The revised version of the manuscript by Zhang et al. definitively improved from a methodological point of view. The predictor selection is now based on a procedure, which automatically identifies pixels of large scale climate variables, which significantly correlate with summer precipitation anomalies. In a next step, the pixel anomalies in specific regions are averaged and different variables are combined by means of a pca analysis. PCA timeseries (explaining 95% of the overall variance), are eventually utilized as predictors for a linear regression based prediction. The procedure is now fully cross-validated which results in poor to moderate prediction results.

While the application of the method is solid (beside of some smaller issues, see specific remarks), the results of the statistical forecast model are worse compared with some dynamical models (and sometimes even worse compared with climatology). Beside of two clusters all correlations between observations and cross-validated forecast are below 0.2, which is not statistically significant for a period of 30 years. Thus in my opinion the study is not yet ready for publication in HESS. However, I see quite a big potential of the study from a hydro-climatological point of view. The major question for me is, why there is such a big difference in the skill of the forecast models, although the clusters are very close to each other. This could either be a statistical artefact or it could be explained by large/regional scale climate mechanisms. I believe that the results are robust and physically interpretable (all clusters with moderate forecast skill are windward of the East-African monsoon), but this should be somehow investigated. In the revised manuscript the authors give some more information on the general climate characteristics and also mention, that different cluster regions are correlated with different climate modes and sst patterns. I still feel that this insufficient for the interpretation of the results. I suggest to give a short literature overview of climatic mechanisms in the introduction. Which anomalies (pressure patterns, moisture fluxes etc) trigger precipitation surplus and deficits? How are these patterns related to ENSO and other potential predictors? There is quite a lot literature on the monsoon in general.

What kind of predictors are important for which cluster? Maybe this could eventually explain the different model skills and support the robustness of your models. Further it would definitively justify the clustering procedure. My impression from the gridded analysis is that it doesn't make a difference?

We truly appreciate the reviewer's help in improving this work. As the reviewer suggested, we added a more detailed literature review of climatic mechanisms affecting JJAS precipitation in Ethiopia at Page 5 Line 21:

"Large-scale climate variables are often evaluated as potential predictors in statistical seasonal precipitation prediction models, commonly including sea surface temperatures (SST) in the equatorial Pacific Ocean representing the well-known of the El Nino-Southern Oscillation (ENSO) (Stone et al., 1996). Sea level pressure (SLP) in the eastern Pacific Ocean at Tahiti as an critical and stable component for measuring an ENSO index (Torrence and Webster, 1999) warrants another potential predictor. For Ethiopia, the ENSO phenomenon is considered a primary indicator of precipitation variability, particularly in the main JJAS rainy season with El Nino/La Nina often associated with deficit/excess of precipitation amount in the study region (e.g. NMSA, 1996, Camberlin, 1997, Bekele, 1997, Segele and Lamb, 2005, Diro et al., 2011, Elagib and Elhag, 2011). Evidences have also shown a more direct moisture transport from the Gulf of Guinea (equatorial Atlantic Ocean), the Indian Ocean, and the Mediterranean Sea affecting Ethiopia's summer precipitation (Viste and Sorteberg, 2013a, Viste and Sorteberg, 2013b). Those moisture fluxes are often related to pressure patterns across the continent. For instance, the St. Helena high over the southern Atlantic Ocean or a high pressure over Gulf of Guinea, and a simultaneous low pressure over

Indian Ocean or a monsoon trough over Arabic Peninsula bring intensified westerlies and south-westerlies that transport moist air across the Congo Basin to the western Ethiopian highlands in the summer (Segele et al., 2009, Williams et al., 2011). Similarly, the southwest Asian monsoon at the Indian Ocean, which has a strong positive relationship with the concurrent JJAS rainfall in the western Ethiopia, is associated with the Mascarene high over the southern Indian Ocean and a low pressure system near Bombay. During this monsoon season, the southeast trades in the southern hemisphere are channeled by the east African highlands while crossing the equator and become a southwest monsoon flow. It is further diverted by the Turkana Channel, enhancing convergence with the westerlies/south-westerlies above the western Ethiopian highlands and bringing moisture to this region (Kinuthia, 1992, Nicholson, 1996, Camberlin, 1997, Slingo et al., 2005, Segele et al., 2009, Nicholson, 2014). In addition, the effect of other climate variables relevant to the aforementioned driven factors, such as the Indian Ocean SST, local and other regional atmospheric pressure systems such as Azores High also have notable influence on Ethiopia's precipitation variability (e.g. Kassahun, 1987, Tadesse, 1994, NMSA, 1996, Shanko and Camberlin, 1998, Goddard and Graham, 1999, Latif et al., 1999, Black et al., 2003, Segele and Lamb, 2005).

Consequently, season-ahead (March-May) or month-ahead (May) large-scale climate variables that are physically relevant in potentially modulating moisture transport to the basin (or cluster) are selected as potential predictors. Four climate variables are selected here for further evaluation based on outcomes of the aforementioned prediction studies: SST, SLP, geopotential height (GH) at 500mb, and surface air temperature (SAT). All climate variables are from the National Centers for Environmental Prediction and National Center for Atmospheric Research (NCEP/NCAR) reanalysis dataset (Kalnay et al., 1996) at a 2.5˚×2.5˚ grid scale. "

Regarding to the reasons why the prediction skills are different across the clusters, it is likely due to the complex climate mechanisms affecting this region, causing great spatial heterogeneity, although we cannot rule out the possibility that the linear structure of the model cannot capture some non-linear relationships, if any, between the climate predictors and JJAS precipitation in some clusters. As the reviewer suspects, the differences in skills are possibly related to the East-African monsoon (the southwest Asian monsoon over Indian Ocean) given that the clusters with moderate prediction skills lie on the windward pathway of the East-African monsoon. We inspected the concurrent correlation maps with sea level pressure (SLP) for each cluster and found that Cluster 5 and 7 – the two clusters with the best skills – are the only ones that are strongly and negatively corrected with SLP near Bombay (indicating a strong East-African monsoon is associated with higher JJAS precipitation amount, and vice versa), and meantime strongly and positively correlated with the SLP at eastern equatorial Pacific Ocean (indicating high surface pressure often accompanied with cold sea surface temperature and a raining pattern over the eastern equatorial Pacific Ocean – a La Nina phenomenon – somehow brings higher JJAS precipitation in western Ethiopia, and vice versa, i.e. El Nino associated with lower precipitation). Cluster 2 – one of the worst predicted clusters – shows moderately strong negative correlation with SLP near Bombay; however, it is also correlated strongly and negatively with SLP at southern Indian Ocean (indicating a possible weak gradient of East-African monsoon) and its correlation with SLP over equatorial Pacific Ocean is nonsignificant. Considering in general El Nino suppresses the monsoon and La Nina increases it (Kumar et al., 2006), the connections with both ENSO and monsoon in the right direction such as for Cluster 5 and 7, indicate a double insurance over their connection with the East-African monsoon. Therefore, we conclude

that clusters which are more affected by East-African monsoon, particularly coupled with the influence of ENSO, show more promises in prediction skills.

Additional discussions are provided at Page 16 Line 22:

"Relatively poor prediction performance is evident in some locations such as southwestern Ethiopia and regions outside Ethiopia, where the hydroclimatic processes that produce precipitation might be driven by local factors or other regional climate patterns rather than large-scale climate variables identified in this study. A previous study (Zhang et al., 2016) has shown that the influence of ENSO on JJAS precipitation in western Ethiopia decreases generally from north to south, and is likely one of the reasons why skills are relatively low in southwestern Ethiopia. Cluster 5 was also identified with the strongest connection to equatorial Pacific SST (Zhang et al., 2016), which is consistent with the highest skill found in this study. Other regions with low prediction skill show relatively strong connections to SST in neighboring oceanic regions. However, connections with those climate patterns appear to be less robust than with ENSO, making the predictions in their associated regions less skillful. This is also consistent with the findings from other studies that even though all three oceans (Indian, Atlantic, and Pacific Ocean) affect the JJAS precipitation in western Ethiopia, the Pacific Ocean still plays the greatest role (Segele et al., 2009, Omondi et al., 2013).

The southwest Asian monsoon over Indian Ocean may also be critical in determining the precipitation, given that the clusters with better prediction skills lie along the pathway of the monsoon. Based on the global concurrent correlation maps between JJAS precipitation and SLP for each cluster, Cluster 5 and 7 – the two clusters with the best skills – are the only ones that are strongly and negatively corrected with SLP near Bombay, and meantime strongly and positively correlated with the SLP at the eastern equatorial Pacific Ocean. The former indicates that a strong southwest Asian monsoon is associated with higher JJAS precipitation amount, and vice versa. The latter indicates that a high surface pressure over the eastern equatorial Pacific Ocean often accompanied with cold SST and a raining pattern – a La Nina phenomenon – also brings higher JJAS precipitation to the western Ethiopia, and vice versa. Cluster 2 – one of the worst predicted clusters – shows moderately strong negative correlation with SLP near Bombay; however, it is also correlated strongly and negatively with SLP at southern Indian Ocean, indicating a possible weak gradient of the southwest Asian monsoon. Moreover, its correlation with SLP over equatorial Pacific Ocean is nonsignificant. Considering in general El Nino suppresses the monsoon and La Nina increases it (Kumar et al., 2006), strong correlations with both ENSO and the monsoon in the correct direction, such as for Cluster 5 and 7, indicate a double insurance over their association with the southwest Asian monsoon. Therefore, clusters which are more affected by the southwest Asian monsoon over Indian Ocean, particularly coupled with the influence of ENSO, are likely to show more promises in their prediction skills."

We admit that except two clusters all correlations between observation and the cross-validated forecast are below 0.2; however, we want to point out that Cluster 5 and 7, which show correlations of 0.51 and 0.35 respectively, occupy a large portion of area inside Ethiopia, which is also the agricultural-richest region. Besides, we can also see positive RPSS values for Cluster 1, 3, and 4, although their correlations are below 0.2. As mentioned in the discussion section, for most clusters inside Ethiopia, the predictions in this study are superior to the current NMA operational forecast at regional scale. As observational

datasets continue to grow, this data-driven cluster analysis and statistical modeling approach may also be expected to improve.

Regarding to the gridded analysis, we want to emphasize that the grid-level prediction would be useful for operational purpose with more accurate spatial representation. While comparing to bias-corrected grid-level predictions based on dynamical models, our results show higher skills over some critical region that has a great deal of agricultural activities, such as Cluster 5. Gridded analysis also enables us to compare across *indirect* and *direct* method: whether cluster-level prediction can be interpreted properly to grid-level prediction helps to justify the clustering analysis – whether the clusters are robust and coherent. More responses regarding to gridded analysis are provided given the reviewer's minor remarks below.

Minor remarks:

- Fig. 1: Please show boarders and rivers to make orientation easier. The seasonal precipitation amounts might be rather shown as a barplot.

We appreciate the reviewer's suggestion and included an updated Fig. 1 in the manuscript (also included here):

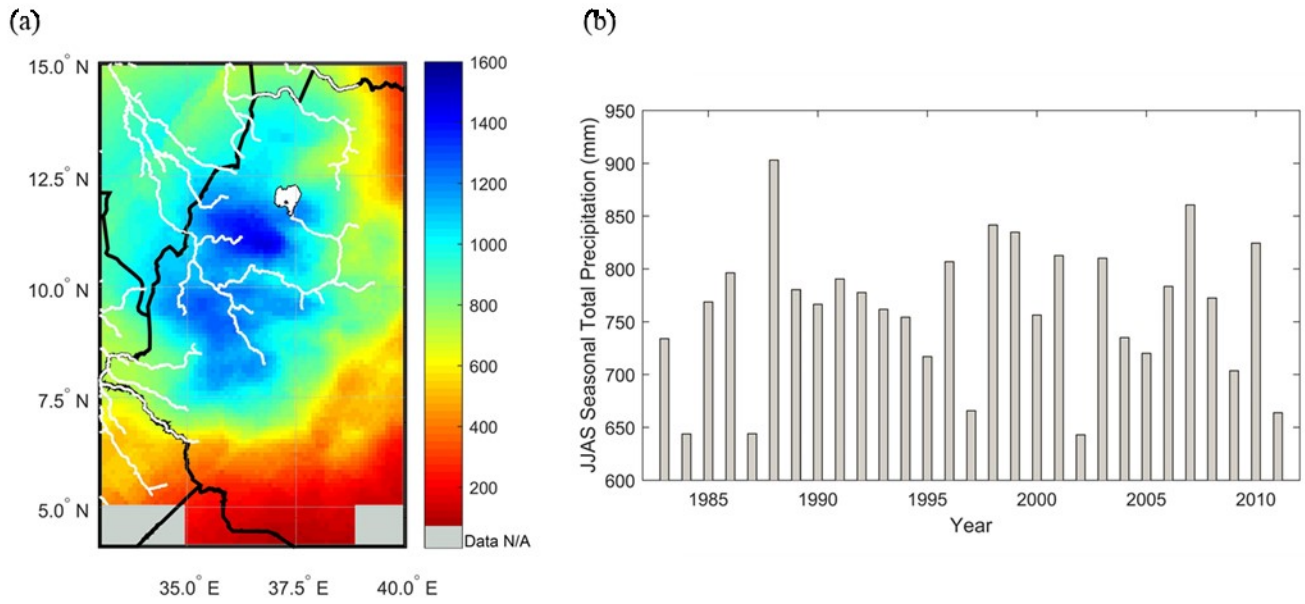


Figure 1: Spatial and temporal variability of June-September seasonal total precipitation in western Ethiopia: (a) spatial pattern of temporal-average, and (b) spatial-average time series.

- P3L24 (and others): for me the term "scenario" stands for future climate change assessments. Do you mean similar large scale conditions? Likewise in Sec. 3 the term is misleading.

To avoid misunderstanding, we have changed to “climate condition” instead of “climate scenario”. As for Sec. 3 and beyond, we kept the word as to our understanding “scenario” can also be used as “model scenarios”.

- There’s a bit more information on the clustering procedure now. However the selection of k is still not clear. WSS is smallest for the largest possible k, right? I see that the authors refer to their previous publication, but very general information should be provided.

We thank the reviewer for the comments. The following sentence is added to provide more information on the selection of *k* (Page 4 Line 25):

“…; the optimal number of clusters is identified by various evaluation metrics based on the within-cluster sum of square errors (WSS), including elbow method with *difference in WSS*, gap statistic with *difference in difference*, and qualitative analysis on post-visualization of clusters.”

- I was a bit lost in the section on the statistical modelling approach. Would it be possible to structure it into predictor selection, model calibration and evaluation?

As model calibration (cross-validation) penetrates the entire statistical modeling process, we find it hard to structure it as suggested. Besides, model evaluation is for both statistical and dynamical models and is therefore separated as a third section “*Performance metrics*” after statistical and dynamical approaches.

- Predictor regions: Why are those regions chosen (Maybe a broader literature review could justify the selection). Please also name the regions in the map.

Thank you. We have provided a broader literature review as in the very first response above and updated the map (also included here):

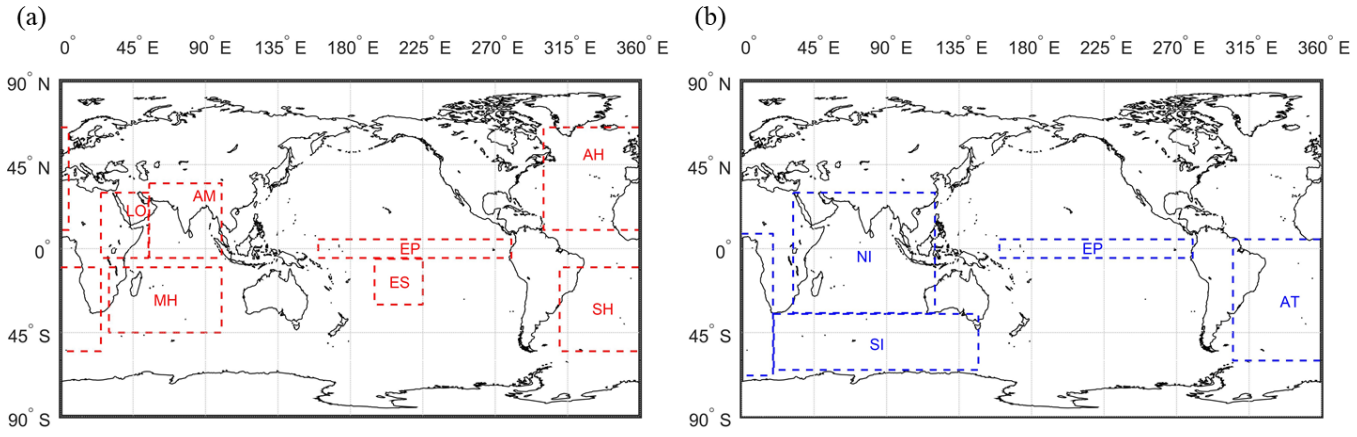


Figure 4: Justifiable climate regions globally for selecting predictors: (a) For SLP and GH at 500 mb with regions including EP, ES, LO, AH, SH, MH, and AM. For SAT, only LO is included. (b) For SST with regions including EP, NI, SI, and AT. Note: EP - equatorial Pacific region, ES – Tahiti island for ENSO measurement, LO - local region, AH - Azores High, SH -

St Helena High, MH - Mascarene High, AM - SW Asian Monsoon, NI - North Indian Ocean, SI - South Indian Ocean, AT - Equatorial/South Atlantic Ocean.

- Predictor selection approach: In step 3, the "best" grid cells are spatially averaged. This is not clear to me. If there are positive and negative correlations in the region, this might average out the predictive skill. For example in the southern Indian Ocean, both positive and negative correlations are detected in Fig. 5. Likewise the North Atlantic SLP domain, which contains not only the Azores (high) but also Iceland (low), might be problematic. Wouldn't it be straight forward, to use all gridcells directly for the PCA analysis?

We thank the reviewer for the comment. For regions containing grid-cells with both positive and negative correlations, the number of grid-cells with significant correlation in each sign is counted. If a greater number of grid-cells is associated with significant positive correlation, for example, we would average over only grid-cells with positive correlations, and vice versa. We use regional average instead of single grid-cell to hopefully filter out some noises.

To avoid confusion, the following sentence is added to the revised manuscript (Page 6 Line 34):

"… Grid-cells within each justifiable region (e.g. equatorial Pacific; Fig. 4) with correlation above the 99% significance level are identified (Fig. 5). For regions containing grid-cells with both positive and negative correlations, the number of the identified grid-cells in each sign is counted. If a greater number of grid-cells is associated with significant positive correlation, for example, only grid-cells with positive correlations are kept for the following steps, and vice versa."

- PCA: No information is given on standardization of predictor variables. If different variables are not standardized, the final PCA might be much more affected by SLP (values around 1000hpa) then SST (around 25°C).

We thank the reviewer for the comment. We did perform standardization first before using PCA. To make it clear, the following sentence is revised (Page 7 Line 1):

"Pre-predictors are standardized, combined, and transformed through principal component analysis (PCA; Jolliffe, 2002) for each cluster or non-cluster, and each dropped-year analysis separately."

- Predictor selection: How many PCA-predictors are used in the linear models at the moment? Are all of them correlated with the predictant? (Theoretically, the stepwise procedure, which the authors tested for the original predictors (author response), could also be used for the PCA based predictors. The predictors used in the linear regression (most likely very few) might then be easier interpretable. Further it might reduce overfitting, which is judged by the cross-validation.)

The number of PCs used in one model could be different for different dropped year under cross-validation. It also differs for different clusters and different model scenarios. The criterion of how many of them are kept as predictors is that the total variance explained by them should reach 95% (as stated at Page 7 Line 3).

PCA extracts the most dominant signals from the group of potential predictors and is independent of the predictand; that is, the selection of PCs does not depend on their correlation with the predictand. Therefore, PCA helps to reduce artificial skills. Combining PCA and stepwise regression could potentially select the PCs with the least dominant signals (likely noises) as predictors, which makes PCA analysis meaningless and could cause overfitting (to noises).

- P10L18: Please only use the term significant if you conducted a test (otherwise remarkable, great etc)

Thank you. We have changed related terms to "remarkable/remarkably", "evident/evidently", or "great" at appropriate places. Please see track-changes document for details.

- Dynamical Models: I am not sure if it's really necessary to do that comparison and I still feel that it is poorly integrated. The heart of the study is the statistical approach – and I think one can easily argue, that statistical predictions are good for operational use (without comparing with complex statistical models?)

As statistical prediction with gridded high spatial resolution in this study is the very first one for this region, it is hard to find similar statistical product to compare with. The NMME dynamical model outputs, in the same format of gridded prediction, within the same region and same time frame (lead time and target season) are therefore used to investigate whether our prediction makes some progress compared to the existing ones. The NMME dynamical models also produce timely outputs available for operational use.

- Cluster Results (Fig. 7) : How is the standardization conducted? The y-axis should be limited to -2/2 to better present the results.

We first standardized the time-series in each grid-cell and then average over all time-series in each cluster and the non-clustered region. We realize that in this way, some information of the target year, which is used at first for standardization at grid level, could leak through the prediction process. We apologize for the mistake (only made for *clustered indirect* case) and have rerun entire process (including grid-level predictions based on the cluster-level prediction) to ensure correct presentation of the method and outcome, even though the grid-level results do not alter much (visualization of grid-level correlation and RPSS maps is almost the same as before).

We have updated figures, tables, and some texts in result analysis (please see track-change document for details). For Fig. 7, instead of presenting standardized results, we present the actual values predicted compared to observations:

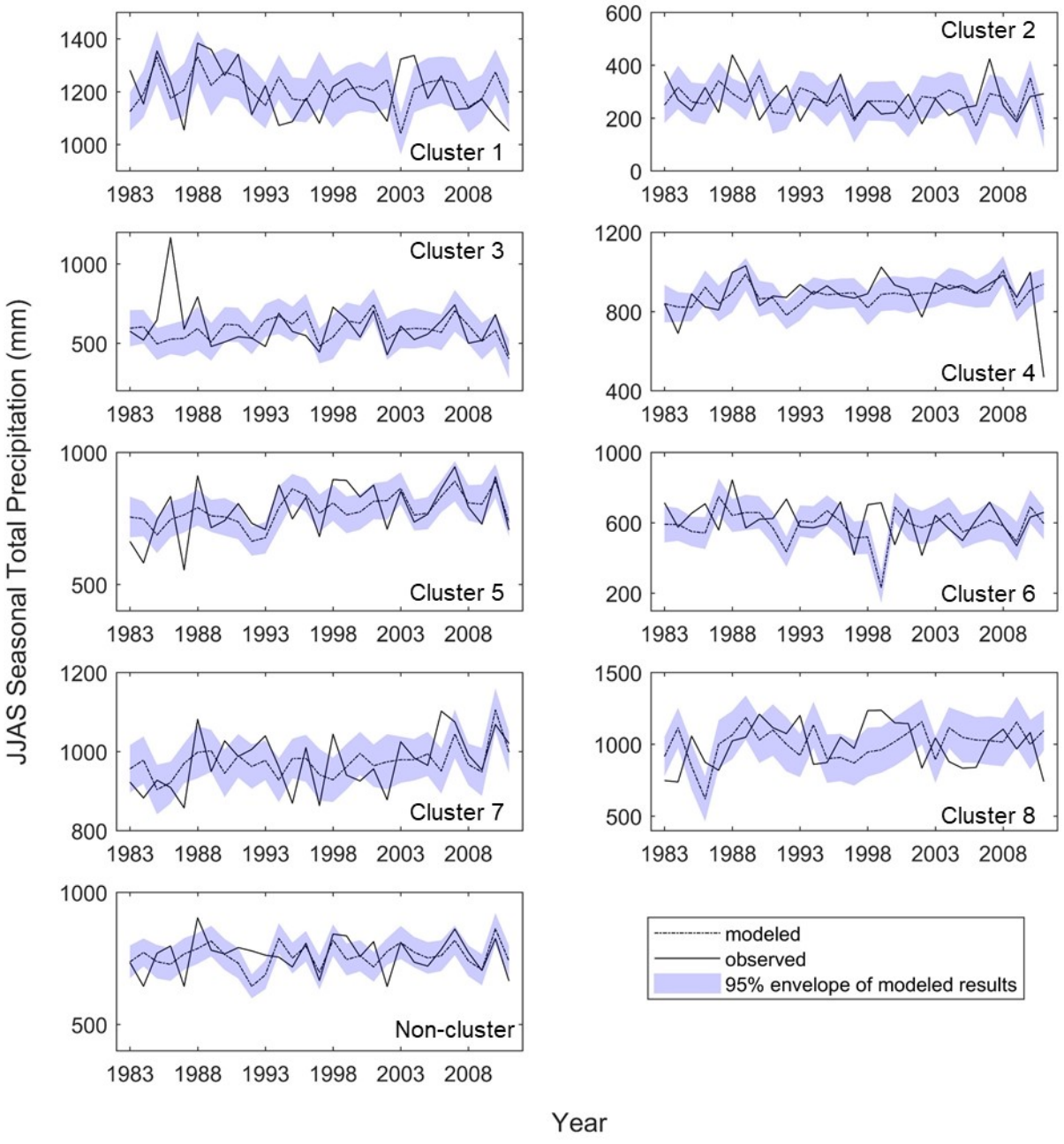


Figure 7: cluster-level predictions and observations under C-I and NC-I scenario, with drop-one-year cross-validation. The 95% envelope shows the 95% confidence interval constructed using model errors.

- Precipitation trends (Fig. 7): The two clusters with skillfull models (5 and 7) seem to have positive precipitation trends. This might lead to an overestimation of skill, since every variable with a similar trend could be used as a predictor. Thus I suggest to test the model for detrended time-series.

We appreciate the comment. We performed linear trend test using student-t statistics and found that Cluster 5 and 7 have a significant trend (slope significantly differs from zero) at 95% significance level, while trends for all others including non-cluster are insignificant.

Ideally, for Cluster 5 and 7, the trend test would be performed under cross-validation, detrend the data, predict, add the trend on, and repeat with other dropped years. It would be also ideal to detrend the global predictors for each pixel, each variable, and each period with a different dropped year. As only 29 years of data is used in this study, the conclusion on significant trend could be unreliable and often data over 30-40 years are recommended for trend analysis in this field. Considering both the complexity of the process and the length of our data, we decided to keep trend analysis aside from the modeling process, but mentioned in the discussion section that if longer time series is available, one should consider using trend analysis first before prediction. We added the following sentence in the revised manuscript (Page 17 Line 10):

"As observational datasets continue to grow, data-driven cluster analysis and statistical modeling approaches may be expected to improve. The growing length of time series and climate change impacts also call for careful analysis on possible significant trends in the data."

From the perspective of possible overestimating of the skills, even though when selecting predictors, a biased region with pseudo-high correlation may be selected, this is not likely to lead to an overestimation of skill, as the PCA detrends data at the first step and produces detrended PCs. Besides, as mentioned above, the PCA process for getting the final predictors is completely independent of predictand; that is, PCs that explain the largest portion of total variance of all selected variables (not predictand) will be remain as predictors. Moreover, cross-validation also avoids overestimated skills. Therefore, the skillful results from Cluster 5 and 7 are more likely due to the climate mechanisms affecting them.

- P23: The argument that the statistical models are of higher resolution than dynamical models is misleading. One could downscale the results the same way, as it is done with the cluster results. Further (as I mentioned during the first review round), the high resolution (indirect) forecast has the same temporal variability as the cluster forecast (due to the univariate linear relationship) and thus does not contain much additional information. The analysis of different statistical relationships at the cluster-level seem to be more relevant (in my point of view).

We agree with the reviewer that one could downscale the results the same way. In fact, we have bias-corrected the dynamical model outputs using our observation data set, and the bias-corrected outputs are in the same resolution as the statistical products. What we meant by "higher resolution" is relative to the original dynamical product. To avoid miscommunications, we have reconstructed the argument to (in abstract and elsewhere):

"The general skill (after bias-correction) of the two best performing dynamical models over the entire study region is superior to that of the statistical models, although the dynamical models issue predictions at a lower resolution and the raw predictions requires bias correction to guarantee comparable skills."

We also agree on univariate linear relationship of cluster-level prediction and regressed grid-level predictions for the *clustered* and *non-clustered indirect* cases. While comparing across *indirect* and *direct*

cases for grid-level prediction show promises. For example, as mentioned in the manuscript, the grid-level predictions under *indirect* cases perform better than *direct* cases in terms of categorical evaluation metrics (RPSS), with a higher percentage of grids demonstrating positive RPSS values.

References

BEKELE, F. 1997. Ethiopian Use of ENSO Information in Its Seasonal Forecasts. *Internet Journal of African Studies*.

BLACK, E., SLINGO, J. & SPERBER, K. R. 2003. An observational study of the relationship between excessively strong short rains in coastal East Africa and Indian Ocean SST. *Monthly Weather Review,* 131**,** 74-94.

CAMBERLIN, P. 1997. Rainfall Anomalies in the Source Region of the Nile and Their Connection with the Indian Summer Monsoon. *J. Climate,* 10**,** 1380-1392.

DIRO, G. T., GRIMES, D. I. F. & BLACK, E. 2011. Teleconnections between Ethiopian summer rainfall and sea surface temperature: part I—observation and modelling. *Climate Dynamics,* 37**,** 103-119.

ELAGIB, N. A. & ELHAG, M. M. 2011. Major climate indicators of ongoing drought in Sudan. *Journal of Hydrology,* 409**,** 612-625.

GODDARD, L. & GRAHAM, N. E. 1999. Importance of the Indian Ocean for simulating rainfall anomalies over eastern and southern Africa. *Journal of Geophysical Research: Atmospheres (1984–2012),* 104**,** 19099-19116.

JOLLIFFE, I. 2002. *Principal component analysis*, Wiley Online Library.

KALNAY, E., KANAMITSU, M., KISTLER, R., COLLINS, W., DEAVEN, D., GANDIN, L., IREDELL, M., SAHA, S., WHITE, G. & WOOLLEN, J. 1996. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American meteorological Society,* 77**,** 437-471.

KASSAHUN, B. 1987. Weather systems over Ethiopia. *Proc. First Tech. Conf. on Meteorological Research in Eastern and Southern Africa.* Nairobi, Kenya: UCAR.

KINUTHIA, J. H. 1992. Horizontal and Vertical Structure of the Lake Turkana Jet. *Journal of Applied Meteorology,* 31**,** 1248-1274.

KUMAR, K. K., RAJAGOPALAN, B., HOERLING, M., BATES, G. & CANE, M. 2006. Unraveling the Mystery of Indian Monsoon Failure During El Niño. *Science,* 314**,** 115-119.

LATIF, M., DOMMENGET, D., DIMA, M. & GRÖTZNER, A. 1999. The role of Indian Ocean sea surface temperature in forcing east African rainfall anomalies during December-January 1997/98. *Journal of Climate,* 12**,** 3497-3504.

NICHOLSON, S. E. 1996. A review of climate dynamics and climate variability in Eastern Africa. *The limnology, climatology and paleoclimatology of the East African lakes***,** 25-56.

NICHOLSON, S. E. 2014. The Predictability of Rainfall over the Greater Horn of Africa. Part I: Prediction of Seasonal Rainfall. *Journal of Hydrometeorology,* 15**,** 1011-1027.

NMSA 1996. Climate and agroclimatic resources of Ethiopia. *NMSA Meteorological Research Report Series.* Addis Ababa, Ethiopia: National Meteorological Services Agency of Ethiopia.

OMONDI, P., OGALLO, L. A., ANYAH, R., MUTHAMA, J. M. & ININDA, J. 2013. Linkages between global sea surface temperatures and decadal rainfall variability over Eastern Africa region. *International Journal of Climatology,* 33**,** 2082-2104.

SEGELE, Z. T. & LAMB, P. J. 2005. Characterization and variability of Kiremt rainy season over Ethiopia. *Meteorology and Atmospheric Physics,* 89**,** 153-180.

SEGELE, Z. T., LAMB, P. J. & LESLIE, L. M. 2009. Seasonal-to-Interannual Variability of Ethiopia/Horn of Africa Monsoon. Part I: Associations of Wavelet-Filtered Large-Scale Atmospheric Circulation and Global Sea Surface Temperature. *Journal of Climate,* 22**,** 3396-3421.

SHANKO, D. & CAMBERLIN, P. 1998. The effects of the Southwest Indian Ocean tropical cyclones on Ethiopian drought. *International Journal of Climatology,* 18**,** 1373-1388.

SLINGO, J., SPENCER, H., HOSKINS, B., BERRISFORD, P. & BLACK, E. 2005. The meteorology of the Western Indian Ocean, and the influence of the East African Highlands. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* 363**,** 25-42.

STONE, R. C., HAMMER, G. L. & MARCUSSEN, T. 1996. Prediction of global rainfall probabilities using phases of the Southern Oscillation Index. *Nature,* 384**,** 252-255.

TADESSE, T. 1994. The influence of the Arabian Sea storms/ depressions over the Ethiopian weather. *Proc. Int. Conf. on Monsoon Variability and Prediction.* Geneva, Switzerland: World Meteorological Organization.

TORRENCE, C. & WEBSTER, P. J. 1999. Interdecadal Changes in the ENSO–Monsoon System. *Journal of Climate,* 12**,** 2679-2690.

VISTE, E. & SORTEBERG, A. 2013a. The effect of moisture transport variability on Ethiopian summer precipitation. *International Journal of Climatology,* 33**,** 3106-3123.

VISTE, E. & SORTEBERG, A. 2013b. Moisture transport into the Ethiopian highlands. *International Journal of Climatology,* 33**,** 249-263.

WILLIAMS, A. P., FUNK, C., MICHAELSEN, J., RAUSCHER, S. A., ROBERTSON, I., WILS, T. H. G., KOPROWSKI, M., ESHETU, Z. & LOADER, N. J. 2011. Recent summer precipitation trends in the Greater Horn of Africa and the emerging role of Indian Ocean sea surface temperature. *Climate Dynamics,* 39**,** 2307-2328.

ZHANG, Y., MOGES, S. & BLOCK, P. 2016. Optimal Cluster Analysis for Objective Regionalization of Seasonal Precipitation in Regions of High Spatial-Temporal Variability: Application to Western Ethiopia. *Journal of Climate,* 29**,** 3697-3717.

Does objective cluster analysis serve as a useful precursor to seasonal precipitation prediction at local scale? Application to western Ethiopia

Ying Zhang[1], Semu Moges[2], Paul Block[1]

[1]Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, 53706, USA
[2]School of Civil and Environmental Engineering, Addis Ababa University, Addis Ababa, 1000, Ethiopia

Correspondence to: Paul Block (paul.block@wisc.edu)

Abstract. Prediction of seasonal precipitation can provide actionable information to guide management of various sectoral activities. For instance, it is often translated into hydrological forecasts for better water resources management. However, many studies assume homogeneity in precipitation across an entire study region, which may prove ineffective for operational and local-level decisions, particularly for locations with high spatial variability. This study proposes advancing local-level seasonal precipitation predictions by first conditioning on regional-level predictions, as defined through objective cluster analysis, for western Ethiopia. To our knowledge, this is the first study predicting seasonal precipitation at high resolution in this region, where lives and livelihoods are vulnerable to precipitation variability given the high reliance on rain-fed agriculture and limited water resources infrastructure. The combination of objective cluster analysis, spatially high-resolution prediction of seasonal precipitation, and a modeling structure spanning statistical and dynamical approaches makes clear advances in prediction skill and resolution, as compared with previous studies. The statistical model improves versus the non-clustered case or dynamical models for a number of specific clusters in northwestern Ethiopia, with ~~some~~ clusters having regional average correlation and RPSS values of ~~approximately~~ up to 0.5 and ~~27~~33%, respectively. The general skill (after bias-correction) of the two best performing dynamical models over the entire study region is superior to that of the statistical models, although the dynamical models issue predictions at a lower resolution and the raw predictions requires bias correction to guarantee comparable skills. ~~The general skill of the two best performing dynamical models over the entire study region is superior to that of the statistical models, although dynamical models issue predictions at a lower resolution.~~

1 Primer on prediction models and cluster analysis

Seasonal precipitation prediction can provide potentially actionable information to guide management of various sectoral activities. For instance, precipitation prediction is often translated into a hydrological forecast, which can be used to optimize reservoir operations, provide early flood or drought warning, inform waterway navigation, etc. As a primary input to soil moisture, precipitation prediction is also essential to agricultural management – farmers can take advantage of anticipated preferable climatic conditions or avoid unnecessary costs under expected undesirable conditions. Two types of models are commonly used for seasonal precipitation prediction: statistical and dynamical. Dynamical models, such as general circulation models (GCMs), include complex physical climate processes, while statistical models are purely data-driven, relating observations and hydroclimate variables directly.

While both modeling approaches have produced skillful seasonal predictions for a variety of applications (e.g. Barrett, 1993, Hammer et al., 2000, Shukla et al., 2016), each has noteworthy drawbacks. Dynamical models often require a ~~significant~~ great amount of time to build and parameterize, whereas statistical models require considerably fewer resources (e.g. Mutai et al., 1998, Gissila et al., 2004, Block and Rajagopalan, 2007, Diro et al., 2008, Diro et al., 2011b, Block and Goddard, 2012). Dynamical models also suffer from their high sensitivity to initial uncertain conditions, particularly given a long lead time. Consequently, a number of simulations are typically produced, each with unique initial conditions, to provide a range of possible outcomes (e.g. Roeckner et al., 1996, Anderson et al., 2007). Furthermore, the outputs from dynamical models often require additional bias correction, typically using statistical methods, to better match observations (e.g. Ines and Hansen, 2006, Block et al., 2009, Teutschbein and Seibert, 2012). Statistical models, on the other hand, are highly dependent on substantial high-quality historical data to capture hydroclimatic patterns and signals, particularly extreme conditions, which is often not available. Additionally, statistical models are often linear by construction, and may not well capture non-linear complex interactions and feedbacks. The physical nature of dynamical models, however, allows for prediction under non-stationary conditions, and also when insufficient historical data is available, whereas statistical models, by construction, typically rely on stationary relationships (Schepen et al., 2012).

The spatial extent selected for statistical seasonal prediction is critical. It is not uncommon to simply assume homogeneity in precipitation across an entire study region; however, this limits addressing potential spatial variability. While this may be suitable for very broad regional planning, it is often ineffectual for operational and local-level decisions, particularly for locations with high spatial variability. This prompts the need for delineation of sub-regional scale homogeneous regions, often defined through cluster analysis. Defining these homogeneous regions, however, is a non-trivial process. There are a variety of methods to delineate homogeneous regions, including comparing annual cycles (e.g. unimodal and bimodal distributions in precipitation) between stations (or grid-cells), comparing station correlations with regional averages, applying empirical orthogonal functions (EOF), various clustering techniques, and other methods of increasing complexity (e.g. Parthasarathy et al., 1993, Mason, 1998, Landman and Mason, 1999, Gissila et al., 2004, Diro et al., 2008, Diro et al., 2011b, Singh et al., 2012). In addition, delineation of the sub-region size is also important to consider. Smaller sized homogeneous sub-regions do not necessarily lead to improved predictions, as the noise at overly small scales can dominate any real signals representing spatial coherency of precipitation. For additional discussion regarding defining homogeneous sub-regions and cluster analysis, the reader is referred to Zhang et al. (2016) and Badr et al. (2015).

2 Application to western Ethiopia and objectives of the study

Precipitation in western Ethiopia peaks in the summer with approximately 70% of annual total precipitation falling during the main raining season - also known as the *Kiremt* season spanning from June to September (JJAS). On average, the seasonal total precipitation in the study region is approximately 760 mm; however in the northwest, precipitation can exceed 1200 mm (Fig. 1a). Along with the high spatial variability in this mountainous region, the temporal variability is also ~~significant~~ remarkable with spatial-average seasonal total precipitation ranging from 650 mm in dry years up to 900 mm in wet years (Fig. 1b). These highly variable spatial and temporal precipitation patterns have made skillful seasonal predictions challenging, particularly at local scales (e.g. Gissila et al., 2004, Block and Rajagopalan, 2007).

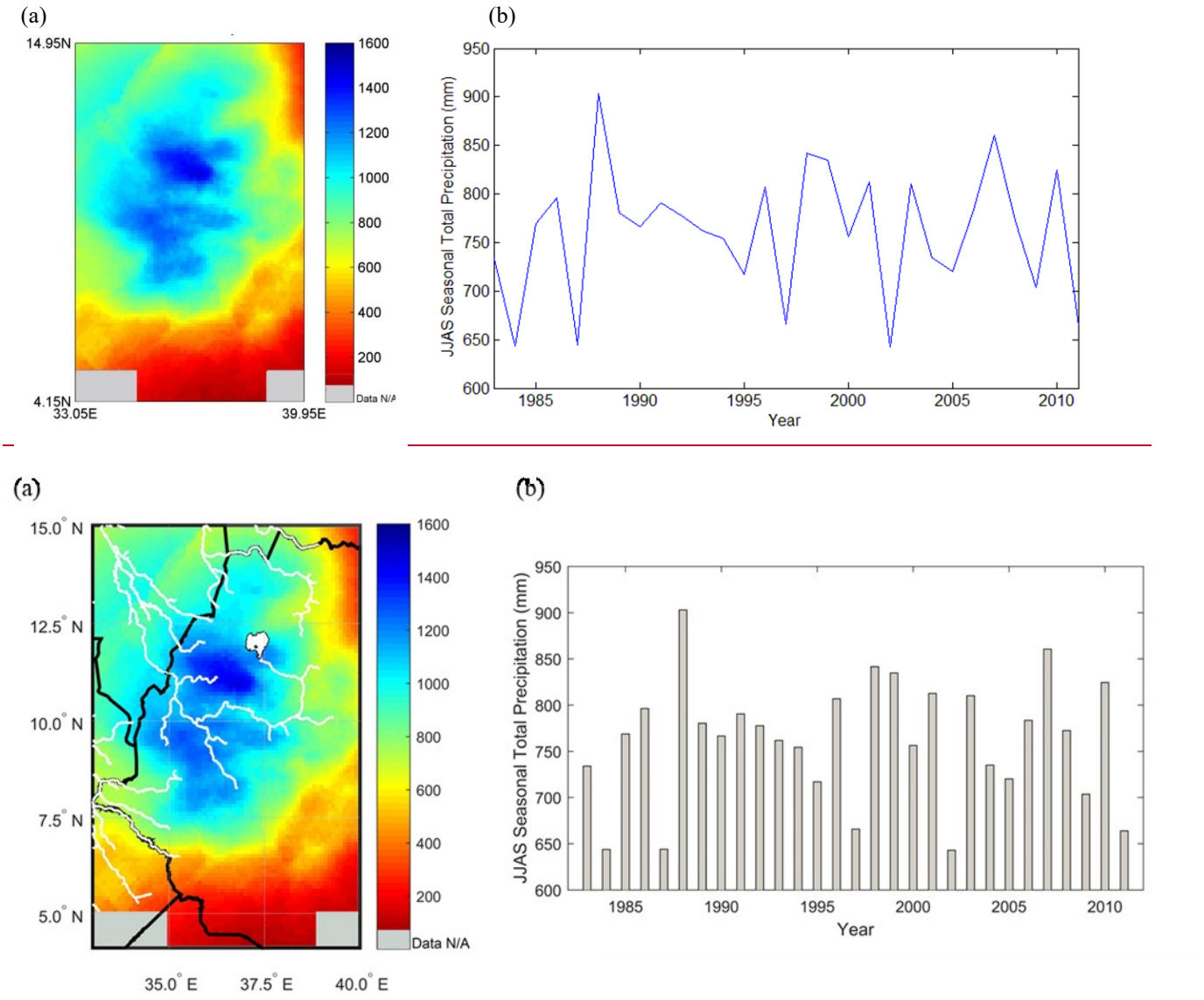


Figure 1: Spatial and temporal variability of June-September seasonal total precipitation in western Ethiopia: (a) spatial pattern of temporal-average, and (b) spatial-average time series.

Ethiopia is vulnerable to fluctuations in precipitation given its reliance on rain-fed agriculture and limited water resources infrastructure. The majority of agriculture and infrastructure are in western Ethiopia, where water resources are relatively rich

compared to other parts of the country (Awulachew et al., 2007). Operational precipitation predictions in Ethiopia have been issued by its National Meteorological Agency (NMA) since 1987 using an analog methodology (i.e. locating a similar climate ~~scenario~~ condition in the past – an analog – to predict future conditions), however this approach has produced only marginally skillful outcomes (Korecha and Sorteberg, 2013). For NMA's prediction, the country is divided into eight homogeneous regions, for which NMA produces independent predictions. Similarly, others have also addressed seasonal prediction in Ethiopia contingent on both temporal and spatial precipitation patterns. Gissila et al. (2004) divide Ethiopia into four regions conditioned on the seasonal cycle and interannual variability coherence prior to prediction, while Diro et al. (2009) apply a similar approach but with dynamic cluster boundaries, allowing for different delineations for each rainy season. Segele et al. (2015) consider statistical precipitation predictions across Ethiopia as a whole, as well as for northeastern Ethiopia and at two Ethiopian cities. Block and Rajagopalan (2007) predict the average summertime (JJAS) precipitation over the upper Blue Nile basin – a region they claim is homogenous at inter-annual time scales. Korecha and Barnston (2007) select an all-Ethiopia average precipitation index to characterize predictability broadly, with minimal attention to operational-level predictions. All of these studies focus on predicting regional average precipitation based on subjective clustering methods applying a limited number of stations or coarsely gridded data; no local predictions at a finer spatial scale are explored.

This study moves forward by exploring local-level seasonal precipitation prediction through the use of regional-level predictions, based on previous cluster analyses over western Ethiopia (Zhang et al., 2016). The advantages of defining homogeneous regions for seasonal prediction at operational (small) scales will be demonstrated by comparing approaches with and without undertaking a cluster analysis *a priori*. The combination of objective cluster analysis, spatially high-resolution prediction of seasonal precipitation, and a modeling structure spanning statistical and dynamical approaches makes clear advances compared to previous studies.

3 Modeling high-resolution seasonal prediction

To evaluate high-resolution seasonal precipitation prediction comparing with versus without cluster analysis *a priori*, statistical models are developed and further compared with bias-corrected dynamical model predictions. Four scenarios are evaluated based on two criteria – (1) *clustered* vs. *non-clustered* and (2) *direct* vs. *indirect*. In the *clustered* case, predictions are produced for each homogeneous region (cluster) given a unique set of predictors. In the *non-clustered* case, the entire study region is considered as one cluster and thus only one set of predictors is utilized for predictions. For the *direct* case, precipitation is predicted directly at the local level (grid scale); for the *indirect* case, the average precipitation within each homogeneous region is predicted first (as an intermediary), and then regressed to local-level (grid scale) predictions. Combinations of the two criteria form four scenarios – *clustered direct* (C-D), *non-clustered direct* (NC-D), *clustered indirect* (C-I), and *non-clustered indirect* (NC-I) predictions.

3.1 Cluster analysis

Using a k-means clustering technique, western Ethiopia – the major agricultural region of the country – is divided into eight homogeneous regions (Fig. 2), conditioned on the interannual variability of total precipitation in JJAS, the same variable that is to be predicted. Precipitation is based on a 0.1˚×0.1˚ gridded precipitation dataset from NMA (Dinku et al., 2014), consisting of 7320 grid-cells across 1983–2011 (29 years). This product has been verified against station data and has been deemed representative of observed precipitation in western Ethiopia (Dinku et al., 2014). Given the high-resolution gridded dataset, k-means clustering is performed for a range of predefined numbers of clusters; the optimal number of clusters is identified by various evaluation metrics

based on the within-cluster sum of square errors (WSS), including elbow method with *difference in WSS*, gap statistic with *difference in difference*, and qualitative analysis on post-visualization of clusters.~~the optimal number of clusters is identified by comparing the within-cluster sum of square errors (WSS)~~. During the clustering process, each grid-cell is assigned and reassigned to clusters until the WSS is minimized. This does not require any subjective delineation or manual delineation of boundaries between clustered stations or grid-cells; instead, an automated and objective delineation is performed. The mean time series of each cluster illustrates high intra-correlation within the cluster and low inter-correlation between any two clusters, indicating strong coherency of the clustering results. For a detailed analysis including a complete correlation table and unique patterns for each cluster-level time series associated with large climate variables, readers are referred to Zhang et al. (2016).

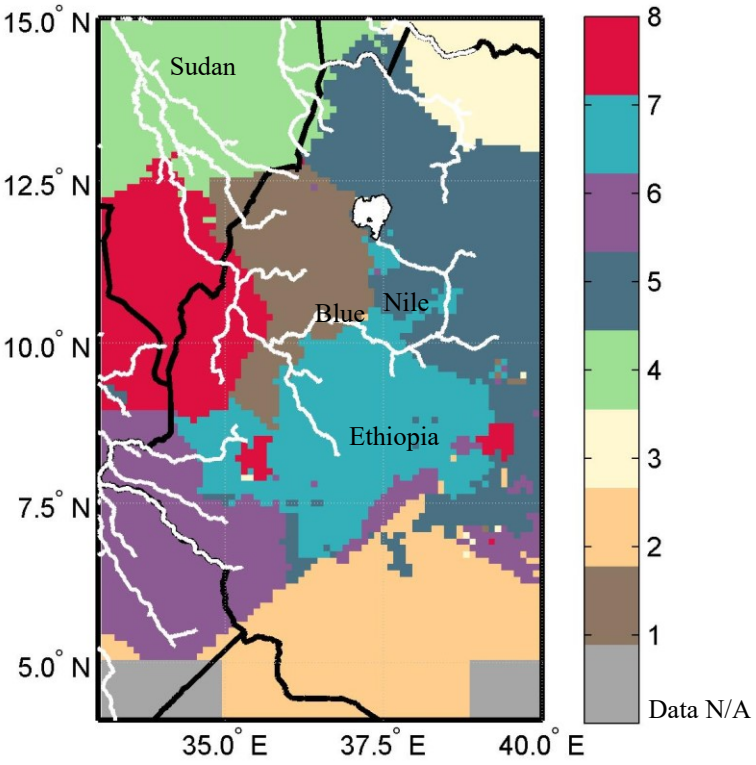


Figure 2: Regionalization map of 8 homogeneous regions marked by different colors, with country boundary and river profile. After Zhang et al. (2016)

3.2 Statistical modeling approach

Many studies have investigated statistical models for seasonal climate prediction. These studies vary by pre-classification of predictor or predictand regime, predictor selection process, and statistical methods. For example, Hertig and Jacobeit (2011) investigate sea surface temperature (SST) regimes as potential predictors for subsequent precipitation and temperature in the Mediterranean region. Through techniques including multiple applications of PCA, 17 stationary SST regimes were identified. Gerlitz et al. (2016) apply a k-means cluster analysis to grid-cells identified with significant correlations in the predictor field in order to facilitate predictor selection. Suárez-Moreno and Rodríguez-Fonseca (2015) investigate stationarity based on a long time series using a 21-year moving correlation window. The statistical prediction models are then applied to each stationary period respectively and the entire period for comparison. Despite diverse methods in seasonal prediction, multiple linear regression (MLR) is favored by many as a statistical modeling approach given its well-developed theory, simple model structure, efficient processing,

and often skillful outcomes (e.g. Omondi et al., 2013, Camberlin and Philippon, 2002, Diro et al., 2008). As mentioned, only a few studies have focused on seasonal precipitation prediction in Ethiopia (Gissila et al., 2004, Block and Rajagopalan, 2007, Korecha and Barnston, 2007, Diro et al., 2008, Diro et al., 2011b, Segele et al., 2015), and almost all of them include the applications of MLR. This study also applies MLR to predict seasonal precipitation, yet differentiates from other studies by applying predictions to pre-defined homogeneous regions and further translating to local-level predictions.

Large-scale climate variables are often evaluated as potential predictors in statistical seasonal precipitation prediction models, commonly including sea surface temperatures (SST) in the equatorial Pacific Ocean representing the well-known of the El Nino-Southern Oscillation (ENSO) (Stone et al., 1996). Sea level pressure (SLP) in the eastern Pacific Ocean at Tahiti as an critical and stable component for measuring an ENSO index (Torrence and Webster, 1999) warrants another potential predictor. For Ethiopia, the ENSO phenomenon is considered a primary indicator of precipitation variability, particularly in the main JJAS rainy season with El Nino/La Nina often associated with deficit/excess of precipitation amount in the study region (e.g. NMSA, 1996, Camberlin, 1997, Bekele, 1997, Segele and Lamb, 2005, Diro et al., 2011, Elagib and Elhag, 2011). Evidences have also shown a more direct moisture transport from the Gulf of Guinea (equatorial Atlantic Ocean), the Indian Ocean, and the Mediterranean Sea affecting Ethiopia's summer precipitation (Viste and Sorteberg, 2013a, Viste and Sorteberg, 2013b). Those moisture fluxes are often related to pressure patterns across the continent. For instance, the St. Helena high over the southern Atlantic Ocean or a high pressure over Gulf of Guinea, and a simultaneous low pressure over Indian Ocean or a monsoon trough over Arabic Peninsula bring intensified westerlies and south-westerlies that transport moist air across the Congo Basin to the western Ethiopian highlands in the summer (Segele et al., 2009, Williams et al., 2011). Similarly, the southwest Asian monsoon at the Indian Ocean, which has a strong positive relationship with the concurrent JJAS rainfall in the western Ethiopia, is associated with the Mascarene high over the southern Indian Ocean and a low pressure system near Bombay. During this monsoon season, the southeast trades in the southern hemisphere are channeled by the east African highlands while crossing the equator and become a southwest monsoon flow. It is further diverted by the Turkana Channel, enhancing convergence with the westerlies/south-westerlies above the western Ethiopian highlands and bringing moisture to this region (Kinuthia, 1992, Nicholson, 1996, Camberlin, 1997, Slingo et al., 2005, Segele et al., 2009, Nicholson, 2014). In addition, the effect of other climate variables relevant to the aforementioned driven factors, such as the Indian Ocean SST, local and other regional atmospheric pressure systems such as Azores High also have notable influence on Ethiopia's precipitation variability (e.g. Kassahun, 1987, Tadesse, 1994, NMSA, 1996, Shanko and Camberlin, 1998, Goddard and Graham, 1999, Latif et al., 1999, Black et al., 2003, Segele and Lamb, 2005).

Consequently, season-ahead (March-May) or month-ahead (May) large-scale climate variables that are physically relevant in potentially modulating moisture transport to the basin (or cluster) are selected as potential predictors. Four climate variables are selected here for further evaluation based on outcomes of the aforementioned prediction studies: SST, SLP, geopotential height (GH) at 500mb, and surface air temperature (SAT). All climate variables are from the National Centers for Environmental Prediction and National Center for Atmospheric Research (NCEP/NCAR) reanalysis dataset (Kalnay et al., 1996) at a 2.5˚×2.5˚ grid scale.

~~Large-scale climate variables are often evaluated as potential predictors in statistical seasonal precipitation prediction models, commonly including sea surface temperatures (SST) in the equatorial Pacific Ocean representing the well-known of the El Nino-Southern Oscillation (ENSO) (Stone et al., 1996). For Ethiopia, the ENSO phenomenon is considered a significant indicator of precipitation variability, particularly in the main JJAS rainy season (e.g. NMSA, 1996, Camberlin, 1997, Bekele, 1997, Segele and~~

~~Lamb, 2005, Diro et al., 2011a, Elagib and Elhag, 2011). In addition to ENSO, the effect of Indian Ocean SST and regional atmospheric pressure systems such as the St. Helena, Azores, and Mascarene Highs also have notable influence on Ethiopia's precipitation variability (e.g. Kassahun, 1987, Tadesse, 1994, NMSA, 1996, Shanko and Camberlin, 1998, Goddard and Graham, 1999, Latif et al., 1999, Black et al., 2003, Segele and Lamb, 2005). Consequently, season ahead (March-May) or month ahead (May) large-scale climate variables that are physically relevant in potentially modulating moisture transport to the basin (or cluster) are selected as potential predictors. Four climate variables are selected here for further evaluation based on outcomes of the aforementioned prediction studies: SST, sea level pressure (SLP), geopotential height (GH) at 500mb, and surface air temperature (SAT). All climate variables are from the National Centers for Environmental Prediction and National Center for Atmospheric Research (NCEP/NCAR) reanalysis dataset (Kalnay et al., 1996) at a 2.5°×2.5° grid scale.~~

To avoid overfitting, the entire process including predictor selection and statistical modeling is processed using cross-validation. To start, drop-one-year precipitation observations for JJAS averaged across the region and each cluster are spatially correlated independently with each global climate variable. As a result, there are total of 1044 global correlation maps given the 29-year time-series, eight clusters plus one non-cluster, and four climate variables. Hence, a program to automatically select highly correlated and justifiable regions as predictors is developed. The following steps describe the subsequent statistical modeling process (Fig. 3):

(1) Grid-cells within each justifiable region (e.g. equatorial Pacific; Fig. 4) with correlation above the 99% significance level are identified (Fig. 5). For regions containing grid-cells with both positive and negative correlations, the number of the identified grid-cells in each sign is counted. If a greater number of grid-cells is associated with significant positive correlation, for example, only grid-cells with positive correlations are kept for the following steps, and vice versa.

(2) The top 10% of the identified grid-cells with the highest correlation in each region is then selected, in order to boost the potential model skill.

(3) For each region, data of the selected grid-cells within the region are spatially averaged (defined as "pre-predictors").

(4) ~~Pre-predictors are combined and transformed (for each cluster or non-cluster, and each dropped-year analysis separately) through principal component analysis~~ Pre-predictors are standardized, combined, and transformed through principal component analysis (PCA; Jolliffe, 2002) for each cluster or non-cluster, and each dropped-year analysis separately.

(5a) The top principal components (PCs) from the PCA with a total of 95% variance explained are used as predictors – the direct inputs into the MLR model, otherwise known as the principal component regression (PCR). For the *direct* case, PCR is used to directly predict the grid-level precipitation; for the *indirect* case, PCR is used to predict the intermediate cluster-level precipitation.

(5b) For the *indirect* case only, cluster-level predictions are regressed to the grid-level. Note that the downscaling of cluster-level predictions to grid-level predictions is also cross-validated to avoid overfitting.

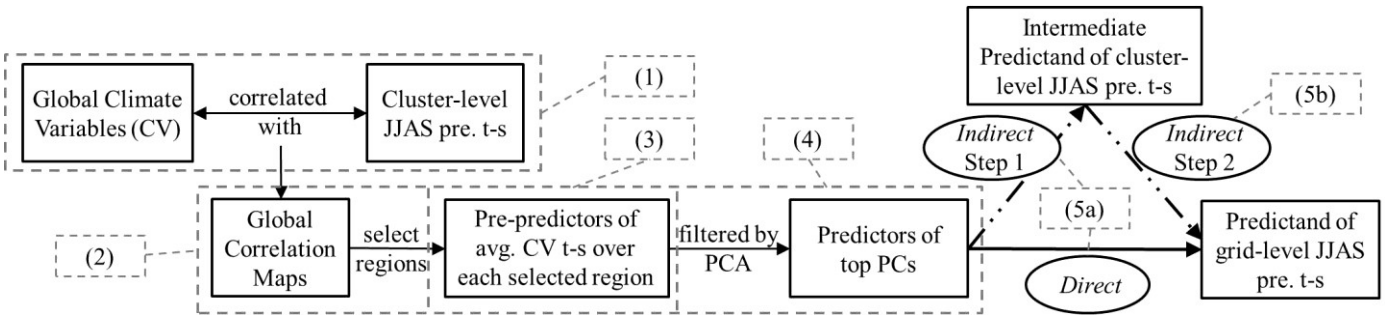


Figure 3: Flow chat of data processing for predictors into the statistical model. Numbers framed by dash lines correspond to the procedures listed in the context. Note: pre. – precipitation, t-s – time-series, avg. – average.

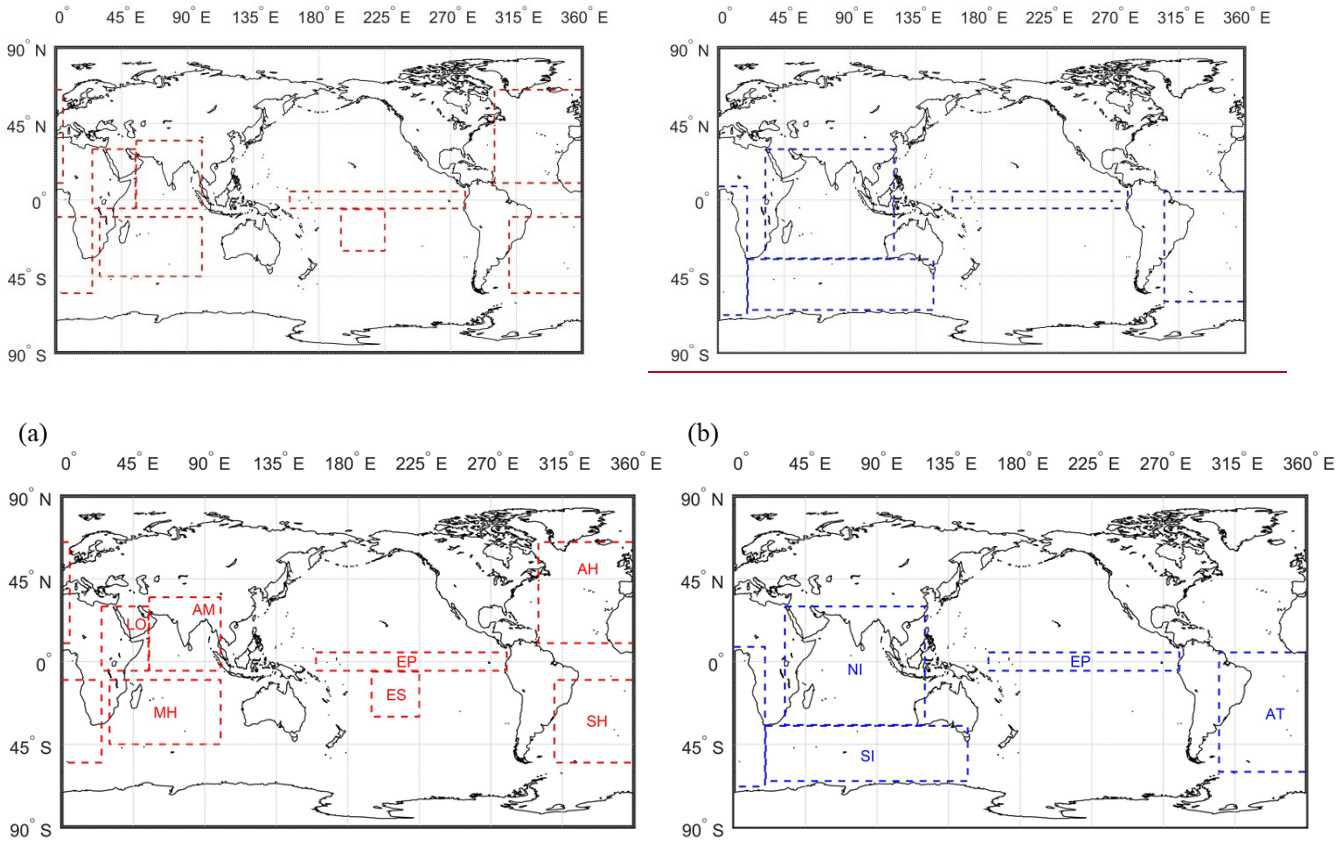


Figure 4: Justifiable climate regions globally for selecting predictors: (a) For SLP and GH at 500 mb with regions including EP, ES, LO, AH, SH, MH, and AM. For SAT, only LO is included. (b) For SST with regions including EP, NI, SI, and AT. Note: EP - equatorial Pacific region, ES – Tahiti island for ENSO measurement, LO - local region, AH - Azores High, SH - St Helena High, MH - Mascarene High, AM - SW Asian Monsoon, NI - North Indian Ocean, SI - South Indian Ocean, AT - Equatorial/South Atlantic Ocean.

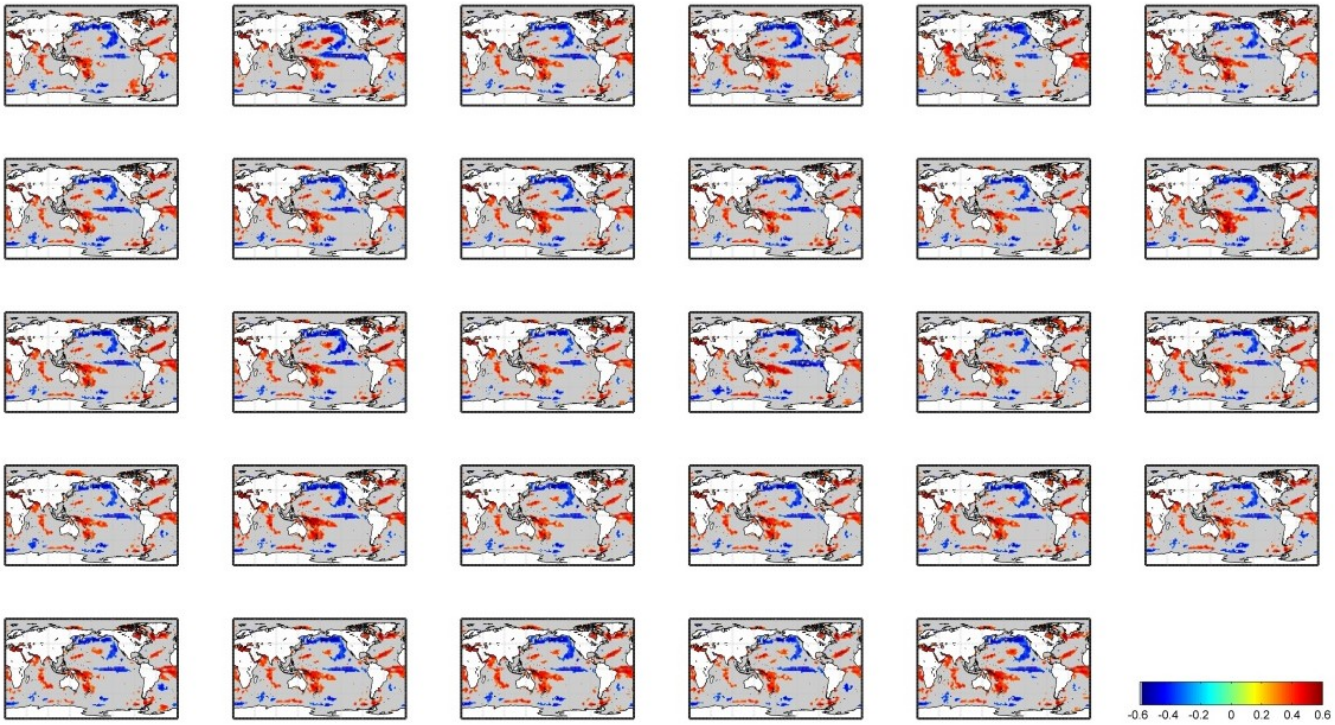


Figure 5: Correlation map between mean JJAS seasonal precipitation time series in Cluster 5 and global SST under cross-validation, with correlations lower than 99% significance level masked out (one-tail test).

PCA is a common approach in climate modeling to reduce the dimensionality of predictors and remove multi-collinearity, while simultaneously extracting the most dominant signals from the potential predictors, typically reflected in the first few PCs. Since PCA is independent of the predictand, retaining the first few PCs as predictors, in lieu of the original variables, also helps to reduce artificial prediction skill.

PCR is performed in a "drop-one-year" cross-validation mode to reduce over-fitting effects and therefore avoid overestimation of prediction skill. This requires reconstructing the principal components for the dropped year, and then multiplying the coefficient estimates with each reconstructed PC respectively in order to obtain the final predicted value for the dropped year (e.g. Block and Rajagopalan, 2009, Wilks, 2011). A 95% confidence interval of the cross-validated predictions is also constructed conditioned on model errors. Q-Q plots are evaluated to verify normally distributed residuals (results not included).

For the four scenarios, the model structures are quite similar but have subtle differences which could lead to ~~significantly~~ evidently different outcomes (Table 1). Under the NC-D (Eq. (1a, b)) and C-D scenarios (Eq. (2a, b)), the time-series of JJAS seasonal total precipitation in each grid-cell (i.e. at local level) is used as the direct predictand ($Y_{i,t}$); however, the NC-D and C-D scenarios differ, as the former uses the same predictors (X_t) across all the grid-cells, while the latter uses different predictors according to the cluster to which the grid-cell is assigned ($X_{j,t}$). In the indirect case, the cluster-level time-series of JJAS seasonal total precipitation (the time-series averaged over all grid-cells that belong to a given cluster, $Y_{m,t}$ or $Y_{j,t}$) is first predicted (Eq. (3a, b) and (4a, b)). The predicted intermediate product ($\widetilde{Y}_{m,t}$ or $\widetilde{Y}_{j,t}$) is then used as the only regressor in the second step to estimate the grid-level precipitation ($\widetilde{Y}_{i,t}$ or $\widetilde{Y}_{i\epsilon j,t}$ for every j; Eq. (3c, d) and (4c, d)). Again, for the C-I scenario, predictors in the first step are unique for

each of the eight clusters and grid-cells within that cluster ($X_{j,t}$), while predictors are identical for all grid-cells (X_t) under the NC-I scenario.

Table 1: Equations of linear regression panel models under four scenarios

	Non-clustered		Clustered	
Direct	$Y_{i,t} = \tilde{\alpha}_i + \tilde{\beta}_i X_t + \varepsilon_{i,t}$	..... (1a)	$Y_{i\epsilon j,t} = \tilde{\alpha}_i + \tilde{\beta}_i X_{j,t} + \varepsilon_{i,t}$	..... (2a)
	$\tilde{Y}_{i,t} = \tilde{\alpha}_i + \tilde{\beta}_i X_t$	..... (1b)	$\tilde{Y}_{i\epsilon j,t} = \tilde{\alpha}_i + \tilde{\beta}_i X_{j,t}$	..... (2b)
Indirect	$Y_{m,t} = \tilde{\alpha} + \tilde{\beta} X_t + \varepsilon_t$	..... (3a)	$Y_{j,t} = \tilde{\alpha}_j + \tilde{\beta}_j X_{j,t} + \varepsilon_{j,t}$	..... (4a)
	$\tilde{Y}_{m,t} = \tilde{\alpha} + \tilde{\beta} X_t$	..... (3b)	$\tilde{Y}_{j,t} = \tilde{\alpha}_j + \tilde{\beta}_j X_{j,t}$	..... (4b)
	$Y_{i,t} = \tilde{\eta}_i + \tilde{\gamma}_i \tilde{Y}_{m,t} + \nu_{i,t}$	..... (3c)	$Y_{i\epsilon j,t} = \tilde{\eta}_i + \tilde{\gamma}_i \tilde{Y}_{j,t} + \nu_{i,t}$	..... (4c)
	$\tilde{Y}_{i,t} = \tilde{\eta}_i + \tilde{\gamma}_i \tilde{Y}_{m,t}$	..... (3d)	$\tilde{Y}_{i\epsilon j,t} = \tilde{\eta}_i + \tilde{\gamma}_i \tilde{Y}_{j,t}$	..... (4d)

where Y- predictand of JJAS seasonal total precipitation; X- two predictors of top two PCs;
ε,ν - error terms; $\tilde{Y}$ - predicted values of JJAS seasonal total precipitation; $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{\eta}$, $\tilde{\gamma}$- estimated coefficients; i- grid-cell index; t- time (year) index; j- cluster index; $i \in j$- grid-cell i that belongs to clusterj; m- mean over entire study region that is equivalently the only one cluster.

3.3 Dynamical modeling approach

The North American Multi-Model Ensemble (NMME; Kirtman et al., 2014) is an experimental multi-model system consisting of coupled dynamical models from various modeling centers in North America. To our knowledge, it is also the most extensive multi-model seasonal prediction archive. The NMME provides gridded climate predictions that cover regions globally and with different lead times. The hindcasts of monthly mean precipitations are easily accessible through the International Research Institute for Climate and Society (IRI) website (http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/), and can be easily aggregated to seasonal totals for comparison with the statistical model results in this study. Therefore, NMME JJAS seasonal precipitation predictions ($1^\circ \times 1^\circ$ grid-cells) are extracted from model ensembles that cover the same time period (1983–2011), geographic region (western Ethiopia), and with the same lead time (predictions made on June 1). A subset of 10 NMME models meet these criteria and are retained for further evaluation: (1) COLA-RSMAS-CCSM3, (2) COLA-RSMAS-CCSM4, (3) GFDL-CM2p1, (4) GFDL-CM2p1-are04, (5) GFDL-CM2p5-FLOR-A06, (6) GFDL-CM2p5-FLOR-B01, (7) IRI-ECHAM-AnomalyCoupled, (8) IRI-ECHAM-DirectCoupled, (9) NASA-GMAO, (10) NCEP-CFSv2. The names are kept the same as on the International Research Institute for Climate and Society (IRI) data repository website.

The NMME predictions for each of the 10 models are bias-corrected by applying probability mapping (e.g. Block et al., 2009, Teutschbein and Seibert, 2012, Chen et al., 2013) under cross-validation, subject to the observational dataset from NMA (Fig. 6). This is performed on a grid-cell by grid-cell basis on standardized data (the NMME dataset is reshaped to $0.1° \times 0.1°$ grid-cells to match the observational NMA dataset grid-cell size). The basic steps include:

(1) Fit gamma distributions to drop-one-year time-series from each observed and NMME grid-cell; for NMME this is performed on an individual model basis using all ensemble members available. (Goodness-of-fit tests indicate gamma distributions are appropriate; results not shown.)

(2) Translate gamma distributions into cumulative distribution functions (CDF).

(3) For any given dynamical model prediction at the grid-cell level, a corrected prediction value is attained by mapping from the modeled CDF to the observed CDF and applying the inverse gamma distribution. This is repeated for all grid-cells, all NMME models, and all dropped years.

After correction, the gamma CDF of predictions and observations approximately match (Fig. 6a). Additionally, each ensemble still retains its variability over time, though the overall ensemble mean is shifted to closely match observation (Fig. 6b).

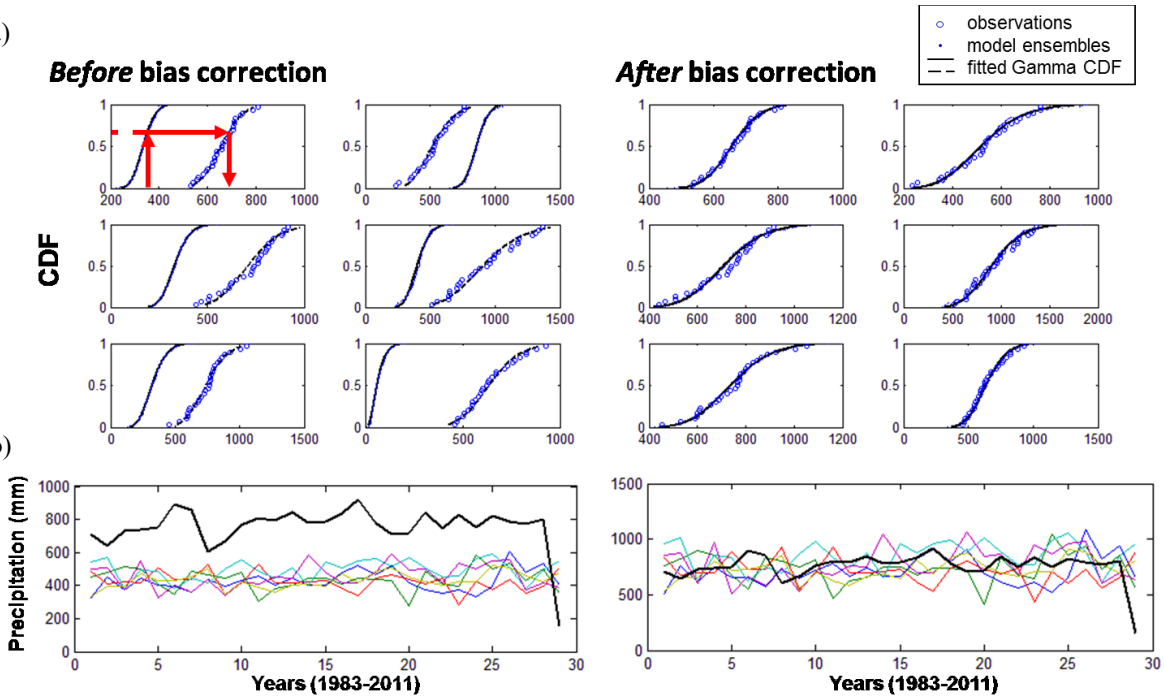


Figure 6: (a) bias correction of NMME predictions using probability mapping; (b) precipitation time-series from NMME (colored lines) before and after correction, compared to observations (black line). Examples are shown for randomly selected six grid-cells.

3.4 Performance metrics

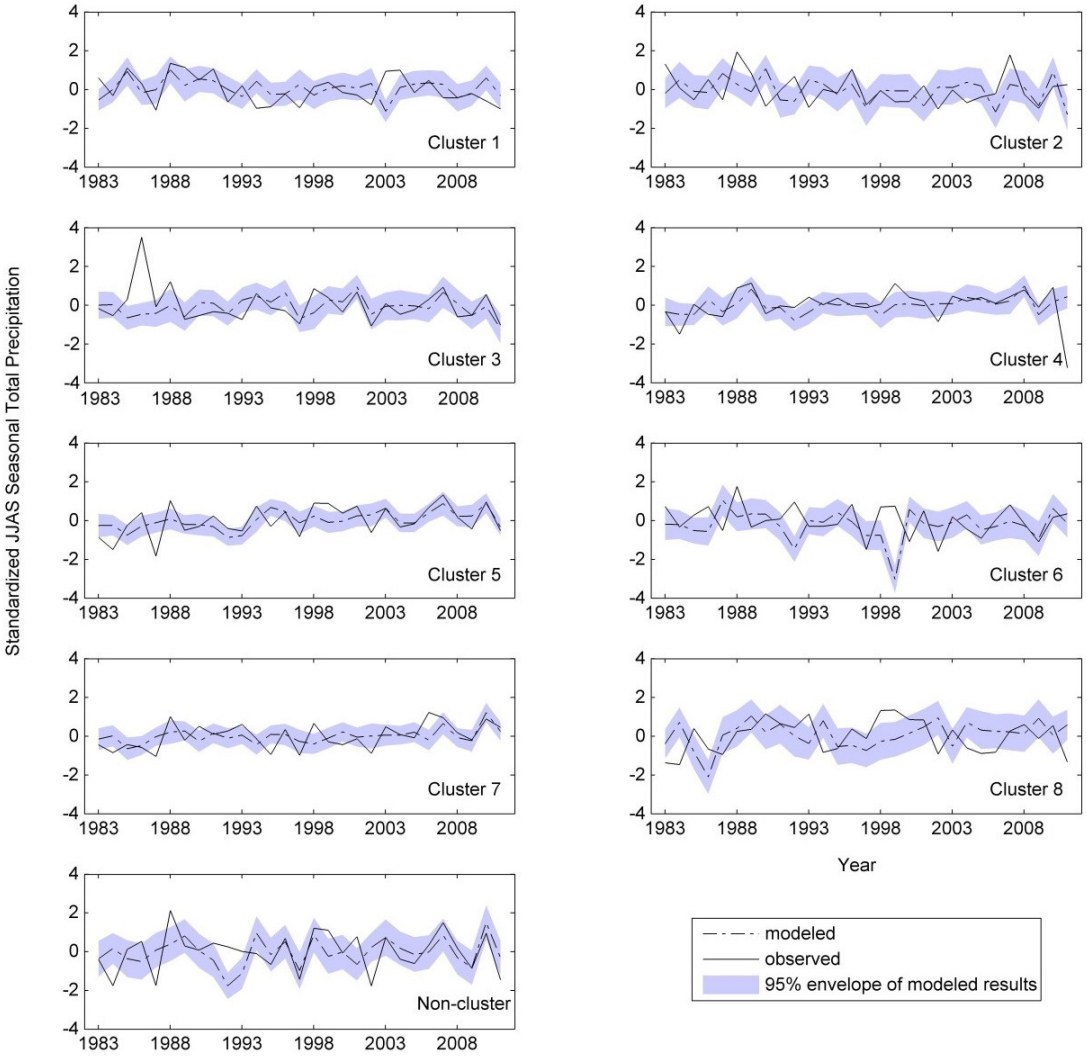
Pearson correlations are used to measure the standardized covariance between observations and predictions. Ranked probability skill scores (RPSS; Wilks, 2011) are also evaluated to determine categorical skill based on probabilistic predictions. Here, the data are split into three equal terciles representing below-normal, near-normal, and above-normal conditions. A perfect prediction yields an RPSS of 100%, and a prediction with less skill than climatology (long-term averages) yields an RPSS of less than zero. Median RPSS values from all 29 years are reported.

4 Results

4.1 Statistical model predictions

Correlations between cluster-level model predictions and observations range from -0.16~~8~~ to 0.51~~0~~, with Cluster 5 having the highest correlation and Cluster 6 the lowest (Table 2). In approximately 1/5 of the 29 years, the observation falls outside the prediction envelope (Fig. 7), indicating model overfitting and an inability of the predictors to capture precipitation variability. For RPSS, 3~~5~~ out of 8 clusters indicate superior prediction skill over climatology (Table 2). Improvement in terms of RPSS over the non-cluster scenario is evident for Cluster 1, 3~~, 5~~ and 7. ~~Among all clusters,~~ Although Cluster 5, in agriculturally rich central-

northwestern Ethiopia (Fig. 2), shows a slightly deteriorated RPSS relative to non-cluster scenario, it still performs outstandingly~~the best,~~ with the highest correlation and a positive RPSS value~~s~~ of 0.51 and ~~27~~10%, respectively. Cluster 2, 4, 6 and 8~~, however,~~ show deteriorated RPSS compared to non-cluster scenario, although those clusters are mainly regions outside Ethiopia and southern Ethiopia (Fig. 2) where water resources and agricultural activities are considerably less (Fig. 1).


Standardized JJAS Seasonal Total Precipitation
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Non-cluster
Year
modeled
observed
95% envelope of modeled results

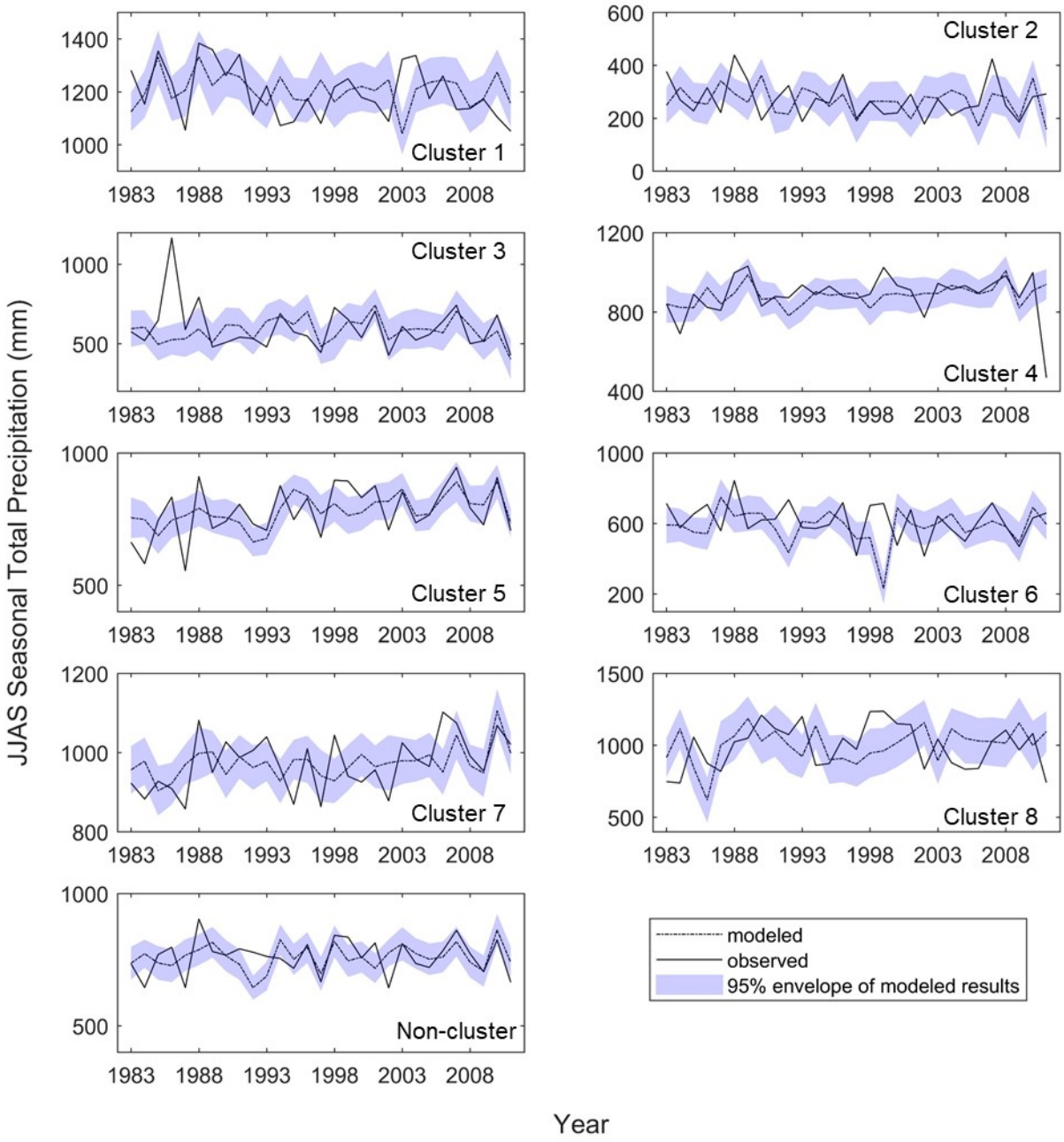
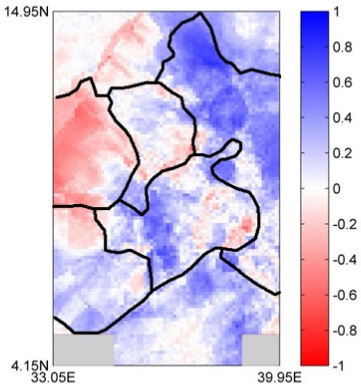
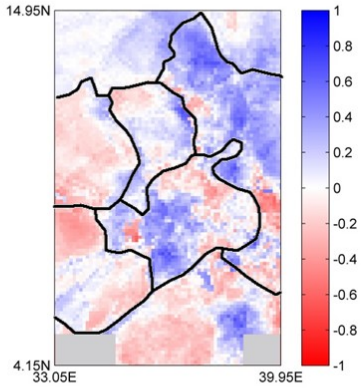
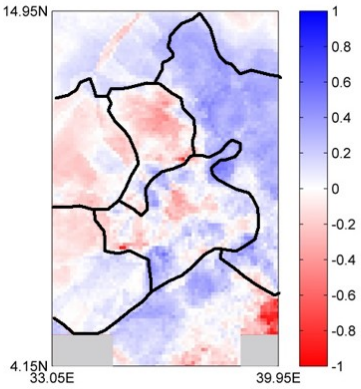
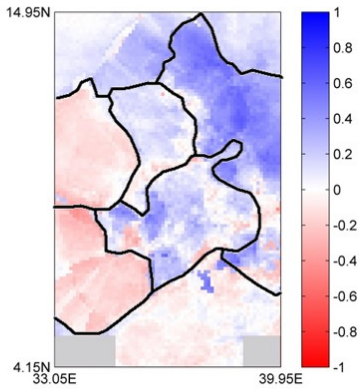


Figure 7: cluster-level predictions and observations under C-I and NC-I scenario, with drop-one-year cross-validation. The 95% envelope shows the 95% confidence interval constructed using model errors.

Table 2 Correlation coefficients (Corr.) and RPSS for predictions (drop-one-year cross-validated) at cluster level compared to observations under C-I and NC-I scenario.

Cluster	C1	C2	C3	C4	C5	C6	C7	C8	Non-cluster
Corr.	0.137~~0.163~~	-0.027~~-0.010~~	0.171~~0.179~~	0.184~~0.188~~	0.514~~0.504~~	-0.157~~-0.180~~	0.353~~0.351~~	-0.108~~-0.122~~	0.297
RPSS(%)	22.88~~33.41~~	-26.14~~-21.66~~	33.32~~43.01~~	12.74~~12.46~~	10.02~~27.40~~	-43.61~~-37.79~~	20.92~~20.63~~	-26.40~~-55.96~~	13.25

At the grid scale, depending on the case (*direct* or *indirect)*, and for different clusters, correlations between predictions and observations can favor the *clustered* case or the *non-clustered* case (Fig. 8). In general, the *indirect* model provides a smoother pattern of correlations, with grid-cells showing a negative correlation in the *direct* case now improved to near or above zero (Fig. 8). For example, Cluster 5 under the *indirect* case illustrates a more consistent positive correlation within the cluster. Some parts of the region reach a correlation over 0.6, such as central-northwestern Ethiopia (Cluster 5), which is consistent with the region of high cluster-level prediction skill. The percentage of grid-cells with correlations passing the 95% significance test is the highest for the NC-D case (Table 3); however, some locations demonstrate the lowest skills among all four scenarios.

	Non-Clustered	Clustered
Direct		
Indirect		

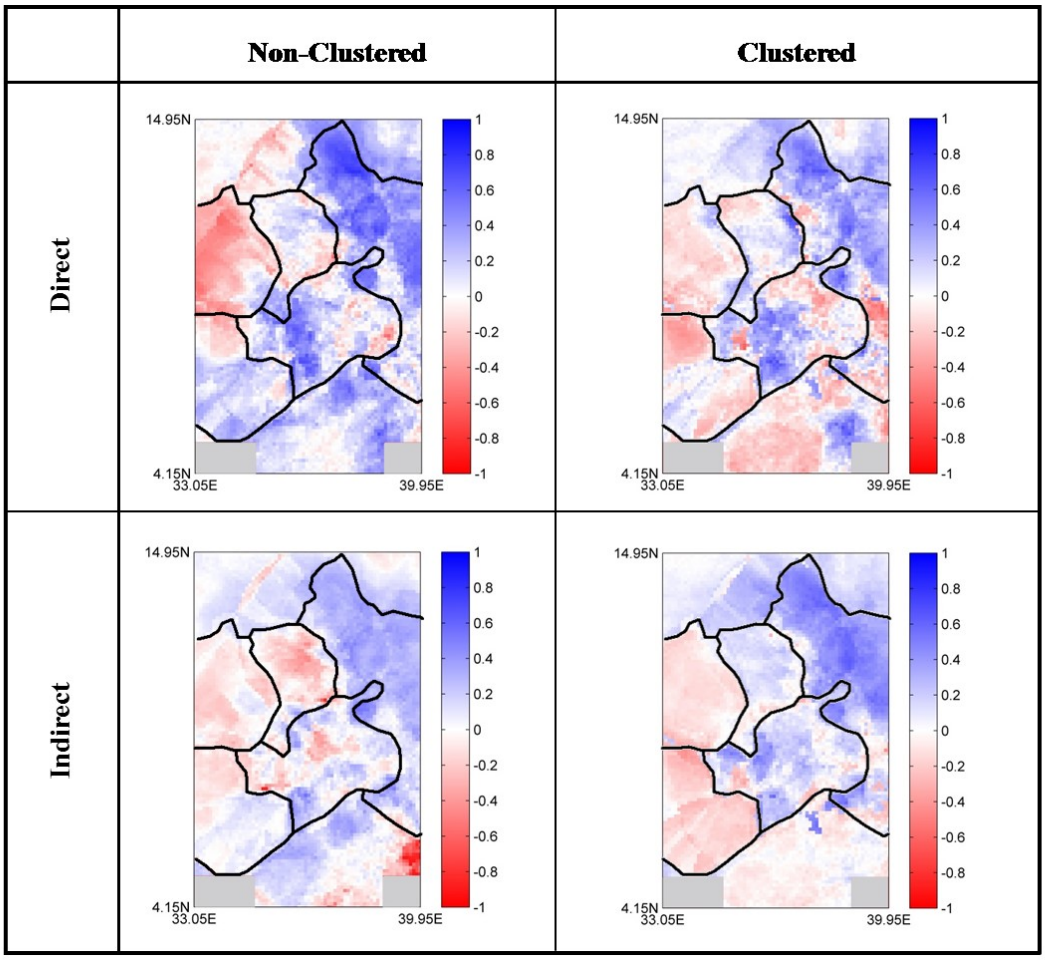


Figure 8: Pearson correlations between grid-level observations and predictions under four scenarios, with the clustering boundary delineated roughly in black.

5

Table 3: Grid-level Pearson correlation and RPSS statistics

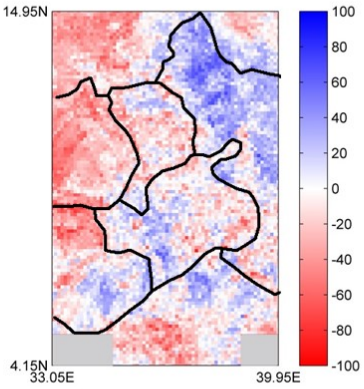
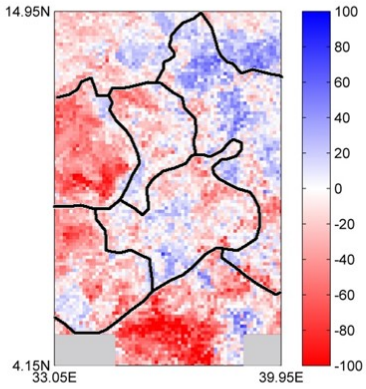
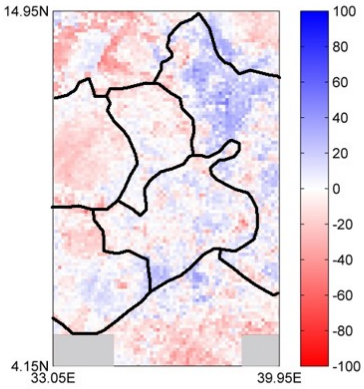
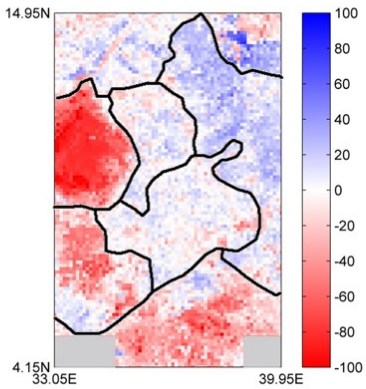
Statistical Model	Grid-level correlations			Grid-level RPSS		
	mean	stdev	significant corr %	mean (%)	stdev (%)	positive RPSS %
NC-D	0.128	0.258	19.3%	-5.21	27.0	42.8%
NC-I	0.063	0.186	3.13%	-2.26	14.6	43.9%
C-D	0.055	0.230	10.6%	-14.0	31.0	33.9%
C-I	0.080~~0.081~~	0.205~~0.206~~	12.3%~~12.4%~~	-9.93~~-9.60~~	29.3~~29.4~~	43.7%~~44.4%~~
Dynamical Model						
(1)	-0.105	0.209	0.51%	-31.4	25.4	5.70%
(2)	0.133	0.171	6.26%	-14.2	24.6	27.0%
(3)	0.086	0.130	2.08%	-14.9	25.2	26.2%
(4)	0.027	0.156	0.38%	-14.4	19.3	22.6%

(5)	0.067	0.170	1.64%	-9.66	17.0	28.4%
(6)	0.139	0.165	6.53%	-5.66	16.7	38.1%
(7)	0.102	0.130	1.67%	-8.64	17.6	31.7%
(8)	0.009	0.185	0.90%	-10.3	14.8	26.7%
(9)	0.244	0.149	23.1%	-2.33	21.8	46.0%
(10)	0.244	0.149	21.2%	-1.09	16.8	48.9%

Similar findings are evident by evaluating the RPSS except for Cluster 8; instead of improving with increased RPSS in the *indirect* case, the grid-scale predictions deteriorate given poor cluster-level prediction (for the C-I case). However, the percentage of grid-cells with positive RPSS values overall for the C-I case is still the second highest after the NC-I case~~for the C-I case~~ (Table 3), indicating the ~~*clustered indirect*~~ *indirect* cases ~~is~~are superior in terms of the number of grid-cells with improved skill compared to using climatology, particularly for grid-cells associated with skillful intermediate cluster-level predictions. The predictions are most skillful for the same region of central-northwestern Ethiopia (Cluster 5; Fig. 9) with 87% of its grid-cells showing positive RPSS and a spatial average RPSS value of 15% under the C-I scenario (Table 4).

Table 4: Grid-level Pearson correlation and RPSS statistics for grid-cells *within Cluster 5*

Statistical Model	Grid-level correlations			Grid-level RPSS		
	mean	stdev	significant corr %	mean (%)	stdev (%)	positive RPSS %
NC-D	0.378	0.211	60.7%	19.1	22.9	80.3%
NC-I	0.265	0.111	12.8%	8.33	14.8	70.3%
C-D	0.229	0.244	30.5%	6.91	24.1	62.3%
C-I	0.346~~0.345~~	0.167~~0.165~~	55.4%~~55.7%~~	14.5~~14.7~~	13.1~~13.3~~	87.0%~~87.1%~~
Dynamical Model						
(9)	0.353	0.110	46.8%	8.21	18.2	65.7%
(10)	0.248	0.130	18.4%	3.92	16.2	59.5%

	Non-Clustered	Clustered
Direct		
Indirect		

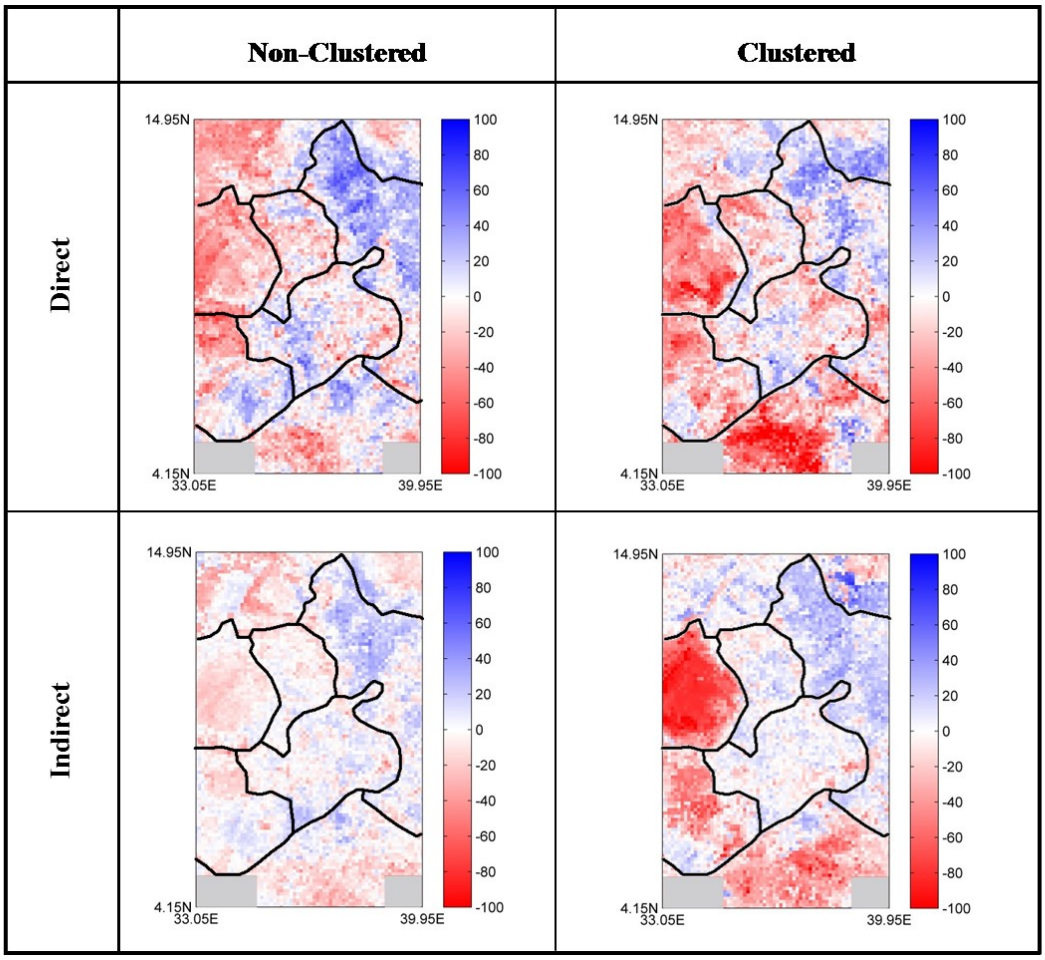


Figure 9: grid-level RPSS (%) under four scenarios using climate variables as predictors, with the clustering boundary delineated roughly in black.

4.2 Dynamical model predictions

The RPSS values based on the prediction ensembles of each dynamical model improve ~~significantly~~ remarkably after bias correction. The median RPSS values over all the grid-cells are now close to zero (Fig. 10) with two models, NASA-GMAO and NCEP-CFSv2, showing the highest RPSS value (-2.3% and -1.1%, respectively; Table 3). These two dynamical models also exhibit generally higher grid-level correlations over the study region (averaging 0.24 for both models; Table 3 and Fig. 11), as compared with other NMME models. The two best performing dynamical models after bias correction show advantage over statistical models, as assessed by correlation and RPSS metrics; however, all other dynamical models are inferior to the statistical models under NC-D and C-I scenarios, particularly given the percent of grid-cells with significant correlation and positive RPSS metrics (Table 3).

Within a certain cluster, statistical models may perform better than all dynamical models. For example, for Cluster 5, all statistical models show higher average RPSS values than that of all dynamical models (Table 4). The percentage of grid-cells with significant correlation reaches 61% for the statistical model under NC-D scenario, compared to the highest value of 47% among all the dynamical models. Similarly, the percentage with positive RPSS achieves 87% under C-I scenario as opposed to 66% for dynamical models. Note that the dynamical models also produce raw predictions in a lower spatial resolution (1°×1°) than the statistical models (0.1°×0.1°) and requires bias correction to guarantee comparable skills.

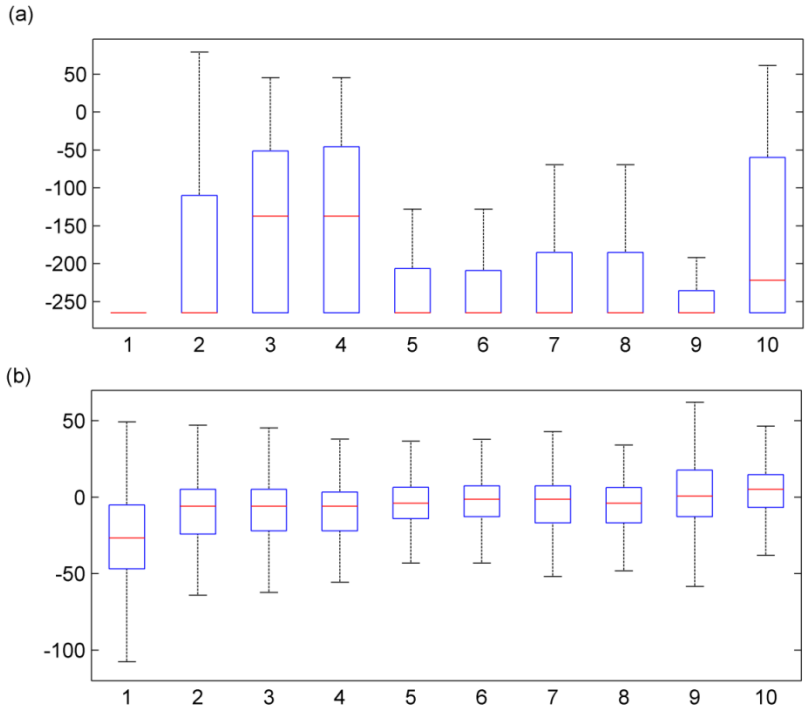


Figure 10: Boxplots of grid-level RPSS (%) for 10 dynamical models from NMME (a) before and (b) after bias correction, labeled with the same number as listed in the context. Note: For each box plot, the line inside the box is the median, the box edges represent the 25th and 75th percentiles, and the whiskers extend to the most extreme data points not considered outliers (outliers not shown).

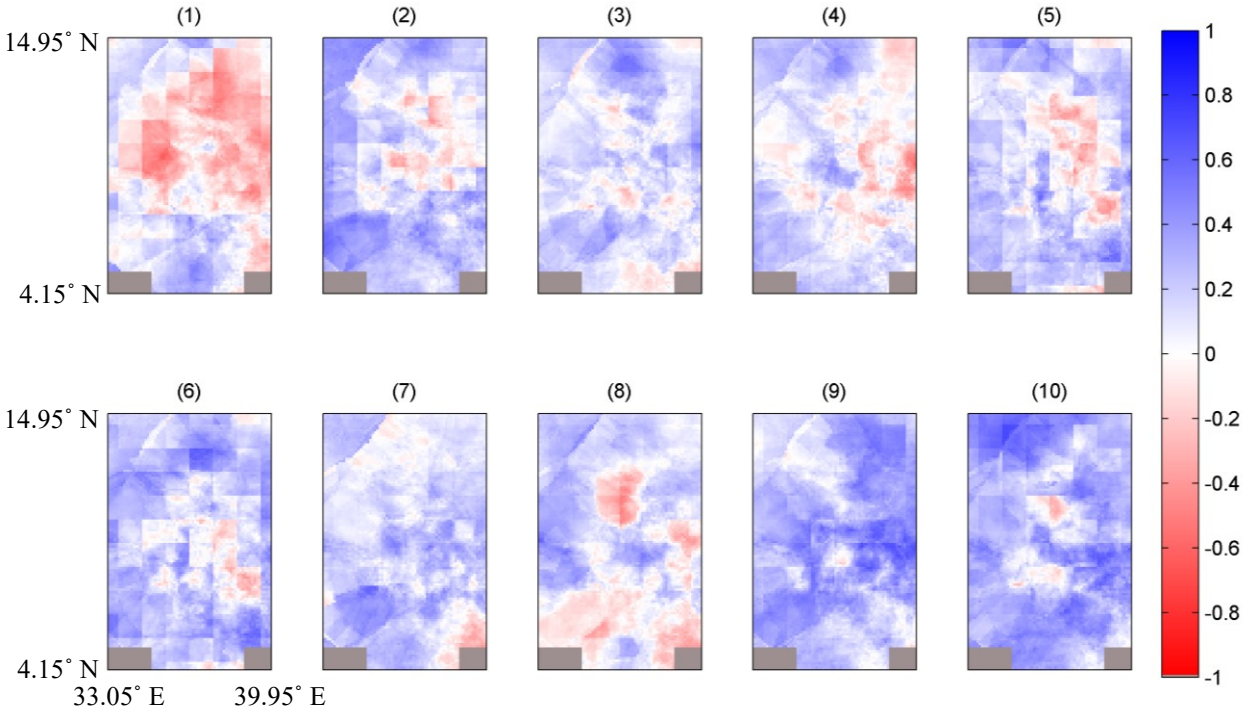


Figure 11: Pearson correlations between grid-level observations and ensemble mean of bias-corrected predictions for 10 dynamical models from NMME, labeled with the same number as listed in the context. Note that the scale ranges from -1 to 1.

5 Conclusions and discussion

This study demonstrates the potential for season-ahead large-scale climate information to produce skillful and credible high-resolution precipitation predictions under a *clustered indirect* approach in western Ethiopia. At the regional scale, the approach shows promise for northwestern Ethiopia (Cluster 1, 3, 5, and 7), particularly compared to current NMA operational forecasts, which are only moderately more skillful than climatology (Korecha and Sorteberg, 2013). The regional average RPSS in this study under the *clustered* case ranges from ~~21~~10% to ~~43~~3% for northwestern Ethiopia, as opposed to values under 6% for NMA operational forecast (Korecha and Sorteberg, 2013). The approach adopted here also advances on previous studies (Gissila et al., 2004, Block and Rajagopalan, 2007, Korecha and Barnston, 2007, Diro et al., 2011b, Segele et al., 2015) by first applying an objective cluster analysis and then conditionally constructing high-resolution predictions. A unique set of predictors is applied to each cluster, which contributes to superior prediction performance at cluster levels in northwestern Ethiopia, as compared with predictions from the *non-clustered* approach. Grid-level prediction under the *clustered indirect* case also reduces the effect of over-fitting relative to the *direct* case and improves negative RPSS values to near or above zeros; that said, the *non-clustered direct* case also illustrates higher correlation and RPSS values on average.

Two out of 10 NMME dynamical models, NASA-GMAO and NCEP-CFSv2, demonstrate overall superior performance to the statistical models; however, for certain regions such as Cluster 5, the performance of statistical models under *clustered indirect* and *non-clustered direct* cases is still superior. It is also worth noting that the statistical model predictions are at a one hundred times finer spatial resolution than the dynamical models providing additional advantages at the local scale, when skillful. Nevertheless, improvements in dynamical models continue and their application to seasonal precipitation prediction is likely to grow (e.g. Palmer et al., 2004, Saha et al., 2006, Lim et al., 2009).

Relatively poor prediction performance is evident in some locations such as southwestern Ethiopia and regions outside Ethiopia, where the hydroclimatic processes that produce precipitation might be driven by local factors or other regional climate patterns rather than large-scale climate variables identified in this study. A previous study (Zhang et al., 2016) has shown that the influence of ENSO on JJAS precipitation in western Ethiopia decreases generally from north to south, and is likely one of the reasons why skills are relatively low in southwestern Ethiopia. Cluster 5 was also identified with the strongest connection to equatorial Pacific SST (Zhang et al., 2016), which is consistent with the highest skill found in this study. Other regions with low prediction skill show relatively strong connections to SST in neighboring oceanic regions. However, connections with those climate patterns appear to be less robust than with ENSO, making the predictions in their associated regions less skillful. This is also consistent with the findings from other studies that even though all three oceans (Indian, Atlantic, and Pacific Ocean) affect the JJAS precipitation in western Ethiopia, the Pacific Ocean still plays the greatest role (Segele et al., 2009, Omondi et al., 2013).

The southwest Asian monsoon over Indian Ocean may also be critical in determining the precipitation, given that the clusters with better prediction skills lie along the pathway of the monsoon. Based on the global concurrent correlation maps between JJAS precipitation and SLP for each cluster, Cluster 5 and 7 – the two clusters with the best skills – are the only ones that are strongly and negatively corrected with SLP near Bombay, and meantime strongly and positively correlated with the SLP at the eastern equatorial Pacific Ocean. The former indicates that a strong southwest Asian monsoon is associated with higher JJAS precipitation amount, and vice versa. The latter indicates that a high surface pressure over the eastern equatorial Pacific Ocean often accompanied with cold SST and a raining pattern – a La Nina phenomenon – also brings higher JJAS precipitation to the western Ethiopia, and vice versa. Cluster 2 – one of the worst predicted clusters – shows moderately strong negative correlation with SLP

near Bombay; however, it is also correlated strongly and negatively with SLP at southern Indian Ocean, indicating a possible weak gradient of the southwest Asian monsoon. Moreover, its correlation with SLP over equatorial Pacific Ocean is nonsignificant. Considering in general El Nino suppresses the monsoon and La Nina increases it (Kumar et al., 2006), strong correlations with both ENSO and the monsoon in the correct direction, such as for Cluster 5 and 7, indicate a double insurance over their association with the southwest Asian monsoon. Therefore, clusters which are more affected by the southwest Asian monsoon over Indian Ocean, particularly coupled with the influence of ENSO, are likely to show more promises in their prediction skills.

~~Relatively poor prediction performance is evident in some locations such as southwestern Ethiopia and regions outside Ethiopia, where the hydroclimatic processes that produce precipitation might be driven by local factors or other regional climate patterns rather than large-scale climate variables identified in this study. A previous study (Zhang et al., 2016) has shown that the influence of ENSO on JJAS precipitation in western Ethiopia decreases generally from north to south, and is likely one of the reasons why skills are relatively low in southwestern Ethiopia. Cluster 5 was also identified with the strongest connection to equatorial Pacific SST (Zhang et al., 2016), which is consistent with the highest skill found in this study. Other regions with low prediction skill show relatively strong connections to SST and pressure systems in neighboring oceanic regions. However, connections with those climate patterns appear to be less robust than with ENSO, making the predictions in their associated regions less skillful. To test the prospects for improving prediction performance by including season-ahead local variables, soil moisture and spring rains were investigated; however, no significant improvement was found; in fact, performance skill deteriorates when adding local predictors may simply serve to introduce more noise and encourage over-fitting.~~

Additional prediction features also warrant future attention, including longer prediction lead times and evaluation of other relevant characteristics (e.g. intra-seasonal dry spells, seasonal onset or cessation, etc.). As observational datasets continue to grow, data-driven cluster analysis and statistical modeling approaches may be expected to improve. The growing length of time series and climate change impacts also call for careful analysis on possible significant trends in the data. Improving predictive capabilities may not be a complete panacea, but it can continue to be an important part of a decisions-maker's portfolio as they cope with hydroclimatic variability and its inherent risks.

6 Data availability

The National Centers for Environmental Prediction and National Center for Atmospheric Research (NCEP/NCAR) reanalysis dataset can be accessed through the National Oceanic & Atmospheric Administration (NOAA) Earth System Research Laboratory (ESRL) website (https://www.esrl.noaa.gov/psd/data/reanalysis/).

The NMME hindcasts are available through the International Research Institute for Climate and Society (IRI) website (http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/).

The gridded precipitation dataset in western Ethiopia is available upon request from NMA (http://www.ethiomet.gov.et/).

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This study was supported by NASA Project NNX14AD30G and NSF PIRE Project 1545874. We acknowledge the National Meteorological Agency of Ethiopia for sharing data. We also want to thank the reviewers for their suggestions in improving this work.

References

ANDERSON, D., STOCKDALE, T., BALMASEDA, M., FERRANTI, L., VITART, F., MOLTENI, F., DOBLAS-REYES, F., MOGENSEN, K. & VIDARD, A. 2007. Development of the ECMWF seasonal forecast System 3. *ECMWF Technical Memoranda,* 503.

AWULACHEW, S. B., YILMA, A. D., LOULSEGED, M., LOISKANDL, W., AYANA, M. & ALAMIREW, T. 2007. *Water resources and irrigation development in Ethiopia*, Iwmi.

BADR, H. S., ZAITCHIK, B. F. & DEZFULI, A. K. 2015. A tool for hierarchical climate regionalization. *Earth Science Informatics*, 1-10.

BARRETT, C. B. 1993. THE DEVELOPMENT OF THE NILE HYDROMETEOROLOGICAL FORECAST SYSTEM1. Wiley Online Library.

BEKELE, F. 1997. Ethiopian Use of ENSO Information in Its Seasonal Forecasts. *Internet Journal of African Studies*.

BLACK, E., SLINGO, J. & SPERBER, K. R. 2003. An observational study of the relationship between excessively strong short rains in coastal East Africa and Indian Ocean SST. *Monthly Weather Review,* 131**,** 74-94.

BLOCK, P. & GODDARD, L. 2012. Statistical and Dynamical Climate Predictions to Guide Water Resources in Ethiopia. *Journal of Water Resources Planning and Management,* 138**,** 287-298.

BLOCK, P. & RAJAGOPALAN, B. 2009. Statistical–Dynamical Approach for Streamflow Modeling at Malakal, Sudan, on the White Nile River. *Journal of Hydrologic Engineering,* 14**,** 185-196.

BLOCK, P. J., FILHO, F. A. S., SUN, L. & KWON, H. H. 2009. A Streamflow Forecasting Framework Using Multiple Climate and Hydrological Models. *Journal of the American Water Resources Association,* 45**,** 828-843.

BLOCK, P. J. & RAJAGOPALAN, B. 2007. Interannual Variability and Ensemble Forecast of Upper Blue Nile Basin Kiremt Season Precipitation. *J. Hydrometeor,* 8**,** 327-343.

CAMBERLIN, P. 1997. Rainfall Anomalies in the Source Region of the Nile and Their Connection with the Indian Summer Monsoon. *J. Climate,* 10**,** 1380-1392.

CAMBERLIN, P. & PHILIPPON, N. 2002. The East African March–May Rainy Season: Associated Atmospheric Dynamics and Predictability over the 1968–97 Period. *Journal of Climate,* 15**,** 1002-1019.

CHEN, J., BRISSETTE, F. P., CHAUMONT, D. & BRAUN, M. 2013. Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resources Research,* 49**,** 4187-4205.

DINKU, T., HAILEMARIAM, K., MAIDMENT, R., TARNAVSKY, E. & CONNOR, S. 2014. Combined use of satellite estimates and rain gauge observations to generate high-quality historical rainfall time series over Ethiopia. *International Journal of Climatology,* 34**,** 2489-2504.

DIRO, G. T., BLACK, E. & GRIMES, D. I. F. 2008. Seasonal forecasting of Ethiopian spring rains. *Meteorological Applications,* 15**,** 73-83.

DIRO, G. T., GRIMES, D. I. F. & BLACK, E. 2011a. Teleconnections between Ethiopian summer rainfall and sea surface temperature: part I—observation and modelling. *Climate Dynamics,* 37**,** 103-119.

DIRO, G. T., GRIMES, D. I. F. & BLACK, E. 2011b. Teleconnections between Ethiopian summer rainfall and sea surface temperature: part II. Seasonal forecasting. *Climate Dynamics,* 37**,** 121-131.

DIRO, G. T., GRIMES, D. I. F., BLACK, E., O'NEILL, A. & PARDO-IGUZQUIZA, E. 2009. Evaluation of reanalysis rainfall estimates over Ethiopia. *International Journal of Climatology,* 29**,** 67-78.

ELAGIB, N. A. & ELHAG, M. M. 2011. Major climate indicators of ongoing drought in Sudan. *Journal of Hydrology,* 409**,** 612-625.

GERLITZ, L., VOROGUSHYN, S., APEL, H., GAFUROV, A., UNGER-SHAYESTEH, K. & MERZ, B. 2016. A statistically based seasonal precipitation forecast model with automatic predictor selection and its application to central and south Asia. *Hydrol. Earth Syst. Sci.,* 20**,** 4605-4623.

GISSILA, T., BLACK, E., GRIMES, D. I. F. & SLINGO, J. M. 2004. Seasonal forecasting of the Ethiopian summer rains. *International Journal of Climatology,* 24**,** 1345-1358.

GODDARD, L. & GRAHAM, N. E. 1999. Importance of the Indian Ocean for simulating rainfall anomalies over eastern and southern Africa. *Journal of Geophysical Research: Atmospheres (1984–2012),* 104**,** 19099-19116.

HAMMER, G. L., NICHOLLS, N. & MITCHELL, C. 2000. *Applications of seasonal climate forecasting in agricultural and natural ecosystems*, Springer Science & Business Media.

HERTIG, E. & JACOBEIT, J. 2011. Predictability of Mediterranean climate variables from oceanic variability. Part II: Statistical models for monthly precipitation and temperature in the Mediterranean area. *Climate Dynamics,* 36**,** 825-843.

INES, A. V. M. & HANSEN, J. W. 2006. Bias correction of daily GCM rainfall for crop simulation studies. *Agricultural and Forest Meteorology,* 138**,** 44-53.

JOLLIFFE, I. 2002. *Principal component analysis*, Wiley Online Library.

KALNAY, E., KANAMITSU, M., KISTLER, R., COLLINS, W., DEAVEN, D., GANDIN, L., IREDELL, M., SAHA, S., WHITE, G. & WOOLLEN, J. 1996. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American meteorological Society,* 77**,** 437-471.

KASSAHUN, B. 1987. Weather systems over Ethiopia. *Proc. First Tech. Conf. on Meteorological Research in Eastern and Southern Africa.* Nairobi, Kenya: UCAR.

KIRTMAN, B. P., MIN, D., INFANTI, J. M., KINTER, J. L., PAOLINO, D. A., ZHANG, Q., VAN DEN DOOL, H., SAHA, S., MENDEZ, M. P., BECKER, E., PENG, P., TRIPP, P., HUANG, J., DEWITT, D. G., TIPPETT, M. K., BARNSTON, A. G., LI, S., ROSATI, A., SCHUBERT, S. D., RIENECKER, M., SUAREZ, M., LI, Z. E., MARSHAK, J., LIM, Y.-K., TRIBBIA, J., PEGION, K., MERRYFIELD, W. J., DENIS, B. & WOOD, E. F. 2014. The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction. *Bulletin of the American Meteorological Society,* 95**,** 585-601.

KORECHA, D. & BARNSTON, A. G. 2007. Predictability of June–September Rainfall in Ethiopia. *Monthly Weather Review,* 135**,** 628-650.

KORECHA, D. & SORTEBERG, A. 2013. Validation of operational seasonal rainfall forecast in Ethiopia. *Water Resources Research,* 49**,** 7681-7697.

LANDMAN, W. A. & MASON, S. J. 1999. Operational long-lead prediction of South African rainfall using canonical correlation analysis. *International Journal of Climatology,* 19**,** 1073-1090.

LATIF, M., DOMMENGET, D., DIMA, M. & GRÖTZNER, A. 1999. The role of Indian Ocean sea surface temperature in forcing east African rainfall anomalies during December-January 1997/98. *Journal of Climate,* 12**,** 3497-3504.

LIM, E.-P., HENDON, H. H., HUDSON, D., WANG, G. & ALVES, O. 2009. Dynamical Forecast of Inter–El Niño Variations of Tropical SST and Australian Spring Rainfall. *Monthly Weather Review,* 137**,** 3796-3810.

MASON, S. 1998. Seasonal forecasting of South African rainfall using a non–linear discriminant analysis model. *International Journal of Climatology,* 18**,** 147-164.

MUTAI, C. C., WARD, M. N. & COLMAN, A. W. 1998. Towards the prediction of the East Africa short rains based on sea-surface temperature–atmosphere coupling. *International Journal of Climatology,* 18**,** 975-997.

NMSA 1996. Climate and agroclimatic resources of Ethiopia. *NMSA Meteorological Research Report Series.* Addis Ababa, Ethiopia: National Meteorological Services Agency of Ethiopia.

OMONDI, P., OGALLO, L. A., ANYAH, R., MUTHAMA, J. M. & ININDA, J. 2013. Linkages between global sea surface temperatures and decadal rainfall variability over Eastern Africa region. *International Journal of Climatology,* 33**,** 2082-2104.

PALMER, T., ALESSANDRI, A., ANDERSEN, U. & CANTELAUBE, P. 2004. Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bulletin of the American Meteorological Society,* 85**,** 853.

PARTHASARATHY, B., KUMAR, K. R. & MUNOT, A. A. 1993. Homogeneous Indian Monsoon rainfall: Variability and prediction. *Proceedings of the Indian Academy of Sciences - Earth and Planetary Sciences,* 102**,** 121-155.

ROECKNER, E., OBERHUBER, J. M., BACHER, A., CHRISTOPH, M. & KIRCHNER, I. 1996. ENSO variability and atmospheric response in a global coupled atmosphere-ocean GCM. *Climate Dynamics,* 12**,** 737-754.

SAHA, S., NADIGA, S., THIAW, C., WANG, J., WANG, W., ZHANG, Q., VAN DEN DOOL, H. M., PAN, H. L., MOORTHI, S., BEHRINGER, D., STOKES, D., PEÑA, M., LORD, S., WHITE, G., EBISUZAKI, W., PENG, P. & XIE, P. 2006. The NCEP Climate Forecast System. *Journal of Climate,* 19**,** 3483-3517.

SCHEPEN, A., WANG, Q. J. & ROBERTSON, D. E. 2012. Combining the strengths of statistical and dynamical modeling approaches for forecasting Australian seasonal rainfall. *Journal of Geophysical Research: Atmospheres,* 117**,** n/a-n/a.

SEGELE, Z. T. & LAMB, P. J. 2005. Characterization and variability of Kiremt rainy season over Ethiopia. *Meteorology and Atmospheric Physics,* 89**,** 153-180.

SEGELE, Z. T., RICHMAN, M. B., LESLIE, L. M. & LAMB, P. J. 2015. Seasonal-to-Interannual Variability of Ethiopia/Horn of Africa Monsoon. Part II: Statistical Multi-Model Ensemble Rainfall Predictions. *Journal of Climate***,** 150129124820009.

SHANKO, D. & CAMBERLIN, P. 1998. The effects of the Southwest Indian Ocean tropical cyclones on Ethiopian drought. *International Journal of Climatology,* 18**,** 1373-1388.

SHUKLA, S., ROBERTS, J., HOELL, A., FUNK, C. C., ROBERTSON, F. & KIRTMAN, B. 2016. Assessing North American multimodel ensemble (NMME) seasonal forecast skill to assist in the early warning of anomalous hydrometeorological events over East Africa. *Climate Dynamics***,** 1-17.

SINGH, A., KULKARNI, M. A., MOHANTY, U. C., KAR, S. C., ROBERTSON, A. W. & MISHRA, G. 2012. Prediction of Indian summer monsoon rainfall (ISMR) using canonical correlation analysis of global circulation model products. *Meteorological Applications,* 19**,** 179-188.

STONE, R. C., HAMMER, G. L. & MARCUSSEN, T. 1996. Prediction of global rainfall probabilities using phases of the Southern Oscillation Index. *Nature,* 384**,** 252-255.

SUÁREZ-MORENO, R. & RODRÍGUEZ-FONSECA, B. 2015. S⁴CAST v2.0: sea surface temperature based statistical seasonal forecast model. *Geosci. Model Dev.,* 8**,** 3639-3658.

TADESSE, T. 1994. The influence of the Arabian Sea storms/ depressions over the Ethiopian weather. *Proc. Int. Conf. on Monsoon Variability and Prediction.* Geneva, Switzerland: World Meteorological Organization.

TEUTSCHBEIN, C. & SEIBERT, J. 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology,* 456–457**,** 12-29.

WILKS, D. S. 2011. *Statistical methods in the atmospheric sciences*, Academic press.

ZHANG, Y., MOGES, S. & BLOCK, P. 2016. Optimal Cluster Analysis for Objective Regionalization of Seasonal Precipitation in Regions of High Spatial-Temporal Variability: Application to Western Ethiopia. *Journal of Climate,* 29**,** 3697-3717.