



Can climate projection uncertainty be constrained over Africa using metrics of contemporary performance?

David P. Rowell¹ · Catherine A. Senior¹ ·
Michael Vellinga¹ · Richard J. Graham¹

Received: 17 July 2015 / Accepted: 2 November 2015 / Published online: 14 November 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Projections of climate change over Africa are highly uncertain, with wide disparity amongst models in their magnitude of local rainfall and temperature change, and in some regions even disparity in the sign of rainfall change. This has significant implications for decision-makers within the context of a vulnerable population and few resources for adaptation. One approach towards addressing this uncertainty is to rank models according to their historical climate performance and disregard those with least skill. This approach is systematically evaluated by defining 23 metrics of model skill and focussing on two vulnerable regions of Africa, the Sahel and the Greater Horn of Africa. Some discrimination in the performance of 39 CMIP5 models is achieved, although divergence amongst metrics in their ranking of climate models implies some uncertainty in using these metrics to robustly judge the models' relative performance. Importantly, when the more capable models are selected by an overall performance measure, projection uncertainty is not reduced because these models are typically spread across the full range of projections (except perhaps for Central to East Sahel rainfall). This suggests that the method's underlying assumption is false, this assumption being that the modelled processes that most strongly drive errors and uncertainty in projected change are a subset of the processes whose errors are observed by standard metrics of historical climate. Further research must now develop an expert judgement approach that will discriminate models using an in-depth understanding of the mechanisms that drive the errors and uncertainty in projected changes over Africa.

Electronic supplementary material The online version of this article (doi:10.1007/s10584-015-1554-4) contains supplementary material, which is available to authorized users.

✉ David P. Rowell
dave.rowell@metoffice.gov.uk

¹ Met Office Hadley Centre, Fitzroy Road, Exeter EX1 2NX, UK

1 Introduction

The population of Africa is particularly vulnerable to climate change. Widespread poverty renders the very foundations of society – such as rainfed agriculture, water supply and urban sanitation – vulnerable to weather and climate extremes, and furthermore leaves little capacity for adaptation. Climate models suggest that substantial changes in rainfall are expected over many tropical regions (e.g. Chadwick et al. 2015). On the other hand, the precise pattern and magnitude of rainfall and temperature changes is highly uncertain – in some regions even the sign of rainfall change is uncertain – with multi-model ensembles envisaging a wide range of scenarios of climate change over Africa (e.g. Niang et al. 2014; James et al. 2014; Kent et al. 2015). This uncertainty at specific locations poses significant challenges for today's decision-makers who formulate long-term strategic development plans.

Similar projection uncertainty elsewhere has motivated numerous studies to develop metrics of climate model performance to attempt to discriminate between the best and worst models (e.g. Hall and Qu 2006; Déqué and Somot 2010, O'Gorman 2012). The hope is that if the models that perform poorly against observational data are excluded, then this also excludes models with unreliable future projections, so that the remaining relatively trustworthy models are more closely clustered, encompassing a narrower range of projection uncertainty. The assumption inherent in this approach is that the poorly-represented processes that lead to the errors measured in the contemporary climate are also the processes that drive the errors and uncertainty in future changes of interest.

In some instances this approach has had some success, such as assessing the reliability of models' projected snow-albedo feedback through validation of their contemporary seasonal cycle (Hall and Qu 2006), or constraining the storage of tropical terrestrial carbon storage through validation of CO₂ growth rates in coupled climate-carbon-cycle models (Cox et al. 2013). However, for regional analysis it has usually (not always) been found that the apparently poorest models are distributed throughout the range of future projections, so that when these are then excluded there is no reduction in projection uncertainty (e.g. Déqué and Somot 2010, Knutti et al. 2010).

This approach has yet to be appraised for Africa. This study therefore aims to assess the trustworthiness of 39 models from version 5 of the Coupled Model Intercomparison Project (CMIP5, Taylor et al. 2012) for two especially vulnerable regions of Africa, using a basket of carefully chosen metrics. This assessment is then integrated into an analysis of projected rainfall and temperature changes over these regions to evaluate the impact on uncertainty.

2 Scope, data and metrics

Our focus is on climate change in two of the most vulnerable regions of Africa, the Sahel and the Greater Horn of Africa (GHA). Over the Sahel, the majority of models suggest a decline in rainfall over its western part during the coming decades, but an increase in rainfall over the Central to East Sahel (Biasutti 2013; Roehrig et al. 2013). Hence, the Sahel (defined following Rowell 2001; 12–18°N, 16°W–35°E) is divided at the zero line of Biasutti's (2013) ensemble mean projected trend map (5°W; Fig. 1), with data averaged over the July to September (JAS) wet season. Over GHA there are two wet seasons, the Long Rains (March to May, MAM) and the Short Rains (October to December, OND). For both seasons the majority of models project an upward trend in rainfall (e.g. Shongwe et al. 2011; James et al. 2014; Kent et al. 2015). We

use Rowell's (2013) definition of GHA (Fig. 1), which also encompasses the area where the wettest month occurs during MAM or OND (Nicholson 2014) and interannual anomalies are broadly consistent (Rowell 2013). Thus, four 'Region-Season Combinations' (RSCs) are studied here: the West Sahel (W.Sahel), the Central to East Sahel (Cen+E.Sahel) (both using JAS wet season averages), the GHA Long Rains (GHA-LR) and the GHA Short Rains (GHA-SR).

We also focus on projections of two climate variables, seasonal mean rainfall and surface air temperature, both of considerable importance to rural and urban communities in Africa.

The models utilised here are those submitted to version 5 of the Coupled Model Intercomparison Project (CMIP5, Taylor et al. 2012). We evaluate all 39 models that archived both rainfall and surface air temperature data (listed in Fig. 2). Evaluation of their current climate capability is undertaken using data from both their 'historical' coupled ocean-atmosphere general circulation model (AOGCM) simulations (driven by realistic anthropogenic and natural forcings) and their atmosphere-only GCM (AGCM) simulations (following the Atmospheric Model Intercomparison Project [AMIP] protocol, in particular driven by observed global sea surface temperatures [SSTs] and atmospheric constituents). For future climate projections, our focus is on those forced by the RCP8.5 scenario (representative concentration pathway 8.5), since these have good data availability, large signal-to-noise ratios, and the greatest consistency with current emission trends which "track the high end of the latest generation of emission scenarios" (Friedlingstein et al. 2014; Clarke et al. 2014). For all experiments (past and future), only the first member of initial condition ensembles is analysed for consistency amongst models. This is also justified in sections 3 and 4 by further analysis of four AOGCMs with 9 to 10-member initial condition historical ensembles (CanCM4, CNRM-CM5, GFDL-CM2.1 and HadCM3). Last, there are two calculations that also require data from the 'pre-industrial control' AOGCM simulations (for which all external forcing is fixed at pre-industrial conditions), except in this case EC-EARTH is omitted because this simulation was not performed.

The observational datasets used to validate the models are the GPCP-Reanalysis v4 precipitation dataset (Global Precipitation Climatology Centre, Schneider et al. 2014), the GPCP v2.1 pentad precipitation dataset (Global Precipitation Climatology Project, Xie et al. 2003), the Berkeley surface air temperature dataset (Rohde et al. 2013), and the HadISST1.1 SST dataset (Hadley Centre Sea Ice and Sea Surface Temperatures, version 1.1, Rayner et al. 2003). These are chosen for their relatively plentiful use of raw data, high spatial resolution (all

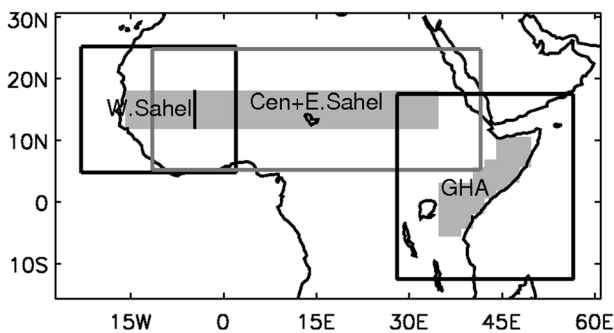


Fig. 1 Definition of the regions used to compute model performance metrics and projected change. Shaded regions are the Region-Season Combinations (RSCs) used for spatial averaging, and boxes are the Large RSCs used to compute spatial RMSEs

available on a 1° grid), sufficient temporal duration and state-of-the-art gap-filling techniques. Use of alternative observational datasets in the context of large modelling uncertainty is unlikely to impact results (e.g. Rowell et al. 2015 for GHA precipitation). All observational and model data are interpolated to a common grid, chosen to be that of the Met Office models, 1.25° latitude by 1.875° longitude. The period used to assess the coupled model data is 1971–2000 (apart from the teleconnections metric, which requires the longer 1922–1994 period, following Rowell 2013), and the period used to assess the AMIP data is 1981–2000.

This study utilises an extensive range of metrics to test both key local processes and remote influences on Africa. These are compiled from a mix of standard climate metrics and a few specialist metrics suggested by recent literature. 23 ‘primary metrics’ are described in Table 1, chosen to meet the following goals:

- Inclusion of a range of ‘processes’: climatological patterns, annual cycle, interannual variance, monsoon onset, and teleconnections.
- Inclusion of a range of spatial scales: from regional patterns to global-scale diagnostics.
- Inclusion of key variables: precipitation, surface air temperature, and SST.
- Inclusion of both AOGCMs and AGCMs, with the former being the models used to project climate change, and the latter being useful to isolate errors due to atmospheric processes alone.

Within most ‘primary metric groups’ a number of sub-metrics are computed for the 4 RSCs or 11–13 SST-SCs (SST-Index-Season Combinations; See Table 1, Note k for detail). The selection and treatment of these sub-metrics depends on the analysis undertaken. Note also that for some models, a number of primary metrics are missing: 14 models did not perform an AMIP simulation, and 6 models did not archive the daily rainfall data required for onset metrics.

3 Discrimination of models

Prior to investigating the capacity for these metrics to constrain projection uncertainty, it is important to determine whether they can successfully discriminate between the capability of CMIP5 models to represent contemporary climate. This is achieved as follows. First, all models are ranked for each sub-metric. Where both absolute (sign removed) and raw (sign included) sub-metrics have been computed, only the absolute sub-metric is used here. Next, the aim is to produce a ranking, across models, for each RSC and each primary metric. This focus on primary metrics is to ensure that the processes they represent have equal weight when aggregated below, so avoiding bias towards processes necessarily represented by a larger number of sub-metrics. The relevant sub-metrics for each primary metric and each RSC are as follows: for the first 16 primary metrics in Table 1, it is simply the single sub-metric for the given RSC; for onset, this is used only for the W.Sahel and Cen+E.Sahel; for SSTs, those sub-metrics of the same season as the given RSC are selected, except that the tropics-wide sub-metrics are selected for all RSCs; and for teleconnections, this single metric is linked to all RSCs (note that it has not been disaggregated to regional sub-metrics due to the sampling concerns raised by Rowell 2013). Thus, the SST primary metrics for each RSC must combine multiple sub-metrics, which is achieved by averaging the sub-metric ranks into a single primary metric rank. Last, for those primary metrics that have missing data for some of the

Table 1 Metrics of model performance used in this study

| Primary Metric | Variable | Model | Sub-Metrics (Regions/Seasons) |
|-------------------------------|---------------|-------|-------------------------------|
| RMSE ^a | Precipitation | AOGCM | 4 x Africa LRSCs ^c |
| RMSE ^a | Precipitation | AGCM | 4 x Africa LRSCs ^c |
| RMSE ^a | Surf Air Temp | AOGCM | 4 x Africa LRSCs ^c |
| RMSE ^a | Surf Air Temp | AGCM | 4 x Africa LRSCs ^c |
| (Abs) Bias ^d | Precipitation | AOGCM | 4 x Africa RSCs ^b |
| (Abs) Bias ^d | Precipitation | AGCM | 4 x Africa RSCs ^b |
| (Abs) Bias ^d | Surf Air Temp | AOGCM | 4 x Africa RSCs ^b |
| (Abs) Bias ^d | Surf Air Temp | AGCM | 4 x Africa RSCs ^b |
| Mean AnnCyc ^e | Precipitation | AOGCM | 4 x Africa RSCs ^b |
| Mean AnnCyc ^e | Precipitation | AGCM | 4 x Africa RSCs ^b |
| Mean AnnCyc ^e | Surf Air Temp | AOGCM | 4 x Africa RSCs ^b |
| Mean AnnCyc ^e | Surf Air Temp | AGCM | 4 x Africa RSCs ^b |
| (Abs) CV ^g | Precipitation | AOGCM | 4 x Africa RSCs ^b |
| (Abs) CV ^g | Precipitation | AGCM | 4 x Africa RSCs ^b |
| (Abs) SD ^f | Surf Air Temp | AOGCM | 4 x Africa RSCs ^b |
| (Abs) SD ^f | Surf Air Temp | AGCM | 4 x Africa RSCs ^b |
| (Abs) Mean Onset ^h | Onset Index | AOGCM | West Africa |
| (Abs) SD Onset ⁱ | Onset Index | AOGCM | West Africa |
| RMSE ^a | SST | AOGCM | Tropics ^j |
| (Abs) Bias ^d | SST | AOGCM | 13 x SST-SCs ^k |
| Mean AnnCyc ^e | SST | AOGCM | 11 x SST-SCs ^k |
| (Abs) SD ^f | SST | AOGCM | 11 x SST-SCs ^k |
| Teleconnections ^l | SST-to-Precip | AOGCM | Africa |

^a *RMSE* ‘root mean squared error’ of the multi-year mean of the pattern in and around the area of interest (LRSC^c)

^b *RSC* ‘region-season combinations’ defined in section 2: W.Sahel (JAS), Cen+E.Sahel (JAS), GHA-LR (MAM) and GHA-SR (OND)

^c *LRSC* ‘large region-season combination’ areas over and around each RSC, combined with the relevant seasons, used to compute RMSE^a, viz. the boxes shown in Fig. 1

^d (*Abs*) *Bias* ‘bias’ (model minus observed) and ‘absolute bias’ (separate sub-metrics, whereby the sign is removed from the latter) of the specified index or regional average

^e *Mean AnnCyc* RMSE of the partial annual cycle of the multi-year mean of the specified index or regional average, using monthly data extended by one month before and after the usual analysis season

^f (*Abs*) *SD* errors in the ‘interannual standard deviation’ (model to observed ratio) and ‘absolute SD’ (absolute value of the logarithm of the model to observed ratio) (separate sub-metrics, whereby symmetry about unity is removed from the latter) of the specified index or regional average

^g (*Abs*) *CV* errors in the ‘interannual coefficient of variation’ (SD divided by the mean, then errors are calculated in the same way as SD errors) and ‘absolute CV’ (separate sub-metrics, whereby symmetry about unity is removed from the latter) of the specified index or regional average

^h (*Abs*) *Mean Onset* bias and absolute bias of the onset date of the West African summer monsoon (Fontaine and Louvet 2006; Vellinga et al. 2013)

ⁱ (*Abs*) *SD Onset* error in the interannual standard deviation and absolute SD (both calculated as #5 above) of the onset date of the West African summer monsoon (Fontaine and Louvet 2006; Vellinga et al. 2013)

^j *Tropics* ocean grid points from 20°S to 20°N using annual means

^k *SST-SC* SST-Index-Season Combinations: tropical Atlantic dipole (JAS; Enfield et al. 1999), equatorial East Atlantic (JAS; Chang et al. 1997), central Indian Ocean (JAS, MAM, OND; Rowell 2013), Indian Ocean dipole (MAM, OND; Saji et al. 1999), Nino-3.4 (JAS, MAM, OND; 5°S–5°N, 170°–120°W), Mediterranean (JAS; Rowell 2013), tropical mean annual mean SSTs (Li and Xie 2012), tropics-wide mean intensity of the equatorial peak of annual mean SSTs (Li and Xie 2012). For the latter two diagnostics, only their (absolute) biases are assessed in keeping with Li and Xie’s (2012) analysis of only the mean state

^l *Teleconnections* proportion of SST-to-AfricanPrecip teleconnections for which the simulated and observed strengths are significantly different (method, regions and seasons follow Rowell 2013)

39 models, the available ranks are inflated to fit a scale of 1 to 39 (note, an alternative policy of ignoring these metrics does not alter our conclusions).

For each model, Fig. 2 shows the histogram of the 23 primary metric ranks, using the Cen+E.Sahel as an example. Results for the other RSCs are shown in the electronic supplemental material Figs. S1–S3. Panels are ordered by the models' overall rank, computed as the average of the available primary metric ranks. To provide some context for these histograms, imagine a hypothetical suite of metrics that give broadly consistent ranks of models and are

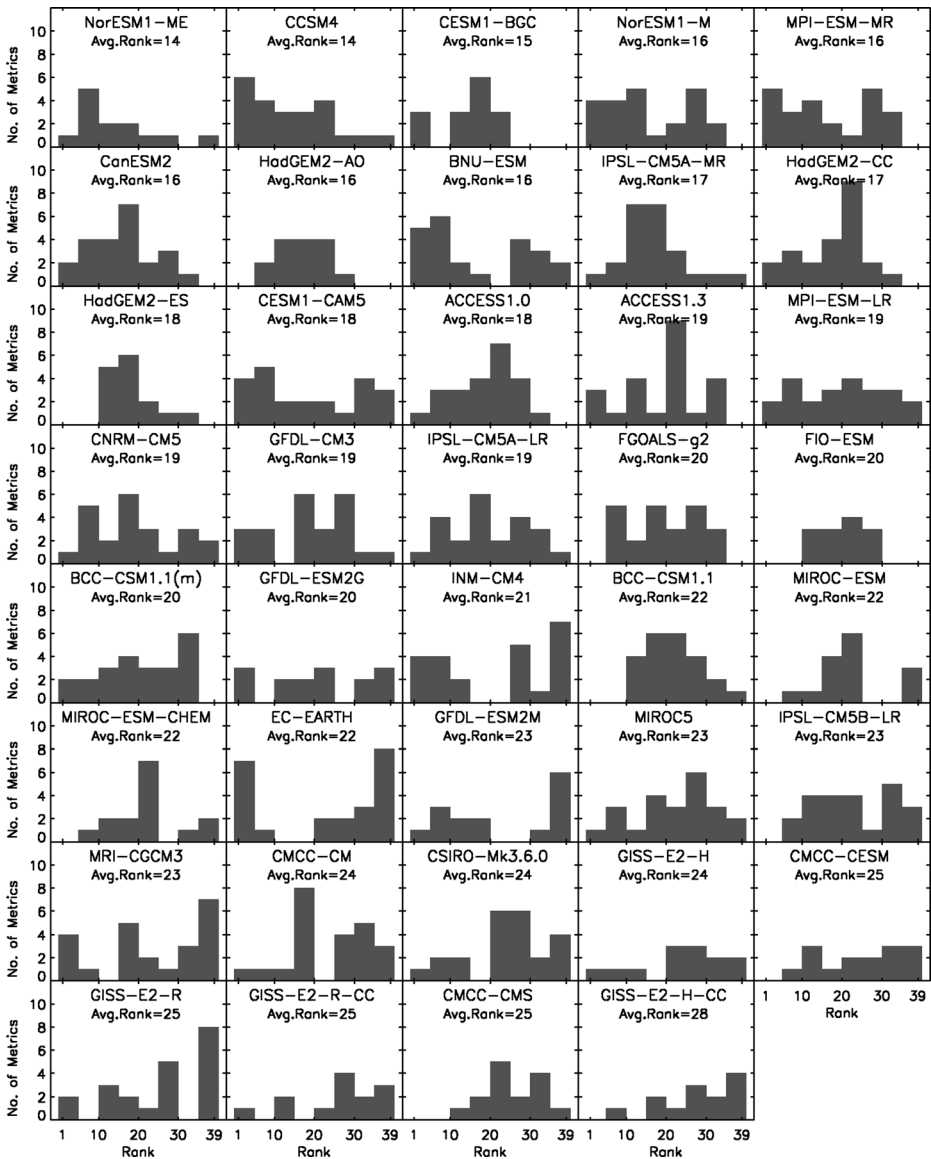


Fig. 2 Histograms of a model's overall ranking for the 23 metrics in Table 1, using all sub-metrics relevant to the Cen+E.Sahel. Low values indicate highest skill, and models are ordered by their overall (average) rank, printed in each panel

able to clearly discriminate the best and worst models. In this case, the histogram for each model would exhibit a narrow peak, with the best models peaking at ranks close to 1, and the worst models peaking at ranks close to 39.

Some discrimination is indeed evident in Fig. 2, with overall ranks ranging from 14 to 28 (with similar ranges for the other RSCs; Figs. S1–S3). However, this discrimination is certainly not as clear as the hypothetical situation above. So we should also consider the opposite situation, which is to test a null hypothesis that the primary metric ranks are scattered at random, so that the discrimination of Fig. 2 arises simply from sampling effects. This could occur, for example, from the chaotic differences between the model and observed phases of natural variability. The above calculations have therefore been repeated with the addition of individual members from four models' 9–10 member initial condition ensembles (using only the AOGCM metrics because AMIP simulations are not available for three of these ensembles). For the Cen+E.Sahel, the average intra-ensemble variance of the simulations' overall rank is found to be just 5 % of the total variance of overall rank. Or, expressed from a different stance, an estimated confidence interval for the overall ranks in Fig. 2 – using twice the standard deviation of the intra-ensemble overall ranks – is ± 3.4 . Broadly similar results are obtained for the other RSCs. More widely, we suggest that any further regions should also be assessed as above, since the contribution of natural variability depends on the location and spatial extent of a region.

In conclusion, the broad range of climate metrics employed here has elicited 'moderate' discrimination of the performance between CMIP5 models over Africa. The impact of random sampling on this discrimination is marginal (also demonstrating that the analysis of a single run from each model has been sufficient). However, there is clearly considerable spread amongst the metrics regarding the ranking of each CMIP5 model, indicating that conclusions about a model's relative contemporary skill can be very dependent on the choice of metrics especially if only a small subset is used. This therefore suggests some uncertainty in the use of these metrics to robustly judge the relative trustworthiness of models' future projections. Last, we also note considerable diversity of the overall model ranking between RSCs (correlation of ranks is 0.39, averaged across all pairs of RSCs), inferring that relative model capability cannot be extrapolated from one region or season to another.

4 Relevance of metrics to climate projections

Prior to utilising this model discrimination to narrow projection uncertainty for our focal regions of Africa, we recall and assess the assumption stated in section 1. This was that those poorly modelled processes that most strongly drive the errors and uncertainty in projected change over Africa are a subset of the processes whose errors are observed by standard metrics of the models' contemporary climate. If this assumption holds true then we might expect a correlation between some of our metrics of the models' present-day performance and their projected regional anomalies. For example, if the erroneous processes that determine shortcomings in some models' climatological rainfall pattern over West Africa are also the key processes that determine the varied W.Sahel response to anthropogenic emissions, then we might expect that the models that perform poorly in this regard also lie at one or both edges of the distribution of projected W.Sahel rainfall changes. These models can then justifiably be excluded from the projection data provided to decision makers.

This assumption is therefore assessed – for each RSC – by correlating, across models, all potentially relevant sub-metrics against projected rainfall or surface air temperature anomalies averaged over the RSC. The sub-metrics relevant for each RSC are selected following sect.3, but here also include the raw sub-metrics (i.e. sign included). Projection anomalies are computed for 2059–2098 minus all years of the pre-industrial control simulations, to maximise signal-to-noise ratios. Histograms of these correlations are presented in Fig. 3. These are plotted against a background theoretical distribution of correlations that would be obtained by taking random samples from two Gaussian populations that have zero correlation, and using the number of model institutes represented in our data (less than the number of available models) to estimate an appropriate size for these random samples. It can be seen that the metric-to-climate-change correlations have a distribution close to that of this null hypothesis (e.g. around 95 % of the measured correlations are within the 95 % confidence limits of the theoretical distribution), so are judged statistically insignificant, i.e. no greater than that expected by chance.

Again, we must also consider the role that natural variability may play in this analysis. Is it possible that chaotic differences between model and observed values of each sub-metric, and chaotic multi-decadal anomalies in the projected climate data, are primarily responsible for the poor correlations shown in Fig. 3? First, assessment of the members of the 9–10 member AOGCM initial condition ensembles reveals that, for the majority of individual sub-metrics, the intra-ensemble variance of model skill is less than 5 % of the total variance of skill across all models. However, for a small minority of sub-metrics (Nino 3.4 variability and MAM Indian Ocean variability) this ratio exceeds 20 %, suggesting a more significant contribution from random variability. Second, following Rowell (2012) and Rowell et al. (2015), the total variance of projected-minus-control anomalies can be separated into components due to natural variability and modelling uncertainty. This reveals that natural variability contributes 2.3–7.5 % of the total variance of 2059–2098 rainfall anomalies for the four RSCs, and 1.2–

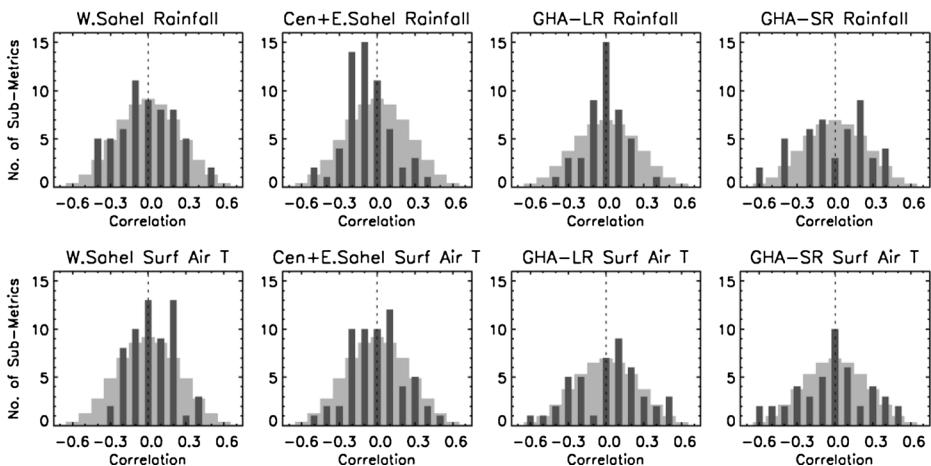


Fig. 3 Histograms of correlations, across models, between the projected regional climate change (2059–2098 minus ‘control’) and each of the relevant sub-metrics (narrow dark grey bars). Results for each of the four RSCs are shown, with upper panels addressing the percentage change in rainfall and lower panels addressing the change in surface air temperature. Also shown (wider light grey bars) is the theoretical distribution of correlations using random samples from two infinite uncorrelated datasets that have a population correlation of zero, with sample size equal to the number of model institutes

2.7 % of the variance of surface air temperature anomalies. Thus, natural variability also contributes relatively little to projection uncertainty at the spatio-temporal scales addressed here. Overall, random sampling of the chaotic component of climate variability does not explain the lack of correlation found in Fig. 3.

In summary, the skill of climate models, measured by any of the individual metrics trialled here, is not linearly related to their projections of rainfall or surface air temperature change. Thus there is no obvious relationship that could be used to constrain projection uncertainty. However, it is possible that (a) large errors in one or more of the zero-bounded metrics (RMSE or ‘Mean AnnCyc’; Table 1) leads to outlying projections of both extremes and hence a highly non-linear relationship between model skill and projection anomalies, or (b) it is necessary to combine all sub-metrics into an overall rank (or metric) to reveal a relationship between model skill and projection anomalies. The approach taken in section 5 addresses these possibilities.

5 Impact on projection uncertainty

Figure 4 shows the CMIP5 model projections for rainfall and surface air temperature for each RSC. A 20-year period centred on 2050 is chosen, which is typical of the timescales considered by decision makers, for example for infrastructure development (e.g. Stafford Smith et al. 2011). Projections for the end of the 21st century lead to the same conclusions, with one exception noted below, and the distinction that anomalies and signal-to-noise ratios are inevitably larger.

It is clear from Fig. 4 that modelling uncertainty of mid-century bidecadal means is substantial, and (*c.f.* section.4) much greater than that due to natural variability shown by the vertical bar (although interannual variability is inevitably much larger due to shorter period fluctuations; not shown). Also shown by the blue bars is the effect of excluding (or culling) the poorest 25 % or 50 % of these models, judged by their overall rank. Recall that this model ranking is identical for the rainfall and temperature projections (i.e. the same models are excluded), but that the ranks differ between RSCs (different models are excluded).

The statistical significance of the reductions in uncertainty due to this model culling is assessed by removing a random selection of models and repeating this 10,000 times to form a comparative null hypothesis distribution. Note that even this random culling of models biases the analysis towards reduced uncertainty, measured by the full maximum-to-minimum range, because no selections can enhance this range. Similarly, random culling is also more likely to reduce than enhance the range of the central 80 % of models (the diagnostic shown by the wide part of the blue bars in Fig. 4), with the amplitude of this tendency being strongly dependent on the distribution of the projection data.

Consider first the projected changes in rainfall over the four RSCs. In the W.Sahel, more than half the models suggest a decline in rainfall over the next 3–4 decades (Fig. 4a; *c.f.* Biasutti 2013; Roehrig et al. 2013). However, a significant minority suggest a small increase in rainfall, and the two models with the largest change in rainfall are extreme positive outliers suggesting a substantial and high-impact increase in rainfall. Culling the poorest quarter or half of models (red X’s and optionally blue ?’s) – judged by standard metrics of contemporary climate – fails to provide any statistically significant reduction in uncertainty or any consensus on the sign of change. These models are distributed throughout the uncertainty range. The apparent reduction arising from the 50 % cull is strongly dependent on the reliability of the ranks of the two outlying models. A physical understanding of the relevance to climate change

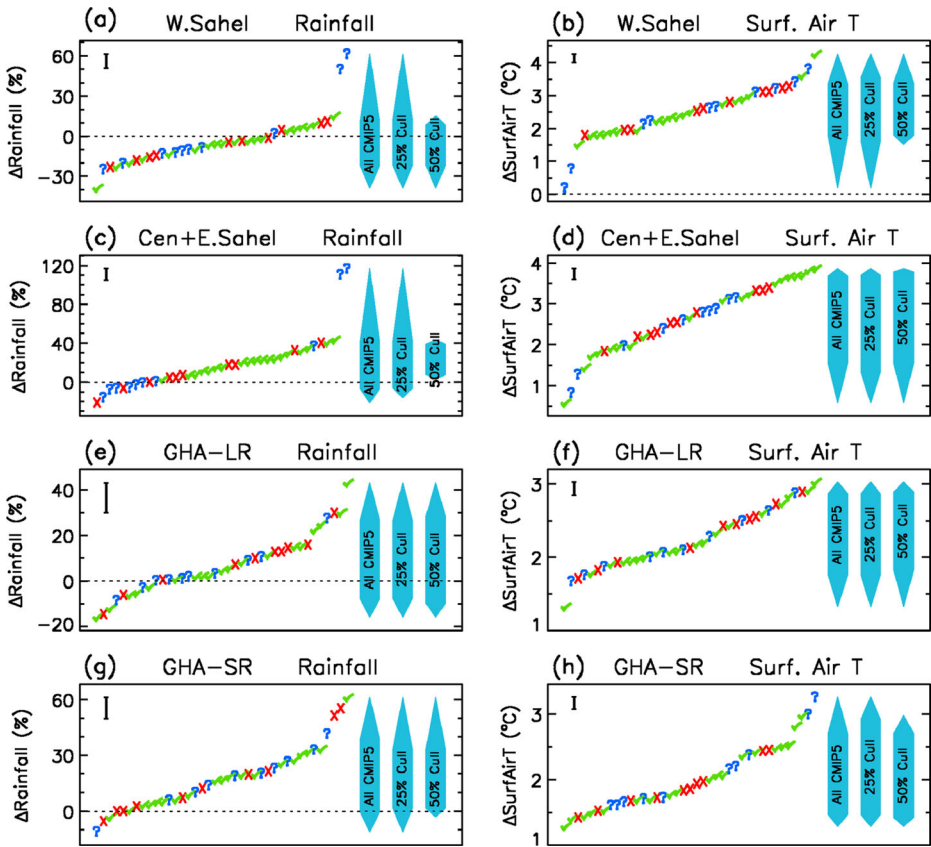


Fig. 4 Projected regional climate change for 2041–2060 from a 1951–2000 baseline for percentage change in rainfall (left-hand panels) and surface air temperature (°C) (right-hand panels), for each of four RSCs. Symbols show the anomalies for each individual model, for the best 50 % of models (using their overall rank, such as displayed in Fig. 2 for Cen+E.Sahel) (green ‘✓’), worst 25 % of models (red ‘X’), and intermediate models (blue ‘?’). Vertical bar at the top left shows the range of natural variability computed as the average of two standard deviations of 20-year means from the models’ corresponding pre-industrial control run. Pale blue shapes show the most extreme anomalies of the ensemble (end points) and the 10th to 90th percentiles (thick part) for all CMIP5 models and for sub-ensembles after excluding the poorest 25 % or 50 % of models (note, this cull is based on different sub-metrics for each RSC, selected to be relevant for that RSC)

of these two models’ moderately poor contemporary performance (ranked 24th and 32nd) is now required.

Over the Cen+E.Sahel, most (but not all) models suggest an increase in rainfall, of up to 40 % over the next 3–4 decades, with two outlying models suggesting that wet season totals could more than double (Fig. 4c). In contrast, a small minority of models suggest either little change or a decline in rainfall of up to 20 %. Culling the poorest quarter of models has little impact on this uncertainty, but further culling the next poorest quarter of models produces a sizable and statistically significant (at the 1 % level) reduction in uncertainty and a robustness in the sign of response. These models (blue ?’s) can be seen to cluster towards both ends of the projected range. The amplitude of this reduction is again heavily reliant on the reliability of the ranking of the two outlying models (24th and 25th in this study), and also dependent on the time period analysed. The analysis of Fig. 3, and further non-linear analysis (not shown), fails

to reveal any relationships between sub-metric values and projected rainfall anomalies, so it appears that the relationship between model skill and rainfall change over the Cen+E.Sahel arises from the combination of all sub-metrics into an overall rank, i.e. a combination of physical processes.

Over the GHA a majority of models suggest an increase in rainfall totals during both seasons, with this confidence in sign being greater during the Short Rains than the Long Rains (Fig. 4c,d; *c.f.* Kent et al. 2015; Rowell et al. 2015). However, the magnitude of this increase varies greatly amongst models, from little change to rather extreme scenarios. Exclusion of the models least able to represent the current climate has no statistically significant impact on this uncertainty (apart from a small and marginally significant impact on the central 80 % range of GHA-SR when 25 % of models are excluded). This reflects the fact that the poorest models (red X's and blue ?'s), as measured by the approach taken here, are scattered throughout the CMIP5 ensemble. Clearly we need to consider whether more sophisticated metrics can be used to discriminate models in a way that usefully narrows projection uncertainty.

The CMIP5 surface air temperature projections are similarly wide-ranging. Over the Sahel, models suggest plausible warming scenarios from 0.5 °C to 4.3 °C by mid-century, and over GHA from 1.3° to 3.3 °C, in both cases implying considerable uncertainty in impacts. For all four RSCs, the apparently poorest models are scattered throughout the ordered anomalies of Fig. 4e-h, so that excluding these models has no statistically significant impact on the uncertainty range.

6 Conclusions

This study has evaluated the contemporary skill of 39 CMIP5 models for two vulnerable regions of Africa, split into four 'Region-Season Combinations': the West Sahel, the Central to East Sahel, and the Greater Horn of Africa's Long Rains and Short Rains seasons. A broad range of performance metrics was compiled to assess these models' climatological patterns, annual cycle, interannual variance, monsoon onset date, and teleconnections, for regional to global spatial scales, different model variables, and different model configurations.

Some discrimination of the most and least skilful CMIP5 models was achieved. However, assessment across 23 'primary metrics' often led to differing conclusions about a model's relative skill, implying some uncertainty in the use of these metrics to assess the relative trustworthiness of the models' future projections. The models' overall rankings also differed between RSCs, inferring spatial and seasonal diversity of the relative strengths and weaknesses of different models.

The motivation for discriminating the CMIP5 models was to assess whether disregarding those with a poor simulation of present-day climate might lead to a narrowing of the uncertainty range of their projected regional changes. Unfortunately, the most deficient models were usually distributed throughout the CMIP5 ensemble, so that excluding these had no significant impact on the large uncertainty of projected rainfall and temperature changes. This was perhaps not surprising after finding that no statistically significant relationships could be identified between the measures of contemporary climate skill and the sensitivity of future regional African climate to anthropogenic emissions. An exception, however, was rainfall projections for the Central to East Sahel, for which rejecting the poorest 50 % of models – based on an overall ranking pooled across all metrics – led to a statistically significant reduction in uncertainty and unanimity

in the sign of change. Nevertheless, a note of caution is that this result is sensitive to both the ranking of the two models that have extreme wet projections and to the precise list of constituent metrics. Further investigation is required to determine whether physical links exist between the models' contemporary climate skill and their projected response over this region, so substantiating or refuting this result.

Thus in general, the comprehensive suite of metrics employed here is unable to address modelling deficiencies in the primary mechanisms that drive regional anthropogenic climate change over Africa and its large uncertainty across the CMIP5 ensemble. To achieve this will require significant research to identify and understand these mechanisms and then uncover ways in which they can be observed in the present climate. This 'expert judgement' approach could discriminate models in a way that is directly relevant to their regional climate change projections, and so help evaluate whether or not the range of projections includes the truth, and if so, provide a basis for narrowing that range.

Acknowledgments DPR, MV and RJG were funded by the U.K. Department for International Development for the benefit of developing countries (the views expressed here are not necessarily those of DFID), and CAS was supported by the Joint UK DECC/Defra Met Office Hadley Centre Climate Programme (GA01101). The many modelling groups listed in Fig. 1 are gratefully acknowledged for producing and making their simulations available, as is the World Climate Research Programme Working Group on Coupled Modelling (WCRP-WGCM), for taking responsibility for the CMIP5 model archive, and the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison (PCMDI) for archiving the model output.

Authors Contribution David P Rowell: Methods, analysis, interpretation, writing, figures.

Catherine A Senior: Discussion of Fig. 4 design, and edits to text and interpretation.

Michael Vellinga: Calculation of onset dates and edits to text.

Richard J Graham: Edits to text and interpretation

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Biasutti M (2013) Forced Sahel rainfall trends in the CMIP5 archive. *J Geophys Res* 118:1–11. doi:10.1002/JGRD.50206
- Chadwick R, Good P, Martin G, Rowell DP (2015) Large rainfall changes expected over tropical land in the coming century. *Nature Climate Change* in press
- Chang P, Ji L, Li H (1997) A decadal climate variation in the tropical Atlantic ocean from thermodynamic air-sea interactions. *Nature* 385:516–518
- Clarke L., K. Jiang, K. Akimoto, M. Babiker, G. Blanford, K. Fisher-Vanden, J.-C. Hourcade, V. Krey, E. Kriegler, A. Löschel, D. McCollum, S. Paltsev, S. Rose, P. R. Shukla, M. Tavoni, B. C. C. van der Zwaan, and D.P. van Vuuren, 2014: Assessing Transformation Pathways. In: *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Edenhofer, O., R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel and J.C. Minx (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Cox PM, Pearson D, Booth BB, Friedlingstein P, Huntingford C, Jones CD, Luke CM (2013) Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature* 494:341–345
- Déqué M, Somot S (2010) Weighted frequency distributions express modelling uncertainties in the ENSEMBLES regional climate experiments. *Clim Res* 44:195–209

- Enfield DB, Mestas-Nunez AM, Mayer DA, Cid-Serrano L (1999) How ubiquitous is the dipole relationship in tropical Atlantic sea surface temperatures? *J Geophys Res* 104(C4):7841–7848
- Fontaine, B. and Louvet, S., 2006: Sudan-Sahel rainfall onset: Definition of an objective index, types of years, and experimental hindcasts. *J. Geophys. Res.*, 111, D20103, doi:10.1029/2005JD007019
- Friedlingstein P, Andrew RM, Rogelj J, Peters GP, Canadell JG, Knutti R, Luderer G, Raupach MR, Schaeffer M, van Vuuren DP, Le Quéré C (2014) Persistent growth of CO₂ emissions and implications for reaching climate targets. *Nat Geosci* 7:709–715
- Hall A, Qu X (2006) Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophys Res Lett* 33:L03502. doi:10.1029/2005GL025127
- James R, Washington R, Rowell DP (2014) African climate change uncertainty in perturbed physics ensembles: implications of warming to 4 °C and beyond. *J Clim* 27:4677–4692
- Kent C, Chadwick R, Rowell DP (2015) Understanding uncertainties in future projections of seasonal tropical precipitation. *J Clim* 28:4390–4413
- Knutti R, Furrer R, Tebaldi C, Messori G, Smerli G, Meehl GA (2010) Challenges in combining projections from multiple models. *J Clim* 23:2739–2758
- Li G, Xie S-P (2012) Origins of tropical-wide SST biases in CMIP multi-model ensembles. *Geophys Res Lett* 39: L22703. doi:10.1029/2012GL053777
- Niang, I., O.C. Ruppel, M.A. Abdrabo, A. Essel, C. Lennard, J. Padgham, and P. Urquhart, 2014: Africa. In: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Barros, V.R., C.B. Field, D.J. Dokken, M.D. Mastrandrea, K.J. Mach, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1199–1265.
- Nicholson SE (2014) A detailed look at the recent drought situation in the greater horn of Africa. *J Arid Environ* 103:71–79
- O’Gorman PA (2012) Sensitivity of tropical precipitation extremes to climate change. *Nat Geosci* 5:697–700
- Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J Geophys Res* 108(D14):4407. doi:10.1029/2002JD002670
- Roehrig R, Bouniol D, Guichard F, Hourdin F, Redelsperger J-L (2013) The present and future of the west African monsoon: a process-oriented assessment of CMIP5 simulations along the AMMA transect. *J Clim* 26:6471–6505
- Rohde R, Muller R, Jacobsen R, Perlmutter S, Rosenfeld A, Wurtele J, Curry J, Wickham C, Mosher S (2013) Berkeley earth temperature averaging process. *Geoinfor Geostat: An Overview* 1:2. doi:10.4172/2327-4581.1000103
- Rowell DP (2001) Teleconnections between the tropical Pacific and the Sahel. *Q J R Meteorol Soc* 127:1683–1706
- Rowell DP (2012) Sources of uncertainty in future changes in local precipitation. *Clim Dyn* 39:1929–1950
- Rowell DP (2013) Simulating SST teleconnections to Africa: what is the state of the art? *J. Climate* 26:5397–5417
- Rowell DP, Booth BBB, Nicholson SE, Good P (2015) Reconciling past and future rainfall trends over east Africa. *J. Climate* in press
- Saji NH, Goswami BN, Vinayachandran PN, Yamagata T (1999) A dipole mode in the tropical Indian Ocean. *Nature* 401:360–363
- Schneider U, Becker A, Finger P, Meyer-Christoffer A, Ziese M, Rudolf B (2014) GPCC’s new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theor Appl Climatol* 115:15–40
- Shongwe ME, van Oldenborgh GJ, van den Hurk B (2011) Projected changes in mean and extreme precipitation in Africa under global warming. Part II: East Africa. *J Climate* 24:3718–3733
- Stafford Smith MS, Horrocks L, Harvey A, Hamilton C (2011) Rethinking adaptation for a 4 °C world. *Phil Trans R Soc A* 369:196–216
- Taylor, K.E., R.J. Stouffer, G.A. Meehl, 2012: An Overview of CMIP5 and the experiment design.” *Bull Am Meteorol Soc*, 93, 485–498
- Vellinga M, Arribas A, Graham R (2013) Seasonal forecasts for regional onset of the west African monsoon. *Clim Dyn* 40:3047–3070
- Xie P, Janowiak JE, Arkin PA, Adler RF, Gruber A, Ferraro RR, Huffman GJ, Curtis S (2003) GPCP pentad precipitation analyses: an experimental dataset based on gauge observations and satellite estimates. *J Clim* 16:2197–2214