Data Regression Strategy to Model Transmission Line Faults on Vegetation

F. V. Lopes, A. M. Britto, E. R. Passos, R. L. A. Reis, K. M. Silva, and K. M. Dantas

***Abstract*—This paper presents a study on data regression strategies which can model real-world cases of transmission line high impedance faults (HIFs) on vegetation. A reference HIF model implemented in the Alternative Transients Program (ATP)/ATPDraw is used as a supporting computational tool to evaluate different data regression formulas, which consider different amounts of parameters to represent the most prominent features of HIF resistance variations caused by the vegetation charring phenomenon. Real cases of HIFs on vegetation that took place on a 500 kV/60 Hz transmission line are investigated, through which a novel promising rational three-parameter regression function is identified. The obtained results reveal that such a novel regression formula further improves the accuracy of other models previously reported in the literature, without requiring complex setting procedures. For instance, comparing it against another existing three-parameter HIF model, the proposed approach reduced the root mean square errors from 17.55 Ω to 8.20 Ω in average, also improving the coincidence rate (measured via R-square coefficient) between real and simulated signals, which increased from 0.883 to 0.975 in average.**

***Keywords*—ATP/ATPDraw, EMTP, high impedance faults, transmission line, power systems, vegetation.**

NOMENCLATURE

ATP	Alternative Transients Program.
EMT	Electromagnetic Transients.
EMTP	Electromagnetic Transients Program.
LOC	Local line terminal.
REM	Remote line terminal.
RMSE	Root Mean Square Error.
HIF	High Impedance Fault.
TACS	Transient Analysis of Control System.
R^2	R-square coefficient.
R_f	Fault resistance.
$R_f(t_x)$	Estimated R_f at instant t_x.
$R_{f,ini}$	Initial high impedance fault resistance value.
$R_{f,end}$	Final high impedance fault resistance value.
R_f^{3p}	HIF R_f value estimated via three-parameter formula (Parameters: $a,\ b$ and τ_1).
R_f^{4p}	HIF R_f value estimated via four-parameter formula (Parameters: $c,\ d,\ \tau_2$ and τ_3).
R_f^{5p}	HIF R_f value estimated via five-parameter formula (Parameters: $f,\ g,\ h,\ \tau_4$ and τ_5).
$R_f^{Rat,2p}$	HIF R_f value estimated via two-parameter rational function (Parameters: r_1 and r_2).
$R_f^{Rat,3p}$	HIF R_f value estimated via three-parameter rational function (Parameters: $s_1,\ s_2$ and s_3).
$R_f^{Rat,5p}$	HIF R_f value estimated via five-parameter rational function (Parameters: $v_1,\ v_2,\ v_3,\ v_4$ and v_5).
t	Time vector.
t_x	Time stamp of the x-th sample.

Felipe V. Lopes and Raphael L. A. Reis are with Federal University of Paraíba (UFPB), Department of Electrical Engineering, João Pessoa-PB, Brazil (e-mails: {felipelopes,raphael.leite}@cear.ufpb.br).

Amauri G. Martins-Britto is with KU Leuven, division Electa & the Etch Competence Hub of EnergyVille, Genk, Belgium (e-mail: amauri.martinsbritto@kuleuven.be).

Eduardo R. Passos and Kleber M. Silva are with University of Brasília (UnB), Department of Electrical Engineering, Brasília-DF, Brazil (e-mails: eduardopassos@lapse.unb.br, klebermelo@unb.br)

Karcius M. Dantas is with Federal University of Campina Grande (UFCG), Department of Electrical Engineering, Campina Grande-PB, Brazil (e-mail: karcius@dee.ufcg.edu.br)

Paper submitted to the International Conference on Power Systems Transients (IPST2025) in Guadalajara, Mexico, June 8-12, 2025.

I. INTRODUCTION

HIGH impedance faults (HIF) on vegetation (such as those on trees) often occur on transmission lines. After the HIF inception, induced arcs lead trees to burn, resulting in varying fault resistances R_f due to the vegetation charring phenomenon. Initially, R_f has a high value, and it gradually decreases as the charred areas increase. Hence, HIFs challenge transmission line monitoring schemes, and it has boosted researches on strategies to model HIFs in Electromagnetic Transients Programs (EMTP) toward supporting studies on transmission line faults that occur on vegetation.

The literature is plenty of works reporting HIF models, being most of them developed in the context of distribution networks. In [1] and [2], experimental measurements are used to fit HIF models for medium voltage networks. Both works divide the fault resistance R_f into two parts. In [2], for instance, the Alternative Transients Program (ATP) is used to simulate HIFs on distribution networks by using two time-varying resistances that emulate HIF signatures for different soil types. One of these resistances varies as proposed in [3], and the other emulates a polynomial formula obtained from regression procedures applied to field R_f measurements.

In [4], the HIF arc resistance dynamic is reproduced via MODELS and Transient Analysis of Control System (TACS) elements in ATP environment. Also, in [5], a Matlab/Simulink model which emulates the HIF R_f behavior as proposed in [6] is presented. Nevertheless, although these HIF models show to be realistic, the great number of used parameters, the lack of detailed and didactic implementation guidelines and the need for experimental data are considered limiting factors.

Still analyzing the literature, one can find research initiatives which have explored other methodologies than those already mentioned, either to improve the HIF model fitting with real HIF data or to detect/locate HIFs. In [7], for example, an automated parameter tuning approach is presented. In [8],

arc-based fault modeling for simulations in ATP environment with nonlinear elements is described. In [9], a polynomial and neural network-driven HIF distance estimation is developed. In [10], HIF-induced high-frequency components are analyzed and a strategy to characterize fault-induced transients is proposed. Also, in [11], a HIF detection method that combines variational mode decomposition algorithms with nonlinear least squares estimation is developed.

From the above-mentioned references, one notices that, while machine-learning-based HIF detection frameworks have been widely addressed in the literature, the primary objectives of most works remain dissociated from the improvement of HIF models. Furthermore, although promising HIF modeling approaches have been reported, significant challenges are identified, particularly due to the need for non-intuitive settings and high computational costs that result mainly from the need for large datasets with physically representative features. Indeed, such features are critical in physics-driven machine learning approaches, where reliance on synthetic data can exacerbate biases and over-fitting problems. Hence, classical regression-based HIF models have regained attention, mainly due to their simplicity and computational efficiency for Electromagnetic Transients (EMT)-type simulations. Besides their straightforward implementation and reduced parameter dependence, these models open possibilities for intuitive setting procedures, allowing to include them in EMTPs, which in turn also permit to emulate noise, circuit sectioning, fault evolution, and so on. This balance between simplicity, accuracy and adaptability makes data regression-based HIF models attractive for EMTPs, specially when high computational overhead is a limiting factor for the EMT simulations of interest.

As mentioned earlier, most works in the literature which report HIF modeling strategies focus on distribution networks. On the other hand, contributions addressing transmission line HIFs have been also presented, but they are scarcer. In [12], transmission line HIF signatures are analytically analyzed, but computational implementation instructions are not reported. In [13], a five-parameter HIF model is proposed. These parameters are obtained from regression procedures based on the Gauss-Newton approach, considering a formula that combines two exponential components and a sine term to emulate R_f decaying and oscillations. Such a HIF model shows to be accurate, but its computational implementation is not explained. Also, some non-intuitive parameters are required, which cannot be directly related to R_f characteristic values, limiting the model reproduction when experimental data is not available. Moreover, in [14], a simplified transmission line HIF model is developed in ATP/ATPDraw. It approximates R_f by a three-parameter exponential function, requiring only intuitive settings, such as initial and final R_f values, as well as R_f decaying time constant. Still in [14], details on the HIF model implementation in ATP/ATPDraw are provided, which consists in an important contribution of such a work. However, as a simplified regression formula is applied, depending on the case, its accuracy can decrease, either during the initial R_f decaying or over the final HIF instants, particularly when the fault takes longer to be cleared.

Based on the presented context, the main goal of this work is to enhance the modeling strategy often used to emulate transmission line HIFs on vegetation in EMT-type simulations. The model must combine accuracy and simplicity, but maintaining its flexibility to represent both real-world HIF cases, when experimental data is available, or fictitious HIF scenarios, just by varying intuitive settings. Hence, this work is fully focused on modeling aspects of transmission line HIFs on vegetation toward obtaining a realistic and simple model, which can support studies on HIF diagnosis methods, whose evaluation is in turn not included in the scope of this paper.

To allow the proposed modeling studies, six regression formulas are analyzed, consisting of three exponential functions and three rational formulas. The R-square coefficients (R^2), as well as the absolute and root mean square errors (RMSE) are evaluated to select four promising regression formulas capable of emulating two representative real HIF cases. ATP/ATPDraw simulations are carried out to assess such modeling approaches, allowing to select the model with best cost-benefit as the proposed modeling strategy.

The analyzed real-world cases of HIFs on vegetation occurred in a 500 kV/60 Hz transmission system, being reproduced via ATP/ATPDraw simulations. The HIF model proposed in [14] is used as reference, being adapted only to reproduce the tested regression formulas. The obtained results reveal that a three-parameter rational function can be used to accurately emulate transmission line HIFs on vegetation. As it requires only three parameters, a direct relationship between the rational function coefficients and characteristic HIF parameters can be obtained. Based on that, a practical application procedure is also developed.

II. Transmission Line HIFs on Vegetation

Transmission line HIFs on vegetation frequently take place on lines with rights-of-way that pass through dense forests. Modeling these events is often demanded during fault studies, and it requires models more sophisticated than single purely resistive and linear elements. Hence, practical and easy-to-use transmission line HIF models have been of a particular interest for professionals in a number of areas, such as those focused on EMT, power system relaying, and disturbance diagnosis.

Line HIFs on vegetation occur due to tree-to-line contacts, which are more likely to take place in hot months, such as the late summer. As the temperatures are higher, power grid costumers use more air conditioning equipment, increasing the system loading. It heats up transmission line conductors, causing their expansion and increasing line sags over the tower spans. Concomitantly, trees grow mainly during the spring and summer months, when, besides the higher temperatures, stronger airflow across the lines are verified [15].

When tree-to-line contacts occur, arcs are established, leading the tree to burn [16], [17]. As the charred areas along the tree increase, the vegetation becomes more conductive [16], [17], resulting in R_f values which gradually decrease, as shown in Fig. 1. This figure shows R_f estimations obtained by means of the methodology reported in [18] for two real-world transmission line HIFs on vegetation, which are referred here as Case 1 and Case 2, for the sake of simplicity.

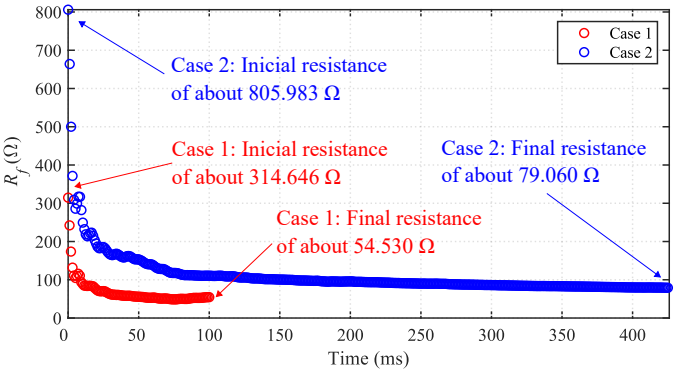

Fig. 1. R_f estimations obtained from the real-world HIF cases 1 and 2.

Both cases depicted in Fig. 1 are related to HIFs on vegetation that took place on a 500 kV/60 Hz double-circuit series compensated transmission line, 333.5 km long, installed in the North region of Brazil. Due to confidentiality reasons, details on terminal substations are omitted, such that they are called here local (LOC) and remote (REM) substations. Surrounding the faulted line, there are other double-circuit series compensated lines, which are fed by strong and weak equivalent sources at each system side. The rights-of-way of these lines pass through the Amazon Rainforest, where very tall trees exist [14]. Case 1 consists of a phase A-to-ground HIF at 18.72 km from LOC substation, which was cleared by the line protection about 100 ms after the fault inception. On the other hand, Case 2 stands for a phase C-to-ground HIF 9.66 km far from REM substation, which was cleared in about 425 ms after the HIF inception. In this paper, measurements from the substations closer to the HIF point are considered, i.e., records from LOC and REM substations are used to obtain R_f in cases 1 and 2, respectively.

According to both scenarios shown in Fig. 1, R_f presents high initial values $R_{f,ini}$, which gradually decrease until a final fault resistances $R_{f,end}$. In Case 1, although the fault period lasts about 100 ms, R_f reduces from $R_{f,ini} \approx 315\ \Omega$ to $R_{f,end} \approx 55\ \Omega$ in approximately 50 ms, remaining in the system for additional 50 ms, without relevant R_f variations. In Case 2, the R_f decreasing rate is smaller than in Case 1, resulting in a slower transition from $R_{f,ini}$ to $R_{f,end}$. In this scenario, R_f takes about 200 ms to vary from $R_{f,ini} \approx 806\ \Omega$ until values smaller than 100 Ω. After that, R_f slowly decreases for 225 ms, resulting in $R_{f,end} \approx 79\ \Omega$.

Although R_f tends to converge to well-defined final values, in some scenarios, like in Case 2, R_f does not completely stabilize in the final instants of the fault. Indeed, R_f remains decreasing in a very slow rate, which is different from the rate observed soon after the HIF inception instant. Based on such features, one concludes that initial and final R_f decreasing rates must be considered in the regression formulas used to reproduce the R_f behavior. Thus, evaluating different regression strategies is necessary to identify those that better emulate R_f over the whole fault period. To do so, the ATP/ATPDraw HIF implementation strategy proposed in [14] is adopted, since it allows to simulate different R_f functions just by adapting very few code lines.

III. Analyzed R_f Regression Formulas

Before defining the regression formulas to be analyzed, it is worth understanding that there are differences between the fault-induced transient signatures typically verified in distribution HIFs and those observed during faults caused by vegetation contact in transmission lines. Distribution HIFs usually result in more complex transient signatures in currents waveforms, resulting in asymmetry, modulations, intermittence, buildup and shoulders [1]–[6]. On the other hand, from real records and analyzing waveforms shown in [12]–[14], one notices that the R_f decaying behavior (buildup) is the most prominent transient signature for transmission line HIFs on vegetation. Thus, while HIF models for distribution grids have been developed to emulate various transient signatures for different soil types, such as sand, asphalt, gravel, cobblestones, grass and so on [2], transmission line HIFs on vegetation can be emulated in a simpler way, allowing to simplify the used modeling strategies toward making transmission line HIF models more attractive for EMTP users.

In [14], real-world records are assessed to demonstrate that transmission line HIFs on vegetation can be properly emulated by considering a single non-linear resistance, provided that representative regression formulas are used to represent R_f variations over the time t. Therefore, in this paper, six regression formulas are assessed with the primary objective of representing the R_f decaying feature, disregarding other smoothed signatures which are less prominent in transmission networks. By doing so, complex simulation structures are avoided, resulting in a simple and accurate model, as proven later on in Section IV. Even so, the proposed modeling strategy is recommended only for studies on transmission line HIFs on vegetation, such that more complex and detailed models are recommended if distribution HIFs are under investigation.

The presented case studies focus on R_f emulation in the real cases 1 and 2 shown in Fig. 1. Pros and cons of each evaluated regression strategy are revealed, proving limitations of less representative expressions. Among the assessed regression formulas, two are chosen based on literature evidences about R_f behavior due to the vegetation charring phenomenon. In [13] and [14], good model fitting results are obtained by approximating R_f by three- and five-parameter exponential functions, respectively, justifying their evaluation. Also, given the good fitting obtained in [14] by using a three-parameter formula, a four-parameter expression is also tested to assess the benefits of adding one more exponential function in the R_f representation. Furthermore, three rational regression formulas are studied, consisting of a two-parameter proper rational function and two improper rational functions with three and five parameters. These formulas are chosen due to their capacity of representing decaying behaviors over the time, as expected for R_f.

A. Three-Parameter Exponential Formula (R_f^{3p})

The evaluated three-parameter exponential regression formula is the one proposed in [14], which is defined as:

$$R_f^{3p} = a + b \cdot e^{(-\tau_1 \cdot t)}\ , \tag{1}$$

where R_f^{3p} represents the R_f estimations calculated via (1), being $R_{f,ini} = a + b$, $R_{f,end} = a$, and τ_1 the R_f decaying time constant.

B. Four-Parameter Exponential Formula (R_f^{4p})

To improve the R_f estimations reported in [14], a four-parameter exponential regression formula was designed. Analyzing the curves shown in Fig. 1, it can be noticed that the decaying rate of R_f may be difficult to be reproduced with only one time constant. Moreover, such a decaying rate seems to change over the time. Hence, by observation of the R_f behavior, the following regression formula was designed:

$$R_f^{4p} = c - d \cdot e^{\left[(-\tau_2 \cdot t) + e^{(-\tau_3 \cdot t)}\right]} \ , \tag{2}$$

being R_f^{4p} the R_f estimations calculated using (2), where $R_{f,ini} = c - d \cdot e^{(1)}$ and $R_{f,end} = c$. Also, τ_2 and τ_3 are time constants which dictate the R_f decaying over the time.

C. Five-Parameter Exponential Formula (R_f^{5p})

The assessed five-parameter exponential formula is inspired (but not equal to) in the one reported in [13], where two independent exponential terms are used in the regression procedure. Such an adaptation is made to better fit the regression terms to the R_f decaying, as follows:

$$R_f^{5p} = f + g \cdot e^{(-\tau_4 \cdot t)} + h \cdot e^{(-\tau_5 \cdot t)} \ . \tag{3}$$

The R_f estimations computed by means of (3) are represented by R_f^{5p}, being $R_{f,ini} = f + g + h$ and $R_{f,end} = f$. Moreover, τ_4 and τ_5 are time constants that emulate the R_f decaying over the time t, such as considered in the other exponential regression formulas studied in this paper.

D. Two-Parameter Rational Function ($R_f^{Rat,2p}$)

Aiming to evaluate a regression formula simpler than those previously described, a two-parameter proper rational function is analyzed, being given by:

$$R_f^{Rat,2p} = \frac{r_1}{t + r_2} \ , \tag{4}$$

where $R_f^{Rat,2p}$ stands for the R_f estimations calculated via (4). $R_{f,ini}$ and $R_{f,end}$ can be obtained by making $t = 0$ s and $t \to \infty$, respectively, as done for the exponential formulas, resulting in $R_{f,ini} = \frac{r_1}{r_2}$ and $R_{f,end} = 0\ \Omega$. Such results reveal that $R_f^{Rat,2p}$ fails in representing final R_f values different from zero, as it will be demonstrated later on in Section IV.

E. Three-Parameter Rational Function ($R_f^{Rat,3p}$)

Such a rational function has the same number of parameters used in the HIF model proposed in [14]. It is obtained by adding one degree in the numerator of (4), leading it to be an improper rational function (same degree in numerator and denominator). As it depends on three parameters, a very similar computational complexity to the one verified in [14] is expected, being the regression formula given by:

$$R_f^{Rat,3p} = \frac{s_1 \cdot t + s_2}{t + s_3} \ , \tag{5}$$

where $R_f^{Rat,3p}$ consists in the R_f estimations obtained from (5), where $R_{f,ini} = \frac{s_2}{s_3}$ and $R_{f,end} = s_1$.

F. Five-Parameter Rational Function ($R_f^{Rat,5p}$)

A five-parameter rational function is also evaluated toward verifying the effects of adding degrees in the numerator and denominator of the rational function. Such a number of parameters is chosen because it allows to evaluate the effects of increasing the number of function zeros and poles, but maintaining a closed mathematical relation between $R_{f,ini}$ and $R_{f,end}$, and the function coefficients. In this sense, the evaluated five-parameter rational formula was defined to be:

$$R_f^{Rat,5p} = \frac{v_1 \cdot t^2 + v_2 \cdot t + v_3}{t^2 + v_4 \cdot t + v_5} \ , \tag{6}$$

being $R_f^{Rat,5p}$ the estimations obtained via (6), in which $R_{f,ini} = \frac{v_3}{v_5}$ and $R_{f,end} = v_1$. As demonstrated, by using the same degree in both numerator and denominator, rational functions with higher orders can be applied, without compromising the direct relation between $R_{f,ini}$ and $R_{f,end}$, and the formula coefficients. However, some intermediate parameters become non-intuitive (such as v_2 and v_4), which is considered a limiting factor.

IV. Evaluation of HIF R_f Regression Formulas

The presented regression formulas are evaluated in two parts. Initially, the real R_f samples shown in Fig. 1 are considered, and each regression formula is fitted to the R_f data by means of Matlab functions based on the trust-region approach [19]. The absolute errors, RMSE and R^2 coefficients are computed in each case to evaluate the fitting quality of each regression formula. Then, EMT-type simulations are carried out in ATP/ATPDraw, considering the most promising regression models, and comparing simulated waveforms with real ones. To do so, real-world transmission line HIF records and validated power network models were required to guarantee reliable evaluations. Hence, since the authors had a limited access to real records and validated power system models, massive case studies are not presented in this paper. Even so, it does not compromise the reliability of the presented results, given that the analyzed cases 1 and 2 are representative for studies on transmission line HIFs on vegetation.

In the proposed studies, the HIF ATP/ATPDraw model described in [14] is used with few adaptations to reproduce the assessed regression formulas. By doing so, a fair comparison between all modeling strategies is provided, allowing to analyze the cost-benefit of each regression strategy in terms of accuracy, simplicity and practicality. For these studies, a validated ATP/ATPDraw power system model used by the Brazilian transmission system operator was taken into account, allowing to realistically compare real and simulated HIF records. However, due to space limitations, further details on the ATP/ATPDraw system validation studies are not detailed.

A. Regression Formula Fitting Quality Evaluation

Obtained coefficients for the regression formulas used to calculate R_f^{3p}, R_f^{4p}, R_f^{5p}, $R_f^{Rat,2p}$, $R_f^{Rat,3p}$ and $R_f^{Rat,5p}$ are shown in Table I, and the R_f behaviors over the time for cases 1 and 2 are shown in Figs. 2 and 3, respectively, where the absolute errors over the time between estimated and real R_f measurements are also presented.

TABLE I
TRANSMISSION LINE HIF MODEL COEFFICIENTS.

R_f Model	Case	Coefficients
R_f^{3p}	1	$a = 57.73,\ b = 229.5,\ \tau_1 = 247.4$ s
	2	$a = 97.22,\ b = 494.5,\ \tau_1 = 89.03$ s
R_f^{4p}	1	$c = 52.01,\ d = -96.58$ $\tau_2 = 77.26$ s, $\tau_3 = 483$ s
	2	$c = 89.86,\ d = -248.1$ $\tau_2 = 32.29$ s, $\tau_3 = 335.3$ s
R_f^{5p}	1	$f = 50.3,\ g = 72.62,\ h = 197.8,$ $\tau_4 = 54.47$ s, $\tau_5 = 611.6$ s
	2	$f = 95.37,\ g = 1587,\ h = -1223,$ $\tau_4 = 55.77$ s, $\tau_5 = 55.72$ s
$R_f^{Rat,2p}$	1	$r_1 = 3.318,\ r_2 = 0.01734$
	2	$r_1 = 23.41,\ r_2 = 0.0758$
$R_f^{Rat,3p}$	1	$s_1 = 46.11,\ s_2 = 0.6105,\ s_3 = 0.001927$
	2	$s_1 = 82.14,\ s_2 = 2.772,\ s_3 = 0.003571$
$R_f^{Rat,5p}$	1	$v_1 = 1.91\text{E}5,\ v_2 = 2.703\text{E}6,\ v_3 = 3.61\text{E}4$ $v_4 = 5.891\text{E}4,\ v_5 = 113.9$
	2	$v_1 = 65,\ v_2 = 11.43,\ v_3 = 0.1048$ $v_4 = 0.06534,\ v_5 = 0.0001266$

To assess the obtained results, in addition Figs. 2 and 3, Table II presents the RMSE and R^2 coefficients for each regression strategy. Hence, by analyzing Figs. 2 and 3, and Table II, conclusions on the best regression strategies among the evaluated ones can be drawn as follows:

- $R_f^{Rat,2p}$ presents relevant errors in the fault beginning and ending moments, which tend to increase as the time goes by. On the other hand, the remaining regression formulas present satisfactory R_f fitting results.
- $R_f^{Rat,2p}$ resulted in errors that reached RMSEs of about 21.73 Ω and 41.36 Ω for cases 1 and 2, respectively, with R^2 coefficients smaller than 0.7 in both scenarios. Thus, due to these high discrepancies, $R_f^{Rat,2p}$ is disregarded in the next studies.
- An overall good HIF emulation is obtained via R_f^{3p}, but it results in estimations that decay faster than the real R_f, so that errors in the first HIF instants are higher than those obtained for the other evaluated regression formulas.
- R_f^{3p} resulted in R^2 coefficients equal to 0.917 and 0.849 for cases 1 and 2, respectively, revealing an overall acceptable fitting. However, the RMSEs in these cases reached the order of 10.85 Ω and 24.42 Ω, respectively, being greater than those obtained from the other regression formulas. Even so, as R_f^{3p} is the simplest approach among the evaluated exponential regression formulas, it will be considered in the next studies for the sake of illustration.
- R_f^{4p}, R_f^{5p}, $R_f^{Rat,3p}$ and $R_f^{Rat,5p}$ presented very good results for Case 1, with RMSEs smaller than 6.0 Ω and R^2 coefficients above 0.98. Nevertheless, in Case 2, R_f^{5p} resulted in an RMSE equal to 26.14 Ω, which is much higher than those obtained from R_f^{4p}, $R_f^{Rat,3p}$ and $R_f^{Rat,5p}$. Thereby, R_f^{5p} is disregarded in the following evaluations.

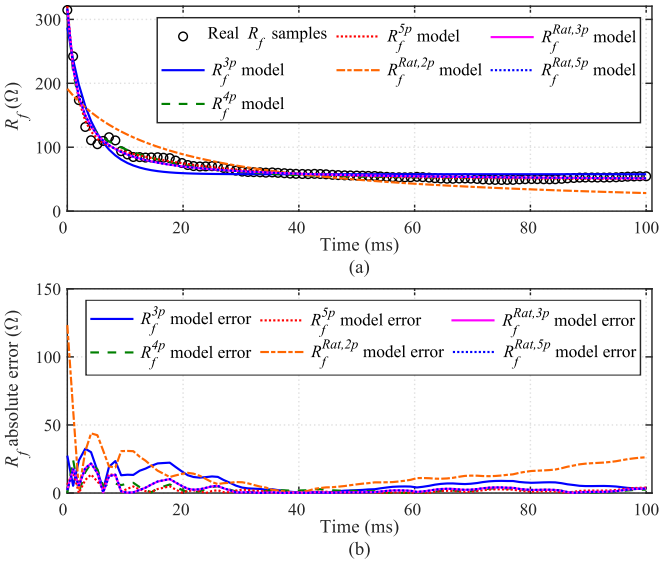

Fig. 2. Case 1 analysis: (a) HIF R_f estimations using the studied regression formulas; (b) Absolute error of estimated R_f values.

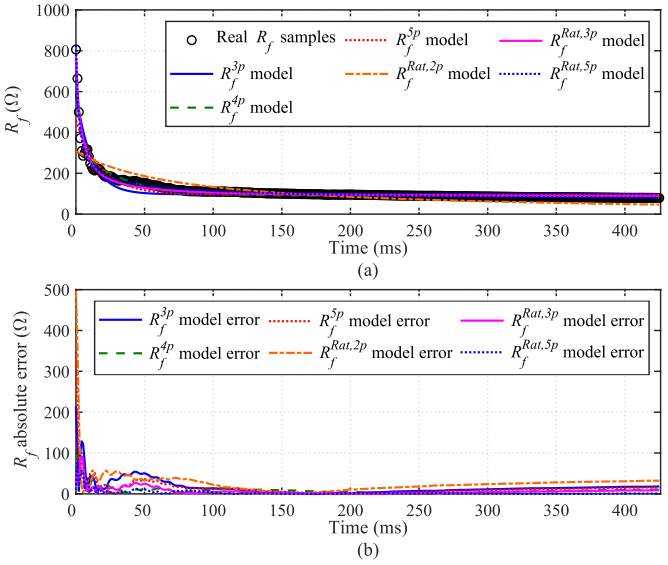

Fig. 3. Case 2 analysis: (a) HIF R_f estimations using the studied regression formulas; (b) Absolute error of estimated R_f values.

- R_f^{3p} and R_f^{4p} were slightly more accurate than $R_f^{Rat,3p}$ and $R_f^{Rat,5p}$ in Case 1, but less accurate for Case 2. Given that, the use of rational functions shows to be promising.
- Analyzing $R_f^{Rat,5p}$, it is noticed that the obtained fitting quality is very similar to that one of $R_f^{Rat,3p}$, with a slight superiority of $R_f^{Rat,5p}$. On the other hand, $R_f^{Rat,3p}$ is simpler than $R_f^{Rat,5p}$, which is interesting for practical applications. Therefore, both approaches are also considered in the next evaluation step, in which R_f^{3p}, R_f^{4p}, $R_f^{Rat,3p}$ and $R_f^{Rat,5p}$ are taken into account.

B. Comparative Analysis via ATP/ATPDraw Tests

In the second evaluation part, R_f^{3p}, R_f^{4p}, $R_f^{Rat,3p}$ and $R_f^{Rat,5p}$ are simulated in ATP/ATPDraw by adapting the HIF model proposed in [14]. For the sake of illustration, the HIF

TABLE II
RMSE AND R^2 COEFFICIENTS FOR THE EVALUATED CASES.

R_f	Case 1		Case 2	
Model	RMSE (Ω)	R^2	RMSE (Ω)	R^2
R_f^{3p}	10.85	0.917	24.42	0.849
R_f^{4p}	4.97	0.983	11.54	0.966
R_f^{5p}	3.76	0.990	26.14	0.828
$R_f^{Rat,2p}$	21.73	0.663	41.36	0.566
$R_f^{Rat,3p}$	5.14	0.981	11.26	0.968
$R_f^{Rat,5p}$	5.21	0.981	7.33	0.986

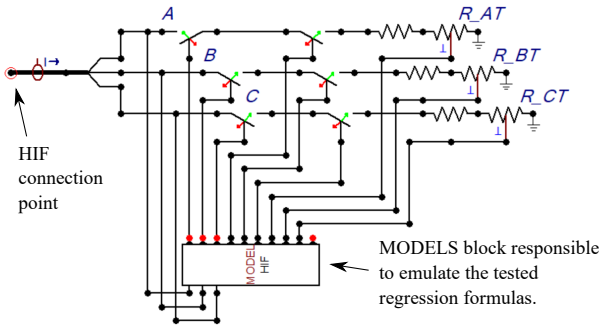

Fig. 4. HIF model implementation in ATP/ATPDraw, as proposed in [14].

model is shown in Fig. 4. Also, the same power system studied in [14] is considered here, such that further details on the HIF model and tested power system available in [14].

Fig. 5 presents the required adaptations in the MODELS code detailed in [14] (see page 6, Fig. 7 in [14]), particularly in the code part on 'HIF resistance configuration' toward simulating R_f^{3p}, R_f^{4p}, $R_f^{Rat,3p}$ and $R_f^{Rat,5p}$. To simplify the implementation, for R_f^{3p} and R_f^{4p}, the HIF model was programmed to receive as inputs $R_{f,ini}$, $R_{f,end}$ and their respective time constants. In R_f^{3p}, the coefficients a and b were directly calculated from $R_{f,ini}$ and $R_{f,end}$ (see Fig. 5(a)), whereas in R_f^{4p}, auxiliary variables `alfa` and `beta` were created to relate the coefficients c and d with $R_{f,ini}$ and $R_{f,end}$ (see Fig. 5(b)). On the other hand, for $R_f^{Rat,3p}$ and $R_f^{Rat,5p}$, the ATP/ATPDraw HIF model was programmed to directly receive as inputs the coefficients of each rational function, whose nomenclature was standardized in the MODELS code to be as follows: in $R_f^{Rat,3p}$, p_1, p_2 and q_1 correspond to s_1, s_2 and s_3, respectively (see Fig. 5(c)), and in $R_f^{Rat,5p}$, p_1, p_2, p_3, q_1 and q_2 stand for v_1, v_2, v_3, v_4 and v_5, respectively (see Fig. 5(d)). The initialization of $R_{f,ini}$ in each model was programmed in accordance to the relations presented in the previous section, being the auxiliary variables declared in MODELS language.

```
-- HIF resistance configuration
IF auxRT=1 THEN
 IF swFa=1 THEN RTA:=RFfinal+(RFinitial-RFfinal)*exp(-tau*(t-ftime)) ENDIF
 IF swFb=1 THEN RTB:=RFfinal+(RFinitial-RFfinal)*exp(-tau*(t-ftime)) ENDIF
 IF swFc=1 THEN RTC:=RFfinal+(RFinitial-RFfinal)*exp(-tau*(t-ftime)) ENDIF
ENDIF
```

(a)

```
-- HIF resistance configuration
IF auxRT=1 THEN
 tx:=t-ftime, alfa:=-RFend, beta:=(RFini+alfa)/(-exp(1.0))
 IF swFa=1 THEN RTA:=-alfa-beta*exp(-taul*tx + exp(-tau2*tx)) ENDIF
 IF swFb=1 THEN RTB:=-alfa-beta*exp(-taul*tx + exp(-tau2*tx)) ENDIF
 IF swFc=1 THEN RTC:=-alfa-beta*exp(-taul*tx + exp(-tau2*tx)) ENDIF
ENDIF
```

(b)

```
-- HIF resistance configuration
IF auxRT=1 THEN
 tx:=t-ftime
 IF swFa=1 THEN RTA:=(p1*tx + p2)/(tx + q1) ENDIF
 IF swFb=1 THEN RTB:=(p1*tx + p2)/(tx + q1) ENDIF
 IF swFc=1 THEN RTC:=(p1*tx + p2)/(tx + q1) ENDIF
ENDIF
```

(c)

```
-- HIF resistance configuration
IF auxRT=1 THEN
 tx:=t-ftime
 IF swFa=1 THEN RTA:=(p1*(tx**2) + p2*(tx) + p3)/((tx**2) + q1*(tx) + q2) ENDIF
 IF swFb=1 THEN RTB:=(p1*(tx**2) + p2*(tx) + p3)/((tx**2) + q1*(tx) + q2) ENDIF
 IF swFc=1 THEN RTC:=(p1*(tx**2) + p2*(tx) + p3)/((tx**2) + q1*(tx) + q2) ENDIF
ENDIF
```

(d)

Fig. 5. Adaptations in ATP-MODELS code presented in [14] (see page 6, Fig. 7 in [14]) to simulate: (a) R_f^{3p}; (b) R_f^{4p}; (c) $R_f^{Rat,3p}$; and (d) $R_f^{Rat,5p}$.

Figs. 6 and 7 show the comparison between real-world and ATP/ATPDraw simulated records of currents in the faulted phase (in pu), for cases 1 and 2, respectively. All tested HIF models presented very satisfactory R_f fitting, proving the accuracy of the evaluated HIF modeling strategies.

A superiority of R_f^{4p}, $R_f^{Rat,3p}$ and $R_f^{Rat,5p}$ over R_f^{3p} is noticed, especially in Case 2, which consists in a fault that takes longer to be cleared. Indeed, in Case 2, discrepancies observed for R_f^{3p} during the initial current increasing and over the final instants of the fault are not verified for R_f^{4p}, $R_f^{Rat,3p}$ and $R_f^{Rat,5p}$. However, differences between the fitting qualities of R_f^{4p}, $R_f^{Rat,3p}$ and $R_f^{Rat,5p}$ are not relevant, which leads to the conclusion that $R_f^{Rat,3p}$ is the best choice when a balance between accuracy and simplicity is considered.

The good fitting quality of $R_f^{Rat,3p}$ reveals also an important finding about the behavior of HIF R_f values. It is known that the $R_f^{Rat,3p}$ formula (Eq. (5)) represents a hyperbola, revealing that R_f does not exactly evolves as a single exponential, as assumed in various works. It explains difficulties found in properly fitting R_f using exponential regression formulas with only one time constant. Moreover, although $R_f^{Rat,3p}$ showed to be more accurate than R_f^{3p}, it requires the same number of settings, i.e., three coefficients. Hence, no additional complexity is observed, which is attractive for EMT-type simulations of line HIFs on vegetation.

V. METHODOLOGY TO APPLY THE $R_f^{Rat,3p}$ MODEL

Although $R_f^{Rat,3p}$ showed to be accurate, the need for regression procedures to fit the HIF model with real R_f measurements might be an issue. Also, for EMT simulations of fictitious cases, setting the HIF $R_f^{Rat,3p}$ model may not be intuitive, so that further guidelines are presented here.

Analyzing the HIF record as reported in [18], R_f samples over the time are estimated. Hence, $R_{f,ini}$, $R_{f,end}$ and R_f samples at different time instants can be identified, which are used here to develop a practical methodology to set $R_f^{Rat,3p}$. To facilitate the application of the proposed HIF modeling strategy via (5), the following relations can be considered:

$$R_{f,ini} = \frac{s_2}{s_3}, \text{ being } s_2 = s_3 \cdot R_{f,ini} \ , \tag{7}$$

$$s_1 = R_{f,end} \ . \tag{8}$$

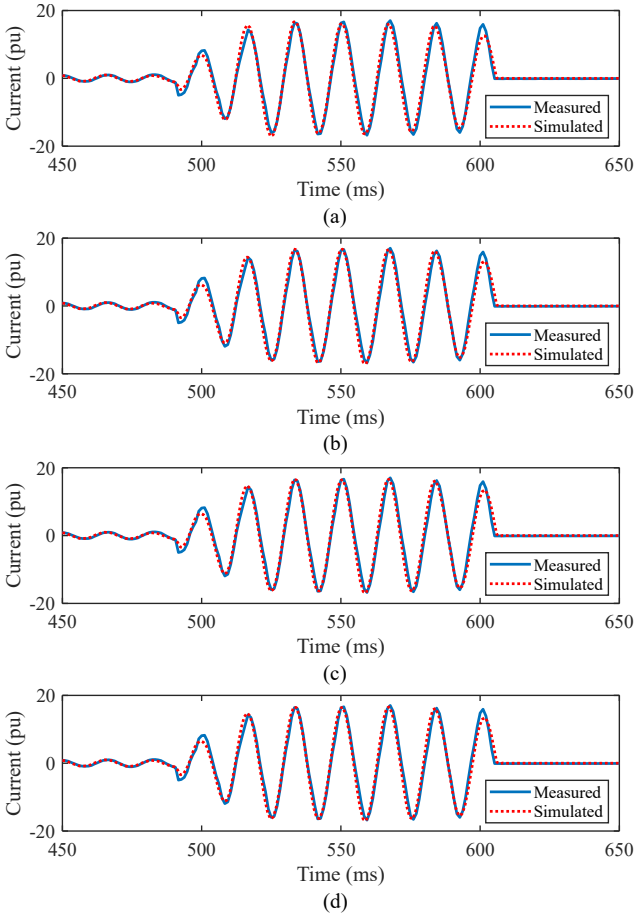

Fig. 6. Currents in Case 1: (a) R_f^{3p}; (b) R_f^{4p}; (c) $R_f^{Rat,3p}$; (d) $R_f^{Rat,5p}$.

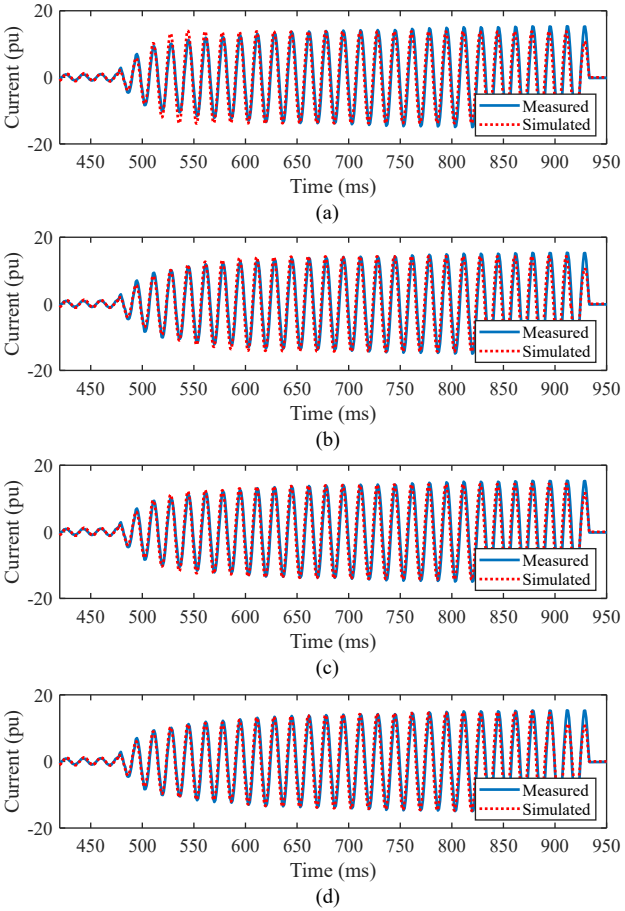

Fig. 7. Currents in Case 2: (a) R_f^{3p}; (b) R_f^{4p}; (c) $R_f^{Rat,3p}$; and (d) $R_f^{Rat,5p}$.

Equations (7) and (8) show that s_1 can be directly set as the final HIF resistance $R_{f,end}$. On the other hand, s_2 and s_3 cannot be directly set, requiring further setting procedures. In summary, assuming that, besides $R_{f,ini}$ and $R_{f,end}$ values, an R_f sample is available at a given instant t_x, i.e., $R_f(t_x)$, substituting t_x and $R_f(t_x)$ in (5), s_3 can be derived as:

$$s_3 = \frac{t_x \cdot [R_{f,end} - R_f(t_x)]}{[R_f(t_x) - R_{f,ini}]} \ . \tag{9}$$

From studies on (9), it was found that reliable estimations of s_3 can be calculated by considering R_f samples obtained when the initial HIF resistance decaying period is ending, around the knee point of the R_f decaying curve (see Fig. 1). For instance, in Case 1, the R_f curve knee point is at about $t_x = 10.4$ ms, when $R_f = 88.692\ \Omega$. On the other hand, in Case 2, the R_f curve knee point is at about $t_x = 20.8$ ms, when $R_f = 186.655\ \Omega$. Thus, substituting these values in (9), the following s_3 coefficients are calculated:

$$s_3^{case1} = \frac{10.4\text{E-}3(54.530 - 88.692)}{(88.692 - 314.646)} = 0.001572 \ , \tag{10}$$

$$s_3^{case2} = \frac{20.8\text{E-}3(79.060 - 186.655)}{(186.655 - 805.983)} = 0.0036135 \ , \tag{11}$$

which are then used to obtain s_2 via (7).

It must be noticed that, although s_3 coefficients calculated using (10) and (11) are close to those shown in Table I, they are slightly different. Indeed, s_3 coefficients listed in Table I result from more complex and accurate data regression procedures, which may not be practical for some potential users of the proposed HIF model. Even so, considering the presented practical setting approach, for the chosen parameters, one obtains RMSE $= 7.94\ \Omega$ and $R^2 = 0.954$ for Case 1, and RMSE $= 11.773\ \Omega$ and $R^2 = 0.965$ for Case 2, which consist in discrepancies comparable to those shown in Table II that attest the reliability and usefulness of such a practical HIF model setting procedure.

In Fig. 8, a flowchart that describes the application of the proposed HIF model based on $R_f^{Rat,3p}$ is presented, explaining setting procedures for both fictitious and real HIF cases. To demonstrate the feasibility of the proposed methodology, Fig. 9 compares R_f curves obtained from real field data (Cases 1 and 2) and from the $R_f^{Rat,3p}$ model adjusted by means of the practical methodology explained in Fig. 8. The results presented in Fig. 9 reveal a good $R_f^{Rat,3p}$ fitting, proving its applicability to emulate HIFs on vegetation in a very simple and accurate way, either in fictitious or real-world scenarios.

VI. Conclusions

A study on six different regression formulas able to emulate transmission line HIFs on vegetation was presented. Real-world cases of 500 kV/60 Hz transmission line HIFs on trees were assessed to evaluate the regression formulas. Four

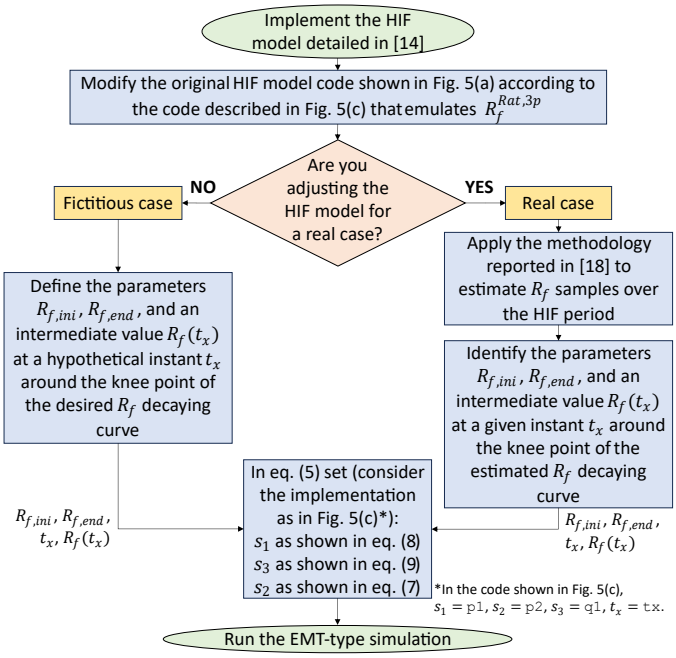

Fig. 8. Flowchart describing the application of the proposed HIF model.

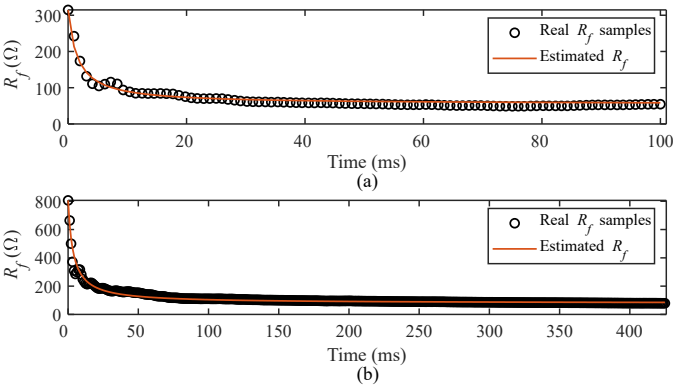

Fig. 9. Practical setting demonstration of $R_f^{Rat,3p}$: (a) Case 1; (b) Case 2.

promising regression strategies were firstly selected based on their fitting quality and practical aspects. Then, the regression formula with the best balance between simplicity and accuracy was chosen to be used in the proposed HIF model.

The chosen regression formula consists of a three-parameter rational function, which improves a previously published model with the same number of coefficients. ATP/ATPDraw simulations were carried out to demonstrate its performance by reproducing field HIF scenarios. Finally, a practical methodology to apply the proposed three-parameter HIF model was presented, which allows to simulate fictitious scenarios and to obtain a good HIF resistance fitting for real-world cases reproduction, without the need for complex regression procedures.

The obtained results reveal that HIFs evolve as a hyperbola function, which is properly reproduced by the proposed model. Furthermore, the results demonstrate that the designed model is reliable, accurate, and easy to set, being flexible to simulate both fictitious and real HIF scenarios. Also, the presented practical application methodology shows to be effective and easy to use, making the proposed modeling strategy even more attractive for researchers and utilities interested in emulating transmission line HIFs on vegetation.

REFERENCES

[1] N. I. Elkalashy, M. Lehtonen, H. A. Darwish, M. A. Izzularab, and A.-m. I. Taalab, "Modeling and experimental verification of high impedance arcing fault in medium voltage networks," *IEEE Trans. on Dielectrics and Electrical Insulation*, vol. 14, no. 2, pp. 375–383, 2007.

[2] W. C. Santos, B. Souza, N. S. Brito, F. B. Costa, and M. Paes, "High impedance faults: From field tests to modeling," *Journal of Control, Autom. and Electrical Systems*, vol. 24, no. 6, pp. 885–896, 2013.

[3] S. Nam, J. Park, Y. Kang, and T. Kim, "A modeling method of a high impedance fault in a distribution system using two series time-varying resistances in emtp," in *2001 Power Engineering Society Summer Meeting. Conference Proceedings*, vol. 2. IEEE, 2001, pp. 1175–1180.

[4] K. M. Shebl, E. A. Badran, and E. Abdalla, "A combined models-tacs atpdraw general model of the high impedance faults in distribution networks," *MEPCON*, vol. 10, pp. 19–21, 2010.

[5] S. Salona, "High impedance fault modelling on 11 kv feeder using matlab simulink," *International Journal of Science and Research*, 2016.

[6] A. Emanuel, D. Cyganski, J. Orr, S. Shiller, and E. Gulachenski, "High impedance fault arcing on sandy soil in 15 kv distribution feeders: contributions to the evaluation of the low frequency spectrum," *IEEE Transactions on Power Delivery*, vol. 5, no. 2, pp. 676–686, 1990.

[7] M. Wei, F. Shi, H. Zhang, F. Yang, and W. Chen, "A high-efficiency method to determine parameters of high impedance arc fault models," *IEEE Trans. on Power Delivery*, vol. 37, no. 2, pp. 1203–1214, 2021.

[8] V. Torres-Garcia, D. Guillen, J. Olveres, B. Escalante-Ramirez, and J. R. Rodriguez-Rodriguez, "Modelling of high impedance faults in distribution systems and validation based on multiresolution techniques," *Computers & Electrical Engineering*, vol. 83, p. 106576, 2020.

[9] P. E. Farias, A. P. De Morais, J. P. Rossini, and G. Cardoso Jr, "Non-linear high impedance fault distance estimation in power distribution systems: A continually online-trained neural network approach," *Electric Power Systems Research*, vol. 157, pp. 20–28, 2018.

[10] D. Gomes, C. Ozansoy, A. Ulhaq, and J. C. Júnior, "The effectiveness of different sampling rates in vegetation high-impedance fault classification," *Electric Power Syst. Research*, vol. 174, p. 105872, 2019.

[11] X. Wei, X. Wang, J. Gao, B. Wang, and D. Yang, "High impedance fault detection based on arc changing tendency and nonlinear least squares for active distribution networks," *IEEE Transactions on Industrial Informatics*, 2024.

[12] S. Maximov, V. Torres, H. Ruiz, and J. Guardado, "Analytical model for high impedance fault analysis in transmission lines," *Mathematical Problems in Engineering*, vol. 2014, 2014.

[13] J. Doria-García, C. Orozco-Henao, R. Leborgne, O. D. Montoya, and W. Gil-González, "High impedance fault modeling and location for transmission line," *Electric Power Systems Research*, vol. 196, p. 107202, 2021.

[14] F. V. Lopes, E. P. Ribeiro, R. L. Reis, K. M. Silva, A. M. Britto, C. M. Moraes, R. L. Agostinho, and M. A. Rodrigues, "Three-parameter atp/atpdraw transmission line high impedance fault model," *Electric Power Systems Research*, vol. 223, p. 109577, 2023.

[15] A. Muir and J. Lopatto, "Final report on the august 14, 2003 blackout in the united states and canada : causes and recommendations," Apr 2004.

[16] A. Ajayi, F. S. Osayi, D. K. Jerome, and A. Omoregie, "Investigating vegetation induced faults on power transmission line: A case study of the irrua-auchi-agenebode 33 kv transmission line edo state nigeria," *Global Journal of Advanced Research*, vol. 3, no. 3, pp. 177–183, 2016.

[17] C. Yang, T. Chen, B. Yang, X. Zhang, and S. Fan, "Experimental study of tree ground fault discharge characteristics of 35 kv transmission lines," in *IEEE Sustainable Power and Energy Conference*, 2021, pp. 2883–2891.

[18] L. Eriksson, M. M. Saha, and G. D. Rockefeller, "An accurate fault locator with compensation for apparent reactance in the fault resistance resulting from remore-end infeed," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-104, no. 2, pp. 423–436, 1985.

[19] T. F. Coleman and Y. Li, "An interior trust region approach for nonlinear minimization subject to bounds," *SIAM Journal on optimization*, vol. 6, no. 2, pp. 418–445, 1996.