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 #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):

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

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

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:

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

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

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

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