Identification of Intraday False Data Injection Attack on DER Dispatch Signals

Jip Kim†, Siddharth Bhela‡, James Anderson†, and Gil Zussman†
†Department of Electrical Engineering, Columbia University, New York, NY 10027
‡ Siemens Technology, Princeton, NJ 08540

***Abstract*— The urgent need for the decarbonization of power girds has accelerated the integration of renewable energy. Concurrently the increasing distributed energy resources (DER) and advanced metering infrastructures (AMI) have transformed the power grids into a more sophisticated cyber-physical system with numerous communication devices. While these transitions provide economic and environmental value, they also impose increased risk of cyber attacks and operational challenges. This paper investigates the vulnerability of the power grids with high renewable penetration against an intraday false data injection (FDI) attack on DER dispatch signals and proposes a kernel support vector regression (SVR) based detection model as a countermeasure. The intraday FDI attack scenario and the detection model are demonstrated in a numerical experiment using the HCE 187-bus test system.**

I. Introduction

Along with the rapid deployment of distributed energy resources (DERs), power system operation heavily relies on information communication technologies (ICT) as shown in Fig. 1 [1]. DERs such as energy storage and small-scale generators receive dispatch signals from the energy management system (EMS) and consumer-side resources such as electric vehicles and demand response also contribute to the increase in the number of ICT devices involved in power grid operations. While the DERs generally provide economic benefits to the electricity suppliers and consumers, the power grid is more exposed to cyber attacks through the ICT devices. For example, in 2015, a coordinated cyber attack against the ICT network and devices in the distribution grid caused blackouts in three different regions in Ukraine for more than seven hours [2].

Meanwhile, renewable integration and decarbonization of the power grid have significantly increased the system's operational difficulties. The intermittent nature of clean, but weather-dependent energy resources such as wind turbine and photovoltaic (PV) generators increase the need in ramping resources to handle the variability and system reserves to address the subsequent uncertainty. At the same time, the profitability of conventional flexibility resources such as coal and gas power plants has been aggravated, making it more challenging to secure an adequate amount of balancing resources [3]. In practice, California ISO which is known for its aggressive PV integration is facing the so-called *duck curve* phenomenon which refers to the shape of the net load profile with a high solar penetration including a rapid ramp-up due to the reduction in solar generation around sunset. The maximum three-hour ramping requirement was predicted to be 13GW/3hr for the year 2020 in 2013, but the actual value turned out to be 17GW/3hr [4]. This trend is currently observed among the renewable pioneers such as power grids in California and Texas, and will accelerate along with the installation of more PV generation. The steep net load ramp-up requires the power grid to operate with small system reliability and resilience margins and increases the system vulnerability to unexpected events such as cyber attacks.

This material is based upon work supported in part by NSF grants CNS-2148128 and EPCN-2144634, and by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy under the Solar Energy Technology Office Award Number DE-EE0008769. The views expressed herein do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

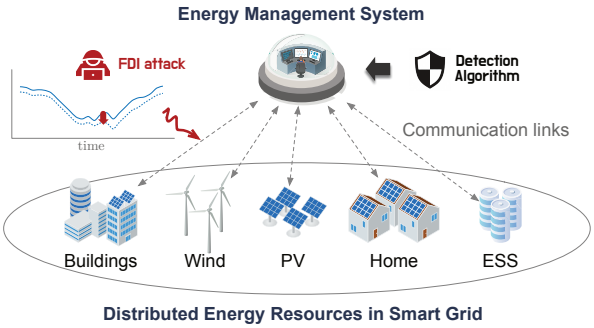


Fig. 1. Various DERs and connecting communication links to the EMS. Potential intraday FDI attacks targeting the communication links and detection algorithm for EMS are also illustrated.

A. Related work

A range of cyber attacks in power grids have been investigated in the literature [5]–[16]. A *denial of service* (DoS) attack that renders the entire communication network in the grid unavailable by injecting meaningless packets was presented in [5]–[7]. A *replay attack* which records the reading of sensors and repeats these historical values was studied in [8]–[11]. The most extensively examined type of attack is a FDI attack [14], [15], [17]. FDI attacks change the system state by injecting falsified data into the communication devices. The aforementioned example in Ukraine falls into this category. Interested readers are referred to review papers [12], [13] for general cyber attacks in power grids. We narrow our attention to FDI attacks in this paper.

FDI attacks (in power grids) can be classified by the target functionality. Rahman *et al.* [14] studied FDI attacks against the state estimation which can be critical as incorrect situational awareness causes a malfunction in reliability applications such as contingency analysis. Isozaki *et al.* [15]

inspected FDI attacks on voltage regulation control causing irregular tap changing by manipulating load measurements and Choeum *et al.* [18] demonstrated a similar FDI attacks on Volt/VAR control which deteriorate the power quality. Also, Khanna *et al.* [17] investigated FDI attacks on optimal power flow by falsifying load measurements so that the resulting dispatch is not N-1 security compliant. However, to the best of our knowledge, there is no existing work that analyzes the FDI attacks associated with the DER dispatch signals.

As diverse as the type of FDI attacks are, proposed detection algorithms are equally diverse [19], [20]. Detection methods can be broadly divided into model-based and model-free, where model-based methods use power system state estimation, exploiting network information and physical system knowledge (e.g., grid topology, line impedances, etc.) to detect anomalies in the observation [21], [22]. In contrast, model-free approaches exploit recent advances in machine learning techniques such as classification and clustering [23]. There are, however, limited detection approaches capable of capturing temporal characteristics. A short-term state forecasting model in [24] takes temporal correlation among different nodal states into account to detect FDI attacks and Karimipour *et al.* [25] presented a dynamic state estimation method accounting for multiple time frames using Kalman filtering. It is noteworthy that none of the existing work has focused on the counter measurements and system vulnerability assessment for intraday FDI attacks.

SVR has gained popularity for time series forecasting in different domains. Thissen *et al.* [26] suggested that the SVR can model nonlinear relations and generate time series predictions. Salcedo-Sanz *et al.* [27] showed how SVR can be used for wind speed prediction, and He *et al.* [28] applied SVR to electricity load prediction. In addition, Feng *et al.* [29] utilized a multiple kernel SVR to capture both local and global information and applied to the traffic flow prediction. Building on these applications, we aim to exploit SVR for considering multi-temporal correlation in the system status and dispatch signals and identifying time-series FDI attacks.

B. Contribution

The contributions of this paper are two-fold. First, this paper reveals the vulnerability of the low-carbon power systems against the intraday FDI attacks by developing an attacker model that is composed of two optimization models: dispatch prediction model and dispatch falsification model. Based on the historical values and the collected knowledge of the target power grid, the dispatch prediction model mimics the functionality of the EMS and predicts the dispatch signals for DERs. Once the dispatch signal prediction is made, the dispatch falsification model solves an optimization problem to determine how to falsify the dispatch signals between DERs and EMS so that the accumulated deviations in DER outputs exceed the system security margin and cause the system power shortage. Second, this paper proposes a kernel SVR based detection model to enhance the reliability of the power grid against cyber attacks. The kernel SVR takes the input of monitored data comprised of multi-interval dispatch

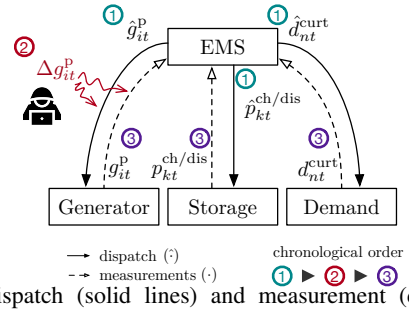


Fig. 2. Dispatch (solid lines) and measurement (dashed lines) signals between EMS and energy resources. The intervention of the attacker is marked in red color.

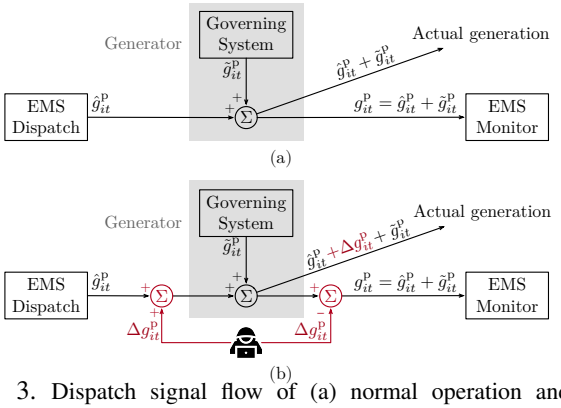


Fig. 3. Dispatch signal flow of (a) normal operation and (b) operation under intraday FDI attack on generation dispatch.

signals and the corresponding network status (nodal voltage magnitudes and phase angles) and predicts the system margin of the time of interest (e.g., two hours ahead). If the predicted margin drops below the threshold, the detection model notifies the system operator. To increase the performance of the kernel SVR, the training data is normalized by feature and transformed into a feature space using a kernel trick.

II. Intraday FDI Attack

This section describes a gradually pervasive FDI attack carried out during the evening net demand ramping-up periods. The attacker manipulates the dispatch signals from the EMS so that the power output of the various DERs (e.g., distributed generator, energy storage, demand response) deviates from the desired output. Once the accumulated deviation in the power output is larger than the system flexibility, then the supply cannot follow the demand ramp-up which can cause system-wide power outages and potentially cascading failures. This attack can be critical as the net demand continues to increase rapidly over the hours which makes the recovery process from the outage more challenging.

A. Overview of intraday FDI attack

The model in the following subsections is generic to intraday FDI attacks against any type of DERs that receive dispatch signals from EMS, while Figs. 2–3 use the attack on generation dispatch for illustrative purposes. Figure 2 illustrates the intraday FDI attack against the interactions between the EMS and the energy resources. First the attacker collects dispatch data and produces the falsification signal ($\Delta g^{\mathrm{p}}_{it}$) for the target time window. The actual attack is carried out by falsifying the dispatch going into the energy

resource ($\hat{g}^{\mathrm{p}}_{it}$) and the monitoring signal (g^{p}_{it}) into the EMS simultaneously. By doing so, the monitoring signal will appear the same from the EMS' perspective as $\hat{g}^{\mathrm{p}}_{it} + \tilde{g}^{\mathrm{p}}_{it}$ under the normal operation in Fig. 3(a) and the one under the intraday FDI attack in Fig. 3(b), while the actual generation is perturbed by $\Delta g^{\mathrm{p}}_{it}$ from the original dispatch regardless of presence of the governor adjustment. Other resources that receive dispatch signals from EMS such as energy storage, demand response can be exposed to the same type of attack.

B. Dispatch prediction model

To generate multi-interval falsifying dispatch signals, the attacker needs to predict the original dispatch signals from the EMS. This requires knowledge of the power grids such as network topology and nodal electricity demands. The attack model in this paper is assumed to have perfect knowledge of necessary information (network topology, line impedances, and thermal limits – see Table I), which can be considered as the most precise attack scenario.[1] The type of information needed for the attack model is summarized in Table I.[2] The dispatch prediction model mimics the power grid operation and is formulated as the second-order cone problem:

$$\min_{\Xi^{\mathrm{P}}} \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}} c^{\mathrm{g}}_i(\hat{g}^{\mathrm{p}}_{it}) + c^{\mathrm{d}}_n(\hat{d}^{\mathrm{p,curt}}_{nt}) + c^{\mathrm{r}}_i(\hat{r}_{it}) \tag{1a}$$

$$f^{\mathrm{p}}_{l|o(l)=n,t} - \sum_{l|r(l)=n}(f^{\mathrm{p}}_{lt} - a_{lt}R_l) - \sum_{i\in\mathcal{I}_n}\hat{g}^{\mathrm{p}}_{it} + D^{\mathrm{p}}_{nt} - \hat{d}^{\mathrm{p,curt}}_{nt} + G_n u_{nt} = 0, \quad \forall n\in\mathcal{N},\ t\in\mathcal{T}, \tag{1b}$$

$$f^{\mathrm{q}}_{l|o(l)=n,t} - \sum_{l|r(l)=n}(f^{\mathrm{q}}_{lt} - a_{lt}X_l) - \sum_{i\in\mathcal{I}_n}\hat{g}^{\mathrm{q}}_{it} + D^{\mathrm{q}}_{nt} - \Gamma_n\hat{d}^{\mathrm{p,curt}}_{nt} - B_n u_{nt} = 0, \quad \forall n\in\mathcal{N},\ t\in\mathcal{T}, \tag{1c}$$

$$u_{o(l),t} - 2(R_l f^{\mathrm{p}}_{lt} + X_l f^{\mathrm{q}}_{lt}) + a_{lt}(R_l^2 + X_l^2) = u_{r(l),t}, \quad \forall l\in\mathcal{L},\ t\in\mathcal{T}, \tag{1d}$$

$$\left[a_{lt} + u_{o(l),t}, a_{lt} - u_{o(l),t}, 2f^{\mathrm{p}}_{lt}, 2f^{\mathrm{q}}_{lt}\right] \in \mathcal{K}^4,\ \forall l\in\mathcal{L}, t\in\mathcal{T}, \tag{1e}$$

$$\begin{bmatrix} F_l \\ \hat{f}^{\mathrm{p}}_{lt} \\ \hat{f}^{\mathrm{q}}_{lt} \end{bmatrix}, \quad \begin{bmatrix} F_l \\ \hat{f}^{\mathrm{p}}_{lt} - \hat{a}_{lt}R_l \\ \hat{f}^{\mathrm{q}}_{lt} - \hat{a}_{lt}X_l \end{bmatrix} \in \mathcal{K}^3, \quad \forall l\in\mathcal{L},\ t\in\mathcal{T} \tag{1f}$$

$$\sum_{i\in\mathcal{I}} \hat{r}_{it} \geq K^{\mathrm{r,sys}} \sum_{n\in\mathcal{N}} D^{\mathrm{p}}_{nt}, \quad \forall t\in\mathcal{T}, \tag{1g}$$

$$R^{\mathrm{dn}}_i \leq \hat{g}^{\mathrm{p}}_{it} - \hat{g}^{\mathrm{p}}_{i,t-1} \leq R^{\mathrm{up}}_i, \quad \forall i\in\mathcal{I}^{\mathrm{C}}, t\in\mathcal{T}, \tag{1h}$$

$$\underline{G}^{\mathrm{p}}_i \leq \hat{g}^{\mathrm{p}}_{it} + \hat{r}_{it} \leq \overline{G}^{\mathrm{p}}_i, \quad \forall i\in\mathcal{I}, t\in\mathcal{T}, \tag{1i}$$

$$\underline{G}^{\mathrm{q}}_i \leq \hat{g}^{\mathrm{q}}_{it} \leq \overline{G}^{\mathrm{q}}_i, \quad \forall i\in\mathcal{I}, t\in\mathcal{T}, \tag{1j}$$

$$\underline{U}_n \leq \hat{u}_{nt} \leq \overline{U}_n, \quad \forall n\in\mathcal{N}, t\in\mathcal{T}, \tag{1k}$$

$$\underline{D}^{\mathrm{p,curt}}_{nt} \leq \hat{d}^{\mathrm{p,curt}}_{nt} \leq \overline{D}^{\mathrm{p,curt}}_{nt}, \quad \forall n\in\mathcal{N}, \tag{1l}$$

$$\hat{e}_{kt} = \hat{e}_{k,t-1} + \hat{p}^{\mathrm{ch}}_{kt}\aleph^{\mathrm{ch}} - \hat{p}^{\mathrm{dis}}_{kt}/\aleph^{\mathrm{dis}}, \quad \forall k\in\mathcal{K}, t\in\mathcal{T}, \tag{1m}$$

$$\underline{E}_{kt} \leq \hat{e}_{kt} \leq \overline{E}_{kt}, \quad \forall k\in\mathcal{K}, t\in\mathcal{T}, \tag{1n}$$

$$\underline{P}_k \leq \hat{p}^{\mathrm{ch}}_{kt}\cdot\aleph^{\mathrm{ch}},\ \hat{p}^{\mathrm{dis}}_{kt}/\aleph^{\mathrm{dis}} \leq \overline{P}_k, \quad \forall k\in\mathcal{K}, t\in\mathcal{T}, \tag{1o}$$

where $\Xi^{\mathrm{P}} := \{\hat{g}^{\mathrm{p}}_{it}, \hat{g}^{\mathrm{q}}_{it}, \hat{r}_{it}, \hat{d}^{\mathrm{p,curt}}_{nt}, \hat{e}_{kt}, \hat{p}^{\mathrm{ch}}_{kt}, \hat{p}^{\mathrm{dis}}_{kt}, \hat{a}_{lt}, \hat{u}_{nt} \geq 0, \hat{d}^{\mathrm{p,net}}_{nt}, \hat{f}^{\mathrm{p}}_{lt}, \hat{f}^{\mathrm{q}}_{lt} \in \mathbb{R}\}$. Power grid is defined with lines $l\in\mathcal{L}$ and nodes $n\in\mathcal{N}$ while time set is denoted as $t\in\mathcal{T}$. The objective function in (1a) minimizes the total cost of generation, load curtailment and reserve. Equations (1b)–(1e) are second-order-conic relaxation of the AC power flow equations [30], where $o(l)$ and $r(l)$ denote the sending and receiving buses of line l. Forward and backward line flow limits are enforced in (1f) with line capacity F_l. Given reserve requirement parameter $K^{\mathrm{r,sys}}$, the minimum system reserve is set proportional to the total demand in (1g). Ramping constraints of flexible generation units ($i\in\mathcal{I}^{\mathrm{C}}$) are imposed in (1h). The active and reactive power limits of generators are enforced in (1i) and (1j). Equations (1k) and (1l) limit lower and upper bounds of nodal voltage magnitudes and load curtailments respectively. Energy storage operation is modeled in (1m)–(1o) where the charging and discharging decisions are denoted as $\hat{p}^{\mathrm{ch}}_{kt}$ and $\hat{p}^{\mathrm{dis}}_{kt}$. The inter-temporal relationship of the state of charge $\hat{e}_{kt}$ is defined in (1m) with charging and discharging efficiency parameters, $\aleph^{\mathrm{ch}}$ and $\aleph^{\mathrm{dis}}$. The lower and upper bounds for the charging and discharging power are in (1o).

C. Dispatch falsification model

Once the prediction of the dispatch is made, the attacker can generate falsification signals. The falsification targets are constrained by the number and type of access points. For simplicity, it is assumed that the FDI attack would be carried out using only a single type of DERs (e.g., generator, storage or demand in Fig. 2). Then the dispatch falsification model is formulated with generic dispatch notation x_{kt} for unit k and attack time $t\in\mathcal{T}^a$ as follows:

$$\min_{\Delta x_{kt}} \sum_{t\in\mathcal{T}}\sum_{k\in\mathcal{K}} \left(|\Delta x_{kt}|^2 + \rho|\Delta x_{kt} - \Delta x_{k,t-1}|^2\right) \tag{2a}$$

$$|\sum_{k\in\mathcal{K}} \Delta x_{kt}| \geq K^{\mathrm{a}}_t \sum_{i\in\mathcal{I}} \hat{r}_{it}, \quad \forall t\in\mathcal{T}^{\mathrm{a}} \tag{2b}$$

$$|\sum_{t\in\mathcal{T}^{\mathrm{a}}}\sum_{k\in\mathcal{K}} \Delta x_{kt}| \geq \sum_{t\in\mathcal{T}^{\mathrm{a}}}\sum_{i\in\mathcal{I}} \hat{r}_{it}, \tag{2c}$$

$$-\epsilon^{\mathrm{a}}\hat{x}_{kt} \leq \Delta x_{kt} \leq \epsilon^{\mathrm{a}}\hat{x}_{kt}, \quad \forall k\in\mathcal{K}, t\in\mathcal{T}^{\mathrm{a}}, \tag{2d}$$

$$-\underline{X}_{kt} \leq \hat{x}_{kt} + \Delta x_{kt} \leq \overline{X}_{kt}, \quad \forall k\in\mathcal{K}, t\in\mathcal{T}^{\mathrm{a}}, \tag{2e}$$

Given the prediction of the target dispatch (e.g., $\hat{x}_{kt} := \hat{g}^{\mathrm{p}}_{it}$ for generator, $\hat{x}_{kt} := \hat{p}^{\mathrm{ch/dis}}_{kt}$ for storage, and $\hat{x}_{kt} := \hat{d}^{\mathrm{p,curt}}_{nt}$ for load curtailment), the attacker determines the falsification signal (e.g., $\Delta x_{kt} = \Delta g^{\mathrm{p}}_{it}$). The objective function in (2a) minimizes the sum of the squared size of the attack and the temporal smoothness regularization term[3] with penalty parameter ρ. Constraint (2b) sets the impact of the attack achieves the target deviation with user-defined parameter K^{a}_t for each time interval and (2c) ensures the accumulated deviation in the dispatch during the attacking windows $t\in\mathcal{T}^{\mathrm{a}}$ exceeds the system reserve. The lower and upper bounds for individual falsification signals are given in (2d) and (2e)

[1] In the 2015 Ukraine case [2] the network information was obtained by reconnaissance operations.

[2] While the reserve allocation for each balancing unit is marked as unknown, the attack model requires the knowledge of the total reserve.

[3] The regularization term in (2a) can be extended to account for other dimensions such as geographical locations (i.e., similar deviations in nearby dispatch signals).

TABLE I. ASSUMPTIONS FOR THE ATTACK AND DETECTION MODELS

Name	Attack Model	Detection Model
Network information:		
topology $(\mathcal{N}, \mathcal{L})$	✓	✓
line impedance (R_l, X_l)	✓	✓
line thermal limit (F_l)	✓	✓
Dispatch signals:		
generation output (g_{it}^{p})	–	P & A
load curtailment (d_{nt}^{curt})	–	P & A
storage dispatch $(p_{kt}^{\mathrm{ch/dis}})$	–	P & A
reserve (r_{it})	–	P & A
Measurements:		
nodal demand (D_{nt}^{p})	P	P & A
nodal voltage (v_{nt}, θ_{nt})	–	✓

* P: prediction, A: actual, ✓: assumed to be known, –: unknown

where ϵ^{a} is an user-defined parameter to confine the attack size and the original dispatch bounds are $\underline{X}_{kt}$ and $\overline{X}_{kt}$.

III. FDI ATTACK DETECTION WITH KERNEL SVR

To detect the intraday FDI attack carried out over several hours, the detection model should be able to capture the temporal changes in dispatch signals as well as the current values. To do so, the detection model requires access to the dispatch signals and network status for the monitoring windows $\mathcal{T}^{\mathrm{m}}$ (in the sequel we shall use a 6-hour window) and the system margin at the time of interest T^{pred} (in our examples, the time of interest is set to 2 hours after the monitoring window closes) as summarized in Table I.

A. Kernel Support Vector Regression

Kernel SVR is a generalization of Kernel support vector machine for real-value function estimation, commonly equipped with ϵ-insensitive loss function (sometimes referred to as soft-margin loss function) [31]. A kernel function maps the original input data into feature space (i.e., $x^i \mapsto \phi(x^i)$) through the use of inner products, which is compatible for the SVR model fitting. The common selection of a kernel is Gaussian radial basis kernel function (RBF) in (3a) and polynomial kernel of degree d in (3b) (See [32] for details):

$$k(x,y) \coloneqq e^{-\frac{\|x-y\|^2}{2\sigma^2}} \tag{3a}$$

$$k(x,y) \coloneqq (x^\top y)^d \tag{3b}$$

Given input data x^i with user-defined parameter C and feature map $\phi(\cdot)$, the Kernel SVR for predicting output: $y_i \in \mathbb{R}$ can be modeled as in (4):

$$\min_{w,b,\xi_i^+\geq 0,\xi_i^-\geq 0} \quad \|w\| + C\sum_{i=1}^{l}(\xi_i^+ + \xi_i^-) \tag{4a}$$

$$\text{s.t.} \quad (w^\top \phi(x^i) + b) - y_i \leq \epsilon + \xi_i^+, \quad \forall i \tag{4b}$$

$$y_i - (w^\top \phi(x^i) + b) \leq \epsilon + \xi_i^-, \quad \forall i \tag{4c}$$

The objective function in (4a) minimizes the sum of the norm of ω and the loss terms, where $\omega^\top x + b = 0$ is a decision boundary and the size of margin is $\frac{2}{||\omega||}$ (i.e., the margin is maximized). Points within an ϵ distance of the support vector do not contribute to the cost.

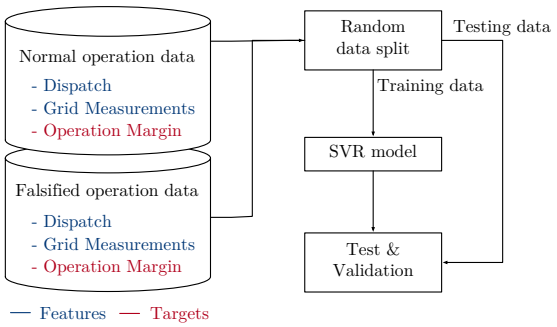


Fig. 4. Training and testing process of the kernel SVR detection model with the normal operation data and falsified dispatch data.

B. Kernel SVR detection model

Figure 4 shows the process of training the kernel SVR for detecting the intraday FDI attacks. To train the SVR model in (4), we first generate *normal operation data* with supply and demand side uncertainties and *falsified operation data*. The supply and demand uncertainties are modeled as a random variable drawn from a Gaussian distribution. In practice, historical data can be used for generating a probability distribution of the supply and demand uncertainties. The power grid measurement data (voltage magnitudes and phase angles) are obtained by solving the power flow equations with fixed power injections based on the supply and demand data. Synthetic falsified operation data can be generated in the same way, while the dispatch signals are fixed as the falsified values from the dispatch falsification model.

Once the training data is prepared, the input vector x^i for each observation i is constructed which is comprised of prediction $(\hat{\cdot})$ and actual $(\cdot)$ values of energy resources over the user-defined monitoring time windows $(t \in \mathcal{T}^{\mathrm{m}})$ and the system status (voltage magnitudes and phase angles):

$$x^i = \begin{bmatrix} [\hat{g}_{it}^{\mathrm{p}}]_{\forall i\in\mathcal{I}, t\in\mathcal{T}^{\mathrm{m}}} \\ [g_{it}^{\mathrm{p}}]_{\forall i\in\mathcal{I}, t\in\mathcal{T}^{\mathrm{m}}} \\ [\hat{d}_{nt}^{\mathrm{curt}}]_{\forall n\in\mathcal{N}, t\in\mathcal{T}^{\mathrm{m}}} \\ [d_{nt}^{\mathrm{curt}}]_{\forall n\in\mathcal{N}, t\in\mathcal{T}^{\mathrm{m}}} \\ [\hat{p}_{kt}^{\mathrm{ch/dis}}]_{\forall k\in\mathcal{K}, t\in\mathcal{T}^{\mathrm{m}}} \\ [p_{kt}^{\mathrm{ch/dis}}]_{\forall k\in\mathcal{K}, t\in\mathcal{T}^{\mathrm{m}}} \\ [v_{nt}]_{\forall n\in\mathcal{N}^{\mathrm{s}}, t\in\mathcal{T}^{\mathrm{m}}} \\ [\theta_{nt}]_{\forall n\in\mathcal{N}^{\mathrm{s}}, t\in\mathcal{T}^{\mathrm{m}}} \end{bmatrix} \in \mathbb{R}^d, \; y_i \in \mathbb{R} \tag{5a}$$

where $d = (2|\mathcal{I}| + 2|\mathcal{N}| + 2|\mathcal{K}| + 2|\mathcal{N}^{\mathrm{s}}|)|\mathcal{T}^{\mathrm{m}}|$. Then the corresponding output (y_i) of the kernel SVR is set as the system operation margin defined as the minimum of the remaining up-ward and down-ward reserves at the time of interest $t = T^{\mathrm{pred}}$ (e.g., 2 hours from now). Formally, the margin at time t is:

$$\min\{r_t^{\mathrm{up}} - \sum_i (x_{it} - \hat{x}_{it}), \; r_t^{\mathrm{dn}} - \sum_i (\hat{x}_{it} - x_{it})\}. \tag{5b}$$

In (5b), the generic notation x_{it} is used to represent all flexible resources such as generators (g_{it}^{p}), storage $(p_{kt}^{\mathrm{ch/dis}})$, load curtailment (d_{nt}^{curt}) and the system margin is defined as the total sum of remaining flexibility in all units.

To increase and validate the model performance, additional steps are added in the kernel SVR. First, the input data is

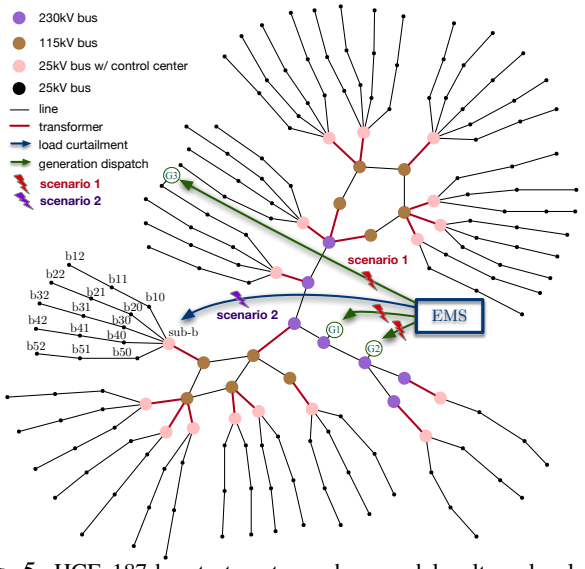


Fig. 5. HCE 187-bus test system where nodal voltage levels are marked color-coded. Two intraday FDI attack scenarios – generation dispatch (scenario 1) and load curtailment (scenario 2).

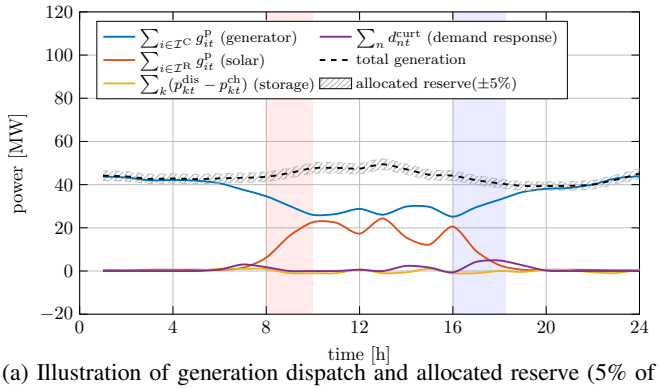
normalized by each feature so that they can be accounted for equally. Once the fitting of the kernel SVR is completed, then the performance is validated with the testing data.

IV. Numerical Experiments

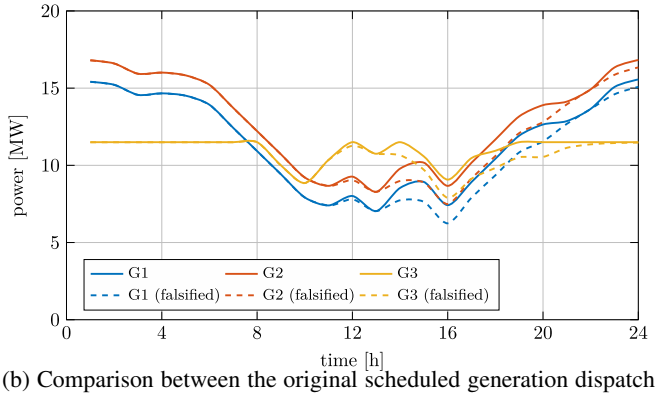
We use the 187-bus test system shown in Fig. 5 with daily generation and load profiles based on real-life data provided by Holy Cross Energy, a power utility in Colorado. The PV generation has been scaled up by a factor of six so that the overall solar penetration is around 15% of total generation, which is similar to the current practice in the state of California (15.43% of the total generation was from solar in 2020) [33]. Demand forecast error is assumed to follow a Gaussian distribution with zero mean and 2% of nominal value as standard deviation. For the attack model in (2), the generation dispatch (Scenario 1) and load curtailment (Scenario 2) are assumed to be falsified respectively. For the detection model in (4), the monitoring window ($\mathcal{T}^{\mathrm{m}}$) and the time of interest (T^{pred}) are set as six hours and two hours from the time of prediction, i.e., $\mathcal{T}^{\mathrm{m}} = [T^{\mathrm{pred}} - 8\mathrm{h},\ T^{\mathrm{pred}} - 2\mathrm{h}]$. The RBF kernel is used for feature map $\phi(\cdot)$. All optimization problems are modeled using Julia/JuMP [34], and solved by Ipopt Solver [35]. The kernel SVR is implemented using the scikit-learn package [36].

A. Scenario 1: intraday FDI attack on generation dispatch

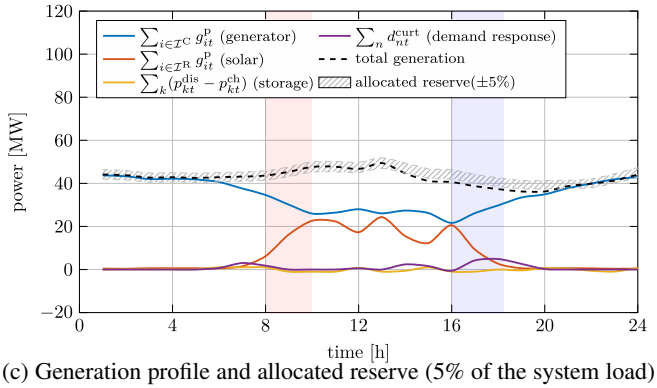
1) Dispatch prediction: Given solar generation and demand forecasts from a cloudy day, the attacker solves the prediction model in (1) and the resulting generation profile is shown in Fig. 6(a). The solar generation (red line) has two valleys around 12:00 and 15:00 (due to the weather conditions such as intermittent clouds), which are offset mainly by the generators (blue line). In order to compensate for the change in solar generation output around sunrise and sunset, controllable resources such as gas generation, energy storage, and load curtailment are utilized. Additionally, the system reserve is allocated to address the system uncertainty.



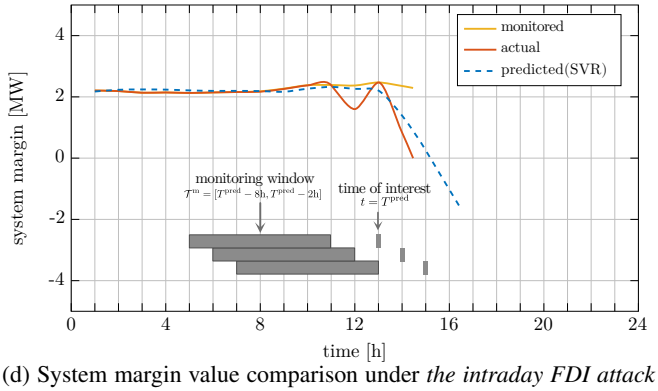
(a) Illustration of generation dispatch and allocated reserve (5% of the system load) on a *cloudy day*. The red and blue boxes represent the net-load morning ramp down and evening ramp up periods.



(b) Comparison between the original scheduled generation dispatch (solid lines) and falsified FDI attack (dashed lines).



(c) Generation profile and allocated reserve (5% of the system load) under *the intraday FDI attack on generation dispatch*.



(d) System margin value comparison under *the intraday FDI attack on generation dispatch.*: values on EMS monitor, actual margin and predicted values from the detection model.

Fig. 6. Illustration of Scenario 1: (a) dispatch prediction, (b) generation dispatch falsification, (c) generation profile under the intraday FDI attack, (d) system margin prediction.

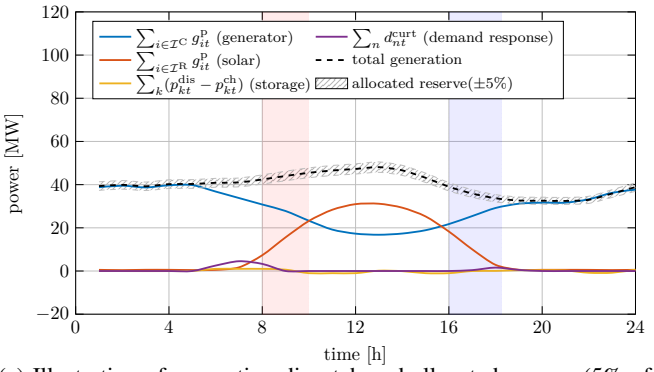
2) Falsification of generation dispatch: The attacker generates falsification signals by solving (2) with the predicted dispatch. Fig. 6(b)–(c) show a scenario of the intraday FDI attacks on generation dispatch g_{it}^{p}. In Fig. 6(b), the original dispatch is planned to provide approximately 10MW/3hr of ramping-up flexibility around sunset (16:00–19:00). The attack signal overrides the original values (solid line) and instead injects reduced generation dispatch (dashed line). As a result, compared to Fig. 6(a), the total system generation (dashed black line) in Fig. 6(c) decreases due to this FDI attack, and the total amount of supply-demand imbalance exceeds the system flexibility (shaded area) which can cause a system-wide power outage.

3) Detecting the FDI attack with kernel SVR: Fig. 6(d) shows how the system margin changes over time when the FDI attack presented in Fig. 6(b) is carried out. While the actual system margin (orange line) drops as a result of the attack and is eventually exhausted at 14:30, the EMS monitor (yellow line) shows a steady margin throughout the entire attack. This is because the attacker injects the falsified signal into the up-link to the EMS monitor as well as the down-link to the DERs. The proposed detection model provides the estimation on the system margin two hours ahead of the time of interest. For example, in Fig. 6(d), the system margin prediction (blue dashed line) plotted for 13:00 is made at 11:00 based on the monitoring window 5:00–11:00 and the prediction gets updated and shifted as time flows (the three gray blocks in the bottom of Fig. 6(d) illustrate how the monitoring window and time of interests change over time). Thus, the grid operator can monitor this prediction and take preventive measures in a timely manner once the margin drops below the reliability threshold, i.e., at 13:00, the operator can know that margin is expected to drop below 1MW in an hour. Note that the monitoring and actual margin values are not recorded after the power outage at 14:30 and therefore the prediction is also available only until 16:30.

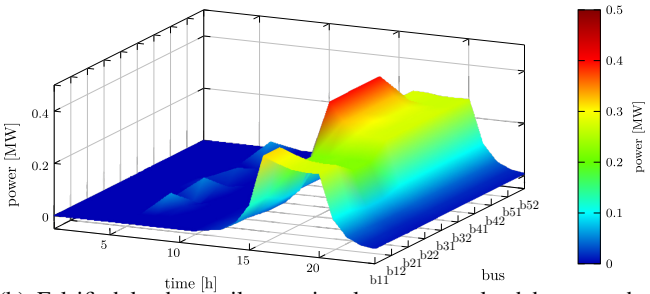
B. Scenario 2: intraday FDI attack on load curtailments

1) Dispatch prediction: Fig. 7(a) shows the generation and demand profile of a sunny day predicted by an attacker. Unlike Scenario 1, the solar generation shows a smooth ramp-up during the sunrise and ramp-down during the sunset, and the generation output covering the variability from solar is also expected to have a smooth shape without a valley.

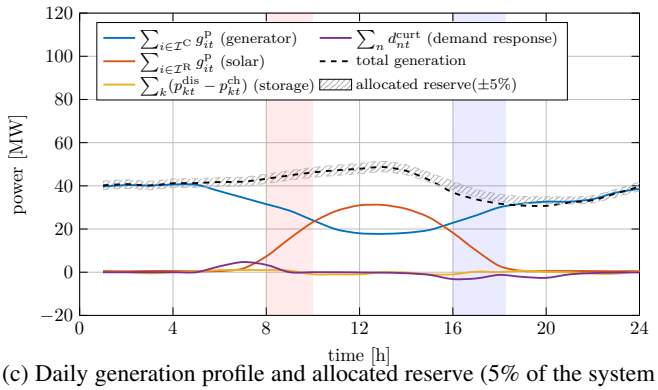
2) Falsification of load curtailment: In this scenario, we assume the control center at *sub-b* in Fig. 5 is hacked and the downstream load curtailment signals (d_{nt}^{curt}) are falsified. In contrast to the falsification of generation dispatch signals in Fig. 6(b), the FDI attack on load curtailment requires manipulation on a greater number of signals that are geographically related (nearby electricity loads have similar patterns). To make the falsification natural, the attacker adds a geographical regularization in addition to the temporal regularization term in the objective function and solves the attack model in (2) to design an attack. Thus, the falsification signals ($\Delta d_{nt}^{\mathrm{curt}}$) on nearby nodes (e.g., *b11* and *b12*, *b51* and *b52*) in Fig. 7(b) have a similar shape over time, which makes



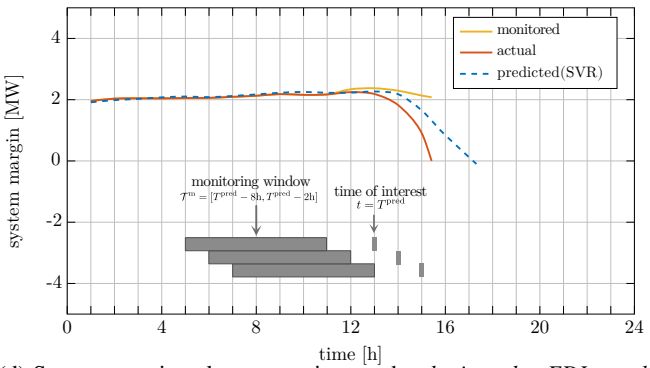
(a) Illustration of generation dispatch and allocated reserve (5% of the system load) on a *sunny day*. The red and blue boxes represent the net-load morning ramp down and evening ramp up periods.



(b) Falsified load curtailment signals on target load buses under *sub-b* substation.



(c) Daily generation profile and allocated reserve (5% of the system load) under *the intraday FDI attack on load curtailments.*



(d) System margin value comparison under *the intraday FDI attack on load curtailments.*: values on EMS monitor, actual margin and predicted values from the detection model.

Fig. 7. Illustration of Scenario 2: (a) dispatch prediction, (b) load curtailment falsification, (c) generation profile under the intraday FDI attack, (d) system margin prediction.

the attack hard to be detected. As a result of the attack, as shown in Fig. 7(c), the total system generation (dashed-line) decreases gradually and the total amount of supply-demand imbalance exceeds the system flexibility (shaded area).

3) Detecting the FDI attack with the kernel SVR: Similar to Scenario 1, the monitored system margin (yellow line) in Fig. 7(d) looks normal throughout all intervals. However, the actual margin (red line) starts to decrease at 13:00 and is fully exhausted at 15:20. The proposed kernel SVR detection model predicts the drop would begin at 14:00 and the margin will be below 1MW at 15:30. In other words, the grid operator will notice the change at 12:00 (two hours ahead of 14:00) and the preventive measure will be taken at 13:30 (two hours ahead of 15:30) if the security threshold is 1MW.

V. Conclusions and Future work

This paper analyzed the vulnerability of power grids with high PV penetration against an intraday FDI attack that falsifies DER dispatch and monitoring signals. Based upon the dispatch prediction and dispatch falsification models, we illustrated how gradual manipulation of DER outputs can cause a power imbalance which exceeds the system reliability margin. To enhance the power grid reliability against the attack scenario, we also proposed a detection model utilizing a kernel SVR which allows a power grid operator to predict the reduction in the system margin ahead of time. The numerical experiments demonstrate the attack scenarios and the performance of the detection model on the HCE test system, which is based on real-world data.

There are several directions for future works. First, we plan to relax the perfect knowledge assumption on the grid conditions. In practice, such information is available only in limited locations and we will investigate how this affects the detection model. We also plan to evaluate the performance of the kernel SVR detection model against other detection methods and carry out the comparative analyses.

Acknowledgement

The authors would like to thank Bruno Leao and Ulrich Muenz at Siemens Technology for helpful discussions regarding the attack scenarios. We thank Chris Bilby at Holy Cross Energy for sharing relevant datasets.

References

[1] P. W. Pong *et al.*, "Cyber-enabled grids: Shaping future energy systems," *Advances in Applied Energy*, vol. 1, p. 100003, 2021.

[2] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "The 2015 ukraine blackout: Implications for false data injection attacks," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3317–3318, 2016.

[3] Utility Dive, "Murray Energy, largest US producer, joins long list of bankrupt coal companies," 2019. [Online]. Available: https://bit.ly/3NfspyW

[4] California ISO, "Today's supply outlook dashboard," 2022. [Online]. Available: http://www.caiso.com/TodaysOutlook/Pages/default.aspx

[5] C.-W. Ten *et al.*, "Vulnerability assessment of cybersecurity for scada systems," *IEEE Trans. Power Syst.*, vol. 23, no. 4, p. 1836, 2008.

[6] J. Liu *et al.*, "Resilience analysis of DC microgrids under denial of service threats," *IEEE Trans. Power Syst.*, 2019.

[7] W. Chen, D. Ding, H. Dong, and G. Wei, "Distributed resilient filtering for power systems subject to denial-of-service attacks," *IEEE Trans. Syst. Man Cybern.: Syst.*, vol. 49, no. 8, pp. 1688–1697, 2019.

[8] X. Jiang, J. Zhang, B. J. Harding, J. J. Makela, A. D. Domı *et al.*, "Spoofing gps receiver clock offset of phasor measurement units," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3253–3262, 2013.

[9] S. Pan, T. Morris, and U. Adhikari, "Developing a hybrid intrusion detection system using data mining for power systems," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 3104–3113, 2015.

[10] T. Huang, B. Satchidanandan, P. Kumar, and L. Xie, "An online detection framework for cyber attacks on automatic generation control," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6816–6827, 2018.

[11] S. Soltan *et al.*, "React to cyber attacks on power grids," *IEEE Trans. Netw. Sci. Eng.*, vol. 6, no. 3, pp. 459–473, 2018.

[12] S. Aoufi *et al.*, "Survey of false data injection in smart power grid: Attacks, countermeasures and challenges," *J. Inf. Secur. Appl*, 2020.

[13] H. Zhang, B. Liu, and H. Wu, "Smart grid cyber-physical attack and defense: A review," *IEEE Access*, vol. 9, pp. 29 641–29 659, 2021.

[14] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks against nonlinear state estimation in smart power grids," in *IEEE Power & Energy Society General Meeting*, 2013.

[15] Y. Isozaki *et al.*, "Detection of cyber attacks against voltage control in distribution power grids with pvs," *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 1824–1835, 2015.

[16] S. Soltan, M. Yannakakis, and G. Zussman, "Power grid state estimation following a joint cyber and physical attack," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 1, pp. 499–512, 2016.

[17] K. Khanna, B. K. Panigrahi, and A. Joshi, "Bi-level modelling of false data injection attacks on security constrained optimal power flow," *IET Gener. Transm. Distrib.*, vol. 11, no. 14, pp. 3586–3593, 2017.

[18] D. Choeum and D.-H. Choi, "OLTC-induced false data injection attack on volt/var optimization in distribution systems," *IEEE Access*, vol. 7, pp. 34 508–34 520, 2019.

[19] F. Pasqualetti *et al.*, "Attack detection and identification in cyber-physical systems," *IEEE Trans. Automat. Contr.*, 2013.

[20] A. S. Musleh, G. Chen, and Z. Y. Dong, "A survey on the detection algorithms for false data injection attacks in smart grids," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2218–2234, 2019.

[21] A. Ashok, M. Govindarasu, and V. Ajjarapu, "Online detection of stealthy false data injection attacks in power system state estimation," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1636–1646, 2016.

[22] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. S. Sastry, "Cyber security analysis of state estimators in electric power systems," in *Proc. IEEE Conf. Decis. Control*, 2010, pp. 5991–5998.

[23] L. Cui *et al.*, "Detecting false data attacks using machine learning techniques in smart grid: A survey," *J. Netw. Comput. Appl*, 2020.

[24] J. Zhao *et al.*, "Short-term state forecasting-aided method for detection of smart grid general false data injection attacks," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1580–1590, 2015.

[25] H. Karimipour and V. Dinavahi, "Robust massively parallel dynamic state estimation of power systems against cyber-attack," *IEEE Access*, vol. 6, pp. 2984–2995, 2017.

[26] U. Thissen *et al.*, "Using support vector machines for time series prediction," *Chemom. Intell. Lab. Syst.*, vol. 69, no. 1, p. 35, 2003.

[27] S. Salcedo-Sanz *et al.*, "Short term wind speed prediction based on evolutionary support vector regression algorithms," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 4052–4057, 2011.

[28] W. He *et al.*, "Model optimizing and feature selecting for support vector regression in time series forecasting," *Neurocomputing*, 2008.

[29] X. Feng *et al.*, "Adaptive multi-kernel svm with spatial–temporal correlation for short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2001–2013, 2018.

[30] M. Farivar and S. H. Low, "Branch flow model: Relaxations and convexification—Part I," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2554–2564, 2013.

[31] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, 2004.

[32] M. Rohmah *et al.*, "Comparison Four Kernels of SVR to Predict Consumer Price Index," in *J. Phys. Conf. Ser.*, 2021.

[33] California Energy Commission, "California electricity data - total system electric generation data," 2020. [Online]. Available: https://www.energy.ca.gov/data-reports/energy-almanac/california-electricity-data/

[34] I. Dunning *et al.*, "Jump: A modeling language for mathematical optimization," *SIAM Review*, vol. 59, no. 2, pp. 295–320, 2017.

[35] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical programming*, vol. 106, no. 1, pp. 25–57, 2006.

[36] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.