A Framework for Unifying Model-based and Data-driven Fault Diagnosis

Hamed Khorasgani[1], Ahmed Farahat[2], Kosta Ristovski[3], Chetan Gupta[4], and Gautam Biswas[5]

[1,2,3,4] *Industrial AI Laboratory, Hitachi America Ltd., Santa Clara, CA, USA*
hamed.khorasgani@hal.hitachi.com
ahmed.farahat@hal.hitachi.com
kosta.ristovski@hal.hitachi.com
chetan.gupta@hal.hitachi.com

[5] *Institute of Software Integrated Systems, Vanderbilt University, TN, USA*
gautam.biswas@vanderbilt.edu

ABSTRACT

Model-based diagnosis methods rely on a model that defines nominal behavior of a dynamic system to detect abnormal behaviors and isolate faults. On the other hand, data-driven diagnosis algorithms detect and isolate system faults by operating exclusively on system measurements and using very little knowledge about the system. Recently, several researchers have combined model-based diagnosis techniques with data-driven approaches to propose hybrid[1]solutions for fault diagnosis. Many researchers have proposed methods to integrate specific approaches. In this paper, we demonstrate that data-driven and model-based diagnosis methods follow a similar procedure and can be represented by a general unifying framework. This unifying framework for fault detection and isolation can be used to integrate different methodologies developed by two communities. As a case study, we use the proposed framework to build a crossover solution for fault diagnosis in a wind turbine benchmark. In this case study, we show that it is possible to achieve a better diagnosis performance by using a hybrid method that follows the proposed framework.

1. INTRODUCTION

Model-based diagnosis methods use analytical redundancies in the system model for fault detection and isolation. Model-based approaches, such as graphical methods (Mosterman & Biswas, 1999; Bregon et al., 2014), observer-based methods (Alcorta-Garcia & Frank, 1997), and parity equations and analytical redundancy relations (ARR) (Gertler, 1998) have been successfully applied for fault detection and isolation (FDI) in dynamic systems. Model-based methods are computationally efficient. Moreover, it is easy to understand and interpret the diagnosis results of these approaches. However, for complex systems, model building can be expensive, and it is often infeasible to derive a sufficiently accurate model for the system that generates correct diagnosis results. When sufficiently accurate models are not available, data-driven diagnosis methods have been developed as a promising alternative to model-based fault detection and isolation (Venkatasubramanian et al., 2003). Data-driven diagnosis approaches use historical data of system operations to find patterns that do not conform to the expected systems behavior. These patterns, called anomalies or outliers, typically correspond to single or a small group of data points that appear to be sufficiently different from the expected operational behaviors of the system (Chandola et al., 2009).

Yang & Rizzoni (2016) compared the performances of a model-based fault detection method based on analytical redundancy relations with a data-driven scheme for fault detection and isolation using linear discriminant analysis applied to an internal combustion engine. Their analysis showed that both methods deliver high detection rates and low false alarm rates for engine diagnosis. However, in many practical cases we cannot expect high diagnosis performance from model-based or data-driven methods alone for the following reasons:

- *Incomplete models*: for complex systems, the system diagnostics models are not easy to develop, and keep updated during the system life-cycle. Therefore, reliable models of these systems are not always available. Even when models are available, they are often incomplete and plagued by uncertainties in tracking system behavior.

[1] Note that in the literature the term hybrid is also used for systems that show combine continuous behaviors interspersed with discrete changes (e.g., due to switching in complex, embedded systems) (Hofbaur & Williams, 2004).

This can lead to high false positive or high false negative rates in model-based diagnosis.

- *Insufficient historical data*: in the real world, we usually do not have access to enough training data to learn the system normal behaviors in all the operating modes. Having access to labeled data for faulty operations is even more expensive and difficult. This can lead to unsatisfying diagnosis performance in data-driven methods. Furthermore, the quality and scope of the data can play a critical role in the performance of the machine learning methods that are employed to derive diagnostic models from the data.

Several researchers have combined model-based diagnosis with data-driven approaches to address some of these problems (Tidriri et al., 2016). Jung et al. (2016) used a model-based approach for fault detection. To achieve better fault isolation, they used the residual outputs from previous fault scenarios to generate models for different fault modes. They trained a one-class support vector machine (1-SVM) to detect each fault. If the new sample does not correspond to the nominal or one of the known fault modes, it is labeled as a likely unknown fault. The classifiers were expected to become more accurate as more data was collected over time. Data-driven methods have also been used to learn uncertain and incomplete models (Talebi et al., 2009).

Benkouider et al. (2012) developed a hybrid approach that combines extended Kalman filtering (EKF) and neural network classifier for fault detection and isolation in chemical reactors. In their method, the system measurements and the estimated parameters by EKF are the inputs to the neural network, while the outputs of the classifier are fault types in the reactor. Sheibat-Othman et al. (2014) used Support Vector Machines (SVM) for fault detection and applied an observer-based diagnosis approach for fault isolation. When the SVM detects no fault in the system, they use the data to update the observer parameters. Narasimhan et al. (2010) applied TRANSCEND diagnosis approach (Mosterman & Biswas, 1999) to reduce the set of possible faults, and then applied a data-driven approach to best distinguish among the remaining faults.

Mack et al. (2017) combined a diagnosis reference model developed by domain experts with a tree augmented naive Bayesian (TAN) learning algorithm to develop a diagnosis method that combines the expert knowledge and historical data to improve the accuracy in differentiating between nominal and faulty situations. Khorasgani & Biswas (2017) developed a hybrid diagnosis approach that combines the use of historical data with the available physics-based knowledge of the system to achieve better diagnosis performance in smart buildings with incomplete models. By combining model-based diagnosis and data-driven anomaly detection, they could overcome the limitations of the incomplete system model and improve the diagnosis accuracy.

This paper makes the following contributions.

- We demonstrate that data-driven and model-based diagnosis methods follow a similar procedure and can be represented by a general unifying framework.
- We use the general unifying framework to build hybrid solutions that integrate techniques developed by different research communities. Unlike previous hybrid diagnosis methods (Tidriri et al., 2016), our proposed framework is not limited to specific model-based and data-driven methods and can be used to combine different methods.

The rest of this paper is organized as follows. Our general framework for unifying data-driven and model-based diagnosis methods is presented in Section 2. Section 3 uses our unifying framework to build a crossover solution for fault diagnosis for a wind turbine benchmark. This section demonstrates diagnosis performance improvement can be achieved through a hybrid solution. Section 4 presents the conclusions of the paper.

2. TOWARDS A UNIFYING FRAMEWORK FOR DATA-DRIVEN AND MODEL-BASED DIAGNOSIS METHODS

In this section, we use the similarities between diagnosis methods, whether model-based or data-driven, to propose a general unifying diagnosis framework. Our common framework can be used for comparison and integration of different diagnosis solutions. Both model-based and data-driven diagnosis solutions apply system measurements for fault detection and isolation. Therefore, data acquisition is the first step in our unifying framework. After the data acquisition, data-driven methods typically apply a feature selection or feature extraction technique to extract a set of relevant features from measurement data or select a subset of measurements that are sensitive to the faults. Among feature extraction methods, Principal Components Analysis (PCA) is the most widely used (Bengio et al., 2012). It generates a set of orthogonal bases in the directions where the data has the greatest variances.

On the other hand, model-based techniques use residuals for fault detection and isolation. A residual is an analytical redundancy relation between known variables in the system such as parameters of the system, process measurements, and inputs. To detect a fault f, model-based approaches generate a residual sensitive to the fault and, at the same time, invariant or at least robust to noise and uncertainties. To isolate a fault f_i from another fault f_j, the model-based methods require a residual sensitive to f_i and at the same time insensitive to f_j (Frank & Ding, 1994). Residuals are used to capture inconsistency among the measurements. Therefore, we can consider a residual as a feature extracted from a set of measurements. In fact, Gertler et al. (1999) showed the equivalence between

PCA, which is a common unsupervised data-driven feature extraction method and parity relations, which is a common model-based residual generation technique for linear systems. Considering each residual as a feature extracted from system model, we define feature extraction as the second step in our unifying diagnosis framework.

In the next step, a diagnosis solution maps the features to the nominal operating mode or different fault modes. We call this step fault diagnosis. Several survey and review articles (Chandola et al., 2009; Hodge & Austin, 2004) have categorized data-driven fault detection approaches based on the method applied for fault classification in this step. Generally, they fall into two main groups: 1) supervised, and 2) unsupervised. In model-based methods, to make a diagnoser robust to noise and uncertainties, typically, a hypothesis test is used to determine if a residual deviation is statistically significant. Then, a fault isolation algorithm, uses a decision logic to generate possible fault candidates based on the hypothesis tests outputs. Hypothesis test followed by decision logic techniques do not include learning normal and faulty classes from training data and, therefore, can be categorized as unsupervised diagnosis methods.

Our proposed common framework to represent data-driven and model-based diagnosis methods is shown in Figure 1. We use this general presentation to develop a framework for unifying data-driven and model-based diagnosis algorithms. Data sets generated from different sources can be used together to achieve a better understanding of system behavior specially with respect to fault modes (Leturiondo, 2016). In the feature extraction step, the set of features can be expanded by using union of features from several methods. It is also possible to combine two different methods by using one of the methods for feature extraction and the other one for fault diagnosis. In the rest of this section, we present a detailed description of each component in our framework and provide a strategy, and guidance to combine different methods in each part of the framework.

2.1. Data acquisition

A key aspect of any fault diagnosis method is the selection of the input data. The input data is generally system measurements collected by sensors that are part of the system, and provide information on system operations. The input data can also include data generated by conducting experimental tests on the physical process, or from simulation data generated by system simulators of sufficient fidelity. In other words, we can categorize the input data for the diagnosis methods into two main groups.

- *Field or experimental data:* Field data is sensor data generated during the system operation. The ultimate goal of any health monitoring strategy is to detect and isolate faults in real operational situations, sometimes in real time as faults occur. Therefore, it is important to use field data for design and validation of diagnosis solutions. Experimental data is sensor data generated during a specific experimental study. Experiments can be done to understand the effect of faults and how they evolve over time. The experimental data can be used for designing fault detectors and diagnosers, and also for testing the existing approaches.

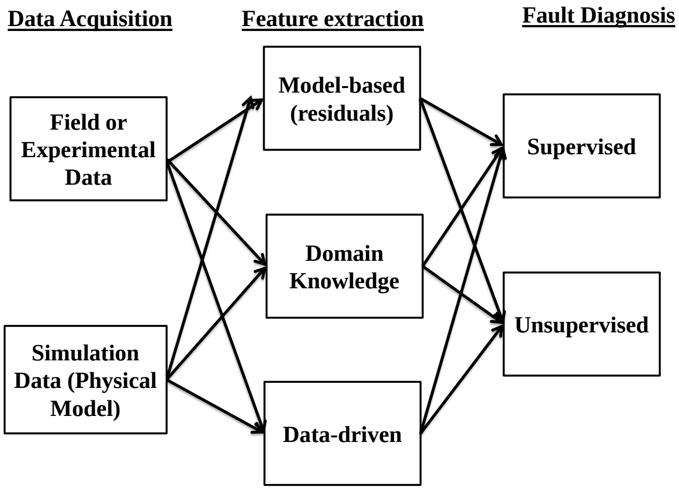

Figure 1. Our unifying framework to represent model-based and data-driven diagnosis methods.

- *Simulation data:* Simulation data is generated using an executable form of a system model (e.g., state space equations) to imitate situations that may occur in the real world. Fault data from actual operations may not always be available. Moreover, in many cases, real normal data is not also available for all the operating modes. In these cases, system experts can use simulators to generate faulty or even normal data, which can then be used to design detectors and diagnosers, much like the case of experimental data.

A combination of data sources can be used to generate hybrid diagnosis methods. Sheibat-Othman et al. (2014) used collected data during the system operations (field data) as normal data and introduce actuator and sensor faults in an experimental study to collect faulty data (experimental data). Leturiondo (2016) proposed a hybrid monitoring approach, which combines field data and data generated using the system physical model (simulation data) to address the lack of sufficient faulty data in real world diagnosis applications.

Since the models are not always completely accurate, the field data and experimental data are typically more reliable than simulation data. For the operating conditions and the fault modes where the field data and experimental data are not available, simulation data can be used to enrich the dataset. Note that simulation data is often generated at different levels of detail compared to field and experimental data. In real systems, there are limited number of sensors with specific sampling rates, however, in simulation experiments, we can record any variable at any desired rate. To blend the simulation data with real world data, the users can only use the variables associated with system sensors and apply a re-sampling approach to create consistent sampling rates in data coming from different sources.

2.2. Feature extraction

Feature extraction is the most critical step in designing a diagnosis algorithm. We categorize feature extraction methods into three main groups.

- *Data-driven:* Typically, data-driven diagnosis methods use sensor data as the set of features for fault detection and isolation. When measurements are noisy or there are irrelevant measurements in the dataset, it becomes challenging to detect and isolate faults by only monitoring the raw data. Therefore, as part of developing a diagnosis approach, we have to devise methods for defining new features, or selecting a subset of measurements that are sensitive to system faults as the features. There are different data-driven methods such as mutual information (Peng et al., 2005) to select a set of features that have maximum relevance to the fault classes.
- *Domain knowledge:* Domain experts can play an important role in feature selection. They can identify important features for detecting and isolating each fault. Moreover, they can help to define new features by providing critical information about nominal behavior of each measurement with respect to others. For example, a bipartite graph developed by domain experts that represents the relationship between features and system faults, called the diagnosis reference model has been used widely in monitoring complex systems, such as subsystems of Boeing 777 aircraft (Mack et al., 2017).
- *Model-based:* The system model can be used to generate residuals which represent analytical redundancy relations among measurements during nominal operation. Residuals are used to capture inconsistency among the measurements. There are three main approaches for residual generation; 1) observer-based (Frank & Ding, 1997), 2) identification methods (Isermann, 1993), and 3) parity equations and analytical redundancy approaches (Gertler, 2012; Samantaray et al., 2006). Model-based methods use different approaches such as sensitivity analysis (Khorasgani et al., 2014) to select a set of residuals that are sensitive to the faults.

The union of domain knowledge-based features, model-based residuals, and data-driven features can be used as the set of features for fault diagnosis. In the case study, we will show additional features can improve diagnosis performance significantly.

2.3. Fault diagnosis

Diagnosis algorithms use the extracted features to detect and isolate faults. We categorize the diagnosis methods into two main groups.

- *Supervised:* Supervised approaches use training data to learn normal and fault classes. These methods typically apply classification methods such as neural networks or Bayesian networks to map the features to the operating modes (Hodge & Austin, 2004).
- *Unsupervised:* Unsupervised approaches do not start with labeled data. These methods typically make the implicit assumption that normal instances are far more frequent than anomalies in the test dataset and use hypothesis tests or clustering approaches to divide up the data points into normal and fault conditions (Li et al., 2011).

As we mentioned earlier model-based methods typically apply hypothesis tests to detect faults, and then apply generic logic-based approaches to generate fault hypotheses and then refine them to isolate faults. A fault hypothesis is a set of faults that is consistent with the observed residual outputs. Data-driven approaches typically use classification methods to map the features to the operating modes. Classification is a supervised diagnosis method that uses labeled training data to learn a set of predefined fault classes. On the other hand, clustering or unsupervised learning methods can be used to detect new modes of operation some of which may be labeled as faulty situations. This analyses can be done without starting with labeled data that is linked to specific faults. Note that classification and clustering techniques are not limited to data-driven methods, and model-based techniques may also use classifiers or clustering algorithms to map the residual outputs to normal and fault classes. Jung et al. (2016) developed a hybrid diagnosis method by using historical data to train a one-class support vector machine (1-SVM) to map the residual outputs to the operating modes.

Table 1 summarizes fault diagnosis methods that can be applied for different scenarios. When training data for both normal and fault modes is available, we can use a classifier, such as support vector machine to distinguish fault modes from nominal operation. The input to the classifier is the set of features and the output is the system operation modes. Using classifiers improves diagnostic performance (Jung et al., 2016). However, faulty data is not always available for training diagnosers. When we only have access to normal data or we have access to the system model to generate normal data, we can apply hypothesis testing to compare nominal feature values (derived from the data), against newly observed values, to identify feature values that are significantly out of range.

Typically, hypothesis testing schemes like the Z-test (Biswas

Table 1. Fault diagnosis methods for different scenarios.

Normal Data	Faulty Data	Fault detection and isolation
Available	Available	Classifier
Available	Unavailable	Hypothesis tests
Unavailable	Unavailable	Clustering

et al., 2003) assume the data is normally distributed, and only require the means and variances of feature values of the nominal data when comparing against new operational data to detect faults. The system is in a fault mode when the hypothesis tests indicate that one or more features are out of their nominal operating ranges. A decision logic unit uses the set of features out of nominal interval to isolate the fault mode. When normal data is not available, it is reasonable to assume that there is no fault in the early stages of operation and therefore, the data in this operational period is normal. However, in many cases the data from the early stages is not available for all the operating modes of the system. In this situation, we can apply a clustering method to detect and isolate fault modes. In the first step, an algorithm such as Calinski and Harabasz method (Calinski & Harabasz, 1974) can be applied to detect the number of clusters in the dataset. Next, we can apply a clustering algorithm such as hierarchical clustering to detect the clusters in the dataset. The small clusters are candidates for the fault modes. To show our proposed framework is general and can represents different diagnosis methodologies, we use the framework to represent previous data-driven, model-based and hybrid diagnosis solutions in Table 2.

3. CASE STUDY

In this section, we use our framework to combine two diagnosis solutions for a wind turbine benchmark competition challenge (Odgaard et al., 2013) and generate a hybrid solution with better diagnosis performance. We use the same notation as reference (Odgaard et al., 2013) when referring to the diagnosis solutions: Gaussian Kernel Support Vector Machine (GKSV) (Laouti et al., 2011), and Estimation-Based (EB) (Zhang et al., 2011). In the next subsection, we use the proposed framework to present the aforementioned approaches.

3.1. Solutions for the wind turbine benchmark competition challenge

A simulator is used to generate data in normal operation and in nine different fault scenarios. This means the physical model approach has been used for data acquisition in all the methodologies. However, each research group chose a different feature extraction and fault diagnosis methodology. A summary of each solution is presented as follows.

- *GKSV* uses domain knowledge to define a set of features and applies SVM to classify fault modes. SVM is a supervised diagnosis method because it uses training data for both normal and anomalous operations.

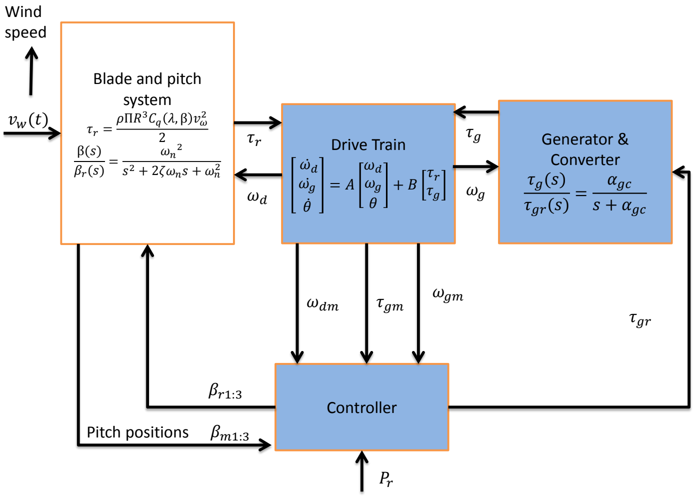
Wind speed
$v_w(t)$
Blade and pitch system
$\tau_r = \frac{\rho \Pi R^3 C_q(\lambda,\beta) v_\omega^2}{2}$
$\frac{\beta(s)}{\beta_r(s)} = \frac{{\omega_n}^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$
τ_r
ω_d
Drive Train
$\begin{bmatrix}\dot{\omega}_d\\ \dot{\omega}_g\\ \dot{\theta}\end{bmatrix} = A\begin{bmatrix}\omega_d\\ \omega_g\\ \theta\end{bmatrix} + B\begin{bmatrix}\tau_r\\ \tau_g\end{bmatrix}$
τ_g
ω_g
Generator & Converter
$\frac{\tau_g(s)}{\tau_{gr}(s)} = \frac{\alpha_{gc}}{s + \alpha_{gc}}$
ω_{dm}
τ_{gm}
ω_{gm}
τ_{gr}
$\beta_{r1:3}$
Pitch positions
$\beta_{m1:3}$
Controller
P_r

Figure 2. Wind turbine block diagram.

- *EB* uses observers-based approach to generate the set of residuals. The hypotheses test is not discussed in the paper.
  Table 2 shows how the two solutions can be expressed using our unifying framework. Note that we can use our proposed general framework to make more specific and useful comparison among different diagnosis methods. For example, Odgaard et al. (2013) only used the overall diagnosis performance of each method for the comparison in the benchmark challenge. However, using our framework it is possible to fix the diagnosis step to the same method for all the solutions and compare the performance of different features in different methods, or use the same features and compare the performance of different diagnosis techniques in classifying the normal and faulty operating modes. In the next subsection, we use our framework to develop a hybrid solution for the wind turbine challenge.

3.2. Combining different diagnosis methods

We demonstrate the effectiveness of our proposed framework for building hybrid solutions by combining the GKSV and the EB solutions for wind turbine competition. The wind turbine as it is shown in Figure 2 has four subsystems: 1) blade and pitch, 2) drive train, 3) generator and converter and 4) con-

Table 2. Different diagnosis solutions presented in our common framework.

Method	Data Acquisition	Feature extraction	Diagnosis
Model-based solution Yang & Rizzoni (2016)	Simulation data (physical model)	Observer-based residuals (model-based)	CUSUM test (unsupervised)
Data-driven solution Yang & Rizzoni (2016)	Simulation data (physical model)	Linear discriminant analysis (LDA) (data-driven)	CUSUM test (unsupervised)
Sheibat-Othman et al. (2014)	Normal filed data + faulty simulation data (hybrid)	Observer-based residuals + sensor data (hybrid)	SVM classifier (supervised)
Benkouider et al. (2012)	Experimental data + simulation data for a hazardous fault	EKF-based residuals + sensor data (hybrid)	Neural networks (supervised)
Jung et al. (2016)	Simulation data (physical model)	Residuals (model-based)	simple threshold + consistency-based analysis + 1-SVM (hybrid)
Mack et al. (2017)	Field data	Experts (domain-knowledge)	TAN classifier (supervised)
Narasimhan et al. (2010)	Simulation data (physical model)	Residuals (model-based)	simple threshold + consistency-based analysis + 1-SVM (hybrid)
GKSV solution Odgaard et al. (2013)	Simulation data (physical model)	Experts features (domain knowledge)	SVR (supervised)
EB solution Odgaard et al. (2013)	Simulation data (physical model)	Observer-based residuals (model-based)	*

troller. In this work, we did not have access to the wind speed data and the complete model of the blade and pitch system, therefore, we only model the three other subsystems (the blue boxes in Figure 2) and assume the blade and pitch subsystem supplies constant torque to the drive train subsystem.

3.2.1. Wind turbine physical model

The drive train model can be represented by the following equations.

$$\begin{aligned}
J_r\dot{\omega}_r(t) &= \tau(t) - K_{dt}\theta_\Delta(t) - (B_{dt}+B_r)\omega_r(t) + \frac{B_{dt}}{N_g}\omega_g(t) \\
J_g\dot{\omega}_g(t) &= \frac{\eta_{dt}K_{dt}}{N_g}\theta_\Delta(t) + \frac{\eta_{dt}B_{dt}}{N_g}\omega_r(t) \\
&\quad -(\frac{\eta_{dt}B_{dt}}{{N_g}^2} + B_g)\omega_g(t) - \tau_g(t) \\
\dot{\theta}_\Delta(t) &= \omega_r(t) - \frac{1}{N_g}\omega_g(t),
\end{aligned} \tag{1}$$

where the generator torque, $\tau_g(t)$, is the input to the subsystem, and ω_r, the rotational speed of the rotor, ω_g, the rotational speed of the generator, and θ_Δ, the torsion angle of the drive train, are the state variables. The subsystem parameters are presented in Table 3.

The generator and converter can be modeled with following first-order equation.

$$\dot{\tau}_g(t) = -\alpha_{gc}\tau_g + \alpha_{gc}\tau_{gr}, \tag{2}$$

where α_{gc} is the generator and converter model parameter and τ_{gr} is the reference torque computed by the controller unit. The power produced by the generator is

$$P_g(t) = -\eta_g\omega_g(t)\tau_g(t), \tag{3}$$

where η_g is the generator's efficiency rate. The controller starts at mode 1. In this mode, we have

$$\tau_{gr}(t) = K_{opt}(\frac{\omega(t)}{N_g})^2, \tag{4}$$

where K_{opt} is the controller parameter. The controller switches to mode 2 if ω_g becomes greater than the nominal generator speed, ω_{nom}, or P_g becomes greater than reference power, P_r. In mode 2, we have

$$\tau_{gr}(t) = \frac{P_r}{\eta_g\omega_g}. \tag{5}$$

There are two sensors for rotor speed, ω_r, two sensors for generator speed, ω_g, a sensor for generator power, P_g, and a sensor for generator torque, τ_g. Each sensor is modeled as a sum of actual variable value and a stochastic noise. The sensor parameters are presented in (Odgaard et al., 2013).

Table 3. Wind Turbine Parameters Used in the Benchmark Model.

Parameter	Name	Value and unit (SI)
Moment of inertia of the low speed shaft	J_r	$55 \times 10^6\ kg.m^2$
Torsion stiffness of the drive train	K_{dt}	$2.7 \times 10^9\ N.m/rad$
Torsion damping coefficient of the drive train	B_{dt}	$775.49\ N.m.s/rad$
Viscous friction of the high speed shaft	B_g	$45.6\ N.m.s/rad$
Gear ratio	N_g	95
Moment of inertia of the high speed shaft	J_g	$390\ kg.m^2$
Efficiency of the drive train	η_g	0.98
Generator and converter model parameter	α_{gc}	$50\ rad./s$
Controller parameter	K_{opt}	1.2171
Reference power	P_r	$4.8 \times 10^6\ W$

3.2.2. Data

We use the wind turbine physical model to generate training and test data sets as described in the competition (Odgaard et al., 2013). We consider the following fault scenarios in the training data:

- f_1 (rotor speed sensor fault): in this fault scenario the first rotor speed sensors has an additive bias equal to 1.4 rad/s from 1500s to 1600s.
- f_2 (generator speed sensor fault): in this fault scenario the first generator speed sensor has a gain factor equal to .9 from 1000s to 1100s.
- f_3 (generator speed sensor fault): in this fault scenario the second generator speed sensor has a gain factor equal to 1.1 from 1200s to 1300s.
- f_4 (generator actuator fault): in this fault scenario the generator generates $100Nm$ more than the reference torque from 3800s to 3900s.

We consider the following fault scenarios in the test data:

- f_1 (rotor speed sensor fault): in this fault scenario the first rotor speed sensors has an additive bias equal to 0.9 rad/s from 1500s to 1800s.
- f_2 (generator speed sensor fault): in this fault scenario the first generator speed sensor has a gain factor equal to 1.1 from 700s to 1000s.
- f_3 (generator speed sensor fault): in this fault scenario the second generator speed sensor has a gain factor equal to 0.8 from 1100s to 1400s.
- f_4 (generator actuator fault): in this fault scenario the generator generates $200Nm$ more than the reference torque from 3700s to 4000s.

More details of the wind turbine model and fault scenarios are presented in (Odgaard et al., 2013). Next, we present the GKSV solution for fault diagnosis.

3.2.3. Hybrid solution for fault diagnosis

Laouti et al. (2011) extracted a set of features and applied several Gaussian kernel support vector regressions to detect different faults. In this paper, we are interested in comparing different methods with our hybrid approach. Therefore, we use the same features as GKSV solution, but to have an identical approach in the second step we use linear support vector classifier for diagnosis. The set of features in GKSV solution are as follows.

$$x = \begin{bmatrix} |\omega_{r,m_1}(k) - \omega_{r,m_2}(k)| \\ |\omega_{r,m_1}(k+1) - \omega_{r,m_1}(k)| \\ |\omega_{r,m_2}(k+1) - \omega_{r,m_2}(k)| \\ |\omega_{g,m_1}(k) - \omega_{g,m_2}(k)| \\ |\omega_{g,m_1}(k+1) - \omega_{g,m_1}(k)| \\ |\omega_{g,m_2}(k+1) - \omega_{g,m_2}(k)| \\ |\frac{\tau_g(k)*\omega_{g,m_1}(k)}{P_g(k)}| \\ |\frac{\tau_g(k)*\omega_{g,m_2}(k)}{P_g(k)}| \\ |\frac{\tau_g(k)*\omega_{r,m_1}(k)}{P_g(k)}| \\ |\frac{\tau_g(k)*\omega_{r,m_2}(k)}{P_g(k)}| \\ ||\tau_{gr}(k) - \tau_g(k)| \\ |\frac{P_g(k)}{t_{gr}} - (\frac{\omega_{g,m_1}(k)+\omega_{g,m_2}(k)}{2})| \end{bmatrix} \tag{6}$$

where $\omega_{r,m_i}(k)$ is the rotor speed sensor i value at sample time k, $\omega_{g,m_i}(k)$ is the generator speed sensor i value at sample time k, $\tau_g(k)$ is the generator torque sensor value at sample time k, $\tau_{gr}(k)$ is the generator reference torque set by the controller at sample time k, and $P_g(k)$ is the generator power sensor value at sample time k.

When we use this feature set to train the classifier and apply the test data to measure the accuracy of the diagnosis approach, we can detect and isolate the faults with 4.5% false positive rate and 19% false negative rate. On the other hand, Zhang et al. (2011) implemented the method described in (Zhang et al., 2008) as a solution for the diagnosis competition. The method uses an observer-based approach to design

Table 4. Diagnosis performances for the wind turbine benchmark

Method	False Positive Rate	False Negative Rate
GKSV solution	4.5%	19%
EB solution	4.6%	16.6%
Hybrid solution	0.5%	6.7%

the following residuals.

$$\begin{aligned}\dot{\hat{\theta}}_{\Delta_1} &= \omega_{r,m_1} - \frac{1}{N_g}\hat{\omega}_{g_1} + k_1(\omega_{g,m_1} - \hat{\omega}_{g_1})\\ \dot{\hat{\omega}}_{g_1} &= \frac{1}{J_g}(\frac{\eta_{dt}K_{dt}}{N_g}\hat{\theta}_{\Delta_1} + \frac{\eta_{dt}B_{dt}}{N_g}\omega_{r,m_1} - (\frac{\eta_{dt}B_{dt}}{{N_g}^2} + B_g)\hat{\omega}_{g_1}\\ &\quad - \tau_g) + k_2(\omega_{g,m_1} - \hat{\omega}_{g_1})\\ r_1 &= \omega_{g,m_1} - \hat{\omega}_{g_1}\end{aligned} \tag{7}$$

$$\begin{aligned}\dot{\hat{\theta}}_{\Delta_2} &= \omega_{r,m_2} - \frac{1}{N_g}\hat{\omega}_{g_2} + k_1(\omega_{g,m_1} - \hat{\omega}_{g_2})\\ \dot{\hat{\omega}}_{g_2} &= \frac{1}{J_g}(\frac{\eta_{dt}K_{dt}}{N_g}\hat{\theta}_{\Delta_2} + \frac{\eta_{dt}B_{dt}}{N_g}\omega_{r,m_2} - (\frac{\eta_{dt}B_{dt}}{{N_g}^2} + B_g)\hat{\omega}_{g_2}\\ &\quad - \tau_g) + k_2(\omega_{g,m_1} - \hat{\omega}_{g_2})\\ r_2 &= \omega_{g,m_1} - \hat{\omega}_{g_2}\end{aligned} \tag{8}$$

$$\begin{aligned}\dot{\hat{\theta}}_{\Delta_3} &= \omega_{r,m_1} - \frac{1}{N_g}\hat{\omega}_{g_3} + k_1(\omega_{g,m_2} - \hat{\omega}_{g_3})\\ \dot{\hat{\omega}}_{g_3} &= \frac{1}{J_g}(\frac{\eta_{dt}K_{dt}}{N_g}\hat{\theta}_{\Delta_3} + \frac{\eta_{dt}B_{dt}}{N_g}\omega_{r,m_1} - (\frac{\eta_{dt}B_{dt}}{{N_g}^2} + B_g)\hat{\omega}_{g_1}\\ &\quad - \tau_g) + k_2(\omega_{g,m_2} - \hat{\omega}_{g_2})\\ r_3 &= \omega_{g,m_2} - \hat{\omega}_{g_3}\end{aligned} \tag{9}$$

$$\begin{aligned}\dot{\hat{\theta}}_{\Delta_4} &= \omega_{r,m_2} - \frac{1}{N_g}\hat{\omega}_{g_4} + k_1(\omega_{g,m_2} - \hat{\omega}_{g_4})\\ \dot{\hat{\omega}}_{g_4} &= \frac{1}{J_g}(\frac{\eta_{dt}K_{dt}}{N_g}\hat{\theta}_{\Delta_4} + \frac{\eta_{dt}B_{dt}}{N_g}\omega_{r,m_2} - (\frac{\eta_{dt}B_{dt}}{{N_g}^2} + B_g)\hat{\omega}_{g_4}\\ &\quad - \tau_g) + k_2(\omega_{g,m_2} - \hat{\omega}_{g_4})\\ r_4 &= \omega_{g,m_2} - \hat{\omega}_{g_4}\end{aligned} \tag{10}$$

$$\begin{aligned}\dot{\hat{\tau}}_g &= -\alpha_{gc}\hat{\tau}_g + \alpha_{gc}\tau_{gr} - \lambda(\hat{\tau}_g - \tau_g)\\ r_5 &= \hat{\tau}_g - \tau_g\end{aligned} \tag{11}$$

where λ, k_1 and k_2 are the observer gain parameters designed to make the residual dynamics stable. In this case study, we select $\lambda = 5$, $k_1 = .05$ and $k_2 = .2$.

Zhang et al. (2011) did not describe the diagnosis step in their method. To be consistent, we use the same linear support vector classifier in the diagnosis step. The EB solution can detect and isolate the faults with 4.6% false positive rate and 16.6% false negative rate. In this case study, we use the general framework to combine the EB solution feature set and the GKSV solution feature set to propose a hybrid solution for the wind turbine diagnosis problem. In this method we use the union of GKSV features and EB residuals as the set of features. To make the comparison easy, we use the same linear SVM in the diagnosis step. The hybrid approach can detect and isolate the faults with 0.5% false positive rate and 6.7% false negative rate. This is a significant improvement compared to the original methods. Table 4 shows a comparison between the two solutions for the wind turbine benchmark competition challenge and the proposed hybrid approach.

4. CONCLUSIONS

In this paper, we proposed a general framework to represent data-driven and model-based diagnosis methods in an unified way. The proposed framework can be used to compare the diagnosis methods in a more specific way. In addition to the traditional approach of comparing the overall results, our framework makes it possible to compare a specific step in each diagnosis method. Moreover, our framework can be used to combine different diagnosis methods and build crossover solutions with better diagnosis results. We demonstrated the application of our unifying framework to propose hybrid solutions through a case study for the wind turbine challenge.

REFERENCES

Alcorta-Garcia, E., & Frank, P. (1997). Deterministic nonlinear observer-based approaches to fault diagnosis : A survey. *Control Engineering Practice 5.5*, 663–670.

Bengio, Y., Courville, A. C., & Vincent, P. (2012). Unsupervised feature learning and deep learning: A review and new perspectives. *CoRR, abs/1206.5538 1.*

Benkouider, A. M., Kessas, R., Yahiaoui, A., Buvat, J. C., & Guella, S. (2012). A hybrid approach to faults detection and diagnosis in batch and semi-batch reactors by using ekf and neural network classifier. *Journal of Loss Prevention in the Process Industries, 25 (4), 694?702.*

Biswas, G., Simon, G., Mahadevan, N., Narasimhan, S., Ramirez, J., & Karsai, G. (2003). A robust method for hybrid diagnosis of complex systems. *In Proceedings of the 5th Symposium on Fault Detection, Supervision and Safety for Technical Processes, pp. 1125-1131..*

Bregon, A., Biswas, G., Pulido, B., Alonso-Gonzalez, C., & Khorasgani, H. (2014, July). A common framework for compilation techniques applied applied to diagnosis of linear dynamic systems. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS*, *44*(7).

Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods 3, no. 1*, 1–27.

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *CM computing surveys (CSUR)*, *41*(3), 15-72.

Frank, P. M., & Ding, X. (1994). Frequency domain approach to optimally robust residual generation and evaluation for model-based fault diagnosis. *Automatica*.

Frank, P. M., & Ding, X. (1997). Survey of robust residual generation and evaluation methods in observer-based fault detection systems. *Journal of process control, 7(6)*, 403–424.

Gertler, J. (1998). *Fault detection and diagnosis in engineering systems*. CRC press.

Gertler, J. (2012). A sequential test selection algorithm for fault isolation. *10th European Workshop on Advanced Control and Diagnosis,Copenhagen, Denmark*.

Gertler, J., Li, W., Huang, Y., & McAvoy, T. (1999). Isolation enhanced principal component analysis. *AIChE Journal 45, no. 2 : 323-334..*

Hodge, V. J., & Austin, J. (2004). A survey of outlier detection methodologies. *Artificial intelligence review 22, no. 2*, 85–126.

Hofbaur, M. W., & Williams, B. C. (2004). Hybrid estimation of complex systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 34, no. 5*, 2178–2191.

Isermann, R. (1993). Fault diagnosis of machines via parameter estimation and knowledge processing– tutorial paper. *Automatica, 29(4)*, 815–835.

Jung, D., Ng, K. Y., Frisk, E., & Krysander, M. (2016). A combined diagnosis system design using model-based and data-driven methods. *In Control and Fault-Tolerant Systems (SysTol), 3rd Conference on, pp. 177-182. IEEE*.

Khorasgani, H., & Biswas, G. (2017). A combined model-based and data-driven approach for monitoring smart buildings..

Khorasgani, H., Jung, D. E., Biswas, G., Frisk, E., & Krysander, M. (2014). Robust residual selection for fault detection..

Laouti, N., Sheibat-Othman, N., & Othman, S. (2011). Support vector machines for fault detection in wind turbines..

Leturiondo, U. (2016). Hybrid modelling in condition monitoring. *PhD diss., Lulea University of Technology..*

Li, L., Gariel, M., Hansman, R. J., & Palacios, R. (2011). Anomaly detection in onboard-recorded flight data using cluster analysis. In Digital Avionics Systems Conference (DASC) IEEE/AIAA 30th, pp. 4A4-1.

Mack, D. L., Biswas, G., & Koutsoukos, X. D. (2017). Learning bayesian network structures to augment aircraft diagnostic reference models. *IEEE Transactions on Automation Science and Engineering 14, no. 1*, 358–369.

Mosterman, P. J., & Biswas, G. (1999). Diagnosis of continuous valued systems in transient operating regions. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, *29*(6), 554–565.

Narasimhan, S., Roychoudhury, I., Balaban, E., & Saxena, A. (2010). Combining model-based and feature-driven diagnosis approaches-a case study on electromechanical actuators. *21st International Workshop on Principles of Diagnosis*.

Odgaard, P. F., Stoustrup, J., & Kinnaert, M. (2013). Fault-tolerant control of wind turbines: A benchmark model. *IEEE Transactions on Control Systems Technology 21, no. 4, 1168-1182*.

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence 27, no. 8, 1226–1238*.

Samantaray, A., Medjaher, K., Bouamama, B., & Staroswiecki, M. G., Dauphin-Tanguy. (2006, Apr). Diagnostic bond graphs for online fault detection and isolation. *Simulation Modelling Practice Theory*, *14*(3), 237–262.

Sheibat-Othman, N., Laouti, N., Valour, J., & Othman, S. (2014). Support vector machines combined to observers for fault diagnosis in chemical reactors. *The Canadian Journal of Chemical Engineering 92, no. 4, 685-695*.

Talebi, H. A., Abdollahi, F., Patel, R. V., & Khorasani, K. (2009). Neural network-based state estimation of nonlinear systems: application to fault detection and isolation. *Vol. 395. Springer*.

Tidriri, K., Chatti, N., Verron, S., & Tiplica, T. (2016). Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges. *Annual Reviews in Control 42 : 63-81*.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., & Yin, K. (2003). A review of process fault detection and diagnosis: Part iii: Process history based methods. *Computers and chemical engineering 27, no. 3*, 327–346.

Yang, R., & Rizzoni, G. (2016). Comparison of model-based vs. data-driven methods for fault detection and isolation in engine idle speed control system. In *In proceedings of the annual conference of the prognostics and health management society, denver, co, usa.*

Zhang, X., Polycarpou, M. M., & Parisini, T. (2008). Design and analysis of a fault isolation scheme for a class of uncertain nonlinear systems. *Annu. IFAC Rev. Control, vol. 3, no. 2, pp. 107?121,.*

Zhang, X., Zhang, Q., Zhao, S., Ferrari, R., Polycarpou, M. M., & Parisini, T. (2011). Fault detection and isolation of the wind turbine benchmark: An estimation-based approach..

BIOGRAPHIES

Hamed Khorasgani received the B. Sc. degree in electronics and electrical engineering from Isfahan University of Technology, Isfahan, Iran, in 2009, the M. Sc. degree in mechatronics engineering from Amirkabir University of Technology, Tehran, Iran, in 2012, and the PhD in electrical engineering from the Institute for Software Integrated Systems, Vanderbilt University, Nashville, TN, USA, in 2017. During his PhD at the Institute for Software Integrated Systems, he has conducted researches in analysis complex systems and their applications in diagnosis, prognostics, and fault tolerant control. His current research projects in Industrial AI Lab at Hitachi America R&D include developing hybrid methodologies and solutions for integrating model-based and data-driven diagnosis methods.

Ahmed K. Farahat holds a Ph.D. degree from the University of Waterloo in Canada and M.Sc. and B.Sc. degrees from Cairo University in Egypt, all in Computer Engineering. Dr. Farahat is currently a senior research scientist at the Industrial AI Laboratory at Hitachi America, Ltd. Previously, Dr. Farahat was a research scientist at the Big Data Laboratory at Hitachi America, Ltd. (2014-2017). He also worked as a postdoctoral fellow at the University of Waterloo (2013-2014), a research assistant at the University of Waterloo (2007-2012) and a research engineer at IBM Egypt (2005-2007). His research interests lie in the areas of machine learning and data mining, and their applications to industrial data analytics. Dr. Farahat is the recipient of the Best Paper Award Runner-Up at the 2013 IEEE International Conference on Data Mining (ICDM), and two best paper awards in specialized workshops at ICDM′09 and NIPS′13.

Kosta Ristovski is a senior research scientist at the Industrial AI Laboratory at Hitachi America, Ltd. He is currently engaged in the development of new technologies for fleet analytics and optimization-based predictive maintenance.

Chetan Gupta is the Chief Data Scientist & Architect, and manages the Industrial AI Lab at Hitachi America R&D. He has more than 15 years of experience in analytics, AI, big data, and related domains. Over his career he has worked both as a machine learning data scientist as well as in designing systems and architectures for big data applications. At Hitachi, he manages a large team of data scientists, architects and developers that is engaged in developing cutting edge solutions and opening new frontiers in the area of industrial analytics. His team builds fundamental horizontal technologies that are then used to build solutions for industry specific verticals. He has led efforts to build horizontal solutions in predictive maintenance, quality, operations monitoring and control, and for verticals such as mobility, mining, building energy management systems, etc. Over the years Chetan has led multiple research and development teams, and mentored young researchers. He has close to 50 patents either granted or under review and more than 40 publications in the area of data mining/machine learning, data stream systems, complex event processing, workload management, etc. Chetan has a Ph.D. in Mathematics and M.S. in Mathematical Computer Science and Chemical Engineering from University of Illinois, Chicago.

Gautam Biswas received the Ph.D. degree in computer science from Michigan State University, East Lansing, MI, USA. He currently holds a Cornelius Vanderbilt Endowed Chair in Engineering and is a Professor of Computer science and Computer engineering in the Department of Electrical Engineering and Computer Science, and a Senior Research Scientist with the Institute for Software Integrated Systems at Vanderbilt University in Nashville, TN, USA. He is involved in research on intelligent systems with primary interest in hybrid modeling, simulation, and analysis of complex embedded systems, and their applications to diagnosis and fault-adaptive control. He is currently working on data-driven methods for anomaly detection in and diagnosis of complex systems. He has over 600 refereed publications and his research has been funded by the AFRL, ARL, NASA, NSF, and DARPA. Prof. Biswas is a Fellow of the IEEE and the PHM society.