Data-driven Global Sensitivity Analysis of Three-Phase Distribution System with PVs

Ketian Ye, Junbo Zhao, Can Huang, Nan Duan, Fei Ding, Rui Yang

***Abstract*—Global sensitivity analysis (GSA) of distribution system with respect to stochastic PV variations plays an important role in designing optimal voltage control schemes. This paper proposes a Kriging , i.e., Gaussian process modeling enabled data-driven GSA method. The key idea is to develop a surrogate model that captures the hidden global relationship between voltage and real and reactive power injections from the historical data. With the surrogate model, the Sobol index can be conveniently calculated to assess the global sensitivity of voltage to various power injection variations. Comparison results with other model-based GSA methods on the IEEE 37-bus feeder, such as the polynomial chaos expansion and the Monte Carlo approaches demonstrate that the proposed method can achieve accurate GSA outcomes while maintaining high computational efficiency.**

***Index Terms*—Distribution system analysis, global sensitivity analysis, Sobol indices, Gaussian process, PVs.**

I. Introduction

With the increased penetration of stochastic and uncertain solar PVs into the distribution systems, there is an emergent concern about the voltage security. Thus, understanding how voltage variations are affected by those stochastic and uncertain resources is important for designing appropriate control schemes. Sensitivity analysis of voltage to uncertain power injections allows us to effectively quantify these effects. Sensitivity analysis includs the local sensitivity analysis (LSA) [1]–[4] and global sensitivity analysis (GSA) [5] methods. For LSA, it mainly focuses on the local impacts from uncertain inputs. The Jacobian matrix is widely used for LSA in power systems. In [1], the Jacobian matrix-based LSA is used for distribution network voltage management. The measurement-based Jacobin matrix estimation for voltage to power sensitivity analysis is also developed in [3], [4]. However, for voltage control, LSA could not reveal the global impacts of various control devices on the voltage changes. To this end, GSA is needed [6]. By establishing sensitivity indices in covering the entire input space, GSA provides more accurate and comprehensive information of the global relationship between inputs and outputs. GSA methods include the Morris method, the Sobol indices, and the Kucherenko indices [7]. Among them, Sobol indices (SI) are widely used in the variance-based analysis [8]. The main idea of SI is to decompose the model into summands that satisfy the orthogonality condition. Then, the influence of the variability of the input on the model response can be conveniently quantified [9]. Note that the Monte Carlo (MC) simulations are usually used for SI calculation, which is time-consuming.

This work was supported by the U.S. Department of Energy through Advanced Grid Modeling program. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. This work was supported by National Science Foundation under grant ECCS 1917308 and Department of Energy Advanced Grid Modernization program. K. Ye and J. Zhao are with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA. C. Huang and N. Duan are with Lawrence Livermore National Laboratory, Livermore, 94550 CA. F. Ding and Y. Rui are with National Renewable Energy Laboratory, Golden, CO 80401 USA. (e-mail: ky291@msstate.edu, junbo@ece.msstate.edu).

To deal with the computational burden of MC-based SI calculations, reduced order model for the original one is developed. This can be achieved using the surrogate modeling techniques [10], [11]. In [5], the polynomial chaos expansion (PCE) based surrogate modeling approach is developed for distribution system GSA. The PCE-based technique is also used for probabilistic power flow [12], [13]. But the calculation of coefficients of PCE requires accurate physical model. This is very difficult to achieve for practical distribution system, especially with the increased penetration of renewable energy, flexible loads and advanced demand response program [14]. It is worth pointing out that LSA has been widely used for distribution system while little efforts have been done for GSA. This paper aims to bridge this gap and proposes a novel data-driven GSA method for distribution system with PVs.

This paper has the following contributions:

- It is data-driven and is not affected by the model and input uncertainties. This is achieved by developing the Kriging-based surrogate model to interact with the SI and the original power flow model is not required;
- It is much more computational efficiency than the MC simulations-based SI calculations without the loss of accuracy. This is because the Kriging-based surrogate model is much cheaper to evaluate as compared to the original complicated physical models;
- It can reveal the global voltage variations to uncertain real and reactive power injections considering the three-phase couplings. The impacts of PV injections with different capacities and distributions are investigated. The proposed method is also robust to measurement noise.

The reminder of the paper is organized as follows: Section II shows the problem statement. Section III illustrates the proposed framework. Section IV analyzes the simulation results and finally Section V concludes the paper.

II. Problem Statement

Let $\boldsymbol{y} = \mathcal{M}(\boldsymbol{x})$ the model with random input vector $\boldsymbol{x} = [x_1, \ldots, x_d]^T$ and its response $\boldsymbol{y}$. In the three-phase distribution systems, the uncertain inputs may include PV injections and loads while the outputs are typically bus voltage

2021 IEEE Power & Energy Society General Meeting (PESGM) | 978-1-6654-0507-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/PESGM46819.2021.9638171

magnitudes V, voltage angles θ, line power flows P_f, etc. Due to the uncertainties from PVs, the power flow results are subject to uncertainties as well. Sensitivity analysis aims to quantify how the power flow model response is affected by each uncertain PV input and their combinations. With the sensitivity analysis outcomes, several practical applications can be achieved, such as voltage regulation, PV inverter control and network partition [1]–[4]. To estimate the sensitivity at the operating point $\tilde{\boldsymbol{x}}$, the corresponding partial derivative $(\partial \boldsymbol{y}/\partial \boldsymbol{x})_{\boldsymbol{x}=\tilde{\boldsymbol{x}}}$ is commonly used for LSA. As for distribution systems, the sensitivity information can be derived from Jacobian matrix. Via the linearization on branch flow equations, [1] formulates the sensitivity of voltage magnitude with respect to power injections as

$$\Delta|V| \approx \sum_i \left(\frac{\partial|V|}{\partial P_i} \times \Delta P_i + \frac{\partial|V|}{\partial Q_i} \times \Delta Q_i\right) \tag{1}$$

where $\partial|V|/\partial P_i$ and $\partial|V|/\partial Q_i$ numerically measure the sensitivity of voltage variations to real and reactive power changes. However, LSA can only capture the local information with individual components and may not reflect the true sensitivity of model response to inputs. The derivatives in Jacobian matrix only consider the buses that are directly connected. By contrast, GSA reveals the global variations of outputs to inputs. One of the widely used GSA methods is the Sobol index-based technique. Sobol indices measure the contribution of uncertain sources and their interactions to predictive uncertainty of model response. The calculations of Sobol indices are often realized via MC simulations that are computational expensive. This calls for the development of more computational efficient methods. In this paper, we develop a data-driven surrogate model $\mathcal{M}'(\boldsymbol{x})$ of the original model $\mathcal{M}(\boldsymbol{x})$ to achieve significant improvement on computational efficiency.

III. Proposed Data-driven GSA Approach

The proposed method consists of two procedures: surrogate modeling and Sobol indices calculation. Specifically, the data-driven Kriging is leveraged to build the surrogate model that captures the mapping relationship between voltage and real and reactive power injections. After that, the surrogate model is used to calculate Sobol indices with MC approach with improved computational efficiency. In this section, we first present the theory of Sobol indices for GSA, followed by the data-driven calculations of them.

A. Sobol Indices

Assume the inputs follow independently uniform distribution with support $\mathcal{D}_x = [0,1]^d$. Based on the idea of decomposing the model with respect to variance, the analysis of variance (ANOVA)-representation of $\mathcal{M}(\boldsymbol{x})$ is defined as:

$$\mathcal{M}(x_1,\ldots,x_d) = \mathcal{M}_0 + \sum_{i=1}^{d} \mathcal{M}_i(x_i) + \sum_{1\le i<j\le d} \mathcal{M}_{ij}(x_i,x_j) + \cdots + \mathcal{M}_{1,\ldots,d}(x_1,\ldots,x_d) \tag{2}$$

under the condition that $\int_0^1 \mathcal{M}_{i_1,\ldots,i_s}(x_{i_1},\ldots,x_{i_s})\,dx_{i_k} = 0$ for $1 \le k \le s, 1 \le s \le d$, where we have

$$\begin{cases} \mathcal{M}_0 = \int_{\mathcal{D}_x} \mathcal{M}(\boldsymbol{x})\,d\boldsymbol{x} \\ \mathcal{M}_i(x_i) = \int_{\mathcal{D}_x^{d-1}} \mathcal{M}(\boldsymbol{x})\,d\boldsymbol{x}_{\sim i} - \mathcal{M}_0 \\ \mathcal{M}_{ij}(z_i,z_j) = \int_{\mathcal{D}_x^{d-2}} \mathcal{M}(\boldsymbol{x})\,d\boldsymbol{x}_{\sim(i,j)} - \mathcal{M}_0 - \mathcal{M}_i - \mathcal{M}_j \end{cases} \tag{3}$$

where $\boldsymbol{x}_{\sim i}$ denotes the subset of $\boldsymbol{x}$ that excludes ith variable x_i and $\mathcal{D}_x^{d-1}$ is the corresponding support. Assuming $\mathcal{M}(\boldsymbol{x})$ is square-integrable, [9] indicates that by squaring (2) and integrating over $\mathcal{D}_x$, one can get

$$\int_{\mathcal{D}_x} \mathcal{M}^2(\boldsymbol{x})\,d\boldsymbol{x} - \mathcal{M}_0^2 = \sum_{s=1}^{d} \sum_{i_1<\cdots<i_s}^{d} \int \mathcal{M}^2_{i_1,\ldots,i_s}\,dx_{i_1}\cdots dx_{i_s}$$

According to the expression of variance, the left-hand side denotes the variance of model response. Thus, we get [10]

$$var(Y) = V = \sum_{s=1}^{d} \sum_{i_1<\cdots<i_s}^{d} \int \mathcal{M}^2_{i_1,\ldots,i_s}\,dx_{i_1}\cdots dx_{i_s} \tag{4}$$

Using the orthogonality condition, the variance of model response can be decomposed and (4) becomes

$$V = \sum_{i=1}^{d} V_i + \sum_{i<j}^{d} V_{ij} + \cdots + V_{1,\ldots,d} \tag{5}$$

Then, the Sobol indices are defined as $S_I = V_I/V$, where $I \subset \{1,\ldots,d\}$. Sobol indices represent the contribution of input components to the output variance. First-order Sobol index S_i measures the effect of individual input while higher-order Sobol indices $S_{i_1\ldots i_s}$ reveal the contribution by the interaction of $x_{i_1},\ldots x_{i_s}$. To further determine the general importance of each input, the total Sobol index is defined as $S_i^T = \sum_{\{i_1,\ldots,i_s\}\supset i} S_{i_1\ldots i_s}$. The calculation of Sobol indices can be achieved via MC-based approach via

$$\begin{cases} \widehat{V}_0 = \frac{1}{N}\sum_{n=1}^{N} \mathcal{M}(\boldsymbol{x}^{(n)}) \\ \widehat{V} = \frac{1}{N}\sum_{n=1}^{N} \mathcal{M}^2(\boldsymbol{x}^{(n)}) - \widehat{V}_0^2 \\ \widehat{V}_i = \frac{1}{N}\sum_{n=1}^{N} \mathcal{M}(x_i^{(n)}, \boldsymbol{x}_{\sim i}^{(n)})\mathcal{M}(x_i^{(n)}, \boldsymbol{x}'^{(n)}_{\sim i}) - \widehat{V}_0^2 \end{cases} \tag{6}$$

where $\boldsymbol{x}_{\sim i}^{(n)}$ is the observation with ith input variable excluded; $\boldsymbol{x}'$ is another realization that is independent with $\boldsymbol{x}$.

B. Proposed Data-Driven Sobol Indices

The Sobol indices are typically calculated based on MC simulations that rely on the original physical power flow models. Reduced order model to speed up the process can be used as well [5]. However, the accurate physical model is assumed, which is challenging to achieve for a practical distribution systems. To deal with that, data-driven surrogate model via Kriging is developed for Sobol indices calculations.

Kriging assumes that the original model is an observation of a Gaussian process (GP), i.e.,

$$\mathcal{M}_{kr}(\boldsymbol{x}) = m(\boldsymbol{x}) + Z(\boldsymbol{x}; \sigma^2, \boldsymbol{\theta}) \tag{7}$$

where $m(\boldsymbol{x}) = \boldsymbol{\beta}^T \boldsymbol{f}(\boldsymbol{x})$ represents the mean function or the trend; $Z(\boldsymbol{x}; \sigma^2, \boldsymbol{\theta})$ is a centered GP with zero mean, variance σ^2, and covariance kernel function $k(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\theta})$. The mean function $m(\boldsymbol{x})$ gives kriging predictor for the model response. The kriging variance $Z(\boldsymbol{x};\sigma^2,\boldsymbol{\theta})$ quantifies the uncertainty at the corresponding point. $\boldsymbol{f}(\boldsymbol{x})$ is typically prescribed while parameters $\boldsymbol{\beta},\sigma^2$, and $\boldsymbol{\theta}$ need to be estimated, such as via the maximum likelihood (ML) estimator. Assume the Kriging model is based on observation $\boldsymbol{X}=\{\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(N)}\}$ with N samples and response $\boldsymbol{y}=\{\mathcal{M}(\boldsymbol{x}^{(1)}),\ldots,\mathcal{M}(\boldsymbol{x}^{(N)})\}$ with pre-determined trends that consist l functions $\{f_j, j=1,\ldots,l\}$. Then, the unknown parameters are estimated using ML estimator [15], [16]:

$$\begin{cases} \widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}(\boldsymbol{\theta}) = (\boldsymbol{F}^T\boldsymbol{K}(\boldsymbol{\theta})^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}^T\boldsymbol{K}(\boldsymbol{\theta})^{-1}\boldsymbol{y} \\ \widehat{\sigma^2} = \sigma^2(\boldsymbol{\theta}) = \frac{1}{N}(\boldsymbol{y}-\boldsymbol{F}\hat{\boldsymbol{\beta}}^T\boldsymbol{K}(\boldsymbol{\theta})^{-1}(\boldsymbol{y}-\boldsymbol{F}\widehat{\boldsymbol{\beta}})) \\ \widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \widehat{\sigma^2}[\det \boldsymbol{K}(\boldsymbol{\theta})]^{1/m} \end{cases} \tag{8}$$

where $\boldsymbol{F}$ is the regression matrix with elements $\{F_{ij}=f_j(\boldsymbol{x}), i=1,\ldots,N, j=0,\ldots,l\}$ and $\boldsymbol{K}(\boldsymbol{\theta})$ is the correlation matrix with unknown hyperparameter $\boldsymbol{\theta}$.

In the prediction step, the model response is approximated by employing the covariance function $k(\boldsymbol{x},\boldsymbol{x'};\boldsymbol{\theta})$ that measures the closeness of the inputs. Specifically, the prediction $\tilde{y}$ is $\tilde{y}\sim\mathcal{N}(\mu_{\tilde{y}}(\tilde{\boldsymbol{x}},\sigma^2_{\tilde{y}}(\tilde{\boldsymbol{x}}))$ with the parameters

$$\begin{cases} \mu_{\tilde{y}} = \boldsymbol{f}^T\widehat{\boldsymbol{\beta}} + \tilde{\boldsymbol{k}}\boldsymbol{K}^{-1}(\boldsymbol{y}-\boldsymbol{F}\widehat{\boldsymbol{\beta}}) \\ \sigma^2_{\tilde{y}} = \widehat{\sigma^2}(1-\tilde{\boldsymbol{k}}\boldsymbol{K}^{-1})\tilde{\boldsymbol{k}} + \boldsymbol{u}^T(\boldsymbol{F}^T\boldsymbol{K}^{-1}\boldsymbol{F}^{-1})\boldsymbol{u} \end{cases} \tag{9}$$

where $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma^2}$ are the generalized least-squares estimates from (8); $\tilde{\boldsymbol{k}}$ is the cross-correlation between $\tilde{\boldsymbol{x}}$ and $\boldsymbol{x}$ with components $\tilde{k}_i = k(\tilde{\boldsymbol{x}},\boldsymbol{x}^{(i)}), i=1,\ldots,N$; $\boldsymbol{K}$ is the correlation matrix of $\boldsymbol{X}$ with elements $K_{ij}=k(\boldsymbol{x}^{(i)},\boldsymbol{x}^{(j)}), i,j=1,\ldots,N$; $\boldsymbol{u}=\boldsymbol{F}^T\boldsymbol{K}^{-1}\tilde{\boldsymbol{k}}-\boldsymbol{f}$. Note that the common trends are linear and quadratic [15] while the kernel functions include exponential kernel, Gaussian kernel, Matérn kernel, etc.

Sobol indices calculation with Kriging: The construction of Kriging model is completed with (8) and the model predictions can be obtained via (9). After that, the MC-based Sobol indices for the voltages can be calculated by replacing $\mathcal{M}$ with $\mathcal{M}_{kr}$ in (6).

TABLE I
LOAD SHAPE AND PV DISTRIBUTIONS IN DIFFERENT SCENARIOS

Scenarios	$P_{inj}(kW)$	$P_{load}(kW)$
Scenario 1	$20\times Beta(3,2.2)$	$\mathcal{N}(P^i_{load},(0.05P^i_{load})^2)$
Scenario 2	$30\times Beta(3,2.2)$	$\mathcal{N}(P^i_{load},(0.05P^i_{load})^2)$
Scenario 3	$30\times Beta(3,2.2)$	$\mathcal{N}(P^i_{load},(0.1P^i_{load})^2)$
Scenario 4	$Weibull(15,3)$	$\mathcal{N}(P^i_{load},(0.1P^i_{load})^2)$

IV. NUMERICAL RESULTS

Numerical results carried out on the modified IEEE 37-bus system considering PVs are used to demonstrate the performance of the proposed method. The schematic is shown in Fig. 1. The inputs include random variations of loads and PV injections while the model responses, voltage magnitudes are used for illustrations. The model-based PCE [5] and the proposed data-driven methods are compared with the benchmark that uses the MC simulations. The relative mean absolute error (RMAE) and root mean square error (RMSE) are used to quantify the prediction accuracy of the model and Sobol indices calculation accuracy, respectively. The overall measurement of surrogate model accuracy is the average value of all RMAEs, denoted as e_M. Similarly, the overall accuracy index for Sobol indices is e_{SI}. All simulations are carried out in MATLAB with 2.60 GHz Intel Core i7-6700HQ.

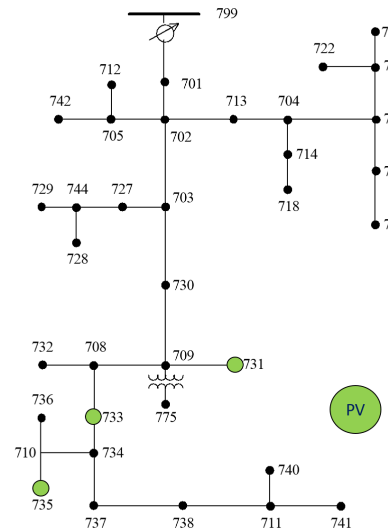


Fig. 1. IEEE 37-bus system with PVs.

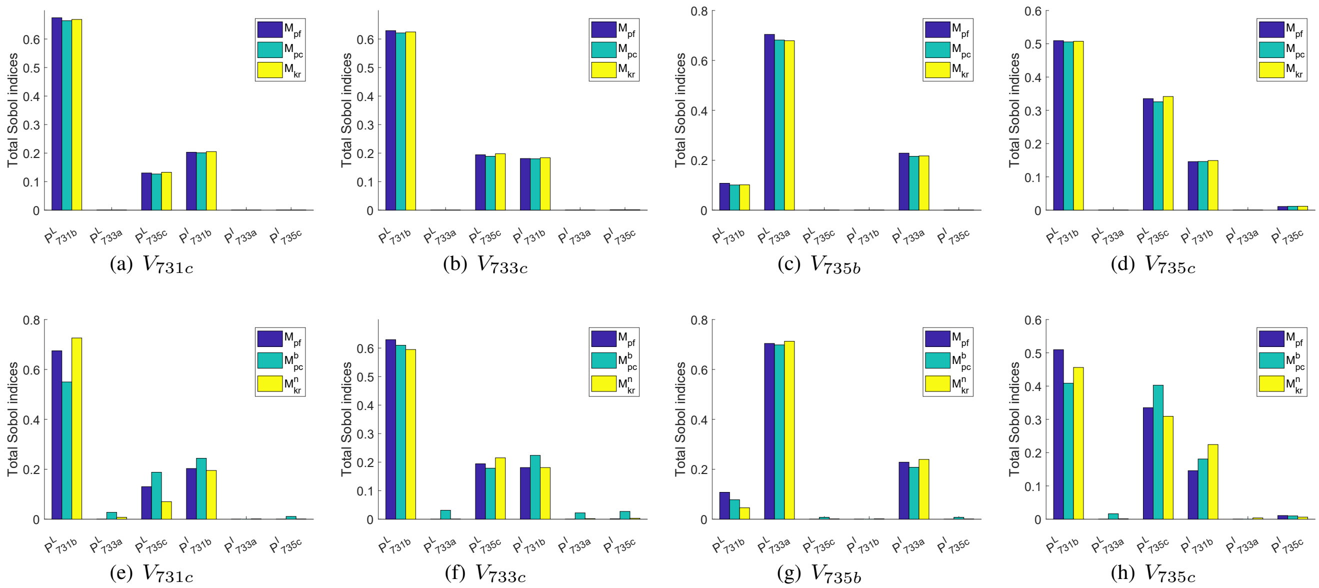
The AC power flow model $\mathcal{M}_{pf}$ is used as the benchmark on which the Sobol indices S^{pf} is obtained via MC simulations. Meanwhile, the inputs $\boldsymbol{X}$ are generated from known distributions and the corresponding model responses $\boldsymbol{y}$ are obtained though $\mathcal{M}_{pf}$. In this paper, all realizations are calculated from OpenDSS [17]. Then, the generated data $\{\boldsymbol{X},\boldsymbol{y}\}$ are used to construct two surrogate models $\mathcal{M}_{pc}$ and $\mathcal{M}_{kr}$ for PCE and proposed Kriging method. Their corresponding Sobol indices are S^{pc} and S^{kr}.

TABLE II
COMPARISON RESULTS OF DIFFERENT MODELS ON 37-BUS FEEDER

Model	Accuracy		CPU time (s)
	$e_M(\times 10^{-5}\%)$	$e_{SI}(\times 10^{-2})$	
$\mathcal{M}_{pf}$	—	—	103
$\mathcal{M}_{qmc}$	—	0.950	61.167
$\mathcal{M}_{pc}$	3.825	0.869	1.243
$\mathcal{M}_{kr}$	3.713	0.811	4.854

A. Sobol Indices Validation

The inputs include six random variables $\boldsymbol{x}=[x_1,\ldots,x_6]$, where $[x_1,x_2,x_3]$ are loads at nodes $[731b,733a,735c]$ and $[x_4,x_5,x_6]$ are power injections from PVs at nodes $[731b,733a,735c]$ respectively. Note that a, b and c are different phases. The corresponding model responses are $\boldsymbol{y}=[y_1,\ldots,y_9]$ whose elements are three-phase voltage magnitudes. The distributions of inputs in Scenario 3 of Table I are used here for illustration. The variances for load distributions are assumed to be $\boldsymbol{\sigma}_L=0.1\boldsymbol{\mu}_L$. PCE and Kriging



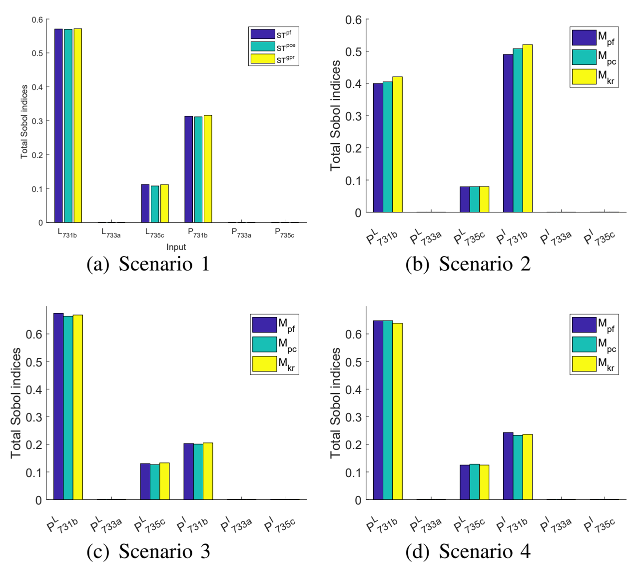
(a) V_{731c} (b) V_{733c} (c) V_{735b} (d) V_{735c}

(e) V_{731c} (f) V_{733c} (g) V_{735b} (h) V_{735c}

Fig. 2. (a)-(d) correspond to the case that there is no error with system model while (e)-(h) are the results considering system model error.

are constructed with $N = 200$ samples, where the degree of PCE is set to $n = 2$ and the linear trend and Gaussian kernel are used for Kriging. Table II shows the comparison results of different methods, where $\mathcal{M}_{qmc}$ stands for quasi-MC method and CPU time denotes the Sobol indices calculation time with 1000 samples. It can be found that the enhanced MC method improves the computational efficiency with slight loss of accuracy. Both PCE and Kriging methods can achieve comparable accuracy as the MC and its enhanced one while being much computational efficiency via the surrogate models. This validates the effectiveness of our data-driven method. Figs.2 (a)-(d) demonstrate the total Sobol indices for V_{731c}, V_{733c}, V_{735b} and V_{735c} by three models. It is observed that only three inputs have noticeable effects on V_{731c} and the dominant factors are load and PV injections at phase b. Results also show that voltage magnitudes at the same phase share a similar pattern of Sobol indices. According to the total index, load fluctuation of 10% has more impacts than PV injections with power rating 30kW in terms of sensitivity.

In practice, the distribution system model is always subject to errors and there are also errors for the measurements. To test the robustness of these methods when the model is subject to errors, we assume there are uncertainties of distribution line parameters. They are assumed to follow independent Gaussian distribution $\mathcal{N}_{LL} \sim (0, (0.05\boldsymbol{\mu}_{LL})^2)$. For input observation $\boldsymbol{X}$, Gaussian noise with zero mean and standard derivations $\boldsymbol{\sigma}_{nx} = 0.01\boldsymbol{\mu}_x$ is added. For data-driven approach, measurement error is introduced for model response with a variance $\boldsymbol{\sigma}_{ny} = 0.01\%\boldsymbol{\mu_y}$. Fig. 2 (e)-(h) display the results and it can be found that the proposed method is only slightly affected. By contrast, due to model errors, the model-based PCE yields much larger errors in Sobol indices, see (e) for example.



(a) Scenario 1 (b) Scenario 2

(c) Scenario 3 (d) Scenario 4

Fig. 3. Sobol indices under different scenarios for $V731c$.

TABLE III
PERFORMANCE OF KRIGING IN DIFFERENT SCENARIOS

Scenarios	**SI error** $e_{SI}(\times 10^{-2})$	**CPU time** $T_{SI}(s)$
Scenario 1	0.684	0.520
Scenario 2	1.331	0.612
Scenario 3	0.811	0.523
Scenario 4	0.960	0.546

B. Sensitivity in Different Scenarios

Further tests, including different PV injections and distributions are used to demonstrate the proposed method, see Table I. Since the conclusions for PCE are similar as previous section, only the proposed method is used for illustrations. Fig. 3 shows the results for different scenarios. From Scenarios 1 and 2, it is observed that the increase of power injections enlarges the Sobol indices of input P_{731b} as expected. The comparisons of Scenarios 2 and 3 show the similar trend in terms of increased load uncertainty. This is because the variation of distribution results in a similar pattern of changing the fluctuation ranges of loads and PV injections. The results in Scenarios 3 and 4 are similar since the parameter setting of PV injections in Scenario 4 yields similar distribution as in Scenario 3. Table III demonstrates the model accuracy under different scenarios, justifying the high accuracy and computational efficiency. These results also demonstrate our proposed method is able to track the sensitivity changes and reveal the complicated global sensitivity relationships between changing inputs and outputs.

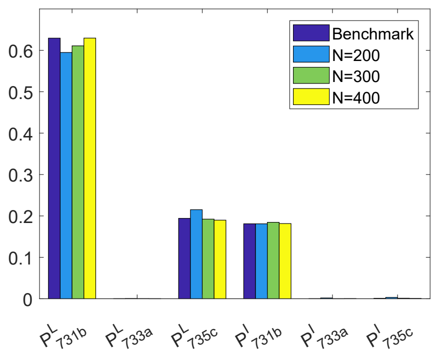


Fig. 4. Robustness to increased measurement noise for V_{735c} sensitivity.

C. Robustness to Noise

This case study is to further assess the robustness of the proposed method using enhanced samples to reduce the impact of noise. Both inputs and outputs are subject to noise $\{\boldsymbol{\sigma}_{nx}, \boldsymbol{\sigma}_{ny}\}$ as discussed in IV-A. The Kriging method are able to handle additive Gaussian noise theoretically [15]. Fig. 4 demonstrates the robustness of the proposed method to noise with moderately increased number of samples, where the V_{735c} sensitivity is used for illustration. Compared with the results shown in Section IV-A, we find that to achieve similar estimation accuracy due to the increased noise level, the number of data samples should be increased. This is expected as handling measurement noise requires a better redundancy. Therefore, in practical applications, the trade-off between robustness to noise and the use of appropriate number of samples should be paid attention. Another mitigation strategy is to pre-process the measurement and filter out the noise, such as the principle component analysis (PCA) or kernel PCA [18].

V. Conclusion

In this paper, a data-driven GSA approach is proposed for three-phase distribution systems with stochastic and uncertain PVs. GSA allows us to quantify the overall effects of uncertain model inputs on model response, i.e., the voltage variations to PVs and loads. The proposed method has two key components, namely the surrogate modeling via data driven Kriging and Sobol indices calculation using MC simulations based on the constructed Kriging model. As a non-parametric method, Kriging enjoys the benefits over parametric approaches, such as requiring less stringent assumptions, and expensive simulations, flexible with prior knowledge embedding. Simulation results on the unbalanced IEEE 37-bus system show that our data driven approach can achieve similar accuracy as the MC simulations but being much more computational efficient. As compared to model-based PCE, it is not affected by model errors and the results reveal that our proposed method is robust to measurement noise. Future work will be on developing closed-loop voltage control algorithm utilizing the global voltage sensitivity analysis outcomes.

References

[1] F. Tamp and P. Ciufo, "A sensitivity analysis toolkit for the simplification of MV distribution network voltage management," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 559-568, Mar. 2014.

[2] H. Xu, A. Dominguez-Garcia, P. W. Sauer, "A data-driven voltage control framework for power distribution systems," *IEEE Power & Energy Society General Meeting*, Portland, OR, 2018, pp. 1-5.

[3] H. Su, *et al.*, "Novel voltage-to-power sensitivity estimation for phasor measurement unit-unobservable distribution networks based on network equivalent," *Applied Energy*, vol. 250, pp. 302-312, 2019.

[4] S. Nowak, C. Chen, L. Wang, "Measurement-based optimal DER dispatch with a recursively estimated sensitivity model," *IEEE Trans. Power Syst*, vol. 35, no. 6, pp. 4792-4802, Nov. 2020.

[5] F. Ni, *et al.*, "Variance-based global sensitivity analysis for power systems," *IEEE Trans. Power Syst*, vol. 33, no. 2, pp. 1670-1682, 2018.

[6] B. Iooss, *et al.*, "A review on global sensitivity analysis methods," arXiv:1404.2405 [math.ST], Apr. 2014.

[7] M. Ye, M. C. Hill, "Global sensitivity analysis for uncertain parameters, models, and scenarios," in *Sensitivity analysis in earth observation modelling*, vol. 10, G. P. Petropoulos and P. K. Srivastava, Ed. Elsevier, 2017, pp. 177-210.

[8] X. Liao, *et al*, "Extended affine arithmetic-based global sensitivity analysis for power flow with uncertainties," *Int. J. Electr. Power Energy Syst.*, vol 115, Feb. 2020.

[9] I. M. Sobol, "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates," *Math. Comput. Simulat.*, vol. 55, iss 1-3, pp. 271-280, Feb. 2001.

[10] L. L. Gratiet, S. Marelli, B. Sudret, "Metamodel-based sensitivity analysis: polynomial chaos expansions and Gaussian processes," arXiv:1606.04273 [stat.CO], Jun. 2016.

[11] S. Balduin, T. Westermann, E. Puiutta, "Evaluating different machine learning techniques as surrogate for low voltage grids," arXiv:2006.12389 [eess.SP], Jun. 2020.

[12] L. Yan, *et al.*, "Gaussian processes and polynomial chaos expansion for regression problem: linkage via the RKHS and comparison via the KL divergence," *Entropy*, vol. 20, iss. 3, Mar. 2018.

[13] Y. Xu *et al.*, "Probabilistic power flow based on a Gaussian process emulator," *IEEE Trans. Power Syst.*, vol. 35, no. 4, pp. 3278-3281, Jul. 2020.

[14] P. Forsyth, O. Nzimako, C. Peters and M. Moustafa, "Challenges of modeling electrical distribution networks in real-time," *2015 International Symposium on Smart Electric Distribution Systems and Technologies (EDST)*, pp. 556-559, Vienna, Sept. 2015.

[15] C. Lataniotis, S. Marelli, B. Sudret, UQLab user manual – Kriging (Gaussian process modelling), Report UQLab-V0.9-105, Chair of Risk, Safety & Uncertainty Quantification, ETH Zurich, 2015.

[16] B. Iooss, "Uncertainty and sensitivity analysis of functional risk curves based on Gaussian processes," *Reliab. Eng. Syst. Saf.*, vol. 187, pp. 58-66, Jul. 2019.

[17] Electric Power Research Institute, Distribution System Simulator, OpenDSS, avialable at: https://sourceforge.net/projects/electricdss

[18] L. Zhao, Y. Liu, J. B. Zhao, Y. Zhang, L. Xu, J. Liu, "Robust PCA-deep belief network surrogate model for distribution system topology identification with DERs", *International Journal of Electrical Power & Energy Systems*, vol. 125 (106441), 2021.