Reliability Engineering and System Safety 232 (2023) 109067

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Sensor fault detection of vehicle suspension systems based on transmissibility operators and Neyman–Pearson test

Ying Wang [a], Xueke Zheng [a], Le Wang [a], Gavin Lu [b], Yixing Jia [b], Kezhi Li [a], Mian Li [c,d,*]

[a] University of Michigan – Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, 800 Dongchuan Road, Minghang district, Shanghai, 200240, China
[b] VMware 1/F, Tower C,Raycom InfoTech Park No. 2 Kexueyuan South Road Haidian District, Beijing, 100190, China
[c] Global Institute of Future Technology, Shanghai Jiao Tong University, 800 Dongchuan Road, Minghang district, Shanghai, 200240, China
[d] Department of Automation, Shanghai Jiao Tong University, 800 Dongchuan Road, Minghang district, Shanghai, 200240, China

ARTICLE INFO

Keywords:
Sensor fault detection
Transmissibility
FIR model
Neyman–Pearson test

ABSTRACT

As sensors become increasingly important and popular in vehicle suspension systems for control and monitoring purposes, the sensor faults endanger the reliability and safety of suspension systems and have raised broad concerns. However, due to the complex structures and unknown inputs of suspension systems, it is difficult to detect sensor faults using purely physics-based or data-driven approaches. This paper proposes a Physics-Informed Machine Learning (PIML) method, combining a residual generation procedure based on the transmissibility operators and a residual evaluation procedure based on the Neyman–Pearson test, to detect sensor faults in suspension systems as early as possible: First, transmissibilities are derived from the physical models to describe the relationship between the outputs of an underlying system; then, noncausal Finite Impulse Response (FIR) models are used to approximate the transmissibilities to generate the model residuals; finally, the generated residuals from noncausal FIR models are evaluated using the Neyman–Pearson test to detect the sensor faults, along with the analysis of the fault detectability and sensitivity. The feasibility and effectiveness of the proposed method are validated using a numerical example and a real-world application, and good detection performance can be obtained in terms of Receiver Operating Characteristic (ROC) curves.

1. Introduction

Suspension systems are key vehicle components to guarantee ride quality, stability, and handling performance [1,2]. Recently, the vibration signals of the suspension systems are increasingly important for control and monitoring purposes, given that the global electronization trend of the automobile industry facilitates efficient signal acquisition through a variety of sensors [3]. For example, the widely used semi-active suspension systems consist of many sensors and actuators to adjust the damping characteristics of the suspension dynamically to deal with the packaging and energy issues [1]; researchers have also investigated structural health monitoring techniques based on the suspension displacements and accelerations in a full vehicle model [4–6]. However, with many sensors being installed, sensor failures are inevitable, which may affect the performance of control or monitoring, and even deteriorate the reliability and safety of the running systems [7–9]. Therefore, it is important to detect the sensor faults in the suspension systems as early as possible.

Sensor faults can be caused by various reasons, such as low voltage, embedded program failure, and communication failure [10]. The most common fault types include complete failure, bias, drift, and precision degradation [7,8,11]. It is easiest to detect the complete failure because the sensor will no longer provide any reading. The other three types of sensor faults are much harder to detect and have attracted considerable attention [9]. In the literature, most sensor fault detection methods can be divided into model-free and model-based methods [12]. Model-free methods assume that faults can be reflected in the measured signals. Thus, the key of those methods is to extract signal features that are sensitive to faults, and based on these features, a diagnostic hypothesis is made [13–15]. Researchers have investigated these methods in the time domain (such as the principal component analysis and canonical variate analysis), frequency domain, and time–frequency domain [15]. However, vehicle systems in real scenarios work under complex driving conditions [16]. It could be difficult to analyze signal characteristics and further make strong hypotheses for diagnosis. Alternatively, the model-based methods have recently attracted strong interest since the development of machine learning and system identification techniques has brought about powerful capabilities in predictive modeling [12].

* Corresponding author.
E-mail address: mianli@sjtu.edu.cn (M. Li).

https://doi.org/10.1016/j.ress.2022.109067
Received 3 January 2022; Received in revised form 16 November 2022; Accepted 26 December 2022
Available online 3 January 2023
0951-8320/© 2023 Elsevier Ltd. All rights reserved.

The basic idea of model-based methods includes the following two steps:

(i) *Residual generation (step 1):* generate the model residuals by using predictive models;
(ii) *Residual evaluation (step 2):* evaluate the generated residuals in step 1 to determine whether sensor faults occur [17].

In the first step of the model-based fault detection scheme, i.e., residual generation, the modeling techniques play an essential role and usually are grouped into two different categories, namely *physics-based* approaches and *machine learning (ML)-based* approaches [18,19].

Physics-based techniques, such as parity space [20,21] and observer-based [22] techniques, usually require knowledge of the accurate first-principle models of suspension systems [23]. However, it is difficult to obtain the first-principle models in some scenarios, where the vehicles are driven under complex environmental and maneuvering conditions [4,5,24]. Moreover, the input of suspension systems (i.e., the excitation from roads) is usually unknown [25], which makes the sensor fault detection based on the first principle even more difficult.

Machine learning (ML)-based approaches, which construct data-driven models using historical data [5], have been widely considered in the recent past and have shown to be successful in many applications [18,19,26]. Particularly, neural networks have been used to detect faults in suspension systems [27,28]. These methods have the ability to automatically extract complex relationship from data for systems that are not fully understood [29]. However, it might be problematic in some scenarios due to their requirement for a large amount of high-quality training data [19]. Also, the ML-based methods may lack generalizability and interpretability, which could limit their engineering applications [30].

As a promising technical direction, Physics-Informed Machine Learning (PIML) techniques have attracted extensive attention and achieved satisfying performance for many reliability and system safety applications [31–33]. PIML methods incorporate physical information with ML approaches to reduce the computational cost and enhance the model performance, interpretability, generalizability, etc. [19,29]. According to the latest literature on PIML [19], the PIML approaches can be divided into two categories, depending on how the physics knowledge is embedded. The first approach is to design a physics-informed loss function, which can enforce known physical constraints into ML models. The second approach is to develop a physics-informed architecture to incorporate physics knowledge into the ML model architecture. In this work, the proposed PIML method falls into the second category, as will be introduced in the next paragraph.

Transmissibilities have been derived from the physical models to describe the relationship between the outputs of the system. In contrast with transmissibilities in the frequency domain, transmissibilities in the time domain have proven to be independent of the initial conditions and excitation of the underlying system [23,34]. This property significantly facilitates the sensor fault detection of suspension systems, where both the initial conditions and excitation are unknown [4,6]. Moreover, derived from Laurent series expansion of transmissibilities, the noncausal Finite Impulse Response (FIR) model structure has a solid mathematical foundation and has been validated as a good model structure to approximate transmissibilities in many sensor-to-sensor identification problems [35–37]. Lastly, the model parameters of the noncausal FIR models can be easily identified from the historical sensor data. Thus, the PIML method, combining the transmissibilities from the physical models with the data-driven sensor-to-sensor FIR models, could be proposed to make the sensor fault detection for suspension systems more efficient, generalizable, and interpretable.

When it comes to the second step of the model-based fault detection scheme, i.e., residual evaluation, we aim to detect any residual deviation from the normal case and then determine whether the sensor faults occur [38]. Model residuals are inevitable given the existence of uncertainties caused by modeling errors, measurement noise, disturbances, etc [39]. The decision rules to distinguish the abnormal residuals from the fault-free ones are usually based on statistical hypothesis testing with common test statistics, such as T^2, Q, and Hotelling T^2 [40]. However, in most previous studies, residual evaluation is usually performed using a fixed threshold defined empirically [1]. This threshold is often difficult to determine without deep domain knowledge in practice.

As a hypothesis testing algorithm, the Neyman–Pearson (NP) test is the criterion of the most powerful test function maximizing the detection probability with a constraint on the false alarm probability [41–43]. In the literature, it has been successfully used to detect and isolate faults in many applications [44,45]. It is worthwhile to note that the NP test provides a way to automatically determine a threshold with the guaranteed low probability of false alarms [38,41]. Thus, the NP test based method could be developed in the sensor fault detection for suspension systems where the threshold associated with the residual evaluation is difficult to determine empirically.

In this work, we aim to detect multiple types of sensor faults in suspension systems as early as possible. A novel and effective PIML method based on transmissibilities, noncausal FIR models, and NP test detectors is proposed to deal with multiple types of sensor faults in a unified framework: the residuals are generated by data-driven noncausal FIR models theoretically developed from sensor-to-sensor transmissibilities, and then the generated residuals are successfully evaluated by NP detectors, which can detect the sensor faults as early as possible. The PIML method inherits the advantages of both physics-based methods and ML-based ones: It is interpretable and generalizable in the context that the FIR model structure is derived from the physical models of the underlying systems. In addition, as linear data-driven models, the FIR models can be efficiently built to provide a good approximation to the sensor-to-sensor system.

The main contributions of this work can be interpreted from the following three perspectives: (1) A novel model-based sensor fault detection framework is proposed for the reliability and safety of vehicle suspension systems in real scenarios. Compared with previous studies with a simplified quarter-car suspension system [46,47], our investigation is carried out with a full vehicle system in a real testing environment, leading to more realistic settings. (2) Two typical sensor fault types, i.e., bias and precision degradation, are addressed under a unified NP detector framework. The threshold associated with the residual evaluation can be obtained automatically without any manual tuning or deep domain knowledge. To the authors' best knowledge, it is the first time that the precision degradation is solved using the NP detector, which has been verified efficient and effective through the experiments. (3) In addition, the concepts of the sensitivity and detectability of sensor faults, which have not been fully addressed in the previous work, are defined and discussed in detail in the context of the NP test, providing more insights into the properties of the proposed method.

A numerical and a real-world application are used to demonstrate the feasibility and effectiveness of the proposed method for the sensor fault detection in suspension systems. It is believed that the proposed method of sensor fault detection in this work will facilitate the development of reliability or safety design of suspension systems.

The rest of the paper is organized as follows: In Section 2, the transmissibility operators, the noncausal FIR models, and the ridge regression are introduced briefly. In Section 3, the sensor fault detection method based on the NP test is discussed in detail. In Sections 4 and 5, a numerical example developed from a quarter-car suspension system and a real-world application are demonstrated respectively to demonstrate the effectiveness of the proposed method. Finally, the conclusion and future works are presented in Section 6.

2. Preliminaries

As introduced in Section 1, the main idea of model-based methods includes two steps: residual generation and residual evaluation. The

preliminaries are centered on the first step, i.e., generating residuals with a predictive model. More specifically, the main purposes of this section are to derive a data-driven model structure from physics-based models and introduce a method to determine the model parameters. Then, the model residuals can be generated based on the constructed models.

In this section, the construction of Single-Input Single-Output (SISO) transmissibility operators from state space models will be introduced briefly in Section 2.1. Then, the sensor-to-sensor models derived from the Laurent Series expansion of transmissibility operators will be introduced in Section 2.2. Lastly, the ridge regression will be used to identify the sensor-to-sensor model parameters in Section 2.3.

2.1. Transmissibility operators

Transmissibility operators are mathematical models that characterize the relationship between the outputs of an underlying system. In other words, the input and output of a transmissibility operator are outputs of the underlying system [23,34]. We here consider a state-space system with the single input and two outputs as follows:

$$\dot{x}(t) = Ax(t) + bu(t), \tag{1}$$

$$x(0) = x_0, \tag{2}$$

$$y_{\mathrm{I}}(t) \triangleq c_{\mathrm{I}}x(t) + d_{\mathrm{I}}u(t), \tag{3}$$

$$y_{\mathrm{O}}(t) \triangleq c_{\mathrm{O}}x(t) + d_{\mathrm{O}}u(t), \tag{4}$$

where $A \in \mathbb{R}^{n_x \times n_x}$, $b \in \mathbb{R}^{n_x}$, $c_{\mathrm{I}}^{\mathrm{T}}, c_{\mathrm{O}}^{\mathrm{T}} \in \mathbb{R}^{n_x}$, and $d_{\mathrm{I}}, d_{\mathrm{O}} \in \mathbb{R}$. The system input $u \in \mathbb{R}$ and the initial state $x_0 \in \mathbb{R}^{n_x}$ are unknown, whereas the system outputs $y_{\mathrm{I}} \in \mathbb{R}$ and $y_{\mathrm{O}} \in \mathbb{R}$ can be measured by sensors.

In order to obtain a transmissibility operator whose pseudo input and pseudo output are denoted by y_{I} and y_{O} respectively, we define the polynomials

$$\Gamma_{\mathrm{I}}(\mathbf{p}) \triangleq c_{\mathrm{I}} \operatorname{adj}(\mathbf{p}\mathbf{I} - A)b + d_{\mathrm{I}}\delta(\mathbf{p}) \in \mathbb{R}[\mathbf{p}], \tag{5}$$

$$\Gamma_{\mathrm{O}}(\mathbf{p}) \triangleq c_{\mathrm{O}} \operatorname{adj}(\mathbf{p}\mathbf{I} - A)b + d_{\mathrm{O}}\delta(\mathbf{p}) \in \mathbb{R}[\mathbf{p}], \tag{6}$$

$$\delta(\mathbf{p}) \triangleq \det(\mathbf{p}\mathbf{I} - A) \in \mathbb{R}[\mathbf{p}], \tag{7}$$

where $\mathbf{p} \triangleq \mathrm{d}/\mathrm{d}t$, $\operatorname{adj}(\mathbf{p}\mathbf{I} - A)$ and $\det(\mathbf{p}\mathbf{I} - A)$ denote the adjugate matrix and the determinant of $\mathbf{p}\mathbf{I} - A$, respectively. Here $\mathbf{I}$ is the identity matrix with the appropriate dimension. With these polynomials, the following equation has been proved by [35,48]

$$\Gamma_{\mathrm{I}}(\mathbf{p})y_{\mathrm{O}}(t) = \Gamma_{\mathrm{O}}(\mathbf{p})y_{\mathrm{I}}(t). \tag{8}$$

Assume that $\Gamma_{\mathrm{I}}(\mathbf{p})$ is not the zero polynomial, Eq. (8) can be rewritten into

$$y_{\mathrm{O}}(t) = \mathcal{T}(\mathbf{p})y_{\mathrm{I}}(t), \tag{9}$$

where the transmissibility operator from the pseudo input y_{I} to the pseudo output y_{O} is defined by [48]

$$\mathcal{T}(\mathbf{p}) = \frac{\Gamma_{\mathrm{O}}(\mathbf{p})}{\Gamma_{\mathrm{I}}(\mathbf{p})}. \tag{10}$$

Different from the complex Laplace variable $\mathbf{s}$, the time-domain operator $\mathbf{p}$ accounts for nonzero initial conditions as explained by [48], and thus the transmissibility operator is independent of both the initial state x_0 and the input u.

Eq. (10) indicates that the information about the zeros of the system but not the poles will influence the properties of $\mathcal{T}(\mathbf{p})$: if the system from u to y_{I} has a nonminimum-phase (unstable) zero, $\mathcal{T}(\mathbf{p})$ will be unstable; if the system from u to y_{O} has more zeros than the system from u to y_{I}, then $\mathcal{T}(\mathbf{p})$ will be noncausal [36]. However, the system information may not be known in advance in some cases. Therefore, a model structure with the ability to approximate the noncausal and unstable transmissibilities has been designed [36], which will be introduced in the next subsection.

Remark 1. Compared with transmissibilities in the frequency domain [49,50], the advantage of transmissibilities in the time domain is that their estimates are independent of the initial condition and the excitation signal acting on the underlying system. This property makes the proposed method robust to various vehicle driving conditions.

2.2. Data-driven model design based on transmissibilities

Since sensor measurements are obtained at discrete time steps, the discrete-time transmissibility operator $\mathcal{T}(\mathbf{q})$ is developed by replacing $\mathbf{p}$ by forward shift operator $\mathbf{q}$ [35,48], i.e., $y_{\mathrm{I}}(k) = \mathbf{q}y_{\mathrm{I}}(k-1)$.

To estimate the transmissibility $\mathcal{T}(\mathbf{q})$ between the output measurements without the information about the dynamics of the system (the state space model, the system excitation, and the initial state), a model structure that can approximate noncausal and unstable transmissibilities is proposed by [51]. The model structure is known as the noncausal FIR model and is derived by truncating the following Laurent series expansion of $\mathcal{T}(\mathbf{q})$ inside the annulus between the asymptotically stable pole of the largest modulus and the unstable pole of the smallest modulus [51]

$$\mathcal{T}(\mathbf{q}) = \sum_{i=-\infty}^{\infty} h_i \mathbf{q}^{-i}, \tag{11}$$

where the Markov parameter $h_i \in \mathbb{R}$ is the ith coefficient of the Laurent expansion of $\mathcal{T}(\mathbf{q})$ in the given annulus.

A noncausal FIR model based on $\mathcal{T}(\mathbf{q})$ can be expressed as

$$\begin{aligned} y_{\mathrm{O}}(k) &= \sum_{i=-d}^{m} h_i \mathbf{q}^{-i} y_{\mathrm{I}}(k) + \epsilon(k) \\ &= h_{-d} y_{\mathrm{I}}(k+d) + \cdots + h_0 y_{\mathrm{I}}(k) + \\ &\quad \cdots + h_m y_{\mathrm{I}}(k-m) + \epsilon(k), \end{aligned} \tag{12}$$

where the non-negative integers m and d denote the order of the causal and noncausal parts of the noncausal FIR model, respectively, and $\epsilon(k)$ is the model residual representing the unmodeled dynamics or the noise. The causal part, i.e., the negative and zero powers of $\mathbf{q}$, approximates the asymptotically stable part of the plant, whereas the noncausal part, i.e., the positive powers of $\mathbf{q}$, approximates the unstable part of the plant [51].

Remark 2. The introduction of transmissibility operators and their Laurent expansion provides a mathematical foundation for using the noncausal FIR models. Suppose physical information about the system is known. It can be applied to the design of the noncausal FIR model, such as the model order determination of causal and noncausal parts. In some cases, there is no information about the dynamics of the system. The orders of the noncausal FIR model can be determined by cross-validation or some classic methods in system identification, such as Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC). [52].

Remark 3. The advantages of the noncausal FIR models based on transmissibilities include: (1) as a PIML model, it has a strong mathematical foundation, which makes it interpretable and generalizable to many similar applications; (2) the simple structure makes the calculation efficient, which is applicable to the scenarios with limited computational resources, such as IoT and edge computing; (3) the linearity of the model makes it feasible to analyze important properties, such as fault sensitivity (i.e., how sensor faults affect the model residuals), as given in Section 3.

2.3. Ridge regression

To obtain the approximated transmissibility operator $\hat{\mathcal{T}}(\mathbf{q})$, the noncausal FIR model parameters need to be estimated using historical data. Eq. (12) can be rewritten as

$$y_{\mathrm{O}}(k) = \phi(k)\theta + \epsilon(k), \quad k = m, \ldots, N - d, \tag{13}$$

where the parameter vector $\theta \triangleq [h_{-d} \ \cdots \ h_m]^{\mathrm{T}}$, the regressor $\phi(k) \triangleq [y_{\mathrm{I}}(k+d) \ \cdots \ y_{\mathrm{I}}(k-m)]^{\mathrm{T}}$, and N is the number of training samples. Stacking the elements of $y_{\mathrm{O}}(k)$ into the vectors and stacking the rows of $\phi(k)$ into the matrices yield the following linear regression problem

$$Y = \boldsymbol{\Phi}^{\mathrm{T}}\theta + E, \tag{14}$$

where

$$Y \triangleq \begin{bmatrix} y_{\mathrm{O}}(m) & \cdots & y_{\mathrm{O}}(N-d) \end{bmatrix}^{\mathrm{T}}, \tag{15}$$

$$\boldsymbol{\Phi} \triangleq \begin{bmatrix} \phi(m) & \cdots & \phi(N-d) \end{bmatrix}, \tag{16}$$

$$E \triangleq \begin{bmatrix} \epsilon(m) & \cdots & \epsilon(N-d) \end{bmatrix}^{\mathrm{T}}. \tag{17}$$

The least-squares estimate of θ in Eq. (14) is given by

$$\hat{\theta} = \left(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}\right)^{-1}\boldsymbol{\Phi}Y \tag{18}$$

provided that $\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}$ is non-singular, which is guaranteed when y_{I} is persistently exciting [35,53]. There are some scenarios where $\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}$ is close to singular, which leads to a poor estimate of θ. In order to tackle this issue, ridge regression as a regularization technique is introduced

$$\underset{\theta}{\text{minimize}} \quad \left\|Y - \boldsymbol{\Phi}^{\mathrm{T}}\theta\right\|_2^2 + \rho\theta^{\mathrm{T}}\theta, \tag{19}$$

where $\rho\theta^{\mathrm{T}}\theta$ is a penalty term, and ρ is a tuning parameter [54]. The closed-form solution to this optimization problem in Eq. (19) is

$$\hat{\theta} = \left(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}} + \rho\mathbf{I}\right)^{-1}\boldsymbol{\Phi}Y, \tag{20}$$

and the hyperparameter ρ can be chosen as

$$\rho = \begin{cases} 0, & \kappa(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}) \le C_{\mathrm{lim}}, \\ \dfrac{\lambda_{\max} - \lambda_{\min}\cdot C_{\mathrm{lim}}}{C_{\mathrm{lim}} - 1}, & \kappa(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}) > C_{\mathrm{lim}}, \end{cases} \tag{21}$$

where $\kappa(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}})$ denotes the condition number of $\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}$ and C_{lim} denotes the acceptable conditioning number. More details about the ridge regression can be found in paper [4,54].

Remark 4. When the physical information about the system is unknown, the selection of the model orders m and d is one issue of using noncausal FIR models. An inappropriate selection of model orders will lead to either a large bias or variance for the model parameter θ. The advantage of regularization techniques is that they can strike a balance between bias and variance [54]. Ridge regression is introduced as a classic regularization method where the choice of the penalty term is squared. In the system identification field, there are other methods to design the regularization terms [55,56], which can be further considered in the future.

3. Methodology

With the ridge regression estimate $\hat{\theta} = [\hat{h}_{-d} \ \cdots \ \hat{h}_m]^{\mathrm{T}}$ of θ being obtained, the residual of the approximated transmissibility $\hat{\mathcal{T}}(\mathbf{q})$ at time k is given by

$$r(k) = y_{\mathrm{O}}(k) - \hat{y}_{\mathrm{O}}(k|\hat{\theta}) = y_{\mathrm{O}}(k) - \sum_{i=-d}^{m} \hat{h}_i y_{\mathrm{I}}(k-i). \tag{22}$$

In this section, the residual evaluation procedure will be conducted in the framework of the NP test under the assumption that the model residual $r(k)$ in the fault-free condition, denoted by $\xi(k)$, can be approximated by a Gaussian distribution with zero mean and variance σ^2, which has been widely used [57]. Note that $r(k)$ will be different from $\xi(k)$ when the sensor faults occur in Sections 3.1 and 3.2.

It will be shown that the sensor faults will change the mean or the variance of the Gaussian distribution of the model residuals. Then, the sensor fault detection problem can be treated as a binary hypothesis testing problem: the null hypothesis H_0 (the fault-free residuals) and the alternative hypothesis H_1 (the abnormal residuals) [38]. Based on this hypothesis test, the NP lemma introduced by [42] provides the optimal test statistic to maximize the detection probability (i.e., the probability of deciding H_1 when H_1 is true) given the false alarm probability (i.e., the probability of deciding H_1 when H_0 is true).

Table 1
Mathematical expressions of four types of classic sensor faults, where s denotes the normal measurements with noise and s' denotes the abnormal measurements corrupted by typical faults. Drift is an offset with an initial offset e_0 and increases with time k. Precision degradation is described by a Gaussian distribution with zero mean and variance σ_e^2.

Types of sensor faults	Mathematical expression
Complete failure	$s' = e, e = \text{const}$
Bias	$s' = s + e, e = \text{const}$
Drift	$s' = s + e_k, e_k = ke_0, e_0 = \text{const}$
Precision degradation	$s' = s + e, e \sim \mathcal{N}(0, \sigma_e^2)$

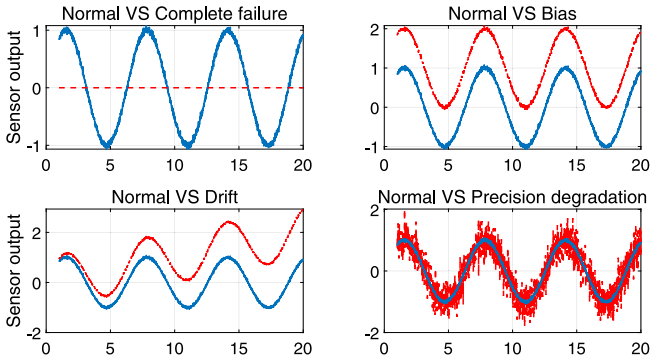


Fig. 1. The comparison of the fault-free sensor data (the blue line) and the abnormal sensor data (the red line) corrupted by the four types of typical sensor faults.

Table 1 and Fig. 1 show four types of sensor faults [58]. Given that it is easiest to detect the sensor with complete failure, whose output is a constant, we will not discuss this fault. From the statistics point of view, the bias and drift affect the mean of measurements, while the precision degradation fault disturbs the variance of measurements [9]. Thus, these three faults are classified into two types, mean-related faults and variance-related faults.

3.1. Mean-related fault detection

In order to analyze the fault detectability, the maximum acceptable mean deviations of the pseudo input sensor and pseudo output sensor are denoted by e_{I} and e_{O}, respectively. Our purpose is that once the deviation of the sensor caused by the bias or drift is larger than e_{I} and e_{O}, the abnormal measurement can be detected as early as possible. The existence of e_{I} and e_{O} is reasonable given that the slight deviation is inevitable due to manufacturing and calibration imperfections, noise, disturbance, etc. Even in some cases, a slight deviation may be acceptable. In practice, e_{I} and e_{O} should be pre-set according to the prior knowledge about the sensors (i.e., the key performance parameters of sensors, such as the temperature drift) or the specific requirements of each application.

There are two important assumptions for the method:

- Model residuals under the fault-free condition are subject to a normal distribution, which will be validated in Section 4.2 and Section 5.2.
- Only one sensor may be faulty during the running of the vehicle. From the engineering experience, the assumption is reasonable because the probability of two sensors being malfunctioned simultaneously is really low, if not impossible.

Faults in the pseudo input and pseudo output sensors will have a different effect on the model residuals. Therefore, the two cases will be discussed separately.

Case 1. The mean-related faults occur in the pseudo input sensor, then the measurement denoted by y'_{I} becomes

$$y'_{\mathrm{I}}(k) = y_{\mathrm{I}}(k) + e_{\mathrm{I}}, \quad e_{\mathrm{I}} = \text{const.} \tag{23}$$

The model residuals affected by the abnormal pseudo input will become

$$\begin{aligned} r(k) &= y_{\mathrm{O}}(k) - \hat{y}_{\mathrm{O}}(k|\hat{\theta}) \\ &= y_{\mathrm{O}}(k) - \sum_{i=-d}^{m} \hat{h}_i y'_{\mathrm{I}}(k-i) \\ &= y_{\mathrm{O}}(k) - \sum_{i=-d}^{m} \hat{h}_i \big(y_{\mathrm{I}}(k-i) + e_{\mathrm{I}}\big) \\ &= -\sum_{i=-d}^{m} \hat{h}_i e_{\mathrm{I}} + \xi(k). \end{aligned} \tag{24}$$

Note that the model parameters $\hat{h}_i,\ i = -d, \dots, m$, are constants after the noncausal FIR models are estimated using historical data. Thus, the mean and variance of the residuals are given by

$$\begin{aligned} &\mathbb{E}[r(k)] = -\sum_{i=-d}^{m} \hat{h}_i e_{\mathrm{I}}, \\ &\mathbb{V}[r(k)] = \sigma^2, \end{aligned} \tag{25}$$

respectively, i.e., the distribution of the model residuals is $\mathcal{N}(-\sum_{i=-d}^{m} \hat{h}_i e_{\mathrm{I}}, \sigma^2)$.

Case 2. The mean-related faults occur in the pseudo output sensor, then the measurement denoted by y'_{O} becomes

$$y'_{\mathrm{O}}(k) = y_{\mathrm{O}}(k) + e_{\mathrm{O}}, \quad e_{\mathrm{O}} = \text{const.} \tag{26}$$

The model residuals affected by the abnormal pseudo output will become

$$\begin{aligned} r(k) &= y'_{\mathrm{O}}(k) - \hat{y}_{\mathrm{O}}(k|\hat{\theta}) \\ &= \big(y_{\mathrm{O}}(k) + e_{\mathrm{O}}\big) - \hat{y}_{\mathrm{O}}(k|\hat{\theta}) \\ &= \xi(k) + e_{\mathrm{O}}. \end{aligned} \tag{27}$$

The mean and variance of the residuals are given by

$$\begin{aligned} &\mathbb{E}[r(k)] = e_{\mathrm{O}}, \\ &\mathbb{V}[r(k)] = \sigma^2, \end{aligned} \tag{28}$$

respectively, i.e., the distribution of the model residuals will become $\mathcal{N}(e_{\mathrm{O}}, \sigma^2)$.

In Case 1 and 2, the mean of the model residuals, denoted by μ, will no longer be zero for the existence of sensor faults. It will become either $-\sum_{i=-d}^{m} \hat{h}_i e_{\mathrm{I}}$ or e_{O}, and thus we conclude that e_{I} and e_{O} have different fault sensitivity, which is used to indicate how sensor faults affect the mean of the model residuals in the context of mean-related fault detection. The conclusion is that for the pseudo output sensor, the sensor faults affect the mean of model residuals in a fixed way; in contrast, for the pseudo input sensor, the fault sensitivity depends on the estimated model parameters.

Once μ is determined, two hypotheses which the NP test is based on can be obtained: H_0 (the fault-free residuals) and H_1 (the abnormal residuals)

$$\begin{aligned} H_0 &: \ r_1, \dots, r_n \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \\ H_1 &: \ r_1, \dots, r_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2), \end{aligned} \tag{29}$$

where $r_i,\ i = 1, \dots, n$, denote n independent samples of model residuals. For the hypothesis testing problem, there are two important metrics: detection rate and false alarm rate. The detection rate denoted by P_D is defined as the probability of deciding H_1 when H_1 is true, while the false alarm rate denoted by P_{FA} is defined as the probability of deciding H_1 when H_0 is true.

The NP lemma states that the likelihood ratio

$$\frac{P(r_1, \dots, r_n|H_1)}{P(r_1, \dots, r_n|H_0)} \tag{30}$$

is the optimal test statistic in the sense that the detection rate P_D is maximized given a false alarm rate P_{FA} [41]. And considering the distribution of r_i in Eq. (29), the NP test can be expressed as

$$\frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(\frac{-1}{2\sigma^2}\left(r_i - \mu\right)^2)}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(\frac{-1}{2\sigma^2} r_i^2)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma, \tag{31}$$

where γ is a positive threshold for this test statistic. Eq. (31) means that if the left-hand side of the inequality is greater than the right-hand side, H_1 is decided. Otherwise, H_0 is decided.

The inequalities are still preserved when a monotonic transformation is applied to both sides, so the expression can be simplified by taking the logarithm. The log-likelihood ratio test is

$$-\frac{1}{2\sigma^2}(-2\mu \sum_{i=1}^{n} r_i + n\mu^2) \underset{H_0}{\overset{H_1}{\gtrless}} \ln\gamma. \tag{32}$$

If $\mu > 0$ (the analysis in the case of $\mu < 0$ is similar to that of $\mu > 0$, and thus it is omitted for the sake of simplicity), Eq. (32) is equivalent to

$$\sum_{i=1}^{n} r_i \underset{H_0}{\overset{H_1}{\gtrless}} \nu, \tag{33}$$

where

$$\nu = \frac{\sigma^2}{\mu} \ln\gamma + \frac{n\mu}{2}. \tag{34}$$

After simplifying the NP test in Eq. (31) into Eq. (33), the optimal test statistic denoted by T can be determined as $\Sigma_{i=1}^{n} r_i$ and ν is its threshold.

The threshold ν can be obtained if a false alarm rate P_{FA} is specified, which provides an advantage to automatically determine the threshold in the application of sensor fault detection. In detail, the test statistic T is also Gaussian since it is the sum of n Gaussian variables. Thus, the hypotheses in terms of T can be expressed as

$$\begin{aligned} H_0 &: T \sim \mathcal{N}(0, n\sigma^2) \\ H_1 &: T \sim \mathcal{N}(n\mu, n\sigma^2). \end{aligned} \tag{35}$$

The false alarm rate P_{FA} can be calculated according to its definition

$$P_{FA} = \int_{\nu}^{\infty} \frac{1}{\sqrt{2n\pi\sigma^2}} \exp(-\frac{T^2}{2n\sigma^2}) dT = Q(\frac{\upsilon}{\sqrt{n\sigma^2}}). \tag{36}$$

Q function is the tail probability of the standard normal distribution, i.e., $Q(z) = \int_{u \geq z} \frac{1}{\sqrt{2\pi}} \exp(\frac{-u^2}{2}) du$ where $z = \frac{\upsilon}{\sqrt{n\sigma^2}}$ in Eq. (36). Since Q function is monotonically decreasing, it has an inverse function $Q^{-1}(\cdot)$. Thus, the threshold ν for the test statistic T can be solved according to Eq. (36) if the false alarm rate P_{FA} is given

$$\nu = \sqrt{n\sigma^2} Q^{-1}(P_{FA}). \tag{37}$$

In practice, the NP test in Eq. (33) and its threshold in Eq. (37) are used to detect the mean-related sensor fault.

Furthermore, the detection rate P_D can be obtained according to its definition

$$P_D = \int_{\nu}^{\infty} \frac{1}{\sqrt{2n\pi\sigma^2}} \exp(-\frac{(T - n\mu)^2}{2n\sigma^2}) dT = Q(\frac{\nu - n\mu}{\sqrt{n\sigma^2}}). \tag{38}$$

Substituting Eq. (37) in Eq. (38), P_D is expressed as a function of P_{FA}

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{\frac{n\mu^2}{\sigma^2}}\right). \tag{39}$$

Remark 5. Note that the above-mentioned detection rate P_D and false alarm rate P_{FA} are theoretical values based on the Gaussian distribution functions. They are different from the empirical values achieved from the validation data set, where P_D is the ratio of the number of successful detection to the total data number and P_{FA} is the ratio of the number of false alarm to the total data number.

Remark 6. Eq. (39) shows that the detection probability increases as μ^2/σ^2 and n increase. There are two ways to increase μ^2/σ^2: (1) increasing the fault magnitude to increase μ; (2) improving the model accuracy to reduce σ^2. As for the hyper-parameter n, although more samples can improve the detection performance theoretically, it could lead to a longer delay time for online detection. Thus, n should be selected carefully based on the trade-off between the detection performance and the delay time in practice.

3.2. Variance-related fault detection

With a slight abuse of notation, e_{I} and e_{O} are used to denote the maximum acceptable sensor precision change of the pseudo input sensor and output sensor, respectively. And once the sensor precision degradation is more significant than e_{I} or e_{O}, the abnormal measurement should be detected. Similarly to the last subsection, e_{I} and e_{O} should be pre-set according to the prior knowledge about the sensors or the specific requirements of each application. Given that our assumption is that only one sensor fails during the same time period, faults in the pseudo input and pseudo output sensors will be discussed separately.

Case 1. The variance-related faults occur in the pseudo input sensor, then the measurement becomes

$$y'_{\mathrm{I}}(k) = y_{\mathrm{I}}(k) + e_{\mathrm{I}}, \quad e_{\mathrm{I}} \sim \mathcal{N}(0, \sigma^2_{e_{\mathrm{I}}}). \tag{40}$$

The model residuals affected by the abnormal pseudo input can be expressed as Eq. (24). Thus, the mean and variance of the residuals will become

$$\begin{aligned} &\mathbb{E}[r(k)] = 0, \\ &\mathbb{V}[r(k)] = \sum_{i=-d}^{m} \hat{h}_i^2 \sigma^2_{e_{\mathrm{I}}} + \sigma^2, \end{aligned} \tag{41}$$

respectively, i.e., the distribution of the model residuals is $\mathcal{N}(0, \sum_{i=-d}^{m} \hat{h}_i^2 \sigma^2_{e_{\mathrm{I}}} + \sigma^2)$.

Case 2. The variance-related faults occur in the pseudo output sensor, then the measurement becomes

$$y'_{\mathrm{O}} = y_{\mathrm{O}} + e_{\mathrm{O}}, \quad e_{\mathrm{O}} \sim \mathcal{N}(0, \sigma^2_{e_{\mathrm{O}}}). \tag{42}$$

The model residuals affected by the abnormal pseudo output can be expressed as Eq. (27). Thus, the mean and variance of the residuals will become

$$\begin{aligned} &\mathbb{E}[r(k)] = 0, \\ &\mathbb{V}[r(k)] = \sigma^2_{e_{\mathrm{O}}} + \sigma^2, \end{aligned} \tag{43}$$

respectively, i.e., the distribution of the model residuals is $\mathcal{N}(0, \sigma^2_{e_{\mathrm{O}}} + \sigma^2)$.

In Case 1 and 2, the variance of the model residuals, denoted by σ_1^2, will increase to $\sum_{i=-d}^{m} \hat{h}_i^2 \sigma^2_{e_{\mathrm{I}}} + \sigma^2$ and $\sigma^2_{e_{\mathrm{O}}} + \sigma^2$, respectively, due to the variance-related faults. The fault sensitivity in the context of variance-related fault detection indicates how sensor faults affect the variance of the model residuals. Thus, we can conclude that the fault sensitivity of the pseudo output sensor is fixed, while the fault sensitivity of the pseudo input sensor is determined by the estimated model parameters. This is similar to the conclusion in the context of mean-related fault detection.

Once σ_1^2 is determined, two hypotheses which the NP test is based on can be obtained: H_0 (the fault-free residuals) and H_1 (the abnormal residuals)

$$\begin{aligned} &H_0 : \ r_1, \dots, r_n \overset{iid}{\sim} \mathcal{N}(0, \sigma_0^2) \\ &H_1 : \ r_1, \dots, r_n \overset{iid}{\sim} \mathcal{N}(0, \sigma_1^2). \end{aligned} \tag{44}$$

The log-likelihood ratio test according to the NP lemma is

$$\frac{n}{2} \ln\left(\frac{\sigma_0^2}{\sigma_1^2}\right) + \left(\frac{1}{2\sigma_0^2} - \frac{1}{2\sigma_1^2}\right) \sum_{i=1}^{n} r_i^2 \underset{H_0}{\overset{H_1}{\gtrless}} \ln \gamma. \tag{45}$$

Simplifying Eq. (45) yields the NP test used in practice

$$\sum_{i=1}^{n} r_i^2 \underset{H_0}{\overset{H_1}{\gtrless}} \nu, \tag{46}$$

where the threshold

$$\nu = 2\left(\frac{\sigma_0^2 \sigma_1^2}{\sigma_1^2 - \sigma_0^2}\right)\left(\ln \gamma + n \ln\left(\frac{\sigma_1}{\sigma_0}\right)\right). \tag{47}$$

Thus, the optimal test statistic based on the NP lemma is $T \triangleq \sum_{i=1}^{n} r_i^2$.

To determine the threshold ν based on the specified false alarm rate P_{FA}, the hypothesis testing is rewritten with the optimal test statistic T to yield

$$\begin{aligned} &H_0 : \sum_{i=1}^{n} \frac{r_i^2}{\sigma_0^2} \sim \chi_n^2 \\ &H_1 : \sum_{i=1}^{n} \frac{r_i^2}{\sigma_1^2} \sim \chi_n^2, \end{aligned} \tag{48}$$

where χ_n^2 denotes the chi-square distribution with n degrees of freedom, which is obtained as the sum of the squares of n independent normally-distributed variables.

According to Eqs. (46) and (48), the false alarm rate P_{FA} is just the probability that a χ_n^2 random variable exceeds ν/σ_0^2, and the detection rate is the probability that a χ_n^2 random variable exceeds ν/σ_1^2, i.e.,

$$P_{FA} = P\left(\frac{T}{\sigma_0^2} > \frac{\nu}{\sigma_0^2}\right), \tag{49}$$

$$P_D = P\left(\frac{T}{\sigma_1^2} > \frac{\nu}{\sigma_1^2}\right), \tag{50}$$

where $P(\cdot)$ denotes the probability of the event in the brackets. If the false alarm rate P_{FA} is given in the application, the threshold ν can be determined by inversely solving Eq. (49). Furthermore, the corresponding theoretical detection rate P_D can be calculated according to Eq. (50).

3.3. Methodology summary

Based on Sections 3.1 and 3.2, the proposed model-based sensor fault detection method includes the following steps:

(i) estimating the transmissibilities with the noncausal FIR models using the historical data;
(ii) achieving H_0, i.e., the Gaussian distribution of model residuals under the fault-free conditions;
(iii) pre-setting e_{I} or e_{O} to obtain H_1, i.e., the distribution of the abnormal model residuals when the maximum acceptable faults occur;
(iv) designing the NP detector, i.e., determining the sample size n as discussed in Remark 6, calculating the threshold based on the pre-set P_{FA}, and then obtaining the theoretical P_D;
(v) assessing the detection performance based on the theoretical values and the empirical values (which are achieved by applying the detector to the validation set) of P_{FA} and P_D. If the detector performs poorly, we can conclude that the detectability of the pre-set e_{I} or e_{O} is low in the context of the NP test.

Remark 7. In this work, the assumption is that e_{I} and e_{O} are pre-set according to prior knowledge. Then the detector is designed, and its detection performance will be evaluated by calculating the false alarm probability and the detection probability. Note that if the pre-set e_{I} and e_{O} are very slight, it may be beyond the fault detection capacity of the detector, i.e., the false alarm probability is higher, or the detection probability is lower than expected. As will be shown in Section 5.3.1, a large e_{I} is set because a smaller e_{I} will make the designed detector perform poorly. Thus, to achieve the desired detection probability and

maintain an acceptable false alarm probability, the question naturally arises as to the minimum magnitude of bias or precision degradation of pseudo input and output sensors that can be detected. More discussion about the question can be found in [59].

As a PIML method, the physical information is embedded within the method in terms of the following two aspects:

(i) The noncausal FIR model structure is determined based on our understanding of the physics systems. It is noticed that in real scenarios, the vehicle systems have unknown system excitation, which varies with the change of the human driver behavior, vehicle velocity, road profile, etc., and what can be measured in the engineering practice are the system response signals [25]. In this case, the transmissibilities and noncausal FIR models are used to approximate the relationship between two sensor signals.

(ii) The pseudo input and output are selected based on physical insights. In this work, the pseudo input and output are the suspension displacement and the acceleration of an unsprung mass, respectively. From physical intuition, the specified pseudo input and output can reflect the vertical dynamics of the vehicles. Also, the mathematical model of a quarter-car suspension system in Section 4.1 does provide us the confidence that the selected pseudo input and output are physically related, and thus meaningful.

To conclude, the physical information helps the determination of the model structure as well as the selection of the model input and output. Thus, the proposed method has advantages in the following aspects:

(i) One major challenge in the PIML model construction is how to choose one among a range of different models that can be used for the specific problem [19]. The proposed method derives the FIR model structure from the physical models, avoiding subjectively choosing one final model among a collection of candidate models.

(ii) Currently, PIML models for online applications have not been extensively investigated because real-time inference requires an algorithm of low computational complexity and fast response [19]. In the proposed method, both the FIR model and the NP test are computationally efficient, making the method applicable and promising to online sensor fault detection.

Now we discuss the applicability and possible future work of the proposed method:

(i) There are fault types that the proposed method cannot handle, e.g., data-loss fault, spike (i.e., intermittent high-amplitude values), and non-Gaussian precision degradation [11]. Due to their random nature, it could be difficult to detect and deal with them in diagnostics algorithms. To be more specific, since the generated model residuals under these faulty cases cannot be approximated by Gaussian distributions, the NP test fails to distinguish the abnormal model residuals from the fault-free ones.

(ii) There is a shortcoming of the proposed method. The method can detect sensor fault, either from the pseudo input sensor or the pseudo output sensor. However, they are not distinguished because faults in the input and output sensors affect model residuals in a similar way. Thus, additional tests need to be conducted to determine exactly which sensor fails. Due to the space limitation, this problem will be addressed in detail in our future work.

4. Numerical validation

In this section, a numerical example developed from a quarter-car suspension system will be used to verify the effectiveness of the proposed sensor fault detection method in linear dynamic systems with unknown inputs.

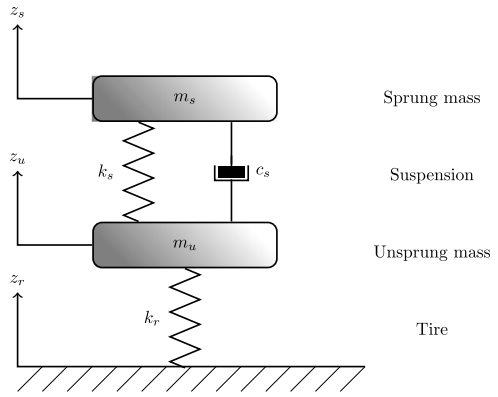


Fig. 2. A quarter-car suspension system.

Table 2
The physical parameters of the suspension system.

Parameters	Units	Numerical values
m_s	kg	300
m_u	kg	80
k_s	N m^{-1}	3.0×10^4
k_r	N m^{-1}	2.2×10^5
c_s	N s m^{-1}	5.0×10^3

4.1. Numerical setup

Fig. 2 shows a quarter-car suspension system, where k_s and c_s are spring and damper coefficients respectively, and k_r is the tire stiffness. The displacements of the sprung mass and unsprung mass are denoted by z_s and z_u, respectively. The input of the system is the displacement denoted by z_r. The system has two outputs: y_O denotes the acceleration of m_u, and y_I denotes the relative displacement between the unsprung mass m_u and sprung mass m_s. Then the state-space form of the quarter-car suspension system in continuous time can be represented by

$$\begin{aligned} \dot{x}(t) &= A_c x(t) + b_c z_r(t) \\ y(t) &= C x(t) + d z_r(t), \end{aligned} \tag{51}$$

where $x(t) = \left[z_s(t)\ \dot{z}_s(t)\ z_u(t)\ \dot{z}_u(t)\right]^{\mathrm{T}}$ are system states, and $y(t) = \left[y_O(t)\ y_I(t)\right]^{\mathrm{T}}$ are measurable system outputs,

$$\begin{aligned} A_c &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ -\frac{k_s}{m_s} & -\frac{c_s}{m_s} & \frac{k_s}{m_s} & \frac{c_s}{m_s} \\ 0 & 0 & 0 & 1 \\ \frac{k_s}{m_u} & \frac{c_s}{m_u} & -\frac{k_s+k_r}{m_u} & -\frac{c_s}{m_u} \end{bmatrix}, \quad b_c = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{k_r}{m_u} \end{bmatrix}, \\ C &= \begin{bmatrix} \frac{k_s}{m_u} & \frac{c_s}{m_u} & -\frac{k_s+k_r}{m_u} & -\frac{c_s}{m_u} \\ 1 & 0 & -1 & 0 \end{bmatrix}, \quad d = \begin{bmatrix} \frac{k_r}{m_u} \\ 0 \end{bmatrix}. \end{aligned} \tag{52}$$

The noncausal FIR model based on the transmissibility operator that we would like to construct is from y_I to y_O. In the experiment, the physical parameters of the suspension system are listed in Table 2.

The system is discretized using a zero-order hold, then we have

$$\begin{aligned} x(t+1) &= Ax(t) + bz_r(t) \\ y(t) &= Cx(t) + dz_r(t), \end{aligned} \tag{53}$$

where $A = \exp(A_c \Delta T), b = A_c^{-1}(A - \mathrm{I})b_c$, and the sampling time ΔT is 0.01 s in this example. The system input z_r is a realization of a zero-mean white Gaussian random process with 0.01 m^2 variance. The measurements y_I and y_O are added by the zero-mean white Gaussian

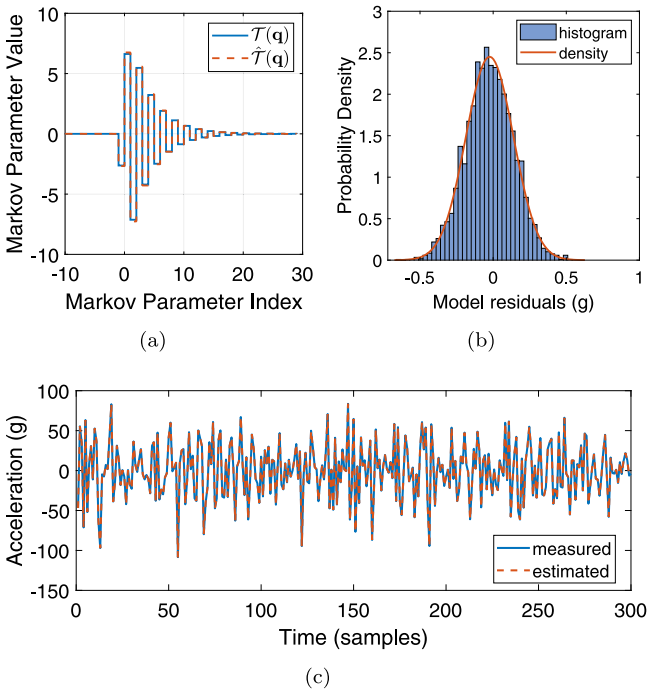


Fig. 3. The performance of noncausal FIR models: (a) the Markov parameters of the transmissibility $\mathcal{T}(\mathbf{q})$ and its estimate $\hat{\mathcal{T}}(\mathbf{q})$, and (b) the histogram of the model residuals and the approximated density, and (c) the measured output and the estimated output.

noise with the SNR being 100 dB. Then, this dynamic system generates the training and validation sets, which contain 1000 and 4000 data points, respectively. Here the unit of the displacement is mm, and the unit of acceleration is g.

4.2. Residual generation with noncausal FIR models

As introduced in Section 2, We use ridge regression with a noncausal FIR model to estimate the transmissibility $\mathcal{T}$ from y_{I} to y_{O}. The hyper-parameter C_{lim} presented in Eq. (21) for ridge regression is set as 1e6. The order of the causal part of the noncausal FIR model is 30, and the order of the noncausal part is 10. Fig. 3(a) shows a plot of the model parameters of $\mathcal{T}(\mathbf{q})$ and its estimate $\hat{\mathcal{T}}(\mathbf{q})$. It is clear that the transmissibility has an obvious noncausal component. The sum and the square sum of the Markov parameters are −0.05 and 182.03 respectively, which will be used for the fault sensitivity analysis later.

Then, the identified noncausal FIR models are used to estimate the pseudo output. Fig. 3(c) is a plot of the measured pseudo output and the estimated pseudo output with 300 data points, which shows good estimation performance of the identified models. In Fig. 3(b), the normal distribution $\mathcal{N}(0, 0.026)$ density fits the histogram of residuals very well, and hence our assumption that the residuals are approximately normally distributed (as mentioned in Section 3) is verified.

4.3. Residual evaluation with Neyman–Pearson test

In this part, the scenario where only a single fault occurs in the pseudo input or output sensor is considered. The NP test will evaluate the model residuals and then detect mean-related faults and variance-related faults separately. Receiver Operating Characteristic (ROC) curves will illustrate the overall performance at all possible values of the test statistic threshold. The ROC space is defined as a two-dimensional space whose horizontal and vertical axes correspond to the false alarm rate and the detection rate, respectively. The area under the ROC curve is a widely used metric to evaluate a classification method, and the larger area means the better detection performance [60].

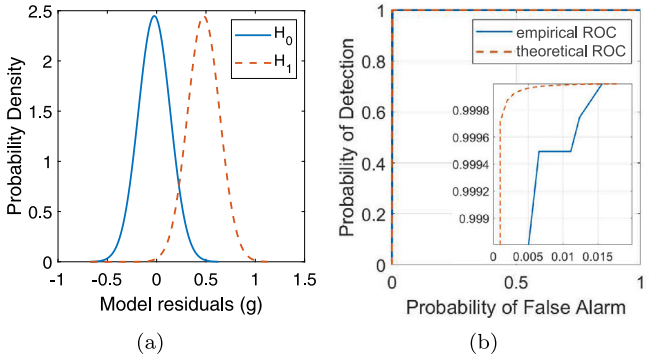


Fig. 4. (a) The two hypotheses, and (b) the fault detection performance of the NP detector when n is 5.

4.3.1. Mean-related faults

From Section 3, the mean of model residuals μ will become $-\sum_{i=-d}^{m}\hat{h}_i e_{\mathrm{I}} = -0.05e_{\mathrm{I}}$ and e_{O} when maximum acceptable deviation e_{I} and e_{O} occur in the pseudo input and output sensors, respectively. In this numerical example, we set e_{I} as 10 mm (a deviation about 5% of the corresponding maximum sensor value) and e_{O} as 0.5 g (about 0.4% of the corresponding maximum sensor value), and then μ is 0.5. Thus, two hypotheses H_0 (the fault-free residuals, whose distribution is obtained in Section 4.2) and H_1 (the abnormal residuals) as shown in Fig. 4(a) will be

$$\begin{aligned} H_0 &: \ r_1, \ldots, r_n \overset{iid}{\sim} \mathcal{N}(0, 0.026) \\ H_1 &: \ r_1, \ldots, r_n \overset{iid}{\sim} \mathcal{N}(0.5, 0.026). \end{aligned} \tag{54}$$

The NP lemma indicates that the optimal test statistic is $T = \Sigma_{i=1}^{n} r_i$, where n is a hyper-parameter denoting the number of sampled model residuals and set as 5 here. According to Section 3, as P_{FA} varies from 0 to 1, the corresponding threshold and theoretical P_D can be calculated, and thus the theoretical ROC curve of this detector can be determined as shown in Fig. 4(b). The empirical ROC curve is used to validate the performance of the NP detector in the simulated data set, whose pseudo output data are injected with 0.5 g bias. By changing the threshold of T gradually, a list of values of P_{FA} and P_D can be obtained to draw the empirical ROC curve. Both the theoretical and empirical ROC curves show that the detection performance is satisfactory because the NP detector can have the high P_D and low P_{FA} simultaneously. Also, the theoretical ROC curve is slightly better than the empirical one due to the finite number of samples of the data set.

To show more details about the NP detector, we inject the bias into the fault-free output data starting from the 2000th data sample, after which the model residuals have an obvious bias immediately, as presented in the top subfigure of Fig. 5. As the sum of five consecutive values of model residuals, the NP test statistic is observed to rapidly exceed the threshold after the bias is added to the measurements. The threshold is calculated to be 1.126 when the given P_{FA} is 0.002, as the solid red line shown in the middle subfigure of Fig. 5. The detection result in the bottom subfigure of Fig. 5 shows that the detector has a quick response to the bias fault of the pseudo output sensor.

4.3.2. Variance-related faults

From Section 3, the variance of model residuals σ_1^2 will increase to $\sum_{i=-d}^{m}\hat{h}_i^2\sigma_{e_{\mathrm{I}}}^2 + \sigma^2$ and $\sigma_{e_{\mathrm{O}}}^2 + \sigma^2$ due to the variance-related faults of the pseudo input and output sensors. Here we set $e_{\mathrm{I}} \sim \mathcal{N}(0, 0.5/182.03)$ (SNR = 58.84 dB) and $e_{\mathrm{O}} \sim \mathcal{N}(0, 0.5)$ (SNR = 33.87 dB), and then variance increment $\sigma_1^2 - \sigma^2$ is 0.5. Thus, two hypotheses H_0 (the fault-free residuals, whose distribution is obtained in Section 4.2) and H_1 (the abnormal residuals) as shown in Fig. 6(a) will be

$$\begin{aligned} H_0 &: \ r_1, \ldots, r_n \overset{iid}{\sim} \mathcal{N}(0, 0.026) \\ H_1 &: \ r_1, \ldots, r_n \overset{iid}{\sim} \mathcal{N}(0, 0.526). \end{aligned} \tag{55}$$

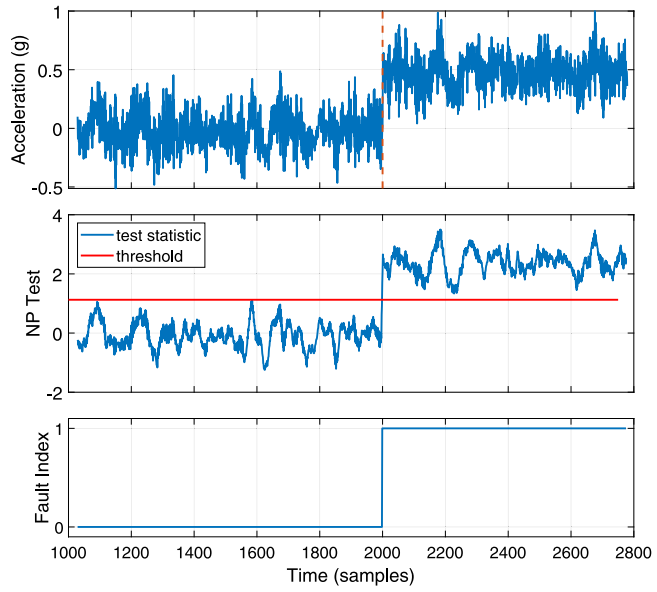


Fig. 5. Top: the model residuals when the bias is injected at the 2000th data sample; Middle: the NP test statistic and the threshold; Bottom: the fault detection result, 0: fault-free, 1: abnormal.

The NP lemma indicates that the optimal test statistic is $T = \sum_{i=1}^{n} r_i^2$, i.e., the square sum of n samples. Here we set the hyper-parameters n as 10. Similarly, to draw the theoretical ROC curve of this detector, P_{FA} is varied from 0 to 1, and the corresponding threshold and theoretical P_D is calculated. Then, the empirical ROC curve can be obtained based on the validation set whose pseudo output data are injected with the precision degradation faults $e_O \sim \mathcal{N}(0, 0.5)$. The ROC curves in Fig. 6(b) show that although the theoretical detection performance could be perfect, the empirical performance is slightly worse because P_{FA} cannot be less than 0.02 when the P_D is acceptable.

To show more details about the NP detector, we inject the precision degradation into the fault-free output data starting from the 2000th data sample, where the variance of the model residuals increases immediately, as shown in the top subfigure of Fig. 7. As the square sum of ten consecutive values of model residuals, the NP test statistic is observed to rapidly exceed the threshold after the faults are added to the measurements. The threshold is calculated as 1.126 when the given P_{FA} is 0.002, as the solid red line shown in the middle subfigure of Fig. 7. The detection result in the bottom subfigure of Fig. 7 shows that the detector has a quick response to the precision degradation faults of the pseudo output sensor.

5. Experimental validation

In this section, the vibration signals collected from a suspension system of a full vehicle system will be used to demonstrate the effectiveness of the proposed fault detection method. At the same time, the Support Vector Machine (SVM) method, as a widely used machine-learning algorithm, is also used to detect sensor faults based on two features: y_I and y_O [26,61].

5.1. Experimental setup

Fig. 8(a) shows the picture of the test vehicle while Fig. 8(b) is the picture of the accelerometer installed in the test vehicle. To show the location of sensors clearly, Fig. 8(c) is a sketch of a quarter-car suspension system. The accelerometer installed at position A will measure the accelerations y_O and the displacement sensor installed between position A and position B will measure the suspension displacement y_I. In the

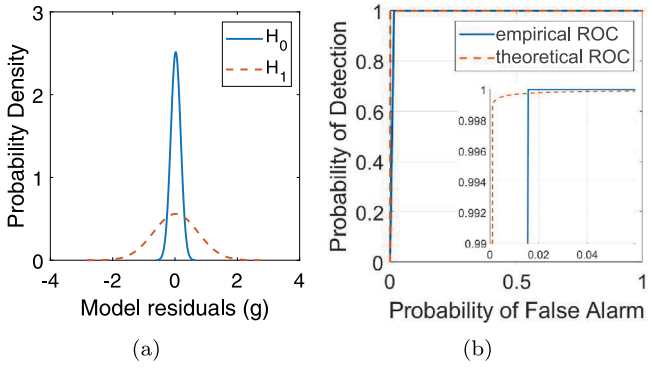


Fig. 6. (a) The two hypotheses, and (b) the fault detection performance of the NP detector when n is 10.

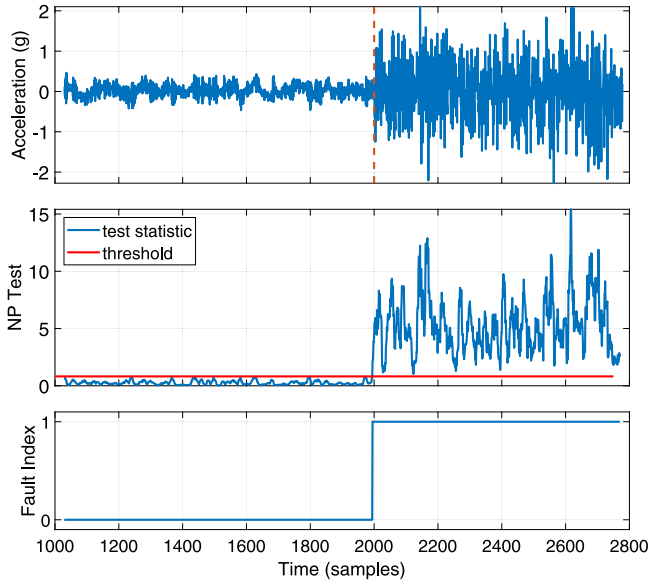


Fig. 7. Top: the model residuals when the precision degradation fault is injected in the 2000th data samples; Middle: the NP test statistic and the threshold; Bottom: the fault detection result, 0: fault-free, 1: abnormal.

experiment, the vehicle runs on the public roads to collect the vibration signals (over 4×10^4 data samples with the sampling frequency 512 Hz are collected).

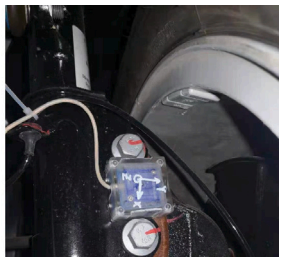
The data set is split into two parts: the training set (1000 data samples) and the validation set (4.2×10^4 data samples). The training set is used for estimating the transmissibility $\mathcal{T}$ from y_I to y_O with a noncausal FIR model. The validation set is used for evaluating the performance of the constructed noncausal FIR model and the proposed fault detector. Note that the sensor faults to be detected are simulated according to the mathematical expressions in Table 1 and injected into the validation set.

5.2. Residual generation with noncausal FIR models

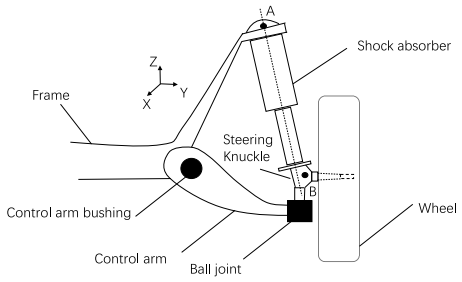
We use ridge regression with a noncausal FIR model to estimate the transmissibility $\mathcal{T}$ from y_I to y_O, with the hyper-parameter C_{lim} being set as 1e6. The order of the causal part of the noncausal FIR model is 20 and the order of the noncausal part is 20 too. Fig. 9(a) shows a plot of the identified parameters of $\mathcal{T}$. It is clear that the transmissibility has some noncausal components. The sum and the square sum of the Markov parameters are −0.0089 and 6.6553, respectively.

Fig. 9(c) is a plot of the measured pseudo output and the estimated pseudo output with 300 data points, which shows good estimation

(a) The test vehicle



(b) The accelerometer installed close to the wheel center



(c) The sketch of a quarter-car suspension system

Fig. 8. The locations of the sensors in the test vehicle.

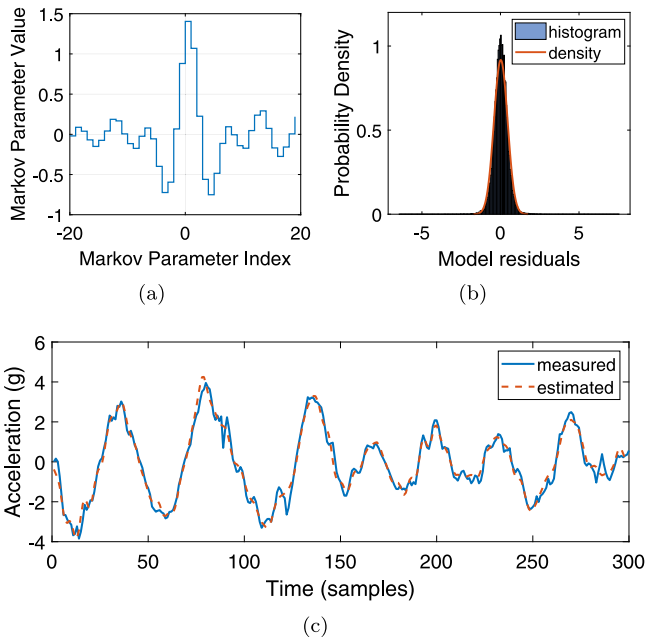


Fig. 9. The performance of noncausal FIR models: (a) the Markov parameters of the estimated transmissibility $\hat{T}(\mathbf{q})$, and (b) the histogram of the model residuals and the approximated density, and (c) the measured output and the estimated output.

performance of the estimated transmissibility. Fig. 9(b) show the histogram of the model residuals, which can be fit well by a Gaussian distribution $\mathcal{N}(0, 0.189)$.

5.3. Residual evaluation with Neyman–Pearson test

In this part, the NP test will be used to evaluate the model residuals and then detect mean-related faults and variance-related faults separately. It is assumed that only a single fault may occur in the pseudo input or output sensor.

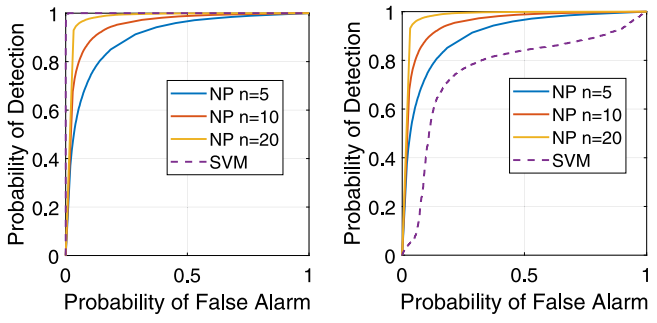


Fig. 10. The ROC curves of NP detectors and SVM when (a) the bias occurs in the pseudo input sensor, (b) the bias occurs in the pseudo output sensor.

5.3.1. Mean-related faults

Here we set e_{I} to be 56.18 mm (a deviation about 98% of the maximum sensor value), since smaller e_{I} will be beyond the fault detection capacity as discussed in Remark 7. And e_{O} is set to be 0.5 g (about 4% of the maximum sensor value). Then the mean of model residuals μ will be 0.5 since $\mu = -0.0089e_{\mathrm{I}}$ and $\mu = e_{\mathrm{O}}$. Two hypotheses H_0 (the fault-free residuals, whose distribution is obtained in Section 5.2) and H_1 (the abnormal residuals) will be

$$\begin{aligned} H_0 &: r_1, \ldots, r_n \overset{iid}{\sim} \mathcal{N}(0, 0.189) \\ H_1 &: r_1, \ldots, r_n \overset{iid}{\sim} \mathcal{N}(0.5, 0.189). \end{aligned} \tag{56}$$

The sample size n influences the detection performance when designing the NP detector and here n is set to be 5, 10 and 20. Faults are injected into the pseudo input and output sensors. The corresponding ROC curves can be obtained: Fig. 10(a) shows the ROC curves when the pseudo input sensor is injected bias 56.18 mm; Fig. 10(b) shows the ROC curves when the pseudo output sensor is injected bias 0.5 g. It shows that more samples to calculate the test statistic will improve the performance of the NP detector, which is consistent with our theoretical analysis in Section 3.

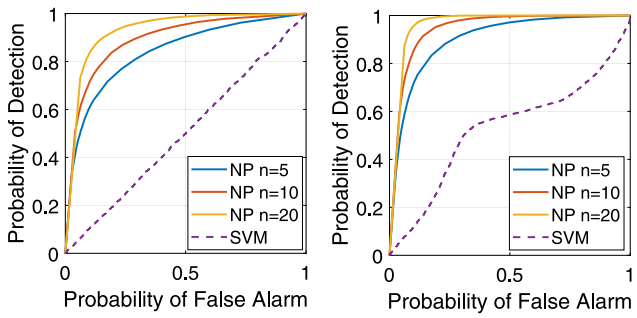
The performance of the SVM classifier is shown as a baseline to compare with that of the NP detector. According to Fig. 10, the SVM has better performance for the pseudo input sensor bias than the NP detector, whereas, for the pseudo output sensor bias, the SVM has worse performance than the NP detector. The result can be explained by the fault sensitivity as discussed in Section 3. The mean of model residuals caused by the pseudo input sensor bias e_{I} is expressed as $\sum_{i=-d}^{m} \hat{h}_i e_{\mathrm{I}} = -0.0089 e_{\mathrm{I}}$. The coefficient -0.0089 shrinks the effect of e_{I} on the mean of model residuals considerably. Thus, to achieve the satisfying performance of the NP detector, e_{I} is set large, which enables SVM to train a classifier with better performance. In comparison, the fault sensitivity is relatively higher for the pseudo output sensor. The slight bias, which is hard for SVM to classify, can easily be detected with the NP detector.

5.3.2. Variance-related faults

Setting $e_{\mathrm{I}} \sim \mathcal{N}(0, 0.5/6.6553)$ (SNR = 35.33 dB) and $e_{\mathrm{O}} \sim \mathcal{N}(0, 0.5)$ (SNR = 9.45 dB), the variance increment of model residuals will be 0.5 since $\sigma_1^2 - \sigma^2 = \sum_{i=-d}^{m} \hat{h}_i^2 \sigma_{e_{\mathrm{I}}}^2 = 6.6553 \sigma_{e_{\mathrm{I}}}^2$ and $\sigma_1^2 - \sigma^2 = \sigma_{e_{\mathrm{O}}}^2$. Two hypotheses H_0 (the fault-free residuals, whose distribution is obtained in Section 5.2) and H_1 (the abnormal residuals) are

$$\begin{aligned} H_0 &: r_1, \ldots, r_n \overset{iid}{\sim} \mathcal{N}(0, 0.189) \\ H_1 &: r_1, \ldots, r_n \overset{iid}{\sim} \mathcal{N}(0, 0.689). \end{aligned} \tag{57}$$

The sample size n is set 5, 10, and 20 in order to investigate its impact on the detection performance when the pseudo input sensor and output sensor are injected with the faults, respectively: Fig. 11(a)



(a) The pseudo input sensor fault occurs. (b) The pseudo output sensor fault occurs.

Fig. 11. The ROC curves of NP detectors and SVM when (a) the precision degradation occurs in the pseudo input sensor, (b) the precision degradation occurs in the pseudo output sensor.

shows the ROC curves when the pseudo input sensor is injected with the precision degradation $e_{\mathrm{I}} \sim \mathcal{N}(0, 0.5/6.6553)$; Fig. 11(b) shows the ROC curves when the pseudo output sensor is injected with the precision degradation $e_{\mathrm{O}} \sim \mathcal{N}(0, 0.5)$. It is indicated that more samples to calculate the test statistic will improve the detection performance. However, for online detection, the delay time will also increase as n increases.

Fig. 11 also shows that the NP detector has superior performance than SVM since SVM can hardly detect the disturbance on the measurements caused by the variance-related faults. Note that the variance increment of model residuals caused by the pseudo input sensor precision degradation is $\sum_{i=-d}^{m} \hat{h}_i^2 \sigma_{e_{\mathrm{I}}}^2 = 6.6553 \sigma_{e_{\mathrm{I}}}^2$. The coefficient indicates high fault sensitivity and even the slight precision degradation in the pseudo input sensor will be amplified by the noncausal FIR models. Thus, the minor sensor fault in this example can be detected with the high accuracy.

Based on the above discussion, two measures can be taken to improve the detection performance of the NP detector and the detectability of e_{I} and e_{O}:

(i) As is discussed in Remark 6, the better detection performance can be achieved by improving the model accuracy to obtain better estimation performance, i.e., decreasing the variance of the model residuals, which is subject to a zero-mean normal distribution;

(ii) For the pseudo input sensor, improving the fault sensitivity has another way. We would like to increase the sum of noncausal FIR model parameters in the case of the mean-related faults and the square sum of the noncausal FIR model parameters in the case of the variance-related faults, to make the mean or variance of the model residuals more sensitive to the sensor faults. However, this way does not work for the pseudo output sensor, whose fault sensitivity is fixed in the proposed method.

6. Conclusion and future works

In this work, the proposed PIML model-based method, combines a transmissibility-based residual generation procedure and a NP-based residual evaluation procedure, to detect sensor faults (i.e., the mean-related and the variance-related faults) in the suspension systems. The fault detectability and sensitivity are defined and analyzed for both the pseudo input sensor and the pseudo output sensor. As a result, the proposed approach offers a generalizable and interpretable solution to the sensor fault detection problems with an automatic manner to determine the threshold when the false alarm rate is given.

A numerical example and a real-world application are used to demonstrate the feasibility and effectiveness of the proposed method using ROC curves as the performance metrics. When compared to the SVM-based sensor fault detection method, the overall performance of the proposed method is better than that of the SVM.

One of the challenges that could be addressed in future works is to estimate transmissibility operators with high fault sensitivity for the pseudo input sensor fault to improve the detection performance. In addition, only mean-related or variance-related faults can be detected and it is difficult to distinguish the pseudo input sensor fault from the pseudo output sensor fault, which are two limitations of the proposed method that could be addressed in our future work.

CRediT authorship contribution statement

Ying Wang: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Conceptualization. **Xueke Zheng:** Writing – review & editing, Project administration, Methodology, Conceptualization. **Le Wang:** Writing – review & editing, Methodology, Conceptualization. **Gavin Lu:** Writing – review & editing, Software, Resources. **Yixing Jia:** Writing – review & editing, Software, Resources. **Kezhi Li:** Writing – review & editing, Validation. **Mian Li:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

The work presented here was supported in part by China and National Natural Science Foundation of China under Grant No. 52275263. Special thanks go to Mr. Yu Qiu from SAIC Motor Ltd. for his support to this work. Such support does not constitute an endorsement by the funding agency of the opinions expressed in the paper.

References

[1] Jeong K, Choi SB, Choi H. Sensor fault detection and isolation using a support vector machine for vehicle suspension systems. IEEE Trans Veh Technol 2020;69(4):3852–63. http://dx.doi.org/10.1109/TVT.2020.2977353.

[2] Pan Y, Sun Y, Li Z, Gardoni P. Machine learning approaches to estimate suspension parameters for performance degradation assessment using accurate dynamic simulations. Reliab Eng Syst Saf 2022;108950. http://dx.doi.org/10.1016/j.ress.2022.108950.

[3] Gandoman FH, Ahmadi A, den Bossche PV, Van Mierlo J, Omar N, Nezhad AE, Mavalizadeh H, Mayet C. Status and future perspectives of reliability assessment for electric vehicles. Reliab Eng Syst Saf 2019;183:1–16. http://dx.doi.org/10.1016/j.ress.2018.11.013.

[4] Zheng X, Cai R, Xiao S, Qiu Y, Zhang J, Li M. Primary-auxiliary model scheduling based estimation of the vertical wheel force in a full vehicle system. 2021, arXiv preprint arXiv:2107.11511.

[5] Elkafafy M, Csurcsia PZ, Cornelis B, Risaliti E, Janssens K. Machine learning and system identification for the estimation of data-driven models: an experimental case study illustrated on a tire-suspension system. In: International conference on noise and vibration engineering 2020. KU Leuven; 2020, p. 3287–301.

[6] Risaliti E, Tamarozzi T, Vermaut M, Cornelis B, Desmet W. Multibody model based estimation of multiple loads and strain field on a vehicle suspension system. Mech Syst Signal Process 2019;123:1–25. http://dx.doi.org/10.1016/j.ymssp.2018.12.024.

[7] Yoo M, Kim T, Yoon JT, Kim Y, Kim S, Youn BD. A resilience measure formulation that considers sensor faults. Reliab Eng Syst Saf 2020;199:106393. http://dx.doi.org/10.1016/j.ress.2019.02.025.

[8] Ji H, He X, Zhou D. Diagnosis of sensor precision degradation using Kullback-Leibler divergence. Can J Chem Eng 2018;96(2):434–43. http://dx.doi.org/10.1002/cjce.22916.

[9] Chen Z, Yang C, Peng T, Dan H, Li C, Gui W. A cumulative canonical correlation analysis-based sensor precision degradation detection method. IEEE Trans Ind Electron 2018;66(8):6321–30. http://dx.doi.org/10.1109/TIE.2018.2873100.

[10] Yang H, Hassan SG, Wang L, Li D. Fault diagnosis method for water quality monitoring and control equipment in aquaculture based on multiple SVM combined with DS evidence theory. Comput Electron Agric 2017;141:96–108. http://dx.doi.org/10.1016/j.compag.2017.05.016.

[11] Saeed U, Jan SU, Lee Y-D, Koo I. Fault diagnosis based on extremely randomized trees in wireless sensor networks. Reliab Eng Syst Saf 2021;205:107284. http://dx.doi.org/10.1016/j.ress.2020.107284.

[12] Samy I, Postlethwaite I, Gu D-W. Survey and application of sensor fault detection and isolation schemes. Control Eng Pract 2011;19(7):658–74. http://dx.doi.org/10.1016/j.conengprac.2011.03.002.

[13] Gao Z, Cecati C, Ding SX. A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches. IEEE Trans Ind Electron 2015;62(6):3757–67. http://dx.doi.org/10.1109/TIE.2015.2417501.

[14] Li C, Luo S, Cole C, Spiryagin M. An overview: modern techniques for railway vehicle on-board health monitoring systems. Veh Syst Dyn 2017;55(7):1045–70.

[15] Aravanis T-C, Sakellariou J, Fassois S. A stochastic functional model based method for random vibration based robust fault detection under variable non–measurable operating conditions with application to railway vehicle suspensions. J Sound Vib 2020;466:115006. http://dx.doi.org/10.1016/j.jsv.2019.115006.

[16] Zhou S, Song G, Sun M, Ren Z. Nonlinear dynamic analysis of a quarter vehicle system with external periodic excitation. Int J Non-Linear Mech 2016;84:82–93.

[17] Ding E, Fennel H, Ding S. Model-based diagnosis of sensor faults for ESP systems. Control Eng Pract 2004;12(7):847–56. http://dx.doi.org/10.1016/j.conengprac.2003.10.009, PC-B02-Process Control IFAC 2002.

[18] Xu Z, Saleh JH. Machine learning for reliability engineering and safety applications: Review of current status and future opportunities. Reliab Eng Syst Saf 2021;211:107530. http://dx.doi.org/10.1016/j.ress.2021.107530.

[19] Xu Y, Kohtz S, Boakye J, Gardoni P, Wang P. Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges. Reliab Eng Syst Saf 2022;108900. http://dx.doi.org/10.1016/j.ress.2022.108900.

[20] Varrier S, Morales-Menendez R, Lozoya-Santos JD-J, Hernandez D, Molina JM, Koenig D. Fault detection in automotive semi-active suspension: experimental results. Tech. rep., SAE Technical Paper; 2013, http://dx.doi.org/10.4271/2013-01-1234.

[21] Börner M, Isermann R, Schmitt M. A sensor and process fault detection system for vehicle suspension systems. Tech. rep., SAE Technical Paper; 2002, http://dx.doi.org/10.4271/2002-01-0135.

[22] Chamseddine A, Noura H. Control and sensor fault tolerance of vehicle active suspension. IEEE Trans Control Syst Technol 2008;16(3):416–33. http://dx.doi.org/10.1109/TCST.2007.908191.

[23] Khalil A, Aljanaideh KF, Rideout G, Al Janaideh M. Fault detection in flexible beams based on output only measurements. In: 2020 American control conference. ACC, IEEE; 2020, p. 5034–9. http://dx.doi.org/10.23919/ACC45564.2020.9147301.

[24] Aravanis T-CI, Sakellariou JS, Fassois SD. On the problem of random vibration based fault detection in railway vehicle suspensions under variable and non-measurable operating conditions. In: The AVT-305 research specialists' meeting on sensing systems for integrated vehicle health management for military vehicles, Athens, Greece. 2018.

[25] Khalil A, Aljanaideh KF, Janaideh MA. On connected autonomous vehicles with unknown human driven vehicles effects using transmissibility operators. IEEE Trans Autom Sci Eng 2022. http://dx.doi.org/10.1109/TASE.2022.3188415.

[26] Wang L, Yu P, Li J, Zhou D, Li J. Suspension system status detection of maglev train based on machine learning using levitation sensors. In: 2017 29th Chinese control and decision conference. CCDC, IEEE; 2017, p. 7579–84. http://dx.doi.org/10.1109/CCDC.2017.7978559.

[27] Capriglione D, Carratu M, Pietrosanto A, Sommella P. Online fault detection of rear stroke suspension sensor in motorcycle. IEEE Trans Instrum Meas 2019;68(5):1362–72. http://dx.doi.org/10.1109/TIM.2019.2905945.

[28] Luo H, Huang M, Zhou Z. A dual-tree complex wavelet enhanced convolutional LSTM neural network for structural health monitoring of automotive suspension. Measurement 2019;137:14–27. http://dx.doi.org/10.1016/j.measurement.2019.01.038.

[29] Willard J, Jia X, Xu S, Steinbach M, Kumar V. Integrating physics-based modeling with machine learning: A survey. 2020, p. 1–34, arXiv preprint arXiv:2003.04919 1 (1).

[30] Kashinath K, Mustafa M, Albert A, Wu J, Jiang C, Esmaeilzadeh S, Azizzadenesheli K, Wang R, Chattopadhyay A, Singh A, et al. Physics-informed machine learning: case studies for weather and climate modelling. Phil Trans R Soc A 2021;379(2194):20200093. http://dx.doi.org/10.1098/rsta.2020.0093.

[31] Ma Z, Liao H, Gao J, Nie S, Geng Y. Physics-informed machine learning for degradation modeling of an electro-hydrostatic actuator system. Reliab Eng Syst Saf 2023;229:108898. http://dx.doi.org/10.1016/j.ress.2022.108898.

[32] Jiang C, Vega MA, Todd MD, Hu Z. Model correction and updating of a stochastic degradation model for failure prognostics of miter gates. Reliab Eng Syst Saf 2022;218:108203. http://dx.doi.org/10.1016/j.ress.2021.108203.

[33] Zhang Q, Xu N, Ersoy D, Liu Y. Manifold-based conditional Bayesian network for aging pipe yield strength estimation with non-destructive measurements. Reliab Eng Syst Saf 2022;223:108447. http://dx.doi.org/10.1016/j.ress.2022.108447.

[34] Aljanaideh KF. Time-domain analysis of sensor-to-sensor transmissibility operators with application to fault detection (Ph.D. thesis), 2015.

[35] Aljanaideh KF, Bernstein DS. Output-only identification of input–output models. Automatica 2020;113:108686. http://dx.doi.org/10.1016/j.automatica.2019.108686.

[36] Khalil A, Aljanaideh KF. Aircraft structural health monitoring using transmissibility identification. IFAC-PapersOnLine 2018;51(15):969–74. http://dx.doi.org/10.1016/j.ifacol.2018.09.068, 18th IFAC Symposium on System Identification SYSID 2018.

[37] Aljanaideh KF, Bernstein DS. Experimental Application of Time-Domain Transmissibility Identification to Fault Detection and Localization in Acoustic Systems. J Vib Acoust 2017;140(2):021017. http://dx.doi.org/10.1115/1.4038436, arXiv:https://asmedigitalcollection.asme.org/vibrationacoustics/article-pdf/140/2/021017/6379870/vib_140_02_021017.pdf.

[38] Aggoune L, Chetouani Y. Neyman–pearson test for fault detection in the process dynamics. J Fail Anal Prev 2016;16(6):999–1005. http://dx.doi.org/10.1007/s11668-016-0186-y.

[39] Dinca L, Aldemir T, Rizzoni G. Fault detection and identification in dynamic systems with noisy data and parameter/modeling uncertainties. Reliab Eng Syst Saf 1999;65(1):17–28. http://dx.doi.org/10.1016/S0951-8320(98)00077-5.

[40] He Z, Zhou H, Wang J, Chen Z, Wang D, Xing Y. An improved detection statistic for monitoring the nonstationary and nonlinear processes. Chemometr Intell Lab Syst 2015;145:114–24. http://dx.doi.org/10.1016/j.chemolab.2015.04.016.

[41] Lv H, Yin C, Cui Z, Zhan Q, Zhou H. Risk assessment of security systems based on entropy theory and the Neyman–Pearson criterion. Reliab Eng Syst Saf 2015;142:68–77. http://dx.doi.org/10.1016/j.ress.2015.04.023.

[42] Neyman J, Pearson ES. On the problem of the most efficient tests of statistical hypotheses. Philos Trans R Soc Lond Ser A 1933;231(694–706):289–337. http://dx.doi.org/10.1098/rsta.1933.0009.

[43] Koizumi Y, Saito S, Uematsu H, Harada N. Optimizing acoustic feature extractor for anomalous sound detection based on Neyman-Pearson lemma. In: 2017 25th European signal processing conference. EUSIPCO, 2017, p. 698–702. http://dx.doi.org/10.23919/EUSIPCO.2017.8081297.

[44] Shashoa NAA, Kvaščev G, Marjanović A, Djurović Ž. Sensor fault detection and isolation in a thermal power plant steam separator. Control Eng Pract 2013;21(7):908–16. http://dx.doi.org/10.1016/j.conengprac.2013.02.012.

[45] Charfi F, Lesecq S, Sellami F. Fault diagnosis using SWT and neyman pearson detection tests. In: 2009 IEEE international symposium on diagnostics for electric machines, power electronics and drives. IEEE; 2009, p. 1–6. http://dx.doi.org/10.1109/DEMPED.2009.5292788.

[46] Jeong K, Choi SB, Choi H. Sensor fault detection and isolation using a support vector machine for vehicle suspension systems. IEEE Trans Veh Technol 2020;69(4):3852–63. http://dx.doi.org/10.1109/TVT.2020.2977353.

[47] Chamseddine A, Noura H. Control and sensor fault tolerance of vehicle active suspension. IEEE Trans Control Syst Technol 2008;16(3):416–33. http://dx.doi.org/10.1109/TCST.2007.908191.

[48] Aljanaideh KF, Bernstein DS. Time-domain analysis of sensor-to-sensor transmissibility operators. Automatica 2015;53:312–9. http://dx.doi.org/10.1016/j.automatica.2015.01.004.

[49] Devriendt C, Guillaume P. Identification of modal parameters from transmissibility measurements. J Sound Vib 2008;314(1):343–56. http://dx.doi.org/10.1016/j.jsv.2007.12.022.

[50] Weijtjens W, De Sitter G, Devriendt C, Guillaume P. Operational modal parameter estimation of MIMO systems using transmissibility functions. Automatica 2014;50(2):559–64. http://dx.doi.org/10.1016/j.automatica.2013.11.021.

[51] Aljanaideh KF, Bernstein DS. Closed-loop identification of unstable systems using noncausal FIR models. Internat J Control 2017;90(2):168–85. http://dx.doi.org/10.1080/00207179.2016.1172733.

[52] Kuha J. AIC and BIC: Comparisons of assumptions and performance. Sociol Methods Res 2004;33(2):188–229. http://dx.doi.org/10.1177/0049124103262065.

[53] Lennart L. System identification: theory for the user. Upper Saddle River, NJ: PTR Prentice Hall; 1999, p. 1–14.

[54] Chen LLT. What can regularization offer for estimation of dynamical systems? IFAC Proc Vol 2013;46(11):1–8. http://dx.doi.org/10.3182/20130703-3-FR-4038.00155, 11th IFAC Workshop on Adaptation and Learning in Control and Signal Processing.

[55] Chen T, Ohlsson H, Goodwin GC, Ljung L. Kernel selection in linear system identification part II: A classical perspective. In: 2011 50th IEEE conference on decision and control and European control conference. 2011, p. 4326–31. http://dx.doi.org/10.1109/CDC.2011.6160722.

[56] Chen T, Ljung L. Implementation of algorithms for tuning parameters in regularized least squares problems in system identification. Automatica 2013;49(7):2213–20. http://dx.doi.org/10.1016/j.automatica.2013.03.030.

[57] Zhang Y, Xiao Z, Li P, Tang X, Ou G. Conservative sensor error modeling using a modified paired overbound method and its application in satellite-based augmentation systems. Sensors 2019;19(12):2826. http://dx.doi.org/10.3390/s19122826.

[58] Jan SU, Koo I. A novel feature selection scheme and a diversified-input svm-based classifier for sensor fault classification. J Sens 2018;2018. http://dx.doi.org/10.1155/2018/7467418.

[59] Imparato D, Teunissen P, Tiberius C. Minimal detectable and identifiable biases for quality control. Surv Rev 2019;51(367):289–99. http://dx.doi.org/10.1080/00396265.2018.1437947.

[60] Truchon J-F, Bayly CI. Evaluating virtual screening methods: Good and bad metrics for the "early recognition" problem. J Chem Inf Model 2007;47(2):488–508. http://dx.doi.org/10.1021/ci600426e.

[61] Saraygord Afshari S, Enayatollahi F, Xu X, Liang X. Machine learning-based methods in structural reliability analysis: A review. Reliab Eng Syst Saf 2022;219:108223. http://dx.doi.org/10.1016/j.ress.2021.108223.