Monitoring of Complex Systems of Interacting Dynamic Systems

Michael E. Cholette, University of Texas at Austin,
email: cholettm@mail.utexas.edu phone:1-248-330-1204

Jianbo Liu, Michigan State University, Lansing, MI (ryukenpa@gmail.com)
Dragan Djurdjanovic, University of Texas at Austin (dragand@me.utexas.edu)
Kenneth A. Marko, University of Michigan, Ann Arbor, MI (kamarko@umich.edu)

Monitoring of Complex Systems of Interacting Dynamic Systems

Michael E. Cholette, University of Texas at Austin (cholettm@mail.utexas.edu)
Jianbo Liu, Michigan State University, Lansing, MI (ryukenpa@gmail.com)
Dragan Djurdjanovic, University of Texas at Austin (dragand@me.utexas.edu)
Kenneth A. Marko, University of Michigan, Ann Arbor, MI (kamarko@umich.edu)

Abstract: Increases in functionality, power and intelligence of modern engineered systems led to complex systems with a large number of interconnected dynamic subsystems. In such machines, faults in one subsystem can cascade and affect the behavior of numerous other subsystems. This complicates the traditional fault monitoring procedures because of the need to train models of the faults that the monitoring system needs to detect and recognize. Unavoidable design defects, quality variations and different usage patterns make it infeasible to foresee all possible faults, resulting in limited diagnostic coverage that can only deal with previously anticipated and modeled failures. This leads to missed detections and costly blind swapping of acceptable components because of one's inability to accurately isolate the source of previously unseen anomalies. To circumvent these difficulties, a new paradigm for diagnostic systems is proposed and discussed in this paper. Its feasibility is demonstrated through application examples in automotive engine diagnostics.

Keywords: Fault detection and diagnosis, distributed anomaly detection, automotive engine diagnostics

1. Introduction

Engineering designs have been equipped with increasingly sophisticated capabilities for system performance assessment and prediction, especially in the case of sophisticated, expensive and safety critical systems. The purpose of these capabilities is to ensure proper functionality of the system.

However, the way diagnostic functions are currently realized in engineered systems of interconnected dynamic systems only has the capability of diagnosing faults, either through *a priori* knowledge or using historical data. More specifically, traditional condition monitoring methods match the behavior of a set of sensor features that are indicative of system performance with their behavior during various types of known abnormal behavior models. Such

models can be obtained from physical models, expert knowledge about the system [1][2] or training data obtained during system operation in the presence of the corresponding fault [3][4]. Recognition of those indications leads to natural isolation of the source of the underlying faulty condition and possibly an appropriate action to mitigate its adverse effects. In a way, this approach to realizing the condition monitoring functionalities in a given system essentially represents part of the system design process, where the condition monitoring (just like the system itself) is carefully designed through selection of hardware and algorithms that facilitate maximal diagnostic coverage of the abnormal behavior modes that are known to exist.

In the case of highly complex systems of interacting dynamic systems, the traditional approach to realizing diagnostic functionality becomes excessively cumbersome because of the need to train the condition monitoring processes to recognize a large number of faults, some of which often cannot be anticipated in advance. Even for the cases one is able to anticipate in advance, many faults manifest themselves very differently under different control inputs and environmental conditions, which makes training of diagnostic units for all possible conditions and all possible faults infeasible. Finally, such systems consist of numerous subsystems, each of which could contain significant non-linearities, with multiple control and environmental inputs, as well as inputs from other subsystems. This situation permits anomalies in one system to cascade and incite anomalous behavior of other subsystems connected to it, which effectively masks the real source of the anomaly.

A new paradigm will be described in this paper. We will describe methods that enable continuous detection of anomalous system behavior, isolation of the source of anomalies and identification of the underlying condition[1]. The novelty of the approach proposed is the pursuit of these functionalities regardless of the control inputs and dynamic state of the monitored system. Additionally, the localization of the source(s) of abnormal behavior is (are) conducted without the need to *a priori* know the entire set of possible faults. The work presented in this paper shows the results of fault diagnosis related research of our overall efforts aimed at realizing fault tolerant operation of complex dynamic systems consisting of multiple interacting dynamic subsystems, as illustrated in Figure 1.

The anomaly detection functionality conceptually follows the traditional model-based diagnostic approach. Abnormal system behavior is characterized as the departure of signatures of performance related sensory features from their models observed during normal system behavior.

[1] The function of performance recovery in spite of the presence of the anomaly in the system is the next logical step in realization of performance self-healing, but it will not be addressed in this paper.

The main conceptual novelty occurs at the fault localization stage, where the subsystem(s) that originated the abnormal behavior is (are) identified without the need to utilize prior signatures corresponding to the underlying fault (in the rest of the paper we will refer to this as "precedent-free fault localization"). Namely, once an anomaly is detected, multiple anomaly detectors connecting to relevant subsystems of the anomalous system will be generated, with each detector that finds an anomaly splitting even further into detectors monitoring subsystems with increased granularity. Such multiplication of anomaly detectors ultimately leads to localization of anomalous subsystems, even if the underlying fault was not observed before, as will be demonstrated in this paper.

The aforementioned precedent-free fault localization based on distributed anomaly detection necessitated the use of novel dynamic modeling and anomaly detection approaches that could separate abnormalities caused by unusual operating conditions (which are not truly anomalies) and those from true anomalies caused by changes in the internal dynamics of the monitored system. Such dynamic modeling and anomaly detection methods are another novelty originating from our work and will be discussed in this paper.

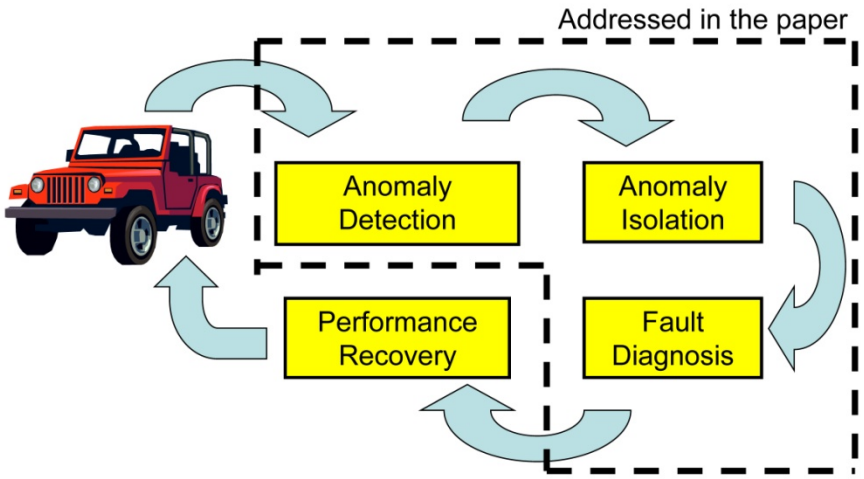


Figure 1: Functionalities for fault tolerance of systems of interacting dynamic systems[1]. The tasks enclosed in the dashed line are addressed in this paper.

Anomaly detection and localization are followed by fault diagnosis via matching of the input-output patterns of the previously isolated anomalous subsystem(s) with the models of various previously seen faults. Each time the observed patterns could not be matched with any of the existing models, a new model needs to be created in order to recognize that particular

situation in the future. This step closely resembles traditional fault diagnosis and will also be discussed in this paper.

Finally, based on the dynamic models identified through the diagnostic process, a control mechanism can be created to augment the nominal controllers in the system and restore as much as possible the original system function in the presence of a fault. Controller adaptation in the presence of a fault is our on-going work and is outside the scope of this paper.

The remainder of the paper is organized as follows. Section 2 briefly describes two different approaches to generic anomaly detection, precedent-free localization of the sources of anomalies and fault diagnosis in systems of interacting dynamic systems. Section 3 describes a framework for isolation of anomalous behavior. Section 4 describes an example of anomaly detection, precedent-free fault localization and fault diagnosis in an automotive electronic throttle system, while Section 4 describes an example of anomaly detection, fault isolation and diagnosis in the Exhaust Gas Recirculation (EGR) system of a Diesel engine. Results in both Section 4 and Section 5 have been obtained using commercially available, high fidelity simulations of automotive engines. Finally, Section 5 offers conclusions of this work and outlines directions of future research.

2. Methods for Dynamic Modeling of System Behavior and Anomaly Detection

A challenge that has been identified in this paper is that modern engineering systems, such as automotive engines or advanced manufacturing equipment, are systems of interacting dynamic subsystems. In that context, it is of primary concern to postulate models that will enable decoupling of anomalies originating from the system itself (real anomalies) from anomalous patterns caused by abnormal influences from other subsystems interacting with the monitored system. To that end, one can utilize “divide and conquer” approaches pursued in [5]-[9], where the space of features describing the behavior of the target system is indexed using features from other systems affecting it (essentially, behavior of any subsystem is indexed by clusters of features emitted by the subsystems affecting the monitored subsystem, i.e. subsystems whose outputs act as inputs to the target subsystem).

Conceptually, the problem of anomaly detection is relatively simple-- one needs to detect anomalous behavior as a statistically significant departure of the current system behavior away from the normal one. Traditional anomaly detection methods are based on global models of system behavior focus on characterizing probability distributions of behavioral features and detecting anomalies as changes in those distributions. For systems whose behavior is not affected by other subsystems of the machine, such anomaly detection approaches are

appropriate. However, interactions with other subsystems of the machine mean that feature distributions can change due to changes in the operating regime too (which should not be seen as an anomaly). Basically, changes in the upstream subsystems cause jumps from one operating regime to another (jumps from one model region to another), while changes in system dynamics (true anomalies) are visible in changes of local distributions of the features of the monitored system.

The “divide and conquer” models [5]-[9] decompose the operating space into regimes of similar dynamics, which allows one to deal with regime-switching induced non-stationarities in performance-related features by postulating relatively tractable models inside each operating regime and creating a set of region-specific anomaly detectors inside those regimes. Figure 2 illustrates the aforementioned approach to piecewise dynamic behavioral modeling and anomaly detection. Each region is equipped with an independent, region-specific detector allowing one to deal with different behavior, modeling accuracies and feature patterns due to regime switching induced non-stationarities. Dealing with such regime-specific behavior is important for performance characterization of modern dynamic systems which can consist of multiple interacting dynamic subsystems, each of which could also be a system of interacting dynamic subsystems. Interactions between various subsystems in such hierarchical dynamic systems can lead to highly non-stationary dynamic behavior, even when everything is normal inside the system.

In the remainder of this section, we will offer more details about modeling and anomaly detection methods based on the use of recently introduced piecewise statistical models of moments of time-frequency signal distributions [5], and piecewise affine models referred to as Growing Structure Multiple Model Systems (GSMMS) [6].

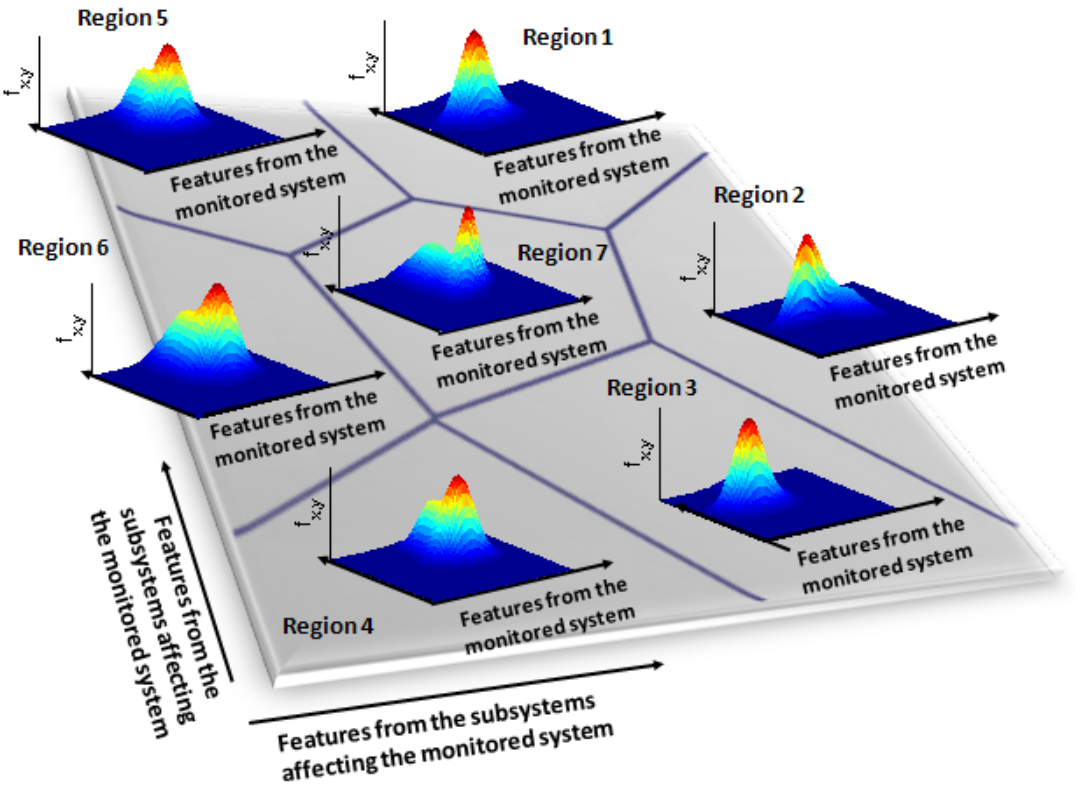


Figure 2: Schematic representation of piecewise behavior models pursued in [5]-[9]. Behavior of the monitored system features is indexed by features emitted from subsystems affecting the target system.

A. Modeling of Dynamic System Behavior and Anomaly Detection using Piecewise Statistical Models of Moments of Time-Frequency Signal Distributions

Cohen's class of time-frequency distributions has long been recognized as a powerful non-stationary signal processing method with a wide range of applications, including radar technology, marine biology, biomedical engineering and manufacturing [10]. Following the general Cohen's formulation, a joint time-frequency distribution (TFD), $C(t,w)$, of a signal $s(t)$ is [10]

$$C(t,w) = \frac{1}{4\pi^2}\iint A(\theta,\tau)\phi(\theta,\tau)e^{-j(w\tau+\theta t)}d\theta d\tau \quad (1)$$

where

$$A(\theta,\tau) = \int s^*(u-\frac{\tau}{2})s(u-\frac{\tau}{2})e^{j\theta u}du \quad (2)$$

is the ambiguity function of the signal, and $\phi(\theta,\tau)$ is the time-frequency kernel, whose properties largely determine the properties of the resulting TFDs.

Without a loss of generality, one can consider $C(t,w)$ as a two dimensional probability density function of a pair of random variables (T,W) denoting time and frequency, respectively. Following [10], one can use the first few lowest-order moments $E[W^pT^q]$ of this probability density function as its descriptive features (where $E[\cdot]$ denotes the mathematical expectation operator with respect to $C(t,w)$ as the probability density function).

Due to asymptotical Gaussianity of these time-frequency moments under constant operating conditions, the ensuing performance characterization and anomaly detection become tractable and straight forward. One possibility is to assess the system behavior corresponding to an observation of time-frequency moments $\mathbf{x}_i$ by calculating the Mahalanobis distance [11]

$$d(\mathbf{x}_i,\boldsymbol{\mu}) = \sqrt{(\mathbf{x}_i-\boldsymbol{\mu})^{\mathrm{T}}\,\Sigma^{-1}(\mathbf{x}_i-\boldsymbol{\mu})} \tag{3}$$

where Σ and μ are respectively the covariance matrix and mean vector of the time-frequency moments of the signals representing normal system behavior. Since time-frequency moments are asymptotically Gaussian, the statistical significance of the drift of a given feature vector can be expressed through the so-called Confidence Value (CV) defined as the probability that the Mahalanobis distance could be larger than the observed one [13]. A significant drop in CVs (jump in Mahalanobis distances) would then signify anomalous behavior, as was done in several manufacturing applications [12].

Nevertheless, for systems of variable interacting subsystems, each subsystem undergoes variable operating conditions due to changes in external excitations (control signals and environmental inputs). Thus, asymptotic Gaussianity of time-frequency moments cannot develop because the system is never in a steady operating mode. Essentially, what we continuously see are changes in system signatures caused by variations in the operating conditions of the system (control and environmental inputs), together with changes in system signatures that are caused by changes in system dynamics. Variations in operating regimes are not descriptive of any system degradation, whereas changes in system dynamics are and need to be detected as anomalies.

In order to decouple operating regime changes from true system anomalies, time-frequency signatures emitted by the system during the time when its behavior is deemed “normal” (“golden dataset” needed to establish the anomaly detector) are grouped according to

the operating conditions of the system. Given that behavior of a dynamic system over an interval of time can be determined by the initial conditions and inputs over that period of time, operating condition of the system at any time sample are distinguished using vectors consisting of past outputs[2], $\boldsymbol{y}$, concatenated with input sequences, $\boldsymbol{u}$, during a certain time interval. This concatenation results in a vector, $\boldsymbol{v}(k)$, which contains the dynamic information (state) of the system and the inputs at time index k defined as

$$\boldsymbol{v}(k) = [\boldsymbol{y}^T(k), \boldsymbol{y}^T(k-1), \dots, \boldsymbol{y}^T(k-n_a+1), \boldsymbol{u}^T(k-n_d), \boldsymbol{u}^T(k-n_d-1) \dots, \boldsymbol{u}^T(k-n_d-n_b+1)]^T \quad (4)$$

where n_a and n_b are the appropriate number of past outputs and inputs respectively, and n_d is the input delay. Each $\boldsymbol{v}(k)$ can be thought of as an "operating-conditions" defining vector [5] and each $\boldsymbol{v}(k)$ in the training set can be associated with the corresponding outputs over the same time interval. One could then build statistical profiles of signatures of outputs associated with "similar" operating conditions defining vectors.

The regionalization of operating conditions is conducted via unsupervised clustering of $\boldsymbol{v}(k)$ using a Kohonen Self-Organizing Map (SOM) [14]. The SOM utilizes a set of weight vectors, $\boldsymbol{\xi}_m$, $m = 1,2, \dots M$, that induce the Voronoi tessellation of the space of $\boldsymbol{v}(k)$, and consisting of disjoint sets V_m corresponding to each SOM weight vector $\boldsymbol{\xi}_m$

$$V_m = \left\{ \boldsymbol{v} : \|\boldsymbol{v} - \boldsymbol{\xi}_m\| \leq \left\|\boldsymbol{v} - \boldsymbol{\xi}_j\right\|, \forall\, j \neq m \right\}. \quad (5)$$

An example of a SOM-induced Voronoi tessellation can be seen in Figure 3. Vectors of inputs and initial conditions that belong to the same Voronoi set are similar and define a steady operating regime within which one could invoke asymptotic Gaussianity of time-frequency moments of the corresponding outputs. Therefore within each operational region (Voronoi set), the corresponding output sequences can be used to create a statistical profile representing the expected system in that region based on their time-frequency moments.

[2] Note that the initial conditions can be replaced by an appropriate number of past output samples, since estimation of high order derivatives can incur large errors under noisy environment

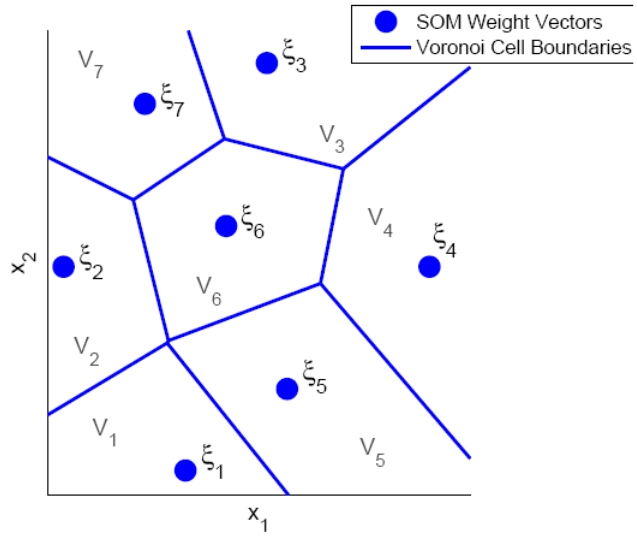


Figure 3: A SOM-induced Voronoi tessellation of $\boldsymbol{x} = [x_1, x_2]^T$. The seven SOM nodes define seven regions of simliar vectors.

During the training phase, the SOM nodes can be recursively adjusted each time the SOM is presented with a new training vector, $\boldsymbol{v}(k)$, as

$$\boldsymbol{\xi}_m(k+1) = \boldsymbol{\xi}_m(k) + \alpha\,(k) h\left(k, \text{dis}\big(m, b(k)\big)\right)[\boldsymbol{v}(k) - \boldsymbol{\xi}_m(k)] \tag{6}$$

where $b(k)$ denotes the SOM node whose weight vector is the nearest to the newly training vector $\boldsymbol{v}(k)$, dis$(m, b(k))$ is the topological distance between region *m* and $b(k)$ on the SOM graph and $h(k,$ dis$(m, b(k))$ is the neighborhood function that enables each weight vector to be updated using training samples in neighboring regions. Usually, the neighborhood function peaks at the best matching node and tapers off away from it. A typical neighborhood function is of the form

$$h\big(k, dis(m, b(k))\big) = \exp\left(\frac{-\text{dis}\,(m, b(k))^2}{2\sigma^2(k)}\right) \tag{7}$$

where the width parameter, $\sigma(k)$, is usually taken to be larger in the initial stages of training and tends to zero as k → ∞ to achieve convergence and global ordering of the SOM [14].

Thus, a SOM can be trained on a set of $\boldsymbol{v}(k)$ emitted during acceptable operation and the time-frequency moments can be utilized to characterize the system performance in each induced Voronoi region, constituting a simple regionalized anomaly detector. Figure 4 summarizes the time-frequency based anomaly detection architecture in the generic framework

proposed in this paper. For any new signal, the anomaly detector first identifies the operational region within which the system is operating by finding the best matching unit in the trained SOM using the concatenated vector of initial conditions and input sequence, $\boldsymbol{v}(k)$. Inside that region, one can compare the time-frequency moments of the corresponding output sequence with the statistical profile time-frequency moments of the outputs observed during training in that region (normal behavior of time-frequency moments in that region).

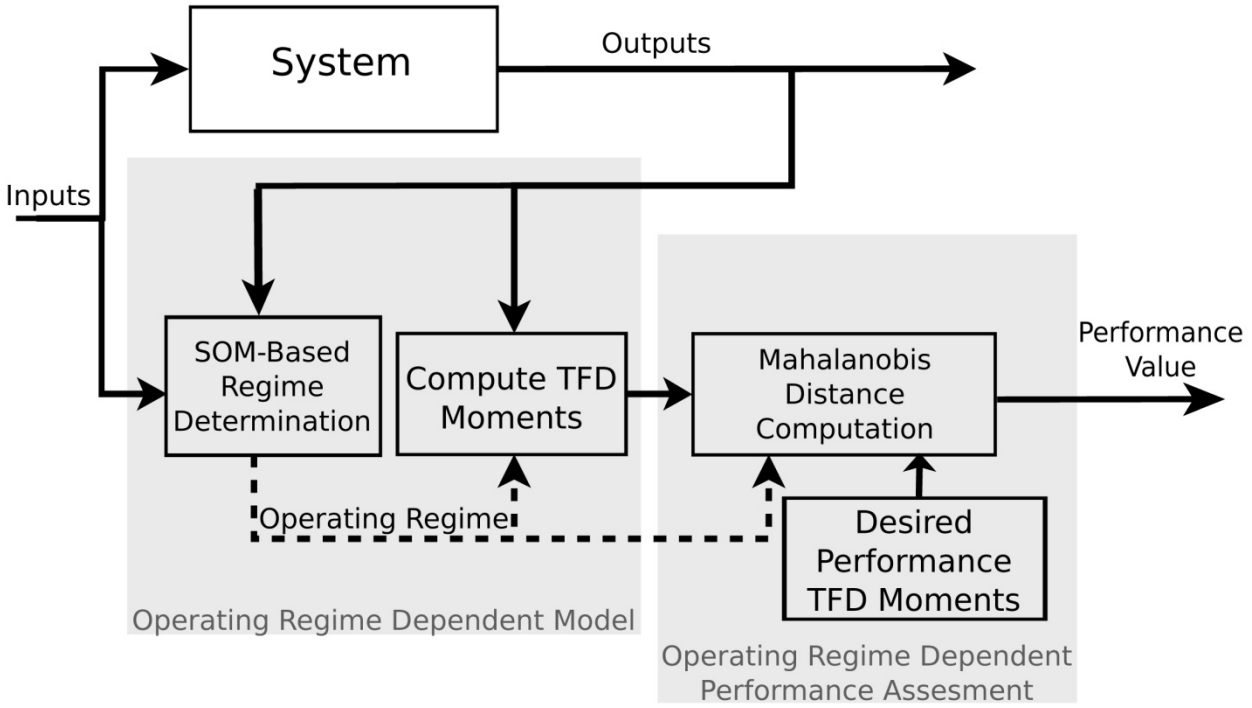


Figure 4: Anomaly detection using local time-frequency analysis based performance assessment.

At the same time, a quantization error is computed as the distance of the vector of inputs and initial conditions away from the weight vector associated with the best matching unit in the SOM. If the quantization error is smaller than a preset threshold, the anomaly detection system is triggered and the time-frequency moments of the corresponding output sequence are compared with their normal statistical profile within that region. If the quantization error is too big, anomaly detection is not invoked. Such triggering based on the quantization error is used to avoid the situations when the current system operating condition has not been experienced before (when quantization errors are too big) and thus avoid making incorrect decisions.

B. Growing Structure Multiple Model Systems based Modeling and Anomaly Detection

The Growing Structure Multiple Model System (GSMMS) [6] can be seen as a collection of local models $\mathbf{F}_i, i = 1,2,\dots,M$, for which the global model for a system with n outputs $\boldsymbol{y}(k) = [y_1(k), y_2(k), \dots, y_n(k)]^T$ and p inputs $\boldsymbol{u}(k) = \left[u_1(k), u_2(k), \dots, u_p(k)\right]^T$ at any time sample k is:

$$\boldsymbol{y}(k+1) = \sum_{m=1}^{M} \nu_m\big(\boldsymbol{s}(k)\big)\, \mathbf{F}_m(\boldsymbol{s}(k)) \tag{8}$$

where

$$\boldsymbol{s}(k) = [\boldsymbol{y}^T(k), \boldsymbol{y}^T(k-1), \dots, \boldsymbol{y}^T(k-n_a+1), \boldsymbol{u}^T(k-n_d), \boldsymbol{u}^T(k-n_d-1) \dots, \boldsymbol{u}^T(k-n_d-n_b+1)]^T,$$

is the similar to $\boldsymbol{v}(k)$ and contains the dynamic input/output information of the system. The local models have a tractable affine form

$$\mathbf{F}_m\big(s(k)\big) = \boldsymbol{a}_m^T \boldsymbol{s}(k) + \boldsymbol{b}_m \tag{9}$$

with local model parameters $\boldsymbol{a_m}$ and $\mathbf{b}_\mathrm{m}$, and the mixing function $\nu_m\big(\mathbf{s}(k)\big)$ defines the regions of validity of each local model. The function $\nu_m\big(\mathbf{s}(k)\big)$ is taken to be a simple gating function

$$\nu_m\big(\boldsymbol{s}(k)\big) = \begin{cases} 1, & \boldsymbol{s}(k) \in V_m \\ 0, & \text{otherwise} \end{cases}$$

with V_m denoting the Voronoi sets induced by the underlying SOM, as in Eq. (5).

The use of a SOM-based unsupervised clustering for partitioning of the operating space overcomes the drawbacks of *ad-hoc* or "variable-by-variable" based approaches to operating space partitioning in "divide-and-conquer" models [15]-[17]. In addition, the use growing mechanisms such as those reported in [18]-[20] enable one to determine the number of local models needed to approximate the underlying non-linear dynamics.

During the training process, both the structural parameters of the model (number and position of weight vectors) and local model parameters must be determined. With each training item $\boldsymbol{s}(k)$, the locations of the SOM weight vectors are updated using the recursive relation

$$\boldsymbol{\xi}_m(k+1) = \boldsymbol{\xi}_m(k) + \alpha\,(k)\zeta_m(k)h\big(k,\mathrm{dis}(m, b(k))\big)[\mathbf{s}(k) - \boldsymbol{\xi}_m(k)] \tag{10}$$

$\boldsymbol{s}(k)$ is the k^{th} training sample, dis$(m, b(k))$ is the topological distance between region *m* and the node $b(k)$ that is nearest (in the Euclidean sense) to the training item $\boldsymbol{s}(k)$ and $h(k,$ dis$(m, b(k))$ denotes the neighborhood function that serves the same purpose and the same form as the neighborhood function, Eq. (7), described in Section 2.A. One also notes the similarity between the SOM updating Eqs. (6) and (10), which both utilize the neighborhood functions, quantization errors and a learning rate. However, since the goal of the GSMMS is minimization of the modeling errors, a penalty term

$$\zeta_m(k) = \frac{e_m(k)}{\sum_{m=1}^{M} e_m(k)} \tag{11}$$

is included in Eq. (10) where $e_m(k)$ are the root-mean square modeling error in region *m* ensures that the weight vectors tend to move towards regions with higher modeling errors, thus leading to a finer partition of the operating space in areas of high nonlinearity.

The local model parameters in region *m* are determined by minimizing

$$J_m(\boldsymbol{a}_m, \boldsymbol{b}_m) = \frac{1}{k}\sum_{i=1}^{k} w_m\big(s(i)\big)\big\|\boldsymbol{y}(i) - \hat{\boldsymbol{y}}_m(i)\big\|^2 \tag{12}$$

where $\boldsymbol{y}(i)$ is the training output at time sample *i* and $\hat{\boldsymbol{y}}_m(i)$ is the predicted output of regional model m at time i. The function $w_m(s(i))$ determines the weight of the modeling error associated with sample i for the cost function in region m, allowing each training sample $s(i)$ to update a neighborhood of regional models. The weighting function

$$w_m(s(i)) = exp\left(\frac{-dis(m,b(i))^2}{2\sigma(k)^2}\right) \tag{13}$$

ensures that each training sample $s(i)$ has the largest effect on the model parameters near the SOM node $b(i)$ that is closest to that training sample, and a smaller effect farther away from it. This cooperation allows the GSMMS to utilize each training sample to affect all local models. Such regional cooperation has been shown to have the ability to speed convergence in the early stages of training as well as increase the bias in later training states [21] (which is why $\sigma(k)$ was reduced to narrow w_m with each subsequent epoch).

During training, the state space partition can be further refined by adding a node half way between the GSMMS node corresponding to the poorest modeled region (largest RMS error) and its immediate neighbor with the largest corresponding RMS error. The growth

mechanism essentially follows the growing cell structure method proposed in [19]. The difference is that it uses modeling errors as the insertion criterion, instead of visitation frequency or quantization errors. Finally, a stopping criterion can be applied to terminate the training, such as the RMS error falling below a pre-determined tolerance, or the number of SOM nodes exceeding a pre-determined number.

One can see that the GSMMS essentially casts the problem of representing the system dynamics into the framework of simple, interconnected dynamic linear models. This structure enables the modeling of a wide variety of complex systems while maintaining analytical tractability and an operating regime decomposition that enables regionalized anomaly detection. The GSMMS approach has been used successfully for modeling of an electronic throttle system in a gasoline engine [6], automotive crankshaft dynamics [7], diesel engine Exhaust Gas Recirculation (EGR) system and its subsystems [8] and electrical portion of an alternating current generator [9].

Once the GSMMS model of normal behavior is built, anomaly detection can be accomplished through comparison of the statistical characteristics of its residuals[3] displayed during normal behavior with characteristics of the most recent modeling residuals. Since the operating space is decomposed into regions of similar dynamic behavior (regions of "similar" input/output vectors $\boldsymbol{s}$), each region of the GSMMS can be equipped with its own decision making scheme that quantifies how close the current residual pattern is to the normal pattern. Following [6], the performance within each region *m* will be quantitatively described in this paper using the concept of regional confidence values (CVs) defined as

$$\mathrm{CV}(m,k) = \frac{|g_m(e,k)\cdot f_m(e)|}{\|g_m(e,k)\|\|f_m(e)\|} \tag{14}$$

where $f_m(e)$ is the probability density function (PDF) of the modeling residuals displayed during normal behavior and $g_m(e,k)$ is the PDF of the residuals corresponding to the current behavior at time k. The regional confidence value, $\mathrm{CV}(m,k)$, describes a normalized area of overlap of the PDFs in that region, a graphical illustration of which can be seen in Figure 5. It is easy to see that $\mathrm{CV}(m,k)$ = 1 if the current residual PDF is exactly the corresponding normal behavior PDF and less than 1 otherwise. The PDF $f_m(e)$ was approximated using Gaussian Mixture Models due to their universal approximation capability [22] and $g_m(e,k)$ was calculated by updating the regional PDFs recursively during operation [23]. Using the Gaussian Mixture Model formulation,

[3] The modeling residuals are differences between the system output and the output of the GSMMS describing the normal system behavior.

a closed form solution for $\mathrm{CV}(m,k)$ can be obtained to quickly compute $\mathrm{CV}(m,k)$ at each time step, *k*.

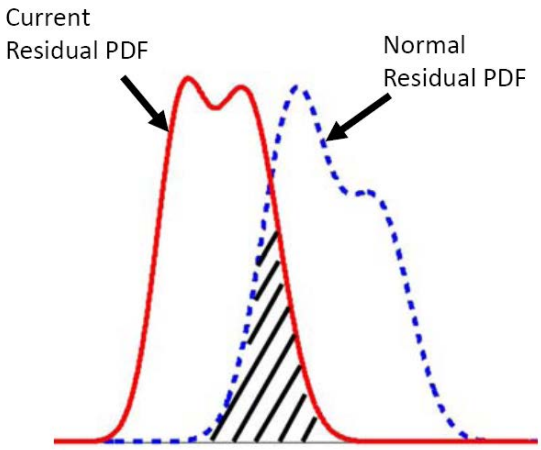


Figure 5: Illustration of the Confidence Value (CV) viewed as an overlap of probability density functions. The hatched area represents the CV.

The GSMMS based anomaly detection approach is summarized in Figure 6, where the SOM-based regionalization and the local model prediction of the GSMMS serve as the operating regime dependent modeling scheme and the regional CVs serve as the performance assessment.

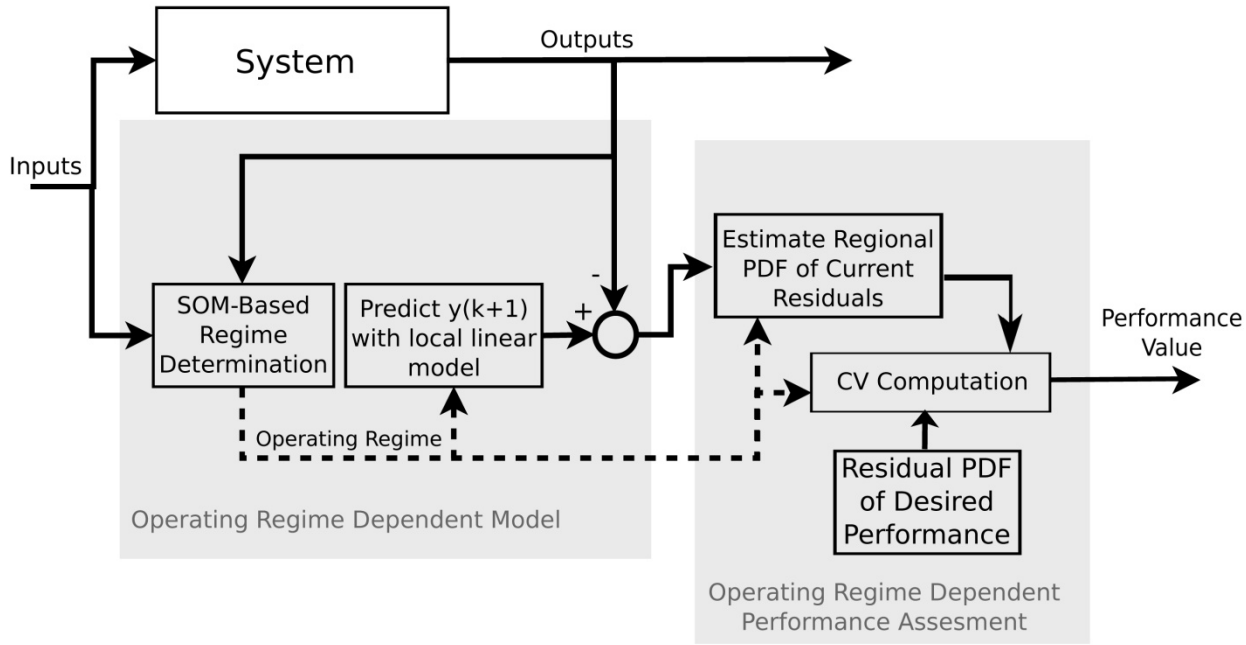


Figure 6: Anomaly detection using the GSMMS.

3. Methods for Isolation of the Source of an Anomaly and Fault Diagnosis

Isolation of the anomaly source can be conducted by reconfiguring and reconnecting anomaly detectors (ADs) to subsystems in the anomalous system. This approach is depicted in Figure 7 for a generic system of 2 subsystems. Once the AD monitoring the entire system detects an anomaly by an unusual drop in CV-s, it is replaced with ADs that monitor the constituent subsystems (in this case, 2 subsystems). Only the CV-s from AD-s monitoring anomalous subsystems will be low. Clearly, this process could be repeated multiple times, until anomalous subsystem(s) is(are) isolated to the smallest possible granularity, i.e. the culprit subsystem or subsystems represent Field Replaceable Units (FRUs) or it is not possible to continue with anomaly detection deeper into pertinent subsystems.

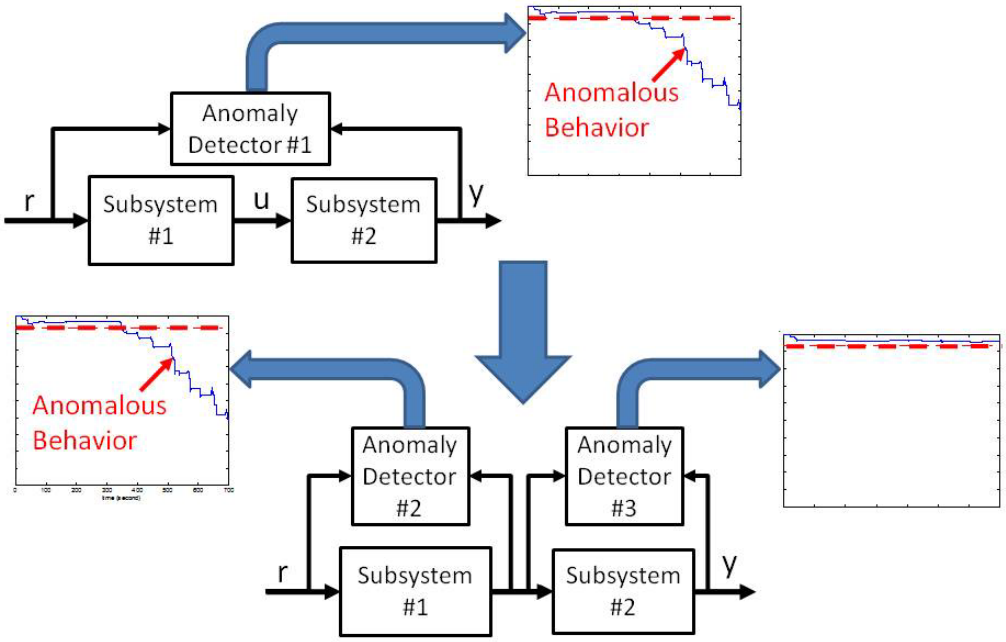


Figure 7: Illustration of fault isolation through distributed anomaly detection. In this example, the fault is in subsystem 1.

Once the anomaly is localized to a subsystem, the next step is to recognize the underlying fault if the model of that fault exists, or to recognize that the underlying fault is an unknown fault for which a model does not exist. A diagnoser for a specific fault can be constructed following essentially the same approach pursued in Section 2 for the purpose of anomaly detection. Signatures emitted in the presence of the fault that the diagnoser needs to recognize can be utilized to estimate the PDFs of the modeling features in the presence of that fault (time-

frequency moments or residuals of the GSMMS corresponding that fault). Proximity of the most recent system behavior to that fault can then be evaluated via the overlap of the PDF characterizing the most recent faulty model features and PDF characterizing faulty model features when that fault is indeed present. Whenever this CV-like value for a specific fault model grows high, one can conclude that the corresponding fault is present and vice versa. If none of the existing diagnosers are displaying a high CV, the presence of an unknown fault can be inferred and a new fault model (time-frequency based or GSMMS) would need to be developed for the current system behavior.

4. Time-Frequency Analysis Based Anomaly Detection, Fault Isolation and Diagnosis in an Automotive Electronic Throttle System

This section details an example of anomaly detection, precedent-free fault localization and fault diagnosis based on the use of piecewise time-frequency analysis models of system behavior described in Section 2.A. This method was applied to an automotive electronic throttle control system simulated using the Gasoline Engine Vehicle Model (GEVM) software package for high fidelity automotive simulations [24]. Note that the controller used in the simulation can be readily embedded into actual automotive engine control units. The target system consists of two interacting subsystems – the throttle system (plant) and the relevant controller. The two subsystems are decoupled by distributing two time-frequency based anomaly detectors to monitor each of the pertinent subsystems, as illustrated in Figure 8.

The signals are collected at the sampling rate of 100 Hz. The anomalies are simulated by modifying the parameters in the controller and the plant. Different levels of parameter changes are introduced to simulate different levels of degradation. Simulations are run using various standard driving profiles. To investigate the effects of noise, 5% multiplicative noise has been imposed on the outputs of the controller and the plant. This corresponds to a signal to noise ratio of 26dB.

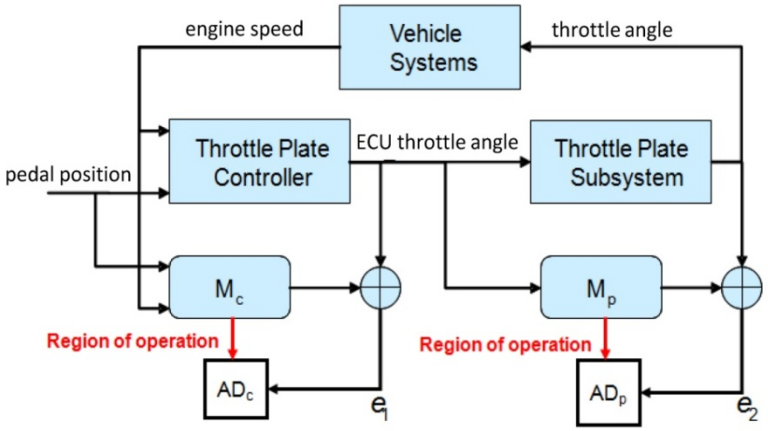


Figure 8: Distributed anomaly detectors for an anomaly detection and isolation in an Electronic Throttle System

A. Anomaly Detection and Fault Isolation

To demonstrate the capability of precedent-free fault isolation through distributed anomaly detection, two anomaly detectors are applied separately to the controller and the plant as shown in Figure 8. Three different test scenarios are introduced by gradually changing the actual system parameters over time to simulate the system degradation, as illustrated in Figure 9. In each scenario only one parameter is varied in time while all the other parameters are kept at their nominal value.

Figure 10 shows the outputs of the two anomaly detectors for the three introduced scenarios. The straight line across the window is the steady state lower 3σ control limits of the EWMA control chart applied to the performance CVs obtained during training [25]. It can be easily observed from Figure 10 that in scenario 1, the CVs from the anomaly detector operating on the controller are high all the time, while the CVs from the anomaly detector on the plant gradually decrease and finally exceed the control limit. This indicates that an anomaly has occurred in the plant, while the controller is still operating normally. Similar behavior can also be observed in scenario 2. In scenario 3, a disturbance to the controller was introduced and, as expected, the performance CVs from the controller anomaly detector eventually exceed the control limits, whereas the performance CVs from the plant detector remain within the control limits. This way, one can readily localize the anomalous behavior to the plant, the controller or both.

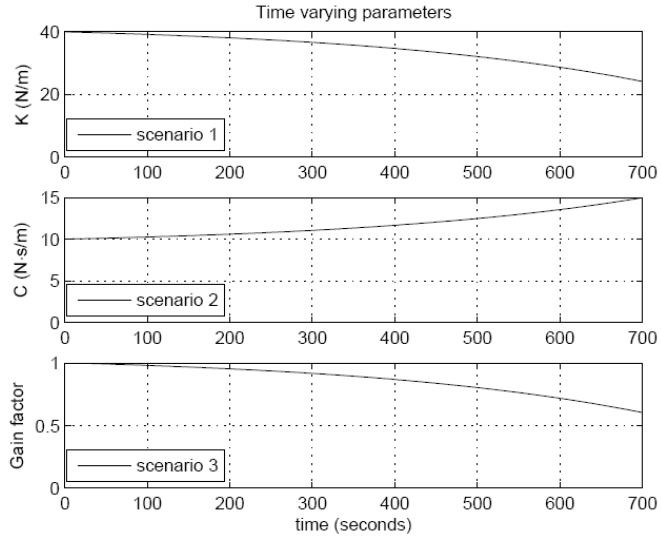
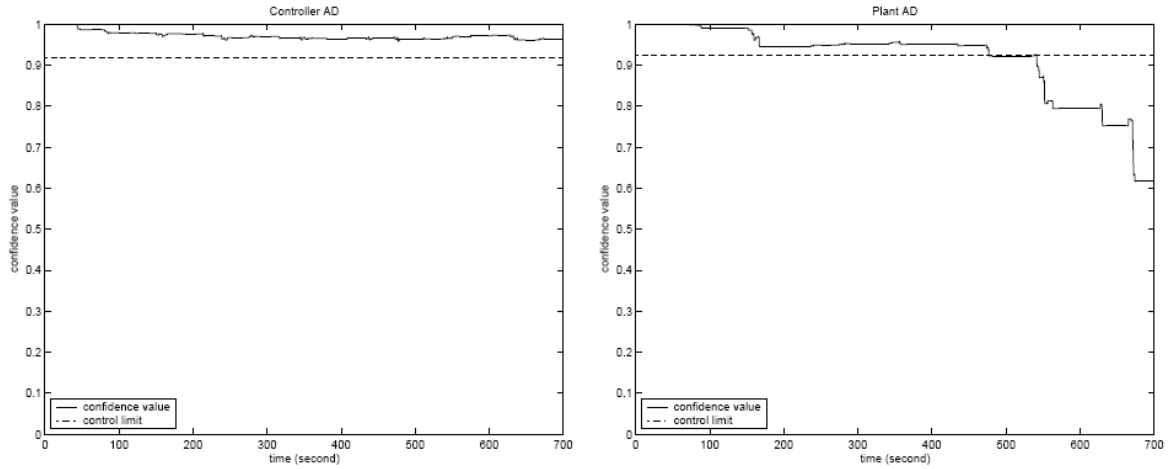
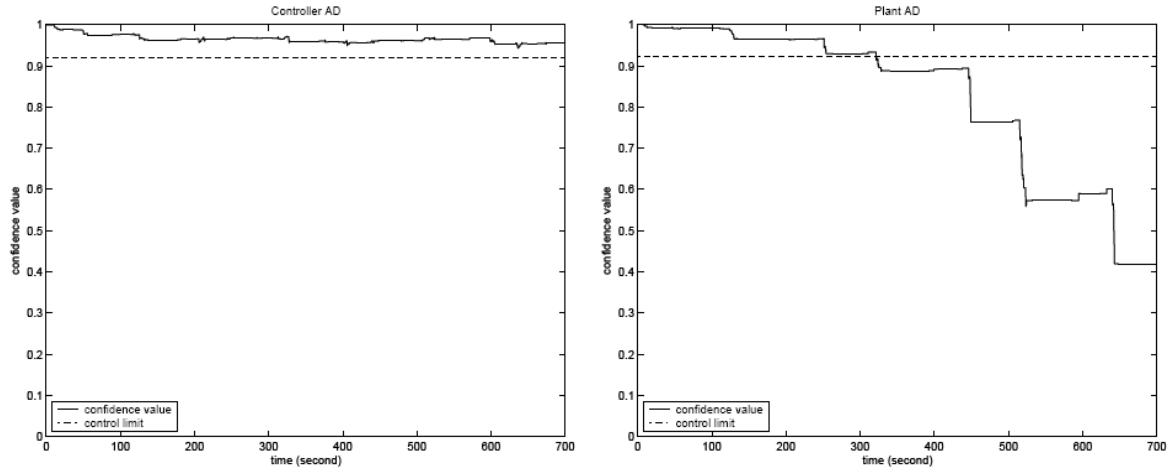


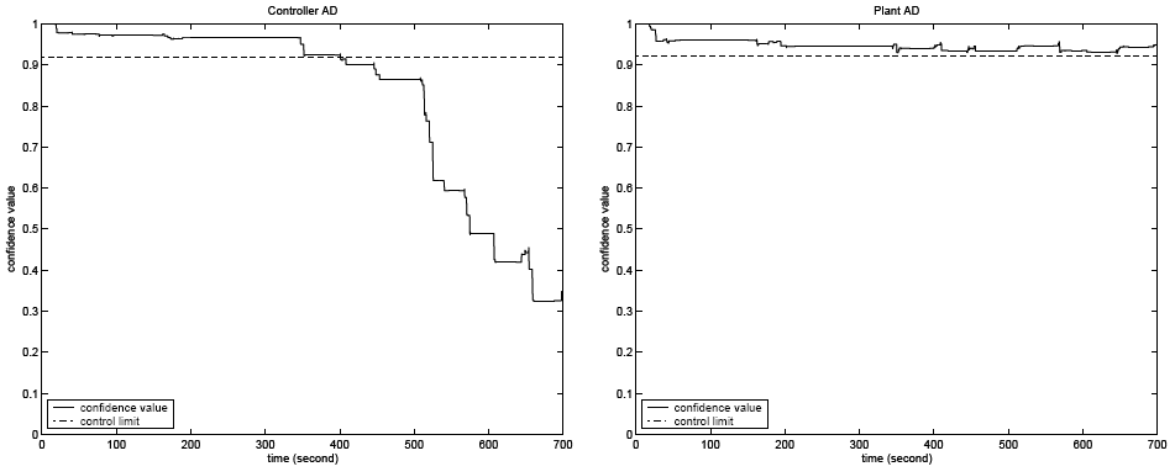
Figure 9: Time varying parameters for scenarios 1, 2 and 3. The nominal values for stiffness (plant parameter), viscous damping (plant parameter) and gain (controller parameter) are 40N/m, 10Ns/m and 1, respectively.



(a) Scenario 1: Stiffness *K* (plant parameter) decreases from its normal value 40N/m to 24N/m in 700s.



(b) Scenario 2: Viscous damping *C* (plant parameter) increases from its normal value 10N·s/m to 15N·s/m in 700s.



(c) Scenario 3: Controller gain decreases from its normal value of 1 to 0.6 in 700s.

Figure 10: Experimental results for the three designed scenarios. Note that the confidence values are updated only when the anomaly detectors are triggered.

B. Fault Diagnosis

As discussed in Section 3, fault isolation is conducted through the use of multiple anomaly detectors distributed throughout a control system. The anomaly can be localized through the identification of the lowest level segment of the system in which an AD has announced the presence of anomalous behavior. This scheme was used in the previous section to isolate the faults between the controller and the plant.

Section 3 also describes a deeper level root cause identification, in which specialized diagnosers identify specific failure modes. In this approach, separate diagnosers are specifically trained to detect specific faults. The fault is then isolated by the simultaneous drop in the confidence value of the anomaly detector (measuring proximity of system behavior to the

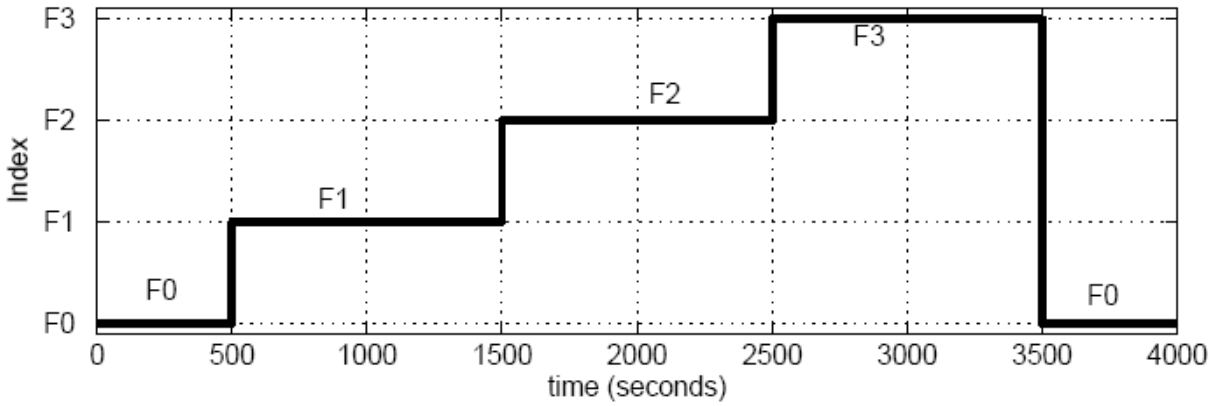
normal behavior), along with the growth in the confidence value of the diagnoser associated with the actual fault (indicating proximity of system behavior to behavior corresponding to that particular fault).

This concept is practically illustrated in the following example. Table 1 shows the operating conditions for various normal and failure modes that were simulated. It was assumed that the input-output signals from only the states F0, F1 and F2 are available for training, while signatures from the state F3 were unknown. Three detectors corresponding to states F0, F1 and F2 were trained as described above, using data obtained from simulations of standard US and European test driving profiles. Each of the detectors is then used to describe the statistical profiles of a particular plant condition.

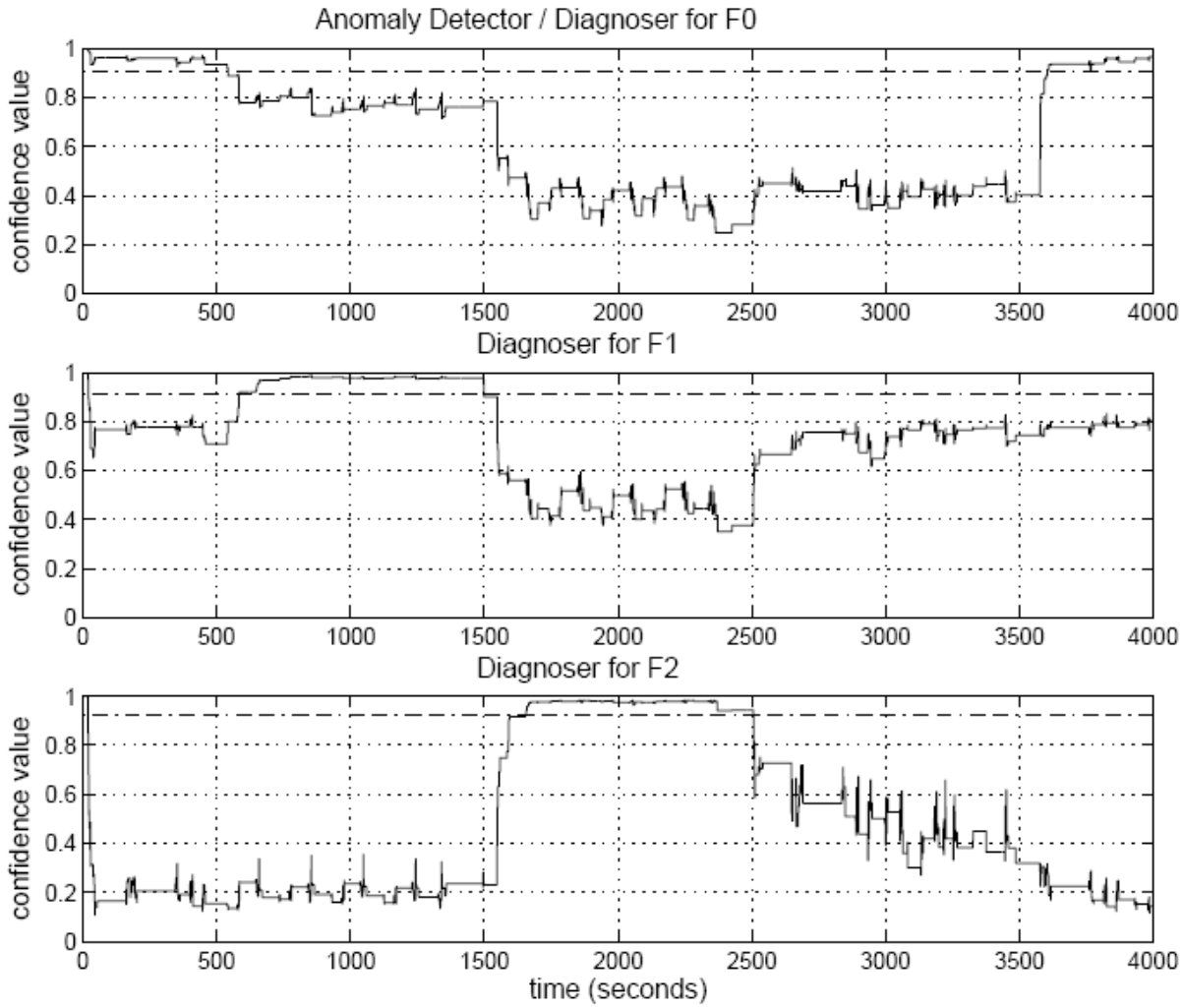
Table 1: Labels for various simulated situations

Fault Label	Description
F0	Normal
F1	Reduced *K*
F2	Increased *C*
F3	Unknown fault (reduced saturation limit on throttle motion, which corresponds to dirt accumulation in the throttle chamber)

Figure 11(a) shows the time intervals during which operating states *F*0, *F*1, *F*2 and *F*3 were introduced. Figure 11(b) shows the confidence values calculated using the diagnosers of the three known situations. In each case, the appropriate diagnoser registers the occurrence of the correct condition within the proper interval. When the unknown condition *F*3 appears, the low CVs from all diagnosers indicate the presence of an anomaly that was not seen before.



(a) Sequence of simulated states in the electronic throttle system.



(b) Performance confidence values calculated from diagnosers specialized for states *F*0, *F*1 and *F*2.

Figure 11: Fault diagnosis in a simulated automotive electronic throttle system.

5. GSMMS based Anomaly Detection, Fault Isolation and Diagnosis in a Diesel Engine Exhaust Gas Recirculation (EGR) System

Exhaust Gas Recirculation (EGR) is widely used as a method to reduce NOx emissions. In an EGR system, a portion of the exhaust gas is introduced into the intake. This proportion is controlled by the EGR valve, which is set by the engine control unit based on the current operating conditions (engine speed and load). A simplified schematic of the EGR system is shown in Figure 12 and the block diagram of the EGR system is shown in Figure 13.

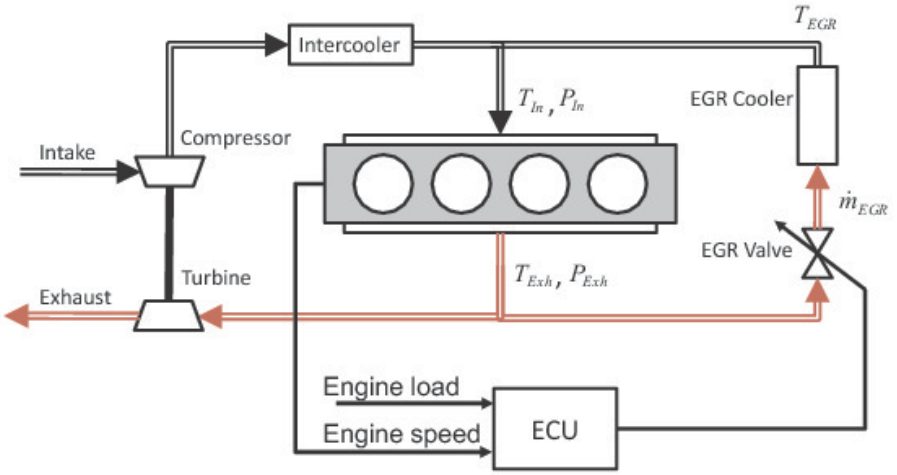


Figure 12: Schematic diagram of a generic EGR system

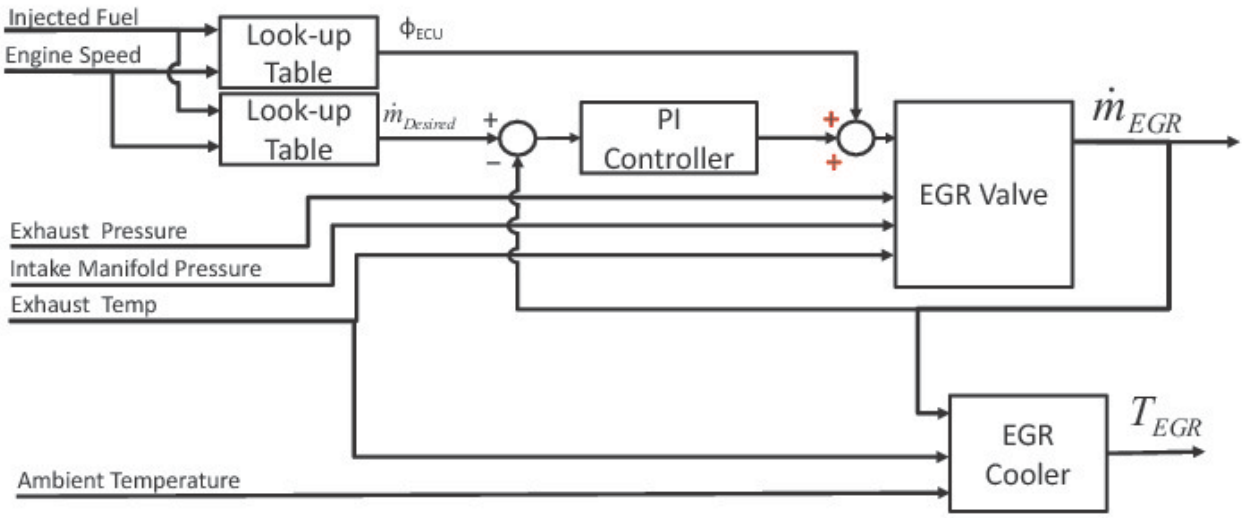


Figure 13: Block diagram a generic EGR system

The GSMMS based anomaly detection, isolation and fault diagnosis procedure outlined in Section 2.B and 3 was applied to the EGR system of a four cylinder turbo charged diesel engine. The software package en-DYNA THEMOS CRTD 2.0 by TESIS was used to conduct the engine simulations [26]. The controller consists of two look up tables determining the mass flow $\dot{m}_{desired}$ and the coarse control for the throttle angle. A proportional-integral (PI) controller is then used to correct the throttle angle such that the desired mass flow is achieved [26]. The look-up table values and PI controller parameters were provided with the default calibration.

En-DYNA was used to generate engine data and the signals were polluted with 2% additive noise. For each of the blocks in Figure 13, a GSMMS for normal behavior was

constructed. Three standard driving profiles (ECE-15, FTP-75 and MVEG-B) were used to generate training data and a fourth profile (Japan 10-15) was used to test the generalization capability of the model.

The GSMMS models of normal and faulty behavior modes were trained offline using inputs and outputs from the relevant subsystems. For each GSMMS, the appropriate inputs were selected and autoregressive orders of the local models were provided. One should note that recent work has shown that even input selection can be automated [27], further decreasing the requisite *a priori* knowledge. The training procedure followed the batch algorithm from [6], with the stopping criterion being a maximum SOM size.

A. Anomaly Detection and Isolation

Several abnormalities were introduced into the plant and controller of the EGR system. The methodology described in Section 2.B and 3 was then used to detect and isolate the anomalies with no prior knowledge of the faults or training based on the anomalous behavior data.

To simplify the monitoring scheme, the regional CVs (computed using Eq. (14)) were merged into a global CV defined as the geometric mean of all regional CVs at any given time. The geometric mean was selected to emphasize individual departures from normal behavior (CV = 1) since a decrease in any CV indicates that the system is behaving abnormally (at least, in the region where the local CVs are low).

An anomaly was introduced to the EGR valve to simulate a clogged (or improperly opening) EGR valve. In the simulations, this anomaly was introduced by modifying the throttle characteristic curve that describes the mass airflow across the throttle as a function of the throttle angle as shown in Figure 14.

Initially, an overall anomaly detector ($AD_{overall}$) was set to monitor the entire EGR system as shown in Figure 15. Once it detected an anomaly, the fault isolation method of Section 3 was applied successively until the anomalies were isolated to the smallest possible subsystem (individual blocks in Figure 13). Thus, after detection, two anomaly detectors $AD_{mass\ flow}$ and AD_{cooler} were formed to monitor the mass flow and cooler subsystems, respectively, as shown in Figure 16. Finally, if $AD_{mass\ flow}$ detected an anomaly, it split into the third level of anomaly detectors shown in Figure 17. On this third level, AD_3 and AD_4 monitored the look-up tables, AD_5 monitored the PI controller and AD_6 was set to track the EGR valve behavior. The results of

$AD_{mass\ flow}$ and AD_{cooler} along with AD_3 through AD_6 are used to isolate the fault to the subsystem with the smallest possible granularity, once $AD_{overall}$ and $AD_{mass\ flow}$ detected an anomaly.

The global CVs for $AD_{overall}$ in the presence of a throttle fault can be seen in Figure 18 while the global CVs for the second level can be seen in Figure 19. Finally the third level of global CVs (AD_3 through AD_6) can be seen in Figure 20. It is obvious from these figures that interpretation of $AD_{mass\ flow}$ and AD_{cooler} and AD_3 through AD_6 localizes the fault to the EGR valve, even though no prior training of models for faulty behavior was done.

An interesting observation is that $AD_{mass\ flow}$ is not as sensitive to plant faults as the valve anomaly detector, AD_6. The reason for this is that $AD_{mass\ flow}$ monitors the controlled EGR system and the PI controller compensates for some of the adverse effects of the valve anomaly.

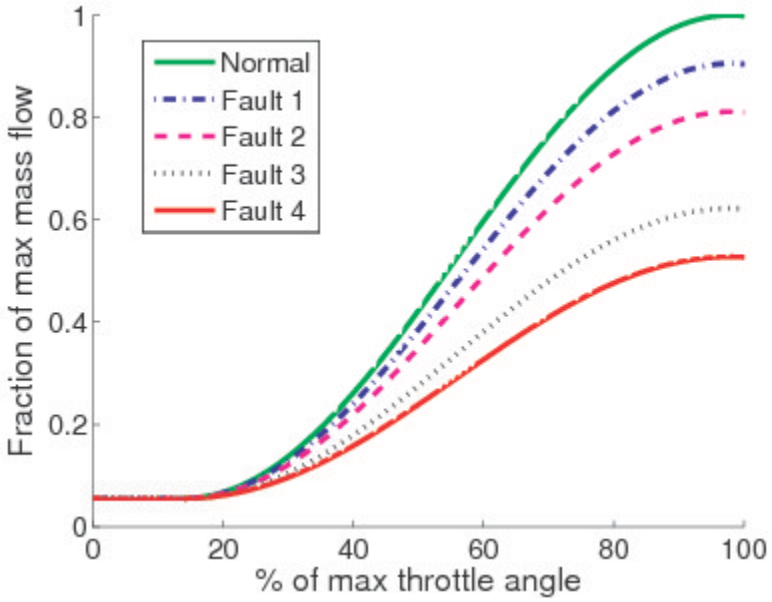


Figure 14: Throttle characteristic curve for various anomalies labeled Faults 1-4

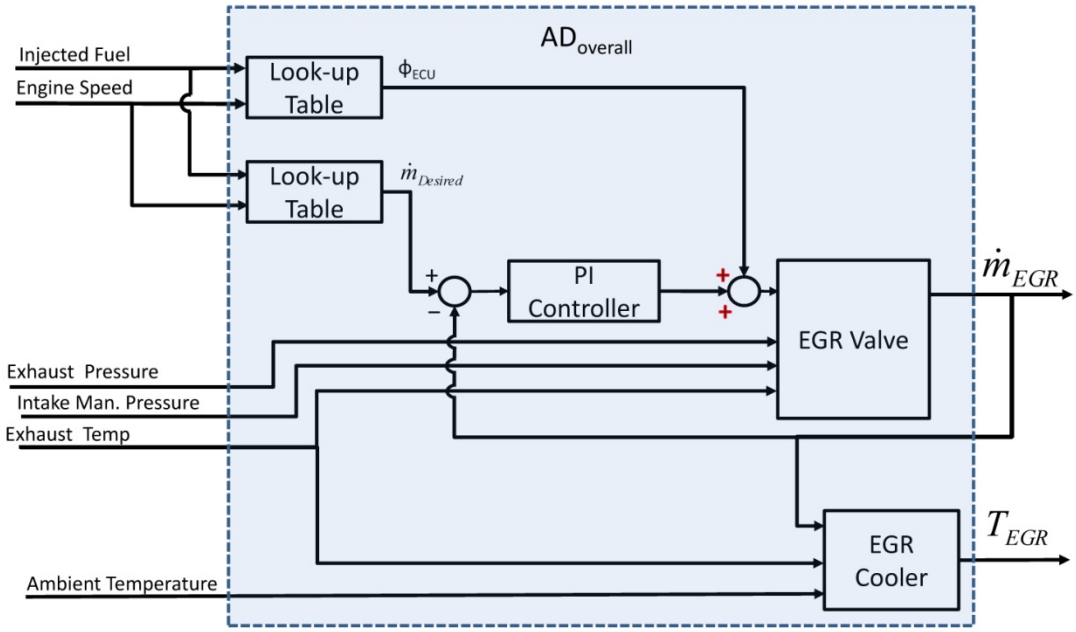


Figure 15: Block diagram with Anomaly Detector monitoring entire EGR system.

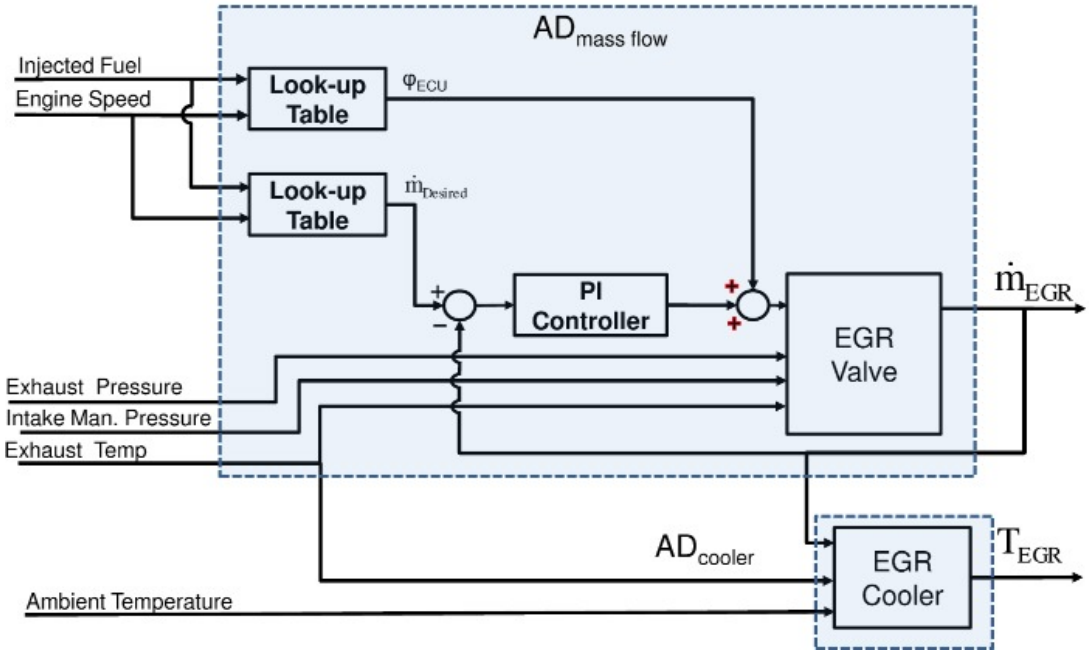


Figure 16: Block diagram with second level distributed anomaly detectors utilized for anomaly isolation.

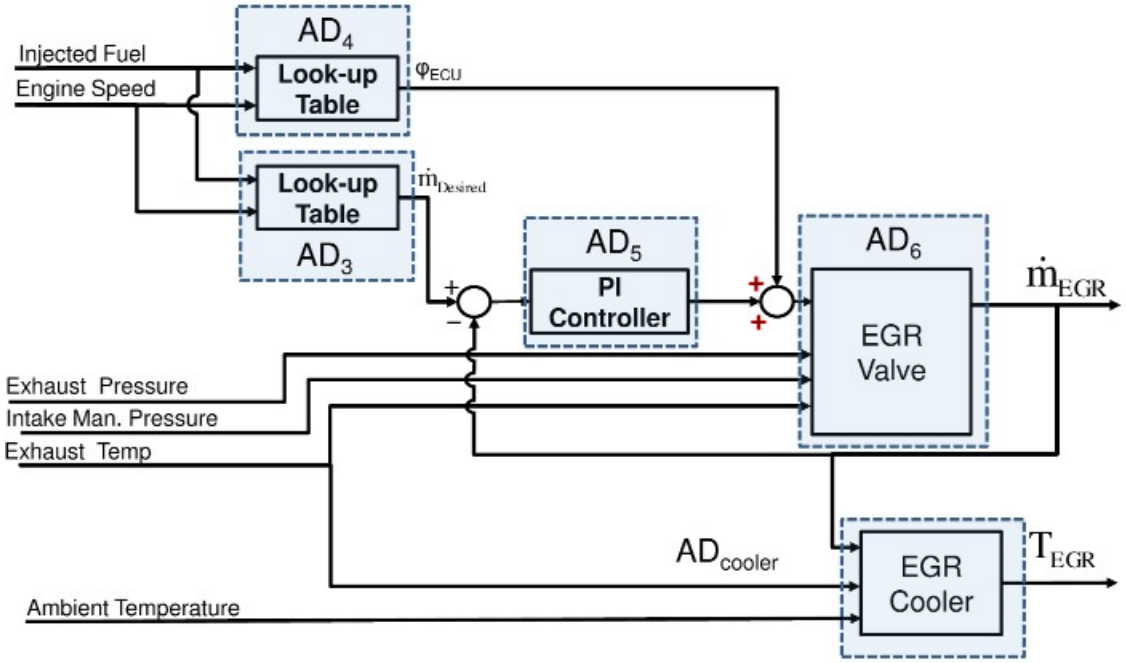


Figure 17: Block diagram with third level distributed anomaly detectors utilized for anomaly isolation.

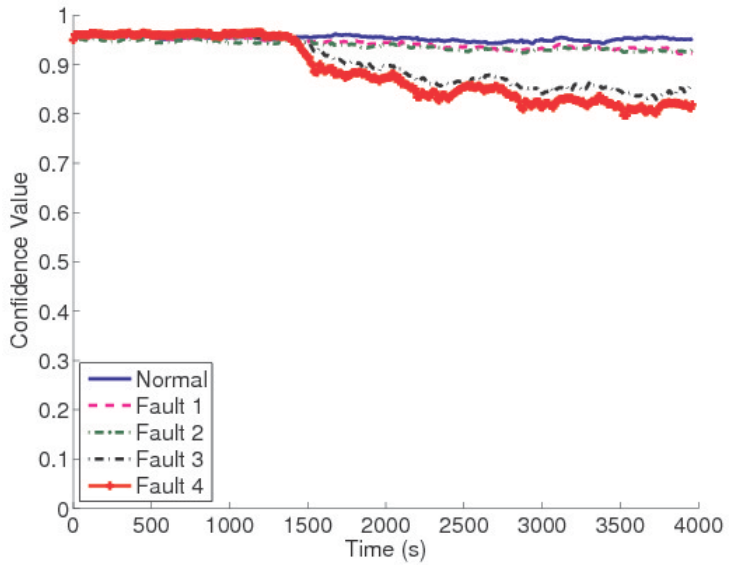


Figure 18: Confidence values for $AD_{overall}$ in the presence of valve faults.

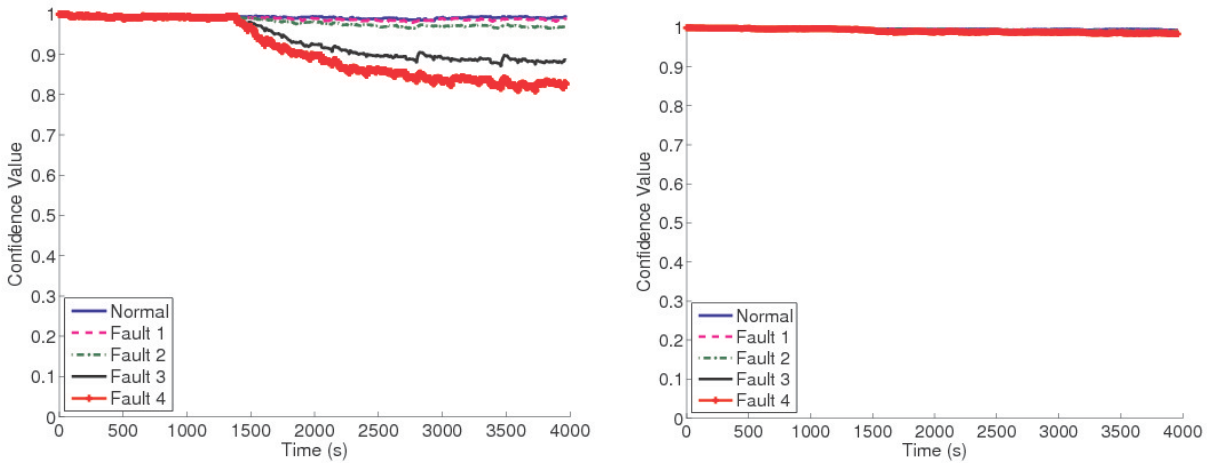


Figure 19: Second level anomaly detection CVs for a controller fault introduced at 1322 seconds. The anomaly is clearly in the mass flow subsystem and thus the third level anomaly detectors will be monitored.

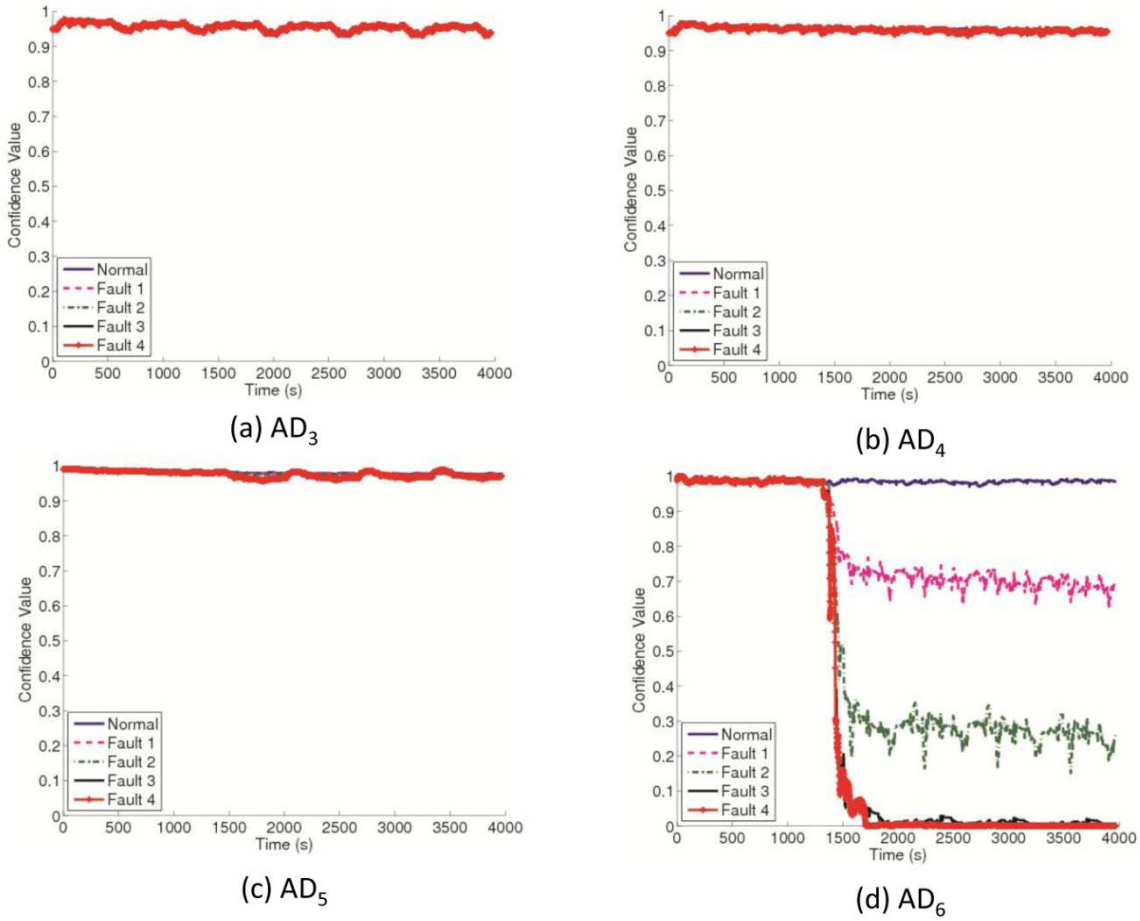


Figure 20: Fault isolation through distributed anomaly detection for a valve anomaly introduced at 1322 seconds. As expected, the EGR Valve AD (AD_6) is the only one indicating an anomaly.

A controller anomaly was simulated by introducing a delay into the control loop as indicated in Figure 21. The increased delay emulates a situation when the control system communication is degraded. Once again, only $AD_{overall}$ is monitored until an anomaly is detected $AD_{mass\ flow}$, AD_{cooler} and AD_3 through AD_6 are utilized to isolate the fault per the methodology in Section 3. Figure 22 shows the CV for $AD_{overall}$ and clearly indicates an anomaly. Figure 23 shows the results for the second level anomaly detectors and Figure 24 shows the results for AD_3 through AD_6 and as expected, only AD_5 indicated an anomaly.

The above anomaly detection and isolation procedure follows the methodology described in Section 2 with one notable modification. Since EGR valve angle saturates ($0^\circ \leq \alpha \leq 90^o$), the PI controller output has a saturation nonlinearity. Thus, the output of the GSMMS was saturated as well so as to maintain a lower number of regions representing the corner introduced by the saturation.

Finally , a set of cooler faults was simulated by modifying the functional relationship between the EGR mass flow and the cooler heat transfer coefficient. This modification simulates a situation where the cooler has experienced fouling, clogging or some other fault that degrades the heat transfer characteristics. These modifications are illustrated in Figure 25. The anomaly detection is demonstrated in Figure 26 where it is clear that the $AD_{overall}$ CV drops for the faults introduced. Isolation is accomplished by the second level anomaly detectors in Figure 27. Note that because the anomaly has been isolated to the cooler, the third level is not needed.

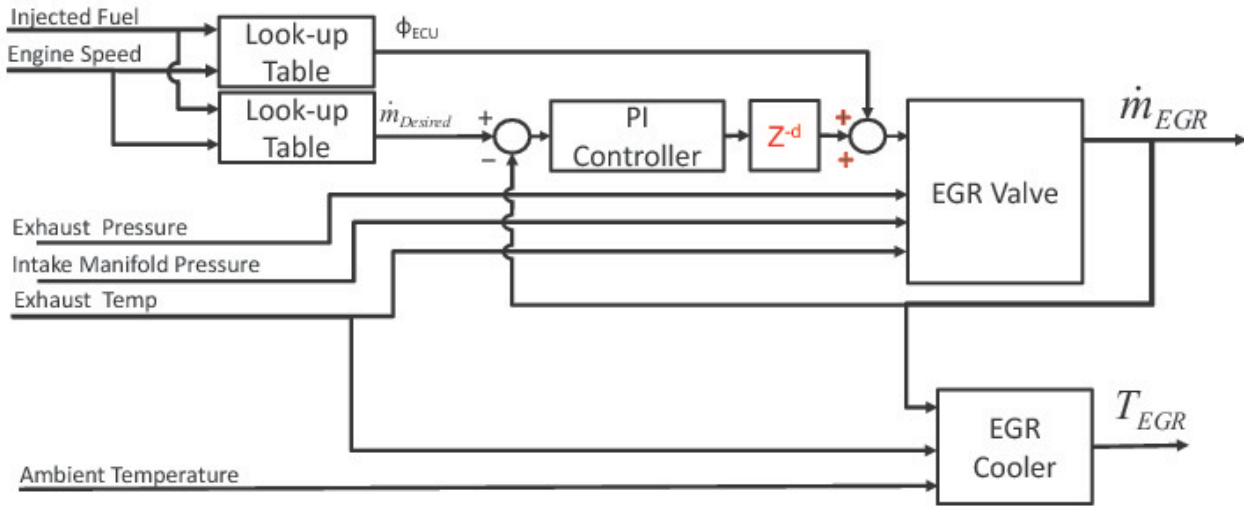


Figure 21: Simulated anomaly in EGR valve controller.

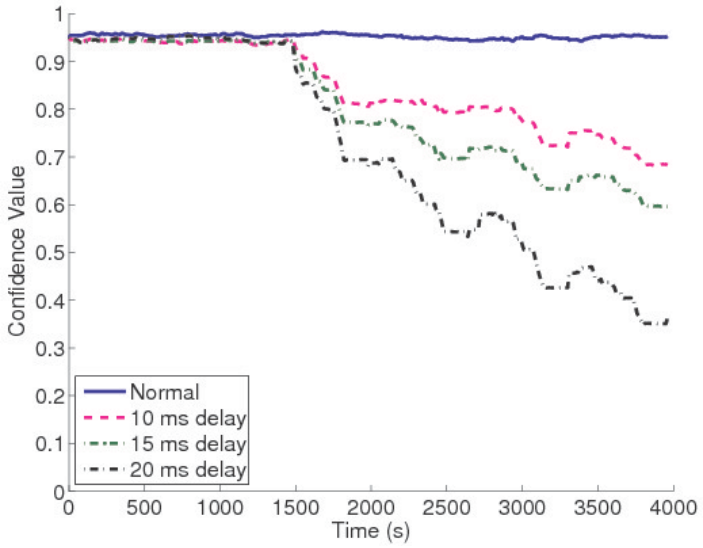


Figure 22: Confidence values for $AD_{overall}$ in the presence of controller faults.

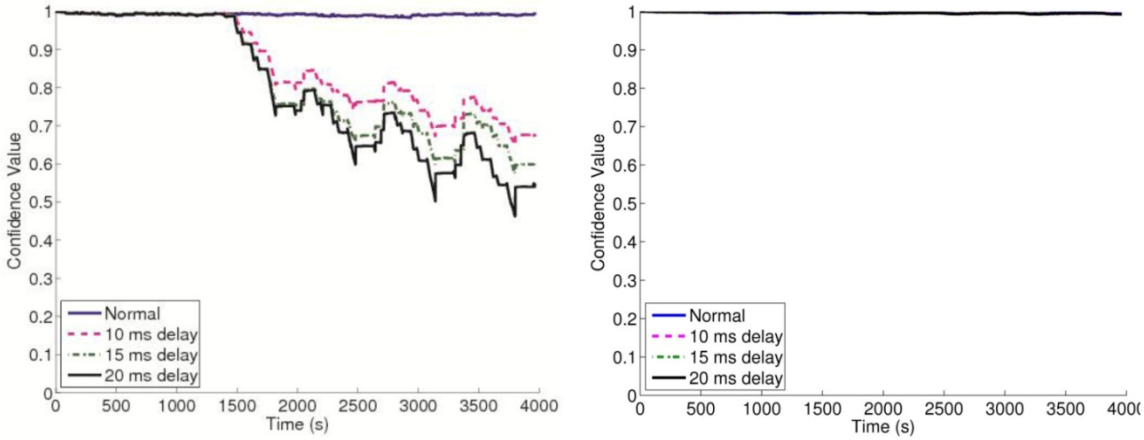


Figure 23: Second level anomaly detection CVs for a controller fault introduced at 1322 seconds. The anomaly is clearly in the mass flow subsystem and thus the third level anomaly detectors will be monitored.

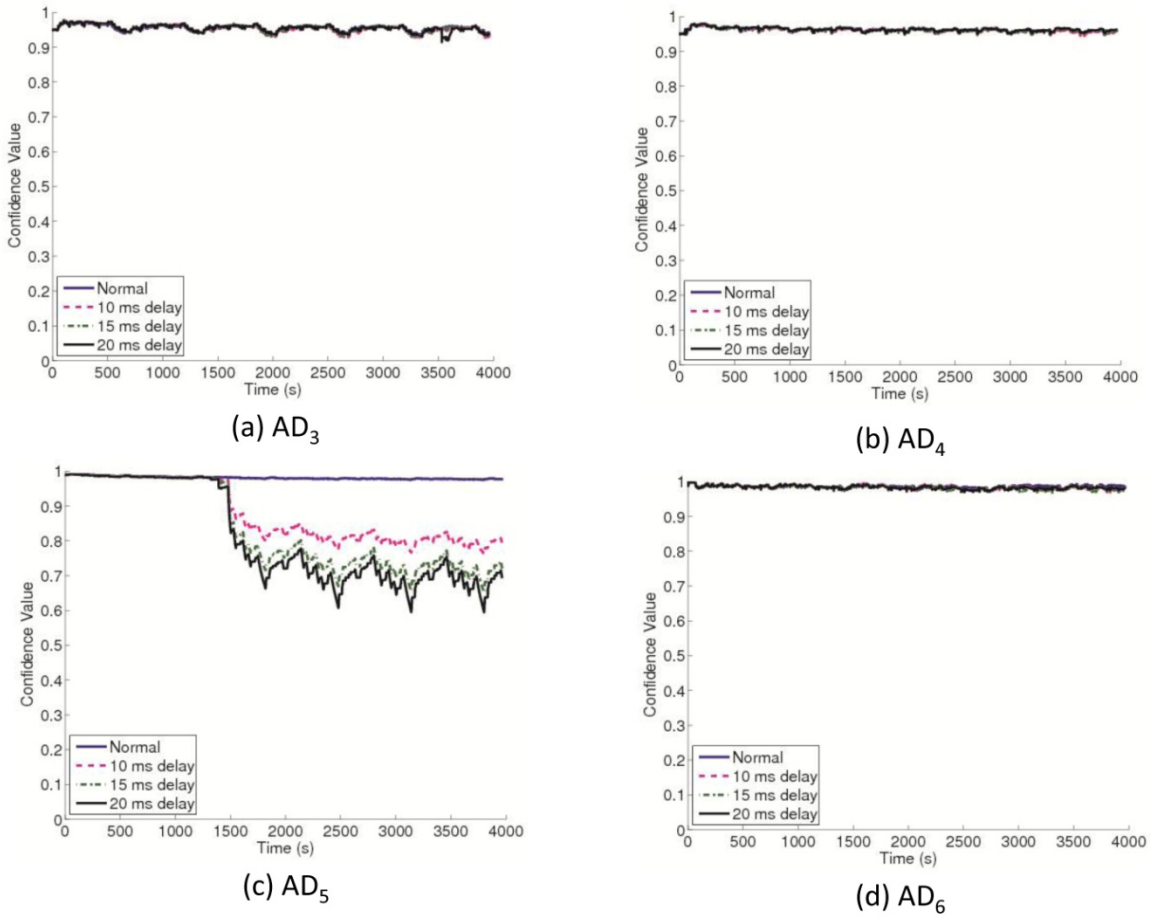


Figure 24: Fault isolation performed by the third level of distributed anomaly detectors with a controller anomaly introduced at 1322 seconds. As expected, the PI controller AD (AD_4) is the only one indicating an anomaly

We also note that global CV for AD_{cooler} appears to be insensitive to cooler Fault 1 (the smallest fault). However, one can take full advantage of the operating space decomposition of the GSMMS and construct regional statistical process control charts to track the number of regions signaling faults. If regional CUSUM charts [25] are constructed for AD_{cooler} we find that zero regions signal faults during the normal period of operation (< 1322 seconds) and at the end of the simulation, three regions are signaling faults for Fault 1. Thus, even the smallest fault is detectable on a regional basis, despite the insensitivity of the global CV.

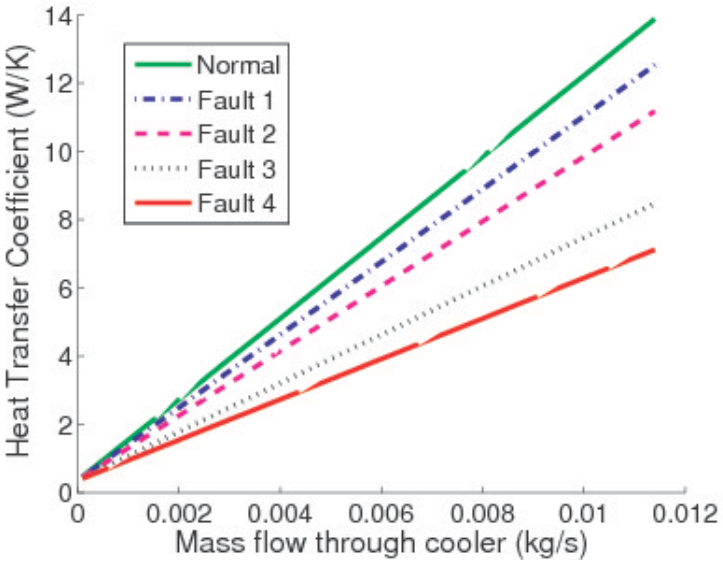


Figure 25: The heat transfer coefficient for the EGR cooler for normal operation and four levels of fouling.

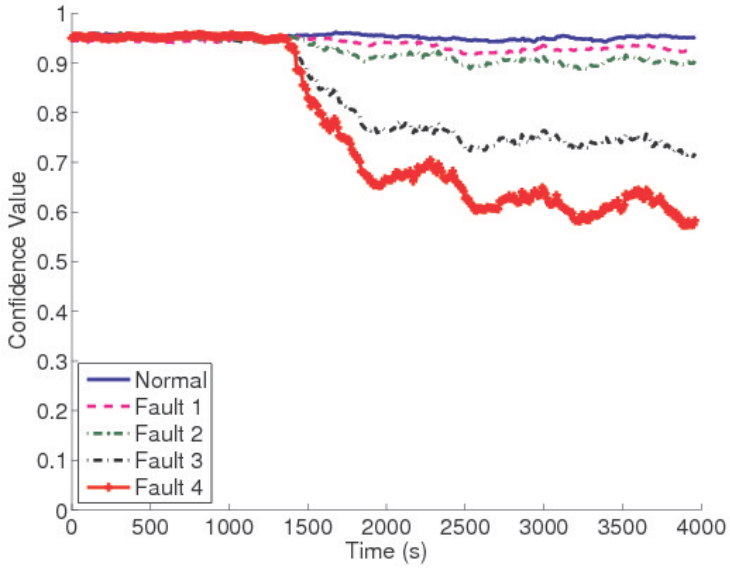


Figure 26: Overall AD for the cooler faults.

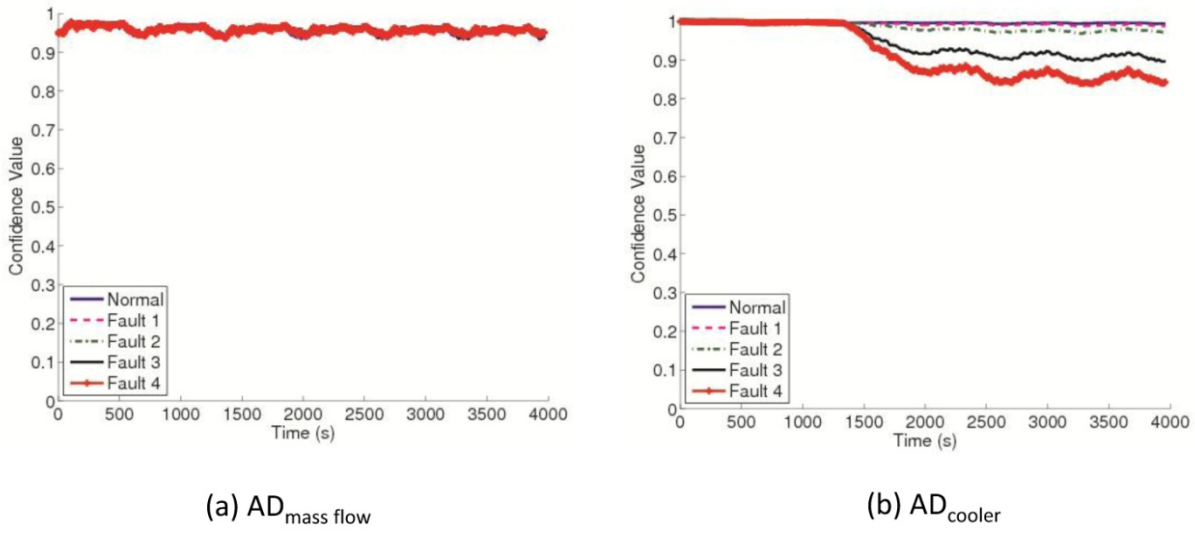


Figure 27: Isolation of the cooler faults conducted by the second level ADs.

B. Fault Diagnosis

The same GSMMS model-based framework used for anomaly detection and isolation is used to diagnose (recognize) faults using the traditional diagnostic methodology, briefly outlined in Section 3. Data emitted by faulty systems was used to construct diagnosers for the valve anomalies in precisely the same manner as the ADs. However, instead of quantifying the similarity to normal behavior, the diagnosers quantify the similarity between the current output patterns and a particular fault.

For valve faults 2 and 4 illustrated in Figure 14, a GSMMS model was constructed utilizing data emitted from the faulty valve operation. Then, as was done for the anomaly detectors, regional residual probability density functions were approximated with local Gaussian Mixture Model and the local CVs were computed. Thus, once an anomaly is detected and isolated to the valve using the aforementioned procedure, these valve diagnoser CVs are then monitored to measures the similarity of the current behavior to behavior see during previous faulty operation. A sample result for diagnosers for the valve faults is shown in Figure 28 where it is clear that Fault 4 is likely present.

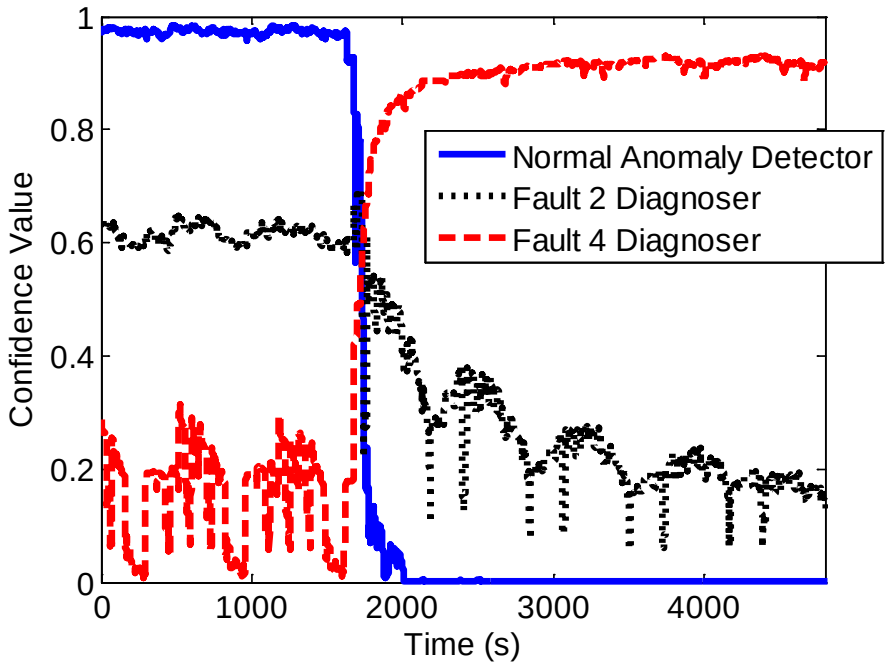


Figure 28: Outputs of the Diagnosers for the throttle fault 4 introduced at 1322 seconds.

6. Conclusions and Future Work

In this paper, a novel approach to anomaly detection is described for anomaly detection, fault localization and fault identification of systems of interacting dynamic systems. Anomalies are detected as statistically significant departures of dynamic modeling errors away from their normal behavioral patterns. The culprit subsystem(s) causing the anomalous behavior are localized through distributed anomaly detection that delves into subsystems of deeper granularity each time an anomaly is detected. Finally, fault identification is accomplished through the traditional diagnostic approaches of matching of identified behavioral models corresponding to known faults with signatures of the currently observed system behavior.

The key enabling method facilitating the above-mentioned vision is the newly developed "divide and conquer" approaches to modeling of complex system dynamics, allowing anomaly detectors and diagnosers to use essentially the same modeling mechanism to successively connect to different inputs and outputs (corresponding to different subsystems), identify behavioral models of the corresponding subsystems and accomplish anomaly detection and diagnosis. Description of system dynamics is cast into the framework of connected multiple models, each of which is relatively simple in nature and can be analyzed in an analytically tractable manner.

Examples of anomaly detection, fault localization and diagnosis in automotive electronic throttle system and diesel engine EGR valve system are presented. In both cases, various faults were inserted into the system, their presence was detected and the subsystems that caused them were correctly identified, even though the diagnostic system was never trained to recognize those faults and thus the corresponding fault models were not available (precedent-free). All results are obtained using high fidelity automotive system simulations developed by a major manufacturer of software for automotive control and diagnostic applications.

Future research will be dedicated to utilizing local model tractability emanating from the "divide and conquer" modeling approaches to devise methods for controller adaptation that will facilitate performance recovery in the presence of faults.

7. Acknowledgment

This work was supported in part by ETAS Inc. and National Science Foundation (NSF) under Grant CMMI 0829237. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

[1] M. Gómez and J. Ventosa, "Expert system hardware for fault detection," *Applied Intelligence*, Sep. 1998.

[2] C. Nott, S.M. Ölçmen, C.L. Karr, and L.C. Trevino, "SR-30 Turbojet engine real-time sensor health monitoring using neural networks, and Bayesian belief networks," *Applied Intelligence*, vol. 26, pp. 251-265, Dec. 2006.

[3] M.P. Feret and J.I. Glasgow, "Combining case-based and model-based reasoning for the diagnosis of complex devices," *Applied Intelligence*, vol. 7, pp. 57–78, 1997.

[4] S. Bahrampour, B. Moshiri, and K. Salahshoor, "Weighted and constrained possibilistic C-means clustering for online fault detection and isolation," *Applied Intelligence*, pp. 1–16, Mar. 2010.

[5] J. Liu, D. Djurdjanovic, K. A. Marko and J. Ni, "A novel method for anomaly detection, localization and fault isolation for dynamic control systems", *Mechanical Systems and Signal Processing*, vol. 23, no. 8, pp. 2488-2499, 2009.

[6] J. Liu, D. Djurdjanovic, K. Marko and J. Ni, "Growing structure multiple model system for anomaly detection and fault diagnosis", *Transactions of ASME, Journal of Dynamic Systems, Measurements and Control*, vol. 131, no. 5, pp. 051001-1 – 051001-13, 2009.

[7] J. Liu, P. Sun, D. Djurdjanovic, K. A. Marko and J. Ni, "Growing structure multiple model system based anomaly detection for crankshaft monitoring," *Proc. of the 2006 International Symposium on Neural Networks (ISNN)*, Chengdu, China, pp. 396-405, 2006.

[8] M. Cholette and D. Djurdjanovic, "Precedent-free fault isolation in a diesel engine EGR system," in review *ASME Journal of Dynamic Systems, Measurement and Control, 2011*.

[9] D. Djurdjanovic, C. Hearn and Y. Liu, "Immune systems inspired approach to anomaly detection, fault localization and diagnosis in a generator", in *Proc. of the 2010 Conf. on Grand Challenges in Modeling and Simulation (GCMS),* Ottawa, ON, paper no. 71, July 11-14, 2010.

[10] L. Cohen, *Time-frequency analysis.* Englewood Cliffs, NJ, Prentice Hall, 1995

[11] R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification.* New York, Wiley, 2001.

[12] J. Lee J. Ni, D. Djurdjanovic, H. Qiu and H. Liao, "Intelligent prognostics tools and e-maintenance", *Computers in Industry*, vol. 57, issue 6, pp. 476-489, 2006.

[13] D. Djurdjanovic, R. Kegg, J. Lee and J. Ni, "Generic multisensor based assessment of performance of manufacturing processes", *Transactions of NAMRI*/SME, vol. 38, pp. 379-386, 2010.

[14] T. Kohonen, *Self Organizing Maps*. Springer Series in Information Sciences. Springer-Verlag, 1995.

[15] J. C. Principe, L. Wang, and M. A. Motter, "Local dynamic modeling with self-organizing maps and applications to nonlinear system identification and control," Proc. IEEE, vol. 86, no. 11, pp. 2240–2258, 1998.

[16] T.A. Johansen and B. A. Foss. "Identification of non-linear system structure and parameters using regime decomposition", *Automatica*, vol. 31, no. 2, pp. 321–326, 1995.

[17] G.A. Barreto and A.F.R. Araujo, "Identification and control of dynamical systems using the self-organizing map", *IEEE Trans. on Neural Networks*, vol. 15, no. 5, pp. 1244–1259, Sept. 2004.

[18] B. Fritzke, "A growing neural gas network learns topologies," in *Advances in Neural Information Processing Systems*. MIT Press,vol. 7, pp. 625–632, 1995.

[19] B. Fritzke, "Growing cell structures - a self-organizing network for unsupervised and supervised learning," *Neural Networks*, vol. 7, no. 9, pp. 1441–1460, 1994.

[20] D. Alahakoon, S.K. Halgamuge, and B. Srinivasan, "Dynamic self-organizing maps with controlled growth for knowledge discovery", *IEEE Trans.* on *Neural Networks*, vol. 11, no. 3, pp. 601–614, May 2000.

[21] J. Liu and D. Djurdjanovic, “Topology preservation and cooperative learning in identification of multiple model systems”, *IEEE Transactions on Neural Networks*, vol. 19, no. 12, pp. 2065–2072, Dec. 2008.

[22] G. McLachlan and D. Peel, *Finite Mixture Models*. John Wiley & Sons, Inc., 2000.

[23] Z. Zivkovic and F. van der Heijden, “Recursive unsupervised learning of finite mixture models”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 651–656, 2004.

[24] ETAS GmbH, “Gasoline Engine Vehicle Model V5.0”, Stuttgart, Germany, 2007.

[25] D. C. Montgomery, *Introduction to Statistical Quality Control*, 4th ed. New York: John Wiley, 2001.

[26] TESIS DYNAware. en-DYNA® THEMOS® 2.0 Block Reference Manual, June 2006.

[27] E. Latronico J. Ni L. Jiang, “A novel method for input selection for the modeling of nonlinear dynamic systems”, *Proc. of the ASME Dynamic Systems and Control Conference*, October 2008.