Response to referee #1

We thank the Referee for his careful reading, and the valuable comments. We considered each of the comments and questions and provided an adequate answer. The comments of Referee #1 are written in bold and the answers in plain text. Sentences indicating a modification to the manuscript are written using italic font.

The latex diff of the revised manuscript, accounting for comments of referees #1 and #2 is provided as an attached file.

General comment

The authors compare simulations made by several M-C codes and come with useful recommendations on simulation procedures (energy cuts, etc..). The analysis is detailed and there are no doubts that all codes in own limits produce rather coherent results. The code verification and comparisons of different code options, as well as different codes, are necessary first step of simulation experiments and constructing of models to be compared with experiment.

We would like to clarify that we do not intend to validate/verify any code in this article. The general purpose code Geant4 has already undergone multiple verification and validation studies in several physical contexts, but nothing guarantees it can be straightforwardly applied to HEAP (also referred as HEPA by the referee) phenomena, in particular concerning the capacity of simulating Relativistic Runaway Electron Avalanches (RREA) that we are extensively testing here. This issue was first raised by (*Skeltved et al.*, 2014). Therefore, we think it is important to provide the researchers of the community with clear tests frameworks, together with reference data, to make it possible for them to benchmark their custom made codes. On our side, with this series of two GMD papers (*Rutjes et al.* (2016) and this one), we are settling the first foundations of a larger, long-term work.

However, we have to understand, that M-C simulations for such a complicated domain as High-energy Physics in Atmosphere (HEPA) is not a precise tool! We don't know the distribution of the electric charges in the cloud and, therefore, strength and elongation of the emerging electric fields. Therefore, very time-consuming and detailed verification of different M-C programs, for opinion of this referee is not too important on the present stage of HEPA progress. As I mention in my review to the first paper of this series in 2016, the validation of the available experimental observations is vital for the progress of HEPA. There are published numerous gamma ray energy spectra observed on the mountain altitudes and few electron energy spectra; why not to try to compare simulations with observations? Continuous simulations with different codes and arbitrary parameters (sometimes nonrealistic, see my comment below to 4.3.1) can make illusion of intense scientific research; however only comparisons with observations and physical inference on the observed phenomena really values. Sure, authors will argue that model validation is out of scope of their paper. And they will be right. However, I can ask, when they will use their verified models for coming with comprehensive model of HEPA? When they will develop models with realistic parameters and compare it with data (energy spectra measured on Earth's surface and in the space)?

Comparison with experimental data is not the goal of this work. However, we think it is worth mentioning here that we are working on other projects in parallel, for which having tested the coherence between these three different codes beforehand is important. Also, we would like to point out that the REAM model, tested in this study, was used to analyze experimental data detected from space and from the atmosphere in many occasions (see *Dwyer et al.* (2012)). Furthermore, compared to our status in 2016, a part of the team involved in the present study (including the first author) is now conducting, in parallel, Geant4 simulations attempts in comparison with experimental data, through two separate studies. They both concern aircraft-based measurements, one from 12 km altitude (ILDAS campaign) and the other from 20 km altitude (ALOFT/FEGS campaign); and there also having first found that there is a form of consensus between the different available RREA models is important to interpret the results. Some of these results have been very recently published *Kochkin et al.* (2018).

[...] Continuous simulations with different codes and arbitrary parameters (sometimes non-realistic, see my comment below to 4.3.1) [...]

See answer below.

To be not too didactic, I'll cite our old paper, where we try for the first time to compare simulated energy spectra with measured ones and establish a TGE model (see Figs 8-12 of Chilingarian, Mailyan and Vanyan, 2012). "With newly estimated thundercloud height, we re-estimate several phenomenological parameters of the RREA process as the following: the most probable height of thundercloud (and electrical field therein) is ~50 m. The number of electrons with energies ° above 1 MeV at the exit from the cloud is 1.97107 electrons/m2/min; if we assume that the radiation region in the thundercloud has a radius of 1 km the total number of ° electrons crossing this region in a minute is ~61013.

Sure, we use not optimized M-C; and, maybe we make some mistakes in our inference. We discuss possible sources of the systematic errors:

"We do not measure the electric field within the thundercloud; near surface electric field is not a good proxy of the intracloud fields accelerating electrons downward. We also do not measure vertical extension of the field and only estimate the height of the cloud. Therefore, simulations of the RREA process in the atmosphere with chosen parameters, although are in an agreement with the available measurements of electric fields in the thunderclouds, cannot be used for direct comparisons with TGE measurements. However, these simulations give us understanding of the RREA scale and MOS processes and expected behavior of the energy spectra."

However, it was the first time we present gamma ray and electron energy spectra along with simulations and achieve overall agreement. Now we develop a new method of cloud height estimation and can approach observations with more realistic simulations with more reliable better parameters. What I want to demonstrate is that simulation should be paired with experimentation; and each should profit from other.

We thank the referee for pointing out and explaining some of the very interesting projects his team has been working on. Fortunately, his review will stay available for future use on the GMD webpage of our article.

After our recently observation of Long Lasting Low energy TGEs (LLL TGEs) – a hours extending flux of gamma rays of 0.3 – 3 MeV energies, we started a cycle of simulations to get answer if remote Extensive Cloud Showers (ECSs) can contribute to this flux, or we should consider stochastic electron acceleration by a "sea" of randomly distributed charges in the thundercloud. Thus, simulations are pairing with observations and with hypothesis testing.

We thank the referee for pointing out this very interesting work in progress.

Specific questions :

5 The emerging field of High Energy Atmospheric Physics studies events producing high energy particles and associated with thunderstorms, such as terrestrial gamma-ray flashes and gamma-ray glows. Without mentioning Thunderstorm Ground Enhancements (TGEs) this statement is biased. The difference in duration between TGF and gamma-ray glows can be explained by two possible different scenario to create runaway electrons. Largest TGE detection prove that long duration can be explained by the continuous acceleration and multiplication of seed electrons entering strong prolonged electric field. Such a condition can sustain for minutes and, so called, extensive cloud showers (ECSs) will produce fluxes of electrons, gamma rays and neutrons on the earth's surface, i.e. TGEs (see Chilingarian et al., 2017).

In the first version of the manuscript, the expression "Thunderstorm Ground Enhancements (TGEs)" is mentioned in the introduction (page 2, line 16 of the non-revised manuscript) but not in the abstract. We first decided not to use it in the abstract for simplification for the reader, as we think they are not an intrinsically different phenomenon from gamma-ray glows. That is also the opinion expressed in the review paper *Dwyer et al.* (2012). However, it is not a problem to present the expression "thunderstorm ground enhancements" in the abstract too.

page 1, line 3 of the revised manuscript : The expression "thunderstorm ground enhancements" was added in the abstract.

In general, this issue touches on present discussions in the HEAP community, but giving a judgment of who is "right" about it, is far from the scope of our article.

1.3 5 "The physics behind TGF, TGF afterglows and gamma ray glows are studied with the help of computer simulations, which necessarily involves model reduction and assumptions."

Hopefully physics is experimental science and most of results are obtained by experimentation, not simulation.

We think that a lot of important results in the HEAP community were obtained by comparing experimental results with simulations, together with analytical calculations. The sentence has been reworded to be more adequate.

page 7, line 4-6 of the revised manuscript : "The physics behind TGF, TGF afterglows and gamma- ray glows are studied with the help of computer simulations, which necessarily involves model reduction and assumptions."
→ "Apart from analytical calculations, the physics behind TGFs, TGF afterglows and gamma-ray glows are also studied with the help of experimental data, computer simulations, and often a combination of both."

4.3.1 5 Figure 7 compares the electron spectra recorded at 128 meters, for an electric field E = 0.80 MV/m, for a RREA generated from 200 seed electrons with ε = 100 keV. Do you especially choose the field never measured in the atmosphere (0.8 MV/m) ?

First of all, we found in the literature one reported measurement from balloon soundings, that shows an electric field of about 200 kV/m at 12 km altitude (see *Stolzenburg and Marshall* (2009), figure 3.2 on page 65). As a reminder this is equivalent to about 0.86 MV/m at sea level, as it scales with density. In this article, such a field is associated with the balloon being close to a lightning leader. All the results are presented at STP for reference, but can be scaled to higher altitude (-> lower densities), the scaling of the different parameters being function of air density.

Furthermore, it is important keep in mind that the codes we are testing here are used by researchers that try to explore extreme hypothetical cases, because it is what may be needed to explain some gamma ray glows, and TGFs. Monte Carlo codes are routinely used by some researcher in the HEAP community in such regimes. The TGF production mechanism in particular is poorly understood, but it may require extreme potentials (>200 MV), leading to electric fields of $E = 0.80$ MV/m or more; maybe only possible for extreme thunderstorms, where measurement are not easy to get (because of rarity, and extreme conditions); maybe only possible from some extreme lightning leaders or thunderstorm configurations.

In addition, an electric field of $E = 0.80$ MV/m is about three time less the classical breakdown field ≈ 3 MV/m at sea level, and we cannot exclude that such fields can exist, at least on a small scale and/or at high altitude (with the proper scaling factor), at which it is rather complicated to obtain in-situ measurements. In particular, simulations investigating the cold runway mechanism can require to simulate localized electric fields of 4 MV/m or more, as presented, e.g., in the recent theoretical work of *Lehtinen and Østgaard* (2018). This last article also points out that 1 MV/m is the typical field in the electrode gap used in laboratory spark experiments (typically 1 meter distance), that is also a possible application of the models used in this study; even if when simulated, these high electric fields usually extend only in a very short range (centimeter or less) and time ($\sim$ micro-second scale).

Page 3, lines 8-13 of the revised manuscript: For completeness, a sentence was added in the introduction about x-ray emissions observed in laboratory spark experiments with a series of interesting references.

Or it is not in atmosphere?

All the electric fields of this study are applied in the earth's atmosphere at sea level, Standard Temperature and Pressure (STP).

Why 128 m? Where do you inject 200 electrons?

For the results presented in this figure, 200 electrons are generated at the origin ($x = 0, y = 0, z = 0$), the record is made at 128 m ($x = 0, y = 0, z = 128$), z being the direction of the electric field, and electrons are accelerated towards positive z.

This was clarified throughout the text, also responding to a comment of referee #2. See page 24, figure 7; page 25 figure 8; page 22, line 19 and page 23, line 21.

This distance of 128 m is chosen because it corresponds to a large enough number of avalanches lengths (about 8.5 avalanche lengths in this case), that corresponds to a maximum multiplication factor of about 5000, where we are sure that RREA is fully developed and has reached the self-similar state. Such multiplication factor were also necessary to be able to produce a database with enough statistic to build, with low noise, the avalanche length and avalanche times curves (figure 3 and 4).

page 22, line 20-21 of the revised manuscript : a sentence was added to justify the record at z=128 meters.

Spectra where the particles are recorded at sorter distances are also presented in the supplementary material, and correspond to snapshots of the RREA evolution with lower multiplication factors.

Furthermore, a multiplication factor of 1000 or more can be obtained by some groups when they attempt to model possible TGF production mechanisms (see, e.g., (*Dwyer*, 2008; *Carlson et al.*, 2010)), and may also be necessary to explain the most extreme gamma-ray glow observations, where increases of the background intensity of this order of magnitude were detected (see, e.g., (*Eack et al.*, 1996; *Eack et al.*, 2000), and some up-coming glow studies).

One way of deciding which model is the most accurate might be to compare these results with experimental measurements. but in the context of TGF and Gamma-ray glows it is complicated to get a proper measurement of electron spectra produced by RREA. Finally, yes, only way to decide which model is true is the comparison with experiment that is missing in this paper.

See response to the general comment above.

Others changes (not directly suggested by referees)

During the revision process, several extra improvements to the paper were suggested by the authors:

- *Page 16, line 26 of the revised manuscript: A citation to the article "Fundamental parameters of the relativistic runaway electrons avalanche in air" by Babich et al. (2004) was added; as it is an important study to mention in the context of this work.*
- *Page 2, line 21 of the revised manuscript: For completeness, we added two citations to the recent gamma-ray glow observations of (Kochkin et al., 2018; Dwyer et al., 2015) in the introduction.*
- *abstract (page 1-2): Improvements in the English, added more details on the Electric field range.*
- *conclusions (page 26-27): Improvements in the English, added more details on the Electric field range, and more details.*
- *The second paragraph of the introduction (page2, line 11-13) was updated to give two more interesting citations about TGF satellite observations (one for RHESSI, one for AGILE and a more recent one for Fermi).*
- *Figure 2: The 10%, 50% and 90% probability contours for the REAM model (red curves) were added, together with a paragraph discussing how it compares with Geant4 (Page 13-14, lines 34-35 and 1-9).*
- *Page 3, lines 29-32 of the revised manuscript: We added a small paragraph clarifying the differences between the different values (between 2.36 and 3.2 MV/m) of the classical breakdown field, that can been seen in the literature.*
- *Page 13, lines 4-5 of the revised manuscript: a sentence was added to justify the use of this particular {E,ϵ} set.*

- *Page 23, lines 1-4 of the revised manuscript:* a sentence indicating why an electric field of $E = 0.80$ MV/m is used for the comparison case was added.
- *Page 13, lines 15-16 of the revised manuscript:* an indication that we provide Geant4 examples source codes with tweakale α_R and $\delta\ell_{max}$ parameters was added.
- Page 12, lines 23-25 of the revised manuscript: We added a sentence about the chosen direction of the initial electrons in the RREA probability simulation and its impact on the probability.

References

Babich, L. P., E. N. Donskoy, R. I. Il'kaev, I. M. Kutsyk, and R. A. Roussel-Dupre (2004), Fundamental parameters of a relativistic runaway electron avalanche in air, *Plasma Physics Reports*, *30*(7), 616–624, doi:10.1134/1.1778437.

Carlson, B. E., N. G. Lehtinen, and U. S. Inan (2010), Terrestrial gamma ray flash production by active lightning leader channels, *Journal of Geophysical Research: Space Physics*, *115*(A10), doi:10.1029/2010JA015647.

Dwyer, J. R. (2008), Source mechanisms of terrestrial gamma-ray flashes, *Journal of Geophysical Research: Atmospheres*, *113*(D10), n/a–n/a, doi:10.1029/2007JD009248, d10103.

Dwyer, J. R., D. M. Smith, and S. A. Cummer (2012), High-Energy Atmospheric Physics: Terrestrial Gamma-Ray Flashes and Related Phenomena, *Space Science Review*, *173*, 133–196, doi:10.1007/s11214-012-9894-0.

Dwyer, J. R., D. M. Smith, B. J. Hazelton, B. W. Grefenstette, N. A. Kelley, A. W. Lowell, M. M. Schaal, and H. K. Rassoul (2015), Positron clouds within thunderstorms, *Journal of Plasma Physics*, *81*(4), 475810405, doi:10.1017/S0022377815000549.

Eack, K. B., W. H. Beasley, W. D. Rust, T. C. Marshall, and M. Stolzenburg (1996), Initial results from simultaneous observation of x-rays and electric fields in a thunderstorm, *Journal of Geophysical Research: Atmospheres*, *101*(D23), 29,637–29,640, doi:10.1029/96JD01705.

Eack, K. B., D. M. Suszcynsky, W. H. Beasley, R. Roussel-Dupre, and E. Symbalisty (2000), Gamma-ray emissions observed in a thunderstorm anvil, *Geophysical Research Letters*, *27*(2), 185–188, doi:10.1029/1999GL010849.

Kochkin, P., D. Sarria, C. Skeie, A. P. J. Deursen, A. I. Boer, M. Bardet, C. Allasia, F. Flourens, and N. Østgaard (2018), In-flight observation of positron annihilation by ildas, *Journal of Geophysical Research: Atmospheres*, *0*(0), doi:10.1029/2018JD028337.

Lehtinen, N. G., and N. Østgaard (2018), X-ray emissions in a multiscale fluid model of a streamer discharge, *Journal of Geophysical Research: Atmospheres*, *123*(13), 6935–6953, doi:10.1029/2018JD028646.

Rutjes, C., D. Sarria, A. B. Skeltved, A. Luque, G. Diniz, N. Østgaard, and U. Ebert (2016), Evaluation of monte carlo tools for high energy atmospheric physics, *Geoscientific Model Development*, *9*(11), 3961, doi:10.5194/gmd-9-3961-2016.

Skeltved, A. B., N. Østgaard, B. Carlson, T. Gjesteland, and S. Celestin (2014), Modeling the relativistic runaway electron avalanche and the feedback mechanism with geant4, *Journal of Geophysical Research: Space Physics*, *119*(11), 9174–9191, doi:10.1002/2014ja020504.

Stolzenburg, M., and T. C. Marshall (2009), *Electric field and charge structure in lightning-producing clouds*, Springer.