Methods to Overcome Lightning Location System Performance Limitations on Spatial and Temporal Analysis: Brazilian Case

VANDOIR BOURSCHEIDT

Brazilian National Institute for Space Research, São Paulo, Brazil

KENNETH L. CUMMINS

Department of Atmospheric Sciences, The University of Arizona, Tucson, Arizona

OSMAR PINTO JR. AND KLEBER P. NACCARATO

Brazilian National Institute for Space Research, São Paulo, Brazil

ABSTRACT

One of the most interesting attributes of Lightning Location Systems (LLSs) data is that they can be analyzed in several ways according to the objectives of the study. However, the quality of the data is governed by the system performance and has some limitations when analyzed at different temporal/spatial scales, and these limitations will depend on the analysis method. This work focuses on approaches to minimize the variations associated with LLS performance. In this way, specific network configurations for the Brazilian Lightning Detection Network (BLDN) were obtained through the reprocessing of selected sensor data, resulting in three distinct datasets. Each dataset was then evaluated using different procedures: trimmed flash (exclusion of low current discharges), thunderstorm days (TDs), and thunderstorm hours (THs). The comparison between TDs obtained from the LLS and TDs available from surface stations shows consistent results with a good correlation of those datasets. An 11-yr analysis of BLDN data also shows that improvement (over time) of the system sensitivity has led to the detection of an increasing number of low peak current events. By eliminating low peak current discharges (less than 19 kA), the sensitivity variation was significantly reduced, partially "normalizing" long-term performance. TDs and THs were the most effective method to normalize temporal variations of the lightning activity, overcoming most of the network performance variations. From the spatial perspective, TDs and THs also seem to produce the most reliable lightning distribution. These results might guide long-term temporal and spatial analysis of lightning data, providing a more stable approach that is independent of system performance.

1. Introduction

The analysis of lightning data from Lightning Location Systems (LLSs) can be carried out in both spatial and temporal domains. In both cases, the performance of the location system is critical and must be evaluated. The network performance is usually quantified by system location accuracy and detection efficiency (DE), and it depends basically on the number and type of sensors working, as well as on the network geometry at any given time (Cummins and Bardo 2004; Cummins et al. 1998; Naccarato and Pinto 2009). Those characteristics are interdependent and their variations may result in unrealistic spatial and temporal characteristics, leading to results that could be more related to the system performance than to physical effects.

The Brazilian Lightning Detection Network (BLDN) has gone through a series of upgrades since its creation (Naccarato et al. 2008; Pinto et al. 2006) and, at the time of this analysis, the network is composed of commercial sensors (from Vaisala) with different technology and sensitivity (lightning sensor models LPATS, IMPACT, LS7000, and LS8000, as shown in Table 1). All of these upgrades, associated with some gaps in the detection for different sensors, lead to variations in the network

Corresponding author address: Vandoir Bourscheidt, P.O. Box 12227-010, Av. Astronautas, Jd. Granja, Brazilian National Institute for Space Research, São José dos Campos 1758, São Paulo, Brazil. E-mail: vandoir@dge.inpe.br

TABLE 1. BLDN network variations over the years (by amount and type of sensors).

Year	Impact	LPATS	LS7000	LS8000	Total
1999	5	17			22
2003	8	17			25
2005	18	27			45
2007	16	27	2		45
2008	16	27	2	2	47
2009	11	17	2	2	32

performance along the years. Special care must be taken to address these effects.

These variations are usually evaluated through the system detection efficiency (absolute and relative). The absolute detection efficiency of the BLDN was determined by Ballarotti et al. (2006) to be around 85% for flashes at that time. Detection efficiency models (DEMs) have been used to overcome the detection efficiency issues worldwide (Rubinstein 1995; Schulz and Diendorfer 1996; Schulz and Cummins 2008). In Brazil, for example, a series of relative DEMs have been developed (Naccarato and Pinto 2009; Naccarato et al. 2006, 2007). Such models are typically used to correct flash/stroke density maps (spatial analysis) and can also be applied to temporal analyses. However, a temporal analysis that accurately reflects time-varying performance of the LLS can be an arduous task. A daily efficiency analysis, for example, might not be practical to obtain.

In this work we show three approaches to overcome system performance variations when dealing with temporal and spatial analysis, without requiring the explicit calculation of detection efficiency. The analysis includes procedures to make the LLS data similar to the station (human) observations of thunderstorm activity—thunderstorm days (TDs) and thunderstorm hours (THs)—as well as trimming the LLS data to remove small peak current events. We refer to TDs from LLS (LSTD) and from surface stations (SSTD). The interesting aspect of using this kind of information based on LLS data is that TDs and THs are more likely to be accurately measured than isolated events, such as flashes or strokes. This comes from increasing the probability to get at least one discharge with sufficiently large peak current.

Validation/comparison between surface station observations and data from LLS has been carried out by others, showing, in general, good agreement between the datasets (Pinto et al. 2011; Reap 1993; Reap and Orville 1990), and indicating that data from location systems might be a good way to analyze THs and TDs. An additional benefit of using LSTDs is its direct climatological relationship with long-term historical TD datasets.

The use of LSTDs has some advantages when compared to the SSTD. The LSTD measurements are based on information from a measured subset of real discharges and do not need spatial interpolation (like for isokeraunic maps) in regions were the surface station observations are not available. Another aspect is that SSTD depends on a number of factors related to audibility and station relocation (Changnon 2001).

2. Methodology

The BLDN network started in southeastern Brazil in 1999 and evolved to the other regions over the years. For this reason, the southeastern region was selected for this analysis, providing an 11-yr dataset (from January 1999 to December 2009). Figure 1 shows the rectangular regions used to calculate the lightning parameters for both spatial (upper right) and temporal (lower left) analysis. A reduced area was used for the temporal analysis to avoid the effect of different thunderstorms occurring in different locations at the same time. For the spatial analysis only five years (January 2003–December 2007) were used due to the limitation of the dataset for the scenarios analysis, which is a reduced dataset that was used previously by Bourscheidt et al. (2010).

a. Network scenarios

To evaluate the impact of network performance, different scenarios (i.e., different network configurations) were created using the reprocessing tools available in the central processor (APA2000) of the BLDN (Vaisala 2003). The scenarios were chosen by considering several factors. Initially, sensor combinations with a reasonable number of working days were selected. The working days were defined as those where the uptime of each sensor exceeded 90%. The scenarios were also selected based on reasonable network geometry. Though not important for this work, a more detailed description of the methodology can be found in Bourscheidt et al. (2010).

Using data from the scenarios and from the entire network, three configurations were defined for this work as follows: one "reference scenario" used all of the sensors of the network, and the two other configurations included configuration 1, a reduced 10-sensor network with 7 LPATS and 3 IMPACT (cgf1), and configuration 2, a network using just nine sensors, removing one important IMPACT sensor (cfg2). The reference scenario is used throughout this work as "the truth," providing an objective reference for relative performance using cfg1 and cfg2.

These scenarios simulate the conditions that many of the very low frequency (VLF)/low frequency (LF)
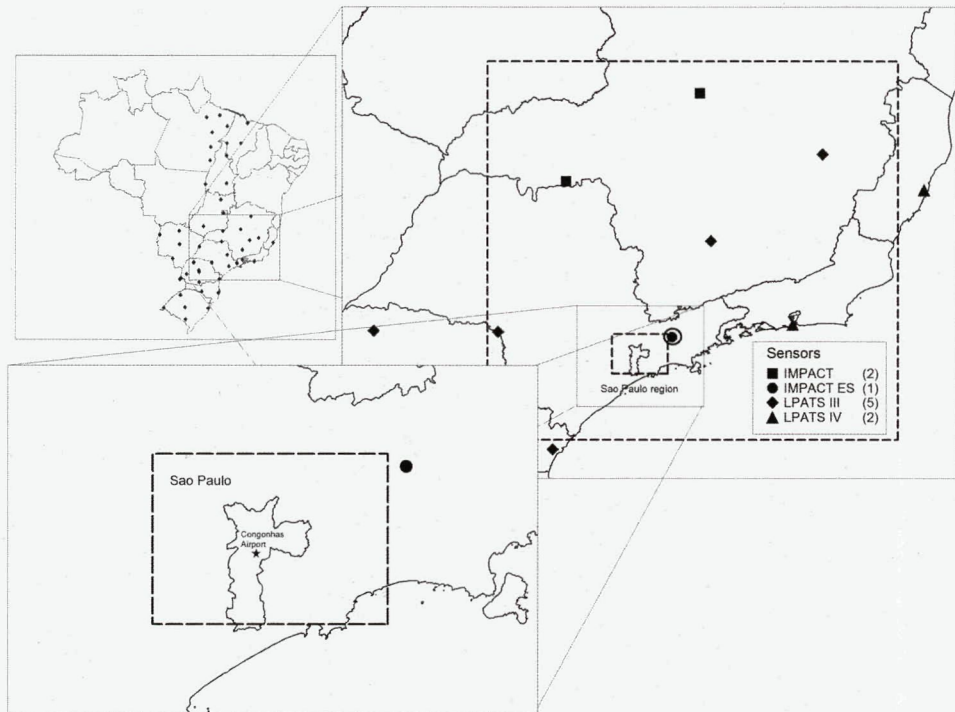
FIG. 1. The study area. (left) The area used to obtain the lightning data (dashed rectangles): (top right) the spatial analysis (large rectangle) and (bottom right) the temporal analysis (small rectangle). The political limits of São Paulo are shown (polygon). The location of Congonhas Airport (star) is indicated, where SSTD observations are available. The circle indicates the sensor that was removed for one of the analysis scenarios (see the text for further details).

networks around the world have experienced, as a result of detection gaps, network expansion, and sensor upgrades during their existence (Brazil, the United States, etc.). In many networks, this also includes improving the sensor sensitivity, allowing the detection of discharges with lower peak currents (e.g., Cummins et al. 1998; Pinto et al. 2006).

b. LLSs data analysis methods

The system performance parameter used on this study is the DE, which, given as the percentage of the total number of events that are detected by the network, depends basically on the peak current of the event (regarding to the signal threshold), on the distance from each sensor, and therefore on the number of reporting sensors. Thus, variations of the network configuration/ geometry result in differences in the system DE, which leads to more or less discharges being detected (Cummins and Bardo 2004).

For this reason, the system performance was evaluated through different scenarios/configurations (described above) using special filters that try to compensate for these instrumentation-related variations in flash counts: "trimmed flashes" (minimum current threshold), THs,

and TDs. THs were obtained by grouping flash data (at least one flash) into hour intervals (60-min steps) and TDs by grouping the lightning data into daily intervals (24-h steps). The spatial scale for grouping depended on the type of analysis, discussed below. These approaches are targeted to mitigate poor detection efficiency, because, as one evolves from flash count to TDs, it is increasingly likely to have at least one discharge with sufficiently large peak current to have one count/ event accumulated in the final analysis. More quantitatively, we can associate the probability of having a TD $[P_{(TD)}]$ to the probability of detecting a flash as follows:

$$P_{(TD)} = 1 - (1 - P_f)^n, \qquad (1)$$

where n is the number of flashes and P_f is the probability of detecting a flash. In this way, the probability of having a TD (or TH) is defined as the probability of detecting at least one flash (or more specifically, not missing all of the flashes). As long as n is larger than one, the TDs will be detected with higher probability than the flashes.

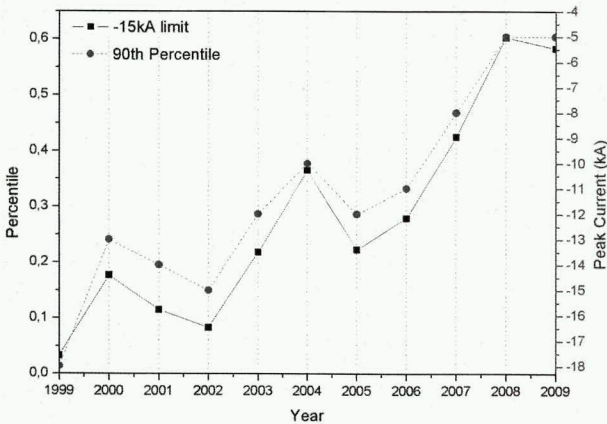The temporal analysis of the effect of system performance variations was carried out using all three "filters"

FIG. 2. Temporal evaluation of network sensitivity: 90th percentile "minimum" current (dashed line with circle) and percent below 15 kA (line with rectangle).

(trimmed flashes, TH, and TD). For the spatial analysis, trimmed flashes were not considered.

To establish the general variation in sensitivity of the BLDN over time, as well as to define a cutoff limit for the trimmed flashes dataset, peak current distributions for the reference scenario were evaluated through an empirical cumulative distribution function (ecdf). The 90th percentile (10% of the currents were smaller) of negative cloud-to-ground (CG)[1] discharges was used as a reference to produce a statistically stable minimum current value, because very low current events usually have a small number of reporting sensors, leading to less reliable peak current estimates. Figure 2 shows how the peak current associated with the 90th percentile varied along the years, as well as the percentage of events with peak currents below 15 kA. The percentage of events that were low current (<15 kA) increased from less than 5% in 1999 to nearly 60% in 2009. The minimum current reduced from about 18 kA in 1999 to 6 kA in 2009. The sensitivity improved rather steadily over the years, related to sensor additions and upgrades[2] (Naccarato and Pinto 2008). This analysis provides a means to normalize network improvements, as well as a way to eliminate any misclassified cloud pulses (Biagi et al. 2007; Cummins and Murphy 2009; Naccarato et al. 2008). Based on these results, a minimum threshold of 19 kA (both positive and negative flashes) was used to equalize the sensitivity for the trimmed flashes temporal analysis. For the spatial

---

[1] Weak positive CG discharges are more susceptible to intracloud (IC) contamination and might not be a good reference to the sensitivity analysis.

[2] The improvements are not related to changes in the peak field to current conversion, which follows the equation described by Cummins et al. (1998) and is the same for the entire studied period.

analysis, as well as the total flash count and both TDs and THs in the temporal analysis, a filter for small positive CG discharges (below 15 kA) was used as a way to mitigate misclassification.

c. Data validation

Ground truth is one important way to validate the data obtained through LLSs. Based on this fact, LSTDs obtained from the BLDN were compared to the SSTDs available from a surface station located near Congonhas Airport, in the core of the urban area of São Paulo, as shown on Fig. 1 (black star). These data, joined with other surface stations, have been used for more than 50 years to build isokeraunic maps. Daily measurements were grouped into monthly averages and then compared to the values obtained through the LLS. Comparisons were made using three different radii around the ground station to emulate variations in hearing distance.

3. Results and discussions

a. Thunderstorm days: Surface station observations versus LLS measurements

Figure 3 shows a scattergram comparison of the TDs obtained through the BLDN (LSTDs) and the human observations (SSTDs) at the Congonhas Airport for the time period from 1999 to 2006. Data are grouped by month. The LSTDs were accumulated over three radii (5, 8, and 10 km). Linear regression fits for each radius are shown as the different lines. The correlation between the two datasets is quite high for the three radii used to compare the data, and the least-biased agreement between the datasets occurs for the radius of 8 km around the station ($R^2 = 0.87$) with a slope of 1.03.

It is also possible to see that as the radius is increased, the number of LSTDs increases in relation to the number of SSTDs (the slope increases). This result is as expected: the hearing and visual ability of the measurer depends strongly on the distance, and the LLS data do not have this limitation. For the 5-km radius, the amount of data from the BLDN is reduced, which affects the number of TDs, decreasing the slope. A similar analysis was carried by Pinto et al. (2011), and they suggest that the small effective radius of the human observer (less than 10 km) is related to the urban/airport noise environment.

b. System performance effects

A monthly temporal analysis of flash counts and the three "filter" parameters is provided in Fig. 4 (accumulated over the temporal analysis domain in Fig. 1).
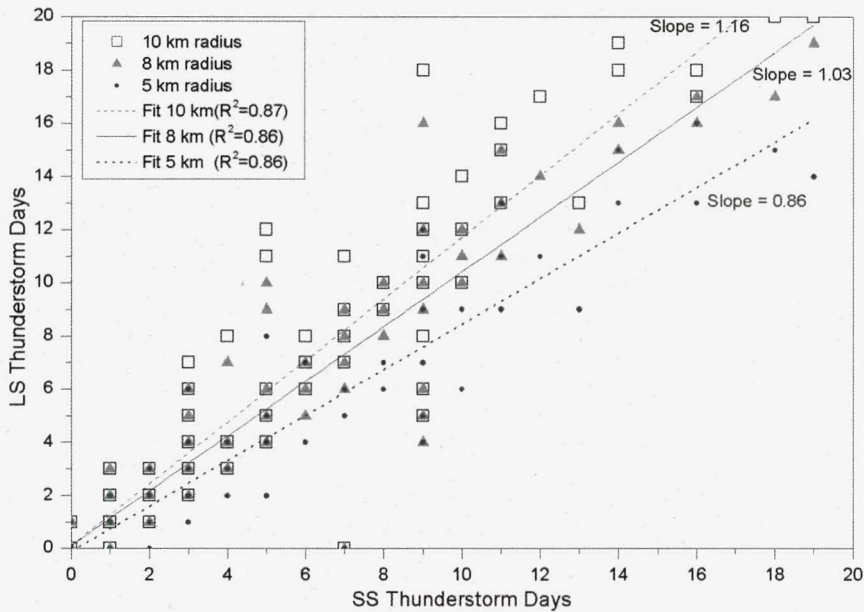
FIG. 3. SSTDs vs LSTDs for different radii around the surface station. Data are grouped by month.

When comparing the monthly flash counts (top panel, filtered for weak positive CG discharges) obtained by the reference network (all sensors) to the first configuration (10-sensor reduced network; gray line), a difference of about 64% (36% of relative efficiency)[3] is observed. This difference is noticeably reduced when trimmed flashes are considered (second panel; 67% average relative efficiency) due to equalization of sensitivity by eliminating low current events. In a similar fashion, when evaluating THs and TDs based on the flash dataset, the differences are further reduced to about 20% and 10%, respectively [78% (TH) and 89% (TD) of relative efficiency], as shown in the third and fourth panels. This analysis shows that even with a reduced network (36% relative efficiency) the THs and TDs values are quiet similar to those produced by the reference configuration.

The results for cfg2 are shown in short dotted gray lines in Fig. 4. This is a very extreme condition, with relative efficiency of only 22%. The relative "performance" increases to 46% for the trimmed dataset and 69% (85%) for TH (TD). Clearly, extreme variations in network performance can be mitigated using TD analysis.

From the spatial perspective, poor LLS performance has a similar effect, as shown in Fig. 5. Two network

conditions were compared: the full network (the reference scenario) and the 10-sensor reduced network cfg1 (scenario 1). The maps enclose the region with the 10 sensors, as described in the methodology (Fig. 1). For these two network configurations, three parameters were analyzed: flash count (FC), THs, and TDs. The TH and TD values were accumulated using a radius of 8 km from the gridcell centroid, based on findings in section 3a. Relative efficiency maps are also shown. The color scale for each parameter (column, in the figure) was selected to produce a similar visual effect for the maps in the first row (full network). To get a more consistent scaling between FC and THs/TDs, equations adapted from previous studies were used (Rakov and Uman 2003; Anderson et al. 1984): $0.11 \times FC^{1/1.4}$ for THs and $0.06 \times FC^{1/1.4}$ for TDs.

The spatial distribution of FC (left column) is poorly reproduced by the reduced network (scenario 1). On the other hand, as one moves to THs and TDs, the spatial distributions become more consistent between the reference condition and scenario 1. This improvement is quantified by the relative efficiency maps (bottom row in Fig. 5). This result is in agreement with the temporal analysis, which was done inside the region shown by the dashed rectangle of Fig. 5. Is important to point out that the relative detection efficiency of TDs and THs is slightly poorer for the spatial analysis than for the temporal analysis, which is likely related to the reduced size of the region used to aggregate the flash data (8-km radius rather than the larger rectangular region). This

---

[3] Relative efficiency based on the ratio between the data from the whole network (reference) to the other analyzed configurations (cfg1/reference and cfg2/reference).
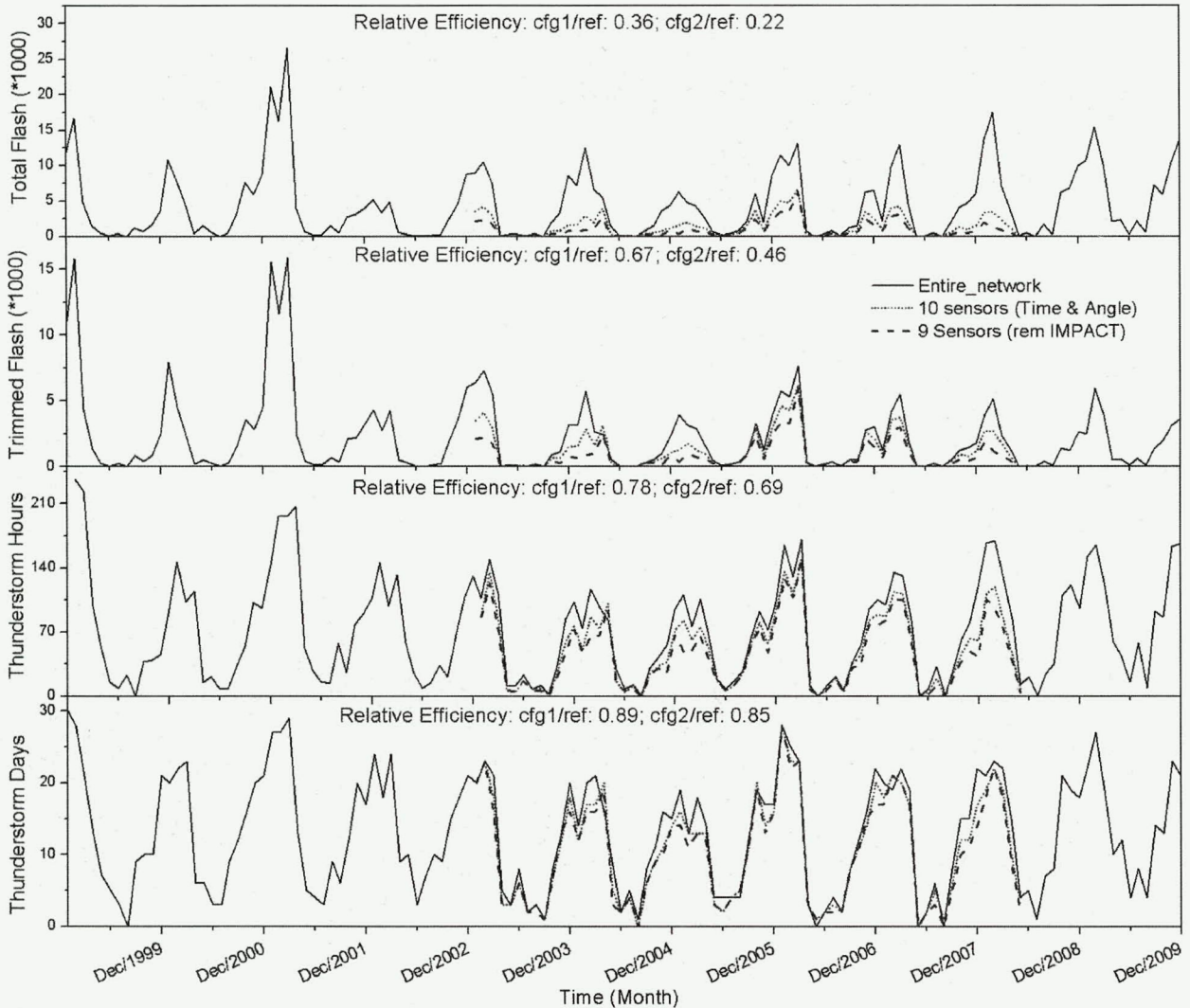
FIG. 4. Temporal evaluation of the network scenarios, using flash data, trimmed flash, THs, and TDs.

effect is associated to the probability equation described in section 2. Note that as the detection efficiency becomes excessively low (mainly at the edges of the reduced network domain), the methods are not able to reproduce the features of the lightning flash density. This is seen in the western corner and northeastern portion of the analysis domain.

4. Conclusions

In this paper we have shown that the effects of the LLSs performance variations on the analysis of lightning data can be minimized by selecting appropriate parameters (filters). Careful selection of a minimum peak current threshold (19 kA in this case) was able to reduce the temporal variation by almost a factor of 2,

even when the flash DE was reduced to less than 20% of the reference performance level. It is possible that smaller variations in performance (maybe in the range of 60%–80% relative DE) could be fully compensated using only a trimmed flash count. TDs provided the least sensitivity to LLS performance variations over time, reducing variations by almost 90%. THs presented comparable behavior, with an almost 80% reduction. From the spatial perspective, a similar improvement was observed, with spatial features being "reasonably" preserved using TD data for scenario 1 (36% relative detection). By using these approaches, temporal and spatial analysis will clearly be less affected by detection efficiency variations.

Another important result is that the TDs obtained from the BLDN are well related to the SSTDs, with the
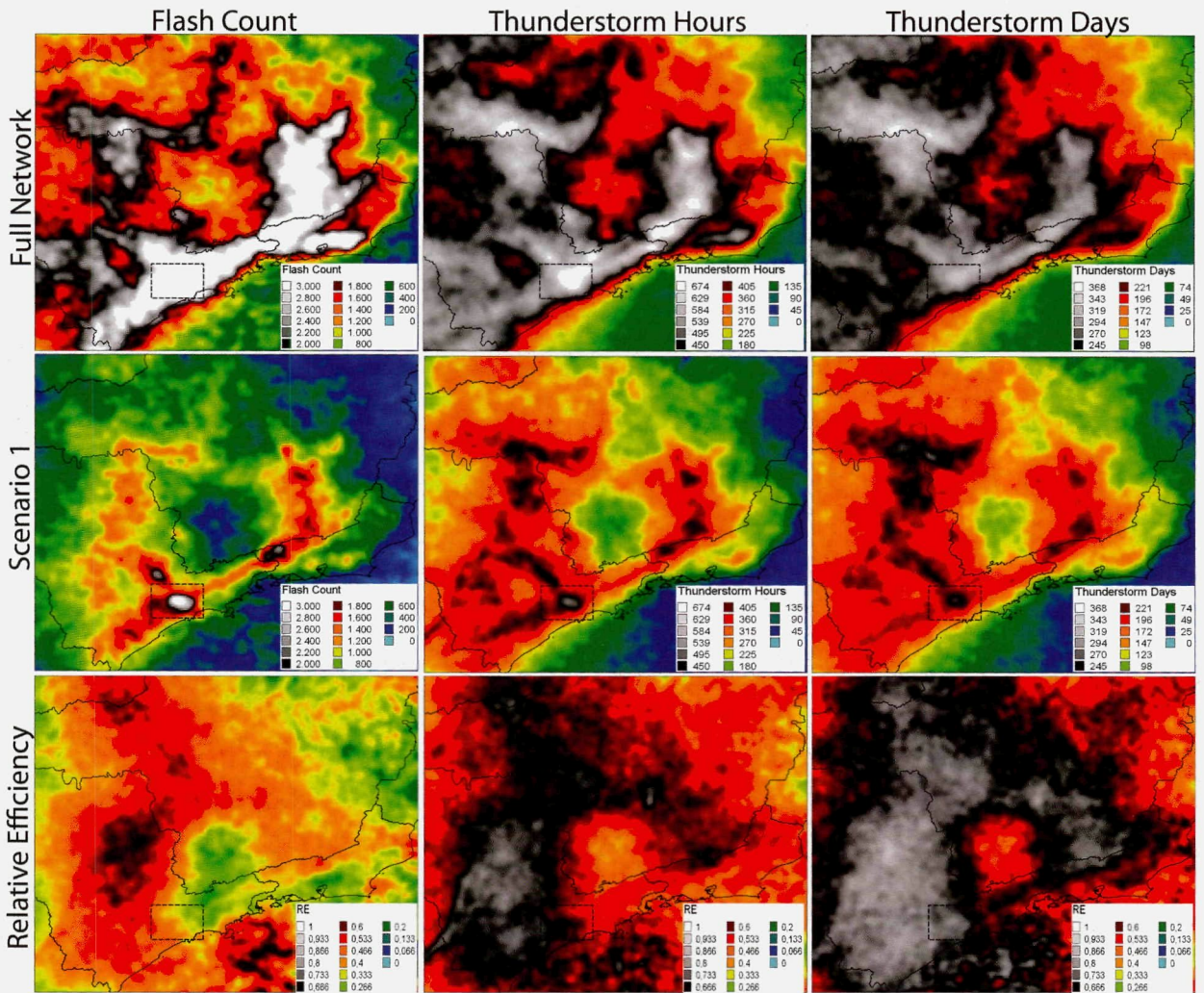
FIG. 5. Spatial evaluation of two network configuration (reference and first scenario) using flash data, THs, and TDs. The region used for the temporal analysis (small dashed rectangle) and (bottom) the relative efficiency are shown.

best agreement for the 8-km radius ($R^2 \geq 0.86$; slope = 1.03). This supports earlier findings that the TDs obtained from the LLSs are in agreement with the expected real conditions.

In conclusion, these methodological procedures might guide users of LLS data toward more reliable climatological spatial and temporal analyses based on THs and TDs, avoiding variations/features associated with (or dependent on) the long-term LLS performance.

REFERENCES

Anderson, R. B., A. J. Eriksson, H. Kroninger, D. V. Meal, and M. A. Smith, 1984: Lightning and thunderstorm parameters. *Proc. Conf. on Lightning and Power Systems,* London, United Kingdom, IEE, No. 236, 5 pp.

Ballarotti, M. G., M. M. F. Saba, and O. Pinto Jr., 2006: A new performance evaluation of the Brazilian Lightning Location System (RINDAT) based on high-speed camera observations of natural negative ground flashes. *Proc. 19th Int. Lightning Detection Conf. (ILDC),* Tucson, AZ, Vaisala. [Available online at http://mtc-m17.sid.inpe.br/col/sid.inpe.br/mtc-m17@80/2006/11.29.16.09/doc/A%20new%20performance%20evaluation.pdf.]

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

Bourscheidt, V., O. Pinto Jr., and K. P. Naccarato, 2010: The effect of different configurations of the Brazilian lightning detection network on the cloud-to-ground lightning spatial distribution. *Proc. 21st Int. Conf. on Lightning Detection/Third Int. Conf. on Lightning Meteorology,* Orlando, FL, Vaisala. [Available online at http://www.vaisala.com/Vaisala

%20Documents/Scientific%20papers/3.Bourscheidt-The%20
Effect%20of%20Different.pdf.]

Changnon, S. A., 2001: Assessment of the quality of thunderstorm data at first-order stations. *J. Appl. Meteor.,* **40,** 783–794.

Cummins, K. L., and E. A. Bardo, 2004: On the relationship between lightning detection network performance and measured lightning parameters. *Proc. Int. Conf. on Grounding and Earthing and First Int. Conf. on Lightning Physics and Effects,* Belo Horizonte, Brazil, Perez Bovolenta, L2, 7 pp.

——, and M. J. Murphy, 2009: An overview of lightning locating systems: History, techniques, and data uses, with an in-depth look at the U.S. NLDN. *IEEE Trans. Electromagn. Compat.,* **51,** 499–518.

——, ——, E. A. Bardo, W. L. Hiscox, R. B. Pyle, and A. E. Pifer, 1998: A combined TOA/MDF technology upgrade of the U.S. National Lightning Detection Network. *J. Geophys. Res.,* **103** (D8), 9035–9044.

Naccarato, K. P., and O. Pinto Jr., 2008: Comparison of LS7000 and IMPACT sensor performance in Paraiba valley (São Paulo) during the summer of 2007/2008. *Proc. Int. Conf. on Grounding and Earthing 2008 and Third Int. Conf. on Lightning Physics and Effects,* Florianopolis, Brazil, Lastro, S29, 451–454.

——, and ——, 2009: Improvements in the detection efficiency model for the Brazilian lightning detection network (BrasilDAT). *Atmos. Res.,* **91,** 546–563.

——, ——, and I. R. C. A. Pinto, 2006: A detection efficiency model for the Brazilian lightning detection network (RINDAT). *Proc. 19th Int. Lightning Detection Conf.,* Tucson, AZ, Vaisala. [Available online at http://www.vaisala.com/Vaisala%20 Documents/Scientific%20papers/A_detection_efficiency_model_ for_the_Brazilian_Lightning_Detection_Network_RINDAT. pdf.]

——, ——, and G. Damata, 2007: Improvements to the detection efficiency model for the Brazilian lightning detection network. *Proc. Int. Conf. on Atmospheric Electricity,* Beijing, China, ICAE/IAMAS/IUGG, 842–845.

——, ——, and M. J. Murphy, 2008: Performance analysis of the BrasilDAT network. *Proc. Int. Conf. on Grounding and Earthing 2008 and Third Int. Conf. on Lightning Physics and Effects,* Florianópolis, Brazil, Lastro, P56, 329–338.

Pinto, O., Jr., K. P. Naccarato, M. M. F. Saba, I. R. C. A. Pinto, S. A. M. Garcia, and A. Cazeta-Filho, 2006: Recent upgrades to the Brazilian integrated lightning detection network. *Proc. 19th Int. Lightning Detection Conf.,* Tucson, AZ, Vaisala. [Available online at http://www.vaisala.com/Vaisala %20Documents/Scientific%20papers/Recent_upgrades_to_the_ Brazilian_Integrated_Lightning_Detection_Network.pdf.]

——, and C. B. Campinho, 2011: Lightning incidence in the southeast Brazil: Comparison of thunderstorm days from different data bases. *Proc. 14th Int. Conf. on Atmospheric Electricity,* Rio de Janeiro, Brazil, INPE/ICAE, No. 40, 4 pp.

Rakov, V. A., and M. A. Uman, 2003: *Lightning: Physics and Effects.* Cambridge University Press, 687 pp.

Reap, R. M., 1993: The use of network lightning data to detect thunderstorms near-surface reporting stations. *Mon. Wea. Rev.,* **121,** 464–469.

——, and R. E. Orville, 1990: The relationships between network lightning locations and surface hourly observations of thunderstorms. *Mon. Wea. Rev.,* **118,** 94–108.

Rubinstein, M., 1995: On the determination of the flash detection efficiency of lightning location systems given their stroke detection efficiency. *Proc. Int. Symp. on Electromagnetic Compatibility,* Zurich, Switzerland Swiss Federal Institute of Technology, 429–432.

Schulz, W., and G. Diendorfer, 1996: Detection efficiency and site errors of lightning location systems. *Proc. Int. Conf. on Lightning Detection,* Tucson, AZ, Global Atmospherics, Inc., 15 pp.

——, and K. L. Cummins, 2008: A method to determine relative stroke detection efficiencies from multiplicity distributions. *Proc. 20th Int. Conf. on Lightning Detection,* Tucson, AZ, Vaisala. [Available online at http://www.vaisala.com/Vaisala %20Documents/Scientific%20papers/A_method_to_determine_ relative_stroke_detection_efficiencies_from_multiplicity_ distributions.pdf.]

Vaisala, Inc., 2003: LP series (LP2000, LP5000) user's guide. Vaisala, M210412EN-B, 258 pp.