Adaptive Regulated Sparsity Promoting Approach for Data-Driven Modeling and Control of Grid-Connected Solar Photovoltaic Generation

Zhongtian Zhang, Javad Khazaei, *Senior Member, IEEE*, and Rick S. Blum *Fellow, IEEE.*

***Abstract*—This paper introduces a new statistical learning technique based on sparsity promotion for data-driven modeling and control of solar photovoltaic (PV) systems. Compared with conventional sparse regression techniques that might introduce computational complexities when the number of candidate functions increases, an innovative algorithm, named adaptive regulated sparse regression (ARSR) is proposed. The ARSR adaptively regulates the hyperparameter weights of candidate functions to best represent the dynamics of PV systems. This method allows for the application of different sparsity-promoting hyperparameters for each state variable, whereas the conventional approach uses the same hyperparameter for all state variables, which may result in not excluding all the unrelated terms from the dynamics. Consequently, the proposed method can identify more complex dynamics with greater accuracy. Utilizing this algorithm, open-loop and closed-loop models of single-stage and two-stage PV systems are obtained from measurements and are utilized for control design purposes. Moreover, it is demonstrated that the proposed data-driven approach can be successfully employed for fault analysis studies, which distinguishes its capabilities from other data-driven techniques. Finally, the proposed approach is validated through real-time simulations.**

***Keywords*—** Photovoltaic (PV) Systems, Single-stage PV, Two-stage PV, Closed-Loop Data-driven Modeling, Adaptive Regulated Sparse Regression.

I. Introduction

AVAILABILITY of high-resolution measurements from field devices and existing challenges for accurately modeling distributed energy resources (DERs) have motivated data-driven modeling and control of smart grid assets. A wide range of system identification approaches have been introduced to extract dynamics from data. Some of these techniques include: dynamic mode decomposition (DMD) [1], [2], Koopman operator [3], [4], and sparse identification of nonlinear dynamics or sparse regression [5], [6]. These approaches have also been applied to power systems [7]–[10]. For instance, dynamic mode decomposition has been employed in [7] for delay-tolerant microgrid control, while the Koopman operator has been utilized in [8] to identify generator dynamics for state estimation purposes.

The DMD is based on the assumption of linear dynamics and cannot effectively handle nonlinear dynamics such as DERs. The Koopman operator is introduced as an infinite-dimensional linear operator that can sometimes transform the dynamics of a nonlinear system into a linear representation. However, the transformation is only guaranteed in certain scenarios and under certain assumptions, which do not generally apply to nonlinear dynamics [3], [4]. In contrast, sparse identification leverages sparse regression to identify the dominant dynamics of candidate functions and has demonstrated promising results in accurately modeling the unknown dynamics of nonlinear systems [11], [12]. One of the key advantages of sparse regression is its implementation simplicity, reduced training time, interpretability, and superior performance compared to other model identification techniques. Numerous studies, including our research, have demonstrated that sparse regression can be efficiently applied in the field of power and energy, including dynamics and impedance modeling of power converters in DC microgrids [13], data-driven nonlinear modeling and feedback linearization control of DERs [14], modularized sparse identification (M-SINDy) of microgrid transient dynamics [15], modeling of naval power systems [16], and power grid parameter estimation [17].

The existing literature underscores the substantial potential of sparse regression in identifying nonlinear dynamics within dynamical systems. However, photovoltaic (PV)-based DERs are instrumental in decarbonizing the power grid, yet a notable research gap exists in developing data-driven models and control strategies that cater to the diverse topologies and dynamics of solar photovoltaic systems, encompassing both single-stage and two-stage configurations. To the best of our knowledge, no existing study has explored data-driven modeling (both open-loop and closed-loop) and control of PV systems using sparse regression. Furthermore, a key hyperparameter (λ in (28)), which acts as a tuning knob for sparsifying the dynamics in conventional sparse regression methods, is fixed and can cause model identification errors in high dimensional systems [5], [6], [15]. The exploration of optimized selection of this key hyperparameter has not been addressed in existing research.

To address these limitations, this paper develops a novel data-driven modeling framework using adaptive regulated sparse regression (ARSR) for characterization of the dynamics of single-stage and two-stage photovoltaic (PV) systems from available data. The approach employs adaptive sparse regression to optimally identify the sparse dynamics from

This research was in part under support from the National Science Foundation under Grant NSF-EPCN 2221784. Zh. Zhang, J. Khazaei, and R. S. Blum are with the Electrical and Computer Engineering department at Lehigh University, PA, USA. (E-mails: $zhz819@lehigh.edu$, $jak921@lehigh.edu$, and $rblum@eecs.lehigh.edu$)

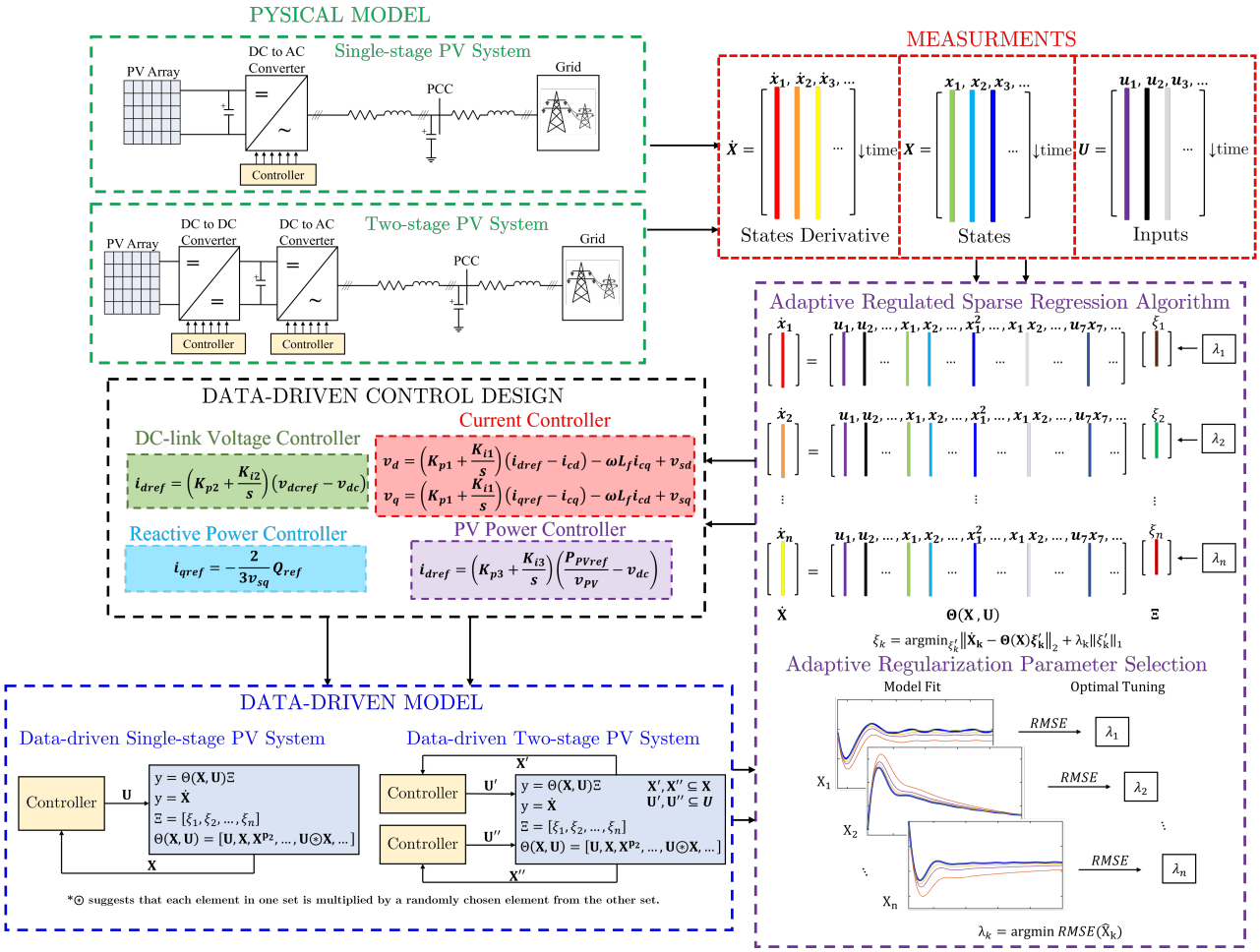

Fig. 1 Proposed adaptive regulated sparse regression for modeling identification and control of PV systems.

a library of candidate functions and construct accurate data-driven models using measurements. Compared to the conventional sparse regression method, which uses a single sparse-promoting hyperparameter for all state variables in the dynamics, the proposed adaptive regularization approach can apply different hyperparameters to different state variables of the system and optimally tune these hyperparameters during the identification process. This enhances the accuracy of data-driven modeling and enables handling of complex dynamics that the conventional method cannot manage, such as closed-loop modeling for PV systems.

This paper makes several key contributions to the field of PV system modeling and control, including: 1) introduction of a fast and scalable sparsity promoting method for dynamic model identification of nonlinear dynamics named as adaptive regulated sparse regression algorithm; 2) data-driven open-loop and closed-loop modeling of single-stage and two-stage PV systems; 3) validation of the proposed method in real-time simulation; 4) exploration of the proposed data-driven method for model-free fault analysis.

The rest of the paper is structured as follows: Section II formulates the problem. The ARSR method is introduced in Section III. Section IV covers the data-driven system modeling and control design. Section V comprises case studies, and Section VI concludes the paper.

II. Problem Formulation

This section presents the open-loop and closed-loop dynamic models of single-stage and two-stage PV systems to be utilized for the data-driven modeling approach in the next sections. Fig. 1 shows the proposed data-driven modeling approach to solve this challenge. An adaptive sparse regression algorithm is employed to identify the dynamics of PV systems using collected data. Subsequently, the outcomes of the identification are harnessed to construct data-driven models for the PV systems. Finally, data-driven controllers are designed and implemented using the obtained data-driven models.

A. Open-loop Dynamic Model and Control of PV systems

*1) **Single-Stage PV**:* The single-stage PV model consists of a PV array, a three-phase voltage source converter (VSC), and a low-pass filter connected to the main grid at a point of common coupling (PCC), as shown in Fig.2. The DC-side of the VSC is connected to the PV array with a capacitor C_{dc} and the AC-side is connected to an LC filter. The components of

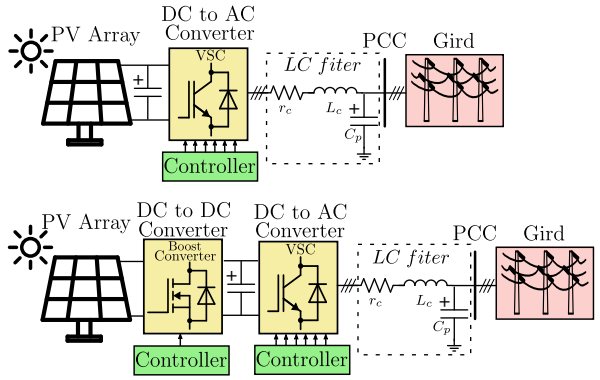

Fig. 2 Schematic of a single-stage and two-stage PV systems.

the filter are represented by r_c, L_c and C_f, and the impedance of the grid is represented by L_g and r_g, which refer to the inductance and resistance of the grid.

a) AC-side Dynamics: The AC-side dynamics of the system after converting to the dq-frame, can be described by

$$\begin{bmatrix} \dot{i}_{cd} \\ \dot{i}_{cq} \\ \dot{i}_{gd} \\ \dot{i}_{gq} \\ \dot{v}_{sd} \\ \dot{v}_{sq} \end{bmatrix} = \begin{bmatrix} -\frac{r_c}{L_c} & \omega_0 & 0 & 0 & -\frac{1}{L_c} & 0 \\ -\omega_0 & -\frac{r_c}{L_c} & 0 & 0 & 0 & -\frac{1}{L_c} \\ 0 & 0 & -\frac{r_g}{L_g} & \omega_0 & \frac{1}{L_g} & 0 \\ 0 & 0 & -\omega_0 & -\frac{r_g}{L_g} & 0 & \frac{1}{L_g} \\ \frac{1}{C_f} & 0 & -\frac{1}{C_f} & 0 & 0 & \omega_0 \\ 0 & \frac{1}{C_f} & 0 & -\frac{1}{C_f} & -\omega_0 & 0 \end{bmatrix} \begin{bmatrix} i_{cd} \\ i_{cq} \\ i_{gd} \\ i_{gq} \\ v_{sd} \\ v_{sq} \end{bmatrix} + g(u)$$

$$g(u) = \begin{bmatrix} \frac{v_{cd}}{L_c} & \frac{v_{cq}}{L_c} & \frac{-v_{gd}}{L_g} & \frac{-v_{gq}}{L_g} & 0 & 0 \end{bmatrix}^T \quad (1)$$

where the output voltage and current of the VSC are represented by v_{cd}, v_{cq}, i_{cd} and i_{cq} [14]. Furthermore, the dq-frame voltages and currents of the grid are represented by v_{gd}, v_{gq}, i_{gd} and i_{gq}. Also, the PCC voltage is represented by v_{sd} and v_{sq} in dq-frame, and ω_0 is the nominal frequency of the grid.

b) DC-side Dynamics: Based on the power balance on the DC and AC sides of the VSC (steady state), the output power of the PV array (P_{PV}) should be equal to the VSC output power (P_g) [18]. The voltage dynamics of the DC-link capacitor can be described as

$$\frac{1}{2}C_{dc}v_{dc} \times v_{dc}s = P_{PV} - P_g \quad (2)$$

$$\dot{v}_{dc} = \frac{i_{PV}}{C_{dc}} - \frac{3}{2}\frac{v_{gd}}{C_{dc}v_{dc}}i_{gd} \quad (3)$$

where v_{gd} and i_{gd} are the voltage and current of the grid along the d-axis; $P_g = \frac{3}{2}v_{gd}i_{gd}$; v_{dc} and C_{dc} are the voltage and capacitance of the capacitor on the DC side of VSC; and i_{PV} is the PV output current.

c) State Space Model: According to the AC-side and DC-side dynamics, the system can be represented in state space form as shown in equations (1) and (3).

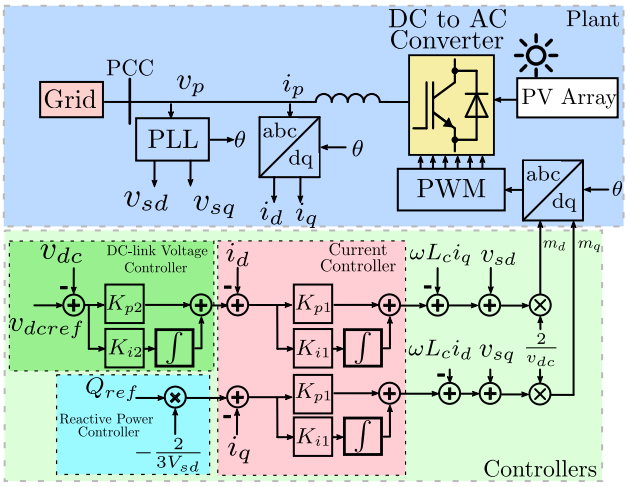

Fig. 3 Control block diagram of single-stage PV system.

d) Current Controllers: To regulate the converter current, a current controller is implemented (see Fig.3), which can be formulated as

$$v_{dref} = (K_{p1} + \frac{K_{i1}}{s})(i_{dref} - i_{cd}) - \omega L_f i_{cq} + v_{sd} \quad (4)$$

$$v_{qref} = (K_{p1} + \frac{K_{i1}}{s})(i_{qref} - i_{cq}) + \omega L_f i_{cd} + v_{sq} \quad (5)$$

where v_{dref} and v_{qref} are the reference voltages in dq-frame while i_{dref} and i_{qref} are the reference currents of the controller in dq-frame [19]. In addition, K_{p1} and K_{i1} are the proportional and integral parameters of the PI regulators in the current controller.

e) Active and Reactive Power Controllers: In grid-tie PV systems, a phase-locked loop is implemented to synchronize the converter frequency to the grid frequency by setting the q component of the converter voltage (v_{cq}) to 0 using a PI controller [20]. Therefore, the dynamics of the active power P and reactive power Q delivered by the system to the grid are presented by

$$i_{dref} = \frac{2}{3v_{sd}}P_{ref}, \quad i_{qref} = -\frac{2}{3v_{sq}}Q_{ref} \quad (6)$$

where the P_{ref} and Q_{ref} are the reference active and reactive powers [19].

f) DC-link Voltage Controllers: In this case, since a DC-link controller is applied to the system, i_{dref} is given by

$$i_{dref} = (K_{p2} + \frac{K_{i2}}{s})(v_{dcref} - v_{dc}) \quad (7)$$

where v_{dcref} is the nominal DC-link voltage and v_{dc} is the DC-link voltage. Parameters K_{p2} and K_{i2} are the proportional and integral parameters of the PI regulator of the DC-link controller [19].

In summary, a single-stage PV system is represented by a nonlinear state space model formulated in (1) and (3), incorporating two current controllers in equations (4) and (5), a power controller in equation (6), and a DC-link voltage controller in equation (7).

*2) **Two-stage PV**:* A two-stage PV system has a similar structure as a single-stage PV system with an additional DC to DC converter as shown in Fig.2.

a) DC to DC Converter: The DC to DC converter amplifies the input voltage from the PV array by utilizing a maximum power point tracking (MPPT) controller to deliver the target power (P_{PVref}). Dynamics of the DC/DC converter are expressed in following equations [19].

$$\frac{di_{PV}}{dt} = \frac{1}{L_b}v_{PV} - \frac{(1-d_{ref})}{L_b}v_{dc} \tag{8}$$

$$\frac{dv_{dc}}{dt} = \frac{(1-d_{ref})}{C_{dc}}i_{PV} - \frac{1}{C_{dc}}i_{dc} \tag{9}$$

where v_{PV} and i_{PV} are the output voltage and current of the PV array, L_b is the inductor within the DC to DC converter, and C_{dc} is the DC-link capacitor. The input current of the VSC, i_{dc}, can be represented by the power balance equation below [19].

$$v_{dc}i_{dc} = \frac{3}{2}(v_{sd}i_{sd} + v_{sq}i_{sq}) \tag{10}$$

$$i_{dc} = \frac{3}{2v_{dc}}(v_{sd}i_{sd} + v_{sq}i_{sq}) \tag{11}$$

where v_{sd}, v_{sq}, i_{sd} and i_{sq} are the voltages and currents in dq-frame at the PCC. In summary, the dynamics of a two-stage PV system are represented by linear dynamics in (1) and the nonlinear part shown below.

$$\dot{i}_{PV} = \frac{1}{L_b}v_{PV} - \frac{(1-d_{ref})}{L_b}v_{dc} \tag{12}$$

$$\dot{v}_{dc} = \frac{(1-d_{ref})}{C_{dc}}i_{PV} - \frac{1}{C_{dc}}\frac{3}{2v_{dc}}(v_{sd}i_{sd} + v_{sq}i_{sq}) \tag{13}$$

b) PV Power Controller: As shown in Fig. 4, besides the current controller, DC-link voltage controller and reactive power controller, a PV power controller is applied to the system. In equations (8) and (9), the new variable d_{ref} is defined as the duty cycle setpoint of the pulse width modulation (PWM) block represented by

$$d_{ref} = (K_{p3} + \frac{K_{i3}}{s})(i_{PVref} - i_{PV}) \tag{14}$$

where i_{PV} is the output current of the PV array and proportional parameter K_{p3} and integral parameter K_{i3} are the parameters of the PI regulators in the current controller. Given that $P_{PVref} = v_{PV}i_{PVref}$, equation (14) can be rewritten as

$$d_{ref} = (K_{p3} + \frac{K_{i3}}{s})(\frac{P_{PVref}}{v_{PV}} - i_{PV}) \tag{15}$$

in which P_{PVref} is the reference power of the PV array, which is given by the MPPT controller, and v_{PV} is the output voltage of the PV array. Since the dynamics of the MPPT controller is much slower than current and DC voltage controller, they are not considered for data-driven modeling [19]. Instead, P_{PVref} is assumed to be known.

Hence, a two-stage PV system can be represented by a nonlinear state space model of equations (1), (12) and (13), incorporating two current controllers in equations (4) and (5), a DC-link voltage controller in equation (7), a reactive power (Q) controller, and a PV power controller in equation (15).

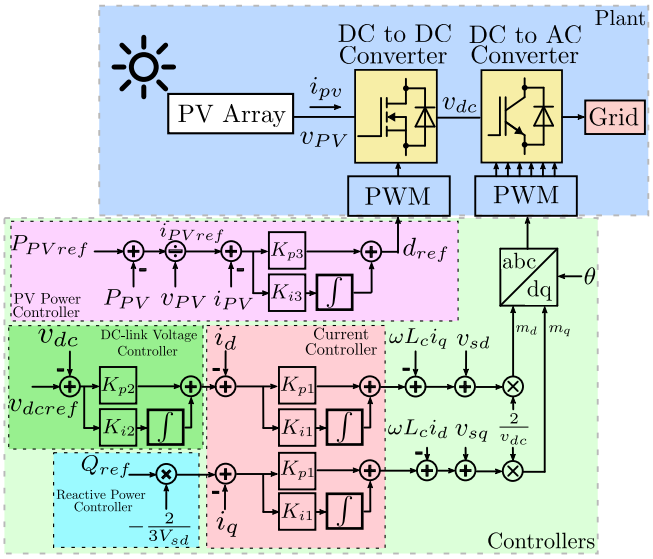

Fig. 4 Control diagram of two-stage PV system.

B. Closed-loop Dynamic Model of PV systems

In an open-loop case, the PV plant model, excluding the controllers, will be replaced by a data driven model. In a closed-loop case, the combination of the PV plant and the controller will be replaced by a data-driven model.

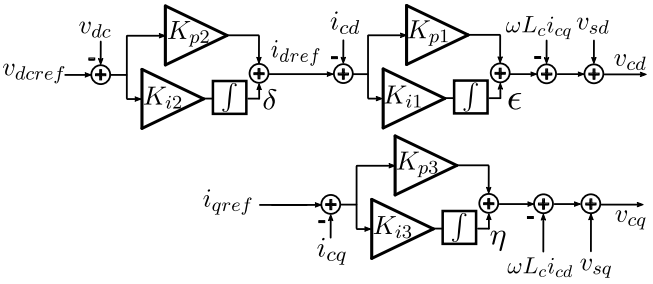

Fig. 5 PI controllers of the PV system.

*a) **Single-Stage System Controller Modeling**:* Fig. 5 illustrates the structure of three PI controllers presented in equations (4), (5), and (7). For modeling of controllers, the outputs of integrators are treated as new state variables of the system, labeled as δ, ϵ and η. Dynamics of these three states are expressed as

$$\dot{\delta} = (v_{dcref} - v_{dc})K_{i2} \tag{16}$$

$$\dot{\epsilon} = ((v_{dcref} - v_{dc})K_{p2} + \delta - i_{cd})K_{i1} \tag{17}$$

$$\dot{\eta} = (i_{qref} - i_{cq})K_{i3} \tag{18}$$

The dynamic of the inverter's voltages v_{cd} and v_{cq} are then defined by the following equations

$$\begin{aligned} v_{cd} =&((v_{dcref} - v_{dc})K_{p2} + \delta - i_d)K_{p1} \\ &+ \epsilon - L_c\omega_0 i_{cq} + v_{sd} \end{aligned} \tag{19}$$

$$v_{cq} = ((i_{qref} - i_{cq})K_{p3} + \eta + L_c\omega_0 i_{cd} + v_{sq} \tag{20}$$

A closed-loop single-stage PV system can be represented by a nonlinear state space model with equations (1), (3), (16), (17),

and (18). This representation is also used in cases studies in Section V.

*b) **Q-V Droop Controller Modeling**:* The proposed approach enables formulation of additional control functions, such as droop control. This section explores the potential of modeling the Q-V droop controller for distributed generation, as defined in IEEE Std. 1547 [21]. A similar approach can be used to model various types of droop controllers.

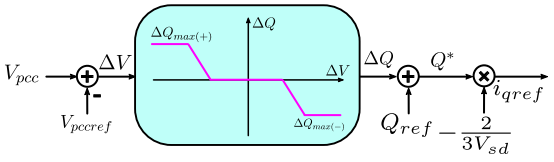

Fig. 6 Schematic of Q-V droop controller.

To regulate the injected reactive power and voltage at the PCC, the Q-V droop controller is implemented by adjusting the reactive power setpoint i_{qref} in Fig. 5 using ΔQ through droop control, as shown in Fig. 6.

$$Q^* = Q_{ref} + \Delta Q \tag{21}$$

where the ΔQ is formulated as [22]

$$\Delta Q = \begin{cases} (-1/m_Q)\Delta Q & |\Delta V| > V_{th} \\ 0 & |\Delta V| \leq V_{th} \end{cases} \tag{22}$$

where ΔV is the threshold voltage determined by the grid, typically 5% of the reference voltage, and m_Q is the droop control coefficient. Additionally, ΔQ should be constrained within the lower and upper limits of the droop output, $\Delta Q_{max}(-)$ and $\Delta Q_{max}(+)$, respectively. Following the droop control, a reactive power controller is formulated to generate the output current of the converter i_{qref},

$$i_{qref} = -\frac{2Q^*}{3V_{sd}} = -\frac{2(Q_{ref} + \Delta Q)}{3V_{sd}} \tag{23}$$

By replacing i_{qref} in (20) with (23), dynamics of the inverter voltage v_{cq} with Q-V droop control can be obtained. Given that V_{PCC} is measurable, one can identify the Q-V droop control dynamics (i.e., the unknown in (22), which is m_Q) using the proposed approach.

To conclude, the closed-loop single-stage PV system with a Q-V droop controller can be represented by a nonlinear state space model with equations (1), (3), (16), (17), and (18), with the Q-V droop controller described by equations (21) (22), and the reactive power controller detailed in equation (5).

III. Sparse Identification of PV Systems

In a real-world scenario, where access to every component in the system is not feasible, data-driven modeling becomes essential. In this research, we developed our approach based on a sparsity promoting method to identify parameters in the models and construct data-driven models for both single-stage and two-stage PV systems. The proposed framework can be generalized to identify any nonlinear dynamics. The proposed sparsity-promoting model identification focuses on finding models with minimum number of parameters, while data fitting algorithms aim to fit a model to data as accurately as possible, often without considering the complexity or sparsity of the model. Sparsity-promoting techniques produce simpler, more interpretable models by reducing the number of parameters, which can enhance generalization and reduce overfitting compared to traditional data-fitting algorithms. This section will primarily introduce the sparse regression method and our proposed methodology.

Consider a dynamic system, i.e., an open-loop single-stage PV system represented by equations (1) and (3), an open-loop two-stage PV system represented by equations (1), (12) and (13), or a closed-loop single-stage PV system represented by (1), (3), (16), (17) and (18), following a general state-space model

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)). \tag{24}$$

Here, the state variable vector of a system at time t is represented by $\mathbf{x}(t) \in \mathbb{R}^n$ and the inputs to the system at time t are represented by $\mathbf{u}(t) \in \mathbb{R}^m$, where n and m denote the dimensions of the state variables and inputs. In this paper, the state variables and inputs of an open-loop single-stage PV system are denoted as: $X_1 = [i_{cd}, i_{cq}, i_{gd}, i_{gq}, v_{sd}, v_{sq}, v_{dc}]$ and $U_1 = [v_{cd}, v_{cq}, v_{gd}, v_{gq}, \omega_0, i_{PV}]$. For an open-loop two-stage PV system, the state variables and inputs are represented as: $X_2 = [i_{cd}, i_{cq}, i_{gd}, i_{gq}, v_{sd}, v_{sq}, v_{dc}, i_{PV}]$ and $U_2 = [v_{cd}, v_{cq}, v_{gd}, v_{gq}, \omega_0, v_{PV}, d_{ref}]$. In the case of a closed-loop single-stage PV system, the state variables and inputs are specified as: $X_3 = [i_{cd}, i_{cq}, i_{gd}, i_{gq}, v_{sd}, v_{sq}, v_{dc}, \delta, \epsilon, \eta]$ and $U_3 = [v_{dcref}, i_{qref}, v_{gd}, v_{gq}, i_{PV}]$. The goal in this section is to identify a data-driven version of the function $\mathbf{f}(\mathbf{x}(t), \mathbf{u}(t))$ represented as $\dot{\mathbf{X}} = \Theta(X, U)\Xi$ for the PV systems from available data. This process is explained in the following. For other assets within smart grids, such as distributed generation or flexible AC transmission systems (FACTS), as long as the state variables and inputs are known and can be measured, the proposed method can be applied.

A. Data Collection

According to (24), the time history of state variables $\mathbf{x}(t)$ and inputs $\mathbf{u}(t)$ need to be collected [11], and the derivative of states $\dot{\mathbf{x}}(t)$ can either be measured or numerically approximated from $\mathbf{x}(t)$. The data then are sampled and organized as three matrices

$$\mathbf{X} = \begin{bmatrix} | & | & & | \\ x_1(t_k) & x_2(t_k) & \cdots & x_n(t_k) \\ | & | & & | \end{bmatrix} time \downarrow \tag{25}$$

$$\mathbf{U} = \begin{bmatrix} | & | & & | \\ u_1(t_k) & u_2(t_k) & \cdots & u_m(t_k) \\ | & | & & | \end{bmatrix} time \downarrow \tag{26}$$

$$\dot{\mathbf{X}} = \begin{bmatrix} | & | & & | \\ \dot{x}_1(t_k) & \dot{x}_2(t_k) & \cdots & \dot{x}_n(t_k) \\ | & | & & | \end{bmatrix} time \downarrow \tag{27}$$

$$\Theta(\mathbf{X},\mathbf{U}) = \begin{bmatrix} | & | & | & | & | & | & | & | \\ 1 & U & \mathbf{X} & (\mathbf{X},\mathbf{U})^{P_2} & (\mathbf{X},\mathbf{U})^{P_3} & \cdots & sin((\mathbf{X},\mathbf{U})) & cos((\mathbf{X},\mathbf{U})) & \cdots \\ | & | & | & | & | & | & | & | \end{bmatrix} time \downarrow \tag{28}$$

$$(\mathbf{X},\mathbf{U})^{P_2} = \begin{bmatrix} | & | & | & | & | & | & | & | & | \\ \mathbf{u_1^2}(t_k) & \mathbf{u_1}(t_k)\mathbf{u_2}(t_k) & \cdots & \mathbf{u_1}(t_k)\mathbf{x_1}(t_k) & \cdots & \mathbf{u_m^2}(t_k) & \mathbf{x_1^2}(t_k) & \mathbf{x_1}(t_k)\mathbf{x_2}(t_k) & \cdots & \mathbf{x_n^2}(t_k)s \\ | & | & | & | & | & | & | & | & | \end{bmatrix} \tag{29}$$

B. Library Construction

The next step is to construct a library $\Theta(\mathbf{X},\mathbf{U})$ containing all the candidate functions of the state variables in the columns of $\mathbf{X}$ and $\mathbf{U}$. These candidate functions may include linear functions of state variables $X = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}$, polynomial functions such as $x_i + x_i^2 + \dots x_i^k \ \forall i \in \{1,2,\dots,n\}$, sinusoidal functions such as $\sin(x_i) \ \forall i \in \{1,2,\dots,n\}$, or polynomial/sinusoidal functions of input/output variables such as $\sin(x_i u_j)$ or $x_i u_j^2 \ \forall i \in \{1,2,\dots,n\}, \ j \in \{1,2,\dots,m\}$. Candidate functions in this paper are illustrated in equation (28) and (29), where $(\mathbf{X},\mathbf{U})^{P_2}$ and $(\mathbf{X},\mathbf{U})^{P_3}$ represent second and third-order polynomials of $\mathbf{X}$ and $\mathbf{U}$, respectively. Specifically, $(\mathbf{X},\mathbf{U})^{P_2}$, with second-order terms, is shown in equation (29). During the regression process, the construction of the library can significantly influence the identification results. It is noted that knowledge of the dynamics (full or partial) can significantly improve the accuracy of the identification, as it removes the unrelated library functions from the process and speeds up the identification step.

To test this hypothesis, we increased the library of candidate functions from 13 to 60 candidate functions. Our findings indicate that by selecting appropriate sparsity-promoting hyperparameters, the algorithm can effectively nullify coefficients for most irrelevant candidate functions, ensuring they do not impact our results. However, adding more library functions increases the computation time costs for identification when using a larger number of candidate functions. In practice, if full knowledge of system dynamics is not available, partial knowledge of dynamics can help identify potential candidate functions and reduce the overall number of candidate functions in the library, which eventually increases the computational efficiency of identification phase. For example, in this paper, it is known that dynamics of PV system will not have a sinusoidal term involved due to the nature of dq transformation. Therefore, the sinusoidal terms from the library function can be removed.

C. Sparse Regression

Each column of $\Theta(\mathbf{X},\mathbf{U})$ represents a potential candidate function for the right-hand side of equation (24). The matrix $\Theta(\mathbf{X},\mathbf{U})$ allows considerable flexibility in selecting the specific nonlinearities for inclusion in the candidate functions. Since it is believed that only a small number of these nonlinearities are active in each row of $\mathbf{f}(\mathbf{x}(t),\mathbf{u}(t))$, a sparse regression problem can be formulated to identify the sparse coefficient vectors $\Xi = [\xi_1,\xi_2,\dots,\xi_k,\dots,\xi_n]$ indicating which nonlinearities are active. The process can be articulated as an optimization problem

$$\xi_k = \underset{\xi_k'}{\operatorname{argmin}} \left\| \dot{\mathbf{X}}_\mathbf{k} - \mathbf{\Theta}(\mathbf{X},\mathbf{U})\xi_\mathbf{k}' \right\|_2 + \lambda \left\| \xi_k' \right\|_0 \tag{30}$$

where $||\cdot||_2$ denotes the $L2$-norm, measuring the overall magnitude of the vector, $\left\| \dot{\mathbf{X}}_\mathbf{k} - \mathbf{\Theta}(\mathbf{X},\mathbf{U})\xi_\mathbf{k}' \right\|_2$ represents the optimization problem aiming to find the coefficient vector ξ_k that minimizes the disparity between the measured data $\dot{X}_k$ and the data-driven model $\Theta(X,U)\xi_k'$. In addition, $||\cdot||_0$ denotes the $L0$-norm, the number of nonzero elements in the coefficient vector [5]. The parameter λ serves as the trade-off between fitting the data well and promoting sparsity in the solution. It penalizes large values and encourages many coefficients to be exactly zero. In other words, the hyperparameter determines which candidate functions are relevant. If a candidate function is deemed irrelevant, it is assigned a coefficient of "0" and is excluded in subsequent iterations. If the hyperparameter is set too small, even relevant candidate functions may be excluded, leading to inaccurate results. Conversely, if the hyperparameter is set too large, irrelevant candidate functions may be included, resulting in a complex and inaccurate outcome. Selecting an appropriate value for λ is crucial in the sparse regression approach, which is thoroughly discussed in section V.A.

D. Adaptive Regulated Sparse Regression

In order to better evaluate the performance of the method, we first introduce our optimization criteria, root mean square error (RMSE), which measures the error between the true and predicted values of the k-th state. It is denoted as:

$$RMSE\{Data(\Xi), Phy\}_k = \sqrt{\frac{\sum_{i=1}^{P}(y_{kpi} - y_{kdi})^2}{P}} \tag{31}$$

where P represents the number of data points, y_{kpi} denotes the measurement of k-th state variable of the physical model at data point i, and y_{kdi} represents the prediction of k-th state variable of the data-driven model for data point i. The RMSE values serve as a quantitative measure of the dissimilarity between the physical model and the data-driven model outputs, providing insight into the accuracy and performance of the data-driven model.

Conventional sparse regression approaches employ one single λ value for all state variables, which poses challenges in managing complex dynamics such as the closed-loop model of PV systems. This challenge arises because the coefficients of distinct terms in the differential equations for each state variable may exhibit notable variations. For instance, in equation (16) for δ, the coefficient K_{i2} is 0.01, while in equation (17) describing the dynamic of ϵ, the coefficient K_{i1} is 500. As discussed in the previous section, to include a coefficient whose value is 500, the hyperparameter "λ" should be set to at least above 500 (if a coefficient has a true value much smaller than λ, it may be eliminated by the sparcification). However, for coefficient with a value 0.01, using the same λ value is too large and may introduce irrelevant functions. Identifying such coefficients proves to be difficult using a single λ. To address this issue, a new algorithm has been developed, which is described as an optimization problem in the following

$$\xi_k = \underset{\xi'_k}{\operatorname{argmin}} \left\| \dot{\mathbf{X}}_\mathbf{k} - \mathbf{\Theta}(\mathbf{X}, \mathbf{U})\xi'_\mathbf{k} \right\|_2 + \lambda_k \left\| \xi'_k \right\|_0 \tag{32}$$

$$\lambda_k = \underset{\lambda_k}{\operatorname{argmin}}\, RMSE\{Data(\Xi), Phy\}_k \tag{33}$$

where ξ_k is the k-th column of coefficient matrix Ξ, $\dot{X}_k$ is the derivative of k-th state and λ_k is the k-th regularization parameter. It is worth mentioning that the selection and combination of state variables in the library function $\Theta(X, U)$ depend on the understanding of the system dynamics. In this paper, we assume a partial knowledge of the dynamics of the system and control (i.e., control mode of the PV system). The key enhancement lies in the algorithm's capability to adapt the regularization parameter λ dynamically while identifying distinct state variables of the system. As shown in equation (32), parameter λ is represented as a vector $\mathbf{\Lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_k, \ldots, \lambda_n]$, where the number of elements corresponds to the number of state variables, and the hyperparameter for k-th state variable, λ_k, is determined by equation (33). The sparse regression method outlined in equation (32) is denoted as $SparseRegression\{\dot{\mathbf{X}}, \Theta(\mathbf{X}, \mathbf{U}), \mathbf{\Lambda}\}$.

Algorithm 1 presents the ARSR, defined as **AdaptiveSINDy**$\{\dot{\mathbf{X}}, \mathbf{\Theta}(\mathbf{X}, \mathbf{U})\}$, which can be summarized in three main steps: 1) Initialization, 2) Data-driven modeling, and 3) RMSE calculation and Optimization loop. It is crucial to emphasize that to compute the RMSE, a data-driven model should be constructed using training data (discussed in Section IV). Next we calculate the RMSEs between the measured state variables and data-driven model predictions and compare them with the test data. Moreover, since the relationship between regularization parameters and state variable error is non-linear, the algorithm updates the parameters using a predefined step unit reaching the minimum state variable error. More specifically, the step change is defined based on the desired accuracy. Initially, a larger step change, such as "1", is used and then gradually reduced , and the process is iteratively continued until the reconstructed data from the identified model ($\hat{X}$) accurately presented the measured data (X). This process continues with progressively smaller step changes until the data-driven model achieves the desired accuracy.

Compared to the conventional sparse regression method, allowing the use of different hyperparameters based on varying state variables enables the proposed ARSR method to maintain its ability to identify simple dynamics and handle more complex systems with greater differences between state variables. Additionally, the proposed method requires less data. In our tests, as few as 8,000 to 200,000 data points were sufficient to ensure accurate results. In contrast, other data-driven methods such as deep learning required 3 million data points to achieve the same accuracy [23]. However, compared to conventional sparse regression [14], the proposed method can manage more complex models with diverse range of nonlinearities such as closed-loop modeling of PV systems. The main reason is that conventional sparse regression utilizes a fixed hyperparameter, which might not capture the coefficients that are significantly different. For example, if dynamics are represented as $\dot{x} = \begin{bmatrix} 0.001 \\ \vdots \end{bmatrix} x + \begin{bmatrix} \vdots \\ 560 \end{bmatrix} u$, due to the large difference between coefficients $(0.001, 560)$, a single hyperparameter in conventional sparse regression cannot accurately represent this system, while our proposed method utilizes two hyperparameters, which can easily account for the large difference between these two coefficients. The performance of the ARSR is further discussed in section V.A and V.B.

IV. Data-Driven Modeling and Control

Using the outcomes of the proposed ARSR, it becomes feasible to construct data-driven controllers for both open-loop and closed-loop PV systems.

A. Open-loop Data-driven Modeling with Controllers

Fig. 7 depicts the control of the physical model of both PV systems consisting of a DC-link voltage controller, a reactive power controller, and a current controller in dq-frame. The two-stage PV system additionally includes a controller for regulating the PV power via a DC-to-DC converter. The dynamics of the data-driven DC-link voltage controller and the reactive power controller should mirror those of the physical model. The outputs of the data-driven controllers will determine the converter voltages in the dq frame.

B. Data-driven Controllers Design

The design of controller gains for these current controllers in Fig. 7 can now be accomplished based on the identified data-driven model using ARSR. As shown in Fig. 8, a simplified block diagram of current control loop is utilized for designing the bandwidth of controllers in the data-driven model [20].

According to [20], the simplified plant model is composed of L and r_c on the AC-side dynamics. The data-driven plant dynamics are denoted by $\hat{L}$ and $\hat{r_c}$, which can be

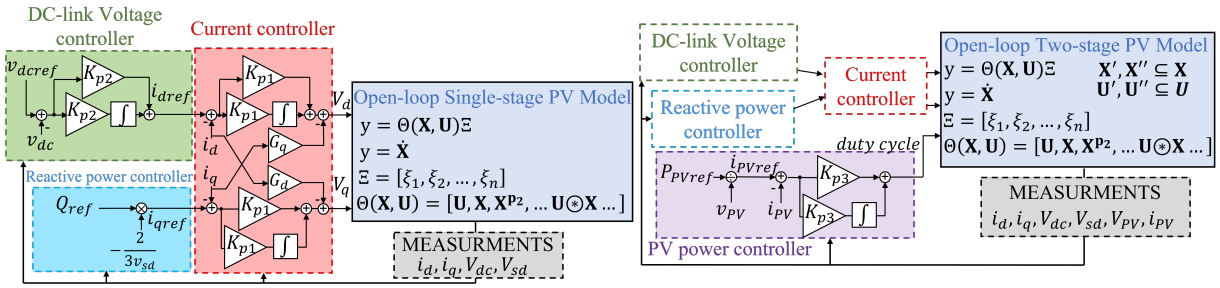

Fig. 7 Data-driven control design for single-stage and two-stage PV systems.

Algorithm 1 Adaptive Regulated Sparse Regression

Inputs: Measured derivatives $\dot{\mathbf{X}}$, library $\Theta(\mathbf{X}, \mathbf{U})$, Optimized states array $StateX$

Output: Sparse regression model coefficients Ξ, Optimal regularization parameter vector Λ.

Initialization :
$\Lambda = [\lambda_1, \lambda_2, ..., \lambda_n]$; ▷ Init λs array
λ_{max}, $Steps$; ▷ set the upper limit and unit steps

AdaptiveSINDy$\{\dot{\mathbf{X}}, \Theta(\mathbf{X}, \mathbf{U})\}$
$\hat{\Xi} = SparseRegression\{\dot{\mathbf{X}}, \Theta(\mathbf{X}, \mathbf{U}), \mathbf{\Lambda}\}$; ▷ find $\hat{\Xi}$ with initial λs
$err = RMSE\{Data(\hat{\Xi}), Phy\}$; ▷ find the RMSEs for each state between physical model and data-driven model with current coefficient matrix $\hat{\Xi}$
$Ind = max(err)$ ▷ find the state with the largest RMSE
while $Ind \leq n$ **do** ▷ n is number of states
  $\hat{\Lambda} = \Lambda$; ▷ update temporary lambda array
  for $\hat{\Lambda}(Ind) = \Lambda(Ind) : Steps(Ind) : \lambda_{max}(Ind)$ **do**
    $\hat{\Xi} = SparseRegression\{\mathbf{X}, \dot{\mathbf{X}}, \Theta(\mathbf{X}), \hat{\Lambda}\}$;
    $err = RMSE\{Data(\hat{\Xi}), Phy\}$;
    if $err(Ind)$ is reduced **then**
      $\Lambda(Ind) = \hat{\Lambda}(Ind)$
    end if
  end for
  $StateX \leftarrow Ind$ ▷ update $StateX$ with optimized states
  $Ind = max(err)$
  while $Ind \in StatesX$ **do**
    $Ind = nextmax(err)$ ▷ find the state with next largest RMSE
    if $length(StateX) = n$ **then**
      $\Xi = \hat{\Xi}$; ▷ update coefficient matrix
      return ▷ The algorithm ends with all states are optimized
    end if
  end while
end while

obtained through the sparse regression algorithm, as specified in equation (1). Taking the d-axis controller as an example, the transfer function of current controller $k_d(s)$ can be expressed as [20]

$$k_d = \frac{k_p s + k_i}{s} \tag{34}$$

where k_p and k_i denote the proportional and integral gains of the current controller, respectively. The loop gain is expressed as

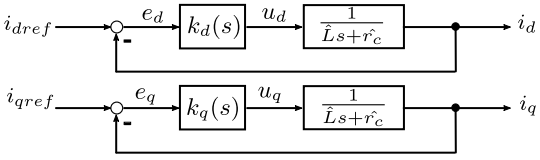

Fig. 8 Simplified closed-loop model of current controllers.

$$l(s) = \left(\frac{k_p}{Ls}\right)\frac{s + k_i/k_p}{s + r_c/L} \tag{35}$$

Due to the pole of the loop gain at $s = -r_c/L$, an ideal loop gain takes the form $l(s) = k_p/Ls$, where the pole is cancelled by the compensator zero at $s = -k_i/k_p$. Consequently, the closed-loop transfer function becomes

$$G(s) = \left(\frac{l(s)}{1 + l(s)}\right) = \frac{k_p/L}{s + k_p/L} = \frac{1}{\tau_i s + 1} \tag{36}$$

where τ_i is the time constant of the current controller represented by $\tau_i = L/k_p$. By replacing L and r_c with their data-driven counterparts obtained via the proposed ARSR method, a data-driven control design can be achieved. Thus the data-driven proportional and integral gains k_p and k_i are obtained.

$$k_p = \frac{\hat{L}}{\tau_i},\ k_i = \frac{\hat{r}_c}{\tau_i} \tag{37}$$

Similar concept can be used to obtain the data-driven controllers design for other controllers such as DC voltage and power controller. By utilizing the constructed open-loop models and controllers, data-driven single-stage and two-stage PV systems can be developed based on the outcomes of the ARSR algorithm.

V. Case Studies

To evaluate the effectiveness of the proposed ARSR approach for modeling and control of PV systems, various case studies are conducted using time-domain and real-time simulations in MATLAB and Opal-RT.

TABLE I Impact of sampling frequency on accuracy of data-driven model identification.

Sample Time	Data Points	States RMSEs						
		x_1	x_2	x_3	x_4	x_5	x_6	x_7
$1 \times 10^{-5}s$	400,000	0.0641	0.0068	0.0196	0.2088	2.0859	0.3564	1.2887
$1 \times 10^{-4}s$	40,000	0.0641	0.0069	0.0196	0.2088	2.0859	0.3567	1.2886
$1 \times 10^{-3}s$	4000	0.0641	0.0068	0.0195	0.2088	2.0858	0.3559	1.2881
$1 \times 10^{-2}s$	400	17.2464	1.0540	226.6993	5.0518	62.6753	2234.8427	945.6372

A. *Impact of Sampling Time*

This case study analyzes the impact of sampling time on the accuracy of the proposed approach for identifying the dynamics of a single-stage PV system. In all scenarios, the simulation was run for 4 seconds and the sampling time of data collection was modified from 10 μs to 10 ms and the RMSE of prediction of state variables is calculated as presented in Table I. It is observed that the RMSE of state prediction was not significantly impacted until the sampling time was reduced to $1e^{-2}$ seconds (resulting in 400 data points). It is noted that the accuracy of the proposed approach significantly reduces when the sampling frequency reduces to 100 Hz (0.01 sampling time). This highlights the need for having phasor measurement units (PMUs) installed for a successful identification or utilization of interpolation techniques if the sampling frequency of measurement devices is not above 100 Hz.

B. *Impact of hyperparameter λ for PV systems*

This section demonstrates the impact of various λ values on the identification of open-loop models for both single-stage and two-stage PV systems.

1) ***Impact of λ in single-stage PV models:*** Fig. 9 and the left part of Table II illustrate the tracking performance of the single-stage PV system with two reference inputs: the DC-link voltage v_{dc} and reactive power Q. As depicted in the figure, the data-driven models closely approximate the physical model in the majority of cases when λ ranges from 1 to 35. Notably, in Table II, the model achieves its best performance when λ ranges from 5 to 35.

2) ***Impact of λ in Two-stage PV Models:*** Fig. 10 and right part of the Table II illustrate the tracking performance of the two-stage PV model with three reference inputs: the DC-link voltage v_{dc}, reactive power Q, and PV array power P_{PV}. When λ ranges from 1 to 30, the data-driven models closely approximate the physical model. It is observed that v_{dc} exhibits optimal performance when λ is set to 30. On the other hand, P_{PV} and Q demonstrate their best performance when λ changes to 5 and 1, respectively.

3) ***Analysis of Adaptive Method:*** To further improve the results, the adaptive method outlined in section III.D is applied to the open-loop modeling process for both single-stage and two-stage PV systems. The resulting regulation parameters are $\mathbf{\Lambda_1} = [5, 5, 5, 5, 30, 30, 20]$ and $\mathbf{\Lambda_2} = [1, 5, 5, 5, 30, 30, 1, 35]$, respectively. Examining the tracking performance in the last row of Table II, it is evident that the single-stage models do not exhibit significant improvement since the conventional method already achieves the optimal performance. However, the two-stage models demonstrate a significant enhancement, indicating that the adaptive method yields better results in modeling complex systems.

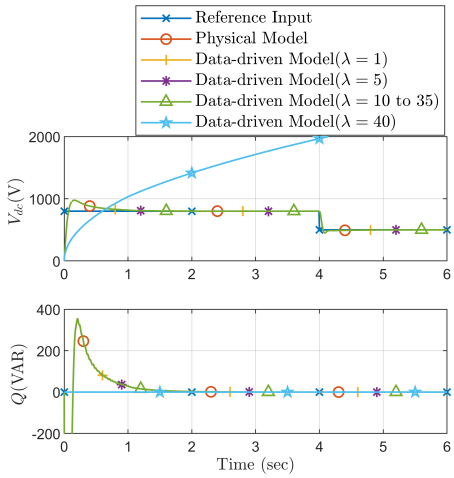

Fig. 9 Impact of varying λ for identification of single-stage PV system.

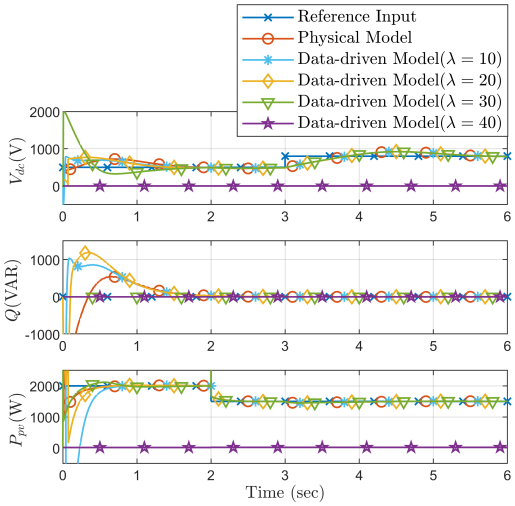

Fig. 10 Impact of λ on identification of two-stage PV system.

In conclusion, the results obtained from the identification process in this section demonstrate that varying the values of λ can significantly impact the performance of the sparse regression approach.

TABLE II References tracking performance (RMSEs) of single-stage and two-stage PV systems.

λ	***Single-stage***		***Two-stage***		
	v_{dc} *error*	Q *error*	v_{dc} *error*	Q ***error***	P_{PV} ***error***
1	11.3752	28.9546	39.9779	30.7711	212.33
5	1.2882	3.2921	39.7971	27.3171	296.048
10	1.2882	3.2921	115.5112	59.6551	1930.4
15	1.2882	3.2921	98.9295	51.8175	1913.9
20	1.2882	3.2921	98.9295	51.8175	1913.9
25	1.2882	3.2921	98.9295	51.8175	1913.9
30	1.2882	3.2921	36.9085	339.3858	403.3931
35	1.2882	3.2921	36.9085	339.3858	403.3931
40	644.5295	358.758	696.4531	339.3858	1651.7
Adpt.	1.2882	3.2921	36.9085	26.6831	6.1833

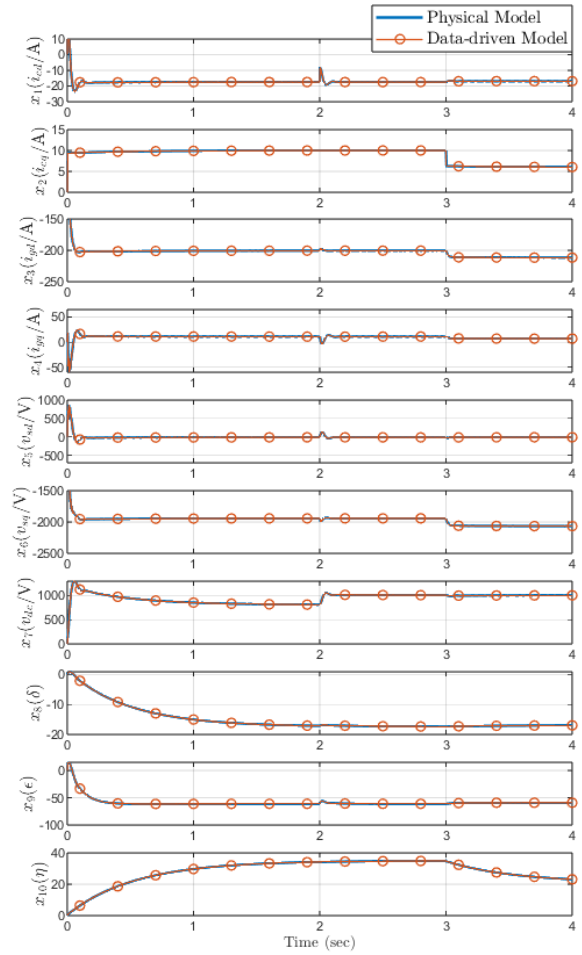

Fig. 11 Comparison between the state variables of physical model and closed-loop data-driven model.

C. Closed-loop Modeling for Single-stage PV Systems

This case study primarily addresses the closed-loop model identification of a single-stage PV system using the proposed ARSR algorithm. After employing measured data from the single-stage PV system and utilizing the obtained sparse matrix of coefficients Ξ via ARSR, the data-driven model is formulated. To validate its effectiveness, a comparison is conducted with the physical model. The outcomes of regulation parameters λ and state errors (RMSEs) are presented in Table III and Fig. 11. The rationale for selecting states to assess the system is based on the differences between open-loop and closed-loop models. In the open-loop model with controllers, multiple reference inputs are used for regulation. However, in the closed-loop model, it is more appropriate to compare all the states of the system to obtain a comprehensive evaluation. The results illustrate that the data-driven model exhibits high accuracy when compared to the original physical model. This conclusion is supported by the small state variable errors detailed in Table III. The outcomes of the closed-loop modeling in this section demonstrate the effectiveness of the ARSR algorithm in closed-loop sparse regression modeling for a single-stage PV system.

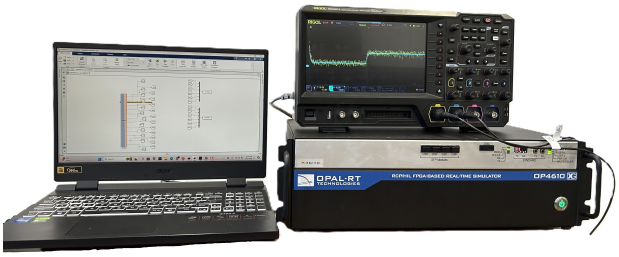

Fig. 12 Real-time verification setup.

D. Real-time Simulation Verification

This case study focuses on real-time simulation of the physical and data-driven models to verify the feasibility and performance of the proposed ARSR approach in real-world settings. The validation process is conducted using the Opal-RT OP4610XG real-time simulator. It involves comparing the single-stage PV system's physical model with its corresponding open-loop data-driven model, as described in Section II. Both the physical model and the data-driven model were constructed in MATLAB Simulink, transformed into real-time models using RT-LAB software, and run in the OP4610XG via an Ethernet connection. Subsequently, the simulation results are observed through the analog outputs of the OP4610XG using a digital oscilloscope, as shown in Fig.12.

a) Verification of Single-stage PV system modeling : As depicted in Fig. 13, screenshots captured from the oscilloscope display six state variables of the two models. The red curves represent the state variables of the data-driven model, which closely match the blue curves that represent the state variables of the physical model. Results confirm the ARSR method's capability and feasibility in accurately replicating real-world physical systems. It is worth mentioning that, there are slight differences between the real-time experiments and simulation results such as high-frequency ripples on the real-time data as shown in Fig. 13, which are attributed to the detailed nature of real-time simulations with power electronics switching functions, which have not been included

TABLE III RMSE of prediction in data-driven ARSR-based closed-loop single-stage PV system.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
λ	1	1	2	1	7	19	1	0.01	1	1
State Error	0.0409	0.0006	0.0244	0.1548	1.5791	0.2592	0.8664	0.0009	0.0181	0.0008

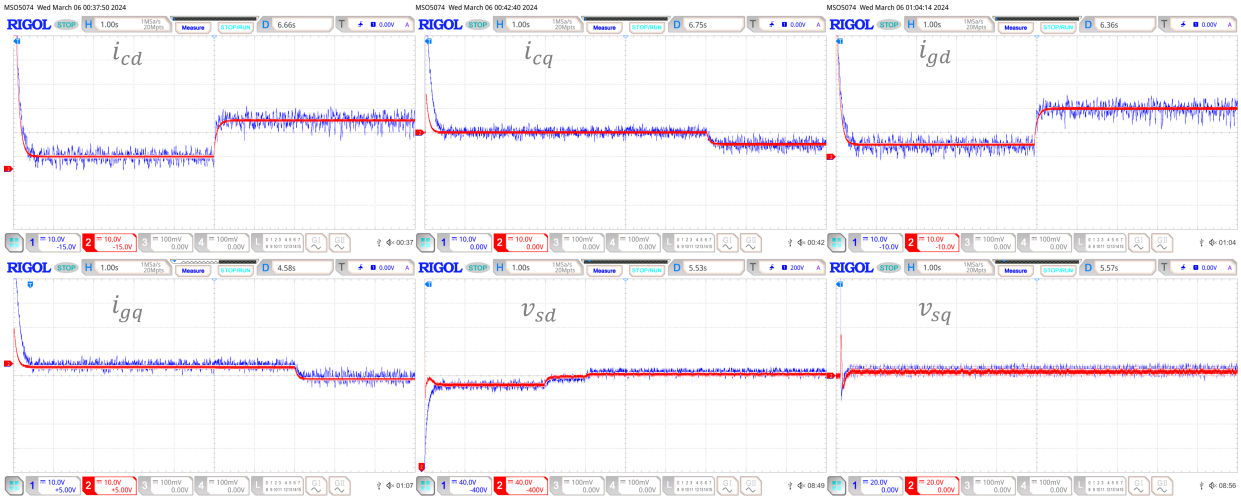

Fig. 13 Real-time verification of 6 states in single-stage PV system.

in the simulations. For example, in real-time simulations, an additional high-frequency ripple is observed, which is related to the switching frequency of the transistors in the power converters. Such switching ripples were not observed in the simulations as averaged converter models were used.

b) Verification of Single-stage PV system under Fault: This case study introduces data-driven fault analysis capability of the ARSR approach in single-stage PV systems. The proposed framework is depicted in Fig. 14. To ensure a fair and consistent comparison, all the reference inputs are set to be constant throughout this section. Figs. 15 and 16 depict the results for the verification of undervoltage and grounded faults, respectively. In these figures, the red curves represent the output currents (i_{cd}, i_{cq}) of the inverter in the data-driven model, while the blue curves represent the same state variables in the physical model. The tests lasted for 4 seconds, during which the grid voltage dropped from 380V to 200V at 2 seconds for the undervoltage fault, and the three-phase was grounded at 2 seconds for the grounded fault. Overall, the red and blue curves in both tests exhibit similar trends, demonstrating that the proposed method captures the characteristics of the physical model under both fault conditions. The slight differences between the curves after the fault are primarily attributed to the data-driven model representing an averaged system without the dynamics of power electronics and switches. The results confirm close agreement between the data-driven model and physical model for capturing the dynamics of the PV system during faults. The results confirm close agreement between the data-driven model and physical model for capturing the dynamics of the PV system during faults.

A major limitation of a model-based approach is its inability to account for real-time system parameter changes. For example, the grid impedance is constantly changing depending on the topology of the grid and that would impact the voltage at the point of common coupling during transients. A model-based PV generation control that utilizes a fixed grid impedance cannot deal with these dynamic changes. On the other hand, a data-driven approach, such as the proposed method, can re-identify the dynamics of the system in close to real-time and capture these parameter changes. As a result, the controller parameters can be updated to account for these changing scenarios.

VI. Conclusion

This paper introduces ARSR algorithm and its application in data-driven modeling for single-stage and two-stage PV systems. Using measurements, sparse identification of nonlinear dynamics is employed to identify the dynamic models of PV systems. Open-loop and closed-loop data-driven models are developed for data-driven modeling of PV systems and the results are compared with physical models. The results confirm a close agreement with optimized root mean squared error of prediction for the data-driven models. The proposed adaptive hyperparameter tuning approach increases the efficiency and accuracy of conventional sparse regression technique by adaptively identifying the hyperparameters that optimizing the error of prediction. Application of the proposed ARSR algorithm in fault study highlights the data-driven model's ability to simulate grid faults accurately. Finally, validations through real-time simulation confirm the applicability of the proposed approach for eliminating the dependency of existing PV controllers' reliance on

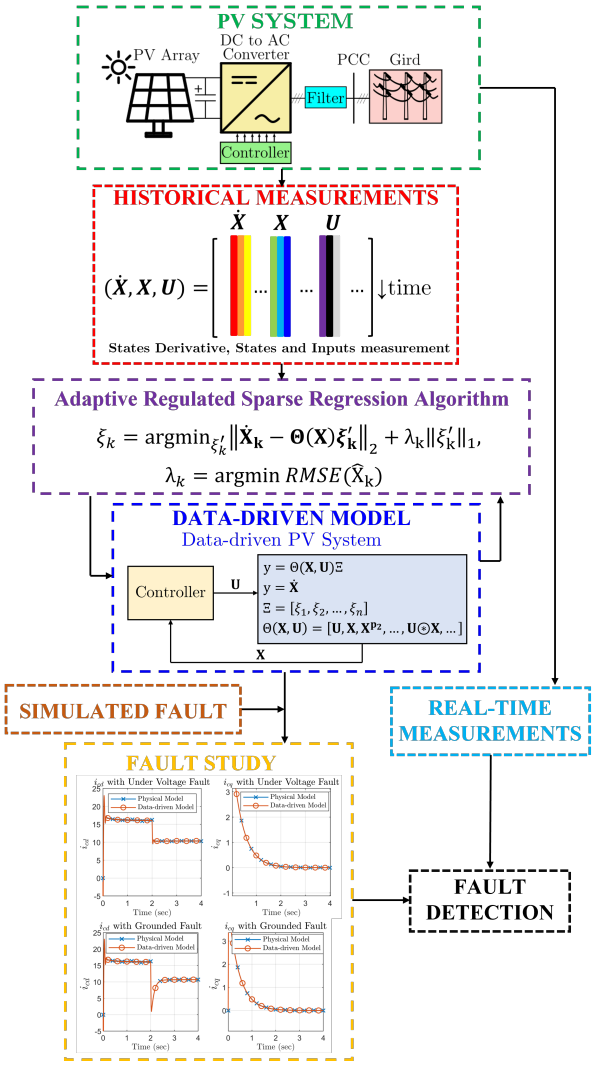

Fig. 14 Fault study and fault detection of single-stage PV system.

known physical models. The proposed method may introduce additional computational costs due to the introduction of a set of sparsity-promoting hyperparameters for dealing with complex dynamics. However, compared to the conventional sparse regression that might not accurately identify the complex dynamics, this step is necessary to optimize the performance. Therefore, the additional computational cost introduced by the proposed adaptive sparsity promoting method is compensated by its superior performance for identifying complex dynamics with diverse coefficients. Future research will aim to create precise data-driven models in noisy environments, focusing on exploring state estimation using the sparse regression approach in PV systems.

REFERENCES

[1] M. Liu, L. Tan, and S. Cao, "Method of dynamic mode decomposition and reconstruction with application to a three-stage multiphase pump," *Energy*, vol. 208, p. 118343, 2020.

[2] H. Lu and D. M. Tartakovsky, "Prediction accuracy of dynamic mode decomposition," *SIAM Journal on Scientific Computing*, vol. 42, no. 3, pp. A1639–A1662, 2020.

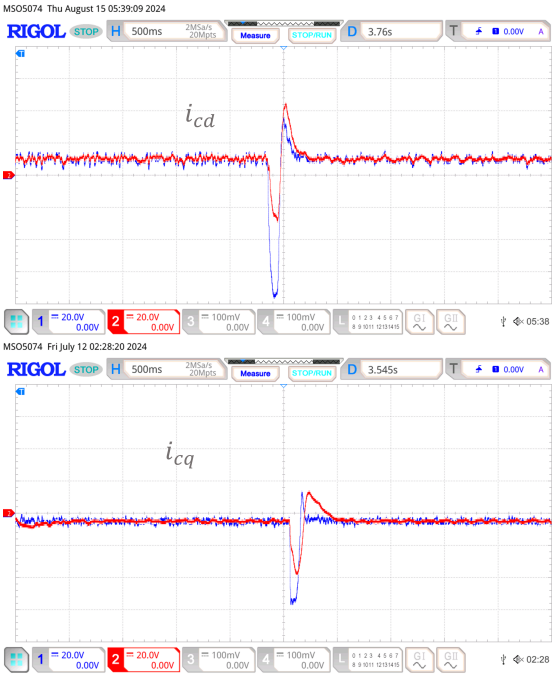

Fig. 15 Real-time verification of undervoltage fault.

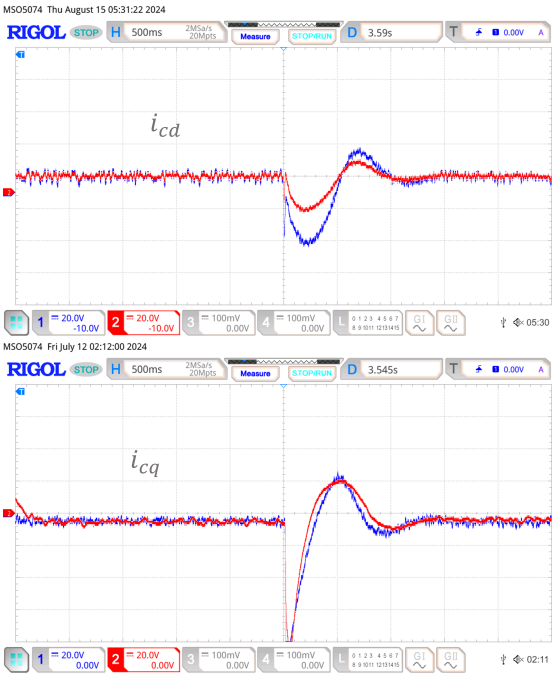

Fig. 16 Real-time verification of grounded fault.

[3] M. Al-Gabalawy, "Deep learning for koopman operator optimal control," *ISA transactions*, 2021.
[4] A. Mauroy, I. Mezić, and Y. Susuki, *The Koopman Operator in Systems and Control: Concepts, Methodologies, and Applications*. Springer Nature, 2020, vol. 484.
[5] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 113, no. 15, pp. 3932–3937, 2016.
[6] U. Fasel, E. Kaiser, J. N. Kutz, B. W. Brunton, and S. L. Brunton, "Sindy with control: A tutorial," *arXiv preprint arXiv:2108.13404*, 2021.
[7] G. Kandaperumal, K. P. Schneider, and A. K. Srivastava, "A data-driven algorithm for enabling delay tolerance in resilient microgrid controls using dynamic mode decomposition," *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 2500–2510, 2022.
[8] M. Netto and L. Mili, "A robust data-driven koopman kalman filter for power systems dynamic state estimation," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7228–7237, 2018.
[9] K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis *et al.*, "Data-driven prediction of battery cycle life before capacity degradation," *Nature Energy*, vol. 4, no. 5, pp. 383–391, 2019.
[10] Y. Li, Y. Liao, X. Wang, L. Nordström, P. Mittal, M. Chen, and H. V. Poor, "Neural network models and transfer learning for impedance modeling of grid-tied inverters," in *2022 IEEE 13th International Symposium on Power Electronics for Distributed Generation Systems (PEDG)*. IEEE, 2022, pp. 1–6.
[11] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the national academy of sciences*, vol. 113, no. 15, pp. 3932–3937, 2016.
[12] ——, "Sparse identification of nonlinear dynamics with control (sindyc)," *IFAC-PapersOnLine*, vol. 49, no. 18, pp. 710–715, 2016.
[13] A. Hosseinipour and J. Khazaei, "Sparse identification for data-driven dynamics and impedance modeling of power converters in dc microgrids," *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*, pp. 1–13, 2023.
[14] J. Khazaei and A. Hosseinipour, "Data-driven feedback linearization control of distributed energy resources using sparse regression," *IEEE Transactions on Smart Grid*, pp. 1–1, 2023.
[15] A. Nandakumar, Y. Li, H. Zheng, J. Zhao, D. Zhao, Y. Zhang, T. Hong, and B. Chen, "Data-driven modeling of microgrid transient dynamics through modularized sparse identification," *IEEE Transactions on Sustainable Energy*, pp. 1–14, 2023.
[16] J. Khazaei and A. Hosseinipour, *Advances in Data-Driven Modeling and Control of Naval Power Systems*. John Wiley & Sons, Ltd, 2022, ch. 21, pp. 453–473. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119812357.ch21
[17] A. Hamid, D. Rafiq, S. A. Nahvi, and M. A. Bazaz, "Power grid parameter estimation using sparse identification of nonlinear dynamics," in *2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP)*, 2022, pp. 1–6.
[18] V. N. Lal and S. N. Singh, "Control and performance analysis of a single-stage utility-scale grid-connected pv system," *IEEE Systems Journal*, vol. 11, no. 3, pp. 1601–1611, 2017.
[19] J. Khazaei, Z. Tu, and W. Liu, "Small-signal modeling and analysis of virtual inertia-based pv systems," *IEEE Transactions on Energy Conversion*, vol. 35, no. 2, pp. 1129–1138, 2020.
[20] A. Yazdani and R. Iravani, *Voltage-sourced converters in power systems: modeling, control, and applications*. John Wiley & Sons, 2010.
[21] T. S. Basso and R. DeBlasio, "Ieee 1547 series of standards: interconnection issues," *IEEE Transactions on Power Electronics*, vol. 19, no. 5, pp. 1159–1162, 2004.
[22] P. Monica and M. Kowsalya, "Control strategies of parallel operated inverters in renewable energy application: A review," *Renewable and Sustainable Energy Reviews*, vol. 65, pp. 885–901, 2016.
[23] X. Ge, A. Hosseinipour, S. Putri, F. Moazeni, and J. Khazaei, "Learning the dynamics of future marine microgrids using temporal convolutional neural network," in *2024 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS)*. IEEE, 2024, pp. 1–7.

Zhongtian Zhang (Student Member, IEEE) received his Master's degree in Electrical and Electronic Engineering from Lehigh University, Pennsylvania, USA, in 2021. Since then, he has been pursuing a Ph.D. at Lehigh University. His research primarily focuses on power systems, with an emphasis on data-driven modeling, nonlinear system identification, and power inverter control design.

Javad Khazaei (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of South Florida, Tampa, FL, USA, in 2016 with focus on power and energy systems. He is currently an Assistant Professor with Electrical and Computer Engineering, Lehigh University, Bethlehem, PA, USA. His research interests include data-driven and model-based control, optimization, and dynamic modeling of cyber-physical power systems and microgrids, smart grid security, and power electronics applications in smart grids and shipboard microgrids.

Rick S. Blum (Fellow, IEEE) received a B.S. in Electrical Engineering from the Pennsylvania State University in 1984 and his M.S. and Ph.D in Electrical Engineering from the University of Pennsylvania in 1987 and 1991.

From 1984 to 1991 he was a member of technical staff at General Electric Aerospace in Valley Forge, Pennsylvania and he graduated from GE's Advanced Course in Engineering. Since 1991, he has been with the Electrical and Computer Engineering Department at Lehigh University in Bethlehem, Pennsylvania where he is currently a Professor and holds the Robert W. Wieseman Endowed Professorship in Electrical Engineering. His research interests include machine learning and signal processing for security, smart grid, communications, sensor networking, radar and sensor processing. He was on the editorial board for the Journal of Advances in Information Fusion of the International Society of Information Fusion. He was an associate editor for IEEE Transactions on Signal Processing and for IEEE Communications Letters. He has edited special issues for IEEE Transactions on Signal Processing, IEEE Journal of Selected Topics in Signal Processing and IEEE Journal on Selected Areas in Communications. He was a member of the SAM Technical Committee (TC) of the IEEE Signal Processing Society. He was a member of the Signal Processing for Communications TC of the IEEE Signal Processing Society and is a member of the Communications Theory TC of the IEEE Communication Society. He was on the awards Committee of the IEEE Communication Society.

Dr. Blum is a Fellow of the IEEE and served two terms as an IEEE Signal Processing Society Distinguished Lecturer. He is an IEEE Third Millennium Medal winner, Eleanor and Joseph F. Libsch Research Award winner, a member of Eta Kappa Nu and Sigma Xi, and holds several patents. He was awarded an ONR Young Investigator Award in 1997 and an NSF Research Initiation Award in 1992. His IEEE Fellow Citation "for scientific contributions to detection, data fusion and signal processing with multiple sensors" acknowledges contributions to the field of sensor networking.