

# SCIENTIFIC REPORTS



OPEN

## Precipitation Biases in CMIP5 Models over the South Asian Region

Raju Pathak<sup>1</sup>, Sandeep Sahany<sup>1,2</sup>, Saroj Kanta Mishra<sup>1,2</sup> & S. K. Dash<sup>2</sup>

Using data from 33 models from the CMIP5 historical and AMIP5 simulations, we have carried out a systematic analysis of biases in total precipitation and its convective and large-scale components over the south Asian region. We have used 23 years (1983–2005) of data, and have computed model biases with respect to the PERSIANN-CDR precipitation (with convective/large-scale ratio derived from TRMM 3A12). A clustering algorithm was applied on the total, convective, and large-scale precipitation biases seen in CMIP5 models to group them based on the degree of similarity in the global bias patterns. Subsequently, AMIP5 models were analyzed to conclude if the biases were primarily due to the atmospheric component or due to the oceanic component of individual models. Our analysis shows that the set of individual models falling in a given group is somewhat sensitive to the variable (total/convective/large-scale precipitation) used for clustering. Over the south Asian region, some of the convective and large-scale precipitation biases are common across groups, emphasizing that although on a global scale the bias patterns may be sufficiently different to cluster the models into different groups, regionally, it may not be true. In general, models tend to overestimate the convective component and underestimate the large-scale component over the south Asian region, although with spatially varying magnitudes depending on the model group. We find that the convective precipitation biases are largely governed by the closure and trigger assumptions used in the convection parameterization schemes used in these models, and to a lesser extent on details of the individual cloud models. Using two different methods: (i) clustering, (ii) comparing the bias patterns of models from CMIP5 with their AMIP5 counterparts, we find that, in general, the atmospheric component (and not the oceanic component through biases in SSTs and atmosphere-ocean feedbacks) plays a major role in deciding the convective and large-scale precipitation biases. However, the oceanic component has been found important for one of the convective groups in deciding the convective precipitation biases (over the maritime continent).

The multi-model mean (widely used in climate change projections by the Intergovernmental Panel on Climate Change) assumes inter-model statistical independence (SI)<sup>1,2</sup>. However, the SI assumption is not quite accurate (Pennell and Reichler<sup>3</sup>; Pincus *et al.*<sup>4</sup>). For example, Pennell and Reichler<sup>3</sup> reported an effective ensemble size much smaller than the actual number of CMIP3 models. SI violations are very large in CMIP5 than CMIP3 models due to multiple reasons, especially similarities in numerical schemes, physical parameterizations, etc. (Pincus *et al.*<sup>4</sup>; Masson and Knutti<sup>5</sup>). For example, IPSL-CM5A-LR, IPSL-CM5A-MR, and IPSL-CM5B-LR were developed by slight modification in resolution and the atmospheric component. Similar is the case with GISS-E2H and GISS-E2R. GFDL-ESM2M and GFDL-ESM2G primarily differ only in their ocean components. MIROC-ESM and MIROC-ESM-CHEM differ in ocean biogeochemistry and atmospheric chemistry. MPI-ESM-LR and MPI-ESM-P were developed from their predecessor by changing resolution and neglecting the feedback between dynamic vegetation and land use.

Although sharing of components between various model versions from the same center may sound obvious, even different modeling centers share large fraction of the model code. For example, CNRM and EC-EARTH use similar atmospheric components (AEPEGE/IFS/ECMWF); ACCESS uses the HadGEM2 atmospheric component; FGOALS uses several physical parameterizations from CCSM; NorESM uses some key components of

<sup>1</sup>Centre for Atmospheric Sciences, Indian Institute of Technology Delhi, Delhi, India. <sup>2</sup>DST Centre of Excellence in Climate Modelling, Indian Institute of Technology Delhi, Delhi, India. Correspondence and requests for materials should be addressed to S.S. (email: [ssahany@cas.iitd.ac.in](mailto:ssahany@cas.iitd.ac.in))

CESM1. All such similarities in codes, schemes and concepts among the various CMIP5 models have been found to be a reason for common biases in the simulated fields<sup>5–8</sup>. Past studies on model similarity and genealogy have reported the use of similar ocean component to be less relevant than the use of similar atmospheric component in producing large model similarity in surface climatology<sup>5,8</sup>. In simpler words, it means that the surface climatology from two models with similar atmospheric components but different ocean components would have greater commonality than that of two models with similar ocean components but different atmospheric components.

Common biases in the simulated fields are found to be usually large in precipitation simulation, with largest over the south Asian region during the southwest summer monsoon<sup>9–13</sup>. Deficiencies in precipitation simulation by models have primarily been due to the persistent errors in the simulation of location and timing, and improving the spatial and seasonal features would provide a better model agreement in historical and future<sup>14</sup>. Regionally, the Arabian Sea cold SST bias during the pre-monsoon season in some of the CMIP5 models has also been found to be important for the simulation of south Asian summer monsoon precipitation<sup>15</sup>. In the context of climate change, precipitation is expected to increase in the future due to increased human influence and anthropogenic emissions<sup>16,17</sup>, and thus changing the water cycle<sup>18</sup>. Reliable precipitation simulation over the south Asian region is crucial for society, and for mitigation and adaptation strategies due to the changes in its pattern and variability under climate change<sup>19–21</sup>.

There have been studies on model genealogy as discussed above, using total precipitation as the variable of interest, but common biases across models in the individual precipitation components (convective and large-scale) have never been analyzed. Evaluating model similarity in biases in precipitation components is critically important because most of the model development efforts have focused on reducing the common biases in total precipitation, leading to an invisible bias in the individual components (e.g., He *et al.*<sup>22</sup>).

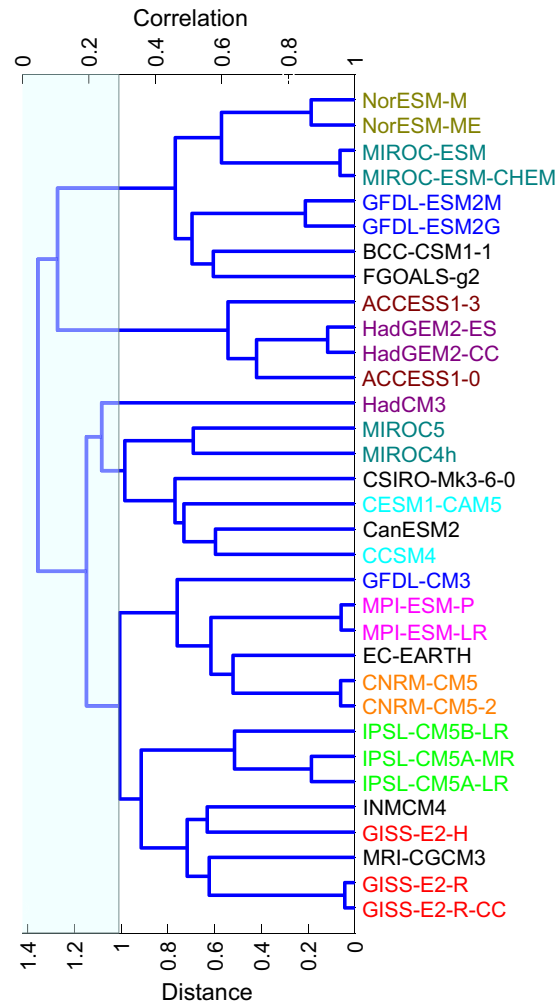
In this paper we investigate common biases in CMIP5 models by using total precipitation as well as its convective and large-scale components, and for each of the three variables we divide the models into broad groups based on the patterns of their biases on a global scale. Subsequently, the ability of models belonging to a particular group in simulating the south Asian summer monsoon and tropical waves is evaluated. In addition, AMIP5 model results have also been used to explore if the reported biases in the corresponding models from CMIP5 behave any different with prescribed sea surface temperatures.

## Results and Discussion

We have carried out a systematic analysis of similarity and dissimilarity in bias structures of CMIP5 and AMIP5 models in simulating the partitioning of precipitation between the convective and large-scale components. Figure 1 shows the hierarchical structure of CMIP5 models for total precipitation simulation. Models developed either at the same center (same color) or at different centers falling in the same branch (see Section 2 for details on methodology used for branching) show large similarity in total precipitation simulation, whereas models in the farthest branches show highest dissimilarity. From Table 1, and from Knutti *et al.*<sup>7,8</sup> and Dai<sup>19</sup>, we found some of the obvious similarities between same center or between different center models shown in Fig. 1 are arising due to the similarities in atmospheric component in spite of having different ocean components or inclusion of ocean biogeochemistry, and atmospheric chemistry. In other words, if two climate models either from the same center or different centers have the same atmospheric component their total precipitation bias is very similar, irrespective of the other components. For example, ACCESS and HadGEM show high similarity, even though they have only the atmospheric component in common. The level of similarity is very similar to model pairs that have lot more common components than that shared between ACCESS and HadGEM. For example, the level of similarity between ACCESS and HadGEM is not very different than that between: (i) MIROC-ESM-CHEM and MIROC-ESM (the former having an additional component in the form of the atmospheric chemistry package), (ii) NorESM1-ME and NorESM1-M (the former having an additional component in the form of the ocean biogeochemistry package), (iii) GISS-E2R-CC and GISS-E2R (the former having an additional component in the form of carbon cycle package), (iv) HadGEM2-ES and HadGEM2-CC (the former having an additional component in the form of atmospheric chemistry package), and (v) GFDL-ESM2G and GFDL-ESM2M (the former having different ocean component). Figure 1 also shows large dissimilarity between model pairs that started with a similar parent atmospheric component that underwent significant modification during model development. For example, dissimilarity seen between (i) MIROC-4h/5 and MIROC-ESM/ESM-CHEM, (ii) GFDL-CM3 and GFDL-ESM2G/2M, and (iii) CCSM4 and CESM-CAM5, are due to significant modifications in the atmospheric components during the model development process.

Above results on model similarity for total precipitation is consistent with findings related to CMIP3 models (e.g., Annamalai *et al.*<sup>9</sup>, Pincus *et al.*<sup>4</sup>, Bollasina *et al.*<sup>23</sup>), and CMIP5 models (e.g., Sperber *et al.*<sup>11</sup>), that state that monsoon precipitation biases in atmosphere-only models (AGCMs) are similar to the atmosphere-ocean coupled models (AOGCMs). Precipitation partitioning between convective and large-scale components are simulated separately by convective and large-scale parameterization schemes of the atmospheric component<sup>22,24,25</sup>. Figures 2 and 3 show the hierarchical structure of CMIP5 models based on their similarity in simulating the precipitation partitioning. As can be seen from Figs 1–3, similarity between models in simulating total precipitation (see Fig. 1) may not necessarily imply similarity in simulating the convective (see Fig. 2) and large-scale components (see Fig. 3).

**Convective and large-scale precipitation parameterizations used in CMIP5 models.** Convective parameterization schemes are generally based on one of the following cloud model types: (i) Spectral cloud ensemble, similar to Arakawa and Schubert<sup>26</sup>, or (ii) Bulk cloud ensemble, or (iii) Combination of spectral and bulk ensemble (Zhang and McFarlane [hereafter ZM]<sup>27</sup>; for more details see Table 2). For example, GFDL and MIROC models are based on approach (i), GISS-E2R/H and GISS-E2R-CC, HaDGEM2-ES/CC and HadCM3, ACCESS, MPI-ESM-LR/P, CNRM-CM5/5-2, CSIRO-Mk3-6-0, and EC-EARTH models are based on approach



**Figure 1.** Hierarchical clustering in CMIP5 models based on the correlation in model biases for mean annual total precipitation (40S–40N; 0–360E). The clustering method is based on weighted pairwise average distance algorithm<sup>33</sup>. The models developed at same center/institution are shown in same color.

(ii), and CCSM4, CESM-CAM5, FGOALS, and NorESM models are based on approach (iii). In addition to cloud model type, the closure and triggering mechanism of convective parameterization (for more details see Table 2) also controls the total precipitation and its partitioning. In CMIP5, most of the models use convective available potential energy (CAPE) or dilute CAPE (DCAPE) based closure and trigger function (as can be seen from Table 2), whereas, few models use moisture convergence-based closure and moisture convergence or relative humidity-based triggers. In CMIP5, most of the models use prognostic cloud condensate-based approach in their large-scale precipitation parameterization (for more details see Table 2), and hence exhibit large similarity in the simulated large-scale precipitation (discussed later).

**Model grouping based on precipitation partitioning.** CMIP5 models are clustered into the various convective groups, namely, GC1, GC2, GC3 and GC4 (see Fig. 2) and large-scale groups, namely, GL1, GL2, and GL3 (see Fig. 3), by computing similar statistics for convective and stratiform precipitation as those computed for total precipitation earlier. Models of a given convective group show similarities in cloud model type, closure assumption, or trigger mechanism. For example, (i) in GC1, MIROC-ESM/ESM-CHEM, HadGEM2-CC/ES, ACCESS1-0, FGOALS-g2, and BCC-CSM1-1 use CAPE based closure, (ii) in GC2, GFDL-ESM2M/2G, NorESM1-M/ME and INMCM4 use CAPE based closure (GISS-E2H-R/R-CC, however, uses moisture convergence based closure), and most of them use CAPE based trigger, (iii) in GC3, most of the models use either CAPE or moisture convergence based closure and CAPE or moisture convergence based trigger (for more details see Table 2), and (iv) in GC4, CNRM-CM5/5-2 use moisture convergence based stability profile for closure, HadCM3, GFDL-CM3, EC-Earth use CAPE based closure, however, most of the models in this group use CAPE based trigger (for more details see Table 2).

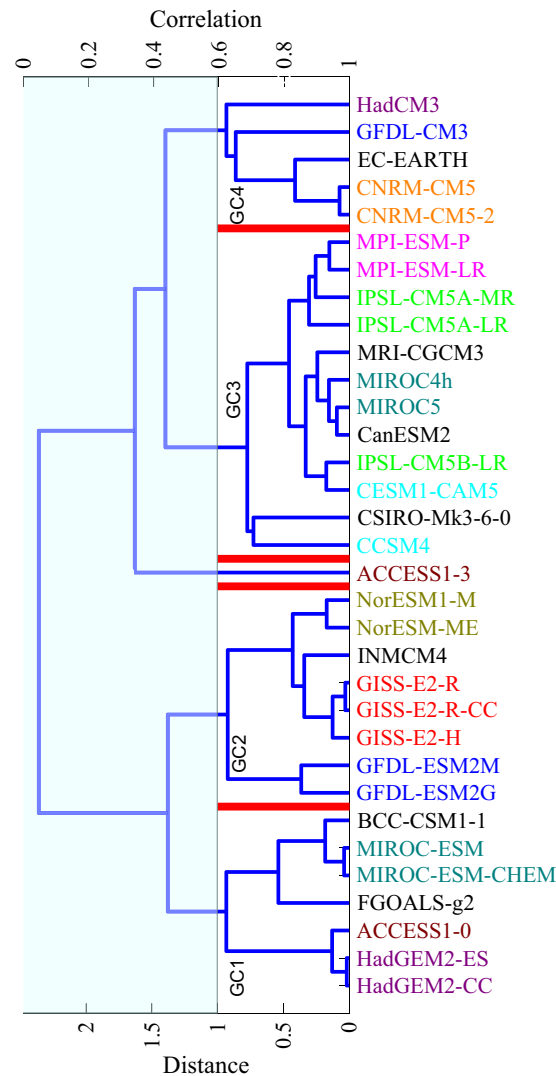
Considering three major aspects of the convection parameterization schemes used in CMIP5 models, namely, (a) cloud model type, (b) closure, and (c) trigger, we find that: (1) based on cloud model type (spectral, bulk and mix), the distribution in GC1 is (2, 3, 2), in GC2 it is (2, 3, 3), in GC3 it is (2, 4, 6), and in GC4 it is (1, 4, 0), respectively, (2) based on closure mechanism (CAPE, moisture convergence, and other methods), the distribution in

Models	Atmosphere, resolution, vertical level, reference	Ocean, resolution, vertical level	Country
GFDL-CM3	CM3, $\sim 1.802 \times 1.802$ , <b>48</b> , Donner [2011]	MOM4.1, $1-2.342 \times 1.802$ , <b>50</b>	USA
GFDL-ESM2G	CM2.1, $2.5 \times 2$ , <b>48</b> , Delworth [2006]	Gold, $1-2.342 \times 1.802$ , <b>50</b>	USA
GFDL-ESM2M	CM2.1, $2.5 \times 2$ , <b>48</b> , Delworth [2006]	MOM4.1, $1-2.342 \times 1.802$ , <b>50</b>	USA
GISS-E2R	$2 \times 2.5$ , <b>40</b>	Russell Ocean, $1 \times 1.25$ , <b>32</b>	USA
GISS-E2H	$2 \times 2.5$ , <b>40</b>	HYCOM Ocean, $0.2-1 \times 1$ , <b>26</b>	USA
GISS-E2R-CC	$\sim 1 \times 1$ , <b>40</b>	Russell Ocean, $1 \times 1.25$ , <b>32</b>	USA
CESM-CAM5	CAM5, $0.9 \times 1.25$ , <b>27</b> , Neale [2010; 2013]	Modified POP2, $1.125 \times 0.27-0.64$ , <b>60</b> , Danabasoglu [2012]	USA
CCSM4	CAM4, $0.9 \times 1.25$ , <b>27</b> , Neale [2010 & 2013]	Modified POP2, $1.125 \times 0.27-0.64$ , <b>60</b> , Danabasoglu [2012]	USA
IPSL-CM5A-LR	LMDZ5A, $1.9 \times 3.75$ , <b>39</b>	$2 \times 2-0.5$ , <b>31</b> , Madec [2008]	France
IPSL-CM5A-MR	LMDZ5A, $1.25 \times 2.5$ , <b>39</b>	$2 \times 2-0.5$ , <b>31</b> , Madec [2008]	France
IPSL-CM5B-LR	LMDZ5B, $1.9 \times 3.75$ , <b>39</b>	$2 \times 2-0.5$ , <b>31</b> , Madec [2008]	France
CNRM-CM5	ARPEGE-Climat v5.2, (IFS), $\sim TL127$ , <b>31</b> , Déqué [1994] and Voldoire [2013]	NEMOv3.2, $\sim 0.7$ , <b>42</b> , Madec [2008]	France
CNRM-CM5-2	ARPEGE-Climat v5.2, (IFS), $\sim TL127$ , <b>31</b> , Déqué [1994] and Voldoire [2013]	NEMOv3.2, $\sim 0.7$ , <b>42</b> , Madec [2008]	France
MIROC4h	AGCM5, $7.0.563 \times 0.563$ , <b>56</b>	COCO3.4, $0.28 \times 0.19$ , <b>48</b> , Hasumi [2000]	Japan
MIROC5	AGCM6, $1.406 \times 1.406$ , <b>40</b> , Nozawa [2007] and Watanabe [2008a]	COCO4.5, $1.4 \times 0.5-1.4$ , <b>50</b> , Hasumi [2006]	Japan
MIROC-ESM	MIROC-AGCM, $2.813 \times 2.813$ , <b>80</b>	COCO3.4, $1.4 \times 0.5-1.4$ , <b>44</b>	Japan
MIROC-ESM-CHEM	MIROC-AGCM, $2.813 \times 2.813$ , <b>80</b>	COCO3.4, $1.4 \times 0.5-1.4$ , <b>44</b>	Japan
CSIRO-Mk3-6-0	$\sim 1.875 \times 1.875$ , <b>18</b> , Gordon [2002; 2010] and Rotstayn [2012]	Modified MOM2.2, $\sim 0.9 \times 1.875$ , Gordon [2002 and 2010]	Australia
ACCESS1-0	HadGEM2, $1.729 \times 1.306$ N96, <b>38</b> , Martin [2011], Bi [2013b] and Rashid [2013]	ACCESS-OM (MOM4p1), $\sim 1 \times 1$ , <b>50</b> , Bi [2013a] and Marsland [2013]	Australia
ACCESS1-3	GAM1.0, $1.729 \times 1.306$ N96, <b>38</b> , Hewitt [2011], Bi [2013b] and Rashid [2013]	ACCESS-OM (MOM4p1), $\sim 1 \times 1$ , <b>50</b> , Bi [2013a] and Marsland [2013]	Australia
HadCM3	HadAM3, $3.75 \times 2.5$ , <b>19</b> , Pope [2000]	HadOM, $1.25 \times 1.25$ , <b>20</b>	UK
HadGEM2-CC	HadGEM2, $1.875 \times 1.25$ , <b>60</b>	$1.875 \times 1.25$ , <b>40</b>	UK
HadGEM2-ES	HadGEM2, $1.875 \times 1.25$ , <b>38</b> , Davies [2005]	$1 \times 1$ 30NS & $1/3^\circ$ at equator, <b>40</b>	UK
MPI-ESM-P	ECHAM6, $\sim 1.8$ T63, <b>47</b> , Stevens [2012]	MPIOM, $\sim 1.5^\circ$ GR15, <b>40</b> , Jungclaus [2013]	Germany
MPI-ESM-LR	ECHAM6, $\sim 1.8$ T63, <b>47</b> , Stevens [2012]	MPIOM, $\sim 1.5^\circ$ GR15, <b>40</b> , Jungclaus [2013]	Germany
NorESM1-M	CAM4-Oslo, $1.9 \times 2.5$ , <b>26</b> , Neale [2010] and Kirkevåg [2013]	NorESM-Ocean, $1.125$ along the equator, <b>53</b>	Norway
NorESM1-ME	CAM4-Oslo, $1.9 \times 2.5$ , <b>26</b> , Neale [2010] and Kirkevåg [2013]	NorESM-Ocean $1.125$ , long the equator, <b>53</b>	Norway
BCC-CSM1-1	BCC-AGCM2.1, $T42 \times T42$ , <b>L26</b> , Wu [2008b; 2010a, and 2012]	MOM4-L40, $\sim 1^\circ$ with $1/3$ at the equator, <b>40</b> , Griffies [2005]	China
FGOALS-g2	GAMIL2, $2.813 \times 2.813$ , <b>26</b> , Wang [2004] and Li [2013b]	LICOM2, $1 \times 1$ with $0.5$ at merid. in tropical, <b>30</b> , Liu [2012a]	China
INM-CM4	$2 \times 1.5$ , <b>21</b>	$1 \times 0.5$ , <b>40</b> , Volodin <i>et al.</i> [2010]; Zalesny <i>et al.</i> [2010]	Russia
CanESM2	<i>Spectral</i> T63, <b>35</b> , von Salzen [2013]	$1.171 \times 1.067$ , <b>40</b> , Merryfield [2013]	Canada
EC-EARTH	IFS-c31r1, $\sim 1.125 \times T159$ , <b>L62</b> , Hazeleger [2012]	NEMO_ecmwf, $\sim 1 \times 1$ , Hazeleger [2012]	Europe

**Table 1.** Model component Description, Resolution (latitude  $\times$  longitude), Vertical level.

GC1 is (7, 0, 0), in GC2 it is (5, 3, 0), in GC3 it is (9, 3, 0), and in GC4 it is (3, 2, 0), respectively, and (3) based on trigger function (CAPE, moisture convergence, other methods), the distribution in GC1 is (5, 0, 2), in GC2 it is (4, 3, 1), in GC3 it is (5, 4, 3), and in GC4 it is (3, 2, 0), respectively. Further, we notice that some models in GC2 use moisture convergence for both trigger as well as closure, whereas, some models in GC3 use moisture convergence for triggering deep convection (similar to GC2) but for closure they use CAPE (for more details see Table 2). We also find that the cloud model type used in the convection parameterization schemes has very limited effect on simulated convective precipitation. For example, FGOALS-g2 and CESM-CAM5, GFDL-CM3 and MIROC-5 model pairs have a common cloud model type but do not show much similarity in their convective precipitation fields. This finding is in line with Yanai *et al.*<sup>28</sup>, wherein it was reported that for tropical convection both spectral and bulk methods were found to produce similar total vertical mass fluxes. Unlike the convective groups, which show large inter-group variations in convective precipitation, the large-scale groups do not show as much variations, likely due to lesser degree of differences in the large-scale precipitation parameterization schemes.

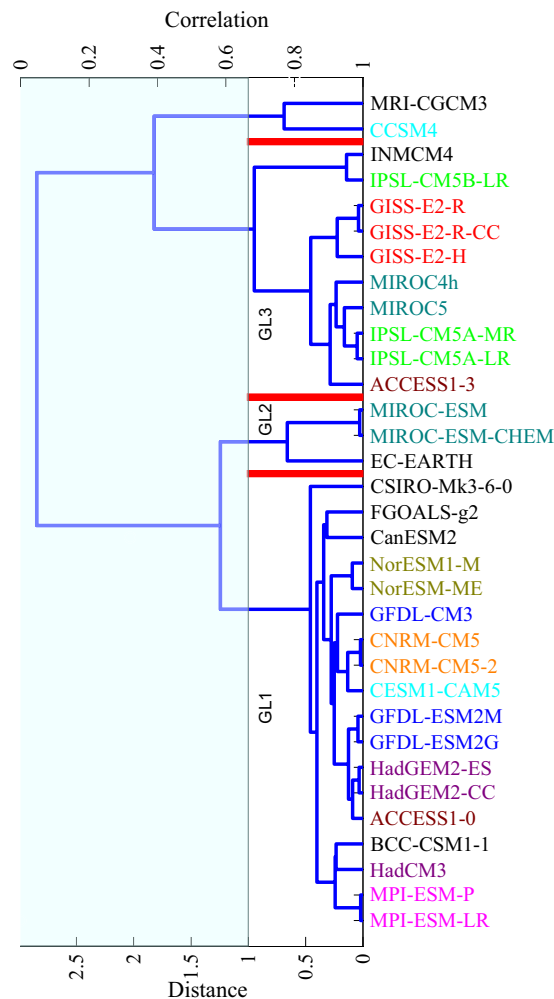
*Convective Precipitation Biases in South Asian Summer Monsoon Simulations.* Figure 4 shows spatial variation of mean seasonal convective precipitation and convective precipitation biases from observation in various convective groups. Observed convective precipitation is found to be highest over Indo-Burmese Mountain, Western



**Figure 2.** Same as Fig. 1 but with clustering done based on convective precipitation.

Ghats (WG), moderate over central India and eastern equatorial Indian Ocean, and lowest over northwest India (Fig. 4a). In GC1, the convective precipitation is found to be highest over Indo-Burmese mountain, eastern Bay of Bengal (BoB), south Arabian Sea (AS) adjoining to WG, and south China Sea (SCS) (Fig. 4b), in GC2, the convective precipitation is found to be high only over northeast India (Fig. 4c), in GC3, convective precipitation is found to be high only over the Indo-Burmese mountain (Fig. 4d), and in GC4, convective precipitation is found to be high over eastern BoB, WG, and central SCS (Fig. 4e). Thus, in GC1, we find large significant overestimation over northern AS, Indo-Burmese mountain, and SCS (Fig. 4f), in GC2, we find large significant overestimation over northeast India (Fig. 4g), in GC3, we find small overestimation over entire south Asian region (Fig. 4h), and in GC4, we find the large significant overestimation over SCS, Indo-Burmese Mountain and eastern BoB (Fig. 4i). Small underestimation in convective precipitation is found over the Indo-Gangetic region in all convective groups except in GC4, with increase in spatial extent of the negative bias from GC1 to GC3. We also notice that model grouping and spatial pattern of convective precipitation biases do not change much irrespective of the changes in observational dataset type (for example, when we change PERSIANN-CDR data with GPCP data, convective model grouping and spatial pattern biases do not change much; figure not shown).

Some of the past studies have reported that large monsoon precipitation biases over the AS, Indian land, and Indo-Burmese mountains could be due to the cold Arabian Sea SST biases<sup>15,29</sup>. In another relevant study Levine and Turner<sup>30</sup> have also shown this using numerical experiments that cool AS SST can delay south Asian summer monsoon and subsequently reduce monsoon precipitation. Next, we investigate how important are SST biases and atmosphere-ocean feedbacks in the convective precipitation biases discussed above, by analyzing the differences in biases in CMIP5 models with their corresponding AMIP5 counterparts. Hatching (in Fig. 4) indicates that the biases are primarily due to the atmospheric component, whereas, stippling indicates that the errors in SSTs and the atmosphere-ocean feedbacks are also important. Over a given grid point if the root mean square error (RMSE) in a given model group from AMIP5 simulations is greater than or equal to 80% of RMSE in the



**Figure 3.** Same as Fig. 1 but with clustering done based on large-scale precipitation.

same model group from CMIP5 simulations the grid point is hatched, whereas, if the RMSE in the AMIP5 simulations is smaller than 20% of the RMSE in the corresponding CMIP5 simulations the grid point is stippled. In addition, we also impose a second condition of two-tailed student-t test for significant bias at 99% on hatching and stippling along with the first condition mentioned above. If a grid point is neither hatched nor stippled it means that the bias is either not significant or is due to both atmospheric and oceanic components. Thus, in all the convective groups, overestimation in convective precipitation over majority of the South Asian region seems primarily to be coming from the atmospheric component. Notably, the significant biases in GC1 models over the maritime continent seem to be coming from the SST biases (since the CMIP5 biases are found to be much higher than the AMIP5 biases). In order to further confirm that the model groups are distinctively different from each other, we analysed the inter-group differences (shown in Fig. 5), and find that there are indeed significant differences between the groups, thus also confirming the robustness of the method used for clustering.

Figure 6 shows the spatial variation of mean wind pattern at 850 hPa from ERA-I and from the various convective groups (i.e. GC1, GC2, GC3 and GC4). Also shown are the corresponding biases for each of the groups. ERA-I shows a well-established cross-equatorial flow and well-established Somali current over northern AS, southern peninsular India and over the northern BoB (Fig. 6a) as reported in the literature<sup>21,31</sup>. The cross-equatorial current and Somali current are also found in all convective groups, however with varying magnitudes (Fig. 6b–e). From the mean wind biases: (a) in GC1 we find a very large cyclonic anomaly over the central equatorial Indian Ocean consistent with the overestimation in precipitation over eastern AS and BoB, easterly wind anomaly over western coast of AS, and consistent with the underestimation in precipitation over northern Indian region<sup>31–33</sup>; easterly wind anomaly over SCS consistent with the overestimation in precipitation over SCS<sup>34,35</sup> (Fig. 6f), (b) in GC2 we find easterly wind anomaly to be low over peninsular India, high over AS and small over SCS, which is thus consistent with the underestimation in precipitation over northern Indian region, and overestimation over SCS. Weak cyclonic anomaly over central equatorial Indian Ocean (IO) is consistent with small overestimation in precipitation over equatorial IO, BoB, and eastern AS (Fig. 6g)<sup>33</sup>, (c) in GC3 we find the easterly wind anomaly over the western AS, northern BoB and westerly wind anomaly over peninsular India consistent with the large underestimation in convective precipitation over Indian region and small overestimation over rest of the domain, in line with previous studies<sup>35,36</sup> (Fig. 6h), and (d) in GC4 we find the large easterly equatorial wind anomaly and



Models	Convective precipitation	Large-scale precipitation	Convective Trigger	Convective Closure
<b>Cloud Model Type: Spectral Cloud Ensemble</b>				
GFDL-CM3	Relaxed Arakawa–Schubert scheme of Moorthiand Suarez [1992] with few modifications in physics from Donner <i>et al.</i> [2011]	Cloud microphysics of Rotstayn [2000] and macrophysics from Tiedtke [1993], stratiform clouds from Golaz <i>et al.</i> [2011]	Cloud work function (CWF) similar to dilute cape (DCAPE)	CAPE closure towards a threshold over a relaxation time scale
GFDL-ESM2G	Relaxed Arakawa–Schubert scheme of Moorthiand Suarez [1992] and Dunne <i>et al.</i> [2012 and 2013]	Same as GFDL-CM3	Cloud work function (CWF) similar to DCAPE	CAPE closure towards a threshold over a relaxation time scale
GFDL-ESM2M	Same as GFDL-ESM2G	Same as GFDL-CM3	Cloud work function (CWF) similar to DCAPE	CAPE closure towards a threshold over a relaxation time scale
MIROC5	Entraining plume model scheme of Chikira <i>et al.</i> [2010] similar to Gregory [2001] with some modification according Pan and Randall [1998]	Prognostic large-scale cloud scheme of Watanabe <i>et al.</i> [2009] and bulk microphysical scheme from Wilson and Ballard [1999]	CAPE	Prognostic convective kinetic energy closure similar to CAPE closure
MIROC4h	Prognostic closure Arakawa Schubert scheme from Pan and Randall [1998] and addition of relativehumidity-based suppression condition by Emori <i>et al.</i> [2001]	Prognostic cloud water scheme of Treutand Li [1991]	Relative humidity	Prognostic convective kinetic energy closure similar to CAPE closure
MIROC-ESM	Same as MIROC4h	Large-scale condensation is diagnosed based on Treut& Li (1991) and simple cloud microphysics scheme	Relative humidity	Prognostic convective kinetic energy closure similar to CAPE closure
MIROC-ESM-CHEM	Same as MIROC4h	Same as MIROC-ESM	Relative humidity	Prognostic convective kinetic energy closure similar to CAPE closure
<b>Cloud Model Type: Bulk Cloud Ensemble</b>				
GISS-E2R	Bulk mass flux scheme by Delgenio& Yao (1993)	Prognostic stratiform cloud based on moisture convergence by Delgenio <i>et al.</i> (1996)	—	Moisture convergence
GISS-E2H	Same as GISS-E2R	Same as GISS-E2R	—	Moisture convergence
GISS-E2R-CC	Same as GISS-E2R	Same as GISS-E2R	—	Moisture convergence
HadCM3	Bulk mass flux scheme by Gregory & Rowntree (1990)	Large-scale precipitation is calculated based on cloud water and ice contents similar to Smith [1990]	Cloud base buoyancy	CAPE
HadGEM2-CC	Same as HadCM3, withan additional adaptive detrainment parameterization by Derbyshire <i>et al.</i> [2011]	Same as HadCM3	Cloud base buoyancy	CAPE
HadGEM2-ES	Same as HadGEM2-CC	Same as HadCM3	Cloud base buoyancy	CAPE
ACCESS1-0	Same as HadGEM2-CC	Same as HadCM3	Cloud base buoyancy	CAPE
ACCESS1-3	Same as in ACCESS1.0, except physical parameterization, which is similar to GAM1.0	Same as HadCM3	—	CAPE
MPI-ESM-LR	Bulk mass flux scheme by Tiedtke [1989] with modifications in deep convection by Nordeng <i>et al.</i> [1994]	Prognostic equations of the water phases, bulk cloud microphysics from Lohmann andRoeckner [1996]	Moisture convergence and buoyant surface air when lifted to the LCL	Moisture convergence/ adjustment type
MPI-ESM-P	Same as MPI-ESM-LR	Same as MPI-ESM-LR	Moisture convergence and buoyant surface air when lifted to the LCL	Moisture convergence/ adjustment type
MRI-CGCM3	Same as MPI-ESM-LR		Moisture convergence	CAPE
CNRM-CM5	Mass-flux scheme of Bougeault [985]	Statistical cloud scheme of Ricard and Royer [1993]	Depends on moisture convergence and stability profile	Moisture convergence
CNRM-CM5-2	Same as CNRM-CM5	Same as CNRM-CM5	Depends on moisture convergence and stability profile	Moisture convergence
EC-EARTH	Bulk mass-flux scheme and Entraining/detraining plume cloud model by Hazeleger <i>et al.</i> [2010]	saturated downdraughtsandsimple microphysics scheme	—	CAPE
CSIRO-Mk3-6-0	Bulk mass flux convection scheme of Gregory and Rowntree [1990] with slightly modified by Gregory [1995]	Stratiform cloud condensate scheme from Rotstayn([000])	—	Stability-Dependent Closure
<b>Cloud Model Type: Mixed Nature of Spectral and Bulk Cloud Ensemble</b>				
CCSM4	Simplified Arakawa and Schubert cumulus ensemble scheme of Zhang and McFarlane <sup>27</sup> with plume dilution of Neale <i>et al.</i> [2008]	Prognostic condensate and precipitation parameterization from Zhang <i>et al.</i> [2003]	CAPE	DCAPE
CESM-CAM5	Same as CCSM4 with few more modifications by Neale <i>et al.</i> [2012]	Same as CCSM4	CAPE	DCAPE
NorESM1-M	Same as CCSM4	Same as CCSM4	CAPE	DCAPE
NorESM1-ME	Same as CCSM4	Same as CCSM4	CAPE	DCAPE
BCC-CSM1-1	Mass flux scheme developed by Zhang and McFarlane [1995], has been adapted as proposed by Wu <i>et al.</i> [2010]	Same as CCSM4	CAPE	CAPE
FGOALS-g2	Mass flux type cumulus convection developed by Zhang and McFarlane <sup>27</sup>	Precipitation occurs whenever the local relative humidity is supersaturated	CAPE	CAPE
Continued				

Models	Convective precipitation	Large-scale precipitation	Convective Trigger	Convective Closure
CanESM2	Mass flux type cumulus convection scheme by Scinocca and McFarlane [2004]	Prognostic cloud liquid water and ice, statistical cloud scheme, interactive with aerosols	CAPE	Cloud base closure
IPSL-CM5A-LR	Episodic mixing and buoyancy sorting scheme by Emanuel [1991] and modified moist convection scheme by Grandpeix <i>et al.</i> [2004]	Cloud cover and in-cloud water deduced from large-scale total water and moisture at saturation from Bony and Emmanuel [2001]	—	CAPE closure
IPSL-CM5A-MR	Same as IPSL-CM5A-LR	Same as IPSL-CM5A-LR	—	CAPE closure
IPSL-CM5B-LR	Same as IPSL-CM5A-LR, with the modification in closure and trigger mechanism by Grandpeix and Lafore [2010]	Same as IPSL-CM5A-LR, with the few modifications by Jam <i>et al.</i> [2011]	Available Lifting Energy	Available Lifting Power
INM-CM4	Lagged convective adjustment after Betts [1986], but modified referenced profile for deep convection	Stratiform cloud fraction is calculated as linear function of relative humidity	—	CAPE

**Table 2.** Description of Convective and Large-scale parameterization, Convective triggers and Convective closures.

westerly wind anomaly over central BoB causing more overestimation in precipitation over equatorial region, BoB and SCS, in line with previous studies<sup>31,35</sup> (Fig. 6i).

*Large-Scale Precipitation Biases in South Asian Summer Monsoon Simulations.* Figure 7 shows spatial variations of mean seasonal large-scale precipitation and large-scale precipitation biases from observations in various large-scale groups. It can be seen from Fig. 7a that observed large-scale precipitation is high over Indo-Burmese mountain, WG, and northeast India. In GL1, highest values are found over Himalayan foothills (Fig. 7b); in GL2, highest values are found over Himalayan foothills, WG and eastern Arabian Sea (Fig. 7c); in GL3, highest values are found over Himalayan foothills and over northeast India (Fig. 7d). From the bias patterns, GL1 shows large underestimation over Indo-Burmese mountain, eastern BoB, and WG (Fig. 7e), GL2 shows large underestimation over Indo-Burmese mountain and eastern BoB (Fig. 7f), and GL3 shows negative biases in line with GL1 and GL2 but with lower magnitudes (Fig. 7g). All large-scale precipitation groups also show the underestimation over central India and eastern equatorial Indian Ocean. In all the large-scale groups, underestimation in large-scale precipitation over majority of the south Asian region seems primarily to be coming from the atmospheric component (see hatching in Fig. 7). The underestimation in large-scale precipitation over WG in all large-scale groups seems to be due to both atmospheric and oceanic components. Similar to the convective precipitation grouping and spatial bias pattern, the large-scale model grouping and spatial bias pattern is also found to be minimally affected by the use of two different observational datasets (PERSIANN-CDR and GPCP; figure not shown). Similar to the analysis carried out to test the distinctiveness of the convective groups discussed above, we analysed the inter-group differences for the large-scale-precipitation-based groups (shown in Fig. 8), and find that the inter-group differences are significant.

## Conclusions

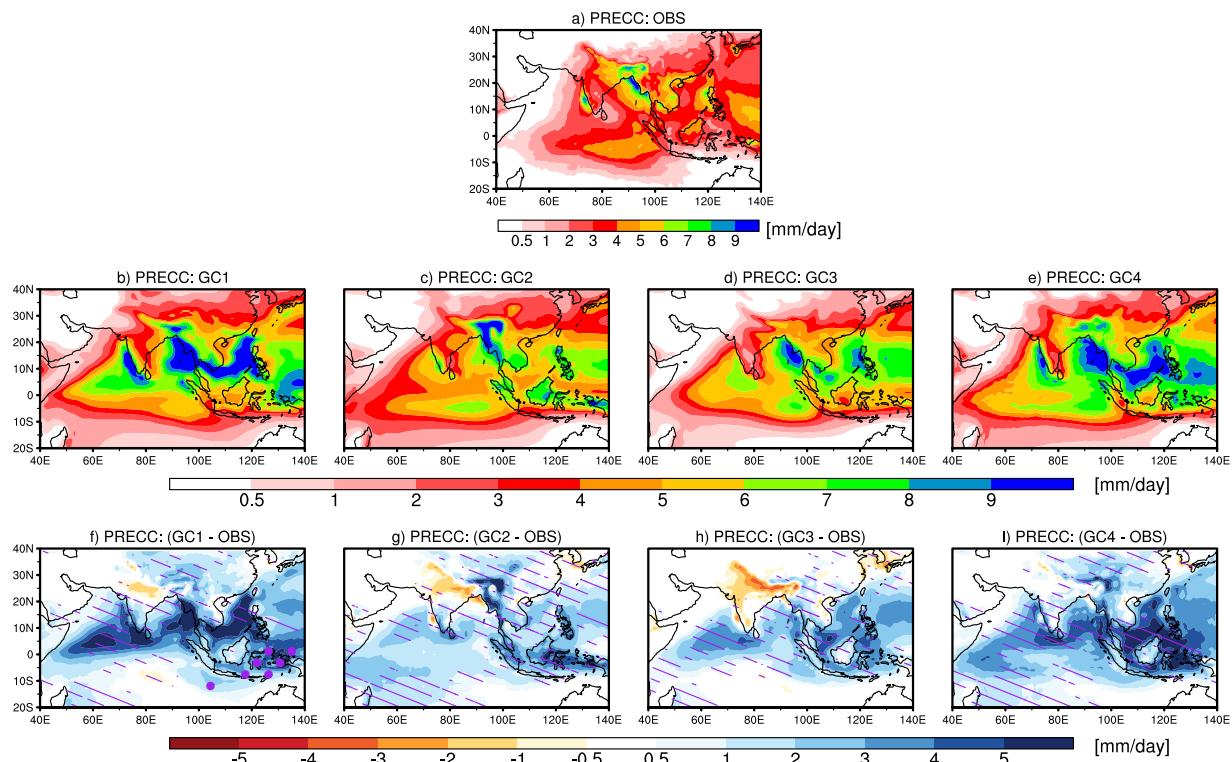
We have carried out a systematic analysis of the structure of precipitation biases in 33 CMIP5 and AMIP5 models, and have grouped them based on the correlation of their biases in total, convective and large-scale precipitation on global scale. We found that the grouping of models is somewhat sensitive to the variable used, i.e., a given pair of models that fall in the same total precipitation bias group may not necessarily fall in the same convective or large-scale precipitation bias group.

By grouping the CMIP5 models based on their convective precipitation biases we find that the similarity in convective precipitation biases in a given group primarily comes from similarity in closure assumptions and trigger mechanisms, and to a lesser extent on the details of the cloud models used in the deep convection parameterization schemes of the models. By grouping the CMIP5 models based on their large-scale precipitation biases we find that the degree of similarity in large-scale precipitation biases among model groups was much higher than that seen in the corresponding convective precipitation biases (based on convective precipitation grouping). Over the south Asian domain, we find many biases that are common across the groups. In general, each of the convective groups show largely positive biases, whereas, each of the large-scale groups show largely negative biases over the south Asian region, with spatially varying magnitudes. We find that the spatial pattern of biases in the convective precipitation in various model groups have prominent signatures in the 850 hPa wind circulation biases as well.

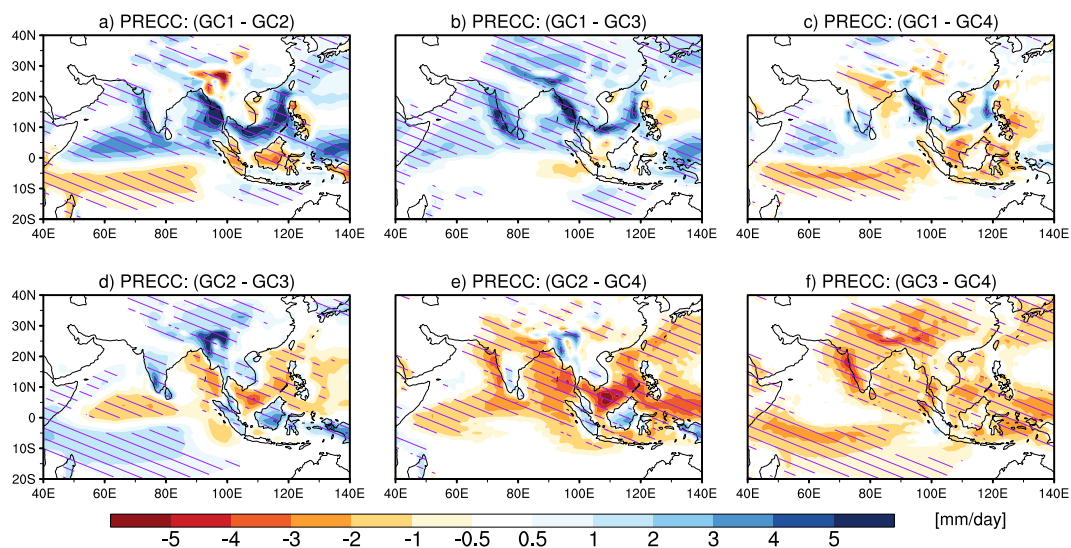
In agreement with some prior studies<sup>5,8</sup>, we find that if 2 models have the same atmospheric component the degree of similarity in their global precipitation bias patterns is quite high, as compared to that if some other component(s) are similar but the atmospheric components (especially the convection scheme) are quite different. This finding highlights the primary role played by the atmospheric component of the model in governing precipitation biases. To investigate this further, we compare the corresponding model biases from CMIP5 and AMIP5 simulations, and conclude that, in general, the precipitation biases primarily depend on the atmospheric component of the models, and to a lesser extent on biases in SSTs or atmosphere-ocean feedbacks, at least on timescales of the current analysis. Notably, we find that there is only one model group wherein the ocean component is primarily responsible for the simulated convective precipitation biases (found over the maritime continent region).

As a first step towards eliminating a given bias in a model it is important to know how the bias structure in the model compares to other models in the same group and models in different groups. Thus, a more informed and efficient model development approach may be designed for achieving improved simulations of global and



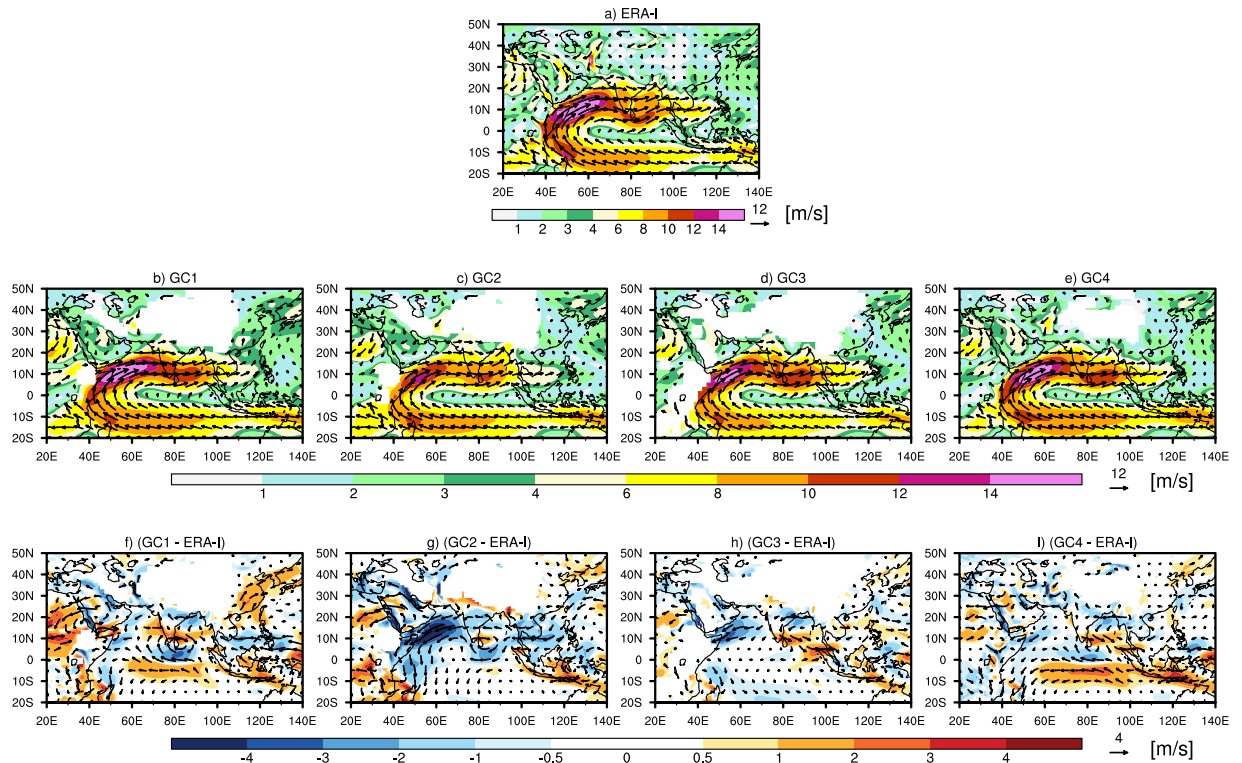


**Figure 4.** The spatial variation of mean JJAS (June–September) convective precipitation over the south Asian region from observation (a), GC1 (b), GC2 (c), GC3 (d), and GC4 (e). The biases in mean JJAS convective precipitation for different groups with respect to observation are shown in (f) for GC1, (g) for GC2, (h) for GC3, and (i) for GC4. Hatching show bias to be coming from atmospheric component and stippling show bias to be coming from oceanic components (i.e. biases in SSTs and atmosphere-ocean feedbacks) and the biases are significant the level of 99%.



**Figure 5.** The differences in mean JJAS convective precipitation of each cluster with the other clusters: (a) GC1 and GC2, (b) GC1 and GC3, (c) GC1 and GC4, (d) GC2 and GC3, (e) GC2 and GC4, and (f) GC3 and GC4. Regions with differences that are statistically significant at 99% are hatched.

regional climate. Not to mention, the spatial resolution used in CMIP5 models is too coarse to resolve weather features such as fronts, atmospheric rivers, cyclone properties, and thus, weather resolving climate models would be required for improving the accuracy of simulations even further<sup>37,38</sup>.



**Figure 6.** The spatial variation of mean JJAS (June – September) 850 hPa wind pattern over the south Asian region from (a) ERA-I, (b) models in convective group GC1, (c) GC2, (d) GC3, and (e) GC4. The biases in mean JJAS 850 hPa wind pattern for different convective groups with respect to reanalysis are shown in (f) for GC1, (g) for GC2, (h) for GC3, and (i) for GC4.

### Data and Methodology

The historical simulation of monthly convective (PRECC) and total precipitation (PRECT) dataset from 33 CMIP5 and AMIP5 models<sup>39</sup> were downloaded from the Earth System Grid Federation (ESGF; <https://esgf-node.llnl.gov/>). We use the r1i1p1 ensemble member for all CMIP5 and AMIP5 models, since some models of CMIP5 and AMIP5 do not provide the individual convective and large-scale components for other ensemble members. Observed monthly total precipitation for 23 years (1983–2005) are from PERSIANN-CDR dataset, which is a high-resolution ( $0.25^\circ \times 0.25^\circ$ ) long-term satellite and observation merged precipitation dataset, developed by the Centre for Hydrometeorology and Remote Sensing (CHRS) at University of California Irvine (<https://chrsdata.eng.uci.edu/>; Nguyen *et al.*<sup>40</sup>). The PERSIANN-CDR dataset is first bilinearly interpolated to the  $0.5^\circ \times 0.5^\circ$  grid of TRMM 3A12 (1998–2013;  $0.5 \times 0.5$  degree)<sup>41</sup>, and then the corresponding convective and large-scale precipitation components are computed from total precipitation, by using Eq. (1). The large-scale precipitation (PRECL) dataset for observation and for model simulations are computed by subtracting the convective components from the total precipitation as shown in Eq. (2). Monthly zonal (u) and meridional (v) wind dataset (1983–2005) at 850 hPa are used from ECMWF (ERA-I) reanalysis (<https://apps.ecmwf.int/datasets/>). The domain used for our analysis is  $0^\circ\text{--}360^\circ\text{E}$ ;  $40^\circ\text{S}\text{--}40^\circ\text{N}$ .

$$PRECC (PERSIANN\_CDR) = PRECT(PERSIANN\_CDR) * \left( \frac{PRECC (TRMM)}{PRECT (TRMM)} \right) \tag{1}$$

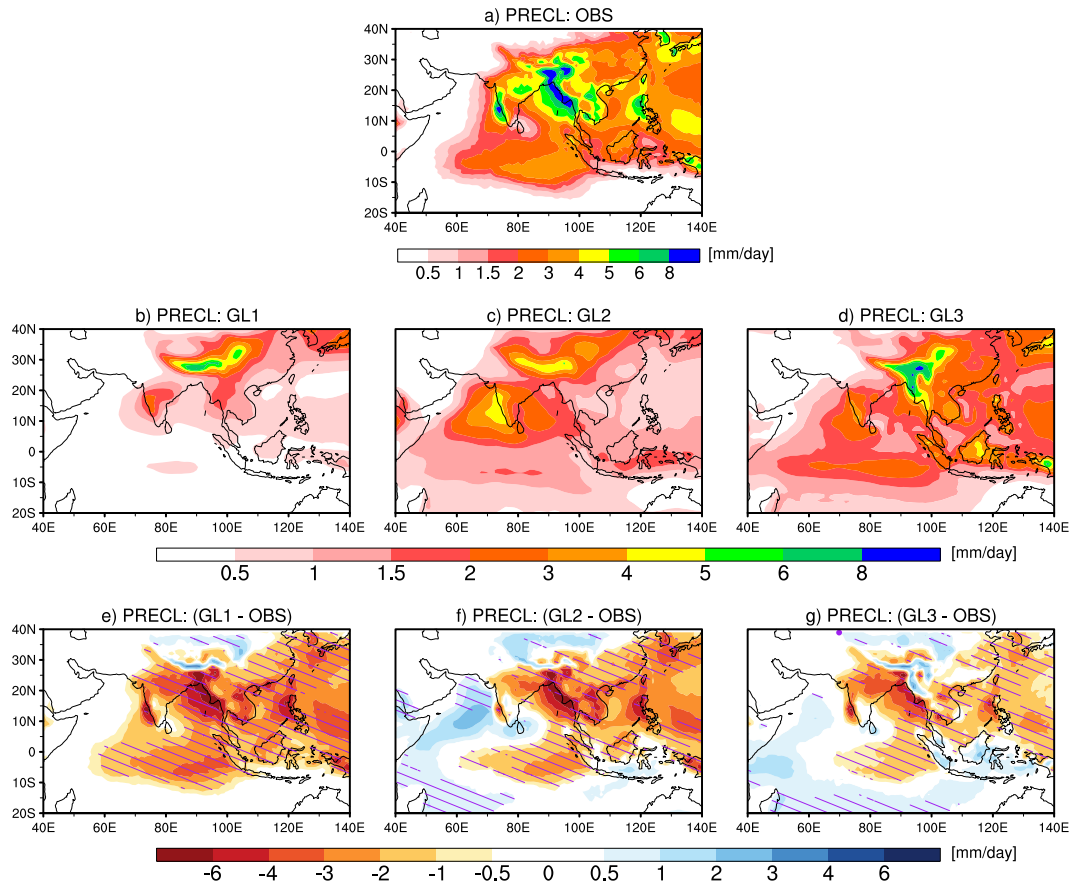
$$PRECL = (PRECT - PRECC) \tag{2}$$

To compute inter-model similarity between the CMIP5 models, we follow a similar method as that used by Pennell and Reichler<sup>3</sup> and Knutti *et al.*<sup>8</sup> which are widely accepted methods for model genealogy studies. Thus, we first compute the normalized bias ( $e_{n,m}$ ) in the mean annual total, convective, and large-scale precipitation for the CMIP5 models by using Eq. (3)<sup>3</sup>:

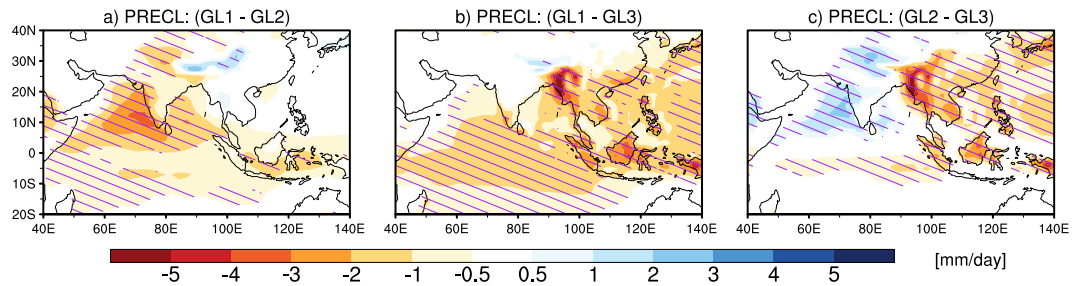
$$e_{n,m} = (f_{n,m} - o_n) / \sigma_n$$

$$n = 1, 2, 3, \dots, N; m = 1, 2, 3, \dots, M \tag{3}$$

where ‘N’ and ‘M’ are the total number of grid points and total number of models used for this study, symbol ‘o’, ‘f’, and ‘σ’ represent observation, model output, and observed standard deviation (σ) for any variable. For example, the normalized bias for a particular model (m) can be written from Eq. (3) as  $e_m = (e_{1,m}, e_{2,m}, \dots, e_{N-1,m}, e_{N,m})$ .



**Figure 7.** Same as Fig. 4 but for large-scale precipitation from observation and from large-scale precipitation groups (GL1, GL2, and GL3).



**Figure 8.** Same as Fig. 5, but for large-scale precipitation.

Subsequently, we compute the common bias total, convective, and large-scale precipitation from the multi-model mean bias ( $\bar{e}$ ) by using Eq. (4)<sup>3</sup>.

$$\bar{e} = \left( \frac{1}{M} \right) \sum_{m=1}^{m=M} e_m \tag{4}$$

The portion of multi-model mean bias ( $\bar{e}$ ) corresponding to each individual model bias (i.e.  $r = \text{Cor.} (e_m, \bar{e})$ ) is then subtracted from the corresponding model bias, to make the individual model biases more dissimilar from each other by using Eq. (5)<sup>3</sup>.

$$d_m = e_m - r\bar{e} \tag{5}$$

Finally, the level of inter-model similarity is computed by applying the Pearson sample linear cross-correlation between the model pair biases [ $\text{Cor.} (d_{m1}, d_{m2})$ ]. The hierarchical structure of CMIP5 models is constructed by converting the correlation matrix into distance matrix by using weighted pair-wise average distance method<sup>3,42</sup>.

## Data Availability

All the data used in this study is in public domain and can be freely downloaded.

## References

- Flato, G. *et al.* Evaluation of Climate Models. In: *Climate Change 2013: The Physical Science Basis*. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T. F. *et al.* (eds)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA (2013).
- Meehl *et al.* Global climate projections. *Climate Change 2007: The Physical Science Basis*. Solomon, S. *et al.* Eds Cambridge Press University 747–846 (2007a).
- Pennell, C. & Reichler, T. On the Effective Number of Climate Models. *Journal of Climate* **24**, 2358–2367, <https://doi.org/10.1175/2010JCLI3814.1> (2011).
- Pincus *et al.* Evaluating the present-day simulation of clouds, precipitation, and radiation in climate models. *Journal of Geophysical Research* **113**(D14209), <https://doi.org/10.1029/2007JD009334> (2008).
- Masson, D. & Knutti, R. Climate model genealogy. *Geophysical Research Letter* **38**(L08703), <https://doi.org/10.1029/2011GL046864> (2011).
- Jun, M., Knutti, R. & Nychka, D. W. Spatial analysis to quantify numerical model bias and dependence: How many climate models are there? *J. Amer Stat Assoc* **103**, 934–947 (2008b).
- Knutti, R., Furrer, R., Tebaldi, C., Cermak, J. & Meehl, J. A. Challenges in Combining Projections from Multiple Climate Models. *J Climate* **23**(10), 2739–2758 (2010).
- Knutti, R., Masson, D. & Gettelman, A. Climate model genealogy: Generation CMIP5 and how we got there. *Geophysical Research Letters* **40**, 1194–1199 (2013).
- Annamalai, H., Hamilton, K. & Sperber, K. R. The South Asian summer monsoon and its relationship with ENSO in the IPCC AR4 simulations. *Journal of Climate* **20**(6), 1071–1092, <https://doi.org/10.1175/JCLI4035.1> (2007).
- Gadgil, S. & Sajani, S. Monsoon precipitation in the AMIP runs. *Clim Dyn* **14**, 659–689 (1998).
- Sperber, R. K. *et al.* The Asian summer monsoon: An intercomparison of CMIP5 versus CMIP3 simulations of the late 20th century. *Clim Dyn* **41**, 2711–2744, <https://doi.org/10.1007/s00382-012-1607-6> (2013).
- Turner, A. *et al.* Modelling monsoons: understanding and predicting current and future behavior, in: *Global Monsoon System: Research and Forecast*, 2nd ed. edited by: C-P C, Ding, Y., Lau, N. C., Johnson, R. H., Wang, B. & Yasunari, T., World Scientific Publishing Co, Singapore 421–454 (2011).
- Turner, A. G. & Annamalai, H. Climate change and the South Asian summer monsoon. *Nat Clim Change* **2**, 587–595, <https://doi.org/10.1038/Nclimate1495> (2012).
- Levy, A. A. L. *et al.* Can correcting feature location in simulated mean climate improve agreement on projected changes? *Geophysical Research. Letter* **40**, 354–358, <https://doi.org/10.1029/2012GL053964> (2013).
- Sandeep, S. & Ajayamohan, R. S. Origin of cold bias over the Arabian Sea in Climate Models. *Scientific Report* **4**, 6403, <https://doi.org/10.1038/srep06403> (2014).
- Sarojini, B. B., Stott, P. A., Black, E. & Polson, D. Fingerprints of changes in annual and seasonal precipitation from CMIP5 models over land and ocean. *Geophysical Research letter* **39**, L21706, <https://doi.org/10.1029/2012GL053373> (2012).
- Sarojini, B. B., Stott, P. A. & Black, E. Detection and attribution of human influence on regional precipitation. *Nature Climate Change* **6**, 669–675, <https://doi.org/10.1038/NCLIMATE2976> (2016).
- Hegerl, G. C. *et al.* Challenges in Quantifying Changes in the Global Water Cycle. *BAMS*, 1097–1115, <https://doi.org/10.1175/BAMS-D-13-00212.1> (2015).
- Dai, A. Precipitation Characteristics in Eighteen Coupled Climate Models. *Journal of climate* **19**, 4605–4630 (2006).
- KrishnaKumar, K., Rupa, K. K., Ashrit, R. G., Deshpande, N. R. & Hansen, J. W. Climate Impacts On Indian Agriculture. *Int J Climatol* **24**, 1375–1393 (2004).
- Menon, A., Levermann, A., Schewe, J., Lehmann, J. & Frieler, K. Consistent increase in Indian monsoon rainfall and its variability across CMIP-5 models. *Earth SystDynam* **4**, 287–300, <https://doi.org/10.5194/esd-4-287-2013> (2013).
- He, J. & Alapaty, K. Precipitation Partitioning in Multiscale Atmospheric Simulations: Impacts of Stability Restoration Methods. *Journal of Geophysical Research: Atmosphere* **123**(18), 185–201, <https://doi.org/10.1029/2018JD028710> (2018).
- Bollasina, M. & Nigam, S. Indian Ocean SST, evaporation, and precipitation during the South Asian summer monsoon in IPCC-AR4 coupled simulations. *Clim Dyn* **33**, 1017–1032, <https://doi.org/10.1007/s00382-008-0477-4> (2009).
- Sun, D.-Z. & Zhang, T. Radiative and Dynamical Feedbacks over the Equatorial Cold Tongue: Results from Nine Atmospheric GCMs. *Journal of Climate* **19**, 4059–4074 (2005).
- Zhu, H. & Hendon, H. Convection and MJO Performance in UM7.1: CAWCR Technical Report No. 022 (2010).
- Arakawa, A. & Schubert, W. H. Interaction of cumulus cloud ensemble with the large-scale environment Part I. *J Atmos Sci* **31**, 674–701 (1974).
- Zhang, G. J. & McFarlane, N. A. Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian climate centre general circulation model. *Atmosphere-Ocean* **33**(3), 407–446, <https://doi.org/10.1080/07055900.1995.9649539> (1995).
- Yanai, M., Chu, J.-H., Stark, T. E. & Nitta, T. Response of deep and shallow tropical maritime cumuli to large-scale processes. *J Atmos Sci* **33**, 976–991 (1976).
- Levine, R. C., Turner, A. G., Marathayil, D. & Martin, G. M. The role of northern Arabian Sea surface temperature biases in CMIP5 model simulations and future projections of Indian summer monsoon rainfall. *Clim Dyn* **41**, 155–172 (2013).
- Levine, R. C. & Turner, A. G. Dependence of Indian monsoon rainfall on moisture fluxes across the Arabian Sea and the impact of coupled model sea surface temperature biases. *Climate Dyn* **38**, 2167–2190 (2012).
- Findlater, J. Cross-equatorial jet stream at low level over Kenya. *Meteorol. Mag.* **95**, 353–364 (1966).
- Day, J. A., Fung, I. & Risi, C. Coupling of South and East Asian Monsoon Precipitation in July–August. *Journal of Climate* **28**, 4330–4356, <https://doi.org/10.1175/JCLI-D-14-00393> (2015).
- Pai, D. S. & Rajeevan, M. Summer monsoon onset over Kerala: New definition and prediction. *J. Earth Syst. Sci.* **118**, 123–135 (2009).
- Wu, R. & Hu, W. Air-Sea Relationship Associated with Precipitation Anomaly Changes and Mean Precipitation Anomaly over the South China Sea and Arabian Sea during the Spring to Summer Transition. *Journal of Climate* **28**, 7161–7181, <https://doi.org/10.1175/JCLI-D-15-0136.1> (2015).
- Goswami, B. N., Chakravorty, S. Dynamics of the Indian Summer Monsoon. *Climate System and Climate Dynamics*, Oxford Research Encyclopaedia for Climate Science, <https://doi.org/10.1093/acrefore/9780190228620.013.613>.
- Taniguchi, K. & Koike, T. Comparison of definitions of Indian summer monsoon onset: Better representation of rapid transitions of atmospheric conditions. *Geophys. Res. Lett.* **33**, L02709, <https://doi.org/10.1029/2005GL024526> (2006).
- Kendon, E. J. *et al.* Heavier summer downpours with climate change revealed by weather forecast resolution model. *Nature Climate Change* **4**, 570–576 (2014).
- Roberts, C. D., Palmer, M. D., McNeall, D. & Collins, M. Quantifying the likelihood of a continued hiatus in global warming. *Nature Climate Change* **5**, 337–342 (2015).



39. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An Overview of CMIP5 and the Experiment Design. *Bull. of Amer. Meteor. Soc.* **93**, 485–498 (2012).
40. Nguyen, P. *et al.* The CHRS Data Portal, an easily accessible public repository for PERSIANN global satellite precipitation data. *Nature Scientific Data* **6**, 180296, <https://doi.org/10.1038/sdata.2018.296> (2019).
41. Huffman, G. J. *et al.* The TRMM Multi-satellite Precipitation Analysis: Quasi-Global, Multi-Year, Combined-Sensor Precipitation Estimates at Fine Scale. *J. Hydrometeor.* **8**, 38–55 (2007).
42. Sokal, R. & Michener, C. A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin.* **38**, 1409–1438 (1958).

## Acknowledgements

This work is partially supported by the DST Center of Excellence in Climate Modeling (<http://dcecm.iitd.ac.in/>) at IIT Delhi, India. We are thankful to ESGF Lawrence Livermore National Laboratory (LLNL) (<https://esgf-node.llnl.gov/>) for providing CMIP5 and AMIP5 model datasets. We are also thankful to ECMWF (<http://apps.ecmwf.int/datasets/>) for providing ERA-Interim reanalysis dataset and to NOAA-ESRL Physical Science Division (PSD) for providing the observed GPCP precipitation dataset. The authors would also like to thank the supercomputing facility at IIT Delhi, India. For data analysis and plotting NCAR Command Language Version-6.4.0 (<https://www.ncl.ucar.edu/>) and MatLab (R2017b) is used. RP would like to thank MHRD for his PhD fellowship.

## Author Contributions

S.S., S.K.M. and R.P. designed the study and jointly carried out the analysis. R.P. generated the figures. All authors contributed in writing the manuscript.

## Additional Information

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as 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. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019