 doi:10.20944/preprints201710.0199.v1

Article

Improving Quantitative Rainfall Prediction Using Ensemble Analogues in the Tropics: Case study of Uganda

Isaac Mugume[1], Michel d. S. Mesquita[2,‡], Yazidhi Bamutaze[1,‡], Didier Ntwali[3,‡], Charles Basalirwa[1,‡], Daniel Waiswa[1,‡], Joachim Reuder[4,‡], Revocatus Twinomuhangi[1,‡], Fredrick Tumwine[1,‡], Triphonia Jakob Ngailo[5] and Bob Alex Ogwang[6]

1 Department of Geography, Geoinformatics and Climatic Sciences, Makerere University, P. O. Box 7062, Kampala, Uganda; imugume@caes.mak.ac.ug
2 Uni Research Climate, Bjerknes Centre for Climate Research, Bergen, Norway; Michel.Mesquita@uni.no
3 Institute of Atmospheric Physics Laboratory for Middle Atmosphere and Global Environmental Observation, University of Chinese Academy of Sciences, P.R. China; ntwalididi@yahoo.fr
4 Geophysical Institute, University of Bergen, Allegaten 70, 5007 Bergen, Norway; Joachim.Reuder@uib.no
5 Department of General Studies, Dar es Salaam Institute of Technology, P.O. Box 2958, Dar-es-Salaam, Tanzania; triphongailo@gmail.com
6 Uganda National Meteorological Authority, P. O. Box 7025, Kampala, Uganda; bob_ogwang@yahoo.co.uk
* Correspondence: imugume@caes.mak.ac.ug; Tel.: +256-701-349477
‡ These authors contributed equally to this work.

Abstract: Accurate and timely rainfall prediction enhances productivity and can aid proper planning in sectors such as agriculture, health, transport and water resources. This study is aimed at improving rainfall prediction using ensemble methods. It first assesses the performance of six convective schemes (Kain–Fritsch (**KF**); Betts–Miller–Janjić (**BMJ**); Grell–Fretas (**GF**); Grell 3D ensemble (**G3**); New–Tiedke (**NT**) and Grell–Devenyi (**GD**)) using the root mean square error (RMSE) and mean error (ME) focusing on the March–May 2013 rainfall period over Uganda. 18 ensemble members are generated from the three best performing convective schemes (i.e. **KF, GF & G3**). The performance of three ensemble methods (i.e. ensemble mean (**EM**); ensemble mean analogue (**EMA**) and multi–member analogue ensemble (**MAEM**)) is also analyzed using the RMSE and ME. The **EM** presented a smaller RMSE compared to individual schemes (EM:10.02; KF:23.96; BMJ:26.04; GF:25.85; G3:24.07; NT:29.13 & GD:26.27) and a better bias (EM:-1.28; KF:-1.62; BMJ:-4.04; GF:-3.90; G3:-3.62; NT:-5.41 & GD:-4.07). The **EMA** and **MAEM** presented 13 out of 21 stations & 17 out of 21 stations respectively with smaller RMSE compared to **EM** thus demonstrating additional improvement in predictive performance. The **MAEM** is a new approach proposed and described in the study.

Keywords: Ensemble mean; Analogue ensemble mean; Multi–member analogue ensemble mean; Quantitative rainfall prediction

---

1. Introduction

Rainfall is a key climatic element that has consequences on key production sectors including agriculture [1,2], health [3], electricity generation [4] and water resources [5,6] among others. Over Eastern Africa, the rainfall distribution and quantity is influenced by many factors such as the Inter–Tropical Convergence Zone, El Niño/La Niña episodes, Indian Ocean Dipole and extra-tropical weather systems [7,8]. The spatial and temporal variability of rainfall makes its quantitative prediction a challenge [8,9]. However, according to He et al. [5] and Jie et al.[10], rainfall can be predicted quantitatively up to 7 days.

Some of the scientific ways of predicting rainfall quantitatively have been suggested such as: the use of radar [8] and the use of Numerical Weather Prediction (NWP) models [6]. The use of radar is considered superior at short–range forecasts due to better spatial representation and assimilation of the initial rainfall estimates. Unfortunately the radar–based method's accuracy degrades with increasing temporal resolution due to poor resolving of the growth and decay of rainfall for long lead times [5]. The NWP models usually have higher skill because they can represent the dynamics and physics of the atmosphere [6]. The limitation of NWP models arise due model's failure to resolve sub–grid processes but it is addressed using parameterization schemes [11]. An additional scientific method is the use of statistical models such as regression [9].

Due to spatial–temporal variability in performance of NWP models (e.g. the presence of orographic bias [12] and the limit to predictability with increasing prediction period [13]), errors often arise that introduce uncertainty. Therefore, a couple of deterministic NWP models can be run to improve quantitative precipitation forecasts (QPF) and help to quantify uncertainty. He et al. [5] noted that ensemble QPF normally produces a higher skill in terms of quantity and occurrence time than the individual ensemble members.

According to the European Center for Medium–range Weather Forecasts (ECMWF [14]), which runs 51–ensemble members, an ensemble prediction consists of simultaneously running multiple forecasts (i.e. 'ensemble members') with varying initial conditions and slightly perturbed physical conditions to represent uncertainty in initial conditions and also produce a range of possible weather. It is different from the model output statistics introduced by Glahn & Lowry[15] which is a method of determining the statistical relationship between the predictand and the variables predicted by a numerical model [16]. The output from the ensemble members are then statistically post–processed to obtain a skillful probabilistic forecast [1,17] which addresses the uncertainty inherent in the initial conditions and associated model imperfections faced by a deterministic NWP model [18,19]. This is because the ensemble spread gives the measure of the uncertainty of the prediction [10,20]. However, quantifying absolute uncertainty presents additional challenge due to the bias inherent in the models used [20].

The ensemble mean is normally has a smaller error compared to the mean error of individual deterministic forecasts [21,22]; it outperforms climatology; quantify uncertainty in prediction [23] and presents high probability of precipitation prediction [10]. For this reason, the ensemble mean is the most widely used tool and is normally used as a deterministic forecast [21]. However studies such as He et al.[5], Coiffier[13], Evans et al.[24] and many others, suggest that much of the results given by the ensemble mean could be obtained using a single deterministic forecast and then improved by statistically correcting it. The major limitation of using this approach, is that, there could be a chance that the statistical formulation process may not accurately represent extreme weather. A wide array of ensemble members giving an adequate ensemble spread was found skillful for short and medium term range prediction (i.e. less than 7 days) and also to assist in dividing ensemble rainfall data into sub–samples [25].

Ensemble members can be obtained in many ways, such as running NWP models with varying initial conditions [22], perturbing the physical parameterization schemes of the model [26], initializing the models at the different times (time–lagged) [10], combining output from different NWP models (multi–model ensemble) [10,27]. It is important to integrate multiple methods of generating ensemble members because, for example, perturbing initial conditions or model physics has been found to be under-dispersive [22,25].

Although ensemble rainfall prediction has been extensively covered by previous studies e.g. He et al.[5], Fritsch et al.[27], Redmond et al.[26] and many others, majority of these studies employed ensemble mean for quantitative rainfall prediction. The ensemble mean analogue has been used in wind speed forecasting (e.g. Vanvyve et al.[28]) and this study employs it in rainfall prediction. The performance of ensemble mean, ensemble mean analogue, including a new approach 'multi–member analogue ensemble' is assessed. 18 ensemble members are derived from the 3 best performing cumulus schemes. The ensemble members are generated from combining the convection schemes, model perturbation, and time–lagging.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 October 2017 doi:10.20944/preprints201710.0199.v1

2. Data and Methods

2.1. Data sources

The study used rain gauge rainfall data of March–May 2013 from 21 weather stations of Uganda (i.e. Figure, 1). The rainfall data were obtained from the Uganda National Meteorological Authority (UNMA) and quality controlled by checking it for completeness. The missing gaps in some stations were filled using normal ratio method described by Mugume et al.[8].

The rainfall data were then compared to the ensemble generated rainfall data over the same period. The input data to initialize the deterministic NWP model were obtained from the National Centers for Environmental Prediction (NCEP) final reanalysis [29] at a resolution of $1^o \times 1^o$, covering the period of study.

2.2. The study area

The study was conducted over Uganda which geographically approximately extends from 29.55^oE to 35.09^oE and 1.50^oS to 4.36 ^{o}N . The country largely experiences tropical savanna climate with isolated cases like the Lake Victoria Basin experiencing equatorial climate [30]. Additional climatological zoning has been carried out by Basalirwa[31] who delineated the country into homogeneous zones. The 21 stations used in the study are representative of the major political regions e.g. the northern region (Arua, Gulu, Lira, & Kitgum); the eastern region (Serere, Soroti, Buginyanya & Tororo); the western region (Masindi, Kasese, Bushenyi, Mbarara & Kabale); the cattle corridor (is a diagonal stretch from south western Uganda to north eastern Uganda, has Ntusi) and the Lake Victoria Basin (Entebbe, Makerere, Kituza, Namulonge, Jinja, Kamenyamigo & Kibanda).

The rainfall over Uganda exhibits large spatial and temporal variations with the MAM rainfall period starting in March in most areas except the norther region where the rains normally start in April [2]. With exception of the northern region, which experience a unimodal rainfall distribution, Uganda generally experience two rainfall seasons (i.e. March–May and September–November). The March–May (MAM) seasonal rainfall over Uganda is generally influenced by the Inter–Tropical Convergence Zone [2,8]; the monsoon winds of East Africa [32,33]; the Indian ocean dipole [2,7]; the humid Congo airmass [6]; the tropical cyclones, semi–permanent subtropical anticyclones and easterly waves [7,8]; the complex topography [33], vegetation and inland water bodies which modulate local rainfall [8,34]. The Inter–Tropical Convergence Zone migrates north and southwards over the equator two times a year, which brings two rainfall seasons in the region i.e. the March–May and September–November seasons.

2.3. Experimental design

The ensemble members were generated from (i) initializing models at different times (i.e. time–lagged ensemble members); (ii) varying the cumulus parameterization schemes; (iii) perturbing the cumulus parameterization schemes; and (iv) combination of (i) to (iii). We first assessed the performance of six cumulus parameterization schemes (i.e. Kain–Fritsch (**KF**); Betts–Miller–Janjić (**BMJ**); Grell–Fretas (**GF**); Grell 3D ensemble (**G3**); New–Tiedke (**NT**) and Grell–Devenyi (**GD**)) using the root mean square error (*A*) and the mean error (*B*) and the three poor performers were eliminated in line with Evans et al. [24]. The description of the cumulus schemes and their strong influence in convective precipitation simulation is presented by Mayor & Mesquita [11],

The time–lagged ensemble members were derived by initializing WRF at 0000UTC, 0600UTC, 1200UTC and 1800UTC while perturbing the schemes was carried out by varying the entrainment rates (i.e. ±25% of the default rate). The study mainly varies cumulus parameterization schemes because of their significant effect on precipitation simulation [11]. All runs are done using the same initial conditions provided by NCEP/NCAR (Section, 2.1) and we used a total of 18 ensemble members.

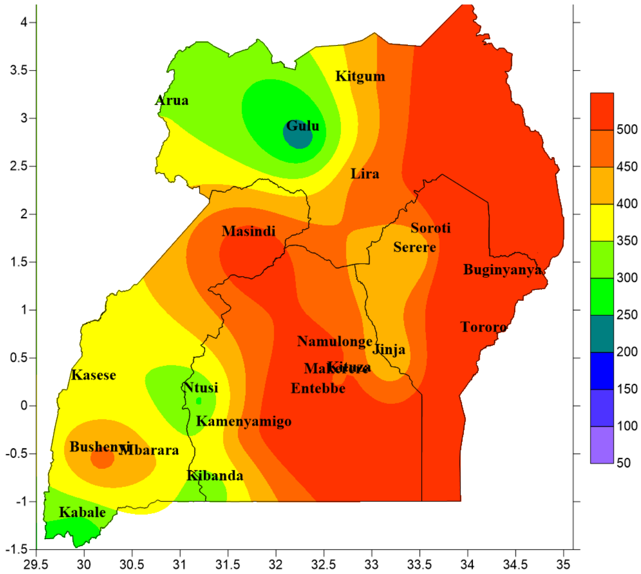


Figure 1. *The figure shows the total MAM 2013 seasonal rainfall over the Uganda.*

The study used three domains as shown in Figure (2):

- the first domain was at a horizontal resolution of 90 km. This domain covered Africa and was deemed sufficiently enough to cover the large scale synoptic systems such as the sub–tropical high pressure systems which are important for rainfall over equatorial region;
- the second domain was at a horizontal resolution of 30 km covering most parts of equatorial region to cater for influx of moisture over Uganda especially the Congo air mass and the moist currents from Mozambique channel;
- the third domain was at a horizontal resolution of 10 km covering Uganda, the study region.

The integration was done for the period 1^{st} March – 31^{st} May 2013 running the experiment for each day with 12 hours as spin–up period and using the same initial conditions for all the experiments. These experiments employed the staggered Arakawa C–grid; 30 vertical layers with model top fixed at 50 hPa; the terrain–following mass coordinate as vertical coordinate that had the capability of allowing variation of vertical grid–spacing and the Runge–Kutta 2^{nd} order integration option. The other physical schemes used are: the WRF Single–Moment 3–class microphyiscal scheme; the Rapid Radiative Transfer Model as the longwave radiation scheme; the Dudhia scheme as the shortwave radiation scheme; the Noah Land Surface Model and the Yonsei University scheme as the planetary boundary layer scheme. A 12 hour spin-up time was allowed to reduce spin–up errors.

2.4. Methods

2.4.1. The ensemble mean

The ensemble mean is the arithmetic average of the ensemble members. It outperforms climatological forecasts by 62% [23]. If we have n ensemble members such that,

$$\phi_1, \phi_2, \cdots, \phi_{n-1}, \phi_n$$

and for each member, i we get the prediction distribution over a total number of m events,

$$a_{1i}, a_{2i}, \cdots, a_{(m-1)i}, a_{mi}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 October 2017 doi:10.20944/preprints201710.0199.v1

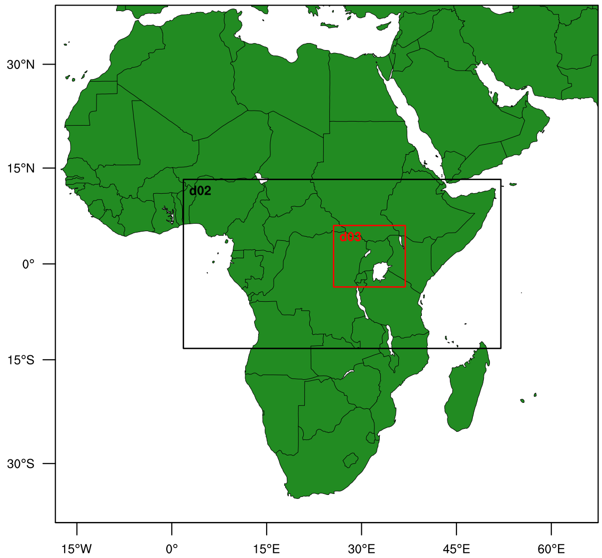


Figure 2. *The figure shows the domains used in running WRF model.*

which we summarize as

$$\begin{pmatrix} a_{11} & a_{21} & . & . & . & a_{m1} \\ a_{12} & a_{22} & . & . & . & a_{m2} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ a_{1n} & a_{2n} & . & . & . & a_{mn} \end{pmatrix}$$

where the rows represent the different ensemble members and the columns are the prediction events. The ensemble mean is thus the arithmetic average of the individual columns. Therefore, for an event t_k within the events space i.e. [1, m], the ensemble average, $\bar{a}_k$ is computed as (eqn. 1).

$$\bar{a}_k = \frac{1}{n}\sum_{i=1}^{n} a_{ki} \tag{1}$$

the ensemble mean, $\bar{a}_k$ is compared with the event, t_k.

2.4.2. The ensemble mean analogue

This *Ensemble Mean Analogue* (EMA) method is proposed by Delle–Monache et al.[35] and used by Vanvyve et al.[28] and Horvath et al.[36] in assessing the wind energy. They argued that it is key in developing wind farm projects due to its ability to provide associated uncertainty.

The EMA method involves obtaining the ensemble mean, $\bar{a}_k$ and looking in the training set consisting of analogues (i.e. $\{O^T : O_1^T, O_2^T, \cdots,\}$) to obtain a corresponding prediction [36]. If the ensemble average, $\bar{a}_k$ corresponds to observed analogue (i.e. O_k^T), this O_k^T is considered as the forecast, but if a corresponding analogue cannot be found, then $\bar{a}_k$ is used and the case is considered a rare event. In the study, the analogues were obtained from prediction of the MAM 2006 rainfall season. This period is considered an appropriate analogue season because it had the same Oceanic Niño Index of -0.2 [37,38] with MAM 2013 and the Inter–tropical Convergence Zone, one of the major rainfall drivers over the study area, is normally over the same region.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 October 2017 doi:10.20944/preprints201710.0199.v1

2.4.3. The multi–member analogue ensemble

In order to improve ensemble rainfall prediction, the **multi–member analog ensemble method** (MAEM) was used and is described here.

According to the MAEM, for a given member, its corresponding analogue is used to give the new ensemble member. The description of this method follows that: for a given prediction ϕ_i making part of the n member ensemble, i.e.

$$\phi_1, \phi_2, \cdots, \phi_i, \cdots, \phi_{n-1}, \phi_n$$

we look for its analogue and consider the observed value corresponding to ϕ_i. This observed value, Φ_i becomes the new ensemble member thus giving a new ensemble, i.e.

$$\Phi_1, \Phi_2, \cdots, \Phi_i, \cdots, \Phi_{n-1}, \Phi_n$$

If more than one analogue is found, an average of the analogues is used, i.e.

$$\overline{\Phi_i} = \frac{1}{M} \sum_{j=1}^{M} \Phi_{ij} \tag{2}$$

where M is the number of obtained analogues. If no corresponding analogue is found, model prediction is used as it could probably indicate a rare event. The prediction now becomes the arithmetic mean, $\overline{\Phi}_k$ of the observed analogues, Φ_k, i.e.

$$\overline{\Phi}_k = \frac{1}{n} \sum_{i=1}^{n} \Phi_{ik} \tag{3}$$

which was then compared with the rainfall event, t_k.

3. Results and discussion

3.1. Overview of the MAM seasonal rainfall totals over Uganda

The total rainfall obtained at the stations used in the study is illustrated in Figure (1). The total MAM 2013 rainfall amount over the entire study area was in the range of 200–900 mm. The stations in northern Uganda received comparatively lower seasonal rainfall amounts of 200–500 mm. This was expected because this region normally receives a unimodal rainfall distribution with rainfall onset around April/May peaking around July/August [2]. Over western Uganda, the rainfall amount was in the range of 270–550 mm while rainfall over Eastern Uganda varied from 400 mm to 900 mm. The Lake Victoria Basin received rainfall in the range of 400–650 mm. The MAM 2013 seasonal rainfall over Uganda thus exhibited large spatial and temporal variations.

 doi:10.20944/preprints201710.0199.v1

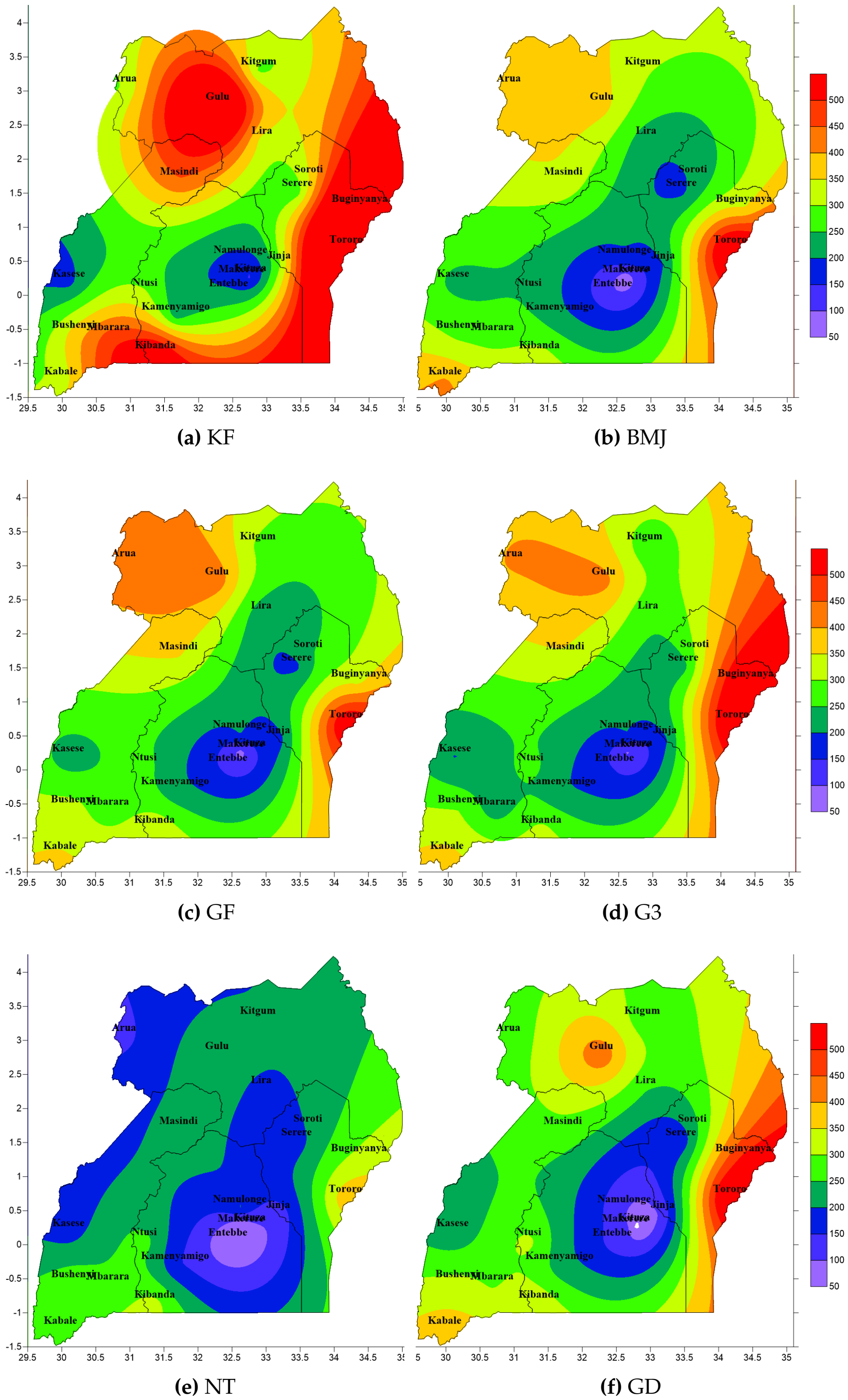


Figure 3. *The figure shows the amount of rainfall simulated by the cumulus schemes for the MAM 2013 rainfall period*

 doi:10.20944/preprints201710.0199.v1

3.2. Performance of the cumulus schemes

The performance of the cumulus schemes is presented using Tables (1 & 2). Table (1) shows the RMSE of the cumulus schemes along with rankings in performance while Table (2) shows the ME of the cumulus schemes and their respective ranking based on the ME. The schemes yielded varied results for different regions in Uganda as shown by Figure (3). For example, the **KF** scheme over estimated rainfall over northern Uganda (i.e. 300–750 mm) but also presented comparable rainfall amounts over Eastern Uganda (i.e. 400–900 mm). The **BMJ**, the **GD** and the **G3** underestimated rainfall amount over most parts of the country especially the eastern region (i.e. 300–600 mm). The **GF** captured rainfall amount over the northern region (i.e. 250–450 mm) but also underestimated rainfall over the Lake Victoria Basin (i.e. 150–300 mm) while the **NT** generally underestimated rainfall amount over most areas of Uganda.

Other studies such as Mayor & Mesquita[11] found the **GF** scheme to represent spatial precipitation distribution and the **BMJ** presenting higher precipitation amount. Related studies by Ratna et al. [39] found the **GD** scheme over–estimating convective rainfall. Our findings show that the **KF**, **GF** and **G3** had the comparatively smaller RMSE and better rankings based on RMSE (Table, 1) and better scores of ME (Table, 2). These schemes were thus selected to generate additional ensemble members.

Table 1. The RMSE of cumulus parametrization schemes.

	RMSE Scores						RMSE Rankings					
	KF	BMJ	GF	G3	NT	GD	KF	BMJ	GF	G3	NT	GD
Arua	12.19	12.57	14.31	14.13	24.18	14.51	1	2	4	3	6	5
Buginyanya	19.66	58.76	59.51	33.10	62.66	42.89	1	4	5	2	6	3
Bushenyi	14.70	21.84	19.25	20.51	21.70	19.03	1	6	3	4	5	2
Entebbe	45.35	55.99	53.03	51.93	61.20	53.22	1	5	4	2	6	3
Gulu	57.98	17.49	24.05	23.04	6.53	25.21	6	2	4	3	1	5
Jinja	14.35	21.62	23.33	22.13	29.31	29.62	1	2	4	3	5	6
Kibanda	31.40	8.21	8.35	7.62	7.86	7.77	6	4	5	1	3	2
Kabale	8.19	15.95	10.04	11.31	6.28	13.00	2	6	3	4	1	5
Kamenyamigo	27.94	29.32	28.67	29.81	34.57	28.28	1	4	3	5	6	2
Kasese	19.18	14.63	15.43	18.02	19.27	16.72	5	1	2	4	6	3
Kitgum	16.44	14.69	16.31	18.33	24.77	18.32	3	1	2	5	6	4
Kituza	40.00	43.14	40.80	42.91	46.13	50.84	1	4	2	3	5	6
Lira	20.42	30.50	31.22	26.30	33.28	26.95	1	4	5	2	6	3
Makerere	33.76	43.77	42.71	41.29	46.78	37.59	1	5	4	3	6	2
Mbarara	11.73	18.17	18.80	22.16	18.68	15.68	1	3	5	6	4	2
Masindi	19.59	27.75	26.33	26.83	38.51	32.24	1	4	2	3	6	5
Namulonge	36.39	35.90	34.72	35.17	34.87	44.21	5	3	1	4	2	6
Ntusi	8.85	9.33	8.81	8.20	8.53	8.70	5	6	4	1	2	3
Serere	14.94	23.23	23.12	19.17	24.27	25.51	1	4	3	2	5	6
Soroti	16.08	27.62	25.25	18.84	28.24	25.93	1	5	3	2	6	4
Tororo	34.11	16.34	18.72	14.72	34.17	15.37	5	3	4	1	6	2
Average	**23.96**	26.04	**25.85**	**24.07**	29.13	26.27	**2.38**	3.71	**3.43**	**3.00**	4.71	3.76

3.3. The performance of ensemble mean

The comparison of the RMSE of the ensemble mean to the individual cumulus parameterization schemes is presented in Figure (4) while comparison of ME of the ensemble mean to individual cumulus parameterization schemes presented in Figure (5). The RMSE for ensemble mean is presented in Table (3, 2nd column) while the results for bias (or ME) presented in Table (4, 2nd column). This study is at 10 Km horizontal resolution, slightly higher than the 18 Km resolution of the ensemble of European Center for Medium–range Weather Forecasts (ECMWF) which could interest ECMWF to investigate potential improvement in its prediction skill using higher resolution.

Table 2. The RMSE of cumulus parametrization schemes.

	ME Scores						ME Rankings					
	KF	BMJ	GF	G3	NT	GD	KF	BMJ	GF	G3	NT	GD
Arua	-0.89	0.91	1.83	1.20	-4.49	-1.97	1	2	4	3	6	5
Buginyanya	-2.08	-11.77	-11.93	-6.00	-12.54	-8.31	1	4	5	2	6	3
Bushenyi	-2.19	-4.03	-3.22	-3.78	-3.99	-3.34	1	6	2	4	5	3
Entebbe	-9.06	-11.36	-10.73	-10.50	-12.44	-10.77	1	5	3	2	6	4
Gulu	11.76	3.39	4.76	4.50	0.64	4.93	6	2	4	3	1	5
Jinja	-2.15	-4.14	-4.58	-4.31	-5.89	-5.93	1	2	4	3	5	6
Kibanda	6.22	-0.04	0.53	-0.18	-0.08	0.34	6	1	5	3	2	4
Kabale	0.83	2.75	1.65	1.90	0.31	2.16	2	6	3	4	1	5
Kamenyamigo	-5.31	-5.58	-5.50	-5.76	-6.78	-5.46	1	4	3	5	6	2
Kasese	-3.55	-2.39	-2.68	-3.31	-3.76	-2.98	5	1	2	4	6	3
Kitgum	-2.90	-2.54	-2.93	-3.38	-4.80	-3.36	2	1	3	5	6	4
Kituza	-7.94	-8.57	-8.08	-8.53	-9.23	-10.24	1	4	2	3	5	6
Lira	-3.43	-5.85	-6.00	-4.94	-6.51	-5.11	1	4	5	2	6	3
Makerere	-6.74	-8.75	-8.51	-8.22	-9.39	-7.39	1	5	4	3	6	2
Mbarara	1.30	-2.96	-3.17	-4.02	-3.29	-2.51	1	3	4	6	5	2
Masindi	-2.34	-4.63	-4.29	-4.50	-7.26	-5.78	1	4	2	3	6	5
Namulonge	-7.22	-7.06	-6.82	-6.89	-6.83	-8.81	5	4	1	3	2	6
Ntusi	0.32	-0.98	-0.54	-0.55	-0.48	0.47	1	6	4	5	3	2
Serere	-2.61	-4.53	-4.51	-3.66	-4.78	-5.04	1	4	3	2	5	6
Soroti	-2.34	-5.30	-4.63	-3.35	-5.41	-4.98	1	5	3	2	6	4
Tororo	6.30	-1.40	-2.51	-1.80	-6.58	-1.49	5	1	4	3	6	2
Average	**-1.62**	-4.04	**-3.90**	**-3.62**	-5.41	-4.07	**2.14**	3.52	**3.33**	**3.33**	4.76	3.90

The performance of the three best cumulus schemes (i.e. **KF, GF** & **G3**) and the ensemble mean (EM) across the regions in the country is illustrated in Figure (4). Spatial analysis of Figure (4) shows that, the RMSE of the ensemble mean was slightly higher for the western and the eastern regions of Uganda (i.e. RMSE:9–15 & 9–16) respectively. The 'cattle corridor' of Uganda and the Lake Victoria Basin had RMSE of 8–10 while the northern region had 8–11.

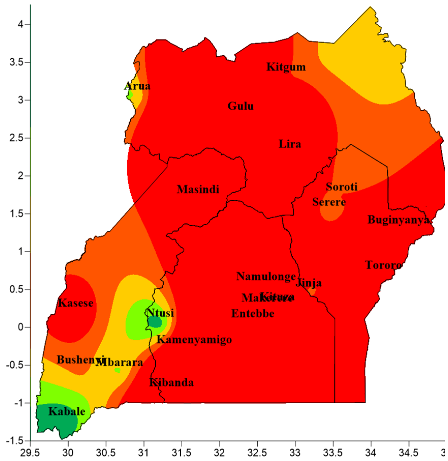
The ensemble mean RMSE results are comparatively smaller than the RMSE results of individual cumulus parameterization schemes. Additional statistical analysis using paired t–test at 99% confidence level also showed improvement in the RMSE of the ensemble mean compared to individual cumulus parameterization schemes (i.e. KF: t=4.73 & $p < 0.001$; GF: t=5.14 & $p < 0.001$ and G3: t=5.41 & $p < 0.001$). The ensemble mean did not show a significant improvement in the ME because, apart from the northern Uganda, we still observed a negative bias over most parts of the country as shown by Figure 5(a–c). So relying on the RMSE, we can argue that there is improvement in performance when using the ensemble mean but not necessarily changes in bias.

3.4. The performance of ensemble mean analogue

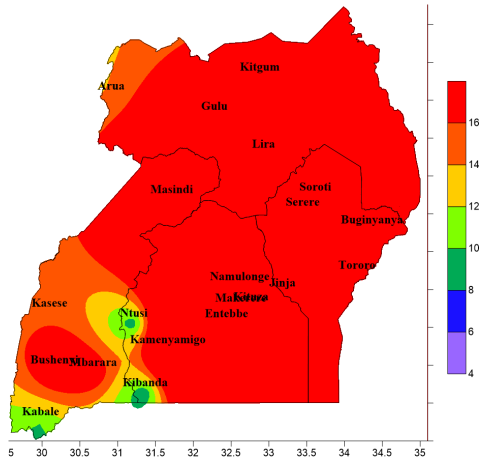
The RMSE results on the performance of the ensemble mean analogue (EMA) are presented in Table (3, 3^{rd} column) while the ME results are presented in Table (4, 3^{rd} column). Results show that although the improvement in simulation as shown by reduction in RMSE was not significant at 95%, 13 out of 21 stations (RMSE shown in italics in Table, 3, the 3^{rd} column) had their RMSE less than RMSE of ensemble mean. A slight improvement in negative bias is observed (t = 1.710; p_value = 0.096) and additional analysis shows that the magnitude of bias of 16 out of 21 stations (Table, 4 with bold values, the 3^{rd} column) was less than a magnitude of 2.00. The results of RMSE and ME confirm that the EMA can present a modest improvement in rainfall prediction.

3.5. The performance of multi–member analogue ensemble method

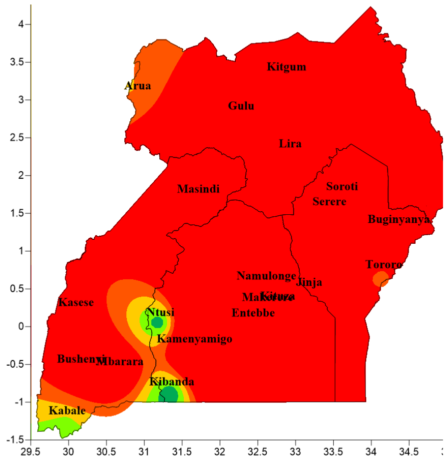
The RMSE results for multi–member analogue ensemble are also presented in Table (3, the 4^{th} column) while the ME results are presented in Table (4, the 4^{th} column). We again observed that



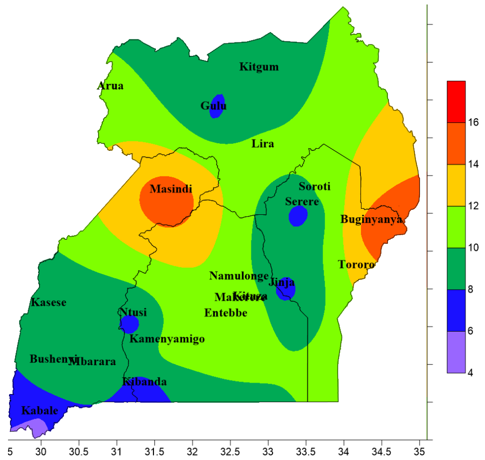
(a) KF



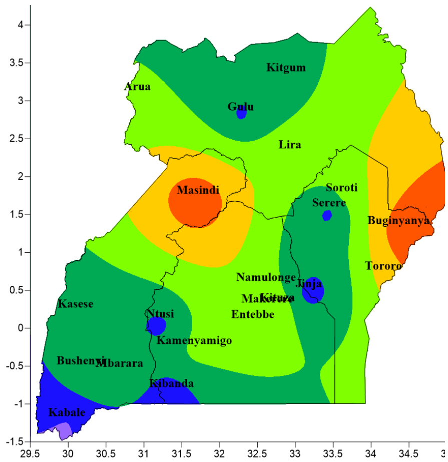
(b) BMJ



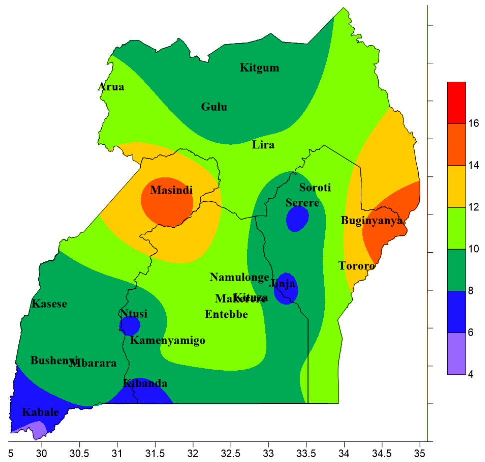
(c) GF



(d) Ensemble Mean (ENS)



(e) Ensemble Mean Analogue (EMA)

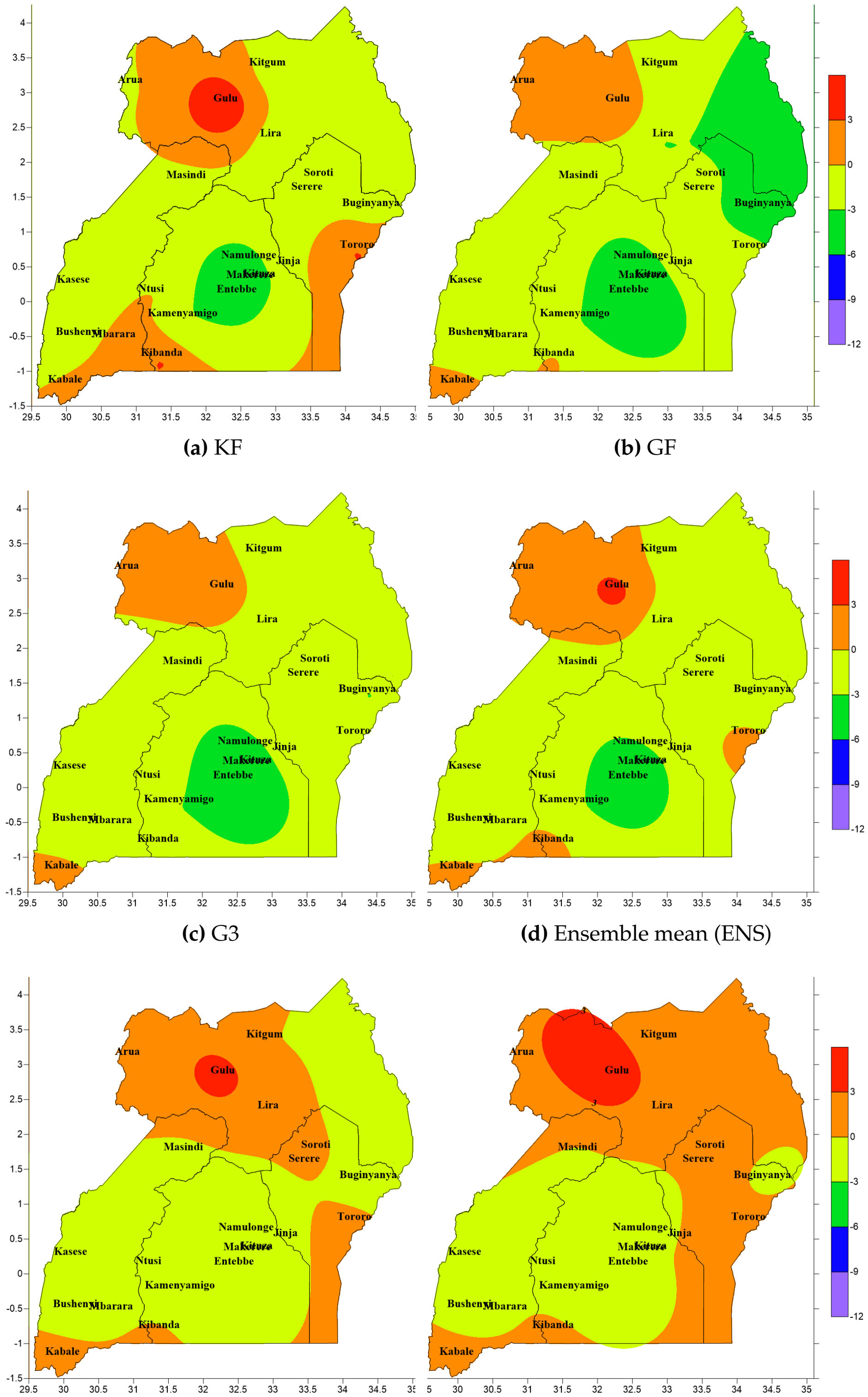


(f) Multi–analogue ensemble mean (MAEM)

.

Figure 4. *The figure shows the RMSE of* **KF***(a),* **GF***(b),* **G3***(c) cumulus parameterization schemes and the RMSE of the ensemble* **ENS***(d) ,* **EMA***(e) &* **MAEM***(e). The schemes* **KF***,* **GF** *&* **G3** *were found to have comparatively smaller RMSE than the other cumulus schemes*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 October 2017 doi:10.20944/preprints201710.0199.v1



(a) KF

(b) GF

(c) G3

(d) Ensemble mean (ENS)

(e) Ensemble Mean Analogue (EMA)

(f) Multi–analogue ensemble mean (MAEM)

Figure 5. *The figure shows the ME of* **KF***(a),* **GF***(b),* **G3***(c) cumulus parameterization schemes and the ME of the ensemble* **ENS***(d),* **EMA***(e) &* **MAEM***(e)*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 October 2017 doi:10.20944/preprints201710.0199.v1

Table 3. *Shows the RMSE for: ensemble mean (ENS); ensemble mean analogue (EMA) and multi–member analogue ensemble (MAEM). The italicized values show improvement in the RMSE of the EMA and the RMSE of the MAEM compared to the RMSE of the ENS. The bold values in the* 4^{th} *column show a reduction in RMSE using MAEM compared to the RMSE of EMA*

Station	**RMSE**		
	ENS	**EMA**	**MAEM**
Arua	10.93	*10.70*	*10.85*
Buginyanya	15.23	15.36	***15.02***
Bushenyi	9.56	*9.33*	*9.38*
Entebbe	11.51	*11.28*	***10.98***
Gulu	7.78	*7.76*	8.04
Jinja	7.23	*7.07*	***6.99***
Kabale	5.81	5.92	5.87
Kamenyamigo	10.84	*10.67*	*10.70*
Kasese	8.37	*8.21*	*8.28*
Kibanda	6.91	*6.89*	***6.90***
Kitgum	8.34	8.96	***8.32***
Kituza	11.71	*11.65*	***11.25***
Lira	10.97	11.8	***10.76***
Makerere	11.02	*10.83*	***10.63***
Masindi	15.84	*15.82*	***15.80***
Mbarara	10.01	*9.99*	*10.00*
Namulonge	11.20	*11.11*	***10.80***
Ntusi	7.43	7.45	***7.40***
Serere	7.30	7.62	***7.18***
Soroti	9.83	10.16	**9.84**
Tororo	12.65	12.82	12.99

although the improvement in simulation as shown by reduction in RMSE was not significant at 95%, 17 out of 21 stations (RMSE shown in italics in Table, 3, the 4^{th} column) had their RMSE less than RMSE of ensemble mean. We further noted a slight improvement in RMSE, though not significant at 95% by employing multi–member analogue ensemble compared to ensemble mean analogue with 13 out of 21 stations (i.e. bold values in Table (3), the 4^{th} column) having a smaller RMSE compared to the RMSE for ensemble mean analogue.

The results further show a significant improvement in negative bias at 95% confidence level (t=2.5285; p_value=0.016). Additional analysis shows that, the magnitude of bias of 18 out of 21 stations (Table, 4 with bold values, the 4^{th} column) was less than a magnitude of 2.00. The results of RMSE and ME confirm that the multi–member analogue ensemble method can improve rainfall prediction.

4. Summary and conclusion

The study investigated the potential for improving rainfall prediction using ensemble methods and considered 18 ensemble members. The study first analyzed the spatial distribution of the MAM 2013 rainfall and found that it was in the range of 200–900 mm. It was also noted that the MAM 2013 rainfall exhibited large spatial and temporal variations over the study region.

The study then assessed the performance of the cumulus schemes and found varying performance over different regions of Uganda with the **KF** scheme over estimating rainfall over northern Uganda; the **BMJ**, the **GD** and the **G3** underestimating rainfall amount over most parts of the country especially the eastern region; the **GF** capturing rainfall amount over the northern region while the **NT** generally underestimating rainfall amount over most areas.

The study further assessed the performance of ensemble mean, ensemble mean analogue and the multi–member analogue ensemble. An improvement in the RMSE was observed while using the ensemble mean compared to using individual cumulus parametrization schemes. There was a non significant change in the ME of the ensemble mean compared to individual parameterization schemes. The ensemble mean analogue presented a reduction in the magnitude of the RMSE and a

 doi:10.20944/preprints201710.0199.v1

Table 4. *Shows the ME (or Bias) for: ensemble mean (ENS); ensemble mean analogue (EMA) and multi–member analogue ensemble mean (MAEM). The bold values indicate a magnitude of ME less than 2.00 for all the three methods.*

Station	Mean error (or Bias)		
	ENS	EMA	MAEM
Arua	**0.42**	**1.45**	2.61
Buginyanya	-2.92	-2.02	**-0.54**
Bushenyi	**-1.62**	**-1.04**	**-1.06**
Entebbe	-5.20	**-1.18**	**-1.94**
Gulu	4.08	4.26	5.13
Jinja	**-1.60**	**-0.78**	**0.08**
Kabale	**0.57**	**1.00**	**1.25**
Kamenyamigo	-2.59	-2.35	-2.19
Kasese	**-1.49**	**-1.04**	**-0.76**
Kibanda	**0.89**	**0.56**	**1.15**
Kitgum	**-1.02**	**0.15**	**0.94**
Kituza	-4.01	**-1.52**	**-0.43**
Lira	**-1.96**	**0.61**	**0.63**
Makerere	-3.92	-2.50	**-1.39**
Masindi	**-1.49**	**-1.03**	**-0.54**
Mbarara	**-0.69**	**-0.71**	**-0.63**
Namulonge	-3.27	-2.13	**-1.63**
Ntusi	**0.03**	**-0.03**	**0.05**
Serere	**-1.15**	**0.17**	**0.34**
Soroti	**-0.94**	**0.52**	**1.39**
Tororo	**1.08**	**1.32**	**1.89**

slight improvement in the ME where 16 out of 21 stations had their magnitude of ME less than 2.00. The multi–member analogue ensemble presented a further reduction in the RMSE and additional improvement in the magnitude of ME. We thus note that although the ensemble mean improves the prediction accuracy, the ensemble mean analogue and the multi–member analogue ensemble presents additional improvement in accuracy with multi–member analogue ensemble giving the best results.

Acknowledgments

The authors are grateful to the project of "Partnership for Building Resilient Ecosystems and Livelihoods to Climate Change and Disaster Risks" (BREAD project: SIDA/331) under the "Swedish International Development cooperation Agency" (SIDA) for the financial support; to the project ”Improving Weather Information Management in East Africa for effective service provision through the application of suitable ICTs” (WIMEA–ICT project: UGA–13/0018) under the Norad’s Programme for Capacity Development in Higher Education and Research for Development (NORHED) for the technical support and to the reviewers for the constructive feedback. We are also grateful to Uganda National Meteorological Authority (https://www.unma.go.ug) for availing the rainfall data used in the study.

Appendix: Performance measures

The study employed two performance measures i.e. the root mean square error (RMSE) and the mean error (ME).

Appendix Root mean square error

The RMSE is obtained from the square root of the mean square differences between predicted (i.e. P) and observed (i.e. O) when paired. It is computed mathematically as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[P_i - O_i]^2} \tag{4}$$

Appendix Mean error

The ME is the mean of the differences (i.e $P_i - O_i$) which is computed as:

$$ME = \frac{1}{n}\sum_{i=1}^{n}[P_i - O_i] \quad (5)$$

where i is the i^{th} data point ordered in time.

References

1. S. Tao, S. Shen, Y. Li, Q. Wang, P. Gao, I. Mugume, Projected crop production under regional climate change using scenario data and modeling: sensitivity to chosen sowing date and cultivar, Sustainability 8 (3) (2016) 214.
2. B. A. Ogwang, H. Chen, X. Li, C. Gao, The influence of topography on east african october to december climate: sensitivity experiments with regcm4, Advances in Meteorology 2014.
3. S. W. Karuri, R. W. Snow, Forecasting paediatric malaria admissions on the kenya coast using rainfall, Global health action 9.
4. A. T. Kabo-Bah, C. J. Diji, K. Nokoe, Y. Mulugetta, D. Obeng-Ofori, K. Akpoti, Multiyear rainfall and temperature trends in the volta river basin and their potential impact on hydropower generation in ghana, Climate 4 (4) (2016) 49.
5. S. He, S. V. Raghavan, N. S. Nguyen, S.-Y. Liong, Ensemble rainfall forecasting with numerical weather prediction and radar-based nowcasting models, Hydrological Processes 27 (11) (2013) 1560–1571.
6. D. Ntwali, B. A. Ogwang, V. Ongoma, The impacts of topography on spatial and temporal rainfall distribution over rwanda based on wrf model, Atmospheric and Climate Sciences 6 (02) (2016) 145.
7. J. Awange, R. Anyah, N. Agola, E. Forootan, P. Omondi, Potential impacts of climate and environmental change on the stored water of lake victoria basin and economic implications, Water Resources Research 49 (12) (2013) 8160–8173.
8. I. Mugume, M. D. Mesquita, C. Basalirwa, Y. Bamutaze, J. Reuder, A. Nimusiima, D. Waiswa, G. Mujuni, S. Tao, T. Jacob Ngailo, Patterns of dekadal rainfall variation over a selected region in lake victoria basin, uganda, Atmosphere 7 (11) (2016) 150.
9. T. Ngailo, N. Shaban, J. Reuder, E. Rutalebwa, I. Mugume, Non homogeneous poisson process modelling of seasonal extreme rainfall events in tanzania, International Journal of Science and Research (IJSR) 5 (10) (2016) 1858–1868.
10. W. Jie, T. Wu, J. Wang, W. Li, T. Polivka, Using a deterministic time-lagged ensemble forecast with a probabilistic threshold for improving 6–15day summer precipitation prediction in china, Atmospheric Research 156 (2015) 142–159.
11. Y. G. Mayor, M. D. Mesquita, Numerical simulations of the 1 may 2012 deep convection event over cuba: sensitivity to cumulus and microphysical schemes in a high-resolution model, Advances in Meteorology 2015.
12. F. Maussion, D. Scherer, R. Finkelnburg, J. Richters, W. Yang, T. Yao, Wrf simulation of a precipitation event over the tibetan plateau, china–an assessment using remote sensing and ground observations, Hydrol. Earth Syst. Sci 15 (2011) 1795–1817.
13. J. Coiffier, Fundamentals of numerical weather prediction, Cambridge University Press, 2011.
14. ECMWF (2017). What is ensemble weather forecasting? https://www.ecmwf.int/en/about/media--centre/fact--sheet--ensemble--weather--forecasting, accessed on 24^{th} October 2017.
15. Glahn, H. R., & Lowry, D. A. (1972). The use of model output statistics (MOS) in objective weather forecasting. Journal of applied meteorology, 11(8), 1203-1211.
16. Scheuerer, M. (2014). Probabilistic quantitative precipitation forecasting using ensemble model output statistics. Quarterly Journal of the Royal Meteorological Society, 140(680), 1086-1096.
17. C. Fraley, A. E. Raftery, T. Gneiting, J. Sloughter, V. J. Berrocal, Probabilistic weather forecasting in R, The R Journal 3 (1) (2011) 55–63.

18. I. Mugume, C. Basalirwa, D. Waiswa, J. Reuder, M. d. S. Mesquita, S. Tao, T. J. Ngailo, Comparison of parametric and nonparametric methods for analyzing the bias of a numerical model, Modelling and Simulation in Engineering 2016.
19. T. Gneiting, A. E. Raftery, Weather forecasting with ensemble methods, Science 310 (5746) (2005) 248–249.
20. S. Hemri, M. Scheuerer, F. Pappenberger, K. Bogner, T. Haiden, Trends in the predictive performance of raw ensemble weather forecasts, Geophysical Research Letters 41 (24) (2014) 9197–9205.
21. J. S. Whitaker, A. F. Loughe, The relationship between ensemble spread and ensemble mean skill, Monthly weather review 126 (12) (1998) 3292–3302.
22. A. E. Raftery, T. Gneiting, F. Balabdaoui, M. Polakowski, Using bayesian model averaging to calibrate forecast ensembles, Monthly Weather Review 133 (5) (2005) 1155–1174.
23. Z. T. Segele, M. B. Richman, L. M. Leslie, P. J. Lamb, Seasonal-to-interannual variability of ethiopia/horn of africa monsoon. part ii: Statistical multimodel ensemble rainfall predictions, Journal of Climate 28 (9) (2015) 3511–3536.
24. J. P. Evans, F. Ji, G. Abramowitz, M. Ekström, Optimally choosing small ensemble members to produce robust climate simulations, Environmental Research Letters 8 (4) (2013) 044050.
25. J. Zhu, F. Kong, L. Ran, H. Lei, Bayesian model averaging with stratified sampling for probabilistic quantitative precipitation forecasting in northern china during summer 2010, Monthly Weather Review 143 (9) (2015) 3628–3641.
26. G. Redmond, K. I. Hodges, C. Mcsweeney, R. Jones, D. Hein, Projected changes in tropical cyclones over vietnam and the south china sea using a 25 km regional climate model perturbed physics ensemble, Climate Dynamics 45 (7-8) (2015) 1983–2000.
27. J. M. Fritsch, R. Carbone, Improving quantitative precipitation forecasts in the warm season: A uswrp research and development strategy, Bulletin of the American Meteorological Society 85 (7) (2004) 955–965.
28. E. Vanvyve, L. Delle Monache, A. J. Monaghan, J. O. Pinto, Wind resource estimates with an analog ensemble approach, Renewable Energy 74 (2015) 761–773.
29. E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, et al., The ncep/ncar 40-year reanalysis project, Bulletin of the American meteorological Society 77 (3) (1996) 437–471.
30. D. H. White, G. A. Lubulwa, K. Menz, H. Zuo, W. Wint, J. Slingenbergh, Agro-climatic classification systems for estimating the global distribution of livestock numbers and commodities, Environment international 27 (2) (2001) 181–187.
31. C. Basalirwa, Delineation of uganda into climatological rainfall zones using the method of principal component analysis, International Journal of climatology 15 (10) (1995) 1161–1177.
32. C. Funk, A. Hoell, S. Shukla, G. Husak, J. Michaelsen, The east african monsoon system: Seasonal climatologies and recent variations, in: The Monsoons and Climate Change, Springer, 2016, pp. 163–185.
33. W. Yang, R. Seager, M. A. Cane, B. Lyon, The annual cycle of East African precipitation, Journal of Climate 28 (6) (2015) 2385–2404.
34. R. Pizarro, P. Garcia-Chevesich, R. Valdes, F. Dominguez, F. Hossain, P. Ffolliott, C. Olivares, C. Morales, F. Balocchi, P. Bro, Inland water bodies in chile can locally increase rainfall intensity, Journal of hydrology 481 (2013) 56–63.
35. L. Delle Monache, F. A. Eckel, D. L. Rife, B. Nagarajan, K. Searight, Probabilistic weather prediction with an analog ensemble, Monthly Weather Review 141 (10) (2013) 3498–3516.
36. K. Horvath, A. Bajić, S. Ivatek-Šahdan, M. Hrastinski, I. Odak Plenković, A. Stanešić, M. Tudor, T. Kovačić, Overview of meteorological research on the project "weather intelligence for wind energy"-will4wind, Hrvatski meteorološki časopis 50 (50) (2016) 91–104.
37. CPC, Cold and warm episodes by season (2017). http://www.cpc.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml,accessedon1stFeb.2017
38. Y. Hafez, Study on the relationship between the oceanic nino index and surface air temperature and precipitation rate over the kingdom of saudi arabia, Journal of Geoscience and Environment Protection 4 (05) (2016) 146.

doi:10.20944/preprints201710.0199.v1

39. S. B. Ratna, J. Ratnam, S. Behera, T. Ndarana, K. Takahashi, T. Yamagata, et al., Performance assessment of three convective parameterization schemes in wrf for downscaling summer rainfall over south africa, Climate dynamics 42 (11-12) (2014) 2931–2953.