

RESEARCH ARTICLE

Evaluating the CMIP5 ensemble in Ethiopia: Creating a reduced ensemble for rainfall and temperature in Northwest Ethiopia and the Awash basin

Ellen Dyer¹  | Richard Washington¹ | Meron Teferi Taye²

¹School of Geography and the Environment, University of Oxford, Oxford, UK

²International Water Management Institute, Addis Ababa, Ethiopia

Correspondence

Ellen Dyer, School of Geography and the Environment, University of Oxford, Oxford, UK.

Email: ellen.dyer@ouce.ox.ac.uk

Funding information

UK Department for International Development (DFID)

Abstract

The purpose of this study was to evaluate the historical skill of models in the Coupled Model Intercomparison Project Phase 5 (CMIP5) in two regions of Ethiopia: northwestern Ethiopia and the Awash, one of the main Ethiopian river basins. An ensemble of CMIP5 models was first selected so that atmosphere-only (Atmospheric Model Intercomparison Project, AMIP) and fully coupled simulations could be directly compared, assessing the effects of coupled model sea surface temperature (SST) biases. The annual cycle, seasonal biases, trends, and variability were used as metrics of model skill. In the Awash basin, both coupled and AMIP simulations had late Belg or March-May (MAM) rainy seasons. In connection to this, most models also missed the June rainfall minimum entirely. Northwest Ethiopia, which has a unimodal rainfall cycle in observations, is shown to have bimodal seasonality in models, even in the AMIP simulations. Significant AMIP biases in these regions show that model biases are not related to SST biases alone. Similarly, a clear connection between model resolution and skill was not found. Models simulated temperature with more skill than rainfall, but trends showed an underestimation in Belg (MAM/April-May (AM)) trends, and an overestimation in Kiremt or July-September (JAS/June-September (JJAS)) trends. The models which were shown to have the most skill in a range of categories were HadGEM2-AO, GFDL-CM3, and MPI-ESM-MR. The biases and discrepancies in model skill for different metrics of rainfall and temperature found in this study provide a useful basis for a process-based analysis of the CMIP5 ensemble in Ethiopia.

KEYWORDS

climate, ensembles, general circulation models, policy, rainfall, seasonal, tropics

1 | INTRODUCTION

Africa's vulnerability to climate change is widely acknowledged to be higher than most other regions

(Niang *et al.*, 2014). This vulnerability can be partitioned into exposure due to the physical changes projected in the climate and the ability to adapt to such changes. Africa's high share of vulnerability is in stark contrast to

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. International Journal of Climatology published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

its contribution to global emissions of greenhouse gases (AFDB, 2011).

The connection between climate and humanitarian disaster has focused on Ethiopia more than almost any other country in Africa, partly because of a number of drought events that have ended in famines which have required humanitarian intervention (Conway and Schipper, 2011). A strong connection between the health of the Ethiopian economy and climate has also been made in the past (Grey and Sadoff, 2007) although the nuances of this connection are complicated (Lewis, 2017; Borgomeo *et al.*, 2018). A number of studies have highlighted the need to understand the meteorology of drought events, the spatial heterogeneity of such events, and what adaptation might be necessary to reduce future risk (Viste *et al.*, 2013; Lewis, 2017). Climate models are key to analysing the projected risk of climate change, including droughts. In most regions of Africa there has been investment in the analysis of global climate models, for example in southern Africa (Dieppoos *et al.*, 2015; Munday and Washington, 2017; Pohl *et al.*, 2017; Munday and Washington, 2018), East Africa (Otieno and Anyah, 2013; Yang *et al.*, 2015; Hirons and Turner, 2018; Ongoma *et al.*, 2018) and the Sahel (Cook and Vizy, 2006; Vizy *et al.*, 2013; Martin *et al.*, 2014; Monerie *et al.*, 2017). However the ability of GCMs to reproduce the historical climate of Ethiopia has not been widely studied, possibly because of the unique climate characteristics of the country such as the complex rainfall seasonality which render it distinct from East Africa to the south, and the Sahel to the west. The model studies that have been done, focus on a particular rainy season alone (Li *et al.*, 2016), or a smaller sub-set of models (Diro *et al.*, 2011; Degefu *et al.*, 2017). Establishing the fidelity of global climate models in different regions of Ethiopia, and for different seasons could yield strategic information about why models can or cannot simulate the climate, and teleconnections. This is an important step towards a better constrained set of future climate projections. In turn, this will increase the possibility of better adaptive capacity and possibly reduce negative exposure to the effects of climate change.

The purpose of this study is to identify global climate models from the Coupled Model Intercomparison Project Phase 5 (CMIP5) that most accurately reproduce rainfall and temperature, but also to characterize patterns of skill in the ensemble of models we are using. We are doing this in two regions of Ethiopia (Figure 1), namely: (a) the north-west of the Rift Valley (Northwest Ethiopia) and (b) one of the main Ethiopian river basins, the Awash basin. Annual rainfall masked for these two regions from the Climate Hazards Group InfraRed Precipitation with Station dataset (CHIRPS) are shown in Figure 1. We chose Northwest Ethiopia, as its climate is distinct from other parts of East Africa, with rainfall seasonality which is closer to the

unimodal distribution of the Sahel. The Awash basin has rainfall seasonality which is on the margin of the bi-modal East African and unimodal Sahelian distributions and was chosen because of its significance to water security in the region (e.g., REACH, 2015; <https://reachwater.org.uk>). The basin experiences droughts and floods which frequently challenge water security of communities and various sectors; changes to the climate are expected to exacerbate the current water insecurity situation. Climate model evaluations are therefore targeting impact studies on water allocation for different sectors through understanding future water availability projections.

There are four parts to the model comparisons. First, there is an inspection of individual model climatologies relative to the reference dataset. Second, the major rainy seasons are identified and average annual and seasonal biases are calculated. Agreement in trend sign and magnitude is then assessed for the same seasons and annual averages. Finally, variability in the full time series is assessed by using Taylor diagrams, which combine standard deviation and time series correlation with a reference dataset (Taylor *et al.*, 2012). The values of the skill score from the Taylor diagrams are not used as thresholds, but rather to identify where models are clustered, and differences between the Atmospheric Model Intercomparison Project (AMIP) and coupled ensembles. From these comparisons, a reduced ensemble is suggested for process-based evaluation (James *et al.*, 2015).

2 | BACKGROUND: ETHIOPIAN CLIMATE

Ethiopia sits at the juncture of the Sahel and East Africa, and it is affected by numerous climatic zones: hot lowlands to the east, cooler and wetter highlands to the west, and numerous large river basins. Ethiopia has three seasons: Kiremt (June–September), Bega (October–January), and Belg (February–May) (Shanko and Camberlin, 1998). However, the intensity and exact timing of these seasons varies regionally. The complexity of the Ethiopian climate is reflected in the state of understanding of its future projections; the fifth intergovernmental panel on climate change (IPCC) assessment report (AR5) states that there is a broad range of changes and a large degree of model disagreement (Niang *et al.*, 2014).

Complexity is part of the reason the Ethiopian climate is understudied. Cheung *et al.* (2008) examined rainfall trends by splitting Ethiopia up into 13 different watersheds and found a wide range of both negative and positive trends. Flohn (1987) and Nicholson (1986) showed that Ethiopian droughts from the 1950s to 1980s were part of a large-scale pattern of African drought in which

the latitude band from West Africa to Ethiopia would face concurrent drought. Conversely, the 2009 drought, a dry Belg followed by a dry Kiremt, in Ethiopia was different, with extreme dryness extending from Ethiopia to the Democratic Republic of the Congo (Viste *et al.*, 2013). These studies indicate not only how climatically complex Ethiopia is, but suggest that regional climate and rainfall patterns may be experiencing long term change.

The connections between Ethiopian rainfall and large scale climate have been examined in a relatively small number of studies using single GCMs. Jury and Funk (2013) used the GFDL model and Degefu *et al.* (2017) used two different versions of the HadGEM2 model. Degefu *et al.* (2017) found that there are teleconnections with the Nino3.4 region, the Indian Ocean Dipole (IOD), and central Indian Ocean sea surface temperatures (SSTs) in observations, but found that teleconnections simulated by the models were much weaker than observations. They also found that the resolution of the models did not impact on the relative strength of these teleconnections. Jury and Funk (2013) were investigating long term historical trends, and future trends, finding that warmer western Indian Ocean SSTs, and a westward shift in the Walker circulation are both important drivers of drying in Ethiopia, but that the orographic effect of the highlands lessens the severity of the drying compared to surrounding lowland regions. The complex teleconnections and varying model skill levels indicate that using models for future projections without first assessing model fidelity is risky.

A comparison of the full ensemble of CMIP5 models has been done by Li *et al.* (2016) for historical period Kiremt (June-September (JJAS)) rainfall. They determined that if models were grouped by grid-spacing that a threshold of 2° could differentiate models better able to capture the annual cycle of rainfall in this season. This threshold does lead to a sub-group of models which can better capture the annual cycle, but the ability of models to reproduce observed variability, and trends outside of the JJAS season were not addressed.

3 | MODELS, REANALYSIS, AND MEASUREMENTS

Both historical, fully coupled simulations and AMIP simulations from the CMIP5 ensemble (Taylor *et al.*, 2012) are used in this assessment. AMIP runs are analysed since they preclude the influence of SST bias. Our CMIP5 subset is therefore limited to models run in both configurations and for the period 1981–2005 which is common to all data sets. The rli1p1 ensemble member for each model is used. A summary of the 24 models examined is included in Table 1.

Our primary reference datasets are the CHIRPS merged satellite and rain gauge dataset (Funk *et al.*, 2015),

and the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis (Dee *et al.*, 2011). CHIRPS is available at a $0.05^\circ \times 0.05^\circ$ resolution and uses an amalgamation of satellite retrieved rainfall and rain gauge data in Africa. CHIRPS incorporates publicly available and private station data from meteorological agencies in sub-Saharan Africa, and therefore includes information from a higher number of gauges than other datasets. This is beneficial in a data sparse region, and this dataset has been shown to perform well in Eastern Africa compared to gauge data and other merged rainfall datasets (Dinku *et al.*, 2018). We also use CRU TS3.22 rainfall (Harris *et al.*, 2014) observations to assess the agreement between reference data sets. We do not use rainfall from reanalysis datasets because their biases can be very high compared to observations (Tsidu, 2012), and therefore we are relying on rainfall observations that incorporate gauges.

We use temperature from the ERA-Interim reanalysis ($0.75^\circ \times 0.75^\circ$ grid spacing). As with rainfall, we also include comparisons with other temperature datasets such as National Centers for Environmental Prediction reanalysis (NCEP2), and Climate Research Unit (CRU) TS 3.22 temperature fields (Kanamitsu *et al.*, 2002; Harris *et al.*, 2014). ERA-Interim and CRU have very similar temperature climatologies, with NCEP2 being cooler than both, and having a May rather than June maximum. Due to the agreement between ERA-Interim and CRU, along with its higher resolution than that available from NCEP2, we use ERA-Interim as the reference for bias differences, trends and variability but compare model temperature climatologies to all three datasets.

4 | RESULTS

4.1 | Rainfall

4.1.1 | Rainfall: Annual cycle

Rainfall climatologies for models and observations over the Awash basin are shown in Figure 2 (individual model annual cycles are shown in Figure A1). While the modelled ensemble averages do show a weak bimodal seasonality, there are some important discrepancies in the averages, and a large spread in ensemble members. In the AMIP ensemble, there is large overestimation in both rainy seasons among a number of models. While most models include a distinct JAS rainy season, the March-May (MAM) rainy season is entirely missed by four out of the 24 models. It is the lack of bimodal nature of some climatologies, due to very low rainfall in MAM, accompanied by a large overestimation of MAM rainfall

in others, which allows the ensemble mean to be near to observed rainfall in MAM. In the coupled ensemble, the MAM rainy season is not evident in a number of models, and the ensemble resembles a unimodal distribution. All but one model (MIROC5) miss March rainfall, and this model greatly overestimates the March rainfall while reproducing the relative seasonality of rainfall quite well.

The separation between the two rainy seasons of low June rain rates is not captured by the AMIP ensemble average. This is a combination of models having a prolonged JJAS season (five out of 24), and others with a delayed MAM rainy season (six out of 24). The AMIP ensemble also has a main rainy season that extends to October. Although this is less pronounced in the coupled ensemble, model rainy seasons are shifted by 1 month to varying degrees with September having a rain rate almost

as high as August. As with the AMIP ensemble average, the June minimum is not dry in the coupled ensemble.

The observed annual cycle of Northwest Ethiopia (Figure 3a) differs from that of the Awash basin in that there is no distinct dry period between two rainy seasons; it is unimodal with the main season covering June to September and an onset in April/May. Maximum monthly rain rates can reach $6 \text{ mm}\cdot\text{day}^{-1}$ in observations, compared to the $4 \text{ mm}\cdot\text{day}^{-1}$ in the Awash basin. Northwest Ethiopia and the Awash basin do have a similarly dry period from November to February with monthly rain rates below $1 \text{ mm}\cdot\text{day}^{-1}$. The annual cycles of the model ensembles for Northwest Ethiopia (individual model annual cycles are shown in Figure A2) have similar characteristics to those for the Awash basin with higher rain rates throughout the year, while the

TABLE 1 The 24 CMIP5 models included in this study

Model	Institution	Horizontal resolution	Vertical levels
ACCESS1-0	CSIRO (Commonwealth Scientific and Industrial Research Organisation, Australia), and BOM (Bureau of Meteorology, Australia)	1.25×1.875	38
ACCESS1-3	CSIRO	1.25×1.875	38
BNU-ESM	Beijing Normal University	2.7906×2.8125	26
CanCM4	Canadian Centre for Climate Modelling and Analysis	2.7906×2.8125	35
CCSM4	National Center for Atmospheric Research (NCAR)	0.9424×1.25	26
CESM1-CAM5	NCAR	0.9424×1.25	26
CMCC-CM	Euro-Mediterranean Center on Climate Change	0.7484×0.75	31
CNRM-CM5	Centre National de Recherches Meteorologiques	1.4008×1.40625	31
CSIRO-Mk3-6-0	CSIRO in collaboration with the Queensland Climate Change Centre of Excellence	1.8653×1.875	18
FGOALS-g2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences; and CESS, Tsinghua University	2.7906×2.8125	26
GFDL-CM3	Geophysical Fluid Dynamics Laboratory	2×2.5	48
GISS-E2-R	NASA Goddard Institute for Space Studies	2×2.5	40
HadGEM2-AO	Met Office Hadley Centre	1.25×1.875	38
IPSL-CM5A-LR	Institut Pierre-Simon Laplace (IPSL)	1.8947×3.75	39
IPSL-CM5A-MR	IPSL	1.2676×2.5	39
IPSL-CM5B-LR	IPSL	1.8947×3.75	39
MIROC5	University of Tokyo, National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	1.4008×1.40625	40
MPI-ESM-LR	Max Planck Institute for Meteorology (MPI-M)	1.8653×1.875	47
MPI-ESM-MR	MPI-M	1.8653×1.875	95
MRI-CGCM3	Meteorological Research Institute	1.12148×1.125	48
NorESM1-M	Norwegian Climate Centre	1.8947×2.5	26
bcc-csm1-1	Beijing Climate Center, China Meteorological Administration (BCC)	2.7906×2.8125	26
bcc-csm1-1-m	BCC	2.7906×2.8125	26
inmcm4	Institute for Numerical Mathematics	1.5×2	21

Note: HadGEM2-AO is listed here and is used as the reference throughout, but for the AMIP simulation the HadGEM2-a model is used. The same convention has been used for CanAM4 which is an AMIP configuration, but both coupled and AMIP versions are labelled as CanCM4 in this study.

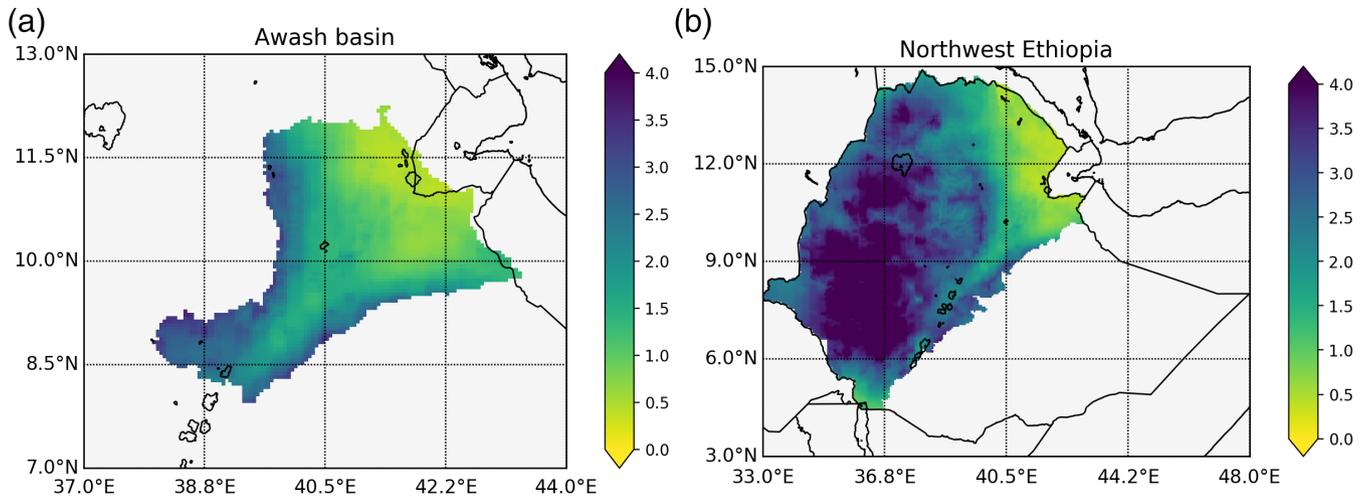


FIGURE 1 Annual average rainfall (mm-day) in the (a) Awash basin and (b) Northwest Ethiopia from CHIRPS (1981–2005) [Colour figure can be viewed at wileyonlinelibrary.com]

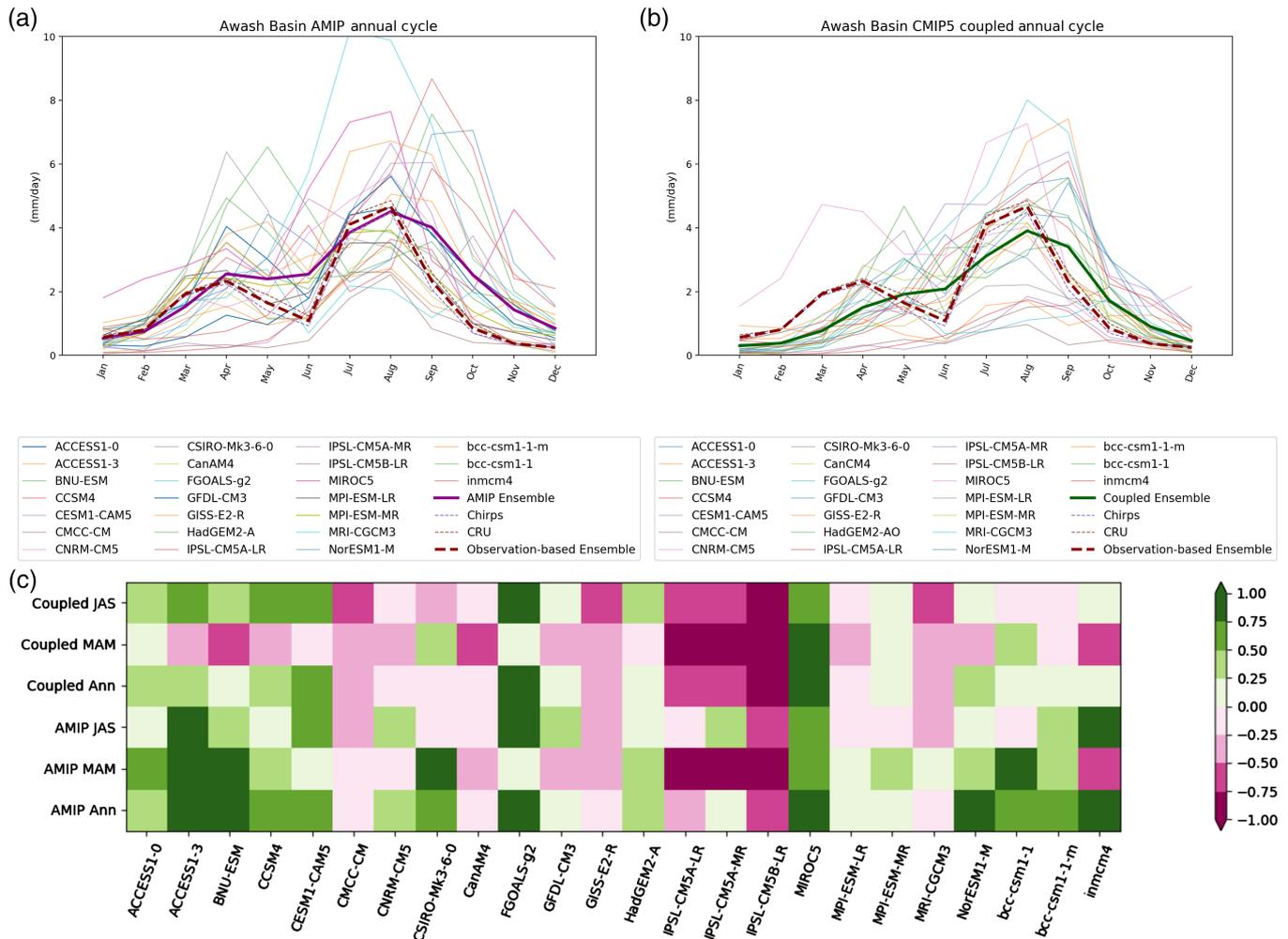


FIGURE 2 Awash basin annual rainfall cycles for (a) AMIP simulations and (b) coupled simulations from CMIP5 with observations. Averages of 1981–2005 rain rates (mm-day^{-1}), averaged over the masked Awash basin region shown in Figure 1a. Model annual cycles are solid lines, while observed annual cycles are dashed lines. (c) Biases in AMIP and coupled simulation rainfall in key rainy seasons and annual average rainfall in the Awash basin. The relative differences ($[\text{Model}-\text{Chirps}]/[\text{Chirps}]$) between model and Chirps rainfall are shown as either drier negative grid spaces or wetter positive grids. Zero indicates perfect agreement. The bottom three rows show atmospheric simulation biases while the upper three rows show coupled model biases [Colour figure can be viewed at wileyonlinelibrary.com]

seasonality of the annual cycles in the observed datasets are quite different in the two regions.

The difference between the Northwest Ethiopian AMIP (Figure 3a) and coupled (Figure 3b) simulations highlights the problematic simulation of MAM rainfall. In AMIP simulations, some models vastly overestimate rainfall and have a distinct MAM or MAMJ season and the AMIP ensemble average overestimates rainfall in this season. The coupled model ensemble average has models which both overestimate and underestimate April–May rainfall. Some miss this onset season entirely, and others have a distinctly bimodal annual cycle leading to the ensemble average matching the observations well in this season.

Unfortunately, this means that nearly 40% of AMIP model simulations have a distinctly bimodal rainfall distribution. For the majority models (six out of nine) with this bimodal distribution, the driest summer month is July which should be one of the wettest months of the year. BNU-ESM, CCSM4, and bcc-csm1-1-m have improved coupled simulations but a dry July is still a feature in CESM-CAM5, NorESM1-M, and bcc-csm1-1. The coupled ensemble has less of a MAM bias and less of an October–November bias but JJAS rain rates are too low.

In selecting skilful models with the annual cycle as a metric, biases in rainfall are not prioritized as much as the annual cycle reproducing the differentiation between the onset and main rainy season, and the relative rainfall between the two seasons, and in particular the existence of a MAM season. A dry June was also a criterion for the Awash basin. These criteria were applied to both the AMIP and coupled simulation of each model; a model which meets the criteria in the AMIP run but does not in the coupled run is not selected and vice versa. The models which perform the best in this category are: MPI-ESM-LR and MPI-ESM-MR, with GFDL-CM3, with HadGEM2-AO also simulating the seasonality of both regions. In the Awash basin ACCESS1-0 also performs well, while in Northwest Ethiopia CMCC-CM preserves the relative rainfall levels in its annual cycle in both simulations.

4.1.2 | Rainfall: Regional biases

Rainy seasons identified from the observed climatologies shown above, along with annual average biases are examined in Figures 2c and 3c for the Awash basin and Northwest Ethiopia, respectively. From the annual cycles we have defined the Belg rainy season as MAM for the Awash basin, and AM in Northwest Ethiopia. We have defined the Kiremt rainy season as JAS in the Awash basin and JJAS in Northwest Ethiopia. Annual average biases are also shown in Figures 2c and 3c, but dry season biases are not included.

In the Awash basin average biases, there is a consistent dry bias in coupled MAM while both wet and dry biases are apparent in AMIP MAM. Most models have a drier coupled MAM than AMIP MAM, with some showing little change, generally those that are already dry. GFDL-CM3 becomes slightly wetter having an AMIP relative difference of -0.49 to the CHIRPS rainfall rate, compared to the coupled relative difference of -0.43 . Otherwise, the only model that has a wetter coupled MAM than AMIP MAM is MIROC5 having an AMIP relative difference of -0.55 and a coupled difference of 1.25 . Applying a threshold of relative difference between -0.50 and 0.50 for annual and seasonal biases in the AMIP and coupled simulations only five models are accepted: CNRM-CM5, GFDL-CM3, HadGEM2-AO, MPI-ESM-LR, and MPI-ESM-MR.

Conversely, for Northwest Ethiopia, there are systematic wet biases in the April–May (AM) season for AMIP simulations which are reduced in the coupled simulations. In AMIP simulations 14 out of 24 models are wetter with a relative difference over 0.30 with 7 models having a relative difference between 0.90 and 1.50 . In coupled simulations seven out of 24 models are wetter with a relative difference over 0.30 with two models having relative differences between 0.90 and 1.50 . This is a reiteration of the overestimation in the onset rainy season shown in the annual cycle. Interestingly models with dry AMIP AM still have dry coupled AM, but the magnitude of the dry biases does not become consistently drier. This is even the case for models with relatively small biases, such as *inmcm4*. This is the case because referring back to the climatologies, the main decrease in rainfall between AMIP and coupled onset seasons is during March, rather than April or May. It is also the case that the JJAS season tends to be relatively dry in many models in both AMIP and coupled simulations, due in part to the dry July bias. Some models stand out for their particularly strong biases: MIROC5 is too wet in all selected timeframes, IPSL-CM5B-LR is too dry in AMIP AM and gets drier in coupled simulations, and finally CSIRO-Mk3-6-0 stands out as the AM season has a strong wet bias in both sets of simulations. There are however a number of models with very small biases. These are CMCC-CM, GFDL-CM3, HadGEM2-AO, IPSL-CM5A-MR, MPI-ESM-LR, MPI-ESM-MR, MRI-CGCM3, and *inmcm4*. All of these models have a relative difference of between -0.5 and 0.5 compared to CHIRPS averages, except for HadGEM2-AO which has a relative difference of 0.53 in its AMIP AM average.

4.1.3 | Rainfall: Trends

Trend calculations are based on a simple linear regression of total seasonal averages for the period 1981–2005.

Trend comparisons in rainy seasons and the annual averages for both regions are shown in Figure 4, while all observed trends are summarized in Table 2. The goal of this metric is to show which models have spurious trends which might lead to a strong, but unrealistic signal of future change. Observed trends are small and mostly not statistically significant even at a 0.1 level. For CHIRPS, the Awash basin annual average trend is small and positive ($0.004 \text{ mm}\cdot\text{day}^{-1}\cdot\text{year}^{-1}$, $p = .437$), the MAM trend is negative ($-0.021 \text{ mm}\cdot\text{day}^{-1}\cdot\text{year}^{-1}$, $p = .135$), and the JAS trend is positive ($0.033 \text{ mm}\cdot\text{day}^{-1}\cdot\text{year}^{-1}$, $p = .090$), and statistically significant at a 0.1 level. CRU trends are slightly larger than CHIRPS (-0.051 , $p = .007$ in MAM; 0.050 , $p = .020$ in JAS) and are statistically significant at a 0.1 level in both these seasons. For Northwest Ethiopia the trend signs are the same as those for the Awash basin (annual: $0.006 \text{ mm}\cdot\text{day}^{-1}\cdot\text{year}^{-1}$, $p = .218$; AM: $-0.007 \text{ mm}\cdot\text{day}^{-1}\cdot\text{year}^{-1}$, $p = .622$; JJAS: $0.017 \text{ mm}\cdot\text{day}^{-1}\cdot\text{year}^{-1}$, $p = .191$) and none of the trends are statistically significant at the 0.1 level for either CHIRPS or CRU. Due to the fact that there is a disagreement between CHIRPS and CRU, and the fact that most observed trends are not significant, we are highlighting models with trends that are statistically significant and large.

In both regions, models generally experience trends that are too large in the onset season (MAM for the Awash basin and AM for Northwest Ethiopia) in AMIP simulations, 10 for the Awash basin but 16 for Northwest Ethiopia. For the Awash basin five out of these 10 models have the same trend sign as CHIRPS and are therefore exhibiting too much drying. These models are ACCESS1-0, CSIRO-Mk3-6-0, MIROC5, MRI-CGCM3, and bcc-csm1-1-m. CCSM4 is the only model with a strong positive trend towards higher rain rates. For JAS in the Awash basin, where the CHIRPS trend is statistically significant, most AMIP models have a positive trend but most are too weak and not statistically significant at the same level. Models with the largest trends in AM for Northwest Ethiopia are the same as for Awash basin MAM trends, but the agreement in JAS/JJAS for the two regions is quite different. It is harder to comment on the JJAS season for Northwest Ethiopia as the observed trends are not significant.

Coupled trends differ between the two regions for both seasons. Northwest Ethiopia coupled trends are relatively similar in magnitude to the small CHIRPS trends, while in the Awash basin the two rainy seasons almost all have smaller trend magnitudes than CHIRPS. The positive JAS trend in CHIRPS is statistically significant, but this is not the case for all but the HadGEM2-AO model.

Discriminating among models based on trend is difficult, but might benefit from breaking the study regions

down further as shown by Viste *et al.* (2013) and Cheung *et al.* (2008). While such a regional breakdown is beyond the scope of this study, and the coarse resolution of most models, we make a selection based on models which do not have large statistically significant trends when the observed trend is not or are not of a different sign when the observed trends are significant. It must be acknowledged that the period over which we are examining trends is short. However, a similar period has been the subject of a great deal of study in East Africa due to the fact that models have spurious trends (Rowell *et al.*, 2015). To certify that individual extreme years were not influencing our reference trends we applied a bootstrapping method to determine how robust these trends are. Single years of data were removed and the trends recalculated to create a distribution of trends. The standard deviation of all trends was an order of magnitude smaller than the average trend in cases where the original full period trend was statistically significant, indicating that sensitivity to individual years in this short period is not particularly high.

The most weight is placed on the rainy season trends. For the Awash basin these models are CNRM-CM5, GFDL-CM3, IPSL-CM5B-LR, and MPI-ESM-MR. For Northwest Ethiopia, the models that fulfil the criteria are CESM1-CAM5, CNRM-CM5, GFDL-CM3, GISS-E2-R, IPSL-CM5B-LR, and NorESM1-M.

4.1.4 | Rainfall: Variability

Ethiopia has been strongly influenced by rainfall variability, in the form of both droughts and flooding. To assess this aspect of model skill we use Taylor diagrams to examine model standard deviation, and correlation relative to CHIRPS, combined to create a skill score (Figure 5). All months of the year are considered in this metric. We further constrain using standard deviation margins, set at 25% as shown in Figure 5, as we are more concerned with a model's ability to capture the scale of characteristic variations rather than the timing of variations, and correlating coupled model time series with observed time series is not a meaningful comparison due to free running SSTs. Our selection of a root mean square error (RMSE) threshold of 1.5 is arbitrary, and the selection is based more on the positioning of models relative to the model cluster and similar skill in both AMIP and coupled simulations.

The distribution of models in the AMIP and coupled simulations is not dissimilar for the Awash basin. There is a collection of models that fall inside the 1.5 contour for both ensembles, nine AMIP and 10 coupled. In the Awash, it is more common for coupled model simulations to have

lower standard deviations at the same level of correlation as AMIP simulations. For the Awash basin models that fall within the 1.5 contour in both configurations, and fall within the 25% range from the reference standard deviation are CanCM4, HadGEM2-AO, MPI-ESM-LR, and MPI-ESM-MR. For Northwest Ethiopia, there are actually more models which fall within the same skill contour in the coupled simulations (13) than AMIP simulations (11). The spread in standard deviations is higher in AMIP than coupled models as well; the coupled simulations are generally more tightly clustered. Models which are within the 1.5 contour for both simulations and within the 25% standard deviation contours are ACCESS1-0, HadGEM2-AO, MPI-ESM-LR, and MPI-ESM-MR with the coupled simulation of GFDL-CM3 just outside the 1.5 contour. Models that are consistently outliers for both regions are ACCESS1-3, MIROC5, and NorESM1-M.

4.1.5 | Rainfall summary

A summary of the evaluation for CMIP5 rainfall in the Awash basin is shown in Table 3). We combined the evaluation of both AMIP and coupled simulations for each metric. Comparing the difference between the performance of simulations without the effects of a freely running ocean means that the ability of the atmospheric model to reproduce the dynamics of regional climate accurately and the potential effects of SST biases are considered. Only MPI-ESM-MR performs well in all

categories (annual cycle, bias, trend, and variability), but there are models that perform well in three out of four of the categories: GFDL-CM3, HadGEM2-AO, and MPI-ESM-LR. The summary of CMIP5 model skill for Northwest Ethiopia (Table 4) shows that a number of the models that have skill in the most categories are the same as for the Awash Basin. GFDL-CM3 appears in all four categories while HadGEM2-AO and MPI-ESM-MR perform well in three out of four categories.

The annual cycles from the coupled simulations are shown in Figure 6 and show that biases remain even in a reduced ensemble. This highlights the need for future process-based analysis to understand why, for example, March rainfall is so low, and models struggle to reproduce June rainfall. It would be useful, in the future, to examine groups of models which have skill in a number of categories and show weakness in an isolated category, but also examine which models had more skill in the more unimodal Northwest Ethiopia than the distinctly bimodal Awash. Although the full list of skilful models in each rainfall skill category is different, the main group of selected models is the same. This is interesting as the rainfall characteristics in these two regions are quite different. Although part of the Awash basin is contained within Northwest Ethiopia it is a climatically distinct region, evident in the average rainfall shown in Figure 2 and rainfall seasonality. But, it is also on the boundary of the predominantly dry lowland region to the east. The fact that the same group of models is most skilful indicates that it may be large scale drivers of climate that are

Annual cycle	Seasonal bias	Trend	Variability
MPI-ESM-MR	MPI-ESM-MR	MPI-ESM-MR	MPI-ESM-MR
MPI-ESM-LR	MPI-ESM-LR	IPSL-CM5B-LR	MPI-ESM-LR
GFDL-CM3	GFDL-CM3	GFDL-CM3	CanCM4
HadGEM2-AO	HadGEM2-AO	CNRM-CM5	HadGEM2-AO
ACCESS1-0	CNRM-CM5		

Note: Bold models are those selected for three or more criteria.

TABLE 2 Awash basin model selection

Annual cycle	Seasonal bias	Trend	Variability
MPI-ESM-MR	MPI-ESM-MR	IPSL-CM5B-LR	MPI-ESM-MR
MPI-ESM-LR	MPI-ESM-LR	CESM1-CAM5	MPI-ESM-LR
GFDL-CM3	GFDL-CM3	GFDL-CM3	GFDL-CM3
HadGEM2-AO	HadGEM2-AO	NorESM1-M	HadGEM2-AO
CMCC-CM	MRI-CGCM3	CNRM-CM5	ACCESS1-0
	inmcm4	GISS-E2-R	
	CMCC-CM		
	IPSL-CM5A-MR		

Note: Bold models are those selected for three or more criteria.

TABLE 3 Northwest Ethiopia model selection

TABLE 4 Rainfall trends in observations for the awash basin and Northwest Ethiopia

	Awash basin MAM	Awash basin JAS	Awash basin annual
CHIRPS	(−0.021, 0.135)	(0.033, 0.090)	(0.004, 0.437)
CRU	(−0.051, 0.007)	(0.050, 0.020)	(0.005, 0.483)
	NW AM	NW JJAS	NW annual
CHIRPS	(−0.007, 0.610)	(0.017, 0.185)	(0.006, 0.218)
CRU	(−0.031, 0.141)	(0.019, 0.335)	(0.002, 0.775)

Note: The format in each cell is (trend, *p* value). Trends are given in mm-day^{−1}.year^{−1}, and Northwest Ethiopia has been abbreviated to NW.

TABLE 5 Temperature trends in observations for the awash basin and Northwest Ethiopia

	Awash basin MAM	Awash basin JAS	Awash basin annual
CRU	(0.056, 0.000)	(0.016, 0.143)	(0.029, 0.000)
ERA-Interim	(0.051, 0.006)	(0.007, 0.725)	(0.031, 0.001)
NCEP2	(0.018, 0.184)	(0.027, 0.009)	(0.025, 0.002)
	NW AM	NW JJAS	NW annual
CRU	(0.051, 0.000)	(0.019, 0.029)	(0.028, 0.000)
ERA-Interim	(0.039, 0.034)	(0.005, 0.564)	(0.017, 0.020)
NCEP2	(0.067, 0.000)	(0.020, 0.018)	(0.034, 0.000)

Note: The format in each cell is (trend, *p* value). Trends are given in mm-day^{−1}.year^{−1}, and Northwest Ethiopia has been abbreviated to NW.

predominantly affecting biases in the region rather than the ability to represent small scale features.

4.2 | Temperature

4.2.1 | Temperature: Annual cycle

Climatologies for modelled and reanalysis temperatures in the Awash basin are shown in Figure 7 (individual model annual cycles are shown in Figure B1). The warmest month of the year is June (NCEP2 peaks slightly earlier in May), with a secondary maximum in September in ERA-Interim and CRU. The AMIP annual cycle closely follows this seasonality, although May is equally as warm as June, and there is a cool bias from November to March. The coupled model ensemble has a similar annual cycle to the AMIP average, with less of a decrease in temperature between June and September. There is still a cool bias in the coupled model ensemble from November to March. The climatologies of individual models are closer to reanalysis in AMIP rather than coupled configurations. Cold biases from November–February are a common

characteristic of almost all of the coupled model climatologies. Certain models stand out: FGOALS-g2 has a bimodal climatology with a relative minimum during July and August and it is biased cold throughout the year in AMIP and coupled simulations, NorESM1-M which is also biased cold, and inmcm4 which has a unimodal annual cycle and temperatures of close to 15°C in December and January where the reanalysis ensemble is 20°C. Most models are cooler than the ERA-Interim and CRU reference temperature for rainy seasons, in the annual average, and for both the AMIP and coupled simulations. Two models are warmer, especially in the JAS season: GISS-E2-R and CMCC-CM. The coolest models are CCSM4, FGOALS-g2, NorESM1-M, and inmcm4.

There is less temperature variation in the observed annual cycle of Northwest Ethiopia (Figure 8) than there is in the Awash basin with the observation-based ensemble reaching 23°C, while the Awash basin peaks at above 25°C. Models tend to be warmer than the observation-based ensemble (individual model annual cycles are shown in Figure B2), with NCEP2 forming almost a lower bound and reducing the temperature of the observation-based ensemble. The exception to this is inmcm4, a consistent outlier. As Northwest Ethiopia, and the highlands in general, is cooler than eastern Ethiopia and the Awash basin, the models' opposite biases indicate that they may not be able to capture the spatial heterogeneity in regional temperature. There is a cool bias from November to February in the coupled ensemble compared to the data ensemble (with the coupled ensemble being cooler than NCEP2 in January), which is less pronounced in the AMIP ensemble. Although smaller in magnitude than that for the Awash basin, this cool bias is nonetheless a consistent feature.

Biases in the rainy seasons are less consistent in Northwest Ethiopia than they are for the Awash basin, and they are also smaller. The only models that have large biases are inmcm4, which is cool, and GISS-E2-R, which is warm. The shift in biases between AMIP and coupled simulations is neither large nor consistent. Unlike rainfall, there is more of a consistent issue with particular models across simulation type rather than a troubling change, consistent for most models, from AMIP to coupled simulations. Furthermore, the annual cycles have highlighted the discrepancy between NCEP2 and ERA-Interim reanalysis temperatures, with ERA-Interim agreeing much more closely with CRU.

4.2.2 | Temperature: Trends

Trends in temperature tend to be more robust than those in precipitation in models. For this reason, we are using

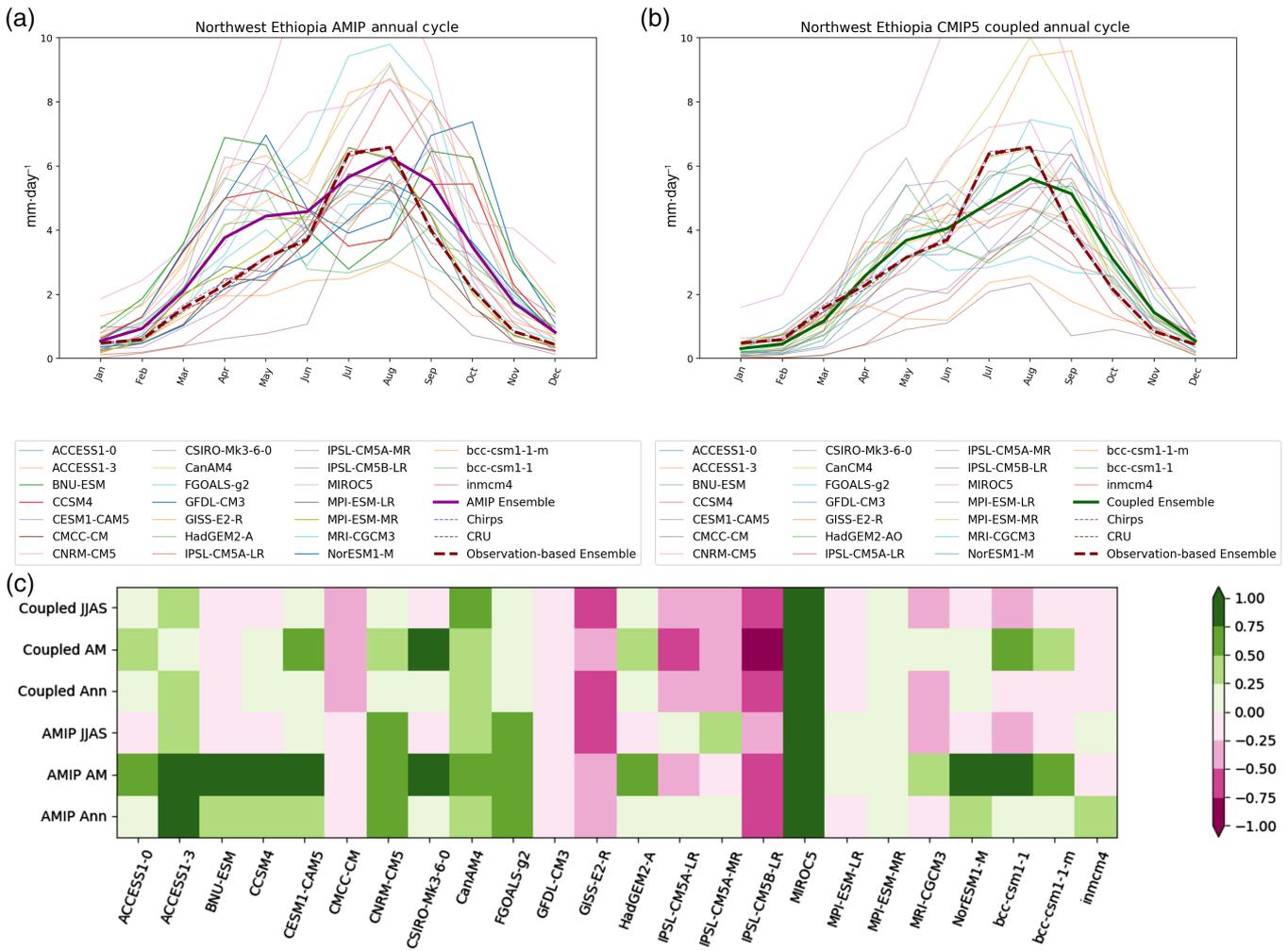


FIGURE 3 Northwest Ethiopia annual rainfall cycles for (a) AMIP simulations and (b) coupled simulations from CMIP5 with observations. Averages are masked as shown in Figure 1b. (c) Biases shown as relative differences ($[\text{Model} - \text{Chirps}] / [\text{Chirps}]$) in AMIP and coupled simulation rainfall in key rainy seasons and annual average rainfall in Northwest Ethiopia, organized as in Figure 2 [Colour figure can be viewed at wileyonlinelibrary.com]

temperature trends to filter out models that are particularly poor at reproducing observed and reanalysis trends. In some cases, trends in observed and reanalysis datasets do not agree, so temperature trends are summarized in Table 5 for reference. The reference dataset used in the relative difference approximations continues to be ERA-Interim. The most important differences to note are that Northwest Ethiopia JJAS in ERA-Interim, and Awash basin MAM and JAS in NCEP2 have different trends from other datasets.

In the Awash basin, ERA-Interim records positive trends for both rainy seasons (MAM: 0.051° per year, $p = .006$; JAS: 0.007° per year, $p = .725$) and the annual mean (0.031° per year, $p = .001$), but only the annual and MAM trends are statistically significant to a 0.1 level. Figure 9a shows most AMIP simulation models do have sign agreement with ERA-Interim trends. Only NorESEM1-M misses agreement in MAM, while CCSM4 has no agreement across seasons. The majority of AMIP

simulations also show statistically significant trends in MAM and annually. Fewer AMIP models show good magnitude agreement with CSIRO-Mk3-6-0, GISS-E2-R, and inmcm4 standing out as having particularly good agreement. It is worth noting here that these were some of the more problematic models in terms of biases. Most models have too weak a trend even when they agree in sign, particularly in MAM. This issue is consistent in the coupled simulations, meaning that models have not captured warming trends in the coupled period. Another issue in the coupled simulations is that over half of the models have significant trends in JAS, but also have the same sign but much larger trend magnitudes than ERA-Interim and AMIP. This would be cause to question future projections of temperature in this area, for the following models in particular: IPSL-CM5a-MR, IPSL-CM5a-LR, MPI-ESM-LR, MPI-ESM-MR, and bcc-csm-1-m.

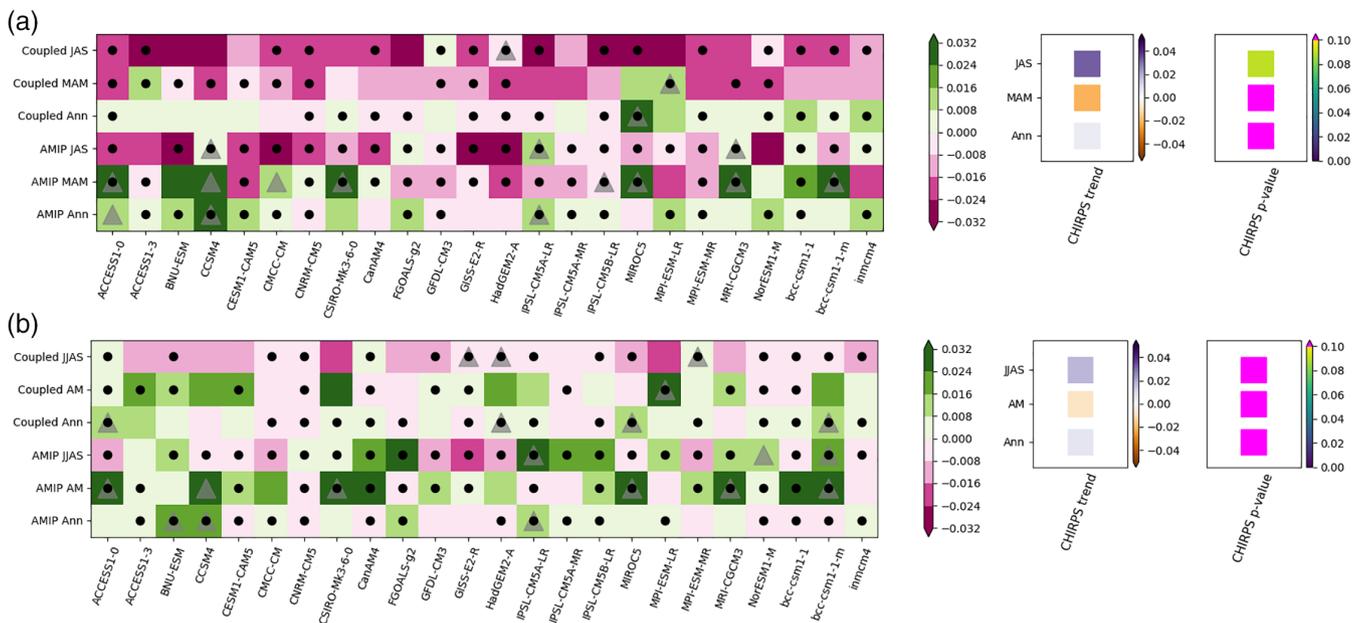


FIGURE 4 Differences in trend magnitude ($\text{mm}\cdot\text{day}^{-1}\cdot\text{year}^{-1}$) and direction relative to the Chirps dataset in (a) Awash basin and (b) Northwest Ethiopia. Grids with a black dot indicate sign agreement between model and Chirps trend direction. Negative grids represent a trend magnitude that is smaller than Chirps, while positive grids indicate a modelled trend magnitude that is larger than Chirps. A grid with value zero indicates perfect agreement between model trend and Chirps trend. Grey triangles represent modelled trends that are statistically significant at p value of less than 0.1. The bottom three rows show atmospheric simulation trend differences, while the top three rows show coupled simulation differences. Reference trends from Chirps for the Awash basin: annual, $0.004 \text{ mm}\cdot\text{day}^{-1}\cdot\text{year}^{-1}$, $p = .437$; MAM, $-0.021 \text{ mm}\cdot\text{day}^{-1}\cdot\text{year}^{-1}$, $p = .135$; JAS, $0.033 \text{ mm}\cdot\text{day}^{-1}\cdot\text{year}^{-1}$, $p = .090$. Reference trends from Chirps for Northwest Ethiopia: annual, $0.006 \text{ mm}\cdot\text{day}^{-1}\cdot\text{year}^{-1}$, $p = .218$; AM, $-0.007 \text{ mm}\cdot\text{day}^{-1}\cdot\text{year}^{-1}$, $p = .622$; JJAS, $0.017 \text{ mm}\cdot\text{day}^{-1}\cdot\text{year}^{-1}$, $p = .191$ [Colour figure can be viewed at wileyonlinelibrary.com]

ERA-Interim temperature trends are also positive in Northwest Ethiopia (Figure 9b), and again the main rainy season, JJAS, has a trend of 0.005° per year ($p = .564$), which is not significant at the 0.1 level while the annual and onset season, AM, trends are both positive, 0.017° per year ($p = .020$) and 0.039° per year ($p = .034$), respectively, which are significant. However, CRU and NCEP2 trends in this season are larger and statistically significant. The main difference between Figure 9a and b is the disagreement in trend magnitude. AMIP annual and AM trends in Northwest Ethiopia have magnitude disagreements that are smaller than they are for the Awash basin. In the coupled simulations the annual trends tend to be too large, while the AM trends are generally too small, although this is less distinct than for MAM in the Awash basin. Coupled model trends for JJAS in Northwest Ethiopia and JAS in the Awash basin are similar in that almost all models have a significant trend of the same sign as the reference with a magnitude that is too big. This is particularly so for IPSL-CM-5A-LR/MR, MPI-ESM-LR, and bcc-csm1-1-m.

In both regions, the onset season is that which experiences the strongest warming trends in observations and reanalysis. ERA-Interim has statistically significant warming trends for both regions, and the Awash basin

average MAM temperature increase has been 1.275°C , and for Northwest Ethiopia the AM increase has been 0.975°C over the historical period from 1981 to 2005. The baseline onset temperature is warmer in the Awash than it is in Northwest Ethiopia, and is experiencing stronger warming trends. This season precedes the driest month of the year, June, and the underestimation of this trend in coupled models is, therefore, a concerning characteristic of the ensemble.

4.2.3 | Temperature: Variability

Unlike the case of rainfall, models cluster quite closely in the Taylor diagrams of temperature, shown in Figure 10a,b, for both AMIP and coupled simulations for the Awash basin. The AMIP cluster tends to have a higher standard deviation than the reference and this is more extreme in the coupled model diagram. Correlations are generally high for AMIP simulations. Again, while the exact skill score is not used as a threshold, they are used to compare distributions with more weight being put on standard deviation. For AMIP, 18 models fall within the 1.6 score, while for coupled models, this has been reduced to 12 models for the Awash basin. All of the

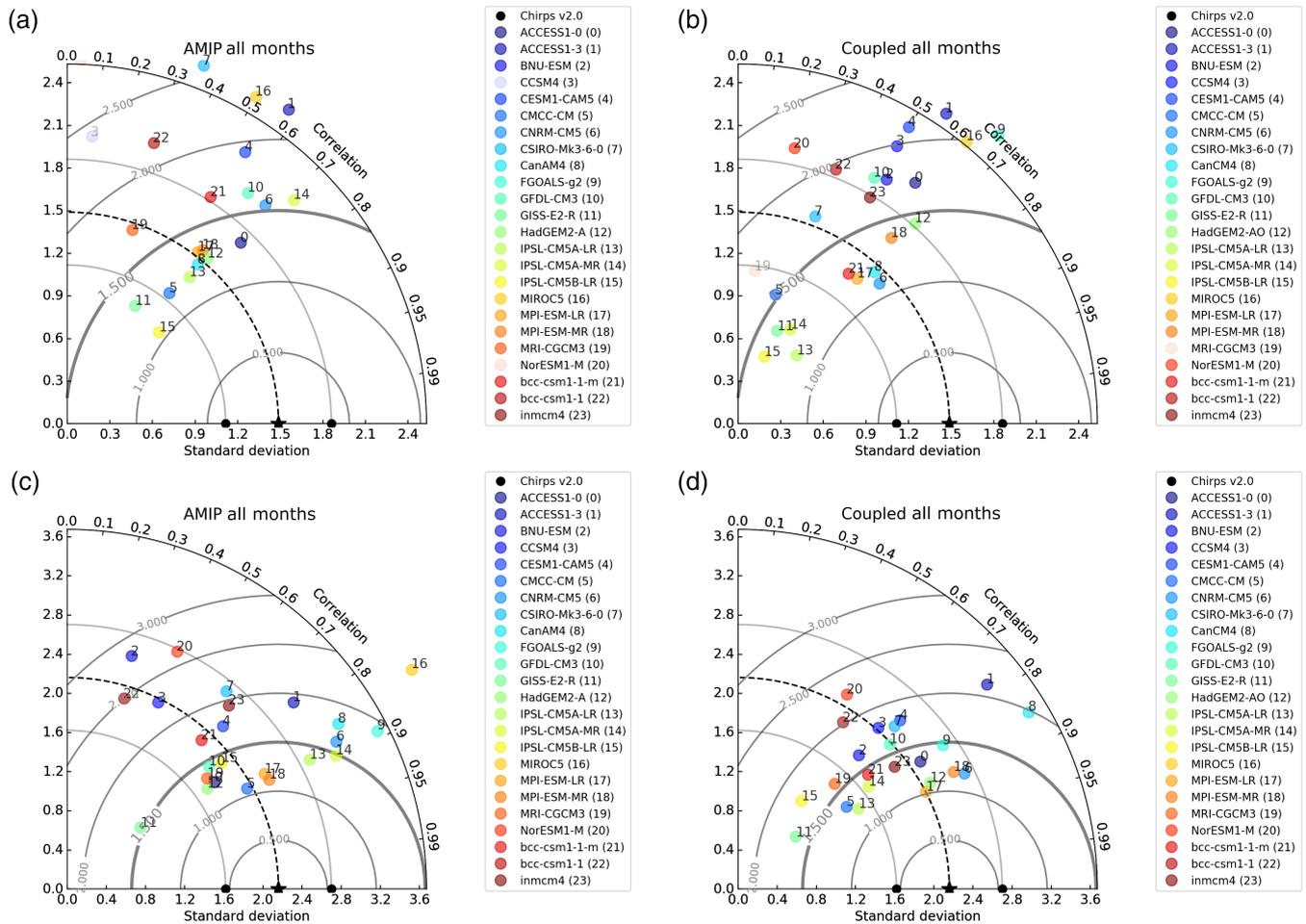


FIGURE 5 Taylor diagrams of rainfall in all months of the year in the Awash basin (a and b) and Northwest Ethiopia (c and d) AMIP simulations, and coupled simulations. The reference time series for the masked region is CHIRPS. Points that are translucent on the diagrams indicate correlations with are not significant to a 95% level. Standard deviation increases in the radial direction with the reference (Chirps) value denoted with a star, while correlations vary azimuthally, decreasing from 0° to 90°. Skill score contours are calculated constant root mean square error (RMSE) which relates the sample standard deviation, reference standard deviation, and sample correlation as in Taylor *et al.* (2012). Models with lower RMSE have higher skill. Standard deviations above and below 25% of the reference value are shown as translucent grey lines, and the RMSE threshold of 1.5 is shown as a bold contour [Colour figure can be viewed at wileyonlinelibrary.com]

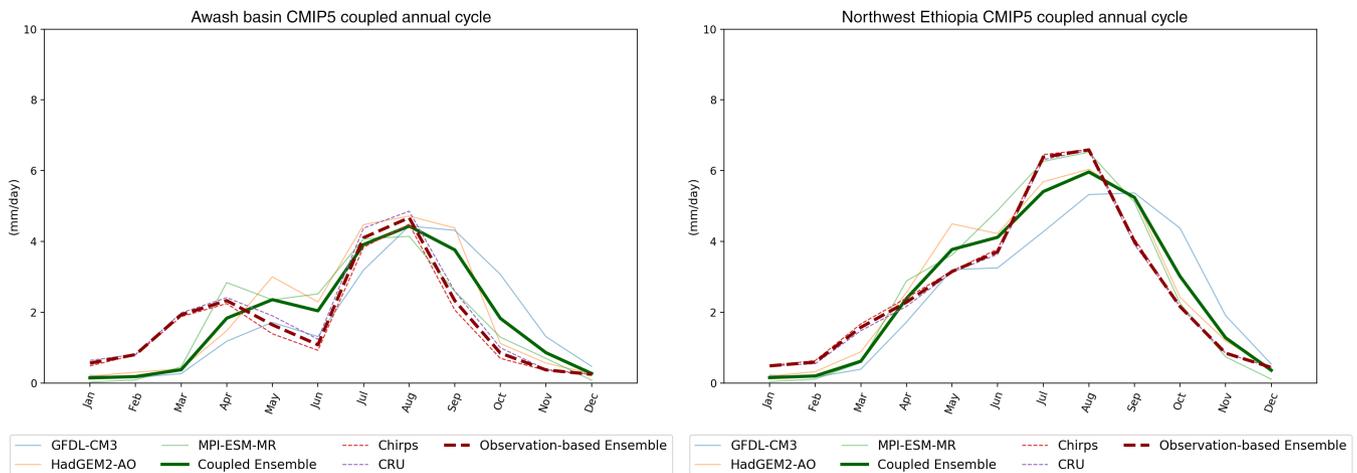


FIGURE 6 Annual rainfall cycles for the coupled simulations of models selected as a potential reduced ensemble as highlighted in Tables 3 and 4. Awash basin annual cycles are shown in (a) and Northwest Ethiopia are shown in (b) [Colour figure can be viewed at wileyonlinelibrary.com]

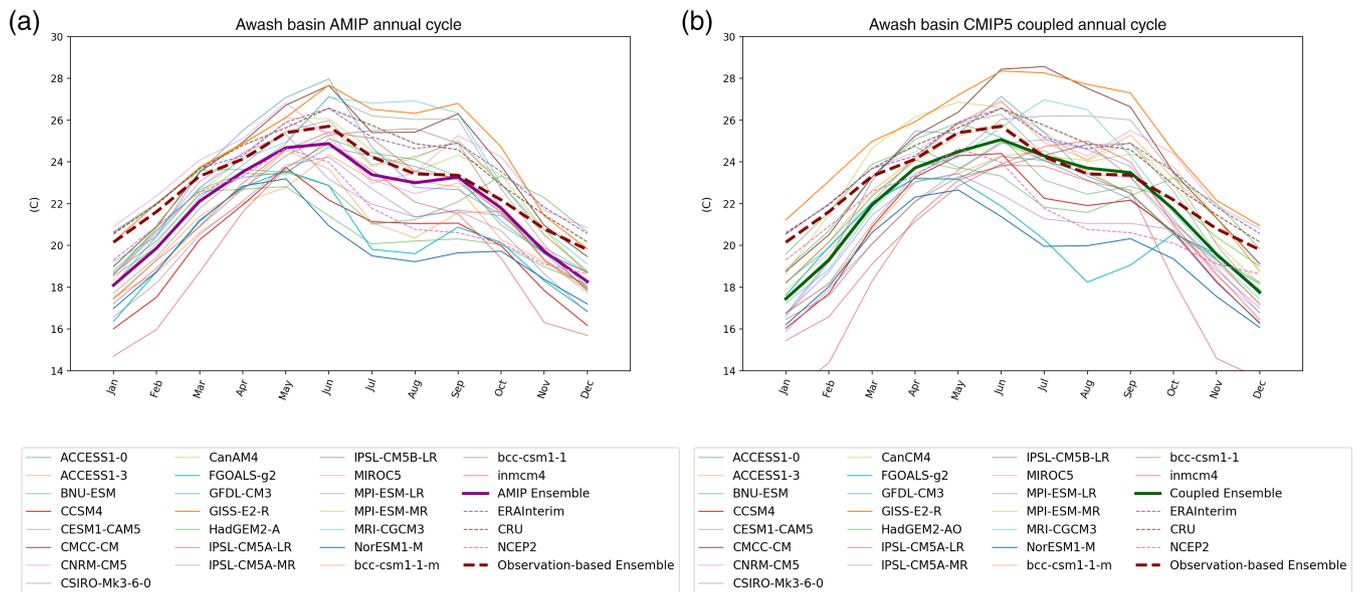


FIGURE 7 Awash Basin annual temperature cycles for a) AMIP simulations, and b) coupled simulations from CMIP5 with observation-based data. Averages of 1981–2005 surface temperature ($^{\circ}\text{C}$), averaged over the masked Awash Basin region. Model annual cycles are solid lines, while reanalysis and observation annual cycles are dashed lines [Colour figure can be viewed at wileyonlinelibrary.com]

coupled models within this score were part of the AMIP group too, except for BNU-ESM which was outside the 1.6 score on the AMIP diagram. In the Northwest Ethiopia diagrams (Figure 10c,d) correlations are lower than for the Awash basin for both model configuration groups, but have a similar tendency towards higher standard deviations than the reference. In fact, all models have equal or greater standard deviations than ERA-Interim. For this case we compare models that fall within the 1.2 skill score. For AMIP simulations, there are 19 models, while for the coupled simulation models there are 12. All 12 coupled models were in the group of 19 AMIP models to fall within this score.

There are a number of models which are outliers in both regions. The GFDL-CM3 model has much higher standard deviation than the reference in Northwest Ethiopia, and is one of the outliers in the Awash basin too. The inmcm4 model is off the range of the plots for all of the Taylor diagrams except for AMIP in the Awash basin. Finally the bcc-csm1-1 model has consistently low correlations and in the Awash basin is one of the only models to have a standard deviation lower than the reference.

4.2.4 | Temperature summary

A selection of models using the application of temperature criteria will not be made as models generally reproduce the temperature characteristics of the chosen regions well. There is also some variation between the reanalysis and observational temperatures, which makes

selecting models more difficult. For instance, models are bounded by the bias between NCEP and ERA-Interim reanalysis temperatures in the annual cycle except for the boreal winter months in some models, and there is also some disagreement between which trends are significant (although there is no sign discrepancy) especially in JJAS for Northwest Ethiopia.

There are, however, some interesting caveats to be added to the list of models selected based on rainfall criteria for each region based on the information in the temperature evaluations. For instance, MPI-ESM-LR has some problematic temperature trends. Therefore, in making the choice between the two models from the same institution it would be best to use MPI-ESM-MR instead. The GFDL-CM3 model may have some issues with representing temperature variability in the study regions, and while it should still be regarded as one of the better performing models for rainfall we might exercise some caution in using it to diagnose changes in extremes in both future rainfall and temperature. However, many of the models that performed poorly in the temperature evaluations were also among those that performed poorly in the rainfall evaluations, especially in the annual cycle and variability categories.

5 | DISCUSSION AND CONCLUSIONS

Unless Ethiopia is the focus for a project, anything but the eastern part of the country is usually excluded from

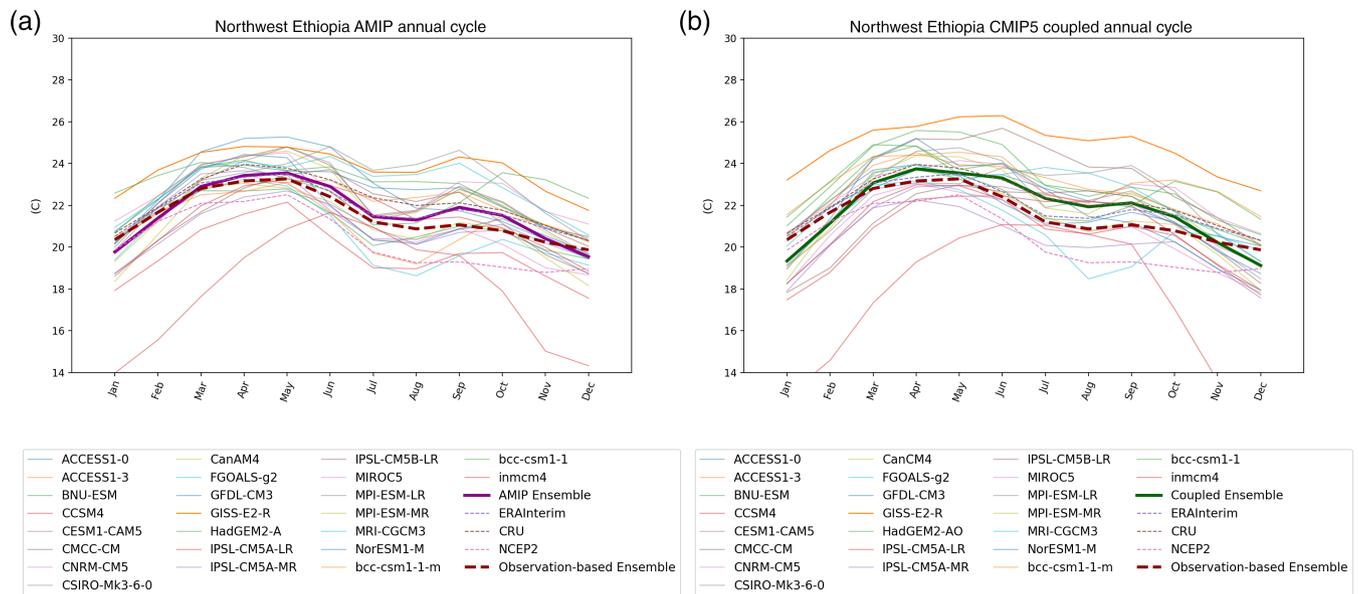


FIGURE 8 Northwest Ethiopia annual temperature cycles for a) AMIP simulations, and b) coupled simulations from CMIP5 with observation-based data. Averages of 1981–2005 surface temperature ($^{\circ}\text{C}$), averaged over the masked Northwest Ethiopia region. Model annual cycles are solid lines, while reanalysis and observation annual cycles are dashed lines [Colour figure can be viewed at wileyonlinelibrary.com]

large model ensemble comparison studies of East Africa, and it is not generally included in studies of the Sahel at the same latitude. This evaluation of rainfall and temperature is a starting point in refining which models in the CMIP5 ensemble may be good candidates to use in examining future climate in the Awash basin and Northwest Ethiopia, and to build future process-based evaluations on. The CMIP model ensemble grows with each IPCC iteration. While the use of the ensemble mean to summarize the model set is convenient, the approach masks the contribution of models with poor skill.

There were some key differences in the model representations of the rainfall in the Awash basin and Northwest Ethiopia. First, in the Awash basin, the onset season was generally too dry in coupled models and less so in AMIP models, while in Northwest Ethiopia the onset season bias was largest in the AMIP ensemble, being too wet. The annual cycle of the Awash basin rainfall was problematic in coupled models because it was shifted to be later in the year than in observation, while in most AMIP models it was too bimodal with the second rainy season occurring later than in observations. In Northwest Ethiopia, the annual cycle was bimodal for both model ensembles even though the observations show a more unimodal annual cycle with a main JJAS rainy season and an AM onset season.

Some of the characteristics of these regional differences and those between the AMIP and coupled ensembles are also present in studies of East Africa, mostly

revolving around model representation of the onset rainy season (AM or MAM). For instance, Rowell *et al.* (2015) showed that coupled CMIP5 have more trouble reproducing trends in this season than AMIP models, while Dunning *et al.* (2017) showed that CMIP5 models simulate late onset in this season. For this reason the transitions from the biases in the Awash basin to the biases in Northwest Ethiopia have proved interesting. The underestimation of coupled CMIP5 rainfall in the MAM rainy season compared to the overestimation in the later OND rainy season has been documented in studies on East Africa (Tierney *et al.*, 2015). This underestimation of MAM rainfall in coupled CMIP5 models is present in the Awash basin too, but as previously discussed, the AMIP representation of MAM rainfall in most models also is not accurate. The seasonality which makes AMIP models better able to capture East African MAM rainfall persists in Northwest Ethiopia; many models have bimodal annual cycles with wet AM biases, and distinct MAM and OND rainfall seasons in some cases, resulting in an ensemble annual cycle which is approaching bimodal. This seasonality is less clear in the coupled ensemble, but this is in part because March rainfall is underestimated in numerous models, while April and May still tend to be overestimated. This is also a holdover of coupled model behaviour of a late long rains season in East Africa more broadly, and contributes to the missing dry June in models in the Awash basin. The feature of a dry June is an important feature in the annual cycle and an

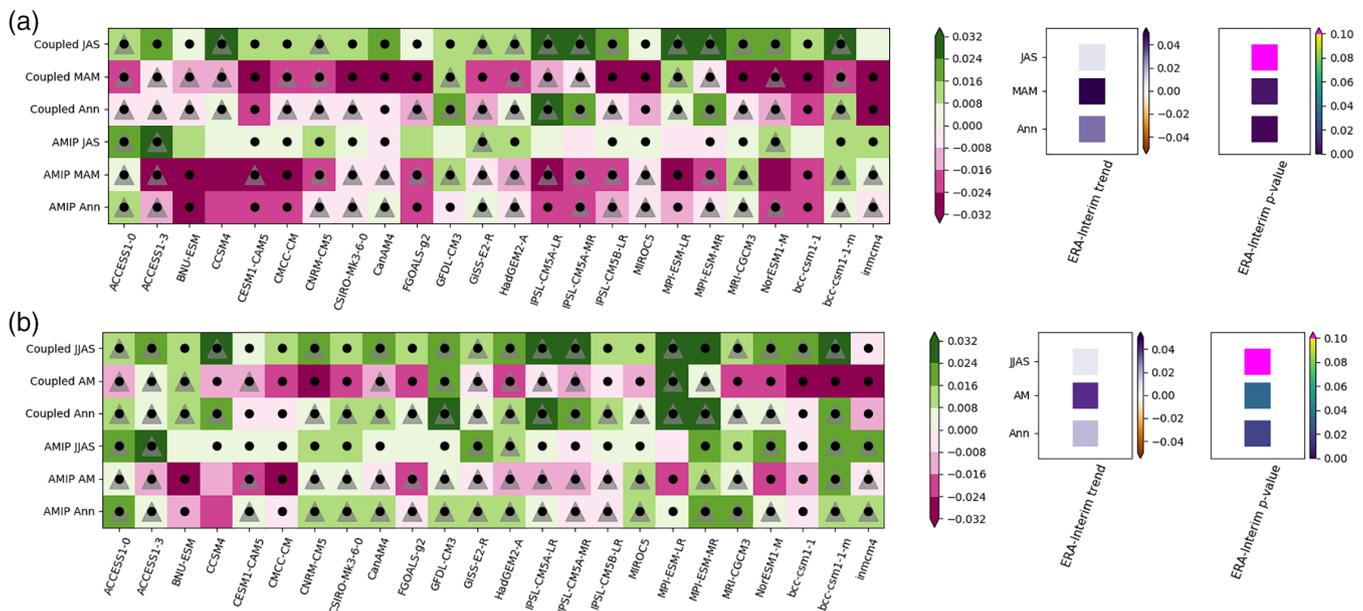


FIGURE 9 Differences in trend magnitude ($^{\circ}\text{C}\cdot\text{year}^{-1}$) and direction relative to ERA-interim temperatures in (a) Awash Basin and (b) Northwest Ethiopia. Grids with a black dot indicate sign agreement between model and ERA-Interim trend direction. negative grids represent a trend magnitude that is smaller than ERA-Interim, while positive grids indicate a modelled trend magnitude that is larger than ERA-Interim. A grid with value zero indicates perfect agreement between model trend and ERA-Interim trend. Grey triangles represent modelled trends that are statistically significant at p value of less than 0.1. The bottom three rows show atmospheric simulation trend differences, while the top three rows show coupled simulation differences. Reference trends from ERA-Interim for the Awash basin: Annual, 0.031° per year, $p = .001$; MAM, 0.051° per year, $p = .006$; JAS, 0.007° per year, $p = .725$. Reference trends from ERA-Interim for Northwest Ethiopia: annual, 0.017° per year, $p = .020$; AM, 0.039° per year, $p = .034$; JJAS, 0.005° per year, $p = .564$ [Colour figure can be viewed at wileyonlinelibrary.com]

important feature for models to capture if being used to investigate future water stress in the region.

The biases in these regions of Ethiopia, the missing June minimum in the Awash basin and the persistent bimodal cycle in Northwest Ethiopia, indicate that the ability of AMIP simulations to reproduce the East African MAM season is due to a bias during that season. The distinctly bimodal annual cycle and high rain rates in MAM persist north of the region for which the MAM peak is found in the observation based datasets. These connections and biases should be explored further, as they may be able to shed some light on biases in East Africa as a whole and not just in these areas of Ethiopia.

There were some unexpected findings, in particular for temperature. First, the cold bias from November to February was a consistent feature in both parts of Ethiopia, both simulation configurations, and for the majority of models. This bias will be essential to quantify from a policy perspective. These months are some of the driest during the year. The compounding effects of dryness and higher temperatures are part of what can cause the accumulating feedbacks to drought events which have hit Ethiopia in the past and potentially into the future. Models that cannot capture higher temperatures in the

dry season may not be useful tools in predicting such events. Like the cold biases, temperature trends have been less widely studied in models but would have important implications for the use of future projections. We found that the strongest warming trend in both regions was in the onset season (AM or MAM), the warmest part of the year, and that the Awash basin had stronger warming trends in the annual and seasonal averages. Coupled models were warming too much in the primary rainy season, while all models were not warming enough in the onset rainy season.

Annual trends in models also tended to overestimate reanalysis trends. Jury and Funk (2013) examined annual temperature trends for the whole of Ethiopia from 1948 to 2006 and found a similar warming trend to our Awash basin trend (0.031° per year) and also that this warming trend was set to continue in the GFDL simulation they used. Given the biased trends of this model and the coupled ensemble in general, this future trend might be too large. The impact of trends in temperature is compounded by issues highlighted in the rainfall trends. The onset rainy season in both regions has lower rain rates, and higher temperatures, than the main rainy season. This season is still used as a planting and growing season

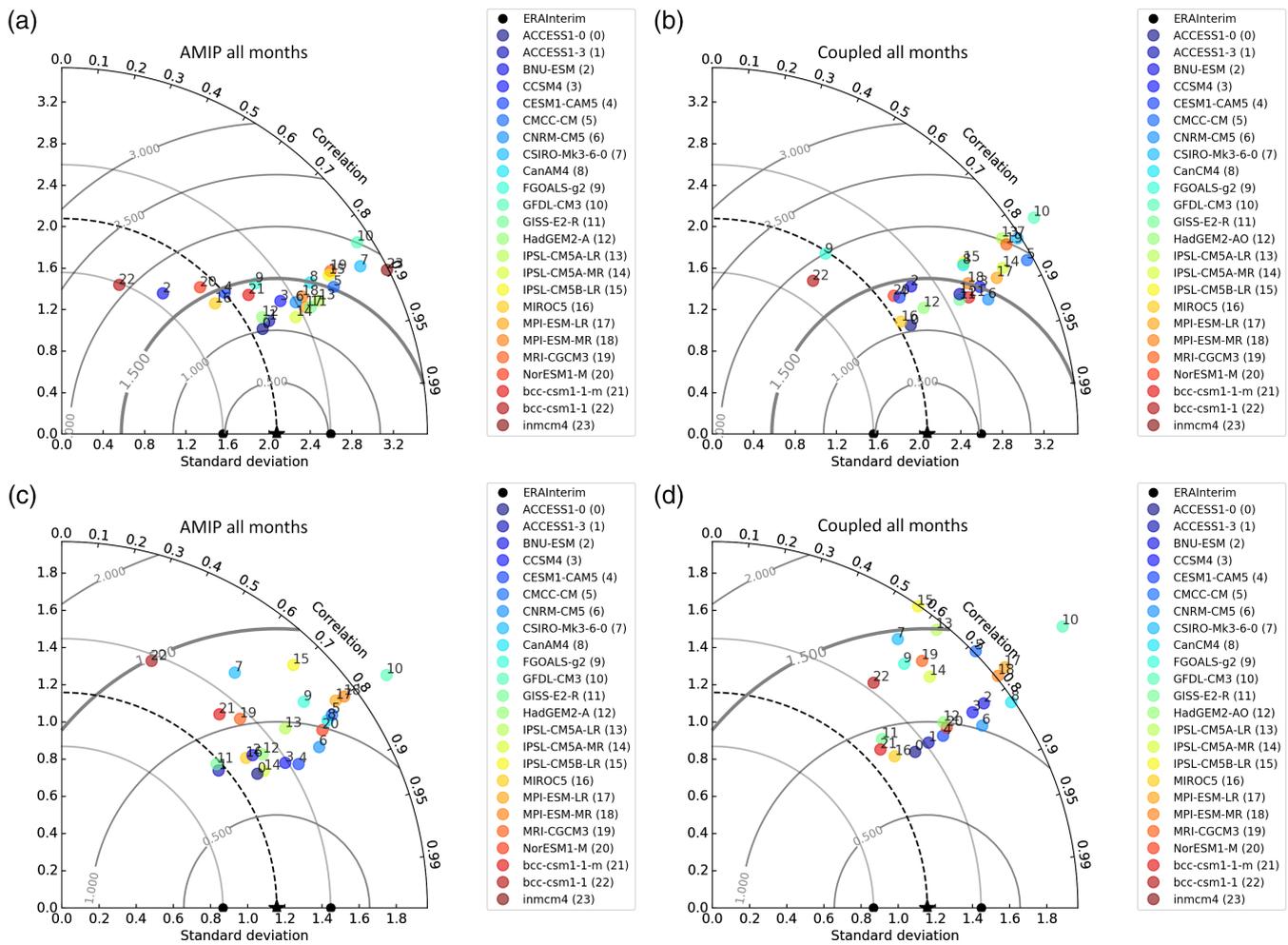


FIGURE 10 Taylor diagrams of temperature in all months of the year in the awash basin (a and b) and Northwest Ethiopia (c and d) AMIP simulations, and coupled simulations. The reference time series for the masked region is ERA-Interim. Points that are translucent on the diagrams indicate correlations with are not significant to a 95% level. Standard deviation increases in the radial direction with the reference (ERA-Interim) value denoted with a star, while correlations vary azimuthally, decreasing from 0° to 90°. Skill score contours are calculated constant root mean square error (RMSE) which relates the sample standard deviation, reference standard deviation, and sample correlation as in Taylor *et al.* (2012). Models with lower RMSE have higher skill. Standard deviations above and below 25% of the reference value are shown as translucent grey lines, and the RMSE threshold of 1.5 is shown as a bold contour [Colour figure can be viewed at wileyonlinelibrary.com]

for specific crops and rainfall is correlated with the success of certain crops in this season (Borgomeo *et al.*, 2018). Changes in the onset rainy season could have dramatic effects on the ability to grow certain crops or require a significant adaptation effort. Biases in rainfall can also influence volumetric flows and flooding with biases in June and October changing the total volume of the Kiremt season indicating more flooding problems than what is presently experienced by the basin giving a false alarm of flooding risk. Similarly, the bias in March gives false indication that water will not be a limiting factor during the Belg season, which is usually affected by drought conditions.

We have identified three models to be used in a process-based analysis for our two study regions: GFDL-

CM3, HadGEM2-AO, and MPI-ESM-MR. These models showed most skill in both AMIP and coupled simulations for rainfall. Most performed well in the temperature evaluation, with a caveat for GFDL-CM3 temperature variability. The MPI-ESM models were also found to perform well over the Greater Horn of Africa by Otieno and Anyah (2013), who also argued for evaluation of model uncertainties in the creation of reduced ensembles, and studies of future change. Our reduced ensemble differs in that it does not include MRI-CGCM3, and does include GFDL-CM3 and HadGEM2-AO, highlighting differences at the regional and sub-regional scale, and the importance of carrying out this kind of evaluation for the region in which modelled information will be used.

Investigating which atmospheric processes affect biases is a natural next step. Beyond this, the relationship between Ethiopian rainfall and regional SSTs needs to be better understood. This would build upon previous work by Degefu *et al.* (2017), Li *et al.* (2016) and Diro *et al.* (2011) but incorporate more models and focus on understanding the dynamics of biases in the onset season along with as the main rainy season. The interactions with surrounding African climate regimes, such as the Sahel band, and East and Central Africa have also been alluded to (Viste *et al.*, 2013), but requires further investigation with relation to GCMs. This is especially important as models are not able to distinguish these regions very distinctly, causing some of the model biases. Finally, the effects of some of the large scale overturning circulations on Ethiopian rainfall may highlight some potential causes of bias in models. In particular investigating the impact of the ability of models to capture the Asian Monsoon circulation on rainfall in Ethiopia from June–September building on the connection between the Tropical Easterly Jet and Ethiopian rainfall shown by Li *et al.* (2016) and on the work of Sperber *et al.* (2013) who did a detailed analysis of the skill of CMIP3/5 models in reproducing the Asian Monsoon.

This study is the necessary first part of a multi-phase process of model evaluation which will not only lead to better constrained projections, but also help to contextualize our understanding and use of these projections. Of course, the ability of models to accurately simulate future climate cannot be diagnosed purely by looking at historical simulations, but it does provide a good baseline for examining model behaviour. Being able to quantify and explain biases and uncertainty, and clearly map connections between regional climates is an important part of developing better modelling tools. Following such an approach is also key to building better, sustainable, adaptive capacity.

ACKNOWLEDGEMENTS

This document is an output from the REACH programme funded by UK Aid from the UK Department for International Development (DFID) for the benefit of developing countries (Aries Code 201880). However, the views expressed and information contained in it are not necessarily those of or endorsed by DFID, which can accept no responsibility for such views or information or for any reliance placed on them.

ORCID

Ellen Dyer  <https://orcid.org/0000-0002-9136-8972>

REFERENCES

- AFDB. (2011) *The Cost of Adaptation to Climate Change in Africa, Tech. Rep.* London, UK: Vivid Economics for Africa Development Bank.
- Borgomeo, E., Vadheim, B., Woldeyes, F.B., Alamirew, T., Tamru, S., Charles, K.J., Kebede, S. and Walker, O. (2018) The distributional and multi-Sectoral impacts of rainfall shocks: evidence from computable general equilibrium modelling for the Awash Basin, Ethiopia. *Ecological Economics*, 146, 621–632. <https://doi.org/10.1016/j.ecolecon.2017.11.038>.
- Cheung, W.H., Senay, G.B. and Singh, A. (2008) Trends and spatial distribution of annual and seasonal rainfall in Ethiopia. *International Journal of Climatology*, 28(13), 1723–1734. <https://doi.org/10.1002/joc.1623>.
- Conway, D. and Schipper, E.L.F. (2011) Adaptation to climate change in Africa: challenges and opportunities identified from Ethiopia. *Global Environmental Change*, 21(1), 227–237. <https://doi.org/10.1016/j.gloenvcha.2010.07.013>.
- Cook, K.H. and Vizy, E.K. (2006) Coupled model simulations of the west African monsoon system: twentieth- and twenty-first-century simulations. *Journal of Climate*, 19(15), 3681–3703. <https://doi.org/10.1175/JCLI3814.1>.
- Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N. and Vitart, F. (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553–597. <https://doi.org/10.1002/qj.828>.
- Degefu, M.A., Rowell, D.P. and Bewket, W. (2017) Teleconnections between Ethiopian rainfall variability and global SSTs: observations and methods for model evaluation. *Meteorology and Atmospheric Physics*, 129(2), 173–186. <https://doi.org/10.1007/s00703-016-0466-9>.
- Dieppois, B., Rouault, M. and New, M. (2015) The impact of ENSO on southern African rainfall in CMIP5 ocean atmosphere coupled climate models. *Climate Dynamics*, 45(9–10), 2425–2442. <https://doi.org/10.1007/s00382-015-2480-x>.
- Dinku, T., Funk, C., Peterson, P., Maidment, R., Tadesse, T., Gadain, H. and Ceccato, P. (2018) Validation of the CHIRPS satellite rainfall estimates over eastern Africa. *Quarterly Journal of the Royal Meteorological Society*, 144, 292–312. <https://doi.org/10.1002/qj.3244>.
- Diro, G.T., Grimes, D.I.F. and Black, E. (2011) Teleconnections between Ethiopian summer rainfall and sea surface temperature: part I-observation and modelling. *Climate Dynamics*, 37(1), 103–119. <https://doi.org/10.1007/s00382-010-0837-8>.
- Dunning, C.M., Allan, R.P. and Black, E. (2017) Identification of deficiencies in seasonal rainfall simulated by CMIP5 climate models. *Environmental Research Letters*, 12(11), 114,001. <https://doi.org/10.1088/1748-9326/aa869e>.
- Flohn, H. (1987) Rainfall teleconnections in northern and north-eastern Africa. *Theoretical and Applied Climatology*, 38(4), 191–197. <https://doi.org/10.1007/BF00867412>.
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. and Michaelsen, J. (2015) The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, 2, 150,066. <https://doi.org/10.1038/sdata.2015.66>.

- Grey, D. and Sadoff, C.W. (2007) Sink or swim? Water security for growth and development. *Water Policy*, 9(6), 545–571. <https://doi.org/10.2166/wp.2007.021>.
- Harris, I., Jones, P.D., Osborn, T.J. and Lister, D.H. (2014) Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 dataset. *International Journal of Climatology*, 34(3), 623–642. <https://doi.org/10.1002/joc.3711>.
- Hirons, L. and Turner, A. (2018) The impact of Indian Ocean mean-state biases in climate models on the representation of the east African short rains. *Journal of Climate*, 31(16), 6611–6631. <https://doi.org/10.1175/JCLI-D-17-0804.1>.
- James, R., Washington, R. and Jones, R. (2015) Process-based assessment of an ensemble of climate projections for West Africa. *Journal of Geophysical Research: Atmospheres*, 120(4), 1221–1238. <https://doi.org/10.1002/2014JD022513>.
- Jury, M.R. and Funk, C. (2013) Climatic trends over Ethiopia: regional signals and drivers. *International Journal of Climatology*, 33(8), 1924–1935. <https://doi.org/10.1002/joc.3560>.
- Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.K., Hnilo, J.J., Fiorino, M. and Potter, G.L. (2002) NCEP-DOE AMIP-II reanalysis (R-2). *Bulletin of the American Meteorological Society*, 83(11), 1631–1643. <https://doi.org/10.1175/BAMS-83-11-1631> (2002)083<1631:NAR>2.3.CO;2.
- Lewis, K. (2017) Understanding climate as a driver of food insecurity in Ethiopia. *Climatic Change*, 144(2), 317–328. <https://doi.org/10.1007/s10584-017-2036-7>.
- Li, L., Li, W., Ballard, T., Sun, G. and Jeuland, M. (2016) CMIP5 model simulations of Ethiopian Kiremt-season precipitation: current climate and future changes. *Climate Dynamics*, 46(9–10), 2883–2895. <https://doi.org/10.1007/s00382-015-2737-4>.
- Martin, E.R., Thorncroft, C. and Booth, B.B. (2014) The multidecadal Atlantic SST-Sahel rainfall teleconnection in CMIP5 simulations. *Journal of Climate*, 27(2), 784–806. <https://doi.org/10.1175/JCLI-D-13-00242.1>.
- Monerie, P.A., Sanchez-Gomez, E. and Boé, J. (2017) On the range of future Sahel precipitation projections and the selection of a subsample of CMIP5 models for impact studies. *Climate Dynamics*, 48(7–8), 2751–2770. <https://doi.org/10.1007/s00382-016-3236-y>.
- Munday, C. and Washington, R. (2017) Circulation controls on southern African precipitation in coupled models: the role of the Angola low. *Journal of Geophysical Research*, 122(2), 861–877. <https://doi.org/10.1002/2016JD025736>.
- Munday, C. and Washington, R. (2018) Systematic climate model rainfall biases over southern Africa: links to moisture circulation and topography. *Journal of Climate*, 31(18), 7533–7548. <https://doi.org/10.1175/JCLI-D-18-0008.1>.
- Niang, I., Ruppel, O., Abdrabo, M., Essel, A., Lennard, C., Padgham, J. and Urquhart, P. (2014) Africa. In: *Climate Change 2014: Impacts, Adaptation and Vulnerability - Contributions of the Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY: Cambridge University Press, pp. 1199–1265. <https://doi.org/10.1017/CBO9781107415386.002>.
- Nicholson, S.E. (1986) The spatial coherence of African rainfall anomalies: interhemispheric teleconnections. *Journal of Climate and Applied Meteorology*, 25, 1365–1381. [https://doi.org/10.1175/1520-0450\(1986\)025<1365:TSCOAR>2.0.CO;2](https://doi.org/10.1175/1520-0450(1986)025<1365:TSCOAR>2.0.CO;2).
- Ongoma, V., Chen, H. and Gao, C. (2018) Evaluation of CMIP5 twentieth century rainfall simulation over the equatorial East Africa. *Theoretical Applied Climatology*, 135, 893–910. <https://doi.org/10.1007/s00704-018-2392-x>.
- Otieno, V.O. and Anyah, R.O. (2013) CMIP5 simulated climate conditions of the greater horn of Africa (GHA). Part 1: contemporary climate. *Climate Dynamics*, 41(7–8), 2081–2097. <https://doi.org/10.1007/s00382-012-1549-z>.
- Pohl, B., MacRon, C. and Monerie, P.A. (2017) Fewer rainy days and more extreme rainfall by the end of the century in southern Africa. *Scientific Reports*, 7(1), 46,466. <https://doi.org/10.1038/srep46466>.
- REACH. (2015) *REACH Country Diagnostic Report: Ethiopia, Tech. Rep.* Oxford, UK: University of Oxford.
- Rowell, D.P., Booth, B.B.B., Nicholson, S.E., Good, P., Rowell, D.P., Booth, B.B.B., Nicholson, S.E. and Good, P. (2015) Reconciling past and future rainfall trends over East Africa. *Journal of Climate*, 28(24), 9768–9788. <https://doi.org/10.1175/JCLI-D-15-0140.1>.
- Shanko, D. and Camberlin, P. (1998) The effects of the southwest Indian ocean tropical cyclones on Ethiopian drought. *International Journal of Climatology*, 18(12), 1373–1388. [https://doi.org/10.1002/\(SICI\)1097-0088\(199810\)18:12<1373::AID-JOC313>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1097-0088(199810)18:12<1373::AID-JOC313>3.0.CO;2-K).
- Sperber, K.R., Annamalai, H., Kang, I.S., Kitoh, A., Moise, A., Turner, A., Wang, B. and Zhou, T. (2013) The Asian summer monsoon: an intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century. *Climate Dynamics*, 41(9–10), 2711–2744. <https://doi.org/10.1007/s00382-012-1607-6>.
- Taylor, K.E., Stouffer, R.J. and Meehl, G.A. (2012) An overview of CMIP5 and the experiment design, 93, 485–498. <https://doi.org/10.1175/BAMS-D-11-00094.1>.
- Tierney, J.E., Ummenhofer, C.C. and DeMenocal, P.B. (2015) Past and future rainfall in the horn of Africa. *Science Advances*, 1(9), e1500,682–e1500,682. <https://doi.org/10.1126/sciadv.1500682>.
- Tsidu, G.M. (2012) High-resolution monthly rainfall database for Ethiopia: homogenization, reconstruction, and gridding. *Journal of Climate*, 25(24), 8422–8443. <https://doi.org/10.1175/JCLI-D-12-00027.1>.
- Viste, E., Korecha, D. and Sorteberg, A. (2013) Recent drought and precipitation tendencies in Ethiopia. *Theoretical and Applied Climatology*, 112(3–4), 535–551. <https://doi.org/10.1007/s00704-012-0746-3>.
- Vizy, E.K., Cook, K.H., Crétat, J. and Neupane, N. (2013) Projections of a wetter Sahel in the twenty-first century from global and regional models. *Journal of Climate*, 26(13), 4664–4687. <https://doi.org/10.1175/JCLI-D-12-00533.1>.
- Yang, W., Seager, R., Cane, M.A. and Lyon, B. (2015) The rainfall annual cycle bias over East Africa in CMIP5 coupled climate models. *Journal of Climate*, 28(24), 9789–9802. <https://doi.org/10.1175/JCLI-D-15-0323.1>.

How to cite this article: Dyer E, Washington R, Teferi Taye M. Evaluating the CMIP5 ensemble in Ethiopia: Creating a reduced ensemble for rainfall and temperature in Northwest Ethiopia and the Awash basin. *Int J Climatology*. 2019;1–22. <https://doi.org/10.1002/joc.6377>

APPENDIX A

RAINFALL ANNUAL CYCLES

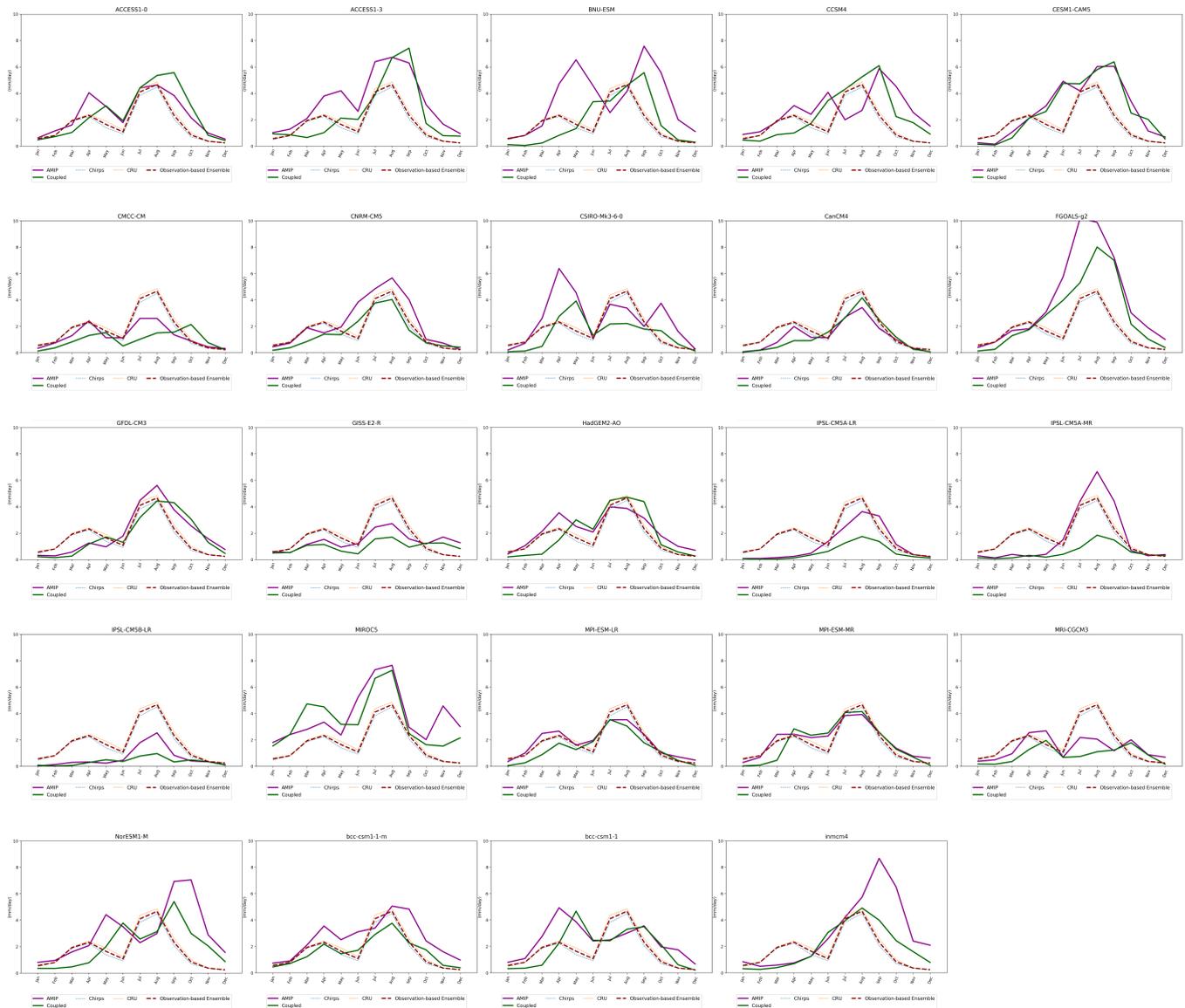


FIGURE A1 Awash basin annual rainfall cycles for AMIP and coupled model simulations and observation-based data (mm/day) [Colour figure can be viewed at wileyonlinelibrary.com]

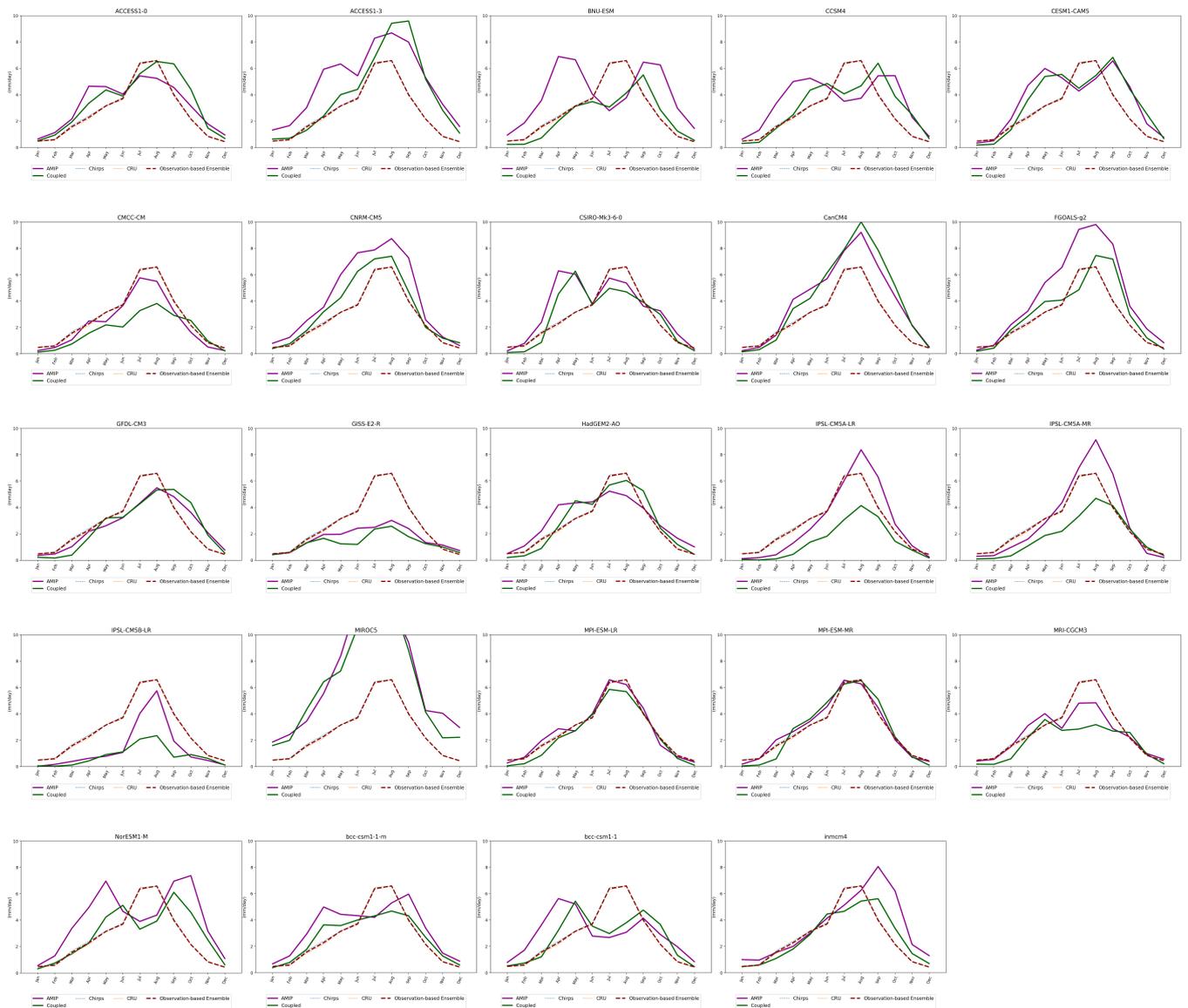


FIGURE A2 Northwest Ethiopia annual rainfall cycles for AMIP and coupled model simulations and observation-based data (mm/day) [Colour figure can be viewed at wileyonlinelibrary.com]

APPENDIX B

TEMPERATURE ANNUAL CYCLES

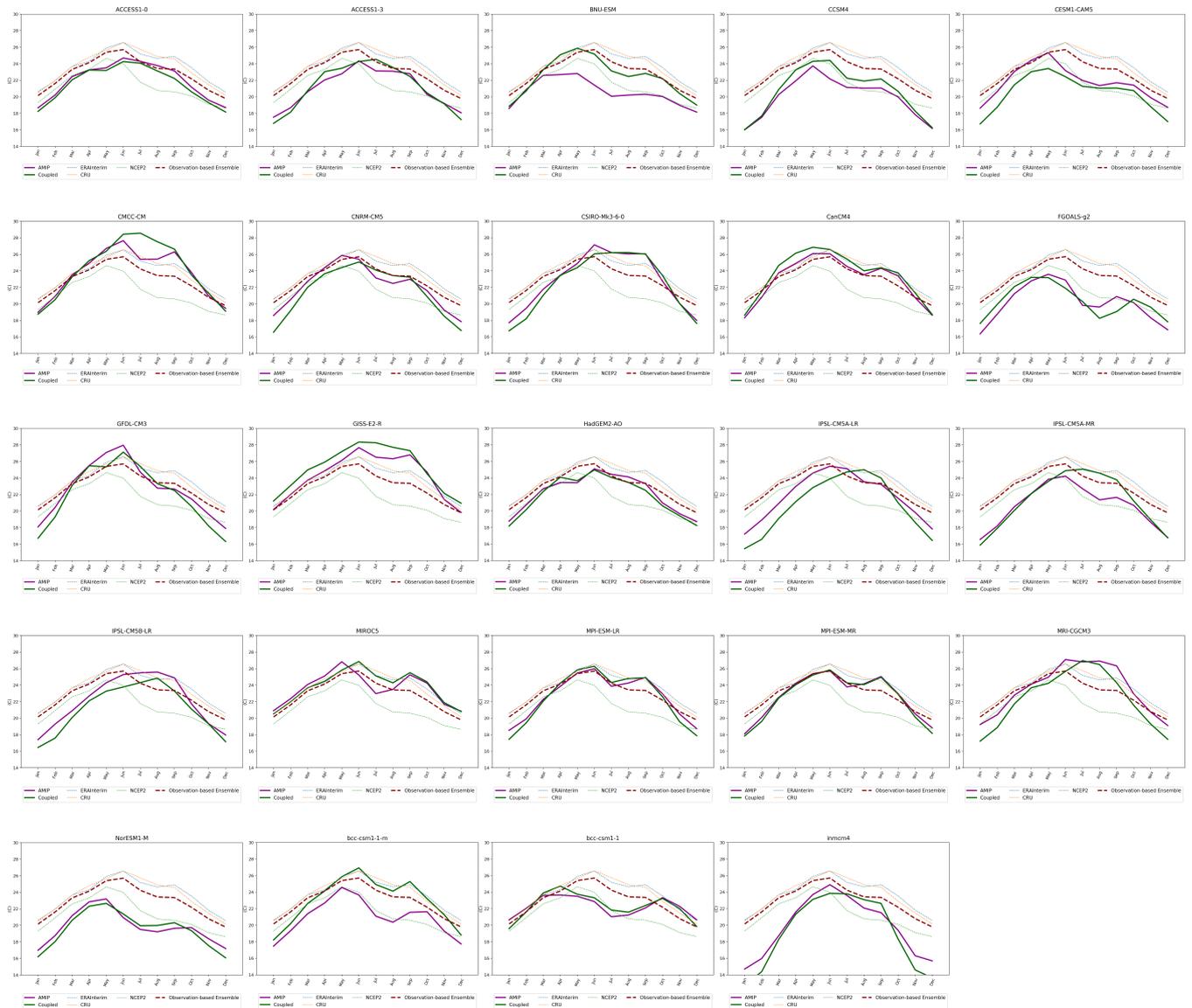


FIGURE B1 Awash basin annual temperature cycles for AMIP and coupled model simulations and observation-based data (C)
[Colour figure can be viewed at wileyonlinelibrary.com]

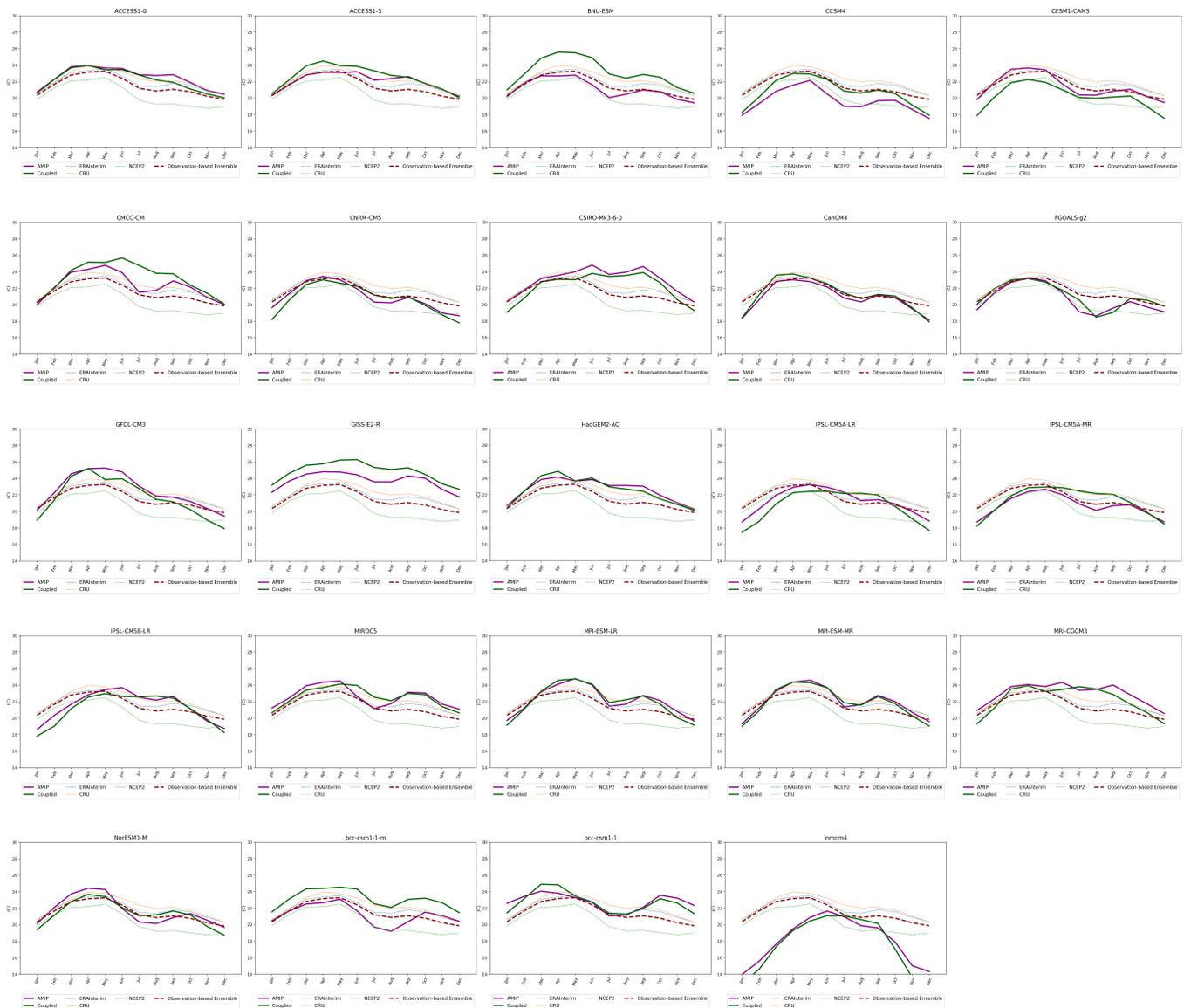


FIGURE B2 Northwest Ethiopia annual temperature cycles for AMIP and coupled model simulations and observation-based data (C) [Colour figure can be viewed at wileyonlinelibrary.com]