Received: 21 June 2020 | Revised: 1 January 2021 | Accepted: 30 January 2021 | Published on: 19 February 2021

DOI: 10.1002/joc.7041

RESEARCH ARTICLE

International Journal of Climatology RMetS

Communicating potentially large but non-robust changes in multi-model projections of future climate

Giuseppe Zappa[1,2] | **Emanuele Bevacqua**[2] | **Theodore G. Shepherd**[2]

[1]Istituto di Scienze dell'Atmosfera e del Clima, Consiglio Nazionale delle Ricerche (ISAC-CNR), Bologna, Italy

[2]Department of Meteorology, University of Reading, Reading, UK

Correspondence
Theodore G. Shepherd, Department of Meteorology, University of Reading, Reading, UK.
Email: theodore.shepherd@reading.ac.uk

Funding information
ERC advanced grant, Grant/Award Number: 339390; H2020 Environment, Grant/Award Number: 820712

Abstract

The future climate projections in the IPCC reports are visually communicated via maps showing the mean response of climate models to alternative scenarios of socio-economic development. The presence of large changes is highlighted by stippling the maps where the mean climate response (the signal) is large compared to internal variability (the noise) and the response is *robust*, that is, consistent in sign, across the individual models. In addition, hatching is used to mark the regions with a small multi-model mean change. This approach may fail to recognize the risk of large changes in regions where the uncertainty is large and the response is not robust. Here, we present a more informative diagnostic to support risk assessment that is obtained by quantifying the mean forced signal-to-noise ratio of the individual model responses, rather than the signal-to-noise ratio of the mean response. This enables us to identify regions where a large future change compared to year-to-year variability is plausible, regardless of whether the signal is robust across the ensemble. For mean precipitation changes, we find that the majority (58% in surface area) of the unmarked regions and a sizeable portion (19%) of the hatched regions from the AR5 projections hid climate change responses to the RCP8.5 scenario that are on average large compared to the year-to-year variability. Based on the newer CMIP6 ensemble, a considerable potential for large annual-mean precipitation changes, despite the lack of a robust multi-model projection, exists over 22% of the surface land area, particularly in Central America, northern South America (including the Amazon), Central and West Africa (including parts of the Sahel), and the Maritime Continent.

KEYWORDS

climate change, climate model, CMIP5, CMIP6, IPCC, precipitation, signal-to-noise, time of emergence

1 | INTRODUCTION

A central part of the Intergovernmental Panel on Climate Change (IPCC) assessment reports describes the response of the physical climate system to the anthropogenic forcing due to greenhouse gas and aerosol emissions, and land use change (Collins *et al*., 2013). The key source of information to these future projections are the climate

model experiments coordinately run as part of the Coupled Model Inter-Comparison Project (CMIP) by all the main world climate modelling groups (Taylor *et al.*, 2012; Eyring *et al.*, 2016). Multi-model projections are intrinsically uncertain, even for a given forcing scenario, due to the presence of both internal climate variability and differences in how models simulate the response to anthropogenic forcing (Hawkins and Sutton, 2011). To visually aid the communication of future climate change, the model projections in the IPCC reports, such as for surface temperature and precipitation, are presented via spatial maps of the multi-model mean change. Additional information on the size of the mean response to climate change (the signal) relative to the amplitude of the unforced internal variability, considered as noise, is overlaid on the maps in the form of stippling and hatching. In particular, stippling highlights regions where the signal-to-noise ratio is large and the response is robust, that is, consistent in sign, across the models. In contrast, hatching highlights regions where the signal-to-noise ratio of the mean response is small (see, e.g., Figure 1a).

The above approach is effective in communicating the aspects of future climate change in which there is more confidence (stippling), but it leaves some ambiguity on the evaluation of the potential for large changes in the presence of non-robust projections. First of all, as discussed in the IPCC 5th Assessment Report (AR5), hatching can either indicate that different models genuinely agree on a small response or that different models project large changes of opposite sign that compensate in the mean (Collins *et al.*, 2013). Furthermore, a number of regions remain neither hatched nor stippled (Figure 1a). This creates some additional ambiguity on whether these

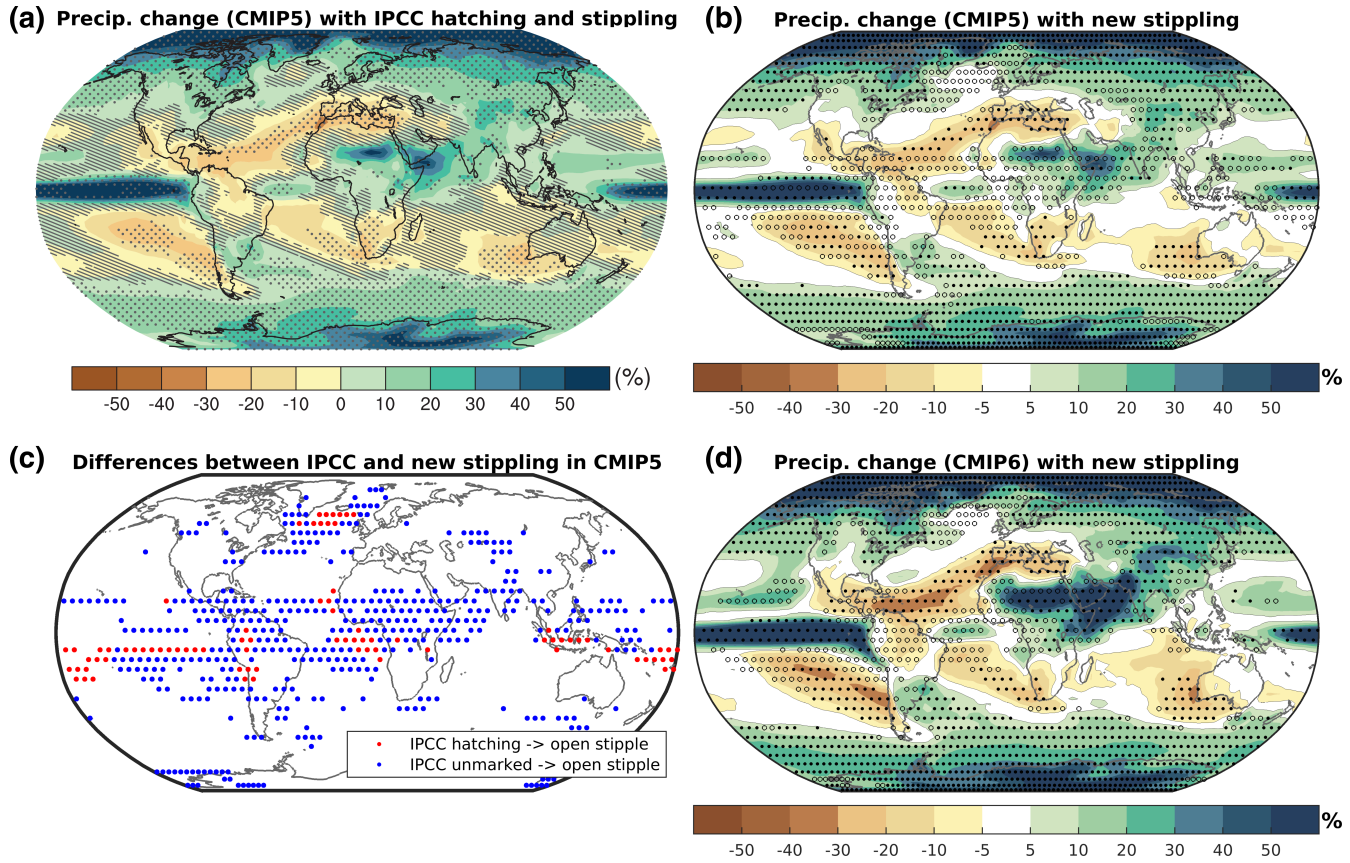

FIGURE 1 Projected change (%) in the annual mean precipitation by 2081–2100 in the RCP8.5 (a–c) and SSP5-8.5 (d) scenarios as portrayed using different approaches and model ensembles. (a) Projection based on the CMIP5 models using the standard hatching and stippling adopted in the IPCC (taken from fig. SPM.8 of the AR5 Summary for Policy Makers [IPCC, 2013]). Stippling reflects changes where at least 90% of models agree on the direction of change (robust response) and the mean change is large, and hatching where the mean change is small, compared to internal variability (see box 12.1 in Collins *et al.*, 2013 for the exact definition of hatching). (b) Projected change in the same set of CMIP5 models analysed in a), but marked using the criteria proposed in this study: Full stippling is a robust response (≥90% of models agree on the direction of change), while the open stippling indicates a plausibly large response compared to year-to-year variability in the presence of a non-robust projection (see text for details). (c) Map of the grid points identified with hatching (red) and unmarked (blue) in the IPCC approach (a) but marked with open stippling in our approach (b). (d) as (b) but based on the CMIP6 models [Colour figure can be viewed at wileyonlinelibrary.com]

unmarked regions reflect consistent changes of intermediate amplitude, or potentially large changes in which models do not agree on the direction of change. These issues become of importance when evaluating climate impacts in risk assessments (Sutton, 2019), since the different range of responses described above would clearly have different implications for the evaluation of the societal risks posed by climate change. The ambiguity is exacerbated for those aspects of climate characterized by large uncertainty and non-robust projections, as for those influenced by the atmospheric circulation, such as precipitation. In these cases, the mean response can show a washed-out signal and risk assessment may better take the form of examining alternative, physically-based, storylines of climate change (Shepherd, 2019; Zappa, 2019). The mean response may be particularly misleading where models have systematic biases in representing the relevant physical processes, hence implying only low confidence in the projected changes.

A number of more sophisticated statistical methods - of varying degrees of complexity - have been proposed to map model agreement and distinguish between cases in which there is agreement on a small change, large changes of opposite sign, and agreement on a significant change (Tebaldi *et al.*, 2011; Knutti and Sedlácek, 2012; Power *et al.*, 2012). The most advanced statistical methods also account for the uncertainty in the magnitude of the change, since a large inter-model spread can still be present in regions where projections are consistent in sign across models (Knutti and Sedlácek, 2012). While acknowledging and comparing these different and more informative approaches, the simpler diagnostic described above was still the one adopted in the AR5 (Collins *et al.*, 2013).

Here, we propose a novel method to communicate and map climate information that is as simple as the standard IPCC approach, but more suitable to provide relevant information for risk assessment. We achieve this by quantifying the mean of the signal-to-noise of the individual model responses, rather than the signal-to-noise of the mean response. The method accounts for the impact of internal variability in inflating the mean signal-to-noise of the projected climate response to anthropogenic forcing. This enables us to identify regions where a large change is plausible, regardless of whether the multi-model projections are robust, by avoiding the compensation arising from discordant individual model responses on the multi-model mean. While the range of existing methods may provide a richer description of model uncertainty, the present one has the advantage of being more intuitive to those users accustomed to the standard IPCC approach, and of being directly useful to identify those regions where the development of physical storylines might be most needed. We show the utility of the method by applying it to examine future projections of precipitation change from the previous (CMIP5) and current (CMIP6) generation of CMIP climate models. In particular, CMIP5 is here instrumental to compare our method against the standard maps presented in the IPCC AR5, while CMIP6 is used to provide an updated assessment of the areas of the globe that may be affected by large—though not robustly projected—changes.

2 | DATA AND METHODS

2.1 | The CMIP5 and CMIP6 data

We analyse monthly mean precipitation output from the 39 CMIP5 models (Taylor *et al.*, 2012) and 34 CMIP6 models (Eyring *et al.*, 2016) listed in Table 1. To be able to compare our results with the IPCC AR5, we use the same set of CMIP5 models and time periods analysed in the AR5. In particular, for each model (m) the climate change response (β_m) is defined as the difference between the mean climate in 2081–2100 in the future scenario and 1986–2005 in the historical experiment. The difference between the two time means includes the forced response to GHGs, aerosols and land use change, as well as the residual unforced internal variability in the 20-year means. Alternative parametric approaches may be more effective at extracting the forced response (e.g., Maraun, 2013), but we here stick to the difference in means to retain consistency with the IPCC. For the future climate, we consider the highest available emissions scenario, namely RCP8.5 in CMIP5, and SSP5-8.5 in CMIP6 (O'Neill *et al.*, 2016). Although the two scenarios have the same nominal radiative forcing by the end of the 21st century, they have slight differences in the time evolution of the forcing and in the balance between the different forcing agents.

Also following the AR5, the magnitude of the unforced internal climate variability is quantified as the variance of non-overlapping Y-year means from the pre-industrial control runs of the models. In order to remove any residual drift, the pre-industrial runs are pre-processed by discarding the first 100 years and by removing long-term trends in the form of a quadratic polynomial. Using this data, we quantify both the variance in the 20-year mean climate ($Y = 20$, $\hat{\sigma}_m^2$), and that in the year-to-year variability ($Y = 1$, σ_m^2). Neglecting the changes in the magnitude of internal climate variability with anthropogenic forcing, the variance in the climate change response from a single realization can then be estimated as $2{\cdot}\hat{\sigma}_m^2$.

One ensemble member is analysed per model, and the models' precipitation output is interpolated via conservative remapping onto a common $2.5 \times 2.5^\circ$ grid prior

TABLE 1 List of CMIP5 and CMIP6 models analysed in this study

Institutions	Country	CMIP5 models	CMIP6 models
Commonwealth Scientific and Industrial Research Organization	Australia	ACCESS1-0	ACCESS-CM2
and Bureau of Meteorology		ACCESS1-3	ACCESS-ESM1-5
Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research	Germany		AWI-CM-1-1-MR
Beijing Climate Centre, China Meteorological Administration	China	bcc-csm1-1	BCC-CSM2-MR
Beijing Normal University	China	BNU-ESM	
Canadian Centre for Climate Modelling and Analysis	Canada	CanESM2	CanESM5
US National Center for Atmospheric Research	USA	CCSM4	CESM2-WACCM
		CESM1-BGC	CESM2
		CESM1-CAM5	
Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici	Italy	CMCC-CMS	CMCC-CM2-SR5
		CMCC-CM	
Centre National de Recherches Meteorologiques and	France	CNRM-CM5	CNRM-CM6-1-HR
Centre Européen de Recherche et Formation Avancées en Calcul Scientifique			CNRM-CM6-1
			CNRM-ESM2-1
Commonwealth Scientific and Industrial Research Organization	Australia	CSIRO-Mk3-6-0	
European Consortium	Europe	EC-Earth	EC-Earth3-Veg
			EC-Earth3
Institute of Atmospheric Physics, Tsinghua University	China	FGOALS-g2	FGOALS-f3-L
			FGOALS-g3
The First Institute of Oceanography	China	FIO-ESM	
Geophysical Fluid Dynamics Laboratory	USA	GFDL-CM3	GFDL-ESM4
		GFDL-ESM2G	
		GFDL-ESM2M	
NASA Goddard Institute for Space Studies	USA	GISS-E2-H (p1, p2, p3)	GISS-E2-1-G
		GISS-E2-R (p1, p2, p3)	
UK Met Office Hadley Centre	UK	HadGEM2-AO	HadGEM3-GC31-LL
		HadGEM2-CC	HadGEM3-GC31-MM
		HadGEM2-ES	UKESM1-0-LL
Russian Institute for Numerical Mathematics	Russia	inmcm4	INM-CM4-8
			INM-CM5-0
Institut Pierre Simon Laplace	France	IPSL-CM5A-LR	IPSL-CM6A-LR
		IPSL-CM5A-MR	
		IPSL-CM5B-LR	
University of Tokyo, National Institute for Environmental Studies,	Japan	MIROC-ESM-CHEM	MIROC-ES2L
and Japan Agency for Marine-Earth Science and Technology		MIROC-ESM	MIROC6
		MIROC5	

TABLE 1 (Continued)

Institutions	Country	CMIP5 models	CMIP6 models
Max Planck Institute for Meteorology	Germany	MPI-ESM-LR	MPI-ESM1-2-HR
		MPI-ESM-MR	MPI-ESM1-2-LR
Meteorological Research Institute	Japan	MRI-CGCM3	MRI-ESM2-0
Nanjing University of Information Science and Technology	China		NESM3
National Institute of Meteorological Sciences	Korea		KACE-1-0-G
Department of Geosciences, University of Arizona	USA		MCM-UA-1-0
Norwegian Climate Centre	Norway	NorESM1-ME	NorESM2-LM
		NorESM1-M	NorESM2-MM
Research Center for Environmental Changes	Taiwan		TaiESM1

Note: Following the IPCC AR5, for the GISS models in CMIP5 we consider the three perturbed physics ensembles (p1, p2, and p3) as different models.

to data analysis. For CMIP5, minor differences with the results presented in the AR5 can arise due to possible differences in the analysed ensemble members and in the adopted interpolation method.

2.2 | The mean forced signal-to-noise ratio

We define the signal-to-noise ratio (γ_m) in the future projection of model m as the ratio between the climate change response (the signal) and the standard deviation in the unforced year-to-year variability (here considered noise), that is, $\gamma_m = |\beta_m/\sigma_m|$. We choose year-to-year variability since this defines a range of variations that societies are already adapted to, and hence constitutes a useful benchmark to quantify large changes for impact purposes (Hawkins and Sutton, 2012; Maraun, 2013). An alternative option, which is used in the stippling of future projections in the AR5, is to evaluate the noise in terms of the internal variability in the mean climate, here given by 20-year means. This enables one to assess the statistical significance of the response, which is informative for climate change detection purposes (Zappa *et al.*, 2015). Hence, the two options are both legitimate and the choice ultimately depends on the question being asked.

Since β_m contains both the response to anthropogenic forcing and unforced fluctuations due to internal variability, taking the multi-model mean of $|\beta_m/\sigma_m|$ would lead to a positive mean signal-to-noise even if all models had no forced climate change response at all. We demonstrate in the Appendix that an unbiased estimator of the mean *forced* signal-to-noise can be obtained if the multi-model mean is computed via a quadratic—rather than arithmetic—average. In this case, the contribution due to internal variability can be accounted for, and the mean forced signal-to-noise of the ensemble (γ_{forced}) can be defined as:

$$\gamma_{\text{forced}} = \sqrt{\left[\frac{1}{M}\sum_{m=1}^{M}\frac{\beta_m^2}{\sigma_m^2}\right] - 2\bar{f}} \qquad (1)$$

where $\bar{f} = \frac{\bar{\sigma}_m^2}{\sigma_m^2}$ is the ratio of the multi-model median variance in the 20-year mean climate to the median variance in the year-to-year variability, and $-2\bar{f}$ is the correction term that accounts for the inflation of the mean signal-to-noise due to internal variability. All models have pre-industrial runs longer than 220 years, so that at least 5 independent 20-year-mean samples per model are available for the estimation of $\hat{\sigma}_m^2$ in $\bar{f}$.

To evaluate the impact of the correction term, we note that the variance ratio ($\bar{f}$) is equal to 0.05 if the year-to-year fluctuations are entirely random, that is, the time-series behave as white noise. This reference value implies a correction on γ_{forced} of about 10% for $\gamma_{\text{forced}} \approx 1$, and of about 25% for $\gamma_{\text{forced}} \approx 0.5$. Deviations from 0.05 are found in regions where atmosphere–ocean coupling creates memory in the system. In particular, for annual-mean precipitation, $\bar{f}$ ranges from 0.02 in the ENSO region to 0.08 in the high latitude oceans (see the Appendix). In the Appendix, we also provide the definition of γ_{forced} for users interested in measuring the signal relative to the noise in the mean climate ($\hat{\sigma}$) rather than in the year-to-year variability (Equation A7).

2.3 | Definition of open and full stippling

We use stippling to mark the future climate projections as follows:

- *full stippling* marks regions where at least 90% of the models agree on the direction of change, that is, a robust response.
- *open stippling*, that is, open circles, marks regions where less than 90% of models agree on the direction of change, but the mean forced signal-to-noise relative to the year-to-year variability is greater than unity ($\gamma_{\text{forced}} \geq 1$). This is interpreted as a plausibly large response in the presence of a non-robust projection.
- *unmarked* marks regions where less than 90% of models agree on the direction of change and $\gamma_{\text{forced}} < 1$. This is interpreted as an uncertain response in which the projected changes are small compared to inter-annual variability.

The rationale behind this choice is the following. Robust projections (full stippling) are of interest on their own, since they define regions where there is more confidence in the sign of the projected climate change. For stippling, the IPCC AR5 further required the mean response to be large compared to the noise on 20-year means, that is, $\bar{\beta}/(\bar{\hat{\sigma}}_m \cdot \sqrt{2}) \geq 2$, which can be interpreted as the mean response being statistically significant at approximately the 5% level. This condition is here dropped with little loss of information, since a 90% agreement on direction of change is itself a strong indication of statistical significance (Power *et al.*, 2012). In regions where projections are not robust, it is instead the magnitude of the changes that becomes more important. Therefore, our attention shifts to quantifying how the forced changes compare with the amplitude of year-to-year variability, which is a useful quantity to assess the potential for impacts. If such signal-to-noise is on average greater than or equal to unity ($\gamma_{\text{forced}} \geq 1$), it then becomes important to communicate the risk of a large change, which is what the open stippling portrays. Note that open stippling can be found both if the signal-to-noise is about unity in all models, or if it is much larger than unity in some models but close to zero in others, so examining the full spread in the model responses remains important in follow-up risk assessments. Finally, unmarked regions show where the risk of a large response compared to inter-annual variability is small. Overall, these conditions define a set of three mutually exclusive and exhaustive categories to classify and communicate the future changes projected from multi-model ensembles.

3 | RESULTS

3.1 | Revisiting the IPCC AR5

We first analyse the projected changes in annual-mean precipitation in the CMIP5 models for the RCP8.5 scenario (Figure 1b). Here, the aim is not to discuss the mean precipitation changes themselves, which have been extensively discussed in the literature, but rather to evaluate the information provided by our diagnostic against that employed by the IPCC. The patterns and magnitudes of precipitation change are indeed nearly identical to those presented in the AR5 (Figure 1a) since the same set of CMIP5 models and time periods are considered.

Comparing Figure 1a,b reveals that the two diagnostic approaches are overall consistent. First of all, the AR5 stippled regions largely coincide with the full stippled regions in our approach. This is to be expected, since robust changes are defined in the same way, and, as previously mentioned, the additional criterion on the magnitude of the mean response included in the AR5 is largely redundant (Power *et al.*, 2012). Furthermore, the majority of hatched regions in the AR5 (small signal in the mean) coincide with areas that are unmarked in our approach (small mean signal-to-noise).

The additional benefit of our approach is that it enables a direct assessment of the regions with a large forced signal-to-noise despite the lack of a robust projection (open stippling). We find that these regions tend to predominantly correspond to the unmarked regions from the AR5 (blue dots in Figure 1c), but the correspondence is not one to one. First of all, we note that there are some unmarked responses in the AR5 map, such as the wetting of the northern contiguous United States and western Canada, that remain unmarked also within our diagnostic. Furthermore, open stippling is also found in some of the hatched regions from the AR5 map (red dots in Figure 1c). This is the case in some oceanic areas, for example, the high-latitude North Atlantic and the Gulf of Guinea, as well as some land areas, such as to the east of the northern Andes in South America, and parts of the Maritime Continent. The implication is that the presence of a small response in the mean (hatching) hides large individual model responses that tend to average out, which is consistent with what was previously found using other statistical approaches (Tebaldi *et al.*, 2011; Knutti and Sedlácek, 2012).

In summary, the majority (58% in global surface area) of the unmarked regions and a sizeable portion (19%) of the hatched regions from the AR5 hid non-robust precipitation responses to climate change that are on average large compared to the year-to-year variability. These

results confirm the presence of some ambiguity in interpreting the magnitude of the responses in the hatched and in the unmarked regions using the standard IPCC scheme. Such ambiguities can be avoided by explicitly considering the mean forced signal-to-noise of the multi-model ensemble as presented in this study.

3.2 | The CMIP6 ensemble

Having found consistent results with those presented for the CMIP5 models in the AR5, we move on to examine the CMIP6 ensemble, in order to characterize the risk of large - though not robustly projected - precipitation changes in this latest set of climate models. As a general note, comparing Figure 1b,d shows a very similar large-scale pattern in the end-of-century annual-mean precipitation change in the CMIP5 and CMIP6 models (the pattern correlation weighted by unit area is .94). This is consistent with the fact that the large-scale precipitation changes are largely controlled by thermodynamic mechanisms, such as wet-get-wetter and land-sea contrast, that are robust to model formulation (Held and Soden, 2006; Byrne and O'Gorman, 2015). The magnitude of the precipitation changes, such as the increase in high-latitude precipitation, is however slightly higher in CMIP6 (19% more in the global-mean) which is consistent with the higher climate sensitivity—and hence warming—of CMIP6 compared to CMIP5 models (Forster *et al.*, 2019; Zelinka *et al.*, 2020).

In spite of the greater climate sensitivities within the CMIP6 ensemble, there is only a slight increase in the global surface area satisfying $\gamma_{\text{forced}} \geq 1$ from CMIP5 to CMIP6. However, the CMIP6 models tend to show a higher inter-model agreement on the direction of change. In particular, the area covered by either full or open stippling remains approximately the same, but while the fraction of Earth's surface showing full stippling increases from 36% in CMIP5 to 45% in CMIP6, the open stippled area decreases from 27 to 17%. This change is most

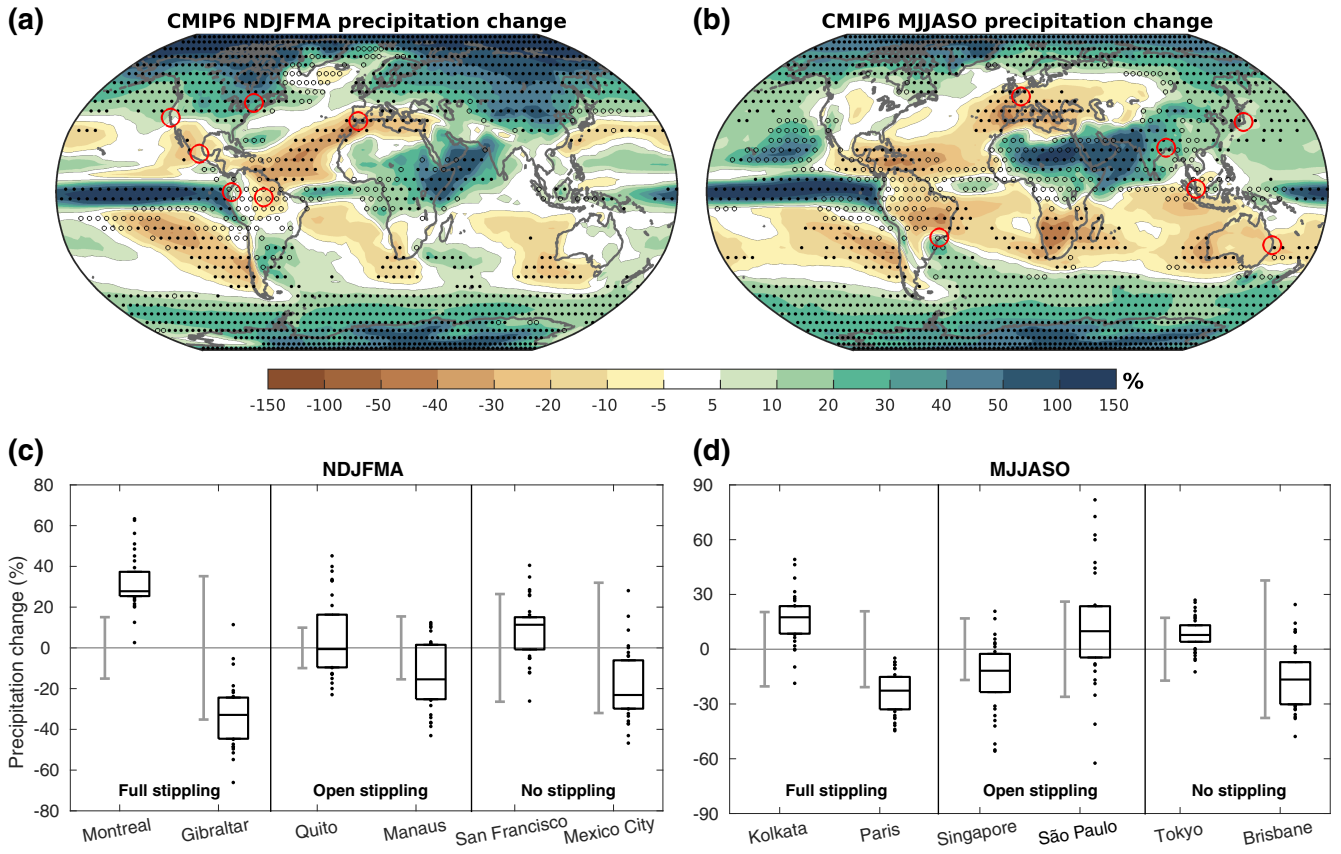

FIGURE 2 (a,b) As in Figure 1d but for the mean precipitation response in the boreal cold season (November–April) and boreal warm season (May–October). (c,d) Distribution of individual model responses at the selected locations around the globe indicated by the red circles in panels (a) and (b). The horizontal line is the multi-model median, the box extends to the inter-quartile range, and the dots show the remaining individual models. The grey vertical bar shows the multi-model median inter-annual variability range ($\pm\bar{\sigma}$). For each season, the panels include two locations marked by full stippling, two marked by open stippling and two that are unmarked. Note that in panels (a) and (b) there is a slight change in palette compared to Figure 1d [Colour figure can be viewed at wileyonlinelibrary.com]

notable in, but not limited to, the band of increasing precipitation extending from eastern Africa to northern India, passing through the Arabian Gulf, which is open stippled in CMIP5 but full stippled, and with a much larger precipitation increase, in CMIP6. The causes for this change will have to be investigated. However, the increasing robustness on a wetting of East Africa should be regarded with caution, given that a drying trend—opposite to the forced response—is so far emerging in the observational record (Rowell *et al*., 2015; Lyon and Vigaud, 2017).

To describe the potential for large future changes in more detail, we examine the seasonal response and separately present the boreal cold-season (November–April, NDJFMA) and warm-season (May–October, MJJASO) in Figure 2a,b. In particular, by focusing on land, we find that:

- Non-robust but potentially large changes (open stippling) are typically found in the tropics. In particular, they are found year-round in the Amazon region, in large parts of central and northern Africa (including the Sahel), of Central America, and of the Maritime Continent in MJJASO, as well as in the remaining tropical part of South America in NDJFMA.
- A robust wetting (full stippling) is primarily found in the mid-to-high latitudes year-round, in central Africa, eastern Africa, and south-east South America in NDJFMA, and in large parts of South-east Asia in MJJASO.
- A robust drying (full stippling) is mostly confined to Mediterranean-like climates in the cold season, particularly the Mediterranean proper and southern Chile in NDJFMA, and western Europe, central Chile, and south-west Australia in MJJASO.

To contrast the different models' behaviour identified by the diagnostic, Figure 2c,d compares the distribution of the model responses against the magnitude of inter-annual variability at different locations marked by full stippling, open stippling and no stippling. The selected locations within the full stippled regions, for example, Montreal (Canada) and Gibraltar (Iberia) in NDJFMA, in addition to showing robust changes, are also characterized by typically large responses compared to inter-annual variability. While this is not an imposed criterion, we find that the condition $\gamma_{\text{forced}} \geq 1$ is indeed satisfied for 85% of the full stippled area in NDJFMA and 81% in MJJASO. At the other end of the spectrum, locations that are unmarked, such as San Francisco and Mexico City in NDJFMA, show projections that are not robust and on average within the typical range found in the year-to-year variability. Note that this does not necessarily imply the percentage changes are small—with values of the order of 10 to 20% being common—so that absence of stippling should not be interpreted as absence of impacts. Finally, the locations marked by open stippling show a range of behaviours in terms of inter-model spread, but they all support the potential for large changes compared to inter-annual variability. For example, Quito (Equador) in NDJFMA shows no change in the mean, but both a large wetting or large drying are possible in the individual models. Manaus (Brazil) in NDJFMA and Singapore in MJJASO show a non-robust tendency towards drying, but with the potential for a very large drying in some models. São Paulo (Brazil) in MJJASO shows a mean tendency towards moderate wetting, but both a large wetting and a large drying compared to inter-annual variations are possible based on the full ensemble spread.

Having described end-of-century projections, we finally explore when such potentially large responses might first appear in the course of the 21st century. For this purpose, the signal-to-noise ratio has been recomputed for a range of 5-year-apart future time windows starting from 2011–2030. The first time-window satisfying $\gamma_{\text{forced}} \geq 1$ is shown in Figure 3a,c, and it provides an alternative way to estimate the mean *time-of-emergence* of a large climate change response (Hawkins and Sutton, 2012; Maraun, 2013; King *et al*., 2015; Zappa *et al*., 2015). In the open-stippled areas, with the notable exception of central Africa in MJJASO, large precipitation responses to climate change tend to be only found in the second-half of the 21st century. For example, among the regions mentioned before, $\gamma_{\text{forced}} \geq 1$ is first met in 2056–2075 for NDJFMA precipitation in Manaus, and in 2071–2090 for MJJASO in Singapore. However, it would be dangerous to conclude from these results that there is ample time to adapt to these uncertain changes. Societal impacts due to climate change can take place well before the changes in the mean climate become as large as year-to-year variability, especially since they are first realized during extreme weather events resulting from the combination of forced changes and internal variability (Seneviratne *et al*., 2012). Allowing for an additional contribution from internal variability, Figure 3b,d shows that the emergence of a large response associated with a lower threshold on the mean forced signal-to-noise ($\gamma_{\text{forced}} \geq 0.5$) is already satisfied over vast areas of the globe by 2050.

4 | SUMMARY AND CONCLUSIONS

We have here revisited the standard approach from the IPCC to present future projections from multi-model

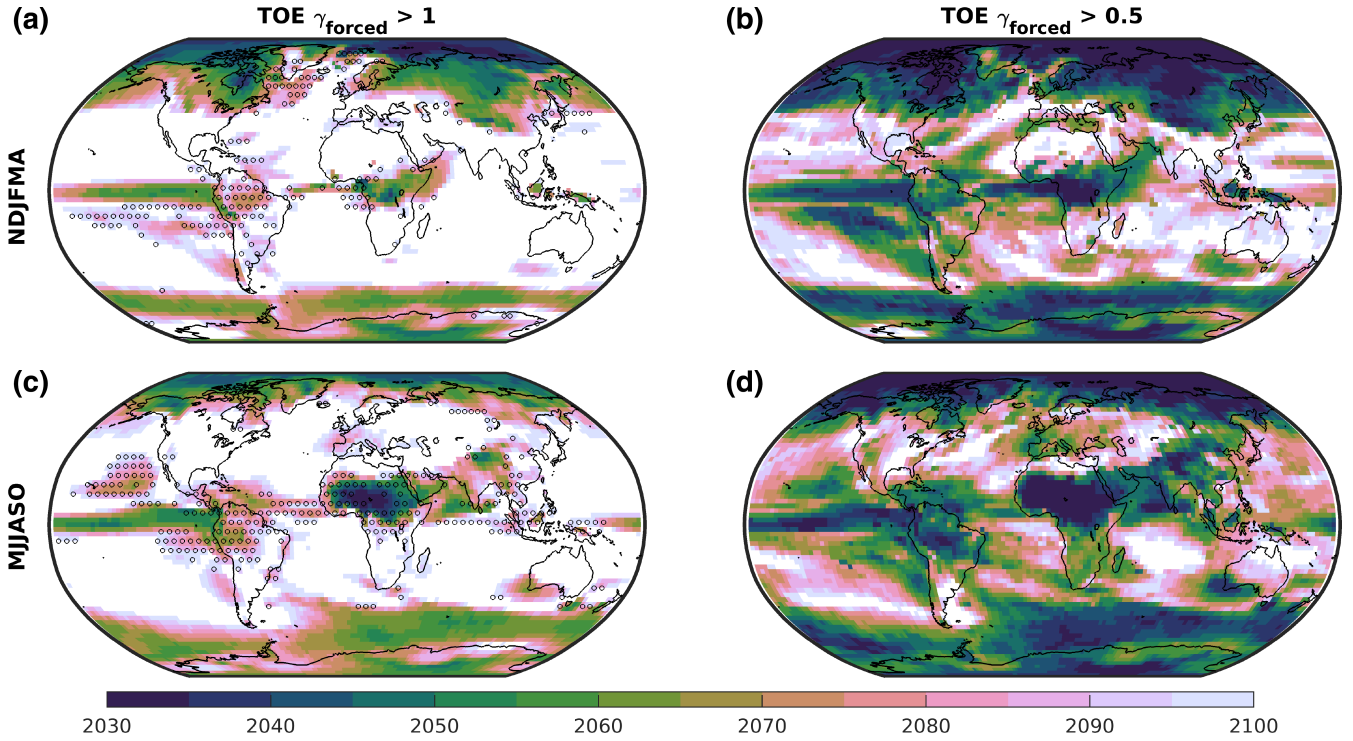

FIGURE 3 First time the forced mean signal-to-noise of the precipitation response in the CMIP6 models exceeds unity ($\gamma_{forced} \geq 1$) in (a) the boreal cold season (NDJFMA) and (c) the boreal warm season (MJJASO). b and d are the same but for $\gamma_{forced} \geq 0.5$. The colour shading displays the last year of the future 20-year time window. The open stippling in panels (a) and (c) is the same as in panels (a) and (b) of Figure 2 [Colour figure can be viewed at wileyonlinelibrary.com]

ensembles with the aim of highlighting the regions with potentially large—though not robustly projected—changes. In the standard approach from the IPCC, the focus is placed on contrasting regions with a robust response to anthropogenic forcing (stippling) against those that feature a small or uncertain response (hatching). The rationale for this may be that adaptation planning requires to have confidence in the response, and distinguishing between a small and an uncertain response may be thought of as a matter of secondary importance. However, from the point of view of risk assessment, recognizing the potential for large future changes is itself of importance regardless of the confidence in the response itself. Such information is directly available in our approach.

In its essence, the new diagnostic mainly differs from the IPCC one by evaluating the mean forced signal-to-noise ratio of the individual model responses, rather than the signal-to-noise ratio of the mean response. Doing so avoids the compensation in the mean response that occurs in regions where models disagree on the direction of change, and hence provides a *real* quantification of the mean forced signal-to-noise from a multi-model ensemble. This enables one to discriminate between areas where the response is robust and those where it is potentially large though non-robust, that is, where climate-related uncertainties need to be taken into account with particular caution. Together with the unmarked areas (small signal-to-noise compared to variability), the approach provides three mutually exclusive and exhaustive categories to unambiguously classify the projected climate change in multi-model ensembles. As a caveat, the method (as does the IPCC one) assumes the distinct CMIP model responses to be independent and equally likely, which is known not to be the case (Abramowitz *et al.*, 2019). Model weighting could be considered to down-weight either similar model versions within the same 'family' (Knutti *et al.*, 2013) and/or the models that are most poorly-performing under present climate conditions (e.g., Gillett, 2015). However, while model dependence certainly affects the assessment of robustness in the response, it has a milder impact on mean statistics, such as the mean forced signal-to-noise.

The approach is applicable to any field and multi-model dataset provided that enough data, either in the form of ensemble members or long runs, is available to estimate the amplitude of internal climate variability. Here, by examining projections of mean precipitation change in the CMIP5 and CMIP6 global climate models, we conclude that:

- For a high emissions scenario, the majority (58% in surface area) of the unmarked regions and a sizeable portion (19%) of the hatched regions from the AR5 map for mean precipitation (based on CMIP5) hid non-robust climate change responses that are on average large compared to the year-to-year variability.
- The agreement on the direction of future mean precipitation changes has increased in the newer CMIP6 ensemble. Whether this truly represents more reliable projections remains to be assessed. Nevertheless, large but non-robust precipitation changes that could emerge before the end of the 21st century still characterize 22% of the global surface land area in CMIP6—particularly including Central America, northern South America (including the Amazon), central and west Africa (including parts of the Sahel) and the Maritime Continent.
- Forced precipitation changes as large as half the standard deviation in the year-to-year variability can potentially emerge by the mid-21st century in the majority of the land surface area.

In these identified regions with large but non-robust responses, it would be useful to develop alternative plausible regional storylines of climate change that illustrate the physical factors that can drive such large—but discordant—responses, together with the impacts that would unfold from such changes (Shepherd, 2019; Zappa, 2019). Using regional climate model output would be particularly valuable given their potential to better inform changes at the local scale, provided that the biases in the large-scale atmospheric circulation and regional processes are accounted for (Maraun *et al*., 2017). Providing such information would help to support any risk assessment.

ACKNOWLEDGEMENTS

The authors were supported by the European Research Council Grant 'Understanding the atmospheric circulation response to climate change' (Grant 339390) and the European H2020 grant 'Remote Climate Effects and their Impact on European sustainability, Policy and Trade' (Grant 820712). The authors are grateful to two reviewers for their insightful comments that have helped to improve the article, and to Susanna Corti and Erich Fischer for useful discussions. The authors acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP5 and CMIP6, and the authors thank the climate modelling groups listed in Table 1 for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP and the ESGF.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

ORCID

Giuseppe Zappa https://orcid.org/0000-0003-4306-7451

REFERENCES

Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., Lorenz, R., Pincus, R. and Schmidt, G.A. (2019) ESD Reviews: Model dependence in multi-model climate ensembles: weighting sub-selection and out-of-sample testing. *Earth System Dynamics*, 10, 91–105. https://doi.org/10.5194/esd-10-91-2019.

Byrne, M.P. and O'Gorman, P.A. (2015) The response of precipitation minus evapotranspiration to climate warming: why the "wet-get-wetter dry-get-drier" scaling does not hold over land. *Journal of Climate*, 28, 8078–8092. https://doi.org/10.1175/jcli-d-15-0369.1.

Collins, M., Reto, K., Arblaster, J., Dufrense, J.L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W.J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A.J. and Wehner, M. (2013) Long-term climate change: projections, commitments and irreversibility. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V. and Midgley, P.M. (Eds.) *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, UK and New York, NY: Cambridge University Press.

Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J. and Taylor, K.E. (2016) Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9, 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016.

Forster, P.M., Maycock, A.C., McKenna, C.M. and Smith, C.J. (2019) Latest climate models confirm need for urgent mitigation. *Nature Climate Change*, 10, 7–10. https://doi.org/10.1038/s41558-019-0660-0.

Gillett, N.P. (2015) Weighting climate model projections using observational constraints. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373, 20140425. https://doi.org/10.1098/rsta.2014.0425.

Hawkins, E. and Sutton, R. (2011) The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dynamics*, 370(1), 407–418. https://doi.org/10.1007/s00382-010-0810-6.

Hawkins, E. and Sutton, R. (2012) Time of emergence of climate signals. *Geophysical Research Letters*, 39, L01702. https://doi.org/10.1029/2011gl050087.

Held, I.M. and Soden, B.J. (2006) Robust responses of the hydrological cycle to global warming. *Journal of Climate*, 19, 5686–5699. https://doi.org/10.1175/jcli3990.1.

IPCC. (2013) Climate Change 2013: The Physical Science Basis. Summary for Policymakers. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J.,

Nauels, A., Xia, Y., Bex, V. and Midgley, P.M. (Eds.) *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, UK and New York, NY: Cambridge University Press. https://doi.org/10.1260/095830507781076194.

King, A.D., Donat, M.G., Fischer, E.M., Hawkins, E., Alexander, L. V., Karoly, D.J., Dittus, A.J., Lewis, S.C. and Perkins, S.E. (2015) The timing of anthropogenic emergence in simulated climate extremes. *Environmental Research Letters*, 10, 094015. https://doi.org/10.1088/1748-9326/10/9/094015.

Knutti, R., Masson, D. and Gettelman, A. (2013) Climate model genealogy: Generation CMIP5 and how we got there. *Geophysical Research Letters*, 400(6), 1194–1199. https://doi.org/10.1002/grl.50256.

Knutti, R. and Sedlácek, J. (2012) Robustness and uncertainties in the new CMIP5 climate model projections. *Nature Climate Change*, 3, 369–373. https://doi.org/10.1038/nclimate1716.

Lyon, B. and Vigaud, N. (2017) Unraveling East Africa's climate paradox. In: *Climate Extremes*. Hoboken, NJ: John Wiley & Sons Inc, pp. 265–281. https://doi.org/10.1002/9781119068020.ch16.

Maraun, D. (2013) When will trends in European mean and heavy daily precipitation emerge? *Environmental Research Letters*, 8, 014004.

Maraun, D., Shepherd, T.G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J.M., Hagemann, S., Richter, I., Soares, P.M.M., Hall, A. and Mearns, L.O. (2017) Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, 7, 764–773. https://doi.org/10.1038/nclimate3418.

O'Neill, B.C., Tebaldi, C., van Vuuren, D.P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G.A., Moss, R., Riahi, K. and Sanderson, B.M. (2016) The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, 9, 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016.

Power, S.B., Delage, F., Colman, R. and Moise, A. (2012) Consensus on twenty-first-century rainfall projections in climate models more widespread than previously thought. *Journal of Climate*, 25, 3792–3809. https://doi.org/10.1175/jcli-d-11-00354.1.

Rowell, D.P., Booth, B.B.B., Nicholson, S.E. and Good, P. (2015) Reconciling past and future rainfall trends over East Africa. *Journal of Climate*, 28, 9768–9788. https://doi.org/10.1175/jcli-d-15-0140.1.

Seneviratne, S.I., Nicholls, N., Easterling, D., Goodess, C.M., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C. and Zhang, X. (2012) Changes in climate extremes and their impacts on the natural physical environment. In: Field, C.B., Barros, V., Stocker, T.F., Qin, D., Dokken, D.J., Ebi, K.L., Mastrandrea, M.D., Mach, K.J., Plattner, G.-K., Allen, S.K., Tignor, M. and Midgley, P.M. (Eds.) *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York, NY: Cambridge University Press, pp. 109–230.

Shepherd, T.G. (2019) Storyline approach to the construction of regional climate change information. *Proceedings of the Royal Society A: Mathematical Physical and Engineering Sciences*, 475, 20190013. https://doi.org/10.1098/rspa.2019.0013.

Sutton, R.T. (2019) Climate science needs to take risk assessment much more seriously. *Bulletin of the American Meteorological Society*, 100, 1637–1642. https://doi.org/10.1175/bams-d-18-0280.1.

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

Tebaldi, C., Arblaster, J.M. and Knutti, R. (2011) Mapping model agreement on future climate projections. *Geophysical Research Letters*, 38, L23701. https://doi.org/10.1029/2011gl049863.

Thompson, D.W.J., Barnes, E.A., Deser, C., Foust, W.E. and Phillips, A.S. (2015) Quantifying the role of internal climate variability in future climate trends. *Journal of Climate*, 28, 6443–6456. https://doi.org/10.1175/jcli-d-14-00830.1.

Zappa, G. (2019) Regional climate impacts of future changes in the mid–latitude atmospheric circulation: a storyline view. *Current Climate Change Reports*, 5, 358–371. https://doi.org/10.1007/s40641-019-00146-7.

Zappa, G., Hoskins, B.J. and Shepherd, T.G. (2015) Improving climate change detection through optimal seasonal averaging: the case of the North Atlantic Jet and European precipitation. *Journal of Climate*, 28, 6381–6397. https://doi.org/10.1175/jcli-d-14-00823.1.

Zelinka, M.D., Myers, T.A., McCoy, D.T., Po-Chedley, S., Caldwell, P.M., Ceppi, P., Klein, S.A. and Taylor, K.E. (2020) Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters*, 47, e2019GL085782. https://doi.org/10.1029/2019gl085782.

How to cite this article: Zappa G, Bevacqua E, Shepherd TG. Communicating potentially large but non-robust changes in multi-model projections of future climate. *Int J Climatol*. 2021;41: 3657–3669. https://doi.org/10.1002/joc.7041

APPENDIX

Derivation of the mean forced signal to noise

The signal-to-noise (γ_m) of the climate change response in model m is here defined as the ratio between the response itself (β_m) and the standard deviation in the unforced year-to-year variability (σ_m). Since the climate change response is evaluated as the difference in means between a future and a present Y-year time slice from a single realization, β_m can itself be regarded as a normally distributed random variable with mean C_m and variance $2\hat{\sigma}_m^2$, i.e., $\beta_m \sim N(C_m, 2\hat{\sigma}_m^2)$, where C_m is the forced response and $\hat{\sigma}_m^2$ is the variance in the Y-year mean climate. Therefore, the multi-model mean of the squares of the signal-to-noise ratios from an ensemble of M models is distributed as follows:

$$\gamma^2 = \frac{1}{M}\sum_{m=1}^{M}\frac{\beta_m^2}{\sigma_m^2} \sim \frac{1}{M}\sum_m \frac{N(C_m, 2\hat{\sigma}_m^2)^2}{\sigma_m^2} = \tag{A1}$$

$$= \frac{1}{M}\sum_m \frac{C_m^2}{\sigma_m^2} + \frac{2}{M}\sum_m \frac{C_m \cdot N(0, 2\hat{\sigma}_m^2)}{\sigma_m^2} + \frac{1}{M}\sum_m \frac{N(0, 2\hat{\sigma}_m^2)^2}{\sigma_m^2}. \tag{A2}$$

We define the parameter $f_m = \hat{\sigma}_m^2/\sigma_m^2$ as the ratio of the Y-year mean variance to the inter-annual variance. The values of f_m depend on the auto-correlation of the data, so that $f_m \approx 1/Y$ if the data is serially independent (white noise), but it will differ if there is some inter-annual memory in the system. As shown in Figure A1, this is the case in the presence of slow multi-decadal variability ($f_m > 1/Y$), or of faster oscillating phenomena such as ENSO ($f_m < 1/Y$). Therefore:

$$\gamma^2 \sim \frac{1}{M}\sum_m \frac{C_m^2}{\sigma_m^2} + \frac{2}{M}\sum_m f_m \frac{C_m \cdot N(0, 2\hat{\sigma}_m^2)}{\hat{\sigma}_m^2} + \frac{1}{M}\sum_m f_m \frac{N(0, 2\hat{\sigma}_m^2)^2}{\hat{\sigma}_m^2}. \tag{A3}$$

We then take f_m out of the sums in the second and third terms by approximating it with the ratio of the multi-model-median variances, $\bar{f} = \bar{\hat{\sigma}}_m^2/\bar{\sigma}_m^2$. This is justified under the assumption that, for any given grid point, the magnitude of the inter-annual variance is unrelated to the structure of the auto-correlation function across the CMIP models. The authors are not aware that such a behaviour is commonly found, and the proposed approximation is still an improvement compared to the common approach of simply treating inter-annual variability as white noise (Thompson *et al.*, 2015). It thus follows that γ^2 is distributed as

$$\gamma^2 \sim \frac{1}{M}\sum_m \frac{C_m^2}{\sigma_m^2} + \frac{2\bar{f}}{M}\sum_m \frac{C_m \cdot N(0, 2\hat{\sigma}_m^2)}{\hat{\sigma}_m^2} + \frac{2\bar{f}}{M}\chi^2(M), \tag{A4}$$

where χ^2 is the chi-squared distribution with M degrees of freedom. The expected value for the mean of γ^2 is:

$$E(\gamma^2) = \left[\frac{1}{M}\sum_m \frac{C_m^2}{\sigma_m^2}\right] + 2\bar{f}. \tag{A5}$$

By rearranging the terms in Equation (A5), we obtain that the multi-model mean forced signal-to-noise of the ensemble can be estimated as

$$\gamma_{forced} = \sqrt{\frac{1}{M}\sum_m \frac{C_m^2}{\sigma_m^2}} = \sqrt{E(\gamma)^2 - 2\bar{f}} = \sqrt{\left[\frac{1}{M}\sum_m \frac{E(\beta_m^2)}{\sigma_m^2}\right] - 2\bar{f}}$$
$$\approx \sqrt{\left[\frac{1}{M}\sum_m \frac{\beta_m^2}{\sigma_m^2}\right] - 2\bar{f}}. \tag{A6}$$

In the last step, we have evaluated the expected value of β_m^2 assuming that only a single realization of the climate change response is available for each model. Otherwise, if multiple ensemble members are available, the estimation of γ_{forced} can be made more accurate by taking an ensemble average, that is, $E(\beta_m^2) \approx 1/R_m \sum_{r=1}^{R_m} \beta_{mr}^2$, where

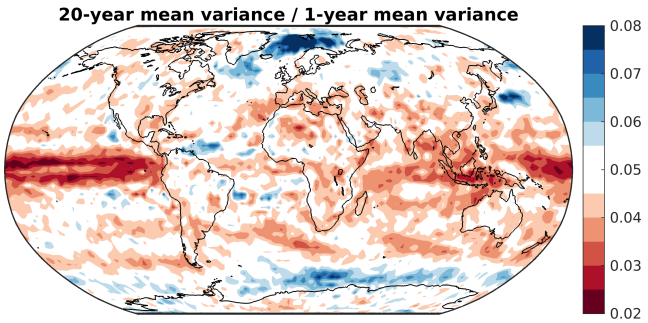

FIGURE A1 Value of $\bar{f}$ for annual-mean precipitation in the CMIP5 models, evaluated as the ratio of the multi-model-median 20-year-mean variance to the multi-model-median of the inter-annual variance. White shading, approx. 0.05, corresponds to a variance ratio consistent with white noise. Blue shading (ratio >0.05) in the high-latitude oceans indicates that slow atmosphere–ocean processes—such as the Atlantic Multi-decadal Variability—generate more decadal variability in the mean precipitation than expected based on year-to-year fluctuations. Red shading (ratio <0.05) in the tropical regions indicates that fast oscillating processes, such ENSO, generate lower multi-decadal variability than expected based on year-to-year variability [Colour figure can be viewed at wileyonlinelibrary.com]

β_{mr} is the climate change response in ensemble member r, and R_m is the number of available ensemble members for model m. However, a substantial benefit is only expected if all models provide multiple members of both historical and future climate, which has not been the case in CMIP. Due to sampling uncertainty, the argument of the square root can be negative where the forced response is close to zero in all models, in which case γ_{forced} is approximated to zero.

Equation (A6) provides the mean forced signal-to-noise relative to the amplitude of the year-to-year variability, which is most useful to assess the magnitude of climate change for risk assessment. If the focus is instead on climate change detection, the response might better be evaluated relative to the noise in the Y-year mean climate, and the expression for γ_{forced} becomes:

$$\hat{\gamma}_{\text{forced}} = \sqrt{\frac{1}{M}\sum_m \frac{C_m^2}{\hat{\sigma}_m^2}} \approx \sqrt{\left[\frac{1}{M}\sum_m \frac{\beta_m^2}{\hat{\sigma}_m^2}\right] - 2}. \qquad \text{(A7)}$$