Supporting Information for "Combination of decadal predictions and climate projections in time: challenges and potential solutions"

D. J. Befort[1], L. Brunner[2,3], L.F. Borchert[4,5], C.H. O'Reilly[6], J. Mignot[4],

A.P. Ballinger[7], G. Hegerl[7], J.M. Murphy[8], A. Weisheimer[9,10]

[1]Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, United Kingdom

[2]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland

[3]Department of Meteorology and Geophysics, University of Vienna, Vienna, Austria

[4]Sorbonne Universités (SU/CNRS/IRD/MNHN), LOCEAN Laboratory, Institut Pierre Simon Laplace (IPSL), Paris, France

[5]École Normale Supérieure, Laboratoire de Météorologie Dynamique (LMD), IPSL, Paris, France

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

[7]School of Geosciences, University of Edinburgh, Edinburgh, United Kingdom

[8]Met Office Hadley Centre, Exeter, United Kingdom

[9]National Centre for Atmospheric Science, University of Oxford, Oxford, United Kingdom

[10]European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

Contents of this file

3. Tables S1

4. Figures S1 to S8

1. Description of variance inflation method (VINF)

The simple calibration technique, called the variance inflation method (VINF), following Doblas-Reyes, Hagedorn, and Palmer (2005), which has been recently applied to large-ensemble projections (O'Reilly et al., 2020) is used here. The uncalibrated (original) ensemble can be expressed as

$$X_{uncalibrated}(t,m) = X_{ens-mean}(t) + X_{ens-anom}(t,m) \tag{1}$$

In this formula, t and m indicate the dependence on time and ensemble member. A calibrated ensemble $X_{calibrated}$ can be obtain by:

$$X_{calibrated}(t,m) = \alpha X_{ens-mean}(t) + \beta X_{ens-anom}(t,m) \tag{2}$$

with α and β as

$$\alpha = \rho \frac{\sigma_{obs}}{\sigma_{ens-mean}} \tag{3}$$

$$\beta = \sqrt{1-\rho^2}\frac{\sigma_{obs}}{\sigma_{ensemble}} \tag{4}$$

Here ρ is the linear Pearson correlation coefficient between the observations and the ensemble mean, σ_{obs} the standard deviation of the observations, $\sigma_{ens-mean}$ the standard deviation of the uncalibrated ensemble mean and $\sigma_{ensemble}$ the square root of the mean variance of the uncalibrated ensemble members. The resulting calibrated ensemble is per construction reliable in the sense that the root-mean-square-error matches the ensemble spread (Fortin et al., 2014). In this study, the VINF method has been applied to the climate projection MME from 1970 to 2014 and separately for each lead year to the decadal prediction MME. The ensemble median of HadCRUT5 is used as a reference/observational dataset, which is a blended product of the CRUTEM5 land-surface air temperature dataset and the HadSST4 sea-surface temperature (SST) dataset (Morice et al., 2021). In accordance with model data, anomalies are calculated to the 1970 until 2014 average.

2. Description of model weighting method

The fundamental idea of model weighting approaches is that a models' future projection is assigned a weight based on the models' ability to reproduce observations in the past (Abramowitz et al., 2019). Applying such weights has been used to quantify model uncertainty for many different applications, including future changes of temperature, precipitation, sea ice, and ozone at different scales from local to global (Massonnet et al., 2012; Knutti et al., 2017; Lorenz et al., 2018; Brunner et al., 2019; Amos et al., 2020; Brunner et al., 2020; Merrifield et al., 2020). Here an adaptation of such a method is described which is used to investigate if weighting can help to reduce inconsistencies when concatenating predictions and projections.

Typically, model weights are based on multiple metrics relevant to the variable of interest to assess if a model is 'fit for purpose' and to avoid overconfident weighting (Lorenz et al., 2018). Here, we generalise the weighting to assign weights not only to model projections but also, lead-year depended, to predictions. The weights are based on five metrics, which are calculated individually for each of the regions investigated. These metrics are chosen to be generally relevant for our target but not individually tailored to each of the regions, leaving room for improvement. We use (1) the area-weighted mean of the root-mean-squared-distance (RMSD) between model and observation time-series from all grid cells in each region (temperature and precipitation), (2) the area-weighted RMSD of the model and observation standard deviations over time (temperature and precipitation), and (3) the area-weighted RMSD of the model and observation linear trends in time (temperature). The metrics are calculated separately for each prediction lead time and the projections, normalised by the median over all models, and combined to generalised distances. In earlier applications these distances were then translated into weights, for example, by using a Gaussian weighting function scaled by a performance shape parameter (Brunner et al., 2020). However, the use of only seven models in this study means that no perfect model test (which needs several tens of models) can be performed and therefore no objective shape parameter can be established to translate distances to weights. Instead we use the simpler approach of just calculating the weights as the inverse distance as stated in the main manuscript.

Table S1. Decadal prediction (dcppA-hindcasts) and climate projections datasets.

Institution	Model	Member (dcppA)	Member (historical)	Reference
CCCma	CanESM5	20 (i1)	20	Sospedra-Alfonso et al. (2021)
EC-Earth-Consortium	EC-Earth3	10 (i1)	10	Bilbao et al. (2021) Haarsma et al. (2020)
IPSL	IPSL-CM6A-LR	10 (i1)	10	Boucher et al. (2020)
MIROC	MIROC6	10 (i1)	10	Kataoka et al. (2020) Tatebe et al. (2019)
MOHC	HadGEM3-GC31-MM	10 (i1)	4	Andrews et al. (2020) Williams et al. (2018)
MPI-M	MPI-ESM1-2-HR	10 (i1)	10	Müller et al. (2018) Mauritsen et al. (2019)
NCC	NorCPM1	10 (i1) 10 (i2)	20	Bethke et al. (2021)

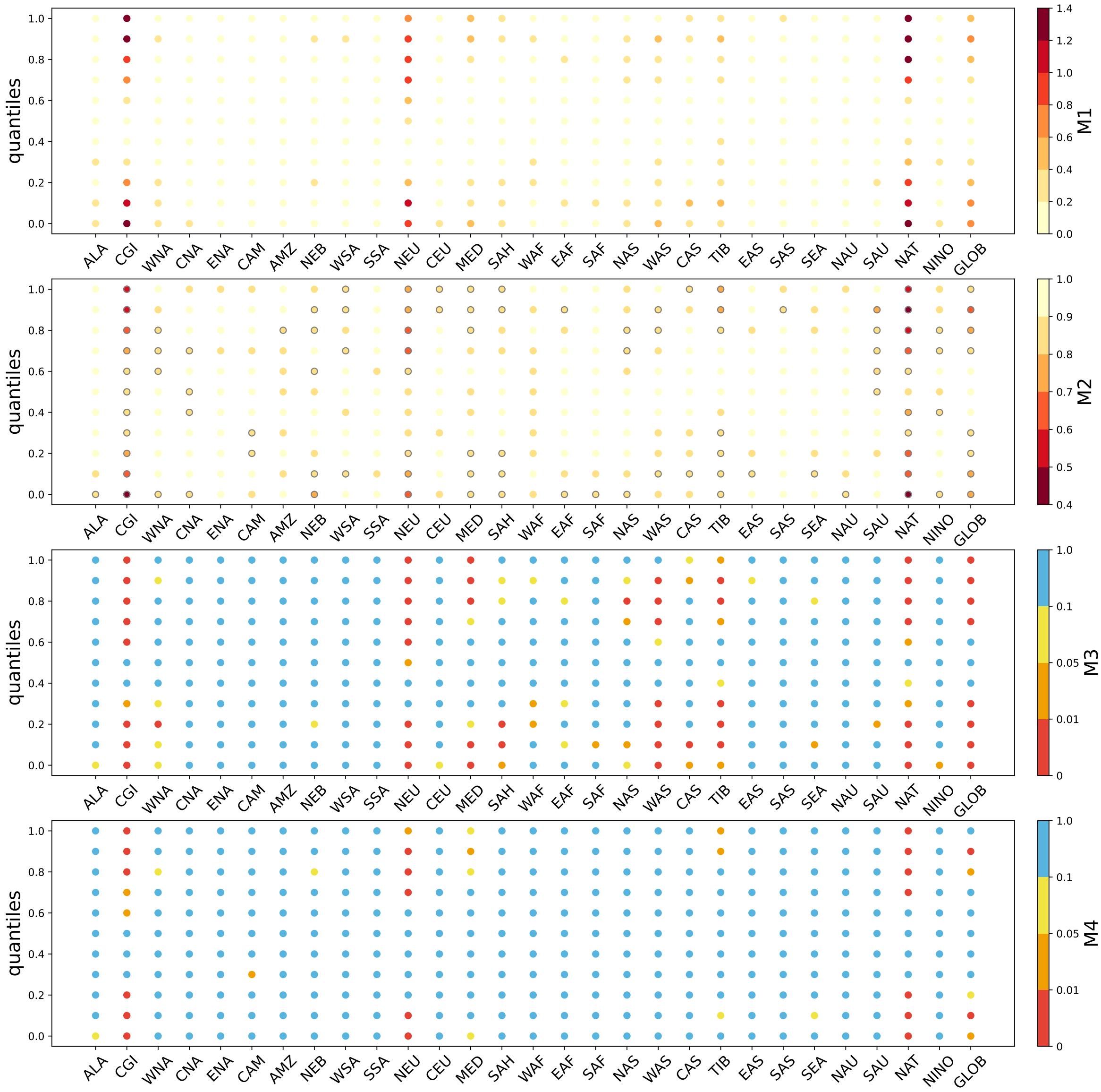

Figure S1. Results for metrics M1 to M4 when concatenating decadal predictions after forecast year ten and climate projections. The circles in M2 indicate those values which are significantly different to the 10th percentile of randomly sampled *baseline* distribution overlaps.

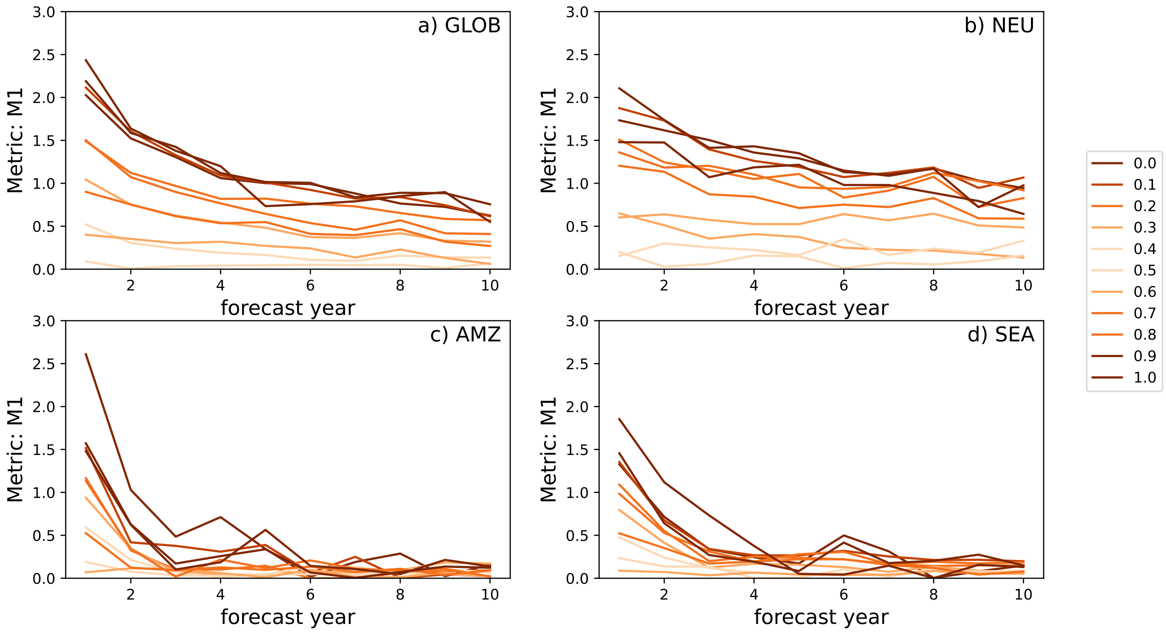

Figure S2. Results for Metric M1 when using different forecast years to concatenate decadal predicitons and climate projections for a) global SAT, b) Northern Europe SAT, c) Amazon SAT and d) South East Asia (SEA) SAT. Colors represent different quantiles of the MME.

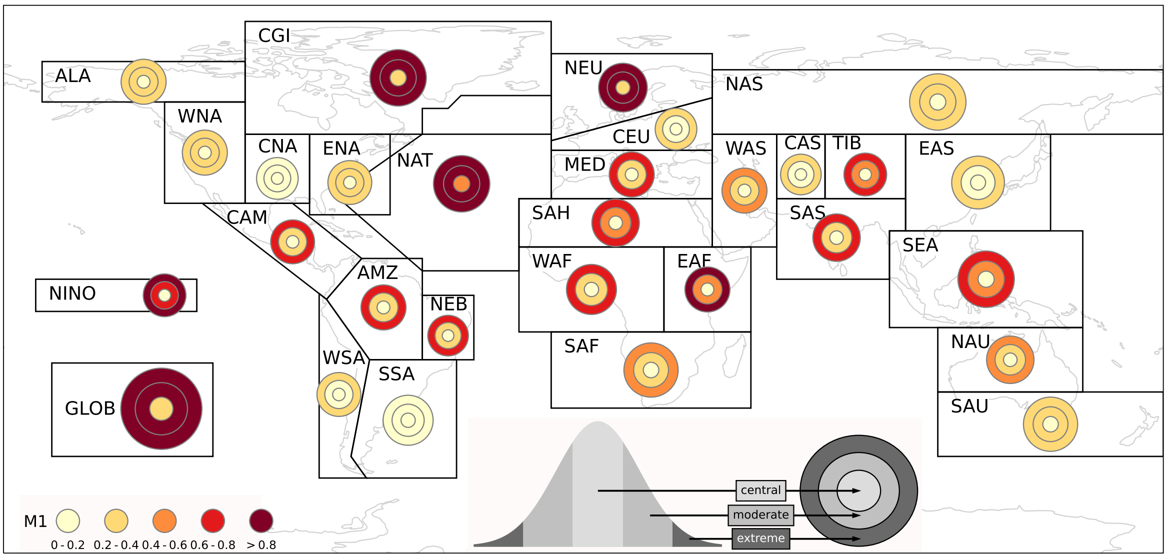

Figure S3. As Figure 3a but when concatenating decadal predictions and climate projections after forecast year 2.

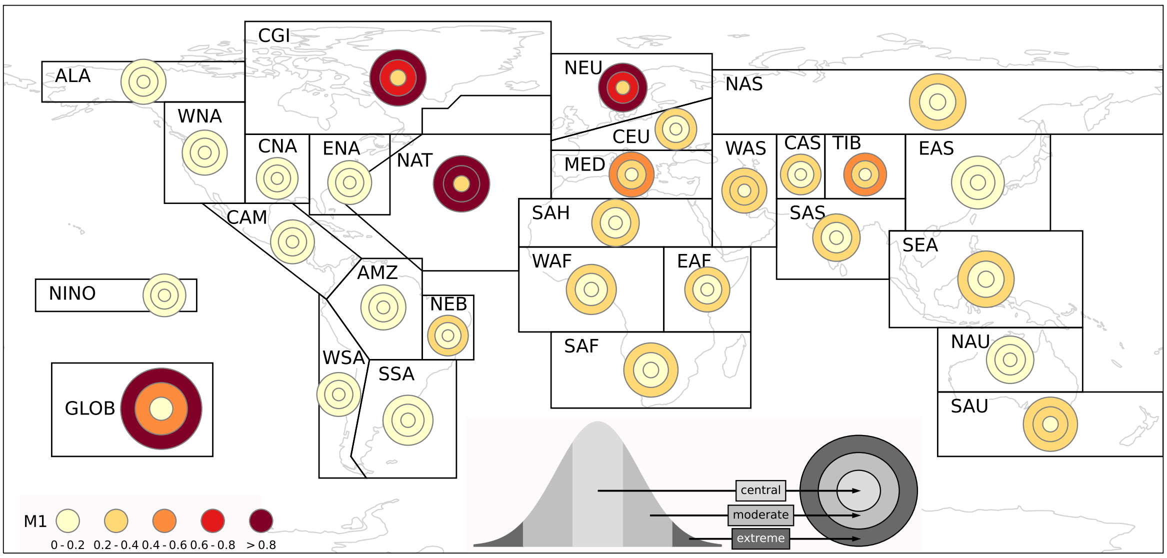

Figure S4. As Figure 3a but when concatenating decadal predictions and climate projections after forecast year 7.

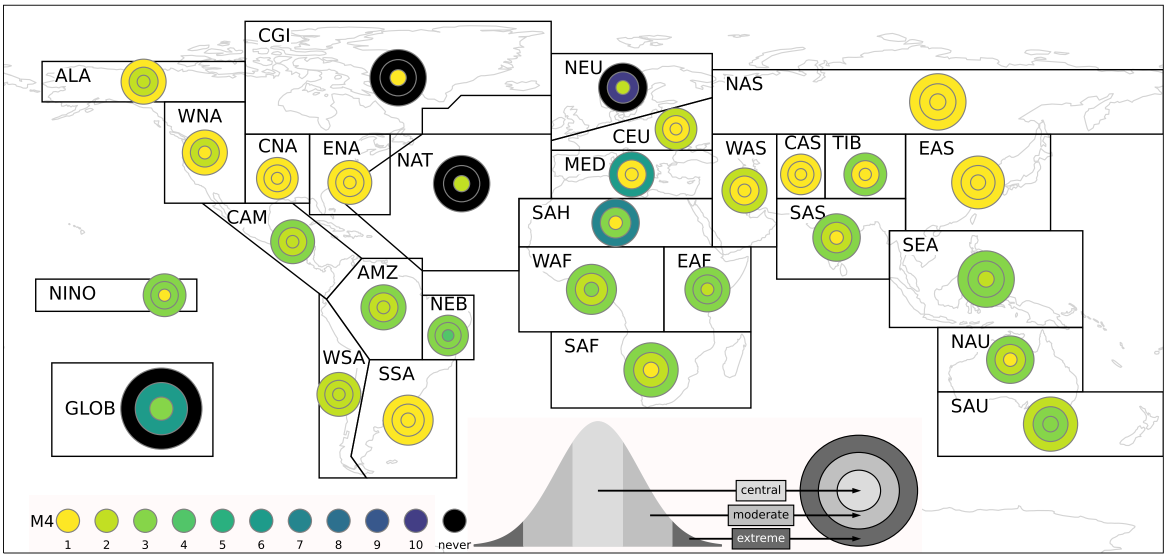

Figure S5. Forecast year at which baseline and merged distributions are indistinguishable based on M4 (ks-test). The first forecast year at which both distributions are non-significantly different from each other has been determined the following: For each forecast year, the percentage of all quantiles with p-values greater than 0.05 is calculated. The value shown is the first year for which at least 50% of all p-values are greater than 0.05 and all following forecast years have at least 50% p-values larger than 0.05. Due to the small sample size this criterion can be violated for one (but not more) subsequent forecast years (meaning that the percentage of p-values greater than 0.05 can be below 50%).

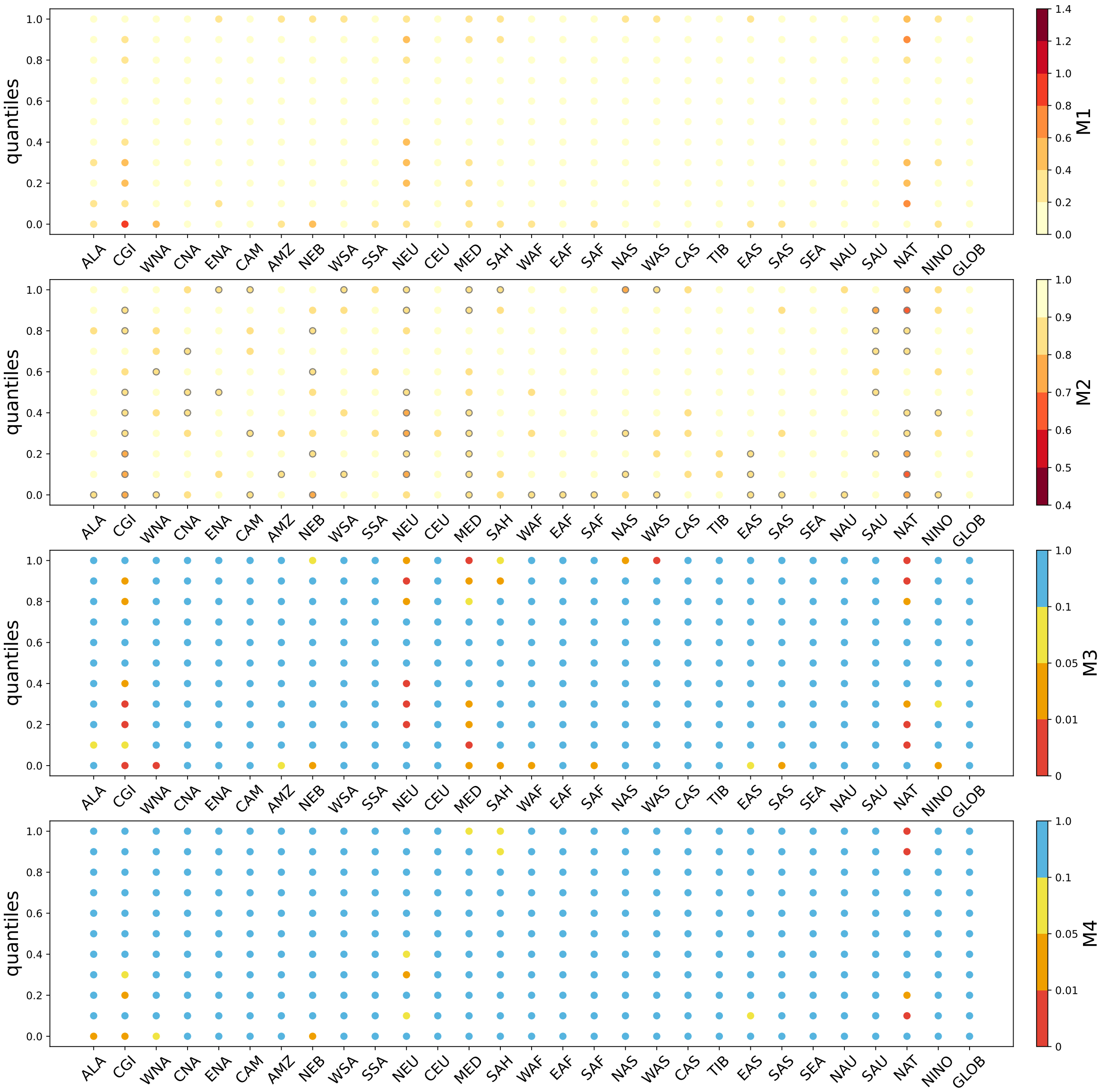

Figure S6. As Fig. S1 but for calibrated decadal predictions and calibrated climate projections.

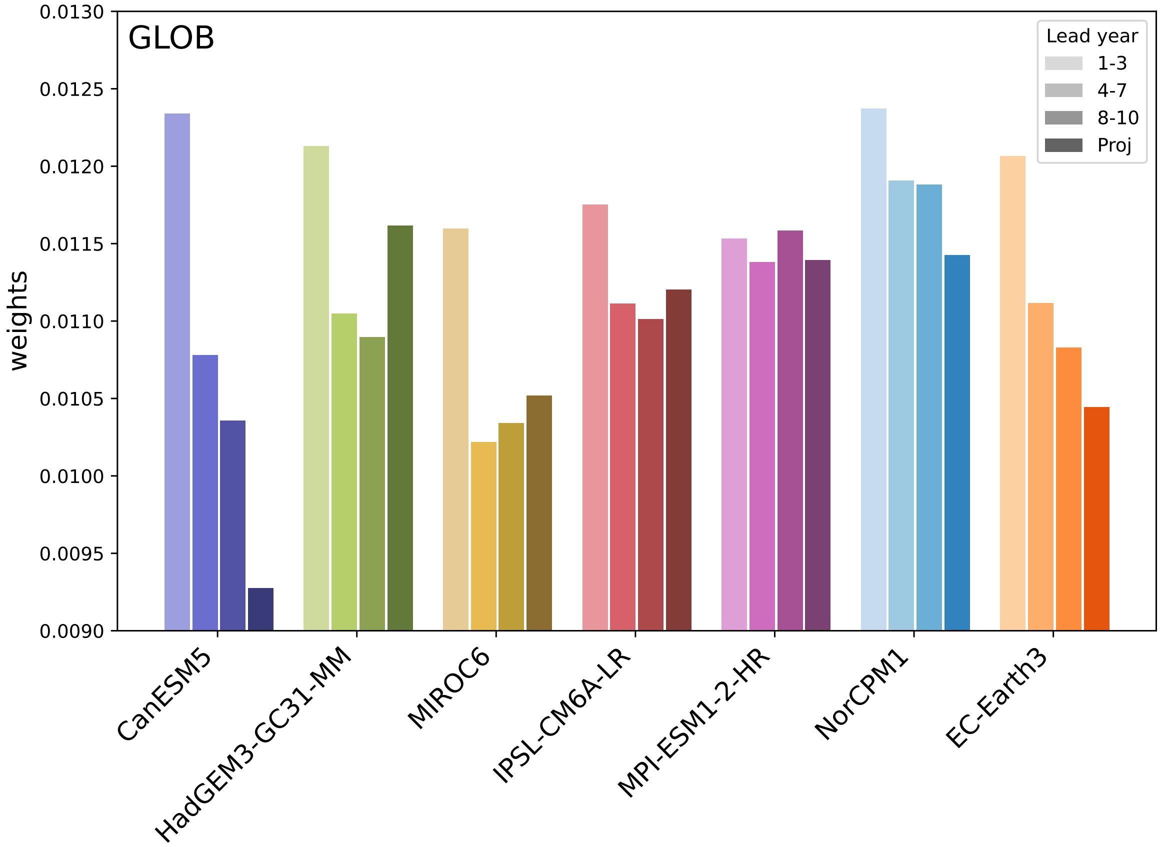

Figure S7. Weights (1/distance) for the global (GLOB) case for the different models used in this study. Weights for the predictions are aggregated for lead-years 1-3, 4-7 and 8-10.

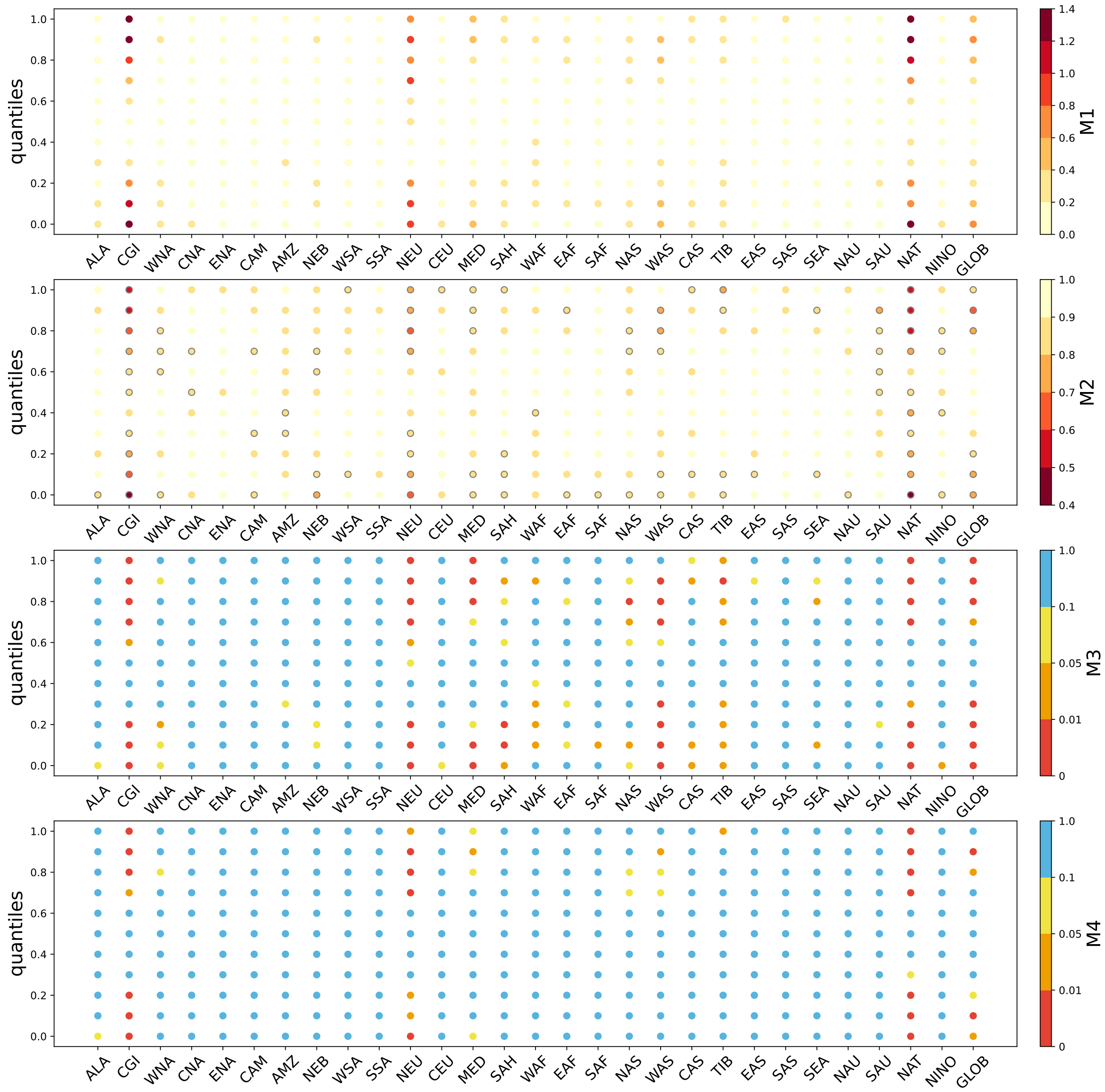

Figure S8. As Fig. S1 but for weighted decadal predictions and weighted climate projections.

References

Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., ... 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*(1), 91–105. Retrieved from `https://esd.copernicus.org/articles/10/91/2019/` doi: 10.5194/esd-10-91-2019

Amos, M., Young, P. J., Hosking, J. S., Lamarque, J.-F., Abraham, N. L., Akiyoshi, H., ... Yamashita, Y. (2020). Projecting ozone hole recovery using an ensemble of chemistry–climate models weighted by model performance and independence. *Atmospheric Chemistry and Physics*, *20*(16), 9961–9977. Retrieved from `https://acp.copernicus.org/articles/20/9961/2020/` doi: 10.5194/acp-20-9961-2020

Andrews, M. B., Ridley, J. K., Wood, R. A., Andrews, T., Blockley, E. W., Booth, B., ... Sutton, R. T. (2020). Historical Simulations With HadGEM3-GC3.1 for CMIP6. *Journal of Advances in Modeling Earth Systems*, *12*(6), e2019MS001995. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001995` (e2019MS001995 10.1029/2019MS001995) doi: https://doi.org/10.1029/2019MS001995

Bethke, I., Wang, Y., Counillon, F., Keenlyside, N., Kimmritz, M., Fransner, F., ... Eldevik, T. (2021). NorCPM1 and its contribution to CMIP6 DCPP. *Geoscientific Model Development*, *14*(11), 7073–7116. Retrieved from `https://gmd.copernicus.org/articles/14/7073/2021/` doi: 10.5194/gmd-14-7073-2021

Bilbao, R., Wild, S., Ortega, P., Acosta-Navarro, J., Arsouze, T., Bretonnière, P.-A., ... Vegas-Regidor, J. (2021). Assessment of a full-field initialized decadal climate prediction system with the CMIP6 version of EC-Earth. *Earth System Dynamics*, *12*(1), 173–196. Retrieved from https://esd.copernicus.org/articles/12/173/2021/ doi: 10.5194/esd-12-173-2021

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., ... Vuichard, N. (2020). Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. *Journal of Advances in Modeling Earth Systems*, *12*(7), e2019MS002010. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS002010 (e2019MS002010 10.1029/2019MS002010) doi: https://doi.org/10.1029/2019MS002010

Brunner, L., Lorenz, R., Zumwald, M., & Knutti, R. (2019, nov). Quantifying uncertainty in European climate projections using combined performance-independence weighting. *Environmental Research Letters*, *14*(12), 124010. Retrieved from https://doi.org/10.1088/1748-9326/ab492f doi: 10.1088/1748-9326/ab492f

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., & Knutti, R. (2020). Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth System Dynamics*, *11*(4), 995–1012. Retrieved from https://esd.copernicus.org/articles/11/995/2020/ doi: 10.5194/esd-11-995-2020

Doblas-Reyes, F. J., Hagedorn, R., & Palmer, T. N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting — II. Calibration and combination. *Tellus A: Dynamic Meteorology and Oceanography*, *57*(3), 234-252. Retrieved from https://doi.org/10.3402/tellusa.v57i3.14658 doi: 10.3402/tellusa.v57i3.14658

Fortin, V., Abaza, M., Anctil, F., & Turcotte, R. (2014). Why Should Ensemble Spread

Match the RMSE of the Ensemble Mean? *Journal of Hydrometeorology*, *15*(4), 1708 - 1713. Retrieved from https://journals.ametsoc.org/view/journals/hydr/15/4/jhm-d-14-0008_1.xml doi: 10.1175/JHM-D-14-0008.1

Haarsma, R., Acosta, M., Bakhshi, R., Bretonnière, P.-A., Caron, L.-P., Castrillo, M., ... Wyser, K. (2020). HighResMIP versions of EC-Earth: EC-Earth3P and EC-Earth3P-HR – description, model computational performance and basic validation. *Geoscientific Model Development*, *13*(8), 3507–3527. Retrieved from https://gmd.copernicus.org/articles/13/3507/2020/ doi: 10.5194/gmd-13-3507-2020

Kataoka, T., Tatebe, H., Koyama, H., Mochizuki, T., Ogochi, K., Naoe, H., ... Watanabe, M. (2020). Seasonal to Decadal Predictions With MIROC6: Description and Basic Evaluation. *Journal of Advances in Modeling Earth Systems*, *12*(12), e2019MS002035. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS002035 (e2019MS002035 2019MS002035) doi: https://doi.org/10.1029/2019MS002035

Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., & Eyring, V. (2017). A climate model projection weighting scheme accounting for performance and interdependence. *Geophysical Research Letters*, *44*(4), 1909-1918. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL072012 doi: https://doi.org/10.1002/2016GL072012

Lorenz, R., Herger, N., Sedláček, J., Eyring, V., Fischer, E. M., & Knutti, R. (2018). Prospects and Caveats of Weighting Climate Models for Summer Maximum Temperature Projections Over North America. *Journal of Geophysical Research: Atmospheres*, *123*(9), 4509-

4526. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017JD027992 doi: https://doi.org/10.1029/2017JD027992

Massonnet, F., Fichefet, T., Goosse, H., Bitz, C. M., Philippon-Berthier, G., Holland, M. M., & Barriat, P.-Y. (2012). Constraining projections of summer Arctic sea ice. *The Cryosphere*, *6*(6), 1383–1394. Retrieved from https://tc.copernicus.org/articles/6/1383/2012/ doi: 10.5194/tc-6-1383-2012

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., ... Roeckner, E. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2. *Journal of Advances in Modeling Earth Systems*, *11*(4), 998-1038. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001400 doi: https://doi.org/10.1029/2018MS001400

Merrifield, A. L., Brunner, L., Lorenz, R., Medhaug, I., & Knutti, R. (2020). An investigation of weighting schemes suitable for incorporating large ensembles into multi-model ensembles. *Earth System Dynamics*, *11*(3), 807–834. Retrieved from https://esd.copernicus.org/articles/11/807/2020/ doi: 10.5194/esd-11-807-2020

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., ... Simpson, I. R. (2021). An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set. *Journal of Geophysical Research: Atmospheres*, *126*(3), e2019JD032361. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD032361 (e2019JD032361 2019JD032361) doi: https://doi.org/10.1029/2019JD032361

Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., ...

Marotzke, J. (2018). A Higher-resolution Version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). *Journal of Advances in Modeling Earth Systems*, *10*(7), 1383-1413. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017MS001217 doi: https://doi.org/10.1029/2017MS001217

O'Reilly, C. H., Befort, D. J., & Weisheimer, A. (2020). Calibrating large-ensemble European climate projections using observational data. *Earth System Dynamics*, *11*(4), 1033–1049. Retrieved from https://esd.copernicus.org/articles/11/1033/2020/ doi: 10.5194/esd-11-1033-2020

Sospedra-Alfonso, R., Merryfield, W. J., Boer, G. J., Kharin, V. V., Lee, W.-S., Seiler, C., & Christian, J. R. (2021). Decadal climate predictions with the Canadian Earth System Model version 5 (CanESM5). *Geoscientific Model Development*, *14*(11), 6863–6891. Retrieved from https://gmd.copernicus.org/articles/14/6863/2021/ doi: 10.5194/gmd-14-6863-2021

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., ... Kimoto, M. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, *12*(7), 2727–2765. Retrieved from https://gmd.copernicus.org/articles/12/2727/2019/ doi: 10.5194/gmd-12-2727-2019

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., ... Xavier, P. K. (2018). The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations. *Journal of Advances in Modeling Earth Systems*, *10*(2), 357-380. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/

2017MS001115 doi: https://doi.org/10.1002/2017MS001115