ARTICLE OPEN

Check for updates

Effects of multi-observations uncertainty and models similarity on climate change projections

Raju Pathak[1,2], Hari Prasad Dasari [ID][1,2], Karumuri Ashok [ID][1,2,3] and Ibrahim Hoteit [ID][1,2 ✉]

Climate change projections (CCPs) are based on the multimodel means of individual climate model simulations that are assumed to be independent. However, model similarity leads to projections biased toward the largest set of similar models and intermodel uncertainty underestimation. We assessed the influences of similarities in CMIP6 through CMIP3 CCPs. We ascertained model similarity from shared physics/dynamics and initial conditions by comparing simulated spatial temperature and precipitation with the corresponding observed patterns and accounting for intermodel spread relative to the observational uncertainty, which is also critical. After accounting for similarity, the information from 57 CMIP6, 47 CMIP5, and 24 CMIP3 models can be explained by just 11 independent models without significant differences in globally averaged climate change statistics. On average, independent models indicate a lower global-mean temperature rise of 0.25 °C (~0.5 °C–1 °C in some regions) relative to all models by the end of the 21st century under CMIP6's highest emission scenario.

INTRODUCTION

Climate change is a serious concern for modern civilization because of its unforeseeable consequences for human societies and ecosystems and its destabilization effect in the case of various earth systems and processes[1,2]. Increasing global mean surface temperatures exacerbate the magnitude and frequency of heat waves and are associated with increasing extreme storms, heavier precipitation, floods, and more intense droughts. Such disasters cause deaths, infrastructural damage, and environmental and economic losses[3]. Climate change projections (CCPs), driven by emissions estimated according to a set of future socioeconomic conditions, can be used to estimate changes in future climate statistics and the changing frequencies and intensities of extreme events. They are useful for risk-based long-term planning[3,4].

CCPs of a particular climate variable (such as temperature, precipitation, etc.) from a single global climate model (GCM) are, generally, the averages of the variable from multiple simulations (realizations) that differ in terms of initial conditions or parameter values[5,6]. Analogously, multimodel-based CCPs are statistical consensus estimates (i.e., averaging) of multiple simulated GCMs[1,3–5]. Such methods often lead to inaccurate estimates of uncertainty associated with model parameters and numerical schemes/physics. For example, equally weighted multimodel mean of the GCM simulations is typically used as the best estimates[3,4]. Multimodel ensembles of several GCMs, wherein each GCM provides statistically independent climate information, can be used to represent structural uncertainty[7]. Typically, the uncertainty associated with a simulated climate variable is expressed by the spread across models[3,4,8]. The consensus methods used to generate CCPs generally smooth out structural uncertainties and ignore disparate physical process assumptions associated with different models[8–11]. Scientists consider consensus estimates along with uncertainties when interpreting CCPs.

Importantly, because GCMs are developed, calibrated, and initialized using observations, any uncertainty in the observations can affect the model-produced weather and climate projections.

Observational uncertainty affects the validation and evaluation of model outputs, leading to incorrect model performance and ranking[12–16]. Validations conducted using single observation-based gridded datasets are subject to data uncertainty and incorrect interpretation[13]. A few studies have discussed observational uncertainties at the regional scale, especially for Europe[12], the United States of America[14], and India[13,15,16]. Using multiple regional climate model simulations, Kotlarski et al.[12] found observational uncertainty comparable to model uncertainty for many European countries, making the model results sensitive to the observations used. In addition, GCMs developed by various institutions share similar/same model physics and components, which inevitably introduces some level of dependence across the models[5,9,10,17–20]. In this study, observational uncertainty and model uncertainty for a particular climate variable refer to, respectively, uncertainty across multiple observations and that across multiple models. In a pioneering study, Pennell and Reichler[9] found that only 8 of the 24 Coupled Model Intercomparison Project (CMIP) Phase-3 (CMIP3) models were independent over the northern extratropics. Collins et al.[5] suggested this was due to the complexity of developing models from scratch. Moreover, CMIP5 model dependency is higher than that of CMIP3, being associated with similar initial conditions and numerical schemes[10] and similarities in physics packages, such as cumulus parameterization[16]. Because such similarities in a group of models lead to common biases, treating all models equally will likely bias multimodel predictions[18] and underestimate uncertainties[11]. Leduc et al.[20] also argued that any result that agrees across multiple models of the same group may not be considered reliable unless there is evidence that these models are independent. Indeed, consensus estimates from weighted CMIP6 models based on performance and dissimilarity indicate relatively reduced future global warming with lower uncertainty than unweighted CMIP6 models[8]. Nevertheless, model-developing centers with many models in CMIP6 (even though they do not differ among themselves by much) contribute more weight to the multimodel

---

[1]Physical Science and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. [2]Climate Change Center, National Center for Meteorology, Jeddah, Saudi Arabia. [3]Centre for Earth, Ocean and Atmospheric Sciences, University of Hyderabad, Hyderabad, India. ✉email: ibrahim.hoteit@kaust.edu.sa

means;[20] the combined model performance is also susceptible to observational uncertainties. This suggests that implications of observational uncertainty and model dependence must be considered for reliable climate projections. In this case, selecting the group of best models that is relatively more robust to the observational uncertainty (i.e., the group of models that score an acceptable rank against the different observations) makes the result more reliable. A further criterion of running these best models through a model independence test should enhance the degree of confidence in climate projections.

Here, we quantify the uncertainties in multi-observations and models and the dissimilarities among GCM simulations for the historical period 1980–1999, for which multiple observations and models are available. Three metrics were used to evaluate the uncertainties, namely, the spatial mean bias, pattern correlation, and interannual variability, which describe different aspects of observational and model uncertainties. These metrics were computed for temperature and precipitation.

We used CMIP6 datasets, along with those from CMIP5 and CMIP3 datasets, to estimate contributions to intermodel similarity (dissimilarity) that arise from modules common (distinct) to other models in terms of physics or features, such as aerosol loading, carbon cycle, ocean biogeochemistry, and resolution. We calculated the "effective" number of climate models after removing the influence of model similarity—this exercise also showed that the intermodel independence assumption was not valid. We then investigated the impact of model similarity on the global and regional future climate change estimates of socioeconomic pathways (SSP) 5–8.5 for the period 2015–2100 against the mean climate of 1980–2014 using three CMIP6 model sets: all models, similar model pairs, and diverse models. We estimated the observational and model uncertainties for historical simulations and their implications. Assuming that these uncertainties persist in the future, we conjectured similar implications from these uncertainties when interpreting future CCPs.

RESULTS

Observation and model dataset uncertainty

Differences across various observational and reanalyzed datasets introduce uncertainties when rating model performance[13] and thus influence the interpretations of multimodel evaluations and model uncertainties. We intercompared annual and seasonal climatological spatial distributions, pattern correlations, and interannual variabilities to gauge the differences among the seven observational rainfall and temperature datasets (Supplementary Table 1). We used the Multi-Source Weather (MSWX) datasets as a reference for both precipitation and temperature because of (1) their high resolution and bias correction with quality-controlled ground and satellite observations[21] and (2) both MSWX temperature and rainfall datasets exhibit high correlation (>0.98) with the averages of all the observations (Supplementary Table 2). We generated differences between the other observations and the MSWX datasets for the intercomparison. Hereafter, unless described otherwise, "difference" means a difference (bias) of an observational or a model dataset compared to the MSWX dataset.

Figure 1a–f shows the differences among temperature and precipitation observations on seasonal and annual timescales for the global and 21 regional areas (see Table 1). Considerable inter-observational differences are seen in the high-topographic, polar, and desert areas for precipitation and temperature. In the case of precipitation (Fig. 1a–c), the inter-observational spread for the spatial mean difference ranges from −30 to 30%, the pattern correlation ranges from 0.7 to 0.99, and the interannual variability ranges from 0.7 to 1.5. The spread of these statistics is greater for some regions, including Northern Europe, the Sahara, South and

North Asia, Tibet, Central America, and Greenland. In the case of temperature (Fig. 1d–f), the inter-observational spread for spatial mean difference, pattern correlation, and interannual variability values are −1.5 °C–1 °C, 0.8–0.99, and 0.75–1.25, respectively. However, the seasonal inter-observational spread is larger over some of the regions. For example, the spatial mean difference values vary by up to −3 °C over Alaska, Greenland, Tibet, and North Asia; the pattern correlation values vary by 0.7 over Alaska, Southern Africa, the Sahara, and Southeast Asia; and the interannual variability values vary by 1.4 over Central America and South Africa. From the perspective of observational density, particularly in data-sparse regions, such as the Sahara and Tibet, the Climate Prediction Center (CPC) and Global Precipitation Climatology Project datasets show the largest differences in precipitation. Notably, the spread was relatively high over the relatively better-sampled region of South Asia for summer. However, the CPC and National Centers for Environmental Prediction datasets exhibit the largest differences in temperature, mainly over data-sparse regions.

In addition to inter-observational differences, we estimated intermodel differences with respect to the MSWX datasets (Fig. 1g–l; see Supplementary Table 3 for the list of models). For precipitation, the global and regional intermodel differences in the spatial mean, pattern correlation, and interannual variability vary from −30% to 60%, 0.5 to 0.99, and 0.5 to 2, respectively, except for Southern Africa, the Sahara, Tibet, South and North Asia, Alaska, Central America, and Australia, where this range is higher for the summer and winter seasons. This indicated that a large set of global climate models is also affected by poor correlations with the observed data and overestimations of interannual precipitation variability in several regions. More than half of the CMIP6 models exhibit a predominantly wet bias. The intermodel spatial mean bias, pattern correlation, and interannual variability for temperature vary from −3 °C to 2 °C, 0.9 to 0.99, and 0.75 to 1.75, respectively. Further, the majority of CMIP6 models overestimate the interannual temperature variability and exhibit a cold bias, which is commensurate with the previously mentioned wet bias. However, we must note that there are also a few models that underestimate the interannual temperature variability and exhibit a warm bias.

In principle, model uncertainty should be greater than the observed uncertainty. Since the robustness of model performance and intermodel performance differences are subject to uncertainty across observational datasets (Supplementary Fig. 3), we estimated observed-to-model uncertainty ratios (Fig. 2), as defined in "Methods: Uncertainty intercomparison." For precipitation, the model uncertainties in spatial mean bias and interannual variability are comparable to the observed uncertainties in all CMIP vintages on a global scale (Fig. 2a). Specifically, for regions such as Greenland, Alaska, Northern Europe, Tibet, North and Central Asia, and Western North America for all CMIP vintages, model uncertainties in spatial mean bias and those in interannual variability are either comparable or lesser to the corresponding observational uncertainties. In contrast, for the rest of the world, model uncertainty dominates over the observed uncertainty (Fig. 2a). For temperature, the model uncertainties for spatial mean difference and interannual variability were greater than four times at the global scale and two times for most regional areas, compared to the observed uncertainty for all CMIP vintages (Fig. 2b). However, the model uncertainties were only slightly comparable to the observed uncertainties, specifically in the interannual variability for South Africa, Central and Western North America (Fig. 2b). Nonetheless, the model uncertainties in the case of pattern correlations for temperature were comparable to observed uncertainties globally and in most regional areas, with the observed uncertainty being more predominant over Alaska, Greenland, Tibet, Sahara, Mediterranean Basin, South, Southeast and Central Asia, Western North America, and South Africa.
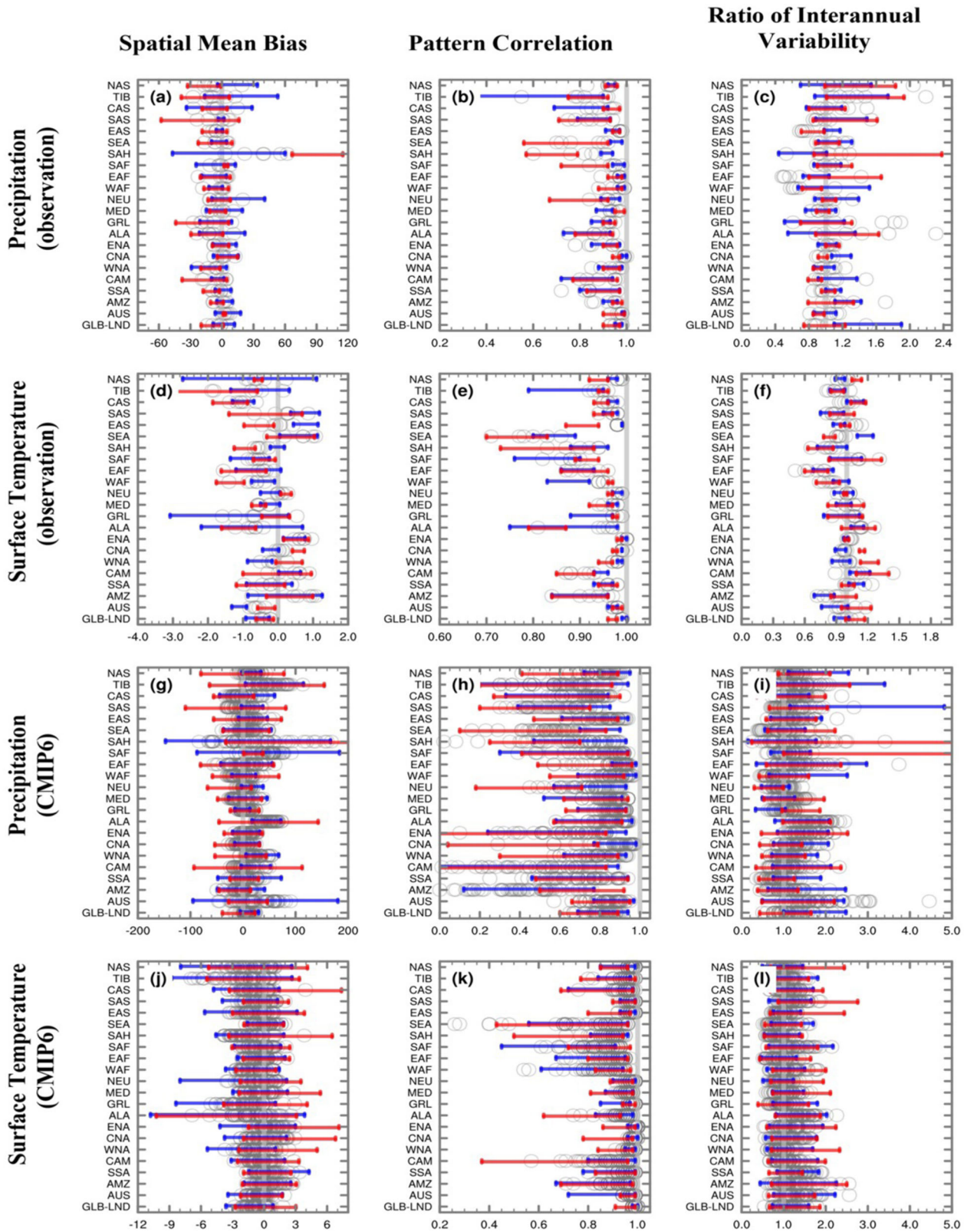
Fig. 1 Inter-observation/model spread for the spatial mean bias, pattern correlation, and ratio of interannual variability. a–c For observed precipitation, **d–f** for observed surface temperature, **g–i** for CMIP6 model simulated precipitation, and **j–l** for CMIP6 model simulated surface temperature. The y-axis shows the inter-observation/model spread for global land and the 21 different regional areas. The gray circles represent the annual mean values of corresponding observations/models. The red and blue lines represent the minimum to maximum of the inter-observation/model spread for June–August and December–February seasonal means, respectively. The spatial mean bias for precipitation and temperature is shown in percentage and degree Celsius, while the pattern correlation and the ratio of interannual variability are unitless.

Observed uncertainty, particularly when greater or comparable to model uncertainty (Supplementary Figs. 4 and 5), is a matter of grave concern[13], as observational uncertainty contributes greatly to the overall model performance uncertainty (see Supplementary Fig. 3 in the Supplementary Material). In this context, it is reasonable to ask whether these high uncertainty ratios are largely due to intermodel similarity.

Model structure similarity

Figure 3 shows a dendrogram comprising CMIP6 and its ancestral models, obtained by applying the hierarchal clustering on the

Table 1. List of 21 regional areas as defined in Giorgi and Francisco[32].

Sr. No.	Name	Acronym	Latitude (°)	Longitude (°)
1.	Australia	AUS	45S-11S	110E-155E
2.	Amazon Basin	AMZ	20S-12N	82W-34W
3.	Southern South America	SSA	56S-20S	76W-40W
4.	Central America	CAM	10N-30N	116W-83W
5.	Western North America	WNA	30N-60N	130W-103W
6.	Central North America	CNA	30N-50N	103W-85W
7.	Eastern North America	ENA	25N-50N	85W-60W
8.	Alaska	ALA	60N-72N	170W-103W
9.	Greenland	GRL	50N-85N	103W-10W
10.	Mediterranean Basin	MED	30N-48N	10W-40E
11.	Northern Europe	NEU	48N-75N	10W-40E
12.	Western Africa	WAF	12S-18N	20W-22E
13.	Eastern Africa	EAF	12S-18N	22E-52E
14.	Southern Africa	SAF	35S-12S	10W-52E
15.	Sahara	SAH	18N-30N	20W-65E
16.	Southeast Asia	SEA	11S-20N	95E-155E
17.	East Asia	EAS	20N-50N	100E-145E
18.	South Asia	SAS	5N-30N	65E-100E
19.	Central Asia	CAS	30N-50N	40E-75E
20.	Tibet	TIB	30N-50N	75E-100E
21.	North Asia	NAS	50N-70N	40E-180E

The region name, acronyms, and the details of latitude and longitude. Only land grid points are considered for the analysis.

weighted pairwise distance of intermodel correlation[22] for temperature and precipitation, as discussed in the methodology. Models under the lowest branches in Fig. 3 are most similar. As we move up the tree, the similarity between the models, or groups of models, progressively decreases. The farther the models or groups of models are from each other, the more they are dissimilar. Remarkably, the model pairings with high similarities (i.e., correlation >0.72) are those that had been developed at the same institute, or they use the same model components/features despite having been developed at different institutes (Fig. 3a). A similar inference has been made for the CMIP3 and CMIP5 models[5,9,10,17] (also see Fig. 3b). For example, the model pairs of the CMCC-CM2-SR5 and CMCC-ESM2; CESM2-FV2 and CESM-WACCM-FV2; and CanESM5 and CanESM5-CanOE, which were developed at the same institute, are significantly similar (Fig. 3a); so are the KACE-1-0-G and ACCESS-CM2; FIO-ESM-2-0 and TaiESM1; and MPI-ESM-1-2-HAM and AWI-ESM-1-1-LR despite having been developed at the different institutes (Fig. 3a). As mentioned above, even when we move up the tree, the signatures of similarity are still present. For instance, owing to the similarity due to shared dynamics or/and physics, the four Max Plank Institute for Meteorology (MPI-M) model variants appear as a broad group under branch 22 (Fig. 3a). We also see in Fig. 3a similar groups of of models derived from the same parent model, such as the six Met Office Hadley Center (MOHC) model-derivatives, the nine European Centers Consortium (EC-Cons) models, and twelve National Center for Atmospheric Research (NCAR) models.

Curiously, the Indian Institute of Tropical Meteorology (IITM) model exhibits significant similarities with the EC-Cons models, as evidenced by it being clustered in branch 20, despite having no direct connection with them (Fig. 3a). This pairing is an obvious exception, unlike all other models that fall into the same group, owing to having the same or similar physics components/schemes or same resolution, or have a common dominant component (e.g., atmospheric or oceanic general circulation model), etc.[23,24] (see Supplementary Table 4 for more details on the groups of CMIP6 models that share large portions of common components/features). We speculate that the afore-discussed grouping of the IITM model into the group dominated by the EC-Cons models (Fig. 3) is probably a statistical artifact.

A similar analysis carried out with all vintages of CMIP models highlights the contribution of the lineage of the model to the similarity with its predecessors across the generations. For example, all NCAR models across the vintages form a broad group (Fig. 3b).
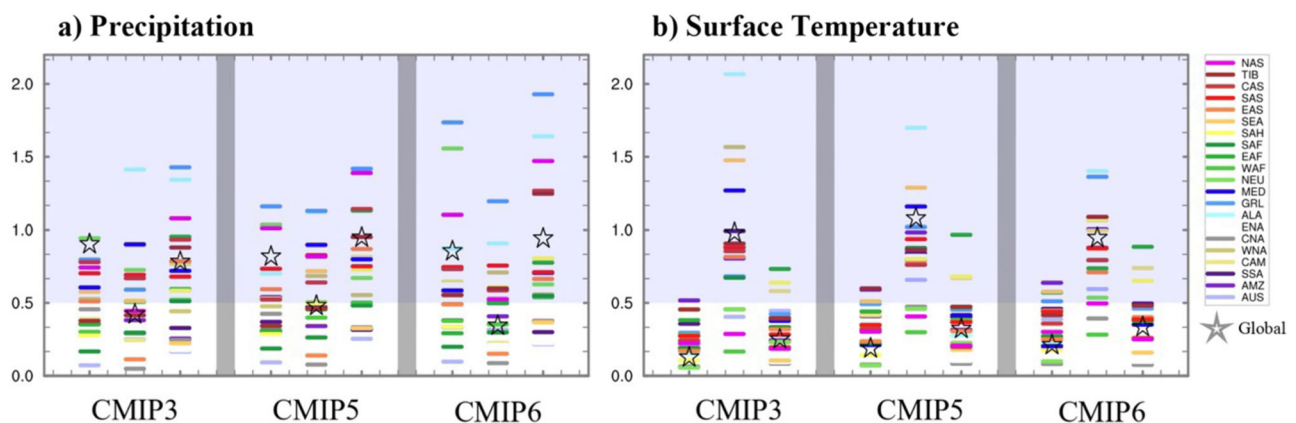

Fig. 2 Global and regional intercomparison of observed-to-model uncertainty ratio. a Annual mean precipitation and **b** surface temperature. The observed-to-model uncertainty ratio is computed for CMIP3, CMIP5, and CMIP6 models for three different metrics, namely, spatial mean bias, pattern correlation, and the ratio of interannual variability (from left to right in each CMIPs). The colored bars represent different regional areas, and the gray star represents the global region. The ratio larger than 0.5 indicates that the observational uncertainty significantly contributes to the model uncertainty at the 95% confidence level (shown in light-blue shading).
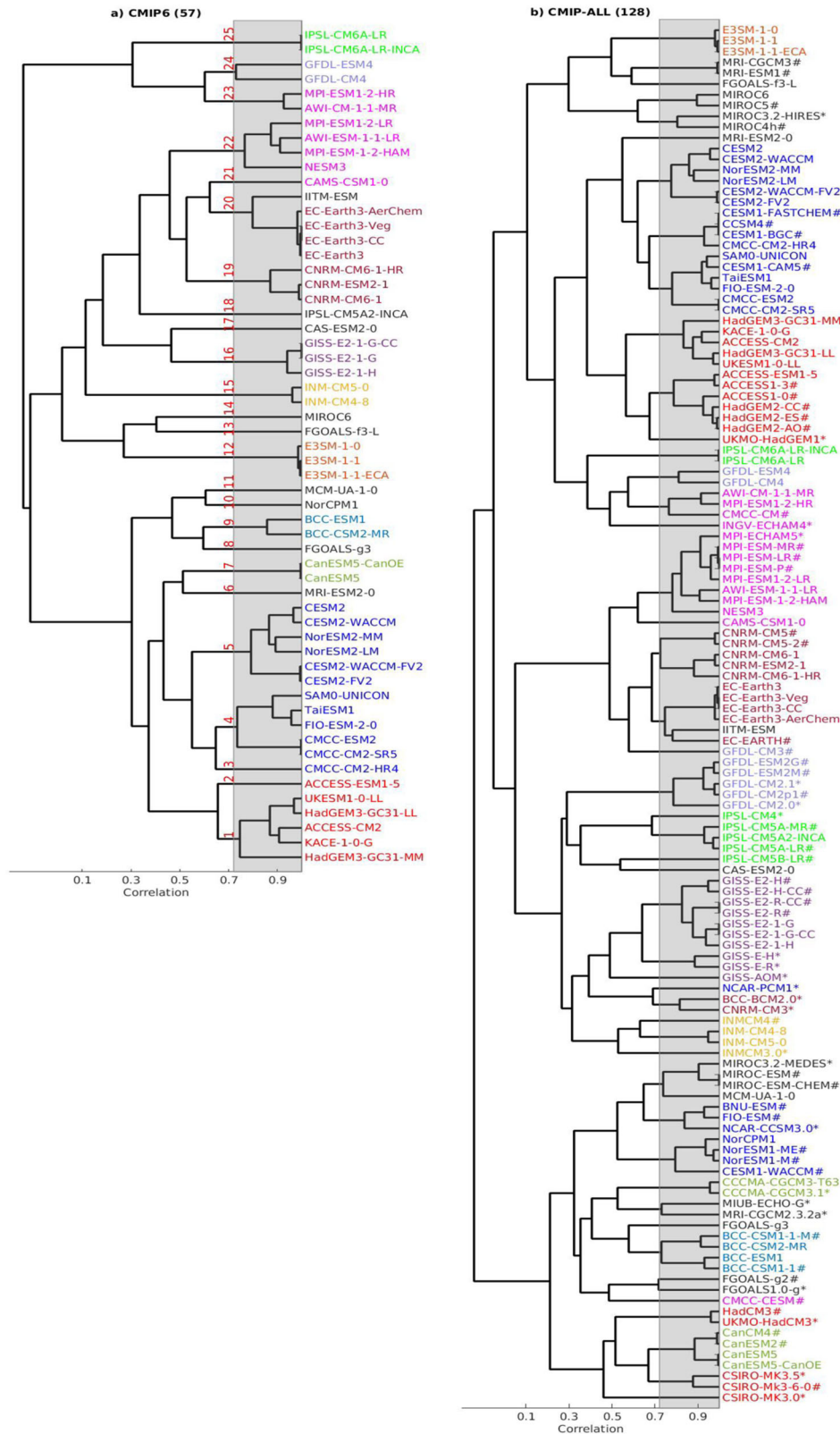
Fig. 3 Hierarchical clustering of different climate models on the global scale. a CMIP6 with 57 models and **b** CMIP6 with the previous two CMIP vintages (i.e., CMIP3 and CMIP5) with 128 models. The models that merged with a high correlation value are highly similar and share a lot of structural characteristics. A weighted pairwise distance algorithm was used for model clustering, where the distance between two similar/dissimilar models was calculated as one minus pattern correlation (see Eq. 5). The values on the x-axis represent equivalent correlations, interpreted as the level of model similarity/dissimilarity. Significantly similar models, indicated by gray shading, merged at correlation ($r > or = 0.72$) with a confidence level of 99%. Models with obvious code similarities or produced by the same institution/center are marked in the same color. The significantly dissimilar models branch in CMIP6 models clustering are marked by the numbers (**a**).
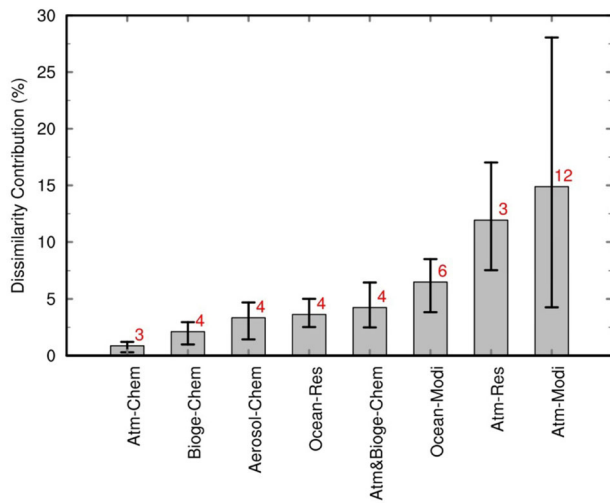
Fig. 4 Dissimilarity contribution by model component. The vertical bar indicates the mean dissimilarity contribution of a particular model component, and the vertical line shows the minimum to maximum dissimilarity range. The number of model pair sets used for a component is shown in red. The dissimilarity contribution was obtained by adding a component in a model to its predecessor/parent model or modifying it with respect to its predecessor/parent model.

Sensitivity of model components to similarity

A few studies have hinted that shared atmospheric components potentially contribute to the similarity of multimodel outputs[5,10,17]. This is likely due to the implementation of similar atmospheric physics schemes, such as, for example, convective parameterization, in many climate models. In fact, models may have the same atmospheric or oceanic modules, which also can be expected to contribute to the model similarity more substantially. For instance, the two NCAR models—the CESM2-FV2 and CESM2-WACCM-FV2 are the same, except that the latter model includes a comprehensive chemistry mechanism with a model top of 130 km, compared to a simplified chemistry module with a top height of 40 km in the first model. This leads to a high similarity score (i.e., low dissimilarity score) between them (Fig. 3a). Note that the intermodel similarity may likewise arise from the atmospheric or oceanic modules having the same resolution.

In this context, we carried out an exercise to quantify the individual contributions to the intermodel similarity by different typical shared components/features, such as convection, resolution, chemistry, etc. For this purpose, we identified pairs of models, which only differ either in the use of a model component/physics package or resolution (Supplementary Table 5). All other features across the pair are the same and could even appear very close to one another in Fig. 3a, b. In other words, one model of each identified pair contains an additional component or modified version of the same component. For example, in the pair of IPSL-CM5A-LR and IPSL-CM5B-LR models, IPSL-CM5B-LR uses a set of different physical parameterizations in the atmospheric component compared to IPSL-CM5A-LR model, while all other features remain the same (Fig. 3b). We then calculated the contribution from this distinct component/feature to the dissimilarity as a dissimilarity index, defined as the average of $D_{i,j}^{mod}$ for all model pairs (Eq. 6), where D is the distance (see "Methods: Model dissimilarity").

Out of all the 128 CMIP model outputs we analyzed (from all three vintages; see Supplementary Table 3), we identified several distinct pairs of models. These pairs and the distinctions are detailed in Supplementary Table 5. Six model pairs (i.e., 12 models) have different/modified ocean components. Thirteen model pairs

have different/modified atmospheric components (in terms of convective parameterizations, cloud model, radiation schemes, etc.). Three sets of four model pairs (i.e., 8 models) have different aerosol treatments, ocean biogeochemistry, and oceanic horizontal resolutions, respectively. Three model pairs have components that differ only in resolution, and three model pairs have different/modified atmospheric chemistry. Given such distinctions, we expected that different/modified or additional components would increase model dissimilarity, and a greater change in the dissimilarity of a model due to different/modified or additional components would indicate a greater contribution of that component to model dissimilarity.

Figure 4 demonstrates the percentage contribution of adding a component/feature in a model-to-model dissimilarity. The contributions of atmospheric chemistry, ocean biogeochemistry, and interactive aerosol to dissimilarity are, respectively, ~0.86%, 2.11%, and 3.34% (Fig. 4). Furthermore, adding atmospheric chemistry and oceanic biogeochemistry together contributed to model dissimilarity by 4.24%; however, this was not a simple linear combination of individual model contributions. These contributions to dissimilarity are relatively less than those arising from distinct atmospheric and/or ocean modules and resolution. For example, the dissimilarity index associated with a modified or different ocean component is a relatively high 6.49%. Furthermore, doubling horizontal ocean resolution results in an additional increase of 4.24%. Likewise, adding a modified/distinct atmospheric component to a model increases its dissimilarity by an average of 13.94%. The dissimilarity associated with atmospheric modification ranges from a relatively low 2.54% due to increasing vertical atmospheric levels to a very high 28.06% related to major parameterization distinctions (Fig. 4). Doubling the horizontal resolution of the atmospheric model contributed to dissimilarity by an average of 11.93%, ranging from 7.54 to 17.02% across three pairs (Fig. 4). Apparently, having distinct atmospheric components and/or resolution introduces relatively higher intermodel dissimilarity compared to atmospheric chemistry or ocean modules. By analogy, having the same atmospheric components and/or resolution can be expected to add to the intermodel similarity equally. As a result, any modest changes in the features of the atmospheric component of a GCM may enhance their dissimilarity level.

Effect of similarity on the assumption of intermodel independence

Given the dependencies associated with model code/component sharing, we quantified the statistical independence of the various models, as discussed in "Results: Sensitivity of model components to similarity." This was estimated by computing the effective number of "independent" climate models M_{eff} compared to the actual number of models considered (M) using Eq. 7. As each CMIP group contained multiple models, we also note that M_{eff} would change based on M. In this context, we estimated the sensitivity of M_{eff} to changes in M. Figure 5 shows the variation of M_{eff} with respect to M. The M_{eff} is 10.7 for CMIP3 but 11 for both CMIP5 and CMIP6, which is surprising given the tremendous improvements and developments in CMIP6 models. Similar M_{eff} statistics were obtained regionally (figure not shown). While approximately 55% of the CMIP3 models were similar, this number increased to 76% for CMIP5 models and 80% for CMIP6 models, indicating a decline in dissimilar CMIP models. Figure 5 also shows a concave-shaped evolution of M_{eff} with respect to M, indicating that the contribution of a model to effective models decreases asymptotically as the model number, over which the averages were obtained, increases. This indicates that information from the additional model to the multimodel ensemble overlapped with prior models, reducing the number of effective climate models due to lesser and lesser new information[9]. For example, although
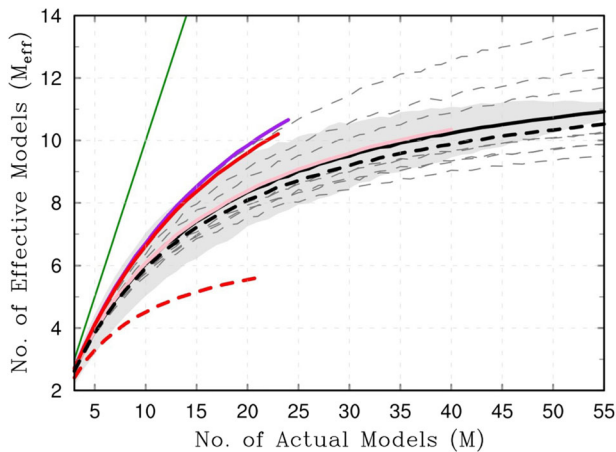
Fig. 5 Sensitivity of the number of effective climate models (M_{eff}) to changes in actual climate models (M) at the global scale. The solid black line represents the M_{eff} averages over precipitation and temperature and over seasons for the CMIP6 models from 1980 to 1999, with the gray shaded area representing 95% confidence intervals. The thick dotted black line represents the M_{eff} averages over precipitation and temperature over seasons for the CMIP6 models from 2015 to 2021 under SSP5–8.6. The dashed gray lines represent the M_{eff} for CMIP6 precipitation and temperature across the four seasons. The thick red and dashed red lines represent the M_{eff} of most dissimilar CMIP6 models and ensembles of models from either the same institute or from a different institute that largely shared the model components, respectively. The solid pink and purple lines represent the M_{eff} values averages over precipitation and temperature over seasons for CMIP5 and CMIP3 models, respectively. The straight green line represents $M_{eff} = M$.

we upscaled from a 50-model ensemble mean to a 55-model ensemble mean, the effective model size only increased by 0.2. However, increasing the number of models from 15 to 20 considerably increased the M_{eff} by 1.

We repeated the computation of M_{eff} using select sets of CMIP6 models to quantify the sensitivity of the M_{eff} to the similarity effect. This was conducted using two ensembles: one for all similar model pairs from three source groups, accounting to 21 models (hereafter referred to as 21CMIP6, with eight models from the NCAR group, six models from MOHC, and seven models from the EC-Cons group) (Supplementary Table 4). In contrast, the other model ensemble is for all dissimilar model pairs, accounting to 23 models (hereafter referred to as 23CMIP6, with only one parent model from a pair/group of similar models; see Fig. 3a/ Supplementary Table 4). The M_{eff} increased rapidly with an increase in dissimilar models up to M 15 (Fig. 5), but the rate of increase flattened beyond this for further increases of M. This was because these models also had some similarities to other models even though they form distinct pairs with them. M_{eff} decreased more rapidly with increasing numbers of similar models in the ensemble (Fig. 5). If one of the models was removed from the same-source model pairs (e.g., models from NCAR or MOHC), there was only a marginal increase in M_{eff}. The effected model similarity became increasingly problematic as shoot-off models with large duplications increased. The resultant reduction in M_{eff} greatly discredited the assumption of model independence.

Similarity effects on climate change estimates

The M_{eff} for the projection period 2015–2021 under SSP5–8.5 is the same as that for the corresponding historical period (1980–1999) (Fig. 5). However, this low M_{eff} may change in far-future simulations if model components become distinct or if models with common components, as seen in CMIP6 vintage

models, behave dissimilarly due to increased forcings[10,13]. This would also be associated with changes in physical processes due to changes in mean climate (e.g., shifting from a cumulus convection regime to a large-scale rain regime). However, in the near future, we expect future climate simulations to retain model similarities.

For this analysis, we used 37 of 57 CMIP6 models that had future surface temperature and precipitation projections for both SSP5–8.5 and SSP1-2.6 scenarios (see Supplementary Table 4 for the list of 37 CMIP6 models). We analyzed multimodel mean temperature and rainfall averages from two model sets: one from 23CMIP6 and the other from all 37 models (hereafter referred to as 37CMIP6). Relative to the 37CMIP6, the 23CMIP6 showed lesser changes in mean temperature and precipitation by the end of the 21st century when compared to the current climate of 1980–2014 (Fig. 6a–d). Specifically, while the mean projections of 37CMIP6 indicate a 4.5 °C rise in temperature and a 7.5% increase in precipitation, the 23CMIP6 mean shows 0.25 °C less warming and 0.75% less precipitation by the end of the 21st century. For the SSP1–2.6 projections, 23CMIP6 shows ~0.2 °C less warming and ~0.4% less precipitation relative to the corresponding 37CMIP6, for which the corresponding values are 1.25 °C for temperature and 2.25% for precipitation (figure not supplied).

Further, to highlight contributions from similar models, we also showed the multimodel averages of climate variables from 21CMIP6. This allows us to unravel how model selection with similar models influences climate change estimates even when the number of sampled models is large. We found that the projected mean temperature and precipitation increase, respectively, by ~0.5 °C ~1% by the end of the 21st century. These projected changes by the 21CMIP6 models are much higher than those projected by the dissimilar models. Among these models, most of the MOHC models particularly exhibited larger changes compared to 37CMIP6 and dominated the results of 21CMIP6 and, thereby, 37CMIP6 (Fig. 6b, d).

Understandably, there is also a modest reduction in uncertainty in 21CMIP6 relative to 37CMIP6, and more so relative to the 23CMIP6 models because the 21CMIP6 models contain mostly similar models (Fig. 6a, c). In particular, the high similarity in the climate variables from the EC-Cons group greatly reduced uncertainty compared to the other groups. These results indicate that having such a high number of similar models reduces the intermodel uncertainty of a larger model group, such as 21CMIP6 or 37CMIP6, and skews the mean estimates toward the dominant subgroup of models. Critically, the intermodel uncertainty across 23CMIP6, the set of dissimilar models, increases over time.

Similarly, on a regional scale, the 23CMIP6, compared to the mean change of 37CMIP6, showed smaller changes in mean surface temperature and precipitation over Central America, Tibet, Australia, South and North Asia, Southern Africa, and the Sahara (0.2 °C–0.7 °C for temperature and 2–5% for precipitation), except for larger changes in precipitation over East Asia and western Africa in the far future (2081–2100) (Fig. 7a, b). The relatively larger multimodel mean surface temperature and precipitation in the far future for most regions projected by the 37CMIP models relative to the 23CMIP6 models are mainly due to the corresponding higher contributions from the 21CMIP6 models to the 37CMIP6 models. To be more specific, the multimodel mean temperature and precipitation from the 21CMIP6 models are higher than those of the 37CMIP6 by 0.2–0.8 °C (Fig. 7a) and 2–4% (Fig. 7b), respectively. This is a clear example of how the common biases in a group of similar models can also bias the projections from a larger ensemble of models containing this group of models. Interestingly, most of these regions have also been reported as hotspots of concern due to observational uncertainty as well (see "Results: Observation and model dataset uncertainty"). The wide uncertainty for 23CMIP6 and the somewhat smaller uncertainty for
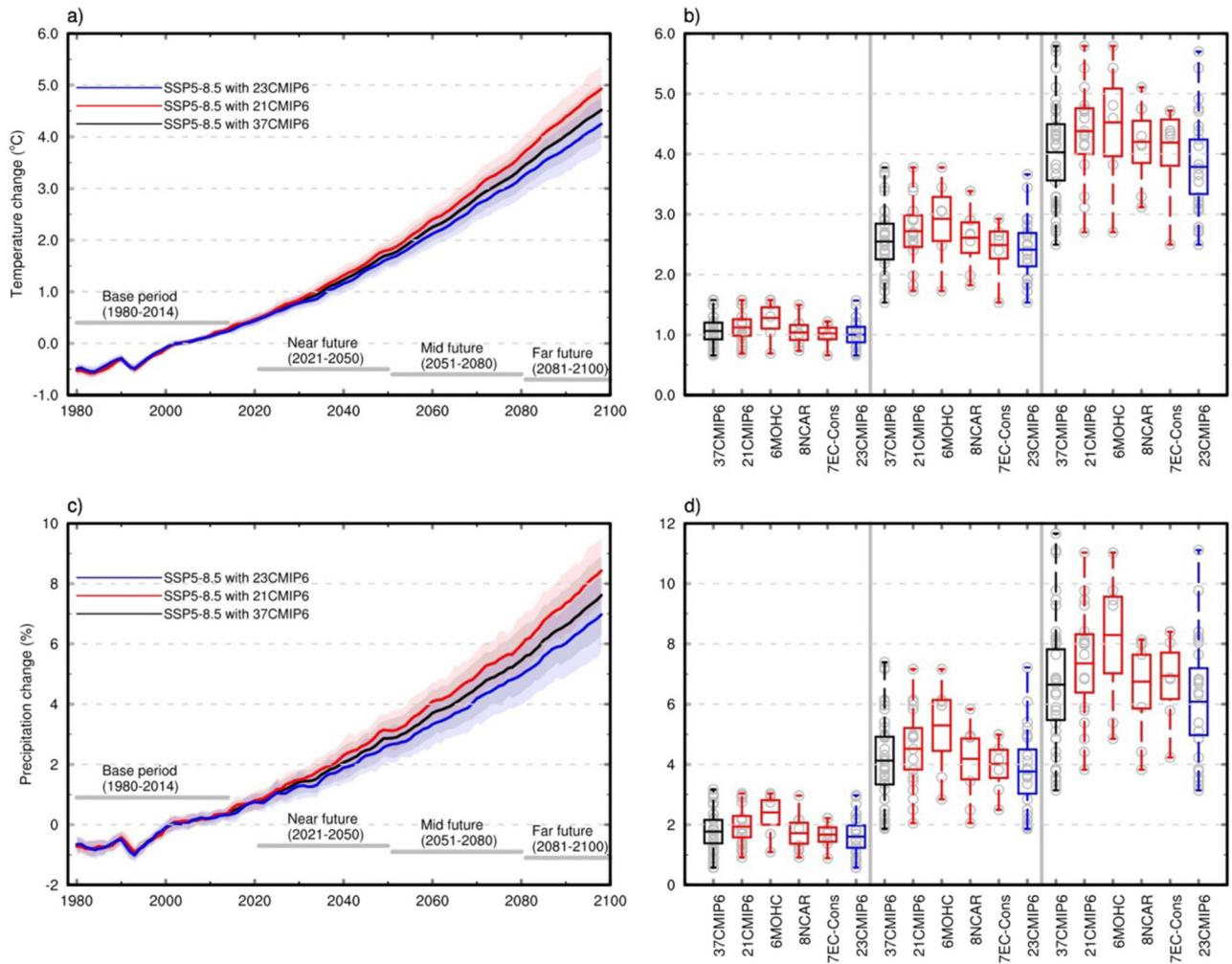
Fig. 6 Projected future changes in global temperature and precipitation relative to the current climate of 1980–2014 under SSP5–8.5. The time-series and box-whisker plots for **a**, **b** surface temperature and **c**, **d** precipitation. The time-series and box-whisker plots in black color indicate the mean values of 37CMIP6 models. The time-series and box-whisker plots in red and blue color indicate the mean values of 21CMIP6 (i.e., all common model pairs belonging to three major groups (MOHC: Met Office Hadley Center, NCAR: National Center for Atmospheric Research, and EC-Cons: European Center Consortium)) and 23CMIP6 models (i.e., all dissimilar models), respectively. The uncertainty in climate change projections from 37CMIP6, 21CMIP6, and 23CMIP6 is reflected by one standard deviation of the intermodel spread in black, red, and blue shadings. The box-whisker plots are plotted for three different projection periods, namely, the near-future (2021–2050), mid-future (2051–2080), and far-future (2081–2100), and indicate means and uncertainties for the respective periods. The individual model projected values are shown by gray circles in the box-whisker plots. The numbers before CMIP6/model groups in the plot indicate the total number of models used for computation.

21CMIP6 could also be seen in most regional areas, as expected from the analysis of M_{eff} (Fig. 5).

DISCUSSION

This study aims to unravel the effects of model and observational data uncertainty and those of intermodel similarity on current climate simulations and future climate projections of temperature and precipitation. We used seven gridded observations/reanalysis datasets and historical and projection datasets of 128 GCMs from CMIP6, CMIP5, and CMIP3 vintages. Our analysis suggests that while the seasonal climatological mean values from seven observational datasets agreed with one another when averaged globally, they showed localized differences. These differences translate into differences in GCM performance scores. In particular, the high-topographic, polar, and desert areas exhibit large seasonal temperature and precipitation differences. On the regional scale, the largest differences in temperature observations are seen over Tibet, North and South Asia, Central America, Alaska,

and Greenland, and precipitation observations for Tibet, North and South Asia, the Sahara, Northern Europe, and Greenland. Although most of the regions where the uncertainties are high are, expectedly, observation-sparse regions, observational uncertainties are also prominent in relatively well-sampled regions, such as South Asia, in terms of summer rainfall.

Our results indicate that the uncertainty associated with the choice of gridded observational temperature datasets is, in general, smaller than that associated with the uncertainty across the models. Unfortunately, unlike for the temperature, the influence of observational uncertainty across the observed precipitation datasets is seen to be large, both in terms of spatial mean as well as the interannual variability. The choice of gridded observational datasets becomes important globally, with the greatest impact over Greenland, Alaska, Northern Europe, Tibet, and North, South and Central Asia. These results suggest that temperature projections from all CMIP vintages are more reliable than corresponding projected rainfall. This means that the currently available multiple temperature observations are
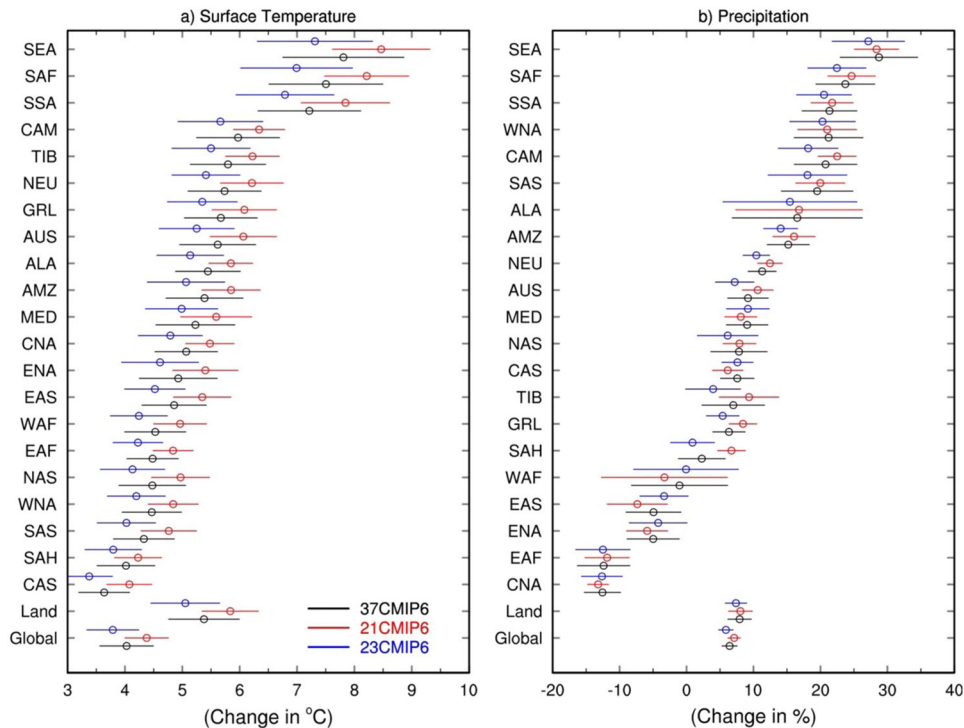
Fig. 7   Regional precipitation and temperature change in the far future (2081–2100) relative to the current climate of 1980–2014 under SSP5-8.5. Changes in global and regional mean **a** surface temperature and **b** precipitation for 37CMIP6 (black), 21CMIP6 (red), and 23CMIP6 (blue) models. The uncertainty in projected far-future climate change estimates for surface temperature and precipitation are reflected by one standard deviation of the intermodel spread in the horizontal line for respective regions. The regional areas in the figure are arranged in ascending order of increasing mean surface temperature change from 37CMIP6 models.

acceptable for evaluating and validating model statistics, such as simulated mean temperature or its variability when they are within the limit of observational uncertainty.

Importantly, from an exhaustive analysis of commonality in various model components, we found that the intermodel similarity increased from CMIP3 to CMIP6. The similarity increased from 55% in 24 CMIP3 through ~76% in 47 CMIP5, reaching a high of 80% in 57 CMIP6 models. While we found 10.7 effective (that is, independent) climate models in the case of CMIP3, the corresponding effective model number for CMIP5 is a marginally higher 11 and remains the same even that for the CMIP6 models. The disparity between the number of effective climate models and the number of actual models, particularly in the case of CMIP6, is due to the high degree of similarity in the contributory models. Most of the similarity is because such models are developed at the same institute or originate from the same predecessor and have many common dynamics and physical parameterizations. Among the current climate model modules, the highest to lowest contributors to the model similarity are the atmospheric, oceanic, aerosol, ocean biogeochemistry, and atmospheric chemistry components. However, the high similarity among CMIP6 models relative to CMIP3 models does not mean that the CMIP6 projections are of no additional importance, as these have better physics and resolutions and, consequently, better climate statistics.

The area-averaged annual temperature and rainfall from the ensemble of all CMIP6 models are projected to increase by 4.5 °C and 91.3 mm (i.e., a 7.5% increase), respectively, by the end of the 21st century under the SSP8.5 scenario, while the dissimilar models suggest projected increases of 4.25 °C in temperature and 84 mm (i.e., 6.75%) in precipitation. On the regional scale, the dissimilar models showed far less increases of 0.2–0.8 °C in temperature and 2–5% in precipitation relative to the ensemble of all models. The disparity in the results of the dissimilar models

relative to those from all models can also be seen in other climate change scenarios, such as SSP1-2.6. These results have implications for the development of future models. In fact, these results call for a multi-institutional collaboration, which should be more focused, coordinated, and targeted to address the similarity sources. Such a carefully designed multi-institutional collaboration would enable to synthesize the available physics packages and dynamics configurations, introduce diverse combinations based on the existing packages, and, most importantly, design physics packages based on ever-increasing observational in-situ and remotely sensed datasets, particularly those from modern campaigns that are ongoing around the globe. Increased intermodel diversity should provide more reliable intermodel uncertainty, which is currently underestimated due to model similarity. In this context, prioritizing the critical model components identified in this work will pave the way for developing advanced model projections with higher reliability. Finally, our results on model similarity have practical implications for application scientists and policymakers who use multimodel outputs for adaptation planning.

METHODS

Data

We used historical precipitation and surface temperature data from 24 CMIP3, 47 CMIP5, and 57 CMIP6 models (Supplementary Table 3) and CMIP6 projection data from the business-as-usual climate change scenario (SSP5–8.5). Most CMIP6 models are updated/improved versions of their predecessors in terms of parameterization, resolution, earth system components, coupling, and tuning[3,25]. The outputs of these climate models were downloaded from the website https://data.ceda.ac.uk/badc, provided by the Center for Environmental Data Analysis of the National Center for Atmospheric Science, United Kingdom.

The observed surface temperatures were taken from the Global Historical Climatology Network, National Oceanic and Atmospheric Administration, University of Delaware (UoDel), Climate Research Unit (CRU), National Centers for Environmental Prediction (NCEP), Climate Prediction Center (CPC), and Multi-Source Weather (MSWX) databases. Although the NCEP and MSWX datasets are essentially reanalysis products, they are commonly used as proxies for observations; therefore, in this study, they were deemed observations when evaluating observational and model uncertainties (e.g., refs. [15,21]). The precipitation datasets we analyzed included the Global Precipitation Climatology Center, Global Precipitation Climatology Project (GPCP), UoDel, CPC, CPC Merged Analysis of Precipitation, CRU, and MSWX datasets. Details of the observations are presented in the Supplementary Materials (Supplementary Table 1). All these observational datasets were downloaded from the website https://psl.noaa.gov/data/gridded/index.html, provided by the Physical Science Laboratory of the National Oceanic and Atmospheric Administration, United States of America.

Notably, we only used surface temperature and precipitation, as these variables have been found to be sufficient for model genealogy investigations (e.g., refs. [10,17]). The model outputs and observational datasets were bilinearly interpolated to a horizontal resolution of one degree (e.g., refs. [26,27]).

Uncertainty intercomparison

We used three performance-based metrics, spatial mean bias, pattern correlation, and the ratio of interannual variability, to compare observational and model uncertainties both globally and regionally over the 21 different regions (see Table 1). The reasons for selecting these three metrics are as follows: the spatial mean bias provides an estimate of how accurate the values from one dataset/model are in comparison to the actual values, while the pattern correlation indicates the degree to which the spatial patterns of two different datasets are structurally similar/dissimilar. In contrast, the ratio of interannual variability indicates how well the interannual variations of one dataset/model are compared to the actual variations, which is important for comprehending the severity of past and future climate events. These metrics were calculated for each model m, variable q, season k, and region r against each observation c. For each model with a total of N grid points, we defined o_n and y_n as monthly observational and climate model data at grid point n in region r. Furthermore, in the following, an overbar indicates the temporal mean across all time steps for a season k in the study period, while two overbars reflect the temporal mean across all time steps in a year that falls into season k and spatial means.

Spatial mean performance (P_S) was evaluated by averaging the climatological bias (e) between a model simulation and an observation over all grid points as follows:

$$e_{n,m,c,r}^{q,k} = \left(\overline{y}_{n,m,r}^{q,k} - \overline{o}_{n,c,r}^{q,k} \right) \tag{1}$$

$$P_S^{q,k,m,c,r} = \sum_{n=1}^{N} e_{n,m,c,r}^{q,k} \tag{2a}$$

Spatial pattern similarity performance (P_R) was assessed using pattern correlation, which, for clarity, was expressed by dropping the indices q, k, m, c, r as follows:

$$P_R = \frac{cov(\overline{y}_n, \overline{o}_n)}{\sigma(\overline{y}_n)\sigma(\overline{o}_n)}, n = 1 \dots . N \tag{2b}$$

where cov and σ denote spatial covariance and standard deviation, respectively.

Interannual variability performance (P_σ) was evaluated using the ratio of the standard deviation of a yearly varying model

simulation and observation as follows:

$$P_\sigma = \frac{\sigma(\overline{\overline{y}})}{\sigma(\overline{\overline{o}})} \tag{2c}$$

Further, to diagnose the relationship between the influence of the choice of reference data on the observational uncertainty and the influence of the choice of a GCM on the model uncertainty, these were quantified using the methodology suggested by Kotlarski et al.[12] Briefly, observational uncertainty was defined as the mean standard deviation of the metric values when a GCM was compared to all seven reference datasets as follows:

$$U_{obs} = \frac{\sum_{m=1}^{M} \sqrt{\frac{1}{6} \sum_{c=1}^{7} \left(P_{m,c} - \frac{1}{7} \sum_{c=1}^{7} P_{m,c} \right)^2}}{M} \tag{3a}$$

where M is 57 for CMIP6, 47 for CMIP5, and 24 for CMIP3, and $P_{m,c}$ is the value of a performance metric for a given variable, season, and region when using the reference dataset c ($c \in \{1, 2, 3, \dots 7\}$ for evaluating GCM m ($m \in \{1, 2, \dots . M\}$. Note that whenever we averaged correlations, we first took the mean over Fisher's z transformed correlation coefficients and then computed their inverses to derive an average correlation (see Supplementary Methods).

Furthermore, when calculating model uncertainty, 150 ($M7$) GCM realizations were created for each CMIP vintage using the bootstrapping without replacement method to ensure consistency with several observational references. Because these realization numbers were deemed sufficient, increasing them had no discernible effect on the uncertainty calculation (Supplementary Fig. 2). The model uncertainty was thus defined as the mean standard deviation of the metric values when all 150 member realizations were compared to a particular reference dataset as follows:

$$U_{mod} = \frac{\sum_{c=1}^{7} \left(\frac{1}{150} \sum_{nr=1}^{150} \sqrt{\frac{1}{6} \sum_{m \in S_{nr}} \left(P_{m,c} - \frac{1}{7} \sum_{m \in S_n} P_{m,c} \right)^2} \right)}{7} \tag{3b}$$

In this case, the ratio of the observed to the model uncertainty defined the robustness of the evaluation, and a ratio greater than 0.5 (i.e., a model uncertainty within two standard deviations of observations) indicated that observational uncertainty significantly contributed to the overall evaluation uncertainty as follows:

$$R = \frac{U_{obs}}{U_{mod}} \tag{3c}$$

Model dissimilarity

We used independent spatial patterns to assess model dissimilarity[28]. As the models with common physics module(s) and codes may have had similar error patterns, a symmetric correlation matrix (S) was created using spatially varying error structures (Eq. 1). As the amounts of precipitation and their variability were higher over the tropics compared to other regions[29], we normalized these errors using the observed interannual standard deviation (σ) at each grid point for each variable, season, and model (Supplementary Fig. 6a, b) as follows:

$$e' = e/\sigma_o \tag{4a}$$

Individual climate models also have common errors (Supplementary Fig. 6c, d) because of unresolved physical processes or/and topography[30]. Importantly, there was a high and statistically significant correlation (~0.72) between individual model errors and the multimodel mean error (MME or $\overline{e} = \sum_{m=1}^{M} e'$) across all fields, indicating a shared systematic bias (Supplementary Fig. 1). Model similarity due to shared conceptual frameworks and codes,

by definition, means that the error patterns of variable q of model m should be related to the error in the corresponding q patterns of the other model. Thus, the MME-associated error pattern of a model can be obtained as the correlation coefficient $r(e', \overline{e})$ between e' and $\overline{e}$. The MME-independent portion of the models, in a linear sense, was determined by subtracting their MME-associated error patterns (Supplementary Fig. 6e, f) as follows:

$$d = e'^* - r(e', \overline{e}) \times \overline{e}^* \qquad (4b)$$

Where * denotes the standardization, which ensures all data fields have uniform scales before they are used in Eq. 4b.

In Eq. 4b, the MME pattern $\overline{e}^*$ and the model error pattern d are unrelated. The correlation between the two models ($S_{i,j}$) for any variable and season was given as follows:

$$S_{i,j} = r(d_i, d_j) \qquad (5)$$

The zero average correlation between model pairs showed that removing the MME signal from a model reduced its combined influence associated with unresolved physical processes or/and topography (Supplementary Fig. 1). The spatial error patterns in the annual precipitation simulated by the GFDL-ESM4 model with and without the MME-associated component are shown in Supplementary Fig. 6c, e, and those simulated by the MCM-UA-1 model are shown in Supplementary Fig. 6d, f, respectively. $S_{i,j}$ were then converted into a distance metric, $D_{i,j} = 1 - S_{i,j}$, which with a smaller (larger) value of $D_{i,j}$ ($S_{i,j}$) indicate high similarity between the two models and vice versa for dissimilar models. Further, the $D_{i,j}$ computed over all model pairs form a symmetric $M \times M$ matrix D. The matrix D is then used to rearrange the models in a hierarchical structure by applying the weighted pairwise algorithm on matrix D. The detailed description of applying the weighted pairwise algorithm for hierarchical clustering of datasets can be found in Sokal et al.[22].

The effect of GCM modification (i.e., modifications of parametrizations, resolution, and addition of a model component) on intermodel similarity was estimated as follows:

$$D_{i,j}^{mod} = \sum_{k=1, k \neq i, k \neq j}^{M} \frac{(D_{i,k} - D_{j,k})}{D_{j,k}} \times 100 \qquad (6)$$

Finally, the effective number of climate models, M_{eff}, following Pennell and Reichler[9], was calculated using eigenvalues of correlation matrix[28] S as follows:

$$M_{eff} = \frac{\left(\sum_{i=1}^{i=M} \lambda_i\right)^2}{\sum_{i=1}^{M} \lambda_i^2} \qquad (7)$$

where λ_i represents the ith eigenvalues of matrix S, M_{eff} varies between one and M, and $M_{eff} = 1(M)$ indicates that all model error structures are identical (independent).

Statistical analysis

We used a one-tailed 99% confidence limit (i.e., $r \geq 0.72$) to indicate significantly similar model pairings. This was done under the assumption that intermodel correlations (i.e., matrix S from Eq. 5) follow a Gaussian distribution, and the mean of all averaged intermodel correlations is zero (Supplementary Fig. 1).

In addition, in the case of computing M_{eff} irrespective of the choice of models, we used the bootstrapping resampling without replacement method[31] to generate 150 samples of randomly chosen models. The number of models in each sample ranged from 3 to 24 for CMIP3, 3 to 47 for CMIP5, and 3 to 57 for CMIP6. Increasing the sample numbers beyond 150 had no further discernible effect on M_{eff} and were, therefore, 150 samples considered to be sufficient for M_{eff} analysis (Supplementary Fig. 2). The interval between the 2.5 and 97.5% levels of the 150

bootstrapped samples was used to calculate the 95% confidence interval of the M_{eff} over model M.

Furthermore, the effect of observational uncertainty on model uncertainty was significant at the 95% confidence level when the model uncertainty was either lower than or equal to two times that of the observed uncertainty (i.e., the ratio of observational uncertainty to model uncertainty was larger than 0.5).

The significant levels of errors in both the multimodel mean and individual models were tested using the two-tailed Student's t-test (Supplementary Fig. 6). The independent sample size ($Neqv$) in a time series of length N (i.e., 20), used in the analysis of the two-tailed Student's t-test, was calculated as a lag-one autocorrelation ($r1$) at the 95% significance level (see Eq. 8) as follows:

$$Neqv = N * \left(\frac{1 - r1}{1 + r1}\right) \qquad (8)$$

DATA AVAILABILITY

In this study, the climate model datasets that were utilized can be accessed publicly at the website https://data.ceda.ac.uk/badc. These datasets have been made available by the Center for Environmental Data Analysis. More detailed insights into the outputs of these climate models are available in Eyring et al.[25]. Furthermore, the various observational datasets are accessible at https://psl.noaa.gov/data/gridded/index.html, provided by NOAA Physical Science Laboratory. Additional information regarding the data related to this paper can be obtained from the corresponding author.

CODE AVAILABILITY

The code used to generate the figures and to compute statistics in this study is available upon request from the corresponding author. Interested individuals can obtain access to the code by contacting Ibrahim Hoteit at ibrahim.hoteit@kaust.edu.sa.

REFERENCES

1. Masson-Delmotte, V. et al. Summary for Policymakers. *Climate Change 2021—The Physical Science Basis*. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge University Press, 2021a).
2. Steffen, W. et al. Trajectories of the Earth system in the Anthropocene. *Proc. Natl Acad. Sci. USA* **115**, 8252–8259 (2018).
3. Masson-Delmotte, V. et al. *Climate Change 2021—The Physical Science Basis*. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge University Press, 2021b).
4. Stocker, T. F. et al. *Climate Change 2013—The Physical Science Basis*. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge University Press, 2013).
5. Collins, M. et al. Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multimodel ensembles. *Clim. Dyn.* **36**, 1737–1766 (2011).
6. Murphy, J. M. et al. Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* **430**, 768–772 (2004).
7. Qian, Y. et al. Uncertainty quantification in climate modeling and projection. *Bull. Am. Meteorol. Soc.* **97**, 821–824 (2016).
8. Brunner, L. et al. Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth Syst. Dyn.* **11**, 995–1012 (2020).
9. Pennell, C. & Reichler, T. On the effective number of climate models. *J. Clim.* **24**, 2358–2367 (2011).
10. Knutti, R., Masson, D. & Gettelman, A. Climate model genealogy: generation CMIP5 and how we got there. *Geophys. Res. Lett.* **40**, 1194–1199 (2013).
11. Steinschneider, S., McCrary, R., Mearns, L. O. & Brown, C. The effects of climate model similarity on probabilistic climate projections and the implications for local, risk-based adaptation planning. *Geophys. Res. Lett.* **42**, 5014–5044 (2015).
12. Kotlarski, S. et al. Observational uncertainty and regional climate model evaluation: a pan-European perspective. *Int. J. Climatol.* **39**, 3730–3749 (2019).
13. Collins, M. et al. Chapter 12—Long-term climate change: projections, commitments, and irreversibility. *Climate Change 2013—The Physical Science Basis*.

Contribution of Working Group I To The Fifth Assessment Report of The Inter-governmental Panel on Climate Change (Cambridge University Press, 2013).

14. Gibson, P. B., Waliser, D. E., Lee, H., Tian, B. & Massoud, E. Climate model evaluation in the presence of observational uncertainty: precipitation indices over the contiguous United States. *J. Hydrometeorol.* **20**, 1339–1357 (2019).

15. Jourdain, N. C. et al. The Indo-Australian monsoon and its relationship to ENSO and IOD in reanalysis data and the CMIP3/CMIP5 simulations. *Clim. Dyn.* **41**, 3073–3102 (2013).

16. Soraisam, B., Karumuri, A. & Pai, D. S. Uncertainties in observations and climate projections for the North East India. *Glob. Planet. Change* **160**, 96–108 (2018).

17. Pathak, R., Sahany, S., Mishra, S. K. & Dash, S. K. Precipitation biases in CMIP5 models over the south Asian region. *Sci. Rep.* **9**, 1–13 (2019).

18. Abramowitz, G. et al. ESD reviews: model dependence in multimodel climate ensembles: weighting, sub-selection and out-of-sample testing. *Earth Syst. Dyn.* **10**, 91–105 (2019).

19. Boé, J. Interdependency in multimodel climate projections: component replication and result similarity. *Geophys. Res. Lett.* **45**, 2771–2779 (2018).

20. Leduc, M., Laprise, R., De Elia, R. & Šeparović, L. Is institutional democracy a good proxy for model independence? *J. Clim.* **29**, 8301–8316 (2016).

21. Beck, H. E. et al. Global 3-hourly 0.1 bias-corrected meteorological data including near-real-time updates and forecast ensembles. *Bull. Am. Meteorol. Soc.* **103**, 710–732 (2022).

22. Sokal, R. & Michener, C. A. statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull.* **38**, 1409–1438 (1958).

23. Prajeesh, A. G. et al. The Indian summer monsoon and Indian Ocean Dipole connection in the IITM Earth System Model (IITM-ESM). *Clim. Dyn.* **58**, 1–21 (2022).

24. Doblas Reyes, F. et al. Using EC-Earth for climate prediction research. *ECMWF Newsl.* **154**, 35–40 (2018).

25. Eyring, V. et al. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).

26. Wang, T., Hamann, A., Spittlehouse, D. L. & Aitken, S. N. Development of scale-free climate data for western Canada for use in resource management. *Int. J. Climatol.* **26**, 383–397 (2006).

27. Rauscher, S. A., Coppola, E., Piani, C. & Giorgi, F. Resolution effects on regional climate model simulations of seasonal precipitation over Europe. *Clim. Dyn.* **35**, 685–711 (2010).

28. Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M. & Bladé, I. The effective number of spatial degrees of freedom of a time-varying field. *J. Clim.* **12**, 1990–2009 (1999).

29. Rasmusson, E. M. & Arkin, P. A. A global view of large-scale precipitation variability. *J. Clim.* **6**, 1495–1522 (1993).

30. Jun, M., Knutti, R. & Nychka, D. W. Local eigenvalue analysis of CMIP3 climate model errors. *Tellus A Dyn. Meteorol. Oceanogr.* **60**, 992–1000 (2008).

31. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap* 1st edn, Vol. 1 (*CRC Press*, 1993).

32. Giorgi, F. & Francisco, R. Uncertainties in regional climate change prediction: a regional analysis of ensemble simulations with the HADCM2 coupled AOGCM. *Clim. Dyn.* **16**, 169–182 (2000).

AUTHOR CONTRIBUTIONS

R.P., H.P.D., and I.H. originated the project. All the authors contributed to the ideas and design of the research. R.P. wrote the first draft of the manuscript with inputs from I.H., A.K., and H.P.D. All the authors contributed to subsequent drafts.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41612-023-00473-5.

Correspondence and requests for materials should be addressed to Ibrahim Hoteit.

Reprints and permission information is available at http://www.nature.com/reprints

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