Mapping Rule Estimation for Power Flow Analysis in Distribution Grids

Jiafan Yu, Student Member, IEEE, Yang Weng, Member, IEEE, and Ram Rajagopal, Member, IEEE

Abstract—The increasing integration of distributed energy resources (DERs) calls for new monitoring and operational planning tools to ensure stability and sustainability in distribution grids. One idea is to use existing monitoring tools in transmission grids and some primary distribution grids. However, they usually depend on the knowledge of the system model, e.g., the topology and line parameters, which may be unavailable in primary and secondary distribution grids. Furthermore, a utility usually has limited modeling ability of active controllers for solar panels as they may belong to a third party like residential customers. To solve the modeling problem in traditional power flow analysis, we propose a data-driven approach to reveal the mapping rules between different variables and recover useful variables based on physical understanding and data mining. Specifically, we illustrate how to use our machine learning tool to conduct 1) system monitoring with asynchronous measurements in distribution grids with DERs and 2) short-term prediction of potential voltage violation for operational planning. We demonstrate the superior performance of our method through extensive numerical validation on different scales of distribution grids. Finally, we analyze the robustness of our method with satisfactory results.

I. INTRODUCTION

Electric grids are undergoing a profound change. Renewables and other distributed energy resources (DERs) are expected to supply more than 50% of electricity demand by 2050 in various parts of the world [1], [2]. Deep penetration of DERs adds new capabilities and significantly affects the operations of distribution grids. In such distribution networks, proper monitoring will be needed for detecting outages [3], cyber attacks [4], and system failures [5]. In addition to monitoring, operational planning is needed for predicting over-voltage, calculating economic dispatch [6], and conducting short-term grid controls [7]–[9].

Power flow-based methods can be used for monitoring and planning in distribution grids [10], [11]. Recent research augments traditional power flow equations by using historical data to initialize state estimators and solvers [12]–[14], modifying the current system models [15], [16] and proposing novel multi-objective optimization formulations [17]. Power flow-based analysis and state-estimation methods require the knowledge of system topology and network parameters. Such knowledge is available in well-maintained primary distribution grids and limited secondary distribution grids.

In many primary and secondary distribution grids, the assumption of complete information does not hold. In secondary distribution grids, only the planned topology and switch locations are known, but real-time changes to the topology can be hard to track. Line parameter profiles are inaccurate or even missing. Even reconstructing the admittance matrix can be hard when using distribution management systems (DMS) such as CYME [18]. Future distribution networks will host a variety of active control devices ranging from voltage regulators to inverters for rooftop solar, EV charging and storage. These assets are usually independently owned and operated outside of the domain of the DMS. The control rules implemented by these devices are unavailable or can be hard to model, making the direct application of power flow analysis difficult and inaccurate [19] even when topology and line parameters are perfectly known. Incomplete system information and limited measurements as shown in Fig. 1 makes the system identification problem hard in practice. The availability of measurements from active devices, line sensors, smart meters, and PMUs [20] is an opportunity to overcome this challenge by designing scalable approaches for system monitoring and analysis relying on new types of data.

In this paper, we focus on data-driven alternatives to power flow analysis when distribution system topology, line parameters, and active device control rules are unavailable and are hard to obtain. We study how to design data-driven mapping rules between power network variables (voltage phasor, current phasor, and complex power) that can be learned from historical data from new types of measurements available today. The challenges to obtain such maps are that they should approximately satisfy the physical behavior of the system (e.g., power flow equations plus device dynamics) and the map should have some ability to generalize beyond the situations captured in the historical data. These requirements limit the types of mapping rules that can be considered. For example, neural networks can have trouble generalizing the map beyond ranges of system values observed in the training data.

We propose designing mapping rules adopting Support
Vector Regression (SVR) \[21\]. Physical laws are embedded in the mapping by careful choice of kernels that can capture power flow equations. The mapping is learned from historical data. The proposed SVR approach brings some advantages over conventional regression approaches \[22\]: the obtained model description is sparse and robust due to the nature of the SVR loss function \[23\]. SVR regressions can be computed very efficiently using interior point methods and distributed computing \[24\], \[25\] and many different kernels can be utilized depending on the applications \[26\].

We illustrate the proposed approach designing and evaluating mapping rules for two practical problems: generation of synchronous measurements for distribution networks based on measurements with different sampling rates \[27\], \[28\] and operational planning to provide over-voltage warnings in distribution networks. We test both the forward mapping estimation and inverse mapping estimation on different scales of distribution grids, e.g., IEEE 8, 123 bus, and systems with bus number between 8 and 123. In all simulations, the estimation results of our SVR model are compared with the results from other state-of-the-art models. The results reveal that the SVR model outperforms other models in comparison to other methods such as linear regression and power flow approaches \[22\]: the obtained estimation results from other state-of-the-art models. The results reveal the existing of a zero solution:

\[
\min ||(p, q) - f(v, \theta)||. \tag{4}
\]

Increasing penetration of DER adds a variety of active controllers whose control algorithms and device models might not be available to the power monitoring systems at the utility. For instance, if bus \(i\) in an \(n\)-bus distribution grid is equipped with a reactive power bank, where injected reactive power follows voltage variation, \(q'_i = h(v, \theta)\), the modified power flow equation at bus \(i\) changes to:

\[
q_i = \sum_{k=1}^{n} |v_i||v_k| (g_{ik} \sin \theta_{ik} - b_{ik} \cos \theta_{ik}) - q'_i, \tag{5a}
\]

where \(q'_i\) is the additional power injection from the reactive power bank, and \(h(\cdot)\) is the control policy. If \(q'_i\) is omitted from the model, an incorrect mapping will be obtained.

As an illustration, we add a reactive power controller at bus 4 for the IEEE 8-bus distribution grid and assume topology and line parameters are known. Fig. 2 shows the significant mean absolute error (MAE) appearing in the traditional power flow analysis when an active device is added but not modeled.

In addition to the problem of unmodeled active devices, traditional power flow analysis may also suffer from inaccurate topology information and missing line parameters in distribution grids. Instead, if historical data is available, the mapping between \((p, q)\) and \((v, \theta)\) can be learned using alternative approaches from statistical learning \[21\]. The challenge for any alternative is to capture the underlying physics of the power flow equations as well as being capable of generalizing the map accurately beyond the original range of the variables. Traditional benchmarks to compare any models are to utilize parameter learning for the power flow equations or polynomial regression models to relate the variables. In the remainder of the paper, we develop an alternative based on support vector regression and demonstrate its applicability.
III. SUPPORT VECTOR REGRESSION FOR POWER FLOW

A. The Basic SVR model

In order to solve (4), the function $f$ corresponding to the power flow can be expressed in a form that emphasizes the unknown coefficients $g_{ik}$ and $b_{ik}$:

$$
p_i = \sum_{k=1}^{n} g_{ik} (|v_i||v_k| \cos \theta_{ik}) + b_{ik} (|v_i||v_k| \sin \theta_{ik}),$$

(6a)

$$
q_i = \sum_{k=1}^{n} g_{ik} (|v_i||v_k| \sin \theta_{ik}) - b_{ik} (|v_i||v_k| \cos \theta_{ik}).
$$

(6b)

Define the power feature map for bus $i$ by $\phi_i(v, \theta)[k] = (|v_i||v_k| \cos \theta_{ik}; |v_i||v_k| \sin \theta_{ik})$. The power flow equations can then be redefined as:

$$
p_i = \langle (g_i, b_i), \phi_i(v, \theta) \rangle,$$

(7a)

$$
q_i = \langle (-b_i, g_i), \phi_i(v, \theta) \rangle,$$

(7b)

where $\langle \cdot , \cdot \rangle$ represents the inner product. Define $x = (v, \theta)$, $\omega^p_i = (g_i, b_i)$ and $\omega^q_i = (-b_i, g_i)$. If $y = p_i$, the real power balance at node $i$ can be written as $y = \langle \omega^p_i, \phi_i(x) \rangle$. If $y = q_i$, the reactive power balance is $y = \langle \omega^q_i, \phi_i(x) \rangle$. This forms the basis of a support vector regression model (SVR) [30] to learn the mapping rule $f$ in (3):

$$
\text{minimize} \quad \frac{1}{2} \|\omega\|^2 + C \sum_{t=1}^{T} \lambda_t (\|\xi_t\|_1 + \|\xi^*_t\|_1),
$$

(8a)

subject to:

$$
y_t - \langle \omega, \phi(x_t) \rangle - b_t \leq \xi_t + \xi^*_t,$$

(8b)

$$
y_t - \langle \omega, \phi(x_t) \rangle + b_t - y_t \leq \xi_t + \xi^*_t,$$

(8c)

$$
\xi_t, \xi^*_t \geq 0,
$$

(8d)

where $t = 1, \ldots, T$ indexes samples from historical data, $\lambda_t = 1$ and $\xi_t = \epsilon_t$. If $b_t = 0$ and $y$ is set to $p_i$, the real power flow for bus $i$ is recovered, and alternatively if it is set to $q_i$ the reactive power flow is recovered. SVR models are solved for real and reactive maps for each node $i$ in the network. An illustration of a typical SVR fit is shown in Fig. 3. The optimization formulation uncovers an $\epsilon$-insensitive zone. Points inside this zone contribute no error to the regression fit so $\xi = \xi^* = 0$. Points outside contribute to the error and define the regression.

B. SVR Power Flow

The SVR optimization in (8) is in general difficult to solve due to the large number of constraints and the dimension of the feature map $\phi(x)$. However, special choices of feature maps lead to a simple representation of the solutions for the SVR regression. These feature maps satisfy the kernel trick property $K(x, y) := \langle \phi(x), \phi(y) \rangle = h(\langle x, y \rangle)$, where $h$ is a positive scalar function [29]. The space of such feature maps satisfying this property is the reproducing Hilbert kernel space (RHKS). If the feature map is in RHKS, the solution to (8) is given by an optimal set of parameters $\alpha_t^*$ with the same dimension of $y$ for each $t$ and:

$$
y = \sum_{t=1}^{T} \alpha_t K(x, x_t),
$$

(9)

which corresponds to the feature mapping $\phi(x) = [x_{n_1}, \ldots, x_1, \sqrt{2}x_{n_1}x_{n_1-1}, \ldots, \sqrt{2}x_{n_1}x_1, \ldots, \sqrt{2}x_1x_1, \sqrt{2}x_1x_{n_1}, \ldots, \sqrt{2}x_1x_1, c]$. Setting $x = [\cos(\theta_1), \sin(\theta_1), \ldots, \sin(\theta_n)]$ enables the SVR map with a quadratic kernel to represent the power flow equations exactly by choosing an appropriate $\omega$. Fig. 3 presents a data dependent power flow mapping that avoids explicit modeling of topology and network parameters. The mapping is exact for each node when the parameters $\alpha_t$ are accurately estimated. In practice, we shift all angles by $\pi/4$ to reduce numerical errors.

The proposed methodology generalizes directly when there are only partial measurements available. In such cases, the map $\{x\}$ can include $p_j$ and $q_j$ ($j \neq i$) for modeling variables in node $i$. In the absence of phase angle measurements, we utilize higher order polynomial kernels setting $x = \{|v_i|\}$, which are accurate if phase angle fluctuations are small.

We use Fig. 4 to show that our learning has the locality property from the power flow equation (1). For example, the power injection at the $i$-th bus in an $n$-bus distribution grid only depends on the voltages at the bus of interest and the directly connected buses (the first red dashed circle in Fig. 4), according to the power flow equation (1).

C. Choosing Tuning Parameters for SVR Power Flow

Cross-validation is typically utilized in SVR to chose the tuning parameters $C$ and $\epsilon$ in (8) [31]. This enables the method to increase robustness towards noise and outliers in the data. It also ensures that SVR has good predictive performance.

The suggested approach for SVR-based Power Flow is to utilize $k$-fold cross-validation [31] (typically $k = 5$) with the
training data to select the optimal choices of $C$ and $\epsilon$, with $k-1$ blocks of data used to train the model and one block utilized to assess validation performance and select tuning parameters. The SVR performance is then assessed in a separate data set. The choice of parameters determines the sparsity of $\alpha_t$ in the kernel representation (10).

D. Generalizing Power Flow SVR: Inverse Maps

In many applications of power flow analysis, we are interested in recovering voltage magnitude and phase angle information from the measurements of real and reactive power. Typically, a calibrated power flow model is utilized and solved. Power flow solutions are not guaranteed to be unique, and thus, locally, it can be approximated by a polynomial flow is a differentiable map as a function of real and reactive power from historical data. The inverse power function of power is then assessed in a separate data set. The choice of parameters determines the sparsity of $\alpha_t$ in the kernel representation (10).

DWSVR utilizes the temporal and spatial correlations simultaneously.

E. Doubly Weighted SVR (DWSVR): Robustness to Outliers

The proposed data-driven approach relies on access to accurate historical data. However, measurements and communication in distribution grids are less reliable than those in transmission grids. Therefore, measurement outliers are more widespread.

Temporal and spatial correlations from the historical data can be utilized to mitigate the impact of data points with large errors. The weighted SVR (12) can be utilized by setting values for each $\epsilon_t$ in (8) to reduce the no penalty tube (the area between the blue dashed curves in Fig. 3) for more recent data points $\epsilon_t = \exp^{-1}\left(-\frac{(\tau_t - \tau_T)^2}{\theta_1^2} - \theta_2 \sin^2(\theta_3(\tau_t - \tau_T))\right)$, (12)

where $\tau_t$ is the timestamp of the $t$-th data. The correlation function captures the feature and is widely used in machine learning. $\theta_1$, $\theta_2$, and $\theta_3$ are hyperparameters that are used to modify the weight modulation intensity. If $\theta_1 \rightarrow \infty$, $\theta_2 = 0$, then the weight $\epsilon_t = 1$ for $t = 1, \ldots, T$, and the problem reduces to a regular SVR problem.

Additional reduction of the impact of outliers can be obtained utilizing the objective function weights $\lambda_t$ in (8). We reduce the penalty of training samples that are far from the mean of training data points. The weights are:

$$
\lambda_t = (1 - \theta_4) \left(1 - \frac{D((p_t, q_t), (\bar{p}, \bar{q})) - D_{\text{min}}}{D_{\text{max}} - D_{\text{min}}}\right)^2 + \theta_4,
$$

(13)

where $D((p_t, q_t), (\bar{p}, \bar{q}))$ is the distance and $(\bar{p}, \bar{q})$ are the average real and reactive power in the training set, respectively. $D_{\text{min}}$ and $D_{\text{max}}$ are the minimum and maximum distances, and $\theta_4$ is a hyperparameter modifying the weights. If $\theta_4 = 1$, then $\lambda_t = 1$ and reduces to regular SVR. (13) is monotonic with respect to distances $\bar{D}$. The resulting DWSVR utilizes the temporal and spatial correlations simultaneously.

IV. Experimental Results

A. Approach

In the following, we illustrate two use-case examples of our learning algorithms.

1) Forward Mapping-System Data Synchronization: In distribution grids, the sampling frequency of Phasor Measurement Units (PMUs) is much higher than the sampling rate of smart meters. The “power flow analysis” is necessary to synchronize the data stream of different types of measurements from PMUs, PV inverters, and smart meters. In particular, the system data synchronization task uses the historical data of voltage phasors and power injections to recover the mapping from voltage phasors to power injections without requiring the topology or parameters. After obtaining the mapping, we can recover the power injection from the voltage measurements.

2) Inverse Mapping-Voltage Violation Warning: In distribution grids, a voltage violation warning is essential for short-term planning for renewable integration. However, existing techniques only enable us to make reasonably good forecasting for the load and distributed energy generation. Therefore, we propose to recover the voltage magnitude information from the measurements of real and reactive power, so that system operators can identify potential voltage violations via load forecasts. Specifically, we deploy regular power injection (electricity load) forecasting and apply the learned mappings from historical data to estimate future voltages at different buses. When the estimated voltage is either higher than 1.05 times the regular voltage or lower than 0.95 times the regular voltage, a voltage violation warning is triggered.

B. Experiment Setup

We test our data-driven power flow approach on a variety of settings and real-world data sets. We use 8, 16, 32, 64, 96, 123-bus test feeders, as well as two Southern California Edison (SCE) distribution networks with different shapes. Here, the
16, 32, 64, 96-bus systems are extracted from the IEEE 123-bus system. The bus power injection data is from primary distribution grids of Southern California Edison (SCE) and secondary distribution grids of Pacific Gas and Electric (PG&E). The real and reactive power injection data are from small and medium business or aggregate of several residential homes. The sampling frequency is one hour. The SCE data set’s period is from January 1, 2015, to December 31, 2015, and the PG&E data set’s period is from August 1, 2010, to July 1, 2011. For IEEE standard test feeders, we run power flow using Matpower package to obtain the associated voltage magnitudes and phase angles at each bus. For SCE distribution networks, the voltage phase angle information are available on some buses. We run power flow to obtain the voltage and phase angle information at other buses. In our experiments, the topology and line parameter information from case files are only used for data preparation via the Matpower-based power flow to obtain the voltage and phase angle information. In all evaluation steps, we assume that the topology information and the line parameters are unavailable. Finally, noises are added to check the robustness of the proposed approach.

C. Forward Mapping: From \( v, \theta \) to \( p, q \)

1) The Effectiveness of SVR Model with 2nd Order Polynomial Kernel: In particular, the power flow equations can be represented exactly by the proposed SVR model with the 2nd order polynomial kernel in (10), when we choose the rectangular coordinate representation of the state vector (11). In this simulation, we build the proposed SVR model with 2nd order polynomial kernel, other SVR models with different kernels, as well as the widely used regression model and the averaging model. The input variables (11), the voltage phasors, and the output variable, power injections at certain buses, are all the same for different models. We use four weeks’ historical data of voltage phasors and power injections to train different models and validate the performance using another three weeks’ data. After estimating the power injection based on the voltage phasors over testing data using different models, we compare the performance of different methods by calculating the root-mean-square error (RMSE) of the power injection estimations for different models.

The result for 123-bus case is shown in Fig. 5. It is clear that the performance of the SVR models are better than the regression model and averaging model. Among the SVR models with different kernels, the 2nd order polynomial kernel provides the smallest RMSE, supporting the theoretical deduction in Section III-B.

2) The Performance Over Different Distribution Grids: We evaluate the performance of the proposed SVR model, the widely used regression method, and the straightforward averaging method on various scales of distribution grids from 8-bus distribution grid to 123-bus distribution grid. We also test the performances on 123-bus distribution grid with mesh network by manually connecting several buses to mimic some urban systems’ weakly meshed structure. For all of the evaluation, four weeks’ data is used for training, and three weeks’ data is used for testing. We randomly choose five different buses to estimate their power injections. RMSE is used for evaluating the performance. The results are shown in Table I. For all test cases, the RMSEs of SVR model are better than regression model and averaging model. For the computational time, training SVR models for all cases is within seconds. This is fast enough for real-time updating.

| Test Case       | RMSE (p.u.)          | Time Cost (s) |
|-----------------|----------------------|--------------|
|                 | SVR        | Reg        | Ave | SVR      | Reg        | Ave |
| 8-Bus           | 0.023      | 0.060      | 0.058 | 14.5     | 0.0010     | 0.0003 |
| 16-Bus          | 0.030      | 0.060      | 0.057 | 13.3     | 0.0011     | 0.0002 |
| 32-Bus          | 0.031      | 0.060      | 0.057 | 13.3     | 0.0025     | 0.0007 |
| 64-Bus          | 0.035      | 0.060      | 0.058 | 14.1     | 0.0010     | 0.0009 |
| 96-Bus          | 0.040      | 0.060      | 0.057 | 14.0     | 0.0002     | 0.0005 |
| 123-Bus         | 0.055      | 0.061      | 0.060 | 15.0     | 0.015      | 0.0005 |
| 123-Bus w/ loop | 0.050      | 0.062      | 0.058 | 15.0     | 0.009      | 0.0006 |

3) Noise Performance: In practice, most of the measurements in electric grids are with measurement errors. Therefore, we conduct experiments to investigate the performance of different models versus the measurement error levels. We manually add zero-mean Gaussian noises with different variances on the training set and test the performance of the learned mapping rule on a noiseless test set. The mean absolute error (MAE) is chosen for performance evaluation. We compare the performance of the proposed SVR model and the widely used regression model. First, the MAE increasingly follows the increase of the measurement error for both methods as we can see in Fig. 6. Furthermore, with small errors, SVR method is much better than the linear model. The linear model is better only when the error completely corrupts the signal as expected.
4) **Topology Availability:** Traditional power flow mapping relies on the knowledge of topology and line parameters. If we know the topology, in the power flow equation (6), $g_{ij}$ and $b_{ij}$ are non-zero only if bus-$i$ and bus-$j$ are neighbors. Therefore, if we know the topology of the distribution grid, we can shorten the input state vector (11) and decrease the computation time. However, the proposed mapping rule estimation approach is purely data-driven. The performance of the proposed SVR model is the same no matter whether the topology information is given or not. The experimental results are shown in Fig. 7, where we compare the performance of forward mapping estimation for both SVR model and regression model with and without topology information. Though the topology information does help a little for regression model, it does not improve the performance of SVR model.

D. The Robustness of the Forward Mapping Estimation

In this section, we test the robustness of the forward mapping estimation by SVR model when (1) training and test data are in different ranges, (2) only partial buses’ measurements are available, and (3) there are outliers in the data.

1) **Extrapolation of SVR Model:** A power system is a dynamic system, and the load values change significantly over time, especially when it is with different DER penetration levels. Therefore, we train the proposed SVR model and the regression model by using a fixed training set, where all the real power injections are within the range of $[-1 \text{ p.u.}, 0]$ to obtain the mapping rule from voltage phasors to power injections. Then, we test the performances of the learned mapping rules in different power injection levels. Fig. 8 demonstrates the performance of the two models, where the mean absolute error (MAE) is used to evaluate the performance.

Fig. 8a shows the MAEs of estimating the real power injection when the actual power injection range in the test set is the same as the training set. When the actual power injection is between $-0.6 \text{ p.u.}$ and $-0.4 \text{ p.u.}$ (same range for training and test), the performances of the regression model and the SVR model are both good. When the actual power injection is around $-1 \text{ p.u.}$ or 0 (slightly different ranges for training and test), the performance of the SVR model is much better than the regression model. In this case, the performance of the regression model is worse but still acceptable (error is smaller than $0.05 \text{ p.u.}$). However, when the range of the testing set is different from the training set, the linear model performs poorly as shown in Fig. 8c. In contrast, the performance of the SVR model is much better than the regression model.

In addition to the variation of PV generations in Fig. 8c, we consider the load variation in Fig. 8b. It shows that the SVR model is very robust in such case, and the associated absolute error is always less than $0.1 \text{ p.u.}$ in various test cases. On the other side, the absolute error of regression model could be as high as $0.3 \text{ p.u.}$ The data for Fig. 8 is in Table II.

2) **Learning the Forward Mapping from Partial Buses in a Distribution Grid:** In addition to exploring the variation in time (temporal domain), we also look into the change of available data from neighbors in the spatial domain. This is important when we only have partial data in a distribution grid. The MAE versus the maximum available degree of neighbors is shown in Fig. 9. As we can see, the first-degree neighbors (the directly connected buses) are more than enough to recover the mapping from $(v, \theta)$ to $p_i$. This is consistent with the
TABLE II: Data Synchronization: Mapping from Voltage Phasors to Power Injection

| Range     | MAE Reg | MAE SVR | Range     | MAE Reg | MAE SVR |
|-----------|---------|---------|-----------|---------|---------|
| [-2.0, -1.8) | 0.2036  | 0.0185  | [-0.4, -0.2) | 0.0135  | 0.0035  |
| [-1.8, -1.6) | 0.1625  | 0.0225  | [-0.2, 0.0) | 0.0245  | 0.0048  |
| [-1.6, -1.4) | 0.1448  | 0.0258  | [0.0, 0.2)  | 0.0467  | 0.0229  |
| [-1.4, -1.2) | 0.1444  | 0.0320  | [0.2, 0.4)  | 0.0631  | 0.0204  |
| [-1.2, -1.0) | 0.0833  | 0.0325  | [0.4, 0.6)  | 0.0792  | 0.0166  |
| [-1.0, -0.8) | 0.0292  | 0.0045  | [0.6, 0.8)  | 0.1019  | 0.0155  |
| [-0.8, -0.6) | 0.0129  | 0.0031  | [0.8, 1.0)  | 0.1102  | 0.0174  |
| [-0.6, -0.4) | 0.0043  | 0.0030  |           |         |         |

In distribution grids, bad data exists in different measurements. Therefore, we evaluate the performance when there are historical outliers. While significant improvement is observed for our method in all testing cases, we choose the learned mapping rule from \((v, \theta)\) to \(p, q\) as an example.

The numerical validation of the support vector selection is shown in Fig. 10. For this test, we randomly add 5% bad data in the training set. Fig. 10 visualizes the support vectors among all training data points. The \(x\)-axis is the training data point index, and the \(y\)-axis is the magnitude of the associated dual Lagrangian multipliers. If a training data point is not a support vector, the dual Lagrangian multiplier is zero. If a training data point is a support vector, the dual Lagrangian multiplier is nonzero. In addition, we mark the outlier data points with a black cross. Fig. 10 shows that whenever a bad data appears (black cross), our coefficient is nonzero. This means that the bad data is located.

Furthermore, we modify the relative weights of training data samples following Section III-B’s formulation to further alleviate the impact of outliers. We observe significant improvements over testing cases when comparing to the regular SVR. Fig. 11 shows the performances versus different percentages of outliers among five different SVR models. The weighted SVR always performs the best.

**E. Inverse Mapping: From \(p, q\) to \(|v_i|\)**

For completeness, we test the performance of the SVR model on inverse mapping: from \(p, q\) to \(|v_i|\) for voltage violation prediction and the robustness on different scenarios.

1) The effectiveness of the SVR model: Section III-D illustrates using SVR models to recover the inverse mapping of the power flow: from power injections to voltage magnitudes. We also validate the effectiveness of the SVR model for estimating the inverse mapping. The settings of the numerical validation are very similar to the settings in Section IV-C1. Four weeks’ data is used for training and another three weeks’ data is used for testing. For SVR models, we use 5-fold cross-validation to determine the best hyper parameters \(\lambda\) and \(\epsilon\). The only difference is that, for inverse mapping, the input vector is \((p, q)\) and the output variable is the voltage magnitude at a certain bus. The result is shown in Fig. 12. Differing from the forward mapping, the SVR model with 1st 2nd order polynomial kernel have the best performance.

2) The Performance Over Different Distribution Grids: We evaluate the performances of different models for inverse mapping (from power injections to voltage magnitudes) on various scales of distribution grids from 8-bus to 123-bus distribution grid. We also test the performance on 123-bus distribution grid with loops. The results are in Table III. Similar to the results of the forward mapping estimation, the proposed SVR model outperforms other models on all test cases. The computational times for different models are also
similar to the forward mapping estimation.

**TABLE III: Benchmark of Inverse Mapping**

| Test Case      | RMSE (10^{-3} p.u.) | Time Cost (s) |
|----------------|---------------------|---------------|
|                | SVR                | Reg           | SVR       | Reg           |
| 8-Bus          | 0.20               | 0.66          | 2.1       | 13.0          | 0.0012        | 0.0005       |
| 16-Bus         | 0.0061             | 0.12          | 3.0       | 13.6          | 0.0016        | 0.0006       |
| 32-Bus         | 0.18               | 0.38          | 2.5       | 13.5          | 0.0068        | 0.0003       |
| 64-Bus         | 0.60               | 1.4           | 1.8       | 11.7          | 0.026         | 0.0005       |
| 96-Bus         | 1.1                | 3.2           | 2.6       | 13.2          | 0.057         | 0.0006       |
| 123-Bus        | 1.9                | 6.0           | 2.8       | 12.0          | 0.095         | 0.0008       |
| 123-Bus w/ loop| 1.9                | 6.5           | 2.8       | 12.2          | 0.10          | 0.0008       |

3) Extrapolation of SVR Model: We also test the extrapolation ability of the SVR model for the inverse mapping from power injections to the voltage magnitude at a certain bus of the distribution grid. Similar to case one, we investigate the performance of the proposed model in different power injection levels. Table [IV] presents the detailed results for inverse mapping estimation. When the training data and testing data are in the same range, the performances of the SVR model is better than the linear regression model, while the error of the regression model is still relatively small, e.g., MAE is less than 0.002 p.u. However, when the power injection range of the testing set is different from the range in the training set, the performance of the regression model degrades significantly, while the SVR model retains good performance.

**TABLE IV: Voltage Estimate: Mapping from Power Injections to Voltage Magnitude**

| Real Power Injection Range | Voltage Magnitude Range | MAE of Regression | MAE of SVR |
|---------------------------|-------------------------|-------------------|------------|
| [-2.0, -1.0)              | [0.75, 0.90)            | 0.0180            | 0.0028     |
|                           | [0.90, 0.95)            | 0.0170            | 0.0016     |
|                           | [0.95, 1.00)            | 0.0167            | 0.0007     |
|                           | [1.00, 1.10)            | 0.0157            | 0.0004     |
| [-1.0, 0.0)               | [0.80, 0.95)            | 0.0007            | 0.0002     |
|                           | [0.95, 1.00)            | 0.0005            | 0.0002     |
|                           | [1.00, 1.05)            | 0.0003            | 0.0003     |
|                           | [1.05, 1.10)            | 0.0004            | 0.0003     |
| [0.0, 1.0)                | [0.90, 1.00)            | 0.0165            | 0.0005     |
|                           | [1.00, 1.05)            | 0.0155            | 0.0004     |
|                           | [1.05, 1.10)            | 0.0145            | 0.0011     |
|                           | [1.10, 1.20)            | 0.0126            | 0.0013     |

4) Partial Information Available: Finally, we show the performances along the maximum available degree of neighbors for inverse mapping in Fig. [13]. Different from the results in forward mapping, learning the mapping rules from power injections to the voltage magnitude is not localized.

**V. Conclusion**

With deep DER penetration in distribution grids, proper monitoring with current sensor capability is needed. As the topology, parameters, and active controller information are usually insufficient in some primary distribution grids and many secondary distribution grids, it is hard to apply the transmission grid monitoring tools directly. In this paper, we propose to use the historical data from smart meters, micro PMUs, and inverter sensors for data-driven monitoring and operational planning in the distribution grids. Especially, we propose an SVR based model to learn the mapping rules of different measurements with embedded physical understanding. To show the application of the proposed approach, we show two use cases: 1) Monitoring: recover system variables from asynchronous measurements; 2) Planning: provide over-voltage warnings for operational planning. Numerical results on different distribution grids show that our method not only works when traditional methods fail due to incomplete system information but also needs only the local information rather than information from the whole network. This makes our method robust and fast for distribution grid automation.

**References**

[1] European Commission, “Energy Roadmap 2050.”

[2] M. Z. Jacobson, M. A. Delucchi, G. Bazouin, Z. A. F. Bauer, C. C. Heavey, E. Fisher, S. B. Morris, D. J. Y. Pickutowski, T. A. Vencill, and T. W. Yeskoo, “100% clean and renewable wind, water, and sunlight (WWS) all-sector energy roadmaps for the 50 United States,” Energy & Environmental Science, vol. 8, no. 7, pp. 2093–2117, 2015.

[3] Y. Zhao, R. Sevlian, R. Rajagopal, A. Goldsmith, and H. V. Poor, “Outage detection in power distribution networks with optimally-deployed power flow sensors,” IEEE PES General Meeting, 2013.

[4] D. Deka, R. Baldick, and S. V. Vishwanath, “Attacking power grids with secure meters: The case for breakers and jammers,” IEEE Conference on Computer Communications Workshops, pp. 646–651, 2014.

[5] C. Rudin, D. Waltz, R. N. Anderson, A. Boulanger, A. Salleb-Aouissi, M. Chow, H. Dutta, P. N. Gross, B. Huang, and S. Jerome, “Machine learning for the new york city power grid,” IEEE transactions on pattern analysis and machine intelligence, vol. 34, no. 2, pp. 328–345, 2012.

[6] A. Vargas and M. E. Samper, “Real-time monitoring and economic volt/var control with distributed power flow solver in smart grid,” IEEE Trans. on Smart Grid, vol. 3, no. 2, pp. 866–877, 2012.

[7] X. Zhang, A. J. Flueck, and C. P. Nguyen, “Agent-based distributed volt/vvar control with distributed power flow solver in smart grid,” IEEE Transactions on Smart Grid, vol. 7, no. 2, pp. 600–607, 2016.
[8] P. Jahangiri and D. C. Atiprantis, “Distributed volt/var control by pv inverters,” IEEE Transactions on power systems, vol. 28, no. 3, pp. 3429–3439, 2013.

[9] H.-G. Yeh, D. F. Gayme, and S. H. Low, “Adaptive var control for distribution circuits with photovoltaic generators,” IEEE Transactions on Power Systems, vol. 27, no. 3, pp. 1656–1663, 2012.

[10] F. C. Schweppe and J. Wildes, “Power system static-state estimation, part i: Exact model,” IEEE Transactions on Power Apparatus and systems, no. 1, pp. 120–125, 1970.

[11] G. Andersson, “Modelling and analysis of electric power systems,” EEH-Power Systems Laboratory, Swiss Federal Institute of Technology (ETH), Zürich, Switzerland, 2008.

[12] Y. Weng, R. Negi, and M. D. Ilic, “A search method for obtaining initial guesses for smart grid state estimation,” IEEE Third International Conference on Smart Grid Communications, pp. 599–604, 2012.

[13] ——, “Historical data-driven state estimation for electric power systems,” IEEE International Conference on Smart Grid Communications, pp. 97–102, 2013.

[14] Y. Weng, R. Negi, C. Faloutsos, and M. D. Ilic, “Robust data-driven state estimation for smart grid,” IEEE Transactions on Smart Grid, vol. PP, no. 99, pp. 1–12, 2016.

[15] A. Abdel-Majeed, S. Tenbohlen, D. Schöllhorn, and M. Braun, “Development of state estimator for low voltage networks using smart meters measurement data,” IEEE Grenoble PowerTech, pp. 1–6, 2013.

[16] S. Wang, W. Gao, and A. S. Meliopoulos, “An alternative method for power system dynamic state estimation based on unscented transform,” IEEE Transactions on Power Systems, vol. 27, no. 2, pp. 942–950, 2012.

[17] S. Sarri, M. Paolone, R. Cherkauki, A. Borghetti, F. Napolitano, and C. A. Nucci, “State estimation of active distribution networks: Comparison between wls and iterated kalman-filter algorithm integrating pnums,” IEEE PES Innovative Smart Grid Technologies Europe, pp. 1–8, 2012.

[18] “CYME International T&D.” http://www.cyme.com.

[19] M. Fan, V. Vittal, G. T. Heydt, and R. Ayyanar, “Probabilistic power flow studies for transmission systems with photovoltaic generation using cumulants,” IEEE Transactions on Power Systems, vol. 27, no. 4, pp. 2251–2261, 2012.

[20] J. Sexauer, P. Javanbakht, and S. Mohagheghi, “Phasor measurement units for the distribution grid: Necessity and benefits,” IEEE PES Innovative Smart Grid Technologies (ISGT), pp. 1–6, 2013.

[21] J. Friedman, T. Hastie, and R. Tibshirani, The elements of statistical learning. Springer series in statistics Springer, Berlin, 2001.

[22] L. S. Moulin, A. A. Da Silva, M. El-Sharkawi, and R. J. Marks, “Support vector machines for transient stability analysis of large-scale power systems,” IEEE Transactions on Power Systems, vol. 19, no. 2, pp. 818–825, 2004.

[23] A. Smola and V. Vapnik, “Support vector regression machines,” Advances in neural information processing systems, vol. 9, pp. 155–161, 1997.

[24] A. K. Menon, “Large-scale support vector machines: algorithms and theory,” Research Exam. UC San Diego, pp. 1–17, 2009.

[25] C.-H. Ho and C.-J. Lin, “Large-scale linear support vector regression,” Journal of Machine Learning Research, vol. 13, pp. 3323–3348, 2012.

[26] C. M. Bishop, “Pattern recognition,” Machine Learning, vol. 128, 2006.

[27] F. Blaabjerg, R. Teodorescu, M. Liserre, and A. V. Timbus, “Overview of control and grid synchronization for distributed power generation systems,” IEEE Transactions on industrial electronics, vol. 53, no. 5, pp. 1398–1409, 2006.

[28] P. Yang, Z. Tan, A. Wiesel, and A. Nehorai, “Power system state estimation using pnums with imperfect synchronization,” IEEE Transactions on Power Systems, vol. 28, no. 4, pp. 4162–4172, 2013.

[29] B. C. Lesieutre, D. K. Molzahn, A. R. Borden, and C. L. DeMarco, “Examining the limits of the application of semidefinite programming to power flow problems,” Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 1492–1499, 2011.

[30] B. Schölkopf and A. J. Smola, Learning with kernels: Data mining, inference, and prediction, Feb 2002.

[31] T. Hastie, R. Tibshirani, and J. Friedman, The elements of statistical learning: Data mining, inference, and prediction,” 2009.

[32] X. Han and L. Clemmensen, “On weighted support vector regression,” Quality and Reliability Eng. Intl., vol. 30, no. 6, pp. 891–903, 2014.

[33] J. Yu, Y. Weng, C.-W. Tan, and R. Rajagopal, “Probabilistic estimation of the potentials of intervention-based demand side energy management,” IEEE International Conference on Smart Grid Communications, pp. 865–870, 2015.

[34] C. Josz, S. Fliscounakis, J. Maeght, and P. Panciatici, “AC power flow data in matpower and qcqp format: itesla, rte snapshots, and pegase,” arXiv preprint arXiv:1603.01533, 2016.