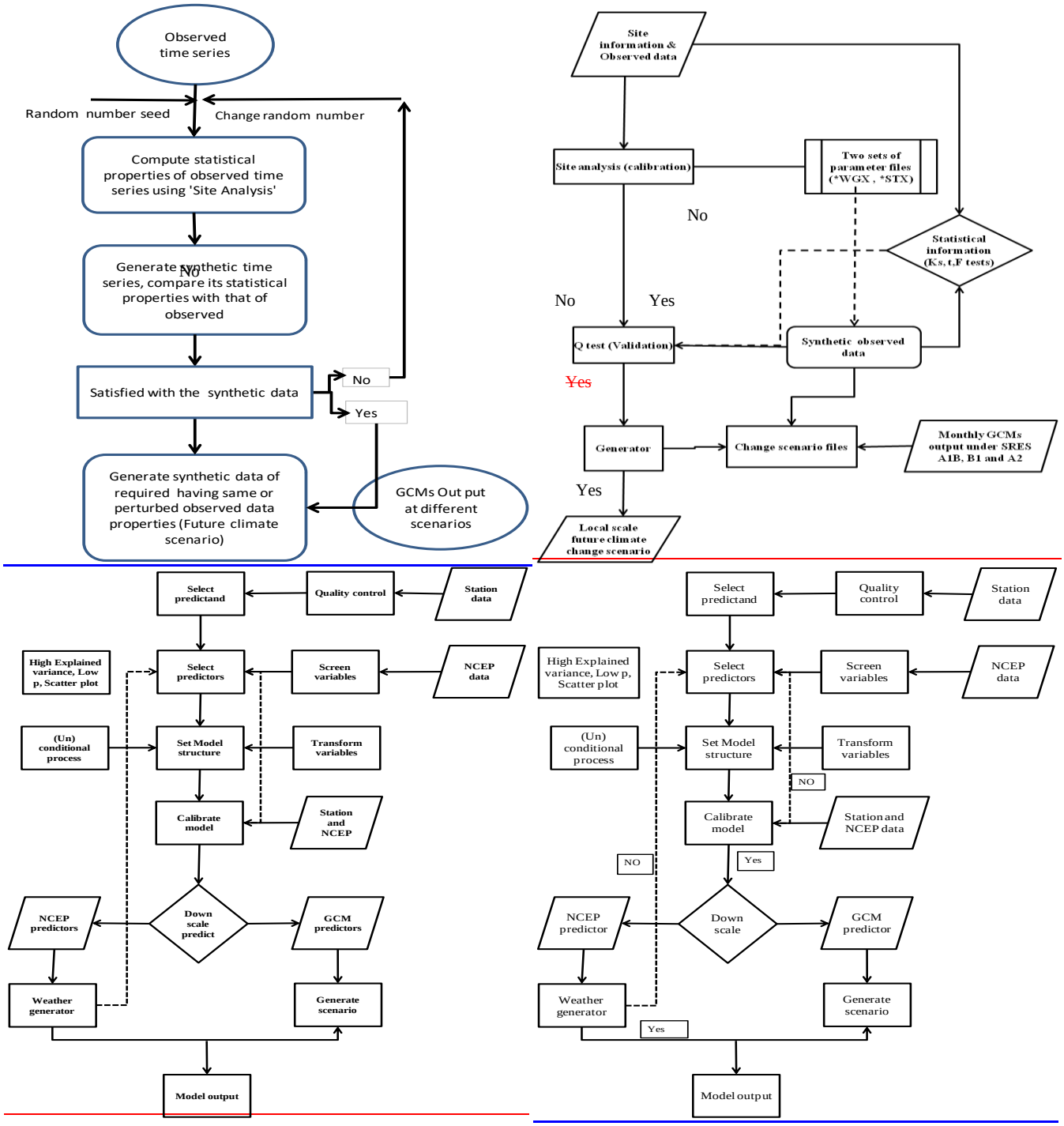


Figure 2~~2~~: Schematic diagram of a) LARS WG analysis b) SDSM analysis source (Wilby *et al.*, 2002)

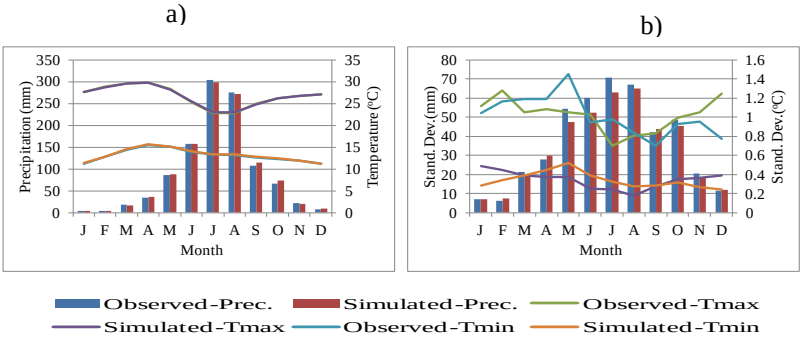


Figure 33: Observed and simulated a) mean monthly precipitation,Tmax and Tmin ; b) standard deviation of precipitation, Tmax and Tmin using LARS-WG

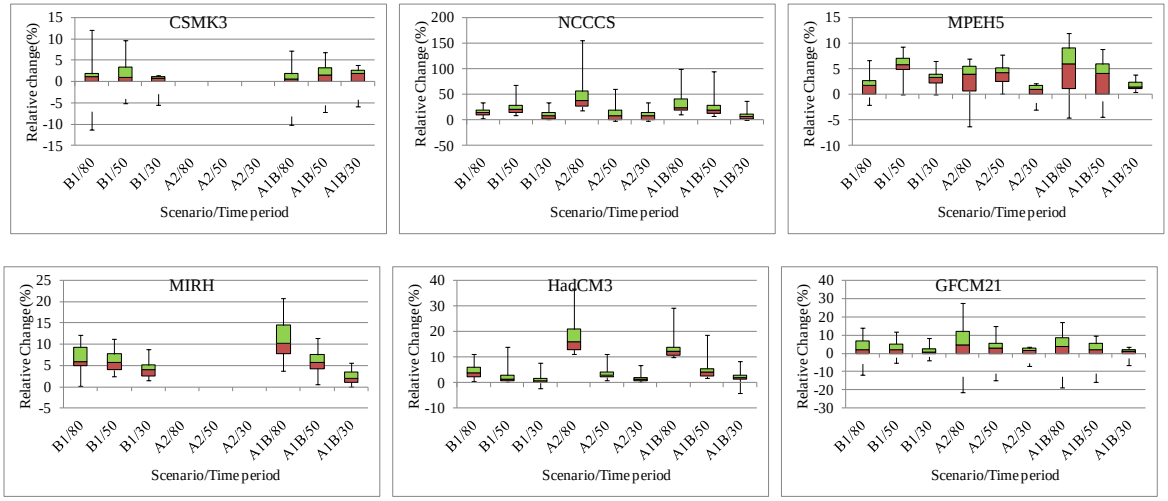
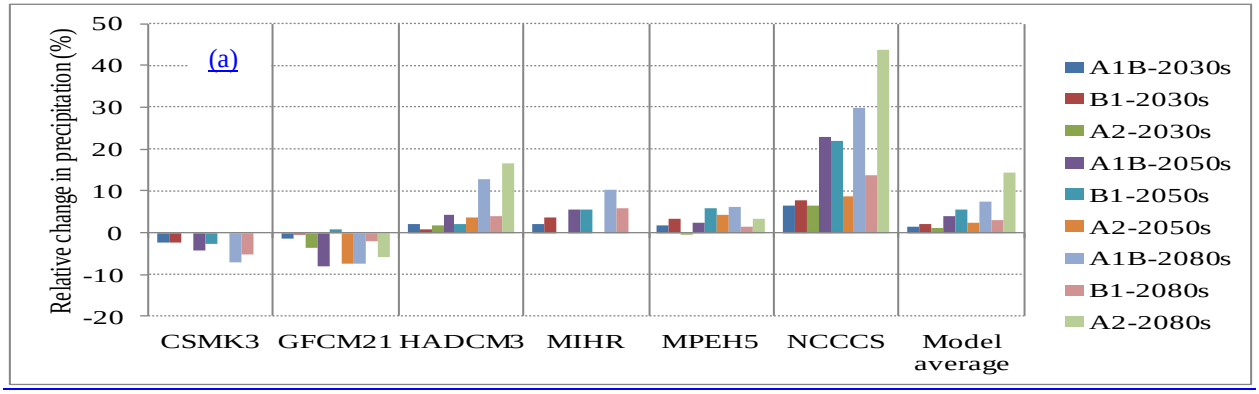


Figure 44: Box plots showing the relative change of precipitation (%) for each six selected GCMs downscaled from 15 stations by using LARS-WG for scenarios (B1, A2 and A1B) during three time periods as compared to the base line. Box boundaries indicate the 25th and 75th percentiles, the line within the box marks the median, whiskers below and above the box indicate the 10th and 90th percentiles,

B1/80: B1 scenario time period 0f 2080s, B1/50: B1 scenario time period 2050s, B1/30: B1 scenario time period 2030s



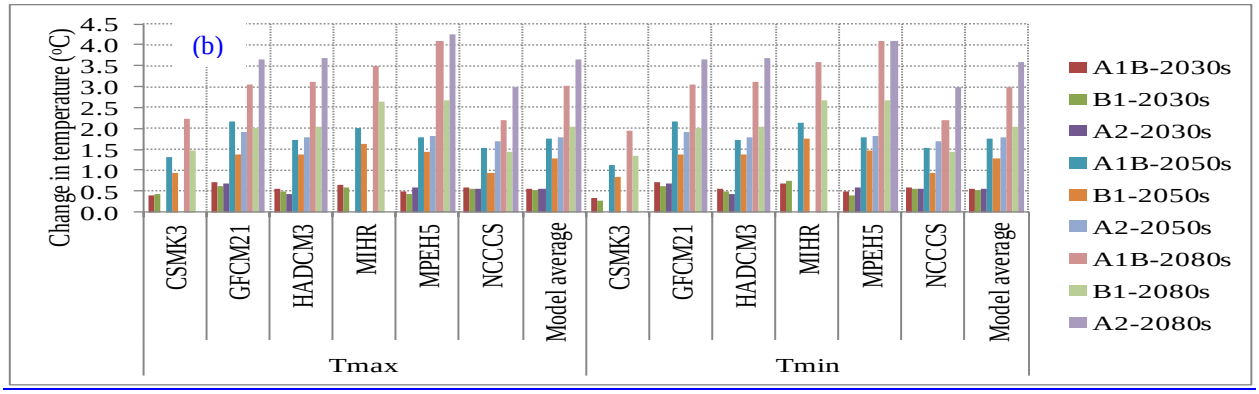


Figure 5: (a) Relative change mean annual precipitation and (b) change in Tmax and Tmin modeled from six GCMs for three time periods of UBNRB as compared from the reference period of 1984-2011 by using LARS-WG

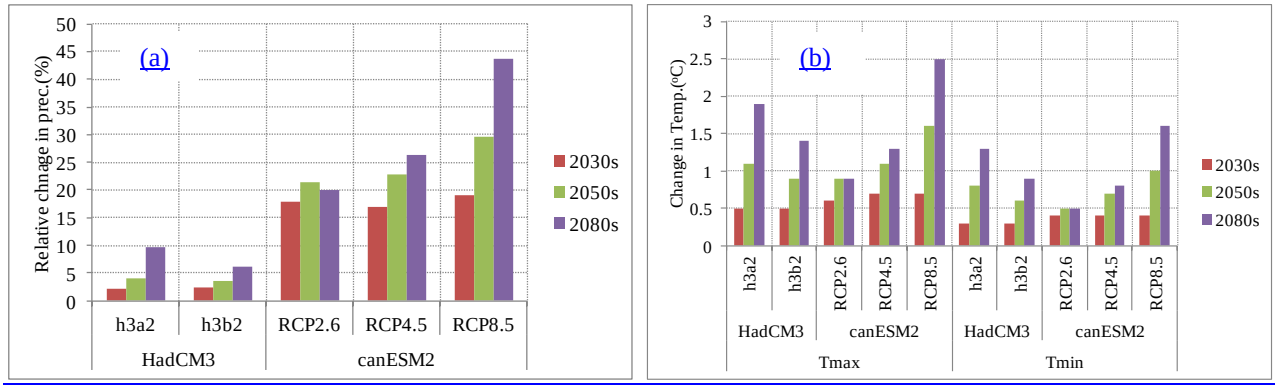


Figure 6: (a) Relative change of mean annual precipitation, and (b) change of mean annual Tmax and Tmin for three time periods as compared to the baseline period of UBNRB using SDSM for HadCM3 and canESM2 GCMs under different scenarios

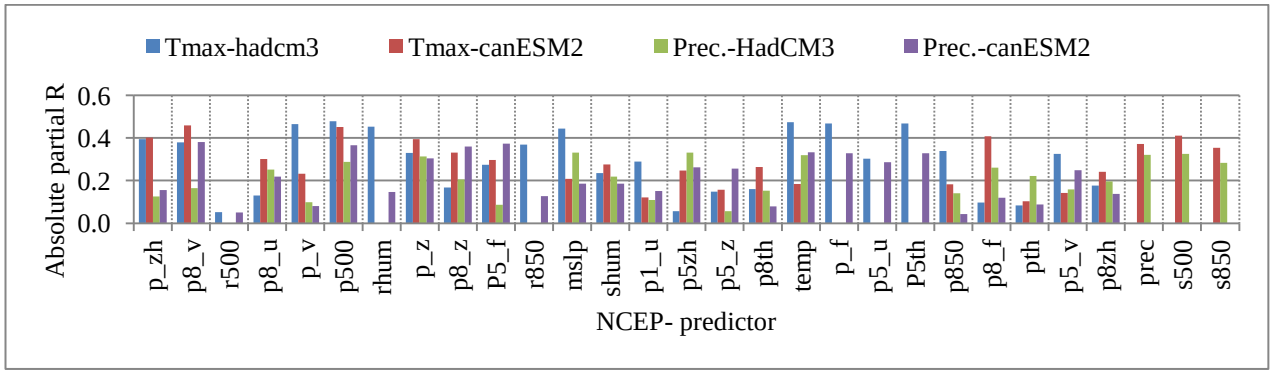


Figure 7~~5~~: Average partial correlation coefficient values of all stations for precipitation and Tmax with NCEP- reanalysis predictors

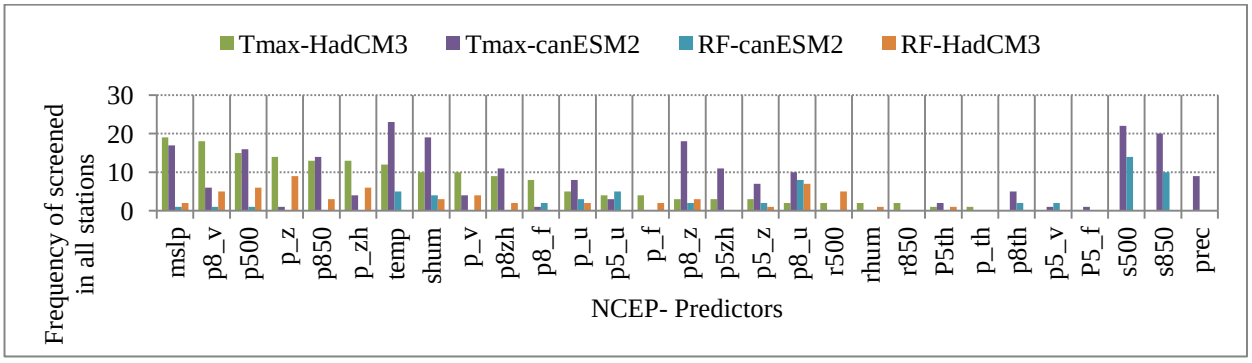


Figure 8~~6~~: Screened NCEP- predictor variables for observed Tmax and precipitation from two GCMs, the maximum frequency is 15 for precipitation and 25 for Tmax

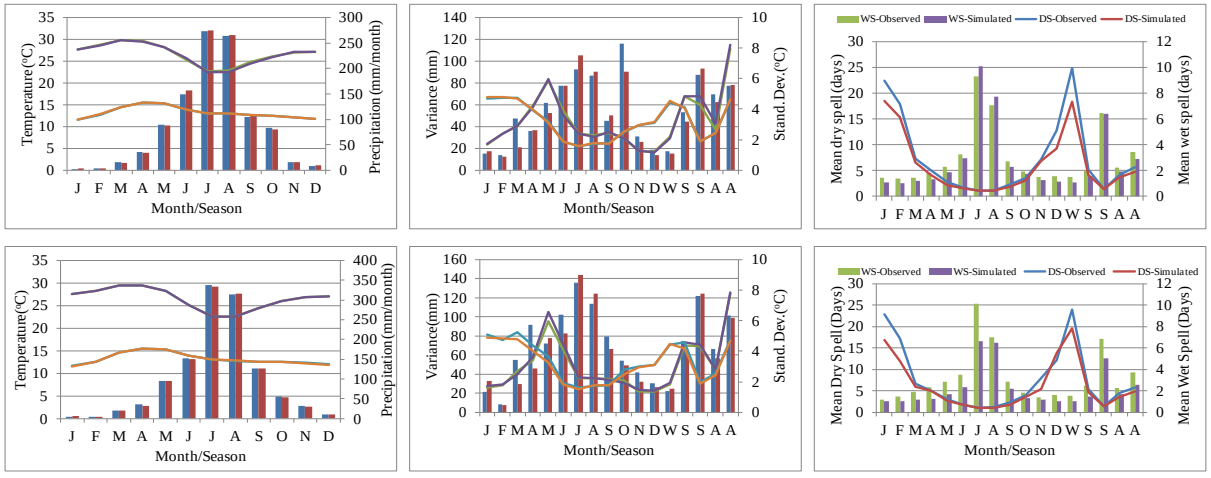
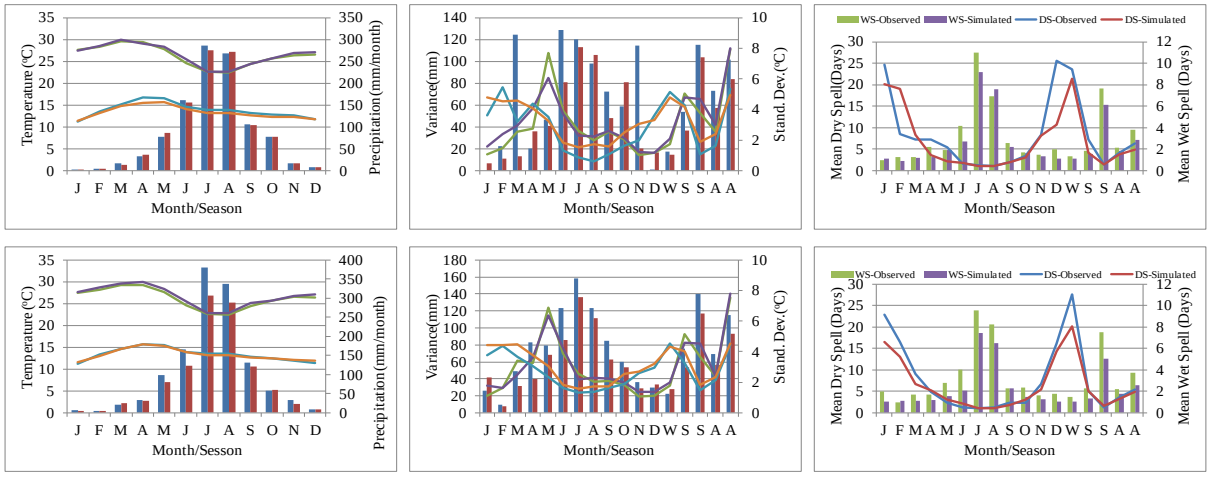


Figure 9~~7~~: calibration of observed and simulated of precipitation, maximum and minimum temperature for the Gondar station using SDSM from canESM2 and HadCM3 from top to bottom



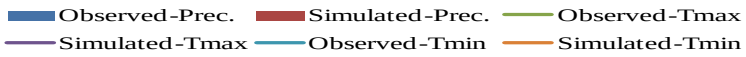


Figure 108: Validation of observed and simulated of precipitation, maximum and minimum temperature for Gondar station using SDSM from canESM2 and HadCM3 from left to right respectively

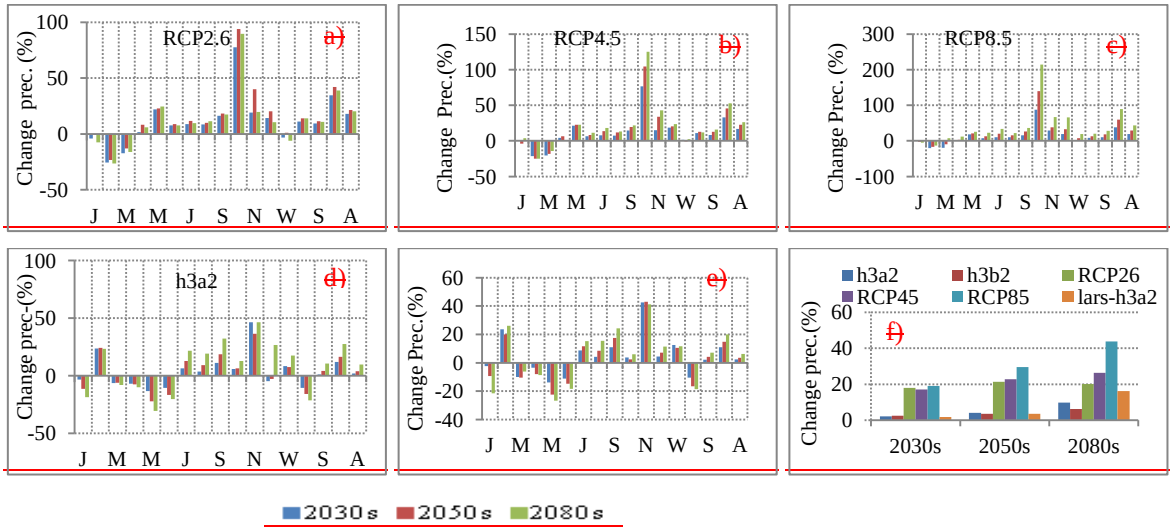
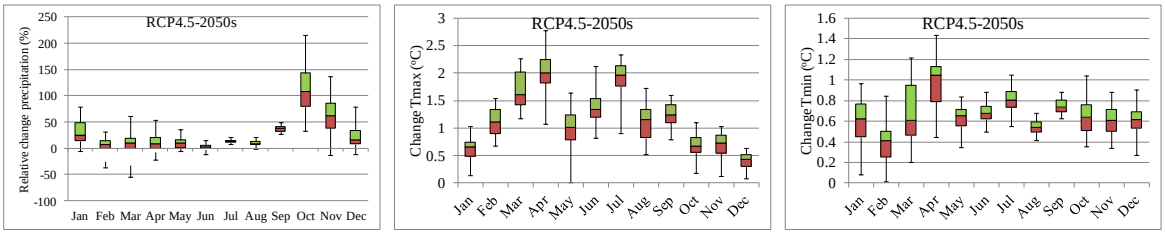


Figure 9: Relative change in monthly mean precipitation for three time periods as compared to the baseline period of UBNRB area a) canESM2/RCP2.6, b) canESM2/RCP4.5, c) canESM2/RCP8.5, d) hadCM3/A2a and e) hadCM3/B2a scenario f) relative change in mean annual precipitation for all scenarios.

(a)



(b)

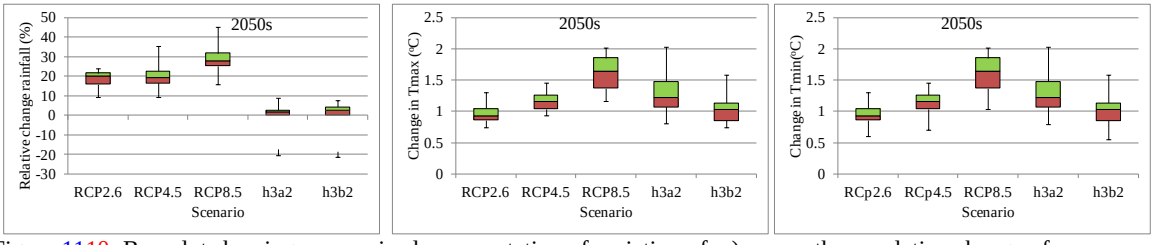
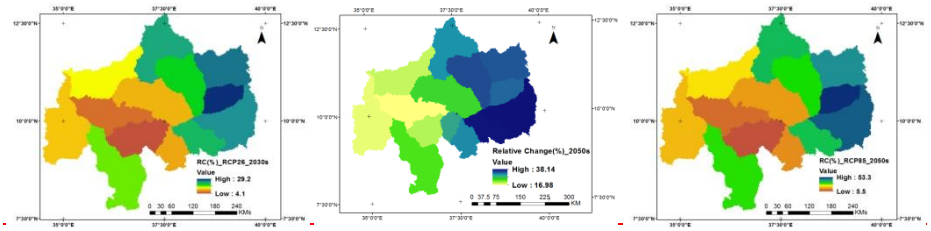


Figure 1110: Box plot showing summarized representation of variation of a) upper three: relative change of mean monthly precipitation, change maximum and minimum temperature from left to right for UBNRB across all stations under RCP4.5 at 2050s Using SDSM b) lower three: Relative change of mean annual precipitation, change in mean annual maximum and minimum temperature from left to right at 2050s for scenarios (RCP2.6, 4. 5 and 8.5 and SRES A2 and B2) for UBNRB using SDSM.



~~**Figure 11: Spatial distribution of the relative change in precipitation under RCP2.6, 4.5 and 8.5 scenario 2050s from left to right by using SDSM**~~

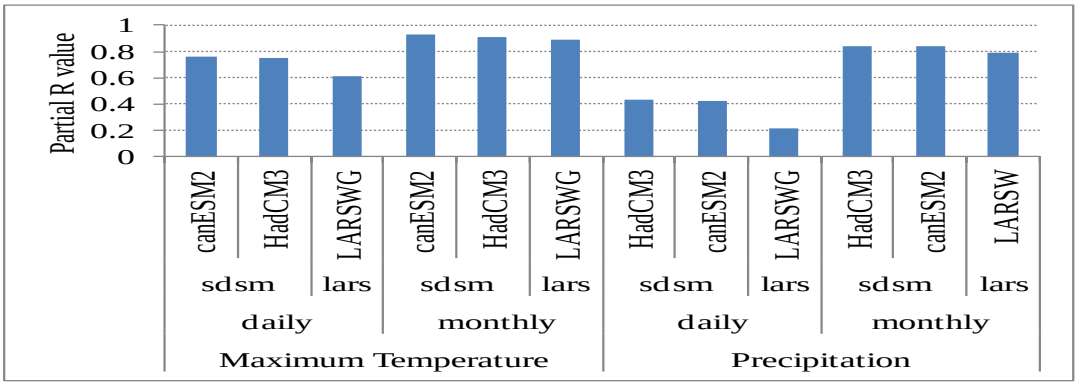


Figure 12~~12~~**:** Performance comparison of LARSWG and SDSM at different time scale

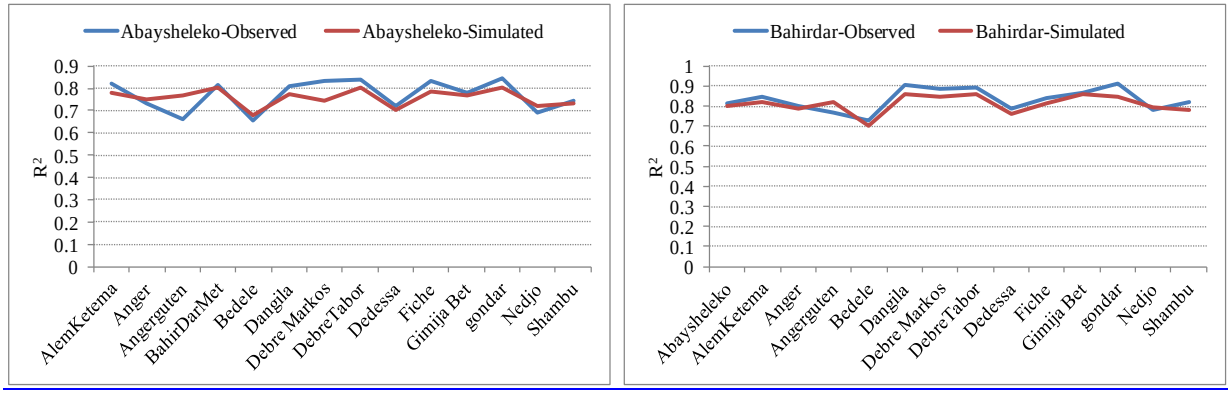
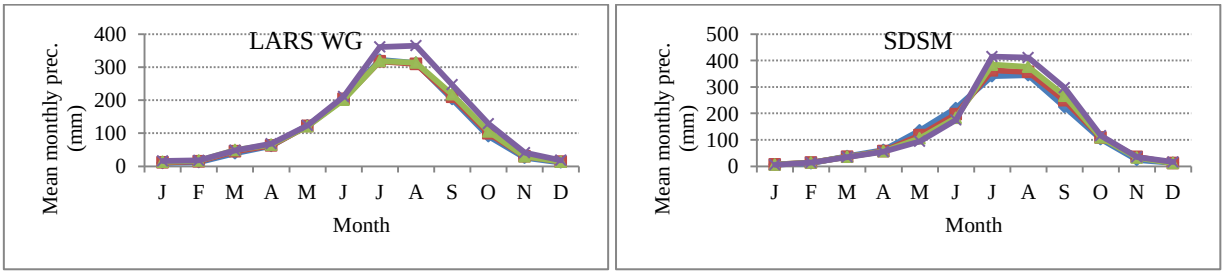


Figure 13: Spatial correlation coefficient (R^2) of Abaysheleko (left) and Bahirdar (right) weather stations with others for monthly precipitation from 1984-2011.



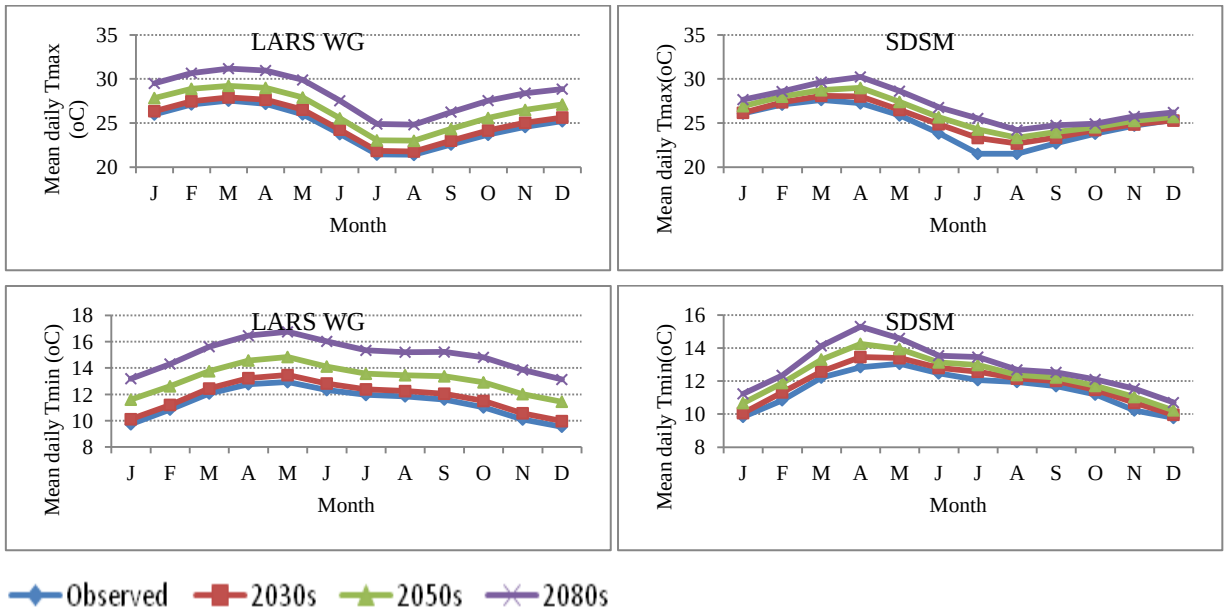


Figure 14~~13~~: General trend in precipitation, Tmax and Tmin at UBNRB corresponding to a climate change scenario downscaled using LARS WG and SDSM from HadCM3 GCM for a2 scenario~~o~~