Article

Weather Research and Forecasting Model Sensitivity to Choice of Parameterization over Ethiopia

Andualem Shiferaw [1],*, Tsegaye Tadesse [2] and Clinton Rowe [3]

[1]   School of Natural Resources, University of Nebraska-Lincoln, 3310 Holdrege Street, Lincoln, NE 68583, USA; andus2@huskers.unl.edu

[2]   National Drought Mitigation Center, School of Natural Resources, University of Nebraska-Lincoln, 3310 Holdrege Street, Lincoln, NE 68583, USA; ttadesse2@unl.edu

[3]   Department of Earth and Atmospheric Sciences (EAS), University of Nebraska-Lincoln, 126 Bessey Hall, Lincoln, NE 68588, USA; crowe1@unl.edu

*   Correspondence: andus@huskers.unl.edu; Tel.:+1-402-419-9605

Abstract: A 3-month (June-August) regime of the year 2002 summer rainfall (JJA2002) was simulated with 30 physics combinations using the Weather Research and Forecasting (WRF) model at 12-km horizontal grid resolution. The objective is to examine summer rainfall sensitivity to parameterization of microphysical, convective, and boundary layer processes and identify an best possible combination of parameterization options that performs relatively better in simulating spatial and temporal distribution of summer rains over Ethiopia. The WRF simulated rainfall was evaluated against station data and satellite rainfall products (CHIRPS and ENACTS) using mean absolute error, Pearson and Pattern correlation coefficients (PCC), pattern correlation and and error in number of rainy days as evaluation metrics. Summer rainfall is found to be most sensitive to choice of cumulus parameterization and least sensitive to cloud microphysics. All simulations captured the spatial distribution of mean seasonal precipitation with PCC ranging from 0.89-0.94. However, all simulations overestimated precipitation amount and number of rainy days. Out of the 30, the simulations that uses a combination of Grell-3D cumulus scheme, ACM2 boundary layer, Lin Microphysics, Dudhai shortwave radiation and RRTM longwave radiation scheme ranked the top and provided the most realistic simulation in terms of amount and spatio-temporal distribution of summer rainfall.

Keywords: RCM, WRF, Downscaling,Parameterization, Sensitivity, Ethiopia

1. Introduction

Advancements in scientific understanding of the climate system and climate modeling have promoted seasonal forecasting to be a well-established operational area at several national centers[1]. As a result, different centres around the world (e.g., the National Center for Environmental Prediction-NCEP, the European Center for Medium Range Weather Forecast-ECMWF, Australian Bureau of Meteorology) run seasonal forecasting systems (e.g., the Climate Forecast System version 2 [CFSV2, 2], seasonal forecast system [SEAS5, 3], and the Australian Ocean-Atmosphere Model for Climate Prediction [POAMA, 4] respectively on a global scale. These seasonal forecast products provide reasonable global perspectives and outlooks of the climate several months in advance. However, despite their potential applications for different socio-economic sectors, the usefulness of these forecasts has been limited because of their coarse spatial resolutions [1,5]. In order for such climate forecasts to be of practical societal value, it is essential for them to be issued at spatial scales appropriate to the decision maker or at the scale needed to exploit them further (e.g., using them as an input for hydrologic or crop simulation models) [5]. To address the scale problem and meet the need for regional information, downscaling seasonal forecast by using regional climate models (RCMs) became an emerging area during the last decade [5–7].

Several studies around the world have demonstrated the potential advantages of using RCMs to downscale coarse resolution climate predictions [e.g., 1,8–15]. Although

fewer in number and scope (i.e., experimental and/or research only), RCMs have been tested over Greater Horn of Africa (GHA) region. For example, in their dynamical seasonal hindcast study over East Africa, [11] found that the Regional Climate Model system 4 (RegCM4) reproduces both spatial and inter-annual variability of seasonal rainfall and captures the teleconnection between El Nino Southern Oscillation (ENSO) and regional precipitation structure. Cheneka *et al.* [16] evaluated downscaling of global seasonal hindcasts from the Max Planck Institute for Meteorology Earth System Model (MPI-ESM) using the COSMO-CLM (CCLM) RCM over East Africa during summer season over ten-year period (2000–2009). They found that although COSMO-CLM didn't remove wet bias in summer precipitation over the Ethiopian highlands and in parts of the lowland, it managed to add value in capturing extreme precipitation years, especially over the Ethiopian highlands.

However, despite such promising results and the widely accepted notion that RCMs can improve the simulation of precipitation compared with global forecasts owing to their more comprehensive representation of the important physical processes at a finer resolution [17–19], dynamical downscaling cannot be applied universally. This is due to a range of options available in RCMs for different physical and dynamical parameterizations. For example, WRF model (which is the focus of this study) currently provides more tha 15 cumulus (CU), 15 planetary boundary layer (PBL) and 20 microphysics (MP) parameterizations options [20]. These range of options are meant to allow users to select physics and dynamics settings that optimise the model for their particular needs [21–23]. However, the the variety of configurations that WRF can be operated can lead to varying results. Selection of an optimum combination of options depends on quality of lateral boundary condition, scales, geographic location, application, domains size, spin up, vertical resolution or nesting architecture [19] and any change in the configuration of these factors can lead to varying result. Hence, it is crucial to test for the most appropriate model setup for a particular purpose over a given region through numerical experiments [21]. Consequently, numerous sensitivity studies have been conducted over different parts of the world to identify an optimum WRF configuration [e.g., 22,24–28].

Despite the crucial importance of sensitivity studies, only a handful of such studies have been conducted over the GHA region [e.g., 9,29]. In addition, the above few studies either cover a specific season or a small portion of the region. Given the high spatial variability of climate over the region and the vast possible combination of physical paramerization options, the available studies are almost insignificant. Thus, in this study a range of physics combinations in WRF model are used to simulate a summer rainfall during drought year of 2002 across GHA (with focus on Ethiopia) for identifying the best possible configurations that would latter be used to dynamically downscale global seasonal precipitation forecast from global models.

2. Materials and Methods

2.1. Verification data

In order to address observation uncertainty, we used three data sets to evaluate the performance of each combination of parameterization options: daily precipitation from 58 meteorological stations obtained from National Meteorology Agency (NMA) archive, gridded monthly precipitation from the Climate Hazards Infrared Precipitation with Stations version 2 [CHIRPS, 30] and gauge-satellite blended rainfall estimate from Enhancing National Climate Services initiative [ENACTS, 31,32]. CHIRPS has a resolution of $0.05°$ $\times 0.05°$ while ENACT has around $0.1° \times 0.1°$ (10km) resolution and were mainly used to evaluate performance in terms of capturing magnitude and spatial distribution of mean seasonal precipitation. To facilitate grid-to-grid comparison with WRF simulation, both CHIRPS and ENACTS datasets were regridded from their native grid to WRF simulation grid configuration using bilinear interpolation routine from the Earth System Modelling Framework (ESMF) in NCAR Command Language (NCL) Version 6.3 [33].

Despite having comparable spatial resolution to WRF simulation, the two gridded datasets exhibit lower spatial details in precipitation compared to WRF likely owing to the sparse network of weather stations used in the blending process (e.g., spatial distribution of mean seasonal precipitation shown in Figure 3). Thus, it is important to note the possible implications of this mismatch on results of evaluations. However, as these datasets are relatively the best available products over the region ([34,35] and the station network over the area is sparse and uneven, they have been utilized for the evaluation. In addition, evaluation of performance related to intensity and frequency of rainfall events performed with respect to weather stations only as the gridded products either lack data or are unreliable on daily time scales. For evaluations involving weather stations, data for WRF grid points nearest to respective stations were extracted.

2.2. Initial and boundary data

The initial and lateral boundary conditions including SST used to drive WRF are obtained from the Climate Forecast System Reanalysis(CFSR) data set [36] where surface variables have $0.312^0 \times 0.312^0$ resolution while pressure level data have $0.5^0 \times 0.5^0$ resolution with model top at 1mb. These datasets (i.e., temperature, humidity, surface pressure, geopotential heights and winds) were prescribed at 6-h intervals starting on 1st May and were obtained from the NCAR Research Data Archive (https://rda.ucar.edu/datasets/).

2.3. Model Description

All experiments in this study were conducted with WRF model, version 3.8.1 [20]. The WRF model is a state-of-the-art, next-generation mesoscale numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs (http://www.wrf-model.org). It is a non-hydrostatic model, with several available dynamic cores as well as many different choices for physical parameterizations suitable for a broad spectrum of applications across scales ranging from meters to thousands of kilometers. The physics package includes microphysics, cumulus parameterization, planetary boundary layer (PBL), land surface models (LSM), longwave (LWR) and shortwave radiation (SWR) radiation. An in depth description of WRF model can be found in [20].

2.4. Model Setup: domain and integration time

We used a two-domain configuration with one-way nesting for all simulations: parent domain (D01) and nested domain (D02) with resolution of 12km and 4km respectively (Fig 1). The parent domain (D01) is centered on Ethiopia and extends from 15^0S to 22^0N and 15^{18}E to 55^0E. It comprises the entire Greater Horn of Africa, central Africa, parts of southern Africa, Red Sea, parts of North Africa and the Middle East, Arabian Sea and western half of Indian Ocean. It is intended to encompass, to the extent feasible, regions that include synoptic features and circulations which directly influence summer climate over Ethiopia [17]. Although the domain is not large enough to incorporate synoptic systems starting from their source regions (e.g., low level flow from Atlantic Ocean to Ethiopia begins from southern Atlantic ocean where St. Elena high is located), findings from previous studies ([e.g., 11,37] suggest that the domain is large enough for systems to develop fully. For example, Diro *et al.* [11] carried out sensitivity experiment using RegCM3 driven by ECMWF seasonal hindcast to quantify the impact of domain size in simulating the spatial pattern of summer rainfall over Ethiopia using two-domain set up, one covering the horn of Africa (23^0E to 57^0E, 5^0S to 23^0N) and the second smaller domain covering only Ethiopia. The comparison showed that the larger domain was far better reproducing the correct precipitation pattern while the smaller domain performed relatively poor to the extent that the location of precipitation maxima is misplaced. The above study, while showing the importance of domain size in simulating summer precipitation over Ethiopia, it also partly justifies that the size of the outer domain used in our study (which extends further south and west by 10 degrees compared to larger domain used in Diro *et al.* [11] is sufficiently large not to be too constrained to reproduce errors in the driving reanalysis [17].

The nested domain (D02) on the other hand covers the whole Ethiopia with few extra grid cells on all sides to account for relaxation zone. Results discussed in subsequent sections will be based on simulation outputs from D02.

As in horizontal resolution, a uniform vertical resolution 40 eta levels spaced closer together in the PBL were used across all simulation. It cover the whole troposphere with resolution decreasing slowly with height in order to allow low-level flow details to be captured. The first 20 levels are inside atmospheric boundary layer (below 1500 m), with the first level at approximately 16 meters, and the domain top at 100 hPa. Although it is recognized that the choice of model horizontal and vertical resolution, size and location of domain boundary and choice of boundary conditions can be equally important to the choice of physics options [e.g., 17,29], identifying the optimum configuration for these options is beyond the scope of this study. This sensitivity experiment is conducted for anomalously dry summer season of 2002 with each simulation covering period from May 1st to August 31st. The first month of the simulation (i.e., May) was considered as model spin-up and only the simulations from June through August were used for model evaluation.
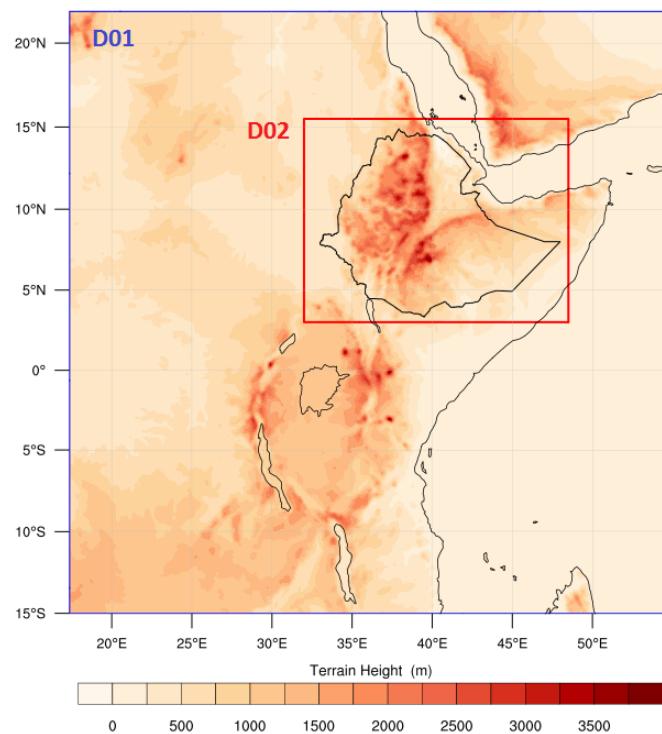

Figure 1. The model domain and topography

2.5. Experimental setup

The accuracy of the model configured with a certain scheme cannot be uniquely attributed to a single parameterization but rather to the combination of them, since feedbacks are usually as important as the schemes themselves[38]. Furthermore, the suitability of a specific configuration strongly depends on the region, the season, or even the particular event considered and hence, there is no single configuration appropriate for every situation. Since testing all the possible combinations of physics options is not computationally affordable, a representative sample of the physics packages was chosen with a different level of complexity and formulation [38].

The WRF model comprises multiple options for most parameterization schemes that can be combined in any different way, enabling the user to optimize the model for a range of spatial and temporal resolutions and climatologically different geographical regions [22]. The options typically range from simple and efficient, to sophisticated and more computationally costly, and from newly developed schemes to well-tried schemes such

as those in current operational models. Since testing the extremely large ensemble of all possible combinations of physics options is not computationally feasible, we considered a small subset based on the most commonly used physics options for cumulus (CU), planetary boundary layer (PBL), microphysics (MP) and radiation parametrization. Figure 2 shows the schemes tested and how the different physics options for each are combined. The CU is the scheme with highest impact on precipitation simulation and is used to predict the collective effects of convective clouds at smaller scales as a function of larger-scale processes and conditions. The CU options tested are Kain-Fritsch [KF, 39], Betts-Miller-Janjic [BMJ, 40,41] and Grell 3D [Grell, 42,43]. The PBL and surface layer schemes define boundary layer fluxes (heat, moisture, momentum) and the vertical diffusion processes. For PBL (and surface layer) the Yonsei University [YSU, 44], Mellor-Yamada-Janjic scheme [MYJ, 40] and Asymmetric Convective Model [ACM2, 45] were tested. In the WRF model, some PBL schemes are tied to particular surface layer schemes [20], so a single common surface layer scheme could not be used here. Thus, Revised MM5 Monin-Obukhov scheme [46] scheme was used with YSU and ACM2 and Monin-Obukhov (Janjic Eta) scheme [47] scheme was used with MYJ scheme. The MP schemes allows to predict water phase transitions in the atmosphere and to consider snow and hail. The MP options tested were the WRF Single-Moment 6-class [WSM6, 48], LIN [Lin, 49] and Morrison double-moment scheme [MOR, 50]. The combination of 3 options from CU, PBL and MP resulted in 27 simulations with longwave (LW) and shortwave radiation (SW) parameterizations set to the Rapid Radiative Transfer Model [RRTM, 51] and Dudhia Shortwave Scheme [Dudh, 52] respectively. In addition to the above 27 simulations, 3 more simulations were conducted to test sensitivity to selection of radiation schemes where the RRTMG Shortwave and Longwave Schemes [RRTMG, 53] was combined with Dudhai and RRTM schemes (Table 1). The following options were used for the other parametrization schemes: KF for CU, YSU for PBL, WSM6 for MP. Unlike the RRTM/Dudhia schemes that consider a binary measure of grid cloudiness, the RRTMG schemes use overlapping cloud fraction algorithms to determine the cloudiness of the grid. Furthermore, RRTMG schemes take into account the concentrations of trace gases, aerosols, ozone,and carbondioxide, and they consider reflected shortwave radiation fluxes [26]. All the 30 experiments use the Noah land surface model [54], and MODIS 21 class landuse data.

CUMULUS		PBL									LW	SW
		YSU			MYJ			ACM2			RRTM	Dudhai
	KF	E1	E2	E3	E4	E5	E6	E7	E8	E9		
	BMJ	E10	E11	E12	E13	E14	E15	E16	E17	E18		
	GRELL3	E19	E20	E21	E22	E23	E24	E25	E26	E27		
		WSM6	LIN	MORR	WSM6	LIN	MORR	WSM6	LIN	MORR		
		MICROPHYSICS										

Figure 2. Experimental set up. Each row contains experiments with KF, BMJ and Grell 3D CU schemes. First 3 columns YSU PBL, columns 4-6 experiments with MYJ PBL scheme and columns 7-9 ACM2 PBL scheme. Columns 1,4,7 WSM6 MP scheme; columns 2,5,8 Lin MP scheme and columns 3,6,9 Morrison MP scheme. All 27 experiments utilize RRTM LW and Dudhai SW radiation schemes.

WRF Run	Cu	PBL	MP	LW	SW
E28	KF	YSU	WSM6	RRTMG	Dudhia
E29	KF	YSU	WSM6	RRTM	RRTMG
E30	KF	YSU	WSM6	RRTMG	RRTMG

Table 1. Additional set of experiments for Radiation parameterization sensitivity

2.6. Evaluation Statistics

The WRF simulated and observed precipitation values from meteorological stations as well as gridded precipitation products are compared using four different statistics: the Mean Absolute Error (MAE), pattern correlation coefficient (PCC), Pearson Correlation Coefficient (R), daily intensity index and frequency of rainy days in the season. Gridded rainfall products (i.e., CHIRPS and ENACT) were used as observational references in calculating the first two statistics only while daily precipitation from meteorological stations is used to calculate all but PCC. MAE is used to measure the closeness of the modeled and observed values. For grid-based comparison, mean seasonal precipitation was first computed for observation and WRF simulations and MAE was calculated over each grid-point pairs with respect to both gridded data sets. For station-based comparison, absolute value of bias is calculated for each day and averaged over 92 days in the season. The PCC is computed from observed and simulated mean JJA precipitation according to the usual Pearson correlation operating on the M grid point pairs from WRF and gridded observations [55]:

$$PCC = \frac{\sum\limits_{m=1}^{M}(y_m - \bar{y})(o_m - \bar{o})}{\left[\sum\limits_{m=1}^{M}(y_m - \bar{y})^2 \sum\limits_{m=1}^{M}(o_m - \bar{o})^2\right]^{1/2}} \tag{1}$$

where y and o are WRF simulated and observed seasonal mean precipitation at the mth grid point and the over-bars refer to these variables averaged over M grid points (i.e., here, M refers to grid points within the boundary of Ethiopia. The PCC ranges from -1 to 1 [56] with values closer to one indicating the higher skill of WRF in capturing the observed spatial patterns of mean seasonal precipitation over Ethiopia. The PCC is calculated with respect to both CHIRPS and ENACT and values reported in this paper are average of the two. The R is used to quantify the ability of WRF simulations in capturing intra-seasonal variation of daily precipitation and is calculated from time series of daily precipitation from June 1st to August 31st, 2002 between gauging stations and WRF grid-points where respective stations fall. The error in frequency of rainy days (ERD) is computed as the difference between simulated and observed total number of rainy days (i.e., daily precipitation is greater than 1 mm) over WRF grid-points and gauging station. Similarly, mean Daily intensity is compared by dividing total precipitation in the season with total number of rainy days in the season.

To rank the different combination of physics options based on performance statistics, a new aggregate score (AS) is defined and computed. As the above statistics have different units, range and orientation, simple manipulations are applied to the scores before aggregation. First, evaluation metrics that are computed over multiple grid-points or stations (i.e., all except PCC) are averaged (spatially) to create one score for each experiment. In terms of score orientation, R and PCC are positively oriented (larger value indicating higher performance) while MAE, MB and ERD are negatively oriented. To facilitate aggregation, R and PCC were converted to negative orientation by subtracting the scores from 1 (1-R or 1-PCC). All statistical scores are then normalized using Equation.2 to change their range to 0-1.

$$X_{norm} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \tag{2}$$

where X is a score and X_{min} and X_{max} are defined by the best and worst of the 27 simulations for any given score.

Thereafter, the AS were computed as the sum of the normalized (X_{norm}) values of R, PCC, MAE, MBias and ERD (Equation 3). As each of the five normalized terms have values ranging from 0 (for the best) to 1 (for the worst), AS will range between 0 (best) and 5 (worst).

$$AS = R_{norm} + PCC_{norm} + MAE_{norm} + MAE2_{norm} + ERD_{norm} \qquad (3)$$

3. Result and Discussion

3.1. Mean seasonal precipitation

According to analysis of long term precipitation records from CHIRPS and ENACTS, northwestern, western, and central mountainous regions of Ethiopia (where summer is the main rainy season), receive 500 mm to 1200 mm of mean seasonal rainfall. The semiarid regions of northeastern, eastern, southeastern and southern Ethiopia on the other hand receive comparably less precipitation during JJA season that accounts for less than 25% their mean annual precipitation respectively. Although 2002 is one of the driest years over Ethiopia, the CHIRPS and ENACTS rainfall distribution for JJA2002 (Figure. 3a and b) is similar to the long term climatology except for the relatively wetter highlands that drier compared to climatology.

Figure. 3 (2nd to 4th column) show mean seasonal precipitation for JJA2002 for 27 WRF simulations shown in Figure 2 and the PCC scores for these simulations with respect to both gridded products are presented in Figure 6. In general, all 27 experiments have reproduced the spatial patterns of JJA precipitation over Ethiopia that includes the north-south and east-west precipitation gradients, precipitation gradient within and either sides of central rift valley, fine scale patterns associated with the topography; the maximum precipitation over the Bale mountain ridges in southern Ethiopia and over the western side of the Semien mountains. This is evident from the PCC ranging from 0.92 to 0.96 (Figure.6a) when compared with CHIRPS and 0.86 to 0.93 when compared with ENACTS. Although the observed fine scale rainfall patterns are well reproduced, the exaggerated details in WRF simulations (compared to CHIRPS and ENACTS that have comparable resolution as WRF simulations) over the mountainous regions suggest an over-sensitivity to the topography. Diro *et al.* [11] also found a similar result while comparing RegCM3 simulations (driven by ERA-Intrim) with TRMM gridded precipitation and suggested that this might be due to misrepresented topographic forcing and the diagnostic nature of convective parametrization schemes that prevent advection of evolving convective systems from one grid cell to another [57].

Although WRF simulations have captured spatial patterns reasonable well, all simulations have consistently overestimated JJA precipitation prominently over the Ethiopian highlands (i.e., west and northwestern Ethiopia as well as Bale Mountains on eastern side of the Rift Valley) and north eastern lowlands to a lesser extent (and for selected experiments) and slightly underestimating over southwestern and northeastern peripheries. Zeleke *et al.* [58] also found a similar pattern where RegCM4 simulations although reproducing climatological rainfall pattern show positive and negative biases over the western mountainous regions and some isolated lowland areas (e.g., area extending from north to south between northwestern highlands and northeastern lowlands respectively(large dry bias),mountainous regions on the eastern side of rift valley) respectively when compared with RegCM4 simulation of JJAS rainfall averaged over period from 1989 to 2005.
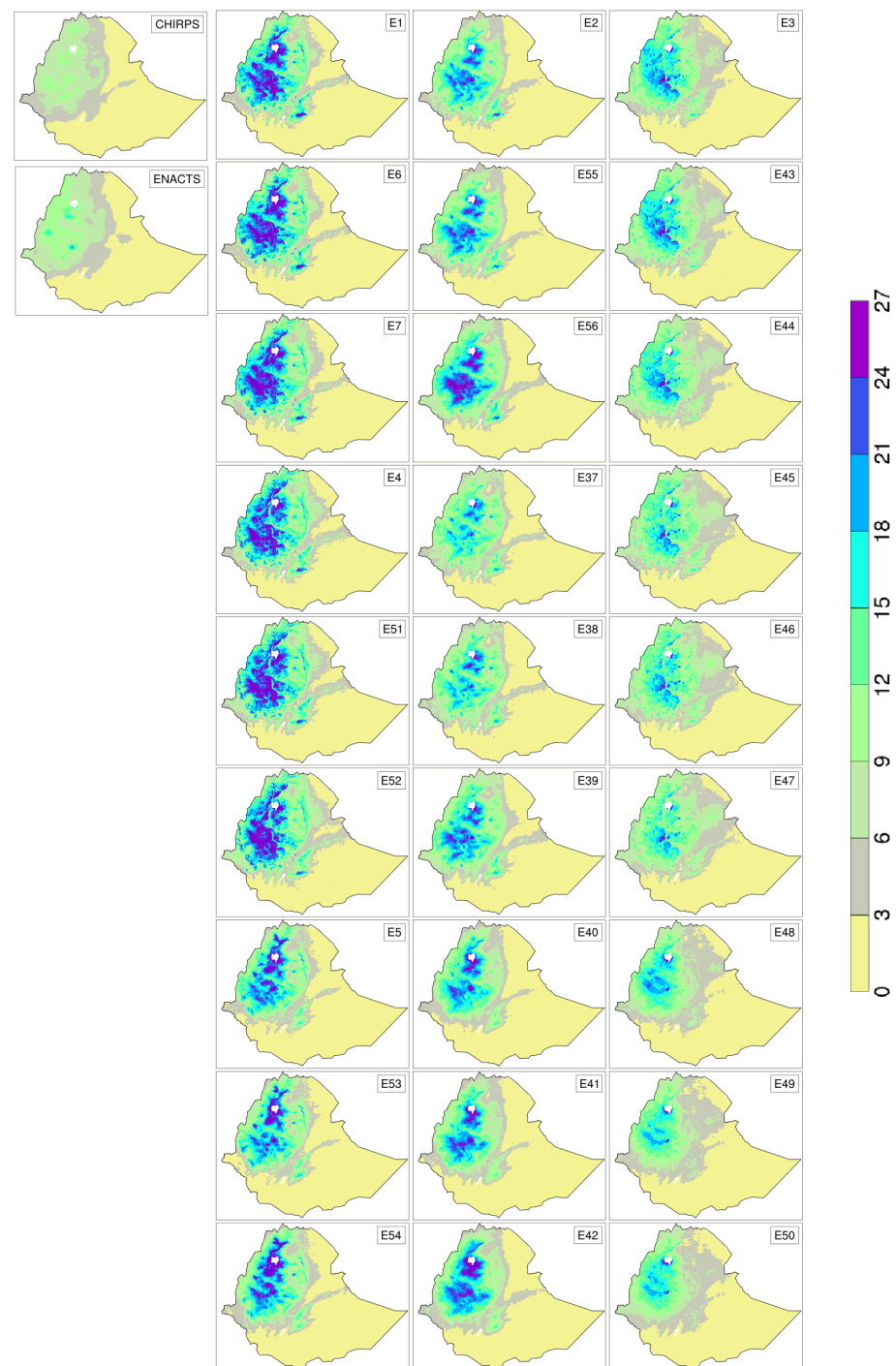
Figure 3. Seasonal (JJA) mean precipitation (mm/day) during 2002 for a) CHIRPS, b)ENACTS and WRF Simulation. First column KF CU scheme, 2nd BMJ CU scheme and 3rd Grell CU scheme. 2-4 row YSU PBL scheme, 5-7 MYJ PBL scheme and 8-10 ACM2 PBL scheme. Rows 2, 5 and 8 WSM6 MP scheme; rows 3,6,and 9 Lin scheme and rows 5,7 and 10 Morrison MP scheme. Refer to Figure 2 for more details.

The wet bias is the highest for simulations using KF CU scheme followed by BMJ and the least for Grell CU scheme (Figure.3 and **??**). Over some of these locations, wet bias exceeds 10mm/day and covers the largest area under KF scheme followed by BMJ that shows localized spots. Such wet bias magnitudes are however seen only at a single location for Grell scheme. Although Grell scheme has outperformed KF and BMJ schemes in terms of mean seasonal precipitation magnitude, it has a false wet spot over northeastern dry

lowlands of Ethiopia where precipitation is overestimated by as much as 6mm/day when compared with both gridded datasets. Although not as extensive and as large as Grell, this wet bias is also exhibited for KF scheme. The wet bias for simulations involving KF scheme show no change regardless of changes in PBL schemes (YSU and MYJ) and MP schemes (all three) except when used together with ACM2 PBL. There is still a considerable wet bias but smaller than the other KF simulations. When KF and Grell CU schemes are used with MYJ PBL scheme results in a notable increase in wet bias over northeastern lowlands (among other locations). For KF CU scheme, the wet bias increases in magnitude and extends further east when used with MYJ PBL scheme. For Grell CU scheme, the wet bias is present across all simulations but magnitude increases when used with MYJ scheme. Unlike KF and Grell , BMJ CU scheme performed better when used with MYJ PBL scheme with significantly smaller wet bias over highlands and better representation of precipitation gradient on either sides of the rift valley. This confirms to the findings of Gbode *et al.* [59] who pointed out that schemes that are developed together tend to perform better. Precipitation is least sensitive to Microphysics schemes with only noticeable response seen when BMJ Cu and YSU PBL schemes are used with Morrison MP scheme which results in a relatively higher wet bias use of WSM6 and Lin MP schemes.
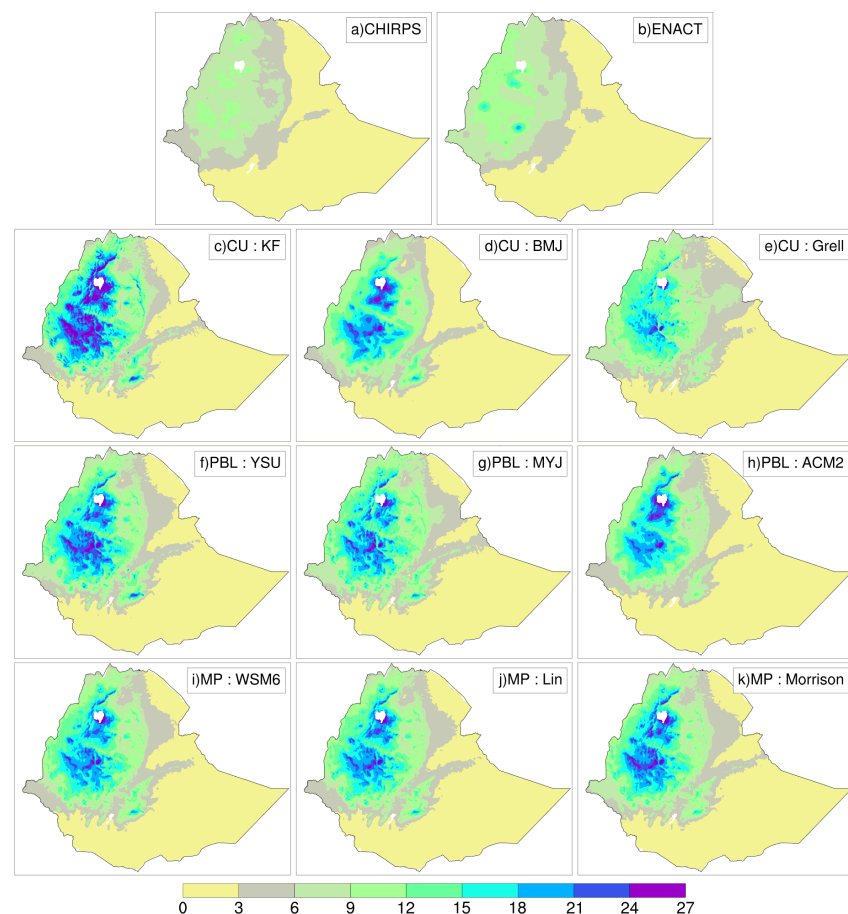

Figure 4. Seasonal (JJA) mean precipitation (mm/day) during 2002 for a) CHIRPS, b)ENACTS, WRF Simulation with c-e) CU sensitivity with KF, BMJ and Grell CU physics option, f-h) PBL sensitivity with YSU, MJY and ACM2 options respectively, i-k) Microphysics sensitivity with WSM6, Lin and Morrison options respectively. The WRF simulated rainfall for each group (CU, PBL and MP) are mean of 9 members with different combinations.
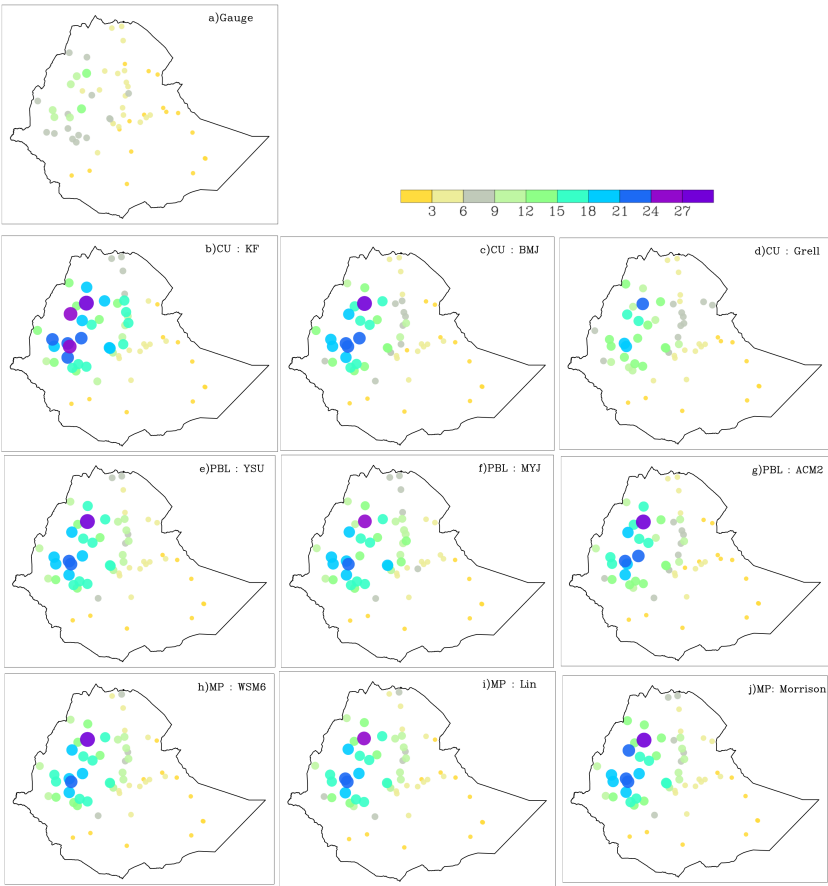
Figure 5. Seasonal (JJA) mean precipitation (mm/day) during 2002 for a) CHIRPS, b)ENACTS, ensemble mean of WRF simulation utilizing c) KF CU, d)BMJ CU, e)Grell CU, f)YSU PBL, g)MYJ PBL, h) ACM2 PBL, i)WSM6 MP, j)Lin MP, k)Morrison MP options. The WRF simulated rainfall for each group (CU, PBL and MP) are mean of 9 members with different combinations.
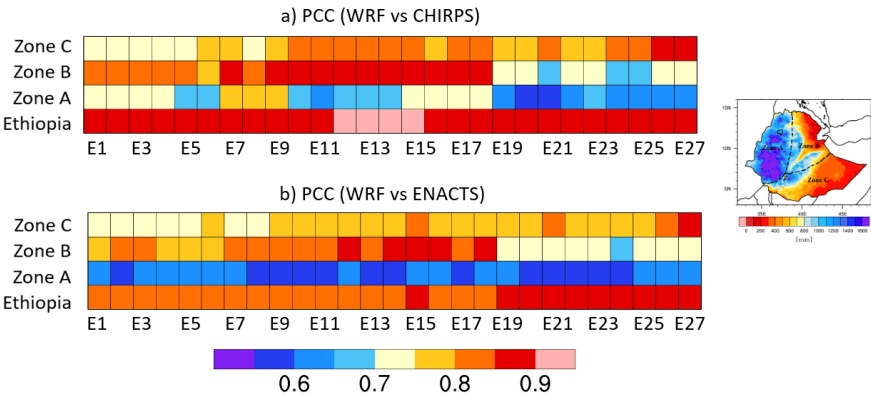

Figure 6. Pattern Correlation Coefficient of seasonal mean JJA precipitation between WRF simulations and CHIRPS a) and ENACTS b) for 3 homogeneous rainfall zones and entire country
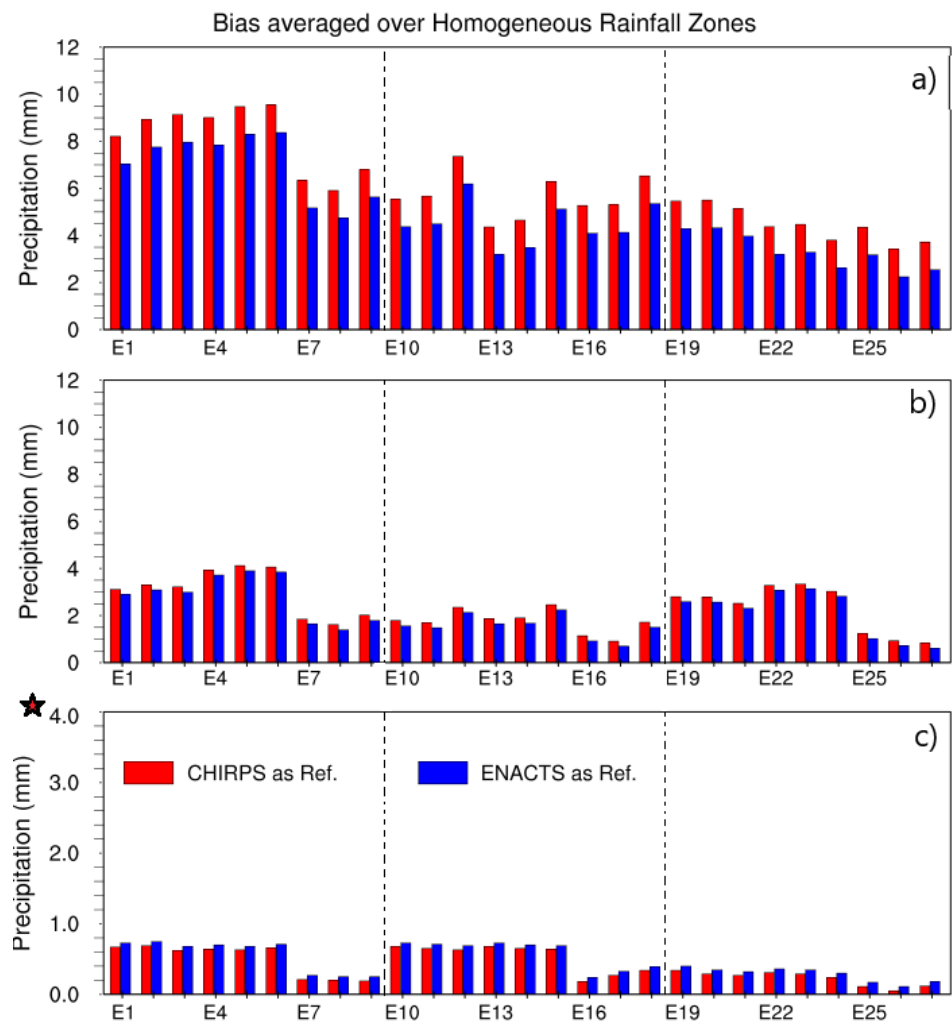
Figure 7. Bias in spatially averaged JJA mean precipitation (mm/day) during 2002 against CHIRPS (red) and ENACTS (blue) datasets for a) Homogeneous zone A, b) and C) Zone C.

The large wet bias over highlands of Ethiopia is in line with previous studies [37, 60,61] that showed biases of simulated rainfall to be large in mountainous regions. For example, Endris *et al.* [61] evaluated the ability of WRF (and other 9 RCMs in CORDEX experiment) to simulate the characteristics of rainfall patterns over East Africa and found that WRF overestimated JJAS rainfall climatology over the Ethiopian highlands. Although the horizontal resolution in this study is quite different from 50km resolution used by Endris *et al.* [61], the model setup used coincides with E1 (i.e., KF CU, WSM5 MP, YSU PBL, Dudhia SW, and RRTM LW radiation schemes).

The performance of WRF simulations in capturing JJA2002 rainfall is further evaluated by spatially averaging seasonal rainfall over three homogeneous rainfall zones (after Diro *et al.* [62]) shown in Fig.6. Figure.7 illustrates simulation bias in JJA2002 mean rainfall spatially averaged over homogeneous zones with respect to CHIRPS and ENACTS datasets. Bias with respect to the two reference datasets show reasonable agreement in terms of both magnitude and inter-simulation differences. However, over Zone A and B, bias magnitude is higher when CHIRPS is used as a reference (by 1.17 mm/day and 0.17 mm/day on average respectively) while for Zone C magnitudes become slightly higher when ENACT is used as reference. As expected, bias is larger for Zone A which encompasses wetter and summer rain receiving areas and ranges from 3.5 mm (E26) to 9.5 mm (E6). For Zone C, which is characterized by a dry summer, all simulations exhibit very similar rainfall rates but slight bias differences can be seen among experiment.

When compared across simulations, the bias is consistently smaller for simulations with Grell CU (i.e., 4.5 mm for zone A) compared to KF and BMJ CU schemes (5.7 mm and 8.1 mm for zone A respectively). For all experiments, bais is smallest when ACM2 PBL is used. The difference is relatively higher when ACM2 is used with KF and BMJ CU schemes. As in the grid-by-grid comparison, no notable pattern can be seen between different MP schemes. The spatial averaging has further demonstrated that KF and Grell CU perform better when used with ACM2 PBL. For example, for zone A out of the 9 simulations using KF CU (i.e., E1-E9) simulations with ACM2 PBL (E7-E9) have an average bias 2.4 mm/day and 3 mm/day smaller than simulations with YSU PBL (E1-E3) and MYJ PBL (E4-E6) respectively. In addition, BMJ CU simulations (E10-18) perform relatively better when used with ACM2 over Zones B and C. Although JJA2002 precipitation is least sensitive to MP schemes and lack notable pattern as compared to CU and PBL schemes, Figures 3, 4 and 7 show some kind of pattern where Morrison MP scheme when used with BMJ CU scheme results in higher magnitude of bias over Zone A. However, a similar pattern seen over Zone B (Figure.7b), E12, E15 and E18) is mostly due to the cancellation of wet and dry biases during spatial averaging.

In terms of sensitivity, the differences exhibited among CU schemes are far more greater than differences seen among PBL and MP schemes indicating that precipitation is most sensitive to changes in CU parameterization. This is expected as the use of different convection schemes leads to substantially different simulations of lower tropospheric circulations and thus significantly affecting simulated rainfall amounts[17,63]. To demonstrate the differences in level of sensitivity among CU, PBL and MP parameterizations, ensemble means are created for three schemes from each parameterization option and presented in Figure 4 and 5 compared to gridded and weather station data respectively. For example, ensemble mean of all 9 experiments using KF, BMJ and Grell CU (Figure 4A-C and Figure 5A-C), regardless of PBL and MP schemes, clearly shows that there is a substantial difference among CU schemes with Grell ensemble showing the least wet bias and KF the largest wet bias. Although there are some notable differences among ensemble means for PBL and MP ensembles, they are significantly smaller than differences among CU schemes. This is expected as the use of different convection schemes leads to substantially different simulations of lower tropospheric circulations and thus significantly affecting simulated rainfall amounts [17,63]. All the above comparisons (i.e., WRF simulation vs gridded observations) were made using weather station precipitation as a reference and show similar bias and sensitivity patterns as in the gridded comparison (Figure. A1 and A2.

Exp	No. of stations in each bias category						Max Bias (in Days)	Avg. Bias (in Days)	Avg. Bias/ physics option
	0-10	10-20	20-30	30-40	40-50	50-70			
E1	5	12	13	10	7	0	46	22.9	KF Only=26.7
E2	6	13	15	7	9	0	47	23.7	BMJ Only=27.8
E3	4	17	10	10	8	0	46	22.7	Grell Only=30.9
E4	4	10	10	8	10	3	53	26.7	YSU Only=26.2
E5	4	9	12	7	13	1	52	27.3	MYJ Only=28.7
E6	5	12	12	6	11	3	54	26.9	ACM2 Only=21.1
E7	12	19	11	6	1	0	44	16.0	WSM6 Only=24.8
E8	13	14	13	5	1	0	42	15.5	LIN Only=25.0
E9	7	15	14	8	2	0	44	18.9	Morr Only=26.2
E10	3	14	12	14	4	1	53	25.6	
E11	4	13	14	17	3	1	52	26.1	
E12	5	9	18	14	4	2	56	27.1	
E13	4	12	19	10	6	2	52	27.1	
E14	4	9	15	16	5	2	53	27.8	
E15	3	12	13	11	8	2	56	29.2	
E16	6	18	17	6	2	0	47	19.3	
E17	10	15	18	5	2	0	43	18.8	
E18	6	14	16	12	4	0	47	23.7	
E19	6	7	10	14	8	6	62	29.2	
E20	4	9	10	14	9	4	59	29.2	
E21	4	10	9	12	10	4	61	29.4	
E22	3	10	10	8	12	6	68	30.9	
E23	3	7	11	8	11	7	65	31.3	
E24	4	9	11	10	11	6	69	30.9	
E25	7	9	16	13	5	2	56	25.4	
E26	3	14	12	14	6	2	56	25.3	
E27	4	13	14	14	4	3	60	26.6	

Table 2. Bias in simulated number of rainy days compared to weather station and number of stations in different rainy day bias categories for 27 WRF simulations. Last column is bias averaged over 9 experiments with similar parameterization schemes for each physics options. For example, KF Only bias is average of 9 experiments with KF CU scheme regardless of parameterization schemes used for PBL and MP.
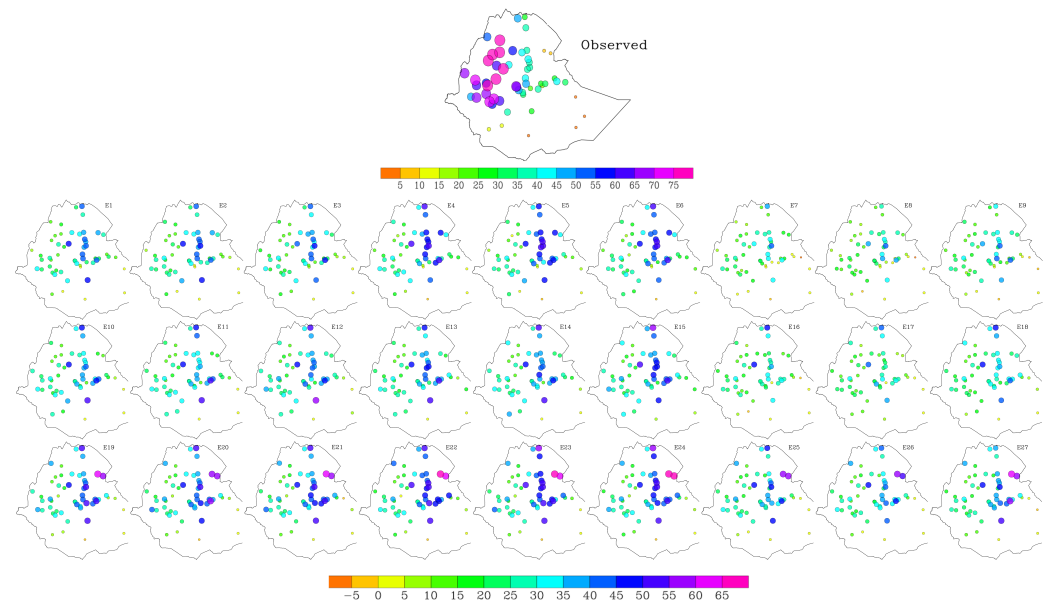
Figure 8. Observed number of rainy days over weather station (top) and bias in WRF simulated number of rainy days during JJA season of 2002 compared to weather stations

3.2. Rainy day frequency and intensity

In this section, WRF simulations are evaluated for their performance in reproducing the observed frequency of rainy days and the various categories of rainfall intensities over the 92-day simulation period. Rainy days are defined as days with the amount of rainfall recorded/simulated at a station/grid-point greater than 1 mm. The observed rainy day frequency (Figure. 8,top) show large spatial variability ranging from 0 days in south and southeastern lowlands where JJA is not main rainy season (e.g., 0, 0, 2 and 3 days over Kebridehar, Gode, Negelle and Degehabour stations) to more than 75 days over stations located in western and northwestern highlands (e.g., Bahirdar, Debremarkos, Chagni, Nekemte and Shambu). The significantly larger magnitude of rainy days over western and northwestern Ethiopia is however expected as these locations receive larger proportion of annual rainfall during summer months. Overall, all WRF simulations have overestimated rainy days except for 4 stations in south and southeast (i.e., Arba Minch, Negelle, Jijiga and Degehabour) that underestimated number rainy days by 1 to 6 days for selected experiments. However, the above cases account only 1.3% of the total station-experiment combination (i.e., 20 cases out of 1512). Out of the 20, 55% occur over Negelle station located in southern Ethiopia while in terms of experiment 50% and 40% involve KF and Grell CU schemes respectively.

Unlike the underestimation, magnitude of the positive bias is significantly large and reaches as high as 69 days over the arid northeastern lowlands for simulation E24 that uses Grell CU, MYJ PBL and MOR MP. As seen in Fig.8, rainy days bias follows a similar spatial pattern across experiments. The most prominent is the larger bias magnitudes over stations located around western escarpment of the Rift Valley (stretching from north to central Ethiopia) and eastern escarpment of Rift Valley (from Bale to eastern highlands) despite these stations having significantly fewer number of observed rainy days (in orders of 30-40 days) compared to those over west and northwest. The other similar pattern is related to stations over south and south eastern lowlands which have the least number of rainy days (less than 15 days) still show the least bias. On the contrary, two station over northeast lowlands show a very high bias that go as high as 69 days for simulations using Grell CU.

Physics Options	Parametrization Schemes	Mean Rainyday Bias (across 54 stations)	ACM2 only Average	Rainyday Bias range
CU	KF	24.82	18.20	16.6 - 30.8
	BMJ	26.29	21.10	18.8 - 31.0
	Grell	31.13	25.80	25.4 - 35.6
PBL	YSU	28.91	NA	25.7 - 32.7
	MYJ	31.64	NA	28.4 - 35.6
	ACM2	21.69	NA	16.6 - 26.0
MP	WSM6	26.95	21.00	17.3 - 35.0
	Lin	26.96	20.30	16.6 - 35.6
	Mor	28.33	23.80	20.8 - 34.7

Figure 9. Summary statistics for rainy day averaged over stations

When comparing rainy days bias among experiments, some notable differences can be seen. In general, all simulations involving Grell CU scheme (E19-E27) show larger bias compared to corresponding experiments utilizing KF and BMJ CU schemes but with similar PBL and MP schemes (Figure 8 and Table 9). When averaged over the 54 stations and experiments, rainy days bias for Grell -only experiments show average bias of around 31 days while KF-only and BMJ-only simulations showing 26.7 and 27.8 days respectively. Rainy days frequency shows an even more sensitivity to choice of PBL schemes where a significant difference in mean bias is seen among experiments using YSU, MYJ and ACM2 (i.e., 28.9, 31.6 and 21.7 days respectively). However, for MP paramerization, the difference among the three schemes is relatively very small (24.8, 25 and 26.2 days for WSM6-, Lin- and Morrison-only schemes respectively). Another notable pattern among experiments is that all CU and MP schemes performed better when used together with ACM2 PBL scheme. For example, for experiments using KF CU and ACM2 PBL (E7, E8 and E9) average bias is 16.8 days and increases to 23.1 days and 27 days when YSU (E1-E3) and MYJ (E4-E6) PBL are is used respectively. As shown above, the same also holds true when bias is averaged over ACM2-, YSU- and MYJ-only experiments. In addition, out of the 9 experiments with the least bias averaged over stations, 6 use ACM2 PBL or KF CU schemes. When considering frequency of rainy days as evaluation criteria, the above findings suggest that E8 (KF CU, ACM2 PBL and Lin MP) followed by E7 (KF CU, ACM2 PBL and Morrison MP) schemes performed relatively better (Table 2).

In terms of daily rainfall intensity, simulations show a general spatial pattern where intensity is overestimated over the highlands (i.e., central, west and northwest Ethiopia) and either slightly overestimated or underestimated the elsewhere (Figure 10. This is somehow a reversed pattern compared to rainy days as areas with larger/smaller rainy day bias have relatively smaller/larger bias in rainfall intensity. This is more evident over station located around the eastern escarpment of the Rift Valley that showed a small positive bias for KF CU simulations and small negative bias for BMJ and Grell CU simulations. As in mean seasonal rainfall, rainfall intensity showed the most sensitivity to choice of CU parameterization options followed by PBL and MP options. This can be seen from the differences among simulations that are the largest when CU options are changed (Table 3). For example, summary of mean absolute bias in intensity (MAB) (Table 3, last column) shows that MAB averaged over similar CU options has a relatively larger differences (1.6mm) compared to differences in PBL (0.4mm) and MP (0.2mm) parameterizations.

Overall, simulations with KF CU showed the largest MAB of 5mm averaged over stations and highest number of stations with MAB > 5mm/day (i.e., 22 stations) and MAB > 10mm/day (9.3 stations). Simulations with Grell CU on the other hand performed relatively better with the lowest MAB 3.4mm/day and lowest number of station with MAB > 5mm/day (15.1 stations) and MAB > 10mm/day (1.4 stations). Even though KF CU simulations have the least bias magnitude in terms of rainy day frequency, this significantly large bias in intensity has resulted in KF simulations to have the highest wet bias in terms of seasonal mean precipitation. Another notable pattern is that BMJ CU performs better when used with MJY PBL (developed together with MYJ CU scheme). For example, MAB increases from 3.2mm/day when BMJ CU is used together with MYJ PBL to 4.2mm/day

and 4.7mm when used with YSU and ACM2 PBL options respectively. The same also holds true for the other statistics in Table 3. When comparing biases in seasonal total precipitation, rainy days, and intensity, the patterns suggested that wet bias over the highlands is still significantly higher compared to the areas showing larger rainy day bias which indicates that source of bias for seasonal precipitation for the highlands is mostly overestimated precipitation intensity while for areas like eastern escapement of the Rift Valley overestimated rainy days mostly accounts for wet bias.

Exp	Bias across stations			MAB	No. Station		MAB/option
	Avg	Min	Max		MAB > 5	MAB >10	
E1	4.1	-8.3	15.1	5.2	24	10	KF-Only =5.0
E2	4.1	-8.0	17.1	5.4	24	10	BMJ-Only=4.0
E3	4.3	-7.0	16.5	5.4	24	9	Grell -Only=3.4
E4	4.0	-8.0	17.3	5.2	22	11	YSU-Only=4.3
E5	4.0	-7.6	16.8	5.1	23	9	MYJ-Only=3.9
E6	4.2	-7.3	17.5	5.3	22	13	ACM2-Only=4.2
E7	3.2	-7.6	14.7	4.4	21	6	WSM6-Only=4.1
E8	2.9	-8.3	15.5	4.7	20	6	Lin-Only=4.1
E9	3.1	-8.9	17.9	4.6	18	10	Morrison-Only=4.3
E10	0.6	-10.4	12.0	3.8	16	3	
E11	0.8	-9.9	9.4	4.2	18	0	
E12	2.0	-9.9	15.0	4.5	21	6	
E13	0.2	-11.0	7.1	3.1	10	2	
E14	0.0	-10.7	6.9	3.0	10	1	
E15	1.2	-10.1	10.9	3.5	14	2	
E16	1.1	-10.2	14.8	4.4	17	4	
E17	1.1	-11.6	15.7	4.8	18	6	
E18	1.9	-9.5	17.9	4.8	22	6	
E19	-0.2	-10.2	8.9	3.4	18	2	
E20	-0.3	-10.3	9.4	3.6	18	2	
E21	-0.6	-11.3	8.2	3.6	19	2	
E22	-0.6	-9.8	7.6	3.3	14	0	
E23	-0.4	-9.3	8.8	3.2	13	0	
E24	-0.9	-10.1	6.9	3.3	12	1	
E25	-1.0	-11.7	9.6	3.7	13	2	
E26	-1.8	-11.7	8.1	3.3	14	2	
E27	-1.8	-12.1	7.9	3.3	15	2	

Table 3. Summary statistics from comparison of simulated and observed (weather stations) rainfall intensity across experiments over grid-points where stations are located in. Last column is bias averaged over physics options (9 members for each).
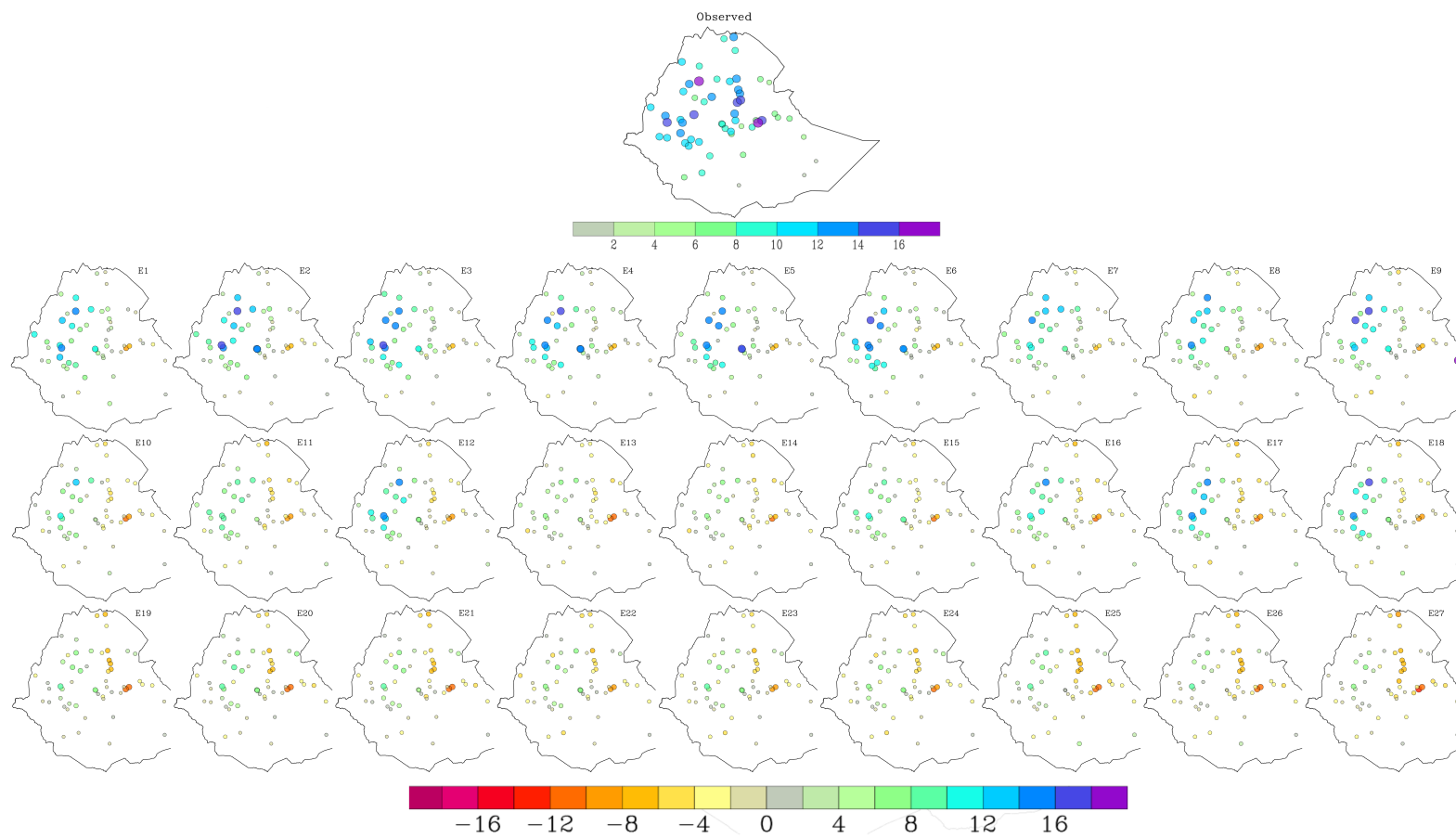
Figure 10. Bias in the mean rainfall intensity (Rainy day total/number of rainy days) during JJA season of 2002 compared to 54 weather stations for 27 WRF experiments

3.3. Radiation scheme

As stated in the methods section, assessing the sensitivity of radiation schemes (long- and short wave) was not part of the primary objectives of this study and additional set of simulations were run (E28, E29 and E30 Table xx) after completion of the other simulations which use fixed options for radiation (i.e., Dudhai for SW and RRTM for LW). All three simulations were compared with E1 which use similar options for CU (KF), PBL (YSU) and MP (WSM6) as the three experiments. E1 uses RRTM scheme for LW and Dudhai for SW; E28 uses RRTMG for LW and Dudhai for SW; E29 used RRTM for LW and RRTMG for SW; E30 uses RRTMG scheme for both LW and SW radiation.

Figures.11 shows comparison of seasonal mean precipitation between radiation experiments with station observation. Given the fact that KF CU scheme is used in all the four simulations, it can be seen that all four simulations have consistently overestimated seasonal mean precipitation with only a single station showing dry bias of 1 mm for experiments E1 and E28. However, significant differences can be observed among them. E29 (RRTM for LW and RRTMG) performed the least with average wet bias of 9.1 mm (range 0-23 mm). Out of the 59 stations used for comparison, E29 showed the highest bias for 41 stations and for 26 out 59 stations show wet bias of 10mm or more. On the other hand, E28 (RRTMG-Dudhai) performed relatively better with wet bias of 5 mm averaged over stations (range -1 mm to 16 mm). In addition, E28 showed the least bias for 43 stations out of 59. E30 and E1 show a similar (close) performance with mean bias of 6 mm (range 0-13mm) and 6.4 mm (range -1-16mm) respectively. It can also be seen that there is a marked difference in wet bias when physics option for SW is changed than LW (i.e., comparing RRTM-Dudhai(Figure.11b) and RRTM-RRTMG (Figure.11d) vs RRTM-Dudhai(Figure.11b) and RRTM-RRTMG (Figure.11c). This result is consistent with previous sensitivity studies over other regions [29](Yuan et al.2012; Awan et al.2011), that have found that shortwave radiation schemes in particular to have a strong precipitation response.

Despite the complex algorithms utilized by RRTMG scheme (i.e.,overlapping cloud fraction for determining grid cloudiness, take into account the concentrations of trace gases, aerosols, ozone,and carbon dioxide[26,53]), the results above do not reflect the complexity of the scheme.(i.e., RRTMG-Dudhai combination better than RRTMG-RRTMG showed better performance). In addition, it deviates from studies (xxx) have suggested that schemes developed together tend to perform better. However, these performances are not necessarily attributed to the limitation of the parameterization schemes but may possibly be the result of other options used for other physics (Cu, PBL and MP). For example, [22] found that the RRTMG radiation, KF cumulus,and YSU PBL physics combination to perform consistently poorly for all their simulations of storm events in Eastern Australia and all our radiation experiments use KF schmeme for CU.

Thus, we performed a preliminary experiment with the following combinations (but keeping RRTMG for both LW and SW), to see how performance is affected by different combinations: E31(KF-YSU-Lin), E32(KF-YSU-Mor), E33(KF-MYJ-WSM6), E34(KF-MYJ-Lin) and E35(KF-MYJ-Mor). Comparison of the mean seasonal precipitation confirms to our initial hypothesis that choices made for other schemes affect the performance. Among the experiments using RRTMG for both LW and SW radiation (i.e., E30-E35), performance seems to be sensitive to choice of Microphysic scheme with those utilizing WSM6 scheme (E30 and E33) showing the least bias (6mm and xxmm respectively). On the other hand, no differences were seen between use of either YSU or MYJ PBL schemes. Although not robust, this might be an indication that RRTMG radiation scheme works well when used with WSM6 MP scheme. However, it should be noted that all radiation experiments use KF for CU and a higher sensitivity might be seen if other CU schemes are tested with RRTMG option for both LW and SW.
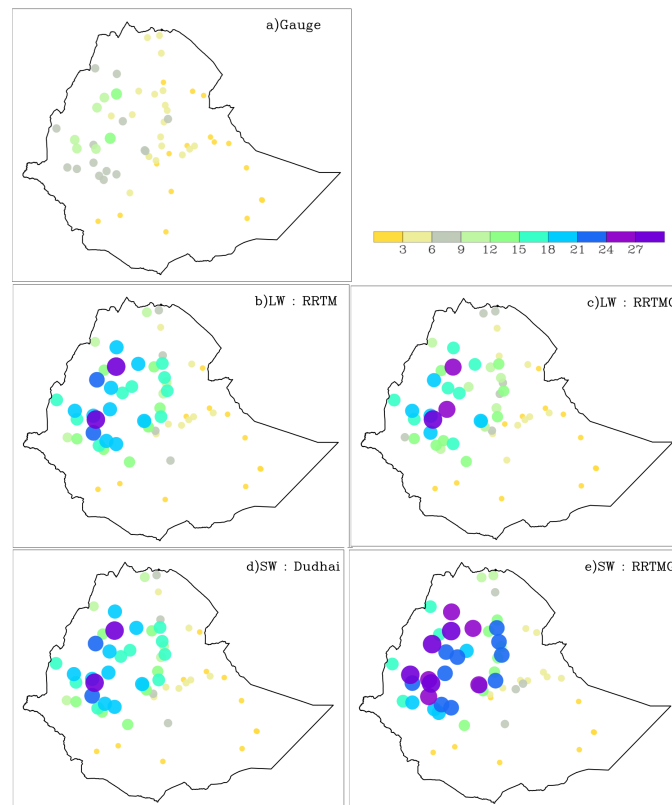
Figure 11. Comparison of Radiation schemes a)CHIRPS, b)LW: RRTM, c)RRTMG, d)SW: Dudhai and e)RRTMG

Comparison of mean seasonal precipitation with gridded products (Figure 12) also showed a similar result as station-based comparison. The RRTM-RRTMG combination (i.e., E29) showed the largest wet bias over western and northwestern parts of the country. The simulated precipitation is two folds of observed precipitation over large portion of the highlands. Although all four radiation experiments overestimated precipitation, the spatial extent of the bias reduces significantly for E1 (RRTM-Dudhai), E30(RRTMG-RRTMG) and E28(RRTMG-Dudhai) in an increasing order. The above results suggest that use of Dudhai scheme for shortwave radiation is the right choice but for longwave radiation, use of RRTM scheme might be questionable or at least require further investigation to choose between RRTM or the more complex RRTMG scheme.
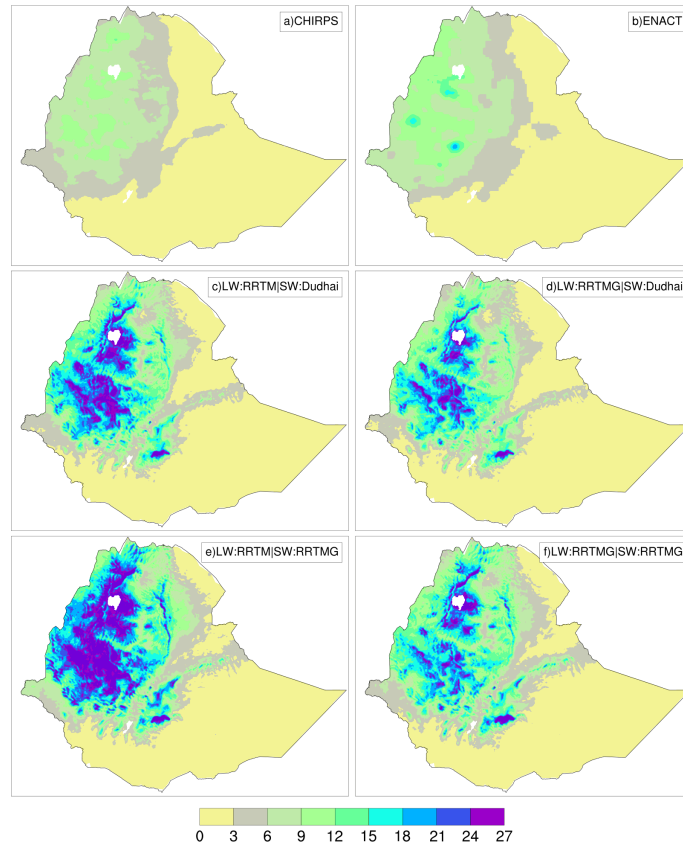
Figure 12. Comparison of Radiation schemes a)CHIRPS, b)ENACTS, c)LW: RRTM, d)RRTMG, e)SW: Dudhai and e)RRTMG

3.4. Ranking

The ranking has been performed using aggregate score that is the summation of normalized MAE_{Stn}, MAE_{Grd}, ERD, R and PCC with equal weight assigned to each. Lower and higher Agg Score indicates better and worse performance respectively. Among the 27 experiments ranked, none of the simulations performed uniformly the best or the worst on all five statistics. However, simulations with Grell CU scheme performed the best in terms of average MAE (both station and grid-based), R and PCC while performed relatively poor in terms of capturing the number of rainy days in a season. It can also be noted that using YSU and MYJ PBL schemes with Grell CU scheme improves performance in capturing the spatial patterns of mean seasonal precipitation but on the expense of deteriorating performance in terms o f rainy days. On the contrary, simulations with KF were poor in terms of MAE and R and performed better in capturing number of rainy days and spatial patterns of seasonal precipitation. The performance of simulations with BMJ CU scheme were in between KF and Grell schemes for most of the statistics but performed the best in terms of R while performance in terms of PCC are relatively poor when when used with ACM2 PBL scheme. The same holds true for KF CU scheme where PCC scores are poor when used with ACM2. However, for statistics other than PCC, simulations that use ACM2 PBL with KF performed better than those using YSU and MYJ.

To better visualize rank scores, we subjectively defined performance category of very good (AS<1.0), good (1.0<AS<2.0), Moderate (2.0<A.0), Poor (3.0<AS<4.0) and very poor (AS>4.0) based on aggregate score and presented them in Figure.13. According to this classification, 22.2%, 55.6%, 18.5% and 3.7% fall under good, moderate, poor and very poor category respectively. It can be seen that there is no significant difference between the MP used as their combinations with the same CU and PBL schemes fall within the same performance category (in most cases). For the CU schemes, 5 out of 9 Grell combinations produce good simulations while KF, on the contrary, performs poorly and very poorly

with YSU and MYJ PBL combinations. KF performance improved to moderate category when combined with ACM2 PBL. Among the three CU physics options, BMJ CU fall under moderate category for all combinations except one that includes WSM6 MP option which falls under good performance category.

Rank	ID	Description	MAE_{Stn}	MAE_{Grd}	ERD	R	PCC	Agg Score
1	E26	Grell -ACM2-Lin	0.00	0.00	0.63	0.15	0.28	1.06
2	E27	Grell -ACM2-Mor	0.00	0.00	0.73	0.16	0.30	1.19
3	E25	Grell -ACM2-WSM6	0.14	0.11	0.65	0.31	0.44	1.65
4	E16	BMJ-ACM2-WSM6	0.33	0.29	0.28	0.00	0.84	1.74
5	E21	Grell -YSU-Mor	0.20	0.31	0.89	0.36	0.05	1.81
6	E20	Grell -YSU-Lin	0.24	0.39	0.87	0.33	0.10	1.93
7	E22	Grell -MYJ-WSM6	0.23	0.33	1.00	0.46	0.03	2.04
8	E14	BMJ-MYJ-Lin	0.27	0.26	0.79	0.39	0.33	2.04
9	E24	Grell -MYJ-Mor	0.19	0.22	0.98	0.59	0.07	2.06
10	E13	BMJ-MYJ-WSM6	0.31	0.22	0.75	0.50	0.29	2.06
11	E19	Grell -YSU-WSM6	0.27	0.39	0.89	0.44	0.10	2.09
12	E17	BMJ-ACM2-Lin	0.36	0.30	0.26	0.21	1.00	2.13
13	E10	BMJ-YSU-WSM6	0.31	0.37	0.67	0.11	0.67	2.13
14	E15	BMJ-MYJ-Mor	0.43	0.45	0.88	0.41	0.04	2.20
15	E23	Grell -MYJ-Lin	0.27	0.34	1.00	0.61	0.00	2.21
16	E7	KF-ACM2-WSM6	0.59	0.45	0.01	0.49	0.75	2.29
17	E8	KF-ACM2-Lin	0.57	0.42	0.00	0.47	0.91	2.38
18	E18	BMJ-ACM2-Mor	0.48	0.40	0.54	0.26	0.75	2.43
19	E12	BMJ-YSU-Mor	0.52	0.56	0.77	0.23	0.40	2.47
20	E11	BMJ-YSU-Lin	0.38	0.38	0.70	0.32	0.71	2.49
21	E9	KF-ACM2-Mor	0.63	0.47	0.20	0.56	0.82	2.69
22	E1	KF-YSU-WSM6	0.83	0.79	0.47	0.56	0.68	3.33
23	E3	KF-YSU-Mor	0.92	0.85	0.45	0.71	0.43	3.36
24	E2	KF-YSU-Lin	0.90	0.88	0.52	0.64	0.66	3.61
25	E4	KF-MYJ-WSM6	0.98	0.93	0.71	0.81	0.35	3.78
26	E6	KF-MYJ-Mor	1.00	0.97	0.72	0.78	0.37	3.83
27	E5	KF-MYJ-Lin	1.00	1.00	0.77	1.00	0.38	4.15

Table 4. Rank of experiments based on Aggregate score. MAE_{Stn} stands for average MAE based on station and MAE_{Grd} stands for mean MAE based on average of all grid points from gridded data. ERD is error in number of rainy days averaged over all weather stations. R is average temporal correlation among all stations while PCC is pattern correlation between observed (gridded) and simulated mean seasonal precipitation. All scores (except AS) are normalized over 27 experiments and AS is the sum of the 5 normalized scores.
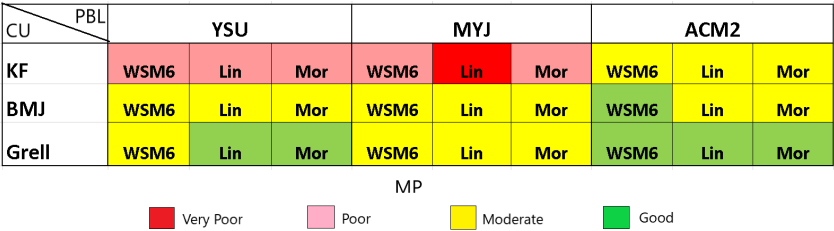
CU \ PBL	YSU			MYJ			ACM2		
KF	WSM6	Lin	Mor	WSM6	Lin	Mor	WSM6	Lin	Mor
BMJ	WSM6	Lin	Mor	WSM6	Lin	Mor	WSM6	Lin	Mor
Grell	WSM6	Lin	Mor	WSM6	Lin	Mor	WSM6	Lin	Mor

MP

Very Poor | Poor | Moderate | Good

Figure 13. Rank category based on Aggregate score

4. Conclusions

In this study, the performance of the WRF model has been investigated with respect to the spatial and temporal rainfall distribution over a domain encompassing Ethiopia (nested domain with 4km) for JJA season of 2002. Different combinations of three CU, three PBL, three MP, two LW and two SW parameterization schemes have been tested in order to select an optimum WRF configuration for further study aimed at assessing the performance of WRF to dynamically downscale seasonal climate prediction from coarse resolution global models over the region. All of the simulations were run over 4-month period from May to August 2002 but evaluation was conducted on 3-month (June-August) with May left as a model spin-up period. WRF simulations were compared with both gridded and gauge observed data. Mean bias, PCC, R and bias in rainy day frequency were used as a criteria to evaluate performance with respect to both spatial and temporal distribution of rainfall.

The findings of this study have demonstrated that a suitable selection of parameterizations can improve performance of simulation which is evident from the range of skill scores among the different experiments (e.g., mean bias xx-xxx, bias in rainy days xxx-xxx). While changing different physical parameterizations, we found out that changes in cumulus parameterization influenced different aspects of summer precipitation the most, with pbl schemes coming next. Out of the 27 simulations (which used Dudhai for SW, RRTM for LW and NOAH land surface model) the combination of Grell -3D convective scheme, ACM2 PBL scheme and Lin Microphysics scheme ranked the top and provided the most realistic simulation in terms of spatial distribution, rainfall totals, and MA bias when compared to observations. A preliminary assessment also revealed that Dudhai shortwave parameterization scheme which was used throughout the 27 simulations, performs better than RRTMG scheme while no conclusive evidence was found between RRTMG and RRTM (default used in this study) longwave schemes. Thus, this set up was chosen for continued study.

We cannot exclude that configurations that were not tested here might potentially perform better. In addition, parameterizations kept fixed in this study, such as (land surface model) or other simulation setups like number of vertical level, size and placement of domains and source of initial and boundary conditions) can also affect the results and likely reduce the model biases. However, any additional consideration of these factors would tremendously increase number of simulations (e.g., consideration of only one additional land surface model would double number of simulations), being prohibitive with respect to the available computational resources and time. In addition, Note that these errors should also be evaluated in view of the uncertainties present in observational data sets,which can be high especially over remote areas or mountainous regions[17]. Nevertheless, this study can serve as a reference for potential WRF users in the region to further investigate use of WRF as a dynamical downscaling tool for seasonal rainfall forecasting or for climate variability and change studies over the region. For climate purposes, although longer periods (including both wet and dry years) are preferable when identifying configurations that describe as correctly as possible the local climate rather than particular periods [38], we have conducted the experiment over a single season (i.e., dry summer of 2002) which might limit the robustness of results. Thus, extending the study with additional "wet" and "normal" seasons will possibly improve reliability of results. In this study, we followed the traditional technique where model results are directly compared with in situ observations,

although this is not a like-with-like comparison. Site-specific measurements describe conditions at single stations affected by very local characteristics, whereas the WRF outputs define average values of the variables over a gridbox[38]. In areas with complex terrain, the "representation error" is of particular importance because the station might be at the extreme of the cell topographical diversity. Thus, other suggested techniques of evaluation such as use of upscaled observations [64] should be tested.

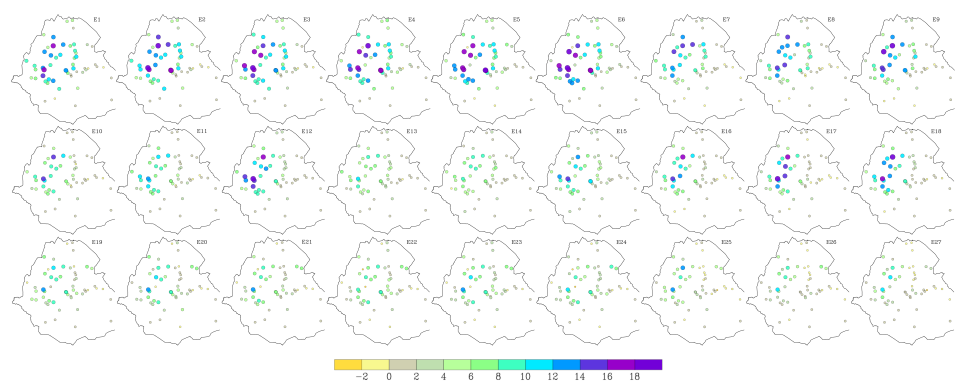## Appendix A Mean seasonal precipitation

Figure A1. Bias in simulated JJA mean precipitation (mm/day) during 2002 compared to weather stations. First column KF CU scheme, 2nd BMJ CU scheme and 3rd Grell CU scheme. 2-4 row YSU PBL scheme, 5-7 MYJ PBL scheme and 8-10 ACM2 PBL scheme. Rows 2, 5 and 8 WSM6 MP scheme; rows 3,6,and 9 Lin scheme and rows 5,7 and 10 Morrison MP scheme.
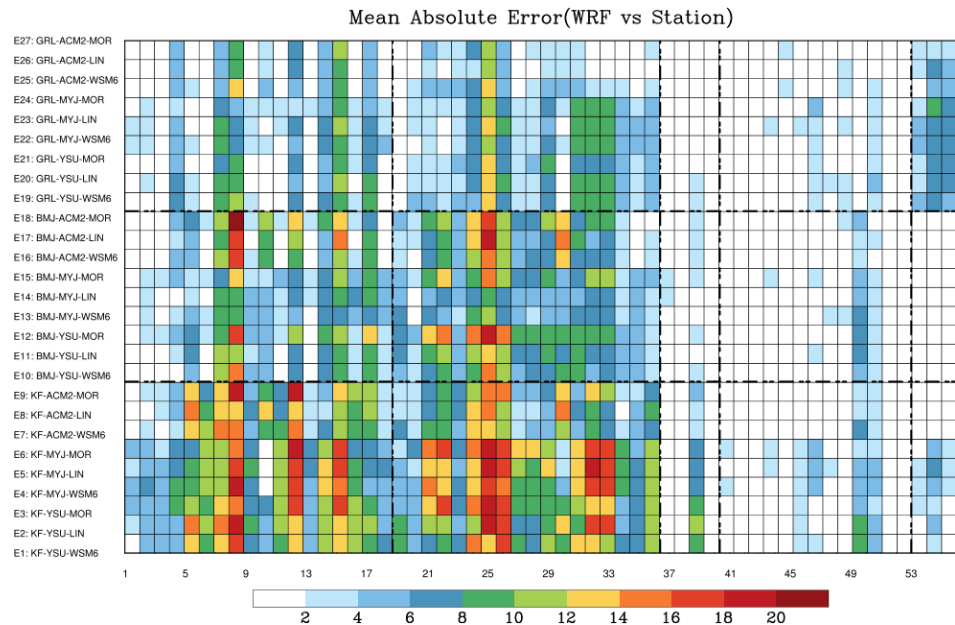
Figure A2. Mean absolute error in WRF simulated mean JJA precipitation (mm/day) during 2002 compared to station data for 58 stations (x-axis) and 27 WRF simulations(y-axis).

References

1. DíEz, E.; Orfila, B.; FríAs, M.D.; FernáNdez, J.; CofiñO, A.; GutiéRrez, J.M. Downscaling ECMWF seasonal precipitation forecasts in Europe using the RCA model. *Tellus A: Dynamic Meteorology and Oceanography* **2011**, *63*, 757–762, [https://doi.org/10.1111/j.1600-0870.2011.00523.x]. doi:10.1111/j.1600-0870.2011.00523.x.

2. Saha, S.; Moorthi, S.; Wu, X.; Wang, J.; Nadiga, S.; Tripp, P.; Behringer, D.; Hou, Y.T.; ya Chuang, H.; Iredell, M.; et al. The NCEP Climate Forecast System Version 2. *Journal of Climate* **15 Mar. 2014**, *27*, 2185 – 2208. doi:10.1175/JCLI-D-12-00823.1.

3. Stockdale, T. SEAS5 user guide **2021**. doi:10.21957/2y67999y.

4. Zhong, A.; Hudson, D.; Alves, O.; Wang, G.; Hendon, H. Predictive Ocean Atmosphere Model for Australia (POAMA). 10th EMS Annual Meeting, 2010, pp. EMS2010–16.

5. Robertson, A.W.; Qian, J.H.; Tippett, M.K.; Moron, V.; Lucero, A. Downscaling of Seasonal Rainfall over the Philippines: Dynamical versus Statistical Approaches. *Monthly Weather Review* **2012**, *140*, 1204–1218. doi:10.1175/mwr-d-11-00177.1.

6. Yuan, X.; Liang, X.Z.; Wood, E.F. WRF ensemble downscaling seasonal forecasts of China winter precipitation during 1982–2008. *Climate Dynamics* **2012**, *39*, 2041–2058. doi:10.1007/s00382-011-1241-8.

7. Davis, N.; Bowden, J.; Semazzi, F.; Xie, L.; Önol, B. Customization of RegCM3 Regional Climate Model for Eastern Africa and a Tropical Indian Ocean Domain. *Journal of Climate* **2009**, *22*, 3595–3616. doi:10.1175/2009JCLI2388.1.

8. Nikulin, G.; Asharaf, S.; Magariño, M.; Calmanti, S.; Cardoso, R.M.; Bhend, J.; Fernández, J.; Frías, M.; Fröhlich, K.; Früh, B.; et al. Dynamical and statistical downscaling of a global seasonal hindcast in eastern Africa. *Climate Services* **2018**. doi:10.1016/j.cliser.2017.11.003.

9. Kerandi, N.M.; Laux, P.; Arnault, J.; Kunstmann, H. Performance of the WRF model to simulate the seasonal and interannual variability of hydrometeorological variables in East Africa: a case study for the Tana River basin in Kenya. *Theoretical and Applied Climatology* **2016**, pp. 1–18. doi:10.1007/s00704-016-1890-y.

10. Diro, G.T. Skill and economic benefits of dynamical downscaling of ECMWF ENSEMBLE seasonal forecast over southern Africa with RegCM4. *International Journal of Climatology* **2015**, *36*, 675–688. doi:10.1002/joc.4375.

11. Diro, G.T.; Tompkins, A.M.; Bi, X. Dynamical downscaling of ECMWF Ensemble seasonal forecasts over East Africa with RegCM3. *Journal of Geophysical Research: Atmospheres (1984–2012)* **2012**, *117*. doi:10.1029/2011JD016997.

12. Yuan, C.; Tozuka, T.; Landman, W.A.; Yamagata, T. Dynamical seasonal prediction of Southern African summer precipitation. *Climate Dynamics* **2014**, *42*, 3357–3374. doi:10.1007/s00382-013-1923-5.

13. Phan-Van, T.; Nguyen-Xuan, T.; Nguyen, H.V.; Laux, P.; Pham-Thanh, H.; Ngo-Duc, T. Evaluation of the NCEP Climate Forecast System and Its Downscaling for Seasonal Rainfall Prediction over Vietnam. *Weather and Forecasting* **2018**. doi:10.1175/waf-d-17-0098.1.

14. Ogwang, B.A.; Chen, H.; Li, X.; Gao, C. Evaluation of the capability of RegCM4.0 in simulating East African climate. *Theoretical and Applied Climatology* **2016**, *124*, 303–313. doi:10.1007/s00704-015-1420-3.

15. Siegmund, J.; Bliefernicht, J.; Laux, P.; Kunstmann, H. Toward a seasonal precipitation prediction system for West Africa: Performance of CFSv2 and high-resolution dynamical downscaling. *Journal of Geophysical Research: Atmospheres* **2015**, *120*, 7316–7339. doi:10.1002/2014jd022692.

16. Cheneka, B.R.; Brienen, S.; Fröhlich, K.; Asharaf, S.; Früh, B. Searching for an Added Value of Precipitation in Downscaled Seasonal Hindcasts over East Africa: COSMO-CLM Forced by MPI-ESM. *Advances in Meteorology* **2016**, *2016*. doi:10.1155/2016/4348285.

17. Giorgi, F.; Mearns, L.O. Introduction to special section: Regional Climate Modeling Revisited. *Journal of Geophysical Research: Atmospheres (1984–2012)* **1999**, *104*, 6335–6352. doi:10.1029/98JD02072.

18. Gao, X.; Shi, Y.; Han, Z.; Wang, M.; Wu, J.; Zhang, D.; Xu, Y.; Giorgi, F. Performance of RegCM4 over major river basins in China. *Advances in Atmospheric Sciences* **2017**, *34*, 441–455.

19. Xue, Y.; Janjic, Z.; Dudhia, J.; Vasic, R.; Sales, F. A review on regional dynamical downscaling in intraseasonal to seasonal simulation/prediction and major factors that affect downscaling ability. *Atmospheric Research* **2014**, *147*, 68–85. doi:10.1016/j.atmosres.2014.05.001.

20. Skamarock, W.; Klemp, J.; Dudhia, J.; Gill, D.; Barker, D.; Duda, M.; Huang, X.; Wang, W.; Powers, J. A description of the advanced research WRF Version 3, NCAR tech note NCAR/TN 475 STR, 125 pp. *Available from: UCAR Communications, PO Box* **2008**, *3000*.

21. Ekström, M. Metrics to identify meaningful downscaling skill in WRF simulations of intense rainfall events. *Environmental Modelling Software* **2016**, *79*, 267–284. doi:10.1016/j.envsoft.2016.01.012.

22. Evans, J.P.; Ekström, M.; Ji, F. Evaluating the performance of a WRF physics ensemble over South-East Australia. *Climate Dynamics* **2012**, *39*, 1241–1258. doi:10.1007/s00382-011-1244-5.

23. Ruiz, J.J.; Saulo, C.; Nogués-Paegle, J. WRF Model Sensitivity to Choice of Parameterization over South America: Validation against Surface Variables. *Monthly Weather Review* **2010**, *138*, 3342–3355. doi:10.1175/2010mwr3358.1.

24. Borge, R.; Alexandrov, V.; Vas, J.d.; Lumbreras, J.; Rodríguez, E. A comprehensive sensitivity analysis of the WRF model for air quality applications over the Iberian Peninsula. *Atmospheric Environment* **2008**, *42*, 8560–8574. doi:10.1016/j.atmosenv.2008.08.032.

25. Flaounas, E.; Bastin, S.; Janicot, S. Regional climate modelling of the 2006 West African monsoon: sensitivity to convection and planetary boundary layer parameterisation using WRF. *Climate Dynamics* **2011**, *36*, 1083–1105. doi:10.1007/s00382-010-0785-3.

26. Kala, J.; Andrys, J.; Lyons, T.J.; Foster, I.J.; Evans, B.J. Sensitivity of WRF to driving data and physics options on a seasonal time-scale for the southwest of Western Australia. *Climate Dynamics* **2015**, *44*, 633–659. doi:10.1007/s00382-014-2160-2.

27. Ratna, S.B.; Ratnam, J.V.; Behera, S.K.; Rautenbach, C.J.d.; Ndarana, T.; Takahashi, K.; Yamagata, T. Performance assessment of three convective parameterization schemes in WRF for downscaling summer rainfall over South Africa. *Climate Dynamics* **2014**, *42*, 2931–2953. doi:10.1007/s00382-013-1918-2.

28. Crétat, J.; Pohl, B.; Richard, Y.; Drobinski, P. Uncertainties in simulating regional climate of Southern Africa: sensitivity to physical parameterizations using WRF. *Climate Dynamics* **2012**, *38*, 613–634. doi:10.1007/s00382-011-1055-8.

29. Pohl, B.; Crétat, J.; Camberlin, P. Testing WRF capability in simulating the atmospheric water cycle over Equatorial East Africa. *Climate Dynamics* **2011**, *37*, 1357–1379. doi:10.1007/s00382-011-1024-2.

30. Funk, C.; Peterson, P.; Landsfeld, M.; Pedreros, D.; Verdin, J.; Shukla, S.; Husak, G.; Rowland, J.; Harrison, L.; Hoell, A.; et al. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data* **2015**, *2*, sdata201566. doi:10.1038/sdata.2015.66.

31. Dinku, T.; Block, P.; Sharoff, J.; Hailemariam, K.; Osgood, D.; Corral, J.d.; Cousin, R.; Thomson, M.C. Bridging critical gaps in climate services and applications in africa. *Earth Perspectives* **2014**, *1*, 15. doi:10.1186/2194-6434-1-15.

32. Dinku, T.; Hailemariam, K.; Maidment, R.; Tarnavsky, E.; Connor, S. Combined use of satellite estimates and rain gauge observations to generate high-quality historical rainfall time series over Ethiopia. *International Journal of Climatology* **2014**, *34*, 2489–2504. doi:10.1002/joc.3855.

33. NCAR. The NCAR Command Language (NCL), 2016. doi:10.5065/D6WD3XH5.

34. Dinku, T.; Funk, C.; Peterson, P.; Maidment, R.; Tadesse, T.; Gadain, H.; Ceccato, P. Validation of the CHIRPS satellite rainfall estimates over eastern Africa. *Quarterly Journal of the Royal Meteorological Society* **2018**, *144*, 292–312, [https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3244]. doi:10.1002/qj.3244.

35. Bayissa, Y.; Tadesse, T.; Demisse, G.; Shiferaw, A. Evaluation of Satellite-Based Rainfall Estimates and Application to Monitor Meteorological Drought for the Upper Blue Nile Basin, Ethiopia. *Remote Sensing* **2017**, *9*. doi:10.3390/rs9070669.

36. Saha, S.; Moorthi, S.; Pan, H.L.; Wu, X.; Wang, J.; Nadiga, S.; Tripp, P.; Kistler, R.; Woollen, J.; Behringer, D.; et al. NCEP Climate Forecast System Reanalysis (CFSR) 6-hourly Products, January 1979 to December 2010, 2010.

37. Segele, Z.T.; Leslie, L.M.; Lamb, P.J. Evaluation and adaptation of a regional climate model for the Horn of Africa: rainfall climatology and interannual variability. *International Journal of Climatology* **2009**, *29*, 47–65. doi:10.1002/joc.1681.

38. Argüeso, D.; Hidalgo-Muñoz, J.M.; Gámiz-Fortis, S.R.; Esteban-Parra, M.J.; Dudhia, J.; Castro-Díez, Y. Evaluation of WRF Parameterizations for Climate Studies over Southern Spain Using a Multistep Regionalization. *Journal of Climate* **2011**, *24*, 5633–5651, [ttps://doi.org/10.1175/JCLI-D-11-00073.1]. doi:10.1175/JCLI-D-11-00073.1.

39. Kain, J.S. The Kain–Fritsch convective parameterization: an update. *Journal of Applied Meteorology* **2004**, *43*, 170–181.

40. Janjić, Z.I. The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Monthly Weather Review* **1994**, *122*, 927–945.

41. Janjić, Z.I. Comments on "Development and Evaluation of a Convection Scheme for Use in Climate Models". *Journal of the Atmospheric Sciences* **2000**, *57*, 3686–3686, [https://doi.org/10.1175/1520-0469(2000)057<3686:CODAEO>2.0.CO;2]. doi:10.1175/1520-0469(2000)057<3686:CODAEO>2.0.CO;2.

42. Grell, G.A. Prognostic evaluation of assumptions used by cumulus parameterizations. *Monthly Weather Review* **1993**, *121*, 764–787.

43. Grell, G.A.; Dévényi, D. A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters* **2002**, *29*.

44. Hong, S.Y.; Noh, Y.; Dudhia, J. A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes. *Monthly Weather Review* **2006**, *134*, 2318–2341, [https://doi.org/10.1175/MWR3199.1]. doi:10.1175/MWR3199.1.

45. Pleim, J.E. A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing. *Journal of Applied Meteorology and Climatology* **2007**, *46*, 1383–1395.

46. Jiménez, P.A.; Dudhia, J.; González-Rouco, J.F.; Navarro, J.; Montávez, J.P.; García-Bustamante, E. A Revised Scheme for the WRF Surface Layer Formulation. *Monthly Weather Review* **2012**, *140*, 898–918, [https://doi.org/10.1175/MWR-D-11-00056.1]. doi:10.1175/MWR-D-11-00056.1.

47. Janjić, Z.I. The Step-Mountain Coordinate: Physical Package. *Monthly Weather Review* **1990**, *118*, 1429–1443, [https://doi.org/10.1175/1520-0493(1990)118<1429:TSMCPP>2.0.CO;2]. doi:10.1175/1520-0493(1990)118<1429:TSMCPP>2.0.CO;2.

48. Hong, S.Y.; Lim, J.O.J. The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc* **2006**, *42*, 129–151.

49. Lin, Y.L.; Farley, R.D.; Orville, H.D. Bulk parameterization of the snow field in a cloud model. *Journal of Climate and Applied Meteorology* **1983**, *22*, 1065–1092.

50. Morrison, H.; Thompson, G.; Tatarskii, V. Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one-and two-moment schemes. *Monthly weather review* **2009**, *137*, 991–1007.

51. Mlawer, E.J.; Taubman, S.J.; Brown, P.D.; Iacono, M.J.; Clough, S.A. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres* **1997**, *102*, 16663–16682.

52. Dudhia, J. Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *Journal of the Atmospheric Sciences* **1989**, *46*, 3077–3107.

53. Iacono, M.J.; Delamere, J.S.; Mlawer, E.J.; Shephard, M.W.; Clough, S.A.; Collins, W.D. Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research: Atmospheres* **2008**, *113*.

54. Tewari, M.; Chen, F.; Wang, W.; Dudhia, J.; LeMone, M.; Mitchell, K.; Ek, M.; Gayno, G.; Wegiel, J.; Cuenca, R. Implementation and verification of the unified NOAH land surface model in the WRF model. 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, 2004, Vol. 1115.

55. Wilks, D.S. *Statistical methods in the atmospheric sciences*; Vol. 100, Academic press, 2011.

56. Pattern Correlation Coefficient. https://www.ncl.ucar.edu/Document/Functions/Contributed/pattern_cor.shtml. Accessed: 25 Apr. 2017. 2017-04-25.

57. Leung, L.R.; Ghan, S. A subgrid parameterization of orographic precipitation. *Theoretical and Applied Climatology* **1995**, *52*, 95–118.

58. Zeleke, T.; Giorgi, F.; Tsidu, G.M.; Diro, G.T. Spatial and temporal variability of summer rainfall over Ethiopia from observations and a regional climate model experiment. *Theoretical and Applied Climatology* **2013**, *111*, 665–681. doi:10.1007/s00704-012-0700-4.

59. Gbode, I.E.; Dudhia, J.; Ogunjobi, K.O.; Ajayi, V.O. Sensitivity of different physics schemes in the WRF model during a West African monsoon regime. *Theoretical and Applied Climatology* **2019**, *136*, 733–751. doi:10.1007/s00704-018-2538-x.

60. Awan, N.K.; Truhetz, H.; Gobiet, A. Parameterization-Induced Error Characteristics of MM5 and WRF Operated in Climate Mode over the Alpine Region: An Ensemble-Based Analysis. *Journal of Climate* **2011**, *24*, 3107–3123. doi:10.1175/2011JCLI3674.1.

61. Endris, H.S.; Omondi, P.; Jain, S.; Lennard, C.; Hewitson, B.; Chang'a, L.; Awange, J.L.; Dosio, A.; Ketiem, P.; Nikulin, G.; et al. Assessment of the Performance of CORDEX Regional Climate Models in Simulating East African Rainfall. *Journal of Climate* **2013**, *26*, 8453–8475, [http://dx.doi.org/10.1175/JCLI-D-12-00708.1]. doi:10.1175/JCLI-D-12-00708.1.

62. Diro, G.; Grimes, D.; Black, E.; O'Neill, A.; Pardo-Iguzquiza, E. Evaluation of reanalysis rainfall estimates over Ethiopia. *International Journal of Climatology* **2009**, *29*, 67–78. doi:10.1002/joc.1699.

63. Pal, J.S.; Giorgi, F.; Bi, X.; Elguindi, N.; Solmon, F.; Gao, X.; Rauscher, S.A.; Francisco, R.; Zakey, A.; Winter, J.; et al. Regional climate modeling for the developing world: the ICTP RegCM3 and RegCNET. *Bulletin of the American Meteorological Society* **2007**, *88*, 1395–1410.

64. Göber, M.; Zsótér, E.; Richardson, D.S. Could a perfect model ever satisfy a naïve forecaster? On grid box mean versus point verification. *Meteorological Applications* **2008**, *15*, 359–365. doi:10.1002/met.78.