

A Data-Driven Global Sensitivity Analysis Framework for Three-Phase Distribution System With PVs

Ketian Ye , Junbo Zhao , Senior Member, IEEE, Can Huang , Senior Member, IEEE, Nan Duan , Senior Member, IEEE, Yingchen Zhang , Senior Member, IEEE, and Thomas E. Field , Senior Member, IEEE

Abstract—Global sensitivity analysis (GSA) of distribution system with respect to stochastic PV and load variations plays an important role in designing optimal voltage control schemes. This paper proposes a data-driven framework for GSA of distribution system. In particular, two representative surrogate modeling-based approaches are developed, including the traditional Gaussian process-based and the analysis of variance (ANOVA) kernel ones. 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 indices can be conveniently calculated through either the sampling-based method or the analytical method to assess the global sensitivity of voltage to variations of load and PV power injections. The sampling-based method estimates the Sobol indices using Monte Carlo simulations while the analytical method calculates them by resorting to the ANOVA expansion framework. Comparison results with other model-based GSA methods on the unbalanced three-phase IEEE 37-bus and 123-bus distribution systems show that the proposed framework can achieve much higher computational efficiency with negligible loss of accuracy. The results on a real 240-bus distribution system using actual smart meter data further validate the feasibility and scalability of the proposed framework.

Index Terms—ANOVA kernel, distribution system analysis, Gaussian process, PVs, global sensitivity analysis, sobol indices.

Manuscript received November 5, 2020; revised February 3, 2021; accepted March 20, 2021. Date of publication March 25, 2021; date of current version August 19, 2021. This work was supported in part by National Science Foundation under Grant ECCS 1917308, in part by the Department of Energy Advanced Grid Modernization program, in part by the U.S. Department of Energy by Lawrence Livermore National Laboratory under Grant DE-AC52-07NA27344, and in part by National Renewable Energy Laboratory (NREL), operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Grant DE-AC3608GO28308. Paper no. TPWRS-01833-2020. (Corresponding author: Junbo Zhao.)

Ketian Ye and Junbo Zhao are with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA (e-mail: ky291@msstate.edu; junbo@ece.msstate.edu).

Can Huang and Nan Duan are with Lawrence Livermore National Laboratory, Livermore, CA 94550 USA (e-mail: can7huang@gmail.com; duan4@llnl.gov). Yingchen Zhang is with National Renewable Energy Laboratory, Golden, CO 80401 USA (e-mail: yingchen.zhang@nrel.gov).

Thomas E. Field is with Entergy Corporation, Jackson, MS 39213 USA (e-mail: tfield@entergy.com).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TPWRS.2021.3069009>.

Digital Object Identifier 10.1109/TPWRS.2021.3069009

NOMENCLATURE

Abbreviations

ANOVA	Analysis of variance
GP	Gaussian process
GSA	Global sensitivity analysis
LSA	Local sensitivity analysis
MC	Monte Carlo
PCE	Polynomial chaos expansion
MAPE	Mean absolute percentage error
RMSE	Root mean square error
SI	Sobol index

Mathematical Symbols

$\beta, \hat{\beta}$	Coefficients of mean function and their estimates
$\theta, \hat{\theta}$	Kernel parameter and its estimate
F	Regression matrix of f
$f(\cdot)$	Basis of mean function
K	Covariance matrix
$k(\cdot)$	Kernel function
$\sigma^2, \hat{\sigma}^2$	Variance of Gaussian process and its estimate
$\tilde{k}$	Cross-variance
$m(\cdot)$	mean function or trend of Gaussian process

Other Symbols

N	Number of samples
n	PCE degree
P^I	Power injection
P^L	Loads

Sobol Indices

$\mathcal{M}$	Original model
$\mathcal{M}_{kA}$	ANOVA kernel-based Kriging surrogate model
$\mathcal{M}_{kr}$	Kriging surrogate model
$\mathcal{M}_{pc}$	PCE surrogate model
$\mathcal{M}_{pf}$	Power flow model
e_M	Model prediction error (using MAPE)
e_{SI}	Sobol indices approximation error (using RMSE)
S	Sobol indices
S^T	Total Sobol indices
V	Variance of output
x'	Another observation that is independent with x
$x_{\sim i}$	Observation of x with i th input variable excluded

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 the sensitivity of voltage variations due to those stochastic and uncertain resources is important for designing appropriate control schemes. Indeed, the local voltage to real and reactive power sensitivity has been widely used to coordinate voltage control, generation resource dispatch, and distribution network management [1]–[4]. Sensitivity analysis (SA) can be divided into two main categories, i.e., local sensitivity analysis (LSA) [1]–[7] and global sensitivity analysis (GSA) [8]–[11]. For LSA, it studies the impact of small perturbations on the model output by estimating the partial derivative with respect to the inputs at a specific point. In power systems, the Jacobian matrix is usually employed to achieve that. For example, in [3], the Jacobian matrix-based LSA is applied to distribution network voltage management. To mitigate the model uncertainties and errors, the measurement-based Jacobian matrix estimation for voltage to power sensitivity analysis is developed in [4]. However, these methods exclude the interactions of inputs on the model outputs and are typically limited to investigating the small variations of uncertain inputs. As a result, LSA fails to reveal the global impacts of various control devices on the voltage changes and therefore GSA becomes necessary [12]. 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 (SIs), and the Kucherenko indices [13]. Among them, SIs are widely adopted in the variance-based analysis framework [14]. The main idea 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 [15]. Note that the Monte Carlo (MC) simulations are usually used for SIs calculation, which can be time-consuming.

To deal with the computational burden of MC-based SIs calculations, the surrogate modeling techniques are advocated. In particular, an approximated reduced order model is derived to replace the original one so that a higher computational efficiency can be reached with only slight loss of accuracy [16], [17]. For example, in [11], the polynomial chaos expansion (PCE) based approach is developed for distribution system GSA. The PCE-based technique is also applied for probabilistic power flow [18], [19]. However, the calculation of PCE coefficients requires accurate physical model. This is very difficult to achieve for practical distribution system, especially with the increased penetration of PVs, flexible loads, and advanced demand response program [20]. Moreover, PCE suffers from the curse of dimensionality and thus the application of it to large-scale systems is limited. Another commonly-used method is the Kriging [21], i.e., Gaussian process regression. As a statistical interpolation method, it can be applied directly on a set of observations and the required number of samples is much smaller than PCE. But the application of it to distribution system GSA is not trivial.

In summary, LSA only investigates the impacts of uncertain inputs within a small region of domain while practical problems

also require the examination of the interactions of inputs on the outputs across the whole input space. The GSA addresses that but existing works are model-based. However, the distribution system models are of large uncertainties and subject to various types of errors. Furthermore, the existing GSA methods are not computationally efficient to handle large number of uncertain inputs in a large-scale distribution system. This paper aims to bridge these gaps and proposes a data-driven GSA framework for distribution system with PVs, yielding the following contributions:

- It is data-driven and the assumptions on accurate physical model as well as the probabilistic distributions of uncertain inputs are not needed.
- It can reflect the global voltage variations to uncertain real and reactive power injections for unbalanced three-phase system. The impacts of PV injections with different capacities and distributions are investigated.
- Under the new framework, two approaches are proposed to calculate the Sobol indices in a computationally efficient manner, including the data-driven Kriging-based method using MC and the data-driven Kriging with analytical analysis of variance (ANOVA) kernel method. The former one constructs a Kriging model, based on which Sobol indices are estimated using MC simulations. This leads to higher computational efficiency than performing MC simulations on the original physical model. The ANOVA kernel-based method, on the other hand, derives the Sobol indices analytically after establishing Kriging model. By avoiding MC simulations, it achieves even higher computational efficiency than the first approach in the presence of large number of uncertain inputs. These are demonstrated via the results in Section IV.

The paper is organized as follows: Section II shows the problem statement and the proposed data-driven GSA framework is elaborated in Section III. Section IV analyzes the test results and finally Section V concludes the paper.

II. PROBLEM STATEMENT

Let $\mathbf{y} = \mathcal{M}(\mathbf{x})$ be the model with random input vector $\mathbf{x} = [x_1, \dots, x_d]^T$ and model response $\mathbf{y}$. For three-phase distribution systems, the uncertain model inputs may include PV injections and loads while the outputs are typically bus voltage magnitudes, voltage angles, line flows, etc. Sensitivity analysis aims to quantify how the model response is affected by each input or their combinations. To estimate the sensitivity at $\tilde{\mathbf{x}}$, the corresponding partial derivative $(\partial \mathbf{y} / \partial \mathbf{x})_{\mathbf{x}=\tilde{\mathbf{x}}}$ is commonly used. This approach belongs to LSA since the perturbations occur in the neighborhood of the nominal values. Therefore, the index is reliable only in a small range of the inputs. As for distribution system LSA, the sensitivity information can be derived from Jacobian matrix. Via the linearization on branch flow equations, [3] formulates the sensitivity of voltage magnitude with respect to real and reactive power injections P and Q as

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

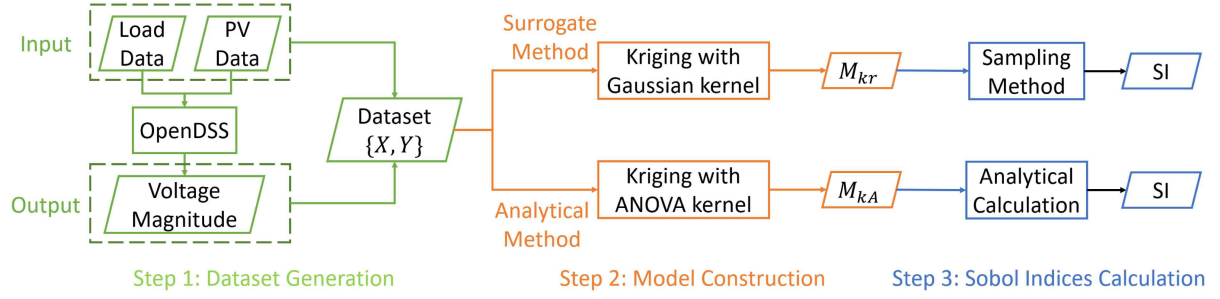


Fig. 1. Framework of the proposed data-driven GSA, where SI represents the Sobol index and two methods are developed for implementations.

Note that LSA index is reliable only in a small range of the inputs. However, the stochastic PV injections can change rapidly within their lower and upper bounds and LSA may not be reliable. Furthermore, LSA can only capture the local information with individual components and does not reflect the true global sensitivity of model response to all inputs. The PV injections can also have joint impacts on the interested node voltages, which are especially the case when multiple PV inverters are coordinated for voltage regulation. This has been ignored by LSA. Finally, for the distribution network, the derivatives in Jacobian matrix only consider the buses that are directly connected to. But those power injections that are not far away can have large impacts on the interested buses, which are shown in the numerical results.

By contrast, GSA reveals the global relationship between input variations and output variations by providing a quantitative importance ranking with respect to each input and the combination thereof. Based on that, the following objectives can be achieved: i) identification of the most influential inputs; ii) characterizing the relationship of the input variation and the output behavior, especially the joint impacts of several uncertain inputs on the model outputs, such as voltage to real and reactive power sensitivity; iii) observability analysis of the parameters for identification. More discussions on the usefulness of GSA for power system applications are shown in Section III-E. One of the widely used GSA methods is the Sobol indices-based technique. Sobol indices measure the contribution of uncertain sources and their interactions to predictive uncertainty of model response. The calculations of SIs are often realized via MC simulations that are computationally expensive. This calls for the development of more computationally efficient methods. Approximate methods, such as the surrogate modeling-based ones, are alternatives of speeding up the calculations although a lot of samples are still needed. On the other hand, analytical method avoids sampling from models and thus further improves the computational efficiency. In this paper, we develop and compare the approximate and analytical methods under the data-driven framework, while achieving significant computational efficiency improvement.

III. PROPOSED DATA-DRIVEN GSA APPROACH

The proposed framework is shown in Fig. 1, where two approaches are developed to perform GSA in a computationally efficient manner, including the data-driven Kriging-based method using MC simulations and the data-driven Kriging with ANOVA

kernel method. From the numerical results shown in Section IV, the latter method is recommended for practical applications. The general steps include the data generation/collection, surrogate modeling and Sobol indices calculation. Specifically, the data-driven surrogate model is realized by traditional Kriging and the ANOVA kernel Kriging. In this section, we first present the theory of Sobol indices for GSA, followed by the two data-driven approaches and their algorithm implementations.

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 ANOVA-representation of $\mathcal{M}(\mathbf{x})$ is defined as [15]:

$$\begin{aligned} \mathcal{M}(x_1, \dots, x_d) = & \mathcal{M}_0 + \sum_{i=1}^d \mathcal{M}_i(x_i) \\ & + \sum_{1 \leq i < j \leq d} \mathcal{M}_{ij}(x_i, x_j) + \dots + \mathcal{M}_{1, \dots, d}(x_1, \dots, x_d) \end{aligned} \quad (2)$$

under the condition that

$$\int_0^1 \mathcal{M}_{i_1, \dots, i_s}(x_{i_1}, \dots, x_{i_s}) dx_{i_k} = 0 \quad (3)$$

for $1 \leq k \leq s, 1 \leq s \leq d$, where we have

$$\begin{cases} \mathcal{M}_0 = \int_{\mathcal{D}_x} \mathcal{M}(\mathbf{x}) d\mathbf{x} \\ \mathcal{M}_i(x_i) = \int_{\mathcal{D}_x^{d-1}} \mathcal{M}(\mathbf{x}) d\mathbf{x}_{\sim i} - \mathcal{M}_0 \\ \mathcal{M}_{ij}(z_i, z_j) = \int_{\mathcal{D}_x^{d-2}} \mathcal{M}(\mathbf{x}) d\mathbf{x}_{\sim (i,j)} - \mathcal{M}_0 - \mathcal{M}_i - \mathcal{M}_j \end{cases} \quad (4)$$

where $\mathbf{x}_{\sim i}$ denotes the subset of $\mathbf{x}$ that excludes i th variable x_i and $\mathcal{D}_x^{d-1}$ is the corresponding support. Assuming $\mathcal{M}(\mathbf{x})$ is square-integrable, [15] indicates that by squaring (2) and integrating over $\mathcal{D}_x$, one can get

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

According to the expression of variance, the left-hand side denotes the variance of model response. Thus, we get [16]

$$var(Y) = V = \sum_{s=1}^d \sum_{i_1 < \dots < i_s} \int \mathcal{M}_{i_1, \dots, i_s}^2 dx_{i_1} \dots dx_{i_s} \quad (5)$$

On account of the orthogonality condition, the variance of model response can be expanded and (5) becomes

$$V = \sum_{i=1}^d V_i + \sum_{i < j}^d V_{ij} + \cdots + V_{1,\dots,d} \quad (6)$$

Then, the Sobol indices are defined as

$$S_I = \frac{V_I}{V} \quad (7)$$

where $I \subset \{1, \dots, d\}$. Sobol indices represent the contribution of input components to the output variance. The first-order Sobol index S_i measures the effect of individual input while higher-order Sobol indices $S_{i_1 \dots i_s}$ quantify the contribution by the interactions of $x_{i_1}, \dots, x_{i_s}$. S_i plays the similar role of LSA while $S_{i_1 \dots i_s}$ allows us to count the joint impacts of inputs on the outputs, which are not available from LSA. To further determine the general importance of each input, the total Sobol index is defined as

$$S_i^T = \sum_{\{i_1, \dots, i_s\} \supset i} S_{i_1 \dots i_s} \quad (8)$$

The MC simulations are typically used to calculate Sobol indices but they are time-consuming. This paper presents data-driven approaches to address this challenge.

B. Kriging-Based Surrogate Model

This paper advocates the use of Kriging method as it has [22]: 1) strong theoretical justifications and proofs in building the surrogate model, 2) data-driven formulation that does not need the physical model and the probabilistic distributions of uncertain inputs, 3) high accuracy using only few samples, which is a huge advantage over other machine learning algorithms that need a large amount of data.

The Sobol indices are typically calculated based on MC simulations that rely on the original physical models. Reduced order model to speed up the process can be used as well [11]. However, the accurate physical model is assumed, which is challenging to achieve for practical distribution systems. To deal with that, data-driven surrogate model via Kriging is developed for Sobol indices calculations. Kriging assumes that the model response is an observation of a Gaussian process (GP), i.e.,

$$\mathcal{M}_{kr}(\mathbf{x}) = m(\mathbf{x}) + Z(\mathbf{x}; \sigma^2, \boldsymbol{\theta}) \quad (9)$$

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

Then, the unknown parameters are estimated using ML estimator [23], [24]:

$$\begin{cases} \hat{\beta} = \beta(\boldsymbol{\theta}) = (\mathbf{F}^T \mathbf{K}(\boldsymbol{\theta})^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{K}(\boldsymbol{\theta})^{-1} \mathbf{y} \\ \hat{\sigma}^2 = \sigma^2(\boldsymbol{\theta}) = \frac{1}{N} (\mathbf{y} - \mathbf{F} \hat{\beta})^T \mathbf{K}(\boldsymbol{\theta})^{-1} (\mathbf{y} - \mathbf{F} \hat{\beta}) \\ \hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \hat{\sigma}^2 [\det \mathbf{K}(\boldsymbol{\theta})]^{1/m} \end{cases} \quad (10)$$

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

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

$$\begin{cases} \mu_{\tilde{y}} = \mathbf{f}^T \hat{\beta} + \tilde{\mathbf{k}} \mathbf{K}^{-1} (\mathbf{y} - \mathbf{F} \hat{\beta}) \\ \sigma_{\tilde{y}}^2 = \hat{\sigma}^2 (1 - \tilde{\mathbf{k}} \mathbf{K}^{-1} \tilde{\mathbf{k}}) + \mathbf{u}^T (\mathbf{F}^T \mathbf{K}^{-1} \mathbf{F})^{-1} \mathbf{u} \end{cases} \quad (11)$$

where $\hat{\beta}$ and $\hat{\sigma}^2$ are the generalized least-squares estimates from (10); $\tilde{\mathbf{k}}$ is the cross-covariance between $\tilde{\mathbf{x}}$ and $\mathbf{x}$ with components $\tilde{k}_i = k(\tilde{\mathbf{x}}, \mathbf{x}^{(i)})$, $i=1, \dots, N$; $\mathbf{K}$ is the covariance matrix of $\mathbf{X}$ with elements $K_{ij} = k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$, $i, j=1, \dots, N$; $\mathbf{u} = \mathbf{F}^T \mathbf{K}^{-1} \tilde{\mathbf{k}} - \mathbf{f}$. Note that the common trends are linear and quadratic [23] while the kernel functions include exponential kernel, Gaussian kernel, Matérn kernel, etc.

Sobol indices calculations: The construction of Kriging model is completed and the model predictions can be obtained. After that, the MC-based Sobol indices can be calculated via the surrogate model as follows:

$$\begin{cases} \hat{V}_0 = \frac{1}{N} \sum_{n=1}^N \mathcal{M}_{kr}(\mathbf{x}^{(n)}) \\ \hat{V} = \frac{1}{N} \sum_{n=1}^N \mathcal{M}_{kr}^2(\mathbf{x}^{(n)}) - \hat{V}_0^2 \\ \hat{V}_i = \frac{1}{N} \sum_{n=1}^N \mathcal{M}_{kr}(\mathbf{x}_i^{(n)}, \mathbf{x}_{\sim i}^{(n)}) \mathcal{M}_{kr}(\mathbf{x}_i^{(n)}, \mathbf{x}'_{\sim i}^{(n)}) - \hat{V}_0^2 \end{cases} \quad (12)$$

where $\mathbf{x}_{\sim i}^{(n)}$ is the observation with i th input variable excluded; $\mathbf{x}'$ is another realization of $\mathbf{X}$ and is independent with $\mathbf{x}$.

C. ANOVA Kernel-Based Kriging

To further improve the efficiency of the previous Kriging-based GSA, a specific type of ANOVA kernel is utilized to obtain the closed-form ANOVA representation of $m(\mathbf{x})$ in (9), where the Sobol indices can be derived analytically. This avoids sampling from surrogate models and thus accelerates the GSA procedure especially in the presence of large number of unknown inputs.

By rewriting the interpolator $m(\mathbf{x})$ in the form of kernel, we get

$$m(\mathbf{x}) = \mathbf{k}(\mathbf{x})^T \mathbf{K}^{-1} \mathbf{y} \quad (13)$$

where $\mathbf{k}(\mathbf{x})$ denotes the vector of covariances between the test point and the training set, i.e. $\mathbf{k}(\cdot)$ is the column vector of $(\mathbf{K}(\mathbf{x}^{(i)}, \cdot))_{1 \leq i \leq N}$; $\mathbf{K}$ represents the Gram matrix with terms $K_{ij} = k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$, where k belongs to symmetric positive definite function, such as Gaussian or Matérn. The original

ANOVA kernel is given by

$$\mathbf{k}(\mathbf{x}) = \mathbf{1} + \sum_{I \subset \{1, \dots, d\}} \bigodot_{i \in I} \mathbf{k}^i(x_i) \quad (14)$$

where $\mathbf{1}$ is the bias term, i.e., a column vector with all ones; d represents the dimension of input; $\bigodot$ denotes entrywise product: $(\bigodot_{i \in I} \mathbf{k}^i(x_i))_j = \prod_{i \in I} k^i(x_i, x_j)$. However, the expression of $m(\mathbf{x})$ based on the original ANOVA kernels does not necessarily satisfy the demanded orthogonality condition as required by ANOVA representation (3). To this end, we resort to [21] and develop the following ANOVA kernels:

$$\mathbf{k}_0(\mathbf{x}) = \mathbf{1} + \sum_{I \subset \{1, \dots, d\}} \bigodot_{i \in I} \mathbf{k}_0^i(x_i) \quad (15)$$

$$k_0(x, y) = k(x, y) - \frac{\int_D k(x, s) d\mu(s) \int_D k(y, s) d\mu(s)}{\int \int_{D \times D} k(s, t) d\mu(s) d\mu(t)} \quad (16)$$

where $D = D_1 \times \dots \times D_d$ is a Cartesian product space of sets $D_i \subset \mathbb{R}$; $\mu = \mu_1 \times \dots \times \mu_d$ is a probability measure over D . The corresponding ANOVA representation of $m(\mathbf{x})$ is:

$$m(\mathbf{x}) = \left[\mathbf{1} + \sum_{i=1}^d \mathbf{k}_0^i(x_i) + \sum_{1 \leq i < j \leq d} \mathbf{k}_0^i(x_i) \odot \mathbf{k}_0^j(x_j) + \dots + \bigodot_{i=1}^d \mathbf{k}_0^i(x_i) \right]^T \mathbf{K}^{-1} \mathbf{y} \quad (17)$$

According to (17), the term of m indexed by I reads as:

$$m_I = \left(\bigodot_{i \in I} \mathbf{k}_0^i(x_i) \right)^T \mathbf{K}^{-1} \mathbf{y} \quad (18)$$

Sobol indices calculations: by combining (7) and (18), the SIs are naturally obtained as:

$$S_I = \frac{\text{var}(m_I)}{\text{var}(m)} = \frac{\mathbf{y}^T \mathbf{K}^{-1} (\bigodot_{i \in I} \mathbf{\Gamma}_i) \mathbf{K}^{-1} \mathbf{y}}{\mathbf{y}^T \mathbf{K}^{-1} (\bigodot_{i=1}^d (\mathbf{1}_{n \times n} + \mathbf{\Gamma}_i) - \mathbf{1}_{n \times n}) \mathbf{K}^{-1} \mathbf{y}} \quad (19)$$

where $\mathbf{\Gamma}_i = \int_{D_i} \mathbf{k}_0^i(x_i) \mathbf{k}_0^i(x_i)^T d\mu_i(x_i)$ and $\mathbf{1}_{n \times n}$ is the $n \times n$ matrix of ones. The integrals of (19) are approximated using a Riemann sum operator in numerical computation [21].

Remark: the key advantage of the ANOVA kernel-based Kriging over the traditional Kriging-based surrogate model for GSA assessment is that it has an approximate linear computational complex increase with the increased number of inputs while this is not the case for the latter. That is essential for the system with high dimension of unknown inputs, which will be demonstrated in the numerical results section.

D. Algorithm Implementation

From Fig. 1, it can be found that the proposed GSA method includes constructing data-driven surrogate models and Sobol indices calculations. Two methods are developed and their main steps of implementations are as follows:

- Step 1: Generate dataset via OpenDSS [25]. Specifically, OpenDSS calculates the power flows with a set of random inputs and generates a set of outputs correspondingly, such as bus voltage magnitudes, real and reactive power injections. This dataset is taken as the input measurement data for the proposed methods in the experimental simulation sections. It is worth noting that for practical distribution system applications, this step is not required, instead the historical smart meter and SCADA measurements are directly used.
- Step 2: Construct Kriging surrogate model. For surrogate model-based Kriging, the data-driven Kriging model is obtained using (10). Alternatively, we can build ANOVA kernel-based method using the ANOVA kernel via (15).
- Step 3: Calculate Sobol indices. For surrogate-based Kriging, MC-based Sobol indices calculations are performed through (12) with (11). For ANOVA kernel-based Kriging, the corresponding Sobol indices are directly obtained in closed form via (19).

E. Potential Applications of GSA

GSA aims to assess the importance of all uncertain sources in a global view, which provides valuable information about how each uncertain PV affects the desired output and to what degree the influence is. This leads to several potential applications, which are summarized as follows:

- *Importance ranking of uncertain sources:* GSA provides a quantitative measure for the importance of each input variable on the model outputs. For example, by ranking the calculated SIs, we can find out the most influential inputs on model outputs, such as those PVs that affect most the voltage variations. There are also some other applications developed in the literature. SA is utilized in [1] to investigate PV buses and tap position adjustments for simulating controllers in distribution systems. The sensitivity analysis using partial derivatives is conducted to check for any violations in the regulated bus voltage magnitudes. SA is also applied to voltage contingency ranking, where the effect of the contingency on the states and stability margin is determined through partial derivatives [9]. Consequently, a severity index is calculated for each contingency. The proposed data-driven GSA is a more general approach that considers the sensitivity in the whole range and can be easily applied to these problems.
- *Overall model sensitivity estimation:* GSA allows exploring the relationship between the input variations and the output behavior. GSA is utilized in [8] for microgrid maximum loadability analysis. The polynomial chaos expansion is used for GSA calculation. In [10], GSA is performed to measure the effect of distributed generation. Using the ranking of each input factor, the voltage sag based fault location can be achieved. These methods are model-based while our framework is data-driven and thus more applicable of addressing the inaccurate physical model issues.
- *Network clustering:* Assume GSA is conducted for all desired outputs in the power system, we can cluster nodes that

have similar sensitivity pattern into several areas, which further helps distributed applications, such as distributed voltage control. A community detection algorithm based on voltage sensitivity matrix is proposed by [5] in order to partition the distribution network into clusters for zonal voltage control. In [6], the relationship between voltage and distributed generation power injections characterized by voltage sensitivity analysis is used to evaluating the electrical distance, which is further used for network partition.

- *Distribution network management*: The GSA results contain useful information of the whole system. The exploration of such sensitivity coefficients leads to further applications, such as distribution network voltage-var optimization, selection of feeders for conservation voltage reduction implementation, etc. The sensitivity coefficients from each participating node or voltage regulator are collected to help distribution management system perform optimization and control [2].
- *Model parameter identifiability assessment*: GSA can assess how the parameter variations affect the quality of the output. This allows us to identify the most influential parameters for identification. The polynomial chaos expansion-based GSA for generator model problematic parameter identification in [26]. The GSA can be also extended to assist the selection of parameters of dynamic loads, DERs, etc., for estimation and calibration.

Remark: If topology changes or reconfigurations happen in distribution systems, the sensitivity is affected and this requires the updating of surrogate model to calculate new SIs. It should be noted that those topology changes or reconfigurations typically affect some local nodes due to the sparsity of the distribution systems. This means that only the SIs associated with those local nodes need to be updated. One excellent characteristic of Kriging-based method is that it needs a few samples to reconstruct and update the surrogate model. This process is also swift to accomplish and thus allows our proposed method to quickly adapt to topology changes and reconfigurations.

IV. NUMERICAL RESULTS

Numerical results carried out on the modified IEEE 37- and 123-bus system considering PVs are used to demonstrate the performance of the proposed method. The schematic is shown in Fig. 2. The inputs include random variations of loads and PV injections while the model responses, voltage magnitudes are used for illustrations. The model-based PCE [11] and the proposed two data-driven methods are compared with the benchmark obtained from the MC simulations. The mean absolute percentage error (MAPE) and root mean square error (RMSE) are used to quantify the surrogate model prediction accuracy and Sobol indices estimation accuracy, i.e.,

$$\begin{cases} MAPE = \frac{1}{N} \sum \left| \frac{y^* - \tilde{y}}{y^*} \right| \times 100\% \\ RMSE = \sqrt{\frac{1}{N} \sum (S^* - \tilde{S})^2} \end{cases} \quad (20)$$

where y^* and $\tilde{y}$ are the true and estimated model response, respectively; S^* and $\tilde{S}$ are the true and estimated SIs, respectively.

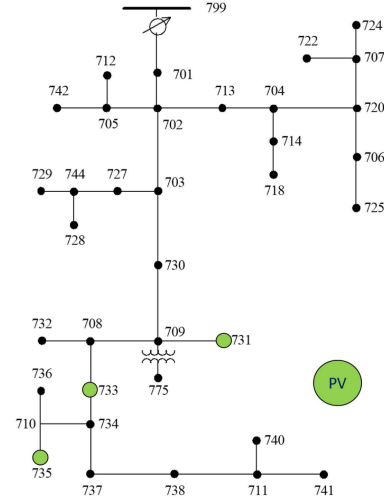


Fig. 2. Single line diagram of IEEE 37-bus system with PVs.

All simulations are carried out in MATLAB with 2.60 GHz Intel Core i7-6700HQ.¹

The AC power flow model $\mathcal{M}_{pf}$ is used as the benchmark to estimate the total SIs S_{pf}^T via MC simulations. Meanwhile, the inputs $\mathbf{X}$ are generated from known distributions and the corresponding model responses $\mathbf{y}$ are obtained through $\mathcal{M}_{pf}$. In this paper, all realizations are calculated from OpenDSS [25]. Then, the generated data $\{\mathbf{X}, \mathbf{y}\}$ are used to construct three surrogate models $\{\mathcal{M}_{pc}, \mathcal{M}_{kr}, \mathcal{M}_{ka}\}$ for PCE, Kriging, and Kriging with ANOVA kernel, respectively. Their corresponding total SIs are $\{S_{pc}^T, S_{kr}^T, S_{ka}^T\}$.

The organization of this section is as follows: Section IV-A and IV-B validate the proposed framework in several scenarios with different numbers and probability distributions of PVs and loads. Section IV-C investigates the robustness to measurement noise and IV-D tests the proposed method in a larger system with more numbers of uncertain inputs. Finally, an actual system with smart meter data is used to validate the scalability of the proposed framework in Section IV-E.

A. Sobol Indices Validation

The inputs include six random variables $\mathbf{x} = [x_1, \dots, x_6]$, 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]. Note that a, b , and c denote different phases. The corresponding model responses are $\mathbf{y} = [y_1, \dots, y_9]$ whose elements are three-phase voltage magnitudes at nodes [731, 733, 735]. The distributions of inputs in Scenario 3 shown in Table I are used here for illustration. All models are constructed with $N = 200$ samples. The degree of PCE is set to $n = 2$. Linear trend and Gaussian kernel are used for $\mathcal{M}_{kr}$ while linear trend and ANOVA kernel are utilized for $\mathcal{M}_{ka}$.

Table II shows the comparisons between three methods. The overall index to quantify surrogate model accuracy is the average

¹The code to implement the proposed method can be obtained by sending email to the corresponding author.

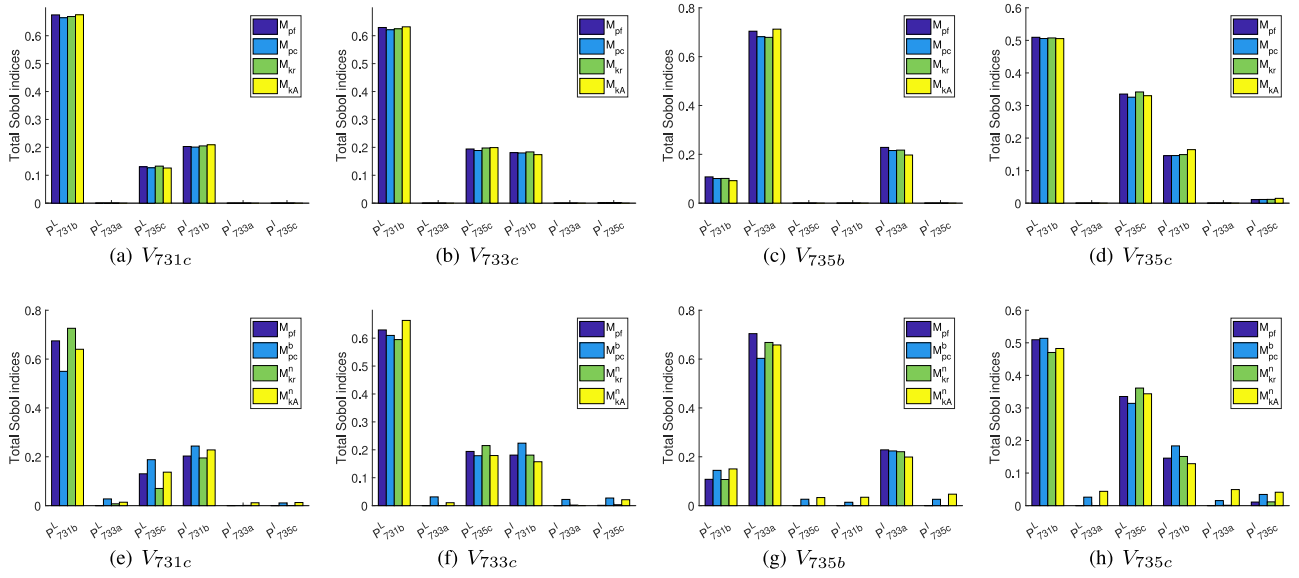


Fig. 3. (a)-(d) correspond to the case that there is no error with system model while (e)-(h) are the results considering system model error.

TABLE I
LOAD AND PV DISTRIBUTIONS FOR SCENARIOS 1-4

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

TABLE II
COMPARISON RESULTS OF DIFFERENT METHODS ON 37-BUS SYSTEM

Model	Accuracy		CPU time (s)
	$e_M (\times 10^{-5} \%)$	$e_{SI} (\times 10^{-2})$	
$\mathcal{M}_{pf}$	—	—	103
$\mathcal{M}_{pc}$	3.825	0.869	1.243
$\mathcal{M}_{kr}$	3.713	0.811	4.854
$\mathcal{M}_{kA}$	3.078	0.752	1.404

value of all MAPEs, denoted as e_M . Similarly, the overall accuracy index for SIs is e_{SI} . Once established offline, the surrogate models can replace the original power flow model and further realize SIs calculations online. The CPU time means the time for Sobol indices online calculations. A number of 10 000 samples is used for MC benchmark. According to Table II, PCE is the fastest method as only a small number of data is used. ANOVA kernel-based Kriging method shows a similar computational efficiency as PCE. The advantage of the proposed analytical method in terms of computational efficiency will become significant when a larger-scale system is tested. Note that all three methods can achieve very accurate results for model response predictions and SIs calculations while being much more computational efficient than the MC-based method. This validates the effectiveness of our data-driven methods. Figs. 3 (a)-(d) demonstrate the total Sobol indices for V_{731c} , V_{733c} , V_{735b} , and V_{735c} by three models for illustrations. It is

observed that only three inputs have noticeable effects on V_{731c} and the dominant factors are power fluctuations at phase b. For voltages at phase b, major influence comes from those at phase a. Results also show that voltage magnitudes at the same phase share a similar pattern of total SIs. According to the total index, load fluctuation of 10% has more impacts than PV injections with power rating 30 kW in terms of sensitivity.

In practice, the distribution system model and the measurements are always subject to errors. To test the robustness of these methods when the model is subject to errors, we assume there are uncertainties of distribution line parameters, following independent Gaussian distribution $\mathcal{N}_{LL} \sim (0, (0.05\mu_{LL})^2)$. For input observation $\mathbf{X}$, Gaussian noise is added with zero mean and standard derivations are set to be a proportion of the true mean: $\sigma_{nx} = 0.01\mu_x$. For the data-driven approaches, measurement error is introduced for model response with a variance $\sigma_{ny} = 0.01\mu_y$. Figs. 3 (e)-(h) display the results, from which it can be observed that the proposed method is only slightly affected. By contrast, due to model errors, the model-based PCE yields much larger errors in total SIs, see (e) and (g) for example.

B. Sensitivity in Different Scenarios

Further tests, including different PV injections and distributions are conducted to examine the proposed method, see Table I. Note that the model uncertainties are not considered in this section and the PCE is added for comparisons. Fig. 4 shows the results of models for different scenarios and all models achieve similar performances. From Scenarios 1 and 2, it is observed that the increase of power injections enlarges the SI of input P_{731b} as expected. The comparison of Scenarios 2 and 3 shows 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

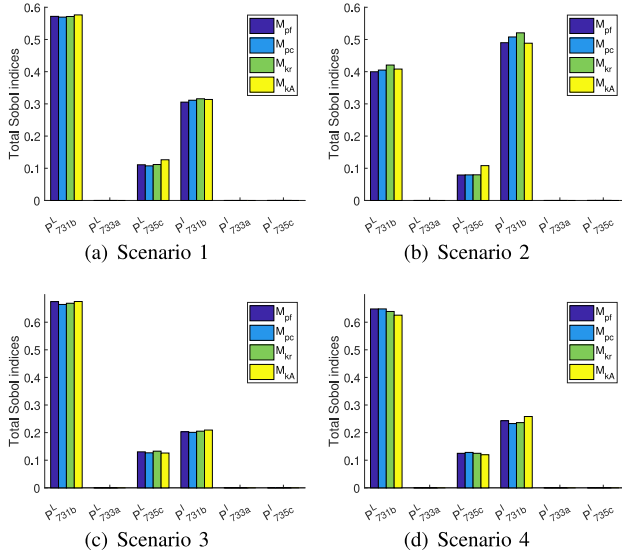


Fig. 4. Total Sobol indices calculated by different methods under scenarios 1-4 for V_{731c} .

TABLE III
MODEL PERFORMANCES IN SCENARIOS 1-4

SI error $e_{SI}(\times 10^{-2})$	$\mathcal{M}_{pc}$	$\mathcal{M}_{kr}$	$\mathcal{M}_{ka}$
Scenario 1	0.617	0.684	0.716
Scenario 2	1.114	1.331	1.172
Scenario 3	0.869	0.811	0.752
Scenario 4	0.891	0.960	1.030

setting of PV injections in Scenario 4 yields analogous distribution as that in Scenario 3. Table III demonstrates the model accuracy under different scenarios, justifying the high accuracy of the proposed methods. These results also demonstrate that our proposed methods are able to track the sensitivity changes and reveal the complicated global sensitivity relationships between changing inputs and outputs.

C. Robustness to Noise

This case study is to assess the robustness of the proposed methods to measurement noise. Both inputs and outputs are subject to noise $\{\sigma_{nx}, \sigma_{ny}\}$ as discussed in IV-A. The proposed methods are Kriging-based methods and therefore are able to handle additive noise [23]. In the model construction step, a good number of samples can help Kriging achieve better surrogate model. Note that for the collected measurements, noise reduction algorithm can be used to preprocess the data. Fig. 5 and 6 demonstrate the robustness of the proposed methods to noise with moderately increased number of samples, where the V_{733c} sensitivity is used for illustration. It can be observed that the proposed methods achieve better capability of filtering out the noise with more samples. The traditional Kriging is more robust to noise than the ANOVA kernel-based method. Compared with the results in Section IV-A, we find that to achieve similar estimation accuracy due to measurement noise, the number of data samples should be increased. This is expected as handling

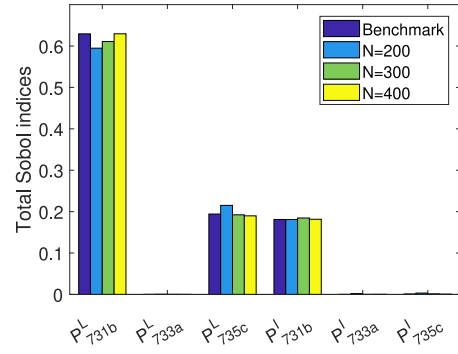


Fig. 5. Robustness of $\mathcal{M}_{kr}$ to measurement noise for V_{733c} sensitivity.

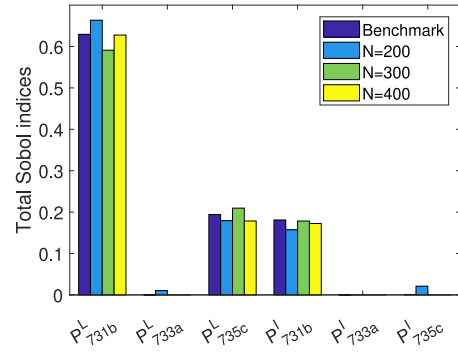


Fig. 6. Robustness of $\mathcal{M}_{ka}$ to measurement noise for V_{733c} sensitivity.

TABLE IV
LOAD AND PV SETTINGS FOR IEEE 123-BUS SYSTEM

Types	Locations
Loads	1a, 22b, 30c, 34c, 37ac, 41c, 45a, 52a, 55a, 58b, 62c, 64b, 68a, 77b, 82a, 87b, 96b, 102c, 107b, 109a
PVs	2b, 24c, 46a, 59b, 66c, 71a, 79a, 83c, 90b, 111a

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 to. For higher noise levels and more inputs, more advanced noise reduction algorithm can be investigated, such as principle component analysis (PCA) or kernel PCA [27].

D. Scalability to Larger Distribution Systems

The proposed Kriging methods are also tested in the IEEE 123-bus system to demonstrate their scalability. For this system, it is assumed that there are 30 uncertain inputs whose detailed descriptions can be found in Table IV. The single line diagram of the 123-bus system with PVs is displayed in Fig. 7. For PCE, its number of degrees is adapted to $n=2:4$ during the model construction stage and the number of samples is increased to be $N_{pc}=1000$ due to the increased number of uncertain inputs and the complexity of system model. By contrast, the hyperparameters of the Kriging methods and the numbers of samplings are the same as those in Section IV-A. Similar to

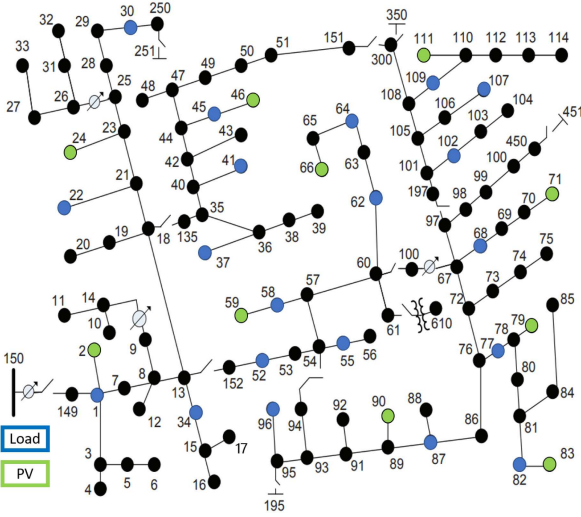


Fig. 7. Single line diagram of IEEE 123-bus system with PVs.

TABLE V
LOAD AND PV DISTRIBUTIONS FOR SCENARIOS 5 AND 6

Scenarios	$P_{inj}(kW)$	$P_{load}(kW)$
Scenario 5	$5 \times Beta(3, 2.2)$	$\mathcal{N}(P_{load}^i, (0.05P_{load}^i)^2)$
Scenario 6	$10 \times Beta(2.5, 3.5)$	$\mathcal{N}(P_{load}^i, (0.05P_{load}^i)^2)$

TABLE VI
COMPARISON RESULTS OF DIFFERENT METHODS IN 123-BUS SYSTEM

Model	Accuracy		CPU time (s)
	$e_M (\times 10^{-5} \%)$	$e_{SI} (\times 10^{-2})$	
$\mathcal{M}_{pf}$	—	—	8286.299
$\mathcal{M}_{pc}$	6.838	0.386	41.677
$\mathcal{M}_{kr}$	6.194	0.378	73.365
$\mathcal{M}_{kA}$	6.271	0.413	5.860

Section IV-A, the Sobol indices calculated by the MC-based approach with 10 000 samples are used for the baseline.

Table VI shows the comparison results for all methods. Compared with the those shown in Table II, it can be observed that the surrogate model prediction accuracy e_M has been decreased due to the increased complexity of the system. However, their level of accuracies are still high and sufficient for practical applications. Figs. 8 and 9 display the results of V_{64c} under Scenarios 5 and 6 for illustrations, where the experimental settings can be found in Table V. In general, all three methods are able to get accurate estimates of total SIs even with increased number of uncertain inputs. On the other hand, similar to Fig. 3, a few inputs have significant impacts on the final outputs. For V_{64c} , the major effects come from the PV injection at node 62c and the load fluctuation on the same node 64b. Power fluctuations at nearby nodes of phases b and c also contribute a certain proportion of total SIs. Furthermore, it can be observed that the increase of PV ratings leads to the increase of total SIs. It is worth noting that the ANOVA kernel-based Kriging method has small errors at nodes where the total SIs are supposed to be close to zero. A possible reason is that the error accumulates during analytical

TABLE VII
COMPARISON RESULTS OF DIFFERENT METHODS ON 240-BUS SYSTEM

Model	Accuracy		CPU time (s)
	$e_M (\times 10^{-5} \%)$	$e_{SI} (\times 10^{-2})$	
$\mathcal{M}_{pf}$	—	—	1462.135
$\mathcal{M}_{pc}$	6.578	0.766	18.331
$\mathcal{M}_{kr}$	5.560	0.637	16.278
$\mathcal{M}_{kA}$	5.087	0.482	2.558

calculation and thus S_{kA}^T tends to be subtle when true total SIs are almost zero. However, such error is negligible and is small enough to be removed through post-processing procedure in practical applications as we are usually looking for the most influential factors for voltage regulations.

In terms of computational efficiency, it can be observed from Table VI that the MC-based method is extremely time-consuming for practical applications. Although the PCE and traditional Kriging-based methods significantly improve the computational efficiency, they are still much more computationally expensive than our ANOVA kernel-based analytical calculation method, i.e., 8 times and 14 times faster than PCE and transitional Kriging method, respectively. Note that with the increased number of uncertain inputs, the PCE and Kriging face the curse of dimensionality issue. The degree of PCE needs to increase as well for larger and more complex systems. Similarly, Kriging has a higher requirement for estimating its hyperparameters. Although Kriging requires less number of training samples than PCE, the size of the covariance matrix K will eventually become unacceptable. By contrast, the ANOVA kernel-based method calculates the SIs analytically and has approximately linear complexity relationship with the number of uncertain inputs. Thus, it is more suitable for practical distribution systems with vast uncertain PVs and loads.

E. Test Results on a Real Distribution System

The proposed method is also tested on a real 240-bus distribution system in the US with smart meter data, see Fig. 10. The data are open public and can be found via [28]. In particular, the feeder C is selected, where there are 11 PVs installed and their locations can be found in Fig. 10. These nodes with PVs are regarded as uncertain resources. The smart meter dataset consists of one-year voltage, real and reactive power data with 8760 samples. The parameters and configurations for three methods are the same as those in Section IV-A. Fig. 11 shows the calculated total SIs for this 240-bus system while Table VII displays the computational efficiency of each method. It can be observed that the proposed analytical method achieves better accuracy than the other approaches while maintaining the best computational efficiency. Take the results for node 3034 as an example, it is found that the voltage of that node is mostly impacted by PVs at nodes 3024, 3085, 3090, and 3095. This is consistent with their distance to the target node and power ratings. Most other nodes have negligible impacts since they are far away from node 3034, such as nodes 3065, 3144 and 3158. Note that the power rating of PV node 3020 is close to zero and that is why it has

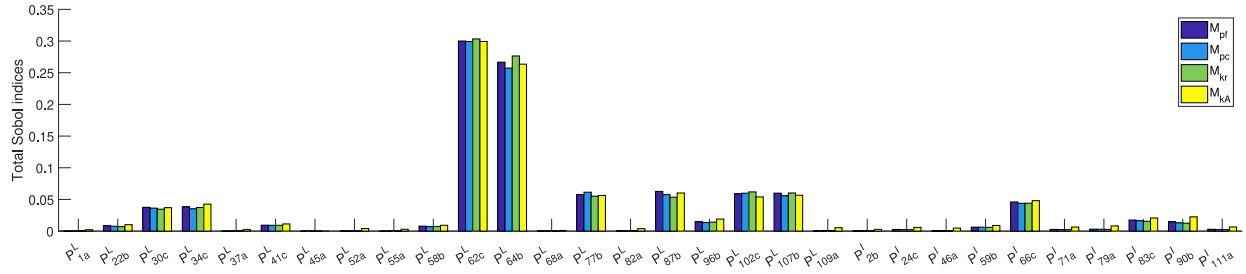


Fig. 8. Total Sobol indices for V_{64c} in the IEEE 123-bus system under Scenario 5.

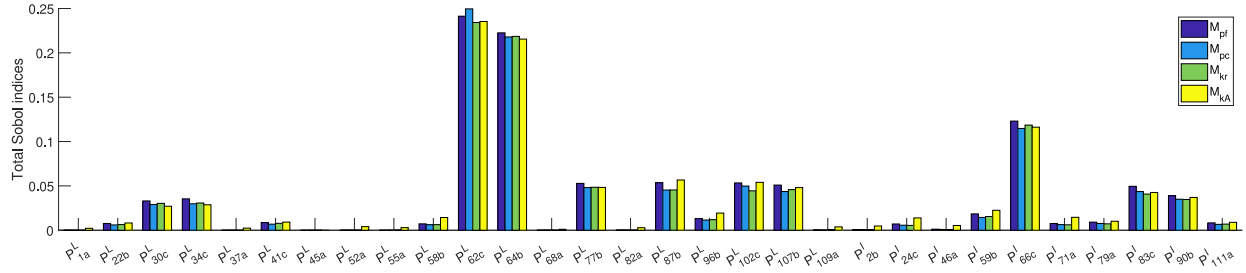


Fig. 9. Total Sobol indices for V_{64c} in the IEEE 123-bus system under Scenario 6.

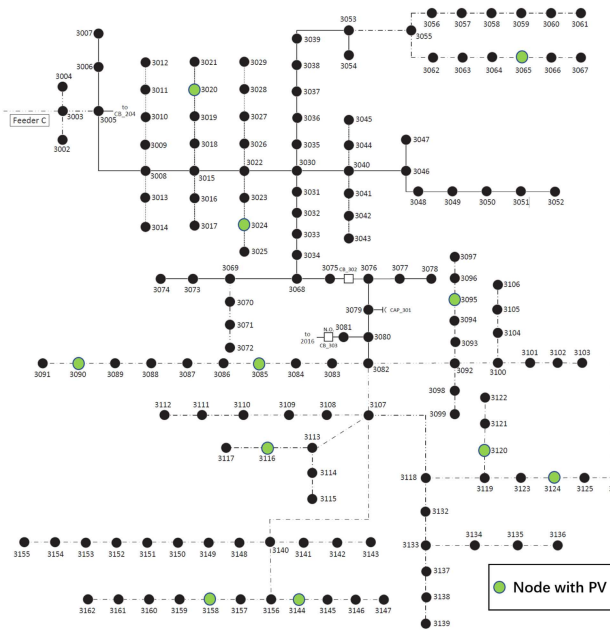


Fig. 10. Single line diagram of feeder C of 240-bus system with PVs.

neglectable impact on node 3034 voltage, which is also reflected by the total SIs.

V. CONCLUSION

In this paper, a data-driven GSA framework is proposed for three-phase distribution systems with stochastic loads and uncertain PVs. GSA allows us to quantify the overall impacts of uncertain inputs on model response, i.e., the voltage variations to PVs and loads. The proposed method has two key

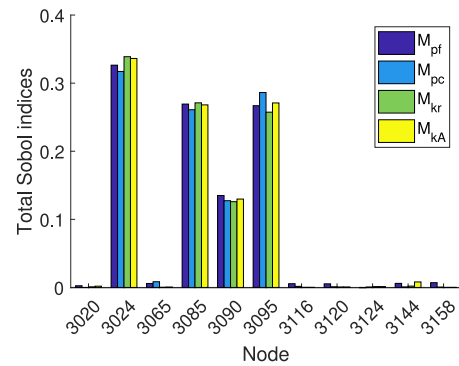


Fig. 11. Total Sobol indices result for V_{3034} .

components, namely the surrogate modeling via data-driven Kriging and Sobol indices calculation. Two approaches have been proposed, namely the traditional Kriging-based and the ANOVA kernel-based Kriging. The former one still requires MC simulations to calculate Sobol indices while the latter analytically derives the Sobol indices from the data and therefore achieves much higher computational efficiency, especially in the presence of large number of uncertain inputs. Simulation results on the unbalanced IEEE 37- and 123-bus systems show that our data-driven approaches can achieve comparable accuracy as the benchmark but being much more computational efficient. The proposed framework is also tested on a real 240-bus system with smart meter data to demonstrate its feasibility and scalability. Future work will be on developing closed-loop voltage control algorithm utilizing the global voltage sensitivity analysis outcomes.

REFERENCES

- [1] I. Dzafic, R. Jabr, E. Halilovic, and B. Pal, "A sensitivity approach to model local voltage controllers in distribution networks," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1419–1428, May 2014.
- [2] K. Youssef, "A new method for online sensitivity-based distributed voltage control and short circuit analysis of unbalanced distribution feeders," *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1253–1260, May 2015.
- [3] F. Tamp and P. Ciufu, "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.
- [4] S. Nowak, C. Chen, and 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] B. Zhao, Z. Xu, C. Xu, C. Wang, and F. Lin, "Network partition-based zonal voltage control for distribution networks with distributed PV systems," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4087–4098, Sep. 2018.
- [6] H. Ruan, H. Gao, Y. Liu, L. Wang, and J. Liu, "Distributed voltage control in active distribution network considering renewable energy: A novel network partitioning method," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4220–4231, Nov. 2020.
- [7] R. Minguez and A. J. Conejo, "State estimation sensitivity analysis," *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1080–1091, Aug. 2007.
- [8] X. Xu *et al.*, "Maximum loadability of islanded microgrids with renewable energy generation," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 4696–4705, Sep. 2019.
- [9] N. Amjadi and M. Esmaili, "Application of a new sensitivity analysis framework for voltage contingency ranking," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 973–983, May 2005.
- [10] P. Chen, V. Malbasa, Y. Dong, and M. Kezunovic, "Sensitivity analysis of voltage sag based fault location with distributed generation," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 2098–2106, Jul. 2015.
- [11] F. Ni, M. Nijhuis, P. H. Nguyen, and J. F. G. Cobben, "Variance-based global sensitivity analysis for power systems," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1670–1682, Mar. 2018.
- [12] B. Iooss, and P. Lemaître, "A review on global sensitivity analysis methods," in *Uncertainty Management in Simulation-Optimization of Complex Systems*, vol. 59, G. Dellino, C. Meloni, Eds., Boston, MA, USA: Springer, 2015, pp. 101–122.
- [13] M. Ye and M. C. Hill, "Global sensitivity analysis for uncertain parameters, models, and scenarios," in *Sensitivity Analysis Earth Observation Modelling*, vol. 10, G. P. Petropoulos, P. K. Srivastava, Eds., Amsterdam, Netherlands: Elsevier, 2017, pp. 177–210.
- [14] X. Liao *et al.*, "Extended affine arithmetic-based global sensitivity analysis for power flow with uncertainties," *Int. J. Elect. Power Energy Syst.*, vol. 115, Feb. 2020, Art. no. 105440.
- [15] I. M. Sobol, "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates," *Math. Comput. Simulat.*, vol. 55, no. 1–3, pp. 271–280, Feb. 2001.
- [16] L. L. Gratiet, S. Marelli, and B. Sudret, "Metamodel-based sensitivity analysis: Polynomial chaos expansions and Gaussian processes," in *Handbook of Uncertainty Quantification*, vol. 38, R. Ghanem, D. Higdon, H. Owaldi, Eds., Cham, Switzerland: Springer, 2017.
- [17] S. Balduin, T. Westermann, and E. Puiutta, "Evaluating different machine learning techniques as surrogate for low voltage grids," *Energy Inform.*, vol. 3, no. 24, Oct. 2020.
- [18] L. Yan, X. Duan, B. Liu, and J. Xu, "Gaussian processes and polynomial chaos expansion for regression problem: Linkage via the RKHS and comparison via the KL divergence," *Entropy*, vol. 20, no. 3, Mar. 2018, Art. no. 191.
- [19] Y. Xu, Z. Hu, L. Mili, M. Korkali, and X. Chen, "Probabilistic power flow based on a Gaussian process emulator," *IEEE Trans. Power Syst.*, vol. 35, no. 4, pp. 3278–3281, Jul. 2020.
- [20] P. Forsyth, O. Nzimako, C. Peters, and M. Moustafa, "Challenges of modeling electrical distribution networks in real-time," in *Proc. Int. Symp. Smart Elect. Distrib. Syst. Technol.*, Sep. 2015, pp. 556–559.
- [21] N. Durrande, D. Ginsbourger, O. Roustant, and L. Carraro, "ANOVA kernels and RKHS of zero mean functions for model-based sensitivity analysis," *J. Multivariate Anal.*, vol. 115, pp. 57–67, Mar. 2013.
- [22] C. Rasmussen and Christopher K. I. Williams, *Gaussian Processes for Machine Learning*, The Cambridge, MA, USA: MIT Press, 2006.
- [23] C. Lataniotis, S. Marelli, and B. Sudret, "UQLab user manual - kriging (Gaussian process modelling)," Chair of Risk, Safety & Uncertainty Quantification, ETH Zurich, 2015, Report UQLab-V0.9-105.
- [24] B. Iooss, "Uncertainty and sensitivity analysis of functional risk curves based on Gaussian processes," *Rel. Eng. Syst. Saf.*, vol. 187, pp. 58–66, Jul. 2019.
- [25] R. C. Dugan, and D. Montenegro, "Reference guide - the open distribution system simulator (OpenDSS)," *Elect. Power Res. Inst., Inc.*, Jun. 2020.
- [26] Y. Xu *et al.*, "Response-surface-based bayesian inference for power system dynamic parameter estimation," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 5899–5909, Nov. 2019.
- [27] L. Zhao, Y. Liu, J. B. Zhao, Y. Zhang, L. Xu, and J. Liu, "Robust PCA-deep belief network surrogate model for distribution system topology identification with DERs," *Int. J. Elect. Power Energy Syst.*, vol. 125, 2021, Art. no. 106441.
- [28] F. Bu, K. Dehghanpour, Y. Yuan, Z. Wang, and Y. Zhang, "A data-driven game-theoretic approach for behind-the-meter PV generation disaggregation," *IEEE Trans. Power Syst.*, vol. 35, no. 4, pp. 3133–3144, Jul. 2020.
- [29] A. Saltelli *et al.*, *Global Sensitivity Analysis: The Primer*. Hoboken, NJ, USA: Wiley, 2008.
- [30] B. Sudret, "Global sensitivity analysis using polynomial chaos expansions," *Rel. Eng. Syst. Saf.*, vol. 93, no. 7, pp. 964–979, 2008.