Docket No. E-2 Sub 1320
Presson Exhibit 14

Identifying Factors Contributing to Poor Performance of Near-Real-Time Power Flow

Gokhan Cakir[1], Ken Crawford[1], Mesut Baran [1], Valentina Cecchi[2], Badrul Chowdhury[2], Oluwatimilehin Adeosun[2], Mariann Thomas[2], Cara DeCoste Chacko[3]

[1] Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, USA 27606
[2] Department of Electrical and Computer Engineering, University of North Carolina Charlotte, Charlotte, NC, USA 28223
[3]Duke Energy, Charlotte, NC, USA 28202

***Abstract*— Accurate near real time monitoring is needed for Distribution Management Systems (DMS), and power flow-based methods are commonly used in practice for this purpose. However, near real time power flow results are not always accurate because of the poor load estimation obtained through the bus load allocation (BLA) procedure used in this approach. This paper focuses on this issue and proposes a data analytics-based method to parse through data and identify factors/parameters that are significant and relevant to poor near real time power flow solution. The proposed method uses a Binary logistic regression method to identify these significant parameters for the feeders having poor BLA performance. Test results show that the significant parameters can help to pinpoint the cause(s) for poor BLA performance.**

***Keywords*— Distribution Management System (DMS), Logistic Regression, Bus Load Allocation (BLA), Machine Learning**

I. INTRODUCTION

Many utilities have recently started deploying a Distribution System State Estimator (DSSE) for providing situational awareness that is needed for applications such as voltage monitoring and control [1]. One of the methods used in practice is the power flow-based distribution system state estimation [2]. We will also call it near real time power flow (RTPF) in this paper. In practice, RTPF is integrated with a Distribution Management System (DMS). Accurate RTPF results are critical to safely and efficiently operate the distribution system.

One of the challenges in RTPF application is that only a limited number of actual measurements are available. Hence, a load estimation method is needed to provide an estimate for the load. This estimated load is then passed to a power flow program to obtain an estimate of node voltages. Since the power flow program depends on the estimated load, performance of RTPF (accuracy of voltage estimates) critically depends on the accuracy of the load estimate [3].

Another common issue that affects the performance of RTPF is the distribution system model used in the power flow program. Most commercial PF programs use detailed three phase circuit models. However, any error in the line models or other devices will have impact on the accuracy of the results. Hence, it becomes very challenging to detect and identify factors when the performance of the RTPF is not satisfactory. This is the problem we focused on in this paper.

We could not find any literature directly related to this practical problem. Most of the research and literature has been on the bad data detection and identification using state estimation [4]-[6]. The methods used for bad data detection will not be very effective for this problem given that there are only very limited real time measurements [7]. Furthermore, bad data detection considers only one snapshot/sample, and thus, it limits the ability to identify long term effects/behavior.

We collaborated with our local utility, Duke Energy, to investigate the poor performance issues and develop a method to address them. Since load allocation is the critical component for RTPF, we first focused on the performance of this component, which is called the Bus load Allocation (BLA). Duke Energy engineers have observed that while BLA converges for most of the feeders, there are some feeders for which BLA does not converge on a consistent basis resulting in inaccurate power flow results, and existing methods of analysis regularly fail to determine the cause of the inaccuracy due to the size and complexity of the distribution system model, and the time-consuming nature of circuit-by-circuit analysis.

This paper focuses on this problem and proposes a methodology for identifying the factors that contribute to poor BLA performance on a given feeder.

The proposed method has two steps. In the first step we adopt a binary regression-based method to check if there are some features associated with a given poorly performing feeder that makes the feeder different from good-performing feeders. We call these features distinguishing features/parameters. In the second step, we perform a detailed analysis on the sample RTPF cases by using the significant parameters in order to further pinpoint the probable factors contributing to the poor performance on that feeder.

Section II introduces the proposed method. The results of a Case study are outlined in section III. Conclusions are given in section IV.

978-1-6654-9071-9/23/$31.00 ©2023 IEEE

II. INDINTFYING FEATURES OF A FEEDER WITH POOR BLA

A. Bus Load Allocation for RTPF

As the first step of RTPF, a Bus Load Allocation (BLA) procedure is used to obtain an initial estimation of load (kW and kVar for each load point/node for a given time of the day RTPF is run. A common method, considered in this paper, uses the “nominal load” values (which are based on historical data) for each load class as initial guess, and then scales them iteratively by using the real-time measurements from SCADA. In each iteration, a power flow is run first to get an estimate for the power flows that are monitored at certain locations on the feeder. The differences between the calculated values and the actual measurements are then calculated as mismatches for convergence check (to see if the load allocation is acceptable). If the difference is within a given tolerance, then BLA converges. If the tolerance is not reached, then the BLA scales the loads by using the mismatches [8]. The flowchart of the BLA process is given in Fig.1.

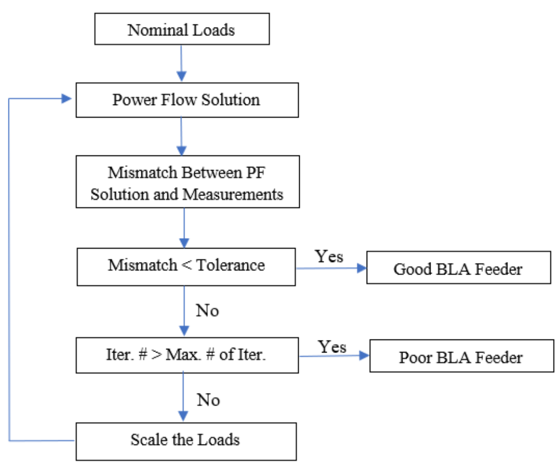

Fig 1. RTPF Flowchart

B. Parameter Selection

As indicated in the introduction, we focus on identifying the factors that contribute to poor BLA performance on a given feeder most of the time. We will call such feeders “poor BLA feeders”. As a first step, we use a regression method to check if there are some features associated with a poor BLA feeder that makes the feeder different from good BLA feeders.

For regression analysis, we first select a set of features/parameters to use as input data. In selecting these parameters, we consider the parameters that will impact the BLA performance. Some of these relevant parameters related to the feeder are feeder characteristics (overhead and underground sections), loading level, load mix etc. We also consider the measurement data itself (feeder head power flow measurements).

The set of parameters selected based on the feedback from Duke Energy engineers are given in Table I.

For regression analysis, a sample data set needs to be collected from the DMS for a set of feeders. RTPF samples (which are called save cases) are taken from real feeders dated in 2020 from the sponsoring utility in a stratified manner. The samples should be cases from a diverse set of feeders that have good BLA performance as well as poor BLA performance on a consistent basis.

TABLE I. SET OF PARAMETERS

1)	Total Load (kW)	18)	Total OH Line Length (ft)
2)	Total Load (kVar)	19)	Total 1ph Load (kW)
3)	Total Residential Load (kW)	20)	Total 1ph Load (kVar)
4)	Total Residential Load (kVar)	21)	Total 2ph Load (kW)
5)	Total Commercial Load (kW)	22)	Total 2ph Load (kVar)
6)	Total Commercial Load (kVar)	23)	Total 3ph Load (kW)
7)	Total Industrial Load (kW)	24)	Total 3ph Load (kVar)
8)	Total Industrial Load (kVar)	25)	Number of Caps ON
9)	Distribution Transformer Loss (kW)	26)	Number of Distribution Transformer
10)	Distribution Transformer Loss (kVar)	27)	Amp Flow Imbalance (%)
11)	Line Losses (kW)	28)	Primary Meter Load (kW)
12)	Line Losses (kVar)	29)	Primary Meter Load (kVar)
13)	Shunt Capacitance (kVar)	30)	Phase A Maximum Volt Drop (V)
14)	Cap Injection (kVar)	31)	Phase B Maximum Volt Drop (V)
15)	Top of Feeder Measurement (kW)	32)	Phase C Maximum Volt Drop (V)
16)	Top of Feeder Measurement (kVar)	33)	Average Primary to Secondary Volt Drop (V)
17)	Total UG Cable Length (ft)		

C. Identfying Distinguishing Parameters

The goal in this step is check whether the selected set of parameters are enough to differentiate a poor BLA feeder from a good BLA feeder. We adopted the Binary Logistic Regression for this purpose. Regression methods (which are supervised machine learning methods) aim at building a mathematical model that captures the relationship between the independent variables (in our case values of parameters as input data) and the observations (in our case whether BLA is converged or not for a given DMS save case as output data). Hence, a properly selected regression method can help us verify that selected parameters can help us to differentiate/estimate if the BLA is mostly converging or not for a given feeder. In other words, logistic regression is a classification algorithm that predicts a binary outcome based on a series of independent variables. Another main output of the logistic regression method is that it can pinpoint which of the parameters used are the so-called *significant parameters*; these are the parameters that contribute the most to the estimated outcome. In our case we used the binary logistic regression to extract the significant parameters rather than predicting the outcome although it is frequently used for prediction of binary outcome with given independent input variables.

1) Binary Logistic Regression

Binary Logistic Regression is a supervised machine learning algorithm that helps in binary classification (separating discrete values, such as Yes/No, Pass/Fail) on a set of observations. Logistic regression is a method that fits a regression curve, y = f(x) on a sample observations y and y is a categorical variable (Yes/No, Pass/Fail). Fig. 2 illustrates this approach.

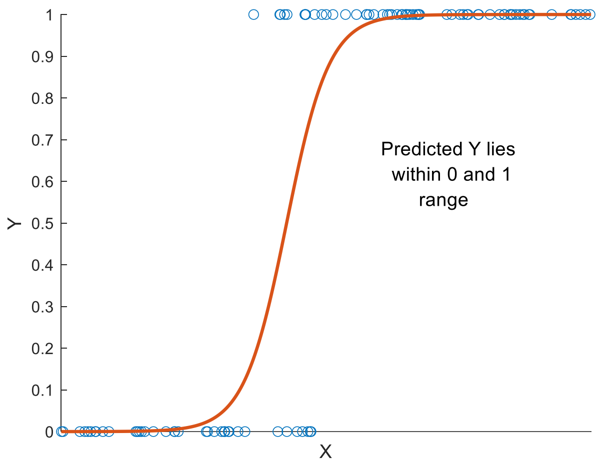

Fig 2. Logistic Regression

In the Logistic Regression model, we need to choose a logistic function [9]. The standard logistic function, for predicting the outcome of an observation given a predictor variable (X), is an s-shaped curve defined as:

$$p(X) = \frac{e^y}{1+e^y} \tag{1}$$

Where $p(X)$ is the probability of event to occur given X, and $y = \beta_0 + \beta_1 X$. To fit the model (1), we use a method called maximum likelihood [10].

When we consider the problem of predicting a binary response using multiple predictors, the logistic function looks like:

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \tag{2}$$

The coefficients $\beta_1, \beta_2, \ldots, \beta_n$ are unknown and must be estimated based on the available data. We seek estimates for $\beta_1, \beta_2, \ldots, \beta_n$ such that plugging these estimates into the model for $p(X)$ yields a number close to 1 for all individuals which are converging, and a number close to 0 for all individuals which are not converging [10]. Probability threshold normally is taken as 0.5. We classify an element in one category or the other depending on whether its probability exceeds the probability threshold or not.

2) Model Selection

Running a regression model with too many variables especially with irrelevant ones will lead to a needlessly complex model.

It is often the case that some or many of the variables used in a multiple logistic regression model are in fact not associated with the response. Including such irrelevant variables leads to unnecessary complexity in the resulting model. Moreover, an enormous number of variables can lead to overfitting and high variance of the coefficient estimates for the logistic regression. Therefore, stepwise methods, which explore a far more restricted set of models, are attractive alternatives to best subset selection [10]. We can obtain a model that is more easily interpreted. Stepwise selection [10] can help to choose the variables to add.

Selection begins with a regression model which includes all candidate variables. Variables are then eliminated from the model one by one until all the variables remaining in the model exceed certain criteria. At each step, the variable showing the smallest improvement to the model is removed. In other words, at each step the variable that gives the greatest additional improvement to the fit is added to the model.

One criterion to select the optimal set of features is Akaike Information Criterion (AIC) [11]. The AIC is:

$$AIC = -2logL + 2k \tag{3}$$

Where L is the likelihood of the candidate model given the data when evaluated at the maximum likelihood estimate, and k is the number of estimated parameters in the model.

This value is calculated for every candidate model after each variable elimination on the logistic regression model and the best model is the model with the smallest AIC [12,13]. AIC estimates the quality of each model relative to the other models and provides a single number score that can be used to determine which of the models is the best fit to the data.

3) Identfying Significant Parameters of a Poor BLA Feeder

We use the model obtained by using binary logistic regression to identify the features/parameters of a poor performing feeder. To illustrate the process, consider the logistic regression results of a sample feeder –named SB41- given below in Table II (details of this case is given in the next section).

TABLE II. THE SUMMARY OF LOGISTIC REGRESSION MODEL FOR SB41

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	3.9651	1.4758	2.687	0.007215	**
Total.Load...kVAR.	-10.6953	4.3059	-2.484	0.012997	*
Total.RES.Load...kVAR.	-17.4742	7.0101	-2.493	0.012677	*
Total.COMM.Load...kVAR.	26.4408	8.0880	3.269	0.001079	**
Total.COMM.Load...kW.	-12.3453	4.6153	-2.675	0.007477	**
Line.Losses...kVAR.	126.8896	52.2947	2.426	0.015248	*
Shunt.Capacitance..Cable.and.Line.Capacitance..kVar	77.0408	47.2296	1.631	0.102849	
Top.of.Feeder.Meas...kW.	14.1221	5.7506	2.456	0.014058	*
Total.Underground.Cable.Length	2.3390	0.5784	4.044	5.25e-05	***
Total.OH.Line.Length	-2.3543	0.6776	-3.474	0.000512	***
Number.of.ON.Caps	-0.8626	0.3769	-2.289	0.022089	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

These summary statistics provide us the estimated model parameters (intercept and coefficients betas), and statistics - standard error, z-value, and p-value – for each model parameter. In the table, the estimate column gives the intercept (β_0) and the coefficient estimates associated with each predictor variable. Std. error column provides the standard error of the coefficient estimates. Z value is the coefficient estimate divided by the standard error of the estimate. Z value is a measure of position that indicates the number of standard deviations a data value lies from the mean. The p-value corresponds to the z-statistic, and it is the likelihood of the calculated Z value. The smaller the p-value, the more significant the estimate is.

Since the smaller the p value the more significant the estimate is, table also show stars (*, **, ***) to indicate the significance level for each parameter according to p-value. We

TABLE IV. PARAMETER VALUES FOR SAVECASES OF SB41

Save Case Day	Total kW Load	Total kVar Load	Dist. Xmer Losses kVar wrt Total Load KVA (%)	Line Losses kVar wrt Total OH Line Length (%)	Cable and Line Capacitance kVar wrt UG Cable Length (%)	Total kVar Losses	Total kVar Losses wrt Total Load KVA %	Cap Injection kVar	Top of Feeder Measurement kVar	kVar Mismatch (Calculated – Measurement)	Total kVar Load to be Allocated (Q_{aloc})
3/17 4 pm	1314.76	121.35	18.1%	0.005%	0.10%	209.51	15.9%	-1419.57	-1339	250.29	-128.9
3/17 2 pm	1397.54	133.45	16.9%	0.005%	0.09%	209.34	14.9%	-1412.2	-1311	241.59	-108.1
5/24 11 am	1369.62	212.41	16.9%	0.006%	0.09%	207.38	15%	-1393.94	-1130	155.85	56.56
3/17 9 am	1962.12	183.66	12%	0.013%	0.09%	220.92	11.2%	-1399.47	-1207	212.11	-28.45
3/17 11 am	1771.23	148.09	13.3%	0.010%	0.09%	215.49	12.1%	-1400.22	-1260	223.36	-75.27
5/13 4 pm	988.33	125.76	23.9%	0.004%	0.10%	207.27	20.8%	-1422.08	-1322	232.95	-107.2
5/20 10 am	961.45	30.26	24.8%	0.004%	0.10%	206.87	21.5%	-1262.57	-1272	246.56	-216
4/24 4 pm	1510.74	162.35	15.7%	0.006%	0.09%	212.52	14%	-1419.27	-1273	228.6	-66.25
4/26 3 pm	961.25	155.35	24.5%	0.003%	0.10%	205.61	21.1%	-158.69	-3	205.27	-49.92
4/29 1 pm	1085.08	147.45	21.9%	0.004%	0.10%	209.19	19.1%	-1428.05	-1284	212.59	-65.14

have used 0.01 significance level; in other words, we selected the parameters with two and three stars as "significant parameters" and thus get the following:

1. Total Underground Cable Length
2. Total OH Line Length
3. Total Commercial Load kVar
4. Total Commercial Load kW

We call these significant parameters the *distinguishing parameters* for this particular feeder with poor BLA performance.

To help assess the overall performance of the regression model we obtained, we use the confusion matrix based on the results. This confusion matrix is given in Table III. NC stands for non-converging feeder, and C stands for converging feeder.

TABLE III. CONFUSION MATRIX FOR SB41

	Predicted NC	Predicted C
Actual NC	3	1
Actual C	0	122

From the Confusion Matrix, the accuracy of the model is 99.2%, which indicates that the method performance is good. This good performance indicates that regression model can distinguish this poor BLA feeder from the other feeders with good BLA performance. Furthermore, significant/distinguishing parameters identified are most likely the parameters that distinguish this feeder from good BLA feeders.

A basic way to make use of these distinguishing parameters is to compare how different these parameters are between that of poor BLA feeders and good BLA feeders. To do this, average values of the significant parameters using poor BLA feeders are determined and compared to that of good BLA feeders. All the parameters given in percentages are normalized with respect to their nominal values. Table V shows this comparison.

Table V compares the average values of the distinguishing parameters of this feeder to that of the feeders with good BLA performance. As seen in the Table V, Total OH Line Length and Total UG Cable Length for feeder SB41 are substantially higher than average values of good BLA feeders. Total Commercial Loads are also substantially lower than average values of good BLA feeders.

TABLE V. AVERAGE VALIUES OF PARAMETERS – SB41

Distinguishing Parameters	Good BLA Feeders Average Values	SB41 Feeder Average Values
Total OH Line Length %	86%	266%
Total UG Cable Length %	85%	133%
Total Commercial Load kVar %	21%	4.6%
Total Commercial Load kW %	38%	21%

Note finally that the distinguishing parameters may not directly point to a cause or causes for the poor BLA performance on the feeder. However, since the parameters are selected such that, they correlate to the difference between good and poor performance, we can use these distinguishing parameters to pinpoint the cause. This is illustrated in the next step.

D. Case Analysis to Pinpoint the Causes

In this second step, the goal is to use the distinguishing parameters we identified for a feeder with poor BLA performance to further pinpoint the cause(s) for this poor BLA convergence. This step involves looking into the save case data we have on the feeder more deeply in order to identify probable cause(s). This process will be illustrated in the following case study.

III. CASE STUDY

Sample data includes 440 RTPF samples (save cases) from the sponsoring utility. We have about 8-10 samples for each feeder. Sample save cases have mostly good BLA, and 4 feeders have poor BLA, which are SB41, SB51, SB52, and LA 45.

Distinguishing parameters are obtained for each poor BLA feeder with logistic regression method by fitting the sample data. We used 70% of consistent save cases for training to fit the model and test the model using the remaining 30% of the data.

A. Feeder LA45

One of the poor BLA feeder is LA45, and the summary of the logistic regression model for LA45 is given in Table VI.

TABLE VI. THE SUMMARY OF LOGISTIC REGRESSION MODEL FOR LA45

```
Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                    4.08690    1.75230   2.332  0.01968 *
Total.RES.Load...kW.         -14.51257    9.79695  -1.481  0.13852
Primary.meter.Size..kW.      -61.80011   16.14958  -3.827  0.00013 ***
Total.RES.Load...kVAR.       -32.47514   11.85130  -2.740  0.00614 **
Total.COMM.Load...kVAR.       16.38510    8.66500   1.891  0.05863 .
Total.IND.Load...kVAR.        15.18761    8.10391   1.874  0.06092 .
Total.IND.Load...kW.         -18.15733    8.38618  -2.165  0.03038 *
Line.Losses...kW.           -486.31392  291.88493  -1.666  0.09569 .
Line.Losses...kVAR.          227.33317  135.57706   1.677  0.09359 .
Top.of.Feeder.Meas...kW.      26.40834   11.16925   2.364  0.01806 *
Top.of.Feeder.Meas...kVAR.     6.71333    3.38466   1.983  0.04732 *
Cap.Injection...kVAR.        -24.64429   11.73043  -2.101  0.03565 *
Total.OH.Line.Length           4.42143    1.74896   2.528  0.01147 *
Number.of.ON.Caps             -1.63736    0.78529  -2.085  0.03707 *
Number.of.Dist.Xmr           -14.52473    6.58698  -2.205  0.02745 *
Amp.flow.imbalance..           0.04008    0.02492   1.608  0.10773
Two.Phase..kW.              -227.39235  151.02524  -1.506  0.13216
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

As the table shows, the distinguishing parameters for this feeder are primary meter size load and residential kVar load. Table VII compares the average values of these significant parameters between the two groups. These results indicate that LA45 feeder has one large industrial load (monitored by a primary meter), and very low residential load. Since the industrial load is monitored, BLA has only the small residential load to adjust to match the power flow measurement. This is the main reason why BLA is not converging.

TABLE VII. AVERAGE VALUES OF PARAMETERS – LA45

Distinguishing Parameters	Good BLA Feeders Average Values	LA45 Feeder Average Values
Primary Meter Size kW Load %	3.06%	58.6%
Total Residential Load kVar %	11.9%	1.3%

B. Feeder SB41

The regression results obtained for feeder SB41 are given in the previous section. As indicated before, the distinguishing parameters obtained are:

1. Total Underground Cable Length
2. Total OH Line Length
3. Total Commercial Load kVar
4. Total Commercial Load kW

For the next step of using these parameters to pinpoint the possible cause(s) for poor BLA performance, we first try to narrow down the cause to one of the main sources of errors: modeling issues, measurement errors, or algorithm issues with RTPF. Given that real power load allocation is acceptable, we can assume that the main measurement at the feeder head is also acceptable. Therefore, we focus next to see if the data indicates any modeling issues.

When the BLA does not converge, it provides a detailed report as part of the output. When we examine these reports from the save cases, we see that the reason for non-convergence for the cases is that there is too little reactive power (kVar) load to be allocated.

To help understand this issue, we first get an estimate of the total kVar load to be allocated to the loads using the DMS power flow results. Total kVar load to be allocated to the loads is estimated using the power balance as follows:

$$Q_{alloc} = Q_{o\_feeder} - Q_{loss} - Q_{cap} \tag{4}$$

Where Q_{o_feeder} is the kVar measurement at feeder head, Q_{loss} is total kVar loss, and Q_{cap} is total capacitor injection. We find that Q_{alloc} is a good indicator, as when Q_{alloc} is negative or very low it means that there is not enough kVar to be allocated based on the kVar measurements. Some of the detailed data for SB41 is given Table IV. It shows that Q_{alloc} is quite negative for all the save cases, and thus, it is clear that the BLA cannot adjust kVar of the loads given the pf constraints.

TABLE VIII. KVAR LOAD ADJUSTMENT BY BLA

Iteration No.	Phase A	Phase B	Phase C
1	488	263	652
2	20.8	10.6	158.4
3	16.1	9.1	86.3
4	16.2	9.1	95.3
5	16.2	9.1	95.9

To confirm this, we also examined how BLA adjusts the kVar load at each iteration. Table VIII show the results for the save case on March 17, 4pm. As seen in the table, BLA keeps decreasing the load kVar in order to reduce the mismatch, and it reaches the minimum kVar limit due the maximum power factor constraints (of 0.9995). The same convergence pattern is observed for other save cases for this feeder. These results clearly indicates that the issue in this case is related to the factors that contribute to the negative Q_{alloc}.

To further pinpoint a cause, a more detailed data from the save cases for this feeder is used. Table IV shows that kVar losses with respect to total kVar load on this feeder are much higher than average value of good BLA feeders (7.8%). Furthermore, distribution transformer losses with respect to total kVar load are much higher for this feeder than the average value of good BLA feeders (8.3%). These observations indicate that one contributing factor is the high distribution transformer kVar losses, partly due to feeder being a long feeder (a distinguishing parameter) with mostly residential loads served by single phase transformers. Interestingly, this contributing factor of too many single-phase distribution transformers is indirectly indicated by the distinguishing parameter of low commercial load on this long feeder.

IV. CONCLUSION

The results on an actual case study indicate that the proposed logistic regression-based method can be used to check if feeders with poor bus load allocation (BLA) have features that separate them from feeders with good BLA. Furthermore, the method can be used to identify the distinguishing parameters (that separate a poor BLA feeder from good BLA feeders) from the long list of parameters initially selected.

The results also illustrate the challenge associated with using the distinguishing parameters to pinpoint the cause for poor BLA performance on a given feeder, as the distinguishing parameters do not directly indicate the cause for poor BLA performance on a feeder. Therefore, the paper proposes a detailed analysis to pinpoint the cause(s) for a feeder with consistent poor BLA performance. The case study illustrates that the distinguishing parameters can help in this search in this second step.

ACKNOWLEDGMENT

The authors gratefully acknowledge the contribution of Duke Energy and Center for Advanced Power Engineering Research (CAPER)

REFERENCES

[1] J. Schoene, M. Humayun, S. Kamma, B. Russel, G. Sun, J. Bui, A. Salazar, N. Badayos, M. Zhong, M. Lak, C. Clarke, M. Zhong, N. Navarro, and A. Zewila" Historical AMI Data for Distribution System State Estimator Deployment," IEEE SoutheastCon , March 2021. pp 1-5

[2] J. Schoene, M. Humayun, B. Russel, G. Sun, J. Bui, A. Salazar, N. Badayos, M. Zhong, M. Lak, and C. Clarke, " Quantifying performance of distribution system state estimators in supporting advanced applications," IEEE Open Access J. Power Energy, vol. 7, pp. 151–162, Apr. 2020.

[3] C. Muscas, S. Sulis, A. Angioni, F. Ponci, A. Monti, “Impact of Different Uncertainty Sources on a Three-Phase State Estimator for Distribution Networks,” IEEE Transactions On Instrumentation And Measurement, Vol. 63, No. 9, September 2014.

[4] D. Wang, X. H. Guan, T. Liu, Y. Gu, Y. Sun, and Y. Liu, “A survey on bad data injection attack in smart grid,” in Proc. IEEE PES AsiaPacific Power and Energy Engineering Conf., Hong Kong, China, 2013, pp. 1–6.

[5] A. Abur and A.G. Exposito, “Power System State Estimation: Theory and Implementation”, (Markel Dekker, Inc.,2004)

[6] Pau, M., Ponci, F., Monti, A. “Analysis of bad data detection capabilities through smart meter based state estimation”, 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I CPS Europe), pp. 1–6 (2018)

[7] A. Primadianto, and C.N. Lu, “A Review on Distribution System State Estimation” IEEE Transactions on Power Systems, vol. 32, no. 5, September 2017

[8] “Distribution Network Analysis Functions (DNAF) Standard Applications Analyst’s Guide”, April, 2016

[9] X. Zou, Y. Hu, Z. Tian, and K. Shen, “Logistic Regression Model Optimization and Case Analysis”, 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), pp 1-5

[10] Gareth James, Daniela Witten,Trevor Hastie, Robert Tibshirani, “An Introduction to Statistical Learning with Applications in R”, Second Edition, 2021

[11] Akaike, H. “A new look at the statistical model identification” IEEE Transactions on Automatic Control, 19(6), 716-723, 1974

[12] F. J. Fabozzi, S. M. Focardi, S. T. Rachev, and B. G. Arshanapalli, “The Basics of Financial Econometrics: Tools, Concepts, and Asset Management Applications”, 2014 John Wiley & Sons, Inc.

[13] M. Wesolawska, “A study on feature selection based on AICc and its application to microarray data”, Final project in the field of informatics, Universitat Polit`ecnica de Catalunya, 2009