Response to reviewer 2
We thank the reviewer for helping us improve our manuscript. Our responses are provided below in blue.

Specific comments:
Page 3, line 6: "2-dimensional swath" is a bit vague (x-y or x-z)?
The CALIPSO lidar laser produces a beam diameter of ~70 m at the surface every 333m along a polar orbit with an inclination of 98.2°, which crosses the equator at approximately 0130 and1330 local time. The formulation used in the manuscript was confusing, so we changed the sentence to: *" However, the narrow swath of the lidar – a beam diameter of 70 m every 333 m along-track – produces a much smaller sample of clouds than passive instruments. "*

Page 4, line 2: By using the nighttime only are you biasing the sampling to a particular part of the diurnal cycle? My understanding is that the GCMs sample all times of the day?
This reference examines the diurnal cycle of marine cloud feedback which might be of interest. It also examines the diurnal features of low marine clouds in some CMIP5 models. Webb, M.J., Lock, A.P., Bodas-Salcedo, A. et al. Clim Dyn (2015) 44: 1419. https://doi.org/10.1007/s00382-014-2234-1
It is true that the diurnal cycle affects LCC, which is maximum in the morning. However, in their supplementary material (text S1), Cesana and Waliser (2016) showed that the difference between all time and gridbox total-column cloud fraction average (cltcalipso) and the same quantity sampled along the CALIPSO orbit is negligible (less than 1%, absolute value) in a sample of four models. As a result, we consider it unlikely that using all times of the day rather than only 01h30 local time would significantly affect the ΔLCC/ΔSST relationship in the models and explain large differences between models and observations. In addition, Webb et al. (2015) found that temporal sub-sampling was not relevant to explaining the multi-model spread in cloud feedbacks. We now acknowledge this at the end of section 2.2: *Although the diurnal cycle of LCC is not fully represented in the observations (sampled at 0130 and 1330 local time), the total-column cloud fraction mean from the lidar simulator is not substantially different from that extracted along the CALIPSO footprint (<1% absolute difference; Cesana and Waliser, 2016) and effects on the strength of the cloud feedback have been found unimportant to understanding multi-model spread in overall cloud feedback (Webb et al., 2015).*
.

Page 4, line 23: Radiative balance during which period?
We added the missing information to the manuscript: *"Over 2007-2015, this version has a small positive radiative imbalance (0.29 W m^{-2}) of a few tenths of a W m^{-2} less than that estimated for the real world in the early 21st Century (0.6 W m^{-2})."*.

Section 2.2: Some time has elapsed since the manuscript was submitted. Is there a GISS-E3 paper available that you might be able to reference that contains most of this information?
Unfortunately, no new paper is available as GISS-E3 is still under development, which is why we provide a somewhat detailed description here.

Section 2.2: Is the required model output available for a longer period of time? It is mentioned a few times in the manuscript that a short time-period is used (a decade). If there is additional model output then is should be possible to indicate how well the 10 years period chosen represents a longer dataset (at least in the models).

We did not mention any of our sensitivity analysis on the time period chosen in this version of the manuscript. Because the COSP simulator is run offline in GISS-E3 it would require not insubstantial effort to analyze another 20 years. However, using the other models, we find (see Figure R1 below) that choosing a different time period (either the full AMIP period or the last 18 or 9 years) may slightly influence the ΔLCC/ΔSST by a few tenths of percent per K (in absolute value) in a subset of two unconstrained and two constrained models, which is far less than the model-to-observation difference, and makes almost no difference for the ΔCRE/ΔSST relationship. This result is now mentioned in the manuscript and in the supporting information (Fig. S1): "*Using a shorter or longer time period may affect the ΔLCC/ΔSST relationship by a few tenths of percent per K (absolute value, Fig. S1), yet it remains much smaller than the models' bias.*"

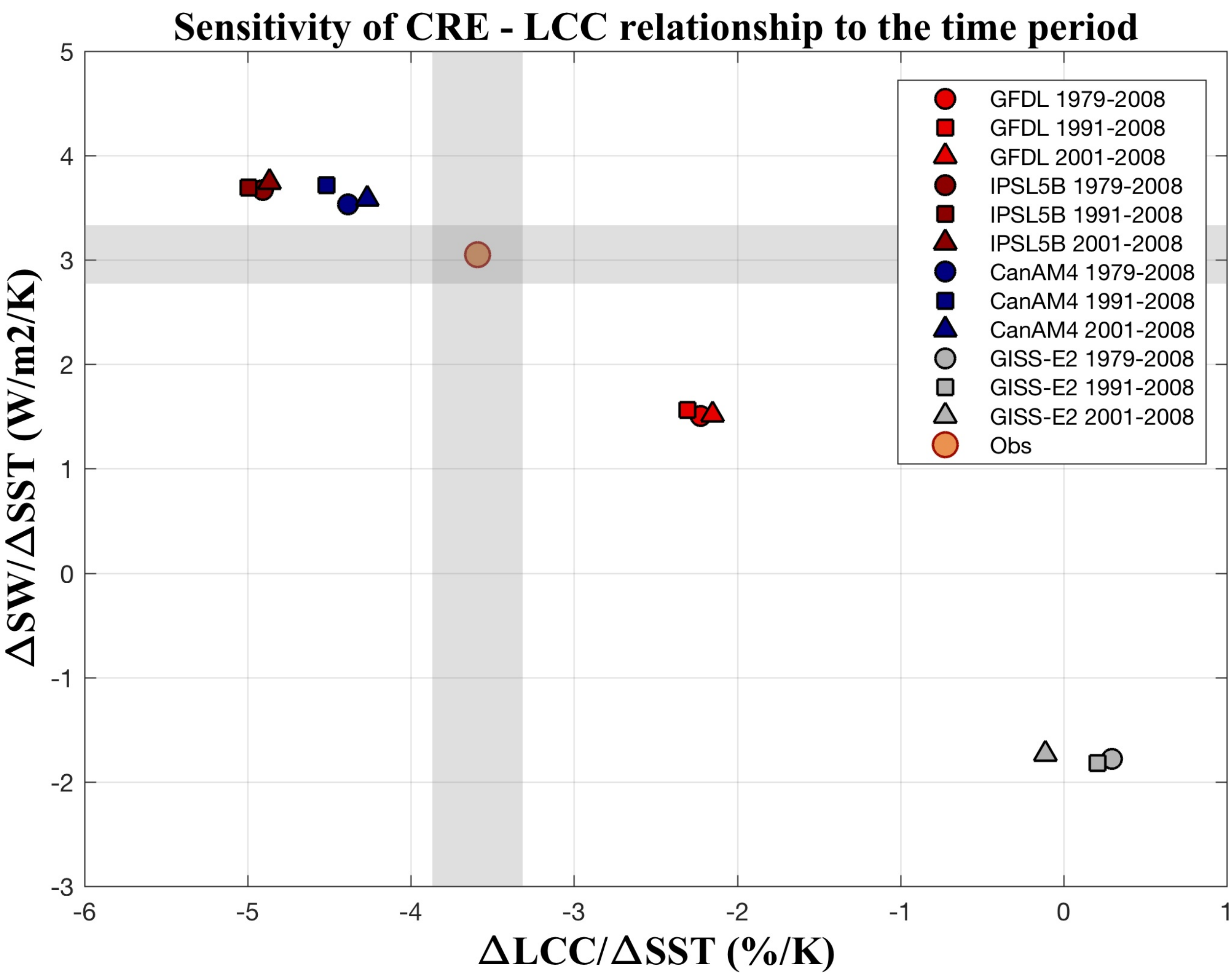

Figure R1: Relationship between ΔLCC/ΔSST (x-axis, % K-1) and ΔCRE/ΔSST (y-axis, W m-2 K-1) for SW radiation as in Fig. 4a of the manuscript. Here we study the sensitivity of that relationship to the chosen time period in four models. The results are shown for three periods of time: the full AMIP period (1979-2008, circles), the last 18 years as used in the manuscript (squares) and the last 9 years as used in GISS-E3 (triangles).

Page 6, line 17: You might want to point out that the low-level cloud fraction is LCC referred to throughout the manuscript.
Done.

Page 8, line 32: "that is a too shallow PBL" -> "too shallow PBL"
Done.

Page 9, line 1: "strong masking effect". I thought the 500 hPa omega filtered out overlaying high clouds? In Figure S2 you show the cloud fraction profiles below 5 km. Could you not extend it vertically?
The reviewer is correct that we chose this omega500 threshold to reduce high-cloud– and it works perfectly in the observation as shown in Fig. S1. However, some models have a large high-cloud bias that may generate more masking effect than in the observations. We vertically extended Fig. S2 as requested by the reviewer to show this.

Page 9, line 9: But in the multimodel mean the response is similar to the observed?
This is correct, we modified the manuscript accordingly: *"As for the mean cloud profiles, ... while the multimodel mean captures the observed shape of $\Delta CF/\Delta SST$ to some extent"*.

Page 12, line 25: "2D" -> cloud-top properties?
We acknowledge that this was ambiguous and changed it as suggested by the reviewer.

Page 13, line 10: Do you get better results with EIS or other variants of LTS? You seem to use EIS later in the analysis (Figure 5).
In GISS-E3, both the EIS and the LTS are well correlated with the LCC in the tropics. While EIS is available from GISS-E3, we could only compute the LTS with the output available from other models. We now mention this information in the manuscript: *"Using estimated inversion strength (EIS) – an LCC predictor that can be also used at mid-latitudes (Wood and Bretherton, 2006) – rather than LTS gives even better correlations in both the observations (see Fig. 5) and E3. However, we could only compute the LTS with the output available from other models."*