Fault Distance Estimation for Transmission Lines with Dynamic Regressor Selection

Leandro A. Ensina[1*], Luiz E. S. de Oliveira[1], Rafael M. O. Cruz[2], George D. C. Cavalcanti[3]

[1*]Department of Informatics (DInf), Federal University of Paraná (UFPR), Curitiba, PR, Brazil.
[2]École de Technologie Supérieure (ÉTS), University of Quebec, Montreal, QC, Canada.
[3]Centre of Informatics (CIn), Federal University of Pernambuco (UFPE), Recife, PE, Brazil.

*Corresponding author(s). E-mail(s): leandro.ensina@ufpr.br;
Contributing authors: luiz.oliveira@ufpr.br;
rafael.menelau-cruz@etsmtl.ca; gdcc@cin.ufpe.br;

Abstract

This Supplementary Material companion the paper entitled "Fault Distance Estimation for Transmission Lines with Dynamic Regressor Selection", published in Neural Computing and Applications. This document is not intended to be self-contained and should be read jointly with the original paper. Section 1 shows complete details about the Gini scores, representing the importance of each feature for the Extremely Randomized Trees used in our method to compose the pool of regressors. In Sections 2 and 3, we present the results considering the variation of the hyperparameter $\boldsymbol{K}$, which determines the size of the region of competence, and the number of trees (estimators) in the pool, respectively.

1 Assessing the Feature Importances

Table 1 shows the features ordered individually by their importance for the Extremely Randomized Trees (ERT) algorithm considering the Gini score, as mentioned in Section 7.4 of the paper.

Table 1 Features ordered by their importance for the ERT algorithm, considering the mean and standard deviation, in parentheses, of the Gini score calculated considering the ten pre-trained ERT models

Ranking	**Feature**	**Gini score**
#1	stdev_phase_A_current	4.2000E-02 (2.6056E-03)
#2	stdev_phase_C_current	4.1800E-02 (2.2051E-03)
#3	stdev_phase_B_current	3.9100E-02 (2.3033E-03)
#4	variance_phase_B_current	2.9300E-02 (1.2778E-03)
#5	variance_phase_A_current	2.8300E-02 (1.7970E-03)
#6	variance_phase_C_current	2.7800E-02 (2.3230E-03)
#7	geometric_mean_phase_B_current	2.7400E-02 (1.6177E-03)
#8	geometric_mean_phase_A_current	2.6000E-02 (2.1522E-03)
#9	geometric_mean_phase_C_current	2.3800E-02 (2.1111E-03)
#10	distance_phase_B_voltage	2.3100E-02 (8.7426E-04)
#11	distance_phase_C_voltage	2.1900E-02 (8.3371E-04)
#12	auc_phase_B_current	2.0600E-02 (1.4745E-03)
#13	auc_phase_C_current	2.0300E-02 (2.3626E-03)
#14	covariance_phase_A_C_current	1.9900E-02 (1.5607E-03)
#15	covariance_phase_B_C_current	1.9600E-02 (1.0864E-03)
#16	distance_phase_A_voltage	1.9500E-02 (6.5629E-04)
#17	covariance_phase_A_B_current	1.9400E-02 (8.5223E-04)
#18	auc_phase_A_current	1.9200E-02 (1.6199E-03)
#19	energy_phase_C_voltage	1.8000E-02 (9.7622E-04)
#20	variance_phase_A_voltage	1.7900E-02 (1.1498E-03)
#21	variance_phase_B_voltage	1.7900E-02 (9.7280E-04)
#22	energy_phase_A_voltage	1.7800E-02 (1.4003E-03)
#23	energy_phase_B_voltage	1.7700E-02 (9.1380E-04)
#24	variance_phase_C_voltage	1.7000E-02 (9.5077E-04)
#25	rms_phase_C_current	1.6100E-02 (1.3401E-03)
#26	rms_phase_B_current	1.5500E-02 (1.8063E-03)
#27	rms_phase_B_voltage	1.5400E-02 (9.9355E-04)
#28	rms_phase_A_voltage	1.5400E-02 (1.3677E-03)
#29	stdev_phase_B_voltage	1.5300E-02 (1.1439E-03)
#30	stdev_phase_A_voltage	1.5200E-02 (1.0014E-03)
#31	rms_phase_A_current	1.5200E-02 (1.2586E-03)
#32	stdev_phase_C_voltage	1.5000E-02 (6.2306E-04)
#33	rms_phase_C_voltage	1.4800E-02 (9.1283E-04)
#34	auc_phase_C_voltage	1.3400E-02 (8.4250E-04)
#35	auc_phase_B_voltage	1.3000E-02 (6.8464E-04)
#36	auc_phase_A_voltage	1.3000E-02 (1.0135E-03)
#37	correlation_phase_A_B_voltage	1.0900E-02 (9.9020E-04)
#38	correlation_phase_A_C_voltage	1.0500E-02 (7.5469E-04)
#39	correlation_phase_B_C_voltage	1.0500E-02 (6.7458E-04)
#40	energy_phase_C_current	1.0200E-02 (8.2585E-04)

Table 1 Continued from the previous page

Ranking	Feature	Gini score
#41	energy_phase_B_current	9.6700E-03 (9.5086E-04)
#42	pk_pk_distance_phase_C_current	8.8800E-03 (1.1211E-03)
#43	energy_phase_A_current	8.8100E-03 (1.0431E-03)
#44	distance_phase_B_current	8.8100E-03 (1.1723E-03)
#45	pk_pk_distance_phase_B_current	8.7700E-03 (1.2930E-03)
#46	distance_phase_A_current	8.7600E-03 (7.7535E-04)
#47	distance_phase_C_current	8.3400E-03 (9.7109E-04)
#48	pk_pk_distance_phase_A_current	8.2100E-03 (9.9163E-04)
#49	covariance_phase_A_C_voltage	8.0500E-03 (3.1886E-04)
#50	covariance_phase_B_C_voltage	7.9600E-03 (4.2570E-04)
#51	covariance_phase_A_B_voltage	7.8800E-03 (3.5044E-04)
#52	maxFrequency_phase_C_current	6.9700E-03 (4.5325E-04)
#53	min_phase_B_current	6.8400E-03 (1.7545E-03)
#54	max_phase_C_current	6.7900E-03 (1.1355E-03)
#55	max_phase_A_current	6.3000E-03 (1.2483E-03)
#56	maxFrequency_phase_A_current	5.9600E-03 (4.9666E-04)
#57	min_phase_A_current	5.6500E-03 (8.0015E-04)
#58	min_phase_C_current	5.6500E-03 (7.3590E-04)
#59	correlation_phase_A_C_current	5.4900E-03 (1.5362E-04)
#60	correlation_phase_B_C_current	5.4700E-03 (1.9980E-04)
#61	maxFrequency_phase_B_current	5.4200E-03 (4.3794E-04)
#62	correlation_phase_A_B_current	5.3400E-03 (3.6046E-04)
#63	max_phase_B_current	5.3000E-03 (5.3685E-04)
#64	geometric_mean_phase_A_voltage	5.1000E-03 (4.7386E-04)
#65	geometric_mean_phase_B_voltage	4.9100E-03 (5.4003E-04)
#66	geometric_mean_phase_C_voltage	4.6300E-03 (2.8865E-04)
#67	pk_pk_distance_phase_B_voltage	4.0900E-03 (8.0593E-04)
#68	pk_pk_distance_phase_C_voltage	3.2300E-03 (5.6128E-04)
#69	pk_pk_distance_phase_A_voltage	2.7600E-03 (4.2818E-04)
#70	maxFrequency_phase_C_voltage	2.2100E-03 (2.0645E-04)
#71	maxFrequency_phase_B_voltage	1.8800E-03 (1.5277E-04)
#72	maxFrequency_phase_A_voltage	1.7900E-03 (1.3512E-04)
#73	kurtosis_phase_A_current	9.7200E-04 (6.7805E-05)
#74	kurtosis_phase_C_current	9.4900E-04 (6.3416E-05)
#75	kurtosis_phase_B_current	8.7200E-04 (7.4582E-05)
#76	max_phase_C_voltage	5.4300E-04 (1.6194E-04)
#77	kurtosis_phase_B_voltage	5.2500E-04 (1.9223E-05)
#78	max_phase_B_voltage	4.6800E-04 (7.5995E-05)
#79	kurtosis_phase_A_voltage	4.1400E-04 (2.5856E-05)
#80	min_phase_B_voltage	4.0500E-04 (1.8585E-04)
#81	kurtosis_phase_C_voltage	3.9300E-04 (1.8621E-05)
#82	min_phase_A_voltage	3.7900E-04 (1.2338E-04)

Table 1 Continued from the previous page

Ranking	Feature	Gini score
#83	min_phase_C_voltage	3.4400E-04 (1.6889E-04)
#84	max_phase_A_voltage	3.1700E-04 (7.8298E-05)
#85	harmonic_mean_phase_C_current	2.3400E-04 (1.3946E-04)
#86	harmonic_mean_phase_B_current	1.8600E-04 (1.5282E-04)
#87	harmonic_mean_phase_A_current	1.5300E-04 (1.0500E-04)
#88	mean_phase_B_current	1.5000E-04 (2.5766E-05)
#89	mean_phase_C_current	1.4900E-04 (3.3544E-05)
#90	shannon_entropy_phase_B_current	1.4800E-04 (5.1165E-05)
#91	shannon_entropy_phase_C_current	1.3300E-04 (3.0557E-05)
#92	mean_phase_A_current	1.3200E-04 (2.8143E-05)
#93	shannon_entropy_phase_A_current	9.8200E-05 (3.6484E-05)
#94	median_phase_A_current	8.3000E-05 (2.8707E-05)
#95	median_phase_C_current	7.7100E-05 (2.3456E-05)
#96	median_phase_B_current	7.3700E-05 (1.7774E-05)
#97	skewness_phase_C_current	6.4600E-05 (4.5346E-06)
#98	median_phase_B_voltage	6.3500E-05 (2.5503E-06)
#99	skewness_phase_A_current	6.3400E-05 (6.1738E-06)
#100	median_phase_A_voltage	6.2900E-05 (3.8083E-06)
#101	skewness_phase_B_current	6.2200E-05 (3.1106E-06)
#102	shannon_entropy_phase_A_voltage	6.2000E-05 (3.2964E-06)
#103	median_phase_C_voltage	6.1400E-05 (2.8213E-06)
#104	shannon_entropy_phase_B_voltage	5.0900E-05 (2.6573E-06)
#105	shannon_entropy_phase_C_voltage	4.9300E-05 (1.0120E-06)
#106	skewness_phase_C_voltage	4.7600E-05 (3.6985E-06)
#107	skewness_phase_A_voltage	4.7500E-05 (3.3593E-06)
#108	skewness_phase_B_voltage	4.6600E-05 (3.6953E-06)
#109	powerBandwidth_phase_C_voltage	4.1800E-05 (2.7130E-06)
#110	powerBandwidth_phase_B_voltage	3.8800E-05 (2.5542E-06)
#111	powerBandwidth_phase_A_voltage	3.4700E-05 (2.6729E-06)
#112	harmonic_mean_phase_B_voltage	3.3500E-05 (9.5276E-07)
#113	harmonic_mean_phase_A_voltage	3.2600E-05 (6.8064E-07)
#114	harmonic_mean_phase_C_voltage	3.1800E-05 (9.5736E-07)
#115	mean_phase_A_voltage	3.0500E-05 (6.2406E-06)
#116	slope_phase_B_voltage	3.0500E-05 (3.0171E-06)
#117	mean_phase_B_voltage	3.0400E-05 (3.3780E-06)
#118	mean_phase_C_voltage	2.9200E-05 (2.3653E-06)
#119	slope_phase_C_voltage	2.8700E-05 (3.4836E-06)
#120	slope_phase_C_current	2.6300E-05 (2.0945E-06)
#121	slope_phase_A_current	2.5700E-05 (1.6138E-06)
#122	slope_phase_B_current	2.5500E-05 (8.6944E-07)
#123	slope_phase_A_voltage	2.4800E-05 (1.6786E-06)
#124	powerBandwidth_phase_C_current	4.5000E-06 (2.1050E-06)
#125	powerBandwidth_phase_A_current	3.3700E-06 (1.0928E-06)
#126	powerBandwidth_phase_B_current	2.9400E-06 (7.2434E-07)

Table 2 shows the features grouped and ordered by their importance for the ERT algorithm considering the Gini score, as also mentioned in Section 7.4 of the paper.

Table 2 Features ordered by their importance for the ERT algorithm, considering the mean and standard deviation, in parentheses, of the Gini score calculated considering the ten pre-trained ERT models. Here, we grouped the features previously separated for each phase (A, B, or C) into a single feature

Ranking	**Feature**	**Gini score**
#1	stdev_current	4.0967E-02 (1.6197E-03)
#2	variance_current	2.8467E-02 (7.6376E-04)
#3	geometric_mean_current	2.5733E-02 (1.8148E-03)
#4	distance_voltage	2.1500E-02 (1.8330E-03)
#5	auc_current	2.0033E-02 (7.3711E-04)
#6	covariance_current	1.9633E-02 (2.5166E-04)
#7	energy_voltage	1.7833E-02 (1.5275E-04)
#8	variance_voltage	1.7600E-02 (5.1962E-04)
#9	rms_current	1.5600E-02 (4.5826E-04)
#10	rms_voltage	1.5200E-02 (3.4641E-04)
#11	stdev_voltage	1.5167E-02 (1.5275E-04)
#12	auc_voltage	1.3133E-02 (2.3094E-04)
#13	correlation_voltage	1.0633E-02 (2.3094E-04)
#14	energy_current	9.5600E-03 (7.0150E-04)
#15	pk_pk_distance_current	8.6200E-03 (3.5930E-04)
#16	distance_current	8.6367E-03 (2.5813E-04)
#17	covariance_voltage	7.9633E-03 (8.5049E-05)
#18	maxFrequency_current	6.1167E-03 (7.8679E-04)
#19	min_current	6.0467E-03 (6.8705E-04)
#20	max_current	6.1300E-03 (7.5941E-04)
#21	correlation_current	5.4333E-03 (8.1445E-05)
#22	geometric_mean_voltage	4.8800E-03 (2.3643E-04)
#23	pk_pk_distance_voltage	3.3600E-03 (6.7446E-04)
#24	maxFrequency_voltage	1.9600E-03 (2.2113E-04)
#25	kurtosis_current	9.3100E-04 (5.2374E-05)
#26	max_voltage	4.4267E-04 (1.1511E-04)
#27	kurtosis_voltage	4.4400E-04 (7.0930E-05)
#28	min_voltage	3.7600E-04 (3.0610E-05)
#29	harmonic_mean_current	1.9100E-04 (4.0731E-05)
#30	mean_current	1.4367E-04 (1.0116E-05)
#31	shannon_entropy_current	1.2640E-04 (2.5548E-05)
#32	median_current	7.7933E-05 (4.7057E-06)
#33	skewness_current	6.3400E-05 (1.2000E-06)
#34	median_voltage	6.2600E-05 (1.0817E-06)
#35	shannon_entropy_voltage	5.4067E-05 (6.9169E-06)
#36	skewness_voltage	4.7233E-05 (5.5076E-07)
#37	powerBandwidth_voltage	3.8433E-05 (3.5642E-06)
#38	harmonic_mean_voltage	3.2633E-05 (8.5049E-07)
#39	mean_voltage	3.0033E-05 (7.2342E-07)
#40	slope_voltage	2.8000E-05 (2.9138E-06)
#41	slope_current	2.5833E-05 (4.1633E-07)
#42	powerBandwidth_current	3.6033E-06 (8.0575E-07)

2 Assessing the Hyperparameter K

Table 3 and Fig. 1 show the results with different values for the hyperparameter K in the kNN algorithm used to define the size of the region of competence. These results justify the use of $K = 30$ since values less than 30 demonstrate a substantially lower performance, while the higher values do not demonstrate significant improvement, especially for the MAE metric. The repeated measures one-way ANOVA test manifested statistical differences among the performances (p-value < 0.0001 for MSE and MAE), while the Tukey post-test confirmed no statistical differences among $K = 30$, $K = 40$, and $K = 50$ for both measures.

Table 3 Mean and standard deviation, in parentheses, of the results in kilometers comparing different values for the hyperparameter K, calculated in ten replications

K	**MSE**	**MAE**
$K = 5$	2.1424 (0.2965)	0.8232 (0.0156)
$K = 10$	2.0140 (0.0821)	0.8187 (0.0102)
$K = 15$	1.7182 (0.1975)	0.7517 (0.0137)
$K = 20$	1.4806 (0.0862)	0.7195 (0.0080)
$K = 30$	1.3973 (0.0942)	0.7086 (0.0068)
$K = 40$	1.3827 (0.0972)	0.7069 (0.0081)
$K = 50$	1.3810 (0.0946)	0.7079 (0.0081)

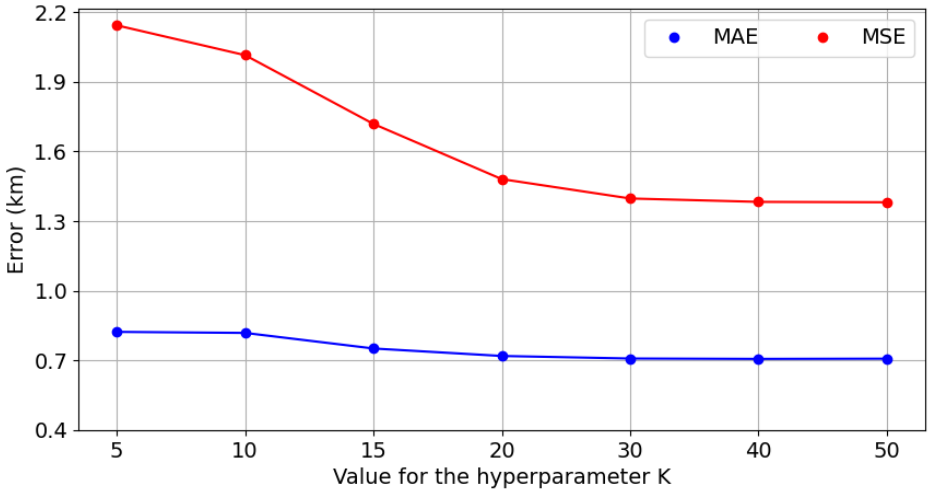

Fig. 1 Performances for the variation of the hyperparameter K

3 Assessing the Number of Estimators

Table 4 and Fig. 2 present the results with different values for the number of estimators for the Extremely Randomized Trees, which trees were used to compose the pool of regressors of our method. The performances were compared with the repeated measures one-way ANOVA test, revealing statistical differences among them (p-value < 0.0001 for MSE and MAE), while the Tukey post-test manifested no statistical differences among $N = 700$, $N = 800$, and $N = 900$ for both measures.

Table 4 Mean and standard deviation, in parentheses, of the results in kilometers comparing different values for the number of trees in the ensemble, calculated in ten replications

Number of trees	**MSE**	**MAE**
$N = 100$	1.8998 (0.0980)	0.8825 (0.0054)
$N = 200$	1.6220 (0.0577)	0.7872 (0.0068)
$N = 300$	1.5311 (0.1209)	0.7503 (0.0059)
$N = 400$	1.6421 (0.3543)	0.7330 (0.0044)
$N = 500$	1.4129 (0.0573)	0.7216 (0.0077)
$N = 600$	1.4035 (0.0969)	0.7132 (0.0076)
$N = 700$	1.3973 (0.0942)	0.7086 (0.0068)
$N = 800$	1.3964 (0.0989)	0.7065 (0.0074)
$N = 900$	1.3969 (0.1155)	0.7059 (0.0070)

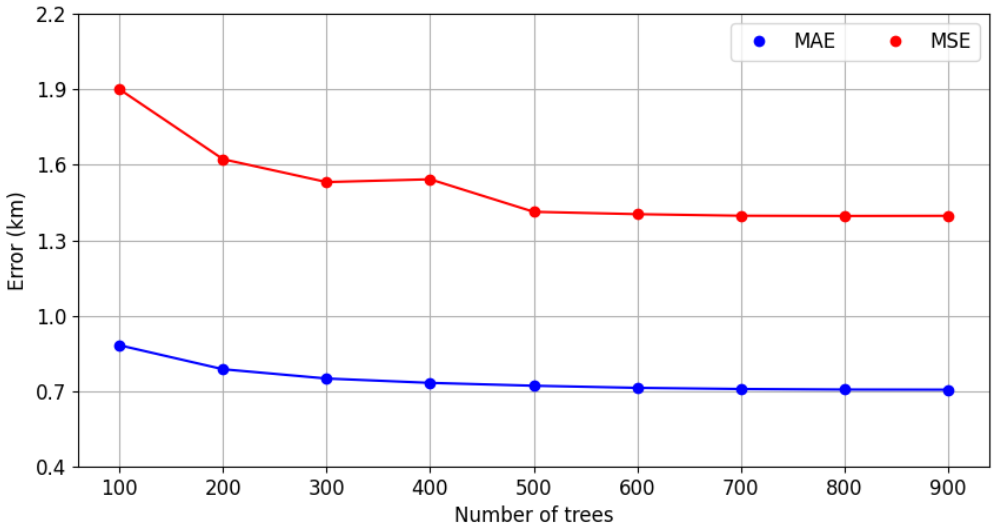

Fig. 2 Performances for the variation of the number of trees in the pool of regressors