INTEGRATING PHYSICAL MODELING, NEURAL COMPUTING, AND STATISTICAL ANALYSIS FOR ONLINE PROCESS MONITORING

by

H. E. Garcia and R. B. Villim*

Nuclear Technology Division
Argonne National Laboratory-West
P. O. Box 2528
Idaho Falls, ID 83403-2528

*Reactor Analysis Division
Argonne National Laboratory-East
9700 S. Cass Ave.
Argonne, IL 60439

The submitted manuscript has been created by the University of Chicago as Operator of Argonne National Laboratory ("Argonne") under contract No. W-31-109-ENG-38 with the U.S. Department of Energy. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.

To be Presented
at the Third ANS International Topical Meeting
on Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies, (NPIC & HMIT) 2000

Washington, D.C.

November 13-17, 2000

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.
DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.
Integrating Physical Modeling, Neural Computing, and Statistical Analysis for Online Process Monitoring

H.E. Garcia and R.B. Vilim
Argonne National Laboratory
P.O. Box 2528
Idaho Falls, ID 83403-2528

Keywords: failure detection, process monitoring, neural computing

ABSTRACT

Two basic approaches can be mentioned to model physical systems. One approach derives a model structure from the known physical laws. However, obtaining a model with the required fidelity may be difficult if the system is not well understood. A second approach is to employ a black-box structure to learn the implicit input-output relationships from measurements in which no particular attention is paid to modeling the underlying processes. This paper describes a method that draws on the respective strengths of each of these two approaches. The technique integrates known first principles knowledge derived from physical modeling with measured input-output mappings derived from neural processing to produce a computer model of a dynamical process. The technique is used to diagnose operational changes of mechanical equipment by statistically comparing, using a likelihood test, the predicted model output for the given measured input with the actual process output. Experimental results with a peristaltic pump are presented here.

1. INTRODUCTION

A complete process model consists of a set of equations that describe the process dynamics and may include correlated experimental data. However, obtaining a good model is often complicated by nonlinearities, system parameter variations, or cost restrictions that does not allow for timely model development. This paper describes a method for obtaining a model when a complete set of equations is not available. The approach assumes that the contribution of the unmodeled behavior to the total process behavior is input-output observable. Neural computing is used to take as inputs the system forcing functions and state vector and return a value that when added to the known model equations reproduces the measured system behavior. This computational intelligence technique has been applied for system identification as seen from numerous publications (e.g., [3]). The approach presented uses likelihood-based methods to deal with the stochastic elements of the model identification and validation problem. In particular, it augments the analytic model with a neural network so that the state equation is the sum of a first principles model and a neural mapping that is to be determined. While the analytic model captures the physical insight known on the system properties, a neural network captures the unmodeled process behavior.
After learning the difference between the true plant response and the analytic model, stochastic parameters representing remaining model uncertainty are estimated by the maximum likelihood method. This stochastic model can be used to detect when the observed process is no longer consistent with the model (incipient failure detection); to estimate the value of an unmeasured process state (process monitoring); and to predict operating limits for proposed actions (process control and prediction).

An online system for equipment state-estimation and operation-detection using the proposed method was implemented, tested and validated in a laboratory-scale experiment. It was used to detect a number of failures introduced into an electric motor-pump system. First principles information consisted of an approximate model for the physical system. The system automatically learned as a mapping the difference between the measured process variables for the electric motor-pump system and the approximate model and then generated the combined model. Experimental runs were conducted to show that the detection sensitivity and reliability were better because the physical system had been modeled more accurately than if the approximate model were used alone.

2. UNCERTAIN SYSTEM MODEL

2.1 Unmodeled Physics

A physical system can be characterized from first principles knowledge in the following generalized form:

\[ 0 = f_{fpm}(x,u,a) \]  \hspace{1cm} (1)

\[ y = g_{fpm}(x,u,a) \]  \hspace{1cm} (2)

where

- \( x \): the state vector \((r \times 1)\),
- \( u \): the input vector \((r \times 1)\),
- \( a \): lumped parameters,
- \( y \): the output vector \((r \times 1)\),

and the subscript \((fpm)\) denotes first principles modeling. Eqs. (1) and (2) are often incomplete known relationships. It is assumed that the unmodeled processes do not introduce additional state variables. Representing the unmodeled dynamics by vector functions \( f_m \) and \( g_m \), Eqs. (1) and (2) become:

\[ 0 = f_{fpm}(x,u,a) + f_m(x,u,w) \]  \hspace{1cm} (3)

\[ y = g_{fpm}(x,u,a) + g_m(x,u,w) \]  \hspace{1cm} (4)

where \( w \) represents the neural network weights. Therefore, \( f_m(\cdot) \) and \( g_m(\cdot) \) represent function approximators. The state, \( x \), and the neural network weights, \( w \), are calculated by fitting to the measured process data. If \( k \) denotes a time at which the forcing functions \( u \) and outputs \( y \) are sampled, the error \( E \) is minimized with respect to \( x \) and \( w \) as follows:
\[
E = \sum_{k=1}^{K} \sum_{i=1}^{m} \left( f_{fpn(i)}[x(k),u(k),\alpha] + f_{nn(i)}[x(k),u(k),w] \right)^2 \left( \frac{1}{moc_i} \right) + \sum_{k=1}^{K} \sum_{j=1}^{n} \left( y(k) - g_{fpn(j)}[x(k),u(k),\alpha] - g_{nn(j)}[x(k),u(k),w] \right)^2 \left( \frac{1}{moc_{m,j}} \right)
\]

where \( moc \) is the maximum operating condition, \( i \) references the elements of \( f_{fpn} \) and \( f_{nn} \), \( j \) references the elements of \( g_{fpn} \) and \( g_{nn} \), and \( K \) is the number of samples. A maximum operating condition \( moc \) is defined for each element of \( f_{fpn} \) and \( (y - g_{fpn}) \) and is the maximum magnitude reached by the element over all samples.

### 2.2 Estimating Uncertainty

Errors enter the model in different ways. It is unlikely that Eqs. (3) and (4) will be identically satisfied after fitting them to the process data. Process dynamics may have been excited during data collection and there may be unmodeled process state variables. Let the resulting error in the state Eq. (3) be \( \xi_x \) and assume it to be a zero mean normally distributed random vector. If \( x_a \) (where "a" denotes approximate) is the value of \( x \) calculated after fitting Eqs. (3) and (4) to the measured order pair \((u, y)\), then

\[
x = x_a + \xi_x
\]

where \( \xi_x \) is a random vector and is dependent on \( \xi_u \). The system model then becomes:

\[
y = g_{fpn}[x_a, u, \alpha, \xi_u, \xi_x] + g_{nn}[x_a, u, w, \xi_u, \xi_x] + \eta
\]

\[
\xi_x = f_{fpn}[x_a, u, \alpha, \xi_u, \xi_x] + f_{nn}[x_a, u, w, \xi_u, \xi_x]
\]

where
- \( \eta \): random zero mean vector representing output noise
- \( \xi_u \): random zero mean vector representing input noise.

The vectors \( \xi_x, \xi_u, \) and \( \eta \) are assumed independent of each other. Eqs. (7) and (8) are then linearized to obtain an error propagation formula [1]. Expanding Eqs. (7) and (8) in a Taylor series and dropping second and higher order derivatives,

\[
\xi_x = A \xi_x + B \xi_u
\]

\[
y = g_{fpn}[x_a, u, \alpha, 0, 0] + g_{nn}[x_a, u, w, 0, 0] + C \xi_x + D \xi_u + \eta
\]

where
Eq. (9) is solved for $\xi_x$ and substituted into Eq. (10) to obtain,

$$y = g_{pm}[x_a,u,\alpha,0,0] + g_{mn}[x_a,u,w,0,0] + CA^{-1}\xi_x + (D-CA^{-1}B)\xi_u + \eta$$  \hspace{1cm} (11)

The error propagation formula of Eq. (11) is used to write the likelihood function for the system. If the vectors $\xi_u$, $\xi_r$ and $\eta$ are normal and their variances are $Q_u$, $Q_r$ and $R$, respectively, then the variance of $y$ is [1]:

$$K = CA^{-1}Q_u(CA^{-1})^T + (D-CA^{-1}B)Q_u(D-CA^{-1}B)^T + R$$  \hspace{1cm} (12)

The likelihood function $y$ is then [2]:

$$f(y) = \frac{1}{(2\pi)^{n/2}(\det K)^{1/2}} e^{-\frac{1}{2}(y-\mu)^T K^{-1} (y-\mu)}$$  \hspace{1cm} (13)

where the mean value $\mu$ is given by

$$\mu = g_{pm}[x_a,u,\alpha,0,0] + g_{mn}[x_a,u,w,0,0]$$

Given ordered pairs of input and output measurements $(u_1, y_1)$, $(u_2, y_2)$, ..., $(u_N, y_N)$ and values for $Q_u$ and $R$, the log likelihood function is equal to [2]:

$$- \ln f(y_1, y_2, ..., y_n) = \frac{nN}{2} \ln 2\pi + \sum_{i=1}^{N} \frac{1}{2} \ln |K_i| + \frac{1}{2}(y_i-\mu_i)^T K_i^{-1} (y_i-\mu_i)$$  \hspace{1cm} (14)
The maximum likelihood method [2] is used to estimate the uncertainty represented by the covariance matrix $Q$. The maximum likelihood solutions are those values that maximize the likelihood function, or equivalently, minimize the negative of the log likelihood function given by Eq. (14). An iterative scheme is used to find these values.

2.3 Model Validation

The described statistical model of the system needs to be validated. Two validation tests were utilized: a batch test and an on line sequential test. The batch test is applied after training is finished but before the model is brought on line. The purpose is to ensure that training was successful and that the final model is a consistent representation of the training data. The sequential test is applied on line at each sample point and continually tests whether the model is a valid representation of the physical process. In both tests, the mathematical representation for the system is the statistical model given by Eq. (11). The on-line test is based on the likelihood ratio test of Wald [4] and summarized as follows. Suppose that a random process, $r$, uncorrelated in time is normally distributed with zero mean and depends on the parameter $\delta$. The likelihood of observing $r_i$ given the parameter value $\delta_j$ is denoted $f(r_i, \delta_j)$. The task is to estimate $\delta$ at each sampling interval. If the choice is between two values, $\delta_0$ and $\delta_1$, then form the following test statistic:

$$S_m = \frac{f(r_1, \delta_1) f(r_2, \delta_1) \ldots f(r_m, \delta_1)}{f(r_1, \delta_0) f(r_2, \delta_0) \ldots f(r_m, \delta_0)}$$

based on the sequence of observations $r_1, r_2, \ldots, r_m$. Next, form two hypotheses: $H_0$ that the parameter value is $\delta_0$, and $H_1$ that the parameter value is $\delta_1$. The decision process is as follows:

$$\begin{align*}
&\text{If } S_m \leq B \quad \text{terminate and accept } H_0, \text{ else} \\
&\text{If } B < S_m < A \quad \text{continue sampling, else} \\
&\text{If } S_m \geq A \quad \text{terminate and accept } H_1
\end{align*}$$

(16)

The likelihood ratio test was used to determine whether the model is a statistically valid representation of the measurements. A drift in the plant response away from the model was used to conclude that the model is no longer valid. The drift is represented by the parameter $\delta$ defined to be the difference between the mean of the model output and the plant output. To monitor this value, the residual vector is first formed, which is defined as the difference between the plant output $y_{\text{meas}}$ and model output $y_i$ for input $u_i$. Let both the model and plant output have the same covariance matrix $K_i$ and let the mean value of the model output be $\mu$ and the mean value of the plant output be $\mu + \delta$. The likelihood function [2] for the residual vector is then given by:

$$f(y_{\text{meas}}, \delta) = \frac{1}{(2\pi)^{n/2}|K_i|^{1/2}} e^{-\frac{1}{2}(y_{\text{meas}} - \mu - \delta)^T K_i^{-1} (y_{\text{meas}} - \mu - \delta)}$$

(17)
If the failure is to be declared when the plant mean has drifted away from the model by an amount \( \delta = \delta_{\text{bad}} \), two hypotheses: \( H_0 \), where \( \delta = 0 \); and \( H_1 \), where \( \delta = \delta_{\text{bad}} \) are defined. The decision as to whether or not a failure has occurred is then given by Eq. (16), where the likelihood function is given by Eq. (17).

3. **REAL-TIME IMPLEMENTATION**

A real-time system was developed that implements the described techniques. The monitoring system has three modes of operation: idle, training, and monitoring. The idle mode is entered on startup at which time the computer communication tasks are initialized. On entering the training mode, a finite set of input-output process data pairs is collected and used to calculate the neural network weights and estimate the stochastic parameters. The result is a fully specified model [Eq. (11)] and likelihood function [Eq. (14)]. In the training mode, no process condition diagnostic is performed. In the monitoring mode, the process output is tested for consistency with the model. The test results are posted on the operator interface display and indicate the condition of the observed process.

This real-time system was implemented in a distributed computer platform. The reading of sensors, filtering of measurements, conversion of measurement units, and the graphical user interface functions are performed on an (I/O) front-end computer. The user interacts with this computer to control the operation of the monitoring system and to observe process behavior. The training, estimation, and test statistic analyses are performed using a monitoring computer. The two computers were networked in a server/client configuration.

4. **EXPERIMENT AND RESULTS**

Tests were conducted to demonstrate the diagnostic technique’s ability to detect abnormal operation of a motor-pump set. Normal operation involved pumping water at a number of different flowrates through a circuit with fixed hydraulic characteristics. Operational anomalies included introducing a mechanical load on the pump shaft to simulate a change in bearing characteristics, switching to a more viscous fluid, and introducing a degraded pump component. Because of the positive displacement characteristic of the peristaltic pump used, the flowrate is proportional to motor-pump speed. By monitoring the change in the electrical power consumed by the motor at a given speed, the reduction in pumping capability is inferred. The goal is to model the relationship between speed and power during normal operation so that abnormal operation can be detected. A diagram of the experimental setup is given in Fig. 1.

Each of these upsets disrupted the normal operation balance between power and back emf as determined during the training session. Two different indices were used to monitor motor-pump condition. In both cases the electrical power to the motor was regarded as the input to the system while the back emf was the output. The first index was the hypothesis test of Eq. (16). The test statistic \( S_m \) was calculated at each sample point and was compared with the two thresholds, \( A \) and \( B \). A value less than \( B \) signals the motor-pump is operating normally, while a value greater than \( A \) signals that a disturbance has occurred. When the
index is between these thresholds, the operating condition cannot be determined. The second index is a visual one and is performed by the experimenter. An observation is made to see if the measured output remains within the 2σ confidence level of the predicted back emf as it should if the motor-pump is operating properly.

![Experimental Setup](image)

**Fig. 1. Experimental Setup**

The objective of the fatigued tube test was to determine whether degradation of the tube wall could be detected as an event lying outside the normal operating envelope. The pump was run for a period long enough to induce visible changes in the silicone tube. The values of the two test measures are shown in Fig. 2. The middle segment (samples 12 through 38) was recorded while the pump ran with a damaged tube while the two ends correspond to a new tube. The data show that both tests correctly signal an operating condition change. The $S_m$ indicator in the hypothesis test (left figure) moves from good to bad and then back. Similarly, the measured back emf moves outside of the 2σ interval when the damaged tube is inserted and then returns when the good tube is put back (right figure).

![Hypothesis Test and Confidence Interval](image)

**Fig. 2. Detection of Degraded Tube Condition**
The prediction from the first principles model only is also shown and is seen during normal operation to lie outside the 2σ confidence level. One can conclude that training the system without augmenting the first principles model with the neural network would have led to a larger 2σ spread, resulting in a less sensitive disturbance detection threshold.

The diagnostic result of changing the fluid being pumped is shown in Fig. 3. In this test, the pump initially drew water from the reservoir in a manner identical to that during the training session. The object was to see if the system could detect a change over to a different fluid (dish soap). A switching manifold installed in the pump inlet line was used to switch to drawing fluid from a soap reservoir. Both the $S_m$ indicator in the hypothesis test (left figure) and the 2σ test (right figure) correctly signaled the change over to the new fluid.

![Hypothesis Test and Confidence Interval](image)

Fig. 3. Detection of Change in Fluid Viscosity

The relative advantage of combining first principles information with measurements is a more reliable model that substitutes unknown physics for curve fitting. Since the motor-pump system is relatively linear, a highly nonlinear simulation was developed for demonstration purposes. This model was then used to generate training data at a number of pump speeds. Two models were fit to the training data. The first model, the approximate combined model, was composed of the training data model with parameters perturbed and a neural network. The second model, the neural network only model, was a neural network. The approximate combined model is more accurate in predicting the measurements than either the first principles only model or the neural network only model. This is achieved by combining first principles information with the mapping capability of a neural network. Because of the greater uncertainty for the neural network only model, a detection capability based on it is less sensitive than one that uses the approximate combined model. This response can be seen from Fig. 4, where the monitored back emf was perturbed by 15% without changing the power. The anomaly is detected by the capability based on the approximate combined model but not that of the neural network only model. The simulation shown in Fig. 4 begins with data samples that correspond to normal operation. The
diagnostic system computes a value that indicates a "Good" operational state when using either the combined model approach or the neural network only model approach. As seen in Fig. 4, when the anomaly is introduced after 30 samples, the diagnostic system using the combined model strategy is able to detect the anomaly and posts an output value that signals a "Bad" operating state. However, when using the neural network only model approach, the diagnostic system is unable to detect the abnormal state. When the anomaly was removed at around 80 samples, the diagnostic system correctly signals a return to a "Good" operating state.

Fig. 4. Detection Sensitivity for Highly Nonlinear Case

5. SUMMARY AND CONCLUSIONS

An approach for modeling physical systems with unknown dynamics was presented. Processes that can be readily interpreted as conservation balances are written as a set of differential equations. Processes that are unknown are represented as an input-output mapping acquired using a neural network. In this way, the proposed method combines both representations into a single model structure. After computing the process output predicted by the model based on measured process inputs, a likelihood test is then performed to decide whether the model and the measurements are consistent with the assumptions. Experimental results show that because the physical system can be modeled more accurately, the detection sensitivity and reliability are improved over a physic-based or neural network model alone.

ACKNOWLEDGMENTS

The methods described in Section 2 were translated into software by F.W. Chen. He also developed a method for analytically determining the derivatives of the outputs of a neural networks with respect to the inputs.
REFERENCES


