Response to RC 2

We appreciate the detailed feedback and comments of the anonymous reviewer. We like to hear that the referee finds our work interesting and supports a publication after some modifications. We address all her/his concerns in the following colored blue.
Stated line numbers refer to the previously submitted manuscript to be consistent with the comments below.

Major comments:
[1] Unless the authors decide to show some plot or table summarizing the performance of their method for all four cases, section 7.2 and Table 2 do not seem relevant. In particular, the detailed description of the three extra cases in Table 2 is not needed. If the authors decide to focus on Case 1 only, the main conclusions (esp. at lines 475, 482 and 497) should be changed accordingly.
We think that it is a worthy addition to the paper to introduce and briefly discuss the other three cases. The conclusions of this work are not based on a single case study. To address the proof-of-concept idea, we want to highlight that different meteorological cases were successfully tested.
We created and included the FSS analysis that combines the four cases. This is to reduce noise in the evaluation of the skill scores.

[2] Since actual MTG-LI observations are not yet available, I would recommend to modify the title of the paper to ”Assimilation of Meteosat Third Generation (MTG) Lightning Imager (LI) pseudoobservations in AROME-France - Proof of Concept”.
We changed the title in our manuscript.

[3] The introduction, although interesting, feels rather lengthy. Even though previous studies do deserve to be cited, some references are cited with excessive detail. As it stands, the introduction could almost make a review paper in itself. I would therefore recommend that the authors try to shorten their review of past literature.
Please see the shortened introduction in the updated paper. We try to keep the citations but with less detail.

[4] In section 3, even though the goal here is to show that NLDN and Meteorage perform in a similar way, the authors should explicitly underline the fact that ground-based networks, on the one hand, and ISS-LIS, on the other hand, are likely to have quite different detection efficiencies for CG and IC flashes, respectively.
We added the information in the beginning of the section:
Line 160 : It should be mentioned that the low earth orbit, optical ISS-LIS observes lightning in a different way than the ground-based LF networks, with different detection efficiencies for CG and IC flashes, respectively (e.g., Erdmann et al., 2020, Zhang et al., 2019).

[5] To define their lightning observation operator, the authors compute a linear regression from the statistical distributions of observed FED and simulated graupel mass values. Their aim is to overcome the issue of frequent point-to-point mismatches between observations and model. It is also mentioned that MTG-LI FED observations are paired with the closest graupel mass values. The authors should clarify whether they specified a maximum distance in both space and time, when searching which model point should be paired with a given observation. If no maximum distance was applied, their pairing approach might be questionable in some situations.

As mentioned in the question, it is always the closest model point and m_g value that is paired with an FED observation (see line 187 and 199 in the manuscript). Hence, the maximum distance between a model point and the paired FED observation equals the model horizontal resolution of 1.3 km.
There is no time difference between paired FED observation and simulated m_g. For training the observation operator regression, only FED observations at the 1-hour forecast time are used, i.e., a 10-minute period center at the analysis time, and combined with m_g values from the analysis of this 1-hour forecast (see lines 184-186).
This first step is necessary to obtain the same amount of datapoints for FED and m_g. When we create the correlation in Fig. 3, both the FED and m_g datapoints are each sorted, and then paired. Hence, plotted pairs of FED and m_g are geographically independent from each other (see line 190). This method is further detailed in Combarnous et. al (2022, doi: 10.5194/nhess-2022-39).
We added to line 188:
"This gives an equal count of FED and column integrated m_g values that are used to train and validate our observation operator."
We changed the manuscript to make this clearer (lines 198-201):
Old: "Figure 3 presents the analyzed linear relationship between FED and column integrated graupel mass m_g that are calculated per 7km×7km pixel whereby mg uses a single model grid point closest to the FED pixel center. It should be noted that all pixels with either FED equal to zero or m_g equal to zero are removed from the data. The observation operator represents the cases when lightning was actually observed."
New: "Figure 3 presents the analyzed linear relationship between FED and column integrated graupel mass m_g. The functions are obtained after sorting both the FED and mg data individually. Paired datapoints in Figure 3 are geographically independent values as described in Combarnous et al. (2022). It should be noted that all pixels with either FED equal to zero or mg equal to zero are removed from the data. The observation operator represents the cases when lightning was actually observed."

[6] The flash detection efficiency of ground-based lightning detection networks, such as Meteorage, is well below 100%, especially for IC flashes. This implies that a no-lightning observation does not necessarily guarantee that (weak) lightning activity was not present in the real world. I believe the assimilation of such ambiguous observation to correct a lightning-active model background might lead to a degradion in the analysis. Could the authors clarify whether/how this possible ambiguity of no-lightning observations is/could be handled in their method?
We use the observations as best estimate of the real atmosphere. If the observation is not correct, i.e., real lightning was not detected, this false observation is still assimilated. In the example, correctly simulated lightning and convection might be depressed as a consequence. However, the flash DE of Meteorage has improved over the years and reaches about 80%, thus, the vast majority of flashes is detected. Missing one flash might be negligible in most cases since we count the flashes in 10 min time frames. In addition, the missing flashes are also part of the observation errors and as such taken into account by our method.
One approach to better handle these situations is the *noCloud*-filter (see line 338-345). If no lightning is observed, and if the background FED is positive, then we modify the background RH only if there is also no lightning within the 3 adjacent FED grid pixels in each direction (within 21 km).
We added the following content to the discussion (line 483):
"Our *noCloud*-filter means one approach to avoid reducing background humidity if lightning is correctly simulated but not observed, e.g., due to the fact that no lightning locating system can detect all flashes."

[7] Could the authors comment on the choice of RH as the sole variable of their 1DBay procedure? How

does their approach address the ambiguity caused by RH's dependence on both temperature and specific humidity? Indeed, the same RH value can be obtained for very different temperature and specific humidity values.
We used RH because it combines information from both specific humidity and temperature, both can be found in the control variables, and RH is related to cloud formation. 100% RH always means condensation is happening no matter what the actual temperature or specific humidity is. Previous studies have shown that pseudo RH profiles can be assimilated with positive results on the forecast (e.g., Caumont et al., 2010, Fierro et al., 2016, 2019, Boderies et al., 2019)

Olivier Caumont, Véronique Ducrocq, Éric Wattrelot, Geneviéve Jaubert & Stéphanie Pradier-Vabre (2010) 1D+3DVar assimilation of radar reflectivity data: a proof of concept, Tellus A: Dynamic Meteorology and Oceanography, 62:2, 173-187, DOI: 10.1111/j.1600-0870.2009.00430.x

Fierro, A. O., Gao, J., Ziegler, C. L., Calhoun, K. M., Mansell, E. R., and MacGorman, D. R.: Assimilation of Flash Extent Data in the Variational Framework at Convection-Allowing Scales: Proof-of-Concept and Evaluation for the Short-Term Forecast of the 24 May 2011 Tornado Outbreak, Monthly Weather Review, 144, 4373–4393, https://doi.org/10.1175/MWR-D-16-0053.1, 2016.

Fierro, A. O., Wang, Y., Gao, J., and Mansell, E. R.: Variational Assimilation of Radar Data and GLM Lightning-Derived Water Vapor for the Short-Term Forecasts of High-Impact Convective Events, Monthly Weather Review, 147, 4045–4069, https://doi.org/10.1175/MWR-D-18-0421.1, 2019.

Borderies, M., Caumont, O., Delanoë, J., Ducrocq, V., Fourrié, N., and Marquet, P.: Impact of airborne cloud radar reflectivity data assimilation on kilometre-scale numerical weather prediction analyses and forecasts of heavy precipitation events, Nat. Hazards Earth Syst. Sci., 19, 907–926, https://doi.org/10.5194/nhess-19-907-2019, 2019.

[8] Which types of observations are assimilated in experiment CTRL?
This information can be found in lines 309 to 312 in the text. Experiment CTRL assimilates all available data, except for the radar reflectivity and radar Doppler velocity and LDA (please see line 324/325).

[9] It would be very interesting to see a comparison between two (appropriately located) vertical crosssections of 3D-Var analysis increments in temperature and specific humidity, the first one from CTRL, and the second one for LDA−CTRL. This would help to check the consistency between the increments due to the lightning assimilation (alone) and those produced by all standard observations (as assimilated in CTRL). In other respects, maps of IWV increments alone do not give any hint on whether the vertical structure of 3D-Var increments is realistic.
We included a new Figure that shows the cross-section of RH for the same time and experiment that is shown in Fig. 6 (LDAnC). In contrast to the reviewer suggestions, we show RH as this is the quantity that we assimilate, and after looking at cross-sections of specific humidity and temperature, RH appears to be more interesting. Also, we show the difference between the analysis and the background, not between experiments CTRL and LDAnC where it is difficult to see the impact of our LDA, because experiments are cycled. The difference between analysis and background is equal to the 3D-Var increments and allows to visualize their vertical structure as requested by the reviewer.

[10] In section 7.1.2, a plot showing the RA maps for experiment CTRL should be added to help assess the relative improvements brought by RDA and LDA.

We added one panel for CTRL to the plot. The figure caption was updated accordingly, so was section 7.1.2.

[11] Figure 7: Unless article length is really an issue, it would be very helpful to add a figure similar to Figure 7, but showing the differences between the FSS of each experiment and the FSS of CTRL.
The curves in Figure 7 are sometimes hard to distinguish, especially for the two lowest RA thresholds.
The primary message of our article here is that the curves are all relatively close to each other. Showing the difference between the CTRL and the other experiments would over-emphasize the noise in the results that comes from under-sampling.

Scientific comments/questions:
[1] Somewhere in the text, it would seem relevant to underline the pros and cons of the proposed two-step (1DBay + 3D-Var) assimilation technique, compared to the direct 3D-Var (or 4D-Var) assimilation of lightning observations.
In general, a direct assimilation is always preferred as the additional 1DBay is an additional source of uncertainty. However, the direct assimilation requires that the assimilated quantity can be found among the control variables of the assimilation system. This is not the case for lightning (or FED) and either for most proxies, i.e., those based on hydrometeor contents or vertical velocity, for most NWP systems, in particular not for AROME-France. The 1DBay is then one method that retrieves a variable that can be assimilated based on the difference between the observation and simulation. For AROME-France, RH is directly related to prognostic variables and has the closest relation to convection.
Assimilating hydrometeor contents when VAR is able to update these variables often results in poor performances because the cross correlations with key variables such as temperature and humidity are poorly represented in climatological B matrices. To mitigate this effect, some recent studies assimilate humidity along with hydrometeor contents (see, e.g., Wang et al. 2013 https://doi.org/10.1175/JAMC-D-12-0120.1, Do et al. 2022, https://doi.org/10.1175/MWR-D-21-0292.1, for radar reflectivity).
The previous statement is added to the article (line 305).
We added a short statement in the beginning of section 5 (line 217):
“FED cannot be assimilated directly since lightning is not predicted by AROME-France. The previously described observation operator relates the FED observations to the prognostic variables of the NWP model and allows for comparing background and observations. The final step prior to the assimilation creates pseudo observations that can be ingested in the 3DVar system. ”

[2] Equation 2: The formulation of the weights assumes that observation and model errors follow Gaussian distributions. Is this assumption valid for FED data?
We added a Figure of innovations (FED observation – background FED) to show that this distribution is bell-shaped.
Even in case of not perfectly Gaussian observation and model errors, the assimilation improves the prediction. Previous studies have shown that LDA can give good results although the assumption of Gaussian errors is not perfectly met. By the construction of our observation operator at least the error distribution is centered around zero FED.

[3] line 248: Here, the authors could simply state that ”Since FED has no vertical dimension, the covariance matrix is reduced to a single value of variance, σ2o , which is assumed to be constant.” There is no need to mention the trace.
The suggestion is added to the text and replaces the previous sentence.

[4] line 254: Is the optimal value of σo expected to be sensitive to the size of the ”vicinity” (currently

set to 500 km) (index i in Equation 1)?
As long as the vicinity is sufficiently large, i.e., it is chosen so that we find non-zero weights for most FED observations while profiles in the vicinity can still be considered as meteorologically related to each other, then σ_o should get similar value even for different vicinities.

[5] line 286: Could the authors give a rough estimate of the percentage of profiles where their Humidity Adjustment needs to be applied?
We verified our assimilated pseudo observation profiles and got the following results for the August case that is detailed in the paper:
Total Bayesian profiles: 19485
Total HA profiles: 312
HA percentage on POs: 1.6

We added to line 286:
"In example, the HA generates 1.6% of almost 20,000 assimilated profiles during 08 Aug. 2018, the study case that is detailed below."

[6] line 332: Why would the thinning be applied in the latitude direction only?
It was an error in the text that is corrected. It is applied in both latitude and longitude direction so that a thinning factor of 2 means 1 in 4 observations, as stated.

[7] Equation 4, lines 418-419: In the definition of O and M, the brackets ([...]2) should not be squared (see Robert and Lean 2008). This is probably a typo.
We corrected it. Thank you for the detailed look at the equations.

[8] line 429: "FSS between 0.8 and 0.9 [...]". This is only true during the first 15 hours of the forecast.
We modified the text and added "FSS between 0.8 and 0.9 during the first 15 hours of the forecast […]" This forecast duration include the first 12 hours of the forecast that were stated as most interesting in this work.

[9] line 501: Could the authors expand a little on their idea of RDA-LDA coupling?
We were thinking of a coupled quantity that is based on FED and radar reflectivity to better identify precipitation and (convective) clouds. This quantity could be treated as an advanced radar *electrical* reflectivity, and there are different ways that could be tested. For a simple example, radar reflectivity could be increased in regions where lightning was observed and the radar reflectivity is lower than expected. The assimilation of this modified reflectivity could be compared to the operational RDA.

Minor comments:
Please note: The authors do not need to provide a formal reply to any of the following minor comments.
We will only comment the minor comments in case that we have a different suggestion or that we have a question.

General remark: The American spelling used throughout the manuscript might not be suitable for a European journal (please check with the Editor).
* line 26: Remove "among others".

The MTG-LI field of view will also include parts of South America, the Caribbean, North America, Asia, the Indian Ocean, the Black Sea, and even the Arctic and Antarctica. Therefore, we think mentioning "among others" is appropriate here.

* lines 26-27: The sentence "GEO lightning sensors [...]" should be removed, since it just repeats what was written on lines 22-23.
* line 30: Remove article "The" (end of line).
* line 32: Add "However, LDA [...]".
* line 38: "Despite those difficulties [...]".
* line 38 and 46: "former" → "previous".
* line 45: Another noteworthy reference would be Allen and Pickering (2002)[doi:10.1029/2002JD002066].
* line 65: "Federico et al. (2014) simulated" (past tense, for consistency).
* line 66: "the most reliable proxy [...]" (add article).
* line 67:"cell-scale" and "domain-scale" (hyphen).
* line 98: "one hour into the forecast".
* line 113: At which vertical levels was this RH-correction applied? Only in the troposphere?

Between cloud base and top height

* line 121: I would suggest to rephrase as follows: "Various approaches in which FED is assimilated together with radar reflectivity and/or Doppler wind [...]".
* line 126: Remove "reviews lightning observation operators and".
* line 128 "among others" → "including".
* line 132: Why not at 15 and 21h UTC? (just out of curiosity).

In general, there are 6 hours between each run. The 0300 UTC is an exception as forecasts in the morning are needed by forecasters for their daily production.

* line 135: "extends up to pressure level 10 hPa".
* line 155: "separation of" → "discrimination between".
* lines 164-165: The sentence "In all cases [...]" ought to be rephrased, to make it clearer.

In all cases, flashes with large extent, long duration, or with a high number of ISS-LIS events or LF strokes/pulses were more likely detected by both LLSs than the small, short duration flashes.

* line 166: "as a suitable input to the GEO [...]".
* line 167: Add a comma after "7 km".
* line 173: The acronym "1DBay" should be defined here, in the main text.
* line 175: "[...] is obvious in Figure 2(b)".
* line 176: "Figure 2 accumulates data of 5 min" sounds redundant, since the authors already mentioned that the period shown in Figure 2 is 13:55-14:00 UTC.
* lines 176-177: It would seem enough to write: "Throughout the domain, the FED pixels have a size of roughly 7 km x 7 km.".
* line 179: "strictly positive" → "non-zero".
* line 178: Rephrasing suggestion: "Figure 2(c) superimposes the zoomed-in simulated FED and the Meteorage [...] used as input to the GEO lightning [...]".
* line 183: I would suggest: "Previous studies suggest that graupel is a reliable [...]".
* line 184: May I suggest: "[...] is trained using the relationship found between MTG-LI FED observations and simulated graupel mass (mg) from hourly forecasts with AROME-France, for 24 days in 2018 (2 days per month)".

Graupel mass is used from the 1-hour forecast as this is the first complete analysis that is created by the model, and the short forecast period should still resample to the real atmosphere. The relationship is not built from hourly forecasts but the analysis of the first forecast hour as most accurate simulation of the real atmosphere.

"[…] is trained using the relationship found between MTG-LI FED observations and simulated graupel mass (m_g) from the 1-hour forecast of AROME-France, for 24 days in 2018 (2 days per month)."
* lines 185-186: "(e.g., as in Fierro et al. 2016 [...]".
* line 190: "without" → "regardless of".
* Figure 3 caption: "white points" and "black points" → "white dots" and "black dots".
* line 209: "to get the first lightning" → "to initiate lightning".
* line 211: "roughly follow".
* line 212: "the values of FED are always [...]" → "observed values of FED are always [...]".
* line 229: σo should be defined for the first time right after Equation 2.
* line 237: "estimation" → "estimate".
* line 251: "approximate 0" → "approach 0".
* line 258: "areF" → "are".
* line 260: Please check the exponent applied to "(7km x 7km 10min)" in the text.
The exponent was correct -1, however, the footnote index 1 was too close. We moved the footnote to index the value 1, not the unit.
* line 267: "The example of 07 Oct. 2018 00:00 UTC [...]".
* line 268: The whole sentence starting by "AROME dBFED (Figure 4c) [...]" is somewhat awkward. It ought to be rephrased and maybe split into two.
New: "Figure 4(c) indicates positive background dBFED in the region of lightning observations, however, also widespread near the center of the domain and over western France. "
* line 269: "[...] in the region of positive dBFED observations [...]".
We changed it to "[…] lightning observations […]" (see previous comment).
* line 270: "The 2D distribution of the 1DBay [...]" → "The map of the 1DBay [...]".
* line 272: "[...] 1DBay retrieves profiles [...]": are we really dealing with profiles, given that the 1DBay seems to be applied to FED values here?
No, FED is not a profile. If the 1DBay is used to get the RH pseudo-observations, we are retrieving profiles.
"[...] 1DBay retrieves profiles [...]" → "[...] 1DBay retrieves pseudo dBFED [...]":
* line 274: "meaning eventually no lightning activity," → "in other words no lightning activity,".
* line 276: "shallow dBFED in northwestern Italy, to the north of Corsica," → "weak dBFED over the Gulf of Genoa" ?
* line 278: "in the vicinity around this regional area" → "in this region" ?
We changed it to: "[…] in the retrieval vicinity". The term "vicinity" was defined for the 1DBay method.
* Figure 4: Remove "in general" at the end of the caption ?
* line 285: "of among others Fierro et al. (2019)" → "of Fierro et al. (2019), among others".
* line 290: "If, however, the specified vicinity were too small, one would need to artificially remove [...]".
* line 294: "has been initially" → "was initially".
* line 304: "by converting the 3D AROME-France outputs into".
* line 308: "allowing the assimilation".
* line 320: I would suggest to rename this section as "Experimental set-up".
* line 328: "[...] the use of all FED observations led to wrong results [...]".
* Table 1: I would suggest to modify the caption: "Set-up of AROME-France assimilation and simulation experiments: [...]".
* line 332: "Here, FED data are thinned by a factor of 2 [...], i.e., 1 in 4 observations".
* line 334: "1 in 64 radar observations".
* line 335: "than our FED data".
* line 336: "The thinning, thus, prevents from assimilating observations with correlated observation

errors , which would otherwise contradict the assumption of a diagonal R matrix.".
* line 344: "as they probably belong to spurious simulated thunderclouds.".
* line 345: "reduce background humidity in spurious convection.".
* line 349: "The reference, LDA and RDA experiments [...] prior to the evaluated analysis".
* line 351: "This time period has been chosen because convection was continuously observed inside the model domain." ?
* line 352: "to efficiently ingest all available observations".

* line 356: "the first 12 hours of the forecast".
* line 357: "the most effects" → "the strongest impacts".
* line 357: This sentence should probably be split into two, for clarity.
* line 366: "the amplifying trough": which trough is referred to, here?
"In the evening of 08 Aug., the trough that was situated over the Atlantic amplified and the associated cold front entered the study domain from the west."
* line 370: "thereafter and until the end of the forecast period": this bit of sentence does not sound right.
"Then, only few cells developed over the northern Mediterranean Sea until the end of the forecast period."
* line 373: "for experiment LDAnC".
* line 373: "Figure 5(g) shows that IWV is added to the background in regions where [...]".
* line 380: Is there a missing word in the sentence: "The 1DBay dBFED (Figure 5b) [...]" ?
Yes, it was also mentioned by RC1. We corrected the sentence:
"Over southern France, the marine blue spots Figure 5(b) indicate that the 1DBay retrieval succeeds in reducing the spurious AROME_dBFED (Figure 5a versus c)."
* line 382: "as shown in Figure 5g." (more concise).
* Figure 5 caption:
- I believe that each panel number should appear before each corresponding description. Please check (also in main text).
- Typo: "where no profiles were retrieved".
- Should not panel (h) rather be described as "1DBay-only IWV", to be consistent with main text?
* line 386: Suggestion: "The 1DBay method aims at finding the profiles that lead to analysis increments which improve the fit of simulated dBFED values to observations.".
* lines 387-389: Splitting this long sentence into two shorter ones would make it more readable.
* line 389: A full stop seems to be missing near the end of this line.
* line 390: "In addition, Figure 5(h) displays [...]".
* Section 7.1.2: Use present tense, when describing the rain validation from the experiments, throughout this section. Also, is there any reference available for Antilope?
We added 2 references for Antilope:
Laurantin, O., 2008: ANTILOPE: hourly rainfall analysis merging radar and rain gauge data. *Int. Symp. on Wea. Radar and Hydrol. (WRaH2008)* (Grenoble and Autrans, France, 10-15 March 2008).
Laurantin, O.: ANTILOPE: hourly rainfall analysis over France merging radar and rain gauges data, *Proceedings of the 11th International Precipitation Conference*, edited by: Leijnse, H. and Uijlenhoet, R., KNMI, Ede-Wageningen, the Netherlands, 30 June to 3 July 2013

* line 396: Remove "ground-based".
* line 397: "RA maps for the first 6 hours of the forecast".

* line 398: Remove ”the” (before experiments).
* Figure 6 caption: ”the first 6 hours of the 30h-range forecast.”
* line 408: ”however” does not seem to be the right word here (rather simply: ”and” ?).
We removed the second part of the sentence as it repeats the information from the previous sentence.
* line 415: Add ”in a statistical way” at the end of the first sentence. Remove the second sentence, which is redundant.
* line 421: ”and 1 otherwise.”.
* line 430: Remove ”conducted”.
* line 434: Add a comma between ”forecast” and ”demonstrating”.
* line 434, second sentence: Suggestion: ”The combination of LDA and RDA only becomes beneficial beyond the 12-hour forecast range”.
* line 436: Remove ”the” before ”LDA”.
* line 437: ”to predict the heavy precipitation amounts” → ”for predicting heavy precipitation”.
* line 437 (end): ”mostly improves” → ”greatly improves” ?
* line 441: ”the other case studies analyzed [...]”.
* line 442: Meteorologically, 1-3 Feb cannot be described as an ”early spring” situation.
Indeed. We changed it to “winter”.
* line 444: The sentence ”The combination of RDA and LDA exhibits [...]” should be rephrased for better clarity.
We modified the order of information:
“Once experiment *RDA* and once *LDAnC* provide the best RA forecast. Experiments combining RDA and LDA exhibit FSS values between *RDA* and *LDA*, thus, mean the best trade-off overall.”
* line 454: Remove ”of” in ” [...] to design of an assimilation [...]”.
* line 456: Please add ”pseudo” before ”MTG-LI flash extent density”.
* line 458: Remove ”the” before ”LDA”.
* line 461: The reference given in brackets is not needed in the conclusions.
* line 479: Missing ”to” after ”the vicinity used”? Is ”expanded” the right word here?
“In detail, a wider vicinity is used to identify vertical profiles in the 1DBay method and a so-called *noCloud*-filter is introduced.”
* line 484: ”i.e., through correlations [...].
* line 488: ”Finally, forecasts of 6-hour rain accumulations [...]” ?
* line 489: ”the established RDA”.
* line 503: Typo: ”explored”.
* line 504: ”also benefit forecasts [...]”.
* line 505: ”our pseudo MTG-LI observations”.