Probabilistic projections of future warming and climate sensitivity trajectories

Philip Goodwin[1]

[1]School of Ocean and Earth Science, University of Southampton, Waterfront Campus, Southampton, SO14 3ZH, UK.

Email corresponding author: p.a.goodwin@soton.ac.uk

Abstract

Projections of future global mean surface warming for a given forcing scenario remain uncertain, largely due to uncertainty in the climate sensitivity. The ensemble of Earth system models from the Climate Model Intercomparison Project phase 6 (CMIP6) represent the dominant tools for projecting future global warming. However, the distribution of climate sensitivities within the CMIP6 ensemble is not representative of recent independent probabilistic estimates, and the ensemble contains significant variation in simulated historic surface warming outside agreement with observational datasets. Here, a Bayesian approach is used to infer joint probabilistic projections of future surface warming and climate sensitivity for SSP scenarios. The projections use an efficient climate model ensemble filtered and weighted to encapsulate observational uncertainty in historic warming and ocean heat content anomalies. The probabilistic projection of climate sensitivity produces a best estimate of 2.9°C, and 5th to 95th percentile range of 1.5 to 4.6 °C, in line with previous estimates using multiple lines of evidence. The joint projection of surface warming over the period 2030 to 2040 has a 50% or greater probability of exceeding 1.5 °C above preindustrial for all SSPs considered: 119, 126, 245, 370 and 585. Average warming by the period 2050 to 2060 has a greater than 50% chance of exceeding 2°C for SSPs 245, 370 and 585. These results imply that global warming is no longer likely to remain under 1.5°C, even with drastic and immediate mitigation, and highlight the importance of urgent action to avoid exceeding 2°C warming.

Keywords: climate sensitivity, global warming, climate projection, climate feedback, climate change, climate models.

1. Projecting future warming and climate feedback trajectories

Projecting global mean surface warming (ΔT in K) and either climate feedback (λ in Wm^{-2}K^{-1}) or climate sensitivity (S in K) responses to anthropogenic forcing scenarios remain key priorities for climate science (e.g. IPCC, 2013; Sherwood et al., 2020). Surface warming since the preindustrial average at time t, $\Delta T(t)$ is related to the time evolutions of radiative forcing, R in Wm^{-2}, the net energy uptake of the Earth system, N in Wm^{-2}, and either λ or S via

$$\lambda(t)\Delta T(t) = R(t) - N(t) = \left(\frac{R_{2\times CO2}}{S(t)}\right)\Delta T(t) \quad (1)$$

where: $R_{2\times CO2}$ is the radiative forcing for a doubling of atmospheric CO_2; the sign convention chosen has positive λ, such that $\lambda(t)\Delta T(t)$ is the additional energy emitted to space due to surface warming in Wm^{-2}; and the most (around 93%) of N leads to the ocean heat content (OHC) anomaly (IPCC, 2013). This study uses warming above the 1850-1900 average to represent warming above preindustrial, $\Delta T(t)$ (e.g. IPCC, 2013).

Some of the terms in the energy (im)balance equation, (1), are historically constrained by observational datasets. Historic surface warming, $\Delta T(t_{hist})$, is well constrained from observational datasets (Fig. 1, black; see e.g. Morice et al., 2020; Rohde et al., 2013; Lenssen et al., 2019) while there is relatively more observational uncertainty in historic ocean heat content anomaly, $\Delta\mathrm{OHC}_{hist}$ (e.g. Cheng et al., 2017; Levitus et al., 2012), giving greater uncertainty in the time average of Earth's historic global energy imbalance, $N(t)$ (Trenberth et al., 2014). To constrain future surface warming for a specified atmospheric composition scenario, we need to also constrain the radiative forcing from atmospheric composition changes, $R(t)$, and the climate feedback, $\lambda(t)$, (or equivalently climate sensitivity, S) response and predict their future evolution, eq. (1).

However, the historic evolutions of λ, S and R have proven difficult to constrain from observational datasets (e.g. Gregory & Andrews, 2016; Gregory et al., 2019; Knutti et al., 2017; Lewis & Curry,

2014; Otto et al., 2013), in part because there is large uncertainty in the magnitude of historic radiative forcing, R, principally from aerosols (Andrews and Forster, 2020; IPCC, 2013; Zelinka et al., 2014). Any climate model ensemble chosen to generate future ΔT projections must also simulate (explicitly or implicitly) historic and future time evolutions of λ, S, R and N, eqs. (1), (2), (3), and historic warming and ocean heat content anomalies. Moreover, eq. (1) reveals that there are multiple historic trajectories that λ and R (or S and R) can take to reproduce the historic trajectories of ΔT and N consistent with observations.

The aim here is to produce joint probabilistic projections of the future trajectory of global mean warming above preindustrial average, $\Delta T_{future}(t)$ in K,

$$p(\Delta T(t_{future})|\Delta T(t_{hist}),\Delta\mathrm{OHC}(t_{hist}),i) \tag{2}$$

and future trajectory of either climate sensitivity, $S(t_{future})$ in K, or climate feedback, $\lambda(t_{future})$,

$$p(S(t_{future})\ |\Delta T(t_{hist}),\Delta\mathrm{OHC}(t_{hist}),i) \tag{3}$$

$$p(\lambda(t_{future})|\Delta T(t_{hist}),\Delta\mathrm{OHC}(t_{hist}),i) \tag{4}$$

for Shared Socio-economic Pathway (SSP) scenarios (O'Neill et al., 2016) given information from observational reconstructions of historic surface warming, $\Delta T(t_{hist})$ (Morice et al., 2020), historic ocean heat content, $\Delta\mathrm{OHC}(t_{hist})$ (Cheng et al., 2017), and other background information, i. This background information, i, will include prior uncertainty distributions for radiative forcing from a range of sources including aerosols (Andrews and Forster, 2020; IPCC, 2013; Zelinka et al., 2014) prior distributions for a range of climate model parameters and the functional form of a range of parameterisations within an efficient climate model with a box-model representation of heat and

carbon fluxes in the atmosphere-ocean system (Goodwin, 2016; 2018; Goodwin and Cael, 2021). To then assess whether the posterior efficient model ensemble is able to capture the qualitative response of the Earth system to forcing, it will be compared to the normalised responses of state-of-the-art CMIP6 models to idealised forcing scenarios,

The primary tools for projecting future ΔT are process-based numerical climate models (e.g. Eyring et al., 2016), whose performance can be in some way assessed against historic observational datasets (e.g. Morice et al., 2020; Cheng et al., 2017; Levitus et al., 2012; Rohde et al., 2013; Lenssen et al., 2019). The dominant climate model variants used to project ΔT by the Intergovernmental Panel of Climate Change (IPCC) Assessment Reports (e.g. IPCC, 2013) are the Climate Model Intercomparison Project (CMIP) ensemble of complex and high-resolution global climate and Earth system models (Eyring et al., 2016), forced by a range of specified future scenarios (O'Neill et al., 2016).

Climate processes relevant to 21st century surface warming and climate feedback projections act over at least 14 orders of magnitude in their spatial and temporal scales (see Stocker, 2011); for example, from cloud droplet collisions up to global mean surface warming on decadal to century timescales. When projecting global mean surface warming to year 2100, computational resources limit state-of-the-art CMIP phase 6 (CMIP6) Earth system models (e.g. Williams et al., 2017) to represent a maximum of around 6 orders of magnitude in time, from a single time-step to a century, and 3 orders of magnitude in horizontal space, from grid-scale to global. The 'high-resolution' Earth system models must therefore contain a great many parameterisations of processes acting on unresolved spatial and temporal scales. These parameterisations include specifying the functional form of equations and their coefficient values.

Parameterisations of unresolved processes significantly affect the historic and future trajectories of simulated warming and climate feedback, demonstrated by the large variation in historic simulations of surface warming from different CMIP models (Fig. 1 for CMIP6; see e.g. IPCC, 2013

for CMIP5), each containing a different set of parameterisations. In their relatively high spatial and temporal resolution, these state-of-the-art models can only run between one to several-tens of simulations each due to their high computational cost, resulting in a total ensemble size for all CMIP6 models of up to a few hundreds of simulations for a given scenario. Thus, the CMIP climate models are unable to perform enough simulations with varied parameterisations to identify all possible combinations of historic pathways of $\lambda(t)$ and $R(t)$ that produce observation-consistent historic pathways of $\Delta T(t)$ and $N(t)$, (eq. 1; Fig. 1a).

Computationally efficient climate models (e.g. Nicholls et al., 2020) are able to utilise computational resources to resolve their own plausible combinations of parameterisations, through the construction of very large model ensembles. However, this parameterisation resolution occurs at the expense of a much-reduced resolution of relevant climate processes. Since the relative simplicity of efficient model representations may not capture the required complexity of the climate system, the *qualitative* nature of their simulated projections may be too simplistic. For example, an efficient climate model may use a single climate feedback value for all time and for all sources of radiative forcing, when in complex models (and the real climate system) climate feedbacks may vary over time and may respond differently to, say, localised aerosol forcing than to well mixed greenhouse gases. Efficient climate models are therefore able to perform a quantitative assessment of uncertainties within their own model framework, but the qualitative form of their responses may not accurately reflect reality.

Since no model can resolve all relevant processes, computational resources introduce an inherent trade-off between resolving relevant climate processes and resolving parameterisation uncertainty. The greater the complexity and resolution of a climate model the less possible parameterisations can be simulated.

This study produces joint probabilistic projections of the future trajectories of global mean warming and climate feedback using an efficient model ensemble. Section 2 considers existing methods for

inferring future probabilistic projections from complex model ensembles. Section 3 presents a novel method for inferring joint probabilistic projections of surface warming and climate feedback from an efficient model ensemble. Section 4 presents the future probabilistic projections generated with section 5 discussing the results and wider context.

2 Inferring probabilistic projections from complex model ensembles

By far the dominant quantitative projections of global mean surface warming in the scientific literature, as used in the Intergovernmental Panel on Climate Change Assessment Reports (IPCC, 2013; 2007), are derived from relatively small ensembles of complex climate model output. The simplest method to develop probabilistic projections of global mean warming and climate sensitivity is to use the raw unadjusted CMIP ensemble simulations. This generation of probabilistic projections of surface warming and climate sensitivity, given the (considerable) background information that goes into the design and integration of the CMIP models, i_{CMIP}, is expressed

$$p(\Delta T(t_{future})|i_{CMIP}) \tag{5}$$

$$p(\mathrm{S}(t_{future})|i_{CMIP}) \tag{6}$$

However, such unadjusted model output (e.g. Fig. 1a, grey lines for a sample of CMIP6 simulations) may give spurious probabilistic inferences, eqns. (5), (6). For example, using the frequency distribution of climate sensitivity values in the CMIP6 ensemble (Zelinka et al., 2020) to infer a probabilistic assessment of S would give results quite different from independent assessments using other lines of evidence (Sherwood et al., 2020; IPCC, 2013): over 20% of CMIP6 models analysed by Zelinka et al. (2020) have climate sensitivity values over 4.7K, while Sherwood et al. (2020) evaluate only a 5% probability that climate sensitivity is greater than 4.7K.

Also, consider the range of simulated CMP6 values of global temperature anomaly above preindustrial for the immediate future when following SSP585. The simulated 11-year running

average surface warming above 1850-1900 in the year 2025 ranges from ΔT = 0.9 to 2.4 °C in the sample of CMIP6 models in Fig. 1a (grey lines). To get to the lower end of these simulated near-future warming levels (Fig. 1a, grey) from our observed warming to the present day (Fig. 1a, black) would require the recent observed decadal rate of warming to reverse, just as radiative forcing is increasing. Meanwhile, to reach the upper end of these simulated near-future warming levels would require the recent decadal rate of warming (Fig. 1a, black) to suddenly increase by a factor of 6. As both of these extremes seem highly improbable, generating plausible future projection appears to require more information than just the (considerable) information put into the CMIP6 ensemble, i_{CMIP} (eqns. 5,6).

To account for any error in simulated historic warming, or bias in climate sensitivity, in CMIP ensemble members (Fig. 1a, grey and black), a method often applied in the literature is to use observations of historic surface warming to inform how the raw CMIP output should be interpreted in a probabilistic sense (e.g. IPCC, 2013; Caldwell et al., 2018; Cox et al., 2018; Hall et al., 2019; Nijsse et al., 2020; Tokarska et al., 2020; Brunner et al., 2020). For temperature projections, this observation adjustment is expressed as a probability of future temperature anomaly given both historic observed temperature anomaly and the background information that goes into the CMIP ensemble,

$$p(\Delta T(t_{future})|\Delta T(t_{hist}),i_{CMIP}) \quad (7)$$

For climate sensitivity, one may use an emergent constraint based on the historic temperature record to produce a probability estimate for climate sensitivity given historic temperature records and background information in the CMIP ensemble (e.g. Cox et al., 2018),

$$p(\mathrm{S}(t_{future})|\Delta T(t_{hist}),i_{CMIP}) \quad (8)$$

The technique for projecting future warming above preindustrial, $\Delta T(t_{future})$, adopted in IPCC (2013) AR5 was simply to add historic observed warming up to the present time t_0, $\Delta T_{obs}(t_0)$, to the additional simulated warming in the future relative to simulated historic warming to the present day in the CMIP5 models, $\Delta T_{sim}(t_{future}) - \Delta T_{sim}(t_0)$ (IPCC, 2013, Table 12.3 therein): $\Delta T(t_{future}) = \Delta T_{obs}(t_0) + (\Delta T_{sim}(t_{future}) - \Delta T_{sim}(t_0))$. One issue with this approach is that simulations are used to support a region of probability space for a different level of warming, $\Delta T(t_{future})$, than the level of warming they simulate, $\Delta T_{sim}(t_{future})$, eq. (7). For example, applied to the CMIP6 ensemble in Fig. 1a, one ensemble member that simulates mean warming during 2020 to 2030 of 2.32 °C above predindustrial would be used to support the probability of warming being only 1.56 °C over that period. Meanwhile, another ensemble member that simulates warming of 1.05 °C during 2020-2030 (Fig. 1a, grey) would be used to support the probability that warming will be 1.38 °C above preindustrial.

Other, more sophisticated, statistical techniques have been proposed for making quantitative projections of observable properties such as warming and/or climate feedback (or climate sensitivity) using the CMIP model output and observational records (eqns, 7, 8), such as the application of emergent constraints (e.g. Caldwell et al., 2018; Cox et al., 2018; Hall et al., 2019; Nijsse et al., 2020; Schlund et al., 2020) and observational adjustments and weightings (e.g. Tokarska et al., 2020; Brunner et al., 2020). Emergent constraints are used to analyse an ensemble of numerical climate simulations, where some functional relationship is thought or known to exist within the model ensemble between one observable property of the climate system and the observable property for which projection is desired (e.g. Hall et al., 2019; Caldwell et al., 2018).

3 Inferring probabilistic future projections from an efficient model ensemble

Efficient climate models are specifically designed to allow the generation of large ensemble simulations, through reduced resolution of relevant climate processes in space and time. Examples

of models designed and configured in this way are found in the Reduced Complexity Model Intercomparison Project (RCMIP: Nicholls et al., 2020).

For the complex model ensembles, the prior expectation for different quantities (such as climate feedback or climate sensitivity) is hidden within the background information put into the development of each of the models, i_{CMIP} in eqns. (5-8). However, the individual model teams do not systematically vary the information put into each model such that the ensemble produces a distribution of each property (e.g. climate sensitivity) equal to some agreed prior belief. Instead, each model is developed independently according to the best estimates of the different model teams. As such, the complex model ensemble may under-sample the less likely values of climate system properties, since no model-team's best estimate is to build a model with those properties.

This study uses the Warming Acidification and Sea level Projector (WASP: Goodwin, 2016; 2018): a box-model representation of heat and carbon fluxes across the atmosphere-ocean system and carbon fluxes between the atmosphere and terrestrial systems. Here, only the atmosphere-ocean components of the WASP model (Goodwin, 2016; 2018; Goodwin and Cael, 2021) are used. This consists of a well-mixed atmosphere box connected to a well-mixed surface ocean box, itself connected to 4 sub-surface ocean boxes representing the upper ventilated region of the ocean, the intermediate water region, the deep ocean water region and the bottom water ocean region.

WASP version 3 (Goodwin, 2018), with time-evolving climate feedbacks, is adopted and configured after Goodwin and Cael (2021) with climate feedback split into three components: the instantaneous (Planck) feedback, λ_P, fast feedbacks (operating on the timescale of water vapour residence in the atmosphere), λ_f, and multidecadal feedbacks (operating on timescales representing the sea-surface temperature warming pattern adjustment), λ_{md} (Appendix B).

Here, the WASP model is utilised to generate a Monte Carlo prior ensemble of 2.4×10^7 efficient model simulations, with 25 model parameters varied independently between ensemble members

(e.g. Fig. 2) and each ensemble member having a unique randomly imposed energy imbalance using AR1 noise to approximate the monthly and yearly observed noise from Trenberth et al. (2014). The computational efficiency of WASP allows all of these 2.4×10^7 to be integrated from preindustrial to the present day in ~ 100 hours on a single Linux cluster node using dual 2.0GHz Intel Skylake processors. The prior simulations are assessed against historical observations (Table 1) and a posterior ensemble of 13340 simulations is extracted and weighted. It then takes ~ 5 minutes on a standard desktop or laptop computer to integrate the 13340 members of the posterior ensemble from preindustrial to year 2300 for a given SSP scenario. When historical output from a complex model ensemble (CMIP5) was used to extract the posterior ensemble in WASP (Goodwin, 2016; Goodwin et al., 2017), the posterior future projection ranges agreed with the complex model ensemble for mitigated and unmitigated scenarios, both for future surface warming (Goodwin, 2016; Goodwin et al., 2017) and future ocean heat uptake (Goodwin et al., 2017). This provides confidence in the future WASP projections when observations are used to extract the posterior ensemble.

The 25 varied parameters include climate feedback magnitudes and timescales for the instantaneous Planck, fast and multidecadal climate feedbacks (Fig. 2), the radiative forcing sensitivity to atmospheric composition for different forcing agents including aerosols (Fig. 2b), and ocean ventilation timescales for different ocean regions. A full list of parameters varied is found in Goodwin and Cael (2021). The prior distributions for the magnitude of climate feedback terms are updated here, with prior distributions for all other terms as described in Goodwin and Cael (2021).

3.1 New prior distributions of climate feedback parameters

The prior model ensemble must contain some prior distributions for climate feedback parameters. The prior distributions for climate feedback and climate sensitivity considered by a recent review (Sherwood et al., 2020) assumed either uniform climate feedback (λ, $Wm^{-2}K^{-1}$) or uniform climate sensitivity (S, K). The difference between these two prior distributions makes a significant difference to the upper percentiles of the posterior distribution of climate sensitivity: the 95th

percentile of climate sensitivity, S, is 1K higher for a uniform-S prior than for a uniform-λ prior (Sherwood et al., 2020). Both uniform-S and uniform-λ priors imply regions with infinite prior probability: a uniform-S prior implies that the probability of λ goes to infinity as λ goes to 0, while a uniform-λ prior implies that the probability of S goes to infinity as S goes to 0. To avoid these infinities, the priors are made discontinuous (Sherwood et al., 2020).

Here, new prior distributions for climate feedback parameters are presented and adopted, based on 5 conditions to prevent both infinities and discontinuities in prior probability. The following conditions on the prior distributions of climate feedback and climate sensitivity are imposed over all timescales:

Condition 1: The climate feedback and climate sensitivity each can take any positive value, $\lambda \in (0, +\infty)$ and $S \in (0, +\infty)$, with both having continuous probability distributions.
Condition 2: The probability of λ tends to zero as λ tends to zero, $p(\lambda) \to 0$ as $\lambda \to 0$, since the system's radiation balance must respond to surface temperature perturbations.
Condition 3: The probability of λ tends to zero as λ tends to infinity, $p(\lambda) \to 0$ as $\lambda \to \infty$, since the system's radiation balance and heat content responses to surface temperature perturbations must be finite.
Condition 4: The probability of S tends to zero as S tends to zero, $p(S) \to 0$ as $S \to 0$, since the system's global mean temperature must respond to a sustained perturbation of the radiation balance.
Condition 5: The probability of S tends to zero as S tends to infinity, $p(S) \to 0$ as $S \to \infty$, since the system's surface temperature responds in a finite way to perturbations in radiation balance.

Uniform-S and uniform-λ priors (e.g. considered in Sherwood et al., 2020) each break at least 3 of the 5 conditions imposed here. Goodwin and Cael (2021) defined three priors in terms of climate feedback terms acting on instantaneous (Planck), multiday (fast) and multidecadal timescales. They adopted a normal distribution for the instantaneous Planck climate feedback, and uniform

distributions for fast and multidecadal climate feedbacks. Again, these prior distributions employed in the WASP model in Goodwin and Cael (2021) break at least 3 of the 5 conditions adopted here.

Here, lognormal-based distributions (Appendix B; Fig. 2) for Planck feedback, λ_P (Fig. 2a), fast feedback, λ_f (Fig. 2b,e), and multidecadal climate feedback, λ_{md} (Appendix B: Fig. 2d,f), are chosen as being able to satisfy the above conditions 1 to 5; being defined only for positive values of total climate feedback and with probability distributions tending to zero as the total climate feedback value tends to both zero and infinity on all timescales.

The resulting prior distribution for equilibrium climate feedback, $\lambda_{equil} = \lambda_P + \lambda_f + \lambda_{md}$, has a median value of 3.3 $Wm^{-2}K^{-1}$, a 66% (17th to 83rd percentile) range of 1 to 9.4 $Wm^{-2}K^{-1}$ and a 95% (2.5th to 97.5 percentile) range of 0.38 to 42 $Wm^{-2}K^{-1}$ (Fig. 2g). The equilibrium climate sensitivity, $S_{equil} = a_{CO2}\ln 2/\lambda_{equil}$, then has a prior distribution with median 1.1 K, 66% range 0.37 to 3.6 K and 95% range 0.07 to 9.6 K (Fig. 2h). The prior distributions used here are less heavily weighted to high climate sensitivity values than a uniform S prior, and less heavily weighted to low climate sensitivity values than a uniform λ prior, and so represent a compromise between the uniform-S and uniform-λ priors considered by *Sherwood et al*. (2020).

The distributions of all other model parameters in the WASP model are set to the distributions used in Goodwin and Cael (2021). For example, the prior distribution of radiative forcing from aerosols in 2011 varies with a skew-normal distribution, with a 95% range from -1.7 to -0.1 Wm^{-2} (Fig. 2b) reflecting uncertainty in direct and indirect aerosol forcing (IPCC, 2013).

3.2 Selecting and weighting the posterior ensemble

Each of the prior ensemble members is integrated from preindustrial to the present day. A cost function is calculated by comparing a set of 11 simulated observable quantities to reconstructions from historical observations (Table 1). For each historical constraint, the probability of the simulated value being consistent with the observations, $p(\mathrm{obs})$, is calculated assuming Gaussian

uncertainty: $p(\text{obs}) \propto e^{-(\mu - x)^2/2\sigma^2}$, where μ is the observational best estimate, σ is the observational standard deviation and x is the simulated value of the observable. Note that Table 1 gives the median and 95% ranges for the 11 historical constraints, corredponding to μ, $\mu + 2\sigma$ and $\mu - 2\sigma$. The overall cost function for the ensemble member is then the product of each simulation's probability of being consistent with the 11 historical constraints,

$$\text{cost function} = \prod_{k=1}^{11} e^{\frac{-(\mu_k - x_k)^2}{2\sigma_k^2}} \quad (9)$$

Note that all 11 historical constraints are applied simultaneously in the cost function (9) so that the filter does not find parameter value combinations corresponding to a local optimum for one historical constraint if that produces very low probability for other constraints (Hourdin et al., 2017).

Here, a new method used to extract and weight a posterior ensemble, replacing the method used in Goodwin & Cael (2021). This new method requires fewer prior ensemble members to be generated, and so improves computational efficiency.

Each prior ensemble member's cost function is calculated (eq. 9). If the cost function is larger than a chosen automatic acceptance limit, then the ensemble member is accepted into the posterior ensemble and weighted by the ratio of the cost function to the automatic acceptance limit. If the cost function is lower than the automatic acceptance limit, then the ensemble member is accepted or rejected into the posterior ensemble as follows. First, the value of the cost function is compared to a number drawn from a random distribution between 0 and the automatic acceptance limit. If the cost function is greater than the random number, the simulation is accepted with a weighting of 1. If the cost function is less than the random number, the simulation is rejected.

Using this method, all simulations whose cost functions (eq. 9) are *x* times larger than the automatic acceptance limit are accepted into the posterior ensemble, and assigned a relative probabilistic weighting of *x*. Simulations whose cost functions are *y* times smaller than the automatic acceptance limit are *not* all accepted into the posterior with a probabilistic weighting of 1/*y*. Instead, they are represented in the posterior by every y^{th} simulation (randomly chosen from those whose cost function is *y* times smaller than the automatic acceptance limit), with a relative weighting of 1.

The automatic acceptance limit is set to give a balance between computational load and data storage requirements: larger values require more prior simulations to be conducted to generate a sufficient number of posterior ensemble members, while smaller values require more data storage for annual outputs of simulations that have very low relative weighting. Here, an automatic acceptance limit of 1×10^6 is setfor the cost function, giving 13340 posterior simulations from a prior ensemble size of 2.4×10^7. This leads to manageably sized datasets (order 30MB for annual output per scenario per simulated property, e.g. ΔT, λ, S) being produced from a manageable computational load (~100 CPU hours for the prior ensemble, and ~5 CPU minutes for subsequent scenarios with the posterior).

The method of posterior ensemble extraction and weighting, from a prior ensemble using a cost function (eq. 9; Table 1), has similarities to climate model tuning (Hourdin et al., 2017), especially Bayesian Uncertainty Quantification methods. However, here the climate model parameters are not being 'tuned' to extract some optimum parameter combination that is to be used for future numerical experiments. Instead, a large probabilistic posterior distribution of parameter value combinations (Fig. 2) are all integrated to generate joint probabilistic projections of future observables (Figs. 1-4).

3.3 Comparison of the posterior ensemble against complex models

For idealised forcing scenarios, the qualitative responses of global surface temperature anomaly, ΔT, and heat imbalance, N, are now explored in complex models and the efficient model ensemble, eq. (1). We are not considering the magnitude of the warming and heating responses, but the qualitative form of their time evolutions in CO_2 only forcing scenarios. Figure 5 shows the normalized warming and normalized heating responses for the efficient model ensemble (Fig. 5, blue lines and shading) and samples of complex model simulations from the CMIP6 ensemble (Fig. 5, grey lines) for sudden quadrupling of CO_2 ($4xCO_2$) and sustained 1 per cent per year increase in CO_2 (1pctCO2) scenarios. The warming is normalized relative to years 95 to 105 in each simulation for both scenarios (Fig. 5a, c), while the heating is normalized relative to years 95 to 105 in the 1pctCO2 scenario (Fig, 5b) and year 1 following CO_2 quadrupling in the 4xCO2 scenario (Fig. 5d).

The majority of complex model simulations lie within the spread of normalized warming and heating behaviours exhibited by the efficient model ensemble (Fig. 5, compare grey lines to blue lines and shaded areas). For the 1pctCO2 scenario, only a single complex model (grey dotted line in Fig 5a, b: NASA GISS-E2-1-G) shows qualitative behaviour significantly outside the variation found within the efficient model ensemble for surface warming: the warming gradient is significantly larger in the first half of the simulation than the second half; behaviour that is not reproduced within the efficient model ensemble.

The responses of the efficient model ensemble are qualitatively similar to complex climate models for idealised CO_2-only forcing scenarios (Fig. 5), and quantitatively similar to observations in the historic period (Fig. 1a, compare blue to black: Table 1), providing confidence in using the efficient model ensemble for making quantitative projections of future global mean surface warming (Fig. 1) and climate feedback (Fig. 3).

4 Inferred probabilistic projections of temperature and climate sensitivity

The joint probabilistic projections of climate sensitivity (eq. 3), climate feedback (eq. 4) and surface warming above preindustrial (eq. 2) from the observation-filtered posterior model ensemble are now presented. The probabilistic projections are inferred from the weighted frequency distributions within the posterior ensemble, and represent the probabilities given prior information, ,i and observational constraints (Table 1: $\Delta T(t_{hist})$,$\Delta \mathrm{OHC}(t_{hist})$).

4.1 Projections of climate feedback and climate sensitivity

The posterior distribution of Planck feedback is similar to the assumed prior (Fig. 2a). The radiative forcing from aerosols has a narrower posterior distribution than the prior (Fig. 2b), with the observational filter rejecting prior ensemble members with the most strongly negative aerosol forcing.

The prior distributions of λ_f and λ_{md} both contain equal amounts of amplifying ($\lambda < 0$) and damping ($\lambda > 0$) ensemble members (Fig. 2c-f), eqns. (9-12). However, the posterior distributions show amplifying fast feedback (Fig. 2c, e, red), with a large peak of weaker amplifying values and a smaller peak of strongly amplifying values. These two peaks correspond to a larger peak of amplifying multidecadal feedback (Fig. 2d, f, red, $\lambda < 0$) and a smaller peak of damping multidecadal feedback (Fig. 2d, f, red, $\lambda > 0$). This gives rise to a posterior distribution of equilibrium climate feedback, $\lambda_{equil} = \lambda_P + \lambda_f + \lambda_{md}$, with a large peak at 0.9 $\mathrm{Wm^{-2}K^{-1}}$ (weaker amplifying fast feedback and amplifying multidecadal feedback) and a smaller peak at 2.1 $\mathrm{Wm^{-2}K^{-1}}$ (stronger amplifying fast feedback and damping multidecadal feedback).

These posterior climate feedback distributions produce a posterior equilibrium climate sensitivity distribution with best estimate (median) value of $S_{equil} = 2.9$ K; a 66% range of 1.7 to 4.1 K and a 95% range of 1.4 to 5.0 K (Fig. 2h). Equilibrium climate sensitivity is often quoted at 90% confidence (IPCC, 2013), for which the range is 1.5 to 4.6 K (5th to 95th percentile). Note that both λ_{equil} and S_{equil} probability distributions here allow feedbacks on multidecadal and shorter

timescales to equilibrate fully to an imposed forcing, but ignore the impact of feedbacks that occur on timescales longer than multidecadal (e.g. Knutti et al., 2017; Rohling et al., 2018; 2012).

Now consider how future climate feedback and sensitivity are projected to evolve over time in response to the two extreme SSP scenarios used: SSP585 and SSP119 (Fig. 3). For SSP585, where the system is continually warming in response to increasing radiative forcing (Fig. 1a), future climate feedback and climate sensitivity are projected to stay close to their present-day uncertainty ranges (Fig. 3, blue), as the radiative forcing increases too quickly for the multidecadal feedbacks to equilibrate. For SSP119, where the rate of warming ceases as radiative forcing is strongly mitigated (Fig. 1c, red) the uncertainty in climate feedback and climate sensitivity increases going into the future as multidecadal feedbacks begin to equilibrate, with the median and uncertainty ranges approaching their equilibrium values towards year 2150 (Fig. 3, red, compare to Fig. 2g, h red).

By mid-21st century (2040-2050 average) surface warming above preindustrial is projected to reach: 1.9 (1.7 to 2.2) K under SSP585 [median (2.5th to 97.5th percentiles)]; 1.8 (1.7 to 2.1) K under SSP370; 1.8 (1.6 to 2.1) K under SSP245; 1.7 (1.4 to 2.0) K under SSP126; and 1.7 (1.3 to 2.0) K under SSP119. Toward the end of the 21st century (2090-2100 average) surface warming reaches: 4.0 (3.3 to 4.8) K under SSP585; 3.5 (2.9 to 4.1) K under SSP370; 2.7 (2.0 to 3.5) K under SSP245; 2.0 (1.3 to 2.7) K under SSP126; and 1.7 (1.0 to 2.3) K under SSP119 (Figs. 1, 4).

The median projections of average surface warming above preindustrial (1850-1900 average) reach or exceed 1.5 K average warming during the period 2030 to 2040 (Figs. 1, 4) for all SSP scenarios considered (Figs. 1, 4), and remain above 1.5K for the rest of the 21st century. Only the SSP119 and SSP126 scenarios, each containing significant and rapid mitigation, have median projections that remain at or below the Paris Accord's 2.0 K target (UNFCCC, 2015) over the 21st century (Figs. 1, 4).

4 Discussion

Consider how the probabilistic projections presented here, eqns. (2-4), compare to projections inferred from raw complex model ensemble output, eqns (5,6), and complex model ensemble output that has been adjusted or weighted by comparison to historical data, eqns. (7,8).

There is (considerably) more background information about the climate and Earth system put into the CMIP6 models, i_{CMIP} (eqns. 5-8), than into the efficient WASP box model, i (eqns. 2-4). A key question is then: Does the relative lack of climate information in the efficient WASP model ensemble impair the projections from capturing the qualitative global warming and heating responses of the climate system to radiative forcing, (eq. 1), relative to the more complex CMIP6 models? Previous studies have found that when the WASP model is trained against historic simulated warming and heat uptake ranges from a complex model ensemble, the future projection-ranges from WASP are in good agreement with those from the complex model ensemble (Goodwin, 2016; Goodwin et al., 2017). Here, this question is explored by comparing the normalised warming and heating responses of the posterior model ensemble to a sample of CMIP6 models for two idealised forcing scenarios (Fig. 5).

The posterior weighted WASP model ensemble is able to reasonably capture the qualitative global mean surface warming, ΔT, and heating, N, responses of the much more complex CMIP6 models for idealised scenarios with 1% per year CO_2 increase and instantaneous $4 \times CO_2$ forcing (Fig. 5, compare blue shading to grey lines; eqn. 1). This ability to reasonably capture the qualitative global average responses of complex climate models to idealised forcing scenarios provides confidence in using the efficient WASP model posterior ensemble for future projections. Future work will explore how to further improve a computationally efficient model ensemble's ability to capture the qualitative global warming and heating responses of complex models to forcing (Fig. 5). It should be noted that the WASP model framework (Goodwin, 2018) cannot currently simulate non-linear transitions in the climate system, and so the future projections (Figs. 1-4) assume that no such

transitions occur. It may be that complex models also underestimate the likelihood of future instabilities (Valdes, 2011).

Now, consider the ability of the posterior WASP ensemble and complex CMIP6 models to quantitatively capture the historically observed trajectories of ΔT and N (eq. 1; e.g. Table 1). The posterior model ensemble is more quantitatively representative of the observed historic global warming (Fig. 1, compare blue to black; Table 1) than the complex CMIP6 ensemble (Fig. 1, compare grey to black). The posterior WASP ensemble is also representative of the historic trajectory of N, as assessed by comparison of ocean heat content anomaly (Table 1). Individual CMIP6 models may agree well with historic surface warming (Fig. 1a). However, the CMIP6 ensemble does not contain enough simulations that agree well to sample all the possible ways that R and λ could have evolved to reproduce the observed trajectories of ΔT and N, eq. (1). Also, each CMIP model represents the best estimate of the model developers, and there is not a coordinated effort to sample parameter space across the different models.

In contrast, the prior ensemble generated here samples tens of millions of possible historic R and λ trajectory combinations (Fig. 2, black dashed lines) using a Monte Carlo approach with some assumed prior information (e.g. eqns. 9-12). From these prior ensemble members, those that lead to historic ΔT and N representative of observed warming and heat content anomalies are extracted and weighted to form the posterior (Fig. 1a; Fig. 2, red; Fig. 3; Table 1).

The 90% range in equilibrium climate sensitivity in the joint probability distribution generated here (1.5 to 4.6 K) is similar to the 'likely' range in IPCC (2013) of 1.5 to 4.5 K. The 50th, 83rd and 95th percentiles obtained here for S_{equil}, of 2.9, 4.1 and 4.6 K, are similar to the baseline calculations in the recent Sherwood et al. (2020) review, of 3.1, 3.9 and 4.7 K respectively. However, the 5th and 17th percentiles obtained here for S_{equil}, of 1.5 and 1.7 K, are less than the Sherwood et al. (2020) values, of 2.3 and 2.6 K respectively, and more in keeping with the earlier IPCC (2013) assessment. This agreement with the posterior probabilistic projection of equilibrium climate

sensitivity (Fig. 2h) to previous analyses (IPCC, 2013; Sherwood et al., 2020) provides confidence with the joint probabilistic projections of surface warming and climate feedback presented here (Figs. 1, 3, 4).

The agreement with observational estimates of warming and independent estimates of climate sensitivity support that the probabilistic projections presented here (Figs. 1-4; eqns. 2-4) should be preferred to any projections inferred directly from the raw CMIP6 output, without making any correction for CMIP6 ensemble's bias in historical simulation relative to observational reconstructions, eqns. (5, 6), (Fig. 1a, grey), or bias in climate sensitivity distribution relative to independent assessments (Zelinka et al., 2020; Sherwood et al., 2020).

The projections of climate sensitivity and surface warming produced here should be considered complementary to projections made by weighting or adjusting the raw CMIP6 output using real world data (eqns. 7, 8; e.g. Tokarska et al., 2020; Cox et al., 2018), because the approaches contains different sources of error. CMIP6 contains more process-based constraints within background information, whereas the projections presented here contain explicit information from observational reconstructions of global temperature and heat content anomalies (Table 1).

A key difference is that here, observational reconstructions of surface warming and ocean heat content anomaly are used in the extraction of the posterior model ensemble. This means that an ensemble member is always and only used to support the probability of surface warming above preindustrial, or climate sensitivity, being equal to the value the ensemble member simulates. In contrast, observational information is used to infer probabilistic projections from the CMIP ensembles (e.g. IPCC, 2013; Cox et al., 2018; Tokarska et al., 2020) after they has been generated. This means that the CMIP models are used to infer projections of surface warming above preindustrial and/or climate sensitivity that are different from the values simulated by those models (e.g. IPCC, 2013; Cox et al., 2018).

Now consider the wider implications of the projections produced here (Figs. 1-4). The posterior distributions find double-peaks in fast and multidecadal climate feedback strengths (Fig. 2). The question arises, how can we distinguish between these double-peaks to further constrain the system and reduce uncertainty in future projections? The larger of the two peaks have total feedback strength that becomes more amplifying from timescales of days to timescales of decades, with weaker amplifying fast feedbacks (Fig. 2c) and amplifying multidecadal feedbacks (Fig. 2d). These larger peaks are consistent with two previous findings from the literature. Firstly, it has been found that the short-timescale cooling from volcanic aerosols may be smaller than expected relative to the decadal trend in warming from greenhouse emissions (Gregory et al., 2016). Secondly, a majority of complex models tested show increasing amplifying feedbacks on multidecadal timescales from the sea surface temperature pattern effect (Andrews et al., 2015). Future work will consider how to rigorously combine the distributions identified here (Fig. 2) with other lines of evidence from the literature (e.g. Gregory et al., 2016; Andrews et al., 2015) to help reduce uncertainty in future warming and climate sensitivity.

The globally agreed Paris Climate Accord aims to restrict global mean surface warming to less 2.0 K warning above preindustrial, and preferably less than 1.5 K above preindustrial (UNFCCC, 2015). The results of this study indicate that the current decade may be the last with average temperature anomaly below the Paris Accord's 1.5 K target (UNFCCC, 2015), even with rapid and significant mitigation implemented as in scenarios SSP119 or SSP126. The priority now is mitigating enough to remain under the 2.0 K target, with an almost 50% chance of 11-year average warming exceeding 2.0 K this century under SSP126 and a very high likelihood (>95%) of exceeding 2.0 K under SSP245.

Appendices:

Appendix A: Choice of CMIP6 models in the sample

This study uses the sub-sample of CMIP6 models considered by *Nijsse et al.* (2020) for the specific scenario and property relevant to each figure. Table A1 contains the list of models included within each figure panel.

Appendix B: Climate feedback for uniform and spatially varying forcing

The global average outgoing radiation (both shortwave and longwave) at the top of the atmosphere, R_{out} in Wm^{-2}, is expressed using the Stefan-Boltzmann constant, σ_{SB} in Wm^{-2}K^{-4}, the global average surface temperature, T in K, the local greenhouse fraction, f (the global average longwave radiation escaping into space divided by the longwave radiation emitted from the surface), the average solar radiation incident per unit area across the Earth's surface, I_{solar} in Wm^{-2}, and the local albedo α, via,

$$R_{out} = 4\sigma_{SB} f T^4 + I_{solar}\alpha \tag{B1}$$

The total climate feedback λ_{total} (for the sign convention used here) is defined as the change in global outgoing radiation per unit global mean surface warming. Therefore, to express λ we differentiate R_{out} with respect to T, noting that σ_{SB} and I_{solar} are unaffected by surface temperatures,

$$\lambda_{total} = \frac{\partial \overline{R_{out}}}{\partial \overline{T}} = 4\sigma_{SB} f T^3 + \sigma T^4 \frac{\partial f}{\partial T} + I_{solar}\frac{\partial \alpha}{\partial T} \tag{B2}$$

where we note that the terms f, T, I_S and α refer to global averages.

It takes some finite amount of time for the greenhouse fraction, f, and the surface albedo, α, to change in response to a surface temperature, T. Thus, the instantaneous climate feedback

response, known as the Planck feedback, is given by $\lambda_P = 4\sigma_{SB} f T^3$, eq. (B2). Using an estimate of $f = 0.62 \pm 0.02$ (from the mean and standard deviation of 5 observational reconstructions in Trenberth et al., 2009), and global mean $T = 287 \pm 0.15$ K (from Jones and Harpham, 2013), gives $\lambda_P = 3.3 \pm 0.2$ Wm^{-2}K^{-1}. The contribution of $\sigma T^4 \frac{\partial f}{\partial T} + I_{solar} \frac{\partial \alpha}{\partial T}$ then evolves over different timescales as fast and multidecadal feedback processes respond. Here, we consider the impact of the instantaneous Planck feedback, λ_P, the fast climate feedback, λ_f, and the multidecadal feedback, λ_{md}, such that on long timescales, $\lambda_{total} = \lambda_P + \lambda_f + \lambda_{md}$.

Appendix C

This study adopts the notation for a lognormal distribution of $Lognormal(\mu,\sigma^2)$, where μ and σ are the mean and standard deviation of the underlying normal distribution such that e^{μ} is the median of the lognormal distribution and $\sigma^2 = \ln\left(1 + \sigma_{LN}^2/\mu_{LN}^2\right)$, where σ_{LN} and μ_{LN} are the standard deviation and mean of the lognormal distribution respectively.

Since the conditions 1 to 5 apply over all timescales, the conditions must apply separately to the each of the prior distributions for λ_P, $\lambda_P + \lambda_f$, and $\lambda_{equil} = \lambda_P + \lambda_f + \lambda_{md}$. For each climate feedback term, λ_P, λ_f and λ_{md}, a translated lognormal distribution is adopted with three choices to set: the distribution's minimum value, median value and uncertainty. For the jth climate feedback term this translated lognormal distribution is written,

$$\lambda_j \sim Lognormal\left(\ln\left(\mu_j - \lambda_{j:min}\right), \ln\left(1 + \frac{\sigma_j^2}{\left(\mu_j - \lambda_{j:min}\right)^2}\right)\right) + \lambda_{j:min} \quad \text{(C1)}$$

where $\lambda_{j:min}$ is the minimum value that the jth climate feedback can take to satisfy the five conditions on climate feedback and sensitivity priors, $\mu_j - \lambda_{j:min}$ is the difference between the distribution's median and the minimum values, and σ_j^2 is the variance of the distribution.

First, consider the Planck feedback term, λ_P. The instantaneous climate feedback must satisfy condition 1, $\lambda_P \in (0, +\infty)$, so the minimum possible value is $\lambda_{P:min} = 0$. From observational constraints on λ_P (Appendix B), we have a best estimate (median) of $\mu_P = 3.3$, and a standard deviation of $\sigma_P = 0.1$ (Sherwood et al., 2020; Goodwin and Cael, 2021). Substituting these values for $\lambda_{P:min}$, μ_P and σ_P into eq. (9) gives the prior probability distribution (Fig. 2a),

$$\lambda_P \sim Lognormal\left(\ln 3.3, \ln\left(1 + \frac{0.1^2}{3.3^2}\right)\right) \quad \text{(C2)}$$

For the fast and multidecadal feedback terms a position of ignorance is adopted to produce uninformed distributions. Now consider the fast feedback distribution, λ_f. Firstly, the total climate feedback on multiday timescales, the sum of the instantaneous plus fast feedback terms, must be positive: $\lambda_{Planck} + \lambda_f \in (0, +\infty)$. Thus, the minimum value that the fast climate feedback term can take depends on the value has been assigned to the Planck feedback, and is equal to minus the Planck feedback, $\lambda_{f:min} = -\lambda_P$.

Secondly, to adopt a prior position of ignorance about whether the fast feedback term will amplify or damp surface warming, the uninformed median of fast climate feedback distribution must be $\mu_f = 0$. This produces a distribution such that fast feedbacks have an equal prior likelihood of amplifying ($\lambda_f < 0$) and damping ($\lambda_f > 0$) surface warming. Thus, the difference between the median and minimum values of fast climate feedback distribution is also dependent on the value of the Planck feedback, $\mu_f - \lambda_{f:min} = \lambda_P$.

Lastly, the uncertainty in the fast feedback distribution must be set. Too small an uncertainty would put too much prior likelihood to small magnitude fast climate feedback values, $|\lambda_f|$, while too large uncertainty would place too much prior likelihood on large magnitude climate feedback values leading to extreme amplifying or damping of warming. To strike an uninformed balance it is assumed that the prior uncertainty in fast feedback scales with the difference between the median

and minimum, $\sigma_f^2 = (\mu_f - \lambda_{f:min})^2 = \lambda_P^2$. Thus, the standard deviation of the distribution is set to be equal to the magnitude of the maximum possible amplifying feedback, $\sigma_f = |-\lambda_P|$.

Substituting these values for $\lambda_{f:min}$, μ_f and σ_f into eq. (C1) gives the prior fast climate feedback distribution (Fig. 2c),

$$\lambda_f \sim Lognormal(\ln \lambda_P, \ln 2) - \lambda_P \quad \text{(C3)}$$

These same arguments are now applied for the multidecadal climate feedback distribution. The total climate feedback on multidecadal timescales, the sum of instantaneous, fast and multidecadal feedback terms, must be positive: $\lambda_P + \lambda_f + \lambda_{md} \in (0, +\infty)$, so $\lambda_{md:min} = -(\lambda_P + \lambda_f)$. A position of ignorance is assumed about whether multidecadal feedback will amplify or damp surface warming. Therefore, a best estimate (median) of $\mu_{md} = 0$ is chosen, such that fast feedbacks have an equal prior likelihood of amplifying warming and damping warming, and the difference between the median and minimum values is $\mu_{md} - \lambda_{md:min} = \lambda_P + \lambda_f$.

The uncertainty in multidecadal feedback scales with the difference between the median and minimum distribution values, $\sigma_{md}^2 = (\mu_{md} - \lambda_{md:min})^2 = (\lambda_P + \lambda_f)^2$. Again, this gives a standard deviation of multidecadal climate feedback distribution equal to the maximum amplifying multidecadal feedback value, $\sigma_{md} = |-\lambda_P - \lambda_f|$. Substituting these values for $\lambda_{md:min}$, μ_{md} and σ_{md} into eq. (9) gives the prior fast climate feedback distribution (Fig. 2d),

$$\lambda_{md} \sim Lognormal(\ln(\lambda_P + \lambda_f), \ln 2) - (\lambda_P + \lambda_f) \quad \text{(C4)}$$

For each simulation in the prior WASP model ensemble, the Planck feedback is selected first from a Lognormal distribution using eq. (C2). Then, the value of the Planck feedback is used to generate a fast climate feedback value from a Lognormal distribution using eq. (C3). Lastly, the

Planck and fast climate feedback values are used to generate a value for the multidecadal feedback from a Lognormal distribution using eq. (C4). This method produces prior distributions that satisfy the five conditions adopted here for climate sensitivity and climate feedback values over all timescales (Fig. 2).

Acknowledgements

PG acknowledges support from UKRI NERC grant NE/T010657/1 and the University of Southampton IRIDIS4 high performance computing cluster, on which the efficient model simulation was performed. The author thanks the groups providing CMIP6 model output used in this study (Table A1) made available via the World Climate Research Programme.

Data and code availability: CMIP6 model data is available for download through the World Climate Research Programme (WCRP) at: https://esgf-node.llnl.gov/projects/cmip6/ . The HadCRUT5 temperature reconstruction dataset is available for download at: https://www.metoffice.gov.uk/hadobs/hadcrut5/data/current/download.html . The efficient climate model (WASP version 3.0: Goodwin et al., 2018; Goodwin & Cael, 2020 discussion paper) used for the example efficient model ensemble is available for download at: https://doi.org/10.5281/zenodo.4088074 .

References:

Andrews, T., and Forster, P.M. (2020) Energy budget constraints on historical radiative forcing. *Nat. Clim. Chang.* 10, 313–316. https://doi.org/10.1038/s41558-020-0696-1

Andrews, T., Gregory, J.M., and Webb, M.J. (2015) The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. *J Clim* 28, 1630–1648. https://doi.org/10.1175/JCLI-D-14-00545.1.

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

Caldwell, P. M., Zelinka, M. D., and Klein, S. A. (2018) Evaluating Emergent Constraints on Equilibrium Climate Sensitivity, *J. Climate*, 31, 3921–3942, https://doi.org/10.1175/JCLI-D-17-0631.1

Cheng, L., Trenberth, K.E., Fasullo, J., Boyer, T., Abraham, J., Zhu, J. (2017) Improved estimates of ocean heat content from 1960 to 2015. *Science Advances*, 3, 3, e1601545, https://doi.org/10.1126/sciadv.1601545

Cox, P. M., Huntingford, C., and Williamson, M. S. (2018) Emergent constraint on equilibrium climate sensitivity from global temperature variability, *Nature*, 553, 319–322, https://doi.org/10.1038/nature25450

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

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S., Aragão, L. E. O. C., Arneth, A., Arora, V., Bates, N. R., Becker, M., Benoit-Cattin, A., Bittig, H. C., Bopp, L., Bultan, S., Chandra, N., Chevallier, F., Chini, L. P., Evans, W., Florentie, L., Forster, P. M., Gasser, T., Gehlen, M., Gilfillan, D., Gkritzalis, T., Gregor, L., Gruber, N., Harris, I., Hartung, K., Haverd, V., Houghton, R. A., Ilyina, T., Jain, A. K., Joetzjer, E., Kadono, K., Kato, E., Kitidis, V., Korsbakken, J. I., Landschützer, P., Lefèvre, N., Lenton, A., Lienert, S., Liu, Z., Lombardozzi, D.,

Marland, G., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Niwa, Y., O'Brien, K., Ono, T., Palmer, P. I., Pierrot, D., Poulter, B., Resplandy, L., Robertson, E., Rödenbeck, C., Schwinger, J., Séférian, R., Skjelvan, I., Smith, A. J. P., Sutton, A. J., Tanhua, T., Tans, P. P., Tian, H., Tilbrook, B., van der Werf, G., Vuichard, N., Walker, A. P., Wanninkhof, R., Watson, A. J., Willis, D., Wiltshire, A. J., Yuan, W., Yue, X., and Zaehle, S. (2020) Global Carbon Budget 2020, *Earth Syst. Sci. Data*, 12, 3269–3340, https://doi.org/10.5194/essd-12-3269-2020

Goodwin, P. and Cael, B.B. (2021) Bayesian estimation of Earth's climate sensitivity and transient climate response from observational warming and heat content datasets, *Earth System Dynamics*, 12, 709–723, 2021, https://doi.org/10.5194/esd-12-709-2021.

Goodwin, P. (2018). On the time evolution of climate sensitivity and future warming, *Earth's Future* 6, EFT2466, https://doi.org/10.1029/2018EF000889.

Goodwin, P., Haigh, I. D., Rohling E. J., and Slangen, A. (2017) A new method for projecting 21st century sea level rise and extremes, *Earth's Future* 5, 240–253, doi:10.1002/2016EF000508. http://onlinelibrary.wiley.com/doi/10.1002/2016EF000508/full

Goodwin, P. (2016) How historic simulation-observation discrepancy affects future warming projections in a very large model ensemble, *Climate Dynamics*, 47, 2219-2233, CLDY-D-15-00368R2, https://doi.org/10.1007/s00382-015-2960-z.

Gregory, J.M., Andrews, T., Ceppi, P., Mauritsen, T., Webb, M.J. (2019) How accurately can the climate sensitivity to CO_2 be estimated from historical climate change? *Climate Dynamics*, 54, 129-157, https://doi.org/10.1007/s00382-019-04991-y

Gregory, J. M., and Andrews, T. (2016) Variation in climate sensitivity and feedback parameters during the historical period, *Geophys. Res. Lett.,* 43, 3911—3920. https://doi.org/10.1002/2016GL068406

Gregory, J.M., Andrews, T., Good, P., Mauritsen, T., and Forster, P.M. (2016) Small global-mean cooling due to volcanic radiative forcing. *Climate Dynamics*, 47, 3979-3991, https://doi.org/10.1007/s00382-016-3055-1.

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

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., Rauser, F., Rio, C., Tomassini, L., Watanabe, M., and Williamson, D. (2017) The Art and Science of Climate Model Tuning. *Bulletin of the American Meteorological Society*, 98(3), 589-602. https://doi.org/10.1175/BAMS-D-15-00135.1

IPCC (2013) Climate Change 2013 The Physical Science Basis, in: *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, UK and New York, NY, USA, https://doi.org/10.1017/CBO9781107415324.

Jones, P. D. and Harpham, C. (2013) Estimation of the absolute surface air temperature of the Earth, *J. Geophys. Res.-Atmos.*, 118, 3213–3217, https://doi.org/10.1002/jgrd.50359

Kennedy, J. J., Rayner, N. A., Atkinson, C. P., and Killick, R. E. (2019) An ensemble data set of sea-surface temperature change from 1850: the Met Office Hadley Centre HadSST.4.0.0.0 data set. *Journal of Geophysical Research: Atmospheres*, 124. https://doi.org/10.1029/2018JD029867

Lenssen, N., Schmidt, G., Hansen, J., Menne, M., Persin, A., Ruedy, R., and Zyss, D. (2019) Improvements in the GISTEMP uncertainty model, *J. Geophys. Res.-Atmos.*, 124, 6307–6326, https://doi.org/10.1029/2018JD029522

Levitus, S., Antonov, J.I., Boyer, T.P., Baranova, O.K., Garcia, H.E., Locarnini, R.A., Mishonov, A.V., Reagan, J.R., Seidov, D., Yarosh, E.S. and Zweng, M.M. (2012) World ocean heat content and thermosteric sea levelchange (0–2000 m), 1955–2010. *Geophysical Research Letters*, 39, https://doi.org/10.1029/2012GL051106

Lewis, N. and Curry, J.A. (2014) The implications for climate sensitivity of AR5 forcing and heat uptake estimates. *Climate Dynamics*, 45, 1009–1023, https://doi.org/10.1007/s00382-014-2342-y

Morice, C.P., Kennedy, J.J., Rayner, N.A., Winn, J.P., Hogan, E., Killick, R.E., Dunn, R.J.H., Osborn, T.J., Jones, P.D., and Simpson, I.R. (2020) An updated assessment of near-surface temperature change from 1850: the HadCRUT5 dataset. *Journal of Geophysical Research*. doi:10.1029/2019JD032361

Nicholls, Z.R.J., Meinshausen, M., Lewis, J., Gieseke, R., Dommenget, D., Dorheim, K., Fan, C.-S., Fuglestvedt, J.S., Gasser, T., Golüke, U., Goodwin, P., Kriegler, E., Leach, N.J., Marchegiani, D., Quilcaille, Y., Samset, B.H., Sandstad, M., Shiklomanov, A.N., Skeie, R.B., Smith, C.J., Tanaka, K., Tsutsui, J., and Xie, Z. (2020) Reduced complexity model intercomparison project phase 1: Protocol, results and initial observations, *Geoscientific Model Development*, 13, 5175–5190, https://doi.org/10.5194/gmd-13-5175-2020.

Stocker, T. (2011) *Introduction to Climate Modelling, Advances in Geophysical and Environmental Mechanics and Mathematics,* Springer, Berlin, Germany, pp179, https://doi.org/10.1007/978-3-642-00773-6

Trenberth, K.E., Fasullo, J.T., and Balmaseda, M.A. (2014) Earth's Energy Imbalance, Journal of Climate, 27, 3129-3144. https://doi.org/10.1175/JCLI-D-13-00294.1.

Trenberth, K.E., Fasullo, J. T., and Kiehl, J. (2009) Earth's global energy budget, *Bulletin of the American Meteorological Society*, 90, 311-324, https://doi.org/10.1175/2008BAMS2634.1

Nijsse, F. J. M. M., Cox, P. M., and Williamson, M. S. (2020) Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models, *Earth Syst. Dynam.*, 11, 737–750, https://doi.org/10.5194/esd-11-737-2020

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

Otto, A., Otto, F.E.L., Boucher, O., Church, J., Hegerl, G., Forster, P.M., Gillet, N.P., Gregory, J., Johnson, G.C., Knutti, R., Lewis, N., Lohmann, U., Marotzke, J.,Myhre, G., Shindell, D., Stevens, B., Allen, M.R. (2013) Energy budget constraints on climate response. *Nature Geoscience*, 6, 415–416, https://doi.org/10.1038/ngeo1836

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., and Mosher, S. (2013) Berkeley Earth Temperature Averaging Process, *Geoinform. Geostat.*, 01, 1–13, https://doi.org/10.4172/2327-4581.1000103

Schlund, M., Lauer, A., Gentine, P., Sherwood, S. C., and Eyring, V. (2020) Emergent constraints on equilibrium climate sensitivity in CMIP5: do they hold for CMIP6?, *Earth Syst. Dynam.*, 11, 1233–1258, https://doi.org/10.5194/esd-11-1233-2020.

Sherwood, S., Webb, M.J., Annan, J.D., Armour, K.C., Forster, P.M., Hargreaves, J.C., Hegerl, G., Klein, S.A., Marvel, K.D., Rohling, E.J., Watanabe, M., Andrews, T., Braconnot, P., Bretherton, C.S., Foster, G.L., Hausfather, Z., von der Heydt, A.S., Knutti, R., Mauritsen, T., Norris, J.R., Proitosescu, C., Rugenstein, M., Schmidt, G.A., Tokarska, K.B., Zelinka, M.D. (2020) An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, 58, https://doi.org/10.1029/2019RG000678

Tokarska, K.B., Stolpe, M.B., Sippel, S., Fischer, E.M., Smith, C.J., Lehner, F., and Knutti, R. (2020) Past warming trend constrains future warming in CMIP6 models, *Science Advances* 6, 12, https://doi.org/10.1126/sciadv.aaz9549

UNFCCC (2015), Adoption of the Paris Agreement FCCC/CP/2015/L.9/Rev.1.

Valdes, P. (2011). Built for stability. *Nature Geoscience*, 4(7), 414-416. https://doi.org/10.1038/ngeo1200

Williams, K.D., Copsey, D., Blockley, E.W., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H.T., Hill, R., Hyder, P., Ineson, S., Johns, T.C., Keen, A.B., Lee, R.W., Megann, A., Milton, S.F., Rae, J.G.L., Roberts, M.J., Scaife, A.A., Schiemann, R., Storkey, D., Thorpe, L., Watterson, I.G., Walters, D.N., West, A., Wood, R.A., Woollings, T., and Xavier, P.K. (2017). 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, 357–380. https://doi.org/10.1002/2017MS001115

Zelinka M.D., Andrews T., Forster P.M., Taylor K.E. (2014) Quantifying components of aerosol-cloud-radiation interactions in climate models. *J Geophys Res* 119(12):7599–7615. https://doi.org/10.1002/2014jd021710

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

Figure 1: 11-year running mean surface temperature anomaly relative to the preindustrial (1850-1900) average from observations and simulations for a range of future scenarios. The HadCRUT5 observational dataset (black) shows the best estimate 11-year running mean (black line) and is shaded to the 11-year running mean of the 2.5th to 97.5th percentile range. Complex model simulations (see Appendix A for model selection) show *n*=125 individual CMIP6 model simulations for ssp585 (a: grey lines). The efficient model simulations show the median (colour lines) and 2.5th to 97.5th percentile ranges (colour shading) of the efficient model posterior ensemble, selected and weighted by an observational filter.

Figure 2: Probability and frequency density distributions for a selection of prior and posterior model properties. Shown are the assumed probability distributions for climate feedback terms (orange lines: a, c, d), frequency density distributions in the prior ensemble (black dotted lines) and weighted frequency density distributions in the posterior ensemble (red lines). Solid orange lines show the assumed lognormal probability distributions of climate feedback terms (panels a, c, d).

Fig. 3: Projected global climate feedback, λ (Wm^{-2}K^{-1}) and climate sensitivity, *S* (K), over time. Shown are the ensemble ranges for SSP585 future forcing (blue, solid line is median, dark shading and dashed lines 17th to 83rd percentile range, and light shading and dotted lines 2.5th to 97.5th percentile range) and SSP119 future forcing (red lines, solid line is median, dotted and

dashed lines are percentile ranges). Values calculated using 11-year average for the values of R, N and ΔT, eq. (1).

Figure 4: Cumulative probability distributions of global surface warming above the 1850-1900 period for SSP scenarios. Colours indicate 11-year average time periods from 2020-2030 to 2090-2100. Dotted lines indicate the time-period has >50% likelihood of warming being less than 1.5K target; dashed lines indicate >50% likelihood that surface warming < 2.0K target; and solid lines indicate >50% likelihood of warming exceeding the 2.0K target. Black dot-dashed line indicates the 2.0K target, and grey dot-dashed line indicates the 1.5K target.

Figure 5: The normalized warming and heating responses in complex and efficient model simulations for the 1pctCO2 and abrupt 4xCO2 idealised forcing scenarios. For both scenarios the normalised global mean surface warming (ΔT) and net downward top of atmosphere heating (N) responses show that the qualitative behaviour of the majority complex model simulations (solid grey lines) are encapsulated within the efficient model ensemble (blue line is ensemble median, dark and light blue shaded areas cover 17th to 83rd and 2.5th to 97.5th percentile ranges repsectively). For the 1pctCO2 scenario, a single complex model is shown as a grey dotted line (panels a and b: NASA GISS-E2-1-G). The normalized heating is the net downward heat flux at the top of the atmosphere (N in Wm^{-2}) and the normalized warming is the global mean surface temperature (ΔT in K). All values of warming and heating in panels (a), (b) and (c) are normalized relative to the average values during years 95 to 105 of the simulations. In panel (d) the net downward heating is normalized relative to the heating in year 1 following CO_2 increase.

Centre	Model	Inclusion in figures
AWI	AIW-CM-1-1-MR	Fig. 1a; Fig. 5c
BCC	BCC-CSM2-MR	Fig. 1a; Fig. 5a,b,c,d
BCC	BCC-ESM1	Fig. 5a,b,c,d
CAMS	CAMS-CSM1-0	Fig. 1a; Fig. 5a,c
CAS	FGOALS-fg-L	Fig. 1a; Fig. 5c
CAS	FGOALS-g3	Fig. 1a;
CCCma	CanESM5	Fig. 1a; Fig. 5a,b,c,d
CCCma	CanESM5-CanOE	Fig. 1a;
CNRM-CERFACS	CNRM-CM6-1	Fig. 1a; Fig. 5a,c
CNRM-CERFACS	CNRM-CM6-1-HR	Fig. 1a;
CNRM-CERFACS	CNRM-ESM2-1	Fig. 1a; Fig. 5a,c
CSIRO	ACCESS-ESM1-5	Fig. 1a; Fig. 5c,d
CSIRO-ARCCSS	ACCESS-CM2	Fig. 1a; Fig. 5c,d
DKRZ	MPI-ESM1-2-HR	Fig. 1a;
DWD	MPI-ESM1-2-HR	Fig. 1a;
E3SM-Project	E3SM-1-0	Fig. 5a,c
EC-Earth-Consortium	EC-Earth3	Fig. 1a; Fig. 5c
EC-Earth-Consortium	EC-Earth3-Veg	Fig. 1a; Fig. 5a,b,c,d
INM	INM-CM4-8	Fig. 1a; Fig. 5c
INM	INM-CM5-0	Fig. 1a; Fig. 5c
IPSL	IPSL-CM6A-LR	Fig. 1a; Fig. 5a,b,c,d
MIROC	MIROC-ES2L	Fig. 1a; Fig. 5c
MIROC	MIROC6	Fig. 1a; Fig. 5a,b,c,d
MOHC	HadGEM3-GC31-LL	Fig. 1a; Fig. 5a,b,c,d
MOHC	HadGEM3-GC31-MM	Fig. 5c,d
MOHC	UKESM1-0-LL	Fig. 1a; Fig. 5a,b,c,d

MPI-M	MPI-ESM1-2-LR	Fig. 1a;
MPI-M	MPI-ESM1-2-HR	Fig. 5c
MRI	MRI-ESM2-0	Fig. 1a; Fig. 5a,c
NASA-GISS	GISS-E2-1-G	Fig. 5a,b,c,d
NASA-GISS	GISS-E2-1-H	Fig. 5a,b,c,d
NCAR	CESM2	Fig. 1a; Fig. 5a,b,c,d
NCAR	CESM2-WACCM	Fig. 1a; Fig. 5a,b,c,d
NCC	NorESM2-LM	Fig. 1a; Fig. 5c
NCC	NorESM2-MM	Fig. 1a;
NOAA-GFDL	GFDL-CM4	Fig. 5a,c
NOAA-GFDL	GFDL-ESM4	Fig. 5c
NUIST	NESM3	Fig. 1a; Fig. 5c
SNU	SAM0-UNICON	Fig. 5a,b,c,d
UA	MCM-UA-1-0	Fig. 1a; Fig. 5c

Table A1: List of CMIP6 models included within this study, and the figure panels each model is included within. Some models contribute multiple realisations to a given figure panel.

Historic observational constraint	**Observational estimate used in cost function, median (2.5th to 97.5th percentile)**	**Weighted distribution in posterior ensemble, median (2.5th to 97.5th percentile)**
Δ*T* for 2008-2018 relative to 1850-1899	0.973 (0.825 to 1.121) K	1.026 (0.880 to 1.121) K
Δ*T* for 2008-2018 relative to 1900-1919	1.042 (0.908 to 1.176) K	0.961 (0.880 to 1.052) K
Δ*T* for 2008-2018 relative to 1920-1939	0.820 (0.698 to 0.942) K	0.804 (0.726 to 0.878) K
Δ*T* for 2008-2018 relative to 1940-1959	0.666 (0.538 to 0.794) K	0.665 (0.576 to 0.744) K
Δ*T* for 2008-2018 relative to 1960-1979	0.706 (0.652 to 0.760) K	0.695 (0.646 to 0.743) K
Δ*T* for 2008-2018 relative to 1980-1999	0.374 (0.323 to 0.420) K	0.400 (0.357 to 0.441) K
ΔSST for 1961-1990 relative to 1850-1899	0.281 (0.071 to 0.491) K	0.365 (0.212 to 0.499) K
ΔOHC, in the upper 700m of the ocean, for 1960-1969 relative to 2006-2015	177.8 (150.2 to 205.4) ZJ	175.5 (152.8 to 198.3) ZJ
ΔOHC, from 700m to 2000m depth, for 1960-1969 relative to 2006-2015	75.6 (51.0 to 100.2) ZJ	86.7 (68.0 to 108.3) ZJ
ΔOHC for the whole ocean, for 2016 relative to 1960	360 (290 to 430) ZJ	323 (280 to 365) ZJ

Whole ocean carbon content increase, from 1982 to the start of 2018	71.2 (22.6 to 119.8) PgC	70.4 (59.6 to 84.9) PgC

Table 1: Table of observational constraints for global mean surface temperature, ΔT, and global mean sea surface temperature, ΔSST, used in the observational filter from the example efficient model ensemble. These observational constraints are based on: the HadCRUT5 for ΔT; the HadSST4 dataset for $\Delta \mathrm{SST}$ (Kennedy et al., 2019); the Cheng et al. (2017) dataset for $\Delta \mathrm{OHC}$; and the Global Carbon Budget (Friedlingstein et al., 2020) for ocean carbon uptake. The likelihood of a simulation being consistent with each constraint is equal to the position of the simulated observable on the normal probability distribution.

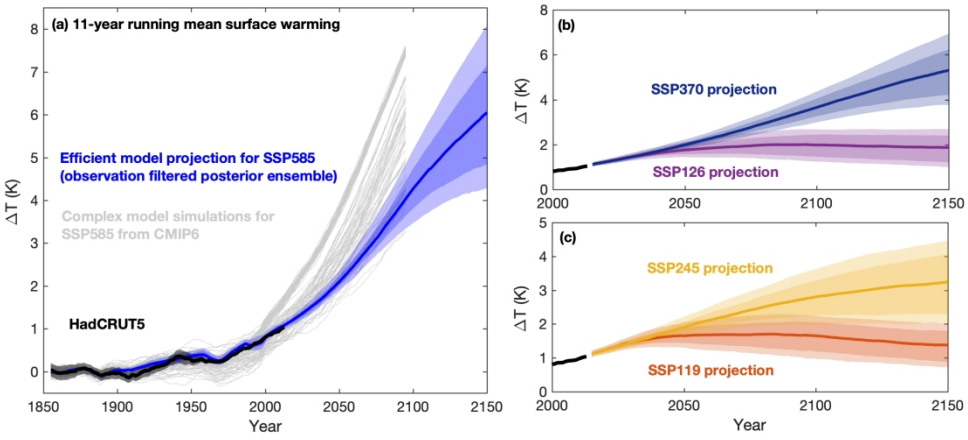

Figure 1

389x184mm (300 x 300 DPI)

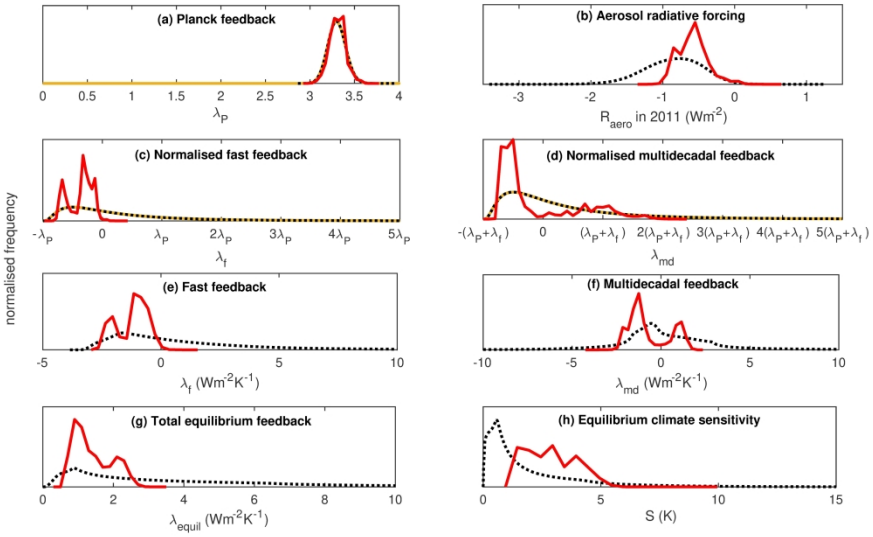

Figure 2

389x225mm (300 x 300 DPI)

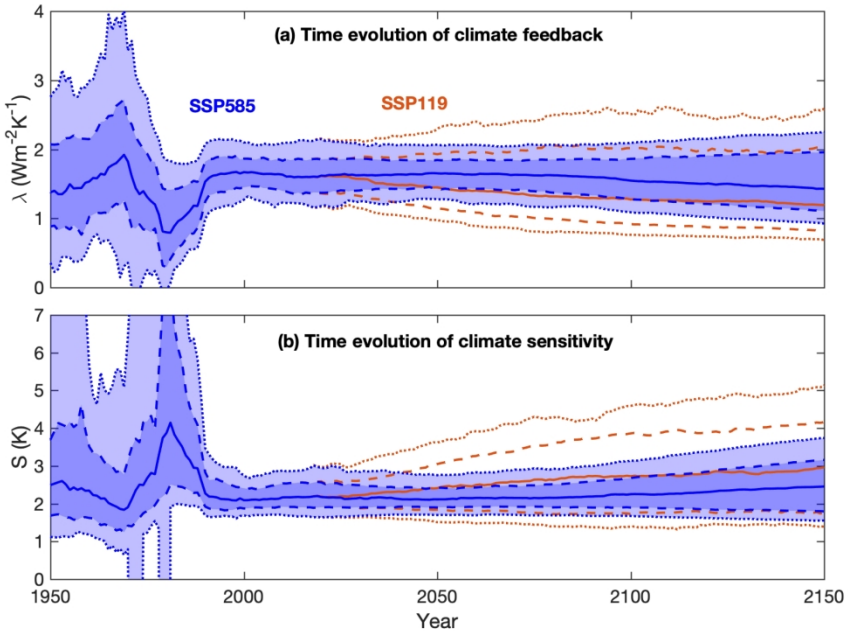

Figure 3

307x214mm (300 x 300 DPI)

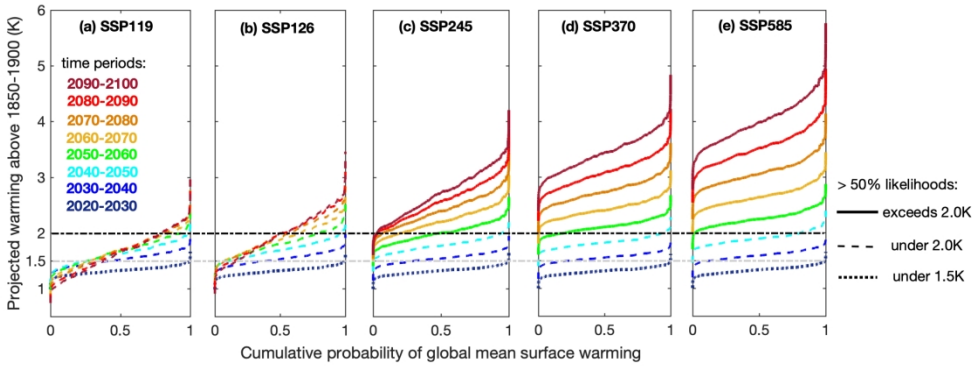

Figure 4

383x149mm (300 x 300 DPI)

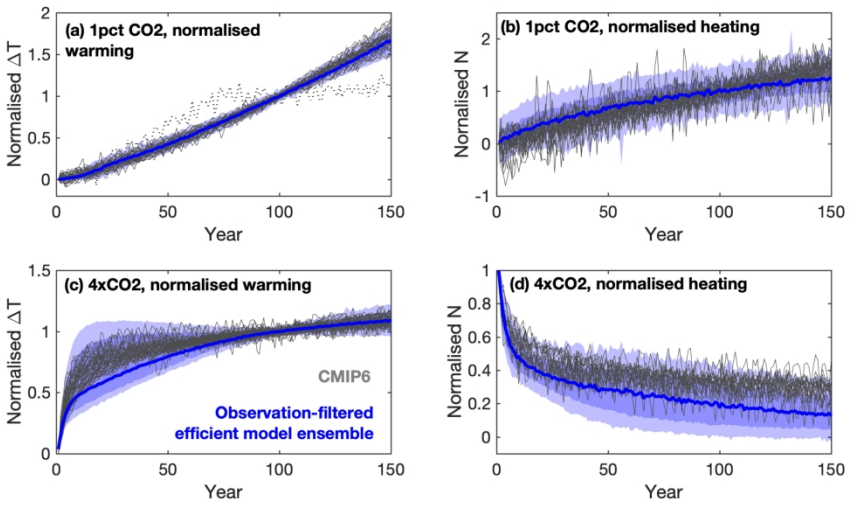

Figure 5

336x183mm (300 x 300 DPI)