USING A MULTIVARIATE DOE METHOD FOR CONGESTION STUDY IN DISTRIBUTION SYSTEMS UNDER IMPACTS OF PLUG-IN ELECTRIC VEHICLES

Masoud ALIAKBAR GOLKAR
K. N. Toosi University of Technology – Iran
golkar@eetd.kntu.ac.ir

Hamed VALIZADEH HAGHI
K. N. Toosi University of Technology – Iran
valizadeh@ieee.org

ABSTRACT

In emerging active distribution networks there is a need to introduce advanced tools for studying new framework. Indeed, the traditional studies might not be able to take the impacts of new system components such as electric vehicles into account well. Therefore, it seems a good idea to examine the possibility of employing more advanced hybrid tools for studying active networks. This paper provides the groundwork for a new design of experiment (DOE) method as employed to the congestion study in active distribution networks considering impacts of plug-in electric vehicles (PEVs). The new DOE method utilizes a multivariate modeling of the huge database of all possible scenarios by using copula approach. Then, the most optimally reduced database could be used for performing power flow calculations followed by a correlation analysis of the results. The final outcome is about knowing which lines will be congested at the same time and to what extent under impacts of PEVs.

INTRODUCTION

Planning distribution systems requires decision tools to determine which part of the network should be developed in the future. In the context of the emerging active distribution networks, undertaking a partial development in the planning stage is further encouraged due to the proliferation of plug-in electric vehicles (PEVs). If a network development decision is not taken at the right time, congestion may appear on the grid. On the other hand, assuming unnecessary network developments may be economically abortive.

Monte Carlo simulation, naturally, provides a very comprehensive tool used to evaluate potential impacts of PEVs based on probabilistic projections of certain aspects of PEV penetration. The Monte Carlo approach is intended to capture both spatial and temporal diversity of PEV integration as customers in varied locations purchase PEVs of varied types and charge them differently. Therefore, the simulations are composed of probabilistic assignment of PEVs to the distribution base case [1]. Each PEV is randomly assigned a location, type, and daily charge profiles based on the provided probability density functions for each characteristic. In this manner, multiple probabilistic scenarios are generated from the system and probability density functions. In other words, there are millions of possible configurations when the chosen factors vary.

To make possible taking into account all of the possible configurations in an acceptable simulation time, a novel multivariate design of experiment (DOE) method is proposed in this paper to create a reduced database. The randomization and the probabilistic modeling make it possible to create an optimal DOE of fewer configurations chosen between the millions of possible configurations. However, unlike DOEs for univariate responses, design for multiple responses has not received much attention [2]. In fact, there is a dependency structure between the responses which make up the database of all configurations/PEV inputs; in other words, the responses for which we are deriving optimal experimental designs. This requires a multivariate distribution underlying a pre-chosen model.

To deal with this characteristic, we consider bivariate DOE for two of the correlated variables in the randomization process, which are the PEVs location and the base typical load profiles. The reason to select these two variables is their obvious dependence based on the customers' behaviours; nonetheless, there are other variables in correlation with each other that have been neglected to make the presentation of the whole idea straightforward. Anyhow, extension of the presented procedure to keep track of more than two correlated variables is practical thanks to the concept of copula. We use copula functions which provide a rich and flexible class of structures to derive joint distributions for bivariate data [3]. Copulas are functions that join multivariate distribution functions to their one-dimensional marginal distribution functions.

Using copulas for modeling purposes includes two straightforward steps: first, the marginal distributions along with their correlation matrix should be modeled; and second, a proper copula should be selected and fitted to the data. This paper proposes an Archimedean copula algorithm based on a Frank copula for case studies. Choosing Frank copula is because the Frank's family permits negative as well as positive dependence. Nonetheless, other types of Archimedean copulas permit only nonnegative correlations because of the limited dependence parameter space [4].

In the next section, the DOE method is described. The modeling needs of applying DOE method to congestion studies are also presented. In addition, the Frank copula is presented in a few words. In conclusion, the proposed method includes four main stages:

1. modeling of uncertainties (database creation),
2. applying multivariate DOE,
3. power flow calculations on the reduced scenarios,
4. and statistical analysis of the results.

Accordingly, it is presented a general perspective of employing bivariate DOE for studying congestion due to proliferation of PEVs in distribution networks; in which the interpretation of the kind of results from the proposed method is illustrated. This short paper is organized to present the theoretical characteristics and background of the proposed method in the EV scenario-generating application with sufficient detail to make it graspable in basics. Therefore, the main focus is not to present a detailed case study (because of the page limit); however, a case study using the recorded data of a real distribution system is available in the last section in brief.

BASICS OF EMPLOYING DOE TECHNIQUE FOR STUDYING ELECTRIC VEHICLES

Fig.1 illustrates a very general model for a process or system. In the presented diagram, output(s) are characteristics as the system's response which could be observed or measured. This could be the load demand in a distribution network due to the electric vehicle battery charging as the system. Controllable variables can be varied during the operational or planning optimization and such variables have a key role to play in the system characterization. Uncontrollable variables are difficult to control; so, they are responsible for variability in system performance or output inconsistency.

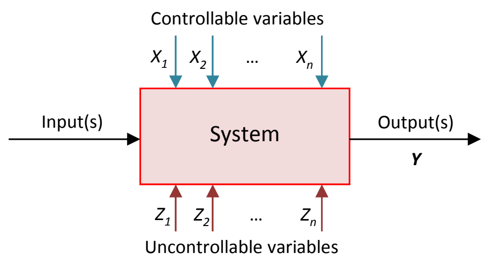

Fig. 1. A very general model of a system

In case of PEVs, one may consider modern tariff structures as controllable variables and battery's state of charge, charging start time and/or its location uncontrollable variables. However, when trying to design a most informative reduced set of scenarios, these variables are better to be treated as controllable variables as well in order to have their part in the final outcome. Indeed, given a variable y and a number of variables $X_1, ..., X_p$ that may be related to y, then such an analysis can be applied to quantify the strength of the relationship between y and the X_j, to assess which X_j may have no relationship with y at all, and to identify which subsets of the X_j contain redundant information about y, thus once one of them is known, the others are no longer informative.

The traditional method of collecting large quantities of data by holding each factor constant in turn until all possibilities have been tested is an approach that quickly becomes impossible as the number of factors increases. Such an increase is about to appear when analyzing PEVs impact. A full factorial scenario design (that is, testing for PEVs' demand at every combination of factors such as location, sate of charge, battery types, charging start time, tariff impacts and etc) is not feasible for extensive power flow calculations. The DOE solves this problem by choosing a set of scenario points that allow estimation of the model with the maximum confidence using just a fraction of the number of scenario runs. These optimally-chosen runs are more than enough to fit the model.

Design of experiments (DOE) is a general technique that could be defined in several ways. One definition is as follows [5]: "A technique to obtain and organize the maximum amount of conclusive information from the minimum amount of work, time, energy, money, or other limited resource." In other terms, it is a strategy to gather empirical knowledge, i.e. knowledge based on the analysis of experimental data and not on theoretical models. It can be applied when investigating a phenomenon in order to gain understanding/modeling. When collecting new data for multivariate modeling, one should pay attention to the *efficiency*, i.e. getting more information from fewer experiments/data and the *focusing*, i.e. collecting only the information that is really needed. Indeed, with today's ever-increasing complexity of models, design of experiment has become an essential part of the modeling process. There are four basic ways to collect data for an analysis:

1. Obtain historical data
2. Collect new data
3. Run specific experiments by disturbing (exciting) the system being studied
4. Design experiments in a structured, mathematical way

More generally, careful sample selection increases the chances of extracting useful information from the data. When one has the possibility to actively perturb the system (making scenarios with the variables), these chances become even greater. The critical part is to decide which variables to change, the intervals for this variation, and the pattern of the experimental points. For example, should we consider charging start time as a variable to change; if so, what intervals for the variation of charging start time should be studied; and what the pattern/distribution of the variation points would be?

Considering the definition by [5], the limited resource here is the computational time required for calculating load flow for all scenarios. In this way, both efficiency and focusing concepts is very helpful for getting a full understanding on how PEVs behave and interact with the network. Furthermore, the data for building initial scenarios is collected from the first and the fourth basic ways above that are using historical data as well as structural or mathematical approach. On the other hand, in order to be able to proceed with DOE technique, a probabilistic model should be fitted the system response. Here, the generalized linear model (GLM) is used. This is because this model works well with the DOE. Also, the GLM presents a regression model that is suitable for modeling PEVs' response in relation to the relevant factors in a discrete distribution framework. The GLM concept is briefly described in the following section.

GENERALIZED LINEAR MODEL (GLM)

In statistics, the generalized linear model (GLM) is a generalization of linear regression by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement to be a function of its expected value [6]. It is very interesting to note that these features are convenient to model the behavior of PEVs; given that a change/shift in the expected value of the total power demand of PEV chargers (maybe due to a shift in timing) correlates with a change in its variance.

On the other hand, modeling PEV demand behavior by a linear or normal regression is simply unrealistic and using a link function would contribute to a much better fit. It should be noted that relying on central limit theorem (CLT) for modeling active network components (e.g. PEVs' response and intermittent renewables) is far more approximate than passive network components (e.g. base load); as a matter of fact such approximate models could even be misleading because the CLT requires all variables to be independent and identical.

In a GLM, each outcome of the dependent variables, **Y**, is assumed to be generated from a particular distribution in the exponential family, a large range of probability distributions that includes the normal, binomial and Poisson distributions, among others. The mean, μ, of the distribution depends on the independent variables, **X**, through [6]

$$E(\mathbf{Y}) = \mu = g^{-1}(\mathbf{X}\beta) \tag{1}$$

where, $E(\mathbf{Y})$ is the expected value of **Y**; $\mathbf{X}\beta$ is the linear predictor, a linear combination of unknown parameters, β and g is the link function. In this framework, the variance is typically a function, V, of the mean

$$\mathrm{Var}(\mathbf{Y}) = V(\mu) = V(g^{-1}(\mathbf{X}\beta)) \tag{2}$$

The unknown parameters, β, are typically estimated with maximum likelihood, maximum quasi-likelihood, or Bayesian techniques.

Accordingly, the GLM consists of three elements:
1. A probability distribution from the exponential family.
2. A linear predictor $\eta = \mathbf{X}\beta$.
3. A link function g such that $E(\mathbf{Y}) = \mu = g^{-1}(\eta)$.

The possibility of using a probability distribution model from the exponential family are considered as a significant advance beyond linear regression models. The exponential family of distributions is a generalization of the exponential family and exponential dispersion model of distributions and includes those probability distributions, parameterized by θ and τ, whose density functions f (or probability mass function, for the case of a discrete distribution) can be expressed in the form [7]

$$f_Y(y \mid \theta, \tau) = h(y, \tau) \exp\left(\frac{\mathbf{b}(\theta)^T \mathbf{T}(y) - A(\theta)}{d(\tau)} \right) \tag{3}$$

τ, called the dispersion parameter, typically is known and is usually related to the variance of the distribution. θ is related to the mean of the distribution. If $\mathbf{b}(\theta)$ is the identity function, then the distribution is said to be in canonical form. The functions $h(y,\tau)$, $\mathbf{b}(\theta)$, $\mathbf{T}(y)$, $A(\theta)$ and $d(\tau)$ are known. Many, although not all, common distributions are in this family.

The second GLM element is a linear predictor. The linear predictor is the quantity which incorporates the information about the independent variables into the model. It is related to the expected value of the data through the link function. η is expressed as linear combinations of unknown parameters β. The coefficients of the linear combination are represented as the matrix of independent variables X. The elements of X are either measured by the experimenters or stipulated by them in the modeling design process. The elements of X are stipulated from tariff and traffic data and by some behavioral forecasts based on the knowledge of planner in case of PEVs.

The link function, on the other hand, is somewhat an arbitrary function. Nevertheless, it should be noted that when using canonical link function the linear predictor may be negative, which would give an impossible negative mean. When maximizing the likelihood, precautions must be taken to avoid this. An alternative is to use a non-canonical link function [6].

MULTIVARIATE DOE BY FRANK COPULA

A design of scenarios/experiments ξ with n support points (to shape the probability distribution of the response) can be written as [2]

$$\xi = \begin{Bmatrix} \boldsymbol{\chi}_1 & \boldsymbol{\chi}_2 & \cdots & \boldsymbol{\chi}_n \\ w_1 & w_2 & \cdots & w_n \end{Bmatrix} \tag{4}$$

where, $\boldsymbol{\chi}_i$ are the support points consisting of the explanatory variables which describe the experimental conditions with weights $w_i \in [0, 1]$ summing to 1. As mentioned in the previous sections, such a design should be made optimal by using some optimization methods. Formulation of this optimization and its indexes are very extensive [2] thus out of the scope of this paper.

Frank Copula

Copula itself is becoming a mature concept while much of its potential applications are still not discovered. The theory of copulas is as comprehensive as a book [4]. Further to what is presented in introduction about copulas in a few words, one may refer to [3] for practical viewpoints and tools. Here, it is just worth to mention the bivariate Frank copula function as it is applied in the congestion study's scenario designing algorithm:

$$C(u_1,u_2;\alpha) = -\alpha^{-1}\log\left(1+\frac{(e^{-\alpha u_1}-1)(e^{-\alpha u_2}-1)}{e^{-\alpha}-1}\right) \quad (4)$$

where, α is the non-zero dependence parameter. Incorporating this copula function in the GLM allows designing a set of optimally reduced scenarios taking into account the dependence between parameters.

CONGESTION STUDY BY USING THE PROPOSED FRAMEWORK

When applying the proposed method to a congestion study under impacts of PEVs, the results could be illustrated as Fig. 2 and Table 1, showing some scenario simulations for five practically correlated feeders/lines in the form of a correlation matrix graph. The DOE is implemented using MATLAB on a 33-bus distribution system test case. A sample base case network for this real distribution area is represented in Fig. 3. The 200 configurations/scenarios have been simulated to put together Fig. 2 as a first round example. The final outcome is about knowing which lines will be simultaneously congested under impacts of PEVs. Both PHEVs and BEVs are included within different scenario simulations. This may help the system operator to take decision in congestion management as well as grid reinforcement.

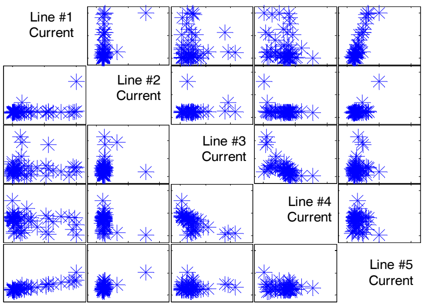

Fig. 2. Some scenario simulations for five practically correlated feeders illustrating their dependence structure under impacts of PEVs.

Table 1. The Rank Correlation Coefficients Together with Confidence Measures (P-values) Corresponding to Fig. 1.

	Line #1	Line #2	Line #3	Line #4	Line #5
Line #1	1.000	0.865 (0.045)	0.172 (0.000)	-0.034 (0.042)	0.903 (0.057)
Line #2		1.000	0.227 (0.004)	0.350 (0.010)	0.005 (0.000)
Line #3			1.000	-0.146 (0.011)	0.202 (0.149)
Line #4				1.000	0.026 (0.000)
Line #5					1.000

Indeed, a correlation analysis applicable to a database of currents in the lines of a distribution system forecast which congestions are correlated. Knowing where congestions will appear in the future is the first step of relevant grid reinforcements. For example, if the system planner has different solutions for the grid reinforcement, he or she could choose the one which removes correlated congestions.

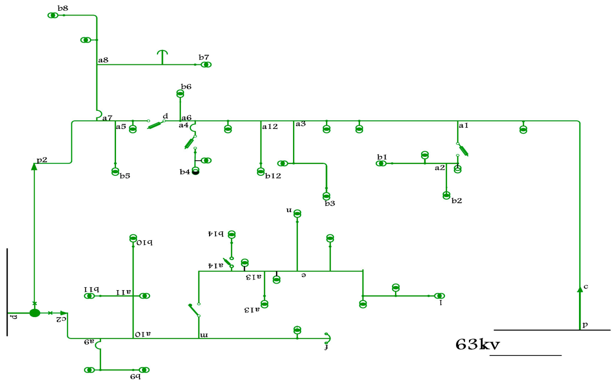

Fig. 3. Base case distribution network for sample analysis of Fig. 1.

CONCLUSIONS

This paper proposes a new statistical framework for modeling the impacts of PEVs in order to make extensive network analyses such as congestion studies feasible and reliable. The proposed statistical methods may be used in congestion studies for either long-term planning or shorter look-ahead time horizons as congestion management by giving the system operator a global view of the grid. For instance, if the system has to be reconfigured to clear congestion in a line, the fact that this line is correlated in inverse with another one has to be taken into account. By knowing the correlated congestions, the system operator may find the most efficient management without studying each line of the grid separately. Furthermore, it should be noted that this paper, besides presenting a congestion study, proposes a technique to take into account the impacts of PEVs in other types of studies through a reduced but more accurate scenario set.

REFERENCES

[1] A. Maitra, J. Taylor, D. Brooks, M. Alexander, and M. Duvall, 2009, "Integrating plug-in electric vehicles with the distribution system," *Proceedings 20th Int. Conf. Electricity Distribution (CIRED)*, Prague, Czech Republic, 8–11.

[2] N.G. Denman, J.M. McGree, J.A. Eccleston, S.B. Duffull, 2011, "Design of experiments for bivariate binary responses modelled by Copula functions," *Comput. Stat. Data An.*, Vol. 55, 1509-1520.

[3] H. Valizadeh Haghi, M. Tavakoli Bina, M.A. Golkar, and S.M. Moghaddas Tafreshi, 2010, "Using Copulas for analysis of large datasets in renewable distributed generation: PV and wind power integration in Iran," *Renew. Energy*, vol. 35, 1991–2000.

[4] R.B. Nelsen, 2006, *An Introduction to Copulas*, 2nd ed., Springer, New York, USA.

[5] L.W. Condra, 2001, *Reliability improvement with design of experiments*, 2nd ed., Marcel Dekker, USA.

[6] P. McCullagh, J. Nelder, 1989, *Generalized Linear Models*, 2nd ed. Chapman and Hall/CRC, Boca Raton.

[7] D.R. Clark, A.T. Charles, 2004, "A Primer on the Exponential Family of Distributions," *CAS Discussion Program*, 117-148.