Using AMIP Simulations to Precondition Coupled Climate Model Behavior: Insights from an atmosphere-only and coupled paired parameter ensemble

TO BE WRITTEN IN AGU's *JAMES* FORMAT

Figures:
https://docs.google.com/presentation/d/13g4UrQulns7bqwxxYBshctHGmuH1k4nmIymFBxkxU_k/edit?usp=sharing

Abstract

1. Introduction

Uncertainty in climate projections remains a central challenge in climate science, largely due to variations in how models represent key physical processes and respond to external forcing (Hourdin et al., 2017; Knutti et al., 2017; Flynn and Mauritsen, 2020). With the emergence of systematic model development frameworks and perturbed parameter ensembles (PPEs), recent efforts have aimed to quantify the impact of parametric uncertainty on simulated climate responses. However, the extent to which atmospheric parameter sensitivities derived from atmosphere-only (AMIP) configurations are represented in coupled simulations remains poorly understood.

One strategy to constrain coupled model behavior has been to use AMIP simulations to precondition tuning. By isolating the atmospheric component from oceanic feedbacks, AMIP PPEs provide a means to test how atmospheric parameter perturbations influence radiation, circulation, and cloud processes without the masking effects of coupled surface adjustments (Vial et al., 2013; Webb et al., 2017; Duffy et al., 2023). These experiments have been widely used to calibrate model parameters, identify stable diagnostics, and explore emergent feedbacks (Mauritsen et al., 2012; Hourdin et al., 2017). This approach also allows partial separation of parametric uncertainty from structural uncertainty, which remains a critical challenge in interpreting ensemble diversity (Williamson et al., 2015; Sanderson et al., 2021; Peatier et al., 2024).

Manual calibration and hand-chosen PPE combinations remain the dominant approach in GCM development, despite growing evidence that it may inadequately sample model diversity in high-dimensional output spaces (Mauritsen et al., 2012; Sanderson et al., 2008). Recent studies have extended these frameworks using emulators, machine learning, and large PPEs to reduce the dimensionality of tuning and enhance physical interpretability. For instance, Peatier et al. (2022; 2024) examined the atmospheric component of CNRM-CM6-1 and revealed challenges in achieving robust tuning across parameter configurations. Eidhammer et al. (2024) introduced an extensible PPE for CAM6 to explore sensitivity in microphysics and aerosol schemes, while Mikkelsen et al. (2025) constrained aerosol–cloud adjustments using surface observations and ensemble output. In parallel, Elsaesser et al. (2025) developed a calibrated physics ensemble (CPE) for GISS ModelE using machine learning, and Yang et al. (2025) proposed a lightweight emulator to map parameter–output relationships.

These hybrid strategies offer alternatives to traditional manual, expert-driven tuning, emphasizing physical diagnostics within statistical frameworks. As shown in Williamson et al. (2013), and more recently by Williamson et al. (2017), Li et al., (2019), Couvreux et al. (2021), Hourdin et al. (2021, 2023), and Yamazaki et al. (2021), this integration of statistical learning and process-based evaluation allows tuning to become more reproducible, scalable, and transferable across AMIP and coupled models. This approach also reveals that structurally distinct coupled climate states can emerge from multiple valid parameter combinations or unresolved structural errors (Hourdin et al., 2023; Peatier et al., 2024).

A persistent question is which variables and diagnostics provide stable, transferable signals for tuning. Studies suggest that the upper troposphere, radiative feedbacks, and extratropical cloud structures offer more reliable targets than precipitation or surface diagnostics alone (Ceppi and Gregory, 2017; Mikkelsen et al., 2025). Parameters affecting TOA clear-sky flux and cloud radiative effect have emerged as key controls on temperature structure and cloud feedback strength in AMIP-based PPEs (Hourdin et al., 2017). However, it remains unclear how such parameter-induced changes interact with sea surface temperature (SST) feedbacks in coupled models and whether the same atmospheric adjustments are retained when running AMIP and coupled PPEs.

This study builds on these advances and addresses several unresolved issues in model calibration using PPEs. Specifically, we assess the limitations of using AMIP-based simulations to precondition coupled model tuning, extending the approach of Hourdin et al. (2023). The motivation of this work is threefold: First, to examine the calibration of AMIP and coupled models, and quantify the role of radiative adjustments and SST-feedbacks in masking, amplifying, or aligning ensemble diversity across configurations. Second, to assess the extent to which AMIP and coupled PPEs exhibit similar ensemble spread across key atmospheric variables, both horizontally and vertically. Finally, to identify the parameters that control ensemble spread in each configuration and evaluate whether these key sensitivities are consistent across AMIP and coupled setups.

2. Methods

2.1 Atmospheric and Coupled Model Descriptions

The tuning exercise is performed using the latest development version of the IPSL coupled model, which is structurally similar to IPSL-CM6A-LR (Boucher et al., 2020), the configuration submitted by IPSL for CMIP6. The coupled model consists of the LMDZ6 atmospheric component (Hourdin et al., 2020), the ORCHIDEE land surface model (Krinner et al., 2005), and the NEMO version 4 ocean model (Madec et al., 2019), coupled via the SI3 sea ice model (Vancoppenolle et al., 2023). The atmospheric resolution is 144 × 143 points in latitude and longitude, with 79 vertical layers extending up to approximately 80 km. The ORCA1 ocean configuration is on a quasi-isotropic global tripolar grid; 1° nominal resolution and increases latitudinal resolution to 1/3° in the equatorial region. Vertically, layer thickness varies from 1 m at the surface, 10 m at 100 m depth, and 200 m in the bottom layers (75 vertical levels).

The design of IPSL-CM6A-LR followed a long three-year sequence of improvements, bug fixes, and expert-driven manual tuning phases, largely focused on improving the sea ice extent and global mean surface temperature (Mignot et al., 2021). In order to avoid repeating this labor-intensive tuning process for the coupled model, and building on the results of Hourdin et al. (2023), this study explores the use of an AMIP-based PPE to precondition the coupled configuration. This approach is intended to reduce both computational cost and human effort while identifying viable parameter configurations for the fully coupled system.

2.2 Perturbed Parameter Ensemble Experiment

As in Hourdin et al. (2023), we apply the Hightune Explorer tool (Couvreux et al., 2021) to facilitate interactive exploration and filtering of PPE. We use the same radiative and precipitation-based diagnostics defined in Hourdin et al. (2023) to visualize ensemble spread and guide history matching across successive tuning waves. While the model configuration is the full atmospheric component of the IPSL coupled model, the tuning begins with forty waves of 1D global preconditioning, followed by five iterations of 3D AMIP simulations.

Unlike Hourdin et al. (2023), this study extends the approach by applying the complete parameter set to both AMIP and coupled configurations, using a shared PPE of 120 members. The list of twelve perturbed parameters and the initial acceptable ranges are provided in Table S1. Eight of which parameters are from within the AMIP component, while two parameters are each from ocean (NEMO4) and sea ice (SI3) components.

To isolate the effect of SST-driven feedbacks on ensemble spread, we construct a "coupled-$SST_{removed}$" ensemble by removing the global mean SST variance via linear regression (described in Section 3.1). Following standard approaches in emulator design and ensemble analysis, we apply principal component analysis (PC) and empirical orthogonal functions (EOFs) to reduce output dimensionality while preserving dominant variance patterns (Wilkinson, 2010; Salter et al., 2019). These methods allow a direct comparison between AMIP and coupled ensembles and helps identify which atmospheric structures, cloud feedbacks, and parameter sensitivities and PPE spread persist under coupling, and which differ with ocean feedbacks.

3. Results

3.1. AMIP and Coupled PPEs and Isolating SST-Feedbacks

In the experimental design, the AMIP and coupled perturbed parameter ensembles (PPEs) are run using the same 120 parametric configurations. This design enables a direct comparison between the atmospheric diversity in the AMIP ensemble and that of the coupled system. he overarching goal of this study is to investigate how AMIP-based preconditioning relates to coupled model behavior, and how SST feedbacks influence this relationship. We will thus first compare the atmospheric adjustment to the parameter variations in the AMIP and coupled PPE.

We begin by examining the top-of-atmosphere (TOA) energy imbalance across the 120-member ensembles. In the AMIP configuration, which uses prescribed SSTs, the ensemble exhibits a substantial net TOA radiative imbalance due to its inability to equilibrate energy through ocean surface warming (Fig. 1a). This imbalance ranges from 0 to 5 W m^{-2} (taken from year 2 of the simulations) across ensemble members In coupled mode, in contrast, the oceanic surface temperature adapts so as to quasi-equilibrate possible TOA imbalance. Global mean SST ranges from 18.6 to 21.5 °C between years 11-20 of the coupled PPE. Comparing the AMIP net TOA imbalance to coupled SSTs yields a tight, positive relationship (Fig. 1a). This pattern reveals a statistically significant linear relationship, with a slope of approximately 1.6 ± 0.1 W m^{-2} $°C^{-1}$ and an R^2 of 0.66, suggesting that the AMIP TOA radiative fluxes can be used to anticipate the global mean SST of the associated coupled model. Interestingly, the relationship between net TOA radiative budget and the SST in the coupled model is not significant when considering the AMIP TOA imbalance, reaching 0.4 ± 0.1 W m^{-2} $°C^{-1}$ and an R^2 of 0.21. Comparing the total net atmospheric adjustment, estimated as the TOA minus surface radiative fluxes, and the relationship between coupled SST, shows the relatively weak heat storage in the atmosphere of the AMIP PPE and that the AMIP-derived imbalance is not statistically significant. These results underscore the relevance of net TOA flux alone in capturing the parametric sensitivity that projects onto coupled SST responses and confirms that

AMIP radiative forcing can be used to anticipate the parametric spread in global SST responses in the coupled ensemble.

To evaluate how this surface warming feeds back on the atmospheric thermal structure in the coupled ensemble, the time-mean SST is regressed onto the zonally averaged atmospheric temperature across the coupled PPE (Figure S1). The resulting pattern shows vertically stratified anomalous warming throughout the troposphere, particularly over the tropics, and a tripolar anomalous cooling pattern in the stratosphere for increased SSTs. This vertically stratified warming and tripolar stratospheric cooling is consistent with past analyses of stratosphere–troposphere coupling in response to radiative forcing changes (Thompson and Solomon, 2005). This structure is also consistent with the inverse thermal relationship described by Lin and Emanuel (2024a), who attribute such patterns to balanced energy adjustment across layers during surface-driven warming. Lin & Emanuel (2024b) further show that tropospheric thermal anomalies force stratospheric responses through quasi-balanced dynamics, which may explain the tripolar vertical pattern seen in the ensemble SST–temperature regressions.

The majority of the PPE tropospheric temperature variance is explained by SST, with R^2 values exceeding 0.7 across most vertical levels, and approaching 1.0 in some layers. Comparatively, the stratospheric signal is notably weaker, with R^2 values up to 0.5 in the mid-latitudes near 150 hPa. These results indicate that a substantial portion of tropospheric atmospheric ensemble spread in the coupled configuration is attributable to ocean-mediated adjustments. Conversely, the upper troposphere and lower stratosphere indicates a significant regression with SST across the PPE, demonstrating that upper-atmosphere PPE spread can result directly from atmospheric parametrizations or indirectly through SST-adjustments in the coupled simulations. These localized relationships suggest that the upper troposphere could be a good region to identify metrics for the AMIP to precondition the coupled model.

Consistent with this pattern of variance explained, the EOF pattern computed across the coupled ensemble shows an anomalous warming in the upper troposphere accompanied by a strong low-level tropical cooling deeply penetration vertically within the tropics (Fig. 2b), consistent with SST-driven adjustment of the troposphere amplification (Fig. S1). This mode explains 89% of the variance. In the AMIP ensemble, the dominant mode of temperature variations across the various configurations explains ~23% of the variance, and its pattern reflects a strong upper-tropospheric and lower-stratospheric warming, paired with modest cooling in the lower atmosphere (Fig. 2a). While this mode is necessarily only driven by the changes in the parametrization, the coupled pattern is strongly constrained by the SST adjustment. In order to isolate the atmospheric ensemble-spread solely due to the parametric changes from the SST-driven signal, we construct a “coupled-$SST_{removed}$” ensemble by subtracting the SST-driven thermal structure identified in Fig. S1 prior to computing the EOF. The resulting pattern closely resembles that of the AMIP, both in vertical structure and spatial amplitude (Fig. 2c). The EOF1 pattern of the coupled-$SST_{removed}$ explains ~51% of the PPE variance.

This visual alignment between the AMIP and coupled-$SST_{removed}$ is quantitatively confirmed by comparing the associated principal component time series (PC1). While there is little linear correlation between AMIP and raw coupled PC1 values ($R^2 = 0.13$), low explained variance and non- 1:1 relationship, the PC1 values in coupled-$SST_{removed}$ are highly correlated with those from AMIP ($R^2 = 0.76$), and lie close to a 1:1 line (Fig. 2d). This confirms that beyond a direct thermal adjustment, the vertical atmospheric thermal structure in the coupled PPE reflects a direct parametric adjustment that is very similar to the adjustment obtained in AMIP mode. This alignment also mirrors past findings that show how AMIP configurations can isolate radiative feedback responses, even as coupled models introduce additional SST-related spread (Siler et al., 2018). Taken together, these results validate the use of the coupled-$SST_{removed}$ framework as a tool for cleanly separating feedback-driven diversity from intrinsic atmospheric responses in a PPE. This enables meaningful comparisons between AMIP-based diagnostics and fully coupled simulations, and provides a foundation for identifying which components of coupled model behavior can be preconditioned through atmospheric tuning alone.

To understand the physical origin of ensemble spread across configurations, we analyze how anomalous cloudiness is associated with PC1 across the PPEs. Focusing first on globally averaged low-level cloudiness, the raw coupled ensemble shows a significant positive relationship with PC1 (R^2 = 0.46), while the AMIP (R^2 = 0.09) and coupled-$SST_{removed}$ (R^2 = 0.22) configurations exhibit weaker, negative regressions. In contrast, PC1 in both AMIP (R^2 = 0.81) and coupled-$SST_{removed}$ (R^2 = 0.97) shows a strong, negative relationship with high-level cloud fraction (Fig. 3b). The raw coupled ensemble exhibits only a weak negative relationship with high clouds (R^2 = 0.11), suggesting reduced sensitivity to upper-level cloud changes. The distinct vertical structure in AMIP versus coupled-SSTremoved clouds is in line with previous work identifying separable cloud feedback regimes within ensemble frameworks (Zelinka et al., 2020).

These differences are further supported by seasonal latitude-height regressions of zonally averaged cloud fraction against PC1 (Fig. S2). In AMIP and coupled-$SST_{removed}$, PC1 is associated with significant reductions in high clouds above 400 hPa, particularly in mid-to-high latitudes, with vertically coherent anomalies extending to the surface. In the coupled configuration, PC1 correlates with increased low-level cloudiness below 850 hPa, most notably in the tropics and midlatitudes. This pattern extends upward to 200 hPa in the tropical regions. Despite this contrast, the coupled ensemble also exhibits a negative high-cloud relationship in polar regions, similar to the AMIP and coupled-$SST_{removed}$ patterns.

These results suggest that AMIP and coupled-$SST_{removed}$ members with stronger positive PC1 signals are characterized by reduced upper cloud cover, contributing to enhanced longwave radiative cooling through cloud cover. This high-cloud signal is seasonally consistent and most pronounced during boreal winter and austral summer (Fig. S2b,e), seasons in which cloud-radiative feedbacks have a substantial influence on polar energy balance. In contrast, the raw coupled ensemble behaves differently: PC1 is positively correlated with tropical and subtropical low-level clouds, highlighting the role of SST-driven warming in enhancing shallow convective cloudiness. Nevertheless, the negative high-cloud–PC1 relationship in polar regions persists in all configurations. This divergence between cloud regimes emphasizes the distinction between intrinsic atmospheric diversity and the feedback-driven diversity introduced through coupling with the ocean.

3.2. The Role of Parameters and Atmospheric Clouds Driving Ensemble Spread

To identify the parameters responsible for driving this spread (see Table S1), we compute Pearson correlation coefficients between PC1 and the 12 tunable model parameters (Fig. 4). In both AMIP and coupled-$SST_{removed}$, the parameters *FALLV* (falling velocity of ice crystals) and *OMEPMX* (ice-to-rain conversion threshold) exhibit the strongest positive correlations with PC1 ($r > 0.6$), identifying them as key modulators of upper-level cloud structure and associated radiative responses. The relevance of *FALLV* and *OMEPMX* in modulating upper-level cloud structure is consistent with prior findings on cloud-ice microphysical schemes to PPE climate diversity (Zelinka et al., 2013; Gettelman et al., 2015; Hourdin et al., 2017). In the raw coupled ensemble, *CLC* (auto-conversion threshold for liquid cloud water) and *RNLC* (non-shear-driven Langmuir vertical mixing) emerge as dominant drivers of PC1 diversity. These parameters are associated with tropical atmospheric boundary-layer and ocean mixing processes, consistent with the enhanced role of low clouds and SST-feedbacks in shaping coupled diversity. Notably, the coupled configuration also retains a significant correlation with *FALLV*. These findings are robust across alternative ranking methods, including SHAP, Sobol, and Spearman analyses.

To further examine the role of *FALLV*, we regress its values onto TOA net cloud radiative effect (CRE) and 2-meter surface air temperature (SAT) across the PPE (Fig. S3). Increasing *FALLV* reduces CRE and cools the surface, particularly in polar regions during winter, highlighting a physically consistent pathway by which upper-level cloud changes affect polar radiative balance. The effect is consistent across both ensembles. This supports the findings in Fig. 3b and underscores how an AMIP-calibrated parameter like *FALLV* can help precondition polar CRE and surface temperature behavior in coupled simulations through upper-level cloud cover.

To assess how well an AMIP simulation can precondition the behavior of the coupled model, we examine how the diversity of temperature structures of the EOF in the AMIP ensemble project onto surface climate variables compared to the coupled PPE. We focus on the coupled-$SST_{removed}$ data to minimize SST-masked signals. Regressions of PC1 onto TOA CRE reveal strong polar signatures across all three configurations during winter (Fig. 5a–f). These polar CRE signals align with previous studies emphasizing their critical role in regulating the polar energy budget (Crook et al., 2011; Taylor et al., 2013; Payne et al., 2015; Stuecker et al., 2018). In both AMIP and coupled-$SST_{removed}$, PC1 is associated with decreased CRE (i.e., enhanced outgoing longwave radiation) at high latitudes, consistent with reduced upper-level cloud cover and radiative feedbacks, likely influenced by *FALLV*. This polar signal is also present in the raw coupled ensemble during winter but diverges in summer, particularly in the tropics and midlatitudes, where coupled responses differ from AMIP and $SST_{removed}$.

CRE anomalies are reflected in surface temperature. As shown in Fig. 5g–l, PC1 regressions onto SAT reveal strong cooling signals in the Southern Ocean and Arctic across all configurations. Spatial agreement is especially pronounced between AMIP and coupled-$SST_{removed}$, indicating a shared polar radiative mechanism. The raw coupled ensemble also displays a strong negative SAT response in the polar regions but shows a broader signal in the midlatitudes and tropics, suggesting an added influence of low-level clouds and SST feedbacks. These results demonstrate that polar cloud and temperature spread are strongly linked in all PPEs, and that AMIP simulations effectively capture the radiative–thermal coupling in polar regions, particularly when SST-feedbacks are removed. However, divergence in the tropics in the raw coupled ensemble points to additional variance driven by SST-coupled processes not captured by AMIP. Such divergence is consistent with results showing that SST pattern feedbacks significantly influence the expression and spatial structure of climate sensitivity (Proistosescu et al., 2018; Dong et al., 2020).

Figure S4 further explores precipitation responses to *FALLV*. While AMIP precipitation patterns are noisy, the coupled and coupled-$SST_{removed}$ ensembles exhibit more spatially coherent relationships, with similar patterns but differing in magnitude. Figure 6 extends this analysis of PC1 to precipitation and zonal surface wind stress. The relationships between the PC1 and precipitation similarly emphasize the differences between the location and magnitude of precipitation ensemble spread in the AMIP and coupled models. Wind stress regressions show moderate alignment between AMIP and coupled-$SST_{removed}$, particularly in the Southern Hemisphere midlatitudes. Zonally averaged regressions of zonal wind velocity (Fig. S5) support these conclusions: both AMIP and coupled-$SST_{removed}$ show strong polar jet responses associated with PC1, especially in the Southern Hemisphere, stretching from the upper-troposphere to the surface. In the Northern Hemisphere, the raw coupled ensemble shows a damped response, likely due to broader meridional SST gradients and feedbacks.

Still, the vertical structure of the jet anomalies is consistent across all configurations, suggesting that zonal winds, both surface and upper-level, can be reliably preconditioned using AMIP simulations, at least in the extra-tropics. The structural robustness of the jet anomalies across configurations is consistent with known persistence in Rossby wave train dynamics and midlatitude jets across model ensembles (Barnes and Hartmann, 2012). However, precipitation anomalies exhibit weaker spatial coherence and diverge in magnitude across configurations. These inconsistencies reflect the masking influence of SST-feedbacks on precipitation, particularly in the tropics, and highlight the potential challenges of using high-frequency metrics like precipitation as reliable tuning targets in AMIP-coupled calibration frameworks. This tuning breakdown is expected in regions where climate feedbacks interact with internal atmospheric signals, as highlighted by Ceppi and Shepherd (2017).

Finally, we assess whether AMIP-based PC1 diversity projects onto sea ice concentration and area. Here, only the raw coupled and coupled-$SST_{removed}$ ensembles are compared. Regressions of PC1 onto sea ice concentration (Fig. 7a–c, e–g) reveal robust seasonal and regional signals, particularly in the Weddell Sea, Ross Sea, and Nordic Seas, in both hemispheres. These spatial patterns and amplitudes are consistent between the raw and $SST_{removed}$ ensembles, indicating that upper-level cloud and CRE processes originating in AMIP can influence the cryospheric state in the coupled model (see also Cesena et al., 2025, on Antarctic cloud-ice interactions).

4. Implications for Climate Model Tuning & Conclusion

This study assesses whether AMIP simulations can effectively capture and precondition the diversity found in coupled climate simulations using a PPE constructed with identical atmospheric parameter configurations. By comparing AMIP and coupled simulations through TOA radiative imbalance, cloud structure, parameter sensitivity, and surface climate diversity, we show that AMIP simulations retain a significant portion of the coupled model's atmospheric ensemble diversity, particularly in the upper troposphere and polar regions, yet primarily masked by the global thermal adjustment of the coupled configurations. The clear expression of extratropical jet responses across AMIP and coupled-SST$_{removed}$ ensembles reflects underlying stratosphere–troposphere coupling processes, recently shown to be amplified through tropospheric pathways (Baldwin et al., 2024). This reinforces broader concerns that tuning efforts must be grounded in physical understanding to avoid compensating errors, as discussed by Sherwood et al. (2020).

To address the implications for climate model tuning strategies and advancements from previous experiments (Hourdin et al., 2023), this study highlights potential additional metrics that can be used to connect AMIP and coupled model behavior and reveal climate sensitivities to particular parameters. Specifically, parameters related to upper-level cloud microphysics, such as ice autoconversion and cloud lifetime, emerged as consistent drivers of ensemble spread across both AMIP and coupled configurations. These results support the selection of upper-tropospheric diagnostics as robust tuning metrics and motivate the introduction of a new AMIP-based preconditioning metric: Northern Hemisphere midlatitude temperature at 150 hPa. This variable shows a near 1:1 relationship between AMIP and coupled simulations (Fig. 8o), comparable in fidelity to TOA-based metrics (Fig. 8a–k), and is minimally affected by feedbacks. In contrast, precipitation-based diagnostics show poor AMIP–coupled alignment (Fig. 8l–n), highlighting limitations in their use for tuning calibration (Ceppi and Gregory, 2017; Mikkelsen et al., 2025).

This limitation is consistent with recent findings from Cesena et al. (2025), who showed that cloud processes, especially those involving high clouds, are stronger predictors of high-latitude diversity than precipitation-based metrics. Their work also highlights the importance of upper-tropospheric radiative feedbacks in shaping Antarctic sea ice extent, complementing our finding that AMIP-derived cloud and CRE patterns project onto cryospheric behavior in coupled simulations.

Indeed, sea ice concentration regressions show that AMIP ensemble structure projects onto regional sea ice patterns in both hemispheres, particularly in the Weddell, Ross, and Nordic Seas (Fig. 7). These results validate the ability of AMIP simulations to precondition the coupled system's polar climate, even in the presence of ocean feedbacks. This is further supported by recent results from Studholme et al. (2025), who link lower stratospheric temperature biases to polar sea ice extent and air-sea coupling strength, highlighting the relevance of the upper-tropospheric and lower-stratospheric layers as effective tuning targets.

To conclude, these findings extend the framework proposed by Hourdin et al. (2023). By expanding this framework to vertically and spatially resolved diagnostics, our results provide a pathway toward more effective, high-confidence tuning strategies. Specifically, this work shows how targeted tuning using AMIP-based diagnostics can isolate parameter impacts that persist into coupled configurations, offering a basis for reproducible, cross-configuration calibration approaches. Together, this work demonstrates that AMIP simulations retain a significant portion of the parametric ensemble diversity found in coupled simulations, particularly for upper-tropospheric, radiative, and high-latitude climate features. The AMIP–coupled alignment for selected diagnostics, combined with tools like hightune-explore, offers a path toward more systematic, transferable model development ahead of CMIP7.

References (NEED TO ORDER AND FINISH ADDING)

Hourdin, F, B. S. Ferster, J. Mignot, J. Deshayes, I. Musat (2023). Towards a machine-assisted tuning that avoids underestimating the uncertainty of climate change projections, Science Advances, https://doi.org/10.1126/sciadv.adf2758.

Thompson, D. W. J., & Solomon, S. (2005). *Recent stratospheric climate trends as evidenced in radiosonde data and coupled models.* Science, 309(5741), 1151–1154. https://doi.org/10.1175/JCLI3585.1

Lin, J., & Emanuel, K. (2024a). *Why the lower stratosphere cools when the troposphere warms.* PNAS, 121. https://doi.org/10.1073/pnas.2319228121

Lin, J., & Emanuel, K. (2024b). *Tropospheric Thermal Forcing of the Stratosphere through Quasi-Balanced Dynamics.* J Atmos Sci 81, 561–582. https://doi.org/10.1175/JAS-D-23-0081.1

Zelinka, M. D., et al. (2020). *Causes of higher climate sensitivity in CMIP6 models.* Geophys. Res. Lett., 47(1), e2019GL085782. https://doi.org/10.1029/2019GL085782

Gettelman, A., Morrison, H., Santos, S., Bogenschutz, P. and Caldwell, P.M., 2015. Advanced two-moment bulk microphysics for global models. Part II: Global model solutions and aerosol–cloud interactions. *Journal of Climate*, *28*(3), pp.1288-1307.

F. Hourdin, T. Mauritsen, A. Gettelman, J.-C. Golaz, V. Balaji, Q. Duan, D. Folini, D. Ji, D. Klocke, Y. Qian, F. Rauser, C. Rio, L. Tomassini, M. Watanabe, D. Williamson, The art and science of climate model tuning. Bull. Am. Meteorol. Soc. 98, 589–602 (2017).

Stuecker, M.F., Bitz, C.M., Armour, K.C., Proistosescu, C., Kang, S.M., Xie, S.P., Kim, D., McGregor, S., Zhang, W., Zhao, S. and Cai, W., 2018. Polar amplification dominated by local forcing and feedbacks. Nature Climate Change, 8(12), pp.1076-1081.

Crook, J.A., Forster, P.M. and Stuber, N., 2011. Spatial patterns of modeled climate feedback and contributions to temperature response and polar amplification. Journal of Climate, 24(14), pp.3575-3592.

Taylor, P.C., Cai, M., Hu, A., Meehl, J., Washington, W. and Zhang, G.J., 2013. A decomposition of feedback contributions to polar warming amplification. Journal of Climate, 26(18), pp.7023-7043.

Payne, A.E., Jansen, M.F. and Cronin, T.W., 2015. Conceptual model analysis of the influence of temperature feedbacks on polar amplification. Geophysical Research Letters, 42(21), pp.9561-9570.

Barnes, E. A., & Hartmann, D. L. (2012). *Detection of Rossby wave trains and the jet stream position from climate model data.* Clim. Dyn., 38(11–12), 2257–2273. https://doi.org/10.1029/2012JD017469

Ceppi, P., & Shepherd, T. G. (2017). *Contributions of climate feedbacks to changes in atmospheric circulation.* J. Climate, 30(23), 9097–9118. https://doi.org/10.1175/JCLI-D-17-0189.1

Sherwood, S. C., et al. (2020). *An assessment of Earth's climate sensitivity using multiple lines of evidence.* Rev. Geophys., 58(4), e2019RG000678. https://doi.org/10.1029/2019RG000678

Baldwin, M. P., Birner, T., & Ayarzagüena, B. (2024). *Tropospheric amplification of stratosphere–troposphere coupling.* QJRMS. https://doi.org/10.1002/qj.4864

Siler, N., Po-Chedley, S. and Bretherton, C.S., 2018. Variability in modeled cloud feedback tied to differences in the climatological spatial pattern of clouds. Climate Dynamics, 50(3), pp.1209-1220.

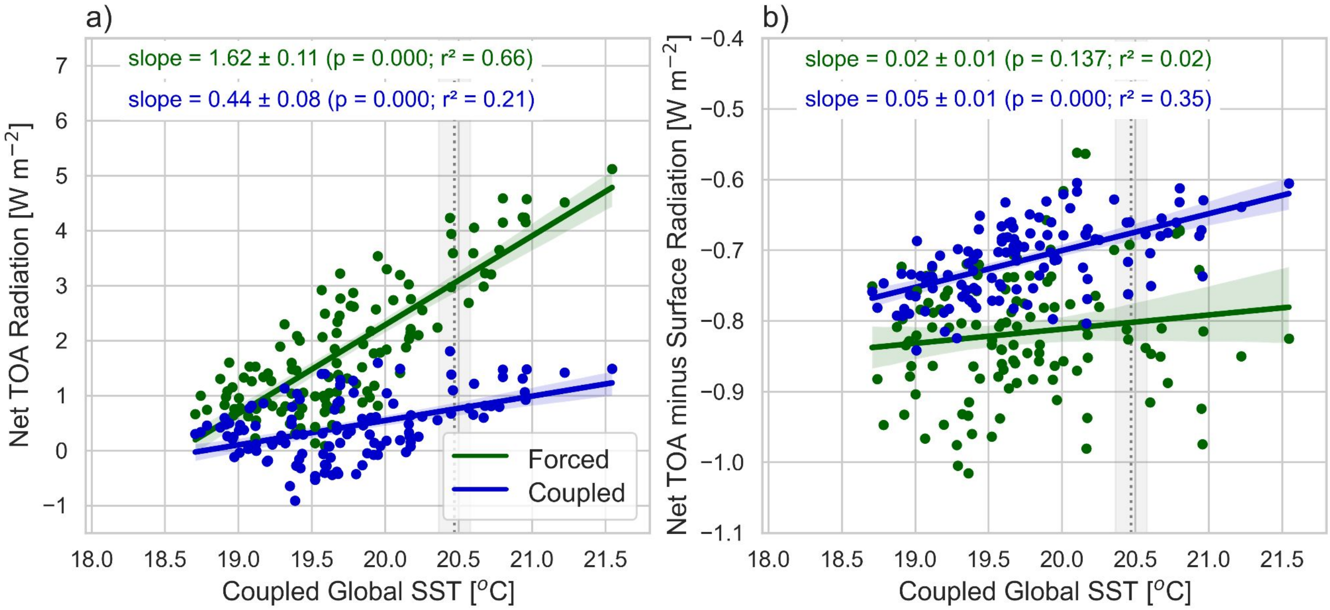
Dong, Y., Armour, K.C., Zelinka, M.D., Proistosescu, C., Battisti, D.S., Zhou, C. and Andrews, T., 2020. Intermodel spread in the pattern effect and its contribution to climate sensitivity in CMIP5 and CMIP6 models. Journal of Climate, 33(18), pp.7755-7775.

Proistosescu, C., Donohoe, A., Armour, K.C., Roe, G.H., Stuecker, M.F. and Bitz, C.M., 2018. Radiative feedbacks from stochastic variability in surface temperature and radiative imbalance. Geophysical Research Letters, 45(10), pp.5082-5094.

Madec Gurvan, Romain Bourdallé-Badie, Jérôme Chanut, Emanuela Clementi, Andrew Coward, Christian Ethé, Doroteaciro Iovino, Dan Lea, Claire Lévy, Tomas Lovato, Nicolas Martin, Sébastien Masson, Silvia Mocavero, Clément Rousset, Dave Storkey, Martin Vancoppenolle, Simon Müeller, George Nurser, Mike Bell, & Guillaume Samson. (2019). NEMO ocean engine. In Notes du Pôle de modélisation de l'Institut Pierre-Simon Laplace (IPSL) (v4.0, Number 27). Zenodo. https://doi.org/10.5281/zenodo.3878122

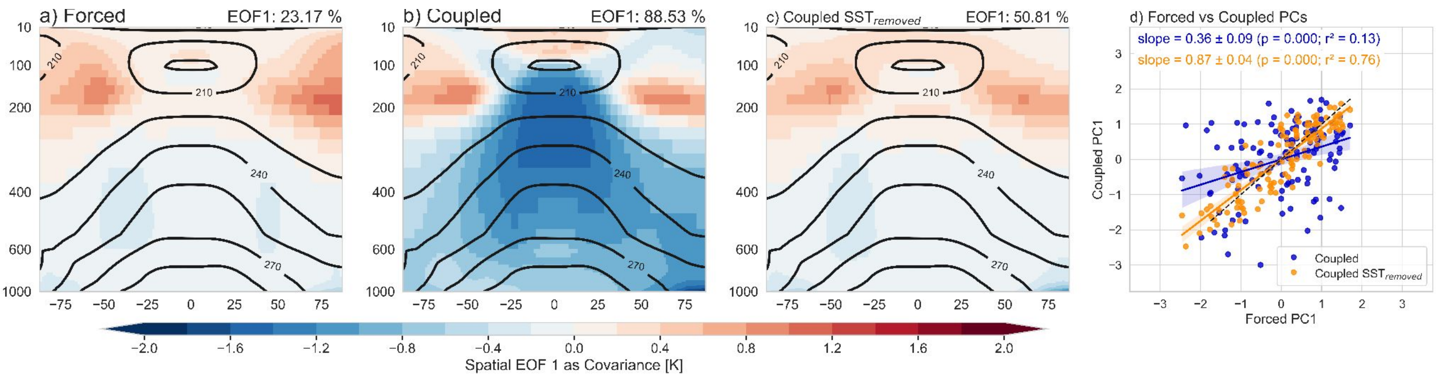
Vancoppenolle, M., Rousset, C., Blockley, E., Aksenov, Y., Feltham, D., Fichefet, T., Garric, G., Guémas, V., Iovino, D., Keeley, S., Madec, G., Massonnet, F., Ridley, J., Schroeder, D., & Tietsche, S. (2023). SI3, the NEMO Sea Ice Engine (4.2release_doc1.0). Zenodo. https://doi.org/10.5281/zenodo.7534900

Figure 1



Regressions of globally averaged annual mean a) net top of the atmosphere (TOA) radiation and b) net TOA minus net surface radiation plotted against the Coupled globally averaged sea surface temperature (SST). Forced values represent year 2 of simulation and Coupled values years 11-20.

Figure 2

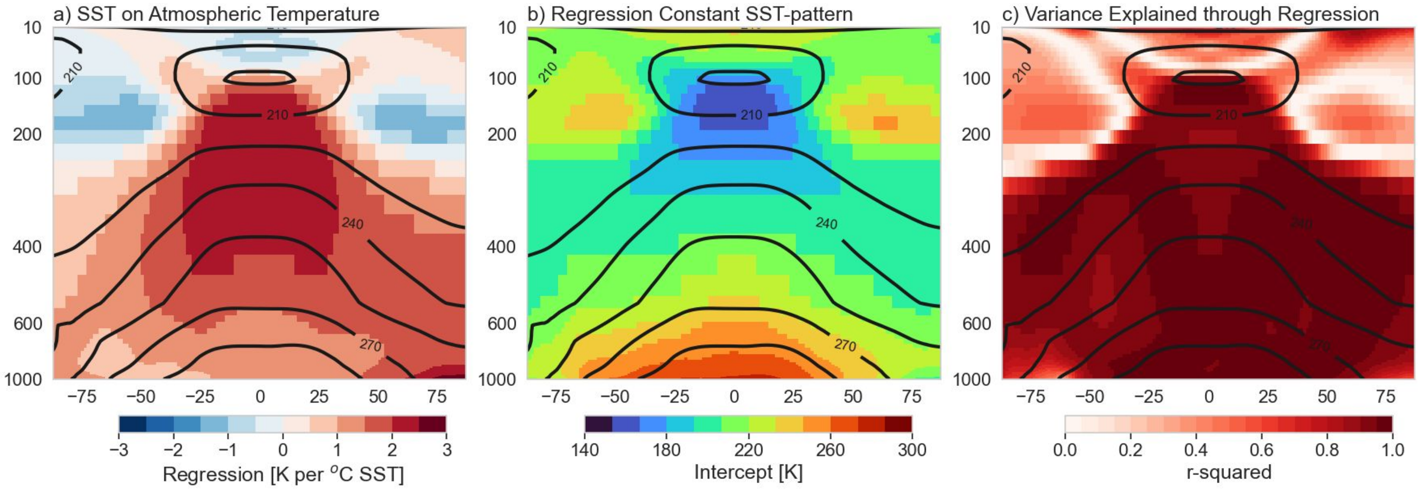


EOF 1 shown as covariance of the annual mean zonally averaged temperature across the ensemble of a) Forced, b) Coupled, and c) Coupled SSTremoved. Forced simulations represent year 2 and the coupled years 11-20. d) Relationship comparing the principle components (PC 1) between the Coupled and Coupled SSTremoved simulations against the Forced, demonstrating the similarity between each member of the PPE. Similar Coupled EOF and PC results using years 41-60.

Results: Correlations of PC and Figure 1: no fig, write in text. If we take the Forced Net TOA radiative balance from Figure 1a, the correlation with the Coupled PC is p = 0.000;r = -0.84; r^2 = 0.70…while the Forced Net TOA and Forced PC is p = 0.006;r = -0.25; r^2 = 0.06. Pattern represents well, Coupled SST and PC are p = 0.000;r = -0.99; r^2 = 0.97

Figure S1 (Fred mentioned to explore)

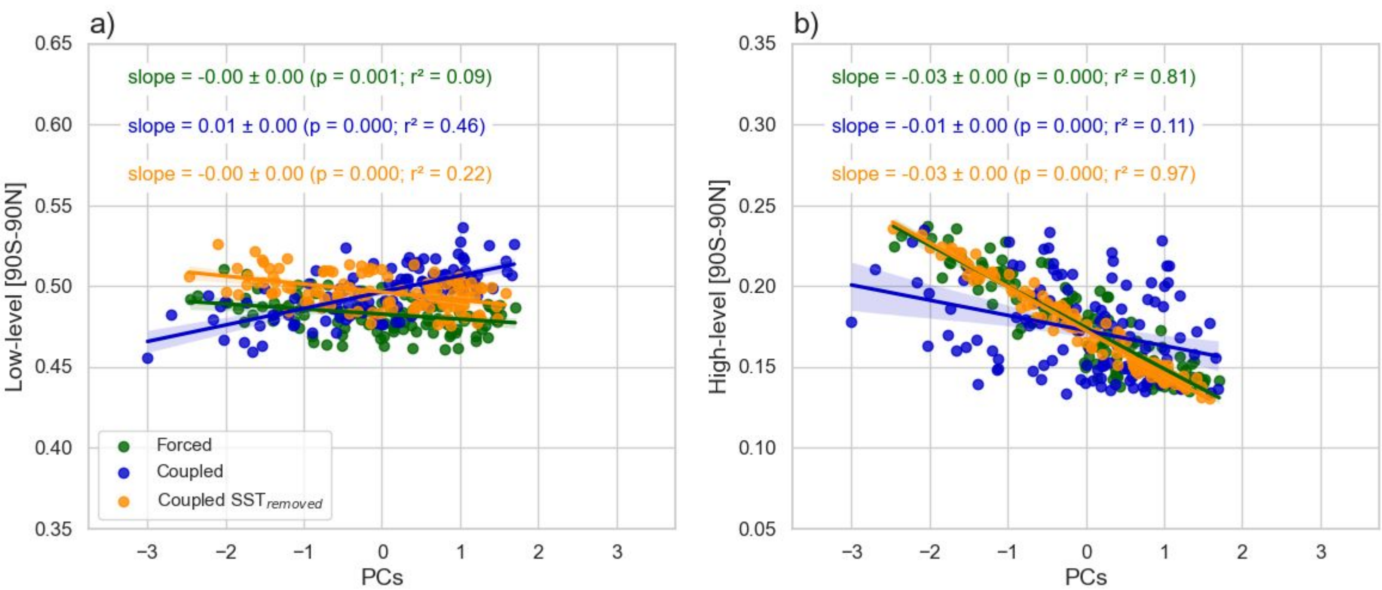
Background anomalous pattern or bias. Panel A link between Fig 1 and 2



SST response explains most of the troposphere response, small strat. Response. Isolates SST.

The a) slope, b) constant, and c) r-squared values representing the linear regression of the Coupled globally averaged sea surface temperature and the zonally averaged atmospheric temperature between years 11-20. The linear regression is done across members of the PPE.

Figure 3



Regressions of annual mean a) low-level and b) high-level cloudiness plotted against the PC 1. Forced values represent year 2 of simulation and Coupled values years 11-20.

Just focus on r-squared values, global average but discuss latitude and height in Figure S1

Figure S2

Regressions of zonally averaged (a-c) JFM and (d-f) JAS cloud fraction and the PC 1 of the Forced, Coupled, and Coupled SSTremoved. The Forced comparisons represent year 2 and the Coupled years 11-20.

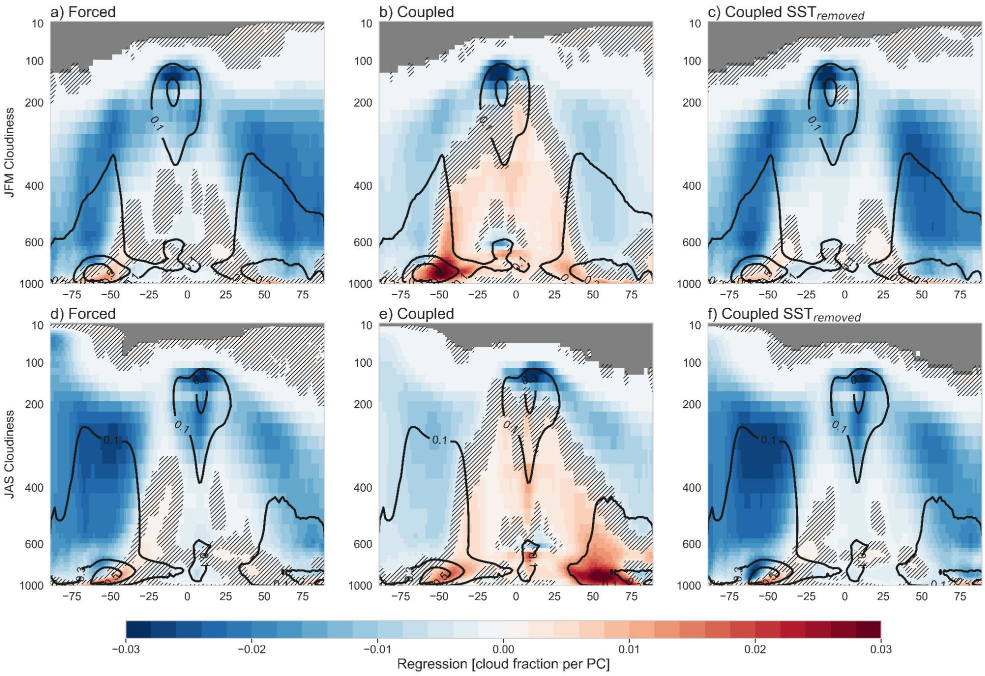
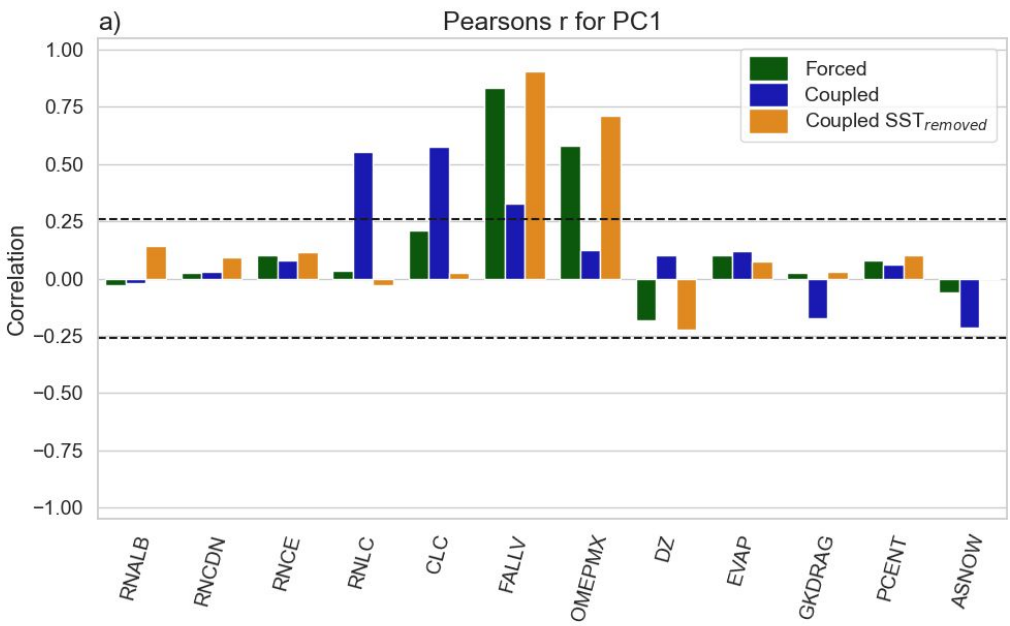


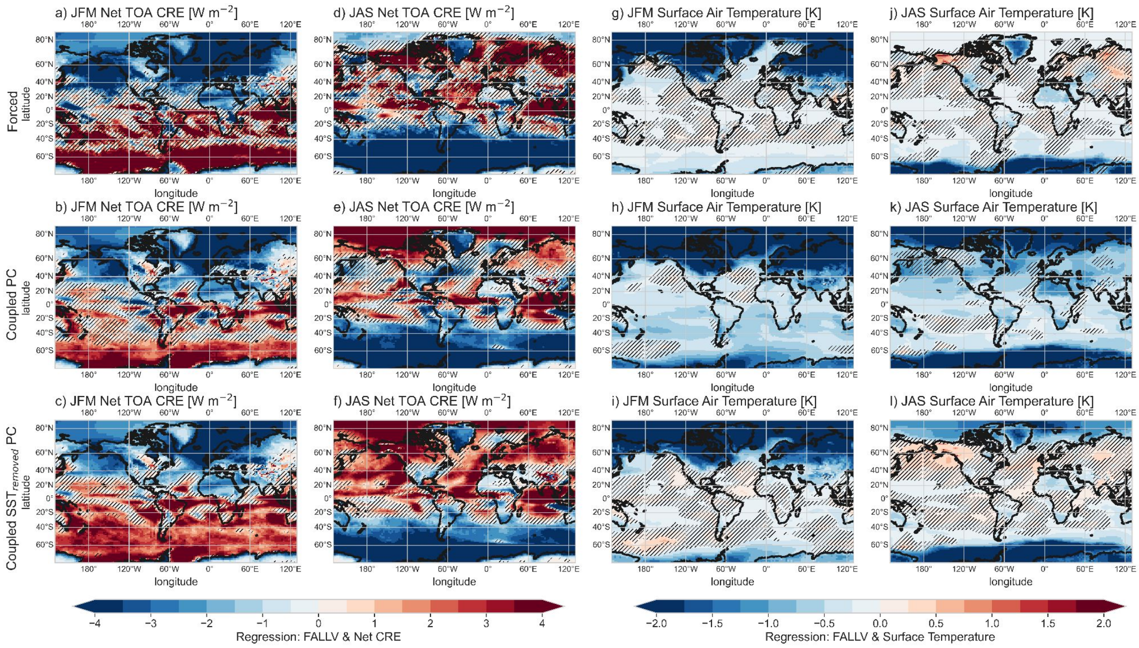
Figure 4



a) Pearson's r of the 12 parameters with PC 1. Forced values represent year 2 of simulation and Coupled values years 11-20. The horizontal dashed lines represent the critical r-value at an alpha of 0.05 for a sample size of 120 simulations. - mention testing SHAP, SOBOL, and Spearman's methods all agree, correlation is simple to explain but potential non-linearities

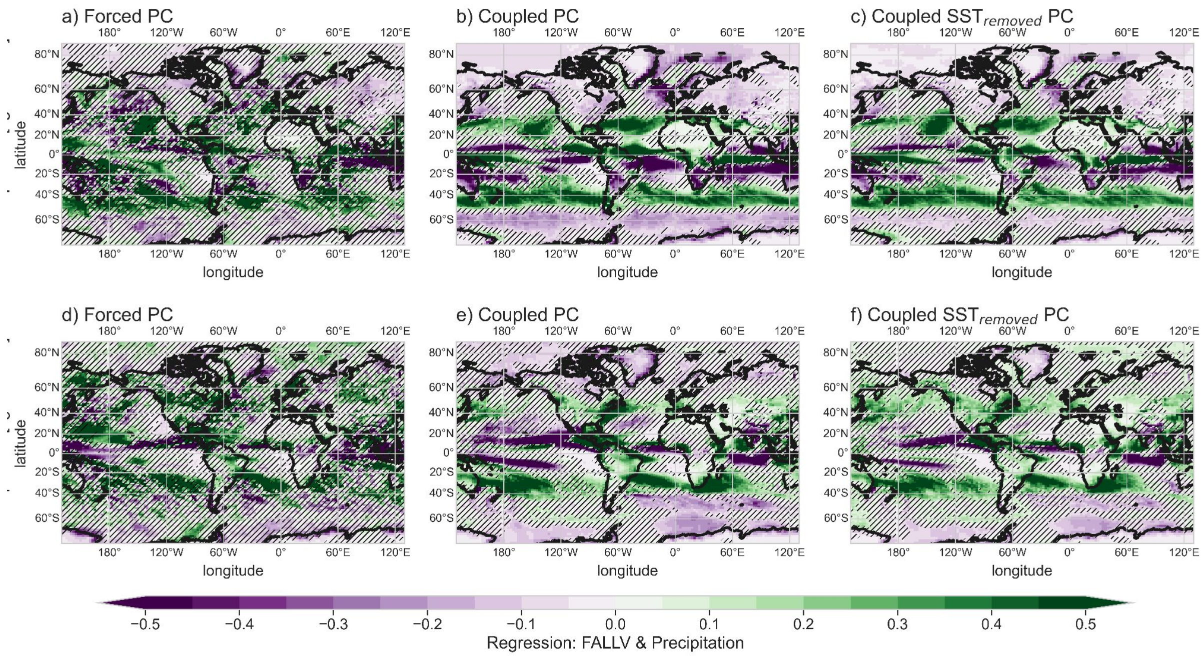
Regress FALLV on precip. Figure S3-S4

Figure S3



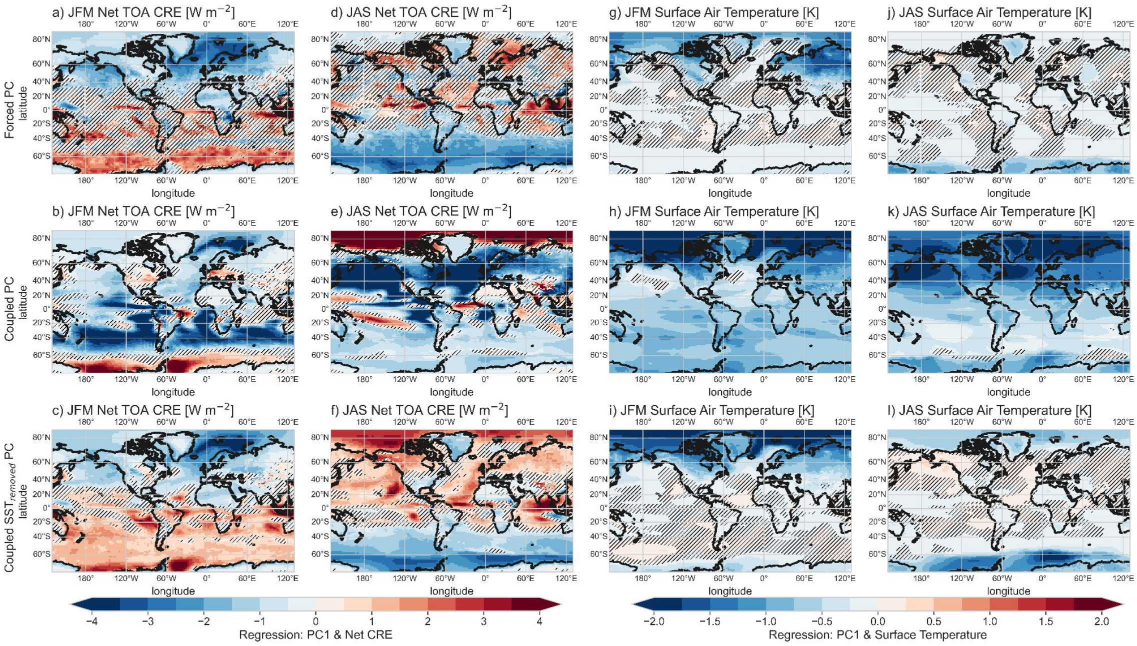
Regression of FALLV onto seasonal Net CRE and surface temperature.

Figure S4



Regression of FALLV onto seasonal precipitation

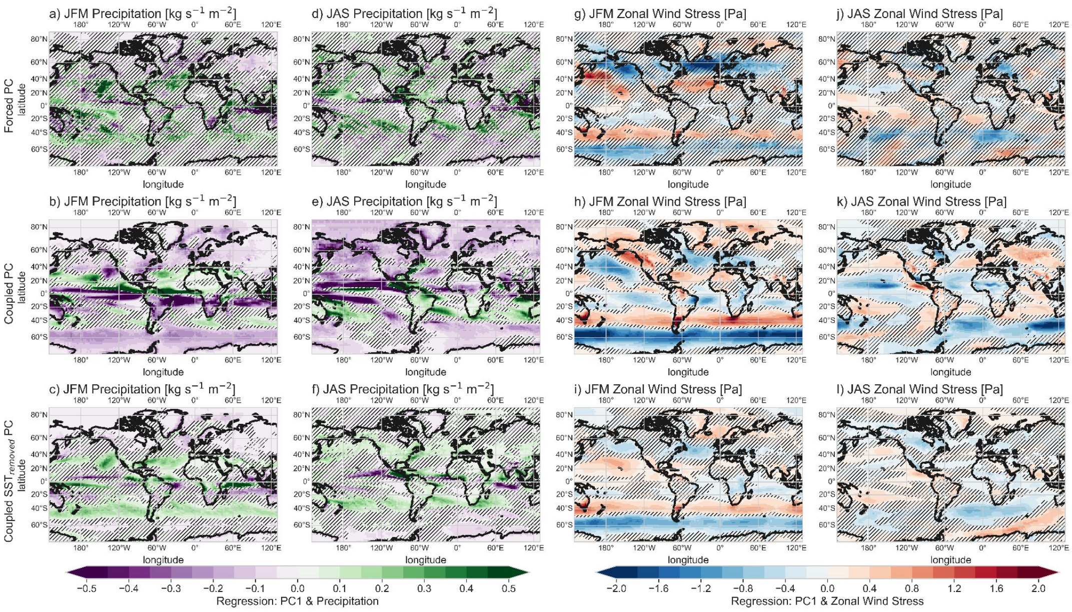
Figure 5



Regressions of (a-c) JFM and (d-f) JAS net CRE and the PC 1 of the Forced, Coupled, and Coupled SSTremoved. Regressions of (g-i) JFM and (j-l) JAS surface temperature and the PC 1 of the Forced, Coupled, and Coupled SSTremoved. The Forced comparisons represent year 2 and the Coupled years 11-20. As these PCs represent diversity in the ensemble, for example, Forced PC regressed across Forced temperature response in the PPE.

Figure 6

Find paper on Zonal winds and AA sea ice - link results to winds preconditioning AA sea ice extent. Link to Figure 7



Regressions of (a-c) JFM and (d-f) JAS precipitation and the PC 1 of the Forced, Coupled, and Coupled SSTremoved. Regressions of (g-i) JFM and (j-l) JAS zonal wind stress and the PC 1 of the Forced, Coupled, and Coupled SSTremoved. The Forced comparisons represent year 2 and the Coupled years 11-20. As these PCs represent diversity in the ensemble, for example, Forced PC regressed across Forced precipitation response in the PPE.

Figure S5

Is 1 year too short to look at winds or is the perturbation itself enough to bring diversity? - Fred

Regressions of zonally averaged (a-c) JFM and (d-f) JAS zonal velocity and the PC 1 of the Forced, Coupled, and Coupled SSTremoved. The Forced comparisons represent year 2 and the Coupled years 11-20.

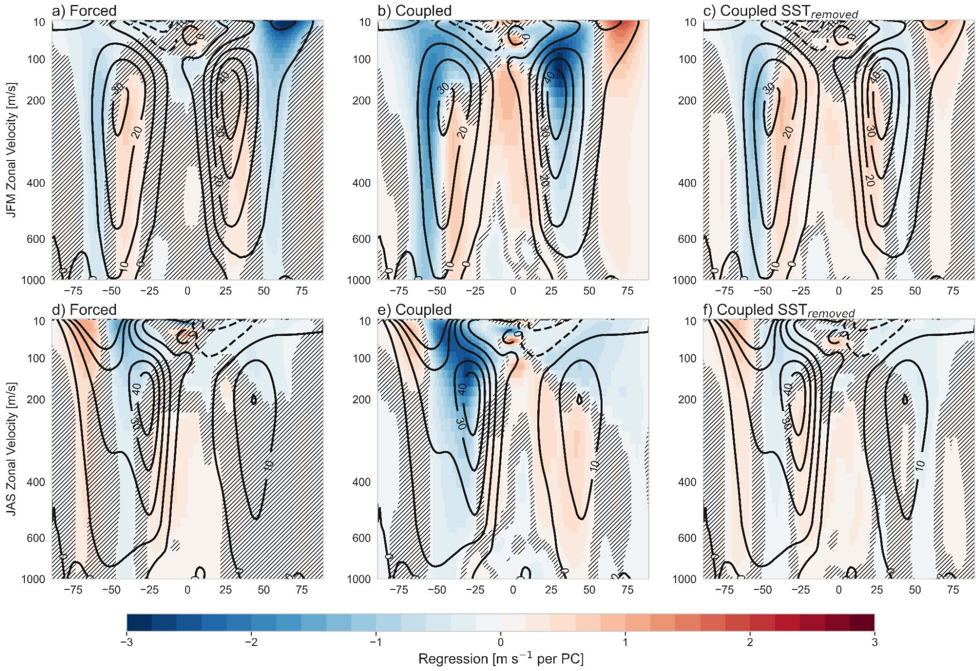
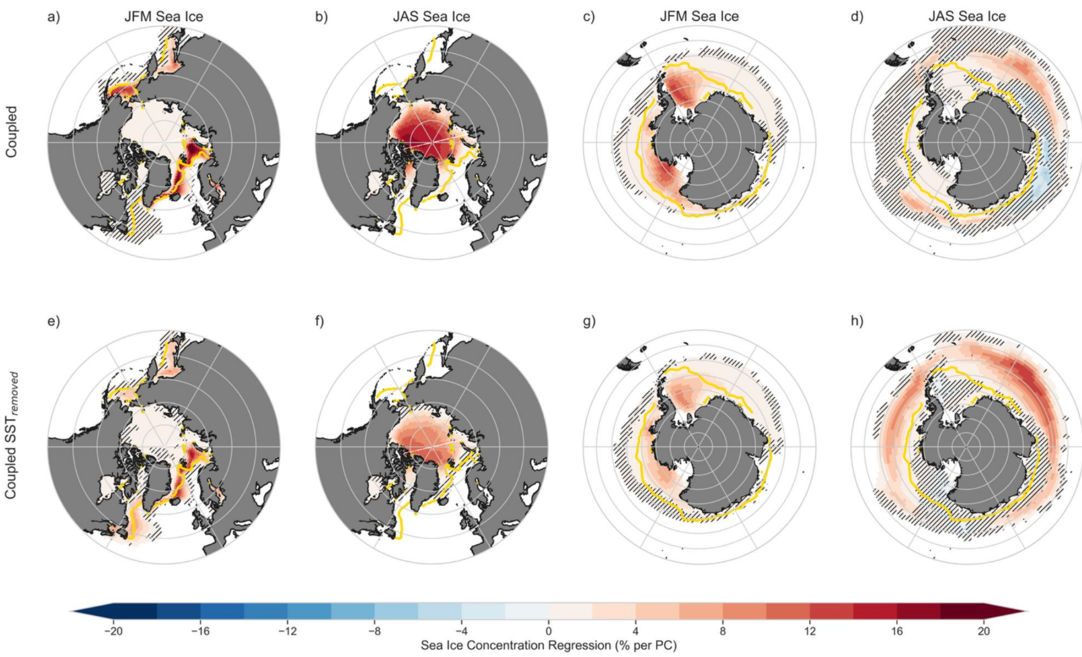
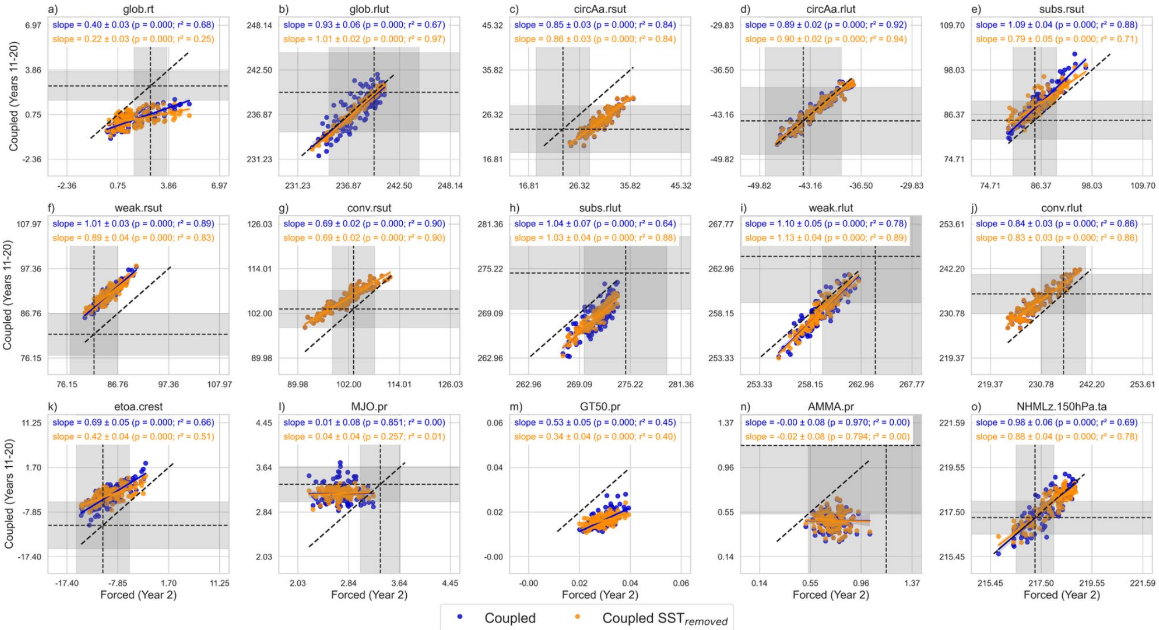


Figure 7



Regressions of PCs onto Coupled JFM (a-c) sea ice concentration and d) Arctic sea ice area. (e-h) Similar, but for Coupled JAS sea ice concentration and Antarctic sea ice area. The Forced PC represent year 2 and the Coupled PCs years 11-20. Forced and Coupled SSTremoved are regressed on the sea ice concentration field with the SST-effect removed.

Figure 8



Scatterplots of the metrics used in the tuning process comparing the Forced (x-axis) and Coupled (y-axis) -derived metrics. The Coupled simulations are shown as both the raw Coupled metrics (blue) and the Coupled metrics with the SST-regression removed (orange). The Forced comparisons represent year 2 and the Coupled years 11-20.

Table S1

Name	Min. Value	Max. Value	Current IPSL Model Value	Model Component	Short Description
CLC	1E-04	1,00E-03	6.5e-4	LMDZ	**Autoconversion threshold for liquid cloud water**.
FALLV	0.3	2.	0.8	LMDZ	**Fall speed of ice crystals (m/s)**.
OMEPMX	0.0003	0.02	0.001	LMDZ	**ice-to-rain conversion efficiency (1 − epmax)**
DZ	0.04	0.12	0.07	LMDZ	**Detrainment rate at the top of thermals**.
EVAP	5E-05	5,00E-04	1,00E-04	LMDZ	**Coefficient for re-evaporation of rain**.
GKDRAG	0.2	2.	0.6	LMDZ	**Subgrid-scale orographic drag coefficient**.
PCENT	0.3	1.	0.8	ORCHIDEE	**Maximum transpiration efficiency based on soil moisture.**
ASNOW	5.	15.	10.	ORCHIDEE	**Snow albedo adjustment**
RNALB	0.	1.	0.50	SI3	**Scaling factor for sea ice albedos**.
RNCND	0.10	0.50	0.31	SI3	**Thermal conductivity of snow over sea ice**.
RNCE	0.06	0.08	0.06	NEMO4	**Eddy diffusivity for mixed-layer processes**.
RNLC	0.05	0.5	0.15	NEMO4	**Langmuir cell mixing parameter**.