

Aims

1. Determine code genealogy of CMIP3, CMIP5 and CMIP6 models with a focus on the atmospheric component and atmospheric physics.
 2. Create a weighting method which takes into account code dependence between the models.
 3. Evaluate climate feedbacks, sensitivity, forcing, historical and projected time series of global mean near-surface temperature under different weighting methods.

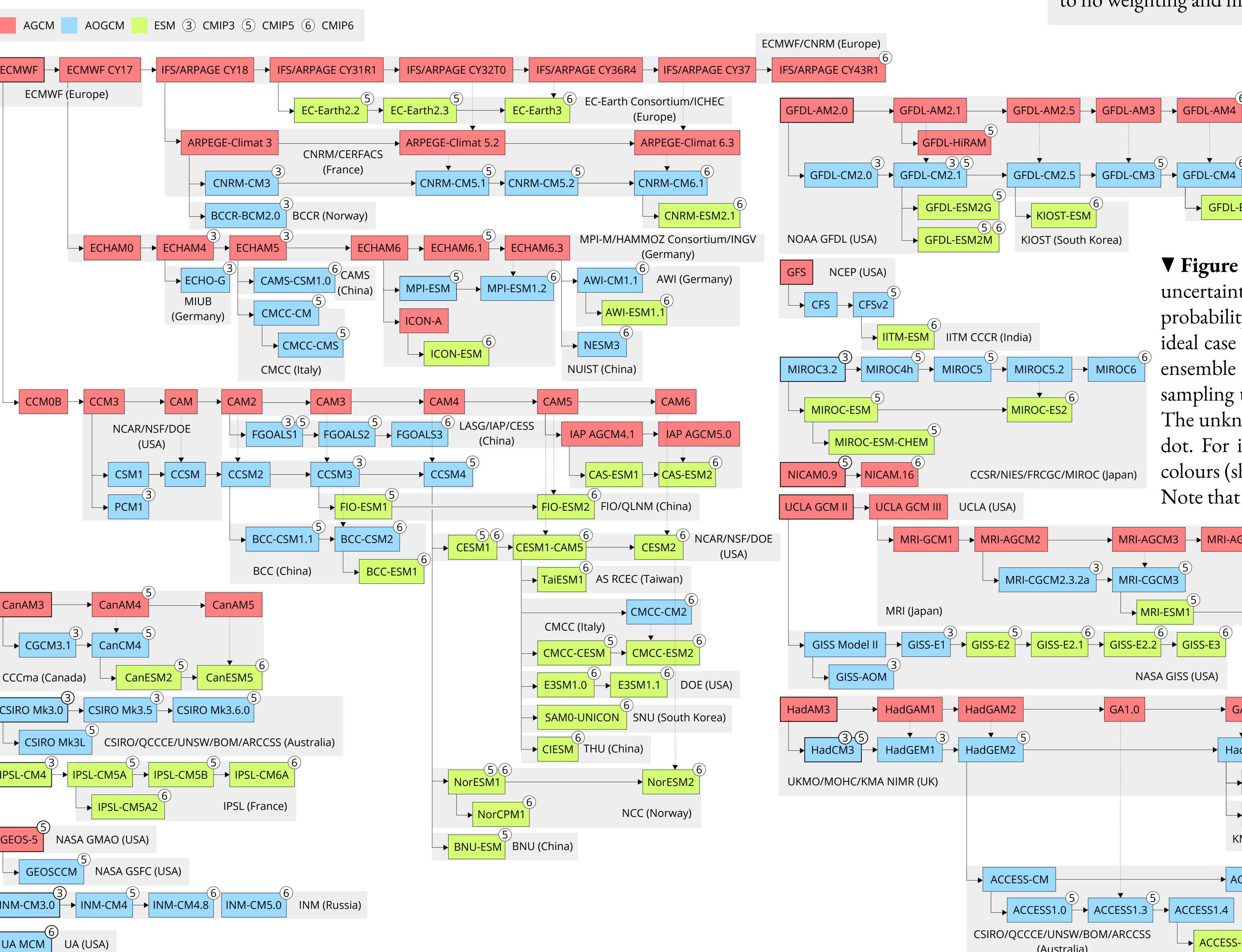
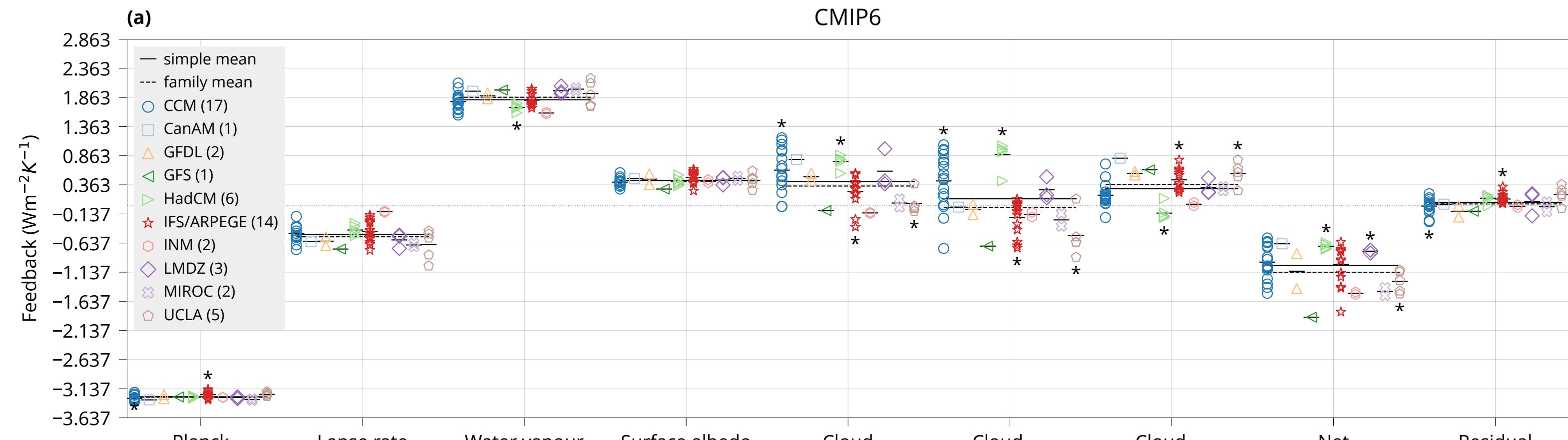


Figure 3: Climate feedbacks, effective climate sensitivity (ECS) and effective radiative forcing (ERF_{2x}) arranged by model family in the Climate Model Intercomparison Project (CMIP) phase 5 (b, d) and 6 (a, c). Model family is identified by the oldest ancestor model. In the legend, numbers in parentheses are the number of models in the family present in the plot. Model families whose *simple* mean is significantly different (with 95% confidence) from the *simple* multi-model mean are marked with an asterisk (*). The underlying data are from Zelinka (2022), described in Zelinka et al. (2020).



(b)

Y-axis: Feedback ($\text{W m}^{-2} \text{ K}^{-1}$)

X-axis categories: Planck, Lapse rate, Water vapour, Surface albedo, Cloud, Cloud_{SW}, Cloud_{LW}, Net, Residual

Legend:

- simple mean
- - - family mean
- CCM (6)
- ▽ CSIRO (1)
- CanAM (1)
- △ GFDL (3)
- ▽ HadCM (3)
- ★ IFS/ARPEGE (4)
- ◇ LMDZ (3)
- ✗ MIROC (2)
- ◇ UCLA (3)

(c)

Y-axis: ECS (K); ERF_{2x} (W m^{-2})

X-axis categories: Planck, Lapse rate, Water vapour, Surface albedo, Cloud, Cloud_{SW}, Cloud_{LW}, Net, Residual

(d)

Y-axis: ECS (K); ERF_{2x} (W m^{-2})

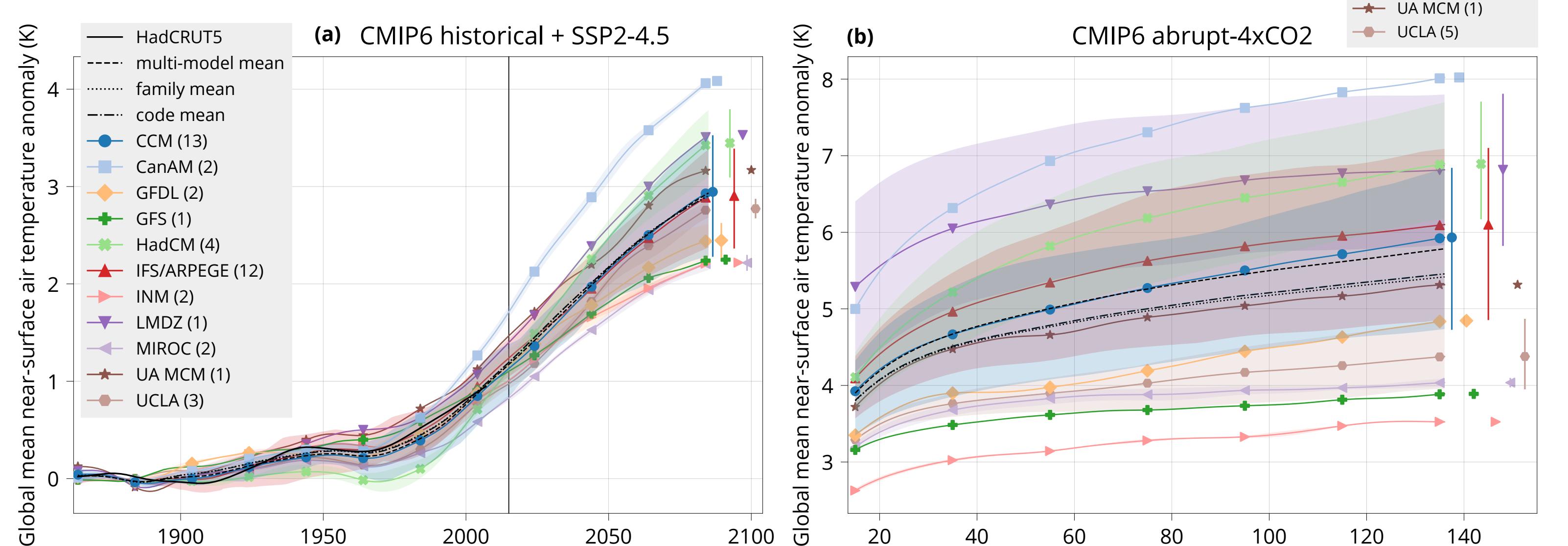
X-axis categories: Planck, Lapse rate, Water vapour, Surface albedo, Cloud, Cloud_{SW}, Cloud_{LW}, Net, Residual

CMIP5 models shown in (b) and (c): CCM (6), CSIRO (1), CanAM (1), GFDL (3), HadCM (3), IFS/ARPEGE (4), LMDZ (3), MIROC (2), UCLA (3).

CMIP6 models shown in (c): CCM (6), CSIRO (1), CanAM (1), GFDL (3), HadCM (3), IFS/ARPEGE (4), LMDZ (3), MIROC (2), UCLA (3).



The *simple* multi-model, *code* and *family* mean. The model family time series are a *simple* mean of models in the family. The time series are smoothed with a Gaussian kernel with a standard deviation of 7 years. The first and the last 14 years of the time series are not shown to avoid artefacts of the smoothing. The values are relative to the mean of the first 30 years of the individual time series in **(a)**, and relative to the mean of the whole individual time series of the *piControl* experiment in **(b)**. Shown are confidence bands representing the 68th percentile range. The vertical divider in the *historical* + SSP2-4.5 plot separates time range of the two experiments. In the legend, the number in the parentheses is the number of models in the family. All CMIP5 and CMIP6 models with necessary data available on the Earth System Grid were included in the plots.



'indings

- A large code dependence exists between most CMIP models.
 - CMIP models can be grouped into about 14 families by code heritage.
 - Code and family weighting can partly reconcile differences between CMIP5 and CMIP6 climate sensitivity.
 - Model families tended to exhibit warm/cold tendencies across CMIP generations.
 - We propose *code* and *family* weighting methods as a more fair weighting for multi-model ensemble studies as an alternative to no weighting and model output similarity and performance weighting.

◀ **Figure 1:** Model code genealogy of models participating in the Climate Model Intercomparison Project (CMIP) phase 3, 5 and 6, including their related common ancestor models. Horizontal arrows indicate inheritance between multiple versions of the same model. Vertical solid arrows indicate inheritance between different models. Vertical dotted arrows indicate inheritance from an AGCM to an AOGCM or ESM (this can also mean that the model is used as a component of the more complex model). Model boxes with a thick outline indicate the oldest model of the model family. The genealogy only traces models necessary for placing the CMIP models and omits versions not included in CMIP.

► **Figure 2:** A theoretical illustrative example of model sampling of the model hypothesis space (model structural uncertainty), representing realisations of physical climate processes (model structure). The shading indicates a probability density function (PDF) quantifying our collective belief that a certain representation is true. In an ideal case (**a**), models are unbiased samples from this PDF, allowing us to estimate the PDF from a multi-model ensemble (MME). In reality (**b**), they form clusters because of structural model dependence (code sharing), sampling the PDF in a biased manner. Weighted sampling is necessary to estimate the PDF from such an MME. The unknown true physical representation, not coinciding with the PDF maximum or mean, is indicated by a red dot. For illustrative purposes only, the hypothesis space is visualised in a 2-dimensional space. Model marker colours (shapes) in (**b**) indicate different hypothetical model families, within which models are structurally related. Note that the PDF represents model structure and might not correlate with model output PDF.

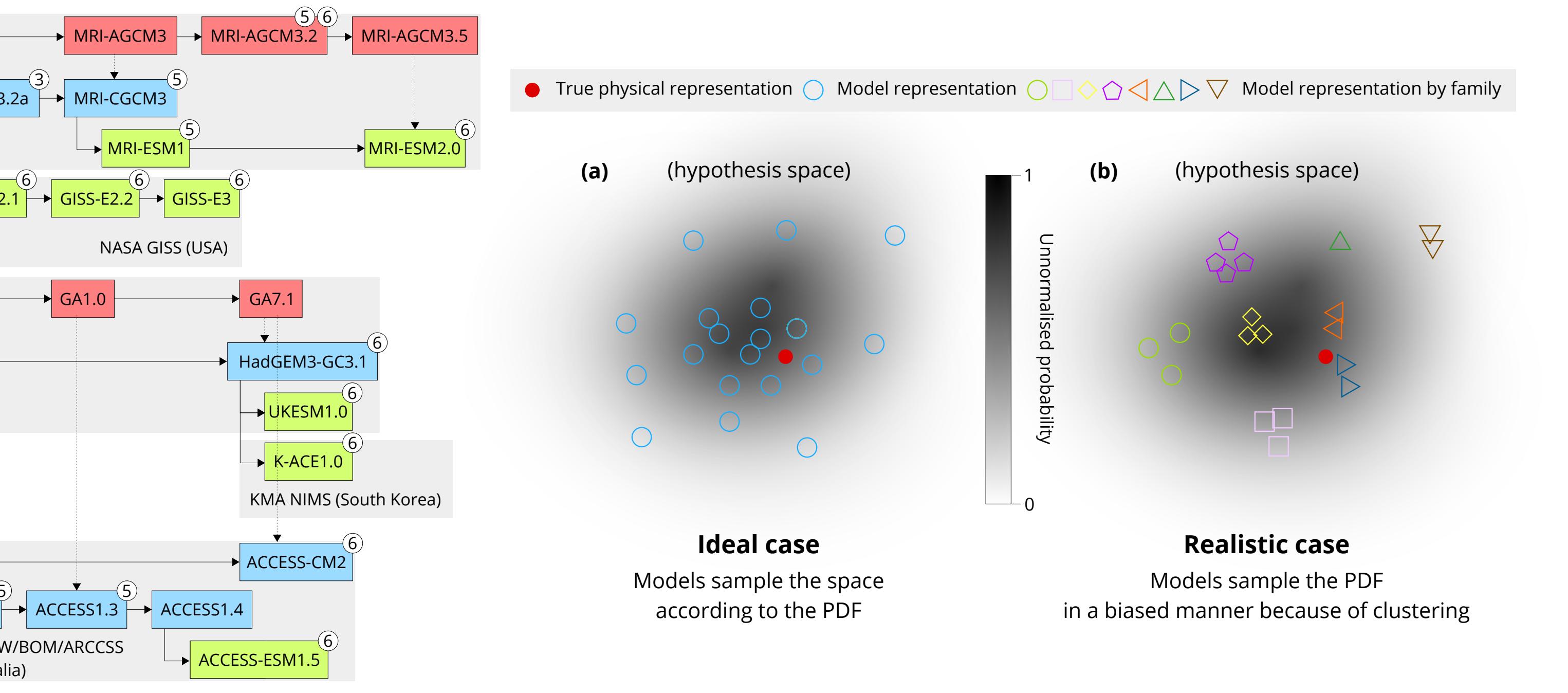


Figure 4: Climate feedbacks, effective climate sensitivity (ECS), effective radiative forcing (ERF_{2x}) in the Climate Model Intercomparison Project (CMIP) phase 5 (**a**) and 6 (**b**) under different weighting methods (*model, institute, country, code* and *nily*) relative to a *simple* mean. (**c**) The difference between CMIP6 and CMIP5. The legend in (**c**) shows root mean square difference (RMSD) between CMIP6 and CMIP5. Climate feedbacks: Planck, water vapour (WV), lapse rate (LR); surface albedo (Albedo); total cloud feedback (Cloud); shortwave cloud feedback (Cloud_{SW}); longwave cloud feedback (Cloud_{LW}); net feedback (Net); residual feedback (Residual). The underlying data are from Zelinka (2022), described in Zelinka et al. (2020).

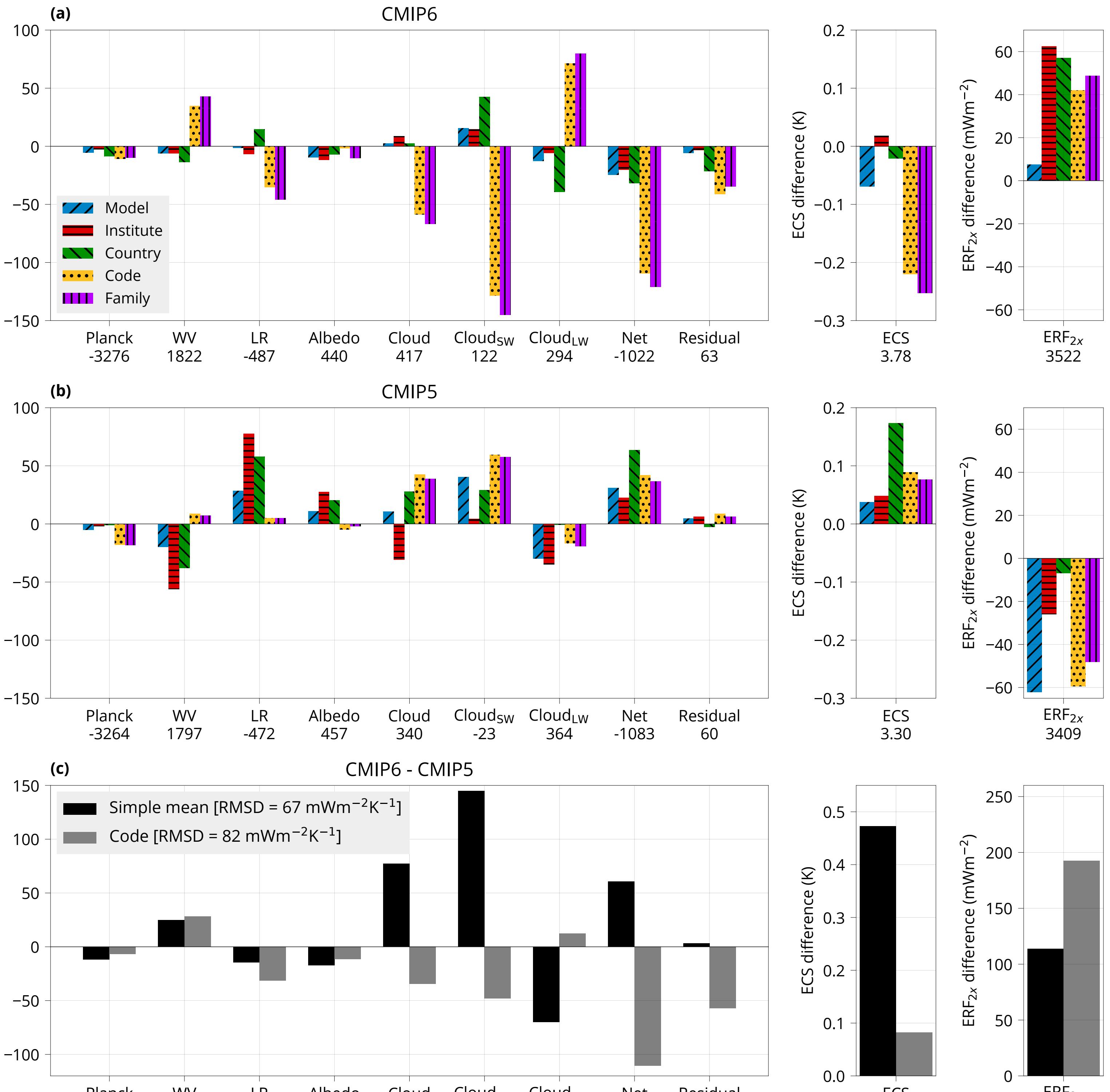


Figure 6: The same as Fig. 5 but for CMIP5, and the RCP4.5 experiment instead of SSP2-4.5

