We thank both reviewers' comments on the manuscript. We have listed our point-by-point responses below with black text showing the reviewers' comments and red text showing our responses.

Reviewer #1

This manuscript discusses the simulated tracer transport from midlatitude to the Arctic in CCMI models. The authors showed that for an idealized tracer CO50, large inter-model spread is found in its concentration over Arctic with or without specified dynamics. Unlike tracer NH5 reported in earlier studies, this inter-model difference cannot be explained by parameterized convection, but is found to be correlated with midlatitude jet position and Hadley cell boundary. Long-range transport of chemically and radiatively important species are important issues that lacks of full understanding. This manuscript provides a useful assessment of the CCMI models' performance on this subject. In this revision, the authors have made some efforts to improve the robustness of their results. I have some remaining minor comments for the authors to consider.

The "20%-40%" inter-model spread of the Arctic CO50 concentration. This number is quoted in the abstract and the conclusion, but it is not very clear where this number comes from. From the paragraph on page 6 line 4-13, it seems this number is calculated as the range (max-min) of the Arctic mean CO50 concentration divided by its multi-model mean. In this paragraph, the range is quoted to be ~11ppbv in DJF. Estimated from Fig. 2, the multi-model mean Arctic CO50 concentration is between 20 to 25 ppmv below 400 hPa. Then the relative inter-model spread would be 44%-55%. Is it because the authors define the inter-model spread as something else than the range? Perhaps the standard deviation among models? In any case, the authors need to articulate how they calculate this inter-model spread.

We define the inter-model spread as the range (max-min) and the numbers of fraction are given as the ratio of such a multi-model range to the multi-model mean. Here, we re-examined the range, mean and fractional ratio for Arctic CO50 concentrations in Table R1. The fractional ratio is about 30%-45% in DJF and 25%-30% in JJA. Therefore, we have made it more clearer on how these fractions are defined and we also revised the associated numbers slightly in the manuscript.

Table R1 Multi-model range, mean, and fraction ratio (range/mean) of the Arctic CO50 concentration (70N-90N).

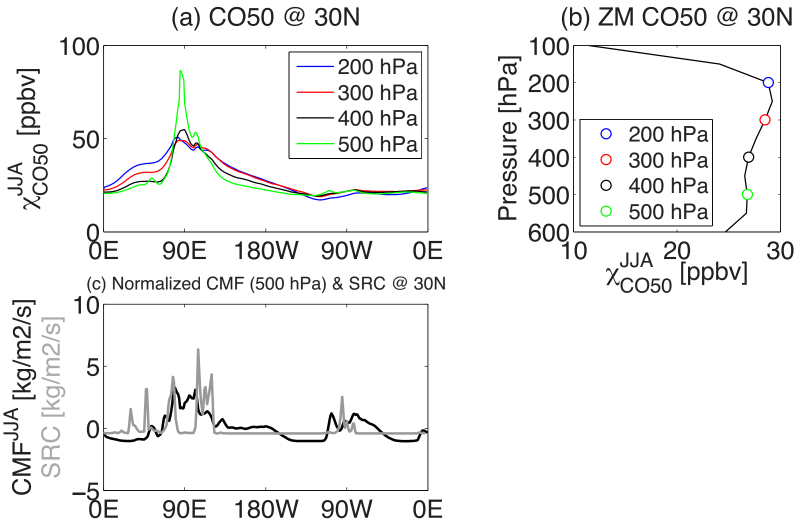
	DJF			JJA		
	Range [ppbv]	Mean [ppbv]	Ratio	Range [ppbv]	Mean [ppbv]	Ratio
500 hPa	9.73	21.72	0.45	4.92	18.81	0.26
550 hPa	9.39	21.95	0.43	4.96	18.43	0.27
600 hPa	9.08	22.13	0.41	5.14	18.03	0.29
650 hPa	8.69	22.35	0.39	5.21	17.58	0.30
700 hPa	8.20	22.65	0.36	5.20	17.10	0.30
750 hPa	7.58	23.04	0.33	5.12	16.60	0.31
800 hPa	6.87	23.59	0.29	4.95	16.02	0.31
500-800 hPa Average	8.51	22.49	0.38	5.07	17.51	0.29

Throughout the manuscript, the authors are doing averages over difference latitudes and pressure levels. It is not always clear why the authors made these choices. For example, CO50 concentration is averaged over 500-800 hPa, but the tracer flux is averaged over 700-1000 hPa, and the convective mass flux and the low level meridional winds are averaged over 800-950 hPa. The tracer mass flux is averaged over 40°N-60°N for JJA but 30°N-50°N for DJF. The authors need to justify their choices.

Thanks for the suggestion, and we have added some justifications throughout the revised manuscript. After the revision, we only apply two different pressure windows for vertical average. One is 500-800 hPa for examining the tracer concentration and zonal wind in the middle and lower troposphere. The middle and lower troposphere is the critical layer for trace gases to exert direct radiative impacts as well as indirect radiative impacts via interaction with clouds, and therefore it is of great interest to examine their concentrations at these levels. For zonal wind, middle and lower troposphere is the layer where the midlatitude jet streams can have a significant impact on tracer transport via wave breaking. The other vertical average zone is 800-950 hPa for meridional wind, flux, and convective mass flux, as all these metrics are found important at low levels. As to the difference of meridional window for tracer mass flux between JJA and DJF, this is because, on average, the maximum poleward zonal-mean flux occurs about 10° further north during JJA than that during DJF (see Fig. 9), and it is more reasonable to measure the tracer flux over regions where a similar distance from the Hadley Cell is kept.

Page 5 Line 27: I don't see how Fig. S1 support the argument in this sentence. Fig. S1 does not show the emission nor convection.

We have made a new Fig. S1 to show relations among CO50, emission, and convection. One important note for interpreting the secondary maximum of CO50 concentration in the subtropics (~30N) during summer is the difference between regional and zonal average. Over the convective plume where CO50 emission and convection are generally collocated, CO50 concentration still decreases as altitudes increase. However, in the perspective of zonal mean as shown in Fig.1 as well as Fig.S1(b) (also as Fig. R1(b)), zonal-mean CO50 concentration is lower at lower levels due to large compensation by low CO50 concentration outside the convection plume, and the zonal-mean concentration is higher at higher levels due to smaller zonal tracer gradient. Although interesting, this argument on the "secondary maximum" is a side note and we decide to keep the statement in a concise format with the newly made Fig. S1 for a better illustration.


(a) CO50 @ 30N

(b) ZM CO50 @ 30N

(c) Normalized CMF (500 hPa) & SRC @ 30N

Fig. R1 Relations among emission, convection, and CO50 concentration during JJA: (a) longitudinal distribution of CO50 at 30N for various vertical levels from 200 hPa to 500 hPa; (b) vertical profile of zonal-mean CO50 concentration at 30N highlighting the values at a few vertical levels that are shown in (a); and (c) 500 hPa CMF (black) and surface CO50 emissions (gray) at 30N.

Page 5 Line 5-6: From Fig. 2, it is not obvious that the inter-model range in JJA peaks around 400 hPa. It seems the JJA inter-model range remains relative constant at all levels below 400 hPa.

Yes, we have revised the text as suggested.

Page 8 Line 4: why the word “strong” needs to be quoted?

We removed the quote.

Page 21 Table 2 caption: Why some variables are scaled? For the SD simulations without a superscript note, are they using a common nudging time scale? What is the difference between the “replay” simulation and the standard nudging simulation?

Some variables in Table 2 are scaled because a few models have not comply with the standard boundary condition of NH50 mixing ratio as 100 ppbv. For example, the boundary condition is set as 100 ppmv in the GEOS-C1/C1SD and therefore needs to be scaled by 0.001. Similarly, CMF output in ACCESS-C1 and NIWA-C1 is multiplied by the gravitational coefficient g ~ 9.8 [m/s^2] and thus a scaling is applied.

Nudging time scales in SD simulations actually vary from different modeling centers, and not as a common time scale (say 50 hrs). We only highlight the two in the WACCM C1SD because WACCM is the only model that implement two different nudging time scales. We decided to remove these footnotes on nudging time scales and instead state them in the caption.

The “replay” simulation involves reading in analyzed field every 6 hrs and recomputing the analysis increment using the same assimilation methods that produce the MERRA reanalysis. Details on the “replay” approach are referred to Orbe et al. (2017).

Orbe, C., Oman, L. D., Strahan, S. E., Waugh, D. W., Pawson, S., Takacs, L. L., & Molod, A. M. (2017). Large-scale atmospheric transport in GEOS replay simulations. Journal of Advances in Modeling Earth Systems, 9. https://doi. org/10.1002/2017MS001053

Page 22 Fig. 1 caption: panel (e) and (f) do not show the “meridional and vertical transport of CO50”. They only show the zonal mean CO50 concentration cross sections.

Revised as suggested.

Reviewer #3

This is an interesting paper aimed at understanding the cross-model variation in the transport in to the Arctic of tracers with mid-latitude sources. Following several other works that have pointed out the diversity in Arctic abundances of these tracers, this work points out convincingly the importance of transport by the zonal mean, meridional circulation for understanding this diversity in the CCMI multi-model ensemble. An important wrinkle to this is that this meridional transport differs even across models that specify the meteorology in some way. Although the role of the mean meridional circulation does not explain all aspects of inter-model disagreement regarding large-scale transport into the Arctic, it seems to be a significant contributing factor, and well worth highlighting.

I find this to be an interesting study well worth publishing. I have some questions about the methodology and some suggestions about presentation, but if these are addressed I would recommend publication. I also found the writing to be a bit difficult to follow at times.

My only more significant comment is regarding the choice of flux decomposition into a zonal mean an eddy component. I was initially somewhat surprised at this choice, rather than decomposing into a stationary (say,monthly mean) and time-varying component, as it seems the latter would be better suited for understanding the local poleward transport in the Pacific sector. The time mean would filter out transport from synoptic scale eddies just as well. I would have thought the zonal mean meridional velocity would be just as, if not more, relevant to NH50 transport as to CO50, but that doesn't seem to be the case (I appreciate the CMF argument and this may well be the story). Perhaps the zonal decomposition is easier to study with the model output available (though this isn't so clear to me). Have you tried this decomposition?

The aim of the decomposition into a zonal mean and an eddy component is to differentiate the roles of mean meridional circulation and atmospheric waves respectively in the poleward transport of tracers. Given that the focus shifts to the zonal-mean component later in the paper, it is also easier to implement the analysis to a larger group of CCMI models with only monthly-mean output. Instead, for a temporal decomposition into a stationary and a transient component, this would be more applied to just the eddies since the zonal-mean tracer flux is mostly contributed by the stationary component, and also the analysis can only be implemented in the GEOS simulations in which daily output are available. As we have shown in Fig. 8, for the total tracer flux difference between GEOS-C1 and GEOS-C1SD, the contribution by difference in the eddy component (stationary + transient) is minor. Therefore, by implementing the temporal decomposition, even if a contribution by synoptic eddies is found, that contribution must be largely cancelled by stationary eddies. This is confirmed in Fig. R2 that we show four components of both temporal and zonal decomposition (compare red dashed for stationary eddies and blue dashed for transient eddies). Note that we have modified the vertical integral range of tracer flux to 800 – 950 hPa (to be consistent with the range where we examine v and CMF) in the revised paper.

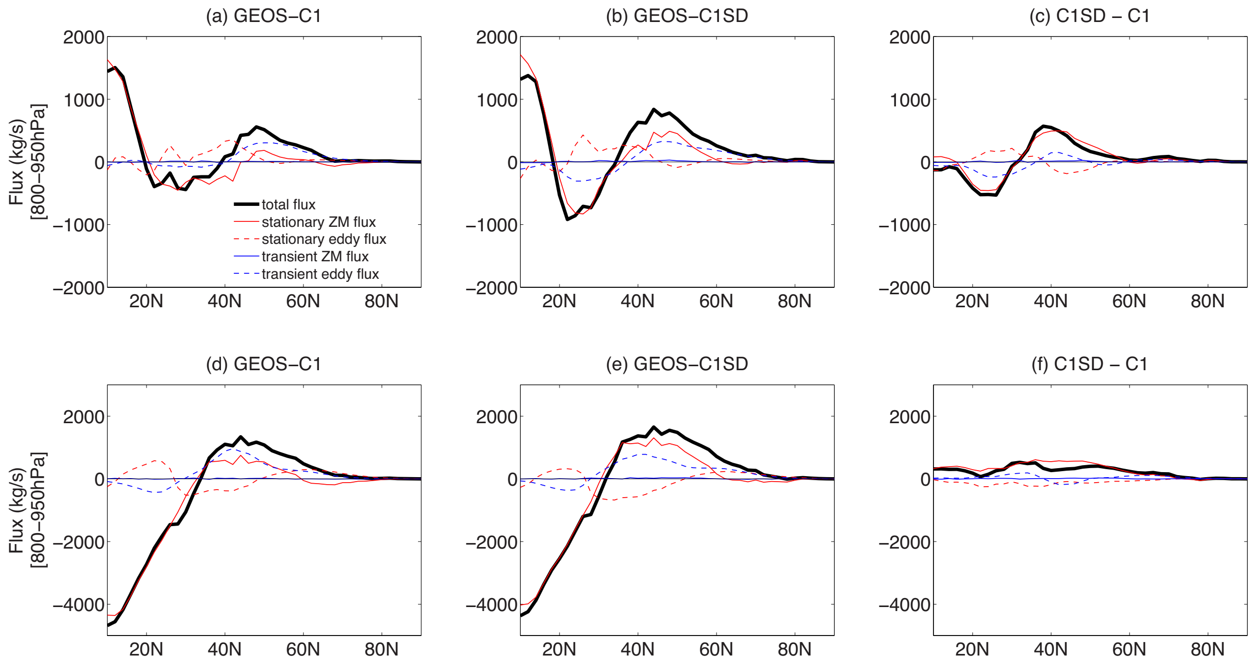


Fig. R2 Similar to Fig.8, but showing all four components from both zonal and temporal decomposition. The four components are: stationary zonal-mean flux (red solid), stationary eddy flux (red dashed), transient zonal-mean flux (blue solid), and transient eddy flux (blue dashed).

Given that no substantial changes are made to conclusions in the paper if imposing an additional temporal decomposition, we decided to keep the discussions on the zonal decomposition.

On a related note, it seems that a second way to quantify the contribution of the meridional transport (beyond inter-model correlations) would be to directly calculate the tracer flux from the time mean (or zonal mean) meridional winds. I would expect this to have substantial variations even in CO and in NH50, even if there are confounding effects that reduce the correlations with overall Arctic burdens. This could be quite helpful for quantifying the importance of this effect for the transport of realistic tracers. Possibly one could even try isolating the role of inter-model differences in meridional velocity by computing the flux with a reference tracer distribution.

We appreciate the idea of directly calculating tracer flux and we have done so for the CO50 tracer (see Fig.9 b,d). As for NH50 and CO, we have not explicitly calculated the zonal-mean flux but we are highly confident that the tracer flux will deliver a similar story as diagnosed by the correlation between Arctic tracer concentration and the Hadley Cell northern edge ($\phi_{v=0}$). Given that NH50 and CO are not the primary foci of the paper, we decide to keep the current form of analysis.

With regards to the presentation, I wonder if it would help to better distinguish between model integrations with constrained meteorology and those with free running dynamics. I have included specific suggestions below.

Please see responses in the “specific comments”.

Specific comments

p3 l29 - Are the source data for this interpolation really on the model native grid? I would be surprised if any model runs in isobaric coordinates rather than hybrid pressure (this seems to be confirmed by Table 1). I suspect it's more likely that the source data have been interpolated by the modeling centers.
Yes, we have modified the text to make this issue more clearer.

p6 l10-11: I was confused by the sentence starting 'However, it is difficult....' Is the point that these integrations are different because of how they implemented the source of CO50, but that they may also differ in their transport?
We apologize for the confusion. We mean that, in EMAC simulations, it is difficult to reduce the Arctic CO50 concentration uncertainty caused by biases in the CO50 emissions with a simple scaling due to complexity in the poleward transport. We have modified this sentence accordingly.

p6 l15: I can see how this source variability increases the model range in winter, but the EMAC models are right in the middle of the model distribution for summer Arctic burdens so the latter assertion wasn't so clear to me.
As discussed above, even inputting equal magnitude of biases (with opposite directions) in CO50 emissions between winter and summer, the responses in the Arctic CO50 concentration may still differ between seasons due to different patterns of long-range poleward transport. As illustrated in Fig.1, poleward transport of CO50 are more along isentropic surfaces during winter while the transport displays a robust convection-related feature during summer and transport at higher levels might also be important. Alternatively speaking, during winter, the midlatitude surface source and the polar middle and lower troposphere are more directly linked via slopes of isentropic surfaces while this connection via relatively fast isentropic mixing seems to be weaker during summer. Therefore, biases in the midlatitude source can be more clearly seen in the Arctic CO50 concentration during winter than summer. Since the issue discussed above is not closely relevant to the statements in P6L15, we decided to make no revisions.

p7 l20-24: I wasn't so convinced by this argument - even if the transport into the Arctic is weaker in the summer, if the sources are at midlatitudes, this transport still seems essential for determining summer burdens, no? Unless a large fraction of the summer is left from winter-time transport, in which case this is probably worth discussing, since transport in the winter and spring would be more important than summer time transport.
Yes, the statement here is mostly speculative. We only provide one plausible explanation on why we have not seen a robust correlation between Arctic CO50 concentration and midlatitude convection during summer even when there is a collocation between summertime convection and CO50 source over lands. It is that isentropic transport is much weaker during summer. Although CO50 concentration over regions slightly higher above the midlatitude source can be highly modulated by variations in the parameterized convection. This variation at the lower end of isentropic slopes can be less dominant in influencing the Arctic CO50 burden at the higher end of isentropic slopes with less efficient isentropic transport during summer. Other processes may also be influencing resulting in the worse correlation between midlatitude convection and Arctic CO50 concentration during summer seen in Fig. 3b. We think the current statement is clear enough so decide to make no revision. As to the reviewer’s comments on leftover effect of

winter-time transport on summertime Arctic burden, we think the chance is smaller as the lifetime of the tracer is 50 days, which means that tracers are well chemically diluted at a seasonal scale. Therefore, it is reasonable to assume that variations in the Arctic tracer burden are by majority contributed from variations in the transport during the same season.

p9 l12: This is inconsistent with the caption of Fig. 8 that gives an upper bound is 200 hPa.
Thanks for pointing out the typo. Note that the integral bounds have been changed to 800 – 950 hPa in the revised paper.

p12 l15: By 'unchanged' do the authors mean that the model anomaly from the ensemble mean NH50 concentrations at high latitudes is consistent with its anomaly at midlatitudes?
Yes.

For all figures I found myself wishing there was a clearer sorting in the legends by specified versus free dynamics. I don't think it is said anywhere explicitly that the open circles are specified dynamics integrations. Also, where correlation coefficients are given it would be informative to show coefficients for only the specified/free dynamics runs. Some of the regression lines (e.g. Fig. 10c) seemed to be obscuring clear differences in behavior; similarly in panels 12d,e,h, for instance, the two sets of integrations seem to follow rather different regression lines. This is discussed at some points in the text, but the presentation could be made clearer.
We appreciate the reviewer's suggestion on sorting the legends by C1 versus C1SD. We have not revised the legend but instead made statement in the captions when firstly noted to make it clearer the difference between lines and marks used to differentiate C1 and C1SD simulations. As to correlation coefficients, we prefer the current way for all models although sometimes such a way could be obscuring. As discussed in the revised paper, the correlation coefficients are only for referencing given that they are derived based on climatologies from a limited number of simulations. Also, we have indicated the difference in correlations based on all simulations versus that is only based on the C1 simulations in Table S1, and we have highlighted this point more throughout the revised manuscript.

Fig. 2,4, and similar panels in the following: Would it be helpful to show ensemble means separate for specified dynamics and free running models?
We aim to examine C1 and C1SD simulations together in these figures, and by showing the ensemble mean for all models, it is easier to derive the multi-model fractional spread (of CO50 concentration) noted in the paper. We tried to explicitly show the separation between C1 and C1SD simulations for the CO50 concentration in Fig. S2. Also, adding more lines may not help in distinguish different lines for different simulations.

Fig. 5: In this figure especially, it is hard to take in these 16 panels - sorting them by specified versus free dynamics would be more useful than by modeling center.
This is a brilliant idea. We have re-arranged the panels and make the eight C1 simulations at the top followed by the other eight C1SD simulations at the bottom.