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 #1

The authors present an approach to use ionization rates derived from electron fluxes calculated by the eVlasiator magnetospheric model in the WACCM atmospheric model for to calculate the response of atmospheric nitrogen oxides and ozone to auroral electron forcing. Traditionally, ionization rates for such atmospheric modelling are derived in proxy-based parameterization so I think that this new approach would allow for a more physical input. As a disclaimer, I don't feel competent to judge the magnetospheric modeling part of the study, but feel more at home with the atmospheric modelling part. However, I see two major issues that I'd like the authors to consider before a publication can be considered regarding a) the motivation for this development, and b) the evaluation of the approach.

Response to general comments: We would like to thank the Reviewer for their expertise and the constructive comments. We appreciate the time devoted to the evaluation of our paper.

a) In the Introduction the authors state that "accurate modelling of the MLTI is required to [...] further our knowledge of the region" and that the complex dynamics lead to "great uncertainties in our understanding of the region". Furthermore it is stated that the lack of knowledge about "auroral electron forcing in climate models could [...] obstruct accurate evaluation of polar climate variability". While this may all be true, it is also very vague, and therefore it is difficult to judge whether the proposed new approach could help to generate new knowledge. What specifically is not known and why do the authors think that their approach can help to fill the knowledge gap? The abstract emphasizes the variability of the electron forcing on the hourly time scale. This point is not addressed in the manuscript anymore, but if it is the ability of eVlasiator to cover such time scales I think it would be necessary to argue why this is important for the stated purposes, e.g., the representation of polar climate variability. The WACCM results show that the eVlasiator electron fluxes lead to quite similar responses of atmospheric NOx and O3 as provided by the traditional Kp-based parameterization. What is the added value of the new approach? If a proper hypothesis about the benefits of the new approach is provided in the introduction, it can be tested and discussed in the manuscript. Instead, in the discussion, the authors put forward the "enhanced information on energy and spatial distribution" provided by the new approach. But if the expected benefits of this are not discussed it is difficult to judge whether the new approach is useful or not. The Discussion mentions that the approach would enable future "near-real-time predictions of the atmospheric response". Isn't this already possible with the existing parameterizations and the available Kp predictions?

Thank you for the feedback on the vague motivation of our paper. Our approach provides a unique methodology for the study of auroral electron precipitation. As pointed out by another Reviewer, there are existing electron forcing data that do cover the auroral electrons, but none specifically focused on the auroral part. We employ a novel method for the production of the precipitating auroral electron fluxes through the use of Vlasiator. Electron precipitation is known to be highly variable energetically, spatially, and temporally. While we do not currently cover the temporal variability of the auroral forcing, this is a first step towards the use of more accurate electron driving in atmospheric simulations. Incorporating several eVlasiator runs with this now-proven methodology could allow for time-variable driving analysis.
We would also like to point out that added value is provided through the agreement of the eVlasiator-derived and the parameterised auroral electron forcing. Since these data are somewhat in agreement, with similar scales and impacts, despite completely independent origins, this works to validate both approaches. This could also act as the basis for a more accurate parameterisation, e.g. using solar wind parameters to drive the ionisation rates in stead of (or in addition to) the Kp index.

Since solar wind can be observed earlier than the geomagnetic activity used to calculate the Kp index, this could lead to the improved, near-real-time predictions alluded to in the Discussion.
We will further address the vagueness of the motivation of our manuscript, both in terms of the Introduction and the Discussion, during the revision.

b) Somewhat related to point a), I'm missing an evaluation of the electron fluxes (or ionization rates) calculated by eVlasiator. This point is related to a) because if the goal is unclear, it is also unclear how to assess whether the presented new approach is useful to reach this goal. Regarding the atmospheric response, I think that the comparison of NOx responses produced by the new and an old approach, as provided in Section 3, is sufficient for the current study. However, the manuscript leaves me uncertain as to whether eVlasiator in its current form provides any useful information. Again, I'm happy to admit that this may be due to my limited knowledge of magnetospheric processes and modeling, but other potential readers from the atmospheric community may share this knowledge gap. In principle I don't see a problem in scaling the eVlasiator output with satellite observations. However, from Figures B1 to B3 in the Appendix, one might get the impression, that the eVlasiator output doesn't provide any similarity to the observations. Often, the observations and the model output differ by factors of several orders of magnitude. Is there any correlation between model and data? If I understand correctly, the magnetospheric simulation is run for fixed boundary conditions that represent a situation similar to that on the two days of the satellite overpasses. Why not check the quality of the scaling by comparing the scaled data with independent observations from a third day with similar conditions. Or if that is not possible, consider using only one of the days for scaling and checkif the observations on the 2nd day are realistically reproduced. There may be many other ways to evaluate the approach, but without any evaluation I find it difficult to assess the usefulness of the approach.

Thank you for raising the point of the evaluation of the eVlasiator fluxes. In the revision, we will clarify that the goal is to demonstrate a new methodology to study the chemical response of the upper and middle atmosphere to particle forcing from space by introducing first-principle approaches to evaluate the precipitating electron fluxes used as an input for the atmospheric model (WACCM-D). The state-of-the-art is currently to use parametrised models, which may have limitations. There are two assets that eVlasiator brings to this aim. First, although it currently does not account for all the relevant processes leading to auroral electron precipitation, eVlasiator is a kinetic model of near-Earth space capable of producing a snapshot of electron parameters. Second, being a global model, eVlasiator provides such a snapshot in a global setup, meaning that the relevant parameters (the precipitating differential fluxes for our purposes here) are available as a 2D map above the polar regions.
The reason for scaling the eVlasiator fluxes with DMSP/SSJ observations is that, as mentioned above and in the manuscript, eVlasiator misses some of the important processes leading to electron pitch-angle scattering (e.g. chorus waves) and acceleration (e.g. field-aligned potential drops above the ionosphere). In our study, we therefore compare the eVlasiator fluxes with DMSP/SSJ fluxes during events for which the driving conditions (in terms of solar wind parameters) are similar to those used in our Vlasiator/eVlasiator runs. The comparison is carried out along the DMSP orbits, both on the dayside and on the nightside. Our aim is to make use of both the 2D extent of the eVlasiator fluxes and the "real-life" (though 1D) DMSP/SSJ observations to produce a forcing dataset (electron fluxes) which can be deemed realistic given the driving conditions.
The reason why we have used the two events presented in the paper to determine the scaling coefficients is that real-life precipitation exhibits strong variability, up to orders of magnitude despite highly similar driving conditions. This is due, among other reasons, to the transient and explosive nature of substorms, which introduce both spatial and temporal variability in the precipitating particle fluxes. Therefore, it is expected that the eVlasiator fluxes scaled based on one DMSP overpass will still exhibit significant differences with the DMSP/SSJ fluxes measured during the second event. Since our objective is not to reproduce a single event but rather to evaluate the atmospheric response to a certain type of forcing conditions over long time scales (a year in our study), we prefer to derive the scaling coefficients for eVlasiator fluxes by combining the measurements from the two events for which we have DMSP data.
That being said – and this is something that is not discussed enough in the initial version of our manuscript – the scaling alone does not enable overcoming all the limitations of eVlasiator. Two aspects, in particular, remain: (i) the fact that the high-energy component of the precipitating spectrum (beyond the cutoff energy originating from the sparse description of

phase-space density in eVlasiator) is still missing after the scaling. This is because the scaling is achieved by multiplying the original eVlasiator fluxes by the determined correction coefficient, and the resulting precipitating spectra are not extrapolated to higher energies. (ii) Currently, only one set of fluxes is available, as there is only one eVlasiator run, corresponding to a single time step in the Vlasiator run. This limitation could be overcome by producing more eVlasiator runs, but the computational cost is very large (millions of CPUh). While introducing time variability in the forcing dataset would certainly be one of the next steps to improve the method, we prefer to keep things as simple as possible in the present study – which already introduces many elements of novelty.

Hence, in the revision, we will pay close attention to (i) clarify the objectives of our study, (ii) highlight the novel aspects introduced by using eVlasiator fluxes as the forcing dataset, (iii) justify better the reasoning behind the choices made for their scaling with DMSP observations, and (iv) discuss more extensively and clearly the limitations of the method in its current version as well as future avenues for improvement.

In the following I will list further minor comments:

L26: "Hence, ... complex dynamics between the neutral atmosphere and the electromagnetic ionosphere". Maybe it is just the wording, but what is meant, here? "Complex interactions? And why "hence"?

The purpose of this sentence was to act as a summary of the preceding introduction, reiterating that the polar MLTI is affected by both the atmospheric and the magnetospheric processes. We will be revising the Introduction, taking into account also the above general comment a), as well as feedback from the other Reviewers.

L39: "in long-term atmospheric and climate simulations although uncertainties are still present in the latter". Do you want to say that only climate and not atmospheric simulations are uncertain?

Thank you for pointing out this inaccuracy. The uncertainty was in reference to there being more uncertainty in radiation belt electron data compared to solar energetic protons, though this is certainly not clear in the current text. The implication that there would be no uncertainty in solar protons data is also inaccurate. We will clarify and correct this in the revised version of the manuscript.

L41: "its impact on the stratospheric ozone balance is to a larger extent affected by polar atmosphere dynamics and is not fully understood". Almost nothing in our field of research is "fully understood". I think this is generally a poor motivation for a study. Please try to be specific on the knowledge gaps.

Thank you for pointing out the overly generic and vague nature of this statement. Specific deficiencies in our understanding of the role of the auroral forcing are already discussed on lines 54–58. We suggest to remove "and not fully understood" from this sentence, and to slightly expand the discussion on lines 54–58.

Fig. 1 caption: "simulation of the 3D-3V magnetospheric simulation"?

Thank you for pointing this error out, the caption has been reworded in the revised manuscript.

Fig1: It's a beautiful figure, but it is hard to extract information from it. What is the color scale of panel a)? Also the grey grids are hard to identify.

We have added the colour scale (proton pressure) for Fig. 1a. The grey grids were chosen for the lack of better alternatives for showcasing the regions of high-resolution solution; Palmroth et al. (2023) includes more details on the underlying Vlasiator run.

Fig. 4: I'd prefer using the same size for all panels. With panel a) larger than the others I find it hard to identify differences, e.g., in "sharpness".

We will update Fig. 4 in the revised manuscript so that the panels will be the same size in order to facilitate more accurate comparisons.

Section 3.1.2, comparison of ionization rates from the parameterization and from VLAS: While I understand that there are differences I'd like to understand which of them are fundamental to the approaches. If, e.g., the missing "sharpness" of the oval is considered a weakness of the parameterization, is this fundamental or couldn't the parameterization be modified accordingly?

Since the auroral electron fluxes from eVlasiator are currently based on a single time step, it is difficult to ascertain which features in the ionisation are characteristic of the physics-based simulation. Theoretically eVlasiator is able to produce the auroral electron fluxes (and thus ionisation rates) dependent on magnetospheric activity that the Kp-driven parametrisation cannot reproduce, since eVlasiator is physics-driven and more complex than the statistical model. This, however, requires (e)Vlasiator runs under multiple different driving conditions. In our current paper we therefore limit the analysis to comparisons with the current parameterisation used in WACCM simulations.
Considering the computational requirements of running Vlasiator, we do not consider it unreasonable that the model might be used to adjust the existing parametrisation in the future. Vlasiator could be useful in creating parametrisations e.g. driven by solar wind parameters, but in order to facilitate such developments, the necessary connections must first be understood.

L294: "realistic auroral electron precipitation fluxes from eVlasiator". This relates to my main point b). "Realistic" by which metric?

The ionisation rates are actually derived from precipitating electron fluxes which are driven by true *first-principles* physical simulations of magnetotail reconnection. We appreciate the need to justify the assertion that the fluxes provided by eVlasiator are "realistic" and we will reevaluate the wording of this statement in the revision.

L303 "Our results also indicate that the current parameterization of aurora may be overestimating the auroral forcing". Yes, may be, but later it is stated that "eVlasiator [...] underestimated the total precipitating fluxes. Is there a reason to believe that the first effect dominates?

These two statements refer to different aspects. The first one is a conclusion of our study, based on the utilisation of eVlasiator fluxes *scaled with DMSP observations* as inputs for electron forcing for WACCM-D. The second one is a discussion about why scaling eVlasiator was necessary, since the model does not include all relevant processes leading to auroral electron precipitation. We will make this difference appear more clearly in the revised manuscript.

L337: "for the detailed study of solar wind – atmosphere interaction". Besides the main issue a) which applies also here, because I find this very vague, I'd also suggest to reconsider the use of the word "interaction", which I usually understand to work in two ways. However the coupling of the magnetospheric and atmospheric models suggested here is clearly one-way, so it would only allow to study effects of auroral electron forcing on the atmosphere. I know that publications in the larger field often use the term "solar-terrestrial interactions" or similar, but I'd prefer an unambiguous use of the language.

Thank you for pointing out that while “interaction” may be a commonly used term, it is indeed ambiguous. It is true that the interaction presented in our study is strictly one-way, and as such we suggest to replace the sentence as follows: “For the future, this work paves the way for a more complete description of auroral electron forcing in atmospheric simulations. Eventually this would enable the detailed study of how the solar wind affects the atmosphere through particle precipitation.” We will further address the vagueness of the Conclusions during the revision.