Electrical thunderstorm nowcasting using lightning data mining

C. A. M. Vasconcellos[1], C. L. Curotto[2], C. Benetti[1], F. Sato[1] & L. C. Pinheiro[1]
[1]*SIMEPAR Technological Institute, Brazil*
[2]*PPGCC/UFPR, Federal University of Paraná, Brazil*

Abstract

This paper presents a study developed at SIMEPAR (Paraná state weather service) using lightning data for electrical thunderstorm nowcasting. Thunderstorm electrical data collected at SIMEPAR such as lightning location, time of occurrence, current intensity, polarity, etc, are stored in real-time in a relational database. As a first step of this study, Microsoft Business Intelligence bundled with SQL Server 2005 Beta was used to access some of these data and create lightning clusters representing electrical thunderstorms. The clusters were continuously monitored to predict electrical thunderstorm displacement and evolution. Work was undertaken to assess suitability and reliability to the process. Algorithm parameters fitting and cases studies are under development and further work will be done using Weka clustering classes. Once approved, the methodology will be integrated to SisRaios - a Java lightning data visualization, analysis and thunderstorm monitoring and forecasting tool (presented in this paper). It is expected that this computational tool enhanced by the data mining study will aid meteorologists and power companies to monitor electrical thunderstorms, supplying information for starting up maintenance teams, as well as providing a better thunderstorm warnings in general and improving SIMEPAR's nowcasting capabilities.
Keywords: data mining, clustering, lightning, SQL Server, Weka.

1 Introduction

The information presented in section 2 of this paper was collected by a Vaisala lightning detection system operated by SIMEPAR.

WIT Transactions on Information and Communication Technologies, Vol 37,
www.witpress.com, ISSN 1743-3517 (on-line)
doi:10.2495/DATA060161

SIMEPAR, CEMIG, FURNAS (power companies) and INPE (national institute for space research) maintains an integrated lightning detection network (RINDAT) covering the whole south, southeast and center-west of Brazil handling real time lightning information.

The lightning information obtained by RINDAT are used to generated applications for forecasting, electric systems analysis and maintenance, power lines and telecommunication protection projects, meteorological reports and analysis generation for insurance and power companies, among other products.

Historical lightning information data allows establishing more adapted project criteria to telecommunication and power installation regarding protection and optimal route choice. This way, historical data also contributes for longer period studies and researches.

Besides long period researches the lightning detection and locating systems also has been helpful to define better operational decisions strategies, reducing collapses occurrence possibility in energy supply, as well as they have been contributing in a significant way to safety's increase and operational costs reduction.

The RINDAT available information have been quite useful among other applications to accomplish operative maneuvers avoiding regional and local energy supply faults; to warn maintenance teams in advance reducing re-establishment power systems times; to aid the power line faults location; to measure power lines performance, as well to evaluate if improvements are being enough.

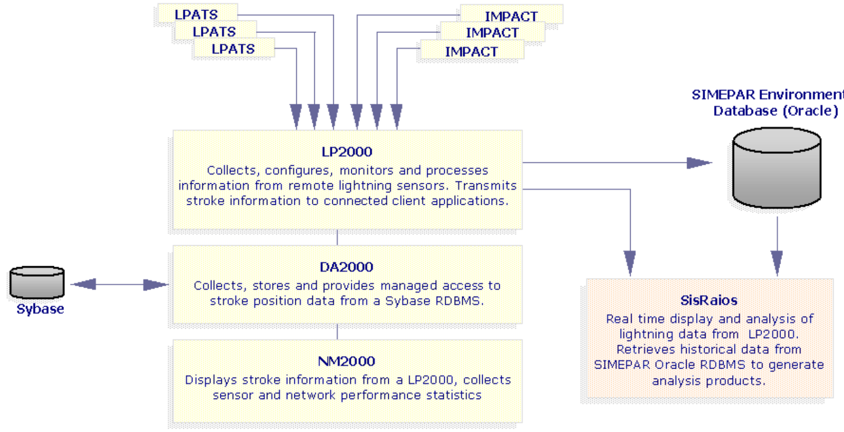

Figure 1: SisRaios and VAISALA-GAI lightning detection and location system integration.

The monitoring and analysis computational environment – SisRaios – works together with the Vaisala lightning detection and locating system operated by SIMEPAR. The set of sensors (LPATS and IMPACT drafted in Figure 1) sends raw lightning data to the central processor unit which collects, configures, monitors and processes that information. These central processors transmits then stroke information (location, peak current, polarity, among other lightning information) to connected client applications such as SisRaios. SisRaios receives this information from central processor and/or SIMEPAR's RDMBS through internet TCP/IP access.

SisRaios receives two types of data: continuously sent LP2000 central unit lightning data and SIMEPAR's and power companies' queried historical database. The first connection type makes SisRaios able to exhibit real time processed lightning. This feature allows electric companies evaluating storm displacement to critical areas, warning maintenance teams for eventual faults power lines. The second connection type is useful in historical analysis and correlation. Queries to SIMEPAR's database provide lightning data for known regions and time.

With this information it is possible to plot lightning density maps (strokes / km²/ year) or keraunics level, for lightning incidence pattern determination. In all cases, data are plotted over customized maps to the power companies, such as counties, power lines, power stations, roads. SisRaios was used to generate figures shown in section 2.

2 Discussion and results

The aim of clustering [1] lightning data [2–4] is to track the clusters displacement allowing thus electrical thunderstorm nowcasting by extrapolation.

First analysis was made with a sample of 2.000 thunderstorm electrical discharges at Paraná state (south of Brazil) with a 30-minute time window. Microsoft Clustering algorithm was employed using SQL Server 2005 Beta [5].

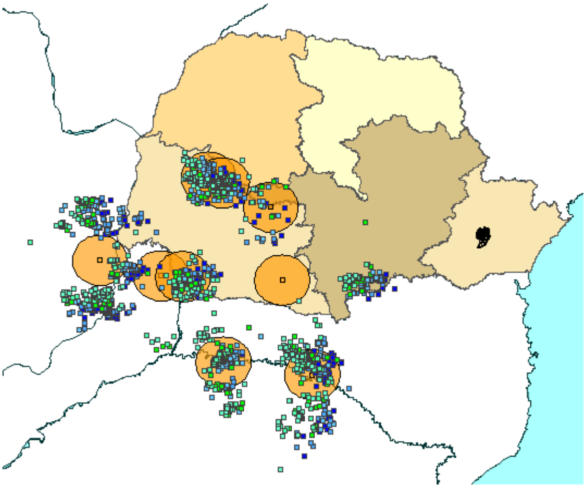

Figure 2: 30-minute time window clusters.

Despite the geographical nature of the event which suggests Euclidean distance measurements K-Means was adopted as clustering method. Scalable

EM (Expectation-Maximization) was also studied because it best handles large data sets [6]. Non scalable K-Means brought better results in position sense.

Figure 1 shows that clusters obtained at this analysis configuration are visually quite representative to the lightning positions. However, the next 30-minute increase in time window showed a poor relationship between clusters from first and second time windows.

In order to be able to track the clusters displacements, a second approach was to reduce the time window to 15 minutes. It was also changed the cluster count parameter from the default value (10) to the best count for each case (selected automatically by the algorithm). Figure 2 shows an example with 3 clusters generated after the changes described above (Time step 1).

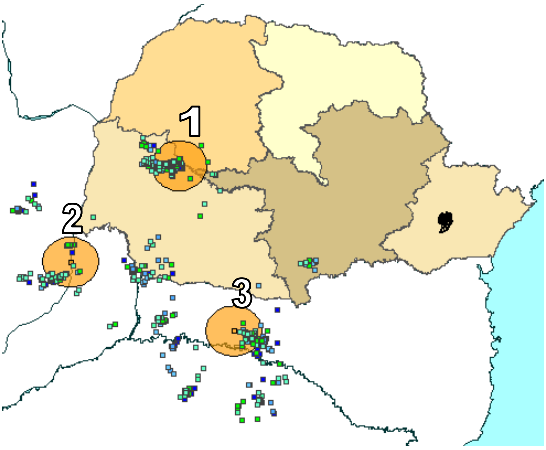

Figure 3: 15-minute time window clusters. Time step 1.

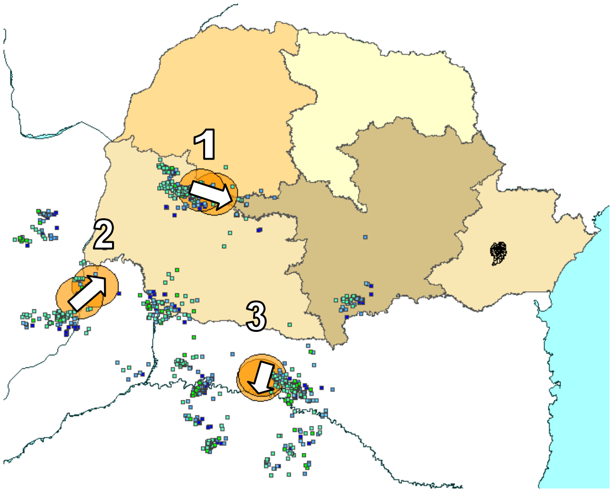

Figure 4: 15-minute time window clusters. Time step 2.

The calculated locations in the next 15-minute step clusters (see Figure 2) corresponded to the actual lightning activity track (except for the cluster 3, which should move to the east but moved to the west). The updated clusters positions are displayed in Figure 3.

Cluster speed and azimuth corresponding to the displacement from time step 1 to time stem 2 are shown in table 1.

Table 1: Clusters displacement from time stem 1 to time step 1.

Cluster	Speed (km/h)	Azimuth
1	72	108º ESE
2	106	50º NE
3	27	197º SSW

A further 15 minutes were increased and in some cases the clusters tracks were lost, as shown in Figure 4 (only 2 clusters plotted).

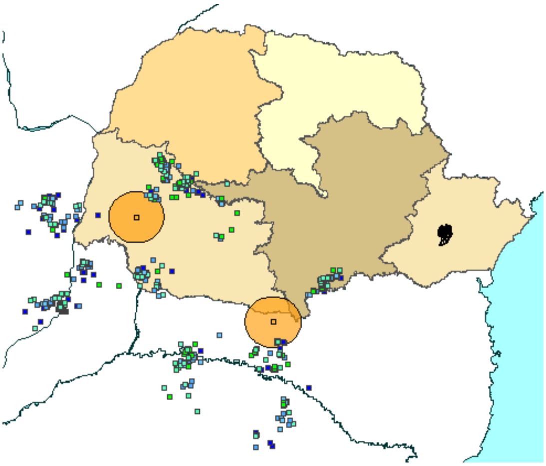

Figure 5: 15-minute time window clusters. Time step 3.

These clusters track losses can be due to the uncoupled phenomenon nature. Other variables (such as orography, cold fronts, radar and satellite data) should be analyzed integrated [7].

3 Conclusions

The clustering analysis using Microsoft Clustering Algorithm lead to good results in well behaved weather conditions (non convective thunderstorms) considering only mathematical data issues. Non scaling K-Means, with automatic cluster counting in 15 minutes time window were the best configuration scheme used.

Other algorithms and parameters adjustments will be evaluated, including Weka Java classes [8] in order to add Data Mining capability at SIMEPAR's visualization and analysis tool (SisRaios).

References

[1] Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. 1996: The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Knowledge Discovery in Communications Of The ACM*. Vol. 39, No. 11.

[2] Lakshmanan, V., and Stumpf, G. J. 2005: A Real-Time Learning Technique to Predict Cloud-to-Ground Lightning. *Fourth Conference on Artificial Intelligence Applications to Environmental Sciences* J5.6, San Diego, USA

[3] Politi, J., Stephany, S., Domingues, M. O., Mendes Jr, O. 2004: Uma metodologia para representação espaço-temporal de ocorrências de descarga nuvem-solo. *Revista Brasileira de Meteorologia*. (submitted)

[4] Politi, J.; Stephany, S.; Domingues, M. O.; Mendes Jr, O. 2004: Implementação de um Ambiente para Mineração de Dados Aplicado ao Estudo de Núcleos Convectivos. *IV WORCAP*. São José dos Campos, Brazil.

[5] Msdn. Accessed in November 2005. URL: http://msdn2.microsoft.com/en-us/library/ms174879.aspx

[6] Bradley, P. S., Fayyad, U. M. and Reina, C. A. 1998: Scaling EM (Expectation-Maximization) Clustering to Large Databases. Microsoft Technical Report. MSR-TR-98-35.

[7] Beneti, C., Calvetti, L., Jusevicius, M. and Gin, R. 2004: The Integration of Radar, Lightning and Satellite Information for Thunderstorm Analysis and Nowcasting. *ILDC*, Helsinki, Finland.

[8] Witten, I. H., Frank, E. 2005*: Data mining : practical machine learning tools and techniques* – 2nd ed. Elsevier, 525 pp.