Reply to Referee 1

We thank the referee for their positive review. The answers to the two questions posed are as follows:

1. The flux limiter implemented in TOMCAT is that described in section 4 of Prather (1986). A comment has been added at the beginning of section 4.5 to highlight this point.

2. We have not attempted a test case with increased vertical resolution, as the vertical resolution is significantly more difficult to change in TOMCAT. However, we do not see any reason why increasing the vertical resolution would cause any loss of accuracy of our algorithm. The column-matrix formulation of the parametrisations means that no essential change to the adjoint formulation is necessary when vertical resolution is increased.

Reply to Referee 2 (Daven Henze)

We thank Dr. Henze for his very thoughtful and careful review. A point by point response follows:

1. *Re: Wilson et al.* We did not become aware of the Wilson et al. submission until after our ACPD manuscript was typeset. Our view is that they have successfully created a sound and practical adjoint (and associated inverse) chemistry transport model in the 'adjoint-of-finite-difference' (AFD) tradition. Their model has advantages over RETRO-TOM, in that they have gone on to implement iterative methods for solving fully nonlinear inverse problems (INVICAT). In our opinion this is a significant achievement.

   However, ATOMCAT (and by extension INVICAT) also have some disadvantages relative to RETRO-TOM. For example, ATOMCAT uses stored data of every chemical species from a corresponding forward run at every time-step, which although clearly necessary for non-linear inversion in the most general case, renders it unsuitable (or at best cumbersome) for applications in which RETRO-TOM is used as an alternative to a back trajectory calculation i.e. in a situation where the relevant chemistry can be modelled by a linear process (see Haines and Esler, 2014, a companion paper now referenced in the manuscript, for an illustration of how RETRO-TOM can be used in this way).

In addition to the numerical advantages discussed in the manuscript, RETRO-TOM also uses the essential machinery of TOMCAT (i.e. the advection scheme, and convection routines) with very little modification, which is a clear advantage for new users of the model. Finally, it is also far from clear in Wilson et al. how ATOMCAT handles the nonlinearity associated with flux limiters, which is well-understood to be an issue for adjoints of advection schemes.

With the above points in mind, we have added the following paragraph to the conclusions (a brief reference is also made in the introduction). "A recent development is that an alternative adjoint model for TOMCAT (ATOMCAT, with associated nonlinear inverse model INVICAT), based on the AFD framework, has been under concurrent development by the Univ. Leeds group (Wilson et al. 2013). The ATOMCAT model has an advantage over RETRO-TOM in that it is coupled to the INVICAT model, which is designed to solve nonlinear inverse problems by iterating forward TOMCAT and adjoint ATOMCAT calculations. As a result, at the present stage of development ATOMCAT / INVICAT are suited to a wider range of applications compared to RETRO-TOM. However, ATOMCAT also has certain disadvantages relative to RETRO-TOM including the disadvantages of AFD numerical schemes detailed above, which may be particularly severe when flux limiters are in use (see discussion above). It is also likely to be the case that RETRO-TOM is much better suited as an alternative to Lagrangian backtracking (e.g. Haines and Esler, 2014), because ATOMCAT requires output from a forward calculation at every time-step, making it difficult to use in problems that are formulated without reference to a forward calculation. Finally, RETRO-TOM has the advantage that, unlike ATOMCAT which relies on code generated by differentiation of TOMCAT's forward code, RETRO-TOM utilizes the machinery of TOMCAT itself (e.g. Prather advection scheme, column matrix approach to parametrisations), making it relatively straightforward for users to adapt and update in conjunction with the forward model. An interesting topic for future study is the question of whether the ATOMCAT 'dynamical core' can be replaced by that of RETRO-TOM without any degradation in model performance, thus creating a full inverse model with the advantages of both schemes."

2. *Re: 1482.7*: We have changed this sentence to include a working definiton of accuracy.
"The numerical accuracy, defined here as the relative difference be-

tween sensitivities calculated by a linear perturbation to a forward calculation and those obtained from the adjoint, as well as the reliability of adjoint models are evidently key to the above applications."

3. *Re: 1483.17-20*: We have added the following comment:
   "In particular, several authors (e.g. Hourdin et al. 2006, Henze et al. 2007, Gou and Sandu, 2011) have noted that the AFD adjoints to nonlinear advection schemes can lead to undesirable results and poor performance."

4. *1485.14*: We agree that Wilson et al.'s reciprocity test is nice, but the accuracy of AFD adjoints is not really in dispute here (only FDA ones). We have nevertheless added the sentence
   "The recent AFD scheme of Wilson et al. (2013) also uses Prather's scheme, and has demonstrated accuracies close to machine precision, but could be subject to the problems of AFD adjoints to nonlinear advection schemes highlighted by e.g. Gou and Sandu (2011)."

5. *Re: 1485.20*: We have changed the text to add a reference to this referee comment (note also *possible* $\rightarrow$ *likely*)
   "Henze et al. (2007) show significantly larger errors for a 2 day integration, although it is likely (see referee comment by D. Henze) that their poorer results are predominantly due to the difficulties of generating an adjoint to the nonlinear piecewise parabolic scheme...".

6. *Re: 1493.25*: As is clear from the companion paper (Haines and Esler, 2014, GRL) meaningful and interesting adjoint problems can be formulated that do not require explicit solution of the forward integration. Obviously, it is still desirable that RETRO-TOM in these cases gives the same sensitivity as corresponding forward runs (if one was to execute them). To get this we need the density at intermediate timesteps between forcing files. Hence, RETRO-TOM is designed to make an economical calculation to get the intermediate time-step densities in the absence of any forward model integration. (We have nothing against check-pointing more generally, we just want to be able to use RETRO-TOM in this single integration mode, as in Haines and Esler).

7. *Re: 1487*: Without giving more considerably more detail about the operator $\mathcal{C}$, it would be a little messy to give the non-density weighted inner product result. Instead, we have added the reference
   "Compare, for example, eqns. (1a) and (5a) of Sandu et al. (2005)."

8. *1497.15*: In TOMCAT and RETRO-TOM there is still an ordering of operations on the concentration (or retro-concentration) field (c.f. Table 2 of Hakami et al.). This test is simply to determine the order of magnitude of the error if the order of the operations is not correctly reversed in RETRO-TOM.

9. Re: Consistency between Prather and assimilation model. This is obviously an important issue, but note that TOMCAT is an offline model not usually used in association with a GCM. TOMCAT is typically forced with ECMWF meteorological products (e.g. operational analyses, ERA-interim etc.), and certainly the ECMWF model has its own advection scheme for e.g. water vapour and other tracers used in generating the analyses. The extent to which this is problematic is somewhat beyond the scope of the present work.

10. *1500.16*: Extra references added.

11. *1500.26*: I think that the issue here is over the definition of accuracy. Given our definition (see comment 2) our statement that "*accurate* adjoint calculations are rendered near- impossible by advective nonlinearities such as flux limiters" is correct. But inaccurate here does not mean impractical, and we did not mean to criticise (implicitly or explicitly) Henze et al. or any subsequent works using GEOS-CHEM. Quite the reverse in fact, the most important point here is that the best possible compromise is obtained using FDA rather than AFD. To clarify matters, we have changed the sentence to read
" Based on the above, our view is that it is near-impossible in the presence of advective nonlinearities such as flux limiters, to obtain adjoint sensitivities that are both accurate (in the sense defined in the introduction) and practical (in that they do not contain spurious and unphysical sensitivities due to the scheme's nonlinearity). "

12. *1501.2 and 1503.20*: The flux limiters used in TOMCAT act to preserve positivity only, not gradients of plumes. However, it is a fair point that if a localised source is sufficiently strong, then flux limiters may still be necessary even if there is a non-zero background. We have added the caveat
"and their localised sources are not too strong."

13. *Re: 1501.7*: The temporal extent of the sources and receptors here are the same as in test cases I-III (e.g. 10 days for test case I). We agree that we have not been clear here, and have changed the text to read

"...the 'worst case scenario' for the spatial stucture of the source, when it is isolated to a single grid-cell...".
We have made a few tests with sources and receptors restricted to single time steps, without significant deterioration in our results.

14. *Re: 1501.25*: We agree that this is a useful alternative way to think about the differences between flux limited and non-flux limited simulations.

15. *Minor Corrections*: Implemented. Thank you.

Other Significant Changes

1. Conclusions: Reference to Haines and Esler (2014, GRL) added.

2. Pg. 1489 l. 5: Reference to Hoyle et al. (2011) added.

3. Section 4.3, first sentence. 'and to assess error growth with time.' Phrase removed.