

Improvements on Lightning Density Estimation Based on Analysis of Lightning Location System Performance Parameters: Brazilian Case

Vandoir Bourscheidt, Osmar Pinto, Jr., and Kleber P. Naccarato

Abstract—Density maps are one of the most common and powerful lightning data applications, and they are more efficient the more detailed they are. When working with CG lightning data from lightning location systems, some aspects must be included in the analysis to overcome network performance variations. Two parameters are typically used to evaluate system performance: detection efficiency (DE) and location accuracy (LA). For the Brazilian National Integrated Lightning Detection Network, DE is typically evaluated by models, and LA is analyzed through confidence ellipses. This paper presents climatological analysis of lightning activity, including the most recent relative DE model developed in Brazil, as well as an adapted kernel smoothing method based on confidence ellipses [called Gaussian kernel based on confidence ellipses (GKBCE)] as approaches to minimize and/or include the spatial variation of the system's performance in the analysis. The maps are produced over the central-south portion of Brazil (mainly along ITAIPU power transmission lines), using 11 years of data available from the network (from January 1999 to December 2009). The model increased density by ~20% over the entire region, without making considerable changes to the spatial pattern. The GKBCE seems to work well in smoothing, obtaining better results than the standard cell count (quadrat) method, by working independently of the grid size (allowing the creation of high-resolution maps), and by including location errors in the analysis. The use of these procedures might result in more detailed maps and more suitable results when analyzing lightning data.

Index Terms—Detection efficiency (DE) models, kernel smoothness, lightning density maps, lightning location systems (LLS), location accuracy (LA), network performance.

I. INTRODUCTION

BASED on the advance in lightning location system (LLS) technologies, significant improvements in the analysis of lightning distribution are possible. Climatological (long-term) data over large areas allowed the association of physical parameters to lightning distribution, and provided powerful tools for different sectors of society. However, the development of good climatologies depends on how system performance is addressed in the analysis.

For the Brazilian National Integrated Lightning Detection Network (RINDAT) and other networks that use similar tech-

nology (U.S. NLDN, ALDIS, EUCLID, and so on), system performance is usually evaluated via detection efficiency (DE) and location accuracy (LA) [1]. RINDAT has gone through a series of upgrades since the beginning of its operation in 1989 [2], [3]. All these upgrades, associated with gaps in the detection for different sensors, lead to variations in network performance. These variations are normally evaluated through DE analysis. In Brazil, for example, many attempts were made to minimize such variations and a series of relative detection efficiency models (RDEMs) were developed [4]–[6]. In addition to these models, other approaches based on thunderstorm days/thunderstorm hours are suggested more recently to overcome the spatial and temporal variation in system performance and are shown to be very efficient, especially for long-term analysis [7]. However, with those methods, one is not able to retrieve quantitative lightning density information, which again requires the use of the RDEMs. Thus, the methods described in this paper and described by [7] are complimentary.

The LA for RINDAT, also for other networks that use the same technology, is produced using confidence ellipses. These ellipses have semimajor axes (SMAs) whose size and orientation are related to the network's geometry, which means that they are directly related to the sensors locating the discharge (essentially related to the distance and azimuth of each sensor from the discharge, and to the standard deviations of the measurements of those sensors). The ellipses may be used in several applications. Grant *et al.* [8], for example, suggest some ways to apply these ellipses with forensic purposes, offering, additionally, a suitable description of the theory behind the confidence ellipses. The ellipses are even used by the LLS itself to determine whether a stroke composes a flash¹ or not [1]. However, its application for mapping purposes is not well documented. It seems the first attempts to use the Gaussian ellipses of lightning data are made along transmission lines in Slovenia [9], using the confidence ellipses as a weighting function over a grid to create high-resolution map, and by the authors near São Paulo-Brazil [10], where lightning density maps are created using probability rings based on the ellipses.

Spatial point pattern analysis (SPPA) tools and methods are available to analyze stochastic events such as lightning. Analysis through density (or intensity) maps usually refer to a first-order analysis [11], i.e., how the information varies in

¹The LLS recognizes each return stroke from a lightning discharge as a stroke and the whole discharge (with every single stroke) is called a flash.

Manuscript received June 20, 2012; revised January 3, 2013; accepted February 19, 2013.

The authors are with the Atmospheric Electricity Group/Earth System Science Center, Brazilian National Institute for Space Research, São José dos Campos 12227010, Brazil (e-mail: vandoir@gmail.com; osmar@dge.inpe.br; kleber.naccarato@inpe.br).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2013.2253109

the space. Some of the most common methods associated with this analysis are the cell count (also referred to as quadrat) and the (adaptive) kernel density estimation. Diendorfer [12] discussed a way to address uncertainty when working with cell count analysis, using a formulation that indicates the amount of lightning events inside each grid cell necessary to obtain achievable results. Using another approach, now for a univariate case, Wilkinson [13] describes the results presented by Freedman and Diaconis [14] and more recently by Scott [15], pointing to the mean integrated square error for histograms in the order of $n^{-2/3}$, where n means the number of events inside each bin, or grid cell. This value changes to $n^{-4/5}$ when considering a univariate kernel, suggesting that errors associated with density estimation are reduced when using kernel analysis. More sophisticated mathematically, the analysis using kernels also show some advantages due its ability to be associated with probability density functions (PDFs). Kernel techniques are applied to lightning density calculation [16], [17], but without the use of the confidence ellipses and their associated (elliptical) Gaussian PDF.

In this paper, we intend to use the latest RDEM from the RINDAT, and the confidence ellipses as a smoothing kernel-like method, to improve lightning density estimation. The techniques are described in detail in Sections II-A and II-B. The density maps are created mainly along the ITAIPU power transmission lines, showing the effects of the system's performance on the large-scale distribution of lightning. A complementary analysis is carried out close to the urban areas of São Paulo and Belo Horizonte using high-resolution maps. Through using the approaches suggested, which includes the spatial variation of the network's performance in terms of both DE and LA, more reliable lightning distribution might be obtained, offering a better tool to evaluate lightning distribution and its social and economic impact.

II. METHODOLOGY

Eleven years (January 1999 to December 2009) of lightning data from the RINDAT are available for this paper. The data set is collected mostly over the south-central portion of Brazil, as shown in Fig. 1. This region is known for its high-network DE, but also for its importance to the power sector, including the ITAIPU² power transmission lines, which extend from the west of Paraná toward the region of São Paulo. The network's sensors are also shown (squares and stars in Fig. 1). As the sensors in the north portion of the network (in gray) are available for only a short portion of the analysis period, only the sensors in black in Fig. 1 are used in this paper. Two other regions are used in this paper to validate the data and/or to create high-resolution maps, and are highlighted by the dashed rectangles in Fig. 1. The lightning density maps are created using two special tools regarding DE and LA, as mentioned earlier. The tools are described in the following sections.

²ITAIPU is considered today the hydroelectric dam with the largest energy production in the world.

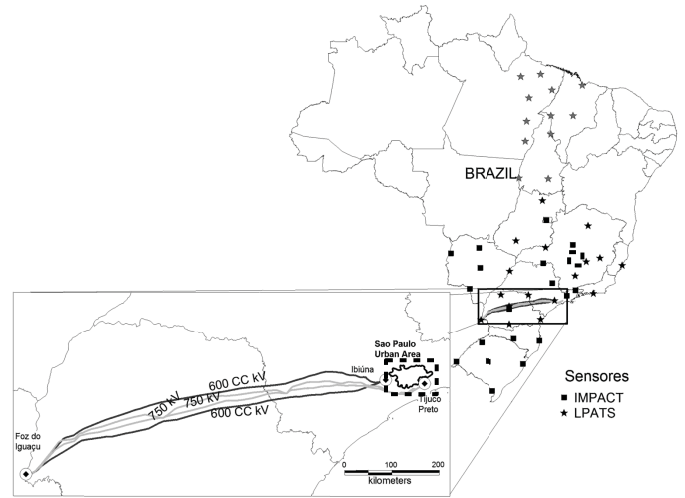


Fig. 1. Study region including the ITAIPU power transmission lines. The black polygon identifies the urban area of São Paulo. Dashed rectangles identify the regions used in the high-resolution analysis: the region of São Paulo (in the zoomed area) and the region of Belo Horizonte (in the large area). Squares and stars identify network sensors. The sensors in gray were not considered in this paper.

A. DE: The Latest DE Model—RDEM4

The latest version of the RDEM aggregates several of the features of the previous versions, described in [4]–[6]. The most important features are summarized as follows.

- 1) Hybrid network based: the calculation of the efficiency (probabilities) considers two angle and four time solutions (IMPACT and LPATS technologies).
- 2) Boundary effects are considered using the idea of essential sensors:³ toward the network boundaries, most of the sensors become crucial for providing a CG lightning solution. Applying this filtering reduces the artificial increase of the relative detection efficiency (RDE) over these outermost areas.
- 3) Angular efficiency: the azimuthal (directional) variations of the sensor relative detection efficiency (SRDE) are considered in this new version.
- 4) Detailed status analysis: the users have the choice of using daily or even higher temporal resolution when evaluating the network's status. The previous version of the model allowed only monthly status analysis.

The major improvement in the current version is the use of directional variations of the DE on the SRDE evaluation and the ability to consider a detailed analysis of the status of each sensor.

To calculate the efficiency, yearly SRDE curves are obtained as a function of distance, angle, and peak current, including the essential sensor and the sensor status⁴ filters. Discharges flagged as intracloud by the LLS are not used in the analysis of the SRDE. Only two peak current intervals (up to 20 kA and

³A sensor is considered essential for a CG lightning solution if this same solution is computed with a minimum number of reporting sensors. Thus, without that sensor, this solution would not exist.

⁴Using a daily uptime analysis. If the sensor works for less than 50% of the day, the information from this day is not used to calculate the SRDE. This percentage includes a weighting function for the afternoon period, when most of the lightning activity takes place. More information can be found in [6].

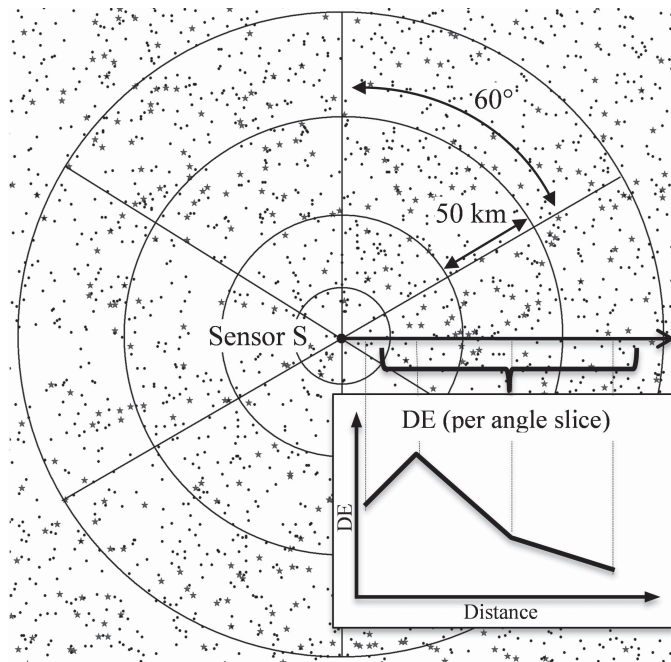


Fig. 2. Illustration of how the relative efficiency of each sensor is obtained as a function of distance (50 km) and direction (60° slices), using data from a specific peak current interval. The efficiency is given by the ratio of the discharges detected by the sensor S (represented by black stars) and the discharges detected by the entire network (black dots) inside each slice and distance step (as shown in the inner graph).

over 20 kA) are used to simplify the calculation, considering that DE changes considerably between these two ranges. The distance and angle bins used in this paper are, respectively, 50 km and 60° (refer to Fig. 2 for more details). Thus, for each angle bin (slice), distance step, and peak current interval, a value of relative efficiency is obtained. These intervals (angle, distance, and peak current) are defined to obtain sufficient data for the relative efficiency (SRDE) curve calculation. Similarly, a minimum of 30 events is required to calculate the RDE for each interval. The curves are then smoothed through a spline function to remove statistical fluctuations, giving more stable results in the final maps. These maps are obtained by integrating the RDE for all the sensors (SRDE) using probability functions, as described by Naccarato *et al.* [5]. The daily status of each sensor, used to calculate the SRDE, is also considered in the calculation of the RDE maps for the entire network. The efficiencies—calculated for each year and for each peak current interval—are then used to correct the density using a proportional weighting method described in Section II-C.

B. LA: A Gaussian Kernel-Like Smoothing Method

As the ellipses are based on a bivariate normal distribution [1], [18], it is possible to compute the discharge location as

$$P(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \times e^{[a(x-x_0)^2 + 2b(x-x_0)(y-y_0) + c(y-y_0)^2]} \quad (1)$$

where

$$a = \frac{\cos^2\theta}{2\sigma_x^2} + \frac{\sin^2\theta}{2\sigma_y^2}$$

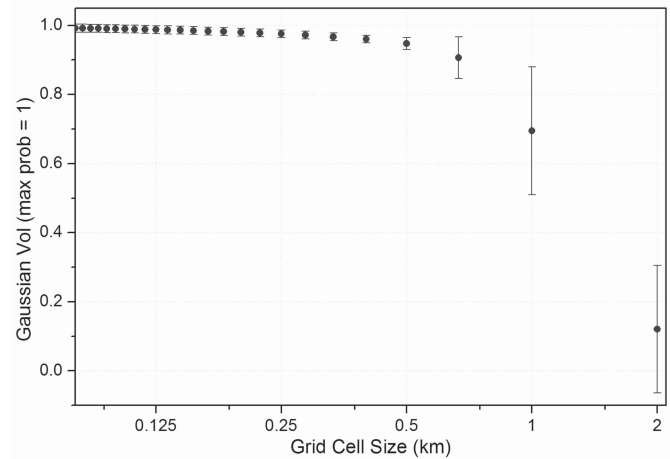


Fig. 3. Monte Carlo-like test showing how resolution affects the precision of the probability, making the volume approximate to one due to the better numerical integration.

$$b = \frac{\sin 2\theta}{4\sigma_x^2} - \frac{\sin 2\theta}{4\sigma_y^2}$$

$$c = \frac{\sin^2\theta}{2\sigma_x^2} + \frac{\cos^2\theta}{2\sigma_y^2}$$

All the unknowns of (1) (σ_x , σ_y , θ , x_0 , y_0) to calculate $P(x, y)$ are given by the LLS. As the default semiaxes length represent sigma (σ_x , σ_y) for a 50% ellipse (i.e., 0.5 of probability that the event is inside the ellipse), this value must be scaled down by 1.177 to represent the one sigma value [19]. Based on (1), a method similar to a simple quadrature numerical integration [20] is used to create the lightning density maps. This method will be called Gaussian kernel based on confidence ellipses (GKBCE). It is similar to the kernel smoothing using the (correlated or uncorrelated) normal distribution [21], except that it is based on the network performance. The main difference lay in the angle and semiaxes, which are different for each discharge. Thus, the density maps resulting from the application of this approach are expected to show a smoothed surface, including the location errors estimated by the LLS.

Each discharge is now spread along the ellipse axes, and the estimation is believed to be less dependent on the grid cell size. If the grid cell size is much smaller than the ellipse (axes) size, the precision of the probability is increased (Gaussian volume approximates to the expected value of one due to the better numerical integration). This is shown in Fig. 3, using a Monte Carlo-like test applied to different grid cell sizes using some solutions from the RINDAT. On the other hand, if the grid cell size is larger than the ellipse size, the chance of having a discharge inside a cell also increases, which reduces the errors associated with the rare event law, as described by Diendorfer [12]. It is important to point out, however, that to obtain consistent results it is always important to have a reasonable number of events over a certain region (e.g., 80 events per grid cell indicated by [12]). As an attempt to identify the effects of both quadrat technique (grid cell count) and the proposed GKBCE on final density maps, a comparison

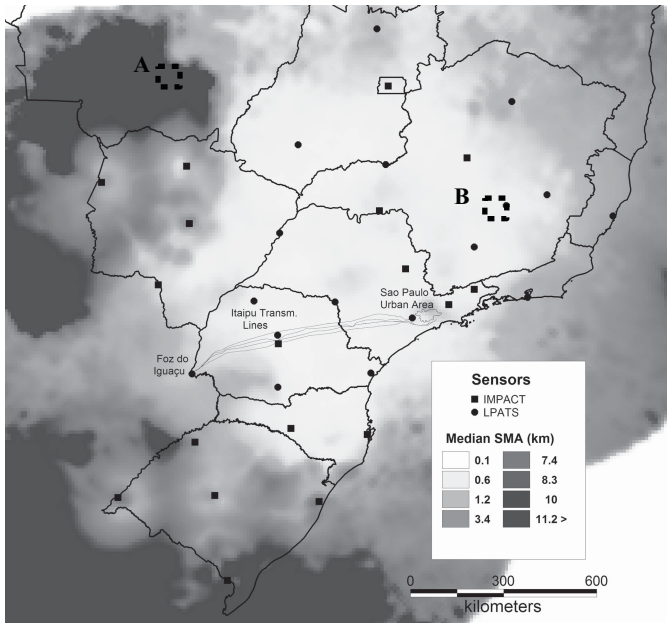


Fig. 4. Map showing the median semi-major axis (SMA) size of the confidence ellipses for the RINDAT based on 29 months of data. Filled black dots and rectangles show the location of the sensors of the network (LPATS and IMPACT technology, respectively). Dashed rectangles indicate the regions used to evaluate the impact of the ellipse size variation on the density maps using the GKBCE: (A) Cuiaba and (B) Belo Horizonte.

is carried out using three different grid cell resolutions: 0.1, 0.5, and 2 km (see discussion of Fig. 9).

Fig. 4 shows the mean size of the SMA of the confidence ellipses given by the LLS. The analysis is carried out using a short period of data (29 months, from January 2006 to May 2008), and indicates how the ellipses behave related to the network geometry: the SMA size increases as one moves toward the boundary of the network. Based on this observation, two regions (dashed rectangles) are selected to evaluate the impact of the ellipse size variation on the density maps using the GKBCE: the region of Cuiaba (A), on the network's edge, where the median SMA size is quite large; and the region of Belo Horizonte (B), where many LPATS sensors are located and the SMA size is reduced.

Over the regions used in this paper, the median SMA size varies by up to 1–2 km. Thus, resolutions from 100 to 500 m are used, giving an accuracy up to 0.99 for the GKBCE (see Fig. 3). To produce the maps, a multiplicative smoothing factor (or bandwidth scaling) can be applied, following the ideas described by Wand and Jones [22]. This factor is constant and only impacts the smoothness of the map (the Gaussian PDF is stretched, reducing the maximum value; each discharge will still have its own ellipse size and orientation). It is an interesting way to analyze coarser grids or, more generally, cases where the size of the grid cell is larger than the size of the SMA of the ellipse. For those cases, the use of the default axes (smoothing factor of 1.0) would reduce the precision of the probability (as shown on Fig. 3) or even produce counts just like other methods, not spreading the Gaussian PDF over multiple grid cells. As a reasonable number of ellipses have an SMA of under 500 m (~15% of the data for the entire country), the maps for this paper are developed using

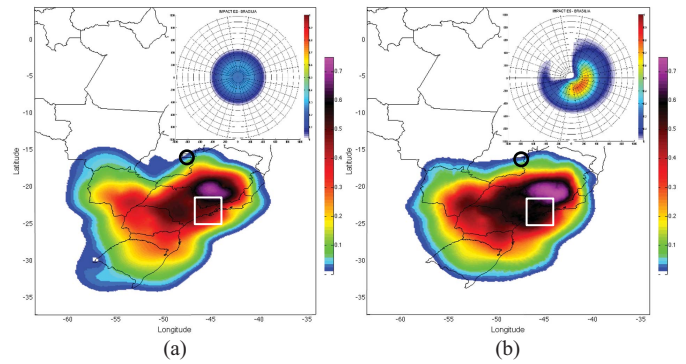


Fig. 5. Average detection efficiency (using the whole period of data) for small peak currents considering (a) only distance and (b) distance and direction (azimuthal). Black circle indicates the location of the highlighted sensor (inner polar graph). Note the increase in the efficiency in the region around São Paulo (white rectangle).

a smoothing factor of two, i.e., the semi-axes twice the initial size.

C. Maps Using Combined Methods

The sum of all the discharges on the grid gives us the final lightning count. To correct the flash count by the DE, the proportion of discharges within each peak current interval for each year is used

$$DE_i = \frac{(\text{Perc}_{i(<20)} \times \text{Ef}_{i(<20)} + \text{Perc}_{i(>20)} \times \text{Ef}_{i(>20)})}{(\text{Perc}_{i(<20)} + \text{Perc}_{i(>20)})} \quad (2)$$

$$FC_{\text{final}} = \sum_{i=1999}^{2009} D_i / DE_i \quad (3)$$

where D_i is the uncorrected flash count, DE_i is the average DE for each year, and FC_{final} is the corrected flash count. Ef and Perc are, respectively, the calculated efficiency and the percentage of discharge inside each peak current interval (< 20 and > 20 kA). Using this technique, regardless of the simple mean of the two peak current intervals, may avoid the overestimation of the efficiency (mainly due the higher efficiency obtained for large peak current discharges). This approach is similar to the procedure described by Diendorfer *et al.* [23], which determined the DE by weighting the probability of occurrence based on the peak current distribution (i.e., the percentage of discharges in each peak current interval). The density in flash · km⁻² · yr⁻¹ is obtained by dividing the total flash count by the area of the grid cell (0.01, 0.25, and 4 km², accordingly to the analysis) and by the number of years (11). Equation (3) is also used to obtain the average DE over the whole period and area.

III. RESULTS AND DISCUSSION

A. Improvements on the RDE Model

Fig. 5 shows the average RDE calculated for events with peak current below 20 kA (using data from the whole period), considering only distance [Fig. 5(a)], or distance and direction [Fig. 5(b)]. The results indicate considerable differences between the two conditions and a significant improvement when the angular variation of the DE is considered.

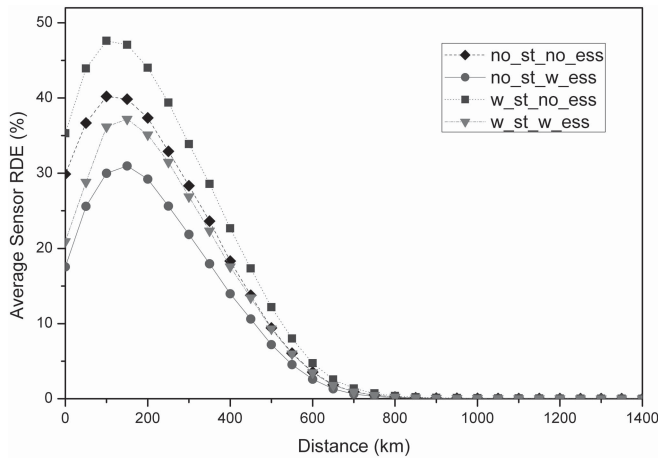


Fig. 6. Variation of the averaged SRDE curves considering (w) and not considering (no) the application of the essential sensor (ess) and status (st) filters. Only data with peak current up to 30 kA and in the period between Nov. 2009 and April 2010 were used.

The efficiency in the network's edge is reduced, especially in the southern and central eastern regions. In addition, it is possible to see that the RDE over the region of São Paulo and its neighborhoods (defined by the white rectangle in Fig. 5) is increased by $\sim 20\%$, suggesting a more realistic condition of the new version of the model when analyzed against the network geometry in this region. The maximum efficiency calculated for the specified peak current interval (below 20 kA) is ~ 0.75 . For larger peak current events, better performance of the network (with higher efficiency) is expected.

DE model improvement relates to the use (or not) of both the status and essential sensor filters, as shown in Fig. 6. This analysis is based on data with peak current up to 30 kA from a single summer (November 2009 to April 2010) with the curves averaged for all the sensors. The results for this case show that the (false) efficiency is reduced when the essential sensor filter is considered, as expected. The status filter, on the other hand, increases the efficiency by considering only days when the sensor is really working to calculate the efficiency values. The variations are $\sim 8\%$ when the filters are considered. In addition, these values may change significantly for each sensor. Thus, the inclusion of these criteria is certainly important in obtaining a more reliable efficiency estimation.

Fig. 7 shows the average RDE calculated for the entire network considering the angle and the filters mentioned before. The average efficiency is calculated using (2) for the two peak current intervals described earlier. The maximum estimated efficiency is at $\sim 85\%$, and the higher efficiency values are concentrated over São Paulo, Minas Gerais, Rio de Janeiro, and Paraná, as expected.

The density maps without correction and corrected by the RDE (using a spatial resolution of 10 km) are shown in Fig. 8(a) and (b), respectively. The variations in the estimated density are at $\sim 30\%$ (DE of 70%) over the main regions used in this paper. The regions with the most variations are in the west portion of the country. Despite the big changes in the density, the main features of the spatial distribution are quite clear and cannot be associated with the system DE variations. The maximum density for this case is ~ 12 flashes \cdot km $^{-2}$ \cdot yr $^{-1}$.

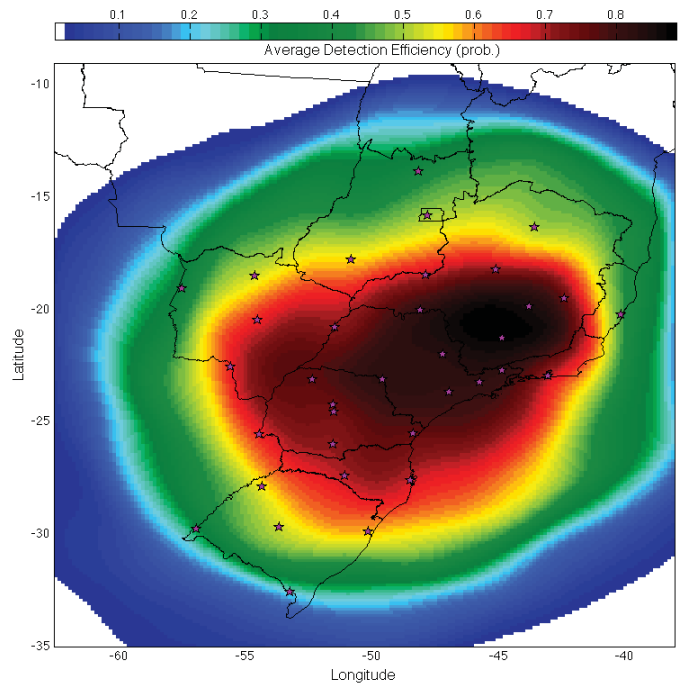


Fig. 7. Average detection efficiency estimated for the whole period (99/2009) using (2).

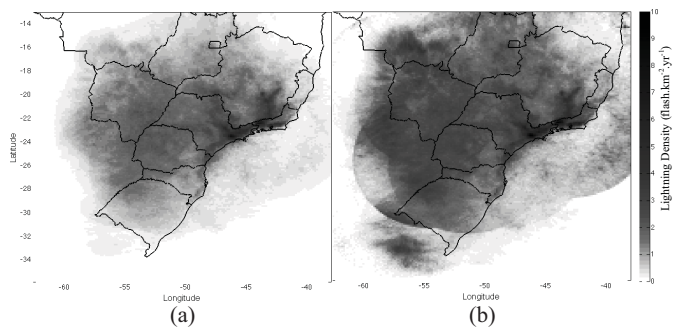


Fig. 8. Density maps without correction (a) and corrected (b) by the DE for the entire network.

It is important to mention that the corrections obtained by most RDE models are not able to compensate for variations of the network's performance over time,⁵ for which the approaches described by Bourscheidt *et al.* [7] could be applied. Thus, this implies that the methods used here and in [7] are connected and may be applied together for a more complete understanding of the system's performance effects on the spatial and temporal analysis of lightning activity.

B. Comparison of Mapping Methods: Cell Count and GKBCE

As described earlier, different methods are available for SPPA and it is necessary to choose the method in accordance with the objectives of this paper. This paper attempts to integrate the detection systems LA for lightning density maps using the confidence ellipses as a kernel with a smoothing purpose. To evaluate the performance of the proposed method (GKBCE), lightning distribution maps are created in two ways:

⁵They would be able to correct for such effects if one has DE estimates for every time the network status changes, but this is not really practical.

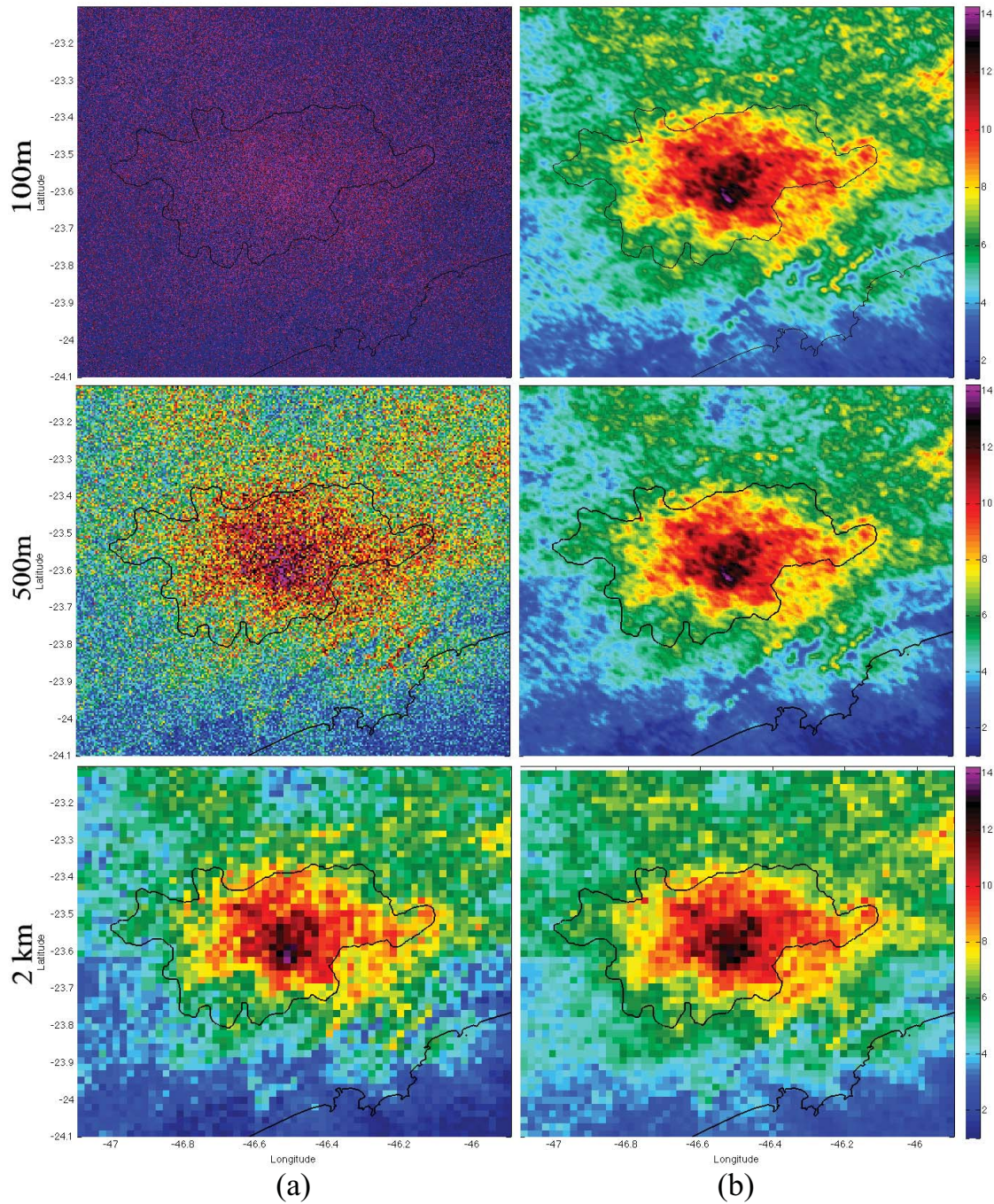


Fig. 9. (a) Lightning density maps using the cell-count method and (b) GKBCCE method for different spatial resolutions (rows, with 0.1, 0.5, and 2 km). The cell-count method shows a strong dependency on the grid resolution, giving unrealistic values at high resolutions (100 m, upper left). GKBCCE seems to be more suitable and independent of the resolution.

using the simple cell count (or quadrat) and using the GKBCCE for three different spatial resolutions (0.1, 0.5, and 2 km), as shown in Fig. 9 (sequentially, from the top to the bottom, are the increasing resolutions; the cell-count method is on the left and the GKBCCE on the right). Values are corrected for DE. The results shown in the figure indicate that the differences increase as the spatial resolution increases (bottom to top), as a result of the number of events inside each grid cell (law of rare events), in agreement with [12]. When high resolution is used

and small counts occur, the noisy aspect is more likely to occur in the cell count method. The same is valid for the density estimation ($\text{flash} \cdot \text{km}^{-2} \cdot \text{yr}^{-1}$): as it depends on the size of the grid cell, the variations will increase accordingly. The density values using the simple count method (left column on Fig. 9), for example, vary from 0 to $\sim 80 \text{ flashes} \cdot \text{km}^{-2} \cdot \text{yr}^{-1}$ when using a 100-m resolution and change to the range within 0 and $14 \text{ flashes} \cdot \text{km}^{-2} \cdot \text{yr}^{-1}$ when the resolution of 2 km is used. For the GKBCCE (right column on Fig. 9), on the other hand, the

TABLE I

DENSITY ESTIMATE VARIATION ACCORDING TO THE METHOD AND THE SPATIAL RESOLUTION USED IN THE ANALYSIS (IN FLASHES $\cdot \text{km}^{-2} \cdot \text{yr}^{-1}$)

Resolution (Km)	Cell Count				GKBCE			
	Min	Max	Median	Std	Min	Max	Median	Std
0.1	0	88.99	5.79	8.85	1.38	14.32	5.97	2.0
0.5	0	21.25	5.49	2.68	1.39	14.28	5.61	2.0
2	1.35	13.97	5.63	2.03	1.38	13.16	5.53	1.97

estimated values stay at the same range for all the resolutions (~ 14 flashes $\cdot \text{km}^{-2} \cdot \text{yr}^{-1}$). This analysis is also shown in Table I.

As the GKBCE spreads the discharges over a probability surface, it becomes independent of the resolution. This allows the creation of high-resolution maps that consider the system performance (LA). It is important to mention that the GKBCE method is similar to the normal kernel method described, for example, by Wand and Jones [21]. The advantage of the GKBCE is that it includes the location errors given by the LLS.

The presented analysis concentrates on high-resolution maps. When low-resolution maps are considered, the amount of data could be large enough so that the noise aspect does not occur. In addition, when the grid cell becomes larger than the ellipse size (as might be the case for the bottom row in Fig. 9), the maps generated by the different methods become similar and any method could be used to produce the density maps.

The GKBCE provide suitable results when considering the random errors (random ellipse size and direction) associated with the location system. However, systematic errors may affect the results, especially when the confidence ellipses have a preferred orientation and large SMA, which could imply significant biases and distortion on the map. These effects are expected to occur near to the edge and/or outside the network. To evaluate the described effects on the final maps, two regions are analyzed (as described in the methodology and highlighted with the black dashed rectangle in Fig. 4): Cuiaba, at the edge of the network, where large ellipses are expected (region A); and Belo Horizonte (region B), in the middle of the network (which usually leads to smaller SMAs). The analysis of these regions is shown in Figs. 10 and 11, which also show the annual variation in median SMA size. This median ellipse size is based on all the discharges occurring inside the analyzed region, and this same procedure is used in the next figures. For the region of Cuiaba, a stronger spatial trend is observed (Fig. 10), with the ellipses having a preferred orientation toward the southeast direction. The median SMA value for this region is 5.7 km, with the smaller SMAs occurring in 2004. This region is in the network's edge and the DE is also slightly reduced.

Fig. 11 shows lightning density and the ellipses' SMA variation over the years for the region of Belo Horizonte (region B in Fig. 4). This region is expected to have the highest DE and LA as it is located in the center of the network (refer to Figs. 4 and 7). The median size of the SMA is of ~ 0.6 km, and this value shows a slight increase in recent years,

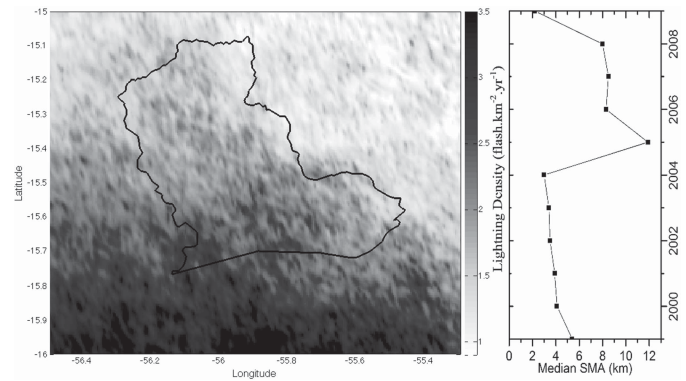


Fig. 10. Map over the region of Cuiaba using the GKBCE. Right panel shows the mean SMA variation over the years.

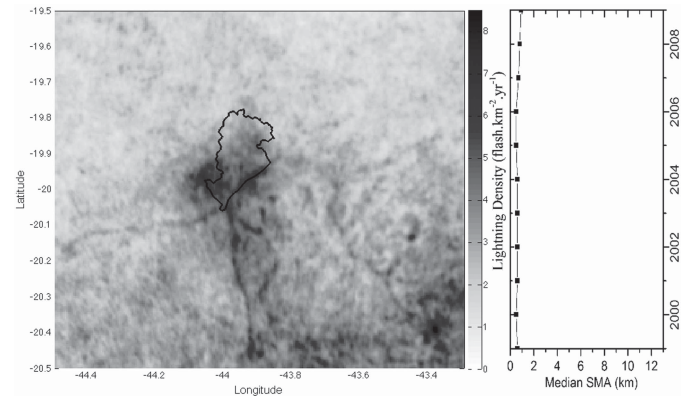


Fig. 11. Map over the region of Belo Horizonte using the GKBCE. Right panel shows the mean SMA variation over the years.

possibly because of the upgrade from IMPACT ESP to more sensitive LS7000 sensors in 2007–2008, thereby increasing the likelihood of reporting low-current discharges that can have larger location errors (to the south, outside of the map limits in Fig. 11). The maps produced over this region did not show any perceptible spatial bias.

The observed results suggest that when most of the ellipses have an SMA under 2 km, almost no spatial trend will be observed in the maps. On the other hand, regions where the SMAs are > 10 km would probably show systematic errors that may restrict the application of the GKBCE method.

Another important aspect that must be considered is the bandwidth scaling adopted: to obtain a suitable smoothness in the final maps, all the analyses performed in this paper considered a bandwidth scaling of two. This value seems reasonable, as the maximum estimated densities for the coarser resolution in Table I (with large amount of data and, thus, more stable and reliable values) are in the same order as the other resolutions for the GKBCE. Nevertheless, variants of this factor may affect the map smoothness and visualization. Hence, each user needs to find the most appropriate values according to the objectives of this paper.

To summarize, the proposed GKBCE method gives satisfactory results and its application may be very useful in analysis using high-resolution maps. However, regions where systematic errors (associated with the detection system) predominate must be analyzed carefully.

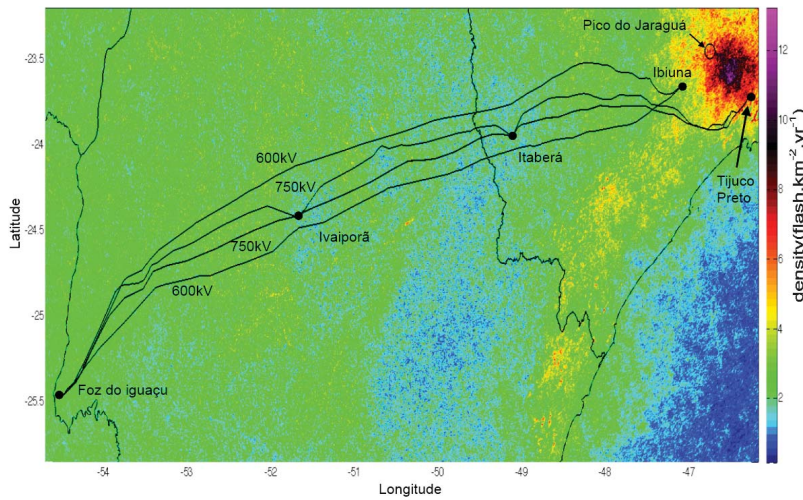


Fig. 12. Annual CG flash density ($\text{flash}\cdot\text{km}^{-2}\cdot\text{year}^{-1}$) along the ITAIPU transmission lines corrected by the detection efficiency. Black circle indicates the region used for the high-resolution analysis.

C. Lightning Density Along the ITAIPU Power Transmission Lines

The main objective of the described methods is their application to create lightning density maps or climatologies. The following analysis is developed considering applications in the electrical sector. Fig. 12 shows the final density obtained in the surroundings of the main transmission line in the country (750 and 600-kV lines) connecting the ITAIPU hydroelectric power plant to the São Paulo metropolitan region (SPMR).

The results are in overall agreement with previous studies for the SPMR (e.g., [24]). Another interesting observation is the relative higher lightning activity along the coast (east side), probably associated with the mountain range in this region. The northeastern limit of the transmission lines coincides with the SPMR and show that the most critical substations—more susceptible to lightning strikes—are in this region. The black circle near to the urban area of São Paulo (Pico do Jaraguá) is explored further using high-resolution analysis, discussed in Section III-D.

The GKBCE method proved suitable for this analysis, exhibiting small effects of systematic errors, even for regions with relatively low DE, which is consistent with the results presented in the previous section. Regarding the performance of the new RDEM, the variations in the density are $\sim 20\%$, with the maximum shifting from ~ 11 to 14 flashes $\cdot\text{km}^{-2}\cdot\text{yr}^{-1}$. This value may change depending on the peak current intervals used in the analysis, as the DE weighting factor described in Section II-C. depends on this information. Fig. 13 shows a profile (cross-section) along one of the transmission lines (upper 750-kV line in Fig. 1). The profile indicates that the intermediary substations (Ivaiporã and Itaberá) are in regions with the lowest flash density. It also shows the higher densities near to São Paulo.

D. High-Resolution Analysis of Lightning Density

Fig. 14 shows lightning density using a high-resolution map (with a spatial resolution of 100 m) overlaid on the elevation for a specific place nearby the SPMR (Pico do Jaraguá, in the

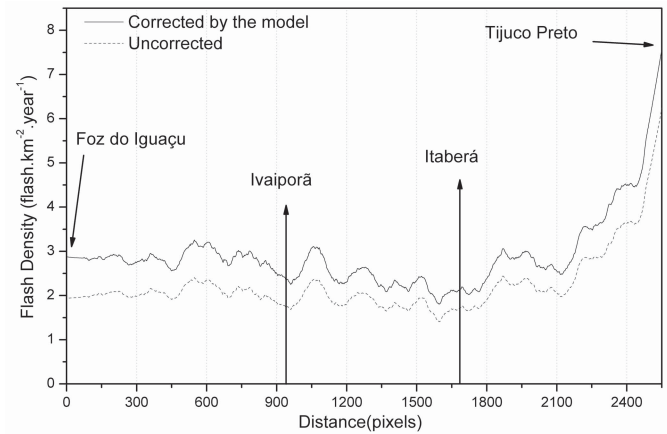


Fig. 13. Cross-section of the upper 750 kV transmission line over the density map.

northwest of São Paulo city, as indicated by the black circle in Fig. 12). The analysis for this case study is carried out using the 3-D view option of Google Earth (GE). The analysis of this region shows a concentration of discharges exactly over the mountain peak. This peak, at 1,135 m above the sea level, is isolated from other mountains and several antennas on top of it, which may facilitate the lightning attachment process. Recent investigations show upward discharges starting at this place [25], which could also indicate that upward leaders (ground-to-cloud) followed by subsequent discharges (α and β pulses) similar to the cloud-to-ground discharges could be being detected by the location systems [23], [26]. This agreement of the discharges with the mountains is also found on other nearby mountain peaks (not shown) and suggests that the LA is fairly high for this region, likely better than 1 km.

A second and complementary analysis using high resolution (with GE) refers to the region near Belo Horizonte (Figs. 15 and 16), where the highest LA of the whole network is expected (refer to Fig. 11, now superimposed for the elevation). The discharges are concentrated mainly along the mountain chain in the north-south direction, as indicated by the dashed lines and arrow in Fig. 15. A concentration of

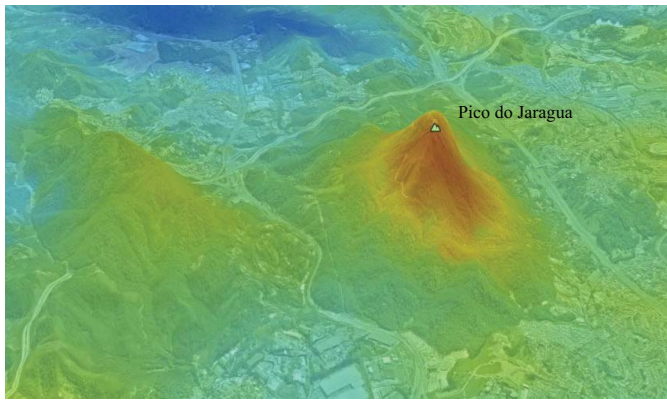


Fig. 14. Detailed view of the Pico do Jaraguá, SP, using a high-resolution map [100(m)].

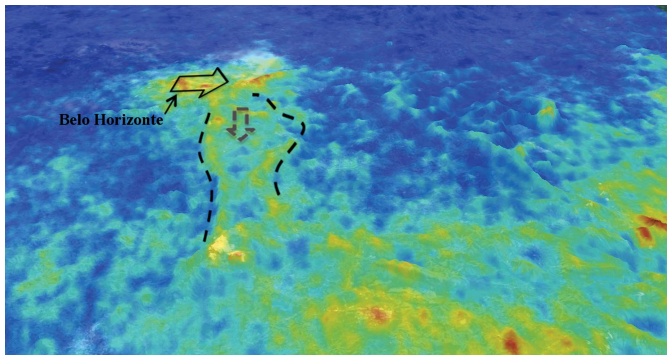


Fig. 15. Flash density map overlaid on the elevation for the region around Belo Horizonte. Dashed lines indicate north-south plateau and black arrow point to mountain chain near urban area (analyzed in detail).

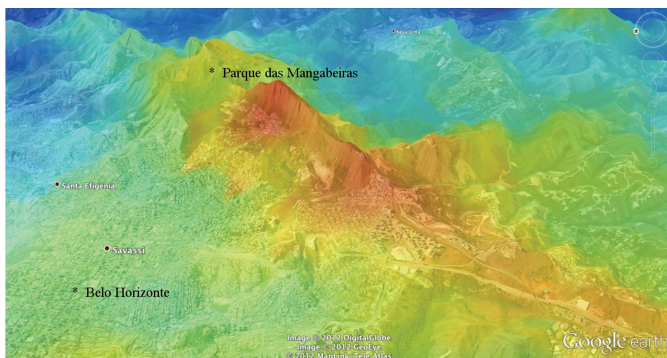


Fig. 16. Detailed analysis of the lightning density in the southern limit of Belo Horizonte's urban area (as indicated by the arrow on Fig. 15).

discharges over the metropolitan region (the area under the black continuous arrow) is also observed, possibly associated with heat island effects [27]. The arrow on Fig. 15 also indicates a peak in lightning activity right over the mountains in the southeast of the urban area (shown in detail on Fig. 16) and illustrates the possible actuation of the terrain on lightning activity, caused either by orographic forcing or by the facilitated discharge attachment process (and the occurrence of upward discharges as well).

IV. CONCLUSION

This paper summarized some new approaches applied to RINDAT data to correct and improve lightning density

estimates. The improvements to the most recent version of the RDEM, as well as a new kernel-like smoothing method, were presented.

The fourth generation RDEM showed substantial improvements in network efficiency estimation by adding angular variations and more efficient sensor status filters. The estimated values were increased in the inner network and reduced outside, which was more consistent. The improvements to the filters minimized the (artificial) efficiency enhancement at the network edges, and also prevented sensors that were down for long periods from being considered in the efficiency curve survey. The variations in lightning density when the DE model was considered were at $\sim 30\%$ (which means 70% efficiency, on average), with the major effect in the west portion of the RINDAT. Even though not the case in this paper, the RDEM was still not able to correct for temporal variations in network performance, because of computational complexity. In this case, other techniques (refer to [7]) could be applied.

Regarding LA, the smoothing method developed by using a GKBCE showed to be of interest and produced new results, allowing the development of high-resolution maps without impacts to density information. The main difference from existent kernel techniques referred to the bivariate axes, which include the LA of each discharge. The method seemed to be very efficient for random errors, but might provide biased density estimates when systematic errors were predominant (usually in regions in the edge or outside the network). The sizes of the SMAs of the ellipses over this paper's regions were < 1 km, in general. This size varied significantly over time, which was possibly related to changes in detection technology over the years.

It was also shown that the final estimated density agreed with previous papers, with the highest values over the urban area of São Paulo, the most critical area along the transmission line. On the other hand, the core substations seemed to be located in regions with lower lightning activity. A band with higher density existed along the east coast and might be related to elevation effects.

Focusing on high-resolution lightning density mapping, a close analysis over São Paulo city showed a clear higher CG lightning density over the Pico do Jaraguá (the highest peak in the city, at 1,135 m). Similarly, the region of Belo Horizonte, expected to have the highest LA, showed a close relation between lightning activity and terrain features. These results suggested that: 1) the LA of the system was pretty high for these regions; 2) the GKBCE method worked well for detailed analysis like this; and 3) mountain ranges and isolated peaks were more likely to have lightning strikes.

Finally, the application of the described methods might be a good way to improve lightning density maps, and also to reduce the effects of the system's performance on lightning spatial distribution analysis.

ACKNOWLEDGMENTS

The authors would like to thank ELETROBRAS FURNAS for the transmission line georeferenced data and for logistical support regarding the DE model. Additionally, Ken Cummins

and Wolfgang Schulz for their helpful comments on several topics discussed in this paper.

REFERENCES

- [1] K. L. Cummins, M. J. Murphy, E. A. Bardo, W. L. Hiscox, R. B. Pyle, and A. E. Pifer, "A combined TOA/MDF technology upgrade of the U.S. National Lightning Detection Network," *J. Geophys. Res. Atmospher.*, vol. 103, no. D8, pp. 9035–9044, Apr. 1998.
- [2] K. P. Naccarato, O. Pinto, Jr., and M. Murphy, "Performance analysis of the BRASILDAT network," in *Proc. 3rd Int. Conf. Lightning Phys. Effects, Int. Conf. Grounding Earthing*, 2008, pp. 329–338.
- [3] O. Pinto, *Lightning in the Tropics: From a Source of Fire to a Monitoring System of Climatic Changes*. Commack, NY, USA: Nova Science Publishers, 2010.
- [4] K. P. Naccarato, O. Pinto, Jr., and I. R. C. A. Pinto, "A detection efficiency model for the Brazilian Lightning Detection Network (RINDAT)," in *Proc. Int. Lightning Detection Conf.*, Tucson, AZ, 2006, pp. 1–6.
- [5] K. P. Naccarato and O. Pinto, "Improvements in the detection efficiency model for the Brazilian Lightning Detection Network (BrasilDAT)," *Atmospher. Res.*, vol. 91, pp. 546–563, Feb. 2009.
- [6] K. P. Naccarato, V. Bourscheidt, and O. Pinto, "The fourth generation of the Brazilian detection efficiency model for BrasilDAT network (RDEM4)," in *Proc. 14th Int. Conf. Atmospher. Electr.*, 2011, pp. 1–3.
- [7] V. Bourscheidt, K. L. Cummins, O. Pinto, and K. P. Naccarato, "Methods to overcome lightning location system performance limitations on spatial and temporal analysis: Brazilian case," *J. Atmospher. Ocean. Technol.*, vol. 29, no. 9, pp. 1304–1311, Sep. 2012.
- [8] M. D. Grant, I. S. McKechnie, I. R. Jandrell, and K. J. Nixon, "Probabilistic interpretation of LDN confidence ellipses with reference to forensic applications," in *Proc. 7th Asia-Pacific Int. Conf. Lightning*, Jan. 2011, pp. 750–754.
- [9] J. Kosmac, G. Lakota Jericek, V. Djurica, and T. Toros, "Evaluation of lightning threat in distribution networks," in *Proc. 20th Int. Conf. Electr. Distrib.*, Jun. 2009, pp. 1–4.
- [10] V. Bourscheidt, O. Pinto, Jr., and K. P. Naccarato, "Using the Gaussian-based confidence ellipse provided by lightning location systems to create more reliable lightning density maps," in *Proc. Int. Conf. Lightning Detection/Int. Conf. Lightning Meteorol.*, 2010, pp. 1–4.
- [11] A. C. Gatrell, T. C. Bailey, P. J. Diggle, and B. S. Rowlingson, "Spatial point pattern analysis and its application in geographical epidemiology," *Trans. Inst. Brit. Geograph.*, vol. 21, no. 1, pp. 256–274, Mar. 1996.
- [12] G. Diendorfer, "Some comments on the achievable accuracy of local ground flash density values," in *Proc. 29th Int. Conf. Lightning Protection*, Jun. 2008, pp. 281–286.
- [13] L. Wilkinson, "Dot plots," *Amer. Stat.*, vol. 53, no. 3, pp. 276–281, Aug. 1999.
- [14] D. Freedman and P. Diaconis, "On the histogram as a density estimator: L_2 theory," *Zeitschrift Für Wahrscheinlichkeitstheorie Und Verwandte Gebiete*, vol. 57, no. 4, pp. 453–476, 1981.
- [15] D. W. Scott, *Multivariate Density Estimation: Theory Practice and Visualization*. New York, NY, USA: John Wiley Sons, 1992.
- [16] G. Amatulli, F. Pérez-Cabello, and J. De La Riva, "Mapping lightning/human-caused wildfires occurrence under ignition point location uncertainty," *Ecol. Model.*, vol. 200, nos. 3–4, pp. 321–333, Jan. 2007.
- [17] W. Liu, W. Shixin, Z. Yi, W. Litao, and Z. Shujie, "Spatial distribution patterns analysis of historical forest fires in DaXingAn Mountains of China," in *Proc. Int. Conf. Comput. Appl. Syst. Model.*, Oct. 2010, pp. V3-634–V3-638.
- [18] R. G. Stansfield, "Statistical theory of DF fixing," *J. Inst. Electr. Eng. Part IIIA, Radiocommun.*, vol. 94, no. 15, pp. 762–770, Mar.–Apr. 1947.
- [19] *LF Sensor LS7000 User's Guide*, Gulf Publishing Co., Houston, TX, USA, 2006.
- [20] C. W. Ueberhuber, *Numerical Computation 2: Methods, Software, and Analysis*. New York, NY, USA: Springer-Verlag, 1997.
- [21] M. P. Wand and M. C. Jones, *Kernel Smoothing*. London, U.K.: Chapman and Hall, 1995.
- [22] M. P. Wand and M. C. Jones, "Comparision of smoothing parametrizations in bivariate kernel density estimation," *J. Amer. Stat. Assoc.*, vol. 88, no. 422, pp. 520–528, 1993.

- [23] G. Diendorfer, M. Bernardi, K. Cummins, F. D. L. Rosa, B. Hermoso, A. Hussein, T. Kawamura, F. Rachidi, and H. Torres, "Cloud-to-ground lightning parameters derived from lightning detection systems: The effects of system performance," in *CIGRE Task Force C4.404A*, Vienna, Austria, 2007, pp. 1–5.
- [24] K. P. Naccarato, O. Pinto, Jr., and I. R. C. A. Pinto, "Evidence of thermal and aerosol effects on the cloud-to-ground lightning density and polarity over large urban areas of southeastern Brazil," *Geophys. Res. Lett.*, vol. 30, no. 13, pp. 71–74, Jul. 2003.
- [25] M. M. F. Saba and J. Alves, "Upward lightning in Brazil," in *Proc. Int. Lightning Detection Conf.*, Apr. 2012, pp. 1–3.
- [26] V. A. Rakov and M. A. Uman, *Lightning: Physics and Effects*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [27] I. R. C. A. Pinto, O. Pinto, Jr., M. A. S. S. Gomes, and N. J. Ferreira, "Urban effect on the characteristics of cloud-to-ground lightning over Belo Horizonte-Brazil," *Ann. Geophys.*, vol. 22, pp. 697–700, Jan. 2004.



Vandoir Bourscheidt received the B.S. degree in geography from the Federal University of Santa Maria, Rio Grande do Sul, Brazil, in 2006, and the M.S. and Ph.D. degrees in space geophysics (with emphases on atmospheric sciences and atmospheric electricity) from the Brazilian National Institute for Space Research (INPE), São Paulo, Brazil, in 2008 and 2012, respectively.

He is currently a Post-Doctoral Fellow with the Atmospheric Electricity Group, INPE, developing a tracking system based on lightning data. His current research interests include the evaluation of lightning location systems performance, and the development of GIS tools applied to lightning data, as well as other climatic and meteorological studies.



Osmar Pinto, Jr. (M'05) was born in Porto Alegre, Brazil, in 1954. He received the B.S. and M.S. degrees in space geophysics from the Brazilian National Institute for Space Research, São Paulo, Brazil.

He is a Senior Scientist of Brazilian National Institute for Space Research (INPE) and conducts experimental and theoretical research in atmospheric electricity, in particular in lightning physics and lightning detection. He is the Head of the Atmospheric Electricity Group, INPE, the

Coordinator of the Brazilian Lightning Detection Network (BrasilDat), the President of the Brazilian Workshop on Atmospheric Electricity and the President of the International Conference on Lightning Physics and Effects. He has advised more than 20 M.Sc. and Ph.D. students. He has authored four books about lightning, about 100 papers in refereed journals and more than 200 papers in international conferences.

Mr. Pinto is a member of the International Commission on Atmospheric Electricity, the Scientific Committee of the International Symposium on Lightning Protection, and the American Geophysical Union. In 1990, he was indicated for the Macelwane Medal of the American Geophysical Union by Prof. James A. Van Allen.



Kleber P. Naccarato was born in São Paulo, Brazil, in 1974. He received the B.S. degree in electrical engineering from the University of São Paulo (USP), São Carlos, in 1996 and the Ph.D. degree in space geophysics, with an emphasis in atmospheric electricity, from the Brazilian National Institute for Space Research (INPE), São José dos Campos, Brazil, in 2005.

He has been an Atmospheric Electricity Scientist at INPE, since 2005, working in collaboration with many other important institutions in Brazil: State University of São Paulo, Technological Centre of Brazilian Air Force, and University of São Paulo. He has authored more than 10 articles related to lightning detection technologies, urban effects, protection of humans against lightning threats, and thunderstorms nowcasting.