Applied Mathematical Modelling 40 (2016) 9067–9078

Contents lists available at ScienceDirect

Applied Mathematical Modelling

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

Robust functional analysis for fault detection in power transmission lines

Gilmar Rosa, Marcelo Azevedo Costa*

Universidade Federal de Minas Gerais, Production Engineering, Av. Antônio Carlos 6 627, Pampulha, 31270901 Belo Horizonte, Minas Gerais, Brazil

ARTICLE INFO

Article history:
Received 25 August 2015
Revised 7 April 2016
Accepted 25 May 2016
Available online 3 June 2016

Keywords:
Functional analysis
Non-linear quantile regression
Fault detection
Power transmission lines

ABSTRACT

Fault detection methods in power transmission lines aim to detect deviations of the electrical signals from the expected behavior of such signals under normal operating conditions. One approach is to model, as accurately as possible, the expected behavior of the electrical signals under normal operating conditions. Furthermore, even under normal conditions, electrical signals are subject to random noises. Therefore, upper and lower limits must be established. The larger the limits, the harder the fault detection. On the contrary, the narrower the limits the more likely to detect false faults. Functional analysis of power transmission lines was originally proposed to represent the behavior of the electrical signals and to estimate the upper and lower limits under normal operating conditions. Nonetheless, the originally proposed estimates are biased and rely on statistical assumptions that do not hold in practice. This work proposes new methods to estimate the parameters of the functional model and new upper and lower limits that do not rely on specific statistical assumptions. Simulated and real case results show that the proposed robust functional analysis reduces bias and provides more accurate false fault detection rates, as compared to the previous method.

1. Introduction

An Electric Power System (EPS) consists of many components that can be damaged over time, either by natural deterioration or misuse of equipment. In general, detection of faults in any EPS component in early stages is crucial, since faults may affect the supply of energy to consumers. Moreover, a fault in an EPS represents an abnormal condition such as an electrical failure in one or more elements in the electrical system.

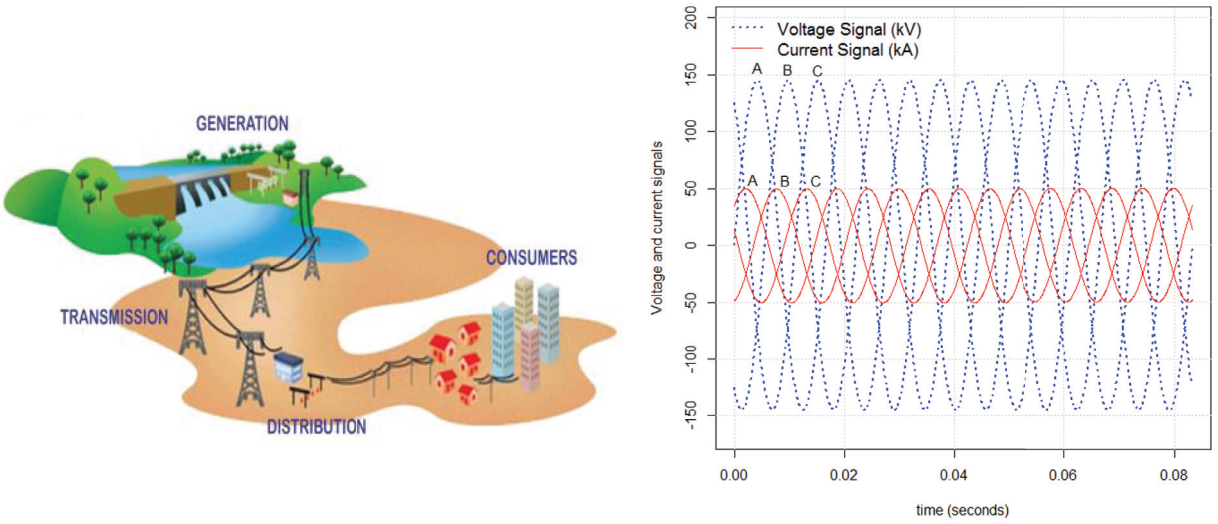
Electricity is transmitted from generation stations to consumers along the transmission lines, which traverse great distances as illustrated in Fig. 1(a). The transmission lines typically operate in a three phase system (A, B and C), conducting alternating currents. This configuration reduces energy loss in power transmission. In general, there is one transmission line for each phase, and each phase is represented by its voltage and current signals. Under normal operating conditions, the expected signals of voltage and current are represented by cosine functions, shown in Eqs. (1) and (2).

$$v_{(t)} = V_0 \cdot \cos(wt), \tag{1}$$

$$i_{(t)} = I_0 \cdot \cos(wt - \varphi), \tag{2}$$

* Corresponding author. Tel.: +55 3134094881; fax:+55 3134094883.
E-mail address: macosta@ufmg.br (M.A. Costa).

http://dx.doi.org/10.1016/j.apm.2016.05.042
0307-904X/



(a) The electrical power system is composed of generation, transmission and distribution systems.

(b) The voltage and current signals of a three phase transmission line are represented by cossine functions.

Fig. 1. Energy distribution system (a) and the expected behavior of the voltage and current signals for each phase (A, B and C) of a transmission line.

where φ is the delay between the current signal $i_{(t)}$ and the voltage signal $v_{(t)}$, V_0 and I_0 are the peak values of the voltage and current signals, respectively, and $w = 2\pi f$ is the angular velocity. In the Brazilian system, $f = 60\,\text{Hz}$ and $\cos(\varphi)$ is also known as the power factor (PF). Eqs. (1) and (2) show that the current and voltage signals are time dependent. Fig. 1(b) shows the voltage and current signals of a three phase transmission line with phases A, B and C. Faults represent conditions in which signal behavior deviates significantly from expected behavior.

In general, wavelet transform and Fourier transform are the dominant methods for feature extraction of electrical signals in transmission lines. The features are used as inputs for different classification methods, in which Artificial Neural Networks (ANN) and Fuzzy Logic (FL) are the most common methods. For example, Rafinha and Moshtagh [1] present a method for fault location in underground lines using wavelet transform and FL. Using simulated scenarios, the authors concluded that wavelet transform effectively represent time characteristics of the line signals, whereas FL can locate and classify faults accurately. Aggarwal et al. [2] present a method for detection and location of faults based on ANN and genetic algorithm (GA), using only current signal information. The authors observed that the spectral energy and peak current are considerably different among different fault types. Sundaravaradan et al. [3] present a survey of the most recent methods applied in fault detection in transmission lines.

Nonetheless, different statistical models have been presented. Morales et al. [4], Samantaray [5] and Morales and Orduna [6] use Principal Component Analysis (PCA) for fault classification. Yusuffa et al. [7] use wavelet transform and statistical regression methods. Upendar et al. [8] proposed a detection and fault classification method using wavelet transform and regression trees (CART). The authors compared the proposed algorithm with the classification results using ANN. The simulation study shows that the proposed classification method is simpler than ANN and achieves high accuracy. Furthermore, statistical models using simple trigonometric functions, instead of complex mathematical equations, have been successfully applied to detect and classify faults in transmission lines [9-11].

In particular, Gomes et al. [10] introduced a novel monitoring system for a single phase of a transmission line in which the mean behaviors of the voltage and current signals under normal or nominal operating conditions are represented as a two-dimensional ellipse. Furthermore, by assuming additive independent Gaussian noises for voltage and current signals, upper and lower limits for the ellipse are estimated. This method can be seen as a non-linear two-dimensional control chart and it allows monitoring of the electrical signal.

Control charts aim at determining if a manufacturing process is in a state of statistical control, i.e., the process is following its expected mean and expected variability under normal operating conditions. These charts can be extended to detect faults in transmission lines, as long as the expected mean and variability under normal operating conditions are known. Therefore, one may think of multivariate control charts for both voltage and current signals. In this case, the Hotelling's T^2 control chart is the most common used tool [12,13]. Robust alternatives to Hotteling's T^2 control chart also have been proposed [14]. In brief, the T^2 control chart is equivalent to the squared Mahalanobis distance, as shown in Eq. (3)

$$\chi_0^2 = n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{x}} - \boldsymbol{\mu}), \tag{3}$$

where n is the sample size, $\bar{\boldsymbol{x}}$ is the observed sample mean, $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ is the covariance matrix. For example, using current and voltage signals, $\boldsymbol{x}_{(t)}^{\mathrm{T}} = [v_{(t)},\ i_{(t)}]$ and $\bar{\boldsymbol{x}}^T = [\bar{v}, \bar{i}]$, $\bar{v} = \sum_{t=1}^{n} v_{(t)}/n$ and $\bar{i} = \sum_{t=1}^{n} i_{(t)}/n$. The parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ must be known in advance.

As shown in Eq. (3), it is commonly assumed that the mean vector is constant under normal operating conditions, i.e., does not change over time. On the contrary, the expected behavior of voltage and current signals under normal operating conditions does change over time, as previously shown. In addition, the method is computationally simpler than most commonly applied fault detection methods in Electric Power Systems, such as Wavelets [1,2,4,7], Fourier Transforms [15] and models of Artificial Intelligence [2,16], and it achieves superior fault detection rates. Furthermore, mathematical and statistical modeling of transmission lines under normal operating conditions are much simpler than modeling specific fault conditions, such as transitory effects produced by atmospheric current discharges [17].

Functional analysis of power transmission lines aims to represent voltage and current signals under normal operating conditions using stochastic models, as accurately as possible. Consequently, faults can be detected as fast as possible, since upper and lower bounds for the signals under normal operating conditions can be statistically estimated. Furthermore, false faults can be controlled. Briefly, the faster a fault is detected the more likely the occurrence of false faults. This is a standard problem in statistical signal analysis, and often neglected by standard fault detection methods. Furthermore, estimates of the functional model may present bias because some of the statistical assumptions do not hold in practice, such as independent Gaussian noise errors. Therefore, the functional model proposed by Gomes et al. [10] has major limitations: it produces both higher false fault rates than originally expected and also biased estimates of the current and voltage signal parameters.

This work proposes improvements to the functional analysis model of transmission lines. A constrained linear optimization problem achieves more accurate estimates of current and voltage signal parameters. Using quantile regression and elliptical equations, new upper and lower bounds are estimated achieving a more precise false fault rate, as compared to the previous approach. Thus, this work evaluates the main limitations of the method proposed by Gomes et al. [10], which are: (i) the estimated confidence intervals are inconsistent with the α-level chosen by the user, and (ii) the estimates of the parameters are biased. In addition, we propose new estimates for the parameters of the ellipse, and new upper and lower limit estimates using non-linear quantile regression [18,19]. Therefore, we do not rely on any statistical assumptions about noise distribution. In brief, to estimate the mean parameter using a sample of size n, one can solve the following optimization problem, $\min_{\mu} \sum_{i=1}^{n} (x_i - \mu)^2$, which leads to $\hat{\mu} = \bar{x}$, where $\bar{x}$ is the sample mean. Similarly, the estimated median is the solution of the following optimization problem: $\min_{m} \sum_{i=1}^{n} |x_i - m|$. In general, the pth quantile, hereafter named q, can be estimated as the solution of an optimization problem presented in Eq. (4).

$$\min_q \frac{1-p}{n} \sum_{y_i<q} |y_i - q| + \frac{p}{n} \sum_{y_i>q} |y_i - q|. \tag{4}$$

Eq. (4) can be applied to estimate quantile regression models, for example, a linear equation model can be estimated using Eq. (4) and $q_{(\beta_0,\beta_1)} = \beta_0 + \beta_1 x$ (see [18,19]). In the present work, a non-linear equation is applied. It is assumed that both the upper and lower limits are represented by ellipse equations.

This paper is organized as follows: Section 2 reviews the functional analysis for power transmission lines, presents new estimates for the functional model and proposes new upper and lower limits using non-linear quantile regression. Section 3 presents results using simulated and real case data sets. Discussion and conclusion are presented in Section 4.

2. Materials and methods

2.1. Monitoring statistic for transmission lines

Although the transmission line has three phases, a monitoring system can be applied to each phase individually, as if they were independent. For each phase, a control chart can be applied to voltage and current signals, separately. Gomes et al. [10] proposed a new approach in which both current and voltage signals are simultaneously controlled using a two-dimensional control region. Following Gomes et al. [10], under normal operating conditions and assuming no noise, voltage and current signals can be represented by the ellipse given by Eq. (5)

$$V_0^2 v_{(t)}^2 - 2V_0 I_0 \cos(\varphi) v_{(t)} i_{(t)} + I_0^2 i_{(t)}^2 - V_0^2 I_0^2 \sin(\varphi) = 0. \tag{5}$$

Eq. (5) can be applied directly to any point $P(v_{(t)}, i_{(t)})$. If the result is zero, then the point $(v_{(t)}, i_{(t)})$ belongs to the ellipse, regardless the value of time t. Furthermore, a control region is created assuming additive random variables for both voltage and current signals, as shown in Eq. (6).

$$\begin{aligned} v_{(t)} &= V_0 \cdot \cos(wt) + \xi_V \cdot V_0, \\ i_{(t)} &= I_0 \cdot \cos(wt - \varphi) + \xi_I \cdot I_0, \end{aligned} \tag{6}$$

where ξ_V and ξ_I are two random variables, with means of zero and variances of σ_V^2 and σ_I^2, respectively. They represent noise components associated to voltage and current signals, respectively.

The monitoring statistics are the standardized values of voltage and current. First, instant values of voltage and current signals are divided by their peak values: $x_{(t)} = \frac{v_{(t)}}{V_0}$, and $y_{(t)} = \frac{i_{(t)}}{I_0}$. Second, the points are rotated using an angle of $\theta = \frac{\pi}{4}$, or $\theta = \frac{3\pi}{4}$. The standardized and rotation operations further simplify the equation of the ellipse, Eq. (6), which is reduced to Eq. (7).

$$\frac{x_{r(t)}^2}{a^2} + \frac{y_{r(t)}^2}{b^2} = 1, \tag{7}$$

where $x_{r(t)}$ and $y_{r(t)}$ are the voltage and current signals after standardization and rotation, respectively, and a and b are constant values related to V_0, I_0 and φ. An alternative representation of Eq. (7) is presented in Eq. (8).

$$\begin{aligned} x_{r(t)} &= a \cdot \sin(wt) + \xi_V, \\ y_{r(t)} &= b \cdot \cos(wt) + \xi_I. \end{aligned} \tag{8}$$

From Eq. (8), and assuming independent Gaussian variables for ξ_V and ξ_I, upper and lower limits for the control region are given by Eq. (9).

$$\begin{aligned} \frac{x_{r(t)}^2}{\left(a + \mathcal{Z}_{\frac{\alpha}{2}} \cdot \sigma_V\right)^2} + \frac{y_{r(t)}^2}{\left(b + \mathcal{Z}_{\frac{\alpha}{2}} \cdot \sigma_I\right)} &\leq 1, \\ \frac{x_{r(t)}^2}{\left(a - \mathcal{Z}_{\frac{\alpha}{2}} \cdot \sigma_V\right)^2} + \frac{y_{r(t)}^2}{\left(b - \mathcal{Z}_{\frac{\alpha}{2}} \cdot \sigma_I\right)} &\geq 1, \end{aligned} \tag{9}$$

where $\mathcal{Z}_{\frac{\alpha}{2}}$ is the $\mathcal{Z}$-score statistic with $1 - \frac{\alpha}{2}$ confidence level. For instance, if $\alpha = 0.3\%$ then $\mathcal{Z}_{\frac{\alpha}{2}} \approx 3$. Fig. 2(a) shows voltage and current signals over time assuming stochastic error components. Fig. 2(b) shows the elliptical behavior of voltage and current signals in one phase of a transmission line. Fig. 2(c) shows the ellipse after standardization and rotation. Fig. 2(d) shows upper and lower limits, assuming independent and Gaussian noises, under normal operating conditions.

2.1.1. Upper and lower limits estimates

Estimates for the V_0, I_0, and $\cos(\varphi)$ parameters have been proposed [10] based on the solution of a linear equation system, which is a simplified version of Eq. (5). In this case, all terms are divided by $V_0^2 I_0^2 \sin(\varphi)$. Let $t = 1, \ldots, N$ be the time period in which the voltage and current signals are under normal operating conditions. Let $\mathbf{v} = [v_{(1)}, v_{(2)}, \ldots, v_{(N)}]^T$ be the voltage signal column vector and $\mathbf{i} = [i_{(1)}, i_{(2)}, \ldots, i_{(N)}]^T$ the current signal vector. Let $\mathbf{X} = [\mathbf{v}^2 \ \mathbf{vi} \ \mathbf{i}^2 \ \mathbf{v} \ \mathbf{i}]$ be a matrix of dimensions $N \times 5$, and $\beta^T = [\beta_1 \ \beta_2 \ \beta_3 \ \beta_4 \ \beta_5]$ be the vector of parameters ($\boldsymbol{\beta}_{5\times 1}$). The estimates of the parameters are the solution of the following linear equation: $\mathbf{X}\beta = 1$, where 1 is the unitary vector. In this case, the solution can be written as $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T 1$. The estimates of V_0, I_0, and $\cos(\varphi)$ are shown in Eq. (10).

$$\begin{aligned} \cos\left(\hat{\varphi}\right) &= \frac{-\hat{\beta}_1}{\sqrt{\hat{\beta}_1 \cdot \hat{\beta}_3}}, \\ \hat{I}_0 &= \frac{1}{\sqrt{\hat{\beta}_1 \cdot \sin^2\left(\hat{\varphi}\right)}}, \\ \hat{V}_0 &= \frac{1}{\sqrt{\hat{\beta}_3 \cdot \sin^2\left(\hat{\varphi}\right)}}. \end{aligned} \tag{10}$$

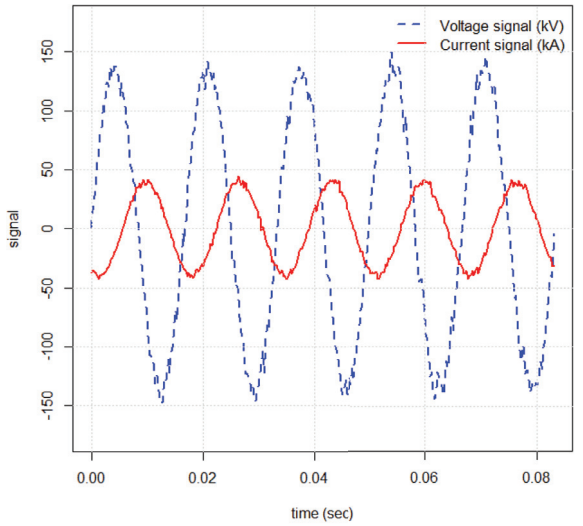
The estimated parameters shown in Eq. (10) are used to standardize both the voltage and current signals, and to estimate parameters a and b (see Eq. (7)). To estimate the variance parameters of the stochastic components, σ_V^2 and σ_I^2, we first estimate the residuals of the standardized and rotated signals using a one-dimensional optimization procedure for each point in the sample, as shown in Eq. (11). Thus, for each value of $i = 1, \ldots, N$, voltage and current residuals are calculated by solving one non-linear optimization problem.

$$t_i^* = \arg\min_t \left\{ \left(x_{r(i)} - \hat{a} \cdot \sin(wt)\right)^2 + \left(y_{r(i)} - \hat{b} \cdot \cos(wt)\right)^2 \right\}, \tag{11}$$

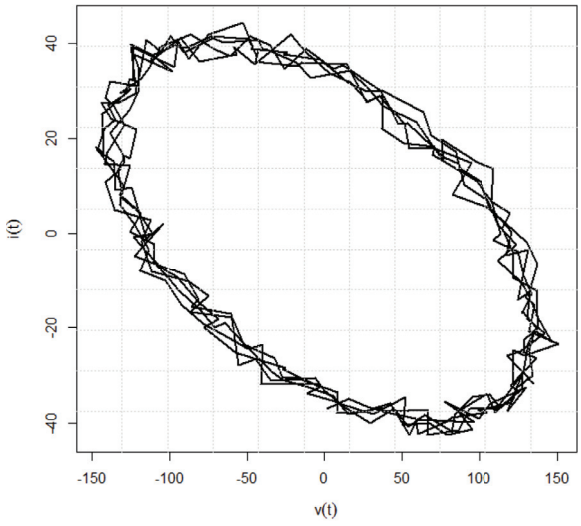
$$\begin{aligned} r_{x_{(i)}} &= x_{r(i)} - \hat{a} \cdot \sin\left(wt_i^*\right), \\ r_{y_{(i)}} &= y_{r(i)} - \hat{b} \cdot \cos\left(wt_i^*\right). \end{aligned} \tag{12}$$

From the estimated residuals, the estimated variances are

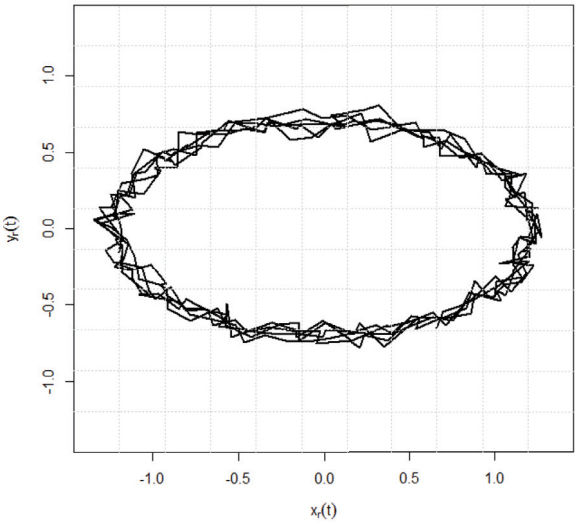
$$\begin{aligned} \hat{\sigma}_V^2 &= \frac{\Sigma_{i=1}^N r_{x_{(i)}}^2}{N-1}, \\ \hat{\sigma}_I^2 &= \frac{\Sigma_{i=1}^N r_{y_{(i)}}^2}{N-1}. \end{aligned} \tag{13}$$



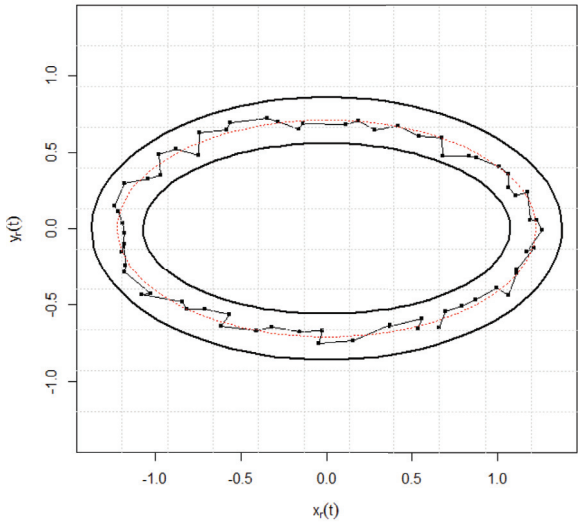
(a) Current and voltage signals over time assuming stochastic noise components.



(b) Voltage and current signals projected into the bi-dimensional space(voltage versus current).



(c) Voltage and current signals after standardization and rotation operations.



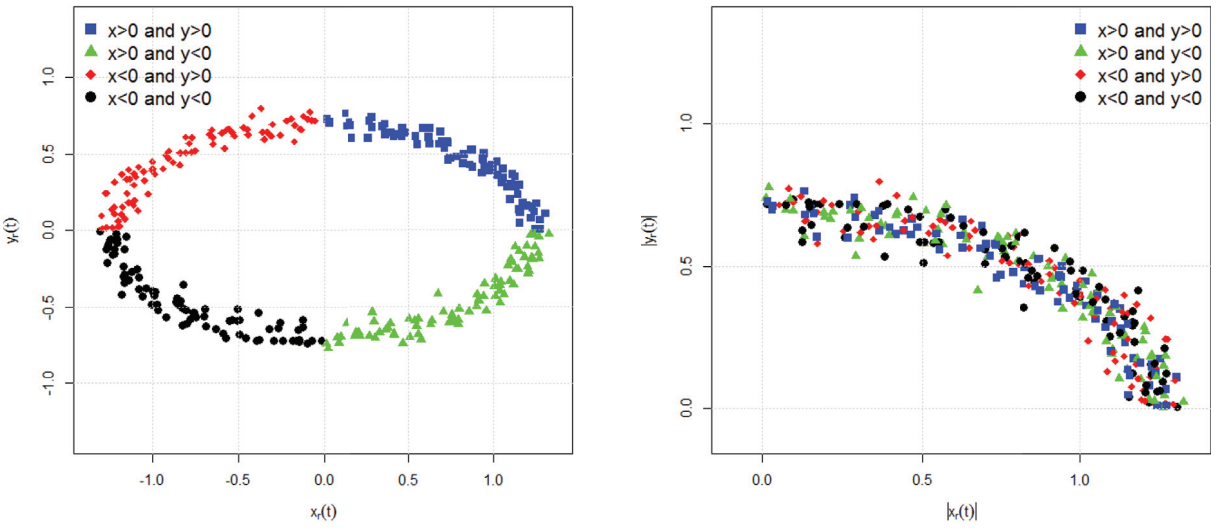
(d) Estimated upper and lower limits assuming stochastic error components under normal operating conditions.

Fig. 2. Voltage and current signals over time (a). Elliptical behavior of voltage and current signals of a phase in a transmission line (b). Voltage and current signals after standardization and rotation (b). Upper and lower limits, assuming independent Gaussian noises under normal operating conditions.

As will be shown in the simulation study, the proposed estimates have major limitations, such as assuming independence between the voltage and current residuals. Furthermore, the estimated values present some bias, and the Gaussian noise assumption does not hold in practice. Therefore, the confidence level of the control limits is much lower than the α-level chosen by the user, which leads to a higher level of false detection rates. Gomes et al. [10] proposed the use of residual percentiles in the upper and lower equations in order to overcome the Gaussian assumption. Nevertheless, this solution further compromises the α-level, which significantly deviates from the original value chosen by the user. We provide a robust approach to estimate the parameters of the ellipse and to estimate upper and lower limits.

2.2. Robust upper and lower limits using non-linear quantile regression

Our proposal does not rely on any stochastic assumption about the voltage and current noises. We do rely on a parametric form for upper and lower limits, which is the ellipse equation. Nevertheless, before estimating upper and lower limits, we propose a more efficient procedure to estimate the parameters of the ellipse equation.



(a) Voltage and current signals after standardization and rotation operations.

(b) Absolute values of the standardized and rotated voltage and current signals .

Fig. 3. The non-linear quantile regression equation (Eq. (16)) is applied to the absolute values of the standardized and rotated signals (b), since they are symmetric with respect to the axis (a).

Using the voltage and current signals, the general equation of a second order conic section is a second order polynomial equation, as shown in Eq. (14).

$$\beta_1 v_{(t)}^2 + \beta_2 v_{(t)} i_{(t)} + \beta_3 i_{(t)}^2 + \beta_4 v_{(t)} + \beta_5 i_{(t)} + \beta_6 = 0. \tag{14}$$

Following Fitzgibbon et al. [20], Eq. (14) represents an ellipse if the following constraint is applied: $\beta_2^2 - 4\beta_1\beta_3 < 0$. Therefore, the parameters of the ellipse can be estimated solving the optimization problem shown in Eq. (15) (see Halir and Flusser [21]).

$$\begin{aligned} &\min\ (\mathbf{X}\beta)^{\mathrm{T}}\mathbf{X}\beta \\ &\text{subjectto}: \beta^{\mathrm{T}}\mathbf{C}\beta = 1, \end{aligned} \tag{15}$$

where $\mathbf{X} = [\mathbf{v}^2\ \mathbf{vi}\ \mathbf{i}^2\ \mathbf{v}\ \mathbf{i}\ 1]$, $\beta^{\mathrm{T}} = [\beta_1\ \beta_2\ \beta_3\ \beta_4\ \beta_5\ \beta_6]$, and

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 2 & \\ 0 & -1 & 0 & \cdots 0 \\ 2 & 0 & 0 & \\ & \vdots & & \ddots \vdots \\ & 0 & & \cdots 0 \end{bmatrix}.$$

Eq. (15) provides robust estimates for the parameters of the ellipse even, if partial data are available. Halir and Flusser [21] provide further information about computing aspects of the estimates.

From Eq. (15) and using Eq. (10), new estimates for V_0, I_0 and φ are achieved. A simulation study presented in Section 3 shows that this approach provides more reliable estimates than the previous one.

From the estimated values, the voltage and current signals are standardized and rotated, which generates data that follows a reduced ellipse equation, as shown in Eqs. (7) and (8). Since the data are centered at the origin and symmetrically dispersed with respect to the axis, we consider the absolute values of the signals, as shown in Fig. 3. Thus, for positive values of voltage and current signals, i.e., $|x_{r(t)}|$ and $|y_{r(t)}|$, Eq. (7) can be rewritten as:

$$y_{r(t)} = \frac{b}{a}\sqrt{a^2 - x_{r(t)}^2}. \tag{16}$$

Different from the previous approach, the radii of the upper and lower limits are no longer estimated from stochastic components, i.e., the residuals. Nevertheless, our approach also assumes that both upper and lower limits are represented as ellipses. We propose to estimate separate values for the parameters a and b in Eq. (17), i.e., $a^{(p)}$ and $b^{(p)}$ are the parameters of the ellipse that leaves $100p\%$ of the points below the ellipse, $p \in \{0, 1\}$. Thus, for the upper limit $p = 0.975$, and for the lower limit $p = 0.025$. In this case, estimated upper and lower limits using quantile regression hold approximately 95% of

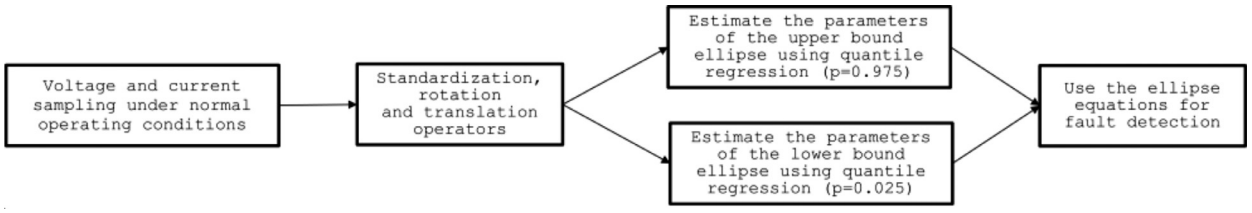


Fig. 4. Flowchart of the proposed fault detection method using ellipse equations.

the sample points under normal operating conditions. Upper and lower limits are estimated using Eq. (17).

$$\left(a^{(p)}, b^{(p)}\right) = \arg\min_{a,b} \frac{1-p}{n} \sum_{y_{r(i)}<q_i} \left|y_{r(i)} - q\left(x_{r(i)}, a, b\right)\right| + \frac{p}{n} \sum_{y_{r(i)}>q_i} \left|y_{r(i)} - q\left(x_{r(i)}, a, b\right)\right|, \tag{17}$$

where $q_i = q(x_{r(i)}, a, b) = \frac{b}{a}\sqrt{a^2 - x_{r(i)}^2}$, and p represents the pth quantile selected by the user. Thus, using non-linear quantile regression, upper and lower limit equations with 95% of statistical confidence can be written as:

$$\frac{x_{r(t)}^2}{\left[a^{(0.975)}\right]^2} + \frac{y_{r(t)}^2}{\left[b^{(0.975)}\right]^2} \leq 1, \tag{18}$$

$$\frac{x_{r(t)}^2}{\left[a^{(0.025)}\right]^2} + \frac{y_{r(t)}^2}{\left[b^{(0.025)}\right]^2} \geq 1.$$

Eq. (18) shows that upper and lower limits are estimated by means of optimization problems rather than stochastic assumptions for the noise distribution.

2.2.1. Design of the control region

In practice, assuming a regular sample period of 32 points per cycle, a single sample is collected every 0.52 ms, or 1920 samples/s. In this case, due to the large number of samples in such short time, the monitoring of the signals is online. Gomes et al. [10] suggest using additional sensitivity criteria of three consecutive points lying outside the control region, in order to detect true faults and minimize false fault detection rates. Thus, assuming a 3-sigma confidence level (99.7%) a false fault occurs with every 40,000,000 sampled points, or at every 6 h, on average. Furthermore, the confidence level and the number of points lying outside the control region can be chosen properly in order to adjust the average run length (ARL) [22,23]. Fig. 4 illustrates the proposed fault detection method using the proposed lower and upper bound ellipse equations.

2.2.2. Computational algorithm for the two-dimensional monitoring system

In addition to Fig. 4, the algorithm shown below provides detailed information of the estimates of the parameters of the upper and lower bound ellipses under normal operating conditions, and the applications of the model to detect faults.

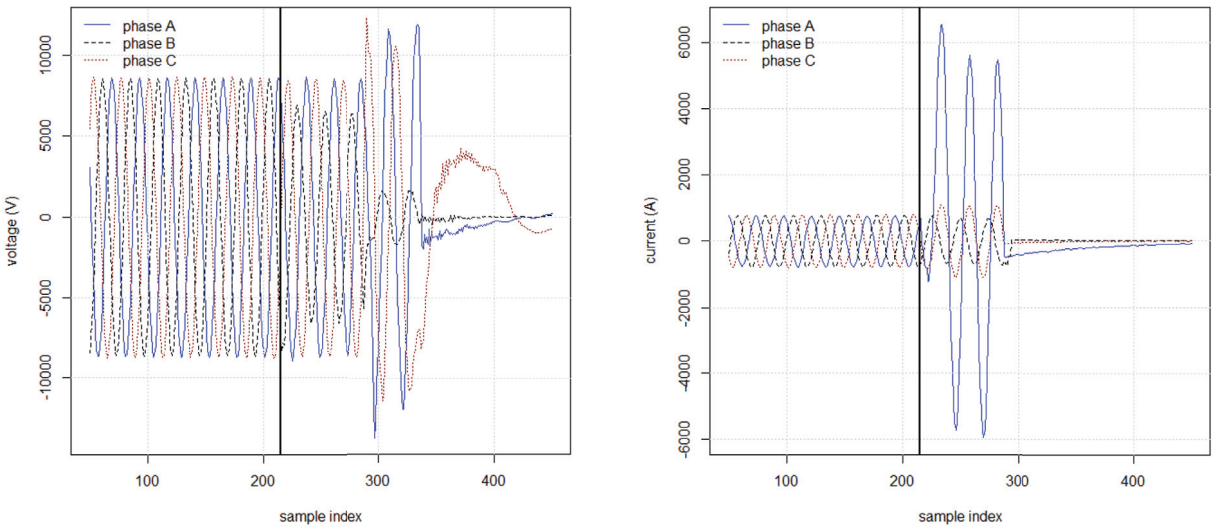
Design of the control limits

- **Step 1:** Using a sample of size **n** of voltage and current signals under normal operating conditions, solve the linear optimization problem presented in Eq. (15). Then, apply Eq. (10) to the estimated coefficients in order to estimate parameters V_0, I_0, and $\cos(\varphi)$.
- **Step 2:** Calculate the standardized voltage and current signals using the estimated voltage and current peak values, $x_{(t)} = \frac{v_{(t)}}{\hat{V}_0}$, and $y_{(t)} = \frac{i_{(t)}}{\hat{I}_0}$.
- **Step 3:** Rotate the standardized signals by an angle $\theta = \frac{\pi}{4}$ or $\theta = \frac{3\pi}{4}$.
- **Step 4:** Calculate upper and lower values for $a^{(p)}$ and $b^{(p)}$ using Eq. (17). Use $p = 0.975$ for upper limit $(a^{(0.975)}, b^{(0.975)})$ and $p = 0.025$ for lower limit $(a^{(0.025)}, b^{(0.025)})$. Thus, a confidence level of 95% is assumed.

The online monitoring system

- **Step 5:** For new values of voltage and current signals, first apply the standardized and rotation operations using the estimated parameters under normal operating conditions (Steps 1–3).
- **Step 6:** Use Eq. (18) and the estimated parameters in Step 4 to test whether the voltage and current signals are within the control limits.
- **Step 7:** If three or more consecutive points lie outside the control region then it is assumed that the system is not under normal operating conditions, i.e., there is evidence that an electric fault occurred in the transmission line.

Next, we describe a simulation study to evaluate our proposed method.



(a) voltage signals before and after the fault. (b) current signals before and after the fault.

Fig. 5. Voltage and current signals for the transmission line (phases A, B and C) after and before a lightning fault.

2.3. The Monte-Carlo simulation study

We propose the following simulation study to compare the original estimates of V_0, I_0 and φ, and the method which we are presenting. Based on a real case scenario, we set $V_0 = 770$ kV, $I_0 = 8.76$ kA, and $\varphi = \frac{2\pi}{3}$ ($\cos(\varphi) = -0.5$). Current and voltage signals were generated using a standard sampling scheme for transmission lines with 32 points per cycle, as shown in Eq. (19).

$$\begin{aligned} v_{(t)} &= V_0 \cdot \cos\left(w \cdot \tfrac{k}{32 \times T}\right) + \xi_V \cdot V_0, \\ i_{(t)} &= I_0 \cdot \cos\left(w \cdot \tfrac{k}{32 \times T} - \varphi\right) + \xi_I \cdot I_0, \end{aligned} \tag{19}$$

where $k = \{0,\ 2, \ldots, n-1\}$ and $T = \frac{1}{f}$. The sample size, n, was set as $n = 320$, or 10 cycles. Random values for ξ_V and ξ_I were generated using independent Gaussian distributions with means of zero and equal variance. We used the following variance values: $\sigma_V^2 = \sigma_I^2 = (0.005^2,\ 0.01^2,\ 0.05^2$ and $0.10^2)$. The total number of Monte Carlo simulations was set at 10,000. For each Monte Carlo simulation voltage and current signals of size $n = 320$ were generated using Eq. (19). Then, estimates for V_0, I_0 and $\cos(\varphi)$ were calculated using methods proposed by Gomes et al. [10] and Fitzgibbon et al. [20]. Finally, upper and lower limits were calculated based on Gaussian and independence assumptions of the residuals, non-Gaussian residuals and using non-linear quantile regression. The proportion of points within the control limits were calculated for each method.

2.4. The Brazilian energy distribution company data set

A real case scenario was obtained from CEMIG, a Brazilian energy distribution company. CEMIG is one of the largest energy companies in Brazil, located in the southeast region. The data represent records of voltage and current signals of a transmission line a few moments before an electrical lightning event which caused a phase-ground fault. The lightning fault happened in 2003 and the data base has 650 observations, which represent a time range of 0.45 s. The data were sampled using a sampling rate of 24 points per cycle, or one point per 0.07 ms. During the initial 215 observations, or 0.15 s, the line is under normal operating conditions. At observation number 216 the lightning fault happens. Due to the opening time of the electrical breaker, which usually starts, on average, within three sine cycles after the fault, or 0.05 s, observation numbers 217–289 represent the voltage and current signals during the fault. We used voltage and current signals from phase A.

Fig. 5(a) and (b) shows voltage and current signals for each phase before and after the occurrence of the fault. It can be seen that the current signal is more sensitive to the fault as compared to the voltage signal.

3. Results

Using the simulated data, Table 1 shows the results for the parameters, estimated using the previous approach [10] and the approach proposed by Fitzgibbon et al. [20]. For smaller values of variances, both methods achieved very similar results. The larger the variances the better are the estimates using Fitzgibbon et al. [20], as expected. In practice, under normal operating conditions the noise variances for both voltage and current signals are very small and close to 0.01^2. Under these circumstances, both methods provide similar results.

Table 1
Estimated peak values of voltage and current signals, and estimated power factor, using simulated data. Best results are shown in bold.

Variance (values)	Statistics	Gomes et al. [10]			Fitzgibbon et al. [20]		
$\sigma_V^2 = \sigma_I^2$		$\cos(\varphi)$	V_0	I_0	$\cos(\varphi)$	V_0	I_0
	True value	−0.50	770	8760	−0.50	770	8760
0.005^2	Mean	**−0.4999**	770.0521	8760.5986	**−0.4999**	**770.0016**	**8760.0239**
	SD	0.0007	0.3909	4.4885	0.0007	0.3908	4.4884
	Error (%)	0.0178	0.0068	0.0068	0.0178	**0.0002**	**0.0003**
0.01^2	Mean	**−0.4996**	770.2150	8762.4826	**−0.4996**	**770.0130**	**8760.1844**
	SD	0.0014	0.7943	9.0024	0.0014	**0.7939**	**8.9950**
	Error (%)	0.0754	0.0279	0.0283	**0.0753**	**0.0017**	**0.0021**
0.05^2	Mean	−0.4903	775.1050	8817.7664	**−0.4905**	**770.1851**	**8761.8014**
	SD	0.0069	3.9318	45.5769	**0.0068**	**3.8706**	**44.8428**
	Error (%)	1.9349	0.6630	0.6594	**1.8986**	**0.0240**	**0.0206**
0.10^2	mean	−0.4635	790.3784	8993.1207	**−0.4659**	**772.0260**	**8784.3151**
	SD	0.0142	7.9376	90.8065	**0.0135**	**7.5240**	**85.6990**
	Error (%)	7.2993	2.6465	2.6612	**6.8297**	**0.2631**	**0.2776**

Absolute error (%): 100×|(real value − estimated value)/real value|.

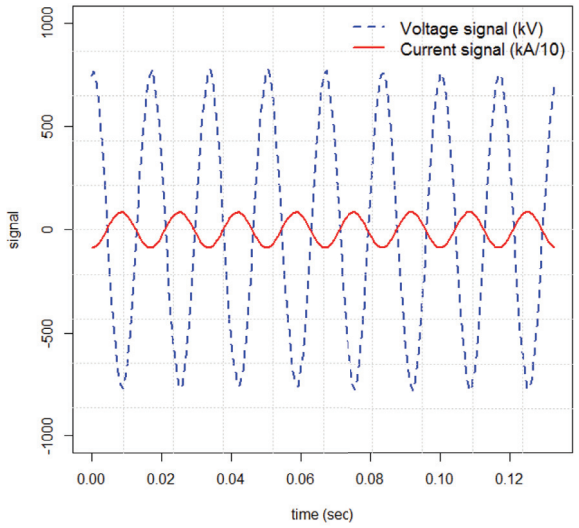
Table 2
Proportion of points within the control limits for each method in 10,000 simulations. Mean values of the proportion of points are indicated as bold.

Variance values $\sigma_V^2 = \sigma_I^2$	Statistics	Independent Gaussian assumption	non-Gaussian assumption	Non-linear quantile regression
0.005^2	Minimum	89.0625	37.8125	92.1875
	Mean	**92.5620**	**45.9676**	**94.6291**
	Maximum	95.9375	54.0625	96.2500
0.01^2	Minimum	89.0625	37.1875	91.8750
	Mean	**92.5672**	**45.9618**	**94.6194**
	Maximum	96.2500	54.0625	95.9375
0.05^2	Minimum	89.0625	36.5625	91.2500
	Mean	**92.5797**	**45.4936**	**94.6096**
	Maximum	96.2500	53.4375	95.9375
0.10^2	Minimum	89.3750	37.1875	91.8750
	Mean	**92.6403**	**44.8731**	**94.6658**
	Maximum	95.9375	52.5000	95.9375

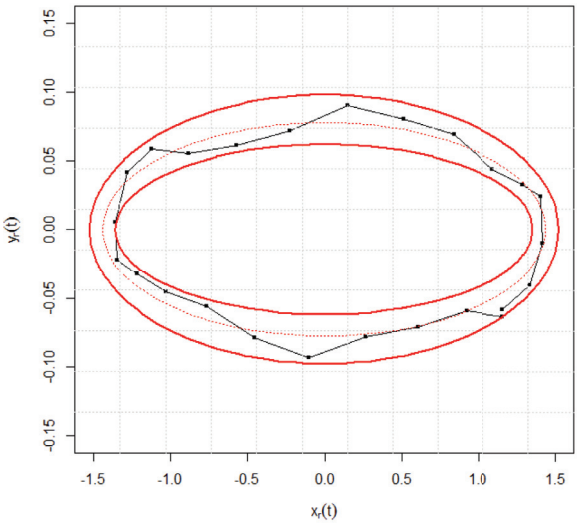
Using previous estimated values, upper and lower limits were calculated assuming Gaussian random variables and using the percentiles of the residuals (non-Gaussian approach). In addition, upper and lower limits were calculated using non-linear quantile regression. Table 2 presents the proportion of points within the control limits using the same simulated data that were applied to estimate the parameters. By doing so, it is expected to find the proportion of points close to $(1-\alpha)\%$, where α was chosen as 0.05 (5%). Results show that control limits, assuming independent Gaussian random variables hold fewer than 95% of points within the control limits. On average, 92.5% of simulated points are within the control limits, even though the simulation and the original method were based on independent Gaussian random variables. As a consequence, 7.5% of the points, on average, are outside the control limits. Using the percentiles of the residuals (non-Gaussian assumption), the proportion of points within the control limits is approximately 45%. This is the smallest value among the evaluated methods and represents almost half of the proposed confidence level (95%). Thus, if this method is applied, approximately 64% of the points will lie outside the control limits. Consequently, this method will frequently detect faults (false faults) even though the transmission line is operating under normal conditions. In practice, we do not recommend such an approach. Finally, the non-linear quantile regression approach holds 94.6% of the points within the control limits, which is much closer to 95% than previous approaches.

Using the Brazilian energy distribution company data set, the parameters of the transmission line were estimated using the first 215 observations under normal operating conditions, as shown in Fig. 6(a). Table 3 shows estimated parameters of voltage and current using the linear equation approach proposed by Gomes et al. [10] and using the penalized linear equation approach [20]. Results show differences with respect to voltage and current peak values. The parameters were estimated using the observations before the fault, i.e., under normal operating conditions. The final estimated voltage and current peak values are $\hat{V}_0 = 768.53$ V and $\hat{I}_0 = 8743.06$ A. The voltage and current standard deviations, for the standardized signals under normal operating conditions, are $s_V = 0.0074$ and $s_I = 0.0107$. It is worth mentioning that the current standard deviation is larger than the voltage standard deviation. As shown in Fig. 5(b), this is because current signals are more sensitive than voltage signals to disturbances in transmission lines, even under normal operating conditions. It is also worth mentioning that the estimated values are similar to the values used in the simulation study.

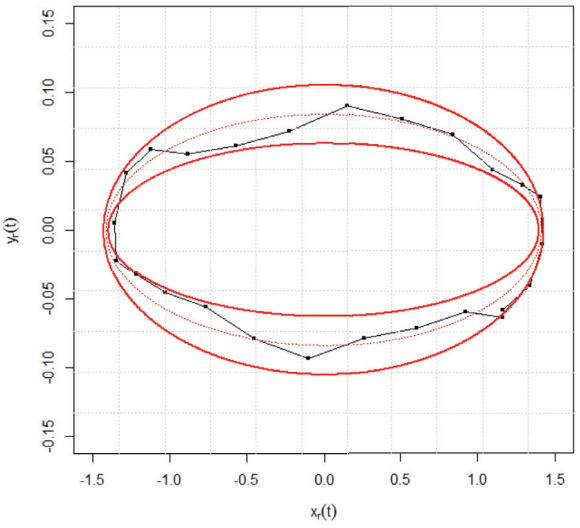
Upper and lower limits with confidence level of 99.7% were estimated using our proposed method, shown in Fig. 6(b); using the Gaussian residuals approach [24,25], shown in Fig. 6 (c); and using the method of non-Gaussian residuals [25,26],



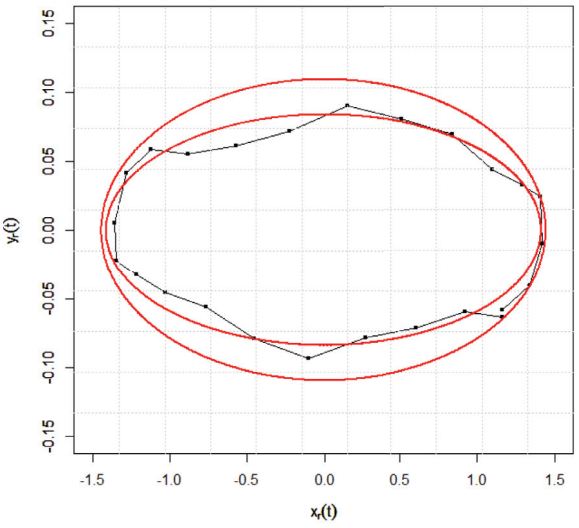
(a) Current and voltage signals under normal operating conditions.



(b) Voltage and current signals under normal operating conditions within the control limits, estimated using quantile regression.



(c) Voltage and current signals under normal operating conditions within the control limits, estimated assuming gaussian residuals [10]



(d) Voltage and current signals under normal operating conditions within the control limits, estimated using non-gaussian residuals [10].

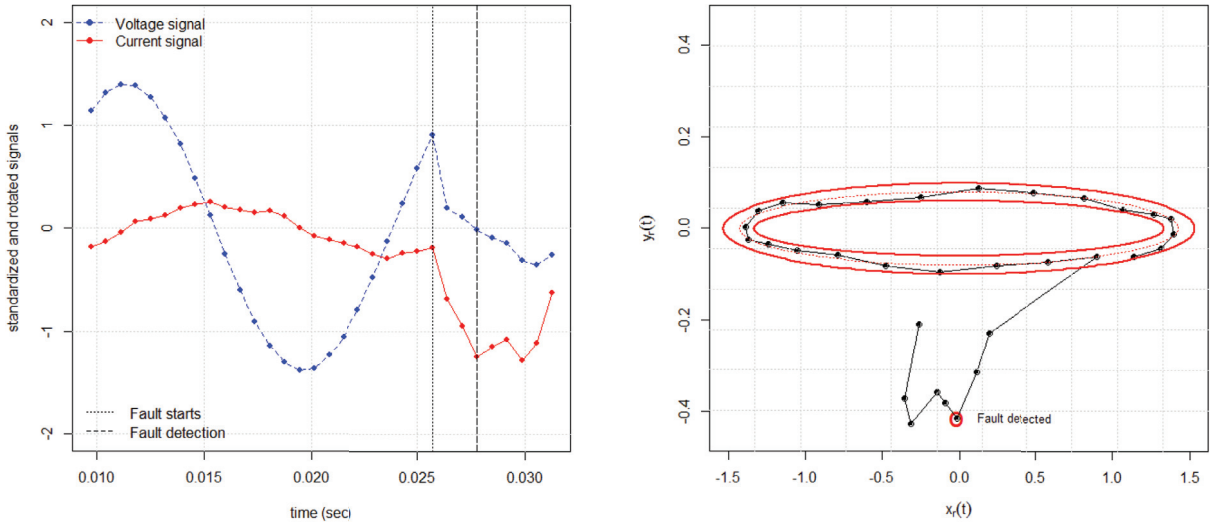
Fig. 6. Real case scenario using voltage and current signals from CEMIG. The signals under normal operating conditions are shown in (a). Control limits with 99.7% confidence level estimated using our proposal are presented in (b). The Gaussian and non-Gaussian control limit estimates are shown in (c) and (d), respectively.

shown in Fig. 6(d). As previously described in the simulation study, the methods using Gaussian and non-Gaussian residuals generate narrow control limits with false fault rates greater than the selected parameter which is 0.3%. This is illustrated in Fig. 6(c) and (d), which show observations lying outside the control region. Our proposed approach, shown in Fig. 6(b), provides a proper fit of the observations under normal operating conditions, as expected using a 99.7% confidence level.

Fig. 7(a) shows current and voltage signals before and after the lightning fault, which is represented by the dotted vertical line. Fig. 7(b) shows that the current and voltage signals leave the control region, representing an abnormal condition. If the criterion of three consecutive points lying outside the control region is applied, then the lightning fault condition is detected within 0.208 milliseconds after the fault, or in one-eighth of a cycle. This value is much smaller than standard fault detection approaches.

Table 3
Estimates of peak value parameters.

Phase	Parameters	Gomes et al [10]	Fitzgibbon et al. [20]
A	$\cos(\varphi)$	0.7376	0.7376
	V_0	8.621 kV	8.743 kV
	I_0	757.8 A	768.5 A
B	$\cos(\varphi)$	0.7377	0.7377
	V_0	8.624 kV	8.723 kV
	I_0	763.5 A	772.3 A
C	$\cos(\varphi)$	0.7384754	0.7384752
	V_0	8.520 kV	8.841 kV
	I_0	767.9 A	796.8 A



(a) Voltage and current signals before and after the lightning fault (vertical dotted line).

(b) Control region and voltage and current signals before and after the lighting fault.

Fig. 7. Voltage and current signals, and the control region, a few moments both before and after a lightning fault. Fault is detected after three consecutive points lying outside the control region.

Table 4
Proportion of points within the control limits for each method under normal operating condition (i.e. the first 215) observation, using a confidence level of 99.7%.

Phase	Independent Gaussian assumption	Non-Gaussian assumption	Non-linear quantile regression
A	84.19%	45.12%	98.14%
B	86.98%	43.26%	98.14%
C	85.12%	41.86%	98.14%

Using the data under normal operating conditions, control regions were estimated using the models proposed in [10] and the proposed non-linear quantile regression approach. For each method the proportion of points lying inside the control regions were calculated. All methods use a confidence level of 99.7%, i.e., it is expected that, theoretically, 99.7% of the points will lye inside the control region. Table 4 shows the proportion of points lying inside the control regions. The independent Gaussian assumption and non-Gaussian assumption models, proposed in [10], achieved a much smaller proportions of points within the control region as compared to the value of 99.7%. Our proposed non-linear quantile regression control region achieved the proportion of points much closer to 99.7% as compared to the previous methods.

4. Discussion and conclusion

This paper proposes a new approach to estimate the parameters of the elliptical model which is applied to detect faults in transmission lines. In addition, it successfully applies non-linear quantile regression to estimate upper and lower control limits. The simulation study and the real case scenario show that our proposal achieves more accurate values for the parameters of the model and provides false detection rates closer to the α-level chosen by the user. Therefore, it provides more reliable control regions than previous approaches.

It is worth mentioning that in energy transmission fault detection methods, voltage and current signals are successively evaluated in order to detect faults as soon as possible. Faults are deviations of the signals from normal operating conditions. Extreme deviations from normal operating conditions are detected quickly and easily. On the contrary, subtle deviations from normal operating conditions are difficult to detect because they approximate normal operating conditions. Therefore, it is crucial to represent normal operating conditions as accurately as possible, thereby establishing a proper boundary between normal and fault conditions. Furthermore, the larger the boundary the more difficult to detect faults. On the contrary, the narrower the boundary, the faster a fault is detected but the greater the chance of false faults. This work proposes new estimate procedures to create a more accurate mathematical representation of the current and voltage signals under normal operating conditions. In addition, new parametric upper and lower bounds are proposed and estimated using quantile regression. This approach does not require any statistical assumption about data uncertainties and create an accurate region under normal operating conditions. The user can change the width of the region in order to adjust the expected number of false faults. A real case scenario shows that this approach can detect faults in one eighth of a cycle, on average.

The main limitation of our proposal, and previous proposals, is the fact that each phase of the transmission line is controlled independently. In a real case scenario, faults may affect more than one phase simultaneously. Therefore, future studies aim at developing a three-dimensional control region which controls, simultaneously, the current signals for all three phases of the transmission line. Furthermore, it is of interest to classify fault conditions, i.e., to use signal information under fault conditions to identify the types of faults. Such conditions may include a fire close to the transmission line, or cable entanglement, or electrical lightning, or falling tree, or any other condition. This information is crucial for preparing maintenance teams for line repair. For instance, statistics calculated from the residuals under fault conditions can be used as input variables for fault type classification models.

Finally, the proposed approach can be easily employed in different research areas as control schemes.

Acknowledgements

The authors thank CAPES , CNPq and FAPEMIG for financial support.

References

[1] A. Rafinia, J. Moshtagh, A new approach to fault location in three-phase underground distribution system using combination of wavelet analysis with ANN and FLS, Int. J. Electr. Power Energy Syst. 55 (2014) 261–274.
[2] R.K. Aggarwal, et al., High frequency fault location technique for transmission lines based on artificial neural networks and genetic algorithms using current signals only, in: 11th International Conference on Developments in Power Systems Protection, 2012, DPSP 2012, University of Bath, 2012.
[3] N. Sundaravaradan, P. Rajaraman, D. Suzith, M.D. Kumar, G. Abilash, M.J.B. Reddy, D.K. Mohanta, Wavelet based transmission line fault analysis: a literature survey, in: 2014 14th International Conference on Environment and Electrical Engineering (EEEIC), IEEE, 2014, pp. 254–259.
[4] J.A. Morales, et al., Comparison between principal component analysis and wavelet transform 'filtering methods for lightning stroke classification on transmission lines, Electr. Power Syst. Res. 118 (2015) 37–46.
[5] S.R. Samantaray, Decision tree-based fault zone identification and fault classification in flexible AC transmissions-based transmission line, IET Gener. Transm. Distrib. 3 (5) (2009) 425–436.
[6] J.A. Morales, E. Orduna, Patterns extraction for lightning transmission lines protection based on principal component analysis, IEEE Lat. Am. Trans. (Rev. IEEE Am. Lat.) 11 (1) (2013) 518–524.
[7] A.A. Yusuff, et al., Fault location in a series compensated transmission line based on wavelet packet decomposition and support vector regression, Electr. Power Syst. Res. 81 (7) (2011) 1258–1265.
[8] J. Upendar, C.P. Gupta, G.K. Singh, Statistical decision-tree based fault classification scheme for protection of power transmission lines, Int. J. Electr. Power Energy Syst. 36 (1) (2012) 1–12.
[9] Z.A. Awan, P.M. Ramos, Improvement of phase difference estimation using modified ellipse fit method, in: 2010 IEEE Instrumentation and Measurement Technology Conference (I2MTC), IEEE, 2010.
[10] A.D.S. Gomes, A.M. Costa, T.G.A. Faria, W.M. Caminhas, Detection and classification of faults in Power transmission lines using functional analysis and computational intelligence, IEEE Trans. Power Delivery 28 (2013) 1402–1413.
[11] L. Sarıbulut, A simple power factor calculation for electrical power systems, Intl. J. Electr. Power Energy Syst. 62 (2014) 66–71.
[12] D.C. Montgomery, Introduction to Statistical Quality Control, John Wiley & Sons, 2007.
[13] H. Hotelling, The Generalization of Students Ratio, Springer, 1992.
[14] M.O.A. Abu-Shawiesh, G. Kibria, F. George, A robust bivariate control chart alternative to the Hotelling's T^2 control chart, Qual. Reliab. Eng. Int. 30 (1) (2014) 25–35.
[15] J. Han, W.-K. Kim, J.-W. Lee, C.-H. Kim, Fault type classification in transmission line using STFT, in: 11th International Conference on Developments in Power Systems Protection, 2012. DPSP 2012, IET, 2012, pp. 1–5.
[16] D. Jwad, P. Leey, Evaluation studies of combined wavelet and neural network applications in high voltage transmission line protection, in: 12th IET International Conference on Developments in Power System Protection (DPSP 2014), 2014.
[17] R.H. Chamié Filho, L.F. Carvalho, P.L. Machado, R.M. de Oliveira, Analysis of voltages induced on power outlets due to atmospheric discharges on radio base stations, Appl. Math. Modell. 37 (9) (2013) 6530–6542.
[18] G. Bassett Jr., R. Koenker, Asymptotic theory of least absolute error regression, J. Am. Stat. Assoc. 73 (363) (1978) 618–622.
[19] R. Koenker, Quantile Regression. Number 38, Cambridge University Press, 2005.
[20] A. Fitzgibbon, M. Pilu, R.B. Fisher, Direct least square fitting of ellipses, IEEE Trans. Pattern Anal. Mach. Intell. 21 (5) (1999) 476–480.
[21] R. Halir, J. Flusser, Numerically stable direct least squares fitting of ellipses, in: Proceedings of 6th International Conference in Central Europe on Computer Graphics and Visualization, WSCG, 98, Citeseer, 1998, pp. 125–132.
[22] P. Goldsmith, H. Whitfield, Average run lengths in cumulative chart quality control schemes, Technometrics 3 (1) (1961) 11–20.
[23] M.B. Khoo, V. Wong, Z. Wu, P. Castagliola, Optimal designs of the multivariate synthetic chart for monitoring the process mean vector based on median run length, Qual. Reliab. Eng. Int. 27 (8) (2011) 981–997.
[24] G. Casella, R.L. Berger, Statistical Inference, 2, Duxbury Pacific Grove, CA, 2002.
[25] M. Fisz. Probability theory and mathematical statistics. Technical report, 1963.
[26] H.A. David, H.N. Nagaraja, Order Statistics, Wiley Online Library, 1970.