## Title
Use-Inspired, Process-Oriented GCM Selection: Prioritizing Models for Regional Dynamical Downscaling



## Journal
Bulletin of the American Meteorological Society, 104(9)

## ISSN
0003-0007

## Authors
Goldenson, Naomi
Leung, L Ruby
Mearns, Linda O
et al.

## Publication Date
2023-09-01

## DOI
10.1175/bams-d-23-0100.1

BAMS Article

# Use-Inspired, Process-Oriented GCM Selection

## Prioritizing Models for Regional Dynamical Downscaling

Naomi Goldenson, L. Ruby Leung, Linda O. Mearns, David W. Pierce, Kevin A. Reed, Isla R. Simpson, Paul Ullrich, Will Krantz, Alex Hall, Andrew Jones, and Stefan Rahimi

**KEYWORDS:**
Downscaling;
Climate models;
Model comparison;
Regional models;
Decision support

**ABSTRACT:** Dynamical downscaling is a crucial process for providing regional climate information for broad uses, using coarser-resolution global models to drive higher-resolution regional climate simulations. The pool of global climate models (GCMs) providing the fields needed for dynamical downscaling has increased from the previous generations of the Coupled Model Intercomparison Project (CMIP). However, with limited computational resources, the need for prioritizing the GCMs for subsequent downscaling studies remains. GCM selection for dynamical downscaling should focus on evaluating processes relevant for providing boundary conditions to the regional models and be inspired by regional uses such as the response of extremes to changes in the boundary conditions. This leads to the need for metrics representing processes of relevance to diverse stakeholders and subregions of a domain. Procedures to account for metric redundancy and the statistical distinguishability of GCM rankings are required. Further, procedures for selecting realizations from ensembles of top-performing GCM simulations can be used to span the range of climate change signals in multiple ways. As a result, distinct weighting of metrics and prioritization of particular realizations may depend on user needs. We provide high-level guidelines for such region-specific evaluations and address how CMIP7 might enable dynamical downscaling of a representative sample of high-quality models across representative shared socioeconomic pathways (SSPs).

https://doi.org/10.1175/BAMS-D-23-0100.1
Corresponding author: Naomi Goldenson, naomi@modelworldconsulting.com
Supplemental material: https://doi.org/10.1175/BAMS-D-23-0100.2
In final form 2 May 2023

**AFFILIATIONS: Goldenson**—Model World Consulting; **Leung**—Pacific Northwest National Laboratory, Richland, Washington; **Mearns and Simpson**—National Center for Atmospheric Research, Boulder, Colorado; **Pierce**—Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California; **Reed**—School of Marine and Atmospheric Sciences, Stony Brook University, State University of New York, Stony Brook, New York; **Ullrich**—University of California, Davis, Davis, California; **Hall, Krantz, and Rahimi**—University of California, Los Angeles, Los Angeles, California; **Jones**—Lawrence Berkeley National Laboratory, and University of California, Berkeley, Berkeley, California

Despite a trend toward high resolution in global climate models (GCMs), regional climate models (RCMs) remain indispensable tools for deriving insights into the climate system (Giorgi 2019; Rummukainen 2010). They further provide a mechanism to deliver high-spatial-and-temporal-resolution data on a near-complete set of atmospheric variables to scientists, decision-makers, and other stakeholders and end-users (Gutowski et al. 2020; Kotamarthi et al. 2021).

The regional-downscaling focus presented here is biased toward the midlatitudes, partly related to previous and ongoing research of the authors on North American regional climate and motivated by the genuine improvements in the representation of midlatitude dynamics in some CMIP6 GCMs (Cannon 2020; Harvey et al. 2020; Priestley et al. 2020; Simpson et al. 2020). In the regional-downscaling community, there are a range of approaches used to construct boundary conditions informed by GCMs: the so-called "pseudo-global warming" approach (Schär et al. 1996) involves simulating historical weather under warmer conditions and is widely applied (e.g., Liu et al. 2017; Xue and Ullrich 2021). Direct dynamical downscaling, practiced notably by several major GCM–RCM programs over North America (Bukovsky and Mearns 2020; Mearns et al. 2012), allows the changes in the large-scale variability (that a GCM might project) to be included in the boundary conditions for the RCM. Thus, improvements to the representation of midlatitude dynamics in GCMs might motivate more direct dynamical downscaling for well-performing GCMs.

There is a long history of debate about whether to choose a subset of GCMs based on model quality for any given regional assessment (e.g., Tebaldi and Knutti 2007; Brekke et al. 2008; Mote et al. 2011). For dynamical downscaling, this has historically been a moot point, since the number of downscalable GCMs (with necessary output) has limited the size of the subset to which this type of refinement could be applied. However, CMIP6's introduction of more downscalable models, but with a wide range of quality, necessitates an ongoing engagement with this question. Distinct efforts are already ongoing on a region-by-region basis [e.g., Sobolowski et al. (2023) for Europe; Krantz et al. (2021) for western North America].

In CMIP6 (Eyring et al. 2016) there are more than a hundred simulations from 53 distinct modeling centers. Of these, 28 GCMs have at least one Shared Socioeconomic Pathway (SSP) simulation with 6-hourly fields on model levels archived on the Earth System Grid Federation (ESGF) at the time of writing, versus 8 from the CMIP5 era. While a GCM can be dynamically downscaled based on interpolation from a few vertical pressure levels, the preferred greater vertical resolution in forcing data are also available from a larger set of GCMs in CMIP6 than in CMIP5. The number of forcing scenarios has increased as well with the new SSPs (O'Neill et al. 2016). Other "MIPs" including HighResMIP (Haarsma et al. 2016) are also a potential source of simulations to downscale for further local refinement.

Finally, much has been written about the range of climate sensitivity in the CMIP6 models. Several studies have pointed out that some of the CMIP6 GCMs exhibit climate sensitivities that appear unrealistically high when compared to observed warming in recent decades (Brunner et al. 2020; Zhu et al. 2020; Tokarska et al. 2020), with explanations generally relating to uncertainty in the representation of clouds (Meehl et al. 2020; Zelinka et al. 2020). For this reason, the Intergovernmental Panel on Climate Change (IPCC 2021) has chosen to constrain the range of realistic global temperature projections with historical observations. But, when climate projections are assessed for a given warming level, they are still informed by the regional response in GCMs with low-probability climate sensitivity. This new distinction, however, introduces an additional burden for users of regionally downscaled data with which our community must grapple. For example, it may be valuable to downscale some GCMs despite unrealistically high climate sensitivity to study the response of regional variability to warmer conditions, but a different set would be suitable for stakeholder applications that require future projections to follow the time evolution of external forcing.

While CMIP evaluation studies abound, these studies do not always focus on evaluating quantities that are most relevant for ensuring good RCM performance. Evaluations targeted toward RCM boundary conditions can be distinct from more general model evaluations. For instance, low-resolution models are largely unable to capture extreme events because they cannot resolve sharp gradients and strong vertical motions (Wehner et al. 2014) even if they can correctly capture the large-scale meteorological patterns and environments responsible for triggering extremes (Barlow et al. 2019; Grotjahn et al. 2016). To drive an RCM, it is important to consider the processes that control the boundary conditions that are linked to key regional climate outcomes. An RCM will integrate its own solution over the region, inevitably modifying GCM biases, both because additional processes and topography are resolved and because the RCM has a distinct representation of atmospheric physics and land–atmosphere interactions with its own biases.

Our aim here is to discuss the selection of GCMs to use as boundary conditions for regional climate models, with a focus on the CMIP6 simulations and their distinct challenges. We emphasize choosing models that are right for the right reasons, while retaining a diversity of projections to capture a nondiscountable range of possibilities. Finally, the question of choosing GCMs to downscale is impossible to address without considering the particular uses to which such data may eventually be put, whether dynamical downscaling or another method is employed.

## Process-oriented selection criteria

In this section we are focused on the types of GCM evaluations that can exclude those models that are not "right for the right reasons" for the processes that data users are most interested in.

We begin by discussing some of the large-scale metrics that measure global and hemispheric-scale model fidelity. With direct dynamical downscaling, we have the opportunity to study how changing teleconnections affect variability inside the domain. For that reason, it makes sense to also select GCMs that capture large-scale modes of global climate variability that explain some portion of climate variability in many regions, like El Niño–Southern Oscillation (ENSO) (Diaz et al. 2001). There are many ways to characterize the realism of ENSO (Planton et al. 2021), but at least some measure of the magnitude and pattern of teleconnections (Patricola et al. 2020) and realism in temporal asymmetry (An et al. 2020) are appropriate, though with all of these, care should be taken to account for uncertainties given the limited historical record (Deser et al. 2017; Wittenberg 2009). Further, one might wish to consider other modes of variability (Fasullo et al. 2020) like the Pacific decadal oscillation (PDO; Mantua et al. 1997) or North Atlantic Oscillation (NAO) depending on the region of interest.

In the midlatitudes, properly capturing the statistics of the jet stream and storm tracks is of critical importance to a wide range of applications particularly those affected by precipitation and wind. Metrics for jet stream and storm tracks along with stationary wave placement and blocking would help identify GCMs that will provide realistic boundary conditions for an RCM. The CMIP6 models on the whole show some improvements in these large-scale midlatitude circulation metrics over CMIP5 (Cannon 2020; Simpson et al. 2020), even in cases where other improvements are less evident (Pierce et al. 2021).

The realism of a solution produced by an RCM can also depend on the mean state. For example, overly cool sea surface temperatures (SSTs) mean that tropical cyclones will not be produced without applying a bias correction prior to downscaling (Bruyère et al. 2014). Even in the midlatitudes, biases in atmospheric moisture, temperature, vertical wind shear, and stability can combine to produce unrealistic amounts of rain, rain–snow partitioning, and spring snowpack, so attention must also be paid to the boundary conditions that influence the thermodynamic environment. [A modest thermodynamic mean state bias correction prior to downscaling can alleviate some of these issues (Rahimi et al. 2023, submitted to *Geophys. Res. Lett.*).] With tropical cyclones, too, biases in variability and circulations in the GCM can lead to biases in the interannual variability of these events (Seager et al. 2019; Wu et al. 2022). Given the proper conditions for a tropical cyclone, one may develop in a higher-resolution RCM simulation. Mean-state biases, at least, can be corrected prior to downscaling in GCMs that otherwise have a good representation of the circulation (Bruyère et al. 2014).

Metrics can be used to identify limitations of particular GCMs for downscaling, but can also be used to pinpoint the processes, regions, or phenomena for which given climate datasets or models are credible (Reed et al. 2022). Figure 1 outlines a sample of large-scale processes that could be relevant to evaluate if one were focused on different subregions of the contiguous United States.

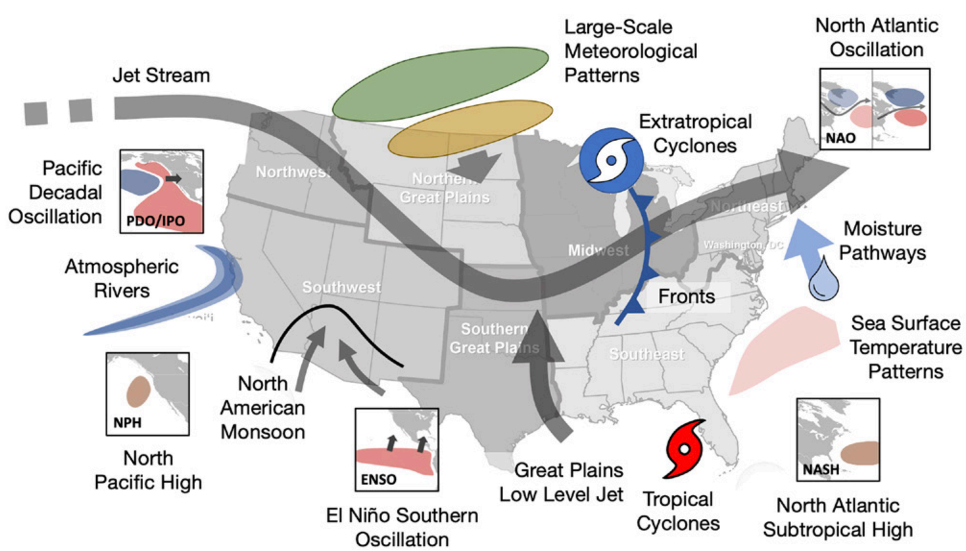


**Fig. 1. A noncomprehensive summary of large-scale processes across the subregions of the continental United States for which GCMs could be evaluated. Large-scale meteorological patterns might be any important weather anomalies of sufficiently large-scale to propagate from the GCM into the RCM.**

## Retaining a diverse set of models

In any of these evaluations, we are mindful of the importance of internal variability (Deser et al. 2012, 2020; Dong et al. 2021) in influencing regional climate, even over decadal time scales. To determine whether a GCM makes a reasonable approximation of historical climate at the regional scale generally requires an ensemble of simulations to sample over natural internal climate variability. As we discuss below, these ensembles can play an important role in the ability to statistically distinguish the performance of different GCMs.

Redundancy in metrics is a natural product of the interconnectedness of the Earth system through relevant process chains and across time scales. For example, in the California region, models that situate the storm track too far to the north also tend to produce precipitation biases that are similarly biased north, and in turn produce conditions that are too dry in Southern California. Identifying these redundancies is important to understand the root cause of model biases and avoid metrics that do not provide additional information to the assessment process.

A cluster analysis applied to the cross-metric dimension can be used to identify metrics, or combinations of metrics, that yield similar results (notably, clustering across models should identify models that are quantitatively similar and allow for better sampling of the model space). Subselection of metrics in this manner is performed using principal component analysis by Pierce et al. (2009, 2021) and using principal feature analysis by Xue and Ullrich (2021). The unsupervised clustering approach and its variants, however, require that the metrics within a given set be all normalized in a consistent fashion (Pierce et al. 2009). The evaluation of GCMs to test for fitness-for-purpose for downscaling often make use of collections of metrics that have many components, such as spatial maps, variable vectors, or time series of the evolution of a variable or process.

It is unlikely for any application that there is such a thing as a "best" model for providing RCM boundary conditions in the region of interest for two reasons. First, based on the type of quantitative assessment just described, most likely a set of models will constitute the best available cohort, with statistically indistinguishable performance. To characterize this, it is essential to assess multiple ensemble members when they are available, as many metrics of interest can be subject to internal variability of similar order to the intermodel differences. Second, even if care is exercised in selecting appropriate metrics, a realistic response to climate change may depend on a phenomenon the modelers did not anticipate. Expert consensus has been studied regarding global measures of climate model quality (Burrows et al. 2018), but the elements of quality that are most important on a region-by-region basis may differ, and those elements may be weighted differently by different users.

Once a cohort of high-performing GCMs has been selected, more simulations may remain than would be practical to downscale. Downscaling groups may wish to select particular realizations to downscale that span a range of climate futures due to a combination of forced responses and internal variability. Additionally, designers of downscaling ensembles often prefer to select simulations from distinct GCMs in order to sample structural uncertainties, and given a fixed number of simulations available for downscaling, selecting single simulations from different GCMs tends to give better skill than selecting multiple simulations from a single GCM (Pierce et al. 2009). This is not as simple as selecting a GCM from a different modeling center given how many components different GCMs can share (Abramowitz et al. 2019). Many GCMs share genealogies (Knutti et al. 2013) which can be invoked to justify selecting a set of GCMs that are more structurally different in their code bases (Climate Change Technical Advisory Group 2015; Sanderson et al. 2015). Indeed, multiple variants of GCMs (e.g., run at different resolutions, with/without biogeochemistry) from the same centers are overrepresented in the subset that have made available the necessary data for direct dynamical downscaling (see Table S1 in the supplemental material; 24 of the 100+ GCMs submitted for CMIP6).

From the science perspective of using dynamical downscaling to build understanding of processes, separating the forced response from internal variance is important. For some climate adaptation planners, on the other hand, it may be more important to simply understand the full range of possible outcomes, regardless of their source. Therefore, if one is selecting from all available ensemble members of the previously identified cohort of GCMs, one may simply wish to span the range of regional temperature and precipitation projections that are possible. By the end of the century in a highly forced scenario, more of that variance across simulations will be due to the anthropogenic forcing, whereas earlier in the century internal variance may dominate, though the extent of this partitioning depends on the region and variable (Goldenson et al. 2018; Hawkins and Sutton 2011; Lehner et al. 2020).

Alternatively, one might also select simulations to downscale that contain the most extreme events to quantify worst case occurrences with especially long return periods (e.g., Huang et al. 2020). Relatedly, one might instead be justified in selecting simulations to capture a range of possible changes in variance or other higher-order statistics of, e.g., extreme precipitation, for stakeholders who are especially interested in anticipating changes in climate variability that have been found in some models (Pendergrass et al. 2017; Swain et al. 2018). Still others have implemented more thorough procedures for ensuring the independence of the simulations chosen across multiple measures (Sexton et al. 2019; Karmalkar et al. 2019).

There is a growing acceptance of the procedure of applying an emergent constraint to the climate sensitivity itself (e.g., Liang et al. 2020; Tokarska et al. 2020; IPCC 2021), essentially downweighting models that produce more warming than has been observed. Given the previously discussed presence of high-climate-sensitivity models in CMIP6, one might be tempted to include the climate sensitivity among the GCM evaluation metrics, to avoid those higher-sensitivity models that were downweighted in the IPCC's global mean temperature estimate. However, in Fig. 2, we plot the equilibrium climate sensitivity (Meehl et al. 2020) against a combined error score for measures of Northern Hemisphere midlatitude circulation and its variability (Simpson et al. 2020) and ENSO (see supplemental material for a full description). The combined error score is the Euclidean distance across the metrics. Because it is based on a set of normalized mean square errors (NMSEs), smaller scores are better. There is, in fact, a significant negative correlation ($R = -0.4$, $p < 0.05$) between the climate sensitivity and this summary error score for the

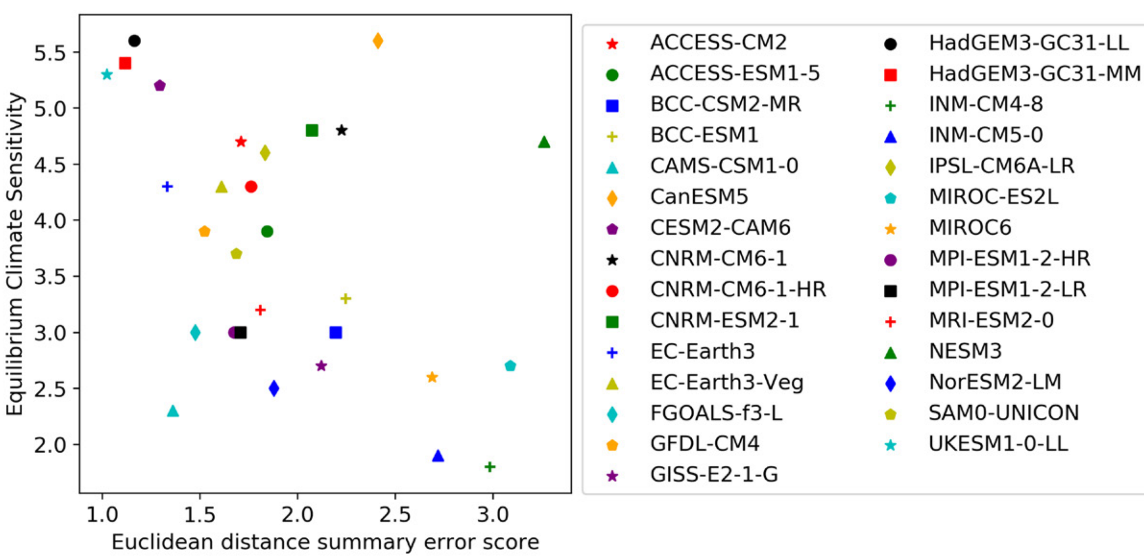


**Fig. 2. A summary NMSE score across a set of Northern Hemisphere circulation fields negatively correlates ($R = -0.4$) with the equilibrium climate sensitivity of the GCMs.**

performance of the historical simulations. Of course, we are not implying that this correlation represents a causal relationship, but that it simply indicates that some of the models that have the highest climate sensitivities are among the best in their representation of other features. Thus, attention must be paid to the GCMs that occupy a space in the upper-left quadrant where a strong representation of midlatitude circulation coincides with some of the highest climate sensitivities out of the set of GCMs examined.

In IPCC (2021), these GCMs are being downweighted by an emergent constraint for global mean temperature change, while still being included in an unweighted mean of the pattern of regional change, with the latter calculated contingent on several global warming levels (IPCC 2021). If a similar approach is adopted in the interpretation of downscaled data, then GCMs need not be excluded based on climate sensitivity.

***Knowing how to use the data.*** There are practical motivations for considering regional climate responses at particular warming levels: 1) as we mentioned, the GCMs with the potentially unrealistic climate sensitivities overlap with those that perform well at certain measures for midlatitude dynamics (see Fig. 3), and 2) different modeling centers have made different choices about which SSP(s) to favor with storage for all of the driving data necessary to force an RCM on many vertical model levels (see Table S1 in the supplemental material). The ability to derive information that can be compared across SSPs is also helpful to ensure that a range of climate responses are sampled. Moreover, this approach allows one to view the climate changes from the point of view of acceptable or unacceptable damage to various resource systems (e.g., IPCC 2018, 2022).

Standardizing outputs against a particular global warming level, however, conflates the role of greenhouse gas and other forcings and neglects that the transient climate response has fast and slow components and is thus rate dependent (Wu et al. 2010). To

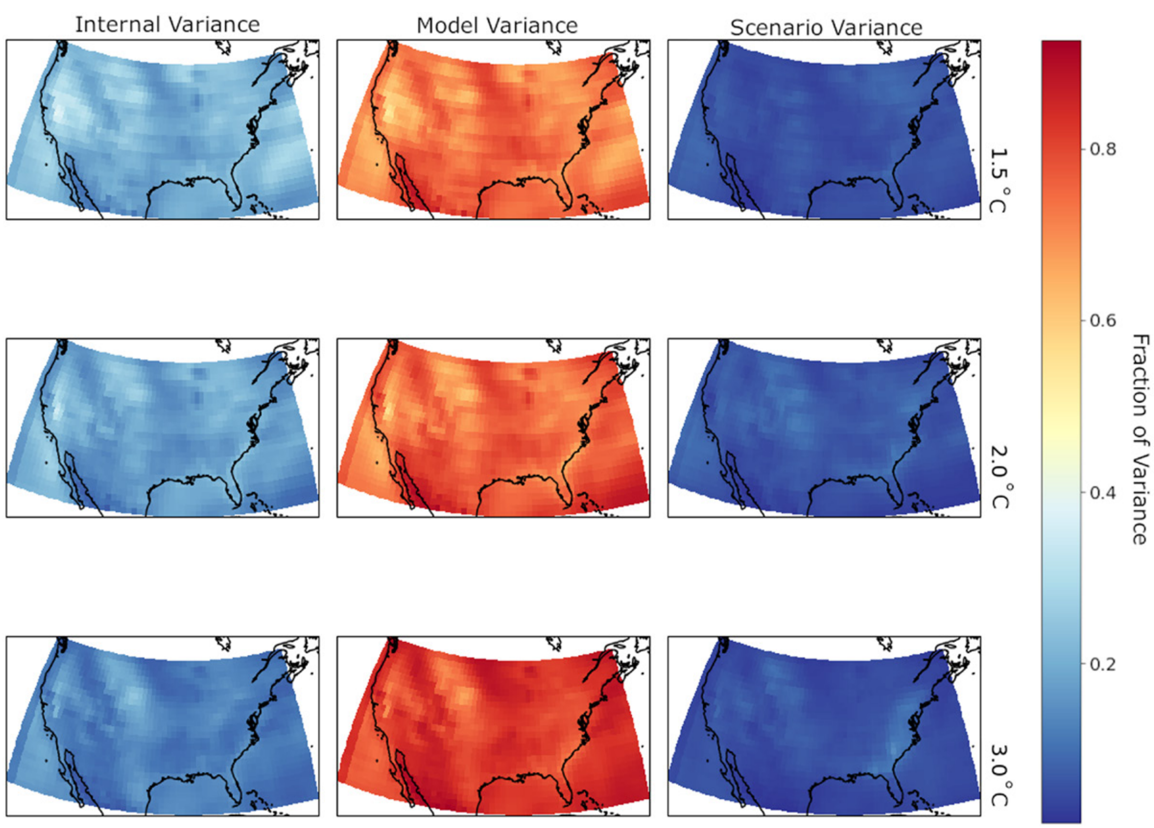


**Fig. 3. Variance fractions over CONUS for near-surface air temperature at several global warming levels based on the available realizations of the 45 models listed in Table S1 of the supplemental material.**

determine whether it is appropriate to treat the warming-level response as independent of the SSP from which it was sampled, we plot the fractions of cross-SSP variance and internal variance for a set of CMIP6 GCMs at a series of warming levels (Fig. 3). The variable is a 20-yr mean of near-surface air temperature centered on the time that the warming level is reached. We see that the internal variance generally exceeds in magnitude any SSP-dependent uncertainty. Examining one's variable of interest, perhaps for a season of interest, can inform whether the warming level assumption is a good one for a particular user of the data. We also emphasize that we would only expect this framework to apply for those SSPs that exhibit a monotonic increase in greenhouse gas concentrations over time, hence we limited our analysis to SSP2–4.5, SSP3–7.0, and SSP5–8.5.

Data users who require projected climate impacts as a function of time within their planning frameworks will still need to grapple with how to weigh the risk of more rapid changes, be they due to high climate sensitivity and/or high emissions scenarios or regional feedback mechanisms, and should consult with climate experts. Nonetheless, when the use case does allow it, disaggregating the time dimension from the regional response can give stakeholders a wide range of scenarios to consider for conditions at a given warming-level benchmark.

## Discussion

We have emphasized how the choice of GCMs to drive regional climate models may depend on the processes relevant to data users in a particular region. This bespoke approach, however, could result in different GCMs being selected by different users for different purposes. For a regional planner at the city or state level, it may be of little concern that their counterpart across the continent finds different GCMs more fit for purpose. On the other hand, data users in finance or insurance may be more concerned that a consistent set of downscaled projections is used globally. For example, to find a set of generally relevant metrics across multiple regions that are not application specific, McSweeney et al. (2015) use 850-hPa wind speed, storm-track density, the annual cycle of surface temperature and precipitation, and key teleconnections taken as a set (across Southeast Asia, Europe, and Africa).

Researchers are often left to their own devices to make these decisions based on their own subjective concept of what would be useful to downstream users (Cash et al. 2006; McNie 2007). No data producers, however, can escape subjective choices. For example, if one wanted to preserve the statistics of the larger CMIP6 ensemble, one might select particular GCM ensemble members that span a range of possibilities, excluding outliers deemed less likely. If, on the other hand, one wanted to explore worst-case scenarios or downscale as much variability as possible to inform emulation, then one might be sure to include those outliers.

For transparency, stakeholders should be engaged in a discussion of where these simulations fall within the larger CMIP ensemble as a dataset of downscaled GCMs develops. If certain regional responses are more likely in the high-performing GCMs, then further scientific analysis is warranted to determine whether there is a credible physical mechanism underpinning the result and that it survives out-of-sample testing before concluding that there is an emergent constraint (Hall et al. 2019) on the possible outcomes.

Interactions with planners can determine whether further scientific investigation is relevant to inform decision-making. This may also require a significant investment of time to explore the data complexities and implications collectively and iteratively for the planning framework in practice (Dilling and Lemos 2011; Lemos et al. 2012; Meadow et al. 2015). Because the final set of downscaled realizations cannot be optimized to everyone's needs, practitioners and climate experts may end up determining that a subset of the downscaled simulations is most appropriate for a particular application.

Finally, we note that regardless of the region, the work does not end with the GCM selection or the downscaling, but continues in the interpretation of the downscaled data, including determination of the differential credibility of the downscaled simulation, for decision-making and other applications in light of the GCM selection choices that were made. Moreover, appropriate provisioning of the data, in a climate services context, is also a key part of the process (McGinnis and Mearns 2021; NSTC 2023)

***Next steps for the modeling community.*** One of the reasons for the warming level approach is that not all modeling centers have saved the necessary output to make a wide range of SSPs and realizations dynamically downscalable. To avoid such gaps in a future CMIP7, we encourage modeling centers to save driving data for RCMs on vertical model levels for multiple scenarios and ensemble members, and to better standardize variables saved and their formats so that the process of creating driving data for RCMs from a variety of GCMs can be streamlined.

While our examples were centered on North America, we believe that more concerted community prioritization is necessary to determine to which subregions the vast computational resources required to downscale many simulations at convective-resolving scales are allocated. With a rapidly developing demand for downscaled climate projections to inform decision-making, resources should be directed to the research community to examine and interpret these datasets for a wide variety of end-uses, especially for those communities that cannot afford to purchase such assistance in the private sector.

**Acknowledgments.** N. G. provided overall leadership of the manuscript, including the general drafting of the manuscript and supervising the review and editing process. W. K., P. A. U., K. A. R., L. M., L. R. L., I. S., and D. P. also contributed substantially to the drafting of the text. All authors were involved in the editing and review of the full-draft manuscript. P. A. U. leads the project administration and supervision of the HyperFACETS project (https://hyperfacets.ucdavis.edu/), which supported N. G., P. A. U., K. A. R., L. M., L. R. L., A. H., A. J., and S. R. HyperFACETS is supported by the U.S. Department of Energy Office of Science Biological and Environmental Research as part of the Regional and Global Model Analysis and Multisectoral Dynamics program areas. PNNL is operated for the Department of Energy by Battelle Memorial Institute under Contract DE-AC05-76RL01830. A. H., D. P., and S. R. are supported on the California Energy Commission grant "Development of Climate Projections for California and Identification of Priority Projections" (Agreement EPC-20-006). I. S. and L. M. were supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement 1852977.

**Data availability statement.** For the CMIP6 data, we acknowledge the World Climate Research Programme's Working Group on Coupled Modelling and the individual modeling groups (listed in Table S1 in the supplemental material) for their roles in making these data available. The data are archived at https://esgf-node.llnl.gov/projects/cmip6/.

## References

Abramowitz, G., and Coauthors, 2019: ESD reviews: Model dependence in multi-model climate ensembles: Weighting, sub-selection and out-of-sample testing. *Earth Syst. Dyn.*, **10**, 91–105, https://doi.org/10.5194/esd-10-91-2019.

An, S.-I., E. Tziperman, Y. M. Okumura, and T. Li, 2020: ENSO irregularity and asymmetry. *El Niño Southern Oscillation in a Changing Climate*, *Geophys. Monogr.*, Vol., 253, Amer. Geophys. Union, 153–172, https://doi.org/10.1002/9781119548164.ch7.

Barlow, M., and Coauthors, 2019: North American extreme precipitation events and related large-scale meteorological patterns: A review of statistical methods, dynamics, modeling, and trends. *Climate Dyn.*, **53**, 6835–6875, https://doi.org/10.1007/s00382-019-04958-z.

Brekke, L. D., M. D. Dettinger, E. P. Maurer, and M. Anderson, 2008: Significance of model credibility in estimating climate projection distributions for regional hydroclimatological risk assessments. *Climatic Change*, **89**, 371–394, https://doi.org/10.1007/s10584-007-9388-3.

Brunner, L., A. G. Pendergrass, F. Lehner, A. L. Merrifield, R. Lorenz, and R. Knutti, 2020: Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth Syst. Dyn.*, **11**, 995–1012, https://doi.org/10.5194/esd-11-995-2020.

Bruyère, C. L., J. M. Done, G. J. Holland, and S. Fredrick, 2014: Bias corrections of global models for regional climate simulations of high-impact weather. *Climate Dyn.*, **43**, 1847–1856, https://doi.org/10.1007/s00382-013-2011-6.

Bukovsky, M. S., and L. O. Mearns, 2020: Regional climate change projections from NA-CORDEX and their relation to climate sensitivity. *Climatic Change*, **162**, 645–665, https://doi.org/10.1007/s10584-020-02835-x.

Burrows, S. M., A. Dasgupta, S. Reehl, L. Bramer, P.-L. Ma, P. J. Rasch, and Y. Qian, 2018: Characterizing the relative importance assigned to physical variables by climate scientists when assessing atmospheric climate model fidelity. *Adv. Atmos. Sci.*, **35**, 1101–1113, https://doi.org/10.1007/s00376-018-7300-x.

Cannon, A. J., 2020: Reductions in daily continental-scale atmospheric circulation biases between generations of global climate models: CMIP5 to CMIP6. *Environ. Res. Lett.*, **15**, 064006, https://doi.org/10.1088/1748-9326/ab7e4f.

Cash, D. W., J. C. Borck, and A. G. Patt, 2006: Countering the loading-dock approach to linking science and decision making: Comparative analysis of El Niño/Southern Oscillation (ENSO) forecasting systems. *Sci. Technol. Human Values*, **31**, 465–494, https://doi.org/10.1177/0162243906287547.

Climate Change Technical Advisory Group, 2015: Perspectives and guidance for climate change analysis. California Department of Water Resources, 142 pp., https://water.ca.gov/-/media/DWR-Website/Web-Pages/Programs/All-Programs/Climate-Change-Program/Climate-Program-Activities/Files/Reports/Perspectives-Guidance-Climate-Change-Analysis.pdf.

Deser, C., R. Knutti, S. Solomon, and A. S. Phillips, 2012: Communication of the role of natural variability in future North American climate. *Nat. Climate Change*, **2**, 775–779, https://doi.org/10.1038/nclimate1562.

——, I. R. Simpson, K. A. McKinnon, and A. S. Phillips, 2017: The Northern Hemisphere extratropical atmospheric circulation response to ENSO: How well do we know it and how do we evaluate models accordingly? *J. Climate*, **30**, 5059–5082, https://doi.org/10.1175/JCLI-D-16-0844.1.

——, and Coauthors, 2020: Insights from Earth system model initial-condition large ensembles and future prospects. *Nat. Climate Change*, **10**, 277–286, https://doi.org/10.1038/s41558-020-0731-2.

Diaz, H. F., M. P. Hoerling, and J. K. Eischeid, 2001: ENSO variability, teleconnections and climate change. *Int. J. Climatol.*, **21**, 1845–1862, https://doi.org/10.1002/joc.631.

Dilling, L., and M. C. Lemos, 2011: Creating usable science: Opportunities and constraints for climate knowledge use and their implications for science policy. *Global Environ. Change*, **21**, 680–689, https://doi.org/10.1016/j.gloenvcha.2010.11.006.

Dong, L., L. R. Leung, F. Song, and J. Lu, 2021: Uncertainty in El Niño-like warming and California precipitation changes linked by the interdecadal Pacific oscillation. *Nat. Commun.*, **12**, 6484, https://doi.org/10.1038/s41467-021-26797-5.

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

Fasullo, J. T., A. S. Phillips, and C. Deser, 2020: Evaluation of leading modes of climate variability in the CMIP archives. *J. Climate*, **33**, 5527–5545, https://doi.org/10.1175/JCLI-D-19-1024.1.

Giorgi, F., 2019: Thirty years of regional climate modeling: Where are we and where are we going next? *J. Geophys. Res. Atmos.*, **124**, 5696–5723, https://doi.org/10.1029/2018JD030094.

Goldenson, N., G. Mauger, L. R. Leung, C. M. Bitz, and A. Rhines, 2018: Effects of ensemble configuration on estimates of regional climate uncertainties. *Geophys. Res. Lett.*, **45**, 926–934, https://doi.org/10.1002/2017GL076297.

Grotjahn, R., and Coauthors, 2016: North American extreme temperature events and related large scale meteorological patterns: A review of statistical methods, dynamics, modeling, and trends. *Climate Dyn.*, **46**, 1151–1184, https://doi.org/10.1007/s00382-015-2638-6.

Gutowski, W. J., and Coauthors, 2020: The ongoing need for high-resolution regional climate models: Process understanding and stakeholder information. *Bull. Amer. Meteor. Soc.*, **101**, E664–E683, https://doi.org/10.1175/BAMS-D-19-0113.1.

Haarsma, R. J., and Coauthors, 2016: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geosci. Model Dev.*, **9**, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016.

Hall, A., P. Cox, C. Huntingford, and S. Klein, 2019: Progressing emergent constraints on future climate change. *Nat. Climate Change*, **9**, 269–278, https://doi.org/10.1038/s41558-019-0436-6.

Harvey, B. J., P. Cook, L. C. Shaffrey, and R. Schiemann, 2020: The response of the Northern Hemisphere storm tracks and jet streams to climate change in the CMIP3, CMIP5, and CMIP6 climate models. *J. Geophys. Res. Atmos.*, **125**, e2020JD032701, https://doi.org/10.1029/2020JD032701.

Hawkins, E., and R. Sutton, 2011: The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dyn.*, **37**, 407–418, https://doi.org/10.1007/s00382-010-0810-6.

Huang, X., D. L. Swain, and A. D. Hall, 2020: Future precipitation increase from very high resolution ensemble downscaling of extreme atmospheric river storms in California. *Sci. Adv.*, **6**, eaba1323, https://doi.org/10.1126/sciadv.aba1323.

IPCC, 2018: *Global Warming of 1.5°C.* V. Masson-Delmotte et al., Eds., Cambridge University Press, 616 pp., https://doi.org/10.1017/9781009157940.

——, 2021: *Climate Change 2021: The Physical Science Basis.* V. Masson-Delmotte et al., Eds., Cambridge University Press, 2392 pp., https://doi.org/10.1017/9781009157896.

——, 2022: *Climate Change 2022: Impacts, Adaptation and Vulnerability.* H.-O. Pörtner et al., Eds., Cambridge University Press, 3056 pp., https://doi.org/10.1017/9781009325844.

Karmalkar, A. V., D. M. H. Sexton, J. M. Murphy, B. B. B. Booth, J. W. Rostron, and D. J. McNeall, 2019: Finding plausible and diverse variants of a climate model. Part II: Development and validation of methodology. *Climate Dyn.*, **53**, 847–877, https://doi.org/10.1007/s00382-019-04617-3.,

Knutti, R., D. Masson, and A. Gettelman, 2013: Climate model genealogy: Generation CMIP5 and how we got there. *Geophys. Res. Lett.*, **40**, 1194–1199, https://doi.org/10.1002/grl.50256.

Kotamarthi, R., K. Hayhoe, L. O. Mearns, D. Wuebbles, J. Jacobs, and J. Jurado, 2021: *Downscaling Techniques for High-Resolution Climate Projections: From Global Change to Local Impacts.* Cambridge University Press, 201 pp.

Krantz, W., D. Pierce, N. Goldenson, and D. Cayan, 2021: Memorandum on evaluating global climate models for studying regional climate change in California (interim deliverable for EPC-20-006). California Energy Commission, 24 pp., https://www.energy.ca.gov/media/7264.

Lehner, F., C. Deser, N. Maher, J. Marotzke, E. M. Fischer, L. Brunner, R. Knutti, and E. Hawkins, 2020: Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth Syst. Dyn.*, **11**, 491–508, https://doi.org/10.5194/esd-11-491-2020.

Lemos, M. C., C. J. Kirchhoff, and V. Ramprasad, 2012: Narrowing the climate information usability gap. *Nat. Climate Change*, **2**, 789–794, https://doi.org/10.1038/nclimate1614.

Liang, Y., N. P. Gillett, and A. H. Monahan, 2020: Climate model projections of 21st century global warming constrained using the observed warming trend. *Geophys. Res. Lett.*, **47**, e2019GL086757, https://doi.org/10.1029/2019GL086757.

Liu, C., and Coauthors, 2017: Continental-scale convection-permitting modeling of the current and future climate of North America. *Climate Dyn.*, **49**, 71–95, https://doi.org/10.1007/s00382-016-3327-9.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on Salmon production. *Bull. Amer. Meteor. Soc.*, **78**, 1069–1080, https://doi.org/10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2.

McGinnis, S., and L. O. Mearns, 2021: Building a climate service for North America based on the NA-CORDEX data archive. *Climate Serv.*, **22**, 100233, https://doi.org/10.1016/j.cliser.2021.100233.

McNie, E. C., 2007: Reconciling the supply of scientific information with user demands: An analysis of the problem and review of the literature. *Environ. Sci. Policy*, **10**, 17–38, https://doi.org/10.1016/j.envsci.2006.10.004.

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

Meadow, A. M., D. B. Ferguson, Z. Guido, A. Horangic, G. Owen, and T. Wall, 2015: Moving toward the deliberate coproduction of climate science knowledge. *Wea. Climate Soc.*, **7**, 179–191, https://doi.org/10.1175/WCAS-D-14-00050.1.

Mearns, L. O., and Coauthors, 2012: The North American Regional Climate Change Assessment Program: Overview of phase I results. *Bull. Amer. Meteor. Soc.*, **93**, 1337–1362, https://doi.org/10.1175/BAMS-D-11-00223.1.

Meehl, G. A., C. A. Senior, V. Eyring, G. Flato, J.-F. Lamarque, R. J. Stouffer, K. E. Taylor, and M. Schlund, 2020: Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. *Sci. Adv.*, **6**, eaba1981, https://doi.org/10.1126/sciadv.aba1981.

Mote, P., L. Brekke, P. B. Duffy, and E. Maurer, 2011: Guidelines for constructing climate scenarios. *Eos, Trans. Amer. Geophys. Union*, **92**, 257–258, https://doi.org/10.1029/2011EO310001.

NSTC, 2023: A Federal Framework and Action Plan for Climate Services: Product of the Fast Track Action Committee on Climate Services. National Science and Technology Council, 24 pp., https://www.whitehouse.gov/wp-content/uploads/2023/03/FTAC_Report_03222023_508.pdf.

O'Neill, B. C., and Coauthors, 2016: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geosci. Model Dev.*, **9**, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016.

Patricola, C. M., J. P. O'Brien, M. D. Risser, A. M. Rhoades, T. A. O'Brien, P. A. Ullrich, D. A. Stone, and W. D. Collins, 2020: Maximizing ENSO as a source of western US hydroclimate predictability. *Climate Dyn.*, **54**, 351–372, https://doi.org/10.1007/s00382-019-05004-8.

Pendergrass, A. G., R. Knutti, F. Lehner, C. Deser, and B. M. Sanderson, 2017: Precipitation variability increases in a warmer climate. *Sci. Rep.*, **7**, 17966, https://doi.org/10.1038/s41598-017-17966-y.

Pierce, D. W., T. P. Barnett, B. D. Santer, and P. J. Gleckler, 2009: Selecting global climate models for regional climate change studies. *Proc. Natl. Acad. Sci. USA*, **106**, 8441–8446, https://doi.org/10.1073/pnas.0900094106.

——, D. R. Cayan, J. Goodrich, T. Das, and A. Munévar, 2021: Evaluating global climate models for hydrological studies of the upper Colorado River basin. *J. Amer. Water Res. Assoc.*, **58**, 709–734, https://doi.org/10.1111/1752-1688.12974.

Planton, Y. Y., and Coauthors, 2021: Evaluating climate models with the CLIVAR 2020 ENSO metrics package. *Bull. Amer. Meteor. Soc.*, **102**, E193–E217, https://doi.org/10.1175/BAMS-D-19-0337.1.

Priestley, M. D. K., D. Ackerley, J. L. Catto, K. I. Hodges, R. E. McDonald, and R. W. Lee, 2020: An overview of the extratropical storm tracks in CMIP6 historical simulations. *J. Climate*, **33**, 6315–6343, https://doi.org/10.1175/JCLI-D-19-0928.1.

Reed, K. A., and Coauthors, 2022: Metrics as tools for bridging climate science and applications. *Wiley Interdiscip. Rev.: Climate Change*, **13**, e799, https://doi.org/10.1002/wcc.799.

Rummukainen, M., 2010: State-of-the-art with regional climate models. *Wiley Interdiscip. Rev.: Climate Change*, **1**, 82–96, https://doi.org/10.1002/wcc.8.

Sanderson, B. M., R. Knutti, and P. Caldwell, 2015: A representative democracy to reduce interdependency in a multimodel ensemble. *J. Climate*, **28**, 5171–5194, https://doi.org/10.1175/JCLI-D-14-00362.1.

Schär, C., C. Frei, D. Lüthi, and H. C. Davies, 1996: Surrogate climate-change scenarios for regional climate models. *Geophys. Res. Lett.*, **23**, 669–672, https://doi.org/10.1029/96GL00265.

Seager, R., M. Cane, N. Henderson, D.-E. Lee, R. Abernathey, and H. Zhang, 2019: Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. *Nat. Climate Change*, **9**, 517–522, https://doi.org/10.1038/s41558-019-0505-x.

Sexton, D. M. H., A. V. Karmalkar, J. M. Murphy, K. D. Williams, I. A. Boutle, C. J. Morcrette, A. J. Stirling, and S. B. Vosper, 2019: Finding plausible and diverse variants of a climate model. Part 1: Establishing the relationship between errors at weather and climate time scales. *Climate Dyn.*, **53**, 989–1022, https://doi.org/10.1007/s00382-019-04625-3.

Simpson, I. R., and Coauthors, 2020: An evaluation of the large-scale atmospheric circulation and its variability in CESM2 and other CMIP models. *J. Geophys. Res. Atmos.*, **125**, e2020JD032835, https://doi.org/10.1029/2020JD032835.

Sobolowski, S., and Coauthors, 2023: EURO-CORDEX CMIP6 GCM selection & ensemble design: Best practices and recommendations. Zenodo, 211pp., https://doi.org/10.5281/zenodo.7673400.

Swain, D. L., B. Langenbrunner, J. D. Neelin, and A. Hall, 2018: Increasing precipitation volatility in twenty-first-century California. *Nat. Climate Change*, **8**, 427–433, https://doi.org/10.1038/s41558-018-0140-y.

Tebaldi, C., and R. Knutti, 2007: The use of the multi-model ensemble in probabilistic climate projections. *Philos. Trans. Roy. Soc.*, **A365**, 2053–2075, https://doi.org/10.1098/rsta.2007.2076.

Tokarska, K. B., G. C. Hegerl, A. P. Schurer, P. M. Forster, and K. Marvel, 2020: Observational constraints on the effective climate sensitivity from the historical period. *Environ. Res. Lett.*, **15**, 034043, https://doi.org/10.1088/1748-9326/ab738f.

Wehner, M. F., K. A. Reed, and F. Li, 2014: The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1. *J. Adv. Model. Earth Syst.*, **6**, 980–997, https://doi.org/10.1002/2013MS000276.

Wittenberg, A. T., 2009: Are historical records sufficient to constrain ENSO simulations? *Geophys. Res. Lett.*, **36**, L12702, https://doi.org/10.1029/2009GL038710.

Wu, P., R. Wood, J. Ridley, and J. Lowe, 2010: Temporary acceleration of the hydrological cycle in response to a $CO_2$ rampdown. *Geophys. Res. Lett.*, **37**, L12705, https://doi.org/10.1029/2010GL043730.

Wu, X., K. A. Reed, P. Callaghan, and J. T. Bacmeister, 2022: Exploring western North Pacific tropical cyclone activity in the high-resolution Community Atmosphere Model. *Earth Space Sci.*, **9**, e2021EA001862, https://doi.org/10.1029/2021EA001862.

Xue, Z., and P. Ullrich, 2021: A retrospective and prospective examination of the 1960s U.S. northeast drought. *Earth's Future*, **9**, e2020EF001930, https://doi.org/10.1029/2020EF001930.

Zelinka, M. D., T. A. Myers, D. T. McCoy, S. Po-Chedley, P. M. Caldwell, P. Ceppi, S. A. Klein, and K. E. Taylor, 2020: Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett.*, **47**, e2019GL085782, https://doi.org/10.1029/2019GL085782.

Zhu, J., C. J. Poulsen, and B. L. Otto-Bliesner, 2020: High climate sensitivity in CMIP6 model not supported by paleoclimate. *Nat. Climate Change*, **10**, 378–379, https://doi.org/10.1038/s41558-020-0764-6.