Feedbacks, Pattern Effects, and Efficacies in a Large Ensemble of HadGEM3-GC3.1-LL Historical Simulations

Harry Mutton[1], Timothy Andrews[1], Leon Hermanson[2], Melissa Seabrook[2], Doug M Smith[1], Mark Adam Ringer[1], and Mark J Webb[3]

[1]Met Office
[2]Met Office Hadley Centre
[3]UK Met Office Hadley Centre

March 22, 2024

Abstract

Climate feedbacks over the historical period (1850–2014) have been investigated in large ensembles of historical, hist-ghg, hist-aer, and hist-nat experiments, with 47 members for each experiment. Across the historical ensemble with all forcings, a range in estimated Effective Climate Sensitivity (EffCS) between approximately 3–6 K is found, a considerable spread stemming solely from initial condition uncertainty. The spread in EffCS is associated with varying Sea Surface Temperature (SST) patterns seen across the ensemble due to their influence on different feedback processes. For example, the level of polar amplification is shown to strongly control the amount of sea ice melt per degree of global warming. This mechanism is responsible for the large spread in shortwave clear-sky feedbacks and is the main contributor to the different forcing efficacies seen across the different forcing agents, although in HadGEM3-GC3.1-LL these differences in forcing efficacy are shown to be small. The spread in other feedbacks is also investigated, with the level of tropical SST warming shown to strongly control the longwave clear-sky feedbacks, and the local surface-air-temperatures and large scale tropospheric temperatures shown to influence cloud feedbacks. The metrics used to understand the spread in feedbacks can also help to explain the disparity between feedbacks seen in the historical experiment simulations and a more accurate modeled estimate of the feedbacks seen in the real world derived from an atmosphere-only experiment prescribed with observed SSTs (termed amip-piForcing).

Feedbacks, Pattern Effects, and Efficacies in a Large Ensemble of HadGEM3-GC3.1-LL Historical Simulations

Harry Mutton[1], Timothy Andrews[1], Leon Hermanson[1], Melissa Seabrook[1], Doug M. Smith[1], Mark A. Ringer[1], Gareth S. Jones[1], Mark J. Webb[1]

[1]Met Office Hadley Centre, Exeter, UK

Key Points:

- Natural variability causes a 3-6K range in Effective Climate Sensitivity in a large single model ensemble of historical simulations.
- Differences in tropical and polar warming strongly influence longwave clear-sky and shortwave clear-sky feedbacks respectively.
- Deficiencies in simulating observed tropical and polar warming cause different feedbacks in historical and amip-piForcing experiments.

Corresponding author: Harry Mutton, `harry.mutton@metoffice.gov.uk`

Abstract

Climate feedbacks over the historical period (1850–2014) have been investigated in large ensembles of historical, hist-ghg, hist-aer, and hist-nat experiments, with 47 members for each experiment. Across the historical ensemble with all forcings, a range in estimated Effective Climate Sensitivity (EffCS) between approximately 3–6 K is found, a considerable spread stemming solely from initial condition uncertainty. The spread in EffCS is associated with varying Sea Surface Temperature (SST) patterns seen across the ensemble due to their influence on different feedback processes. For example, the level of polar amplification is shown to strongly control the amount of sea ice melt per degree of global warming. This mechanism is responsible for the large spread in shortwave clear-sky feedbacks and is the main contributor to the different forcing efficacies seen across the different forcing agents, although in HadGEM3-GC3.1-LL these differences in forcing efficacy are shown to be small. The spread in other feedbacks is also investigated, with the level of tropical SST warming shown to strongly control the longwave clear-sky feedbacks, and the local surface-air-temperatures and large scale tropospheric temperatures shown to influence cloud feedbacks. The metrics used to understand the spread in feedbacks can also help to explain the disparity between feedbacks seen in the historical experiment simulations and a more accurate modeled estimate of the feedbacks seen in the real world derived from an atmosphere-only experiment prescribed with observed SSTs (termed amip-piForcing).

Plain Language Summary

Understanding how the Earth's climate responds to an imposed forcing such as an increase in greenhouse gases or aerosols is an important issue relevant to climate mitigation and adaptation policies on the global scale. One way we can understand this is by analysing the historical period (1850–2014), a period over which the climate has already changed substantially due to human induced forcings, and also a period over which observations allow us to compare modeled changes in climate with the changes seen in the real world. Here, we use a large ensemble of climate model simulations of the historical period were we aim to understand a) how natural variability causes differences in the global temperature response to the same imposed forcing, b) what causes different forcing agents (e.g. greenhouse gases or aerosols) to be more or less effective at warming or cooling the planet, and c) whether historical simulations - where the climate model simulates its own sea surface temperatures - capture the same response to historical forcings as an atmosphere-only simulation prescribed with observed sea surface temperatures. We find that the pattern of sea surface temperatures (particularly the levels of tropical and polar warming) is key to understanding each of these points.

1 Introduction

Climate sensitivity and feedbacks provide valuable information about how the Earth's temperature changes in response to an imposed forcing such as an increase in greenhouse gases, aerosols, or volcanic emissions (Sherwood et al., 2020; Forster et al., 2021). Typically, equilibrium climate sensitivity (ECS) is defined as the equilibrium global temperature increase in response to a doubling of CO_2 and can be related to CO_2 forcing and climate feedbacks using a simple energy balance model (Equation 1) (e.g. Sherwood et al. (2020)).

$$ECS = -F_{2\times CO_2}/\lambda \tag{1}$$

Here, $F_{2\times CO_2}$ is the radiative forcing associated with a doubling of CO_2 and the feedback parameter λ is the radiative response per degree of global temperature change. Currently, the assessed likely range of ECS extends from 2.5°C – 4.0°C (Forster et al., 2021). Since constraining ECS is important for improving our understanding of how the Earth's climate is likely to change in the future, informing climate related mitigation and adaptation policy on the global scale, improving our understanding of different climate feedbacks and why they vary is vital.

The feedback parameter λ can be defined using Equation 2 (e.g. Gregory et al. (2004)).

$$\lambda = d(N - F)/dT_s \tag{2}$$

Here F is the radiative forcing, N is the top of atmosphere radiative flux, and T_s is the surface-air-temperature (in this case, all terms are global mean quantities).

In Atmosphere-Ocean General Circulation Models (AOGCMs), λ and ECS are typically estimated via a linear regression of global T_s and N over the first 150 years of an abrupt-4xCO2 simulation (T. Andrews et al., 2012; Dong et al., 2021; Gregory et al., 2004). The abrupt-4xCO2 simulation is an AOGCM experiment where the atmospheric concentration of CO_2 is abruptly quadrupled and then held constant. This regression method is used in favour of calculating ECS directly from two equilibrium states due to the long timescales needed to equilibrate the deep ocean and the substantial computational cost associated with this (T. Andrews et al., 2022; Rugenstein et al., 2019). ECS estimates produced from these non-equilibrium states are called the Effective Climate Sensitivity (EffCS) (Dong et al., 2021; Sherwood et al., 2020; T. Andrews et al., 2015; Rugenstein & Armour, 2021).

λ and EffCS can also be estimated from simulations of the historical record (1850 to present day), estimating λ over the historical period and applying this to Equation 1 where $F_{2\times CO_2}$ has been diagnosed from an abrupt-4xCO2 run (Gregory et al., 2020). These estimates tend to produce an EffCS smaller than that predicted solely from an abrupt-4xCO2 experiment, largely due to the time variations in λ caused by evolving SST patterns and the different timescales involved in the response to an imposed forcing (T. Andrews et al., 2019; Gregory et al., 2020; Proistosescu & Huybers, 2017). This "pattern effect" describes how a different global radiative response can be generated by the same global temperature change due to different patterns of SSTs (Rugenstein & Armour, 2021; Gregory & Andrews, 2016). In this context, the pattern effect is often quantified as the difference in λ between historical and abrupt-4xCO2 experiments (T. Andrews et al., 2018).

Estimates of λ from historical and abrupt-4xCO2 simulations may also differ due to the different forcing agents involved (Marvel et al., 2015). Whilst the abrupt-4xCO2 experiment is only forced by increases in CO_2 concentrations, the historical simulations are also influenced by changes in aerosols and natural forcings such as volcanic emissions (C. J. Smith & Forster, 2021; Salvi et al., 2023). These different forcing agents may vary in how effective they are at warming or cooling the planet; this is called forcing efficacy (Marvel et al., 2015; Richardson et al., 2019; Hansen et al., 2005). Again AOGCMs can be used to investigate this, with experiments simulating the historical period but only applying the forcing for individual forcing agents. Salvi et al. (2022) use this approach to demonstrate that, in the multi-model mean, greenhouse gases tended to have a more stabilising feedback (lower EffCS) compared to aerosols, although substantial variation across different models exists. It is suggested that across different forcing agents, variations in SST pattern changes lead to differing feedbacks (Haugstad et al., 2017). Ceppi and Gregory (2019) suggest that the changes in atmospheric stability induced by these differing SST patterns is a key factor determining the efficacy of a particular forcing (Salvi et al., 2023). Assuming temperature changes and the radiative responses to each forcing agent add linearly, understanding each component of the full historical forcing can help inform our interpretation of historical feedbacks and how they relate to future climate change.

Historical estimates of a model's EffCS can also be deduced from an Atmosphere only General Circulation Model (AGCM) experiment with prescribed SSTs and sea ice from observations between 1870 and 2014 and atmospheric constituents set to pre-industrial levels, termed amip-piForcing (Gregory & Andrews, 2016; Gregory et al., 2020). Because this experiment is forced with observed SSTs it is able to more accurately simulate historical changes in climate compared to the coupled AOGCMs (Gregory & Andrews, 2016). It is found that the EffCS calculated using the amip-piForcing experiment tends to produce an EffCS smaller than that derived from AOGCM historical experiments (i.e. amip-piForcing has a larger pattern effect relative to abrupt-4xCO2) (Gregory et al., 2020; T. Andrews et al., 2019). Again, this difference is often attributed to differences in SST patterns between the two experiments, with coupled historical simulations struggling to simulate observed SST patterns (Gregory et al., 2020; Wills et al., 2022). Over recent years, observed SSTs demonstrate a marked cooling in the East Pacific and Southern Ocean and more warming over the West Pacific, leading to more negative feedbacks and a lower EffCS. The inability of AOGCM simulations to capture observed trends in SST patterns is a key issue currently facing the scientific community and raises questions regarding how this impacts our understanding of climate sensitivity and feedbacks. The "peculiar" trend in SST patterns as termed by Fueglistaler and Silvers (2021) may have occurred through unforced variability and it may then be by chance that the real world SSTs have evolved in a way that induces a more strongly stabilising feedback. Or, it is possible that the trend is forced, e.g. by aerosols or volcanic emissions (D. Smith et al., 2016; Gregory et al., 2020; Hwang et al., 2024), and our AOGCMs struggle to simulate the real world SSTs accurately due to limitations in our current modelling capabilities.

To date, most of the work examining radiative feedbacks, pattern effects and efficacies has been limited to idealised experimental designs or small ensembles of historical AOGCM simulations with a single model, or via model intercomparisons such as the Coupled Model Intercomparison Project (CMIP) (Eyring et al., 2016), where still only relatively small ensemble sizes are available. Questions remain on the influence of natural variability in historical climate change on diagnosed estimates of feedbacks, the quantification of the forced response to different forcings and whether radiative feedback simulated in AOGCM historical simulations are consistent with observed estimates. Large initial condition ensembles with a single model are useful to address this. For example, previously, large ensembles have been shown to provide valuable insight into the separation of forced climate change and internal variability (Kay et al., 2015). From a sea ice sensitivity perspective, Kay et al. (2011) demonstrate that using an ensemble to quantify internal variability shows that recent trends in sea ice decline cannot be reproduced from modeled internal variability alone. Adams and Dessler (2019) employ a 100 member ensemble of historical simulations to show that internal variability could be a key contributor to the difference in Transient Climate Response (TCR) estimates between models and observations. Applying the analysis of this 100 member ensemble to the study of climate sensitivity and feedbacks over the historical period, Dessler et al. (2018) highlight a large range in EffCS estimates between 2.1 and 3.9K. They note that given that the real world 20$^{\mathrm{th}}$ century is just one realisation of a range of possible realities, due to that large internal variability, we should not expect estimates of EffCS from observations to be a precise guide to the real world's forced response. Alongside this, they note that that different forcing efficacies, imperfect observations, and uncertainty in 20th century forcing also pose challenges for interpreting EffCS from the historical period. Gregory et al. (2020) also noted the high levels of internal variability over the historical record showing how this variability contributed to uncertainty to estimates of EffCS.

In this paper we use a new set of four large ensembles of HadGEM3-GC3.1-LL historical and single forcing simulations performed for the Large Ensemble Single Forcing Model Intercomparison Project (LESFMIP) (D. Smith et al., 2022), aiming to address the following questions.

1. how does natural variability cause differences and spread in climate feedbacks in response to the same imposed forcing?
2. What causes different efficacies of different historical forcing agents?
3. Can AOGCM historical simulations – where the model simulates it's own SSTs - capture the radiative feedback and EffCS estimated from AGCM experiments prescribed with observed SSTs?

Previously, T. Andrews et al. (2019) investigated EffCS and feedbacks in HadGEM3-GC3.1-LL in a 4 member ensemble of historical simulations, finding a net feedback (λ) of -0.86 $\pm$ 0.4 $Wm^{-2}K^{-1}$ (5-95%). This ensemble mean estimate is more negative than the abrupt-4$\times$CO2 feedback in HadGEM3-GC3.1-LL of -0.63 $Wm^{-2}K^{-1}$, although the 5-95% confidence range does extend up to -0.46 $Wm^{-2}K^{-1}$. The large spread in λ was found to be partly caused by considerable variations in Antarctic sea ice. This variability in sea ice inhibited accurate evaluation of the model's historical forced EffCS. There, T. Andrews et al. (2019) were limited to an ensemble of only 4 simulations, so questions remain about whether the full diversity of variability was sampled. Here we investigate this with a much larger ensemble of 47 members.

In the following section we describe the model and experimental setup used. Section 3 presents the results and Section 4 provides a discussion and conclusions.

2 Methods

2.1 HadGEM3-GC3.1-LL

The analysis in this paper uses simulations performed using HadGEM3-GC3.1-LL, an AOGCM with an atmospheric resolution of 135 km with 85 vertical levels and an ocean resolution of 1°and 75 vertical levels (M. B. Andrews et al., 2020). Further details can be found in Williams et al. (2017) where a description of the model's configuration is given.

2.2 Large Historical Ensemble

In this analysis, ensembles of historical, hist-ghg, hist-aer, and hist-nat experiment are used, with 47 members of each experiment mostly consisting of simulations performed for LESFMIP. These experiments are AOGCM simulations analysed between 1850–2014 with atmospheric constituents set to historical levels. Here, the historical experiment includes all forcing agents, whilst the hist-ghg, hist-aer, and hist-nat contain only the forcing associated with well mixed greenhouse gases, anthropogenic aerosols, and natural forcings respectively (Gillett et al., 2016). Each ensemble member differs only in their initial conditions branching from the piControl experiment at different times (1850, 1885, and every 10 years between 1860 and 2300). The piControl experiment is an AOGCM experiment with atmospheric constituents set to pre-industrial levels. The 47 ensemble members consist of 45 simulations performed as part of the LESFMIP ensemble (D. Smith et al., 2022), and two simulations previously analysed in T. Andrews et al. (2019). Only two of the four simulations used in T. Andrews et al. (2019) were analysed here since the other two members had identical branch times to members of the LESFMIP ensemble.

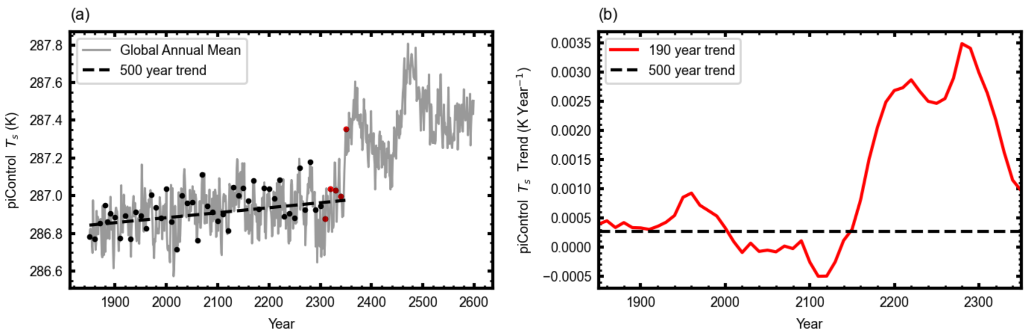

Figure 1. (a) Timeseries of global annual mean T_s in the piControl experiment (grey line), 500 year trend (dashed black line), and branch times for each of the historical and single forcing experiment ensemble members (dots). Red dots indicate the ensemble members that have been excluded due to the strong warming seen in the piControl experiment. (b) 190 year piControl trend for each ensemble member branch date (red), and 500 year piControl trend (horizontal black dashed line).

2.3 piControl and Detrending

To compare ensemble members in the 47 member ensembles, the control drift must be removed from each simulation. For this analysis, this drift is removed by calculating the trend over the first 500 years of the piControl experiment via linear regression and subtracting the corresponding time period from each ensemble member. The piControl timeseries of global annual mean T_s and the 500 year trend is shown in Figure 1a where the dots depict the branch dates for each member of the historical ensemble. This method of control drift removal is chosen in favour of subtracting the piControl year by year to avoid unnecessarily introducing more noise into the historical simulations. The 500 year trend is also preferred above subtracting the 190 year trend across the corresponding piControl period due to issues introduced towards the end of the piControl simulation, where a marked global warming is seen at around 2350. This warming has been previously documented by Ridley et al. (2022) where it is attributed to the onset of deep convection in the Weddell and Ross Sea gyres due to a destabilising of the Southern Ocean. When removing the control drift from the historical ensemble, any drift removed is assumed to be present in the historical ensemble member. For the trend seen over the first 500 years of the control run this is a reasonable assumption, however in the case of the large warming seen around 2350, this assumption may not hold. The impact that this warming has on the 190 year control trend for the respective historical ensemble branch dates is shown in Figure 1b. Here, unsurprisingly, a strong positive trend is seen for ensemble members that branch after the year 2150. We found no evidence that the warming seen in the piControl experiment is present in historical ensemble members initiated up to 2300, but to avoid this feature contaminating the comparison of ensemble members, the last 5 ensemble members have been removed from the analysis. This is why although the LESFMIP ensemble consists of 50 members, only 45 of them are used here.

2.4 Diagnosing Historical Forcing

Whilst λ can be calculated for the abrupt-4xCO2 and amip-piForcing experiments from only T_s and N (since the F is constant), the time varying F over the historical period means that in order to estimate λ, we must first diagnose F.

Table 1. Description of experimental setup used.

Experiments				
Experiment Name	Atmospheric Constituents	SSTs	Run Time	Ensemble Size
Coupled experiments				
piControl	pre-industrial	free running	1850 – 3850	1
abrupt-4xCO2	pre-industrial $CO_2 \times 4$	free running	1850 – 2350	1
historical	historical	free running	1850–2014	47
hist-ghg	historical well mixed greenhouse gases	free running	1850–2014	47
hist-aer	historical aerosols	free running	1850–2014	47
hist-nat	historical natural forcing	free running	1850–2014	47
Atmosphere-only experiments				
amip-piForcing	pre-industrial	historical observed	1870 – 2014	1
piClim-control	pre-industrial	piControl	1850 – 1890	3
piClim-histall	historical to 2014 then ssp-245 to 2100	piControl	1850 – 2100	3
piClim-histghg	historical well mixed greenhouse gases only to 2014 then ssp-245 to 2100	piControl	1850 – 2100	3
piClim-histaer	historical aerosols only to 2014 then ssp-245 to 2100	piControl	1850 – 2100	3
piClim-histnat	historical natural forcing only to 2014 then ssp-245 to 2100	piControl	1850 – 2100	3

Typically, the historical F is diagnosed using RFMIP experiments piClim-control and piClim-histall (Forster et al., 2016; Pincus et al., 2016). These are two AGCM experiments with prescribed SSTs from the piControl simulation. For piClim-control, atmospheric constituents are set to pre-industrial levels and the experiment is run for 30 years. Averaging over the 30 years provides the control state. For piClim-histall atmospheric constituents are set to historical levels between 1850 – 2014 and to ssp-245 levels between 2015 and 2100. Subtracting the 30 year mean piClim-control top of atmosphere radiative flux from the 1850 – 2100 piClim-histall top of atmosphere flux provides F, with years 1850–2014 relevant for the analysis of the historical period.

In order to diagnose F for the individual forcing components that correspond to the hist-ghg, hist-aer, and hist-nat experiments, a similar experimental setup to the piClim-histall experiment is used but only the forcing from the relevant component is applied. These experiments are termed piClim-histghg, piClim-histaer, and piClim-histnat (Forster et al., 2016; Pincus et al., 2016).

A summary of the setup for each experiment used in this paper is presented in Table 1.

3 Results

3.1 Diagnosing Feedbacks in Historical and Single Forcing Ensembles

As discussed in the introduction, the feedback parameter (λ) can be estimated via linear regression of global annual mean surface-air-temperatures (T_s) against top of atmosphere radiative fluxes (N) minus the changes in flux associated with the radiative forcing (F). Timeseries of these diagnostics are presented in Figure 2, where 2a and b show the anomalous global annual mean T_s and anomalous global annual mean N respectively in every ensemble member and in each experiment, and 2c shows the global annual mean F associated with each experiment. From Figure 2a it can be seen that the cooling effect of anthropogenic aerosols and natural forcings is approximately offset by the warming effect of increased greenhouse gases between 1850 and 1990. Here, the F associated with greenhouse gases and aerosols gradually increase, however, after approximately 1990 the aerosol F remains relatively constant (around -1.5 Wm^{-2}) whilst the F associated with greenhouse gases continues to increase (Figure 2c) (T. Andrews et al., 2019). This leads to a net positive F after 1990 in the historical experiment which results in an increase in global mean T_s, warming by approximately 0.8 K by 2014. A detailed analysis of HadGEM3-GC3.1-LL historical simulations is presented in M. B. Andrews et al. (2020). An example of how λ is calculated from these timeseries of T_s, N, and F is presented in Figure 2d, where, for the first ensemble member in the historical experiment, a feedback parameter of -0.85 $\pm$ 0.15 Wm^{-2}K^{-1} is estimated. There the uncertainty is estimated as $\pm$1.645 standard deviations, calculated from the standard error of the linear fit.

One assumption made when estimating λ using timeseries of T_s, N, and F is that the changes in global mean T_s associated with the forcing is zero (i.e. the surface-air-temperature change between piClim-control and piClim-histall is zero). This is generally a reasonable assumption to make, given that the prescribed SSTs do not warm and therefore any changes in land surface temperatures are constrained to be small (Lambert et al., 2011). However, despite this temperature change being small, taking this into account can substantially affect the values of λ estimated. This caveat is noted in Hansen et al. (2005) and Vial et al. (2013) and becomes a particularly relevant issue when comparing feedbacks in the historical experiment to feedbacks in the amip-piForcing experiment, since there is no forced temperature change in the amip-piForcing experiment where $F = 0$ by construction. To handle this issue, in this paper, λ has been calculated accounting for this forced temperature change (Equation 3).

$$\lambda = d(N - F)/d(T_s - \delta T_{s_{forced}}) \tag{3}$$

Where $\delta T_{s_{forced}}$ is calculated as the change in global surface-air-temperature between piClim-control and the relevant piClim-hist experiment used to diagnose F. To simplify the notation, we refer to ($T_s - \delta T_{s_{forced}}$) simply as T_s. Similarly, later when analysing atmospheric temperatures (T_a), we refer to ($T_a - \delta T_{a_{forced}}$) simply as T_a.

To summarise the feedbacks seen across the different experiments analysed, boxplots of feedbacks in the historical and single forcing experiments and markers showing the feedbacks in both amip-piForcing and abrupt-4xCO2 experiments are shown in Figure 3b. Here the net feedback has been decomposed into shortwave clear-sky (SW_{cs}), longwave clear-sky (LW_{cs}), and cloud radiative effect (cre) components. Such a decomposition is useful since it can help isolate the different processes and feedback mechanisms involved. λ_{SWcs}, λ_{LWcs}, and λ_{cre} are calculated by decomposing N and F into the relevant fluxes when applying Equation 3. From Figure 3b, a large spread in feedbacks across the historical ensemble can be seen, ranging from approximately -0.7 to -1.3 Wm^{-2}K^{-1}. Using a 2$\times CO_2$ F of 4.05 Wm^{-2} for HadGEM3-GC3.1-LL (T. Andrews

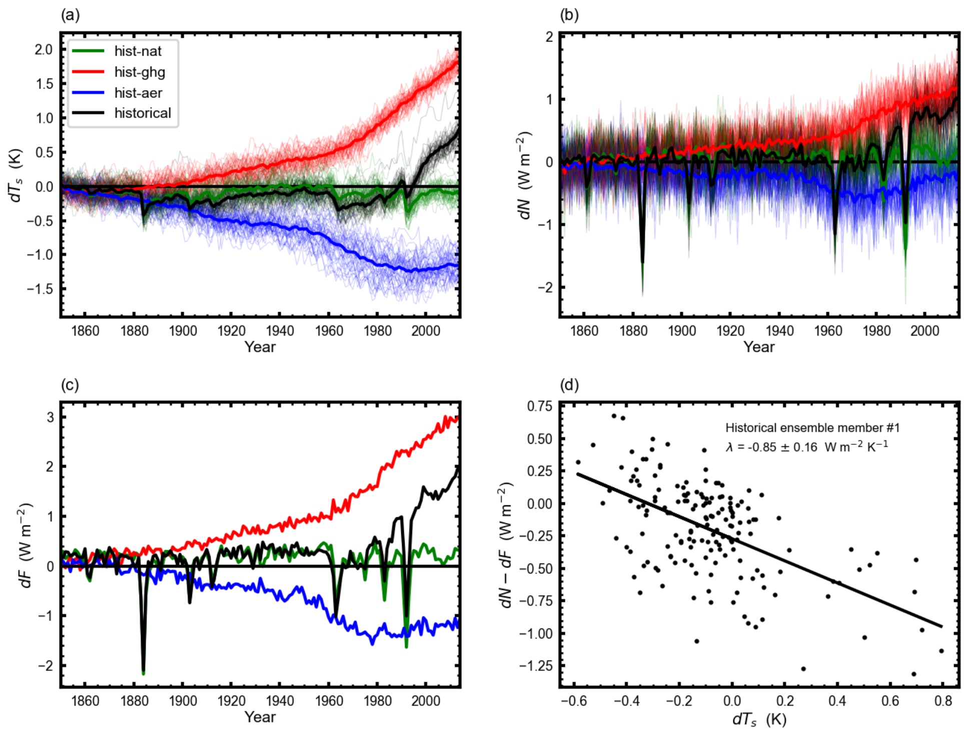

Figure 2. (a) Timeseries of anomalous global annual mean T_s in the historical and single forcing experiments. Thick lines indicate the ensemble mean and thinner lines represent each individual ensemble member. (b) Timeseries of anomalous global annual mean N in the historical and single forcing experiments. Again, thick lines indicate the ensemble mean and thinner lines represent each individual ensemble member. (c) Timeseries of global annual mean F for historical and single forcing scenarios averaged across the three ensemble members for each experiment. (d) Example of method used to estimate λ, where λ is calculated by linearly regressing T_s against $(N - F)$. Each dot represents a year in the historical experiment and the black line shows regression line where the slope (λ) is estimated to be -0.85 $\pm$ 0.15 Wm^{-2}K^{-1}. Here, the uncertainty is estimated as $\pm$1.645 standard deviations, calculated from the standard error of the linear fit.

et al., 2019), and applying Equation 1, such a range in feedbacks leads to an estimate of EffCS between approximately 3 and 6K (Figure 3a). This highlights the role of internal variability in causing different feedback and EffCS estimates over the historical period. The spread in feedbacks seen in the historical and single forcing experiments is largest in the hist-nat experiment and smallest in the hist-ghg experiment, possibly due to the varying signal to noise ratios across the different experiments. The T_s changes in the hist-nat experiment are generally small (Figure 2a), and the natural F is also small with an occasional strong but short-lived signal caused by volcanic emissions (Figure 2c). This causes the regression of T_s against $(N - F)$ to be relatively noisy compared to the hist-ghg experiment where both T_s and $(N - F)$ have a much stronger signal. This is also consistent with the contrast in estimated uncertainty of the linear fit of T_s and $(N - F)$ where for each experiment, the standard error of the linear fit of every ensemble member has be estimated. The estimation of λ_{net} in the hist-ghg experiment has an average 5-95% interval of $\pm$ 0.066 Wm^{-2}K^{-1}, whereas for hist-nat, the mean 5-95% interval is $\pm$0.25 Wm^{-2}K^{-1}.

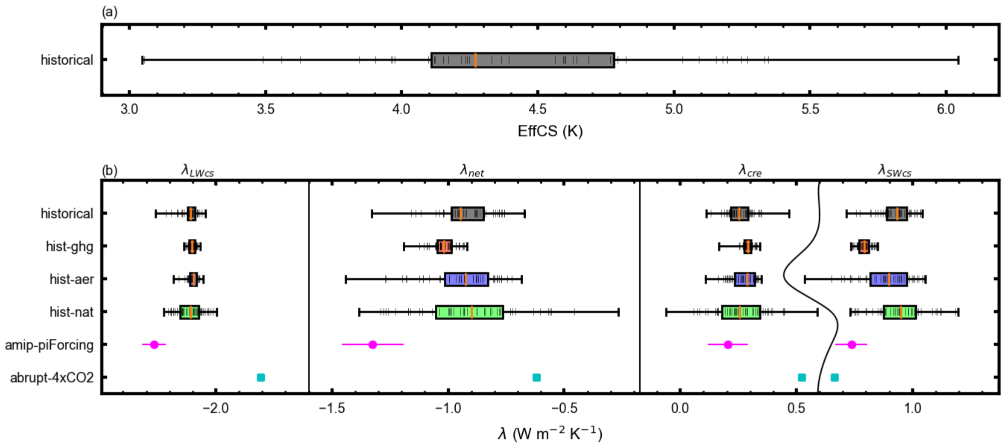

Figure 3. (a) Boxplot of EffCS across the historical ensemble (1850–2014). (b) Boxplots of feedbacks in the historical and single forcing ensembles (1850–2014), amip-piForcing experiment (1870–2014), and abrupt-4xCO2 experiment (first 150 years). For each boxplot, the vertical black lines indicate each ensemble member, the whiskers indicate the maximum and minimum feedbacks seen in the ensemble, the boxes indicate the interquartile range, and the vertical orange line represents the median value. Error bars on amip-piForcing indicate the 5-95% confidence interval, calculated from the standard error of the linear fit.

A further decomposition of λ_{cre} into shortwave and longwave components is shown in Figure S1. There, the largest contribution to the spread in λ_{cre} comes from the shortwave component, consistent with the strong influence of low cloud feedbacks, and the cancelling of the longwave and shortwave response to changes in high cloud.

The feedbacks seen in each historical and single forcing experiment are largely consistent with each other (i.e. differing forcing efficacies do not appear to be strongly evident in HadGEM3-GC3.1-LL), although a slightly more negative median feedback is seen in the hist-ghg experiment, consistent with the findings of Salvi et al. (2022). In Figure 3, the more negative median feedback in the hist-ghg experiment is shown to be caused by a weaker λ_{SWcs}, although due to the large spread in historical, hist-aer, and hist-nat feedbacks, the lower tails of the feedbacks in these experiments extend to be more negative than the lower tail of the hist-ghg experiment. The amip-piForcing and abrupt-4xCO2 feedbacks are also shown in Figure 3b. For each component of λ_{net}, the amip-piForcing feedback lies towards the lower tail of the historical ensemble, a behaviour most strongly seen in the λ_{SWcs}, and λ_{LWcs} components.

Maps of the ensemble mean feedbacks and amip-piForcing feedbacks are shown in Figure 4 to help identify where different feedbacks are located and to highlight regions where feedbacks differ across the range of experiments analysed. The spatial feedback map is calculated by regressing the local $(N-F)$ against the global mean T_s changes. Here the ensemble mean feedbacks are calculated by taking the regression of the mean rather than calculating the feedback for each ensemble member and averaging across the ensemble. This was done to help reduce the noise in the regression of $(N-F)$ and T_s when calculating the feedbacks.

From Figure 4, it can be seen that different feedbacks dominate in different regions. For example, in general λ_{SWcs} is strongly positive at higher latitudes and small at lower latitudes. This is because the sea ice feedback is a key feedback affecting the SW_{cs} fluxes. The strong positive λ_{SWcs} seen over the northern hemisphere land masses is likely related to snow and land ice feedbacks, and the strong negative λ_{SWcs} seen in the Southern Ocean in the hist-aer experiment may be caused by ocean convective events that bring warmer water to the surface due to destabilization of the ocean, similar to those discussed in (Ridley et al., 2022).

With the exception of the Southern Ocean feature seen in the hist-aer experiment, the λ_{LWcs} is generally negative everywhere across all experiments, although a few small regions in the amip-piForcing experiment also have positive λ_{LWcs}. The λ_{LWcs} is largely composed of the Planck, lapse rate, and water vapour feedbacks. This term is generally large and negative due to the strong Planck response. Over the Southern Ocean in the hist-aer experiment, since this region warms, which is of opposite sign to the cooling seen over the rest of the planet, the λ_{LWcs} is strongly positive in this region. In the tropics, the lapse rate and Planck feedbacks are typically negative, therefore the positive λ_{LWcs} regions in the amip-piForcing experiment over the tropics are likely caused by the water vapour feedback (Stephens et al., 2016).

λ_{cre} exhibits relatively large spatial variations. In the historical and single forcing experiments (particularly hist-aer) a strongly positive λ_{cre} is seen over the North Pacific, highlighting the role of positive cloud feedbacks in the sub-tropical cloud decks in subsidence regions. Again, λ_{cre} has been decomposed into longwave and shortwave components (Figure S2). The strong λ_{cre} over the North Pacific is caused by shortwave cloud feedbacks, and over tropical high cloud regions, e.g. the Indo-Pacific warm pool region, strong shortwave and longwave cloud feedbacks cancel, causing the relatively weak λ_{cre} over much of the tropics.

From these maps of feedbacks, it can be seen that although in the global mean, different efficacies are not particularly large in HadGEM3-GC3.1-LL, spatially, large variations do exist between the different experiments.

As mentioned in the introduction, differences in feedbacks across experiments and ensemble members are generally thought to be fundamentally caused by differing SST patterns. Therefore, to help understand the differences in feedbacks seen in Figure 4, ensemble mean T_s patterns are shown in Figure 5. Similar to the maps of λ, these have been calculated by regressing the ensemble mean local changes in T_s against the ensemble mean global mean T_s, written as $dT_s/d\bar{T_s}$, where the bar indicates a global mean. In Figure 5, the strongest regions of $dT_s/d\bar{T_s}$ occur in the Arctic, with weaker more spatially uniform $dT_s/d\bar{T_s}$ seen over the tropics. Over the Southern Ocean, large variations in $dT_s/d\bar{T_s}$ are seen across the different experiments. Here, hist-nat exhibits the strongest $dT_s/d\bar{T_s}$ whilst hist-aer exhibits a negative $dT_s/d\bar{T_s}$ (i.e. although global mean T_s is decreasing in the hist-aer experiment, the southern ocean warms). As previously mentioned, this may be caused by ocean convective events that bring warmer water to the surface due to destabilization of the ocean (Ridley et al., 2022). In the northern hemisphere high latitudes, hist-aer exhibits the strongest $dT_s/d\bar{T_s}$, possibly due to the aerosol F being predominantly based in the northern hemisphere. Over the tropics $dT_s/d\bar{T_s}$ is relatively consistent across each experiment.

Since one of the key aims of this paper is to understand the ensemble spread in feedbacks, maps of the standard deviation in λ in the historical experiment help to highlight the regions that contribute most to this spread (Figure 6). From Figure 6 it can be seen

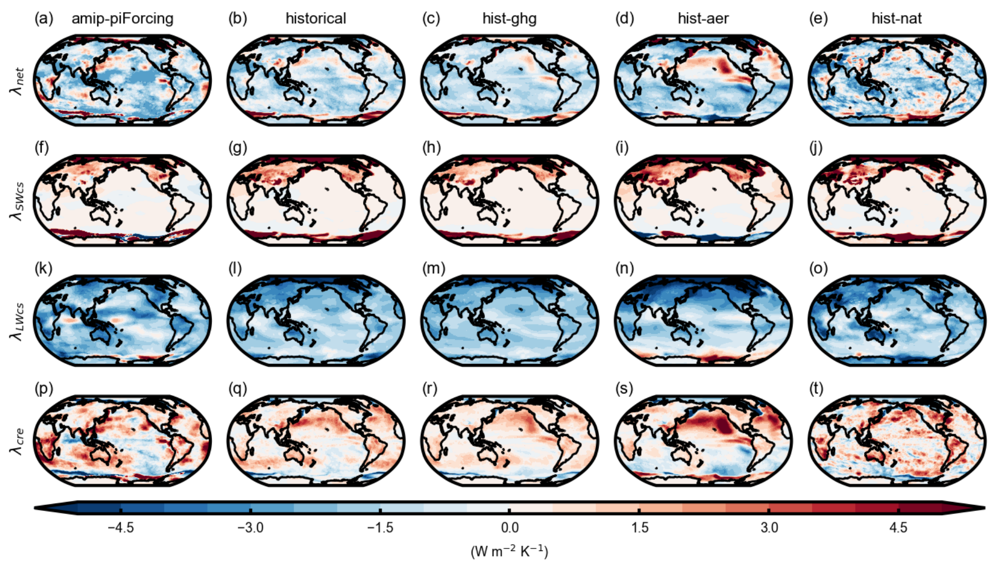

Figure 4. Maps of ensemble mean λ_{net}, λ_{SWcs}, λ_{LWcs}, and λ_{cre} in amip-piForcing, historical, hist-ghg, hist-aer, and hist-nat experiments. Here, λ has been calculated by regressing the ensemble mean local annual mean $(N - F)$ against the ensemble mean global annual mean T_s between 1850 – 2014 for historical and single forcing experiments, and 1870 – 2014 for amip-piForcing.

that for λ_{SWcs} most of the spread comes from the higher latitudes. In contrast, for λ_{cre}, variations in cloud feedbacks across the tropics and subtropics contribute to the spread. λ_{LWcs} exhibits the smallest standard deviations suggesting that this component contributes less to the ensemble spread in feedbacks. This is likely due to the fact that the Planck, lapse rate and water vapour feedbacks are highly constrained by model physics.

The three main scientific aims of this paper were to a) understand how natural variability causes different feedbacks in response to the same imposed forcing (for example, what is it that causes one historical ensemble member to have an net feedback of -1.3 Wm^{-2}K^{-1} whilst another has a feedback of -0.7 Wm^{-2}K^{-1}?), b) understand what causes different efficacies across different forcing agents, and c) investigate whether the AOGCM historical simulations - where the model simulates its own SSTs - can capture the radiative feedback and EffCS estimated from AGCM experiments prescribed with observed SSTs (i.e. are the feedbacks seen in the historical experiment consistent with those seen in amip-piForcing?). To address these questions, the different components of λ_{net} are investigated in isolation, with Section 3.2 investigating λ_{SWcs}, Section 3.3 investigating λ_{LWcs}, and Section 3.4 investigating λ_{cre}.

3.2 Processes Affecting Shortwave Clear-sky Feedbacks (λ_{SWcs})

This section aims to understand λ_{SWcs} in the historical and single forcing experiments, addressing the cause of the ensemble spread, the disparity between historical and amip-piForcing, and the cause of different efficacies across the different forcing agents. Figure 3 shows that λ_{SWcs} is a key contributor to the ensemble spread in λ_{net}, and the correlation between the two feedbacks is 0.82 across the historical experiment ensemble. Both the maps of λ_{SWcs} and standard deviation in λ_{SWcs} (Figure 4 and Figure 6b) indicate that most of the signal and spread in λ_{SWcs} comes from the higher latitudes, a region where the sea ice albedo feedback is a key process. We suggest that this feedback

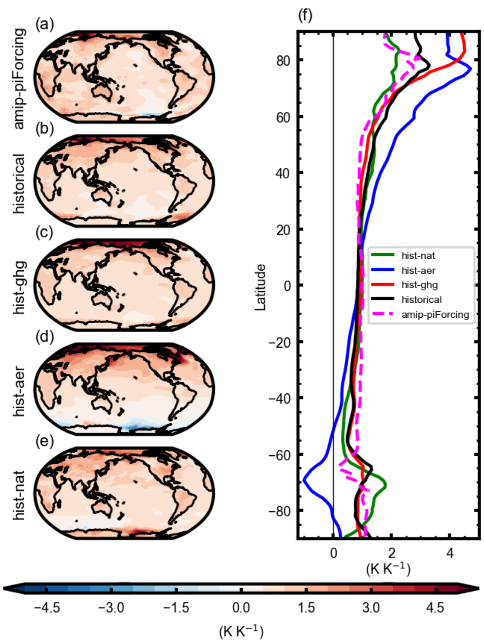

Figure 5. (left) maps of $dT_s/d\bar{T}_s$ in KK^{-1} in each experiment; amip-piForcing, historical, hist-ghg, hist-aer, and hist-nat. Here, $dT_s/d\bar{T}_s$ has been calculated by regressing the ensemble mean local annual mean T_s against the ensemble mean global annual mean T_s between 1850 – 2014 for historical and single forcing experiments, and 1870 – 2014 for amip-piForcing. (right) Zonal mean of maps to the left.

is a key contributor to the spread in λ_{SWcs} and a scatter plot of λ_{SWcs} against global sea ice fraction change per degree of warming ($d(Sea\,Ice)/d\bar{T}_s$) shown in Figure 7a confirms this. There, a correlation of -0.84 is seen between the two variables in the historical experiment over the full time period from 1850 – 2014. As previously mentioned, ultimately, the cause of differing feedbacks can be explained through variations in SST patterns. To understand the varying $d(Sea\,Ice)/d\bar{T}_s$ and λ_{SWcs} across the ensemble, scatter plots of polar $dT_s/d\bar{T}_s$ against global $d(Sea\,Ice)/d\bar{T}_s$ and λ_{SWcs} are shown in Figure 7b and c respectively. Here polar $dT_s/d\bar{T}_s$ is characterised by averaging over latitudes greater than 60°N and lower than 60°S. From Figure 7b and c, a strong relationship between polar $dT_s/d\bar{T}_s$ and both $d(Sea\,Ice)/d\bar{T}_s$ and λ_{SWcs} can be seen. This suggests that the spread in λ_{SWcs} can be understood by the degree of polar amplification across the ensemble.

Figure 7a also indicates that the sea ice albedo feedback is a key reason for the differences in λ_{SWcs} between the historical and amip-piForcing experiments. Here, the amip-piForcing experiment has been analysed only between 1980 and 2014 due to the unrealistic evolution of sea ice in the amip-piForcing experiment prior to 1980 when sea ice observations were sparse (Titchner & Rayner, 2014; T. Andrews et al., 2018). It is therefore important to note that much of the absolute difference in λ_{SWcs} and $d(Sea\,Ice)/d\bar{T}_s$ between the amip-piForcing and historical experiments in Figure 7 may be due to the different time frames analysed. The historical experiment has also been analysed between 1980 and 2014 (Figure 7 non-filled circles) and no substantial change in the relationship between each variable is seen. This does not rule out the possibility that the amip-piForcing evolution of sea ice, polar temperatures, and λ_{SWcs} may have been different for the longer

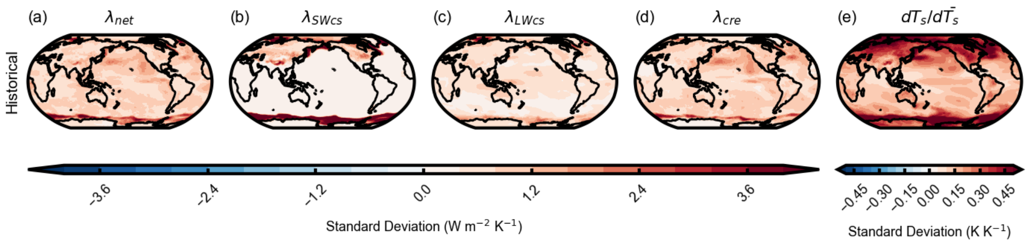

Figure 6. Maps of standard deviation in λ_{net}, λ_{SWcs}, λ_{LWcs}, λ_{cre}, and $dT_s/d\bar{T}_s$ in the historical experiment. Here, λ has been calculated by regressing the local changes in $(N - F)$ against the global mean T_s change, and $dT_s/d\bar{T}_s$ is the local T_s regressed against global mean T_s.

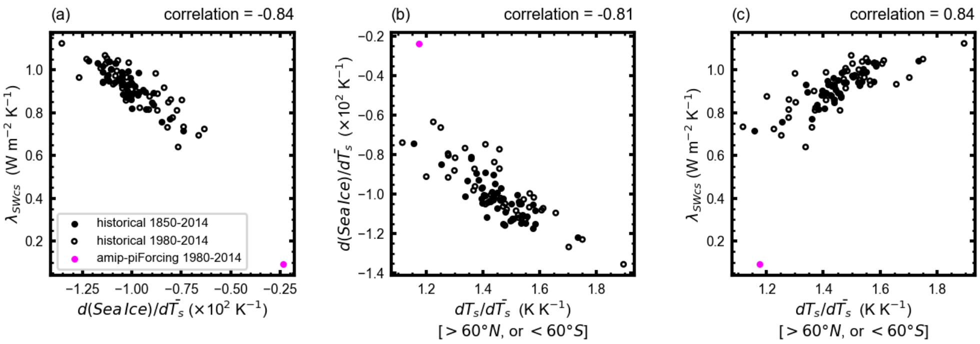

Figure 7. Scatter plots of (a) change in global sea ice per degree of warming against λ_{SWcs}, (b) change in T_s at latitudes greater than 60°N or lower than -60°S per degree of global warming against change in global sea ice per degree of global warming, and (c) change in T_s at latitudes greater than 60°N or lower than 60°S per degree of global warming against λ_{SWcs}. Here, each black dot represents a historical ensemble member where values are calculated between 1850–2014 for the filled black dots, and 1980–2014 for the unfilled black dots. The magenta dots represent the amip-piForcing experiment calculated between 1980–2014 (due to sparse sea ice observations prior to 1980).

period, however, the fact that the amip-piForcing experiment is consistent with the relationship seen in the historical experiment (demonstrated in Figure 7a) would suggest that differences in λ_{SWcs} between historical and amip-piForcing experiments can be explained through this mechanism, and the smaller λ_{SWcs} in amip-piForcing is related to the smaller $d(Sea\,Ice)/d\bar{T}_s$. The fact that in 7b the amip-piForcing experiment does not fit the historical ensemble relationship between polar $dT_s/d\bar{T}_s$ and $d(Sea\,Ice)/d\bar{T}_s$ suggests that that the AOGCMs simulation of the relationship between SSTs and sea ice melt is not the same as the observed relationship in the real world (assuming the relationship seen in amip-piForcing is a good analogue for the real world).

Thus far the ensemble spread and the disparity between historical and amip-piForcing estimates of λ_{SWcs} has been investigated. It is shown that the sea ice albedo feedback is a key process responsible for both, with the level of arctic amplification providing the link between ensemble spread in λ_{SWcs} and T_s patterns. Previously, Dessler (2020) also investigated changes in sea ice and its impact on feedbacks. Consistent with the results shown in Figure 7, Dessler (2020) also found sea ice variability to cause a large spread in λ_{SWcs} in their historical ensemble with a different model, where these feedback variations were linked to changes in different modes of ocean variability. Since Figure 7 highlights a strong relationship between polar SSTs and sea ice, understanding causes of polar SST change and how they are predicted to evolve in a future climate is important.

Other processes could also contribute to the spread in λ_{SWcs}, such as snow melt. This could be responsible for the strong λ_{SWcs} seen over the Northern Hemisphere land masses in Figure 4 f, g, h, i, and j, and the spread in λ_{SWcs} seen in Figure 6b. However, this process is not investigated further here since the strongest spread in λ_{SWcs} is seen over the Arctic and Southern Oceans.

With the understanding gained from Figure 7, the different efficacies of each forcing agent are investigated. Maps of ensemble mean λ_{SWcs} and $dT_s/d\bar{T}_s$ are shown in Figure 8. Here, the hist-ghg experiment is shown and each of the other experiments are shown relative to the hist-ghg values. This enables clearer identification of the differences between each forcing agent.

From Figure 8 the spatial pattern of $dT_s/d\bar{T}_s$ and λ_{SWcs} are shown to be similar, suggesting that the regional change in $dT_s/d\bar{T}_s$ leads to regional changes in λ_{SWcs} due to the close relationship between T_s and sea ice. This is true for both the northern and southern hemisphere and also across each of the experiments. The spatial correlations between $dT_s/d\bar{T}_s$ and λ_{SWcs} across all experiments and each hemisphere are between 0.64 – 0.88, further highlighting the strong coupling between local T_s patterns and local feedbacks. For the historical experiment, in the southern hemisphere, a stronger λ_{SWcs} is associated with a larger Southern Ocean $dT_s/d\bar{T}_s$ relative to hist-ghg. The northern hemisphere maps in 8b show contrasting feedbacks between the Arctic Ocean regions and the slightly lower latitude regions around the Labrador Sea. Over the Arctic Ocean hist-ghg has a stronger λ_{SWcs} compared to the historical simulations, whereas around the Labrador Sea, the historical experiment has the stronger λ_{SWcs}. This is reflected in the $dT_s/d\bar{T}_s$ patterns, where the historical experiment has a weaker $dT_s/d\bar{T}_s$ over the Arctic Ocean, but a stronger $dT_s/d\bar{T}_s$ over the Labrador Sea. This northern hemisphere pattern in λ_{SWcs} and $dT_s/d\bar{T}_s$ relative to hist-ghg is similar to that seen in the hist-aer and hist-nat experiment, where the hist-aer experiment demonstrates the largest positive λ_{SWcs} values and also extends these positive values furthest south.

In the southern hemisphere, unlike the historical experiment, the hist-aer experiment shows strongly negative λ_{SWcs} and $dT_s/d\bar{T}_s$ relative to the hist-ghg experiment. As previously mentioned, this may be due to ocean convection in the Southern Ocean triggered by the ocean becoming unstable (Ridley et al., 2022). This convection could bring warmer water up from below, warming the surface, melting sea ice, and resulting in a negative λ_{SWcs}.

Here, it has been shown that the sea ice albedo feedback and the level of arctic amplification is a key process in producing the large spread in λ_{SWcs} across the ensemble and is also a key reason for the different feedback seen in the historical and amip-piForcing experiments. It has also been shown that the different efficacies seen across the different historical and single forcing experiments can be explained through differing SST patterns (in agreement with Haugstad et al. (2017)), with the experiments with a stronger λ_{SWcs} locally, also exhibiting a larger $dT_s/d\bar{T}_s$.

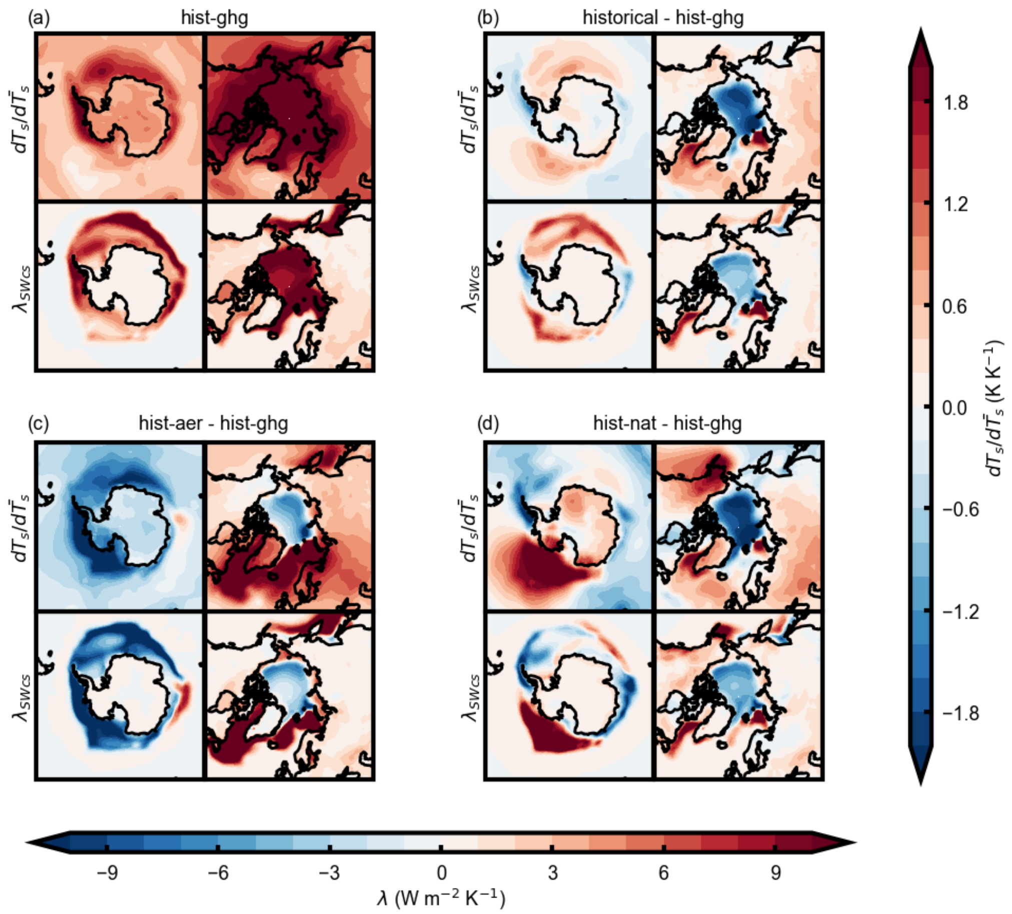

Figure 8. Maps of (top rows) surface warming pattern (KK^{-1}) and (bottom rows) λ_{SWcs} over the (right columns) northern and (left columns) southern hemisphere poles in the (a) hist-ghg experiment and (b) historical, (c) hist-aer and (d) hist-nat experiments relative to hist-ghg.

3.3 Processes Affecting Longwave Clear-sky Feedbacks (λ_{LWcs})

From Figure 3 it can be seen that whilst the λ_{LWcs} does not contribute much to the different efficacies seen in each of the historical and single forcing experiments, it does contribute to the spread in λ_{net} and is also a large source of disparity between the historical and amip-piForcing experiments. Understanding the spread in λ_{LWcs} and the disparity between the historical and amip-piForcing experiments is the aim of this section.

λ_{LWcs} is determined by a combination of the Planck feedback, the water vapour feedback and the lapse rate feedback (T. Andrews & Webb, 2018). The water vapour and lapse rate feedbacks have been shown to be strongest in the tropical troposphere (Soden et al., 2008; T. Andrews & Webb, 2018), since the tropical atmosphere closely follows a moist adiabatic lapse rate and therefore any warming at the surface is amplified vertically in the atmosphere (Po-Chedley et al., 2018). To investigate the λ_{LWcs} in the historical ensemble, first, plots of zonal mean atmospheric temperature regressed against global mean T_s ($dT_a/d\bar{T}_s$) are analysed (Figure 9). Note that as previously discussed, here, the atmospheric temperature (T_a) has had any changes associated with the forcing subtracted from it (see discussion following Equation 3). This means that the CO_2

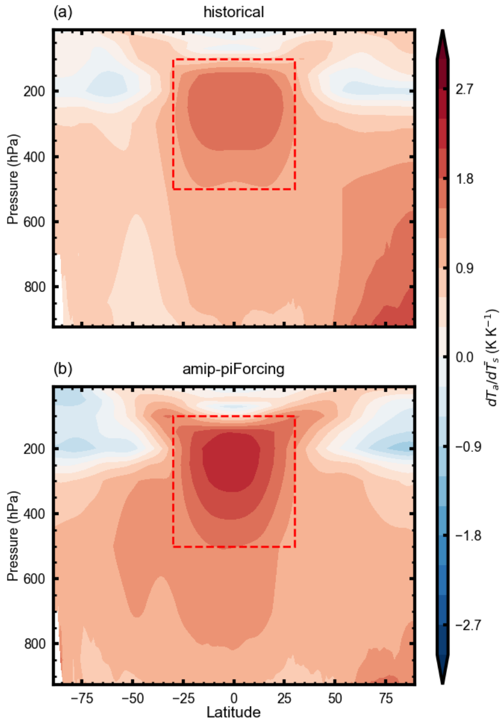

Figure 9. Zonal mean changes in temperature per degree of global warming in the (a) historical and (b) amip-piForcing experiments.

driven stratospheric cooling in the historical experiment is removed, and a more accurate comparison between historical and amip-piForcing experiments can be made.

From Figure 9 the pattern of $dT_a/d\bar{T}_s$ seen in both the historical and amip-piForcing experiments demonstrates a marked warming over the tropical troposphere. Comparing Figure 9b and c it can be seen that this tropospheric $dT_a/d\bar{T}_s$ is stronger in amip-piForcing compared to the historical experiment. The amip-piForcing experiment also exhibits a stronger $dT_a/d\bar{T}_s$ over the southern hemisphere troposphere, whilst the historical experiment has a larger $dT_a/d\bar{T}_s$ signal over the northern hemisphere high latitudes. This is potentially due to the different T_s patterns seen in the historical and amip-piForcing experiments, with the subtropical $dT_s/d\bar{T}_s$ being slightly greater in the Northern Hemisphere in the historical ensemble and in the Southern Hemisphere in amip-piForcing (Figure 5f).

Since the tropical troposphere is a key region in causing variations in λ_{LWcs}, a region between 30°S – 30°N and between 100 – 500 hPa has been analysed further. A scatter plot of tropical tropospheric $dT_a/d\bar{T}_s$ against λ_{LWcs} is shown in Figure 10a. There it can be seen that a strong correlation between the two variables exists with a correlation coefficient of -0.8, consistent with physical expectations that a larger upper tropical tropospheric temperature results in a larger lapse rate feedback and a more negative λ_{LWcs} (T. Andrews & Webb, 2018). The amip-piForcing tropical tropospheric $dT_a/d\bar{T}_s$ and λ_{LWcs} has also been indicated in Figure 10a, where it can be seen that the tropical tropospheric $dT_a/d\bar{T}_s$ does well to capture why the feedbacks in historical and amip-piForcing experiments differ.

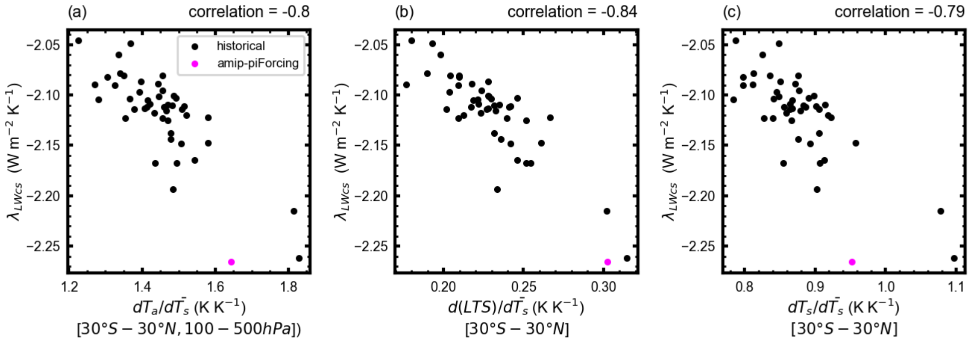

Figure 10. Scatter plots of (a) tropical tropospheric $dT_a/d\bar{T}_s$ against λ_{LWcs}, (b) tropical Lower Tropospheric Stability (LTS) change per degree of global warming ($d(LTS)/d\bar{T}_s$) against λ_{LWcs}, and (c) tropical $dT_s/d\bar{T}_s$ against λ_{LWcs}. Here the tropics have been characterised by averaging between 30°S and 30°N, and the tropical troposphere has used the same latitudinal bounds and averaged between 100–500 hPa (see red boxes in Figure 9). In each plot, black dots represent the historical ensemble and amip-piForcing values are represented by a magenta dot.

Since the spread in feedbacks can ultimately be derived from differing SST patterns, and given the strong relationship between tropical tropospheric temperature and λ_{LWcs}, the relationship between tropical mean $dT_s/d\bar{T}_s$ and λ_{LWcs} has been investigated (Figure 10c). Figure 10c follows a similar analysis to that performed by Soden and Held (2006). There, they demonstrated that across a range of models, due to the approximately adiabatic lapse rate of the tropical atmosphere, the strong coupling between the surface and free troposphere in the tropics, and the relatively weak coupling present over higher latitudes, the ratio between tropical and global warming was a good metric for determining the inter-model spread in lapse rate feedback. In Figure 10c it is shown that across the historical ensemble, the tropical $dT_s/d\bar{T}_s$ is well correlated with λ_{LWcs} with a correlation coefficient of -0.79. It is clear that ensemble members with a stronger warming over the tropics relative to the global mean also have a more strongly negative λ_{LWcs}.

As well as explaining the ensemble spread in λ_{LWcs}, tropical $dT_s/d\bar{T}_s$ changes can also be used to explain the disparity between amip-piForcing and historical experiments. Figure 10c shows that the amip-piForcing experiment has a strong $dT_s/d\bar{T}_s$ in the tropics and also has a strong negative λ_{LWcs}.

3.4 Processes Affecting Cloud Feedbacks (λ_{cre})

Although the historical ensemble used in this paper indicates that λ_{cre} is not the feedback with the largest spread (λ_{SWcs} has a standard deviation of 0.073 Wm^{-2}K^{-1} whilst λ_{cre} has a standard deviation of 0.06 Wm^{-2}K^{-1}), for long term estimates of EffCS across different models, cloud feedbacks are the largest source of uncertainty and are the least understood (Forster et al., 2021; Ceppi & Nowack, 2021; Zelinka et al., 2016; Ceppi et al., 2017). Because of this, over recent years, cloud feedbacks have been the focus of many studies. Cloud controlling factor analyses such as Ceppi and Nowack (2021) and Blanco et al. (2023) aim to relate changes in clouds to other meteorological factors, such as free tropospheric humidity (van der Dussen et al., 2015), SSTs (Bretherton & Blossey, 2014), surface wind speed (Brueck et al., 2015) and inversion strength (Qu et al., 2015; Klein et al., 2017; Kawai et al., 2017). By better understanding what factors cause clouds to change, it is possible to understand differences in cloud feedbacks across models/ensembles.

In this section, λ_{cre} is investigated, primarily focusing on the spread across the historical experiment ensemble. Previously, Salvi et al. (2022) suggested that the different efficacies of well mixed greenhouse gases and aerosols were linked to changes in clouds due to differing changes in stability (although a large variability is seen across different models and a relatively small ensemble of 7 models was used). However here, the results shown in Figure 3 would suggest that for HadGEM3-GC3.1-LL, λ_{cre} does not contribute substantially to different forcing efficacies in the global mean. To understand the spatial distribution of λ_{cre}, Figure 4q is analysed. Here, strong positive cloud feedbacks are seen over the North Pacific and North Atlantic, and slightly weaker cloud feedbacks are seen over the Southern Indian Ocean and South Atlantic (each caused by positive shortwave cloud feedbacks - Figure S2). To understand the spread in λ_{cre}, maps of standard deviation in λ_{cre}, λ_{SWcre}, and λ_{LWcre} and standard deviation in $dT_s/d\bar{T_s}$ are shown in Figure 11. From Figure 11a it is possible to identify regions where the spread in λ_{cre} is largest and therefore which regions contribute most to the spread seen in Figure 3. The regions with the largest spread in λ_{cre} are the North Pacific and North Atlantic, due to a large spread in λ_{SWcre}. The Southern Ocean and low cloud deck regions off the east coast of South America, Australia and Southern Africa, also exhibit a moderately large standard deviation in λ_{cre}, again due to shortwave cloud feedbacks. The map of standard deviation of λ_{LWcre} shows a large spread in feedbacks over the tropical ascent regions, however as previously discussed, in these regions, longwave and shortwave responses to changes in cloud cancel, and therefore the standard deviation in net cloud feedbacks in these regions is generally small.

The spatial distribution of the standard deviation in $dT_s/d\bar{T_s}$ shown in Figure 11f is relatively similar to the pattern of standard deviation in λ_{cre}. Calculating the spatial correlation between Figures 11a and f, a correlation coefficient of 0.47 is found. Given surface temperatures are a key cloud controlling factor, as shown by Ceppi and Nowack (2021), we would expect the spread in λ_{cre} to be partly controlled by the spread in $dT_s/d\bar{T_s}$.

To better understand the cause of the spread in λ_{cre} shown in Figure 3b and 11a, two key cloud controlling factors are investigated; changes in T_s and changes in Lower Tropospheric Stability (LTS), both of which have strong statistical relationships with changes in clouds (Cutler et al., 2022; Klein & Hartmann, 1993; Ceppi & Nowack, 2021). Here LTS is defined as the 700hPa potential temperature minus the surface potential temperature (Cutler et al., 2022). Regarding the physical mechanisms of these relationships, LTS has been shown to influence cloud changes by controlling the amount of entrainment between the moist boundary layer and the drier free troposphere. The physical mechanism whereby surface temperatures effect cloud changes is less well established. Webb et al. (2024) investigate a range of possible mechanism relating surface temperatures to changes in cloud, such as the impact of surface latent heat flux changes, vertical gradients in humidity or moist static energy, or changes in downwelling longwave radiation caused by changing free tropospheric humidity. It was found that different mechanisms were plausible in some models and not in others. For HadGEM3-GC3.1-LL, only one suggested mechanism was not ruled out based on the models behaviour. This mechanism involved a reduction in low cloud due to a warming and a decrease in specific humidity due to an increase in upward longwave radiation from the surface (Ogura et al., 2023).

To relate changes in LTS and surface temperatures to changes in λ_{cre}, first two regions are investigated, the North West (NW) Pacific and North East (NE) Pacific (see Figure 11 boxes). These two regions were selected as being regions with a strong λ_{cre}

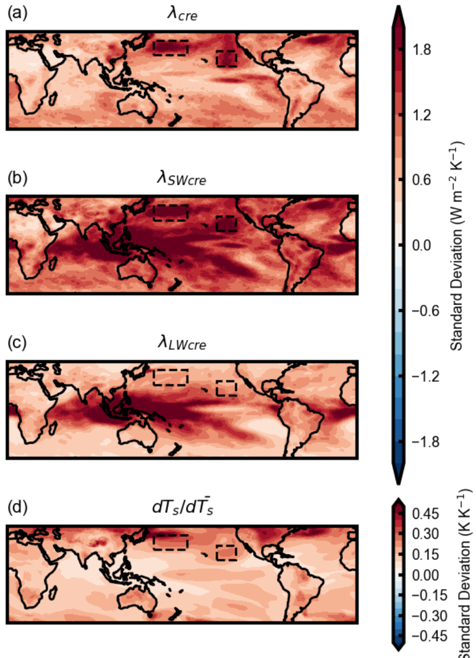

Figure 11. Maps of standard deviation in (a) λ_{cre}, (b) λ_{SWcre}, (c) λ_{LWcre}, and (d) $dT_s/d\bar{T}_s$ across the historical ensemble. Dashed black boxes indicate regions analysed in Figure 12 with the NW Pacific region extending from 150–185°E and 26–41°N, and the NE Pacific region extending from 215–235°E and 15–30°N.

signal (Figure 4q) and spread (Figure 11a). The two regions also capture different climatological regimes, with the NE Pacific a region of climatological subsidence where the surface is decoupled from the free troposphere due to a strong inversion, whereas the NW Pacific region is a region of climatological ascent where the surface is not decoupled from the free troposphere. Scatter plots of $d(LTS)/d\bar{T}_s$ and $dT_s/d\bar{T}_s$ against λ_{cre} over the NW Pacific and NE Pacific regions are shown in Figure 12a, b, c, and d. Here, it can be seen that in both the NE and NW Pacific there is a strong correlation between $dT_s/d\bar{T}_s$ and λ_{cre}, and $d(LTS)/d\bar{T}_s$ and λ_{cre}. This is consistent with Ceppi and Nowack (2021). Although the amip-piForcing and historical estimates of λ_{cre} were not particularly different, for completeness, amip-piForcing values have also been indicated in Figure 12. Here it can be seen that the amip-piForcing values fit the historical relationship between λ_{cre} and both $dT_s/d\bar{T}_s$ and $d(LTS)/d\bar{T}_s$ suggesting that any differences in λ_{cre} between historical and amip-piForcing experiments in these regions can be explained through these cloud controlling factors.

Since the LTS is defined as the 700hPa potential temperature minus the surface potential temperature, it is possible that the strong correlations between $d(LTS)/d\bar{T}_s$ and λ_{cre} exist primarily because of the strong relationship between λ_{cre} and $dT_s/d\bar{T}_s$. To investigate this, scatter plots of 700hPa $dT_a/d\bar{T}_s$ against λ_{cre} are shown in Figure 12e and f. Here, differing relationships between the two variables exist over the two regions analysed. Over the NW Pacific, a strong correlation remains with a correlation coefficient of 0.84. Over the NE Pacific however, this is not the case and a weak correlation

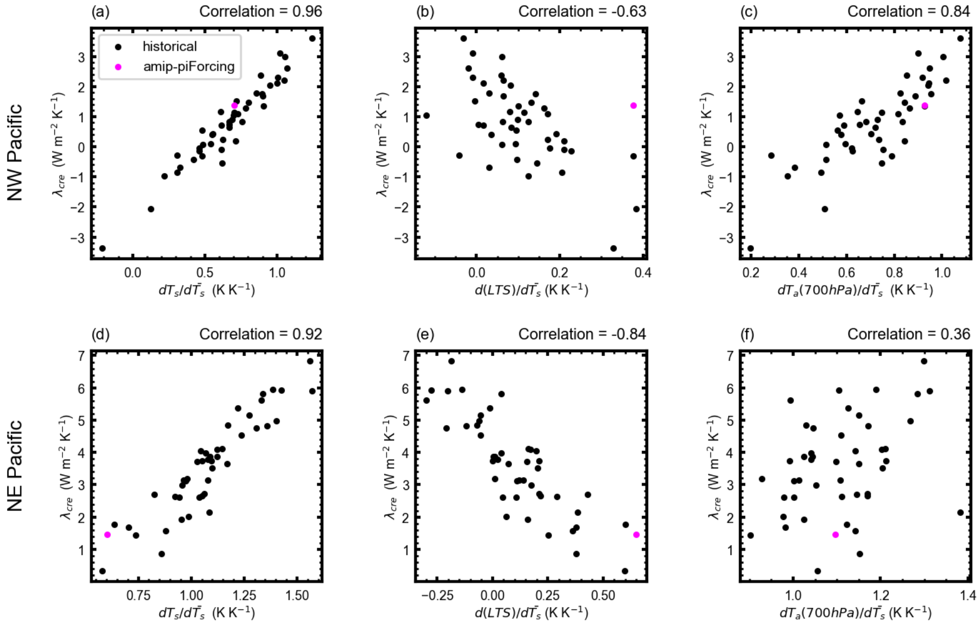

Figure 12. Scatter plots of (a and b) $dT_s/d\bar{T}_s$, (c and d) $d(LTS)/d\bar{T}_s$, and (e and f) 700hPa $dT_a/d\bar{T}_s$ against λ_{cre} over the (a, c, and e) NW Pacific region, and (b, d, and f) NE Pacific region. Black dots represent the historical ensemble and magenta markers indicate amip-piForcing values.

of 0.36 is seen. This differing relationship may be due to the different convective regimes that exist over the two regions. Over the NE Pacific, the strong inversion and the decoupling between the boundary layer and the free troposphere means that any surface warming in this region will be trapped under the strong inversion. Over the NW Pacific, this is not the case and surface warming can be transported efficiently into the free troposphere. Therefore, to some degree, over the NW Pacific the 700hPa temperature is still controlled by the temperatures at the surface.

An alternative approach is taken in Figure 13. Here, the local effect of surface warming and the remote effect of large scale stability changes on λ_{cre} is investigated using maps of the correlation across the historical ensemble between local λ_{cre} and either the local $dT_s/d\bar{T}_s$ or the 50°S – 50°N mean 700hPa $dT_a/d\bar{T}_s$. These latitudinal bounds were previously used by Ceppi and Gregory (2019) and Salvi et al. (2023) to capture large scale tropospheric stability.

From Figure 13 it can be seen that generally, the local $dT_s/d\bar{T}_s$ is the most strongly correlated, with many regions exhibiting correlations greater than 0.7. The correlations between λ_{cre} and the 50°S – 50°N mean 700hPa $dT_a/d\bar{T}_s$ tend to be weaker, although the subtropical cloud deck regions over the East Pacific and the Indian Ocean do exhibit positive correlations (note these are not statistically significant at the 95% confidence range). A decomposition of Figure 13 into shortwave and longwave components is shown in Figure S3. Here the strong correlations seen in the low cloud deck regions in Figure 13 are associated with the shortwave cloud feedbacks, and similar to Figure 11 and S2, the tropical ascent regions exhibit relatively strong correlations with both local $dT_s/d\bar{T}_s$ and 50°S – 50°N mean 700hPa $dT_a/d\bar{T}_s$ in the shortwave and longwave, however these

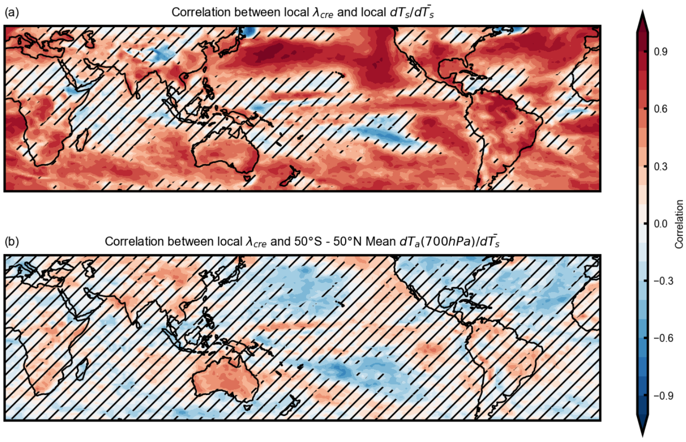

Figure 13. Maps of correlation between local λ_{cre} against (a) local $dT_s/d\bar{T}_s$, and (b) 50°S – 50°N mean 700hPa $dT_a/d\bar{T}_s$ across the historical ensemble. Hatching indicates where correlations are not significant at the 95% confidence interval (i.e. p values are greater than 0.05). Here the p value approximately indicates the probability of two random distributions producing a correlation coefficient at least as great as those indicated in the colored contours.

two components cancel, resulting in the net cloud feedback correlation being relatively weak in those regions in Figure 13.

To summarise, cloud feedbacks are the largest source of uncertainty in EffCS across models, however within the HadGEM3-GC3.1-LL historical ensemble, λ_{SWcs} contributes more to the spread in λ_{net}. Spread in λ_{cre} can be explained through the cloud controlling factors of T_s and LTS where $dT_s/d\bar{T}_s$ is positively correlated with λ_{cre} and $d(LTS)/d\bar{T}_s$ is negatively correlated with λ_{cre}. Finally, it is shown that the local influence of $dT_s/d\bar{T}_s$ on λ_{cre} is much stronger than the remote effect of changes in large scale atmospheric stability.

4 Conclusion

In this paper the feedbacks across a 47 member ensemble of historical and single forcing simulations have been analysed. Across the historical ensemble, EffCS ranges between 3–6K, highlighting the large spread in estimated feedbacks caused by internal variability. The aims of this work have been to understand the main causes of this spread in feedbacks across the ensemble, to understand if and why different forcing agents have different forcing efficacies, and finally to understand why the coupled historical simulations struggle to capture the feedbacks seen in AGCM simulations forced by observed SSTs. To address these aims, three components of λ_{net} were investigated (λ_{SWcs}, λ_{LWcs}, and λ_{cre}).

The analysis found that the ensemble spread in λ_{SWcs} is largely caused by varying degrees of sea ice melt per degree of global warming. Ensemble members that showed a large reduction in sea ice per degree of global warming also exhibited a strong λ_{SWcs}, with a correlation of -0.84 (consistent with Dessler (2020)). It was shown that this relationship was due to varying SST patterns, with ensemble members simulating stronger polar amplification also exhibiting more sea ice melt and a stronger λ_{SWcs} (with a correlation of 0.84 between polar SSTs and λ_{SWcs}). This relationship between λ_{SWcs}, sea ice melt, and polar amplification is also shown to be the reason for a much weaker λ_{SWcs} in the amip-piForcing experiment. Here, weaker polar amplification resulted in less sea ice melt per degree of global warming and a smaller λ_{SWcs}. Finally, the different λ_{SWcs} between the different single forcing experiments was investigated, since λ_{SWcs} was found to be the biggest source of differing forcing efficacies across the different forcing agents. It was shown that different patterns of surface warming were the main cause of different feedbacks across each experiment, with spatial correlations of 0.64 – 0.88 between patterns of T_s change per degree of global warming and λ_{SWcs} across all experiments and each hemisphere.

Previously, Salvi et al. (2022) also investigated different forcing efficacies between different forcing agents, also finding the hist-aer experiment to exhibit more strongly amplifying feedbacks compared to hist-ghg. There they focused on influence of stability changes on changes in cloud feedbacks, however here, we find that for HadGEM3-GC3.1-LL, changes in sea ice and polar T_s play a larger role in causing different forcing efficacies.

The ensemble spread in λ_{LWcs} was also investigated. Here it was shown that both tropical tropospheric temperature changes per degree of global warming and tropical T_s changes per degree of global warming were a key factor in causing the spread in λ_{LWcs}. Here, increased tropical surface warming caused warming in the tropical troposphere which has previously been shown to cause a stronger lapse rate feedback (T. Andrews & Webb, 2018). This relationship between λ_{LWcs} and tropical T_s also captures why the λ_{LWcs} is much stronger in the amip-piForcing experiment compared to the historical simulations, with the amip-piForcing experiment exhibiting a stronger tropical surface warming per degree of global warming compared to most historical ensemble members. Given that the amip-piForcing experiment is prescribed with observed SSTs, this shows how AOGCM biases in tropical SST patterns can impact on the estimated λ_{LWcs}.

The final feedback to be investigated was λ_{cre}. In contrast to the primary role of λ_{cre} in causing uncertainty in long term estimates of climate sensitivity, in the HadGEM3-GC3.1-LL historical ensemble, other feedbacks have a larger spread. Investigating λ_{cre}, it was shown that both T_s and LTS are key factors affecting changes in cloud feedbacks. It is also shown that although amip-piForcing and historical cloud feedbacks are not too dissimilar, both the LTS and T_s are useful metrics for understanding how amip-piForcing cloud feedbacks relate to those seen in the historical simulations. The analysis concludes by investigating the relative importance of local effect of varying T_s or the remote effect of large scale changes in atmospheric stability. Here it is shown that the local T_s is the most important, whilst the large scale stability plays a non-negligible role over the subtropical cloud deck regions.

This work provides useful insight into the different feedbacks seen across different forcing experiments and also provides information as to why coupled historical simulations struggle to capture the feedbacks seen in the amip-piForcing experiment. To take this work further, this large ensemble could be used to better understand the temporal evolution of feedbacks. In recent years, the amip-piForcing experiment demonstrates a marked decrease in λ_{net} (T. Andrews et al., 2022), and this ensemble could be used to investigate whether a similar behaviour is captured in any of the ensemble members. This work could then be used shed light on the causes and mechanisms involved in transient feedbacks.

5 Open Research

Data used in this analysis consists of HadGEM3-GC3.1-LL model simulations performed as part of the Met Office's contribution to CMIP6 (Eyring et al., 2016) and LESFMIP (D. Smith et al., 2022) and can be accessed from the ESGF CEDA data node https://esgf-index1.ceda.ac.uk/search/cmip6-ceda/.

Acknowledgments

All authors were supported by the Met Office Hadley Centre Climate Programme funded by DSIT. Doug Smith was also supported by the European Commission Horizon 2020 CONSTRAIN project.

References

Adams, B. K., & Dessler, A. E. (2019). Estimating transient climate response in a large-ensemble global climate model simulation. *Geophysical Research Letters*, *46*, 311-317. doi: 10.1029/2018GL080714

Andrews, M. B., Ridley, J. K., Wood, R. A., Andrews, T., Blockley, E. W., Booth, B., ... Sutton, R. T. (2020). Historical simulations with hadgem3-gc3.1 for cmip6. *Journal of Advances in Modeling Earth Systems*, *12*. doi: 10.1029/2019MS001995

Andrews, T., Andrews, M. B., Bodas-Salcedo, A., Jones, G. S., Kuhlbrodt, T., Manners, J., ... Tang, Y. (2019). Forcings, feedbacks and climate sensitivity in hadgem3-gc3.1 and ukesm1. *Journal of Advances in Modeling Earth Systems*, *11*, 4377–4394. doi: 10.1029/2019MS001866

Andrews, T., Bodas-Salcedo, A., Gregory, J. M., Dong, Y., adn D. Paynter, K. C. A., Lin, P., ... Liu, C. (2022). On the effect of historical sst patterns on radiative feedback. *Journal of Geophysical Research: Atmospheres*, *127*. doi: 10.1029/2022JD036675

Andrews, T., Gregory, J. M., Paynter, D., Silvers, L. G., Zhou, C., Mauritsen, T., ... Titchner, H. (2018). Accounting for changing temperature patterns increases historical estimates of climate sensitivity. *Geophysical Research Letters*, *45*, 8490-8499. doi: 10.1029/2018GL078887

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

Andrews, T., Gregory, J. M., Webb, M. J., & Taylor, K. E. (2012). Forcings, feedbacks and climate sensitivity in cmip5 coupled atmosphere-ocean climate models. *Geophysical Research Letters*, *39*. doi: 10.1029/2012GL051607

Andrews, T., & Webb, M. J. (2018). The dependence of global cloud and lapse rate feedbacks on the spatial structure of tropical pacific warming. *Journal of Climate*, *31*, 641–654. doi: 10.1175/JCLI-D-17-0087.1

Blanco, J. E., Caballero, R., Datseris, G., Stevens, B., Bony, S., Hadas, O., & Kaspi, Y. (2023). A cloud-controlling factor perspective on the hemispheric asymmetry of extratropical cloud albedo. *Journal of Climate*, *36*, 1793–1804. doi: 10.1175/JCLI-D-22-0410.1

Bretherton, C. S., & Blossey, P. N. (2014). Low cloud reduction in a greenhouse-warmed climate: Results from lagrangian les of a subtropical marine cloudiness transition. *Journal of Advances in Modeling Earth Systems*, *6*, 91–114. doi: 10.1002/2013MS000250

Brueck, M., Nuijens, L., & Stevens, B. (2015). On the seasonal and synoptic time-scale variability of the north atlantic trade wind region and its low-level clouds. *Journal of the Atmospheric Sciences*, *72*, 1428–1446. doi: 10.1175/JAS-D-14-0054.1

Ceppi, P., Brient, F., Zelinka, M. D., & Hartmann, D. L. (2017). Cloud feedback mechanisms and their representation in global climate models. *WIREs Climate Change*, *8*. doi: 10.1002/wcc.465

Ceppi, P., & Gregory, J. M. (2019). A refined model for the earth's global energy balance. *Climate Dynamics*, *53*, 4781–4797. doi: 10.1007/s00382-019-04825-x

Ceppi, P., & Nowack, P. (2021). Observational evidence that cloud feedback amplifies global warming. *PNAS*, *118*. doi: 10.1073/pnas.2026290118

Cutler, L., Brunke, M. A., & Zeng, X. (2022). Re-evaluation of low cloud amount relationships with lower-tropospheric stability and estimated inversion strength. *Geophysical Research Letters*, *49*. doi: 10.1029/2022GL098137

Dessler, A. E. (2020). Potential problems measuring climate sensitivity from the historical record. *Journal of Climate*, *33*, 2237–2248. doi: 10.1175/JCLI-D-19-0476.1

Dessler, A. E., Mauritsen, T., & Stevens, B. (2018). The influence of internal variability on earth's energy balance framework and implications for estimating climate sensitivity. *Atmospheric Chemistry and Physics*, *18*, 5147–5155. doi: 10.5194/acp-18-5147-2018

Dong, Y., Armour, K. C., Proistosescu, C., Andrews, T., Battisti, D. S., Forster, P. M., ... Shiogama, H. (2021). Biased estimates of equilibrium climate sensitivity and transient climate response derived from historical cmip6 simulations. *Geophysical Research Letters*, *48*. doi: 10.1029/2021GL095778

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, *9*, 1937–1958. doi: 10.5194/gmd-9-1937-2016

Forster, P. M., Richardson, T., Maycock, A. C., Smith, C. J., Samset, B. H., Myhre, G., ... Schulz, M. (2016). Recommendations for diagnosing effective radiative forcing from climate models for cmip6. *Journal of Geophysical Research: Atmospheres*, *121*, 12,460-12,475. doi: 10.1002/2016JD025320

Forster, P. M., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., ... co authors (2021). The earth's energy budget, climate feed- backs, and climate sensitivity. In *Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change* (p. 93). Cambridge, UK and New York, USA: Cambridge University Press.

Fueglistaler, S., & Silvers, L. G. (2021). The peculiar trajectory of global warming. *Journal of Geophysical Research: Atmospheres*, *126*. doi: 10.1029/2020JD033629

Gillett, N. P., Shiogama, H., Funke, B., Hegerl, G., Knutti, R., Matthes, K., ... Tebaldi, C. (2016). The detection and attribution model intercomparison project (damip v1.0) contribution to cmip6. *Geoscientific Model Development*, *9*, 3685–3697. doi: 10.5194/gmd-9-3685-2016

Gregory, J. M., & Andrews, T. (2016). Variation in climate sensitivity and feedback parameters during the historical period. *Geophysical Research Letters*, *43*, 3911-3920. doi: 10.1002/2016GL068406

Gregory, J. M., Andrews, T., Ceppi, P., Mauritsen, T., & Webb, M. J. (2020). How accurately can the climate sensitivity to co2 be estimated from historical climate change? *Climate Dynamics*, *54*, 129–157. doi: 10.1007/s00382-019-04991-y

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., ... Williams, K. D. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, *31*. doi: 10.1029/2003GL018747

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., ... Zhang, S. (2005). Efficacy of climate forcings. *Journal of Geophysical Re-*

search: Atmospheres, *110*. doi: 10.1029/2005JD005776

Haugstad, A. D., Armour, K. C., Battisti, D. S., & Rose, B. E. J. (2017). Relative roles of surface temperature and climate forcing patterns in the inconstancy of radiative feedbacks. *Geophysical Research Letters*, *44*, 7455-7463. doi: 10.1002/2017GL074372

Hwang, Y.-T., Xie, S.-P., Chen, P.-J., Tseng, H.-Y., & Deser, C. (2024). Contribution of anthropogenic aerosols to persistent la niña-like conditions in the early 21st century. *Proceedings of the National Academy of Sciences*, *121*. doi: 10.1073/pnas.2315124121

Kawai, H., Koshiro, T., & Webb, M. J. (2017). Interpretation of factors controlling low cloud cover and low cloud feedback using a unified predictive index. *Journal of Climate*, *30*, 9119—9131. doi: 10.1175/JCLI-D-16-0825.1

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., … Vertenstein, M. (2015). The community earth system model (cesm) large ensemble project. *Bulletin of the American Meteorological Society*, 1333—1349. doi: 10.1175/BAMS-D-13-00255.1

Kay, J. E., Holland, M. M., & Jahn, A. (2011). Inter-annual to multi-decadal arctic sea ice extent trends in a warming world. *Geophysical Research Letters*, *38*. doi: 10.1029/2011GL048008

Klein, S. A., Hall, A., Norris, J. R., & Pincus, R. (2017). Low-cloud feedbacks from cloud-controlling factors: A review. *Surves in Geophysics*, *38*, 1307–1329. doi: 10.1007/s10712-017-9433-3

Klein, S. A., & Hartmann, D. L. (1993). The seasonal cycle of low stratiform clouds. *Journal of Climate*, *6*, 1587–1606. doi: 10.1175/1520-0442(1993)006⟨1587:TSCOLS⟩2.0.CO;2

Lambert, F. H., Webb, M. J., & Joshi, M. M. (2011). The relationship between land–ocean surface temperature contrast and radiative forcing. *Journal of Climate*, *24*, 3239 - 3256. doi: 10.1175/2011JCLI3893.1

Marvel, K., Schmidt, G. A., Miller, R. L., & Nazarenko, L. S. (2015). Implications for climate sensitivity from the response to individual forcings. *Nature Climate Change*, *6*, 386–389. doi: 10.1038/NCLIMATE2888

Ogura, T., Webb, M. J., & Lock, A. P. (2023). Positive low cloud feedback primarily caused by increasing longwave radiation from the sea surface in two versions of a climate model. *Geophysical Research Letters*, *50*. doi: 10.1029/2023GL104786

Pincus, R., Forster, P. M., & Stevens, B. (2016). The radiative forcing model intercomparison project (rfmip): experimental protocol for cmip6. *Geoscientific Model Development*, *9*, 3447–3460. doi: 10.5194/gmd-9-3447-2016

Po-Chedley, S., Armour, K. C., Bitz, C. M., Zelinka, M. D., Santer, B. D., & Fu, Q. (2018). Sources of intermodel spread in the lapse rate and water vapor feedbacks. *Journal of Climate*, *31*, 3187 - 3206. doi: 10.1175/JCLI-D-17-0674.1

Proistosescu, C., & Huybers, P. J. (2017). Slow climate mode reconciles historical and model-based estimates of climate sensitivity. *Science Advances*, *3*. doi: 10.1126/sciadv.1602821

Qu, X., Hall, A., Klein, S. A., & DeAngelis, A. M. (2015). Positive tropical marine low-cloud cover feedback inferred from cloud-controlling factors. *Geophysical Research Letters*, *42*, 7767–7775. doi: 10.1002/2015GL065627

Richardson, T. B., Forster, P. M., Smith, C. J., Maycock, A. C., Wood, T., Andrews, T., … Watson-Parris, D. (2019). Efficacy of climate forcings in pdrmip models. *Journal of Geophysical Research: Atmospheres*, *124*, 12824-12844. doi: 10.1029/2019JD030581

Ridley, J. K., Blockley, E. W., & Jones, G. S. (2022). A change in climate state during a pre-industrial simulation of the cmip6 model hadgem3 driven by deep ocean drift. *Geophysical Research Letters*, *49*. doi: 10.1029/2021GL097171

Rugenstein, M., & Armour, K. C. (2021). Three flavours of radiative feedbacks and their implications for estimating equilibrium climate sensitivity. *Geophysical Research Letters*, *48*. doi: 10.1029/2021GL092983

Rugenstein, M., Bloch-Johnson, J., Abe-Ouchi, A., Andrews, T., Beyerle, U., Cao, L., … Yang, S. (2019). Longrunmip: Motivation and design for a large collection of millennial-length aogcm simulations. *Bulletin of the American Meteorological Society*, *100*, 2551–2570. doi: 10.1175/BAMS-D-19-0068.1

Salvi, P., Ceppi, P., & Gregory, J. M. (2022). Interpreting differences in radiative feedbacks from aerosols versus greenhouse gases. *Geophysical Research Letters*, *49*. doi: 10.1029/2022GL097766

Salvi, P., Gregory, J. M., & Ceppi, P. (2023). Time-evolving radiative feedbacks in the historical period. *Journal of Geophysical Research: Atmospheres*, *128*. doi: 10.1029/2023JD038984

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., … Zelinka, M. D. (2020). An assessment of earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, *58*(4). doi: 10.1029/2019RG000678

Smith, C. J., & Forster, P. M. (2021). Suppressed late-20th century warming in cmip6 models explained by forcing and feedbacks. *Geophysical Research Letters*, *48*. doi: 10.1029/2021GL094948

Smith, D., Booth, B., Dunstone, N., Eade, R., Hermanson, L., Jones, G. S., … Thompson, V. (2016). Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. *Nature Climate Change*, *6*, 936–940. doi: 10.1038/nclimate3058

Smith, D., Gillett, N. P., Simpson, I. R., Athanasiadis, P. J., Baehr, J., Bethke, I., … Ziehn, T. (2022). Attribution of multi-annual to decadal changes in the climate system: The large ensemble single forcing model intercomparison project (lesfmip). *Frontiers in Climate*. doi: 10.3389/fclim.2022.955414

Soden, B. J., & Held, I. M. (2006). An assessment of climate feedbacks in coupled ocean–atmosphere models. *Journal of Climate*, *19*, 3354–3360. doi: 10.1175/JCLI3799.1

Soden, B. J., Held, I. M., Colman, R., Shell, K. M., Kiehl, J. T., & Shields, C. A. (2008). Quantifying climate feedbacks using radiative kernels. *Journal of Climate*, *21*, 3504–3520. doi: 10.1175/2007JCLI2110.1

Stephens, G. L., Kahn, B. H., & Richardson, M. (2016). The super greenhouse effect in a changing climate. *Journal of Climate*, *29*, 5469–5482. doi: 10.1175/JCLI-D-15-0234.1

Titchner, H. A., & Rayner, N. A. (2014). The met office hadley centre sea ice and sea surface temperature data set, version 2: 1. sea ice concentrations. *Journal of Geophysical Research: Atmospheres*, *119*, 2864-2889. doi: 10.1002/2013JD020316

van der Dussen, J. J., de Roode, S. R., Gesso, S. D., & Siebesma, A. P. (2015). An les model study of the influence of the free tropospheric thermodynamic conditions on the stratocumulus response to a climate perturbation. *Journal of Advances in Modeling Earth Systems*, *7*, 670-691. doi: 10.1002/2014MS000380

Vial, J., Dufresne, J. L., & Bony, S. (2013). On the interpretation of inter-model spread in cmip5 climate sensitivity estimates. *Climate Dynamics*, *41*, 3339–3362. doi: 10.1007/s00382-013-1725-9

Webb, M. J., Lock, A. P., & Ogura, T. (2024). What are the main causes of positive subtropical low cloud feedbacks in climate models? *Journal of Advances in Modeling Earth Systems*, *16*. doi: 10.1029/2023MS003716

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., … 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*, 357—380. doi: 10.1002/2017MS001115

Wills, R. C. J., Dong, Y., Proistosecu, C., Armour, K. C., & Battisti, D. S. (2022).

Systematic climate model biases in the large-scale patterns of recent sea-surface temperature and sea-level pressure change. *Geophysical Research Letters*, *49*. doi: 10.1029/2022GL100011

Zelinka, M., Zhou, C., & Klein, S. A. (2016). Insights from a refined decomposition of cloud feedbacks. *Geophysical Research Letters*, *43*. doi: 10.1002/2016GL069917

Feedbacks, Pattern Effects, and Efficacies in a Large Ensemble of HadGEM3-GC3.1-LL Historical Simulations

Harry Mutton[1], Timothy Andrews[1], Leon Hermanson[1], Melissa Seabrook[1], Doug M. Smith[1], Mark A. Ringer[1], Gareth S. Jones[1], Mark J. Webb[1]

[1]Met Office Hadley Centre, Exeter, UK

Key Points:

- Natural variability causes a 3-6K range in Effective Climate Sensitivity in a large single model ensemble of historical simulations.
- Differences in tropical and polar warming strongly influence longwave clear-sky and shortwave clear-sky feedbacks respectively.
- Deficiencies in simulating observed tropical and polar warming cause different feedbacks in historical and amip-piForcing experiments.

Corresponding author: Harry Mutton, `harry.mutton@metoffice.gov.uk`

Abstract

Climate feedbacks over the historical period (1850–2014) have been investigated in large ensembles of historical, hist-ghg, hist-aer, and hist-nat experiments, with 47 members for each experiment. Across the historical ensemble with all forcings, a range in estimated Effective Climate Sensitivity (EffCS) between approximately 3–6 K is found, a considerable spread stemming solely from initial condition uncertainty. The spread in EffCS is associated with varying Sea Surface Temperature (SST) patterns seen across the ensemble due to their influence on different feedback processes. For example, the level of polar amplification is shown to strongly control the amount of sea ice melt per degree of global warming. This mechanism is responsible for the large spread in shortwave clear-sky feedbacks and is the main contributor to the different forcing efficacies seen across the different forcing agents, although in HadGEM3-GC3.1-LL these differences in forcing efficacy are shown to be small. The spread in other feedbacks is also investigated, with the level of tropical SST warming shown to strongly control the longwave clear-sky feedbacks, and the local surface-air-temperatures and large scale tropospheric temperatures shown to influence cloud feedbacks. The metrics used to understand the spread in feedbacks can also help to explain the disparity between feedbacks seen in the historical experiment simulations and a more accurate modeled estimate of the feedbacks seen in the real world derived from an atmosphere-only experiment prescribed with observed SSTs (termed amip-piForcing).

Plain Language Summary

Understanding how the Earth's climate responds to an imposed forcing such as an increase in greenhouse gases or aerosols is an important issue relevant to climate mitigation and adaptation policies on the global scale. One way we can understand this is by analysing the historical period (1850–2014), a period over which the climate has already changed substantially due to human induced forcings, and also a period over which observations allow us to compare modeled changes in climate with the changes seen in the real world. Here, we use a large ensemble of climate model simulations of the historical period were we aim to understand a) how natural variability causes differences in the global temperature response to the same imposed forcing, b) what causes different forcing agents (e.g. greenhouse gases or aerosols) to be more or less effective at warming or cooling the planet, and c) whether historical simulations - where the climate model simulates its own sea surface temperatures - capture the same response to historical forcings as an atmosphere-only simulation prescribed with observed sea surface temperatures. We find that the pattern of sea surface temperatures (particularly the levels of tropical and polar warming) is key to understanding each of these points.

1 Introduction

Climate sensitivity and feedbacks provide valuable information about how the Earth's temperature changes in response to an imposed forcing such as an increase in greenhouse gases, aerosols, or volcanic emissions (Sherwood et al., 2020; Forster et al., 2021). Typically, equilibrium climate sensitivity (ECS) is defined as the equilibrium global temperature increase in response to a doubling of CO_2 and can be related to CO_2 forcing and climate feedbacks using a simple energy balance model (Equation 1) (e.g. Sherwood et al. (2020)).

$$ECS = -F_{2\times CO_2}/\lambda \tag{1}$$

Here, $F_{2\times CO_2}$ is the radiative forcing associated with a doubling of CO_2 and the feedback parameter λ is the radiative response per degree of global temperature change. Currently, the assessed likely range of ECS extends from 2.5°C – 4.0°C (Forster et al., 2021). Since constraining ECS is important for improving our understanding of how the Earth's climate is likely to change in the future, informing climate related mitigation and adaptation policy on the global scale, improving our understanding of different climate feedbacks and why they vary is vital.

The feedback parameter λ can be defined using Equation 2 (e.g. Gregory et al. (2004)).

$$\lambda = d(N - F)/dT_s \tag{2}$$

Here F is the radiative forcing, N is the top of atmosphere radiative flux, and T_s is the surface-air-temperature (in this case, all terms are global mean quantities).

In Atmosphere-Ocean General Circulation Models (AOGCMs), λ and ECS are typically estimated via a linear regression of global T_s and N over the first 150 years of an abrupt-4xCO2 simulation (T. Andrews et al., 2012; Dong et al., 2021; Gregory et al., 2004). The abrupt-4xCO2 simulation is an AOGCM experiment where the atmospheric concentration of CO_2 is abruptly quadrupled and then held constant. This regression method is used in favour of calculating ECS directly from two equilibrium states due to the long timescales needed to equilibrate the deep ocean and the substantial computational cost associated with this (T. Andrews et al., 2022; Rugenstein et al., 2019). ECS estimates produced from these non-equilibrium states are called the Effective Climate Sensitivity (EffCS) (Dong et al., 2021; Sherwood et al., 2020; T. Andrews et al., 2015; Rugenstein & Armour, 2021).

λ and EffCS can also be estimated from simulations of the historical record (1850 to present day), estimating λ over the historical period and applying this to Equation 1 where $F_{2\times CO_2}$ has been diagnosed from an abrupt-4xCO2 run (Gregory et al., 2020). These estimates tend to produce an EffCS smaller than that predicted solely from an abrupt-4xCO2 experiment, largely due to the time variations in λ caused by evolving SST patterns and the different timescales involved in the response to an imposed forcing (T. Andrews et al., 2019; Gregory et al., 2020; Proistosescu & Huybers, 2017). This "pattern effect" describes how a different global radiative response can be generated by the same global temperature change due to different patterns of SSTs (Rugenstein & Armour, 2021; Gregory & Andrews, 2016). In this context, the pattern effect is often quantified as the difference in λ between historical and abrupt-4xCO2 experiments (T. Andrews et al., 2018).

Estimates of λ from historical and abrupt-4xCO2 simulations may also differ due to the different forcing agents involved (Marvel et al., 2015). Whilst the abrupt-4xCO2 experiment is only forced by increases in CO_2 concentrations, the historical simulations are also influenced by changes in aerosols and natural forcings such as volcanic emissions (C. J. Smith & Forster, 2021; Salvi et al., 2023). These different forcing agents may vary in how effective they are at warming or cooling the planet; this is called forcing efficacy (Marvel et al., 2015; Richardson et al., 2019; Hansen et al., 2005). Again AOGCMs can be used to investigate this, with experiments simulating the historical period but only applying the forcing for individual forcing agents. Salvi et al. (2022) use this approach to demonstrate that, in the multi-model mean, greenhouse gases tended to have a more stabilising feedback (lower EffCS) compared to aerosols, although substantial variation across different models exists. It is suggested that across different forcing agents, variations in SST pattern changes lead to differing feedbacks (Haugstad et al., 2017). Ceppi and Gregory (2019) suggest that the changes in atmospheric stability induced by these differing SST patterns is a key factor determining the efficacy of a particular forcing (Salvi et al., 2023). Assuming temperature changes and the radiative responses to each forcing agent add linearly, understanding each component of the full historical forcing can help inform our interpretation of historical feedbacks and how they relate to future climate change.

Historical estimates of a model's EffCS can also be deduced from an Atmosphere only General Circulation Model (AGCM) experiment with prescribed SSTs and sea ice from observations between 1870 and 2014 and atmospheric constituents set to pre-industrial levels, termed amip-piForcing (Gregory & Andrews, 2016; Gregory et al., 2020). Because this experiment is forced with observed SSTs it is able to more accurately simulate historical changes in climate compared to the coupled AOGCMs (Gregory & Andrews, 2016). It is found that the EffCS calculated using the amip-piForcing experiment tends to produce an EffCS smaller than that derived from AOGCM historical experiments (i.e. amip-piForcing has a larger pattern effect relative to abrupt-4xCO2) (Gregory et al., 2020; T. Andrews et al., 2019). Again, this difference is often attributed to differences in SST patterns between the two experiments, with coupled historical simulations struggling to simulate observed SST patterns (Gregory et al., 2020; Wills et al., 2022). Over recent years, observed SSTs demonstrate a marked cooling in the East Pacific and Southern Ocean and more warming over the West Pacific, leading to more negative feedbacks and a lower EffCS. The inability of AOGCM simulations to capture observed trends in SST patterns is a key issue currently facing the scientific community and raises questions regarding how this impacts our understanding of climate sensitivity and feedbacks. The "peculiar" trend in SST patterns as termed by Fueglistaler and Silvers (2021) may have occurred through unforced variability and it may then be by chance that the real world SSTs have evolved in a way that induces a more strongly stabilising feedback. Or, it is possible that the trend is forced, e.g. by aerosols or volcanic emissions (D. Smith et al., 2016; Gregory et al., 2020; Hwang et al., 2024), and our AOGCMs struggle to simulate the real world SSTs accurately due to limitations in our current modelling capabilities.

To date, most of the work examining radiative feedbacks, pattern effects and efficacies has been limited to idealised experimental designs or small ensembles of historical AOGCM simulations with a single model, or via model intercomparisons such as the Coupled Model Intercomparison Project (CMIP) (Eyring et al., 2016), where still only relatively small ensemble sizes are available. Questions remain on the influence of natural variability in historical climate change on diagnosed estimates of feedbacks, the quantification of the forced response to different forcings and whether radiative feedback simulated in AOGCM historical simulations are consistent with observed estimates. Large initial condition ensembles with a single model are useful to address this. For example, previously, large ensembles have been shown to provide valuable insight into the separation of forced climate change and internal variability (Kay et al., 2015). From a sea ice sensitivity perspective, Kay et al. (2011) demonstrate that using an ensemble to quantify internal variability shows that recent trends in sea ice decline cannot be reproduced from modeled internal variability alone. Adams and Dessler (2019) employ a 100 member ensemble of historical simulations to show that internal variability could be a key contributor to the difference in Transient Climate Response (TCR) estimates between models and observations. Applying the analysis of this 100 member ensemble to the study of climate sensitivity and feedbacks over the historical period, Dessler et al. (2018) highlight a large range in EffCS estimates between 2.1 and 3.9K. They note that given that the real world 20$^{\text{th}}$ century is just one realisation of a range of possible realities, due to that large internal variability, we should not expect estimates of EffCS from observations to be a precise guide to the real world's forced response. Alongside this, they note that that different forcing efficacies, imperfect observations, and uncertainty in 20th century forcing also pose challenges for interpreting EffCS from the historical period. Gregory et al. (2020) also noted the high levels of internal variability over the historical record showing how this variability contributed to uncertainty to estimates of EffCS.

In this paper we use a new set of four large ensembles of HadGEM3-GC3.1-LL historical and single forcing simulations performed for the Large Ensemble Single Forcing Model Intercomparison Project (LESFMIP) (D. Smith et al., 2022), aiming to address the following questions.

1. how does natural variability cause differences and spread in climate feedbacks in response to the same imposed forcing?
2. What causes different efficacies of different historical forcing agents?
3. Can AOGCM historical simulations – where the model simulates it's own SSTs - capture the radiative feedback and EffCS estimated from AGCM experiments prescribed with observed SSTs?

Previously, T. Andrews et al. (2019) investigated EffCS and feedbacks in HadGEM3-GC3.1-LL in a 4 member ensemble of historical simulations, finding a net feedback (λ) of -0.86 $\pm$ 0.4 $Wm^{-2}K^{-1}$ (5-95%). This ensemble mean estimate is more negative than the abrupt-4$\times$CO2 feedback in HadGEM3-GC3.1-LL of -0.63 $Wm^{-2}K^{-1}$, although the 5-95% confidence range does extend up to -0.46 $Wm^{-2}K^{-1}$. The large spread in λ was found to be partly caused by considerable variations in Antarctic sea ice. This variability in sea ice inhibited accurate evaluation of the model's historical forced EffCS. There, T. Andrews et al. (2019) were limited to an ensemble of only 4 simulations, so questions remain about whether the full diversity of variability was sampled. Here we investigate this with a much larger ensemble of 47 members.

In the following section we describe the model and experimental setup used. Section 3 presents the results and Section 4 provides a discussion and conclusions.

2 Methods

2.1 HadGEM3-GC3.1-LL

The analysis in this paper uses simulations performed using HadGEM3-GC3.1-LL, an AOGCM with an atmospheric resolution of 135 km with 85 vertical levels and an ocean resolution of 1°and 75 vertical levels (M. B. Andrews et al., 2020). Further details can be found in Williams et al. (2017) where a description of the model's configuration is given.

2.2 Large Historical Ensemble

In this analysis, ensembles of historical, hist-ghg, hist-aer, and hist-nat experiment are used, with 47 members of each experiment mostly consisting of simulations performed for LESFMIP. These experiments are AOGCM simulations analysed between 1850–2014 with atmospheric constituents set to historical levels. Here, the historical experiment includes all forcing agents, whilst the hist-ghg, hist-aer, and hist-nat contain only the forcing associated with well mixed greenhouse gases, anthropogenic aerosols, and natural forcings respectively (Gillett et al., 2016). Each ensemble member differs only in their initial conditions branching from the piControl experiment at different times (1850, 1885, and every 10 years between 1860 and 2300). The piControl experiment is an AOGCM experiment with atmospheric constituents set to pre-industrial levels. The 47 ensemble members consist of 45 simulations performed as part of the LESFMIP ensemble (D. Smith et al., 2022), and two simulations previously analysed in T. Andrews et al. (2019). Only two of the four simulations used in T. Andrews et al. (2019) were analysed here since the other two members had identical branch times to members of the LESFMIP ensemble.

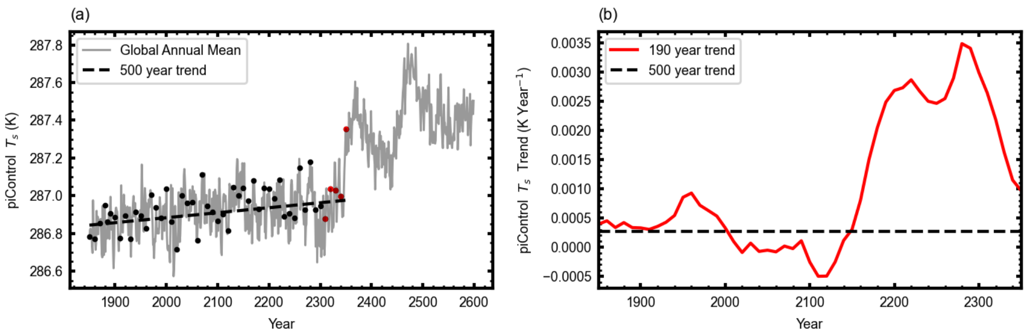

Figure 1. (a) Timeseries of global annual mean T_s in the piControl experiment (grey line), 500 year trend (dashed black line), and branch times for each of the historical and single forcing experiment ensemble members (dots). Red dots indicate the ensemble members that have been excluded due to the strong warming seen in the piControl experiment. (b) 190 year piControl trend for each ensemble member branch date (red), and 500 year piControl trend (horizontal black dashed line).

2.3 piControl and Detrending

To compare ensemble members in the 47 member ensembles, the control drift must be removed from each simulation. For this analysis, this drift is removed by calculating the trend over the first 500 years of the piControl experiment via linear regression and subtracting the corresponding time period from each ensemble member. The piControl timeseries of global annual mean T_s and the 500 year trend is shown in Figure 1a where the dots depict the branch dates for each member of the historical ensemble. This method of control drift removal is chosen in favour of subtracting the piControl year by year to avoid unnecessarily introducing more noise into the historical simulations. The 500 year trend is also preferred above subtracting the 190 year trend across the corresponding piControl period due to issues introduced towards the end of the piControl simulation, where a marked global warming is seen at around 2350. This warming has been previously documented by Ridley et al. (2022) where it is attributed to the onset of deep convection in the Weddell and Ross Sea gyres due to a destabilising of the Southern Ocean. When removing the control drift from the historical ensemble, any drift removed is assumed to be present in the historical ensemble member. For the trend seen over the first 500 years of the control run this is a reasonable assumption, however in the case of the large warming seen around 2350, this assumption may not hold. The impact that this warming has on the 190 year control trend for the respective historical ensemble branch dates is shown in Figure 1b. Here, unsurprisingly, a strong positive trend is seen for ensemble members that branch after the year 2150. We found no evidence that the warming seen in the piControl experiment is present in historical ensemble members initiated up to 2300, but to avoid this feature contaminating the comparison of ensemble members, the last 5 ensemble members have been removed from the analysis. This is why although the LESFMIP ensemble consists of 50 members, only 45 of them are used here.

2.4 Diagnosing Historical Forcing

Whilst λ can be calculated for the abrupt-4xCO2 and amip-piForcing experiments from only T_s and N (since the F is constant), the time varying F over the historical period means that in order to estimate λ, we must first diagnose F.

Table 1. Description of experimental setup used.

Experiment Name	Atmospheric Constituents	SSTs	Run Time	Ensemble Size
Coupled experiments				
piControl	pre-industrial	free running	1850 – 3850	1
abrupt-4xCO2	pre-industrial $CO_2 \times 4$	free running	1850 – 2350	1
historical	historical	free running	1850–2014	47
hist-ghg	historical well mixed green-house gases	free running	1850–2014	47
hist-aer	historical aerosols	free running	1850–2014	47
hist-nat	historical natural forcing	free running	1850–2014	47
Atmosphere-only experiments				
amip-piForcing	pre-industrial	historical observed	1870 – 2014	1
piClim-control	pre-industrial	piControl	1850 – 1890	3
piClim-histall	historical to 2014 then ssp-245 to 2100	piControl	1850 – 2100	3
piClim-histghg	historical well mixed green-house gases only to 2014 then ssp-245 to 2100	piControl	1850 – 2100	3
piClim-histaer	historical aerosols only to 2014 then ssp-245 to 2100	piControl	1850 – 2100	3
piClim-histnat	historical natural forcing only to 2014 then ssp-245 to 2100	piControl	1850 – 2100	3

Typically, the historical F is diagnosed using RFMIP experiments piClim-control and piClim-histall (Forster et al., 2016; Pincus et al., 2016). These are two AGCM experiments with prescribed SSTs from the piControl simulation. For piClim-control, atmospheric constituents are set to pre-industrial levels and the experiment is run for 30 years. Averaging over the 30 years provides the control state. For piClim-histall atmospheric constituents are set to historical levels between 1850 – 2014 and to ssp-245 levels between 2015 and 2100. Subtracting the 30 year mean piClim-control top of atmosphere radiative flux from the 1850 – 2100 piClim-histall top of atmosphere flux provides F, with years 1850–2014 relevant for the analysis of the historical period.

In order to diagnose F for the individual forcing components that correspond to the hist-ghg, hist-aer, and hist-nat experiments, a similar experimental setup to the piClim-histall experiment is used but only the forcing from the relevant component is applied. These experiments are termed piClim-histghg, piClim-histaer, and piClim-histnat (Forster et al., 2016; Pincus et al., 2016).

A summary of the setup for each experiment used in this paper is presented in Table 1.

3 Results

3.1 Diagnosing Feedbacks in Historical and Single Forcing Ensembles

As discussed in the introduction, the feedback parameter (λ) can be estimated via linear regression of global annual mean surface-air-temperatures (T_s) against top of atmosphere radiative fluxes (N) minus the changes in flux associated with the radiative forcing (F). Timeseries of these diagnostics are presented in Figure 2, where 2a and b show the anomalous global annual mean T_s and anomalous global annual mean N respectively in every ensemble member and in each experiment, and 2c shows the global annual mean F associated with each experiment. From Figure 2a it can be seen that the cooling effect of anthropogenic aerosols and natural forcings is approximately offset by the warming effect of increased greenhouse gases between 1850 and 1990. Here, the F associated with greenhouse gases and aerosols gradually increase, however, after approximately 1990 the aerosol F remains relatively constant (around -1.5 Wm^{-2}) whilst the F associated with greenhouse gases continues to increase (Figure 2c) (T. Andrews et al., 2019). This leads to a net positive F after 1990 in the historical experiment which results in an increase in global mean T_s, warming by approximately 0.8 K by 2014. A detailed analysis of HadGEM3-GC3.1-LL historical simulations is presented in M. B. Andrews et al. (2020). An example of how λ is calculated from these timeseries of T_s, N, and F is presented in Figure 2d, where, for the first ensemble member in the historical experiment, a feedback parameter of -0.85 $\pm$ 0.15 Wm^{-2}K^{-1} is estimated. There the uncertainty is estimated as $\pm$1.645 standard deviations, calculated from the standard error of the linear fit.

One assumption made when estimating λ using timeseries of T_s, N, and F is that the changes in global mean T_s associated with the forcing is zero (i.e. the surface-air-temperature change between piClim-control and piClim-histall is zero). This is generally a reasonable assumption to make, given that the prescribed SSTs do not warm and therefore any changes in land surface temperatures are constrained to be small (Lambert et al., 2011). However, despite this temperature change being small, taking this into account can substantially affect the values of λ estimated. This caveat is noted in Hansen et al. (2005) and Vial et al. (2013) and becomes a particularly relevant issue when comparing feedbacks in the historical experiment to feedbacks in the amip-piForcing experiment, since there is no forced temperature change in the amip-piForcing experiment where $F = 0$ by construction. To handle this issue, in this paper, λ has been calculated accounting for this forced temperature change (Equation 3).

$$\lambda = d(N - F)/d(T_s - \delta T_{s_{forced}}) \tag{3}$$

Where $\delta T_{s_{forced}}$ is calculated as the change in global surface-air-temperature between piClim-control and the relevant piClim-hist experiment used to diagnose F. To simplify the notation, we refer to ($T_s - \delta T_{s_{forced}}$) simply as T_s. Similarly, later when analysing atmospheric temperatures (T_a), we refer to ($T_a - \delta T_{a_{forced}}$) simply as T_a.

To summarise the feedbacks seen across the different experiments analysed, boxplots of feedbacks in the historical and single forcing experiments and markers showing the feedbacks in both amip-piForcing and abrupt-4xCO2 experiments are shown in Figure 3b. Here the net feedback has been decomposed into shortwave clear-sky (SW_{cs}), longwave clear-sky (LW_{cs}), and cloud radiative effect (cre) components. Such a decomposition is useful since it can help isolate the different processes and feedback mechanisms involved. λ_{SWcs}, λ_{LWcs}, and λ_{cre} are calculated by decomposing N and F into the relevant fluxes when applying Equation 3. From Figure 3b, a large spread in feedbacks across the historical ensemble can be seen, ranging from approximately -0.7 to -1.3 Wm^{-2}K^{-1}. Using a 2×CO_2 F of 4.05 Wm^{-2} for HadGEM3-GC3.1-LL (T. Andrews

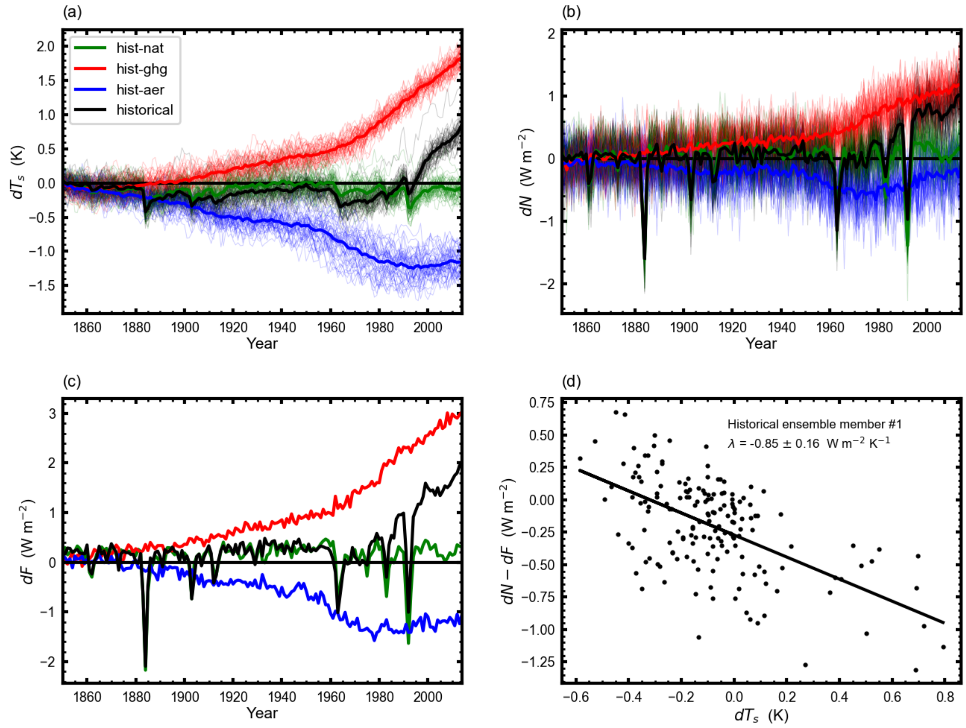

Figure 2. (a) Timeseries of anomalous global annual mean T_s in the historical and single forcing experiments. Thick lines indicate the ensemble mean and thinner lines represent each individual ensemble member. (b) Timeseries of anomalous global annual mean N in the historical and single forcing experiments. Again, thick lines indicate the ensemble mean and thinner lines represent each individual ensemble member. (c) Timeseries of global annual mean F for historical and single forcing scenarios averaged across the three ensemble members for each experiment. (d) Example of method used to estimate λ, where λ is calculated by linearly regressing T_s against $(N - F)$. Each dot represents a year in the historical experiment and the black line shows regression line where the slope (λ) is estimated to be -0.85 $\pm$ 0.15 $\mathrm{Wm^{-2}K^{-1}}$. Here, the uncertainty is estimated as $\pm$1.645 standard deviations, calculated from the standard error of the linear fit.

et al., 2019), and applying Equation 1, such a range in feedbacks leads to an estimate of EffCS between approximately 3 and 6K (Figure 3a). This highlights the role of internal variability in causing different feedback and EffCS estimates over the historical period. The spread in feedbacks seen in the historical and single forcing experiments is largest in the hist-nat experiment and smallest in the hist-ghg experiment, possibly due to the varying signal to noise ratios across the different experiments. The T_s changes in the hist-nat experiment are generally small (Figure 2a), and the natural F is also small with an occasional strong but short-lived signal caused by volcanic emissions (Figure 2c). This causes the regression of T_s against $(N-F)$ to be relatively noisy compared to the hist-ghg experiment where both T_s and $(N-F)$ have a much stronger signal. This is also consistent with the contrast in estimated uncertainty of the linear fit of T_s and $(N-F)$ where for each experiment, the standard error of the linear fit of every ensemble member has be estimated. The estimation of λ_{net} in the hist-ghg experiment has an average 5-95% interval of $\pm$ 0.066 $\mathrm{Wm^{-2}K^{-1}}$, whereas for hist-nat, the mean 5-95% interval is ± 0.25 $\mathrm{Wm^{-2}K^{-1}}$.

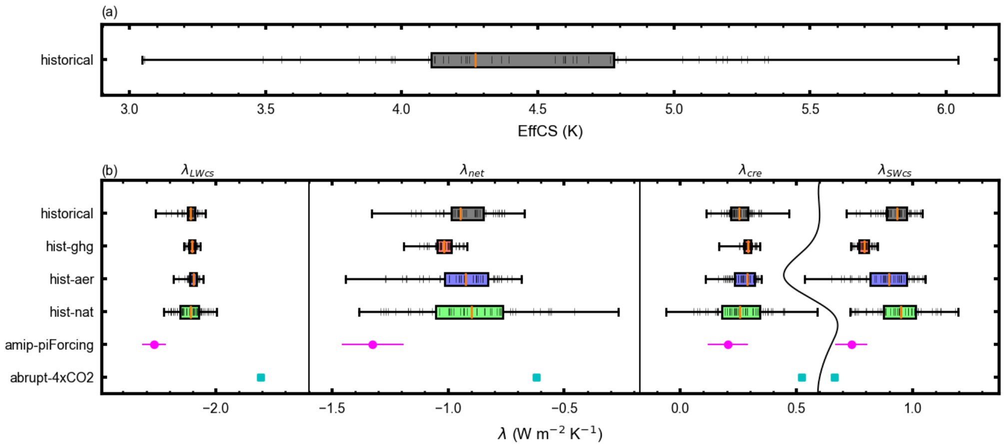

Figure 3. (a) Boxplot of EffCS across the historical ensemble (1850–2014). (b) Boxplots of feedbacks in the historical and single forcing ensembles (1850–2014), amip-piForcing experiment (1870–2014), and abrupt-4xCO2 experiment (first 150 years). For each boxplot, the vertical black lines indicate each ensemble member, the whiskers indicate the maximum and minimum feedbacks seen in the ensemble, the boxes indicate the interquartile range, and the vertical orange line represents the median value. Error bars on amip-piForcing indicate the 5-95% confidence interval, calculated from the standard error of the linear fit.

A further decomposition of λ_{cre} into shortwave and longwave components is shown in Figure S1. There, the largest contribution to the spread in λ_{cre} comes from the shortwave component, consistent with the strong influence of low cloud feedbacks, and the cancelling of the longwave and shortwave response to changes in high cloud.

The feedbacks seen in each historical and single forcing experiment are largely consistent with each other (i.e. differing forcing efficacies do not appear to be strongly evident in HadGEM3-GC3.1-LL), although a slightly more negative median feedback is seen in the hist-ghg experiment, consistent with the findings of Salvi et al. (2022). In Figure 3, the more negative median feedback in the hist-ghg experiment is shown to be caused by a weaker λ_{SWcs}, although due to the large spread in historical, hist-aer, and hist-nat feedbacks, the lower tails of the feedbacks in these experiments extend to be more negative than the lower tail of the hist-ghg experiment. The amip-piForcing and abrupt-4xCO2 feedbacks are also shown in Figure 3b. For each component of λ_{net}, the amip-piForcing feedback lies towards the lower tail of the historical ensemble, a behaviour most strongly seen in the λ_{SWcs}, and λ_{LWcs} components.

Maps of the ensemble mean feedbacks and amip-piForcing feedbacks are shown in Figure 4 to help identify where different feedbacks are located and to highlight regions where feedbacks differ across the range of experiments analysed. The spatial feedback map is calculated by regressing the local $(N-F)$ against the global mean T_s changes. Here the ensemble mean feedbacks are calculated by taking the regression of the mean rather than calculating the feedback for each ensemble member and averaging across the ensemble. This was done to help reduce the noise in the regression of $(N-F)$ and T_s when calculating the feedbacks.

From Figure 4, it can be seen that different feedbacks dominate in different regions. For example, in general λ_{SWcs} is strongly positive at higher latitudes and small at lower latitudes. This is because the sea ice feedback is a key feedback affecting the SW_{cs} fluxes. The strong positive λ_{SWcs} seen over the northern hemisphere land masses is likely related to snow and land ice feedbacks, and the strong negative λ_{SWcs} seen in the Southern Ocean in the hist-aer experiment may be caused by ocean convective events that bring warmer water to the surface due to destabilization of the ocean, similar to those discussed in (Ridley et al., 2022).

With the exception of the Southern Ocean feature seen in the hist-aer experiment, the λ_{LWcs} is generally negative everywhere across all experiments, although a few small regions in the amip-piForcing experiment also have positive λ_{LWcs}. The λ_{LWcs} is largely composed of the Planck, lapse rate, and water vapour feedbacks. This term is generally large and negative due to the strong Planck response. Over the Southern Ocean in the hist-aer experiment, since this region warms, which is of opposite sign to the cooling seen over the rest of the planet, the λ_{LWcs} is strongly positive in this region. In the tropics, the lapse rate and Planck feedbacks are typically negative, therefore the positive λ_{LWcs} regions in the amip-piForcing experiment over the tropics are likely caused by the water vapour feedback (Stephens et al., 2016).

λ_{cre} exhibits relatively large spatial variations. In the historical and single forcing experiments (particularly hist-aer) a strongly positive λ_{cre} is seen over the North Pacific, highlighting the role of positive cloud feedbacks in the sub-tropical cloud decks in subsidence regions. Again, λ_{cre} has been decomposed into longwave and shortwave components (Figure S2). The strong λ_{cre} over the North Pacific is caused by shortwave cloud feedbacks, and over tropical high cloud regions, e.g. the Indo-Pacific warm pool region, strong shortwave and longwave cloud feedbacks cancel, causing the relatively weak λ_{cre} over much of the tropics.

From these maps of feedbacks, it can be seen that although in the global mean, different efficacies are not particularly large in HadGEM3-GC3.1-LL, spatially, large variations do exist between the different experiments.

As mentioned in the introduction, differences in feedbacks across experiments and ensemble members are generally thought to be fundamentally caused by differing SST patterns. Therefore, to help understand the differences in feedbacks seen in Figure 4, ensemble mean T_s patterns are shown in Figure 5. Similar to the maps of λ, these have been calculated by regressing the ensemble mean local changes in T_s against the ensemble mean global mean T_s, written as $dT_s/d\bar{T}_s$, where the bar indicates a global mean. In Figure 5, the strongest regions of $dT_s/d\bar{T}_s$ occur in the Arctic, with weaker more spatially uniform $dT_s/d\bar{T}_s$ seen over the tropics. Over the Southern Ocean, large variations in $dT_s/d\bar{T}_s$ are seen across the different experiments. Here, hist-nat exhibits the strongest $dT_s/d\bar{T}_s$ whilst hist-aer exhibits a negative $dT_s/d\bar{T}_s$ (i.e. although global mean T_s is decreasing in the hist-aer experiment, the southern ocean warms). As previously mentioned, this may be caused by ocean convective events that bring warmer water to the surface due to destabilization of the ocean (Ridley et al., 2022). In the northern hemisphere high latitudes, hist-aer exhibits the strongest $dT_s/d\bar{T}_s$, possibly due to the aerosol F being predominantly based in the northern hemisphere. Over the tropics $dT_s/d\bar{T}_s$ is relatively consistent across each experiment.

Since one of the key aims of this paper is to understand the ensemble spread in feedbacks, maps of the standard deviation in λ in the historical experiment help to highlight the regions that contribute most to this spread (Figure 6). From Figure 6 it can be seen

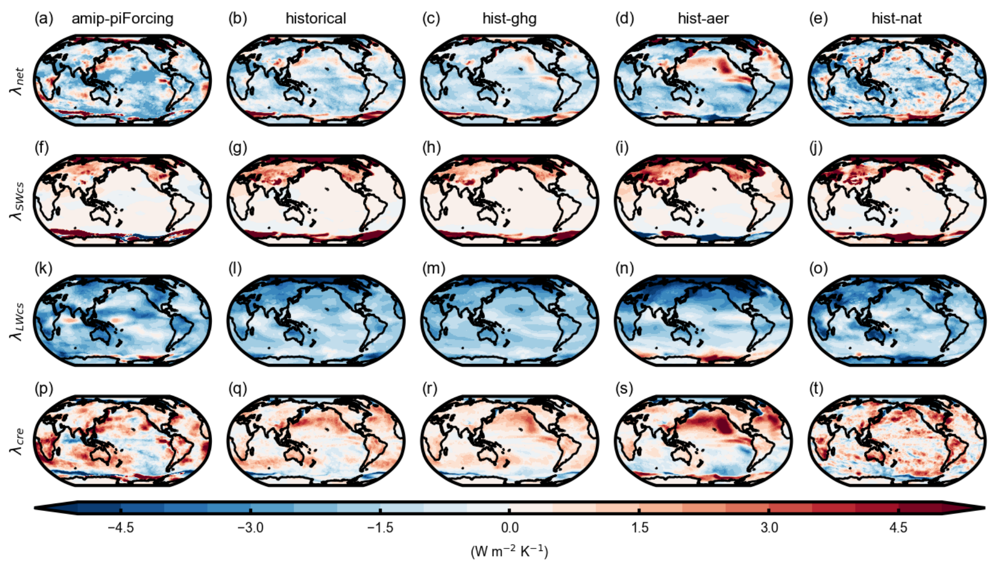

Figure 4. Maps of ensemble mean λ_{net}, λ_{SWcs}, λ_{LWcs}, and λ_{cre} in amip-piForcing, historical, hist-ghg, hist-aer, and hist-nat experiments. Here, λ has been calculated by regressing the ensemble mean local annual mean $(N - F)$ against the ensemble mean global annual mean T_s between 1850 – 2014 for historical and single forcing experiments, and 1870 – 2014 for amip-piForcing.

that for λ_{SWcs} most of the spread comes from the higher latitudes. In contrast, for λ_{cre}, variations in cloud feedbacks across the tropics and subtropics contribute to the spread. λ_{LWcs} exhibits the smallest standard deviations suggesting that this component contributes less to the ensemble spread in feedbacks. This is likely due to the fact that the Planck, lapse rate and water vapour feedbacks are highly constrained by model physics.

The three main scientific aims of this paper were to a) understand how natural variability causes different feedbacks in response to the same imposed forcing (for example, what is it that causes one historical ensemble member to have an net feedback of -1.3 $\mathrm{Wm^{-2}K^{-1}}$ whilst another has a feedback of -0.7 $\mathrm{Wm^{-2}K^{-1}}$?), b) understand what causes different efficacies across different forcing agents, and c) investigate whether the AOGCM historical simulations - where the model simulates its own SSTs - can capture the radiative feedback and EffCS estimated from AGCM experiments prescribed with observed SSTs (i.e. are the feedbacks seen in the historical experiment consistent with those seen in amip-piForcing?). To address these questions, the different components of λ_{net} are investigated in isolation, with Section 3.2 investigating λ_{SWcs}, Section 3.3 investigating λ_{LWcs}, and Section 3.4 investigating λ_{cre}.

3.2 Processes Affecting Shortwave Clear-sky Feedbacks (λ_{SWcs})

This section aims to understand λ_{SWcs} in the historical and single forcing experiments, addressing the cause of the ensemble spread, the disparity between historical and amip-piForcing, and the cause of different efficacies across the different forcing agents. Figure 3 shows that λ_{SWcs} is a key contributor to the ensemble spread in λ_{net}, and the correlation between the two feedbacks is 0.82 across the historical experiment ensemble. Both the maps of λ_{SWcs} and standard deviation in λ_{SWcs} (Figure 4 and Figure 6b) indicate that most of the signal and spread in λ_{SWcs} comes from the higher latitudes, a region where the sea ice albedo feedback is a key process. We suggest that this feedback

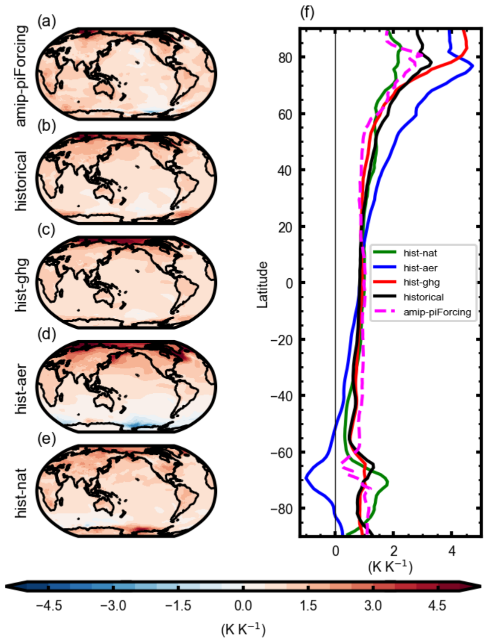

Figure 5. (left) maps of $dT_s/d\bar{T}_s$ in KK^{-1} in each experiment; amip-piForcing, historical, hist-ghg, hist-aer, and hist-nat. Here, $dT_s/d\bar{T}_s$ has been calculated by regressing the ensemble mean local annual mean T_s against the ensemble mean global annual mean T_s between 1850 – 2014 for historical and single forcing experiments, and 1870 – 2014 for amip-piForcing. (right) Zonal mean of maps to the left.

is a key contributor to the spread in λ_{SWcs} and a scatter plot of λ_{SWcs} against global sea ice fraction change per degree of warming ($d(Sea\,Ice)/d\bar{T}_s$) shown in Figure 7a confirms this. There, a correlation of -0.84 is seen between the two variables in the historical experiment over the full time period from 1850 – 2014. As previously mentioned, ultimately, the cause of differing feedbacks can be explained through variations in SST patterns. To understand the varying $d(Sea\,Ice)/d\bar{T}_s$ and λ_{SWcs} across the ensemble, scatter plots of polar $dT_s/d\bar{T}_s$ against global $d(Sea\,Ice)/d\bar{T}_s$ and λ_{SWcs} are shown in Figure 7b and c respectively. Here polar $dT_s/d\bar{T}_s$ is characterised by averaging over latitudes greater than 60°N and lower than 60°S. From Figure 7b and c, a strong relationship between polar $dT_s/d\bar{T}_s$ and both $d(Sea\,Ice)/d\bar{T}_s$ and λ_{SWcs} can be seen. This suggests that the spread in λ_{SWcs} can be understood by the degree of polar amplification across the ensemble.

Figure 7a also indicates that the sea ice albedo feedback is a key reason for the differences in λ_{SWcs} between the historical and amip-piForcing experiments. Here, the amip-piForcing experiment has been analysed only between 1980 and 2014 due to the unrealistic evolution of sea ice in the amip-piForcing experiment prior to 1980 when sea ice observations were sparse (Titchner & Rayner, 2014; T. Andrews et al., 2018). It is therefore important to note that much of the absolute difference in λ_{SWcs} and $d(Sea\,Ice)/d\bar{T}_s$ between the amip-piForcing and historical experiments in Figure 7 may be due to the different time frames analysed. The historical experiment has also been analysed between 1980 and 2014 (Figure 7 non-filled circles) and no substantial change in the relationship between each variable is seen. This does not rule out the possibility that the amip-piForcing evolution of sea ice, polar temperatures, and λ_{SWcs} may have been different for the longer

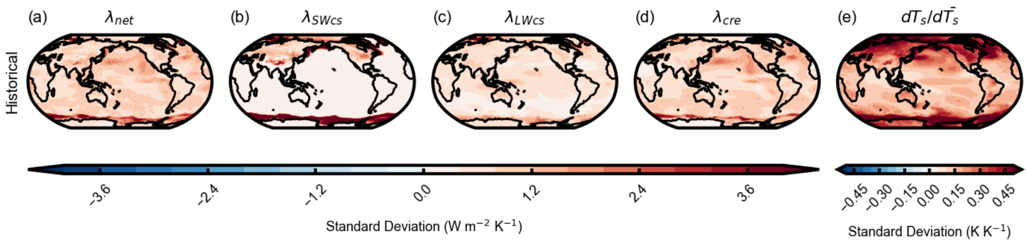

Figure 6. Maps of standard deviation in λ_{net}, λ_{SWcs}, λ_{LWcs}, λ_{cre}, and $dT_s/d\bar{T}_s$ in the historical experiment. Here, λ has been calculated by regressing the local changes in $(N - F)$ against the global mean T_s change, and $dT_s/d\bar{T}_s$ is the local T_s regressed against global mean T_s.

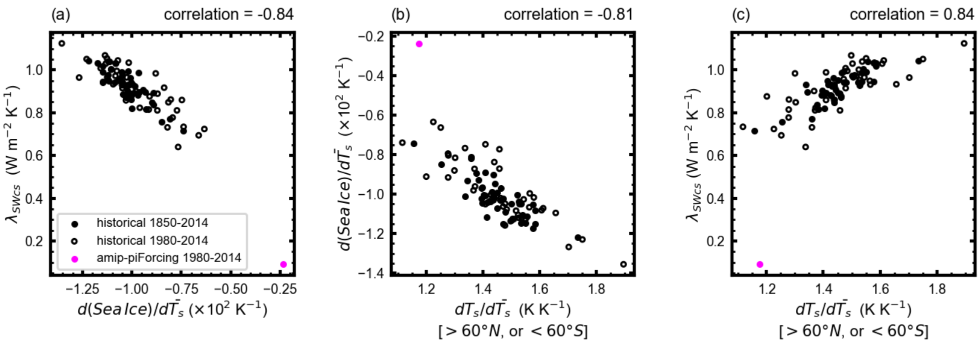

Figure 7. Scatter plots of (a) change in global sea ice per degree of warming against λ_{SWcs}, (b) change in T_s at latitudes greater than 60°N or lower than -60°S per degree of global warming against change in global sea ice per degree of global warming, and (c) change in T_s at latitudes greater than 60°N or lower than 60°S per degree of global warming against λ_{SWcs}. Here, each black dot represents a historical ensemble member where values are calculated between 1850–2014 for the filled black dots, and 1980–2014 for the unfilled black dots. The magenta dots represent the amip-piForcing experiment calculated between 1980–2014 (due to sparse sea ice observations prior to 1980).

period, however, the fact that the amip-piForcing experiment is consistent with the relationship seen in the historical experiment (demonstrated in Figure 7a) would suggest that differences in λ_{SWcs} between historical and amip-piForcing experiments can be explained through this mechanism, and the smaller λ_{SWcs} in amip-piForcing is related to the smaller $d(Sea\,Ice)/d\bar{T}_s$. The fact that in 7b the amip-piForcing experiment does not fit the historical ensemble relationship between polar $dT_s/d\bar{T}_s$ and $d(Sea\,Ice)/d\bar{T}_s$ suggests that that the AOGCMs simulation of the relationship between SSTs and sea ice melt is not the same as the observed relationship in the real world (assuming the relationship seen in amip-piForcing is a good analogue for the real world).

Thus far the ensemble spread and the disparity between historical and amip-piForcing estimates of λ_{SWcs} has been investigated. It is shown that the sea ice albedo feedback is a key process responsible for both, with the level of arctic amplification providing the link between ensemble spread in λ_{SWcs} and T_s patterns. Previously, Dessler (2020) also investigated changes in sea ice and its impact on feedbacks. Consistent with the results shown in Figure 7, Dessler (2020) also found sea ice variability to cause a large spread in λ_{SWcs} in their historical ensemble with a different model, where these feedback variations were linked to changes in different modes of ocean variability. Since Figure 7 highlights a strong relationship between polar SSTs and sea ice, understanding causes of polar SST change and how they are predicted to evolve in a future climate is important.

Other processes could also contribute to the spread in λ_{SWcs}, such as snow melt. This could be responsible for the strong λ_{SWcs} seen over the Northern Hemisphere land masses in Figure 4 f, g, h, i, and j, and the spread in λ_{SWcs} seen in Figure 6b. However, this process is not investigated further here since the strongest spread in λ_{SWcs} is seen over the Arctic and Southern Oceans.

With the understanding gained from Figure 7, the different efficacies of each forcing agent are investigated. Maps of ensemble mean λ_{SWcs} and $dT_s/d\bar{T}_s$ are shown in Figure 8. Here, the hist-ghg experiment is shown and each of the other experiments are shown relative to the hist-ghg values. This enables clearer identification of the differences between each forcing agent.

From Figure 8 the spatial pattern of $dT_s/d\bar{T}_s$ and λ_{SWcs} are shown to be similar, suggesting that the regional change in $dT_s/d\bar{T}_s$ leads to regional changes in λ_{SWcs} due to the close relationship between T_s and sea ice. This is true for both the northern and southern hemisphere and also across each of the experiments. The spatial correlations between $dT_s/d\bar{T}_s$ and λ_{SWcs} across all experiments and each hemisphere are between 0.64 – 0.88, further highlighting the strong coupling between local T_s patterns and local feedbacks. For the historical experiment, in the southern hemisphere, a stronger λ_{SWcs} is associated with a larger Southern Ocean $dT_s/d\bar{T}_s$ relative to hist-ghg. The northern hemisphere maps in 8b show contrasting feedbacks between the Arctic Ocean regions and the slightly lower latitude regions around the Labrador Sea. Over the Arctic Ocean hist-ghg has a stronger λ_{SWcs} compared to the historical simulations, whereas around the Labrador Sea, the historical experiment has the stronger λ_{SWcs}. This is reflected in the $dT_s/d\bar{T}_s$ patterns, where the historical experiment has a weaker $dT_s/d\bar{T}_s$ over the Arctic Ocean, but a stronger $dT_s/d\bar{T}_s$ over the Labrador Sea. This northern hemisphere pattern in λ_{SWcs} and $dT_s/d\bar{T}_s$ relative to hist-ghg is similar to that seen in the hist-aer and hist-nat experiment, where the hist-aer experiment demonstrates the largest positive λ_{SWcs} values and also extends these positive values furthest south.

In the southern hemisphere, unlike the historical experiment, the hist-aer experiment shows strongly negative λ_{SWcs} and $dT_s/d\bar{T}_s$ relative to the hist-ghg experiment. As previously mentioned, this may be due to ocean convection in the Southern Ocean triggered by the ocean becoming unstable (Ridley et al., 2022). This convection could bring warmer water up from below, warming the surface, melting sea ice, and resulting in a negative λ_{SWcs}.

Here, it has been shown that the sea ice albedo feedback and the level of arctic amplification is a key process in producing the large spread in λ_{SWcs} across the ensemble and is also a key reason for the different feedback seen in the historical and amip-piForcing experiments. It has also been shown that the different efficacies seen across the different historical and single forcing experiments can be explained through differing SST patterns (in agreement with Haugstad et al. (2017)), with the experiments with a stronger λ_{SWcs} locally, also exhibiting a larger $dT_s/d\bar{T}_s$.

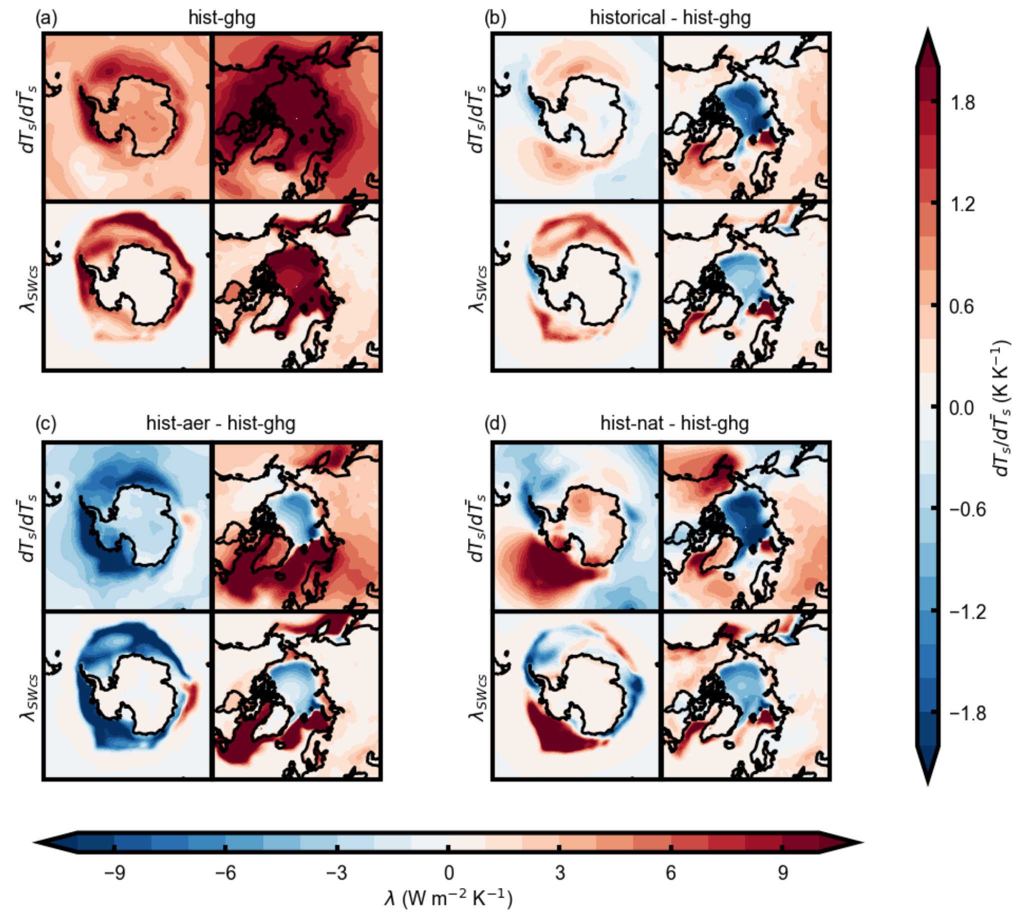

Figure 8. Maps of (top rows) surface warming pattern (KK^{-1}) and (bottom rows) λ_{SWcs} over the (right columns) northern and (left columns) southern hemisphere poles in the (a) hist-ghg experiment and (b) historical, (c) hist-aer and (d) hist-nat experiments relative to hist-ghg.

3.3 Processes Affecting Longwave Clear-sky Feedbacks (λ_{LWcs})

From Figure 3 it can be seen that whilst the λ_{LWcs} does not contribute much to the different efficacies seen in each of the historical and single forcing experiments, it does contribute to the spread in λ_{net} and is also a large source of disparity between the historical and amip-piForcing experiments. Understanding the spread in λ_{LWcs} and the disparity between the historical and amip-piForcing experiments is the aim of this section.

λ_{LWcs} is determined by a combination of the Planck feedback, the water vapour feedback and the lapse rate feedback (T. Andrews & Webb, 2018). The water vapour and lapse rate feedbacks have been shown to be strongest in the tropical troposphere (Soden et al., 2008; T. Andrews & Webb, 2018), since the tropical atmosphere closely follows a moist adiabatic lapse rate and therefore any warming at the surface is amplified vertically in the atmosphere (Po-Chedley et al., 2018). To investigate the λ_{LWcs} in the historical ensemble, first, plots of zonal mean atmospheric temperature regressed against global mean T_s ($dT_a/d\bar{T}_s$) are analysed (Figure 9). Note that as previously discussed, here, the atmospheric temperature (T_a) has had any changes associated with the forcing subtracted from it (see discussion following Equation 3). This means that the CO_2

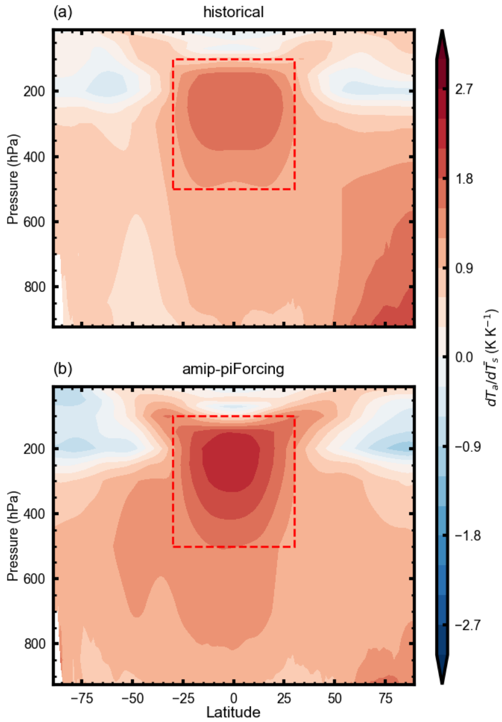

Figure 9. Zonal mean changes in temperature per degree of global warming in the (a) historical and (b) amip-piForcing experiments.

driven stratospheric cooling in the historical experiment is removed, and a more accurate comparison between historical and amip-piForcing experiments can be made.

From Figure 9 the pattern of $dT_a/d\bar{T}_s$ seen in both the historical and amip-piForcing experiments demonstrates a marked warming over the tropical troposphere. Comparing Figure 9b and c it can be seen that this tropospheric $dT_a/d\bar{T}_s$ is stronger in amip-piForcing compared to the historical experiment. The amip-piForcing experiment also exhibits a stronger $dT_a/d\bar{T}_s$ over the southern hemisphere troposphere, whilst the historical experiment has a larger $dT_a/d\bar{T}_s$ signal over the northern hemisphere high latitudes. This is potentially due to the different T_s patterns seen in the historical and amip-piForcing experiments, with the subtropical $dT_s/d\bar{T}_s$ being slightly greater in the Northern Hemisphere in the historical ensemble and in the Southern Hemisphere in amip-piForcing (Figure 5f).

Since the tropical troposphere is a key region in causing variations in λ_{LWcs}, a region between 30°S – 30°N and between 100 – 500 hPa has been analysed further. A scatter plot of tropical tropospheric $dT_a/d\bar{T}_s$ against λ_{LWcs} is shown in Figure 10a. There it can be seen that a strong correlation between the two variables exists with a correlation coefficient of -0.8, consistent with physical expectations that a larger upper tropical tropospheric temperature results in a larger lapse rate feedback and a more negative λ_{LWcs} (T. Andrews & Webb, 2018). The amip-piForcing tropical tropospheric $dT_a/d\bar{T}_s$ and λ_{LWcs} has also been indicated in Figure 10a, where it can be seen that the tropical tropospheric $dT_a/d\bar{T}_s$ does well to capture why the feedbacks in historical and amip-piForcing experiments differ.

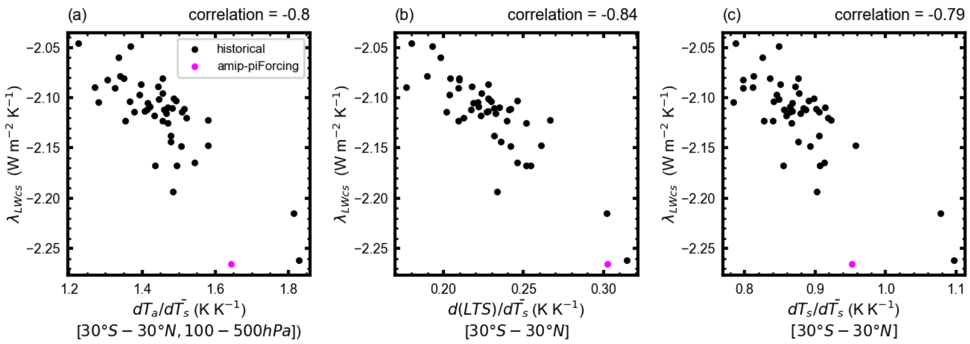

Figure 10. Scatter plots of (a) tropical tropospheric $dT_a/d\bar{T}_s$ against λ_{LWcs}, (b) tropical Lower Tropospheric Stability (LTS) change per degree of global warming ($d(LTS)/d\bar{T}_s$) against λ_{LWcs}, and (c) tropical $dT_s/d\bar{T}_s$ against λ_{LWcs}. Here the tropics have been characterised by averaging between 30°S and 30°N, and the tropical troposphere has used the same latitudinal bounds and averaged between 100–500 hPa (see red boxes in Figure 9). In each plot, black dots represent the historical ensemble and amip-piForcing values are represented by a magenta dot.

Since the spread in feedbacks can ultimately be derived from differing SST patterns, and given the strong relationship between tropical tropospheric temperature and λ_{LWcs}, the relationship between tropical mean $dT_s/d\bar{T}_s$ and λ_{LWcs} has been investigated (Figure 10c). Figure 10c follows a similar analysis to that performed by Soden and Held (2006). There, they demonstrated that across a range of models, due to the approximately adiabatic lapse rate of the tropical atmosphere, the strong coupling between the surface and free troposphere in the tropics, and the relatively weak coupling present over higher latitudes, the ratio between tropical and global warming was a good metric for determining the inter-model spread in lapse rate feedback. In Figure 10c it is shown that across the historical ensemble, the tropical $dT_s/d\bar{T}_s$ is well correlated with λ_{LWcs} with a correlation coefficient of -0.79. It is clear that ensemble members with a stronger warming over the tropics relative to the global mean also have a more strongly negative λ_{LWcs}.

As well as explaining the ensemble spread in λ_{LWcs}, tropical $dT_s/d\bar{T}_s$ changes can also be used to explain the disparity between amip-piForcing and historical experiments. Figure 10c shows that the amip-piForcing experiment has a strong $dT_s/d\bar{T}_s$ in the tropics and also has a strong negative λ_{LWcs}.

3.4 Processes Affecting Cloud Feedbacks (λ_{cre})

Although the historical ensemble used in this paper indicates that λ_{cre} is not the feedback with the largest spread (λ_{SWcs} has a standard deviation of 0.073 $\mathrm{Wm^{-2}K^{-1}}$ whilst λ_{cre} has a standard deviation of 0.06 $\mathrm{Wm^{-2}K^{-1}}$), for long term estimates of EffCS across different models, cloud feedbacks are the largest source of uncertainty and are the least understood (Forster et al., 2021; Ceppi & Nowack, 2021; Zelinka et al., 2016; Ceppi et al., 2017). Because of this, over recent years, cloud feedbacks have been the focus of many studies. Cloud controlling factor analyses such as Ceppi and Nowack (2021) and Blanco et al. (2023) aim to relate changes in clouds to other meteorological factors, such as free tropospheric humidity (van der Dussen et al., 2015), SSTs (Bretherton & Blossey, 2014), surface wind speed (Brueck et al., 2015) and inversion strength (Qu et al., 2015; Klein et al., 2017; Kawai et al., 2017). By better understanding what factors cause clouds to change, it is possible to understand differences in cloud feedbacks across models/ensembles.

In this section, λ_{cre} is investigated, primarily focusing on the spread across the historical experiment ensemble. Previously, Salvi et al. (2022) suggested that the different efficacies of well mixed greenhouse gases and aerosols were linked to changes in clouds due to differing changes in stability (although a large variability is seen across different models and a relatively small ensemble of 7 models was used). However here, the results shown in Figure 3 would suggest that for HadGEM3-GC3.1-LL, λ_{cre} does not contribute substantially to different forcing efficacies in the global mean. To understand the spatial distribution of λ_{cre}, Figure 4q is analysed. Here, strong positive cloud feedbacks are seen over the North Pacific and North Atlantic, and slightly weaker cloud feedbacks are seen over the Southern Indian Ocean and South Atlantic (each caused by positive shortwave cloud feedbacks - Figure S2). To understand the spread in λ_{cre}, maps of standard deviation in λ_{cre}, λ_{SWcre}, and λ_{LWcre} and standard deviation in $dT_s/d\bar{T_s}$ are shown in Figure 11. From Figure 11a it is possible to identify regions where the spread in λ_{cre} is largest and therefore which regions contribute most to the spread seen in Figure 3. The regions with the largest spread in λ_{cre} are the North Pacific and North Atlantic, due to a large spread in λ_{SWcre}. The Southern Ocean and low cloud deck regions off the east coast of South America, Australia and Southern Africa, also exhibit a moderately large standard deviation in λ_{cre}, again due to shortwave cloud feedbacks. The map of standard deviation of λ_{LWcre} shows a large spread in feedbacks over the tropical ascent regions, however as previously discussed, in these regions, longwave and shortwave responses to changes in cloud cancel, and therefore the standard deviation in net cloud feedbacks in these regions is generally small.

The spatial distribution of the standard deviation in $dT_s/d\bar{T_s}$ shown in Figure 11f is relatively similar to the pattern of standard deviation in λ_{cre}. Calculating the spatial correlation between Figures 11a and f, a correlation coefficient of 0.47 is found. Given surface temperatures are a key cloud controlling factor, as shown by Ceppi and Nowack (2021), we would expect the spread in λ_{cre} to be partly controlled by the spread in $dT_s/d\bar{T_s}$.

To better understand the cause of the spread in λ_{cre} shown in Figure 3b and 11a, two key cloud controlling factors are investigated; changes in T_s and changes in Lower Tropospheric Stability (LTS), both of which have strong statistical relationships with changes in clouds (Cutler et al., 2022; Klein & Hartmann, 1993; Ceppi & Nowack, 2021). Here LTS is defined as the 700hPa potential temperature minus the surface potential temperature (Cutler et al., 2022). Regarding the physical mechanisms of these relationships, LTS has been shown to influence cloud changes by controlling the amount of entrainment between the moist boundary layer and the drier free troposphere. The physical mechanism whereby surface temperatures effect cloud changes is less well established. Webb et al. (2024) investigate a range of possible mechanism relating surface temperatures to changes in cloud, such as the impact of surface latent heat flux changes, vertical gradients in humidity or moist static energy, or changes in downwelling longwave radiation caused by changing free tropospheric humidity. It was found that different mechanisms were plausible in some models and not in others. For HadGEM3-GC3.1-LL, only one suggested mechanism was not ruled out based on the models behaviour. This mechanism involved a reduction in low cloud due to a warming and a decrease in specific humidity due to an increase in upward longwave radiation from the surface (Ogura et al., 2023).

To relate changes in LTS and surface temperatures to changes in λ_{cre}, first two regions are investigated, the North West (NW) Pacific and North East (NE) Pacific (see Figure 11 boxes). These two regions were selected as being regions with a strong λ_{cre}

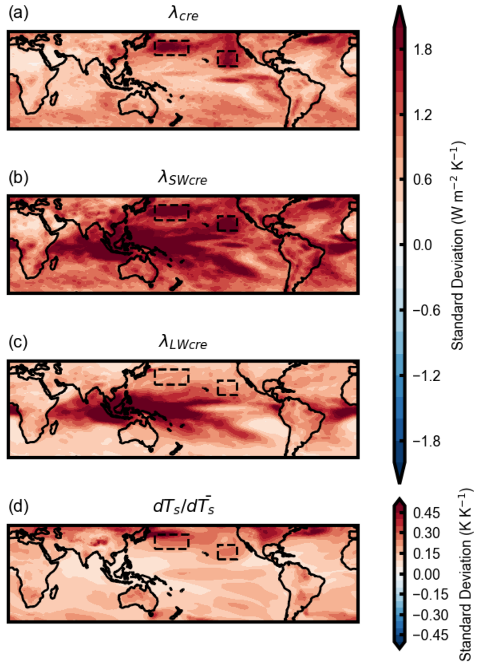

Figure 11. Maps of standard deviation in (a) λ_{cre}, (b) λ_{SWcre}, (c) λ_{LWcre}, and (d) $dT_s/d\bar{T}_s$ across the historical ensemble. Dashed black boxes indicate regions analysed in Figure 12 with the NW Pacific region extending from 150–185°E and 26–41°N, and the NE Pacific region extending from 215–235°E and 15–30°N.

signal (Figure 4q) and spread (Figure 11a). The two regions also capture different climatological regimes, with the NE Pacific a region of climatological subsidence where the surface is decoupled from the free troposphere due to a strong inversion, whereas the NW Pacific region is a region of climatological ascent where the surface is not decoupled from the free troposphere. Scatter plots of $d(LTS)/d\bar{T}_s$ and $dT_s/d\bar{T}_s$ against λ_{cre} over the NW Pacific and NE Pacific regions are shown in Figure 12a, b, c, and d. Here, it can be seen that in both the NE and NW Pacific there is a strong correlation between $dT_s/d\bar{T}_s$ and λ_{cre}, and $d(LTS)/d\bar{T}_s$ and λ_{cre}. This is consistent with Ceppi and Nowack (2021). Although the amip-piForcing and historical estimates of λ_{cre} were not particularly different, for completeness, amip-piForcing values have also been indicated in Figure 12. Here it can be seen that the amip-piForcing values fit the historical relationship between λ_{cre} and both $dT_s/d\bar{T}_s$ and $d(LTS)/d\bar{T}_s$ suggesting that any differences in λ_{cre} between historical and amip-piForcing experiments in these regions can be explained through these cloud controlling factors.

Since the LTS is defined as the 700hPa potential temperature minus the surface potential temperature, it is possible that the strong correlations between $d(LTS)/d\bar{T}_s$ and λ_{cre} exist primarily because of the strong relationship between λ_{cre} and $dT_s/d\bar{T}_s$. To investigate this, scatter plots of 700hPa $dT_a/d\bar{T}_s$ against λ_{cre} are shown in Figure 12e and f. Here, differing relationships between the two variables exist over the two regions analysed. Over the NW Pacific, a strong correlation remains with a correlation coefficient of 0.84. Over the NE Pacific however, this is not the case and a weak correlation

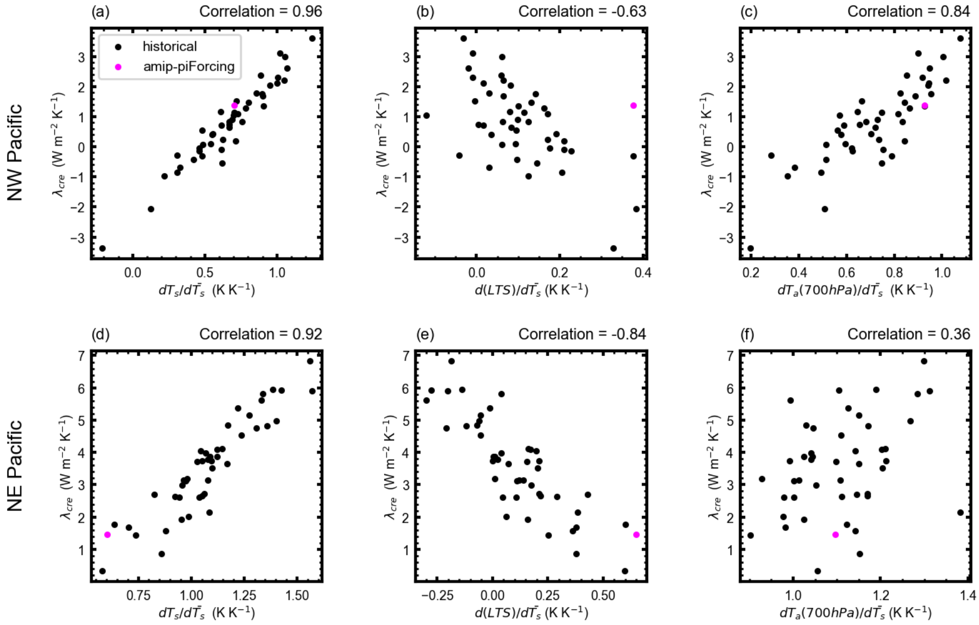

Figure 12. Scatter plots of (a and b) $dT_s/d\bar{T}_s$, (c and d) $d(LTS)/d\bar{T}_s$, and (e and f) 700hPa $dT_a/d\bar{T}_s$ against λ_{cre} over the (a, c, and e) NW Pacific region, and (b, d, and f) NE Pacific region. Black dots represent the historical ensemble and magenta markers indicate amip-piForcing values.

of 0.36 is seen. This differing relationship may be due to the different convective regimes that exist over the two regions. Over the NE Pacific, the strong inversion and the decoupling between the boundary layer and the free troposphere means that any surface warming in this region will be trapped under the strong inversion. Over the NW Pacific, this is not the case and surface warming can be transported efficiently into the free troposphere. Therefore, to some degree, over the NW Pacific the 700hPa temperature is still controlled by the temperatures at the surface.

An alternative approach is taken in Figure 13. Here, the local effect of surface warming and the remote effect of large scale stability changes on λ_{cre} is investigated using maps of the correlation across the historical ensemble between local λ_{cre} and either the local $dT_s/d\bar{T}_s$ or the 50°S – 50°N mean 700hPa $dT_a/d\bar{T}_s$. These latitudinal bounds were previously used by Ceppi and Gregory (2019) and Salvi et al. (2023) to capture large scale tropospheric stability.

From Figure 13 it can be seen that generally, the local $dT_s/d\bar{T}_s$ is the most strongly correlated, with many regions exhibiting correlations greater than 0.7. The correlations between λ_{cre} and the 50°S – 50°N mean 700hPa $dT_a/d\bar{T}_s$ tend to be weaker, although the subtropical cloud deck regions over the East Pacific and the Indian Ocean do exhibit positive correlations (note these are not statistically significant at the 95% confidence range). A decomposition of Figure 13 into shortwave and longwave components is shown in Figure S3. Here the strong correlations seen in the low cloud deck regions in Figure 13 are associated with the shortwave cloud feedbacks, and similar to Figure 11 and S2, the tropical ascent regions exhibit relatively strong correlations with both local $dT_s/d\bar{T}_s$ and 50°S – 50°N mean 700hPa $dT_a/d\bar{T}_s$ in the shortwave and longwave, however these

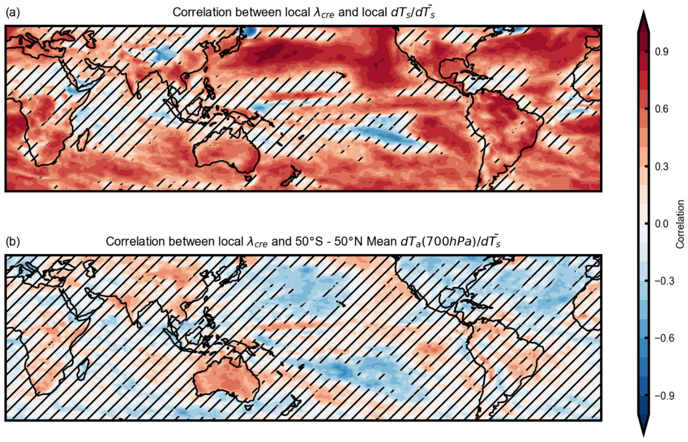

Figure 13. Maps of correlation between local λ_{cre} against (a) local $dT_s/d\bar{T_s}$, and (b) 50°S – 50°N mean 700hPa $dT_a/d\bar{T_s}$ across the historical ensemble. Hatching indicates where correlations are not significant at the 95% confidence interval (i.e. p values are greater than 0.05). Here the p value approximately indicates the probability of two random distributions producing a correlation coefficient at least as great as those indicated in the colored contours.

two components cancel, resulting in the net cloud feedback correlation being relatively weak in those regions in Figure 13.

To summarise, cloud feedbacks are the largest source of uncertainty in EffCS across models, however within the HadGEM3-GC3.1-LL historical ensemble, λ_{SWcs} contributes more to the spread in λ_{net}. Spread in λ_{cre} can be explained through the cloud controlling factors of T_s and LTS where $dT_s/d\bar{T_s}$ is positively correlated with λ_{cre} and $d(LTS)/d\bar{T_s}$ is negatively correlated with λ_{cre}. Finally, it is shown that the local influence of $dT_s/d\bar{T_s}$ on λ_{cre} is much stronger than the remote effect of changes in large scale atmospheric stability.

4 Conclusion

In this paper the feedbacks across a 47 member ensemble of historical and single forcing simulations have been analysed. Across the historical ensemble, EffCS ranges between 3–6K, highlighting the large spread in estimated feedbacks caused by internal variability. The aims of this work have been to understand the main causes of this spread in feedbacks across the ensemble, to understand if and why different forcing agents have different forcing efficacies, and finally to understand why the coupled historical simulations struggle to capture the feedbacks seen in AGCM simulations forced by observed SSTs. To address these aims, three components of λ_{net} were investigated (λ_{SWcs}, λ_{LWcs}, and λ_{cre}).

The analysis found that the ensemble spread in λ_{SWcs} is largely caused by varying degrees of sea ice melt per degree of global warming. Ensemble members that showed a large reduction in sea ice per degree of global warming also exhibited a strong λ_{SWcs}, with a correlation of -0.84 (consistent with Dessler (2020)). It was shown that this relationship was due to varying SST patterns, with ensemble members simulating stronger polar amplification also exhibiting more sea ice melt and a stronger λ_{SWcs} (with a correlation of 0.84 between polar SSTs and λ_{SWcs}). This relationship between λ_{SWcs}, sea ice melt, and polar amplification is also shown to be the reason for a much weaker λ_{SWcs} in the amip-piForcing experiment. Here, weaker polar amplification resulted in less sea ice melt per degree of global warming and a smaller λ_{SWcs}. Finally, the different λ_{SWcs} between the different single forcing experiments was investigated, since λ_{SWcs} was found to be the biggest source of differing forcing efficacies across the different forcing agents. It was shown that different patterns of surface warming were the main cause of different feedbacks across each experiment, with spatial correlations of 0.64 – 0.88 between patterns of T_s change per degree of global warming and λ_{SWcs} across all experiments and each hemisphere.

Previously, Salvi et al. (2022) also investigated different forcing efficacies between different forcing agents, also finding the hist-aer experiment to exhibit more strongly amplifying feedbacks compared to hist-ghg. There they focused on influence of stability changes on changes in cloud feedbacks, however here, we find that for HadGEM3-GC3.1-LL, changes in sea ice and polar T_s play a larger role in causing different forcing efficacies.

The ensemble spread in λ_{LWcs} was also investigated. Here it was shown that both tropical tropospheric temperature changes per degree of global warming and tropical T_s changes per degree of global warming were a key factor in causing the spread in λ_{LWcs}. Here, increased tropical surface warming caused warming in the tropical troposphere which has previously been shown to cause a stronger lapse rate feedback (T. Andrews & Webb, 2018). This relationship between λ_{LWcs} and tropical T_s also captures why the λ_{LWcs} is much stronger in the amip-piForcing experiment compared to the historical simulations, with the amip-piForcing experiment exhibiting a stronger tropical surface warming per degree of global warming compared to most historical ensemble members. Given that the amip-piForcing experiment is prescribed with observed SSTs, this shows how AOGCM biases in tropical SST patterns can impact on the estimated λ_{LWcs}.

The final feedback to be investigated was λ_{cre}. In contrast to the primary role of λ_{cre} in causing uncertainty in long term estimates of climate sensitivity, in the HadGEM3-GC3.1-LL historical ensemble, other feedbacks have a larger spread. Investigating λ_{cre}, it was shown that both T_s and LTS are key factors affecting changes in cloud feedbacks. It is also shown that although amip-piForcing and historical cloud feedbacks are not too dissimilar, both the LTS and T_s are useful metrics for understanding how amip-piForcing cloud feedbacks relate to those seen in the historical simulations. The analysis concludes by investigating the relative importance of local effect of varying T_s or the remote effect of large scale changes in atmospheric stability. Here it is shown that the local T_s is the most important, whilst the large scale stability plays a non-negligible role over the subtropical cloud deck regions.

This work provides useful insight into the different feedbacks seen across different forcing experiments and also provides information as to why coupled historical simulations struggle to capture the feedbacks seen in the amip-piForcing experiment. To take this work further, this large ensemble could be used to better understand the temporal evolution of feedbacks. In recent years, the amip-piForcing experiment demonstrates a marked decrease in λ_{net} (T. Andrews et al., 2022), and this ensemble could be used to investigate whether a similar behaviour is captured in any of the ensemble members. This work could then be used shed light on the causes and mechanisms involved in transient feedbacks.

5 Open Research

Data used in this analysis consists of HadGEM3-GC3.1-LL model simulations performed as part of the Met Office's contribution to CMIP6 (Eyring et al., 2016) and LESFMIP (D. Smith et al., 2022) and can be accessed from the ESGF CEDA data node https://esgf-index1.ceda.ac.uk/search/cmip6-ceda/.

Acknowledgments
All authors were supported by the Met Office Hadley Centre Climate Programme funded by DSIT. Doug Smith was also supported by the European Commission Horizon 2020 CONSTRAIN project.

References

Adams, B. K., & Dessler, A. E. (2019). Estimating transient climate response in a large-ensemble global climate model simulation. *Geophysical Research Letters*, *46*, 311-317. doi: 10.1029/2018GL080714

Andrews, M. B., Ridley, J. K., Wood, R. A., Andrews, T., Blockley, E. W., Booth, B., ... Sutton, R. T. (2020). Historical simulations with hadgem3-gc3.1 for cmip6. *Journal of Advances in Modeling Earth Systems*, *12*. doi: 10.1029/2019MS001995

Andrews, T., Andrews, M. B., Bodas-Salcedo, A., Jones, G. S., Kuhlbrodt, T., Manners, J., ... Tang, Y. (2019). Forcings, feedbacks and climate sensitivity in hadgem3-gc3.1 and ukesm1. *Journal of Advances in Modeling Earth Systems*, *11*, 4377–4394. doi: 10.1029/2019MS001866

Andrews, T., Bodas-Salcedo, A., Gregory, J. M., Dong, Y., adn D. Paynter, K. C. A., Lin, P., ... Liu, C. (2022). On the effect of historical sst patterns on radiative feedback. *Journal of Geophysical Research: Atmospheres*, *127*. doi: 10.1029/2022JD036675

Andrews, T., Gregory, J. M., Paynter, D., Silvers, L. G., Zhou, C., Mauritsen, T., ... Titchner, H. (2018). Accounting for changing temperature patterns increases historical estimates of climate sensitivity. *Geophysical Research Letters*, *45*, 8490-8499. doi: 10.1029/2018GL078887

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

Andrews, T., Gregory, J. M., Webb, M. J., & Taylor, K. E. (2012). Forcings, feedbacks and climate sensitivity in cmip5 coupled atmosphere-ocean climate models. *Geophysical Research Letters*, *39*. doi: 10.1029/2012GL051607

Andrews, T., & Webb, M. J. (2018). The dependence of global cloud and lapse rate feedbacks on the spatial structure of tropical pacific warming. *Journal of Climate*, *31*, 641–654. doi: 10.1175/JCLI-D-17-0087.1

Blanco, J. E., Caballero, R., Datseris, G., Stevens, B., Bony, S., Hadas, O., & Kaspi, Y. (2023). A cloud-controlling factor perspective on the hemispheric asymmetry of extratropical cloud albedo. *Journal of Climate*, *36*, 1793–1804. doi: 10.1175/JCLI-D-22-0410.1

Bretherton, C. S., & Blossey, P. N. (2014). Low cloud reduction in a greenhouse-warmed climate: Results from lagrangian les of a subtropical marine cloudiness transition. *Journal of Advances in Modeling Earth Systems*, *6*, 91–114. doi: 10.1002/2013MS000250

Brueck, M., Nuijens, L., & Stevens, B. (2015). On the seasonal and synoptic time-scale variability of the north atlantic trade wind region and its low-level clouds. *Journal of the Atmospheric Sciences*, *72*, 1428–1446. doi: 10.1175/JAS-D-14-0054.1

Ceppi, P., Brient, F., Zelinka, M. D., & Hartmann, D. L. (2017). Cloud feedback mechanisms and their representation in global climate models. *WIREs Climate Change*, *8*. doi: 10.1002/wcc.465

Ceppi, P., & Gregory, J. M. (2019). A refined model for the earth's global energy balance. *Climate Dynamics*, *53*, 4781–4797. doi: 10.1007/s00382-019-04825-x

Ceppi, P., & Nowack, P. (2021). Observational evidence that cloud feedback amplifies global warming. *PNAS*, *118*. doi: 10.1073/pnas.2026290118

Cutler, L., Brunke, M. A., & Zeng, X. (2022). Re-evaluation of low cloud amount relationships with lower-tropospheric stability and estimated inversion strength. *Geophysical Research Letters*, *49*. doi: 10.1029/2022GL098137

Dessler, A. E. (2020). Potential problems measuring climate sensitivity from the historical record. *Journal of Climate*, *33*, 2237–2248. doi: 10.1175/JCLI-D-19-0476.1

Dessler, A. E., Mauritsen, T., & Stevens, B. (2018). The influence of internal variability on earth's energy balance framework and implications for estimating climate sensitivity. *Atmospheric Chemistry and Physics*, *18*, 5147–5155. doi: 10.5194/acp-18-5147-2018

Dong, Y., Armour, K. C., Proistosescu, C., Andrews, T., Battisti, D. S., Forster, P. M., ... Shiogama, H. (2021). Biased estimates of equilibrium climate sensitivity and transient climate response derived from historical cmip6 simulations. *Geophysical Research Letters*, *48*. doi: 10.1029/2021GL095778

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, *9*, 1937–1958. doi: 10.5194/gmd-9-1937-2016

Forster, P. M., Richardson, T., Maycock, A. C., Smith, C. J., Samset, B. H., Myhre, G., ... Schulz, M. (2016). Recommendations for diagnosing effective radiative forcing from climate models for cmip6. *Journal of Geophysical Research: Atmospheres*, *121*, 12,460-12,475. doi: 10.1002/2016JD025320

Forster, P. M., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., ... co authors (2021). The earth's energy budget, climate feed- backs, and climate sensitivity. In *Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change* (p. 93). Cambridge, UK and New York, USA: Cambridge University Press.

Fueglistaler, S., & Silvers, L. G. (2021). The peculiar trajectory of global warming. *Journal of Geophysical Research: Atmospheres*, *126*. doi: 10.1029/2020JD033629

Gillett, N. P., Shiogama, H., Funke, B., Hegerl, G., Knutti, R., Matthes, K., ... Tebaldi, C. (2016). The detection and attribution model intercomparison project (damip v1.0) contribution to cmip6. *Geoscientific Model Development*, *9*, 3685–3697. doi: 10.5194/gmd-9-3685-2016

Gregory, J. M., & Andrews, T. (2016). Variation in climate sensitivity and feedback parameters during the historical period. *Geophysical Research Letters*, *43*, 3911-3920. doi: 10.1002/2016GL068406

Gregory, J. M., Andrews, T., Ceppi, P., Mauritsen, T., & Webb, M. J. (2020). How accurately can the climate sensitivity to co2 be estimated from historical climate change? *Climate Dynamics*, *54*, 129–157. doi: 10.1007/s00382-019-04991-y

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., ... Williams, K. D. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, *31*. doi: 10.1029/2003GL018747

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., ... Zhang, S. (2005). Efficacy of climate forcings. *Journal of Geophysical Re-*

search: Atmospheres, *110*. doi: 10.1029/2005JD005776

Haugstad, A. D., Armour, K. C., Battisti, D. S., & Rose, B. E. J. (2017). Relative roles of surface temperature and climate forcing patterns in the inconstancy of radiative feedbacks. *Geophysical Research Letters*, *44*, 7455-7463. doi: 10.1002/2017GL074372

Hwang, Y.-T., Xie, S.-P., Chen, P.-J., Tseng, H.-Y., & Deser, C. (2024). Contribution of anthropogenic aerosols to persistent la niña-like conditions in the early 21st century. *Proceedings of the National Academy of Sciences*, *121*. doi: 10.1073/pnas.2315124121

Kawai, H., Koshiro, T., & Webb, M. J. (2017). Interpretation of factors controlling low cloud cover and low cloud feedback using a unified predictive index. *Journal of Climate*, *30*, 9119—9131. doi: 10.1175/JCLI-D-16-0825.1

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., ... Vertenstein, M. (2015). The community earth system model (cesm) large ensemble project. *Bulletin of the American Meteorological Society*, 1333—1349. doi: 10.1175/BAMS-D-13-00255.1

Kay, J. E., Holland, M. M., & Jahn, A. (2011). Inter-annual to multi-decadal arctic sea ice extent trends in a warming world. *Geophysical Research Letters*, *38*. doi: 10.1029/2011GL048008

Klein, S. A., Hall, A., Norris, J. R., & Pincus, R. (2017). Low-cloud feedbacks from cloud-controlling factors: A review. *Surves in Geophysics*, *38*, 1307–1329. doi: 10.1007/s10712-017-9433-3

Klein, S. A., & Hartmann, D. L. (1993). The seasonal cycle of low stratiform clouds. *Journal of Climate*, *6*, 1587–1606. doi: 10.1175/1520-0442(1993)006⟨1587:TSCOLS⟩2.0.CO;2

Lambert, F. H., Webb, M. J., & Joshi, M. M. (2011). The relationship between land–ocean surface temperature contrast and radiative forcing. *Journal of Climate*, *24*, 3239 - 3256. doi: 10.1175/2011JCLI3893.1

Marvel, K., Schmidt, G. A., Miller, R. L., & Nazarenko, L. S. (2015). Implications for climate sensitivity from the response to individual forcings. *Nature Climate Change*, *6*, 386–389. doi: 10.1038/NCLIMATE2888

Ogura, T., Webb, M. J., & Lock, A. P. (2023). Positive low cloud feedback primarily caused by increasing longwave radiation from the sea surface in two versions of a climate model. *Geophysical Research Letters*, *50*. doi: 10.1029/2023GL104786

Pincus, R., Forster, P. M., & Stevens, B. (2016). The radiative forcing model intercomparison project (rfmip): experimental protocol for cmip6. *Geoscientific Model Development*, *9*, 3447–3460. doi: 10.5194/gmd-9-3447-2016

Po-Chedley, S., Armour, K. C., Bitz, C. M., Zelinka, M. D., Santer, B. D., & Fu, Q. (2018). Sources of intermodel spread in the lapse rate and water vapor feedbacks. *Journal of Climate*, *31*, 3187 - 3206. doi: 10.1175/JCLI-D-17-0674.1

Proistosescu, C., & Huybers, P. J. (2017). Slow climate mode reconciles historical and model-based estimates of climate sensitivity. *Science Advances*, *3*. doi: 10.1126/sciadv.1602821

Qu, X., Hall, A., Klein, S. A., & DeAngelis, A. M. (2015). Positive tropical marine low-cloud cover feedback inferred from cloud-controlling factors. *Geophysical Research Letters*, *42*, 7767–7775. doi: 10.1002/2015GL065627

Richardson, T. B., Forster, P. M., Smith, C. J., Maycock, A. C., Wood, T., Andrews, T., ... Watson-Parris, D. (2019). Efficacy of climate forcings in pdrmip models. *Journal of Geophysical Research: Atmospheres*, *124*, 12824-12844. doi: 10.1029/2019JD030581

Ridley, J. K., Blockley, E. W., & Jones, G. S. (2022). A change in climate state during a pre-industrial simulation of the cmip6 model hadgem3 driven by deep ocean drift. *Geophysical Research Letters*, *49*. doi: 10.1029/2021GL097171

Rugenstein, M., & Armour, K. C. (2021). Three flavours of radiative feedbacks and their implications for estimating equilibrium climate sensitivity. *Geophysical Research Letters*, *48*. doi: 10.1029/2021GL092983

Rugenstein, M., Bloch-Johnson, J., Abe-Ouchi, A., Andrews, T., Beyerle, U., Cao, L., … Yang, S. (2019). Longrunmip: Motivation and design for a large collection of millennial-length aogcm simulations. *Bulletin of the American Meteorological Society*, *100*, 2551–2570. doi: 10.1175/BAMS-D-19-0068.1

Salvi, P., Ceppi, P., & Gregory, J. M. (2022). Interpreting differences in radiative feedbacks from aerosols versus greenhouse gases. *Geophysical Research Letters*, *49*. doi: 10.1029/2022GL097766

Salvi, P., Gregory, J. M., & Ceppi, P. (2023). Time-evolving radiative feedbacks in the historical period. *Journal of Geophysical Research: Atmospheres*, *128*. doi: 10.1029/2023JD038984

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., … Zelinka, M. D. (2020). An assessment of earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, *58*(4). doi: 10.1029/2019RG000678

Smith, C. J., & Forster, P. M. (2021). Suppressed late-20th century warming in cmip6 models explained by forcing and feedbacks. *Geophysical Research Letters*, *48*. doi: 10.1029/2021GL094948

Smith, D., Booth, B., Dunstone, N., Eade, R., Hermanson, L., Jones, G. S., … Thompson, V. (2016). Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. *Nature Climate Change*, *6*, 936–940. doi: 10.1038/nclimate3058

Smith, D., Gillett, N. P., Simpson, I. R., Athanasiadis, P. J., Baehr, J., Bethke, I., … Ziehn, T. (2022). Attribution of multi-annual to decadal changes in the climate system: The large ensemble single forcing model intercomparison project (lesfmip). *Frontiers in Climate*. doi: 10.3389/fclim.2022.955414

Soden, B. J., & Held, I. M. (2006). An assessment of climate feedbacks in coupled ocean–atmosphere models. *Journal of Climate*, *19*, 3354–3360. doi: 10.1175/JCLI3799.1

Soden, B. J., Held, I. M., Colman, R., Shell, K. M., Kiehl, J. T., & Shields, C. A. (2008). Quantifying climate feedbacks using radiative kernels. *Journal of Climate*, *21*, 3504–3520. doi: 10.1175/2007JCLI2110.1

Stephens, G. L., Kahn, B. H., & Richardson, M. (2016). The super greenhouse effect in a changing climate. *Journal of Climate*, *29*, 5469–5482. doi: 10.1175/JCLI-D-15-0234.1

Titchner, H. A., & Rayner, N. A. (2014). The met office hadley centre sea ice and sea surface temperature data set, version 2: 1. sea ice concentrations. *Journal of Geophysical Research: Atmospheres*, *119*, 2864-2889. doi: 10.1002/2013JD020316

van der Dussen, J. J., de Roode, S. R., Gesso, S. D., & Siebesma, A. P. (2015). An les model study of the influence of the free tropospheric thermodynamic conditions on the stratocumulus response to a climate perturbation. *Journal of Advances in Modeling Earth Systems*, *7*, 670-691. doi: 10.1002/2014MS000380

Vial, J., Dufresne, J. L., & Bony, S. (2013). On the interpretation of inter-model spread in cmip5 climate sensitivity estimates. *Climate Dynamics*, *41*, 3339–3362. doi: 10.1007/s00382-013-1725-9

Webb, M. J., Lock, A. P., & Ogura, T. (2024). What are the main causes of positive subtropical low cloud feedbacks in climate models? *Journal of Advances in Modeling Earth Systems*, *16*. doi: 10.1029/2023MS003716

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., … 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*, 357—380. doi: 10.1002/2017MS001115

Wills, R. C. J., Dong, Y., Proistosecu, C., Armour, K. C., & Battisti, D. S. (2022).

Systematic climate model biases in the large-scale patterns of recent sea-surface temperature and sea-level pressure change. *Geophysical Research Letters*, *49*. doi: 10.1029/2022GL100011
Zelinka, M., Zhou, C., & Klein, S. A. (2016). Insights from a refined decomposition of cloud feedbacks. *Geophysical Research Letters*, *43*. doi: 10.1002/2016GL069917

Supporting information for

Feedbacks, Pattern Effects, and Efficacies in a Large Ensemble of HadGEM3-GC3.1-LL Historical Simulations

Harry Mutton[1], Timothy Andrews[1], Leon Hermanson[1], Melissa Seabrook[1], Doug M. Smith[1], Mark A. Ringer[1], Gareth S. Jones[1], Mark J. Webb[1]

[1]Met Office Hadley Centre, Exeter, UK

Contents of this file

- Figures S1 to S3

Introduction

The figures presented in this document show the decomposition of the cloud feedback analysis in Figures 3, 4, and 13 into longwave and shortwave components.

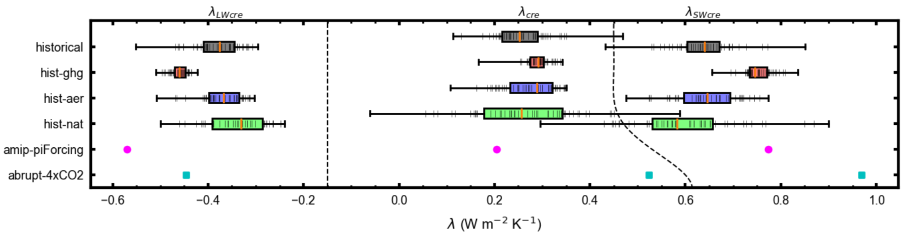

Figure S1. Boxplots of feedbacks in the historical and single forcing ensembles (1850–2014), amip-piForcing experiment (1870–2014), and abrupt-4xCO2 experiment (first 150 years). For each boxplot, the vertical black lines indicate each ensemble member, the whiskers indicate the maximum and minimum feedbacks seen in the ensemble, the boxes indicate the interquartile range, and the vertical orange line represents the median value.

Corresponding author: Harry Mutton, harry.mutton@metoffice.gov.uk

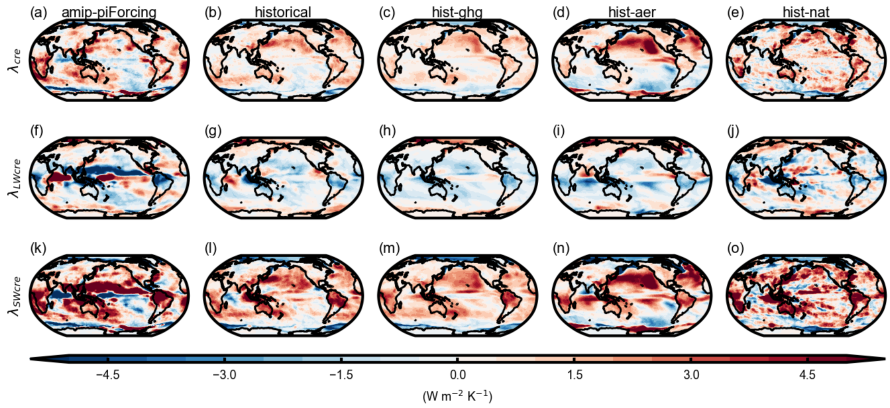

Figure S2. Maps of ensemble mean λ_{cre}, λ_{LWcre}, and λ_{SWcre} in amip-piForcing, historical, hist-ghg, hist-aer, and hist-nat experiments. Here, λ has been calculated by regressing the ensemble mean local annual mean $(N - F)$ against the ensemble mean global annual mean T_s timeseries between 1850 – 2014 for historical and single forcing experiments, and 1870 – 2014 for amip-piForcing.

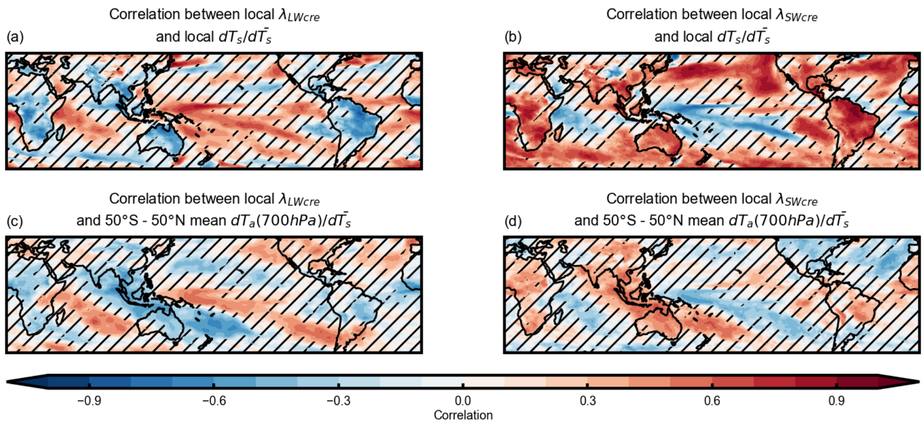

Figure S3. Maps of correlation between local λ_{LWcre} and λ_{SWcre} against local T_s changes per degree of global warming, and 50°S – 50°N mean 700hPa temperature change per degree of global warming across the historical ensemble.Hatching indicates where correlations are not significant at the 95% confidence interval (i.e. p values are greater than 0.05). Here the p value approximately indicates the probability of two random distributions producing a correlation coefficient at least as great as those indicated in the colored contours.