Reviewer 1

We thank the reviewer for their comments. We agree with the reviewer that the evidence to support some of the conclusions regarding the mechanism is not as strong as we would like. However, as the reviewer commented, this is unavoidable given the limitation of a multi-model analysis. We have softened some of our conclusions and included more discussion/analysis of the robustness of our results. Responses to the reviewer's explicit comments are listed below (in red).

1. Most conclusions of this paper is based on correlation analysis of climatologies among models. As the authors already pointed out, these correlation are sensitive to the choices of models included into the analysis. Thus, even a seemly high and statistically significant correlation may not be robust after all. For example, the correlation between phi_v=0 and X_CO50 in JJA is 0.65 (Fig. 10b). But if excluding CAM-C1, WACCM-C1 and GEOS-C1, the correlation becomes much weaker and may even changes sign.

    We agree with the reviewer that the correlations are somewhat sensitive to the choice of the models. We had acknowledged this in the original version, but have made this clearer in the revised version. However, we think the correlations are actually fairly robust. We have calculated the correlation coefficients for different subsets of models, e.g., excluding SD, excluding models with similar heritage, and there are generally only small changes in correlations and often a higher correlation than using all models (e.g. using just the free-running models results in higher correlations between $\phi_{\mathrm{v=0}}$ and CO-50, see Table R1). Excluding CAM-C1, WACCM-C1 and GEOS-C1 does have a big impact but we do not think this is a reasonable thing to do as these simulations are not physical outliers and their exclusion greatly reduces the range of $\phi_{\mathrm{v=0}}$ and leaves only 2 independent FR models, so that resulting correlation is determined only by SD models. As we now make clear in the manuscript, the relationships are much less robust if only SD models are considered.

2. The authors argued the mechanisms leading to the correlation between phi_v=0 and X_CO50 is that if the divergence zone locates north of the emission, then the northward transport from the source region is limited. This mechanism should work on the interannual time scales as well as the climatologies. Yet, the interannual variation within each model does not show similar correlations to the climatologies among different models. This is evident in Fig. 10e, which shows that the interannual variations of phi_v=0 within each model is comparable to the inter-model spread of climatology, but the corresponding Arctic CO concentration does not show any negative correlation. In models such as CMAM-C1SD, the correlation from inter-annual variations even seems to be positive.

    We agree with the reviewer that the interannual variation between different metrics (ϕ_{jet}, mean flux, and $\phi_{\mathrm{v=0}}$) and CO50 for individual models generally doesn't show the same relationship as between models. This is now discussed in detail in the revised manuscript. There are two main reasons for this disagreement: (i) multiple processes influence the transport of CO50 into the Arctic, and although the mid-latitude mean meridional flux may dominate the variations between the models, other processes can have a larger influence on the interannual variation. (ii) The metrics used to define jet location and Hadley Cell edge can have large uncertainties for monthly or even seasonal data (e.g., jet can be weak with no distinct peak or v can be close to 0 over a broad latitude range, for given year) and differences in the metrics between two years may not be representative of the differences in the jet and HC width. Averaging over multiple years is required to get robust values of these metrics.

    An example of another process that influences the transport of CO50 into the Arctic is the meridional transport at high latitudes. The climatological meridional velocity and mean flux poleward of 70N are similar among models as compared to the intermodal variations in the midlatitudes (see Figures 9 and 10) and hence the transport difference due to advection by this flow does not contribute to the spread among models (see Table R2). However, there can be large interannual variations in the high latitude meridional flow/flux within individual models (see Figure R1). This plays an important role in the interannual variations in Arctic CO50 in most of the models with high or moderate interannual correlations between high-latitude meridional flux and CO50 (see Table R2). This is one example, and there are likely other processes that make a larger contribution to interannual variations than to the spread among models.

As indicated above, another reason for the differences between interannual correlations and correlations among model climatologies is the large uncertainty / non-representativeness of metrics for individual seasons. A clear example of this is the $\phi_{v=0}$ during JJA. In many models, v is close to 0 for a wide range of latitudes in this season, and although there are only small differences in the mean meridional flow (and mean meridional flux) over this region between the years there can be large variations in $\phi_{v=0}$, see Figure R1. In particular, in the SD models the $\phi_{v=0}$ can vary between 30°N and 40°N for individual years, even though for all years there is a very weak meridional flow over this range of latitudes. As a result, the interannual spread in $\phi_{v=0}$ is not representative of interannual differences in meridional transport of CO50 between the years, and the interannual variations in $\phi_{v=0}$ (shown in Figure 10b) are generally larger than those in mean fluxes (Figure 9b).

3. The authors argue the importance of the zonal mean circulation for the tracer transport by comparing the zonal mean and eddy component in two models: GEOS-C1 and GEOS-C1SD. But as shown in Fig. 9a, GEOS-C1 is clearly an outlier in terms of the zonal mean contribution to the tracer flux. If comparing GEOS-C1SD versus WACCM- C1 instead of GEOS-C1, the difference in the zonal mean flux would be much smaller. But the Arctic CO concentration is similar in WACCM-C1 and GEOS-C1, this would implies that the eddy component may be more important to explains the difference between the two models.

   We have followed the reviewer's suggestion below (comment 4) and now show in Figure 9 the vertically integrated mean flux calculates over the lowest levels (700-1000 hPa). GEOS-C1 is no longer an outlier, and the zonal-mean transport in GEOS-C1 is not structurally different from other models (e.g., GEOS-C1SD). Fig. 8 is also adjusted accordingly.

4. Since the argument is about the low level divergence zone (800-950 hPa), why the tracer flux is integrated over the whole troposphere (200-1000 hPa) rather than just the low level? The mean circulation pointing to opposite direction in the lower and upper troposphere, and hence there may be canceling effects when integrating over the whole troposphere.

   This is a good point, and we have modified the figure as the reviewer suggested. There are no major changes to the conclusions, but the new figure should help ease the reviewer's concern on GEOS-C1 being as an outlier.

5. Other Editorial comments

   We have revised the manuscript following reviewer's suggestions.

Reviewer 2

We thank the reviewer for their comments. We agree with the reviewer that the evidence to support some of the conclusions regarding the mechanism is not as strong as we would like, due to the limited number of models, and we have softened some of our conclusions and included more discussion/analysis of the robustness of our results. Responses to the reviewer's explicit comments are below (in red).

1. While the physical mechanisms proposed by the authors are plausible, the supporting evidence is based on correlations of decadal averages within a very limited number of models rather than a clear causal indication of the process. It is clear that the number of models that can be examined conclusions can be reached using these methods. In particular, the number of independent pieces of information is unclear. WACCM and CAM-1 seem nearly identical in the scatter plots. The EMAC-47 level model and EMAC-90 level model are definitely not independent, and considering the frequent grouping of these with their SD runs in the scatter plots, it would be useful to know how many independent data points these four represent. Since, as the authors note, the correlations can be easily influenced easily by which models are included the choice of weighting the single ACCESS run as equivalent to one of the EMAC runs, for example, could change the results significantly.

   We agree that it is not possible to make strong conclusions given the limited number of models, and we have softened our language so we are saying the analysis is suggestive rather than proves or shows. We also agree that the models are not all independent, and we have included discussion of the relations between models in the section 4.1. Specifically, we have now noted in the text the similarity between WACCM and CAM, between ACCESS and NIWA, and different EMAC versions. We have also calculated the correlation coefficients only using one of each of these pairs and the values are essentially the same to those using all the models (see Table R1).

2. I would expect that the mechanisms the authors point to would work on shorter timescales than decadal, so it is strange that the small dots of the annual averages do not fall closer to the relationships that are calculated. Perhaps an annual average is too short for the process to be valid, but why should a decadal average be necessary? Some filtered time series showing that this process is valid within a model and not just between models would be helpful, as it would help convince the reader that this physical mechanism is correct. Maybe GEOS has a longer record that could be used? And if the relationship only holds between models (i.e. the explanation of the jet and the tracer concentration only explains the difference between models, not the physical processes within the models), then it just implies that biases in models between transport of tracers and large scale circulation near the tracer sources are correlated.

   We agree with the reviewer that the interannual variation between different metrics (ϕ_{jet}, mean flux, $\phi_{\mathrm{v=0}}$) and CO50 for individual models generally doesn't show the same relationship as between models. Please see our response to comment #2 from reviewer 1, who raised the same issue.

3. The statistical methodology is problematic. The correlations are done using least squares, which necessarily holds one variable as dependent and one as independent. This is why all of the "fits" where there is no correlation are horizontal lines. Since there is no fundamental reason to expect that either quantity should be an independent variable, this method is not sufficient. Either the fit needs to be done with the x and y reversed (in y=ax+b) and that slope should be plotted as well, or a reduced major axis regression, with each quantity scaled by its variance, should be employed. With small r values (e.g., Figure 3), no fit line should be shown at all.

   We followed the reviewer's suggestion and used a total least square method (Petráš and Bednárová, 2010) to calculate the linear fit. Note that the Pearson's correlation coefficient is not affected by the method of fitting. We have also followed the reviewer's suggestion and do not show the linear fit when the correlation is not significant.

4. The analysis of the differences between GEOS and GEOS SD is fine but since GEOS is shown to have quite different behavior from the other models in Figure 9, the conclusions drawn from this comparison are not obviously going to apply to other models.

   We have followed comment #4 by reviewer 1 and now show in Figure 9 the vertically integrating the mean flux only over low levels (700-1000 hPa), which is the most relevant for transport away from the source. In the new Figure 9 GOES-C1 is no longer an outlier, and therefore the proposed mechanism of location shift in mean meridional circulations and associated changes in low-level meridional flow is still applicable to the difference between GEOS-C1 and GEOS-C1SD.

5. My final concern is that analyzing transport in SD runs seems inherently problematic, since it is unclear whether they are actually conserving tracers. I assume that, since the authors are analyzing these, they have reason to believe that it is not a problem. A discussion of the extent to which the SD runs do or do not conserve mass and tracer concentrations would be helpful. I'm not sure if such a study already exists or if the authors could do some analysis of their own with the GEOS-SD model.

   Mass conservation is not inherently problematic for the SD runs. These simulations used the same numerical framework as the free-running (FR) counterpart, so have the same transport algorithm and same implementation of emissions and chemistry as the FR. The only difference between a pair of SD and FR simulations is the winds that are used in the transport. While there may be conservation issues with the winds directly from the reanalyses, in all but one of the SD runs the horizontal winds are only nudged towards the reanalyses (and not set directly equal to the reanalyses) and winds are constrained together with the vertical velocity to conserve mass. The one exception is the GEOS-CTM, the only CTM analyzed in this study, which uses horizontal and vertical winds directly from the reanalyses, and adjusts surface pressure to conserve mass. To reassure the reviewer, we calculated the global CO50 burden in each CCMI simulation (see Table R3), and only small differences can be found among models (except GEOS-CTM), with these small differences likely due to due to interpolation of the emissions onto different resolutions of model grids. In particular, when considering the C1-C1SD pairs within the same base model, the differences are very small (less than 1%).

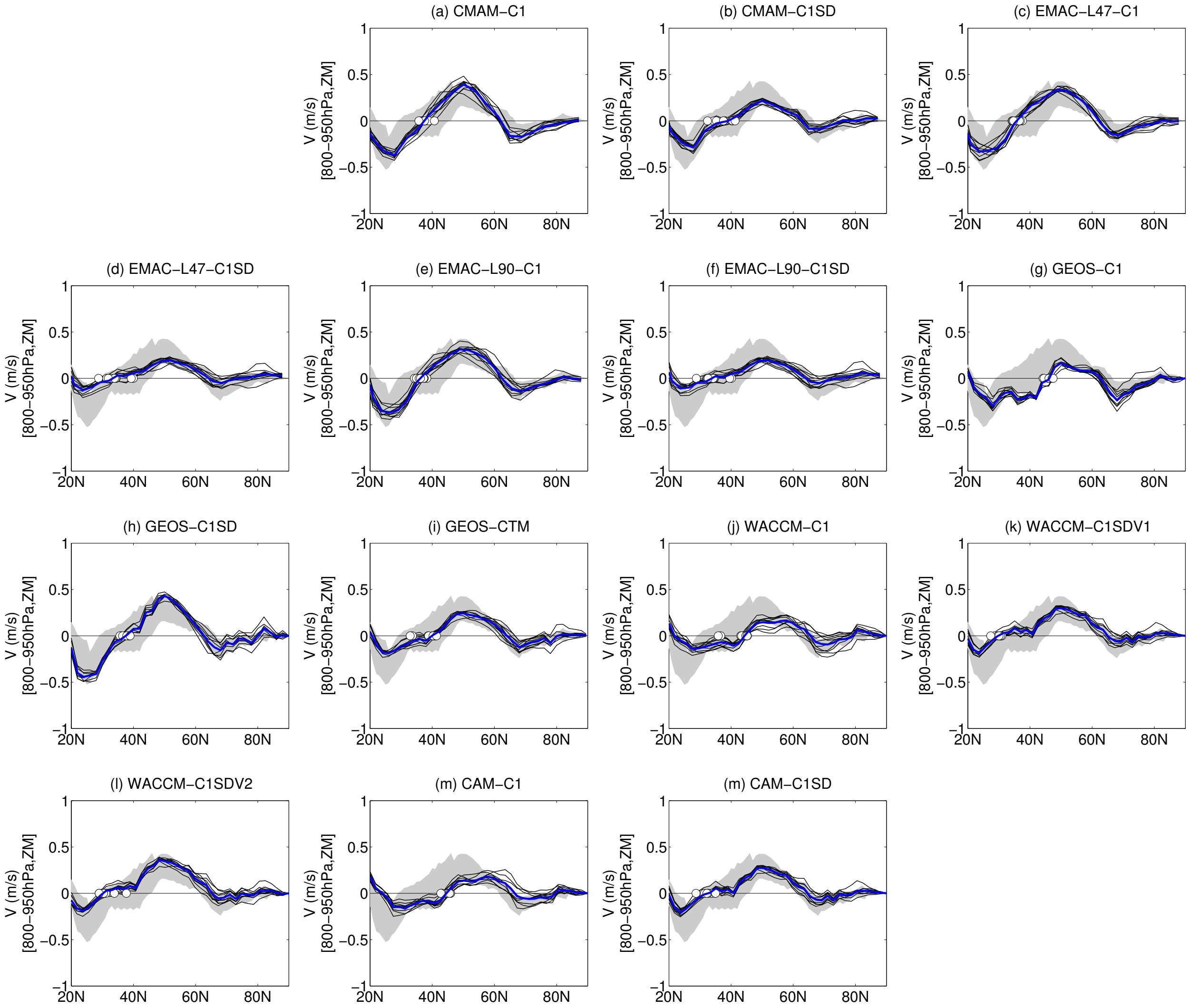

Figure R1 Latitudinal profile of JJA low-level zonal mean meridional wind v (800 – 950 hPa) in each simulation. The black lines and circles denote the interannual variations of v and $\phi_{\mathrm{v}=0}$ within the model, while gray shades give the multi-model spread among model climatology, as shown in Figure 10(a). The blue thick line denotes the climatological v of the simulation.

Table R1 Correlation coefficients between Arctic CO50 concentration and physical process related metrics (such as ϕ_{jet}, mean flux, and $\phi_{\text{v=0}}$ as listed in the 1st column) calculated using all available models versus only the free-running (C1) models versus clustered models (i.e., exclude NIWA-C1 for similarity to ACCESS-C1, exclude EMAC-L90-C1/C1SD for similarity to EMAC-L47-C1/C1SD, exclude WACCM-C1SDV2 for similarity to WACCM-C1SDV1, and exclude CAM-C1/C1SD for similarity to WACCM-C1/C1SDV1). Calculation of correlation is based on climatology, and those are statistically significant (95%) are marked in **bold**.

	DJF			JJA		
	All	C1	Cluster	All	C1	Cluster
ϕ_{jet}	**-0.63**	**-0.92**	**-0.64**	**-0.84**	**-0.79**	**-0.83**
mean flux	**0.69**	0.68	**0.69**	**0.78**	**0.96**	**0.79**
$\phi_{\text{v=0}}$	**-0.76**	**-0.95**	**-0.76**	**-0.58**	**-0.70**	**-0.51**

Table R2 Correlation coefficients between mean meridional flux over high latitudes (60°N-80°N) and Arctic CO50 concentrations (500-800 hPa, 70°N-90°N, zonal mean). For mean meridional flux over high latitudes, it is vertically integrated in the low levels (700-1000 hPa) during DJF but in the upper levels (300-500 hPa) during JJA, considering differences in the CO50 vertical maximum between seasons (Fig. 1 and Fig. 2(c,d)). The correlation among model climatologies are shown in the 2nd row (similar to results shown in Fig. 9 (b,d)), while interannual correlations in individual simulations are shown in the 3rd – 16th rows. Coefficients that are statistically significant (95%) are marked in **bold**.

	DJF	JJA
Multi-model Climatology	0.28	0.37
CMAM-C1	**0.67**	-0.26
CMAM-C1SD	0.35	**0.74**
EMAC-L47-C1	0.50	0.22
EMAC-L47-C1SD	0.43	0.60
EMAC-L90-C1	0.51	-0.25
EMAc-L90-C1SD	0.43	0.55
GEOS-C1	**0.88**	**0.70**
GEOS-C1SD	0.12	**0.75**
GEOS-CTM	0.46	**0.84**
WACCM-C1	-0.02	0.52
WACCM-C1SDV1	0.09	**0.78**
WACCM-C1SDV2	0.53	**0.83**
CAM-C1	0.50	0.51
CAM-C1SD	0.15	**0.75**

Table R3 Climatological global CO50 burden (units: $\times 10^{10}$ kg) in each CCMI simulation. C1-C1SD pairs in the same base model are highlighted in the same box. Note that molecular weight of CO50 is assumed to be the same as CO (28g/mol).

Model	Global CO50 burden ($\times 10^{10}$ kg)
ACCESS-C1	6.18
CMAM-C1	6.39
CMAM-C1SD	6.40
EMAC-L47-C1	6.35
EMAC-L47-C1SD	6.37
EMAC-L90-C1	6.38
EMAC-L90-C1SD	6.40
GEOS-C1	6.42
GEOS-C1SD	6.46
GEOS-CTM	5.95
WACCM-C1	6.22
WACCM-C1SDV1	6.21
WACCM-C1SDV2	6.20
CAM-C1	6.24
CAM-C1SD	6.23
NIWA-C1	6.18