Authors' response to referee comments on "Atmospheric odd nitrogen response to electron forcing from a 6D magnetospheric hybrid-kinetic simulation" by Häkkilä et al.

Please find below our answers (in blue) to the comments (in black).

Response to the comments of Referee #3

Review of "Atmospheric odd nitrogen response to electron forcing from a 6D magnetospheric hybrid-kinetic simulation" by Häkkilä, T. et al.
The manuscript presents a novel approach to combine magnetospheric modelling with whole-atmosphere modelling to investigate the auroral production of odd nitrogen. The magnetospheric model drives the precipitating electrons, and the whole-atmosphere model WACCM is used for the production of the odd nitrogen species, NO in this case.
The study is well thought out and an interesting approach that could shed more light on current discrepancies between measurements and current climate model simulations of the lower thermosphere. Thus it would be a worthy contribution for Ann. Geophys. However, I could recommend publication after minor revision of the manuscript. I feel some of the arguments need to be improved and made more clear, some of the issues might be caused by the language used and might be resolved with a slight rewrite or rephrasing.

Response to general feedback: We would like to thank the Reviewer for their positive comments and the incredibly detailed feedback. We appreciate the time devoted to the evaluation of our paper.

General comments
1. I understand that for numerical reasons the electron mass is different from the real value. However, this changes also the (kinetic) energy for the same speed, or the speed for the same kinetic energy. The energy and energy flux affect the precipitation characteristics, and the velocity is part of the Vlasov equation. How are these corrected back to realistic values corresponding to real electrons? Apparently the mass used is roughly 46 times the real electron mass, which would results in almost a factor of 7 for the velocities. Although [1] briefly discusses the impact of the different mass, the main arguments and justification could be stated here to justify the choice.
In addition, the changed electron mass is given relative to the proton mass, I suggest the authors also state it relative to the real electron mass, maybe denoted by me0 or similar, like it is done in [1]. I think it would make it clearer that it is the electron mass that is scaled and not the proton mass.
Also, currently there are two values listed, 5.11 MeV, citing [1], and what amounts to roughly 23.46 MeV a bit further down in the text. Which one was used? I suggest to remove the other to avoid confusion.

We have clarified the discussions on the electron mass in the manuscript. In response to the specific points, the choice of the mass ratio does not affect the evaluated differential number fluxes (Eq. (1). The mass ratio in the 2D run in [1] was indeed larger than in the 3D eVlasiator run shown here. The change for the 3D simulation was made to reduce computational costs. Differences between runs with different mass ratios were observed to be small, consisting of minor shifts in spatial structures. The change in the mass ratio is now explained explicitly in the manuscript.

2. I see that the authors use classical quantities for the kinetic energy (after Eq. (1)). [1] lists the range of velocities up to 42000 km/s per dimension, which leads to a maximum speed of about 73000 km/s, or roughly 25% light speed. How large is the error introduced here by using the classical approximation instead of the relativistic description? Why not use relativistic quantities throughout? At least some of the impacts and reasons should be discussed briefly (best with a reference).
A related question is why Vlasiator uses velocity and not the momentum to model the phase space more directly? One could then use for example the Liouville theorem to constrain the solutions further. I am not suggesting to rewrite Vlasiator for the current study, but the authors and model developers might want to think about that for future developments.

We would like to emphasise that the maximum velocity space extents are not actually modelled (due to our sparsity approach), and no particle velocities actually modelled will ever even reach the single-dimension maximum value. In [1] some particles do travel at over 10,000 km/s which indeed goes into the relativistic regime, but in the current study, due to more massive electrons being used, velocities remain lower (as shown in panels c,d,e of Fig. 1), and a non-relativistic solution is considered sufficient.
We thank the Referee for an interesting point of discussion. Using momentum directly in Vlasiator is an interesting suggestion and has been investigated in the past. It would not be completely impossible, but would require a re-write of the whole simulation model, as the SLICE-3D decomposition of acceleration into three shear motions would need to be also made relativistic, and this is not straightforward in the least. As regular Vlasiator simulations are constrained to the non-relativistic regime, this has thus not been a worthwhile avenue of code development.

3. Please list the absolute changes in addition to the relative ones, e.g. as 200% (from x to y). That would make it easier to judge if the change is from 1 to 2 (or 3?) molecules, or from 10^8 to 3×10^8. If not in the abstract, then at least in the results and conclusions sections.

We will add the changes in concentrations in absolute values to the Results and Conclusions wherever appropriate. For example, for the Abstract, the approximate NOx changes in the SH are 200 % (from 1.6×10^{14} to 5.1×10^{14} molecules per cm^2), 50 % (from 3.1×10^{14} to 4.9×10^{14} molecules per cm^2), and 7 % (from 1.6×10^{15} to 1.7×10^{15} molecules per cm^2) in the lower thermosphere, mesosphere, and upper stratosphere, respectively.

4. I understand that maybe due to limited resources, only single WACCM runs have been performed and not ensemble runs. However, I did not find that mentioned anywhere in the manuscript. Please add some words of discussion, and maybe add it to the "next steps", as only this would allow to assess the quality of the resulting NOx values.

Since we use specified dynamics in our simulations, below about 60 km an ensemble run would show no differences between the members. At altitudes above, based on our previous work, we found that depending on the season of the year NO variability between ensemble members can be up to 5 % near 80 km (NOx minimum) and smaller at other altitudes. Figure S1 shows the variability of a 10-member WACCM ensemble using specified dynamics. The data in Fig. S1 is from the simulations used in Verronen et al. (2020, https://doi.org/10.5194/angeo-38-833-2020), though this figure does not appear in the paper.
Since our WACCM simulations show more deviation in NOx from the reference than the previously determined 5 %, especially at thermospheric and mesospheric altitudes, we believe the results to be valid without the need for ensemble runs. We will add a brief discussion on this to Section 3.2 of the revised manuscript.

Specific comments
L. 13: See general comment #3, absolute changes would help.

We will add the absolute changes to the Results and Conclusions in the revised manuscript.

L. 22–23: "The polar MLTI depends on solar radiation" is a bit vague. Please revise to make it clearer to what aspect of the MLTI the authors refer to.

We will clarify the Introduction to address the vagueness during the revision, also taking into account the feedback from the other Reviewers.

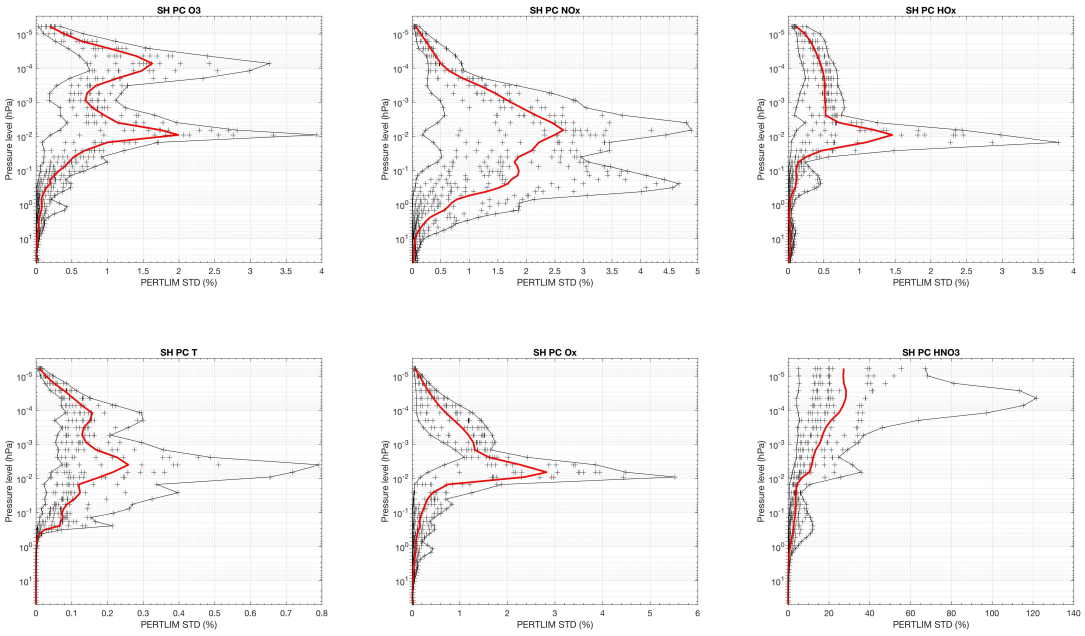

Figure S1. Standard deviation (SD) of the Southern Hemisphere (SH) polar-cap monthly O3, NOx, HOx, T, Ox, and HNO3 anomalies from a 10-member ensemble of SD-WACCM-D APEEP REF simulations, relative to the ensemble mean. Plus signs (+) indicate SDs for the individual months of January, February, March, October, November, and December (summer); circles are for individual months from April to September (winter); black lines are the minimum and maximum SD at each pressure level; and the red thick line is the median of all monthly SDs. See Verronen et al. (2020) for ensemble description.

L. 48: "... simplistic realization ..." only applies to the CMIP6 setup. There are other parametrizations available (e.g. [2]), which have been compared to each other and to measurements (e.g. [3]; [4]). However, so far not explicitly focussing on the auroral part.

Thank you for pointing this out; we will add discussion to the Introduction on existing approaches to auroral electron forcing in atmospheric simulations. As pointed out by the Reviewer, most of these methods do not specifically focus on the auroral part of the electron precipitation. This is the novelty of our study: we present a new methodology for the production of the precipitating electrons at auroral energies, and explicitly evaluate the atmospheric impact of the auroral forcing.

L. 75–76: In other words, the "6D" space is what is also known as the "velocity phase space".

There are many different ways of phrasing this, but yes, we model six-dimensional phase-space density in three spatial and three velocity dimensions. The manuscript has been amended on line 70 to reflect this.

L. 81: What are the "fields" here? Is this solution obtained iteratively? Please be more specific.

The manuscript has been amended to describe how we model both electric and magnetic fields. Our field solver approach is based on a finite volume approach described in Palmroth et al. (2018), and a reference was added here.

L. 84: What is “the ion inertial length”, can the authors give specific numbers here?

The ion inertial length $d_i = c/\omega_{pi}$ where c is the speed of light and ω_{pi} is the ion plasma frequency. This varies as a function of ion density, so a single value for the whole magnetospheric domain cannot be provided, but values can be of the order of, e.g., a hundred km.

L. 86: Which magnetic field model was used?

As the Vlasiator domain inner boundary is at nearly 5 RE, a simple dipole moment is sufficient. This has been clarified in the manuscript.

L. 95: What are “connected field lines”? Aren’t field lines “connected” by definition? Maybe just “along field lines” is enough.

Field lines which connect back to the ionosphere without reaching the Vlasiator computational domain (in equatorial regions) are not connected to the Vlasiator domain. However, we agree that this distinction is not necessary, and have removed that word.

L. 96: See general comment #1, please state the actual value used in the study and explain why it was chosen.

We have updated the manuscript to clarify this.

L. 100: Please add $\mathcal{F}_e$ to the text: “... number flux $\mathcal{F}_e$ at position r and energy E is given by ...”

We have made this correction.

L. 104: Which quantities are averaged? From reading Eq. (1), this would be θ and ϕ. If this is correct, please add it to the description.

We have updated the manuscript to clarify this.

L. 110: How typical are these conditions? Was the driving force applied constantly over the whole simulation period, or intermittently to simulate reality? Please explain the rationale for choosing those numbers.

These driving conditions are constant throughout the Vlasiator simulation, which has been added to the manuscript. These conditions depict fast solar wind stream conditions, ensuring fast initialisation and more simulation time for actual global dynamics modelling. Dynamic driving conditions are supported by Vlasiator, but were not included in this run, as each new driving condition takes time to propagate throughout the magnetospheric domain, requiring significant computational resources to model.

L. 115: Here, suddenly m_p/m_e is different from what was described before, please make these numbers consistent.

The manuscript has been amended to clarify this.

L. 129: “... fluxes ... are lower than ...” is difficult to assess without knowing the true flux. I suggest to use a more careful wording: “... fluxes ... might differ from reality.”

Thank you for this suggestion; we will make the change in the revised paper.

L. 131–132: In my opinion, the authors are stretching the notion of "calibration" here. Typically instruments are calibrated by comparing to a measurement standard to adjust the scale over a wide range of values. Here, only two days with the same conditions are compared. A few questions arise, first, are these typical conditions, in the sense that they occur very often? Second, how would this comparison hold up against a third day with similar conditions, and third, how would it hold up in a comparison with different driving conditions? Ideally, a "calibration" would have to be carried out based on the average fluxes (from SSJ or any other source) during different driving conditions. Unless this is fixed, I suggest to use a better terminology, maybe "adjustment" instead of "calibration". At least another comparison on a different day should be checked if the author's adjustment holds up.

Thank you for pointing this out. We agree that the term of "calibration" was poorly chosen. We will replace it with "scaling" in the revised manuscript.
The driving conditions corresponding to the Vlasiator (and eVlasiator) runs are not very typical, and correspond to a solar wind high-speed stream with a particularly large solar wind speed and low density. These driving parameters are chosen as a result of the computational constraints associated with running Vlasiator.
Our goal here is primarily to present a new methodology to evaluate the effects of auroral electron precipitation on the upper-atmospheric chemistry. We want to investigate to what extent using inputs for WACCM-D (i.e. precipitating electron fluxes) coming from first-principle simulations of the near-Earth plasma environment differs from using a Kp-driven parametrisation. We are aware that our method suffers from multiple limitations, such as the fact that, at the moment, we can only use one set of electron fluxes from a single time step simulated with eVlasiator. However, the Kp-based parametrisation also consists of a single set of forcing fluxes as a function of geomagnetic latitude, magnetic local time, and electron energy. In that sense, comparing the chemical response of the upper atmosphere to the forcing from eVlasiator to that from various Kp values in the parametric approach already provides valuable insight, although the eVlasiator fluxes (after scaling with DMSP observations during similar solar wind driving conditions) cannot realistically aim at fully reproducing real events. Therefore, while there are still many further developments needed to reach a point where eVlasiator fluxes (scaled with observations or not) can be readily used to study atmospheric chemistry under auroral electron forcing, this paper presents the first step towards this goal. We will clarify these objectives in the revised manuscript to underline that we cannot (and do not) claim that the current version of the method can be considered fully validated.

L. 145–149: At the first reading, this part was hard to understand, it only became clear after reading the appendix. It is not fully correct either, the "distributions" are nowhere shown, only the median and the percentiles are shown in the appendix. Please revise for readability, and include all the necessary information, the median/percentile of what exactly? "2nd and 3rd order polynomial" could be stated explicitly instead of vaguely referring to "a polynomial function". Does the "integrated energy flux" take into account the different electron masses between SSJ and eVlasiator? See also general comment #1. And what is finally done with the adjusted ratio, are the eVlasiator fluxes simply multiplied/divided by that factor? Please be more specific about how the final values are obtained.

Thank you for pointing out that this part is difficult to understand without reading the appendix. We will revise the text to better guide the reader.
We will also add the requested elements to ease the understanding and the improve the accuracy of the description of the scaling process.

L. 167: What altitude is "above its altitude range"? How can WACCM account for any effect above the altitude range? Although NOEM is derived for up to 200 km, the main NO distribution is centred around 105–110 km. How is this combined with the auroral forcing, both from the "normal" 2 keV "Kp-model" and the eVlasiator input to avoid double counting? Please be more specific about how the individual sources are combined.

NOEM is used to set the upper boundary condition of NO concentration in WACCM simulations. This accounts for the production of NO at altitudes not covered by the WACCM domain, and it's exclusion would lead to a lack of NO at high altitudes. NOEM is a standard component of WACCM simulations, and we are not aware of any discrepancies arising from its use as the NO upper boundary condition. In our revised analysis we have also limited the energy range of the eVlasiator-derived auroral electron fluxes to energies greater than 100 eV. This prevents overlap with NOEM, as electrons at energies below 100 eV would precipitate at altitudes above 140 km.
The use of NOEM in WACCM makes it necessary to set the Kp value for the VLAS simulation, as explained in Section 2.4.2 of the manuscript. We will clarify this further in the revised manuscript.

L. 175: See general comment #1, does this electron energy take into account the changed electron mass? How is it converted to the real electron mass used and energy used as input for WACCM?

The electron energies are calculated using the eVlasiator electron mass. Given the reduced mass ratio constraint of eVlasiator, we obtain the electron differential number flux in terms of energy, for the eVlasiator output. We assume the energisation of these high-mass electrons in eVlasiator is still representative. We take this differential number flux $\mathcal{F}_e(E, \mathbf{r})$ of eVlasiator electrons and apply it as the input of WACCM as-is, assuming the diff. number flux of real-mass electrons has the same distribution in energy as the eVlasiator output. We now state this explicitly in the manuscript.

L. 176: The parametrization in [5] was derived for an electron energy range from 100 eV to 1 MeV, How did the authors extend that range to lower energies of 50 eV? Is the energy grid in log-space, i.e. in log(energy)?

Thank you for pointing this out. Indeed, spectral energy range of 50 eV to 50 keV is used in the ionisation rate calculation. The energy grid has 32 log-spaced points most of which can be seen in Fig. 2. Three of these energy grid points are below 100 eV. We have now removed those three points and calculated the atmospheric ionisation rates again. At the model altitude range considered here, i.e. below 140 km, the removal of $<$100 eV electrons from the spectrum does not change ionisation. As pointed out by another Reviewer, electrons with such low energy cannot penetrate below the model's upper altitude limit. We have also re-examined the ionisation rates at the higher end of the auroral electron energy spectrum. Since the medium-energy electron (MEE) forcing recommended by CMIP6, which was used in our WACCM simulations, already accounts for electron forcing at energies greater than 30 keV, we have also removed the eVlasiator derived auroral electron fluxes at energies 30–50 keV. This avoids the possibility that the eVlasiator auroral forcing and the MEE forcing might overlap. The eVlasiator derived electron fluxes at these energies were so small, that this removal also had no considerable impact on the results of our study.
The final energy range of the eVlasiator-derived auroral electron input used in WACCM is 100 eV – 30 keV. As explained above, the adjustment of the auroral electron precipitation spectral energy range has negligible impact on the calculated ionisation rates and corresponding atmospheric response. We will add a brief discussion about this issue to Section 2.4.1 of the manuscript.

L. 177: Why was NRLMSIS-00 used for the ionization rates, and not the WACCM atmosphere? NRLMSIS-00 provides a good climatology, but here it might be a good idea to use WACCM itself, especially in terms of consistency. This choice needs to be motivated better, especially since CMIP6 is not part of the presented paper, and the method described in [6] and [7] is used for medium-energy electrons.

The ionisation rate calculation requires an atmosphere which for the calculation was taken from the NRLMSISE-00 model. However, according to the CMIP6 procedure (Matthes et al. 2017), the ionization rates are then divided by the MSIS mass density which effectively removes the atmospheric "signature". When the rates are used in WACCM, they are multiplied by the WACCM mass density profiles making the forcing consistent with the WACCM atmosphere.
We will note this in Section 2.4.1 of the manuscript.

L. 186: “WACCM-D runs from ... 2005 to ... 2006” is confusing and needs more explanation. The study compares the impact of events in 2011 and 2015, yet the WACCM runs are for an entirely different time period. What is the purpose then, and how do the authors reconcile consistency?

The primary goal of our study is to present a new methodology for the evaluation of the atmospheric impacts of precipitating auroral electrons. We are aware that the method currently suffers from many limitations, including the use of only a single set of auroral electron fluxes derived from eVlasiator. As such, we are not interested in the evaluation of any specific event, rather the type of forcing. Indeed this is the reason we use observations from two distinct periods to scale the eVlasiator electron fluxes, to mitigate the strong variability exhibited by real-life particle precipitation.
We therefore argue that the choice of the simulation period for the WACCM runs is of lesser importance. The fact that for each WACCM simulation the auroral electron forcing is kept constant already presents an unrealistic premise, but is in line with our goal of evaluating the type of forcing of the atmosphere. Since all the WACCM simulations include the same components, with the exception of the auroral electron forcing, they are comparable to each other, and differences between the runs arise from differences in the precipitating auroral electron data.
We will add a brief discussion on the choice of the WACCM simulation period to the manuscript in the revision.

L. 190: “turns off Kp ...”, what about the NOEM input which is also aurora related? Is it switched off as well for the “no-aurora” reference run?

NOEM is still included in the VLAS and REF simulations. Since the ionization rates derived from eVlasiator can only affect the NOx production within the altitude range of WACCM, NOEM is necessary to set the upper boundary NO of the WACCM simulations. This ensures comparability with the Kp-driven auroral forcing, while excluding NOEM results in a lack of NOx at high altitudes in WACCM. We will make this clearer in the revised manuscript.

L. 197: “... no Kp-driven ... aurora” depends on if NOEM has been switched off or not, otherwise there will be remnants of Kp-driven aurora via NOEM.

As NOEM is necessary in WACCM simulations to set the upper boundary of NO concentration, it is true that some auroral electron forcing effect is present in the REF simulation. We will clarify in the revised manuscript that NOEM is still included in the REF simulation, as this ensures comparability of the WACCM simulations.

L. 203: For completely “no aurora”, NOEM should be switched off too. It is not clear from the text.

NOEM is included in all our WACCM simulations. We will clarify this explicitly in the revised manuscript.

L. 206: Why is Kp 2 needed for eVlasiator? Shouldn’t it drive all the aurora itself? What Kp-level would correspond to the solar wind conditions used in the Vlasiator simulations? Table 1: The difference between REF and Kp0 could be made a little clearer.

It is necessary to set the Kp index value for the VLAS run, since it is used by NOEM to set the NO upper boundary condition. The auroral electron forcing derived from eVlasiator therefore only accounts for the ionization occurring withing WACCM’s altitude range. Processes above/at the top of WACCM’s altitude range are accounted for through the NO concentration set by NOEM at WACCM’s upper boundary.
We have chosen the Kp index for the VLAS simulation based on the Kp index during the DMSP satellite observations used to scale the eVlasiator electron fluxes.
The difference between the REF and KP0 simulations is the inclusion of the Kp-driven parameterization of the auroral electron forcing. The REF run uses the same modification as the VLAS run to disable the Kp parameterization, while the KP0

includes the parameterization. As can be seen in Fig. 4b, while the parameterized ionization rates may be low, setting the Kp index to 0 does still produce some auroral electron forcing.
We will clarify these aspects of the WACCM simulations in the revised manuscript.

L. 214: Does the "precipitating electron integrated energy flux" account for the heavier electron mass?

As stated in earlier replies, the precipitating electron integrated energy flux calculation uses the real electron mass.

L. 222: Please state the altitude range of the vertical integration.

The ionization rates in Fig. 4 are integrated from the surface to the top of the model (140 km). Since the auroral electron forcing only occurs at the highest altitudes, much of this altitude range does not contribute to the total integral. We will state this in the revised manuscript.

L. 223: The described features are very difficult to see with the chosen colourmap. See also below my comment and suggestion for this figure.

We will be updating Fig. 4, including the colourmap, during the revision.

L. 227: The limited magnetospheric domain only explains the hard cutoff at the lower latitudes. The VLAS ionization rates also do not reach as high latitudes as the KP2 run, between 10^{-4} and 10^{-5} hPa. What would be the reason for this?

The poleward edge of the precipitation corresponds to the open-closed field line boundary, which is captured by (e)Vlasiator at a specific point in time, and this is realistic (see, e.g. Newell et al., 1996, doi:10.1029/95JA03516). In contrast, statistical models contain contributions from varying conditions, blurring the otherwise-sharp poleward edge.

L. 233: I would assume that the "lower boundary" is because of the fixed 2 keV energy for the Kp-WACCM, and probably the limited altitude range used to derive it.

Yes, this is likely to be the case.

L. 237: Has the "polar cap average" been calculated with area weighting (cos (lat) weighting)? Please state it if so, otherwise not weighting correctly will over-emphasise the low values at high latitudes.

The polar averages have been calculated using cos-weighting to account for the differences in area represented by grid points at different latitudes. We now state this in the revised manuscript.

L. 240: See general comment #3, please add the absolute changes for reference.

We will add absolute values of the changes to the revised manuscript.

L. 249: See comment above, why is Kp needed for the Vlasiator run?

As stated in previous responses, it is necessary to set the Kp index value for the VLAS run, since it is used by NOEM. NOEM sets upper boundary condition of NO concentration in WACCM.

L. 254: As the authors point out, there was a NH SSW happening in 2006, which makes it even more confusing why the authors have chosen 2005 and 2006 for the WACCM simulations, and not 2011 or 2015 as for the comparison with SSJ.

As stated previously, our approach was specifically not to evaluate any specific events, and the use of constant auroral forcing throughout the WACCM simulation period would have made such assessments unrealistic. Since the polar vortex is generally known to be more unstable in the NH than the SH, this likely would have created differences between the hemispheric winters no matter the selected simulation period.

L. 258: The authors could show the polar-cap average time-altitude distribution in both hemispheres showing the downward transport for comparison with earlier studies.

Thank you for the suggestion. We will consider adding a figure depicting the NOx descent during the revision of the manuscript.

L. 261: I could not find any profiles presented as line plots. From the maps it is difficult to figure out what the authors are discussing here. Please add at least one figure actually showing profiles for a visual comparison.

Thank you for pointing out that the text and Fig. 7 do not adequately correspond to each other. We will revise the text and consider adding a figure specifically depicting NOx profiles to the revised manuscript.

L. 268: How would “solar irradiation” work in polar winter?

Thank you for pointing out this oversight. NOx production during the polar winter is driven by geomagnetic activity, including energetic particle precipitation. We will correct this in the revised manuscript.

L. 271: “... adds to the effect.” What effect? It is unclear what the authors are trying to point out here.

This refers to the off-pole nature of the distribution of NOx in the NH due to the weaker polar vortex. Since SSW leads to a less stable polar vortex, the SSW also leads to the less symmetrical distribution of NOx in the NH (Fig. 7d–f) compared to the SH (Fig. 7a–c). We will clarify this in the revision.

L. 277–278: “higher REF levels of NOx”, higher than what? Please clarify.

Thank you for pointing out this ambiguous language. This is in reference to the difference in the lower thermospheric NOx concentrations in the SH (Fig. 7a) and NH (Fig. 7d) in the REF simulation. The difference in REF NOx concentrations between the hemispheres helps explain the differences in the VLAS impacts in Fig 7g,j. We will clarify this in the revised manuscript.

L. 280: “difference in the REF background levels”, difference between what? Please clarify.

As above, this is in reference to the differences in the REF NOx concentrations between the two hemispheres, the SH (Fig. 7b) and NH (Fig. 7e). We will clarify this in the revised manuscript.

L. 293–298: I suggest to move these sentences to the conclusions since they provide more of a summary than a discussion.

We will make the suggested changes in the revision.

L. 303: Add “in WACCM”: “current parametrization of aurora in WACCM ...”. How do the authors arrive at the conclusion that the “standard” auroral forcing overestimates it? Overestimation compared to what? So far, no comparison to NOx measurements is presented.

We will add “in WACCM” in the revision. The assessment of the overestimation was made based on the parametrised ionisation rates on average being higher than the ionisation rates derived from the DMSP-scaled eVlasiator electron fluxes. It is true that further validation is needed for this statement, which is why the sentence is phrased as a hypothetical in the manuscript. We will reconsider this wording and its justification in the revision.

L. 310: What makes the authors so sure that they are “underestimated”?

As eVlasiator does not include all sources that are known to produce precipitating auroral electrons, the electron fluxes are expected to be lower than the total, but it is true that this might not be the case. As with line 129, we suggest to rephrase this and the following sentence as follows: “As pointed out in Sect. 2.3.1, eVlasiator does not model all sources of precipitating auroral electrons, and therefore the obtained electron fluxes might differ from reality. We have mitigated the possible discrepancy in this study by using the DMSP observations to scale the electron fluxes.”

L. 323–325: See general comment #3, please provide absolute changes as well.

We will add the absolute changes to the text.

L. 333: I suggest to replace “made” by “carried out”.

We will make the suggested changes in the revision.

L. 336: I suggest to replace “makes way” by “paves the way”.

We will make the suggested changes in the revision.

L. 350: I might be nitpicking here, but since a field line is an integral curve of the vector field (axial vector field in the case of **B**), it is already “traced”. Thus “tracing a field line” seems tautological, and “tracing the field” or “field lines are calculated” would already be unambiguous enough.

We will replace “field lines are traced” by “field lines are followed” in the revised version of the text.

L. 351: I suggest to use the term “initial point” instead of “seed point”, as I assume that is what the authors are describing.

We will replace “seed point” with “start point” in the revised manuscript, following a suggestion from another Reviewer.

L. 359: “follow the geomagnetic field” is unambiguous enough. But again, what magnetic field model was employed here?

We will modify the phrasing as suggested in the revision. As stated in the sentence, the magnetic field model employed here is the Tsyganenko 2001 model to account for the external magnetic field contributions, combined with the untilted dipole accounting for the internal contributions. More details can be found in the referenced paper (Grandin et al., 2023).

L. 365: “Better self-consistency” feels a bit strange, better in what sense? In my opinion, things either are consistent or they are not.

Thank you for pointing this out. In the revision, we will replace this phrase with “to be consistent”.

L. 365: I suggest to remove “lines” and “file” here, it is not clear what this “file” is and how it can be obtained.

We will make the suggested changes in the revision.

L. 368: It might just be a poor choice of words, but to my understanding a field line can only end on a surface, not in free space except when the (magnetic) field becomes exactly zero. Please revise for clarity.

We will modify the phrasing as follows in the revision: “and either the point where the magnetic field tracing was stopped (point C_2)”.

L. 369: I suggest to remove “therefore”.

We will remove “therefore” in the revision.

L. 385: The distinction could also refer to NH and SH instead of dayside and nightside, since the dayside passes are in the NH, and the nightside passes are in the SH. Only an additional comparison during NH nighttime and SH daytime would allow to use one or the other.

In this section, we are interested in scaling the precipitating fluxes on the dayside and on the nightside separately, as the dependence of auroral electron fluxes is predominantly related to magnetic local time. While interhemispheric asymmetries could be studied if suitable satellite overpasses were available to measure the particle fluxes in both hemispheres at the same MLT, they can be expected to be marginal in comparison to the dayside-vs-nightside differences. Therefore, we believe that referring to NH and SH in this paragraph would misportray the approach and further confuse the reader.

L. 388–391: A figure marking the selections for Oct 2015 is missing, in addition to Figs. B1–B3.

Since the purpose of Fig. B3 is solely to illustrate the MLAT–MLT domains retained as “regions of interest” for both events, we believe that adding the version with the data of the 10 October 2015 event in background would not bring more information and would only contribute to making the appendix harder to digest for the reader. We will however add a sentence clarifying that the regions of interest apply to both events in the revised text.

L. 392–393: This sentence essentially repeats the statement above it and can be removed.

Thank you for pointing this out; we will remove this sentence in the revision.

L. 395–408 (Appendix B3): This appendix describes a method that is not used in the end and should be removed to reduce confusion, including Fig B4. A (more extensive) comment about the issue with mean/median can be included in appendix B4. However, a better description is needed about over which dimension (of Fig. B3e,f) the quantiles are calculated. I can only assume that the authors mean “along the DMSP orbit”, or equally “time”. If that is correct, please state it in the text to avoid too much guessing there. Since the percentile is presented on a log-log plot, how do the mean and median of the log-log

behave? Are the distributions still skewed as in the linear case? Are the correction factors applied to the eVlasiator fluxes or to their logarithm?

Thank you for the suggestion to remove Appendix B3. However, we strongly believe that the scaling method, already difficult enough to follow for the reader, would become even more confusing if we were not to first introduce the reasoning by taking the example of using the median values. If we were to state from the beginning that we use the 61st and 67th percentiles to apply the scaling, those numbers would seem completely arbitrary. This is why we first explain how the energy-dependent correction is designed on differential number fluxes, to only later evaluate whether the correction was sufficient in the sense of integrated energy fluxes. We will clarify in the revision that the percentile values are indeed calculated along the DMSP orbits. The correction factors in their polynomial form are converted back to the linear domain before being applied to the eVlasiator fluxes; this will be explicitly stated in the revision.
We are unsure as to what the Reviewer means by "Since the percentile is presented on a log-log plot, how do the mean and median of the log-log behave? Are the distributions still skewed as in the linear case?"

L. 410–414: This description is very hard to understand, please revise. The 0.9 quantile of what exactly, and why should its ratio be 1?

Thank you for pointing out that this paragraph is not clear; we will rephrase and expand it in the revision. In particular, we will be more precise in explaining which parameter we consider the 90th percentile of and the reasoning behind the choices.

L. 411: "Insufficient" for what?

We will add the following precision in the revised text: "However, this increase is still insufficient to be representative of the energy input into the upper atmosphere associated with auroral electron precipitation as obtained in the DMSP observations".

L. 412: Referring to my earlier remarks, does this comparison of (integrated) energy fluxes account for the different electron masses used in eVlasiator and DMSP/SSJ?

Please see the earlier replies on this subject.

L. 410–419 and main text, either replace "quantile" with "percentile" or adapt the numbers to real quantiles between 0 and 1 (divide by 100).

Thank you for pointing out this language issue; we will replace "quantile" with "percentile" throughout the revised manuscript.

L. 418: What is the final ratio of the integrated energy fluxes?

The obtained ratios of the integrated energy fluxes are 0.99 and 1.03 for the cusp and nightside regions, respectively. We will add this information in the revision.

L. 422, Figs. B4, B5: The authors fit a third-order polynomial in log(energy), based on an energy range from 50 eV to about 5 keV (hard to identify in the figures). However, the "final" energy range is from 50 eV to 50 keV, and the 3rd order polynomial reaches a correction factor of about 10^5 already at 10 keV. Is there any "safety" mechanism that keeps the correction factor in a sensible range? Using an empirical model (here the polynomial) outside the parameter range used for fitting is dangerous, unless the authors can ensure that their model is valid outside that range.

There is no such risk, given that the high energies for which the ratio (blue line) is not defined are those for which the eVlasiator fluxes are zero. Therefore, whichever the value of the polynomial at those energies, the corrected fluxes will still be zero. Thank you for pointing this out, though, as it is something that we will clarify in the revision.

Figures

Fig. 4: I suggest to find a better colourmap, the current one makes it very difficult to distinguish VLAS from KP0–KP2. Maybe overlay contour lines from VLAS over all panels for comparison.

We will be updating Fig. 4 for the revised manuscript, including the colourmap and scale. To further highlight the differences between the eVlasiator-derived and the Kp-driven auroral electron forcing, we will also include momentary snapshots of the auroral ionisation rates, in addition to the currently shown daily averages.

Fig. 5: A suggestion would be to show a geomagnetically averaged zonal mean instead of a fixed longitude. For direct comparison, one could add the contours from one panel to the other.

Thank you for the suggestions. We will consider these during the revision of the manuscript.

Fig. 6: Please replace the bright yellow line by a different colour, it is very difficult to see on white paper. Please add the altitude ranges in to the panels or their titles so that they are easier to identify. The authors might consider removing the KP5 line so that the y-axis scale can be reduced, making the other lines easier to distinguish. Again, showing only relative differences is only part of the story, an additional comparison of absolute values would be helpful. And if the authors can find, real NO data would be very interesting to compare to as well.

We have updated the colours used in Figs. 6 and 8, and added the approximate altitude ranges to panel titles. The KP5 simulation has been replaced with KP3 in Figs. 6 and 8, and the scales have been adjusted accordingly. We will add the absolute NOx changes to the text in the revised manuscript.

Fig. 7: Please revise the colourmap of the panels (a)-(f), the features described in the text are difficult to see in the figure. In particular the black contour lines and labels are unreadable.

We will adjust the colourmap used in Fig. 7a–f to make it clearer in the revised manuscript.

Fig. 8: See comment about Fig. 6 and replace the bright yellow line by a different colour. Similar, reducing the Kp range to 0–2, the y-axis scale can be adapted to better distinguish the other lines. Also, please add the absolute changes for reference. Please label the panels clearly with “SH” for (a) and “NH” for (b) to make that easier to find.

We have updated the colours used in Figs. 6 and 8 and added the approximate altitude ranges to panel titles. The KP5 simulation has been replaced with KP3 in Figs. 6 and 8, and the scales have been adjusted accordingly. The hemispheres have been added to Fig. 8. We will add the absolute O3 changes to the text in the revised manuscript.

Fig. B5: Instead of only showing a line for the selected percentile, the authors could add contours for the underlying distributions, including additional lines for the mean and the median, with an emphasis on the percentile line that was used in the end. This would make it clear how skewed these distributions are, since the authors claim that using the mean or median are not good choices.

Thank you for this suggestion; we will consider it when preparing the revision of the manuscript.

References

[1] Alho, M. et al. (2022), Electron Signatures of Reconnection in a Global eVlasiator Simulation, Geophys. Res. Lett., 49(14). doi:10.1029/2022gl098329.

[2] Wissing, J.M. and Kallenrode, M.-B. (2009), Atmospheric Ionization Module Osnabrück (AIMOS): A 3-D model to determine atmospheric ionization by energetic charged particles from different populations, J. Geophys. Res. Space Phys., 114(A6), A06104. doi:10.1029/2008ja013884.

[3] Tyssøy, H.N. et al. (2021), HEPPA III Intercomparison Experiment on Electron Precipitation Impacts: 1. Estimated Ionization Rates During a Geomagnetic Active Period in April 2010, J. Geophys. Res. Space Phys., 126, e2021JA029128. doi:10.1029/2021ja029128.

[4] Sinnhuber, M. et al. (2021), Heppa III Intercomparison Experiment on Electron Precipitation Impacts: 2. Model-Measurement Intercomparison of Nitric Oxide (NO) During a Geomagnetic Storm in April 2010, J. Geophys. Res. Space Phys., 126, e2021JA029466. doi:10.1029/2021ja029466.

[5] Fang, X. et al. (2010), Parameterization of monoenergetic electron impact ionization, Geophys. Res. Lett., 37(22), L22106. doi:10.1029/2010gl045406.

[6] Kamp, M. van de et al. (2016), A model providing long-term data sets of energetic electron precipitation during geomagnetic storms, J. Geophys. Res. Atmos., 121(20), 12, 520–12, 540. doi:10.1002/2015jd024212.

[7] Matthes, K. et al. (2017), Solar forcing for CMIP6 (v3.2), Geosci. Model Dev., 10, 2247–2302. doi:10.5194/gmd-10-2247-2017.