Report

Understanding and quantifying extreme precipitation events in South Asia

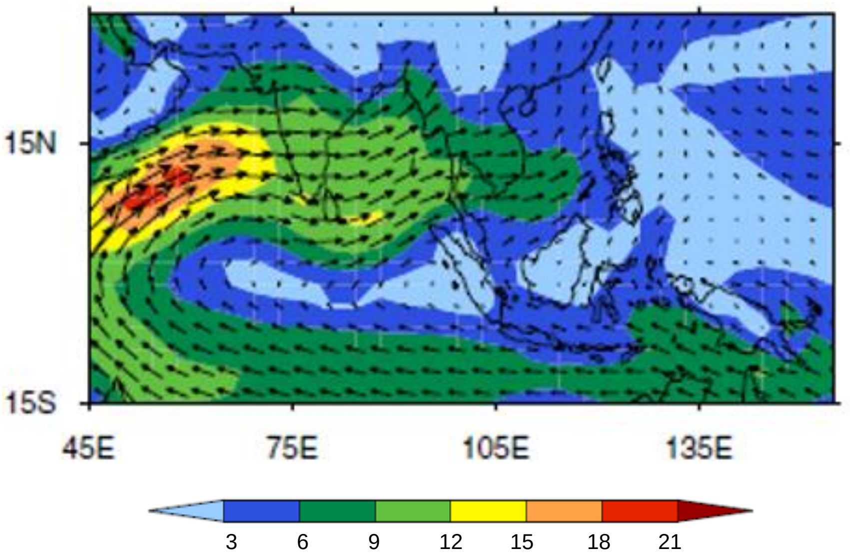
Part II – Process-based evaluation of climate model simulations for South Asia

CARISSA Activity 4: Climate services for the water and hydropower sectors in South Asia

December 2020

Katy Richardson

Reviewed by Joseph Daron and Tamara Janes



Observations of the climatological monsoon circulation (McSweeney et al., 2015)

Delivery Partners:

Contents

Delivery Partners:

1. Introduction

This report constitutes Part II in a series of reports documenting analysis of extreme precipitation in South Asia. This work contributes to two workstreams within the CARISSA (Climate Analysis for Risk Information and Services in South Asia)[1] Work Package of the ARRCC (Asia Regional Resilience to a Changing Climate)[2] programme; workstream 4 focused on developing climate services for the water and hydropower sectors, and workstream 6 focused on developing climate information for food security assessments.

During Year 1 of the CARISSA project, a regional workshop was held in Nepal bringing together users and providers of climate information in the water and hydropower sectors across South Asia (Met Office & ICIMOD, 2019). In addition, initial work was conducted for a pilot study to provide climate information to the hydropower sector in Nepal (Met Office, 2020a). The outcomes of this exploratory work highlighted current and future changes to extreme precipitation as a primary concern amongst stakeholders, as extreme precipitation leads to flooding and other hazards that have wide-ranging impacts on the water and hydropower sector in the region.

Improved understanding of the causes of extreme precipitation events in the region, and how well climate models capture these, is required to improve confidence in climate information products focused on future changes in extreme precipitation. Therefore, Year 2 of the programme has focussed on underpinning analysis of extreme precipitation events with a view to informing the planned development of climate information products for the water, hydropower and food security sectors across the ARRCC focal countries; Afghanistan, Bangladesh, Nepal and Pakistan.

The overall aim of this work is to determine a set of plausible futures for extreme precipitation in South Asia, to support policy and planning in the identified sectors. This report builds on the outcomes of Part I of this series of reports which examined case studies of extreme precipitation events in the ARRCC focal countries to identify the large-scale climate processes associated with such events (Met Office, 2020b). This report constitutes Part II in the series, providing the results of an evaluation of available climate model projections and identifying models that poorly represent extreme precipitation and associated large-scale climate processes identified in Part I.

Simulating the regional climate of South Asia is challenging due to the complex topography and monsoon system. It is known that climate models have difficultly simulating the regional distribution and variability of monsoon rainfall (Singh et al., 2017; Sperber et al., 2013a; Turner & Annamalai, 2012). We therefore evaluate the ability of climate models to capture the large-scale dynamics and driving processes of

[1] https://www.metoffice.gov.uk/services/government/international-development/climate-analysis-for-risk-information--services-in-south-asia-carissa
[2] https://www.metoffice.gov.uk/services/government/international-development/arrcc

the regional climate – i.e., a ‘process-based’ evaluation – as these processes are often better represented compared to the direct precipitation outputs. If the large-scale processes are well captured, then we can have greater confidence in model outputs over the region. However, if these processes are not well simulated then projected changes are unlikely to provide a plausible representation of the future climate, and these model simulations should be excluded from the ensemble of projections used to construct a set of plausible future climate conditions (McSweeney et al., 2012; McSweeney et al., 2015 - hereafter referred to as Mc15).

A number of studies have evaluated global climate model (GCM) simulations from the Coupled Model Intercomparison Project Phase 5 (CMIP5), used to inform the most recent Intergovernmental Panel on Climate Change (IPCC) Assessment Report (AR5; IPCC, 2013), in representing a range of metrics that capture key monsoon processes. In addition to these GCM simulations, a number of these have been dynamically or statistically downscaled to generate regional climate projections for South Asia (Giorgi & Gutowski, 2015; Janes et al., 2019). However, in some cases there is a lack of evaluation of the models selected to be downscaled. Furthermore, it is not necessarily true that the regional projections provide added value over the GCMs and in some cases the regional climate models (RCMs) perform worse (Singh et al., 2017). Here we draw together assessments from the literature, including assessments of those GCMs which have been downscaled to generate regional climate projections for South Asia, to evaluate the range of climate projections available. We conduct our evaluation assessment with a focus on the ability of models to capture processes that lead to extreme precipitation, such as the intra-seasonal variability in the monsoon as identified in our case study assessment in Part I (Met Office, 2020b).

The methods and data that are used for the process-based evaluation are presented in Section 2, the results of the evaluation and assessment of available climate model simulations are shown in Section 3, and a summary and recommendations for future work are provided in Section 4.

2. Data and methods

2.1 Climate model simulations evaluated

This study focuses on evaluating the capability of existing climate model simulations in capturing the relevant processes driving extreme rainfall in South Asia. We evaluate the capability of GCMs from the CMIP5 phase of climate model simulations (see Table 1 for details), and a range of regionally downscaled CMIP5 projections from different downscaling experiments (Table 1).

Table 1 – Details of the four climate model projection datasets considered in this assessment.

Name	Details	Reference
CMIP5	GCM simulations from the World Climate Research Project (WCRP) Coupled Model Intercomparison Project Phase 5 (CMIP5)	Taylor et al. (2012)
CORDEX South Asia	RCM simulations for the South Asia domain from the WCRP CoOrdinated Regional climate modelling Downscaling EXperiment (CORDEX). These are dynamically downscaled CMIP5 model simulations from three different RCMs. There are two sets of simulations at different resolutions; 17 simulations at 50km resolution (WAS-44) and 9 more recent simulations at 25km resolution (WAS-22) which use more recent versions of the RCMs. The CORDEX WAS domain is shown in Figure 1 (left panel) and the specific GCM-RCM combinations are shown in Table 2.	Giorgi & Gutowski (2015)
DECCMA	RCM simulations for the South Asia domain from the DECCMA (Deltas, vulnerability and Climate Change: Mitigation and Adaptation) project. These are dynamically downscaled CMIP5 model simulations with the Met Office HadRM3P RCM. The domain used in shown in Figure 1 (right panel) and the CMIP5 models downscaled are shown in Table 2.	Janes et al. (2019)
Climate change scenarios for Nepal (hereafter Nepal scenarios)	Statistically downscaled CMIP5 model simulations used in the climate scenarios developed for the Nepal National Adaptation Plan. The CMIP5 models downscaled are shown in Table 2.	Ministry of Forests and Environment (2019)

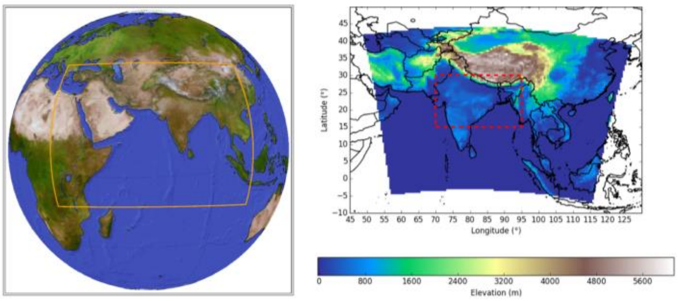


Figure 1 – Domains of the regional climate model simulations considered; CORDEX South Asia domain (left panel) and domain used in the DECCMA project (right panel).

The regional climate projections listed in Table 1 represent both dynamical and statistical methods for downscaling the CMIP5 projections. A different subset of CMIP5 model simulations was used in each downscaling experiment, resulting in 15 CMIP5 models that have been downscaled for South Asia. When assessing the capability of the CMIP5 GCMs we are therefore particularly interested in how well those models that have been downscaled capture relevant processes. Table 2 shows the subset of 15 CMIP5 models that have been downscaled by the different experiments outlined in Table 2 and the specific method used. Each of these models has been assigned a colour shading for ease of identifying the models in subsequent assessments.

Table 2 – Table of CMIP5 models that have been regionally downscaled for South Asia showing which models have been downscaled by each regional climate model/method. Model names are assigned a colour shading for ease in identifying them in later analysis.

	Downscaling experiment							
	CORDEX WAS-44			**CORDEX WAS-22**			**DECCMA project**	**Nepal scenarios**
CMIP5 model	**RCA4**	**RegCM 4-4**	**REMO 2009**	**COSMO-crCLIM**	**RegCM 4-7**	**REMO 2015**	**HadRM3P**	**Statistical**
bcc-csm1-1								✓
CanESM2	✓	✓						✓ [3]
CNRM-CM5	✓	✓					✓	
CSIRO-Mk3-6.0	✓	✓						
EC-EARTH	✓			✓				
GFDL-CM3							✓	
GFDL-ESM2M	✓	✓						✓
HadGEM2-ES	✓					✓	✓	
IPSL-CM5A-LR		✓						
IPSL-CM5A-MR	✓							
MIROC5	✓				✓			
MIROC-ESM-CHEM								✓
MPI-ESM-LR	✓		✓	✓		✓		
MPI-ESM-MR		✓			✓			
NorESM1-M	✓			✓	✓	✓		

[3] The r2 and r5 ensemble members of the CanESM2 model were used for the Nepal scenarios. It is assumed that the r1 ensemble member was used in all other cases here.

2.2 Process-based evaluation method

Methods for sub-selecting GCMs for regional climate assessments usually involve the selection of a small sample of GCMs to be downscaled, generating a new set of regional climate projections. This is because high-resolution RCM simulations are computationally expensive and so downscaling a subset of GCMs that represent the plausible range of future projections reduces the computational cost of such experiments. Criteria for selecting GCMs for downscaling include the model's ability to simulate the baseline climate and large-scale climate features that are important to the focus region, and that the range of plausible future projections is sufficiently represented by the selected subset.

Instead of selecting models for downscaling and generating new regional climate projections, in this study we evaluate existing GCM and regional projections that are already available in the public domain (i.e., those in Table 1). The aim of this approach is to inform the appropriate use of the existing simulations in identifying plausible future climate scenarios of extreme precipitation in South Asia. The methods used to select the GCMs that were previously downscaled for South Asia vary across the different downscaling experiments, ranging from no evaluation of the GCMs to full process-based evaluation. Therefore, it is important to consider the specific combinations of driving GCM and regional downscaling model or method to fully evaluate the suitability of these available projections (discussed further in Section 3.2).

Although we are not sub-selecting model simulations for downscaling, we apply relevant aspects of these methods for conducting the process-based evaluation and recommendations for appropriate use of the available projections. We follow the approach presented in Mc15 (described in Figure *2* and Table 3) where the first stage is to evaluate the representation of key physical processes in the GCMs. A subset of GCMs to downscale is then identified based on the range of future projections using a decision-making matrix as shown in Table 3.

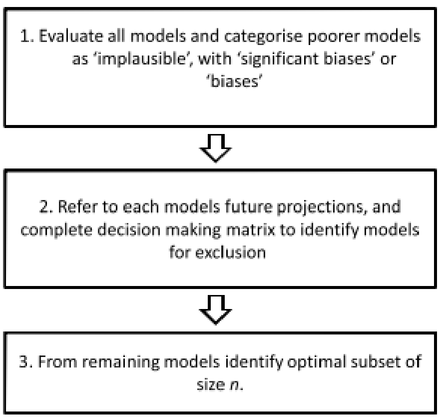


Figure 2 *– Methodology steps from McSweeney et al. (2015)*

Table 3 – *Decision making matrix for potential elimination of ensemble members, from McSweeney et al. (2015)*

Model performance	Model projections	
	Outlier	Other models predict similar outcomes too.
Model suffers sufficient shortcoming that it significantly reduce our confidence in its projections ('Implausible')	*Exclude: we should carefully document justification for this, however, as exclusion will affect the range of outcomes*	*Exclude: We can avoid using these models without much affecting the range of projected outcomes*
Model suffers significant shortcomings which we cannot clearly link to confidence in its projections ('Biases/Significant Biases')	Include: we do not have strong enough evidence to exclude these outcomes from the projections	*Exclude: We can avoid using these models without much affecting the range of projected outcomes*
Model performance is satisfactory ('Satisfactory')	Include	Include

Other methods for sub-selecting model projections for regional climate studies include an envelope-based approach by Lutz et al. (2016), which was used to select the CMIP5 GCMs to statistically downscale for the Nepal scenarios. This method first selects models that represent the 'four corners' (e.g. 'hot and dry', 'cold and wet') of the projected range in mean temperature and precipitation for a specific region. Projected changes in extremes metrics are then used to reduce the subset further, followed by an assessment of the model skill in the climatological annual cycles of temperature and precipitation for the specified region. In this approach the focus is on assessing the outputs of the models in the region of interest, and no consideration is given to the ability of the models to simulate the key processes affecting the climate of the wider region.

In this study we take a process-based evaluation approach (as per stage 1 of the Mc15 approach) to assessing the suitability of the available climate model projections (both global and regional) for use in generating information about future changes in extreme precipitation in South Asia. We first focus on evaluating the capability of GCM simulations in representing the relevant large-scale driving processes. We consider the following key processes that drive extreme precipitation in South Asia, as documented in Part I of this series of reports (Met Office, 2020b):

1. Monsoon circulation (speed and direction of the 850 hPa flow),
2. Large-scale drivers of the monsoon circulation, such as the teleconnection with the El Niño Southern Oscillation (ENSO) which governs inter-annual variability of monsoon precipitation,
3. Drivers of intra-seasonal variability in monsoon precipitation, such as the Boreal Summer Intra-Seasonal Oscillation (BSISO).

We draw on skill assessments of these processes in the published peer-reviewed literature. Three main studies are considered, these are:

1. McSweeney et al. (2015; referred to as Mc15) which assesses flow characteristics of the south-west monsoon in a subset of 35 CMIP5 models using a qualitative approach to identify poor performing models that should be excluded for use in regional climate assessments.

2. Sperber et al. (2013; hereafter referred to as Sp13) which undertakes a thorough skill assessment of a subset of 25 CMIP5 models using metrics for the climatology, climatological annual rainfall cycle and inter-annual variability of the South Asian monsoon.

3. Sabeerali et al. (2013; hereafter referred to as Sa13) which assesses the skill of a subset of 32 CMIP5 models to represent important aspects of the BSISO.

A variety of different metrics are used across these studies and we consider all relevant metrics in this assessment as no model can be excluded or preferred based on a single metric. The Sp13 and Sa13 studies both take the approach of identifying the best performing models for the metrics considered. In this assessment we consider these studies from the perspective of identifying and excluding the poorest performing models, as per the Mc15 approach. We assign each model simulation an assessment category of 'red', 'orange', 'yellow' or 'green' details of the categories are provided in Table 4.

Table 4 – Assessment categories for the GCM process-based evaluation

Assessment category	**Criteria**
Red	Model simulations that fail on two or more metrics considered and should be excluded from the available projections
Orange	Model simulations which fail on one metric and should be used with caution
Yellow	Model simulations which do not fail on any metrics, but where significant biases are present (either an orange or yellow assessment against one or more metrics) or where at least one metric was not assessed
Green	Model simulations performed satisfactorily against all metrics considered

We then discuss the available regional climate projections considering the assessment category of the driving GCMs and the specific downscaling method applied.

Having identified model simulations to be excluded, used with caution or those which are satisfactory to use, appropriate use of these projections will depend on the intended application and the question they are answering. Considering the range of future projections across the set of model simulations and using the decision-making matrix from Mc15 is applicable for deciding whether global or regional projections are suitable to use. However, this process requires tailoring the evaluation to a specific region of interest, and therefore this will be conducted in follow-on work when future projections of the model simulations in specific regions are considered.

3. Process-based evaluation of climate model projections for South Asia

3.1 Assessment of CMIP5 models for key processes governing extreme precipitation in South Asia

3.1.1 Characteristics of the monsoon flow

The first stage in our assessment is to consider the ability of model simulations to represent key characteristics of the South Asian monsoon. Mc15 undertake a thorough assessment of the ability of available CMIP5 models to capture the broad-scale characteristics of the monsoon flow. This qualitative assessment considers the 850 hPa winds and looks for the following key characteristics of the flow, as shown in the observations in Figure 3:

- strongest flow in the core of the Somali Jet,
- flow is largely
  - westerly across Indian Peninsula,
  - south-westerly across the Bay of Bengal (BoB),
  - westerly across continental southeast Asia,
  - southerly before reaching the Philippines.

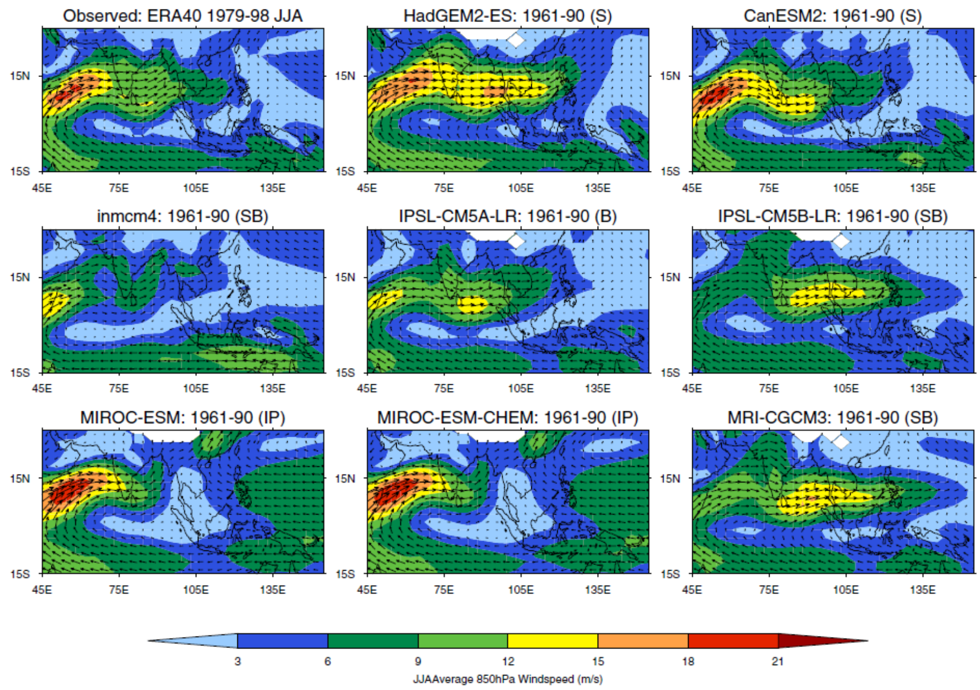


Figure 3 – Assessment of monsoon flow characteristics from McSweeney et al. (2015). Maps show monsoon circulation in 850 hPa flow for JJA for observations from ERA40 and a selection of CMIP5 models. Categories from the assessment are given in brackets for the CMIP5 models: IP – implausible, SB – significant biases, B – biases, S – satisfactory

They find that most models capture these broad-scale flow characteristics. However, two models are categorised as 'implausible' (*MIROC-ESM* and *MIROC-ESM-CHEM)*, three models have 'significant biases' (*inmcm4, IPSL-CM5B-LR,* and *MRI-CGCM3*), and eight models have 'biases' (*MIROC5, ACCESS1-3, FGOALS-g2, IPSL-CM5A-LR, GISS-E2-H, GISS-E2-H-CC, GISS-E2-R* and *GISS-E2-R-CC*) – see Table 3. A list of these models and the reasons for categorising them in this way is given in Table 5 and maps from a selection of these poorest performing models and some examples of better performing models are shown in Figure 3. The assessment focuses on flow characteristics but there are also three models identified as having biases in temperature (*bcc-cms1-1-m, ACCESS1-3* and *EC-EARTH*).

Table 5 – McSweeney et al. (2015) assessment of the poorest performing CMIP5 models in representing the monsoon circulation patterns. The categories assigned are from the decision-making matrix in Table 3. The coloured shading of the model names refers to the models that have been downscaled to generate regional projections, as per Table 2.

McSweeney et al. (2015) assessment category	**Model**	**Reason**
Implausible	MIROC-ESM	Unrealistic representation of the large-scale characteristics of the South West monsoon
	MIROC-ESM-CHEM	
Significant biases	INMCM4	Significantly weaker 850 hPa flow than observed
	IPSL-CM5B-LR	Very weak Somali jet combined with different direction flow over southern Asia to that observed (westerly not southerly around southern India becoming south-westerly in Bay of Bengal)
	MRI-CGCM3	
Biases	MIROC5	Flow is directed too southerly over continental southeast Asia
	ACCESS1-3	Underestimates the strength of the Somali jet
	FGOALS-g2	Flow significantly too westerly across BoB.
	IPSL-CM5A-LR	
	GISS-E2-H	Weak Somali jet and substantially too-strong southerly component of flow into the Bay of Bengal
	GISS-E2-H-CC	
	GISS-E2-R	
	GISS-E2-R-CC	
	Bcc-cms1-1-m	Largest warm biases
	ACCESS1-3	
	EC-EARTH	Cool bias (opposite to all other models). Significantly much weaker seasonal cycle of temperature than observations

A simulation from the *MIROC-ESM-CHEM* model is categorised as 'implausible' due to having an unrealistic representation of the large-scale characteristics of the monsoon. This model was statistical downscaled for the Nepal scenarios but not the dynamically downscaled regional climate model projections for the region (Table 2). No models found to have 'significant biases' have been downscaled over the region. Four models identified as having 'biases' (*MIROC5, IPSL-CM5A-LR, bcc-cms1-1-m* and *EC-EARTH)* have been downscaled for the region (Table 2) and these regional projections should therefore be further assessed and used with caution.

3.1.2 Inter-annual variability: ENSO-monsoon teleconnection

Sp13 consider several metrics to assess monsoon representation in 25 CMIP5 models. They assess how well the models capture the precipitation and 850 hPa wind climatologies, the annual rainfall cycle (results included in Table A1 the Appendix), and the representation of the inter-annual and intra-seasonal monsoon (shown in Table 6).

Table 6 – Assessment of model skill for 25 CMIP5 models, focused on representation of the Indian monsoon and the boreal summer intraseasonal variability (BSISO from Sp13. The best performing models identified in the original assessment are in bold font. The coloured shading for model names shows models that have been downscaled to generate regional projections, to correspond with Table 2. Red shading indicates the poorest performing models from Mc15.

	Indian monsoon		**BSISO**	
	AIR/N3.4	**Pr**	**Variance**	**Life cycle**
Observations	-0.533	0.798	0.995	0.893
Model				
CMIP5 Multi-Model Mean	--	0.616	0.888	0.766
bcc-csm-1	-0.25	-0.14	--	--
CanESM2	-0.273	0.014	0.846	0.651
CCSM4	-0.556	0.337	--	--
CNRM-CM5	-0.307	0.245	--	--
CSIRO-Mk3.6.0	-0.487	0.162	0.809	0.645
FGOALS-g2	-0.052	0.238	--	--
FGOALS-s2	0.114	0.096	0.734	0.608
GFDL-CM3	-0.442	0.192	--	--
GFDL-ESM2G	-0.289	0.251	0.753	0.643
GFDL-ESM2M	-0.187	0.251	--	--
GISS-E2-H	-0.094	0.254	--	--
GISS-E2-R	-0.366	0.379	--	--
HadCM3	-0.299	0.18	--	--
HadGEM2-CC	-0.335	-0.068	**0.857**	0.641
HadGEM2-ES	-0.344	0.216	**0.862**	0.651
INM-CM4	-0.033	0.11	0.639	0.562
IPSL-CM5A-LR	**-0.7**	**0.611**	0.791	0.654
IPSL-CM5A-MR	**-0.763**	**0.636**	0.827	0.635
MIROC-ESM	0.088	0.061	0.548	0.516
MIROC-ESM-CHEM	-0.104	0.045	0.554	0.528
MIROC4h	-0.327	**0.529**	0.736	0.625
MIROC5	-0.321	0.01	0.805	**0.691**
MPI-ESM-LR	-0.291	0.401	**0.874**	**0.681**
MRI-CGCM3	-0.274	0.338	0.782	0.628
NorESM1-M	**-0.69**	**0.522**	0.833	0.627

The focus of the Sp13 assessment was to identify the best performing models for the range of metrics considered. However, as we are interested in identifying models that do not adequately capture key monsoon processes, we relate these findings to the Mc15 assessment to identify the poorest performing models.

There are two different metrics used to assess model performance in capturing the inter-annual variability of the Indian monsoon in Sp13 which focus on the relationship between monsoon rainfall and ENSO. These are:

1. pattern correlations between JJAS anomalies of all-India rainfall and Niño3.4 SST (AIR/N3.4 in Table 2; further details given in Figure 4),
2. pattern correlations of JJAS precipitation anomalies with JJAS anomalies of Niño3.4 SST (Pr in Table 2; further details given in Figure 4).

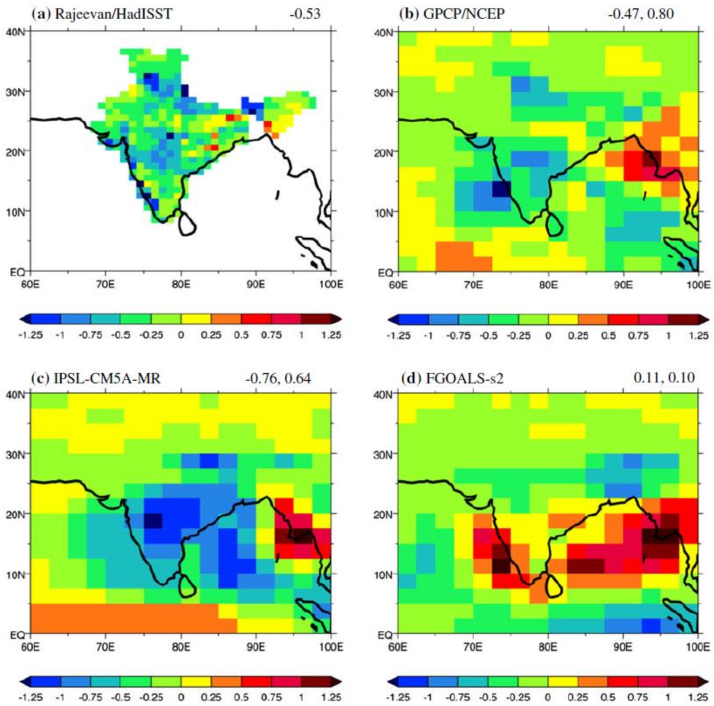


Figure 4 – Indian monsoon metrics from Sperber et al. (2013): panel a shows the observed AIR/N3.4 metric using the (Rajeevan et al., 2006) *rainfall observations for India (land points only) and the HadISST SST data* (Rayner et al., 2003) *for the period 1961-1999, panel b shows the observed Pr metric using GPCP vs SST from the NCEP/NCAR reanalysis (1979-2007), panels c and d show examples of the good I and poor (d) performing CMIP5 models . All metrics used the region 65°E–95°E, 7°N–30°N.*

The poorest performing models for these two metrics were identified in Mc15 and are shaded in red in Table 6. The five poorest performing models for the AIR/N3.4 metric are *FGOALS-g2, GISS-E2-H, inm-cm4, MIROC-ESM, and MIROC-ESM-CHEM*. Of these five models only one has been downscaled to generate regional projections (*MIROC-ESM-CHEM)*, statistical downscaled for the Nepal scenarios (Table 1).

Six models are identified as the poorest performing for the Pr metric; these are *bcc-csm-1, CanESM2, HadGEM2-CC, MIROC-ESM, MIROC-ESM-CHEM* and *MIROC5*. Again, the *MIROC-ESM-CHEM* model performs poorly, along with *bcc-csm1-1*, both of which have been statistically downscaled for the Nepal scenarios. In addition, *CanESM2* performs poorly for this metric and this model has been dynamically downscaled with two different RCMs in the CORDEX South Asia dataset. These GCM-RCM combinations should therefore be further assessed and used with caution.

3.1.3 Intra-seasonal variability: BSISO

Although most GCMs simulate the monsoon characteristics reasonably well, they struggle to capture the regional and intra-seasonal distribution of monsoon precipitation (Turner & Annamalai, 2012). However, CMIP5 showed improvements over the previous generation of models (CMIP3; Sp13), and initial results indicate further improvement in some early available CMIP6 models (Gusain et al., 2020). Caution is advised on the use of simple metrics, such as pattern correlations, to assess model performance of the BSISO as these do not capture the specifics of the spatial pattern and some models can have realistic spatial patterns but perform poorly in the pattern correlation, and vice versa (Sperber & Annamalai, 2008).

Sp13 and Sa13 use different metrics of the BSISO to assess how well CMIP5 models capture intra-seasonal variability in the monsoon. They assess different subsets of CMIP5 models; for some of the models the relevant data to construct these metrics was not available. Here we consider the range of metrics to identify models which have poor representation of the intra-seasonal variability and should be excluded in the development of climate information products showing changes to extreme precipitation events, where intra-seasonal variability is a key driving factor, e.g. in Nepal (Met Office, 2020b).

Results from the Sp13 assessment of intra-seasonal variability are shown in Table 6. They consider two metrics of the BSISO;

1. Variance: the 20-100 day bandpass filtered variance in outgoing longwave radiation (OLR, a proxy for convection) – maps of the observations and examples of some of the model simulations and pattern correlation with observations are shown in Figure 5.

2. Life cycle: the lag regression of the 20-100 day bandpass filtered OLR with a principal component time series (PC4; Sperber & Annamalai, 2008) of the BSISO over eight five day intervals of the BSISO life cycle (day -15 to day 20), as shown in Figure 6.

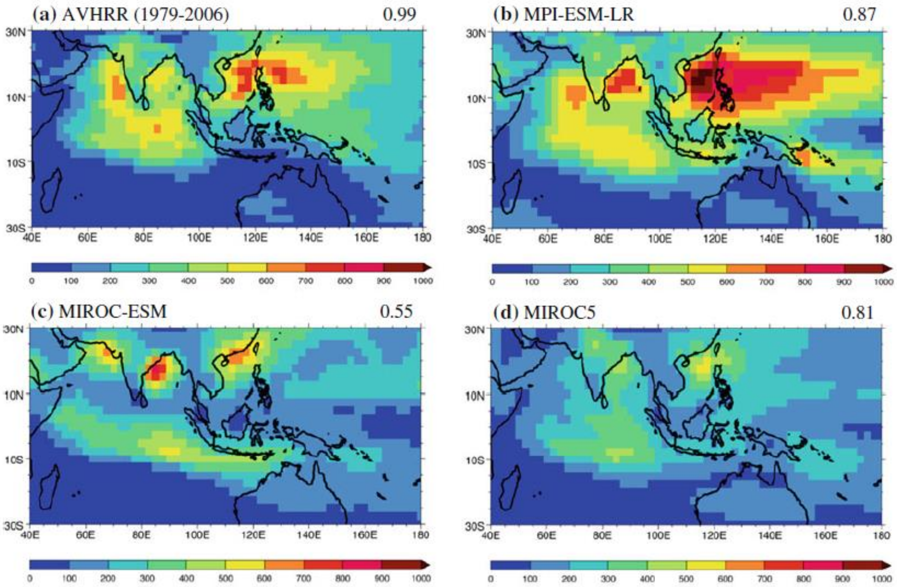


Figure 5 – BSISO variance metric from Sperber et al. (2013): panel a shows the observed BSISO variance metric using the AVHRR OLR observations and pattern correlation value, and panels b-d show the representations of this metric in the three poorest performing models.

OLR data was not available for 9 of the 25 CMIP5 models considered in the Sp13 study: four which have been downscaled for the South Asia region by a variety of models and methods (*bcc-csm-1, CNRM-CM5, GFDL-CM3* and *GFDL-ESM2M;* see Table 2), and five which have not been downscaled (*CCSM4, FGOALS-g2, GISS-E2-H, GISS-E2-R* and *HadCM3*).

For the remaining 16 models, the better performing models capture the observations for the BSISO variance metric better than the BSISO life cycle metric (~12% error from observations for the variance metric compared to ~22% for the life cycle metric; Table 6 and also see sorted skill scores in Table A2 in the Appendix). However, the poorer performing models represent 30-45% error from observations and are the same set of models for both metrics: these are *MIROC4h, FGOALS-s2, inm-cm4, MIROC-ESM-CHEM* and *MIROC-ESM* (Figure 5). The majority of models that have been regionally downscaled (for which there are data to construct these metrics) are ‘better performing models’ for these two metrics. The one model that poorly represents the monsoon variance for these metrics is *MIROC-ESM-CHEM,* which is consistent with previous assessments of this model’s suitability that indicate the model should be excluded in further analysis.

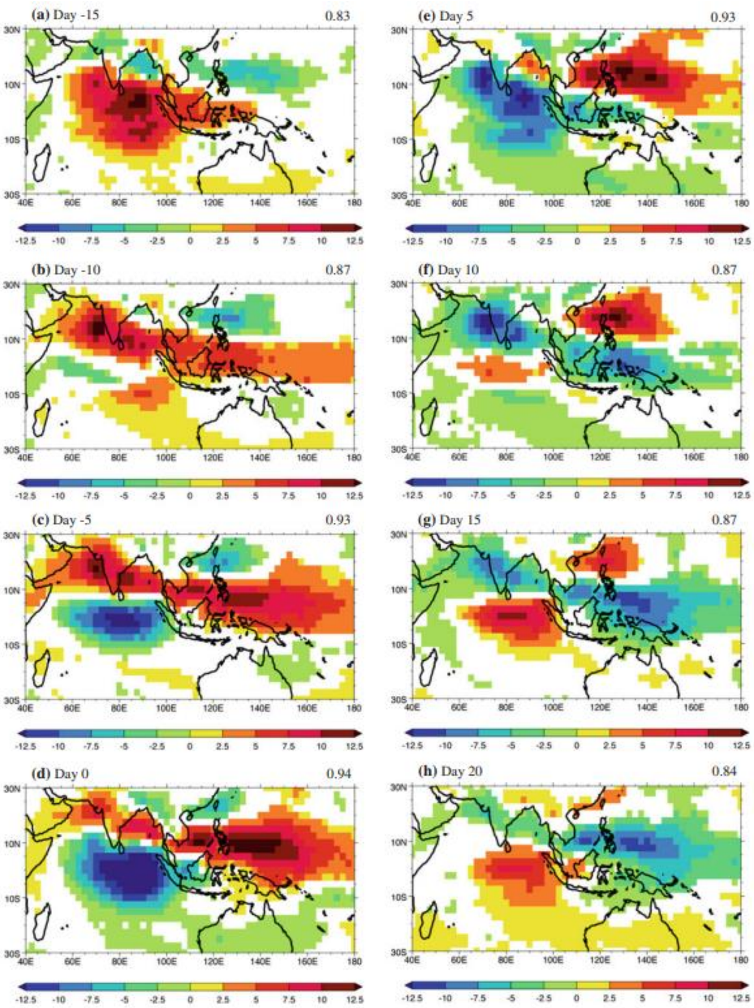


Figure 6 – Observed life cycle of the BSISO for day -15 to day 20 (a-h) used to assess model capability to simulate the BSISO life cycle in Sperber et al. (2013). The maps show the lag regression of the 20-100 day bandpass filtered OLR with a principal component time series of the BSISO (PC-4; Sperber & Annamalai, 2008) at 5 day intervals of the life cycle.

Sa13 conduct a more thorough investigation of the BSISO representation in 32 CMIP5 models. Their assessment focuses on the 20-100 day bandpass filtered precipitation anomalies rather than OLR, allowing more models to be used in the study compared to Sp13. In addition to variance and life cycle metrics they also consider the representation of the propagation features of the BSISO, i.e. the eastward propagating mode over the equatorial Indian Ocean, the northward propagating mode and the tilted rain band. They take a combined approach of qualitative assessment of spatial patterns, including the use of animations of model outputs to assess the evolution of the BSISO life cycle and also quantitative correlation metrics. This results in a binary assessment of whether the models capture these metrics or not (e.g. ‘yes’ or ‘no’, ‘correct’ or ‘wrong’). Results from the study are presented in Table 7.

Table 7 – Sabeerali et al. (2013) assessment of the capability of 32 CMIP5 models to represent key features in the BSISO. The coloured shading for the model names refers to the models that have been downscaled to generate regional projections, as per Table 2. Red shading indicates the metrics that are not captured well for models which fail on five or six of the six metrics. Orange shading indicates the metrics that are not captured well for models which fail on four of the size metrics. The top five models identified in Sabeerali et al. (2013) are highlighted with bold, italic font. The + symbol indicates the models which capture the three peak centres of monsoon variance.

Model	Realistic Spatial Pattern of BSISO Variance	Eastward Propagating Mode Over the Equatorial Indian Ocean	Realistic Northward Propagation of the BSISO	Realistic Space Time Structure of Northward Propagating Mode	Tilted Rain Band	Evolution of BSISO Life Cycle
ACCESS1.0	No	Yes	Yes	No	No	correct
ACCESS1.3	No	Yes	No	No	No	correct
bcc-csm-1	No	Yes	No	No	No	correct
BNU-ESM	No+	Yes	No	No	No	wrong
CanCM4	No	Yes	Yes	No	No	correct
CanESM2	No	Yes	Yes	No	No	correct
CCSM4	No	Yes	No	No	No	wrong
CESM1(BGC)	No	No	No	No	No	wrong
CESM1(FAST CHEM)	No	Yes	No	Yes	No	wrong
CMCC-CM	***No+***	***Yes***	***Yes***	***Yes***	***Yes***	***correct***
CNRM-CM5	No	Yes	Yes	Yes	Yes	wrong
CSIRO-Mk3.6.0	Yes	Yes	No	Yes	No	correct
FGOALS-s2	No	Yes	No	No	No	wrong
GFDL-CM3	***No***	***Yes***	***Yes***	***Yes***	***Yes***	***correct***
GFDL-ESM2G	No	Yes	Yes	No	Yes	correct
GFDL-ESM2M	No	Yes	Yes	No	Yes	correct
HadCM3	No	Yes	Yes	No	No	correct
HadGEM2-CC	No	Yes	No	No	No	correct
HadGEM2-ES	No	Yes	No	No	No	correct
INM-CM4	No	Yes	No	No	No	wrong
IPSL-CM5A-LR	***No***	***Yes***	***Yes***	***Yes***	***Yes***	***correct***
IPSL-CM5A-MR	No	Yes	No	No	No	correct
IPSL-CM5B-LR	No	Yes	No	No	No	wrong
MIROC4h	No	No	No	No	No	wrong
MIROC5	***No+***	***Yes***	***Yes***	***Yes***	***Yes***	***correct***
MIROC-ESM	No	Yes	No	Yes	No	correct
MIROC-ESM-CHEM	No	Yes	No	Yes	No	wrong
MPI-ESM-LR	***Yes+***	***Yes***	***Yes***	***Yes***	***Yes***	***correct***
MPI-ESM-MR	Yes+	Yes	Yes	No	Yes	wrong
MPI-ESM-P	Yes+	Yes	Yes	No	Yes	wrong
MRI-CGCM3	No	Yes	No	No	No	correct
NorESM1-M	No	Yes	No	No	No	wrong

Similar to the Sp13 assessment, Sa13 aim to identify the best performing models for the range of BSISO metrics considered. They identify five best performing models (bold text in Table 7); four of these are models have also been downscaled for South Asia (*GFDL-CM3* downscaled by *HadRM3P, IPSL-CM5A-LR* downscaled by *RegCM4-4, MIROC5* downscaled by *RCA4* and *RegCM4-7*, and *MPI-ESM-LR* downscaled by *COSMO-crCLIM, RCA4, REMO2009,* and *REMO2015*).

Here we consider the Sa13 results to identify the poorest performing models for the BSISO to exclude these from the set of available projections for studies focusing on extreme precipitation. They find that many of the CMIP5 models are unable to capture the spatial pattern of the BSISO variance in precipitation anomalies (Table 7). The eastward propagating mode is well captured in the majority of models, along with the life cycle. However, only a few models accurately capture metrics of the northward propagating mode and tiled rain band (Table 7).

Of the 32 models considered, two fail on all six of the BSISO metrics; these are *CESM1(BGC)* and *MIROC4h* (shaded red in Table 7). Six models only capture the eastward propagating mode and fail on the remaining 5 metrics, these are *BNU-ESM, CCSM4, FGOALS-s2, inm-cm4, IPSL-CM5B-LR, NorESM1-M* (also shaded red in Table 7). Of these models *NorESM1-M* is the only one which has been downscaled for South Asia (by *RCA4* in CORDEX, Table 2).

Eight models accurately capture the eastward propagating mode and one other metric but fail on the remaining four metrics. Six models capture the life cycle (*ACCESS1.3, bcc-csm-1, HadGEM2-CC, HadGEM2-ES, IPSL-CM5A-MR, MRI-CGCM3*) and the other two capture the realistic space time structure of the northward propagating mode (*CESM1(FAST-CHEM)*, *MIROC-ESM-CHEM*). These eight models are shaded orange in Table 7 for the metrics they fail on to indicate that they capture some aspects of the BSISO but not all. Of these models, three have been downscaled for South Asia: *bcc-csm1-1* and *MIROC-ESM-CHEM* for the Nepal scenarios, and *HadGEM2-ES* by *RCA4* in CORDEX and *HadRM3P* in DECCMA.

3.1.4 Summary assessment for CMIP5 models

Table 8 brings together all assessments of model performance for the characteristics of the monsoon flow (Section 3.1.1), inter-annual variability (Section 3.1.2) and intra-seasonal variability (Section 3.1.3). This table includes all CMIP5 models considered across the three studies and grey shading indicates where metrics are not considered for those models. A final assessment category colour has been assigned using the category definitions presented in Table 4.

We conclude that five CMIP5 models are not suitable for regional climate studies focused on extreme precipitation in South Asia. These are *FGOALS-s2, inm-cm4, MIROC4h, MIROC-ESM* and *MIROC-ESM-CHEM* (red shading in final column of Table 8). Of these five GCMs, one has been statistically downscaled to generate regional climate projections for the Nepal scenarios (*MIROC-ESM-CHEM)*. None of the other four models have been used in the regional downscaling studies considered: CORDEX South Asia, DECCMA and Nepal scenarios.

11 further models should be used with caution as they fail on at least one of the metrics considered. These are *bcc-csm-1, BNU-ESM, CanESM2, CCSM4, CESM1(BGC), FGOALS-g2, GISS-E2-H, HadGEM2-CC, IPSL-CM5B-LR, MIROC5, NorESM1-m*

(orange shading in final column of Table 8). Of these 11 models, two were downscaled for the Nepal scenarios (*bcc-csm*-1 and *CanESM2*[4]), and three have been dynamically downscaled with different RCMs over South Asia (*CanESM2, MIROC5* and *NorESM1-M*). Further investigation of the capability of simulations from these specific GCM-RCM combinations is advised before using these projections; an initial assessment is conducted in Section 3.2.

Four models perform satisfactorily for all metrics considered. These are *CSIRO-Mk3-6.0*, *GFDL-ESM2G, IPSL-CM5A-MR* and *MPI-ESM-LR* (green shading in final column of Table 8). All of these models have been downscaled over South Asia, except *GFDL-ESM2G.* However, a related model has been downscaled and performs satisfactorily in all metrics except for the Sp13 BSISO assessment where no data was available for *GFDL-ESM2M* but *GFDL-ESM2G* performed reasonably well. Four models also perform satisfactorily for all the metrics they were assessed against, but they were not able to be assessed against all metrics (*CNRM-CM5*, *GFDL-CM3, MPI-ESM-MR* and *MPI-ESM-P*).

Many of the CMIP5 models that have been used to drive RCM projections for South Asia perform satisfactorily for the majority of metrics considered (yellow shading in Table 8), and three of the models are satisfactory for all metrics (green shading in Table 8) making them good candidates for plausible projections of future changes in extreme precipitation.

[4] Note that the r2 and r5 ensemble members were used in the Nepal scenarios whereas it is assumed (and often not documented) that the assessments in Mc15, Sp13 and Sa13 used ensemble member r1.

Table 8 – Summary of all CMIP5 models assessed using metrics from McSweeney et al. (2015), Sabeerali et al. (2013) and Sperber et al. (2013), using results from Table 5-Table 7 (indicated with the yellow, orange and red shading). Grey shading shows models that were not assessed for the specific metrics. Shading for model names is used for models that have been downscaled to generate regional projections, as per Table 2. Our final assessment based on these metrics is shown in the final column: red shading - models that fail on two or more metrics; orange shading - models which fail on one metric; yellow shading - models which do not fail on any metrics but have significant biases or where at least one metric was not assessed; green shading – models that performed satisfactorily against all metrics.

CMIP5 model	**Downscaled by**	**Monsoon circulation** (Mc15)	**Pr-ENSO correlation** (Sp13)	**BSISO** (Sp13)	**BSISO** (Sa13)	**Final assess-ment**
ACCESS1.0	--					
ACCESS1.3	--	Biases			Only captures eastward mode and life cycle	
bcc-csm1-1	Nepal scenarios	Biases	Pr metric		Only captures eastward mode and life cycle	
bcc-csm1-1-m	--					
BNU-ESM	--				Only captures eastward mode	
CanCM4	--					
CanESM2	CORDEX WAS-44: RCA4 CORDEX WAS-44: RegCM4-4 Nepal scenarios (r2 & r5)		Pr metric			
CCSM4	--				Only captures eastward mode	
CESM1(BGC)	--				All metrics	
CESM1(FAST CHEM)	--				Only captures eastward and northward mode	
CESM1-CAM5	--					

Delivery Partners:

CESM1-WACCM	--					
CMCC-CESM						
CMCC-CM	--					
CMCC-CMS						
CNRM-CM5	CORDEX WAS-44: RCA4 CORDEX WAS-44: RegCM4-4 DECCMA: HadRM3P					
CSIRO-Mk3-6.0	CORDEX WAS-44: RCA4 CORDEX WAS-44: RegCM4-4					
EC-EARTH	CORDEX WAS-44: RCA4 CORDEX WAS-22: COSMO-crCLIM	Biases				
FGOALS-g2	--	Biases	AIR/N3.4 metric			
FGOALS-s2	--			Life cycle & variance metrics	Only captures eastward mode	
FIO-ESM						
GFDL-CM3	DECCMA: HadRM3P					
GFDL-ESM2G	--					
GFDL-ESM2M	CORDEX WAS-44: RCA4 CORDEX WAS-44: RegCM4-4 Nepal scenarios					
GISS-E2-H	--	Biases	AIR/N3.4 metric			
GISS-E2-H-CC	--	Biases				
GISS-E2-R	--	Biases				
GISS-E2-R-CC	--	Biases				
HadCM3	--					
HadGEM2-CC	--		Pr metric		Only captures eastward mode and life cycle	
HadGEM2-ES	CORDEX WAS-44: RCA4 CORDEX WAS-22: REMO2015 DECCMA: HadRM3P				Only captures eastward mode and life cycle	
HadGEM2-AO	--					

inm-cm4	--	Significant biases	AIR/N3.4 metric	Life cycle & variance metrics	Only captures eastward mode	
IPSL-CM5A-LR	CORDEX WAS-44: RegCM4-4	Biases				
IPSL-CM5A-MR	CORDEX WAS-44: RCA4					
IPSL-CM5B-LR	--	Significant biases			Only captures eastward mode	
MIROC4h	--			Life cycle & variance metrics	All metrics	
MIROC5	CORDEX WAS-44: RCA4 CORDEX WAS-22: RegCM4-7	Biases	Pr metric			
MIROC-ESM	--	Implausible	AIR/N3.4 and Pr metrics	Life cycle & variance metrics		
MIROC-ESM-CHEM	Nepal scenarios	Implausible	AIR/N3.4 and Pr metrics	Life cycle & variance metrics	Only captures eastward and northward	
MPI-ESM-LR	CORDEX WAS-44: RCA4 CORDEX WAS-44: REMO2009 CORDEX WAS-22: COSMO-crCLIM CORDEX WAS-22: REMO2015					
MPI-ESM-MR	CORDEX WAS-44: RegCM4-4 CORDEX WAS-22: RegCM4-7					
MPI-ESM-P	--					
MRI-CGCM3	--	Significant biases			Only captures eastward mode and life cycle	
NorESM1-M	CORDEX WAS-44: RCA4 CORDEX WAS-22: COSMO-crCLIM CORDEX WAS-22: RegCM4-7 CORDEX WAS-22: REMO2015				Only captures eastward mode	
NorESM1-ME	--					

Delivery Partners:

3.2 Discussion on regional climate model projections for South Asia

In this section we discuss the regionally downscaled climate model projections (Table 1) and their suitability in generating information about future changes in extreme precipitation in South Asia in the context of the process-based evaluation in Section 3.1. There are two main factors that influence regional climate projections: the method or model used to downscale the GCM projection, and the capability of the GCM to simulate the relevant processes.

Most of the regional climate projections for South Asia considered here are dynamically downscaled with RCMs. RCMs are limited area models which take the outputs from GCMs as lateral boundary conditions and dynamically downscale the global projections at a higher resolution and with improved orography. They are often thought to provide improved projections at a finer scale, but this is not always the case as the projections are highly dependent on the capability of the regional model itself and the driving GCM. For South Asia, in some cases the regional projections provide no improvement based on some measures, or even worsen the projections from the driving GCMs (Singh et al., 2017). For example, many RCM experiments for the South Asia region present dry biases over the monsoon season, but wet biases over the Himalayas that are sometimes larger than those seen in the driving GCM simulations (Janes et al., 2019). In this assessment we also consider a set of statistically downscaled projections for Nepal. Although statistical downscaling methods are comparatively fast and have low computational requirements, they are limited as they assume current statistical relationships will exist in the future and are heavily dependent on the driving GCM capability.

As presented in Section 3.1, there are several regional projections for the South Asia region which have been generated by downscaling GCM projections from models which do not effectively capture the key processes influencing extreme precipitation. In the following sections we discuss each of the downscaling experiments listed in Table 1 in turn. We consider:

1. The downscaling method applied,
2. The specific GCMs downscaled: how they were selected, and the evaluation assessment category assigned in Section 3.1, particularly those that were advised ‘exclude’ (i.e., red shading) or ‘use with caution’ (i.e., orange shading) as these are of most concern (more time is required to consider those in the yellow category),
3. Relevant literature evaluating downscaled observations and/or the GCM projections.

The aim of this discussion is to identify any regional climate model projections that should be excluded from any assessment of future changes in extreme precipitation, based on the assessments made here and in any relevant literature evaluating the regional projections. We also note that most literature evaluating regional climate projections involves comparing the precipitation outputs of the models to observations. As there are large variations in observed precipitation datasets (as discussed in Part III of this series of reports), it is unclear which is the best dataset to use for evaluation and different datasets and evaluation metrics are used by different evaluation assessments. Therefore, we consider the available evaluation assessments with caution and only aim to identify any poorly performing regional projections. Although out of scope for this work, evaluating model simulations against multiple observation datasets to account for observational uncertainty is important and is recommended for future model evaluation assessments.

3.2.1 CORDEX WAS-44

The aim of the CORDEX project is to generate consistent sets of RCM simulations across common domains (Giorgi & Gutowski, 2015). However, although the domains are consistent within the project, different RCMs have been run using different domains at different resolutions, and with different driving GCMs. The initial set of simulations at 0.44˚ resolution (around 50km) for the CORDEX WAS domain produced 17 RCM projections from three RCMs (*RCA4, RegCM4-4* and *REMO*, see Table 9). Various different GCMs were downscaled by the different RCMs and there was no systematic process for selecting the GCMs to downscale (Ashfaq et al., 2020).

Table 9 – RCM models used in CORDEX WAS-44 and the GCMs that were downscaled. The coloured shading indicates the assessment of the driving GCMs made in Table 8.

Project	**RCM**	**Driving GCM**
CORDEX WAS-44	RCA4	CSIRO-Mk3-6.0
		IPSL-CM5A-MR
		CNRM-CM5
		EC-EARTH
		GFDL-ESM2M
		HadGEM2-ES
		MPI-ESM-LR
		CanESM2
		MIROC5
		NorESM1-m
	RegCM4-4	CSIRO-Mk3-6.0
		CNRM-CM5
		GFDL-ESM2M
		IPSL-CM5A-LR
		MPI-ESM-MR
		CanESM2
	REMO2009	MPI-ESM-LR

RCA4

The *RCA4* model has been used to downscale ten CMIP5 projections for CORDEX WAS-44 at 0.44˚ (~50km) resolution (Table 9). There was no skill-based selection of these GCMs; they were the first ten models available for downscaling from the CMIP5 archive (Rana et al., 2020). The same set of ten GCMs have been downscaled for other CORDEX domains.

Iqbal et al. (2017) find that the mean seasonal climate and the position and strength of the jet streams are relatively well captured in the *RCA4* model. However, evaluation of the *RCA4* model when driven by ERA-Interim reanalysis shows a warm bias over northwest India and Pakistan and a cold bias over the Himalayan region when compared with the CRU observation dataset (Rana et al., 2020; Figure 7). Some of the biases in the GCMs are reduced in the downscaled simulations (such as for *EC-EARTH* which has a strong cold bias), but some are increased. For precipitation, the *RCA4* downscaling of ERA-Interim reanalysis removes the wet bias in the Himalayas and amplifies the dry bias in central India (Figure 8). This bias pattern is similar for all the downscaled GCM projections, irrespective of the bias in the driving GCM (Rana et al., 2020). Rana et al. (2020) also find that the *RCA4* model tends to reduce the spread in the projected changes in precipitation compared to the spread in the driving GCMs.

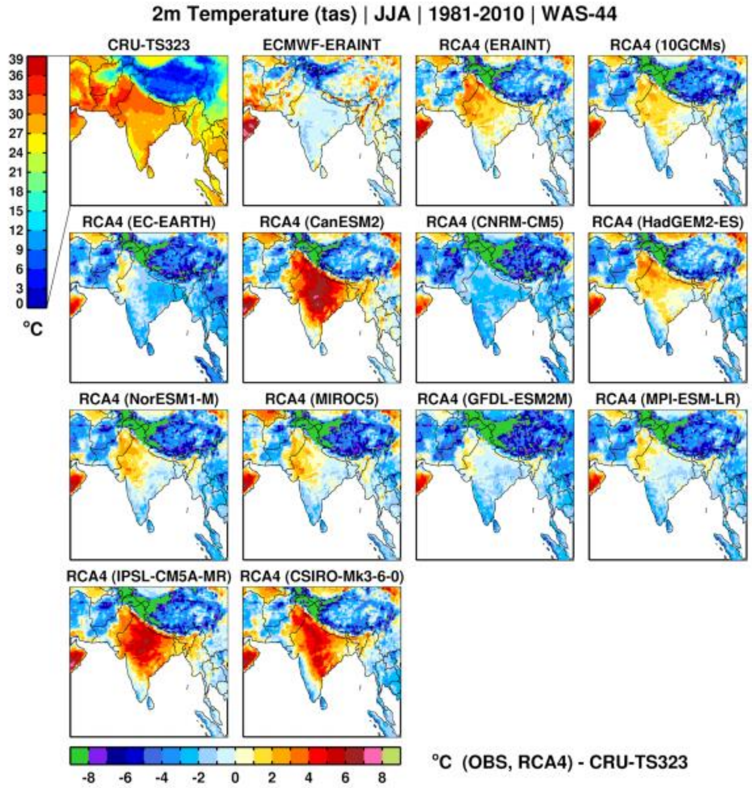


Figure 7 – Rana et al. (2020) assessment of JJA mean temperature anomalies for 1981-2010 (compared with CRU) for the RCA4 simulations from CORDEX WAS-44.

Of the ten RCM projections with *RCA4*, two of the driving GCMs performed satisfactorily for all metrics considered in Section 3.1 (green shading, Table 9), and five performed satisfactorily for most of the metrics they were assessed on (yellow shading, Table 9). The remaining three models failed on at least one metric (orange shading, Table 9) and therefore we assess these downscaled simulations further. The downscaled *CanESM2* simulation presents a significant warm bias across the region, and wet biases in northwest and central India. Interestingly the downscaled *IPSL-CM5A-MR* and *CSIRO-Mk3-6.0* projections show the same biases, and these are the two models in this subset of driving GCMs which performed satisfactorily for all metrics considered in Section 3.1.

The other two models advised to 'use with caution' in this subset are *MIROC5* and *NorESM1-m* and the downscaled simulations of these models show biases similar to the rest of the ensemble, suggesting that the RCM is dominating the biases.

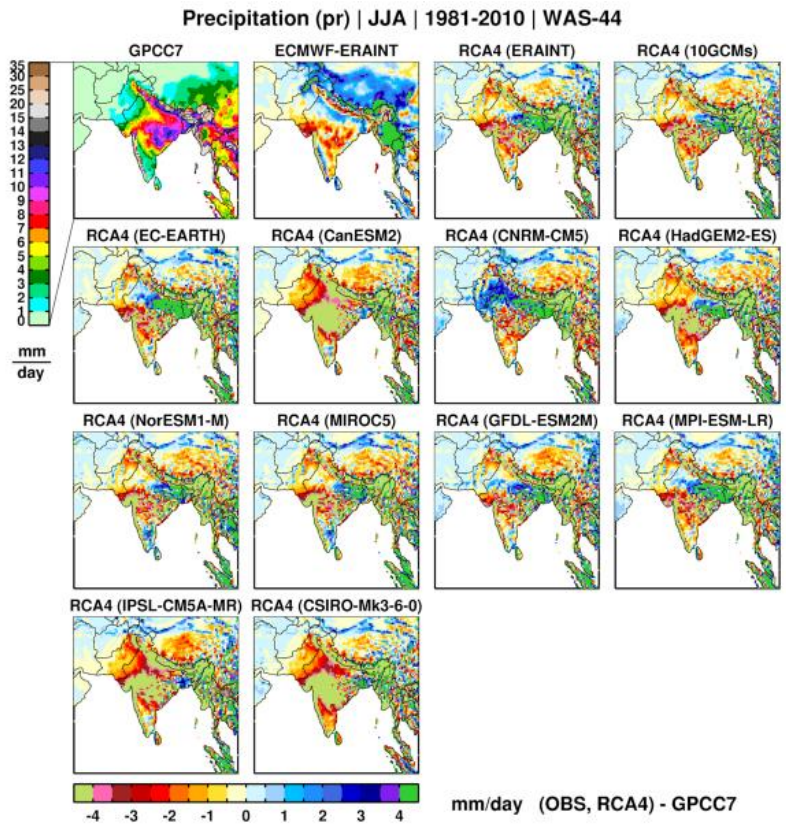


Figure 8 – Rana et al. (2020) assessment of JJA mean precipitation anomalies for 1981-2010 (compared with GPCC) for the RCA4 simulations from CORDEX WAS-44.

RegCM4-4

The *RegCM4-4* model has been used to generate six regional climate projections for CORDEX WAS-44 at 0.44˚ (~50km) resolution (Table 9). These projections used the *RegCM4-4* version of the model. As with *RCA4*, there is no evidence in the literature of any skill-selection for the six driving GCMs used in the WAS-44 projections,

Much of the published literature on *RegCM4-4* focuses on the performance of different convection parametrisation schemes used in the model to effectively simulate the monsoon (Kumar & Dimri, 2020; Maurya et al., 2018; Raju et al., 2015; Sinha et al., 2019). There is a lack of assessment of the specific individual model simulations used in WAS-44 so it is difficult to assess the biases in each of these individual simulations based on the published literature.

Only one of the six driving GCMs was advised to ‘use with caution’ from the assessment in Section 3.1 (orange shading, Table 9). Further assessment is therefore required to assess the validity of the *RegCM4-4 – CanESM2* simulation to examine if the projected changes are outliers when considering specific spatial regions, as per stage 2 in the Mc15 method (Figure 2, Table 3).

REMO2009

The *REMO2009* model has been used to generate one regional climate projection for CORDEX WAS-44 at 0.44˚ (~50km) resolution (Table 9). The GCM used, *MPI-ESM-LR,* performs satisfactorily for all metrics assessed in Section 3.1 (Table 9). Teichmann et al. (2013) evaluate this simulation and find a cold bias over the WAS domain, particularly over the Himalayas, compared to a slight warm bias over much of this region in the original *MPI-ESM-LR* simulation (Figure 9, top panels). The *REMO2009 – MPI-ESM-LR* simulation seems to reduce the dry bias over much of the South Asia region, but the wet bias over high elevation is still pronounced (Figure 9, bottom panels). An early onset in the monsoon is also found (Figure 10; Remedio et al., 2019).

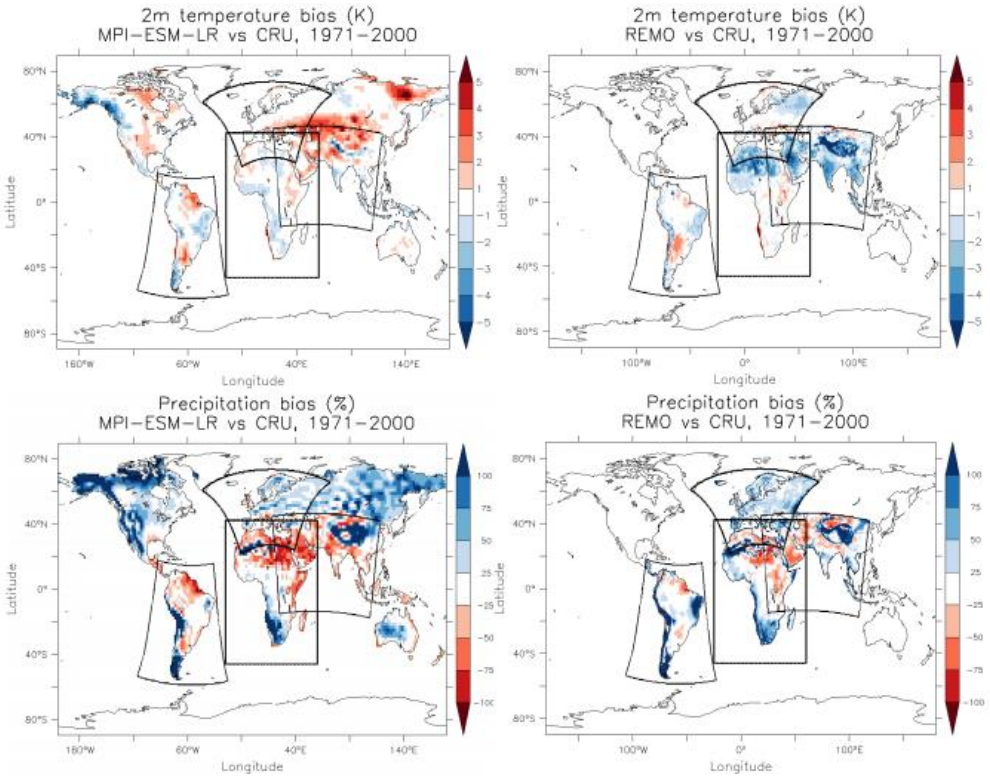


Figure 9 – Teichmann et al. (2013) evaluation of the REMO20–9 – MPI-ESM-LR WAS-44 simulation.

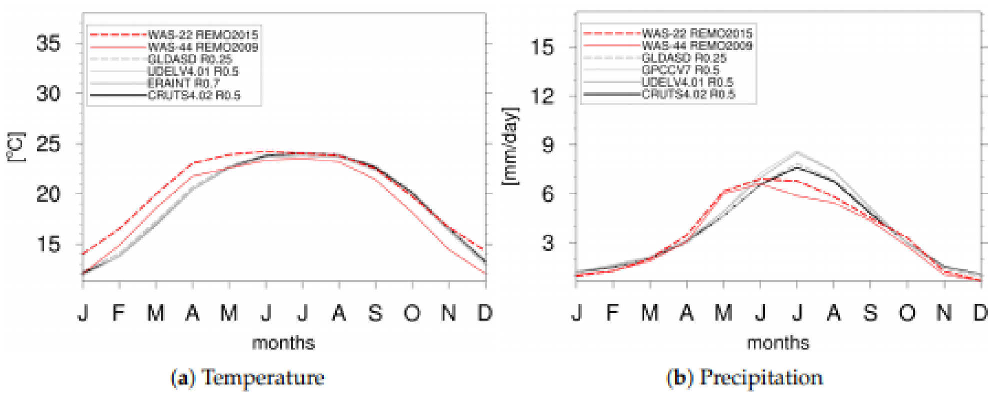


Figure 10 – Comparison of seasonal cycles in the REMO2009, REMO2015 models and observations, from Remedio et al. (2019).

3.2.2 CORDEX WAS-22

A more recent set of projections for the WAS domain are now available from the CORDEX-CORE project which aims to provide consistent RCM projections at 0.22˚ (around 25km) resolution for a subset of driving GCMs. In this project a set of GCMs were recommended that represent the range of equilibrium climate sensitivity (ECS) in the driving models (Ashfaq et al., 2020). These recommended GCMs were selected based on their ECS and applied across multiple CORDEX simulations in different domains. As such this method of selecting driving GCMs does not account for the ability of the models to capture important climate processes for the region.

The specific RCMs used the GCMs downscaled are shown in Table 10. Two of the RCMs used in the WAS-22 simulations were more recent versions of the *RegCM4* and *REMO* models (*RegCM4-7* and *REMO2015)*, and an additional RCM was also used, *COSMO-crCLIM*. The project recommended three driving CMIP5 GCMs as representatives of high (*HadGEM2-ES*), medium (*MPI-ESM-LR/MPI-ESM-MR*) and low (*NorESM1-m*) ECS[5], and 'backup' models (*MIROC5, EC-EARTH* and *GFDL-ESM2M* respectively) for each of the categories also identified. The nine WAS-22 simulations therefore come from three different RCMs forced with high, medium and low ECS GCMs, but the specific GCMs used are different for each RCM as some modelling centres chose to use different combinations of the recommended/backup models.

Table 10 – RCM models used in CORDEX WAS-22 and the GCMs that were downscaled . The coloured shading indicates the assessment of the driving GCMs made in Table 8.

Project	RCM	Driving GCM
CORDEX WAS-22	COSMO-crCLIM	MPI-ESM-LR
		EC-EARTH
		NorESM1-M
	RegCM4-7	MPI-ESM-MR
		MIROC5
		NorESM1-M
	REMO2015	MPI-ESM-LR
		HadGEM2-ES
		NorESM1-M

COSMO-crCLIM

The GCMs downscaled by the *COSMO-crCLIM* model included one model in each of the green, yellow and orange assessment categories (Table 10).

Maharana et al. (2020) consider the full set of WAS-22 projections across India and find that the *COSMO-crCLIM* model has the smallest biases over this region compared to the other two models (Figure 11). They conduct evaluation of the intra-seasonal

[5] https://cordex.org/experiment-guidelines/cordex-core/cordex-core-simulations/

variability in the model simulations by looking at the frequency of active and break spells in the monsoon (Figure 12). The ensemble mean of the *COSMO-crCLIM* model simulations significantly underestimates the frequency of active spells, and slightly overestimates the frequency of break spells (Figure 12).

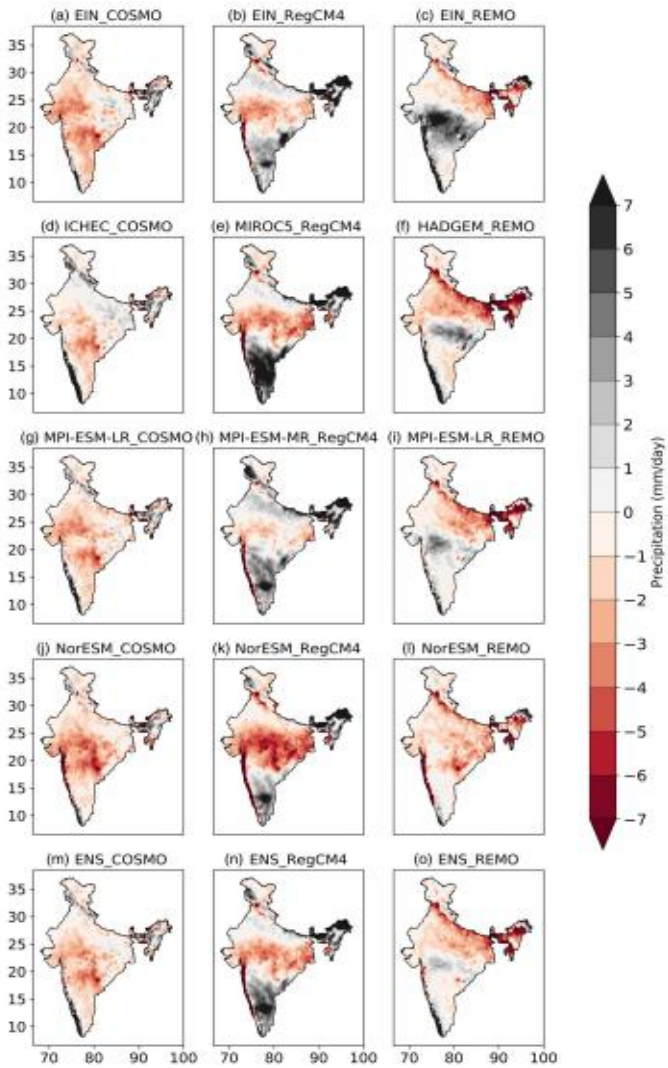


Figure 11 – Evaluation of precipitation from the CORDEX WAS-22 simulations over India from Maharana et al. (2020)

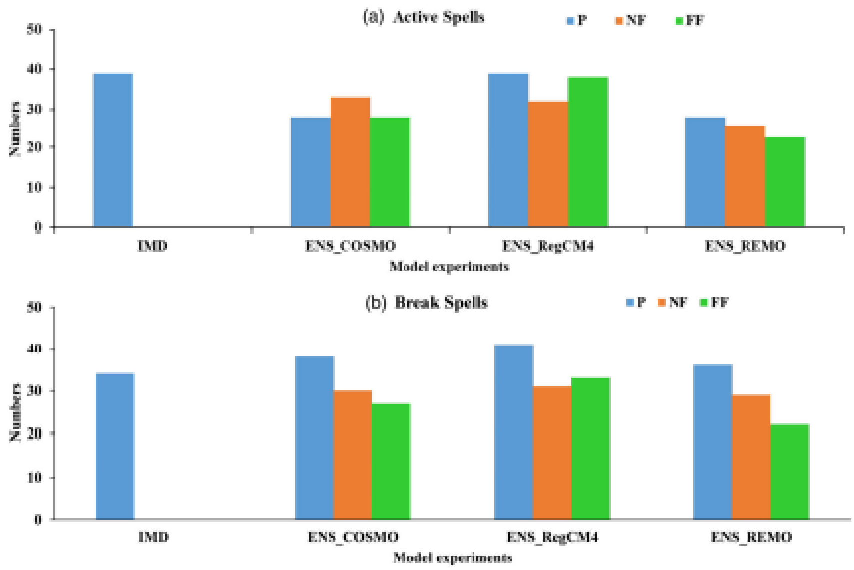


Figure 12 – Frequency of active and break spells over India in the CORDEX WAS-22 simulations for the present-day (P), near future (NF) and far future (FF), from Maharana et al. (2020).

RegCM4-7

The *RegCM4-7* model was used to downscale one GCM in the yellow category and two GCMs in the orange category (Table 10). Evaluation of the *RegCM4-7* model used in the CORDEX CORE WAS-22 simulations by Ashfaq et al. (2020) finds that the model tends to produce excessive convectively driven precipitation resulting in overestimation of the seasonality of precipitation over higher elevations, such as the Himalayas. It also tends to underestimate the observed seasonal precipitation in some parts of South Asia. The timing is generally well simulated, but with a slight delay in the onset of the monsoon. Ashfaq et al. (2020) only consider the ensemble mean of the *RegCM4-7* WAS-22 simulations, making the performance of the individual model simulations difficult to ascertain.

The Maharana et al. (2020) assessment of intra-seasonal variability shows the ensemble mean of the *RegCM4-7* model simulations represents the frequency of active spells reasonably well, but has the highest overestimation of break spells across the three model ensembles (Figure 12).

REMO2015

The GCMs downscaled by the *REMO2015* model included one model in each of the green, yellow and orange assessment categories (Table 10). Evaluation of the *REMO2015* simulations from CORDEX CORE in Remedio et al. (2019) found similar biases to the *REMO2009* model; warm and wet bias over most of the region, but cold bias over the Himalayas (Figure 13). As with *REMO2009,* there is an early onset in the monsoon (Figure 10; Remedio et al., 2019).

The Maharana et al. (2020) assessment of intra-seasonal variability shows the ensemble mean of the *REMO2015* model simulations significantly underestimates the frequency of active spells, similarly to the *COSMO-crCLIM* model, but represents the frequency of break spells reasonably well (Figure 12).

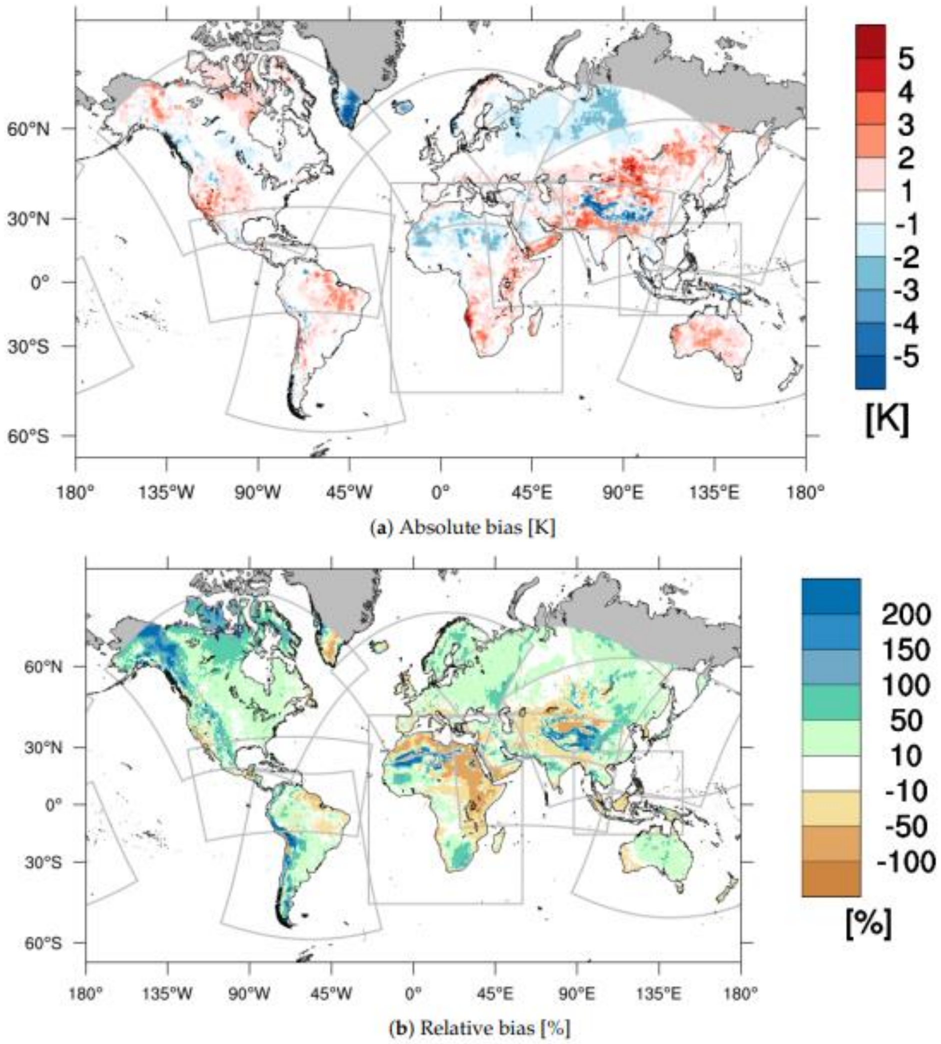


Figure 13 – Temperature (top panel) and precipitation (bottom panel) biases in REMO2015 from Remedio et al. (2019).

3.2.3 DECCMA – HadRM3P

As part of the DECCMA project, three RCM experiments were conducted by downscaling CMIP5 models with *HadRM3P* at 25km resolution (Janes et al., 2019). The GCM selection process followed Mc15 by assessing model performance in capturing the monsoon characteristics and range of future projections. The three GCMs used performed satisfactorily in this assessment of representation of processes driving extreme precipitation, however none of the model simulations were assessed for all metrics (i.e., yellow shading).

Table 11 – RCM models used in DECCMA and the GCMs that were downscaled. The coloured shading indicates the assessment of the driving GCMs made in Table 8.

Project	RCM	Driving GCM
DECCMA	HadRM3P	CNRM-CM5
		GFDL-CM3
		HadGEM2-ES

Janes et al. (2019) find that the three simulations with *HadRM3P* show a cold bias over much of the region, with the strongest cold biases over the Himalayas. Also, there is a slight dry bias during the monsoon season over most of the region, but a wet bias in the higher altitude parts of the Himalayas.

3.2.4 Nepal scenarios – statistical downscaling

Five GCM simulations have been statistically downscaled over Nepal to generate the Nepal Scenarios (Ministry of Forests and Environment, 2019; Table 12). They were selected based on the Lutz et al. (2016) envelope-based method for sub-selecting GCMs. This approach considered multiple ensemble members from the CMIP5 simulations, resulting in projections from two different ensemble members of the *CanESM2* GCM being downscaled, which have not been assessed in the studies considered in the GCM assessment in Section 3.1. It is therefore not possible to assign an assessment category for these GCM simulations (i.e., shaded grey in Table 12). The other GCM simulations that were statistically downscaled include one assigned the yellow category, one assigned the orange category, and one assigned the red category.

Table 12 – GCMs that were statistically downscaled in the Nepal Scenarios. The coloured shading indicates the assessment of the driving GCMs made in Table 8. The grey shading indicates that these specific ensemble members were not included in the assessment in Table 8 and therefore no assessment category is assigned.

Project	Driving GCM	Ensemble member for RCP4.5	Ensemble member for RCP8.5
Nepal Scenarios	GFDL-ESM2M	r1i1p1	r1i1p1
	bcc-csm1-1	r1i1p1	r1i1p1
	MIROC-ESM-CHEM	r1i1p1	r1i1p1
	CanESM2	r2i1p1	r5i1p1

4. Summary and recommendations

Understanding the current and future risk of extreme precipitation events is key to building resilience to climate change in key sectors in South Asia, such as the water and hydropower sectors (Met Office & ICIMOD, 2019). To help provide relevant and plausible climate information for these sectors, we have reviewed existing peer-reviewed literature to assess the ability of available climate model simulations to represent key processes that lead to extreme precipitation events.

Following identification of the key driving processes of extreme precipitation (Met Office, 2020b), this report presents an assessment of the ability of global climate model (GCM) projections to effectively simulate these processes. We follow the method presented in McSweeney et al. (2015) to identify those model simulations that poorly represent these processes and should therefore be excluded from future climate studies investigating changes to extreme precipitation in South Asia. We pay particular attention to the CMIP5 GCM simulations which have been downscaled to generate regional climate projections for South Asia. We also discuss the suitability of using these downscaled projections to explore changes in extreme precipitation, based on the capability of the driving GCMs and the downscaling method used.

The GCM assessment focuses on the models' representation of monsoon flow characteristics, the inter-annual variability in monsoon precipitation through the monsoon-El Niño Southern Oscillation teleconnection, and metrics for the Boreal Summer Intra-Seasonal Oscillation which governs intra-seasonal variability. An assessment category is assigned to each model simulation and a summary of the results is shown in Table 13.

Table 13 - Summary table of the evaluation of CMIP5 GCM simulations considered in this assessment, as presented in Table 8. Models are ordered by assessment category (green, yellow, orange and red) and a brief description of the reasons for the assigned category is provided. Blue shading of the GCM model names indicates those that have been downscaled to generate regional projections, as per Table 2.

Model	**Assessment category and reason**
CSIRO-Mk3-6.0	Satisfactory for all metrics
GFDL-ESM2G	Satisfactory for all metrics.
IPSL-CM5A-MR	Satisfactory for all metrics
MPI-ESM-LR	Satisfactory for all metrics.
ACCESS1.0	Satisfactory for all metrics but not considered in Sp13.
bcc-csm1-1-m	Satisfactory in Mc15 but not considered in Sp13 or Sa13.
CanCM4	Satisfactory in Sa13 but not considered in Mc15 or Sp13
CESM1-CAM5	Satisfactory in Mc15 but not considered in Sp13 or Sa13.
CESM1-WACCM	Satisfactory in Mc15 but not considered in Sp13 or Sa13.
CMCC-CESM	Satisfactory in Mc15 but not considered in Sp13 or Sa13.
CMCC-CM	Satisfactory in Mc15 and Sa13 but not considered in Sp13.
CMCC-CMS	Satisfactory in Mc15 but not considered in Sp13 or Sa13.
CNRM-CM5	Satisfactory for all metrics except Sp13 BSISO assessment where no data available.

EC-EARTH	Cold bias identified in Mc15 and not considered by other studies.
FIO-ESM	Satisfactory in Mc15 and Sa13 but not considered in Sp13.
GFDL-CM3	Satisfactory for all metrics except Sp13 BSISO assessment where no data available.
GFDL-ESM2M	Satisfactory for all metrics except Sp13 BSISO assessment where no data available.
GISS-E2-H-CC	Biases in Mc15 and not considered in Sp13 or Sa13.
GISS-E2-R	Biases in Mc15, satisfactory in Sp13 ENSO correlation metrics and Sa13 but not considered in Sp13 BSISO metrics.
GISS-E2-R-CC	Biases in Mc15 and not considered in Sp13 or Sa13.
HadCM3	Satisfactory in Mc15, Sp13 ENSO correlation metrics and Sa13 but not considered in Sp13 BSISO metrics.
HadGEM2-ES	Satisfactory for most metrics and only captures eastward mode and life cycle in Sa13 BSISO metrics.
HadGEM2-AO	Satisfactory in Mc15 but not considered in Sp13 or Sa13.
IPSL-CM5A-LR	Satisfactory for all metrics but Mc13 identify biases due to too westerly flow across BoB.
MPI-ESM-MR	Satisfactory for Mc15 and Sa13 but not considered in Sp13.
MPI-ESM-P	Satisfactory for Mc15 and Sa13 but not considered in Sp13.
NorESM1-ME	Satisfactory for Mc15 but not considered in Sp13 or Sa13.
ACCESS1.3	Only captures eastward mode and life cycle in Sa13 BSISO metrics, poor Somali jet identified in Mc15.
bcc-csm1-1	Fail on Sp13 Pr metric, only captures eastward mode and life cycle in Sa13 BSISO metrics, warm bias identified in Mc15.
BNU-ESM	Only captures eastward mode in Sa13 BSISO metrics, not considered in Sp13.
CanESM2	Fail on Sp13 Pr metric.
CCSM4	Only captures eastward mode in Sa13 BSISO metrics, not considered in Sp13 BSISO metrics.
CESM1(BGC)	Fail on all Sa13 metrics, satisfactory for Mc15 but not considered in Sp13
CESM1(FAST CHEM)	Only captures eastward and northward mode in Sa13, satisfactory in Mc15 but not considered in Sp13.
FGOALS-g2	Fail on Sp13 AIR/N3.4 metric, biases in Mc15 and not considered in Sa13.
GISS-E2-H	Fail on Sp13 AIR/N3.4 metric and not considered in Sp13 BSISO metric. Biases in Mc15 and not considered in Sa13.
HadGEM2-CC	Fail on Sp13 Pr metric, only captures eastward mode and life cycle in Sa13 BSISO metrics.
IPSL-CM5B-LR	Only captures eastward mode in Sa13 BSISO metrics, Mc15 identify significant biases – weak Somali jet and wrong flow direction.
MIROC5	Fail on Sp13 Pr metric and Mc15 identify biases due to too southerly flow over southeast Asia.
MRI-CGCM3	Only captures eastward mode and life cycle in Sa13 BSISO metrics, Mc15 identify significant biases due to weak Somali jet and wrong flow direction.
NorESM1-M	Only captures eastward mode of BSISO in Sa13 but satisfactory for other metrics
FGOALS-s2	Fail on Sp13 BSISO metrics, only captures eastward mode in Sa13, not considered in Mc15.

inm-cm4	Fail on Sp13 AIR/N3.4 and BSISO metrics, only captures eastward mode in Sa13 BSISO metrics, Mc15 identify significant biases – weak 850 hPa flow.
MIROC4h	Fail on Sp13 BSISO metrics and all Sa13 metrics, satisfactory for others.
MIROC-ESM	Fails on all Mc15, Sp13 assessments
MIROC-ESM-CHEM	Fails on all Mc15 and Sp13 metrics and only captures eastward and northward mode in Sa13 BSISO metrics.

We find that five CMIP5 simulations (*FGOALS-s2, inm-cm4, MIROC4h, MIROC-ESM* and *MIROC-ESM-CHEM*) were assigned the red category as they perform poorly across all metrics considered. It is recommended that these simulations are not to be used for future climate assessments of projected changes in extreme precipitation in South Asia as they do not accurately represent the key driving processes.

Of the remaining 41 CMIP5 simulations evaluated, four were assigned the green category as they passed on all metrics considered and are therefore suitable for use in assessments of future changes in extreme precipitation in South Asia. 23 of the model simulations passed on most metrics but were not assessed for all of them (yellow category) and 14 failed on at least one metric (orange category).

Recommendations resulting from this analysis include further assessment of the model simulations in the yellow assessment category to evaluate the missing metrics and to recategorize them as green or orange. Model simulations in the orange assessment category should be used with caution in any assessments of future changes of extreme precipitation in South Asia as they do not represent all of the key processes that drive extreme precipitation. Appropriate use of these GCM model simulations will depend upon the intended use of them, e.g., whether the direct outputs from these simulations are to be used or whether they are used to drive regionally downscaled projections or impacts models. It is advised to consider whether the future projections from these simulations in the region of interest represent outliers compared to other projections, following the Mc15 approach in Table 3.

Of the 15 GCM simulations have already been downscaled for the South Asia region, there are three in the green category, seven in the yellow category, four in the orange category and one in the red category (see Table 14).

Table 14 – Summary table of the CMIP5 GCM simulations that have been regionally downscaled by the different downscaling experiments considered in this assessment. The CMIP5 GCMs are shaded and ordered by their assessment category (as per Table 8, Table 13).

CMIP5 model	**Downscaling experiment**							
	CORDEX WAS-44			**CORDEX WAS-22**			**DECCMA project**	**Nepal scenarios**
	RCA4	**RegCM 4-4**	**REMO 2009**	**COSMO-crCLIM**	**RegCM 4-7**	**REMO 2015**	**HadRM3P**	**Statistical**
CSIRO-Mk3-6.0	✓	✓						
IPSL-CM5A-MR	✓							
MPI-ESM-LR	✓		✓	✓		✓		
CNRM-CM5	✓	✓					✓	
EC-EARTH	✓			✓				
GFDL-CM3							✓	
GFDL-ESM2M	✓	✓						✓
HadGEM2-ES	✓					✓	✓	
IPSL-CM5A-LR		✓						
MPI-ESM-MR		✓			✓			
bcc-csm1-1								✓
CanESM2	✓	✓						
MIROC5	✓				✓			
NorESM1-M	✓			✓	✓	✓		
MIROC-ESM-CHEM								✓
CanESM2 r2								✓
CanESM2 r5								✓

Based on the assessment of the GCMs driving those regional downscaling experiments we conclude:

- It is recommended that the downscaled projection of the *MIROC-ESM-CHEM* simulation (assigned the red category) is excluded from assessments of future changes in extreme precipitation as the driving GCM does not pass the assessment for capturing the relevant processes.
- The four GCM simulations in the orange category that have been regionally downscaled are *bcc-csm1-1, CanESM2, MIROC5* and *NorESM1-M*, see Table 14. As the driving GCMs have been advised to use with caution as they do not capture all of the key driving processes, regionally downscaled projections from this GCM simulations should also be used with caution. The Mc15 decision-making matrix is also applicable here regarding whether the regional downscaling represents an outlier in the future projections for the region of interest.

- The seven GCM simulations in the yellow category that have been regionally downscaled are *CNRM-CM5, EC-EARTH, GFDL-CM3, GFDL-ESM3M, HadGEM2-ES, IPSL-CM5A-LR* and *MPI-ESM-MR*, see Table 14. These model simulations require further analysis to evaluate the missing metrics in the GCM assessment and recategorize the GCM simulations as green or orange. In absence of this, the Mc15 decision-making matrix can also be applied regarding whether the regional downscaling represents an outlier in the future projections for the region of interest.

- The three GCM simulations in the green category that have been regionally downscaled are *CSIRO-Mk3-6.0, IPSL-CM5A-MR* and *MPI-ESM-LR.* Although the driving GCM simulations are assessed as being plausible making downscaled experiments of these good candidates for providing useful information about the projected changes in extreme precipitation, the specific method used to conduct the downscaling also plays a key part the suitability of these projections. It is therefore recommended that the Mc15 decision-making matrix is also applied regarding whether the regional downscaling represents an outlier in the future projections for the region of interest.

- We also note that two of the GCM simulations downscaled as part of the Nepal Scenarios are from ensemble members not assessed in the GCM assessment here. It is therefore not possible to provide an assessment of these specific simulations (shaded grey in Table 14), but we note that the first ensemble member of the *CanESM2* model was categorised as orange as it did not capture at least one of the metrics of the key driving processes.

This analysis represents part of the technical assessment stages of the climate information distillation process (Met Office, 2019) in selecting which climate datasets are fit for purpose and considering multi-model, multi-method climate information. The next steps involve re-engaging with partners and stakeholders to co-develop future climate information that helps better inform their decisions. With more information about the use of the projections and the specific regions and metrics of interest, a suitable selection of plausible futures can be identified using the assessment and recommendations presented here. Although this process is challenging given the needs of the users and the capability of the climate projections, this analysis provides a route to developing plausible future climate scenarios that can help address climate change challenges. Analysis to quantify the present-day risk of extreme rainfall is also being undertaken, and this work will help to bridge the timescales of the current and future risk, and to support communication activities with relevant stakeholders.

References

Ashfaq, M., Cavazos, T., Reboita, M. S., Torres-Alavez, J. A., Im, E. S., Olusegun, C. F., et al. (2020). Robust late twenty-first century shift in the regional monsoons in RegCM-CORDEX simulations. *Climate Dynamics*, (123456789). https://doi.org/10.1007/s00382-020-05306-2

Giorgi, F., & Gutowski, W. J. (2015). Regional Dynamical Downscaling and the CORDEX Initiative. *Annual Review of Environment and Resources*, *40*(1), 467–490. https://doi.org/10.1146/annurev-environ-102014-021217

Gusain, A., Ghosh, S., & Karmakar, S. (2020). Added value of CMIP6 over CMIP5 models in simulating Indian summer monsoon rainfall. *Atmospheric Research*, *232*(September 2019). https://doi.org/10.1016/j.atmosres.2019.104680

IPCC. (2013). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovern-mental Panel on Climate Change*. [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]: Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

Iqbal, W., Syed, F. S., Sajjad, H., Nikulin, G., Kjellström, E., & Hannachi, A. (2017). Mean climate and representation of jet streams in the CORDEX South Asia simulations by the regional climate model RCA4. *Theoretical and Applied Climatology*, *129*(1–2), 1–19. https://doi.org/10.1007/s00704-016-1755-4

Janes, T., McGrath, F., Macadam, I., & Jones, R. (2019). High-resolution climate projections for South Asia to inform climate impacts and adaptation studies in the Ganges-Brahmaputra-Meghna and Mahanadi deltas. *Science of the Total Environment*, *650*, 1499–1520. https://doi.org/10.1016/j.scitotenv.2018.08.376

Kumar, D., & Dimri, A. P. (2020). Sensitivity of convective and land surface parameterization in the simulation of contrasting monsoons over CORDEX-South Asia domain using RegCM-4.4.5.5. *Theoretical and Applied Climatology*, *139*(1–2), 297–322. https://doi.org/10.1007/s00704-019-02976-9

Lutz, A. F., ter Maat, H. W., Biemans, H., Shrestha, A. B., Wester, P., & Immerzeel, W. W. (2016). Selecting representative climate models for climate change impact studies: an advanced envelope-based selection approach. *International Journal of Climatology*, *36*(12), 3988–4005. https://doi.org/10.1002/joc.4608

Maharana, P., Kumar, D., Das, S., & Tiwari, P. R. (2020). Present and future changes in precipitation characteristics during Indian summer monsoon in CORDEX-CORE simulations, (September), 1–17. https://doi.org/10.1002/joc.6951

Maurya, R. K. S., Sinha, P., Mohanty, M. R., & Mohanty, U. C. (2018). RegCM4 model sensitivity to horizontal resolution and domain size in simulating the Indian summer monsoon. *Atmospheric Research*, *210*(April), 15–33. https://doi.org/10.1016/j.atmosres.2018.04.010

McSweeney, C. F., Jones, R. G., Booth, B. B. B., McSweeney, C. F., Jones, R. G., & Booth, B. B. B. (2012). Selecting Ensemble Members to Provide Regional Climate Change Information. *Journal of Climate*, *25*(20), 7100–7121. https://doi.org/10.1175/JCLI-D-11-00526.1

McSweeney, C. F., Jones, R. G., Lee, R. W., & Rowell, D. P. (2015). Selecting CMIP5 GCMs for downscaling over multiple regions. *Climate Dynamics*, *44*(11–12), 3237–3260. https://doi.org/10.1007/s00382-014-2418-8

Met Office. (2019). *ARRCC CARISSA: Practical guide to distillation in ARRCC CARISSA*.

Met Office. (2020a). *ARRCC CARISSA Work Stream 4. Pilot study progress report: Climate information for the hydropower sector in Nepal*.

Met Office. (2020b). *ARRCC CARISSA Work Stream 4. Understanding and quantifying extreme precipitation events in South Asia: Part I – Understanding climate drivers of extreme precipitation through case studies*.

Met Office, & ICIMOD. (2019). *Regional workshop on climate services for the water and hydropower sectors in South Asia, Proceedings*. Retrieved from https://lib.icimod.org/record/34708

Ministry of Forests and Environment. (2019). *Climate change scenarios for Nepal for National Adaptation Plan (NAP)*. Kathmandu.

Rajeevan, M., Bhate, J., Kale, J., & Lal, B. (2006). High resolution daily gridded rainfall data for the Indian region: analysis of break andactive monsoon spells. *Current Science*, *91*, 296–306.

Raju, P. V. S., Bhatla, R., Almazroui, M., & Assiri, M. (2015). Performance of convection schemes on the simulation of summer monsoon features over the South Asia CORDEX domain using RegCM-4.3. *International Journal of Climatology*, *35*(15), 4695–4706. https://doi.org/10.1002/joc.4317

Rana, A., Nikulin, G., Kjellström, E., Strandberg, G., Kupiainen, M., Hansson, U., & Kolax, M. (2020). Contrasting regional and global climate simulations over South Asia. *Climate Dynamics*, *54*(5–6), 2883–2901. https://doi.org/10.1007/s00382-020-05146-0

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., & Rowell, D. P. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*, 4407.

Remedio, A. R., Teichmann, C., Buntemeyer, L., Sieck, K., Weber, T., Rechid, D., et al. (2019). Evaluation of new CORDEX simulations using an updated köppen-trewartha climate classification. *Atmosphere*, *10*(11), 1–25. https://doi.org/10.3390/atmos10110726

Sabeerali, C. T., Ramu Dandi, A., Dhakate, A., Salunke, K., Mahapatra, S., & Rao, S. A. (2013). Simulation of boreal summer intraseasonal oscillations in the latest CMIP5 coupled GCMs. *Journal of Geophysical Research Atmospheres*, *118*(10), 4401–4420. https://doi.org/10.1002/jgrd.50403

Singh, S., Ghosh, S., Sahana, A. S., Vittal, H., & Karmakar, S. (2017). Do dynamic regional models add value to the global model projections of Indian monsoon? *Climate Dynamics*, *48*(3–4), 1375–1397. https://doi.org/10.1007/s00382-016-3147-y

Sinha, P., Maurya, R. K. S., Mohanty, M. R., & Mohanty, U. C. (2019). Inter-comparison and evaluation of mixed-convection schemes in RegCM4 for Indian summer monsoon simulation. *Atmospheric Research*, *215*(August 2018), 239–252. https://doi.org/10.1016/j.atmosres.2018.09.002

Sperber, K. R., & Annamalai, H. (2008). Coupled model simulations of boreal summer intraseasonal (30-50 day) variability, Part 1: Systematic errors and caution on use of metrics. *Climate Dynamics*, *31*(2–3), 345–372. https://doi.org/10.1007/s00382-008-0367-9

Sperber, K. R., Annamalai, H., Kang, I.-S., Kitoh, A., Moise, A., Turner, A., et al. (2013a). 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

Sperber, K. R., Annamalai, H., Kang, I. S., Kitoh, A., Moise, A., Turner, A., et al. (2013b). *The Asian summer monsoon: An intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century*. *Climate Dynamics* (Vol. 41). https://doi.org/10.1007/s00382-012-1607-6

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Teichmann, C., Eggert, B., Elizalde, A., Haensler, A., Jacob, D., Kumar, P., et al. (2013). How does a regional climate model modify the projected climate change signal of the driving GCM: A study over different CORDEX regions using REMO. *Atmosphere*, *4*(2), 214–236. https://doi.org/10.3390/atmos4020214

Turner, A. G., & Annamalai, H. (2012). Climate change and the South Asian summer monsoon. *Nature Climate Change*, *2*(8), 587–595. https://doi.org/10.1038/nclimate1495

Appendix

Table A1 – Assessment of model skill of 25 CMIP5 models to represent the June-September climatology and climatological annual cycle for South Asia from Sperber et al., (2013). The top five models for each metric across the wider assessment of both CMIP5 and CMIP3 models are in bold font, therefore not all the best performing models are listed in this table as some of the them come from the earlier CMIP3 generation of climate models, not considered here.

	Climatology		**Climatological annual cycle rainfall**						
Model	**Pr**	**850 hPa**	**T–Lat**	**Onset**	**Peak**	**Withd.**	**Duration**	**Hit rate**	**Threat**
Observations	0.927	0.986	0.887	0.748	0.834	0.83	0.671	0.893	0.744
CMIP5 MMM	0.898	0.976	0.674	0.664	0.786	0.792	0.605	0.844	0.625
bcc-csm-1	0.808	0.928	0.338						
CanESM2	0.815	0.951	0.552	0.298	0.451	0.543	0.164	0.782	0.517
CCSM4	**0.849**	0.952	0.678	0.581	**0.717**	**0.798**	**0.57**	**0.836**	**0.619**
CNRM-CM5	**0.852**	**0.974**	0.567	**0.674**	0.638	0.75	**0.656**	0.796	0.513
CSIRO-Mk3.6.0	0.713	0.896	0.232	0.006	0.451	0.729	0.331	0.762	0.497
FGOALS-g2	0.766	0.923	0.455						
FGOALS-s2	0.807	0.916	0.613	**0.601**	0.596	0.649	0.531	0.812	0.537
GFDL-CM3	**0.844**	0.941	**0.742**	0.458	0.407	0.546	0.406	0.796	0.532
GFDL-ESM2G	0.821	**0.955**	**0.727**	0.37	0.56	0.66	0.328	**0.841**	**0.615**
GFDL-ESM2M	0.828	**0.958**	0.676	0.49	**0.714**	0.73	0.383	0.824	0.586
GISS-E2-H	0.631	0.902	0.318						
GISS-E2-R	0.73	0.912	0.235						
HadCM3	0.773	0.931	0.55	0.555	0.447	0.519	0.452	**0.873**	**0.675**
HadGEM2-CC	0.795	0.927	0.376	0.526	0.659	0.634	0.317	0.777	0.543
HadGEM2-ES	0.8	0.933	0.356	0.562	0.62	0.648	0.367	0.769	0.538
INM-CM4	0.742	0.864	0.561	0.153	0.616	0.649	0.224	0.81	0.56
IPSL-CM5A-LR	0.797	0.926	0.442	0.399	0.54	0.712	0.482	0.798	0.515
IPSL-CM5A-MR	0.809	0.935	0.501	0.421	0.575	**0.769**	**0.591**	0.787	0.501
MIROC-ESM	0.617	0.824	0.518	0.391	0.61	0.666	0.394	0.756	0.434
MIROC-ESM-CHEM	0.642	0.831	0.538	0.518	0.669	0.653	0.423	0.752	0.433
MIROC4h	0.802	0.94	0.573	**0.674**	0.626	0.766	**0.62**	**0.843**	**0.611**

MIROC5	0.842	0.94	**0.778**	0.362	**0.778**	**0.851**	**0.652**	0.808	0.531
MPI-ESM-LR	0.792	0.949	0.664	0.316	0.579	0.652	0.472	0.781	0.535
MRI-CGCM3	0.752	0.886	0.195	0.024	0.619	0.535	-0.014	0.751	0.465
NorESM1-M	**0.848**	0.913	0.634	0.558	**0.723**	**0.791**	0.565	**0.838**	**0.624**

Table A2 - Sorted skill scores and percentage error from observations for Indian monsoon and BSISO evaluation from Sperber et al. 2013)

Indian monsoon – AIR/N3.4			**Indian monsoon - Pr**			**BSISO - Variance**			**BSISO – life cycle**		
Model	**Skill score**	**% error from obs**	**Model**	**Skill score**	**% error from obs**	**Model**	**Skill score**	**% error from obs**			
Observations	*-0.533*		*Observations*	*0.798*		*Observations*	*0.995*		*Observations*	*0.893*	
IPSL-CM5A-MR	**-0.763**	**43.2**	IPSL-CM5A-MR	**0.636**	-20.3	*CMIP5 MMM*	*0.903*	*-9.2*	*CMIP5 MMM*	*0.766*	*-14.2*
IPSL-CM5A-LR	**-0.7**	**31.3**	*CMIP5 MMM*	*0.616*	*-22.8*	MPI-ESM-LR	**0.874**	**-12.2**	MIROC5	**0.691**	**-22.6**
NorESM1-M	**-0.69**	**29.5**	IPSL-CM5A-LR	**0.611**	-23.4	HadGEM2-ES	**0.862**	**-13.4**	MPI-ESM-LR	**0.681**	**-23.7**
CCSM4	-0.556	4.3	MIROC4h	**0.529**	-33.7	HadGEM2-CC	**0.857**	**-13.9**	IPSL-CM5A-LR	0.654	-26.8
CSIRO-Mk3.6.0	-0.487	-8.6	NorESM1-M	**0.522**	-34.6	CanESM2	0.846	-15.0	HadGEM2-ES	0.651	-27.1
GFDL-CM3	-0.442	-17.1	MPI-ESM-LR	0.401	-49.7	NorESM1-M	0.833	-16.3	CanESM2	0.651	-27.1
GISS-E2-R	-0.366	-31.3	GISS-E2-R	0.379	-52.5	IPSL-CM5A-MR	0.827	-16.9	CSIRO-Mk3.6.0	0.645	-27.8
HadGEM2-ES	-0.344	-35.5	MRI-CGCM3	0.338	-57.6	CSIRO-Mk3.6.0	0.809	-18.7	GFDL-ESM2G	0.643	-28.0
HadGEM2-CC	-0.335	-37.1	CCSM4	0.337	-57.8	MIROC5	0.805	-19.1	HadGEM2-CC	0.641	-28.2
MIROC4h	-0.327	-38.6	GISS-E2-H	0.254	-68.2	IPSL-CM5A-LR	0.791	-20.5	IPSL-CM5A-MR	0.635	-28.9
MIROC5	-0.321	-39.8	GFDL-ESM2M	0.251	-68.5	MRI-CGCM3	0.782	-21.4	MRI-CGCM3	0.628	-29.7
CNRM-CM5	-0.307	-42.4	GFDL-ESM2G	0.251	-68.5	GFDL-ESM2G	0.753	-24.3	NorESM1-M	0.627	-29.8
HadCM3	-0.299	-43.9	CNRM-CM5	0.245	-69.3	MIROC4h	0.736	-26.0	MIROC4h	0.625	-30.0
MPI-ESM-LR	-0.291	-45.4	FGOALS-g2	0.238	-70.2	FGOALS-s2	0.734	-26.2	FGOALS-s2	0.608	-31.9
GFDL-ESM2G	-0.289	-45.8	HadGEM2-ES	0.216	-72.9	INM-CM4	0.639	-35.8	INM-CM4	0.562	-37.1
MRI-CGCM3	-0.274	-48.6	GFDL-CM3	0.192	-75.9	MIROC-ESM-CHEM	0.554	-44.3	MIROC-ESM-CHEM	0.528	-40.9
CanESM2	-0.273	-48.8	HadCM3	0.18	-77.4	MIROC-ESM	0.548	-44.9	MIROC-ESM	0.516	-42.2
BCC-CSM-1	-0.25	-53.1	CSIRO-Mk3.6.0	0.162	-79.7	BCC-CSM-1			BCC-CSM-1		
GFDL-ESM2M	-0.187	-64.9	INM-CM4	0.11	-86.2	CCSM4			CCSM4		
MIROC-ESM-CHEM	-0.104	-80.5	FGOALS-s2	0.096	-88.0	CNRM-CM5			CNRM-CM5		
GISS-E2-H	-0.094	-82.4	MIROC-ESM	0.061	-92.4	FGOALS-g2			FGOALS-g2		
FGOALS-g2	-0.052	-90.2	MIROC-ESM-CHEM	0.045	-94.4	GFDL-CM3			GFDL-CM3		
INM-CM4	-0.033	-93.8	CanESM2	0.014	-98.2	GFDL-ESM2M			GFDL-ESM2M		
MIROC-ESM	0.088	-117	MIROC5	0.01	-98.7	GISS-E2-H			GISS-E2-H		
FGOALS-s2	0.114	-121	HadGEM2-CC	-0.068	-109	GISS-E2-R			GISS-E2-R		
CMIP5 MMM			BCC-CSM1-1	-0.14	-118	HadCM3			HadCM3		