

Fault Diagnosis and Temperature Sensor Recovery for an Air-Handling Unit

Won Yong Lee, Ph.D.

John M. House, Ph.D.
Associate Member ASHRAE

Dong Ryul Shin, Ph.D.

ABSTRACT

This paper describes the use of a two-stage artificial neural network for fault diagnosis in a simulated air-handling unit. The stage one neural network is trained to identify the subsystem in which a fault occurs. The stage two neural network is trained to diagnose the specific cause of a fault at the subsystem level. Regression equations for the supply and mixed-air temperatures are obtained from simulation data and are used to compute input parameters to the neural networks. Simulation results are presented that demonstrate that, after a successful diagnosis of a supply air temperature sensor fault, the recovered estimate of the supply air temperature obtained from the regression equation can be used in a feedback control loop to bring the supply air temperature back to the setpoint value. Results are also presented that illustrate the evolution of the diagnosis of the two-stage artificial neural network from normal operation to various fault modes of operation.

INTRODUCTION

The presence of faults and the influence they have on system operation is a real concern in the heating, ventilating, and air-conditioning (HVAC) community. A fault can be defined as an inadmissible or unacceptable property of a system or a component. Unless corrected, faults can lead to increased energy use, shorter equipment life, and uncomfortable and/or unhealthy conditions for building occupants.

Faults are not a new problem in the HVAC industry; however, technological advances have helped create both a need and an avenue for the development of fault detection and diagnostic tools. The need has been created by the ever-increasing complexity of the HVAC systems and control strategies that are being installed in buildings today. In many cases the complexity of the systems exceeds the understanding of the building operator. When this occurs, faults may go undetected or, perhaps

worse, may be "corrected" by introducing changes to the system that compensate for the fault rather than eliminating it. The latter scenario could lead to energy waste and possibly to subsequent faults that are related to the initial (and still existing) fault. Technological advances have also made it possible to monitor these complex systems, thus providing the information that is needed to characterize and understand the current operating status of the systems. Fault detection and diagnostic methods can provide a bridge between possessing information and understanding its meaning.

One of the main purposes of fault detection and diagnosis is to detect failures of actuators and sensors that are used in the control systems. To improve the operational reliability of systems in general, it is necessary to validate measured sensor data, isolate failed sensors, and recover the failed measurement. Hence, sensor recovery is an important aspect of comprehensive fault detection and diagnostic methods.

In previous papers, Lee et al. (1996a, 1996b) describe methods for fault detection and diagnosis in a variable-air-volume (VAV) air-handling unit (AHU). One approach used in those studies was to define residuals that provide a measure of the difference between the existing state of the system and the normal state. Residuals that are significantly different from zero represent the occurrence of a fault. Lee et al. (1996b) described the use of a single artificial neural network (ANN) trained on idealized residual patterns to diagnose faults in various subsystems of the AHU. A similar approach was applied by Merchawi and Kumara (1994) to diagnose faults in a nuclear reactor and by Watanabe et al. (1994) to diagnose faults in a chemical reactor. Training a network such as the one described by Lee et al. (1996b), which accounts for all considered faults, can require extensive computational resources due in part to the number of inputs, hidden neurons, and outputs necessary to discriminate each pattern. In addition, if a new fault is added to the existing set, it is necessary to retrain the ANN because the

Won Yong Lee is a senior researcher at the Korea Institute of Energy Research, Taejon, Korea. John M. House is a mechanical engineer in the Mechanical Systems and Controls Group, Building Environment Division, Building and Fire Research Laboratory, National Institute of Standards and Technology, Gaithersburg, Md. Dong Ryul Shin is a principal researcher at the Korea Institute of Energy Research, Taejon, Korea.

THIS PREPRINT IS FOR DISCUSSION PURPOSES ONLY. FOR INCLUSION IN ASHRAE TRANSACTIONS 1997, V. 103, Pt. 1. Not to be reprinted in whole or in part without written permission of the American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., 1791 Tullie Circle, NE, Atlanta, GA 30329. Opinions, findings, conclusions, or recommendations expressed in this paper are those of the author(s) and do not necessarily reflect the views of ASHRAE. Written questions and comments regarding this paper should be received at ASHRAE no later than February 14, 1997.

knowledge stored in the network (values of weights and biases) is probably not adequate to discriminate the new fault.

The objective of this paper is twofold. The first objective is to describe an architecture for a two-stage ANN for fault diagnosis that can alleviate, to a certain degree, the problems discussed in the previous paragraph. The second objective is to describe the use of regression equations for sensor recovery of failed temperature sensors. The two-stage ANN and the sensor recovery method are demonstrated using data obtained from a simulation model of a laboratory-scale AHU. Results based on experimental data are not presented here.

The first sections of this paper provide a brief description of the laboratory-scale AHU, models of the AHU components that are used in the simulations, residuals used in the fault diagnosis, and the faults under consideration. The fault diagnosis and sensor recovery methods are then discussed and results from the methods are presented. Finally, conclusions and recommendations for future work are presented.

AHU COMPONENT MODELS

The AHU simulation model is a simplified dynamic model based on steady-state characteristic equations and approximate first-order dynamics. Component models developed in IEA Annex 17 (Wang 1992) have been modified to fit experimental data obtained from a laboratory-scale AHU.

A schematic diagram of the laboratory-scale VAV AHU is shown in Figure 1. This system was used in papers by Lee et al. (1996a, 1996b). The AHU consists of fans, dampers, a cooling coil, sensors, and controllers. The static pressure in the main supply duct is maintained at a constant setpoint value of 249 Pa (1.0 in. water) by sensing the static pressure and controlling the rotational speed of the supply fan. The supply air temperature is controlled by modulating the cooling water control valve to maintain a constant setpoint value of 14.5°C (58.1°F). The airflow rate difference between the supply and return airstreams is controlled by the variable-speed return fan to maintain a constant setpoint value of 472 L/s (1,000 cfm). A PID algorithm is used to control the cooling water valve, and PI algorithms are used to control the supply and return fan speeds.

The ensuing subsections give a brief overview of the component models used in the AHU simulation model. Numerous simplifying assumptions have been introduced into the simulation model; however, the model successfully captures the trends associated with various modes of operation of the laboratory-scale AHU. Thus, for purposes of this study, the model is sufficient.

Cooling Coil Characteristics

The air temperature at the exit of the cooling coil, T_{ao} , is obtained by combining a steady-state model of the cooling coil with approximate first-order dynamics. The response of T_{ao} is approximated by

$$\frac{T_{ao}}{T_{aos}} = \frac{1}{1 + \tau s} \exp(-\tau_d s) \quad (1)$$

where T_{aos} is the steady-state value of T_{ao} , τ is the process time constant, τ_d is the process delay time, and s is the Laplace variable. To determine T_{aos} , it is first necessary to compute the steady-state value of the moist air enthalpy at the exit of the cooling coil (h_{aos}), which is given by

$$h_{aos} = h_{ai} + \varepsilon(h_{wi} - h_{ai}), \quad (2)$$

where h_{ai} is the moist air enthalpy at the inlet to the cooling coil, h_{wi} is the moist air enthalpy at the inlet chilled-water temperature, and ε is the effectiveness of the cooling coil. T_{aos} is then calculated from the expression

$$h_{aos} = c_{pa}T_{aos} + w_{ao}(c_{pv}T_{aos} + h_{g,o}) \quad (3)$$

where c_{pa} is the specific heat of dry air, c_{pv} is the specific heat of water vapor, w_{ao} is the humidity ratio of the air at the outlet to the cooling coil, and $h_{g,o}$ is the enthalpy of saturated water vapor at a reference temperature T_o . For a single-pass cross-flow heat exchanger where the fluid having the minimum capacitance (air, in this case) is mixed and the water is unmixed, the effectiveness is given by (Incropera and DeWitt 1985):

$$\varepsilon = 1 - \exp(-C_r^{-1} \{1 - \exp[-C_r(\text{NTU})]\}) \quad (4)$$

where C_r is the heat capacity rate, given by

$$C_r = C_{min}/C_{max}, \quad (5)$$

and NTU is the number of transfer units, given by

$$\text{NTU} = UA/C_{min}. \quad (6)$$

As stated previously, C_{min} and C_{max} are the capacitances for air and water, respectively. UA is the overall heat transfer coefficient of the coil.

Braun et al. (1989) describe a more detailed cooling coil model that includes an analysis of dry coil heat transfer characteristics, wet coil heat transfer characteristics, and combined wet and dry coil heat transfer characteristics. The main difference in the model used here and the model described by Braun et al. (1989) is in the computation of the air capacitance when the coil is wet.

Damper Characteristics

The airflow rates through the exhaust, recirculation, outdoor, and VAV box dampers are given by

$$Q = K\sqrt{\Delta P} \quad (7)$$

where K is a resistance coefficient, Q is the volumetric airflow rate, and ΔP is the pressure drop across the dampers. The resistance coefficients are determined from least-squares regression of the experimental data for varying damper blade angles. The resulting third-order polynomial equations are of the general form

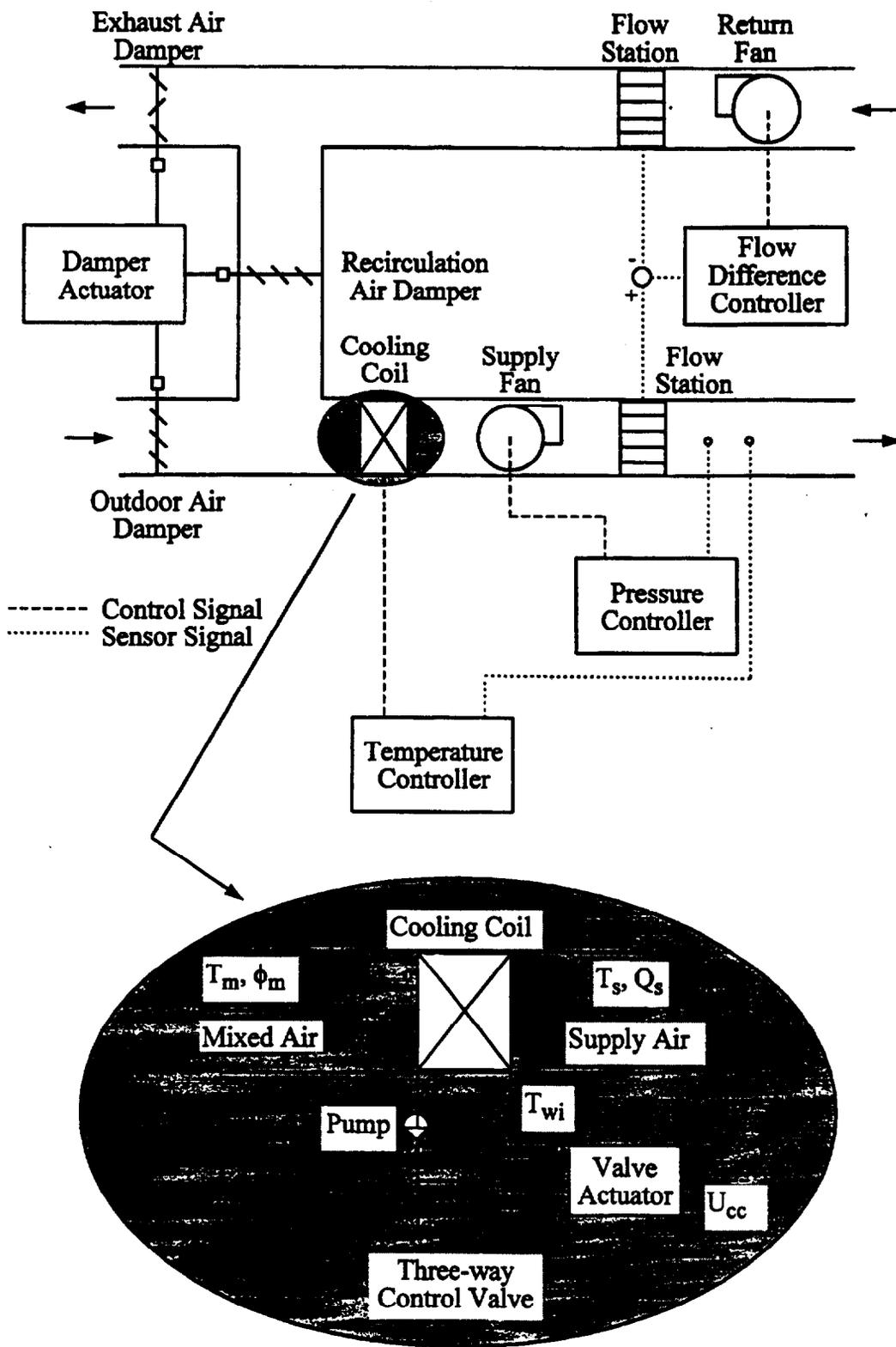


Figure 1 System model for a VAV AHU.

$$K = C_0 + C_1\theta + C_2\theta^2 + C_3\theta^3 \quad (8)$$

where θ is the damper angle. Assuming the air density is constant throughout the system, conservation of mass yields

$$Q_{rec} = Q_{ret} - Q_{exh} \quad (9)$$

$$Q_{rec} = Q_s - Q_{out} \quad (10)$$

and

$$Q_s = \sum_{i=1}^n Q_{vav,i} \quad (11)$$

where the subscript *rec* denotes recirculation air, *ret* denotes return air, *exh* denotes exhaust air, *s* denotes supply air, *out* denotes outdoor air, and *vav,i* denotes airflow through the *i*th VAV box, and where *n* is the total number of VAV boxes.

Sensor Model Characteristics

Sensors convert physical signals into electric signals. The response of the electric signal may exhibit a time delay from the change of the physical value. The sensor response is simulated using a first-order model with a simple time constant, that is

$$\frac{T_o}{T_i} = \frac{1}{1 + \tau s} \quad (12)$$

Fan Characteristics

Fan performance is characterized using the following dimensionless coefficients:

$$\Phi = \frac{Q}{D^3 N} \quad (13)$$

$$\Psi = \frac{\Delta P}{\rho(ND)^2} \quad (14)$$

where Q is the volumetric airflow rate of air through the fan, D is the fan impeller diameter, N is the rotational speed of the fan, ΔP is the pressure rise across the fan, and ρ is the density of air. The dimensionless parameters in Equations 13 and 14 are related by the polynomial expression

$$\Psi = \sum_{i=0}^m a_i \Phi^i \quad (15)$$

where the coefficients a_i ($i = 0$ to m) are determined from least-squares regression.

Air Filter Characteristics

The airflow rate and pressure drop across the air filter and the duct, including the coils, are related by

$$Q_f = K_f \sqrt{\Delta P_f} \quad (16)$$

where K_f is a friction coefficient that is treated as being constant. In real systems, K_f will vary due to fouling of the filter and the coils.

SIMULATION SOLUTION PROCEDURE

The simplified model equations are solved by considering the pressure and flow equations independently of the equations that characterize the cooling coil response. Initial values of the damper positions, valve positions, and fan speeds are selected first. The pressure and flow equations are then solved simultaneously to determine the pressure and airflow rates throughout the system. Next, the equations governing the cooling coil are solved to determine the supply air temperature and the inlet and exit cooling water temperatures. The supply air temperature and pressure and the volumetric airflow rate difference between the supply and return airstreams are then used to compute the control signals to the fans and the cooling coil valve. Finally, the control signals are converted to new values of the supply and return fan speeds and the cooling coil valve position. The solution procedure is then repeated. The simulation requires solving 15 equations (12 for the pressure and flow characteristics, 3 for the cooling coil) for 15 unknowns.

RESIDUAL DEFINITION

The fault diagnosis method described by Lee et al. (1996b) identifies patterns of residuals that can be used as signatures for various faults. An ANN was trained with these patterns and then used to diagnose the status of the AHU for actual experimental data. The set of residuals used in this study is slightly different from that used in Lee et al. (1996b). The residuals used here are:

$$R_{T_s1} = T_s - T_{s,ev} \quad (17)$$

$$R_{T_s2} = T_s - T_{s,sp} \quad (18)$$

$$R_{T_m} = T_m - T_{m,ev} \quad (19)$$

$$R_{P_s} = P_s - P_{s,sp} \quad (20)$$

$$R_{Q_d} = Q_d - Q_{d,sp} \quad (21)$$

and

$$R_{V_{cc}} = V_{cc} - U_{cc} \quad (22)$$

where

- R = residual,
- T = temperature,
- P = pressure,
- Q = volumetric airflow rate,
- U = control signal,
- V = valve position,
- m = mixed air parameters,
- d = flow difference parameters,
- cc = cooling coil parameters,
- sp = setpoint values, and
- ev = expected values calculated using regression equations.

To compute residuals $R_{T_{s1}}$ and R_{T_m} , models are needed for the expected value of the supply air temperature and the mixed air temperature. For this study, regression equations are used to estimate the current values of T_s and T_m for normal operating conditions. A schematic diagram of the cooling coil and the cooling coil valve subsystem is shown in the expanded portion of Figure 1. T_m and ϕ_m are the mixed air temperature and relative humidity, respectively; Q_s is the supply airflow rate; and T_{wi} is the temperature of the cooling water at the inlet to the cooling coil. The other variables retain their previous definitions.

The input and output variables for the regression equation for T_s are:

Inputs

$$\begin{aligned} &Q_s(i), Q_s(i-1), Q_s(i)^2, Q_s(i-1)^2 \\ &T_m(i), T_m(i-1) \\ &T_{wi}(i), T_{wi}(i-1) \\ &\phi_m(i), \phi_m(i-1) \\ &U_{cc}(i), U_{cc}(i-1), U_{cc}(i-2), U_{cc}(i)^2, U_{cc}(i-1)^2 \\ &Q_s(i) T_m(i), Q_s(i-1) T_m(i-1) \\ &U_{cc}(i) T_{wi}(i), U_{cc}(i-1) T_{wi}(i-1) \end{aligned}$$

Output

$$T_{s,ev}(i)$$

where (i) refers to the current discrete time value and $(i-1)$ refers to the previous value.

The coefficients of the regression equation are computed using simulation data obtained as the system operates in a normal mode. The training data consist of 1,000 points. The correlation coefficient for $T_{s,ev}$ is 0.98, indicating that the regression model of the supply air temperature accounts for 98% of the variability of the simulated supply air temperature for normal conditions. A model for the expected value of the mixed air temperature, $T_{m,ev}$, is obtained in a similar manner. The correlation coefficient for $T_{m,ev}$ is 0.95.

It should be pointed out that the regression models for $T_{s,ev}$ and $T_{m,ev}$ could be replaced by other types of models such as physical models or neural network models. Regression models were chosen because of their limited complexity and because physical models are already used in the simulation model of the air-handling unit. The physical models currently used in the simulation model could be used in a laboratory or a real building study to provide an estimate of these temperatures. Regardless of the type of model that is used, laboratory and/or real building studies require a certain amount of training to compute regression coefficients or unknown physical parameters.

FAULT DESCRIPTION

Eleven faults and their dominant symptoms are described in the following paragraphs. The dominant symptoms are identified by simulating each of the faults and observing the response of the residuals. Faults are introduced when the system is operating at normal, steady-state conditions and the dominant symptoms correspond to the steady-state conditions after a fault has occurred. Fasolo and Seborg (1994) applied fault detection and

diagnostic methods to a simulated heating coil subsystem. Several of the faults they considered are described below.

Fault 1 is a failure of the supply fan. The dominant residuals for fault 1 are R_{P_s} , R_{Q_d} , $R_{T_{s1}}$, and R_{T_m} . The fault causes the supply air pressure to decrease to zero and the control signal to the supply fan to increase to its maximum value in an attempt to offset the decreasing supply air pressure. The control signal for the return fan decreases to zero in an attempt to maintain the flow difference between the supply and return air ducts at the setpoint value. The two temperature residuals are large because both $T_{s,ev}$ and $T_{m,ev}$ are affected by the abrupt change in the supply airflow rate and/or rotational speeds of the fans.

Fault 2 is a failure of the return fan. The dominant residual for fault 2 is R_{Q_d} . The fault causes the return fan rotational speed to decrease to zero, thereby causing the flow difference between the supply and return ducts to increase.

Fault 3 is a failure of a local feed water pump. The dominant residual for fault 3 is $R_{T_{s1}}$. The fault causes the water flow rate to decrease, but not to zero, since the main supply pumps continue to operate normally. The decrease in the flow rate of cooling water causes the supply air temperature to increase, with the resultant effect being that the cooling coil valve opens further. By opening the cooling coil valve, it is possible to bring the supply air temperature back to the setpoint value (unless the fault is too severe, in which case $R_{T_{s2}}$ could also be a dominant symptom); however, the control signal to the cooling coil valve will be different from the normal condition. This causes $T_{s,ev}$ to be different from T_s .

Fault 4 is a stuck cooling coil valve. The dominant residuals for fault 4 are $R_{T_{s1}}$, $R_{T_{s2}}$, and $R_{V_{cc}}$. A load change occurring after the introduction of the fault will cause the control signal to the cooling coil valve to saturate at either the minimum or the maximum voltage because the valve is unable to respond to the control input. $R_{T_{s1}}$ and $R_{T_{s2}}$ are dominant because $T_{s,ev}$ decreases (increases) as the cooling coil valve control input increases (decreases), while T_s increases (decreases) and $T_{s,sp}$ remains constant.

Fault 5 is a complete failure of the supply air temperature sensor. The dominant residuals for fault 5 are $R_{T_{s1}}$ and $R_{T_{s2}}$. A temperature sensor failure typically results in a voltage signal that varies randomly between large positive and negative values. If this occurs, the temperature is automatically set to zero so that the temperature residuals given by Equations 17 and 18 do not fluctuate. Setting the supply air temperature signal to zero causes the cooling coil control valve to close, thereby causing $T_{s,ev}$ to increase to some value greater than $T_{s,sp}$.

Fault 6 is a second type of failure of the supply air temperature sensor. The dominant residuals for fault 6 are $R_{T_{s1}}$ and $R_{T_{s2}}$. In this case, the sensor drops from its supporting harness onto the floor of the duct, giving an incorrect temperature reading. Because the duct surface is assumed to be at a temperature that is 5°C (9°F) higher than the air flowing through the duct, the controller attempts to compensate by opening the cooling coil control valve. The sensed supply air temperature at time i is given by

$$T_s(i) = T_s(j) + 5[1 - \exp(-[i-j]/100)] \quad (23)$$

where $T_s(j)$ is the sensed supply air temperature at time j when the fault occurs and the time constant for the response is assumed to be 100 seconds.

Fault 7 is a third type of supply air temperature sensor failure and is due to sensor drift. The dominant residual for fault 7 is R_{T_s1} . This type of failure is classified as a performance degradation rather than a complete fault and would be difficult to detect in its early stages. To simplify the discussion for this fault, the period over which the sensor degrades is short, approximately 30 minutes. The fault is simulated by creating a linear decrease in the supply air sensor reading until the maximum offset value of 1.5°C (2.7°F) is achieved. The cooling coil valve controller compensates for this fault by gradually closing the control valve. Thus, the sensed value of the supply air temperature is maintained at the setpoint value, although the actual temperature is higher than the setpoint value.

Fault 8 is a failure of the supply air pressure transducer. The dominant residuals for fault 8 are R_{P_s} and R_{T_m} . When this failure occurs, a zero reading is obtained for the supply air pressure. This causes the control signal to the supply fan to increase to its maximum value, thus causing the supply fan speed to increase to its maximum value. The change in the supply fan speed produces a large change in $T_{m,ev}$.

Fault 9 is a failure of the supply airflow station. The dominant residual for fault 9 is R_{Q_d} . When this fault occurs, a zero reading is obtained for the supply flow station and the return fan controller decreases the return fan speed to its minimum value in an attempt to maintain the flow difference between the supply and return ducts at the setpoint value.

Fault 10 is a failure of the return fan flow station. The dominant residual for fault 10 is R_{Q_d} . When this fault occurs, a zero reading is obtained for the return flow station and the return fan controller increases the return fan speed to its maximum value in an attempt to maintain the flow difference between the supply and return ducts at the setpoint value.

Fault 11 is a failure of the mixing box damper linkage. The dominant residual for fault 11 is R_{T_m} . The fault causes a discrepancy between actual and expected values of the airflow rates for the airstreams entering and exiting the mixing box. The discrepancy in the airflow rates leads to discrepancies in the actual and expected temperatures in the mixing box. It is assumed that the recirculation damper is closed when the fault occurs.

FAULT DIAGNOSIS METHOD

Steady-State Detection

The fault diagnosis method is developed with the assumption that steady-state conditions exist when classification of the symptoms is performed. Thus, a steady-state detector is used to filter the data. If steady-state conditions prevail, the fault diagnosis algorithm is invoked.

Three variables are used in the steady-state detector, namely, supply air temperature, supply air pressure, and the flow

rate difference between the supply and return air ducts. Using least-squares regression, straight lines are fit through the current and five previous values of each of these variables. If the absolute value of the slope of each line is less than its associated threshold value, the system is deemed to be in steady state. Otherwise, the system operation is unsteady and no further classification is possible. The threshold value for the slope of a regression line is approximately three times the value of the average slope for a particular variable under normal conditions, with obvious transient periods removed from consideration in determining this slope.

Normalization

Residuals are normalized so that the dominant symptom residuals have approximately the same magnitude for the different fault cases. The residuals are normalized using the following expression:

$$\bar{R} = \frac{|R| - R_{min}}{R_{max} - R_{min}} \begin{cases} \text{If } \bar{R} < 0 \text{ then set } \bar{R} = 0 \\ \text{If } \bar{R} > 1 \text{ then set } \bar{R} = 1 \end{cases} \quad (24)$$

where R_{min} and R_{max} must be specified for each residual. A deadband exists between 0 and R_{min} in which \bar{R} is taken to be zero. Values of \bar{R} greater than R_{max} yield values of \bar{R} equal to unity. Values of R_{min} and R_{max} could be determined by computing the standard deviation of each residual under normal, steady-state operating conditions. As an example, R_{min} could be defined as 3σ and R_{max} as 6σ , where σ is the standard deviation for a particular residual. This approach assumes that the normal data are representative of all normal operating conditions, which may not be a valid assumption. Alternatively, expert knowledge could be used to assign values to R_{min} and R_{max} . The second approach is utilized in this study. For instance, for the temperature residuals, R_{min} and R_{max} are defined as 0.5°C (0.9°F) and 1.5°C (2.7°F), respectively. The assignment of values for R_{min} and R_{max} is an important aspect of the fault diagnosis method and to a large extent defines the type and severity of faults that can be diagnosed.

Two-Stage Artificial Neural Network

To use an ANN for fault diagnosis, the ANN must first be trained using data that represent the normal condition and the various fault conditions. Lee et al. (1996b) used a single ANN to classify the operating status of the AHU. Nine possible modes of operation were considered, namely, the normal mode and eight separate fault modes. As stated previously, training a network such as that, which includes patterns for all considered faults, can require fairly extensive computational resources. In addition, the ANN must be retrained to discriminate new faults. To lessen the impact of such problems, an architecture for a two-stage ANN is proposed here. Stage one is used to classify the subsystem in which a fault is occurring. Stage two is used to diagnose the cause of a fault on the subsystem level. Using this architecture, less information is required for diagnosis at a given stage. In this study the subsystem classifications are the pressure control subsystem, the flow control subsystem, the cooling coil subsystem, and the

mixing box damper subsystem. Figure 2 shows a block diagram of the proposed two-stage network for the case where n_s subsystems are considered, and the number of outputs of the i th stage two ANN is m_i . Although four subsystems are considered in this study, stage two ANN results are presented only for the cooling coil subsystem. Hence, reference to stage two ANN parameters or results throughout the remainder of this paper will imply those of the cooling coil subsystem.

Idealized patterns of normalized residuals are specified for training by considering the dominant symptoms of each fault (Lee et al. 1996b; Watanabe et al. 1994). The idealized input/output training patterns for the stage one ANN are given in Table 1, and idealized training patterns for the stage two ANN to diagnose faults in the cooling coil subsystem are given in Table 2. The input

TABLE 1 Idealized Stage One ANN Training Patterns for the AHU Fault Diagnosis

Net Inputs				Net Outputs						Fault Diagnosis
\bar{R}_{Ps}	\bar{R}_{Qd}	\bar{R}_{Ts1}	\bar{R}_{Tm}	y_1	y_2	y_3	y_4	y_5	y_6	
0	0	0	0	1	0	0	0	0	0	Normal
1	0	0	1	0	1	0	0	0	0	Pressure control subsystem
0	1	0	0	0	0	1	0	0	0	Flow control subsystem
0	0	1	0	0	0	0	1	0	0	Cooling coil subsystem
0	0	0	1	0	0	0	0	1	0	Mixing box damper subsystem
?	?	?	?	0	0	0	0	0	1	Unknown

patterns are based on conditions expected to exist after the system has reached steady state. A dominant symptom residual is assigned a value of 1, and all other symptoms are assigned a value of 0. Each output training pattern consists of five values of 0 and one value of 1 for the stage one ANN. Outputs in the stage one ANN are denoted y_i , where $i = 1$ to 6, and outputs in the stage two ANN are denoted z_j , where $j = 1$ to 5. An output pattern of [1 0 0 0 0] signifies normal operation, [0 1 0 0 0] signifies a fault in the pressure control subsystem, and so on. An output pattern of [0 0 0 0 1] signifies unknown operation. There is no ideal input pattern for unknown operation, thus the inputs in Tables 1 and 2 for unknown operation are designated with question marks. This classification is used when the input pattern does not demonstrate the characteristics of any of the ideal input training patterns. For the stage two ANN, each output training pattern consists of four values of 0 and one value of 1.

TABLE 2 Idealized Stage Two ANN Training Patterns for the Cooling Coil Subsystem Fault Diagnosis

Net Inputs			Net Outputs					Fault Diagnosis
\bar{R}_{Ts1}	\bar{R}_{Ts2}	\bar{R}_{Vcc}	z_1	z_2	z_3	z_4	z_5	
0	0	0	1	0	0	0	0	Normal
1	1	0	0	1	0	0	0	Temperature sensor failure (fault #5 or #6)
1	0	0	0	0	1	0	0	Temperature sensor degradation (fault #7) or pump failure (fault #3)
1	1	1	0	0	0	1	0	Valve failure (fault #4)
?	?	?	0	0	0	0	1	Unknown

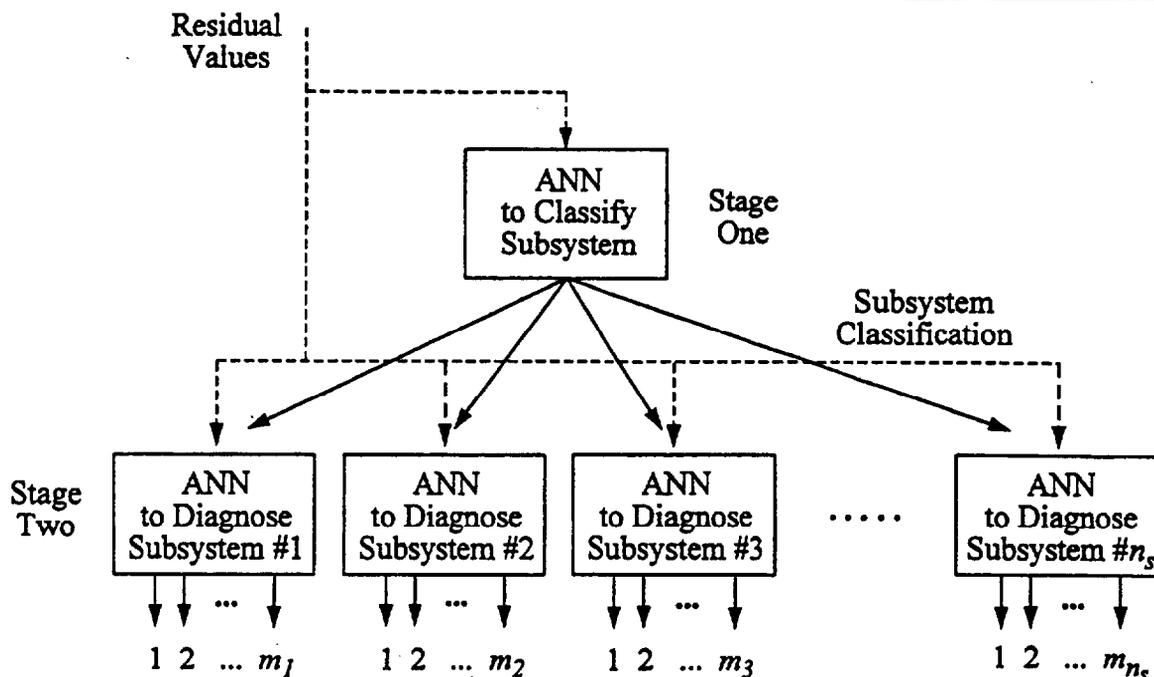


Figure 2 Two-stage ANN for fault diagnosis of an AHU.

Lee et al. (1996b) and Watanabe et al. (1994) used only the idealized input/output training patterns to train the ANNs used for classifying system operation. In those studies, the number of weights and biases of the neural networks are larger than the number of training patterns. In this study, the idealized training patterns are replaced with less ideal patterns that are obtained by replacing residuals that are ideally unity with a random value taken from a uniform distribution ranging from 0.5 to 1 and by replacing residuals that are ideally zero with a random value taken from a uniform distribution ranging from 0 to 0.5. The idealized output patterns are also replaced with patterns that classify the given input pattern as belonging, for example, partially to one or more subsystem faults and partially to the unknown classification. This approach retains the trends associated with the idealized patterns, while simultaneously expanding the input space so that the trained ANN is capable of logically classifying patterns that fall somewhere between idealized input patterns.

To clarify the method for generating the training patterns, the algorithm used to construct patterns for the pressure control subsystem is summarized below.

Typical Input Pattern

\bar{R}_{Ps} = random number between 0.5 and 1 taken from a uniform distribution

\bar{R}_{Qd} = random number between 0 and 0.5 taken from a uniform distribution

$\bar{R}_{T_{s1}}$ = random number between 0 and 0.5 taken from a uniform distribution

\bar{R}_{T_m} = random number between 0.5 and 1 taken from a uniform distribution.

Typical Output Pattern

$[y_1 y_2 y_3 y_4 y_5 y_6]$ where

$$y_1 = 0$$

$$y_2 = 1 / \{1 + \exp[-10(\{\bar{R}_{Ps} + \bar{R}_{T_m}\} / 2 - 0.5)]\}$$

$$y_3 = 0$$

$$y_4 = 0$$

$$y_5 = 0$$

$$y_6 = 1 - y_2$$

Thus, for values of \bar{R}_{Ps} and \bar{R}_{T_m} near unity, such as the idealized pattern for faults of this subsystem, y_2 approaches unity and y_6 approaches zero. For values of \bar{R}_{Ps} and \bar{R}_{T_m} near 0.5, y_2 and y_6 approach 0.5. This classification is interpreted to mean that the pattern is like that of a fault of the pressure control subsystem to a degree 0.5 and is unlike any of the known patterns (thus classified as unknown operation) to a degree 0.5. This approach is used to generate 100 training patterns for

pressure control subsystem faults. Similar algorithms are used to generate 100 patterns for normal operation and 100 patterns for each of the other subsystem faults.

Because the idealized patterns consist of four residuals for the stage one ANN, there are a total of 2^4 possible binary patterns, with five of these patterns being listed in Table 1. Algorithms similar to that described for the pressure control subsystem are used to generate 50 training patterns for each of the other 11 ideal binary input patterns. In all cases, the value of y_6 ranges from 0.5 to 1, indicating that the pattern is not like any of the known patterns to a degree between 0.5 and 1.

The stage one and stage two ANN architectures are $4 \times 12 \times 8 \times 6$ and $3 \times 15 \times 5$, respectively, where the first number is the number of inputs, the last number is the number of outputs, and the middle number(s) is the number of neurons in the hidden layer(s). Note that the stage one ANN has two hidden layers containing 12 and 8 neurons, respectively. For the stage one ANN, 1050 input/output patterns are used for training, and for the stage two ANN, 800 patterns are used for training. The networks are trained until the sum-of-squares error is less than 3. A commercial ANN software package was used for the training (Demuth and Beale 1992).

SENSOR RECOVERY METHOD

When a critical sensor reading is found to be erroneous, it is necessary to estimate its true value using correlated measurements. A simple approach is to have one estimating relation for each sensor reading that needs to be recovered. To demonstrate this concept, the regression equation used to compute the expected value of the supply air temperature $T_{s,ev}$ is used to recover an estimate of the supply air temperature when this sensor fails. Once the fault is diagnosed, the estimate of the supply air temperature is used in the feedback control loop to recover control of the actual supply air temperature.

The sensor recovery method is expected to be successful assuming the inputs to the regression equation are not erroneous. That is, if the conditions in the system deviate from normal conditions (for instance, if the cooling coil valve input signal saturates at its minimum value) because of the temperature sensor fault, the regression equation should yield an accurate estimate of the actual supply air temperature and, therefore, be useful for regaining control of the system. If, however, the conditions deviate because of another type of fault and this fault affects the inputs to the regression equation, the regression equation will not provide an accurate estimate of the actual supply air temperature. The difference between the estimated and actual supply air temperatures will be used to diagnose the fault. Hence, the regression equation is only used for sensor recovery after that particular temperature sensor has been diagnosed as faulty. Results demonstrating sensor recovery are presented in the following section.

RESULTS AND DISCUSSION

Fault Diagnosis

The training phase of the two-stage ANN used input patterns adapted from the idealized input patterns given in Tables 1 and 2. In the testing phase, data obtained from a simulation program based on simplified AHU component models are used as inputs. Faults are introduced through modifications of the computer algorithm. Faults are introduced at $t = 2000$ seconds, except where noted otherwise. Stage one fault diagnoses for the pressure control subsystem, the flow control subsystem, and the mixing box subsystem are presented first. The two-stage diagnosis for faults in the cooling coil subsystem are then presented.

The ANN output associated with normal operation (y_1), unknown operation (y_6), and pressure control subsystem faults (y_2) are plotted as functions of time in Figure 3 for the pressure control subsystem faults (faults 1 and 8). The other ANN output values are typically near zero and, therefore, are not presented. In the results that follow, only the dominant outputs, and sometimes the associated dominant residuals, are plotted to simplify the presentation and discussion of the results. In some of the results, residuals and ANN output values greater than unity are observed. This is strictly an artifact of the plotting package, as these values are bounded by zero and unity in the fault diagnosis method.

During the initial 2000 seconds of the simulation of the supply fan fault (fault 1), the ANN output indicates that the system operation is normal (see Figure 3a). At approximately $t = 2250$ seconds, y_1 begins to decrease, while y_2 and y_6 gradually increase. Eventually the diagnosis evolves to the point where the degree of belief that the operation represents a pressure control subsystem fault is slightly greater than 0.5 and the degree of belief that the operation is unknown is slightly less than 0.5. For the supply pressure transducer fault (fault 8), Figure 3b reveals that the diagnosis evolves to where the degree of belief that oper-

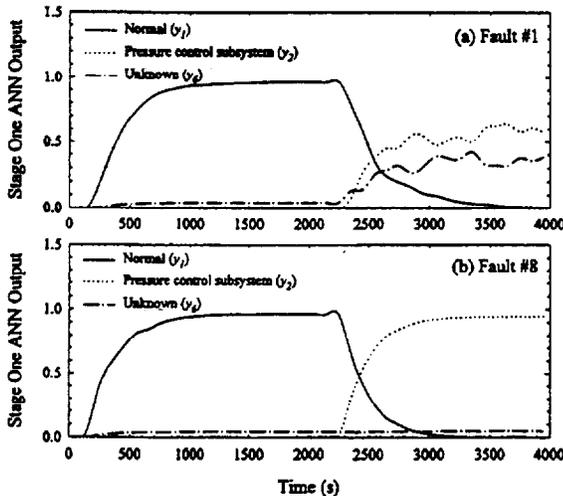


Figure 3 Fault diagnosis for pressure control subsystem faults (faults 1 and 8).

ation represents a fault in the pressure control subsystem is nearly equal to unity.

The results in Figure 3 demonstrate two important points. First, even though the fault occurs at $t = 2000$ seconds, the first indication of the fault appears approximately 250 seconds later. This is because the steady-state detector classifies the system operation as unsteady from $t = 2000$ to 2250 seconds, and therefore no new ANN classifications take place during this time. The second important point has to do with the gradual increases and decreases that are seen in the ANN outputs in Figure 3 despite the fact that abrupt faults have occurred. These gradual changes are created by the use of a filter that calculates a moving average of each ANN output value. This particular filter tends to suppress abrupt diagnoses and makes it necessary for the symptoms of a given fault to persist for an extended period before the diagnosis is made. The advantage of using this filter is that the diagnoses are not as likely to be impacted by spikes in the data. The disadvantage is that the time required to obtain a diagnosis is prolonged.

Results of the diagnoses for faults in the flow control subsystem are shown in Figure 4 and the diagnosis of the mixing box subsystem fault is shown in Figure 5. The diagnoses in Figure 4 are all similar and in each case, the correct diagnosis is made. In all cases, a gradual transition from a diagnosis of normal operation to a diagnosis of a fault in the flow control subsystem occurs after $t = 2000$ seconds. The

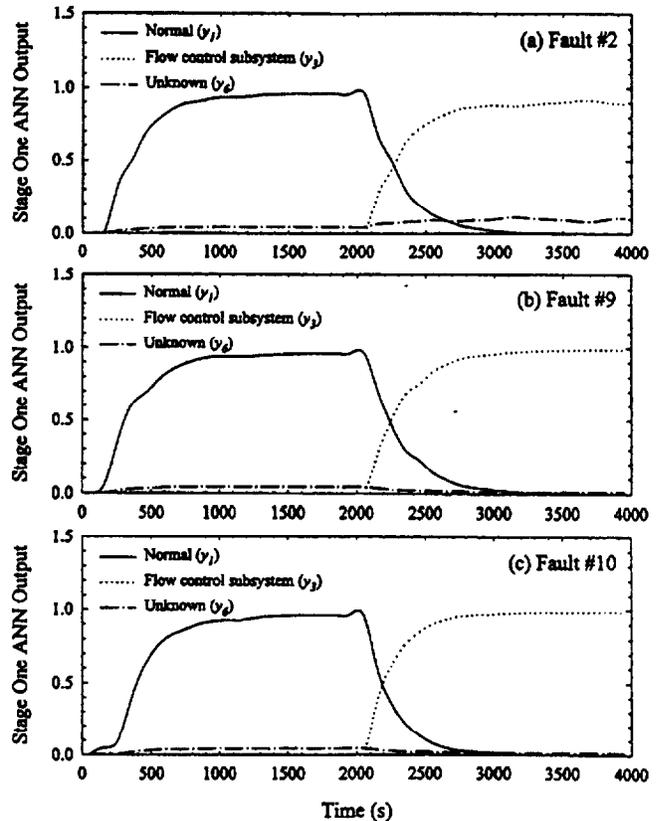


Figure 4 Fault diagnosis for flow control subsystem faults (faults 2, 9, and 10).

correct diagnosis is also made for the mixing box subsystem fault. The faults in the flow control subsystem and the mixing box subsystem fault all have diagnoses with small degrees of belief that the operation is unknown.

Results of the diagnoses for faults in the cooling coil subsystem are shown in Figures 6 through 9. In each figure,

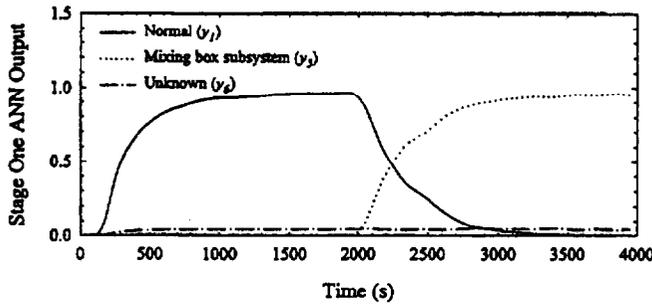


Figure 5 Fault diagnosis for a mixing box subsystem fault (fault 11).

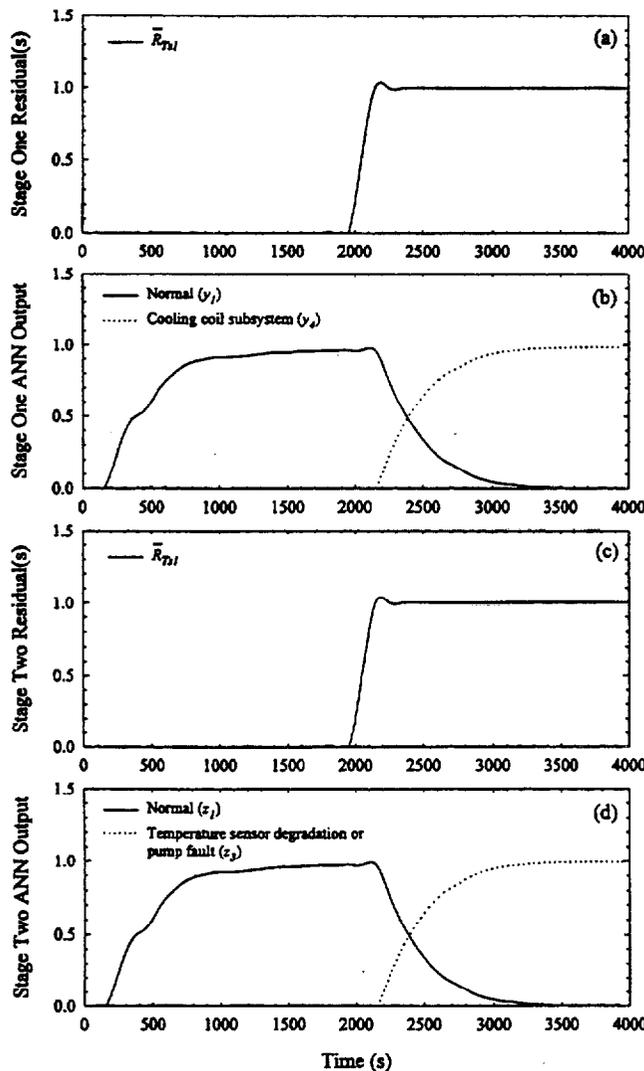


Figure 6 Fault diagnosis for a pump fault (fault 3).

subplot a shows the dominant stage one residuals, subplot b shows the stage one ANN output, subplot c shows the dominant stage two residuals, and subplot d shows the stage two ANN output. Figure 6 shows the dominant residuals and the ANN diagnoses for a pump fault. In both stages, the dominant residual is $\bar{R}_{T,1}$. The stage one ANN output shows that the diagnosis gradually changes from normal operation to a fault in the cooling coil subsystem. The stage two ANN output shows an evolution of the diagnosis in the cooling coil subsystem from normal to a diagnosis of either a supply temperature sensor degradation fault or a pump fault. No further diagnosis is possible with the current two-stage ANN.

The diagnosis of a stuck cooling coil control valve fault (fault 4) is shown in Figure 7. The fault is introduced at $t = 1000$ seconds and a linear change in the system load is introduced beginning at $t = 2000$ seconds. If the valve is well controlled prior to the fault, it will be difficult to diagnose the fault until the load change or some other event occurs to upset the conditions in the system. This kind of behavior is seen in Figure 7, where the

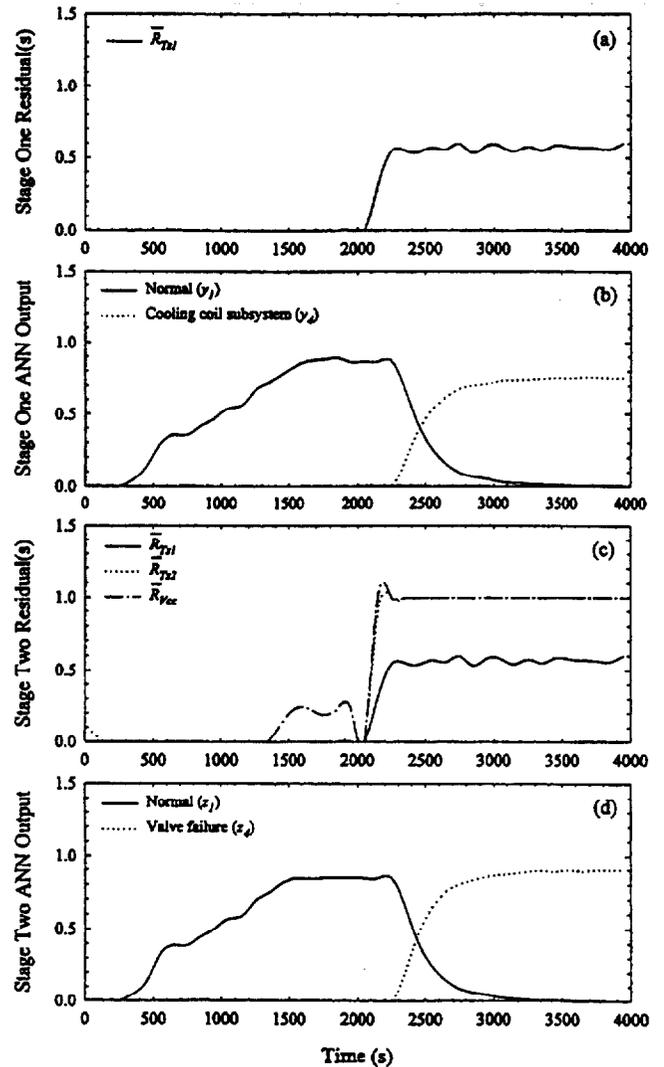


Figure 7 Fault diagnosis for a control valve fault (fault 4).

symptoms do not become evident until $t \approx 2100$ seconds. The ANN output for the initial 1500 seconds of the simulation is different from that seen for other faults. For this fault, the degree of belief that the system operation is normal increases slowly until it reaches a value near 0.9. This is due to the selection of load conditions that cause the control signal to the cooling coil valve to have a small degree of oscillation during the initial 1000 seconds. This, in turn, causes several of the residuals to have values between 0 and 0.3 during this period. A well-tuned control valve would eliminate these characteristics. Near $t = 2100$ seconds, \bar{R}_{Ts1} , \bar{R}_{Ts2} , and \bar{R}_{Vcc} become dominant and the stuck cooling coil control valve is diagnosed correctly.

Results for a complete supply air temperature sensor fault (fault 5) are shown in Figure 8. Results are not shown for fault 6 (5°C [9°F] offset in supply air temperature sensor reading) because they are almost identical to those for fault 5. In Figure 8, the dominant residual for stage one is \bar{R}_{Ts1} and the dominant residuals for stage two are \bar{R}_{Ts1} and \bar{R}_{Ts2} . Both residuals undergo a step change almost immediately after the fault occurs

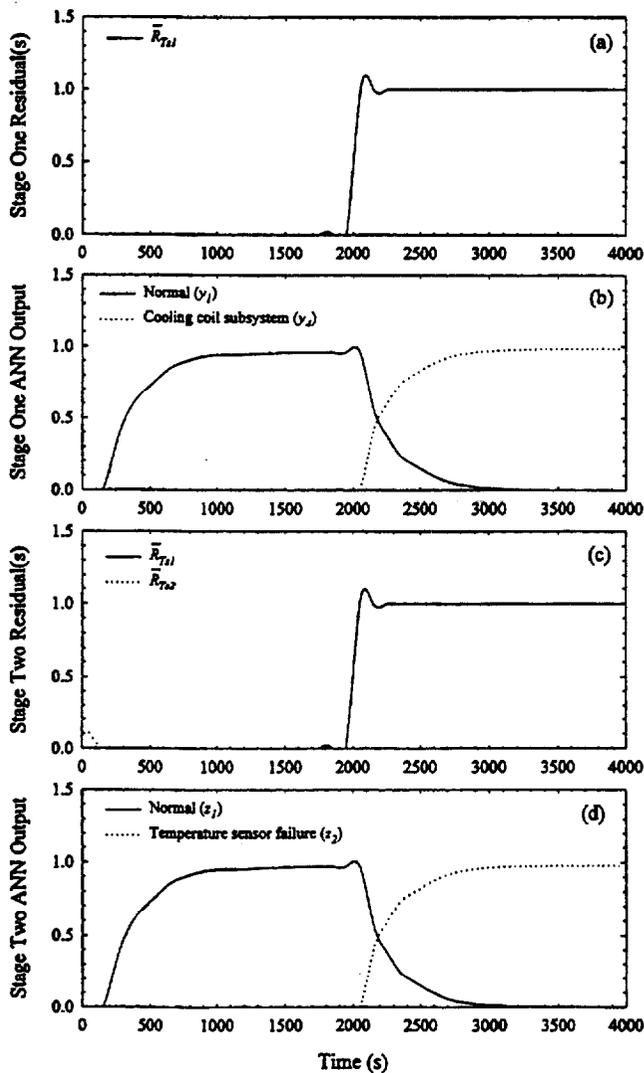


Figure 8 Fault diagnosis for a complete temperature sensor fault (fault 5).

because the sensor value is set equal to 0°C (32°F). The stage one and stage two ANN outputs correctly indicate the subsystem where the fault occurs and the type of fault, respectively.

Results for a supply air temperature sensor degradation fault (fault 7) are given in Figure 9. The dominant residual for both stages one and two is \bar{R}_{Ts1} . Note that although the fault is introduced at $t = 2000$ seconds by subtracting 1.5°C [$[(t - 2000\text{ s})/2000\text{ s}]$] from the actual sensor value, the residual remains equal to zero until $t \approx 2800$ seconds due to the deadband region between $R_{Ts1} = 0$ and $R_{Ts1} = R_{min}$. At $t \approx 3000$ seconds, the stage one ANN output for normal operation begins to decrease and the outputs for a cooling coil subsystem fault and unknown operation begin to increase. The degree of belief that the fault is in the cooling coil subsystem continues to increase to a final value that is slightly greater than 0.5. The stage two ANN diagnosis is similar to that in stage one. As t approaches 4000 seconds, the degree of belief that the fault is either a supply temperature sensor degradation fault or a pump fault increases to a value slightly greater

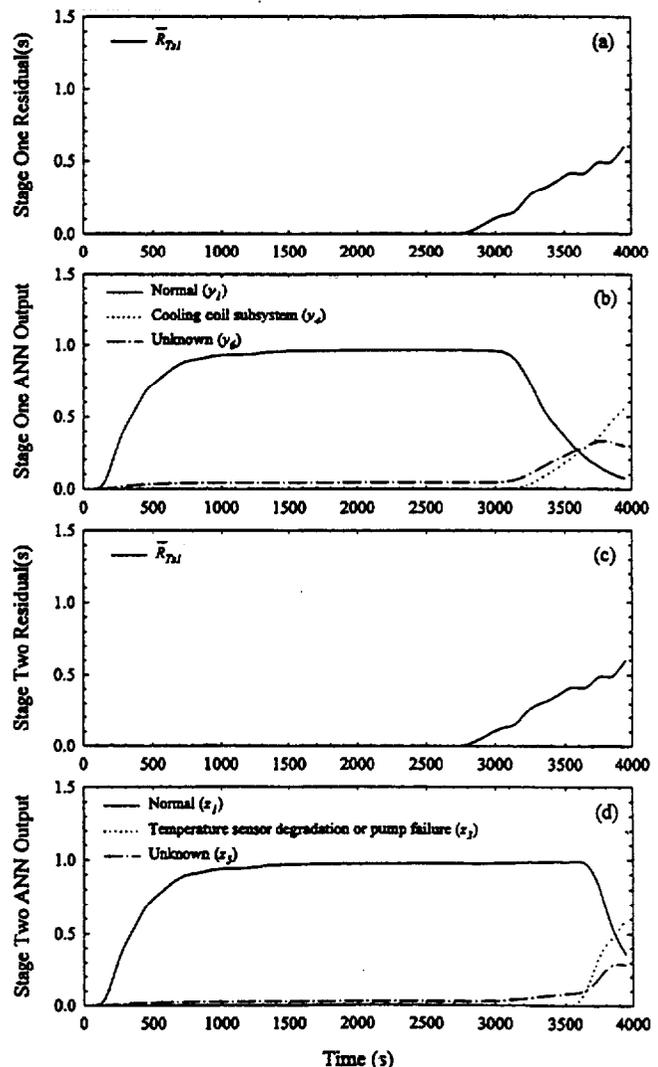


Figure 9 Fault diagnosis for a temperature sensor degradation fault (fault 7).

than 0.5. As in the pump fault case, no further diagnosis is possible with the current two-stage ANN.

The results presented in this section demonstrate the capability of the two-stage ANN, trained with patterns adapted from idealized fault/symptom relationships, to correctly diagnose the faulty subsystem for 11 faults in an AHU and to further diagnose the faulty component for the 5 faults that occurred in the cooling coil subsystem. The study by Lee et al. (1996b) demonstrated the capability of ANNs to generalize from idealized input data to noisy lab data. The two-stage approach simplifies the generalization by replacing a single ANN that encompasses all considered faults with a number of less complex ANNs, each one dealing with a subset of the residuals and symptoms associated with a complete diagnosis of all faults. As more faults are considered, more stages could be added. In addition, this kind of architecture would make it possible to limit retraining to only select ANNs. Retraining should also require fewer computational resources because the ANNs would not be as complex as in the case of a single ANN used for all faults. Finally, this approach is beneficial from a standpoint of understanding the diagnosis and could be utilized by a building operator to step through the reasoning behind the fault diagnosis.

Sensor Recovery

Sensor recovery results for two temperature sensor faults (faults 5 and 6) are shown in Figures 10 and 11. For fault 5, the supply air temperature sensor experiences a complete failure. Real (simulated: T_s) and estimated values ($T_{s,est}$) of the supply air temperature and the supply air temperature sensor signal are

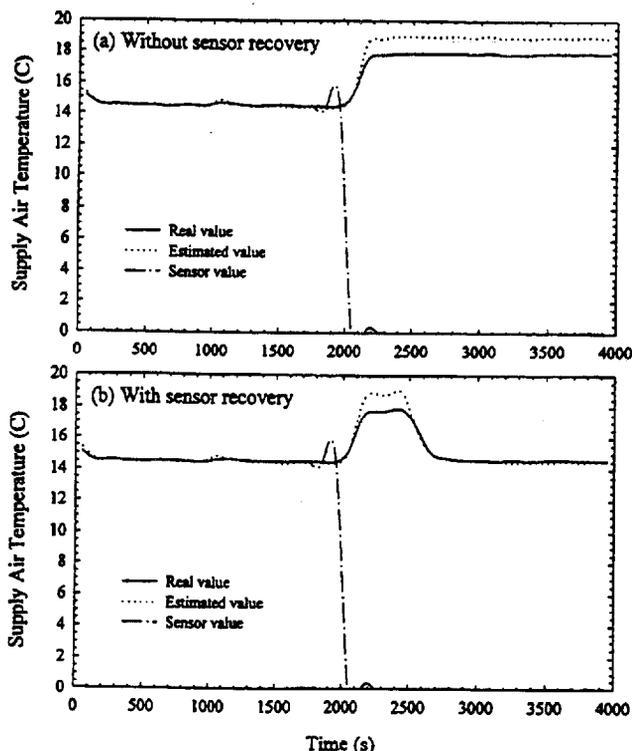


Figure 10 Sensor recovery for supply air temperature sensor: complete failure (fault 5).

plotted as functions of time in Figure 10a for the case where the temperature sensor is not recovered. When the sensor fails at $t = 2000$ seconds, the sensed value of T_s becomes 0°C (32°F) and the cooling coil valve closes in an attempt to make the sensed supply air temperature increase to the setpoint value of 14.5°C (58.1°F). This action causes the actual supply air temperature to increase to approximately 18°C (64.4°F), while the estimated value increases to approximately 19°C (66.2°F). Figure 10b shows that the failed sensor can be recovered within a short time after the fault is detected. This is accomplished by switching the control of the cooling coil valve from the sensor output to the estimated value of the supply air temperature after the fault is detected. That is, using $T_{s,est}$ obtained from a regression equation as the feedback signal to the cooling coil valve controller, the real value of the supply air temperature (T_s) can be recovered to a value near the setpoint. In these results, it is assumed that the fault is diagnosed at $t = 2500$ seconds.

Fault 6 is the case in which the supply air temperature sensor drops from its supporting harness to the floor of the air duct. Simulation results for this fault are shown in Figure 11a. Because the air duct surface is assumed to be at a higher temperature than the air flowing through the duct, the cooling coil valve controller responds as if a large disturbance has upset the process. In particular, the controller attempts to compensate for the fault by opening the control valve. This causes the real and estimated values of the supply air temperature to decrease; however, the wall temperature is assumed to be constant and the sensor tempera-

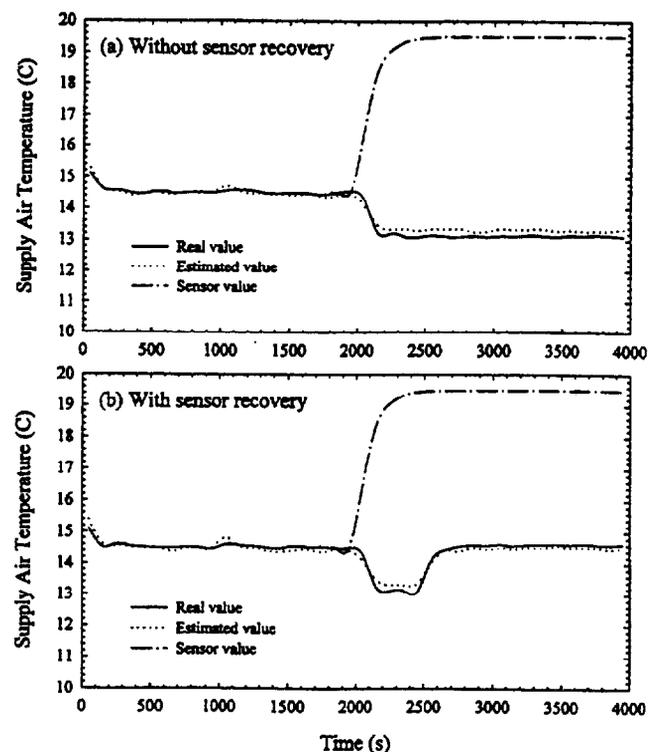


Figure 11 Sensor recovery for supply air temperature sensor: 5°C (9°F) offset due to contact with duct (fault 6).

ture is equal to that of the wall. Figure 11b shows that the failed sensor can be recovered using the estimated value of the supply air temperature.

CONCLUSIONS

The objectives of this paper were to describe an architecture for a two-stage ANN for fault diagnosis and to describe the use of regression equations for sensor recovery of failed temperature sensors. The stage one ANN was trained to classify the subsystems in which faults are occurring, and the stage two ANN was trained to diagnose the cause of faults at the subsystem level. The architecture can be extended in a straightforward manner to consider additional faults such as the faults in the VAV boxes, which can be accommodated with an additional stage two ANN. It is likely that this will require the introduction of additional residuals to the analysis.

To train the ANNs, residuals of system variables were selected that could be used to quantify the dominant symptoms of fault modes of operation. Idealized steady-state patterns of these residuals were then defined for each mode of operation studied, and patterns adapted from the idealized patterns were subsequently used for training. The trained ANNs were applied to simulation data for various faults and successfully identified each fault or the subsystem of the fault.

A regression equation was used to recover an estimate for the supply air temperature when the supply air temperature sensor yields erroneous measurements. Although the agreement between the actual and predicted temperature signals during faulty operation was not perfect, the regression model was adequate for identifying fault modes of operation. It was shown that the estimates of the sensor measurement can be used for control purposes.

Future work related to this study will include implementing a method for fault detection that can classify the system operation as either normal or faulty. The residuals that are used as inputs to the ANNs could also be used for fault detection by identifying allowable tolerances for each residual. If the tolerances are exceeded, a fault would be detected. Additional future work includes the implementation of the two-stage ANN method in real buildings to establish its capabilities, strengths, and weaknesses.

ACKNOWLEDGMENTS

This research was partially supported by the Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy.

REFERENCES

- Braun, J.E., S.A. Klein, and J.W. Mitchell. 1989. Effectiveness models for cooling towers and cooling coils. *ASHRAE Transactions* 95(1): 164-174.
- Demuth, H., and M. Beale. 1992. *Neural network toolbox: User's guide*. Natick, Mass.: The MathWorks, Inc.
- Fasolo, P., and D.E. Seborg. 1994. An SQC approach to monitoring and fault detection in HVAC control systems. *Proceedings of the American Control Conference*, pp. 3055-3059. New York: Institute of Electrical and Electronics Engineers.
- Incropera, F.P., and D.P. DeWitt. 1985. *Fundamentals of heat and mass transfer*, 2d ed. New York: Wiley and Sons.
- Lee, W.Y., C. Park, and G.E. Kelly. 1996a. Fault detection of an air-handling unit using residual and recursive parameter identification methods. *ASHRAE Transactions* 102(1): 528-539.
- Lee, W.Y., J.M. House, C. Park, and G.E. Kelly. 1996b. Fault diagnosis of an air-handling unit using artificial neural networks. *ASHRAE Transactions* 102(1): 540-549.
- Merchawi, N.S., and S.R.T. Kumara. 1994. *Artificial neural networks for intelligent manufacturing*, chap. 16, C.H. Dagli, ed. London: Chapman & Hall.
- Wang, S. 1992. Modeling and simulation of building and HVAC system—Building and HVAC system and component models used in emulation exercise C.3. IEA ANNEX 17 working paper.
- Watanabe, K., S. Hirota, L. Hou, and D.M. Himmelblau. 1994. Diagnosis of multiple simultaneous faults via hierarchical artificial neural networks. *AIChE Journal* 40(5): 839-848.