Electric Power Systems Research xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

An integrated monitoring system and automatic data analysis to correlate lightning activity and faults on distribution networks

André Eugenio Lazzaretti [a,*], Signie Laureano França Santos [b], Kristie Kaminski Küster [b], Luiz Felipe Ribeiro Barrozo Toledo [b], Marcelo Antônio Ravaglio [b], Alexandre Piantini [c], Cleverson Luiz da Silva Pinto [d]

[a] Federal University of Technology – Paraná (UTFPR), Avenida Sete de Setembro 3165, Postal Code 80050-315 Curitiba, PR, Brazil
[b] Institute of Technology for Development (LACTEC), Avenida Comendador Franco 1341, Postal Code 80215-090 Curitiba, PR, Brazil
[c] Institute of Energy and Environment, University of São Paulo, Avenida Professor Luciano Gualberto, 1289, Postal Code 05508-010 São Paulo, SP, Brazil
[d] Energy Company of Paraná (COPEL), Rua José Izidoro Biazetto 158, Postal Code 81200-240 Curitiba, PR, Brazil

ARTICLE INFO

Article history:
Received 11 May 2016
Received in revised form 25 July 2016
Accepted 3 October 2016
Available online xxx

Keywords:
Lightning-correlated faults
Lightning-induced overvoltage
Lightning location systems
Power distribution networks
Power quality analysis

ABSTRACT

Overhead power distribution lines are usually subject to voltage transients during thunderstorms. Lightning strokes, whether direct or indirect, are among the main causes of such power quality disturbances. In order to assist the analysis of lightning effects on power quality indices, a monitoring network and an integrated database analysis, including voltage waveform recorders, local electric field measurements and information from lightning location systems (LLS) is presented. Based on the results of the monitoring systems, a data series of lightning-correlated network faults collected over a period of 15 months within the area covered by the monitoring system was obtained. In order to identify whether a lightning occurrence could have caused a network fault, two pattern recognition methods based on novelty detection approaches have been implemented and compared using the collected data: Parzen density estimator (PDE) and support vector data description (SVDD). The SVDD method is able to identify the impacts of a given lightning occurrence with a 85% success rate when examples from non-related-fault strokes are included in the training procedure. With the proposed data-based approach, it is possible to correlate faults and lightning events with reduced computational effort, include positive discharges in the correlation process, and comprise general features (e.g. soil, slopes etc.) of the region under analysis. These can be highlighted as the main advantages of the proposed procedure in comparison with typically used Monte Carlo-based methods.

1. Introduction

Several events are responsible for variations in voltage and current waveforms in electrical power systems. In the particular case of voltage waveforms (oscillographic records) in a power distribution system, there is a range of events with relevant impacts regarding equipment failure or consumer damage. Therefore, the correct identification of such events is desirable, with the following ones being of particular importance: short-circuits, lightning overvoltages, switching transients along the line, and switching events in distributed power generation [4,2,5].

Numerous international standards specify limits for the power quality indices that indicate the expected quality of service to be received by customers from local distribution networks. If the power utilities fail to comply with the limits set by regulatory agencies, they may incur fines. Lightning flashes are usually the main cause of disturbances and faults in overhead power distribution systems and consequently degrade the power quality indices of the electrical utilities. Some power utilities report that lightning is responsible for about 30% of the outages and 20% of the total interruption time [9,17]. Besides degrading power quality indices, lightning discharges may cause damage to customers' appliances and even reduce the lifetime of equipment connected to the power network. In this sense, some power companies relate around 20% of failures in distribution transformers to lightning activity nearby the overhead distribution systems [9,10]. In some specific regions

* Corresponding author.
E-mail addresses: lazzaretti@utfpr.edu.br (A.E. Lazzaretti), signie@lactec.org.br (S.L. França Santos), kristie.kuster@lactec.org.br (K.K. Küster), luiz.toledo@lactec.org.br (L.F. Ribeiro Barrozo Toledo), marcelo@lactec.org.br (M.A. Ravaglio), piantini@iee.usp.br (A. Piantini), cleverson@copel.com (C.L. da Silva Pinto).

http://dx.doi.org/10.1016/j.epsr.2016.10.005
0378-7796/

characterized by very high ground flash densities, approximately 47% of the total number of distribution transformer failures have been attributed to lightning overvoltages [20,18].

Different methods have been proposed in the literature to correlate power line faults and lightning events. A method that correlates lightning events and relays operations based on the integrated monitored system and calculation results obtained by means of a simulation tool (LIOV-EMTP) for the accurate estimation of lightning-induced voltages (LIOV) on complex power networks using the electromagnetic transient program (EMTP) has been proposed by Bernardi et al. [1] and by Napolitano et al. [13]. A procedure aimed at achieving the best fit between measurements and calculations, which takes into account the uncertainties associated with LLS data has been proposed in Bernardi et al. [1]. In that work, a statistical approach (based on Monte Carlo simulations) is proposed for the evaluation of the probability that a lightning flash detected by a Lightning Location System could cause a fault and, therefore, relay interventions, generally associated with voltage dips. Analysis of lightning overvoltages caused by direct and indirect strokes on overhead distribution lines with surge arresters or a periodically grounded shield wire can be found in Paolone et al. [15], Piantini and Janiszewski [19].

In order to establish a correlation between faults along the distribution system and lightning strokes, a monitoring network, designed to measure fast electromagnetic transients on energized distribution networks (both on high and low voltage circuits) and the local atmospheric electric field, has been installed in the city of Guaraniaçu, Brazil. This network has been developed in the framework of a research and development project whose main aim is the evaluation of the effects of lightning on power quality indices of overhead power distribution networks. To achieve this goal, the monitoring network also includes the integration with a Lightning Location System. With this integrated monitoring system, a lightning occurrence data series was collected over a period of 15 months within the monitored area.

By using the results of this monitoring network, the idea of this paper is to estimate whether a lightning event could have caused a network fault based on the data series collected. To do that, two classification models have been compared: a Parzen density estimator and a support vector data description [22]. Both methods perform the classification by modeling a closed boundary around one of the classes, allowing a novelty detection procedure. In the PDE, the probability density function is estimated using a nonparametric model and the probability value is used as a reference for the classification process. On the other hand, the SVDD model is based on a boundary optimization problem and the distance to the boundary provides the novelty score.

The data series constitutes of chosen 196 identified lightning occurrences, 84 of which are timely correlated with fault events detected by the monitoring system, and 112 were lightning strokes that did not provoke faults. The network fault data are then the observed outcomes in the overhead lines protected with surge arresters covering a particular distribution feeder (named *Cazela*). When compared to Monte Carlo-based methods, the main advantages of the proposed methodology are:

- **Reduced computational effort.** Monte Carlo-based methods normally perform several simulations considering different parameters and electromagnetic models, which results in a computational expensive approach. In the proposed procedure, the classification is based on a simple novelty detector, whose parameters are estimated in an off-line phase;
- **Include positive discharges in the correlation process.** Electromagnetic simulation models are normally designed to simulate downward negative flashes. In the proposed procedure, it is possible to include both negative and positive strokes, as the dataset used to estimate the classification model contains both polarities;
- **Comprise general features of the region under analysis.** In a full Monte Carlo procedure, many different parameters need to be varied in order to take into account the uncertainties related to the LLS, soil characteristics, terrain slopes etc. In a data-based approach, all these features are grouped and represented in the underlying classification model.

This paper is organized as follows. In Section 2, the monitoring systems are presented. In the sequence, Section 3 presents the final dataset distribution and the features that are used in the classification process. Section 4 describes the two classification models (PDE and SVDD), highlighting the benefits and drawbacks of the each method. Section 5 presents the results and, in Section 6, the main conclusions and proposals for future works are outlined.

2. Monitoring systems

The monitoring network consists of a Lightning Location System, four field mill (FM) sensors, and four waveform recorders located in the city of Guaraniaçu, Brazil. The choice for the location of each waveform recorder was based on a historic dataset of ground flash density and the number of interruptions of each feeder plus the number of lightning-correlated distribution transformer failures. In this way, four different places have been selected, as presented at the bottom right of Fig. 1.

The electric field sensors have been installed considering different aspects of the region, such as the location that maximizes the area that can be covered, considering the limitations of the sensor (up to 15 km) and the presence of mobile data services (e.g. General Packet Radio Service). The location of the sensors is also presented in Fig. 1. The stations are named *Campo Bonito*, *Santa Luzia*, *Ibema*, and *Belarmino*. The ground flash density of this region is about 5 flashes/km^2/year. The analysis is initiated when a new waveform is recorded (with total time of 400 ms). Then the waveform is segmented and, for each segment, the automatic classification is performed and the FM and LLS databases are analyzed. In the following subsections, we present the details of each module and the procedure to establish the correlations.

2.1. Waveform monitoring systems

The system has four units and each unit is composed of an oscillographer, a set of three-phase high-voltage capacitive dividers and a three-phase low-voltage probe set connected to a multichannel digital 100 MHz oscillograph. This configuration allows simultaneous sampling of the voltages on the high and low voltage circuits, including the neutral. Each monitoring system is composed of capacitive voltage dividers with adequate frequency response for measuring lightning impulse voltages, the acquisition system (Yokogawa oscillographer, model SL-1000, with sampling frequency adjusted to 10 MHz), uninterruptible power supply, and the communication system, which also enables to synchronize the oscillographer through a Network Time Protocol with the LLS.

2.2. Waveform segmentation and classification

The set of oscillograph loggers was designed and installed in the distribution network in such a way that every time the voltage exceeds some defined threshold, the referred waveform is logged for subsequent analysis. A notorious problem with this method is the large amount of data logged by the systems, which ends up being unfeasible to be manually analyzed. This scenario suggests the need of an automatic data classification scheme, similarly to the ones that have been used for power quality classification [12].