Review of Erdmann and Poelman paper on Lightning jumps and dives

Please note the authors' comments on the review highlighted in green.

This is my second review of this paper, and I find not much has changed in the paper. From the title it appears that the focus on the paper is on how lightning jumps and dives may be used to supply nowcasts of severe weather (hail, tornadoes, wind, etc.).
We regret that the significant changes made to the revised manuscript may not have been fully recognized or adequately valued.

We would like to clarify that the current title "Insights into Thunderstorm Characteristics from Geostationary Lightning Jump and Dive Observations" accurately reflects the focus of our study, which is to characterize storms that have GLM Lightning Jumps (LJs) and Lightning Dives (LDs). We would like to stress once more that the manuscript does not address the application of LJs and LDs in nowcasting.

However, the authors themselves show that ~60% of all lightning jump storms are NOT associated with severe weather (~70% are no for lightning dives). Hence, their own analysis shows that using the GEO lightning data we CANNOT use such parameters for nowcasting of severe weather.
As already mentioned, this study does not address the usability of LJs/LDs for nowcasting. That topic was thoroughly discussed in Erdmann and Poelman (2023).
In any case, the reviewers' comment would be valid if the goal were to use GLM LJ as a standalone severe weather warning tool. However, this is hardly ever the case. The information provided by GLM LJs is meant to complement existing tools, such as radar data.
This should have been the conclusion of the paper – that lightning jumps/dives based on satellite lightning data do not show skill in forecasting severe weather. This negative answer is also important to know.
The authors showed in the previous paper (Erdmann and Poelman, 2023), that the GLM LJs, as a predictor of severe weather, reach a CSI of 0.4 (POD of 0.58, FAR of 0.44) for LJ and severe weather within one storm cell, CSI of 0.48 (POD of 0.65, FAR of 0.37) for matching of LJs and severe weather reports that are close in space/time.
We can compare these results with studies using ground-based LJs for detecting severe weather:

- Schultz et al. (2016, doi: 10.15191/nwajom.2016.0407.): LMA data as GLM proxy data that maintain the flash rate of the LMA data -> found a POD of 69%, FAR of 63 %
- Miller et al. (2015, doi: 10.15191/nwajom.2015.0308): ENTLN ground-based LJs -> CSI of about 0.1 (POD > 0.8, FAR > 0.85)
- Murphy (2017, In Eighth Conf. on the Meteorological Application of Lightning Data): NLDN ground-based LJs -> CSI of about 0.1 (POD of 0.5-0.7, FAR > 0.8)
- Farnell et al. (2017, doi: 10.1016/j.atmosres.2016.08.021): counted the thunderstorms, LJ of predictor of severe weather -> found a CSI of 0.58
- Wapler et al. (2017, doi: 10.1016/j.atmosres.2017.04.009): hail-producing thunderstorms had moderate LJs in about 70% of the cases (FAR not calculated)
- Nisi et al. (2020, doi:10.1002/qj.3897): POD for LJ-based hail warning of 0.45, FAR of 0.3
- Tian et al. (2022, doi: 10.1016/j.atmosres.2022.106404): Chinese lightning network, 2-sigma LJ algorithm to predict hail events -> CSI of 0.41 (POD of 1.0, FAR of 0.59)

And with the skill of other approaches to nowcast severe weather and rare events:

- James et al. (2018, doi: 10.1175/WAF-D-18-0038.1): operational and future thunderstorm and severity warning tools at the German weather service, POD's of 0.4-0.9 (depending the lead time) and FAR of about 0.7
- Yao et al. (2022, doi: 10.1109/JSTARS.2022.3203398): deep learning based nowcasting of heavy precipitation, CSI about 0.4-0.45 (PODs of about 0.8, FAR of about 0.6)

But then the authors go on a fishing expedition to look for connections between lightning jumps/dives and many other cloud parameters, weather parameters, with no real goal of how to use these relationships. I remind the authors that the journal is about natural hazards and hence the extensive analysis NOT related to severe weather may be better suited for another journal. I would suggest reducing the number of parameters analysed, and to focus on a key research question. Can lightning from GEO be used to supply early warnings of severe weather. Yes or no.
NHESS is a journal dedicated to research on natural hazards, and lightning itself is a natural hazard per definition. The research question you mentioned, 'can lightning from GEO be used to supply early warnings of severe weather,' has been thoroughly addressed in Erdmann and Poelman (2023). As such, we believe it is unnecessary to revisit this topic in our current manuscript focusing on thunderstorm characteristics.

The authors use qualitative descriptions of their results, while for scientific evidence we need quantitative numbers. For example, "overall similar characteristics" (line 8), "can be useful for nowcasting" (line 190), "found to resemble" (line 320), "potentially the most dangerous" (line 324), "resemble" (line 342, 346)
The authors will review the manuscript and reduce those qualitative descriptions. Please note that all quantitative values can be found in the figures and the new Table 4.
Table 3 shows that most severe weather occurs WITHOUT LJ, while most LJ would give a false alarm if used for warnings of severe weather.
Please see the papers cited above. The studies demonstrate that LJs from ground-based lightning networks give similar or higher FAR when nowcasting severe weather events than observed for the GLM LJ in Erdmann and Poelman (2023). FARs of this work are common and similar to FARs found in other studies on nowcasting of severe weather. Severe weather in general is difficult to predict, and a combination of nowcasting tools is needed as stated in the current manuscript.
Hence, there is no benefit (and maybe even a danger) in using LJ/LD in nowcasting for severe weather warnings. Why not state this?
The nowcasting performance is not the objective of this manuscript and has been addressed thoroughly in Erdmann and Poelman (2023).

Figure 1 (you mean 2?): How can we compare parameters when the normalization hides the actual values?
The normalized values allow for a conclusion which one is higher or lower. In addition, we provide values with physical units in the text and new Table 4.
The authors added a new table to the revised manuscript that displays the mean value for each characteristic and thunderstorm category. We also removed the normalized mean/median values from Fig 2 as they are not discussed in the text and complicate the readability of the figure. Also, the panel showing LDs is removed and LDs are just briefly discussed in Sect. 3.2 as we noticed that the LJs are more useful and the LD detection algorithm needs further tuning.
Line 203 states the thunderstorms with LJs cover larger areas than storms with LDs. How can we see this from normalized areas? The min and max values will be different for each subset of data.
No, the min and max are identical for each subset as the overall min and max of the entire dataset are used for the normalization.
The manuscript states (lines 152-153): "The minimum and maximum values for each characteristic are taken from all analyzed thunderstorms and do not depend on the TS category" and "X" in Eq. 1 is given as the values from all categories, hence, the entire dataset.

Most of the analysis of the cloud parameters is not new, although the lightning data set is new. We know all the connections between severe weather and cloud top, ice, overshooting tops, updrafts, etc. Why another paper on this?

The paper demonstrates that the expected results are obtained for severe thunderstorms, confirming the effectiveness of the applied methods. Additionally, in a new and interesting finding, the study reveals that thunderstorms with GLM LJs (and, to some extent, LDs) exhibit characteristics typically associated with severe thunderstorms.

In conclusion, I think this paper is very weak in results, and needs either major revision with a new focus, or should be referred to another more appropriate journal.

We would like to point out again that no one has previously studied the satellite cloud characteristics of storms with GEO-detected LJs. Our work shows that GLM LJs occur in thunderstorms with the potential to produce severe weather, although not exclusively, and not all of these thunderstorms were reported as severe.

Dear Reviewers,

Please find below my comments/suggestions regarding the manuscript entitled "Characterizing lightning jump and dive producing thunderstorms from geostationary observations".
We would like to thank the reviewer for his detailed review and constructive criticism.
Please find the authors' responses to the comments highlighted in green.

In general, I think that the Authors try to answer to many questions but I can't see a previous work for adjusting the algorithm to the requierements of the new tool, considering a similar configuration than the used in terrestrial lightning detection networks.
We would like to point out that the lightning jump algorithm in use was adapted to the GEO lightning sensor and is described in detail in Erdmann and Poelman (2023). It does not use the same configuration as the LJ algorithms for terrestrial lightning location systems (LLSs).
The commonly used algorithms was developed for lightning mapping array (LMA) with radar-based storm tracking. Erdmann and Poelman (2023) tuned the algorithm parameters to be used with GLM lightning observations and satellite-based storm cell tracking.

My specific comments are (those starting with "**" are the more rellevant ones):

- L9: "consequently leading to more structured updrafts." -> "As a consequence of more structured and intense updrafts."
OK
- L11: "GEO LJs throughout their lifecycle exhibit the most and strongest OTs, signifying highly organized updrafts, extremely cold cloud tops, and highest CRRs." Add some values.
We revised the abstract and full manuscript to provide more quantitative values. We also added the new Table 4 as a quantitative comparison of thunderstorm categories. For example, the lines above are replaced by:
“While non-severe thunderstorms have a mean cloud top temperature of 236 K, cloud tops are about 20 K colder for severe storms as well as those producing LJs and LDs. Overshooting tops (OTs) in storms producing LJs, LDs and in severe storms were about 3.4 K, 1.9 K, and 2.6 K colder, respectively, than the cloud cell as a consequence of structured and intense updrafts. On the other hand, OTs are rare and shallow in the non-severe, and thunderstorms without LJs and LDs. Accordingly, the convective rain rates (CRRs) of the LJ (23 mm/h), LD (20 mm/h) producing storms and severe storms (20 mm/h) are on average more than 3 times higher than in non-severe thunderstorms and storms without LJs or LDs. Thunderstorms experiencing multiple GEO LJs during their lifecycle feature average cloud top temperatures of 213 K, with an average of 0.5 OTs being 4.8 K colder than the anvil, and the CRR mean exceeds 26.4 mm/h.”
- L17: "lightning observations can be used to locate these deep convective systems." Add some references
We added one reference (Avila et al. 2010) that cites other papers studying the relationship between lightning and convective clouds.
- L20: "lifecycle of the storm." -> Add references
We added 3 references that analyze variability of lightning activity in convective systems: Hayden et al. (2021), Borque et al. (2020), Goodman and MacGorman (1986)
Please note that all following references about LJs also show the variable flash rates in thunderstorms.
- L22: "Previous studies" -> There are many more updated refs. Please, include some of them
We removed the sentence. References are provided in the following and do not need to be repeated here.
- L25: "significant hail, or severe wind." -> Add the specific definition of each phenomenon
We added the specific values given by the NWS:

“The National Weather Service (NWS) defines severe weather as conditions involving tornadoes, significant hail (with a diameter of at least 2.54 cm or 1 inch), or winds of at least 93 km/h.”
- L25: Add Farnell et al 2017

OK
- L31: "decrease in lightning activity" Some interesting references:

Murphy, M. J., & Demetriades, N. W. (2005, January). An analysis of lightning holes in a DFW supercell storm using total lightning and radar information. In Extended Abstracts, Conf. on Meteorological Applications of Lightning Data (p. 52).

We decided that we won’t cite the reference as Murphy and Demetriades (2005) analyze lightning holes on a spatial scale 1-2km. Our manuscript considers the storm scale. In addition, GLM lightning data have a spatial resolution of 8-12km and, thus, features like the detailed lightning holes in Murphy and Demetriades cannot be seen at all.

Pineda, N., Rigo, T., Montanyà, J., & van der Velde, O. A. (2016). Charge structure analysis of a severe hailstorm with predominantly positive cloud-to-ground lightning. Atmospheric Research, 178, 31-44.

Pineda et al. (2016) analyze one specific storm in Catalonia, that showed a slight decrease in the total flash rate prior to hail reports twice. We added it as reference.
- L38: "appear to be" -> "are"

OK
- L39: "CG records" -> Add Rigo, T., & Farnell, C. (2022). Characterisation of thunderstorms with multiple lightning jumps. Atmosphere, 13(2), 171.

OK
- L53: "severe weather" -> Add Farnell et al (2017)

We believe the provided reference does not fit the context of the sentence as Farnell et al. (2017) did not optimize the LJ algorithm but just applied it.
- L66: "found" -> add Farnell et al 2018

We acknowledge the Referee's significant contributions to the studies suggested for citation (e.g., Rigo and Farnell 2022; Farnell et al. 2018, 2017; Pineda et al. 2016; Rigo and Llasat 2016; Rigo et al. 2010). While we have now cited several of these works, particularly the more recent ones, we believe that it is not necessary to include all of them for the purposes of our manuscript..
- General: add a list of acronyms

All acronyms are defined upon their first use, in accordance with the journal's guidelines.
- L151: "characteristics are normalized" -> "similar to Rigo and Llasat (2016) or Rigo et al (2010)"

Min/max normalization is a widely used approach and is not unique to the suggested reviewer's papers, which even employ a different normalization method. Therefore, we do not believe it would be beneficial to cite these papers.
- L164: "increase level (RIL) algorithm." -> add references

The algorithm type was developed in Erdmann and Poelman (2023), and the reference is given in the previous sentence.

We updated the LJ algorithm description so that the revised manuscript does not mention all the different algorithm types. We are simply providing the name and short description of the FRarea LJ algorithm that was developed and optimized for the data used in the manuscript. Hence, no further adaption was needed.

** L162: "Erdmann and Poelman (2023) optimized the LJ algorithm for GLM lightning records" -> Have hou considered the limitations of GLM? Have you compared the FR with the obtained by the NLDN or a LMA, for example?

Yes, Erdmann and Poelman (2023) considers GLM as a new tool, taking into account the strengths and weaknesses of optical lightning detection from space. Comparing the FR to ground-based systems would not provide added value for the LJ detection (The comparison of GLM to terrestrial LLSs is provided in much detail by other authors [e.g., Marchand et al. 2019; Zhang and Cummins 2020, Murphy and Said 2020, Rutledge et al. 2020].). Erdmann and Poelman (2023) tested a variety of different FR thresholds and picks the most suitable one for the GLM data, and the same FR

threshold is used in this submitted manuscript. In addition, also the most suitable sigma threshold is identified for the GLM lightning records in combination with satellite-based cell tracking.
We are aware that GLM performs less effectively over the northern CONUS compared to the southeastern CONUS, and that situations involving high flash rates, short durations, and small flashes present challenges for optical lightning detection from space. The appendix of the submitted manuscript addresses the issues of spatial variation in the GLM detection efficiency and shows that there is small impact on the LJ detection.
** L166: "The FRarea LJ algorithm first smoothens" Why? In the case of large thunderstorms, you limit the algorithm capabilities
The FRarea algorithm is defined like that as it is a modification of the sigma LJ algorithm that smoothens the time series to 2-minute steps (Gatlin and Goodman 2010, Schultz et al. 2009).
The normalization per cell area does not bring any drawback since the calculation uses the ratio. If a cell maintains its surface area, then the areas simply cancel out and the algorithm is identical to the sigma LJ algorithm. In addition, the FRarea algorithm better accounts for the merging/splitting of cells that change size rapidly from one time step to the next. Erdmann and Poelman (2023) show that this normalization increases the skill slightly compared to the original sigma LJ algorithm.
** L171: "This study uses the FRarea LJ algorithm with FR threshold of 15" This threshold is for the observed FR or the normalized FR?
The FR threshold is always based on the observed flash rate. Only the sigma calculation takes the normalization into account (since the history is important for sigma only). We added the unit in the paper, so it is clear that the FR threshold is not normalized:
"The LJ algorithm used in this study is the FRarea LJ algorithm, optimized for GLM lightning records as detailed by Erdmann and Poelman (2023). With a FR threshold of 15 flashes per minute and a sigma level of 1.0, the algorithm first checks …"
** L186: "an LJ also had an LD" -> Why?
We assume that most storms with a sufficiently high flash rate to trigger the LD algorithm feature a reduction of the flash rate (FR) during the decaying phase, and that this reduction often creates LDs. Since the FR threshold of the LD algorithm was set lower than for the LJ algorithm (after analyzing CSI values), LDs are detected for all the LJ storms. However, LDs have never been studied, and further research would be needed to understand a possible correlation between LJs and LDs.
Please note that these lines were removed and LDs are now discussed in the Sect. 3.2.
- L187: "However, it is possible that these storms did produce severe weather that was not reported." There are many references about the limitations of severe weather databases.
Hulton, F., & Schultz, D. M. (2024). Climatology of large hail in Europe: characteristics of the European Severe Weather Database. Natural Hazards and Earth System Sciences, 24(4), 1079-1098.
Schroeter, S., Richter, H., Arthur, C., Wilke, D., Dunford, M., Wehner, M., & Ebert, E. (2021). Forecasting the impacts of severe weather. Australian Journal of Emergency Management, The, 36(1), 76-83.
Thank you, we added the references to the manuscript.
"However, it is possible that these storms did produce severe weather that was not reported since severe weather databases have documented limitations (e.g., Hulton and Schultz 2024, Schroeter et al. 2021)."
** L195: "The key questions to be answered are (i) Do thunderstorms with GLM LJs and/or LDs feature particular characteristics?, (ii) How do the severe thunderstorms compare to the thunderstorms with GLM LJs and/or LDs?, (iii) Do GLM LDs provide added value?, and (iv) Is the number of GLM LJs or LDs important?" These lines should be moved to the end of the Introduction, as the goals of your research
OK, we moved them to the introduction. We also modified the research questions to better match the manuscript content and objective:
"(i) What do GLM LJs tell us about the storms structure from a satellite point of view?, (ii) Are GLM LJs useful to assess thunderstorm severity?, (iii) Do GLM LDs provide additional information about the thunderstorms?"

Sect. 4 is adapted accordingly.
- L201: "Figure 2 compares the normalized characteristics" Why this comparison?
We want to identify cloud characteristics of storm with LJs (and LDs), and we use the thunderstorms without LJs as a reference to know what distinguishes a thunderstorm with a GLM LJ from a thunderstorm that did not produce a LJ. This is among the main objectives of this paper.
** L203: "LJs (Figure 2b) cover a larger area than the storms with LDs" But all cells with LJ have LD... I can't understand this result
The storms with LJs are a subset of the storms with LDs. Hence, there are several storms that have an LD but no LJ, and these storms cause the difference.
- L204: "((Figure 2a))." Typo
Corrected
** L205: "large anvils for CTs near the tropopause" -> Explain better
Satellite-based cell detection and tracking sees the cloud tops. If a thunderstorm reaches the tropopause and an anvil forms, this cell appears larger than a thunderstorm that features mainly vertical development and has no anvil.
We modified the sentence in the manuscript:
"This could be related to the formation of large anvils for CTs near the tropopause that acts as a natural ceiling. If a thunderstorm grows up to the tropopause, vertical development is hampered and moist air is forced horizontally. The satellite-based cell detection sees the resulting anvil and those cells appear larger than thunderstorms that grow mainly vertically."
** L206: "Storms with LJs had an average area of 15,780 km2 (median 6,995 km2), whereas non-severe thunderstorms and those without LJs typically covered about 2,000 km2 on average (with medians around km2)." -> Considering that most of supercells are associated with severe weather (Farnell, C., Rigo, T., & Pineda, N. (2018). Exploring radar and lightning variables associated with the Lightning Jump. Can we predict the size of the hail?. Atmospheric Research, 202, 175-186.), do you have an approximation of the area of this convective mode in the region of study?
We use only satellite data and did not separate the thunderstorms by type/ convective mode. That means we do not know which thunderstorms are supercells (single cells, multi cells, or embedded in MCSs). An idea of the overall area of the storms is provided in Table 2 of Erdmann and Poelman (2023) that uses basically the same dataset as this manuscript:
Mean (median) area of thunderstorms: 2988.5km2 (762.8km2)
Mean (median) area of of all cloud cells: 619.9km2 (146.6km2)
-> the LJ storms are exceptionally large in their footprint area.
- L209: "LJs, LDs, and/or NCEI reports." Could it be related with the way as you consider the FR?
We don't think the FR has an impact on the cell area. It's possible that some cells exist with extensive cloud shields that have low flash rates and do not produce severe weather.
** Section 3.1.1: You could compare your results with:
Bedka, K. M. (2011). Overshooting cloud top detections using MSG SEVIRI Infrared brightness temperatures and their relationship to severe weather over Europe. Atmospheric Research, 99(2), 175-189.
Bedka, K., Brunner, J., Feltz, W., & Dworak, R. Overshooting Top and Enhanced-V Detections Objective Day/Night Overshooting Top and Enhanced-V Detections Using MODIS, AVHRR, and MSG Imagery in Preparation for GOES-R ABI.
The references given above mention a correlation between OTs and severe weather. We added another reference of this author in the following sentence as a reference for typical BTs of cloud tops:
"Coldest CT temperature is found for the multiLJ TSs (mean of 213 K). Typical BT thresholds for the detection of overshooting tops are in the range of 215 K and high anvils with 225 K (Autones et al, 2020, Bedka et al, 2016). Hence, CTs of the multiLJ storms are extremely cold."
** L239 "Overshooting tops (OTs) define a region of" How do you detect OTs?
This is part of the NWCSAF software. We added the following description to the manuscript, in a new paragraph below the section describing the NWCSAF software:

“OTs are detected in the NWCSAF RDT package through the application of several temperature and BTD criteria. The software identifies extremely cold cloud pixels (colder than -223 K in the mid-latitudes), and compares them to the surrounding pixels to identify the depth (as the temperature difference, DT) and horizontal extent of the OT. The BTDs take channels of WV6.2, WV7.3, and IR10.8 into account. Satellite pixels can also be identified as OT if they are at least 5 K colder than the tropopause. A detailed description of the OT detection algorithm can be found in Autones et al. (2020, p. 49-50).”

** L240: "above an anvil" -> Include at least one reference: Bedka, K. M., & Khlopenkov, K. (2016). A probabilistic multispectral pattern recognition method for detection of overshooting cloud tops using passive satellite imager observations. Journal of Applied Meteorology and Climatology, 55(9), 1983-2005.

We moved this description of OTs to the methods and add the reference there.

- L240: "Sometimes OTs even break through the tropopause. OTs are usual transient features, so this study analyzes the maximum OT activity of each thunderstorm. OT development needs a strong force manifestated as a strong, persistent updraft in thunderstorms. The air gets accelerated vertically and can overshoot the level of thermal equilibrium. Hence, OTs are indicative of dynamical thunderstorm cells with strong updrafts that are usually well organized. Given that strong updrafts frequently play a crucial role in the formation of tornadoes and large hail, storms with these characteristics are especially significant for nowcasting." These lines should be removed from results.

OK. We moved these lines to the methods and new paragraph about detection of OTs, and shortened this general part.

- Section 3.1.2: Include at least one reference: Bedka, K. M., & Khlopenkov, K. (2016). A probabilistic multispectral pattern recognition method for detection of overshooting cloud tops using passive satellite imager observations. Journal of Applied Meteorology and Climatology, 55(9), 1983-2005.

We moved this description of OTs to the methods section and add the reference there.

** Section 3.1.3.: "The max CRR in Figure 2 reveals that thunderstorms with LJs experience higher rain rates than storms with LDs, while CRRs 260 of the latter are still significantly higher compared to the noLJ storms. Furthermore," Explain how the CRR are estimated or reference it

This is also part of the NWCSAF nowcasting software. We added the reference to the documentation in the manuscript, under the method section describing the NWCSAF software:
“The NWCSAF software also includes a dedicated package to estimate convective rain rates (CRRs). This estimation uses analytic function calibrated based on radar data as ground truth, and also takes lightning observations into account. The complex algorithm to estimate CRRs is detailed in Lahuerta et al. (2020, p. 22-41).”

- L293: "but about half of the LD storms had no LJ. " Why?

The number of LDs storms was much higher than the number of LJ storms. This is a finding based on the applied detection algorithms that were optimized to match severe weather reports. Different thresholds in the algorithms would change this ratio.

Other questions:

- Which is the Lead time between LJ and severe weather?

Please see Erdmann and Poelman (2023, section 4c). Leadtimes are very variable, with mean (median) leadtimes of 37.5 min (34.0 min).
We added the information in the introduction:
“Erdmann and Poelman 2023 were among the first to optimize the LJ detection specifically for GLM lightning records in the central and eastern contiguous United States (CONUS) and found that GLM LJs as severe weather predictors reach a critical success index of about 0.5, with leadtimes averaging about half an hour.”

- which is the Lead Time between LD and severe weather?

We have not analyzed the leadtime for LDs yet as this work shows that the LD algorithm should be further modified and also that LJs seems to be more meaningful than the LDs.

- Have you considered the limitations of the occurence area? Problems in the limits of the region, or in higher latitudes?

Please see the paper appendix.

- Have you considered the size of the thunderstorms? In some cases it is possible that thunderstorm have small dimensions, specially in non warm season.

Why is the size of the storms important? We understand that it will be meaningful for the overall flash rate. If we look for LJs, however, the relative rate of change for one storm cell is interest. It should not matter too much whether the detected cell is initially small or large.

Best regards