Signal to Symbol Transformation Techniques for Robust Diagnosis in TRANSCEND

Eric J. Manders
Department of
Electrical and Computer Engineering
Vanderbilt University
Nashville, TN 37235.
manders@vuse.vanderbilt.edu

Pieter J. Mosterman
Institute of Robotics and
System Dynamics
DLR Oberpfaffenhofen
D-82230 Wessling, Germany.
Pieter.J.Mosterman@dlr.de

Gautam Biswas
Knowledge Systems Laboratory
Gates Computer Science Bldg, 2A
Stanford University
Palo Alto, CA 94305.
biswas@ksl.stanford.edu

Abstract

TRANSCEND, our system for monitoring and diagnosis of complex dynamic systems, uses qualitative transient analysis methods to overcome the difficulties that arise in numerical processing, especially for nonlinear systems. Future behavior of hypothesized faults is predicted in the form of signatures, and analyzed by a progressive monitoring scheme. Generating qualitative features from real signals, the *signal to symbol transformation* problem, is a challenging task. This paper discusses model-driven methods for generating symbolic feature descriptions of magnitude and slope changes in noisy continuous data. Experiments have been successfully conducted on faults introduced into the cooling system of an automobile engine.

Introduction

Model-based approaches to diagnosis in continuous systems employ functional redundancy for fault detection and isolation (FDI). Traditional diagnosis methods are essentially *state estimation* (Frank 1987) and *parameter estimation* (Isermann 1984) schemes. In this classification, our approach is basically a qualitative parameter estimation scheme for diagnosis from fault transients. A system model that describes the dynamics of normal system operation and a set of observed variable values are used to estimate parameter values associated with system components. When parameters deviate from their expected values, the components associated with these parameters are considered to be faulty. Numerical parameter estimation methods for diagnosis suffer from convergence, precision, and computational complexity problems (Mosterman & Biswas 1999; Patton, Frank, & Clark 1989). To avoid these pitfalls, we have adopted a qualitative reasoning methodology for fault isolation. This requires a *signal to symbol* transformation step that computes symbolic features from continuously sampled data. The symbolic feature values are the input to the diagnosis algorithms.

We have developed TRANSCEND, a comprehensive model-based approach for diagnosis of complex dynamic systems. The diagnosis model captures dynamic characteristics of the dependency relations between component parameters and the observed variables in the form of a *temporal causal graph* (TCG). The TCG is a rich uniform framework for representing algebraic and temporal constraints among system variables (Mosterman & Biswas 1997; 1999). Component parameter values and their temporal influences on system behavior are defined as attributes of the TCG edges. Our work has focused on the analysis of abrupt faults in continuous dynamic systems that result in transient behaviors. It is critical to analyze these transient behaviors for accurate fault isolation (Mosterman & Biswas 1999).

Figure 1 shows the TRANSCEND architecture. Vector u is the input to the physical process under diagnostic scrutiny, and vector y is the set of observations made on the system. Process behavior is tracked by an observer scheme. The observer consists of a differential equation model to generate the expected system behavior $\hat{y}$, and an observer scheme that tracks the residuals $r = y - \hat{y}$, The observer corrects small discrepancies in the estimated state vector $\hat{x}$ using a standard gain matrix scheme. The residuals, r, are input to the fault detection module that uses sophisticated statistical techniques for discrepancy detection. Discrepancies are reported in symbolic form as *abrupt changes* and *magnitude deviations*.

Discrepancy detection triggers a fault isolation mechanism that consists of hypothesis generation and hypothesis refinement. The hypothesis generation unit uses the model, m, and the symbolic residuals, r_s, to generate a set of hypothesized fault candidates, f_h, from observed deviations, and to predict transients and steady state behavior for each candidate, p. The hypothesis refinement unit employs a monitoring scheme unique to TRANSCEND to prune spurious candidates from the set. Our qualitative characterization of transients facilitates tracking continuous dynamic behavior in metric time [1],

[1] See (Brusoni *et al.* 1998) for characterization of temporal model based diagnosis methodologies

and simplifies the fault isolation task without compromising accuracy. The goal is to continue monitoring until the true fault is isolated.

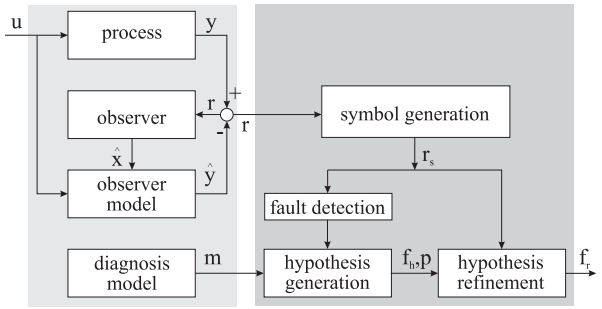


Figure 1: TRANSCEND system architecture.

The emphasis of our previous work has been on developing system models and diagnosis algorithms to achieve accuracy and precision in the fault isolation task (Mosterman & Biswas 1997; 1999; Narasimhan, Mosterman, & Biswas 1998). The focus of this paper is on developing *model-driven signal to symbol transformation* techniques that are essential to making the diagnosis methodology work with real systems. We expect that model generated predictions of system behavior of hypothesized faults can be exploited to select those signal to symbol transformation techniques that give the most timely or robust results in the context of signal properties. To demonstrate the effectiveness of this approach we performed a number of experiments on an automobile engine testbed. This paper discusses the results of some of these experiments.

Diagnosis from Transients

Qualitative diagnosis of abrupt faults relies heavily on the characterization of the transients that occur immediately after the occurrence of the fault (Mosterman & Biswas 1999). *Time constants* play a key role in characterizing transients. Some variable values undergo an instantaneous change due to an abrupt fault. For other variables, energy storage elements (capacitive and inertial effects) act as buffers that introduce propagation delays in signal value propagation. As a result, changes in these variable values take longer to manifest. Computer-based monitoring of physical systems necessarily involves sampling to generate a discrete time representation of the continuous signals. In theory, if the signals are sampled at rates that are faster than the Nyquist rate, the transient characteristics can be fully reproduced. However, for the purpose of analyzing signals to extract relevant features, sampling at rates faster than the Nyquist rate provides an improved signal representation (Young 1988).

We assume that abrupt faults cause changes in parameter values that occur at time scales much faster than the smallest nominal time constant in the system, causing signal value changes on a time scale smaller than the sampling rate. Such signal value changes are labeled as *discontinuities*.

In our qualitative characterization framework, a measurement is considered *normal* if its value is within a predefined threshold of its nominal value predicted by the model. The choice of the threshold value depends on the signal to noise ratio (SNR) and the acceptable false alarm rate. This is discussed in greater detail in the next section. A signal that exceeds the threshold is said to be *deviating*. For dynamic systems, a normal observation at any given time may actually be a slowly changing value that has not yet crossed the threshold, and, therefore, is not labeled as deviating. Therefore, unlike traditional consistency based diagnosis methods (e.g., (deKleer & Williams 1987)), normal measurements cannot necessarily be used to refute faults during transient analysis (Mosterman & Biswas 1999). Only when a fault is predicted to cause an abrupt change in the signal value, can a reported normal measurement be used to refute that fault. This is contingent on the fact that abrupt changes can be detected reliably, an issue that is discussed in detail in the next section. A temporal ordering of the manifestation of first and higher order effects in studying deviations from normal is also generally impossible. In the current implementation, deviating observations are individually analyzed to generate sets of single fault hypotheses.

Individual signal features are the prime discriminating information between competing fault hypotheses. In noisy environments, the accuracy of magnitude change measurements (above (+) and below (−) normal) is governed by the properties of the associated sensors and the change detection techniques employed. Reliable techniques can also be developed for qualitative estimates of slopes or first order derivatives, i.e., increasing (+), steady (0), and decreasing (−), from measured signals. However, it is increasingly difficult to estimate higher order derivatives in the presence of noise (Chantler *et al.* 1996). In our system we limit measurement analysis to recording qualitative magnitude and first derivative values.

Generating Fault Hypotheses and Predicting Future Behavior

Three algorithms make up the core of the fault isolation procedure:

1. fault hypothesis generation by component parameter implication,
2. prediction by generation of transient signatures for each hypothesized fault, and
3. fault isolation or hypothesis refinement by further monitoring.

Hypothesis generation and prediction are the topic of this section. A more detailed discussion can be found in (Mosterman & Biswas 1999). The monitoring methodology is presented in the next section.

Component Parameter Implication For the first recorded discrepancy between measurements and their corresponding nominal value, a backward propagation algorithm is invoked on the temporal causal graph to implicate component parameters. Observed deviating values are propagated backward along the directed edges and consistent '−' and '+' deviation labels are assigned sequentially to vertices along the path based on the edge relation. A component is implicated when an edge corresponding to that system component parameter(s) is traversed. The component parameter deviation is labeled '−' or '+', depending on the value assigned to the last vertex and the edge relation. This procedure generates a set of hypothesized single faults that are consistent with each reported deviating observation.

Signature Generation We constrain the problem space by making the assumption that faults do not cause changes in system configuration, and the system model remains valid even after faults occur in the system. The prediction algorithm employs the system model to compute the qualitative transient behavior of the observed variables for the hypothesized individual fault conditions. Transient behavior is expressed as a tuple of qualitative values for magnitude, 1^{st} order time-derivative and higher order effects. The range of qualitative values are: '+', '−', '0' or '·'. '·' implies that the value is unknown; this can be attributed to opposing qualitative influences on the variable. The tuple is called the *signature* for the variable (Mosterman & Biswas 1997; 1999).

The algorithm propagates the effects of a hypothesized fault to establish a signature for all observations. Energy storage elements cause time integrating effects and introduce temporal edges. In such situations, the cause variable affects the derivative of the effect variable. Propagation of a deviation starts with a 0^{th} order effect, i.e., a magnitude change. When an integrating edge is traversed, the magnitude change becomes a 1^{st} order change, i.e., the first derivative of the affected quantity changes. Similarly, a first order change propagating across an integrating edge produces a second order change, and so on. The highest predicted derivative order required is a design consideration. For some faults, signatures may differ only for higher order derivatives. However, higher order effects also take longer to manifest, thus increasing the time required to isolate the fault uniquely. This issue, and its relation to the measurement selection problem is discussed in detail in (Mosterman 1997; Narasimhan, Mosterman, & Biswas 1998).

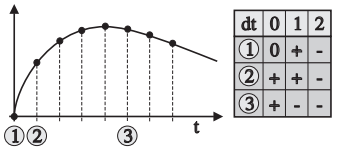


Figure 2: Progressive monitoring.

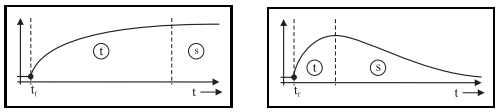
Monitoring

The monitoring mechanism compares actual observations with the predicted behaviors to prune hypothesized faults during transient analysis. It also checks to see if signals have achieved their steady state values so that transient analysis may be replaced by steady state diagnosis.

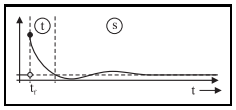
Progressive Monitoring Transient characteristics at the time point of failure change over time as other phenomena in the system affect the measured variables. Therefore, fault signatures evolve dynamically. For example, a fault in the system may have no effect on the initial magnitude of a variable, but it may affect the first derivative. Immediately after the fault occurs the variable value will be observed to be normal, but as time progresses the magnitude will follow the direction of the first derivative. The notion of using higher order derivatives in the analysis of measured values during monitoring is referred to as *progressive monitoring* (Mosterman 1997; Mosterman & Biswas 1999). Figure 2 illustrates the process. When an observed variable does not match a predicted *normal* value, the comparison is extended to predicted higher order derivatives in the signature. If the higher order derivatives match the observed value the hypothesized fault is still retained. A prediction of magnitude deviation that has a '−' or '+' value before progressive monitoring is applied indicates an abrupt or discontinuous change. A discontinuous change in magnitude is different from a magnitude deviation that manifests over time. The fault isolation process is significantly enhanced if discontinuous changes can be reliably detected and differentiated from a continuous magnitude change.

Temporal Behavior Two distinct phases of signals in response to fault disturbances, transient behavior and steady state behavior, carry the distinctive discriminative information for diagnosis. Transient phenomena at the time of failure disappear after a time interval, and so it is important to determine when the transient detection phase terminates. At this point, the system should switch to the steady state detection/verification mode.

Palowitch (Palowitch 1987) reports that signals may exhibit a *compensatory* (Figure 3(a)) or an *inverse response* (Figure 3(b)). A compensatory response exhibits a decreasing slope and gradually moves towards a new steady state value. For an inverse response, after an initial increase or decrease,



(a) compensatory (b) inverse



(c) reverse

Figure 3: Elementary transient signals with different qualitative behavior. A vertical marker indicates the point where transient monitoring is started and suspended.

the signal may reverse direction. An additional phenomenon resulting from abrupt faults can be categorized as a *reverse response* (Figure 3(c)). A reverse response occurs if a discontinuous signal exhibits overshoot, and its qualitative magnitude reverses sign. For example in Figure 3(c) the signal goes from above normal to below normal.

In the qualitative analysis framework, the transition to steady state analysis is detected from an initial magnitude deviation by noting that:

- For a compensatory response the slope eventually becomes 0.
- For an inverse response there is no discontinuous change of magnitude at t_f, the time point of failure. The switch from transient to steady state detection occurs when the magnitude and slope deviations take on opposing signs. Eventually the slope may become 0.
- For a reverse response the signal has a discontinuous initial magnitude deviation. The switch to steady state detection occurs when the magnitude deviation changes sign.

When one of these situations is detected, transient verification for that particular signal is suspended, and steady state detection is initiated. This is indicated by the successive stages of the transient in Figure 3). Steady state is detected when a first order derivative becomes 0 for a sufficient period of time. The sufficient period of time is usually based on design information. Techniques specific to steady state detection are not discussed in this paper

Signal to Symbol Transformation

Reflection on diagnosis methodology discussed in the last section implies a two step process in the generation of symbolic symbolic features from signal transients:

1. *discrepancy detection*, which includes identifying the start point of the transient, classifying the change as abrupt or smooth, and labeling it as a + or −, and
2. *qualitative feature extraction* after discrepancy detection as magnitude and slope values of signals.

We present reliable implementation methods for discrepancy detection and feature extraction in this section. As discussed earlier, it is assumed that the measured signals are oversampled, i.e., the sampling rate is faster than the Nyquist rate. Because the bandwidth of the signals is not known, especially when faults occur, the oversampling factor is chosen empirically.

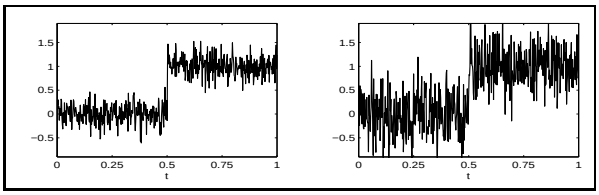
Discrepancy Detection

The *signal to noise ratio* (SNR) of the measurements is a critical factor in the design of the discrepancy detection schemes. When the SNR is high, simple detection and feature extraction methods may suffice. For noisy signals (low SNR), more sophisticated methods that incorporate noise models are required.

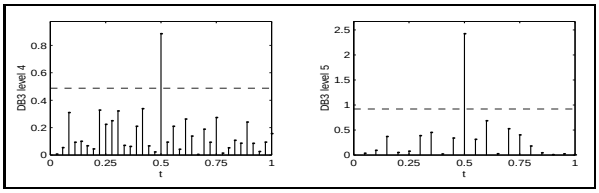
In TRANSCEND, discrepancy detection requires identifying the start of a fault transient, which manifests as one or more measured signal values deviating from their nominal value. Given that the measured signals are noisy, the detection scheme must be designed not to exceed a prespecified false alarm rate. An appropriate threshold value is derived from the false alarm rate, and a signal value is defined to be deviating only if its value crosses the threshold value. To further reduce the number of spurious decisions, the detector may be set up to flag a deviation only if the signal value remains above the threshold for more than a predefined number of measured samples. This is equivalent to a non-linear technique implemented as as a short length rank order filter with the threshold decision scheme. In addition, the direction of the change can be easily noted to label the initial magnitude change as + or −.

However, the threshold crossing scheme is not sufficient because it does not provide information on whether the change was abrupt or not. In a previous version of TRANSCEND, a heuristic analysis scheme was used to label a detected change as abrupt or not abrupt (Mosterman 1997). It was successfully implemented in a number of simulation studies (e.g., (Mosterman & Biswas 1999)), but the technique is unreliable with noisy data. We are investigating several more sophisticated and robust techniques for discontinuity detection that have a sound theoretical basis. Two of these methods are outlined in this section, and their performance characteristics are analyzed.

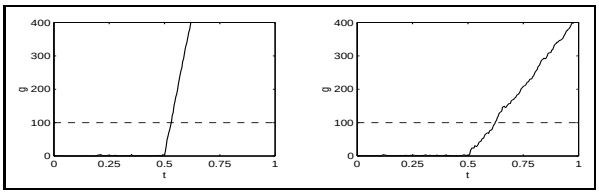
Time frequency analysis An abrupt change in a signal implies a local high frequency component



(a) Unit step function with Gaussian noise with $\sigma = 0.2$ (left) and $\sigma = 0.4$ (right).



(b) Discrete Wavelet Transform (DWT). The threshold (dashed line) is applied at the coefficients at level 4 (for $\sigma = 0.2$) and 5 (for $\sigma = 0.4$) in a nine level decomposition.



(c) Generalized Likelihood Ratio (GLR) innovation sequence (g). The change is detected by a threshold on g (dashed line).

Figure 4: Comparison of two methods for the detection of a step change at t = 0.5.

in the signal. For diagnosis of dynamic systems it is important to accurately determine both the nature of the discrepancy and the time point at which it occurred. Comparing the relative ordering of the measured signals, and their temporal characteristics helps to discriminate among possible fault hypotheses. Classic frequency analysis techniques using Fourier Transform (FT) methods provide useful signal characteristics, but they are not suitable for local signal analysis because the complex exponentials used as basis functions for the signal representation in the frequency domain have infinite duration. On the other hand, transient analysis requires the extraction of frequency characteristics of signals over short time durations. A set of transform techniques known as *time-frequency* methods address this problem by employing basis functions that have compact support, i.e., they can be constrained to finite time intervals. Of particular interest to us is the *wavelet transform* that has been proposed as a method for change detection (Mallat & Hwang 1992; Wang 1995). Change detection is formulated as a hypothesis testing problem with the two alternate hypotheses:

$$\begin{cases} H_0 : \text{the signal is differentiable} \\ \qquad \text{(i.e., it represents a smooth function),} \\ H_1 : \text{the signal has discontinuities} \\ \qquad \text{and/or sharp cusps.} \end{cases}$$

A formulation of this description based on signal regularity can be found in (Wang 1995).

Let Ψ be an appropriate wavelet basis function[2] and $\Psi_{s,u}(t) = s^{-1/2}\Psi((t-u)/s)$, represent scaled and translated versions of Ψ. Note that this function includes the scale or frequency parameter, s, and the position of the signal on the time line, u. (s,u) defines the scale space or the time-frequency plane for the analysis. The wavelet transform of a signal f at s and u is computed by taking the inner product of f with $\Psi_{s,u}$, i.e., $Wf(s,u) = \langle f, \Psi_{s,u} \rangle$. A sharp cusp or a discontinuity at u implies that $Wf(s,u)$ will be proportionally much larger than $Wf(s,u)$ values when f is continuous. Based on this observation, a test statistic, the maximum of the absolute value of coefficients of $Wf(u,s)$, can be formulated. If the test statistic exceeds a predefined threshold, hypothesis H_1 is true, otherwise H_0 is true. The value of u for the test statistic represents the starting time point of the transient.

In the discrete time case the Discrete Wavelet Transform is applied and the threshold selected is proportional to the *median absolute deviation* of the wavelet coefficients at the finest level (Wang 1995). The finest level coefficients model the high frequency components in the signal, and this includes the noise. The threshold reflects a balance between the highly variable components of the signal and the smoother components of the signal. As the SNR of the signals decreases, the decomposition level for applying the threshold corresponds to a larger time scale. The precision with which the start of the transient can be detected depends on the level of the decomposition where the threshold is applied. The ability to localize the change point is thus diminished when noise levels increase. There is no theoretical way to decide that level but in a practical solution a maximization is applied over all levels. The detection delay is determined by the window size of transform. A smaller window implies a smaller detection delay, therefore, more compact wavelets are desirable. Figure 4(b) illustrates an experiment for detecting a step change in noisy signals. The detector for the lower SNR signal requires a higher threshold and a larger time scale.

[2]In the wavelet literature, this is often called the *mother* wavelet.

An advantage of the wavelet transform based detector is that it does not require an underlying signal model. However, it is most suitable for detecting transients with sharp peaks or discontinuities. A challenge in online monitoring and diagnosis applications is to come up with systematic methods for dynamically selecting a proper wavelet basis and the window size. In our experiments, we have followed the results cited in (Wang 1995) and employed the Daubechies-3 wavelet. The problem of selecting the best wavelet basis function is still a topic of study.

Statistical Signal Processing This approach is based on representing the signal as a random process with a known probability distribution. An abrupt change in a signal is modeled as a change in a parameter value of the probability distribution. The signal measurements are assumed to be an independent random variable sequence y_k with probability density function $p_\theta(y_k)$, where θ is the signal model parameter that is being monitored to detect abrupt changes in its value. The abrupt change detection problem can again be formulated as a hypothesis testing problem (Basseville & Nikiforov 1993):

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1, \end{cases}$$

where θ_0 represents the initial parameter value, and θ_1 the changed parameter value. The central quantity in constructing the test statistic is the *log-likelihood ratio*, $s(y) = \ln \frac{p_{\theta_1}(y)}{p_{\theta_0}(y)}$. It is well known that the cumulative log likelihood ratio, $S_j^k = \sum_{i=j}^{k} s_i$, where $s_i = \ln \frac{p_{\theta_1}(y_i)}{p_{\theta_0}(y_i)}$ and j, k define a window in the time sample space, shows a negative drift before a change occurs in θ_0, and a positive drift after the change. This property is exploited in the CUSUM algorithm with S_j^k as the decision function. Letting $m_k = \min_{j \le i \le k} S_i$, the detection time is determined using the stopping rule: $t_a = \min\{k : S_j^k \ge m_k + h\}$, where $m_k + h$ (h is predefined) is an adaptive threshold that is modified online. In diagnosis problem solving, the changed parameter value θ_1 is unknown (but the parameter θ is assumed to be known), the parameter θ_1 can be replaced by its maximum likelihood estimate, which results in the Generalized Likelihood Ratio (GLR) algorithm. The GLR involves a double maximization function (see (Basseville & Nikiforov 1993) for details):

$$g_k = \max_{1 \le j \le k} \sup_{\theta_1} S_j^k.$$

The stopping rule is given by $t_a = \min\{k : g_k \ge h\}$, where h is a predefined threshold and t_a is again the detection time.

Figure 4(c) shows the output of the GLR algorithm, where the parameter θ is the mean μ of a Gaussian process. The GLR for this particular case is computed as (Basseville & Nikiforov 1993):

$$g_k = \frac{1}{2\sigma^2} \max_{1 \le j \le k} \frac{1}{k - j + 1} \left[\sum_{i=j}^{k} (y_i - \mu_0) \right]^2,$$

where σ is the standard deviation of the Gaussian process. In our diagnosis framework, this is equivalent to looking for an unknown abrupt change in the value of a constant signal (initial value μ_0) with superimposed white noise (mean = 0, and standard deviation = σ).

Unlike the wavelet based detectors, this method has the advantage that the decision threshold, h, is directly related to the false alarm rate of the signal, and a clear tradeoff exists between detection delay and sensitivity of the detector. Also, the GLR algorithm shows good performance for signals with low SNR, although the detection delay increases with decreasing signal to noise ratio. This is illustrated in Figure 4(c).

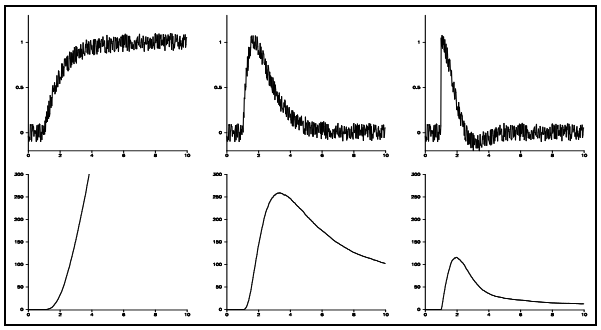


Figure 5: Generalized Likelihood Ratio (GLR) applied to compensatory, inverse, and reverse behaviors.

However, unlike the wavelet detector, the GLR algorithm explicitly incorporates the statistical signal model in defining the test statistic. Therefore, it becomes difficult to interpret the results of this detector when it is applied to change analysis of arbitrary signals with unknown models. This is illustrated in Figure 5, where the GLR detector is applied to the compensatory, inverse, and reverse transients presented in Figure 3 in the previous section. Compensatory response (Figure 3(a)) is close to the signal change detection model used in this section, therefore, the GLR algorithm correctly detects the change in mean, but information about the dynamics of the transient cannot be derived from the value of the innovation function. The signal models for the reverse and inverse response do not correspond to the model employed in the GLR detector, and no meaningful information about the signal dynamics (expect, possibly for the time of the initial change) can be drawn from the innovation function.

Evaluation Our study of the two change detection techniques reveals the advantages and disadvantages for both. A wavelet detector, in some sense, is more general and flexible, because it does not require predefined models for interpreting signal characteristics, but we have found it difficult to come up with systematic procedures (as opposed to trial and error methods) for linking the threshold function to design characteristics, such as the SNR. The GLR detector requires predefined models for interpretation of its innovation function, but it has the distinct advantage that its adaptive threshold function can be directly linked to the SNR. In our experimental studies with the engine testbed, discussed in the next section, we chose the GLR based detector because its signal model fit best with the transients we have encountered thus far for the faults associated with the engine coolant system. As we extend our experimental studies to other faults and other subsystems, we plan to use model based prediction techniques to hypothesize possible fault dynamics, and then attempt to select detectors that best fit predicted dynamic model.

Qualitative Feature Extraction

The analysis of the transient after the initial change requires deriving the qualitative magnitude and slope values for subsequent measured values for the signal. Deriving the qualitative value of the signal magnitude (+ or −) can be accomplished using a simple filter that avoids spurious fluctuations around the zero crossing.

To design a suitable slope estimator we first consider an ideal discrete-time differentiator with frequency response: $H(e^{j\omega}) = j\omega, \quad |\omega| \leq \pi$. The associated filter that exhibits this response is noncausal and has infinite length. Therefore, a simple differentiator can only approximate ideal behavior. The simplest approximation, the *first order difference operator*, $y'(n) = y(n) - y(n-1)$, has frequency response $H(e^{j\omega}) = 1 - cos\,\omega + j\,sin\,\omega$. This approximates the ideal response for low frequencies, but when ω approaches π, the response deviates significantly from the ideal filter (e.g., (Chantler *et al.* 1996)). We can improve performance through oversampling of the signal, but when the signal is noisy, the performance is generally unacceptable. An intuitive approach is to compute a *piecewise linear approximation* of the signal data using least squares regression on successive data segments to approximate the slope. A better approach that derives a *Finite Impulse Response (FIR) filter* that provides an unbiased minimum variance derivative estimate. This method assumes that the signal is a piecewise polynomial with the same degree as the order of the desired derivative (Vainio, Renfors, & Saramaki 1997).

Figure 6 compares the three methods without and with noise. The top figure shows the actual signal.

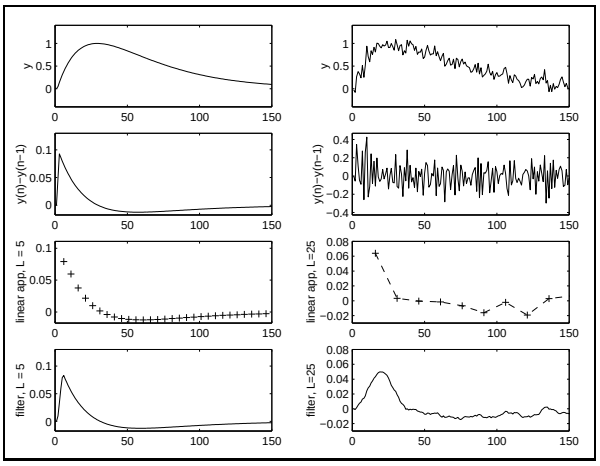


Figure 6: Comparison of slope estimation methods for a signal without noise (left) and with noise (right).

This is followed by the first order difference operator. This operator performs well for the noise free signal but generates meaningless results for the noisy signal. The third and fourth set of results illustrates the performance of the linear approximation method and and the FIR filter, respectively. Note that the length of the segment used for approximation and the filter length are larger for the noisy signal. The performance of both the piecewise linear approximation filter and the FIR filter are comparable for this example. However, the former method implicitly subsamples the derivative signal at the segment boundaries.

The Engine Cooling System

We have performed fault isolation studies with real signals by constructing a testbed using a V-8 internal combustion engine of a Chevrolet automobile. The system is engineered to include a set of pressure, temperature, and flow sensors in the cooling system (Manders *et al.* 1999).

Model of the Cooling System

Figure 7 shows a schematic of the engine block and cooling system that includes all the modeled components. Figure 8 presents the TCG of the model.

An automotive cooling system uses a liquid coolant pumped through passageways in the engine block by a centrifugal pump that is driven by the crank axis. The inertia of the rotor in the pump is represented by mass m_1. The hot coolant flows through the radiator (heat exchanger) where it is cooled by air blown through the radiator by a fan. The coolant is then pumped back into the lower part of the engine block where it moves towards the back through passageways with fluid resistance R_{hy-blk}. The fluid capacitance is represented as a capacitor,

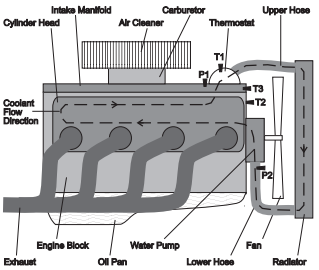


Figure 7: Engine schematic.

C_{hy-blk}. The coolant then flows through the passageways in the cylinder heads up into the intake manifold and towards the front of the engine. If the engine is operating at its desired temperature, the thermostat, modeled as a combination of a fluid inertia, I_{tstat}, and fluid resistance, R_{tstat}, is open and the coolant circulates through the upper hose, represented by fluid resistance, R_{u-hose}, back to the radiator modeled as capacitance, C_{hy-rad}. It is assumed that all hoses and similar fluid paths provide a linear resistance to flow. Resistance R_{l-hose} represents the fluid energy losses that occur during coolant flow through the lower hose. The small outlet of the radiator introduces a fluid inertia, $I_{rad-out}$. A T-split coupling in the lower hose is used to introduce a leak. Even when it is closed, it allows a small flow of coolant, modeled by resistance R_{leak}.

Conductive heat transfer occurs at two primary locations. First, heat is transferred from the combustion chamber of the engine to the liquid coolant. The heat capacitance of the coolant in the cylinder head and block, modeled by C_{th-blk}, is a function of the specific heat of the coolant, the thermal capacitance per unit mass. The overall capacity to absorb heat is a function of the specific heat and the mass flow rate of the coolant. The heat capacitance of the coolant in the radiator is modeled by C_{th-rad}. Convective heat transfer from the engine block to the radiator is attributed to the mass transfer of coolant from the engine block to the radiator, and is modeled by a modulated flow source, MSf_1, which is a function of the fluid flow rate in the upper hose. Analogously, convective heat transfer from the radiator to the engine block through the lower hose is modeled by MSf_2. The second location where conductive heat transfer takes place is in the radiator, where the mass of coolant liquid is cooled by the air blown through the radiator by the fan. The heat transfer occurs through the resistive junction R_{th-rad}, which is a function of thermal conductivity and radiator geometry. The outside air is represented as a sink of entropy at constant temperature. The non-linearity introduced by the centrifugal water pump model is further analyzed by incorporating knowledge of nominal values during normal operation (Mosterman & Biswas 1999).

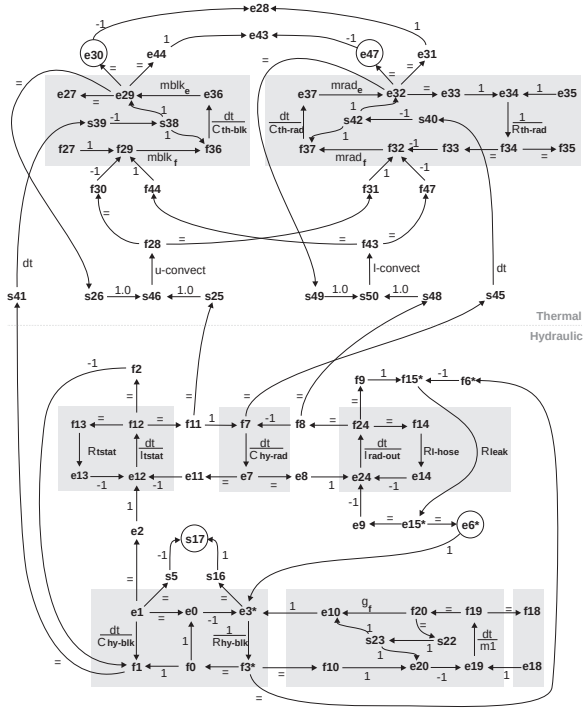


Figure 8: Temporal causal graph for the engine cooling system. Circled vertices are measured variables.

Experimental Setup

Several faults, that include thermostat, belt, hose, and radiator failures, can be introduced into the cooling system without damaging the engine, provided the temperature of the engine block does not exceed certain limits. Since we are dealing with a combined thermal and fluid flow problem, it is important to collect temperature and pressure values at various points in the cooling system loop. We have installed four sensors, two temperature thermocouples T1 and T2, and two pressure transducers, P1 and P2, at expert selected measurement points based on discriminating ability and ease of installation (see Figure 7). Three measured variables {T1,P1,P2} are used in this experiment, corresponding to vertices {e30,s17,e6}, respectively, in the TCG.

Experimental Results: Punctured Hose

Our testbed is set up to run the engine system to a steady state operation (or close to that), then to introduce a desired fault, and to collect fault data for diagnostic analysis. The punctured hose is simulated using a T-split coupling in the lower radiator to drain coolant from the system. The coupling has a large inner diameter to enable a large outflow. A

time	T1	P1	P2
4	(0,.)	(0,.)	(0,.)
5	(0,.)	(0,.)	(0,.)
6	(0,.)	(-,.,*)	(-,.,*)
7	(+,+)	(-,-)	(-,+)
8	(+,+)	(-,-)	(-,0)

Table 1: Output of the Signal to symbol transformation.

large hose puncture is simulated by a lever operated gate valve attached to the coupling that can be switched from closed to open very quickly to make the fault look abrupt. The TCG parameter R_{leak} corresponds to a leak in the lower radiator hose.

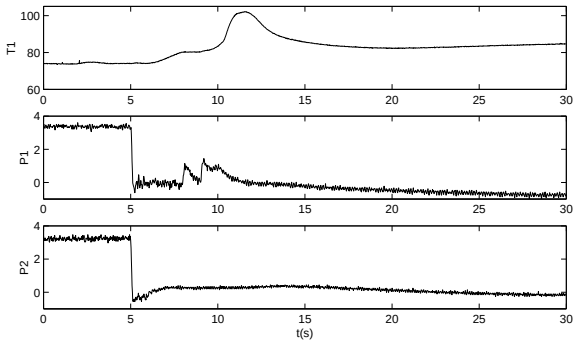


Figure 9: Abrupt loss of coolant through punctured lower radiator hose at t = 5 sec.

Figure 9 shows the T1, P1, and P2 variable measurements as a result of the punctured lower hose fault that was introduced at t = 5 sec. The valve was kept open for several seconds till a large volume of coolant was drained from the system. The sampling time chosen for the experiment was 0.02 sec. The signals were first passed through a median filter of length 5 to remove the few outliers in the temperature data. Symbol values for the measurement were computed every second, using 50 samples for the least squares linear approximation of the signal to determine the slope. A step in the monitoring/diagnosis algorithm was executed for every new set of symbolic measurements that are input by the monitoring scheme.

Table 1 shows the results of the signal to symbol transformation from step 4 to step 8 (the fault occurs at step 5). The slope values are marked as *unknown* ('.') until an initial deviation has occurred. The tuple is extended with a '$\star$' to indicate an abrupt change was detected during a step.

The engine model is nonlinear and contains algebraic loops in the causal graph. This resulted in unknown signatures for second and higher order derivatives for all of the predicted faults. Therefore, only magnitude and slope values were predicted for this model.

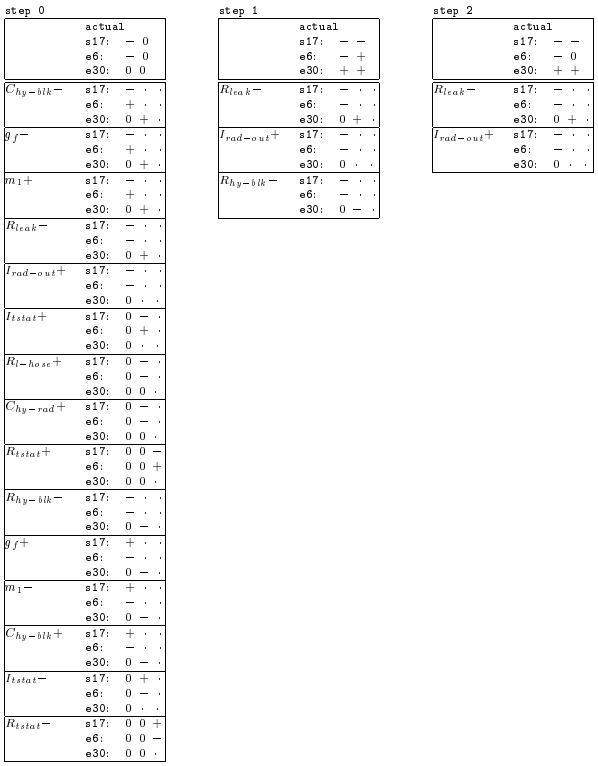


Figure 10: Monitoring and diagnosis results for $R_{leak}-$.

Figure 10 shows the results generated by the hypothesis generation and refinement procedures. The steps are indexed starting at 0. At step 0 (actually time step 6 in Table 1) the first deviations were detected (abrupt drop in both P1 and P2). The resulting candidates generated by the hypothesis generation algorithm and their corresponding fault signatures generated by the prediction algorithm are listed in the first column of Figure 10. For some components, because of the non-linearity of the water pump model, both an increase and decrease in fault magnitudes were hypothesized. This caused an unknown value for a vertex during back propagation for candidate hypothesis generation, and resulted in branching behavior during candidate generation.

Recognition of abrupt changes resulted in the elimination of a large number of hypothesized faults at step 1. Without abrupt change detection, $R_{lhose}+$ and $C_{hy-rad}+$ could not be eliminated at step 1 nor at any later time step and would remain in the set of hypothesized faults. At step 2 the $R_{hy-blk}-$ fault was eliminated by applying the progressive monitoring algorithm. The predicted *slope* was negative for measurement e30, which did not match the observed magnitude deviation.

The final result at step 2 in Figure 10 shows that the diagnosis was accurate because it included the actual fault $R_{leak}-$. $I_{rad-out}+$ was the only spu-

rious candidate that remained, because there were no discriminating measurements to separate it from $R_{leak}-$. In other work (Narasimhan, Mosterman, & Biswas 1998), we have shown how a systematic measurement selection algorithm can be employed to generate the minimal set of required measurements for complete diagnosability, i.e., all faults can be uniquely discriminated. This analysis has not yet been performed on this system.

Discussion and Conclusions

This paper has demonstrated the coupling of signal to symbol transformation techniques with a qualitative approach to model-based fault detection and isolation. The signal to symbol transformation scheme incorporates robust techniques for abrupt change detection in signals, and extracting qualitative magnitude and slope values from noisy signals. The effectiveness of the overall system has been demonstrated by experimental studies on the cooling system of our engine testbed.

The use of symbolic signal representation techniques has been explored by several other researchers. A comprehensive approach to the representation of process trends of complex physiochemical systems has been developed by (Cheung & Stephanopoulos 1990; Bakshi & Stephanopoulos 1994). This work develops a formal representation for process trends based on geometric distinguished time points, outlines the derivation of trends in noisy signals as a multiscale multiresolution problem, and then presents robust and efficient methodologies for extraction of significant temporal features from real data using time-frequency methods. They have applied these methods to pattern-based supervisory control and fault diagnosis in chemical plants. Our overall monitoring and diagnosis approach closely mirrors this approach with some significant differences. Our signal analysis and model based diagnosis approaches are more closely coupled, in that the signal analysis techniques are model driven, allowing us to dynamically select transformation methods during online processing, and set the parameters for these methods (e.g., the thresholds for statistical testing) based on model predictions given the hypothesized fault situations. Further, we believe our fault signature representation and the use of higher order derivatives in progressive monitoring may be more general than the process trend representation methods employed in their work.

The current version of TRANSCEND limits the signal analysis of transients to the extraction of magnitude changes and slope of the measurements. However, we can employ derivative estimation techniques that explicitly incorporate noise models, thus allowing us to compute higher order derivatives for some of the measurements in a reliable manner. Our system models are analytic, therefore, the prediction step can be enhanced to provide characterizations of more complex behavior trends in response to faults, such as oscillations. In response to the more complex fault signatures, more sophisticated signal analysis and interpretation techniques will need to be developed. Further, the analysis techniques could also be extended to derive more quantitative phenomena, such as time constants associated with trends. Mapping back such information into the diagnosis framework will enable us to derive quantitative estimates of degraded and faulty parameters. The problems of traditional quantitative parameter estimation schemes will be significantly reduced, because the discrepancy detection schemes should pinpoint the time point of the fault occurrence fairly accurately, and the estimation process is simplified to the analysis of local trends as opposed to dealing with the complete system model.

From the experiments on the abrupt change detection we see that the detection of a transient should be matched to the model of the transient. To investigate this further our work in transient detection will include methods where a likelihood ratio is computed based on the transform coefficients. The statistical hypothesis test for transient analysis now reflects whether the data is close to a signal model represented by the transform coefficients. This may be especially promising when the basis function of the transform are exponentially damped sinusoidal systems, which frequently occur in the fluid-thermal systems that we work with (Friedlander & Porat 1989).

Acknowledgments

Our research is supported by funds from Hewlett-Packard Laboratories, Palo Alto, California, USA. G. Biswas has been supported by NIST grant 70NANB6H0075-05 while on sabbatical at Stanford University. We gratefully acknowledge the help of Dr. Lee Barford in developing ideas that contributed to this research. We also thank the anonymous reviewers for their comments.

References

Bakshi, B. R., and Stephanopoulos, G. 1994. Representation of process trends. III. multiscale extraction of trends from process data. *Computers & Chemical Engineering* 18(4):267–302.

Basseville, M., and Nikiforov, I. 1993. *Detection of abrupt changes: theory and applications.* Prentice-Hall.

Brusoni, V.; Console, L.; Terenziani, P.; and Dupre, D. T. 1998. A spectrum of definitions for temporal model-based diagnosis. *Artificial Intelligence* 102(1):39–79.

Chantler, M. J.; Daus, S.; Vikatos, T.; and Coghill, G. M. 1996. The use of quantitative dynamic mod-

els and dependency recording for diagnosis. In *Seventh International Workshop on Principles of Diagnosis*, 59–68.

Cheung, J. T. Y., and Stephanopoulos, G. 1990. Representation of process trends. I. a formal representation framework. *Computers & Chemical Engineering* 14(4-5):495–510.

deKleer, J., and Williams, B. C. 1987. Diagnosing multiple faults. *Artificial Intelligence* 32(1):97–130.

Frank, P. M. 1987. Fault diagnosis in dynamic systems via state estimation: A survey. In Tzafestas; Singh; and Smith., eds., *System Fault Diagnostics, Reliability,and Related Knowledge-Based Approaches*, volume 1. Reidel Press, UK.

Friedlander, B., and Porat, B. 1989. Detection of transient signals by the Gabor representation. *IEEE transactions on acoustics speech and Signal processing* 37(2):169–180.

Isermann, R. 1984. Process fault detection based on modeling and estimation methods: A survey. *Automatica* 20:387–404.

Mallat, S., and Hwang, W. L. 1992. Singularity detection and processing with wavelets. *IEEE Trans. on Information Theory* 38:617–643.

Manders, E. J.; Biswas, G.; Mosterman, P. J.; Barford, L.; Ram, V.; and Barnett, J. 1999. Signal interpretation for monitoring and diagnosis, a cooling system testbed. In *Proceedings of the IMTC99*. To appear.

Mosterman, P. J., and Biswas, G. 1997. Monitoring, prediction, and fault isolation in dynamic physical systems. In *AAAI-97*, 100–105.

Mosterman, P. J., and Biswas, G. 1999. Diagnosis of continuous valued systems in transient operating regions. *IEEE Trans. on Systems, Man and Cybernetics*. To appear.

Mosterman, P. J. 1997. *Hybrid Dynamic Systems: A hybrid bond graph modeling paradigm and its application in diagnosis*. PhD dissertation, Vanderbilt University.

Narasimhan, S.; Mosterman, P. J.; and Biswas, G. 1998. A systematic analysis of measurement selection algorithms for fault isolation in dynamic systems. In *Ninth International Workshop on Principles of Diagnosis*, 94–101.

Palowitch, B. 1987. *Fault Diagnosis of Process Plants using Causal Models*. PhD dissertation, Massachusetts Institute of Technology.

Patton, R.; Frank, P.; and Clark, R., eds. 1989. *Fault Diagnosis in Dynamic Systems: Theory and Applications*. Prentice-Hall, UK.

Vainio, O.; Renfors, M.; and Saramaki, T. 1997. Recursive implementation of FIR differentiators with optimum noise attenuation. *IEEE Trans. on Instrumentation and Measurement* 46(5):1202–1207.

Wang, Y. 1995. Jump and sharp cusp detection via wavelets. *Biometrika* 82(2):385–397.

Young, I. T. 1988. Sampling density and quantitative microscopy. *Analytical and Quantitative Cytology and Histology* 10(4):269–275.