2012 International Conference on Lightning Protection (ICLP), Vienna, Austria

Using 20 years of lightning data in ground flash density statistics in France

Stéphane Pédeboy
Météorage
Pau, France
sp@meteorage.com

Abstract— **Most of lightning statistics used by the lightning protection community are derived from lightning locating systems. Such systems collect flash data over large regions and during long period of time. In their life, those systems are likely to undergo several major changes as the state of the art in remote sensing techniques improves leading to some impacts on the data. In addition the flash data measured by lightning locating systems underestimate the actual lightning risk since it accounts only for one ground contact per flash that is well known nowadays not correct. Then the different evolutions in observing systems and underestimation introduced by flash data must be addressed and compensated in order to give a more accurate and relevant information for lightning risk assessment.**

The French national lightning locating system operated by Météorage, has collected more than 20 years of lightning data all over the country. This system is no exception since from its inception several major changes, in either technology or system settings, have significantly modified the lightning detection performances affecting the homogeneity of the data and the relevancy of the GFD statistics.

The work presented in this paper is based on this long duration French dataset. It attempts to define a method which comes around the inhomogeneity introduced by lightning detection performance evolutions and suggests the use of the flash ground contacts multiplicity instead of flash data only. To achieve this goal the cumulative peak current distribution method developed by the CIGRE task force C404 is used to determine the compensation factors to correct the statistics for detection efficiency effect. In addition, it is suggested the use of ground contacts data instead of flash data for lightning risk statistics. This parameter is derived from the lightning data collected in France on 2011 with a program developed by Météorage based on a clustering algorithm so called 'k-means'.

The method suffers from some necessary assumptions depending on the Météorage's LLS history and operational background, but the final new ground contacts density parameter derived from the longest observation period available in France looks more realistic and reliable. This work is a first attempt that must be extended in the future to compute high spatial resolution statistics supporting new applications for lightning risk assessment.

Keywords-component; risk assessment; ground flash density; lightning locating system; ground contacts; detection efficiency; lightning protectioncommunity;

I. INTRODUCTION

The French national low frequency Lightning Locating System (LLS) was setup by Météorage in 1987. One of the goals for this project was to collect lightning data continuously and countrywide to provide accurate and reliable services to end users. In particular, a database has been archiving lightning data located in real time by the system for more than 20 years providing precious historical information on lightning activity in France. One application for this long term dataset is the computation of the Ground Flash Density (GFD) that is equal to the number of ground flashes hitting a square kilometer per year. This parameter is of great importance for the lightning protection community as it helps determining the risk level for people, buildings or other assets to be struck by lightning. Since lightning is driven by meteorological and topographical conditions, the GFD parameter can exhibit large regional and inter-annual variations. Diendorfer [6] has demonstrated the confidence level of GFD statistics is dependent on the number of lightning data used in the calculation. Then, the use of the longest historical data is of course highly recommended and sometimes necessary when the region of interest exhibits little thunderstorm activity, in order to insure relevant statistics for a better assessment of the lightning risk.

However, the handling of long-time series data must be taken with care because of discrepancies introduced by the different evolutions LLS may undergo during its life as remote sensing state of the art improvement, changes in topology of the sensors array, number of sensors in the network, settings and parameterization of the system. In addition to these factors, LLS may suffer from operational failures producing breaks in the real-time data flow. All these factors contribute to affect the overall quality and homogeneity of the data leading to possible bias in trend or seasonality analysis.

In general GFD is provided by LLS operators without any compensation of these possible errors. In addition, based on flash data, this parameter is likely to underestimate the actual lightning risk which is directly related to the number of striking points to the ground. A flash is represented by one ground contact that is usually the location of its first return stroke. Unfortunately this is not consistent with what nature really produces since half flashes exhibit more than one ground contact.

Météorage is no exception, provides its users with GFD statistics based on flash data on the last decade, throwing out the remaining oldest data from the calculation. However it turned out 10 years of data is not long enough to get rid of the annual natural variation in lightning activity and get stable statistics.

The goal of this work is to find a method permitting the production of relevant statistics to the lightning protection community, derived from long time series lightning dealing with multiple ground contacts instead of flash data only. Recent improvements in the location accuracy of the Météorage's LLS make the identification of the separate ground contacts in flashes possible, giving a way for translating GFD to Ground Contact Density (GCD).

This method is detailed in this paper and some corresponding results obtained in France are presented to illustrate the various steps.

II. Background

Before presenting the method and its results, it will be helpful to provide some background on the different causes that impact the quality of LLS-derived lightning data. Also, it is necessary to present the key points of the history from Météorage's LLS in order to clearly understand the breaks in homogeneity of the data that will be commented later on.

A) Review of causes producing discrepancy

LLS use electromagnetic remote sensing techniques to measure and provide information about lightning events, thus there is no direct measurement [2]. On one hand this allows safely monitoring of lightning activity over wide areas, but on the other hand the quality of the remote observations relies on quite a few parameters like:

- Measurement errors: lightning shape (polarity, magnitude and duration) of electromagnetic signals produced by lightning return strokes, attenuation and distortion of the signal due to finite soil conductivity, interferences produced by radio transmitters close to sensors sites …
- Technology and level of the state of the art in remote sensing techniques: type and model of sensors, electronics and software, location algorithm and data processing...
- Operational factors: sensors baselines, network geometry, system calibration and settings and reliability of the technical infrastructure used by the LLS, mainly electrical power and telecom links.

All these factors have various impacts on the overall performances of LLS. As an example telecoms or power failures at sensor sites degrade the local performance generally for short periods of time assuming the operator reacts rapidly to troubleshoot its system. However, lightning detection limitations due technology affect measurements and observations durably on all the observed coverage until the system is upgraded with an up-to-date technology or a more efficient set of settings.

Practically, because of the real life operation and evolutions, long-period lightning data series are not homogeneous because of, technological or operational factors. This is a problem when statistical analyses try to determine the trend of GFD comparing the lightning activity from year to year. It can be assumed the effect of the operational failures is negligible as long as one deal with yearly datasets over large regions. However, sensor technology, settings and network configuration evolutions are of particular importance since their effect may impact the system performances over long periods and wide areas. As a result, each change in these fields tends to produce a break in the continuity of the data homogeneity that must be compensated before using them in reliable statistics.

B) History on the French national LLS

Météorage is, at the same time, the name of the French National LLS and the name of the company which owns and operates the system. This section provides sufficient information about the history of the LLS which helps to explain the breaks in data homogeneity.

Roughly, Météorage underwent three major technological upgrades since its inception, leading to the following ages:

- From 1987 to 1996: First age period…

This is the very beginning of the operational system, started to deploy in 1987. At this time it is made of a 17 ALDF sensors array, manufactured by Lightning Location Protection Inc (LLP), covering the country. The sensors are direction finders' meaning only the direction of the lightning signal is used by the location processor (APA280) to locate flashes. In 1995 data from some sensors in Italy were added in order to improve detection efficiency over neighboring regions and expand the initial LLS coverage. The first three years of data are not usable as the system was being built and not fully operational yet.

- From 1996 to 2008: Middle age period…

In December 1996 the French sensors were upgraded with the IMPACT 141T model from Global Atmospherics Inc

(GAI). A big improvement is gained with the integration of the time of arrival technique. The hardware and electronics are also improved. In 1999, the lightning processor is upgraded from AP283-T to LP2000 which is a software running on a Solaris workstation, enhancing the strokes processing capability. Subsequent strokes can now be located. In June 1999, the peak current calibration factor changed in order to match the manufacturer recommendation. Prior this date the calibration factor was set to match Berger's findings at the Monte Salvadore in Switzerland. In order to keep peak current values consistency with the new settings, all previous amplitudes in the database were recalculated according to the new calibration factor. In 2009 again, news changes in peak current calculation were done, concerning this time the calibration factor and the signal attenuation model [1]. These parameters were taken from results obtained on the Gaisber tower in Austria [16].

Similarly to Italy, agreements for exchanging neighboring sensors raw data are signed in 1999 with Spain and Germany in 2006.

- Since 2008: the modern age…

From mid of 2008 to early 2009, the French sensors were upgraded with the LS7001 model that is a fully digital sensor with no measurement dead time. The performance is dramatically improved in term of detection efficiency and location accuracy as well. At the same time the LP2000 was replaced with TLP131 which includes an improved location algorithm.

In addition, it is important to note the upgrade of the central analyzer underwent in 1999 introduced a major change in the flash data consistency because of the change in the flash grouping algorithm. Prior this date, subsequent strokes are not localized but only reported meaning only the first stroke is used to localize flashes, subsequent strokes being grouped on angle measurements basis (by default +/-2.5 degrees of azimuth). The flash peak current is based on the highest return stroke peak current. Since, 1999, subsequent strokes are all located and grouped when occurring within one second and 10 km apart from each other. This algorithm is described in details by Cummins [1]. The flash peak current is based on the first return stroke peak current. As a result this change highly affects the flash multiplicity and flash peak current.

This brief history of the Météorage's system practically illustrates the different evolutions, in its configuration (the number of sensors being concentrated) and in the remote sensing state of the art (improvement of sensors sensitivity, events discrimination, processing rate…) a long lasting LLS can experience during its operational life.

C) Multiple ground contacts flashes

Several studies, based on video records, have demonstrated about 50% of negative flashes produce in average up to 1.7 contacts on the ground [10][11][12][13]. This means not all return strokes of a flash go through the same channel, but in some conditions some of them tend to produce new channels creating forked lightning flash with separated attachment point to the ground. A recent study has shown the number of ground contacts is highly dependent on the terrain showing an increasing of about 20% in some high terrain and high terrain gradient conditions, reaching 2.1 ground contacts per negative flashes[4]. As a result the number of ground contact in flashes can vary depending on the geographical areas of concern.

Despite this statement, today, most of GFD statistics are still derived from flash data collected by LLS leading to underestimated statistics in regard of real threat represented by lightning for material assets or living beings.

III. Presentation of the method

As introduced previously the statistics provided to the lightning protection community for risk assessment suffers from two main points: discrepancy in the flash data source and underestimation of the real ground contacts. Based on these statements, the method suggested by Météorage is designed to get rid of these limitations.

A) Step 1 : Dataset compensation for FDE effect

It is firstly necessary to compensate the initial raw flash data, which serves in GFD calculation, for the evolution of the observing system in flash detection efficiency (FDE). Together with the location accuracy FDE is one of the most important performance parameters of LLS. It represents the fraction of flashes detected, in respect with the actual number of flashes occurring. Ideally, estimation of this parameter need to compare lightning data located by LLS to ground truth data collected with another observation system known to be efficient enough to be considered as a reference. It is then possible to determine the absolute detection efficiency of a system with instrumented towers [5] or rocket triggered lightning [8]. However both methods are difficult and expensive to setup, furthermore on large areas covered by national LLS.

Since determination of the absolute detection efficiency requires ground-truth information, it is simpler to assess the relative DE as it can be derived from the lightning dataset itself. Such a method was developed by the CIGRE working group C4.404 in the framework of a study on the effects of LLS performances on observed lightning parameters [9]. It is based on comparisons between reference and tested cumulative peak current distributions, since there is a tight link between both peak current and detection efficiency. C4.404 group suggests two distributions curves can be scaled to match for peak current values higher than a certain common limit I_0 which depends on the detection efficiency. It is important to note this method is designed to determine the stroke detection efficiency, so it is not directly applicable for FDE which is dependent on the number of strokes in a flash and the stroke detection efficiency. However, in the early dataset only flash, meaning the first return stroke, data is available, and then it is impossible to compute the stroke detection efficiency. The choice is made to use the first return stroke to approximate the FDE in agreement with C4.404 recommendation.

Fig. 1 presented below shows four scaled negative peak current cumulative distributions for different flash detection efficiency. It is possible after scaling to read the relative FDE values on the vertical axis.

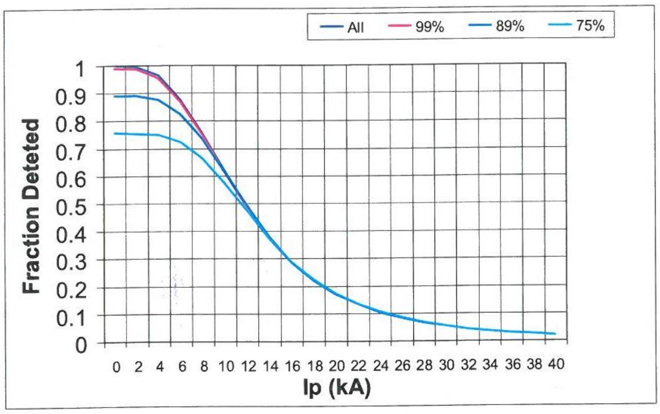

Figure 1 : Example of scaled cumulative Ipeak distribution

However this method must follow some strict rules. Firstly, only negative flashes long term observation periods must be involved in such distributions. In addition the parameters used by the system for computation of peak currents must not have changed during the compared tested and reference periods. Finally LLS performances must be consistent over both periods without any significant changes. Clearly, this is hard to achieve in operational systems.

This method is particularly well adapted for this work but some care must be taken since one important rule on the peak current is violated in 1999. Prior to this date, peak currents assigned to flashes correspond to highest stroke peak currents, whereas the peak current amplitude of the first stroke is assigned to flashes after the 283 upgrade (see chapter II-B). As a result there is a significant break in the homogeneity of the flash data that must be compensated for.

The analysis of the initial dataset revealed also the location algorithms tend to duplicate flashes in some circumstances that had to be removed before relative FDE analysis. The criteria adopted for tracking and removing duplicated flashes are similar to those in use now in the flash grouping algorithm, meaning all flashes within 1 s and closer than 10 km were considered as duplicated and removed from the dataset. Note this treatment was done on the negative flashes only since the positive rarely exhibit subsequent strokes that are likely to split flashes. Furthermore, only negative flashes were used in the relative FDE calculation according to the C4.404 method. In addition the flashes of both polarities exhibiting SMA (estimate location error) larger than 50km were considered as outliers and removed from the dataset too.

Fig. 2 shows a comparison of the initial raw flash data set and the same dataset after filtering out the duplicated and outliers flashes. Curves represent the number of flashes located over France distributed per year.

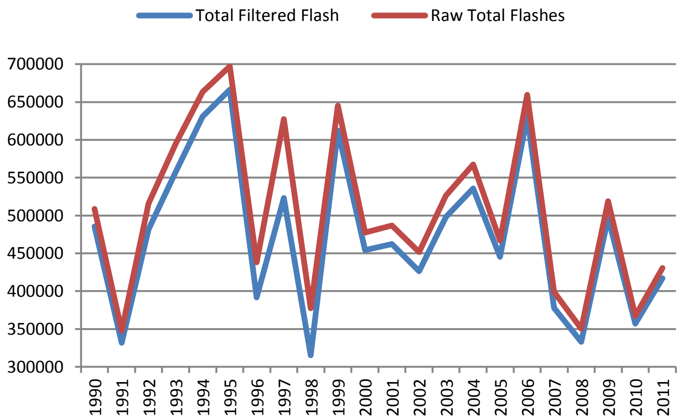

Figure 2 : Annual flash distribution in France

Fig. 3 represents the amount of data removed from the initial dataset. The blue bar graph is related to the total number of removed flashes, and the red curve shows the 'outliers' removal only. The presence of 'outliers' between 1999 and 2006 is related to the change in the settings of the SMA maximum limit. This parameter is used by the location algorithm to filter out poorly located flashes or strokes. Its value has changed from 90 km at the early stages of the system down to 7 km nowadays. The peak reached in 1999 is probably linked to the central processor upgrade and the use of a default value not optimized for an efficient filtering of outliers. Similarly a new upgrade in 2004 seemed to increase the number of outliers which proportion reduced in between. Since 2007 the SMA maximum limit has been tighten to limit 'outliers' in response of customers' requests.

The interesting and unexpected, result is the steadiness in the amount of duplicated negative flashes, as about 4% of flashes are duplicated each year taking into account 'outliers'. This seems to be independent on the central analyzer, and surprisingly this process continues today, although some decrease can be noted since LS7001 and TLP were put in operation in 2008.

The explanation may be related to the sensitivity of the flash grouping algorithm to the angle measurements before 1999 and the location accuracy since this date. The weakest subsequent return strokes are more difficult to measure because their signals often reach the sensor background noise level, affecting the measurement with errors. As a result, these measurement errors may lead the location algorithm to commit stroke location errors producing new flashes located apart the root flash. The improvement in location accuracy introduced by the new generation of sensors (LS7001) tends to limit this process, but a possible identified bug in the flash grouping algorithm may artificially separate a flash as soon as one intra cloud stroke overlap with return strokes of the flash. Finally a big increase of duplicated flashes can be seen from 1996 to 1999 possibly because the new IMPACT 141T sensors were configured with tighter flash-grouping angle criterion. This was corrected in 1999 with the change to the new central processor that processed all strokes individually. During the 20 years of observation about 6% of flashes are removed when are considering ‘outliers’ and duplicated flashes.

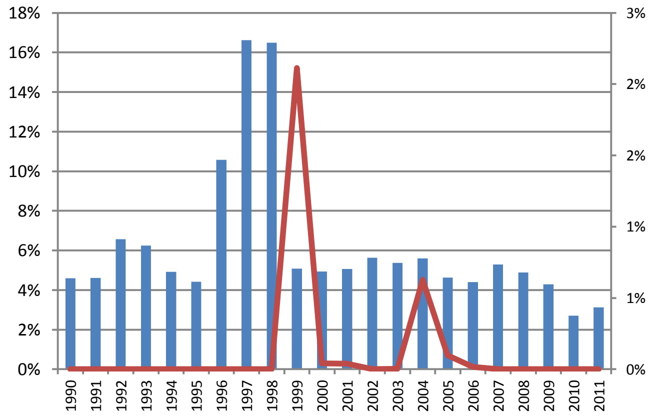

Figure 3 : Percentages of losses after filtering

Once the flash dataset is filtered out from fake data, the relative FDE computation can start. Several options are available to determine the area on which the computation can be done. Ideally the area must be as small as possible to get high resolution results. However, it is important to understand areas surfaces cannot be too small because of two main reasons. First, the scaling of the cumulative distributions is fully manual. This may be a piece of work when a lot of areas are to be studied and it turns out areas size limit is driven by practical reasons. In addition, the relative FDE estimation needs a minimum of data to be reliable implying a minimum size for the area of concern, particularly when the region naturally exhibit low thunderstorm activity. The choice made in this work is to study the yearly relative FDE evolution on each of the 23 administrative regions of France (see Fig. 4).

Their size range from 8200 to 45000 km² (average of 24500 km²) representing a good enough tradeoff for defining areas fitting the requirements of the C4.404 method. One can note the surfaces can vary a lot from a region to another but the size of the smallest one is expected to be sufficient to provide reliable results despite the smallest regions are located in the north where thunderstorms are less frequents.

Figure 4 : Administrative regions in France

The C4.404 method was applied to the 23 negative flash data subsets individually, with 2011 as the reference year. The negative peak current flashes located in each region were distributed in bins of 2kA ranging from 0 up to 100kA in a cumulative yearly distribution. Then one peak current distribution per year, from 1990 to 2011 was made and scaled in respect with 2011, which is assumed to be the best quality dataset ever collected by the system thanks to the latest LS7001 software upgrade [3][15].

In order to get an overall result for France, the 23 subsets where summed as one and the relative FDE computed based on this new global dataset, with always 2011 as reference. To illustrate the result obtained with the method, Fig. 5 shows cumulative distributions for France in 1992, 1998 and 2011 after scaling. Relative FDE values can be read directly on the vertical axe.

Each curve represents the scaled cumulative peak current distribution for one of the main technological ages of the Météorage’s LLS previously detailed. One can notice the big improvements in performance along the years thank to the evolution of the state of the art in terms of technology and rigorous operational procedures constantly improved.

The CIGRE method strives to identify an "I_0" current, above which the cumulative distributions match. For example, in Fig. 5, values are 14kA and 20kA respectively for 1998 and 1992. Note that the curves do not match perfectly in some peak current ranges above the I_0 value. This expected limiting effect is due to the different parameters which were used in the system to estimate the return stroke peak current, combined with some contamination by cloud discharges. This effect can be seen in Fig. 5 where the blue curve that is the reference is shifted upward in respect to the two others curves. It is reasonable to think that Cloud-to-Cloud misclassification can explain this point which is enhanced because of the high sensitivity of the newest LS7001 sensors.

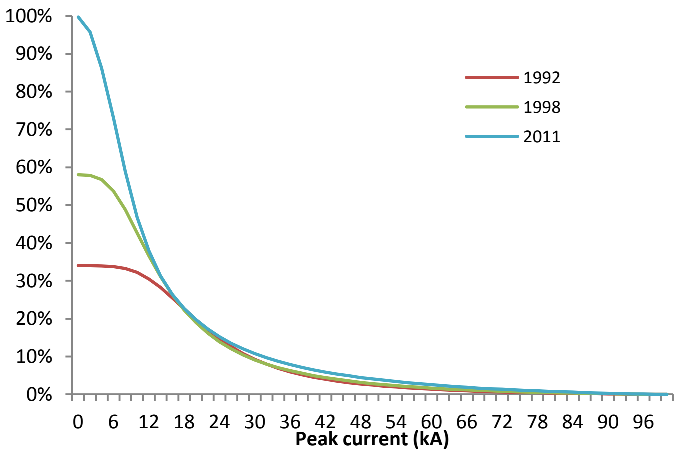

Figure 5 : Relative FDE scaled cumulative distribution

A more critical issue is the flash grouping algorithm used by APA283 during the 1990 to 1998 period, because the flash peak current was assigned with the highest return strokes amplitude. As result, the cumulative distribution cannot be compared directly without violating a fundamental rule of the CIGRE C4.404 method, meaning the peak current calculation must not vary in the tested and reference dataset. Due to this it is reasonable to think the relative FDE values for this period is highly underestimated as one can see on the blue bars in Fig. 6 representing the relative FDE in respect with 2011. This effect must be corrected before going further.

Assuming during this period only the flash grouping algorithm changed in 1999, it is possible to perform a re-scaling of the cumulative distribution in respect with the year 1999. Unfortunately, a lot of changes in the settings were done at this time affecting the overall performance that prevents the use of this dataset as a reliable reference. In replacement, it was decided to use 2000 because the LLS operation was stabilized at this time. The red curve in Fig. 6 represents the relative FDE after the correction factors for the early years were applied. It can be noted the relative FDE values better correspond, after the re-scaling, to expectations of this parameter for systems made of ALDF and APA283.

The evolution of detection efficiency performances is constant when not considering 1999 and 2007-2009 where big changes happened in the system. The first corresponds to the upgrade of APA283 to LP2000 as presented before. The second break is due to a combination of several factors. In 2007, sensors with time of arrival only (LPATS) were installed in France and used by Météorage in the operational system. As a result, those sensors mixed with IMPACT 141T generated a lot of outliers leading to drastic changes in the settings of the central analyzer in order to limit the fake flashes. Some good ones seem to have been removed too, explaining the drop in the increasing performances trend. In 2008 the upgrade of sensors to LS7001 model started, to finish in 2009 with the upgrade of the central analyzer to TLP131, explaining again the huge break in relative FDE.

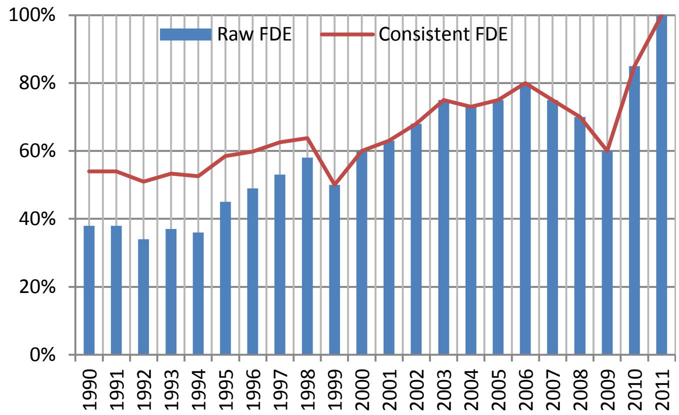

Figure 6 : Harmoized relative FDE

It is interesting to notice that any critical changes in LLS configuration may have a big effect on the general performances of the system. Monitoring FDE is then highly recommended to assess the actual effect of any change in the system.

As a conclusion to this step, the relative FDE correction factors were determined using the C4.404 method. They can be used for compensating the initial filtered dataset. Note, only results for all of France are presented here, but every region gets its own set of annual values.

B) STEP 2: GFD CALCULATION

The next step of the method consists of using the previously computed relative FDE correction factors to compensate the annual number of negative flashes. The positive flash counts are added to the corrected negative flashes in order to get the complete flash dataset for GFD calculation. The positive flashes are not compensated for relative FDE whereas rigorously they should since they suffer from detection efficiency limitation too. However, the proportion of this type of lightning is low, less than 10% so the error committed is expected to be weak. Furthermore there is no way to compute relative FDE for positive flashes, so it would have been a guess to apply any correction factor to this population of flashes.

GFD is obtained by dividing the compensated number of flashes per the surface area of each region of interest. Fig. 7 shows the yearly evolution of GFD in France with two curves. In yellow, GFD based on raw data without any treatment and filtering is plotted in order to be compared with the final GFD result after relative FDE compensation, in red on the graph.

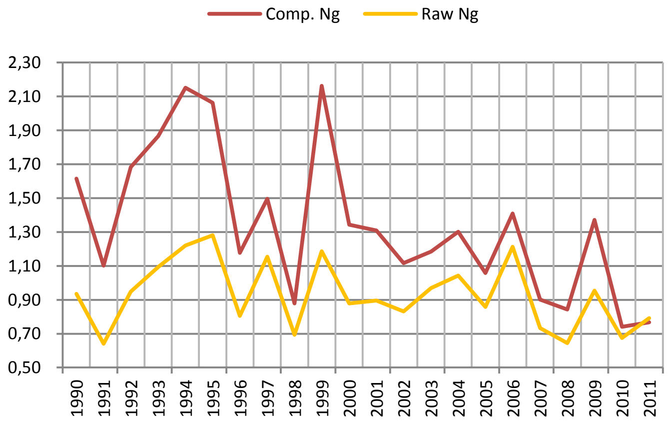

Figure 7 : GFD in France distributed by year

C) Step 3 : Transfoming GFD to GCD

It becomes obvious the number of attachment points must be taken into account in the risk assessment calculation. Since July 2011 a new method for the time of arrival measurement at sensors side is used [15] in France, leading to a significant improvement in the stroke location accuracy down to 110 m, making it possible to identify many of the separate ground contacts in flashes. To benefit from this, Météorage has developed a method to identify the multiple ground contacts in flashes from its LLS lightning dataset. It is based on a statistical clustering method, so called 'k-means', that groups return strokes in ground contacts [14] making possible the estimation of the number of ground contact per flash.

An analysis was carried out on lightning data collected from August to December 2011 in each of the 23 administrative regions in France to determine the correcting factor for multiple ground contacts per flash. Assuming this parameter does not vary from year to year it is possible to transform GFD to GCD with regional multiple ground contacts average values. Of course the unit of this new parameter must be adapted, becoming the "number of ground contacts per km² per year (GC/km²/year)".

Fig. 8 below presents the average number of ground contacts per region. This parameter varies from a region to another as expected. It is interesting to note Alsace (in orange on fig. 8) that is a rather flat region reports 1.47 ground contacts per flashes, to compare with 1.98 for Corsica (in red on fig. 8), where high mountains are located in the center of the island. This last result is fairly consistent with the study realized by K. Cummins in 2012 in the Colorado region [4]. The average number of ground contacts for France is 1.63.

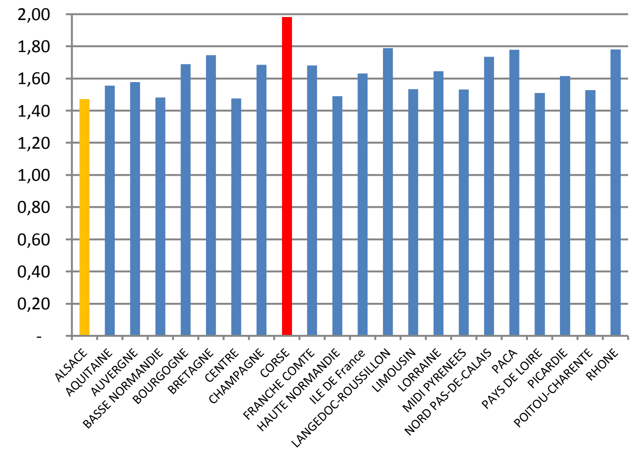

Figure 8: Ground contacts per flash average values per region.

This study was realized on a dataset restricted to half the year in 2011 that could lead to biased results if the ground contact multiplicity is dependent seasons. This result must be confirmed on a whole year dataset when data is available.

As the compensated GFD and ground contact multiplicity parameters are estimated for each region in France, it is possible to determine GCD that is equal to GFD multiplied by the ground contact multiplicity. As a result the GCD curve can be derived from the red curve in Fig. 7 while the values must simply be multiplied by 1.67, scaling the whole curve upward.

Globally, GCD average value in France between 1990 and 2011 is 2.24 GC/km²/year.

IV. Discussion

The analysis methods presented here provides our current best estimate of GCD for France at a national scale. This result has no real operational usage since GFD (or GCD) statistics are usually provided with high spatial resolution, dealing with areas of about some square kilometers only. Results on France were just to illustrate the method proposed for improving the lighting risk assessment. This method can work for any area of concern as long as the considered surface for relative FDE compensation factor calculation fit the rules of the CIGRE C4.404 method.

The computation of ground contacts multiplicity average values was done on the same regions as for relative FDE estimation, but it is reasonable to compute this parameter at very much smaller scale, even down to sub-km scales. This has the advantage not to average too much this parameter, as it is in this study, since it may vary a lot depending on local terrain effects.

There is a clear limitation for relative FDE estimation particularly when the LLS configuration has changed from time to time. The C4.404 method works well on consistent datasets, which is the case in France for more than 10 years. However, careful attention must be paid when this rule is violated. Of course the assumption made in this study for re-scaling the FDE in the early years may introduce some errors. However, the global shape of the re-scaled relative FDE curve looks good and the values obtained for ALDF/APA283 configuration fit expectations.

Finally this method allows any operator with a sufficient knowledge of their system to benefit from the whole observation dataset available leading to the use of very long term time series data and then more reliable statistics.

Surprisingly Fig. 7 shows a small trend towards lower values on recent years, meaning lightning activity in France appears to be decreasing. This may be related to a possible bias due to assumptions made for FDE calculation in the early years of the system. However, looking at the raw initial flash data collected by Météorage you can see GFD varying around a quasi-steady average value of about 0.9 Flash/km²/year, showing no obvious increase in the total number of flashes in later years despite LLS performance evolution (ALDF to LS). It can even be noted the maximum annual lightning count occurred during an early year of the system (1995).

One of possible explanation could be the fact that there were more ‘outliers’ in the early years, but after several analyses of the data quality this statement could not be validated. Another potential explanation could be duplicate flash reports, but again this hypothesis failed to be supported. Once duplicated flashes and ‘outliers’ were removed (see Fig. 2), the shape of GFD evolution curve (before compensating for FDE) showed little change.

As a first conclusion, this decreasing trend seems to be real since the FDE correction factors are consistent even though they will have errors due to assumptions in their calculation. However, this trend must be evaluated in a future work with climatological data in order to validate this result.

V. CONCLUSION

The goal of this work was to define a new way to provide reliable lightning statistics derived from long term time series observation to the lightning protection community. Nowadays, most of those statistics are derived from LLS data collected over long period of time. The issue of the impact of the evolution of LLS performances on the final GFD values becomes necessary to address in order to get reliable statistics. In addition, the flash data used to compute GFD statistics underestimates the real risk since it is well known that more than half flashes produce more than on ground contact.

This work is based on the 20 years of lightning data collected by the Météorage’s LLS which is the French national lightning observation system. During its life, several major changes in configuration and technology affected the homogeneity of the flash dataset leading to the necessity to “normalize” it before computing GFD. Then, the first step of the method consists of filtering out ‘outliers’ and duplicated flashes that pollute the initial flash dataset and may contribute to a fake enhancement of GFD. In addition, the relative FDE effect must be addressed in order to compensate the flash data from the limitations of the observing system and correct for the corresponding lack of data. This task is achieved with the use of the method developed by the CIGRE Task Force C4.404 allowing the assessment of relative relative FDE. This parameter is computed for each year of the considered period on the administrative regions dividing France in several parts. The method gives good results as long as the peak current calculation is consistent, that is the case since 1999. However, some attention must be paid for the earlier stages of the LLS because the flash peak current calculation was different leading to a re-scaling of relative FDE for 1990-1998 periods to guarantee the consistency of this crucial parameter.

Also the method suggests the use of the multiple ground contacts in the lightning density instead of flash only because of the underestimation of the risk. To do so, Météorage has developed a new algorithm based on a statistical clustering method, the ‘k-means’ that takes advantage of the recent location accuracy improvements in 2011. Regionals average yearly number of ground contacts was computed in order to transform GFD to GCD. Of course this translation leads to a new unit that is the “number of ground contacts per km² per year.

As a conclusion, this suggested method seems to give reliable and consistent results. However compensating a raw flash dataset is not only purely a matter of mathematics and models but require also a very good knowledge of the history of the observing system in order to interpret correctly some unexpected results and correct them if necessary. Of course this task is generally more difficult at the early stage of the systems but this allows using the entire information contained in databases.

The future work consists of applying this method to build a high spatial resolution lightning statistics database that will serve cartography or statistical analysis tools for the lightning protection community. The relative FDE will remain computed on the same administrative areas, but the ground contacts will be determined at the flash scale. In parallel the cross-correlation of the decreasing trend found in the compensated GFD with climatological data will be carried out.

ACKNOWLEDGMENT

Many thank to K. Cummins for his grateful help, his support and his great experience in lightning location techniques and data analysis.

REFERENCES

[1] K.L. Cummins, M.J. Murphy, “A combinated TOA/MDF technology upgrade of the US NLDN”, Journal of Geophysical Research, appendix A, 1998.

[2] K.L. Cummins, M.J. Murphy, “An overview of lightning locating systems: history, techniques, and data uses, with an in-depth look at US NLDN. IEEE Transactions on Electromagnetic Compatibility 51, 499–518.

[3] K.L. Cummins, A.E. Pifer, N. Honma, “Improved detection of winter lightning in the Tohoku region of Japan using Vaisala’s LS700X technology”, Ground’ 2010 & 4th LPE, November 2010.

[4] K.L. Cummins, “Analysis of multiple ground contacts in cloud-to-ground flashes using LLS data: the impact of complex terrain”, International Lightning Detection Conference, April 2012.

[5] G. Diendorfer, W. Schulz, M. Mair, “Evaluation of a LLS based on lightning strikes to an instrumented tower”, International Lightning Detection Conference (ILDC), Tucson, Arizona, Nov. 2000.

[6] G. Diendorfer,”Some comments on the achievable accuracy of local ground flash density values”, International Lightning Detection Conference, June 2008.

[7] A. Nag et al, “First versus subsequent return-stroke current and field peaks in negative cloud-to-ground lightning discharges”, Journal of Geophysic Research, 113, D19112,doi:10.1029/2007JD009729

[8] V.A. Rakov, “Lightning discharges triggering using rocket and wire techniques”, Recent Res. Devel. Geophysics p. 141-171, 1999.

[9] Working Group C4.404 “Cloud to ground lightning parameters derived from lightning locating systems”, CIGRE report 376, p. 89-117, April 2009

[10] M.F. Saba, M.G. Ballarotti, O. Pinto Jr,”Negative cloud to ground lightning properties from high speed video observations”, Journal of Geophysical Research vol.111, D03101, 2006.

[11] C.A. Stall, K.L. Cummins, E.P. Krider,J.A. Cramer,”Detecting multiple ground contacts in cloud to ground lightning flashes”,Journal of Atmospheric and oceanic technology vol 26, 2009.

[12] V.A. Rakov, M.A Uman,”Some properties of negative cloud to ground lightning flashes versus stroke order”, Journal of Geophysic Research 95 (D5),5447-5453,2009.

[13] W.C. Valine, E.P. Krider, “Statistics and characteristics of cloud to ground lightning with multiple ground contacts”, Journal of Geophysic Research vol 107, 2002.

[14] S. Pédeboy, “Identification of the multiple ground contacts in flashes with lightning location systems”, International Lightning Detection Conference, April 2012.

[15] N. Honma, K.L. Cummins, M.J. Murphy, « Improved lightning locations in the Tohoku region of Japan using propagation and waveform onset corrections”, 3rd International Symposium on Winter Lightning, 2011.

[16] Schulz W. and Diendorfer G.: Lightning peak currents measured on tall towers and measured with lightning location systems, ILDC, 2004a.

[17] Schulz W. and Diendorfer G.: Lightning field peaks radiated by lightning to tall towers, GROUND, 2004b.