The authors would like to take the opportunity to thank the reviewers for their comments and for taking the time to offer them. We believe the manuscript has been improved with the helpful input.

Response to Reviewer 2

The authors use results from the WACCM model together with observations of CO from two sensors to investigate how well descent rates can be derived from a chemically (nearly) inert tracer with a strong vertical gradient in the altitude range 45-85 km (i.e., the mesosphere). It is found that considering corrections due to horizontal advection, turbulence, and chemical loss can imply differences in the descent rates derived from CO of more than 1 km/day particularly around strong sudden stratospheric warmings. Credibility is provided by a comparison of the modelled CO to the two observation data sets which generally show a good agreement. Considering that descent rates derived from these methods mostly lie in the range of 100-300 m/day, this is quite a large margin of error. Inert tracers are widely used to derive descent rates in the polar winter middle atmosphere – not only in the mesosphere to estimate the input of thermospheric tracers, but also in the stratosphere to derive chemical ozone loss rates – and the paper provides an important caveat for these methods. I found the paper very clearly structured and well written, and recommend publication in ACP with a few minor changes.

Page 10, line 12: what does it mean if "w* corrected" derived from modeled CO using corrections from the model itself does not provide the model w*? If equation 1 is a correct description of all terms affecting CO in the model, then "w* corrected" should provide w* in a self-consistent way. I would say that this means that the terms in Equation 1 do not reflect what the model does to CO. I would expect that in the model, the resolved eddy term (Xedd) is not treated separately but as part of the advection scheme; in which case it is counted double if subtracted for derivation of "w* corrected". Does this make sense?

The terms in equation 1 do represent the changes in CO VMR, but in the Transformed Eulerian Mean (TEM) representation of the atmosphere. It is correct to say that what the model "does" in simulating the atmosphere is not the same as following the TEM equations. Equation 1 is a highly derived equation (Andrews et al., 1987), the terms of which are calculated using the output from SD-WACCM.

The TEM offers a way to represent the atmosphere as an interaction of a mean flow with disturbances, i.e., eddies and waves, imposed upon it, and it contains a description of a mean meridional flow in the atmosphere. The interaction is generally a two-way process and the disturbances feed back into the mean flow through non-linear effects.

So the calculation of w*, using the TEM formalism and using CO VMRs, will have some differences due to the way that each is calculated. The calculation of w*_corrected is also somewhat crude (as mentioned in the paper and now expanded upon in the new manuscript), as it combines tendencies in the TEM formalism with values derived using CO VMRs in the atmosphere. These points are now emphasized in Section 4 and it is made clearer that the goal

of w*_corrected is to get a qualitative estimate of the errors that may be incurred by assuming pure vertical advection when using tracers to calculate w*.
"The resulting rate is called w_{CO} $corrected$*. This could be considered a crude approach, combining daily averaged CO output with CO tendencies calculated using the TEM formalism, but the aim here is to provide an estimate of the errors that may be incurred by neglecting influences on CO other than vertical advection. In any case, the results involving* w_{CO} $corrected$ *are discussed in a qualitative manner, instead of for quantitative error analysis."*

The discussion in Section 6 offers some other reasons for the discrepancy: mainly parameterization of gravity waves in the model, which play a role in the strength of w* and also in the parameterised eddy flux divergence (Xkzz), and the time resolution of the model output. Meraner and Schmidt (2016) found differences in calculated w* when using 6-hourly average model output compared to daily average output.

As I understand the term, the middle atmosphere comprised the stratosphere and mesosphere. As you really focus on the mesosphere here (the altitude range from 45-85 km) you might want to change the title of your paper to "polar mesospheric descent".
The term was chosen because altitude of the stratopause can be around 55 km, which would mean that a 10 km layer of the analysis is within the stratosphere.
Because the analysis does not cover the whole of the middle atmosphere, the first line of the abstract has been edited to express the altitude range. Combined with the title, the reader will now immediately know the area under investigation: *"We investigate the reliability of using trace gas measurements from remote sensing instruments to infer polar atmospheric descent rates during winter in the 46 – 86 km altitude range."*

The altitude range is explicitly mentioned in the conclusions now also.

Page 1, lines 16-17: "The relative importance of vertical advection is lessened : : :" that means that other processes become more important, could you add a sentence which? (Turbulence, horizontal advection,: …?)
The sentence has been edited to read:
"The relative importance of vertical advection is lessened during periods directly before and after a sudden stratospheric warming, mainly due to an increase in eddy transport"

Page 1, lines 25 and following: dynamical tracers have also been used (quite extensively) to derive stratospheric descent rates: to distinguish chemical ozone loss from dynamical processes.
The following statement and references have been added to Section 1: *"This technique has often been used in combination with ozone measurements to separate chemical and dynamical influences when determining ozone depletion (e.g. Proffit et al., 1990, 1993; Müller et al., 1996, 2003; Salawitch et al., 2002; Rösevall et al., 2007)."*

Page 2, line 2-3: you could also reference Funke et al, 2014a, b; and Funke et al., 2017.

These references have been added to the main text and to the references section.

Page 2, line 16: … "and photochemical destruction in the upper mesosphere" limits the altitudes at which it can be used to the stratosphere and lowermost mesosphere.
We would prefer not to provide such a definition because the referenced paper (Lee et al., 2011) does not make such a definite statement.
Lee et al. (2011) states: "However, it is not as good a tracer as CO for diagnosing vertical motions throughout the middle atmosphere due to the relative complexity of photochemical sources and sinks in the stratosphere."

Page 4, section 2.2, line 6: can you also state the approximate altitude range of MLS (in km)?
The sentence in question now reads: *"These CO profiles cover a pressure range of 215 – 0.0046 hPa (approximately 11 – 86 km) …"*

Page 4, section 2.3, line 28: What exactly does daily output mean – once per day at a specific global time (a global snapshot with varying solar zenith angle) or at a specific local time (a global snapshot with nearly fixed solar zenith angle), or output of daily averages? For a dynamical tracer this probably does not make a big difference apart from some impact of the tidal phase in the upper mesosphere.
The model output Is daily averages. This sentence has been edited to read: *"Model output of daily averages from 2008 to 2014 are used for this study."*

Page 5, lines 1-2, discussion of Figure 1: Figure 1 is too small – in my A4 one page per page printout each panel is about 1 cm high, making it very hard to distinguish the lines. You could more than double the vertical range of the panels without filling the page. Please do.
The panels have been enlarged, and superfluous date labels have been removed to give more space for the panels. The figure has also been changed to landscape layout so that the panels can be made larger. The background was also changed to a light colour at the request of Reviewer #1.

Page 5, line 7: ": … but a systematic change in the results … isn't found …" despite the very cramped figure (see my previous comment) I do see a systematic difference between MLS and WACCM in early and late winter, i.e., in the buildup and decrease of the winter maximum: the winter maximum starts earlier and lasts longer in WACCM than in MLS, at least above 66 km.
That particular sentence regards the difference of using bilinear interpolation of the model data, or not.
However, your point about the systematic difference between the model and data remains valid. The following lines have been added to the section to make the point: *"A systematic difference is evident between MLS and SD-WACCM during the times of year when CO VMRs are rapidly increasing or decreasing, with SD-WACCM showing larger values of CO. The difference is*

most pronounced at higher altitudes and is predominantly during August/September and April/May."

Page 6, line 2-3: "The Prandtl number is 2 for the model runs in this work" I am not quite clear what this means. My understanding is that the Prandtl number describes a physical property of a gas or liquid, namely the relation between momentum diffusivity and thermal diffusivity; as such it should be an exact quantity. The Prandtl number of gases is usually given as lower than 1; for air, values around 0.7-0.8 are given. Does this change around the mesopause (where molecular diffusion becomes more important) compared to the lower atmosphere, or is this really used as a scalable fudge factor in WACCM? – I am aware that this is a feature of WACCM which has been implemented for a good reason; I'm not suggesting that this is changed. I am just curious what it means.

The Prandtl number (Pr) as used in GW parameterizations describes the ratio of momentum flux to heat flux, and is properly thought of as a "turbulent Prandtl number"; in particular, it is not a property of a gas or liquid, but of the process whereby gravity waves dissipate when they "break".

For breaking gravity waves, we really do not know what Pr should be, so to some extent it has been used as an adjustable parameter in GW parameterizations. This is the way Pr is used in WACCM and, in fact, in all models that parameterize GW breaking and attempt to derive turbulent mixing due to such breaking.

Section 3.1 in the edited manuscript now clarifies this with the following sentences: *"The value of K_{zz} calculated with SD-WACCM depends, among other things, upon the Prandtl number (or more properly, the "turbulent Prandtl number"), which describes the ratio of momentum flux to heat flux. The Prandtl number is a property of the process whereby gravity waves dissipate when they "break" (see e.g., Fritts and Dunkerton, 1985, for a more details). The Prandtl number is 2 for the model runs in this work (see Sect. 6) and is used in SD-WACCM to parameterise gravity wave breaking (Garcia et al., 2007)."*

Page 6, line 5: The terms of Eq. 1 are "renamed" here. "Rewritten" suggests that you adapted the terms mathematically.

This has been changed.

Page 6, lines 18 and following: I found it quite intriguing that air parcels ending above 66 km actually have their origin in the summer hemisphere. Maybe you can add a mention of this here.

Section 3.2 has been edited and now contains the following information: *"The magnitude of the TEM wind is larger for the higher altitudes, as also shown in Smith et al. (2011), and the air parcels that arrive above 66 km altitude originate in the summer hemisphere. The parcels that arrive below this, which could be considered as part of the Brewer Dobson circulation (Brewer, 1949), originate at latitudes closer to the equator."*

Page 8, line 17 and following, discussion of Figure 4: again, I was intrigued to see that differences between wco and wco corrected sometimes are larger than 1 km / day: two

to ten times larger than (most) estimates of descent rates based on tracers as given in Table 1. That really is a big discrepancy.

The differences between w* and w*_corrected can be quite large, but it is hard to compare them quantitatively to the values listed in Table 1. There are two points here:

The first point is that the differences between w* and w*_corrected were used to provide more of a qualitative estimation of the errors that can be incurred by neglecting processes other than vertical advection. This is mainly due to the crudeness of combining information derived in two different ways: changes in daily averaged CO from the model/instrument, and TEM tendencies. Section 4 of the edited manuscript now clarifies this point:

"The resulting rate is called w_{CO} $corrected$*. This could be considered a crude approach, combining daily averaged CO output with CO tendencies calculated using the TEM formalism, but the aim here is to provide an estimate of the errors that may be incurred by neglecting influences on CO other than vertical advection. In any case, the results involving* w_{CO} $corrected$ *are discussed in a qualitative manner, instead of for quantitative error analysis."*

The second point is that the values in Table 1 come from a variety of analyses, some of which are averages in space, time, or both. Averages in altitude generally provide lower values for w* than those that are seen at altitudes above about 70 km.

Straub et al. (2012), for example, quote a value of w* of ~ 325 m/day, from instrument, model, and trajectory analysis. The value is an average over Feb/March and also over an altitude range of 0.6 hPa - .06 hPa (approx. 52 – 68 km). The modelled and trajectory analysis w* profiles, from which the averages are made, often show values of ~1200 m/s at 0.06 hPa.

Table 1 has been edited to state where averaging has been performed, and also the altitude ranges that were used in the studies. A point is now made in Section 1 about noting the effect different averaging techniques.

"The altitude range over which the rates were determined, and whether averaging was performed, is also shown in Table 1. It is important to note that an average over altitude can mask the higher descent rates that are found in the mesosphere. For example, Straub et al. (2012) show a descent rate of -325 m/day from averaged modelled wind profiles, between 0.6 hPa (~52 km) and 0.06 hPa (~68 km), whereas the individual wind profiles often show descent rates larger than -1000 m/day at 0.06 hPa.

."

Page 10, lines 7-12: here you compare w* from the model (Figure 8) with values derived from tracer observations (Figure 4) – it would certainly be easier to follow your argument here if a) the panels in Figure 4 were larger, and b) more importantly, the scale of the colour bars was the same in Figure 4 and 8. It is difficult to appreciate that the values of "w* corrected" provided by tracer observations in 60-90 (Figure 4) is really smaller than the values provided from model wind fields in Figure 8, as the scale in Figure 8 actually covers a smaller range (-1 to 1 km/day compared to -2 to 2 km/day in Figure 4).

Figure 8 from the current manuscript has been edited to have the same colour limits as Figure 4. The scenarios from Figure 4 (spot, zonal mean, and polar mean) have been split into separate figures and the panels made larger so that the plots are clearer.

Page 10, line 18: see my comment above – what does a Prandtl number of 2-4 mean?
This is addressed above in response to the referenced comment.