Review of “A stratospheric prognostic ozone for seamless Earth System Models: performance, impacts and future” by Monge-Sanz et al.

In this study the authors evaluate the implementation of a new stratospheric ozone model in the ECMWF system, both in terms of the fidelity of the simulated ozone and in terms of the resulting ozone feedbacks on radiation and dynamics in the model. Using a combination of both medium-range (10 day-long) and long-range (7 month) ensemble forecasts, the authors show that the new ozone linearization produces enhanced results, compared to the default ozone implementation in the IFS system. Quite importantly, this improved performance is exhibited in extreme dynamical conditions such as stratospheric sudden warmings, with improved coupling to the surface (manifest in correlations in the NAO). Overall, the manuscript is well motivated, well written and thorough. Therefore, it is, in my opinion, deserving of publication pending minor revisions. At the same time, however, there are several points where certain details of the new ozone implementation need to be clarified, specifically as they compare to the “default” implementation. These clarifications are needed as they may impact the extent to which certain purported “improvements” in the ozone tracer (often described throughout the manuscript as resulting from a more complete implicit incorporation of heterogenous chemical processes) may actually reflect other differences between the schemes (i.e. resolution of the corresponding photochemical box model, etc.).

Major Comments:

#1: I find it somewhat misleading when the authors claim that the BMS scheme is a fundamentally “new model” compared to the CD scheme. Perhaps this is just an issue of semantics, but if my understanding of the differences is correct isn’t the main point the fact that the tendencies for the BMS scheme are derived from a box model with updated chemistry (i.e. one that includes heterogenous chemical effects), unlike the default scheme where these processes need to be included as an additional (crudely parameterized) term? It seems to me, therefore, that it’s not really true that the linearization itself is fundamentally changed (except for the exclusion of a term that is now not needed); rather, what is new is the source of the tendencies that are being input to the linearization. I suppose that this modified input (now from a 3-D CTM with an improved chemical mechanism) constitutes part of the model and therefore represents a “new model” but I would think a more conventional interpretation of “new model” is one that involves additional terms to equations (1-2) and/or changes in functional dependencies (i.e. not simply changes in the input). My suggestion here is to be a bit more clear that what is meant by a ”new model” is really mainly a reflection of new tendencies that derive from a parent GCM that now happens to incorporate heterogenous chemical processes. The functional form of the linearized ozone tracer itself has not really changed.

#2. Since one of the main points of the study is to demonstrate the improvement of the new ozone tracer compared to the default approach, more details are needed about the exact differences between the BMS and CD approaches. Therefore, while it may be true that some of the details that are wanting can be found in Monge-Sanz (2011) I feel that a subset can be repeated here. The reason for this is that the authors tend to attribute most of the improved ozone performance to the incorporation of implicit heterogenous chemical effects but there may be other factors at play as well. In particular, what is the vertical resolution of the 2D photochemical model in the default ozone implementation? The coefficients for the BMS scheme are derived from the box model version of a 3D CTM. They are calculated at 24 levels (and then presumably interpolated somehow to the native vertical grid of the IFS forecasts, although this detail is not provided). Are the coefficients from the default ozone linearization also available at these levels? If the treatment of vertical gradients is still more simple in that model it is not obvious to me that the BMS and CD main differences reside in the treatment of heterogeneous chemistry. Please, I strongly suggest the authors add more details to Section 2.1 that address resolution (and forcing set) differences between the box models separately used to produces the coefficients for these different ozone tracers. Moreover, speculate on how these aspects of the ozone tracer linearization may be contributing to the differences seen (in the conclusions).

Another point that needs clarification in Section 2.1 is the meteorological fields that are used to initially derive the coefficients in both the BMS and default ozone linearization. What is the limitation of using coefficients based on box model runs using input meteorology for one representative year? If the results are based on coefficients on a year with mild temperatures one might expect that this would restricts the ability of the scheme to reproduce realistic ozone loss during colder winters and vice versa. Is this an issue for both ozone tracers?

#3. Major Comment: One issue that is not discussed is the increased computational expense of the BMS scheme, which is introduced by needing to run the full 3-D CTM (and feed into the box model of the CTM to produce the coefficients). While this may be only a "one time" step it nonetheless is more expensive than simply running a 2-D box model using, for example, observed ozone climatologies that already exist. Indeed, the authors even suggest that it may be needed to constantly run the full CTM in order to provide a new "carefully chosen climatology term…from an updated run of the parent full-chemistry CTM." While that certainly will be good (to the extent that the CTM itself does a good job at reproducing observed ozone) it is nonetheless not a small computational ask. I suggest that the authors discuss this trade-off between performance and computational cost in the conclusions.

Minor Comments:

Line 29: A semi-colon (not comma) is needed after "exothermic reactions".
Line 99: A "the" is needed in "can simulate time".
Line 103: More description here is needed regarding the "tendencies derived from the full-chemistry CTM runs". My specific concern regards ambiguities about the forcings and boundary conditions used for the CTM runs from which these full tendencies are derived. In particular, over what time period are these CTM runs performed? More generally, what are the boundary conditions and compositional forcings used in the CTM simulations and how consistent are these with those used in the IFS forecast ensembles?
Line 122: The "to" after "sunlight" should be removed.

Line 133: I'm not sure I agree with this statement beginning with "In addition, the coefficients…". While it may be true that the coefficients are derived from a 3D model, what is used in the BMS scheme are the zonally averaged fields. mPlease reconcile this discrepancy.
Line 141: I am not sure I agree with the sentence "The IFS configuration in each pair … differs only in the scheme used to model stratospheric ozone." Is this really true? Later on in the manuscript, you write that the old scheme does not include radiative feedbacks from ozone so isn't this a key difference with the new BMS scheme? I understand that you ran experiments with the BMS scheme in which feedbacks are/are not incorporated but please indicate here that this is another key difference between the schemes.
Line 145: Given that the box model coefficients are inferred from a much coarser CTM output (7.5 horizontal resolution and 24 levels, according to Monge-Sanz et al. (2011)) how do you account for the increased resolution of the native model grid? Do you simply interpolate? Wouldn't it be worth running the CTM at a higher resolution (albeit with the additional expense) to get more realistic coefficients?
Line 164: Remove "an" before "horizontal resolution".
Line 166: Are these ensembles also launched in May and November of each year over the period 1981-2010? It is not clear when these are initialized.
Line 171: An "and" is needed before "the National Oceanic …".
Line 185: Experiment names (from Table 1) need to be indicated as it is not clear if the only difference here is the linearization scheme or the interactions with the radiation. Simply providing the experiment names will help clarify this. In particular, are you comparing exp001bms and exp001cd or exp002bms and exp001cd?
Line 200: Typo in "characterstics".
Line 209: What do you mean exactly by "the more realistic link with temperature in the new scheme"? Is this because of the implicit incorporation of heterogenous chemistry in the coefficients or are you comparing the schemes with and without feedbacks to radiation? Again, clarification about which exact experiments are being analyzed is needed. See my previous comment.
Line 216: Is there a reason why these experiments, which are described earlier, are being described again? Why the need to repeat these experiment details? Or are these different runs?
Line 229: I'm not sure "manifests" is the right word here.
Line 261: I'm not convinced that the use of the "a recent observation-based climatology" would necessarily be ideal. Isn't one of the main improvements afforded from the BMS scheme the fact that it better ensures internal consistency between the gas phase and heterogenous chemistry? To this end, therefore, wouldn't replacing the climatology term with one derived from observations break consistency with the other terms (which are derived from the CTM)? If consistency is a desirable feature, I would think that using a combination of model and observationally derived coefficients would be problematic.
Line 285: Again, why can't this comparison be cleaner? You should do either BMS w/wout radiative feedback (or using CD). What does that show? I suggest the authors replace this text here with a discussion of those results since this is not a proper apples-to-apples comparison.
Line 480: This is a more of comment that could be addressed in the conclusions but have you considered modifying BMS coefficient calculation to account for zonal variations in ozone? If this

is what is already meant by “the BMS scheme coefficients could be provided in a 3D grid” then please clarify that, indeed, this is the case.