Gaussian Process Learning-based Probabilistic Optimal Power Flow

Pareek, Parikshit; Nguyen, Hung D.

2021

Pareek, P. & Nguyen, H. D. (2021). Gaussian Process Learning-based Probabilistic Optimal Power Flow. IEEE Transactions On Power Systems, 36(1), 541-544.

<https://dx.doi.org/10.1109/TPWRS.2020.3031765>

<https://hdl.handle.net/10356/150725>

<https://doi.org/10.1109/TPWRS.2020.3031765>

© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at:
<https://doi.org/10.1109/TPWRS.2020.3031765>

Downloaded on 26 Apr 2023 07:11:27 SGT

Gaussian Process Learning-based Probabilistic Optimal Power Flow

Parikshit Pareek, *Student Member, IEEE*, and Hung D. Nguyen* *Member, IEEE*

Abstract—In this letter, we present a novel Gaussian Process Learning-based Probabilistic Optimal Power Flow (GP-POPF) for solving POPF under renewable and load uncertainties of arbitrary distribution. The proposed method relies on a non-parametric Bayesian inference-based uncertainty propagation approach, called Gaussian Process (GP). We also suggest a new type of sensitivity called Subspace-wise Sensitivity, using observations on the interpretability of GP-POPF hyperparameters. The simulation results on 14-bus and 30-bus systems show that the proposed method provides reasonably accurate solutions when compared with Monte-Carlo Simulations (MCS) solutions at different levels of uncertain renewable penetration and load uncertainties. The proposed method requires a lesser number of samples and elapsed time. The non-parametric nature of the proposed method is manifested by testing uncertain injections that follow various distributions in the 118-bus system. The low error value results verify the proposed method's capability of working with different types of input uncertainty distributions.

Index Terms—Gaussian Process Regression, Probabilistic Optimal Power Flow, Interpretable Models

I. INTRODUCTION

THE upsurge in the renewable penetration and dynamic loads has made *Probabilistic Optimal Power Flow* (POPF) a necessary tool providing necessary uncertainty description in the decision and state variables [1]. Existing POPF methods fall under analytical, approximate, and Monte-Carlo Simulation (MCS) based categories [1], [2]. These methods have various limitations, like dependencies on approximate power flow formulations, complicated implementation, or lacking data-based guarantees and substantial sample set requirements. Further, mostly these methods are developed to handle specific input uncertainty distribution such as normal, beta, etc. This becomes a bottleneck in case of limited uncertainty information or when a random variable does not follow any such distribution. The second case is prevalent with Solar Photo-Voltaic (PV) based renewable generation and Electric Vehicle load as their uncertainty forecasting remains a challenge.

We introduce a novel POPF solution method relying on the Bayesian inference based uncertainty propagation technique, Gaussian Process (GP). The GP is used extensively as a supervised learning tool in various machine learning applications [3], [4]. The proposed *Gaussian Process Learning-based Probabilistic Optimal Power Flow* (GP-POPF) method can handle main issues of uncertainty description requirement and a large number of sample requirements of data-based methods. GP-POPF can be employed to handle arbitrary input uncertainty as, during the training of GP models, specific

details of uncertainty distribution are not required. This independence from input uncertainty descriptions is due to the non-parametric nature of the GP [4]. This feature allows GP-based probabilistic method to obtain output variable distributions for various input uncertainty distribution types [5]. Further, GP provides mean and variance information for a random variable at testing without requiring the distribution information of the random input. As the GP's foundation is the Bayesian inference, the associated interpretability makes GP suitable for physical network learning such as that in the proposed GP-POPF. To the best of our knowledge, the GP has not been explored for the task of uncertainty quantification task in OPF yet. Here, we do not attempt to learn the deterministic relationship between any one input and output. Our attempt is to learn the output behaviour under input uncertainty. The proposed GP-POPF— compared with other ML methods used for power system optimization [6]— offers advantages in terms of providing closed-form relationship, non-parametric nature [5] and model interpretability [7].

The main contributions of this letter can be summarized as:

- Developing a novel GP-POPF which is a non-parametric in nature, thus be free from a need of pre-defining class of uncertainties.
- Developing the foundation of interpretability in POPF by defining covariance function hyperparameter based on the newly introduced concept of *Subspace-wise Sensitivity*.

We introduce the basic Gaussian Process Regression below.

A. Gaussian Process Regression (GPR)

The GPR is a non-parametric modeling method providing some level of interpretability of models and allowing to model prior understanding in the data-based models [4], [8]. The GPR interpretability means that upon learning, the covariance matrix contains subtle information about the function it was destined to learn. This information can be then utilized to predict the function behavior inside as well as outside of the input domain within a limited range. Further, as GPR allows us to provide prior distribution, understanding of the physical network obtained using classical methods can help greatly to improve this data-based method.

A general GP regression, with a training data set $\mathcal{D} = \{\mathbf{x}^i, \hat{f}(\mathbf{x}^i)\}_{i=1}^N$ where $\hat{f}(\mathbf{x}^i) = \hat{y}(\mathbf{x}^i)$ is the measured function value at input $\mathbf{x}^i \in \mathbb{R}^n$ at the i -th step, is given as [4]:

$$\hat{y}(\mathbf{x}^i) = y(\mathbf{x}^i) + \varepsilon^i, \quad i = 1 \dots N. \quad (1)$$

In the POPF problem, vector $\mathbf{x}$ contains uncertain power injections, load demands, and renewable generations, while $\hat{y}$ refers to the optimal ACOPF output corresponding to the input point $\mathbf{x}$. The optimal output $\hat{y}$ can be optimal cost, generator dispatch, and node voltage values obtained with ACOPF

*Corresponding Author

Authors are with School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore. *paree0001,hunghtd@ntu.edu.sg*

solution. In (1), ε^i are independent and identically distributed noise variable with zero-mean, σ_n standard deviation normal distribution. Interested reader can look into [4] for details of GP fundamentals and compact expressions of mean and variance values of the posterior distribution. The GPR establishes the relationship between input and output variables via the covariance function $k(\mathbf{x}^i, \mathbf{x}^j)$. The kernel function is selected based upon the prior understanding of the problem at hand. We use squared exponential (SE) covariance to model the POPF under the load and renewable generation uncertainties due to the smoothness of kernel. The SE covariance is given as:

$$k(\mathbf{x}^i, \mathbf{x}^j) = \sigma_f^2 \exp \{ -0.5(\mathbf{x}^i - \mathbf{x}^j)^T M(\mathbf{x}^i - \mathbf{x}^j) \}. \quad (2)$$

Here, $M = l^{-2}I$ is a matrix having *characteristic length* (l) while σ_f^2 is the normalised *scaling factor*. Both, l and σ_f are collectively called *hyperparameters*. These *hyperparameters* contain the subtle information about the function, providing interpretations on behaviour.

II. PROPOSED GP-POPF

The basic deterministic OPF problem can be expressed as:

$$\begin{aligned} \min \quad & c(\mathbf{y}) \\ \text{s.t.} \quad & g(\mathbf{x}, \mathbf{y}) = 0; \quad h(\mathbf{x}, \mathbf{y}) \leq 0 \end{aligned} \quad (3)$$

Here, $\mathbf{y}$ is the output set including generation and voltage setpoints while $c(\cdot)$ is the associated cost function. $\mathbf{x}$ represents the input vector combining the load and renewable generations. $g(\mathbf{x}, \mathbf{y})$ refers to the non-convex power balance constraint while $h(\mathbf{x}, \mathbf{y})$ represents other operational inequality constraints like the upper and lower limits of the variables (see [9]). With random $\mathbf{x}$, the deterministic OPF in (3) can be cast as Probabilistic OPF.

In the proposed GP-POPF, GPR is used as an uncertainty propagation method, i.e., obtaining the output variable distribution of OPF for a given input uncertainty. For this, the first stage of the proposed GP-POPF is to construct a learning data set $\mathcal{D}_l = \{\mathbf{X}, \hat{\mathbf{Y}}\}$ by solving (3) for N different input vectors respectively. Here, $\mathbf{X} \in \mathbf{R}^{N \times n}$ with each row being one n -dimensional uncertain input vector, $\mathbf{x} = [\mathbf{P}_r^T \mathbf{P}_d^T \mathbf{Q}_d^T]$ with $\mathbf{P}_r$, $\mathbf{P}_d$ and $\mathbf{Q}_d$ are uncertain renewable generation, real and reactive demand vectors. The output matrix $\hat{\mathbf{Y}} \in \mathbf{R}^{N \times m}$ contains columns of m different output variables as optimal cost $c(\mathbf{P}_g^o)$, optimal generator set points $(\mathbf{P}_g^o, \mathbf{Q}_g^o)$ and node voltages $\mathbf{V}^o$, corresponding to each uncertain input vector in $\mathbf{X}$. The GPR is used to learn all these output variables independently and in parallel for a better computational performance. For the construction of uncertain input set, we use uniform distribution with a box type uncertainty description as $\mathbf{x} \in [\mathbf{x}^-, \mathbf{x}^+]$. In the present work, we assume that the network topology is unchanged, and thus the admittance matrix is constant. Further, we do not limit the sampling at particular power injection locations. Therefore, we can vary the sampling of the power injection vector to reflect both the location of uncertain power injections and the amount of power injections. By maximizing the Log marginal likelihood [4], we obtain the mapping $f_j : \mathbf{R}^n \mapsto \mathbf{R}$ for each individual output variable as $y_j = f_j(\mathbf{x})$ for $j = 1, \dots, m$.

One can learn the voltage magnitude relationship, for example, at j -th bus, with random renewable generation vector $\mathbf{P}_r$. Then based on the function-space view of GPR (Chapter 2 [4]) the mean prediction of voltage magnitude output V_j for an arbitrary, sample input vector $\mathbf{P}_r^s \in [\mathbf{P}_r^-, \mathbf{P}_r^+]$ is given as:

$$V_j(\mathbf{P}_r^s) = k_j(\mathbf{X}, \mathbf{P}_r^s)^T \boldsymbol{\alpha}_j. \quad (4)$$

Here, $\boldsymbol{\alpha}_j = (K_j + \sigma_n I)^{-1} \hat{\mathbf{y}}_j$ with $K \in \mathbf{R}^{N \times N}$ being the covariance matrix obtained on input variables from $\mathcal{D}_l$, and $k_j(\mathbf{X}, \mathbf{P}_r^s) \in \mathbf{R}^{N \times 1}$ is kernel vector. Similarly, the variance at test input vector $\mathbf{P}_r^s$ can be obtained. Here, it is important to highlight that once training is done, $\boldsymbol{\alpha}_j$ remains constant. Thus, repeated matrix inversion is not required. The form of function $V_j(\mathbf{P}_r^s)$ in (4) can be understood as an optimal linear combination approximation of “basic kernel” $k_j(\mathbf{X}, \mathbf{P}_r^s)$. The set of $\boldsymbol{\alpha}_j$ ’s represent the optimal weights or coefficients that best fits the concerned function $y_j = f_j(\mathbf{x})$. For a new input point $\mathbf{P}_r^s$, one just needs to plug in such new value of $\mathbf{P}_r^s$ into (4) to calculate the optimal output solution $V_j(\mathbf{P}_r^s)$. Similarly, voltage-reactive power and other relationships can be learnt.

The main advantages of the proposed GP-POPF method over existing techniques can be summarized as follows:

- At the learning stage, no assumption is made on the type of uncertainty distribution for the input vector $\mathbf{x}$.
- During prediction, the GP-POPF method provides output distributions for any input uncertainty distribution.
- Unlike other uncertainty propagation methods, GP provides variance information directly, which can be used to obtain probabilistic limits on output variations.
- GP is an Interpretable Machine Learning (IML) tool. Thus, *hyperparameters* can be used to obtain more insights into GP-POPF such as relative output variations.
- Proposed GP-POPF makes no assumption in power flow equations for uncertainty propagation. Full ACOPF models, with different objectives and constraints, can be utilized to understand system behavior.

III. RESULTS AND DISCUSSION

In this section, we present the results obtained on the IEEE 14-bus and 30-bus system [10]. The 14-bus system has three renewable generators connected at bus 7, 9, and 14 with total of 11 uncertain load nodes. The 30-bus system has five renewable generators at 6, 9, 22, 25, and 28-th bus. All 21 load buses uncertain apparent power demand. All the renewable generators have an uncertainty of 100%, thus covering entire uncertain space and does not require any specification of the type of generator and power distribution. The MATPOWER ACOPF solutions (from *runopf*) [11] are used to obtain OPF solutions using $\mathbf{x} = [\mathbf{P}_r^T \mathbf{P}_d^T \mathbf{Q}_d^T]$. The testing is performed with MCS on 10^4 random input samples. As the proposed Bayesian inference-based GPR is a non-parametric method, the proposed POPF method can be used to get inference for any distribution of uncertainty once the model is trained.

Table I contains the error in mean and standard deviation of optimal cost $c(\mathbf{P}_g^o)$ distribution considering different levels of renewable penetration with $\pm 10\%$ load uncertainty. The error values show that the proposed GP-POPF has been able

TABLE I

ERROR IN OPTIMAL COST PARAMETERS WITH $\pm 10\%$ LOAD UNCERTAINTY

System	Penetration	% Error in $\mu(cost)$	% Error in $\sigma(cost)$
14-Bus	10.42%	0.000066	0.001700
14-Bus	17.37%	0.000124	0.003000
30-Bus	8.82%	0.000085	0.000108
30-Bus	24.70%	0.000004	0.001400

to achieve very high accuracy in cost estimation. Further, we define $\%L_1$ error as the L_1 norm distance between MCS solution $\hat{\mathbf{y}}$ and GP-POPF solution $\mathbf{y}$ as $\|\hat{\mathbf{y}} - \mathbf{y}\|_1 / \|\hat{\mathbf{y}}\|_1 \times 100$. This error indicates the mismatch between true and estimated solution. The Fig. 1 presents $\%L_1$ error distribution for vector $\mathbf{P}_g$ while Fig. 2 shows the error distribution corresponding to $\mathbf{V}$ for different system cases. Clearly, the mean of error distribution is close to zero, while the number of samples decreases very fast as we move to higher error bins. Further, the low values of $\%L_1$ error establish the applicability of the proposed GP-POPF as the predicted output decision, and state variable will require very fewer adjustments in case they do not satisfy physical and operational constraints. The detailed adjustment method development is not in the scope of this letter and will be developed subsequently.

A prominent feature of the proposed GP-POPF is its applicability with different types of input uncertainty distributions. The results depicting this non-parametric nature of the proposed method are shown in Table II. The table contains the 118-bus system results wherein three 260MW distributed generators (DGs) are located at bus 59, 80, and 90. Further, we present the error in the optimal cost distribution's mean and standard deviation in Table II. We use Weibull distribution for wind speed with parameters and correlation given in [2]. As listed in Table VII of [2], the PEM based method in [2] obtains in the mean value of cost distribution as 0.033% and error in standard deviation as 6.972% with three correlated wind speed Weibull distributions and 10% load uncertainty. Thus, Table II shows that the proposed GP-POPF method achieves smaller error than the PEM based method, particularly for standard deviation. We also perform a simulation with normal distribution for wind speed in the 118-bus system. We recorded errors in cost distribution mean as 0.00053% and standard deviation as 0.0212%. These results establish the non-parametric nature, i.e., the capability to handle various types of input uncertainty distributions of the proposed GP-POPF method.

Fig. 1. $\%L_1$ error in $\mathbf{P}_g$, with $\pm 10\%$ load uncertainty, for 10^4 sample MCS

TABLE II

ERROR IN OPTIMAL COST PARAMETERS WITH $\pm 10\%$ LOAD UNCERTAINTY FOR 118-BUS SYSTEM

Approach	DG Distribution	% Error $\mu(cost)$	% Error $\sigma(cost)$
PEM [2]	Weibull	0.033	6.972
GP-POPF	Weibull	0.0023	0.0122

The Stochastic OPF (SOPF) problem, particularly chance-constrained formulation, has also been used to calculate the mean and standard deviation of the OPF outputs under various uncertainties. The linear approximation methods are used as analytical methods [12], [13] while the Polynomial Chaos Expansion (PCE) method is used to handle the non-linearity in power flow equation in chance-constrained formulation [14]. Table III shows the error in mean and standard estimation for 30-Bus system for using the same setting as in [14] with standard deviation 0.10 and results reported in Table-IV of [14]. The error definition for real power mean is [14] $\|\Delta\mu\|_\infty = \|\mathbb{E}[P_g]_{AC} - \mathbb{E}[P_g]_{GP}\|_\infty$. The proposed GP-POPF method shows improved results compared to the linearized AC power flow-based method [15] and equivalent results with the PCE-based method without requiring distribution information during training. This error comparison indicates that the proposed GP-POPF can replace the PCE method in SOPF, particularly with unknown uncertainty distribution type.

We report the computation time with unoptimized codes below. As training can be done offline and in parallel, GP will not be a direct bottleneck. For the simulations, we require $N \leq 300$ training samples for the systems at different levels of renewable penetration. GP-POPF takes 4.28s for the 14-bus system and 4.67s for the 30-bus system to complete training and prediction over 10^4 testing samples using parallel execution by unoptimized codes. On the contrary, the MCS takes 245.82s for 14-bus and 357.08s for 30-bus system for 10^4 sample results. In Table II, the optimal cost function learning step takes 1.61s, and the inference for 10^4 samples takes 0.186s on the average. The same number of deterministic OPF samples run as MCS takes about 211.28s, which is much higher than the computational time used by the proposed method. The training is done with $N = 350$ for the 118-bus system. Further, when GP-POPF is once trained, it can be used for many predictions using (4). For elevating the elapsed time issue, methods reviewed in [16] can be considered. We use the GPML toolbox with MATLAB 2018b on PC having Intel Xeon E5-1630v4@3.70 GHz, 16 GB RAM. We opt to use the OPF function provided in the MATPOWER, which can be executed with *runopf* function, as it is used by various works to solve deterministic OPF. Other deterministic methods can also be utilized for individual deterministic instances. We record 8.59s for running 350 instances of *runopf* used for learning results of 118-bus system given in Table II. We suggest to learn the proposed GP-POPF on real data from the

TABLE III
ERROR IN COMPUTED MOMENTS USING PCE, LINEARIZATION AND PROPOSED GP-POPF FOR 30-BUS SYSTEM

Method	P_g		$ V $	
	$\ \Delta\mu\ _\infty$	$\ \Delta\sigma\ _\infty$	$\ \Delta\mu\ _\infty$	$\ \Delta\sigma\ _\infty$
PCE [14]	10.1E-5	0.4E-5	2.2E-5	1.0E-5
Lin. AC [15]	411.0E-5	12.9E-5	105.8E-5	7.5E-5
GP-POPF	0.81E-5	0.12E-5	0.27E-5	2.46E-5

Fig. 2. $\%L_1$ error in $|V|$, with $\pm 10\%$ load uncertainty, for 10^4 sample MCS OPF solutions solved by the system operator.

A. Observations on Interpretability: Subspace-wise Sensitivity

As GP is an Interpretable Machine Learning tool, the proposed GP-POPF can be used to understand the steady-state behavior of the power system under uncertainty. One such observation is on the range of variations in generator setpoints with uncertain x . The interpretation is targeted to check if there exists any relationship between kernel *hyperparameters*, l and σ_f in (2), and P_g variation, $\Delta P_g = \max\{P_g\} - \min\{P_g\}$, over input uncertain subspace where the ΔP_g gets affected simultaneously by the cost function, non-linearity of power balance and various inequalities in POPF. The two different *hyperparameters* are indicative of different features of the function $f_j(x)$. The *characteristic length* l indicate "wiggle" lengths and higher value of l means the function varies slower (or less) with respect to variations in input. The σ_f indicates the average distance of function from its mean. As both *hyperparameters* are obtained together using optimization, it is important to interpret them with each other. Therefore, we define ratio of these hyperparameters as $\gamma = l/\sigma_f$ which indicates the normalised variations in $f_j(x)$ within the input subspace. The lower value of γ means that the function is highly non-linear and has large variations from its mean. Figure 3 depicts observation establishing this interpretation which shows an inverse relationship between γ and ΔP_g , for all generators. This means that a generator with less non-linearity i.e., higher γ , will have fewer variations in P_g . γ can be considered as the output sensitivity over an input subspace as *hyperparameters* are obtained by optimization over the entire input subspace $x \in [x^-, x^+]$. This is much different than the traditional derivative-based sensitivity calculations which are valid only at that operating point. We call it *Subspace-wise Sensitivity* γ and will be explored in the future.

IV. CONCLUSION

Different from existing analytical methods, the proposed GP-POPF method does not rely on uncertainty information and linearization assumptions on the power flow. Compared to data-based methods, the proposed method does not require extensive training samples of POPF solutions, thus reducing computation time. The Bayesian inference-based GP is used for uncertainty propagation and allows one to explore a new

Fig. 3. Inverse relationship of ΔP_g with *Subspace Sensitivity* $\gamma = l/\sigma_f$ for all the generators in 30-bus system at different load uncertainties.

path for interpretability for the first time through the concept of *Subspace-wise Sensitivity*. The latter opens up the possibility of exploring interpretable models for uncertainty handling in power system operation and control applications for small to medium scale networks. Some applications of the proposed method (not limited to) are: 1) conversion of the probabilistic chance-constraints into deterministic ones, 2) bound tightening via learning 3) handling multiple uncertainties via non-parametric formulation of micro-grid dispatch, and 4) constraint screening etc.

REFERENCES

- [1] W. Sun, M. Zamani, M. R. Hesamzadeh, and H.-T. Zhang, "Data-driven probabilistic optimal power flow with nonparametric bayesian modeling and inference," *IEEE Trans. on Smart Grid*, 2019.
- [2] Y. Li, W. Li, W. Yan, J. Yu, and X. Zhao, "Probabilistic optimal power flow considering correlations of wind speeds following different distributions," *IEEE Trans. on Power Systems*, vol. 29, no. 4, pp. 1847–1854, 2014.
- [3] R. M. Sanner and J. E. Slotine, "Gaussian networks for direct adaptive control," *IEEE Transactions on Neural Networks*, vol. 3, no. 6, pp. 837–863, 1992.
- [4] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.
- [5] Y. Xu, Z. Hu, L. Mili, M. Korkali, and X. Chen, "Probabilistic power-flow calculation based on a novel gaussian process emulator," *IEEE Trans. on Power Systems*, 2020.
- [6] G. Ruan *et al.*, "Review of learning-assisted power system optimization," *arXiv preprint arXiv:2007.00210*, 2020.
- [7] C.-K. Lu, S. C.-H. Yang, X. Hao, and P. Shafto, "Interpretable deep gaussian processes with moments," in *International Conference on Artificial Intelligence and Statistics*, 2020, pp. 613–623.
- [8] P. Pareek and H. D. Nguyen, "Probabilistic robust small-signal stability framework using gaussian process learning," *Electric Power Systems Research*, vol. 188, 2020.
- [9] S. Misra, D. K. Molzahn, and K. Dvijotham, "Optimal adaptive linearizations of the ac power flow equations," in *2018 Power Systems Computation Conference (PSCC)*. IEEE, 2018, pp. 1–7.
- [10] S. Babaeinejadsarookolae *et al.*, "The power grid library for benchmarking ac optimal power flow algorithms," *arXiv preprint arXiv:1908.02788*, 2019.
- [11] Zimmerman *et al.*, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. on power systems*, vol. 26, no. 1, pp. 12–19, 2010.
- [12] T. Mühlfordt, T. Faulwasser, and V. Hagenmeyer, "A generalized framework for chance-constrained optimal power flow," *Sustainable Energy, Grids and Networks*, vol. 16, pp. 231–242, 2018.
- [13] L. Roald, F. Oldewurtel, B. Van Parys, and G. Andersson, "Security constrained optimal power flow with distributionally robust chance constraints," *arXiv preprint arXiv:1508.06061*, 2015.
- [14] T. Mühlfordt *et al.*, "Chance-constrained ac optimal power flow: A polynomial chaos approach," *IEEE Trans. on Power Systems*, vol. 34, no. 6, pp. 4806–4816, 2019.
- [15] L. Roald and G. Andersson, "Chance-constrained ac optimal power flow: Reformulations and efficient algorithms," *IEEE Trans. on Power Systems*, vol. 33, no. 3, pp. 2906–2918, 2017.
- [16] H. Liu, Y.-S. Ong, X. Shen, and J. Cai, "When gaussian process meets big data: A review of scalable gps," *IEEE Trans. on Neural Networks and Learning Systems*, 2020.