We thank the reviewers for their constructive feedback. In the responses below, the reviewers' comments are presented in black, while our replies are provided in blue. Text added to the revised manuscript is indicated in *italics*, with a red colour for emphasis.

Reviewer #1

Review of "Lightning in global chemistry-climate models: A review of parameterizations, evaluations, and future projections" by Krause et al.

This paper reviews the lightning parametrisations used in chemistry-climate models, along with the trends in their changing use, evaluations of skill, and the climate projections made in the literature.

I like this review. I think it is fairly straightforward but plenty of work has gone into it and it is valuable. More importantly, it is timely. Much development and evaluation of lightning parametrisations has happened over the low-earth orbit observation era that we've just had. Now we are moving into a geostationary satellite observation era for lightning, and this review takes stock of where we got to with those datasets. There is also increasing use of machine learning models, and this model takes stock of where (largely) traditional parametrisation techniques got to, given the data available. Finally, storm-resolving models on global and regional scales are now becoming useful for climate change projections. Whilst, they're unlikely to be used in chemistry-climate models soon, they can contribute to our understanding of how lightning will change in future.

All my comments are minor, even the major one. Once the authors have considered them, I consider this paper is ready for publication.

Yours sincerely,

Declan Finney

We are happy the reviewer likes our paper and are grateful for the detailed review.

(sort of) Major comment

Possible ways forward (or lack of) – This work is a useful stock-take, I just wonder if the authors have ideas for how to push its value further by proposing recommendations to the community of how to make progress? The article touches on machine-learning approaches, and on the new geostationary satellites, but doesn't provide any specific directions for how to overcome the large uncertainties in, not only projecting lightning changes, but even parameterising it in the present-day. I provide a few examples of the kinds of things that come to my mind, but I do not consider necessary to include any of these. If the authors don't have their own thoughts, then the paper can be published without enhancing the recommendations.

- It is great that Cecil2026 has developed a new OTD/TRMM-LIS/ISS-LIS dataset, which kind of is the final thing from LEO satellites for now. Great progress has been made because of the ease of use (i.e. gridded, well characterised, netcdfs) of the LIS/OTD data. I would say we need the same coherent datasets from the lightning-observing GEO satellites now. Something that merges the datasets into something that's easy to pick up and evaluate or create parametrisations with. It's ok if there's missing data – we now have the opportunity to get to relevant storm timescales (daily or 3-hourly). Evaluating at these scales allows us to properly start questioning the physics connecting the input variables and the lightning diagnostic.

We thank the reviewer for these great suggestions and agree that the development of coherent, well-characterized gridded datasets from geostationary lightning imagers represents an important next step for lightning parameterization development and evaluation. We have therefore substantially revised Section 3.1 to provide a broader overview of ground-based, LEO, and GEO lightning observations and to discuss their respective advantages and limitations. In particular, we now highlight the substantially higher temporal sampling of GEO instruments and the resulting opportunity to evaluate parameterizations at daily and sub-daily timescales relevant to convective systems. We also identify the development of coherent, multi-satellite GEO datasets as an important direction for future work. The revised text states:

"In contrast to LEO instruments, GEO lightning imagers continuously monitor a fixed portion of the Earth's surface, providing substantially higher temporal sampling and enabling observations at timescales relevant to individual convective systems. Examples include the Geostationary Lightning Mapper (GLM) on the GOES-R series over the Americas (Rudlosky et al., 2019) and the Lightning Imager (LI) on Meteosat Third Generation (MTG) over Europe, Africa, and adjacent regions (Bliznák and Sokol, 2026). Like LEO optical sensors, these instruments detect total lightning without directly distinguishing between CG and IC flashes, although efforts have been made to separate the two (Ringhausen et al., 2021). Their continuous observations provide new opportunities for evaluating lightning parameterizations beyond climatological spatial patterns, including the diurnal cycle and the timing and evolution of lightning associated with individual convective systems. GEO lightning observations are already being used for regional evaluations; for example, GLM observations have revealed biases in the spatial distribution and diurnal cycle of lightning simulated by CTH (Dang et al., 2025). However, the DE of GEO instruments is not spatially uniform across the GEO field of view and can decrease substantially at large viewing angles and toward its edges (Murphy and Said, 2020; Wu et al., 2023; Bliznák and Sokol, 2026). In addition, lightning detection is affected by variations in solar illumination associated with the day–night cycle and by the optical thickness of clouds above the lightning flashes. Differences in viewing geometry, instrument characteristics, DE, and data processing therefore need to be considered when combining observations from different geostationary sensors.

Both LEO and GEO optical lightning sensors primarily detect emissions near the 777 nm atomic oxygen line. However, observations at this wavelength do not capture all forms of electrical activity within thunderstorms. In particular, corona discharges, which comprise large numbers of streamers, produce little or no detectable emission at 777 nm and may therefore be missed by conventional space-based lightning imagers (Soler et al., 2020). Instead, their optical emission is concentrated near 337 nm and is accompanied by bursts of Very High Frequency (VHF) radio emission. Space-borne observations from the Fast On-orbit Recording of Transient Events (FORTE, Jacobson and Heavner, 2005) LEO satellite and the Atmosphere–Space Interactions Monitor (ASIM, Soler et al., 2020) have shown that such discharges can be detected from orbit through these optical or radio signatures. The planned Constellation of Radio Frequency Sensors (CubeSpark, Gatlin et al., 2023) aims to combine optical measurements at 777 and 337 nm with VHF observations, potentially providing additional information on thunderstorm evolution beyond conventional lightning detection (Soler et al., 2024).

All lightning observation systems thus have measurement and detection characteristics that can influence both the development and evaluation of parameterizations. These include differences in DE, sensor type and viewing geometry, temporal and spatial sampling, and the algorithms used to identify and group lightning signals into events, strokes, or flashes. For optical satellite sensors, detection is affected by background illumination, cloud optical thickness, and thresholds used to distinguish lightning signals from noise or cosmic-ray events, whereas LLSs are influenced by network geometry, signal propagation, and location algorithms. Consequently, observations from different

systems may differ in their spatial and temporal coverage, the types of lightning signals they detect, and their sensitivity to individual events or flashes, and are therefore not necessarily directly comparable across products. While the availability of coherent, well-characterized gridded LEO datasets has greatly facilitated the evaluation and development of lightning parameterizations, comparable multi-satellite datasets from GEO observations are an important remaining opportunity. In particular, their ability to resolve lightning at daily or sub-daily timescales could enable more direct evaluation of the physical relationships between environmental predictors and lightning activity and thus provide a stronger basis for developing and testing physically based parameterizations."

- Something that I think can benefit projections is to break apart lightning intensity from storm occurrence. Finney2020 fig2 provides an example of this – I haven't seen it very often in other climate-lightning studies. This can be done fairly easily in models by outputting one more diagnostic of lightning-days, or lightning-hours, as well as total monthly lightning. Then you can talk about lightning-days, lightning intensity (fl/day), and total monthly lightning. Again this gets down closer to the timescale that storms actually exist upon, even if not the process timescales. I think we'll find lightning intensity more reliably increases with climate change across the world, and is tied to thermodynamic profiles and boundary layer conditions. Lightning-days is probably the complicated one, as it will depend more on synoptic patterns and global circulation changes. But pulling the metrics apart is a useful starting point for knowing how to build confidence in projections.

We thank the reviewer for this valuable suggestion and agree that separating lightning occurrence from lightning intensity could provide a useful framework for assessing projected changes in lightning. Following the approach illustrated by Finney *et al.* (2020), we have added this distinction to the discussion of future research directions. Specifically, we now highlight the potential value of evaluating the frequency of lightning-producing days or hours separately from the amount of lightning produced during those events. This would allow projected changes in total lightning to be decomposed into changes in storm occurrence and changes in lightning intensity, potentially providing a more physically interpretable basis for assessing projection uncertainty.

We agree that this decomposition could also help distinguish processes operating at different timescales. As the reviewer suggested, lightning intensity during individual events may be more directly related to local thermodynamic and boundary-layer conditions, whereas the frequency of lightning-producing events may additionally depend on synoptic-scale weather patterns and changes in atmospheric circulation. We have therefore added this approach as a promising direction for future model evaluation and projection studies, while noting that the relative contributions of these factors remain to be investigated systematically. We added this text to the Conclusions and outlook:

"In the meantime, extending existing lightning observations by new products from GEO satellites and ground-based LLSs will provide new opportunities to evaluate lightning parameterizations. The higher temporal resolution of these new products could also allow changes in total lightning to be separated into changes in the frequency of lightning-producing events and changes in lightning intensity, for example by considering the number of lightning-producing days or hours and the amount of lightning produced during these events (Finney et al., 2020). Such a decomposition could provide a more physically interpretable basis for assessing projected changes, as lightning intensity may be more directly related to thermodynamic and boundary-layer conditions, whereas the frequency of lightning-producing events may additionally depend on synoptic weather patterns and changes in atmospheric circulation. Together, these observational constraints could help identify the physical mechanisms underlying projected changes and thereby provide a basis for assessing confidence in different approaches to predicting lightning under future climate conditions."

(If the reviewer did not intend these suggestions to be incorporated into the review, we would of course be happy to remove the respective passages.)

Minor comments

L9 – I see no requirement for chemistry to be present in a climate model in order to project lightning activity. I suggest you change this this to "Climate models…" and add a sentence or two to explain that chemistry is one of the main motivations for inclusion of lightning params in climate models, and that much of the evaluation is done from that perspective. Therefore it's useful to look at chemistry-climate models.

We appreciate the reviewer comment and agree that while lightning is often studied because of its effects on atmospheric chemistry, it is not a requirement. We changed the wording to "Climate models…". We added the following sentence to section 2:
"While lightning itself does not require atmospheric chemistry to be simulated, lightning is often included in chemical transport models because of its impacts on atmospheric composition and chemistry."

L13 – "times" seems a bit colloquial. I think either just list the two variables, or say the "the product of.."

We changed the wording to "*the product of*".

L43 – Is "strong influence on…" "..climate" justified? The radiative effects resulting from lightning effects on ozone and methane are opposing, so I'm not sure how certain we are of the climate effect. Certainly atmospheric chemistry is strongly influenced.

The reviewer is right that the net climate impact may be quite small. We changed the sentence to:
"Lightning is an important natural source of nitrogen oxides (NO_X) in the middle and upper troposphere, an important precursor of ozone (O_3) and the hydroxyl radical (OH), thereby reducing the lifetime of methane (CH_4), and exerting strong influences on atmospheric chemistry, with implications for climate (Schumann and Huntrieser, 2007; Pickering et al., 2026)."

L57 – I find the ", while…." part of this sentence to be odd. As I understand it, this is not a caveat to the charge separation statement, but just one of the papers in the debate mentioned at the beginning of the sentence. Why just mention one paper in the debate? If you want to get into it, then I suggest you'll need to give a couple of sentences to it. Or you could just include a number of citations at the point of "debate".

Agreed. We decided to not go into the details and changed the sentence to:

"Although the precise mechanism responsible for lightning initiation remains an open question (e.g., Pasko et al., 2025), it is well established that lightning requires strong electric fields generated by charge separation within clouds (Lopez, 2016; Saunders, 1993)."

L59 – Whilst I am happy with the term "electrification", I wonder if you would keep better connection to the earlier part of the paragraph to say "explanation for charge separation"

We changed the wording accordingly.

L127 – Need to check nothing missed from "parametrization" and "parametrisation". I imagine WoS is smart enough to return them anyway, but you should check.

https://trends.google.com/trends/explore?date=all&q=parameterisation,parameterization,parametrization, parametrisation and

https://english.stackexchange.com/questions/79847/parametrise-or-parameterise-a-curve

Thanks for pointing this out. We added these search terms in our analysis and revised the text accordingly.

L132 - There is no need for them to be global. You can use regional models to project regional lightning changes. I think you might need to give a paragraph to tighten up you discussion of models. All kinds of climate model can be used to study this including GCMs and ESMs, and they may or may not include chemistry. Convection-permitting or cloud-resolving models on global or regional domains could be used (e.g. Romps2019 or Finney2020). So please have a think about how you want to describe what tools are used for projecting lightning.

While the reviewer is right that different models, including convection-permitting and regional climate models, can be used to simulate and project lightning, our review focuses on global lightning parameterizations and projections (see also line 296 in the original manuscript). We have now clarified this distinction and the chemistry issue in this paragraph:

"Global climate models, Earth System Models, and chemical transport models are our main tools for estimating future changes in global lightning activity. While lightning itself does not require atmospheric chemistry to be simulated, lightning is often included in chemical transport models because of its impacts on atmospheric composition and chemistry."

L237 – Finney2018 supplementary fig S1 also shows that they attempted to account for shifting isotherms under warming climate.

Thank you for pointing this out. We revised the text:

"Finney et al. (2018) subsequently adjusted the sampling level from 440 hPa to 390 hPa for future climate simulations, to account for the expected rise of the tropopause and associated shift in the vertical temperature profile under warming. Romps (2019) later modified this scheme further by using the upward cloud ice flux at the 260 K isotherm (IFLUX_T), which lies within the mixed-phase region, instead of a fixed pressure level, arguing that a temperature-based reference is more physically appropriate for climate change simulations, where the height of cloud charge separation can shift with warming."

L300 – I'd say this statement is already out of date but certainly wont stand the test of time well. I think you'd be best acknowledging that convection-permitting models are becoming more feasible for use in climate projection (there is no reason that they have to be applied globally anyway). E.G. I don't know if the ICON simulations had a lightning parameterisation but it would be lightweight so there's no reason they couldn't https://link.springer.com/article/10.1007/s00382-026-08235-8 I don't think you have to change your scope, just define it well. You could say that you are focusing on studies that have included a lightning param and a chemistry model. Then you just need to acknowledge that there could be higher resolution simulations in the near future, but that this is the state-of-the-art in CCMs.

We thank the reviewer for pointing this out. We changed the text to:

"However, while computational power has increased significantly in recent years and large-scale models are beginning to explicitly represent storm-scale convective circulations (Uribe et al., 2026),

running these models is still computationally expensive, especially for global, long-term climate simulations (Field et al., 2018). An alternative is so-called superparameterization, in which a two-dimensional cloud-resolving model is embedded within each grid column of a conventional climate model (Charn and Parishani, 2021). Nevertheless, summarizing all existing lightning parameterizations is beyond the scope of this review. Instead, we focus here on schemes that have been applied and evaluated in global-scale modelling studies using models that do not explicitly resolve convection."

Section 2 – this is lengthy, I think dividing it into 2-3 subsections would be pleasant. Perhaps one for CTH-related params, one for other non-microphysical params, and one for microphysics-based params. But whatever is fine, just break it up a bit.

We agree. We have substantially reorganized Section 2, also taking into account feedback from reviewer #2. Following these questions and suggestions, we divided parameterizations into four subsections based on the physical basis of the parameterizations rather than their chronological development:

- Cloud-depth-based parameterizations, covering CTH, its modifications, and CCD
- Convective intensity-based parameterizations, covering approaches based on precipitation, mass flux, and CAPE
- Microphysical parameterizations, covering approaches that explicitly represent ice-phase processes
- Multi-predictor statistical and machine-learning parameterizations, covering approaches that use multiple environmental variables and machine-learning approaches. Here, we now also describe two own parameterizations (ENV_XGB and ENV_Linear) which were introduced in a recently accepted paper (Krause *et al.*, accepted).

This restructuring reduces the impression of a chronological order and provides a clearer progression from empirical proxies of convective development toward parameterizations that represent increasingly specific physical controls on lightning.

L335 – "…though it doesn't distinguish between whether a flash is CG or IC" or something like that? Current sentence might imply the satellite measurement sees them as different

We made this clearer in the revised text:

"Because the 777-nm emission is produced by both CG and IC lightning, these instruments observe total lightning activity without directly discriminating between the two flash types."

and in the GEO paragraph

"Like LEO optical sensors, these instruments detect total lightning without directly distinguishing between CG and IC flashes, although efforts have been made to separate the two (Ringhausen et al., 2021)."

L350 – I feel like this sentence is leading to a "..and therefore…" there is not high confidence in their ability to simulate such trends?

We agree with the reviewer and added the following statement to the end of the sentence:

"… leaving the ability of models to reproduce such trends poorly constrained."

L369 – "Related to…" sounds odd here. The extratropical underestimate introduced in the previous sentence is not really to do with this intra-tropical differences. I think you mean "The tropical

underestimation occurs primarily over central Africa, with compensating overestimation from the Amazon and martime continent"?

We meant that the tropical overestimation is primarily driven by the overestimation over the Amazon and the Maritime Continent. In contrast, lightning over Central Africa is underestimated by all parameterizations and therefore does not contribute to the tropical overestimation. We made this clearer in the revised text:

"A common bias across many schemes, including CTH, is the overestimation of lightning in the tropics and underestimation in sub- and extratropical regions (Finney et al., 2014; He et al., 2022; Ghosh et al., 2025). Within the tropics, the overestimation is particularly pronounced over the Amazon and the Maritime Continent, whereas lightning over Central Africa is underestimated by all parameterizations."

L551 – I don't know if you want to comment on potential reason for this insensitivity but I would hypothesise that it is heavily determined by the tropopause shift, which is reasonably well understood radiative response. Given it is a fairly direct response (at least in the tropics) to GHGs/water vapour, it probably is fairly similar across models. I suspect it's something that is checked a lot for whether it sensible, and may well be part of the model tuning process. The cloud top height baseline may also be something that's tuned to observations across models in the present-day simulations – that well- established baseline might also help achieve similar projected changes in CTH/lightning.

We thank the reviewer for these thoughts. We agree that providing a potential explanation for the relatively similar sensitivity across climate models and scenarios is desirable given the strong nonlinear dependence on CTH. We have now added a brief discussion of a possible physical explanation:

"This apparent consistency may partly reflect the relatively robust response of tropopause height to warming (Santer et al., 2003), which may contribute to similarities in projected changes in CTH, particularly in the tropics."

L600 – "presumably" – If you don't feel confident to make a robust statement, then I would change this to "for a more complex, 2-moment scheme".

We changed the wording accordingly.

Technical comments

L107 – "sensitivities" typo?

Indeed, fixed.

L377 – "intermediate"?

Good catch! Fixed.

L534 – "sensitive" typo?

We changed the sentence to:

"Whether CG lightning responds more strongly to climate change than total lightning is uncertain as the few studies that reported changes for both lightning types (all computing CG lightning from CTH-derived total lightning multiplied by the CCD-derived CG fraction) came to different conclusions:"

Reviewer #2

This manuscript reviews the parameterization schemes used to represent lightning flash rates in global chemistry-climate models. It introduces 16 schemes and summarizes their evaluation against observations and their projected future changes. This work is mainly a summary of previous studies and includes little new analysis. It would be a useful paper for newcomers to learn the current status of lightning parameterization research and to find the relevant earlier work, but it lacks in-depth insight or outlook. Overall, it falls within the scope of ACP and serves as a supplement to the classic review by Schumann and Huntrieser (2007) from nearly 20 years ago. Before it can be considered for publication, the following comments need to be addressed.

We thank the reviewer for this overall assessment of our manuscript. The main contribution of the study is indeed its systematic synthesis of the lightning parameterization schemes currently used in global climate models, their evaluation against observations, and their projections under climate change. Given the substantial development of lightning observations, parameterizations, and climate projections since the Schumann and Huntrieser (2007) review, we believe that an updated synthesis is timely and provides a useful entry point as well as a critical overview for researchers working with lightning in global climate models. In response to the reviewer's critique, we provide now a more in-depth synthesis and a clearer outlook.

Major comments:

1. Lightning flash observations play an important role in this review, as they serve as the ground truth against which the model parameterizations are evaluated (Figure 1). However, the manuscript does not describe these datasets in sufficient detail. Please provide a proper introduction to at least the OTD/LIS climatology and the FLASHMAP dataset used in Figure 1: how each dataset is constructed and validated, and what their main limitations are. In addition, please explain why these particular datasets were chosen for Figure 1 rather than the other products introduced in Section 3.1.

We thank the reviewer for pointing this out. We agree that the observational datasets used to evaluate the parameterizations, particularly those shown in Figure 1, warrant a more detailed description. We have therefore substantially expanded Section 3.1 to describe the main classes of lightning observations (ground-based LLS, Schumann Resonance, LEO optical sensors, GEO optical sensors), including detailed descriptions of the LIS/OTD climatology and FLASHMAP. For both datasets, we now describe their construction, spatial and temporal coverage, detection characteristics, and main limitations.

For the LIS/OTD climatology, we now explain that the product combines observations from the Optical Transient Detector (OTD) and Lightning Imaging Sensor (LIS), applies detection-efficiency corrections, and aggregates the observations into gridded climatologies. We also discuss the intermittent sampling inherent to the LEO platforms and the fact that the instruments detect total lightning without directly distinguishing between CG and IC flashes:

"Space-based optical instruments detect short-duration emissions from the neutral oxygen triplet near 777 nm produced by lightning, observed as changes in pixel brightness from above (Blakeslee et al., 2020; Cecil et al., 2026b). These optical signals are then clustered into groups and flashes using spatiotemporal criteria. Because the 777-nm emission is produced by both CG and IC lightning, these instruments observe total lightning activity without directly discriminating between the two flash types. The DE of space-based optical instruments is generally high for both CG and IC lightning, but depends on factors such as time of the day, cloud thickness, discharge altitude, and field of view (Blakeslee et al., 2020; Cecil et al., 2026a; Murphy and Said, 2020). Space-based observations are obtained from Low Earth Orbit (LEO) or geostationary (GEO) satellites.

The most widely used LEO sensors for global lightning observations are the Optical Transient Detector (OTD) on board of the Orbview-1 satellite and the Lightning Imaging Sensor (LIS) on the Tropical Rainfall Measurement Mission (TRMM). OTD provided near-global coverage but was only operational from 04/1995 to 03/2000. In contrast, LIS only covered regions between ±38° but was active for a longer period (12/1997-04/2015). LIS also achieved a higher DE, recently estimated at 59-71%, compared to 22-52% for OTD (Cecil et al., 2026a). Data from both sensors are included in the joint LIS/OTD products, which are consequently more robust in the tropics. More recently, the "successor" of TRMM-LIS, ISS-LIS, operated on board the International Space Station (ISS) from 03/2017 to 11/2023 and extended observations to approximately 55°S–55°N (Cecil et al., 2026a). LIS/OTD are among the most extensively characterized global lightning observations, with their detection performance evaluated through coincident comparisons with ground-based lightning networks and other independent observations, as well as dedicated instrument and algorithm validation. The LIS/OTD products have been widely used to develop and evaluate lightning parameterizations because they provide near-global observations of total lightning, include explicit corrections for DE, and are available as standardized, freely accessible gridded products at spatial and temporal scales compatible with global models. However, LEO instruments sample a given region only occasionally. Obtaining robust gridded estimates of lightning density therefore generally require temporal and/or spatial smoothing to compensate for sparse sampling. Consequently, these datasets are particularly well suited to evaluating climatological spatial patterns, seasonal variability, and other relatively long-term statistics, but are less suitable for assessing the evolution of individual convective systems."

For FLASHMAP, we now describe its construction from spatially merged observations from Vaisala's GLD360 and the U.S. National Lightning Detection Network (NLDN), its evaluation, its aggregation into daily and monthly gridded products, and the lightning information retained in the dataset. We also discuss the spatially heterogeneous detection characteristics of the underlying ground-based networks and the resulting limitations when interpreting the product as a global observational reference:

"The Flash Location Aggregation from Strokes into a High-resolution Multi-scale Analysis Product (FLASHMAP) is a recently developed, freely available gridded lightning product based on spatially merged observations from Vaisala's Global Lightning Detection Network (GLD360) and the U.S. National Lightning Detection Network (NLDN) (Qu et al., 2026). It provides aggregated information on both CG and IC lightning, including flash counts and multiplicity statistics, as well as stroke counts and peak-current statistics. Around 52% of the recorded flashes are classified as CG, indicating a substantial underestimation of IC flashes. FLASHMAP was validated globally against WWLLN and regionally against lightning detection networks in Australia, Spain, and Alaska. While FLASHMAP detects a lot more strokes than WWLLN over land and most ocean regions, WWLLN has higher stroke densities in some oceanic regions The spatial patterns of FLASHMAP CG lightning are broadly consistent with those observed in LIS/OTD total lightning and ENTLN CG lightning datasets (Krause et al., accepted). The product is available at 0.25° daily and 0.5° monthly resolution for 2019-2025 (Qu et al., 2026)."

We also provide our reasoning why these two datasets were selected for Figure 1: OTD/LIS provides a well-established, near-global observational benchmark for total lightning that is independent of ground-based network coverage and is available as a standardized gridded climatology at spatial scales appropriate for evaluating global model output. FLASHMAP complements OTD/LIS by providing continuous, global CG lightning observations at high temporal resolution, allowing us to assess whether the main spatial patterns identified from the LEO observations are also evident in a contemporary ground-based dataset. These two datasets in our view are currently the most reliable

global products for total and CG lightning, respectively. We have added a short explanation for showing these datasets to the caption of Fig. 1:

"LIS/OTD and FLASHMAP are used here because they provide well-characterized, gridded global observations that are particularly suitable for comparison with global model output. LIS/OTD provides a near-global benchmark of total lightning independent of ground-based network coverage, while FLASHMAP provides a more recent, high-resolution product of CG lightning."

2. Section 2 reads as a chronological catalogue of equations rather than a synthesis, and it lacks the physical reasoning that should tie the schemes together. The section lists one parameterization after another largely in time order, but it does not explain the underlying physics of lightning occurrence or how the proxy in each scheme is connected to that physics. In particular:

- Why was CTH chosen to represent flash rate in the first place, and what is the physical justification for the proxy variable used in each of the other schemes?

We agree and have revised the section throughout to emphasize the physical motivation of the different parameterizations. We now introduce the parameterizations according to which aspect of convection or electrification they represent: cloud depth, convective intensity, microphysics, and machine-learning approaches. For each major parameterization, we briefly explain why the predictor is expected to be related to lightning.

For example, we clarify that CTH was initially used as an empirical proxy for convective development because taller convective clouds generally indicate deeper and more vigorous convection. We also now explicitly distinguish CTH from updraft velocity, noting that updraft velocity provides a more direct measure of convective intensity, whereas CTH was particularly attractive for global modelling because it was readily available from atmospheric models:

"The first global lightning parameterization used convective cloud top height (CTH) as a proxy for lightning flash rate (Price and Rind, 1992). In terms of the processes outlined above, CTH represents lightning only indirectly: deeper convective clouds tend to be associated with stronger convection and a greater vertical extent over which electrification can occur, but CTH does not explicitly describe updraft intensity or mixed-phase microphysics. Its use as a lightning predictor was motivated by both theoretical scaling arguments and observations showing a strong relationship between storm height and lightning activity. The theoretical basis originates from Vonnegut (1963) and was subsequently further developed by Williams (1985). In this framework, a thunderstorm is conceptualized as an electric dipole with two equal but opposite charge volumes. Assuming approximately constant charge density and geometrical similarity of thunderstorms, the characteristic electric field scales with storm size L, which is assumed to be proportional to CTH, while the potential difference scales as L^2. The charging current I scales with the product of charge transport velocity V and cloud cross-sectional area ($I \propto VL^2$), such that electrical power P scales as $P \propto VL^4$. Assuming further that charge transport velocity scales with convective velocity, that convective velocity increases approximately linearly with storm height ($V \propto L$), and that lightning flash rate is proportional to electrical power gives the fifth-power relationship $F \propto L^5$. Williams (1985) found this scaling to be consistent with observations from New England, Florida, and New Mexico, which yielded exponents of 5.0, 4.7, and 4.9, respectively.

Price and Rind translated this relationship into a parameterization suitable for global models, using CTHs from the International Satellite Cloud Climatology Project (ISCCP) to estimate global lightning distribution. They evaluated the resulting distribution against global lightning observations from the Defense Meteorological Satellite Program and regional sferics observations from Southern Africa. Because no comparable direct observations of flash rate as a function of CTH were available for maritime thunderstorms, Price and Rind (1992) derived the ocean relationship indirectly. They related

maximum updraft velocity to CTH separately for continental and maritime clouds and assumed that the relationship between updraft velocity and flash rate derived over land was also applicable over the ocean. This yielded a separate relationship for ocean grid boxes:

$$F_L = 3.44 \times 10^{-5}\, CTH^{4.9} \quad (1)$$

$$F_O = 6.40 \times 10^{-4}\, CTH^{1.73} \quad (2)$$

where F_L and F_O denote lightning flash rate per land or ocean grid-cell [flashes min^{-1}] and CTH is expressed in km above ground. Thus, rather than requiring CTH itself to explain the land–ocean contrast, the parameterization incorporates this contrast empirically through separate relationships. Price and Rind (1992) also proposed an alternative formulation based on peak updraft velocity, which is more directly related to convective intensity but is not universally available from model convection schemes. CTH therefore offered the practical advantage of being readily available from both satellite observations and global models."

> • What drives the distinct lightning characteristics over land versus ocean? Why do some schemes capture this land–ocean contrast while others do not, and why do some schemes fail to reproduce it even when separate coefficients are used for land and ocean (see Section 3.2, Fig. 3c)?

We agree that the land–ocean contrast warrants further discussion. As described in the Introduction, its physical origin is not fully understood, but differences in thermodynamic conditions, convective dynamics, and aerosol–cloud interactions are likely to contribute. Parameterizations based on a single equation cannot reproduce this contrast if the predictors do not adequately capture the processes responsible for the different lightning efficiencies of continental and maritime convection (e.g. see the revised description of the CAPE×P parameterization).

It is less obvious why parameterizations using separate land and ocean relationships may still fail to reproduce the observed global contrast. The original studies usually do not provide a specific explanation, but several factors could contribute: First, the calibration may not have been designed to reproduce the mean land and ocean flash rates specifically. If other characteristics, such as spatial variability or the overall lightning distribution, were prioritized, the resulting integrated land and ocean means may still be biased. Second, the predictor fields or lightning observations used for calibration may differ systematically from those used when the parameterization is subsequently applied or evaluated. This may be particularly relevant when relationships calibrated using reanalysis data are applied in a GCM, because biases in the simulated predictor propagate through the parameterization. Similarly, different lightning observational products exhibit substantially different land–ocean contrasts. For example, the contrast across ENTLN, FLASHMAP, and LIS/OTD ranges from approximately 2.2:1 to 4.8:1. A parameterization may therefore reproduce the contrast of its calibration dataset reasonably well but appear biased when evaluated against another observational product. This issue was particularly relevant for early parameterizations of maritime lightning, when observations over the oceans were sparse before the availability of global satellite measurements (see the revised description of the CTH parameterization). We have added a short discussion of these possible explanations to Section 3.2.

"Even when separate land and ocean relationships are used, the observed contrast may not be reproduced if the calibration was not specifically constrained by the respective mean flash rates. Differences between the predictor fields used for calibration and application may introduce additional biases, particularly when relationships derived from reanalysis data are applied in climate models. Moreover, observational lightning products themselves differ substantially in their land–ocean contrast, such that a parameterization calibrated against one dataset may appear biased when

evaluated against another. These issues were compounded for early parameterizations by the limited observations available over maritime regions."

3. Section 2: For each parameterization equation, please briefly explain how it was derived — based on what theory or dataset, and from what period. Simply stating that the authors "came up with an equation" (e.g., Line 165) is not sufficient.

We agree and have extensively revised the descriptions of the individual parameterizations to state their basis. For example, we now specify that the original CTH relationships of Price and Rind (1992) were derived from observations of continental thunderstorms in the United States and were motivated by earlier theoretical and empirical work on the relationship between cloud development and lightning. We have also replaced informal descriptions such as "came up with an equation" with wording that identifies whether a relationship was empirically derived, theoretically motivated, or developed as a modification of an existing parameterization.

4. Section 2: When introducing variants of existing schemes, please explain why each modification was considered necessary and how much it improved on the original, ideally with quantitative numbers. For example, He et al. (2022) modified the Lopez (2016) scheme based on findings from McCaul et al. (2009): what were those findings, and how does the modified scheme perform relative to the original?

We have added the physical motivation for the modifications and – if this information was provided – their improvements compared to the original parameterizations in quantitative numbers. In particular, we now explain the findings of McCaul et al. (2009) and how these were included in the He et al. (2022) parameterization, as well as the improvements in quantitative numbers:

"Lopez (2016) noted that Q_R bears some similarity to one of two lightning proxies developed by McCaul et al. (2009) from convection-resolving simulations. McCaul et al. related lightning either to the upward graupel flux at the -15 °C level, which better captured temporal variability, or to the vertically integrated amount of graupel, snow, and cloud ice in each model column, which better represented the spatial extent of lightning. While the latter sums the abundance of different ice hydrometeors, Lopez (2016) instead used the product of graupel and other hydrometeor concentrations as a proxy for their interactions and associated charge generation. He et al. (2022) subsequently modified the Lopez scheme to more closely follow the second approach of McCaul et al. (2009), replacing the collision-based Q_R proxy with the column-integrated amount of cloud ice, snow, and graupel within the 0 to −25 °C charging region (Q_{Ra}) and combining it with CAPE (Q-CAPE):

$$f_L = a_1 \, Q_{Ra} \, CAPE^{1.3} \qquad (17)$$

$$f_O = a_2 \, Q_{Ra} \, CAPE^{1.3} \qquad (18)$$

where f is the lightning density [flashes m^{-2} s^{-1}], a_1 and a_2 are constants obtained after calibration against LIS/OTD lightning observations, Q_{Ra} in kg m^{-2}, and CAPE in J kg^{-1}. Thus, compared with Q-CAPE-CBH, the modification replaces the product of interacting hydrometeors with their summed column abundance, removes the dependence on CBH, and increases the dependence on CAPE from $CAPE^{0.5}$ to $CAPE^{1.3}$. These modifications increased the global spatial correlation with LIS/OTD from r=0.79 for Q-CAPE-CBH to r=0.83 for Q-CAPE. More recently, Michibata (2024) extended this approach by coupling Q-CAPE to a two-moment microphysics scheme with prognostic graupel, allowing the mass and number concentration of graupel to evolve explicitly through processes including riming, accretion, melting, and sedimentation rather than diagnosing graupel from frozen precipitation."

5. Section 3.1: Geostationary satellite instruments for lightning detection should also be described.

We agree that geostationary satellite observations are an important complement to the LEO-based datasets discussed in the original manuscript. We have therefore expanded Section 3.1 to include a description of geostationary lightning instruments. We now describe their optical measurement principle, continuous temporal sampling, and the advantages and limitations of GEO observations compared with LEO instruments:

"In contrast to LEO instruments, GEO lightning imagers continuously monitor a fixed portion of the Earth's surface, providing substantially higher temporal sampling and enabling observations at timescales relevant to individual convective systems. Examples include the Geostationary Lightning Mapper (GLM) on the GOES-R series over the Americas (Rudlosky et al., 2019) and the Lightning Imager (LI) on Meteosat Third Generation (MTG) over Europe, Africa, and adjacent regions (Bliznák and Sokol, 2026). Like LEO optical sensors, these instruments detect total lightning without directly distinguishing between CG and IC flashes, although efforts have been made to separate the two (Ringhausen et al., 2021). Their continuous observations provide new opportunities for evaluating lightning parameterizations beyond climatological spatial patterns, including the diurnal cycle and the timing and evolution of lightning associated with individual convective systems. GEO lightning observations are already being used for regional evaluations; for example, GLM observations have revealed biases in the spatial distribution and diurnal cycle of lightning simulated by CTH (Dang et al., 2025). However, the DE of GEO instruments is not spatially uniform across the GEO field of view and can decrease substantially at large viewing angles and toward its edges (Murphy and Said, 2020; Wu et al., 2023; Bliznák and Sokol, 2026). In addition, lightning detection is affected by variations in solar illumination associated with the day–night cycle and by the optical thickness of clouds above the lightning flashes. Differences in viewing geometry, instrument characteristics, DE, and data processing therefore need to be considered when combining observations from different geostationary sensors."

We also discuss the potential of their high temporal resolution for evaluating lightning parameterizations at daily and sub-daily timescales and identify the development of coherent multi-satellite GEO datasets as an important direction for future work (also see response to reviewer #1).

6. Section 3.3, Lines 442–444: Why does the seasonal performance of the different schemes depend on their ability to simulate the land–ocean contrast? Please explain. Isn't it rather a matter of how well the proxy variable represents lightning activity across seasons, and of whether the model can reproduce the seasonal cycle of the proxy itself?

Both are actually important. While capturing the land–ocean contrast is critical — as its misrepresentation would lead to an incorrect spatial distribution of lightning and could therefore prevent the parameterization from reproducing the NH summer peak — it is not sufficient on its own. The reviewer is correct that even a model that accurately captures the climatological land–ocean contrast may fail to reproduce the seasonal cycle if the proxy variables do not themselves exhibit the correct seasonality. In practice, both aspects are essential: a parameterization must represent both the spatial structure of lightning, including the land–ocean contrast, and the seasonal dynamics of the underlying meteorological predictors.

We have clarified this dual dependence in the revised manuscript, emphasizing that both the land–ocean contrast and the seasonality of the proxy variables contribute to the ability to reproduce the seasonal cycle of lightning activity:

"Lightning activity generally follows the shift of the intertropical convergence zone, resulting in peak global lightning activity during Northern Hemisphere (NH) summer, as lightning is concentrated over land areas. Consequently, the parameterizations' ability to reproduce lightning seasonality depends on two key factors: (1) their ability to simulate the land-ocean contrast, which is facilitated by the use

of separate equations for land and ocean grid boxes — and (2) their ability to capture the seasonality of the predictor variable."

7. Section 4.1 and Abstract: for the values 4.4±7.6% and 7.5±7.9% K^{-1}, what does the standard deviation represent? I would suggest not reporting the average over all projections, since — as the authors themselves note — this value is heavily biased toward CTH. Instead, please report a separate average for each scheme, with the standard deviation reflecting the spread across different models or scenarios.

The standard deviation for 4.4 ± 7.6% K^{-1} was calculated across the mean projected lightning sensitivities of the individual parameterizations, with each parameterization contributing one mean value. In contrast, the standard deviation for 7.4 ± 7.8% K^{-1} was calculated across all individual projections. We agree that the latter estimate is strongly influenced by the large number of CTH-based projections. However, we decided to retain this value to highlight exactly the point that projections available in the literature are heavily biased towards CTH, which may not be representative of broader range of parameterization behaviour. We have clarified this distinction in the revised Abstract:

"The mean across all parameterizations is +5.5±8.7% K^{-1} (one value per parameterization), while the mean across all individual projections (which are strongly biased towards CTH) is +7.7±8.3% K^{-1}."

We agree that reporting the mean and standard deviation separately for each parameterization is important and have included these values in Section 4.1. However, we consider a scheme-by-scheme breakdown too detailed for the Abstract, where we instead report the overall spread across parameterizations and the literature projections.

Minor comments:

8. Abstract, Line 25: do these values represent global-mean changes?

Yes, we clarified this in the revised manuscript.

9. Introduction: consider removing the subsection titles. They are more of a distraction than a help, and read like a ChatGPT-suggested edit.

We removed the subsection titles from the Introduction.

10. Introduction, Lines 75–76: on what reference or dataset is this statement based?

Values around 25% are widely cited in the literature (e.g., Whaley *et al.*, 2024) and also confirmed when dividing the number of FLASHMAP CG lightning strikes by the number of LIS/OTD total lightning strikes. However, some papers also assume a ratio of around 20% (e.g., Mackerras *et al.*, 1998). We therefore changed the sentence to:

"Around 20-25% of all flashes reach the ground, while the remainder occur within or between clouds (Mackerras et al., 1998; Whaley et al., 2024)."

11. All equations should be numbered.

Agreed. All equations in Section 2 are now numbered.

12. Section 2, Lines 151–152: please provide the reference and briefly explain why this scheme is less widely used.

This statement was based on our own literature review. The reason for these formulas not being used in global models is likely related to the difficulty of representing maximum updraft velocity

consistently at GCM scales, where it is generally not a directly available model variable. In contrast, CTH is easily available from the models. We have added this explanation to the revised manuscript:

"Price and Rind (1992) also proposed an alternative formulation based on peak updraft velocity, which is more directly related to convective intensity but is not universally available from model convection schemes. CTH therefore offered the practical advantage of being readily available from both satellite observations and global models."

13. Section 2, Lines 152–153 ("To account for the large differences in hypothesized updraft velocities"): this seems to imply that updraft velocity is the more direct proxy for flash rate, and that CTH is used only as a proxy for updraft velocity. Since CTH cannot capture the land–ocean difference in updraft velocity, separate equations are then needed for land and ocean. Is this the intended reasoning? If so, please state it explicitly

We thank the reviewer for pointing out this potential ambiguity. The reviewer is correct that updraft velocity provides the more direct physical link to convective intensity and lightning in the theoretical framework underlying Price and Rind (1992), while CTH provides an indirect proxy for convective intensity and updraft velocity.

For continental thunderstorms, Price and Rind (1992) derived an empirical relationship between observed lightning flash rate and CTH directly. For the marine environment, however, direct flash-rate–CTH observations were not available to derive an analogous relationship. Price and Rind (1992) therefore used observed relationships between maximum updraft velocity and CTH for continental and marine clouds, which differed substantially, and combined these with the land-derived relationship between updraft velocity and flash rate. Thus, the separate land and ocean equations reflect the different relationships between CTH and convective updraft velocity, together with the lack of direct marine flash-rate–CTH observations.

We have revised Section 2.1 to make this reasoning explicit and to clarify that, although updraft velocity provides the more direct physical link to convective intensity, CTH was used as a practical and more readily available variable for global applications

14. Section 2, Lines 169–171: are the authors suggesting that the original CTH scheme is inconsistent with theoretical expectations or storm observations? Which expectations and which observations are meant here?

We thank the reviewer for pointing out that our original statement was too vague. We did not intend to suggest that the original CTH parameterization was generally inconsistent with observations. Rather, we were referring to two specific limitations identified in subsequent studies. First, Boccippio (2002) identified a formal inconsistency in the scaling relationships underlying the original derivation, which could lead to nonphysical behaviour when the relationships were inverted. Second, the very different land and ocean CTH exponents in Price and Rind (1992) arose from the empirical relationships used to derive the ocean parameterization, which was constrained by the scarcity of maritime lightning observations available at the time. The resulting ocean parameterization was subsequently found to substantially underestimate oceanic lightning (Luhar et al., 2021).

We have therefore removed the original general statement and replaced it with a more explicit discussion of these specific issues, including the theoretical inconsistency identified by Boccippio (2002) and the observational limitations of the original ocean relationship:

"However, the ocean parameterization (Eq. 2) suffered from a scarcity of maritime lightning observations at the time of its derivation and later turned out to substantially underestimate lightning over the oceans. Boccippio (2002) subsequently showed that the derivation contained a

formal inconsistency in the underlying scaling relationships and could produce nonphysical behaviour when the relationships were inverted. Boccippio (2002) therefore re-examined the scaling framework and formulated self-consistent relationships between flash rate, updraft velocity, and storm height based on the original Vonnegut framework. Combining the generator-power scaling with empirical relationships between updraft velocity and CTH resulted in substantially more similar CTH exponents over land and ocean than in Price and Rind (1992), 5.09 and 4.38, respectively, compared with 4.9 and 1.73. Boccippio (2002) also considered an alternative framework in which flash rate scales with generator current rather than electrical power, implying a weaker dependence on storm geometry and a stronger dependence on the rate of charge generation. More recently, Luhar et al. (2021) revisited these relationships and derived updated parameterizations following Boccippio's framework. When implemented in the ACCESS-UKCA chemistry–climate model, the revised oceanic parameterization substantially improved the representation of oceanic lightning compared with the original CTH parameterization. Thus, the strongly different exponents in the original CTH land and ocean relationships can be interpreted primarily as a consequence of the empirical assumptions used to derive the original ocean relationship rather than evidence for fundamentally different physical scaling laws over land and ocean."

15. Section 2, Line 161: what kind of spatial calibration factor? Please specify.

We have clarified the form and purpose of the spatial calibration introduced by Price and Rind (1994b). The calibration is an empirical, resolution-dependent correction that compensates for the smoothing and loss of localized maximum CTH when convective systems are spatially averaged over increasingly large model grid-cells. Without this correction, applying the strongly nonlinear CTH–flash-rate relationship to spatially averaged CTH would yield artificially low lightning activity compared with calculating lightning from the underlying fine-scale convective structures. We have revised the text accordingly:

"A further difficulty arises when relationships derived for individual thunderstorms are applied at the scale of climate model grid-cells. Price and Rind (1994b) noted that diagnosed CTH depends on horizontal resolution because deep convective cores occupy only a fraction of a grid-cell. When convective fields are spatially averaged over increasingly larger areas, the maximum CTH is progressively smoothed, reducing the representation of localized deep convection. Because flash rate depends nonlinearly on CTH, calculating lightning from this spatially averaged CTH can yield substantially lower flash rates than calculating lightning separately from the underlying fine-scale convective structures. Price and Rind (1994b) therefore derived an empirical resolution-dependent correction to the CTH-based flash rate, based on the decrease in maximum CTH when convective fields were averaged over progressively larger spatial areas. This correction compensates for the loss of resolved convective extremes at coarse resolution and reduces the resulting dependence of simulated lightning on model grid size."

16. Section 2, Lines 163–164: what method, and for what purpose? It is difficult to tell what Stolz et al. (2021) actually contributed here.

We have clarified both the method and purpose of the Stolz et al. (2021) approach. The original Price and Rind (1992) CTH parameterization estimates flash rate per thunderstorm, whereas its application in global models requires conversion to a flash density representative of an entire model grid-cell. Although the spatial correction of Price and Rind (1994) accounts for the effect of grid-cell averaging on CTH, it does not explicitly account for the number of thunderstorms occurring within a grid-cell. Stolz et al. (2021) addressed this scale mismatch by estimating the number of thunderstorm complexes from the modelled convective area and an assumed characteristic storm area and using

this number to convert the per-storm flash rate to a grid-cell flash density. We changed the text accordingly:

"More recently, Stolz et al. (2021) addressed a related but distinct scale mismatch in applying the CTH parameterization to coarse-resolution models. The original Price and Rind (1992) relationships predict flash rate per thunderstorm, whereas global models require a flash density representative of the entire grid-cell. While the spatial correction of Price and Rind (1994b) accounts for the reduction in CTH resulting from spatial averaging, it does not explicitly account for the number of thunderstorms within a grid-cell. Stolz et al. (2021) therefore estimated the number of thunderstorm complexes from the modelled convective area and an assumed characteristic storm area (1,000 km² over land and 10,000 km² over ocean). They then multiplied the CTH-derived flash rate per storm by the estimated number of storms and normalized by grid-cell area to obtain grid-cell flash density."

17. Section 2, Line 210: typo — "lightsning" should be "lightning".

Good catch! Fixed.

18. Section 2, Lines 211–219: are the authors suggesting that Yoshida et al. (2009) consider cloud depth a better proxy for flash frequency than CTH? What is the theoretical basis for this?

We have clarified the motivation and theoretical basis for the CCD approach. Yoshida et al. (2009) argued that CCD provides a more physically relevant measure of the vertical extent of the cloud associated with electrification than CTH. The rationale is that thunderstorm electrification and charge separation primarily occur through interactions among ice-phase hydrometeors in the mixed-phase region above the freezing level. Consequently, the vertical distance between the freezing level and cloud top is more directly related to the depth of the ice-containing part of the cloud than CTH measured relative to the surface. This distinction may be particularly important in a changing climate because the freezing level can rise together with cloud top, such that an increase in CTH does not necessarily imply a corresponding increase in the depth of the region relevant for electrification.

We have revised the text to explain this physical motivation and now describe how Yoshida et al. (2009) derived their empirical CCD–lightning relationships:

"Yoshida et al. (2009) further refined the geometrical measure by modifying how cloud depth is defined. They argued that the full cloud depth is not necessarily the most relevant quantity for electrification because charge separation is primarily associated with interactions among ice-phase hydrometeors in the mixed-phase region. They therefore introduced cold-cloud depth (CCD), defined as the vertical distance between the radar-derived storm top and the 0 °C melting level. Unlike CTH, CCD accounts for variations in the altitude of the freezing level and more directly represents the vertical extent of the ice-containing part of the cloud. This distinction is also relevant under climate warming, as an increase in CTH accompanied by a rising freezing level does not necessarily imply a corresponding increase in the depth of the cloud region relevant for electrification. Using eight years (1998–2005) of coincident TRMM Precipitation Radar and LIS observations, Yoshida et al. related CCD to the number of lightning flashes per second per convective cloud (NFSC). They found an approximately fifth-power relationship across all three regions considered, but with substantially different prefactors:

$$NFSC_L = 1.58 \times 10^{-6}\ CCD^{4.9} \quad (6)$$

$$NFSC_O = 1.26 \times 10^{-7}\ CCD^{5.1} \quad (7)$$

$$NFSC_{MaritimeContinent} = 5.01 \times 10^{-7}\ CCD^{5.1} \quad (8)$$

where NFSC is the number of flashes per second per convective cloud and CCD is in km. The coefficients correspond to the mean logarithmic prefactors of −5.8, −6.9, and −6.3 reported for land, ocean, and the Maritime Continent, respectively. Thus, while the exponent is remarkably consistent among regions, the approximately order-of-magnitude difference in the prefactor between land and ocean indicates substantially greater lightning activity in continental clouds for a given CCD. These relationships were subsequently implemented in the global CAM5 model by Clark et al. (2017). The observed present-day spatial lightning distribution was captured almost as well as with the CTH parameterization (spatial correlations of 0.80 and 0.83, respectively)."

19. Section 2, Line 218: what does F_MC denote?

F_{MC} refers to lightning over the Maritime Continent. We made this clear in the revised manuscript.

20. Section 2, Line 275: what is meant by a "two-dimensional" deep learning framework?

Two-dimensional referred here to spatial fields, i.e. longitude and latitude. We changed the sentence to:
"Yin et al. (2026) developed a deep-learning lightning scheme based on a convolutional architecture (ResUNet). CAPE is the model's sole predictor variable but is provided as a two-dimensional spatial field, such that the predicted lightning at a given grid-cell can depend on CAPE both locally and in other grid-cells. This contrasts with conventional grid-cell-based parameterizations, in which lightning at each grid-cell is calculated solely from predictor values within that grid-cell."

21. Section 3, Line 336: detection efficiency also depends on time of day (daytime vs. nighttime observation). Please add this.

Thank you for highlighting this point. We added this information to the text.

22. Section 3, Line 340: where does the estimated CG fraction come from? Please provide references and details on the assumed CG-fraction distribution and its justification.

We added this explanation:
"When CG parameterizations such as CP_Linear, CP_Poly or CAPE×P are compared with total lightning observations from space, the predicted CG lightning is usually scaled by an estimated CG fraction within each grid-cell — e.g., derived from cold-cloud depth (Price and Rind, 1993) or latitude (Prentice and Mackerras, 1977) — which can distort evaluation metrics."

23. Figure 3: what do the labels T1, EC, ZH, ZHW, B1, 1-m, and 2-m mean? Please clarify.

We clarified the meaning in the Figure caption:

"The abbreviations T1, EC, ZH, ZHW, and B1 in Tost et al. (2007) denote different convection schemes. In Charn and Parishani (2021), the labels 1-m and 2-m refer to one-moment and two-moment microphysics schemes, respectively. The number of moments indicates the degrees of freedom used to represent the size distribution of precipitating hydrometeors."

24. Section 3, Line 455: "NH South America" — this appears to be a typo. Please check.

No, the northernmost part of South America is actually located on the Northern Hemisphere. In this region, the CTH-CCD parameterization performs very poorly, see Fig. S4 in Clark *et al.* (2017).

25. Section 3: other studies have examined monthly time-series / seasonal / diurnal performance using geostationary instruments (e.g., Wu et al., 2023; Dang et al., 2025). Please consider including them.

We thank the reviewer for these suggestions and it is true that geostationary instruments have been used to evaluate lightning parameterizations. However, our review focuses on global-scale evaluations. Geostationary instruments only cover parts of the Earth's surface and, without merging different products, can therefore not be used for global evaluations. We therefor do not discuss these references in Section 3. Instead, we included the suggested references in Section 2.1 when discussing GEO observations.

26. Section 3, Lines 498–499: please provide quantitative numbers to support the claim of "impressive performance" by the machine-learning approaches.

We added this sentence:

*"*For instance, *Verjans and Franzke (2025) reported that ELM-tree explained 72% of the variance in daily LIS/OTD lightning observations, while Yin et al. (2026) reported 89% for ResUNet."*

27. Section 3, Lines 503–505 and Figure 5: it is difficult to follow how this analysis relates to the reliability of the schemes. Please explain the reasoning more explicitly.

We have now investigated this issue in considerably more detail. While this in-depth analysis confirmed our concerns and yielded results that are highly relevant for this section, we believe that the full analysis should be published in a separate study. We have therefore submitted it as a separate manuscript (Krause and Rammig, submitted), with a preprint expected to become available through the ESS Open Archive within the next few days. Accordingly, we have removed Fig. 5 and revised the text to summarize the main findings and their implications, while referring readers to the preprint for a detailed discussion:
"Recently, data-driven machine learning approaches achieved impressive statistical performance when evaluated against observations, with the test years explicitly excluded from the training data (Verjans and Franzke, 2025; Yin et al., 2026). For instance, Verjans and Franzke (2025) reported that ELM-tree explained 72% of the variance in daily LIS/OTD lightning observations, while Yin et al. (2026) reported 89% for ResUNet. However, such high evaluation scores do not necessarily demonstrate that these highly flexible models have learned transferable relationships between atmospheric conditions and lightning activity.

Krause and Rammig (submitted) showed that an even higher predictive skill can be achieved without using any meteorological predictors. Specifically, a simple climatological benchmark that predicts, for each grid-cell and day of the year, the corresponding mean of the training data explained 92% of the variance in daily LIS/OTD lightning. The benchmark also reproduced spatial patterns and hemispheric seasonal cycles with performance comparable to or better than the much more complex machine-learning models. This demonstrates that, for strongly smoothed lightning products such as LIS/OTD, test observations are already highly predictable from the spatiotemporal climatology contained in the training data. This issue may be particularly important for convolutional neural networks and other approaches that receive information from other grid-cells, because recurring large-scale spatial patterns in predictor fields can provide implicit information about location and season. Such models could learn to reconstruct recurring training data patterns without being appropriately sensitive to changes in meteorological drivers. Evaluating machine-learning lightning models for climate projections therefore requires tests that go beyond conventional spatial and seasonal skill metrics and explicitly assess whether their responses to changing environmental conditions are physically meaningful and transferable beyond the climate represented in the training data (Krause and Rammig, submitted)."

References:

Dang, R., Jacob, D. J., Wang, H., Nowlan, C. R., Gonzalez Abad, G., Chong, H., . . . Liao, H.: High-resolution geostationary satellite observations of free tropospheric NO2 over North America and implications for lightning emissions, Proceedings of the National Academy of Sciences, 122, e2510535122, doi:10.1073/pnas.2510535122, 2025.

Wu, Y., Pour-Biazar, A., Koshak, W. J., and Cheng, P.: LNOx Emission Model for Air Quality and Climate Studies Using Satellite Lightning Mapper Observations, Journal of Geophysical Research: Atmospheres, 128, e2022JD037406, https://doi.org/10.1029/2022JD037406, 2023.

Literature

Clark S K, Ward D S and Mahowald N M 2017 Parameterization-based uncertainty in future lightning flash density *Geophys Res Lett* **44** 2893-901

Finney D L, Marsham J H, Wilkinson J M, Field P R, Blyth A M, Jackson L S, Kendon E J, Tucker S O and Stratton R A 2020 African Lightning and its Relation to Rainfall and Climate Change in a Convection-Permitting Model *Geophys Res Lett* **47**

Krause A, Gregor K, Qu Y, Veraverbeke S and Rammig A accepted Large Projected Increase in Cloud-to-Ground Lightning Under Future Climate *Earth's Future*

Krause A and Rammig A submitted A Simple Climatology Benchmark Outperforms Machine-Learning Lightning Parameterizations

Mackerras D, Darveniza M, Orville R E, Williams E R and Goodman S J 1998 Global lightning: Total, cloud and ground flash estimates *J Geophys Res-Atmos* **103** 19791-809

Schumann U and Huntrieser H 2007 The global lightning-induced nitrogen oxides source *Atmos Chem Phys* **7** 3823-907

Whaley C, Etten-Bohm M, Schumacher C, Akingunola A, Arora V, Cole J, Lazare M, Plummer D, von Salzen K and Winter B 2024 A new lightning scheme in the Canadian Atmospheric Model (CanAM5.1): implementation, evaluation, and projections of lightning and fire in future climates *Geosci Model Dev* **17** 7141-55