

## Research

# Research on detection and defense methods for false data injection attacks in power systems based on state-space decomposition

Chao Hong<sup>1,2,3</sup> · Zhihong Liang<sup>1,3</sup> · Yiwei Yang<sup>1,3</sup> · Pandeng Li<sup>1,3</sup> · Lin Chen<sup>1,3</sup> · Leyi Bi<sup>1,3</sup> · Yunan Zhang<sup>1,3</sup>

Received: 19 March 2025 / Accepted: 3 June 2025

Published online: 10 July 2025

© The Author(s) 2025 **OPEN****Abstract**

With increasing renewable energy integration, load frequency control (LFC) faces security risks from false data injection attacks (FDIAs). Existing detection methods struggle to distinguish control input attacks from measurement attacks, affecting system stability. This paper formulates a novel state-space decomposition strategy for decoupling control input and measurement attacks in the context of LFC systems with high renewable penetration; introduces a sliding mode observer (SMO) tailored for real-time attack estimation; and integrates an attack-resilient control (ARC) scheme based on  $H_\infty$  control theory to actively suppress attack impacts. Simulations show that the proposed method reduces AE mean squared error by nearly 30% and improves frequency response stability. These results confirm its effectiveness in detecting FDIAs and enhancing power system security.

**Highlights**

- The detection accuracy is improved by transforming the system into a new coordinate space through state-space decomposition, enabling independent identification of control input and measurement attacks;
- Attack signal estimation is enhanced using a sliding mode observer (SMO), which demonstrates higher robustness against modeling uncertainties and measurement noise compared to conventional observers such as Kalman-based filters;
- The attack-resilient control (ARC) strategy under the  $H_\infty$  control framework is proposed to maintain system frequency stability in the presence of disturbances and targeted FDIAs.

**Keywords** Load frequency control (LFC) · False data injection attack (FDIA) · State-space decomposition · Sliding mode observer (SMO) · Anti-attack control

---

✉ Chao Hong, hongchao2@csg.cn; Zhihong Liang, liangzh@csg.cn; Yiwei Yang, yangyw@csg.cn; Pandeng Li, lipd@csg.cn; Lin Chen, chenlin1@csg.cn; Leyi Bi, bily@csg.cn; Yunan Zhang, zhangyn2@csg.cn | <sup>1</sup>Electric Power Research Institute, China Southern Power Grid, Guangzhou, Guangdong, China. <sup>2</sup>School of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong, China. <sup>3</sup>Guangdong Provincial Key Laboratory of Power System Network Security, China Southern Power Grid, Guangzhou, Guangdong, China.



## 1 Introduction

In recent years, with the large-scale integration of renewable energy sources, the inertia level of power systems has gradually decreased. The stochastic nature of wind and photovoltaic power generation has intensified system frequency fluctuations and significantly increased the complexity of regulation, posing greater challenges for load frequency control (LFC). Moreover, the digitalization and intelligence of power systems have heightened the risk of cyberattacks, particularly false data injection attacks (FDIAs), which may compromise LFC measurement data and control signals, thereby affecting the stable operation of the system [1].

Modeling FDIAs in power systems with high renewable integration is of great significance [2]. In the model-driven field, [3] formulated AC-based FDIA construction as a sparsity-constrained nonlinear optimization and used semidefinite programming to reveal vulnerable state subspaces. [4] proposed a complementarity-based reformulation for PMU-only estimation to derive optimal stealthy attack sets. In the context of DLR systems, [5] developed a detection model combining feature extraction and machine learning classification; [6] introduced a reliability modeling framework incorporating redundancy and cyber-physical characteristics; and [7] designed a probabilistic DLR forecasting algorithm to enhance robustness under data uncertainty. [8] demonstrated the impact of PMU placement strategies on observability and FDIA resilience in flexible-rated systems.

In the probabilistic and state estimation domain, [9] proposed an interval estimation framework based on probabilistic forecasting and an enhanced Krawczyk operator, while [10] developed a square-root unscented Kalman filter scheme enhanced via dynamic time warping and generalized likelihood ratio testing. [11] applied canonical variate analysis to exploit correlation structures in AC estimations, and [12] proposed blind attack identification via matrix reconstruction and subspace learning. To improve stealth, [13] utilized CANDECOMP/PARAFAC tensor decomposition, and [14] further adopted robust tensor principal component analysis to generate undetectable FDIA vectors.

In deep learning, [15] introduced a self-supervised deep clustering technique combining autoencoders and trainable layers. Topology-invariant detection was addressed by [16], who fused variational graph autoencoders, TCNs, and GRUs. [17] presented a spatiotemporal CNN combining Kalman filters and Gaussian process regression. For localization, [18] developed a method using time-frequency domain features with spectral GCNs, while [19] adopted a cascade boosting forest to improve accuracy.

Graph-based and signal processing strategies were also explored. [20] introduced localization through spatiotemporal graph-wavelet CNNs. [21] integrated denoising diffusion models with graph learning under Kirchhoff constraints. [22] reconstructed clean measurement data via matrix encoding techniques, and [23] presented a polynomial-time secure estimator using a decomposed state-space representation.

At the communication layer, vulnerabilities were comprehensively analyzed. [24] examined ICT component failures in synchrophasor-based DTR and SIPS; [25] emphasized the impact of network availability on measurement reliability; and [26] surveyed CPS-layer reliability in electrical networks. Broader security contexts were addressed in [27], which analyzed DoS-induced frequency control degradation using energy-function stability models. [28] introduced topology-aware multichannel anomaly detection based on density estimation and signal fusion, and [29] developed a secure control framework founded on quantitative risk metrics.

Building on these contributions, this paper proposes an attack detection and estimation method based on state-space decomposition, integrated with an active robust control mechanism to enhance system stability under FDIA. The proposed work focuses on the modeling, detection, and mitigation of FDIAs in LFC systems incorporating high levels of renewable energy and energy storage, using observer- and control-based approaches grounded in state-space theory.

Compared to existing methods, the proposed approach offers the following key contributions: it enhances detection accuracy by transforming the system into a new coordinate space through state-space decomposition, thereby enabling independent identification of control input and measurement attacks. Furthermore, attack signal estimation is significantly improved through the use of a sliding mode observer (SMO), which provides increased robustness against modeling uncertainties and measurement noise compared to conventional observers such as Kalman-based filters. Finally, an attack-resilient control (ARC) strategy is developed under the  $H_\infty$  control framework to ensure system frequency stability, even in the presence of disturbances and targeted false data injection attacks (FDIAs).

The structure of this paper is as follows. Section 2 provides an overview of the LFC model in power systems with high renewable energy penetration and establishes the mathematical formulation of FDIAs. Section 3 presents an attack detection and estimation method based on state-space decomposition. By constructing a new coordinate

system, independent identification of control input attacks and measurement attacks is achieved, and a SMO is designed to enhance detection accuracy. Section 4 proposes an attack-resilient method based on  $H_\infty$  control, which employs robust control strategies to mitigate attack impacts and ensure stable system operation. Section 5 validates the effectiveness of the proposed methods through simulation experiments, analyzes the system's dynamic response under different attack scenarios, and evaluates the performance of the detection and estimation methods.

## 2 FDIA modeling in power systems with renewable energy penetration

### 2.1 LFC system model

LFC plays a crucial role in secondary frequency regulation of power systems. Its primary objective is to maintain system frequency stability and ensure that power exchange across interconnections remains within the planned or permissible deviation range. When load variations or fluctuations in renewable energy output occur, LFC restores system balance by adjusting generator output power. Failure to make timely adjustments may lead to significant frequency deviations, potentially causing generator disconnection or even large-scale blackouts.

In a typical third-order LFC model, each control area is represented by an equivalent generator that characterizes its frequency and power dynamics [30]. The system state is described by the following set of differential equations:

1. Frequency deviation equation

$$\frac{d\Delta f_i}{dt} = \frac{1}{2H_i}(\Delta P_{g_i} - D_i\Delta f_i - \Delta P_{tie_i} - \Delta P_{L_i} + \Delta P_{RES_i}) \quad (1)$$

2. Turbine equation

$$\frac{d\Delta P_{g_i}}{dt} = \frac{1}{T_{t_i}}(\Delta X_i - \Delta P_{g_i}) \quad (2)$$

3. Governor equation

$$\frac{d\Delta X_i}{dt} = \frac{1}{T_{g_i}}\left(\Delta P_{c_i} - \Delta X_i - \frac{\Delta f_i}{R_i}\right) \quad (3)$$

where Eq. (1) represents the frequency deviation equation, in which  $2H_i$  is the rotational inertia constant,  $H_i$  denotes the system inertia constant, and  $D_i$  represents the damping coefficient. The term  $\Delta P_{g_i}$  refers to the deviation in generator mechanical output,  $\Delta P_{L_i}$  denotes the load disturbance,  $\Delta P_{RES_i}$  represents the fluctuation in renewable energy output, and  $\Delta P_{tie_i}$  is the deviation in real power exchange across interconnections.

Equation (2) describes the turbine dynamics, where  $T_{t_i}$  is the turbine time constant, and  $\Delta X_i$  represents the deviation in the governor valve or guide vane opening.

Equation (3) defines the governor dynamics, where  $T_{g_i}$  is the governor time constant,  $\Delta P_{c_i}$  is the control power setpoint determined by the control center, and  $R_i$  represents the frequency droop coefficient.

In practical applications, the frequency droop coefficient  $R_i$  reflects the sensitivity of generator power output to frequency changes and contributes to primary frequency control. It is typically specified by grid operation codes or derived from generator parameters. In most utility-scale systems,  $R_i$  ranges from 2% to 5%, which corresponds to a droop value between 0.02 p.u. and 0.05 p.u. depending on generator configuration and regulatory frameworks.

By selecting the state variables as  $x_i = [\Delta f_i, \Delta P_{g_i}, \Delta X_i]^T$ , the control input as  $u_i = \Delta P_{c_i}$ , and the disturbance vector as  $d_i = [\Delta P_{L_i}, \Delta P_{RES_i}, \Delta P_{tie_i}]^T$ , the state-space equation can be expressed as:

$$\frac{dx_i}{dt} = A_i x_i + B_i u_i + E_i d_i \quad (4)$$

where

$$A_i = \begin{bmatrix} -\frac{D_i}{2H_i} & \frac{1}{2H_i} & 0 \\ 0 & -\frac{1}{T_{ti}} & \frac{1}{T_{ti}} \\ -\frac{1}{R_i T_{gi}} & 0 & -\frac{1}{T_{gi}} \end{bmatrix} \quad (5)$$

$$B_i = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{T_{gi}} \end{bmatrix} \quad (6)$$

$$E_i = \begin{bmatrix} -\frac{1}{2H_i} & \frac{1}{2H_i} & -\frac{1}{2H_i} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (7)$$

The measurement output equation is defined as:

$$y_i = C_i x_i \quad (8)$$

The matrix  $C_i$  can be set according to measurement requirements. For instance, choosing  $C_i = [1, 0, 0]$  allows measurement of the frequency deviation  $\Delta f_i$ , while selecting  $C_i = [0, 1, 0]$  enables measurement of the generator power output deviation  $\Delta P_{gi}$ .

## 2.2 Renewable energy models

### 1. Wind turbine model

The output power of a wind turbine depends on the wind speed  $V_W$ :

$$P_W = \frac{1}{2} \rho A C_p(\lambda, \beta) V_W^3 \quad (9)$$

where  $P_W$  represents the output power of the wind turbine,  $\rho$  denotes the air density,  $A$  is the swept area of the turbine, and  $C_p(\lambda, \beta)$  is the power coefficient, which theoretically reaches a maximum at the Betz limit of approximately 0.59 but is typically within the range of 0 to 0.5 in engineering applications. The parameter  $\lambda$  represents the tip speed ratio, while  $\beta$  denotes the pitch angle.

Due to the strong randomness of wind speed, short-term wind speed variations and the resulting power fluctuations can be modeled using methods such as Markov chains and autoregressive moving average (ARMA) models. The long-term statistical distribution of wind speed is commonly described by the Weibull distribution:

$$f(V_W) = \frac{k}{c} \left( \frac{V_W}{c} \right)^{k-1} e^{-(V_W/c)^k} \quad (10)$$

where  $k$  represents the shape parameter, and  $c$  denotes the scale parameter. This distribution effectively characterizes the statistical properties of wind speed over a given period and can be used to assess the range of wind power fluctuations.

However, it should be noted that the Weibull distribution, while widely adopted for wind speed modeling, primarily encapsulates long-term variability and may fail to accurately represent abrupt gusts or extreme wind events that can provoke significant transient disturbances in power systems. To remedy this, future research may incorporate high-resolution stochastic models or event-driven simulations, such as autoregressive conditional heteroskedasticity (ARCH) models or extreme value theory, to better capture these phenomena.

### 2. Photovoltaic model

The output power of a photovoltaic system is influenced by solar irradiance  $G_T$  and temperature  $T$ :

$$P_{PV} = \eta A G_T (1 - \gamma(T - T_{ref})) \quad (11)$$

where  $P_{PV}$  represents the output power of the PV system,  $\eta$  denotes the conversion efficiency of the PV modules,  $A$  is the total area of the PV panels or arrays,  $G_T$  represents the total irradiance,  $\gamma$  denotes the temperature coefficient,  $T$  is the ambient temperature, and  $T_{ref}$  represents the reference temperature.

The probability distribution of solar irradiance can be described using the Beta distribution:

$$f(G_T) = \frac{G_T^{\alpha-1} (1 - G_T)^{\beta-1}}{B(\alpha, \beta)} \quad (12)$$

where  $\alpha$  and  $\beta$  represent the shape parameters, and  $B(\alpha, \beta)$  denotes the Beta function.

In addition, rapid cloud coverage can introduce high-frequency fluctuations, which can be analyzed using wavelet transforms or Fourier transforms for more refined modeling.

Likewise, the Beta distribution is limited in its ability to account for short-term irradiance anomalies, such as those caused by sudden cloud movements. To address high-frequency fluctuation modeling, this paper has included a brief mention of wavelet and Fourier transform techniques. Nonetheless, future iterations of this work may further investigate hybrid statistical-physical models or real-time data assimilation from remote sensing systems to improve resilience under anomalous solar conditions.

### 3. Energy storage model

The energy storage system can be approximated as a first-order inertial system:

$$\frac{dP_{ESS}}{dt} = -\frac{1}{T_{ESS}} (P_{ESS} - P_{ref}) \quad (13)$$

where  $P_{ESS}$  represents the actual power output of the energy storage system,  $T_{ESS}$  denotes the time constant of the energy storage system, and  $P_{ref}$  is the reference power.

By integrating the energy storage system into the LFC framework, the primary and secondary frequency regulation capabilities of the system can be enhanced in high-renewable-penetration scenarios.

If the power system incorporating renewable energy and energy storage is expressed in state-space form:

$$\frac{d(x)}{dt} = Ax + Bu + Ed \quad (14)$$

some elements of the matrix  $A$  will change due to reduced inertia and variations in damping. The system eigenvalues are determined by:

$$\det(A - \lambda I) = 0 \quad (15)$$

As the proportion of renewable energy increases, system inertia decreases, causing the eigenvalues to shift closer to the imaginary axis. This results in more sensitive frequency responses and larger oscillation amplitudes. Proper energy storage and control strategies can expand the system's stability margin and mitigate these adverse effects.

### 2.3 FDIA model

In power systems, state estimation is typically represented by the following measurement equation:

$$z = Hx + v \quad (16)$$

where  $z$  represents the measurement vector,  $x$  denotes the system state vector,  $H$  is the measurement matrix, and  $v$  represents the measurement noise. If an attacker constructs an attack vector  $a$  such that the compromised measurement

$$z' = z + a \quad (17)$$

still satisfies

$$z' = H(x + c) + v = Hx' + v \quad (18)$$

then traditional detection methods based on measurement residuals may fail to identify such attacks.

LFC relies on remote measurement data for control calculations, making FDIAs particularly detrimental to control systems. The primary impacts of FDIA on LFC include the following aspects:

#### (1) Measurement tampering

LFC primarily depends on the measured system frequency deviation  $\Delta f_i$  and the inter-area power exchange  $\Delta P_{tie_i}$  to compute the control command  $\Delta P_{c_i}$ . If an attacker alters these measurements in the communication link, the control center will compute load adjustment commands based on incorrect system states. This may cause generator outputs to deviate from actual demand, thereby affecting system frequency stability.

#### (2) Control command tampering

Another possible attack involves directly modifying the control command  $\Delta P_{c_i}$ , introducing a deviation  $a_c$  in the control signal received by the governor. If this deviation is sufficiently large or aligns with the system's dynamic characteristics, it may lead to sustained frequency oscillations or even exceed safe operational limits, endangering system stability.

#### (3) Inter-area power exchange misguidance

In multi-area systems, an attacker may inject false data into tie-line power measurements or control signals, leading to erroneous "power exchange" commands between areas. This can disrupt load distribution. In High Voltage Direct Current (HVDC) transmission systems, tampering with DC power exchange data can induce frequency fluctuations across different regions. Such attacks can rapidly alter the power balance in certain areas, causing abrupt frequency deviations and potentially destabilizing the system.

FDIA can be regarded as an additional disturbance or control deviation that alters the system's state equation. For LFC, the modified state equation can be expressed as

$$\frac{dx'}{dt} = Ax' + B[u + a_c] + Ed \quad (19)$$

where  $A$  represents the system state matrix,  $B$  denotes the input matrix,  $u$  represents the normal control input,  $Ed$  accounts for external disturbances, and the attack vector  $a_c$  acts as an additional disturbance to the system. If the attack's impact is equivalently mapped to changes in system parameters, a correction matrix  $A_a$  can be introduced, modifying the system characteristic equation as

$$\det(A + A_a - \lambda I) = 0. \quad (20)$$

This equation determines system stability. If the attack causes certain eigenvalues to enter the right half-plane, meaning their real parts become positive, the system will exhibit instability.

FDIA can stealthily manipulate system measurements and control data, making it difficult to detect under traditional monitoring mechanisms. This may lead to LFC failure, disrupt inter-area power balance, and even compromise overall system stability. Therefore, establishing effective detection and defense mechanisms is crucial for ensuring the secure operation of power systems.

## 3 Attack detection and estimation method based on state-space decomposition

### 3.1 State-space decomposition for network attack detection

In LFC systems, the state variables typically include frequency deviation, tie-line power deviation, and the control states of governors or generating units. When the system is subjected to a FDIA, traditional state variables cannot directly distinguish between the effects of "normal system dynamics" and "attack signals." Therefore, an appropriate transformation of the state space is required to construct a new coordinate system, enabling the separate characterization of "control input attacks" and "measurement attacks." This representation facilitates the design of observers for effective attack signal detection and estimation.

Let the system state vector be  $x(t)$ , the control input be  $u(t)$ , the measurement output be  $y(t)$ , and the external disturbance be  $d(t)$ . When the system is under attack, the attack signals can be categorized into those acting on the control

input, denoted as  $a_c(t)$ , and those affecting the measurement output, denoted as  $a_m(t)$ . Under these conditions, the system state equation can be expressed as

$$\frac{d}{dt}x(t) = Ax(t) + Bu(t) + Ed(t) + F\phi(x(t), t) + Ba_c(t), \quad (21)$$

where  $A$  is the system state matrix,  $B$  is the input matrix,  $E$  is the external disturbance coupling matrix, and  $F\phi(x(t), t)$  represents possible nonlinear dynamics in the system. The attack signal  $a_c(t)$  directly affects the input.

The measurement output equation is given by

$$y(t) = Cx(t) + Da_m(t), \quad (22)$$

where  $C$  is the measurement matrix, and  $D$  characterizes the coupling of the measurement attack  $a_m(t)$  in the output observations.

To achieve independent identification of “control input attacks” and “measurement attacks,” assume the existence of a state transformation matrix  $T$  and an output transformation matrix  $S$ . Define new state variables and measurement outputs as

$$\zeta(t) = Tx(t), \quad (23)$$

$$\omega(t) = Sy(t). \quad (24)$$

Transforming the original system equations into the new coordinate system yields

$$\frac{d}{dt}\zeta(t) = TAT^{-1}\zeta(t) + TBU(t) + TBa_c(t) + TE d(t) + TF\phi(T^{-1}\zeta(t), t), \quad (25)$$

$$\omega(t) = SCT^{-1}\zeta(t) + SDa_m(t). \quad (26)$$

By appropriately selecting  $T$  and  $S$ , the system matrix  $TAT^{-1}$  and the output matrix  $SCT^{-1}$  can be transformed into a specific block structure, enabling the separation of attack signals:

$$TAT^{-1} = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}, \quad (27)$$

$$TB = \begin{bmatrix} B_1 \\ 0 \end{bmatrix}, \quad (28)$$

$$SCT^{-1} = \begin{bmatrix} C_1 & 0 \\ 0 & C_4 \end{bmatrix}, \quad (29)$$

where  $A_1, A_2, A_3, A_4, B_1, C_1$ , and  $C_4$  are submatrices of appropriate dimensions. If  $TB = \begin{bmatrix} B_1 \\ 0 \end{bmatrix}$ , then the control input attack  $a_c(t)$  affects only the upper subsystem, while the measurement attack  $a_m(t)$  influences only the lower subsystem. Consequently, the system can be decomposed into two independent submodules.

**Module I: Control input attack detection**

The state equations for Module I, which is primarily affected by the control input attack  $a_c(t)$ , are given by

$$\frac{d}{dt}\zeta_1(t) = A_1\zeta_1(t) + A_2\zeta_2(t) + B_1u(t) + a_c(t) + E_1d(t) + F_1\phi(T^{-1}\zeta(t), t), \quad (30)$$

$$\omega_1(t) = C_1\zeta_1(t). \quad (31)$$

The measurement output  $\omega_1(t)$  is associated with  $\zeta_1(t)$  and is used to design an observer for identifying the impact of  $a_c(t)$ .

**Module II: Measurement attack detection**

The state equations for Module II, which is primarily affected by the measurement attack  $a_m(t)$ , are given by

$$\frac{d}{dt}\zeta_2(t) = A_3\zeta_1(t) + A_4\zeta_2(t) + E_2d(t) + F_2\phi(T^{-1}\zeta(t), t), \quad (32)$$

$$\omega_2(t) = C_4\zeta_2(t) + D_2a_m(t). \quad (33)$$

where  $D_2$  characterizes the coupling of the attack signal  $a_m(t)$  in the measurement output.

By decomposing the state space, independent identification of “control input attacks” ( $a_c$ ) and “measurement attacks” ( $a_m$ ) can be achieved in the transformed coordinate system. This method significantly simplifies the structure of observers or filters, enhancing the performance and robustness of attack detection.

#### Algorithm 1 FDIA detection: state transformation and decomposition

---

**Require:** System matrices  $A, B, C, D, E, F$ ; Nonlinear term  $\phi(x(t), t)$ ; transformation matrices  $T, S$

**Require:** Disturbance  $d(t)$ , control input  $u(t)$ , output measurements  $y(t)$

**Ensure:** Transformed state  $\zeta(t)$ , transformed output  $\omega(t)$

- 1: Compute transformed variables:
- 2:  $\zeta(t) = Tx(t)$
- 3:  $\omega(t) = Sy(t)$
- 4: Derive transformed system:
- 5:  $\frac{d}{dt}\zeta(t) = TAT^{-1}\zeta(t) + TBU(t) + TBA_c(t) + TE d(t) + TF\phi(T^{-1}\zeta(t), t)$
- 6:  $\omega(t) = SCT^{-1}\zeta(t) + SDa_m(t)$
- 7: Partition matrices:
- 8:  $TAT^{-1} = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}, TB = \begin{bmatrix} B_1 \\ 0 \end{bmatrix}, SCT^{-1} = \begin{bmatrix} C_1 & 0 \\ 0 & C_4 \end{bmatrix}$
- 9: Decompose state/output:
- 10:  $\zeta(t) = \begin{bmatrix} \zeta_1(t) \\ \zeta_2(t) \end{bmatrix}, \omega(t) = \begin{bmatrix} \omega_1(t) \\ \omega_2(t) \end{bmatrix}$
- 11: Define subsystem dynamics:
- 12: Module I (control attack):
- 13:  $\frac{d}{dt}\zeta_1(t) = A_1\zeta_1(t) + A_2\zeta_2(t) + B_1u(t) + a_c(t) + E_1d(t) + F_1\phi(T^{-1}\zeta(t), t)$
- 14:  $\omega_1(t) = C_1\zeta_1(t)$
- 15: Module II (measurement attack):
- 16:  $\frac{d}{dt}\zeta_2(t) = A_3\zeta_1(t) + A_4\zeta_2(t) + E_2d(t) + F_2\phi(T^{-1}\zeta(t), t)$
- 17:  $\omega_2(t) = C_4\zeta_2(t) + D_2a_m(t)$

---

### 3.2 Design and optimization of a SMO for FDIA

For the state equation of Module I,  $\zeta_2(t)$  can be considered a known input or obtained through other estimation methods. A SMO is designed for  $\zeta_1(t)$  as follows [31]:

$$\frac{d\hat{\zeta}_1(t)}{dt} = A_1\hat{\zeta}_1(t) + A_2\bar{\zeta}_2(t) + B_1u(t) + K_1[\omega_1(t) - C_1\hat{\zeta}_1(t)] + v_1(t) \quad (34)$$

where  $K_1$  represents the conventional gain matrix,  $v_1(t)$  denotes the sliding mode compensation signal, and  $\bar{\zeta}_2(t)$  is the known or estimated value of  $\zeta_2(t)$ . The observation error is defined as

$$e_1(t) = \zeta_1(t) - \hat{\zeta}_1(t). \quad (35)$$

The corresponding error dynamics can be expressed as

$$\frac{de_1(t)}{dt} = [A_1 - K_1C_1]e_1(t) + ac(t) + \rho_1(t) - v_1(t) \quad (36)$$

where  $\rho_1(t)$  accounts for the effects of external disturbances  $E_1 d(t)$  and nonlinearities  $F_1 \phi(\cdot)$ . To ensure that  $e_1(t)$  converges to a small neighborhood within a finite time, the sliding surface is defined as

$$s_1(t) = C_1 e_1(t). \quad (37)$$

The sliding mode compensation strategy is chosen as

$$v_1(t) = -K_s \text{sign}(s_1(t)) \quad (38)$$

where  $K_s$  is appropriately tuned to suppress the influence of  $a_c(t)$  on the error. When the sliding surface  $s_1(t)$  converges,  $v_1(t)$  can be used to estimate  $a_c(t)$ :

$$\hat{a}_c(t) \approx -\frac{1}{K_s} v_1(t). \quad (39)$$

For the state equation of Module II, the measurement attack  $a_m(t)$  affects only the output and can be estimated using an unknown input observer or a SMO. The SMO is designed as follows:

$$\frac{d\hat{\zeta}_2(t)}{dt} = A_4 \hat{\zeta}_2(t) + A_3 \bar{e}_1(t) + L_2 [\omega_2(t) - C_4 \hat{\zeta}_2(t)] + v_2(t). \quad (40)$$

The observation error is defined as

$$e_2(t) = \zeta_2(t) - \hat{\zeta}_2(t). \quad (41)$$

The corresponding error dynamics are given by

$$\frac{de_2(t)}{dt} = [A_4 - L_2 C_4] e_2(t) + D_2 a_m(t) + \rho_2(t) - v_2(t) \quad (42)$$

where  $\rho_2(t)$  reflects the effects of external disturbances  $E_2 d(t)$  and nonlinearities  $F_2 \phi(\cdot)$ . The sliding surface is selected as

$$s_2(t) = C_4 e_2(t) \quad (43)$$

and the sliding mode compensation strategy is applied as

$$v_2(t) = -K_m \text{sign}(s_2(t)) \quad (44)$$

where  $K_m$  represents the sliding mode gain parameter. In sliding-mode observer design, the compensation term  $v_2(t)$  is introduced to suppress the unknown disturbance induced by the measurement attack  $a_m(t)$ . The negative sign preceding the gain  $K_m$  ensures that the discontinuous control term  $v_2(t)$  is directed opposite to the measurement sliding surface  $s_2(t)$ , defined as  $C_4 e_2(t)$ . This alignment with the direction of the reaching condition promotes convergence of the error dynamics to the sliding manifold and stabilizes the estimation error. The essential role of the negative sign is, therefore, to enforce the corrective action required for robustness against  $a_m(t)$ , consistent with standard sliding-mode formulations. If the matrix  $[A_4 - L_2 C_4]$  is a Hurwitz matrix and  $K_m$  is appropriately chosen,  $e_2(t)$  can rapidly converge under the sliding mode effect, enabling the estimation of  $a_m(t)$ :

$$\hat{a}_m(t) \approx -\frac{1}{K_m} v_2(t). \quad (45)$$

The parameters  $K_1$ ,  $L_2$ ,  $K_s$ , and  $K_m$  can be optimized using linear matrix inequalities or adaptive tuning strategies to enhance the observer's convergence speed and system robustness.

The observer gain matrices  $K_1$  and  $L_2$  are first designed through pole assignment techniques such that the matrices  $(A_1 - K_1 C_1)$  and  $(A_4 - L_2 C_4)$  become Hurwitz, satisfying necessary stability conditions for exponential convergence of observer error dynamics. In practice, the desired poles are selected to provide a trade-off between fast convergence and noise sensitivity, typically placing them at  $-5$  to  $-10$  in the complex plane. After linear stability is established, the sliding mode gains  $K_s$  and  $K_m$  are selected to drive the error trajectories to the sliding manifold in finite time. These gains are tuned based on Lyapunov stability arguments and bounded disturbance rejection. Specifically, the gains are chosen to satisfy

$$V_5(t)e(t) \geq \eta \|e(t)\| \quad (46)$$

where  $\eta > 0$  ensures finite-time convergence despite nonlinear and attack-induced perturbations. In our implementation, initial values of  $K_1$  and  $L_2$  are obtained via pole placement, and then iterative simulations are used to tune  $K_s$  and  $K_m$  adaptively based on trial conditions and robustness margins.

#### Algorithm 2 FDIA detection: sliding mode observer design

---

**Require:** Input  $u(t)$ , output  $\omega_1(t)$ , estimated  $\bar{\zeta}_2(t)$ , initial state  $\hat{\zeta}_1(0)$

**Ensure:** Estimate of  $a_c(t)$

1: **Module I – Estimating Control Input Attack**

2: Observer:

3:  $\frac{d}{dt}\hat{\zeta}_1(t) = A_1\hat{\zeta}_1(t) + A_2\bar{\zeta}_2(t) + B_1u(t) + K_1[\omega_1(t) - C_1\hat{\zeta}_1(t)] + v_1(t)$

4: Estimation error:  $e_1(t) = \zeta_1(t) - \hat{\zeta}_1(t)$

5: Sliding surface:  $s_1(t) = C_1e_1(t)$

6: Sliding mode:  $v_1(t) = -K_s \text{sign}(s_1(t))$

7: Estimate:  $\hat{a}_c(t) \approx -\frac{1}{K_s}v_1(t)$

**Require:** Output  $\omega_2(t)$ , estimated  $\bar{\zeta}_1(t)$ , initial state  $\hat{\zeta}_2(0)$

**Ensure:** Estimate of  $a_m(t)$

8: **Module II – Estimating Measurement Attack**

9: Observer:

10:  $\frac{d}{dt}\hat{\zeta}_2(t) = A_4\hat{\zeta}_2(t) + A_3\bar{\zeta}_1(t) + L_2[\omega_2(t) - C_4\hat{\zeta}_2(t)] + v_2(t)$

11: Estimation error:  $e_2(t) = \zeta_2(t) - \hat{\zeta}_2(t)$

12: Sliding surface:  $s_2(t) = C_4e_2(t)$

13: Sliding mode:  $v_2(t) = -K_m \text{sign}(s_2(t))$

14: Estimate:  $\hat{a}_m(t) \approx -\frac{1}{K_m}v_2(t)$

---

## 4 Attack detection and estimation method based on state-space decomposition

In the previous discussion, independent identification of “control input attacks” and “measurement attacks” was achieved through state-space decomposition. A SMO was designed to detect and estimate attack signals. However, merely detecting attack signals using an observer cannot completely eliminate the interference caused by cybersecurity threats to system operation. Therefore, building upon this foundation, this paper further designs an ARC mechanism based on  $H_\infty$  control theory to achieve a robust control strategy that adapts to FDIA environments, ensuring normal system operation under attack conditions [32].

In  $H_\infty$  control theory, the primary objective is to design a control law that optimizes system performance under the worst-case scenario. This control framework provides a systematic approach to handling malicious attacks and system uncertainties. Consider the following state-space system under attack:

$$\frac{d}{dt}x(t) = Ax(t) + Bu(t) + F_a a(t) \quad (47)$$

$$z(t) = Qx(t) + Ru(t) \quad (48)$$

where  $x(t)$  represents the state vector,  $u(t)$  denotes the control input, and  $a(t)$  is the attack signal. The performance output  $z(t)$  reflects the impact of the attack on the system and is used to evaluate the system’s attack resilience.  $Q$  and  $R$  are weighting matrices. The objective is to design an appropriate control law  $u(t)$  such that the performance index  $J$  is optimized. The performance index is defined as

$$J = \int_0^{\infty} (z^T z + \gamma^2 a^T a) dt \quad (49)$$

where  $\gamma$  represents the upper bound of the attack signal energy's impact on the system. To ensure that the controller suppresses the attack impact under worst-case conditions, the closed-loop system must satisfy the  $H_{\infty}$  performance criterion:

$$\|z\|_2^2 < \gamma^2 \|a\|_2^2 \quad (50)$$

where  $\|\cdot\|_2$  denotes the  $L_2$ -norm of a vector signal. When this constraint is satisfied, the system can limit performance degradation caused by attack signals.

To achieve this objective, the ARC based on  $H_{\infty}$  control requires solving the corresponding Riccati equation. Define the system performance matrix  $P$ , which satisfies the closed-loop stability condition in the following algebraic Riccati equation:

$$A^T P + PA + Q - PBR^{-1}B^T P + \gamma^{-2}PF_a F_a^T P = 0 \quad (51)$$

where  $P$  is a positive definite matrix. If this equation has a positive definite solution, the optimal control law can be constructed as

$$u(t) = -R^{-1}B^T Px(t). \quad (52)$$

This control law suppresses state deviations through state feedback, effectively compensating for performance degradation and instability caused by the attack signal  $a(t)$ . Consequently, the system achieves two core objectives: (1) The performance output  $z(t)$  remains within the prescribed performance bounds even under attack conditions, ensuring robustness against attack disturbances. (2) The system's tolerance to attack signals is enhanced, and gain adjustment allows flexible adaptation to different attack intensities.

It should be noted that while the proposed control framework demonstrates significant improvements under model-based simulation assumptions, it operates under the assumption of linear system dynamics. This approximation may limit detection sensitivity and control optimality in environments characterized by strong nonlinear coupling or discrete-time hybrid dynamics. Moreover, in designing the sliding mode observer, the presence of discontinuities in the estimation mechanism may introduce chattering, especially under high-frequency measurement noise, thereby necessitating compromise between estimation responsiveness and noise tolerance during gain tuning. On the control side, although the  $H_{\infty}$  methodology offers strong worst-case guarantees, its implementation depends on the successful solution of an associated Riccati equation, which may not be computationally tractable in distributed or geographically dispersed networks with evolving topologies. These challenges highlight the importance of future work in robust nonlinear observer design and distributed control synthesis for real-world deployment.

**Algorithm 3** Attack-resilient control via  $H_\infty$  strategy

---

**Require:** System matrices  $A, B, Q, R, F_a$ , Attack inputs  $a(t)$   
**Ensure:** Controller output  $u(t)$

- 1: Define system under attack:
- 2:  $\frac{d}{dt}x(t) = Ax(t) + Bu(t) + F_a a(t)$
- 3:  $z(t) = Qx(t) + Ru(t)$
- 4: Performance objective:
- 5:  $J = \int_0^\infty (z^T z + \gamma^2 a^T a) dt$
- 6:  $H_\infty$  criterion:
- 7:  $\|z\|_2^2 < \gamma^2 \|a\|_2^2$
- 8: Solve Riccati equation:
- 9:  $A^T P + PA + Q - PBR^{-1}B^T P + \gamma^{-2}PF_a F_a^T P = 0$
- 10: **if** Positive definite solution  $P$  exists **then**
- 11:     Compute control law:
- 12:      $u(t) = -R^{-1}B^T P x(t)$
- 13: **end if**
- 14: **Real-Time Control Loop**
- 15: **while** system is running **do**
- 16:     Receive measurements  $y(t)$
- 17:     Compute transformed output:  $\omega(t) = Sy(t)$
- 18:     Update observers with  $\omega_1(t), \omega_2(t)$
- 19:     Estimate  $\hat{a}_c(t)$  and  $\hat{a}_m(t)$
- 20:     Compute  $u(t)$  using  $H_\infty$  controller
- 21:     Apply  $u(t)$  to the physical system
- 22: **end while**

---

## 5 Case analysis and experimental validation

### 5.1 Attack scenario configuration

Simulation tests were conducted to evaluate the effectiveness and applicability of the proposed FDIA detection method. The experimental environment was built using Simulink, as illustrated in Fig. 1. Three scenarios were constructed based on two different system topologies. In the figure, blue solid lines represent conventional AC tie lines, orange solid lines denote AC tie lines with controllable phase-shifting transformers, and green solid lines indicate high-voltage DC tie lines. All simulations and numerical experiments presented in this work were conducted using MATLAB/Simulink R2022b on a personal computer equipped with an Intel Core i7-11800H processor (2.30 GHz, 8 cores), 32 GB RAM, and running Windows 11 Pro 64-bit operating system.

Scenario 1 employs a two-region structure, where a single conventional AC tie line connects the two regions. At  $t = 5$  s, a 1% p.u. step load disturbance was applied to Region 1, considering power fluctuations caused by renewable energy sources. Due to the single tie line, the system exhibits low disturbance redundancy, making it relatively vulnerable to load disturbances and FDIA.

**Fig. 1** Multiregion power system topology

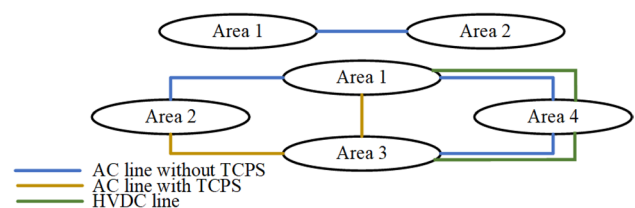


**Fig. 2** System dynamic response characteristics, **a** system frequency response trajectory; **b** system key characteristics

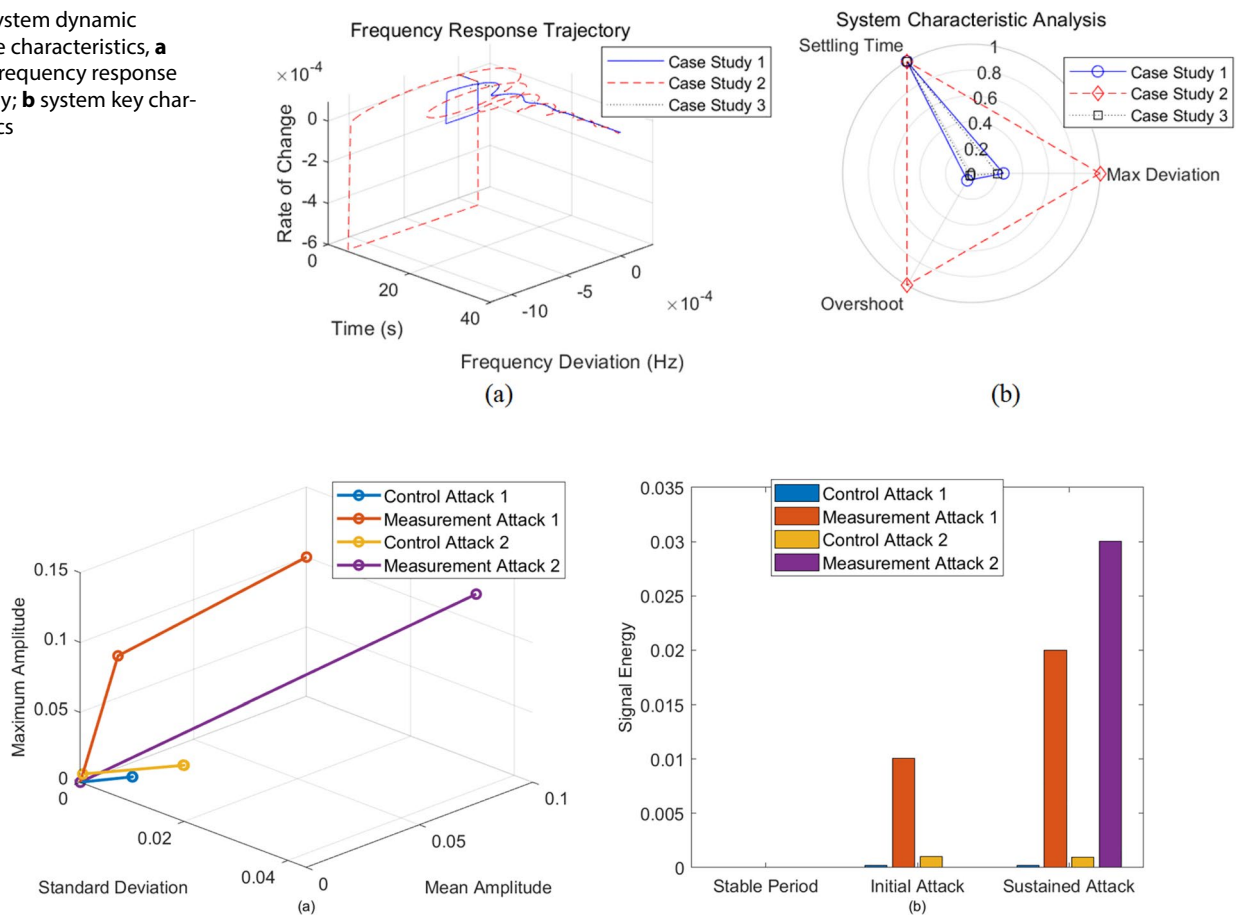


**Fig. 3** Attack vector characteristics and energy distribution, **a** three-dimensional distribution of attack vector characteristics; **b** energy distribution of attack vector in each time window

In Scenarios 2 and 3, the system was expanded to a four-region structure, interconnected through multiple tie lines of different types, representing the characteristics of a complex power system. In Scenario 2, a 5% p.u. step load disturbance was applied to Region 1 at  $t = 5$  s, while also considering the impact of renewable energy variations. In Scenario 3, only the influence of renewable energy variations was considered, without additional load disturbances.

The dynamic response characteristics of the three scenarios are shown in Fig. 2. Subfigure a presents the frequency response trajectories, illustrating the relationship between frequency deviation, rate of change, and time. The oscillation amplitude in Scenario 2 is the largest, indicating that the combined effect of the step load disturbance and renewable energy fluctuations results in the most significant transient response. In contrast, Scenario 1 exhibits the smallest oscillation amplitude, with minor deviations around the origin, suggesting that the system's frequency response remains relatively stable when only renewable energy fluctuations are present.

Subfigure b characterizes the system features in terms of maximum deviation, settling time, and overshoot. Scenario 2 exhibits the highest values for maximum deviation and overshoot, indicating that the 5% p.u. step load disturbance has a significant impact on system stability. Scenario 3 has the lowest values, demonstrating that renewable energy variations alone introduce relatively minor disturbances to the overall system.

## 5.2 Performance analysis of attack detection methods

To evaluate the effectiveness of the proposed FDIA detection and estimation method, four types of attack vectors were constructed. The feature space of these attack vectors is illustrated in Fig. 3a, while their energy distribution is shown in Fig. 3b.

(1) Control Attack 1: This attack vector targets the control signal. It has a relatively small standard deviation, a moderate average amplitude, and a high maximum amplitude. In the initial phase, it exhibits a periodic sinusoidal pattern, but

at specific moments, it generates large instantaneous peaks. (2) Measurement Attack 1: This attack vector targets the measurement signal. Its distribution is relatively concentrated, indicating that after the attack is triggered, the amplitude fluctuates around a steady-state level. (3) Control Attack 2: This attack vector also targets the control signal. It has a large standard deviation and maximum amplitude, exhibiting strong random pulse or noise characteristics. (4) Measurement Attack 2: This attack vector targets the measurement signal and exhibits a linearly accumulating attack trend, with its impact increasing over time.

All four types of attacks can be divided into three phases. During the steady-state phase (0–20 s), the energy of all attack signals remains close to zero, indicating that the system is in a normal state. In the initial attack phase (20–40 s), the energy of Control Attack 1 and Control Attack 2 increases significantly, whereas the energy of Measurement Attack 1 and Measurement Attack 2 grows at a relatively slower rate. During the sustained attack phase (40–60 s), the energy of Measurement Attack 1 rises rapidly and stabilizes, while the energy of Measurement Attack 2 increases linearly. Meanwhile, the energy of Control Attack 2 fluctuates rapidly, leading to unstable dynamic variations.

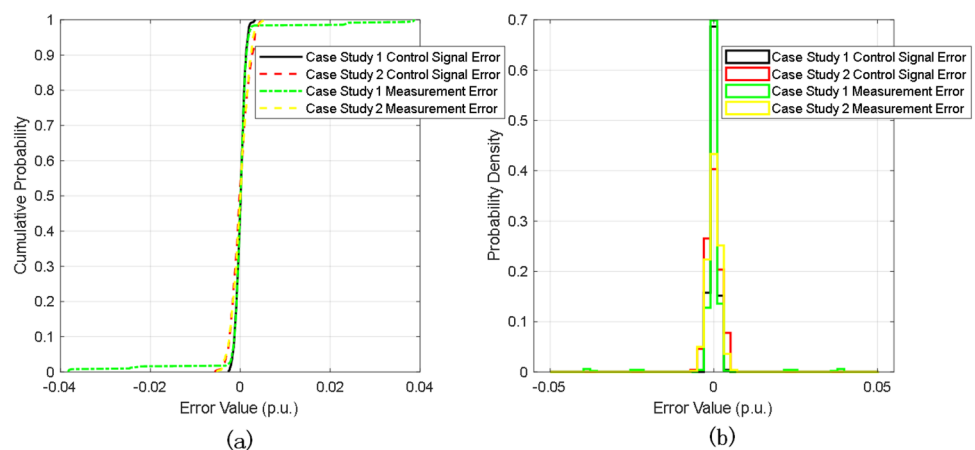
For these four attack vectors, detection and estimation were performed. To quantify the accuracy, the statistical characteristics of estimation errors were analyzed using the cumulative distribution function and probability density function. The estimation bias under control and measurement attack signals in Scenario 1 and Scenario 2 is shown in Fig. 4, where (a) represents the cumulative distribution of attack estimation (AE) errors, and (b) illustrates the probability density distribution of AE errors.

The cumulative distribution results indicate that all error curves exhibit a steep increase near zero, suggesting that estimation errors are primarily concentrated within a small range. Compared to the traditional Kalman filtering method, the proposed estimation method reduces the mean squared error by approximately 30%, demonstrating its high estimation accuracy across various scenarios. Specifically, in Scenario 1, the cumulative probability of control attack signal errors within  $\pm 0.01$  p.u. reaches 95%, while in Scenario 2, the cumulative probability within  $\pm 0.015$  p.u. exceeds 90%, further validating the robustness of the proposed method in complex environments.

The probability density analysis of estimation errors reveals key characteristics of the error distribution. The overall error distribution approximates a normal distribution, with a high concentration near the peak. For control attack signals, the error distribution is narrow and concentrated, with a peak significantly higher than that of measurement attack signals. This characteristic indicates smaller fluctuations in estimation errors and higher accuracy. In contrast, the distribution curve of measurement attack signals is relatively wider. Notably, in Scenario 1, a sudden deviation occurs around  $t = 35$  s, resulting in a long-tail characteristic in the measurement error distribution. Based on the above analysis, it can be concluded that the proposed method ensures that the estimation mean remains close to zero, with a concentrated error distribution and a small standard deviation, thereby achieving high-precision consistent estimation.

It is noted that several control parameters directly affect the attack estimation performance and dynamic response. For the SMO, the selection of  $K_s$  and  $K_m$  governs the intensity of the sliding mode correction term: large values enhance convergence speed and suppress disturbance components, but may introduce higher-frequency fluctuations in the estimation signal; smaller values yield smoother estimates but reduce sensitivity to fast-varying attacks. Similarly, the matrix gains  $K_1$  and  $L_2$  influence the error dynamics and observer stability margin—choices ensuring sufficient eigenvalue damping (e.g., via pole placement) improve robustness under varying noise and attack levels. In the  $H_\infty$  control structure, lower  $\gamma$  values enhance disturbance rejection and produce more aggressive control actions, which, while

**Fig. 4** Analysis of statistical characteristics of AE errors, (a) cumulative distribution of AE errors; (b) probability density distribution of AE errors



reducing frequency deviation under sustained attack conditions, may elevate control signal energy and response overshoot. Conversely, larger  $\gamma$  values relax the control penalty, resulting in slower dynamic response but reduced energy consumption of control execution. The weighting matrices  $\mathbf{Q}$  and  $\mathbf{R}$  likewise balance performance tracking and control effort in frequency regulation under adversarial conditions.

To examine the boundary at which system resilience deteriorates, additional observations were made under varying attack magnitudes without altering the scenario configurations. It is observed that when the effective amplitude of control attacks exceeds approximately 0.05 p.u., the estimation error variance increases significantly, resulting in a discernible rise in frequency deviation beyond 0.15 Hz in Scenario 2. These deviations exceed typical acceptable bounds for primary frequency control, indicating that the system begins to lose robustness under higher-intensity attack inputs.

To analyze the impact of the ARC mechanism on system stability, the power spectral density under three conditions—"ARC enabled," "ARC disabled (control attack)," and "ARC disabled (measurement attack)"—was examined for Scenarios 1, 2, and 3, as shown in Fig. 5.

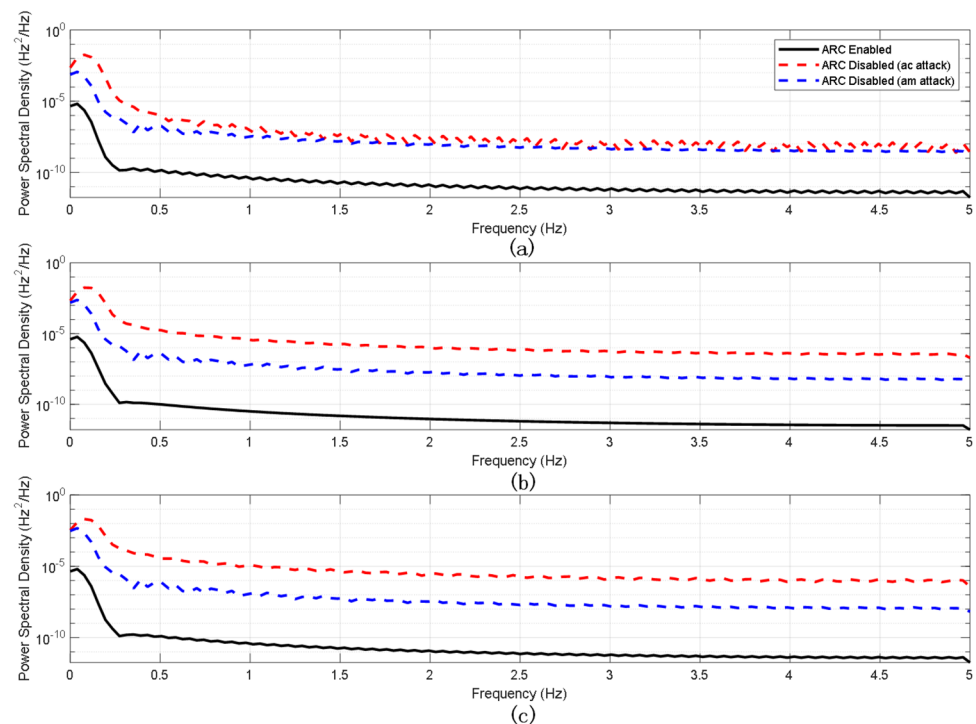
The system with ARC enabled exhibits favorable spectral characteristics. The energy is primarily concentrated in the 0.1–0.3 Hz range, with an amplitude maintained at the  $10^{-5}$  level, indicating that the primary dynamic response of the system is suppressed, thereby enhancing low-frequency stability. Additionally, the low-frequency components are minimal, reducing the likelihood of large steady-state deviations. Meanwhile, the energy in the high-frequency range ( $> 0.5$  Hz) is significantly suppressed, improving the system's oscillation damping capability and overall dynamic stability.

In the absence of ARC and under control attacks, the system's spectral characteristics change significantly. A prominent energy peak of approximately  $10^{-2}$  appears in the 0.4–0.6 Hz range, reflecting high-frequency oscillations induced by the additional control signals introduced by the attack. These oscillations may lead to coupling oscillations between generating units and the load side. In contrast, measurement attacks have a weaker impact on this frequency range but can still cause disturbances in the mid-frequency range, posing a risk to system stability.

When ARC is disabled and measurement attacks occur, energy accumulates in the low-frequency range ( $< 0.1$  Hz), indicating a prolonged deviation of the system from its equilibrium point, which increases the risk of steady-state deviation. Although no significant high-frequency peaks are observed, measurement attacks have a more persistent impact. In Scenario 3, measurement attacks may induce progressive deviations, further deteriorating system frequency stability. Without additional control strategies, these deviations will continue to amplify, ultimately affecting the long-term operational reliability of the system.

A comparative analysis of different operating conditions reveals that the ARC mechanism effectively suppresses high-frequency oscillations while significantly reducing steady-state deviations. Across all tested scenarios, enabling ARC

**Fig. 5** Frequency response energy distribution characteristics of the power system, **a** Scenario 1; **b** Scenario 2; **c** Scenario 3



results in a 1–2 order-of-magnitude reduction in power spectral density, with a more concentrated energy distribution, demonstrating strong frequency disturbance suppression capability. These findings confirm that the ARC mechanism enhances system stability in the presence of FDIA.

To systematically validate the superior performance of the proposed method in both attack detection and estimation, four representative existing approaches were selected for comparative analysis: Standard Kalman Filter (KF) [33], Extended Kalman Filter (EKF) [34], Extended State Observer (ESO) [35], and Two-Stage Denoising Diffusion Model (TSDM) [36]. The comparison focuses on key performance metrics, including attack detection success rate, estimation mean squared error (MSE), maximum frequency deviation, and system settling time. The results are presented in Table 1.

As shown in Table 1, the proposed approach outperforms all other methods across all evaluated metrics. Specifically, the detection success rate of the proposed method reaches 95.2%, representing a significant improvement over the Standard Kalman Filter (87.3%), Extended Kalman Filter (89.7%), Extended State Observer (91.5%), and Two-Stage Denoising Diffusion Model (92.6%). Even under low-amplitude and stealthy attack scenarios, the proposed approach maintains a high detection rate, demonstrating superior sensitivity and robustness.

In terms of estimation accuracy, the proposed method achieves a mean squared error of  $6.28 \times 10^{-3}$ , which is substantially lower than KF ( $1.35 \times 10^{-2}$ ), EKF ( $1.12 \times 10^{-2}$ ), and ESO ( $9.87 \times 10^{-3}$ ). Compared to the more accurate TSDM ( $8.70 \times 10^{-3}$ ), the MSE is further reduced by 27.8%, confirming the high precision of the proposed method in attack reconstruction.

Regarding the dynamic performance of the system, the proposed method restricts the maximum frequency deviation to 0.053 Hz, which is clearly superior to that of KF (0.078 Hz), EKF (0.072 Hz), ESO (0.068 Hz), and TSDM (0.065 Hz). This demonstrates its ability to effectively suppress frequency disturbances caused by attacks, thereby improving the overall security and stability of the power system. Moreover, the settling time of 17.4 s achieved by the proposed method represents at least a 14.7% reduction compared to traditional methods, all of which required more than 20 s, highlighting its faster recovery capability.

In summary, the Standard Kalman Filter suffers from poor detection performance under modeling errors and subtle attacks. While the Extended Kalman Filter can handle some nonlinearities, it is still limited when facing stealthy attacks. The Extended State Observer offers certain robustness but struggles to distinguish between control input and measurement attacks. Although the Two-Stage Denoising Diffusion Model adapts well to multimodal data variations, its real-time performance is limited. In contrast, the proposed method not only maintains real-time capability but also accurately differentiates between control input and measurement attacks. It delivers optimal detection accuracy, reconstruction precision, and system stability, thereby validating the effectiveness and superiority of the proposed strategy.

## 6 Conclusion

This paper presents a novel FDIA detection and mitigation strategy for power system LFC based on state-space decomposition, SMO design, and  $H_\infty$ -based ARC. By transforming system coordinates, the method effectively decouples control input attacks and measurement attacks, enabling their independent estimation through two dedicated SMOs. Simulation results show that the proposed approach reduces attack estimation error by approximately 30% compared to traditional Kalman filtering and achieves high estimation accuracy under various operational conditions.

The integration of the estimated attack signals into the  $H_\infty$  controller forms a closed-loop ARC schemes that enhances system resilience. Spectral analysis confirms that ARC significantly suppresses both high-frequency oscillations and low-frequency deviations, reducing power spectral density by 1–2 orders of magnitude. The entire framework is model-based,

**Table 1** Comparison of detection model performance

| Method          | Attack detection success rate (%) | MSE                   | Max. frequency deviation (Hz) | Settling time (s) |
|-----------------|-----------------------------------|-----------------------|-------------------------------|-------------------|
| KF              | 87.3%                             | $1.35 \times 10^{-2}$ | 0.078                         | 23.5              |
| EKF             | 89.7%                             | $1.12 \times 10^{-2}$ | 0.072                         | 21.8              |
| ESO             | 91.5%                             | $9.87 \times 10^{-3}$ | 0.068                         | 20.6              |
| TSDM            | 92.6%                             | $8.70 \times 10^{-3}$ | 0.065                         | 20.1              |
| Proposed method | 95.2%                             | $6.28 \times 10^{-3}$ | 0.053                         | 17.4              |

adaptive, and does not require prior attack data, making it transparent, interpretable, and easy to implement in practical grid applications.

In summary, the proposed method improves both the detection accuracy and control robustness of LFC systems under FDIA conditions. Future work will extend the approach to nonlinear models, incorporate intelligent tuning mechanisms, and explore distributed detection in multi-area systems to further enhance cyber resilience.

**Author contributions** Conceptualization, Chao Hong and Zhihong Liang; methodology, Chao Hong; software, Chao Hong; validation, Chao Hong, Yiwei Yang, and Pandeng Li; formal analysis, Chao Hong; investigation, Chao Hong; resources, Chao Hong; data curation, Chao Hong; writing-original draft preparation, Chao Hong; writing-review and editing, Chao Hong, Lin Chen, and Leyu Bi; visualization, Chao Hong; supervision, Zhihong Liang and Yunan Zhang; project administration, Zhihong Liang; funding acquisition, Yiwei Yang. All authors have read and agreed to the published version of the manuscript.

**Funding** This research was funded by the Science and Technology Project of Guizhou Power Grid Company, grant number GZKJXM20222346.

**Data availability** The datasets generated and analyzed during the current study are not publicly available due to ongoing research and potential sensitivity of the power system configurations. However, the data that support the findings of this study are available from the corresponding author upon reasonable request.

## Declarations

**Ethics approval and consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Competing interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

1. Lin F, Mei Y, Zhu Y, et al. Overview of the entire process influence of cyber attack on typical scenarios of power systems. *South Power Syst Technol.* 2023;17(11):61–75.
2. Li P, Liu N, Hu Q. Commentary on special issue of reviews on key technologies for digitalization of new power system. *Autom Electric Power Syst.* 2024;48(6):1–12.
3. Jin M, Lavaei J, Johansson KH. Power grid ac-based state estimation: vulnerability analysis against cyber attacks. *IEEE Trans Autom Control.* 2019;64(5):1784–99. <https://doi.org/10.1109/TAC.2018.2852774>.
4. Alexopoulos TA, Korres GN, Manousakis NM. Complementarity reformulations for false data injection attacks on pmu-only state estimation. *Electric Power Syst Res.* 2020;189: 106796. <https://doi.org/10.1016/j.epsr.2020.106796>.
5. Lawal OA, Teh J, Alharbi B, Lai C-M. Data-driven learning-based classification model for mitigating false data injection attacks on dynamic line rating systems. *Sustain Energy Grids Netw.* 2024;38: 101347. <https://doi.org/10.1016/j.segan.2024.101347>.
6. Lawal OA, Teh J. A framework for modelling the reliability of dynamic line rating operations in a cyber-physical power system network. *Sustain Energy Grids Netw.* 2023;35: 101140. <https://doi.org/10.1016/j.segan.2023.101140>.
7. Lawal OA, Teh J. Dynamic line rating forecasting algorithm for a secure power system network. *Expert Syst Appl.* 2023;219: 119635. <https://doi.org/10.1016/j.eswa.2023.119635>.
8. Jimada-Ojuolape B, Teh J, Lai C-M. Enhancing power grid reliability with pmu placement in flexibly rated cyber-physical systems. *Electric Power Syst Res.* 2025;241: 111327. <https://doi.org/10.1016/j.epsr.2024.111327>.
9. Wei S, Wu Z, Xu J, Hu Q. Multiarea probabilistic forecasting-aided interval state estimation for FDIA identification in power distribution networks. *IEEE Trans Industr Inf.* 2024;20(3):4271–82. <https://doi.org/10.1109/TII.2023.3321098>.
10. Wei S, Xu J, Wu Z, Hu Q, Yu X. A false data injection attack detection strategy for unbalanced distribution networks state estimation. *IEEE Trans Smart Grid.* 2023;14(5):3992–4006. <https://doi.org/10.1109/TSG.2023.3235945>.
11. Pei C, Xiao Y, Liang W, Han X, Wang W. Canonical variate analysis for detecting false data injection attacks in alternating current state estimation. *IEEE Trans Netw Sci Eng.* 2024;11(4):3332–45. <https://doi.org/10.1109/TNSE.2024.3370649>.
12. Yang H, He X, Wang Z, Qiu RC, Ai Q. Blind false data injection attacks against state estimation based on matrix reconstruction. *IEEE Trans Smart Grid.* 2022;13(4):3174–87. <https://doi.org/10.1109/TSG.2022.3164874>.

13. Yang H, Zhang W, Liang Z, Wang Z, Chung CY, Wang Q. Parameter-free false data injection attack against ac state estimation: a canonical polyadic decomposition based approach. *IEEE Trans Power Syst.* 2025;40(3):2153–64. <https://doi.org/10.1109/TPWRS.2024.3465874>.
14. Yang H, Zhang W, Chung CY, Wang Z, Qiu W, Liang Z. Ac false data injection attack based on robust tensor principle component analysis. *IEEE Trans Industr Inf.* 2024;20(8):9887–97. <https://doi.org/10.1109/TII.2024.3390389>.
15. Bhattacharjee A, Mondal AK, Verma A, Mishra S, Saha TK. Deep latent space clustering for detection of stealthy false data injection attacks against ac state estimation in power systems. *IEEE Trans Smart Grid.* 2023;14(3):2338–51. <https://doi.org/10.1109/TSG.2022.3216625>.
16. Wang J, Chen H, Si Y, Zhu Y, Zhu T, Yin S, Liu B. FDIA localization and classification detection in smart grids using multi-modal data and deep learning technique. *Comput Electr Eng.* 2024;119: 109572. <https://doi.org/10.1016/j.compeleceng.2024.109572>.
17. Zhang G, Li J, Bamisile O, Cai D, Hu W, Huang Q. Spatio-temporal correlation-based false data injection attack detection using deep convolutional neural network. *IEEE Trans Smart Grid.* 2022;13(1):750–61. <https://doi.org/10.1109/TSG.2021.3109628>.
18. Peng S, Zhang Z, Deng R, Cheng P. Localizing false data injection attacks in smart grid: a spectrum-based neural network approach. *IEEE Trans Smart Grid.* 2023;14(6):4827–38. <https://doi.org/10.1109/TSG.2023.3261970>.
19. Xi L, Tian X, Yu T, Cheng C. Locational detection of false data injection attack in power grid based on relevant features multi-label cascade boosting forest. *South Power System Technol.* 2024;18(5):39–50.
20. Qu Z, Dong Y, Li Y, Song S, Jiang T, Li M, Wang Q, Wang L, Bo X, Zang J, Xu Q. Localization of dummy data injection attacks in power systems considering incomplete topological information: a spatio-temporal graph wavelet convolutional neural network approach. *Appl Energy.* 2024;360: 122736. <https://doi.org/10.1016/j.apenergy.2024.122736>.
21. He Y, Wang J, Yang C, Shi D. A graph and diffusion theory-based approach for localization and recovery of false data injection attacks in power systems. *Electric Power Syst Res.* 2025;239: 111184. <https://doi.org/10.1016/j.epsr.2024.111184>.
22. Liu M, Zhang X, Zhao C, Deng R. Matrix coding enabled impact mitigation against primary false data injection attacks in cyber-physical microgrids. *IEEE Trans Power Syst.* 2025. <https://doi.org/10.1109/TPWRS.2025.3528322>.
23. Lu A-Y, Yang G-H. A polynomial-time algorithm for the secure state estimation problem under sparse sensor attacks via state decomposition technique. *IEEE Trans Autom Control.* 2023;68(12):7451–65. <https://doi.org/10.1109/TAC.2023.3278839>.
24. Jimada-Ojuolape B, Teh J. Composite reliability impacts of synchrophasor-based DTR and SIPS cyber-physical systems. *IEEE Syst J.* 2022;16(3):3927–38. <https://doi.org/10.1109/JSYST.2021.3132657>.
25. Jimada-Ojuolape B, Teh J. Impacts of communication network availability on synchrophasor-based DTR and SIPS reliability. *IEEE Syst J.* 2022;16(4):6231–42. <https://doi.org/10.1109/JSYST.2021.3122022>.
26. Jimada-Ojuolape B, Teh J. Surveys on the reliability impacts of power system cyber-physical layers. *Sustain Cities Soc.* 2020;62: 102384. <https://doi.org/10.1016/j.scs.2020.102384>.
27. Al Faruque MA, Vatanparvar K. Energy management-as-a-service over fog computing platform. *IEEE Internet Things J.* 2016;3(2):161–9. <https://doi.org/10.1109/JIOT.2015.2471260>.
28. Kim J, Tong L. On topology attack of a smart grid: undetectable attacks and countermeasures. *IEEE J Sel Areas Commun.* 2013;31(7):1294–305. <https://doi.org/10.1109/JSAC.2013.130712>.
29. Teixeira A, Sou KC, Sandberg H, Johansson KH. Secure control systems: a quantitative risk management approach. *IEEE Control Syst Mag.* 2015;35(1):24–45. <https://doi.org/10.1109/MCS.2014.2364709>.
30. Gulzar MM, Sibtain D, Alqahtani M, Alismail F, Khalid M. Load frequency control progress: a comprehensive review on recent development and challenges of modern power systems. *Energ Strat Rev.* 2025;57: 101604. <https://doi.org/10.1016/j.esr.2024.101604>.
31. Bravo J, Ortiz L, García E, Ruiz M, Aguila A. An on-line sensor fault detection system for an ac microgrid secondary control based on a sliding mode observer model. *Energies.* 2024. <https://doi.org/10.3390/en17153808>.
32. Yue D, Hu S, Cheng Z. Resilient load frequency control design: nonperiodic dos attacks against additional control loop. 2021. pp. 225–50. [https://doi.org/10.1007/978-981-33-6730-2\\_11](https://doi.org/10.1007/978-981-33-6730-2_11).
33. Zhang Z, Hu J, Lu J, Cao J, Alsaadi FE. Preventing false data injection attacks in lfc system via the attack-detection evolutionary game model and kf algorithm. *IEEE Trans Netw Sci Eng.* 2022;9(6):4349–62. <https://doi.org/10.1109/TNSE.2022.3199881>.
34. Li X, Wang Z, Zhang C, Du D, Fei M. A novel dynamic watermarking-based EKF detection method for FDIAs in smart grid. *IEEE/CAA J Automatica Sinica.* 2022;9(7):1319–22. <https://doi.org/10.1109/JAS.2022.105704>.
35. Ahmad S, Ahmed H. Robust intrusion detection for resilience enhancement of industrial control systems: an extended state observer approach. *IEEE Trans Ind Appl.* 2023;59(6):7735–43. <https://doi.org/10.1109/TIA.2023.3305361>.
36. Pei J, Wang J, Shi D, Wang P. Improved efficient two-stage denoising diffusion power system measurement recovery against false data injection attacks and data losses. 2023. arXiv preprint [arXiv:2312.04346](https://arxiv.org/abs/2312.04346).

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.