Determining lightning outliers based on Belgian radar data to evaluate the performance of EUCLID

Poelman, D. R. , Goudenhoofdt E., Delobbe L.

Royal Meteorological Institute of Belgium
Brussels, Belgium
dieter.poelman@meteo.be

Schulz W.

OVE-ALDIS
Vienna, Austria
w.schulz@ove.at

Abstract—**In this study, cloud-to-ground (CG) and cloud-to-cloud (CC) stroke data are superimposed on corresponding quantitative precipitation estimations (QPE) derived from radar observations in order to extract the percentage of lightning outliers, i.e. 'fake' or 'ghost' strokes, based on the distance between each lightning event and the nearest precipitation. Applying this to a large dataset from 2006-2015 it is possible to analyze the behavior of outliers over time with respect to the performance of the EUCLID network. We find that the introduction of the newest sensor technology has a positive impact on the occurrence of outliers over the years with a clear drop from 2011 onwards. Outside the European summer thunderstorm season the percentage of outliers tends to increase somewhat. This increase results from an underestimation of the precipitation by the radar at the outer radar observation boundary. The latter in its turn could be due to the fact that in general winter storms are less vertically developed compared to summer storms. In addition, it is shown that the majority of the semi-major axis (SMA) assigned to a lightning discharge is much smaller for non-outlier events compared to the SMA of outliers retrieved by this method.**

Keywords—lightning detection; radar; outlier

I. INTRODUCTION

Present-day lightning location systems (LLS) are the result of continuous development over the years with improved location accuracy, peak current estimation and type classification for each observed lightning event. However, despite the great progress made to determine those properties amongst others, occasionally some remain poorly determined. The reason for these anomalies is not straightforward to find out, but is generally related to an unfavorable network geometry, signal interferences from power lines, radio frequencies or other site-specific disturbances.

A direct way to determine the quality of a network, and therefore the values assigned to each lightning event, is by comparing the data against so-called ground-truth observations. Those observations examine for instance lightning strikes to instrumented towers [Diendorfer et al., 2000a, 2000b; Pavanello et al., 2009; Romero et al., 2011; Schulz et al., 2012; Schulz et al., 2013; Cramer and Cummins et al. 2014], control the time and location of a lightning discharge to ground by triggering it through the launch of a rocket [Jerauld et al., 2005; Nag et al., 2011; Chen et al., 2012; Mallick et al., 2014a, 2014b, 2014c], and/or are recorded by high-speed video and E-field measurements in open field [Biagi et al., 2007; Chen et al., 2012; Poelman et al., 2013a, Schulz et al., 2015]. Although best practice to retrieve in-depth information on a networks' performance, the aforementioned methods are quite labor-intensive in order to acquire a large enough dataset to retrieve a reliable output. Other methods exist, such as intercomparison studies between different LLS within regions of overlapping coverage [Said et al., 2010; Pohjola and Mäkelä, 2013; Poelman et al., 2013b]. However, the main disadvantage of those studies is the assumption of one network as being "ground-truth". In reality this is hardly the case for any existing LLS, except maybe for the short-baselined, lightning mapping arrays (LMA) [Rison et al., 1999; Thomas et al., 2004; van der Velde et al., 2013; Defer et al., 2015].

In this particular study we are interested in the amount of lightning outliers, or sometimes also referred to in the literature as fake or ghost strokes. Besides determining outliers by checking the used information of individual sensors, different statistical approaches exist. One can either superimpose the lightning discharges on top of satellite cloud imagery, radar reflectivity data or a combination of both. In the course of this paper, we solely make use of radar precipitation observations to distinguish between outlier and non-outlier lightning events.
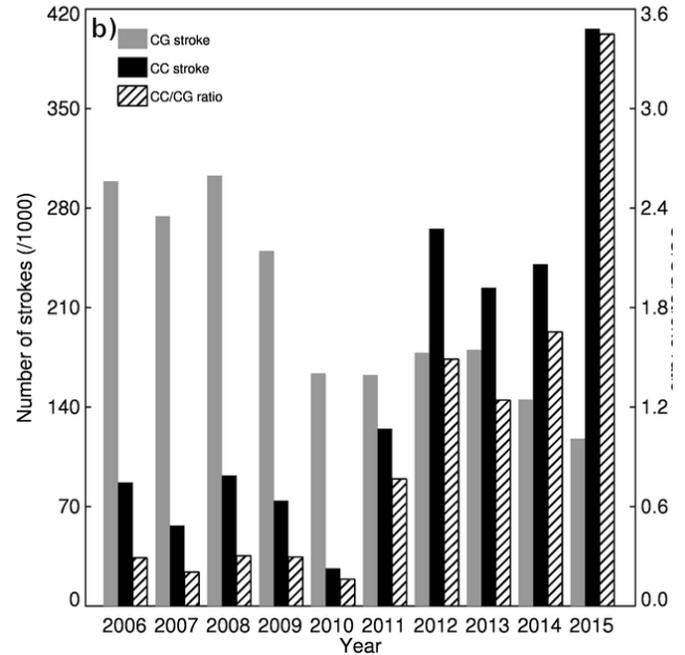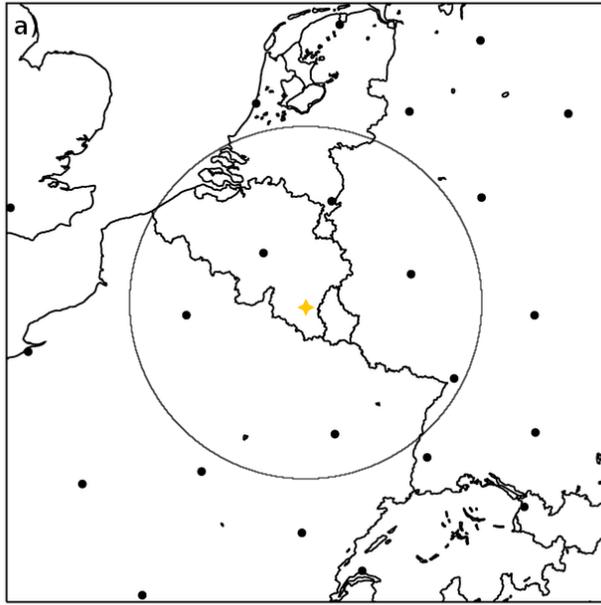
Fig. 1: a) Some of the EUCLID sensors in and around Belgium are indicated (black dots), as well as the position of the C-band radar at Wideumont (yellow star) in the middle of the circle indicating the 240 km observation radius. b) Variation of the annual CG and CC stroke counts and CC/CG ratio, as detected within the radar coverage.

II. DATA AND METHODOLOGY

Lightning data are based on observations made by the European lightning location system EUCLID [Schulz et al., 2015; Poelman et al., 2015]. This network has been operational since 2001 and processes in real-time data of 153 sensors (as of October 2015) to provide European wide lightning data of high and nearly homogeneous quality. This network has been tested continuously over the years against ground-truth data from direct lightning current measurements at the Gaisberg Tower (GBT) in Austria and data from E-field and video recordings in Austria, France and Belgium [Schulz et al., 2015]. It has been found that the location accuracy (LA) dropped steadily over the years down to the present LA in the range of 100 m. The locations of some of the EUCLID sensors in and around Belgium are plotted in Fig. 1a, while Fig. 1b presents the CG and CC stroke count as well as the CC/CG ratio as observed within the radar coverage domain.

Three radars are located in Belgium, of which two are operated by the Royal Meteorological Institute of Belgium (RMIB). One of these radars, operational since 2001, is positioned in Wideumont (49.9°N, 5.5°E) at 592 m above sea level in the southeast of Belgium, see Fig. 1. This particular radar is a single-polarization C-band Doppler radar and performs a 5-elevation scan every 5 minutes producing reflectivity measurements up to 240 km. The radar thus covers Belgium, Luxembourg as well as parts of France, the Netherlands and Germany. Quantitative precipitation estimation (QPE) on a 500 by 500 meters grid is derived from reflectivity measurements after applying clutter filtering, beam blockage correction and reflectivity profile correction.

Subsequently, the five minutes resolution rain rates are then accumulated to 10-min and hourly precipitation rates used further in this study. Note that the radar reflectivity threshold is set at 7 dBZ, corresponding to a rainfall threshold of 0.1 mm/h, below which rainfall is considered as nonexistent in this study.

Cloud-to-ground (CG) as well as cloud (CC) strokes in the
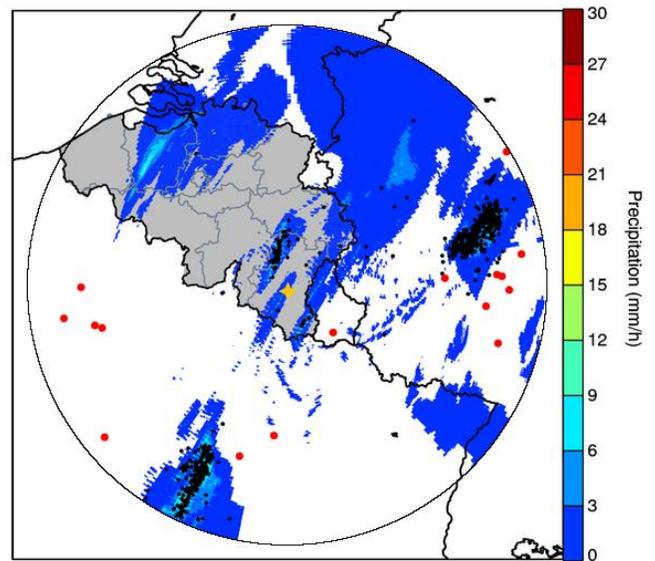
Fig. 2: Example of an hourly quantitative precipitation estimation (QPE) superimposed with the lightning events observed during the same time interval. The 'true' events are indicated as black dots, whereas the derived outliers are plotted in red.
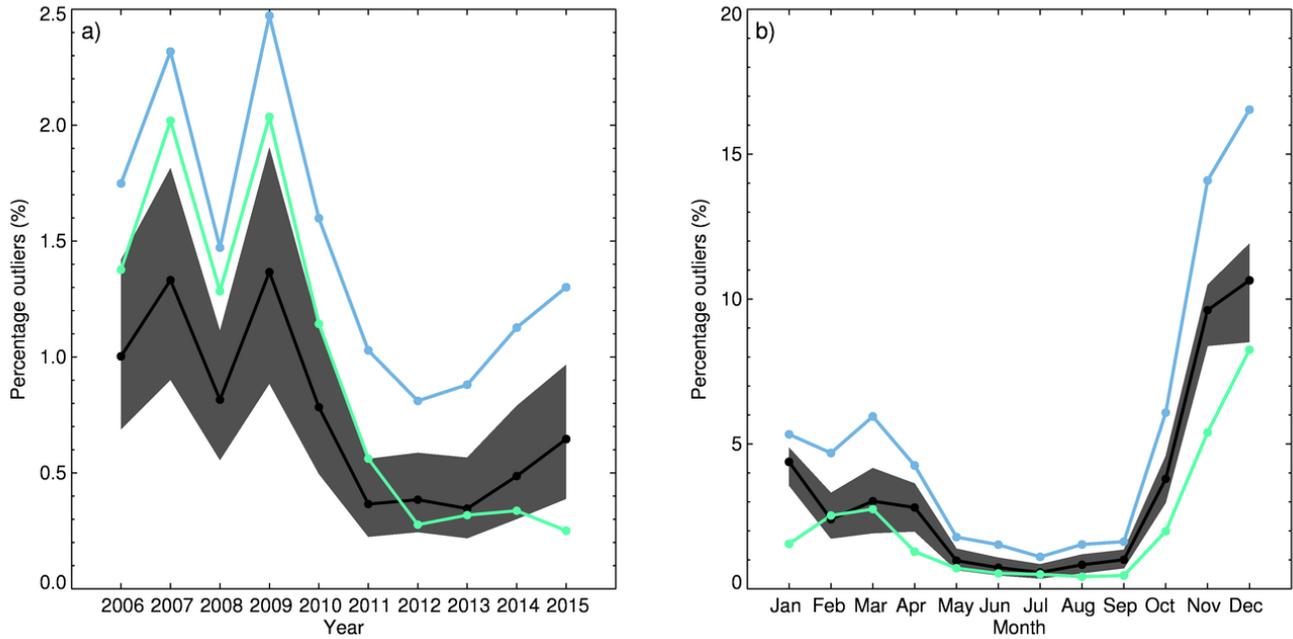
Fig. 3: a) Annual and b) monthly percentage of outliers based on 2006-2015 EUCLID stroke data. The black line represents the CG outliers based on hourly intervals and a 5 km search radius, whereas the upper and lower boundary of the grey area results from using a 2 km and 10 km radius, respectively. The percentage of CC outliers based on hourly intervals and a 5 km search radius is shown in green. The blue line depicts the CG outliers when using 10-min intervals and a 5 km search radius.

corresponding time interval are then superimposed on the QPE. Subsequently, a stroke is defined as an outlier when no precipitation within a certain distance has been observed. This distance is somewhat chosen arbitrary. Different runs are performed applying a distance of 2, 5, and 10 km, respectively. An example of this method is visualized in Fig. 2. On the left the hourly QPE is plotted at a particular time, while on the right hand side all the lightning strokes are superimposed as black dots with the retrieved outliers in red for clarity. In this figure an event is defined as an outlier when no precipitation is observed within a radius of 5 km. During this particular hour, 16 out of the 1347 strokes are flagged as outliers, or 1.2% of the total.

III. RESULTS

Fig. 3a plots the percentage of outliers for a given year between 2006 and 2014. The black line represents the percentage of CG outliers based on hourly intervals and a 5 km search radius, whereas the upper and lower boundary of the grey area results from using a 2 km and 10 km search radius, respectively. It is clear that a smaller search radius will increase the percentage of outliers, and vice versa. The percentage of CC outliers based on hourly intervals and a 5 km search radius is shown in green. The blue line depicts the CG outliers using 10-min intervals and a 5 km search radius. On average 0.8% of the CG events per year are classified as outliers based on hourly QPE and a 5 km search radius by the adopted method and this value increases to 1.6% based on 10-

min precipitation accumulations. Using 10-min intervals an increased number of outliers is expected because the hourly intervals may hide some of those due to the temporal movement of the storm within this hour. In a similar way, one expects a further increase when 5-min intervals are used. Nevertheless, a similar trend is visible in that during the years 2006-2010 the percentage on average is higher than what is found for the years 2011-2014. The drop from 2009-2011 by a factor of about two is apparent and stems partly from the fact that during that time sensors of the older type LPATS and IMPACT were changed into the newer LS700x type of sensors. As a result the location accuracy improved from that moment onwards. In addition, some quality parameters in the central processor 'TLP' were changed in 2010 which impacts the results. However, in 2015 the percentage of outliers increases visibly and can be traced back to the introduction of a new algorithm to locate CG events at that time. Although not shown in this figure, it is worthwhile to notice that reprocessing the 2015 CG data with the current and latest updated algorithm brings back the percentage of outliers at the same value as is found within the period 2011-2014. As regards to CC detections, on average over the years 0.7% are outliers and follows a similar trend compared to the CG outliers. Note that the CC detections augmented drastically by the introduction of LS700x technology in the network from 2011 onwards as seen in Fig. 1b. Hence, the relatively low amount of CC outliers during those particular years has a large impact on the overall average.
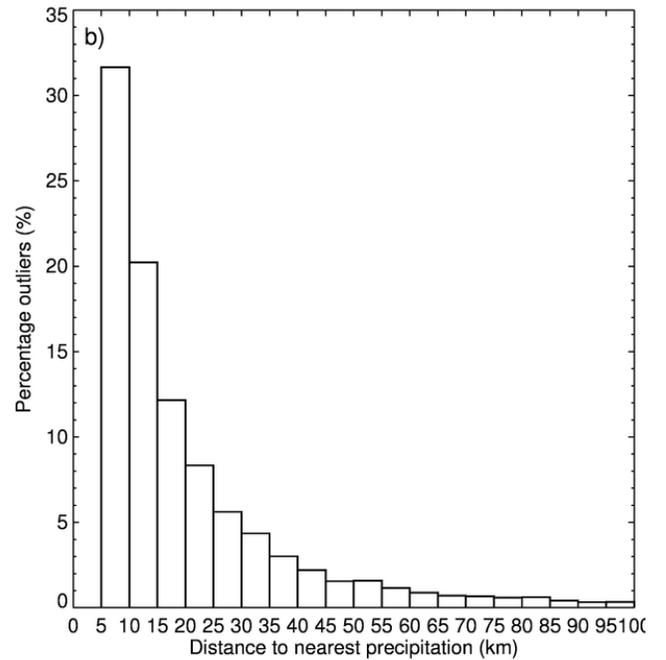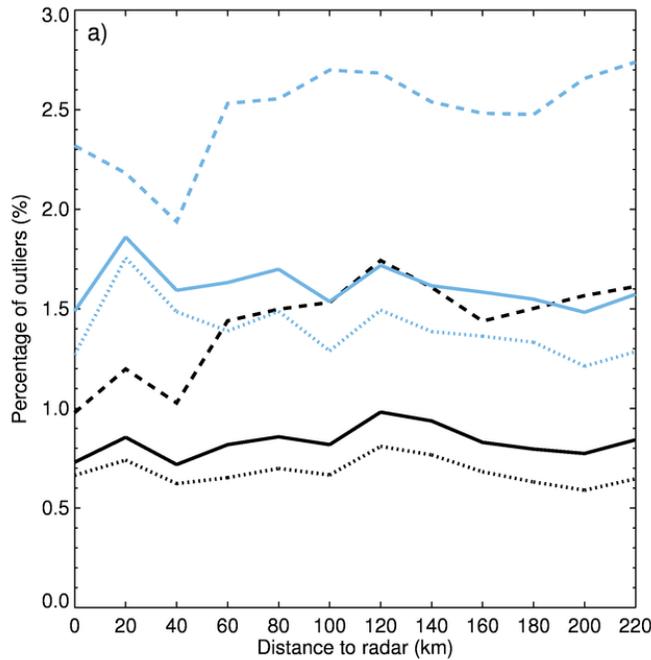
Fig. 4: a) Relative amount of CG outliers as a function of distance to the radar based on hourly (black) and 10-min (blue) intervals for the whole year (solid), Sept-May (dashed) and June-Aug (dotted). b) Percentage of CG outliers as a function of distance to the nearest precipitation.

Fig. 3b visualizes the monthly variation of outliers. An obvious decrease is observed in the percentage of outliers during May-Sept, compared to the other months of the year. A reasonable amount of precipitation will be detected without

Fig. 5: All CG outliers are plotted in black as a function of distance to the radar and to the nearest precipitation. In addition, for each distance interval to the radar of 20 km a corresponding whisker box is plotted in blue indicating the lower (10% and 25%), median and upper (75% and 90%) percentiles of the distance to the nearest precipitation. In addition, the green dot indicates the mean value.

any problem close or far away from the radar, while it is the light precipitation that tends to be harder to detect at large distances. While during summer most of the storms are associated with large amounts of precipitation in vertically extended clouds, winter storms are peculiar in nature and tend to occur with lower precipitation amounts and at somewhat lower altitudes. Hence, the precipitation can be undetected especially during winter storms due to overshooting further away from the radar. This in turn influences the outlier classification with the method employed.

In Fig. 4a, the percentage of CG outliers within consecutive rings with a width of 20 km from the radar is plotted. Black lines are based on hourly intervals, whereas blue lines result from the 10-min data. The solid lines represent the average over the full year, whereas dotted lines are used to represent the result for Jun-Aug, while dashed lines for Sept-May. We choose to split the year in this particular way since this results into a comparable lightning density from the smallest surface inner ring to the largest surface outer ring. One notices the same behavior for the hourly as well as the 10-min data in that the percentage of outliers is quasi stable over the radar coverage when averaged over all the months of the year. However, this percentage rises slightly with increasing distance from the radar during Sept-May. Similar as to what is found in Fig. 3b, this could result from to the fact that the radar underestimates or does not detect rain at larger distances. Note that during Sept-May the lightning activity accounts for only about 20% of the total lightning activity during the year in Belgium [Poelman, 2014], hence it has minor influence on the overall behavior. A very similar behavior is found for the CC outliers.
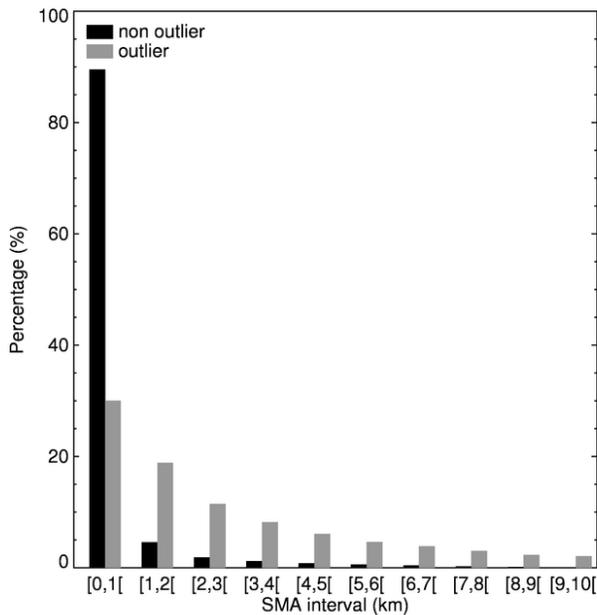
Fig. 6: Distribution of the semi-major axis (SMA) for increasing intervals of 1 km for CG outliers (grey) and non-outliers (black) based on 10 min QPE intervals.

In Fig. 4b the percentage of CG outliers as a function of distance to its nearest precipitation is plotted, based on 10-min intervals and a 5 km search radius. It is found that in more than 30% of the cases, the distance between an outlier and its nearest precipitation is situated within]5, 10] km and drops off steadily for the other intervals. About 50% of all suspected outliers have precipitation located within 15 km. A very similar behavior is found based on hourly intervals and for CC outliers.

In Fig. 5 the CG outliers as identified by the adopted method are positioned as a function of distance to the radar and to the nearest precipitation based on 10-min intervals and a 5 km search radius. On top of this for each 20 km interval from the radar a whisker box is plotted in blue, indicating the lower (10% and 25%), median and upper (75% and 90%) percentiles of the distance to the nearest precipitation, whereas the green dot indicates the mean value. Within a distance of 100 km from the radar the mean value of the distance to the nearest precipitation is stable at around 20 km. This value increases slightly further away from the radar and peaks at 28 km at the outer boundary of the radar coverage. The median values are somewhat lower and vary around 13-14 km independent of the distance to the radar, except for the outer ring where it increases towards 16.5 km. A very similar behavior is found based on hourly intervals and for CC outliers.

Finally, the central processor TLP assigns each lightning event to a value of the semi-major axis (SMA) of the 50% confidence ellipse. This value can be used as a quality indicator of the detection. The distribution of SMA is plotted in Fig. 6 and split into outlier and non-outlier events found by this study. Whereas for the non-outlier events the majority, i.e. 90%, of the SMA falls in the range of 0-1 km, this is only the case for 30% of the outliers. Hence, outliers have on average larger SMA values compared to non-outlier detections. A similar

behavior is found based on hourly intervals and for CC outliers.

Up to now, a lightning event is classified as an outlier based solely on its distance to the nearest detected precipitation. However, when SMA is taken into account, some events previously classified as outliers could turn out to be 'true' lightning events after all. This is so when the difference between the distance to the nearest precipitation and SMA is smaller than the adopted search radius. Hence, in case of a 5 km search radius, it is found that the amount of CG (/CC) outliers drops 20% (/15%) when the supplementary SMA information is utilized to differentiate between outlier and non-outlier events.

IV. SUMMARY

In this study all lightning events detected by the EUCLID network during 2006 and 2015 that fall within the coverage of the radar at Wideumont, Belgium, are classified as outliers or non-outliers based on their distance to the nearest precipitation. It is found that the percentage of the outliers drops over time, with the lowest values found from 2011 onwards. This may well reflect the improved performance of the EUCLID network when more and newer type of sensors became operational in and around Belgium. The percentage of outliers increases by a factor of two when 10-min compared to hourly intervals are used. On average, the amount of CG and CC outliers is similar. However, keep in mind that only from 2011 onwards EUCLID improved its capability to detect CC signals by the implementation of LS700x sensors. Outside the European summer thunderstorm season the percentage of outliers tend to increase somewhat and also further away from the radar. This increase could be related to the fact that the radar underestimates or might not detect to some extent precipitation at the outer radar boundary. This in its turn is related to the fact that in general winter storms are less vertically developed compared to summer storms. Hence, at large distances from the radar, the lowest radar beam can overshoot regions where indeed precipitation is present. In addition, it is found that in general the SMA of non-outliers is much smaller compared to the SMA belonging to outliers.

The method used in this work can be easily extended for example by expanding the area of coverage through the use of composite data of the three different radars located in Belgium. Also, the shortest time scale used in this study is the 10-min interval period due to computing time considerations. Ideally, however, this method should be applied to the time scale of one radar volume scan, i.e. 5 minutes. In addition, one could think of combining radar as well as satellite data to discriminate between outlier and non-outlier events.

REFERENCES

Biagi, C. J., K. L. Cummins, K. Kehoe, and E. P. Krider (2007), National lightning detection network (NLDN) performance in southern Arizona, Texas, and Oklahoma in 2003–2004, J. Geophys. Res., 112, D05208, doi:10.1029/2006JD007341.

Chen, M., D. Zheng, Y. Du, and Y. Zhang (2013), A statistical method for evaluating detection efficiency of lightning location network and its application, Atmos. Res., 128, 13–23.

Cramer, J. A., and K. L. Cummins (2014), Evaluating location accuracy of lightning location networks using tall towers, in 23rd International Lightning Detection Conference & 5th International Lightning Meteorology Conference, Vaisala Inc., Tucson, Arizona.

Defer, E., Pinty, J.-P., Coquillat, S., Martin, J.-M., Prieur, S., Soula, S., Richard, E., Rison, W., Krehbiel, P., Thomas, R., Rodeheffer, D., Vergeiner, C., Malaterre, F., Pedeboy, S., Schulz, W., Farges, T., Gallin, L.-J., Ortéga, P., Ribaud, J.-F., Anderson, G., Betz, H.-D., Meneux, B., Kotroni, V., Lagouvardos, K., Roos, S., Ducrocq, V., Roussot, O., Labatut, L., and Molinié, G. (2015), An overview of the lightning and atmospheric electricity observations collected in southern France during the HYdrological cycle in Mediterranean EXperiment (HyMeX), Special Observation Period 1, Atmos. Meas. Tech., 8, 649–669, doi:10.5194/amt-8-649-2015.

Diendorfer, G., M. Mair, W. Schulz, and W. Hadrian (2000a), Lightning current measurements in Austria—Experimental setup and first results, in 25th Intl. Conf. on Lightning Protection, pp. 44–47, ICLP Centre, Rhodes, Greece.

Diendorfer, G., W. Schulz, and M. Mair (2000b), Evaluation of a LLS based on lightning strikes to an instrumented tower, in 16th International Lightning Detection Conference, Tucson, Ariz.

Goudenhoofdt, E., and L. Delobbe (2009), Evaluation of radar-gauge merging methods for quantitative precipitation estimates, Hydrol. Earth Syst. Sci., 13, 195–203.

Jerauld, J., V. Rakov, M. Uman, K. Rambo, D. Jordan, K. Cummins, and J. Cramer (2005), An evaluation of the performance characteristics of the US National Lightning Detection Network in Florida using rocket-triggered lightning, J. Geophys. Res., 110, D19106, doi:10.1029/2005JD005924.

Mallick, S., V. Rakov, T. Ngin, W. Gamerota, J. Pilkey, J. Hill, M. Uman, D. Jordan, A. Nag, and R. Said (2014a), Evaluation of the GLD360 performance characteristics using rocket-and-wire triggered lightning data, Geophys. Res. Lett., 41, 3636–3642, doi:10.1002/2014GL059920.

Mallick, S., et al. (2014b), Performance characteristics of the NLDN for return strokes and pulses superimposed on steady currents, based on rocket-triggered lightning data acquired in Florida in 2004–2012, J. Geophys. Res. Atmos., 119, 3825–3856, doi:10.1002/2013JD021401.

Mallick, S., et al. (2014c), Performance characteristics of the ENTLN evaluated using rocket-triggered lightning data, Electr. Power Syst. Res., 118, 15–28, doi:10.1016/j.epsr.2014.06.007.

Nag, A., et al. (2011), Evaluation of US National Lightning Detection Network performance characteristics using rocket-triggered lightning data acquired in 2004–2009, J. Geophys. Res., 116, D02123, doi:10.1029/2010JD014929.

Pavanello, D., F. Rachidi, W. Janischewskyj, M. Rubinstein, V. Shostak, C. A. Nucci, K. L. Cummins, A. M. Hussein, and J. S. Chang (2009), On the current peak estimates provided by lightning detection networks for lightning return strokes to tall towers, IEEE Trans. Electromagn. Compat., 51(3), 453–458.

Poelman, D. R., W. Schulz, and C. Vergeiner (2013a), Performance characteristics of distinct lightning detection networks covering Belgium, J. Atmos. Oceanic Technol., 30, 942–951, doi: 10.1175/JTECH-D-12-00162.1.

Poelman, D. R., F. Honoré, G. Anderson, and S. Pedeboy (2013b), Comparing a regional, subcontintental, and long-range lightning location system over the Benelux and France, J. Atmos. Oceanic Technol., 30, 2394–2405, doi:10.1175/JTECH-D-12-00263.1.

Poelman, D.R. (2014), A 10-year study on the characteristics of thunderstorms in Belgium based on cloud-to-ground lightning data, Mon. Wea. Rev., 142, 4839–4849.

Poelman, D.R., W. Schulz, G. Diendorfer, and M. Bernardi (2015), The European lightning location system EUCLID – Part 2: Observations, Nat. Hazards Earth Syst. Sci. Discuss., 15, 5357–5381, doi:10.5194/nhessd-15-5357-20152015.

Pohjola, H., and A. Mäkelä (2013), The comparison of GLD360 and EUCLID lightning location systems in Europe, Atmos. Res., 123, 117–128.

Rison, W., Thomas, R. J., Krehbiel, P. R., Hamlin, T., and Harlin, J. (1999), A GPS-based three-dimensional lightning mapping system: initial observations in central New Mexico, Geophys. Res. Lett., 26, 3573–3576.

Romero, C., M. Paolone, F. Rachidi, M. Rubinstein, A. Rubinstein, G. Diendorfer, W. Schulz, M. Bernardi, and C. A. Nucci (2011), Preliminary comparison of data from the säntis tower and the euclid lightning location system, XIth International Symposium on Lightning Protection (SIPDA), Fortaleza, Brazil, 3–7 Oct.

Said, R., U. Inan, and K. Cummins (2010), Long-range lightning geolocation using a VLF radio atmospheric waveform bank, J. Geophys. Res., 115, D23108, doi:10.1029/2010JD013863.

Schulz, W., C. Vergeiner, H. Pichler, G. Diendorfer, and S. Packet (2012), Validation of the Austrian Lightning Location System ALDIS for negative flashes, in Proc. CIGRE Symposium.

Schulz, W., H. Pichler, G. Diendorfer, C. Vergeiner, and S. Pack (2013), Validation of detection of positive flashes by the Austrian Lightning Location System (ALDIS), in 12th Intl. Symposium on Lightning Protection (XII SIPDA), paper 2.2, Belo Horizonte, Brazil.

Schulz, W., G. Diendorfer, S. Pedeboy, and D.R. Poelman (2015), The European lightning location system EUCLID – Part 1: Performance validation, Nat. Hazards Earth Syst. Sci. Discuss., 15, 5325–5355, doi:10.5194/nhessd-15-5325-2015.

Thomas, R. J., Krehbiel, P. R., Rison, W., Hunyady, S. J., Winn, W. P., Hamlin, T., and Harlin, J. (2004), Accuracy of the lightning mapping array, J. Geophys. Res., 109, D14207, doi:10.1029/2004JD004549.

van der Velde, O., and Montanya, J. (2013), Asymmetries in bidirectional leader development of lightning flashes, J. Geophys. Res. Atmos., 118, 13, 504–13, doi:10.1002/2013JD020257.