29th International Conference on Lightning Protection

23rd – 26th June 2008 – Uppsala, Sweden

SOME COMMENTS ON THE ACHIEVABLE ACCURACY OF LOCAL GROUND FLASH DENSITY VALUES

Gerhard Diendorfer
OVE-ALDIS
g.diendorfer@ove.at
Kahlenberger Str. 2A, 1190 Vienna, Austria

Abstract: Today, values of local ground flash density (GFD) are estimated from data from lightning location systems. Lightning is a stochastic phenomenon and it's occurrence at a given location can be described by a so-called Poisson distribution. Assuming pure random nature of the lightning events from the Poisson distribution we can estimate the achievable accuracy of GFD values as a function of observation period and grid cell size. An accuracy of about ± 20% is achievable when on average more than about 80 events occurred in each grid cell. This finding suggests (1) to adjust the grid cell size A_{cell} according to the expected GFD and available observation period T_{obs} and (2) to consider an uncertainty range of at least ± 20% for any N_g value that is based on LLS data by counting lightning events in defined grid cells.

1 INTRODUCTION

National and international standards for lightning protection (e.g. EN 62305-2:2006 [1]) provide methods and tools to evaluate the risk of a lightning strike to a given object. Ground flash density (N_g), defined as the number of lightning flashes per km^2 and per year, is one of the fundamental input parameters for such risk analyses. In many areas of the world N_g is derived from data form lightning location systems (LLS). If LLS data are not available, in temperate regions N_g may be estimated by:

$$N_g \approx 0.1\ T_d \qquad (1)$$

where T_d is then number of thunderstorm days per year obtained from isokeraunic maps. Typical values for N_g in temperate regions are between 1 and 5. In tropical regions N_g values of up to 47 flashes per km^2 and per year are reported by Pinto et al. [2].

Lightning is a highly stochastic phenomenon and hence we have to consider some fundamental rules for such random events. In this presentation we will discuss some effects that finally limit the achievable accuracy of the value of N_g when determined from LLS data. For all the following calculations we assume that lightning flashes occur purely random in a given area and no meteorological and/or topographical effects are causing local variations of flash density.

2 POISSON DISTRIBUTION AND LAW OF RARE EVENTS [3]

Poisson distribution is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed period of time if these events occur with a known average rate and independently of the time since the last event. Hence this distribution is an appropriate model for the occurrence of lightning flashes.

If the expected number of occurrences in a given interval is λ, then the probability that there are exactly *k* occurrences (*k* being a non-negative integer, *k* = 0, 1, 2, ...) is equal to

$$f(k,\lambda) = \frac{\lambda^k . e^{-\lambda}}{k!} \qquad (2)$$

where
k is the number of occurrences of an event
λ is a positive real number, equal to the expected number of occurrences observed during the given interval.

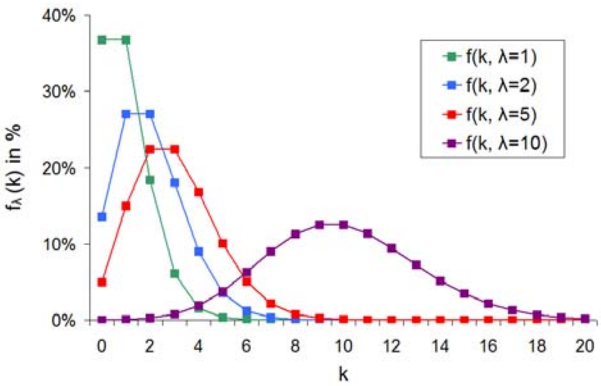

Fig. 1a: Poisson Distribution function for λ = 1, 2, 5, 10

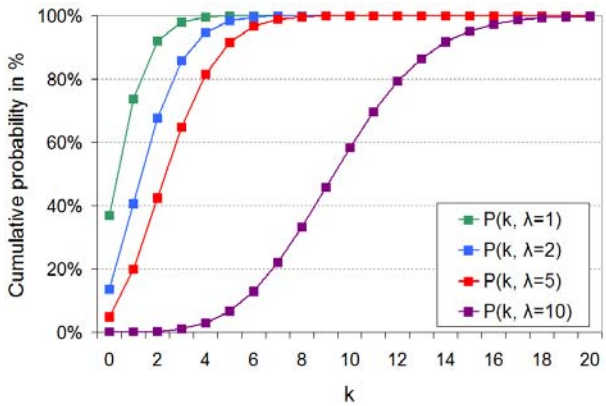

Fig. 1b: Cumulative distribution function for λ = 1, 2, 5, 10

For sufficiently large values of λ, the normal distribution with mean $\mu=\lambda$ and variance $\sigma^2 = \lambda$, is an excellent approximation to the Poisson distribution.

In terms of ground flash density N_g, when we assume an average of $N_g = 1$ flash per year in a given area and when we are interested in the number of events occurring in a one year interval in that area, we have to use the Poisson distribution with $\lambda = 1/1 = 1$ (green line in Fig 1). When we consider a 5 years observation period in the same area the line for $\lambda = 5/1 = 5$ (red line) is applicable. The same red line ($\lambda = 5$) is applicable, when we have a one year observation period in an area of $N_g = 5$ flashes per year. This equivalence (k years of observation and $N_g = 1$ being equal to one year of observation and $N_g = k$) is a consequence of the basic assumption that all events are occurring completely independent from each other.

In Table 1 the probability for k events being observed in an area of $\lambda = 1$, $\lambda = 2$ and $\lambda = 5$ expected events are listed. In case of $\lambda = 2$ (e.g. we monitor over a 1 year period an area of 1 km^2 assuming $N_g = 2$ flashes per year per km^2) Table 1 and also Fig. 1 show, that there is a probability of 14 % of seeing NO flash in the area, a 27 % probability to observe only 1 flash, a 27 % probability to monitor the expected 2 flashes, an 18 % probability to observe 3 flashes, and so on. There is still a 1 % chance to observe 6 flashes in the given 1 km^2 area.

Table 1: Probability of k events

k	λ = 1	λ = 2	λ = 5
0	37%	14%	1%
1	37%	27%	3%
2	18%	27%	8%
3	6%	18%	14%
4	2%	9%	18%
5	0%	4%	18%
6	0%	1%	15%
7	0%	0%	10%
8	0%	0%	7%
9	0%	0%	4%
10	0%	0%	2%

With a simple simulation routine we can demonstrate the random nature of N_g values. Fig. 2a shows one example of such simulation run when 10.000 flashes are randomly placed over an area of 10.000 km^2. The assumptions represent a uniform N_g of 1 flash per km^2 over the entire area. Fig. 2b shows the resulting histogram when we count the number of cells with k = 0, 1, 2, 3., ... 10 flashes placed in each of the 100.000 cells by the random simulation. In this simulator run 3.661 (36,6 %) of the 10.000 cells were without any lightning (k = 0), 3.705 (37 %) cells showed 1 flash, 1.844

(18,4 %) showed 2 flashes, etc. These results are in almost perfect agreement with the calculated numbers for the "Poisson Distribution" for λ = 1 in Table 1. In this particular run of the simulator two 1 km^2 cells showed 7 and one cell actually showed 8 flashes, respectively, being 7 or 8 times more than the assumed N_g of 1 flash/km^2. This clearly shows that there is some chance of "unexpected" high numbers of lightning strikes to single locations as a result of the random nature of lightning discharges. In statistics this effect is referred to as "Poisson Clumping".

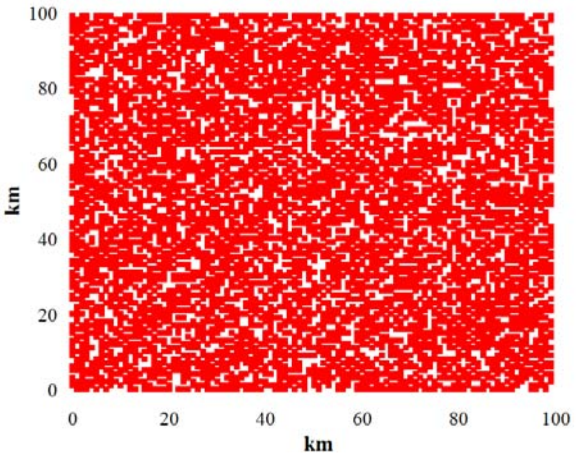

Fig. 2a: Simulation result when 10.000 flashes are randomly distributed over a 10.000 km^2 area (N_g = 1)

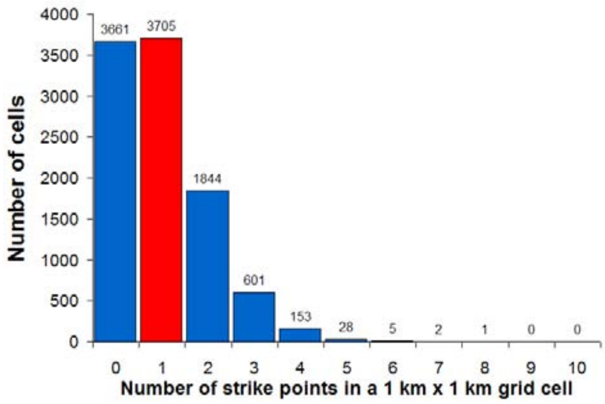

Fig. 2b: Histogram of cell counts as a function of strike points within a grid cell

Fig. 2b also represents the so called "Law of Rare Events" corresponding to the Poisson distribution for λ=1. Only about 1/3 of the cells are actually struck by 1 flash as expected when 10.000 flashes are placed on 10.000 km^2. About 8 % of the cells show $N_g \geq 3$.

For a Poisson distribution the parameter λ is not only the mean number of occurrences k, but also its variance and thus the number of observed occurrences fluctuates about its mean λ with a standard deviation σ_k

$$\sigma_k = \sqrt{\lambda} \qquad (3)$$

Now we can use the "Coefficient of Variation" (CV), which is a measure of dispersion of a probability distribution. It is defined as the ratio of the standard deviation σ to the mean λ:

$$CV = \frac{\sigma}{\lambda} = \frac{\sqrt{\lambda}}{\lambda} \qquad (4)$$

CV is a dimensionless number and is often reported as a percentage (%) value by multiplying the above calculation in Eq. (4) by 100 and this is referred to as the "Relative Standard Deviation (RSD)". Fig. 3 shows a plot of RSD as a function of the mean value λ and obviously there is a fast decay of RSD for values of λ < 10.

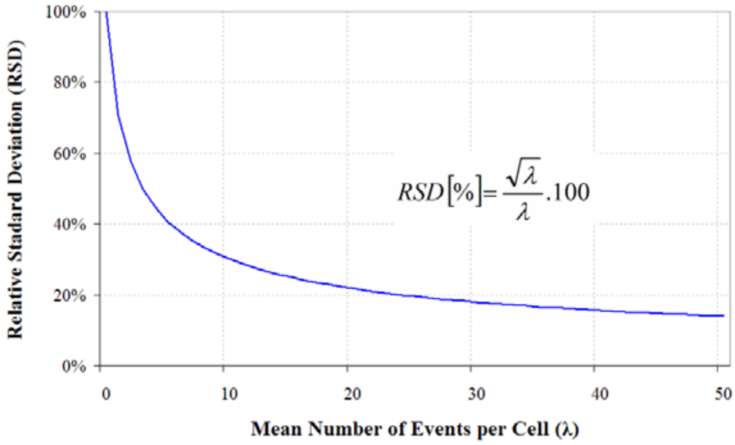

Fig. 3: Relative Standard Deviation (RSD %) as a function of mean number of events per cell

For a Poisson distribution we can estimate the 90% - Confidence Interval ($\lambda_{min} \leq \lambda \leq \lambda_{max}$) for λ with Eq. (5) (see Eq. 1.270 in [3]) as

$$\lambda_{min} = \left(\frac{1{,}645}{2} - \sqrt{k}\right)^2 \leq \lambda \leq \left(\frac{1{,}645}{2} + \sqrt{k+1}\right)^2 = \lambda_{max} \qquad (5)$$

where k is the number of observed events. To see the relative uncertainty as a function of observed events k, we have calculated and plotted in Fig. 4 the lower and upper percentage limits of the relative confidence interval in percent.

As an example in case of k = 20 observed events (20 flashes in a given area and time), with Eq. (5) we calculate λ_{min} = 13.3 and λ_{max} = 29.2. This means that there is a 90% probability that the true mean value of λ is in the range from 13.3 to 29.2. This is equal to k.(1 - 0.33) ≤ λ ≤ k.(1 + 0.46) or -33% less or +46% more than the observed number k = 20, also indicated in Fig. 4.

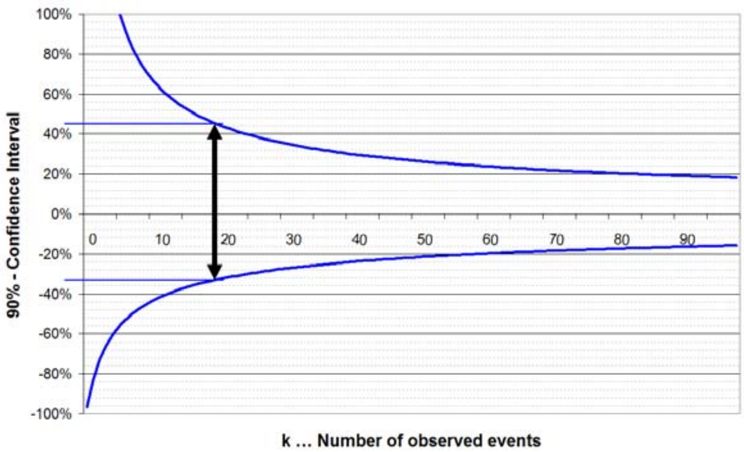

Fig. 4: 90% - Confidence Interval as a function of observed number of events k, given as percentage of the observed number of events k

For larger numbers of observation (k > 50) the 90% - Confidence interval becomes about ± 20%. For smaller numbers of events (e.g. k = 5) the relative uncertainty is rapidly increasing and exceeds +100%. In Table 2 we have summarized a few calculated values.

Table 2: 90% - Confidence Interval range (see also Eq. 5 and Fig. 4)

k	λ_{min}	λ_{max}	neg. confidence interval	pos. confidence interval
5	2.00	10.71	-60 %	+114 %
10	5.47	17.13	-45 %	+71 %
15	9.31	23.26	-38 %	+55 %
20	13.32	29.21	-33 %	+46 %
30	21.67	40.84	-28 %	+36 %
50	39.04	63.42	-22 %	+27 %
80	**65,96**	**96,48**	**-18 %**	**+20 %**
100	84.23	118.21	-16 %	+18 %

3 GROUND FLASH DENSITY ESTIMATES

Regional N_g values are typically computed from a large data set of flashes located by a LLS by dividing the region of interest into small cells (rectangles or squares) and accumulating the total number of flashes occurring in each grid cell over a time interval of interest. Now this brings us to the problem of selecting an adequate grid size for this regional N_g analysis. The average location accuracy of the LLS in the selected region is a lower bound for the selectable grid size. Today's LLS achieve median location accuracies in the range of 500 – 1000 m and hence the finest grid size applicable for N_g estimated should not be smaller than 1 km x 1 km.

The time period covered by the analyzed lightning data should be as long as possible to average temporal fluctuations of lightning activity. During the selected period the LLS should have had more or less uniform and constant flash detection efficiency (DE).

Typical N_g values in temperate regions are in the range of 1 to 5 flashes per km^2 and per year and hence the number of events per grid cell is relatively small when we select a 1 km x 1 km grid size. If we require an uncertainty of less than ±20 % (90% - confidence interval) for N_g values we should have more than 80 events (see Fig. 4) per grid cell and hence

$$N_g \cdot T_{obs} \cdot A_{cell} \geq 80 \qquad (6)$$

is suggested, where
T_{obs} is the observation period in years, and
A_{cell} is the area of a grid cell in km^2.

Hence, in a region of true value N_g=5 about 80 events per cell could be achieved either by an observation period of 15 years with a 1 km^2 grid cell area or with a reduced observation period and adequately increased grid size (e.g. 4 years for 4 km^2 grid size).

Fig. 5 shows a histogram of the annual number of flashes located by ALDIS within the 84.000 km^2 territory of Austria.

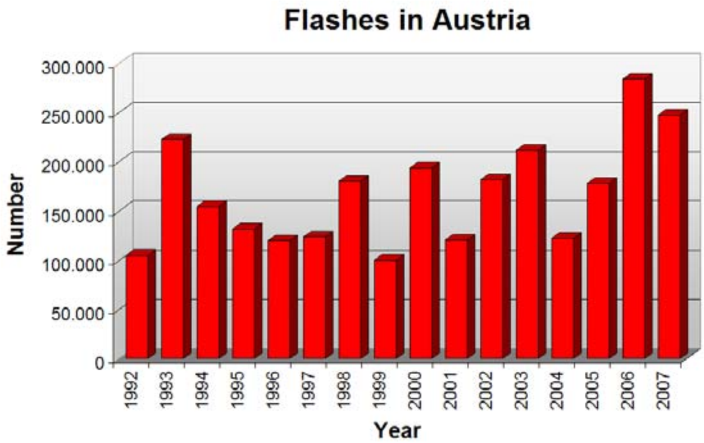

Fig.5: Annual number of flashes located in Austria

Obviously the annual number of flashes in Austria ranges from 100.000 to 270.00, by a ratio of 2.7, as a result of the annual variations of regional weather conditions affecting the occurrence, duration and intensity of thunderstorms. For the data shown in Fig. 5 we determine a mean of 167.300 flashes per year (STD = 54.500).

4 CONCLUSIONS

Based on the random nature of lightning any "measured" ground flash density is of limited accuracy. Achievable uncertainty is in the range of ±20% even when pure random occurrence of the lightning strikes is assumed and topographical and meteorological effects are neglected. Those effects eventually introduce additional uncertainties in the measured GFD values. These limits of accuracy should be considered when GFD values are applied in complex risk analysis algorithms. We have to note that there are other parameters affecting the accuracy of N_g values from LLS data, as the detection efficiency of the LLS or the spatial and temporal parameters applied in the stroke-to-flash grouping algorithm (see e.g. [4]).

5 REFERENCES

[1] EN 62305-2:2006, Protection against lightning - Part 2: Risk management

[2] Osmar Pinto Jr., Iara R.C.A. Pinto, Kleber P. Naccarato, Simone A. de M. Garcia (2007) Maximum Cloud-to-Ground Lightning Densities Obeserved by Lightning Location Networks in the Tropics. IX International Symposium on Lightning Protection (SIPDA), Brazil

[3] Sachs, L. (1991), Angewandte Statistik, 7. Auflage, Springer Verlag

[4] Schulz, W., K. Cummins, G. Diendorfer, M. Dorninger: Cloud-to-ground Lightning in Austria: A 10-years Study using data from a lightning location system. J. Geophys. Res., 110, D09101, doi:10.1029/2004JD005332.