We thank the referee for taking the time for reading our manuscript and their helpful comments!

General changes

- We have considerably changed the text throughout the manuscript to improve the logical order of the text and to improve the explanations and comprehensibility. We added new subsections and improved the use of the English language.

General comments

- **Combined reply to the following:**

  2. From the work presented it becomes obvious that validation, and specifically the validation of the core component — the residence times during convective updrafts — is very difficult.

  4. Admitting that the validation problem is largely inherent and not easily overcome, I think the paper could be acceptable if it would limit itself to a description of the algorithm implemented together with tests conducted so far, while including a clear characterisation of the limitations and the way how a more robust testing and/or tuning will be done, and making it at least plausible that the scheme will be superior to simpler alternatives.

  We agree that more discussion of these issues was needed. Currently, the large uncertainties in emissions, chemistry, microphysics and measurements of many short-lived species do not allow for a quantitative assessment whether our scheme improves the simulation of these short-lived species, even if this is suggested by the more realistic simulation of the time spent in convective clouds. Rather, our scheme allows for estimating the uncertainties in the simulation of these species associated with different parameterizations of vertical transport in convective updrafts. These uncertainties generally pose a challenge for the validation of the simulation of short-lived species, and there is a clear need to improve on this situation (as also noted by e.g. Forster et al., 2007).

  In addition, the globally constant lifetime of radon does not allow to validate the parameterization of the time spent in convective updrafts. Nevertheless, currently radon is probably still the species most suitable for the validation of convective transport models, since there is a lack of good alternatives.

  We have added discussion to section 4.4 on how well the results of other studies compare to radon measurements to put the comparison of our model to radon measurements into perspective. Other studies show differences between their models and the radon measurements of a similar order of magnitude (Jacob et al., 1997, Collins et al., 2002, Forster et al., 2007, Feng et al., 2011).

For many physical parameterizations in GCMs and CTMs there is no sufficient data for validation. The only way to make it more plausible that they are superior is to state that the physical assumptions are closer to reality.

Changes to the manuscript: We extended the discussion in the introduction and conclusions to discuss the large uncertainties in the validation of short-lived species as outlined above and to discuss the validation with radon. In addition, we added discussion in section 4.4.4 (section 4.2 in original manuscript) how well other models compare to the radon measurements and on the uncertainties in radon emissions, simulations and measurements.

- **2. The claim of the paper of a successful validation appears to be not sufficiently supported.**

We are aware that validation of the model is difficult and paid attention to a careful formulation of the results. The only occasion in the original manuscript, where we speak of an "successful validation" is at page 22, line 14 in the conclusions. This only refers to the technical part of the validation, i.e., mass conservation and reproduction of the convective mass fluxes and detrainment rates from the reanalysis. Since this part of the sentence is not really needed, we deleted it to avoid confusion.

Changes to the manuscript: Deleted "The algorithm is successfully validated by showing that" from the sentence.

3. The usefulness of the scheme in the context of the whole model will also depend on how well the chemical environment inside a convective cloud is actually modelled. The manuscript is not giving much attention to this aspect, which probably depends strongly on the model resolution (i.e. number of Lagrangian parcels). In addition, it should be compared to the option of just parameterising key reactions such as the heterogeneous oxidation in convective clouds.

The chemistry scheme is a part of the model which is independent from the transport scheme, and we think that a discussion of chemistry schemes is better suited to a separate study, which may for example study the effects that the different model components have in a complete GCM or CTM.

This is a technical paper presenting a new algorithm for a convective transport scheme. While it is certainly very interesting and important, it is out of the scope of this study to perform a detailed comparison of complex chemistry schemes or to discuss the chemistry of short-lived species like SO_2 in detail.

This model was originally developed as part of a larger study of the chemistry and transport of SO_2 from the troposphere to the stratosphere. An important part of this study is how the numerous uncertainties in SO_2

chemistry, convection, transport and microphysics translate into uncertainties in the SO_2 mixing ratios. It was decided to split the publication of this study into two papers. The combined study would have been too extensive and it is not a good idea to start a study about SO_2 with a long technical description of a convection model.

Unfortunately, a meaningful validation of the model is difficult with these SO_2 simulations and measurements. There are so many uncertainties that the results always can be tuned to agree with the measurements.

Specific comments

- **1. It would be good to include a brief introduction to the ATLAS model and how it works, so that the paper can be understood well without first reading other papers, as there is no easy or natural method to include complex chemistry into a Lagrangian model.**

  The ATLAS model is a model consisting of several independent modules. In this study, only the trajectory module is used. The chemistry module and the mixing module are not used.

  Radon and SO_2-like tracer mixing ratios are calculated with a simple exponential decay and fixed lifetimes. The more sophisticated chemistry model, which is implemented in the full ATLAS model and uses a system of coupled differential equations, is not employed.

  Changes to the manuscript: We changed the text in several locations (abstract, introduction, section 4, conclusions) to make clear that only the trajectory module is used. Added that the trajectory module uses a 4th order Runge-Kutta scheme.

- **2. Page 4 L 1ff: These sentences are not sufficiently precise, for example, it is not possible to speak about the mass of a trajectory.**

  It probably was not clear what the discussion was aiming at.

  We agree that there is no natural way to assign a mass to a single trajectory air parcel. One could argue that a trajectory air parcel only refers to an infinitesimal volume and that only intensive quantities like density are well defined for an air parcel, while extensive quantities like mass are not well defined.

  However, in a global model, where the model domain is filled with trajectory air parcels, this looks different. Here, the volume of the model domain can be divided into smaller subvolumes that make up the complete volume. Each subvolume can be associated with a trajectory air parcel, with the air parcel mass given by the product of density of air and air parcel volume. The same constant mass can be assigned to each trajectory air parcel, which implies that the associated volume is increasing with decreasing air density. Since the subvolumes should not overlap to avoid that the same air volume is counted twice, this means that trajectory air parcels are distributed uniformly over pressure (but exponentially decreasing over altitude).

This is not merely a theoretical consideration, but becomes important when e.g. the global mass of a chemical species is calculated, or the mass flux of a chemical species through a control surface (as the tropopause).

Changes to the manuscript: We considerably extended the discussion in section 2.1 as outlined above and moved the discussion to a new subsection 2.2.

- **3. Figures 1 and 6: The blue colour does print well.**

  You probably mean "does not print well"? A darker blue is used now.

- **4. Page 5, Eq 4: The equation of state should contain moisture**

  For a worst case scenario with a temperature of 300 K and a relative humidity of 100 %, the change in density compared to the dry density is 2.2 %. This is negligible given the uncertainties of the method.

  Changes to the manuscript: We added a note to the text.

- **5. Page 5, Eq 5 ff. One would better use just c as subscript.**

  Thanks for noting this. That was inconsistent throughout the manuscript, sometimes c was used, and sometimes "conv".

  Changes to the manuscript: We changed the subscript to "conv" consistently (see also below).

- **6. Page 6 Eq. 7 ff: Better not to use (long) words as subscripts.**

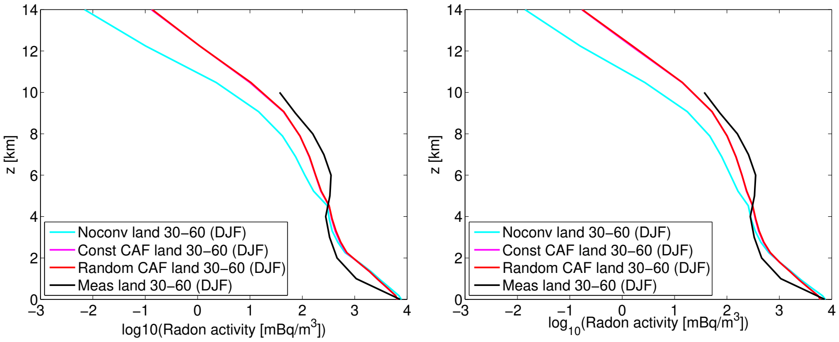
  In our opinion, short words as subscripts help to understand the equations. We agree that very long words (e.g. "subsidence") make the equations hard to read.

  Changes to the manuscript: We changed all subscripts of all variables consistently to consist of short words.

- **7. Page 10, L 22: It is not clear why an artificially degraded resolution of 2 degrees is used for the meteorological input from ERA-Interim.**

  The difference is due to computational constraints. The long-time run comprises more than 15 years. Simulation time is considerably reduced by changing the resolution from the original resolution of 0.75° x 0.75° to a resolution of 2° x 2° without changing the results significantly.

  The results of the long-time runs are not particularly sensitive to the resolution of the reanalysis data. 1-year runs with a time step of 10 minutes, 0.75° x 0.75° resolution of the analysis and a mean distance of the trajectories of 75 km have been performed to demonstrate that the results do not change significantly (a related comment of reviewer 1 asked for the difference that the change in time step from 10 min in the simplified run to 30 min in the radon run would cause). The runs with a time resolution of 30 min, a horizontal resolution of 2^o x 2^o and a mean distance of 150 km give nearly identical results (see figure, left: 2^o x 2^o, 30 min from Fig. 10 manuscript, right: 0.75^o x 0.75^o, 10 min).



The idealized runs from section 4.1 and the SO_2 run, which comprises a shorter time period, are based on ERA Interim data with a resolution of 0.75^o x 0.75^o now.

Changes to the manuscript: We added discussion of the 1-year runs to section 4.4.1 (section 4.2 in original manuscript). We increased the resolution to 0.75^o x 0.75^o in the simplified runs in section 4.1 and for the SO_2 runs in section 5.

- **8. Figure 4 and others: It would be good to frame figures (with tick marks on the upper and right axis)...**

  Changes to the manuscript: Done.

  ...and to use secondary ticks as appropriate (in Fig. 4, for each day).

  We are sorry that this is not feasible. Our software does only allow automatic placement of secondary tick marks, but there is no control over the spacing.

  The number of digits given should not vary along one axis.

  It is common practice that digits vary. For example, we do not think it makes sense to label the pressures "0800", "0900", "1000" or the mass flux "0.025", "0.030".

- **9. Page 14, L 10–11: I am wondering why trajectories were initialised at random positions rather than on an equal-area grid.**

  The random positioning is the default for trajectory initialization in the ATLAS model. It is normally used to avoid that an initialization on a regular grid can have any systematic effect on the results. It was used here for simplicity. An equal-area grid would probably work equally well for the application in this study.

Changes to the manuscript: We added that this is the default initialization scheme of ATLAS and that it is normally used to avoid any systematic effects to the paragraph in section 4.4 (4.2 in original manuscript).

Also, the 150 km horizontal resolution seems to be add odds with a random positioning.

This indeed needs a better explanation.

Changes to the manuscript: We changed the text to "Trajectories are initialized at random positions (both horizontally and in pressure) between 1100 hPa and 50 hPa. The number of trajectories is chosen in such a way that the mean horizontal distance of the trajectories is 150 km in reference to a layer of a width of 50 hPa."

- **10. Page 14, L 28 ff: "Radon is distributed evenly over these parcels by assuming a well-mixed boundary layer" Wording is not good.**

  Changes to the manuscript: Rephrased the sentence to "Radon is emitted into all trajectory air parcels that are in the boundary layer by assuming a well-mixed boundary layer, and a volume mixing ratio x of…"

  Eq. 13 is not an equation.

  Changes to the manuscript: Changed the text to "volume mixing ratio x" and the equation to $x = \dots$

  The emission rate would better not be denoted by e in a context where thermodynamic variables appear, it might be confused with vapour pressure.

  The disadvantage of using a letter different from e is that the association with the starting letter of "emission" is lost, so this is a compromise. ε is already used for the entrainment probability in the text, and E is used for the entrainment rate.

  It is also interesting to learn at this place that parcels transport volume mixing ratios, whereas in other places it was said that they represent masses.

  This is no contradiction. The basic assumption behind the concept of an "air parcel" is that it contains the same set of atoms at any given time. It follows that the mixing ratio of a given species is conserved along a trajectory (given that no chemical reactions take place) and that the mass of air is conserved.

- **11. Page 14–15, para. starting with line 33: The argument is not very clear. It would appear that an artificial minimum boundary-layer height of 500 m would systematically overestimate the input of Rn into the free atmosphere over land during winter, where probably the emission is already overestimated because of the snow cover effects.**

Our approach may cause some Radon which would be "trapped" in the boundary layer to end up in the free troposphere in the simulation and may cause some differences of the simulation to the Radon measurements.

However, assuming a minimum boundary layer height (or some similar measure) is unavoidable in global trajectory models, since the required number of trajectories needed for a model run which resolves the boundary layer by far exceeds any reasonable number that is computationally feasible.

The mass of radon emitted into the boundary layer per time period and area is still the same as with the actual boundary layer height and is not overestimated. This is accomplished by dividing by the boundary layer height z_{BL} in Equation 13.

Changes to the manuscript: We added discussion to the paragraph along the lines outlined above.

- **12. Page 15 L 17: I would not call this agreement "reasonable". Especially in Fig. 11 it is not good.**

  We agree that a better explanation is needed why the agreement is called "reasonable, given the large uncertainties in measurements and emissions". We think that there are good reasons to keep this formulation.

  We have now added discussion to section 4.4 (section 4.2 in the original manuscript) on how well the results of other studies compare to radon measurements to put the comparison of our model to radon measurements into perspective. Other studies show differences between their models and the radon measurements of a similar order of magnitude (e.g. Mahowald et al., 1995, Jacob et al., 1997, Collins et al., 2002, Forster et al., 2007, Feng et al., 2011). This suggests that a better agreement cannot be expected, given the uncertainties in measurement, emission and the simulation. The wording in other studies describing the agreement is comparable. E.g. Feng et al. states that their results "agree reasonably well" to the radon measurements. Their Figs. 13 and 14 show that the differences are comparable. Currently radon is probably still the species most suitable for the validation of convective transport models, since there is a lack of good alternatives.

  The underestimation of radon by the simulation in Fig. 11 has also been observed in other studies (e.g. Jacob et al., 1997, Forster et al., 2007). This may be due to uncertainties in emission and due to the fact that measurements from coastal areas are included, where horizontal radon gradients are high and difficult to model (see Forster et al., 2007).

  Changes to the manuscript: We extended the discussion as outlined above. Discussion was added to the introduction and conclusions, discussion in section 4.4 (4.2 in the original manuscript) was extended, and a discussion of the differences seen in Fig. 11 was added.

One is also wondering why no comparisons with single flights were done in the 1990ies there are ERA-Interim data.

The uncertainties of both the simulation and the radon measurements are so large that the data need to be averaged to obtain meaningful results. This is the common approach in most studies (e.g. Forster et al., 2007, Feng et al., 2011).

- **13. Page 16, Figure 8: It is not clear what "Points per layer" means.**

  Changes to the manuscript: Changed to "trajectory air parcels per layer".

- **14. Page 16 ff, Figures 9-12: It would be more instructive to show mixing ratios rather than concentrations.**

  The plots show the frequency of radioactive decay events (mBq) per volume (m^3), which is proportional to concentrations. This is the standard unit for radon, which is found in the majority of the publications (see e.g. Mahowald et al., 1995, Collins et al., 2002, Feng et al., 2011). For the reason of being comparable to other studies, we would like to stick to the units.

- **15. Page 18 L 9 ff: Do not repeat explanation of the colour of curves in the text.**

  We do not see a disadvantage. We would like to keep the text as is.

- **16. Page 18 ff, Section 4.3: The implications of choosing a specific cut-off value for the vertical velocity need to be discussed.**

  We substantially extended and rephrased this discussion. Part of the problem is caused by the conceptual problem of defining what a convective updraft is in the measurements. It is common to apply a lower threshold to the vertical updraft velocities to define convective situations in the measurements. Typically, this threshold is between $0\,\mathrm{m\,s^{-1}}$ and $1.5\,\mathrm{m\,s^{-1}}$ and may have a significant effect on the results (e.g. Kumar et al., 2015). Note that the 0.6 m/s cut-off is applied in Fig. 15 only for comparison. It does not appear in the model formulation.

  Replacing the simulated vertical updraft velocities by the measured vertical updraft velocities in the model would increase the average residence time between entrainment and detrainment. In turn, this would lead to a lower concentration of a short-lived species like SO_2 in the upper troposphere.

  Changes to the manuscript: Substantially expanded and rephrased the discussion in section 4.2 (4.3 in the original manuscript) as outlined above.

  Would it help to use cumulative frequency distributions rather than probability densities?

No. Since the cumulative frequency distribution is the integral of the probability density, changes at the small values of velocity will affect the values of the cumulative frequency distribution at large velocities.

- **17. Page 21, Figure 14: A step function or just symbols should be used, not continuous curves, as the data represent binned values.**

  In this case the binned data is used to approximate a curve which should be continuous in theory (by using an infinite number of bins). For this plot, which shows 30 bins, there is hardly any difference to a "continuous" curve.

- **18. Page 22, L 15–16: The Rn simulation is not suitable to demonstrate the proper long-term stability of mass distribution as radon has a short lifetime.**

  This is a misunderstanding. Radon is not used to demonstrate the long-term stability of the mass distribution. The long-term run is used for two separate purposes: a) To demonstrate the stability of the mass distribution, and b) to validate the model with radon. The radon mixing ratios are not needed to demonstrate the stability of the mass distribution, and the positions of the trajectories are sufficient for this. The stability of the mass distribution is demonstrated by counting the trajectory air parcels in a given altitude layer. Since every trajectory parcel is associated with a constant mass, this is equivalent to determining the mass in a layer.

  Changes to the manuscript: Changed the text in several locations to avoid misunderstandings: Added a new section 4.4.3 with the title "Conservation of vertical mass distribution". We changed the text in section 4.4.3 (originally section 4.2, page 14, lines 19–24) by including: "We revisit the issue of the conservation of the vertical mass distribution in this more realistic setup (compared to the idealized setup in Section 4.1)". We changed "mass distribution" in the sentence "The number of trajectories … at the start … compares well with the mass distribution at the end" to "number of trajectories". Added "conservation of vertical mass distribution of air (not of radon)" to the description in the text.

- **19. a) Authors should pay more attention to upper vs. lower case.**

  Changes to the manuscript: Changed.

  b) Page 2 L 2: It is surprising to see species in a CTM called "tracers"

  Changes to the manuscript: Changed "tracers" to "species".

- **20. Code and data accessibility**

  We would be happy to provide the source code to you by creating an account on our repository for you, if you feel this is necessary.

As far as we understand it from the ”model and data policy” statement, we are obliged to make the source code available to the editor, so that would have been the designated point of contact to our understanding.

It would also be nice if authors make available the old measurement data on-line in digital form (in which they must have them already), if it is legally possible, rather than pointing to printed publications.

We have no permission to do that.