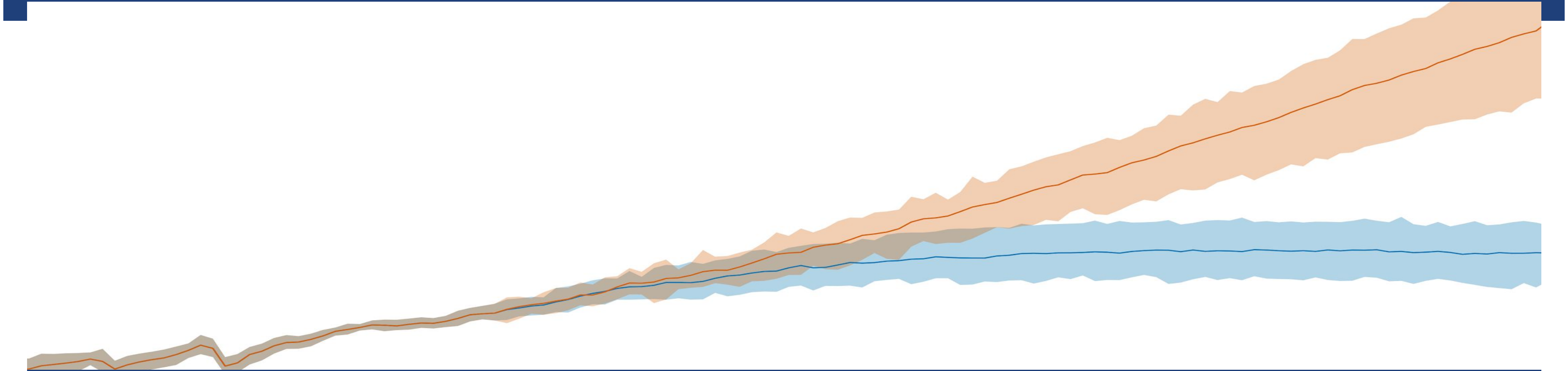


Weighting models by performance and independence
Effects on projections of future climate

Lukas Brunner | Wegener Center Common Space | October 21st 2021

With contributions from Reto Knutti, Ruth Lorenz, Angeline G. Pendergrass, Flavio Lehner, Anna L. Merrifield and many others

What are climate models?

- *A model is an informative **representation** of an object, person or system.* Wikipedia
- *Climate models simulate the interactions of the **important** drivers of climate.* Wikipedia
- Climate model are used to
  - simulate historical climate
  - understand (parts of) the climate system and interactions
  - **project future climate**
  - etc...

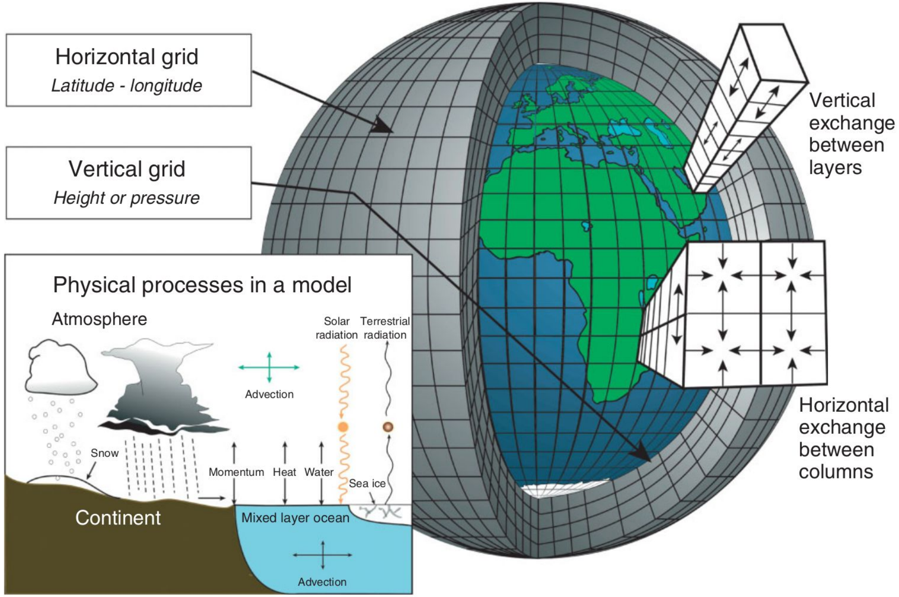
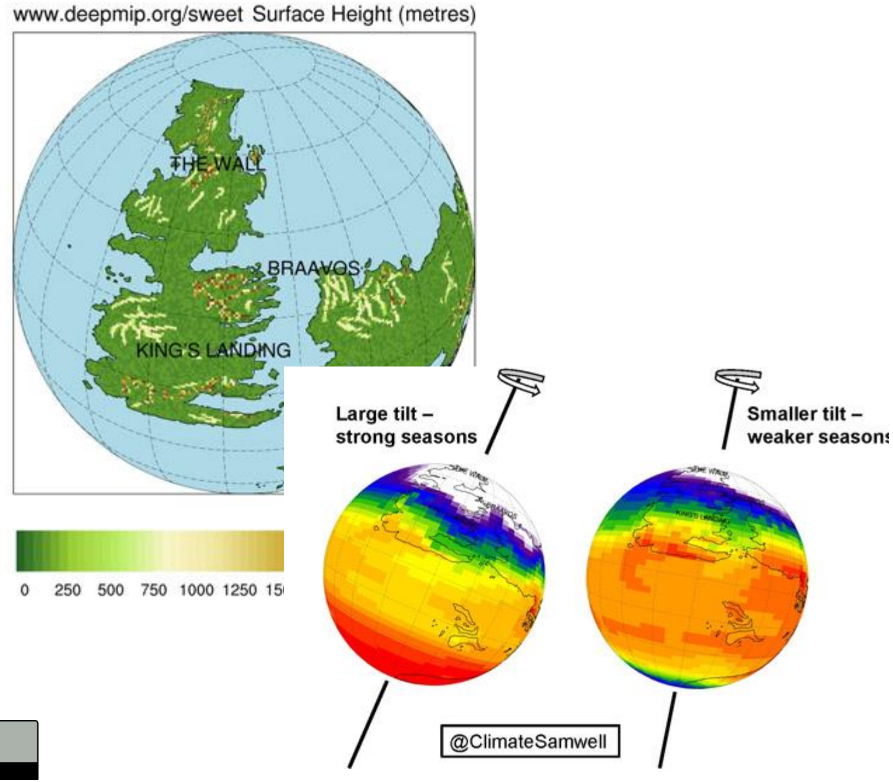


Figure: Schematic representation of a general circulation model. Edwards (2011)

What are climate models used for?

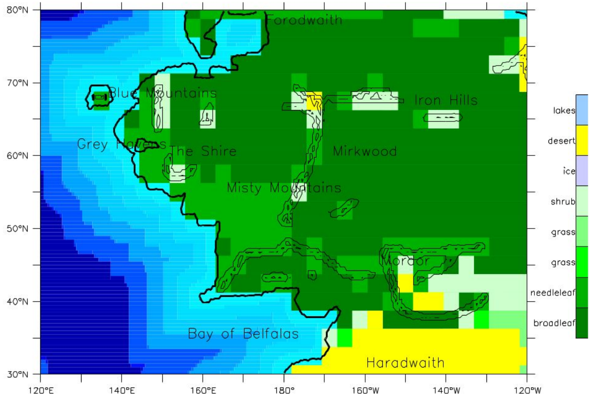
The world of Game of Thrones @ClimateSamwell


www.deepmip.org/sweet Surface Height (metres)
THE WALL
BRAAVOS
KING'S LANDING
0 250 500 750 1000 1250 15
Large tilt – strong seasons
Smaller tilt – weaker seasons
@ClimateSamwell

The Climate of Middle Earth

Radagast the Brown[1,2]

[1]Rhosgobel, nr. Carrock, Mirkwood, Middle Earth.
[2]The Cabot Institute, University of Bristol, UK.



Why do we need reliable projections of future climate?

- Within science
  - investigation of feedbacks
  - regional studies
  - impact assessments
- Infrastructure planning
- Climate adaptation
- Climate mitigation decisions
- ...

Figure: Damaged water pipeline due to thawing permafrost in Norway. CC-BY-NC-ND Rakesh Rao / Climate Visuals Countdown

Uncertainty in model projections of future climate

- Different socio-economic developments are represented by **scenario uncertainty**
- Structural differences in models lead to **model uncertainty**
- The chaotic behavior of the climate system leads to **internal variability**

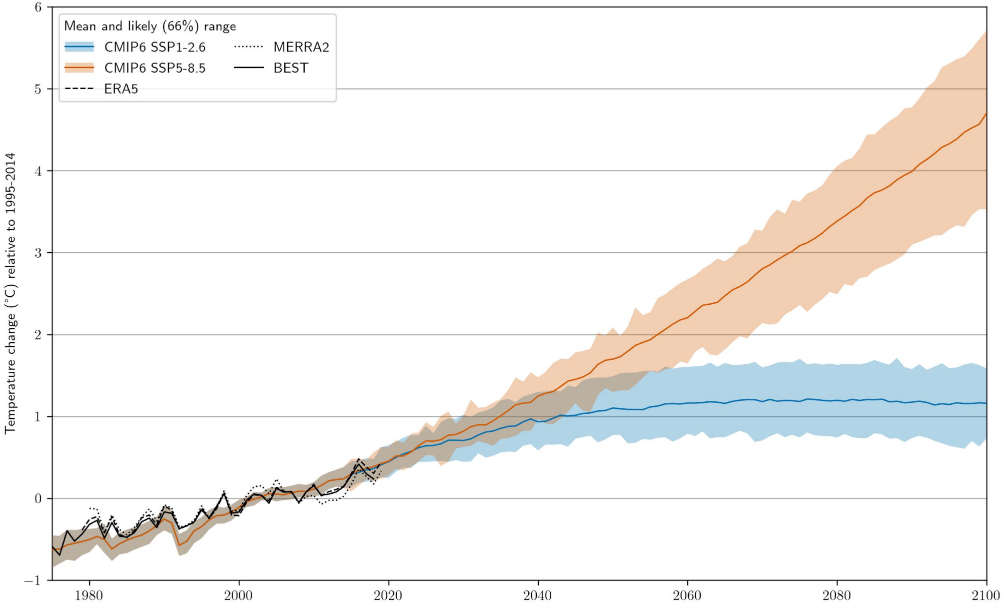


Figure: Global mean, annual mean temperature change (relative to 1995-2014) from CMIP6. Brunner et al. (2020a)

Known and unknown model uncertainty

- Model uncertainty arises when looking at **multi-model ensembles**
- Model uncertainty ≠ actual uncertainty (e.g., IPCC AR5 & 6)
    - ○ there might be process es not covered by any model (not considered here)
    - ○ **not all models are equally ‘good’**
    - ○ **not all model are independent**

→ Here we look at uncertainty from model spread

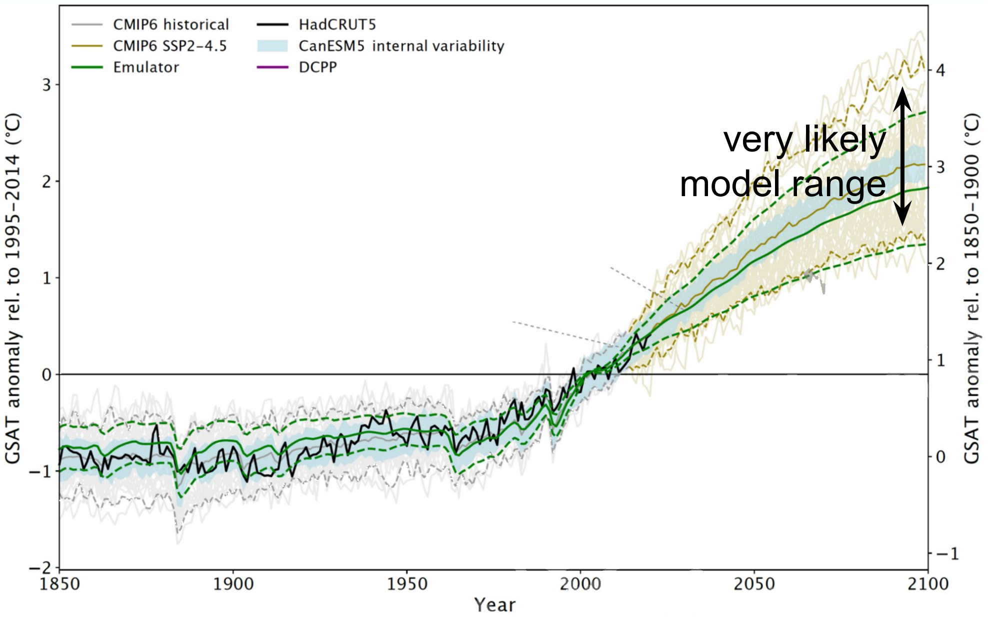


Figure: Global mean, annual mean temperature change based on 39 CMIP6 models. The dashed brown lines indicate the 90% model range. IPCC AR6

Not all models are equally ‘fit for purpose’

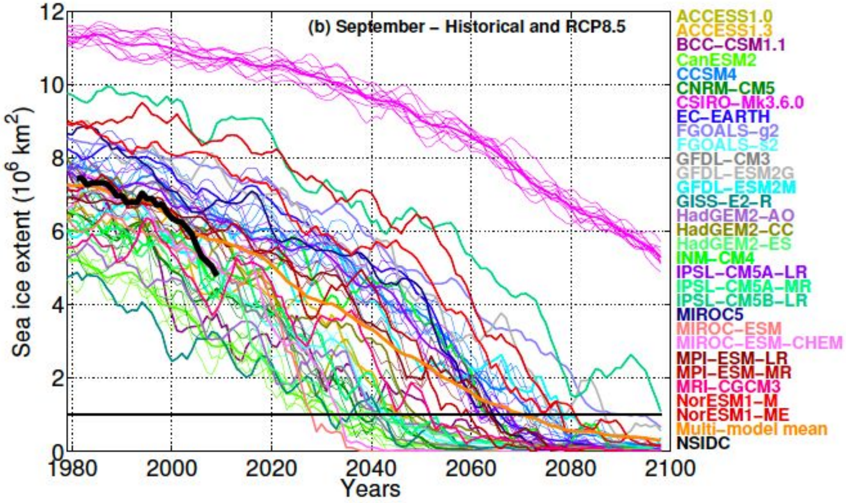


Figure: September Arctic sea ice extent in CMIP5 historical / RCP8.5 runs and observations. Massonnet et al. (2012)

Not all models are equally ‘fit for purpose’

- we might want to trust models less if they are far away from observations → **weighting by performance**
- need a way to **convert** model-observation **distance into weights**
    - if we are very strict: strong weighting leaving us only with few models
    - if we are very generous: weak weighting not doing anything
- weights should be based on **metrics relevant to the target**

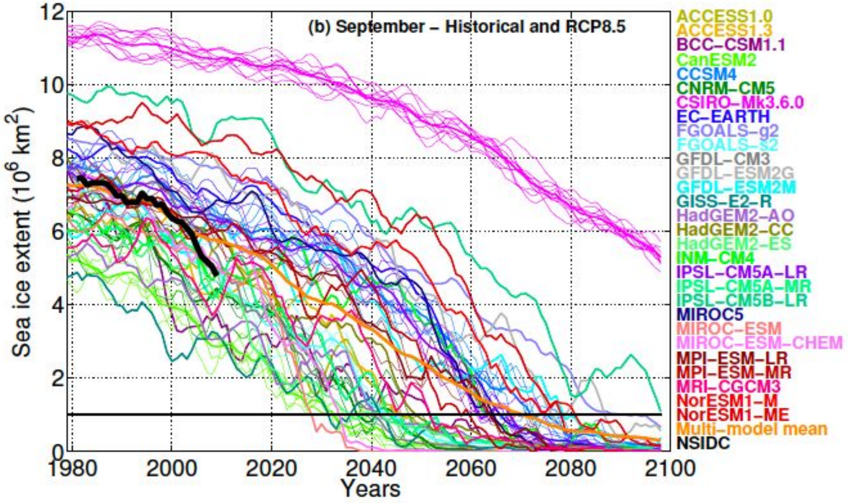


Figure: September Arctic sea ice extent in CMIP5 historical / RCP8.5 runs and observations. Massonnet et al. (2012)

Not all models are independent

- Multi-model studies often draw on **all available models**
- the CMIP multi-model ensembles are not designed to only include independent models (**‘ensembles of opportunity’**)
  - Several models are closely related (one different component, resolution)
  - Models have been branched from each other
  - Some models share components

→ **weighting by independence**

The AGCM family tree

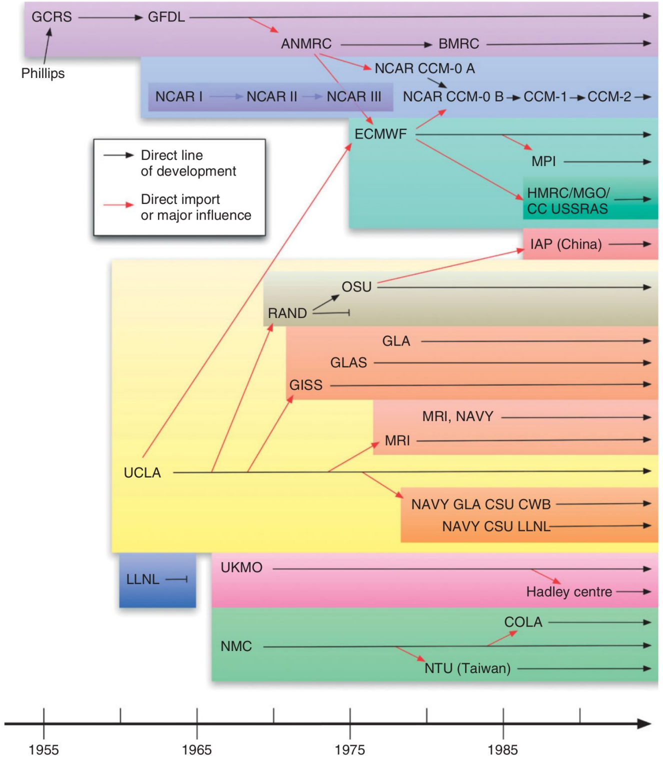
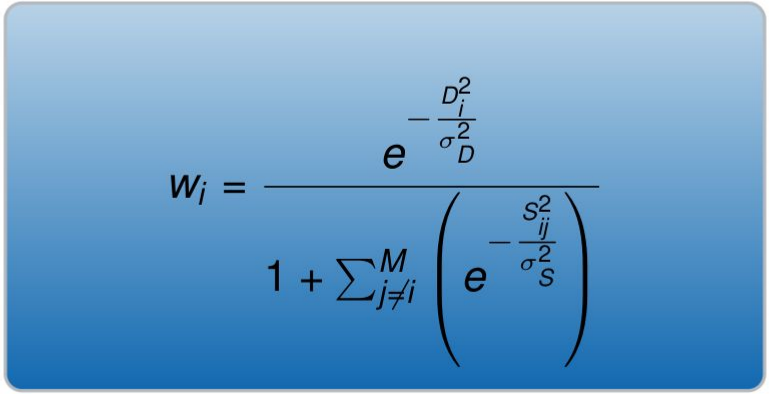


Figure: Development and dependencies for several climate models. Edwards (2010)

Putting it all together: calculation of model weights

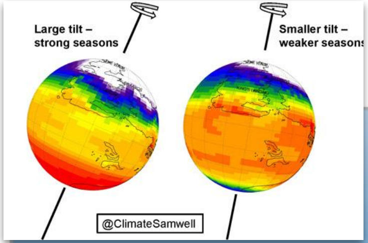


$$w_i = \frac{e^{-\frac{D_i^2}{\sigma_D^2}}}{1 + \sum_{j \neq i}^{M} \left(e^{-\frac{S_{ij}^2}{\sigma_S^2}} \right)}$$

Knutti et al. (2017)

- w_i : weight for model i
- D_i : generalised distance of model i to observations (performance diagnostics)
- σ_D : performance shape parameter
- M: number of models
- S_{ij} : generalised distance between model pair (independence diagnostics)
- σ_S : independence shape parameter

Weighting model by performance: Westeros



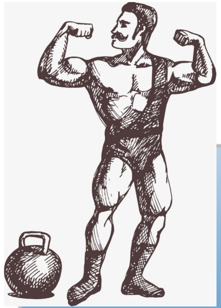
$$w_i = \frac{e^{-\frac{D_i^2}{\sigma_D^2}}}{1 + \sum_{j \neq i}^{M} \left(e^{-\frac{S_{ij}^2}{\sigma_S^2}} \right)}$$

D_i → ∞ (simulations from Westeros are pretty far away from observations on Earth)

$e^{-\infty} = 0 \rightarrow w_i = 0$

A model which simulates a climate very far away from what we observe gets weight zero.

Weighting model by performance: Super model



$$w_i = \frac{e^{-\frac{D_i^2}{\sigma_D^2}}}{1 + \cancel{\sum_{j \neq i}^{M} \left(e^{-\frac{S_{ij}^2}{\sigma_S^2}} \right)}}$$

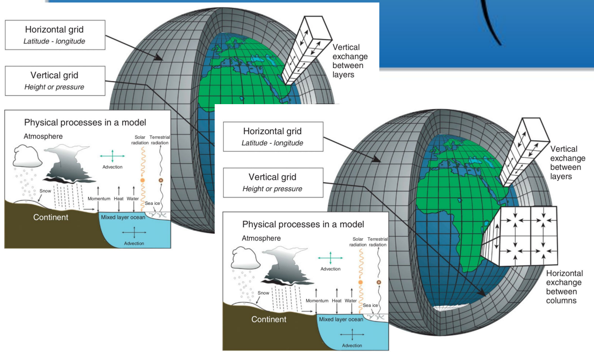
$D_i = 0$ (a super model is perfectly simulating observations on earth)

$e^{-0} = 1 \rightarrow w_i = 1$ (without independence)

A model which perfectly simulates Earth’s climate gets the highest weight of one.

Weighting model by independence: Two identical models

$$w_i = \frac{\cancel{e^{-\frac{D_i^2}{\sigma_D^2}}}}{1 + \sum_{j \neq i}^{M} \left(e^{-\frac{S_{ij}^2}{\sigma_S^2}} \right)}$$



S_{ij} = 0 (the distance of a model to a identical copy of itself is zero)

$e^{-0} = 1 \rightarrow w_i = 1/(1+1) = ½$ (without performance)

If a multi-model ensemble contains the same model twice both instances get only half of the weight.

Recap: Introduction

- Projections of future climate by climate models have three main sources of uncertainty:
  - emission scenario uncertainty
  - model uncertainty
  - internal variability
- Here I focus on **model uncertainty**
- Weighting to better quantify model uncertainty
  - accounting for model dependencies (**Part I**)
  - downweighting models which are not ‘fit for purpose’ (**Part II**)
- Finally I check if things improved (**Part III**)

Part I: Model Independence

Model independence weighting: basic assumption

Structural model similarity can be inferred from model output similarity

- Models with multiple **shared components** have **similar output** (e.g. temperature climatologies)
- We can check this by looking at models which we know are similar
- **Two variables are enough to cluster/separate models**

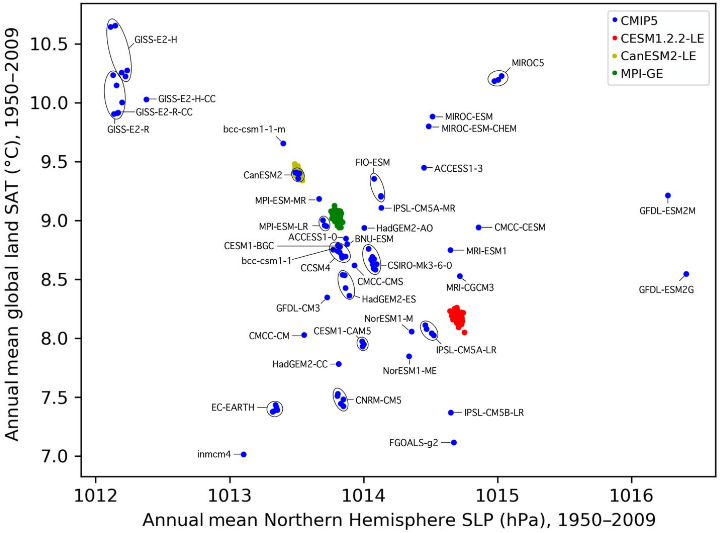


Figure: Clustering of CMIP5 models based on mean temperature and sea level pressure. Merrifield et al. (2020)

CMIP6 model family tree

- The tree structure on the right-hand side is only based on model output
- Model branching further to the left are closer to each other in output space

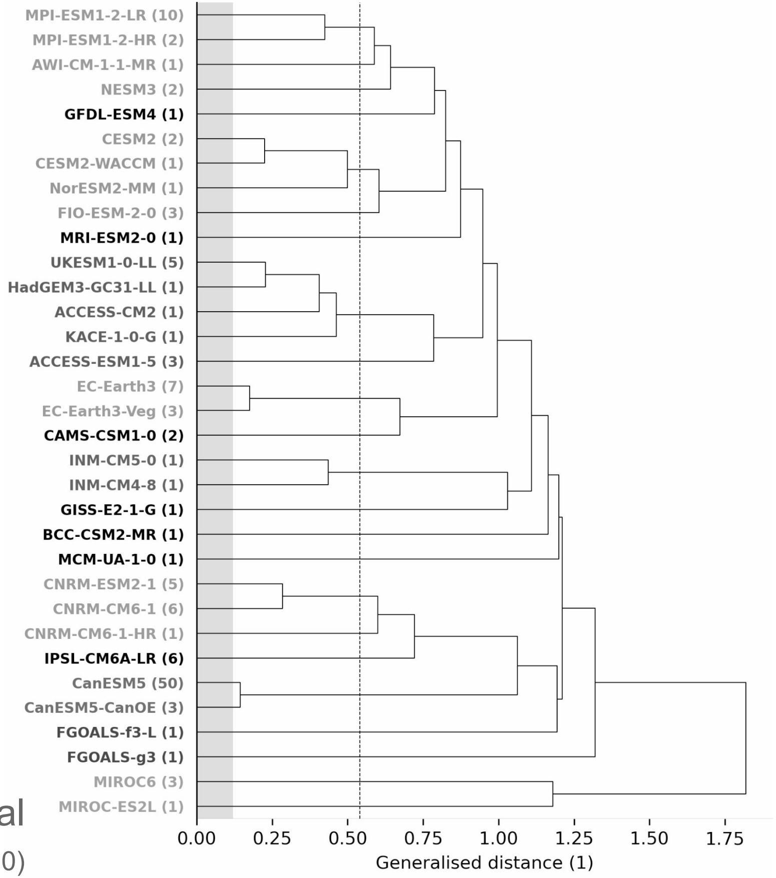


Figure: Model family tree for CMIP6, based on global temperature and sea level pressure. Brunner et al. (2020)

CMIP6 model family tree

- The tree structure on the right-hand side is only based on model output
- Model branching further to the left are closer to each other in output space
- Label colors based on expert knowledge of model components

→ **Models know to be similar are clustered together based on their output**

→ transfer generalised distance to independence weights (**shape parameter**)

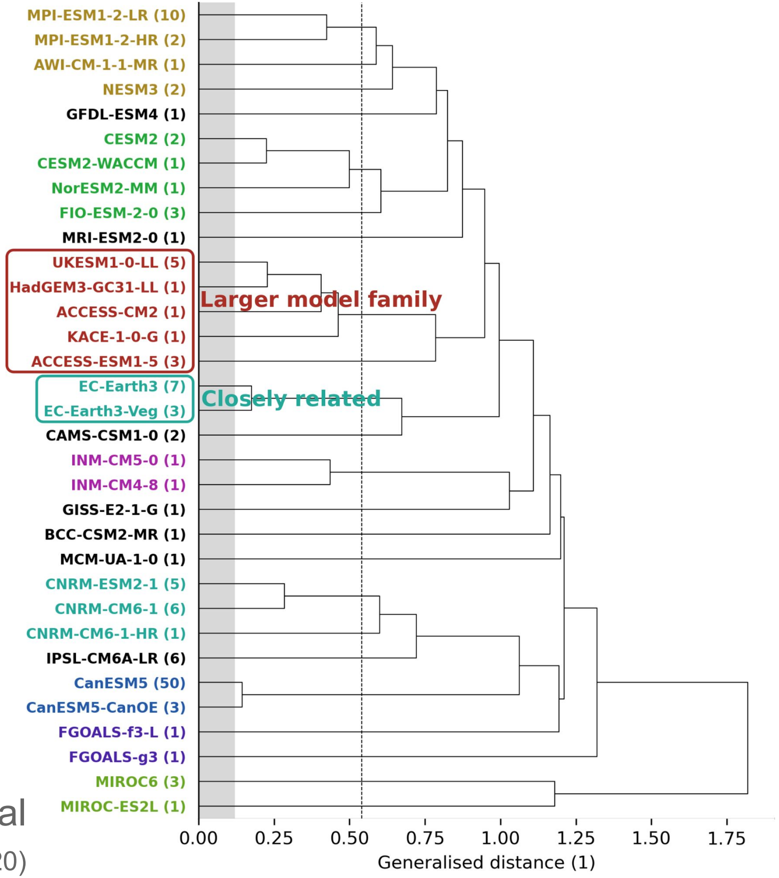


Figure: Model family tree for CMIP6, based on global temperature and sea level pressure. Brunner et al. (2020)

Part II: Model Performance

Model-observation distances

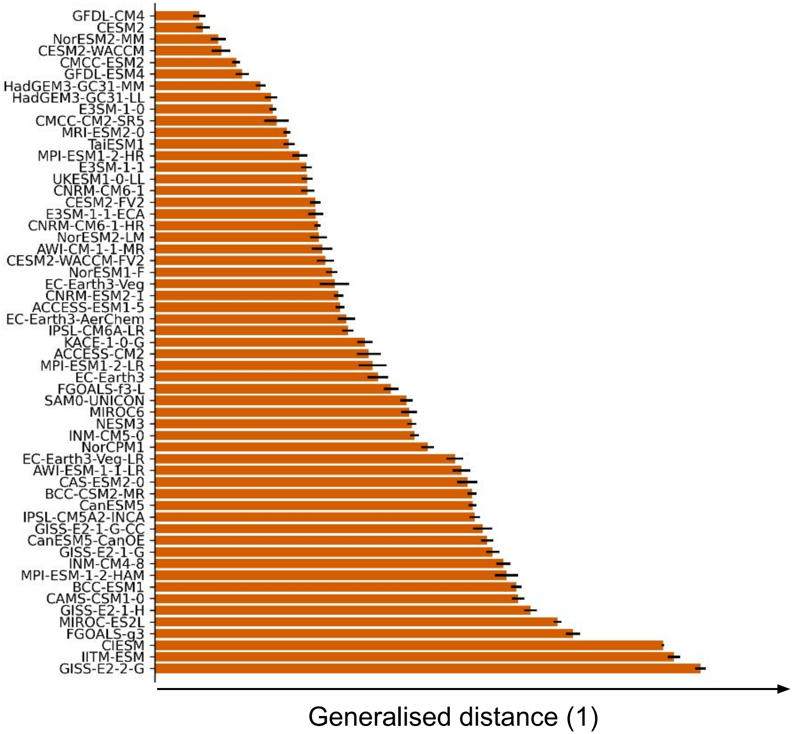
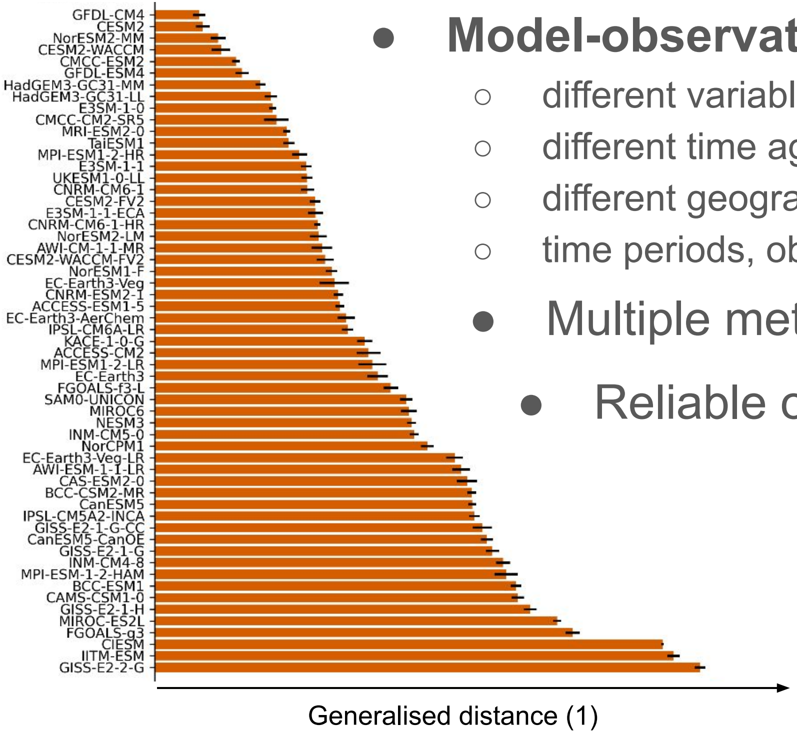


Figure: Generalized distance to observations (ERA5) for CMIP6 models. Based on 21-year climatology of temperature and precipitation. Brunner et al. (in prep)

Model-observation distances



- **Model-observation distance** can be based on
  - different variables (temperature, precipitation, sea level pressure, …)
  - different time aggregations (climatology, variability, trend)
  - different geographical regions (that can differ from the target region)
  - time periods, observational datasets, resolutions, etc.
- Multiple metrics can be combined (**generalized distance**)
- Reliable observations are needed as reference

Figure: Generalized distance to observations (ERA5) for CMIP6 models. Based on 21-year climatology of temperature and precipitation. Brunner et al. (in prep)

Model-observation distances

- **Model-observation distance** can be based on
  - different variables (temperature, precipitation, sea level pressure, …)
  - different time aggregations (climatology, variability, trend)
  - different geographical regions (that can differ from the target region)
  - time periods, observational datasets, resolutions, etc.
- Multiple metrics can be combined (**generalized distance**)
- Reliable observations are needed as reference
- Weighting: metrics should be **relevant for the target**

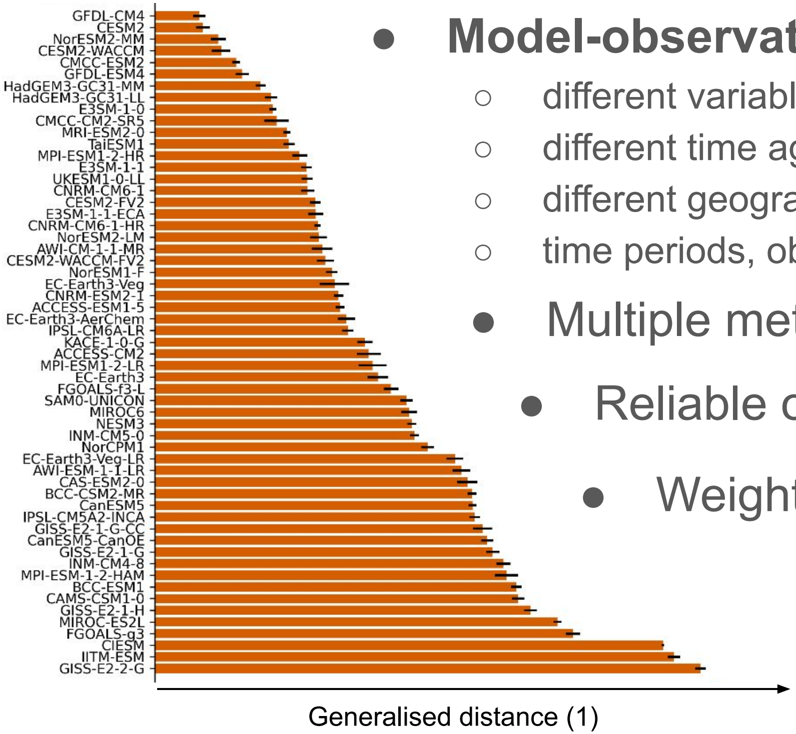
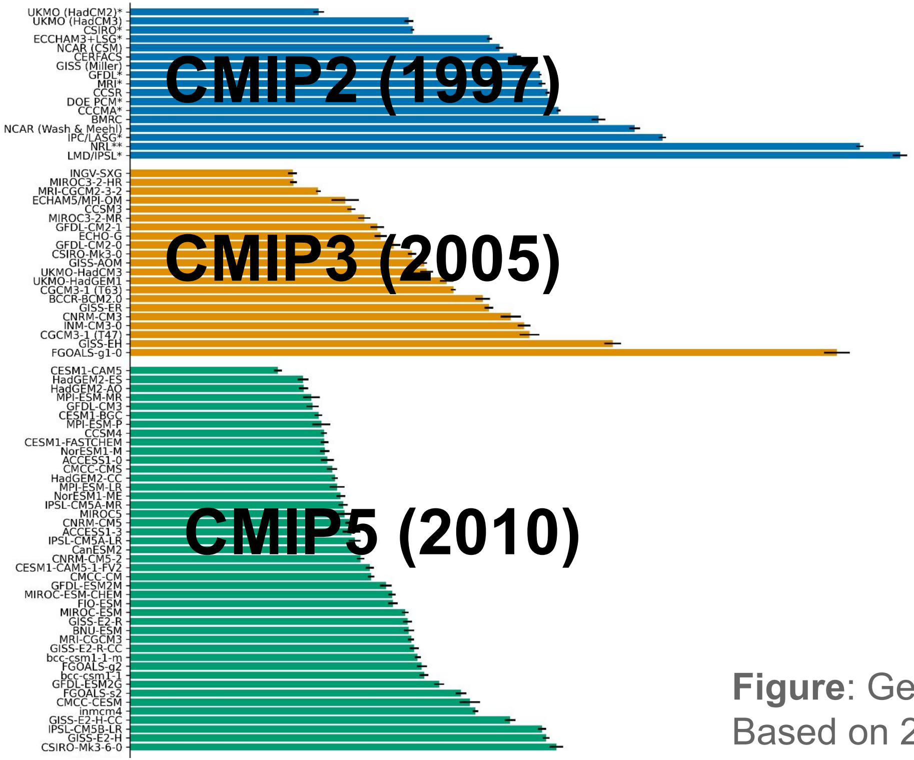


Figure: Generalized distance to observations (ERA5) for CMIP6 models. Based on 21-year climatology of temperature and precipitation. Brunner et al. (in prep)

Detour: Distances across CMIP generations



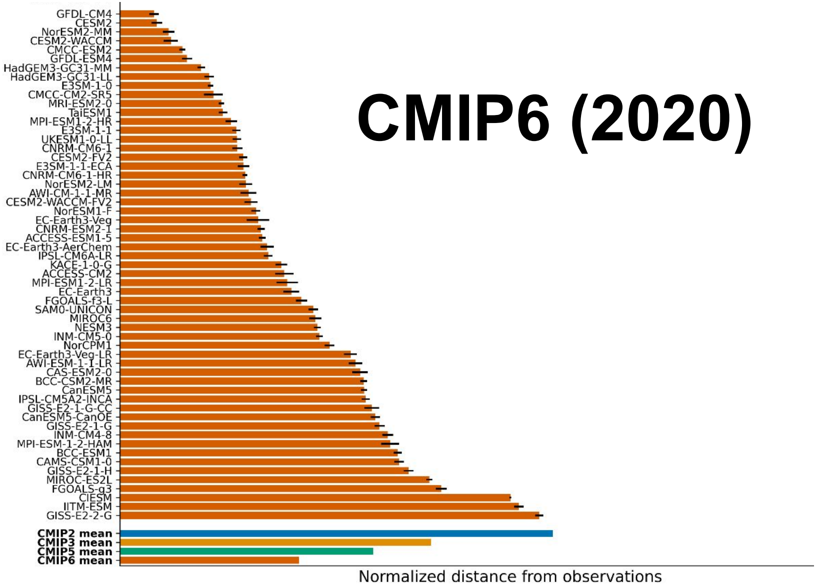


Figure: Generalized distance to observations (ERA5). Based on 21-year climatology of temperature and precipitation. Brunner et al. (in prep)

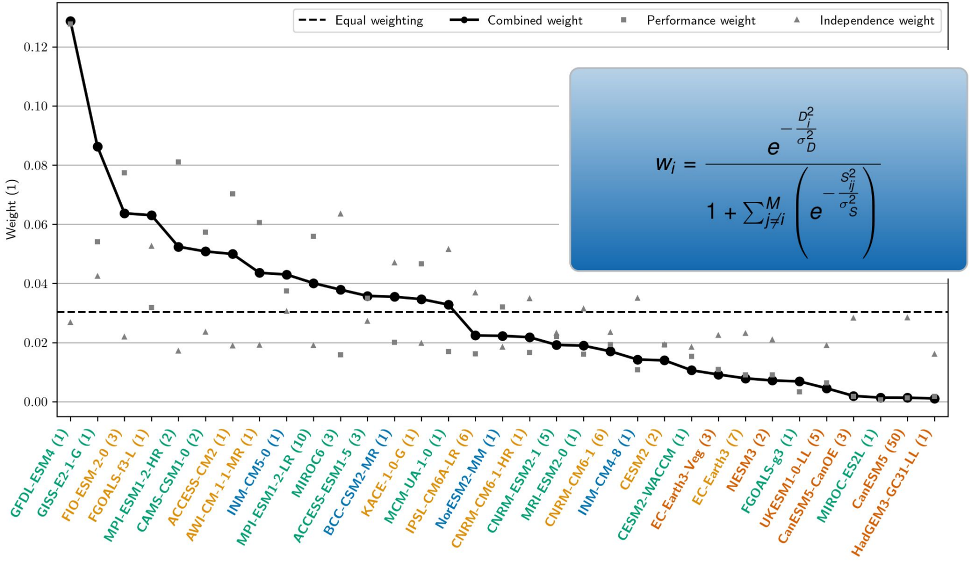
Please don’t share

Translating distances to weights: shape parameter

The **shape parameter** σ_D needs to be carefully chosen to provide confident and meaningful weights

- small values lead to strong, selecting only a few models
- large values lead to equal weighting

→ **model-as-truth test**



$$w_i = \frac{e^{-\frac{D_i^2}{\sigma_D^2}}}{1 + \sum_{j \neq i}^{M} \left(e^{-\frac{S_{ij}^2}{\sigma_S^2}} \right)}$$

Figure: Weights for 33 CMIP6 models based on **five performance** and **two independence metrics** chosen for weighting global temperature. Brunner et al. (2020a)

Effect of weighting CMIP6 projections of future climate

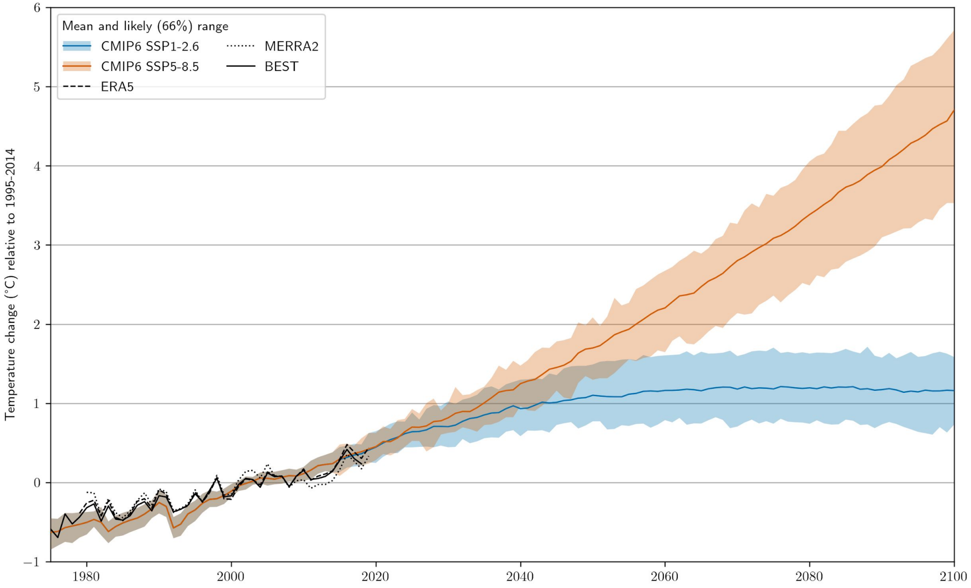


Figure: Global mean, annual mean temperature change (relative to 1995-2014) from 33 CMIP6. Brunner et al. (2020a)

Effect of weighting CMIP6 projections of future climate

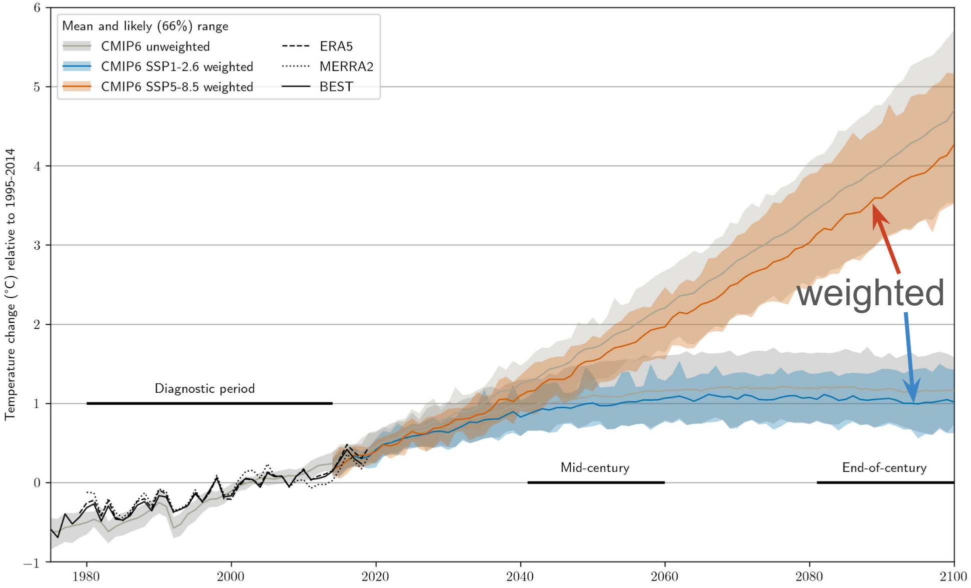


Figure: Weighted global mean, annual mean temperature change (relative to 1995-2014) from 33 CMIP6 models.
Brunner et al. (2020a)

- The weighted distribution shows **reduced mean warming from CMIP6** models broadly consistent with other studies
  - Nijsse et al. (2020)
  - Tokarska et al. (2020)
  - Ribes et al. (2021)
- **Reduction of uncertainty** by 10%-20% for the likely range due to a constraining of the upper percentiles

Recap: Performance and independence weighting

- Using the model range directly as uncertainty range disregards that
  - not all model are independent
  - not all models are qualy ‘fit for purpose’
- Model weighting can help to account for that
- Distances are translated into weights assuming that
  - model similarity can be inferred from output similarity
  - future model performance can be inferred from past model performance
- The translation from distances to weights is done via two shape parameters

Part III: Does weighting improve future projections?

Measuring the benefit of weighting climate models

From weather forecasting: “What Is a Good Forecast?” (Murphy 1993)

- **Accuracy**: level of agreement between forecast and truth
- **Skill**: accuracy relative to a reference forecast
- **Reliability**: average agreement between forecasts and truth
- **Sharpness**: tendency of the forecast to predict specific values (counter-example: the climatology has no sharpness)

Quality

- **Consistency**: forecast is consistent with prior knowledge
- **Value**: degree to which the forecast helps decision makers

Measuring the benefit of weighting climate models

What Is a Good Weighting? - **we don’t know the ‘truth’**

✘ **Accuracy**: level of agreement between **weighted projection** and ‘truth’

✘ **Skill**: accuracy relative to the **unweighted projection**

✘ **Reliability**: average agreement between **weighted projections** and ‘truth’

✔ **Sharpness**: tendency of the **weighted projections** to reduce model uncertainty compared to the **unweighted projections**

✔ **Consistency**: is **weighting** consistent with other methods

✔ **Value**: degree to which the **weighted projection** helps users

Measuring the benefit of weighting climate models

What Is a Good Weighting? - we don’t know the ‘truth’

- ✔ **Sharpness**: determined by the performance shape parameter σ_D: smaller σ_D leads to sharper results but might no longer be **reliable**
- ✔ **Value**: determined by the users

- ✔ **Consistency**: **quantify** by comparing methods using a **common setup** (Brunner et al. 2020b, Hegerl et al. 2021, O’Reilly et al. in prep.)
- ✔ **Accuracy, Skill, Reliability**: we don’t know the true climate in the future and there will be only one realisation → **model-as-truth approach**

Consistency: comparing methods to constrain projections

No **coordinated framework** to compare methods exist. They might differ for a range of reasons independent of the methods itself:

- variable (temperature vs precip)
- region (global vs Europe)
- season and time period
- models included
- uncertainties included
- ...

15 May 2002 GIORGI AND MEARNS 1141

Calculation of Average, Uncertainty Range, and Reliability of Regional Climate Changes from AOGCM Simulations via the "Reliability Ensemble Averaging" (REA) Method

Filippo Giorgi

1 March 2007 KETTLEBOROUGH ET AL. 843

Estimates of Uncertainty in Predictions of Global Mean Surface Temperature

J. A. Kettleborough

British Atmospheric Data Centre, Rutherford Appleton Laboratory, Chilton, United Kingdom

B. B. B. Booth

Hadley Centre for Climate Prediction and Research, Met Office, Exeter, United Kingdom

1 July 2015 SANDERSON ET AL. 5171

A Representative Democracy to Reduce Interdependency in a Multimodel Ensemble

Benjamin M. Sanderson

National Center for Atmospheric Research,* Boulder, Colorado

Reto Knutti

Clim Dyn (2017) 48:367–386
DOI 10.1007/s00382-016-3079-6

CrossMark

A new statistical approach to climate change detection and attribution

Aurélien Ribes[1] · Francis W. Zwiers[2] · Jean-Marc Azaïs[3] · Philippe Naveau[4]

Geosci. Model Dev., 10, 2379–2395, 2017
https://doi.org/10.5194/gmd-10-2379-2017
© Author(s) 2017. This work is distributed under the Creative Commons Attribution 3.0 License.

Geoscientific Model Development

Skill and independence weighting for multi-model assessments

Benjamin M. Sanderson[1], Michael Wehner[2], and Reto Knutti[3,1]

AGU100 ADVANCING EARTH AND SPACE SCIENCE

JAMES Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2019MS001639

Special Section:
The Max Planck Institute for Meteorology Earth System Model version 1.2

The Max Planck Institute Grand Ensemble: Enabling the Exploration of Climate System Variability

Nicola Maher[1], Sebastian Milinski[1,2], Laura Suarez-Gutierrez[1,2], Michael Botzet[1], Mikhail Dobrynin[3,4], Luis Kornblueh[1], Jürgen Kröger[1], Yohei Takano[1], Rohit Ghosh[1], Christopher Hedemann[1], Chao Li[1], Hongmei Li[1], Elisa Manzini[1], Dirk Notz[1], Dian Putrasahan[1], Lena Boysen[1], Martin Claussen[1,3], Tatiana Ilyina[1], Dirk Olonscheck[1], Thomas Raddatz[1], Bjorn Stevens[1], and Jochem Marotzke[1]

https://doi.org/10.5194/esd-2019-69
Preprint. Discussion started: 21 November 2019
© Author(s) 2019. CC BY 4.0 License.

Earth System Dynamics Discussions

A weighting scheme to incorporate large ensembles in multi-model ensemble projections

Anna L. Merrifield[1], Lukas Brunner[1], Ruth Lorenz[1], and Reto Knutti[1]

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

Correspondence: Anna L. Merrifield (anna.merrifield@env.ethz.ch)

Abstract. Multi-model ensembles can be used to estimate uncertainty in projections of regional climate, but this uncertainty often depends on the constituents of the ensemble. The dependence of uncertainty on ensemble composition is clear when single model initial condition large ensembles (SMILEs) are included within a multi-model ensemble. SMILEs introduce

SCIENTIFIC REPORTS
natureresearch

OPEN Future continental summer warming constrained by the present-day seasonal cycle of surface hydrology

F. M. Selten[1*], R. Bintanja[1,2], R. Vautard[3] & B. J. J. M. van den Hurk[4,5]

Present-day land temperatures simulated by state-of-the-art global climate models exhibit considerable uncertainty. Generally it is assumed that these temperature biases do not affect the

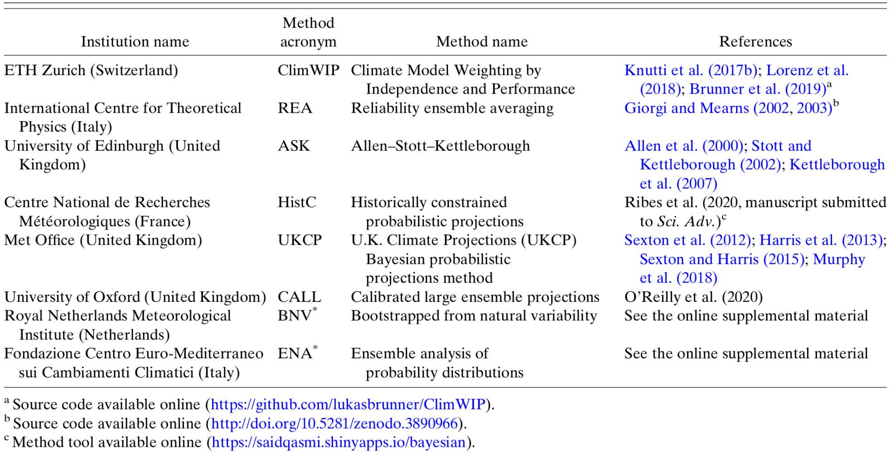
Figures: Comparing (top) methods and (right) apples and oranges right: CC-BY M. Johnson

A consistent framework for method comparison

We brought together **8 groups** working on constraining and developed a **level playing field for comparison**

2 conditions for participation:

1. quantify uncertainty in future projections
2. able to handle common settings



Institution name	Method acronym	Method name	References
ETH Zurich (Switzerland)	ClimWIP	Climate Model Weighting by Independence and Performance	Knutti et al. (2017b); Lorenz et al. (2018); Brunner et al. (2019)[a]
International Centre for Theoretical Physics (Italy)	REA	Reliability ensemble averaging	Giorgi and Mearns (2002, 2003)[b]
University of Edinburgh (United Kingdom)	ASK	Allen–Stott–Kettleborough	Allen et al. (2000); Stott and Kettleborough (2002); Kettleborough et al. (2007)
Centre National de Recherches Météorologiques (France)	HistC	Historically constrained probabilistic projections	Ribes et al. (2020, manuscript submitted to *Sci. Adv.*)[c]
Met Office (United Kingdom)	UKCP	U.K. Climate Projections (UKCP) Bayesian probabilistic projections method	Sexton et al. (2012); Harris et al. (2013); Sexton and Harris (2015); Murphy et al. (2018)
University of Oxford (United Kingdom)	CALL	Calibrated large ensemble projections	O'Reilly et al. (2020)
Royal Netherlands Meteorological Institute (Netherlands)	BNV*	Bootstrapped from natural variability	See the online supplemental material
Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (Italy)	ENA*	Ensemble analysis of probability distributions	See the online supplemental material

[a] Source code available online (https://github.com/lukasbrunner/ClimWIP).
[b] Source code available online (http://doi.org/10.5281/zenodo.3890966).
[c] Method tool available online (https://saidqasmi.shinyapps.io/bayesian).

Table: Participating institutions, methods, and references. Brunner et al. (2020b)

Comparing future Central European temperature change

- Trade-off between number of methods and the **fairness of the comparison**
- **Fairest** comparison:
  4/8 methods could participate
- All methods **narrow the uncertainty** range
- All methods agree on slightly **less warming**

→ not all cases look that nice

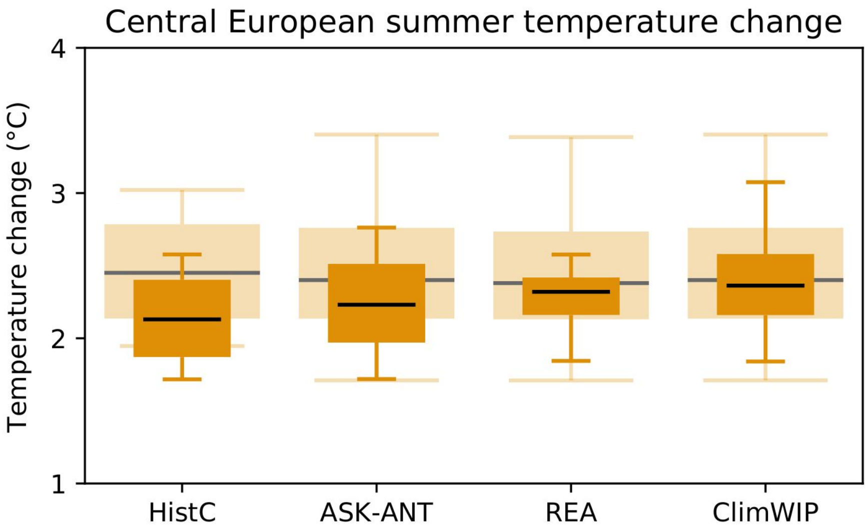


Figure: Unconstrained (light) and constrained (dark) Central European summer temperature change (2041-60 relative to 1995-2014) from CMIP5. Brunner et al. (2020b)

Take home messages

- **Uncertainty in projections of future climate** comes from
  - emission scenario uncertainty
  - climate model uncertainty
  - internal variability
- **Model spread** can be translated to **model uncertainty** but
  - **not all models are independent** estimates of the future
  - **not all models are equally ‘fit for purpose’**
- **Model weighting** can help to account for this
- Weighting is consistent with other methods

Thank you for your attention!
Questions?

Literature

- Brunner, L., Lorenz, R., Zumwald, M., & Knutti, R. (2019). Quantifying uncertainty in European climate projections using combined performance-independence weighting. Environmental Research Letters, 14(12), 124010. https://doi.org/10.1088/1748-9326/ab492f
- Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., & Knutti, R. (2020a). Reduced global warming from CMIP6 projections when weighting models by performance and independence. Earth System Dynamics, 11(4), 995–1012. https://doi.org/10.5194/esd-11-995-2020
- Brunner, L., McSweeney, C., Ballinger, A. P., Befort, D. J., Benassi, M., Booth, B., Coppola, E., de Vries, H., Harris, G., Hegerl, G. C., Knutti, R., Lenderink, G., Lowe, J., Nogherotto, R., O'Reilly, C., Qasmi, S., Ribes, A., Stocchi, P., & Undorf, S. (2020). Comparing Methods to Constrain Future European Climate Projections Using a Consistent Framework. Journal of Climate, 33(20), 8671–8692. https://doi.org/10.1175/JCLI-D-19-0953.1
- Brunner, L. et al. (in preparation): Evolution of climate models in past, present, and future
- Edwards, P. N. (2011). History of climate modeling. Wiley Interdisciplinary Reviews: Climate Change, 2(1), 128–139. https://doi.org/10.1002/wcc.95
- Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., Knutti, R., & Hawkins, E. (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. Earth System Dynamics, 11(2), 491–508. https://doi.org/10.5194/esd-11-491-2020
- 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. https://doi.org/10.1002/2016GL072012
- 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. Cryosphere, 6(6), 1383–1394. https://doi.org/10.5194/tc-6-1383-2012
- 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. https://doi.org/10.5194/esd-11-807-2020

Bonus Slides

Distribution of uncertainty

- The contribution of each source of **uncertainty depends on various parameters**:
  - lead time
  - variable
  - region
  - models used
- Scenario uncertainty can be eliminated by making projections conditional to a scenario
- Internal variability can, for example, be investigated using so-called SMILEs
- Leaves us with model uncertainty...

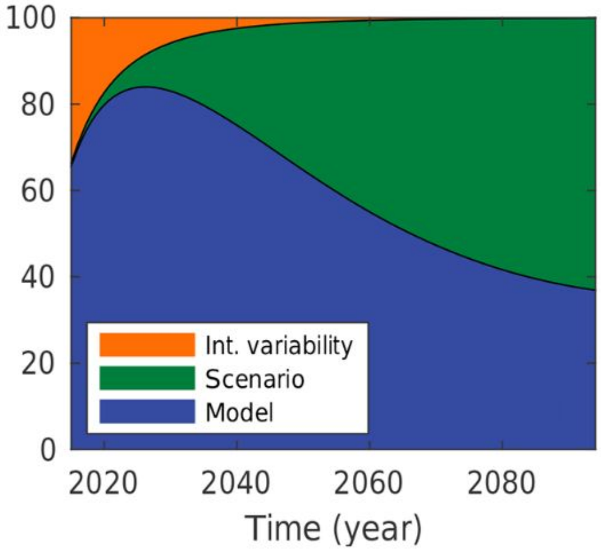


Figure: Fractional contribution to total uncertainty for 10-year running mean of global mean, annual mean temperature from CMIP6. Lehner et al. (2020)

Distribution of uncertainty - dependence on region

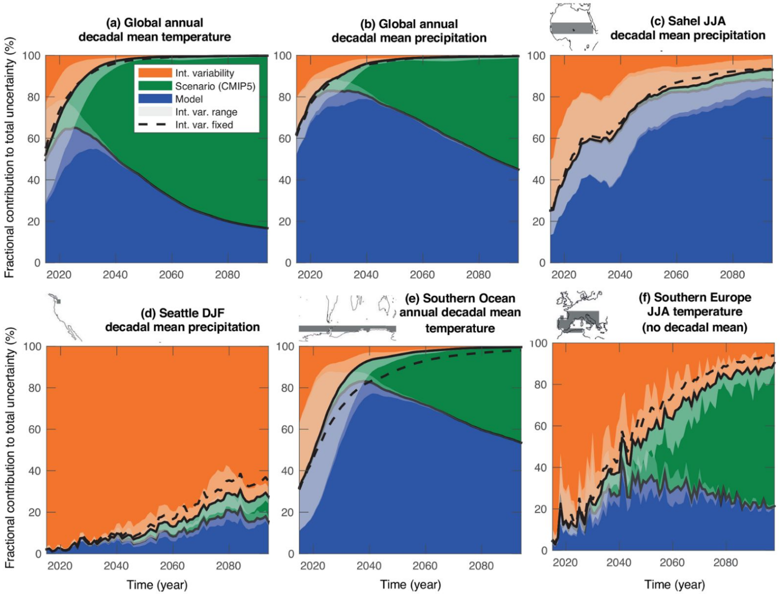


Figure 7. Sources of uncertainty from SMILEs (using scenario uncertainty from CMIP5) for different regions, seasons and variables. The solid black lines indicate the borders between sources of uncertainty; the slightly transparent white shading around those lines is the range of this estimate based on different SMILEs. The dashed line marks the dividing line if internal variability is assumed to stay fixed at its 1950–2014 multi-SMILE mean. All panels are for decadal mean projections, except **(f)** southern Europe June–August temperature, to which no decadal mean has been applied.

Weighting regional summer temperature

- **The effect of the weighting depends on the case**
- Different set of models and relevant metrics for this case!
- For Mediterranean summer the weighted distribution shows **stronger warming** from CMIP5
- The interquartile range is reduced by about 24% by the end of the century

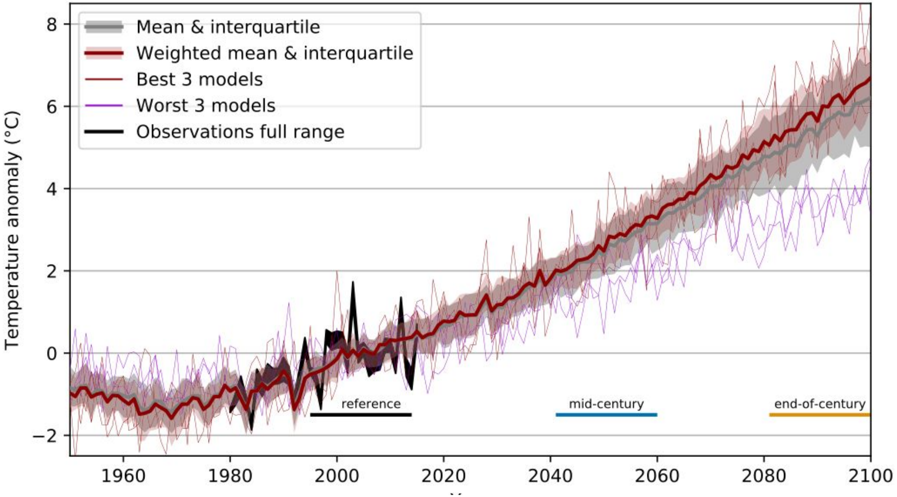


Figure: Weighted **Mediterranean summer temperature** anomaly (relative to 1995-2014) based on 37 CMIP5 models (79 realizations).

Brunner et al. (2019)

A common framework for method comparison

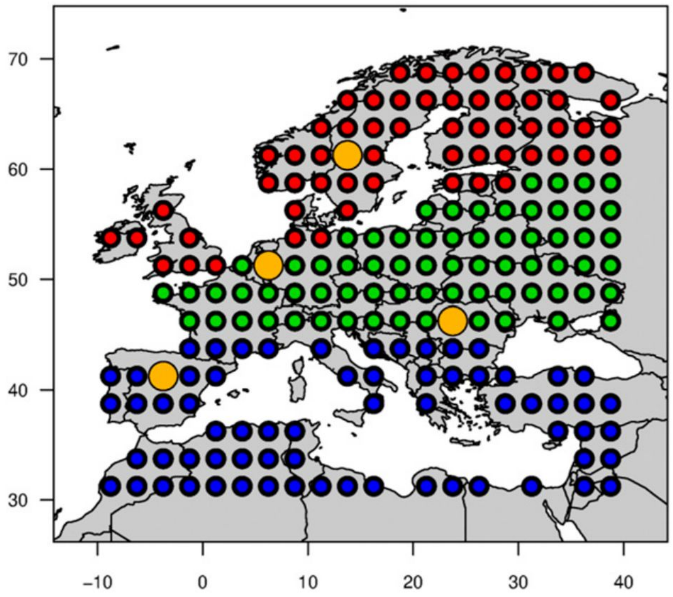


Figure: Resolution, regions, and sea mask. Brunner et al. (2020b)

We brought together **8 groups** working on constraining and developed shared settings:

- temperature and precipitation
- Europe, SREX regions, 4 grid points
- summer (June, July, August)
- change in 2041-60 relative to 1995-2014
- CMIP5 models under RCP8.5
- same model pool (if possible)
- including 20-year internal variability
- median, 50%, and 80% range
- 2.5x2.5 horizontal resolution

Projections for Central European summer temperature

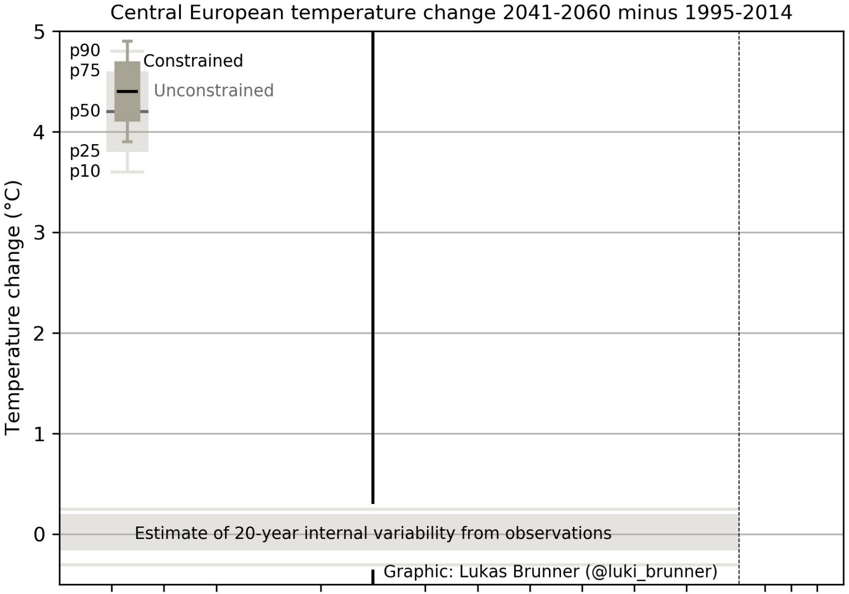


Figure: Unconstrained and constrained change in Central European summer temperature.
Brunner et al. (2020b)

Projections for Central European summer temperature

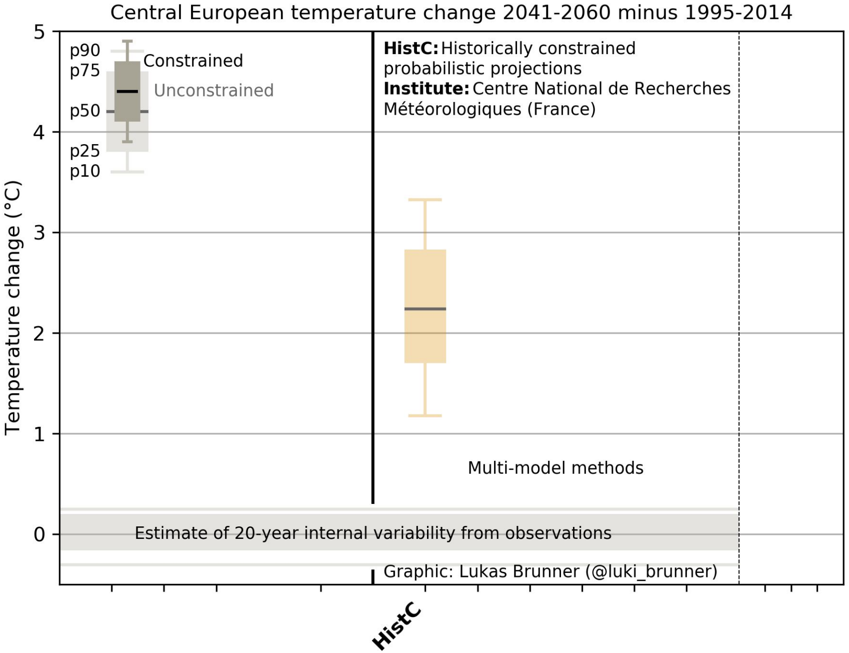


Figure: Unconstrained and constrained change in Central European summer temperature.
Brunner et al. (2020b)

Projections for Central European summer temperature

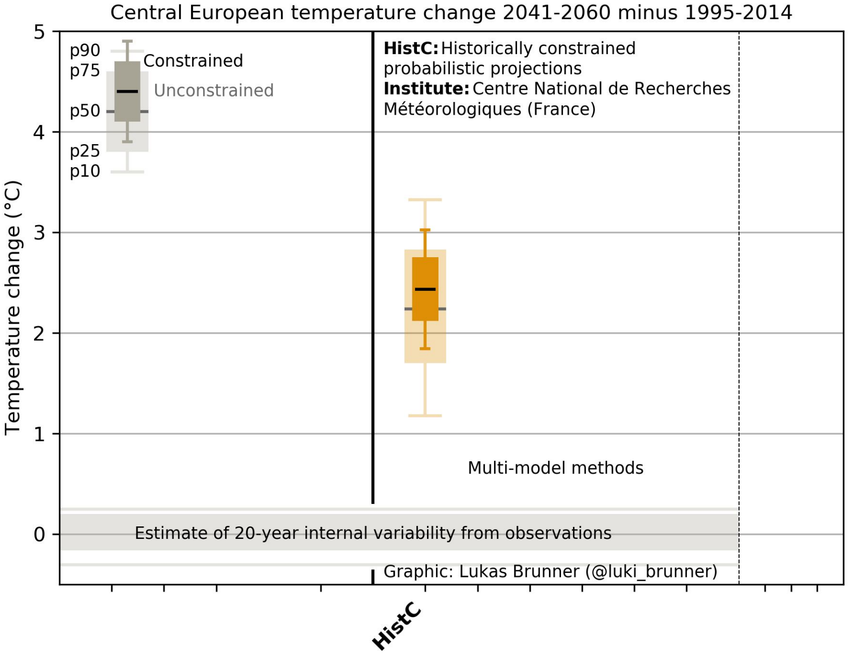


Figure: Unconstrained and constrained change in Central European summer temperature.
Brunner et al. (2020b)

Projections for Central European summer temperature

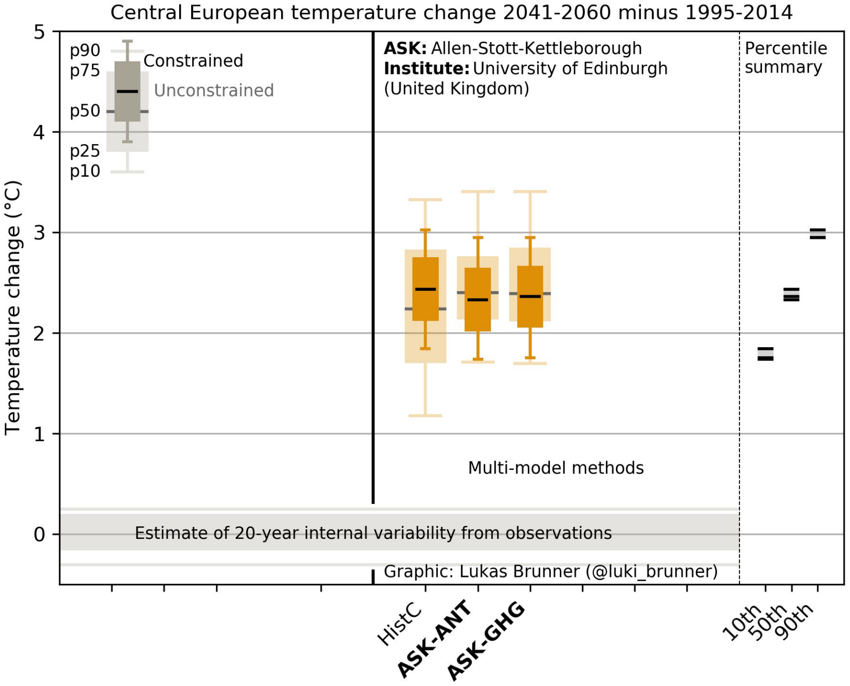


Figure: Unconstrained and constrained change in Central European summer temperature.
Brunner et al. (2020b)

Projections for Central European summer temperature

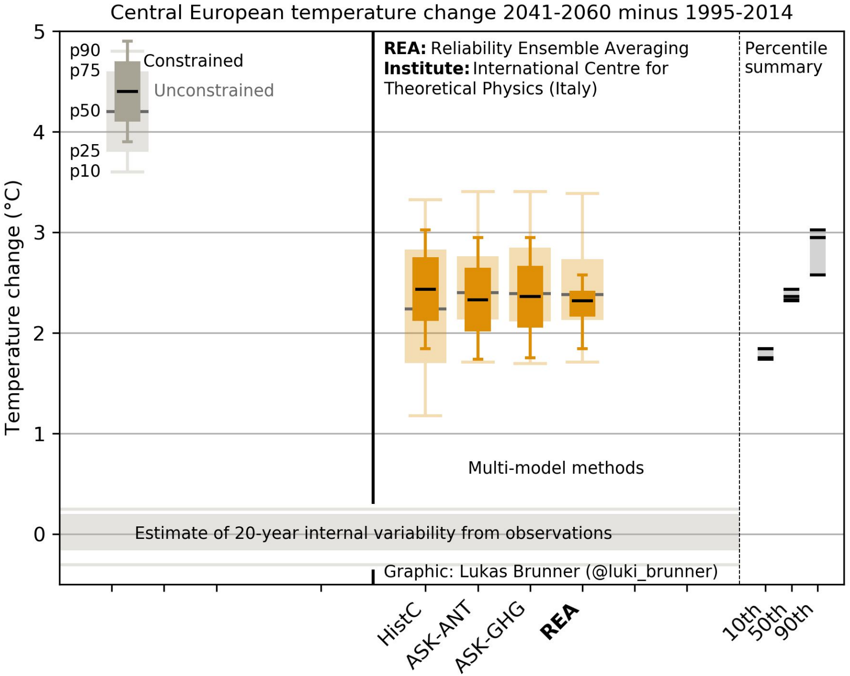


Figure: Unconstrained and constrained change in Central European summer temperature.
Brunner et al. (2020b)

Projections for Central European summer temperature

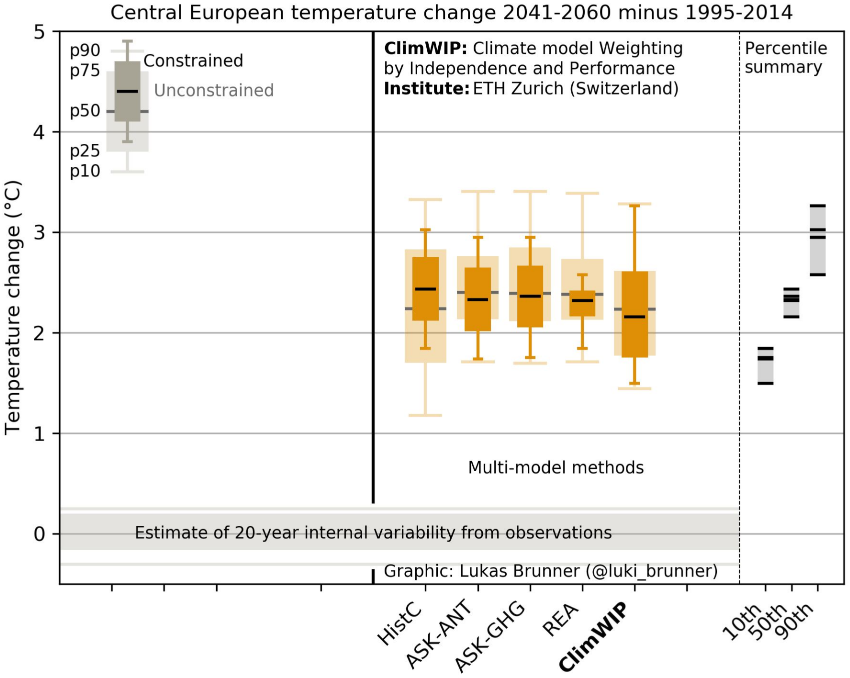


Figure: Unconstrained and constrained change in Central European summer temperature.
Brunner et al. (2020b)

Projections for Central European summer temperature

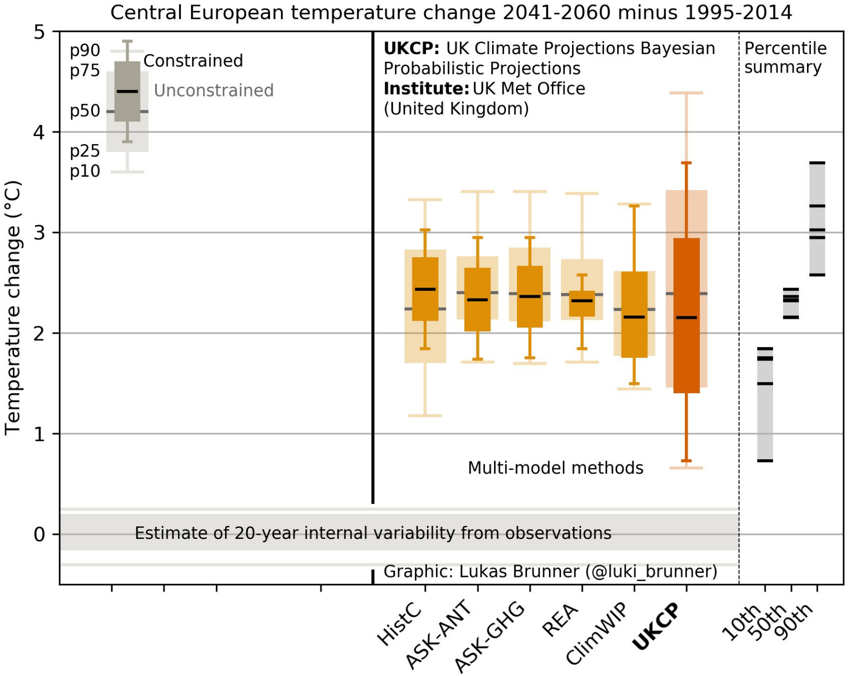
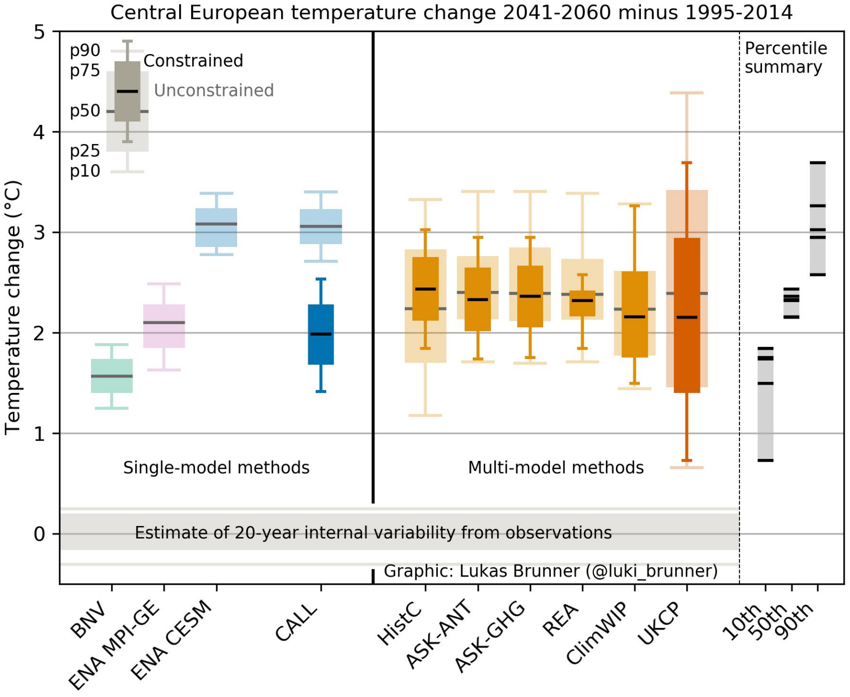


Figure: Unconstrained and constrained change in Central European summer temperature.
Brunner et al. (2020b)

Projections for Central European summer temperature



- Most methods show a slightly lower constrained median warming
- Most methods show a reduction in model spread
- More agreement in the central estimate than in extremes
- **Different models used**: unconstrained distributions differ

Figure: Unconstrained and constrained change in Central European summer temperature.
Brunner et al. (2020b)

Central European temperature - same model subset

- Using the same 29 models
- Excludes some methods but starting distributions are now almost identical
- **Methods consistently narrow the uncertainty range and agree on slightly less warming** → not all cases look that nice

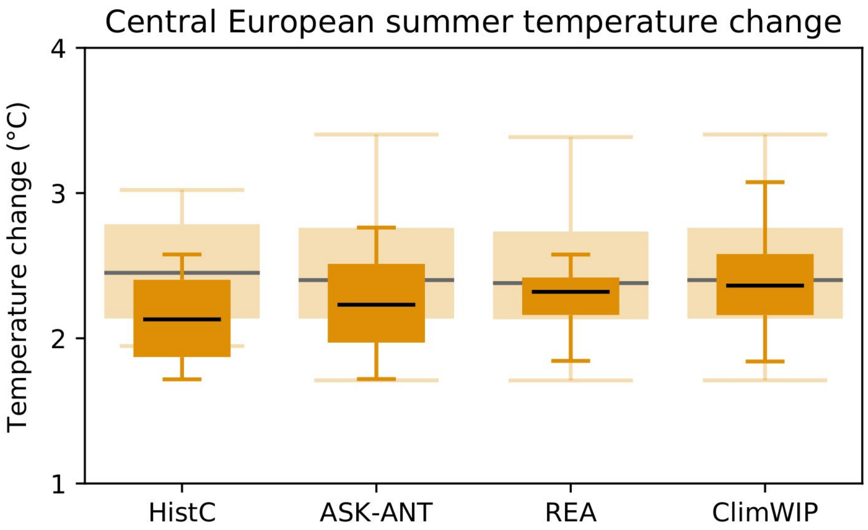
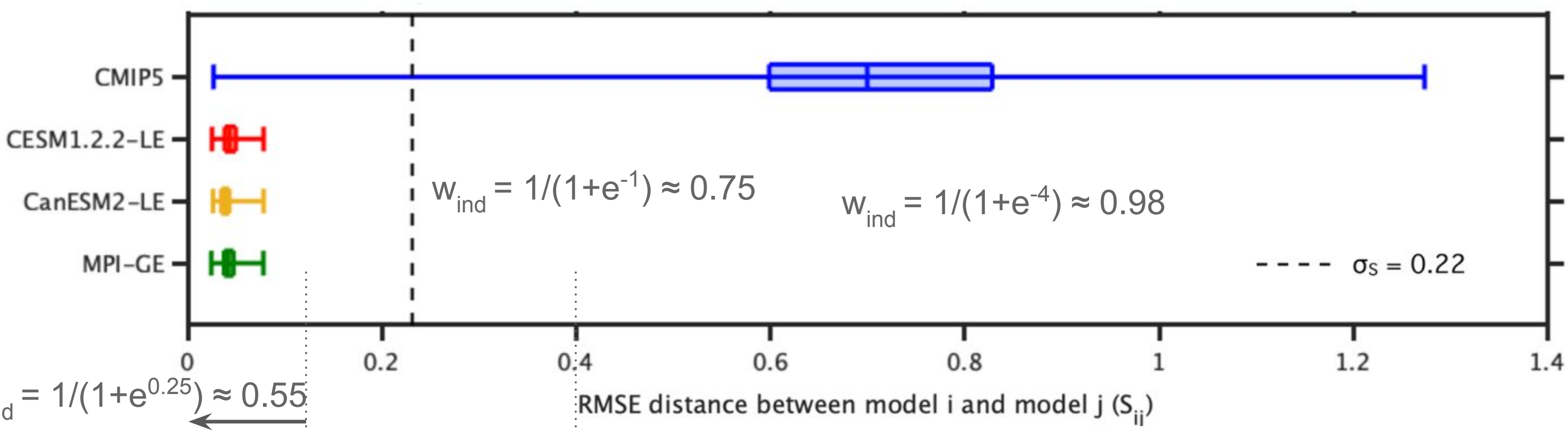


Figure: Unconstrained and constrained change in Central European summer temperature.
Brunner et al. (2020b)

Selection of the independence shape parameter



$w_{ind} = 1/(1+e^{-1}) \approx 0.75$

$w_{ind} = 1/(1+e^{-4}) \approx 0.98$

$w_{ind} = 1/(1+e^{0.25}) \approx 0.55$

Identical models (different only due to internal variability)

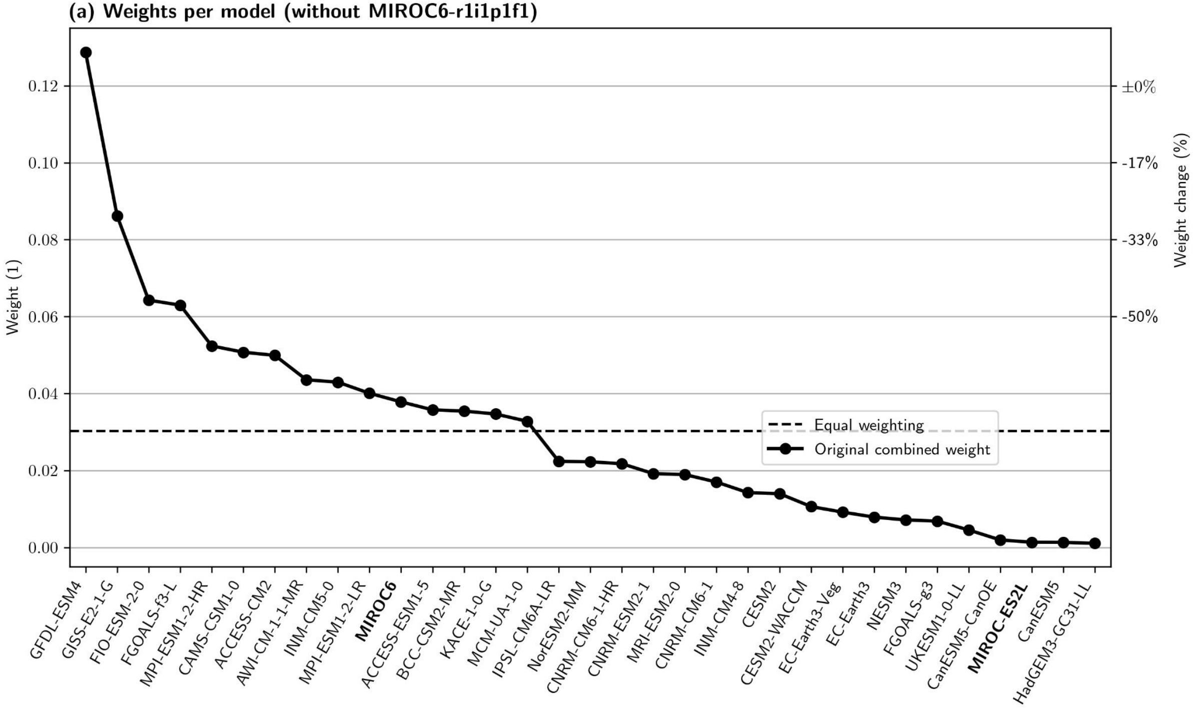
Partly dependent models

Mostly independent models

Merrifield et al. (2020)

$$w_i = \frac{e^{-\frac{D_i^2}{\sigma_D^2}}}{1 + \sum_{j \neq i}^{M} \left(e^{-\frac{S_{ij}^2}{\sigma_S^2}} \right)}$$

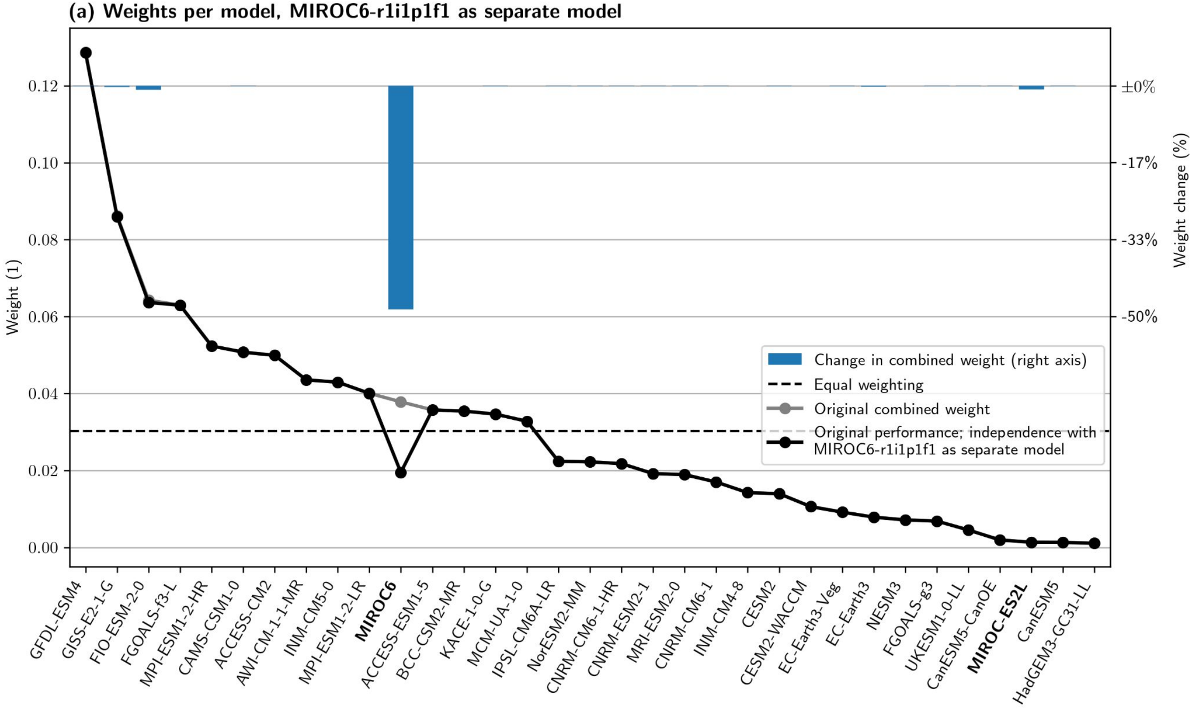
Validation of the independence weighting



How do the weights change if we add a model identical to a model already in the ensemble (apart from internal variability)?

Figure: Combined performance-independence weights, one member of MIROC6 withheld (Brunner et al. 2020.)

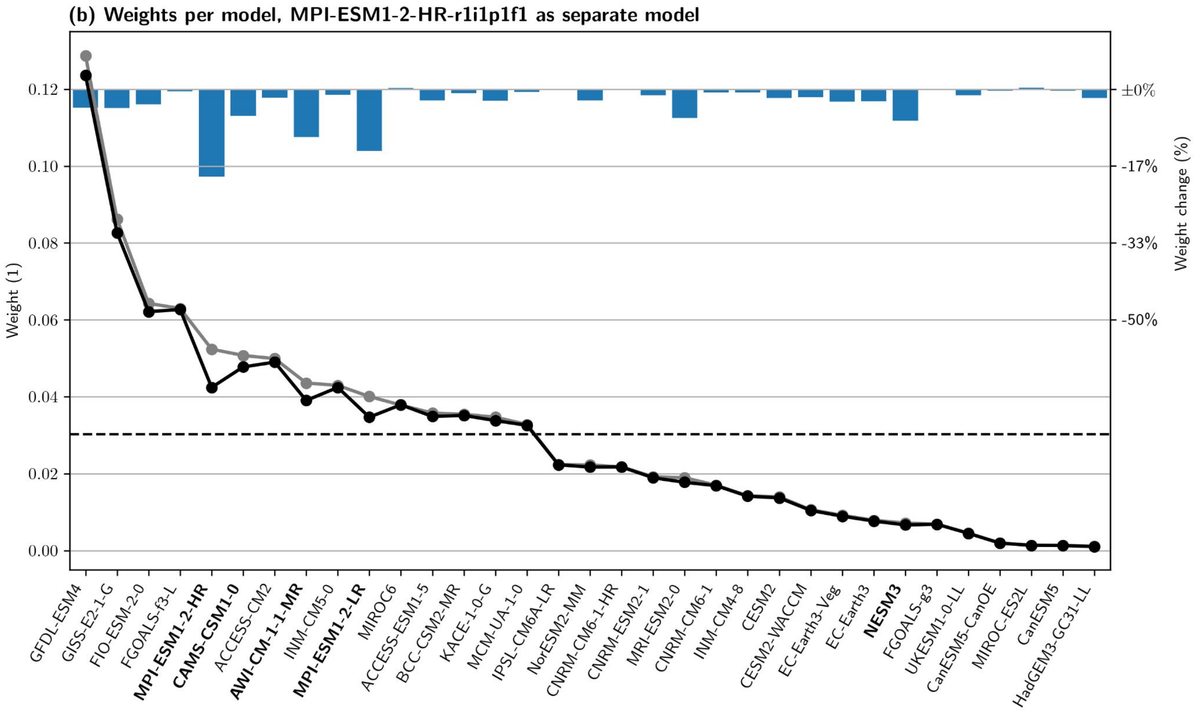
Validation of the independence weighting


(a) Weights per model, MIROC6-r1i1p1f1 as separate model

How do the weights change if we add a model identical to a model already in the ensemble (apart from internal variability)?

Figure: Combined performance-independence weights, one member of MIROC6 as separate model (Brunner et al. 2020.)

Validation of the independence weighting



Adding a model already in the ensemble can also affect multiple other models (if they are also dependent on it.

Figure: Combined performance-independence weights, one member of MPI as separate model (Brunner et al. 2020.)

Model-observation distance metrics I

For each model the weight is calculated based on metrics defined as the euclidean distance for

- a variable (e.g., temperature, precipitation, etc.)
- in a period (e.g., 1979-2020)
- aggregated over time (e.g., mean,trend, etc.)
- in a region (e.g., global, Central Europe, etc.)
- with or without land/sea mask
- to a reference (one or multiple observations)

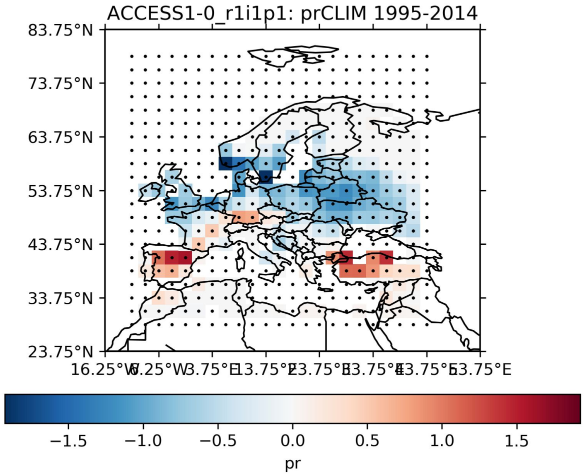
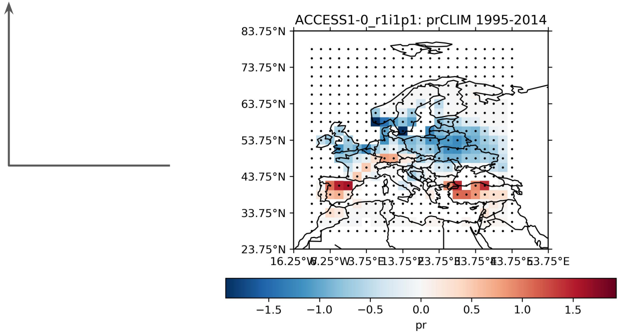


Figure: Example of a model-observation difference for the time mean precipitation in the period 1995-2014 over European land. For the metric the field still needs to be aggregated to a single value.

Model-observation distance metrics II

- For each model-observation pair d(lat, lon) the **area-weighted root-mean-square error** is calculated

$$\Delta = \sqrt{\frac{\sum_{\text{lat}} \sum_{\text{lon}} w(\text{lat}, \text{lon})\, d(\text{lat}, \text{lon})^2}{\sum_{\text{lat}} \sum_{\text{lon}} w(\text{lat}, \text{lon})}}$$


ACCESS1-0_r1i1p1: prCLIM 1995-2014

- If different metrics *k* are used they are normalised and combined and can be assigned different importance

$$\Delta_{\text{total}} = \frac{1}{\sum_k w_k} \sum_k \frac{w_k \Delta_k}{\overline{\Delta_k}}$$

Skill of the weighting for two examples

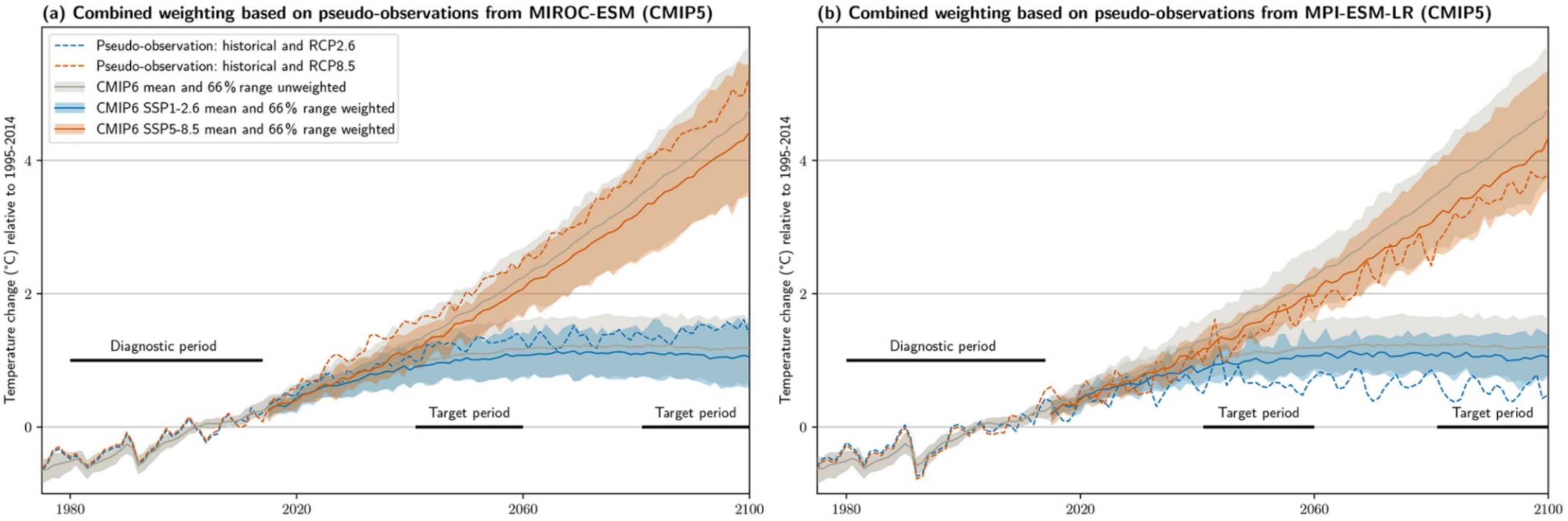


Figure: Temperature change for unweighted and weighted CMIP6 as well as the CMIP5 models serving as pseudo-observations. Cases with (a) decrease and (b) increase in skill. Brunner et al. (2020a)

Skill of the weighting: CMIP5 models

Weighting is applied to CMIP6 based on CMIP5 models used as **pseudo-observation**

- Weight based on pseudo observations (from a CMIP5 model) in the past
- Evaluation of ensemble forecast skill in the future (compared to same CMIP5 model)
- **Median skill increase: 10-20%**
- Remaining risk of skill decrease through weighting

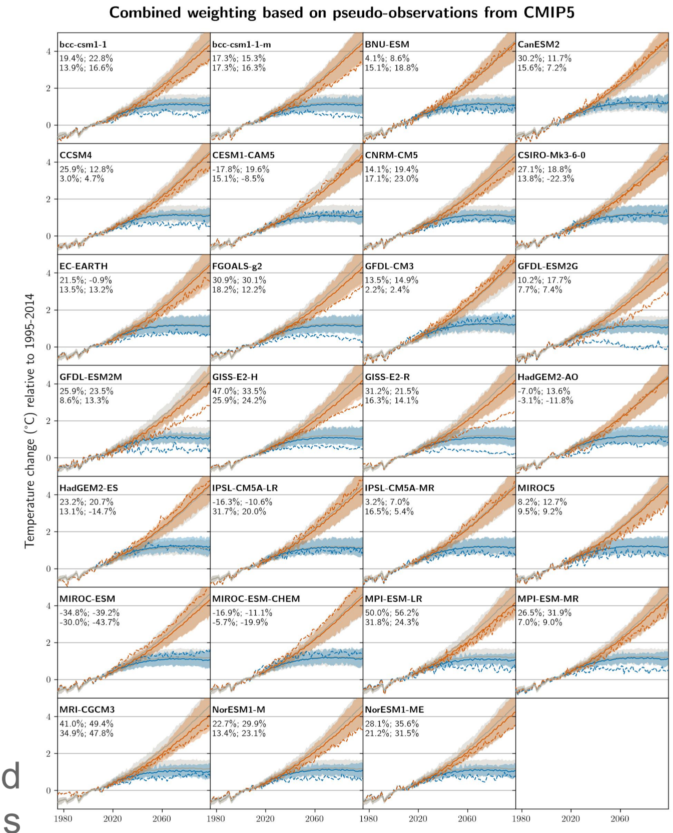


Figure: Temperature change for unweighted and weighted CMIP6 as well as the CMIP5 models serving as pseudo-observations. Brunner et al. (2020a)

Model-as-truth testing

- Other names: **perfect model test** or using models as **pseudo-observations**.
- A similar concept in statistics is a **leave-one-out cross-validation**
- The aim is to evaluate the **skill and reliability** of the weighting method in constraining a given model ensemble (e.g., CMIP6) by
  - ...weighting based on pseudo-observations (from another model - from CMIP6 or CMIP5) in the past
  - ...evaluating the weighting in the future against the ‘truth (from the same model) compared to the unweighted case

Median skill increase: 10-20%

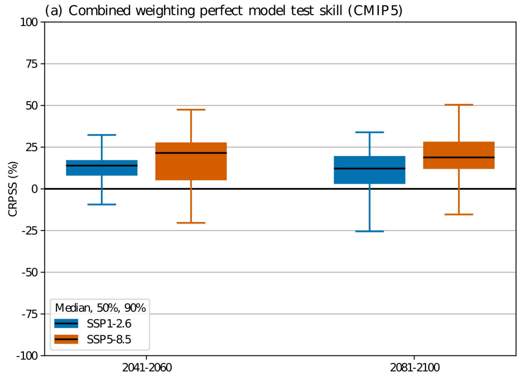


Figure: Continuous ranked probability skill score (CRPSS) for CMIP6 relative to the unweighted ensemble using perfect models from CMIP5.
Brunner et al. (2020a)

Combining model-as-truth testing and method comparison

- Several methods provide individual skill estimates based on model-as-truth tests but they are not comparable
- Comparing method results can increase confidence (if they agree) but can tell us nothing about which method is ‘right’
- **Coordinated model-as-truth test using a level playing field**

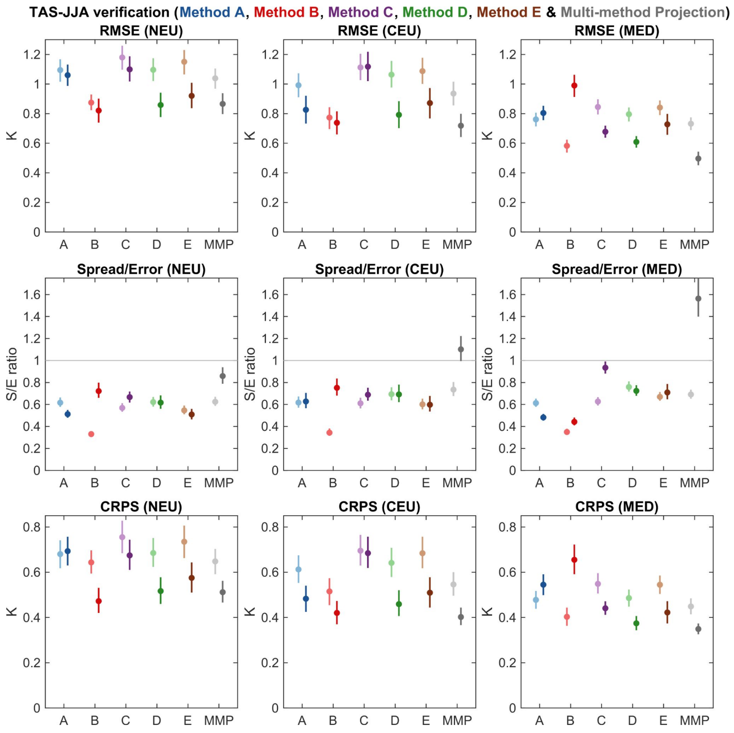


Figure: RMSE, S/E, CRPS for five different methods based on a consistent model-as-truth test. O’Reilly et al. (in prep)

Reliability of the weighting

The **reliability** of the weighting is **~100% in a model world** by definition

- The performance shape parameter σ_D is selected to lead to reliable weighting based on a leave-one-out model-as-truth test
- Caveats:
  - all models might have common biases → overconfident in the real world
  - computationally expensive

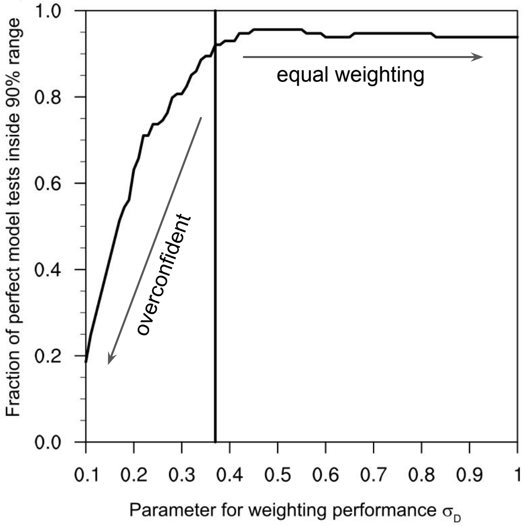
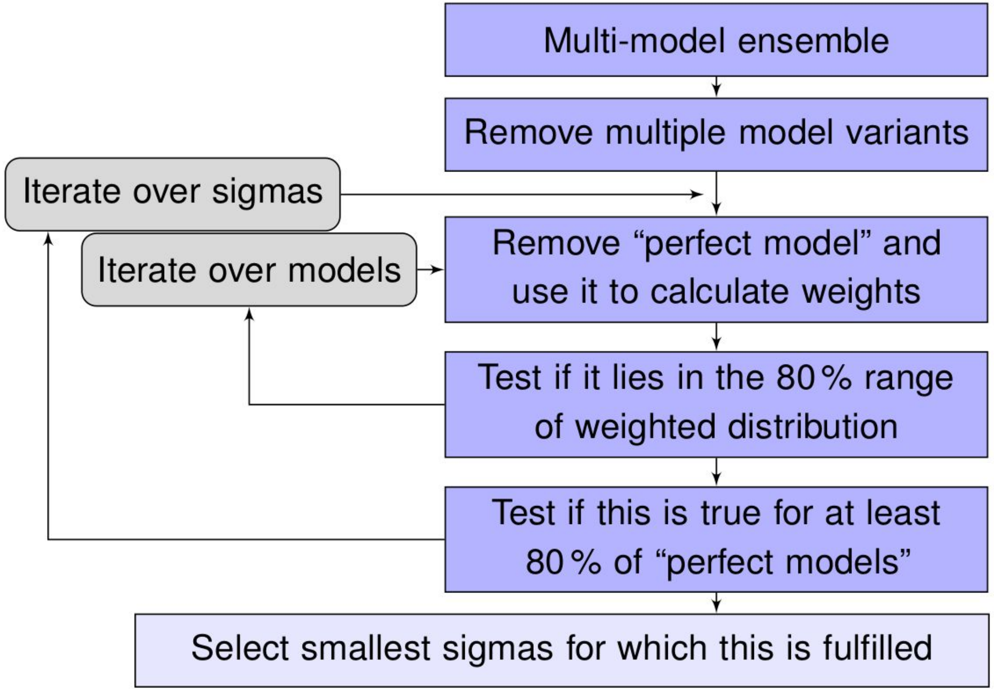
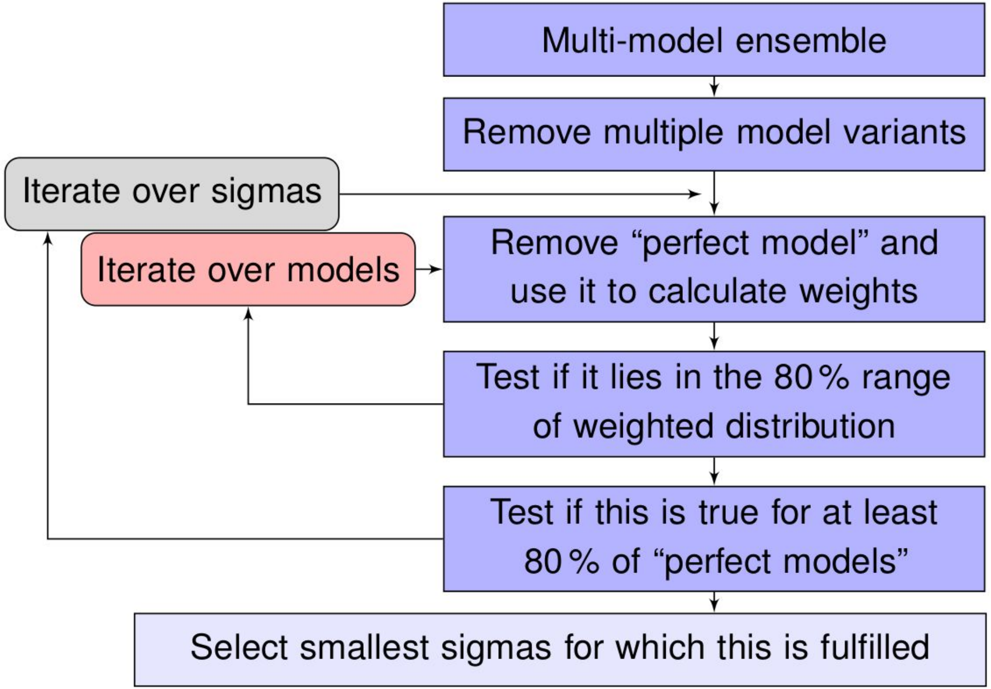


Figure: Fraction of cases when the ‘truth’ is in the 5–95% range predicted by weighting all other models. Knutti et al. (2017)

Reliability of the weighting: model-as-truth test


Multi-model ensemble
Remove multiple model variants
Iterate over sigmas
Iterate over models
Remove “perfect model” and use it to calculate weights
Test if it lies in the 80 % range of weighted distribution
Test if this is true for at least 80 % of “perfect models”
Select smallest sigmas for which this is fulfilled

Reliability of the weighting: model-as-truth test



Multi-model ensemble

Remove multiple model variants

Iterate over sigmas

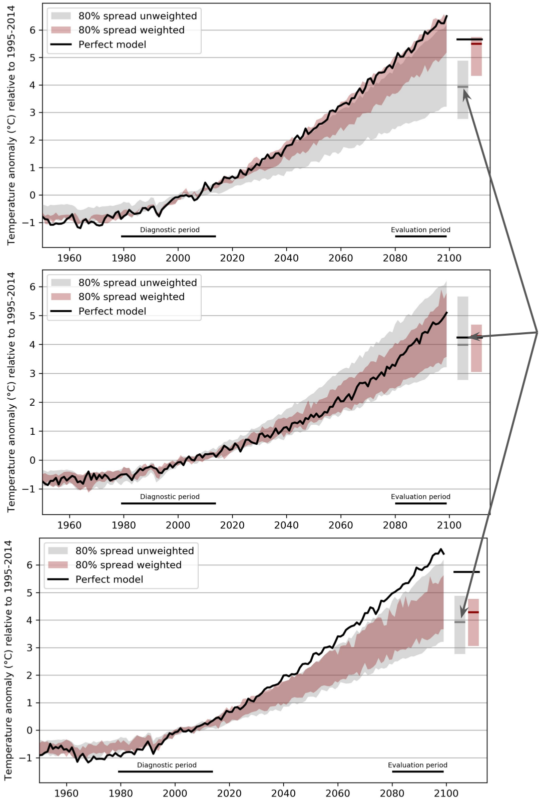
Iterate over models

Remove "perfect model" and use it to calculate weights

Test if it lies in the 80 % range of weighted distribution

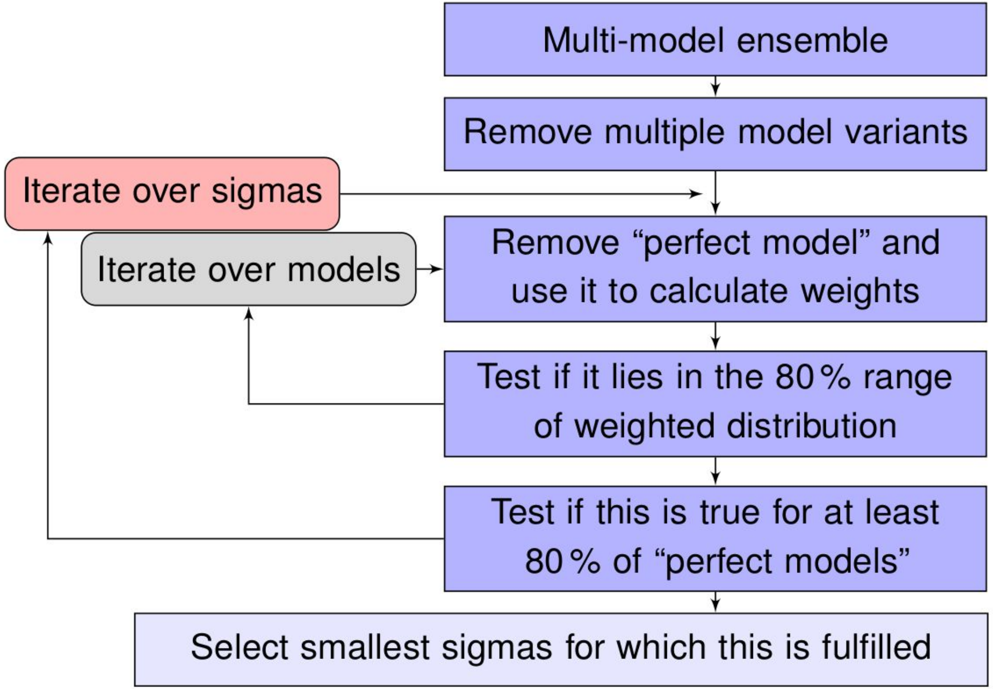
Test if this is true for at least 80 % of "perfect models"

Select smallest sigmas for which this is fulfilled



In the **unweighted** case we expect **~80%** of perfect models to fall inside of the distribution

Reliability of the weighting: model-as-truth test


Multi-model ensemble

Remove multiple model variants

Iterate over sigmas

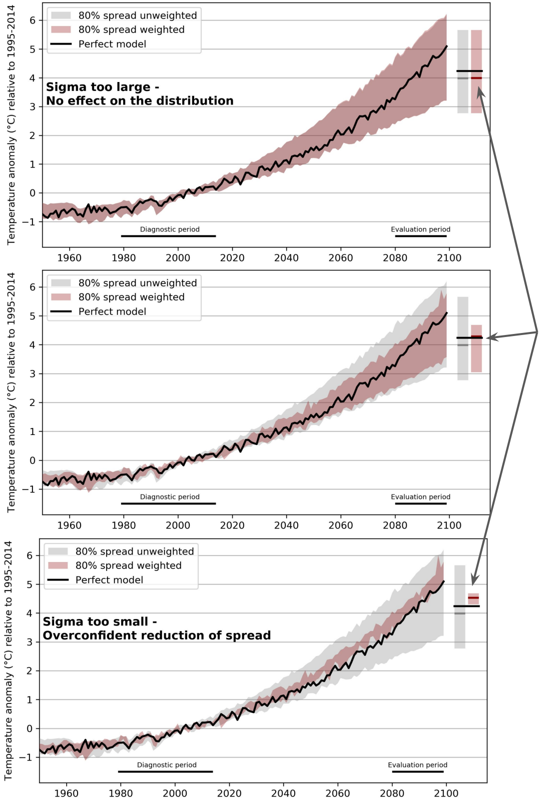
Iterate over models

Remove “perfect model” and use it to calculate weights

Test if it lies in the 80 % range of weighted distribution

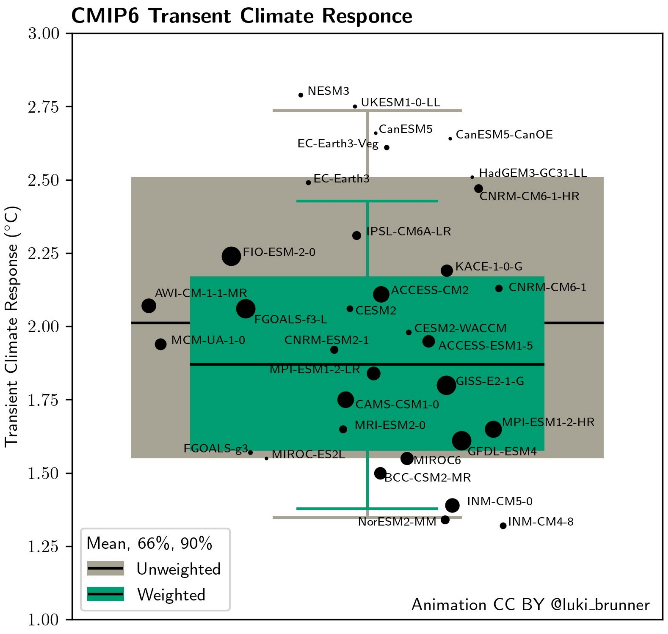
Test if this is true for at least 80 % of “perfect models”

Select smallest sigmas for which this is fulfilled



The sigma must be chosen so that weighting **does not decrease** that value

Weighting Transient Climate Response



- A measure for a **models transient response** to a **doubling of CO_2**
- Frequently used measure for the models' **climate sensitivity** independent of time
- Effect of weighting models (mean / 66%):
  - Unweighted: 2.01 (1.55-2.51)°C
  - Weighted: 1.87 (1.58-2.17)°C
  - Change: -0.14°C (-37.5%)
- Other studies (66% range):
  - Nijsse et al. (2020): 1.3°C-2.1°C
  - Tokarska et al. (2020): 1.2°C-2.0°C
  - Sherwood et al. (2020): 1.5°C-2.2°C