Response to referee #4.

Referee #4: The authors developed a novel POPEM parameterization and applied it to CESM to enhance the realism of global climate modeling by improving the direct representation of human activities and climate. They argued that modeling CO2 emissions and pollutants directly at model grid points is a better approach. As such, their new approach will help understand the potential effects of localized pollutant emissions on long-term global climate statistics, thus assisting adaptation and mitigation policies. The topic is interesting and the approach is provoking.

Reply: Thank you for your positive feedback.

Referee #4: However, I am not quite convinced by the validation part (Part 3.2). I therefore recommend major revision.

Reply: We followed your recommendations. We have expanded section 3.2 and added a new subsection; *3.3 Validation against ESPI and ONI indices.* Please, see following comments for a detailed revision of the updates. Hope the changes can solve your concerns.

Referee #4: First, I cannot find a remarkable improvement using POPEM based on the comparison of precipitation and temperature biases. There are some differences between POPEM and CONTROL but these differences are buried in the large biases in either set.

Reply: We have made clearer in the paper that we do not claim to solve the problem of homogenous emissions versus point-wise estimates. We did not state that our contribution produces a remarkable improvement. What we have achieved by now is far more modest: we have shown that including our more-realistic forcings preserves the model ability to produce realistic fields. Nonetheless, some improvements can be seen (we have included additional figures to illustrate the improvements). We agree that the improvements are limited, but given the small model sensitivity to this forcing (the logic of RCP85 is to somehow 'exaggerate' the emissions to increase the signal), one cannot expect major changes. In other words, the actual signal is too faint to be affected by a more realistic emission pattern. Indeed, the reason for having a distributed method is to be able to evaluate 'what-if' scenarios (i.e. what happens if China cuts off emissions, or the like). We have added a paragraph at the end of the section 3.1 to explain why the approach is valuable in spite of the marginal improvements compared with validation data.

As referee #5 says, we also believe that the use of local population projections to project emissions at each grid point is novel, and is advantageous to the current practice of using global emissions projections to drive ESMs.

The added paragraph reads:

Potential applications of POPEM include not only sensitivity analyses of local CO_2 emissions policies, but also the added feature of performing tests for 'what-if' scenarios. One interesting example would be the climate response under the hypothesis that China and India –the most populated countries in the world- reach US CO_2 per capita emissions rates. Another 'what-if' scenario would be the climate response of an increasingly urbanized world. In both cases, POPEM provides a flexible framework for testing the alternative hypotheses.

Referee #4: *It is true that observations have uncertainties and a new parameterization does not have to improve the model performance in every aspect. Nevertheless, could the authors show some improvements more robust than the current ones (precipitation and temperature) for validation? Maybe TOA radiation balance, ENSO index, Arctic sea ice, etc?*

Reply: We agree that the analysis of Artic sea ice response would be a good addition. Unfortunately, sea ice was not a focus of our research when we ran the simulations and now it is too late to do so. Same about TOA. However, in order to satisfy this requirement, we have included two additional validation metrics using two ENSO indices: namely the ENSO Precipitation Index (ESPI) and the Oceanic el Niño Index (ONI).

We have chosen the ESPI index, which estimates the gradient of the anomalies across the Pacific basin (Curtis and Adler, 2000). It compares well with SST-and pressure-based indices and is widely used by the scientific community (Figure 13 now). The Oceanic el Niño Index is a SST index developed by NOAA as a principal measure for monitoring, assessing and predicting ENSO (Kouski and Higgins, 2007).

We have made two new figures and added a table: Figure 13 for ESPI index, El Niño (EI) and La Niña (LI), and Table 1 and Figure 14 for ONI.

The new section reads as follows:

3.3 Validation against ESPI and ONI indices

The El Niño-Southern Oscillation (ENSO) is the most dominant inter-annual climate variation in the tropics. It occurs when seasonally averaged sea surface temperature anomalies in the eastern Pacific Ocean exceed a given threshold and cause a shift in the atmospheric circulation (Trenberth 1997). Historically, the definition of ENSO does not include precipitation because of the limitations of stations (Ropelewski and Halpert, 1987), but recent work with satellites has confirmed that this phenomenon is a major driver of global precipitation variability (Haddad et al., 2004).

A major advantage of satellite-derived precipitation indices over more conventional ones is the description of the strength and position of the Walker circulation (Curtis and Adler, 2000). Under that assumption, Curtis and Adler

(2000) derived three satellite-based precipitation indices: the ENSO precipitation index (ESPI); El Niño index (EI); and La Niña index (LI). Precipitation anomalies are averaged over areas of the Equatorial Pacific and Maritime Continent -where the strongest precipitation anomalies associated with ENSO are found- to construct differences or basin-wide gradients (Curtis, 2008).

Figure 13 shows a comparison of GPCP, CONTROL, and POPEM for the ESPI, EI and LI indices.
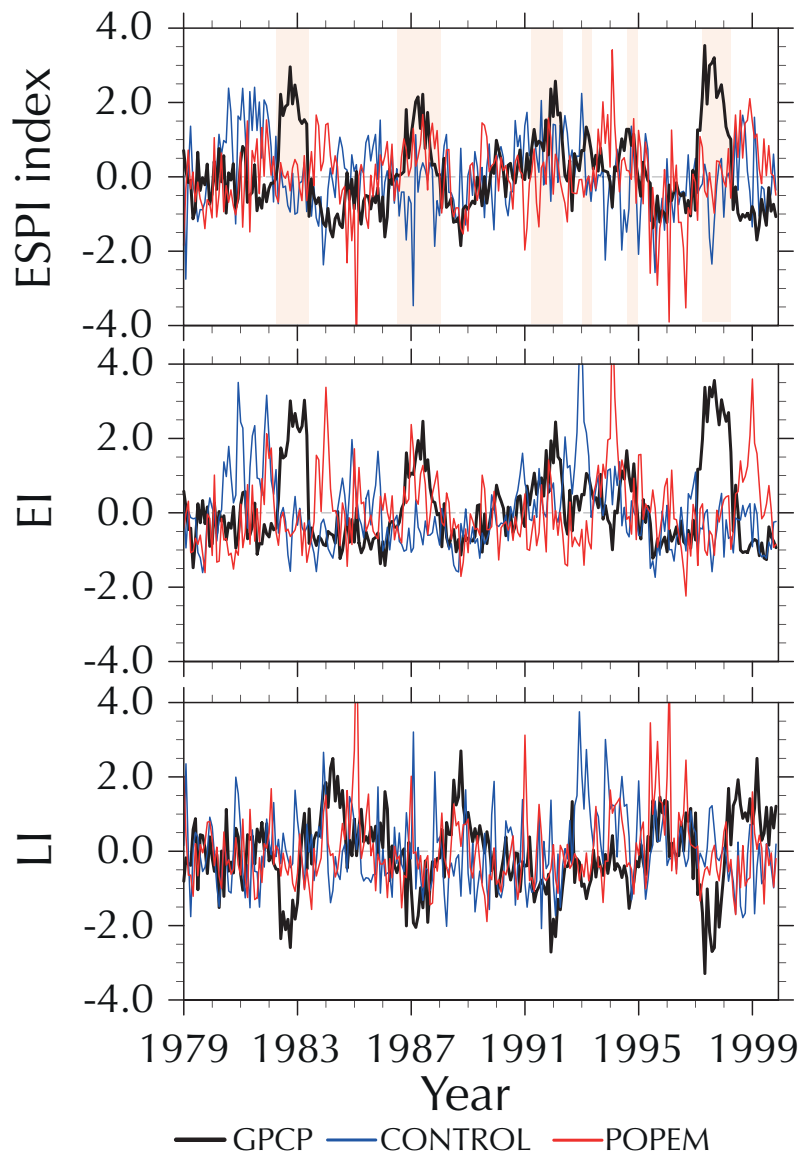
Figure 13: Time-series of precipitation anomalies for the ENSO region after Curtis and Adler (2000). (Top) ENSO Precipitation Index (ESPI); (Middle) El Niño Index (EI); and (Bottom) La Niña Index (LI). The Black line shows GPCP data, the blue line is the CONTROL case, and the red line is the POPEM case. Orange shading denotes El Niño years defined as consecutive months (minimum 3) with NIÑO3.4 sea surface temperature anomalies (5N–5S, 170–120W) greater than +0.5^{o} C.

Unfortunately, CONTROL and POPEM cases have difficulty simulating the precipitation patterns associated with ENSO. Figure 13 shows that bias increases in 82-83 and 97-98 El Niño years. The same bias emerges when comparing the EI and LI indices. In that case, the CESM model produces stronger El Niño/La Niña events than the observed data. Consequently, we can consider that CESM is unable to obtain a precise estimate of precipitation patterns, suggesting that current climate models are far from generating realistic simulations of the precipitation field (Dai, 2006).

Another widely used ENSO index is the Oceanic Niño Index (hereafter ONI). ONI was developed by the NOAA Climate Prediction Center (CPC) as the principal means for monitoring, assessing and predicting ENSO (Kousky and Higgins, 2007). This index is defined as 3-month running-mean values of SST departures from the average in the Niño-3.4 region. It is computed from a set of homogeneous historical SST analyses (Kousky and Higgins, 2007, Smith et al. 2002).
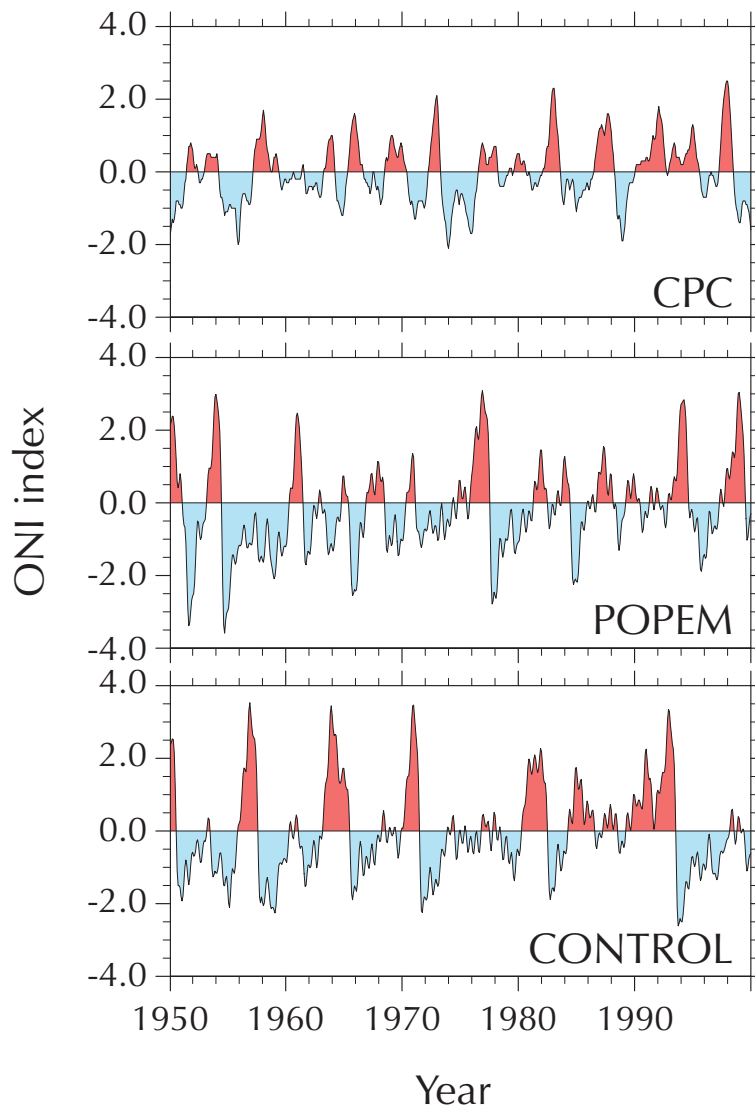
Figure 14: Comparison of the Oceanic el Niño Index (ONI) for CPC (top), POPEM (middle), and CONTROL (bottom) cases. El Niño and La Niña are defined according to Kousky and Higgins (2007): 3-month running mean with anomalies greater than $+0.5^{o}$C (or-0.5^{o}C) for at least five consecutive months in NIÑO3.4 region. The base period for computing SST departures is 1971–1999.

Figure 14 compares the ONI index for CPC, POPEM and CONTROL cases. It is clear from the figure, that POPEM produces a more realistic representation of the ENSO, especially if we focus on the 1992-1999 period. POPEM also obtains better results than CONTROL in the number of simulated el Niño events (see Table 1). The improvement is also noticeable in the intensity. The CONTROL case exhibits an overly strong ENSO -a common bias in CESM (Tang et al., 2016)- but POPEM reduces this bias (0.22^{o} C versus 0.59^{o} C).

Table 1. Comparison of the ONI index for the period 1950-1999. The table compares the ability of the models to reproduce the number, strength, and duration of el Niño events.

Source	Number of events	Agreement[1]	Disagreement[2]	Intensity Bias$_{avg.}$[3]	Duration$_{avg}$[4]
CPC	14				10.3
CONTROL	7	33	121	0.59 °C	19.4
POPEM	10	37	121	0.22° C	11.4

[1] The number of months that CPC and CESM agree on El Niño. [2] Disagreement defined as the number of months where CPC and CESM obtain opposite results. [3] Intensity: (|CESM ONI| − |CPC ONI|)/number of cases (units in degrees Celsius). [4] Mean duration of El Niño event (in months).

Another important indicator is the mean duration of El Niño events. Table 1 shows that POPEM obtains better results according to observations (11 months in CPC, 10 months in POPEM, and 19 months in CONTROL).

Referee #4: Actually, I am somewhat interested in the Arctic sea change. It is known that climate models (like CESM CONTRL) cannot capture a rapid observed decline of Arctic sea ice during recent decades. In Fig. 5(B), POPEM is colder than CONTROL over the Barents Sea area. Will this mean that Arctic sea ice decline in POPEM is even slower than that in CONTROL?

Reply: It's true that the POPEM parameterization produces colder temperatures in that area and that might reinforce the bias of a slower Artic sea ice decline. Unfortunately, we can't contrast this hypothesis because we did not keep the sea ice outputs for our simulations. Sorry about that.

The bias is less evident when confronted with GISTEMP annual mean anomalies for that area. It is seen from the Figure 11 (top) that CONTROL and POPEM cases have a similar margin error. In other words, the original CESM model is not really good in capturing this feature. Our approach slightly improves the situation in some cases (Bering Sea from 1975 to 1990, Figure 11 (middle)) but we cannot expect a major overall improvement.

We have added a paragraph and a figure to clarify this point.

The text now reads:

The bias is also reproduced when compared with temperature anomalies for a specific region. Thus, for instance, CESM gives poor scores in the Barents Sea area (Figure 11; top) while POPEM obtains better results for the Bering Sea, especially in the Russian part (Figure 11; middle). Here, POPEM gives more realistic values for the period 1970-1998 but, even with the improvement, the model still overestimates the temperature anomaly.
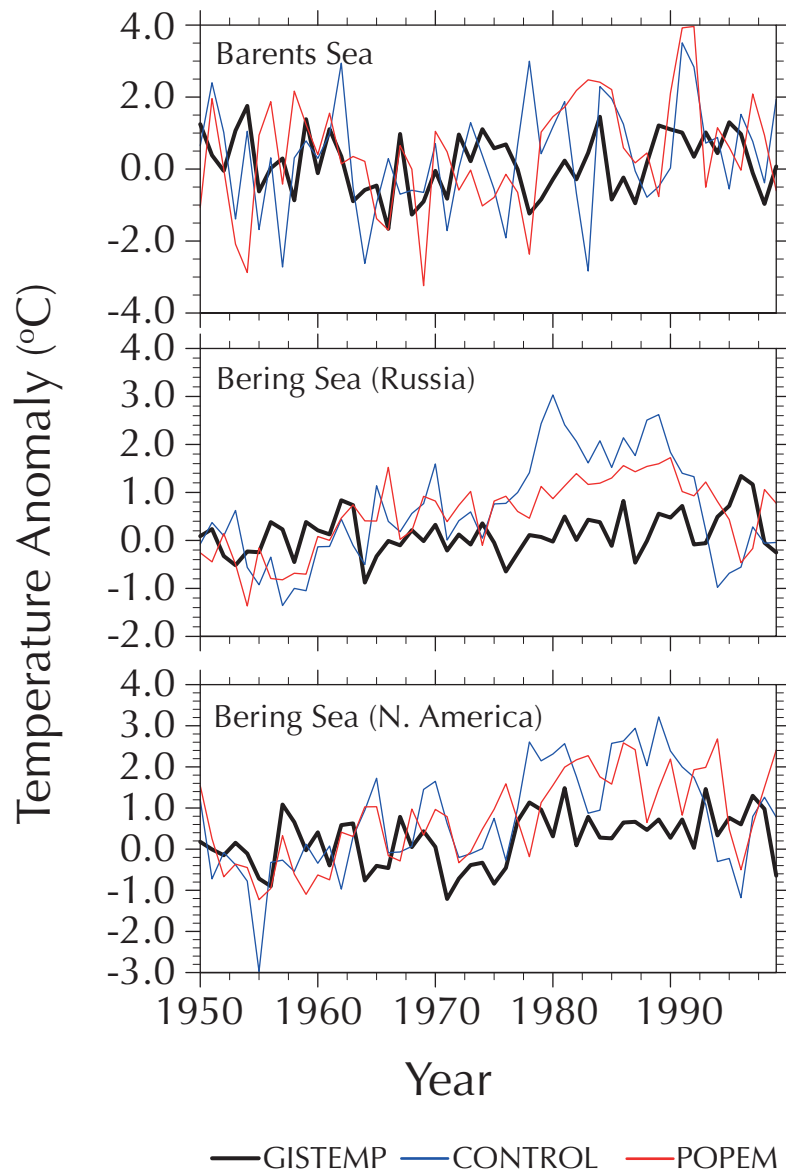
Figure 11: *A comparison of the annual mean surface temperature anomaly between GISTEMP, CONTROL and POPEM from 1950 to 1999. (Top) represents the Barents Sea (68N-80N, 19E-68E); (middle) Russian part of the Bering Sea (50N-65N, 150E-180E); and (bottom) American part of the Bering Sea (50N-75N, 140W-180W). The black line represents observational data (GISTEMP), the blue line is the CONTROL case, and the red is the POPEM case. Anomaly was referenced to 1951-1980 period.*

We also calculated the temperature anomalies with monthly data (attached as a supplementary material). However, the noise is high and it is difficult to distinguish any clear pattern other than the consistency between the series. Only in Figure EXT2(top) we see that POPEM more frequently yields extreme values.
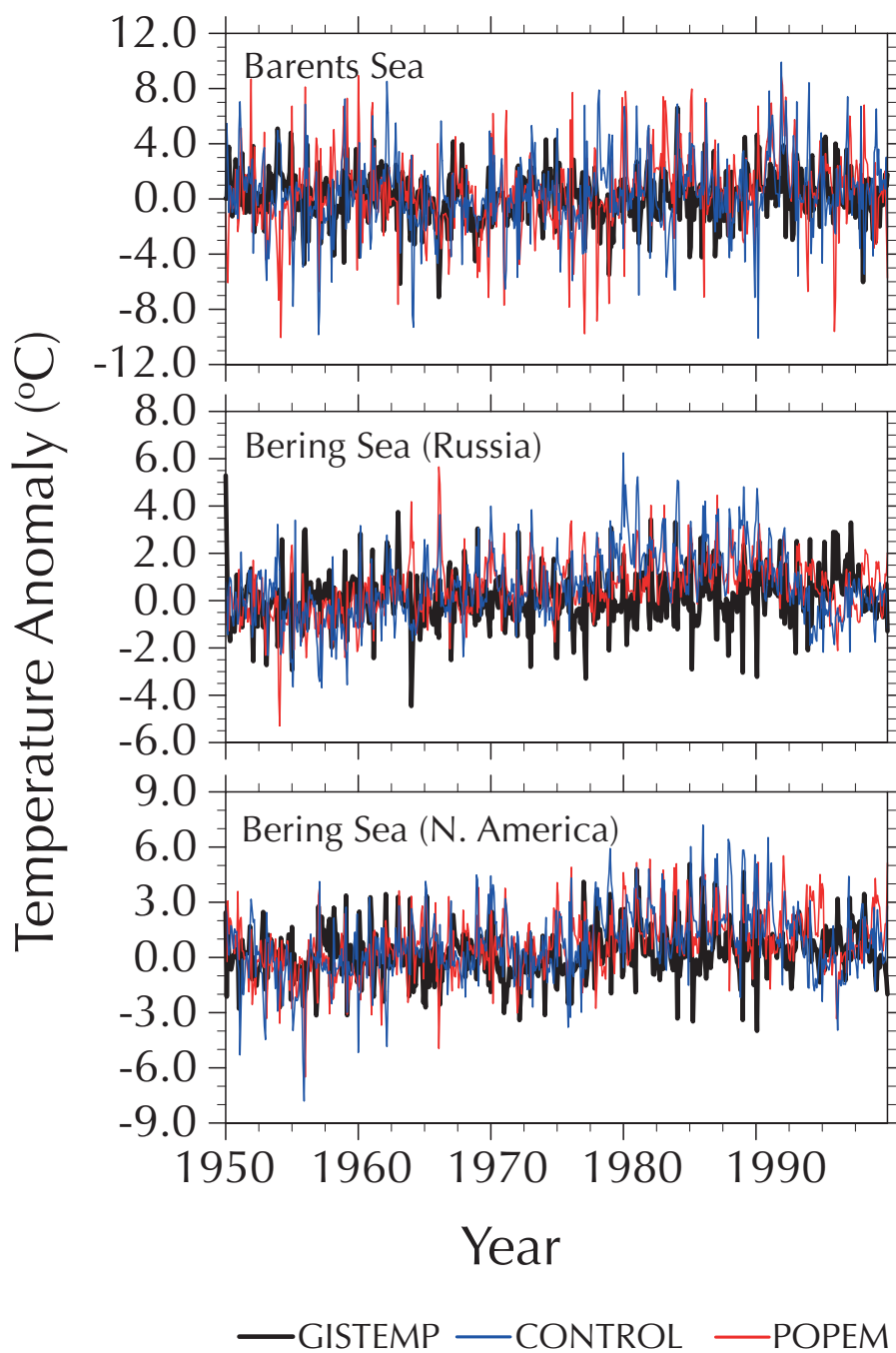
Figure EXT2: The same as Figure 11 but using monthly means.

Reply: Thanks for the suggestion. As you seen in the previous comment we included GISTEMP in several figures and also made a brief description of the source in section 2.4.3.

The new subsection reads:

> ### 2.4.3 GISTEMP data set
> *NASA's GISTEMP (GISS Surface Temperature Analysis) is a global surface temperature change dataset (Hansen and Lebedeff, 1987; see Hansen et al. 2010 for an updated version). It combines land and ocean surface temperatures to create monthly temperature anomalies at 2° x 2° degrees of spatial resolution. The use of anomalies reduces the estimation error in those places with incomplete spatial and temporal coverage (Hansen and Lebedeff, 1987). The anomalies are calculated over a fixed base period (1951-1980) that makes the anomalies consistent over long periods of time.*
> *The first version was originally conceived only for land areas (Hansen and Lebedeff, 1987) but in 1996 marine surface temperatures were added (Hansen et al., 1996). The updated version of GISTEMP includes satellite-observed nightlights to identify stations located in extreme darkness and adjust temperature trends of urban stations for non-climatic factors (Hansen et al. 2010). Just like CRUTS, GISTEMP is commonly used to validate climate models because of its coverage and confidence levels (Baker and Taylor, 2016; Brown et al., 2015; Neely et al., 2016, Peng et al., 2015).*

Additionally, we used GISTEMP to analyze temperature anomalies for regional (previous comment; Figure 11) and global scales (Figure 12).

The results of Figure 12 were discussed in the section 3.2 of the manuscript:

The new paragraph reads as follows:

> *If we focus on global temperature anomalies, CESM simulations are able to reproduce the progressive increase in the temperature anomaly (Figure 12; top). However, the CONTROL case simulates a sharp drop at the end of the period (1990-1999), while POPEM portrays this change as smooth, in agreement with the observations.*
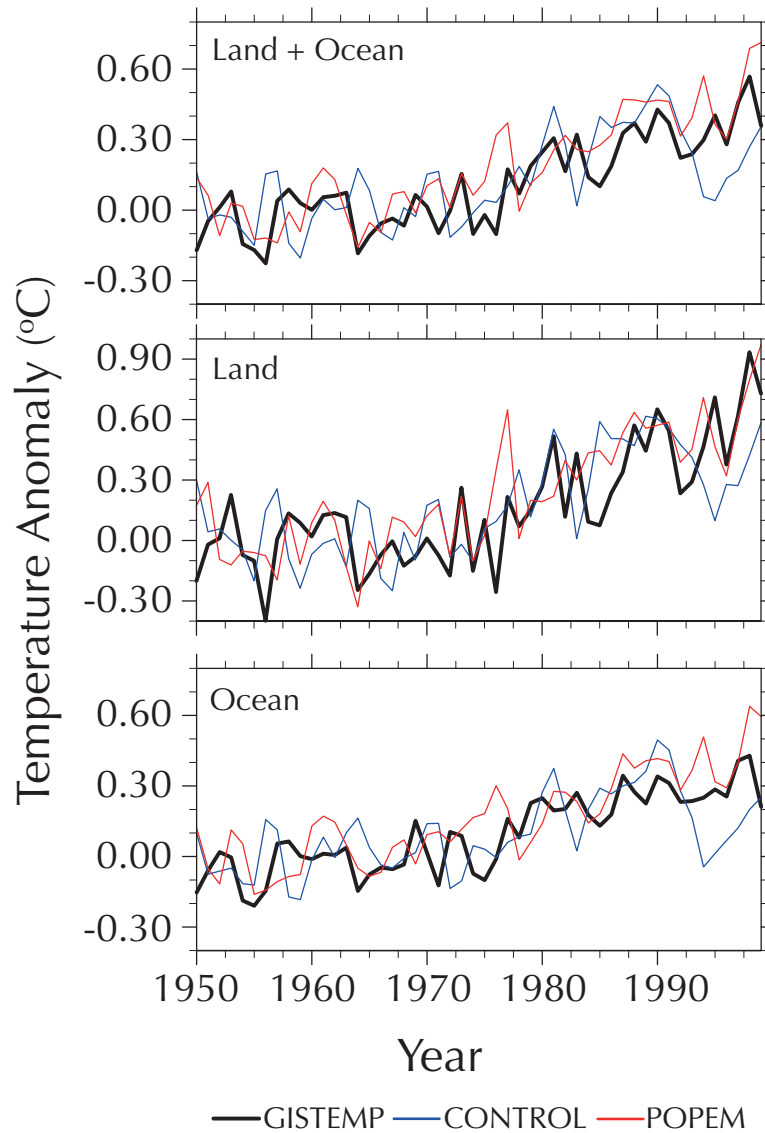
Figure 12: A comparison of the global annual mean surface temperature anomaly between GISTEMP, CONTROL, and POPEM from 1950 to 1999. (Top) global; (middle) land; and (bottom) ocean. The black line represents observational data (GISTEMP), the blue line is the CONTROL case, and the red is the POPEM case. Anomaly was referenced to 1951-1980 period.

The differences between CONTROL and POPEM are better demonstrated when comparing land and ocean separately (Figure 12; middle and bottom). While the temperature anomalies for land are quite similar in both cases, POPEM provides a better representation of the ocean tendency from 1992 onwards, and that translates to an overall improvement (Figure 12, top).

We also computed monthly mean temperature anomalies. However, is difficult to appreciate the differences between models, especially for cases A and B. The figure is therefore included as a supplementary material.
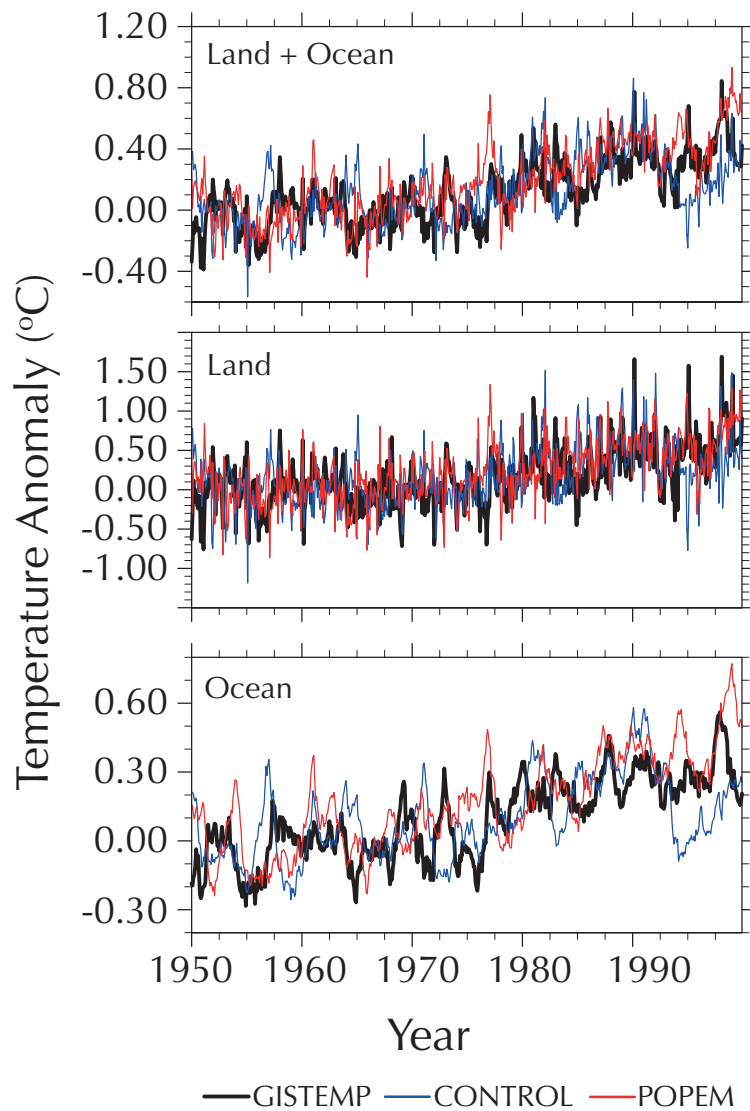

Figure EXT3: Same as Figure 12 but for monthly mean temperature anomaly. The main tendency is consistent albeit differences exists. Thus for instance the POPEM model clearly improves over CONTROL from 1992 onwards.

References

Curtis, S., Adler, R., Curtis, S. and Adler, R.: ENSO Indices Based on Patterns of Satellite-Derived Precipitation, J. Clim., 13(15), 2786–2793, doi:10.1175/1520-0442(2000)013<2786:EIBOPO>2.0.CO;2, 2000.

Kousky, V. E. and Higgins, R. W.: An Alert Classification System for Monitoring and Assessing the ENSO Cycle, Weather Forecast., 22(2), 353–371, doi:10.1175/WAF987.1, 2007.