THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Felix ERDMANN

Le 9 décembre 2020

Préparation à l'utilisation des observations de l'imageur d'éclairs de Météosat Troisième Génération pour la prévision numérique à courte échéance

---

Ecole doctorale : **SDU2E - Sciences de l'Univers, de l'Environnement et de l'Espace**

Spécialité : **Océan, Atmosphère, Climat**

Unité de recherche :
CNRM - Centre National de Recherches Météorologiques

Thèse dirigée par
Olivier CAUMONT et Eric DEFER

Jury

M. Eric BRUNING, Rapporteur
Mme Vassiliki KOTRONI, Rapporteure
Mme Wiebke DEIERLING, Examinatrice
M. Sylvain COQUILLAT, Examinateur
M. Serge SOULA, Examinateur
M. Thomas FARGES, Examinateur
M. Olivier CAUMONT, Directeur de thèse
M. Eric DEFER, Co-directeur de thèse

Abstract

The European Meteosat Third Generation Lightning Imager (MTG-LI) will provide continuous and large-area total lightning observations over Europe, Africa, and the Atlantic. This work prepares and develops the assimilation of such data in the French operational convection-permitting NWP model AROME-France and includes generating pseudo MTG-LI observations.

Lightning locating systems (LLSs) covering France, i.e., the low frequency (LF) Meteorage, the very high frequency SAETTA Lightning Mapping Array (LMA), and the low Earth orbit Lightning Imaging Sensor (LIS) on the International Space Station (ISS) are first studied to understand features of different types of lightning observation. This rare multi instrumental intercomparison documents, through a novel matching algorithm, the physical characteristics of coincident flashes such as extent, duration, detected flash components, and signal strengths. Among others, a degradation of LIS flash detection efficiency (DE) relative to Meteorage is found for low altitude flashes, and for small and short duration flashes.

With the advent of the Geostationary Lightning Mappers (GLMs), similar to MTG-LI, on GOES-East/West, a unique resource to simulate MTG-LI-like data became available. Coincident GLM and LF National Lightning Detection Network (NLDN) flashes serve to train so-called target generators that use various (i.e., 196) machine learning models (e.g., random forest, neural network, support vector machine) to produce pseudo-GLM flash characteristics. The method simulates pseudo-GLM flash extent, duration, and number of events using 6 NLDN flash characteristics. Pseudo-GLM events are then constructed using the simulated characteristics. Eventually, the most performant GEO lightning pseudo-observation generator, i.e., a linear Support Vector Machine-based generator that standardizes the used flash characteristics and applies a multi-step approach to sequentially predict the three pseudo-GLM flash characteristics, is identified through comparing statistics of pseudo-GLM flash characteristics and inferred Flash Extent Density (FED) products to those of the operational GLM observations.

As this model is trained for data in the US, ISS-LIS observations are used as a common reference to indirectly compare Meteorage and NLDN records. The good agreement in Meteorage and NLDN flash characteristics statistics (relative to ISS-LIS) encourages applying the generator without further adaptation in order to generate pseudo MTG-LI data from Meteorage records over France. As a final step before MTG-LI data assimilation can be explored, FED is derived from pseudo MTG-LI data.

Then, AROME-France background FED is inferred from a simple linear relation between column graupel mass (as proxy) and pseudo MTG-LI FED. A 1D Bayesian approach provides the best estimation of the relative humidity (RH) profile in the vicinity of each pseudo MTG-LI FED observation. The Bayesian approach redistributes model humidity and minimizes the risk of wet or dry biases. Indeed, background integrated water vapor can be reduced where the background FED exceeds the observed FED to suppress spurious convection. Humidity can also be added to the background in regions of higher observed FED than background FED to promote convection. The corresponding RH profiles are then assimilated in the 3D variational (3DVar) system of AROME-France. During this final assimilation, it is found that

the 3DVar analysis contradicts the FED-derived RH-profiles as the analyzed Integrated Water Vapor (IWV) increases relative to the background FED despite the FED-derived IWV indicates a IWV reduction in the same region. As a consequence, the experiments using a lightning data assimilation show lower Fractions Skill Scores of 6-hour accumulated rainfall than the control experiment and experiments using radar data assimilation for one test case.

Résumé

L'imageur d'éclairs (LI) de la mission spatiale européenne Meteosat Troisième Génération (MTG) fournira des observations de l'activité électrique totale de manière continue et sur un grand domaine couvrant l'Europe, l'Afrique et l'Atlantique. Ce travail prépare et développe l'assimilation de ces nouvelles données par le modèle opérationnel de prévision numérique du temps AROME-France de Météo-France et comprend la conception et la validation d'un générateur de pseudo-observations du capteur MTG-LI.

Dans un premier temps, les systèmes de localisation des éclairs (SLE) couvrant la France, à savoir le réseau Météorage opérant dans le domaine des basses fréquences radio (LF), le réseau SAETTA opérant dans le domaine VHF et le détecteur spatial optique Lightning Imaging Sensor (LIS) installé sur la station spatiale internationale sont étudiés pour comprendre les caractéristiques des observations, de nature différente, des éclairs. Cette rare comparaison multi-instrumentale documente, grâce à un nouvel algorithme d'appariement, les caractéristiques physiques des éclairs simultanément observés tels que l'étendue, la durée, les types de composantes détectées, ou encore les courants, radiances et niveaux de rayonnement électromagnétique. L'analyse de ces appariements a mis en évidence, entre autres, une dégradation de l'efficacité de détection des éclairs de l'instrument LIS par rapport à Météorage pour des éclairs situés à basse altitude et pour des éclairs peu étendus spatialement et de courte durée.

L'avènement des détecteurs américains d'éclairs Geostationary Lightning Mappers (GLM), qui fournissent des observations similaires à celles de MTG-LI, sur les satellites géostationnaires GOES-East/West, offre une ressource unique pour simuler des données de type MTG-LI. La seconde partie de la thèse se focalise sur la génération de données synthétiques MTG-LI. Pour cela, les caractéristiques des éclairs simultanément détectés par GLM et le réseau basse fréquence National Lightning Detection Network (NLDN) servent à entrainer 196 générateurs d'éclairs qui utilisent différents modèles d'apprentissage automatique (par exemple Random Forest, Neural Network, Support Vector Machine) pour produire des caractéristiques électriques réalistes. Chaque méthode simule à partir de 6 caractéristiques fournies par NLDN, et pour chaque éclair, l'étendue d'un éclair de type GLM synthétique, sa durée et le nombre de pixels illuminés. Les répartitions spatio-temporelles des pixels illuminés de chaque éclair synthétique sont ensuite réalisées en utilisant les caractéristiques simulées. L'analyse a identifié comme meilleur générateur de données synthétiques GLM le générateur employant la méthode linear Support Vector Machine appliquée sur des variables normées, et construit sur une approche en plusieurs étapes pour prédire séquentiellement les trois caractéristiques d'un éclair synthétique GLM. Cette analyse de performances a été conduite sur l'ensemble des 196 générateurs en comparant de manière statistique les caractéristiques des éclairs synthétiques et de leur densité d'étendue à celles déduites des observations GLM coïncidentes.

Dans un troisième temps, comme ce modèle est construit à partir de données collectées sur le territoire américain, les observations du détecteur spatial LIS sont utilisées comme référence commune pour comparer indirectement les performances des deux réseaux terrestres Météorage et NLDN. La bonne concordance entre les statistiques issues des observations communes

Météorage-LIS d'une part, et de NLDN-LIS d'autre part, valide l'application du modèle sans adaptation profonde aux observations du réseau français Météorage afin de générer des données synthétiques MTG-LI, moyennant une simple adaptation aux résolutions spatio-temporelles de MTG-LI. Enfin, ce jeu de données synthétiques MTG-LI est utilisé pour dériver une densité d'étendue des éclairs (FED, pour *Flash Extent Density*) en vue de l'assimilation des données MTG-LI dans le modèle de prévision météorologique AROME.

Enfin, la FED de l'ébauche d'AROME-France est déduite d'une simple relation linéaire entre la masse de graupel intégrée sur la verticale (comme *proxy* ou variable intermédiaire) et la pseudo FED de MTG-LI. Une approche bayésienne 1D fournit la meilleure estimation de profil d'humidité relative au voisinage de chaque pseudo-observation de FED de MTG-LI. L'approche bayésienne redistribue l'humidité du modèle et minimise le risque de biais humide ou sec. En effet, la vapeur d'eau intégrée de l'ébauche peut être réduite là où la FED de l'ébauche dépasse la FED observée pour supprimer la convection parasite. L'humidité peut également être ajoutée à l'ébauche dans les régions où la FED observée est plus élevée que la FED de fond pour favoriser la convection. Les profils d'humidité relative correspondants sont ensuite assimilés dans le système variationnel 3D (3DVar) d'AROME-France. Au cours de cette assimilation finale, on constate que l'analyse 3DVar contredit les profils d'humidité relative dérivés de la FED observée, car le contenu intégré en vapeur eau (IWV) analysé augmente par rapport à la FED de l'ébauche alors que l'IWV dérivé de la FED observée indique une réduction d'IWV dans la même région. En conséquence, les expériences utilisant une assimilation de données d'éclairs montrent des performances en termes de FSS (*Fractions Skill Score*) pour les précipitations cumulées sur 6 h inférieures à celles de l'expérience de contrôle et des expériences utilisant l'assimilation de données radar pour un cas de test.

Acknowledgements and Special Thanks

This thesis was quite an adventure. Everything started with an interview from Dubai, and later a German guy arrived to France. I must say that I enjoy the mentality of the people in Toulouse in general. I've never had a problem communicating even during the first time with broken French, everybody was sympathetic. Then there was a period when all was smooth sailing, until the Covid pandemic changed life completely. The adventure became even more challenging, and the year 2020 was not going to plan at all. However, I am happy to say that we – many supporters and I – made the best out of the new conditions. Finally, this thesis is the result of all the efforts and it wouldn't be possible without support.

First of all, a "géant merci" to my PhD supervisors Olivier Caumont and Eric Defer. You always have an open ear and I was very lucky to have both of you as my advisors. Your support started when I came to France, you provided expertise at work, and also for organizing my future helping me with research proposals and applications. We had a very good time, both at work and personally. Together, we managed the difficulties of the year 2020 and I hope you are as happy as I am with the result of this thesis. Regarding science, I could benefit from Olivier and his expertise in everything related to numerical models and the prediction of thunderstorms. You were always available for my questions, and I learnt a lot about how you attack scientific questions and challenges. This will be a very valuable skill during my scientific career. Your help included technical support with the AROME assimilation system and LaTeX, too. Eric, thanks a lot for your immense support in the beginning, accompanying me to agencies and other places. I am not sure whether I would have mastered the French administration without your support. Then, I benefited from your knowledge of lightning and microphysics, and learnt a lot as your Padawan. Due to this thorough knowledge, I could be more critical with my own work. Our work trip to the NCAR and the MSFC was a real pleasure, a great success both regarding work and having an amazing time. I am thankful for critical reviews of our papers and my thesis chapters although it meant sometimes a lot of work. I could clearly see where I can improve my academic writing skills. I hope that the three of us will stay in touch, not only to discuss work but especially as friends.

I thank the CNES and Météo-France for the funding of my thesis. I got to known Pierre Tabary who followed my thesis as a CNES scientific expert. Thanks Pierre for accepting the invitation to my jury and to the thesis defense. I enjoyed discussing with you. I thank Anne-Marie Cerrato for her administrative effort during the entire three years and to prolong my PhD contract for the months of November and December 2020 as a response to the difficulties enforced by the Covid-19 pandemic.

I also thank the Université Toulouse III Paul Sabatier and the Ecole Doctorale SDU2E for accepting and supporting my PhD.

All members of my jury were friendly and fair evaluating my work, with Sylvain Coquillat as president. A special thanks goes to Vassiliki Kotroni and Eric Bruning for their reviews. Thanks for not hesitating to take part at my defense that took place very early for Wiebke Deierling and Eric B., and quite late for Christelle Barthe. Thank you also Serge Soula for accepting

to be a jury member despite a very short preparation time. Thomas Farges, Christelle, and Sylvain were also members of my comité de these, thanks for your support during the entire three years.

I spent the first 14 months of my thesis at the Laboratoire d'Aérologie, an excellent place to work. Jean-Pierre Pinty shared an office with me, and I am thankful for this experience, some short French lessons, and an always enjoyable atmosphere. I found new friends at the LA - Carlos, Hervé, Irene, Laurent, Romulo, Thomas, Thibaut, Zaida, Zoé – we really had some fun times in Toulouse. I would also like to mention in particular the SAETTA team: Sylvain Coquillat, Eric Defer, Pierre de Guibert, Jean-Pierre Pinty, Veronique Pont, Serge Prieur, and Evelyne Richard, as well as the younger members Keun-Ok Lee, Pauline, and Ronan – I love(d) being part of this little family. I will always keep our field campaign in good memory. On Corsica I also met new friends such as Stavros, Michael K., and Jeoffrey, and experts like the SAFIRE team. Let me also say thank you to Chien Wang who arrived at the LA and shared his experience in machine learning applications with me.

Then, my office moved from the LA to the CNRM. All the people involved made the transition as easy as it could possibly be. Thanks to the entire group PRECIP at the CNRM with their warm welcome. I loved working here because there are always happy and smiling colleagues around. I enjoyed the ambiance and especially the coffee breaks that also included the PHY-NH team. If I try to name all the pleasant people here, I might forget to mention someone. Please feel my thankfulness if you are/were a member of those working groups. I'd like to thank Laurent in addition for his guidance to receive data files and for his help with IT issues. I thank everybody who provided technical support for the assimilation and Olive, especially Véronique Mathiot, and PAR for all their help with the administration. César, thanks for welcoming me to "your" office, making it a very smooth transition from the LA to the CNRM. Your help during the first weeks was very appreciated, you introduced me to the group of young researchers at the CNRM. You also provided some guidance for the format of the manuscript while you finished your own thesis. The entire group of young researchers was a pleasure to spend time with, talk, have lunch breaks, and just have fun - Alexane, Clément, Damien, Erfan, Marc, Mary, Natalia, Sabine, and all the others.

During my thesis, I was in the lucky position to present my work at conferences in Lecce (Italy), Phoenix, AR (USA), Boston, MA (USA), Boulder, CO (USA), and Huntsville, AL (USA). Here, I met many inspiring scientists, experts in atmospheric electricity, including authors of papers that I just read, I had stimulating discussions, and could establish contacts that will help me during my career. I'm thankful that I got to know many great people in person; among others Wiebke Deierling, Jim Dye and Scott Ellis from NCAR, NASA experts Rich Blakeslee, Hugh Christian, and Bill Koshak, and experts such as Eric Bruning, Paul Krehbiel and Ron Thomas.

Finally, my family and friends in Germany: We are always in contact and I am very happy to have all of you. My cousin Gernot was also helping me to find an apartment in Toulouse and to go my first steps in France in general. There are some more difficult phases during the course of studying and writing a thesis. I know I can always call you, talk to you, and get a motivating feedback. Despite the physical distance, this social support was a very important factor to keep me going and earn my PhD. I cannot say how happy I am to hear your reactions, to hear that you are proud of me. This achievement was possible due to your support.

Contents

Introduction

Contents of this Part

Introduction

Lightning strikes are fascinating and powerful phenomena. They have been observed by humans for thousands of years. Historically, people could not understand such phenomena, therefore, lightning and related thunder were associated with higher forces and gods, e.g., Zeus and Thor. Electricity was discovered in the 18th century. The famous experiments of Benjamin Franklin (1752), who published the results of using a kite with iron rod discovering that clouds can be electrified, aroused interest of scientists in natural and cloud electricity and encouraged further research.

Lightning is a very energetic phenomenon. It emits not only an optical signal, but also other electromagnetic radiation on a broad spectrum, from very low frequency (VLF) of a few hertz to very high frequency (VHF) of about 100 MHz (Nag et al., 2015). In addition, lightning discharges heat the air in and near the channel to temperatures up to 30,000 K, which is roughly five times hotter than the surface of the sun. The heating happens rapidly and causes the air to expand. The triggered shockwave travels as acoustic signal and can be heard as thunder. Scientists use a variety of instruments to detect and locate the various signals of lightning. There are, among others, (V)LF ground based networks, e.g., the National Lightning Detection Network (NLDN; Cummins and Murphy, 2009) in the USA and Meteorage (Schulz et al., 2016) in France, and VHF ground-based networks, e.g., Lightning Mapping Arrays (LMAs; e.g., Thomas et al., 2004, Coquillat et al., 2019). Cameras with high frame rates can capture light emitted by lightning (just as the human eye can see lightning). Such cameras are also installed on some satellites to observe lightning from space. The Optical Transient Detector (OTD) on MicroLab-1 satellite and the Lightning Imaging Sensor (LIS) of the Tropical Rainfall Measurement Mission (TRMM) and nowadays on the International Space Station (ISS) proved the idea of lightning observation from space. Recently, geostationary (GEO) satellites were equipped with lightning locating systems (LLSs), such as the Geostationary Lightning Mappers (GLMs; Goodman et al., 2013) on the newest Geostationary Operational Environmental Satellite (GOES) Series over the Americas. The Lightning Imager (LI) of the upcoming Meteosat Third Generation (MTG) satellites (Dobber and Grandell, 2014) will provide GEO lightning observations over Europe, the Mediterranean Sea, Africa, wide parts of the Atlantic, and parts of South America.

The GEO lightning observation sensors on GOES, MTG, and Fengyun-4 (Yang et al., 2017) will provide in real time a near-global coverage for continuous space-based lightning observation as soon as MTG is launched. The optical sensors observe the total lightning activity, including cloud-to-ground (CG) and inter- and intra-cloud (IC) lightning. The large area coverage, especially over the oceans, with relatively high detection efficiency of total lightning (e.g., Murphy and Said, 2020, Marchand et al., 2019) is an important advantage of the GEO technique over conventional ground-based lightning observations. The continuous GEO observation of a region offers a continuous monitoring of the storm activity in contrast to Low Earth Orbit (LEO) satellite sensors which only provide 90 s snapshots of the lightning activity underneath the spaceborne platform. As a consequence, the GEO lightning sensors create a unique dataset

to not only track the thunderstorms in real time but also to document the properties of the atmosphere for numerical weather prediction (NWP). The use of observation data to tune the initial state of NWP models is referred to as assimilation.

This thesis describes the complex process of developing an MTG-LI data assimilation method for the AROME-France model of Météo-France (Seity et al., 2011, Brousseau et al., 2016). The Lightning Data Assimilation (LDA) has the objective to improve the forecast of convective clouds and convective rainfall. In particular, the MTG-LI data assimilation aims at a better representation of existing convection in the initial state of AROME-France as basis for each forecast. Recent LDA studies demonstrated that GEO lightning data can indeed help to improve the forecast of convective clouds, rainfall, and cold pools (e.g., Allen et al., 2016, Fierro et al., 2019, Hu et al., 2020a, Liu et al., 2020). However, the studies reported that assimilating lightning data can cause a wet bias or spurious convection in the model. This thesis develops an assimilation method for MTG-LI data based on a Bayesian retrieval that is capable of both promoting convection where observed and suppressing spurious convection (e.g., Caumont et al., 2010). Pseudo-observations of relative humidity profiles are retrieved and then assimilated as vertical sounding data in the 3-dimensional variational (3DVar) assimilation system of the operational AROME-France.

To date, AROME-France gains information about the location of convection mainly from radar observations that resolve the convective scale (Brousseau et al., 2014). Lightning observation can become another data source to initialize convective clouds in the initial state. That is because the dynamics and microphysics within and near convective clouds are fundamental for cloud electrification as requirement for the occurrence of lightning discharges. For example, laboratory studies showed that the presence of graupel plays an important role for cloud electrification (e.g., Takahashi, 1978, Brooks et al., 1997). Further studies investigated relationships between cloud dynamical and microphysical properties and lightning activity in order to diagnose lightning from NWP proxies. The occurrence of lightning correlates with dynamics such as the updraft strength and volume (e.g., Deierling et al., 2008, Bovalo et al., 2019). As suggested in the laboratory studies and numerical investigations, it was also found that observed or modeled graupel mass (or volume) are well correlated to observed or simulated lightning activity (e.g., Deierling and Petersen, 2008, Mansell, 2014, Allen et al., 2016, Buiat et al., 2017).

The conversion of model prognostic variables to a product comparable to a given observation type is referred to as observation operator. One could also say the observation operator links the model space to the observation space. The representation of MTG-LI data in AROME-France is computed by a lightning observation operator. The aforementioned correlations between lightning activity and model variables are further investigated to derive this operator. As AROME-France does not yet include a diagnostic of lightning, defining the lightning observation operator is one key task of this thesis for the development of the MTG-LI assimilation method.

It was mentioned that MTG-LI will be launched in the near future, thus, the data is currently not available. The development of the assimilation method needs generating MTG-LI pseudo-observations, that are synthetic data that mimic the upcoming MTG-LI lightning data as realistic as possible. It is examined which source of information can be used to achieve the goal of a realistic pseudo MTG-LI dataset. Some former studies, (e.g., Höller and Betz, 2010, Bateman, 2013, Mansell, 2014), provide first ideas to simulate GEO lightning data. Their approaches are revisited and refined in this thesis.

This thesis is structured in three main parts:

I Concepts and the framework for the research conducted during this thesis. The physics of cloud electrification and lightning discharges are explained, the lightning process is

detailed, and lightning locating systems (LLSs) are introduced.

II Generation of pseudo-observations to mimic the MTG-LI. This part includes the understanding and intercomparison of operational LLS observations and the application of those lightning observations in one complex algorithm termed the GEO lightning pseudo-observation generator. The generator is trained by using the records of the operational US LLSs NLDN and GLM to be applied to Meteorage lightning data for France.

III Assimilation of lightning data in the operational limited-area convective-scale NWP model AROME-France. An observation operator based on graupel mass is used to compare AROME-France background and Flash Extent Density (FED) observations. A one-dimensional Bayesian (1DBay) retrieval is used to invert the FED observations into a pseudo-observation of relative humidity that can be assimilated. One case of convective storms near Corsica Island is simulated with different initial conditions to evaluate the effect of the LDA. Therefore, AROME-France forecasts are initialized without radar and lightning data assimilation (control experiment), with only radar data assimilation (current operational setup), with only the LDA, and with a combination of radar and lightning data assimilation.

Figure 1 illustrates the different work phases of the thesis. First, records from the French LLS Meteorage and the SAETTA LMA are compared to LEO lightning observations from the ISS-LIS (Intercomparison in Figure 1; Chapter II.1). This intercomparison allows for understanding different lightning data, for developing methodologies, and validating data analysis tools. ISS-LIS observations are then compared to the US network NLDN (red arrows in Figure 1; Chapter II.2) enabling an indirect intercomparison between Meteorage in France and NLDN in the US with ISS-LIS as common reference. The similarity of Meteorage and NLDN observations is a crucial assumption since records from both networks are used for our GEO lightning pseudo-observation generator. The new operational GLM data became available in 2018 during the thesis. GLM observations are analyzed and compared to the well-known NLDN observations (Chapter II.3). With the operational LLSs GLM and NLDN, a GEO lightning pseudo-observation generator is trained (Training arrow in Figure 1; Chapter II.4). Its application for Meteorage records yields pseudo MTG-LI data (Simulation arrow in Figure 1), that can in turn be assimilated in AROME-France (Chapter III.4). The preparation of tools necessary to assimilate MTG-LI data, marked in red as LDA in Figure 1, constitutes the main objective of this thesis.

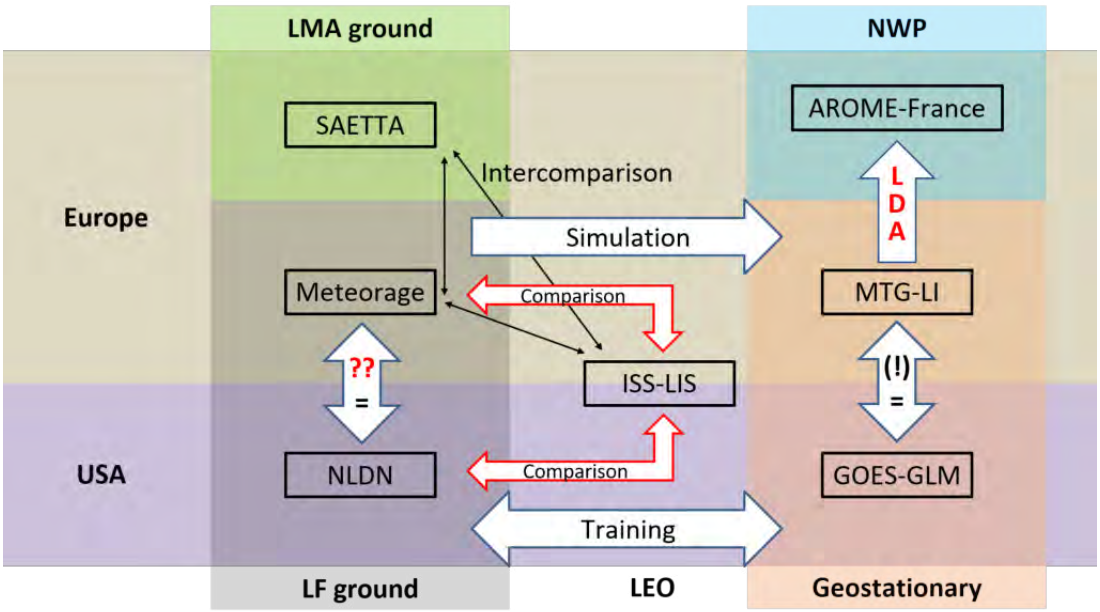


Figure 1: Flow chart of work phases during this thesis, with LLSs used in Europe (olive) and the USA (purple). The LLS types are marked light green (LMA), gray (LF network), white (LEO), and light red (GEO). NWP with AROME-France in light blue.

Introduction Générale

Les éclairs sont des phénomènes fascinants et puissants. Ils ont été observés par les humains depuis des milliers d'années. Historiquement, les gens ne pouvaient comprendre de tels phénomènes et associaient foudre et tonnerre à des forces et des dieux supérieurs, par exemple Zeus et Thor. L'électricité a été découverte au 18^{e} siècle. Depuis, les recherches sur l'électricité atmosphérique ont suscité un intérêt grandissant avec entre autres les célèbres expériences de Benjamin Franklin (1752), qui a démontré que les nuages pouvaient être électrisés à partir de l'utilisation d'un cerf-volant et d'une clef métallique.

L'éclair est un phénomène très énergétique. Il émet non seulement un signal optique, mais aussi un rayonnement électromagnétique sur un large spectre, de la très basse fréquence (VLF) d'environ 1 kHz à la très haute fréquence (VHF) d'environ 100 MHz (Nag et al., 2015). De plus, les décharges chauffent l'air le long des canaux électriques à des températures allant jusqu'à 30,000 K, soit une température cinq fois plus grande que celle de la surface du soleil. Le chauffage se produit rapidement et provoque l'expansion de l'air. L'onde de choc créée se déplace comme un signal acoustique et peut être entendue, c'est le tonnerre. Les scientifiques utilisent une variété d'instruments pour détecter et localiser les différents signaux des éclairs. Il existe, entre autres, des réseaux VLF-LF terrestres, par exemple le réseau National Lightning Detection Network (NLDN ; Cummins et Murphy, 2009) aux États-Unis ou encore Météorage (Schulz et al., 2016) en France. D'autres réseaux terrestres fonctionnent dans la bande VHF comme par exemple les réseaux de type Lightning Mapping Array (LMAs ; e.g., Thomas et al., 2004, Coquillat et al., 2019). Les caméras avec des fréquences d'images élevées peuvent aussi capturer la lumière émise par les éclairs, tout comme l'œil humain peut voir la foudre. De telles caméras sont également installées sur certains satellites pour observer l'activité électrique depuis l'espace comme cela a été démontré par le détecteur optique Optical Transient Detector (OTD) de la mission MicroLab-1 ou encore l'imageur Lightning Imaging Sensor (LIS) de la mission Tropical Rainfall Measuring Mission (TRMM) et sur la station spatiale internationale (ISS). Désormais les satellites géostationnaires (GEO) sont équipés de systèmes de localisation des éclairs (LLS), tels que les caméras Geostationary Lightning Mappers (GLMs ; Goodman et al., 2013) installées sur la toute dernière série de satellites Geostationary Operational Environmental Satellite (GOES) couvrant le continent américain. De son côté, le détecteur Lightning Imager (LI) des futurs satellites géostationnaires Meteosat Troisième Génération (MTG) (Dobber et Grandell, 2014) fournira des observations d'éclairs sur l'Europe, la mer Méditerranée, l'Afrique, de vastes régions de l'Atlantique et certaines régions de l'Amérique du Sud. En incluant le détecteur d'éclairs chinois de la mission géostationnaire Fengyun-4 (Yang et al., 2017), tous ces instruments fourniront en temps réel une couverture quasi mondiale pour l'observation continue des éclairs depuis l'espace, et ce dès le lancement de MTG-LI (mi-2022). Ces caméras optiques observent l'activité électrique totale, c'est-à-dire les éclairs nuage-sol (CG) et les éclairs inter et intra-nuage (IC). La couverture étendue, en particulier sur les océans, avec une efficacité de détection relativement élevée de l'activité électrique totale (e.g., Murphy et Said, 2020, Marchand et al., 2019) est un avantage important par rapport aux observations terrestres classiques. De

plus l'observation continue d'une région depuis l'orbite géostationnaire offre une surveillance continue de l'activité électrique par rapport aux instantanés de l'activité électrique (d'une durée de l'ordre 90 secondes) que mesurent les détecteurs en orbite terrestre basse au sein de leur fauchée. En conséquence, les détecteurs d'éclairs en orbite géostationnaires offrent des observations uniques pour non seulement suivre les orages en temps réel mais aussi pour documenter les propriétés de l'atmosphère en particulier pour la prévision numérique du temps (PNT). L'utilisation d'observations pour déterminer l'état initial des modèles de prévision numérique du temps est appelée assimilation.

Cette thèse décrit le procédé relativement complexe qu'il a fallu développer pour préparer l'assimilation des données MTG-LI dans le modèle AROME-France de Météo-France (Seity et al., 2011, Brousseau et al., 2016). L'assimilation des données d'éclairs a pour objectif d'améliorer la prévision des nuages convectifs et des précipitations convectives. En particulier, l'assimilation des données MTG-LI vise à une meilleure représentation de la convection existante dans l'état initial de chaque prévision du modèle AROME-France. De récentes études sur l'assimilation de données d'éclairs LDA ont démontré que les données spatiales d'éclairs peuvent effectivement aider à améliorer la prévision des systèmes convectifs, des précipitations et des poches d'air froid (e.g., Allen et al., 2016, Fierro et al., 2019, Hu et al., 2020a, Liu et al., 2020). Cependant, il a été signalé que l'assimilation de données d'éclairs peut provoquer un biais humide ou une convection parasite dans le modèle. Cette thèse développe une méthode d'assimilation de données MTG-LI se fondant sur une restitution bayésienne capable à la fois de favoriser la convection observée et de supprimer la convection parasite. Les pseudo-observations des profils d'humidité relative sont restituées puis assimilées sous forme de données de sondage vertical par le système d'assimilation variationnelle tridimensionnelle (3DVar) du modèle opérationnel AROME-France.

À ce jour, le modèle opérationnel AROME-France exploite des informations sur la localisation de la convection principalement à partir d'observations radar qui résolvent les échelles convectives (Brousseau et al., 2014). L'observation « éclair » peut devenir une autre source de données pour initialiser les nuages convectifs dans les conditions initiales du modèle. En effet, la dynamique et la microphysique à l'intérieur et à proximité des nuages convectifs sont fondamentales pour l'électrisation des orages, elle-même condition requise pour le déclenchement de tout éclair. Par exemple, des études en laboratoire ont montré que la présence de graupel joue un rôle important dans l'électrisation des nuages (e.g., Takahashi, 1978, Brooks et al., 1997). D'autres études ont exploré les relations entre les propriétés dynamiques et microphysiques des nuages et l'activité électrique afin de produire un diagnostic d'« éclair » à partir de variables météorologiques prévues par les modèles. L'occurrence des éclairs est corrélée à la dynamique tout comme la force et le volume du vent ascendant (e.g., Deierling et al., 2008, Bovalo et al., 2019). Comme suggéré par des études de laboratoire ou numériques, il a également été constaté que la masse (ou le volume) de graupels observée ou modélisée était bien corrélée à l'activité électrique mesurée ou simulée (e.g., Deierling et Petersen, 2008, Mansell, 2014, Allen et al., 2016, Buiat et al., 2017).

La conversion des variables pronostiques du modèle en un produit comparable à un type d'observation donnée est assurée par l'opérateur d'observation. On peut également dire que l'opérateur d'observation relie l'espace du modèle à l'espace des observations. La représentation des données MTG-LI dans le modèle opérationnel AROME-France est calculée par un opérateur d'observation d'éclairs. Les corrélations susmentionnées entre l'activité électrique et les variables du modèle font l'objet d'une étude plus approfondie pour dériver cet opérateur. Comme le modèle opérationnel AROME-France n'intègre pas encore de diagnostic de l'activité électrique, la définition de l'opérateur d'observation d'éclairs est donc une tâche clé de cette thèse pour le

développement de la méthode d'assimilation des données du détecteur MTG-LI.

Il a été mentionné que MTG-LI sera lancé dans un proche avenir, les données ne sont donc pas encore disponibles. Le développement de la méthode d'assimilation nécessite de générer des observations synthétiques similaires à celles du détecteur MTG-LI à venir et aussi réalistes que possible. Ainsi il est examiné quelles sources d'information et méthodes peuvent être utilisées pour créer de telles données. Certaines études antérieures (e.g., Höller et Betz, 2010, Bateman, 2013, Mansell, 2014) fournissent des premières idées. Leurs approches sont revues et affinées dans cette thèse.

Cette thèse est structurée en trois parties principales :

I Les concepts et le cadre de la recherche menée au cours de cette thèse. La physique de l'électrisation des nuages d'orage et des décharges électriques naturelles est dans un premier temps expliquée, les processus de formation des éclairs sont ensuite détaillés et différentes techniques de localisation des éclairs sont décrites.

II La génération d'observations synthétiques imitant celles du détecteur MTG-LI. Cette partie comprend la compréhension et l'inter-comparaison des observations issues de différents instruments spatiaux et terrestres d'observation des éclairs. Elle détaille aussi l'utilisation de ces observations pour construire à partir d'un algorithme complexe un générateur d'observations synthétiques de données spatiales d'éclairs. Ce générateur, basé sur des méthodes d'apprentissage automatique entraînées à partir d'observations coïncidentes de NLDN et de GLM, utilise des données du réseau terrestre Météorage pour créer des données synthétiques proches de celles du futur détecteur spatial MTG-LI.

III L'assimilation des données d'éclairs dans le modèle opérationnel de prévision numérique du temps AROME-France à échelle convective et à domaine limité. Un opérateur d'observation basé sur la masse de graupel est utilisé pour comparer l'ébauche d'AROME-France et les observations de densité d'étendue d'éclair. Une restitution bayésienne unidimensionnelle (1DBay) est utilisée pour inverser la densité d'étendue d'éclair en pseudo-observation d'humidité relative qui peut être assimilée. Un cas d'orages près de la Corse est simulé avec différentes conditions initiales pour évaluer l'effet de l'assimilation des données d'éclairs. Les prévisions du modèle opérationnel AROME-France sont ainsi initialisées sans assimilation de données radar et d'éclairs (contrôle), avec seulement l'assimilation des données radar (configuration opérationnelle actuelle), avec uniquement l'assimilation des données d'éclairs selon la méthode développée durant la thèse, et avec une combinaison d'assimilation de données radar et d'éclairs.

La Figure 2 illustre les différentes phases de travail de la thèse. Tout d'abord, les enregistrements des réseaux français terrestres de détection des éclairs Météorage et SAETTA sont comparés aux observations issues de l'instrument spatial ISS-LIS (flèche noir fine « Intercomparison » à la Figure 2 ; Chapitre II.1). Cette inter-comparaison permet non seulement de comprendre les différentes données d'éclairs exploitées durant la thèse mais aussi de développer des outils numériques d'appariement et des méthodes d'analyse de données. Les observations du détecteur ISS-LIS sont ensuite comparées à celles du réseau américain terrestre NLDN (flèche rouge « Comparison » à la Figure 2 ; Chapitre II.2), analyse permettant ainsi une inter-comparaison indirecte entre les réseaux Météorage en France et NLDN aux États-Unis en considérant l'instrument ISS-LIS comme référence commune. La similitude des observations des réseaux Météorage et NLDN est une hypothèse cruciale puisque les enregistrements de ces deux réseaux sont utilisés par notre générateur de données synthétiques d'éclairs. Les données

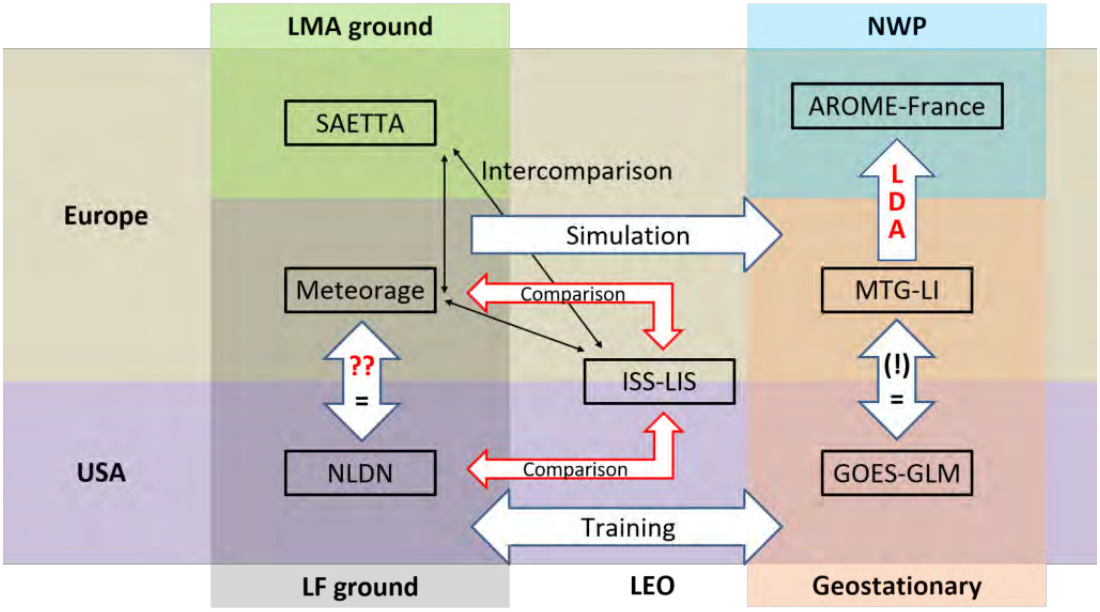


FIGURE 2 : Stratégie développée durant la thèse reliant les observations des différents réseaux de détection d'éclairs terrestres et spatiaux opérationnels ou à venir en Europe (olive) et aux États-Unis (violet). Les types de réseaux de détection sont indiqués en vert clair (réseau terrestre SAETTA), gris (réseau terrestre LF), blanc (capteur optique spatial en orbite défilante), rouge clair (capteur optique spatial en orbite géostationnaire). NWP (pour *Numerical Weather Prediction*, prévision numérique du temps) avec AROME-France en bleu clair.

du détecteur spatial GLM sont devenues disponibles au cours de la thèse en 2018. Les observations du capteur spatial GLM sont donc analysées et comparées aux observations du réseau terrestre NLDN (Chapitre II.3). À partir des observations des instruments GLM et NLDN, un générateur de données synthétiques d'éclairs est construit (flèche « Training » sur la Figure 2 ; Chapitre II.4). L'application de ce générateur aux enregistrements du réseau terrestre Météorage fournit les données synthétiques de type MTG-LI (flèche « Simulation » à la Figure 2), qui peuvent à leur tour être assimilées dans le modèle opérationnel AROME-France (Chapitre III.4). La préparation des outils nécessaires à l'assimilation des données MTG-LI (flèche « LDA » à la Figure 2) constitue l'objectif principal de cette thèse.

Part I

Lightning Physics and Observation

Lightning and thunder have spurred fantasy to seek for an explanation of such phenomena. Early on, they were attributed to gods, for example Zeus in Greek, Thor In Nordic, and Raijin in Japanese mythology. As electricity was discovered in the 18th century, scientists soon suggested that clouds can somehow generate electricity and that lightning and thunder are a product of some kind of discharge processes. Kites, balloons, and insulated metal poles were used in dangerous experiments to confirm that thunderstorms are electrified. Although lightning has been studied for almost 300 years by now, science does still not understand all details of processes involved in the electrification of clouds and lightning discharges.

Contents of this Part

I.1 | The Electrical Nature of Thunderstorms

This chapter describes the current knowledge about cloud electrification and the discharge processes. Basic ideas of cloud formation and hydrometeor types are introduced. Then, the role of hydrometeors for the cloud electrification is emphasized. Important concepts and theories of cloud electrification are reviewed and hypotheses for the transition from electric fields (E-fields) to distinct discharges, i.e., lightning, are explained. It should be mentioned, however, that the microphysical processes behind all of the theories are still subject of current research. Studies agree on the need of hydrometeors. The hydrometeors might collide, exchange and redistribute electrical charges, and cause discontinuities of the conductivity. On the other hand, it is not fully understood how the charge transfer between different hydrometeor types happens, and which process(es) cause(s) the charging of individual particles and forming of charging zones. Another active topic of research is the initiation of lightning in natural E-fields and the lightning propagation through virgin air. The section reflects theories of lightning initiation and propagation, for example, how lightning can propagate over several tens of kilometers within the atmosphere. Finally, conclusions to the different types of lightning are drawn.

I.1.1 A brief introduction to cloud formation

Clouds consist of liquid water droplets and/or frozen ice particles. The vertical temperature profile is one factor responsible for the forming of either droplets, ice particles, or both at different heights. In general, the temperature decreases with altitude within the troposphere and thus along the vertical extent of clouds. However, due to their small size with high surface tension and dissolved particles, e.g., aerosols, droplets persist in temperature regions significantly below freezing. Droplets and ice particles may coexist in a layer of about $-10\,°C$ to $-35\,°C$, with mostly droplets in warmer temperatures below and mostly ice crystals aloft. The formation of clouds is often related to lifting processes that raise water vapor to higher altitudes and colder temperature, saturate an air parcel, and cause condensation and freezing. Clouds can be classified by the nature of the generating lifting processes as (i) convective clouds and (ii) stratiform clouds.

Convective lifting is characterized by relatively local vertical motion, with upward wind speeds in the order of magnitude of a few meters per second. The level of free convection (LFC), where an air parcel ascends freely, can be situated at the surface or can be elevated. An air parcel can reach the LFC either directly through buoyancy (thermal convection) or by dynamical forcing, e.g., an air mass boundary or the topography (forced convection). The temperature and the humidity of the lifted air parcel relative to the ambient air determine whether the vertical motion stops quickly or the parcel can reach high altitudes. The parcel is assumed to rise relatively fast and without the exchange of heat with the environment (adiabatic motion).

It cools according to the dry adiabatic lapse rate until it is saturated with vapor. The level of saturation determines the cumulus condensation level (CCL) and the lifting condensation level (LCL) for thermal and forced convection, respectively, and also defines the visible cloud base. After saturation, latent heat release due to condensation reduces the cooling with height and the parcel ascents moist adiabatically. One parameter that describes the potential of the air parcel to gain vertical acceleration and reach high altitudes is defined by the Convective Available Potential Energy (CAPE). As the name might imply, it gives the potential energy that would be available for the parcel to create kinetic energy based on its temperature and humidity relative to the local atmospheric profile. A positive CAPE indicates potential for the air parcel to rise. However, this potential energy might just become available at a certain height, that is referred to as level of free convection (LFC). The parcel then needs a certain amount of energy to overcome that barrier and reach its LFC. The energy barrier is called convective inhibition (CIN) and can be interpreted as a layer of negative buoyancy. A parcel that reaches the LFC can accelerate depending on the CAPE. In (conditionally) unstable conditions, e.g., thunderstorm environments, vertical motions in the order of $10\,\mathrm{m\,s^{-1}}$ are observed that can even exceed $50\,\mathrm{m\,s^{-1}}$ (Bluestein, 2014, p. 112). The regions of highest upward wind speeds are called updrafts. Their maximum height is again defined by the buoyancy of the air parcel and the storm environment. At a certain altitude, the temperature of the consistently rising and cooling air parcel matches the ambient temperature. This is the equilibrium level (EL) where there is no net vertical force on air parcel. The following Section I.1.2 explicitly elucidates thunderstorms as a form of deep convective clouds.

Synoptical scale (length scale 100 km to 1000 km) stratiform clouds form in regions of slow lifting processes with vertical motion of about 1 % of convection. These lifting processes result from synoptic scale air masses. The slightly less dense, warmer air mass gets lifted at the border to a denser, colder air mass as the warmer air gradual upglides on the dense, cold air at the surface. The air cools adiabatically until it is saturated, condensation (or freezing) takes place, and droplets (or ice particles) form. The resulting stratiform clouds appear as widespread cloud shield that can reach from low levels up to the cirrus level. The slow lifting processes can be seen by increases in temperature (inversion) and dew point with height, in a saturated layer. An inversion defines a very stable vertical profile, and vertical motion can only happen through the forcing of warmer air above the colder air mass. Typical weather phenomena include drizzle and continuous rain with small drop diameters. Stratiform clouds on smaller scales are observed in mesoscale convective systems (MCSs). They are formed by local mechanisms for slow ascent or directly related to the convective clouds of the MCS. Lightning in MCSs may propagate over long distances entering the stratiform clouds.

I.1.2 Thunderstorm electrification life cycle

Lightning and corresponding thunder can be observed all over the globe. Figure I.1.1 shows the average annual flash rate as observed from a low Earth orbit (LEO) satellite [1]. Christian et al. (2003) reports an average of 44 ± 5 lightning flashes (intracloud and cloud-to-ground combined) occurring around the globe every second. Flash rate hotspots exist over North- and South-America, Africa, southeast Asia, and northern Australia (Figure I.1.1). In general, much more flashes are observed over land than over oceans. If a seasonal climatology was shown, the reader would also see that most flashes occur during the summer in extra-tropical regions (e.g., Holle et al., 2016, Zhang et al., 2018, Cecil et al., 2014). The diurnal cycle of flash activity

[1] Latitudes are limited to about 52°N to 52°S

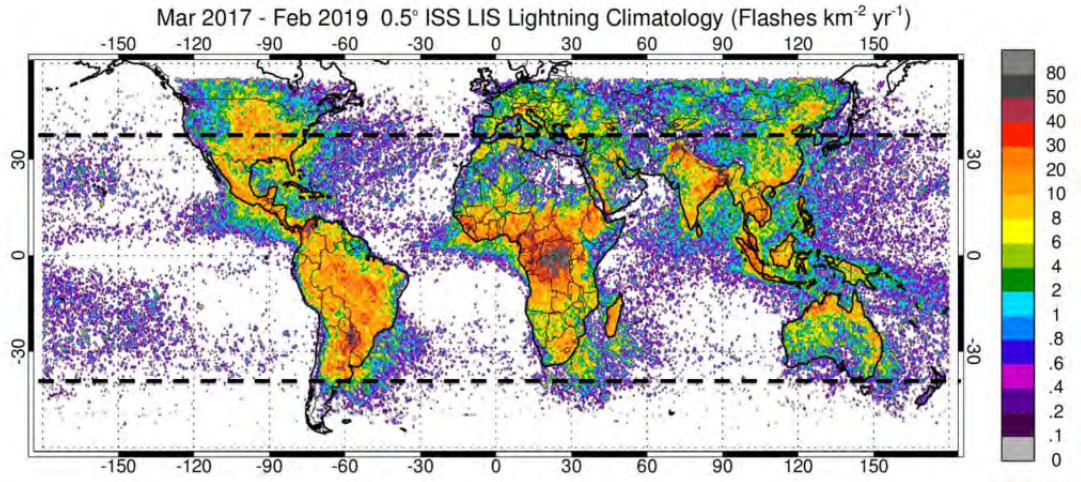


Figure I.1.1: Annual global flash rate as observed with the Lightning Imaging Sensor (LIS) on board the International Space Station (ISS) from March 2017 to February 2019 (adapted from Blakeslee et al., 2019).

peaks at local afternoon over land (see also Albrecht et al., 2016). These facts indicate that insolation plays an important role. The sun can effectively heat the continental surface, and cause more unstable conditions or even thermal convection. Over the ocean, surface heating is less efficient due to heat flow into the ocean and the high heat capacity of water, however, the air is often water vapor saturated (especially near the surface) and sea salt acts as aerosol. The nature of instability and causes of lifting are locally specific, e.g., there is no topography over oceans. Deep convection and cloud hydrometeors can feature different characteristics over land and ocean, such as updraft strength and ice content (e.g., Matsui et al., 2016, Zipser et al., 2006, Williams and Stanfill, 2002).

In (conditionally) unstable environments, air can get lifted and then condense and rise until the EL. The result is towering clouds that might grow up to the tropopause, where the vertical temperature gradient reverses, temperatures increase with height and create the stable stratospheric layer. A cloud may start as Cumulus (Cu), grow as Cu-congestus, and finally build a Cumulonimbus (Cb). Anvils may form if the rising air reaches the tropopause or upper level winds disperse cloud hydrometeors, especially in regions where the vertical wind shear is favorable for thunderstorms[2]. Sometimes, overshooting tops of Cbs are indicative of the strong updrafts within these deep convective clouds. The following sections will explain the microphysical and dynamical processes within updrafts initiating lightning. Before that, the phenomenology of the electrification life cycle of clouds is discussed. Observations revealed that the occurrence of strong natural E-fields and high flash rates often coincides with the location of deep convection. For example, Avila et al. (2010) found Pearson correlation coefficients of 0.9 and 0.6 between deep convection and LEO lightning observations over land between the equator and 35°N (northern hemisphere) and 35°S (southern hemisphere), respectively.

Even the biggest thunderstorms start as shallow convective clouds. Here, the phenomenology of cloud electrification and the life cycle an ordinary cell thunderstorm are described as shown in Figure I.1.2. The updraft begins to lift humid air until water condenses and then also freezes in temperatures well below 0 °C. The presence of both frozen and liquid hydrometeors in the mixed phase of the cloud and the convective forces lead to collisions and redistribution

[2]Strong upper level winds may also cause the updraft to tilt, which leads to a separation of the updraft and downdraft, and to a longer lifetime of the storm.

of different hydrometeor types inside the cloud. Details about hydrometeor types and the electrification of the cloud are provided in the following Sections I.1.3 and I.1.4. Usually, light ice particles with positive electrical charges dominate the upper portion of the cloud, while heavier, negatively charged ice particles remain in the lower portion. During the early thunderstorm development, graupel may also be positively charged, however, graupel in colder environments of intensifying storms is mostly negatively charged. During the phase of development, the updraft still intensifies and the cloud gains height. About 15 min after the first charging processes, the E-field is sufficient to cause breakdown inside the cloud, and an intracloud (IC) flash discharge occurs. The cloud is now, by definition, a thundercloud as it produces lightning and thunder. Precipitation has started as hydrometeors grew and reached regions apart from the updraft. Within the next 15 min the thundercloud grows to its maximum size. The updraft further intensifies and so do the E-field and IC activity. A few minutes after the first IC flashes, that often occurs in the upper portion of the cloud, the E-field is strong enough to trigger a discharge to the ground. Most cloud-to-ground (CG) flashes emerge from the lower negative charge region. The mature stage is characterized by the highest flash rate, with a maximum of IC flashes and significant CG lightning. Strong updrafts may force the growing of heavy graupel and hail due to coalescence and deposition of vapor from supercooled droplets. Heavy rain and hail, that also produce downdrafts, may then be observed at the surface. If precipitation falls into the updraft (in particular with a lack of vertical wind shear), cold air downdrafts and cooling due to evaporation, melting, or sublimation decrease the instability as driving factor of the updraft. The decay of the thunderstorm begins. However, the maximum CG activity is usually observed at the beginning of the decay phase. With a weakening updraft, fewer hydrometeors are suspended. Heavy rain and possibly some small size hail are the consequence. Downdrafts and microburst coincide with the heavy precipitation. The lightning activity decreases with time, until just few IC flashes remain. After about one hour from the first electrification, the storm dissipates. Gentle rainfall, some last IC flashes, and weak downdrafts are the last phenomena produced by the storm.

Thunderstorms like the ordinary single-cell storm occur in environments of low to moderate CAPE and weak vertical wind shear. Downdrafts and microbursts can, however, reach the surface and flow out laterally as a gust front and density current. In low CIN environments and whenever there are cold pools and vertical wind shear favorable to support lifting (Bluestein, 2014), gust fronts can trigger new storm cells. Thunderstorms can apparently stay at the same location for several hours, while actually new cells are triggered by existing cells. Storms could also appear to *jump* forward if new cells are formed upstream of the triggering cell. Independent cells with own updrafts, that often form due to a single trigger (e.g., a front, a convergence line, a dry line, or a gust front) are referred to as multicell thunderstorms. New cells may be triggered more rapidly as existing cells decay, and cause an extending storm system. As there is usually a preferred direction of new cells to form relative to the motion of the triggering cell and vertical wind shear, linear storm patterns can develop. Squall lines are large, linearly organized systems with strong cells (even supercells) embedded. The superposition of gust fronts enforces an additional propagation velocity along the gust front, where new cells are continuously triggered. The motion of the storms may deviate from the mean wind. For example, multi-storm systems with new cell development along the right flank propagate to the right of the mean wind. This is often observed as the region of greatest CAPE also tends to be to the right of the mean wind. The linear organization can also be altered by the ambient wind field, especially if the lines reach a length so that different wind conditions prevail on both ends.

Multicell storms that grow in size to mesoscale, i.e., a contiguous area of precipitation that

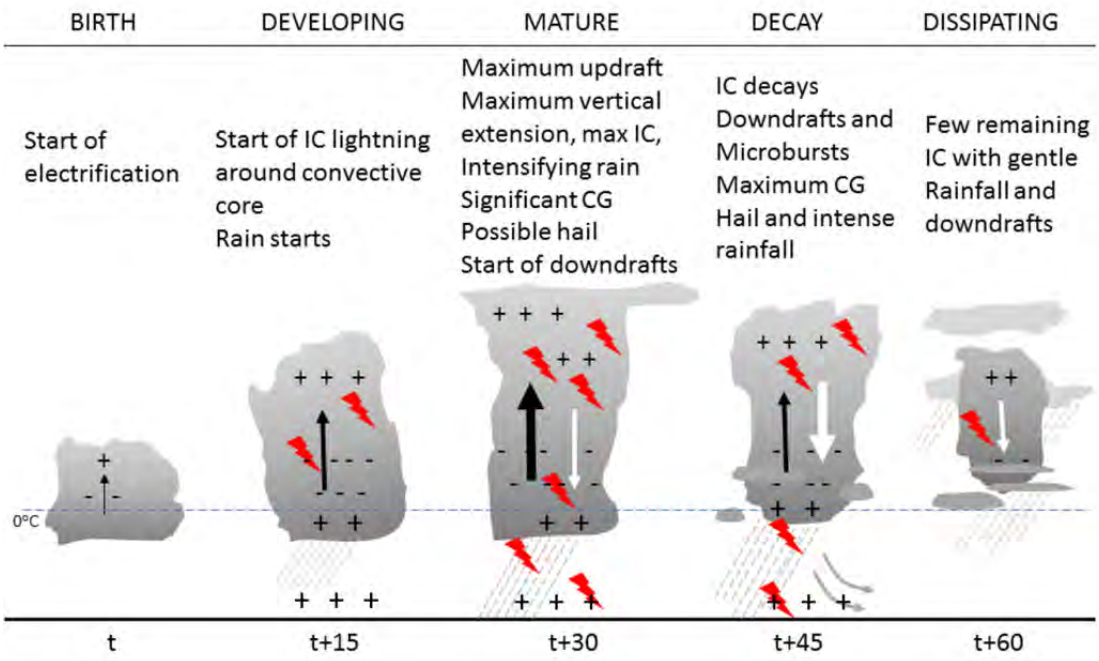


Figure I.1.2: Thunderstorm life cycle of a typical single cell storm. Charge (+,-), updraft (black arrow), lightning activity (red), and downdraft (white arrow) are indicated symbolically. The time axis on the bottom defines the time from the start of electrification (t) in minutes (Colin Price, personal communication through Eric Defer, 2019).

stretches at least 100 km along at least one dimension (Bluestein, 2014), are referred to as mesoscale convective systems (MCSs). MCSs include usually both isolated thunderstorms or storm complexes and squall lines. In addition, stratiform cloud regions surround or trail convective cells in MCSs. MCSs can have various appearances depending on the environmental conditions and the formation process. Bluestein (2014) states that most MCSs begin as squall lines that show four major ways of formation: (i) "broken line" formation from a line of discrete cells, (ii) "back-building" formation where a single cell triggers other cells upstream relative to the storm motion, (iii) "broken areal" formation with separated convective cells becoming organized as a line, and (iv) "embedded areal" formation where a convective line appears within an area of stratiform precipitation. Supercells and multicells may also be embedded within MCSs. Due to their size, mesoscale forces such as pressure gradients, fronts, jet streams, and Coriolis affect MCSs. They can exist for a long time period, and move fast or be quasi-stationary. The former brings strong to severe wind gusts, and the latter precipitates continuously over the same region and enhances the risk of (flash) flooding. Long lived and often fast moving MCSs that often bring dangerous straight-line winds are named derechos (e.g., Johns and Hirt, 1987).

In the most unstable conditions (i.e., CAPE $> 1500\,\mathrm{J\,m^{-2}\,kg^{-1}}$) and in a moderate to strong vertical wind shear with a trend to rotating winds with height in the lower troposphere, the most dangerous thunderstorms can develop: supercells. They are characterized as long-lived single cells with rotating updrafts. The rotating updraft creates an individual mesocyclone (or meso-anticyclone) and a relatively stable, long-living storm system. Supercells grow much larger than ordinary single cell storms, and their trajectory may differ from the prevailing wind direction. Such long-lived cells often move more slowly and off to the right of mean wind direction while they maintain their appearance longer than other thunderstorms (Bluestein, 2014). Polarimetric radar observations of supercell storms revealed that precipitation and

especially hail are formed relatively high in the cloud. The particle growth takes time and in strong (i.e., about $50\,\mathrm{m\,s^{-1}}$) updrafts a significant size occurs at higher altitudes. Surrounding the updraft core, larger particles can form at lower altitudes. Schematic models of supercells show a forward flank downdraft (FFD) and a rear flank downdraft (RFD). When cooler, more humid air from the FFD enters the updraft, the condensation level is lowered and a wall cloud forms. A typical feature of supercells in radar images is the hook echo, that is located adjacent to the RFD. RFD gust fronts may also trigger flanking line convection. The most intense precipitation, including large hail, occurs in the downshear direction from the main updraft, and might wrap around the mesocyclone.

In general, the electrification and lightning occurrence depend on the environment and thus are case-specific. Ordinary single cell storms may feature flash rates as high as the ones of some supercells. In addition, the transition between storm types is smooth, and convective systems may exhibit properties of different types during their lifetime. Nevertheless, the different convective systems may influence the lightning activity. For example, extended MCSs which might also include stratiform regions may allow flashes to propagate over long distances within the system itself. Dotzek and Price (2009) conclude that „Severe weather appears to be characterized by anomalous lightning activity, whether through lightning frequencies, lightning polarity, multiplicity, peak currents, or spatial patterns.“ Supercells often exhibit very high flash rates (more than 40 flashes per minute), and the majority of flashes are relatively small and of type IC (e.g., Zhang et al., 2017).

I.1.3 Hydrometeors

Water exists in its three states within the atmosphere. As vapor, water is an invisible gas that can, however, be felt and sensed as humidity of the air. Every cloud first contains small liquid water droplets, that may persist, fuse to raindrops, or freeze in cold temperatures (and under lower pressure) to form ice crystals. Phenomena consisting of liquid and frozen water particles within the atmosphere are referred to as hydrometeors. According to the World Meteorological Organization (WMO) (2017), hydrometeors may be suspended or fall in the atmosphere, raised from the surface by wind, or are deposited on objects. In addition, phenomena like tornadoes and spouts may suspend even heavy hydrometeors in the atmosphere. In this thesis the term hydrometeor means the embedded liquid and solid water particles.

Hydrometeors usually form around small particles and clumps of molecules (aerosols) that serve as cloud condensation nuclei (CCN) and ice nuclei (IN) for water droplets and ice crystals, respectively. That is because small hydrometeors that might form freely have a strong convex surface curvature and high surface tension. The saturation water vapor over such surfaces is relatively high. Hence, small pure water hydrometeors evaporate (resublimate) often directly after their formation. If, however, CCN (IN) are involved in the formation of hydrometeors, the saturation water vapor is lower than over pure water surfaces and the hydrometeors can persist (Mölders and Kramm, 2014, Kraus, 2004, p. 109-117, p. 200). Indeed, CCN and IN contribute to the hydrometeor formation.

Some hydrometeor types and forms are involved in the cloud electrification and charging mechanisms. They are listed in the following (information from Kraus, 2004):

Drops and droplets Liquid, often round shaped liquid water particles. They may be suspended, i.e., cloud droplets and fog, or precipitate. If their temperature is below freezing, they get the attribute of supercooled liquid. Precipitation of liquid water reaches the

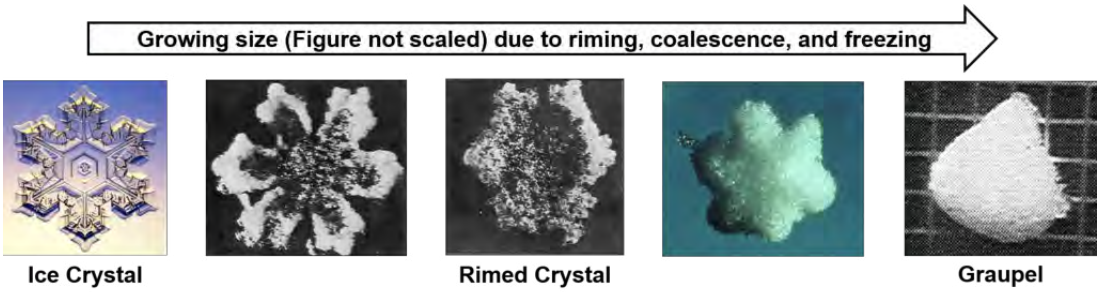


Figure I.1.3: Growth of graupel due to coalescence, riming, and freezing of a small ice crystal (Sylvain Coquillat, personal communication, 2019).

surface as rain (diameter>0.5 mm) or drizzle (diameter<0.5 mm). Rain and drizzle might freeze at the surface (freezing rain) or before they reach the surface (ice pellets).

Ice crystals Particles that either form directly on IN or through freezing of small droplets. Ice crystals are less than 2 mm in diameter. They can either be suspended (as cloud ice particles) or fall as ice prisms and snow.

Graupel and snow pellets Mostly precipitating, conical or rounded ice particles with diameter usually between 2 mm to 5 mm. They are formed by coalescence of small supercooled droplets and ice crystals and may grow by riming at the surface and dissipation of vapor from liquid water to the ice particle. Riming can also cause small graupel in clouds with weak vertical winds (snow grains). Graupel appears with a rough surface and mainly opaque. A transition from graupel to hail is referred to as small hail. It has a rimed core inside a glaciated, smooth layer at the outside, and the diameter can exceed 5 mm.

Hail (Hailstones) Precipitating ice particles with diameters generally exceeding 5.0 mm. They can grow to sizes in the order of 10 cm and are solid, densely packed ice. Usually, an ice nucleus collects (supercooled) drops and droplets. Hail is most common in thunderstorms with significant updrafts and supercooled droplets. Hail surfaces are smooth, while the hailstone can have different forms and appearance. Hail with diameters of at least one inch (2.54 cm) at the surface is considered as a severe weather event.

Figure I.1.3 shows the transition from an ice crystal to a graupel (size not scaled). Ice crystals have typical diameters smaller than one millimeter, whereas graupel diameters reach several millimeters. The graupel is much larger than the ice crystal as graupel is usually an aggregate of smaller hydrometeors due to coalescence of small supercooled droplets, raindrops, and ice crystals. The figure illustrates how the clear, smooth surface of the ice crystal becomes more and more rough as a result of accretion. Meanwhile, the particle becomes opaque as it grows in size.

I.1.4 Natural cloud charging mechanisms

This section introduces three types of charging processes, that are commonly considered when the electrification of clouds is studied. There are many more hypotheses of mechanisms, with detailed microphysical approaches, such as the electrical dipole moment of water molecules or electrical double layers at the surface of droplets, however, it would be beyond the scope of thesis to discuss all processes in detail. The reader is also referred to MacGorman and Rust (1998) and Saunders (2008) and references cited therein for further details.

I.1.4.1 Non-inductive graupel-ice charging

Noninductive charging means all charging processes that do not require an initial E-field and polarization of hydrometeors. The most prominent mechanism describes charge transfer during the collision of larger and heavier graupel (or hail) and smaller, lighter ice crystals. Due to their different size and mass, the forces such as gravity and convection cause opposite motions, with graupel and hail usually falling relative to ice crystals. MacGorman and Rust (1998) state that the graupel-ice charging is the only noninductive mechanism that studies showed is capable of producing enough charge to electrify clouds and initiate lightning. Although the empirical nature of this mechanism is known, the microphysical processes of the direction of charge transfer and the dependence on environmental conditions is not fully understood.

The graupel-ice charging is schematically described in Figure I.1.4. It requires the presence of graupel and ice crystals that gain different terminal vertical motion enforced by the acting forces, e.g., the updraft and gravitation. The general idea states that if the collision of heavy graupel and small ice crystals happens at temperatures above the Charge Reversal Temperature (CRT), the graupel charges positively and ice crystal charges negatively. If, however, graupel and ice crystals collide in environments colder than the CRT, the graupel gains negative charge while the ice crystals transport net positive charge.

The exact mechanism of charge separation and conditions under which graupel charge reversal occurs has been being studied for several decades. One empirical hypothesis, the so-called Relative Diffusional Growth Rate hypothesis, combines several environmental effects (Baker et al., 1987). It states that the growth rate of ice particles due to deposition of water vapor determines whether one particle acquires positive or negative charge after the collision. In particular, the particle that grows faster due to diffusion of vapor at the moment of the collision will charge positively. Model studies (e.g., Dash et al., 2001) and recent laboratory studies support this hypothesis (e.g., Emersic and Saunders, 2020).

Takahashi (1978) were among the first to find dependencies of the graupel charge on temperature and liquid water content inside the cloud. Many studies followed, that are summarized in recent papers of, e.g., Luque et al. (2020) and Jayaratne and Saunders (2016). Takahashi et al. (2017) presents the Figure I.1.5 that demonstrates the variety of studies, and in turn the lack of a consensus. All those studies point out that temperature and liquid water content are likely the most important influencing factors. Brooks et al. (1997) refined the relationships, for example by introducing the effective liquid water content defined as droplets in the path of falling graupel. They pointed out that the rime accretion rate, that increases similarly with higher liquid water content and an increase in velocity, might be the key parameter for graupel charge polarities.

The resulting average charge of graupel after collisions also depends on the size of involved particles (e.g., Avila et al., 1998). Luque et al. (2020) studied in particular the charge reversal temperature for collisions with larger graupel of 1 cm diameter. They found a critical temperature of −15 °C. Rimed collision targets are charged negatively for colder and positively for warmer ambient temperatures, independent of the liquid water content. It is important to note that liquid water was present in all cases, with contents of $0.5\,\mathrm{g\,m^{-3}}$ to $5.0\,\mathrm{g\,m^{-3}}$.

To electrify the cloud, differently charged ice and graupel must separate after the collision in order to transport the charges. Jayaratne and Saunders (2016) found that a separation of colliding ice particles can still occur in the wet growth of particles. The surface of hail was not uniformly wet but had dry spots where colliding particles would not stick but rebound. Nevertheless, the charge transfer during the wet growth of graupel is less efficient than during the dry growth. Lighezzolo et al. (2010) measured positive charge on fewer than 1 % of graupel after the collision with ice particles. Their experiments were conducted in temperatures between

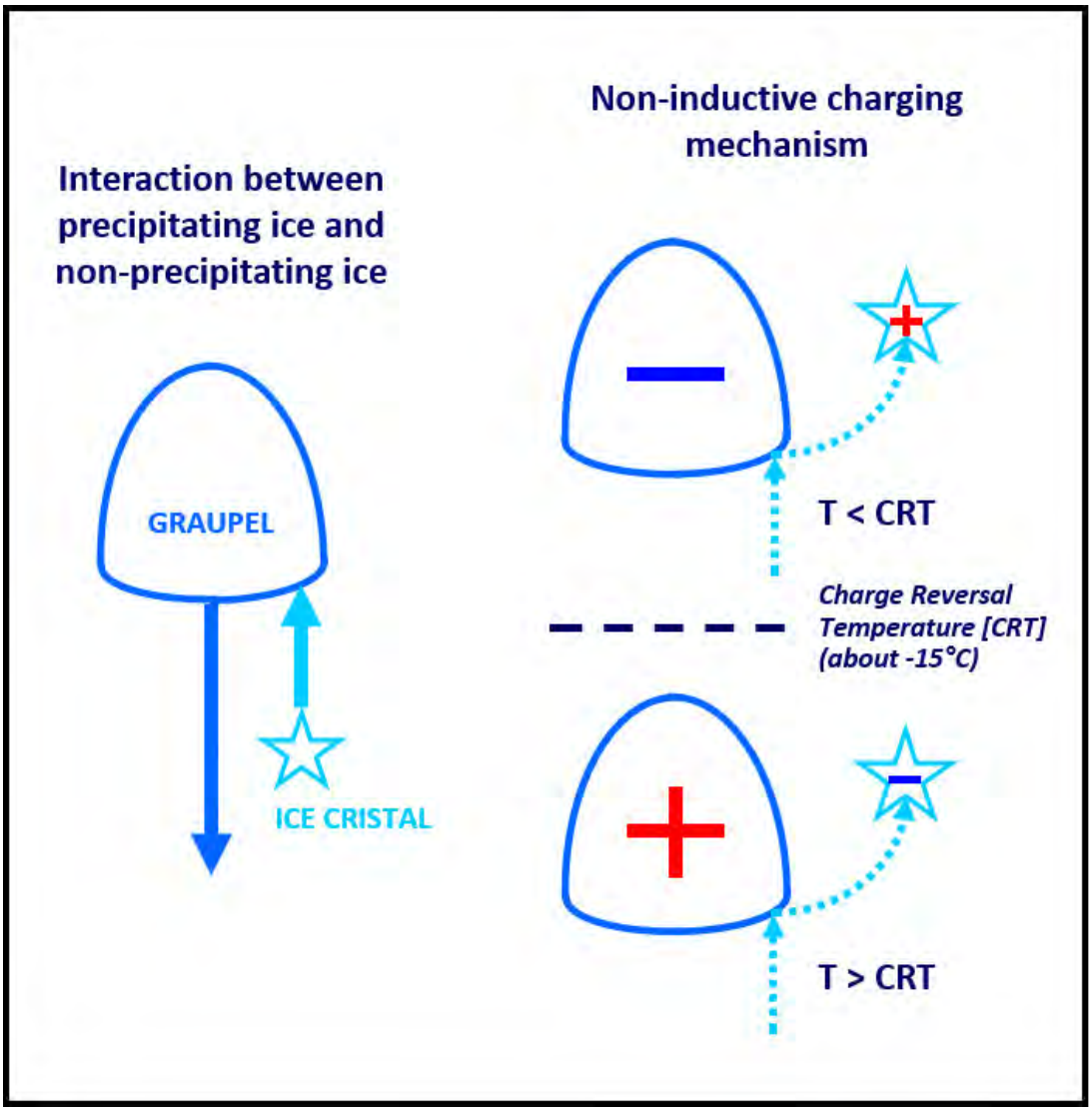


Figure I.1.4: Scheme of the noninductive graupel-ice charging mechanism. The interaction of ice crystals and graupel leads to charging of the hydrometeors. The polarity of the charges depends on the altitude of the collision relative to the Temperature Charge Reversal Temperature (CRT) that is located at about $-15\,°C$ (Eric Defer and Sylvain Coquillat, personal communication, 2019).

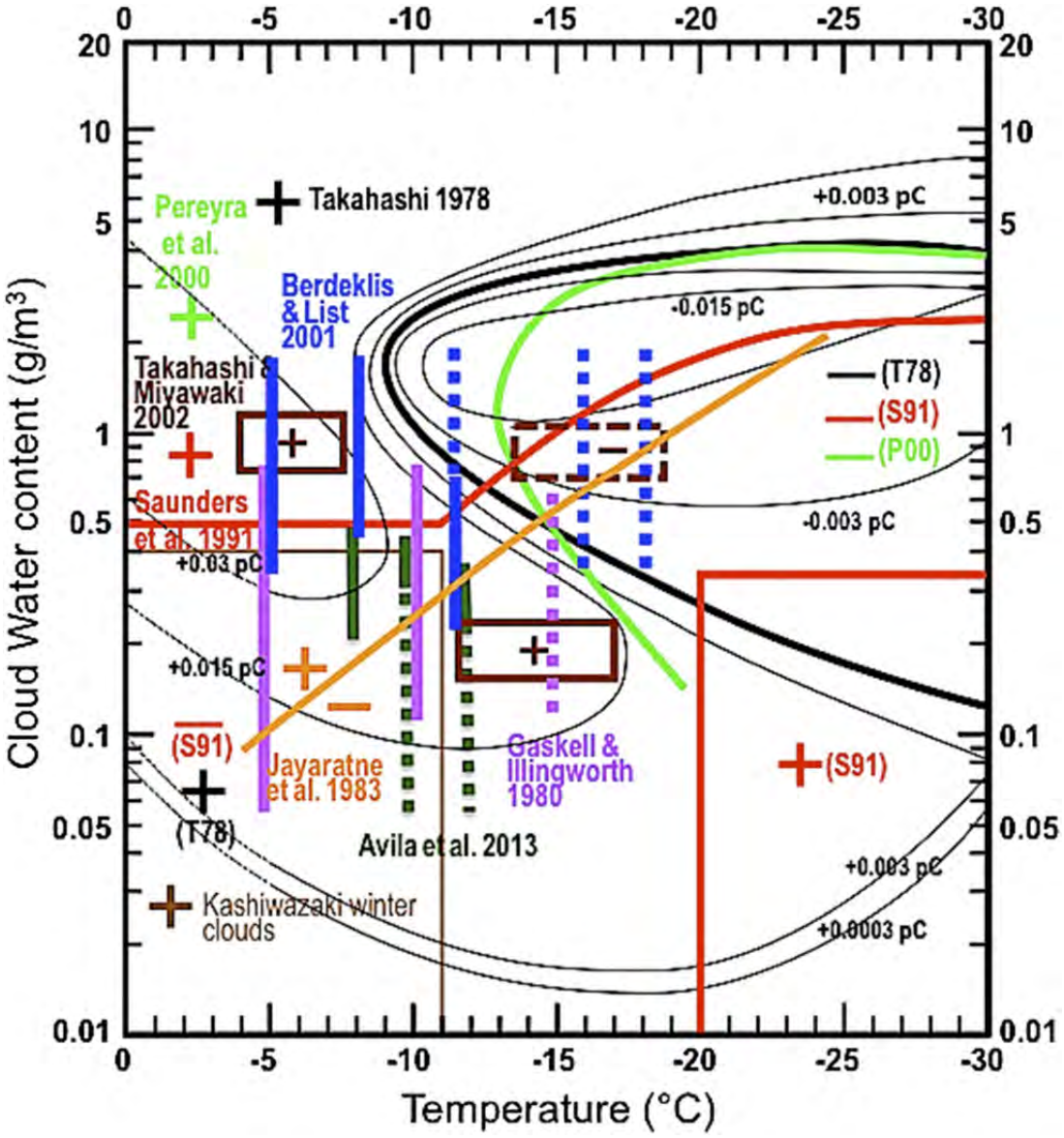


Figure I.1.5: Hypothesis and models for the positive and negative charge regimes of graupel after collision to ice crystals as a function of temperature and cloud water content. From Takahashi et al. (2017) (please see the entire list of citations therein).

−2 °C to −10 °C and with different (effective) liquid water contents, thus likely above the charge reversal temperature and for wet growth of graupel.

There are additional hypotheses on the charge separation during cloud electrification. For example, the chemical properties and composition of aerosols can affect the charge of the rimed hydrometeor of collisions (e.g., Jungwirth et al., 2005). Molecular research found ion defects on the surface of ice. Nelson and Baker (2003) modeled charges at ice-vapor interfaces and found relatively more negative OH^- at sharp edges (i.e., at ice crystals) than rounded surfaces (i.e., of graupel). Devlin (2011) also suggest that the proton density at the surface of larger ice particles is higher than at the surface of smaller ice particles. As a result, larger particles are more likely to transfer positive charge to the smaller particles during the collision.

In conclusion, both graupel and ice crystals can charge negatively and positively during collisions. Under most conditions, collisions at temperatures below (above) a critical charge reversal temperature cause negatively (positively) charged graupel. Typical charge reversal temperatures are −10 °C to −20 °C. The charge transfer at the collision of graupel and ice crystals appears to more efficient if supercooled liquid water is present. In addition, high liquid water contents tend to reduce the charge reversal temperature, and graupel charges positively for high liquid contents. Negative graupel charges occur for colder temperatures and intermediate liquid water content; conditions often present in mid-levels of thunderstorm clouds. Hence, the main negative charge region is formed. The discussed results consider the average graupel charge. Single graupel particles can in fact charge positively or negatively. The ambient temperature and liquid water content also influence the growth regime, i.e., wet or dry growth, of graupel and the velocity of the particle growth. Nowadays, noninductive charging is accepted as the primary charging mechanism in most thunderstorms. More than 25 years ago, Ziegler et al. (1991) already demonstrated the primary role of the noninductive charging in a cloud model and in-situ measurements. Model results of Mansell et al. (2005) support the noninductive charging as primary mechanism especially for initial cloud electrification. However, the noninductive mechanism is believed to act not alone, and it can barely explain tropical thunderstorms with no or low mass of frozen hydrometeors.

I.1.4.2 Inductive charging

Although Mansell et al. (2005) included just one inductive scheme (compared to five noninductive schemes) in their model, they suggest an important, secondary role of inductive cloud electrification. Simulation of four charge separation mechanisms by Takahashi (1979) support this idea. However, inductive charging alone produces E-fields of one order of magnitude lower than what is observed in thunderstorm clouds and needed for lightning to be triggered (e.g., Christian et al., 1980). More recently, Shi et al. (2016) tested the effect of inductive charging mechanisms in a 2D cloud model with electrification and lightning processes. The noninductive charging process alone was sufficient to produce a charge separation into an upper main positive and lower main negative charge region. The inductive processes, however, led to the classical tripolar cloud structure (Section I.1.5) including a lower positive charge layer, where common negative CG flashes occurred. In addition, they suggest that the most important inductive charging involves graupel-ice crystal interactions.

Inductive charging means all mechanisms that require an existing E-field. All particles, that contain free electrons and/or protons, will be polarized under the forcing of that E-field. Positive and negative charges align according to the E-field and create two oppositely charged poles of each particle. The ambient E-field has an additional effect on space charge and gaseous ions, as their net charge enforces a net force on those particles. The inductive mechanisms can

therefore be classified into two key types: (i) particle-particle interaction, and (ii) ion capture by particles.

Particle-particle inductive charging usually means collisions between falling hydrometeors and cloud droplets. Both particle types are likely to be polarized within an E-field, and they carry initially neutral net charge. Thus, the E-field does not affect their motion. At collisions, the smaller particle usually takes away charge from the lower portion of the falling particle. The particles carry away opposite charge and might enhance the existing E-field or reduce the E-field for breakups of colliding raindrops (e.g., Canosa and List, 1993). The E-field within clouds has usually a major vertical component that aligns with relative motions of different particles under the influence of gravity and convection. As the E-field acts similarly on all particles, the top of light cloud droplets has opposite charge than the bottom of falling particles. Although electrical attraction is assumed to be weak between the colliding particles, particles do not always rebound after the collision. This is especially observed for two liquid particles and ice within relatively warm ambient temperature. The separation of colliding particles is the first necessary condition for this mechanism to be effective (MacGorman and Rust, 1998). A second condition requires the time of interaction between the particle to be sufficient to allow charge transfer. As, for example, collisions between unrimed ice particles happen rapidly, those types of particle-particle interactions are also inefficient. As a consequence, inductive charging is considered an important mechanism for collisions between frozen hydrometeors and cloud droplets, as well as collisions involving rimed particles (i.e., graupel).

The ion capture mechanism includes a neutrally charged, polarized hydrometeor and a charged ion. The falling hydrometeor attracts ions of opposite charge than its bottom, and repels ions of the same charge. Hence, the hydrometeor collects net charge of the polarity of the attracted ions (selective ion capture). This process is most efficient if the ions motion driven by the E-field and updraft is slower than the fall speed of the hydrometeors. Otherwise, ions can be captured at both the bottom and the top of the hydrometeor. The captured ions are then of opposite charge and reduce or even nullify the charge gain of the hydrometeor. The process also limits the maximum charge one hydrometeor can gain as the capture of ions reduces the attraction to ions of the same charge. Fair-weather ion densities are in general too low to explain thunderstorm electrification solely by ion captures. It can have a significant distribution to the electrification if there is a surplus of one ion type, e.g., positive ions from corona discharges beneath the thunderstorm. Another mechanism considers the discontinuity of conductivity and the cloud boundaries where cloud hydrometeors capture ions from the always present fair-weather current. It is discussed in the next section.

Melting can cause another inductive charging mechanism, as proposed by Simpson and Walker (1910). Falling, frozen hydrometeors can melt and shed droplets. The separating particles become polarized and exchange charge as describe for the particle-particle mechanism. Since droplets emerge from the top, the melting mechanism is usually not expected to enhance charge regions in thunderstorms. It can be important for stratiform cloud charge near the melting layer (e.g., Stolzenburg and Marshall, 2008, MacGorman and Rust, 1998).

I.1.4.3 Convective charging

Point discharges (corona) at the ground create positive fair-weather space charge (MacGorman and Rust, 1998, p. 35). Developing clouds ingest space charge that is carried by updrafts inside the cloud (Grenet, 1947, Vonnegut et al., 1962), where hydrometeors quickly capture the positive charge. The ability to transport charge (conductivity) of clear air is higher than inside the clouds. Therefore, there is a discontinuity in conductivity at the boundaries of clouds,

that causes ions to be captured. The net positive charge attracts negative ions at the cloud boundaries and a negative screening layer develops. The negative screening layer charge is then carried downward due to cooling and the convective circulation with upward motion in the center and downward motions of the cloud boundaries. The resulting region of negative charge in the lower cloud can cause corona at the surface and increases the flux of positive space charge into the cloud and then upward in the updraft. The E-field enhances. The convective charging mechanism is thus independent of falling hydrometeors and collisions.

Although convective charging can contribute to early cloud electrification, the process is unlikely responsible for strong E-fields and lightning: (i) The space charge density is often limited as key source of charge for this mechanism; (ii) The deposition of negative charge by downward motion varies with the convection itself and ambient wind. It is questionable whether the observed main negative charge layer could form under such variations. (iii) With sufficient negative charge transported at the cloud boundaries in order to explain observed charging rates in thunderstorm, the resulting E-field between inner positive region and negative cloud boundaries would be strong enough to initiate lightning. Initial lightning can indeed discharge between screening layer and upper positive charge of a storm (MacGorman and Rust, 1998, p. 73-74). Hence, the efficiency of the mechanism is further limited. For example, the detailed model of Helsdon Jr. et al. (2002) simulated a small and a severe thunderstorm. In both cases, the convective mechanism alone could not explain the E-field strength required to initiate lightning. Only with the noninductive mechanism activated, organized strong E-fields were simulated during the mature and dissipative phases of the storm cases.

The mechanism overall highlights the significance of convective motion and external (to the cloud) currents. Screening layer charge contributes to some part to the thunderstorm charge structure and electrification. It might also be an important mechanism in tropical regions where clouds contain less ice particles and the noninductive charging is less efficient than in extratropical thunderstorms.

I.1.5 Charge structures of thunderclouds

Thunderstorms are formed by complex dynamical and microphysical processes. The cloud electrification is still not completely understood; however, negative and positive charge separate to build an ambient E-field. Both positive and negative charge regions are observed within thunderclouds. Figure I.1.6 shows one concept of the main charge regions within the convective core of a thundercloud. In close proximity to the updraft, the charge regions are most separated, while the vertical charge structure becomes more complicated outside the updraft, and even more outside the convective core (not shown). The provided temperatures indicate that the lower positive charge region is found below the typical CRT, and the main negative charge region is situated in colder environments within the cloud. In the following, further concepts of the thundercloud charge structure are presented.

A classical model describes the thunderstorm charge structures as a tripole. The main negative charge region, comprised of negatively charged graupels, is situated in the lower portion of the cloud. Both supercooled liquid water and frozen hydrometeors are needed to effectively charge graupel negatively (Section I.1.4.1). Therefore, the main negative charge region is found in temperatures of $-10\,°C$ to $-25\,°C$. Depending on the storm type, region, and season, the altitude above ground can vary between 2 km to 8 km above sea level (Dwyer and Uman, 2014). Just above the main negative charge region, there is the main positive charge region that extends to the cloud top, and for example also into anvils of Cbs. Mainly light, positively charged ice crystals are present in this cold portion of the cloud. A lower positive charge region formed

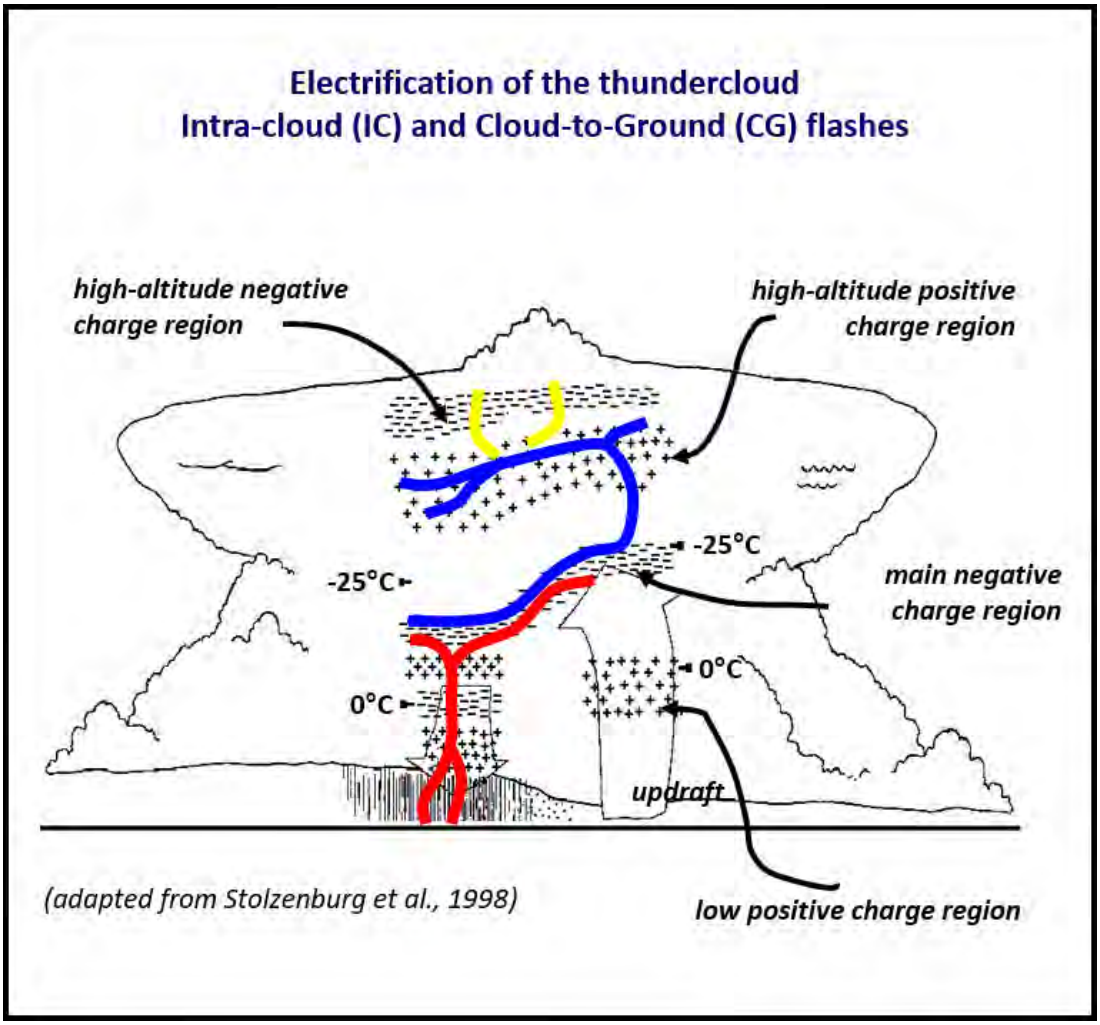


Figure I.1.6: Charge regions within a thundercloud, with a typical negative CG discharge (red), an IC discharge between the main charge regions with negative and positive leaders (blue), and IC discharges between high altitude charge regions (yellow). (Eric Defer and Sylvain Coquillat, personal communication 2019, after Stolzenburg and Marshall, 2008).

by positively charged graupel and some positively charged precipitation, is found between the visible bottom of the cloud and the main negative charge region. The charge density in this region is lower than for the main negative and main positive charge regions, that can be referred to as main charge dipole. Williams (1989) summarized evidence for the tripole structure of thunderstorms.

In situ measurements support this basic tripolar structure in the updraft of convective clouds, however, a fourth negatively charged screening layer at the cloud top is added to the tripole vertical charge structure (e.g., Stolzenburg and Marshall, 2008, 1998). Figure I.1.6 shows the corresponding model of a thundercloud, and a typical vertical profile of the E-field, temperature, relative humidity, and inferred charge regions is shown in Figure I.1.7(a). The resulting 4-layer structure in the updraft of convective systems is described by other authors (e.g., MacGorman and Rust, 1998, Soula, 2012, Dwyer and Uman, 2014). The altitude of the main negative charge region is correlated to the updraft strength and, thus, to the cloud type. For example, supercells with strong updrafts feature relatively high altitude main negative charge regions. The studies also emphasize that the charge structure outside the updrafts is usually more complicated. Anvils often feature positive charge surrounded by negative screening layers. At least six alternating charge layers were reported by Stolzenburg and Marshall (2008) outside the updraft but still within the convective system (Figure I.1.7b). The horizontal extent of charge regions is in general much larger than the vertical extent. Discharge processes, i.e., lightning, can rearrange the charges and thus cause a more complex structure (as indicated in Figure I.1.6 near the CG flash). In addition, the charge structure is not static but evolves during the stages of a thunderstorm, especially during the development and decay. The idealized charge structure with four horizontal charge layers approximates the observed charge structures only during the mature stage with a relatively stable updraft. According to Stolzenburg and Marshall (2008), the strongest E-field is observed in upper cloud parts during the early electrification. Their polarity determines the first lightning that is often IC. The initial charging fits the noninductive mechanism.

Further observational evidence revealed that inverted polarity thunderstorms exist (Rust et al., 2005). The main charge dipole consists of a lower main positive charge region and main upper negative charge region. Fuchs et al. (2015) and Chmielewski et al. (2018) studied conditions under which inverted polarity storms can form. They suggest that high cloud base and high CAPE environments enhance the chance of inverted charging structures. A shallow warm cloud depth and strong updraft would transport more supercooled liquid water in the mixed-phase region of the cloud and allows for positive graupel charge. Hence, a lower main positive charge region can form. Dry air entrainment is also suggested as an influencing factor, as it affects liquid water content, ambient temperature, and instability.

I.1.6 Lightning initiation

The process of lighting initiation remains one of the biggest mysteries in atmospheric science. Recently, Lyu et al. (2019) studied 31 IC flashes using a VHF interferometer and found that the so-called fast positive breakdown (FPB) process is responsible for some of the lightning. However, the majority of the studied flashes did not show the FPB-typical continuous VHF emissions during initiation and, thus, were initiated through unknown processes. Despite several laboratory studies, in-situ measurements, and theories, there is no real evidence how the electric field inside clouds can cause a breakdown. The maximum measured E-fields inside clouds are usually at least one order of magnitude smaller than what is believed to be the critical E-field causing a discharge. Babich et al. (2016) report an external E-field required for stable streamer

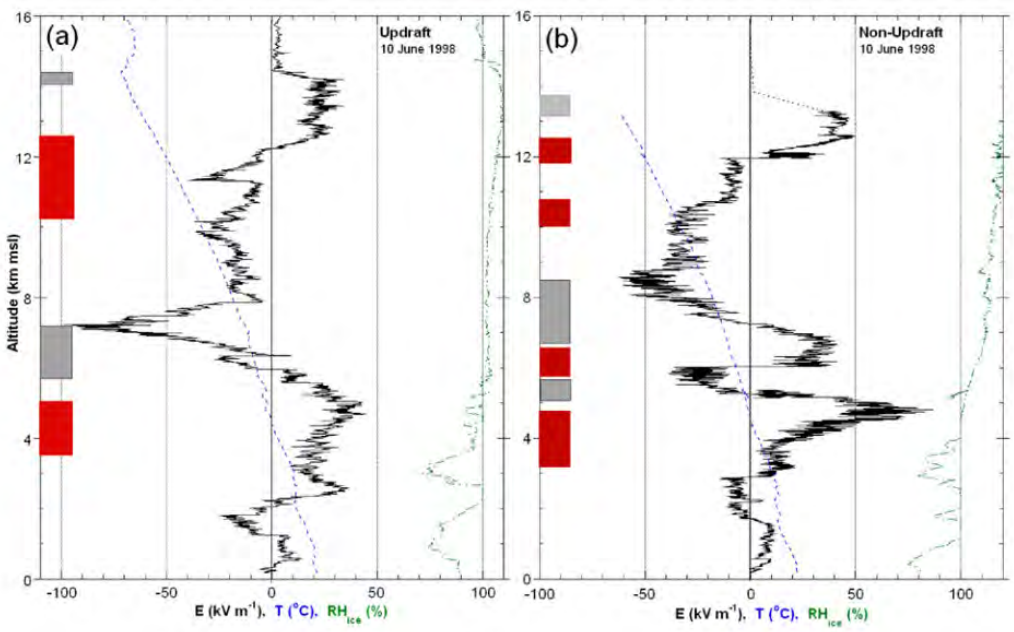


Figure I.1.7: Vertical profiles of electric field (E, black), temperature (T, blue), and relative humidity (RH_{ice}, green) with the inferred charge (red: positive, gray: negative) measured from balloon soundings. Four charge regions are seen in the updraft (a), and seven here outside the updraft (b). From Stolzenburg and Marshall (2008).

development near needle-shape ice crystals (from numerical simulations) must be a factor of 2.0 to 2.7 higher than the maximum observed E-fields. Hypothesis suggest that there are locally enhanced E-fields inside clouds (Zeng et al., 2016). Measurements might just not have captured so far the local maximums. The enhancement might result from modified E-fields near and around hydrometeors with their curved surfaces. The curvature of the surfaces causes a local growth of the ambient E-field that might in turn lead to corona from the hydrometeor surfaces. The corona removes charge from the hydrometeor and enhances the E-field locally. Another hypothesis considers high energy electrons. The so-called runaway electrons can create high conductivity near their paths, and might ionize molecules at collisions. Although such electrons enter Earth's atmosphere continuously from space, there still need to be local mechanisms to cause a relativistic runaway electron avalanche (RREA) and affect the conductivity in clouds. The RREA requires an E-field about one order of magnitude lower than the conventional electric breakdown in natural clouds. Dwyer and Uman (2014, p. 163-172) details processes involved in lightning initiation. It should be emphasized that there is not a clear understanding of how and what are the environmental conditions where lightning initiation occurs inside clouds. Some evidence was found that most lightning origin in the main negative and/or main positive charge region (Dwyer and Uman, 2014, Coleman et al., 2008, Shao and Krehbiel, 1996). CG lightning propagates downward and towards the ground if the lowest charge region is insufficient as barrier. IC lightning might propagate through both main charge regions, with a (single) vertical connection channel. The initial breakdown, often a bipolar pulse, constitutes the first leader propagation and can produce a significant amount of electromagnetic radiation comparable to return strokes.

I.1.7 Lightning propagation

The locally enhanced E-field creates thermally-ionized conducting channels. These hot channels can elongate within the ambient E-field as a (stepped or continuous, depending on the charge type) leader can propagate, while the type of propagation depends on the polarity and environment. If the conductive channel causes an electrical breakdown, the E-field collapses and charge is exchanged. Charge can be transferred inside a cloud, between clouds, between clouds and the clear air, and between clouds and the ground. If a leader connects between the cloud and the ground, a short circuit is created with high potential differences. Current flows rapidly ($>1 \times 10^8\,\mathrm{m\,s^{-1}}$) upward and a bright light bold, the return-stroke, becomes visible. Similar discharges process can occur within the cloud, where the bright signal is referred to as K-change. In the following, leader propagation is discussed in detail for leaders between the cloud and the ground, as those are mainly studied. Leader propagation inside clouds, also referred to as streamer propagation, is similar to that discussed between the cloud and the ground, although the presence of hydrometeors, and lowering pressure and temperature with altitude likely influence the leader speed. Dwyer and Uman (2014) provide an extensive literature review and most information is extracted from this publication.

Lightning propagation may begin with an initial breakdown defined as a sequence of channels extending randomly from the source, and one of which becomes the stepped leader to the ground. The stepped leader as a unipolar conducing channel can be of negative or positive nature, although positive leaders usually propagate smoothly. The propagation of those differently charged leaders is detailed in the following. Typical negative leader speed ranges from initially $4 \times 10^5\,\mathrm{m\,s^{-1}}$ to $18 \times 10^5\,\mathrm{m\,s^{-1}}$ and then decreases, however, leader processes such as dart leaders and K- and M-changes can reach speeds of over $1 \times 10^7\,\mathrm{m\,s^{-1}}$ (e.g., Shao et al., 1995, Shao and Krehbiel, 1996, Zhang et al., 2008). Single segments of the negative stepped leader last 20 µs to 100 µs. Downward moving negative stepped leaders are the common type, however, also positive downward leaders, and upward leaders of both polarities can draw the path through virgin air. A special type is referred to as dart-stepped leaders. Its upper portion propagates along return stroke channels and a stepped leader is formed in the lower portion. Often leaders elongate bidirectionally, with different polarity of both ends and with different characteristics of both branches (e.g., Shao and Krehbiel, 1996), depending on the environment and the ambient E-field.

Downward negative leaders, as responsible for most CG flashes, are extensively studied. Their reported average speed approximates $2 \times 10^5\,\mathrm{m\,s^{-1}}$ near the ground and up to $1 \times 10^6\,\mathrm{m\,s^{-1}}$ for their upper portion. Although they propagate 3-dimensionally (3D), the highest speed component is by far along the vertical direction. Observed leader durations of some tens of milliseconds imply primary channel lengths of a few kilometers. The channel is, however, not continuous but formed in steps. This is due to the breakdown that is required ahead of the leader tip to propagate. Typical step lengths range from 5 m to 100 m, with about 10 µs to 100 µs interstep intervals. The evidence was found because leader steps produce weak pulses prior to return strokes that can be measured. A single step lasts about 1 µs. The propagation speed of the electrical well within each step is in excess of $5 \times 10^7\,\mathrm{m\,s^{-1}}$. Creating single leader steps involves luminous so-called space stem in corona near (i.e., some meters) the actual leader tip. Space stems are bipolar, isolated plasma elements. They get heated by streamer currents and propagate bidirectional as space leaders. The new leader step is created as space leaders connect to the leader tip (Zeng et al., 2016, Dwyer and Uman, 2014, p. 177). Recently, similar X-ray emissions were measured for different types of negative leaders indicating common physics (Hill et al., 2011, Howard et al., 2008).

Positive leaders connect less frequently to the ground, are less luminous, and produce less electromagnetic radiation (sensed by instruments) than negative leaders. Therefore, few detailed studies investigated positive leaders. It is known that they can step or move in more continuous or pulsing (intermediate form) fashion. In contrast to negative leaders, free electrons are attracted by the tip of positive leaders and produce ionization. The corona and leader propagation require less leader tip voltage and lower E-field magnitude than for negative leaders. One could say that the positive leader leaves a fainter electrical impression than a negative leader. Despite the difference, (clear air) leader speeds, step lengths and interstep intervals (if stepped propagation) are similar to negative leaders (e.g., Saba et al., 2008). The downward positive leader speed increases towards the ground, whereas upward positive leader speed increases towards the cloud. The upward positive leader might propagate an order of magnitude slower than downward leaders, with slightly shorter step lengths (4 m to 40 m) and similar interstep intervals (20 µs to 120 µs). Whether positive leaders propagate as stepped or continuous leaders depends on the quantity of available free electrons and apparently also the humidity. If there is steady corona triggered by electrons ahead of the leader tip, the leader moves more continuously.

I.1.8 Phases of a flash and lightning types

The previous sections described the processes leading to lightning more or less at the microscopic scale. Propagating from an initial source, there are four types of leaders that also define four different types of cloud-to-ground (CG) flashes; (i) negative downward, (ii) negative upward, (iii) positive downward, and (iv) positive upward. Anthropogenic constructions such as tall structures, airplanes, and rockets can influence and even trigger CG flashes.

Negative downward leaders produce negative CG flashes. As illustrated in Figure I.1.8(1), they often also have a positive leader that propagates as IC component. This is the most common type of CG flashes. The stepped leader builds several branches seeking for the channel of lowest resistance to connect the cloud and the ground. Figure I.1.9 shows one example of a negative CG flash that lasted about 16.5 ms from the first sign of the stepped leader until the return stroke. As mentioned, negative stepped leaders propagate over several kilometers, despite the fact that there must be a significant E-field to allow for a breakdown ahead of the leader tip. The steps become fainter and propagate slower as the leader approaches the ground. Streamers can emerge from sharp points to facilitate the connection of the leader with the ground. For downward negative leaders, upward positive streamers may attach to the leader tip. Once connected to the ground, the narrow, ionized conducting channel is activated. Negative charge is lowered from the cloud to the ground by an upward return-stroke (Figure I.1.8(2)). The return-stroke moves continuously fast in the opposite direction of the charge transfer. Even after this first, effective discharge, the conducting channel remains active and there is often IC activity (Figure I.1.8(3)). Subsequent dart leaders, that move continuously and about two orders of magnitude faster than the stepped leader, can use the existing ionized channel to transfer further charge from the cloud to the ground (Figure I.1.8(4)). Hence, dart leaders cause subsequent return-strokes in multi-stroke flashes (Figure I.1.8(5)). The entire process is what usually is called a flash. It takes only some milliseconds from the initiation to the return-stroke (see for example Figure I.1.9), and the phases proceed too fast for being captured by the human eye without technical support, e.g., high speed cameras. Even if the phases are of very short duration, a flash can last up to several seconds if leaders merge, separate, or travel long distances. Recently, Stolzenburg et al. (2020) studied negative CG flashes with multiple return strokes that emerged from so-called subsequent stepped leaders (SSLs). The SSLs may

reach the ground directly or intercept a channel of a previous stepped leader.

Negative CG flashes are also produced by initial upward positive leaders. In consequence of a lack of sharp edges as initiation points of corona discharges and streamers in the cloud, there is usually no streamer coming down from the cloud. Hence, the upward positive leader attaches directly to the cloud. Upward positive leaders start, as the downward negative leader, discharges processes that serve to drain negative charge from the cloud to the ground. In consequence, a subsequent downward negative leader develops and then initiates the upward return stroke. Dart leaders and subsequent return-strokes are also possible after positive CG flashes.

Similar processes with opposite polarity exist for downward positive leaders (as in Figure I.1.10) and upward negative leaders. They transport positive charge from the cloud to the ground and produce positive CG flashes. Positive CG flash usually consist of the positive downward leader and a negative IC leader as shown in Figure I.1.10(1). The return strokes move in opposite direction of the charge transfer upward into the cloud following the ground connection of the positive leader. This flash type is rare for normal polarity thunderstorms, however, the vast majority of CG flashes in inverted polarity storms lower positive charge to the ground.

Cloud flashes are the most frequent flash type. On average, at least two thirds of all lightning remain in the sky, and only one third reaches the ground. Nevertheless, there are less studies on cloud flashes than CG lightning, as the latter directly impacts mankind. Cloud flashes include discharges between the different charge regions inside a cloud or of different clouds, and discharges that start within a cloud and terminate in the clear air. Like for CG flash, an IC flash starts by a bi-leader, i.e., a negative (positive) leader propagating toward the positive (negative) charge region. Figure I.1.11 illustrates the two main components of an IC flash: (1) a bi-leader discharge and (2) the junction discharges composed of K-change processes. Fast recoil streamer discharges are also called K-changes and constitute the brightest cloud flash types. The physics are the same as for dart leaders (e.g., Shao et al., 1995, Zhang et al., 2008), except that the K-changes propagate partially or completely within the channels initially ionized by both negative and positive leaders. K-changes that connect to an existing continuing current conducting channel may light up the channel and are referred to as M-changes.

Air discharges can propagate over long distances at speeds of about $1 \times 10^3\,\mathrm{m\,s^{-1}}$ to $1 \times 10^5\,\mathrm{m\,s^{-1}}$ (MacGorman and Rust, 1998). Hence, they are relatively slow and one can often determine the direction of propagation. This thesis will use the term intracloud (IC) flash to summarize all kinds of flashes that do not connect to the ground if not stated differently.

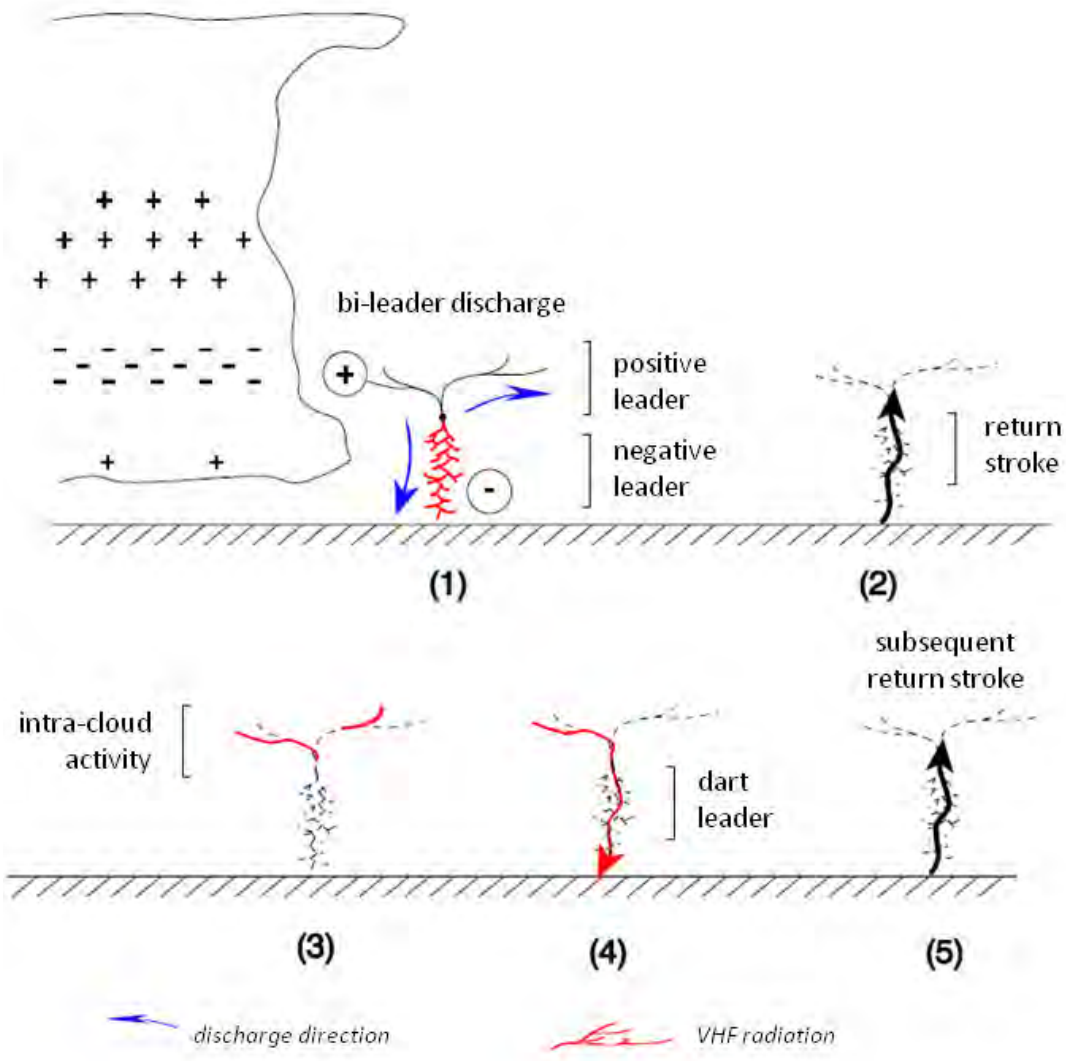


Figure I.1.8: Idealized phases (1-5) of a negative CG flash with a negative leader (red) to the ground. Leader propagation is indicated as blue arrows, return strokes as black arrows. The positive and negative cloud charge is illustrated on the left as "+" and "-", respectively, defining the typical cloud charge layers most important for lightning. The negative components emit strong very high frequency (VHF) radiation. (Eric Defer, personal communication, 2020).

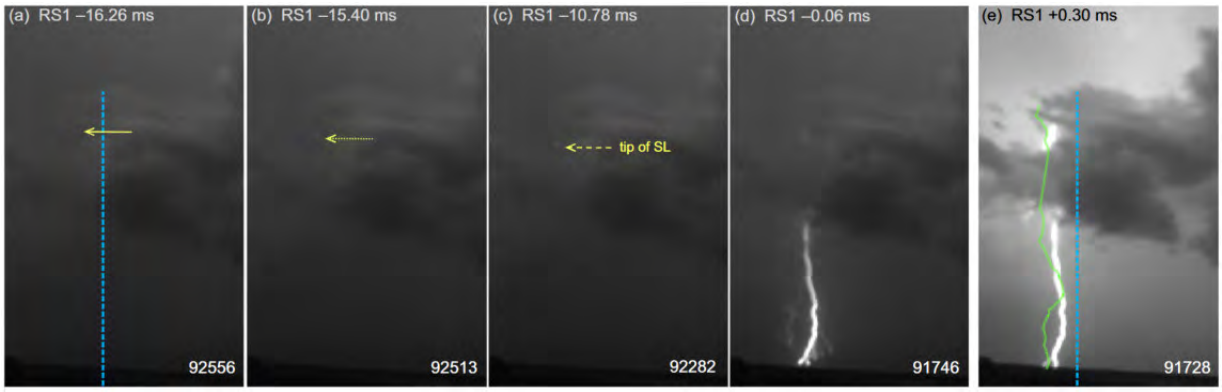


Figure I.1.9: The stepped leader (SL) propagation (a-d) and the first return stroke RS1. Vertical blue line as reference. The time relative to RS1 is given in the top left corner, and the camera frame in the bottom right corner. The green line in (e) indicates a return stroke that occurred 132.86 ms after RS1 from a subsequent stepped leader. Adapted from Fig. 4 of Stolzenburg et al. (2020).

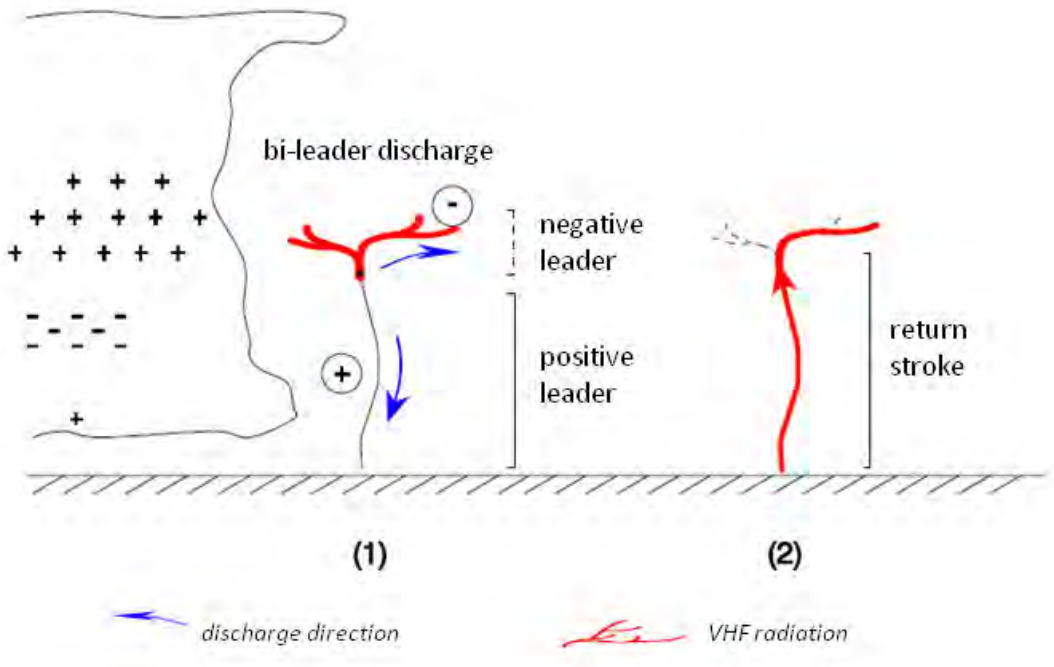


Figure I.1.10: As Figure I.1.8 for a positive CG flash with positive downward leader and negative IC leader. (Eric Defer, personal communication, 2020).

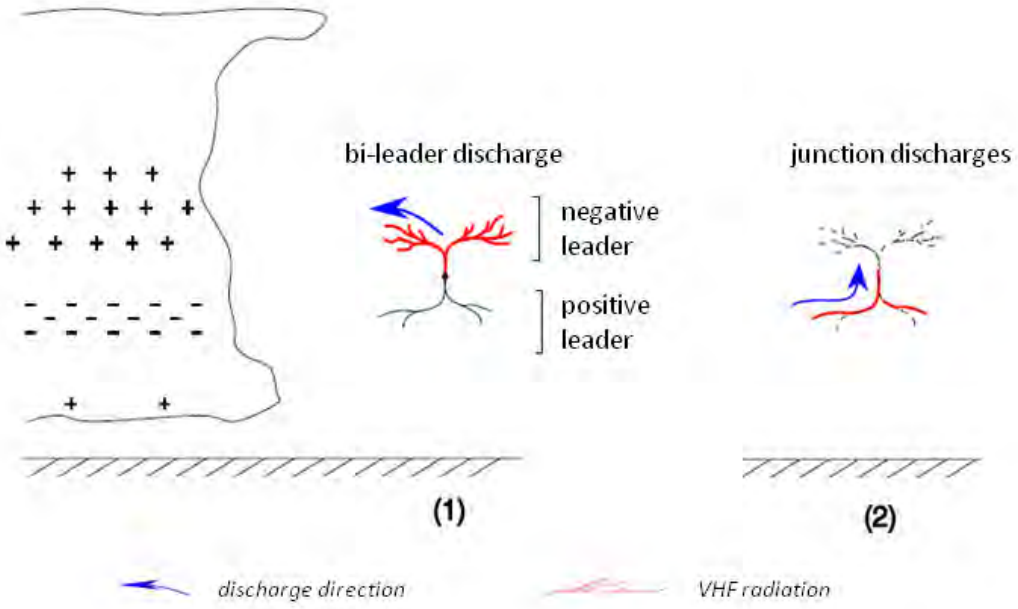


Figure I.1.11: As Figure I.1.8 for a bidirectional IC flash (1) and a junction discharge (2). Leaders of different polarity may propagate within the main charge regions of opposite polarity (Eric Defer, personal communication, 2020).

I.2 | Techniques to Observe and Locate Lightning

Everybody, who has ever seen a flash or heard a thunder, is in fact a lightning observer. The observation of flashes by eye or by hearing thunder is the oldest and easiest way to detect lightning. If the location, i.e., latitude and longitude, and the time are estimated and registered, the method can already provide important insights into occurrence and frequency of flashes in an area. Obviously, researchers use more dedicated techniques and have automatized the process of lightning location. A continuous, accurate locating of all lightning over the entire globe would provide an ideal base for lightning research, however, it is an ambitious goal. Different types of Lightning Locating Systems (LLSs) have been developed to locate different signals emitted by lightning discharges. This includes sensor at Earth's surface, on planes, and in space. Nag et al. (2015) and Cummins and Murphy (2009) present summaries of existing lightning LLSs, with their characteristics, advantages, and disadvantages. MacGorman and Rust (1998) explain background information about electromagnetic radiation in general and from lightning, the transfer of energy released by lightning, and measurements of E-fields, currents, cloud charge, and lightning. In this section, automated techniques are introduced to detect, geolocate, and register lightning. There is a variety of instrument; emphasis is put on recent operational LLSs, and instruments of interest to this thesis. The section starts with a brief discourse on the electric and magnetic field and electromagnetic radiation, and describes different measurement principles thereafter.

I.2.1 The electric and magnetic field

Electromagnetic radiation contains, as the name might imply, both electric and magnetic field components. The electric field vector $\vec{E}$ for a vertically oriented, infinitesimal channel at the point r,ϕ, and $z = 0$ at the ground and at time t is defined as

$$\begin{aligned}\vec{E}\left(r,\phi,0,t\right) = \frac{1}{2\pi\epsilon_0}\Bigg[&\int_{H_B}^{H_T}\frac{2z^2-r^2}{R^5}\cdot\int_0^t i\left(z,\tau-\frac{R}{c}\right)\mathrm{dtdz}\\ &+\int_{H_B}^{H_T}\frac{2z^2-r^2}{cR^4}i\left(z,t-\frac{R}{c}\right)\mathrm{dz}\\ &-\int_{H_B}^{H_T}\frac{r^2}{c^2R^3}\frac{\partial i\left(z,t-\frac{R}{c}\right)}{\partial t}\mathrm{dz}\Bigg]\hat{z}\end{aligned} \tag{I.2.1}$$

where ϵ_0 is the permittivity of vacuum, H_B the height of the leader's lower tip, H_T the leader origin, z the height of the current $i\left(z,t\right)$, R the distance from the source to the point of observation, c is the speed of light in vacuum, and $\tau = t - \frac{R}{c}$ the time required for the radiation

to propagate from the source point to the point of observation. The unit vectors $\hat{r}$, $\hat{\phi}$, and $\hat{z}$ create a right-handed coordinate system with $\hat{r}$ pointing in direction of r (MacGorman and Rust, 1998).

The corresponding magnetic field vector $\vec{B}$ is then defined as

$$\begin{aligned}\vec{B}\left(r,\phi,0,t\right) =& \frac{\mu_0}{2\pi}\left[\int_{H_B}^{H_T}\frac{r}{R^3}i\left(z,t-\frac{R}{c}\right)\mathrm{dz}\right.\\ &\left.+\int_{H_B}^{H_T}\frac{r}{cR^2}\frac{\partial i\left(z,t-\frac{R}{c}\right)}{\partial t}\mathrm{dz}\right]\hat{\phi}\end{aligned} \tag{I.2.2}$$

with the magnetic permeability of vacuum μ_0. It can be noted that $\vec{E}$ and $\vec{B}$ are orthogonal to the channel and to each other.

The expression of the electric field (Equation (I.2.1)) has components from the electrostatic field (first line), induction (second line), and the radiation (third line). The electrostatic field term dominates close to the source and shows the strongest dependency on the distance to the source. The induction term contains the current and is the next most dependent on the distance to the source. Finally, the radiation component that contains the derivative of the current dominates the far field. The magnetic field components (Equation (I.2.2)) are induction (first line) and radiation (second line). The energy distribution of the electric and magnetic fields has wave characteristics, and thus depends on the medium and the wavelength/frequency. The characteristics of the signal propagation influence the measurement techniques, e.g., the range and the discharge processes detected.

I.2.2 Electromagnetic radiation from lightning

Measurements of lightning wave forms allow for monitoring electric field changes induced by lightning. The manifestation of leader propagation and return-strokes appears as peaks in the records. For example, a negative CG flash lowers the E-field magnitude and creates a negative peak measurable E-field change as a result to the transport of negative charges from the cloud to the ground. The E-field and the E-field changes can be measured with two different types of antennas, i.e., slow and fast antennas. Slow antennas with slow decay times reproduce the E-field change from an entire flash, while the fast antennas with short decay times capture fast E-field changes induced by flash components (MacGorman and Rust, 1998). Slow initial breakdown processes inside clouds, that can be measured with the slow antennas, are also referred to as J-changes. K-changes (or recoil streamers) are rapid pulses within the J-changes, that use pre-ionized channels. They are similar to dart leaders that do not reach the ground and produce the strongest discharges within the cloud. Continuing currents are slow E-field changes that can illuminate the channel much longer than return strokes, however, less intensely. They are mainly responsible for lightning caused fires, whereas return strokes have more an explosive effect at the striking point. M-changes are a hybrid form that start as K-changes but reach the ground through a continuing current channel.

Besides the different phases of a lightning flash that emit signals of different currents and duration, a wide frequency range of electromagnetic waves is associated to lightning emissions. Lightning electromagnetic signatures range from a few hertz (long continuous current) to gamma rays. The peak energy from lightning discharges is found at a few to 10 kHz (Nag et al., 2015). Figures I.2.1 and I.2.2 show that lightning locating systems (LLSs) are sensitive to certain phases of a flash according to the frequency bands. The location of lightning

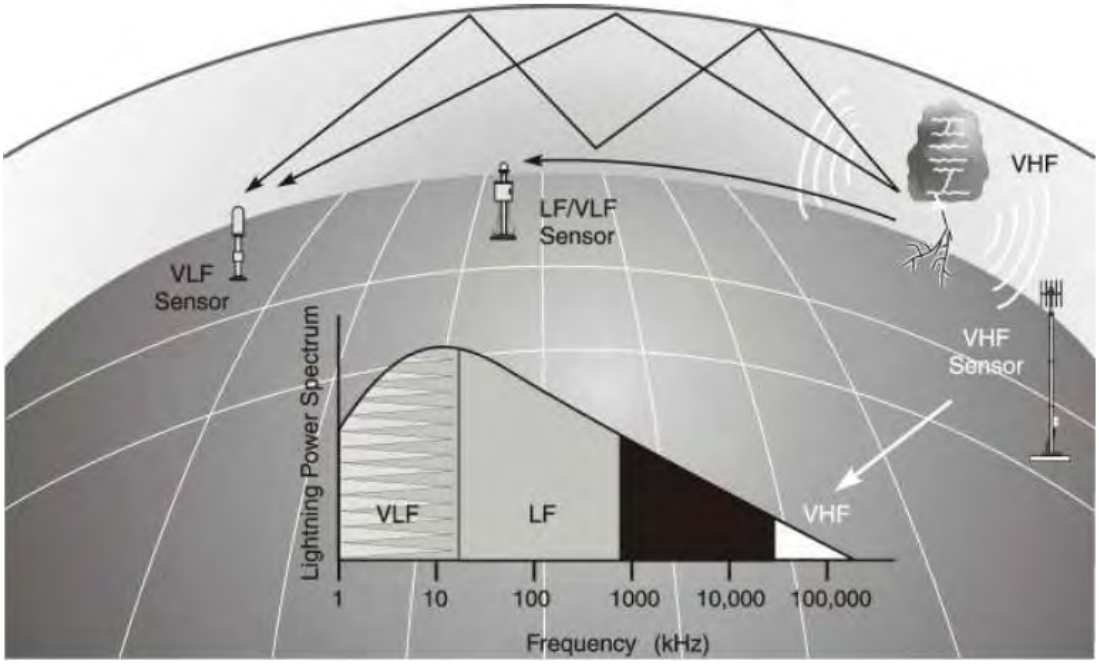


Figure I.2.1: The spectrum of electromagnetic radiation and the LLSs detecting different frequency ranges of radiation emitted by lightning (adapted from Figure 2 of Cummins and Murphy, 2009).

as sources of electromagnetic radiation and the observation techniques are explained in the following sections.

I.2.3 Geolocation of lightning emissions

Lightning flashes exhibit a 3D structure that sometimes can be rather complex according to the distribution of the electric charge regions inside the thundercloud. Some lightning detection techniques offer the possibility to map in details this 3D structure (see Section I.2.5), others are only mapping the flashes in two dimensions (see Section I.2.6) while some others are providing, by detection concept, punctual locations of some flash components (see Section I.2.4). The LLSs, sensing waveforms of the (electromagnetic or acoustic) signal emitted by lightning flashes, use, uniquely or synergistically, direction finding (DF), time of arrival (TOA), and arrival-time difference (ATD) locating methods.

DF instruments consist of orthogonal loop antennas, often oriented north-south and east-west, respectively. Radiation from lightning induces a signal in each loop. The ratio of the amplitude of a given waveform feature, e.g., the peak, at the orthogonal loop can be used to determine the azimuth of the source relative to the antennas. With a third orthogonal antenna, also the elevation can be measured. Triangulation of azimuth (and elevation) vectors from two stations to the source provides the source location. Magnetic DF (MDF) systems detect the vertical section of a lightning channel that produces a magnetic field with an azimuthal component only. DF systems are particularly sensitive to the portion of CG leaders and return strokes close to the ground, as they are mostly vertically oriented channels. Horizontal channels with radial and vertical field components cause errors in the measured channel bearing. The main source of errors, however, are site errors due to reradiation at terrain and man-made structures. As they are constant for a certain direction, they can be corrected at a given site

Table 1. Characteristics of the Lightning Processes Commonly Detected and Geolocated by Different LLSs[a]

Frequency Range	Approximate Wavelength Range	Discharge Process(es) / Lightning Event(s)	Typical Spectral Amplitude in dB (V/m/Hz) Normalized to 100 km[b]	Mode of Signal Propagation	Typical Number of Emissions per Flash (Order of Magnitude)	LLS Type / Comments
VLF (3–30 kHz)	10–100 km	CG RS, CIDs	−76 to −86 (10 kHz) for return strokes	Ground wave and Earth-ionosphere waveguide (affected by dispersion due to finite soil conductivity and characteristics of ionosphere)	1–10	Long range, ground based
		Large-amplitude cloud pulses including PB	—		1–100	
LF and MF (30 kHz to 3 MHz)	0.1–10 km	CG RS	−92 to −110 (100 kHz)	Ground wave and Earth-ionosphere waveguide (affected by dispersion due to finite soil conductivity and characteristics of ionosphere)	1–10	Medium range or short range, ground based, depending on sensor sensitivity and baselines
		Cloud pulses including PB, CIDs, and K-changes	−105 to −125 (100 kHz)		1–100	
		CG leader steps	−116 to −126 (300 kHz)		10	
HF (3–30 MHz)	10–100 m	Various in-cloud and leader processes	−136 to −146 (3 MHz)	"Mixed" ground wave and line of sight (affected by blockage due to presence of objects in line of sight)	10–100	Short range, ground based
VHF (30–300 MHz)	1–10 m	Breakdown of "virgin air" during channel formation	−160 to −190 (rough estimate for frequency ranges from 20–400 MHz)	Line of sight (affected by blockage due to presence of objects in line of sight)	10–1000	Short range, ground based, using TOA
		Dart leaders and K-changes, CIDs			10–100	Short range, ground based using ITF
Near IR / optical (30–300 THz)	10^{-7} to 10^{-6} m	Hot current-carrying channels	NA (different unit of measurement)	Line of sight (affected by scattering in clouds)	10–100 groups per flash	Satellite based

[a]CG RS = cloud-to-ground return stroke; CID = compact intracloud discharge; PB = preliminary breakdown; TOA = time of arrival; ITF = interferometry.
[b]Spectral amplitude data gleaned from *Serhan et al.* [1980], *Weidman et al.* [1981], *Horner* [1964], *Oetzel and Pierce* [1969], *Kosarev et al.* [1970], and *Willett et al.* [1989a, 1989b] and then normalized to a distance of 100 km assuming a 1/distance dependence.

Figure I.2.2: Characteristics of different LLS (adapted from Table 1 of Nag et al., 2015).

(MacGorman and Rust, 1998, p. 160). Location errors in general increase with the distance between the antennas and the lightning. Redundant information from additional stations can reduce the location errors, especially in direction of the baseline. Low frequency networks (Section I.2.4) and interferometers (Section I.2.5) may use DF geolocation methods.

TOA systems consist typically of seven to twelve stations. Information from four stations is needed to reconstruct the location of the signal source in 3D. The occurrence of a specific feature of the lightning emitted waveform at two stations is used to determine the time delay of arrival of this feature at these stations. The time delay between a pair of stations defines a hyperbola of constant time difference. The intersection of at least four hyperbolas leads to unique latitude and longitude location. Five hyperbolas enable to locate the source in 3D. This technique crucially relies on exactly synchronizing all stations of the network. The location error is directly proportional to the time error at each station. Another important aspect concerns the identification of the common feature of the waveform. In situation with very high signal rates, the system might not be able to distinguish information from superpositions of signals. To reduce processing times, TOA systems use the dominant waveform feature in a certain time window, usually in the order of 10 µs to 100 µs. This technique can be applied at all frequency ranges. Adjustment is needed for long range networks due to modification of the electromagnetic waves emitted by lightning during propagation. In addition, the hyperbola gets altered due to the curvature of the Earth. This effect must be considered to identify the correct intersection point.

Some LLSs, e.g., the National Lightning Detection Network (NLDN) and the French Meteorage network, use a combination of DF and TOA geolocation techniques. This is referred to as a hybrid geolocation technique. Such systems need only two sensors to locate the lightning source (as DF), however, the additional TOA information reduces the location error compared to DF-only systems.

The ATDnet network can be seen as a TOA system. It is referred to as a different technique because time differences are not calculated between single features of the waveform, but from cross-correlations of the entire waveform signal at the stations. The UK MetOffice developed this automated system to locate thunderstorms over thousands of kilometers. It uses a low frequency of 2 kHz to 23 kHz and seven stations with baselines of 250 km to 2200 km. As for other long-baseline TOA systems, measurements from four stations are required to locate a source without ambiguity.

I.2.4 Low frequency networks

This section summarizes all instruments that sense frequencies below 30 kHz as low frequency (LF) LLSs. In the literature, there is often a separation of extremely low frequency (ELF), very low frequency (VLF), and low frequency (LF) in the range from 30 Hz to 30 kHz. For simplicity, only the LF definition is used in the thesis as stated above.

Short duration currents in lightning channels and longer continuing currents produce LF radiation. LF sensors usually detect short processes, e.g., return strokes and M-changes (CG), initial breakdown and K-changes (IC). The radiation of the strongest currents related to lightning are within the LF band. Hence, LF systems allow long-range detection of high energy emission that propagates through the Earth-ionosphere wave guide over a few thousand kilometers as they reflect between the ionosphere and the ground (Cummins and Murphy, 2009). In addition, CG flashes trigger ground waves that constitute a propagation of the electromagnetic radiation over the Earth's surface. Ground waves experience stronger attenuation for higher

than for lower frequency components and, thus, LF signals can propagate relatively far via ground wave. LF networks exist on different scales:

Global coverage (land and ocean) systems, e.g., the World Wide Lightning Locating Network (WWLLN, Lay et al., 2004) and the Global Lightning Detection network (GLD360, Said et al., 2010), sense the lowest frequencies. In the detected frequency range, return strokes are responsible for the greatest emissions. In consequence, global LF networks locate CG flashes significantly more reliably than IC flashes, however, even the CG flash detection efficiency (DE) of about 10 % (Nag et al., 2015) remains low compared to LF networks with shorter baseline distances. Global LF networks have long baseline distances of up to few thousands of kilometers.

Country-wide (national) and regional coverage (medium-range) is provided for frequencies near the energy peak of emission from lightning (about 10 kHz). Typical network baseline distances are in the order of 100 km. Such LF networks feature high CG flash DE. About 50 % of IC flashes can be detected and located, and this value increases for lower baseline distances. Examples are the French and US national LF networks Meteorage and NLDN, respectively.

LF systems use geolocation as MDF, TOA, or a combination of the two. They can classify lightning flashes as CG and IC. The locating accuracy depends mainly on the baseline distance. Nag et al. (2015) approximates the accuracy as 1 km to 10 km within the network, and higher errors (lower DE) appear outside the network geometry. The location accuracy can be significantly higher for CG strokes within the network. For example, Cummins and Murphy (2009) give about 250 m median location accuracy for NLDN, and Pédeboy (2015) states up to 120 m location accuracy for Meteorage CG strokes. In contrast, Pédeboy et al. (2018) found about 1.64 km location accuracy for Meteorage IC pulses.

Short baseline LF networks can map single lightning channels and provide altitudes as they detect even relatively weak LF radiation of short duration pulses. As mentioned, the performance depends mainly on the baseline distance. The Huntsville Alabama Marx Meter Array (HAMMA, Bitzer et al., 2013) is a 10 m to 15 m baseline LF network with 7 stations, with about 100 m altitude variation. It measures E-field changes and can be considered hybrid slow-fast antenna due to the sensor time constant of 100 ms. HAMMA is sensitive to the radiation component rather than only electrostatic field changes. Each Marx meter has an instrumental response in the frequency range from 1 Hz to 4×10^5 Hz. The network can continuously sample E-field changes for high flash rate storms. CG flashes are mapped in detail and almost to the ground. Average source location errors are in the range of a few hundred meters in the horizontal and about 700 m in the vertical. The TOA system has about 1 µs temporal errors that directly influence the location accuracy.

Figure I.2.3 compares the observations of the HAMMA system to Lightning Mapping Array (LMA; see following section) records, and NLDN observations. The example shows one hybrid flash on 25 Oct. 2010 near Huntsville, AL. The short baseline LF HAMMA captures much more detail than the national scale LF NLDN, and it provides additional altitude information. The HAMMA source locations are well correlated with the very high frequency (VHF) LMA sources. Whereas the LMA locates the IC negative leader channel in detail, the HAMMA observes the return strokes until the ground.

I.2.5 Very high frequency measurements

This section and the remainder of the thesis consider all frequencies from 3 MHz to 3000 MHz as Very High Frequency (VHF). VHF systems depend on the direct line-of-sight and cover regional areas. Their effective range is limited to about 200 km. VHF radiation from lightning comprises

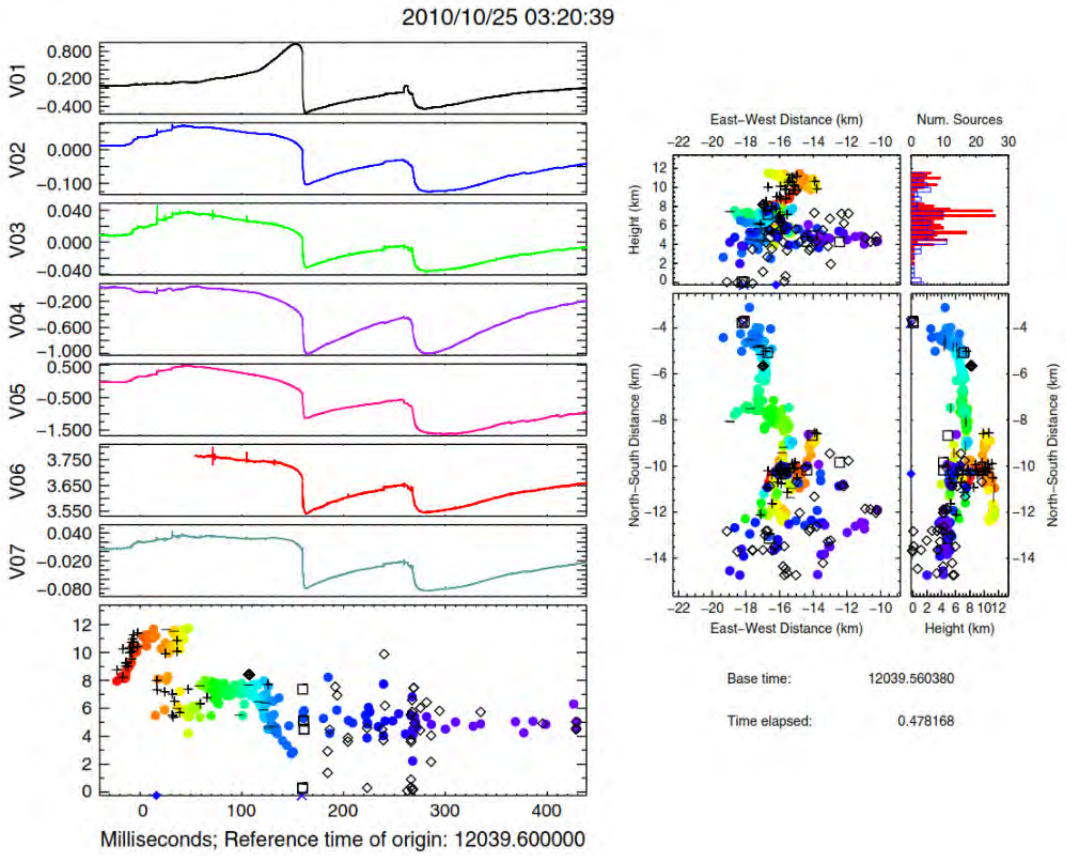


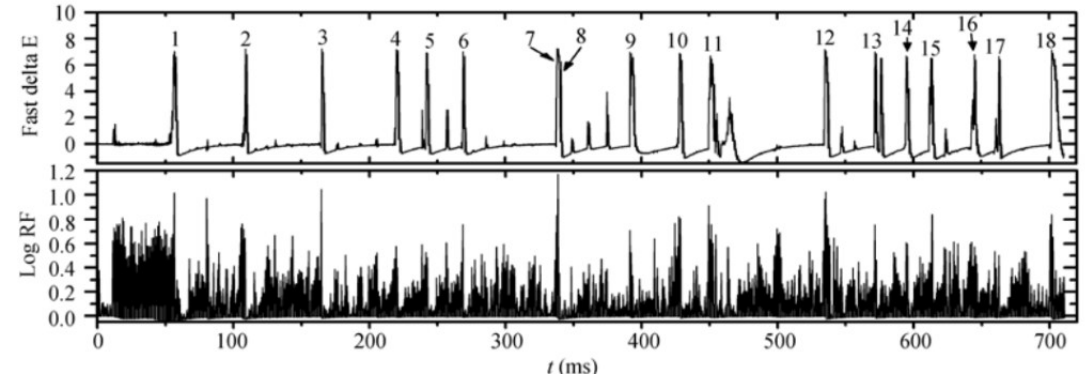
Figure I.2.3: Comparison of HAMMA station voltage measurements (V01-V07) as waveforms, and reconstructed HAMMA sources with positive ("+") and negative ("-") initial polarity, and HAMMA return strokes (black squares) in the bottom plot on the left. Colored dots are North Alabama Lightning Mapping Array (NALMA) sources. Blue "x" shows NLDN CG strokes, blue diamond for one NLDN IC pulse. On the right: Spatial distribution with maps of latitude-longitude, latitude-altitude, longitude-altitude. In addition, the histogram of sources per height is shown. From Bitzer et al. (2013).

much less energy than the emissions in the LF range (Figure I.2.1), and VHF amplitudes feature no direct relation to the peak current of lightning. Most VHF radiation is emitted by channel formation with the breakdown of virgin air and leader processes in existing channels, e.g., dart leaders. The first process produces short, well-defined VHF pulses at the microsecond-scale, while the latter results in quasi-continuous VHF signatures for a few milliseconds without distinct peaks (Nag et al., 2015). VHF systems are equally sensitive to CG and IC lightning processes. They can map lightning channels but do not classify flash types. The dependence on the line-of-sight together with the Earth's curvature result in a lack of low altitude source detection at long range.

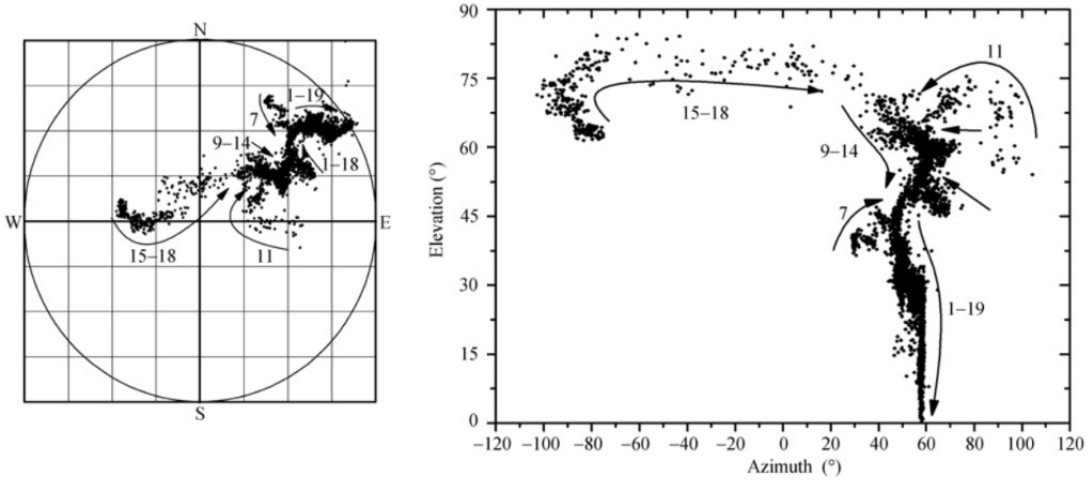
VHF TOA systems are more likely to detect the VHF peaks than the quasi-continuous VHF radiation. Common baselines distances equal some tens of kilometers. For flat networks with only horizontal baselines, the vertical errors are much larger than the horizontal errors. The vertical errors can be reduced if some stations of the network are situated at higher altitude above the sea relative to the remaining stations. Early systems were provided by Proctor (1981, 1983) (at 355 MHz) and the Lightning Detection and Ranging system (LDAR, Poehler and Lennon, 1979) (at 63 MHz). As the TOA VHF systems detect the VHF pulses, there are often several source detections for a lightning channel. This could result in issues due to tedious processing time for the early systems, however, recent computers can usually handle such data. A new LDAR system (Lennon and Maier, 1991, Maier et al., 1995) was developed that measures with 6 MHz bandwidth from 60 MHz to 66 MHz or 222 MHz to 228 MHz. It analyzes the VHF peak amplitude in 80 µs time windows. At least 4 antennas must detect the same VHF source to provide a location. Researchers at the New Mexico Tech refined the VHF TOA system and developed the Lightning Mapping Array (LMA) (e.g., Thomas et al., 2000). The LMA technique will be presented in detail in Section II.1.2. Modern operational VHF TOA LLSs locate lightning in 3D with an accuracy of some tens of meters in the interior of the network. The location errors increase with distance outside the network geometry. Flash DE approaches 100 % but also decreases outside the network geometry. It is reminded here that continuous positive leaders produce less TOA VHF amplitude than negative stepped-leaders. Therefore, VHF TOA systems might indicate the charge structure within clouds. Due to their sensitivity to fast VHF pulses, sources in negative leaders result in stronger VHF TOA signals than positive leader propagation.

The described VHF TOA technique was tested with short network baselines of about 10 m (MacGorman and Rust, 1998, p. 154). The technique showed relatively high uncertainty (approximately 0.5° in both azimuth and elevation) and a lack of confidence for the location of some VHF sources. Combinations of short-baseline systems could be used to triangulate source locations and to increase the location accuracy.

A second type of VHF LLSs uses DF based on interferometry to geolocate VHF sources. One sensor of a VHF interferometer consists of an array of (at least two) closely spaced antennas (distances much smaller than the distance to the lightning). The phase difference between signals induced by a VHF source at different antennas is a function of direction angle of the VHF source relative to the inter-sensor baseline between the antennas. Two orthogonal inter-sensor baselines are needed for azimuth and elevation angles. The antennas must be placed in a way to avoid phase ambiguity (referred to as fringes, MacGorman and Rust, 1998), especially for narrowband interferometers that are sensitive to a particular frequency range. In general, short inter-sensor baselines reduce the number of fringes at the cost of increasing angular errors. Broadband interferometers may lower the risk of phase ambiguity. Shao et al. (1995) and Shao and Krehbiel (1996) investigated E-field changes and discharges related to CG and IC flashes, respectively. They located lightning using an interferometer, and could



(a)



(b)

Figure I.2.4: Records of one CG flash of (a) waveform time series of E-field changes recorded by a fast antenna (top) and the logarithmic radio frequency (Log RF) radiation measured by the VHF receiver, and (b) the plane projection view (left) and azimuth-elevation plot of reconstructed interferometer VHF sources. Adapted from Zhang et al. (2008).

study lightning discharges and processes. For example, details about the nature of dart-leaders, attempted leaders, K-, and M-changes are provided. They found that IC flashes often have a bilevel structure connected by one single vertical channel. The interferometer also allowed for analyzing propagation speeds of different discharge processes.

An example of fast antenna measurements and narrowband interferometry is provided in Figure I.2.4(a) and (b), respectively. Zhang et al. (2008) used a combination of short and long baseline phase values to deal with the fringe ambiguities. The two panels show the same flash, and the leader processes are numbered continuously in the E-field changes and projection views. The example shows a CG-flash that connects to the ground at about 57° azimuth (Figure I.2.4b right). All 19 leader processes went down to the ground along the same channel. Figure I.2.4(a) indicates additional E-field changes that can be attributed to parts of K- and M-changes. Some are studied by Zhang et al. (2008) and the reader is referred to this paper for details. E-field changes are best identified in the fast antenna records, and can be located using the narrowband interferometer.

The VHF source can be located in 2D by a single sensor. 3D locations of VHF sources are reconstructed via triangulation with at least two or more interferometer sensors that are

separated by up to 150 km for reliable source altitudes. VHF interferometers integrate typically over 100 µs of radiation to find phase changes. Hence, they detect preferably continuous leader processes (Nag et al., 2015).

The French SAFIR system is a VHF interferometer that can detect both CG and IC. It measures within a band of 1 MHz width selected between 110 MHz to 118 MHz. Each station operates an array (two arrays for 3D location) of antennas to determine the direction to a source via interferometry. A central processor triangulates signals from at least two stations to locate the VHF source. Typically, 3 stations are separated by 10 km to 100 km. VHF DF was also tested on LEO satellites, however, there is no system currently operational. One example of a former interferometer in space is the VHF broadband digital interferometer as part of the electromagnetic payload of the Global Lightning and sprIte MeasurementS (GLIMS) mission at the Exposed Facility of Japanese Experiment Module (JEM-EF) of the ISS (Morimoto et al., 2016).

I.2.6 Visible light detection – ground and space

Visible light emitted from lightning can be detected by cameras. At the ground, high frequency cameras have been utilized to not only record CG flashes and striking points, but also single steps of CG lightning leaders. Camera records are valuable resource for validating other LLSs as they see all CG lightning in their field of view when exiting the cloud. However, cameras are less reliable for IC lightning, both location and detection, as the light is scattered, reflected, and refracted by the cloud hydrometeors, or the cloud is even completely opaque for light.

Optical sensors are operated on LEO and geostationary (GEO) satellites to map lightning from space. These instruments monitor cloud top illuminations. LEO satellites capture only snapshots for about 100 s of one region and storm. GEO satellites continuously monitor storms in their field of view (FOV) from development to dissipation of a storm.

The heating and ionization of molecules produce visible radiation at distinct atomic lines. The strongest emissions are found at 777.4 nm (oxygen line) and 886.3 nm (nitrogen line) in the near IR. Operational satellite sensors, e.g., the International Space Station's (ISS) Lightning Imaging Sensor (LIS), the Geostationary Lightning Mappers (GLMs) on the Geostationary Observational Environmental Satellites (GOES 16/17), and the Lightning Mapping Imager (LMI) on the Fengyun-4 (FY-4) satellite, and planned instruments, e.g., the MTG-LI, measure at 777.4 nm (dissociation of molecular to atomic oxygen by heating) with a narrowband filter of few nanometers. Visible light emitted from lightning produces illuminated pixels within the camera image. One pixel defines the smallest measurable unit (an event). The location accuracy is subject to sensor specification, spatial resolution, and altitude performances of the satellite, and ranges typically from 4.5 km to 16 km. It might increase due to satellite pointing accuracy issues and parallax effects. Satellite LLSs map the spatial extent of flashes (with less spatial resolution than VHF instruments).

Figure I.2.5 shows typical images of lightning as seen from space. The lightning observations (bright, yellow) are superimposed on the Advanced Baseline Imager (ABI, Schmit et al., 2005, 2008) visible image with landmass, ocean, and clouds. The lightning sources illuminate the cloud tops and thus can be located by the GLM camera sensor. One situation on 14 Apr. 2018 was further analyzed by Peterson et al. (2020a). Without going into detail at this point, Figure I.2.6 shows Flash Extent Density (FED) observed for 1-hour activity of a matured MCS. The colored pixels indicate the number of flashes per minute and approximate the GLM image pixels. The figure distinguishes four flash types that are described in Peterson et al. (2020a).

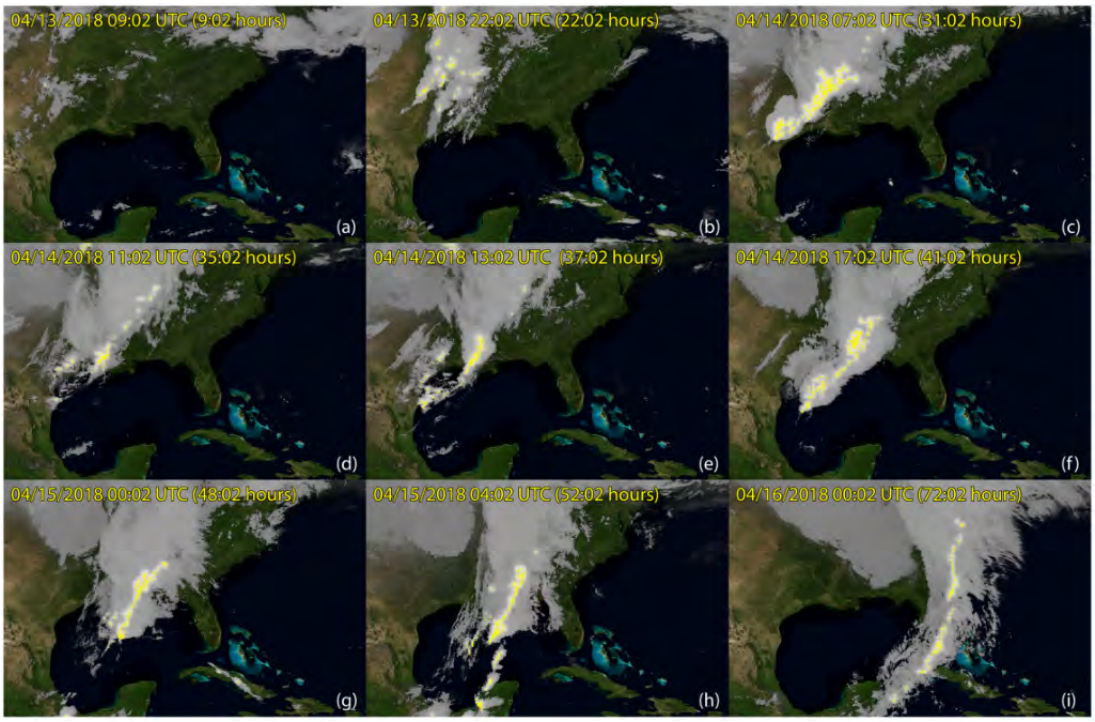


Figure I.2.5: Images of the GLM and the Advanced Baseline Imager (ABI) at (a) 09:02 UTC on 13 April 2018, (b) 22:02 UTC on 13 April 2018, (c) 07:02 UTC on 14 April 2018, (d) 11:02 UTC on 14 April 2018, (e) 13:02 UTC on 14 April 2018, (f) 17:02 UTC on 14 April 2018, (g) 00:02 UTC on 15 April 2018, (h) 04:02 UTC on 15 April 2018, and (i) 00:02 UTC on 16 April 2018. The number of hours since 13 April 2018 0:00 UTC are also specified. From Peterson et al. (2020a).

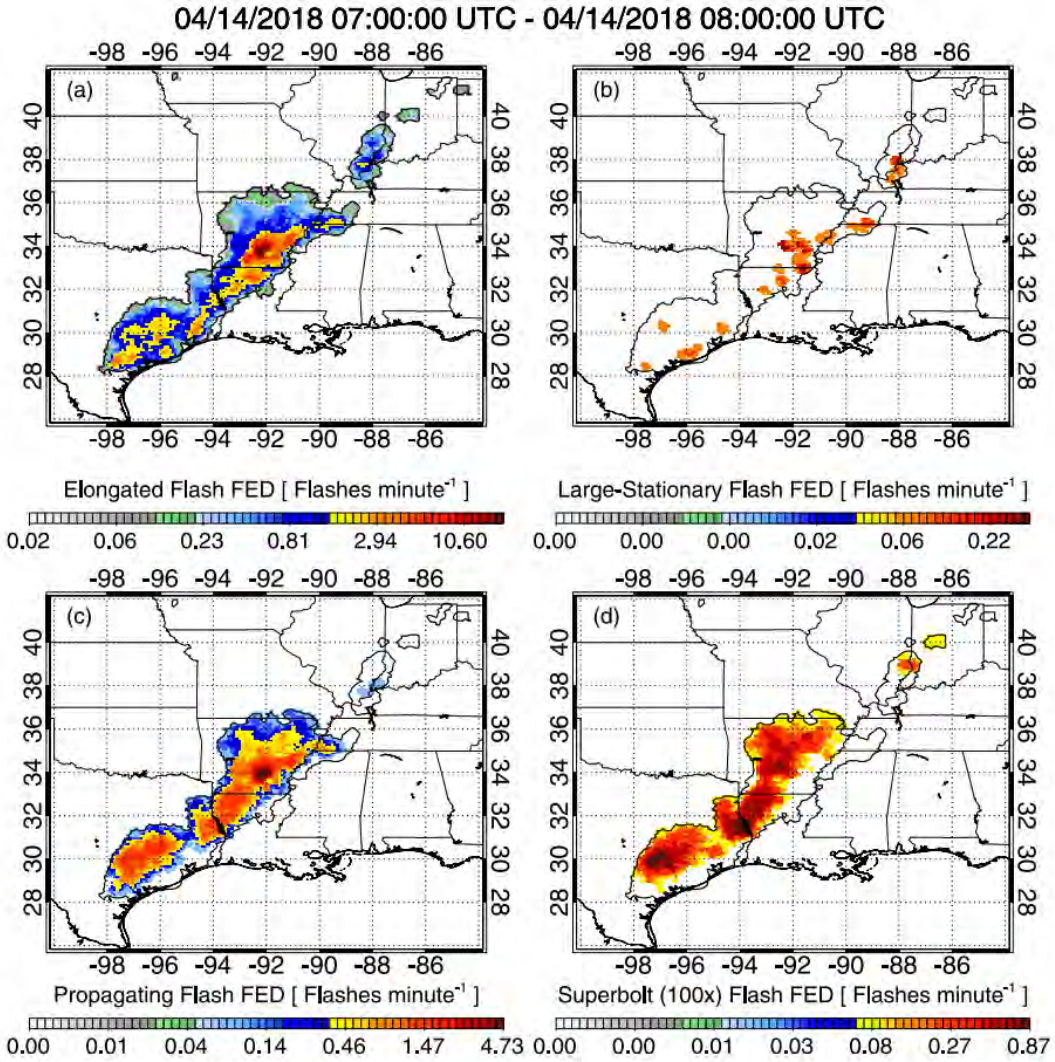


Figure I.2.6: GLM Flash Extent Density (FED) for 1-hour records of a matured MCS. Flash types (a) elongated flashes, (b) large-stationary flashes, (c) propagating flashes, and (d) superbolts (FED times 100) are separated. Colored pixels approximate the GLM image pixels. From Peterson et al. (2020a).

Measured optical energy is integrated over the time frame and is influenced along the path of photons through the atmosphere. There is no direct relation to the emitted peak currents of the lightning source, and the optical sensors do not measure the charge polarity. Optical cloud top illumination does not provide altitude information. Although approaches were made to discriminate flash types from optical characteristics of observed flashes (Koshak, 2010), currently flash types are not separated. The systems map CG and IC flashes with total flash detection efficiency (DE) greater than 50 % (Nag et al., 2015), but usually lower at day than at night (details in following chapters). Thresholding and filtering allow detection of visible light from lightning even during the day.

I.2.7 Thunder and acoustic lightning location

Thunder emerges from strong heating within the lightning conducting channel. The air expands rapidly (<10 µs). The contraction and cooling after the expansion happens relatively slowly (some tens of milliseconds) (MacGorman and Rust, 1998, p. 98). The discharge process creates a shock wave that triggers an acoustic wave propagation. Audible thunder from heating has a peak frequency at about 100 Hz. The sound from one flash might last from a few seconds to one minute while the signal amplitudes vary. Acoustic waves propagate at the speed of sound (about 330 $\mathrm{m\,s^{-1}}$ in the atmosphere). If reflection and scattering are not taken into account, the duration of thunder is the time sound needs to travel from the closest and the furthest point, relative to the observer, of the flash to the observer. It can be used to estimate the minimal extent of the flash (minimal because only the two points determine the duration, not all channels). As the lightning channel acts as series of acoustic sources, complex waveforms are registered for thunder as superpositions of several acoustic waves. The recorded signal also depends on the position of the observer relative to the lightning channel. As for electromagnetic waves, lower frequency signals can maintain the wave amplitude over longer distances than higher frequency acoustic waves (especially above 100 Hz) (MacGorman and Rust, 1998, p. 103). Therefore, the frequency decreases with the distance to the source. Furthermore, vertical temperature and density changes cause refraction and upward curvature of the acoustic ray and thereby limit the maximum distance possible of propagation. A good estimation of the maximum range to hear thunder is about 25 km, however, it depends on the state of the atmosphere (Gallin et al., 2016, Lacroix et al., 2018). A second type of acoustic signals related to thunderstorms is infrasonic thunder from electrostatic sources. It may be recorded directly underneath thunderstorms. Little is known in detail about infrasonic thunder. It is not used by LLSs.

Acoustic lightning mapping techniques were first adapted from under water techniques, e.g., sonar and oil exploration. Any rugged outdoor microphone sensitive to flat frequency responses between a few hertz and a few hundred hertz can be used to record thunder (MacGorman and Rust, 1998, p. 149). As electromagnetic waves propagate much faster than acoustic waves within the atmosphere, the pulse's distance can be estimated from the time delay between E-field change and acoustic pulse. Acoustic location and mapping techniques include Thunder Ranging and Ray Tracing. The former needs at least 3 microphones in a noncollinear array. Typical distances between the microphones are 1 km to 3 km. Pulse ranges to the microphone are computed as the speed of sound (that can be modified according to the propagation) and propagation time. The intersection of ranges of a certain acoustic pulse to each microphone provides the source (i.e., lightning) location in 3D (latitude, longitude, altitude). The propagation time requires the identification of common features (i.e., used as certain pulses) in E-field changes and thunder waveforms at each of the microphones.

Ray tracing uses at least 3 microphones separated by only a few tens of meters. A cross-correlation analysis finds commonly recorded waveform features at each microphone. Then, the propagation time of common pulses is computed between each pair of microphones. The propagation times and baselines are used to infer the initial azimuth and elevation of the source. The total path length results from the delay between the occurrence of the flash and arrival of the acoustic pulse at the microphones. Both initial azimuth and elevation and the total path length are inferred from a model atmosphere to trace the signal back to the pulse source location. This is necessary as paths are curved due to temperature and humidity variations, and the wind. Several thunder pulse sources within one lightning channel may be detected.

Thunder ranging is much faster than ray tracing and can be applied to provide an idea of the direction and general location of lightning channels. Ray tracing gives more details about lightning channels and more reliable source locations. Limitations of all acoustic lightning mapping techniques arise from the limited range of thunder within the atmosphere and a lack of information about the charge or development of the lightning source. Typical location errors are about 10 % of the distance from the microphone to the source.

Part II

Meteosat Third Generation (MTG) Lightning Imager (LI) Pseudo-Observations

This part of the thesis is dedicated to generating pseudo-observations of the future MTG-LI sensor. Realistic MTG-LI data are required in the final Part III developing a method to assimilate lightning data in the regional, convection-permitting numerical weather prediction (NWP) model of Météo-France (AROME-France). The generation of MTG-LI pseudo-observations includes the analysis of both French and US LLSs, the detailed analysis and intercomparison of ground-based and space-borne LLSs, and eventually a sophisticated algorithm referred to as GEO lightning pseudo-observation generator.

Contents of this Part

II.1 | Study of Lightning Observations Recorded by Operational Lightning Locating Systems in France

The previous part of this thesis introduced the physical nature of lightning and different techniques to detect and locate lightning discharges. The LIS on board the ISS is to date the only space-borne instrument providing lightning observations over France. Studying the measured optical flashes can reveal first impressions of what can be expected from the future GEO instrument MTG-LI being constructed to observe lightning at the same 777.4 nm-band as ISS-LIS. Ground-based lightning observation is more common in France. There are LF networks, e.g., Meteorage and LINET, there is the SAETTA LMA on Corsica Island, and there are other research LLSs, e.g., locating acoustic signals from lightning. Concurrent records of the ISS-LIS and those ground-based LLSs, that are in general well understood, can provide insight into the specifications of the optical lightning observation over France.

Lightning records also gain growing interest of scientists because the occurrence of lightning correlates well with regions of deep convection, often heavy precipitation, and potentially severe weather. Avila et al. (2010) shows that the occurrence of deep convection and lightning (over land) are closely related in a statistical sense and on a global scale. With advances in computer power and refined spatial and temporal resolutions of numerical weather prediction (NWP) models, deep convection is nowadays resolved in several operational NWP models, e.g., AROME-France. Therefore, lightning observations might be implemented to help the model locate convection in the initial state and consequently improve the prediction of convection and the related weather phenomena. This topic is detailed in the following chapters of the thesis.

In 2017, when work at the present thesis started, there has been no data of GEO LLSs available yet, but the analysis and use of GEO lightning observations are prepared in this chapter. Figure II.1.1 puts this chapter in the context of this thesis. Three different types of operational LLSs in France are introduced, i.e., (i) the LF network Meteorage, (ii) the VHF LMA SAETTA, and (iii) the LEO ISS-LIS. The LLSs and their observations are examined to develop data processing, flash merging, and flash matching algorithms, to understand specific features of ground- and spaced-based lightning observations, and to gain knowledge of the capabilities of the analyzed LLS techniques both individually and relative to each other. The conducted multi-instrument intercomparison, marked by the yellow frame, is the first of its kind in Europe and at extratropical latitudes in general. The work included experts from French research institutions and collaboration with experts for the LIS in the USA.

The following chapters will make use of the algorithms and methods developed and presented here (see Figure II.1.1). ISS-LIS that observes lightning on a LEO is used as a common reference

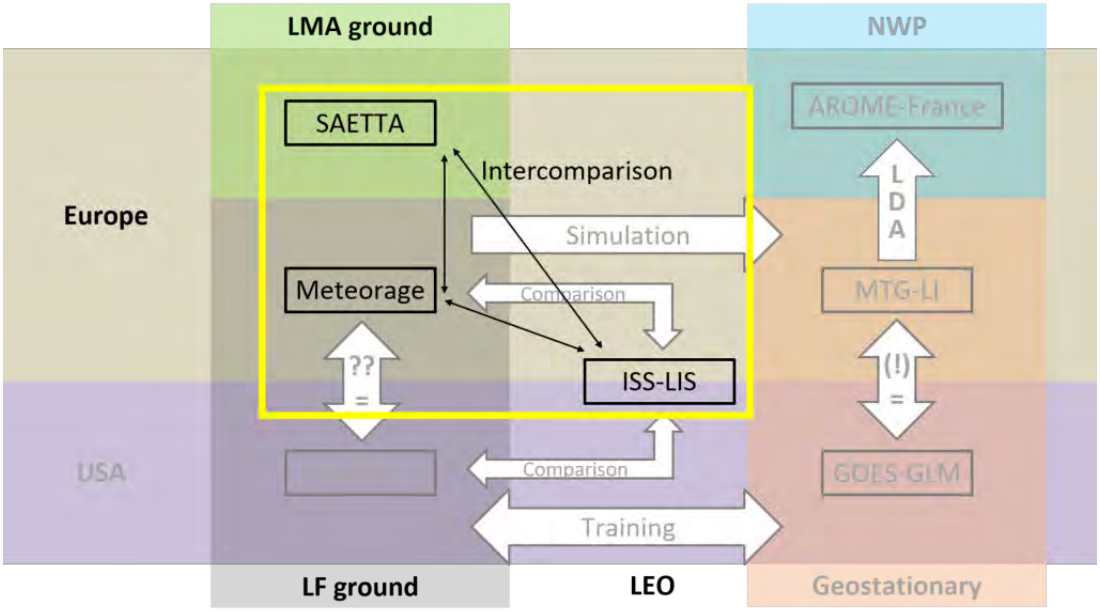


Figure II.1.1: Flow chart of work phases during this thesis as Figure 1. The LLS and phases detailed in this chapter are highlighted in the yellow frame, while other phases are shaded.

LLS to indirectly compare French and US LF networks (Chapter II.2). Knowledge about their similarities and differences is crucial for this thesis analyzing then GEO lightning observations of the GLMs in the southeastern (SE) USA (Chapter II.3). The GLMs provide the same kind of lightning observation as MTG-LI will do, however, over the Americas and adjacent oceans rather than over France. Understanding of and confidence in the LSS performances is needed to apply any method developed based on US NLDN and GLM lightning observations in other parts of the world. In particular, a GEO lightning pseudo-observation generator is developed based on NLDN and GLM observed flash characteristics (detailed in Chapter II.4), and then applied to Meteorage records to simulate MTG-LI pseudo-observations. In summary, this chapter presents the first step towards generating realistic, synthetic pseudo MTG-LI data for studying the assimilation of GEO lightning observations (see Chapter III.4).

II.1.1 Paper abstract

The new space-based Lightning Imager (LI) onboard the Meteosat Third Generation (MTG) geostationary satellite will improve the observation of lightning over Europe, the Mediterranean Sea, Africa and the Atlantic Ocean from 2021 onwards. In preparation for the use of the upcoming MTG-LI data, we compare observations by the Lightning Imaging Sensor (LIS) on the International Space Station (ISS), which applies an optical technique similar to MTG-LI, to concurrent records of the low-frequency (LF) ground-based network Meteorage. Data were analyzed over the northwestern Mediterranean Sea from 1 March 2017 to 20 March 2018. Flashes are detected by ISS-LIS using illuminated pixels, also called events, within a given (2.0 ms) frame and during successive frames. Meteorage describes flashes as a suite of intra-cloud and cloud-to-cloud (IC) pulses and/or cloud-to-ground (CG) strokes. Both events as well as pulses and strokes are grouped to flashes using a novel in-house algorithm.

In our study, ISS-LIS detects about 57 % of the flashes detected by Meteorage. The flash detection efficiency (DE) of Meteorage relative to ISS-LIS exceeds 80 %. Coincident matched

flashes detected by the two instruments show a good spatial and temporal agreement. Both peak and mean distances between matches are smaller than the ISS-LIS pixel resolution (about 5.0 km). The timing offset for matched ISS-LIS and Meteorage flashes is usually shorter than the ISS-LIS integration time frame (2.0 ms). The closest events and the pulses and strokes of matched flashes achieve sub-millisecond offsets. Further analysis of flash characteristics reveals that longer-lasting and more spatially extended flashes are more likely detected by both ISS-LIS and Meteorage than shorter-duration and smaller-extent flashes. The ISS-LIS relative DE is lower for daytime versus nighttime as well as for CG versus IC flashes.

A second ground-based network, the very high-frequency (VHF) SAETTA Lightning Mapping Array (LMA), further enhances and validates the lightning pairing between ISS-LIS and Meteorage. It also provides altitude information on the lightning discharges and adds a detailed lightning mapping to the comparison for verification and better understanding of the processes. Both ISS-LIS and Meteorage flash detections feature a high degree of correlation with the SAETTA observations (without altitude information). In addition, Meteorage flashes with ISS-LIS match tend to be associated with discharges that occur at significantly higher altitudes than unmatched flashes. Hence, ISS-LIS flash detection suffers from degradation, with low-level flashes resulting in lower DE.

II.1.2 Paper: Concurrent satellite and ground-based lightning observations from the Optical Lightning Imaging Sensor (ISS-LIS), the low-frequency network Meteorage and the SAETTA Lightning Mapping Array (LMA) in the northwestern Mediterranean region

Reference: Erdmann, F., Defer, E., Caumont, O., Blakeslee, R. J., Pédeboy, S., and Coquillat, S.: Concurrent satellite and ground-based lightning observations from the Optical Lightning Imaging Sensor (ISS-LIS), the low-frequency network Meteorage and the SAETTA Lightning Mapping Array (LMA) in the northwestern Mediterranean region, Atmos. Meas. Tech., 13, 853–875, https://doi.org/10.5194/amt-13-853-2020, 2020.

Atmos. Meas. Tech., 13, 853–875, 2020
https://doi.org/10.5194/amt-13-853-2020

Atmospheric Measurement Techniques
Open Access
EGU

Concurrent satellite and ground-based lightning observations from the Optical Lightning Imaging Sensor (ISS-LIS), the low-frequency network Meteorage and the SAETTA Lightning Mapping Array (LMA) in the northwestern Mediterranean region

Felix Erdmann[1,2], Eric Defer[1], Olivier Caumont[2], Richard J. Blakeslee[3], Stéphane Pédeboy[4], and Sylvain Coquillat[1]

[1]Laboratoire d'Aérologie, Université de Toulouse, CNRS, UPS, Toulouse, France
[2]CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France
[3]NASA George C. Marshall Space Flight Center/NSSTC, Huntsville, AL, USA
[4]Météorage, Pau, 2 avenue du Président Pierre Angot – CS 8011 64053 Pau CEDEX 9, France

Correspondence: Felix Erdmann (erdmann.professional@gmx.de)

Received: 12 April 2019 – Discussion started: 4 July 2019
Revised: 25 November 2019 – Accepted: 5 December 2019 – Published: 20 February 2020

Abstract. The new space-based Lightning Imager (LI) onboard the Meteosat Third Generation (MTG) geostationary satellite will improve the observation of lightning over Europe, the Mediterranean Sea, Africa and the Atlantic Ocean from 2021 onwards. In preparation for the use of the upcoming MTG-LI data, we compare observations by the Lightning Imaging Sensor (LIS) on the International Space Station (ISS), which applies an optical technique similar to MTG-LI, to concurrent records of the low-frequency (LF) ground-based network Meteorage. Data were analyzed over the northwestern Mediterranean Sea from 1 March 2017 to 20 March 2018. Flashes are detected by ISS-LIS using illuminated pixels, also called events, within a given (2.0 ms) frame and during successive frames. Meteorage describes flashes as a suite of intra-cloud and cloud-to-cloud (IC) pulses and/or cloud-to-ground (CG) strokes. Both events as well as pulses and strokes are grouped to flashes using a novel in-house algorithm.

In our study, ISS-LIS detects about 57 % of the flashes detected by Meteorage. The flash detection efficiency (DE) of Meteorage relative to ISS-LIS exceeds 80 %. Coincident matched flashes detected by the two instruments show a good spatial and temporal agreement. Both peak and mean distances between matches are smaller than the ISS-LIS pixel resolution (about 5.0 km). The timing offset for matched ISS-LIS and Meteorage flashes is usually shorter than the ISS-LIS integration time frame (2.0 ms). The closest events and the pulses and strokes of matched flashes achieve submillisecond offsets. Further analysis of flash characteristics reveals that longer-lasting and more spatially extended flashes are more likely detected by both ISS-LIS and Meteorage than shorter-duration and smaller-extent flashes. The ISS-LIS relative DE is lower for daytime versus nighttime as well as for CG versus IC flashes.

A second ground-based network, the very high-frequency (VHF) SAETTA Lightning Mapping Array (LMA), further enhances and validates the lightning pairing between ISS-LIS and Meteorage. It also provides altitude information on the lightning discharges and adds a detailed lightning mapping to the comparison for verification and better understanding of the processes. Both ISS-LIS and Meteorage flash detections feature a high degree of correlation with the SAETTA observations (without altitude information). In addition, Meteorage flashes with ISS-LIS match tend to be associated with discharges that occur at significantly higher altitudes than unmatched flashes. Hence, ISS-LIS flash detection suffers from degradation, with low-level flashes resulting in lower DE.

Published by Copernicus Publications on behalf of the European Geosciences Union.

1 Introduction

Lightning defines electrical discharges within the atmosphere. The discharges can happen within a cloud, between clouds (IC), or between a cloud and the ground (CG). The total lightning activity (IC+CG) is of interest for, e.g., numerical weather prediction (NWP) as lightning serves as tracer for deep convection. The total lightning flash rate is associated with storm intensity features. For example, Deierling and Petersen (2008) found a strong correlation between the updraft volume above the −5 °C level in clouds and total lightning activity. Deierling et al. (2008) show a fairly stable relationship and strong correlation between the precipitation ice mass flux, the non-precipitation ice mass flux and their product on the one hand and the total lightning flash rate on the other. Graupel and small hail ice mass correlate especially well with the mean total lightning rate in their study. Among others, Mattos et al. (2017) investigated the life cycle of thunderstorms and processes leading to the different discharge types. They found in their analysis of 46 isolated thunderstorms that in 98 % of their cases, the first CG flash is preceded by IC lightning by approximately 6 min on average.

At this time, lightning observations in Europe are mainly made with ground-based sensors. To maximize the impact of lightning data on assimilation in NWP systems, total lightning should be observed continuously over large areas. In a few years, the new Lightning Imager (LI) onboard the Meteosat Third Generation (MTG) satellite (Stuhlmann et al., 2005) will provide continuous lightning observation over Europe, the Mediterranean Sea, Africa, the Atlantic Ocean and parts of Brazil. The satellite sensor will be able to detect total lightning including CG and IC flashes when launched in the 2021 time frame. The Lightning Imaging Sensor (LIS) on the International Space Station (ISS) (Blakeslee and Koshak, 2016) creates a unique opportunity to provide proxy data to help prepare research and operational applications for the MTG-LI data. It overpasses, among others, wide parts of Europe, including the entire Mediterranean region. ISS-LIS is in principle similar to the planned MTG-LI so that ISS-LIS data can to some extent mimic the upcoming MTG-LI data. In addition, a comparison between European ground-based lightning observation networks and ISS-LIS should improve the understanding of ground- and space-based lightning observations. All instruments and networks are hereafter simply referred to as lightning locating systems (LLSs).

These comparisons focus on the spatial and temporal coincidence of flashes reported by the various systems, resulting in measures of detection efficiency (DE) as a function of the flash parameters. This study uses the term relative DE. It is defined as the ratio of the number of matched flashes to the number of flashes in the other (reference) LLS, expressed as a percentage.

An LIS was previously operational on the Tropical Rainfall Measurement Mission (TRMM) satellite (e.g., Christian et al., 1999; Cecil et al., 2005). Several LLSs comparisons exist for regions covered by TRMM-LIS. The focus of the following (not exhaustive) literature review is on observational analyses rather than laboratory experiments; e.g., Boccippio et al. (2002). Ground-based LLSs observe different frequency ranges of the lightning radio signal. They are classified as, e.g., very low-frequency (VLF) and low-frequency (LF) LLSs as well as very high-frequency (VHF) LLSs (e.g., Fig. 2; Cummins and Murphy, 2009). A summary of detection characteristics, (dis)advantages and the range of the various ground-based LLSs is provided in Nag et al. (2015). VLF–LF systems detect lightning at middle to long ranges. Their DE is somewhat limited. It varies for different networks and flash types (CG flash DE is usually higher than IC flash DE for VLF–LF LLSs) but increases in general with lower baseline distance.

Thompson et al. (2014) aimed to explore suitable proxy data for the Geostationary Lightning Mapper (GLM) (Goodman et al., 2013). They report a pulse and stroke DE maximum for two long-range LLSs, the World Wide Lightning Location Network (WWLLN) and the Earth Networks Total Lightning Location Network (ENTLN), of 18.9 % and 63.3 %, respectively, relative to 18-month records of TRMM-LIS groups (a combination of adjacent illuminated pixels in the optical image that occur in the same 2 ms time frame). The maxima were found over the Pacific Ocean for WWLLN and near North America for ENTLN (within the analyzed region with the highest sensor density) in 2010 and 2011. They did not study how many WWLLN and ENTLN pulses and/or strokes had coincident TRMM-LIS groups.

Rudlosky et al. (2017) analyzed the performance of the Global Lightning Dataset 360 (GLD360) relative to TRMM-LIS from 2012 to 2014 in different regions. GLD360 was able to detect 63.6 % of the TRMM-LIS flashes in North America in 2014, the maximum DE reported in their study. The performance steadily increased from 2012 to 2014. The relative DE of GLD360 increased with the TRMM-LIS flash duration, flash extent and group number. The mean (median) location offset of the nearest GLD360 stroke to the matched TRMM-LIS flashes was 8.7 km (7.0 km). Rudlosky et al. (2017) applied the assumption that TRMM-LIS would detect all flashes within its field of view but did not study the reverse problem, i.e., the relative DE of TRMM-LIS to the GLD360 flashes or strokes.

Defer et al. (2005) used both the UK Met Office long-range VLF arrival time difference (ATD) system and TRMM-LIS to study the lightning activity in the eastern Mediterranean Sea for 20 d during winter 2008–2009. For their investigation of the flash scale, they developed and employed their own algorithm for TRMM-LIS flashes. The flash density analysis exhibits a general agreement between ATD and TRMM-LIS. The relatively small dataset, and the fact that ATD detected mostly CG lightning, limited the ability to gain overall statistics.

Bitzer et al. (2016) tested a Bayesian approach on the DE of TRMM-LIS and ENTLN by implementing the conditional

DEs of the two LLSs relative to each other. They found a relative conditional group-to-pulse DE of 52 % (27 %) for TRMM-LIS to ENTLN (ENTLN to TRMM-LIS near North America in 2013). They also addressed peak timing differences and distances for the collocated discharges (again LIS groups, ENTLN pulses; see Sect. 3.2 for details). Bitzer et al. further tested the effect of assimilating one dataset into the other on the detected number of discharges; i.e., 23.6 % of discharges could be added to the total number of observed flashes after combining the datasets.

While the previous papers focused on the DE, Höller and Betz (2010) analyzed TRMM-LIS and a VLF–LF lightning location network (LINET) in order to generate random proxy optical data from a given set of LINET data using model distribution functions. The outcomes are of specific interest for proxy data for the MTG-LI. Besides the relative DEs (approximately 50 % for both LLSs), they investigated the distribution functions and correlations between the TRMM-LIS group and the LINET pulse and stroke number per flash, flash extent, and duration as well as between LINET pulse and stroke amplitude and TRMM-LIS group radiance. Although the Pearson correlation coefficients remained low, the approach can be further refined for high-fidelity MTG-LI proxy data.

VHF LLSs are sensitive to lightning channel formation and leader processes, which occur multiple times during a single flash. Hence, VHF LLSs typically feature high DE performances and three-dimensional (3-D) mapping of lightning channel propagation and spatial extent (Thomas et al., 2004). VHF LLSs depend on direct line-of-sight detection, and thus the range suffers from the Earth's curvature and terrain shading effects.

Thomas et al. (2000) presented a case study of a storm in Oklahoma, USA, at local nighttime. The storm was observed by both the local VHF lightning mapping array (LMA) and TRMM-LIS; 108 of the 128 LMA lightning discharges were detected by TRMM-LIS, and the LMA detected all TRMM-LIS flashes. The lightning missed by TRMM-LIS was mainly confined to low-altitude discharges, i.e., below 7.0 km. Optical signals of lightning discharges that propagated via scattering to the upper part of the cloud were easily detected by TRMM-LIS.

Blakeslee et al. (2013) studied the São Paulo LMA (SP LMA) dataset and its capability to serve as GLM proxy data. TRMM-LIS events were in good agreement with the concurrent SP LMA, ENTLN and LINET observations regarding latitude, longitude and timing. The records showed as expected more VHF (SP LMA) sources than VLF pulses and strokes (ENTLN and LINET) per flash.

Due to the TRMM satellite orbit, the comparisons of TRMM-LIS and ground-based LLSs records are restricted to tropical and subtropical regions between about 38° N and 38° S. As a result of its higher-inclination orbit, ISS-LIS now allows for the observation of extratropical thunderstorms to extend to 55° N and 55° S. The higher-latitude storms might show different behaviors to their tropical and subtropical counterparts due to modified cloud vertical extent and forcing like the general wind field, average temperature and temperature gradients. Our study concentrates on the characteristics of lightning flashes over the northwestern (NW) Mediterranean Sea and should contribute to a better understanding of both European storms and European LLSs. This allows for the first time an intercomparison of LIS and European LLSs. Three LLSs operating in different spectral regions (near-IR, VLF–LF, VHF) are compared: the satellite-based ISS-LIS operational since March 2017, the French Meteorage VLF–LF LLS and the VHF SAETTA LMA on Corsica. The relative DE of ISS-LIS to Meteorage (and reverse) is analyzed, while SAETTA is used to verify and understand the results. Indeed, the spatially and temporally high resolution of SAETTA's measurements capture the structure and the life cycle of each lightning flash and gather additional information, i.e., discharge altitude, to more thoroughly assess ISS-LIS and Meteorage strengths and weaknesses. Besides the commonly investigated relative DEs, distances and timing offsets, this work also examines specific characteristics of matched ISS-LIS and Meteorage flashes. It aims to provide the basis for mimicking optical satellite-based lightning data from a VLF–LF LLS.

In Sect. 2, ISS-LIS, Meteorage and SAETTA are introduced as are the data processing, developed algorithms and the investigation methodology. Results are presented in Sect. 3. A brief summary and a discussion are given in Sect. 4.

2 Instrumentation and methodology

This paper aims to identify the individual lightning detection characteristics by the satellite-based ISS-LIS, the VLF–LF Meteorage and the VHF SAETTA LLSs. ISS-LIS, installed on the International Space Station in 2017, has been acquiring data since 1 March 2017. Our intercomparison of the LLSs covers the period from 1 March 2017 until 20 March 2018. The region was limited to 40.5 to 44.0° N and 7.0 to 11.0° E around the island of Corsica in the NW Mediterranean Sea. Figure 1 shows the domain with accumulated data of one overpass (a), an infrared (IR) satellite picture (b), and the example of one flash recorded by ISS-LIS, Meteorage and SAETTA (c). The three instruments are introduced within this section. In total, the ISS-LIS field of view (FOV) intersected the region of interest 851 times during the study period, with 26 of the overpasses exhibiting lightning activity. In this work, all times are given in Coordinated Universal Time (UTC). Altitudes are defined above sea level (a.s.l.). Distances are calculated using Vincenty's formulae (Vincenty, 1975) based on the WGS 84 reference ellipsoid, which is more accurate on Earth than, for example, great circle distances (assumes the Earth as an oblate sphere rather than a sphere). The term detection efficiency

(DE) means in the following the DE for flashes, not the event or pulse–stroke DE.

2.1 ISS-LIS

The ISS operates in low Earth orbit (LEO) and overpasses one region on the surface up to three times a day (up to two times in the tropics). Lightning observation of a specific point lasts up to 90 s per overpass due to the ISS orbit characteristics and the LIS FOV of approximately $655 \times 655\,\mathrm{km}^2$. The optical lightning detection is performed at a wavelength of 777.4 nm at the atomic oxygen line. ISS-LIS observes both IC and CG discharges but cannot distinguish the lightning type. ISS-LIS captures an image of the Earth every 2 ms, referred to as a frame. The LIS focal plane consists of a 128×128 pixel charge-coupled device (CCD) that is read out every 2 ms. The pixel FOV ranges between 4.5 km (nadir) and 6.2 km at the edges (Dennis Buechler, personal communication 2019). Blakeslee and Koshak (2016) apply a four-step filtering approach, involving a spatial, spectral, temporal and background subtraction filter, to identify pixels with lightning activity. This is required to detect the lightning during daytime when the sunlight reflected off the cloud tops otherwise overwhelms and masks the lightning signal (i.e., it is daytime lightning detection that drives the design of space-based lightning detectors such as LIS and the new MTG-LI). An illuminated pixel that breaks a predefined threshold in a given 2 ms frame is identified as an event. Events define the smallest units of the optical signals in the ISS-LIS dataset. Their latitude and longitude correspond to the pixel center. A group is the next unit of ISS-LIS data. An ISS-LIS group contains one or more events occurring within the same time frame and in adjacent pixels of the ISS-LIS image (Christian et al., 2000). Next, groups are organized into flashes so that a flash can consist of one or multiple groups. A weighted Euclidean distance (WED) employs spatial and temporal clustering with 330 ms and 5.5 km, respectively, to merge groups in flashes (Mach et al., 2007). The locations of groups and flashes are defined by the radiance-weighted average positions of their events and groups, respectively. Finally, an area contains all flashes with distances of less than 16.5 km to each other. The National Aeronautics and Space Administration (NASA) provides the ISS-LIS at different post-processing levels. In the latest available version, P0.2, the quality control is already close to its (expected) final stage, but the data may contain some undetected minor errors (Blakeslee et al., 2017). The main difference will concern the detection efficiency. The fully validated flash density should not differ more than 5.0 % to 10.0 % from version P0.2 (Richard J. Blakeslee, personal communication, 2018). LIS data comprise 2 ms scientific data, e.g., the time, latitude, longitude and optical amplitude count of events and instrument, platform or external errors to verify the data quality, and housekeeping data. The available ISS-LIS P0.2 version data do not yet include the (background-)calibrated radiance. The strength of the optical signal is defined by the raw amplitude count. It depends somewhat on the background value, but in general the radiance increases with the raw count (Dennis E. Buechler, NASA MSFC, personal communication, 2019). The housekeeping data, received every second, contain among others LIS viewtime with information about the FOV at a certain time. It is provided on a $0.5° \times 0.5°$ grid. ISS-LIS viewtime is fundamental for the intercomparison to continuous observations at the ground.

The original ISS-LIS data contain times in International Atomic Time with reference to the 1 January 1993 format (TAI93). For intercomparison of the LLSs, times are converted to UTC while taking the missing leap seconds into account. The ISS-LIS times include a time-of-flight (TOF) correction accounting for the time photons need to travel from the optical source at cloud top to the satellite.

2.2 Meteorage

The Meteorage LF LLS uses Vaisala LS7002 sensors (Vaisala, 2013) at a frequency between 1 and 350 kHz. It includes 21 ground sensors across France and contributes to the European Cooperation for Lightning Detection (EUCLID). EUCLID comprises lightning sensors all over Europe and helps to improve the performance of national LLSs (Schulz et al., 2016). The LS7002 sensors measure the signals related to CG strokes as well as IC pulses and thus the total lightning. Vaisala has a CG DE of 95 % and a DE for IC of 50 %. Pédeboy et al. (2018a) stated that indeed 97 % of the CG flashes and 56 % of the IC flashes were detected by Meteorage (68.3 % overall DE relative to LMA flashes). The theoretical median location accuracy is approximately 250 m and improves inside the network to about 150 m. Pédeboy et al. (2018a) found a reduced median location accuracy for IC flashes of 1.64 km. Time synchronization applies a GPS receiver with an accuracy of 50 ns to UTC. The lightning location needs at least two sensors (each provides a time and an angle of arrival) by applying combined magnetic direction finding and time-of-arrival techniques. Lightning can be detected at a distance up to 1500 km from a sensor. In practice, the use of ionospheric reflection is avoided, hence limiting the sensor range to about 625 km. It ensures that the ground plane wave front of the signal is measured rather than a reflected wave of the lightning-related signal. Our study makes use of Meteorage lightning pulse and stroke data. For each pulse and stroke, the occurrence time, latitude, longitude, the amplitude with polarity and the type (IC/CG) are provided. Meteorage observes lightning continuously within its range. In the intercomparison with ISS-LIS, data are disregarded if observation space or time do not fit the corresponding ISS-LIS viewtimes.

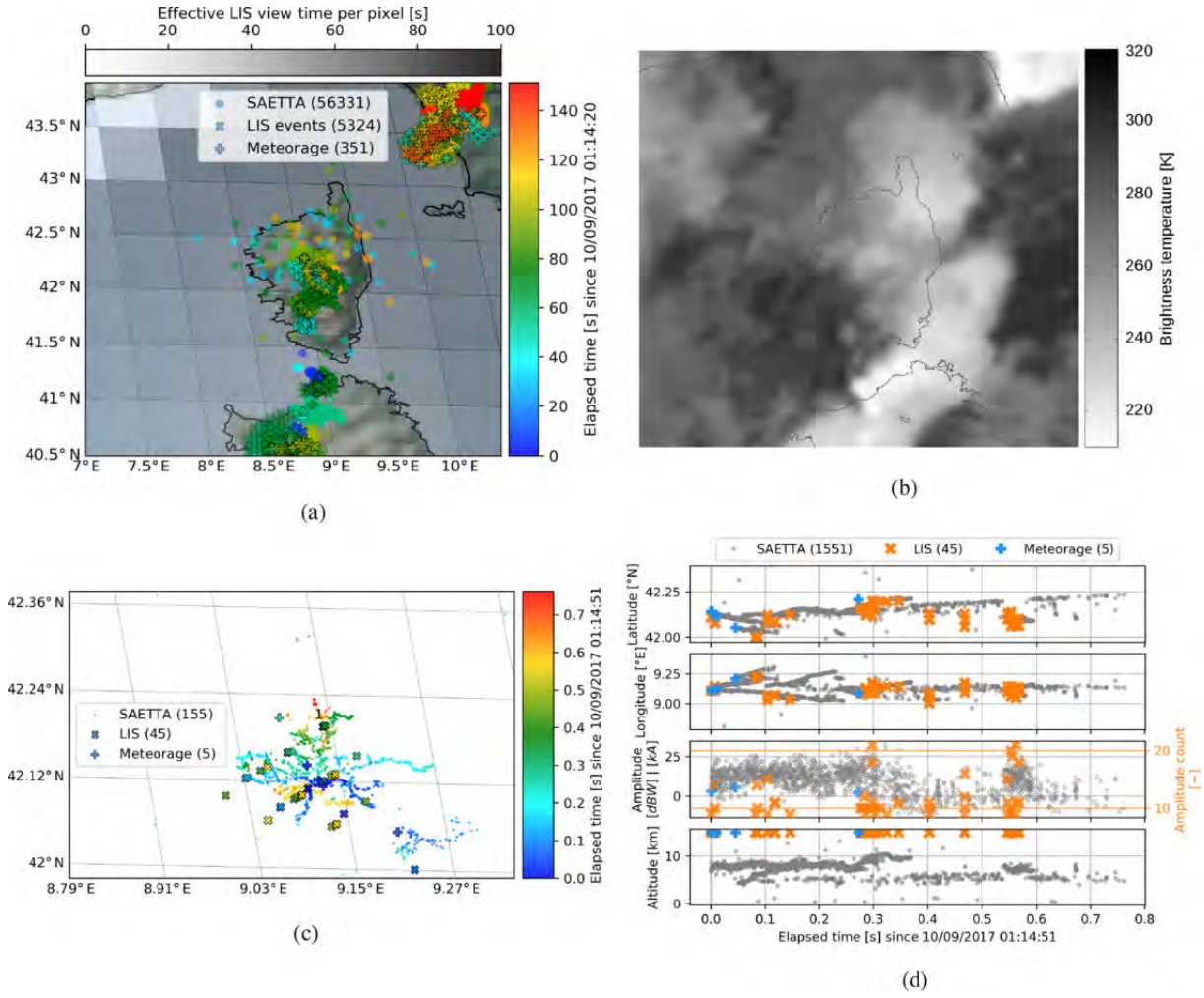


Figure 1. Observations of ISS-LIS events (as pixel centers), Meteorage pulses and strokes, and SAETTA VHF sources (as indicated) during one ISS overpass over Corsica on 10 September 2017 **(a)**. The ISS-LIS viewtime is presented as grayscale in the background. Numbers in parentheses give the number of SAETTA VHF sources, ISS-LIS events, and Meteorage pulses and strokes, respectively. Panel **(b)** shows the infrared (IR 10.8 µm) satellite image of the same day at 01:15:00 UTC (data visualization provided by the AERIS/ICARE Data and Services Center). One flash over Corsica detected by the three LLSs during the same ISS overpass is shown in **(c)** as a map and **(d)** as time series of latitude, longitude, signal strength amplitude (dBW for SAETTA, kA for Meteorage and amplitude count for ISS-LIS) and altitude (LIS and Meteorage set to 15 km).

2.3 SAETTA (Suivi de l'Activité Electrique Tridimensionnelle Totale de l'Atmosphère)

The LMA technology was developed by New Mexico Tech (Rison et al., 1999). The SAETTA LMA operates in the 60–66 MHz VHF band, with an 80 µs analysis window (Coquillat et al., 2014), and consists of 12 LMA stations distributed over the island of Corsica. The distance between the network's northernmost and southernmost (westernmost and easternmost) stations is approximately 180 km (70 km). The station altitude ranges from 3.8 to 1950.2 m a.s.l. SAETTA maps the total lightning activity. A minimum of six stations is needed to capture a lightning source in 3-D. Redundant information from more stations improves the location accuracy and consequently decreases the chance of mislocation and possible noise (e.g., single VHF sources in Fig. 1c). As a drawback, fewer VHF sources and flashes are detected simultaneously by more than six stations. Aiming at a high flash DE, coincident signals at six stations are sufficient for the LMA data in this study.

SAETTA data include the time, latitude, longitude, altitude and amplitude of each lightning source. Lightning location reaches up to a radius of 350 km from the center of the network. The VHF LLS depends on the direct line of sight to a lightning discharge. The altitude of the lowest detectable VHF source increases with the distance to the LMA due to Earth curvature. For example, sources at 100 km (200 km) of distance to a station at sea level must be at least 0.8 km

(3.1 km) in altitude to be visible to that station. Equation (6) of Koshak et al. (2018) is applied here. Their study also investigates the effects of the LMA network geometry, mainly on the altitude errors. For SAETTA, Coquillat et al. (2019a) show that the displacement of two stations in 2016 markedly reduced the radial error and increased the altitude error over wide parts of the studied domain (Fig. 3 in Coquillat et al., 2019a). Therefore, when different sets of at least six stations are involved in the reconstruction of the VHF source position one would expect a different geometry of the network, which influences the location precision. In general the SAETTA location uncertainty increases with the distance to the network center. According to the theoretical model of Thomas et al. (2004), the radial, azimuthal and altitude errors are, at best for VHF sources at 10 km of altitude, 15, 8 and 40 m, respectively, within 50 km from the center of the network (Coquillat et al., 2019a). These theoretical errors reach about 300, 20 and 400 m, respectively, at the borders of the present study domain (Fig. 1a). SAETTA location errors are of the same order of magnitude as those of Meteorage CG location, while the LMA should capture lightning in more detail than the LF LLS.

SAETTA data are employed for locations and times of coincident ISS-LIS or Meteorage observations. Therefore, they are analyzed in space and time regarding the detected ISS-LIS and Meteorage lightning activity. A combined space–time filter identifies SAETTA sources within 0.2° (both latitude and longitude) and (simultaneously) 0.3 s of corresponding ISS-LIS events and Meteorage pulses and strokes. The filtering per (ISS-LIS or Meteorage) flash allows for analyzing the concurrent VHF measurements with, e.g., altitude information. Furthermore, SAETTA data are not used to exclude any ISS-LIS or Meteorage observations, and neither ISS-LIS nor Meteorage data are confined to any SAETTA data condition. They are, however, used to verify the applied data processing approaches, i.e., grouping elements (events, pulses and strokes) to flashes and the analysis of possible false alarms within the lightning detection of ISS-LIS and Meteorage. The maximum altitude of SAETTA sources is bounded at 15.0 km a.s.l., and the maximum reduced χ^2, which defines a measure for the overall uncertainty of the time-of-arrival-based system (Thomas et al., 2004), is set to 0.5.

This study uses SAETTA source altitudes to define flash mean, minimum and maximum altitudes. The flash mean altitude is the true mean of all altitudes of sources coincident with the (ISS-LIS or Meteorage) flash. The minimum altitude is defined as the 10th percentile of the altitudes of concurrent SAETTA sources rather than the true minimum. It is aimed at reducing the influence of noise in the data. In the same manner the flash maximum altitude equals the 90th percentile of the concurrent SAETTA source altitudes instead of the true distribution maximum.

2.4 Flash-grouping algorithm

The NASA LIS flash-clustering algorithm distinguishes events, groups and flashes (Sect. 2.1). It makes use of a WED with a maximum difference of 5.5 km in space and 330 ms in time. The WED analyzes group centroids and not the events in the group to determine if two groups are considered part of the same flash. One flash cannot last longer than 2.0 s (Christian et al., 2000). An analysis of the (P0.2) NASA LIS flash-clustering algorithm revealed that it tends to separate flashes when compared to concurrent SAETTA observations. Similar results were observed by Defer et al. (2005). Consequently, a new algorithm is developed to merge the ISS-LIS events to flashes. It has the additional advantage of treating both ISS-LIS events and Meteorage pulses and strokes. The lightning elements sensed by each LLS, which are the smallest available lightning signals (events as well as pulses and strokes), are merged into flashes. More explicitly, an event of ISS-LIS (pulse and/or stroke of Meteorage) should belong to exactly one flash, and a flash is defined as a collection of events (pulses and strokes). Flash characteristics are derived from the underlying element characteristics; e.g., the positions of its elements are used instead of the mean flash location. This study makes use of the elementary ISS-LIS event data as provided by NASA prior to any data merging. It is accepted that ISS-LIS events do not have a direct representation in the Meteorage-like data. Former studies have claimed that LIS groups roughly correspond to the physical processes detected by VLF–LF LLSs (e.g., Bitzer et al., 2016; Höller and Betz, 2010). Nevertheless, those studies found significantly more groups than pulses and strokes within the same region and time period. Bitzer et al. (2016) found for the number of TRMM-LIS groups to ENTLN pulses and strokes a factor of about 28.4 globally and even 3.7 in North America in 2013. Höller and Betz (2010) analyzed 6.7 groups per pulse or stroke on average. Due to those results, LIS optical groups emerge from both discharge processes measured by VLF–LF sensors but also processes lacking significant VLF–LF radiation. In addition, the detected lightning sources of the applied VHF LLS comply more with the LIS events than the groups. Using events rather than group centroids improves in particular the finding of the coincident LMA data. The analysis of flash extents profits from the use of events in that the extent of an ISS-LIS flash corresponds to the full illuminated area rather than the ISS-LIS group centroid locations. The representation of the flash extent (density) will influence the future assimilation of lightning data in NWP models. A statistical analysis of (ISS-LIS) events and LF strokes and pulses will also be of interest for creating a proxy optical dataset, e.g., for MTG-LI, derived from LF data.

Our grouping algorithm analyzes the elements (events or pulses and strokes) and groups them based on their relative location and time of occurrence to each other. First, the spatial and temporal constraints, $\mathrm{d}s_{\mathrm{merge}}$ and $\mathrm{d}t_{\mathrm{merge}}$, for elements within one flash must be determined. Then, a com-

bined space–time test merges the elements into flashes. It starts with the first available element (in the data of one LLS) and identifies all elements (of the same LLS data) within the range of the constraints. Thereby, an element can only belong to the same flash if both the distance to any element of the flash is less than ds_{merge} and the time difference (to the same element) is shorter than dt_{merge}. All elements identified for a flash (including the initial element) are classified as used. For each used element within a flash, the test is repeated until no unused element can be added to the flash. This step allows for considering propagating flashes and the potentially increasing extent and duration of a flash while adding new elements. The algorithm continues until all elements are classified as used. Our algorithm does not limit the duration of a flash. The number of elements per flash also remains free to the algorithm.

The algorithm verification includes a sensitivity study for ds_{merge} and dt_{merge} (Fig. 2) as well as a comparison to NASA's algorithm and concurrent SAETTA observations (Fig. 3).

Figure 2 gives the number of flashes analyzed from all observations of the approximately 1-year period by using different ds_{merge} (panels a and c) and dt_{merge} (b, d) for ISS-LIS (a, b) and Meteorage (c, d). In general, as expected, smaller ds_{merge} and dt_{merge} increase the flash numbers because fewer individual elements are part of a given flash, and thus more flashes exist for the same elements. LIS flash numbers range from 236 to 4567 for the ds_{merge} (dt_{merge}) between 50.0 and 1.0 km (1.0 and 0.1 s), respectively. For the same constraints, Meteorage flash numbers vary from 340 to 1720 flashes.

The ISS-LIS flash number decreases rapidly for ds_{merge} between 0 and 10 km (Fig. 2a). The rapid change depends on the pixel size within the ISS-LIS image. Hence, it is expected that events of one flash are partitioned within the same frame if the ds_{merge} becomes smaller than the ISS-LIS image pixel size. ISS-LIS flash numbers remain constant for ds_{merge} greater than 15 km for all tested dt_{merge}; 0.3 s balances the need for consistency and the wish for a strict dt_{merge} (Fig. 2b). The resulting flashes are verified against concurrent 3-D SAETTA sources, which supported the choice of our constraints and the identification of resulting flashes. The spatial constraint (ds_{merge} of 15 km) refers directly to event locations and not to group centroids (as for NASA's algorithm). The chosen time constraint for ISS-LIS flashes (dt_{merge} of 300 ms) is similar to the P0.2 NASA flash-clustering algorithm (330 ms).

The same algorithm is applied to group the Meteorage pulses and strokes into flashes. It needs, however, modified constraints ds_{merge} and dt_{merge} since physical processes producing Meteorage pulses and strokes do not always correspond to ISS-LIS events and occur with significantly lower counts. Meteorage pulses and strokes do not cover the full structure and duration of a lightning flash. Figure 2b and d are analyzed for Meteorage flash numbers as demonstrated for ISS-LIS flash numbers in Fig. 2a and c. To find constant

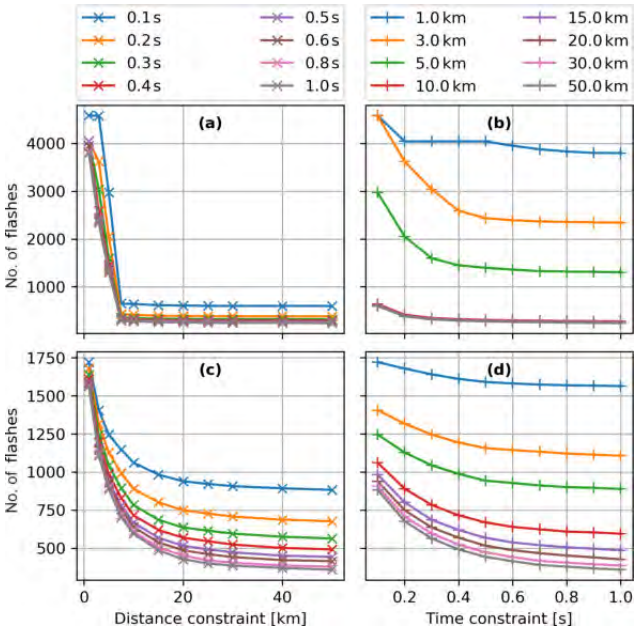


Figure 2. Total flash number based on constant, equal time constraint dt_{merge} (line color) with varying distance ds_{merge} for elements of ISS-LIS **(a)** and Meteorage **(c)** flashes. Panels **(b)** and **(d)** are with constant, equal distance ds_{merge} (line color) and varying dt_{merge} for ISS-LIS and Meteorage, respectively.

constraints suitable for various situations (e.g., vertical cloud structures, severity of a storm, flash rate), resulting flashes for different constraints are verified manually using SAETTA observations. Ultimately, Meteorage pulses and strokes belong to the same flash if they are detected within 20 km and 0.4 s. Due to the limited number of pulses and strokes, Meteorage ds_{merge} and dt_{merge} are coarser than the ISS-LIS merging constraints. Our constraints (20 km, 0.4 s) are consistent with Meteorage's flash-grouping algorithm using a separation distance of less than 10 km for subsequent CG strokes and 20 km if IC pulses are involved. The delay between subsequent discharges of the same flash must be smaller than 0.5 s in Meteorage's algorithm. Höller and Betz (2010) provided a clustering of LINET VLF–LF pulses and strokes to flash scale with 10 km and 1.0 s in space and time, respectively. Hence, their merging constraints for a flash are finer in space but coarser in time.

The determination of ds_{merge} and dt_{merge} does not ensure a perfect arrangement of the elements in flashes. The objective is to find constraints leading to statistical representations of flashes in the ISS-LIS and Meteorage data. Therefore, all identified flashes are double-checked against concurrent 3-D SAETTA observations. Even if it is sometimes challenging to separate the flashes in the SAETTA data, the detailed VHF mapping helps us to understand the processes leading to the identification of the ISS-LIS and Meteorage flashes. The

SAETTA data can also be used to find possible false alarms in the ISS-LIS and Meteorage data.

Figure 3 demonstrates the behavior of NASA's flash-merging algorithm and our developed algorithm for one example. It shows a short time period (6.5 s) during one ISS overpass on 10 September 2017. In Fig. 3a, there is a map of flash locations from the P0.2 NASA flash-merging algorithm and our developed algorithm as well as concurrent VHF SAETTA sources. The ISS-LIS events (not plotted) coincide generally well with the SAETTA observations in both location and time. The mapped observations are presented in latitude, longitude and altitude time series in Fig. 3b; 20 flashes from NASA's algorithm are confronted with 11 flashes from our developed algorithm for the same ISS-LIS events. NASA's merging algorithm somehow splits some flashes, e.g., the flash between 5.3 and 6.2 s where NASA's algorithm identifies five flashes. Our algorithm finds a single flash for that period, and the concurrent SAETTA observations support this result.

Our developed algorithm was additionally tested versus GLM flash-scale data. The GLM algorithm uses the distance between events and not (as ISS-LIS) between group centroids in order to merge events and/or groups to flashes. WED time and spacing for events and/or groups of one GLM flash are 330 ms and 16.5 km, respectively (Goodman et al., 2013). The GLM flash-scale data agree very well with the flashes identified by our algorithm utilizing the underlying GLM events.

2.5 Flash-matching algorithm

ISS-LIS and Meteorage detect lightning in a different way. It was described how the different signals can be merged into a common entity, namely a flash. The intercomparison of LLSs uses the flash scale to find concurrent observations. Individual flashes of both LLSs are sorted into one of the four following categories: LIS detected by both (i.e., an ISS-LIS flash has a coincident Meteorage flash), LIS-only (i.e., no coincident Meteorage observations), Meteorage detected by both (i.e., a Meteorage flash with concurrent ISS-LIS events) or Meteorage-only (i.e., ISS-LIS does not detect the flash). Matching criteria in space (ds_{match}) and time (dt_{match}) are specified. The criteria ds_{match} and dt_{match} do not address the flash mean position and time, respectively, but the single events or pulses and strokes within a flash. Two flashes observed by different LLSs are defined as matched if at least two elements (one per flash) meet both ds_{match} and dt_{match}. A given flash of the reference LLS does not necessarily correspond to exactly one matched flash. It is also possible that a flash meets the matching criteria of more than one given flash (and is collocated with more than one flash). Hence, the two categories, which are LIS detected by both and Meteorage detected by both, are expected to have different counts.

The criteria ds_{match} and dt_{match} are determined through a sensitivity study of the relative DEs of ISS-LIS and Meteorage (Fig. 4). A spatial criterion lower than 10.0 km reduces the relative DE of both ISS-LIS and Meteorage rapidly (Fig. 4a: ISS-LIS, c: Meteorage). In general, the ISS-LIS relative DE is more sensitive to both ds_{match} and dt_{match} than the Meteorage relative DE. This result is triggered by the low number of Meteorage pulses and strokes (compared to the number of ISS-LIS events) effectively hampering the finding of suitable elements, i.e., pulses and strokes, for a collocation. The ISS-LIS relative DE decreases within the entire range of investigated times dt_{match}. The most sensitive behavior occurs for dt_{match} up to 1.5 s (Fig. 4b). Meteorage appears to be sensitive to dt_{match} only up to 0.5 s (Fig. 4d). Despite the differences in sensitivity to the criteria between ISS-LIS and Meteorage, the aim is to use the same ds_{match} and dt_{match} for both LLSs. Finally, ds_{match} of 20 km and dt_{match} of 1.0 s are chosen to balance the individual sensitivities of the LLSs to the criteria. They allow for the identification of matches if, for example, ISS-LIS detects primary IC discharges of a flash and Meteorage only detects a CG stroke occurring during the final stage of the same flash. Our criteria are relatively coarse compared to some former studies (Sect. 1). Höller and Betz (2010) applied the same dt_{match} but an even coarser ds_{match} (i.e., 30 km) to match LINET VLF–LF flashes and TRMM-LIS flashes. Further investigation of the matched flashes, e.g., the distributions of the distances and timing offsets, will demonstrate to what extent matches rely on the fairly coarse criteria.

A detailed analysis of distances and timing offsets between matched flashes refines the matching algorithm further: the number of cases in which one flash is matched to multiple flashes of the second LLS should be reduced. Therefore, the refined algorithm initiates with finer matching criteria, i.e., 1 % of both ds_{match} and dt_{match}. It searches for one element detected by the second LLS that meets the finer criteria for any element of the given flash. Only if no match is found does the allowed distance and time difference increase by 1 % of ds_{match} and dt_{match}, respectively. The process repeats iteratively until either a match is found or the allowed distance (timing offset) exceeds the original ds_{match} (dt_{match}). In the latter case, the algorithm stops and the flash is labeled unmatched (note: the refined analysis is performed for matched flashes only; however, the algorithm can also treat the unmatched flashes). One or more matches for the given flash are still possible because of the discrete increments from one iteration to the following. There might also be flashes within an equal distance and equal time offset to the given flash.

3 Results

The different LLSs detect flashes in different ways and with distinct characteristics. In this section, flash observations are compared and analyzed. As an example, the ISS overpass with the corresponding observations of ISS-LIS, Meteorage and SAETTA in Fig. 1 comprises (almost) the entire study

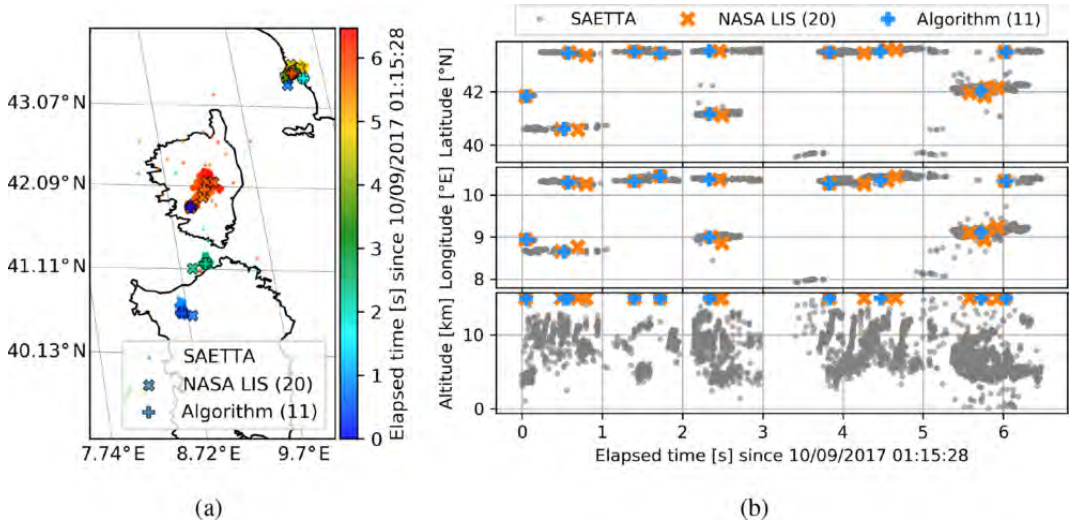


Figure 3. The map **(a)** and time series **(b)** of SAETTA observations and the mean flash positions based on the NASA LIS algorithm and our developed algorithm for one situation on 10 September 2017. The numbers in parentheses in the legend indicate the number of identified flashes (both algorithms analyzed the same LIS events). Colors in **(a)** represent the elapsed time from the initial lightning activity. Note: flash altitudes in **(b)** are not known for ISS-LIS flashes but arbitrarily plotted at 15 km.

region. It lasted 169 s from FOV entering to leaving the region. The effective viewtime per 0.5° × 0.5° grid box is indicated in grayscale in Fig. 1a. Wide parts of the domain have been seen for at least 60 s. Figure 1b additionally shows an IR satellite image indicating the cloud tops. The example of a single flash observed by all three LLSs during this overpass is given in Fig. 1c. SAETTA captures the most detail of the flash structure, and there are significantly more ISS-LIS events than Meteorage pulses and strokes. All but the first Meteorage signals indicate an IC pulse. Since the first stroke is of type CG, the entire flash is characterized as a CG flash.

First, relative DEs of ISS-LIS and Meteorage are elucidated. The comparisons of matched flash location and timing differences are discussed, and finally characteristics of flashes, with a special interest in differences between matched and unmatched flashes, are analyzed.

3.1 Detection comparison

This DE analysis is realized on the flash scale. Flashes were preliminarily identified by our in-house algorithm, which merges ISS-LIS events as well as Meteorage pulses and strokes according to their locations and times of occurrence. Further investigations break the flash scale down into events and pulses and strokes, e.g., for the flash characteristics.

The period of observation spans from 1 March 2017 to 20 March 2018. In total, 330 ISS-LIS flashes and 569 Meteorage flashes are identified by our algorithm.

Besides the DE, the probability of false alarm (POFA) characterizes the quality of detection. Quantifying the POFA requires knowledge about the truth, which is the real number of flashes. SAETTA could provide a reference value to quantify the POFA; however, not all stations have operated continuously for the entire study period. Signals from at least six stations are needed to reconstruct and locate a discharge signal, and 31 of 330 (89 of 569) ISS-LIS (Meteorage) flashes were not detected by SAETTA. SAETTA's detection efficiency and accuracy also decrease with distance to the network's center. A vast majority of more than 90 % of all flashes occurred outside a distance of 100 km from SAETTA's center. There was evidence of VHF activity for the vast majority (95.0 %) of the flashes not reported by SAETTA (missed flashes). In 59.2 % of the cases, at least one and fewer than six stations recorded signals. SAETTA lightning was observed just outside our matching criteria (0.2 °, 0.3 s) for 28.3 % of the missed flashes. The result indicates that the POFA is low for both ISS-LIS and Meteorage (although it cannot be quantified).

Only three (12) of the ISS-LIS (Meteorage) flashes missed by SAETTA are located within 100 km of SAETTA's center. Due to the low total flash number within this close domain to SAETTA, a statistical analysis is ambiguous; 13 of the 15 missed flashes near SAETTA were missed due to station downtimes or filter criteria (spatial, temporal, reduced χ^2; Sect. 2.3). Pédeboy et al. (2018b) reported Meteorage flashes missed by SAETTA with (absolute) peak currents exceeding 100 kA; 2 of the 12 missed Meteorage flashes close to SAETTA exhibit an (absolute) current above 100 kA. SAETTA data can, in fact, not provide the desired true flash numbers, mainly due to station downtimes, and the POFA of Meteorage and ISS-LIS cannot be calculated.

A total of 60.7 % (54) of the Meteorage flashes without concurrent SAETTA sources contain one pulse or stroke only. Those flashes with only one pulse or stroke (or one

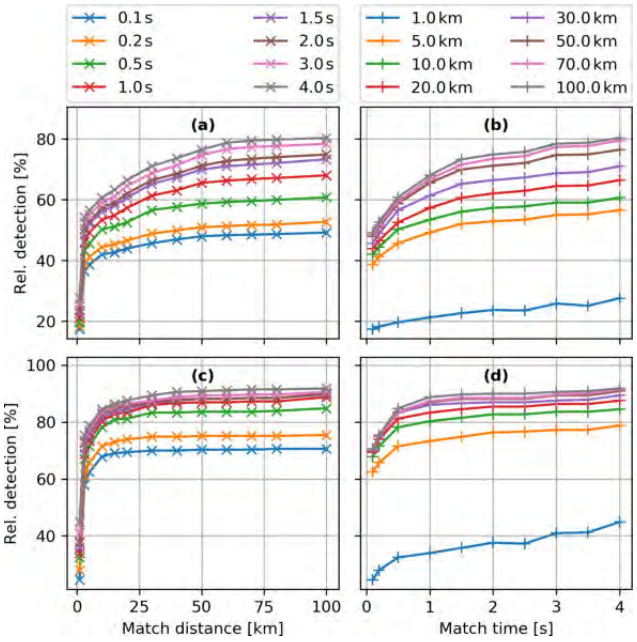


Figure 4. Relative detection efficiency based on constant, equal time criterion dt_{match} (line color) with varying distance ds_{match} for ISS-LIS **(a)** and Meteorage **(c)**. Panels **(b)** and **(d)** are with constant, equal distance ds_{match} (line color) and varying dt_{match} for ISS-LIS and Meteorage, respectively.

event in the case of ISS-LIS) are referred to as single-element flashes. Missing SAETTA observations for a single-element flash might be indicative of a locating and timing problem of the ISS-LIS event or Meteorage pulse and/or stroke (possible false alarm). The DE analysis distinguishes results for the complete dataset and excluded single-element flashes; 316 ISS-LIS and 367 Meteorage flashes remain after excluding the single-element flashes. Thus, the ISS-LIS (Meteorage) flash number is reduced by 14 (202) flashes compared to the overall count. ISS-LIS single-element flashes are rare, while there is a significant amount of Meteorage single-element flashes. The result is related to the differences in optical and LF lightning detection. The entire dataset contains 16 881 ISS-LIS events and 2144 Meteorage pulses and strokes (487 CG, 1657 IC); 15 578 events (92 %) are distributed over the ISS-LIS flashes with a match. For Meteorage, 1439 pulses and strokes (271 CG, 1168 IC) constitute the flashes with matches (67 %). Hence, 55.6 % (70.5 %) of the CG strokes (IC pulses) belong to flashes with coincident ground–space detection. Despite coarser ds_{merge} and dt_{merge} for a Meteorage flash than an ISS-LIS flash, Meteorage observed 239 flashes more than ISS-LIS within similar regions and time frames.

Figure 5a presents a histogram of the total flash detection counts within the four categories introduced in section 2.5. The number of single-element flashes is marked. Additionally, Fig. 5 includes a map of the locations of the flashes within each category. Figure 5b maps the flashes as a 2-D histogram on a $0.1° \times 0.1°$ grid. Flashes are detected all over the study domain for both ISS-LIS and Meteorage without any apparent pattern, even if the number of flashes is not sufficient to be statistically representative.

Table 1 summarizes all relative DEs. Daytime covers the time from 05:00 to 17:00 UTC. Nighttime flashes are defined between 17:00 and 05:00 UTC.

ISS-LIS was able to detect 326 of the 569 recorded Meteorage flashes from 1 March 2017 to 20 March 2018, a relative DE of 57.3 %. If the notable number of Meteorage single-element flashes is neglected, ISS-LIS detected 229 of the remaining 367 Meteorage flashes (62.4 %). ISS-LIS shows a low relative DE of less than 54 % for daytime flashes; 58.7 % of the Meteorage nighttime flashes are detected by ISS-LIS. In particular, the nighttime relative DE cannot reach the literature expectations of over 90 % for LIS (Boccippio et al., 2002). The ISS-LIS relative DE significantly depends on the Meteorage flash type. A flash with at least one CG stroke, referred to as a CG flash, is detected in only 53.5 % of the cases, while a pure IC flash is detected with 59.3 % relative DE. ISS-LIS could detect 68.8 % of the occurring Meteorage IC flashes with at least two pulses. If flashes with at least two pulses and/or strokes are considered, the relative DE of IC flashes surpasses that of CG flashes by almost 14 % and increases compared to the total IC flash relative DE by 9.5 %. Hence, CG flashes and single-pulse IC flashes especially decrease the total DE of ISS-LIS. All relative DEs use ds_{match} of 20 km and dt_{match} of 1.0 s. Finer criteria would further decrease the relative DE of ISS-LIS (higher sensitivity to the criteria than Meteorage; Fig. 4a).

Out of the total 330 ISS-LIS flashes, Meteorage detected 275 (83.3 %). The DE of Meteorage relative to ISS-LIS flashes with at least two events equals 83.9 % (265 of 316 flashes). The relative DE of the VLF–LF Meteorage LLS appears to be significantly higher than in former studies (Sect. 1) using LF LLSs and TRMM-LIS. It is assumed that the ISS-LIS detection efficiency is similar to that of TRMM-LIS in general (Richard J. Blakeslee, personal communication, 2019), and thus Meteorage provides a high-quality LF LLS. Moreover, the Meteorage detection efficiency in particular appears to be quite resistant to changes in ds_{match} and dt_{match}. For example, halving both criteria (10 km in space, 0.5 s in time) results in a relative detection efficiency of about 78 %. More details about the sensitivity to the matching criteria can be found in Sect. 2.5.

Meteorage detected 80.0 % of the 100 ISS-LIS daytime flashes. Its relative DE reaches 84.8 % for 230 ISS-LIS nighttime flashes. The relative DE depends on the performance of the LLS itself but also the performance and locating accuracy relative to the reference LLS. As ISS-LIS detects flashes optically, the influence of different lighting on ISS-LIS daytime and nighttime accuracy is investigated as part of the following section.

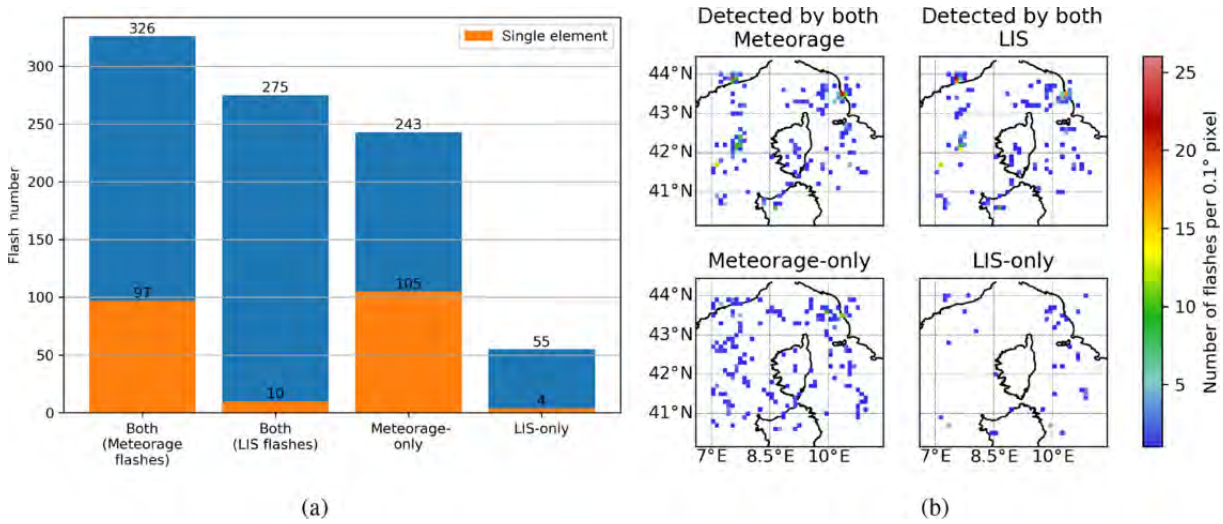


Figure 5. Flash category histogram **(a)** and spatial distribution **(b)** while matching ISS-LIS and Meteorage flashes for the available ISS overpasses from March 2017 to March 2018. Categories show the number of Meteorage flashes seen by ISS-LIS: both (Meteorage flashes), the number of ISS-LIS flashes also detected by Meteorage (both LIS flashes), and flashes detected either by Meteorage (Meteorage-only) or ISS-LIS (LIS-only). The numbers of single-element flashes (event for ISS-LIS, pulse or stroke for Meteorage) are marked for each category as indicated.

Table 1. Relative detection efficiencies (DEs) of Meteorage and ISS-LIS. The values in parentheses give the relative DEs for flashes with at least two elements. The flash numbers (100 %) to calculate the DEs are indicated. Note: ISS-LIS (Meteorage) DE uses Meteorage (ISS-LIS) flash numbers.

	Overall	Daytime	Nighttime	IC flash	CG flash
ISS-LIS DE (%)	57.3 (62.4)	53.9 (60.8)	58.7 (63.0)	59.3 (68.8)	53.5 (55.4)
Meteorage flash number	569 (367)	167 (102)	402 (265)	369 (192)	200 (175)
Meteorage DE (%)	83.3 (83.9)	80.0 (80.2)	84.8 (85.5)	–	–
ISS-LIS flash number	330 (316)	100 (96)	230 (220)	–	–

3.2 Distances and timing offsets between collocated flashes

In this section, the matched ISS-LIS and Meteorage flashes are studied regarding their relative location and time of occurrence. For each element of a flash detected by one LLS the closest (in time or in space, not a combined filter here) element of the matched flash(es) accounts for the statistic. One element can be closest to multiple elements of the second LLS. The entirety of elements of flashes with matches is analyzed statistically. Figure 6 presents the results for distances (a) and timing offsets (b) between events as well as pulses and strokes.

Figure 6a shows histograms of the distance between a given ISS-LIS event (Meteorage pulse and/or stroke) and the closest pulse and/or stroke (event) of a matched flash. The distribution given an ISS-LIS event peaks primarily between 2.50 and 3.00 km and secondarily at about 4.50 km, with a median (mean) of 4.74 km (5.68 km). The distribution given a Meteorage pulse or stroke has a broad maximum from 0.75 to 2.75 km, with a median (mean) of 2.31 km (3.60 km). The Meteorage pulse–stroke distance distribution features a more pronounced (if wider) peak for less distance than the distribution given an ISS-LIS event. This is due to the calculation method and the numbers of available events as well as pulses and strokes. The higher number of (and smaller distance between) ISS-LIS events allows in general for finding a closer event to a given Meteorage pulse or stroke than vice versa. The cumulative distribution functions (CDFs) within the plotted interval (Fig. 6aii) show that the distance distribution given an ISS-LIS event has a larger tail than the distribution given a Meteorage pulse or stroke. The 60th percentile is found at approximately 5.5 and 2.6 km for a given ISS-LIS event and Meteorage pulse or stroke, respectively. Both Meteorage IC pulses and CG strokes exhibit similar distributions to the overall Meteorage pulses and/or strokes (also in the CDFs; Fig. 6aii), with the peak between 0.75 and 2.75 km. The median (mean) distance for IC pulses and CG strokes and their match equals 2.36 km (3.63 km) and 2.22 km (3.51 km), respectively. Hence, CG strokes feature a slightly lower distance to matched events than IC pulses.

Distances are calculated between the closest ISS-LIS events (not groups or flashes as in former studies) and LF pulses and/or strokes. Events provide a finer resolution of the lightning discharge than groups, while group centroids use radiance-weighted event locations and are interpolated to sub-pixel locations. Bitzer et al. (2016), who used TRMM-LIS groups and ENTLN pulses, found for both conditional distributions median (mean) differences in location between 7.0 and 7.2 km (7.6 and 7.9 km) in North America. Those values are in accordance with the distances observed by Rudlosky et al. (2017) between TRMM-LIS flashes and GLD360 strokes. Zhang et al. (2019) thoroughly investigated TRMM-LIS performance and found additional latitudinal location offset resulting from TRMM yaw maneuvers. Their analysis suggests a correction for TRMM-LIS group locations that improves the TRMM-LIS group distance to NLDN pulses and/or strokes for the summers of 2012 and 2013. The distance distribution peaks for 1–2 km after the correction (5–6 km before the correction). The distribution peak distances before the correction were similar to those reported by Bitzer et al. (2016) and Rudlosky et al. (2017). Hence, it is assumed that distances between TRMM-LIS groups and VLF–LF LLS pulses and/or strokes are similar or slightly smaller than distances comparing ISS-LIS events and Meteorage LF pulses and/or strokes in this work. It should be mentioned that the TRMM-LIS pixel size is slightly smaller than that of ISS-LIS, i.e., 4.3 km (3.7 km) nadir after (before) TRMM boost versus 4.5 km nadir.

The optical ISS-LIS sensor might be affected by different lighting. Therefore, the accuracy of ISS-LIS flashes relative to ground-based LLSs is explicitly investigated during day and night (not shown as a figure). Daytime flash distances are concentrated mainly between 2.0 and 5.0 km, and the distribution peaks at about 3.5 km. The ISS-LIS nighttime flash distribution peaks at about 5.5 km of distance to matched Meteorage flashes. Given an ISS-LIS flash, the CDF distribution also rises faster for daytime than for nighttime flash distances. Hence, distances between coincident flashes are in fact smaller during daytime than during nighttime. The comparison of ISS-LIS flashes to SAETTA reveals a small difference of up to 0.05° latitude and longitude during both day- and nighttime. ISS-LIS flashes tend to occur slightly south and west of the corresponding SAETTA observations. The small location difference, considering ds_{match} of 20 km and ISS-LIS spatial resolution of 4.5 km (nadir), does not significantly influence our results. In particular, ISS-LIS maintains its locating accuracy during daytime and during nighttime.

The timing offset subtracts the time of the matched element from the time of the given element. It yields positive and negative values according to which element occurred first, with a positive value indicating that the given element occurred later than its match. Again, the two conditions, which are given an ISS-LIS event and given a Meteorage pulse or stroke, are applied. The resulting distribution (Fig. 6b) peaks between −0.5 and 0.5 ms for a given ISS-LIS event and between −1.0 and 1.0 ms for a given Meteorage pulse or stroke. The distribution tails, with an absolute timing offset longer than 10 ms and up to 1.0 s, are not plotted. They are larger for a given ISS-LIS event than for a given Meteorage pulse or stroke. It is observed that Meteorage pulses and strokes often do not cover the entire duration of a flash. ISS-LIS events reflect the actual flash duration (reference to concurrent SAETTA sources) better than the Meteorage pulses and strokes. Hence, given an ISS-LIS event and looking for a matched Meteorage pulse or stroke, the number of available pulses and/or strokes is often limited. Several events can have the same closest pulse or stroke even if the events occurred in different time frames. It increases the probability of larger timing offsets, especially for a given ISS-LIS event compared to a given Meteorage pulse or stroke. The CDFs (Fig. 6bii) reveal that about 20 % (5 %) of the ISS-LIS events (Meteorage pulses and/or strokes) shown here exhibit timing offsets of more than 2.5 ms. About 20 % (10 %) of ISS-LIS events (Meteorage pulses and/or strokes) have values lower than −2.5 ms. In the overall distribution (not shown), time offsets exceed 10.0 ms for 43 % (22 %) of ISS-LIS events (Meteorage pulses and/or strokes). Negative time offsets exceed −10.0 ms for 25 % (22 %) of ISS-LIS events (Meteorage pulses and/or strokes). The distribution given an ISS-LIS event is slightly skewed towards positive time offsets (given that the ISS-LIS event occurred later than its best match stroke or pulse). The overall median (mean) values yield 2.36 ms (54.60 ms) and −0.00 ms (2.70 ms) given an ISS-LIS event and Meteorage pulse or stroke, respectively. The mean for a given ISS-LIS event is an artifact of the skewed distribution (also in the tails). Considering the ISS-LIS integration frame time of 2.0 ms, the remaining average statistics are close to the temporal accuracy of ISS-LIS. Both conditional distributions given ISS-LIS and given Meteorage show an overall similar shape (Fig. 6b). The matched element, considering both the ISS-LIS and Meteorage distributions, occurs with similar probability earlier or later (or simultaneously) than the element itself, and the distribution peak is centered at zero time offset. This is an interesting finding since, e.g., Höller and Betz (2010) and Bitzer et al. (2016) found that TRMM-LIS detected lightning on average 1 to 2 ms later than the ground-based LLSs. This is not the case for ISS-LIS in our study (and again one must consider the ISS-LIS integration time frame of 2.0 ms), although the order of magnitude of the time offsets agrees well with our results. Timing differences can in fact be directly compared to those studies as the closest event provides the same time as the closest group (groups merge several events within the same time frame and in adjacent pixels of ISS-LIS).

The distribution given an IC pulse is also symmetric around zero and shows a maximum between −1.0 and 1.0 ms (Fig. 6b). Its median (mean) is 0.00 ms (4.29 ms). For CG strokes, however, the distribution peaks between −1.0 ms and 0.0 ms. The negative distribution peak and median (mean) of −0.07 ms (−4.32 ms) indicate that ISS-LIS

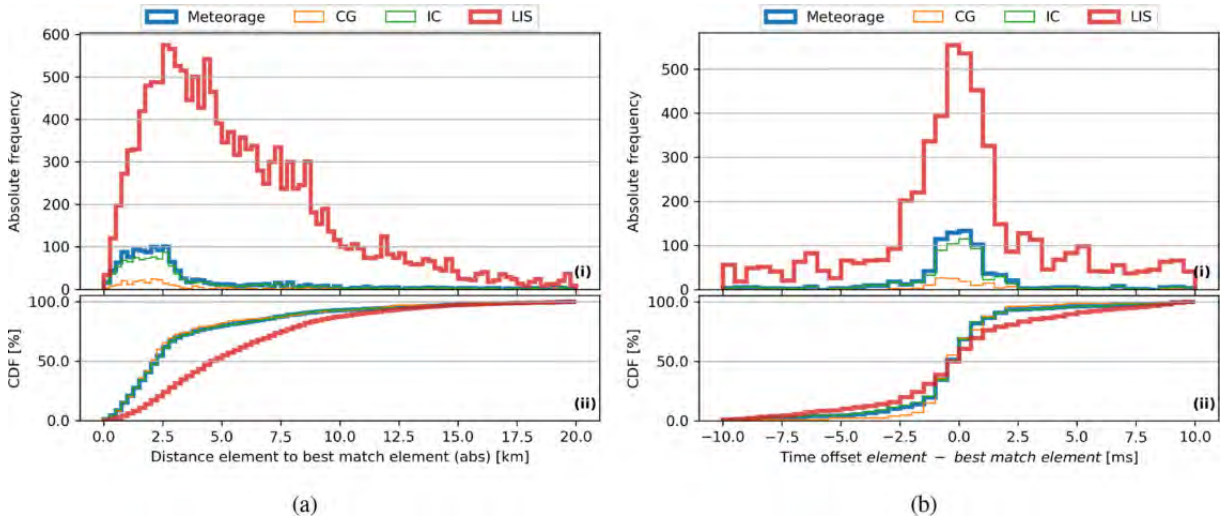


Figure 6. Best match distance **(a)** and time offset **(b)** between a given Meteorage pulse or stroke (LIS event) and the closest ISS-LIS event (Meteorage pulse or stroke). Histogram (i) and cumulative distribution function (ii) with bins of 0.25 km **(a)** and 0.5 ms **(b)**. Analyzed pulses and strokes (events) belong to flashes with matches (spatial and temporal filters), while the pulses and/or strokes (events) of unmatched flashes are not considered. For Meteorage, the discharge types (CG, IC) are distinguished. Only elements with absolute timing offsets of less than 10 ms are included in the plotted time offset distribution. A positive time indicates that the given element occurred later than the best match.

detected CG lightning slightly later than Meteorage. It might account for the time the light of the CG lightning needs to propagate towards the higher parts of the cloud and to become visible from space.

3.3 Characteristics of detected flashes

The previous sections dealt with the relative DEs, location and times of coincident ISS-LIS and Meteorage records. In this section the unmatched flashes (42.7 % Meteorage, 16.7 % ISS-LIS) are also considered to investigate the following flash characteristics: the number of elements (events, pulses and strokes) per flash, flash extent, flash duration, flash mean absolute (pulse–stroke) amplitudes and individual pulse–stroke amplitudes, flash mean (event) amplitude count, flash maximum (event) amplitude count, and the flash mean, minimum and maximum altitudes based on SAETTA observations. They are separated per matched and unmatched flash, per daytime (05:00 to 17:00 UTC) and nighttime (17:00 to 05:00 UTC), and per flash type (IC, CG). The ISS-LIS flash IC or CG attribute depends on the type of the matched Meteorage flash. There is no flash type associated with ISS-LIS-only flashes. As explained in Sect. 2.4, ISS-LIS events are analyzed. The statistical results obtained would be similar using groups instead of events, except for the flash extent and maximum amplitude count per flash. It should be noted that the number of daytime CG flashes is particularly limited (24 ISS-LIS and 42 Meteorage, meaning < 10 %). Flash extents add the north–south and the east–west distance of a flash. The north–south distance uses the maximum and minimum latitude of the flash elements. The east–west distance of a flash is defined as the distance between the longitudinal maximum and minimum of the elements at the mean latitude (as that distance also depends on the latitude). Flash durations, or the times from the first to the last element of a flash, are only limited by the viewing time of ISS-LIS. Theoretically, one flash could last for up to 90 s. Meteorage flash durations are not limited.

Statistics for ISS-LIS flash characteristics are summarized in Table 2. It contains the overall average and the averages, minima and maxima observed for matched and unmatched ISS-LIS flashes. ISS-LIS flashes have on average 51.2 events. Matched ISS-LIS flashes had more than twice as many events as the unmatched flashes (56.6 versus 23.7). The detailed event number distributions for matched and unmatched ISS-LIS flashes are shown in Fig. 7a and b, respectively. They include the histogram (i) and the CDF (ii). Daytime and nighttime flashes are distinguished for the flash types (only for matched flashes). The histogram bars add the numbers of the different categories for the corresponding bin. All following figures make use of the same layout. ISS-LIS nighttime flashes have about two times more events than daytime flashes. The background subtraction threshold for the optical signal is usually greater during daytime than during nighttime, and the sensor acquisition is less sensitive during the day (minimum event amplitude count of 10.0 and 9.0 during daytime and nighttime, respectively). This influences the number of events per flash, with a relative reduction of event numbers on bright backgrounds (daytime) compared to dark backgrounds (nighttime). For example, 16.3 % (2394 of 14 710) of nighttime events observed in this study have an

amplitude count of 9.0 (value less than the daytime threshold). Additionally, ISS-LIS CG flashes comprise on average approximately 11 % more events than IC flashes.

In accordance with the difference in event numbers, the matched ISS-LIS flashes feature a larger extent and longer duration than the unmatched ISS-LIS flashes (Table 2). ISS-LIS flash extents range from 0 km (single events) to about 92 km, as shown in the distributions in Fig. 8. Peterson et al. (2017), who studied the evolution and structure of extreme flashes observed by TRMM-LIS, found an LIS flash with a maximum event separation of 162 km. This size likely results from an elongation due to scattering of optically bright discharges.

This work found coincident Meteorage flashes for all but four ISS-LIS flashes with extents exceeding 40 km. ISS-LIS nighttime flashes are on average 13.2 km larger than daytime flashes (and comprise more events). The result is likely caused by the higher background subtraction threshold during the day than during the night. It could also result from an optical elongation of nighttime flashes. Large flashes with the maximum event separations in Peterson et al. (2017) also occurred at nighttime, but the groups of these flashes were not separated by a significant fraction of the event separation. Fundamentally different cloud structures or types during day and night might also influence the results. We would need additional information, e.g., measuring infrared brightness temperatures for the cloud tops, to verify this hypothesis. Referring to the flash types, the mean extent of ISS-LIS CG flashes is 6.1 km longer than for ISS-LIS IC flashes; however, the longest ISS-LIS flash is a nighttime IC flash.

One observed ISS-LIS flash lasted about 1.7 s (a CG nighttime flash), the longest duration found in this study. Peterson et al. (2017) found spurious flash durations up to 28 s in convective clouds, which result from high flash rates and slow storm motion. One large propagating flash lasted 5.04 s in their study. Figure 9 presents the duration distributions for the matched (a) and unmatched (b) ISS-LIS flashes. Matched ISS-LIS flashes last on average almost twice as long as ISS-LIS-only flashes, i.e., 0.35 s versus 0.20 s (also Table 2). About 10 % (25 %) of the matched (unmatched) flashes were recorded during a single LIS frame. Long-lasting flashes (duration longer than 0.5 s) were detected by both LLSs with high probability (92.6 %). ISS-LIS nighttime flashes last statistically 0.1 s longer than the daytime flashes. The result is in accordance with the ISS-LIS higher relative DE and more detected events at nighttime than during the day.

The averages, minima and maxima of Meteorage flash characteristics are summarized in Table 3. Meteorage flashes contain on average 3.8 (but up to 54) strokes and/or pulses. The distributions of pulse and stroke numbers per matched and unmatched flash are presented in Fig. 10a and b, including the stacked histogram (i) and the CDFs (ii) as explained for Fig. 7. Meteorage flashes seen by ISS-LIS are composed of 4.4 pulses and/or strokes on average. Meteorage-only flashes contain 2.9 pulses and/or strokes on average; 29.8 % of Meteorage flashes with a coincident ISS-LIS flash have only one pulse or stroke (10 CG, 87 IC). Single-element flashes constitute about 43.2 % of the unmatched Meteorage flashes (15 CG, 90 IC). As for ISS-LIS flashes, Meteorage flashes with a match not only contain more pulses and/or strokes, but also extend and last longer than the unmatched flashes (Table 3). The flash extent distributions in Fig. 11a and b show a mean (maximum) of 12.1 km (147.5 km) and 6.9 km (109.2 km) for matched and unmatched flashes, respectively. ISS-LIS detected all IC Meteorage flashes with extents above 32 km. The longest flashes are categorized as CG nighttime. In general, Meteorage CG flashes extend further than IC flashes. The mean extent equals 18.2 km (11.6 km) and 9.2 km (3.9 km) for matched (unmatched) CG and IC flashes, respectively. This is particularly small for unmatched IC flashes (as ISS-LIS can detect the longer IC flashes).

Meteorage flash durations support the findings, with matched flashes lasting on average (maximal) 0.22 s (2.3 s) and unmatched flashes lasting on average (maximal) 0.11 s (1.0 s). Figure 12a and b provide the duration distributions for Meteorage flashes. Distributions of both matched (a) and unmatched (b) flashes are sharply peaked for flashes shorter than 0.05 s (first bin; including single-element flashes, maximum of 13 pulses and/or strokes per flash). The CDF (Fig. 12aii and bii) illustrates that Meteorage CG flashes (mean 0.28 s) last statistically longer than Meteorage IC flashes (mean 0.11 s).

To conclude, matched flashes contain more elements, are more extended and last longer than unmatched flashes for both ISS-LIS and Meteorage records. Meteorage flashes appear to be on average both smaller in extent and shorter in duration than ISS-LIS flashes. The finding is in accordance with the different expectations for optical (LIS) and LF (Meteorage) signals.

Seven (two ISS-LIS, five Meteorage) exceptionally long flashes (extent > 90 km or duration > 1.5 s) are analyzed using concurrent SAETTA observations. The VHF sources highlight the fact that there can be concurrent flashes that either merge and form one flash or propagate at different height levels. ISS-LIS and Meteorage detect both types as continuous flashes as the LLSs capture the flashes two-dimensionally. They can, in particular, not distinguish the different altitudes for the latter.

Then, the VHF SAETTA LLS is used to determine the altitude range of each ISS-LIS and Meteorage flash. The flash mean altitude, the flash minimum altitude and the flash maximum altitude (as defined in Sect. 2.3) constitute three additional flash characteristics. Since not every ISS-LIS and Meteorage flash was detected by SAETTA, flash numbers are reduced compared to the ones discussed in Sect. 3.1 to 256 ISS-LIS flashes with a match, 43 ISS-LIS-only flashes, 292 Meteorage flashes with a match and 188 Meteorage-only flashes. ISS-LIS mean event amplitude count and Meteorage mean pulse–stroke amplitude distributions are examined for

Table 2. Overall average and the average, minimum and maximum values of ISS-LIS flash characteristics for matched and unmatched flashes. The minimum (maximum) altitude represents the 10th (90th) percentile of concurrent SAETTA sources. Amplitude count is recorded per event.

	Average	Average matched	Average unmatched	Minimum matched	Minimum unmatched	Maximum matched	Maximum unmatched
Number of events	51.2	56.6	23.7	1	1	518	116
Extent (km)	27.9	29.5	19.8	0.0	0.0	92.3	57.8
Duration (s)	0.32	0.35	0.20	0.0	0.0	1.69	0.89
Mean altitude (km)	8.2	8.2	8.2	2.9	3.6	10.9	11.6
Minimum altitude (km)	6.3	6.2	6.7	1.9	2.2	9.6	10.5
Maximum altitude (km)	9.9	10.0	9.6	2.9	4.6	12.8	12.9
Mean amplitude count	17.9	18.2	16.0	9.0	9.0	40.2	32.5
Maximum amplitude count	50.6	53.3	36.9	9.0	9.0	127.0	88.0

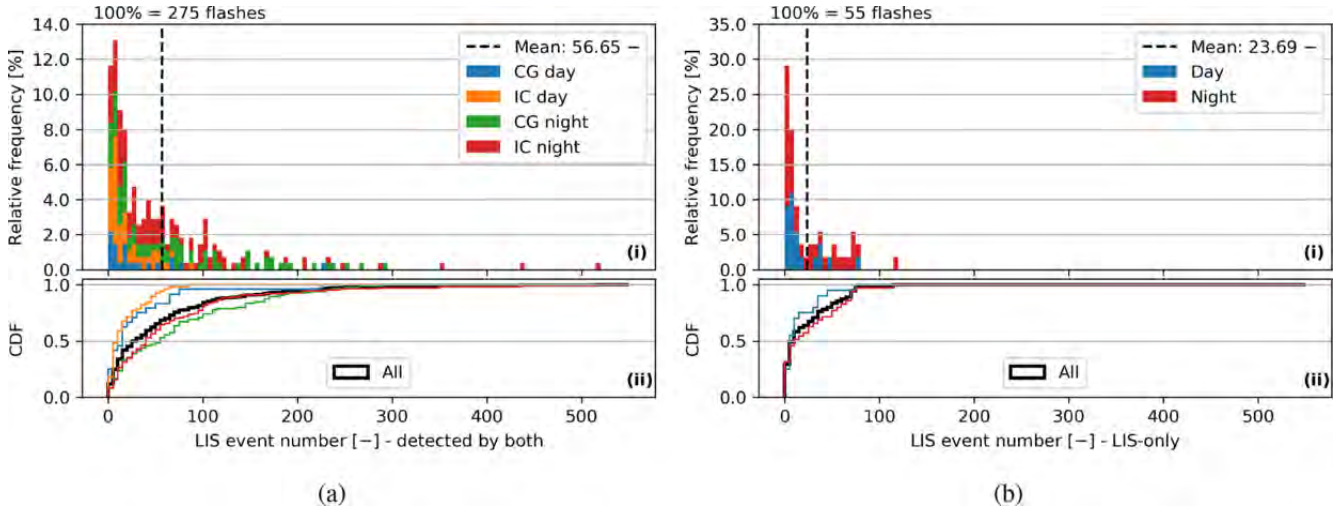


Figure 7. LIS event numbers of ISS-LIS flashes with a coincident Meteorage flash **(a)** and unmatched ISS-LIS flashes **(b)**. Daytime, nighttime and flash type, IC and CG, are indicated by the colors. The histogram (i) and corresponding CDF (ii) use the same colors. The CDF additionally shows a black curve for all data. The histogram bin width is constant at five events. Note: the CG–IC attribute for an ISS-LIS flash needs the matched Meteorage flash and does not exist for ISS-LIS-only flashes. The mean value is plotted as a dashed line. The total number of flashes is indicated above the histogram.

the analyzed altitude levels of flashes. The lowest altitude of detectable VHF sources increases with distance to the LMA network, mainly due to Earth's curvature and also due to shading by the relief, especially in the south of the domain (Coquillat et al., 2019a). Hence, flash minimum (and mean) altitudes might suffer from undetected VHF sources at low altitudes.

Figure 13 presents the mean altitude of matched (a) and unmatched (b) ISS-LIS flashes as histograms (i) and CDFs (ii). The distributions of the flash mean amplitude counts in each altitude bin are included as a blue box plot (with the mean marked as a diamond; outliers not plotted). The distribution of mean altitudes for unmatched ISS-LIS flashes fits that of the matched ISS-LIS flashes (although the number of unmatched flashes is low). The mean flash altitudes average about 8.2 km (Table 2).

The overall ISS-LIS flash mean altitude distribution, which is dominated by 83.3 % flashes with a match, peaks at about 9.5 km, as shown in the histogram in Fig. 13ai. The daytime distribution has a second mode near 5.0 km of altitude. ISS-LIS flashes reach on average altitudes of 9.9 km and were observed up to almost 13 km of altitude (Table 2) (a noteworthy high value considering the tropopause at 10 to 12 km of altitude).

Differences between matched and unmatched ISS-LIS minimum flash altitudes are approximately 0.5 km, with matched flashes showing lower minima (distributions not shown). The difference is significant as it exceeds the predicted SAETTA altitude error (about 0.2 km over wide parts of the domain); 89.7 % of the 126 ISS-LIS flashes with minima less than (or equal to) 6.0 km of altitude have a coincident Meteorage flash. ISS-LIS flashes with minima above

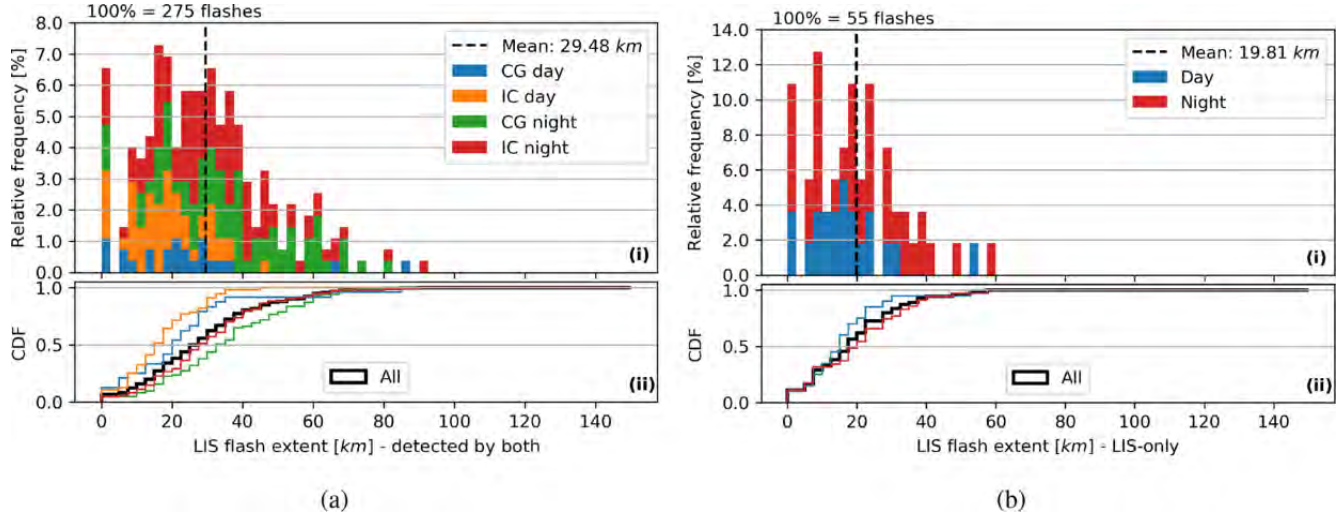


Figure 8. Flash extents of ISS-LIS flashes with a coincident Meteorage flash **(a)** and unmatched ISS-LIS flashes **(b)**. Daytime, nighttime and flash type, IC and CG, are indicated by the colors. The histogram (i) and corresponding CDF (ii) use the same colors. The CDF additionally shows a black curve for all data. The histogram bin width is constant at 2.5 km. Note: the CG–IC attribute for an ISS-LIS flash needs the matched Meteorage flash and does not exist for ISS-LIS-only flashes. The mean value is plotted as a dashed line. The total number of flashes is indicated above the histogram.

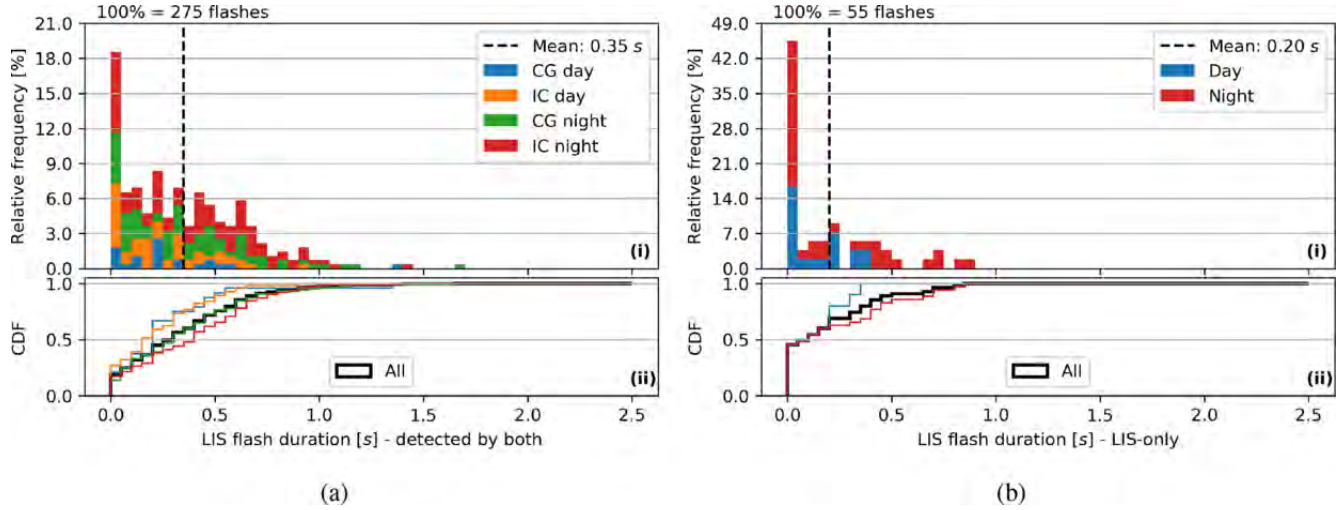


Figure 9. Flash duration of ISS-LIS flashes with a coincident Meteorage flash **(a)** and unmatched ISS-LIS flashes **(b)**. Daytime, nighttime and flash type, IC and CG, are indicated by the colors. The histogram (i) and corresponding CDF (ii) use the same colors. The CDF additionally shows a black curve for all data. The histogram bin width is constant at 0.05 s. Note: the CG–IC attribute for an ISS-LIS flash needs the matched Meteorage flash and does not exist for ISS-LIS-only flashes. The mean value is plotted as a dashed line. The total number of flashes is indicated above the histogram.

6.0 km (173) are detected by Meteorage in 82.7 % of the cases. Overall, Meteorage better detected low-altitude ISS-LIS flashes than ISS-LIS flashes restricted to middle and high levels.

The amplitude count of ISS-LIS flashes increases in general with the mean altitude (Fig. 13ai and bi). The highest observed flash mean altitudes occur mainly for pure IC flashes and show statistically high-amplitude counts. They likely originate within high-reaching convective clouds like cumulus congestus and cumulonimbus. Similar results regarding the amplitude count distributions were identified for the ISS-LIS flash maximum altitude distributions and maximum event amplitude count per flash at different altitudes (not shown). Matched and unmatched ISS-LIS flashes feature almost similar mean amplitude counts (Table 2). The overall brightest event (127.0) occurred during nighttime for a matched flash. The strongest optical signal during the day (105.0) is also recorded within a matched flash. Accordingly,

Table 3. Overall average and the average, minimum and maximum values of Meteorage flash characteristics for matched and unmatched flashes. The minimum (maximum) altitude represents the 10th (90th) percentile of concurrent SAETTA sources. The maximum pulse–stroke amplitude is the maximum current observed for the flash (negative or positive).

	Average	Average matched	Average unmatched	Minimum matched	Minimum unmatched	Maximum matched	Maximum unmatched
Number of pulses and/or strokes	3.8	4.4	2.9	1	1	54	26
Extent (km)	9.9	12.1	6.8	0.0	0.0	147.5	109.2
Duration (s)	0.17	0.22	0.11	0.0	0.0	2.32	0.98
Mean altitude (km)	7.6	8.1	6.7	2.7	2.1	11.2	11.0
Minimum altitude (km)	5.7	6.1	5.1	1.7	1.6	10.3	9.4
Maximum altitude (km)	9.2	9.8	8.2	3.2	2.1	12.8	12.9
Mean absolute amplitude (kA)	9.5	8.0	11.6	1.3	1.1	98.1	102.9
Maximum pulse–stroke amplitude (kA)	−4.8	−2.2	−8.2	−149.0	−144.0	150.0	128.5

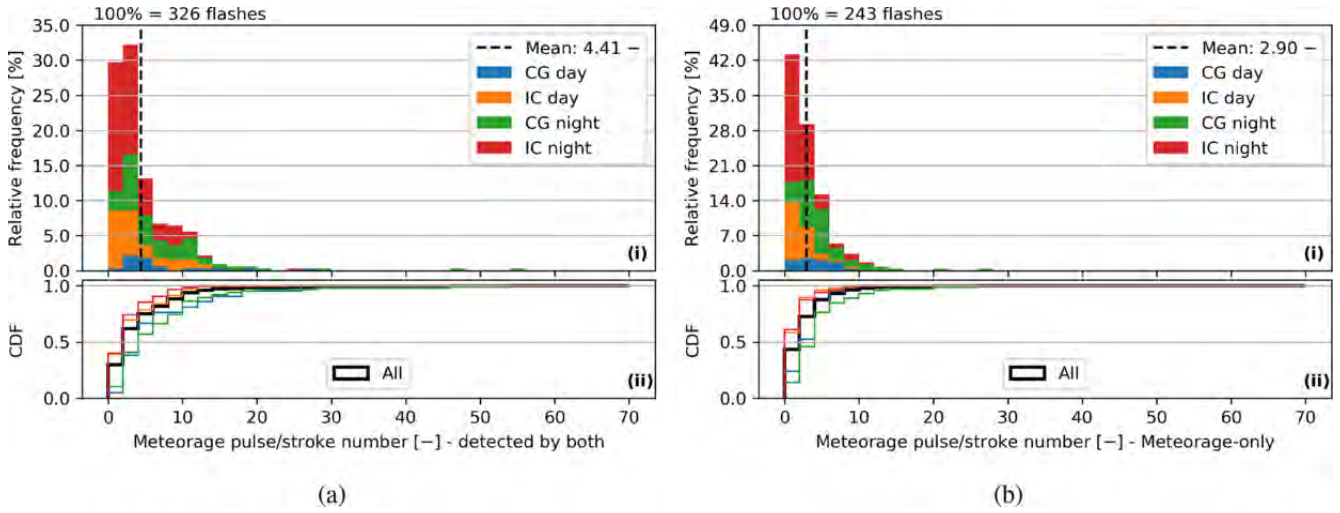


Figure 10. Pulse–stroke number of Meteorage flashes with a coincident ISS-LIS flash **(a)** and unmatched Meteorage flashes **(b)**. Daytime, nighttime and flash type, IC and CG, are indicated by the colors. The histogram (i) and corresponding CDF (ii) use the same colors. The CDF additionally shows a black curve for all data. The histogram bin width is constant at two pulses and/or strokes. The mean value is plotted as a dashed line. The total number of flashes is indicated above the histogram.

maximum event amplitude counts averaged over all flashes of about 53.3 and 36.9 characterize matched and unmatched flashes, respectively (Table 2). Flashes containing the optically brightest events have a higher chance of producing significant LF signals and being detected by Meteorage than the optically darker flashes.

The comparison of altitudes of Meteorage flashes with and without ISS-LIS matches aims to study how ISS-LIS can detect low-altitude flashes. The flash mean (absolute) amplitude and maximum pulse–stroke amplitude per flash are additionally analyzed for each altitude bin in the histograms. The maximum amplitude per flash can either show a positive or negative current. Results are presented in Fig. 14 for the altitude maximum with the mean absolute amplitude per flash. Mean flash altitudes average 8.1 km for Meteorage flashes with a match (Table 3). They are on average 1.4 km lower for the Meteorage-only flashes. The mean altitude of matched flashes is similar to that of ISS-LIS matched flashes (Fig. 13a). The unmatched flashes, however, differ by about 1.5 km of altitude (difference well above the SAETTA altitude uncertainty). Meteorage flash maximum altitudes confirm this result (Fig. 14): flashes with a coincident ISS-LIS flash reach on average 9.8 km of altitude. The Meteorage-only flashes feature a maximum altitude of 8.2 km on average. The maximum altitude distribution peaks, as for the ISS-LIS matched flashes, at about 11.0 km of altitude (Fig. 14). For the Meteorage-only flashes, another mode exists between 6.5 and 7.0 km of altitude. The daytime distribution of unmatched Meteorage flashes peaks at the lower altitudes. It is indicative of the ISS-LIS reduced DE for low-altitude flashes (during all times and even more pronounced in daytime than during the night). Meteorage flashes with maxima exceeding 10.0 km (248) are detected by ISS-LIS in 75.4 % of the cases. The ISS-LIS relative DE for Meteorage flashes with maxima

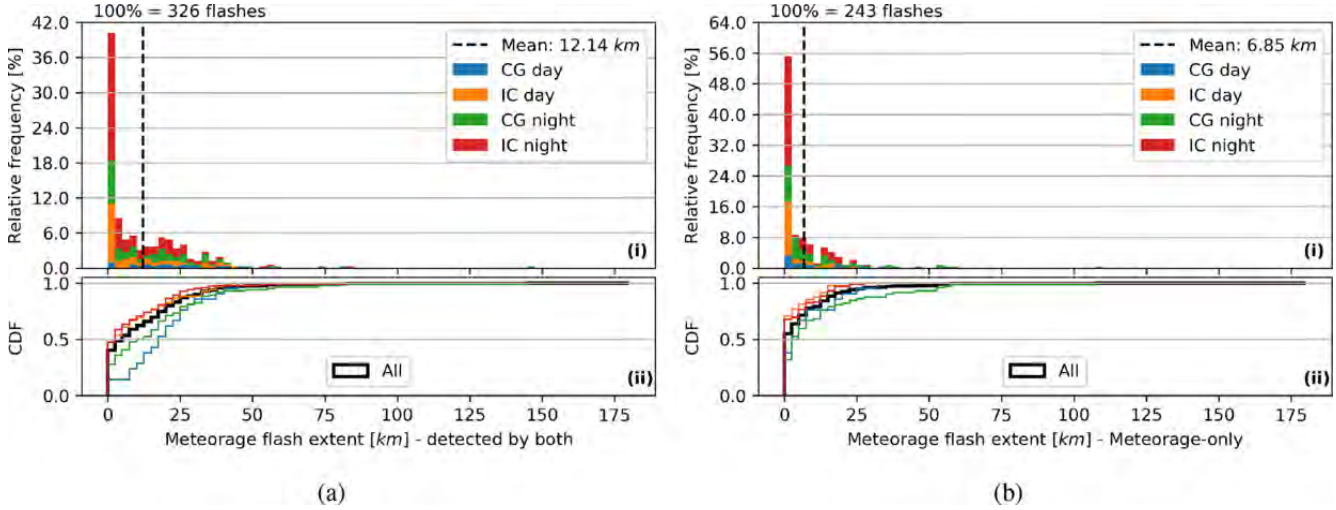


Figure 11. Flash extent of Meteorage flashes with a coincident ISS-LIS flash **(a)** and unmatched Meteorage flashes **(b)**. Daytime, nighttime and flash type, IC and CG, are indicated by the colors. The histogram (i) and corresponding CDF (ii) use the same colors. The CDF additionally shows a black curve for all data. The histogram bin width is constant at 2.5 km. The mean value is plotted as a dashed line. The total number of flashes is indicated above the histogram.

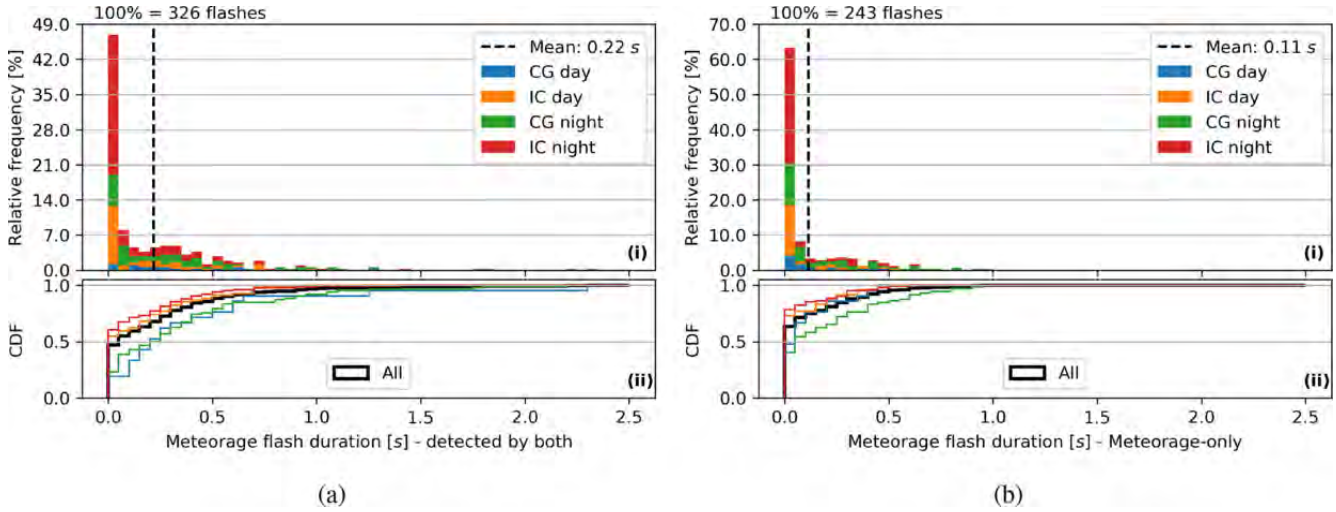


Figure 12. Meteorage flash duration with a coincident ISS-LIS flash **(a)** and unmatched Meteorage flashes **(b)**. Daytime, nighttime and flash type, IC and CG, are indicated by the colors. The histogram (i) and corresponding CDF (ii) use the same colors. The CDF additionally shows a black curve for all data. The histogram bin width is constant at 0.05 s. The mean value is plotted as a dashed line. The total number of flashes is indicated above the histogram.

lower than (or equal to) 10.0 km (232) is only 45.3 %. This trend still influences the flash mean and minimum altitudes (Table 3). Hence, it is confirmed that ISS-LIS flash detection declines from high- to low-altitude flashes. The result agrees with the case study of Thomas et al. (2000), who found significantly less skill of TRMM-LIS for (CG) discharges near the cloud base than for lightning channels propagating to near the top of the clouds.

Figure 15 shows the distribution of minimum flash altitudes with the maximum (pulse–stroke) amplitude per flash in each altitude bin. Low-altitude flashes (minimum altitudes below 5.0 km) feature statistically higher flash mean (not plotted) and maximum amplitudes than flashes occurring above 5.0 km of altitude (Fig. 15). Those flashes are mainly identified as CG flashes. The analysis of the flash maximum amplitude shows that those low-altitude flashes are dominated by negative maximum currents. The flashes with minimum altitudes above 5.0 km exhibit statistically more positive than negative maximum currents. Further investigation reveals that about 94 % of the (absolute) currents above 22.5 kA belong to CG strokes. The strongest currents reach up to 150.0 kA (both negative and positive currents)

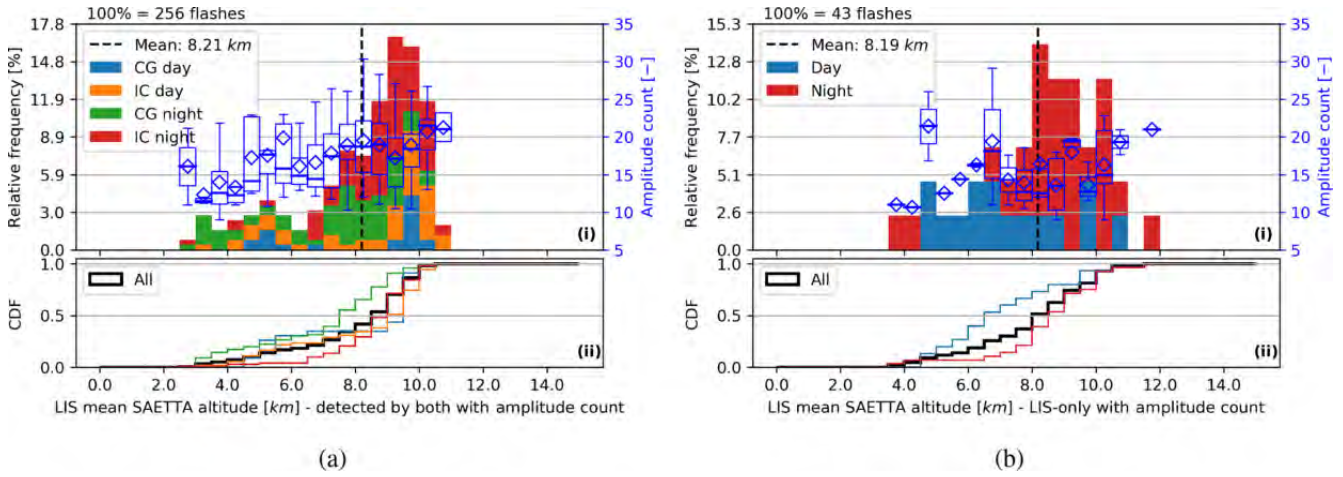


Figure 13. Flash mean altitude of ISS-LIS flashes (from concurrent SAETTA observations) with a coincident Meteorage flash **(a)** and unmatched ISS-LIS flashes **(b)**. Daytime, nighttime and flash type, IC and CG, are indicated by the colors. The histogram (i) and corresponding CDF (ii) use the same colors. The CDF additionally shows a black curve for all data. The histogram bin width equals 0.5 km. Note: the CG–IC attribute for an ISS-LIS flash needs the matched Meteorage flash and does not exist for ISS-LIS-only flashes. The mean value is plotted as a dashed line. The total number of flashes is indicated above the histogram. The blue box plots (median as a line, mean as a diamond, interquartile range – IQR – as a box, 1.5 IQR as whiskers; outliers not plotted) represent the distributions of the ISS-LIS mean event amplitude count per flash for each altitude bin.

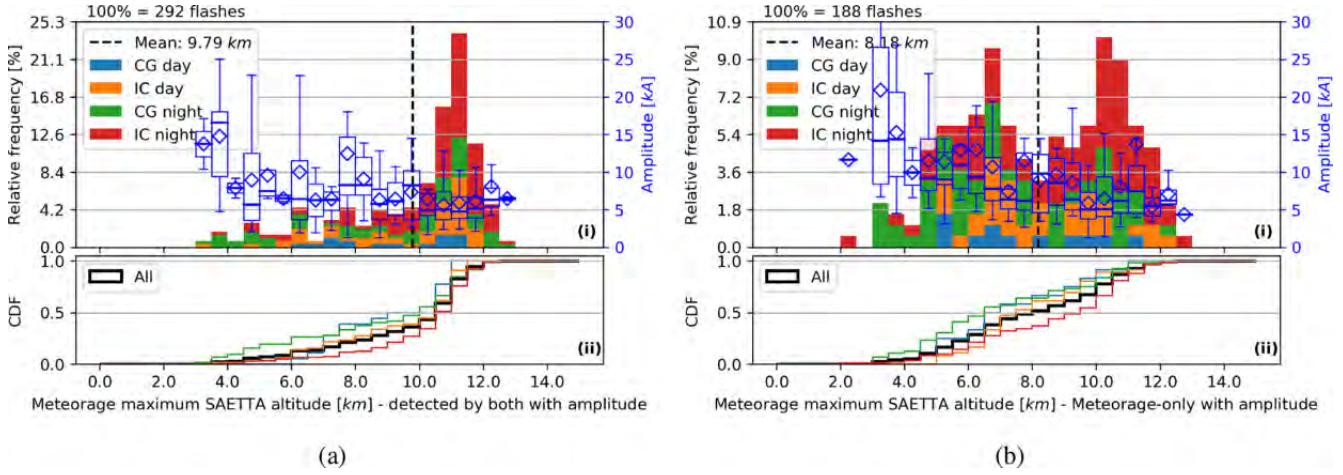


Figure 14. Flash maximum altitude of Meteorage flashes (from concurrent SAETTA observations) with **(a)** and without **(b)** a coincident ISS-LIS flash. Daytime, nighttime and flash type, IC and CG, are indicated by the colors. The histogram (i) and corresponding CDF (ii) use the same colors. The CDF additionally shows a black curve for all data. The histogram bin width equals 0.5 km. The mean value is plotted as a dashed line. The total number of flashes is indicated above the histogram. The blue box plots (median as a line, mean as a diamond, IQR as a box, 1.5 IQR as whiskers; outliers not plotted) represent the distributions of the Meteorage mean absolute amplitude per flash for each altitude bin (scale fixed range from 0 to 30 kA).

and are related to CG strokes. The vast majority (90.6 %) of the CG strokes have negative currents in this study. IC pulse currents do not exceed 50 kA. About 90 % of pulses and/or strokes with an amplitude below 10.0 kA are IC pulses. A similar result is provided by Cummins and Murphy (2009). They found that 90 % of positive LF currents with less than 10.0 kA belong to IC pulses. Negative currents are observed for approximately 26 % of the IC pulses.

The Meteorage mean (maximum) flash absolute amplitude equals 8.0 kA (13.2 kA) and 11.6 kA (18.1 kA) for matched and unmatched flashes, respectively. The difference between matched and unmatched flashes is attributed to some low- to mid-level flashes producing strong currents and not being detected by ISS-LIS (compare panels a and b in Figs. 14 and 15). However, the overall distributions of absolute flash

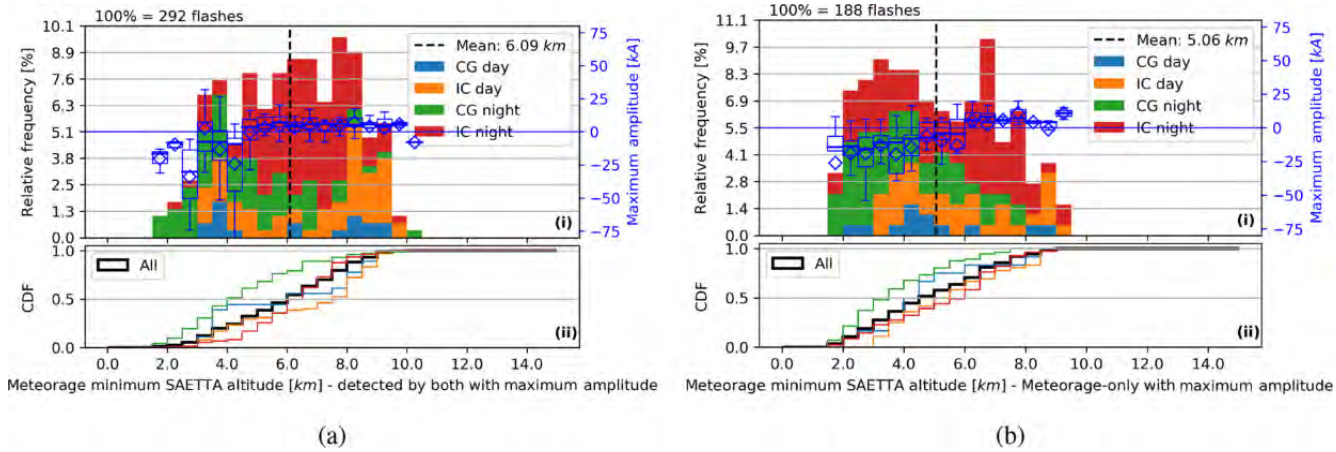


Figure 15. As Fig. 14 for the minimum altitude of Meteorage flashes with **(a)** and without **(b)** a match. Here, the blue box plots represent the distributions of the Meteorage maximum (pulse–stroke) amplitude per flash (positive or negative currents) for each altitude bin.

amplitudes appear to be similar for matched and unmatched Meteorage flashes.

Flashes observed in this study show a statistical relationship between the polarity of the maximum (LF) current and the altitude. The relationship was detailed for the flash minimum altitudes and also appears for the flash maximum altitudes. In this study, flashes with maximum altitudes below 10.0 km exhibit mainly negative maximum currents. As it was found that the ISS-LIS DE is 30 % higher for flashes with maximum altitudes above 10.0 km than for flashes restricted to lower levels, the polarity of the flash maximum current might provide the first information on whether a flash is detected by ISS-LIS. This finding is probably specific for the storm types and flashes analyzed in this study (and region). The observed relationship between the polarity of the maximum current of a flash and its altitude might change for inverted-polarity storms or hybrid (IC+CG) flashes.

4 Conclusions

This study compares the results of the LF ground-based Meteorage LLS, the satellite sensor ISS-LIS and the VHF ground-based LMA SAETTA. The study domain is bounded to a region near Corsica in the Mediterranean Sea where SAETTA data are available. As ISS-LIS has been operating since March 2017, the period is confined to about 1 year from 1 March 2017 to 20 March 2018.

A new algorithm is developed to group ISS-LIS events as well as Meteorage pulses and strokes to flashes. The algorithm is validated using concurrent SAETTA observations and the results of the existing NASA LIS algorithm.

ISS-LIS detected in total 16 881 events distributed over 330 flashes during its overpasses of the study domain. Meteorage data are filtered for the times of ISS overpasses. It contains 2144 pulses and/or strokes (487 CG, 1657 IC) in 569 flashes. ISS-LIS detected about 57.3 % of the Meteorage flashes. Cloud-to-ground (CG) flashes and single-pulse intracloud and cloud-to-cloud (IC) flashes especially decrease the overall relative detection efficiency (DE) of ISS-LIS. A relative DE of 53.9 % was observed for flashes detected by Meteorage during daytime. LIS detected Meteorage IC flashes with about a 6 % higher relative DE than CG flashes. The LF Meteorage LLS was able to detect more than 80 % of all occurring ISS-LIS flashes.

Distances and timing offsets between matched ISS-LIS and Meteorage flashes are analyzed. A mode (median) distance (given a Meteorage flash) of about 1.8 km (2.3 km) indicates a fairly accurate collocation of the flashes. Given an ISS-LIS flash, the mode (median) distance equals about 3.0 km (4.7 km). The majority of distances between matched flashes are within the ISS-LIS pixel resolution (4.5 km nadir, 6.2 km at the edge of the field of view). The absolute timing offset distribution between a given Meteorage flash and the matched ISS-LIS flash is sharply peaked for less than 1.0 ms. Considering the ISS-LIS frame integration time of 2.0 ms, this is a very satisfying result. An analysis of the closest elements (events and pulses and/or strokes) reveals that, with similar probability, ISS-LIS or Meteorage detected lightning first, while the mode timing offsets remain within the LIS frame integration time. For CG strokes, however, ISS-LIS tended to detect the lightning activity later than Meteorage. All offsets increase relatively from the distribution given a Meteorage flash to the distribution given an ISS-LIS flash.

This finding is likely caused by the significantly lower number of pulses and strokes than the number of events. Thus, it is more likely to find an event close to a pulse or stroke than vice versa.

For an enhanced understanding of the flash detection by ISS-LIS and Meteorage, characteristics of the flashes are investigated. In accordance with, e.g., Rudlosky et al. (2017) and Zhang et al. (2019), the probability of a match increases with a larger flash extent and flash duration. A matched flash extended on average almost twice as wide and lasted twice as long as a flash not seen by both ISS-LIS and Meteorage. In a similar manner, the matched flashes contained on average twice the number of elements as a flash observed by only one of the LLSs. ISS-LIS is sensitive to optical signals, while Meteorage detects LF signals of electrical discharges. Nevertheless, ISS-LIS flashes with at least one very bright event were more likely to be detected by Meteorage than optically darker flashes. Using the 3-D lightning location of concurrent SAETTA observations, ISS-LIS and Meteorage flash altitudes are compared. Altitude-related behaviors are likely driven by the range of IC and CG flash altitudes. Detailed results for the flash types are given in Sect. 3.3. Matched flashes of both ISS-LIS and Meteorage feature similar mean altitudes near 8.2 km on average. Unmatched Meteorage flashes occurred on average 1.4 km lower than Meteorage flashes seen by ISS-LIS. The maximum altitude of a flash significantly influenced the detectability by ISS-LIS (compare, e.g., Thomas et al., 2000). Meteorage flashes with maxima exceeding 10.0 km of altitude were detected by ISS-LIS in 75.4 % of the cases. The ISS-LIS relative DE for Meteorage flashes with maxima lower than 10.0 km of altitude is only 45.3 %. The Meteorage flash detection depended slightly on the flash minimum altitude; 89.7 % of the ISS-LIS flashes with minima less than 6.0 km of altitude had a coincident Meteorage flash. ISS-LIS flashes with minima above 6.0 km of altitude had a coincident Meteorage flash in 82.7 % of the cases.

Further investigation revealed that the optical brightness of ISS-LIS flashes is somewhat correlated with the flash altitude, with increasing (both mean and maximum event) amplitude counts for increasing flash altitudes. Meteorage amplitudes increased statistically with decreasing flash altitudes. The polarity and the current of the strongest pulse or stroke within a Meteorage flash showed particular potential to gain qualitative flash altitude information. Flashes with maximum currents of -10 kA or lower remained mainly below 10.0 km of altitude. As stated earlier, the ISS-LIS relative DE was 30 % higher for those flashes than for flashes with maximum altitudes above 10.0 km. This finding will need additional proof, but it can be useful for mimicking satellite lightning products using LF LLSs.

This study analyzes satellite observed lightning over an extratropical region and compares the observations to ground-based LLSs. Our results, including the statistics, use about 1 year of data within the limited region around Corsica island. This results in a limited number of lightning cases. The limited region enables the direct unique comparison of not only ISS-LIS and LF Meteorage but also the VHF SAETTA LLS. Hence, ISS-LIS and Meteorage flash detection is investigated in more detail, e.g., considering the concurrent SAETTA lightning source altitudes. The coincidences between ISS-LIS and Meteorage flashes do not always have a one-to-one correspondence. It is, in addition, an artifact of the relatively coarse match constraints of 20.0 km in space and 1.0 s in time. The constraints are validated and their influence on the results is seen in the matched distance and timing offset distributions. It should be mentioned that the available ISS-LIS data are the provisional P0.2 version for this work, which represents close to but not quite the fully validated data of ISS-LIS. Due to our limited number of cases, all ISS-LIS data are treated in the same way independent of the position within the ISS-LIS field of view (FOV). It is known that the ISS-LIS pixel (event) resolution and the DE decrease near the edge of the FOV. However, it was decided not to filter or reduce the observed cases further in order to allow for a statistical analysis. Our method can be applied to geostationary satellite LLSs, i.e., GLM and the future MTG-LI, and the comparison of their observations to ground-based LLSs. It is planned to study GLM and NLDN lightning observations in America using our methodology. The geostationary satellite observes one region continuously, and thus there will be many more cases for the statistics. The results might be compared to our results of the comparison of ISS-LIS and Meteorage.

Data availability. ISS-LIS provisional science data are available via NASA and HyDRO Search at the following DOI: https://doi.org/10.5067/LIS/ISSLIS/DATA204 (Blakeslee et al., 2017). Fully validated ISS-LIS data are provided by NASA and HyDRO Search. SAETTA data are available to members of HyMeX on the HyMeX website and can be provided on demand. They are also available on the AERIS/SEDOO/HyMeX database (https://www.aeris-data.fr/, Coquillat et al., 2019b). Meteorage data are provided by and are the property of Meteorage as a company.

Author contributions. FEr, EDe and OCa designed the methodologies to merge and analyze the lightning data. FEr implemented the methods and verified both the code and results. FEr created all the plots. RJB, SPe and SCo provided the lightning data and some expertise on their quality. FEr wrote the paper. All authors contributed to revisions of the paper.

Competing interests. The authors declare that they have no conflict of interest.

Special issue statement. This article is part of the special issue "Hydrological cycle in the Mediterranean

(ACP/AMT/GMD/HESS/NHESS/OS inter-journal SI)". It is not associated with a conference.

Acknowledgements. Felix Erdmann thanks the CNES and Météo-France for funding his PhD. This work is a contribution to the HyMeX program through the EXAEDRE project, grant ANR-16-CE04-0005, funded by the French research foundation ANR and the SOLID project, funded by CNES. Acknowledgements are also addressed to the CORSiCA-SAETTA main sponsors (Collectivité Territoriale de Corse through the Fonds Européen de Développement Régional of the European Operational Program 2007-2013 and the Contrat de Plan Etat Région, HyMeX/MISTRALS, Observatoire Midi-Pyrénées, Laboratoire d'Aérologie, CNES) and many individuals and regional institutions in Corsica that host the 12 stations of the network or helped us to find sites. The authors also want to thank the SAETTA team. EUMETSAT MSG/SEVIRI data were provided by SATMOS (Météo-France/CMS). We thank the AERIS/ICARE Data and Services Center for providing access to and visualization for the data used in Fig. 1b.

Financial support. This research has been supported by the Centre National d'Etudes Spatiales (grant no. SOLID).

Review statement. This paper was edited by Domenico Cimini and reviewed by Kenneth Cummins and two anonymous referees.

References

Bitzer, P. M., Burchfield, J. C., and Christian, H. J.: A Bayesian Approach to Assess the Performance of Lightning Detection Systems, J. Atmos. Ocean. Tech., 33, 563–578, https://doi.org/10.1175/JTECH-D-15-0032.1, 2016.

Blakeslee, R. and Koshak, W.: LIS on ISS: Expanded Global Coverage and Enhanced Applications, The Earth Observer, 28, 4–14, 2016.

Blakeslee, R. J., Bailey, J. C., Carey, L., Goodman, S. J., Rudlosky, S., Albrecht, R., Morales, C. A., Anselmo, E., and Neves, J.: São Paulo Lightning Mapping Array (SP-LMA): Network Assessment and Analyses for Intercomparison Studies and GOES-R Proxy Activities, J. Atmos. Ocean. Tech., 19, 1318–1332, 2013.

Blakeslee, R. J., Mach, D. M., Stewart, M. F., Buechler, D., and Christian, H.: Non-Quality Controlled Lightning Imaging Sensor (LIS) on International Space Station (ISS) Provisional Science Data P0.2, https://doi.org/10.5067/LIS/ISSLIS/DATA204, dataset available online from the NASA Global Hydrology Center DAAC, Huntsville, Alabama, USA, 2017.

Boccippio, D., Koshak, W. J., and Blakeslee, R. J.: Performance Assessment of the Optical Transient Detector and Lightning Imaging Sensor. Part I: Predicted Diurnal Variability, J. Atmos. Ocean. Tech., 19, 1318–1332, https://doi.org/10.1175/1520-0426(2002)019<1318:PAOTOT>2.0.CO;2, 2002.

Cecil, D. J., Goodman, S. J., Boccippio, D. J., Zipser, E. J., and Nesbitt, S. W.: Three Years of TRMM Precipitation Features. Part I: Radar, Radiometric, and Lightning Characteristics, Mon. Weather Rev., 133, 543–566, https://doi.org/10.1175/MWR-2876.1, 2005.

Christian, H. J., Blakeslee, R. J., Goodman, S. J., Mach, D. A., Stewart, M. F., Buechler, D. E., Koshak, W. J., Hall, J. M., Boeck, W. L., Driscoll, K., and Boccippio, D. J.: The Lightning Imaging Sensor, 11th International Conference on Atmospheric Electricity, NASA Conf. Publ. NASA/CP-1999-209261a, 746–749, 1999.

Christian, H. J., Blakeslee, R. J., Goodman, S. J., and Mach, D. M.: Algorithm Theoretical Basis Document (ATBD) For the Lightning Imaging Sensor (LIS), Earth Observing System (EOS) Instrument Product, National Aeronautics and Space Administration George C. Marshall Space Flight Center, 2000.

Coquillat, S., Defer, E., Jarnot, C., Lambert, D., Martin, J.-M., Pinty, J.-P., and Prieur, S.: SAETTA: fine-scale observation of the lightning activity in the framework of the CORSiCA Atmospheric Observatory, 8th HyMeX workshop, Valletta, 15–18 September, Valletta, Malta, 2014.

Coquillat, S., Defer, E., de Guibert, P., Lambert, D., Pinty, J.-P., Pont, V., Prieur, S., Thomas, R. J., Krehbiel, P. R., and Rison, W.: SAETTA: high-resolution 3-D mapping of the total lightning activity in the Mediterranean Basin over Corsica, with a focus on a mesoscale convective system event, Atmos. Meas. Tech., 12, 5765–5790, https://doi.org/10.5194/amt-12-5765-2019, 2019a.

Coquillat, S., Defer, E., De Guibert, P., Lambert, D., and Prieur, S.: SAETTA LMA L1 lightning data, AERIS/SEDOO/HyMeX database, available at: https://www.aeris-data.fr/, last access: 23 October 2019b.

Cummins, K. L. and Murphy, M. J.: An Overview of Lightning Locating Systems: History, Techniques, and Uses, With an In-depth Look at the U.S. NLDN, IEEE Trans. Electromag. Compat., 51, 499–518, https://doi.org/10.1109/TEMC.2009.2023450, 2009.

Defer, E., Lagouvardos, K., and Kotroni, V.: Lightning activity in the eastern Mediterranean region, J. Geophys. Res., 110, D24210, https://doi.org/10.1029/2004JD005710, 2005.

Deierling, W. and Petersen, W. A.: Total lightning activity as an indicator of updraft characteristics, J. Geophys. Res., 113, D16210, https://doi.org/10.1029/2007JD009598, 2008.

Deierling, W., Petersen, W. A., Lathman, J., Ellis, S., and Christian, H. J.: The relationship between lightning activity and ice fluxes in thunderstorms, J. Geophys. Res., 113, D15210, https://doi.org/10.1029/2007JD009700, 2008.

Goodman, S. J., Blakeslee, R. J., Koshak, W. J., Mach, D., Bailey, J., Buechler, D., Carey, L., Schultz, C., Bateman, M., McCaul Jr., E., and Stano, G.: The GOES-R Geostationary Lightning Mapper (GLM), Atmos. Res., 125–126, 34–49, https://doi.org/10.1016/j.atmosres.2013.01.006, 2013.

Höller, H. and Betz, H.-D.: Study on Inter-comparison of LIS and Ground-Based Lightning Location System Observations, Report ITT No. 09/996, EUMETSAT, Deutsches Zentrum für Luft- und Raumfahrt, 2010.

Koshak, W. J., Mach, D. M., and Bitzer, P. M.: Mitigating VHF Lightning Source Retrieval Errors, J. Atmos. Ocean. Tech., 35, 1033–1052, https://doi.org/10.1175/JTECH-D-17-0041.1, 2018.

Mach, D. M., Christian, H. J., Blakeslee, R. J., Boccippio, D. J., Goodman, S. J., and Boeck, W. L.: Performance assessment of the Optical Transient Detector and Lightning Imaging Sensor, J. Geophys. Res., 112, D09210, https://doi.org/10.1029/2006JD007787, 2007.

Mattos, E. V., Machado, L. A. T., Williams, E. R., Goodman, S. J., Blakeslee, R. J., and Bailey, J. C.: Electrification life cycle of incipient thunderstorms, J. Geophys. Res.-Atmos., 122, 4670–4697, https://doi.org/10.1002/2016JD025772, 2017.

Nag, A., Murphy, M. J., Schulz, W., and Cummins, K. L.: Lightning locating systems: Insights on characteristics and validation techniques, Earth Space Sci., 2, 65–93, https://doi.org/10.1002/2014EA000051, 2015.

Peterson, M., Rudlosky, S., and Deierling, W.: The Evolution and Structure of Extreme Optical Lightning Flashes, J. Geophys. Res.-Atmos., 122, 13370–13386, https://doi.org/10.1002/2017JD026855, 2017.

Pédeboy, S., Barnéoud, P., Defer, E., and Coquillat, S.: Analysis of the Intra-Cloud lightning activity detected with Low Frequency Lightning Locating Systems, 25th International Lightning Detection Conference, 7th International Lightning Meteorology Conference, 12–15 March, Ft. Lauderdale, Fl, USA, 2018a.

Pédeboy, S., Defer, E., Kolmašová, I., Coquillat, S., Guibert, P. D., Lambert, D., Pinty, J.-P., Prieur, S., Santolík, O., Lán, R., and Uhlíř, L.: Analysis of the initiation phase in negative lightning flashes exhibiting an intense return stroke peak current, XVI International Conference on Atmospheric Electricity, 17–22 June, Nara City, Nara, Japan, 2018b.

Rison, W., Thomas, R., Krehbiel, P., Hamlin, T., and Harlin, J.: A GPS-based Three-Dimensional Lightning Mapping System: Initial Observations in Central New Mexico, Geophys. Res. Lett., 26, 3573–3576, https://doi.org/10.1029/1999GL010856, 1999.

Rudlosky, S. D., Peterson, M. J., and Kahn, D. T.: GLD360 Performance Relative to TRMM LIS, J. Atmos. Ocean. Tech., 34, 1307–1322, https://doi.org/10.1175/JTECH-D-16-0243.1, 2017.

Schulz, W., Diendorfer, G., Pedeboy, S., and Poelman, D. R.: The European lightning location system EUCLID – Part 1: Performance analysis and validation, Nat. Hazards Earth Syst. Sci., 16, 595–605, https://doi.org/10.5194/nhess-16-595-2016, 2016.

Stuhlmann, R., Rodriguez, A., Tjemkes, S., Grandell, J., Arriaga, A., Bézy, J.-L., Aminou, D., and Bensi, P.: Plans for EUMETSAT's Third Generation Meteosat geostationary satellite programme, Adv. Space Res., 36, 975–981, https://doi.org/10.1016/j.asr.2005.03.091, 2005.

Thomas, R. J., Krehbiel, P. R., Rison, W., Hamlin, T., Boccippio, D. J., Goodman, S. J., and Christian, H. J.: Comparison of ground-based 3-dimensional lightning mapping observations with satellite-based LIS observations in Oklahoma, Geophys. Res. Lett., 27, 1703–1706, https://doi.org/10.1029/1999GL010845, 2000.

Thomas, R. J., Krehbiel, P. R., Rison, W., Hunyady, S. J., Winn, W. P., Hamlin, T., and Harlin, J.: Accuracy of the Lightning Mapping Array, J. Geophys. Res., 109, D14207, https://doi.org/10.1029/2004JD004549, 2004.

Thompson, K. B., Bateman, M. G., and Carey, L. D.: A Comparison of Two Ground-Based Lightning Detection Networks against the Satellite-Based Lightning Imaging Sensor (LIS), J. Atmos. Ocean. Tech., 31, 2191–2205, https://doi.org/10.1175/JTECH-D-13-00186.1, 2014.

Vaisala: Vaisala Thunderstorm Advanced Total Lightning Sensor LS7002, available at: https://www.vaisala.com/sites/default/files/documents/WEA-LS7002-Datasheet-B211284EN-A-LOW.pdf (last access: 28 June 2018), 2013.

Vincenty, T.: Geodetic inverse solution between antipodal points, DMAAC Geodetic Survey Squadron, https://doi.org/10.5281/zenodo.32999, 1975.

Zhang, D., Cummins, K. L., Bitzer, P. M., and Koshak, W. J.: Evaluation of the Performance Characteristics of the Lightning Imaging Sensor, J. Atmos. Ocean. Tech., 36, 1015–1030, https://doi.org/10.1175/JTECH-D-18-0173.1, 2019.

II.1.3 Important points to remember

The work of Section II.1.2 was conducted prior to the availability of any GEO lightning data. It constitutes the first analysis of lightning observations within this thesis. Similarities and differences between records of the ground-based Meteorage network and LMA SAETTA as well as the LEO ISS-LIS are investigated in detail as a crucial requirement for the further research within this thesis. It was found that the satellite-based ISS-LIS detects flashes with large extent and long duration well. In contrast, the majority of the smaller flashes with relatively short extent and duration are missed by ISS-LIS. Flashes that occur in lower altitudes, i.e., below 7 km, are particularly difficult to locate from space. The results also indicate that the LF or VHF signal strength at the ground does not significantly impact the flash detection of ISS-LIS. The methods of lightning data handling and processing, i.e., reading the different data formats and matching of flashes observed by the different LLSs, developed here are used and adapted for US LLSs during the course of this thesis. For example, the following chapter presents the comparison of observations of ISS-LIS to those of the US LF network NLDN applying the same methodology as in this chapter.

II.2 | Analyzing LLSs NLDN and ISS-LIS in the USA

This chapter presents the work necessary to verify the application of tools that use data from different parts of the world, e.g., like our MTG-LI pseudo-observation generator will be trained in the USA and applied in France. The previous chapter analyzed operational LLSs in France. The LF network Meteorage provides total lightning observations and coverage of the entire European territory of France. National LF ground-based total lightning observations are assumed to be the most suitable source for a GEO pseudo-observation generator as they cover larger domains than for example LMAs, feature higher DE than global scale LF networks, and similar networks operate in both the USA and France.

The US equivalent to Meteorage is called NLDN. Technical specifications of NLDN can be found in Section II.4.2. Here, it is sufficient to recognize that NLDN and Meteorage locate the same lightning processes as both Meteorage and NLDN use the same Vaisala LS702 sensors. NLDN (Meteorage) covers the entire contiguous US [CONUS] (France) and locates both CG and IC lightning. On the other hand, thunderstorms in the USA have different characteristics than in Europe. The US landmass is significantly greater than the European landmass. Whereas latitudinal mountain ranges, i.e., the Pyrenees, the Alps, and the Carpathians (from west to east) prevent direct exchange of subtropical and polar air masses in Europe, tropical air from the Gulf of Mexico can directly collide with cold air from Canada over the USA, especially in the region of the Great Plains and southeastern (SE) US. It may also be considered that most regions of the USA are situated at lower latitudes than most of Europe. Therefore, the summertime atmosphere is usually warmer, the tropopause is found at higher altitudes, and clouds might grow thicker in the CONUS than in Europe. The cloud top height and cloud thickness likely influence the detection and location of lightning of both ground-based and satellite LLSs since hydrometeors scatter radiation and especially optical light.

Figure II.2.1 illustrates that this chapter draws the connection between the lightning observation in Europe and in the USA. LEO LLSs can be utilized to compare lightning observations over the two continents assuming that LEO LLSs exhibit constant performances at the same local times. The ISS-LIS provides coverage of the entire CONUS and all French territories. It is the common reference to compare satellite observed lightning over the USA and France, and for an indirect comparison of the LF networks NLDN and Meteorage. A mission summary of three years of ISS-LIS was published by Blakeslee et al. (2020). They emphasize the possible use of ISS-LIS as reference network to assess characteristics of the flash detection of other LLSs.

First, the relative detection efficiency (DE) of ISS-LIS and NLDN are presented. Flash characteristics are then analyzed, and finally the results are compared to the results of the intercomparison of French LLSs (Erdmann et al., 2020a, see Chapter II.1). It should be mentioned that the same algorithms and methods explained in the previous section are applied here. Any differences in the findings are attributed to the LLSs and meteorological reasons. The indirect comparison of the LF networks NLDN and Meteorage (marked red in Figure II.2.1) is an

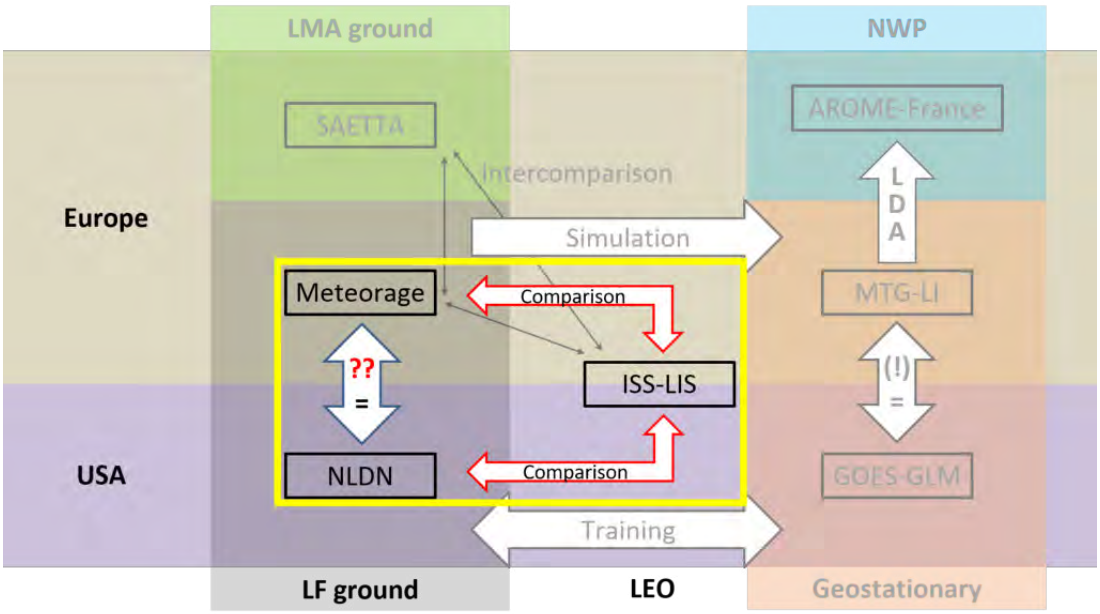


Figure II.2.1: Flow chart of work phases during this thesis as Figure 1. The LLS and phases detailed in this chapter are highlighted in the yellow frame, while other phases are shaded.

important step while developing the GEO lightning pseudo-observation generator. It will use NLDN and GLM as training data, and should be applied to produce pseudo MTG-LI records over France with Meteorage data as input. The method requires similar observations of NLDN and Meteorage, or an adjustment of the input dataset to fit the trained algorithm.

II.2.1 Datasets and methods

Six months of data are analyzed for a region in the SE US, from March 2018 to September 2018. The domain is restricted to less than 700,000 km^2 (Vaisala data policy requirements). In order to identify a suitable domain, a preliminary analysis of lightning activity is conducted. As the same NLDN data serve as database to study GLM observations (see Chapter II.3), a lightning climatology of NDLN and early GLM records, i.e., few months after the shift of GOES-16 to its operational GOES-E position, prior to the GOES-17 (now GOES-W) launch, and with an intermediate cloud top height assumption, provide the regions with high lightning activity within the entire CONUS and parts of the Gulf of Mexico. On that basis, the 700,000 km^2 domain was selected (see Figure II.2.2). The domain selection took also into account the perspective to simulate MTG-LI over France, a mainly landmass area. In addition, first studies using the early GLM data found that GLM flash DE was relatively high over the southeast (SE) CONUS, with relatively high location accuracy (e.g., Schultz et al., 2018, Rudlosky et al., 2018). In consequence, a domain between 30°N and 35°N latitude and from 95°W to 82°W longitude is studied.

Using the merging algorithm developed in Erdmann et al. (2020a) (see Chapter II.1), a flash database is created by merging ISS-LIS events (NLDN strokes and pulses) to flashes. The spatial and temporal criteria for merging events (strokes and pulses) to one flash are 15 km and 0.3 s (20 km and 0.4 s), respectively, thus, the same as for the comparison of ISS-LIS and Meteorage. The NLDN ground-based data are only considered at times when the corresponding region was within the FOV of the LEO ISS-LIS and when both instruments operated. Then,

the matching algorithm of Erdmann et al. (2020a) (see Chapter II.1) is applied. A coincident *matched* flash has at least one element, that is an optical event or an LF source, detected by both LLSs within a distance of 20 km and a time offset of 1.0 s. Flashes without a match are called *unmatched*. Then, flash characteristics such as flash extent, flash duration, the number of elements per flash, and flash energetics defined as mean and maximum amplitude count (for ISS-LIS) and LF amplitude (for NLDN) per flash are inferred and compared between matched and unmatched flashes.

II.2.2 Detection efficiency

The relative flash detection efficiency (DE) calculates the percentage of flashes detected by a reference LLS that are detected by the LLS in question. Figure II.2.2 provides DE within $1° \times 1°$ pixels. The area covered by the colored pixels indicates the study domain. The pixel color represents the DE value, and the greyscale contours mark the 0th (1 flash), 50th , 80th , and 95th percentiles of the flash number distribution per $0.25° \times 0.25°$ pixel. The NLDN DE surpasses the ISS-LIS DE in the region and for most individual pixels (Figure II.2.2). Low ISS-LIS DE is in particular observed in regions of low flash counts, e.g., for pixels near 90°W and for some pixels in the northeast of the domain. NLDN DE remains below 50 % for two pixels with high ISS-LIS flash counts. The observed flashes of those two pixels could constitute flashes in relatively high altitudes that are in general well detected by ISS-LIS but more challenging for LF ground-based networks.

A full summary of NLDN and ISS-LIS flash DE, also for day- and nighttime as well as CG and IC flashes can be found in Table II.2.1. ISS-LIS detected 58.9 % (4272 of 7258 flashes) of all flashes that NLDN detected during the ISS-LIS overflights. This overall flash DE agrees well with the ISS-LIS flash DE as reported over Corsica (Erdmann et al., 2020a, see Chapter II.1). Nevertheless, there are differences for observations in the SE US and over Corsica. A high nighttime ISS-LIS DE is observed relative to NLDN in the SE US (about 80 %), compared to about 59 % relative to Meteorage over Corsica. The daytime DE, however, is only 49.0 % over the SE US region and, thus, even lower than relative to Meteorage over Corsica (53.9 %). ISS-LIS DE for LF CG flashes equals 61.3 % in the SE US and 53.5 % over Corsica. The analysis of the corresponding flash DEs for LF IC flashes yield 57.5 % and 59.3 %, respectively, and thus similar values within the two regions. ISS-LIS features a higher DE for CG flashes than for IC flashes in the SE US region for the studied period. This result is counter-intuitive, as space-borne sensors usually detect high altitude flashes better than flashes at lower altitudes (Thomas et al., 2000, Erdmann et al., 2020a, see Chapter II.1). Further research revealed that the NLDN CG flashes of our dataset extend on average (maximum) 11.7 km (100.7 km). The mean (maximum) flash duration yields 0.30 s (3.24 s), and the mean (maximum) number of strokes+pulses per flash equals 5.8 (43) for the CG flashes. The IC flashes of our dataset are much smaller in all regards, with a mean (maximum) extent of 4.5 km (79.6 km), mean (maximum) duration of 0.09 s (1.30 s), and mean (maximum) number of strokes+pulses per flash of 2.3 (19). As shown in the following, ISS-LIS detects large, long lasting flashes with higher DE than small and short duration flashes (see also Erdmann et al., 2020a, Chapter II.1), thus, explaining the high DE of the CG flashes in the SE US.

It should be mentioned here that ISS-LIS data includes a variable to identify data issues, i.e., the data flags with warning and fatal attributes. A so-called fatal flag filter is used to remove all ISS-LIS events with active fatal flag attributes. This procedure is consistent with the methods in Erdmann et al. (2020a) (see Chapter II.1); however, these flags seem to mask reasonable

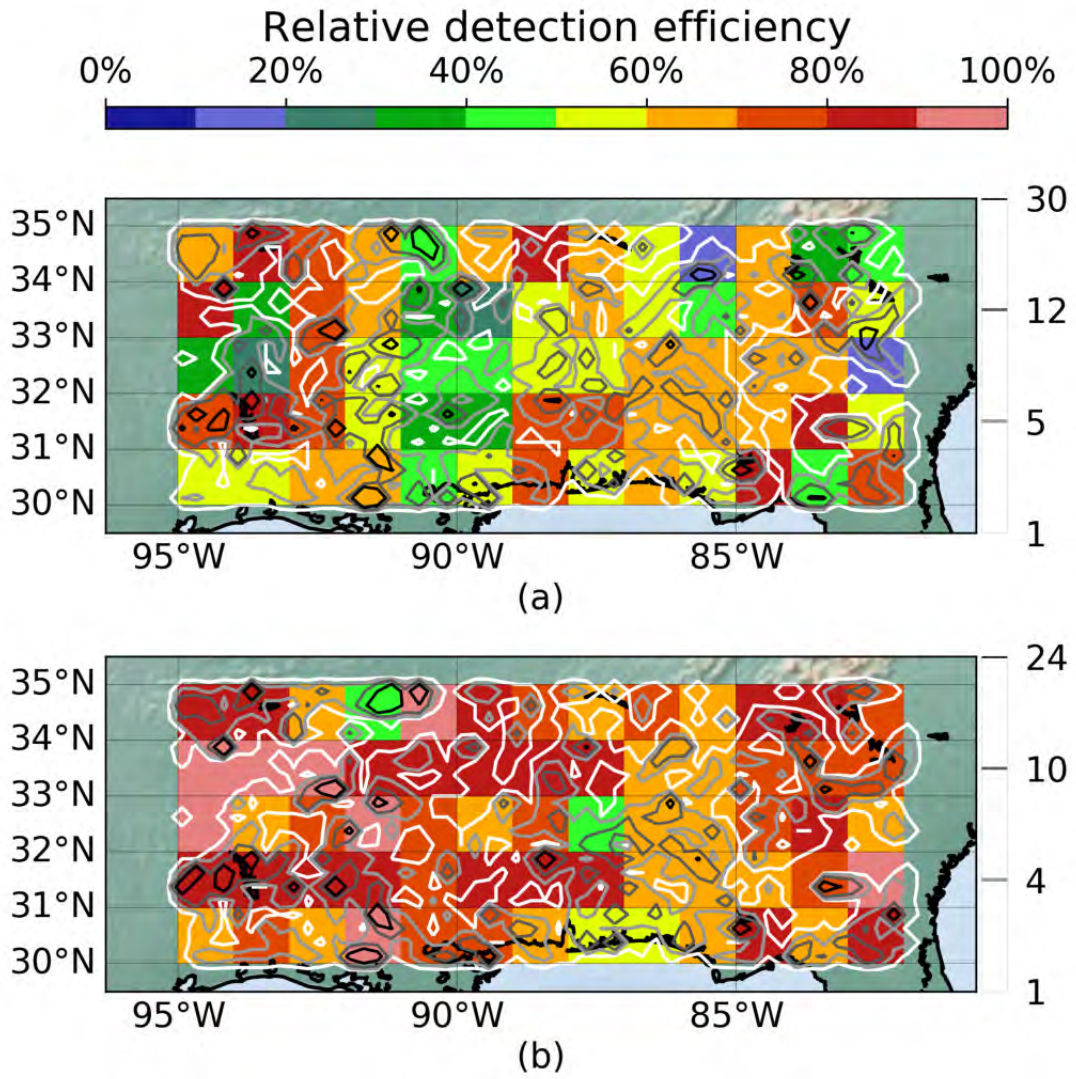


Figure II.2.2: Relative detection efficiency of ISS-LIS (a) and NLDN (b) with respect to the other instrument for $1° \times 1°$ pixels (colors). Grayscale contours mark the 0th (1 flash), 50th, 80th, and 95th percentiles of the flash number distribution per $0.25° \times 0.25°$ pixel.

Table II.2.1: ISS-LIS and NLDN relative detection efficiency as different categories and with flash number in each category. Values in parentheses excluding the single event (stroke/pulse) ISS-LIS (NLDN) flashes.

	Overall	Daytime	Nighttime	IC-flash	CG-flash
ISS-LIS DE [%]	58.9 (62.6)	49.0 (52.7)	80.3 (82.9)	57.5 (62.0)	61.3 (63.4)
NLDN flash number	7258 (4598)	4972 (3086)	2286 (1512)	4685 (2464)	2573 (2134)
NLDN DE [%]	75.7 (76.7)	74.6 (75.5)	77.1 (78.2)	-	-
ISS-LIS flash number	5352 (5170)	3058 (2941)	2294 (2229)	-	-

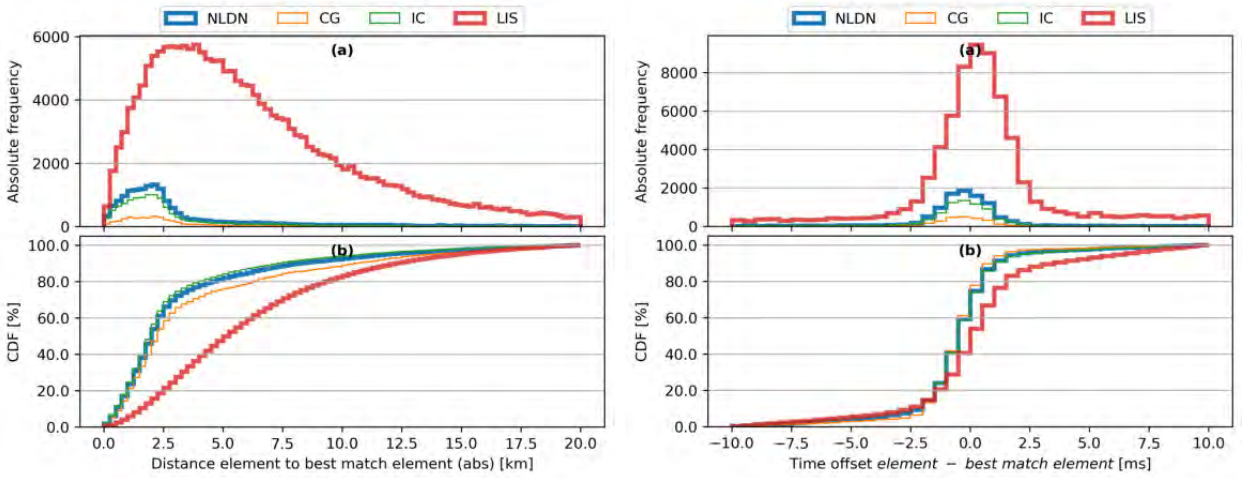
Table II.2.2: ISS-LIS and NLDN relative detection efficiency as different categories and with flash number in each category. Values in parentheses excluding the single event (stroke/pulse) ISS-LIS (NLDN) flashes.

	Overall	Daytime	Nighttime	IC-flash	CG-flash
ISS-LIS DE [%]	65.5 (70.4)	60.3 (66.0)	76.8 (79.5)	64.6 (70.9)	67.1 (69.7)
NLDN flash number	7628 (4795)	5218 (3232)	2410 (1563)	4916 (2558)	2712 (2237)
NLDN DE [%]	75.6 (76.6)	75.3 (76.2)	76.1 (77.3)	-	-
ISS-LIS flash number	6230 (6012)	3904 (3758)	2326 (2254)	-	-

ISS-LIS events for the US dataset, that reduce in particular the daytime DE significantly. Table II.2.2 shows the resulting DEs as presented in Table II.2.1 but including all ISS-LIS events, i.e., deactivating the fatal flag filter. With all events being used to create ISS-LIS flashes and match them to NLDN flashes, the daytime flash DE of ISS-LIS in the SE US exceeds 60 %. Hardly any change of the DE was observed in Europe as ISS-LIS operated well during the rare overpasses whether events with fatal flag attributes were removed or kept within the data (not shown). The NLDN flash number here increases, too, as using the fatal flag filter removed all the ISS-LIS events during some overpasses, and the corresponding times were then excluded from the comparison.

The DE of NLDN equals 75.7 % (4049 of 5352 flashes) relative to ISS-LIS with the fatal flag filter in the SE US. In general, NLDN DEs of all flash types (i.e., day- and nighttime, CG, IC) provided in Table II.2.1 are about 5 % lower than for the corresponding flash types of Meteorage records over Corsica. NLDN flash DE remains constant during the diurnal cycle and does not depend on local day- or nighttime hours.

While flash DE is investigated, the distances and timing offsets between coincident observations become an interesting subsequent topic. As in Erdmann et al. (2020a) (see Chapter II.1) the closest elements (events and LF strokes and pulses for ISS-LIS and NLDN, respectively) per matched flash lead to the following statistics. Figure II.2.3 presents the distances and time offsets between matched flashes as histograms (top, (a)a and (b)a) and CDFs (bottom, (a)b and (b)b). The distributions depend on whether the given element was observed by ISS-LIS (red) or NLDN (blue, CG yellow, IC green), and there is significantly more data for the ISS-LIS events than NLDN strokes and pulses. The median distance between elements of coincident flashes equals 2.13 km and 5.26 km given an NLDN stroke or pulse and an ISS-LIS event, respectively. The median values of the distributions are in general more resistant to extreme values and outliers than the means. Hence, mean distances provide larger offsets of 3.40 km and 6.23 km given an NLDN stroke or pulse and an ISS-LIS event, respectively. NLDN CG strokes were on average further apart from the closest ISS-LIS event than IC pulses, with a median (mean) of 2.34 km (4.06 km) and 2.06 km (3.16 km), respectively. The distribution peak, that defines the most frequently observed value, is found at smaller values than the average statistics. This is caused by a relatively large right tail of the distributions ending at the specified matching distance of 20 km. Peak distances of 1.75 km to 2.25 km (given an NLDN stroke or pulse) and 2.50 km to 4.00 km (given an ISS-LIS event) can be reported for the observed distributions (Figure II.2.3(a)). The flatter CG curve than the overall and IC curves is indicative of the slightly higher geolocation offset for CG strokes. The CG strokes might be located close to convective cores, thus, under a thick cloud, and the emitted light might be detected by ISS-LIS as scattered to cloud edges or an anvil. It was also observed that NLDN CG flashes extended on average significantly further than IC flashes, with 11.5 km and 4.4 km, respectively, which might affect the distances being calculated for events and strokes+pulses of coincident flashes.



(a) (b)

Figure II.2.3: Matching distance (a) and time offset (b) for the closest events (ISS-LIS) and strokes and pulses (NLDN) of matched flashes. Time offsets greater than ±10 ms are not plotted.

All distance offsets are in the same order of magnitude as the largest geolocation uncertainty, i.e., the ISS-LIS pixel size (about 5 km) plus possible pointing errors.

Time offsets between elements of coincident flashes can be positive or negative as shown in Figure II.2.3(b). A positive time offset means that the given element occurred later than the matched element, while a given element was observed earlier than the match for a negative time offset. The median time offsets are smaller than the ISS-LIS integration time frame of 2 ms. A given NLDN pulse or stroke occurred with a median offset of 0.3 ms (0.2 ms CG, 0.4 ms IC) earlier than the matched event, with similar results for the CG strokes and IC pulses. The given ISS-LIS event was observed 1.2 ms after the matched NLDN pulse or stroke on median. In general, ISS-LIS detected the first source of a lightning flash slightly later than NLDN, although within the limits of instrument uncertainty. Peak time offsets to the closest matched elements range from −0.5 ms to 0.0 ms and 0.0 ms to 0.5 ms for a given NLDN stroke/pulse and a given ISS-LIS event, respectively (Figure II.2.3(b)). CG and IC time offset distributions are similar. The distribution mean is in general more sensitive to extreme values of the distribution, i.e., exceptionally large time offsets. Our matching criterion allows coincident flashes up to time offsets of 1.0 s. The distribution outliers increase the mean time offsets compared with the medians to −6.3 ms (32.3 ms) between a given NLDN stroke or pulse (ISS-LIS event) and the closest element of the matched flash. This trend is also observed for given CG strokes and IC pulses, with mean time offsets of 2.0 ms and −9.3 ms, respectively. It is recalled that the negative IC time offset implies that ISS-LIS detected the closest (in time) event of the coincident flash later than the NLDN IC pulse. As the overall ISS-LIS time offset features a relatively high positive value (ISS-LIS event after NLDN pulse or stroke), some IC flashes producing weak cloud illumination but strong initial LF radiation deemed responsible for the observed mean time differences. Such processes are mainly caused by fast vertical discharges, e.g., K- and M-changes. They could have occurred mainly during the day when the ISS-LIS detection threshold requires a higher visible luminosity than during the night, however, this hypothesis is not examined.

II.2.3 Flash Characteristics

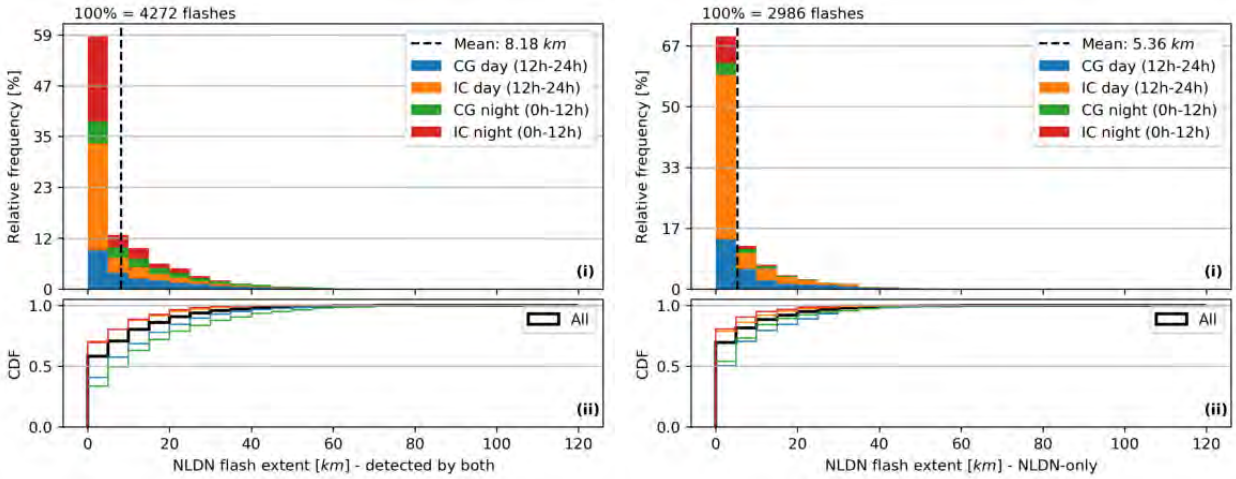
In order to compare NLDN and Meteorage observations in the following section, statistics of NLDN flash characteristics are presented here. Our US database comprises ISS-LIS and NLDN data only. Hence, there is no altitude information. ISS-LIS flash characteristics can be directly compared to the flash characteristics as observed over Corsica as the same space-borne sensor is used. Any statistical difference found between ISS-LIS observations over the SE US and ISS-LIS observations over Corsica may influence the indirect comparison of NLDN and Meteorage flash characteristics where ISS-LIS should serve as the common reference. In the following, the statistics are presented for distributions of matched flashes detected by both the given and the reference LLS and unmatched flashes (only detected by the given LLS). This separation conforms with the methods in Erdmann et al. (2020a) (Chapter II.1) and allows to analyze the lightning observations of ISS-LIS and the ones of the LF network in more detail and relative to each other. The figures in this and the following chapter separate, as mentioned, matched and unmatched flashes. The flashes are categorized regarding their time of occurrence (local day, night) and their type (CG, IC). Hence, 4 categories are distinguished and plotted with different color. All panels include a histogram (i, top) with stacked bars for the 4 categories and the CDF (ii, bottom).

Flash extent is defined as the sum of North-South (NS) and West-East (WE) distances. Those of NLDN flashes yield on average 8.18 km and 5.36 km with and without coincident ISS-LIS, respectively. Figure II.2.4 shows the distributions for NLDN flashes, and Figure II.2.5 for ISS-LIS flashes. Both the distributions for matched (a) and for unmatched flashes (b) peak for a flash extent from 0 km to 5 km. Matched ISS-LIS flashes extended on average 29.82 km. The unmatched ISS-LIS flashes were significantly shorter with an average extent of 19.03 km. This result is confirmed by the distribution peak extent of 15.0 km to 17.5 km for both matched and unmatched flashes, but with a secondary peak from 0.0 km to 2.5 km only in the unmatched flash distribution. The longest NLDN flash was a daytime CG flash with an extent of 100.7 km. The longest ISS-LIS flash extended 181.4 km at nighttime as an IC flash. ISS-LIS flashes are in general more extended than NLDN flashes due to the different design of the LLSs. Indeed, ISS-LIS maps the cloud top illuminations, whereas NLDN delivers punctual observations.

The flash duration is defined as the time period in between the first and last element of one flash. The distributions of flash duration are shown in Figures II.2.6 and II.2.7 for NLDN and ISS-LIS, respectively. The matched flashes (a) lasted on average longer than the unmatched flashes (b) for both LLSs. Matched (unmatched) NLDN flashes feature an average duration of 0.19 s (0.12 s). ISS-LIS flashes lasted on average longer than the NLDN flashes with average ISS-LIS flash duration of 0.32 s and 0.20 s for matched and unmatched flashes, respectively. Flash duration distributions peak at the smallest distribution bin from 0.00 s to 0.05 s in all cases. The longest NLDN and ISS-LIS flashes lasted 3.24 s and 2.93 s, respectively [1].

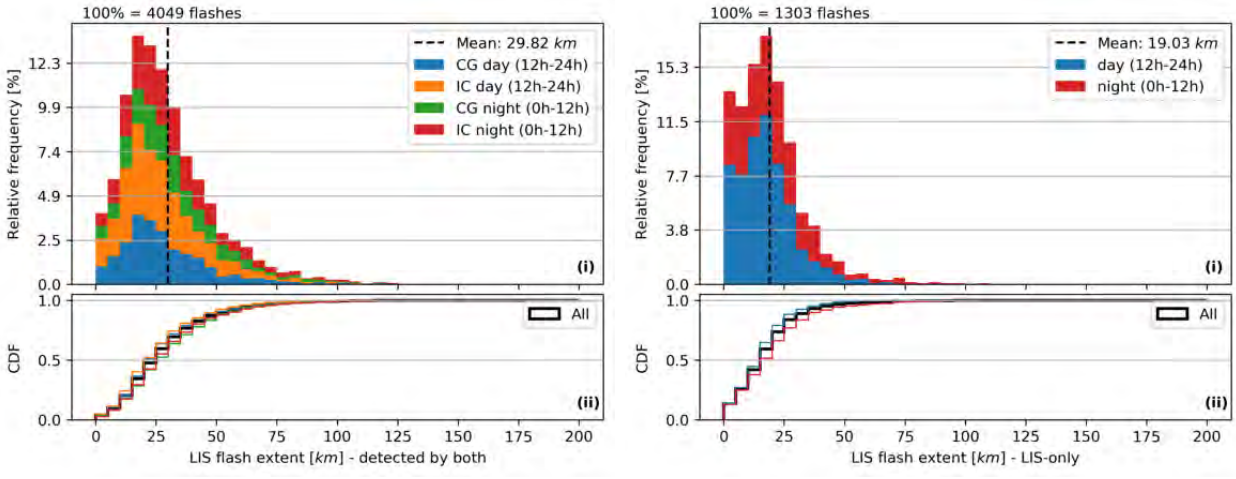
Flash extent and duration usually increase with the number of elements per flash. In accordance with the previous results, matched NLDN flashes comprise on average more strokes+pulses than NLDN-only flashes (3.94 and 3.09, respectively) as shown in the corresponding distributions in Figure II.2.8. In particular, there are relatively more unmatched flashes with one or two strokes+pulses (first bin, about 42.5 %) than for matched flashes (about 32.6 %). The ISS-LIS flashes with coincident NLDN observation had on average 49.61 events (see Figure II.2.9a). The average event number per flash equals 23.72 for ISS-LIS-only flashes (see Figure II.2.9b). The mapping of ISS-LIS observations comprises significantly more optical events than the recorded LF sources of NLDN per flash, because of the different processes

[1] Maxima not included in Figure II.2.6 and Figure II.2.7 for convenience



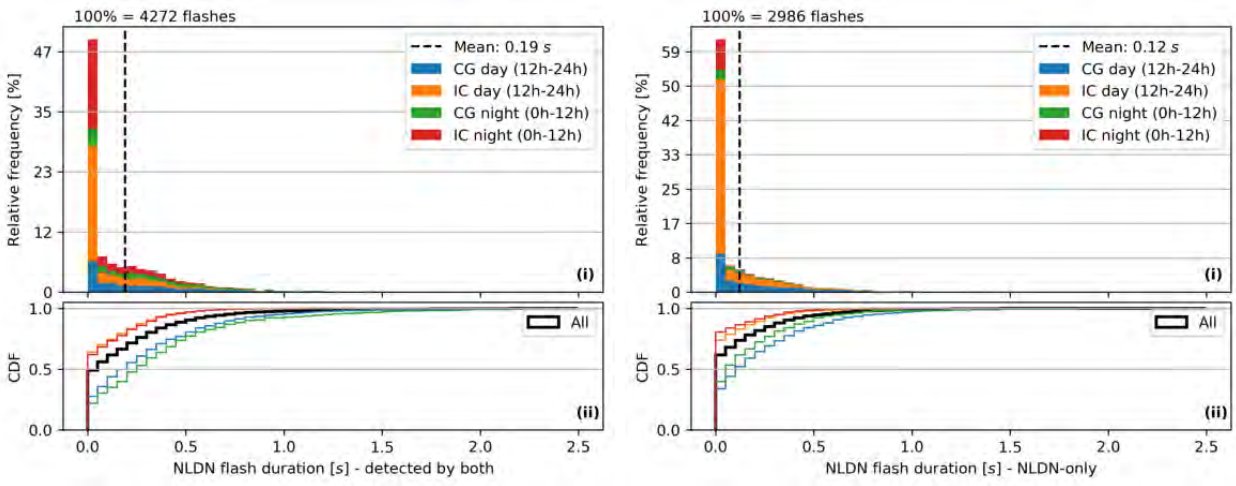
(a) (b)

Figure II.2.4: NLDN flash extent of matched (a) and unmatched (b) flashes, with the histogram in (i) and the cumulated histogram in (ii). Colors show categories of day, night, CG, and IC flashes as indicated. The numbers are stacked in (i) for the 4 categories indicated in the legend.



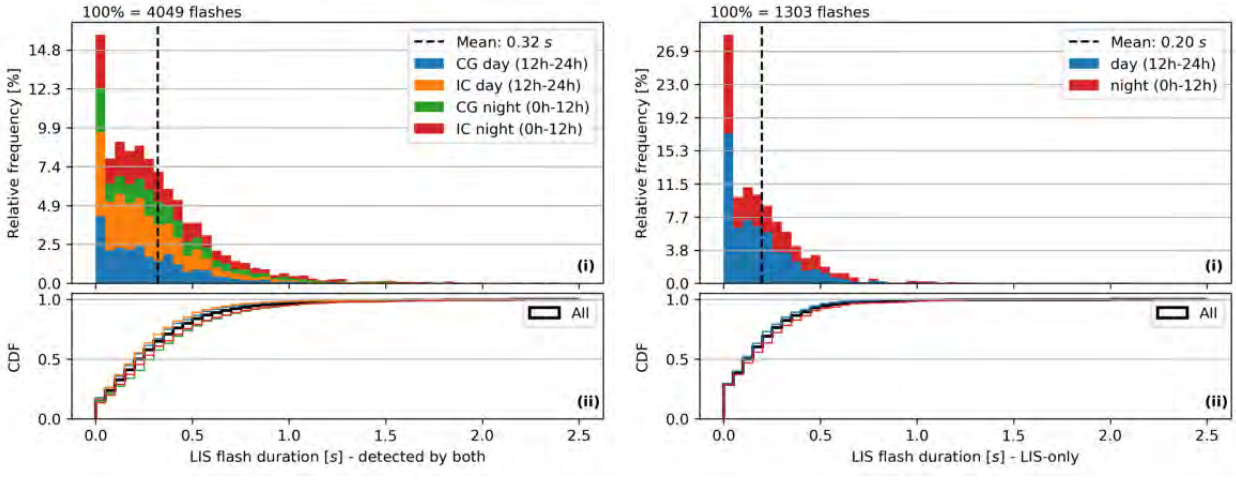
(a) (b)

Figure II.2.5: ISS-LIS flash extent of matched (a) and unmatched (b) flashes, with the histogram in (i) and the cumulated histogram in (ii). Colors show categories of day, night, CG, and IC flashes as indicated. The numbers are stacked in (i) for the 4 categories indicated in the legend. Note: No flash type (IC, CG) for ISS-LIS-only flashes



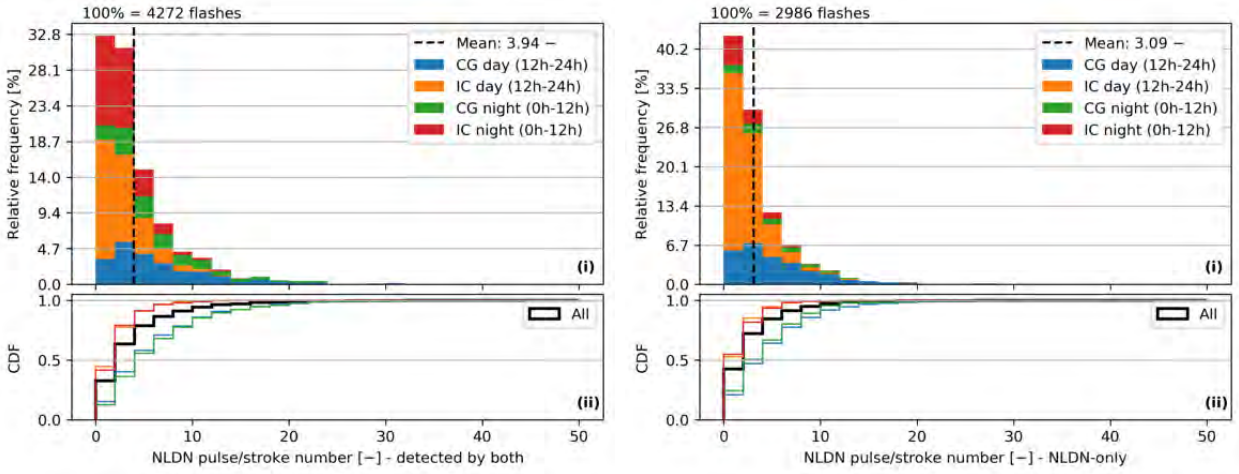
(a) (b)

Figure II.2.6: As Figure II.2.4 for NLDN flash duration.



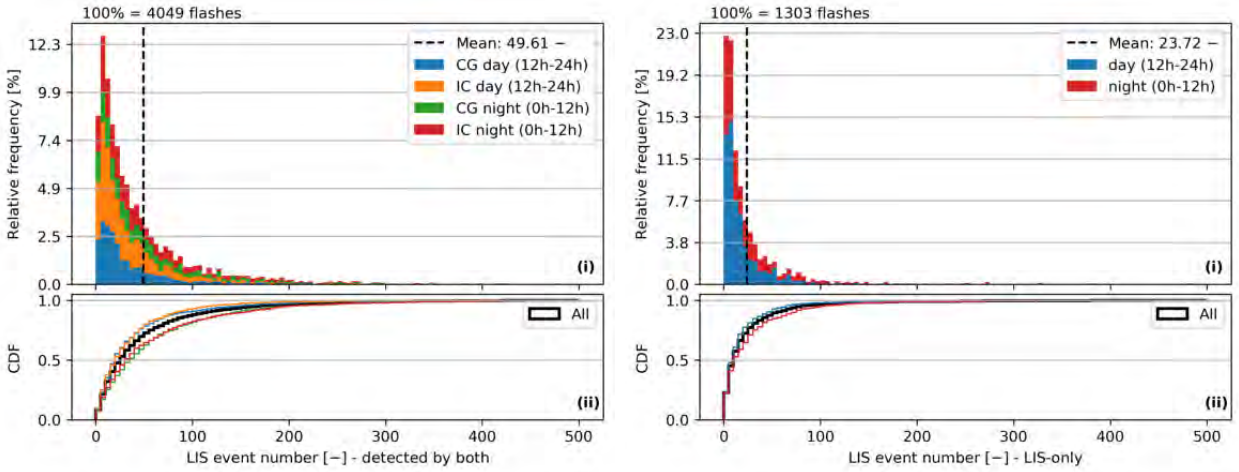
(a) (b)

Figure II.2.7: As Figure II.2.5 for ISS-LIS flash duration.



(a) (b)

Figure II.2.8: As Figure II.2.4 for NLDN stroke+pulse number per flash.



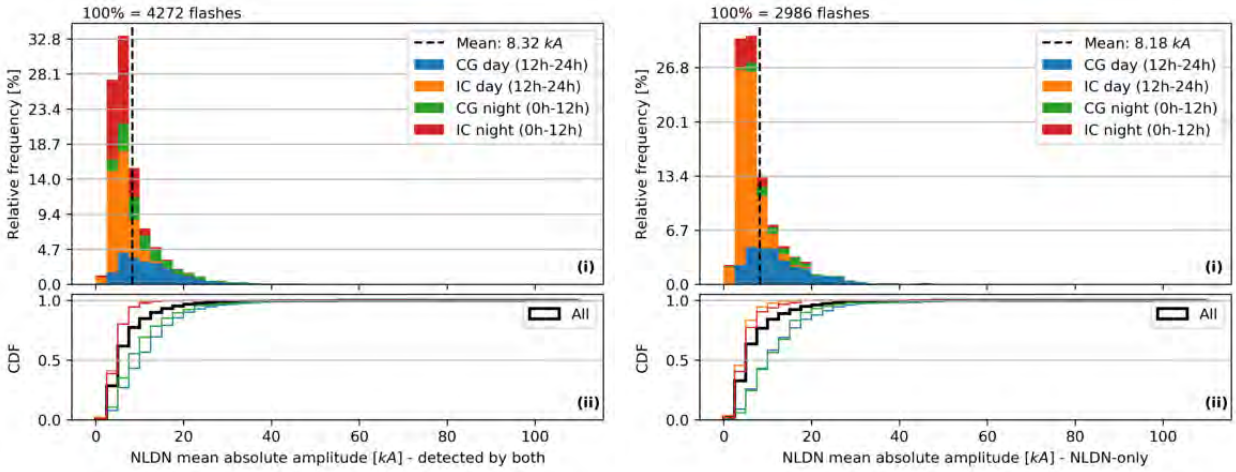
(a) (b)

Figure II.2.9: As Figure II.2.5 for ISS-LIS event number per flash.

detected and located by the two LLSs, and because of the detection techniques applied. All element number distributions peak within the two smallest bins. Hence, few elements per flash are the general case, however, there can be a high number of both events and strokes+pulses per flashes as proved by identified maxima of 1161 events for an ISS-LIS nighttime IC flash[2] and 43 strokes+pulses for an NLDN nighttime CG flash.

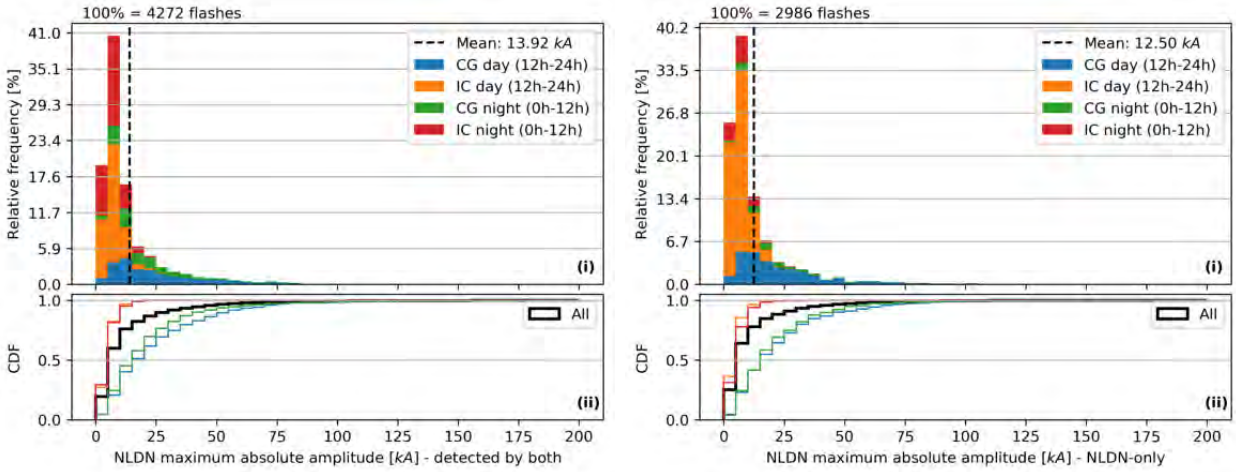
The analysis of flash energetics includes the average and maximum LF current (amplitude count) of NLDN pulses/strokes (ISS-LIS events) per flash. The average LF current (amplitude count) per flash is defined as the mean value of the recorded LF currents (amplitude counts) of all strokes and pulses (events) of that given flash. Equivalently, the maximum equals the highest LF current (amplitude count) measured for any pulse or stroke (event) of the flash. Then, the mean of the average LF currents over all flashes equals 8.32 kA and 8.18 kA for flashes with and without coincident ISS-LIS observation, respectively (distribution in Figure II.2.10). The corresponding mean of the maximum LF currents over all flashes are 13.92 kA for matched

[2]Maximum not included in Figure II.2.9 for convenience



(a) (b)

Figure II.2.10: As Figure II.2.4 for NLDN mean absolute amplitude per flash.

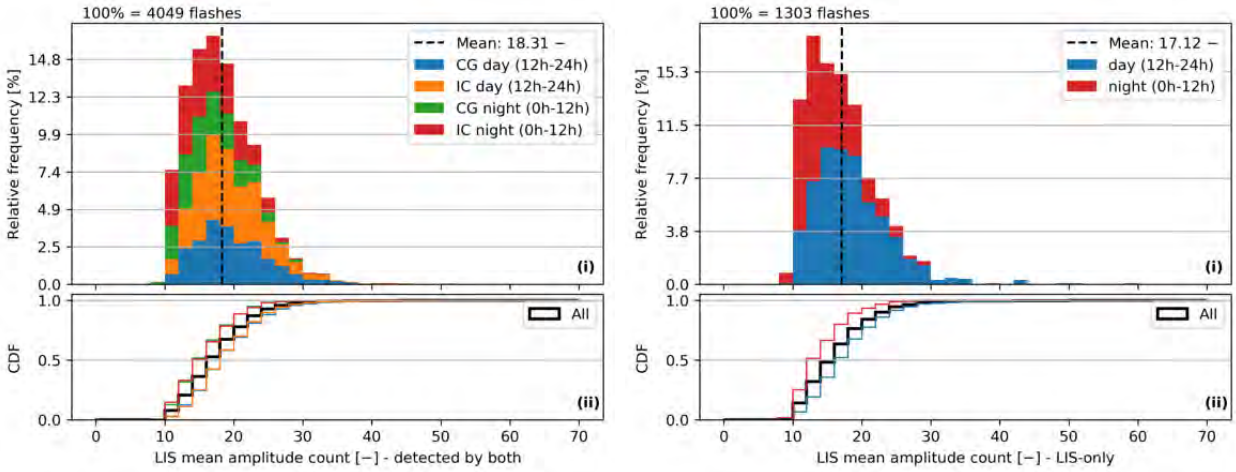


(a) (b)

Figure II.2.11: As Figure II.2.4 for NLDN maximum absolute amplitude per flash.

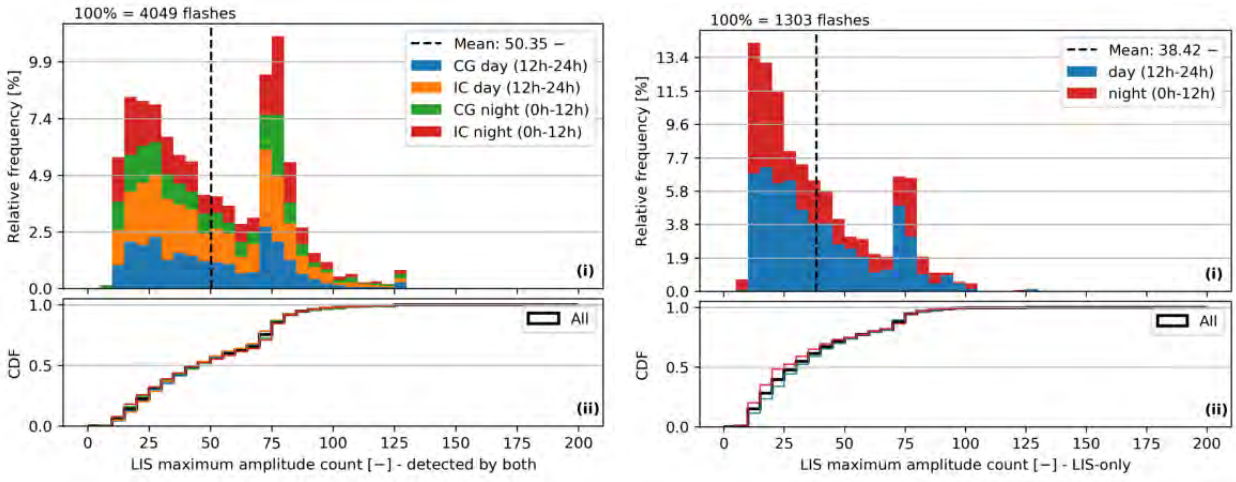
and 12.50 kA for unmatched flashes. Both the mean LF current distributions for matched and unmatched NLDN flashes peak at about 5 kA (Figure II.2.10). The maximum LF currents feature the most frequent bins from 5 kA to 10 kA as shown in Figure II.2.11 (a) and (b). All in all, NLDN flashes with and without ISS-LIS match feature similar statistics of flash energetics. Hence, ISS-LIS detects flashes rather independently of the LF current measured at the ground. This may become an important finding for the development of the pseudo-GEO lightning data generator, that will also use optical lightning observations, however, from a GEO sensor.

In the present study, ISS-LIS flash energetics use the raw amplitude counts, which are mostly proportional to the radiance recorded within the 777.4 nm, as radiance had not been available in the Provisional ISS-LIS data yet when the study was conducted. The mean and maximum amplitude counts are plotted in Figures II.2.12 and II.2.13, respectively. The maximum amplitude count per flash has a significant impact on the existence of a coincident LF source. Matched ISS-LIS flashes feature 50.35 maximum amplitude counts on average, while the ISS-LIS-only flashes appeared significantly darker (38.42 maximum amplitude count on



(a) (b)

Figure II.2.12: As Figure II.2.5 for ISS-LIS mean amplitude count per flash.



(a) (b)

Figure II.2.13: As Figure II.2.5 for ISS-LIS maximum amplitude count per flash.

average). This behavior can be seen in the distribution shapes and peak values. The matched flash distribution has a primary peak from 75 to 80 and the secondary peak occurs at 15 to 20 amplitude count. Both low and high value peaks feature slightly lower amplitude counts in the unmatched flash distribution, whereby the lower amplitude counts are even more frequent than the amplitude counts near the high value peak. The mean amplitude counts per flash support these trends, although the differences between matched and unmatched ISS-LIS flashes are less pronounced (Figure II.2.12). In summary, it is likely that ISS-LIS flashes detected by Meteorage are brighter than flashes only detected by ISS-LIS.

II.2.4 Comparison of the US NLDN with the French Meteorage LF network

Erdmann et al. (2020a) (see Chapter II.1) analyzed Meteorage and ISS-LIS observations in detail. Although the same methodology is applied when comparing NLDN and ISS-LIS observations in this chapter, the results depend on meteorological differences between the analyzed cases. The region in the SE US is approximately four times larger than the domain used for the Corsica study. All statistics and distributions contain much more cases for the NLDN-ISS-LIS comparison (USA) than for Meteorage-ISS-LIS one (France). Periods of 6 months and 13 months are studied in the US and in France, respectively. The majority of flashes that were observed by ISS-LIS during the study period occurred at local daytime in the US domain. In contrast, more nighttime than daytime flashes were observed during the ISS-LIS overpasses over the Corsica domain. It should be noticed that the US comparison does not include LMA observations and consequently altitude information. The US region is situated further south and data are gathered only from March to September during spring and summer. During the active storm season in the USA, higher fraction of cases with deep convection and severe storms are expected relative to the storm activity in the Corsica proximity. Severe storms are often characterized by high flash rates, thick clouds, and by a high fraction of IC relative to CG flashes. High flash rates are expected to reduce the ISS-LIS DE relative to an LF network, as found by Zhang and Cummins (2020) for the optical LLS GLM. Long optical paths through thick clouds may also reduce the ISS-LIS DE, but it also depends on the charge structure and location of emissions by lightning within the cloud. A higher fraction of IC flashes in the US region than over Corsica could then mean relatively short optically paths for the majority of flashes, and thus an increase in DE relative to the Corsica study (Thomas et al., 2000, Erdmann et al., 2020a, ; see Chapter II.1).

ISS-LIS' DE was slightly higher over the SE US than over Corsica (compare DEs in Table 1 of Section II.1.2 for ISS-LIS and Meteorage, Table II.2.1 for ISS-LIS and NLDN). In particular, the nighttime DE of ISS-LIS was significantly better (80 % versus 59 %) in the SE US than over Corsica. Contrarily, the daytime DE of ISS-LIS was lower in the SE US than over Corsica (49 % versus 54 %). If, however, all events are used without the fatal flag filter, the daytime DE increases to 60 % in the SE US while it remains similar over Corsica. This result indicates a possible issue with the provisional version of ISS-LIS data that was available at the time of the study. Correct ISS-LIS events might be flagged erroneously. The CG flash DE was higher in the SE US than over Corsica, and the IC flash DE appears to be slightly lower for ISS-LIS records in the SE US than over Corsica. Relatively high CG flash DE of ISS-LIS could be attributed to the large average size and long duration of the CG flashes relative to the IC flashes in the SE US.

The DE of NLDN was somewhat lower than that of Meteorage overall, at daytime, and at night. The difference is likely attributed to the longer baselines of NLDN and thus potentially lower IC DE than Meteorage. It should be reminded that a higher fraction of IC flashes, that are usually more difficult to be detected by a ground-based LF network than CG flashes, was expected, however, it was not observed by the LF networks. The IC-fraction equals 64.5 % for the NLDN data in the SE US study, compared to 64.8 % of IC flashes in the Meteorage data during the Corsica study. It yet remains the dependency of the IC flash DE on the baseline distance of LF networks (see Section I.2.4) and meteorological reasons such as storm types, cloud and charge structure, and lightning characteristics to explain lower NLDN than Meteorage flash DE.

A general finding is that differences between ISS-LIS daytime and nighttime flash characteristics are less pronounced over the SE USA with more cases than over Corsica where only few daytime flashes occurred for the overpasses during the studied period. These rare daytime ISS-LIS flashes over Corsica appeared smaller, with shorter duration and with less events during the studied period than the US daytime ISS-LIS flashes and than the nighttime flashes in both regions. As the daytime flashes during the Corsica study are related to just few storm cases, this finding is likely case specific and not representative in a climatological sense.

Distance offsets between coincident flashes showed the same order of magnitude but were slightly greater over the US than over Corsica (see Figure 6 of Section II.1.2 and Figure II.2.3). It is observed that CG flashes featured shorter distance offsets than IC flashes over Corsica, while the SE US CG flashes exhibited higher distance offsets than the IC flashes. The fraction of CG and IC on the overall distributions is similar in both regions. Meteorage data were provided with nanosecond time accuracy, whereas the NLDN data used here has millisecond time resolution. This difference affects the details of the statistics, but it is not decisive for the comparison of NLDN and Meteorage results. Time differences remained mostly within the range of timing uncertainty of the LLSs. There is a shallow trend that NLDN detected flashes a few milliseconds earlier than ISS-LIS. Such a trend could not be observed for Meteorage and ISS-LIS over Corsica.

Flash extent distribution and flash extent average of ISS-LIS flashes were very similar for both flash categories with and without coincident LF observation over Corsica and over the SE USA for both distribution mean and peak values (see Figure 8 of Section II.1.2 and Figure II.2.5). The LF ground-based networks within the regions showed some differences: Meteorage flashes were on average more extended than NLDN flashes (9.9 km versus 7.0 km). The detailed analysis of the data reveals that this is a result of rare, large nighttime flashes over the Corsica domain that influence the distribution relatively strongly due to the limited number of cases. The distribution shapes and CDFs of NLDN and Meteorage flash extent agree better than the average values might imply (see Figure 11 of Section II.1.2 and Figure II.2.4). For example, about 57 % and 62 % of all Meteorage and NLDN flashes, respectively, had an extent of less than 10 km. This difference likely results from the slightly higher number of Meteorage IC pulses increasing the flash extent compared with NLDN, although this hypothesis cannot be proven with the available data.

Distributions and averages of ISS-LIS both matched and unmatched flash duration feature very similar mean and peak values for both studied regions (see Figure 9 of Section II.1.2 and Figure II.2.7). The same holds for NLDN and Meteorage observed flashes, although the mean flash duration slightly increased for Meteorage records compared to NLDN records due to few long-lasting CG flashes with duration exceeding 1 s (see Figure 12 of Section II.1.2 and Figure II.2.6).

ISS-LIS flashes comprise similar amounts of events per flash over the US and over Corsica. The statistics of matched ISS-LIS yield 49.6 and 56.7 events on average for flashes with NLDN match and for flashes with Meteorage match, respectively (see Figure 7 of Section II.1.2 and Figure II.2.9). The difference can be attributed to few nighttime flashes with more than 250 events over Corsica. As mentioned before, extreme cases significantly alter the statistics of a distribution with small sample size. The mean event number per LIS-only flash is almost identical over Corsica and the SE US. The distribution and average of Meteorage and NLDN strokes+pulses per flash are nearly identical over the US and Corsica region if flashes with and without coincident ISS-LIS observation are compared (see Figure 10 of Section II.1.2 and Figure II.2.8). It was yet observed that a Meteorage flash contains on average slightly more strokes+pulses than an NLDN flash (3.8 versus 3.5).

A first result analyzing flash energetics showed that the gap found in the distributions of ISS-LIS mean amplitude count per flash over Corsica is not present for ISS-LIS amplitude counts over the SE US. It was, thus, caused by the limited number of cases and potentially a binning issue. Average and most frequent values of mean LF amplitudes per flash yield mostly similar values for both matched and unmatched flashes within both studied regions with approximately 8.0 kA. Only for the Meteorage flashes not detected by ISS-LIS a higher average of the mean measured LF current per flash is obtained with about 11.6 kA. It might be caused by low level CG flashes that produced high LF radiation and that were not detected by ISS-LIS. The distribution shapes of NLDN and Meteorage mean and maximum stroke/pulse currents per flash are overall similar. The difference in the average over all flashes between the unmatched Meteorage flashes and the other categories of LF flashes appears to be even more pronounced for the maximum LF currents per flash, e.g., with on average 18.1 kA and 12.5 kA for Meteorage-only and NLDN-only flashes, respectively.

In conclusion, remarkably similar flash statistics and characteristics are found for ISS-LIS observations as a reference over the two regions. Hence, it can be expected that Meteorage and NLDN records feature similar statistics, too. Indeed, the similarities between the LF networks are sufficient to assume that the GEO lightning pseudo-observation generator can be developed using concurrent NLDN and GLM observations, and can then be applied to simulate pseudo MTG-LI data from Meteorage records in France. Meteorage flashes were on average slightly more extended, lasted longer duration, and had more strokes+pulses than NLDN flashes. However, the differences are mainly attributed to few large CG flashes that were observed over Corsica with up to almost 150 km extent and 55 strokes+pulses. Measured LF currents were almost identical for the flashes with and without coincident ISS-LIS observation. An interesting finding for further research pinpoints that flashes detected by Meteorage but missed by ISS-LIS have significantly higher maximum currents over Corsica than the matched flashes, however, NLDN-only flashes have slightly lower LF currents than the flashes with ISS-LIS match.

II.2.5 Chapter conclusions

In this chapter, the domain is defined for all analysis of US LLSs within this thesis. A climatology of NLDN and early GLM data was used to identify regions with high lightning activity. Eventually, the almost 700,000 km^2 domain was selected between 30°N and 35°N latitude and from 95°W to 82°W longitude. It is situated solely on the continent. Hence, tools as our GEO lightning pseudo-observation generator developed for this domain may lead to less accurate results in other parts of the world, especially if the study domain would include wide parts of ocean.

This chapter applies then the methods developed in Erdmann et al. (2020a) (see Chapter II.1). This time, ISS-LIS records are compared to US LF NLDN observations. Good spatial and temporal agreement between coincident ISS-LIS and NLDN flashes is found, as was also the case for ISS-LIS to Meteorage and SAETTA flashes. The median spatial and temporal offsets of matched flashes are in all cases smaller than the ISS-LIS pixel size and integration time frame, respectively. Statistics of ISS-LIS (NLDN) flashes with and without NLDN (ISS-LIS) match are compared, with special focus on the difference between NLDN to ISS-LIS flash characteristics that is then also compared to the difference between Meteorage and ISS-LIS flash characteristics. This analysis reveals that ISS-LIS (NLDN) flashes with long extent, long duration, and high number of events (strokes+pulses) are more likely detected by NLDN (ISS-LIS), too. The NLDN flash energetics exhibit similar statistics for flashes with and without ISS-LIS

match. Matched ISS-LIS flashes exhibit higher amplitude counts than the ISS-LIS-only flashes, indicating that optically bright flashes often produce enough LF radiation to be detected at the ground. Similar results were reported by Erdmann et al. (2020a) for ISS-LIS and Meteorage observations. All in all, the results of Chapter II.1 and of this chapter support the conclusion that Meteorage and NLDN observe flashes similarly as demonstrated in a statistical analysis. The development of the GEO lightning pseudo-observation generator thus assumes that an algorithm trained with NLDN data can use Meteorage data as input without any prior data adjustment. The next chapter examines GLM observations, and then the data generator based on NLDN and GLM will be explained.

II.3 | Operational Lightning Observations from the Geostationary Orbit

In analogy to the previous chapter, characteristics of lightning observations are investigated for NLDN and a space-borne instrument. This time, however, the satellite sensor is not the LEO ISS-LIS instrument but the Geostationary Lightning Mappers (GLMs) on the GOES E/W (former GOES 16/17) satellites. GLM data became available in late 2018, and have since been utilized within this thesis. This section includes the same analyses as presented for NLDN and ISS-LIS previously. The same NLDN data is used and compared to the new GLM lightning observations. The section starts with an introduction of work recently conducted using the operational GLM. It also includes a brief review of the benefits of using total (CG+IC) lightning observations. Then, our results are presented. The relative detection efficiencies are calculated, and flash physics and energetics are analyzed for both matched and unmatched both GLM and NLDN flashes. As can be seen in Figure II.3.1, this chapter focuses on records of the US LLSs GLM and NLDN. The resulting data base and knowledge of coincident GEO and LF lightning observations build the framework for our GEO lightning pseudo-observation generator that will be introduced in the following Section II.4.2. The training will use the characteristics as derived in this chapter.

II.3.1 State of the art

During the last decades, lightning observation and in particular the observation of total (CG+IC) lightning became more common for applications such as numerical weather prediction and assimilation. Total lightning activity is recognized as a valuable information to predict thunderstorms and in particular potentially harmful, severe events. Total lightning sensors on low Earth orbit satellites observe one storm for a maximum of 100 s. There are many studies using and validating the low Earth orbit Optical Transient Detector (OTD) (Boccippio et al., 2000) and LIS (e.g., Lavigne et al., 2019, Rudlosky et al., 2017, Bitzer et al., 2016, Thompson et al., 2014, Thomas et al., 2000, Erdmann et al., 2020a), therefore, this not-comprehensive literature review focuses on recent papers mostly concerning GLM, and studies of total lightning related to severe weather.

Former studies linked total lightning activity and severe weather events (e.g., Williams et al., 1999). Severe weather means, according to the US NWS, the presence of one or more of three phenomena: (i) hail at the ground at least one inch (2.54 cm) in diameter, (ii) winds of at least 50 kn (93 km h^{-1}), (iii) a tornado. Hence, the definition does not include rain rates. Heavy rain usually accompanies severe thunderstorms and it may lead to dangerous flash floods. Schultz et al. (2011) demonstrated the value of total lightning data here: Total lightning, as now available

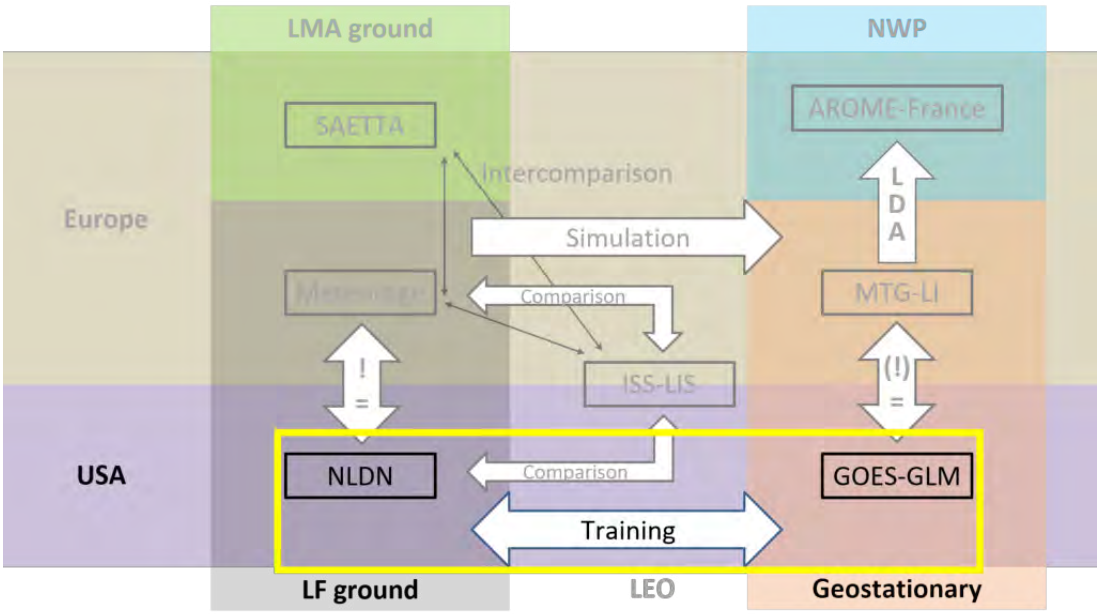


Figure II.3.1: Flow chart of work phases during this thesis as Figure 1. The LLS and phases detailed in this chapter are highlighted in the yellow frame, while other phases are shaded.

over the entire CONUS from GLM, led to higher probability of detection, lower false alarm rate, and longer lead times than using only CG strokes when relating lightning trends to severe weather occurrence. Herzog et al. (2014) presented a climatology of total lightning for three different LMAs in the US. Total lightning expressed as flash rate and FED showed moderate to strong correlations with radar-derived storm intensity attributes. Schultz et al. (2017) found increases in mixed-phase updraft volume and peak speed, and graupel mass (less significant) prior to North Alabama LMA total lightning jumps. The lightning jump corresponds to a rapid increase in the flash rate. Severe storms feature likely at least one lightning jump per hour, and changes in flash rates can indicate an enhanced probability for a tornadic storm (Steiger et al., 2007a,b, Rudlosky and Fuelberg, 2013). Although some parameters might not be resolved due to GLM spatial resolution storm scale information derived from total lightning observation can complement existing data especially for nowcasting applications and NWP. The cited studies used mainly local ground-based instruments. The GLMs and GEO LLS in general provide total lightning for a wider domain.

The GLMs sense cloud-top illuminations to provide optical observations of total lightning as single illuminated pixels (events), that are clustered to groups and flashes (Mach, 2020). They feature continuous observation of one region with near storm-scale spatial and high temporal resolution. Storms can be tracked during their entire lifetime in order to identify phases of lightning activity and changes of electrical (i.e., optical) characteristics with time.

Recent studies show that GLM performance changes geographically. Murphy and Said (2020) and Marchand et al. (2019) found higher DE relative to NLDN, GLD360, and Earth Network Total Lightning Network (ENTLN) over the southern and eastern CONUS than over the northern and western CONUS. The location within the GLM field of view (FOV) and incidence angle at the GLM influence the DE and correlation to ground-based instruments due to instrument characteristics, spatial resolution, and parallax effects (e.g., Bruning et al., 2019) that grow, as for all satellite products, towards the edges of the FOV. Furthermore, the ground type influences lightning seen by GLM as Rudlosky et al. (2018) analyzed that GLM flashes over the sea are on average larger, last longer, and appear brighter than flashes over land.

Rutledge et al. (2020) investigated the GLM storm-scale DE for normal and inverted-polarity storms with the Colorado and the North Alabama LMA as references. GLM DE varies with the geometric size of the flash with higher DE for larger flashes, and the cloud water path as a function of flash altitude and cloud water content whereby the DE is higher for a lower cloud water path. Light intensity at cloud top can drop below the GLM detection threshold for large cloud water path, as Rutledge et al. (2020) found minimum GLM DE for cloud water paths exceeding about 0.1 kg/m^2 in Colorado and Alabama. Hence, flashes at lower altitudes (i.e., below 7 km) and large cloud water content degrades the GLM DE. Marchand et al. (2019) and Zhang and Cummins (2020), who used GLM observations and Kennedy Space Center LMA data, noticed that flash size and flash duration are critical parameters for GLM DE. GLM tended to miss a significant amount of small and short duration flashes. As there is often an inverse relationship between flash rate and size, the results agree with the reduced GLM DE (relative to NLDN) for high flash rates reported by Murphy and Said (2020). The correlation of GLM and NLDN lightning jumps varied for their few study cases, however, the spatial differences in GLM DE were not considered during this analysis.

A higher GLM DE was noticed for positive LF pulses than for negative LF pulses, which is likely a combined effect of peak current (energy) and the flash altitude (Murphy and Said, 2020, Marchand et al., 2019). GLM DE also depended on time of day, flash energetics, cloud (electrical) structure and microphysics. A general result shows that GLM DE is higher during nighttime hours, whereas optical detection thresholds are lower than during the day. Velde et al. (2020) analyzed the optical energy measured by Atmosphere-Space Interactions Monitor (ASIM), ISS-LIS, and GLM. ASIM is sensitive to both the 777.4 nm-band (used by GLM and ISS-LIS) and the 337.0 nm-band and detects in general more optical emissions than both GLM and ISS-LIS. They found a slightly better pixel sensitivity for GLM than ISS-LIS. GLM detected 14 % of the ASIM optical sources versus the 2.5 % for ISS-LIS. As a result, the GLM flash area was larger and the flash duration longer than for the ISS-LIS data. Investigating the optical signal sensed by the GLMs and cloud structures, Peterson et al. (2020a) reported relatively bright and large area optical signals near the edges of convective clouds with light reflecting from cloud surfaces, and strong optical radiation over rainy stratiform clouds and high-altitude clouds such as anvils. Peterson et al. (2020b) further used the results to develop a GLM-based algorithm for partition of convective and non-convective clouds. The key parameters are frequency and ratio of horizontally propagating lightning per pixel, that is common only in stratiform clouds. Further application of GLM data includes data assimilation (e.g., Hu et al., 2020b, Fierro et al., 2019) and the validation of numerical lightning predictions (e.g., Blaylock and Horel, 2020, McCaul et al., 2020).

All cited studies demonstrate the potential of using total lightning locating GEO sensors in research and for operational applications. The gain of information can be of unprecedented value for forecasting and nowcasting thunderstorms. It is one main objective of this thesis and in particular the following sections to develop and evaluate the use of GEO lightning observations for that purpose.

II.3.2 Datasets and methods

The GLMs provide total lightning observations over the Americas and adjacent oceans. A suitable ground-based reference network is NLDN, that has been introduced and compared to ISS-LIS (Chapter II.2). This section uses the same NLDN data base and geographic domain as for the ISS-LIS versus NLDN comparison. However, in the present study both NLDN and GLM data is available continuously (ISS-LIS coverage is occasional during overpasses). To

reduce the large amount of data, ten days (see Section II.4.2 for details) are selected based on the analysis of the lightning activity. The selected days feature high lightning activity within the region. At least one day is selected of each month within the 6-month period from March to September 2018. Yet, the statistics comprise much more cases (almost a factor 200 overall) than the ISS-LIS versus NLDN comparison.

GLM data as provided by NOAA contains both event and flash level data. As some issues were detected for the ISS-LIS clustering algorithm (Erdmann et al., 2020a, see Section II.1.2), the NASA L2 clustering algorithm is compared to our clustering algorithm developed for both ISS-LIS and LF flashes. Figure II.3.2 shows 10 s of GLM and NLDN (green circle) lightning observations. It compares the GLM flash centroids for the same events (grey point) using NASA's L2 clustering algorithm (orange cross) and using the clustering algorithm developed during this thesis for ISS-LIS (blue plus). Our algorithm (the NASA L2 algorithm) uses 15 km and 0.3 s (16.5 km and 330 ms) as spatial and temporal merging criteria, respectively. NASA flash centroids are energy weighted locations, while our algorithm simply computed the event mean locations. This difference in the approach explains the displacement of some flash centroids. In general, both clustering algorithms perform very similarly. Slight differences exist as the spatial and temporal criteria are not exactly the same, e.g., for the flashes near 31.5°N in the beginning of the interval. NASA's GLM merging algorithm is detailed in Section II.4.2. At about 2.5 s and about 5 s in Figure II.3.2, our algorithm detects two small flashes (at 35°N) that are missing in the NASA L2 algorithm flashes. Those are single event flashes, and the NASA L2 algorithm currently treats single event flashes as noise (Mach, 2020). The GLM flash centroids are mostly in good agreement with the NLDN flash centroids, too. Overall, the NASA L2 algorithm provides trustworthy clustering of events. Eventually, this thesis uses the GLM flash level directly as provided in the data. Our clustering algorithm is still used for the NLDN flash level data.

Matching of GLM and NLDN flashes uses the algorithm developed by Erdmann et al. (2020a) (see Chapter II.1). Hence, the matching algorithm is the same as for comparing ISS-LIS to Meteorage (Chapter II.1) and to NLDN (Chapter II.2). Although the GLM resolution (pixel size) is coarser than that of ISS-LIS, GLM DE is significantly higher due to its apparent better sensitivity and its lower detection threshold of optical energy (e.g., Velde et al., 2020). Therefore, the main uncertainty for matching arises from the limited number of NLDN strokes and pulses rather than the GEO satellite sensor spatial resolution, and the same matching criteria are kept.

II.3.3 Detection efficiency

For the 10 days of interest, GLM featured a very high flash DE throughout the entire studied region with on overall (day and night, CG and IC) 87 % and almost 90 % if NLDN single stroke or pulse flashes are neglected). Figure II.3.3 provides the GLM DE relative to NLDN (and NLDN DE relative to GLM) in the SE US accumulated for ten selected days in 2018. High GLM DE in this region is in accordance with papers cited in Section II.3.1.

Table II.3.1 summaries relative DEs of both GLM and NLDN overall, for day and night, and different flash types. Slightly higher GLM DE is observed at night than during daytime and for IC than CG flashes. The diurnal differences are overall small and GLM can be considered as a consistently well performing LLS. NLDN averaged about 84 % of flash DE relative to all GLM flashes. The LF network detected about 82 % and 86 % of the GLM daytime and nighttime flashes, respectively.

As performed in previous chapters of this thesis (Chapters II.1 and II.2), the distance

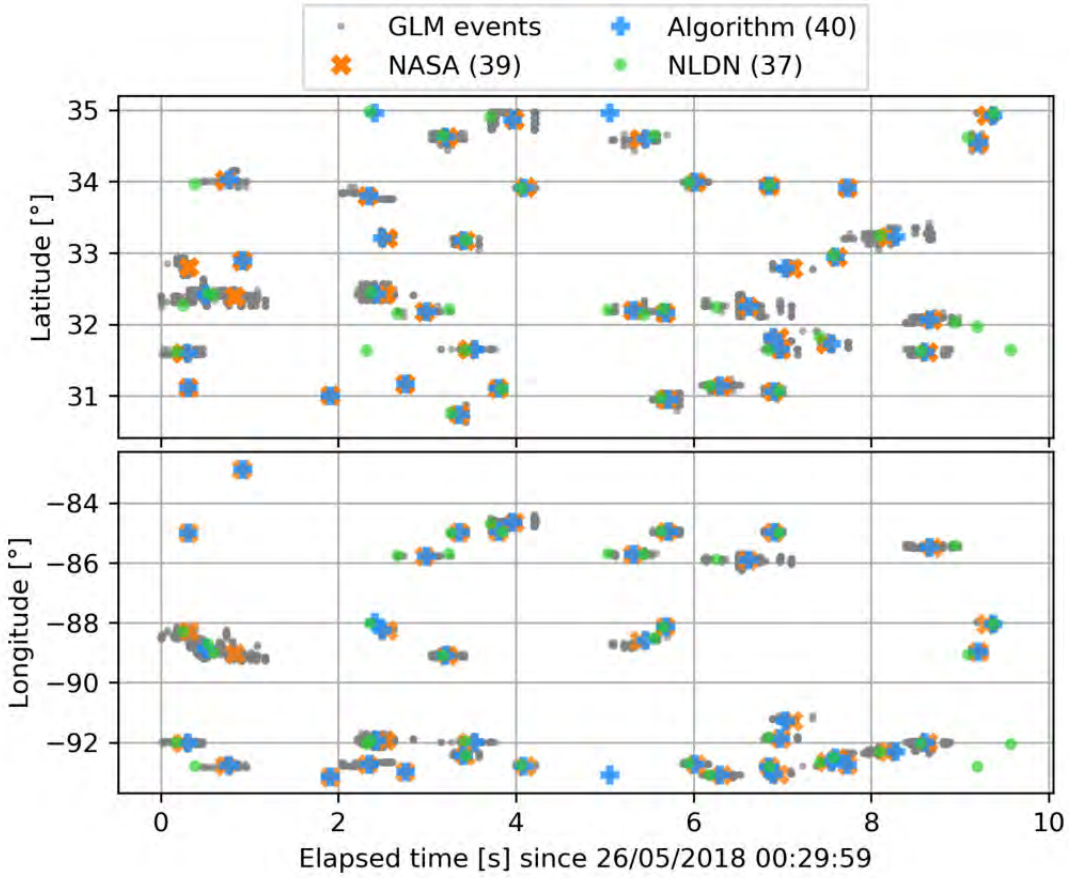


Figure II.3.2: A 10-second time series of the latitudes (top) and longitudes (bottom) of observed GLM events, flash centroids of NASA's L2 clustering algorithm, and the clustering algorithm developed during this thesis for ISS-LIS (as indicated in the legend). NLDN flash centroids are plotted for the same time period.

Table II.3.1: GLM and NLDN relative detection efficiency as different categories and with flash number in each category. Values in parentheses excluding the single event (stroke/pulse) GLM (NLDN) flashes.

	Overall	Daytime	Nighttime	IC-flash	CG-flash
GLM DE [%]	87.0 (91.2)	85.7 (90.4)	88.6 (92.2)	87.7 (93.1)	86.0 (89.3)
NLDN flash number	1,115,585 (715,193)	608,508 (381,591)	507,077 (333,602)	692,968 (364,086)	422,417 (351,107)
NLDN DE [%]	83.8 (83.8)	82.1 (82.1)	86.1 (86.1)	-	-
GLM flash number	1,132,051 (1,132,051)	638,014 (638,014)	494,037 (494,037)	-	-

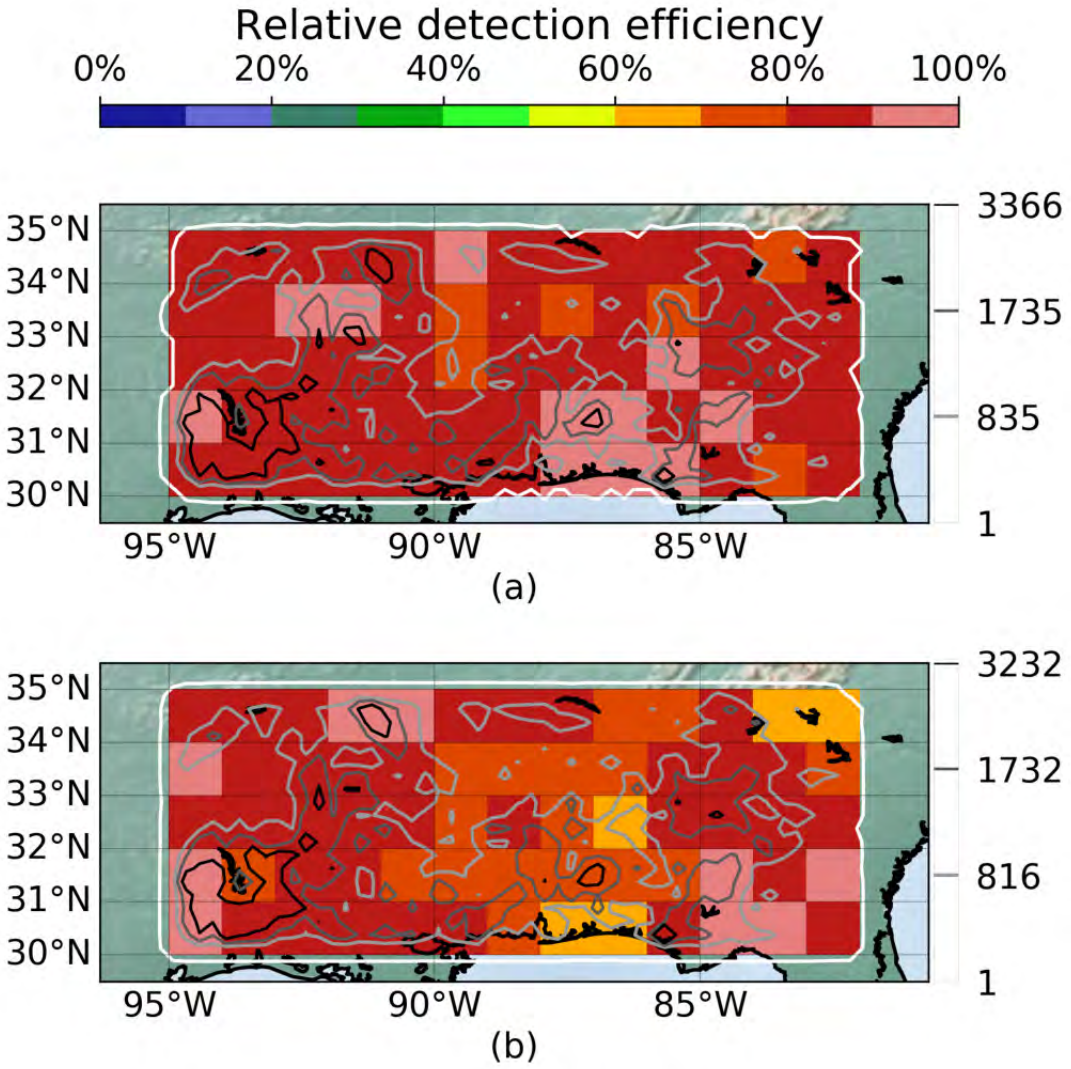
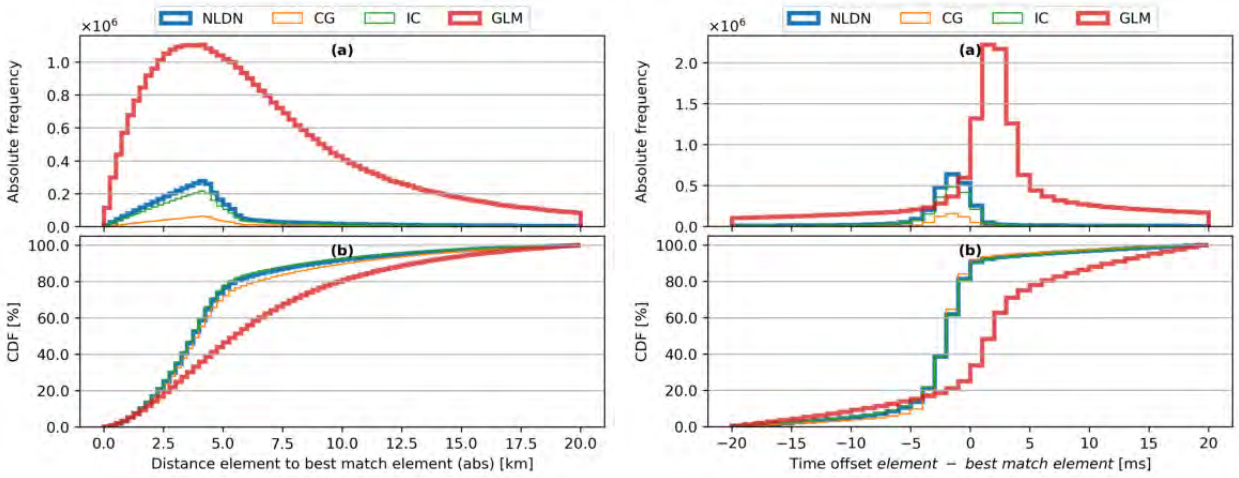


Figure II.3.3: Relative detection efficiency of GLM (a) and NLDN (b) with respect to the other instrument for $1° \times 1°$ pixels (colors). Grayscale contours mark the 0th (1 flash), 50th, 80th, and 95th percentiles of the flash number distribution per $0.25° \times 0.25°$ pixel.



(a) (b)

Figure II.3.4: Matching distance (a) and time offset (b) for the closest events (GLM) and strokes and pulses (NLDN) of matched flashes.

between the closest matched GLM event (NLDN stroke or pulse) and corresponding NLDN stroke or pulse (GLM event) are analyzed. Given a GLM event, the median distance to the closest NLDN stroke/pulse approximates 5.59 km and is significantly smaller than the GLM pixel size (about 8.7 km in the domain). A median distance of about 3.90 km (4.04 km;3.86 km) is found between a given NLDN (CG;IC) stroke or pulse and the closest matched GLM event. Mean distances yield 6.58 km and 4.67 km (5.03 km;4.56 km) given GLM and given NLDN (CG;IC), respectively. Figure II.3.4 presents the results of matching distances (a) and time offsets (b). The panels (a) and (b) include the histogram and the CDF of the distributions for both plots. The distribution peak for the distances is located at 3.25 km to 4.25 km and 4.00 km to 4.25 km (similar for given CG and given IC) given a GLM event and a NLDN stroke or pulse, respectively. These results feature the same order of magnitude as between ISS-LIS and NLDN (see Figure II.2.3), hence, despite the larger pixel size, the GLM locating accuracy relative to NLDN is similar to that of ISS-LIS. The difference in accuracy for the detection of NLDN CG strokes and IC pulses by GLM appears to be small, in particular, less significant than reported for ISS-LIS.

Photons emitted by lightning need approximately 100 ms to reach a GEO LLS, which is referred to as time-of-flight (TOF). The actual value varies with the distance between the emission and the sensors. The GLM L2 data do contain the real time when an event was recorded. Hence, GLM times are TOF corrected here, with a dynamic correction that depends on the geolocation of the optical emission. Figure II.3.4(b) shows the timing offset between matched GLM events and NLDN strokes and pulses. Obviously, the curves for GLM (red) and NLDN (blue) peak for different time offsets. The peak time offset is positive (1 ms to 3 ms) given GLM, and negative (−1 ms to −2 ms) given NLDN. Median and mean time offsets corroborate this finding. The median (mean) time offset yields 4.9 ms (41.2 ms) given a GLM event. Given a NLDN stroke or pulse (CG stroke; IC pulse), the median and mean time offsets approximate −1.9 ms (−1.9 ms; −1.9 ms) and −28.2 ms (−26.8 ms; −28.7 ms), respectively. Again, means are sensitive to the outliers of the distribution that includes matches with up to 1 s time offsets. GLM detected the events slightly later, but usually within few milliseconds, than NLDN detected coincident strokes and pulses. The median and peak time offsets are significant as the feature the same order of magnitude as the GLM integration time frame.

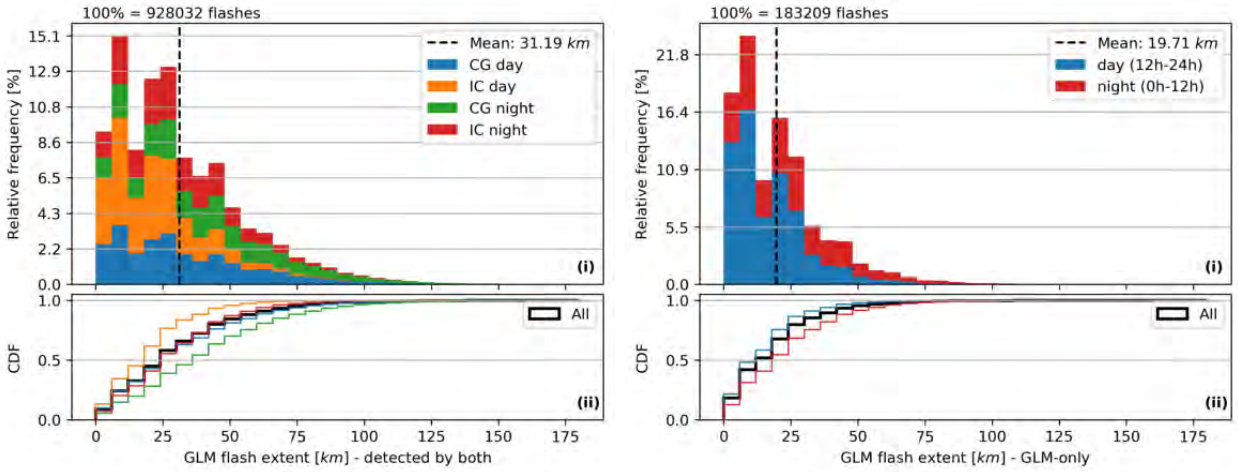
However, it is only few milliseconds higher than the time offset between ISS-LIS and NLDN (Section II.2.2). Blakeslee et al. (2020) reported that there is a peak time offset between GLM and ISS-LIS of about 1 ms to 2 ms which agrees well with our result shown in Figures II.2.3 and II.3.4. It might be an artifact of the relatively early version of GLM L2 data downloaded November 2018. NASA and NOAA updated the GLM data since the download, but it is not known whether event timings were adjusted.

II.3.4 Flash characteristics

The characteristics of the GLM and NLDN flashes are of importance for the simulation of GEO lightning pseudo-observations as NLDN flash characteristics will be related to GLM flash characteristics. In consistency with the analyses comparing ISS-LIS to both Meteorage and NLDN, matched and unmatched flash are separated. It should be noticed that the flash numbers for 100 % in the histograms of this section are slightly lower than flash numbers discussed in Section II.3.3. This section uses only those flashes with centroids inside a domain excluding the 0.3° latitude and longitude margin at each side of the full domain. This so-called buffer zone is applied to avoid excessive cutting of flashes near the domain borders. The width of the zone is chosen to approximate the mean extent of observed GLM flashes. The remaining flashes should thus be studied including the majority of their events and strokes+pulses. In detail, 1,111,241 (out of 1,132,051) GLM and 1,111,844 (out of 1,115,585) NLDN flashes remain for the statistics. This methodology is important for the next chapter where machine learning (ML) approaches are utilized to analyze flash characteristics. Artificially cut flashes might distort the ML analysis in Chapter II.4.

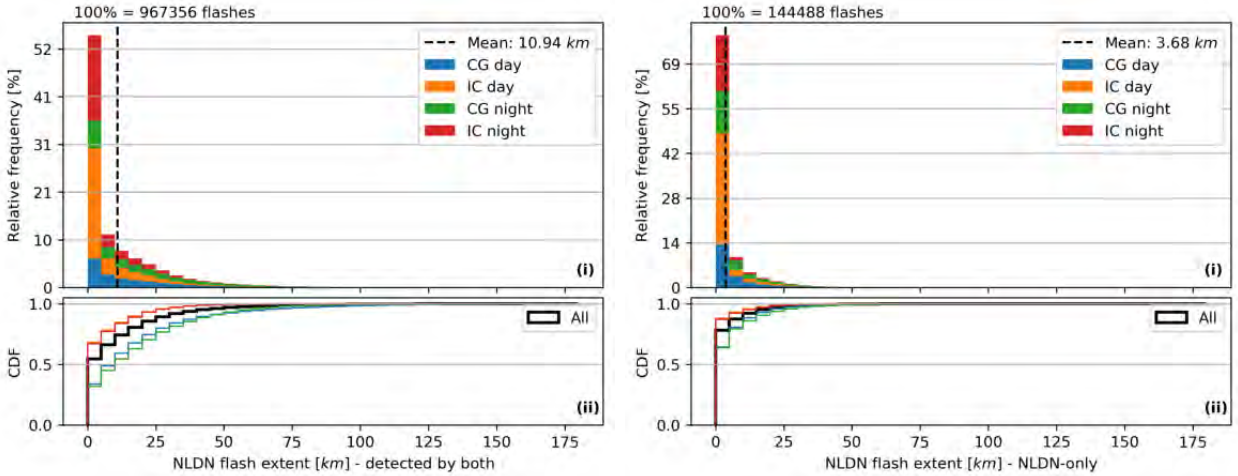
The mean flash extent of GLM flashes with match (31.2 km) is significantly greater than for GLM flashes that were not detected by NLDN (19.7 km). The distributions are shown in Figure II.3.5, with the histogram and CDF and the same color coding for daytime, nighttime, CG, and IC flashes as in the previous chapter. The figures of this section keep the same layout. It should be mentioned, that the smallest GLM flashes with two events are distributed over the first two bins as one bin equals 5 km and one GLM pixel has a side length of approximately 8.7 km in the studied domain. The smallest GLM flash extent was observed for daytime IC flashes (orange), and the largest for nighttime CG flashes (green). Statistics and the distribution of the GLM flash extent are similar to those reported for ISS-LIS (see Figure II.2.5) in spite of the about 200 times more GLM cases. NLDN flashes extended on average 10.9 km if with a coincident GLM observation and only 3.7 km if observed by NLDN only. The vast majority of NLDN flashes features an extent of less than 5 km, as seen in Figure II.3.6. NLDN CG flashes (blue, green) are larger than NLDN IC flashes (orange, red), with average extent of 17.1 km and 5.7 km, respectively. One can draw the same conclusions as for the LF and LEO comparison: matched flashes extend further than flashes without coincident observations. Indeed, GLM missed almost exclusively very small NLDN flashes with total extent (NS+EW) below 25 km. For both GLM and NLDN, few extreme flashes with extent over 200 km can occur (not shown).

Results for GLM flash duration are shown in Figure II.3.7. GLM flashes with NLDN match last on average 0.39 s, while the GLM-only flashes feature a mean duration of 0.23 s. The shortest GLM flash extent are daytime IC flashes and longest GLM flashes appear as nighttime CG flashes. The mean flash duration of GLM flashes is on average 0.07 s longer than for ISS-LIS flashes (see Figure II.2.7). This may be due to the lower sensitivity of ISS-LIS and to its limited viewtime. Figure II.3.8(a) for the matched (b, unmatched) NLDN flash duration demonstrates that almost half (about 70 %) of the NLDN flashes last less than 50 ms. Matched and unmatched NLDN flashes lasted on average 0.24 s and 0.08 s, respectively. The GLM



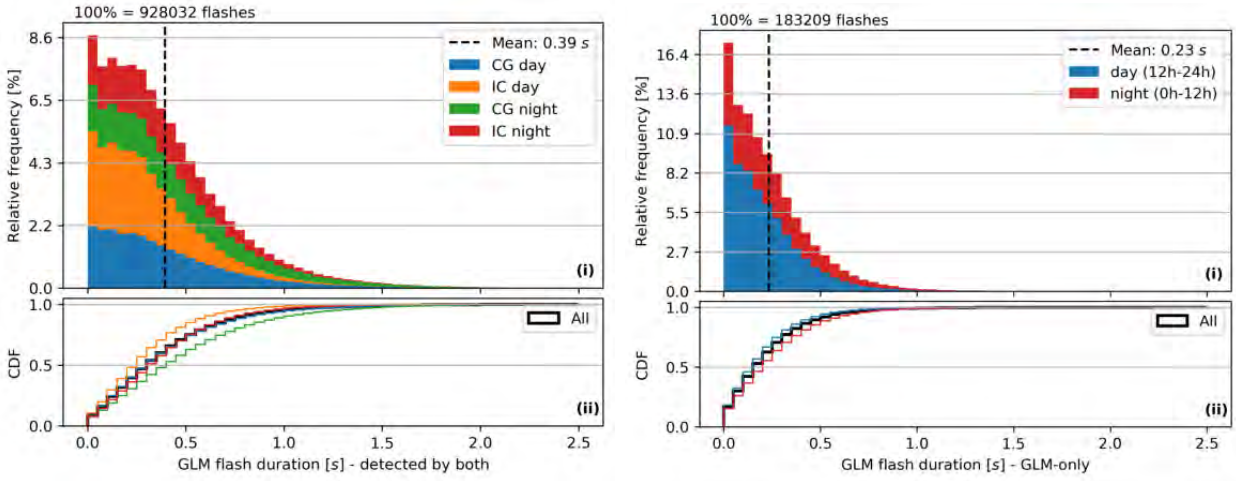
(a) (b)

Figure II.3.5: GLM flash extent of matched (a) and unmatched (b) flashes, with the histogram in (i) and the cumulated histogram in (ii). Colors show categories of day, night, CG, and IC flashes as indicated. The numbers are stacked in (i) for the 4 categories indicated in the legend. Note: No flash type specification (CG, IC) for GLM-only flashes.



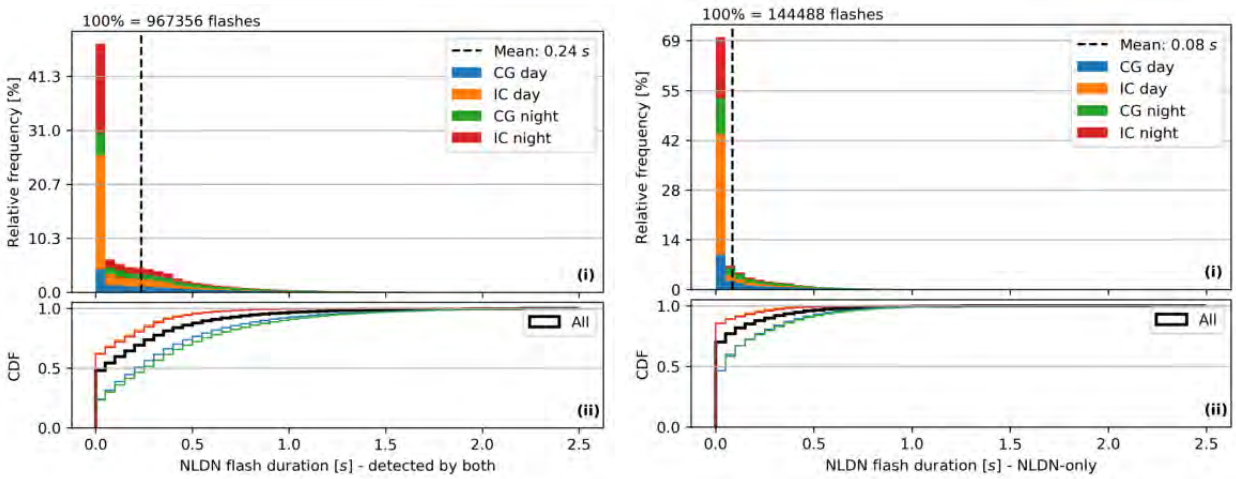
(a) (b)

Figure II.3.6: NLDN flash extent of matched (a) and unmatched (b) flashes, with the histogram in (i) and the cumulated histogram in (ii). Colors show categories of day, night, CG, and IC flashes as indicated. The numbers are stacked in (i) for the 4 categories indicated in the legend.



(a) (b)

Figure II.3.7: As Figure II.3.5 for the GLM flash duration.



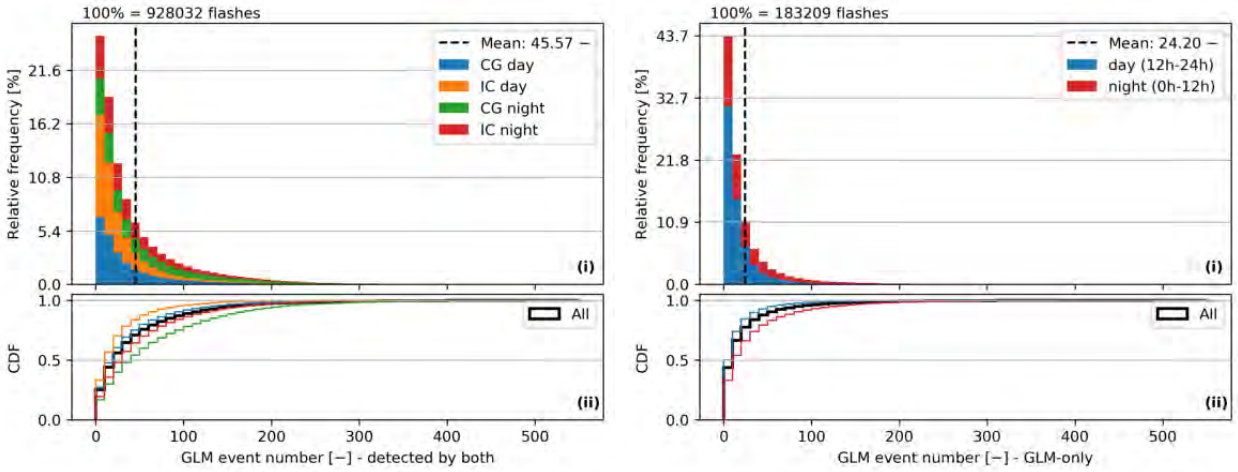
(a) (b)

Figure II.3.8: As Figure II.3.6 for the NLDN flash duration.

mainly missed particularly short duration NLDN flashes. NLDN CG flashes were detected for longer periods than NLDN IC flashes as indicated by the blue+green (CG) and orange+red (IC) bars, and average duration of 0.40 s and 0.11 s, respectively. In conclusion, the trends for the statistics of flashes with and without match are similar as seen for LF and LEO observation comparisons Sections II.1.2 and II.2.3. The longer lasting flashes are more likely detected by both instruments. As for the flash extent, it should be mentioned that few cases of extremely long-lasting flashes with duration of about 15 s were observed[1].

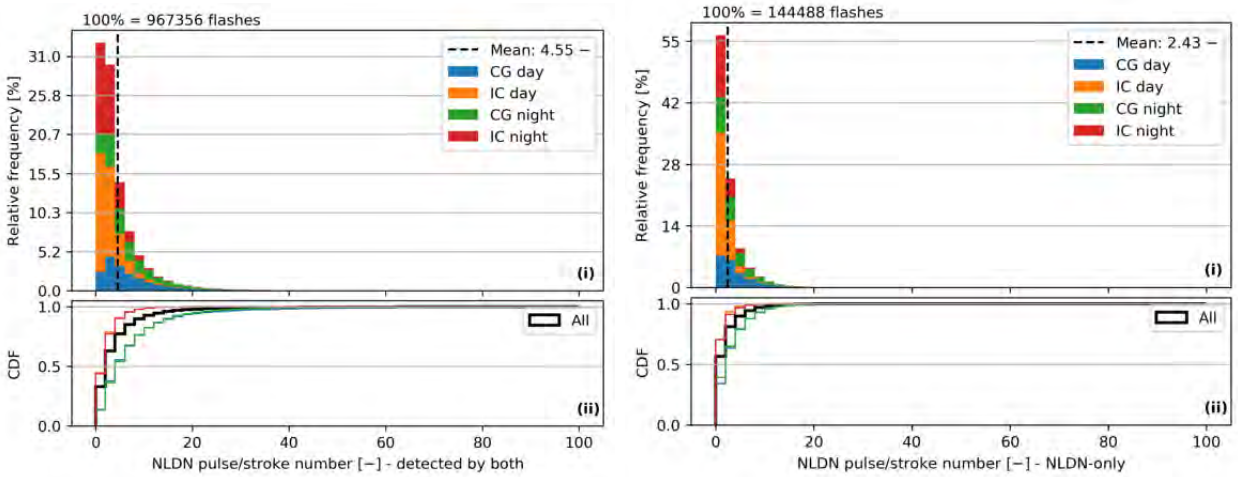
The distributions for the GLM event number and NLDN stroke+pulse number per flash are presented in the Figures II.3.9 and II.3.10. The distributions and statistics prove once again that larger flashes, i.e., with more events or strokes+pulses, are more likely detected by both LLSs. The mean event number per flash for GLM flashes with match equals 45.6, and for GLM flashes without match it equals 24.2. GLM nighttime flashes usually contain more events than daytime flashes. GLM IC have more events per flash than the CG flashes. The GLM event

[1]The large values are too rare to appear in the plotted distributions that are therefore cut for convenience.



(a) (b)

Figure II.3.9: As Figure II.3.5 for the GLM event number per flash.

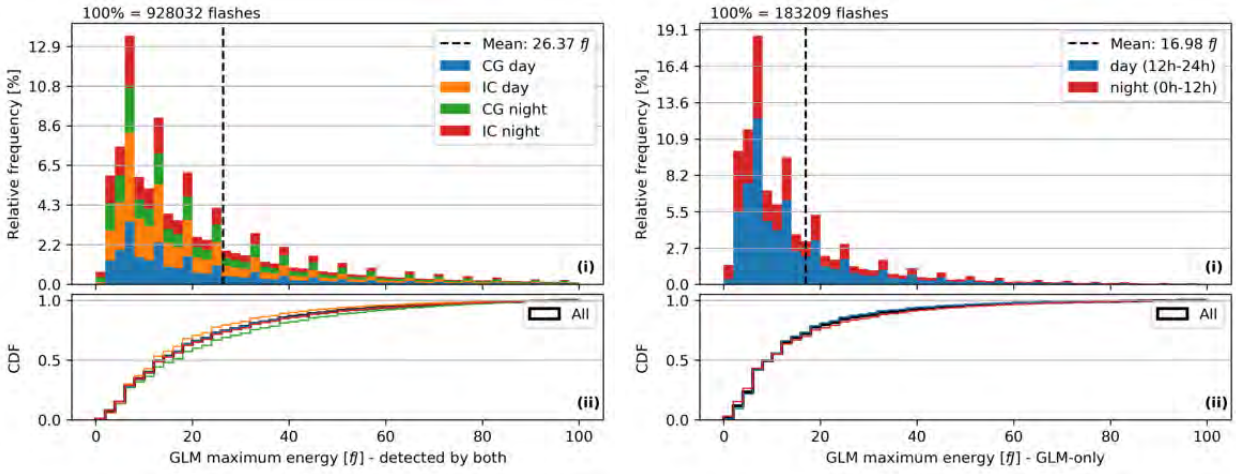


(a) (b)

Figure II.3.10: As Figure II.3.6 for the NLDN stroke+pulse number per flash.

number per flash is on average about 4 events smaller than that of ISS-LIS (see Figure II.2.9). However, one must consider that a GLM event (as pixel) is larger than an ISS-LIS event. The NLDN stroke+pulse number per flash is also almost twice as high for flashes with match (4.6) than for flashes without coincident GLM detection (2.4). NLDN CG (IC) flashes contained on average 7.3 strokes+pulses (2.5 pulses) and the CG flashes constitute the main part of the right tail of the distribution. The difference in event or stroke+pulse number between matched and unmatched GLM and NLDN flashes is more significant than observed for the LF and LEO comparisons, indicating that the GLM flash DE degrades statistically only for the smallest NLDN flashes.

The comparison of flash energetics includes the mean and maximum optical energy for GLM as well as the mean and maximum absolute LF current amplitude per flash. The mean energetics per flash (not shown) feature hardly any difference between matched and unmatched flashes. GLM flashes detected by both GLM and NLDN had mean optical energy of 5.1 fJ, and the GLM-only flashes emitted on average 4.7 fJ measured by GLM. The GLM tends to

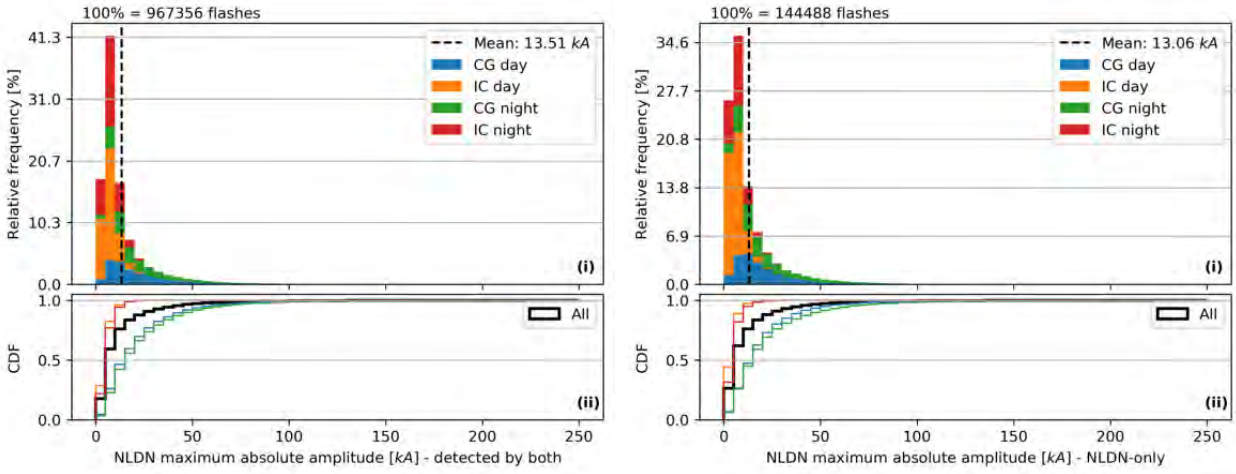


(a) (b)

Figure II.3.11: As Figure II.3.5 for the GLM maximum radiant energy per flash.

measure slightly higher mean optical energy during the day than during the night, possibly due to the higher threshold value during daytime. The lowest energy as recorded during the night is thus not in the distribution for daytime radiant energy, which likely leads to the increase of the mean. NLDN flashes with (without) coincident GLM flash exhibit on average 8.2 kA (9.2 kA) mean LF amplitudes. NLDN CG flash mean LF amplitudes are higher than for pure IC flashes. Some differences between matched and unmatched flashes exist for the maximum energetics per flash, as shown by the distributions for GLM and NLDN in the Figures II.3.11 and II.3.12. The GLM flashes with coincident NLDN flash appear brighter (26.4 fJ) than the GLM-only flashes (17.0 fJ). No difference was observed between GLM daytime and nighttime flashes, as also the CDFs in panels (ii) of Figure II.3.11(a and b) reveal. The observed higher peaks, that occur periodically above about 6.5 fJ could not be explained yet. NASA experts (Richard J. Blakeslee, Douglas Mach, and William Koshak) were approached in order to explain this phenomenon.

The NLDN maximum LF amplitude per flash is higher for CG flashes than for IC flashes as seen by the blue and green colored tail of the distributions in Figure II.3.12. This result is expected as CG return strokes emit strong LF radiation. When the distributions for matched (Figure II.3.12a) and unmatched (Figure II.3.12b) NLDN flashes are compared, similar shapes are found. Mean values equal 13.5 kA and 13.1 kA for the matched and unmatched flashes, respectively. Despite the fact that the NLDN flashes with coincident GLM flash featured larger extent, longer duration, and more strokes+pulses, their detected maximum LF amplitudes per flash are similar to those of NLDN-only flashes. Considering mean and maximum LF amplitude, no trend is seen to separate NLDN flashes that were detected by GLM from those that were not detected by GLM based on the measured LF amplitude. This finding is in particular important for Chapter II.4, as it indicates that GLM flash detection might not depend on the mean and maximum of the LF signals emitted by the discharge processes of the flash. On the other hand, optically bright flashes (considering the maximum optical energy) were more likely to produce enough LF radiation to be detected by NLDN than optically dim flashes.



(a) (b)

Figure II.3.12: As Figure II.3.6 for the NLDN maximum absolute amplitude per flash.

II.3.5 Chapter conclusions

A first analysis of GEO GLM lightning observations is presented in this chapter. The tools and methods developed for ISS-LIS are successfully adapted to process the GLM L2 data. Aiming at understanding the features of the GLM observations, they are compared to NLDN in a similar manner as performed for the comparison between ISS-LIS and NLDN observations (see Chapter II.2). As one result, the GLM flash DE relative to NLDN of 87 % is more than 20 % higher than ISS-LIS' flash DE relative to the same set of NLDN data, however, not the same cases are investigated[2]. Although reasons for the difference between GLM DE and ISS-LIS DE were not further studied, they might be related to sensor sensitivity, i.e., lower optical energy threshold defining a GLM event than a LIS event. However, the reasoning is likely more complex as flash rates and sizes should be considered (Zhang and Cummins, 2020), and also different cases were studied. The high flash DE of both GLM and NLDN is consistent throughout the studied domain in the SE US. It should, however, be noted, that only ten selected days are studied here. Nevertheless, the number of flashes analyzed in this chapter is almost 200 times higher than in the comparison of the LEO ISS-LIS and NLDN for the entire 6-month period.

Coincident GLM and NLDN flashes exhibit sub-pixel median geolocation offsets, however, GLM detects the events of the matched flashes few milliseconds later than NLDN detects the closest stroke or pulse. The GLM (NLDN) flashes with coincident NLDN (GLM) flash are statistically more extended, last longer, and comprise more events (strokes+pulses) than the flashes detected only by the given LLS. In addition, the GLM flashes with NLDN match are brighter than GLM-only flashes. The recorded absolute LF amplitudes as mean and maximum per flash are similar for NLDN flashes with and without GLM match. The results of this chapter, especially the characteristics of matched GLM and NLDN flashes, are essential for the development of our GEO lightning pseudo-observation generator in the next chapter. Correlations between the NLDN and GLM flash characteristics would promote the training of our GEO lightning pseudo-observation generator.

[2]The comparison of ISS-LIS and NLDN records analyzes only times with ISS-LIS overpasses. The GLM study comprises 10 days during the 6-month period.

II.4 | A Generator of GEO Lightning Pseudo-Observations

Operational LLLs covering France (Chapter II.1) and the US (Chapter II.2 and Chapter II.3) were introduced and analyzed in the previous chapters. GLM and NLDN detect a similar amount of flashes overall, however, they do not always detect the same flashes coincidently. In this chapter, the coincidently detected flashes and their characteristics are further processed. It is studied if any relation, i.e., correlation, exists between the characteristics of observed NLDN flashes and the coincident GLM flashes. The objective is to use NLDN observations to generate GLM pseudo-observations. Therefore, an algorithm, the so-called GEO lightning pseudo-observation generator, needs to be designed and trained on both NLDN and GLM observations.

This chapter describes the last step before generating pseudo MTG-LI data. As it can be seen in Figure II.4.1, the records of French operational LLS Meteorage has been compared to ISS-LIS and the LMA SAETTA lightning observations (Section II.1.2). Data of US LLSs such as NLDN and GLM can now be used to build, train, and test our GEO lightning pseudo-observation generator as detailed in the following Section II.4.2. Due to the similar performances of NLDN and Meteorage relative to the ISS-LIS observations, that served as a common reference, the trained GEO pseudo-observation generator will directly be applied to Meteorage data to simulate pseudo MTG-LI data. Pseudo MTG-LI FED constitutes the input for the LDA in Chapter III.4. The present chapter, however, considers only NLDN and GLM data. Indeed, the generator is built on NLDN and GLM records, and then is applied to simulate pseudo-GLM data that are compared to GLM observations in order to evaluate the generator performance for a test dataset independent of the training dataset.

II.4.1 Paper summary

The simulation of pseudo-GLM flashes from NLDN observations is performed in 2 parts as shown in Figure II.4.2 (center, yellow). First, our generator uses the GLM-NLDN flash data base with the coincident GLM and NLDN flashes and their characteristics. This part is referred to as target generator. It is built on statistical relationships between the NLDN flash characteristics and the GLM flash characteristics. This analysis is conducted using different machine learning (ML) approaches. They include simple linear regressions as well as more sophisticated ML models. The generator is trained using up to six NLDN flash characteristics and up to two GLM flash characteristics as pseudo-features (left, Training in Figure II.4.2). The second part simulates pseudo-GLM events using the simulated flash characteristics. Based on the pseudo-GLM events, Flash Extent Density (FED) is derived as the final product that can then be assimilated in an NWP model (in Chapter III.4 of this thesis). A verification with independent data is conducted on both the target generators, i.e., for simulated and observed distributions,

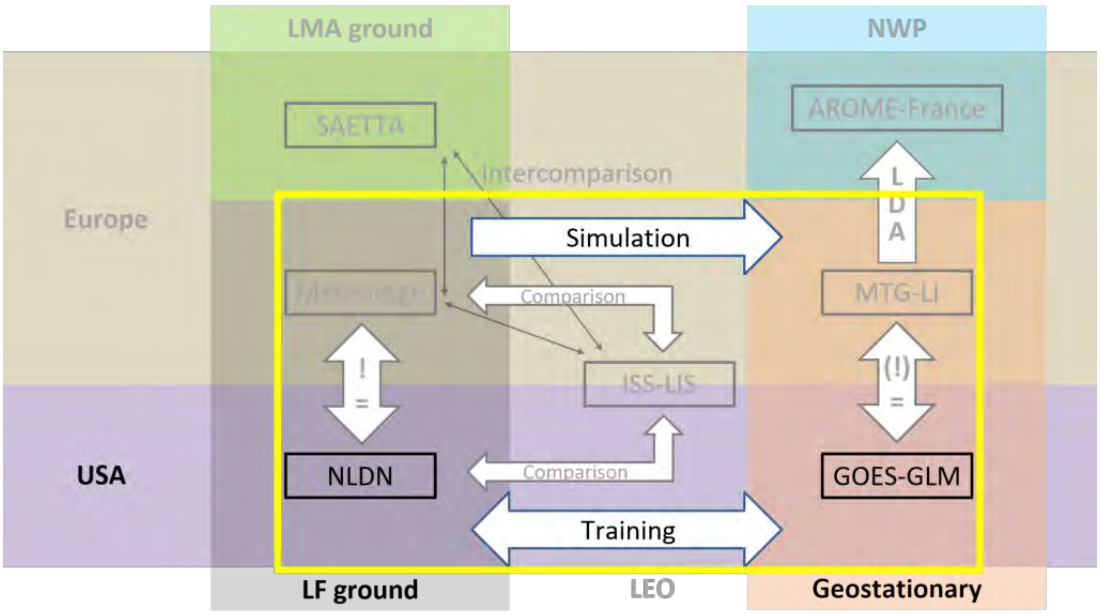


Figure II.4.1: Flow chart of work phases during this thesis as Figure 1. The LLS and phases detailed in this chapter are highlighted in the yellow frame, while other phases are shaded.

and final FED-based products (right, Test in Figure II.4.2), i.e., for hourly FED sum and electrified area. The recommended GEO lightning pseudo-observation generator applies a linear support vector regressor (linSVR). It uses all six available, normalized NLDN features and two normalized GLM pseudo-features applying a multi-step method. This generator provides a low bias of only 2 % for the hourly FED sum. However, there are times where the simulation differs from the observed hourly FED sum as the accumulated absolute difference can reach 25 %. The hourly electrified area is mostly underestimated by the recommended, linSVR-based generator by about 21 % when cumulated over the 72-hour test period. Some generators based on other ML techniques, e.g., Bagging or Multi-Layer Perceptron (MLP), could provide a better estimate of the electrified area with an accumulated absolute difference to the observation area of only 7 % to 8 %.

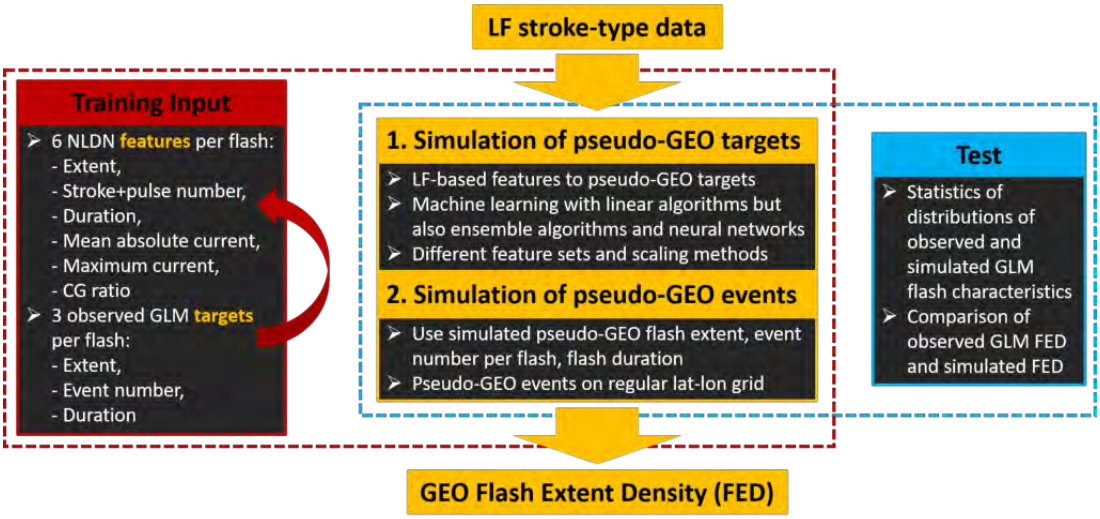


Figure II.4.2: Flow chart of the GEO lightning pseudo-observation generator.

II.4.2 Paper: A geostationary lightning pseudo-observation generator utilizing low frequency ground-based lightning observations

The following paper was submitted to the Journal of Atmospheric and Ocean Technology (JTECH) with the internal manuscript number JTECH-D-20-0160 (Erdmann et al., 2020b)ṪThe PDF file was created using the official JTECH template.

Reference: Erdmann, F., Caumont, O., and Defer, E.: A geostationary lightning pseudo-observation generator utilizing low frequency ground-based lightning observations, submitted to Journal of Atmospheric and Oceanic Technology.

Journal of Atmospheric and Oceanic Technology

A geostationary lightning pseudo-observation generator utilizing low frequency ground-based lightning observations

--Manuscript Draft--

Manuscript Number:	
Full Title:	A geostationary lightning pseudo-observation generator utilizing low frequency ground-based lightning observations
Article Type:	Article
Corresponding Author:	Felix Erdmann, M.Sc. CNRM, Université de Toulouse, Météo-France, CNRS Toulouse, FRANCE
Corresponding Author's Institution:	CNRM, Université de Toulouse, Météo-France, CNRS
First Author:	Felix Erdmann, M.Sc.
Order of Authors:	Felix Erdmann, M.Sc.
	Olivier Caumont
	Eric Defer
Abstract:	Coincident Geostationary Lightning Mapper (GLM) and National Lightning Detection Network (NLDN) observations are used to build a generator of realistic lightning optical signal in the perspective to simulate Lightning Imager (LI) signal from European NLDN-like observations. Characteristics of GLM and NLDN flashes are used to train different machine learning (ML) models, that predict simulated pseudo-GLM flash extent, flash duration, and event number per flash (targets) from several NLDN flash characteristics. Comparing statistics of observed GLM targets and simulated pseudo-GLM targets, the most suitable ML-based target generators are identified. The simulated targets are then further processed to obtain pseudo-GLM events and flashes. In the perspective of lightning data assimilation, Flash Extent Density (FED) is derived from both observed and simulated GLM data. The best generators simulate accumulated hourly FED sums with a bias of 2 % to the observation, while cumulated absolute differences remain of at least 22 %. A visual comparison reveals that hourly simulated FED features local maxima at the similar geolocations as the FED derived from GLM observations. However, the simulated FED often exceeds the observed FED in regions of convective cores and high flash rates. The accumulated hourly area with FED>0 simulated by some pseudo-GLM generators differs by only 7 % to 8 % from the observed values. The recommended generator uses a linear support vector machine model to create pseudo-GLM FED. It provides the best balance between target simulation, hourly FED sum, and hourly electrified area. p { margin-bottom: 0.25cm; line-height: 115%; background: transparent none repeat scroll 0% 0%; }
Suggested Reviewers:	Eric Bruning Associate Professor, Texas Tech eric.bruning@ttu.edu
	Kenneth L. Cummins cummins@atmo.arizona.edu
	Douglas Mach dmach@nasa.gov

A geostationary lightning pseudo-observation generator utilizing low frequency ground-based lightning observations

Felix Erdmann*

CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France & Laboratoire d'Aérologie, Université de Toulouse, CNRS, UPS, Toulouse, France

Olivier Caumont

CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

Eric Defer

Laboratoire d'Aérologie, Université de Toulouse, CNRS, UPS, Toulouse, France

* *Corresponding author*: Felix Erdmann, Erdmann.professional@gmx.de

ABSTRACT

Coincident Geostationary Lightning Mapper (GLM) and National Lightning Detection Network (NLDN) observations are used to build a generator of realistic lightning optical signal in the perspective to simulate Lightning Imager (LI) signal from European NLDN-like observations. Characteristics of GLM and NLDN flashes are used to train different machine learning (ML) models, that predict simulated pseudo-GLM flash extent, flash duration, and event number per flash (targets) from several NLDN flash characteristics. Comparing statistics of observed GLM targets and simulated pseudo-GLM targets, the most suitable ML-based target generators are identified. The simulated targets are then further processed to obtain pseudo-GLM events and flashes. In the perspective of lightning data assimilation, Flash Extent Density (FED) is derived from both observed and simulated GLM data. The best generators simulate accumulated hourly FED sums with a bias of 2 % to the observation, while cumulated absolute differences remain of at least 2 %. A visual comparison reveals that hourly simulated FED features local maxima at the similar geolocations as the FED derived from GLM observations. However, the simulated FED often exceeds the observed FED in regions of convective cores and high flash rates. The accumulated hourly area with FED>0 simulated by some pseudo-GLM generators differs by only 7 % to 8 % from the observed values. The recommended generator uses a linear support vector machine model to create pseudo-GLM FED. It provides the best balance between target simulation, hourly FED sum, and hourly electrified area.

1. Introduction

Lightning is defined as electrical discharges within the atmosphere, more particularly within and between clouds (intra- and intercloud, IC) or between clouds and the ground (CG). While cloud electrification and lightning initiation are still subject of studies, it is widely accepted that cloud ice and graupel are necessary to separate charges within clouds (e.g., Luque et al. 2020; Emersic and Saunders 2020; Lyu et al. 2019; Kolmasova et al. 2019; Takahashi et al. 2017; MacGorman and Rust 1998; Brooks et al. 1997). In particular, convection creates favorable conditions for lightning, and the updraft strength can be well correlated to the total lightning rate (e.g., Deierling and Petersen 2008). Ávila et al. (2010) found a high correlation between the occurrence of deep convection and lightning over land at a global scale. Hence, lightning is an effective tracer of convection.

The new generation of geostationary (GEO) satellites carries optical lightning sensors, among other instruments. The Geostationary Lightning Mapper (GLM) of the Geostationary Operational Environmental Satellite (GOES) R-series, the Lightning Mapping Imager (LMI) on board the Chinese Fengyun-4 satellites (Yang et al. 2017), and the upcoming Meteosat Third Generation Lightning Imager (MTG-LI, Dobber and Grandell 2014) will provide GEO lightning observations at a global scale. This satellite-based, large-scale, continuous observation of lightning offers new information for climate monitoring and studies. In addition, the assimilation of GEO lightning data in the Numerical Weather Prediction (NWP) can help improving the initial state of the model. Most of recent lightning data assimilation studies use gridded Flash Extent Density (FED), for example Allen et al. (2016); Fierro et al. (2019).

To assimilate new observation types in NWP models, e.g., MTG-LI records in AROME, it is desired to develop an assimilation scheme prior to the instrument launch and data availability. The simulation of appropriate realistic pseudo-observations precedes the development of any assimilation scheme, especially when the sensor is not yet in operation. Such synthetic observations can be derived from existing GEO sensors over other regions, i.e., GLM, and ground-based Lightning Locating Systems (LLSs). In addition, Low Earth Orbit (LEO) missions such as the Lightning Imaging Sensor (LIS) on the Tropical Rainfall Measurement Mission (TRMM) satellites (e.g., Christian et al. 1999; Cecil et al. 2005) and on board the International Space Station (ISS) (Blakeslee and Koshak 2016; Blakeslee et al. 2020) provide space-based lightning observations. One can also use ground-based networks, e.g., the National Lightning Detection Network (NLDN) (e.g., Cummins and Murphy 2009), Météorage (e.g., Schulz et al. 2016; Erdmann et al. 2020), and Lightning Mapping Arrays (LMAs) (e.g., Rison et al. 1999; Thomas et al. 2004; Coquillat et al. 2019). While the satellite sensors detect visible light of lightning at 777.4 nm, the ground-based networks are operated at different frequencies that match electromagnetic radiation emitted by different lightning processes. NLDN and Météorage use low frequency (LF) sensors that are most sensitive to discharge processes such as return strokes for CG flashes. Most LF networks can distinguish CG and IC signals. The CG flash detection (with return strokes) is usually reliable, whereas the IC flash detection efficiency increases within the network and for shorter baselines given one LF sensor type (personal communication, Stéphane Pedeboy, 2020/21). Global LF networks have lower detection efficiency and accuracy than national and regional LF networks (e.g., Nag et al. 2015). LMA stations sense very high frequency (VHF) signals of lightning leader propagation and allow for 3-dimensional (3D) channel mapping (e.g.,

Rison et al. 1999). Their drawback is the limited range. A LMA network provides coverage within a radius of typically a few hundreds of kilometers (e.g., Coquillat et al. 2019).

Biron et al. (2008) resampled TRMM-LIS data on a MTG-LI-like grid to assess the potential performance of the MTG-LI with emphasis on the influence of varying minimal detectable radiant energy. However, this method relying on LEO lightning data is not suitable for producing continuous pseudo-observations in the same area for operational applications because of the poor revisiting time. Stano (2013) demonstrated a simple method to create pseudo-GLM gridded products using LMA data. The pseudo-GLM data served to train forecasters on the use of GLM data products. GLM's Algorithm Working Group (AWG) investigated a transformation function that transforms LMA sources to optical lightning observations. The technique combines TRMM-LIS flash statistics and observed LMA flashes (Bateman 2013). The same method was applied by Schultz et al. (2016) to study automated storm tracking and lightning jump algorithms using GLM pseudo-observations. Höller and Betz (2010) presents a simple statistical model for transforming stroke-type data of the LF network LINET (Betz et al. 2009) to pseudo-MTG-LI optical events. The statistical relations were studied comparing LINET to concurrent TRMM-LIS groups. Then, they created a pixel matrix of the future MTG-LI and used TRMM-LIS statistics of radiance and event number per group to obtain pseudo-MTG-LI events. Their work aimed to propose a statistics-based method to create optical pseudo-observations of lightning from a given set of LF strokes. The available satellite lightning data solely emanated from the LEO TRMM-LIS mission, and in addition the number of cases was fairly limited (705 coincident flashes).

Recent studies assessing the GLM performance have shown that the DE varies within the field of view. GLM detects almost 90 % of the flashes in the south-eastern USA (e.g., Marchand et al. 2019; Murphy and Said 2020). The flash DE is statistically lower in other regions

like Colorado. Rutledge et al. (2020) showed that the GLM performance depends on the charge structure and the hydrometeor distribution. In particular, electrically "anomalous" storms led to degrading GLM flash DE.

This paper introduces in-depth techniques and results of creating GEO lightning pseudo-observations. The GEO lightning pseudo-observation generator is developed using NLDN records in the US and can be applied to all NLDN-like ground-based LLSs, e.g., Meteorage in France. The generator simulates the GEO lightning pseudo-observations on the flash level including events (and thus flash extent). FED grids can be derived from the generated pseudo-observations to serve as assimilation input data. This work prepares in particular the assimilation of pseudo-MTG-LI data in AROME in France. As MTG-LI will produce GLM-like data, and the French Meteorage network observes lightning similarly as NLDN in the US, the developed GEO lightning pseudo-observation generator can be used to simulate realistic pseudo-MTG-LI data.

The main objective of this study is the generation of a realistic GEO lightning FED field. It does not aim at reproducing individual flashes correctly, but the FED product. Therefore, the most important characteristics are the overall flash number and the flash extents. There is no direct dependency of FED on the flash duration and event number per flash, neither on flash energetics. The developed generator should provide a product that allows a usage of the simulated data similar to Fierro et al. (2019). Ultimately, our paper serves for presenting the methodology and preparing a specific assimilation study (not in the scope of the present paper). The presented GEO lightning generator is aimed at simulating MTG-LI FED over France. The application in our study is not intended for an operational use even though the developed algorithm could be used for operational application or for teaching forecasters and users.

Section 2 introduces both NLDN and GLM instruments. It also describes the dataset with coincident GLM and NLDN flashes. Section 3 explains in-depth the strategy to mimic GLM data from NLDN observations. This includes a 2-part GEO lightning pseudo-observation generator and different machine learning (ML) models to relate GLM and NLDN flash characteristics. Section 4 presents pseudo-GLM observations, their comparison to real GLM observations, and the evaluation of the 2-part generator. FED from real and pseudo-GLM observations is compared for the different ML-based generators. Finally, recommendations for suitable GEO lightning pseudo-observation generators are given.

2. Instruments and Data

GLM and NLDN make use of different lightning detection and locating techniques. This section introduces important specifications of both instruments and the studied dataset. It briefly describes the developed methods to match and compare GLM and NLDN observations, and to infer flash characteristics needed for training ML models.

a. Geostationary Lightning Mapper (GLM)

The GLM is an optical sensor on board the GOES R-Series (currently GOES-16 at 75°W and GOES-17 at 137°W). This study uses the GOES-16 GLM data only. The GLM detects total lightning including IC and CG during day and night. Although it cannot directly distinguish IC from CG signals, Koshak and Solakiewicz (01 Jan. 2015) show that some ICs and CGs can be statistically differentiated. Especially due to the difficulty of the detection of daytime lightning against bright, sunlit clouds, thresholds and filters are applied to separate the lightning optical signal from background and other light sources. Lightning is detected in a narrow (1 nm) band centered at the 777.4 nm oxygen line in the near infrared. The

wide field-of-view (FOV) image is focused on a high speed Charge Coupled Device (CCD) focal plane with a nearly hemispheric FOV coverage (1372×1300 pixels). The variable pitch pixel CCD allows for resulting pixels of about 8 km at nadir and only 14 km at the edge of the FOV (Goodman et al. 2013). Images are produced continuously and in time frames of 2 ms.

NASA's GLM lightning data algorithm produces Level 2 data with lightning information as events, groups and flashes. The x,y-coordinates of the focal plane are transformed to latitude and longitude coordinates of an estimated cloud top ellipsoid (with a height of 14 km at the equator and 6 km at the poles). Bruning et al. (2019) describes the effects of using this ellipsoid on GLM parallax with respect to any known ground-relative reference. GLM events are single illuminated pixels that pass the optical filters and are thus identified as lightning signals. Their location is defined as the center of the illuminated pixel. Adjacent events observed in the same 2 ms time frame are merged to form a group. Next, groups are combined into flashes. NASA's clustering algorithm uses a Weighted Euclidean Distance (WED) with spatial (16.5 km in latitude and longitude) and temporal (330 ms) constraints. Two groups with a WED of less than one are assigned to the same flash. The WED criterion is tested for pairs of events with one event in each group (Mach 2020). GLM data before 26 September 2018 need a time-of-flight (TOF) correction that takes into account the time lightning photons need to travel from the cloud tops (approximated at 10 km of altitude) to the GLM orbit. Our study applies a dynamical TOF correction with values ranging from 122.8 ms to 124.9 ms in the region of interest.

The reader is referred to Goodman et al. (2013), the GLM Product Performance Guide for Data Users (Koshak et al. 2010), and Goodman et al. (2012) for further information.

b. The National Lightning Detection Network (NLDN)

The NLDN (Cummins and Murphy 2009) consists of more than 100 Vaisala LS7002 ground sensors in the contiguous US (CONUS). It detects low-frequency (LF) electromagnetic signals generated by fast lightning discharges such as return strokes or by intracloud components. Due to a combination of magnetic direction finding and time-of-arrival techniques, only two sensors are needed to construct the horizontal location (latitude and longitude, no altitude) and time of a signal. NLDN locates total lightning, including CG and IC discharges. According to Vaisala (2013), up to 95 % and better than 50 % of all CG and IC lightning, respectively, is detected. Zhu et al. (2016) found that one third of 153 IC pulses were detected by NLDN, and 86 % were classified correctly. NLDN detected 92 % of 367 return strokes, and also 92 % were correctly classified as CG. The median location accuracy approximates 250 m for CG strokes in the interior of the network. Lightning can be located at long range (1500 km), but the location accuracy in the interior of the network is significantly higher than outside. NLDN measures also the peak amplitude of the LF source. NLDN data used in this study include time (resolved at 1 ms), the location as latitude and longitude, the peak amplitude [kA], the polarity, the type (CG or IC) of the LF source, and quality parameters, e.g., the location error ellipse axes. Although Vaisala merges strokes to flashes (within 10 km and 1 s), the raw stroke-pulse-type data is used in this work.

c. Database for the current study

The general dataset consists of six months of GLM and NLDN records, from March 15, 2018 to September 15, 2018. NLDN data were provided in a region between 30°N and 35°N and 95°W and 82°W.

In order to handle the large amount of GLM data and hence to limit the data processing time, a reduction of the 6-month dataset was necessary. The complete lightning dataset is studied to identify lightning-active days (start and end at 00 UTC), defined by the number of GLM flashes and the number of GLM events. Ten days with significant lightning activity and different storm types during both day and night are selected. Table 1 summarizes the number of GLM events and flashes as well as NLDN pulses and strokes and flashes recorded in the region during each of the ten selected days. Table 1 also states the dominant weather situation during each of the ten days. At least one day per month is selected to represent possible climatological differences of the lightning within the region. All further analyses use these ten days in order to reduce the immense amount of GLM event scale data. The resulting dataset comprises 1,133,671 GLM flashes and 1,115,675 NLDN flashes (NLDN flash level data as in Erdmann et al. 2020). Missing data is identified through an analysis of instrument activity during 20 s time windows equal to those of the GLM L2 data files. The amount of flashes is reduced to 1,132,051 GLM flashes and 1,115,585 NLDN flashes due to possible[1] short periods of instrument inactivity. Hence, the difference in the number of observed flashes is less than 2 % of the flash counts, and both instruments operated continuously during the selected days. As the effect of downtimes of an instrument can be disregarded, the following analysis uses all available data. Three among the ten days are chosen to test the ML models with uncorrelated data and to assess the variability in the results (test days). The test days (07 April 2018, 26 May 2018, and 31 July 2018) feature both thermally driven convection and dynamic forcing at air mass boundaries. In the

[1]We do not know instrument downtimes from the data. Data may come with flags, but they do not give reliable information about the instrument status. We used a two-step approach to identify downtimes; (i) the DE is less than 50 % and (ii) the number of flashes observed is less than 10 % of the reference LLS.

following, the different weather regimes with different lightning activity are briefly described for the test days as the final FED product is in fact only analyzed for these three days.

For instance on 07 April 2018, the weather was dominated by a major cold front that traversed the region from northwest to southeast on 07 April 2018. Temperatures dropped by about 10 K behind the front. The strong dynamic forcing caused a mesoscale convective system (MCS) with linear structure. This system produced the vast majority of flashes observed during the test period of 07 April 2018 until it left the studied region at about 12 UTC.

26 May 2018 was characterized by relatively warm surface temperatures with slightly decreasing temperatures from west to east within the region. Moisture was induced into the region by a weak tropical depression over Cuba and later southern Florida. Convection occurred mainly in the local afternoon as a result of surface heating. Well defined cells formed and propagated slowly southward in the cyclonic flow.

Daytime temperatures widely exceeded 30 °C and remained at about 25 °C at night within the region on 31 July 2018. Moisture was advected into the region from the Gulf of Mexico while a dry line approached from the northwest. A multicell storm cluster formed in the convergence zone at local nighttime and propagated eastward driven by short baroclinic wave aloft. The second peak of lightning activity results from thermal convection in the eastern portion of the region before the dry air moved in and inhibited further convection.

d. Data processing algorithms – Flash scale data and identification of matches

NLDN and GLM observe lightning independently of each other. The comparison of the two LLSs needs, however, coincident observations. This work uses the matching algorithm as introduced by Erdmann et al. (2020). Coincident observations are defined at the flash

scale for flashes within 20 km and 1.0 s. The criteria are tested for any pair of events and pulses/strokes. Two parent flashes are matched if one event (pulse/stroke) meets both the spatial and the temporal criteria to any pulse/stroke (event) of the given flash. The algorithm does not analyze the flash mean position but the event and pulse/stroke locations. NLDN flash scale data are retrieved using the algorithm developed by Erdmann et al. (2020) for Meteorage records in France. Hence, pulses/strokes are merged into a flash if they occur within both 20 km and 0.4 s. The resulting NLDN flashes are compared to GLM observations and validated through a visual inspection of NLDN pulses/strokes for three different days. In most cases the algorithm performs well relative to the GLM flashes, in particular, with the selected spatial and temporal merging criteria. A more thorough validation of the algorithm would require the use of LMA-like mapping of all lightning flash channels.

GLM flash level data are included in the GLM L2 science data. Mach (2020) found recently that NASA's GLM clustering algorithm was quite stable for different spatial and temporal merging criteria (mainly for storms with flash rates below about 40 flashes per minute). In comparison, our algorithm is applied to cluster GLM events by using 15 km and 0.3 s as spatial and temporal criteria, respectively. A very high degree of accordance is found between NASA's GLM L2 and our algorithm clustering the same GLM events (tested for 1 hour on 26 May 2018 with 15,505 flashes by our algorithm and 15,303 flashes by NASA's operational algorithm; difference mainly due to single event flashes that are removed by NASA's algorithm but remained with our algorithm). Using group centroids to calculate distances between groups (Goodman et al. 2013; Mach 2020), NASA's L2 GLM flash scale algorithm succeeds to merge many events and to detect large flashes. Nevertheless, the GLM operational algorithm limits the maximum size of flashes due to computational restrictions,

however, such cases are rare and do hardly influence statistical results. In the following, GLM flash level data always emanate from NASA's L2 clustering algorithm.

The matching of GLM and NLDN flashes (for the 10-days dataset) leads to 948,872 GLM and 971,102 NLDN flashes with match. Some flashes are matched to more than one flash, and it happens more often that one GLM flash matches multiple NLDN flashes than vice versa. Considering the total number of GLM (NLDN) flashes, the relative flash detection efficiency is defined as ratio of flashes observed by both given and reference LLSs to the total number of flashes observed by the reference LLS. It yields 87.0 % (of 1,115,585 NLDN flashes) and 83.8 % (of 1,132,051 GLM flashes) for GLM and NLDN, respectively. Figure 1 illustrates the DE of both GLM and NLDN within the studied region, along with 2D density of observed flash centroids (gray iso-contour). The DE remains consistent within the entire domain. The local minimum in the North-East is caused by a low number of observed flashes for the two 1° x 1° pixels in Figure 1. The high DE of GLM agrees with the results of Marchand et al. (2019), who found the GLM DE relative to ground-based Earth Networks Total Lightning Network (ENTLN) flashes exceeding 80 % for most of the southeastern CONUS. They used 35 km and 330 ms as spatial and temporal matching criteria, respectively. Murphy and Said (2020) compared among others GLM and NLDN relative DE, matching flashes within 20 km between GLM flash centroids and the first NLDN pulse/stroke per flash and 200 ms between the flash time windows between the start and end times, and report similar DE values on the large scale in the southeastern CONUS. A new approach to the GLM flash DE and false alarm ratio (FAR) is introduced by Bateman and Mach (2020); Bateman et al.: Combining several ground-based networks to provide reference data, and using coarse matching criteria of 50 km and 10 min, they found flash DE of over 90 % and FAR just above 5 % for the GLM on GOES-16.

3. Methods

This section defines the concepts and the strategy to generate GEO pseudo lightning observations. The methods are designed to use NLDN data and evaluated using real GLM observations. MTG-LI will provide total lightning observations with similar data structure as GLM observations. It will also consist of events, groups, and flashes. Although MTG-LI's spatial resolution (4.5 km at nadir versus 8.0 km at nadir) and the temporal resolution (1 ms versus 2 ms) will be slightly higher than those of GLM, the methods presented here can still be applied to simulate MTG-LI observations.

a. Definition of the Flash Extent Density (FED)

Flash extent density is a gridded product that can be derived from every lightning observation. It needs a flash level data product with subsequent information about the extent of the flash. The extent might be an illuminated area (e.g., for GLM) or simply distances between point sources (e.g., for NLDN). FED is then defined on a regular grid, with a given size of each pixel. Observed flashes are superimposed and transformed to that FED grid. Pixels with any lightning observation are identified, while pixels with multiple observations (e.g., multiple NLDN pulses/strokes) are counted once per flash. This gives a grid of pixels with either lightning (value 1) or no lightning (value 0) for each flash. The FED product considers all flashes within a given time integration period and sums up the occurrence of flashes per pixel. Hence, the FED product can have values greater than or equal to one flash per pixel. It shows the spatial distribution of lightning activity within the given time period, such as the most electrically active regions. The propagation of convective cores can be then tracked over several successive time integration periods.

Deriving FED needs knowledge about flash locations or, in case of satellite observations, the positions of lightning data pixels. GLM products do not come with this necessary information. Therefore, the real GLM grid is reconstructed locating the centers of all events of the full half-a-year dataset. A time invariant real GLM grid is assumed. As individual pixels appear to wobble locally with time and do not appear on a regular grid; due to micro-vibrations of the satellite platform, spacecraft jitter, and variable pitch CCD; a k-means-clustering analysis is performed to identify the statistical mean location of each pixel center. Corner points of pixels are then defined as the mean locations between the centers of the 4 pixels adjacent to each point. It is assumed that corner points can be connected by straight lines in order to represent the pixel shapes. This assumption is not entirely true, as the regular CCD grid is projected on the Earth (more precisely, on the cloud top ellipsoid, Section 2.a), however, the FED should be less impacted by this assumption than by assuming a regular GLM grid. Shapes of GLM events do usually not match the FED grid pixel shapes. One GLM event with average side length of 8.7 km can overlap multiple FED pixels with side length of 5 km to some part. The fractions of the GLM event within each pixel of the FED grid are summed up while integrating over the time period. This slicing of GLM events reduces the effect of producing gaps or double counts of GLM pixel when transformed to the regular FED grid, as recently described by Bruning et al. (2019).

The FED in this study is calculated on a regular latitude longitude grid with an average pixel size of 5 km x 5 km. This pixel size approximates the expected MTG-LI pixel size at nadir. To obtain the regular latitude longitude grid, the distance of 5 km is transformed to latitudinal and longitudinal distance as of the pixel at the center of the study region. In the present study, FEDs are analyzed per 60 min time integration periods. The 1-hour period maintains information to locate tracks of convective cores and most electrified regions while

it is also long enough to capture several storms distributed within the full domain. There might be, however, multiple storms at one location during 60 min. The FED integration period can be changed as needed since our GEO lightning pseudo-observation generator simulates data at the flash level. The sum of multiple short FED periods is equal to the FED of a corresponding long period, but the computation of one long period is more efficient. Hence, this work uses hourly FED for illustration and computational reasons. It should be mentioned, however, that other FED time integration periods are currently under investigation, and the assimilation of MTG-LI will use a shorter FED time integration period.

b. Work flow – The simulation of GEO pseudo-observations of FED

The simulation of pseudo-GLM flashes from NLDN observations is performed in 2 parts. First, our GEO lightning pseudo-observation generator uses the flash database with the coincident GLM and NLDN flashes and their characteristics. This part called target generator employs ML techniques. It is based on statistical relationships between the NLDN characteristics (features) and the characteristics of the concurrently observed GLM flashes (targets). The target generator is detailed in the following section. This part is conducted using different approaches, which will be explained thereafter. They include simple linear regressions as well as more sophisticated ML models. The second part of the GEO lightning pseudo-observation generator, described in the last section here, simulates pseudo-GLM events using the simulated GLM flash characteristics.

1) Simulate pseudo-GLM flash characteristics

Coincident NLDN and GLM flashes are analyzed regarding their characteristics including the flash extent and flash duration (both GLM and NLDN) as well as the event number per flash (GLM) or pulse+stroke number (NLDN) per flash. The flash extent sums up the distance between the lowest and highest latitude (the North-South [NS] extent) and the distance between lowest and highest longitude (the West-East [WE] extent) of events or pulses/strokes of the flash. GLM flash extent relies on the pixel center position but do not include the pixel extensions (single pixel GLM flashes have an extent of 0.0 km). Single stroke or pulse NLDN flashes do also feature a flash extent of 0.0 km. The maximum and mean signal strengths, defined from the LF peak currents and radiant energies as measured by NLDN and GLM, respectively, are evaluated per flash to represent flash energetics. In addition, a CG stroke ratio is calculated for NLDN flashes dividing the number of CG strokes of the flash by the total pulse+stroke number. Previous studies (e.g., Thomas et al. 2000; Marchand et al. 2019; Erdmann et al. 2020; Murphy and Said 2020; Rutledge et al. 2020) found that characteristics of flashes observed by optical satellite LLSs depend among others on the flash altitude. A flash with multiple ground connections is more likely to propagate in the lower part of the clouds than flashes with less CG strokes. In total, there are 5 GLM flash characteristics (flash duration, event number per flash, flash extent, mean and maximum event radiant energy per flash) and 6 NLDN flash characteristics (flash duration, stroke+pulse number per flash, flash extent, mean and maximum LF amplitude per flash, CG stroke ratio).

The total flash database is used for training each model because the mean, median and standard deviation of each of the CG and IC distributions by day or night all approximate the

overall distribution (not shown). Then, linear regressions between any two GLM and NLDN flash characteristics were analyzed (not shown). It was found that GLM flash duration shows Pearson correlation coefficients R above 0.64 to NLDN flash duration and the number of strokes+pulses per flash. On the contrary, GLM event number per flash and GLM flash extent feature R of 0.08 to 0.43 to the complete set of features. Mean and maximum event radiant energy per GLM flash are not correlated to any NLDN flash characteristic on the flash scale and, hence, excluded from the ML targets. The remaining targets (GLM flash duration, event number per flash, and flash extent) are sufficient to derive the FED.

Building the GEO lightning pseudo-observation generator requires independent ML and test data for the ML model design and for the verification of the generated product, respectively. The ML data consist of seven days and the test data of the remaining 3 days of the full dataset (see Section 2.c and Table 1). Features (input data) of the ML data are the six NLDN characteristics, and targets (output data) are GLM flash duration, event number per flash, and flash extent. Feature and target sample sizes are given as the number of matched flashes detected by GLM and NLDN, respectively, and are not equal in general (Section 2.d). Since training the models requires the same sample size for the features and targets, two (or more) flashes matched to the same flash of the other LLS are merged, and characteristics of the merged flashes are combined. The resulting ML data consists of 672,794 flashes, each sample with six NLDN features and three GLM targets. The ML part further splits this set of ML data randomly into independent ML training and ML validation (or development) data at a ratio of 90 % to 10 %. The ML models are thus trained with 605,515 flashes. The ML validation data serves to calculate goodness-of-fit scores for each applied ML technique. Then, the different ML models are compared and the model parameters are tuned based on the scores (not shown).

The 3-day test dataset is used to evaluate each generator as a whole and not the specific ML technique applied. Indeed, as the objective of the generators is to create realistic GEO observations in regions where no GEO observations are available yet, e.g., over France, their evaluation must be conducted in the same observational configuration, i.e., without any truth on the occurrence of coincident LF and GEO records. As a result, the test exercise necessarily exploits both observed GLM and generated, NLDN-based pseudo-GLM datasets as two independent populations. Hence, the generator simulates one pseudo-GLM flash for each of the 340,712 observed NLDN flashes. The realistic character of the generated pseudo-observations is assessed using all 338,579 GLM flashes observed during the same 3-day period. Statistics of GLM targets and FED fields inferred from the generated pseudo-GLM flashes are compared to those from observed GLM flashes.

Simulated GLM targets are generated for each NLDN flash of the test data. The comparisons of statistics of the observed and simulated targets include the distribution mean, median, minimum, and maximum. The root mean squared error (RMSE) between characteristics of individual (simulated and real) GLM flashes is also computed, but only for the 295,313 NLDN flashes with GLM match (representing a GLM flash DE of 86.7 % for the test days). The evaluation makes in addition use of two statistical scores that are defined for the cumulative (in fact empirical) distribution functions (CDFs): The Kolmogorov-Smirnov statistic (KS, Massey 1951) and the Cramer-von Mises criterion (CvM, Anderson 1962) measure the distance between the observed and simulated CDFs of the targets. Both the KS and the CvM tests can verify the null-hypothesis that two samples belong to the same population.

2) Multi-step Approach

Targets of a multi-target ML training can be correlated, e.g., GLM event number per flash is strongly correlated to GLM flash extent with R of 0.74. To the best knowledge of the authors, none of Python's scikit-learn library (Pedregosa et al. 2011) models takes advantage of correlations between targets. Indeed, the so-called single target (ST) approaches do not consider correlations between targets, however, such correlations can help to improve the prediction of ML models. Borchani et al. (2015) summarize methods to deal with multi-target regressions and take advantage of correlations between targets. The paper compares the ST approach to multiple multi-target approaches, e.g., multi-target regressor stacking (MTRS), regression chains (RC), multi-output support vector regression, multi-target regression trees, and rule methods. Spyromitros-Xioufis et al. (2016) introduced the stacked ST (SST) and ensemble RC (ERC). These methods can be computationally complex with high memory costs (Mastelini et al. 2019). As Aguiar et al. (2019) state, choosing the most suitable approach needs previous testing and depends on the task. The methods cited here are computationally expensive.

The flow chart in Figure 2 shows a computationally efficient multi-target approach that simplifies the SST. As a starting point, there are NLDN features and GLM targets as input for the ML training. The approach combines four ST models (Figure 2a) of three classes (colored) for the training. The application case only uses the NLDN features as first input. Therefore, a multi-step approach is required. An application example is shown in Figure 2b. More details about our approach can be found in Appendix 5.

In summary, the multi-step approach modifies the input and thus the configuration of a ML model. It is a form of multi-target regression that can take advantage of correla-

tions between the ML targets. With six NLDN features and three GLM targets, several configurations might be realized. Figure 2b shows just one example. Section 4 will show whether the additional GLM pseudo-features can help to tune the pseudo-GLM simulation towards observed GLM data. Appendix B0.b summarizes the available sets of features used as different configurations for the ML models.

3) ML-based target generators relating NLDN flash characteristics to GLM flash characteristics

In the previous sections, the GEO lightning pseudo-observation generator was explained. Appendix B0.a briefly describes the different ML models used during the ML-based part of this generator. The ML-based algorithms relate NLDN flash characteristics to GLM flash characteristics in this work. Hence, all ML models are supervised models with the same training data. The models use Python's scikit-learn library (Pedregosa et al. 2011).

This study uses seven different ML model types: Multivariate linear regressions, third-degree polynomial regressions, Extra Trees Regressors as a form of a Random Forests, Bagging with k-nearest neighbor regressors, Multilayer Perceptron neural networks, linear Suppor Vector Regressors, and Histogram Gradient Boosting Regressors.

4) Applied Scaling Methods

Distributions of NLDN features exhibit more variable and less continuous shapes than the three target distributions (not shown). After testing five different scaling methods, this study (i) normalizes features to the [0,1] range with a Min-Max-scaler:

$$X_N = \frac{X - min(X)}{max(X) - min(X)} \tag{1}$$

where X is a data vector, $min(X)$ and $max(X)$ define the minimum and maximum of X, respectively, and the resulting normalization X_N is scaled from 0 to 1.

(ii) The targets are scaled with a common standard-scaler (also called z-value scaling) defined as

$$Z = \frac{X - mean(X)}{std(X)} \tag{2}$$

where X is a data vector, $std(X)$ means the standard deviation of X and the resulting standardization Z is centered around 0.

The Min-Max-Scaler is an alternative standardization method that is more robust to small standard deviations and for different feature ranges than the common standard-scaler (sklearn documentation). Some ML models performed well with unscaled data (i.e., direct input of data with physical units) that is also used as a reference input method during the ML part. All results presented in the following have been re-scaled to physical units.

Including the multi-step and scaling appraoches, 28 different configurations of each of the seven ML model types (Section 3.b.3) are tested giving a total of 196 different GEO lightning pseudo-observation generators. Abbreviation of ML model types and configurations can be found in Table A1 in Appendix B0.b.

5) Generate pseudo-GLM events

The studied domain is separated into regular adjustable size latitude longitude pixels that represent the pseudo-GLM pixel matrix. Any given latitude-longitude position is projected on that pixel matrix to determine the corresponding pixel and thus the shape of one pseudo-GLM event. Using a regular grid simplifies and speeds up the simulation of pseudo-GLM events significantly. Each regularly shaped pseudo-GLM event covers an area equal to the average size of the observed, irregularly shaped GLM events in the region of interest. Ana-

lyzing simulations built on this regular pseudo-GLM grid should lead to statistically similar results as for the irregular grid of the GLM observations.

Each NLDN flash is used to generate one pseudo-GLM flash. The target generator of the GEO lightning pseudo-observation generator simulates the targets based on the given NLDN flash characteristics. These pseudo-GLM targets provide the information to derive individual pseudo-GLM events. As the target generator may produce targets with values smaller than the observed (and physical) limits, the targets are adjusted to account for the known thresholds. For instance, negative flash extent and flash duration are set to zero, and there are at least 2 pseudo-GLM events per flash in accordance with NASA GLM data processing (Mach 2020). Pseudo-GLM flash NS and WE extents are calculated based on the simulated pseudo-GLM flash extent applying the same ratio as the NS and WE extents of the corresponding NLDN flash. If the NLDN flash contains a single pulse or stroke, the NS to WE ratio is set to one.

First, the location of pseudo-GLM events are generated. Using the simulated pseudo-GLM flash extent and its NS and WE components, a rectangular sub-domain on the pseudo-GLM pixel matrix is defined. The center of this sub-domain houses the NLDN flash position centroid and the corresponding pixel constitutes the first event of the pseudo-GLM flash. Any pixel within the sub-domain may also become a pseudo-GLM event of this pseudo-GLM flash. Three constraints have been designed to generate subsequent pseudo-GLM events: (i) each event of the flash has at least one adjacent or diagonal neighbor within one flash, thus, avoiding spatial gaps; (ii) pixels are primarily selected starting at the first event and propagating (meaning increasing distance to the first event) towards the sub-domain border to approximate the simulated flash extent; and (iii) additional pixels can be selected randomly within the rectangular area until the simulated event number is reached.

In consequence one single pixel of the sub-domain can contain more than one pseudo-GLM event.

Then, the pseudo-GLM events get time stamped. In the present study, the matching of GLM and NLDN flashes revealed that the median time offset between the mean time of a given NLDN flash and the mean time of the matched GLM flash was about 8 ms (not shown). The NLDN and GLM average flash duration were 0.24 s and 0.39 s, respectively. Hence, the mean time of matched NLDN and GLM flashes are relatively close while GLM flashes last on average longer than NLDN flashes. As a consequence, the mean time of the NLDN flash defines the mean time of the pseudo-GLM flash that is also the time stamp of the first pseudo-GLM event. Our generator is built to generate realistic FED fields. Only the spatial distribution of the events is needed to infer FED. Hence, the temporal occurrences of pseudo-events are uniformly and arbitrary distributed during the duration of one flash. Pseudo-event times are then rounded to the time frames of the mimicked GEO LLS, i.e., to 2-ms-frames for pseudo-GLM data. The only constraint is that any adjacent pixel occurs within 330 ms (i.e., the time criterion to separated flashes in NASA's GLM L2 algorithm). One 2-ms-frame contains often several pseudo-GLM events.

4. Results

Figure 3 shows the example of one simulated pseudo-GLM flash created with the final GEO lightning pseudo-observation generator based on a *linSVR* model, the corresponding GLM and NLDN observations, and the observed and simulated GLM flash characteristics. One can see the difference between the real GLM grid and the regular pseudo-GLM grid of the simulation (Figure 3(c)). The difference between observed and simulated flash extent is within the size of one GLM pixel for this example. The simulated flash duration exceeds

the observed flash duration significantly. There is also an overestimation of the number of GLM events by the generator.

Results are obtained from the 3-day test dataset. It contains 340,712 NLDN flashes that are used to simulate the same number of pseudo-GLM flashes. Statistics of the pseudo-GLM flashes are compared to the statistics of all 338,579 observed GLM flashes. First, the distributions of the simulated and observed GLM flash extent, flash duration, and event number per flash are compared. In a second time, the best target generators are used to simulate pseudo-GLM events and eventually compute the pseudo-FED product. The FED is analyzed statistically for both observed and simulated GLM data of the three test days. The minimum discrepancy between observation and simulation will indicate the most suitable target generator configuration for the final GEO lightning pseudo-observation generator.

a. Evaluating the target generators – Distributions of GLM flash extent, flash duration and event number per flash

In a general sense, a wide range of values is observed for all target distributions. The GLM flash duration ranged from 0.0 s to 16.4 s. Flash duration is defined as the time between the frames; therefore, a single frame features a flash duration of 0 s. Observed GLM flashes comprised between 2 and 1395 events. The test data features GLM flash extent between 0 km and 277 km. An extent of 0 km means that all events of the flash occurred at the same pixel. The target generators, that use ML models, should handle these ranges of values and predict targets statistics similar to the statistics of observed GLM flashes.

Tables 2, 3, and 4 summarize the findings, with statistics, the KS, and the CvM of the distributions of observed and simulated GLM flash duration, event number per flash, and flash extent on 07 April 2018, 26 May 2018, and 31 July 2018, respectively. The tables

contain distribution statistics for the respective target generator with smallest difference between observed and simulated characteristics for one target on that day. This analysis was also conducted for the full test data combining the 3 days and results are presented in Table 5. In the following, statistics of the simulated pseudo-GLM and the observed distributions are referred to as simulated statistics and observed statistics, respectively.

The majority of the target generator ML types (considering all tested models) feature mean values similar to the observed means for all three target characteristics. The simulated medians, however, exceed the observed medians in most cases, especially for the number of events per flash, suggesting a tendency to overestimate the target values (Tables 3 and 4). Results for the 07 April 2018 test case differ from that behavior (Table 2). That day saw exceptionally large flashes with high event numbers per flash that likely occurred within the MCS and the squall line. In consequence, the ML-based target generators underestimated the means of the observed flash characteristics for that test case, but the medians of simulated and observed targets are similar.

The previously described behavior is true for all ML types except for the *linSVR*. *linSVR* models filter the dataset in advance to build the prediction on the support vectors (Appendix B0.a.0)(vi)). That results (in this study) in lower differences between the simulated and observed median values as compared to using the other ML model types. On the other hand, the mean value of *linSVR*-based predictions is often smaller than the observed mean, especially for the event number per flash. To detail one example, the recommended *linSVR*-based generator (as explained later) underestimates median and mean flash extent by about 4.5 % and 11.7 %, respectively. The mean event number per flash is also underestimated by about 29.6 %, however, the median event number per flash is overestimated by 20 %. The *linSVR*-based generator creates, compared to the observations, not enough flashes with an

event number in the tails of the distribution, i.e., close to the observed minimum and maximum event numbers. Hence, it cannot mimic the full range of the observed event numbers per flash. This *linSVR*-based generator outperforms all other ML types with respect to the median considering the full test data.

Some general conclusions can be drawn regarding the generator performances for the observed range and variability of the target values. The target generator minimum often approximates or slightly exceeds the observed minimum, whereas the maximum is underestimated in most cases (see Tables 2 to 5). This particular behavior can even be seen for the best target generators because the number of small flashes with characteristics close to the minimum observed target values is relatively high. The rare, highest observed values, however, are often underrepresented in the statistical approach. It is further found that observed GLM flash statistics can vary for a given set of the six observed NLDN features. This is the case as our six NLDN features cannot completely explain the range of target values even if the statistics derived here are significant in terms of the large sample size. The large values of the RMSE per flash (Tables 2 to 5) result from the deterministic nature of the ML models in combination with this lack of information in the features, e.g., cloud properties. The RMSE values of the GLM flash extent are similar to the mean values, whereas they reach twice the mean for both GLM flash duration and event number per flash (Tables 2 to 5). Here, the optimization of our GEO lightning pseudo-observation generator for FED that depends mostly on the flash extent is evident. A relatively wide range of target values is in particular found for small NLDN flashes with NLDN pulse+stroke number, extent, and duration near the lower end of the distributions (not shown). Large (meaning long extent, long duration, and many strokes+pulses or events) NLDN flashes usually coincide with large GLM flashes. As the NLDN features are somewhat correlated to the GLM targets, the high

RMSE due to a small NLDN flash as input also leads to a high RMSE when predicting small GLM flashes.

KS and CvM assign a quantitative value to measure the distance between two samples. While KS is normalized (values of 0 to 1), the CvM value depends in general on the distance between simulated and observed CDFs and the sample size. As the sample size is kept constant for all generators, CvM in fact provides a common measure of the agreement between observed and simulated targets. Both KS and CvM feature lower values for the GLM flash duration than for both the GLM flash extent and the GLM event number per flash considering the full test dataset (Table 5). This result is in accordance with the strong correlation coefficients between observed GLM flash duration and NLDN features (not shown, see also Section 3.b.1). KS and CvM reach their highest values, i.e., when comparing the 3 target distributions, for the GLM event number per flash, for which the weakest correlations to features were observed.

Tables 2 to 5 also indicate that the choice of the target generator can be situational. The objective now is to find a configuration that approximates the observed GLM flashes for the entire test data and all three targets. Therefore, the differences between the simulated and observed statistics (i.e., mean, median, minimum, maximum, RMSE, KS, CvM) are calculated and normalized for each statistic. The normalization divides each absolute difference by the maximum absolute difference of all target generators for a given statistic. A value of 1 represents the worst target generator for the given statistic, while a value of 0 indicates no difference to the observation. In addition, and to summarize all the information, the normalized absolute differences and scores of the statistics are averaged in the so-called Normalized Difference Average (NDA). The perfect generator would yield an NDA of zero.

NDAs of the target generators can be directly compared in order to identify the highest performer. NDA is calculated per target and for all three targets overall.

Figure 4 shows the normalized differences and scores of different target generators for the GLM flash duration for the 3 test days combined. The GLM flash duration distribution is equally well simulated by a variety of ML-based target generators as the narrow spread of the medians (green line) indicates. In detail, a *linSVR*-based generator and a *MLP*-based generator can predict the mean well, a *MLP*-based generator and an *ETR*-based generator are best for the maximum, while a *linSVR*-based generator exhibits the lowest differences for the median as well as both KS and CvM scores. In total, a *linSVR*-based target generator best approximates the observed distribution of the GLM flash duration in this comparison with an NDA value of 0.30.

For the GLM event number per flash in Figure 5, *linSVR* and *BAGR KNN dist* models make the best target generators. The lowest NDA of 0.45 is obtained for several *linSVR* and *BAGR KNN dist*-based generators. For test day 07 April 2018 (a dominant mesoscale system with above-average mean and median GLM event numbers per flash), all generators underestimated the event number per flash. As generators using a *linSVR* usually predict lower values than the other generators, they underestimate the observed statistics even more on 07 April 2018. Nevertheless, for the full test data, there are *linSVR*-based generators that predict the mean event number equally well as the best target generator, i.e., *MLP*-based, as demonstrated by the lower whiskers in Figure 5. *LinSVR*-based generators are again most suitable to predict the event number distribution median.

One can analyze Figure 6 in a similar manner, this time for the GLM flash extent, and compare it to the tables. Tables 2 to 4 indicate that a *BAGR KNN dist*-based generator provides the best simulation for the 07 April 2018 test case, and that a *linSVR*-based

generator performed best for 26 May 2018 and 31 July 2018 test cases. Figure 6 supports these findings, as the boxplot minima for the *linSVR* type generators are the closest to zero for most statistics. *BAGR KNN dist*-based generators feature the second lowest values of KS and CvM. Some boxplots exhibit a wide range of outcomes. The range shows that all ML model types are sensitive to the configuration. The NDA of the best *linSVR* and of the best *BAGR KNN dist*-based generators equal 0.28 and 0.34, respectively.

Overall NDAs for all three targets range from 0.35 for the *linSVR num ext raw* generator to 0.87 for the *MLP num ext(a) raw* generator. The best (i.e., lowest NDA) 24 target generators all use a *linSVR*, and the performance of the best target generators varies only within the range of uncertainty given in Section 4.a.1. For example, the difference between the 1st and 10th ranked target generator is only 0.04 NDA. The NDA ranking of target generators reveals a clustering explained by the ML model type, with *linSVR*-based generators performing the best, followed by *BAGR KNN dist*-based, *ETR*-based, and polynomial regression-based generators. MLP- and HGBR-based generators exhibit the highest NDAs.

The NDA values are further evaluated for the target generator configuration and sets of features. The lowest 20 NDA values are obtained for 14 generators using the multi-step approach. Eleven of the best 20 target generators use all 6 NLDN features, and 7 use unscaled features. To obtain results independent of the ML model type, the best 10 *linSVR*-based generators are compared. Eight among them use the multi-step approach, 7 use all available features, and 6 use unscaled features.

In a brief conclusion, the ML model type has in fact the highest impact on the simulation of pseudo-GLM flashes and thus on the target generator. It is further found that most target generators performed best using the multi-step approach explained in Section 3.b.2. Generators using unscaled data performed surprisingly well in some cases. Nevertheless,

feature and target scaling generalize the used ML models with respect to feature and target ranges.

1) Confidence in the results

The confidence in the outcomes is evaluated for the two parts of the GEO lightning pseudo-observation generator. Doing so, the uncertainty of the outcomes is expressed as the range of outcomes given the same configuration. First, (only) three selected ML models with constant configuration are trained 10 times using the same full training dataset (for computational efficiency). Herewith, the training variability of the ML model is assessed. Figure 7 shows the distributions (boxplots) of targets for the full test data for the 3 models (x-axis) each trained trained 10 times for pseudo-GLM flash duration (a), pseudo-GLM number of event per flash (b), and pseudo-GLM flash extent (c), respectively. The predicted target range of the 10 trained models is smaller than the variability due to different ML model types and due to different configurations of one ML type. The 10 *BAGR KNN dist* based simulations feature a very narrow range of outcomes for all statistics. The 10 trainings of both a *linSVR* and an *MLP* yield a range of values of 0.2 to 0.4 normalized absolute difference for most statistics. The range of the minimum event number per flash (Figure 7b) and the minimum flash extent (Figure 7c) reaches about 0.5 and up to 0.7 for the *linSVR* and *MLP*-based generators, respectively. This range is in the order of magnitude as the variability enforced by using different ML model types. The range of normalized absolute difference for the maximum event number predicted based on 10 equally configured *linSVR* models is also about 0.6. In addition, the range of normalized absolute differences is always wider for the mean than for the median. Despite a relatively high uncertainty in some statistics, the overall trends as described in the previous section remain valid. Statistics sensitive to distribution

outliers, i.e., the mean and minimum, exhibit higher uncertainties due to the ML training than more robust statistics, i.e., the median, KS, and CvM. Some target generators, i.e., the *BAGR KNN dist*-based one, appear to provide very robust predictions. The uncertainty range is usually smaller than the overall range of values for each statistic.

The test of the variability in the results enforced by the second part of the GEO lightning pseudo-observation generator, i.e., generating pseudo-GLM events (not shown) is much smaller than for the ML part. Hence, the overall range of targets for a given generator configuration is similar to those shown in Figure 7.

b. Evaluating observed and simulated Flash extent density (FED)

Hourly FED maps are calculated for both GLM observed and simulated flashes. They will be referred to as observed and simulated FED, respectively, in the following. The evaluation includes the hourly FED summed-up over the domain (termed FED sum), the electrified areas defined as pixels with positive FED (i.e., greater than 0), and a visual inspection of convective cores. As the choice of the ML model type has the highest impact on the overall performance of the GEO lightning pseudo-observation generator, the results are mainly discussed with respect to the ML model types.

Figure 8 presents the observed FED (a) to the simulated FED of 3 selected generator configurations (b-d) for the example of 26 May 2018, 20:00 to 21:00 UTC. The three generator configurations capture the geographical distribution of the observed FED relatively well. One can identify the most active regions (highest FED values), that are situated at similar locations for the observed and simulated FEDs. The numbers in the top corners of the panels (a)-(d) indicate the number of lightning pixels with FED>0 on the left and the FED sum on the right. The product of the number of lightning pixels and the area per

pixel yields the electrified area. The *linSVR* (Figure 8b) uses GLM duration as additional feature when simulating GLM number, and then GLM duration and GLM event number to simulate GLM extent. This *linSVR*-based generator performs among the best for the simulation of GLM targets overall, and it appears to be among the best also for the FED sum. It underestimated the electrified areas in most cases (as in the example in Figure 8a,b). The *MLP*-based simulation of the FED of Figure 8c uses unscaled features and targets. GLM flash extent and flash duration relate only to 4 NLDN features (without mean amplitude and CG stroke ratio) and the GLM event number per flash uses the GLM flash duration and flash extent as additional features. This *MLP*-based generator performs among the best for the electrified area, but overestimates GLM flash extent, GLM event number per flash, and eventually the FED sum. Figure 8(d) maps the FED as simulated by a *BAGR KNN dist* based generator that predicts the pseudo-GLM flash extent and pseudo-GLM flash duration directly from the NLDN features (no additional pseudo-GLM feature). Both targets are then used to simulate the pseudo-GLM event number per flash. It is the best performing generator using the *BAGR KNN dist* type. Although the underlying *BAGR KNN dist* model overestimates the target medians and FED sum, the configuration is found among the best 50 target generators for both FED sum and and electrified area. It performed best for 07 April 2018 test case with the dominant squall line that produced most of the large-extent lightning flashes. In general, all 3 generators overestimate the 1-hour FED sum in Figure 8. The *linSVR*-based generator simulates an FED sum significantly closer to the observed FED sum than using both the *MLP* and the *BAGR KNN dist*. The *linSVR*, however, underestimates the number of lightning pixels, that is best simulated by the *MLP*-based generator.

The results are further investigated for the 3-day test period by comparing pixel-to-pixel simulated and observed hourly FED. Figure 9 shows the 2D-histograms, computed for the entire 3-day test dataset, for the same *linSVR* (a), *MLP* (b), and *BAGR KNN dist* (c) based generators as used in Figure 8. In general, the Pearson correlation coefficients R of 0.91 to 0.92 indicate well correlated distributions of observed and simulated FED. Figure 9 also shows the range of simulated FED is wider than the range of observed FED (grey box). The corresponding trend to overestimate the FED in the simulation is proofed by the regression lines (light green) that feature steeper slopes than the equal-value line (black). In particular, the *MLP* (Figure 9b) and the *BAGR KNN dist* model (Figure 9c) overestimate the FED usually stronger than the *linSVR* model (Figure 9a). Y-intercepts near 0 indicate good agreement for regions without lightning activity. These findings agree with the example in Figure 8.

To summarize and quantify the evaluation of both FED sum and electrified area, the metrics *normalized difference* D_{real} and *absolute normalized difference* D_{abs} are defined:

$$(a)\ D_{real} = \frac{\sum_{n=1}^{H} S_n - O_n}{\sum_{n=1}^{H} O_n} \quad \text{and} \quad (b)\ D_{abs} = \frac{\sum_{n=1}^{H} |S_n - O_n|}{\sum_{n=1}^{H} O_n} \tag{3}$$

where S_n is the simulated hourly FED sum or electrified area, O_n is the observed hourly FED sum or electrified area, and H is the total number of time steps (here 72 hours). D_{real} and D_{abs} can be used to compare the different GEO lightning pseudo-observation generators and identify the generator with the lowest difference to the observation.

As the ML part of the generator enforces significantly higher differences than the derivation of pseudo-GLM events (the second part), again results are mainly discussed regarding the different ML configurations. D_{real} and D_{abs} are calculated for the 3-day test period. For the FED sum, the 28 *linSVR*-based generators tested are ranked as first 28 configurations

in the comparison, i.e., lowest D_{abs}. The best GEO lightning pseudo-observation generators exhibit an D_{abs} of 22 % to 25 %, while D_{real} is close to zero (i.e., balance between situations with over- and underestimated FED sum). The worst ML model types (some of *MLP* and *ETR*) lead to almost twice as high FED sum as the observed values. Similar, positive values of FED sum D_{real} and D_{abs} for most generators mean that most generators overestimate the FED sum. This agrees well with Figure 9. The exception is found for the *linSVR* type generators that often underestimate the FED sum with D_{real} ranging from -22 % to 39 %. Figure 9a shows one example of a *linSVR* with positive D_{real}).

As mentioned, the best 28 generators for the FED sum are all of type *linSVR*. The best 10 *linSVR*-based generators use the multi-step approach. Three among those 10 generators use all available NLDN features, and five use unscaled features.

Results for the electrified area are in general closer to the observation than the FED sum. The generators with the lowest D_{abs}, *HGBR* type, differ absolutely by about 7.5 % from the observed electrified area. The vast majority of all tested target generators underestimate the electrified area (negative D_{real}). Multiple generators of various types feature D_{abs} of less than 10 %, e.g., using *HGBR*, *Poly*, *BAGR KNN dist*, *ETR*, or *MLP* models. The *linSVR*-based generators, that performed best for the FED sum, exhibit the highest differences to the observation here with D_{abs} from 15 % to 35 % (all with negative D_{real}). For example, the best performer for the FED sum is ranked as third worst for the electrified area with a high underestimation of the area.

The best 20 generators for the electrified area take advantage of the multi-step approach in 15 cases. Also 15 of those 20 ML-based generators use all NLDN features, and 9 do not apply feature and target scaling. Comparing only the *linSVR*-based generators, all 10

leading generators use 6 rather than only 4 NLDN features. This result strengthens the meaning of including all NLDN features and of the multi-step approach.

The recommended ML model for the GEO lightning pseudo-observation generator to balance the simulation of all pseudo-GLM target distributions, FED sum, and electrified area is the *linSVR num ext(a) plus* model. This configuration features an overall NDA of 0.39, and an D_{abs} to observed FED sum and electrified area of 24.9 % and 21.3 %, respectively. This generator used all available features and utilizes the multi-step approach. First, GLM flash duration is predicted from all six NLDN features, and then used as additional pseudo-feature to predict the event number per flash. Finally, the pseudo-GLM flash extent is simulated from NLDN features and the pseudo-features GLM flash duration and event number. Both features and targets are standardized.

Figure 10 presents hourly FED sum (a) and electrified area (b) with the overall value (1) and the difference to the observation (2) for 31 July 2018 test case. The observed FED and results for the 10 generators with lowest D_{abs} are plotted. The figure shows in addition the number of hourly simulated pseudo-GLM flashes (histogram). Similar figures for the other two test days are also evaluated but not shown here because identical conclusions are drawn. The absolute values (Figure 10 panels 1) show that the FED sum (a) reacts directly to the number of (simulated) flashes. The electrified area curves (b) appear to have a time offset relative to changes in the flash number, suggesting that within 1 h a lower number of relatively large flashes can electrify a similar area as a higher number of smaller flashes. An increasing (decreasing) flash rate during the development (decay) of convective storms does not automatically mean a larger (smaller) electrified area, since even less flashes can still illuminate a large portion of the cloud via scattering. The simulated FED adapts this

behavior very well. In particular, the simulated FED features similar hours with highest FED and electrified area as the observed FED.

It is observed that the simulated FED sum usually exceeds the corresponding observation during the phases of highest flash amounts within the region (Figure 10a). This could mean that NLDN detects significantly more flashes than GLM during these times, and thus the number of simulated flashes is significantly higher than the number of observed GLM flashes. These findings agree with Zhang and Cummins (2020), who found that the GLM DE decreases for high flash rates and with shorter extent and duration flashes, which are observed during the mature phase of a thunderstorm.

It should be noted that the absolute values (Figure 10a1 and b1) and difference to the observation (Figure 10a2 and b2) for the FED sum (Figure 10a) have the same order of magnitude for the FED sum. In contrast, the difference (10b2) is one order of magnitude smaller than the absolute values (Figure 10b1) for the electrified area. Hence, the difference to observed FED and also the spread between generators with different configurations are much greater for the FED sum than for the electrified area. Therefore, it is decided to put more weights on the ranking of the FED sum than on the ranking of generators by electrified area when choosing the recommended generator. Eventually, the *linSVR*-based generator returns as the recommendation in an overall evaluation context. If, however, for a certain objective the electrified area is most important, several *HGBR*, *MLP* or even *ETR* based generators perform better than the recommended *linSVR*-based generator.

In a Monte Carlo approach, FEDs for 10 of in total 100 realizations of the recommended *linSVR* model are calculated for the three test days. Figure 11 illustrate the median (line) and range (shaded) of FED sum and electrified area as an example. The variability of both the FED sum and the electrified area has the same order of magnitude as the difference

between the leading generators (Figure 10). Figure 11 also confirms that the *linSVR*-based generator tends to underestimate the electrified area. The vast majority of the time, all 10 realization simulate lower electrified areas than the GLM observations indicate. On the bright side, all 10 realizations remain relatively close to the observed FED sum at most times (except for the cases with intense convection, as discussed earlier). It should be noted that this *linSVR*-based generator does not appear among the best 10 generators for the electrified area (Figure 10b).

5. Summary

This study analyzed in detail the simulation of GEO lightning pseudo-observations in two parts: First pseudo-GLM flash characteristics are simulated and then pseudo-GLM events are derived. The data generator uses only LF ground-based data. There is no additional cloud information used in the generator. The entire process is non-trivial because relations (correlations) between characteristics of coincident LF ground-based and optical satellite lightning observations are often weak at the flash scale.

A multivariate analysis using several features and targets is conducted to achieve more robust flash characteristics. Simulated GEO flash characteristics (targets) are obtained via machine learning (ML) models. Targets include GLM flash extent, GLM flash duration, GLM event number per flash. An independent test data set is then introduced to compare the statistics of simulated pseudo-GEO flashes to the observed GEO, i.e., GLM, flash characteristics. In a second part, the simulated targets are used to mimic individual GEO events on a regular latitude longitude grid.

After testing different ML models used in the first part of our generator, a linear SVR (*linSVR*) based GEO lightning pseudo-observation generator is recommended. The results of

multiple *linSVR* configurations turned out to be similar. In more detail, our recommendation is to use a *linSVR* with feature and target scaling, that uses all available features (including NLDN and pseudo-GLM characteristics) and pseudo-features in a multi-step approach.

The type of the ML model chosen in the first part of our GEO lightning pseudo-observation generated has the major impact on the simulated flashes. In fact, the performance ranking of tested target generators reveals clusters per ML model type. Whereas the vast majority of generators produces pseudo-GLM flashes with flash characteristic means close to the observed ones, they simultaneously overestimate the medians of flash characteristics. Therefore, they produce not enough small flashes as compared to the GLM observations. Only *linSVR*-based generators were able to simulate pseudo-GLM flash characteristics with distribution medians close to the observation for the 3-day test dataset. This gain is achieved at the expense of slightly underestimating the target means. It is then found that FED sums from *linSVR*-based generators are closer to the observed FED sum than for all other generators, however, the electrified area is at least 10 % smaller than the observed electrified area.

Besides the type of the ML model, the set of features and the feature scaling impact the results. In particular, including (pseudo) GLM flash characteristics in the set of features improved the predictions of most ML models as target generators and thus the overall performance of the GEO lightning pseudo-observation generator.

In general, generators that perform well for the FED sum exhibit high D_{abs} for the electrified area and vice versa. For example, the best generator for the electrified area with D_{abs} (D_{real}) of 7 % (−2 %) highly overestimates the FED in most cases with D_{abs} and D_{real} of 75 % and 72 %, respectively. On the other hand, the best generator for the FED sum with

D_{abs} (D_{real}) of 22 % (2 %) always underestimates the electrified area with D_{abs} (D_{real}) of 27 % (−27 %).

The developed GEO lightning pseudo-observation generator provides exactly one pseudo-GEO flash for each LF flash. It does not distinguish whether an LF flash, i.e., an NLDN flash, is detected by the GEO LLS, i.e., GLM. During the application of the generator, there is no information whether a given NLDN flash could be detected by the GEO LLS. Additional assumptions, e.g., using flash characteristics, would then be needed to distinguish the LF flashes with and without GEO match. In addition, our GEO lightning data generator does not include a specific part to simulate GEO flashes that are not directly coincident to any LF flash. Here, the pragmatic approach of using all LF flashes as input is justified with similar DE of the LF (i.e., NLDN) and the GEO (i.e., GLM) LLS thus giving overall similar amounts of GLM and NLDN flashes. Then, NLDN and GLM flashes without any coincident observation are analyzed. They are referred to as NLDN-only and GLM-only flashes, respectively. It was observed that both the NLDN-only and GLM-only flashes occurred mostly in proximity to the convective cores and regions of overall high flash rates. The number of observed GLM-only and NLDN-only flashes was in general of the same order of magnitude. It is assumed that pseudo-GLM flashes simulated from the NLDN-only flashes substitute the observed GLM-only flashes. It should be mentioned that some simulated pseudo-GLM flashes might overlap as the pseudo-GLM flash extent is usually greater than the NLDN flash extent. Overlapping pseudo-GLM should actually be merged, however, this is not further studied here. As one possible consequence, the simulated pseudo-GLM FED can be somewhat higher than the observed GLM FED (as seen for most configurations of generators). In particular, the simulated hourly FED values are often higher than observed in situations when many NLDN flashes were observed. On the other hand, lower simulated

than observed FED at the rim of cells indicate that NLDN flashes cannot represent the scattering of light as seen by GLM. Peterson et al. (2020) showed that optically detected flashes can appear large near storm edges due to light reflected off nearby clouds. Simulated FED (based on NLDN observations) could then be closer to the actual flash channel extent as derived from LMA-type observations than the observed FED, especially at the rim of cells. Nevertheless, the simulation might differ from what the satellite sensor sees.

Our methodology is configured and refined for NLDN Vaisala sensors. NLDN flash statistics were compared to coincident GLM flashes and their extent, duration, and event number. For an application in other regions than the US and/or with different LF networks, NLDN operational specification and observations might be compared with the ones of the other LF network in order to identify the necessity for adapting the input data. This comparison can be of direct (e.g., NLDN and GLD360) or indirect (e.g., NLDN and Meteorage compared to ISS-LIS as common reference) nature.

The studied dataset is limited to a region in the SE USA and for the months of March to September. GLM features high flash detection efficiency (e.g., Marchand et al. 2019; Murphy and Said 2020) in this region satisfying our objective to build a high-fidelity generator to simulate GEO lightning data. However, the limited dataset lacks winter storms that may have different characteristics. For the application of our generator in Europe, this should be a minor limitation as winter storms rarely occur here. The SE US region features mostly normal polarity storms while storms with different charge structure occur more often in other parts of the US. For example, Rutledge et al. (2020) show that flash characteristics and GLM flash DE are altered for storms with anomalous charge structure. In addition, the data used to train our GEO lightning data generator were recorded in this region well covered and far from the edges of the GLM's (on GOES-16) field of view. Simulating data

of a GEO LLS near the edges of the field of view needs caution regarding parallax effects and an increase in the area one event covers.

The GLM data includes a parallax correction. Our GEO lightning pseudo-observation generator assumes that GLM observations are correctly located. The simulated flashes are placed according to the LF lightning data. If the GEO LLS that should be mimicked uses a different parallax correction than GLM, an adaption may become necessary to obtain realistic data of this LLS.

A comparison of GLM and NLDN during day and night, and for intra-cloud (IC) and cloud-to-ground (CG) flashes revealed similar relationships between NLDN and GLM flash characteristics. The dataset for the ML includes all observed flashes, without a separation of these flash types. In addition, all applied ML models aim to optimize average characteristics. This study uses deterministic approaches without a definition of a confidence interval of the outcomes. As one result, the tails of the characteristics' distributions, e.g., exceptionally small flashes, are underrepresented in the simulation compared to the observation.

Supplementary data might improve the present GEO lightning data generator. Cloud information and brightness temperature data could provide additional features for the ML, e.g., cloud top height, and also information about more likely scattering directions, e.g, in anvils of convective clouds or in stratiform cloud lightning. (Doppler-)Radar data would provide even more versatile possibilities to include cloud structures, dynamics and microphysics.

Data availability statement. NLDN data are available from Vaisala, and data as presented in this paper were provided by Ronald L. Holle.

GLM data are in general available from NOAA: GOES-R Algorithm Working Group and GOES-R Series Program, (2018): NOAA GOES-R Series Geostationary Lightning Mapper (GLM) Level 2 Lightning Detection: Events, Groups, and Flashes. Flash and Event subsets used. NOAA National Centers for Environmental Information. doi:10.7289/V5KH0KK6. Access date: 01 November 2018. The GLM data as presented in this paper were downloaded from the French AERIS/ICARE Data and Services Center of the Université Lille where the files were stored 19 April 2018.

Acknowledgments. This work is part of the Ph.D. thesis of Felix Erdmann funded by the Centre National de l'Etudes Spatiales (CNES) and Météo-France. This article is funded by Météo-France, the SOLID project (Funding ID: n/a), and the EXAEDRE project (Funding ID: ANR-16-CE04-0005). This work was supported by the French National program LEFE (Les Enveloppes Fluides et l'Environnement), project ASMA. We thank the AERIS/ICARE Data and Services Center for providing access to GLM data.

The authors thank Ronald L. Holle (Vaisala) for providing the NLDN data and reviewing the NLDN specific information. We would like to acknowledge the guidance of Chien Wang (MOPGA Recipient Scientist at Laboratoire d'Aérologie, Toulouse, France) regarding machine learning applications.

The authors would like to thank the 3 anonymous reviewers and Eric Bruning for their detailed and constructive critics.

The authors declare that there is no conflict of interests.

APPENDIX A

The multi-target multi-step approach

This section describes a multi-target regression that simplifies the idea of the stacked single target (SST) approach (Spyromitros-Xioufis et al. 2016). In this study, there are six NLDN features (as physical input) and three GLM targets (as physical simulated variables) per sample, i.e., per flash. The three GLM targets are denoted T_i, T_j, and T_k. Targets that are used like features are referred to as pseudo-features, i.e., T_j and T_k in Figure 2a. With this dataset, there are 4 different ways to simulate the target T_i. The 4 ST models are shown as the training part in Figure 2a. There are 3 classes of models: Yellow is the model class (1) without pseudo-features, gray indicates model class (2) using 1 pseudo-feature, and the red for model class (3) uses 2 pseudo-features. The model $M_{\rightarrow i}$ constitutes the common ML model, i.e., class (1), with only the NLDN features as input. One (i.e., T_j or T_k) or two (i.e., T_j and T_k) of the three targets can be added to the input as pseudo-features in order to simulate the target T_i. The resulting models $M_{j\rightarrow i}$ (using T_j as pseudo-feature), $M_{k\rightarrow i}$ (using T_k with the features), and $M_{j,k\rightarrow i}$ (using T_j and T_k with the features) may indeed take advantage of correlations between the predicted target and the targets that are used as pseudo-features.

The application case only uses the NLDN features as first input. Therefore, a multi-step approach is required. Figure 2b presents the example application for a 3-step approach that first predicts the pseudo-GLM flash duration, then the pseudo-GLM event number per flash, and finally the pseudo-GLM flash extent. This configuration is denoted *num ext(a)* (see Appendix 5 for details on the configuration naming). The first step, $M_{\rightarrow i}$, uses the NLDN features and predicts the first pseudo-GLM characteristic $M_{\rightarrow i}\,(NLDN)$, i.e., pseudo-GLM flash duration. The second step, $M_{i\rightarrow j}$, uses the NLDN features and the result of the first step, $M_{\rightarrow i}\,(NLDN)$, i.e., the pseudo-feature GLM flash duration. This model of class (2) predicts the second pseudo-GLM characteristic $M_{i\rightarrow j}\,[NLDN, M_{\rightarrow i}\,(NLDN)]$, i.e.,

the pseudo-GLM event number per flash. Both predicted pseudo-GLM characteristics (i.e., GLM flash duration and event number per flash) can then be used as pseudo-features to predict the third target with the class (3) model $M_{i,j\rightarrow k}$. Hence, the final target prediction $M_{i,j\rightarrow k}\left\langle NLDN, M_{i\rightarrow j}\left[NLDN, M_{\rightarrow i}\left(NLDN\right)\right]\right\rangle$ depends on the NLDN features and both previous predictions for this configuration. In general, a model of class (3) can also use two pseudo-features produced by two models of class (1). Also, 2 models of class (2) could be used to simulate the remaining 2 targets after the first step. Utilizing 3 times a model of class (1) is equal to the common ML ST approach. Hence, several combinations of models of different classes are possible and have been investigated here.

The ML training for the multi-step approach can be performed in parallel for the models $M_{\rightarrow i}$, $M_{j\rightarrow i}$, and $M_{j,k\rightarrow i}$. The approach can use all ML model types as the training creates independent learners. Our multi-step approach adapts the idea of the SST, but uses observations instead of simulated targets during the ML training. It has the advantage of faster training for computation efficiency.

Although the correlations between the NLDN features and both GLM flash extent and event number per flash are relatively weak, the NLDN features improve the prediction during each step as seen through a feature drop tests (not shown). Indeed, all features have a positive effect on the model score. Due to strong correlations between GLM flash duration and NLDN features flash duration and stroke+pulse number, and to reduce the number of ML-based target generators, only the multi-step approaches which predict the GLM flash duration in the first step ($M_{\rightarrow}i$) are considered. The GLM flash duration is also weakly correlated to both GLM flash extent and event number (R of about 0.10 and 0.17), and GLM flash extent and event number per flash are well correlated (R of about 0.74). Our multi-step approach aims at producing more realistic pseudo-GLM flash extent and event

number per flash than using the NLDN features alone. The NLDN features also remain important as the correlations between some targets are weak.

Up to two (depending on the configuration) additional steps follow the first step to simulate the remaining GLM targets. The first step always provides the pseudo-GLM flash duration from the NLDN flash characteristics as features (as explained). The second step uses the simulated flash duration in addition to the NLDN features to simulate one or both of pseudo-GLM flash extent and event number per flash. A potential third step may simulate the last GLM target based on NLDN features and the two remaining simulated pseudo-GLM characteristics as additional pseudo-features. The impact of both using only the GLM duration (strongest correlations) or using GLM duration and a second target as additional pseudo-features to simulate the remaining target (GLM flash extent or event number per flash) is investigated. These multi-step strategies are considered as new configurations of each ML model to simulate all three GLM targets (besides using the NLDN features only).

APPENDIX B

Definitions of the machine learning (ML) algorithms

a. ML model types

This section defines the seven ML model types that are trained in the study. The basic idea of each ML model type is introduced, and specifications and important parameters for their tuning are briefly described. As mentioned in Section 3.b.3, Python's sklearn-package is used. Model names are given as they appear in the sklearn library and documentation (`https://scikit-learn.org/stable/`) that provides further details.

(i) Multivariate Linear Regression. The first approach is the most commonly used linear regression *sklearn.linear_model.LinearRegression*. It is applied simultaneously to all features and targets and is, thus, a multivariate linear regression (*LinReg*). The algorithm seeks for the minimum sum of squared errors between the features and the targets by using linear functions. It is an ordinary least square fit in a space with dimensions equal to the number of features times the number of targets.

(ii) Multivariate Polynomial Regression. The Polynomial Regression (*Poly*) is an adjustment to the multivariate linear regression. It fits a polynomial of degree 3.0 (rather than a linear function) to minimize the sum of squares between predicted targets and the corresponding observations in the validation dataset. The cubic polynomial model is chosen based on the initial correlation analysis with relations between any one feature and one target. The low polynomial degree allows fast computation.

(iii) Random Forest Regressor. A Random Forest (*RF*) is a ML algorithm using bootstrapping and applying single decision trees to each bootstrap sample. The overall result is the average of the outcomes of all the decision trees. The minimum leaf size defines the minimum size at the end of the decision tree. A specific form of the *RF* is called Extra Trees *sklearn.ensemble.ExtraTreesRegressor* (*ETR*, Geurts et al. 2006). *ETR* enforces randomness by not only selected random features in each subset but also splits depending on the best randomly produced thresholds instead of looking for the most distinctive threshold (as in *RF*). *ETR* usually reduces the variance and increases the bias of the model compared to *RF*. In general, a higher number of trees improves the performance but also the computation time. Our *RF* implementation uses a *ETR* model with 50 decision trees. The number of decision trees results from a sensitivity test (*ETRs* with 5, 10, 20, 50, 100, and 500 trees

were tested, not shown) between performance as R^2-score (see skilearn documentation) and computational effort. Here, the full ML dataset with independent training and validation (i.e., calculating the R^2-scores) data (see Section 2.c) is used. The minimum leaf size is set to two, i.e., a remaining sample of two data points defines the end of the branch. Single point leaf size would increase the variability of the trees and would lead to a higher likelihood of overfitting.

(iv) Bagging Regressor with k-Nearest Neighbor Regressor. Bootstrap Aggregation, short Bagging (Breiman 1996), uses subsamples drawn by bootstrapping from the entire dataset. This step is similar to the *RF* regressor. The algorithm used to treat the subsamples can, however, be chosen (not always a decision tree). This paper applies the bagging regressor *sklearn.ensemble.BaggingRegressor* combined with a K-Nearest Neighbor (*KNN*) regressor (e.g., Altman 1992) *sklearn.neighbors.KNeighborsRegressor* on each of 50 subsamples. The number of neighbors to use by default is set to the 5 closest points and distance weighting is applied for Euclidean distances. The *KNN* finds closest neighbors with a K-dimensional tree (KD tree) method (Bentley 1975). It reduces the number of distance calculations compared to a brute-force approach calculating distances between all data points. The *KNN* regressor in combination with distance weighting should represent the actual range of the subsample training data better than a decision tree (as used in *RF* and *ETR*). The expense might be an increase in overfitting of the data.

(v) Multilayer Perceptron Neural Network. MultiLayer Perceptrons (*MLPs*) are a form of Neural Networks in supervised ML (Glorot and Bengio 2010). They consist of different layers of neurons, where the input layer neurons represent the features and the output layer neurons represent the simulated targets. An adjustable number of hidden layers can connect the

input and output layers. Each neuron initially transforms the values from the previous layer in a weighted linear summation. Then, a (non-)linear activation function is used. Parameters of our *MLP* model *sklearn.neural_network.MLPRegressor* were determined after testing different configurations to balance computation time and accuracy. It uses one hidden layer with 50 neurons. The activation function is the rectified linear unit function. Additionally, an early stopping criterion is applied if there is no improvement over 20 consecutive iterations. The early stopping requires splitting the training dataset randomly, whereby 10 % are used to verify the improvement of the model and 90 % remain as actual training dataset. The tolerance for the stopping criteria is reduced from default 10^{-4} to 10^{-8} to allow a higher number of iterations. Furthermore, the default Adam solver (Kingma and Ba 2014) and a constant learning rate are used, along with adjusted parameters beta1 (0.7), beta2 (0.9), and epsilon (10^{-10}) for the decay rates and the numerical stability in the Adam solver.

(vi) Support Vector Regressor. The Support Vector Regressor (*SVR*) is based on Support Vector Machine (*SVM*) algorithms. A set of hyper-planes is constructed. Therefore, a defined kernel function is applied to achieve a separation of data clusters (by the hyper-planes) for the regression. The kernel function can be a linear or non-linear function (i.e., polynomial or Radial Basis Function). Linear *SVR* (*linSVR*) is faster and uses less memory than *SVR* with non-linear kernel-functions. Non-linear *SVR* provides usually better separation of different clusters in the data and thus a higher score than linear SVR. The distances of the nearest data points to the hyper-planes (so-called functional margins) are maximized. Points with a larger functional margin lead to less uncertainty for the prediction than data close to the hyper-planes. *SVM* in general analyzes all data while the cost function (L1 loss) depends on a subset of the training data, referred to as support vectors. Support

vectors are a set of data points with some distance from the target values that still allow the correct prediction. The systematic reduction of the training data makes this model type fundamentally different from the remaining model types of this study. Further information is also provided by Smola and Schölkopf (2004).

Due to our large sample size (672,794 flashes), only the *linSVR sklearn.svm.LinearSVR* is used in this study in its default configuration. As for the *MLP*, an early stopping criterion is used for a lack of improvement between consecutive iterations.

(vii) Histogram-based Gradient Boosting Regression Tree. Boosting is, besides bagging, another approach to reduce overfitting of ML models. It combines an ensemble of weak learners to one strong learner. The Histogram Gradient Boosting Regression *sklearn.ensemble.HistGradientBoostingRegressor* (*HGBR*) is much faster than regular Gradient Boosting Regressors. Data is first binned into 256 integer-valued bins. The algorithm can then leverage histograms instead of relying on sorted continuous values when building the decision trees. The number of splitting points is reduced and the algorithm becomes time efficient, inspired by LightGBM (Ke et al. 2017). The first step of the *HGBR* averages the target values and calculates residuals (average difference of observation to prediction) with a least-squares loss function. Based on these residuals, a small decision tree is built, along with a learning rate. The learning rate limits the influence of a single small decision tree in the final ensemble to avoid overfitting. Then, new predictions are computed using the averages and the decision tree for residuals. Based on new predictions, new residuals are calculated and a new decision tree is created. The final model combines several of these decision trees to pull the target averages towards the observations. The used maximum

number of iteration is 500 and the early stopping criteria kicks in after 50 iterations without significant improvement of the loss value.

b. Naming convention for the GEO lightning pseudo-observation generator configurations

This section defines the meaning of names given to different configurations of a target generator. The names and abbreviations of the ML model types can be found in Table A1. The given ML model types are used in the first part of the GEO lightning pseudo-observation generator referred to as target generator. Table A2 summarizes the feature usage that is available for each ML model type available for the target generator. The names are used in Tables 2 to 5 and in the Figure 4 to 6. The feature usage called *NLDN* is the default configuration as described. Generations with extension of only *default* or only *plus* are single-step approaches, i.e., using 3 times the model of class (1) in Figure 2. Multi-step simulations are performed in the order of the table columns from left to right. The GLM flash duration is always obtained from the first step here, followed by the the number of events per flash, and the GLM flash extent during the last step. The extension *num* indicates one additional step only for the pseudo-GLM event number per flash using the pseudo-GLM flash duration as pseudo-feature. GEO lightning pseudo-observation generator configurations with extension *num ext* and *num ext(a)* have two additional steps using different pseudo-features as shown in Table A2. The *num ext(a2)* generators use only the GLM flash duration as pseudo-feature, thus two models of class (2) as of Figure 2. Various combinations of the given name extensions are possible, e.g., an unscaled model with NLDN mean LF amplitude and CG stroke ratio as additional features that uses the GLM flash duration as pseudo-feature for the event number per flash gets the extension *num(a) raw plus*.

References

Aguiar, G. J., E. J. Santana, S. M. Mastelini, R. G. Mantovani, and S. B. Jr, 2019: Towards meta-learning for multi-target regression problems. 1907.11277.

Allen, B. J., E. R. Mansell, D. C. Dowell, and W. Deierling, 2016: Assimilation of pseudo-GLM data using the ensemble Kalman filter. *Monthly Weather Review*, **144**, 3465–3486, doi:10.1175/MWR-D-16-0117.1.

Altman, N. S., 1992: An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, **46 (3)**, 175–185, doi:10.1080/00031305.1992.10475879, URL https://amstat.tandfonline.com/doi/abs/10.1080/00031305.1992.10475879, https://amstat.tandfonline.com/doi/pdf/10.1080/00031305.1992.10475879.

Anderson, T. W., 1962: On the distribution of the two-sample cramér-von mises criterion. *The Annals of Mathematical Statistics*, **33 (3)**, 1148–1159, URL http://www.jstor.org/stable/2237885.

Bateman, M., 2013: A high-fidelity proxy dataset for the geostationary lightning mapper (GLM). AMS Sixth Conference on the Meteorological Application of Lightning Data, 725, austin, TX, USA.

Bateman, M., and D. Mach, 2020: Preliminary detection efficiency and false alarm rate assessment of the Geostationary Lightning Mapper on the GOES-16 satellite. *Journal of Applied Remote Sensing*, **14 (3)**, 1 – 10, doi:10.1117/1.JRS.14.032406, URL https://doi.org/10.1117/1.JRS.14.032406.

Bateman, M., D. Mach, and M. Stock, ????: Further investigation into detection efficiency & false alarm rate for the geostationary lightning mappers aboard goes-

16 and goes-17. *Earth and Space Science*, **n/a (n/a)**, e2020EA001 237, doi:https://doi.org/10.1029/2020EA001237, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020EA001237, e2020EA001237 2020EA001237, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020EA001237.

Bentley, J. L., 1975: Bagging predictors. *Multidimensional Binary Search Trees Used for Associative Searching*, **18 (9)**, 509–517, doi:10.1.1.160.335.

Betz, H. D., K. Schmidt, P. Laroche, P. Blanchet, W. P. Oettinger, E. Defer, Z. Dziewit, and J. Konarski, 2009: LINET—an international lightning detection network in europe. *Atmospheric Research*, **91 (2)**, 564 – 573, doi:https://doi.org/10.1016/j.atmosres.2008.06.012, URL http://www.sciencedirect.com/science/article/pii/S0169809508002305, 13th International Conference on Atmospheric Electricity.

Biron, D., L. D. Leonibus, P. Laquale, D. Labate, F. Zauli, and D. Melfi, 2008: Simulation of Meteosat Third Generation-Lightning Imager through tropical rainfall measuring mission: Lightning Imaging Sensor data. *Remote Sensing System Engineering*, P. E. Ardanuy, and J. J. Puschell, Eds., SPIE, International Society for Optics and Photonics, Vol. 7087, 77 – 88, doi:10.1117/12.794764.

Blakeslee, R., and W. Koshak, 2016: Lis on iss: Expanded global coverage and enhanced applications. *The Earth Observer*, **28**, 4–14.

Blakeslee, R. J., and Coauthors, 2020: Three years of the lightning imaging sensor onboard the international space station: Expanded global coverage and enhanced applications. *Journal of Geophysical Research: Atmospheres*, **125 (16)**, e2020JD032 918, doi:https://doi.org/10.1029/2020JD032918, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/

10.1029/2020JD032918, e2020JD032918 2020JD032918, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020JD032918.

Borchani, H., G. Varando, C. Bielza, and P. Larrañaga, 2015: A survey on multi-output regression. *WIREs Data Mining and Knowledge Discovery*, **5 (5)**, 216–233, doi:10.1002/widm.1157, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1157, https://onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1157.

Breiman, L., 1996: Bagging predictors. *Machine Learning*, **24(2)**, 123–140, doi:10.1023/A:1018054314350.

Brooks, I., C. Saunders, R. Mitzeva, and S. Peck, 1997: The effect on thunderstorm charging of the rate of rime accretion by graupel. *Atmospheric Research*, **43 (3)**, 277 – 295, doi:https://doi.org/10.1016/S0169-8095(96)00043-9, URL http://www.sciencedirect.com/science/article/pii/S0169809596000439.

Bruning, E. C., and Coauthors, 2019: Meteorological imagery for the geostationary lightning mapper. *Journal of Geophysical Research: Atmospheres*, **124 (24)**, 14 285–14 309, doi:10.1029/2019JD030874, URL https://agupubs-onlinelibrary-wiley-com-s.docadis.ups-tlse.fr/doi/abs/10.1029/2019JD030874, 10.1029/2019JD030874, https://agupubs-onlinelibrary-wiley-com-s.docadis.ups-tlse.fr/doi/pdf/10.1029/2019JD030874.

Cecil, D. J., S. J. Goodman, D. J. Boccippio, E. J. Zipser, and S. W. Nesbitt, 2005: Three years of trmm precipitation features. part i: Radar, radiometric, and lightning characteristics. *Monthly Weather Review*, **133**, 543–566, doi:10.1175/MWR-2876.1.

Christian, H. J., and Coauthors, 1999: The lightning imaging sensor. 11th International Conference on Atmospheric Electricity, NASA Conf. Publ. NASA/CP-1999-209261a, 746–

749.

Coquillat, S., and Coauthors, 2019: SAETTA: high-resolution 3-d mapping of the total lightning activity in the mediterranean basin over corsica, with a focus on a mesoscale convective system event. *Atmospheric Measurement Techniques*, **12 (11)**, 5765–5790, doi: 10.5194/amt-12-5765-2019, URL https://amt.copernicus.org/articles/12/5765/2019/.

Cummins, K. L., and M. J. Murphy, 2009: An overview of lightning locating systems: History, techniques, and uses, with an in-depth look at the u.s. nldn. *IEEE Trans. Electromag. Compat.*, **51(3)**, 499–518, doi:10.1109/TEMC.2009.2023450.

Deierling, W., and W. A. Petersen, 2008: Total lightning activity as an indicator of updraft characteristics. *Journal of Geophysical Research*, **113**, D16 210, doi:10.1029/2007JD009598.

Dobber, M., and J. Grandell, 2014: Meteosat Third Generation (MTG) Lightning Imager (LI) instrument performance and calibration from user perspective. InProceedings of the 23rd Conference onCharacterization and Radiometric Calibration for Remote Sensing (CALCON), 11-14 August 2014,Utah State University, Logan, Utah, USA. 13 pages.

Emersic, C., and C. Saunders, 2020: The influence of supersaturation at low rime accretion rates on thunderstorm electrification from field-independent graupel-ice crystal collisions. *Atmospheric Research*, **242**, 104 962, doi:https://doi.org/10.1016/j.atmosres.2020.104962, URL http://www.sciencedirect.com/science/article/pii/S0169809519312803.

Erdmann, F., E. Defer, O. Caumont, R. J. Blakeslee, S. Pédeboy, and S. Coquillat, 2020: Concurrent satellite and ground-based lightning observations from the optical lightning imaging sensor (ISS-LIS), the low-frequency network meteorage and the SAETTA

lightning mapping array (LMA) in the northwestern Mediterranean region. *Atmospheric Measurement Techniques*, **13 (2)**, 853–875, doi:10.5194/amt-13-853-2020, URL https://www.atmos-meas-tech.net/13/853/2020/.

Fierro, A. O., Y. Wang, J. Gao, and E. R. Mansell, 2019: 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, doi:10.1175/MWR-D-18-0421.1.

Geurts, P., D. Ernst, and L. Wehenkel, 2006: Extremely randomized trees. *Machine Learning*, **63**, 3–42, doi:10.1007/s10994-006-6226-1.

Glorot, X., and Y. Bengio, 2010: Understanding the difficulty of training deep feedforward neural networks. Vol. 9 of JMLR: W& CP 9, 13th International Conference on Artificial Intelligence and Statistics, Sardinia, Ital, 249–256.

Goodman, S., D. Mach, W. Koshak, and R. Blakeslee, 2012: GLM lightning cluster-filter algorithm. Algorithm theoretical basis document, NOAA NESDIS Center for Satellite Application and Research.

Goodman, S. J., and Coauthors, 2013: The GOES-R Geostationary Lightning Mapper (GLM). *Atmospheric Research*, **125-126**, 34 – 49, doi:https://doi.org/10.1016/j.atmosres.2013.01.006, URL http://www.sciencedirect.com/science/article/pii/S0169809513000434.

Höller, H., and H.-D. Betz, 2010: Study on inter-comparison of lis and ground-based lightning location system observations. Report ITT No. 09/996, EUMETSAT, Deutsches Zentrum für Luft- und Raumfahrt.

Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, 2017: Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., Curran Associates, Inc., 3146–3154, URL http://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.pdf.

Kingma, D. P., and J. Ba, 2014: Adam: A method for stochastic optimization. URL https://arxiv.org/abs/1412.6980, 1412.6980.

Kolmasova, I., T. Marshall, S. Bandara, S. Karunarathne, M. Stolzenburg, N. Karunarathne, and R. Siedlecki, 2019: Initial breakdown pulses accompanied by vhf pulses during negative cloud-to-ground lightning flashes. *Geophysical Research Letters*, **46**, 5592–5600, doi:10.1029/2019GL082488.

Koshak, W., D. Mach, M. Bateman, P. Armstrong, and K. Virts, 2010: GOES-16 GLM level 2 data full validation data quality - product performance guide for data users. Guide, NASA Marshall Space Flight Center.

Koshak, W. J., and R. J. Solakiewicz, 01 Jan. 2015: A method for retrieving the ground flash fraction and flash type from satellite lightning mapper observations. *Journal of Atmospheric and Oceanic Technology*, **32 (1)**, 79 – 96, doi:10.1175/JTECH-D-14-00085.1, URL https://journals.ametsoc.org/view/journals/atot/32/1/jtech-d-14-00085_1.xml.

Luque, M. Y., F. Nollas, R. G. Pereyra, R. E. Bürgesser, and E. E. Ávila, 2020: Charge separation in collisions between ice crystals and a spherical simulated graupel of centimeter size. *Journal of Geophysical Research: Atmospheres*, **125 (3)**, e2019JD030 941, doi:10.1029/2019JD030941, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/

2019JD030941, e2019JD030941 2019JD030941, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019JD030941.

Lyu, F., S. A. Cummer, Z. Qin, and M. Chen, 2019: Lightning initiation processes imaged with very high frequency broadband interferometry. *Journal of Geophysical Research: Atmospheres*, **124**, 2994–3004, doi:10.1029/2018JD029817.

MacGorman, D. R., and W. D. Rust, 1998: *The electrical nature of storms*. 1st ed., Oxford University Press, 198 Madison Avenue, New York, New York 10016.

Mach, D. M., 2020: Geostationary lightning mapper clustering algorithm stability. *Journal of Geophysical Research: Atmospheres*, **125 (5)**, e2019JD031 900, doi:10.1029/2019JD031900, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD031900, e2019JD031900 2019JD031900, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019JD031900.

Marchand, M., K. Hilburn, and S. D. Miller, 2019: Geostationary lightning mapper and earth networks lightning detection over the contiguous united states and dependence on flash characteristics. *Journal of Geophysical Research: Atmospheres*, **124 (21)**, 11 552–11 567, doi:10.1029/2019JD031039, URL https://agupubs-onlinelibrary-wiley-com-s.docadis.ups-tlse.fr/doi/abs/10.1029/2019JD031039, 2019JD031039, https://agupubs-onlinelibrary-wiley-com-s.docadis.ups-tlse.fr/doi/pdf/10.1029/2019JD031039.

Massey, F. J., 1951: The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, **46 (253)**, 68–78, URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1951.10500769.

Mastelini, S. M., V. G. T. da Costa, E. J. Santana, F. K. Nakano, R. C. Guido, R. Cerri, and S. Barbon, 2019: Multi-output tree chaining: An interpretative modelling and lightweight multi-target approach. *Journal of Signal Processing Systems*, **91 (2)**, 191–215, doi:10.1007/s11265-018-1376-5, URL https://doi.org/10.1007/s11265-018-1376-5.

Murphy, M. J., and R. K. Said, 2020: Comparisons of lightning rates and properties from the u.s. national lightning detection network (nldn) and gld360 with goes-16 geostationary lightning mapper and advanced baseline imager data. *Journal of Geophysical Research: Atmospheres*, **125 (5)**, e2019JD031 172, doi:10.1029/2019JD031172, URL https://agupubs-onlinelibrary-wiley-com-s.docadis.ups-tlse.fr/doi/abs/10.1029/2019JD031172, e2019JD031172 2019JD031172, https://agupubs-onlinelibrary-wiley-com-s.docadis.ups-tlse.fr/doi/pdf/10.1029/2019JD031172.

Nag, A., M. J. Murphy, W. Schulz, and K. L. Cummins, 2015: Lightning locating systems: Insights on characteristics and validation techniques. *Earth and Space Science*, **2**, 65–93, doi:10.1002/2014EA000051.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830.

Peterson, M., S. Rudlosky, and D. Zhang, 2020: Changes to the appearance of optical lightning flashes observed from space according to thunderstorm organization and structure. *Journal of Geophysical Research: Atmospheres*, **125 (4)**, e2019JD031 087, doi:10.1029/2019JD031087, URL https://agupubs-onlinelibrary-wiley-com-s.docadis.ups-tlse.fr/doi/abs/10.1029/2019JD031087, e2019JD031087 2019JD031087, https://agupubs-onlinelibrary-wiley-com-s.docadis.ups-tlse.fr/doi/pdf/10.1029/2019JD031087.

Rison, W., R. Thomas, P. Krehbiel, T. Hamlin, and J. Harlin, 1999: A gps-based three-dimensional lightning mapping system: Initial observations in central new mexico. *Geophysical Research Letters*, **26 (23)**, 3573–3576, doi:10.1029/1999GL010856.

Rutledge, S. A., K. A. Hilburn, A. Clayton, B. Fuchs, and S. D. Miller, 2020: Evaluating geostationary lightning mapper flash rates within intense convective storms. *Journal of Geophysical Research: Atmospheres*, **125 (14)**, e2020JD032 827, doi:10.1029/2020JD032827, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020JD032827, e2020JD032827 2020JD032827, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020JD032827.

Schultz, E. V., C. J. Schultz, L. D. Carey, D. J. Cecil, and M. Bateman, 2016: Automated storm tracking and the lightning jump algorithm using GOES-R geostationary lightning mapper (GLM) proxy data. *Journal of Operational Meteorology*, **4(7)**, 92–107, doi:10.15191/nwajom.2016.0407.

Schulz, W., G. Diendorfer, S. Pedeboy, and D. R. Poelman, 2016: The european lightning location system euclid – part 1: Performance analysis and validation. *Nat. Hazards Earth Syst. Sci.*, **16**, 595–605, doi:10.5194/nhess-16-595-2016.

Smola, A. J., and B. Schölkopf, 2004: A tutorial on support vector regression. *Statistics and Computing*, **14**, 199–222, doi:10.1023/B:STCO.0000035301.49549.88.

Spyromitros-Xioufis, E., G. Tsoumakas, W. Groves, and I. Vlahavas, 2016: Multi-target regression via input space expansion: treating targets as inputs. *Machine Learning*, **104 (1)**, 55–98, doi:10.1007/s10994-016-5546-z, URL https://doi.org/10.1007/s10994-016-5546-z.

Stano, G. T., 2013: Fusing total lightning data with aviation weather center and storm prediction center operations during the GOES-R visiting scientist program. AMS Sixth Conference on the Meteorological Applications of Lightning Data, 724, Austin, TX, USA.

Takahashi, T., S. Sugimoto, T. Kawano, and K. Suzuki, 2017: Riming Electrification in Hokuriku Winter Clouds and Comparison with Laboratory Observations. *Journal of the Atmospheric Sciences*, **74 (2)**, 431–447, doi:10.1175/JAS-D-16-0154.1, URL https://doi.org/10.1175/JAS-D-16-0154.1, https://journals.ametsoc.org/jas/article-pdf/74/2/431/3869861/jas-d-16-0154_1.pdf.

Thomas, R. J., P. R. Krehbiel, W. Rison, T. Hamlin, D. J. Boccippio, S. J. Goodman, and H. J. Christian, 2000: Comparison of ground-based 3-dimensional lightning mapping observations with satellite-based lis observations in oklahoma. *Geophysical Research Letters*, **27 (12)**, 1703–1706, doi:10.1029/1999GL010845.

Thomas, R. J., P. R. Krehbiel, W. Rison, S. J. Hunyady, W. P. Winn, T. Hamlin, and J. Harlin, 2004: Accuracy of the lightning mapping array. *Journal of Geophysical Research*, **109**, D14 207, doi:10.1029/2004JD004549.

Vaisala, 2013: Vaisala thunderstorm advanced total lightning sensor ls7002. URL https://www.vaisala.com/sites/default/files/documents/WEA-LS7002-Datasheet-B211284EN-A-LOW.pdf, accessed on 2018-06-28.

Ávila, E. E., R. E. Bürgesser, N. E. Castellano, A. B. Collier, R. H. Compagnucci, and A. R. Hughes, 2010: Correlations between deep convection and lightning activity on a global scale. *Journal of Atmospheric and Solar-Terrestrial Physics*, **72 (14)**, 1114 – 1121, doi:https://doi.org/10.1016/j.jastp.2010.07.019, URL http://www.sciencedirect.com/science/article/pii/S1364682610002154.

1371 Yang, J., Z. Zhang, C. Wei, F. Lu, and Q. Guo, 2017: Introducing the new generation of chi-
1372 nese geostationary weather satellites, fengyun-4. *Bulletin of the American Meteorological*
1373 *Society*, **98 (8)**, 1637–1658, doi:10.1175/BAMS-D-16-0065.1.

1374 Zhang, D., and K. L. Cummins, 2020: Time evolution of satellite-based op-
1375 tical properties in lightning flashes, and its impact on glm flash detection.
1376 *Journal of Geophysical Research: Atmospheres*, **125 (6)**, e2019JD032 024, doi:
1377 10.1029/2019JD032024, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/
1378 2019JD032024, e2019JD032024 2019JD032024, https://agupubs.onlinelibrary.wiley.com/
1379 doi/pdf/10.1029/2019JD032024.

1380 Zhu, Y., V. A. Rakov, M. D. Tran, and A. Nag, 2016: A study of national lightning
1381 detection network responses to natural lightning based on ground truth data acquired
1382 at log with emphasis on cloud discharge activity. *Journal of Geophysical Research:*
1383 *Atmospheres*, **121 (24)**, 14,651–14,660, doi:https://doi.org/10.1002/2016JD025574,
1384 URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JD025574, https://
1385 agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016JD025574.

LIST OF TABLES

1415 TABLE 1. Study dates (year 2018) with the amounts of GLM and NLDN data. The rightmost
1416 columns indicate whether the data are used for ML-based target generator (indicated ML) or the
1417 test part, the time of most lightning activity in the region (D: local daytime, N: local nighttime),
1418 and the primary forcing for storm development and lightning.

Date	number of GLM events	number of GLM flashes	number of NLDN pulses, strokes	number of NLDN flashes	Usage	Time	Trigger
19 Mar	4,053,599	79,420	315,854	78,351	ML	D+N	cyclone, cold front
29 Mar	2,611,064	35,822	122,772	37,931	ML	D+N	stationary front, MCS
07 Apr	5,854,407	94,447	494,686	113,978	Test	D+N	short wave trough, front
14 Apr	8,610,567	142,587	729,622	169,181	ML	D+N	cold front
26 May	4,364,985	130,632	422,193	120,608	Test	D	thermal convection
03 Jun	6,103,693	204,295	825,601	188,330	ML	D+N	cold front
21 Jul	5,541,425	150,363	943,644	142,023	ML	D+N	squall line, outflow boundary
31 Jul	4,885,532	114,133	391,602	106,142	ML	D	dry line, thermal convection
07 Aug	5,283,358	153,671	472,369	137,963	Test	D	thermal convection
13 Sep	1,015,483	28,301	61,124	21,168	ML	D	Hurricane Florence

1419 TABLE 2. Comparison of distribution statistics for observed GLM data and the best generator
1420 for each target on 07 April 2018. Details about the target generator names are provided in the
1421 Appendix B0.b.

Generator	Mean	Median	Minimum	Maximum	RMSE per flash	KS	CvM
GLM flash duration [s]							
Observed	0.62	0.45	0.00	16.44	0.00	0.00	0.0
BAGR KNN dist num ext default	0.57	0.46	0.01	10.56	0.92	0.25	949.2
GLM event number per flash [-]							
Observed	73.5	46.0	2	1395	0.0	0.00	0.0
BAGR KNN dist num raw	57.8	48.0	3	467	99.7	0.24	996.6
GLM flash extent [km]							
Observed	38.5	34.8	0.0	277.0	0.0	0.00	0.0
BAGR KNN dist num ext(a2) raw plus	41.24	35.1	0.0	166.0	33.7	0.24	582.7

1422 TABLE 3. Comparison of distribution statistics for observed GLM data and the best generator
1423 for each target on 26 May 2018. Details about the target generator names are provided in the
1424 Appendix B0.b.

Generator	Mean	Median	Minimum	Maximum	RMSE per flash	KS	CvM
GLM flash duration [s]							
Observed	0.34	0.26	0.00	7.42	0.00	0.00	0.0
linSVR raw	0.41	0.25	0.01	4.73	0.56	0.21	599.9
GLM event number per flash [-]							
Observed	34.8	19.0	2	775	0.0	0.00	0.0
linSVR num ext(a) raw plus	32.2	29.0	3	341	54.0	0.45	3593.1
GLM flash extent [km]							
Observed	28.0	20.2	0.0	218.6	0.0	0.00	0.0
linSVR num ext raw	27.4	26.1	0.0	154.8	27.8	0.28	1029.2

1425 TABLE 4. Comparison of distribution statistics for observed GLM data and the best generator
1426 for each target on 31 July 2018. Details about the target generator names are provided in the
1427 Appendix B0.b.

Generator	Mean	Median	Minimum	Maximum	RMSE per flash	KS	CvM
GLM flash duration [s]							
Observed	0.39	0.29	0.00	9.21	0.00	0.00	0.0
linSVR num ext(a) plus	0.39	0.30	0.02	5.34	0.58	0.32	1307.9
GLM event number per flash [-]							
Observed	45.7	23.0	2	883	0.0	0.00	0.0
linSVR num ext(a) raw plus	33.9	30.0	2	316	76.4	0.40	2441.8
GLM flash extent [km]							
Observed	33.8	27.5	0.0	242.2	0.0	0.00	0.0
linSVR raw	29.6	26.2	0.0	180.6	33.5	0.22	745.2

1428 TABLE 5. Comparison of distribution statistics for observed GLM data and the best generator
1429 for each target during the full test period. Details about the target generator names are provided
1430 in the Appendix B0.b.

Generator	Mean	Median	Minimum	Maximum	RMSE per flash	KS	CvM
GLM flash duration [s]							
Observed	0.43	0.31	0.00	16.44	0.00	0.00	0.0
linSVR num ext raw	0.46	0.30	0.00	9.41	0.77	0.15	656.5
GLM event number per flash [-]							
Observed	49.3	25.0	2	1395	0.0	0.00	0.0
linSVR num ext(a) raw plus	35.3	30.0	2	411	79.8	0.38	6687.6
GLM flash extent [km]							
Observed	32.9	27.5	0.0	277.0	0.0	0.00	0.0
linSVR num ext(a2)	30.1	26.3	0.0	157.3	30.9	0.24	3479.2

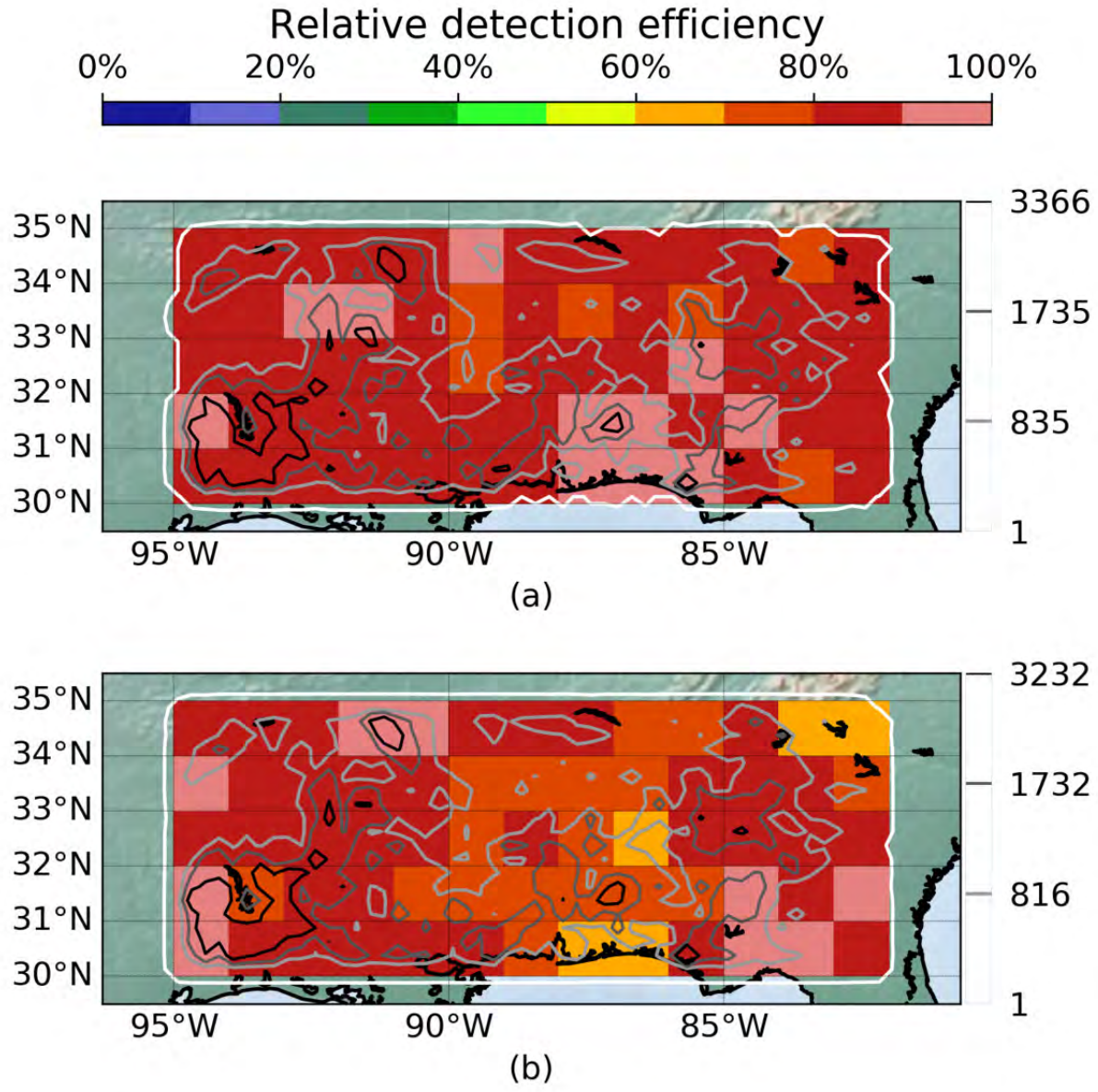
Table A1. ML model types with abbreviation.

ML model type	Abbreviation
Multivariate Linear Regression	*LinReg*
Multivariate Polynomial Regression	*Poly*
Random Forest Regressor	*RF*
Random Forest Extra Trees Regressor	*ETR*
Bagging Regressor with k-Nearest Neighbor Regressor (distance weighting)	*BAGN KNN dist*
Multilayer Perceptron Neural Network	*MLP*
(Linear) Support Vector Regressor	*SVR* (*linSVR*)
Histogram-based Gradient Boosting Regression Tree	*HGBR*

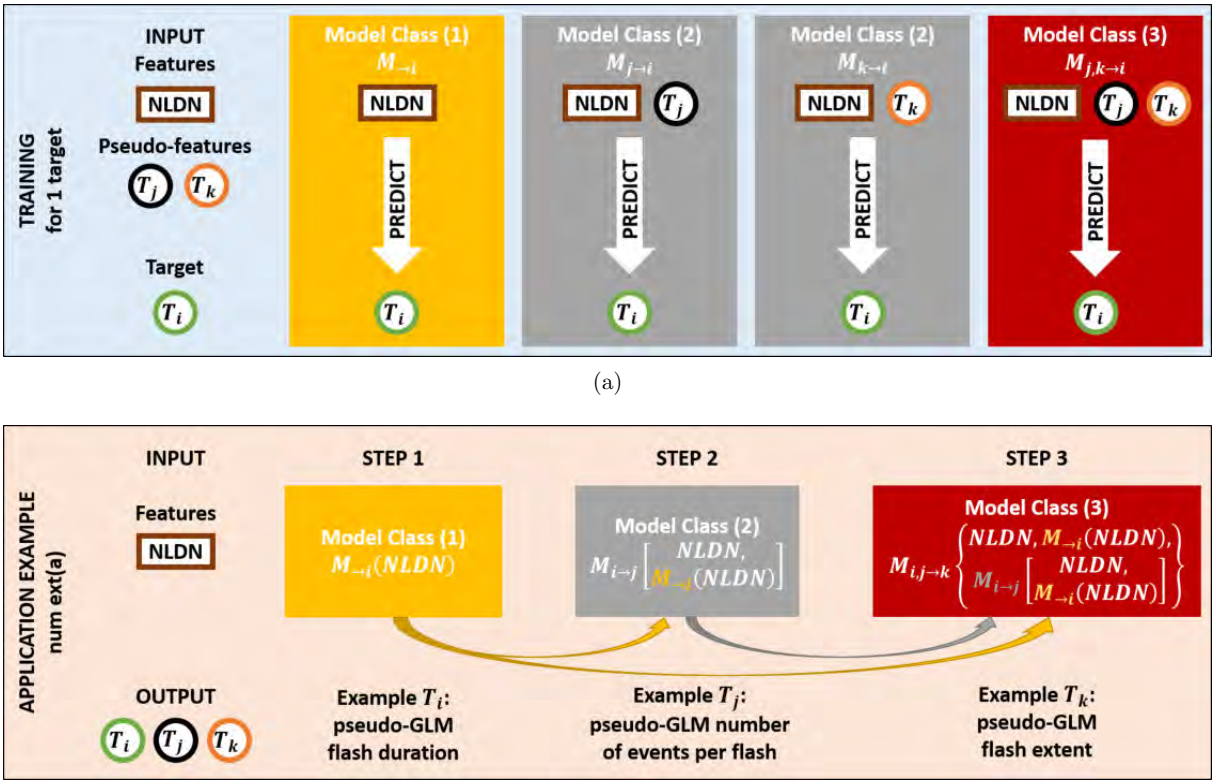
Table A2. Naming conventions of used target generator configurations. The name extensions as of column 1 are used following the ML model type. The three columns indicate the utilized features during the ML training for each the of three targets GLM flash duration (Flash Duration), number of events per flash (Event Number), and GLM flash extent (Flash Extent). NLDN indicates that NLDN flash duration, the number of strokes+pulses per flash, NLDN flash extent, and the maximum LF amplitude are used as features. Combination of different extensions are possible.

Name extension	Flash Duration	Event Number	Flash Extent
default	NLDN	NLDN	NLDN
(a)	Training with default configuration. Model predictions within the training uncertainty of the default model (Initial training step while training further alternative models)		
plus	NLDN with mean LF amplitude and CG fraction	NLDN with mean LF amplitude and CG fraction	NLDN with mean LF amplitude and CG fraction
raw	features and target not scaled	features and target not scaled	features and target not scaled
num	NLDN	NLDN + Flash Duration + Flash Extent	NLDN
num(a)	NLDN	NLDN + Flash Duration	NLDN
num ext	NLDN	NLDN + Flash Duration + Flash Extent	NLDN + Flash Duration + Event Number
num ext(a)	NLDN	NLDN + Flash Duration	NLDN + Flash Duration + Event Number
num ext(a2)	NLDN	NLDN + Flash Duration	NLDN + Flash Duration

LIST OF FIGURES



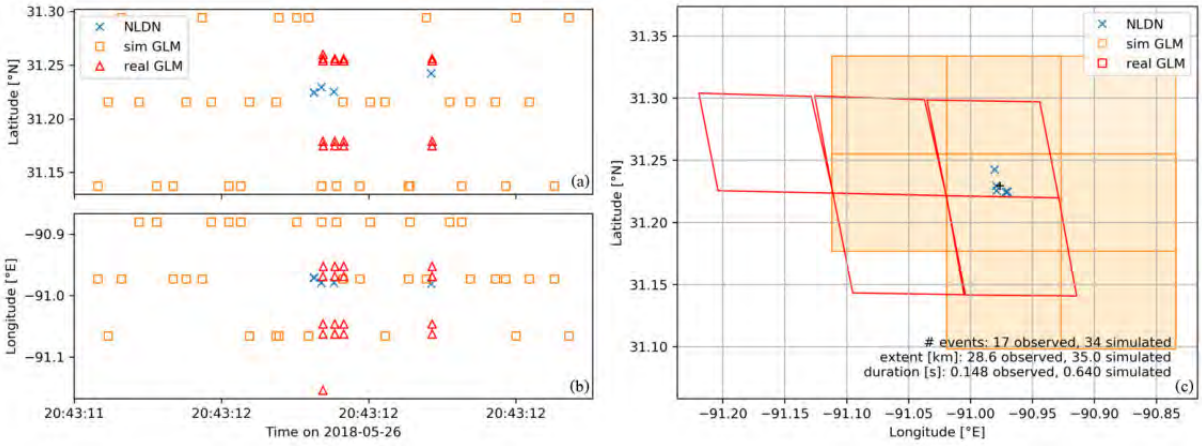
1487 FIG. 1. Relative detection efficiency per $1^\circ \times 1^\circ$ pixel (color) for the full 10-day dataset. In (a)
1488 for GLM and in (b) for NLDN. Grey-scale lines contour the flash number at the 0th (1 flash), 50th,
1489 80th, and 95th percentile of the flash number distribution per $0.25^\circ \times 0.25^\circ$ pixel (only for pixels
1490 with flash activity).



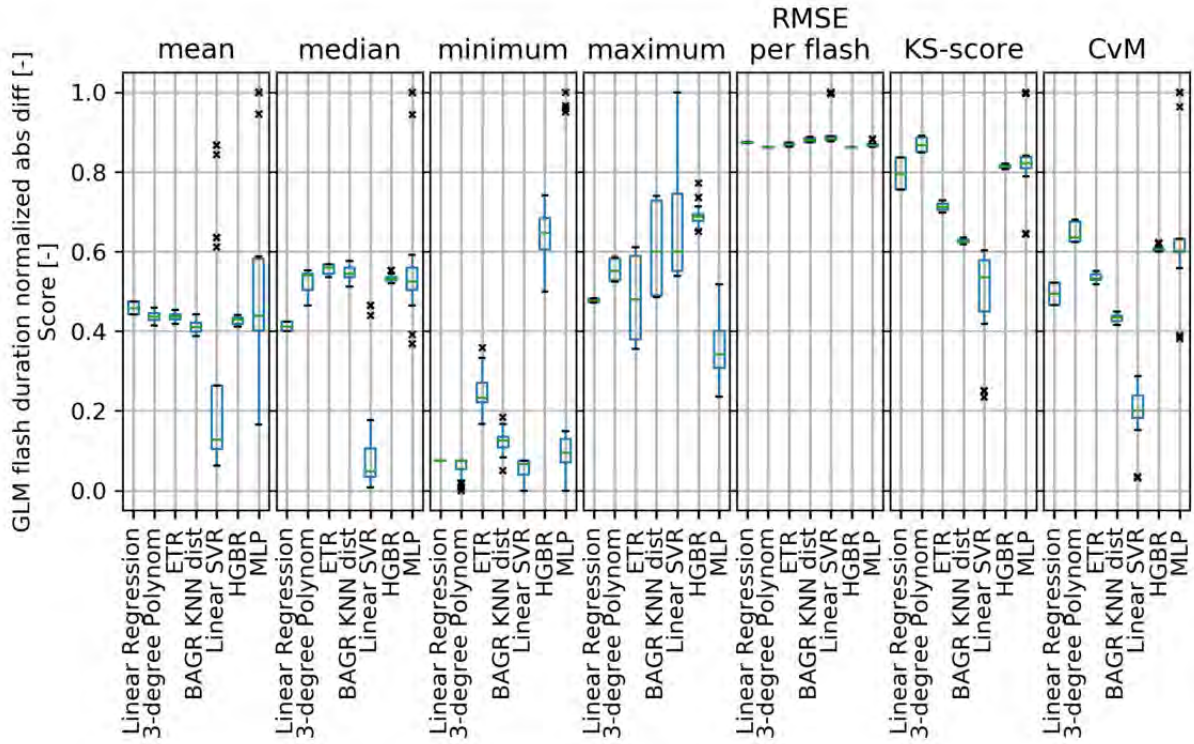
(a)

(b)

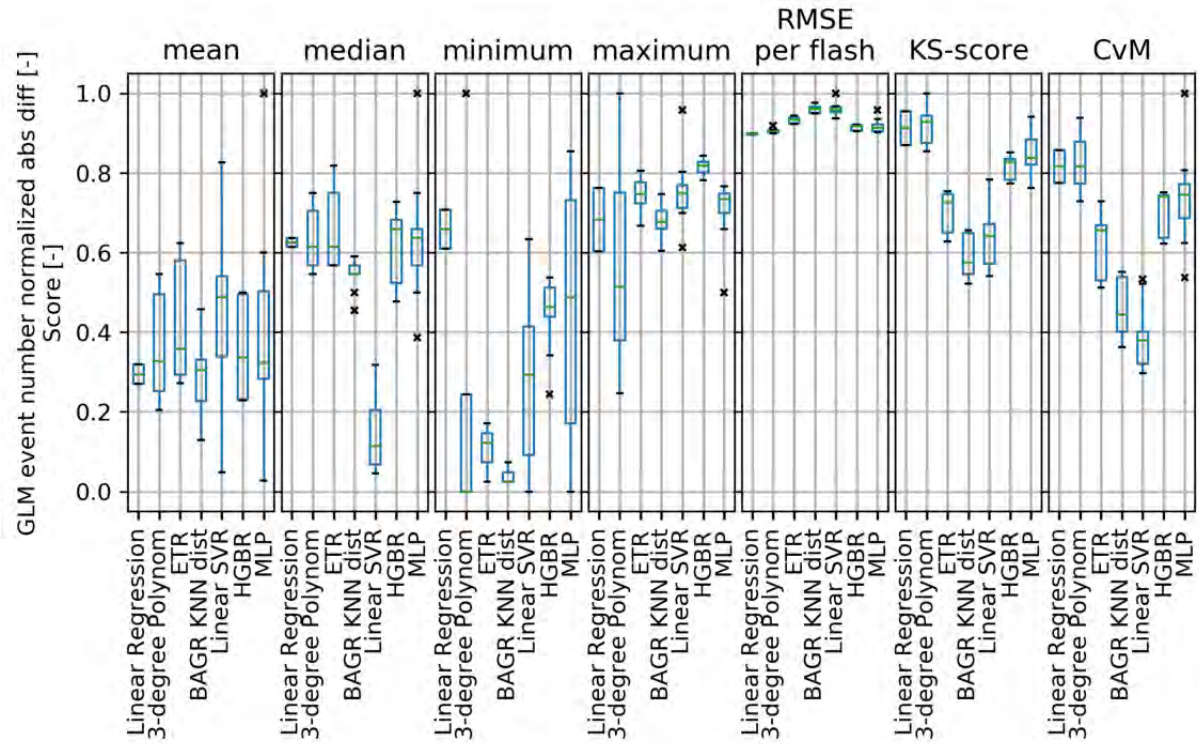
FIG. 2. Flow chart of the multi-step approach illustrating the possible predictions of a given target using different combinations of features and pseudo-features (a, Training). The Application (b) shows the example of the *num ext(a)* configuration (Appendix B0.b).



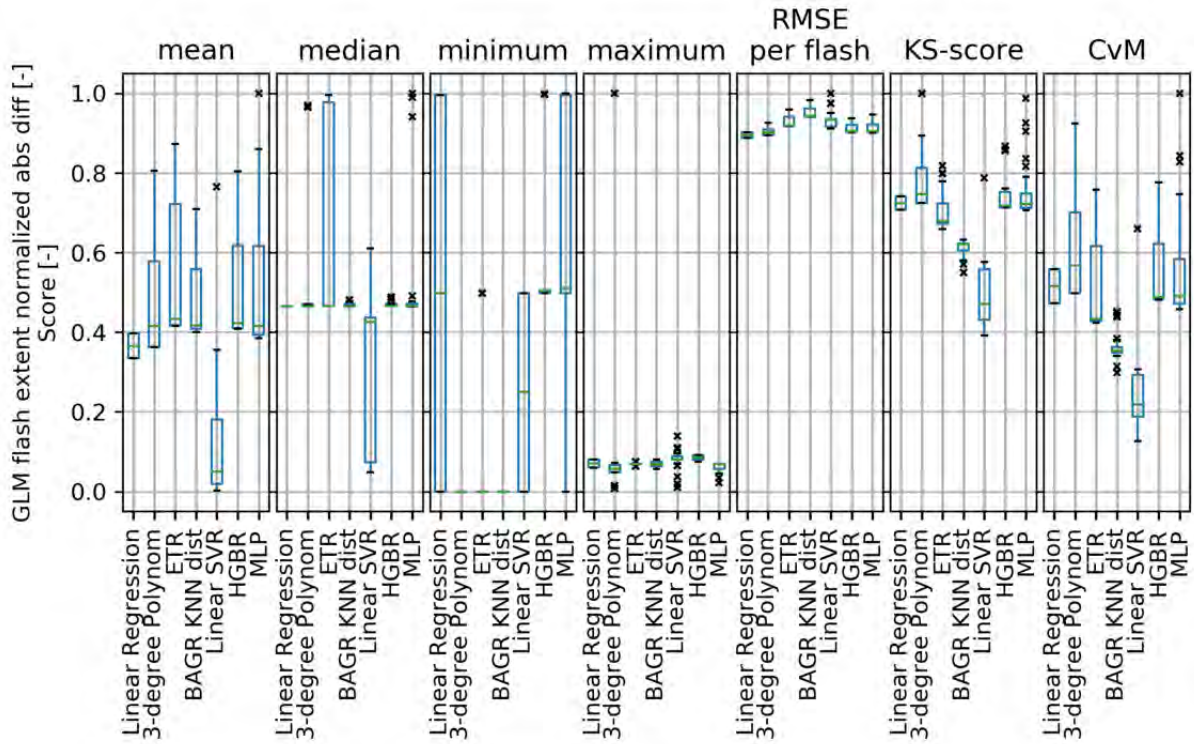
1494 FIG. 3. An example of one simulated flash with corresponding GLM and NLDN observations
1495 on 26 May 2018. The final GEO lightning pseudo-observation generator is used including a linear
1496 SVR model. Time series of latitudes (a), longitudes (b), and a map (c). The map (c) includes
1497 characteristics of the observed and simulated GLM flash. The time interval shown matches the
1498 simulated flash duration of 640 ms



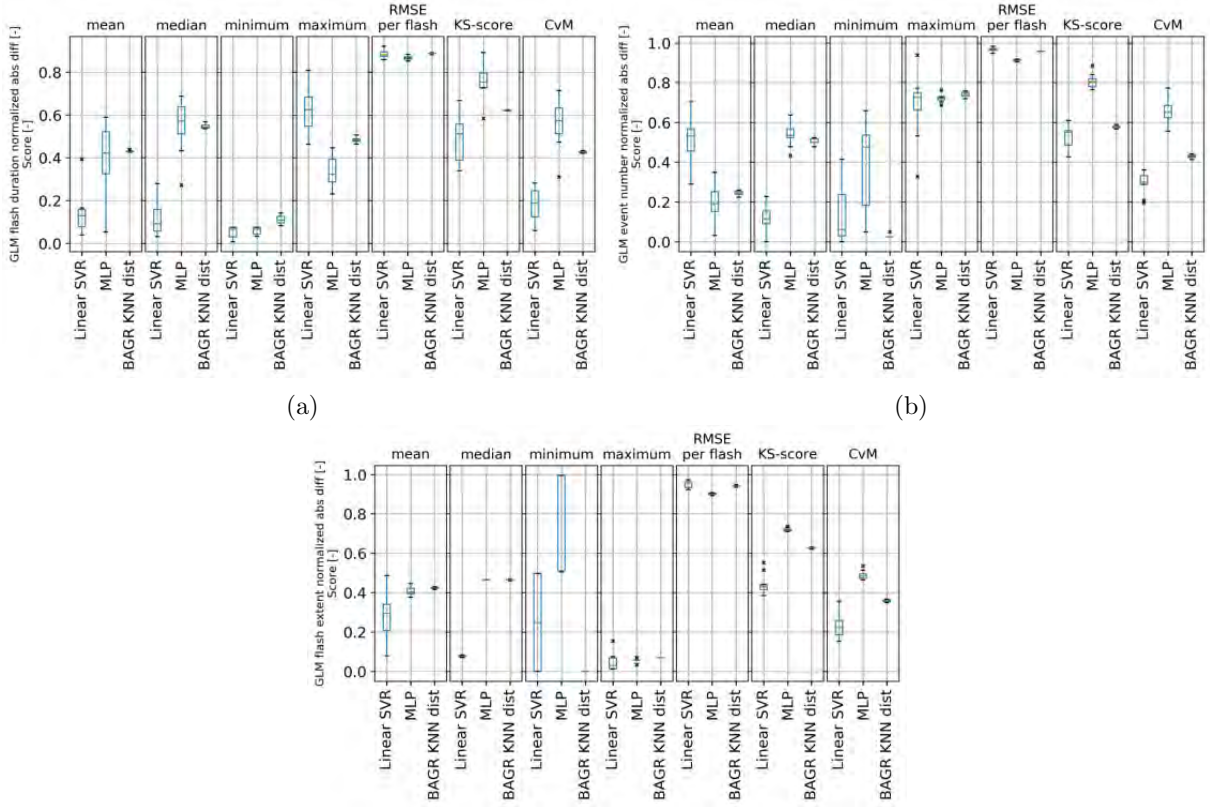
1499 FIG. 4. Normalized absolute difference of statistics and scores (titles) between distributions of
1500 observed and simulated GLM flash duration (0 means equal to observation, 1 represents the worst
1501 simulation). The boxplots represent the distributions of 28 target generator results per ML type
1502 (x-axis) including the Inter-Quartile-Range (IQR, blue box), 1.5 times the IQR (whiskers), and
1503 outliers (black cross). The horizontal green line give the median. Results for the full test dataset.
1504 ML type abbreviations provided in Table A1.



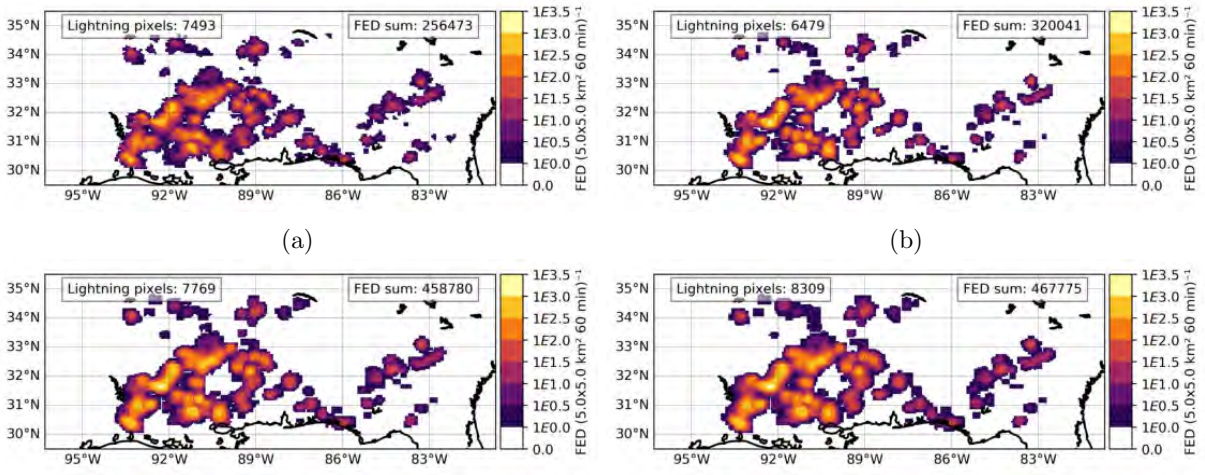
1505 FIG. 5. As Figure 4 but for the normalized absolute difference of statistics and scores (titles)
1506 between distributions of observed and simulated GLM event number per flash (0 means equal to
1507 observation, 1 represents the worst simulation).



1508 FIG. 6. As Figure 4 but for the normalized absolute difference of statistics and scores (titles)
1509 between distributions of observed and simulated GLM flash extent (0 means equal to observation,
1510 1 represents the worst simulation).



1511 FIG. 7. Normalized absolute difference of statistics and scores (titles) between distributions of
1512 observed and simulated GLM flash duration (a), event number per flash (b), and flash extent (c);
1513 0 means equal to observation, 1 represents the worst simulation as of Figure 4 to 6, respectively.
1514 Boxplots (as in 4) represent the distribution for training the same model (x-axis) 10 times during
1515 the first step of the simulation. ML type abbreviations provided in Table A1.



(a) (b)

(c) (d)

1516 FIG. 8. Observed (a) and simulated hourly FED using one *linSVR* (b), one *MLP* (c), and one
1517 *BAGR KNN dist* (d) model on 26 May 2018, 20:00 to 21:00 UTC. The FED grid uses pixels of
1518 5 km × 5 km. ML type abbreviations provided in Table A1.

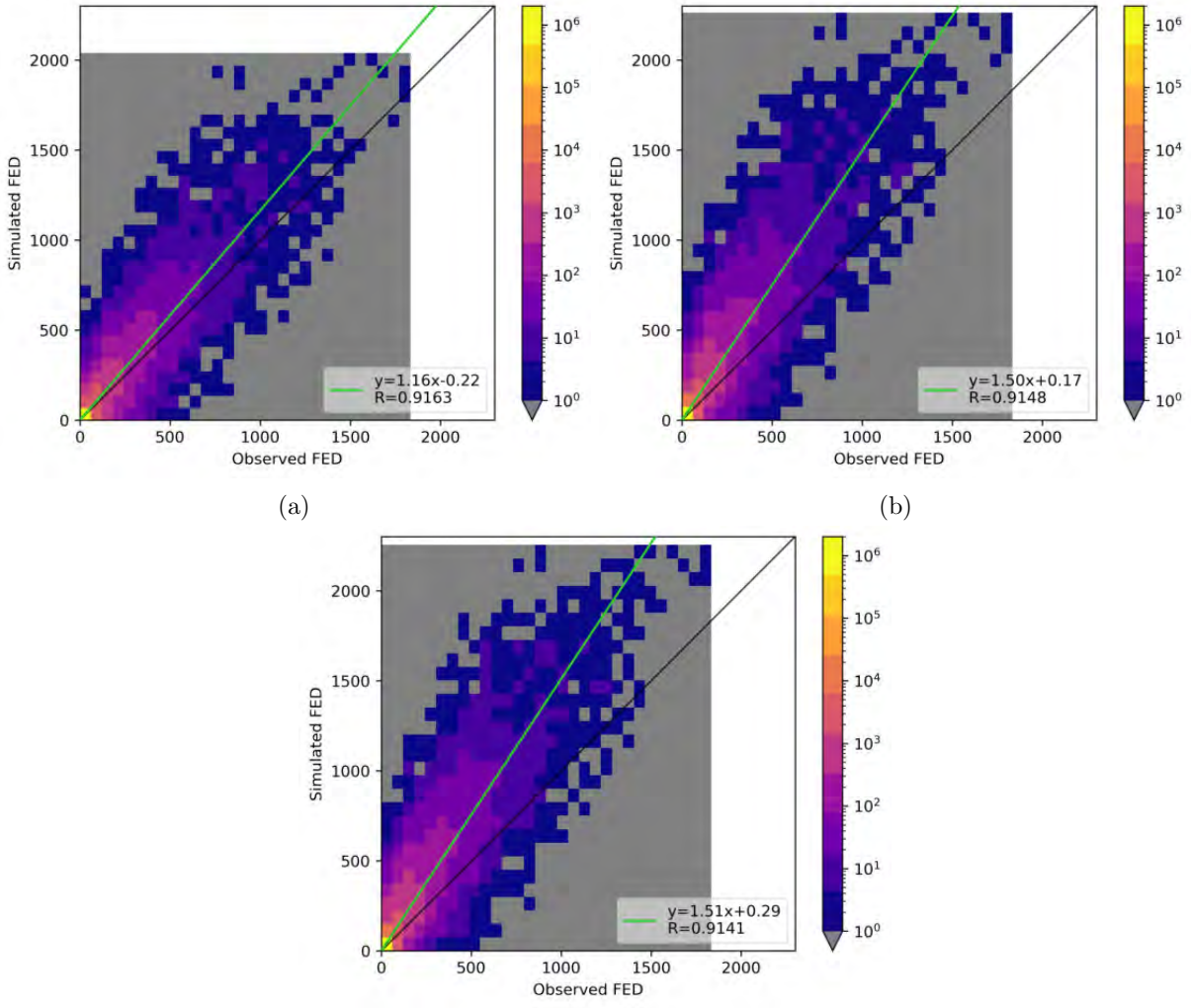
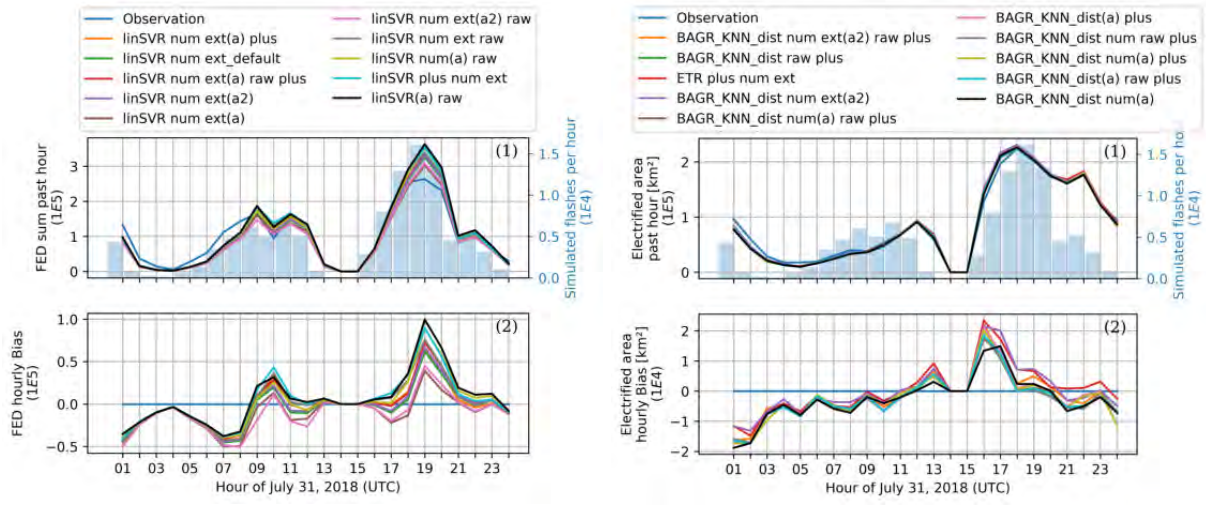
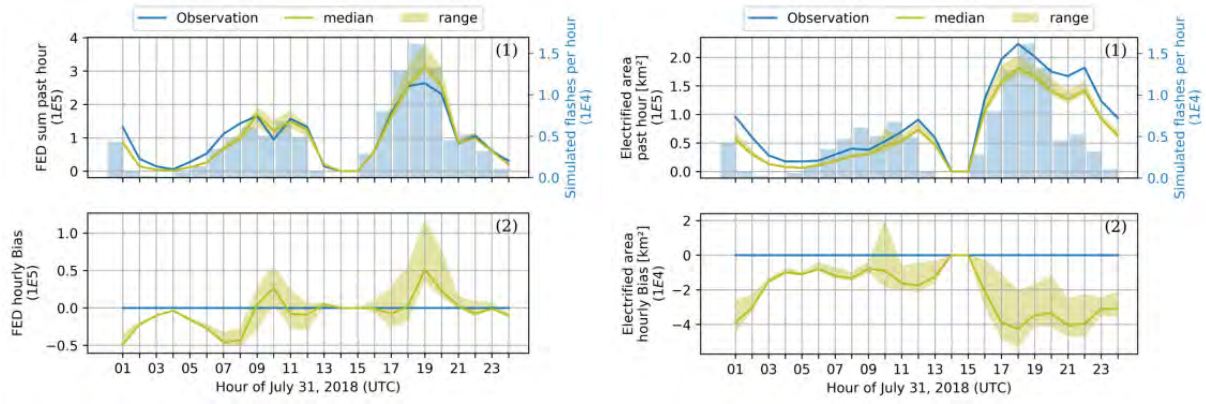


FIG. 9. Pixel-to-pixel (5 km × 5 km) simulated versus observed hourly FED for the 3-day test period using the same *linSVR* (a), *MLP* (b), and *BAGR KNN dist* (c) models as in Figure 8. The gray box and white margins indicate the upper limits of distributions on each axis. ML type abbreviations provided in Table A1.



(a) (b)

1523 FIG. 10. Hourly sum of FED (a) and hourly electrified area (b) within the region of interest.
1524 Top (1): Absolute values and number of simulated flashes per hour. Bottom (2): Difference of
1525 simulation minus observation. The observation is plotted in blue, the remaining colors represent
1526 the 10 best models for both FED sum (a) and electrified area (b), respectively. Results for 31 July
1527 2018. Details about the model names are provided in Appendix B0.b.



1528 FIG. 11. As Figure 10 with 10 repetitions of the selected *linSVR*-based generator. Median (line)
1529 and range (shaded) of 10 generator repetitions for 31 July 2018.

II.4.3 Conclusions for the LDA

Operational GEO lightning data are available over the Americas and China. To date, such observations are not provided over Europe. Our GEO lightning pseudo-observation generator has been developed using GLM and NLDN in the US. Chapter II.2 revealed that the French LF network Meteorage observes lightning flashes in a similar manner as NLDN does. In particular, Meteorage can serve as a suitable ground-based lightning data source to generate over France pseudo-optical observation from our GEO lightning pseudo-observation generator. By design, the algorithm is capable of producing pseudo-observations with adjustable spatial and temporal resolution. MTG-LI will have somewhat higher spatial resolution than GLM, and the integration time frame will be 1 ms rather than the 2 ms of GLM, hence, the algorithm should then be adapted to MTG-LI parameters. This pseudo-synthetic dataset provides the basis for the assimilation of GEO lightning data in AROME-France. The pseudo-observations are generated on a regular latitude-longitude grid. Real MTG-LI observations will use an irregular grid with irregular pixel shapes that result from the projection of the regular pixel grid on Earth. The pseudo MTG-LI data is generated based on LF observations that might suffer problems such as data sparse regions, e.g., over mountains, and limited detection range. MTG-LI will not be affected by similar problems. A future comparison between the pseudo MTG-LI and real MTG-LI observation may unveil the necessity for adapting the LDA scheme that is built on the GEO lightning pseudo-observations as will be elucidated in Chapter III.4. Nevertheless, the tools for GEO LDA in AROME-France are developed for the real GEO data. It should be also reminded that the GEO lightning pseudo-observation generator was trained for continental terrain, while the AROME-France domain contains parts of the Atlantic Ocean and Mediterranean Sea. The consequent impact on the results of the generator is not known and should be the subject of future investigations.

Our GEO lightning pseudo-observation generator processes and produces flash level data since FED can be derived from the generated pseudo MTG-LI flashes. Moreover, the flash level allows for defining several flash characteristics for both the input LF data, i.e., the features, and the generated GEO data, i.e., the targets. Features and targets are related in the multivariate training algorithm with the objective of creating pseudo GEO flashes that do not only approximate real GEO lightning observations in terms of location and time of occurrence, but also in terms of the physical flash characteristics. The flash level approach is the prevailing difference to an existing pseudo MTG-LI created for EUMETSAT. Höller and Betz (2010) developed a pseudo MTG-LI generator on the pulse/stroke-group level using LINET and TRMM-LIS observations. As only physical measurement, besides location and time of a LINET stroke or pulse, they used the peak LF current to simulate the radiance of an optical pseudo-group coincident to each pulse/stroke. The correlation between the optical radiance and the LF peak current was, however, rather weak, as Höller and Betz (2010) state a Pearson correlation coefficient of 0.26. A weak relation between measured LF currents and the observations of optical flashes from space was also concluded in Section II.1.2 and chapters II.2 and II.3. Höller and Betz (2010) then simulated additional groups around each observed LINET pulse/stroke based on a Gaussian distribution fitted on the observed TRMM-LIS group times relative to LINET pulses/strokes, and observed TRMM-LIS group statistics as they found on average 7 groups per pulse/stroke. Eventually group footprints were inferred directly from the simulated radiance. As stated above, a pulse/stroke-group level generator is limited in that a single measurement is used, that only provides location, time, and LF current. In addition, the TRMM-LIS and LINET data were only available in tropical regions, i.e., in Brazil, and restricted to times with TRMM-LIS overpasses while lightning occurred. The results might in consequence be not

valid in extratropical regions such as France, and the statistical analysis comprised only 705 coincident flashes overall (Höller and Betz, 2010).

It should be mentioned that Höller and Betz (2010) suggest a pixel matrix using an orthogonal projection to transform group locations and footprints to pseudo MTG-LI events. This approach is more sophisticated than the regular latitude-longitude event pixel matrix used by our GEO lightning generator for reasons of (i) computational efficiency, and (ii) a domain limited to France instead of the full MTG-LI FOV. The approach of Höller and Betz (2010) might be applied to create an advanced pixel matrix that would then give to our generator the capability of producing GEO lightning pseudo-observations on any given grid.

Our GEO lightning pseudo-observation generator considers and mimics the area of cloud illuminations that is usually wider than the area covered by the corresponding LF observations. This behavior was also seen in Erdmann et al. (2020a) (Chapter II.1), Chapter II.2, and Chapter II.3 where the flashes observed from the satellite instruments extended further than the coincident LF flashes.

The recommended GEO lightning pseudo-observation generator provides the best trade-off with respect to minimal differences between simulation and observation in the target statistics, FED sum, and electrified area. FED sums cumulate the hourly FED values within the studied regions, while the electrified areas are obtained from the number of FED pixels with lightning activity. In general, the generators that perform well for the FED sum underestimate the electrified area, and generators that predict the electrified area well overestimate the FED sum. As the differences between observation and simulation are one order of magnitude smaller relative to the overall observed values for the electrified area than for the FED sum, the recommended generator is more focused on estimating the FED sum correctly. The generation of pseudo MTG-LI FED for the assimilation might explore in addition to the recommended *linSVR*-based generator the use of *BAGR KNN dist*-based or *MLP*-based generators that perform superiorly for the electrified area.

The work of Section II.4.2 is heavily built on machine learning (ML) methods. The paper uses an immense amount of different approaches, in total 196, including ML mode types, scaling techniques, and a newly developed multi-step approach to benefit from correlation between ML targets, i.e., the GLM flash characteristics. The development of the approaches and the evaluation of all results are very time consuming. There might be room for improvement, especially to include alternative, existing ML methods that take target correlations into account (examples are discussed in Section II.4.2).

Part III

Assimilation of Flash Extent Density (FED)

In 1904, Vilhelm Bjerknes published a paper (Bjerknes, 2009) that can be seen as the first in-depth vision of Numerical Weather Prediction (NWP) (e.g., Lackmann, 2012). The paper includes a set of governing equations and ideas of an NWP process. Among others, two necessary and sufficient conditions for predicting the state of the system that develops from an initial state following physical laws are defined: (i) A sufficiently accurate knowledge of the state of the atmosphere at the initial time, and (ii) sufficiently accurate knowledge of the laws according to which one state of the atmosphere develops from another.

The latter is nowadays part of NWP that includes solving the set of governing equations, parameterization, and approximation (e.g., scaling). The former condition can be addressed by using observations during the production of the initial state. The observations should ideally be available continuously in space and time, have no errors, and represent all quantities predicted by the NWP model. Obviously, that is not the case in practice. A good estimation of the initial state can thus not only rely on observations but needs additional information. Data Assimilation (DA) is the process of combining different sources of information to estimate the initial state of a system, i.e., of the atmosphere in weather prediction. Such information may contain (i) NWP prognostic equations, (ii) Observations, (iii) Model background (the a priori state, usually a short term forecast valid at the time of the initial state), and (iv) Climatology. Ensemble forecasts can build the weather forecast upon an ensemble of predictions that are initialized with slightly different initial conditions. A deterministic forecast, on the other hand, produces always the same forecast given a specific set of initial conditions.

Contents of this Part

III.1 | Definitions and Concepts of the Data Assimilation

The following sections introduce the important variables and concepts of the data assimilation in a general fashion. These theoretic constructs will be used to explain different data assimilation (DA) methods and in particular the lightning data assimilation (LDA) as objective of this thesis.

III.1.1 Variables

Before important assimilation concepts are introduced, the main variables used in the equations describing those concepts are defined. The nomenclature adapts the one from Ide et al. (1997) and Olzhabaev (2013). The NWP model, its assimilation system, and the detailed assimilation work are introduced in the following chapters (Chapter III.2 to Chapter III.4). The definitions given hereafter are general and do not include dimensions or information specific to the NWP model used in this thesis:

x The state vector x comprises all model prognostic variables. Data assimilation systems usually use a subset of the state vector during the analysis constituted of the so-called control variables. For simplicity, x is used as equivalent to the control variables to define the concepts of the data assimilation.

x_t The true state vector x_t, also called the real state, describes the actual state of the atmosphere. It serves as reference to evaluate an NWP. However, it is not known for all prognostic variables nor at all model grid points.

x_b The background state vector x_b defines the initial state to launch an NWP. It includes all prognostic variables and is usually based on a short-term forecast.

$\epsilon_b = x_b - x_t$ The background error ϵ_b as the difference the difference between x_b and x_t. The relations between all background errors are expressed in the background error covariance matrix $\mathbf{B}$.

x_a The analysis state vector x_a connects x_b with recent observations and constitutes the output of the DA. The new NWP forecast starts with x_a.

$\epsilon_a = x_a - x_t$ The analysis error ϵ_a is defined as difference between x_a and x_t, with analysis error covariance matrix $\mathbf{A}$.

y_o The observation vector y_o includes all observations that are valid at the given time and should be assimilated in the NWP model.

ϵ_o The observation error ϵ_o describes uncertainties and possible errors, e.g., of the instruments, but is sometimes difficult to quantify. The observation error covariance matrix is referred to as $\mathbf{R}$.

$H(x) + \epsilon_o = y_o$ The observation operator H relates the observation space represented as discrete vector y_o to a model state vector x and takes the observation errors ϵ_o into account. If H can be linearized, it can be expressed as matrix $\mathbf{H}$.

III.1.2 Basic relationships

Data assimilation needs relations between the observation space and the model space. A function h projects a state vector x in the model space on a vector in the observation space. The image of h is only a subset of the observation space. It is related to a set of actual observations at a given time that is referred to as the observation vector y_o. Each observation has a measurement error ϵ_m that must be considered. One vector of y_o is defined as

$$y_o = h(x) + \epsilon_m \tag{III.1.1}$$

Both the observations and the model are discrete, so there is no continuous function $h(x)$ in real problems. A set of calculations is introduced to relate the discrete observation vector and model state vector. The observation operator H summarizes this set of calculations. The observation operator is usually not perfect to link the true state x_t perfectly to the observations. An additional error term, the representation error ϵ_r, must be taken into account. For simplicity, this error term is usually added to the measurement error to form the observation error $\epsilon_o = \epsilon_m + \epsilon_r$. Equation (III.1.1) becomes

$$y_o = H\left(x_t\right) + \epsilon_o \tag{III.1.2}$$

This relation is used to compare the most recent forecast, also called the background state x_b, with the observation. The forecast (or background) error ϵ_b is added to the equation. Equation (III.1.2) can now be written as

$$y_o = H\left(x_b\right) + \epsilon_o + \epsilon_b \tag{III.1.3}$$

The NWP model alters an initial state of system x_k to the subsequent state x_{k+1}. The model forecasts are discrete in time. Therefore, there is not a continuous function but a model operator M that links the initial state and the subsequent state. The operator has a model error η. It is applied to the best estimation of the true state x_t, that is the analysis x_a (defined as the result of data assimilation that combines the background x_b and the observation vector y_o). The resulting forecast of x_t is also the new background for the subsequent data assimilation.

$$x_{k+1,b} = M_{k,k+1}\left(x_{k,a}\right) + \eta_{k,k+1} \tag{III.1.4}$$

III.1.3 Error statistics

This short interlude defines the mathematical meaning of error statistics as used in the following. The notation $E(\epsilon)$ means the (best) estimation for the error matrix ϵ that is used because the real error matrix is usually not known. Unbiased errors vanish when their average value is

taken. There can be non-zero errors in specific situations, such as short time periods or locally, however, unbiased errors fulfill the equation

$$E(\epsilon) = 0 \tag{III.1.5}$$

It is straightforward to give the definition of two uncorrelated errors ϵ_1 and ϵ_2

$$E\left(\epsilon_1 \epsilon_2^T\right) = 0 \tag{III.1.6}$$

Non-trivial errors mean that the error covariance matrix $Cov(\epsilon)$ is positive definite. The covariance matrix is defined as

$$Cov(x) = E\left([x - E(x)][x - E(x)]^T\right) \tag{III.1.7}$$

III.1.4 Modern data assimilation techniques

Data assimilation exists in a variety of forms and techniques. The early techniques changed the NWP equations to tune the model towards the observations. Other techniques aim at leaving the model equations untouched and allowing the model to adapt to the observations by itself. A brief summary of assimilation techniques that are in use is provided in the following. The basic ideas are given along with some of the main advantages and drawbacks of each technique. More details are provided by among others Blayo et al. (2011), Bouttier and Courtier (2002), and Bannister (2017).

III.1.4.1 Nudging

Nudging, or Newtonian relaxation (Anthes, 1974, Hoke and Anthes, 1976, Davies and Turner, 1977, Krishnamurti et al., 1988), is one of the oldest and less costly assimilation techniques. The model itself with its equations is modified to tune the (last) model forecast towards the observation. The principle is shown in Equation (III.1.8). $M(x)$ defines in that case a specific model forcing operator that corrects the prognostic variables toward the observations. The difference between the model state vector and the observation vector is included in the NWP equations with a specific scaling factor α that depends on the model and the type of forecast.

$$\frac{\delta x}{\delta t} = M(x) + \alpha \cdot (y - x) \tag{III.1.8}$$

Nudging corrects the model state towards the observation by adding a non-physical diffusive term (e.g., Zou et al., 1992). The Newtonian relaxation (Wang and Warner, 1988, Hoke and Anthes, 1976) enables 4D nudging and incremental modification of the model prognostic fields towards observations at each time step. A major concern about nudging is that the forcing of the model happens without solid theoretical foundation and physical balance in the model (e.g., Wang et al., 2017b). In addition, nudging only assimilates the model prognostic variables and thus limits the set of observations that can be assimilated. It also does not take into account any model and measurement errors.

III.1.4.2 Best Linear Unbiased Estimator (BLUE)

Under certain assumptions, an optimal assimilation can be defined. First, it is assumed that both background errors (ϵ_b, $\mathbf{B}$) and observation errors (ϵ_o, $\mathbf{R}$) are unbiased and non-trivial.

The ϵ_b and ϵ_o are uncorrelated. It is further assumed that H is a linear operator and thus can be expressed as a matrix $\mathbf{H}$. The solution to this least square estimation is referred to as Best Linear Unbiased Estimator (BLUE). The BLUE solution is an optimization in a least squares and minimal variance sense. It can use a direct stochastic computation, e.g., Kalman Filter (KF) or a variational optimization problem (Var). They yield identical results as long as linear operators are used. The analysis x_a can be expressed as a linear combination of background estimate and observation vector with minimal variance (diagonal of $\mathbf{A}$).

$$x_a = \mathbf{L}x_b + \mathbf{K}y_o \tag{III.1.9}$$

where $\mathbf{L}$ can be expressed in terms of $\mathbf{K}$ as

$$x_a = (\mathbf{1} - \mathbf{KH})\, x_b + \mathbf{K}y_o = x_b + \mathbf{K}\,(y_o - \mathbf{H}x_b) \tag{III.1.10}$$

with the gain or weight matrix $\mathbf{K}$ equal to

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T \left(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right)^{-1}$$

In the specific case of BLUE with the optimal least square gain $\boldsymbol{K}$, the analysis covariance matrix becomes

$$\mathbf{A} = (\mathbf{1} - \boldsymbol{K})\,\mathbf{B} \tag{III.1.11}$$

The BLUE equations could be solved, but usually non-linearity and correlated errors cause deviations from this optimal estimation. The concept can still be used with a non-linear model operator and H if it is adjusted, e.g., the extended Kalman Filter and Incremental Variational method (Blayo et al., 2011).

One relatively simple solution is called Optimal Interpolation (OI) analysis. Here, it is assumed that each variable x_i of x depends only on a limited number of observations. This simplifies the computation of $\boldsymbol{K}$. The lines of $\boldsymbol{K}$ can be calculated from matrices with limited dimensions for each x_i. The analysis is computed using Equation (III.1.10) and $\boldsymbol{K}$. In NWP applications, it is critical to choose the observations and to use a different set of observations for each model variable.

III.1.4.3 (Ensemble) Kalman Filter (KF)

The Kalman Filter (KF) continues the formulation of the BLUE and presents one solution. The matrix $\mathbf{K}$ from Equation (III.1.10) is often termed Kalman gain. As introduced, it weights the difference between observations and background correction. The complete set of equations describing the Kalman Filter should not be shown here since this thesis does not apply a KF approach (see e.g., Bouttier and Courtier, 2002). KF applies the concept of Equation (III.1.4) that a background state is provided by a forecast from the previous analysis. $\mathbf{K}$ is then computed sequentially from a forecast of error covariances using a linearized model forecast operator.

The standard KF has the issue of a large matrix $\mathbf{K}$, usually with too many dimensions to be computed. An ensemble of possible forecasts can help to estimate the background error covariance matrix. Hence, an ensemble of analyses with analysis error statistics is created. This approach defines the Ensemble KF (EnKF). The flow dependent background error statistics defining the covariance matrix of the background are not constant (as a climatology would be) but are calculated for every time step. In practice, the number of ensemble members is limited due to high computational costs. As a result, ensemble-derived correlations are often noisy. Recently, localization is applied to mitigate the problem of noise (Bannister, 2017, and references therein). Variances are underestimated as the ensemble collapses to one single trajectory (Bannister, 2017). The correction of the variance is an active field of research (i.e., inflation).

III.1.4.4 Variational data assimilation

An alternative solution to a stochastic computation of BLUE is a Variational (Var) approach. It estimates a single initial state (3DVar) or model trajectory (4DVar). 3D refers to the three dimensions of the physical space and 4D refers, in addition, to the temporal dimension that is solely included in the 4DVar. These Var techniques became popular as they allow direct assimilation of various observation types, e.g., conventional and remote sensing data. The formulation involves the minimization of a cost function J containing background J_b and observation J_o (i.e., the representation of observations in the model) errors. The cost functions are presented in Equation (III.1.12) for 3DVar and Equation (III.1.13) for 4DVar. The analysis x_a is the solution that minimizes the cost function $J(x)$.

$$J(x) = \overbrace{\frac{1}{2}\left(x - x_b\right)^T \mathbf{B}^{-1}\left(x - x_b\right)}^{J_b(x)} + \overbrace{\frac{1}{2}\Big[y_o - H(x)\Big]^T \mathbf{R}^{-1}\Big[y_o - H(x)\Big]}^{J_o(x)} \tag{III.1.12}$$

$$J(x) = \frac{1}{2}\left(x - x_b\right)^T \mathbf{B}^{-1}\left(x - x_b\right) + \frac{1}{2}\sum_k \Big[y_{k,o} - H_k(x_k)\Big]^T \mathbf{R_k}^{-1}\Big[y_{k,o} - H_k(x_k)\Big] \tag{III.1.13}$$

Where x_k results from the model operator applied from time step 0 to k.

The minimization of $J(x)$ requires the gradient of the cost function. Those gradients can usually not be calculated directly and are thus computed in a numerical iterative calculation. 3DVar can use for example an iterative descent method until the gradient is close enough to zero or the maximum number of iterations is reached. First, it computes J with an initial x. Then, the gradient of J is computed. It follows the descent and update of x for the subsequent iteration until x_a is obtained. The 4DVar time dependency requires the use of an adjoint model to compute the gradient of J. The adjoint model(s) can include non-linearities, non-differentiabilities, and iterative solvers (Blayo et al., 2011). 4DVar techniques utilize a backward integration of the model with the adjoint model during the assimilation time window. This process is computationally expensive. A comparison of 4DVar and 3DVar methods using different assimilation windows and both a high- and low-resolution model is provided by Rabier et al. (e.g., 1998). The study found in particular an advantage of 4DVar methods with 6- and 12-hour assimilation windows over the 3DVar method.

Figure III.1.1 illustrates the idea of 3DVar and 4DVar. The 3DVar technique corrects the background (X_b in the Figure) towards the observation (obs) to estimate the analysis (X_a in the Figure). The two components of the cost function (J_b and J_o in the Figure) are indicated. The 4DVar technique assimilates several observations within an *Assimilation window* on the time axis. The trajectory of the previous forecast is tuned towards the discrete observations and yields a new, corrected forecast. In practice, there is also an assimilation window in 3DVar which corresponds to the observation times around the analysis time for which corresponding observations are assimilated as if they were observed at the assimilation time.

These Var techniques use parameterized background and model error statistics. The matrix $\mathbf{B}$ is in general too large to be computed (dimension 10^6 or larger). In 3DVar, background covariances are inferred from a climatology. Therefore, 3DVar features a high computational efficiency. The background covariance matrix evolves with the flow in the 4DVar techniques. As mentioned, linearized and adjoint versions of the model are required to compute the solution x_a. 4DVar techniques are much more expensive computationally than the 3DVar techniques. Hence, 4DVar is still rarely used in convective scale operational NWP models (e.g., the 4DVar of the UK Met Office).

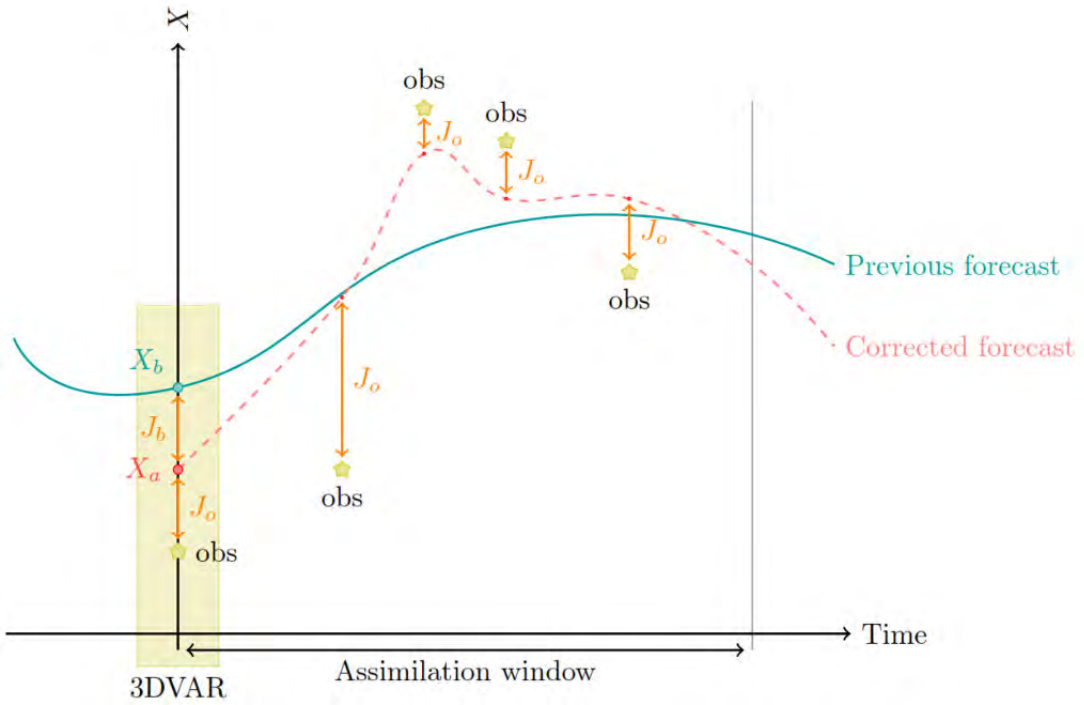


Figure III.1.1: Scheme of the 3DVar and 4DVar assimilation methods. From Blayo et al. (2011).

III.1.4.5 Ensemble variational (EnVar) and hybrid methods

Some assimilation techniques combine EnKF and Var in so-called Ensemble Variational (EnVar). They feature the benefit of a flow dependent background error covariance matrix **B** as the ensemble DA, and the robustness of variational methods for the assimilation. Var and EnKF scheme run in parallel and exchange information. For example, the ensemble mean should equal the Var analysis or the background error covariance in Var comes from the ensemble technique. Bannister (2017) and Desroziers et al. (2014) provide a detailed summary of EnVar techniques. There are further hybrid EnVar methods that do not only infer the **B** from the ensemble but from a combination of the ensemble and a climatology. At this point, some remarks of Bannister's paper are directly cited with adjustments (Bannister, 2017, p. 615):

- EnVar does not "... require a background error covariance model as Var does. This is important when modelling background errors in **B** involving processes that are too complicated, nonlinear or when geophysical balances are not relevant." For example, that is the case for storm-scale severe weather prediction as a subject of this thesis.
- The vector of the control variables has less elements for the EnVar than for the variational DA methods.
- Some EnVar techniques do not need linearized models and, thus, avoid the use of linear and adjoint models. This may result in a significant cost gain during the assimilation. Supplementary, linearized models are often difficult to derive (e.g., because of nonlinearities in physical parametrizations) and expensive to maintain.
- "The low-rank property of the implied background-error covariance matrix in EnVar means that sampling error problems will inevitably arise when N [the number of ensemble members] is small. This usually requires some kind of mitigation such as localization ... " This problem is inherited from the EnKF.

- Despite the advantages of the EnVar, it is still expensive to run (e.g., computing the ensemble) and not yet commonly used as operational data assimilation technique at the convective scale.

III.2 | Data Assimilation in AROME-France

In the previous chapter, the theoretic framework for assimilation and different assimilation techniques are provided. This chapter introduces the NWP model Applications de la Recherche l'Opérationnel à Méso-Echelle France (AROME-France). First, model characteristics are briefly explained and references for detailed information are provided. The second section details the assimilation system as implemented in AROME-France. It includes a description of the assimilation technique, the assimilated observations, and the determination of background errors during the assimilation.

III.2.1 The numerical model AROME-France

AROME-France is the convective-scale, operational, limited area model run by Météo-France since 2008 (Seity et al., 2011, Brousseau et al., 2016). It provides 36- to 42-hour forecasts five times a day (00 UTC, 03 UTC, 06 UTC, 12 UTC, 18 UTC). After an update in 2015, the model grid comprises 1440 grid points in the horizontal with uniform 1.3 km horizontal resolution. Figure III.2.1 illustrates the physical domain of AROME-France with the model topography. In the vertical, the lowest model level is situated at 5 m above ground. Each column reaches up to the highest level at 10 hPa. The vertical coordinate follows the terrain at the lowest levels and isobars at the highest ones. The vertical resolution is refined homogeneously from top to bottom by a factor of 1.5 for each layer. In total, 90 vertical levels (33 levels below 2000 m)are computed. Model time steps equal 50 s. Model dynamics are non-hydrostatic, semi-implicit, and semi-Lagrangian. AROME-France predicts twelve 3D prognostic variables: 2 components of the horizontal wind (U and V), temperature T, specific contents of water vapor q_v, and of 5 hydrometeors, turbulent kinetic energy TKE, and two non-hydrostatic variables, as well as the 2D surface pressure p (Seity et al., 2011).

AROME-France uses a mixed-phase microphysical scheme including riming processes and graupel (Seity et al., 2011). In particular, the ICE3 scheme is adapted from the research model Meso-NH (mesoscale non-hydrostatic, Lac et al., 2018) and upgraded by a condensation and sedimentation scheme. The microphysics scheme of AROME-France separates five prognostic hydrometeor variables; those are specific contents of precipitating species rain (q_r), snow (q_s), and graupel (q_g) and the two non-precipitating species ice crystals (q_i) and cloud droplets (q_c). Borderies (2018, Section 2.2.3) provides details about the computation of the different species and microphysical processes. In addition, the water vapor specific content q_v (also termed specific humidity) is computed. Hail is assumed to behave as large graupel particles. The size distribution of each hydrometeor species uses a generic gamma distribution. Overall, more than 25 processes are parameterized into the microphysics scheme (Lascaux et al., 2006).

Solid hydrometeors, i.e., ice crystals and graupel, are important contributors to the cloud

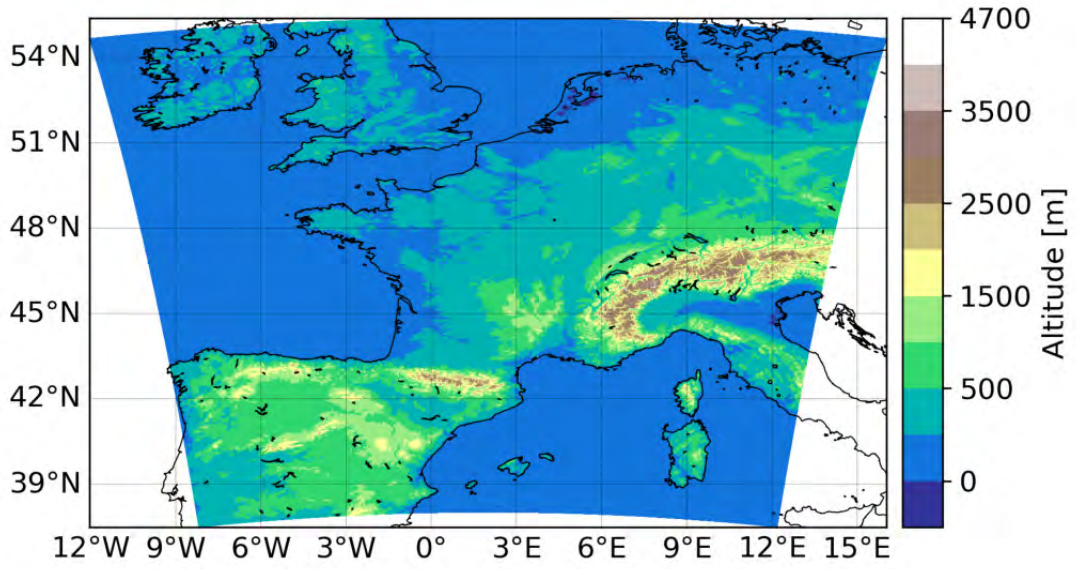


Figure III.2.1: AROME-France physical domain and model topography. The AROME-France grid is equidistant at 1.3 km resolution.

electrification. ICE3 distinguishes three ice types, (i) the pristine or primary ice phase, (2) the aggregation or snowflakes type as lightly rimed crystals or dry assemblages, and (iii) more heavily rimed crystals which are graupel, frozen drops, or hail. The third type could be split to distinguish graupel and hail (ICE4) but is currently considered as one broad class of heavily rimed particles in ICE3 (Bechtold et al., 2011). The scheme contains a prognostic equation for the primary ice mixing ratio, the snowflakes mixing ratio, and the rimed crystal mixing ratio. A diagnostic approach provides the number concentration of primary ice and then of snowflakes and of rimed ice crystals. The mass and terminal speed velocity of each ice type are estimated via power law relationships to the diameter, and, in AROME-France, the diameters are estimated as the gamma distributions. The details about the computation and the distribution can be found in Bechtold et al. (2011, p. 107-141). In general, primary ice has very slow terminal fall speed compared to the aggregates and rimed particles that are considered precipitating ice.

The cold processes initiating ice crystals comprise the heterogenous formation of ice embryos in a supersaturated environment over ice (deposition) and freezing of supercooled droplets subsequent to the attraction of aerosol particles (contact). A third process, the homogeneous freezing, kicks in at temperatures below −35 °C where small cloud droplets freeze immediately (Bechtold et al., 2011). It is also assumed that raindrops freeze immediately to graupel in such a cold environment. Deposition and sublimation cause then growing or depletion of the ice crystals depending on the environment. Aggregates form as a consequence of autoconversion of primary ice crystals. Raindrop contact freezing and heavy riming of snowflakes can create graupel. The aggregate and graupel can also grow through collecting small pristine ice crystals and by partial riming of cloud droplets and raindrops. The scheme includes the dry (i.e., collected ice crystals remain frozen and droplets and raindrops can freeze as the graupel surface temperature remains cold enough) and wet (i.e., a shallow liquid layer at the surface of the graupel has formed due to latent heat release and any excess of accreted liquid condensate is shed away) growth of graupel. The wet growth corresponds to the formation of hail (Bechtold et al., 2011). Shed liquid is converted into raindrops. The scheme compares the rates of dry and wet growth and only the prevailing process is activated at a time. While primary ice crystals melt directly at temperatures warmer than a critical value, snowflakes are transferred into the

graupel category and graupel melt to raindrops. It should be mentioned that most clouds are mixed-phase clouds, thus, also warm processes like condensation, evaporation, accretion, and autoconversion of cloud droplets are included in the microphysical scheme and influence indirectly the ice contents as the determine the content of liquid cloud water (see Bechtold et al., 2011).

AROME-France physics further include a 1-dimensional (1D) turbulence parameterization. It combines a prognostic TKE equation with a diagnostic mixing length. The parameterization is applied to compute the exchange coefficients for momentum, potential temperature, and humidity. An Externalized Surface (SURFEX) scheme separates surface interactions for tiles land, towns, sea and inland waters. It provides diagnoses of 2-m temperature, 2-m humidity, and 10-m wind at every time step. An inserted turbulence scheme for below the lowest model level (5 m in current AROME-France) predicts temperature, horizontal wind, specific humidity, and TKE. Details about AROME-France physics can be found in Seity et al. (2011). AROME-France uses the European Center for Medium-Range Weather Forecasts (ECMWF) radiation parameterization with computations of cloud optical properties and cloud cover, and a climatology of aerosol and ozone contents. Deep convection is expected to be resolved on the model grid (Fischer et al., 2018). Parameterization of sub-grid scale shallow convection is based on Pergaud et al. (2009). Lateral boundary conditions (LBCs) are extracted from the global model Action de Recherche Petite Echelle Grande Echelle (ARPEGE, Courtier et al., 1991). AROME-France features a Davies relaxation (Davies, 1976) coupling and ARPEGE synchronization. The initial conditions rely on a 3DVar data assimilation technique (following section).

III.2.2 The data assimilation system of AROME-France

AROME-France adapts the assimilation scheme of Aire Limitée Adaption dynamique Développment InterNational model (ALADIN, Fischer et al., 2005) for the AROME-France domain and resolution. A 3DVar technique (Section III.1.4.4) is used. Updates in the assimilation cycles are computed every hour. AROME-France employs 1-hour assimilation windows (Brousseau et al., 2016) meant to maximize the amount of assimilated observations. The high update frequency should partially balance the lack of the temporal dimension in the 3DVar system. Data assimilation techniques involving the temporal dimension, e.g., 4DVar or 4DEnVar, remain too expensive for AROME-France to date. The shortening of the assimilation cycle from originally three hours to one hour required the reduction of spin-up time, i.e., to reduce the effect of non-realistic effects from existing imbalances in the model during the first hours of the forecast. The full resolution background covariances reduce the model spin-up time (Brousseau et al., 2016).

Background error covariances $\mathbf{B}$ result from an offline AROME-France ensemble assimilation (six members) and are available as climatological training data at model resolution. The background error covariance matrix is calculated from multivariate formulation based on the control variables (vorticity, divergence, temperature, surface pressure, and specific humidity). In particular, the control variables do not include crucial variables related to deep convection and lightning such as hydrometeor specific contents and the vertical velocity. Hence, these crucial variables are left unchanged during the analysis. Standard deviations are uniform horizontally and vary in the vertical. Correlations vary with both horizontal grid point and altitude. These variations are horizontally homogeneous and isotropic (Brousseau et al., 2014) and cross-correlation are calculated via scale-dependent regression operators.

AROME-France long forecasts are launched 90 min after the analysis time to take advantage

of the hourly analysis of the assimilation cycle. The observation cut-off times range from 45 min to 105 min and observations within the time window until the cut-off are used for the hourly assimilation. As mentioned, control variables that are computed with every model analysis include the temperature, the specific humidity, the surface pressure, and the horizontal wind components (as vorticity and divergence). The following types of observation are currently assimilated in AROME-France:

- Automated ground-stations, ships, and buoys with data of surface pressure, 2 m-temperature, 10 m-wind, and humidity. Continuously available.
- Aeronautic airplane measurements of temperature and wind along the track.
- Global Positioning System (GPS) zenith tropospheric delay (ZTD) data (Yan et al., 2009, Mahfouf et al., 2015) with continuous information of column integrated water vapor (IWV).
- Radio soundings with vertical profiles of pressure, temperature, wind, and humidity. Only 2 to 4 times per day (or irregular timing) in few locations.
- Ground radar data of Doppler radar velocity (since 2008) (Montmerle and Faccani, 2009) and reflectivity (since 2010) (Caumont et al., 2010, Wattrelot et al., 2014). Continuously available over land and within radar range (about 200 km) at the coasts.
- Brightness temperature from satellite instruments Advanced Tiros Operational vertical Sounder(ATOVS), Infrared Atmospheric Sounding Interferometer (IASI), the Atmospheric Infrared Sounder (AIRS), the Cross-track Infrared Sounder (CrIS), the Advanced Technology Microwave Sounder (ATMS), the Special Sensor Microwave Imager Sounder (SSMIS), the Microwave Humidity Sounder (MHS), and Global Precipitation Measurement (GPM) Microwave Imager (GMI) in LEO, and the Spinning Enhanced Visible and infrared Imager (SEVIRI) on Meteosat Second Generation (MSG) in GEO
- Wind measurements from atmospheric motion vectors (AMVs, e.g., Brousseau et al., 2014, Velden et al., 2017) in MSG images and scatterometers. Continuously available.

Observations from fixed instruments are only considered at the time closest to the analysis time to avoid smoothing (i.e., through averaging) the information of particular observations. Observations from a moving platform like an aircraft or a LEO satellite exist at different locations during the assimilation window. Therefore, all such observations can be assimilated until the observation cut-off time (Brousseau et al., 2016).

III.3 | 1DBay and 3DVar Assimilation Method

Lightning observation is related to the AROME-France control variables in a non-linear, indirect manner. The lightning data assimilation (LDA, Chapter III.4) needs a complex strategy. This thesis adapts the 1D+3DVar assimilation method developed by Caumont et al. (2010) and implemented by Borderies et al. (2019) for radar reflectivity data. First, FED observations are inverted to profiles of relative humidity (RH) through a 1D Bayesian (1DBay) approach. The so-called pseudo-observations of created RH profiles are in turn assimilated in the 3DVar system of AROME-France as sounding data. This section details the concept of the 1DBay retrieval and briefly explains advantages of the assimilation of FED in AROME-France. Figure III.3.1 shows the flow chart of the entire assimilation scheme developed in this thesis. The included processes are detailed in the sections of this chapter.

III.3.1 Interlude: The 1DBay+3DVar method for LDA

At this point a short discussion about the application of the Bayesian retrieval method for the assimilation of lightning, i.e., FED, data is provided. The 1DBay has been tested for microwave and radar reflectivity retrievals of humidity (e.g., Olson et al., 1996, Kummerow et al., 2001, Caumont et al., 2010, Borderies et al., 2019).

Despite the established radar reflectivity assimilation in the operational AROME-France (Wattrelot et al., 2014), the implementation for FED data might be not as straightforward. FED is a 2D variable and 3D data should be retrieved for the assimilation. Furthermore, FED data are different from both microwave and radar reflectivity in that there are discrete values and that there is no threshold for a cloud. Although a positive FED always means that a cloud is present, an FED equal to zero does not mean that there is no cloud, but only no lightning. This fact may lead to difficulties when retrieving humidity profiles as locations with FED equal to zero might match both humid and dry regions. The LDA 1DBay retrieval uses the background graupel mass and RH to retrieve RH pseudo-observations (POs). The method is still more promising than nudging the FED-inferred graupel mass directly since previous studies showed that the graupel falls out quickly without any persisting effect on the forecast (Ge et al., 2013). In addition, AROME-France does currently not allow the assimilation of hydrometeor contents as they are not included in the control variables. It would be needed to extend the set of control variables and thus modify the model itself. However, former studies showed that adjusting relative humidity (RH) can create convection in accordance with observations (e.g., Ducrocq et al., 2002). RH can be assimilated in the AROME-France 3DVar as the specific humidity and temperature are in the control variables.

The objective here is to retrieve the best estimation of the RH profile based on the predicted multi-variate model columns of the background state. Indeed, the expected RH profile allows

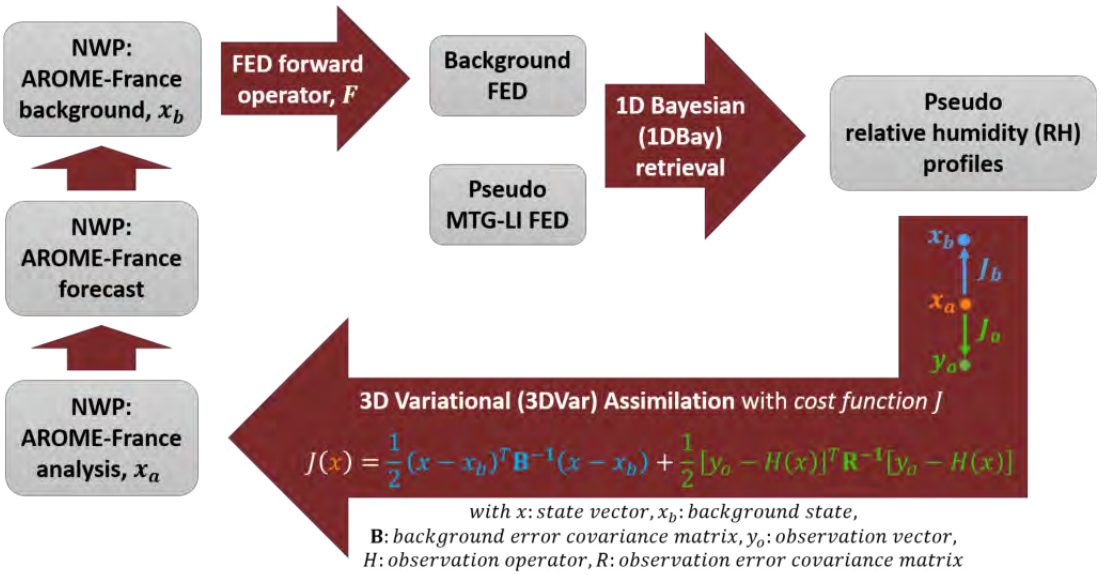


Figure III.3.1: Processes and steps included in the lightning data assimilation scheme developed in this thesis.

both to increase and to reduce the humidity of the background. Therefore, it is expected that convection can be triggered and spurious convection can be suppressed according to the observation. However, the use of the model background information in both the Bayesian retrieval of POs and the assimilation may cause problems that are yet unknown. This subject may be studied to assess the validity of the 1DBay method in a general sense, e.g., for various data types.

III.3.2 1DBay retrieval of pseudo-RH profiles

The 1DBay retrieval method is equivalent to a best estimate approach based on Bayes' theorem. The 1DBay problem can be solved in a similar way to 3DVar, except that only the vertical dimension exists. Here, it is assumed that the model is capable of producing analyses profiles that approximate the observation related profiles within a given region of the observation. Under this assumption, the database for the 1DBay approach includes the observation vector and the set of model forecasts that are situated in the vicinity of each observation. *Vicinity* is defined as a domain that includes a sufficient number of grid points to estimate the pseudo-observations, however, small enough to still represent the meteorological conditions at the location of the observation. Borderies et al. (2018) proposes a square with sides of 160 km centered at the observation for the assimilation of radar data in AROME-France. The same vicinity area is initially used here for the assimilation of FED in AROME-France. Since however, the FED pixels (side 7 km) are significantly larger than the radar reflectivity pixels (side 1 km) and thus a lower number of values is available, the vicinity can be adapted to the specifications of the FED data as described in Chapter III.4. The associated model variables are expected to be always consistent with the meteorological conditions at the observation time. The theoretic framework for the 1DBay approach is provided in the following.

The best estimation of the RH profile x^{RH}, $y_{po}^{\mathrm{RH}} = E(x^{\mathrm{RH}})$ is given by (Kummerow et al., 2001)

$$y_{\mathrm{po}}^{\mathrm{RH}} = E\left(x^{\mathrm{RH}}\right) = \int x^{\mathrm{RH}} \mathrm{PDF}\left(x^{\mathrm{RH}}\right) dx^{\mathrm{RH}} \tag{III.3.1}$$

where the probability density function (PDF) is proportional to the conditional probability that x^{RH} represents the true RH profile x_t^{RH} given that y equals the observation vector y_o

$$\mathrm{PDF}(x^{\mathrm{RH}})dx^{\mathrm{RH}} \propto P\left(x^{\mathrm{RH}} = x_t^{\mathrm{RH}} | y = y_o\right) \tag{III.3.2}$$

Applying Bayes' Theorem, the conditional probability can be expressed as

$$P\left(x^{\mathrm{RH}} = x_t^{\mathrm{RH}} | y = y_o\right) = \frac{P\left(y = y_o | x^{\mathrm{RH}} = x_t^{\mathrm{RH}}\right) P\left(x^{\mathrm{RH}} = x_t^{\mathrm{RH}}\right)}{P\left(y = y_o\right)} \tag{III.3.3}$$

It is further assumed that the errors of the observations and the simulated observations $H(x)$, with the observation operator H, are Gaussian, with zero mean, and uncorrelated. Then

$$P\left(y = y_o | x^{\mathrm{RH}} = x_t^{\mathrm{RH}}\right) \propto \exp\left\{-\frac{1}{2}\Big[y_o - H(x)\Big]^T (\mathbf{O} + \mathbf{S})^{-1}\Big[y_o - H(x)\Big]\right\} \tag{III.3.4}$$

with the observation and simulated observation error covariance matrices $\mathbf{O}$ and $\mathbf{S}$, respectively.

At this point, it is assumed that the natural PDF of RH is described by the RH of model columns in the vicinity of the observation (Olson et al., 1996). With the assumption that the pseudo-observations can be retrieved from a discrete number i, j of model grid points in the vicinity of the observation, the final expression for $y_{\mathrm{po}}^{\mathrm{RH}}$ yields

$$y_{\mathrm{po}}^{\mathrm{RH}} = \sum_i x_i^{\mathrm{RH}} \frac{W_i}{\sum_j W_j} \tag{III.3.5}$$

with the weights

$$W_i = \exp\left\{-\frac{1}{2}\Big[y_o - H(x_i)\Big]^T \mathbf{R}^{-1}\Big[y_o - H(x_i)\Big]\right\} \tag{III.3.6}$$

where $\mathbf{R}$ is the matrix of combined observation and simulated observation error covariances. It is chosen as a diagonal matrix so that $tr(\mathbf{R}) = n\sigma_o^2$, where n is the number of vertical levels in the observation and σ_o is the standard deviation of observation and observation operator errors. The weights are then obtained from

$$W_i = \exp\left\{-\frac{1}{2}\frac{[y_o - H(x_i)]^2}{n\sigma_o^2}\right\} \tag{III.3.7}$$

In the specific case of the assimilation of FED, n equals 1 as there is only one observation level in the vertical. The σ_o is inferred from a sensitivity study where the 1DBay method is applied to create FED pseudo-observations (see Section III.4.6 for details). The optimum is received for a value of 2 dB[1]. The observation operator H as FED forward operator F applies a simple linear regression as defined in Section III.4.5 (a more sophisticated operator F is tried, see Section III.4.12). $F(x)$ defines then the simulated FED.

The 1DBay method can fail to produce pseudo-observations if all the W_i approximate zero meaning that no simulated FED is close enough to the observed one. This is the case if (i) the observed FED is greater than zero and all simulated FED in the vicinity equal zero, or (ii) the observed FED equals zero and the simulated FED is positive for all points in the vicinity. The former case is treated using a humidity adjustment (HA). At the convective scale, the presence of lightning should correspond to a model column featuring a cloud. The humidity profile is

[1]FED is used in units of dB.

assumed to be saturated within the cloud. The HA saturates the column of pseudo-RH in a layer between the lifted condensation level and a constant altitude of 13 km (as artificial cloud top). Saturation means RH is set to 100 % if the background RH is less than 100 %. This idea is similar to among others Fierro et al. (2019). Case (ii) occurs if the model simulates lightning in a region where no lightning is observed. The related spurious convection should be removed from the model. However, as it appears to be unlikely that the background FED (calculated from model graupel mass, see Section III.4.5) indicates lightning at all grid points within the vicinity area if no lightning was observed, this case has never occurred so far. It has not been included in the retrieval yet.

The HA is also performed if all the FED values from the background are at least one order of magnitude smaller than the observed FED in the center of the corresponding vicinity area. In this case, the sum of the weights W_i can yield a value greater than zero and the retrieval can estimate a profile, however, this profile would be much drier than the expected profile considering the FED observation. The HA can address this issue and add humidity to the model background.

Finally, no pseudo-observation is retrieved if both observed FED and the closest (geolocation distance) background FED equal zero. In that case, there is no useful information to identify the expected profile and to use the 1DBay retrieval.

III.3.3 3DVar assimilation of pseudo-RH profiles

The pseudo-RH profiles from the 1DBay retrieval are assimilated as sounding data in the 3DVar assimilation system of AROME-France (Section III.2.2). Instead of repeating the technical details of the 3DVar assimilation, a brief reasoning is provided about assimilating FED besides the currently assimilated observations. Despite all the different types of observations currently assimilated, only radar data are relevant on horizontal length scales below 200 km (Brousseau et al., 2014). Thus, forecasts of local convection do not benefit directly from the vast majority of assimilated data. The locations of convective cells in the initial state are only provided by the radar reflectivity data. This thesis adds the GEO lightning data assimilation (Chapter III.4) as new source of information for the locations of convective cells. It aims at improving in particular the prediction of convective clouds and related phenomena such as precipitation and winds. GEO lightning data can indicate the location of convection and thunderstorms in regions that are barely covered by weather radars, e.g., over mountains and over the sea. A specific example for France features storms that arrive from Spain. They can be masked for the weather radars as the radar beam is blocked by the Pyrenees mountain range. Also, storms arriving from the Atlantic might be out of range of the weather radars at the time of assimilation, while they would be already visible for the GEO satellite at this time.

III.4 | Lightning Data Assimilation (LDA) in AROME-France – Proof of Concept

The final chapter of this thesis applies the methods developed until this point, as illustrated in Figure III.4.1. All requirements are met to simulate pseudo MTG-LI data. Chapter II.2 has proven that Meteorage and NLDN do not only utilize the same Vaisala sensors but also observe lightning in a statistically similar manner. It was concluded that they constitute two similar LF networks in different parts of the world. Our GEO lightning pseudo-observation generator that was originally designed using NLDN and GLM data (Section II.4.2) can thus use Meteorage observations as input data. Doing so, the generator mimics MTG-LI observations and FED over France. This chapter describes the assimilation of the pseudo MTG-LI data in AROME-France. The following sections will eventually be published as a paper. Therefore, some important concepts that were detailed in the previous chapters are repeated in a concise fashion.

III.4.1 Introduction

Convective weather phenomena such as thunderstorms threaten the society by producing severe weather and related impacts, e.g., flash floods, large hail, tornadoes, and strong winds. Cloud electrification and subsequent lightning discharges are caused by interactions of different ice particles inside convective clouds. The process makes lightning an effective tracer of deep convection. A new type of lightning locating systems (LLSs) on geostationary (GEO) satellites allows for continuous large-scale monitoring of lightning activity. The Geostationary Lightning Mappers (GLMs) on the GOES-R series satellites cover the Americas and the adjacent oceans (Goodman et al., 2013). A similar instrument, the Meteosat Third Generation (MTG) Lightning Imager (LI), will be launched in the 2022 time-frame to monitor lightning over among others Europe, Africa, and wide parts of the Atlantic Ocean. The advantage of GEO lightning sensors over existing observations of convective clouds, e.g., by radars, is the availability of GEO data also in data-sparse regions with limited access, such as over oceans and mountainous terrain. GEO lightning data may become important for numerical weather prediction (NWP), potentially improving the initial state of the model and the accuracy of the predicted storm location, timing, and intensity. The lightning data assimilation (LDA) addresses this objective.

The specific objective of this work is to further improve the convective-scale prediction (<3 km horizontal resolution) of thunderstorm forecasts in terms of location, timing, and structure of the events. This also includes related precipitation and wind fields, especially maxima. A metric taking displacement errors into account i.e., the Fractions Skill Score (FSS, Roberts and Lean, 2008), is analyzed to compare different assimilation approaches. Those include the

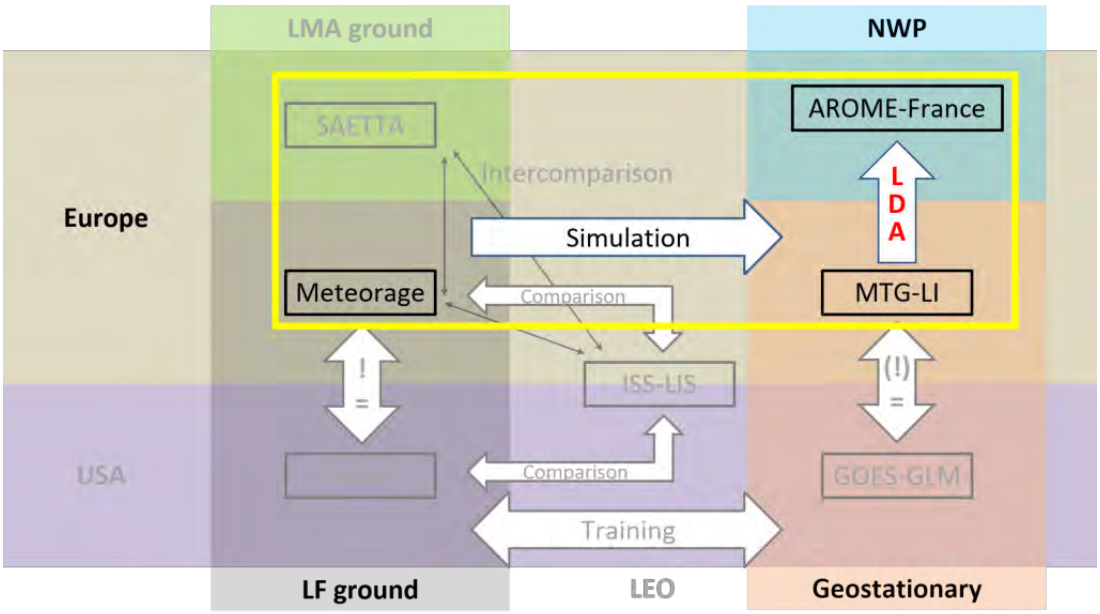


Figure III.4.1: Flow chart of work phases during this thesis as Figure 1. The LLS and phases detailed in this chapter are highlighted in the yellow frame, while other phases are shaded.

presence and absence of radar reflectivity, Doppler wind speeds, and FED in the set of assimilated observations. Pseudo-observations of the MTG-LI are generated to form the lightning data base (Section III.4.4). Then, the pseudo-observations are assimilated in the regional, operational, convective-scale model of Météo-France . The benefits and problems of the GEO lightning data assimilation scheme that is developed here are assessed in the final section.

The following section describes work to date conducted in the context of LDA. The NWP model configuration and lightning data as used for this work are briefly explained thereafter (Sections III.4.3 and III.4.4, respectively). Section III.4.5 reviews lightning observation operators and introduces the lightning observation operator developed during this thesis. Then, our LDA method and the model experiments are explained in Sections III.4.6 and III.4.7. Sections III.4.8 and III.4.9 describe the AROME-France analysis resulting from LDA. One case study is presented using different assimilation experiments, among others the new LDA (Section III.4.10), followed by the conclusions.

III.4.2 LDA review

Lightning data assimilation (LDA) can be challenging since lightning is not resolved in most NWP models, and does not directly involve any control variable to be compared to the observations and adjusted in the analysis. Other difficulties of LDA include: lightning is an ephemeral phenomenon compared to the life cycle of a thunderstorm, time-space shifts between observations and the background which are usual at the storm scale challenge up-to-date data assimilation systems (zero-spread, zero-gradient problems), background errors are not Gaussian, and lightning is more related to lower-impact variables such as hydrometeor content and vertical velocity than to thermodynamic variables. Despite the difficulties, former studies have shown the benefits of LDA for forecasts of convection and related phenomena.

Papadopoulos et al. (2005) were among the first to use lightning intensity without prior rain rate (RR) retrievals. Their LDA nudged model generated humidity profiles toward empirical

profiles as a function of observed CG lightning intensity. Thus, the activation of convection in the convective parametrization scheme (CPS) was controlled. The LDA resulted in improved convective RR forecasts. A similar LDA technique was used by Lagouvardos et al. (2013) to improve the forecast of precipitation and locations of precipitation field maxima of a heavy precipitation event over southern France. Giannaros et al. (2016) developed and evaluated an LDA method that controls the triggering of the CPS in the Weather Research and Forecasting (WRF) model. For eight cases over Greece, they found the LDA led to higher precipitation forecast scores, especially for high precipitation values, and better spatial agreement between forecast and observation of precipitation. Mansell et al. (2007) directly related lightning to the existence of deep convection. The lightning data included NLDN CG and LMA total lightning observations and was used as lightning rate per grid cell. The lightning rate controlled the (de)activation of the CPS in their study, while the model physics controlled other model variables following the forced or depressed convection. The nudging added small amounts of moisture in the parcel source layer when the CPS was too weak to produce convection where lightning was observed. The, the increase in buoyancy and thermal instability forced lightning-consistent updraft strengths and cloud heights above the freezing layer. In regions without lightning, convection could be suppressed by, e.g., weakening updrafts through increasing of entrainment or inhibiting the CPS. The LDA successfully generated observed cold pools at the time of initialization. Pessi and Businger (2009) nudged the model's latent heating rates according to a relationship between observed lightning and convective rainfall to force convection production by the CPS of their model. The method improved pressure and wind forecasts of a rapidly deepening extratropical cyclone. These studies used NWP models with parametrized convection.

At a convection-permitting scale, a nudging technique was also used by Fierro et al. (2012), who increased the water vapor mixing ratio in the 0 °C to −20 °C layer where lightning was observed and the relative humidity (RH) of the background was less than 81 %. They also introduced a smooth function to calculate water vapor mixing ratio (q_v) from flash rate and simulated graupel mass. Precipitating convection was better correlated with observed reflectivity fields for the LDA than for the control experiment. The method was further tested by Fierro et al. (2014). They increased the water vapor mass proportional to 10-minute FED and inversely proportional to model graupel mixing ratio. Results showed similar improvements on a derecho forecast as 3DVar radar data assimilation (RDA). Lynn et al. (2015) refined the nudging technique of Fierro et al. by holding the temperature constant and gradually increasing the water vapor mixing ratio to increase the virtual temperature perturbation and favor static instability that can lead to convective clouds. They also introduced an extension to suppress spurious convection. A different nudging approach was used by Marchand and Fuelberg (2014). They enhanced convection at locations with lightning observations (as gridded flash rates) and no modeled convection by warming of the source layer to temperatures associated with the cumulus condensation level (CCL, from the most unstable level per grid cell). According to their discussion on enhanced instability through modification of the source layer, increasing latent heat does not work at convective scales, while increasing the humidity may lead to a moist bias in the model and reduced CAPE in the mid- and upper troposphere. Lightning data were used by Wang et al. (2014) to generate pseudo radar reflectivity which was subsequently assimilated using the physical initialization method to adjust model variables. In a case study, this LDA method improved the forecast for about 3 hours. Qie et al. (2014) constructed empirical formulas between total lightning and graupel, ice, and snow mixing ratios based on thunderstorms over northern China. The nudging function added to the WSM6 microphysical scheme of WRF adjusted the mixing ratio of the ice-phase particles between the 0 °C and −20 °C isotherms.

Forecast of a mesoscale convective system (MCS) showed precipitation center, amount, and coverage closer to observation when using their nudging LDA method, with best results on the short term (1 hour). Dixon et al. (2016) used the well-established nudging of water vapor mixing ratio toward saturation. However, they increased q_v in a region within 10 km of lightning activity without taking flash rates into account. The LDA was performed within the first 3 hours of the forecast, and subsequent times were evaluated. Emphasis was put on the LDA comparison for both deterministic and ensemble forecasts. The general improvement of reflectivity forecasts was stronger for the deterministic than ensemble forecasts. A LINET total lightning as strokes per grid cell LDA was deployed by Stefano et al. (2017) using water vapor substitution. They applied the relationship given by Fierro et al. (2012) to get water vapor mixing ratio from lightning rate and simulated graupel mixing ratio. Wang et al. (2017a) retrieved graupel mixing ratio q_g fields from total lightning and empirical q_g profiles. They used a linear relationship between column integrated graupel mass m_g and the total lightning rate. Their nudging-based assimilation of retrieved q_g fields and related latent heat release due to the formation of the retrieved graupel promotes and sustains convection at observed lightning locations (conceptionally similar to Fierro et al., 2012). Improvements were found for forecasts of convection-related cloud parameters, lightning, and precipitation especially in the first 2 hours. Wang et al. (2018) used time-lagged ensemble forecasts to retrieve q_g from lightning rates by utilizing empirical vertical profiles from previous forecast cycles. Increments of model state variables are nudged by tuning the prognostic equations during the LDA. Their LDA method could recover observed convective cells, suppress spurious convection, and correct the geolocations of convective systems. Forecast skills were improved in convective regions using the LDA, with less impact on stratiform regions.

Adapting their previous nudging method, Fierro et al. (2016) compared a 3DVar LDA with RDA. Pseudo-q_v profiles were created through saturation of modeled q_v-profiles between the lifted condensation level (LCL) and 15 km if 10-minute FED exceeded 50 flashes per pixel. They found that the RDA yields better forecasts of convective cells during the first 30 min of the forecast, while the LDA gives better storm structures at about one-hour forecast. The combination of both radar and lightning observation provides the highest forecast skill. All these studies used ground-based LLSs. Another 3DVar LDA was applied by Wang et al. (2017b). They used the relationship given by Fierro et al. (2012) to retrieve water vapor mixing ratio q_v from total flash rates with 9 km gridded resolution. The q_v is then transformed to RH in the form of sounding data and assimilated as such. Improvement in reflectivity and precipitation forecasts were reported along with improved model profiles of temperature, dew point, and RH after 7 hours of forecast. They noted that the method still needs improvement to suppress spurious convection.

In preparation of a hybrid variational-ensemble LDA technique for GLM, Apodaca et al. (2014) assimilated World Wide Lightning Location Network (WWLLN) data in WRF. LDA had a positive impact on model variables during several assimilation cycles, but no clear improvement was found for a 6-hour forecast. However, the use of WWLLN as a GLM proxy is critical as WWLLN mainly detects CG flashes, whereas GLM detects total lightning.

Recent studies have been exploring the use of (pseudo-)GLM data. Mansell (2014) used synthetic GLM total lightning represented as flash extent density (FED) in an Ensemble Kalman Filter (EnKF) assimilation with positive effects on predicted spurious convection, updraft locations, and reflectivity. The idea was also tested by Allen et al. (2016) with success to improve reflectivity forecast of a multicell and a supercell case. Fierro et al. (2019) tested their 3DVar assimilation technique developed for ground-based data with real GLM total lightning observations. LDA of flash origin density was also compared with RDA again. The LDA method

adjusts water vapor mass mixing ratio within a fixed layer depth above the LCL by assuming nearly water-saturated conditions at observed lightning locations. It adds q_v in regions of lightning regardless of flash rate by setting RH to 95 % in a layer of 3 km above the LCL if the background RH is less than 95 %. One new model experiment balanced the total water vapor mass added by the LDA by an equal removal outside observed lightning locations. Therefore, the q_v mass added by the LDA is divided by the number of grid points without lightning, weighted by the air density value at each grid point, and then subtracted from the background q_v at each grid point. This experiment ensured q_v mass balance in the model, while only low values of q_v were subtracted per grid point (as usually many points without lightning) with negligible effect on the simulation apart from lightning locations. Both LDA and RDA improved the short-term accumulated precipitation and radar reflectivity composite. However, the LDA scheme of Fierro et al. (2019) is not able to suppress spurious convection.

Hu et al. (2020b) adopted the technique of Fierro in their cloud-scale model with 1.5 km horizontal resolution and found, as also previously reported, a wet bias in the model that increased with the forecast time. The method still misses a suppression of spurious convection in regions without lightning. They also conducted a layer depth sensitivity study with similar results for adding q_v in layers of 2 km to 10 km depth. Other sensitivity tests revealed the best setup with 10 min of lightning rate accumulation up to the analysis time (no future observations of lightning), assimilation of GLM-derived q_v every 15 min, and a horizontal decorrelation length scale of 3 km. Kong et al. (2020) present an LDA of real GLM data in an EnKF framework. FED at 10 km pixel resolution is assimilated using both graupel mass and graupel volume-based observation operators. Both LDA could produce model-consistent analyses of the state variables and improved the forecast especially for regions of deep moist convection in a MCS environment.

Liu et al. (2020) assimilated Chinese satellite Fengyun-4 (FY4) lightning and cloud top data. Pseudo-RH profiles were created by setting the RH to 90 % where lightning was observed and the background RH was less than 90 %. The cloud top heights were used to determine the maximum height of humidity adjustment in pseudo-RH profiles, with LCL as base of the RH adjustment. Another approach used layers between the LCL and a fixed height of 15 km, and a third approach adjusted the humidity between the 0 °C and −20 °C isotherms. The two methods using the LCL improve the equitable threat score for precipitation forecasts the most, however, the approach with fixed top height produced spurious convection in one case (not seen for the approach with cloud top height). Chen et al. (2020) also assimilated FY4 lightning data, but they retrieved maximum proxy-reflectivity from the lightning data via a logarithmic relationship. Maximum proxy-reflectivity is then extended to 3D pseudo-reflectivity profiles based on observed radar reflectivity profiles, and eventually assimilated. The cycling LDA could adjust the humidity fields, and indirectly alter the temperature, pressure, and wind fields. 6-h FSSs of precipitation improved similar for the LDA experiments and additional RDA experiments.

Early LDA was conducted using nudging techniques. Nowadays, models usually use Var, KF, or EnVar methods. Many studies agree in the modification of the RH field of the model, whereby humidity is often added in regions of lightning observations. Most studies report improvements especially in regions of deep convection and for the location and estimation of precipitation maxima. GLM lightning observations, e.g., as FED fields, were assimilated and the model experiments showed similar skill as RDA experiments. Recent Var LDAs using GLM data, e.g., Fierro et al. (2019) and Hu et al. (2020b) successfully promote convection, but cannot suppress spurious convection. This capability is the subject of current research.

III.4.3 Model configuration

AROME-France is the convective-scale, limited-area model run operationally by Météo-France since 2008 (Seity et al., 2011). It provides 36 to 42-hour forecasts five times a day (00 UTC, 03 UTC, 06 UTC, 12 UTC, 18 UTC). After an update in 2015, the model grid comprises 1440 × 1536 grid points in the horizontal with uniform 1.3 km horizontal resolution. The physical model domain and model topography are shown in Figure III.2.1. In the vertical, the lowest model level is situated at 5 m above ground. Each column reaches up to the highest level at 10 hPa. The vertical resolution is refined homogeneously from top to bottom by a factor of 1.5. In total, 90 vertical levels (33 levels below 2000 m) are computed. Model time steps equal 50 s. Model dynamics are non-hydrostatic, semi-implicit, and semi-Lagrangian. Lateral boundary conditions (LBCs) are extracted from the global model ARPEGE. AROME-France features a Davies relaxation (Davies, 1976) coupling and ARPEGE synchronization. The initial conditions rely on a 3DVar data assimilation technique (following section).

Deep convection is expected to be mostly resolved on the model grid (Fischer et al., 2018). parametrization of sub-grid scale shallow convection is based on Pergaud et al. (2009). AROME-France uses a mixed-phase microphysical scheme with riming processes and graupel (Seity et al., 2011). In particular, the microphysics scheme of AROME-France separates five prognostic hydrometeor variables, that are specific content of precipitating species rain (q_r), snow (q_s), and graupel (q_g) and the two non-precipitating species ice crystals (q_i) and cloud droplets (q_c). In addition, the water vapor specific content q_v (also termed specific humidity) is computed. Hail is assumed to behave as large graupel particles. Overall, more than 25 processes are parametrized into the microphysics scheme (Lascaux et al., 2006). AROME-France physics include a 1-dimensional (1D) turbulence parametrization as combination of a prognostic turbulent kinetic energy (TKE) equation with a diagnostic mixing length. An Externalized Surface (SURFEX) scheme and the European Center for Medium-Range Weather Forecasts (ECMWF) radiation parametrization are other components of AROME-France model physics. Details can be found in Seity et al. (2011) and Brousseau et al. (2016).

III.4.4 Lightning data

This work adapts the GEO lightning pseudo-observation generator as developed by Erdmann et al. (2020b) (see Section II.4.2). The low frequency (LF) ground-based lightning observations are provided by the French network Meteorage (Schulz et al., 2016, Pédeboy, 2015) as input. Meteorage locates total lightning with a separation of CG strokes and IC pulses. They are clustered to the flash level data using the same method as Erdmann et al. (2020a), i.e., a spatiotemporal clustering with criteria of 20 km and 0.4 s. A statistical analysis of both Meteorage and NLDN observations with ISS-LIS data as common reference (Chapter II.2) was performed to validate Meteorage as a suitable source for the input to the GEO lightning pseudo-observation generator. Pseudo-observations of MTG-LI are then generated on a regular latitude-longitude grid with average pixel resolution of 7 km which approximates the expected MTG-LI resolution over France (personal communication Bartolomeo Viticchie, 2020). It is straightforward to calculate flash extent densities (FEDs) on the same regular grid. FEDs are calculated in short 5-minute intervals that can then be added as needed for the assimilation. This work uses a short 10-minute interval of FED data in the LDA that is centered at the time of the analysis (see, e.g., Fierro et al., 2016, Hu et al., 2020b). The short period around the analysis reduces displacement errors in the analysis. The domain is limited to 40°N to

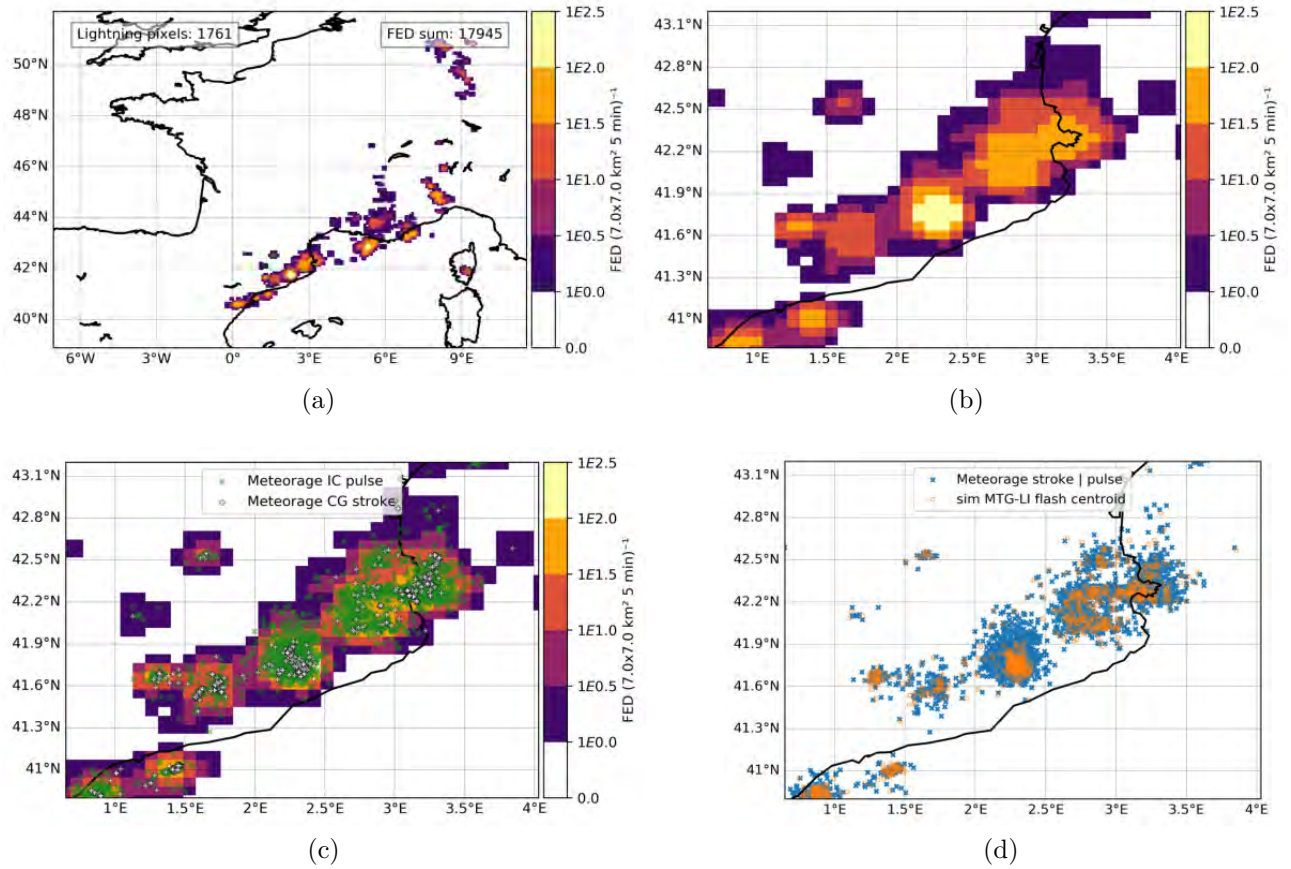


Figure III.4.2: Simulated pseudo MTG-LI FED of the entire domain (a) and zoomed on the region of the maximum FED value (b) on 7 km × 7 km grid. The Meteorage strokes and pulses that were used to generate the pseudo MTG-LI flashes are added to the FED in (c). Pseudo MTG-LI flash centroids and the corresponding Meteorage strokes+pulses in (d). Example for the period from 09 Aug. 2018, 13:55 UTC to 14:00 UTC.

51°N and 5.5°W to 10°E, which is inside the AROME-France physical model domain. The pseudo MTG-LI FED is referred to as FED observation hereafter to avoid confusion with the pseudo-observations created by the 1DBay retrieval.

An example of simulated MTG-LI is provided in Figure III.4.2(a) for the entire study domain, and Figure III.4.2(b) zoomed on the FED maximum at 09 Aug. 2018, 13:55 UTC to 14:00 UTC. The regular latitude-longitude grid of the FED is obvious. The example accumulates data of 5 min, and the FED pixels have a size of about 0.0630° latitude and 0.0896° longitude that corresponds to a pixel size of 7 km times 7 km in the center of the domain between 40°N to 51°N and 5.5°W to 10°E. Whereas the pixel size in longitudinal direction remains constant, it increases southward for the latitudinal direction. Hence, the FED pixel size has an exact range from 6.287 km times 7.000 km to 7.648 km times 7.000 km. For simplicity, the latitudinal length of one pixel is considered as being constant at 7 km from here on. The reader should remember, however, that this is not exactly true.

Figure III.4.2(c) shows the zoomed-in simulated FED and adds the Meteorage CG strokes and NLDN pulses that were used as input for our GEO lightning pseudo-observation generator. The vast majority of strokes and pulses lies within the area of positive FED. Figure III.4.2(d) MTG-LI illustrates that the simulated MTG-LI flash centroids are situated within the corresponding Meteorage strokes+pulses (combined here).

III.4.5 Lightning observation operator

Lightning is not a prognostic variable in most operational NWP models. An observation operator is needed to relate the model variables to the electric field or lightning activity and to find the most suitable representation of lightning for LDA applications. For example, Fierro et al. (2012) put forward a widely used lightning proxy calculating water vapor mixing ratio from simulated graupel mass and observed flash rates. The lightning observation operator draws the connection between the model prognostic variables and the lightning activity, e.g., GLM FED. The relationship can be linear or non-linear. Mansell (2014) used both the simulated flash rate and a linear relationship between total lightning and graupel volume as observation operators. Simulation with both types sharpened the location of reflectivity echoes and the spatial location probability of convective updrafts. LDA could also limit spurious deep convection. Allen et al. (2016) tested observation operators in the sense of a linear best fit operator. They found that graupel mass and graupel volume yield better results than an operator based on the non-inductive charging rate. Deierling et al. (2008) also found a strong correlation between lightning activity and ice mass fluxes. Deierling and Petersen (2008) showed a robust relationship between total lightning rate and updraft characteristics, especially the updraft volume. Buiat et al. (2017) analyzed CloudSat Cloud Profiling Radar (CPR) to relate cloud ice water content (IWC) and effective radius (ER) to LINET (Betz et al., 2009) lightning strokes. They found that high IWC and high ER are favorable for CG strokes. Graupel contributes to both high IWC and high ER. Strong on-going updraughts can cause high IWC. Hence, the findings agree with the studies of Deierling et al. (2008) and Deierling and Petersen (2008).

Another approach for a lightning proxy relates the cloud top height to the lightning activity. The Price and Rind lightning parametrization is one example (Price and Rind, 1992). Price and Rind (1993) suggest that the IC/CG ratio of flashes is related to the thickness of the cold cloud region (0 °C to cloud top) rather than only the height of the freezing level. Wong et al. (2013) describe some issues using the Price and Rind lightning parametrization especially for the simulation of IC/CG ratio. A technical problem was the use of cloud top from convective parametrization in their study. Based on 10 single-day case studies in Greece, Giannaros et al. (2015) suggest that the Price and Rind lightning parametrization in WRF needs model-related variables related to the microphysics and thermodynamics to prevent lightning overprediction. In fact, combining ice content, the maximum vertical velocity, and the convective available potential energy helped the most to correct the lightning forecasts. Karagiannidis et al. (2019) tested alternatives to the Price and Rind lightning parametrization, i.e., using cloud top height, cloud top pressure, and cold cloud depth as predictors to estimate the lightning density. They report overall successful results for continental Europe during summer.

McCaul et al. (2009) brought forward two lightning proxies based on ice-phase hydrometeor fields in WRF. "One method is based on upward fluxes of precipitating ice hydrometeors in the mixed-phase region at the −15 °C level, while the second method is based on the vertically integrated amounts of ice hydrometeors in each model grid column." They related domain-wide statistics of peaks in the proxy fields to peaks of the flash rate density. Overall, a blended solution using both proxies is suggested to reproduce both temporal variability and areal coverage of the lightning observations. A simulation of Mediterranean lightning events revealed that lightning data were hardly correlated with the temporal distribution of convective rainfall, while the temporal distribution and maximum of simulated solid hydrometeors correlated well with those of lightning occurrences. The correlations were stronger during the development of the storms than during the decay, where lightning activity decreased faster than the solid hydrometeor concentrations. Yair et al. (2010) introduced a lightning potential index (LPI) as

kinetic energy of the updraft in a developing thundercloud, scaled by the potential of charge separation based on ratios of ice and liquid water between 0 °C to −20 °C. LPI is calculated using the simulated grid-scale vertical velocity and simulated hydrometeor mass mixing ratios of liquid water, cloud ice, snow, and graupel.

There are also NWP studies that investigated the use of different model parameters for a lightning parametrization. For example, Barthe et al. (2010) looked at precipitation ice mass, ice water path, ice mass flux product, updraft volume, maximum vertical velocity, and cloud top height used to predict lightning rates with WRF. As a result of two case studies, one severe and one ordinary storm, they concluded that no proxy could predict the lightning flash rates and trends well for both cases. The most reliable results were still obtained for the maximum updraft velocity, the precipitation ice mass, and the ice flux product. Bovalo et al. (2019) compared simulated lightning to model proxy such as total graupel mass, graupel volume, updraught volume and maximum speed, and a product of precipitating and non-precipitating ice mass flux. They tested linear relations between predicted flash rate and the proxies for eight storms. Graupel mass and ice mass flux appeared to be the best estimators. Eventually, only graupel mass led to reasonable results for a test case, while remaining proxies overestimated flash rates and produced spurious lightning (likely due to positive y-intercepts of linear regressions for updraft volume and ice mass flux). Their cell scale approach resulted in more robust relationships than a domain scale approach (i.e., for graupel mass-flash rates Pearson correlation coefficients of r=0.82 and r=0.53, respectively). Formenton et al. (2013) related simulated lightning and model microphysics in a 1D model. The ice-ice noninductive charging was important with a key role of graupel in cloud electrification. The IC activity in the charging zone was strongly correlated with graupel mixing ratio, and CG with the presence of graupel (negative charge). Highest IC rates were observed in correspondence with strong updrafts and presence of supercooled water in higher parts of the cloud causing the upper level positive charge layer due to ice crystals (high IC lightning). A given threshold of simulated flash rate required a minimum threshold of both graupel mass and updraft in the column. A lightning simulation based on graupel carrying negative charge and ice carry positive charge was used by Stefano et al. (2014). Graupel and ice areas act as capacitor planes in their cloud-resolving model. Several studies cited in Section III.4.2 also demonstrated successful LDA using proxy graupel mass (e.g., Qie et al., 2014, Wang et al., 2017a, 2018, Kong et al., 2020).

In consequence of the former studies, graupel mass appears to be a reliable proxy for lightning. Our observation operator is trained based on an observed relation between pseudo MTG-LI FED observations and AROME-France simulated graupel mass (m_g) for 24 days in 2018 (2 per month). FED time periods of 10 min are used (like, e.g., Fierro et al., 2016, Hu et al., 2020b) centered at the corresponding time of the AROME-France analysis. The graupel mass profile is extracted from the AROME-France grid point closest to the FED pixel center and transformed to the FED grid. Following Deierling et al. (2008), column integrated graupel mass is taken for layers where the temperature was below −5 °C. Graupel mass combines graupel and hail. All FED and column graupel mass values are further processed as climatological distributions without observation location and time yielding a very strong Pearson correlation coefficient of 0.96 (see Figure III.4.3). 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. This approach is different from Deierling et al. (2008) and Barthe et al. (2010) who used storm-based relationships of case studies, and from McCaul et al. (2009) who used the domain-wide peak values. Pixel-to-pixel m_g and FED were barely correlated here, with about 0.09 Pearson correlation coefficient, likely as a consequence of a typical displacement of convection in the model by more than the FED spatial resolution of

7 km. The second approach with domain-wide peak values was not further tested as it reduces the sample size for a regression analysis drastically. Our observation operator uses a simple linear regression between (pseudo-)observed FED and simulated column graupel mass. Our approach optimizes both the slope factor and the y-intercept of the regression, whereas, e.g., McCaul et al. (2009) only used a proportionality between FED and a proxy. An advantage of the linear regression-based observation operator is that no bias correction is needed as the linear regression is unbiased by definition.

Figure III.4.3 presents the analyzed linear relationship between FED and column integrated graupel mass m_g that are calculated per 7 km × 7 km pixel whereby m_g uses a single model grid point closest to the pixel center. The training data (24 days) results are shown in Figure III.4.3(a) and the results for 06 and 07 October 2018 as validation period can be seen in Figure III.4.3(b). The Pearson correlation coefficient equals 0.97 for the training data and 0.92 for the validation data.

Analyzing the training data, the linear relationship fits the majority of the data well. A discrepancy is identified for the largest graupel mass and FED values. Here, the observation operator tends to underestimate the FED for a given graupel mass. The high values (of both FED and graupel mass) are rare relative to the lower values as indicated by the colored pixels in Figure III.4.3(a). Hence, this observed discrepancy does minorly affect the Pearson correlation coefficient. It should still be considered that high FED values are systematically underestimated. It is further noted that the y-intercept is negative, meaning a certain mass of graupel is required to get the first lightning. This result is well in accordance with the widely accepted non-inductive charging as main electrification processes in extratropical storms.

The validation data (Figure III.4.3(b)) do not exactly conform the training data, i.e., the values of FED are always lower for a given m_g than the training data implies. The number of FED-m_g-pairs is significantly lower for the 48-hour validation case than for the 24-day training dataset. Furthermore, the validation data comprises one single meteorological situation, while the training data is gathered as a climatological dataset. The observed discrepancy is considered during the evaluation of results. In fact, the AROME-France background FED will be higher on average than the observed FED at 06 and 07 Oct. 2018.

III.4.6 1DBay+3DVar assimilation method

The observation operator is used in the new lightning data assimilation method. Unlike Fierro et al. (2019) and Hu et al. (2020b) and others who used an empirical method to adjust moisture in thunderstorms, the expected water vapor pseudo-observation is retrieved for each model grid column (applying Bayes' Theorem) in a 1D-Bayesian (1DBay) approach. This allows to i) replace the humidity field in spurious convection areas with that of their environment, and thus abolish the wet bias that results from data assimilation techniques that only consider the occurrence of lightning (e.g., Fierro et al., 2012, 2019, Hu et al., 2020b), and ii) make use of the FED value to modulate the humidity field in observed lightning areas. Technically speaking, synthetic profiles are created and assimilated as sounding data. A great advantage of doing so is to avoid tedious coding in the AROME-France data assimilation system while providing agility and flexibility, which allow fast validation and effective research.

The 1DBay retrieval of pseudo-observations of relative humidity RH, $y_{\mathrm{po}}^{\mathrm{RH}}$, is defined as

$$y_{\mathrm{po}}^{\mathrm{RH}} = \sum_i x_i^{\mathrm{RH}} \frac{W_i}{\sum_j W_j} \tag{III.4.1}$$

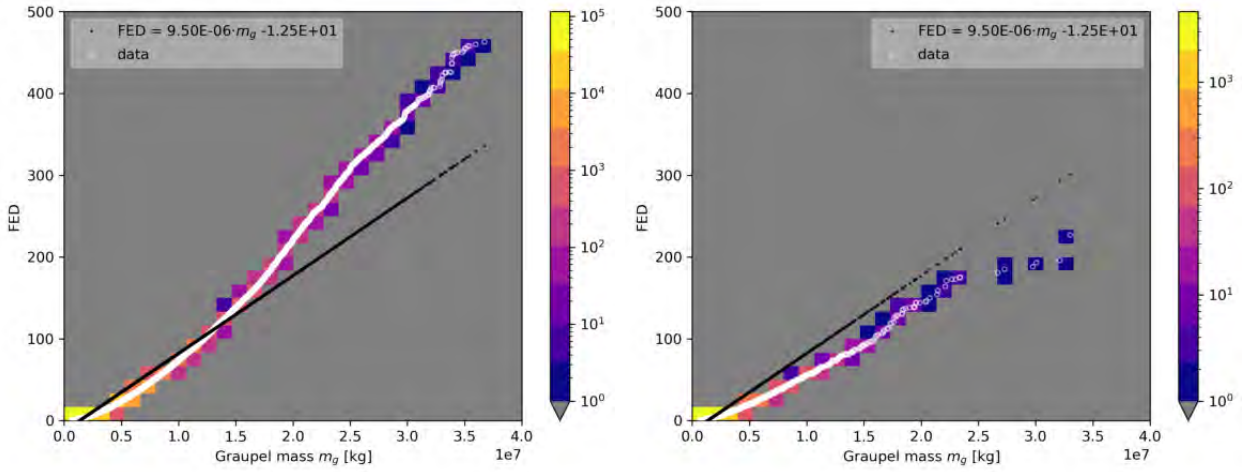


Figure III.4.3: The observed FED distribution versus the distribution of AROME-France graupel mass m_g at model grid points closest to each FED observation. Grid points with any of FED or m_g equal to zero are not considered. (a) shows the training of a linear regression for 24 days in 2018, and (b) shows the results of a validation for independent data of 2 additional days in 2018. Colors indicate the number of samples, the white points plot the data points, and black points applied the m_g values in the given linear regression.

with the weights W_i for each point as

$$W_i = \exp\left\{-\frac{1}{2}\frac{[\text{FED} - F(x_i)]^2}{\sigma_o^2}\right\} \tag{III.4.2}$$

where i, j are counters for the grid points within the defined vicinity area, and F means the FED forward operator as specific observation operator. $F(x_i)$ then defines the simulated AROME-France FED.

Equations (III.4.1) and (III.4.2) can be used with different model variables or diagnostics, i.e., the sensitivity test for σ_o as described in the following used x_i^{FED} rather than x_i^{RH}.

Each RH pseudo-observation is the best estimation from a weighted linear combination of RH profiles taken from the model background in the vicinity of the FED observation.

The method was proposed by Caumont et al. (2010), used in the operational model AROME-France (Wattrelot et al., 2014), and applied by Borderies et al. (2019) for radar reflectivity data and Duruisseau et al. (2019) for microwave radiances. Here, weights are calculated based on differences between observed and simulated FED (instead of reflectivity). The 1DBay retrieves the best estimation of RH at the center of each observed FED pixel from the background using Bayes' Theorem. It is expected that the model can predict a quantity similar to the observation within a certain area in most cases. The area (termed vicinity hereafter) is initially fixed to a square of 160 km × 160 km centered at the observation. This size follows the suggestion of Borderies et al. (2018, 2019). Since however, the FED pixels (about 7 km) are significantly larger than the radar reflectivity pixels (about 1 km), a lower number of values is available to retrieve the expected profile. Hence, the vicinity is adapted to the specifications of the FED data in different LDA experiments as described in Section III.4.7. The database comprises model forecasts at the observation time and in the vicinity of the observation location. The covariance matrix of observation and observation operator is chosen diagonal with a trace of $n \cdot \sigma_o^2$, where n is the vertical dimension of the observations and equals 1 for the 2D variable

FED. A small σ_o means that the retrieval favors columns with values close to the FED observation. This can produce accurate retrievals; however, no retrieval can be created the difference between all simulated FED in the vicinity and the observed FED is large relative to σ_o (see Equation (III.4.2) where weights yield 0 in that case). Large values of σ_o cause smoothing over all grid points in the vicinity of the observation. The likelihood to retrieve pseudo-observations is high, at the cost of a less accurate retrieval potentially independent of the observation. The value for the standard deviation σ_o of the observation and observation operator is inferred from a sensitivity study. It aims at minimizing the root mean square error (RMSE) between observed and simulated (using the observation operator and 1DBay retrieval) FED for the 24 training days also used in Section III.4.5. The FED is used in units of dB ($10 \cdot \log_{10}(\text{FED})$) to account for different scales. FED equal to zero is transformed to -10 dB which corresponds to a value between 0 and 1, knowing that FED is expressed as an integer in linear unit. Figure III.4.4(a) shows the curve of the RMSE between FED observation (from pseudo MTG-LI) and retrieved pseudo-observations from the 1DBay retrieval for different σ_o and as a reference the AROME-France background FED, referred to as AROME_FED, directly obtained from the observation operator. The 1DBay retrieval produces pseudo-observations much closer to the FED observation than the AROME_FED. The minimum RMSE for the retrieval is found at σ_o of 2.0 dB. This value is used for retrieving the pseudo-RH profiles.

Case studies of fields of observed, background, and retrieved FED are conducted to visualize the effect of σ_o on the pseudo-observations. One example of a 07 Oct. 2018, 00:00 UTC case is presented in Figure III.4.4(b-d). The FED observation Figure III.4.4(b) shows lightning activity mainly in the south center of the domain. AROME_FED (Figure III.4.4c) exhibits positive FED in the region of FED observation, however, also widespread in the northern center of the domain. The 1DBay retrieved FED in Figure III.4.4(d) demonstrates that the method effectively reduces the spurious FED in the north. The marine blue indicates that relatively high values from the AROME_FED are decreased to values close to -10 dB. Furthermore, the FED values and areas of positive FED in the south center of the domain closely approximate the FED observation.

An additional method, the humidity adjustment (HA), is applied if lightning (i.e., positive FED) is observed but all background FED values in the vicinity are zero. In this case, there is no estimated profile and all W_i equal zero. The HA is also applied if all background FED values within the vicinity of the FED observation are at least one order of magnitude smaller than the observed FED value. Although the sum of W_i can become greater than zero in this case, the retrieval would generate a profile that is too dry with respect to the FED observation. To produce RH pseudo-observations, the layer between lifted condensation level (LCL) and 13 km is saturated (i.e., RH set to 100 %) if the modeled RH is less than 100 %. This is conceptually similar to the method of among others Fierro et al. (2019). However, the HA is only applied for few pseudo-RH profiles where the 1DBay method did not retrieve pseudo-observations (POs).

Another case with $\sum_i W_i = 0$ may occur if the pseudo MTG-LI FED equals zero and the background FED is positive for all grid points in its vicinity. One would need to remove spurious convection and humidity from the model in this case. However, it has never been observed here[1]. It is unlikely that all background grid points exhibit lightning activity as long as the vicinity comprises a sufficient area to effectively apply the 1DBay retrieval. Eventually, no pseudo-RH profiles are created if both the pseudo MTG-LI FED as observation and the closest (in space) AROME-France background FED equal zero.

It should be mentioned that the 1DBay retrieval has been successfully applied to retrieve

[1]The case occurred for the initial 160-km vicinity, but not within the 320-km and 500-km vicinity, Section III.4.7.

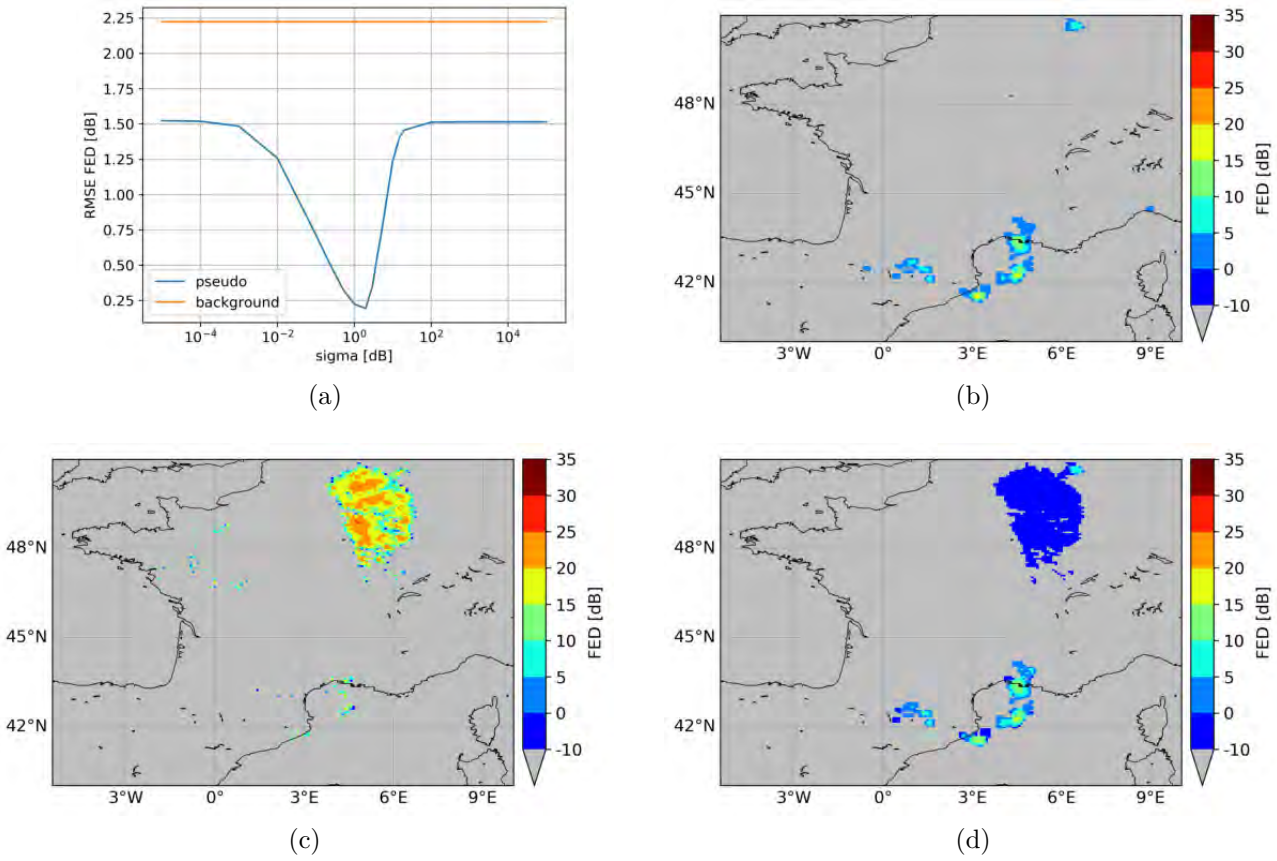


Figure III.4.4: (a) The sensitivity test for standard deviation σ_o of the 1DBay retrieval as inferred from the pixel-to-pixel RMSE between the FED observation and the AROME_FED (background) as well as the 1DBay retrieved FED (pseudo). (b-d) 07 Oct. 2018, 00:00 UTC, case with the FED observation as of pseudo MTG-LI data (b), the model background AROME_FED (c), and the 1DBay retrieved FED (d).

humidity and cloud profiles from microwave and radar reflectivity data (e.g., Olson et al., 1996, Kummerow et al., 2001, Caumont et al., 2010, Wattrelot et al., 2014, Borderies et al., 2019), however, it has not yet been applied to retrieve humidity profiles from FED data. Whereas an FED greater than zero is always related to the presence of graupel and thus a RH profile with a cloud, a FED equal to zero does not necessarily mean a location without cloud coverage. Indeed, there is no FED lower threshold to specify whether a cloud is present or not, as is the case for both microwave data and radar reflectivity. In addition, FED is a 2D variable and the 1DBay retrieves a 3D RH profile.

The retrieved pseudo-RH profiles are assimilated as sounding data in the 3DVar assimilation system of AROME-France. AROME-France uses a one-hour assimilation window. The short assimilation cycles aim to partially overcome the missing temporal dimension that 4DVar takes into account, and to allow an assimilation of more high-frequency observations that can improve the initial conditions especially on the convective scale. AROME-France operationally assimilates conventional (e.g., ground-stations, ships, buoys) and aircraft measurements, Global Positioning System (GPS) Zenith Tropospheric Delay (ZTD) data, satellite brightness temperatures of several polar orbiting satellites and from Meteosat Second Generation (MSG) SEVIRI, satellite-based atmospheric motion vectors, and radar velocity and reflectivity data. The control variables are temperature, specific humidity, surface pressure, and horizontal wind components. The 3DVar systems minimizes a classical 3DVar cost function J of the state vector x:

$$J(x) = \frac{1}{2}(x - x_b)^T \mathbf{B}^{-1}(x - x_b) + \frac{1}{2}[y_o - H(x)]^T \mathbf{R}^{-1}[y_o - H(x)] \tag{III.4.3}$$

With the state vector of the background x_b, the observation vector y_o, the observation operator H, and the observation error covariance matrix $\mathbf{R}$. The climatological background error covariance matrix $\mathbf{B}$ is inferred from offline AROME-France ensemble assimilation as a multivariate set of calculations for the control variable covariances and cross-covariances (Brousseau et al., 2014).

III.4.7 Model simulation experiments

This work expands the set of the observations assimilated in AROME-France in order to compare and investigate the effect of LDA relative to RDA. Since the application of the 1DBay retrieval for FED data constitutes a new approach, different experiments of AROME-France with respect to the LDA are conducted. Table III.4.1 lists the eight different assimilation experiments. First, there is a control experiment used as reference without RDA, without Doppler wind speed assimilation, and without LDA. It is called *R*. The second experiment is the current operational AROME-France using RDA and Doppler wind speed assimilation. It is referred to as *RDA*. A similar assimilation but without the Doppler wind speed is called *RDA_noUV*. Using LDA with the dynamic 1DBay vicinity of 160 km or 320 km, the experiment *LDA* is defined. If RDA (without Doppler wind speed assimilation) is added to the experiment *LDA*, the experiment *RDA_LDA* (*RDA_noUV_LDA*) is created. Finally, there are two additional pure LDA experiments, i.e., like the control experiment without any RDA but with LDA. The first, *LDA_R500*, uses a vicinity of 500 km in the 1DBay retrieval. The second experiment with LDA and a vicinity in the 1DBay enlarged to 500 km, *LDA_R500nC*, adds a *noCloud*-filter. This *noCloud*-filter is utilized for locations where the observed FED equals zero but the AROME_FED exhibits lightning activity. Then, the distance to the closest positive FED observation, d_{FED}, is computed. If d_{FED} remains within 21 km, i.e., three FED pixels, it is assumed that the profile is still situated within the same thundercloud being responsible for

Table III.4.1: Simulation and assimilation techniques of the different AROME-France experiments conducted. All model experiments the observations given in Section III.4.6. They differ by the use of radar data assimilation (RDA) and lightning data assimilation (LDA).

Experiment	RDA and LDA	Forecast period	1DBay vicinity	Note
R	Reference (control experiment)	30 h	-	-
RDA	RDA	30 h	-	-
RDA_noUV	RDA, no Doppler wind	30 h	-	-
LDA	LDA	30 h	160 km to 320 km	-
RDA_LDA	RDA, LDA	30 h	160 km to 320 km	-
RDA_noUV_LDA	RDA, no Doppler wind, LDA	30 h	160 km to 320 km	-
LDA_R500	LDA	30 h	500 km	-
LDA_R500nC	LDA	30 h	500 km	*noCloud*-filter

the positive FED observation. In this case, the RH profile of the background is kept to avoid reducing the RH if FED equals zero but the location is likely associated with a cloud. In the case where d_{FED} exceeds 21 km, AROME-France profiles within 21 km are not considered in the 1DBay as they might belong to the spurious thundercloud that is modeled but not observed. Hence, the *noCloud*-filter should also help to effectively reduce the background humidity.

III.4.8 The assimilation time period and background modifications

AROME-France simulations are analyzed for 06 and 07 October 2018. A major cyclone entered the model domain from the Atlantic Ocean, bringing widespread precipitation on 06 October. The cyclone then dissipated while maintaining precipitation mainly over mountainous terrain and along the occluded cold front that moved eastward over the Mediterranean Sea and southern France. The center of the former cyclone was situated near the French-Swiss border in the end of the forecast period. The actual forecast period starts 07 October 2018, 00:00 UTC and runs for 30 hours. All model experiments are initiated one day prior to the start of the model forecast. During the first 24 hours, the 3DVar assimilation system of AROME-France creates 23 analyses (there is no analysis for the first time step). Therefore, reference experiment, the LDA, and RDA are conducted for 23 hours in advance of the evaluated forecast. This time period is chosen as convection was observed within the model domain continuously. The long assimilation period allows AROME-France to well ingest the available observations.

This section describes the results of the LDA to modify and update the model background. Figure III.4.5 presents the FED observation (a), the FED inferred from the AROME-France background using our observation operator (b), the background integrated water vapor (IWV) field (c), and the difference between the background IWV corrected by the POs, called PO IWV, and the background IWV itself (d). The example illustrates the situation on 07 October

2018, 00:00 UTC for model experiment *LDA_R500*. The FED observation (Figure III.4.5a) and the AROME_FED (Figure III.4.5b) indicate that the model background contains spurious convection for a large region in the northern center of the domain, and that the lightning activity is underestimated by AROME-France between 41.5°N and 44°N and 1°W and 5°E. The background IWV (Figure III.4.5c) shows that both those regions are the most humid zones within the studied domain. The assimilation should still increase the humidity between 41.5°N and 44°N and 1°W and 5°E, and reduce the humidity in the northern center of the domain. Indeed, the difference between the PO IWV and the background IWV (Figure III.4.5b) exhibits positive values where the observed FED shows lightning. In detail, there are both positive (red) and negative (blue) differences in the region between 41.5°N and 44°N and 1°W and 5°E that result from slightly different locations of the FED maxima that are about 0.2° to 0.5° further west in the AROME_FED than in the observation. Hence, IWV is reduced in the PO IWV compared to the background IWV on the western edge of the positive FED there, and increased for eastern part of the positive FED observations that are not seen in the AROME_FED. In the northern center, the 1DBay retrieval mainly reduces the IWV in the region of spurious FED, however, also higher PO IWV than background IWV occurs at some locations in this region of spurious FED. The former is the desired result as it helps to suppress spurious convection in AROME-France. The latter is likely associated to the fact that locations with FED equal to zero can still be found for cloudy vertical model columns. A detailed look reveals that the red zones indicating higher PO than background humidity (Figure III.4.5d) match the zones of the local, i.e., for this system, minima in the background IWV (Figure III.4.5c). As the 1DBay retrieval might find both grid points with FED equal to zero within the cloud system and points with FED equal to zero outside the cloud system, the IWV of local background IWV minima in a region of spurious convection can still be increased. To address this issue, the *noCloud*-filter as described in Section III.4.7 is applied.

Figure III.4.6 shows the model experiment *LDA_R500nC* at the same time as Figure III.4.5 for *LDA_R500*. Each model experiment uses a vicinity of 500 km in the 1DBay. *LDA_R500nC* applies the *noCloud*-filter that may force either no change to the background IWV or excluding profiles in the close vicinity of the FED observation from the 1DBay retrieval. The difference of PO IWV minus background IWV in Figure III.4.6(d) shows that the background IWV is almost exclusively reduced for model profiles with spurious convection. In particular, *LDA_R500nC* seems to enhance the experiment *LDA_R500* in that regard. It should be mentioned, however, that the AROME-France background also depends on the previous hours of the assimilation cycle. Therefore, the presented AROME_FED and background IWV in the Figures III.4.5 and III.4.6 are different for the model experiments *LDA_R500* and *LDA_R500nC*, respectively. A direct comparison is thus not possible for the long term forecast here. One might note that the discrepancy in the relation of FED and m_g (Sec. III.4.5) for high values is somewhat balanced by the Bayesian approach itself. Nevertheless, the retrieved profiles of RH are likely associated with lower m_g than the background value as the trained observation operator usually overestimates the FED for a given m_g on the dates used in Figures III.4.5 and III.4.6. The retrieved RH profiles contain usually less m_g and often also less humidity than one would expect from the given FED observations during the validation period.

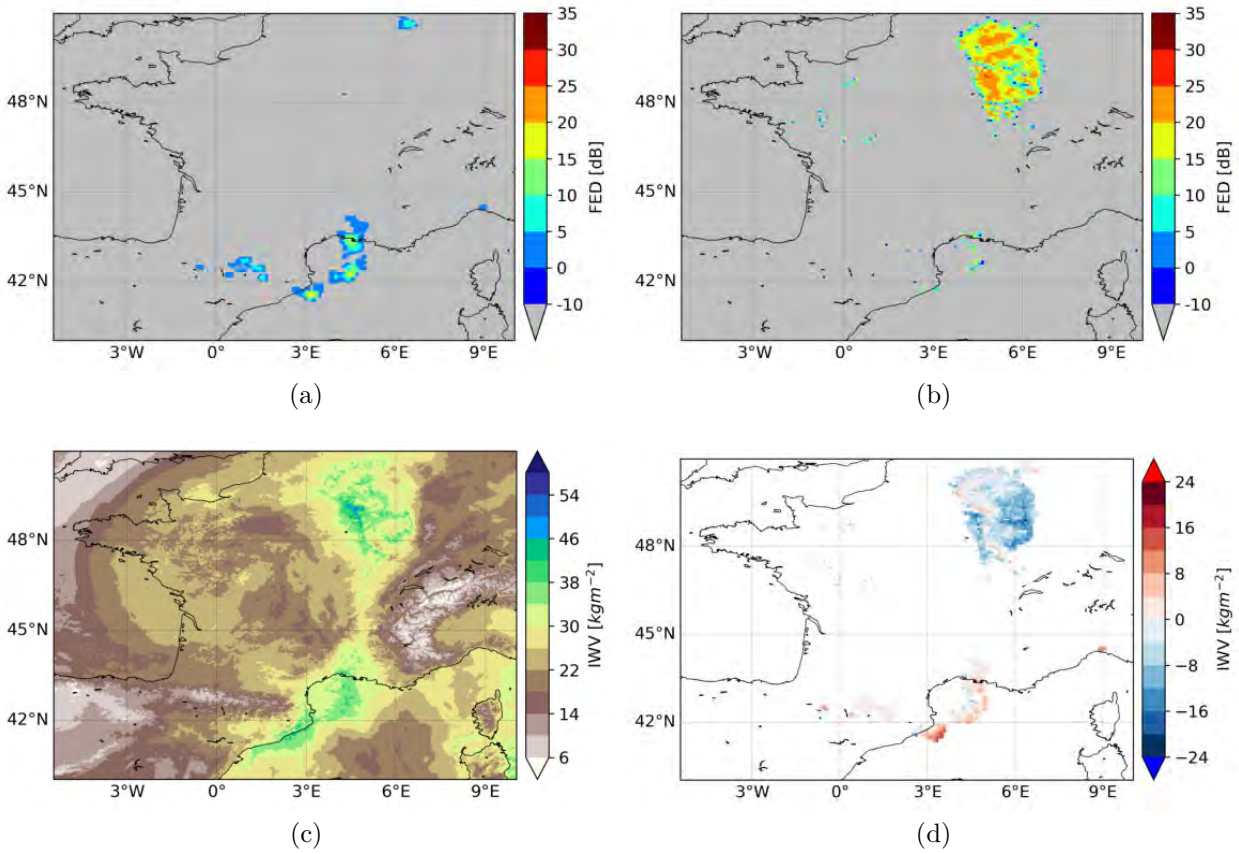


Figure III.4.5: FED observation as of pseudo MTG-LI data (a), the background AROME_FED (b), the AROME-France background IWV (c), and the difference between the IWV of the background corrected by the 1DBay retrieved RH profiles (referred to as PO IWV) minus the background IWV (d). Results for 07 Oct. 2018, 00:00 UTC and model experiment *LDA_R500*.

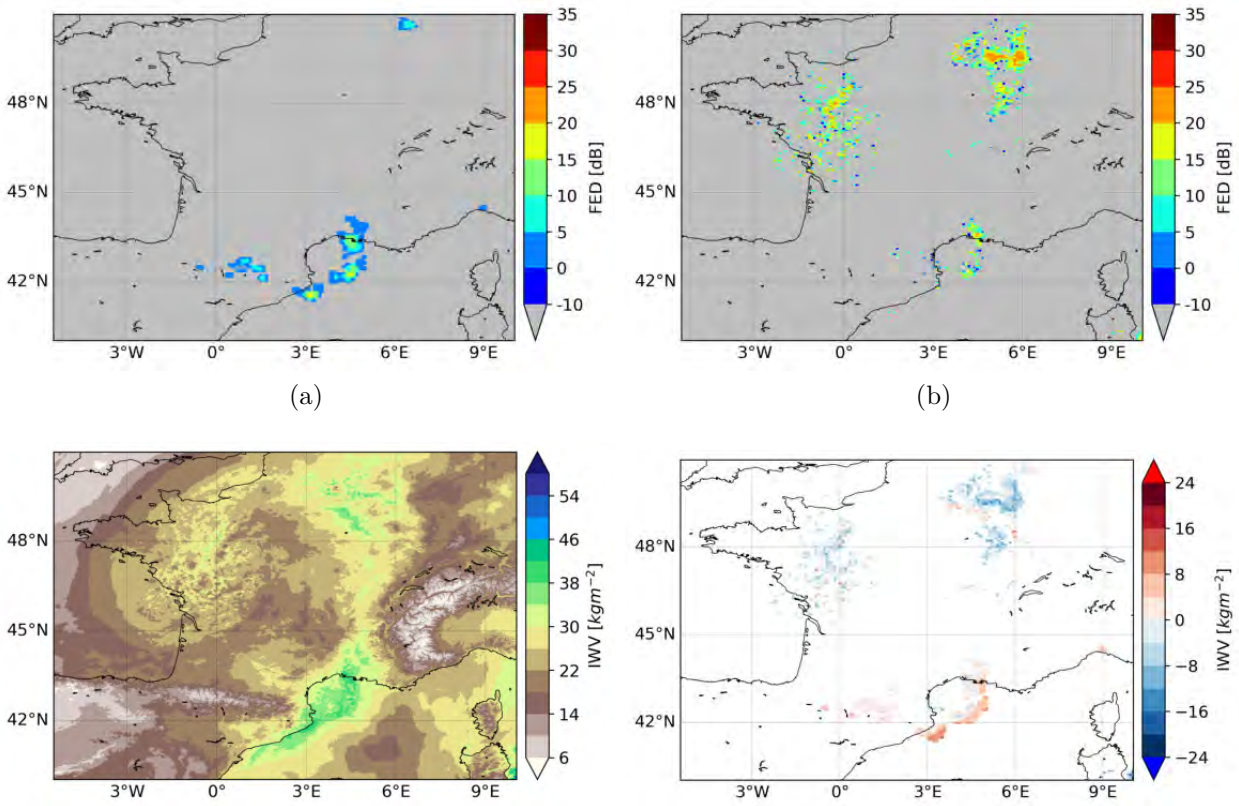


Figure III.4.6: As Figure III.4.5 but for model experiment *LDA_R500nC* using the *noCloud*-filter.

III.4.9 The AROME-France analysis using FED observations

The 1DBay retrieval shows the potential to increase the model humidity in regions of missing convection and to reduce the background humidity in regions of spurious convection. The logical next step is the implementation of the 1DBay+3DVar LDA technique in AROME-France. The 1DBay-based PO RH profiles are added to the set of observations in AROME-France and assimilated as dropsonde data; however, the analysis does not change in the way that one must expect from the POs and FED observations when compared to the model background.

One prominent example is presented in Figure III.4.7 for the model experiment *LDA_R500nC* on 06 Oct. 2018, 13:00 UTC. The first row of the figure (a-c) shows the AROME_FED (a), the 1DBay FED (b), and the FED observation (c). From these plots one can expect that the 1DBay retrieval adds humidity to the background for the positive FED observation near 47°N and 4°W. In addition, the 1DBay FED (Figure III.4.7b) indicates that profiles are retrieved in the south-east (SE) to effectively reduce the spurious FED of the AROME_FED (Figure III.4.7a). Thus, IWV should be reduced in the SE. The PO IWV (Figure III.4.7e) clearly decreases compared to the background IWV (Figure III.4.7d). The difference of PO IWV minus background IWV (Figure III.4.7g) supports this finding. Therefore, the 1DBay retrieval finds the profiles that should lead to an analysis with the expected increase and decrease of IWV. Despite that finding, the analysis IWV (Figure III.4.7f) adds humidity to the background in the SE, as can also be seen by the difference of analysis IWV minus background IWV (Figure III.4.7i). Figure III.4.7(h) shows in addition the IWV as pure output of the 1DBay retrieval (background IWV not included). It equals the PO IWV in the SE of the domain; hence, complete profiles are retrieved there rather than for example using the HA.

In conclusion, the analysis changes the humidity field in the opposite direction as the AROME_FED and FED observations indicate, and even more contrarily to the assimilated 1DBay-based POs. Similar behavior is found for all model experiments with LDA, including experiments that use both LDA and RDA. The issue has not been identified for the control experiment or pure RDA experiments. None of the analyses related to the control experiment or pure RDA exhibit IWV values as high as the LDA experiments. Hence, the assimilated observations besides the FED-derived RH profiles do not explain why such high IWV are found in the LDA experiment analyses, and neither why the analysis IWV increases despite a PO IWV lower than the background IWV.

III.4.10 Case study and evaluation of the LDA in AROME-France

One case study is conducted evaluating the results of the LDA using the 1DBay+3DVar technique. In a 30-hour forecast beginning 07 Oct. 2018, 00:00 UTC, eight AROME-France experiments with different assimilation (see Section III.4.7) are evaluated by a visual inspection of rainfall accumulations RAs and the objective calculation of the 6 h-RA FSSs. The FSS was introduced by Roberts and Lean (2008): The FSS can be calculated as skill score from the mean squared error (MSE) for the observed and forecast fraction $O_{(n)}$ and $M_{(n)}$, respectively,

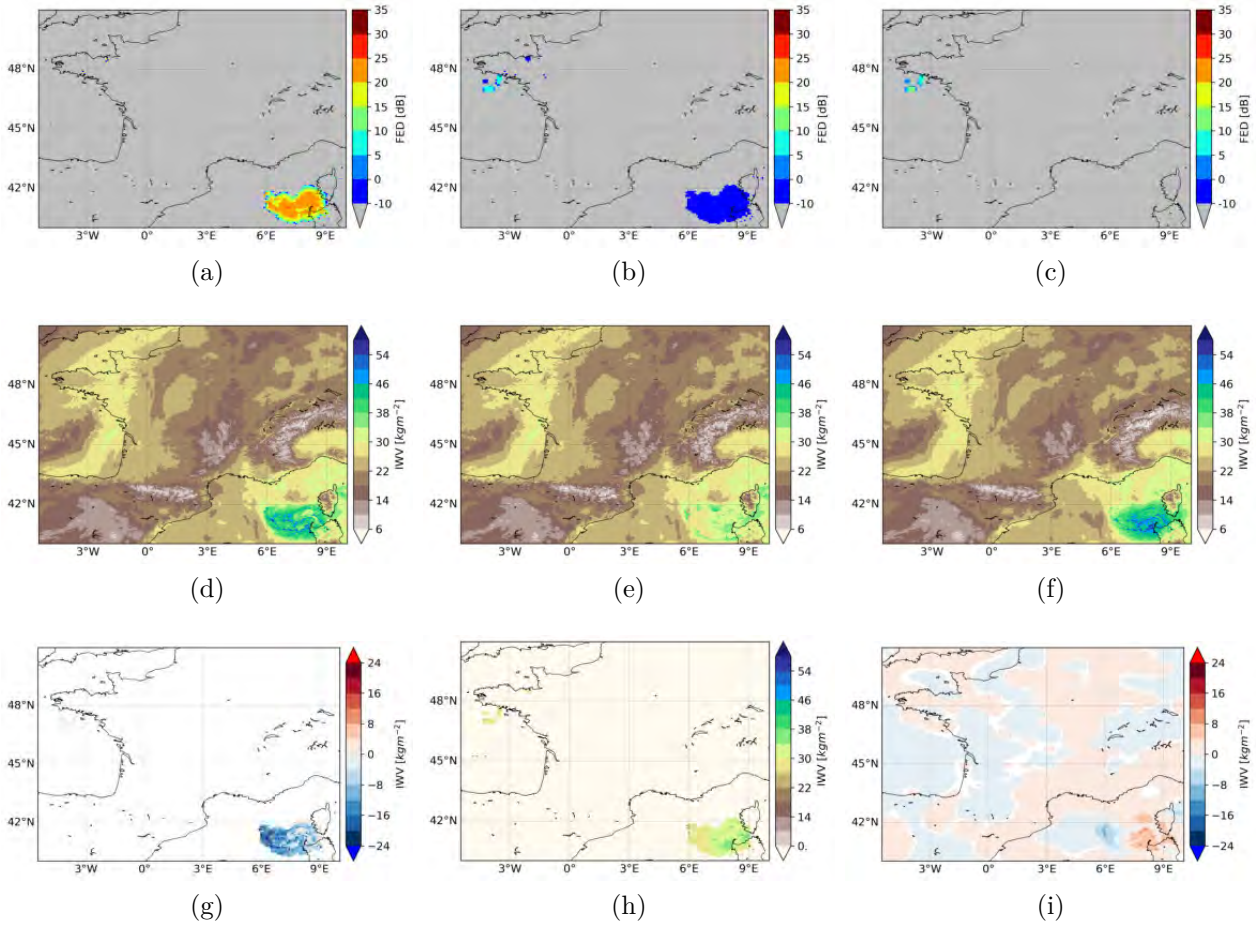


Figure III.4.7: The background AROME_FED (a), the 1DBay retrieved FED (b), the FED observation as of pseudo MTG-LI data (c), the AROME-France background IWV (d), PO IWV (e), the IWV of the analysis (f), the difference between the PO IWV minus the background IWV (g), and the pure PO IWV including only points where RH was retrieved (h), and the difference in IWV of analysis minus background (i). Results for 06 Oct. 2018, 13:00 UTC and model experiment *LDA_R500nC*.

from a neighborhood of length n as

$$\mathrm{FSS}_{(n)} = 1 - \frac{\mathrm{MSE}_{(n)}}{\mathrm{MSE}_{(n)ref}}$$

with

$$\mathrm{MSE}_{(n)} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[O_{(n)i,j} - M_{(n)i,j} \right]^2$$

$$\mathrm{MSE}_{(n)ref} = \frac{1}{N_x N_y} \left[\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} O^2_{(n)i,j} + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} M^2_{(n)i,j} \right] \tag{III.4.4}$$

and

$$O_{(n)i,j} = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} I_O \left[i + k - 1 - 1\frac{n-1}{2}, j + l - 1 - \frac{n-1}{2} \right]^2$$

$$M_{(n)i,j} = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} I_M \left[i + k - 1 - 1\frac{n-1}{2}, j + l - 1 - \frac{n-1}{2} \right]^2$$

with the observation and forecast binary fields I_O and I_M that equal 0 if the field value is smaller than the threshold, and 1 in all other cases. Our implementation uses the fast calculation of FSS in Python as proposed by Faggian et al. (2015). In this section, selected results are presented that indicate the current problems of the LDA in AROME-France.

The analyzed test case includes a 30-hour forecast that is initiated 07 Oct. 2018, 00:00 UTC using AROME-France. This specific case is selected based on a follow-up of the airborne field campaign EXAEDRE (Defer and EXAEDRE Team, 2019) studying lightning near Corsica island in September and October 2018. In particular, there were daily weather briefings to reserve flight zones for the upcoming days. After consulting the AROME-France forecasts, no flight time was reserved for 07 Oct. 2018 as the model indicated a fairly weak potential for deep convection and thunderstorms.

Figures III.4.8 and III.4.9 present the observed (a,b) and predicted (c-j) RA fields for two 6-hour periods during the 30-hour forecast, i.e., at the start of the forecast accumulated until 07 Oct. 2018, 06:00 UTC and more towards the end of the forecast as accumulated over 6-hours until 08 Oct. 2018, 00:00 UTC. The observation of RAs combines the data of operational Doppler radar network of Météo-France and ground-based rain gauges to get the best estimation of the RA on a 0.01° resolution grid at each hour. This product is called *Antilope*. Figure III.4.8 (a) and (b) show the observed 6-hour RAs and also indicate the data is not available for the entire domain but only within the range of the radars. The areas without data coverage are kept white in Figures III.4.8 and III.4.9.

The RA observation shows a widespread region with rain over France in the beginning of the 30-hour forecast (Figure III.4.8a). A prevailing cyclone over western France that just reached its mature state was responsible for the precipitation. The cold front with the RA maxima was situated over eastern France and just entered the Mediterranean Sea at its southern end. During the course of the 30-hour forecast, the cyclone weakened and moved slowly south-southeastward. The 6-hour time period until 08 Oct. 2018, 00:00 UTC (Figure III.4.8b) saw localized precipitation over the Pyrenees mountain range and convection over SE mainland France and to the west of Corsica island. The center of the almost dissipated cyclone was located over the Pyrenees mountain range, with the rest of the occluded cold front being located in the region of the maximum RA.

Both the AROME-France experiments *R* and *RDA* capture the RA field in the beginning of the 30-hour forecast (Figure III.4.8 c and e, respectively). However, the RA maximum that

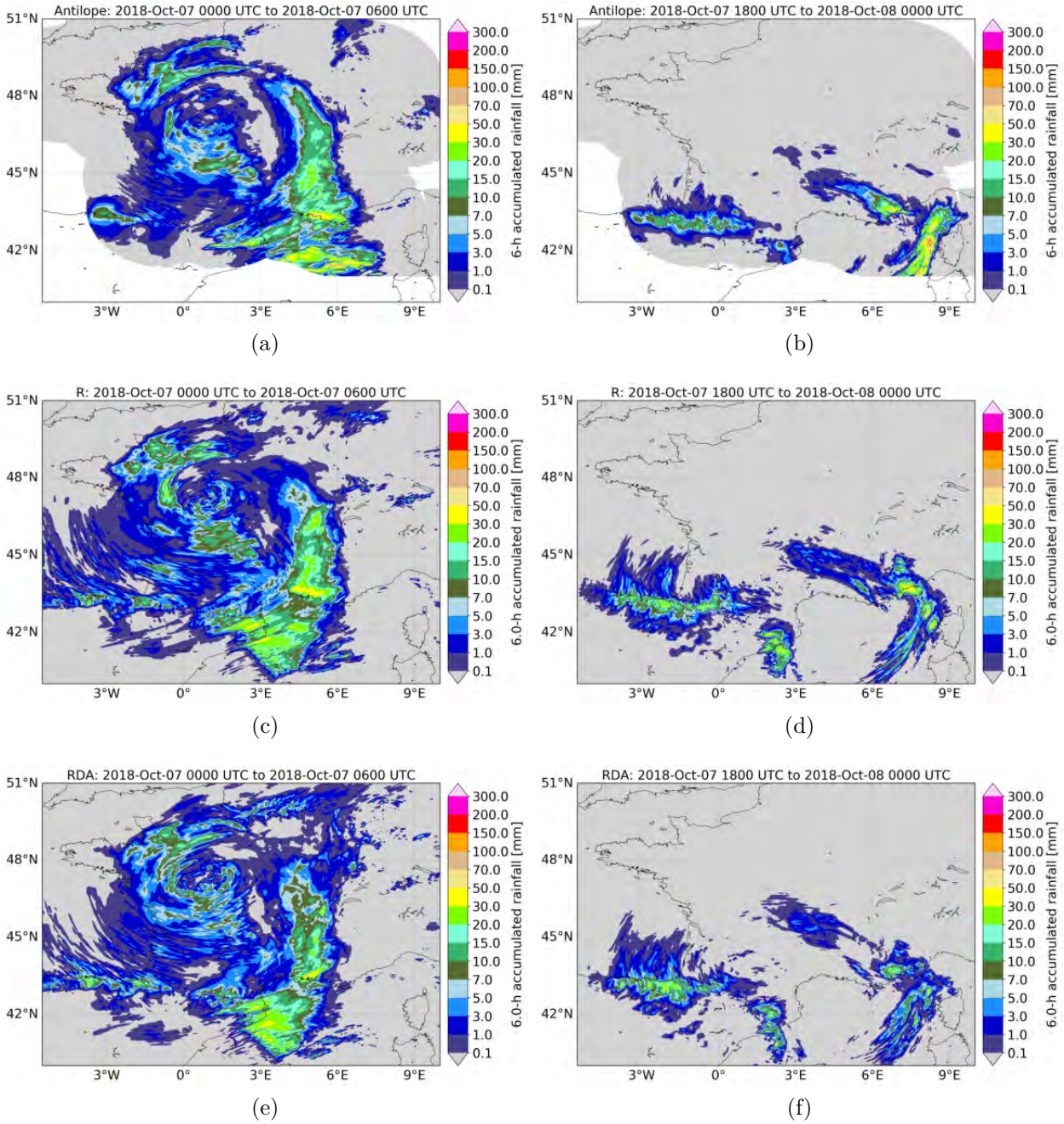


Figure III.4.8: 6-hour rainfall accumulations (RAs) for the period until 07 Oct. 2018, 06:00 UTC, i.e., the first 6 forecast hours (a,c,e,g,i) and the period until 06 Oct. 2018, 13:00 UTC, i.e., forecast hours 18 to 24 (b,d,f,h,j) of the 30-hour forecast initiated on 07 Oct. 2018, 00:00 UTC. The rows are the RA observations (a,b), model experiment *R* (c,d), *RDA* (e,f) [g-j in Figure III.4.9].

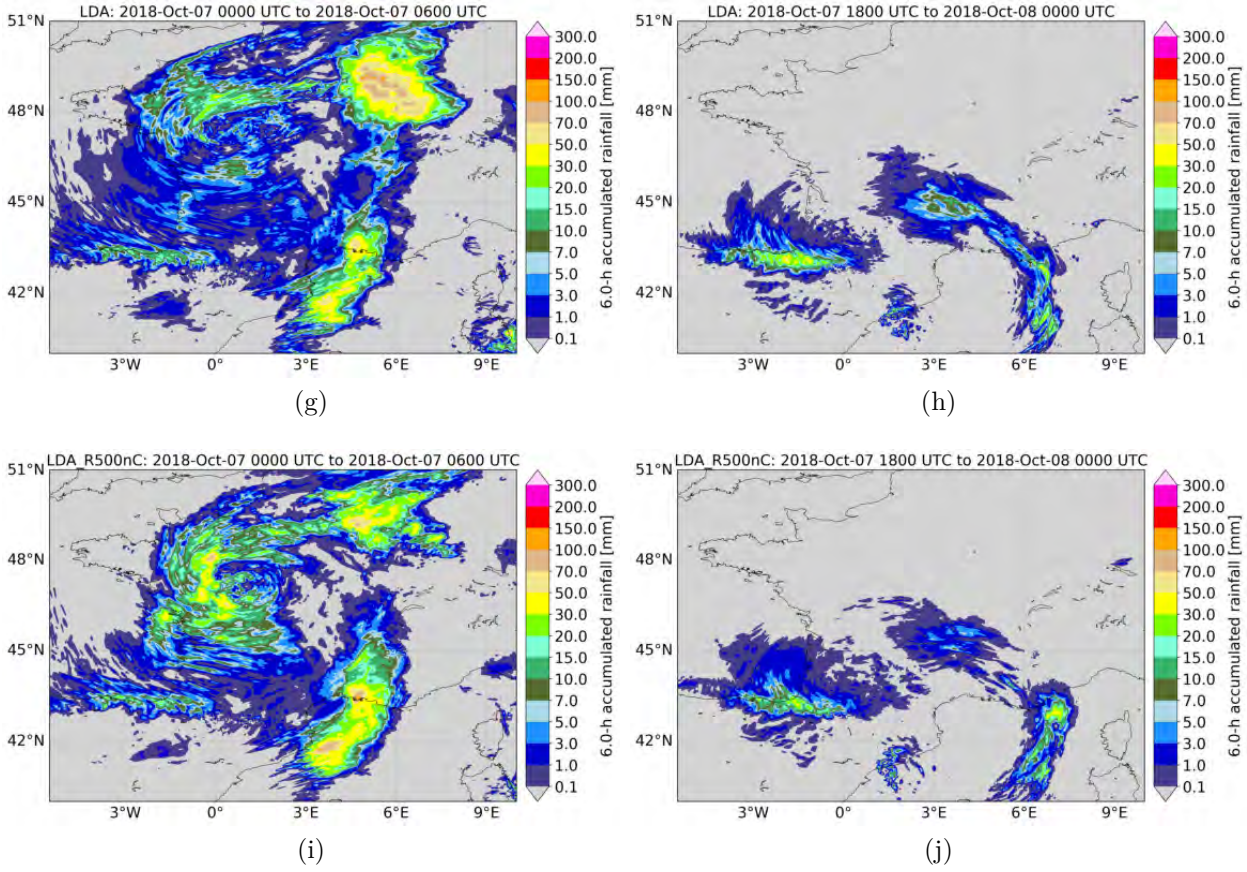


Figure III.4.9: Continuation of Figure III.4.8 for the AROME-France experiments *LDA* (g,h) and *LDA_R500nC* (i,j) with LDA.

is observed at about 41.5°N and 6°E and the high RA further to east are not predicted. The AROME-France experiments *R* and *RDA* predict but slightly overestimate the 6-hour RA over the Pyrenees mountain range until 08 Oct. 2018, 00:00 UTC. Differences exist between the RA over southern mainland France and to the west of Corsica. Both experiments lack the high RA observed to the west of Corsica. Here, the control experiment is still closer to observed RA maxima than the experiment *RDA*. The RA value is also better predicted by the control than the experiment *RDA*, however, placed about 0.5° too far to the east compared to the observation.

For the AROME-France with LDA in Figure III.4.9, higher maximum RAs are observed than for the RA observations and the experiments with RDA for the first period shown (Figure III.4.9g and i). Especially the experiment *LDA* (Figure III.4.9g) overestimated the RA over northeast (NE) France. The experiment *LDA_R500nC* also overestimated the RA over NE France (Figure III.4.9i), which is in accordance with the observed FED and AROME_FED as shown in Figure III.4.6. An overestimation of the RA is also seen at the southern end of the cold front for both LDA experiments. For the period from 18 to 24 hours of forecast, the AROME-France experiment *LDA_R500nC* predicts well the RA over the Pyrenees mountain range (Figure III.4.9j). The RA here is overestimated by the experiment *LDA* (Figure III.4.9h). Both presented AROME-France experiments with LDA place the precipitation to the west of Corsica further west compared to the observation. The results for the LDA are influenced by the analyzed issue in producing the AROME-France analyses while using FED observations (Section III.4.9). Further research regarding the production of the analysis will address this issue. Then, a similar study as presented here will show the true potential of the LDA in AROME-France.

The FSS is calculated hourly for 6-hour RAs, thus, as a sliding 6-hour time window, for the 30-hour forecast period. Figure III.4.10 presents the results for the RA thresholds of 0.1 mm (a), 1.0 mm (b), 4.0 mm (c), and 10.0 mm (d). The FSS also allows the determination of the size of the neighborhood that is used when comparing the observation and the forecast fields (Equation (III.4.4)). Here, neighborhoods of 0.1° (solid lines in Figure III.4.10) and 1.0° (dashed lines in Figure III.4.10) are shown. The AROME-France experiments in the legend are summarized in Table III.4.1. Interpreting the results, it should be mentioned that this is one case only, i.e., a 30-hour forecast that is initiated 07 Oct. 2018, 00:00 UTC.

As can be expected from reported problems with the LDA in Section III.4.9, the pure LDA experiments (pink, red, and black in Figure III.4.10) exhibit the lowest 6 h-RA FSS in the comparison of the model experiments. This is in general the case for all four RA thresholds and the two FSS neighborhoods. It is, however, still interesting that the FSS for the pure LDA using the small, dynamic vicinity of 160 km or 320 km in the 1DBay (*LDA*, pink curves) and the pure LDA using a vicinity of 500 km in the 1DBay (*LDA_R500*, red curves) yield very similar FSS for the four thresholds and two neighborhoods and throughout the entire duration of the forecast. Furthermore, the pure LDA with our *noCloud*-filter and a vicinity of 500 km in the 1DBay (*LDA_R500nC*) outperforms the other two pure LDA AROME-France experiments, especially for the lower RA thresholds of 0.1 mm and 1.0 mm. The result might be a first hint that the *noCloud*-filter indeed helps to effectively reduce the background humidity for spurious convection and prevent a humidity reduction if an observed FED equal to zero is actually matched to a location with a cloud. However, the 3DVar showed unexpected behavior that prevents a cogent conclusion.

The control AROME-France experiment performs surprisingly well relative to all model experiments with RDA and/or LDA for this case. That is true for all the four tested RA thresholds and the two FSS neighborhoods. Only the experiment *RDA_noUV* features slightly

higher FSS than the control AROME-France experiment. Towards the end of the 30-hour forecast period, the control experiment outperforms all AROME-France experiments using RDA and/or LDA for the higher 6-hour RA thresholds of 4.0 mm and 10.0 mm.

At this point, the results for the LDA and RDA are not further compared as there is still the unsolved issue with the analysis production assimilating the FED-inferred PO RH profiles in AROME-France (see Section III.4.9).

III.4.11 Conclusions

The objective of this work is the design of an assimilation technique for the upcoming MTG-LI data for the regional, convection-permitting model AROME-France. To date, AROME-France applies a 3DVar assimilation system. A 1DBay+3DVar assimilation technique (Caumont et al., 2010) is used to assimilate pseudo MTG-LI flash extent density (FED) in AROME-France. The same assimilation technique is currently used operationally for radar reflectivity data assimilation in AROME-France, but has not yet been tested for the lightning data assimilation (LDA).

This work first generated pseudo MTG-LI data that are used to create the FED observations. Then, an observation operator for FED is developed based on a linear, climatological relationship between observed FED and the column integrated AROME-France graupel mass above the −5 °C isotherm (m_g; as suggested by Deierling et al., 2008). The operator is trained for 24 days in 2018, and validated for 2 independent days in 2018. Pearson correlation coefficients of 0.97 and 0.92 for the training and validation data, respectively, reveal a very strong relation between the distributions of observed FED and model m_g. Nevertheless, the observation operator systematically overestimates the FED for m_g values greater than 1.5×10^7 kg per AROME-France grid cell of 1.3 km × 1.3 km. More sophisticated observation operators are currently tested but have not been included in this work yet.

The observation operator is then used to compare AROME-France-derived background FED (AROME_FED) to the observations. The 1DBay method identifies the best estimation of the background model locations that can in turn be used to create pseudo-observations (POs) of relative humidity (RH) profiles based on both the FED observation and the AROME-France background fields. It is found that the 1DBay retrieval leads to reduction of humidity where the observed FED equals zero and the AROME_FED is positive, thus in regions of spurious convection, i.e., represented as substantial m_g. The PO RH profiles do in addition add humidity to the AROME-France background where the observed FED exceeds the AROME_FED. Hence, our LDA technique is consistent with the meteorological situation and model physics as background model profiles are processed. It is capable of both promoting convection in regions with lightning and suppressing spurious convection. FED exhibits the highest values near the convective core of a thunderstorm and the lightning activity does not always cover the entire cloud size. In fact, zero FED exists at cloudy locations. In order to address the specific nature of FED data, the 1DBay retrieval is adapted. In detail, the vicinity used to look for profiles in the 1DBay is expanded and a so-called *noCloud*-filter is introduced. First results reveal that these changes to the method can help to more effectively reduce the background humidity in regions of spurious convection, and to avoid a reduction of the background humidity if the profile occurs at the location of a cloud.

Despite the promising results after applying the 1DBay retrieval and comparing the POs to the AROME-France background, the AROME-France analysis fields do not follow the changes in humidity suggested by the FED-derived POs. In particular, the AROME-France analysis includes regions where the humidity expressed as integrated water vapor (IWV) is increased

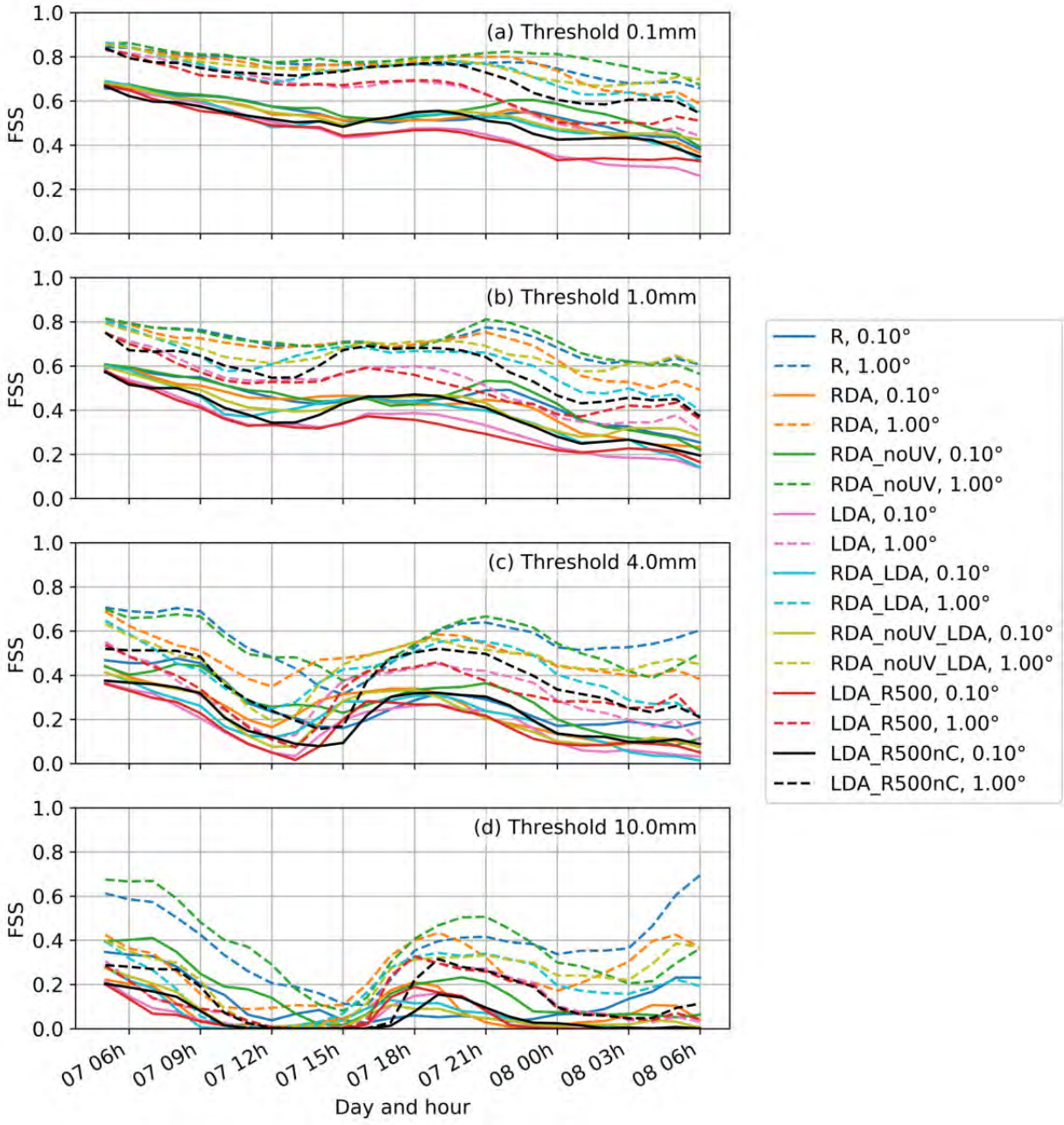


Figure III.4.10: FSS for 6-hour RAs calculated hourly for a sliding window. The colors indicate the model experiments as defined in Section III.4.7. 4 different RA thresholds are used, 0.1 mm (a), 1.0 mm (b), 4.0 mm (c), and 10.0 mm (d). The line style indicates the size of the domain used to calculate the FSS, i.e., 0.1° and 1.0° (as indicated in the legend).

compared to the AROME-France background despite a reduced IWV inferred from combining directly the POs and background (PO IWV) relative to the background IWV. In principle, the analysis should result in a value between the background and the observations. Opposite behavior is found for the LDA in AROME-France. It could be caused by the assumption that the covariance matrix in the AROME-France 3DVar assimilation system is a diagonal matrix assuming that the observation errors are not correlated. However, correlations between observation errors may exist if the distance between two points of observations is small and observations are measured by the same instrument. To address the analysis production issue, thinning of the FED observations might reduce potential correlations between the observation errors and thus help to correct the analysis. Other leads to correct the analysis production may include, e.g., the thorough examination of humidity profiles themselves, as well as other fields such as temperature and pressure that are most likely altered, too, when RH is assimilated. A first step could be to assimilate only one pseudo-RH profile in a region where the AROME-France analysis shows unexpected behavior. The increments in the model variables may reveal the nature of the problem.

The last section presents a case study of a 30-hour forecast that was initiated 07 Oct. 2018, 00:00 UTC. AROME-France experiments with and without both radar data assimilation (RDA) and LDA are conducted. The presented 6-hour rainfall accumulation and the calculated Fractions Skill Scores (FSS) exhibit the lowest forecast skill for the pure LDA model experiments, likely as a consequence of the problems in producing the analysis when FED-derived PO RH profiles are assimilated. Further research will address in particular the analysis production. It will be investigated why the PO RH profiles are currently not assimilated in a proper sense and in fact the AROME-France analysis does not show the desired changes compared to the AROME-France background. As soon as this problem can be solved, the presented model experiments for the 07 Oct. 2018, 00:00 UTC case and potentially further cases will be repeated to investigate the real potential of the LDA in AROME-France.

III.4.12 Discussion: Towards a better FED observation operator

A first FED observation operator is presented in Section III.4.5. This operator constitutes one fairly easy to implement approach. In detail, using the closest AROME-France grid point to the FED pixel center is one method to realize the transformation of the AROME-France model grid to the observation, i.e., pseudo MTG-LI FED, grid (for simplicity only referred to as model grid transformation hereafter). It is also tested to use all AROME-France grid points within each FED pixel and calculate the mean, and to use the point of the maximum graupel mass within each FED pixel. Furthermore, the linear regression between FED and background graupel mass m_g as proxy is just one method to build the observation operator. More sophisticated approaches are examined to relate the FED observation and m_g, i.e., Random Forest (RF), Multilayer Perceptron, linear Support Vector Machine, and Histogram Gradient Boosting machine learning regressors. The most promising one bases on a RF model that uses the AROME-France grid point of maximum m_g within each FED pixel as grid transformation. As no significant improvement was seen between a RF with 20 and a RF with 50 decision trees, the smaller RF is further used as it is faster to compute and takes less memory for stockage.

The evaluation of the observation operator models uses the coefficient of determination,

usually denoted as R^2. It is defined for the prediction $\hat{p}$ and the truth p with n samples as

$$R^2(p, \hat{p}) = 1 - \frac{\sum_{i=1}^{n} (p_i - \hat{p}_i)^2}{\sum_{i=1}^{n} (p_i - \bar{p})^2} \tag{III.4.5}$$

with the predicted value $\hat{p}_i$ of the i-th sample and the corresponding true value p_i, and the mean of all true values $\bar{p}$

The linear regression yields R^2 scores (Section III.4.5) of 0.927 and 0.557 for training (24 days in 2018, 2 per month) and validation (06 and 07 October 2018), respectively. A tested RF that uses the same grid transformation as the previously analyzed linear regression-based observation operator produces R^2 scores of 1.000 and 0.652 for training and validation data, respectively. If the linear regression uses the grid transformation with the point of maximum m_g, R^2 scores of 0.822 (training) and 0.424 (validation) are found. Hence, the maximum m_g grid transformation leads to worse predictions of FED for the linear regression than using the closest AROME-France profile to each FED pixel center. This behavior is different for the RF based observation operator. Here, the R^2 scores for the training (validation) data increase to 1.000 (0.719). Even a training R^2 score of 0.724 is found when keeping the times and geolocations, i.e., a pixel-to-pixel approach, and of 0.907 if data is summed up for the domain but time information is maintained[2]. However, negative R^2 scores for the latter two RF regressors applied to the validation data indicate that these trained RFs suffer from overfitting if any geolocation or time information is maintained.

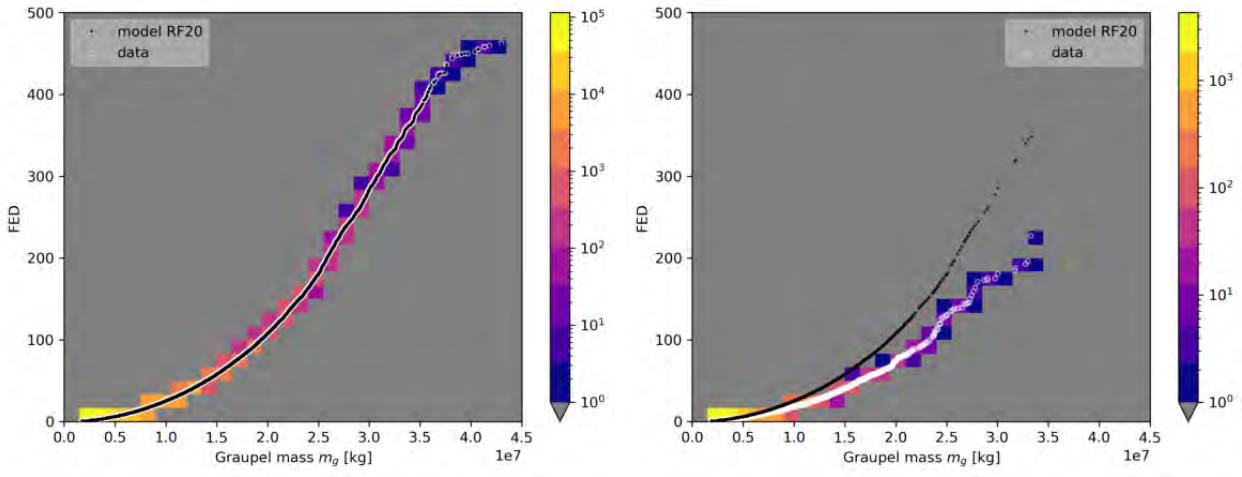
Building the climatological distribution seems to be a necessary step to get reliable observation operator relations. Storm-based data might also yield robust relations (see, e.g., Deierling et al., 2008, Barthe et al., 2010), but were not included in this study (see also Section III.4.5). All approaches trying to maintain a temporal or geolocation, i.e., pixel-to-pixel, information of the data result in overfitting. This result shows that displacements of thunderstorms in the simulation must be taken into account, and that data associated with a certain thunderstorm situation cannot characterize thunderstorms in a general sense. The direct relations between hydrometeor (i.e., graupel mass) and lightning activity (i.e., FED) are particular to each storm. In a climatological sense, however, the FED is well correlated to the graupel mass above the $-5\,°C$ isotherm with Pearson correlation coefficients of 0.96 and 0.92 for our training and validation data, respectively.

Using the mean m_g within each FED pixel for the grid transformation resulted always in lower R^2 scores than the 2 other approaches, i.e., using the closest AROME-France grid point to the FED pixel center or the point of maximum m_g within the FED pixel.

The RF with the maximum m_g grid transformation yields overall the best observation operator as tested for our training and validation data. Figure III.4.11 shows the training (a) and validation (b) of this RF-based observation operator. In comparison to the linear regression (Figure III.4.3), the high FED values are still fit by the RF model during the training. The validation dataset shows, as mention previously, that the observed FED for a given m_g is lower on 06 and 07 Oct. 2018 than for the distribution based on the 24 training days. Nevertheless, the RF-based observation operator with maximum m_g grid transformation can predict the lowest FED for low m_g quite well, and in consequence the R^2 score for the validation dataset is somewhat higher than using the linear regression model using the closest m_g (0.719 versus 0.557, respectively).

It should be mentioned that all observation generators have used linear FED and m_g up to this point. It could be tested whether even more reliable results, especially for the highest FED values, can be obtained using dB(FED).

[2] A similar approach as McCaul et al. (2009), but instead of the peak value in the domain, the mean is used.



(a) (b)

Figure III.4.11: The observed FED distribution versus the distribution of maximum AROME-France graupel mass m_g within each FED observation pixel of size 7 km × 7 km. Grid points with any of FED or m_g equal to zero are not considered. (a) shows the training of a random forest (RF) regressor with 20 decision trees (RF20) for 24 days in 2018, and (b) shows the results of a validation for independent data of 2 additional days in 2018. Colors indicate the number of samples, the white points plot the data points, and black points applied the m_g values in the given RF regressor.

Conclusions

Contents of this Part

Conclusions and perspectives

The launch of the first new Meteosat Third Generation (MTG) satellites is planned in the 2022-time frame. The MTG series will carry, among others, a Lightning Imager (LI) monitoring lightning from a geostationary (GEO) orbit. The GEO lightning observations will provide valuable information about the location and the severity of thunderstorms over a large geographical domain (Europe, Mediterranean Sea, Africa and Atlantic Ocean). The MTG-LI observations will be available also in data sparse regions with, e.g., limited radar coverage such as over mountains and oceans. Therefore, the GEO lightning records are definitely of special interest to numerical weather prediction (NWP). This thesis develops an assimilation method of MTG-LI Flash Extent Density (FED) data in the regional, convective-scale NWP model of Météo-France (AROME-France). The work is structured in three main parts: (i) a theoretical part reviewing lightning physics and observation techniques, (ii) the analysis and intercomparison of existing operational lightning locating systems (LLSs) to eventually generate MTG-LI pseudo-observations, and (iii) the proof of concept of the MTG-LI pseudo-observation assimilation in AROME-France.

The first part introduces key concepts of lightning discharges. Emphasis is put on the signals emitted by lightning and different techniques to detect and locate those signals. The second chapter of this part explains among others the measurement principles of ground-based low frequency (LF) and very high frequency (VHF) networks, as well as the space-based optical lightning detection and location from low Earth orbit (LEO) and GEO orbit. LF, VHF, and optical observations delivered by operational instruments are used in the following part of the thesis.

The first chapter of the second part of the thesis summarizes the intercomparison of operational LLSs and their observations over France. The French LF Meteorage network records are compared to data of both the optical (777.4 nm) satellite Lightning Imaging Sensor (LIS) of the International Space Station (ISS) and the VHF lightning mapping array (LMA) SAETTA within a domain centered on Corsica island in the northwestern Mediterranean Sea. The main results reveal that the ISS-LIS flash detection efficiency (DE) relative to Meteorage degrades for flashes below 7 km of altitude (measured by SAETTA) and for single stroke or pulse Meteorage flashes. It is found that both ISS-LIS and Meteorage DEs increase with longer extent and longer duration of the flashes. Coincident flashes were located with distances and time offset lower than the nominal instrument uncertainties of the coarser instrument, i.e., ISS-LIS. The intercomparison is the subject of the first paper published during the thesis (Erdmann et al., 2020a).

Then, the methodology developed in Erdmann et al. (2020a) is applied to compare observations of the United States (US) National Lightning Detection Network (NLDN) to both ISS-LIS and Geostationary Lightning Mapper (GLM) observations. NLDN uses the same LF Vaisala technology as Meteorage. The comparison to ISS-LIS is conducted to investigate the characteristics of NLDN relative to the same reference LLS, here ISS-LIS, as previously realized with Meteorage. Both LF networks surprisingly agreed well, given the different meteorological and

geographical conditions in the US and in France, in terms of statistics of flash characteristics, with in addition the flash detection relative to ISS-LIS data featuring itself surprisingly similar overall flash statistics in the US and in France. Extending the analysis to GEO lightning observations, the comparison of NLDN and GLM records revealed a significantly higher DE of GLM compared to ISS-LIS DE, with about 87.0 % and 66 % for GLM and ISS-LIS (considering all ISS-LIS events) relative to NLDN, respectively. The statistical distributions of GLM flash characteristics appear within the expected range of the ISS-LIS flash statistics even if the lower number of GLM events per flash compared to its ISS-LIS counterpart is attributed to the larger pixel size within the GLM image than within the ISS-LIS image (about 9 km × 9 km versus about 5 km × 5 km). The consistent flash properties and the relative instrument performances of the studied LLSs confirm the validity to apply a NLDN-GLM-based transfer function on Meteorage observations to simulate realistic GLM-like optical signals over France to develop a lightning-based assimilation scheme in preparation for MTG-LI.

That NLDN-GLM-based transfer function uses machine learning (ML) models trained on coincident GLM and NLDN flashes. The objective is to simulate the GLM targets, i.e., flash extent, flash duration, and number of events per flash based on the LF ground-based lightning observations. The simulated targets are further processed to generate GEO lightning pseudo-observations at a flash level including events. This allows for eventually generating a pseudo-FED product. The methodology and algorithm, referred to as GEO lightning pseudo-observation generator, are detailed in Chapter II.4 and are content of a scientific paper submitted to JTECH (Section II.4.2). The lightning data generator is used to mimic MTG-LI observations over France based on Meteorage records. The thorough analysis conducted on the performances of 196 ML-based algorithms recommends a linear Support Vector Regressor (SVR)-based generator with feature and target scaling, using all available features (i.e., six NLDN flash characteristics) and pseudo-features (i.e., the simulated GLM targets) in a multi-step approach, as detailed in Section II.4.2. Hence, a realistic lightning data generator now exists and has been used to develop the assimilation of (pseudo) MTG-LI data in AROME-France.

The final part of this thesis assimilates pseudo MTG-LI FED in AROME-France. Here, the 1DBay+3DVar assimilation technique (Caumont et al., 2010), that is operationally used for radar reflectivity data, is adapted to lightning FED data. The 1DBay retrieval provides the best estimation of a pseudo-observation (PO) relative humidity (RH) profile for a given FED observation. To relate the FED observations to AROME-France control variables, a FED observation operator is created based on the background graupel mass m_g above the −5 °C isotherm, as suggested by Deierling et al. (2008). The assimilation experiment conducted in this thesis uses a simple linear regression between the FED and m_g distributions. The AROME-France background grid is projected on the observation, i.e., FED grid using the closest AROME-France grid point to each center of an FED pixel. The approach is justified with a strong Pearson correlation coefficient of 0.97 between FED and m_g.

In the following, background AROME-France FED (referred to as AROME_FED) is inferred by applying the observation operator. The AROME_FED can then be compared to the FED observation in the 1DBay retrieval. Doing so, the grid points for the best estimation of the observed FED by the AROME_FED is identified as a linear combination of the background values. The weights are used to retrieve the PO RH profiles that can be assimilated as sounding data. For the first time, a 1DBay retrieval uses FED data. Promising results are obtained as the background integrated water vapor (IWV) is reduced in regions where AROME_FED indicates higher lightning activity than the observed FED. In addition, IWV can be increased using the 1DBay retrieval in regions where the FED observation exceeds the AROME_FED.

When the FED-derived PO RH-profiles are assimilated in AROME-France, however, the analysis changes in the opposite direction as suggested by the PO RH-profiles, especially when the background humidity should be reduced. Since radar data assimilation experiments and a control experiment do not show IWV values as high as those of the Lightning Data Assimilation (LDA) analysis, this is an unexpected result. The high IWV values in the LDA experiments are caused by assimilating the FED-derived RH profiles despite those RH profiles suggest a reduction of the background humidity. Further research will be necessary to study in detail the production of the AROME-France analysis when FED-derived PO RH profiles are assimilated. Despite this problem, an MTG-LI-like lightning data assimilation scheme has been prototyped during the thesis, a first step for deeper studies and tunings of the LDA scheme in the perspective of the operational use of MTG-LI observations for better weather prediction.

Perspectives

The methodology, i.e., flash level clustering and matching, developed in this thesis paves the way to the MTG-LI validation to compare satellite and ground-based LLSs. Locations and times of coincident ground-based and space-based lightning flashes agree usually well, with offsets for both space and time lower than the measurement uncertainty. Nevertheless, the flash level analysis of different LLS types revealed that the corresponding records are often specific to the type and technique of the LLS. For example, LF flash energetics expressed as mean and maximum of the LF currents recorded for the strokes and pulses of one flash are overall similar for LF flashes with and without coincident optical flash observed from space. On another topic, GLM flashes contain more events and extend further than the ISS-LIS ones, which is likely due to the GLM lower energy threshold to record lightning relative to the ISS-LIS one. Hence, even a LEO LLS that passes through the field of view of a GEO LLS does not necessarily record all GEO lightning discharges suggesting that the comparison between LEO lightning and MTG-LI measurements should be conducted with caution by considering lightning physics, the cloud effects, and the observational setups.

Analyses of lightning observations are conducted on the flash level in this thesis because (i) FED data must be generated for assimilation purposes, and (ii) the flash level allows for defining different macroscopic flash characteristics of both LF and optical GEO observed flashes, like flash extent and flash duration, that can be related on the point of view of physics. The idea for a pulse/stroke-group level generator, as suggested in Höller and Betz (2010), could be further investigated. The outputs of such a generator would need to be merged in flashes to produce FED values. However, a pulse/stroke-group level data generator will use only LF currents as physical measurements, besides location and time, to create GEO flashes which might not be sufficient to simulate realistic GLM/MTG-LI-like flashes.

Our flash level GEO lightning pseudo-observation generator can become a valuable tool to provide pseudo MTG-LI for further research activities, preparation of nowcasting applications, and potential training of weather forecasters in the use of the future MTG-LI lightning data. The algorithm might be further refined as discussed in the following:

- In general, this thesis does not separate land and ocean regions. The focus was on continental areas as the majority of the AROME-France domain consists of mainland France. In particular, the US region where the GEO lightning pseudo-observation generator was trained contains solely continental locations. As the AROME-France domain includes parts of the Atlantic Ocean and Mediterranean Sea, a distinct analysis of the GEO lightning pseudo-observation generator over land and over the ocean may indicate whether

this separation could become necessary. There is also the potential issue of different meteorological conditions in the regions of training and application (i.e., USA and France, respectively). In addition, the FED observation operator might use different relations over land and over the ocean as, for example, cloud and hydrometeor characteristics are affected by the surface type. When MTG-LI data will be available, a comparison between observed and simulated MTG-LI observations will show whether the developed observation operator and in consequence 1DBay retrieval need to be adapted.

- Ground-based weather radars and satellite cloud imagers could provide complementary information about cloud structure and properties that could help to better simulate a 2D mapping of optical flashes. Such information about the possible propagation of a flash within the parent cloud, can hardly be inferred from the punctual LF lightning data. Parameters like cloud thickness and cloud top height could then be used as features in the ML-part of the generator. The complementary cloud data can also be useful in cloudy regions without lightning observations, e.g., to locate shallow convection prior to the first electrical discharges.

- The ML part of the GEO lightning pseudo-observation generator could be further refined. First, a good recommendation here would be to evaluate the performance of one of the leading GEO lightning pseudo-observation generators with respect to the electrified area, as the recommended generator applied in this thesis performs well for the FED sum, but systematically underestimates the electrified area. A more extended area of positive FED values might indeed affect the 1DBay retrieval and maybe lead to better LDA performances. The results will also reveal the most important parameter regarding the LDA for tuning the ML models. For example, the applied multistep approach may use the direct output of the previous training and prediction cycle for the subsequent training. On the downside, the improvement of the ML can become time consuming.

- Another comment should be made on the grid of the MTG-LI pseudo-observations. It is known from the GLM records that GEO lightning events feature irregular shapes. In spite of the expected MTG-LI event shapes, our GEO lightning pseudo-observation generator produces data on a regular latitude-longitude grid mainly for computational reasons, with an average size per pixel that can be specified. The algorithm is already capable of producing pseudo GEO lightning on an irregular grid, however, the grid must be defined and given as input.

The second main achievement of the thesis is an LDA scheme for AROME-France. This is the first approach to assimilate lightning data in AROME-France, and there is room for further investigation and improvement. In fact, the most important and first upcoming step should be the in-depth research of the AROME-France analysis when FED-derived RH profiles are assimilated. The analysis increases the IWV compared to the background even if background IWV directly corrected by the FED-derived RH profiles shows a reduction relative to the background IWV. The assimilation of single FED-derived RH profiles in region of unexpected behavior of the analysis will help to identify and solve this issue. Increments of the other model control variables that are also altered when assimilating RH, e.g., temperature and pressure, will support accomplishing this task. A thinning of the FED data and thus an increase in the distance between assimilated FED-derived RH profiles might help to reduce possible observation error correlations that contradict the diagonal observation covariance matrix in the 3DVar system of AROME-France. With a corrected analysis, the actual performance of the LDA can be evaluated to further elaborate the LDA.

Another interesting topic concerns an optimal FED integration period as this optimal period will be a trade-off between a short period around the analysis time to avoid displacement errors and sufficient FED accumulation to illuminate thunderclouds as much as possible. FED for the 1DBay retrieval is currently integrated over 10 min, 5 min before and 5 min after the time of the analysis. As this period is already relatively short, the effect on the LDA of longer FED integration periods up to one hour could be studied. Besides the temporal aspect, the vicinity used to invert FED in the 1DBay retrieval may become a subject of further examination. An optimal size and shape, e.g., excluding clouds, of this vicinity can most certainly help to avoid situations where the FED-derived RH profiles increase the background humidity for a column with observed FED smaller than the background FED. It may be in particular interesting to determine the positions of clouds. An FED equal to zero can be found for both profiles with cloud cover and apart from clouds, whereas the cloud profile is usually more humid than the no-cloud profile. As a result, the retrieved profile can get too humid (dry) if the FED of zero was observed apart from the cloud (inside a cloud), but the 1DBay retrieval finds also FED of zero at location of cloud profiles (of no-cloud profiles). In order to retrieve the estimated RH profile only from the type of profiles as at the point of the FED observation, a cloud mask may be applied.

While preparing the LDA, some ideas were given towards a more sophisticated FED observation operator. Especially a random forest (RF) ML model replacing the linear regression model showed potential for a stronger relationship between FED and the column integrated graupel mass m_g, more robustness to various weather situations, and a better reproduction of the highest FED values. Moreover, other proxy variables may be tested for the FED observation operator. For example, Deierling and Petersen (2008) and Bovalo et al. (2019) found strong relations between the updraft volume and lightning activity, Price and Rind (e.g., 1993), Wong et al. (2013), and Karagiannidis et al. (2019) investigated the cloud top height and lightning, and Deierling et al. (2008) also suggested an ice flux product. Using one of such proxies or a combination may further improve the already relatively strong relations that do, however, underestimate the highest FED values. One might consider in addition to study distinct storm cases and relate the model variables to FED observations (see for example Barthe et al., 2010, Bovalo et al., 2019).

Finally, the assimilation of the retrieved RH profiles may benefit from the change of the AROME-France assimilation system from 3DVar to four-dimensional ensemble-variational (4DEnVar) which is supposed to be operational around 2023. The additional time component of 4DEnVar compared to the current 3DVar system will enable to assimilate humidity profiles related to different phases of the development and dissipation of the occurring thunderstorms and help the model to find the best trajectory to simulate the full thunderstorm life cycle. Hence, it can be expected that the convection will be more efficiently promoted and suppressed as the modeled convective regions are closer to the observation. The ensemble technique will provide a flow-dependent $\mathbf{B}$ matrix. An inclusion of among others hydrometeor specific contents in the control variables of EnVar data assimilation systems will allow to retrieve RH and also other variables like the hydrometeor contents with the 1DBay. A direct assimilation of hydrometeor specific contents (if used to express FED) could even become possible but would need the development of a linear version of the corresponding observation operator and its adjoint. Moreover, with a 3DVar assimilation system, Ge et al. (2013) have shown that the assimilation of hydrometeors has the lowest impact compared to the assimilation of horizontal wind or humidity on the prediction of a thunderstorm. Research will reveal whether cross correlations in the flow-dependent $\mathbf{B}$ matrix enable the model to build consistent environments favorable for convection.

Conclusions et perspectives

Le lancement du premier satellite Meteosat Troisième Génération (MTG) est prévu mi-2022. La série MTG emportera en orbite géostationnaire, entre autres, le détecteur d'éclairs *Lightning Imager* (LI). Ces nouvelles observations spatiales d'éclairs fourniront des informations précieuses sur la localisation et la violence des orages sur un vaste domaine géographique (Europe, Méditerranée, Afrique et océan Atlantique). Les observations MTG-LI seront également disponibles dans les régions où les données météorologiques sont clairsemées comme par exemple celles limitées en couverture radar comme les montagnes ou les océans. Par conséquent, les observations des éclairs depuis l'espace sont certainement d'un intérêt particulier pour la prévision numérique du temps (PNT). Cette thèse développe une méthode d'assimilation des données MTG-LI, et plus spécifiquement de la densité d'étendue des éclairs (*Flash Extent Density* ; FED) dans le modèle opérationnel de PNT à l'échelle convective de Météo-France (AROME-France). Le travail est structuré en trois parties principales : (i) une partie théorique passant en revue la physique des éclairs et des orages, et les techniques d'observation, (ii) l'analyse et l'inter-comparaison d'observations de systèmes opérationnels de localisation des éclairs (Lightning Locating Systems, LLS) existants pour générer in fine des données synthétiques similaires à celles du détecteur MTG-LI, et (iii) la preuve de concept de l'assimilation des données synthétiques de type MTG-LI dans le modèle opérationnel AROME-France.

La première partie du manuscrit présente les concepts clés des décharges électriques. L'accent est mis sur les signaux émis par les éclairs et sur les différentes techniques pour les détecter et les localiser. Le deuxième chapitre de cette partie explique entre autres les principes de mesure des réseaux terrestres basse fréquence (LF, *Low Frequency*) et très haute fréquence (VHF, *Very High Frequency*), ainsi que ceux de la détection optique et de localisation des éclairs depuis l'espace depuis une orbite défilante ou une orbite géostationnaire. Ces différentes observations LF, VHF et optiques sont utilisées par la suite tout au long de la thèse.

Le premier chapitre de la deuxième partie de la thèse résume l'inter-comparaison des détecteurs d'éclairs et leurs observations sur la France. Ainsi les enregistrements du réseau terrestre français Météorage LF sont comparés aux données du détecteur optique (777.4 nm) Lightning Imaging Sensor (LIS) de la Station spatiale internationale (ISS) et du réseau de cartographie des éclairs SAETTA opérant dans le domaine radio VHF. Le réseau SAETTA couvre un domaine centré sur l'île de Corse au nord-ouest de la mer Méditerranée. Un des résultats les plus importants de cette analyse révèle que l'efficacité de détection des éclairs (DE) de l'instrument ISS-LIS par rapport au réseau terrestre Météorage se dégrade pour des éclairs situés à des altitudes inférieures à 7 km (altitudes mesurées par le réseau SAETTA) et pour les éclairs vus par le réseau Météorage comme présentant soit une seule connexion au sol ou une seule composante intra-nuage. On constate aussi que les efficacités de détection des deux instruments ISS-LIS et Météorage augmentent lorsque les éclairs présentent une plus longue étendue et aussi une durée plus longue. Les éclairs coïncidents (donc observés par ces deux instruments) ont été localisés avec des écarts de distance et des décalages temporels inférieurs aux incertitudes nominales de l'instrument le plus grossier, à savoir ISS-LIS. L'inter-comparaison fait l'objet du premier

article publié au cours de la thèse (Erdmann et al., 2020a).

Ensuite, la méthodologie développée par Erdmann et al. (2020a) est appliquée pour comparer les observations du réseau terrestre américain National Lightning Detection Network (NLDN) aux observations des détecteurs spatiaux ISS-LIS et Geostationary Lightning Mapper (GLM). Le réseau américain NLDN utilise la même technologie Vaisala et le même principe de mesure que le réseau français Météorage. Cette comparaison entre observations d'éclairs est menée pour étudier les caractéristiques du réseau terrestre NLDN par rapport à celles de l'instrument spatial ISS-LIS, de la même manière que pour les deux instruments Météorage et ISS-LIS. Les deux réseaux LF présentent étonnamment de très proches propriétés, étant données les conditions météorologiques et géographiques différentes entre les États-Unis et la France, en termes de statistiques des caractéristiques des éclairs et de détection des éclairs par rapport à l'instrument ISS-LIS. En étendant l'analyse aux observations géostationnaires d'éclairs, la comparaison des enregistrements des instruments NLDN et GLM a révélé une efficacité de détection du détecteur GLM par rapport au réseau terrestre NLDN significativement plus élevée (87.0 %) que celle du capteur ISS-LIS en orbite défilante (66 %). Les distributions statistiques des caractéristiques des éclairs détectés par le capteur spatial GLM apparaissent dans la plage attendue des statistiques des éclairs mesurés par le capteur ISS-LIS même si le nombre plus faible de pixels GLM illuminés par éclair par rapport à celui mesuré par l'instrument ISS-LIS est attribué à la taille de pixel plus grande du détecteur GLM que celle du détecteur ISS-LIS (environ $9\,\mathrm{km} \times 9\,\mathrm{km}$ contre environ $5\,\mathrm{km} \times 5\,\mathrm{km}$). La cohérence des propriétés des éclairs et les performances relatives des instruments des différents détecteurs d'éclairs étudiés ici confirment la validité d'appliquer aux observations de Météorage une fonction de transfert construite à partir de la synergie de données NLDN-GLM pour simuler des signaux optiques réalistes, ce qui consitue une première étape afin de développer un schéma d'assimilation des futures données du détecteur spatial MTG-LI.

Pour développer cette fonction de transfert, les caractéristiques d'éclairs simultanément détectés par les instruments GLM et NLDN sont utilisées pour entraîner des modèles d'apprentissage automatique (ML). L'objectif ici est de simuler différents paramètres, appelés cibles (targets), décrivant les éclairs et mesurés par le détecteur spatial GLM, à savoir l'étendue spatiale de l'éclair, la durée de l'éclair et le nombre de pixels illuminés par flash en fonction des observations du réseau terrestre, elles-mêmes appelées caractéristiques (features). Les cibles simulées sont ensuite traitées pour générer des pixels lumineux synthétiques. Cette approche permet ainsi de générer *in fine* un produit synthétique de densité d'étendue des éclairs (FED). La méthodologie et l'algorithme, appelés générateur d'observations synthétiques, sont détaillés au Chapitre II.4 et sont le sujet principal d'un article scientifique soumis au JTECH (Section II.4.2). Ce générateur est donc utilisé pour imiter à partir des enregistrements du réseau terrestre Météorage les observations du futur détecteur MTG-LI au-dessus de la France. L'analyse approfondie menée sur les performances de 196 algorithmes développés durant la thèse recommande d'utiliser un générateur basé sur un modèle de type linear Support Vector Regressor avec normalisation des caractéristiques et des cibles, utilisant toutes les caractéristiques (c.-à-d. les six caractéristiques des éclairs fournies par le réseau terrestre NLDN), et restituant les différentes cibles par une approche en plusieurs étapes, comme détaillé dans la Section II.4.2. Ainsi, un générateur de données d'éclairs réaliste existe maintenant et a été utilisé pour explorer l'assimilation de données synthétiques de MTG-LI dans le modèle opérationnel AROME-France.

La dernière partie de la thèse explore l'assimilation de la densité (synthétique) de FED dans le modèle opérationnel AROME-France. Ici, la technique d'assimilation 1DBay + 3DVar (Caumont et al., 2010), qui est opérationnellement utilisée pour les données de réflectivité radar, est adaptée aux données de FED. La restitution 1DBay fournit la meilleure estimation

de profil d'humidité relative pour une observation de FED donnée. Pour relier les observations de FED aux variables de contrôle AROME-France, un opérateur d'observation de la FED est créé sur la base de la masse de graupel de l'ébauche m_g au-dessus de l'isotherme $-5\,°C$, comme suggéré par Deierling et al. (2008). L'expérience d'assimilation menée dans cette thèse utilise une régression linéaire simple entre les distributions de densité FED et de masse de graupel m_g. La grille d'ébauche AROME-France est projetée sur l'observation, c'est-à dire sur la grille de densité FED en utilisant le point de grille AROME-France le plus proche de chaque centre d'un pixel de FED. L'approche se justifie par un fort coefficient de corrélation de Pearson de 0,97 entre FED et m_g. Dans ce qui suit, l'ébauche de FED d'AROME-France (appelée AROME_FED) est déduite en appliquant l'opérateur d'observation. AROME_FED peut alors être comparé à l'observation FED dans la restitution 1DBay. Ce faisant, la meilleure estimation par AROME_FED de la densité FED observée résulte d'une combinaison linéaire des valeurs de l'ébauche. Les poids sont utilisés pour restituer les profils d'humidité relative qui peuvent être assimilés en tant que données de sondage. Des résultats prometteurs sont obtenus car la vapeur d'eau intégrée de l'ébauche (IWV, *Integrated Water Vapor*) est réduite dans les régions où la FED observée est égale à zéro, alors que AROME_FED présente une activité électrique plus importante que la FED observée. En outre, l'IWV peut être augmentée à l'aide de la restitution 1DBay dans les régions où l'observation FED dépasse AROME_FED.

Cependant, lorsque les profils d'humidité relative dérivés de la densité FED sont assimilés dans AROME-France, l'analyse change dans la direction opposée à celle suggérée par les profils d'humidité relative pseudo-observés, en particulier lorsque l'humidité de l'ébauche doit être réduite. Comme les expériences assimilant les données radar et une expérience de contrôle n'affichent pas des valeurs d'IWV aussi élevées que celles de l'analyse résultant de l'assimilation des données d'éclairs (LDA, pour *Lightning Data Assimilation*), ceci constitue un résultat inattendu. Les fortes valeurs d'IWV dans les expériences de LDA sont dues à l'assimilation de profils d'humdité relative déduites de la FED alors même que ces profils d'humidité relative suggèrent une réduction de l'humidité de l'ébauche. Des recherches supplémentaires seront nécessaires pour étudier en détail la production de l'analyse AROME-France, en particulier lorsque les profils d'humidité relative dérivés de la densité FED observée sont assimilés. Malgré ce problème, un schéma d'assimilation de données d'éclairs de type MTG-LI a été prototypé au cours de la thèse, ce qui constitue une première étape pour des études plus approfondies et des ajustements du schéma LDA dans la perspective de l'utilisation opérationnelle des observations MTG-LI pour une meilleure prévision météorologique.

Perspectives

La méthodologie développée durant cette thèse, c'est-à-dire le regroupement et la mise en correspondance des observations à l'échelle de l'éclair pourra être appliquée lors de la validation du détecteur spatial MTG-LI à partir d'autres observations d'éclairs issues de réseaux terrestres ou d'instruments spatiaux. Les positions et les datations des éclairs coïncidents concordent généralement bien, avec des écarts spatio-temporels inférieurs aux incertitudes de mesure. Néanmoins, l'analyse à l'échelle de l'éclair a révélé que les enregistrements correspondants sont souvent spécifiques au type et à la technique de localisation. Par exemple, les courants de l'éclair mesurés par les techniques de basse fréquence LF, et exprimés tant par leur moyenne par éclair que par leur maximum par éclair sont, à l'échelle de l'éclair, globalement similaires pour les éclairs présentant une signature basse fréquence avec ou sans signal lumineux coïncident mesuré depuis l'espace. Sur un autre sujet, les éclairs mesurés par le détecteur spatial GLM contiennent plus de pixels illuminés et s'étendent sur de plus grandes surfaces que ceux mesurés par le capteur

spatial ISS-LIS, ce qui est probablement dû à des sensibilités différentes des deux capteurs et à des seuils d'acquisition plus bas pour l'instrument GLM. Par conséquent, un détecteur défilant traversant le champ de vue d'un détecteur géostationnaire ne voit pas nécessairement le même éclair de la même façon. Cela suggère que toute comparaison entre mesures spatiales de missions en orbite défilante et de MTG-LI doit être menée avec prudence en considérant la physique des éclairs, les effets des nuages et les méthodes d'observation.

Les analyses réalisées durant la thèse ont été menées à l'échelle de l'éclair car (i) les données de densité FED doivent être générées à des fins d'assimilation, et (ii) des données regroupées à l'échelle de l'éclair permettent de définir différentes caractéristiques macroscopiques des éclairs mesurables tant en fréquence radio que dans le domaine optique, comme l'étendue des éclairs ou encore la durée des éclairs, qui peuvent être davantage reliées du point de vue de la physique. L'idée d'un générateur construit sur les groupes lumineux - ensemble de pixels lumineux adjacents en espace et en temps - et sur les observations ponctuelles des réseaux terrestres basse fréquence LF, comme exploré dans Höller et Betz (2010), pourrait être étudiée plus en avant. Les sorties d'un tel générateur devraient alors être fusionnées pour former des éclairs afin de produire des valeurs de densité FED. Cependant, un tel générateur de données utilisera uniquement les courants mesurés par les réseaux terrestres comme mesure physique, en plus de la position et de la datation, pour créer des éclairs synthétiques, ce qui pourrait ne pas être suffisant pour simuler des éclairs réalistes du type de ceux mesurés par GLM ou MTG-LI.

Notre générateur de données synthétiques à l'échelle de l'éclair est un outil précieux pour fournir des observations de type MTG-LI pour des activités de recherche, pour la préparation d'applications de prévision immédiate, mais aussi pour la formation des prévisionnistes à l'utilisation des futures données de la mission MTG-LI. L'algorithme peut être davantage affiné comme indiqué ci-dessous :

- En général, ce travail de recherche ne sépare pas les régions terrestres des régions maritimes. L'accent a été mis sur les zones continentales car la majorité du domaine AROME-France se compose de la France métropolitaine. En particulier, la région des États-Unis sur laquelle le générateur d'éclairs synthétiques a été entraîné est quasiment continentale. Une analyse distincte des propriétés des éclairs créés par le générateur au-dessus de la terre et de l'océan peut indiquer si une séparation terre-mer est nécessaire. Il y a aussi le possible problème des différentes conditions météorologiques entre la région d'entraînement (les États-Unis) et celle de l'application (la France). Dès que les données MTG-LI seront disponibles, une comparaison entre observations réelles et observations synthétiques montrera si le schéma d'assimilation développé doit être adapté. En outre, l'opérateur d'observation pour la densité FED peut utiliser des relations différentes sur terre et sur mer, car, par exemple, les caractéristiques des nuages et des hydrométéores sont affectées par le type de surface. Lorsque les données MTG-LI seront disponibles, une comparaison entre les observations MTG-LI observées et simulées montrera si l'opérateur d'observation développé et, par conséquent, la restitution 1DBay doivent être adaptés.

- Les radars météorologiques terrestres et les radiomètres sensibles aux nuages pourraient fournir des informations complémentaires sur la structure et les propriétés des nuage afin de mieux simuler en 2D l'extension du signal lumineux émis par les éclairs. Une telle information décrivant la propagation éventuelle d'un éclair au sein du nuage parent, peut difficilement être déduite des données ponctuelles des réseaux LF. Des paramètres tels que l'épaisseur et la hauteur du sommet des nuages pourraient donc être utilisés comme caractéristiques dans la partie ML du générateur. Les données complémentaires sur les nuages peuvent également être utiles dans les régions nuageuses sans éclair observé, par exemple

pour localiser la convection peu profonde avant les premières décharges électriques.

- La partie modèle d'apprentissage automatique du générateur de données synthétiques pourrait être encore affinée. Premièrement, une bonne recommandation ici serait d'évaluer la performance d'un des meilleurs générateurs de pseudo-observations d'éclairs GEO par rapport à la zone électrisée, car le générateur recommandé appliqué dans cette thèse fonctionne bien pour la somme de FED, mais sous-estime systématiquement la zone électrisée. Une zone plus étendue de valeurs de FED positives pourrait en effet affecter la restitution 1DBay et peut-être conduire à de meilleures performances de la LDA. Les résultats révéleront également le paramètre le plus important concernant la LDA pour le réglage des modèles ML. Par exemple, l'approche multi-étapes appliquée peut utiliser les résultats directs du cycle d'entrainement et de prédiction précédent pour l'entraînement suivant. Un inconvénient est que l'amélioration du ML peut prendre beaucoup de temps.

- Un autre commentaire doit être fait sur la grille des données synthétiques du capteur spatial MTG-LI. Il est connu d'après les enregistrements du détecteur spatial GLM que les pixels illuminés présentent des formes irrégulières. Cependant, le générateur développé durant la thèse produit des données sur une grille régulière en latitude et longitude, avec une taille moyenne par pixel qui peut être spécifiée.

La deuxième réalisation principale de la thèse est le développement d'un schéma d'assimilation de données d'éclairs (LDA) pour le modèle opérationnel AROME-France. Il s'agit de la première approche pour assimiler les données électriques dans AROME-France, et il y a de la place pour davantage d'investigations et d'améliorations. En fait, l'étape la plus importante et la plus urgente devrait être un travail de recherche détaillé sur l'analyse AROME-France lorsque les profils d'humidité relative dérivés de la densité FED sont assimilés. L'analyse augmente l'IWV par rapport à l'ébauche alors que l'IWV de l'ébauche directement corrigée par les profils d'humidité relative dérivés de la densié FED montre une réduction de l'IWV par rapport à l'IWV de l'ébauche. Un écrémage des données de densité FED et donc une augmentation de la distance entre les profils d'humidité relative assimilés pourraient contribuer à réduire les éventuelles corrélations d'erreur d'observation qui s'opposent à l'hypothèse d'une matrice de covariance d'observation diagonale dans le système 3DVar d'AROME-France. Avec une correction de l'analyse, les performances réelles du schéma d'assimilation de données d'éclairs LDA pourront être évaluées afin de poursuivre l'élaboration de la LDA.

Un autre sujet intéressant concerne la période optimale d'intégration de la densité FED, car cette période optimale doit être un compromis entre une courte période autour de la période d'analyse pour éviter les erreurs de déplacement et un cumul suffisant de densité FED pour illuminer les nuages d'orage autant que possible. La FED pour la restitution 1DBay est actuellement intégrée sur 10 min, 5 min avant et 5 min après le moment de l'analyse. Comme cette période est déjà relativement courte, l'effet sur le schéma LDA de périodes d'intégration de FED plus longues, jusqu'à une heure, pourrait être étudié. Outre l'aspect temporel, le voisinage utilisé pour inverser la densité FED dans la restitution 1DBay pourrait faire l'objet d'un examen plus approfondi. Une taille et une forme optimales de ce voisinage, par exemple en excluant les nuages, peuvent très certainement aider à éviter les situations où les profils d'humidité relative dérivés de la densité FED augmentent l'humidité de l'ébauche pour une colonne avec une densité FED observée plus petite que la densité FED de l'ébauche. Il peut être particulièrement intéressant de trouver un moyen de déterminer les positions des nuages. Une densité FED égale à zéro peut être trouvée pour des profils avec couverture nuageuse et en ciel clair, alors que le profil du nuage est habituellement plus humide que le profil sans nuage. En

conséquence, le profil restitué peut devenir trop humide (sec) si la FED nulle a été observée en ciel clair (à l'intérieur d'un nuage), mais la restitution 1DBay trouve également une FED nulle dans les profils nuageux (dans les profils en ciel clair). Afin de récupérer le profil d'humidité relative estimé uniquement à partir du type de profil au point d'observation de la FED, un masque de nuage peut être appliqué.

Lors de la préparation du schéma LDA, certaines idées ont été émises en faveur d'un opérateur d'observation plus sophistiqué pour la densité FED. En particulier, un modèle ML de *Random Forrest* (RF) remplaçant le modèle de régression linéaire a montré le potentiel d'une relation plus forte entre la densité FED et la masse de graupel intégrée à la colonne m_g, une plus grande robustesse aux diverses situations météorologiques et une meilleure reproduction des valeurs les plus élevées de la densité FED. En outre, d'autres variables intermédiaires peuvent être testées pour l'opérateur d'observation pour la densité FED. Par exemple, Deierling et Petersen (2008) et Bovalo et al. (2019) ont trouvé des relations fortes entre le volume du courant ascendant et l'activité électrique, Price et Rind (e.g., 1993), Wong et al. (2013) et Karagiannidis et al. (2019) ont étudié la hauteur du sommet des nuages et l'activité électrique, et Deierling et al. (2008) ont également suggéré un produit de flux de glace. L'utilisation de l'un ou l'autre de ces indicateurs ou d'une combinaison de ceux-ci pourrait améliorer encore les relations déjà relativement fortes. On pourrait envisager en outre d'étudier des cas distincts des orages et de relier les variables du modèle aux observations de la densité FED (voir par exemple Barthe et al., 2010, Bovalo et al., 2019).

Enfin, l'assimilation des profils d'humidité relative restitués pourrait bénéficier du changement du système d'assimilation AROME-France de 3DVar à 4DEnVar (quadri-dimensionnel ensembliste-variationnel) qui est censé être opérationnel vers 2023. La composante temporelle supplémentaire du 4DEnVar par rapport au système 3DVar actuel permettra d'assimiler les profils d'humidité liés aux différentes phases de développement et de dissipation des orages qui se produisent et aidera le modèle à trouver la meilleure trajectoire pour simuler le cycle de vie complet d'un orage. On peut donc s'attendre à ce que la convection soit plus efficacement favorisée et supprimée à mesure que les régions convectives modélisées se rapprocheront de l'observation. La technique d'ensemble fournira une matrice **B** dépendant de l'écoulement. L'inclusion, entre autres, de teneurs spécifiques en hydrométéores dans les variables de contrôle des systèmes d'assimilation de données EnVar permettra de restituer l'humidité relative et d'autres variables comme les teneurs spécifiques en hydrométéores avec la restitution 1DBay. Une assimilation directe de la teneur spécifique en hydrométéores (si elle est utilisée pour exprimer la densité FED) pourrait même devenir possible mais nécessiterait le développement d'une version linéaire tangente de l'opérateur d'observation correspondant et son adjoint. En outre, avec un système d'assimilation 3DVar, Ge et al. (2013) ont montré que l'assimilation d'hydrométéores a le plus faible impact en comparaison avec l'assimilation du vent horizontal ou de l'humidité sur la prévision d'un orage. La recherche révélera si les corrélations croisées dans la matrice **B** dépendante de l'écoulement permettent au modèle de construire des environnements cohérents favorables à la convection.

Bibliography

Albrecht, R. I., S. J. Goodman, D. E. Buechler, R. J. Blakeslee, and H. J. Christian, 2016: Where Are the Lightning Hotspots on Earth? *Bulletin of the American Meteorological Society*, **97**11, 2051–2068. ISSN: 0003-0007. DOI: 10.1175/BAMS-D-14-00193.1 (cit. on p. 15).

Allen, B. J., E. R. Mansell, D. C. Dowell, and W. Deierling, 2016: Assimilation of pseudo-GLM data using the ensemble Kalman filter. *Monthly Weather Review*, **144**, 3465–3486. DOI: 10.1175/MWR-D-16-0117.1 (cit. on pp. 4, 8, 220, 224).

Anthes, R. A., 1974: Data Assimilation and Initialization of Hurricane Prediction Models. *Journal of the Atmospheric Sciences*, **31**3, 702–719. ISSN: 0022-4928. DOI: 10.1175/1520-0469(1974)031<0702:DAAIOH>2.0.CO;2 (cit. on p. 203).

Apodaca, K., M. Zupanski, M. DeMaria, J. A. Knaff, and L. D. Grasso, 2014: Development of a hybrid variational-ensemble data assimilation technique for observed lightning tested in a mesoscale model. *Nonlin. Processes Geophys.*, **21**, 1027–1041. DOI: 10.5194/npg-21-1027-2014 (cit. on p. 220).

Avila, E., G. Caranti, N. Castellano, and C. Saunders, 1998: Laboratory studies of the influence of cloud droplet size on charge transfer during crystal-graupel collisions. *Journal of Geophysical Research: Atmospheres*, **103**D8, 8985–8996. DOI: 10.1029/97JD03115 (cit. on p. 20).

Avila, E. E., R. E. Bürgesser, N. E. Castellano, A. B. Collier, R. H. Compagnucci, and A. R. Hughes, 2010: Correlations between deep convection and lightning activity on a global scale. *Journal of Atmospheric and Solar-Terrestrial Physics*, **72**14, 1114 –1121. ISSN: 1364-6826. DOI: https://doi.org/10.1016/j.jastp.2010.07.019 (cit. on pp. 15, 51).

Babich, L., E. Bochkov, and T. Neubert, 2016: The role of charged ice hydrometeors in lightning initiation. *Journal of Atmospheric and Solar-Terrestrial Physics*, **154**. DOI: 10.1016/j.jastp.2016.12.010 (cit. on p. 27).

Baker, B., M. B. Baker, E. R. Jayaratne, J. Latham, and C. P. R. Saunders, 1987: The influence of diffusional growth rates on the charge transfer accompanying rebounding collisions between ice crystals and soft hailstones. *Quarterly Journal of the Royal Meteorological Society*, **113**478, 1193–1215. DOI: 10.1002/qj.49711347807 (cit. on p. 20).

Bannister, R. N., 2017: A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **143**703, 607–633. DOI: 10.1002/qj.2982 (cit. on pp. 203–204, 206).

Barthe, C., W. Deierling, and M. C. Barth, 2010: Estimation of total lightning from various storm parameters: a cloud-resolving model study. *Journal of Geophysical Research: Atmospheres*, **115**D24. DOI: 10.1029/2010JD014405 (cit. on pp. 225, 244, 253, 260).

Bateman, M., 2013: A high-fidelity proxy dataset for the Geostationary Lightning Mapper (GLM). In: Austin, TX, USA. AMS Sixth Conference on the Meteorological Application of Lightning Data, 725 (cit. on pp. 4, 9).

Bechtold, P., S. Belair, P. Bougeault, J. Carrière, J.Cuxart, V. Ducrocq, C. Fischer, M. Georgelin, P. Héreil, J. Lafore, C. Liousse, C. Mari, I. Mallet, P. J. Mascart, V. Masson, J. Pinty, E. Richard, K. Suhre, J.Stein, P. Tulet, J. V.-G. de Arellano, D. Barbary, Y. Bouteloup, O. Caumont, J.-P. Chaboureau, F. Couvreux, O. Geoffroy, O. Nuissier, C. Lac, T. Mari'c, P. L. Moigne, M. Leriche, J. Pergaud, F. Solmon, G. Tanguy, and M. Tomasini, 2011: *The Meso-NH Atmospheric Simulation System: Scientific Documentation-Part III: Physics*. Scientific Documentation Meso-NH Masdev4.9. Meteo-France, Centre National de la Recherche Scientifique (cit. on pp. 210–211).

Betz, H. D., K. Schmidt, P. Laroche, P. Blanchet, W. P. Oettinger, E. Defer, Z. Dziewit, and J. Konarski, 2009: LINET—an international lightning detection network in europe. *Atmospheric Research*, **91**2. 13th International Conference on Atmospheric Electricity, 564–573. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2008.06.012 (cit. on p. 224).

Bitzer, P. M., J. C. Burchfield, and H. J. Christian, 2016: A bayesian approach to assess the performance of lightning detection systems. *Journal of Atmospheric and Oceanic Technology*, **33**, 563–578. DOI: 10.1175/JTECH-D-15-0032.1 (cit. on p. 95).

Bitzer, P. M., H. J. Christian, M. Stewart, J. Burchfield, S. Podgorny, D. Corredor, J. Hall, E. Kuznetsov, and V. Franklin, 2013: Characterization and applications of vlf/lf source locations from lightning using the huntsville alabama marx meter array. *Journal of Geophysical Research: Atmospheres*, **118**8, 3120–3138. DOI: 10.1002/jgrd.50271 (cit. on pp. 40–41).

Bjerknes, V., 2009: The problem of weather prediction, considered from the viewpoints of mechanics and physics. *Meteorologische Zeitschrift*, **18**6, 663–667. DOI: 10.1127/0941-2948/2009/416 (cit. on p. 199).

Blakeslee, R., D. Mach, and K. Virts, 2019. In: July 12-16, Atlanta, Ge, USA. ISS R&D conference (cit. on p. 15).

Blakeslee, R. J., T. J. Lang, W. J. Koshak, D. Buechler, P. Gatlin, D. M. Mach, G. T. Stano, K. S. Virts, T. D. Walker, D. J. Cecil, W. Ellett, S. J. Goodman, S. Harrison, D. L. Hawkins, M. Heumesser, H. Lin, M. Maskey, C. J. Schultz, M. Stewart, M. Bateman, O. Chanrion, and H. Christian, 2020: Three years of the lightning imaging sensor onboard the international space station: expanded global coverage and enhanced applications. *Journal of Geophysical Research: Atmospheres*, **125**16. e2020JD032918 2020JD032918, e2020JD032918. DOI: https://doi.org/10.1029/2020JD032918 (cit. on pp. 79, 102).

Blaylock, B. K. and J. D. Horel, 2020: Comparison of lightning forecasts from the high-resolution rapid refresh model to geostationary lightning mapper observations. *Weather and Forecasting*, **35**2, 401–416. DOI: 10.1175/WAF-D-19-0141.1 (cit. on p. 97).

Blayo, E., E. Cosmel, M. Nodet, and A. Vidard, 2011: *Introduction to Data Assimilation*. Last Accessed on 2020-08-13 (cit. on pp. 203–206).

Bluestein, H. B., 2014: *Severe Convective Storms and Tornadoes - Observations and Dynamics*. 1st ed. Springer Heidelberg New York Dordrecht London: Springer-Verlag Berlin Heidelberg. ISBN: 978-3-642-05380-1. DOI: 10.1007/978-3-642-05381-8 (cit. on pp. 14, 16–17).

Boccippio, D. J., W. Koshak, R. Blakeslee, K. Driscoll, D. Mach, D. Buechler, W. Boeck, H. J. Christian, and S. J. Goodman, 2000: The optical transient detector (otd): instrument characteristics and cross-sensor validation. *Journal of Atmospheric and Oceanic Technology*, **17**4, 441–458. DOI: 10.1175/1520-0426(2000)017<0441:TOTDOI>2.0.CO;2 (cit. on p. 95).

Borderies, M., O. Caumont, C. Augros, E. Bresson, J. Delanoë, V. Ducrocq, N. Fourrié, T. L. Bastard, and M. Nuret, 2018: Simulation of w-band radar reflectivity for model validation and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **144**711, 391–403. DOI: 10.1002/qj.3210 (cit. on pp. 214, 227).

Borderies, M., O. Caumont, J. Delanoë, V. Ducrocq, N. Fourrié, and P. Marquet, 2019: Impact of airborne cloud radar reflectivity data assimilation on kilometre-scale numerical weather prediction analyses and forecasts of heavy precipitation events. *Natural Hazards and Earth System Sciences*, **19**4, 907–926. DOI: 10.5194/nhess-19-907-2019 (cit. on pp. 213, 227, 230).

Borderies, M., 2018: *Assimilation de données de radar à nuages aéroporté pendant la campagne de mesures HyMeX*. at Météo-France. PhD thesis. Institut National Polytechnique de Toulouse (Toulouse INP): Sciences de l'Univers de l'Environnement et de l'Espace (SDUEE) (cit. on p. 209).

Bouttier, F. and P. Courtier, 2002: *Data Assimilation Concepts and Methods* (cit. on pp. 203–204).

Bovalo, C., C. Barthe, and J.-P. Pinty, 2019: Examining relationships between cloud-resolving model parameters and total flash rates to generate lightning density maps. *Quarterly Journal of the Royal Meteorological Society*, **145**720, 1250–1266. DOI: 10.1002/qj.3494 (cit. on pp. 4, 8, 225, 253, 260).

Brooks, I., C. Saunders, R. Mitzeva, and S. Peck, 1997: The effect on thunderstorm charging of the rate of rime accretion by graupel. *Atmospheric Research*, **43**3, 277 –295. ISSN: 0169-8095. DOI: https://doi.org/10.1016/S0169-8095(96)00043-9 (cit. on pp. 4, 8, 20).

Brousseau, P., G. Desroziers, F. Bouttier, and B. Chapnik, 2014: A posteriori diagnostics of the impact of observations on the arome-france convective-scale data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **140**680, 982–994. DOI: 10.1002/qj.2179 (cit. on pp. 4, 8, 211–212, 216, 230).

Brousseau, P., Y. Seity, D. Ricard, and J. Léger, 2016: Improvement of the forecast of convective activity from the AROME-France system. *Quarterly Journal of the Royal Meteorological Society*, **142**699, 2231–2243. ISSN: 0035-9009. DOI: 10.1002/qj.2822 (cit. on pp. 4, 8, 209, 211–212, 222).

Bruning, E. C., C. E. Tillier, S. F. Edgington, S. D. Rudlosky, J. Zajic, C. Gravelle, M. Foster, K. M. Calhoun, P. A. Campbell, G. T. Stano, C. J. Schultz, and T. C. Meyer, 2019: Meteorological imagery for the geostationary lightning mapper. *Journal of Geophysical Research: Atmospheres*, **124**24. 10.1029/2019JD030874, 14285–14309. DOI: 10.1029/2019JD030874 (cit. on p. 96).

Buiat, M., F. Porcu, and S. Dietrich, 2017: Observing relationships between lightning and cloud profiles by means of a satellite-borne cloud radar. *Atmospheric Measurement Techniques*, **10**, 221. DOI: 10.5194/amt-10-221-2017 (cit. on pp. 4, 8, 224).

Canosa, E. F. and R. List, 1993: Measurements of inductive charges during drop breakup in horizontal electric fields. *Journal of Geophysical Research: Atmospheres*, **98**D2, 2619–2626. DOI: 10.1029/92JD02389 (cit. on p. 24).

Caumont, O., V. Ducrocq, E. Wattrelot, G. Jaubert, and S. 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 (cit. on pp. 4, 212–213, 227, 230, 241, 250, 256).

Cecil, D. J., D. E. Buechler, and R. J. Blakeslee, 2014: Gridded lightning climatology from trmm-lis and otd: dataset description. *Atmospheric Research*, **135-136**, 404 –414. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2012.06.028 (cit. on p. 14).

Chen, Y., Z. Yu, W. Han, J. He, and M. Chen, 2020: Case study of a retrieval method of 3d proxy reflectivity from FY-4A lightning data and its impact on the assimilation and forecasting for severe rainfall storms. *Remote Sensing*, **127**, 1165. ISSN: 2072-4292. DOI: 10.3390/rs12071165 (cit. on p. 221).

Chmielewski, V. C., E. C. Bruning, and B. C. Ancell, 2018: Variations of thunderstorm charge structures in west texas on 4 june 2012. *Journal of Geophysical Research: Atmospheres*, **123**17, 9502–9523. DOI: 10.1029/2018JD029006 (cit. on p. 27).

Christian, H., C. R. Holmes, J. W. Bullock, W. Gaskell, A. J. Illingworth, and J. Latham, 1980: Airborne and ground-based studies of thunderstorms in the vicinity of langmuir laboratory. *Quarterly Journal of the Royal Meteorological Society*, **106**447, 159–174. DOI: 10.1002/qj.49710644711 (cit. on p. 23).

Christian, H. J., R. J. Blakeslee, D. J. Boccippio, W. L. Boeck, D. E. Buechler, K. T. Driscoll, S. J. Goodman, J. M. Hall, W. J. Koshak, D. M. Mach, and M. F. Stewart, 2003: Global frequency and distribution of lightning as observed from space by the Optical Transient Detector. *Journal of Geophysical Research: Atmospheres*, **108**D1, ACL 4–1–ACL 4–15. DOI: 10.1029/2002JD002347 (cit. on p. 14).

Coleman, L. M., M. Stolzenburg, T. C. Marshall, and M. Stanley, 2008: Horizontal lightning propagation, preliminary breakdown, and electric potential in new mexico thunderstorms. *Journal of Geophysical Research: Atmospheres*, **113**D9. DOI: 10.1029/2007JD009459 (cit. on p. 28).

Coquillat, S., E. Defer, P. de Guibert, D. Lambert, J.-P. Pinty, V. Pont, S. Prieur, R. J. Thomas, P. R. Krehbiel, and W. Rison, 2019: SAETTA: high-resolution 3-d mapping of the total lightning activity in the mediterranean basin over corsica, with a focus on a mesoscale convective system event. *Atmospheric Measurement Techniques*, **12**11, 5765–5790. DOI: 10.5194/amt-12-5765-2019 (cit. on pp. 3, 7).

Courtier, P., C. Freydier, J.-F. Geleyn, F. Rabier, and M. Rochas, 1991: The ARPEGE project at Meteo France. In: *Seminar on Numerical Methods in Atmospheric Models, 9-13 September 1991*. Vol. II. ECMWF. Shinfield Park, Reading: ECMWF, 193–232 (cit. on p. 211).

Cummins, K. L. and M. J. Murphy, 2009: An overview of lightning locating systems: history, techniques, and uses, with an in-depth look at the u.s. nldn. *IEEE Trans. Electromag. Compat.*, **51(3)**, 499–518. DOI: 10.1109/TEMC.2009.2023450 (cit. on pp. 3, 7, 35, 37, 39–40).

Dash, J. G., B. L. Mason, and J. S. Wettlaufer, 2001: Theory of charge and mass transfer in ice-ice collisions. *Journal of Geophysical Research: Atmospheres*, **106**D17, 20395–20402. DOI: 10.1029/2001JD900109 (cit. on p. 20).

Davies, H. C., 1976: A lateral boundary formulation for multi-level prediction models. *Quarterly Journal of the Royal Meteorological Society*, **102**432, 405–418. DOI: 10.1002/qj.49710243210 (cit. on pp. 211, 222).

Davies, H. C. and R. E. Turner, 1977: Updating prediction models by dynamical relaxation: an examination of the technique. *Quarterly Journal of the Royal Meteorological Society*, **103**436, 225–245. DOI: 10.1002/qj.49710343602 (cit. on p. 203).

Defer, E. and the EXAEDRE Team, 2019: *The EXAEDRE campaign for a better understanding of the microphysical, dynamical and electrical processes in thunderstorms*. 12th HyMeX Workshop (cit. on p. 237).

Deierling, W. and W. A. Petersen, 2008: Total lightning activity as an indicator of updraft characteristics. *Journal of Geophysical Research*, **113**, D16210. DOI: 10.1029/2007JD009598 (cit. on pp. 4, 8, 224, 253, 260).

Deierling, W., W. A. Petersen, J. Lathman, S. Ellis, and H. J. Christian, 2008: The relationship between lightning activity and ice fluxes in thunderstorms. *Journal of Geophysical Research*, **113**, D15210. DOI: 10.1029/2007JD009700 (cit. on pp. 4, 8, 224–225, 241, 244, 250, 253, 257, 260).

Desroziers, G., J.-T. Camino, and L. Berre, 2014: 4DEnVar: link with 4d state formulation of variational assimilation and different possible implementations. *Quarterly Journal of the Royal Meteorological Society*, **140**684, 2097–2110. DOI: 10.1002/qj.2325 (cit. on p. 206).

Devlin, J. P., 2011: Relating the current science of ion-defect behavior in ice to a plausible mechanism for directional charge transfer during ice particle collisions. *Phys. Chem. Chem. Phys.*, **13** (44), 19707–19713. DOI: 10.1039/C1CP21593K (cit. on p. 23).

Dixon, K., C. Mass, G. Hakim, and R. Holzworth, 2016: The impact of lightning data assimilation on deterministic and ensemble forecasts of convective events. *Journal of Atmospheric and Oceanic Technology*, **33**. DOI: 10.1175/JTECH-D-15-0188.1 (cit. on p. 220).

Dobber, M. and J. Grandell, 2014: Meteosat Third Generation (MTG) Lightning Imager (LI) Instrument Performance and Calibration from User Perspective. In: InProceedings of the 23rd Conference onCharacterization and Radiometric Calibration for Remote Sensing (CALCON), 11-14 August 2014,Utah State University, Logan, Utah, USA. 13 pages (cit. on pp. 3, 7).

Dotzek, N. and C. Price, 2009: Lightning characteristics of extreme weather events. In: *Lightning : Principles, Instruments and Applications*. Ed. by U. S. H.D. Betz and P. Laroche. Springer Publications. Chap. 22, 487–508 (cit. on p. 18).

Ducrocq, V., D. Ricard, J.-P. Lafore, and F. Orain, 2002: Storm-scale numerical rainfall prediction for five precipitating events over france: on the importance of the initial humidity field. *Weather and Forecasting*, **17**6, 1236 –1256. DOI: 10.1175/1520-0434(2002)017<1236:SSNRPF>2.0.CO;2 (cit. on p. 213).

Duruisseau, F., P. Chambon, E. Wattrelot, M. Barreyat, and J.-F. Mahfouf, 2019: Assimilating cloudy and rainy microwave observations from SAPHIR on board Megha Tropiques within the ARPEGE global model. *Quarterly Journal of the Royal Meteorological Society*, **145**719, 620–641. DOI: 10.1002/qj.3456 (cit. on p. 227).

Dwyer, J. R. and M. A. Uman, 2014: The physics of lightning. *Physics Reports*, **534**4. The Physics of Lightning, 147 –241. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2013.09.004 (cit. on pp. 25, 27–29).

Emersic, C. and C. Saunders, 2020: The influence of supersaturation at low rime accretion rates on thunderstorm electrification from field-independent graupel-ice crystal collisions. *Atmospheric Research*, **242**, 104962. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2020.104962 (cit. on p. 20).

Erdmann, F., E. Defer, O. Caumont, R. J. Blakeslee, S. Pédeboy, and S. Coquillat, 2020: Concurrent satellite and ground-based lightning observations from the optical lightning imaging sensor (ISS-LIS), the low-frequency network meteorage and the SAETTA lightning mapping array (LMA) in the northwestern Mediterranean region. *Atmospheric Measurement Techniques*, **13**2, 853–875. DOI: 10.5194/amt-13-853-2020 (cit. on pp. 79–81, 83, 85, 91, 93–95, 98, 197, 222, 249, 256).

Erdmann, F., O. Caumont, and E. Defer, 2020: "A geostationary lightning pseudo-observation generator utilizing low frequency ground-based lightning observations". submitted to Journal of Atmospheric and Oceanic Technology as JTECH-D-20-0160 (cit. on pp. 111, 222).

Faggian, N., B. Roux, P. Steinle, and B. Ebert, 2015: Fast calculation of the Fractions Skill Score. *Mausam*, **66**, 457–466 (cit. on p. 237).

Fierro, A. O., J. Gao, C. L. Ziegler, K. M. Calhoun, E. R. Mansell, and D. R. MacGorman, 2016: 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. DOI: 10.1175/MWR-D-16-0053.1 (cit. on pp. 220, 222, 225).

Fierro, A. O., J. Gao, C. L. Ziegler, E. R. Mansell, D. R. MacGorman, and S. R. Dembrek, 2014: Evaluation of a cloud-scale lightning data assimilation technique and a 3dvar method for the analysis and short-term forecast of the 29 June 2012 derecho event. *Monthly Weather Review*, **142**, 183–202. DOI: 10.1175/MWR-D-13-00142.1 (cit. on p. 219).

Fierro, A. O., E. R. Mansell, C. L. Ziegler, and D. R. MacGorman, 2012: Application of a lightning data assimilation technique in the WRF-ARW model at cloud-resolving scales for the tornado outbreak of 24 May 2011. *Monthly Weather Review*, **140**, 2609–2627. DOI: 10.1175/MWR-D-11-00299.1 (cit. on pp. 219–220, 224, 226).

Fierro, A. O., Y. Wang, J. Gao, and E. R. Mansell, 2019: 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. DOI: 10.1175/MWR-D-18-0421.1 (cit. on pp. 4, 8, 97, 216, 220–221, 226, 228).

Fischer, C., F. Bouyssel, P. Brousseau, R. El Khatib, P. Pottier, Y. Seity, E. Wattrelot, and A. Joly, 2018: Les modèles opérationnels de prévision numérique à aire limitée de Météo-France. *La Météorologie*, **100**. Numéro Spécial Anniversaire 25 ans, 18–28. DOI: 10.4267/2042/65139 (cit. on pp. 211, 222).

Fischer, C., T. Montmerle, L. Berre, L. Auger, and S. E. Stafanescu, 2005: An overview of the variational assimilation in the aladin/france numerical weather-prediction system. *Quarterly Journal of the Royal Meteorological Society*, **131**613, 3477–3492. DOI: 10.1256/qj.05.115 (cit. on p. 211).

Formenton, M., G. Panegrossi, D. Casella, S. Dietrich, A. Mugnai, P. Sano, F. Di Paolo, H.-D. Betz, C. Price, and Y. Yair, 2013: Using a cloud electrification model to study relationships between lightning activity and cloud microphysical structure. *Natural Hazards and Earth System Sciences*, **13**, 1085–1104. DOI: 10.5194/nhess-13-1085-2013 (cit. on p. 225).

Franklin, B., 1752: Xcv. a letter of benjamin franklin, esq; to mr. peter collinson, f. r. s. concerning an electrical kite. *Philosophical Transactions of the Royal Society of London*, **47**, 565–567. DOI: 10.1098/rstl.1751.0096 (cit. on pp. 3, 7).

Fuchs, B. R., S. A. Rutledge, E. C. Bruning, J. R. Pierce, J. K. Kodros, T. J. Lang, D. R. MacGorman, P. R. Krehbiel, and W. Rison, 2015: Environmental controls on storm intensity and charge structure in multiple regions of the continental united states. *Journal of Geophysical Research: Atmospheres*, **120**13, 6575–6596. DOI: 10.1002/2015JD023271 (cit. on p. 27).

Gallin, L.-J., T. Farges, R. Marchiano, F. Coulouvrat, E. Defer, W. Rison, W. Schulz, and M. Nuret, 2016: Statistical analysis of storm electrical discharges reconstituted from a lightning mapping system, a lightning location system, and an acoustic array. *Journal of Geophysical Research: Atmospheres*, **121**8, 3929–3953. DOI: 10.1002/2015JD023745 (cit. on p. 47).

Ge, G., J. Gao, and M. Xue, 2013: Impacts of Assimilating Measurements of Different State Variables with a Simulated Supercell Storm and Three-Dimensional Variational Method. *Monthly Weather Review*, **141**8, 2759–2777. ISSN: 0027-0644. DOI: 10.1175/MWR-D-12-00193.1 (cit. on pp. 213, 253, 260).

Giannaros, T. M., V. Kotroni, and K. Lagouvardos, 2015: Predicting lightning activity in greece with the weather research and forecasting (WRF) model. *Atmospheric Research*, **156**, 1 – 13. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2014.12.009 (cit. on p. 224).

— 2016: WRF-LTNGDA: a lightning data assimilation technique implemented in the WRF model for improving precipitation forecasts. *Environmental Modelling & Software*, **76**, 54 –68. ISSN: 1364-8152. DOI: https://doi.org/10.1016/j.envsoft.2015.11.017 (cit. on p. 219).

Goodman, S. J., R. J. Blakeslee, W. J. Koshak, D. Mach, J. Bailey, D. Buechler, L. Carey, C. Schultz, M. Bateman, E. McCaul, and G. Stano, 2013: The GOES-R Geostationary Lightning Mapper (GLM). *Atmospheric Research*, **125-126**, 34 –49. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2013.01.006 (cit. on pp. 3, 7, 217).

Grenet, G., 1947: Essai d'explication de la charge éléctrique des nuages d'orages. *Ann. Geophys.*, **3**. For English translation see Atmos. Res. 30, 176-179, 306–307 (cit. on p. 24).

Helsdon Jr., J. H., S. Gattaleeradapan, R. D. Farley, and C. C. Waits, 2002: An examination of the convective charging hypothesis: charge structure, electric fields, and maxwell currents. *Journal of Geophysical Research: Atmospheres*, **107**D22, ACL 9–1–ACL 9–26. DOI: 10.1029/2001JD001495 (cit. on p. 25).

Herzog, B. S., K. M. Calhoun, and D. R. MacGorman, 2014: *Total Lightning Information in a 5-Year Thunderstorm Climatology*. XV International Conference on Atmospheric Electricity, Norman, OK, USA. accessed 22 May 2020 (cit. on p. 96).

Hill, J. D., M. A. Uman, and D. M. Jordan, 2011: High-speed video observations of a lightning stepped leader. *Journal of Geophysical Research: Atmospheres*, **116**D16. DOI: 10.1029/2011JD015818 (cit. on p. 29).

Höller, H. and H.-D. Betz, 2010: *Study on Inter-comparison of LIS and Ground-Based Lightning Location System Observations*. Report ITT No. 09/996. EUMETSAT, Deutsches Zentrum für Luft- und Raumfahrt (cit. on pp. 4, 9, 196–197, 251, 258).

Hoke, J. E. and R. A. Anthes, 1976: The Initialization of Numerical Models by a Dynamic-Initialization Technique. *Monthly Weather Review*, **104**12, 1551–1556. ISSN: 0027-0644. DOI: 10.1175/1520-0493(1976)104<1551:TIONMB>2.0.CO;2 (cit. on p. 203).

Holle, R. L., K. L. Cummins, and W. A. Brooks, 2016: Seasonal, Monthly, and Weekly Distributions of NLDN and GLD360 Cloud-to-Ground Lightning. *Monthly Weather Review*, **144**8, 2855–2870. ISSN: 0027-0644. DOI: 10.1175/MWR-D-16-0051.1 (cit. on p. 14).

Howard, J., M. A. Uman, J. R. Dwyer, D. Hill, C. Biagi, Z. Saleh, J. Jerauld, and H. K. Rassoul, 2008: Co-location of lightning leader x-ray and electric field change sources. *Geophysical Research Letters*, **35**13. DOI: 10.1029/2008GL034134 (cit. on p. 29).

Hu, J., A. Fierro, Y. Wang, J. Gao, and E. R. Mansell, 2020: An Evaluation of the Impact of Assimilating GLM-Observed Total Lightning Data on Short-Term Forecasts of High-Impact Convective Events. In: January 12-16, Boston, Ma, USA. 100th AMS Annual Meeting (cit. on pp. 4, 8).

Hu, J., A. O. Fierro, Y. Wang, J. Gao, and E. R. Mansell, 2020: Exploring the assimilation of glm-derived water vapor mass in a cycled 3dvar framework for the short-term forecasts of high-impact convective events. *Monthly Weather Review*, **148**3, 1005–1028. DOI: 10.1175/MWR-D-19-0198.1 (cit. on pp. 97, 221–222, 225–226).

Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Unified notation for data assimilation : operational, sequential and variational. *Journal of the Meteorological Society of Japan. Ser. II*, **75**1B, 181–189. DOI: 10.2151/jmsj1965.75.1B_181 (cit. on p. 201).

Jayaratne, E. R. and C. P. R. Saunders, 2016: The interaction of ice crystals with hailstones in wet growth and its possible role in thunderstorm electrification. *Quarterly Journal of the Royal Meteorological Society*, **142**697, 1809–1815. DOI: 10.1002/qj.2777 (cit. on p. 20).

Johns, R. H. and W. D. Hirt, 1987: Derechos: widespread convectively induced windstorms. *Weather and Forecasting*, **2**1, 32 –49. DOI: 10.1175/1520-0434(1987)002<0032:DWCIW>2.0.CO;2 (cit. on p. 17).

Jungwirth, P., D. Rosenfeld, and V. Buch, 2005: A possible new molecular mechanism of thundercloud electrification. *Atmospheric Research*, **76**1. Atmospheric Electricity, 190 –205. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2004.11.016 (cit. on p. 23).

Karagiannidis, A., K. Lagouvardos, S. Lykoudis, V. Kotroni, T. Giannaros, and H.-D. Betz, 2019: Modeling lightning density using cloud top parameters. *Atmospheric Research*, **222**, 163 –171. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2019.02.013 (cit. on pp. 224, 253, 260).

Kong, R., M. Xue, A. O. Fierro, Y. Jung, C. Liu, E. R. Mansell, and D. R. MacGorman, 2020: Assimilation of GOES-R Geostationary Lightning Mapper Flash Extent Density Data in GSI EnKF for the Analysis and Short-Term Forecast of a Mesoscale Convective System. *Monthly Weather Review*, **148**5, 2111–2133. ISSN: 0027-0644. DOI: 10.1175/MWR-D-19-0192.1 (cit. on pp. 221, 225).

Koshak, W. J., 2010: Optical Characteristics of OTD Flashes and the Implications for Flash-Type Discrimination. *Journal of Atmospheric and Oceanic Technology*, **27**11, 1822–1838. ISSN: 0739-0572. DOI: 10.1175/2010JTECHA1405.1 (cit. on p. 47).

Kraus, H., 2004: *Die Atmosphäre - Eine Einführung in die Meteorologie*. 3rd ed. Springer-Verlag Berlin Heidelberg: Springer-Verlag Berlin Heidelberg. ISBN: 3-540-20656-6 (cit. on p. 18).

Krishnamurti, T. N., H. S. Bedi, W. Heckley, and K. Ingles, 1988: Reduction of the Spinup Time for Evaporation and Precipitation in a Spectral Model. *Monthly Weather Review*, **116**4, 907–920. ISSN: 0027-0644. DOI: 10.1175/1520-0493(1988)116<0907:ROTSTF>2.0.CO;2 (cit. on p. 203).

Kummerow, C., Y. Hong, W. S. Olson, S. Yang, R. F. Adler, J. McCollum, R. Ferraro, G. Petty, D.-B. Shin, and T. T. Wilheit, 2001: The Evolution of the Goddard Profiling Algorithm (GPROF) for Rainfall Estimation from Passive Microwave Sensors. *Journal of Applied Meteorology*, **40**11, 1801–1820. ISSN: 0894-8763. DOI: 10.1175/1520-0450(2001)040<1801:TEOTGP>2.0.CO;2 (cit. on pp. 213–214, 230).

Lac, C. et al., 2018: Overview of the Meso-NH model version 5.4 and its applications. *Geoscientific Model Development*, **11**5, 1929–1969. DOI: 10.5194/gmd-11-1929-2018 (cit. on p. 209).

Lackmann, G., 2012: *Midlatitude Synoptic Meteorology - Dynamics, analysis, and forecasting*. 2nd ed. 45 Beacon Street, Boston, Massachusetts 02108: American Meteorological Society. ISBN: 978-1-878220-10-3 (cit. on p. 199).

Lacroix, A., T. Farges, R. Marchiano, and F. Coulouvrat, 2018: Acoustical measurement of natural lightning flashes: reconstructions and statistical analysis of energy spectra. *Journal of Geophysical Research: Atmospheres*, **123**21, 12,040–12,065. DOI: 10.1029/2018JD028814 (cit. on p. 47).

Lagouvardos, K., V. Kotroni, E. Defer, and O. Bousquet, 2013: Study of a heavy precipitation event over southern France, in the frame of HYMEX project: Observational analysis and model results using assimilation of lightning. *Atmospheric Research*, **134**, 45 –55. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2013.07.003 (cit. on p. 219).

Lascaux, F., E. Richard, and J.-P. Pinty, 2006: Numerical simulations of three different map iops and the associated microphysical processes. *Quarterly Journal of the Royal Meteorological Society*, **132**619, 1907–1926. DOI: 10.1256/qj.05.197 (cit. on pp. 209, 222).

Lavigne, T., C. Liu, and N. Liu, 2019: How does the trend in thunder days relate to the variation of lightning flash density? *Journal of Geophysical Research: Atmospheres*, **124**9. 2018JD029920, 4955–4974. DOI: 10.1029/2018JD029920 (cit. on p. 95).

Lay, E. H., R. H. Holzworth, C. J. Rodger, J. N. Thomas, O. Pinto Jr., and R. L. Dowden, 2004: WWLL global lightning detection system: regional validation study in Brazil. *Geophysical Research Letters*, **31**3. DOI: 10.1029/2003GL018882 (cit. on p. 40).

Lennon, C. and L. Maier, 1991: Lightning mapping system. In: *Proceedings of the International Aerospace and Ground Conference on Lightning and Static Electricity*. Cocoa Beach, FL. 01 August, NASA Conference Publication 3106, Vol. 2, 89.1–89.10 (cit. on p. 42).

Lighezzolo, R. A., R. G. Pereyra, and E. E. Avila, 2010: Measurements of electric charge separated during the formation of rime by the accretion of supercooled droplets. *Atmospheric Chemistry and Physics*, **10**4, 1661–1669. DOI: 10.5194/acp-10-1661-2010 (cit. on p. 20).

Liu, P., Y. Yang, J. Gao, Y. Wang, and C. Wang, 2020: An approach for assimilating FY4 lightning and cloud top height data using 3DVAR. *Frontiers in Earth Science*, **8**, 288. ISSN: 2296-6463. DOI: 10.3389/feart.2020.00288 (cit. on pp. 4, 8, 221).

Luque, M. Y., F. Nollas, R. G. Pereyra, R. E. Bürgesser, and E. E. Ávila, 2020: Charge separation in collisions between ice crystals and a spherical simulated graupel of centimeter size. *Journal of Geophysical Research: Atmospheres*, **125**3. e2019JD030941 2019JD030941, e2019JD030941. DOI: 10.1029/2019JD030941 (cit. on p. 20).

Lynn, B. H., G. Kelman, and G. Ellrod, 2015: An Evaluation of the Efficacy of Using Observed Lightning to Improve Convective Lightning Forecasts. *Weather and Forecasting*, **30**2, 405–423. ISSN: 0882-8156. DOI: 10.1175/WAF-D-13-00028.1 (cit. on p. 219).

Lyu, F., S. A. Cummer, Z. Qin, and M. Chen, 2019: Lightning initiation processes imaged with very high frequency broadband interferometry. *Journal of Geophysical Research: Atmospheres*, **124**, 2994–3004. DOI: 10.1029/2018JD029817 (cit. on p. 27).

MacGorman, D. R. and W. D. Rust, 1998: *The electrical nature of storms*. 1st ed. 198 Madison Avenue, New York, New York 10016: Oxford University Press. ISBN: 0-19-507337-1 (cit. on pp. 19–20, 24–25, 27, 31, 35–36, 39, 42, 47).

Mach, D. M., 2020: Geostationary lightning mapper clustering algorithm stability. *Journal of Geophysical Research: Atmospheres*, **125**5. e2019JD031900 2019JD031900, e2019JD031900. DOI: 10.1029/2019JD031900 (cit. on pp. 96, 98).

Mahfouf, J.-F., F. Ahmed, P. Moll, and F. N. Teferle, 2015: Assimilation of zenith total delays in the arome france convective scale model: a recent assessment. *Tellus A: Dynamic Meteorology and Oceanography*, **67**1, 26106. DOI: 10.3402/tellusa.v67.26106 (cit. on p. 212).

Maier, L., C. L. Lennon, T. Britt, and S. Schaefer, 1995: Lightning Detection and Ranging (LDAR) System Performance Analysis. In: *Proceedings of the 6th Conference on Aviation Weather and Systems*. Paper 8.9. Amercian Meteorological Society, Boston, MA (cit. on p. 42).

Mansell, E. R., 2014: Storm-scale ensemble kalman filter assimilation of total lightning flash-extent data. *Monthly Weather Review*, **142**, 3683–3695. DOI: 10.1175/MWR-D-14-00061.1 (cit. on pp. 4, 8–9, 220, 224).

Mansell, E. R., D. R. MacGorman, C. L. Ziegler, and J. M. Straka, 2005: Charge structure and lightning sensitivity in a simulated multicell thunderstorm. *Journal of Geophysical Research: Atmospheres*, **110**D12. DOI: 10.1029/2004JD005287 (cit. on p. 23).

Mansell, E. R., C. L. Ziegler, and D. R. MacGorman, 2007: A lightning data assimilation technique for mesoscale forecast models. *Monthly Weather Review*, **135**, 1732–1748. DOI: 10.1175/MWR3387.1 (cit. on p. 219).

Marchand, M., K. Hilburn, and S. D. Miller, 2019: Geostationary lightning mapper and earth networks lightning detection over the contiguous united states and dependence on flash characteristics. *Journal of Geophysical Research: Atmospheres*, **124**21. 2019JD031039, 11552–11567. DOI: 10.1029/2019JD031039 (cit. on pp. 3, 7, 96–97).

Marchand, M. R. and H. E. Fuelberg, 2014: Assimilation of Lightning Data Using a Nudging Method Involving Low-Level Warming. *Monthly Weather Review*, **142**12, 4850–4871. ISSN: 0027-0644. DOI: 10.1175/MWR-D-14-00076.1 (cit. on p. 219).

Matsui, T., J.-D. Chern, W.-K. Tao, S. Lang, M. Satoh, T. Hashino, and T. Kubota, 2016: On the Land–Ocean Contrast of Tropical Convection and Microphysics Statistics Derived from TRMM Satellite Signals and Global Storm-Resolving Models. *Journal of Hydrometeorology*, **17**5, 1425–1445. ISSN: 1525-755X. DOI: 10.1175/JHM-D-15-0111.1 (cit. on p. 15).

McCaul, E. W., G. Priftis, J. L. Case, T. Chronis, P. N. Gatlin, S. J. Goodman, and F. Kong, 2020: Sensitivities of the wrf lightning forecasting algorithm to parameterized microphysics and boundary layer schemes. *Weather and Forecasting*, **0**0. Discussion paper, 0. DOI: 10.1175/WAF-D-19-0101.1 (cit. on p. 97).

McCaul Eugene W., J., S. J. Goodman, K. M. LaCasse, and D. J. Cecil, 2009: Forecasting Lightning Threat Using Cloud-Resolving Model Simulations. *Weather and Forecasting*, **24**3, 709–729. ISSN: 0882-8156. DOI: 10.1175/2008WAF2222152.1 (cit. on pp. 224–226, 244).

Mölders, N. and G. Kramm, 2014: *Lectures in Meteorology*. 3rd ed. Springer Cham Heidelberg New York Dordrecht London: Springer International Publishing Switzerland. ISBN: 978-3-319-02144-7. DOI: 10.1007/978-3-319-02144-7 (cit. on p. 18).

Montmerle, T. and C. Faccani, 2009: Mesoscale Assimilation of Radial Velocities from Doppler Radars in a Preoperational Framework. *Monthly Weather Review*, **137**6, 1939–1953. ISSN: 0027-0644. DOI: 10.1175/2008MWR2725.1 (cit. on p. 212).

Morimoto, T., H. Kikuchi, M. Sato, T. Ushio, A. Yamazaki, M. Suzuki, R. Ishida, Y. Sakamoto, K. Yoshida, Y. Hobara, T. Sano, T. Abe, and Z.-I. Kawasaki, 2016: An overview of VHF lightning observations by digital interferometry from ISS/JEM-GLIMS. *Earth, Planets and Space*, **68** (1), 145. ISSN: 1880-5981. DOI: 10.1186/s40623-016-0522-1 (cit. on p. 44).

Murphy, M. J. and R. K. Said, 2020: Comparisons of lightning rates and properties from the u.s. national lightning detection network (nldn) and gld360 with goes-16 geostationary lightning mapper and advanced baseline imager data. *Journal of Geophysical Research: Atmospheres*, **125**5. e2019JD031172 2019JD031172, e2019JD031172. DOI: 10.1029/2019JD031172 (cit. on pp. 3, 7, 96–97).

Nag, A., M. J. Murphy, W. Schulz, and K. L. Cummins, 2015: Lightning locating systems: insights on characteristics and validation techniques. *Earth and Space Science*, **2**, 65–93. DOI: 10.1002/2014EA000051 (cit. on pp. 3, 7, 35–36, 38, 40, 42, 44, 47).

Nelson, J. and M. Baker, 2003: Charging of ice-vapor interfaces: applications to thunderstorms. *Atmospheric Chemistry and Physics*, **3**4, 1237–1252. DOI: 10.5194/acp-3-1237-2003 (cit. on p. 23).

Olson, W. S., C. D. Kummerow, G. M. Heymsfield, and L. Giglio, 1996: A method for combined passiveactive microwave retrievals of cloud and precipitation profiles. *Journal of Applied Meteorology*, **35**10, 1763–1789. ISSN: 0894-8763. DOI: 10.1175/1520-0450(1996)035<1763:AMFCPM>2.0.CO;2 (cit. on pp. 213, 215, 230).

Olzhabaev, T., 2013: *Datenassimilation Ein Überblick*. Master Of Science course, Universität Hamburg, Last Accessed on 2020-08-14 (cit. on p. 201).

Papadopoulos, A., T. Chronis, and E. Anagnostou, 2005: Improving convective precipitation forecasting through assimilation of regional lightning measurements in a mesoscale model. *Monthly Weather Review*, **133**, 1961–1977. DOI: 10.1175/MWR2957.1 (cit. on p. 218).

Pédeboy, S., 2015: Analysis of the French lightning locating system location accuracy. In: Balneário Camboriú, Brazil, 28th Sept. to 2nd Oct. 2015, 337–341. DOI: 10.1109/SIPDA.2015.7339299 (cit. on pp. 40, 222).

Pédeboy, S., P. Barnéoud, E. Defer, and S. Coquillat, 2018: Analysis of the Intra-Cloud lightning activity detected with Low Frequency Lightning Locating Systems. In: March 12-15, Ft. Lauderdale, Fl, USA. 25th International Lightning Detection Conference| 7th International Lightning Meteorology Conference (cit. on p. 40).

Pergaud, J., V. Masson, S. Malardel, and F. Couvreux, 2009: A parameterization of dry thermals and shallow cumuli for mesoscale numerical weather prediction. *Boundary-Layer Meteorology*, **132**, 83–106. DOI: 10.1007/s10546-009-9388-0 (cit. on pp. 211, 222).

Pessi, A. T. and S. Businger, 2009: The impact of lightning data assimilation on a winter storm simulation over the North Pacific Ocean. *Monthly Weather Review*, **137**, 3177–3195. DOI: 10.1175/2009MWR2765.1 (cit. on p. 219).

Peterson, M., S. Rudlosky, and D. Zhang, 2020: Changes to the appearance of optical lightning flashes observed from space according to thunderstorm organization and structure. *Journal of Geophysical Research: Atmospheres*, **125**4. e2019JD031087 2019JD031087, e2019JD031087. DOI: 10.1029/2019JD031087 (cit. on pp. 44–46, 97).

— 2020: Thunderstorm cloud-type classification from space-based lightning imagers. *Monthly Weather Review*, **148**5, 1891–1898. DOI: 10.1175/MWR-D-19-0365.1 (cit. on p. 97).

Poehler, H. A. and C. L. Lennon, 1979: *Lightning Detection and Ranging system LDAR system description and performance objectives*. Technical Memorandum NASA-TM-74105. NASA (cit. on p. 42).

Price, C. and D. Rind, 1992: A simple lightning parameterization for calculating global lightning distributions. *J. Geophys. Res.*, **97**, 9919–9933. DOI: 10.1029/92JD00719 (cit. on p. 224).

Price, C. and D. Rind, 1993: What determines the cloud-to-ground lightning fraction in thunderstorms? *Geophysical Research Letters*, **20**6, 463–466. DOI: 10.1029/93GL00226 (cit. on pp. 224, 253, 260).

Proctor, D. E., 1981: VHF radio pictures of cloud flashes. *Journal of Geophysical Research: Oceans*, **86**C5, 4041–4071. DOI: 10.1029/JC086iC05p04041 (cit. on p. 42).

— 1983: Lightning and precipitation in a small multicellular thunderstorm. *Journal of Geophysical Research: Oceans*, **88**C9, 5421–5440. DOI: 10.1029/JC088iC09p05421 (cit. on p. 42).

Qie, X., R. Zhu, T. Yuan, X. Wu, W. Li, and D. Liu, 2014: Application of total-lightning data assimilation in a mesoscale convective system based on the WRF model. *Atmospheric Research*, **145-146**, 255 –266. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2014.04.012 (cit. on pp. 219, 225).

Rabier, F., J.-N. Thépaut, and P. Courtier, 1998: Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **124**550, 1861–1887. DOI: 10.1002/qj.49712455005 (cit. on p. 205).

Roberts, N. and H. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review - MON WEATHER REV*, **136**, 78–97. DOI: 10.1175/2007MWR2123.1 (cit. on pp. 217, 235).

Rudlosky, S. D. and H. E. Fuelberg, 2013: Documenting storm severity in the mid-atlantic region using lightning and radar information. *Monthly Weather Review*, **141**9, 3186–3202. DOI: 10.1175/MWR-D-12-00287.1 (cit. on p. 96).

Rudlosky, S. D., S. J. Goodman, K. S. Virts, and E. C. Brunning, 2018: Initial geostationary lightning mapper observations. *Geophysical Research Letters*, **45**. DOI: 10.1029/2018GL081052 (cit. on pp. 80, 96).

Rudlosky, S. D., M. J. Peterson, and D. T. Kahn, 2017: Gld360 performance relative to trmm lis. *Journal of Atmospheric and Oceanic Technology*, **34**, 1307–1322. DOI: 10.1175/JTECH-D-16-0243.1 (cit. on p. 95).

Rust, W., D. MacGorman, E. Bruning, S. Weiss, P. Krehbiel, R. Thomas, W. Rison, T. Hamlin, and J. Harlin, 2005: Inverted-polarity electrical structures in thunderstorms in the severe thunderstorm electrification and precipitation study (steps). *Atmospheric Research*, **76**, 247–271. DOI: 10.1016/j.atmosres.2004.11.029 (cit. on p. 27).

Rutledge, S. A., K. A. Hilburn, A. Clayton, B. Fuchs, and S. D. Miller, 2020: Evaluating geostationary lightning mapper flash rates within intense convective storms. *Journal of Geophysical Research: Atmospheres*, **125**14. e2020JD032827 2020JD032827, e2020JD032827. DOI: 10.1029/2020JD032827 (cit. on p. 97).

Saba, M. M. F., K. L. Cummins, T. A. Warner, E. P. Krider, L. Z. S. Campos, M. G. Ballarotti, O. Pinto Jr., and S. A. Fleenor, 2008: Positive leader characteristics from high-speed video observations. *Geophysical Research Letters*, **35**7. DOI: 10.1029/2007GL033000 (cit. on p. 30).

Said, R. K., U. S. Inan, and K. L. Cummins, 2010: Long-range lightning geolocation using a VLF radio atmospheric waveform bank. *Journal of Geophysical Research: Atmospheres*, **115**D23. DOI: 10.1029/2010JD013863 (cit. on p. 40).

Saunders, C., 2008: Charge separation mechanisms in clouds. *Space Science Reviews*, **137**335, 335–353. DOI: 10.1007/s11214-008-9345-0 (cit. on p. 19).

Schmit, T. J., M. M. Gunshor, W. P. Menzel, J. J. Gurka, J. Li, and A. S. Bachmeier, 2005: Introducing the next-generation Advanced Baseline Imager on GOES-R. *Bulletin of the American Meteorological Society*, **86**8, 1079–1096. DOI: 10.1175/BAMS-86-8-1079 (cit. on p. 44).

Schmit, T. J., J. Li, J. Li, W. F. Feltz, J. J. Gurka, M. D. Goldberg, and K. J. Schrab, 2008: The GOES-R Advanced Baseline Imager and the continuation of current sounder products. *Journal of Applied Meteorology and Climatology*, **47**10, 2696–2711. DOI: 10.1175/2008JAMC1858.1 (cit. on p. 44).

Schultz, C. J., M. Bateman, K. Virts, G. Stano, P. Meyer, and D. Mosier, 2018: *Geostationary Lightning Mapper Performance, Capabilities, and Applications*. University of Alabama Geological Sciences Seminar, 02 November 2018. last accessed 08 October 2020 (cit. on p. 80).

Schultz, C. J., L. D. Carey, E. V. Schultz, and R. J. Blakeslee, 2017: Kinematic and microphysical significance of lightning jumps versus nonjump increases in total flash rate. *Weather and Forecasting*, **32**1, 275–288. DOI: 10.1175/WAF-D-15-0175.1 (cit. on p. 96).

Schultz, C. J., W. A. Petersen, and L. D. Carey, 2011: Lightning and severe weather: a comparison between total and cloud-to-ground lightning trends. *Weather and Forecasting*, **26**5, 744–755. DOI: 10.1175/WAF-D-10-05026.1 (cit. on p. 95).

Schulz, W., G. Diendorfer, S. Pedeboy, and D. R. Poelman, 2016: The european lightning location system euclid – part 1: performance analysis and validation. *Nat. Hazards Earth Syst. Sci.*, **16**, 595–605. DOI: 10.5194/nhess-16-595-2016 (cit. on pp. 3, 7, 222).

Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson, 2011: The AROME-France Convective-Scale Operational Model. *Monthly Weather Review*, **139**3, 976–991. ISSN: 0027-0644. DOI: 10.1175/2010MWR3425.1 (cit. on pp. 4, 8, 209, 211, 222).

Shao, X. M. and P. R. Krehbiel, 1996: The spatial and temporal development of intracloud lightning. *Journal of Geophysical Research: Atmospheres*, **101**D21, 26641–26668. DOI: 10.1029/96JD01803 (cit. on pp. 28–29, 42).

Shao, X. M., P. R. Krehbiel, R. J. Thomas, and W. Rison, 1995: Radio interferometric observations of cloud-to-ground lightning phenomena in florida. *Journal of Geophysical Research: Atmospheres*, **100**D2, 2749–2783. DOI: 10.1029/94JD01943 (cit. on pp. 29, 31, 42).

Shi, Z., H.-Q. Tang, and Y. Tan, 2016: Effects of the inductive charging on the electrification and lightning discharges in thunderstorms. *Terrestrial, Atmospheric and Oceanic Sciences*, **27**, 241. DOI: 10.3319/TAO.2015.12.10.01(A) (cit. on p. 23).

Simpson, G. C. and G. T. Walker, 1910: On the electricity of rain and snow. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, **83**564, 394–404. DOI: 10.1098/rspa.1910.0027 (cit. on p. 24).

Soula, S., 2012. *Aerospace Lab Issue 5*, 1–10 (cit. on p. 27).

Stefano, F., E. Avolio, M. Petracca, G. Panegrossi, P. Sanò, D. Casella, and S. Dietrich, 2014: Simulating lightning into the rams model: implementation and preliminary results. *Natural Hazards and Earth System Sciences Discussions*, **2**. DOI: 10.5194/nhessd-2-3351-2014 (cit. on p. 225).

Stefano, F., G. Panegrossi, M. Petracca, C. Transerici, and S. Dietrich, 2017: Impact of the assimilation of lightning data on the precipitation forecast at different forecast ranges. *Advances in Science and Research*, **14**, 187–194. DOI: 10.5194/asr-14-187-2017 (cit. on p. 220).

Steiger, S. M., R. E. Orville, and L. D. Carey, 2007: Total lightning signatures of thunderstorm intensity over north texas. part i: supercells. *Monthly Weather Review*, **135**10, 3281–3302. DOI: 10.1175/MWR3472.1 (cit. on p. 96).

— 2007: Total lightning signatures of thunderstorm intensity over north texas. part ii: mesoscale convective systems. *Monthly Weather Review*, **135**10, 3303–3324. DOI: 10.1175/MWR3483.1 (cit. on p. 96).

Stolzenburg, M., T. Marshall, and S. Karunarathne, 2020: Inception of subsequent stepped leaders in negative cloud-to-ground lightning. *Meteorol Atmos Phys*, **132**, 489–514. DOI: 10.1007/s00703-019-00702-8 (cit. on pp. 30, 33).

Stolzenburg, M. and T. C. Marshall, 1998: Charged precipitation and electric field in two thunderstorms. *Journal of Geophysical Research: Atmospheres*, **103**D16, 19777–19790. DOI: 10.1029/98JD01675 (cit. on p. 27).

— 2008: Charge structure and dynamics in thunderstorms. *Space Science Reviews*, **137**, 355–372. DOI: 10.1007/s11214-008-9338-z (cit. on pp. 24, 26–28).

Takahashi, T., 1978: Riming Electrification as a Charge Generation Mechanism in Thunderstorms. *Journal of the Atmospheric Sciences*, **35**8, 1536–1548. ISSN: 0022-4928. DOI: 10.1175/1520-0469(1978)035<1536:REAACG>2.0.CO;2 (cit. on pp. 4, 8, 20).

— 1979: Warm Cloud Electricity in a Shallow Axisyminetric Cloud Model. *Journal of the Atmospheric Sciences*, **36**11, 2236–2258. ISSN: 0022-4928. DOI: 10.1175/1520-0469(1979)036<2236:WCEIAS>2.0.CO;2 (cit. on p. 23).

Takahashi, T., S. Sugimoto, T. Kawano, and K. Suzuki, 2017: Riming Electrification in Hokuriku Winter Clouds and Comparison with Laboratory Observations. *Journal of the Atmospheric Sciences*, **74**2, 431–447. ISSN: 0022-4928. DOI: 10.1175/JAS-D-16-0154.1 (cit. on pp. 20, 22).

Thomas, R. J., P. R. Krehbiel, W. Rison, T. Hamlin, D. J. Boccippio, S. J. Goodman, and H. J. Christian, 2000: Comparison of ground-based 3-dimensional lightning mapping observations with satellite-based lis observations in oklahoma. *Geophysical Research Letters*, **27**12, 1703–1706. DOI: 10.1029/1999GL010845 (cit. on pp. 42, 81, 91, 95).

Thomas, R. J., P. R. Krehbiel, W. Rison, S. J. Hunyady, W. P. Winn, T. Hamlin, and J. Harlin, 2004: Accuracy of the lightning mapping array. *Journal of Geophysical Research*, **109**, D14207. DOI: 10.1029/2004JD004549 (cit. on pp. 3, 7).

Thompson, K. B., M. G. Bateman, and L. D. Carey, 2014: A comparison of two ground-based lightning detection networks against the satellite-based lightning imaging sensor (lis). *Journal of Atmospheric and Oceanic Technology*, **31**, 2191–2205. DOI: 10.1175/JTECH-D-13-00186.1 (cit. on p. 95).

Velde, O. A. van der, J. Montanyà, T. Neubert, O. Chanrion, N. Østgaard, S. Goodman, J. A. López, F. Fabró, and V. Reglero, 2020: Comparison of high-speed optical observations of a lightning flash from space and the ground. *Earth and Space Science*, **7**10. e2020EA001249 2020EA001249, e2020EA001249. DOI: 10.1029/2020EA001249 (cit. on pp. 97–98).

Velden, C., W. E. Lewis, W. Bresky, D. Stettner, J. Daniels, and S. Wanzong, 2017: Assimilation of High-Resolution Satellite-Derived Atmospheric Motion Vectors: Impact on HWRF Forecasts of Tropical Cyclone Track and Intensity. *Monthly Weather Review*, **145**3, 1107–1125. ISSN: 0027-0644. DOI: 10.1175/MWR-D-16-0229.1 (cit. on p. 212).

Vonnegut, B., C. B. Moore, R. G. Semonin, J. W. Bullock, D. W. Staggs, and W. E. Bradley, 1962: Effect of atmospheric space charge on initial electrification of cumulus clouds. *Journal of Geophysical Research (1896-1977)*, **67**10, 3909–3922. DOI: 10.1029/JZ067i010p03909 (cit. on p. 24).

Wang, H., Y. Liu, W. Y. Y. Cheng, T. Zhao, M. Xu, Y. Liu, S. Shen, K. M. Calhoun, and A. O. Fierro, 2017: Improving lightning and precipitation prediction of severe convection using lightning data assimilation with NCAR WRF-RTFDDA. *Journal of Geophysical Research: Atmospheres*, **122**22, 12,296–12,316. DOI: 10.1002/2017JD027340 (cit. on pp. 220, 225).

Wang, H., Y. Liu, T. Zhao, Y. Liu, M. Xu, S. Shen, Y. Jiang, H. Yang, and S. Feng, 2018: Continuous assimilation of lightning data using time-lagged ensembles for a convection-allowing numerical weather prediction model. *Journal of Geophysical Research: Atmospheres*, **123**17, 9652–9673. DOI: 10.1029/2018JD028494 (cit. on pp. 220, 225).

Wang, W. and T. T. Warner, 1988: Use of Four-Dimensional Data Assimilation by Newtonian Relaxation and Latent-Heat Forcing to Improve a Mesoscale-Model Precipitation Forecast: A Case Study. *Monthly Weather Review*, **116**12, 2593–2613. ISSN: 0027-0644. DOI: 10.1175/1520-0493(1988)116<2593:UOFDDA>2.0.CO;2 (cit. on p. 203).

Wang, Y., Y. Yang, D. Liu, D. Zhang, W. Yao, and C. Wang, 2017: A case study of assimilating lightning-proxy relative humidity with WRF-3DVAR. *Atmosphere*, **8**, 20. DOI: 10.3390/atmos8030055 (cit. on pp. 203, 220).

Wang, Y., Y. Yang, and C. Wang, 2014: Improving forecasting of strong convection by assimilating cloud-to-ground lightning data using the physical initialization method. *Atmospheric Research*, **150**, 31 –41. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2014.06.017 (cit. on p. 219).

Wattrelot, E., O. Caumont, and J.-F. Mahfouf, 2014: Operational Implementation of the 1D+3D-Var Assimilation Method of Radar Reflectivity Data in the AROME Model. *Monthly Weather Review*, **142**5, 1852–1873. ISSN: 0027-0644. DOI: 10.1175/MWR-D-13-00230.1 (cit. on pp. 212–213, 227, 230).

Williams, E. and C. R. Stanfill, 2002: The physical origin of the land-ocean contrast in lightning activity. *C. R. Pysique*, **3**, 1277–1292 (cit. on p. 15).

Williams, E., B. Boldi, A. Matlin, M. Weber, S. Hodanish, D. Sharp, S. Goodman, R. Raghavan, and D. Buechler, 1999: The behavior of total lightning activity in severe florida thunderstorms. *Atmospheric Research*, **51**3, 245 –265. ISSN: 0169-8095. DOI: 10.1016/S0169-8095(99)00011-3 (cit. on p. 95).

Williams, E. R., 1989: The tripole structure of thunderstorms. *Journal of Geophysical Research: Atmospheres*, **94**D11, 13151–13167. DOI: 10.1029/JD094iD11p13151 (cit. on p. 27).

(WMO), World Meteorological Organization, 2017: *International Cloud Atlas (WMO-No. 407)*. URL: https://cloudatlas.wmo.int/en/general-classification-of-meteors-hydrometeors.html (visited on 10/02/2020) (cit. on p. 18).

Wong, J., M. C. Barth, and D. Noone, 2013: Evaluating a lightning parameterization based on cloud-top height for mesoscale numerical model simulations. *Geoscientific Model Development*, **6**2, 429–443. DOI: 10.5194/gmd-6-429-2013 (cit. on pp. 224, 253, 260).

Yair, Y., B. Lynn, C. Price, V. Kotroni, K. Lagouvardos, E. Morin, A. Mugnai, and M. d. C. Llasat, 2010: Predicting the potential for lightning activity in Mediterranean storms based on the weather research and forecasting (WRF) model dynamic and microphysical fields. *Journal of Geophysical Research: Atmospheres*, **115**D4. DOI: 10.1029/2008JD010868 (cit. on p. 224).

Yan, X., V. Ducrocq, G. Jaubert, P. Brousseau, P. Poli, C. Champollion, C. Flamant, and K. Boniface, 2009: The benefit of gps zenith delay assimilation to high-resolution quantitative precipitation forecasts: a case-study from cops iop 9. *Quarterly Journal of the Royal Meteorological Society*, **135**644, 1788–1800. DOI: 10.1002/qj.508 (cit. on p. 212).

Yang, J., Z. Zhang, C. Wei, F. Lu, and Q. Guo, 2017: Introducing the new generation of chinese geostationary weather satellites, fengyun-4. *Bulletin of the American Meteorological Society*, **98**8, 1637–1658. DOI: 10.1175/BAMS-D-16-0065.1 (cit. on pp. 3, 7).

Zeng, R., C. Zhuang, X. Zhou, S. Chen, Z. Wang, Z. Yu, and H. Jinliang, 2016: Survey of recent progress on lightning and lightning protection research. *High Voltage*, **1**, 2–10. DOI: 10.1049/hve.2016.0004 (cit. on pp. 28–29).

Zhang, D. and K. L. Cummins, 2020: Time evolution of satellite-based optical properties in lightning flashes, and its impact on glm flash detection. *Journal of Geophysical Research: Atmospheres*, **125**6. e2019JD032024 2019JD032024, e2019JD032024. DOI: 10.1029/2019JD032024 (cit. on pp. 91, 97, 107).

Zhang, G., Y. Zhao, X. Qie, T. Zhang, Y. Wang, and C. Chen, 2008: Observation and study on the whole process of cloud-to-ground lightning using narrowband radio interferometer. *Science in China Series D: Earth Sciences*, **51** (5), 694. ISSN: 1862-2801. DOI: 10.1007/s11430-008-0049-9 (cit. on pp. 29, 31, 43).

Zhang, W., Y. Zhang, D. Zheng, L. Xu, and W. Lyu, 2018: Lightning climatology over the northwest pacific region: an 11-year study using data from the world wide lightning location network. *Atmospheric Research*, **210**, 41 –57. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2018.04.013 (cit. on p. 14).

Zhang, Z., D. Zheng, Y. Zhang, and G. Lu, 2017: Spatial–temporal characteristics of lightning flash size in a supercell storm. *Atmospheric Research*, **197**, 201 –210. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2017.06.029 (cit. on p. 18).

Ziegler, C. L., D. R. MacGorman, J. E. Dye, and P. S. Ray, 1991: A model evaluation of noninductive graupel-ice charging in the early electrification of a mountain thunderstorm. *Journal of Geophysical Research: Atmospheres*, **96**D7, 12833–12855. DOI: 10.1029/91JD01246 (cit. on p. 23).

Zipser, E. J., D. J. Cecil, C. Liu, S. W. Nesbitt, and D. P. Yorty, 2006: WHERE ARE THE MOST INTENSE THUNDERSTORMS ON EARTH? *Bulletin of the American Meteorological Society*, **87**8, 1057–1072. ISSN: 0003-0007. DOI: 10.1175/BAMS-87-8-1057 (cit. on p. 15).

Zou, X., I. M. Navon, and F. X. Ledimet, 1992: An optimal nudging data assimilation scheme using parameter estimation. *Quarterly Journal of the Royal Meteorological Society*, **118**508, 1163–1186. DOI: 10.1002/qj.49711850808 (cit. on p. 203).

Abstract As an initial analysis, an intercomparison of lightning observations over Corsica from the Lightning Imaging Sensor on the International Space Station (ISS-LIS), the Low Frequency (LF) Meteorage network, and the SAETTA Lightning Mapping Array (LMA) reveals that coincident flashes of all three lightning locating systems can be identified. Large and long-duration flashes are more likely detected by both ISS-LIS and Meteorage than small and short-duration flashes. Using the information provided by SAETTA, it is found that the flash detection efficiency of ISS-LIS degrades for flashes detected by Meteorage that do not extend over 7 km of altitude. This intercomparison methodology is further applied to analyze records of ISS-LIS relative to National Lightning Detection Network (NLDN) observations over the southeastern USA. Overall, the flash characteristics analyzed in both French and US regions are not only similar from ISS-LIS records, but also when comparing their statistics as depicted by Meteorage and NLDN. It is concluded that Meteorage and NLDN detect and locate lightning similarly.

With the advent of the Geostationary Lightning Mapper (GLM) concurrent geostationary (GEO) GLM and ground-based NLDN lightning observations are analyzed in detail to develop a complex algorithm to generate GEO lightning pseudo-observations from NLDN records. The so-called GEO lightning pseudo-observation generator first relates NLDN and GLM flash characteristics to train machine learning models, and secondly creates pseudo-GEO events from the simulated GEO flash characteristics. Finally, this generator is applied to simulate synthetic Meteosat Third Generation (MTG) Lightning Imager (LI) observations over France using Meteorage records as input. Eventually, Flash Extent Density (FED) is inferred from that pseudo MTG-LI data.

Pseudo MTG-LI FED serves as data source for a new lightning data assimilation (LDA) scheme in the French operational model AROME-France. Here, a 1-dimensional Bayesian (1DBay) retrieval inverts the FED observations and provides relative humidity (RH) profiles. The 1DBay proves to suppress spurious convection and promote convection in regions with positive FED. As a last step, retrieved RH profiles are assimilated using the 3D variational (3DVar) system of AROME-France. Despite promising results of the 1DBay, the AROME-France analysis contradicts the retrieved RH profiles in that humidity is increased in some regions where the retrieved RH profiles suggest a reduction of the background humidity.

Résumé En guise d'analyse initiale, une intercomparaison d'observations d'éclairs au-dessus de la Corse issues du détecteur Lightning Imaging Sensor de la Station Spatiale Internationale (ISS-LIS), du réseau de Météorage de basse fréquence (LF) et du réseau Lightning Mapping Array (LMA) SAETTA révèle que des enregistrements coïncidents des trois systèmes de localisation des éclairs peuvent être identifiés. Les éclairs de grande extension et de longue durée sont plus susceptibles d'être simultanément détectés par ISS-LIS et Météorage que les éclairs de petite extension et de courte durée. En utilisant les informations fournies par SAETTA, on constate que l'efficacité de détection des éclairs de l'instrument spatial ISS-LIS se dégrade pour les éclairs détectés par Météorage qui ne s'étendent pas sur plus de 7 km d'altitude. Cette méthodologie d'intercomparaison est aussi appliquée pour analyser les enregistrements du capteur spatial ISS-LIS par rapport aux observations du réseau National Lightning Detection Network (NLDN) sur le sud-est des États-Unis. Dans l'ensemble, les caractéristiques des éclairs analysées dans les deux régions ne sont pas seulement similaires quand elles sont comparées aux enregistrements du détecteur spatial ISS-LIS, mais aussi lorsque l'on compare leurs statistiques telles que décrites indépendamment par Météorage et NLDN. Il est conclu que Météorage et NLDN détectent et localisent les éclairs de la même manière.

Avec l'avènement du détecteur spatial géostationnaire (GEO) Geostationary Lightning Mapper (GLM), les observations coïncidentes de ce même détecteur GLM avec des observations du réseau terrestre NLDN sont analysées en détail pour construire un algorithme complexe générant des données synthétiques géostationnaires d'éclairs à partir des données du réseau NLDN. Ce générateur de données synthétiques d'éclairs utilise d'abord différentes caractéristiques des éclairs déduites des observations NLDN et GLM pour entraîner des modèles d'apprentissage automatique, et crée ensuite les différents pixels lumineux constituant chaque éclair synthétique à partir des caractéristiques de ce même éclair. Enfin, ce générateur est appliqué aux enregistrements du réseau français Météorage afin de simuler des observations synthétiques de l'imageur Lightning Imager (LI) de la mission Meteosat Troisième Génération (MTG) au-dessus de la France. Finalement, la densité d'étendue des éclairs (FED) est calculée à partir de ces données synthétiques MTG-LI.

La FED sert ensuite de source de données pour un nouveau schéma d'assimilation de données d'éclairs (LDA) dans le modèle opérationnel français AROME-France. Ici, une restitution bayésienne à 1 dimension (1DBay) inverse la densité FED et fournit des profils d'humidité relative. La méthode 1DBay s'avère efficace pour supprimer la convection parasite et pour favoriser la convection dans les régions à FED positive. En dernier lieu, les profils d'humidité relative restitués sont assimilés à l'aide du système variationnel 3D (3DVar) du modèle AROME-France. Malgré les résultats prometteurs de la méthode 1DBay, l'analyse AROME-France contredit les profils d'humidité relative restitués dans la mesure où l'humidité est augmentée dans certaines régions où les profils d'humidité relative restitués suggèrent une réduction de l'humidité de l'ébauche.