


**RESEARCH ARTICLE**

10.1029/2020MS002165

**Special Section:**

Using Radiative-Convective Equilibrium to Understand Convective Organization, Clouds, and Tropical Climate

**Key Points:**

- Climate feedback parameter ( $\lambda$ ) is impacted by change of convective aggregation with temperature, not by its average value
- Combined temperature dependencies of convective aggregation and shallow cloud fraction explain 70–80% of intermodel spread in  $\lambda$
- $\lambda$  is more negative in GCMs than in CRMs because only GCMs show a mean increase of convective aggregation with warming

**Supporting Information:**

- Supporting Information S1

**Correspondence to:**

T. Becker,  
tobias.becker@mpimet.mpg.de

**Citation:**

Becker, T., & Wing, A. A. (2020). Understanding the extreme spread in climate sensitivity within the Radiative-Convective Equilibrium Model Intercomparison Project. *Journal of Advances in Modeling Earth Systems*, 12, e2020MS002165. <https://doi.org/10.1029/2020MS002165>

Received 8 MAY 2020

Accepted 9 SEP 2020

Accepted article online 14 SEP 2020

# Understanding the Extreme Spread in Climate Sensitivity within the Radiative-Convective Equilibrium Model Intercomparison Project

Tobias Becker<sup>1</sup> and Allison A. Wing<sup>2</sup>

<sup>1</sup>Max Planck Institute for Meteorology, Hamburg, Germany, <sup>2</sup>Department of Earth, Ocean and Atmospheric Science, Florida State University, Tallahassee, FL, USA

**Abstract** The Radiative-Convective Equilibrium Model Intercomparison Project (RCEMIP) consists of simulations at three fixed sea-surface temperatures (SSTs: 295, 300, and 305 K) and thus allows for a calculation of the climate feedback parameter based on the change of the top-of-atmosphere radiation imbalance. Climate feedback parameters range widely across RCEMIP, roughly from  $-6$  to  $3 \text{ W m}^{-2} \text{ K}^{-1}$ , particularly across general-circulation models (GCMs) as well as global and large-domain cloud-resolving models (CRMs). Small-domain CRMs and large-eddy simulations have a smaller range of climate feedback parameters due to the absence of convective self-aggregation. More than 70–80% of the intermodel spread in the climate feedback parameter can be explained by the combined temperature dependencies of convective aggregation and shallow cloud fraction. Low climate sensitivities are associated with an increase of shallow cloud fraction (increasing the planetary albedo) and/or an increase in convective aggregation with warming. An increase in aggregation is associated with an increase in outgoing longwave radiation, caused primarily by mid-tropospheric drying, and secondarily by an expansion of subsidence regions. Climate sensitivity is neither dependent on the average amount of aggregation nor on changes in deep/anvil cloud fraction. GCMs have a lower overall climate sensitivity than CRMs because in most GCMs convective aggregation increases with warming, whereas in CRMs, convective aggregation shows no consistent temperature trend.

**Plain Language Summary** To determine how much Earth will warm in response to anthropogenic greenhouse gas emissions, we need to understand the atmospheric response to this forcing. The amount of warming in response to a given forcing is called climate sensitivity. Although global climate models are a useful tool to estimate climate sensitivity, estimates remain uncertain, in particular because the response of tropical clouds to warming is uncertain. The weakness of climate models is their coarse grid spacing, with which they cannot resolve important aspects of the weather like clouds and convection. In this study, we use a popular idealization for the tropics, the radiative-convective equilibrium setup, to compare climate sensitivities across a wide range of models including global climate models and cloud-resolving models. We find that more than 70–80% of variations in climate sensitivity across these models result from changes in shallow cloud fraction and changes in the spatial organization of convection with warming. Our results indicate that climate sensitivity might be underestimated by global climate models, in which the amount of spatial organization of convection mostly increases with warming, leading to low climate sensitivities, while the cloud-resolving models show no consistent trend in spatial organization, and thus have higher climate sensitivities.

## 1. Introduction

Climate sensitivity describes the amount of global warming that occurs in response to an external radiative forcing. Therefore, climate sensitivity controls many aspects of the climate response to global warming, both globally and locally, like the probability of droughts and flooding (IPCC, 2018). Thus, the value of halving the uncertainty in climate sensitivity may be in the trillions of dollars (Hope, 2015). The earliest estimates of *equilibrium* climate sensitivity (ECS), defined as the global-mean equilibrium surface air temperature response to a doubling of CO<sub>2</sub> concentration in the atmosphere, were made by Arrhenius (1896). His first ECS estimate of 5–6 K, which he later corrected to 4 K, is remarkably close to

©2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

modern estimates. For example, Charney et al. (1979) estimated ECS to be in the range of 1.5 to 4.5 K, with a best estimate of 3 K. The range of possible equilibrium climate sensitivities has not been significantly reduced over the last decades (e.g., Knutti et al., 2017), despite the use of increasingly complex models. In this study, we return to a less complex model setup that was used in many early studies to estimate ECS (e.g., Ramanathan & Coakley, 1978), the radiative-convective equilibrium (RCE) framework, to get a better quantitative understanding of the processes that explain most of the intermodel spread in climate sensitivity.

Generally, a large fraction of the spread in model-based climate sensitivity estimates can be attributed to the response of clouds, and particularly low clouds, to warming (e.g., Bony & Dufresne, 2005; Sherwood et al., 2014; Vial et al., 2013; Webb et al., 2013). In global climate models, clouds as well as the majority of processes that lead to their formation, such as moist convection, which is a main source of precipitation, are not resolved by the coarse grid and thus need to be parameterized. In a recent study, Fiedler et al. (2020) showed that the representation of tropical precipitation has not substantially improved over three generations of global climate models participating in the Coupled Model Intercomparison Project (CMIP; Meehl et al., 2007; Taylor et al., 2012; Eyring et al., 2016). To escape this deadlock, Stevens et al. (2019) proposed to replace general-circulation models (GCMs), which parameterize deep convection, with models that explicitly resolve deep convection, for example, a global cloud-resolving model (GCRM; Satoh et al., 2019). However, due to computational constraints, GCRM simulations are limited to short time scales and to kilometer-scale resolution, with the consequence that clouds, whose correct representation is crucial for climate sensitivity, are still poorly represented. Thus, to develop a deeper understanding of complex multiscale systems, it is beneficial to use a hierarchy of models (Bony, Stevens, et al., 2013; Held, 2005; Jeevanjee et al., 2017; Medeiros et al., 2008), for example, by linking GCRMs to coarse-resolution GCMs, to limited-area cloud-resolving models (CRMs) and to hectometer-scale large-eddy simulations (LES).

RCE, a balance between convective heating and radiative cooling, is a surprisingly accurate simplification of the tropical troposphere (e.g., Dines, 1917; Manabe & Strickler, 1964). The RCE framework is accessible to many types of atmospheric models, and there is probably no other framework that is so relevant for understanding climate change and so accessible to a hierarchy of models (e.g., Jeevanjee et al., 2017). Early RCE studies were conducted with single-column models (SCMs; e.g., Ramanathan & Coakley, 1978, and references therein), and later with CRMs (e.g., Held et al., 1993; Nakajima & Matsuno, 1988; Tompkins & Craig, 1998) and GCMs (e.g., Popke et al., 2013), for example, to improve the understanding of how clouds and convection couple to circulations and to contribute to the development and evaluation of atmospheric models and their parameterizations (e.g., Becker et al., 2018; Held et al., 2007; Reed & Medeiros, 2016).

A weakness of past RCE studies is that the analysis is usually limited to a single model and that different boundary conditions are used, which means that results have been difficult to compare across studies. These limitations are overcome by the Radiative-Convective Equilibrium Model Intercomparison Project (RCEMIP; Wing et al., 2018), which defines a standardized RCE protocol for many types of atmospheric models: GCMs, SCMs, GCRMs, CRMs, and LES. RCEMIP was motivated by three goals: (1) to investigate the robustness of the RCE state, (2) to analyze the response of clouds to warming and the impact on climate sensitivity, and (3) to understand how convective self-aggregation depends on temperature. One intriguing initial result of RCEMIP is that in those simulations that permit spatial organization of convection, there is a very wide range of climate sensitivities (Wing et al., 2020). This motivates our present study, which seeks to develop an understanding of the extreme spread in climate sensitivity in the RCEMIP ensemble. The focus of our investigation is on changes in cloud amount and spatial organization, which are expected to affect climate sensitivity (e.g., Bony et al., 2015; Cronin & Wing, 2017).

In addition to the ability to compare RCE across model types, the absence of large-scale heterogeneities in boundary conditions or forcing allows for an undistracted analysis of convection and its spatial organization in a setup that, due to the lack of external constraints, reveals differences between atmospheric models particularly well. Indeed, Wing et al. (2020) found that there was a wide range in the representation of mean profiles of temperature, humidity, and cloudiness in the RCEMIP ensemble. Convective self-aggregation, which is the spontaneous organization of convection despite homogeneous boundary conditions and forcings, was found to occur in nearly all large-domain RCEMIP simulations, though the spatial structure and degree of aggregation varied across models (Wing et al., 2020). Consistent with prior studies (e.g., Bretherton et al., 2005), there is widespread agreement across the RCEMIP ensemble that self-aggregation affects the mean simulated climate, by drying the non-convective regions and enhancing

outgoing longwave radiation (OLR) (Wing et al., 2020). Similar relationships between convective aggregation (of any type) and its large-scale environment are found in observations (Lebsack et al., 2017; Stein et al., 2017; Tobin et al., 2012, 2013). In addition, Bony et al. (2020) showed that the observed interannual variance of the tropical top-of-atmosphere (TOA) radiative imbalance depends on the degree of convective aggregation. Because of the impact of self-aggregation on the mean climate and radiative imbalance, self-aggregation may modulate climate sensitivity (e.g., Becker et al., 2017; Becker & Stevens, 2014; Cronin & Wing, 2017). However, there is no scientific consensus on how the propensity of convection to self-aggregate depends on temperature (reviewed by Wing, 2019). Thus, this study quantifies which intermodel differences contribute to the extreme spread in climate sensitivity across the different types of RCEMIP models, with a focus on the temperature dependencies of cloud fraction and convective self-aggregation.

## 2. What Is the Climate Sensitivity Across the RCEMIP Simulations?

Estimating equilibrium climate sensitivities from climate models that include a dynamical ocean component takes several thousand modeling years (e.g., Li et al., 2013; Rugenstein et al., 2019; Senior & Mitchell, 2000). Thus, an effective climate sensitivity is often used, which extrapolates from the transient temperature increase and TOA imbalance assuming certain behaviors of radiative feedbacks (e.g., Gregory et al., 2004; Murphy, 1995). To estimate climate sensitivity in the RCEMIP simulations, a third established method must be used, the Cess sensitivity method (Cess & Potter, 1988). In this method, the climate feedback parameter ( $\lambda$ ) is estimated, which can be transformed to ECS if the CO<sub>2</sub> forcing ( $F_{\text{CO}_2}$ ) is known:

$$\text{ECS} = -\frac{F_{\text{CO}_2}}{\lambda}, \quad (1)$$

where  $\lambda$  is calculated from the rate of change of the TOA radiative imbalance ( $R$ ) across simulations at different fixed surface temperatures ( $T$ ),

$$\lambda = \frac{dR}{dT}. \quad (2)$$

$R$  is the sum of net incoming shortwave ( $R_{\text{SW}}$ ) and longwave ( $R_{\text{LW}}$ ) radiation, so a positive value indicates an energy flux into the atmosphere. In RCEMIP,  $\lambda$  can be calculated in each model for a higher and a lower temperature range because each model provides three simulations at different fixed sea-surface temperatures (SSTs), at 295, 300, and 305 K. These are atmosphere-only aquaplanet simulations with no planetary rotation. In addition to homogeneous surface conditions, the insolation is everywhere equal to the tropical annual mean (409.6 W m<sup>-2</sup>), and all trace-gas concentrations are fixed and spatially uniform. More details on the boundary conditions and RCEMIP model setup are described in Wing et al. (2018, 2020).

RCEMIP simulations were performed with large- and small-domain configurations. Depending on model type, the large domain is either global (GCMs), global with reduced Earth radius (GCRMs), or an elongated channel of  $\sim 6,000 \times \sim 400$  km<sup>2</sup> (CRMs and one GCM), while the small domain is either the SCM version of a GCM or a square of  $\sim 100 \times \sim 100$  km<sup>2</sup> (CRMs). GCMs have horizontal and vertical grids similar to their CMIP6 configuration (Eyring et al., 2016), GCRMs and large-domain CRMs have a horizontal grid spacing of  $\sim 3$  km, and small-domain CRMs have a horizontal grid spacing of 1 km, but there are also two additional RCEMIP model types on the small square domain, with approximately doubled vertical (hereafter VERT) and additionally quintupled horizontal (hereafter LES) resolution. This information is summarized in Table 1, and the snapshots of OLR in Figure S1 in the supporting information visualize differences across the different RCEMIP model types. For further visualizations, in form of snapshots (of OLR and precipitable water) and movies for all RCEMIP models, the reader is referred to Wing et al. (2020). In Figure 1, we compare climate feedback parameters across 11 global-domain GCMs, six versions of WRF-GCM on the elongated channel, three GCRMs with reduced Earth radius ( $R_E/8$  for MPAS;  $R_E/4$  for NICAM and SAM-GCRM) merged with 14 CRMs on the elongated channel, 16 CRMs on the small square domain, six VERT models, and six LES models. All models except the GCMs resolve convection explicitly. We compute  $\lambda$  based on averages of  $R$  over the entire simulation, excluding only the first 75 days because within this period of time, an equilibrium state is reached in the RCEMIP simulations (see Figure 1 in Wing et al., 2020). The exception is the LES simulations, in which we only exclude the first 25 days because the LES simulations are only run for 50 days. CRMs and GCRMs are run for 100 days and GCMs are run for  $\sim 1,000$  days.

**Table 1**

Characteristics of RCEMIP Model Types With Respect to Approximate Domain Size (in  $\text{km}^2$ ), Horizontal Grid Spacing and Number of Vertical Levels, as Well as the Average and Standard Deviation of the Climate Feedback Parameter Based on Change of Top-of-Atmosphere All-Sky ( $\lambda$ ) and Clear-Sky ( $\lambda_{\text{CLR}}$ ) Radiative Imbalance With Warming (in  $\text{W m}^{-2} \text{K}^{-1}$ ), Calculated Across Both the Lower (300–295 K) and Higher (305–300 K) SST Range for Each Model Type

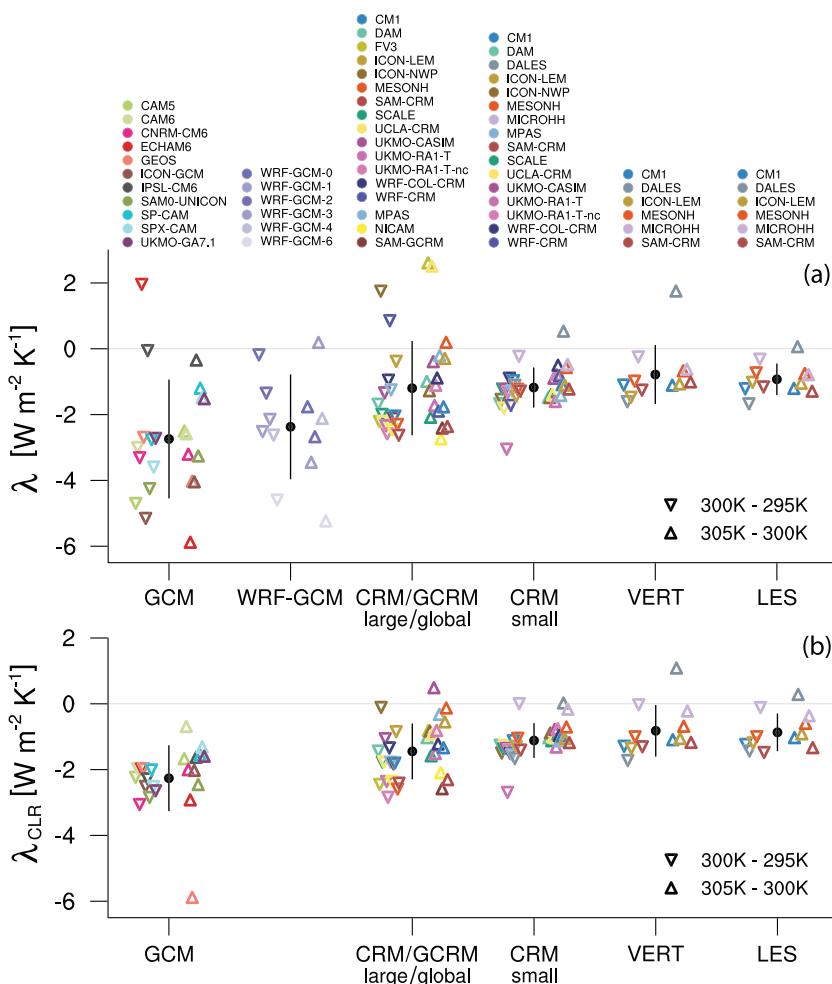
	GCM	WRF-GCM	GCRM	CRM large	CRM small	VERT	LES
Domain	Global	$6,000 \times 400$	Global <sup>a</sup>	$6,000 \times 400$	$100 \times 100$	$100 \times 100$	$100 \times 100$
Grid spacing	$\sim 1^\circ$	50 km	$\sim 3\text{--}4$ km	3 km	1 km	1 km	200 m
Vert. levels	26–91	48	~74	~74	~74	~146	~146
avg( $\lambda$ )	-2.7	-2.4		-1.2	-1.2	-0.8	-0.9
std( $\lambda$ )	1.8	1.6		1.4	0.6	0.9	0.5
avg( $\lambda_{\text{CLR}}$ )	-2.3	—		-1.4	-1.1	-0.8	-0.9
std( $\lambda_{\text{CLR}}$ )	1.0	—		0.8	0.5	0.8	0.6

Note. Clear-sky values are not available in WRF-GCM and WRF-CRM. Results are merged for GCRMs and large-domain CRMs.

<sup>a</sup> With reduced Earth radius:  $R_E/8$  for MPAS and  $R_E/4$  for NICAM and SAM-GCRM.

A comparison of the average climate feedback parameters (in Figure 1 and Table 1) across the different RCEMIP model types shows that  $\lambda$  is on average more negative in GCMs than in GCRMs and CRMs. Interestingly, the average climate feedback parameter is the same in the small-domain CRMs and in the large-domain CRMs (including the three GCRMs). Note that Cronin and Wing (2017) found a more negative climate feedback parameter (lower climate sensitivity) in a large domain than in a small domain. This contrast can be reproduced by about 50% of the large/small pairs of CRM simulations in RCEMIP (Wing et al., 2020), but the other 50% of CRMs behave differently and compensate in such a way that the average climate feedback parameter is the same in the large and small domain. This is also subtly different than in Wing et al. (2020), who found that in models with explicit convection, the average climate feedback parameter was slightly more negative in large-/global-domain simulations than in small-domain simulations (see Figure 17 in Wing et al., 2020). This is because the six VERT and LES models, which on average have the least negative feedback parameter (highest climate sensitivity), were included in the explicit group of small-domain simulations, whereas here, we consider them separately. These six models also have a similar climate feedback parameter in the default small-domain CRM setup, which means that climate sensitivity is almost independent of vertical and horizontal resolution across the analyzed resolutions in these small-domain simulations. In most models,  $\lambda$  becomes less negative with increasing temperature, in particular in the setups with explicit convection. In these setups,  $\lambda$  is on average  $0.5 \text{ W m}^{-2} \text{ K}^{-1}$  less negative between 305 and 300 K than between 300 and 295 K, and when  $\lambda$  is based on clear-sky radiative fluxes, the change with warming is even slightly larger ( $0.6 \text{ W m}^{-2} \text{ K}^{-1}$ , Figure 1b). This indicates that the temperature dependency can be linked to clear-sky mechanisms like an intensification of the water vapor feedback with increasing temperature, as explained by Held and Soden (2000). As an aside, Wing et al. (2020) found that  $\lambda$  was much more negative in GCMs than in SCMs (comparing the large and small versions of models with parameterized convection), but we don't consider SCMs here because our focus is on three-dimensional models that can represent circulations.

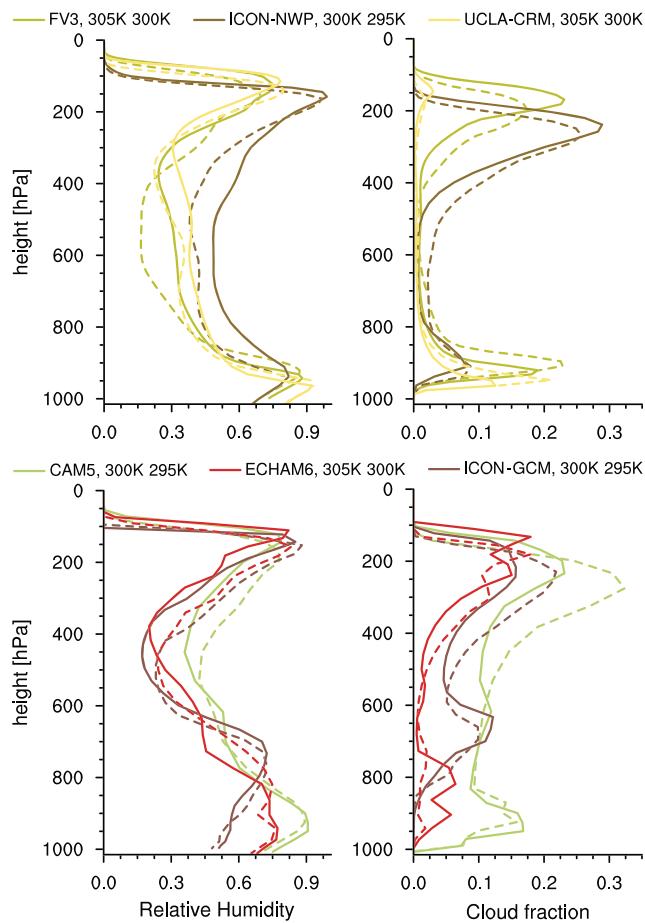
In the RCEMIP simulations, only a rough estimate of ECS can be made with Equation 1 because the  $\text{CO}_2$  concentration is not varied, and thus, the  $\text{CO}_2$  forcing is unknown and needs to be estimated. Here, we assume  $F_{2x\text{CO}_2} = 3.5 \text{ W m}^{-2}$ , which is the average  $\text{CO}_2$  forcing in CMIP6 (see Figure 1b in Zelinka et al., 2020). Across RCEMIP, this leads to an average ECS of  $\sim 1.3$  K in GCMs and  $\sim 2.9$  K in large- and small-domain CRMs. Compared to previous RCE studies (e.g., Cronin & Wing, 2017; Popke et al., 2013; Ramanathan & Coakley, 1978), these two values are at the lower and upper limit of the expected range. The average ECS in CMIP6, however, is higher, 3.9 K (Zelinka et al., 2020). This is expected because RCE climate feedbacks should be compared to local climate feedbacks over the tropical oceans in more comprehensive models; amplified warming over the poles and over land is absent in RCE (Popke et al., 2013). Nonetheless, some RCEMIP models have a positive climate feedback parameter, which would be associated with an unstable climate state and an infinite ECS. In the remainder of the paper, we refrain from making an assumption on the  $\text{CO}_2$  forcing and only discuss  $\lambda$  rather than ECS.



**Figure 1.** Climate feedback parameter based on change of top-of-atmosphere (a) all-sky ( $\lambda$ ) and (b) clear-sky ( $\lambda_{CLR}$ ) radiative imbalance with warming (Equation 2), calculated across both the lower (300–295 K) and higher (305–300 K) SST range. Downward fluxes are defined positive. All RCEMIP models are ordered by type. The average and the  $\pm 1$  standard deviation, calculated across both SST ranges for each model type, are indicated by the black circle and line. Clear-sky values are not available in WRF-GCM and WRF-CRM. Results are merged for large-domain CRMs and GCRMs.

A comparison of the intermodel spread in climate feedback parameters across the different RCEMIP model types shows that intermodel spread is largest in the GCMs, where  $\lambda$  ranges from  $-6$  to  $2\text{ W m}^{-2}\text{ K}^{-1}$  (Figure 1a), which translates to a standard deviation of  $1.8\text{ W m}^{-2}\text{ K}^{-1}$  (Table 1). Intermodel spread is second largest in the different versions of WRF-GCM and third largest in the large-/global-domain CRMs/GCRMs. All small-domain setups, which are too small for convection to self-aggregate (Wing et al., 2020), have a comparably small intermodel spread in  $\lambda$ , with a standard deviation of  $0.6\text{ W m}^{-2}\text{ K}^{-1}$  (Table 1). The large spread in WRF-GCM is remarkable, given that the only difference is which convective parameterization is used. Nonetheless, the remainder of our analysis focuses on the other model types because the main purpose of WRF-GCM is to bridge the gap between global GCMs and limited-domain CRMs. WRF-GCM was therefore configured with a very small number of model grid points, which means that robust results are more difficult to find in this setup. Note that ECHAM6 has both the highest climate feedback parameter of all GCMs (between 300 and 295 K), and the most negative climate feedback parameter (between 305 and 300 K). This reflects the nonlinear temperature dependency of the governing processes within individual models. In total, eight RCEMIP models have positive climate feedback parameters for at least one of the two 5 K SST ranges.

To get a general idea of the processes that are responsible for the extreme intermodel spread, Figure 2 shows the domain-mean vertical profiles of relative humidity and cloud fraction for the three CRMs with the



**Figure 2.** Relative humidity (left) and cloud fraction (right) versus height, for the three CRMs that have the most positive (top) and the three GCMs that have the most negative (bottom) climate feedback parameters. Solid lines show the simulations at higher SST and dashed lines at lower SST. Cloud fraction either results from the cloud parameterizations, or, in case of FV3 and UCLA-CRM, is calculated based on the cloud condensate thresholds defined by the RCEMIP protocol (Wing et al., 2018).

most positive (high climate sensitivity) and the three GCMs with the most negative (low climate sensitivity) climate feedback parameters. In the high climate sensitivity cases, the relative humidity in the middle troposphere and clouds in the upper troposphere increase with warming while clouds in the lower troposphere decrease with warming. The opposite tendencies can be found in the cases with low climate sensitivity. This is in line with expectations from previous work (e.g., Held & Soden, 2000; Manabe & Wetherald, 1967). In these previous studies, an increase in relative humidity with warming is associated with a strong water vapor feedback, which results in a TOA longwave radiation imbalance ( $R_{\text{LW}}$ ) that is less negative. An increase in high cloud fraction also means that  $R_{\text{LW}}$  is less negative, while an increase of shallow clouds increases the atmospheric albedo and thus results in a TOA shortwave radiation imbalance ( $R_{\text{SW}}$ ) that is less positive. If the climate feedback parameter is based on the clear-sky TOA radiative fluxes ( $\lambda_{\text{CLR}}$ ), the intermodel spread is 45% smaller in GCMs and 41% smaller in large-/global-domain CRMs/GCRMs (Table 1 and Figure 1); this is another indication that cloud changes have a significant contribution to the intermodel spread in climate sensitivity across the large-/global-domain RCEMIP models, as further discussed in the next section. However, note that for all small-domain simulations, the intermodel spread in  $\lambda_{\text{CLR}}$  and  $\lambda$  is similar (Table 1), indicating that, on the small domain, clear-sky processes are responsible for intermodel differences in  $\lambda$ . On the large/global domain, about 60% of the average difference in  $\lambda$  between the GCMs and CRMs/GCRMs can be attributed to differences in  $\lambda_{\text{CLR}}$  (Table 1), emphasizing the important contribution of clear-sky processes to general differences between these two model types.

### 3. What Explains the Extreme Spread in Climate Sensitivity?

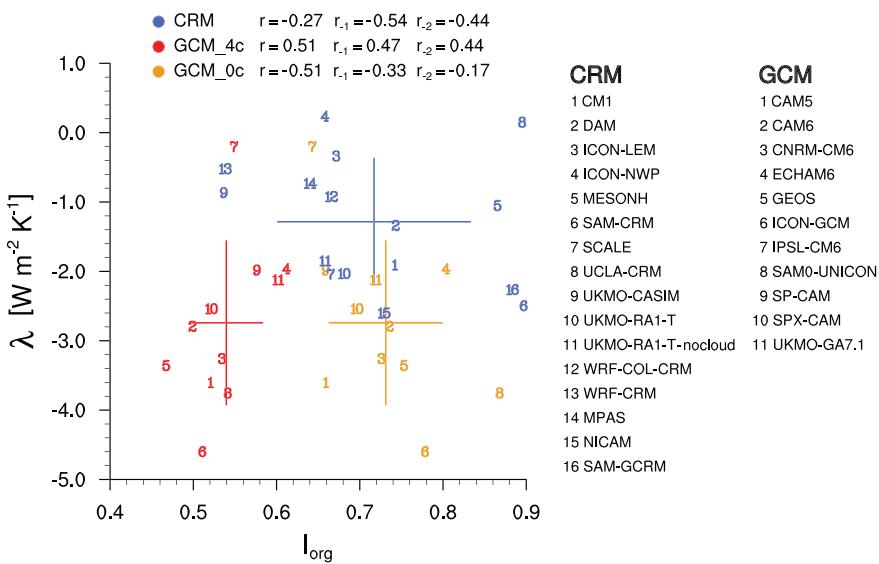
This section investigates the mechanisms that are responsible for the extreme spread in climate sensitivity across RCEMIP. The analysis will focus on the GCMs and the large-/global-domain CRMs/GCRMs (hereafter collectively referred to as CRMs for simplicity) because the intermodel spread in the climate feedback parameter  $\lambda$  is much larger in these simulations than in the small-domain simulations (Figure 1a). As described by Wing et al. (2020), convective self-aggregation only occurs in the global and large-domain simulations. Given that the degree of aggregation and its changes with warming vary across models (Wing et al., 2020), we investigate differences in convective aggregation as a possible explanation for the extreme spread in climate sensitivity (Section 3.1). In addition, changes in cloud fraction are known to affect climate sensitivity (see Section 2), which we investigate further in Section 3.2. The combined influence of convective aggregation and changes in cloud fraction is discussed in Section 3.3.

Because of missing model output, DAM will not be considered when indices are dependent on 500 hPa vertical velocity, WRF-CRM will not be considered for clear-sky radiative fluxes, and FV3 will not be considered at all due to a general shortage of two-dimensional model output. If not stated otherwise, hourly averages will be analyzed, excluding the first 75 days, except in IPSL-CM6 where only daily output is available.

#### 3.1. Convective Aggregation

In this study, we quantify convective aggregation primarily with the organization index,  $I_{org}$ , which is a clustering metric in which nearest neighbor distances of convective entities are compared to that expected from a random distribution.  $I_{org}$  has been used to characterize the spatial organization of deep convection across a variety of both modeling (e.g., Cronin & Wing, 2017; Tompkins & Semie, 2017; Wing et al., 2020) and observational (Bony et al., 2020) studies.  $I_{org}$  has the advantage that its computation is temperature independent and easy to interpret; values greater than 0.5 represent aggregated convection. However,  $I_{org}$  is integrated over multiple spatial scales and thus covers both large-scale organization related to moisture gradients and smaller-scale clustering within a region of convective activity. Our calculation of  $I_{org}$  is identical to that in the RCEMIP overview paper (Wing et al., 2020) and uses a threshold value of OLR ( $R_{LW,up} < 173 \text{ W m}^{-2}$ ) to define a convective column, following the RCEMIP protocol (Wing et al., 2018). To identify convective columns that are part of the same convective entity, 4-point connectivity (considering straight neighbors) is employed. Our results are insensitive to instead using 8-point connectivity (also considering diagonal neighbors). With respect to GCMs,  $I_{org}$  has some weaknesses and limitations. For example, to avoid problems with varying grid spacing in spherical models, we only consider grid points between  $-30^\circ$  and  $30^\circ$  latitude. In the NICAM simulation at 305 K, convection occurs almost entirely outside of this latitude band between simulation day 75 and 80, so this time interval is excluded when calculating  $I_{org}$ . In addition, due to the coarse resolution of GCMs, it may be just as justified to consider each convective column as its own entity (zero connectivity), rather than identifying connected columns. Therefore, we test the sensitivity of our results for GCMs to using 4-point or zero connectivity. We also compare to two other metrics of aggregation: subsidence fraction and the spatial variance of precipitable water scaled by its mean value. The latter is similar to spatial variance of column relative humidity, which was used in Wing et al. (2020) together with  $I_{org}$  and subsidence fraction to quantify convective self-aggregation.

Convective aggregation can affect climate sensitivity due to its influence on the mean state or by changing with warming (Becker et al., 2017; Cronin & Wing, 2017). Both the average degree of convective aggregation and its temperature dependence may vary across models. In both cases, our initial expectation is that, if convective aggregation is high or increases with warming, climate sensitivity will be small because of a weak water vapor feedback due to drying or a dry mean state. However, we note that the overall cloud feedback is less certain because while it is agreed upon that high clouds decrease with aggregation (Bony et al., 2016), changes in low clouds are ambiguous. To evaluate the dependence of climate sensitivity on the average degree of convective aggregation, we calculate the correlation between  $\lambda$  (based on the entire 10 K SST range) and  $I_{org}$  (averaged over the three different SSTs) for the CRMs and GCMs (Figure 3). To quantify the robustness of the results, we use two complementary approaches, a *t* test and an outlier exclusion method. A *t* test assumes that data must be randomly sampled from the population of interest and that the data variables follow a normal distribution, which is not necessarily fulfilled for the RCEMIP models, in particular within one family of models. Keeping this caveat in mind, the sign of any correlation discussed in the following paragraphs is correct with 99% confidence when correlation coefficients are higher than approximately 0.5 to 0.6 in most cases. Generally, confidence is a bit higher for the CRMs than for the GCMs because the

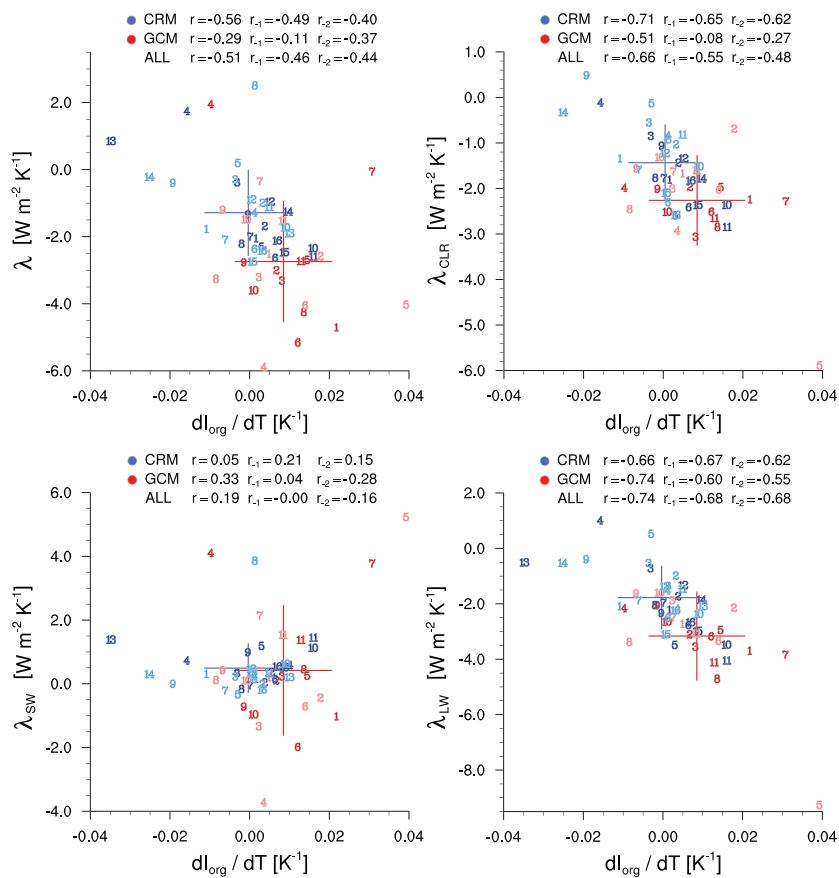


**Figure 3.** Climate feedback parameter (based on the entire 10 K SST range) versus organization index (averaged over the three different SSTs), for all GCMs and large-domain CRMs (including GCRMs). GCM\_4c and GCM\_0c stand for 4-point and zero connectivity when calculating  $I_{org}$ . Horizontal and vertical lines illustrate the  $\pm 1$  standard deviation in each dimension for each model type, and their intersection is the mean value. Correlation coefficients ( $r$ ) are given for CRMs and GCMs, using both  $I_{org}$  approaches for the latter. To test robustness, the largest as well as the two largest outliers (product across both dimensions) are excluded in  $r_{-1}$  and  $r_{-2}$ .

correlations are based on a larger number of models. Because of the aforementioned caveats with the  $t$  test and because the  $t$  test only estimates uncertainty in the sign of the correlation, but not in the strength of the correlation, our second approach is to estimate uncertainty by recalculating the correlation twice, once after removing the largest and once after removing the two largest outliers. We define the outliers based on the product of deviations from the model-category mean across both dimensions, in this case  $\lambda$  and  $I_{org}$ .

Based on these two approaches,  $\lambda$  does not show a robust dependency on the average degree of convective aggregation (Figure 3). In the CRMs,  $\lambda$  is on average slightly more negative when  $I_{org}$  is large, in line with our initial expectation, but the correlation is weak, so the  $t$  test fails. The  $t$  test also fails for the GCMs, and the sign of the correlation is even opposite depending on whether the 4-point or zero connectivity approach is used in the  $I_{org}$  calculation. The results remain similarly inconclusive for the correlation of  $I_{org}$  with  $\lambda$  based only on longwave fluxes ( $\lambda_{LW}$ ; Figure S2) and when using different metrics to quantify convective aggregation (spatial variance of scaled precipitable water and subsidence fraction; Figure S3). The only correlation we find that is significant with 95% confidence is that between  $\lambda$  and subsidence fraction in CRMs. The climate feedback parameter also varies across the small-domain CRM simulations, for reasons that are unrelated to aggregation. Therefore, in an attempt to remove this baseline spread in  $\lambda$  across CRMs and to investigate the effects of convective aggregation in isolation, we scale the large-domain climate feedback parameter by the feedback parameter in the small, non-aggregated domain. However, the correlation of this scaled feedback parameter with  $I_{org}$  is even weaker (Figure S4). In sum, despite the initial expectation of a smaller climate sensitivity in simulations with on average more aggregated convection, climate sensitivity does not show a clear dependency on  $I_{org}$ . In line with this, the small- and large-domain CRMs have on average the same climate feedback parameter (Figure 1a and Table 1). The expectation that simulations with aggregated convection have lower climate sensitivities than simulations with non-aggregated convection is only fulfilled if  $\lambda$  is based on the clear-sky TOA radiative fluxes:  $\lambda_{CLR}$  is more negative in the large-domain CRMs than in the small-domain CRMs (Figure 1b and Table 1), but the imprint on the all-sky climate feedback parameter is compensated by effects attributable to clouds.

Though  $\lambda$  does not correlate with the *average* amount of convective aggregation, it does seem to depend on the *change* of convective aggregation with warming, as Figure 4 shows. Climate feedback parameters are more negative in models in which convective aggregation increases with warming and climate feedback parameters are less negative or even positive in models in which convective aggregation decreases with warming. When calculated with all-sky fluxes,  $\lambda$  shows a robust dependence on  $dI_{org}/dT$  across the entire



**Figure 4.** Climate feedback parameter based on the TOA all-sky (top left), clear-sky (top right), shortwave (bottom left), and longwave (bottom right) radiative fluxes versus change of organization index with warming, for GCMs and CRMs, numbered according to Figure 3. Darker colors show change between the two simulations at colder SST (300–295 K), and lighter colors at warmer SST (305–300 K). Horizontal and vertical lines illustrate the  $\pm 1$  standard deviation in each dimension for each model type, and their intersection is the mean value. Correlation coefficients ( $r$ ) are given for CRMs and GCMs both separately and combined. To test robustness, the largest as well as the two largest outliers (product across both dimensions) are excluded in  $r_{-1}$  and  $r_{-2}$ .

ensemble of models, as well as when considering CRMs alone. These results are insensitive to the choice of aggregation metric (Figure S5) and the choice of connectivity in the  $I_{\text{org}}$  calculation in GCMs (Figure S6). Additionally, Figure 4 shows that  $dI_{\text{org}}/dT$  correlates most strongly with  $\lambda$  calculated using longwave ( $\lambda_{\text{LW}}$ ) or clear-sky ( $\lambda_{\text{CLR}}$ ) fluxes. The latter indicates that cloud-radiative effects (CREs) are at most of secondary importance in how climate sensitivity is affected by convective aggregation, with the primary importance given to clear-sky mechanisms like enhanced cooling from dry regions. Even though  $dI_{\text{org}}/dT$  has a similar correlation with  $\lambda_{\text{LW}}$  in CRMs and GCMs, the correlation with  $\lambda$  is significantly smaller and not robust in GCMs, indicating that another process that affects shortwave radiation is masking the overall dependency on aggregation in the GCMs. Interestingly, averaged over all GCMs, convective aggregation increases with warming, while averaged over all CRMs there is no temperature dependency. This result is also backed up by the other two aggregation metrics in Figure S5, spatial variance of scaled precipitable water and subsidence fraction, which both indicate an increase of convective aggregation with warming in the GCMs, while the CRMs have an opposite trend. The different dependency of convective aggregation on temperature might be the main reason why within RCEMIP, the GCMs generally have more negative climate feedback parameters than the CRMs. This also explains why about 60% of the average difference in  $\lambda$  between GCMs and CRMs is apparent in  $\lambda_{\text{CLR}}$  (Table 1), and why almost 100% of the average difference in  $\lambda$  is apparent in  $\lambda_{\text{LW}}$ , as a comparison of the horizontal lines (indicating the mean value for each model type) across the different panels in Figure 4 shows.

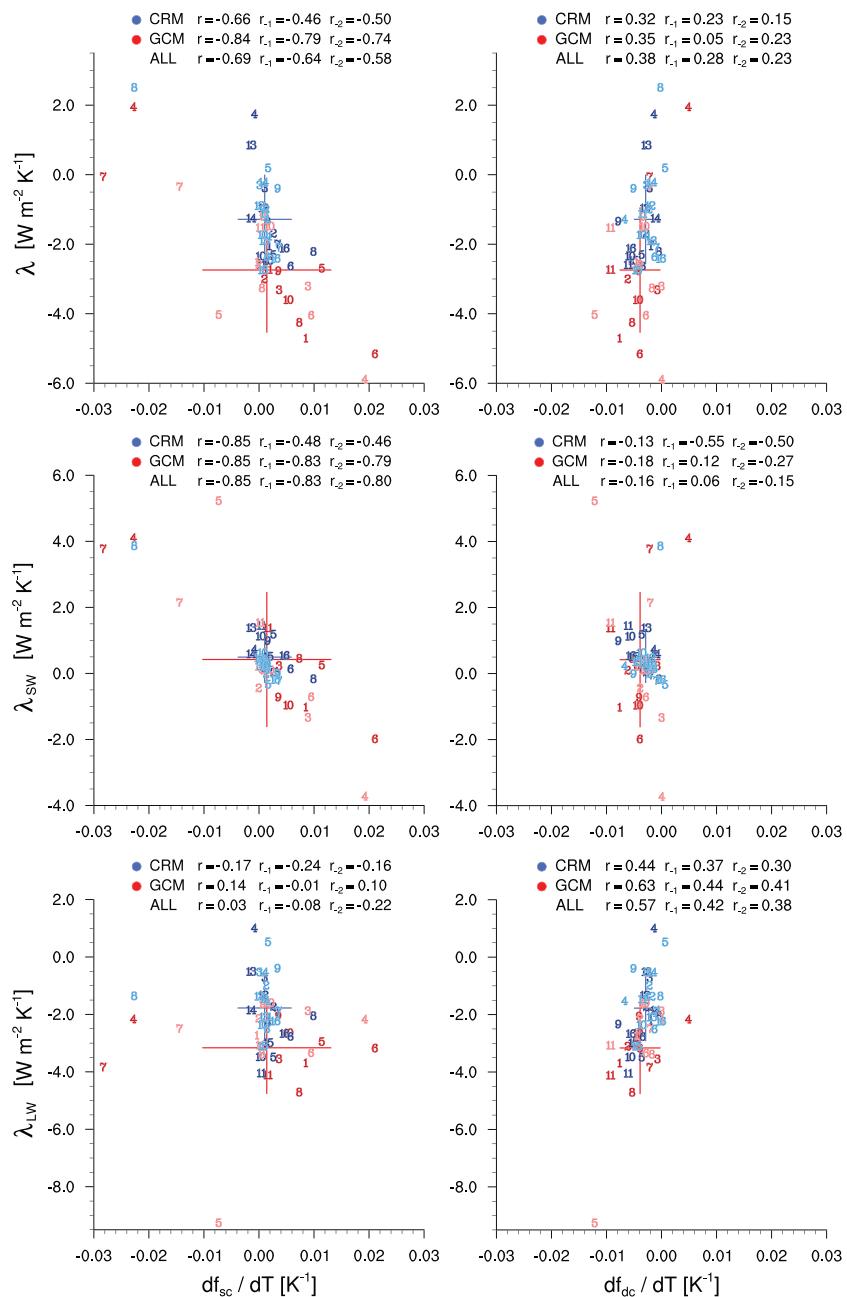
Three GCMs are noticeable outliers in Figure 4: ECHAM6, GEOS, and IPSL-CM6. ECHAM6, which has the largest difference in  $\lambda$  between its 305–300 and 300–295 K estimate of all the RCEMIP models, has a decrease in  $I_{org}$  with warming in the case of the positive  $\lambda$  value, and an increase in  $I_{org}$  with warming in the case of the extremely negative  $\lambda$  value, so this is in line with the correlation we find. However, the magnitude of the change in  $I_{org}$  is relatively low, indicating that a second process must be responsible for the extreme differences in  $\lambda$  within that model. GEOS (between 305 and 300 K) and IPSL-CM6 (between 300 and 295 K) are the two models with the strongest increase in  $I_{org}$  with warming of all RCEMIP models. In GEOS, this causes OLR to increase much more strongly with warming than in any of the other models, leading to an extremely negative  $\lambda_{LW}$  and  $\lambda_{CLR}$ , but also overall to a strongly negative  $\lambda$ . Even though  $\lambda_{LW}$  is also more negative in IPSL-CM6 (between 300 and 295 K) than in most other models,  $\lambda$  is overall close to zero, causing a decrease in the correlation between  $dI_{org}/dT$  and  $\lambda$  for the GCMs.  $\lambda$  being close to zero in IPSL-CM6 (between 300 and 295 K) is related to a strong decrease in reflected shortwave radiation with warming, a tendency that is also apparent in GEOS (between 305 and 300 K) and in ECHAM6 (between 300 and 295 K). This tendency can be explained by a decrease in cloud fraction, as the next section will show.

### 3.2. Cloud Fraction

We define the area covered by shallow and deep clouds based on threshold values of the TOA radiative flux. This definition, which reflects the effect of clouds on radiation, is better suited for our interest in climate sensitivity than defining cloud fraction based on a threshold value of cloud condensate that captures optically thin clouds, as was used in Wing et al. (2020). Our conditions for shallow clouds are (1)  $R_{SW,up} > 100 \text{ W m}^{-2}$ , thus ensuring the presence of clouds that reflect a significant amount of shortwave radiation, and (2)  $R_{LW,up} > 250 \text{ W m}^{-2}$ , thus excluding areas of deep convection. Note that a fixed threshold for  $R_{SW,up}$  is only justified in simulations with uniform insolation and surface albedo in space and time and depends on  $R_{SW,down}$ , which is set to  $409.6 \text{ W m}^{-2}$  in the RCEMIP simulations. The threshold for  $R_{LW,up}$  is in line with the low-level cloud threshold used in Fiedler et al. (2020), and tests confirmed that the presented results are robust across a range of different threshold values for  $R_{LW,up}$ . The shallow cloud definition neglects shallow clouds that are concealed by high clouds, but these clouds are also much less relevant for the radiation imbalance. Our definition of deep clouds is the same as what we use to define grid points with deep convective activity for  $I_{org}, R_{LW,up} < 173 \text{ W m}^{-2}$ . Deep clouds are thus considered as the sum of deep convective clouds and high stratiform clouds like anvil clouds. According to these definitions, deep cloud fraction ranges from 0.01 to 0.18 across the GCMs and CRMs, while shallow cloud fraction ranges from 0.00 to 0.22 (Figure S7). These values, which for each model are averaged over the three different SSTs, do not show any robust correlation with the climate feedback parameter.

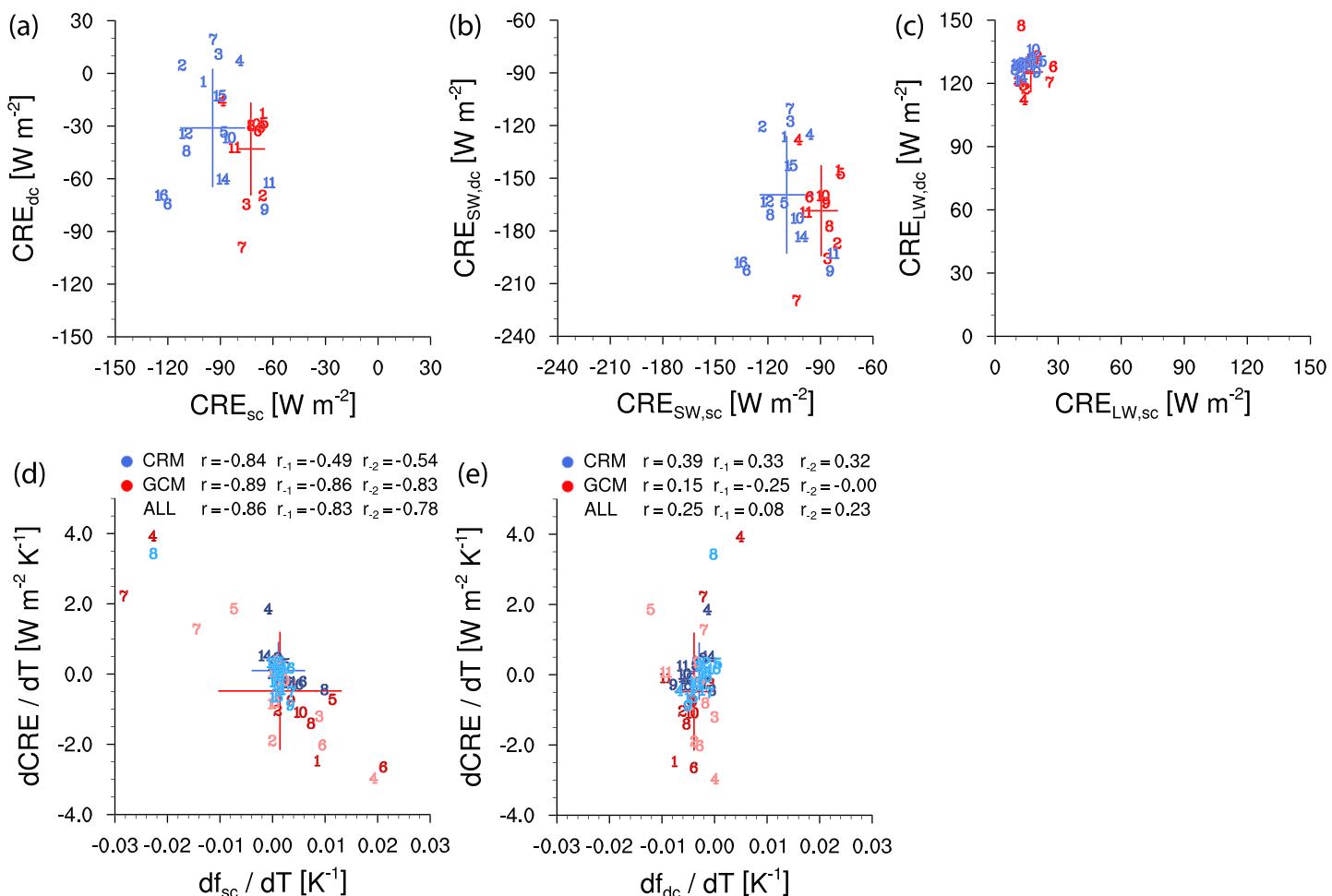
Instead, changes in shallow cloud fraction with warming correlate robustly with  $\lambda$  and with the climate feedback parameter based on shortwave fluxes ( $\lambda_{SW}$ ), while there is no correlation with  $\lambda_{LW}$  (Figure 5). There is no significant correlation between changes in deep cloud fraction with warming and  $\lambda$ . However, there is a weak tendency of  $\lambda$  to be less negative when deep cloud fraction increases with warming. This is particularly true when considering  $\lambda_{LW}$ , which indicates that high clouds reduce OLR. Averaged over the respective model types, deep cloud fraction decreases with warming in GCMs and slightly decreases with warming in CRMs, in line with the stability iris effect (Bony et al., 2016). Shallow cloud fraction, on the other hand, slightly increases with warming both in GCMs and CRMs, but the increase is almost zero compared to the intermodel spread. Though these results are not directly comparable to those presented in the RCEMIP overview paper due to the different definition of clouds (Wing et al., 2020), the general tendency for a reduction in deep cloud fraction with warming is consistent. Overall, comparing the relative contribution of changes in shallow clouds and deep clouds to climate sensitivity, changes in shallow cloud fraction correlate more strongly with  $\lambda$  and the intermodel spread of changes in shallow cloud fraction is also larger than for deep clouds, particularly so in one CRM, UCLA-CRM, and in the GCMs. In ECHAM6, the extremely high climate sensitivity between 295 and 300 K as well as the extremely low climate sensitivity between 300 and 305 K is related to a shallow cloud fraction that is much smaller at 300 K than at the other two SSTs (Figure 5).

The aforementioned correlations, of course, do not guarantee causality. For example, it is plausible that changes in deep cloud fraction only correlate with  $\lambda_{LW}$  because changes in deep cloud fraction correlate with changes in convective aggregation. Indeed, this correlation is significant for GCMs (CRMs:  $r = -0.2$ ; GCMs:  $r = -0.6$ ), in line with expectations that the coverage of deep clouds is reduced when convection is more



**Figure 5.** Climate feedback parameter based on the TOA all-sky (top), shortwave (middle), and longwave (bottom) radiative fluxes versus change of shallow (left) and deep (right) cloud fraction with warming, for GCMs and CRMs, numbered according to Figure 3. Darker colors show change between the two simulations at colder SST (300–295 K), and lighter colors at warmer SST (305–300 K). Horizontal and vertical lines illustrate the  $\pm 1$  standard deviation in each dimension for each model type, and their intersection is the mean value. Correlation coefficients ( $r$ ) are defined like in Figure 4.

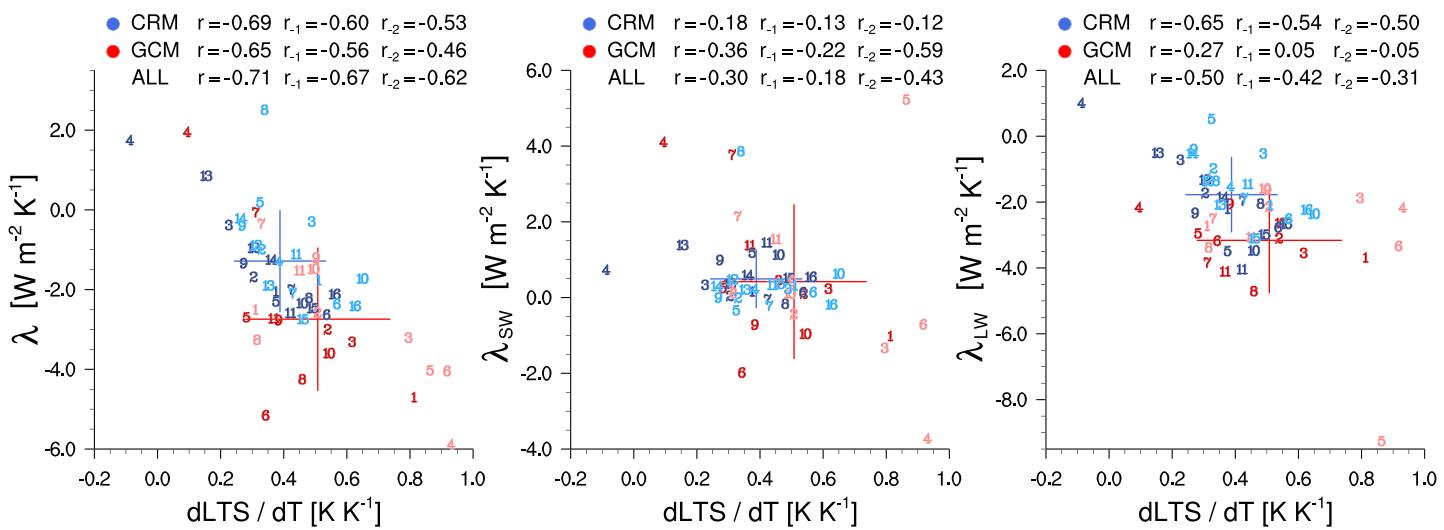
aggregated (e.g., Bony et al., 2016). The increase of  $I_{org}$  with warming in GCMs might also explain why the tendency of deep cloud fraction to decrease with warming is stronger in GCMs than in CRMs. To get a better understanding of causality, we also examine the effect of clouds on the TOA radiation imbalance more directly by investigating the CRE (CRE =  $R - R_{CLR}$ ) of shallow and deep clouds. Figure 6d shows that the correlation of changes in CRE with changes in shallow cloud fraction is strongly negative, while the correlation with changes in deep cloud fraction is weakly positive (Figure 6e). The latter is surprising given that



**Figure 6.** Cloud-radiative effect averaged over all grid columns covered by deep versus shallow clouds for (a) all-sky, (b) shortwave, and (c) longwave radiative fluxes, and change of cloud-radiative effect with warming versus change of (d) shallow and (e) deep cloud fraction with warming, for GCMs (red) and CRMs (blue), numbered according to Figure 3. The three top panels show the average over the three different SSTs, and the two bottom panels show the change between the two simulations at colder SST (300–295 K) in the darker color, and at warmer SST (305–300 K) in the lighter color. Horizontal and vertical lines illustrate the  $\pm 1$  standard deviation in each dimension for each model type, and their intersection is the mean value. Correlation coefficients ( $r$ ) are defined like in Figure 4.

the CRE of deep clouds is negative in most CRMs, and in all GCMs (Figure 6a), so the direct effect of increasing the deep cloud fraction with warming should be a decrease in climate sensitivity. This indicates that, in particular in the GCMs, climate sensitivity depends only indirectly on deep cloud fraction itself. The dominant factor is instead that a large deep cloud fraction is associated with weak convective aggregation. Figure 6a also shows that the CRE is two to three times more negative for shallow than for deep clouds. This is because in most models, the longwave CRE is much less positive for shallow than for deep clouds (Figure 6c), while the shortwave CRE is only moderately less negative for shallow than for deep clouds (Figure 6b). Note that a general caveat of looking at the CRE is that it does not provide the full picture of how clouds and cloud changes affect climate sensitivity because the overall cloud feedback also depends on and interacts with changes in clear-sky radiation. For an accurate analysis, radiative kernel calculations should be used, which can generally differ from the CRE by about  $0.3\text{--}0.4 \text{ W m}^{-2} \text{ K}^{-1}$  (Soden et al., 2004). This might be a focus of future analysis of the RCEMIP simulations.

Overall, shallow clouds affect climate sensitivity more strongly than deep clouds, both because their CRE is larger (more negative), and because the temperature dependency of shallow cloud fraction is more inconsistent across models, thereby introducing more spread in climate sensitivity. This is why we concentrate on shallow clouds in the next section, where we investigate the combined effect of changes in convective aggregation and shallow cloud fraction on climate sensitivity.



**Figure 7.** Climate feedback parameter based on the TOA all-sky (left), shortwave (middle), and longwave (right) radiative fluxes versus change of lower-tropospheric stability with warming, for GCMs and CRMs, numbered according to Figure 3. Darker colors show change between the two simulations at colder SST (300–295 K), and lighter colors at warmer SST (305–300 K). Horizontal and vertical lines illustrate the  $\pm 1$  standard deviation in each dimension for each model type, and their intersection is the mean value. Correlation coefficients ( $r$ ) are defined like in Figure 4.

### 3.3. Combining Convective Aggregation and Shallow Cloud Fraction

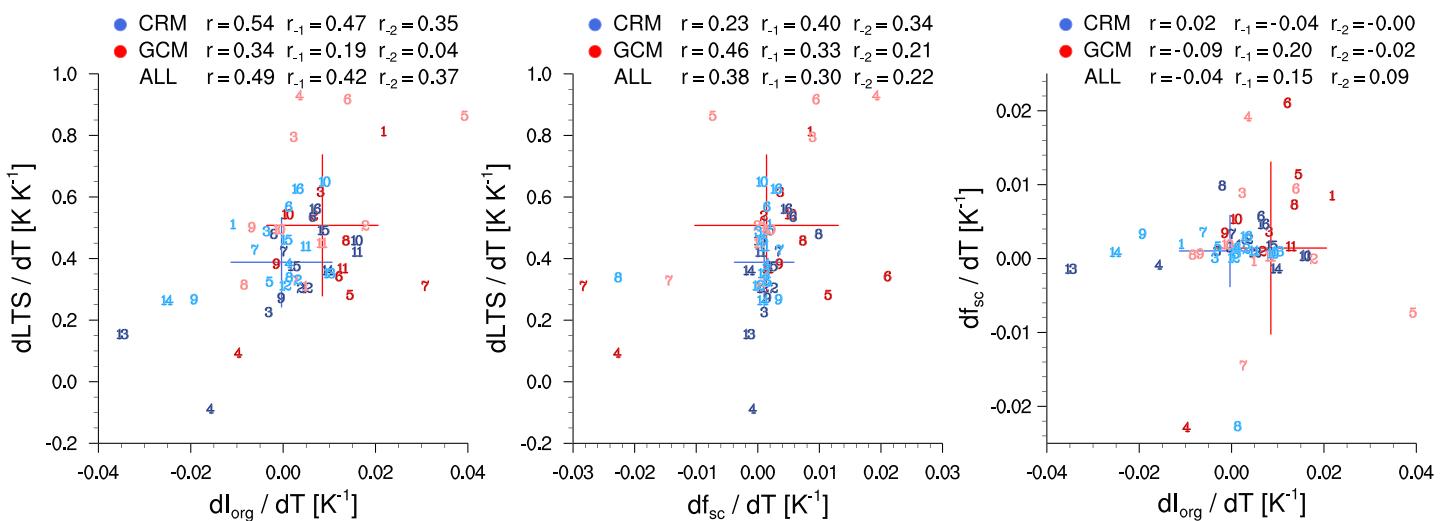
In the previous two subsections, we identified a significant correlation both between  $\lambda$  and changes in convective aggregation, where changes in the longwave fluxes are the main contributor, and between  $\lambda$  and changes in shallow cloud fraction, where changes in the shortwave fluxes are the main contributor. This indicates that both of these metrics are relatively independent of each other, which implies that more intermodel variance in  $\lambda$  can be explained when combining both metrics.

In principle, changes in shallow cloud fraction could be related to changes in convective aggregation via their respective relationships with lower-tropospheric stability (LTS): high LTS is associated with both higher shallow cloud fraction (e.g., Bony et al., 2020; Drótos et al., 2020) and increased convective aggregation (Wing et al., 2020). The explanation is that aggregated convection is triggered at a higher boundary layer moist static energy, leading temperatures to follow a warmer moist adiabat and thus be more stable. Across the RCEMIP models, changes in LTS, defined as the potential temperature difference between 700 and 1,000 hPa, correlate with  $\lambda$  (Figure 7). In GCMs, this correlation is more robust with respect to shortwave fluxes, in particular after excluding the model GEOS (Figure 7), and, in line with this, changes in LTS correlate more robustly with changes in shallow cloud fraction than in CRMs (Figure 8). In CRMs, in contrast, the correlation of changes in LTS with  $\lambda$  is mainly with respect to longwave fluxes (Figure 7), and consistent with this, changes in LTS correlate more strongly with changes in  $I_{\text{org}}$  than in GCMs (Figure 8). However, the correlation between changes in LTS and changes in  $I_{\text{org}}$  is weak in GCMs, and the correlation between changes in LTS and changes in shallow cloud fraction is weak in CRMs. This explains why, despite the relationship between each of  $I_{\text{org}}$  and shallow cloud fraction with LTS, there is almost no correlation between changes in  $I_{\text{org}}$  and changes in shallow cloud fraction (Figure 8).

As changes in  $I_{\text{org}}$  are not correlated with changes in shallow cloud fraction ( $f_{\text{sc}}$ ), we may be able to explain much more variance in the climate feedback parameter by combining the two metrics. We use a multivariate linear regression to find the partial regression coefficients  $b_1$  and  $b_2$  of the combined index:

$$\beta = b_1 \frac{dI_{\text{org}}}{dT} + b_2 \frac{df_{\text{sc}}}{dT}, \quad (3)$$

separately for the GCMs and for the CRMs. Figure 9 shows that the correlation of the combined index  $\beta$  with  $\lambda$  is indeed very high,  $r = 0.91$  for GCMs and  $r = 0.85$  for CRMs. Even when only half of the GCMs and CRMs are used to calculate  $b_1$  and  $b_2$ , and the correlation between  $\beta$  and  $\lambda$  is calculated for the other half of models, the correlation still remains larger than 0.8, both for GCMs and CRMs (not shown). When calculating the correlation between  $\beta$  and  $\lambda$  for the sum of all GCMs and CRMs (but still using separate



**Figure 8.** Temperature dependency of lower-tropospheric stability and organization index (left), lower-tropospheric stability and shallow cloud fraction (middle), and shallow cloud fraction and organization index (right), for GCMs and CRMs, numbered according to Figure 3. Darker colors show change between the two simulations at colder SST (300–295 K), and lighter colors at warmer SST (305–300 K). Horizontal and vertical lines illustrate the  $\pm 1$  standard deviation in each dimension for each model type, and their intersection is the mean value. Correlation coefficients ( $r$ ) are defined like in Figure 4.

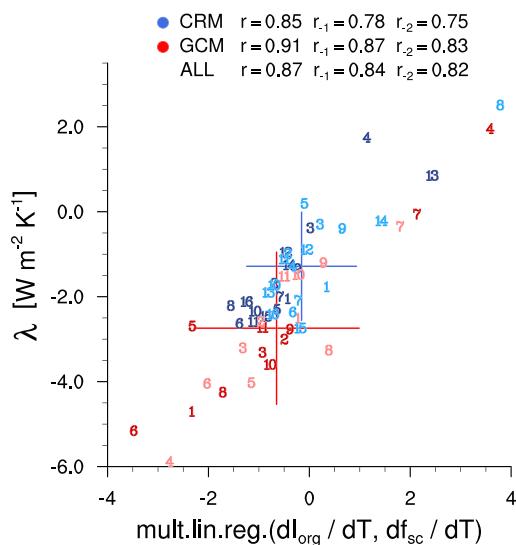
$\beta$  coefficients for the GCMs and CRMs), the correlation is  $r = 0.87$  (Figure 9). However, if an additional parameter  $b_0$  for the vertical offset is allowed in the multivariate linear regression, then  $r = 0.91$  for the sum of GCMs and CRMs (not shown).

When considered separately, changes in shallow cloud fraction explain 43% of the variance in  $\lambda$  across CRMs, and 70% across GCMs, while changes in  $I_{org}$  explain 31% of the variance in  $\lambda$  across CRMs but only 8% across GCMs. Thus, it is not surprising that the partial regression coefficients are different as well:  $b_1 = -63.38$  and  $b_2 = -170.52$  for the CRMs, and  $b_1 = -54.11$  and  $b_2 = -134.05$  for the GCMs. The standardized regression coefficients  $B_1$  and  $B_2$  in units of standard deviation follow after multiplying by the standard deviation of each index and dividing by the standard deviation in  $\lambda$ . For the CRMs,  $B_1 = -0.54$  and  $B_2 = -0.65$ , and for the GCMs,  $B_1 = -0.36$  and  $B_2 = -0.87$ . This means that when shallow cloud fraction, for example, increases by 1 standard deviation in GCMs,  $\lambda$  decreases by 0.87 standard deviations. Therefore, since changes in shallow cloud fraction explain more variance in  $\lambda$  than changes in  $I_{org}$ , they get more weight in the regression. However, since changes in  $I_{org}$  explain *relatively* more variance across the CRMs than across the GCMs,  $I_{org}$  gets *relatively* more weight in the regression for CRMs. As an aside, results using the  $I_{org}$  calculated with zero connectivity for GCMs are qualitatively similar (Figure S6), but with a smaller relative contribution of  $I_{org}$  to the regression.

In summary, the intermodel spread in climate feedback parameter in the large-domain RCEMIP simulations may be explained by the combination of changes in convective aggregation and changes in shallow cloud fraction with warming. The combined index  $\beta$ , calculated separately for GCMs and CRMs, explains 83% of the variance ( $r^2$ ) of  $\lambda$  across GCMs and 73% across CRMs. This impressively high amount of explained variance underlines the utility of idealized frameworks like RCE, where it is easier to explain the behavior than in some other studies that analyzed variance in the radiative budget. For example, combined variations in  $I_{org}$  and LTS explain  $\sim 60\%$  of the observed interannual variance of the tropical radiation budget (Bony et al., 2020) and combined variations in small- and large-scale lower-mid-tropospheric mixing explain almost 50% of the variance in climate sensitivity across 43 climate models (Sherwood et al., 2014).

#### 4. How Does Convective Aggregation Affect Longwave Radiation?

The previous section has shown that an increase of convective aggregation with warming leads to a low climate sensitivity, mostly because of an increase of OLR. However, the question of what causes the negative correlation between changes in convective aggregation and changes in TOA longwave radiation ( $\lambda_{LW}$ ) remains unanswered. We have shown that this correlation is mostly captured by clear-sky processes and



**Figure 9.** Climate feedback parameter versus multivariate linear regression of change of organization index and change of shallow cloud fraction with warming, for GCMs and CRMs, numbered according to Figure 3. Darker colors show change between the two simulations at colder SST (300–295 K), and lighter colors at warmer SST (305–300 K). Horizontal and vertical lines illustrate the  $\pm 1$  standard deviation in each dimension for each model type, and their intersection is the mean value. Correlation coefficients ( $r$ ) are defined like in Figure 4.

confirming that both metrics are indeed responding to changes in convective aggregation (Figures 11 and S10). When considering the average values of the metrics, rather than their change, subsidence fraction significantly correlates both with the average value and with the temperature dependency of longwave radiation in CRMs, while in GCMs, MTH correlates very strongly with the average value of longwave radiation ( $r = 0.89$ ; Figure S11).

In summary, the drying of subsidence regions has the strongest contribution to the increase in OLR when convective aggregation increases, while the expansion of subsidence regions is of secondary importance. We hypothesize that models with enhanced OLR have drier subsidence regions (compared to the other models) either in response to ( $b_1$ ) higher subsidence velocities or ( $b_2$ ) larger individual patches of subsidence. In contrast to the above discussion, we do not investigate subsidence fraction here, but rather the relation between the size of *individual* patches of subsidence and dryness.

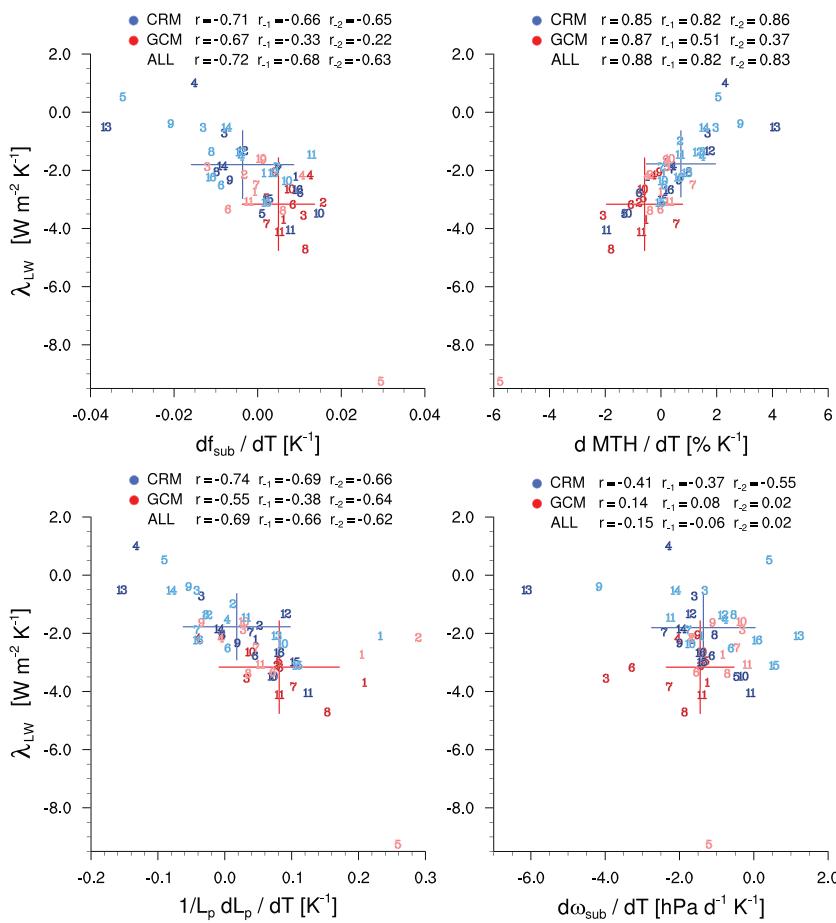
To address this hypothesis, we introduce two new metrics. Subsidence velocity,  $\omega_{\text{sub}}$ , is defined as the daily-mean large-scale vertical velocity at 500 hPa (downward defined as positive), using the same averaging as in the calculation of subsidence fraction. The second metric is the non-convective area patch length,  $L_p$ . To diagnose  $L_p$ , the distance between every non-convective grid point in space and time and the closest grid point that is part of a convective region is calculated, where convective regions are defined according to the threshold  $R_{\text{LW},\text{up}} < 173 \text{ W m}^{-2}$ . The mean distance to the closest convective grid point then serves as an estimate of the size of individual subsidence regions. The analysis is limited to one hour per day (every 24th output time step) for computational efficiency. When looking at the temperature dependency of these two metrics, and scaling the change in  $L_p$  by its value at the lower SST, an increase of OLR with warming is significantly correlated with a widening of individual non-convective regions, but is not significantly correlated with increasing subsidence velocities, in particular not in GCMs (Figures 10 and 11).

The schematic in Figure 11 provides a summary of the correlations between all discussed metrics and illustrates that many of the metrics co-vary. Overall, changes in  $\omega_{\text{sub}}$  with warming show a weaker correlation with the other metrics (changes in  $I_{\text{org}}$  and changes in MTH) than changes in  $L_p$  with warming, particularly for GCMs. It is not surprising that the correlation between changes in  $L_p$  and changes in  $I_{\text{org}}$  is significant because  $L_p$  could also be used as a metric to measure convective organization. Note that unlike  $I_{\text{org}}$ ,  $L_p$

therefore hypothesize that enhanced OLR under conditions of enhanced convective aggregation is due to (a) larger regions covered by subsidence or (b) drier subsidence regions in the more aggregated state.

To investigate the contributions of subsidence region size and dryness, we examine subsidence fraction and mid-tropospheric humidity (MTH). Subsidence fraction, which was introduced as an alternative metric for convective aggregation in Section 3.1, is the daily-mean area fraction of the domain covered by large-scale (CRMs averaged in space over  $100 \times 100 \text{ km}^2$  blocks, GCRMs over  $1^\circ \times 1^\circ$  blocks, GCMs on their native grid) subsidence at 500 hPa, and MTH is defined as the domain-mean mass-weighted vertical average of relative humidity between 300 and 700 hPa. Figure S8 shows that changes in domain-mean MTH with temperature are similar to changes in MTH over the subsidence region. We thus use the domain-mean MTH as a proxy for MTH in the subsidence region because the latter requires calculation from 3D output, which is available only for 60% of the models and for only the last 25 days of each simulation, whereas the domain-mean MTH is available for the entire simulation period for all models.

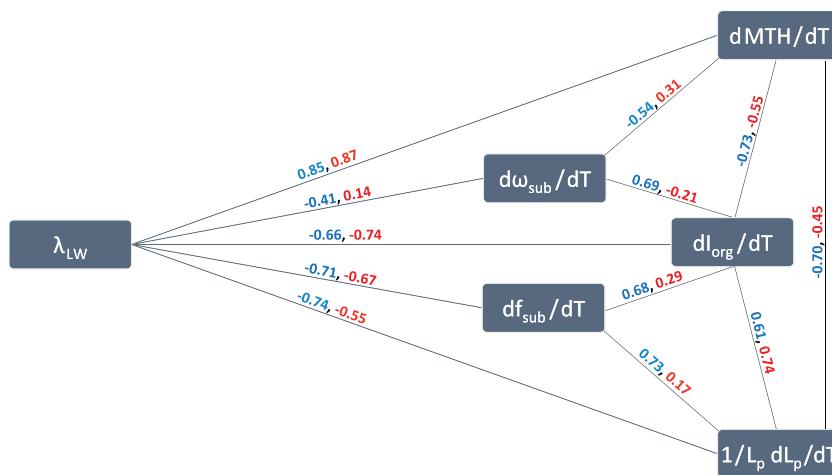
We find that both the expansion of subsidence regions and mid-tropospheric drying show a significant correlation with  $\lambda_{\text{LW}}$ , but MTH dominates, with correlation coefficients of at least  $r = 0.85$  (Figures 10 and 11). Changes in subsidence fraction and MTH are also strongly correlated with each other, which means that the correlation with  $\lambda_{\text{LW}}$  does not further increase when combining both indices in a multivariate linear regression (Figure S9). Changes in both metrics do not only correlate with  $\lambda_{\text{LW}}$  but also to a slightly smaller degree with changes in  $I_{\text{org}}$ ,



**Figure 10.** Climate feedback parameter based on the TOA longwave fluxes versus temperature dependency of subsidence fraction (top left), mid-tropospheric humidity (top right), scaled non-convective area patch length (bottom left), and subsidence velocity (bottom right), for GCMs and CRMs, numbered according to Figure 3. Darker colors show change between the two simulations at colder SST (300–295 K), and lighter colors at warmer SST (305–300 K). Horizontal and vertical lines illustrate the  $\pm 1$  standard deviation in each dimension for each model type, and their intersection is the mean value. Correlation coefficients ( $r$ ) are defined like in Figure 4.

focuses on the non-convective region and is mainly affected by large-scale organization, while  $I_{org}$  focuses on convecting areas and integrates over all length scales. However, the most important conclusion from Figure 11 is that  $\lambda_{LW}$  primarily depends on changes in MTH, in response to changes in convective aggregation. Changes in MTH significantly correlate with changes in  $L_p$  (and also with changes in subsidence fraction). MTH depends on the patch size because when  $L_p$  is larger, a larger part of the subsidence region is isolated from convective moistening through detrainment, thus reducing MTH in the subsidence region. However, changes in MTH are only weakly correlated with changes in subsidence velocity in CRMs, and uncorrelated in GCMs. The latter is presumably because the exchange of air between convective and non-convective regions happens more slowly in RCE simulations with GCMs than with CRMs, which gives even weak subsidence enough time to dry the non-convective regions. Thus, in RCE simulations with GCMs, subsidence velocity is not a limiting factor of the dryness of the non-convective region. As an aside, the role of vertical velocity is different when using the tropical circulation index  $I$ , defined as the mean upward velocity minus the mean downward velocity at 500 hPa (Bony, Bellon, et al., 2013), as a measure of vertical velocity. Changes in  $I$  correlate well with  $\lambda_{LW}$  (Figure S12), but this is misleading because, across the RCExMIP simulations, differences in the tropical circulation index are dominated by differences in the convective region (i.e., by changes in the mean upward velocity).

The average values (rather than tendencies) of non-convective patch size and subsidence velocity have only weak correlations with the average value of longwave radiation and its tendency (Figures S13 and S14). One reason why temperature dependencies, as in Figure 11, are more informative than average values is



**Figure 11.** Schematic of how the climate feedback parameter based on the TOA longwave fluxes correlates with the change of different metrics with warming, and how these metrics correlate with each other. The metrics are organization index, subsidence fraction, mid-tropospheric humidity, subsidence velocity, and scaled non-convective area patch length. The correlation coefficients, which are based on Figures 10 and S10, are given in blue for the CRMs and in red for the GCMs.

that base state differences between the different RCEMIP simulations, which often are linked to specific problems in parameterizations, play less of a role when only the change in a metric is considered. However, unrelated to these issues, the average values of MTH stand out as most important contributors to intermodel spread in longwave radiation (Figure S14), and changes in MTH stand out as most important contributors to intermodel spread of changes in longwave radiation ( $\lambda_{\text{LW}}$ , Figure 11).

## 5. Summary and Conclusion

In this study, we analyzed the variability of climate sensitivities across 31 models, run at three fixed SSTs (295, 300, and 305 K) that participated in the RCEMIP (Wing et al., 2018, 2020). Estimates of the climate feedback parameter ( $\lambda$ ) have an extremely large range across the RCEMIP models, from  $-6$  to  $3 \text{ W m}^{-2} \text{ K}^{-1}$ . The intermodel spread in  $\lambda$  is particularly large in simulations configured on a domain that is large enough to allow convection to self-aggregate; that is, CRMs run on an elongated channel and GCMs/GCRMs run on a global domain. High climate sensitivities are primarily associated with a decrease of shallow cloud fraction with increasing temperature, and secondarily with a decrease of convective aggregation with increasing temperature, and vice versa. The average amount of convective aggregation, on the other hand, plays no significant role in modulating climate sensitivity. Because changes in shallow cloud fraction and in convective aggregation are almost independent from each other, they can together explain, when combined with a multivariate linear regression approach, more than 80% of the total variance in climate sensitivity across the GCMs and more than 70% across the CRMs. An increase in shallow cloud fraction enhances the planetary albedo and thus affects the shortwave radiation imbalance, while an increase in convective aggregation primarily enhances clear-sky OLR. This finding has similarities to the iris effect, but with the difference that in most other studies changes in high-level cloud fraction play an important role in the iris effect (e.g., Mauritsen & Stevens, 2015), while in RCEMIP, changes in high-level cloud fraction are of minor importance for the intermodel spread in climate sensitivity. Instead, OLR increases when convection aggregates primarily because of mid-tropospheric drying and secondarily because the subsidence regions expand, both individually and in the aggregate. The mid-tropospheric drying is only weakly related to higher subsidence velocities in CRMs and not at all in GCMs. This indicates that in RCE, the exchange of air between convective and non-convective regions happens slowly enough to even give weak subsidence enough time to dry the subsidence region. Another possible mechanism that can cause drier subsidence regions, which was not analyzed in this study, is an increase of precipitation efficiency when convection aggregates, as discussed by (Lindzen et al., 2001).

The comparison of the mean characteristics of GCMs and CRMs with respect to climate sensitivity also offer some interesting conclusions. In GCMs,  $\lambda$  is significantly more negative than in large-domain CRMs, which

is almost entirely related to differences in the longwave radiation imbalance. About 60% of these differences can be associated with changes in the clear-sky radiative fluxes. In agreement with this, shallow cloud fraction has no overall temperature trend when averaged over the respective sets of GCM and CRM simulations. Deep cloud fraction decreases slightly more strongly with warming in GCMs than in CRMs, consistent with the stability iris effect (Bony et al., 2016), which relates the decrease of anvil cloud fraction with warming to increasing stability and a weakening overturning circulation. However, in our study this effect cannot explain the significantly more negative climate feedback parameter in GCMs than in CRMs because there is no indication that this mechanism works substantially differently in GCMs than in CRMs, and the net CRE of deep clouds is relatively small. Instead,  $\lambda$  is significantly more negative in GCMs than in CRMs mainly because, averaged over all CRMs, the temperature dependency of convective self-aggregation is close to zero, while convective self-aggregation increases with warming in most GCMs, causing a mid-tropospheric drying and enhanced OLR in subsidence regions. Interestingly, Retsch et al. (2019) also found that climate sensitivity is strongly linked to the temperature dependency of mid-tropospheric humidity in subsidence regions (using ICON aquaplanet simulations). However, in their simulations, mid-tropospheric humidity increases with warming when convection is parameterized, causing higher climate sensitivities compared to simulations with explicit convection.

Given that the temperature dependency of convective self-aggregation shows a large spread across the CRMs, it would be plausible that the offset between the average temperature dependency of convective self-aggregation in GCMs and in CRMs is due to shortcomings in some of the CRMs, for example, because shallow clouds in CRMs are poorly represented at kilometer-scale resolution. However, the climate feedback parameters in CRMs are supported by very similar results in the higher-resolution LES experiments. Hence, it is more likely that caveats in GCMs explain the GCM-CRM difference in the temperature dependency of convective self-aggregation. We speculate that the moisture-memory feedback, in which convection favors locations that have been pre-moistened by convection (Tompkins, 2001), explains some of the discrepancy between GCMs and CRMs. The moisture-memory feedback may become more important with warming because of an increase in the saturation deficit, which increases the efficiency with which updraft buoyancy is reduced through entrainment. In this regard, the important difference between GCMs and CRMs is that GCMs usually entrain the grid-cell mean air, from an area that spans  $\sim 100 \times 100 \text{ km}^2$ , while CRMs feel less of this effect because they entrain air much more locally, from within their moist shell (Becker et al., 2018). Even though there is also some evidence that convective parameterizations are not sensitive enough to environmental humidity due to too-low entrainment (e.g., Mapes & Neale, 2011; Sherwood et al., 2013), our results suggest that the increase in convective aggregation with warming in many GCMs may be an artifact of convective parameterization. If true, this implies that many GCMs underestimate climate sensitivity, not only in the RCE setup but also in less idealized model configurations.

This study is an important step toward using idealized frameworks like RCE in a hierarchy of models to constrain climate sensitivity. We have identified the temperature tendency of shallow cloud fraction and convective aggregation as key contributors to the intermodel spread in climate sensitivity across RCE simulations with GCMs, GCRMs, and large-domain CRMs. An important target for future work is the question of what processes and details in parameterizations are responsible for the different temperature dependencies of these key factors across the different models. Further work could focus on targeted parameter perturbation simulations, as planned for the possible second phase of RCEMIP. As an example, the suite of WRF-GCM simulations, while not included in our main analysis here, shows that the change in shallow cloud fraction with temperature explains most of the spread in  $\lambda$  across different convective parameterizations (Figure S15). This type of targeted physics comparison, in which it may be tractable to determine why a particular parameterization has a different change in clouds, ideally in combination with radiative kernel calculations of climate feedbacks, could be extended to other model setups. When comparing the different RCEMIP model types, temperature dependencies seem to be weaker in GCRMs than in CRMs, both with respect to climate sensitivity and to the discussed metrics, indicating that some of the extreme behavior that we find is setup dependent. Thus, it will be interesting to investigate whether the key contributors to intermodel spread that we identified are also key contributors to intermodel spread across other, less idealized intercomparison projects, such as CMIP6 (Eyring et al., 2016) or the DYAMOND project (Stevens et al., 2019).

## Acknowledgments

We thank the German Climate Computing Center (DKRZ) for hosting the RCEMIP data, which is publicly available at <https://hdl.handle.net/21.14101/d4bee8e-6996-453e-bbd1-ff53b6874c0e>. Secondary data and scripts used in the analysis are archived by the Max Planck Institute for Meteorology and can be obtained by contacting publications@mpimet.mpg.de. T.B. acknowledges support from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 820829 (CONSTRAIN Project), and A.A.W. acknowledges support from NSF grant 1830724. We thank all co-authors on the RCEMIP overview paper (Wing et al., 2020) for providing the RCEMIP simulations; we thank Hauke Schulz and Catherine Stauffer for their important contributions to defining the organization index; and we thank Bjorn Stevens, the internal reviewer Maria Rugenstein, and three external reviewers (Christopher Holloway and two anonymous reviewers) for their thoughtful and helpful comments to the manuscript.

## References

- Arrhenius, S. (1896). On the influence of carbonic acid in the air upon the temperature of the ground. *Philosophical Magazine Series 5*, 41, 237–276. <https://doi.org/10.1086/121158>
- Becker, T., Bretherton, C. S., Hohenegger, C., & Stevens, B. (2018). Estimating bulk entrainment with unaggregated and aggregated convection. *Geophysical Research Letters*, 45, 455–462. <https://doi.org/10.1002/2017gl076640>
- Becker, T., & Stevens, B. (2014). Climate and climate sensitivity to changing CO<sub>2</sub> on an idealized land planet. *Journal of Advances in Modeling Earth Systems*, 6, 1205–1223. <https://doi.org/10.1002/2014MS000369>
- Becker, T., Stevens, B., & Hohenegger, C. (2017). Imprint of the convective parameterization and sea-surface temperature on large-scale convective self-aggregation. *Journal of Advances in Modeling Earth Systems*, 9, 1488–1505. <https://doi.org/10.1002/2016MS000865>
- Bony, S., Bellon, G., Klocke, D., Sherwood, S., Fermepin, S., & Denvil, S. (2013). Robust direct effect of carbon dioxide on tropical circulation and regional precipitation. *Nature Geoscience*, 6(6), 447–451. <https://doi.org/10.1038/NGEO1799>
- Bony, S., & Dufresne, J.-L. (2005). Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophysical Research Letters*, 32, L20806. <https://doi.org/10.1029/2005GL023851>
- Bony, S., Semie, A., Kramer, R. J., Soden, B., Tompkins, A. M., & Emanuel, K. A. (2020). Observed modulation of the tropical radiation budget by deep convective organization and lower-tropospheric stability. *AGU Advances*, 1(3), e2019AV000155. <https://doi.org/10.1029/2019AV000155>
- Bony, S., Stevens, B., Coppin, D., Becker, T., Reed, K. A., Voigt, A., & Medeiros, B. (2016). Thermodynamic control of anvil cloud amount. *Proceedings of the National Academy of Sciences of the United States of America*, 113(32), 8927–8932. <https://doi.org/10.1073/pnas.1601472113>
- Bony, S., Stevens, B., Frierson, D. M. W., Jakob, C., Kageyama, M., Pincus, R., et al. (2015). Clouds, circulation and climate sensitivity. *Nature Geoscience*, 8(4), 261–268. <https://doi.org/10.1038/ngeo2398>
- Bony, S., Stevens, B., Held, I. H., Mitchell, J. F., Dufresne, J.-L., Emanuel, K. A., et al. (2013). Carbon dioxide and climate: Perspectives on a scientific assessment. In G. R. Asrar & J. W. Hurrell (Eds.), *Climate science for serving society* (pp. 391–413). Dordrecht: Springer. [https://doi.org/10.1007/978-94-007-6692-1\\_14](https://doi.org/10.1007/978-94-007-6692-1_14)
- Bretherton, C. S., Blossey, P. N., & Khairoutdinov, M. F. (2005). An energy-balance analysis of deep convective self-aggregation above uniform SST. *Journal of the Atmospheric Sciences*, 62(12), 4273–4292. <https://doi.org/10.1175/JAS3614.1>
- Cess, R. D., & Potter, G. L. (1988). A methodology for understanding and intercomparing atmospheric climate feedback processes in general circulation models. *Journal of Geophysical Research*, 93(D7), 8305–8314.
- Charney, J. G., Arakawa, A., Baker, D. J., Bolin, B., Dickinson, R. E., Goody, R. M., et al. (1979). *Carbon dioxide and climate: A scientific assessment*. Washington, DC: National Academies of Sciences Press.
- Cronin, T. W., & Wing, A. A. (2017). Clouds, circulation, and climate sensitivity in a radiative-convective equilibrium channel model. *Journal of Advances in Modeling Earth Systems*, 9, 2883–2905. <https://doi.org/10.1002/2017MS001111>
- Dines, W. H. (1917). The heat balance of the atmosphere. *Quarterly Journal of the Royal Meteorological Society*, 43(182), 151–158. <https://doi.org/10.1002/qj.49704318203>
- Drótó, G., Becker, T., Mauritsen, T., & Stevens, B. (2020). Global variability in radiative-convective equilibrium with a slab ocean under a wide range of CO<sub>2</sub> concentrations. *Tellus A: Dynamic Meteorology and Oceanography*, 72(1), 1–19. <https://doi.org/10.1080/16000870.2019.1699387>
- 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(5), 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>
- Fiedler, S., Crueger, T., DAgostino, R., Peters, K., Becker, T., Leutwyler, D., et al. (2020). Simulated tropical precipitation assessed across three major phases of the Coupled Model Intercomparison Project (CMIP). *Monthly Weather Review*, 148(9), 3653–3680. <https://doi.org/10.1175/MWR-D-19-0404.1>
- Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., et al. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, 31, L03205. <https://doi.org/10.1029/2003GL018747>
- Held, I. M. (2005). The gap between simulation and understanding in climate modeling. *Bulletin of the American Meteorological Society*, 86(11), 1609–1614. <https://doi.org/10.1175/BAMS-86-11-1609>
- Held, I. M., Hemler, R. S., & Ramaswamy, V. (1993). Radiative-convective equilibrium with explicit two-dimensional moist convection. *Journal of the Atmospheric Sciences*, 50, 3909–3927. [https://doi.org/10.1175/1520-0469\(1993\)050h3909:RCEWETi2.0.CO;2](https://doi.org/10.1175/1520-0469(1993)050h3909:RCEWETi2.0.CO;2)
- Held, I. M., & Soden, B. J. (2000). Water vapor feedback and global warming. *Annual Review of Energy and the Environment*, 25(1), 441–475. <https://doi.org/10.1146/annurev.energy.25.1.441>
- Held, I. M., Zhao, M., & Wyman, B. (2007). Dynamic radiative-convective equilibria using GCM column physics. *Journal of the Atmospheric Sciences*, 64(1), 228–238. <https://doi.org/10.1175/JAS3825.1>
- Hope, C. (2015). The \$10 trillion value of better information about the transient climate response. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373(2054), 20140429. <https://doi.org/10.1098/rsta.2014.0429>
- IPCC (2018). Summary for Policymakers. In V. Masson-Delmotte et al. (Eds.), *Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty* (pp. 1–32). Geneva, Switzerland: World Meteorological Organization.
- Jeevanjee, N., Hassanzadeh, P., Hill, S., & Sheshadri, A. (2017). A perspective on climate model hierarchies. *Journal of Advances in Modeling Earth Systems*, 9, 1760–1771. <https://doi.org/10.1002/2017MS001038>
- Knutti, R., Rugenstein, M. A. A., & Hegerl, G. C. (2017). Beyond equilibrium climate sensitivity. *Nature Geoscience*, 10(10), 727–736. <https://doi.org/10.1038/ngeo3017>
- Lebsack, M. D., LEcuyer, T. S., & Pincus, R. (2017). An observational view of relationships between moisture aggregation, cloud, and radiative heating profiles. *Surveys in Geophysics*, 38(6), 1237–1254. <https://doi.org/10.1007/s10712-017-9443-1>
- Li, C., von Storch, J.-S., & Marotzke, J. (2013). Deep-ocean heat uptake and equilibrium climate response. *Climate Dynamics*, 40(5), 1071–1086. <https://doi.org/10.1007/s00382-012-1350-z>
- Lindzen, R. S., Chou, M. D., & Hou, A. Y. (2001). Does the earth have an adaptive infrared iris? *Bulletin of the American Meteorological Society*, 82, 417–432. [https://doi.org/10.1175/1520-0477\(2001\)082h0417:DTEHAAi2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082h0417:DTEHAAi2.3.CO;2)
- Manabe, S., & Strickler, R. F. (1964). Thermal equilibrium of the atmosphere with a convective adjustment. *Journal of the Atmospheric Sciences*, 21(4), 361–385. [https://doi.org/10.1175/1520-0469\(1964\)021h0361:TEOTAWi2.0.CO;2](https://doi.org/10.1175/1520-0469(1964)021h0361:TEOTAWi2.0.CO;2)
- Manabe, S., & Wetherald, R. T. (1967). Thermal equilibrium of the atmosphere with a given distribution of relative humidity. *Journal of the Atmospheric Sciences*, 24(3), 241–259. [https://doi.org/10.1175/1520-0469\(1967\)024h0241:TEOTAWi2.0.CO;2](https://doi.org/10.1175/1520-0469(1967)024h0241:TEOTAWi2.0.CO;2)

- Mapes, B., & Neale, R. (2011). Parameterizing convective organization to escape the entrainment dilemma. *Journal of Advances in Modeling Earth Systems*, 3, M06004. <https://doi.org/10.1029/2011MS000042>
- Mauritsen, T., & Stevens, B. (2015). Missing iris effect as a possible cause of muted hydrological change and high climate sensitivity in models. *Nature Geoscience*, 8(5), 346–351. <https://doi.org/10.1038/ngeo2414>
- Medeiros, B., Stevens, B., Held, I. M., Zhao, M., Williamson, D. L., Olson, J. G., & Bretherton, C. S. (2008). Aquaplanets, climate sensitivity, and low clouds. *Journal of Climate*, 21(19), 4974–4991. <https://doi.org/10.1175/2008JCLI1995.1>
- Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J. F. B., et al. (2007). The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bulletin of the American Meteorological Society*, 88(9), 1383–1394. <https://doi.org/10.1175/BAMS-88-9-1383>
- Murphy, J. M. (1995). Transient response of the Hadley Centre coupled ocean-atmosphere model to increasing carbon dioxide. Part III: Analysis of global-mean response using simple models. *Journal of Climate*, 8(3), 496–514. [https://doi.org/10.1175/1520-0442\(1995\)008<0496:TROTHC>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<0496:TROTHC>2.0.CO;2)
- Nakajima, K., & Matsuno, T. (1988). Numerical experiments concerning the origin of cloud clusters in the tropical atmosphere. *Journal of the Meteorological Society of Japan Series II*, 66(2), 309–329. <https://doi.org/10.2151/jmsj1965.66.2309>
- Popke, D., Stevens, B., & Voigt, A. (2013). Climate and climate change in a radiative convective equilibrium version of ECHAM6. *Journal of Advances in Modeling Earth Systems*, 5, 1–14. <https://doi.org/10.1029/2012MS000191>
- Ramanathan, V., & Coakley, J. A. (1978). Climate modeling through radiative-convective models. *Reviews of Geophysics and Space Physics*, 16(4), 465–489. <https://doi.org/10.1029/RG016i004p00465>
- Reed, K. A., & Medeiros, B. (2016). A reduced complexity framework to bridge the gap between AGCMs and cloud-resolving models. *Geophysical Research Letters*, 43, 860–866. <https://doi.org/10.1002/2015GL066713>
- Retsch, M. H., Mauritsen, T., & Hohenegger, C. (2019). Climate change feedbacks in aquaplanet experiments with explicit and parametrized convection for horizontal resolutions of 2,525 up to 5 km. *Journal of Advances in Modeling Earth Systems*, 11, 2070–2088. <https://doi.org/10.1029/2019MS001677>
- Rugenstein, M., Bloch-Johnson, J., Abe-Ouchi, A., Andrews, T., Beyerle, U., Cao, L., et al. (2019). LongRunMIP: Motivation and design for a large collection of millennial-length AOGCM simulations. *Bulletin of the American Meteorological Society*, 100(12), 2551–2570. <https://doi.org/10.1175/BAMS-D-19-0068.1>
- Satoh, M., Stevens, B., Judt, F., Khairoutdinov, M., Lin, S.-J., Putman, W. M., & Düben, P. (2019). Global cloud-resolving models. *Current Climate Change Reports*, 5, 172–184. <https://doi.org/10.1007/s40641-019-00131-0>
- Senior, C. A., & Mitchell, J. F. B. (2000). The time-dependence of climate sensitivity. *Geophysical Research Letters*, 27(17), 2685–2688. <https://doi.org/10.1029/2000GL011373>
- Sherwood, S. C., Bony, S., & Dufresne, J.-L. (2014). Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature*, 505(7481), 37–42. <https://doi.org/10.1038/nature12829>
- Sherwood, S. C., Hernández-Deckers, D., Colin, M., & Robinson, F. (2013). Slippery thermals and the cumulus entrainment paradox. *Journal of the Atmospheric Sciences*, 70(8), 2426–2442. <https://doi.org/10.1175/JAS-D-12-0220.1>
- Soden, B. J., Broccoli, A. J., & Hemler, R. S. (2004). On the use of cloud forcing to estimate cloud feedback. *Journal of Climate*, 17(19), 3661–3665. [https://doi.org/10.1175/1520-0442\(2004\)017h3661:OTUOCF2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017h3661:OTUOCF2.0.CO;2)
- Stein, T. H. M., Holloway, C. E., Tobin, I., & Bony, S. (2017). Observed relationships between cloud vertical structure and convective aggregation over tropical ocean. *Journal of Climate*, 30(6), 2187–2207. <https://doi.org/10.1175/JCLI-D-16-0125.1>
- Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., et al. (2019). DYAMOND: The DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains. *Progress in Earth and Planetary Science*, 6(1), 61. <https://doi.org/10.1186/s40645-019-0304-z>
- Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4), 485–498. <https://doi.org/10.1175/BAMS-D-11-00094.1>
- Tobin, I., Bony, S., Holloway, C. E., Grandpeix, J.-Y., Sèze, G., Coppin, D., et al. (2013). Does convective aggregation need to be represented in cumulus parameterizations? *Journal of Advances in Modeling Earth Systems*, 5, 692–703. <https://doi.org/10.1002/jame.20047>
- Tobin, I., Bony, S., & Roca, R. (2012). Observational evidence for relationships between the degree of aggregation of deep convection, water vapor, surface fluxes, and radiation. *Journal of Climate*, 25, 6885–6904. <https://doi.org/10.1175/JCLI-D-11-00258.1>
- Tompkins, A. M. (2001). Organization of tropical convection in low vertical wind shears: The role of water vapor. *Journal of the Atmospheric Sciences*, 58(6), 529–545. [https://doi.org/10.1175/1520-0469\(2001\)058<0529:OOTCIL>2.0.CO;2](https://doi.org/10.1175/1520-0469(2001)058<0529:OOTCIL>2.0.CO;2)
- Tompkins, A. M., & Craig, G. C. (1998). Radiative-convective equilibrium in a three-dimensional cloud-ensemble model. *Quarterly Journal of the Royal Meteorological Society*, 124(550), 2073–2097. <https://doi.org/10.1002/qj.49712455013>
- Tompkins, A. M., & Semie, A. G. (2017). Organization of tropical convection in low vertical wind shears: Role of updraft entrainment. *Journal of Advances in Modeling Earth Systems*, 9, 1046–1068. <https://doi.org/10.1002/2016MS000802>
- Vial, J., Dufresne, J.-L., & Bony, S. (2013). On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates. *Climate Dynamics*, 41(11–12), 3339–3362. <https://doi.org/10.1007/s00382-013-1725-9>
- Webb, M. J., Lambert, F. H., & Gregory, J. M. (2013). Origins of differences in climate sensitivity, forcing and feedback in climate models. *Climate Dynamics*, 40(3), 677–707. <https://doi.org/10.1007/s00382-012-1336-x>
- Wing, A. A. (2019). Self-aggregation of deep convection and its implications for climate. *Current Climate Change Reports*, 5(1), 1–11. <https://doi.org/10.1007/s40641-019-00120-3>
- Wing, A. A., Reed, K. A., Satoh, M., Stevens, B., Bony, S., & Ohno, T. (2018). Radiative-convective equilibrium model intercomparison project. *Geoscientific Model Development*, 11(2), 793–813. <https://doi.org/10.5194/gmd-11-793-2018>
- Wing, A. A., Stauffer, C. L., Becker, T., Reed, K. A., Ahn, M.-S., Arnold, N. P., et al. (2020). Clouds and convective self-aggregation in a multi-model ensemble of radiative-convective equilibrium simulations. *Journal of Advances in Modeling Earth Systems*, 12, e2020MS002138. <https://doi.org/10.1029/2020MS002138>
- Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., et al. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters*, 47, e2019GL085782. <https://doi.org/10.1029/2019GL085782>