Probabilistic interpretation of LDN confidence ellipses with reference to forensic applications

Michael D. Grant*†, Ian S. McKechnie*‡, Ian R. Jandrell*, and Ken J. Nixon*
*School of Electrical and Information Engineering
University of the Witwatersrand, Johannesburg, Republic of South Africa
Email: michael.grant@wits.ac.za
†CBI-electric: low voltage
Isando, Republic of South Africa.
‡Innopro, PO Box 9288, Centurion, 0046, Republic of South Africa.

***Abstract*—Data from lightning detection networks is often used for forensic purposes: to validate insurance claims or even to determine the cause of death. Stroke location and current estimates are subject to measurement error, and the dilution of precision is reported in terms of a median confidence ellipse and χ^2 distribution. A method is presented to derive the probability density function from the reported statistics, and hence calculate cumulative probability densities. Two hypothetical cases are presented, with real lightning detection network data, to illustrate the obstacles constraining direct forensic application of the data.**

I. Introduction

The understanding of the climatology of lightning has been greatly enhanced through the analysis of data recorded by Lightning Detection Networks (LDNs) around the globe. With some networks claiming flash detection efficiencies in excess of 90% and location accuracies below 500 m [1], the data has been applied to a wide variety of applications (like fault correlation on long transmission lines); although these detection systems were not primarily designed for such applications, but rather for meteorological investigations [2], [3].

Data, with sufficient accuracy for climatological studies, is often seen as *"the smoking gun"* on scenes where lightning damage or injury has occurred. Unfortunately many misconceptions about the operation and limitations of these LDNs and their use, continue to persist; this paper investigates the specific use of this data in forensic analysis.

Two hypothetical scenarios are presented: that of damage to an industrial plant and personal injury; using data with the same parameters as that detected by combined time of arrival and magnetic direction finding LDNs. This paper is intended for engineering professionals who incorporate lightning detection data in their forensic studies.

II. Summary of LDN data

A. Flash vs. Stroke, and aggregation

A lightning flash consists of several discrete stages: first the formation of the ionised channel, followed by the impulse current flowing through the channel. LDN sensors detect the electromagnetic wave radiated from the impulsive current flowing through the channel; and may be coupled to either the VLF, LF, or VHF parts of the spectrum [4], [5].

The first current impulse is termed the return stroke and usually has the largest peak current. There may be more charge in the vicinity of the origin of the initiating charge centre, in this case a dart leader follows the previously ionised channel and shares the previous attachment point on the the ground. The impulsive current that flows after a dart leader is termed a subsequent stroke. Subsequent strokes produce similar electromagnetic impulses to return strokes and may also be detected by LDN sensors.

The typical duration of a lightning flash, consisting of a few lightning strokes, is in the order of a few hundred milliseconds. It is for this reason that strokes are aggregated into flashes which provides a more intuitive description for a lightning event. The aggregation algorithm employed by most LDNs is described in detail by Cummins [1], [5]. The algorithm aggregates strokes that occur close together in both time (total flash time $<$1 s and preceding stroke $<$500 ms) and space (computed location $<$10 km and confidence region overlap within 50 km) into a single flash [1], [5].

B. Detection efficiency

The Detection Efficiency (DE) of an LDN is a measure of how many lightning events are recorded by the network versus the true number of lightning events. Since network sensors respond to the radiated electromagnetic impulse from the discharge, the propagation of the wave over ground (or water) where conductivity, permittivity, and permeability change has the greatest influence.

DE is improved by reducing the baseline distance between sensors and locating sensors around severe terrain.

Since LDN sensors primarily detect strokes, a flash will have a higher DE because there are multiple detection opportunities. Individual strokes have a much lower DE. For example the North American Lightning Detection Network (NALDN) has a stroke DE in the region of about 52% to 85% (depending on the type of lightning) [5]. Similarly the NALDN has a flash DE in the region of 91% to 95% [5]. As the number of strokes in a lightning flash is reduced, so the flash detection efficiency approaches that of a stroke. Therefore single stroke flashes (only consisting of a return stroke) have a relatively low detection efficiency.

978-1-4577-1466-5/11/$26.00 ©2011 IEEE

C. Location accuracy

Location accuracy is represented by three parameters: the semi-major and semi-minor axes of the median confidence ellipse, and the χ^2 parameter.

1) Probability model: The probability model employed in many combined Time of Arrival (TOA) and Magnetic Direction Finding (MDF) LDNs is based on the work of Stansfield who assumed that the random error obeyed a Gaussian distribution [1], [6].

LDN sensors that determine position by comparing the difference in arrival time of the radiated electromagnetic impulse, use a time measurement which is synchronised with a GPS clock installed in the sensor. It has recently been shown that the error of these GPS clocks is not Gaussian [7].

D. Confidence ellipse

Stroke location is reported in terms of the fitted agreement between the time of arrival and direction measurements. Around this point an ellipse may be drawn which provides a graphical estimation of the quality of location fit.

The ellipse is determined from a probability distribution: a bivariate Gaussian distribution is fitted to the parameters and scaled as a function of the quality of fit. This distribution is described by the following equation:

$$P(\mathbf{x}) = \frac{1}{2\pi\sqrt{|\mathbf{\Sigma}|}} e^{\frac{-(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\,\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}} \tag{1}$$

Where:

$\mathbf{x}$	Position matrix: $\mathbf{x} = \begin{bmatrix} x & y \end{bmatrix}$
$\boldsymbol{\mu}$	Mean matrix: $\boldsymbol{\mu} = \begin{bmatrix} \mu_x & \mu_y \end{bmatrix}$
$\mathbf{\Sigma}$	Co-variance matrix: $\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$
$P(\mathbf{x})$	Probability density at point x, y

The confidence ellipse is calculated as the intersection of this distribution through a median plane. Hence there are equal probabilities that the real stroke occurred inside the ellipse and outside the ellipse.

The confidence ellipse is described in terms of the lengths of the semi-minor and semi-major axes, and the bearing of the semi-major axis; and by scaling these lengths it is possible to increase the ellipse to an arbitrary confidence interval (99% for example). A factor of ∞ is required to scale the ellipse to 100%; and this has no physical interpretation. Table I lists median scaling factors to increase or decrease the ellipse probability.

As the ellipse lengths are decreased, the closer the ellipse approximates the centre point, or the stroke "location". Hence:

$$\lim_{A \to 0} \iint_{A} P(x, y)\,\mathrm{d}x\mathrm{d}y = 0 \tag{2}$$

Where:

A	Area within confidence ellipse
$P(x, y)$	Probability density at point x, y

So it is possible to say, with absolute certainty, that the stroke did not attach at the computed stroke location.

TABLE I
CONFIDENCE ELLIPSE SCALING FACTORS, FROM MEDIAN CONFIDENCE ELLIPSE. ADAPTED FROM [8].

Confidence ellipse size (%)	Scaling factor
50.000	1.000
75.000	1.414
90.000	1.823
95.000	2.079
99.000	2.578
99.900	3.157
99.990	3.646
99.999	4.076

1) Physical interpretation of median ellipse: Figure 1 shows the computed stroke location as a black dot, surrounded by the median confidence ellipse. The computed stroke location does not lie on the object of interest, which is represented by a rectangle. The rectangle represents the physical footprint of the object of interest–it is considered that lightning may have attached directly to this object. There is a section of common area enclosed both by the median confidence ellipse and the outline of the object of interest i.e. the intersection of these areas.

However, the ellipse is only that of the median, therefore there are equal probabilities that the stroke occurs within the ellipse, or outside of it. There is no direct physical interpretation for the common area, described by the intersection of the ellipse and object areas.

If the confidence ellipse were scaled by one of the factors in Table I (99% ellipse for example) then there would be a larger area in the intersection between the ellipse and object of interest. However there would also be more unrelated area covered by the ellipse. Regardless of the probability enclosed by the confidence ellipse (median, or even 99%) there is no direct physical interpretation for this measure.

It is possible, however, to integrate the spatial probability density function, described by Equation 1, over the object of interest. This number then yields a direct probability that the stroke occurred within the area of the object of interest, which may be compared with other strokes computed to have occurred near the object of interest. The integral of Equation 1 does not have an analytical solution, and this integration must therefore be performed numerically.

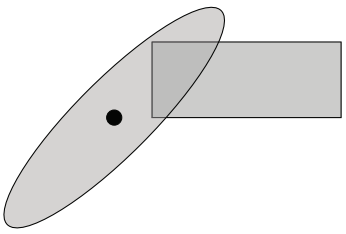

Fig. 1. Illustration of intersection between median confidence ellipse and object of interest (rectangular). Computed stroke location shown as black dot in the centre of the ellipse.

2) Determination of the probability density function from confidence ellipse parameters: Points of equal probability density, as described by Equation 1 form concentric ellipses, centred at $\boldsymbol{\mu}$.

These ellipses are described by the following equation:

$$c^2 = (\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \tag{3}$$

The LDN system reports confidence ellipses in terms of the lengths of the semi-major and semi-minor axes; and the bearing of the semi-major axes. Using the principal axis theorem it is possible to compute $\boldsymbol{\Sigma}$ (the co-variance matrix). The mean matrix, $\boldsymbol{\mu}$, is simply the stroke location.

No clear procedure to compute the parameters of the probability density function from the parameters of the confidence ellipse is available in literature. The following method is proposed and discussed in the subsequent sections, and consists of finding a set of basis vectors that describe the bearing of the semi-major axis. The lengths of the semi-major and semi-minor axes correspond to the eigenvalues of the co-variance matrix. These steps may be generalised into m dimensions.

An orthogonal matrix, consisting of unit orthonormal basis vectors is first computed from the bearing:

$$u_{11} = \frac{\tan\theta}{\sqrt{1 + \tan^2\theta}} \tag{4}$$

$$u_{12} = \pm\sqrt{1 - u_{11}^2} \tag{5}$$

Now $\mathbf{u_2}$ must be orthogonal to $\mathbf{u_1}$ so $\mathbf{u_1} \times \mathbf{u_2} = 0$. To expand into additional dimensions, the condition $\prod_{i=1}^{m} \mathbf{u_i} = 0$ must be satisfied.

$$u_{22} = \frac{1}{\sqrt{1 + \left(\frac{u_{12}}{u_{11}}\right)^2}} \tag{6}$$

$$u_{21} = \pm\sqrt{1 - u_{22}^2} \tag{7}$$

The orthogonal matrix is then $\mathbf{S} = \begin{bmatrix}\mathbf{u_1} & \mathbf{u_2}\end{bmatrix}$, and the ellipse can be expressed in terms of the basis vectors:

$$c^2 = \left(\mathbf{S}^{\mathrm{T}} (\mathbf{x} - \boldsymbol{\mu})\right)^{\mathrm{T}} \mathbf{D} \left(\mathbf{S}^{\mathrm{T}} (\mathbf{x} - \boldsymbol{\mu})\right) \tag{8}$$

Where:

$\mathbf{D}$ Diagonal matrix of co-variance matrix eigenvalues. $\mathbf{D} = \begin{bmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{bmatrix}$

In this form it is apparent that the semi-major and semi-minor axes lengths occur when the component of either of the basis vectors is zero. And so the eigenvalues correspond with these lengths:

$$l_{\text{major}} = \sqrt{\frac{c^2}{\lambda_1}} \tag{9}$$

$$l_{\text{minor}} = \sqrt{\frac{c^2}{\lambda_2}} \tag{10}$$

Equation 8 may be rewritten in the same form as Equation 3, and thus an expression for the co-variance matrix is derived:

$$c^2 = (\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}} \mathbf{S}\mathbf{D}\mathbf{S}^{\mathrm{T}} (\mathbf{x} - \boldsymbol{\mu}) \tag{11}$$

$$\Rightarrow \boldsymbol{\Sigma}^{-1} = \mathbf{S}\mathbf{D}\mathbf{S}^{\mathrm{T}} \tag{12}$$

Thus all the parameters of the probability density function, given by Equation 1 are satisfied.

E. Peak current

Peak current of a lightning stroke, is estimated from the Range Normalised Signal Strength (RNSS) [1]. More precisely, the magnitude of the vertical component of the electric field is normalised against a 100 km standard and peak current is estimated from this value.

In forensic analysis, peak current is an important parameter because it is an estimate of the energy of the impulse. However few LDNs have been compared against direct lightning current measurements; the maximum directly recorded peak currents are yet to exceed 60 kA [2], [5], [9], [10].

III. Hypothetical cases

Two hypothetical cases are presented in this section, and the scenarios are illustrated in Figure 2. The scenarios have been constructed as a means to illustrate typical problems encountered during forensic investigations. Real LDN data has been used, the data was selected by choosing an area a thunder storm had passed through and arbitrarily placing two hypothetical objects in the path.

The plant was placed near a cluster of 16 strokes, aggregated by the LDN into 5 flashes: three consisting of four strokes each, and two double stroke flashes. These strokes all occurred within a 10 minute window. To the south of a plant a three stroke flash occurred, 35 minutes after the cluster.

The four strokes to the west of the person are aggregated by the LDN into a single flash; this flash occurred 15 minutes before the cluster near the plant. The remaining strokes are distributed throughout the hour window.

A. Industrial damage

In this scenario the case of an industrial plant is considered. It was reported that the failure of the plant occurred as the result of a direct lightning flash. Radar data from the national meteorological service shows a storm crossing the area of the plant. A set of lightning strokes for the area (including confidence ellipses, bearings, and χ^2 values) was also supplied by the national meteorological service. The modelled median location accuracy for the area is 500 m and flash detection efficiency is predicted to be in excess of 95%.

All reported strokes within 10 km of the plant on the day in question are plotted in Figure 2. None of the computed stroke locations are within the area of the plant. Two median confidence ellipses intersect the plant, and the corresponding computed stroke locations for these two ellipses are the closest of the 16 stroke cluster to the plant. Both of these strokes are aggregated by the network into the same four stroke flash.

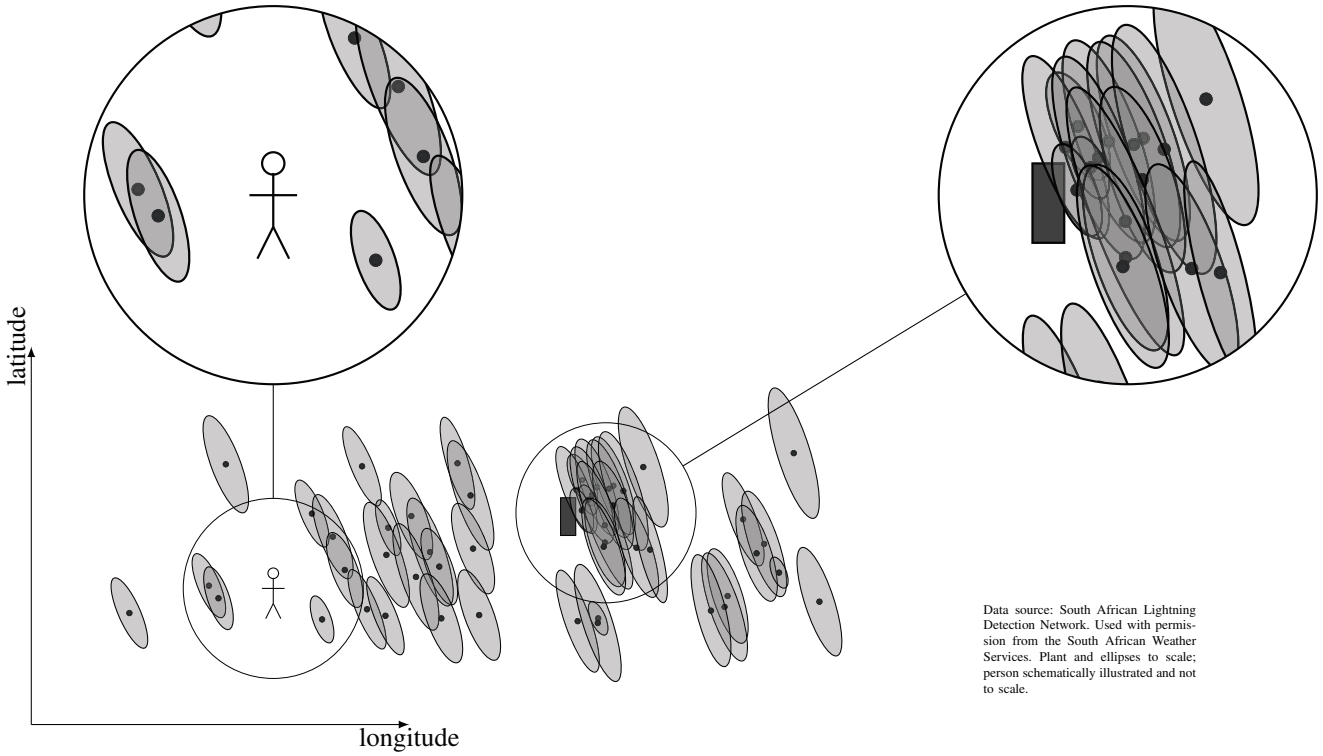

Fig. 2. Schematic lightning strokes occurring in a 1 hour window and within 10 km of a plant and person, including computed locations and median confidence ellipses. The flash density of the area is between 10 to 12 flashes/km^2/year.

If the confidence ellipses are scaled from 50% to a 99% confidence interval, then 10 ellipses intersect or enclose the plant area. These 10 strokes are aggregated by the network into four separate flashes.

Computing the cumulative probability density within the area of the plant determines that the two median confidence ellipses that intersect the area of the plant have the highest cumulative area probabilities. In the zoomed view of the problem geometry, shown in Figure 2, a small confidence ellipse to the immediate north west of the plant corresponds to the third most cumulative probability density.

B. Personal injury

In this scenario the body of a 1.6 m tall person was found in the open savannah. There were no eye witnesses to the event, and cause of death was determined, by autopsy, as direct lightning attachment to the person.

The position of the person is not found within any of the median confidence ellipses. Since the location of the person effectively represents a point source the same limit restriction described by Equation 2 is present when calculating any of the strokes' cumulative probability integrals, yielding nonsensical cumulative probabilities. None of the scaled 99% confidence ellipses enclose the position of the person either.

IV. The "smoking gun"?

The two hypothetical scenarios illustrate some of the obstacles constraining the direct application of LDN data in forensic investigations.

In the case of the plant, there are numerous stroke candidates clustered tightly around the plant. Direct physical interpretation of the stroke median confidence ellipses is not possible: the two median ellipses that intersect the outline of the plant may seem like more likely candidates, however three strokes have similar *cumulative* probabilities and for one stroke the median confidence ellipse does not intersect the plant. In this case the aggregation of two strokes into the same flash may provide a higher cumulative flash probability. Although it must be remembered that networks claim median location accuracies in the order of 500 m, but use even wider flash aggregation criteria [1]. A more appropriate approach would be the ranking of the strokes based on their cumulative stroke and flash probabilities, but it is not possible to select a single stroke or flash that may have been responsible.

In the case of the person, there are no likely stroke or flash candidates. The cumulative probability approach is not applicable as the numerical integration area presented by the person effectively is a point; and therefore has a zero cumulative probability. What can be inferred with a high degree of confidence is that the conditions were severe, with lightning occurring all around the person. Thus the risk to life was potentially grave, but no overlapping lightning stroke information exists. The absence of overlapping LDN data does not preclude the involvement of lightning. This may in fact be an indication that the DE in a specific area is less than that predicted by the performance model or that the event falls within the non-detection performance characteristic. This of

course applies to both scenarios.

Each stoke must be considered as the full set of computed location, confidence ellipse, χ^2, and peak current. The data set may be reduced by removing extreme outliers, where strokes occurring some distance away from an event may be ignored. Correlation with calibrated clock sources provides a most effective selector, but is not always available as field information may be scarce.

V. Conclusion

A single smoking gun does not exist; but rather a list of probable strokes. Even if the case of a single unambiguous stroke to a point existed, the probability of an undetected stroke being the real cause is not zero.

It is not possible to select a stroke based on the intersection of a collective area and confidence ellipse because there is a non-constant probability density within the ellipse–direct interpretation is not possible. A median confidence ellipse is also a very small probability on which to base an analysis, since there is a coin toss between being inside or outside the ellipse. A more appropriate analysis includes calculating the cumulative probability density within a collection area.

Detection efficiency is not perfect and this measure of network performance varies geographically due to variations in inter-sensor spacing and even terrain features. Detection efficiency, in conjunction with other performance measures like location accuracy, need to be verified since one network does not have the same performance the same as another network. Hence verification studies are critically important to confirm the performance of specific LDN's.

This paper has noted particular and limiting constraints on the use of lightning detection networks in forensic applications, particularly when such data is considered in isolation.

The holistic investigation of the entire scene, including the presence of direct physical evidence of lightning attachment and other corroborating evidence is required to forensically analyse the hypothesis of an event or incident potentially ascribed to lightning.

Acknowledgment

The authors would like to acknowledge the South African Weather Services for use of their data in the hypothetical scenarios, and thank them for their support of the Lightning/EMC research group. The authors would like to thank Eskom TESP, THRIP and the NRF for funding this research. The South African Air Force is thanked for their assistance in the triggered-lightning experiments and Doug Kay from Lectro-Tech for the supply of materials used to construct the lightning protection system for the triggered-lightning experiments. Campbell Scientific Africa is thanked for their generous donation of meteorological equipment. The authors also express their gratitude to CBI-electric: low voltage for funding this work, the Lightning/EMC Research group and for support of the CBI-electric Chair of Lightning.

References

[1] K. Cummins, M. Murphy, E. Bardo, R. Hiscox, W.L. Pyle, and A. Pifer, "A combined TOA/MDF technology upgrade of the U.S. National Lightning Detection Network," *Journal of Geophysical Research*, vol. 103, no. D8, pp. 9035–9044, April 1998.

[2] R. Wacker and R. Orville, "Changes in measured lightning flash count and return stroke peak current after the 1994 U.S. National Lightning Detection Network Upgrade: 1. Observations," *Journal of Geophysical Research*, vol. 104, no. D2, pp. 2151–2157, January 1999.

[3] R. Orville, "Development of the National Lightning Detection Network," *Bulletin of the American Meteorological Society*, pp. 180–190, February 2008.

[4] V. Rakov and M. Uman, *Lightning Physics and Effects*, 1st ed. The Edinburgh Building, Cambridge, CB2 2RU, United Kingdom: Cambridge University Press, 2003, ISBN 0 521 58327 6.

[5] K. Cummins and M. Murphy, "An overview of lightning locating systems: History, techniques and data uses, with an in-depth look at the U.S. NLDN," *IEEE Transactions on Electromagnetic Compatibility*, vol. 51, no. 3, pp. 499–518, August 2009.

[6] R. Stansfield, "Statistical theory of df fixing," *Proceedings of Institute of Electrical Engineers*, pp. 762–770, March 1947.

[7] S. Bednarz, "Adaptive modeling of gps receiver clock for integrity monitoring during precision approaches," Master's thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, 2004.

[8] Vaisala, "Vaisala thunderstorm total lightning sensor," *User guide*, January 2008.

[9] V. Idone, A. Saljoughy, R. Henderson, P. Moore, and R. Pyle, "A reexamination of the peak current calibration of the national lightning detection network," *Journal of Geophysical Research*, vol. 98, no. D10, pp. 18 323–18 332, July 1993.

[10] J. Jerauld, V. Rakov, M. Uman, K. Rambo, and D. Jordan, "An evaluation of the performance characteristics of the U.S. National Lightning Detection Network in Florida using rocket-triggered lightning," *Journal of Geophysical Research*, vol. 110, no. D19106, pp. 1–16, October 2005.