A Data-Driven Uncertainty Quantification Method for Stochastic Economic Dispatch

Xiaoting Wang, Rong-Peng Liu, Xiaozhe Wang, Yunhe Hou, François Bouffard

Abstract—This letter proposes a data-driven sparse polynomial chaos expansion-based surrogate model for the stochastic economic dispatch problem considering uncertainty from wind power. The proposed method can provide accurate estimations for the statistical information (e.g., mean, variance, probability density function, and cumulative distribution function) for the stochastic economic dispatch solution efficiently without requiring the probability distributions of random inputs. Simulation studies on an integrated electricity and gas system (IEEE 118bus system integrated with a 20-node gas system) are presented, demonstrating the efficiency and accuracy of the proposed method compared to the Monte Carlo simulations.

Index Terms—Data-driven, economic dispatch, polynomial chaos expansion (PCE), uncertainty quantification.

I. INTRODUCTION

The intermittency of power generation from renewable energy sources (RES) results in great challenges in power system daily operation like economic dispatch (ED). To address the challenge of uncertainty in the ED problem, various formulations have been proposed such as stochastic programming, i.e., stochastic ED (SED), robust optimization formulation, chanceconstrained optimization, etc. A large number of Monte Carlo (MC)-based simulations are needed to solve the optimization problem using realizations from stochastic models, which unfortunately yields prohibitive computational efforts [1].

To lower the computation burden, uncertainty quantification methods have been applied to the SED problem. Safta et al. [1] applied the polynomial chaos expansion (PCE) method to build a surrogate model for the SED problem, which can achieve accurate results efficiently with fewer samples compared to MC-based methods. Li et al. [2] further adopted a sparse PCE (SPCE) method to alleviate "the curse of dimensionality". However, the surrogate models in [1], [2] are constructed by assuming that the random inputs follow Gaussian distributions, which may not be true in practice. More recently, a Gaussian process emulator-based approach was proposed in [3] to solve the SED problem. Yet, it has been discussed in [4] that the PCE method may outperform the Gaussian process emulatorbased method when the probability distribution of the response (e.g., the SED solution) tends to be multimodal.

In this letter, we propose a data-driven sparse PCE (DDSPCE) method that builds a surrogate model directly from a raw data set of random variables, without any prior assumptions on the marginal distributions of the random variables and output responses. The DDSPCE-based surrogate

model can accurately estimate the statistical information (e.g., mean, variance, probability density function (PDF), cumulative distribution function (CDF)) of the SED solution efficiently (e.g., 33 times faster than MC simulations on the integrated IEEE 118-bus power system and 20-node gas system [5]), even when the PDF of the SED solution is multimodal.

II. FORMULATION OF STOCHASTIC ECONOMIC DISPATCH

The two-stage models are typically used to solve the stochastic unit commitment (UC) problem, where the first stage is to find the UC decisions in the day-ahead electric market, and the second stage is to find the dispatch decisions [5]. In this letter, we assume that the first stage UC decisions for conventional generating units have already been determined from the day-ahead UC model similar to [1], [3]. Our goal is to solve the multi-period ED problem with fixed UC decisions while considering the uncertainties from RES. The uncertainties introduced by the power output of renewable generating units turn the production cost $Q(P_g, P_w)$ into a random variable, where P_g denotes the generator power output vector; P_w is a vector of random variables, e.g., wind power generation in this paper. The multi-period SED problem under fixed UC decisions can be represented by:

$$Q(\boldsymbol{P_g}, \boldsymbol{P_w}) = \min_{\boldsymbol{P_g}} \sum_{t \in T} \sum_{g \in G} C_g(\boldsymbol{P_g^t})$$
(1)

s.t.

$$\sum_{g \in G} P_g^t + \sum_{w \in W} P_w^t = \sum_{d \in D} P_d^t \quad \forall t \in T$$
(2a)

$$\underline{P_l} \le \sum_{g \in G} k_{lg} P_g^t + \sum_{w \in W} k_{lw} P_w^t - \sum_{d \in D} k_{ld} P_d^t \le \overline{P_l} \quad \forall t \in T \quad (2b)$$

$$\begin{split} P_{g}^{\min} x_{g}^{t} &\leq P_{g}^{t} \leq P_{g}^{\max} x_{g}^{t} \quad \forall g \in G, t \in T \\ &- R_{g}^{RD} x_{g}^{t} - R_{g}^{SD} (x_{g}^{t-1} - x_{g}^{t}) - P_{g}^{\max} (1 - x_{g}^{t-1}) \\ &\leq P_{g}^{t} - P_{g}^{t-1} \leq R_{g}^{RU} x_{g}^{t-1} + R_{g}^{SU} (x_{g}^{t} - x_{g}^{t-1}) \\ &+ P_{g}^{\max} (1 - x_{g}^{t}) \quad \forall g \in G, t \in T \end{split}$$
(2d)

where (1) is the objective function, i.e., to minimize the total production cost; t is the specific time period in the time periods set $T = \{1, \dots, T_m\}$ (e.g., 24-hour period in the simulation study); g is the generator index and G is the generator set. The operational and physical constraints are given in (2) based on direct current (DC) power flow model. Equation (2a) is the power balance constraint, where P_d^t is the d-th load demand at time t. Constraint (2b) denotes the power flow limits, where k_{lg} , k_{lw} and k_{ld} are the sensitivity coefficients for the *l*-th transmission line with respect to the traditional generator g, wind generator w and load d, respectively [5]. $\underline{P_l}$ and $\overline{P_l}$ are thermal limits of transmission line *l*. Constraint (2c) denotes the generation capacity limits with x_g^t being the predetermined UC decision for generator g at time t. Constraint (2d) describes the ramping capability constraint of generator g, where R_g^{RD} and R_g^{RU} denote the ramping down and up

This work was supported partially by Natural Sciences and Engineering Research Council (NSERC) Discovery Grant, NSERC RGPIN-2016-04570 and partially by the Fonds de Recherche du Québec-Nature et technologies under Grant FRQ-NT PR-253686.

rate; R_g^{SD} and R_g^{SU} denote the shut-down and start-up ramp rate. Note that in this letter the gas system is integrated into the power system for case study. Additional constraints of gas network are considered (see Appendix), yet the essence of the SED problem is not affected. Please check [5] for more details.

As can be seen, the SED problem by nature is a complex constrained optimization problem. MC simulations are typically adopted to estimate the cost from a finite set of realizations from stochastic models of load variation, wind generation, etc., which nevertheless lead to high computation efforts even with scenario reduction techniques. In this letter, we propose a DDSPCE-based surrogate model to estimate the expected minimum cost using much less samples compared to MC-based approaches. The proposed method can effectively estimate the statistics as well as the PDF and CDF of the minimum cost $Q(P_q, P_w)$ purely from the wind power data.

III. DATA-DRIVEN BASED POLYNOMIAL CHAOS EXPANSION

In this section, a DDSPCE-based surrogate model, i.e., a linear combination of multivariate orthogonal polynomials, will be developed to represent the relationship between the input random variables P_w and the minimum cost $Q(P_g, P_w)$.

A. Polynomial Chaos Expansion

Consider an independent random vector $\boldsymbol{\xi} = \{\xi_1, ..., \xi_M\} \in \mathbb{R}^M$, serving as the input to the system, then the model response $Y = f(\boldsymbol{\xi})$, being a function of $\boldsymbol{\xi}$, is also stochastic. In the SED problem, $\boldsymbol{\xi}$ can be obtained by decorrelating the wind power P_w through, for example, the principal component analysis (PCA), while the model response Y corresponds to the minimum production cost Q. It is claimed by the Cameron–Martin theorem that the model output Y can be expressed by K numbers of the expansion terms constructed by the orthogonal polynomial basis of $\boldsymbol{\xi}$ [2]:

$$Y = f(\boldsymbol{\xi}) \approx \sum_{i=1}^{K} c_i \Phi_i(\boldsymbol{\xi}) \tag{3}$$

where c_i are the unknown expansion coefficients to be determined; $\Phi_i(\boldsymbol{\xi})$ are the multidimensional polynomial basis, which are orthogonal to the joint marginal distribution of $\boldsymbol{\xi}$. Note that when $K \to \infty$, the series in (3) converges in the sense of the \mathcal{L}^2 -norm. For practical implementation, $\Phi_i(\boldsymbol{\xi})$ is generally truncated to a finite number of expansion terms:

$$\Phi_i(\boldsymbol{\xi}) = \prod_{j=1}^{M} \phi_j^{(i_j)}(\xi_j)$$
(4)

where $\Phi_i(\boldsymbol{\xi})$ is formed by a tensor product of univariate orthogonal polynomial basis $\phi_j^{(i_j)}(\xi_j)$, $j = \{1, \dots, M\}$. $\phi_j^{(i_j)}(\xi_j)$ is the one-dimensional polynomial basis of ξ_j with degree i_j at *i*-th expansion term. i_j denotes the degree of *j*-th univariate polynomial basis at *i*-th expansion term.

Note that for random inputs with thick tailed distributions or nonlinear correlations, the PCA may provide bias results. To overcome this, vine copula and Rosenblatt transform may be applied to model the dependence and decouple the components of the input data [6]. The kernal PCA can also be applied to handle nonlinear correlations.

The essence of the proposed DDSPCE method and its relation to the original SED problem (1)-(2) is presented in Fig. 1. Instead of running MC simulations on the SED problem (1)-(2) using a large number of scenarios, we first evaluate $Q(P_g, P_w)$ using a small number of scenarios of P_w . Next,

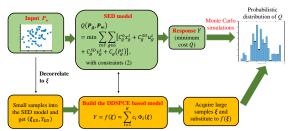


Fig. 1. Relation between the SED and DDSPCE-based surrogate model we use the obtained inputs P_w and outputs $Q(P_g, P_w)$ to build a DDSPCE-based surrogate model (3), a pure algebraic equation. Finally, we acquire a large number of scenarios of P_w , the production cost $Q(P_g, P_w)$ of which can be quickly calculated by substituting the detailed numbers into the established DDSPCE-based surrogate model (3). Compared to the SED problem (1)-(2), the DDSPCE-based surrogate model is much faster to evaluate and thus can save significant computational efforts. To build the DDSPCE-based surrogate model, our main tasks are to build the multidimensional polynomial basis $\Phi_i(\boldsymbol{\xi})$ using the available data of P_w and then to calculate the coefficients c_i in (3).

B. Constructing the Polynomials from Moments

For simplicity, we use $\phi_j^{(l)}(\xi_j)$ to replace $\phi_j^{(i_j)}(\xi_j)$ hereafter. The polynomial $\phi_j^{(l)}(\xi_j)$ of degree l is defined as:

$$\phi_j^{(l)}(\xi_j) = \sum_{k=0}^{l} p_{k,j}^{(l)} \xi_j^k, \quad l = \{0, 1, \cdots, D\}$$
(5)

where $p_{k,j}^{(l)}$ are the coefficients of polynomial $\phi_j^{(l)}(\xi_j)$ in the *k*-th degree. *D* is the order of the PCE-based model.

Traditionally, the orthogonal polynomial $\phi_j^{(l)}(\xi_j)$ can be determined by the Weiner-Askey scheme based on some typical distribution types of ξ_j [7]. However, exact knowledge of probability distribution of random inputs may hard to be acquired in practice, while raw data sets (e.g., wind power [8]) are more likely to be available. In this letter, the main idea of the data-driven SPCE method is to calculate the coefficients $p_{k,j}^{(l)}$ based on a few statistical moments that can be directly estimated from limited data. The exact probability distributions of ξ_j are not required to be known or even exist. ξ_j can be either continuous, discrete, or mixed and can be specified either by raw data sets, histogram or probability distributions.

The polynomial basis $\phi_j^{(l)}(\xi_j)$ needs to satisfy the orthogonality condition

$$\int_{\Omega} \phi_j^{(m)}(\xi_j) \phi_j^{(l)}(\xi_j) d\Gamma(\xi_j) = 0 \quad \forall m \neq l$$
(6)

where Γ is the marginal cumulative distribution function of ξ_j . If we let the coefficients of leading terms for all polynomials to be 1, i.e., $p_{l,j}^{(l)} = 1, \forall l$ in (5), and utilize the definition of the k-th raw moment of ξ_j :

$$\iota_{k,j} = \int_{\xi_j \in \Omega} \xi_j^k d\Gamma(\xi_j) \tag{7}$$

It can be shown that the coefficients $p_{k,j}^{(l)}$ in (5) can be calculated based on the following matrix equation [9]:

$$\begin{bmatrix} \mu_{0,j} & \mu_{1,j} & \dots & \mu_{l,j} \\ \mu_{1,j} & \mu_{2,j} & \dots & \mu_{l+1,j} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{l-1,j} & \mu_{l,j} & \dots & \mu_{2l-1,j} \\ 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} p_{0,j}^{(l)} \\ p_{1,j}^{(l)} \\ \vdots \\ p_{l-1,j}^{(l)} \\ p_{l}^{(l)} \\ p_{l}^{(l)} \\ p_{l}^{(l)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$
(8)

where $\mu_{m,j}$ is the raw statistical moments of ξ_j with $m = \{0, \dots, 2l - 1\}, j = \{1, \dots, M\}, l = \{0, 1, \dots, D\}$. $\mu_{m,j}$ can be easily calculated from the given raw data set or probabilistic distribution of ξ_j . Besides, the raw moment matrix in (8) has to be nonsingular. This condition is satisfied when finite statistical moments up to 2D-1 order exist as well as D or more distinct values are included in the data set, if $\mu_{m,j}$ is evaluated directly from data. Generally, $\boldsymbol{\xi}$ represented by a data set can satisfy these conditions easily, since the data points are finite and the degree D is small (typically \leq 5).

points are finite and the degree D is small (typically ≤ 5). Once the coefficients $p_{k,j}^{(l)}$ are solved from (8), the polynomial basis $\phi_j^{(l)}(\xi_j)$ and the multidimensional polynomial basis $\Phi_i(\boldsymbol{\xi})$ can be obtained from (5) and (4) in sequence. To build the PCE model (3), we need to solve the coefficients c_i . To this end, the orthogonal matching pursuit (OMP) is applied to find the best polynomial sets and the corresponding coefficients.

C. Calculating c_i by Orthogonal Matching Pursuit (OMP)

OMP is an iterative algorithm, which selects regressors that are most correlated to current approximation residual and adds them to the active set of basis in each iteration and then updates the coefficients c_i for all active regressors by the ordinary least squares (OLS) method. During the selection, the leave-one-out (LOO) error estimator ((1.26) in [10]) is chosen as a criterion for the model order D and the sparse candidate basis. Detailed description of the OMP procedure can be found in [10]. The OMP algorithm can achieve a sparse PCE-based surrogate model, which reduces the effort while guaranteeing the accuracy.

IV. SIMULATION STUDIES

In this section, we test the proposed DDSPCE method on an integrated electricity and gas system (IEGS), i.e., the IEEE 118-bus system integrated with a 20-node gas system. 5 wind farms are added into the system at bus {2, 33, 51, 81, 108} using the NREL's Western Wind Data Set [8]. The time period considered is 24 hours, e.g., $T = \{1, \dots, 24\}$. Thus, the wind generator output P_w is a 120-dimension random vector (24 time periods for each wind farm). Further configuration of the IEGS (i.e., the power and gas network data, the wind power data, and the load profile) can be found on: https://github.com/ TxiaoWang/DDSPCE-based-Stochastic-ED.git.

To test the performance of the proposed DDSPCE-based surrogate model, we compare the probabilistic characteristics (e.g., mean μ , standard deviation σ , the PDF and the CDF) of the minimum cost Q estimated from the DDSPCE-based surrogate model with those by the SPCE method [7] (e.g., with inferred PDF [11]) and those from the benchmark 10,000sample MC simulations. Note that the SPCE-based model in this simulation is built based on the PDF inferred from the available data while the proposed method is built directly from data. It can be seen in Table I and Fig. 2 that both the proposed DDSPCE method and the SPCE method can provide good estimations for the probabilistic characteristics of the minimum cost Q, though the proposed method possesses a better accuracy in the estimations.

Particularly, the DDSPCE-based surrogate model (3) is constructed using only 1, 100 samples (e.g., number of training sample $\mathcal{N} = 1100$). The time consumption of the DDSPCE method is only about $\frac{1}{9}$ of the time consumed by the MC simulations. However, if only mean and variance are needed while the detailed PDF is not of interest, only 300 samples ($\mathcal{N} = 2.5M$) are required to achieve accurate estimations (see Table II)), which takes only about $\frac{1}{33}$ of the time required by the MC simulations. Besides, the SPCE method takes 8.83s more than the proposed method due to the additional PDF inferring procedure.

Comparison of the estimated statistics of Q by the MC simulations, the DDSPCE and the SPCE methods with $\mathcal{N}=1100.$

Index	MC	DDSPCE	SPCE
μ	7.3276×10^{6}	7.3276×10^{6}	7.3277×10^{6}
σ	4.7069×10^4	4.7159×10^{4}	4.8600×10^4
$\frac{\Delta\mu}{\mu_{\rm MC}}\%$		2.0739×10^{-5}	1.8503×10^{-3}
$\frac{\Delta \sigma}{\sigma_{\rm MC}} \%$		$1.9218 imes10^{-1}$	3.2529

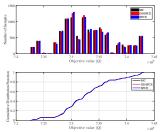


Fig. 2. Comparison of the PDF and CDF of the minimum cost Q by the MC simulations, the DDSPCE, and the SPCE methods with $\mathcal{N} = 1100$.

Remark: From our simulation experience, $\mathcal{N} \approx 2.5M$ is typically sufficient for accurate mean and variance. While for an accurate PDF estimation for a *unimodal* system response, $\mathcal{N} \approx 5M$ is typically required. For an accurate PDF estimation for a *multimodal* system response, more samples (e.g., $\mathcal{N} \approx 9M$ in this letter) may be required.

TABLE II Comparison of the estimated statistics of Q by the MC simulations, the DDSPCE and the SPCE methods with $\mathcal{N} = 300$.

Index	MC	DDSPCE	SPCE
μ	7.3276×10^{6}	7.3276×10^{6}	7.3277×10^{6}
σ	4.7069×10^{4}	4.7228×10^{4}	5.0379×10^{4}
$\frac{\Delta \mu}{\mu_{\rm MC}}\%$		4.8945×10^{-4}	1.7092×10^{-3}
$\frac{\Delta \sigma}{\sigma_{\rm MC}}\%$		$3.3939 imes 10^{-1}$	7.0318

V. CONCLUSIONS

In this letter, we propose a data-driven sparse PCE (DDSPCE)-based surrogate model for the SED problem. The probability characteristics of the SED solution (i.e., minimum production cost) can be approximated by the DDSPCE-based surrogate model without pre-assumed probability distribution of random inputs. Simulation results on an IEGS system (the IEEE 118-bus system integrated with a 20-node gas system) verify that the proposed DDSPCE method can provide accurate estimations for the mean, variance, PDF and CDF of the SED solution accurately and efficiently, even when the probability distribution of the SED solution is multimodal.

Particularly, compared to the method in [9], orthogonal matching pursuit (OMP) is applied in this letter to find the coefficients of the DDSPCE method, which may achieve a faster convergence than the method in [9]. Besides, unlike in [9] where both the random inputs and the model response

(Total Transfer Capability) are unimodal, the random inputs from real-world data and the model response (the minimum cost Q) of this letter are both multimodal, showing the accuracy and efficiency of the DDSCPE method in handling more generalized situations.

APPENDIX

For the IESG system, the multi-period SED problem under fixed UC decisions can be represented by:

$$Q(\boldsymbol{P_g}, \boldsymbol{P_w}) = \min_{\boldsymbol{P_g}} \sum_{t \in T} \left(\sum_{g \in G} C_g(\boldsymbol{P_g}^t) + \sum_{s \in S} C_s(\boldsymbol{g_s^t}) \right)$$
(9)

s.t.

(

Constraints
$$(2a) - (2d)$$

$$G_s^{\min} \le g_s^t \le G_s^{\max} \quad \forall s \in S, t \in T$$
 (10a)

$$G_a^{\min} \le \pi_a^t \le G_a^{\max} \quad \forall a \in G_a, t \in T$$
(10b)

$$g_b^t = W_b \sqrt{((\pi_{e(b)}^t)^2 - (\pi_{a(b)}^t)^2)} \quad \forall b \in G_b, t \in T$$
(10a)

$$\pi_{e(c)}^t \le \alpha_c \pi_{a(c)}^t \quad \forall c \in G_c, t \in T$$
(10d)

$$0 \le g_b^t \le G_b \quad \forall b \in G_b, t \in T$$
 (10e)

$$0 \le g_c^t \le G_c \quad \forall c \in G_c, t \in T \tag{10f}$$

$$\sum_{s \in G_s} g_{s(a)}^t + \sum_{b_1 \in G_b} g_{b_1(a)}^t - \sum_{b_2 \in G_b} g_{b_2(a)}^t + \sum_{c_1 \in G_c} g_{c_1(a)}^t \\ - \sum_{c_2 \in G_c} g_{c_2(a)}^t = \sum_{d \in G_d} G_{d(a)}^t + \sum_{g \in G} \Theta_g P_{g(a)}^t \quad t \in T$$
(10g)

where $C_s(\cdot)$ is the cost of gas well. Gas network constraints are shown in (10), where (10a) denotes the output capacity limits of a gas well; constraint (10b) denotes the nodal pressure range; constraint (10c) describes the gas flow g_b^t in gas passive pipeline b; constraint (10d) is the simplified gas compressor model; a(c); constraints (10e)-(10f) denote the gas flow transmission capacity limits in a gas passive pipeline and gas compressor, respectively; constraint (10g) denotes the gas nodal balance. The detailed notations can be found in [5].

REFERENCES

- C. Safta, R. L.-Y. Chen, H. N. Najm, A. Pinar and J.-P. Watson, "Efficient uncertainty quantification in stochastic economic dispatch," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2535–2546, Jul. 2017.
- [2] J. Li, N. Ou, G. Lin and W. Wei, "Compressive sensing based stochastic economic dispatch with high penetration renewables," *IEEE Trans. Power Syst.*, vol. 34, no. 2, pp. 1438–1449, March 2019.
- [3] Z. Hu, Y. Xu, M. Korkali, X. Chen, L. Mili and J. Valinejad, "A Bayesian approach for estimating uncertainty in stochastic economic dispatch considering wind power penetration," *IEEE Trans. Sustain. Energy*, vol. 12, no. 1, pp. 671–681, Jan. 2021.
- [4] M. Rajabi, "Review and comparison of two meta-model-based uncertainty propagation analysis methods in groundwater applications: polynomial chaos expansion and Gaussian process emulation," *Stoch Environ Res Risk Assess*, vol. 33, no. 2, pp. 607–631, 2019.
- [5] R. Liu, Y. Hou, Y. Li, S. Lei, W. Wei and X. Wang, "Sample robust scheduling of electricity-gas systems under wind power uncertainty," *IEEE Trans. Power Syst.*, 2021.
- [6] E. Torre, S. Marelli, P. Embrechts, and B. Sudret, "A general framework for data-driven uncertainty quantification under complex input dependencies using vine copulas," *Probabilistic Engineering Mechanics*, vol. 55, pp. 1–16, 2019.
- [7] H. Sheng and X. Wang, "Applying polynomial chaos expansion to assess probabilistic available delivery capability for distribution networks with renewables," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 6726–6735, Nov. 2018.
- [8] National Renewable Energy Laboratory (NREL). Western Wind Data Set. Accessed on: Jan 18, 2021. [Online]. Available: https://www.nrel. gov/grid/western-wind-data.html

- [9] X. Wang, X. Wang, H. Sheng and X. Lin, "A data-driven sparse polynomial chaos expansion method to assess probabilistic total transfer capability for power systems with renewables," *IEEE Trans. Power Syst.*, vol. 36, no. 3, pp. 2573-2583, May 2021.
- [10] S. Marelli and B. Sudret, "UQLab user manual polynomial chaos expansions," Report UQLab-V1.1-104, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, 2018.
- [11] E. Torre, S. Marelli, B. Sudret, "UQLab user manual Statistical inference," Report UQLab-V1.4-114, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, 2021.