HVAC&R Research

ISSN: 1078-9669 (Print) 1938-5587 (Online) Journal homepage: http://www.tandfonline.com/loi/uhvc20

Detection of Rooftop Cooling Unit Faults Based on Electrical Measurements

Peter R. Armstrong , Christopher R. Laughman , Steven B. Leeb & Leslie K. Norford

To cite this article: Peter R. Armstrong , Christopher R. Laughman , Steven B. Leeb & Leslie K. Norford (2006) Detection of Rooftop Cooling Unit Faults Based on Electrical Measurements, HVAC&R Research, 12:1, 151-175

To link to this article: http://dx.doi.org/10.1080/10789669.2006.10391172

Published online: 28 Feb 2011.

Detection of Rooftop Cooling Unit Faults Based on Electrical Measurements

Peter R. Armstrong, PhD **Christopher R. Laughman**
Student Member ASHRAE

Steven B. Leeb, PhD **Leslie K. Norford, PhD**
Member ASHRAE

Received January 17, 2005; accepted August 1, 2005

Nonintrusive load monitoring (NILM) is accomplished by sampling voltage and current at high rates and reducing the resulting start transients or harmonic contents to concise "signatures." Changes in these signatures can be used to detect, and in many cases directly diagnose, equipment and component faults associated with rooftop cooling units. Use of the NILM for fault detection and diagnosis (FDD) is important because (1) it complements other FDD schemes that are based on thermo-fluid sensors and analyses and (2) it is minimally intrusive (one measuring point in the relatively protected confines of the control panel) and therefore inherently reliable. This paper describes changes in the power signatures of fans and compressors that were found, experimentally and theoretically, to be useful for fault detection.

INTRODUCTION

Hardware degradation and control failures are the source of a large fraction of US HVAC operational and maintenance costs (Li and Braun 2004). Comfort and consequent productivity loss may represent significant additional costs. Inefficient performance from lack of maintenance or improper maintenance is ubiquitous in unitary equipment (Breuker and Braun 1998a). Many rooftop unit (RTU) faults go undetected by occupants and maintenance staff. A "wait for complaints" strategy is often taken in spite of its many well-known negative side effects. There are two generally-accepted paths to correcting this situation:

1. Test-and-measurement-based preventive maintenance (including such activities as recommissioning or continuous commissioning).
2. Automated online fault detection.

The key to cost-effective online fault detection and diagnosis (FDD) is finding the optimal mix of sensors and automated analysis for a given target system. Much attention has focused on temperature-measurement-based fault detection because of low sensor cost (Rossi and Braun 1997). However, temperature sensor location is critical; sensors must be installed in specific locations regardless of the resulting susceptibility to harsh environments or damage during inspection and service activities. Consequently, more expensive but less intrusive sensor types, such as pressure and power, may be justified by improved reliability.

Peter R. Armstrong is a senior research engineer at Pacific Northwest National Laboratory, Richland, Wash. **Christopher R. Laughman** is a graduate research assistant, **Steven B. Leeb** is a professor of electrical engineering, and **Leslie K. Norford** is a professor of building technology at Massachusetts Institute of Technology, Cambridge, Mass.

A device called the nonintrusive load monitor (NILM) measures and analyzes electrical signals that may be useful in fault detection (Leeb 1992; Shaw et al. 1998; Laughman et al. 2003). Results of recent tests, reported here, show that some RTU thermo-fluid faults can be detected by the NILM alone. Certain mechanical (compressor and fan) faults and most electrical (supply, controls, motor) faults can also be detected. This paper reviews previous research and characterizes the responses of steady and transient electrical loads to faults artificially introduced in an operating unit. New methods of detecting bypass leakage, fan imbalance, and liquid ingestion faults are described, and fault detection implementation issues are discussed. We suggest that the modest incremental cost of sampling and online analysis of transient, as opposed to only steady-state, data is well justified in this FDD application.

PREVIOUS RESEARCH

The frequency of faults in unitary cooling equipment has been assessed from substantial insurance claim and service record data bases. Stouppe and Lau (1989) compiled 15,716 failure records and attributed 76% of faults to electrical components, 19% to mechanical, and 5% to refrigerant circuit components. Of the electrical failures, 87% were motor winding failures from deterioration of insulation (vibration, overheating), unbalanced operation, and short cycling (overheating). The reported numbers refer to events; costs of repair and the relative proportion of failures attributable to package versus built-up equipment were not published. Breuker and Braun (1998a) analyzed "a statistically representative subset of the data" drawn from "over 6,000 separate fault cases" to estimate repair costs associated with different faults in packaged equipment.

Rossi and Braun (1997) developed a statistical, rule-based method for detecting and diagnosing faults in vapor-compression equipment. Breuker and Braun (1998a, 1998b) evaluated the detection sensitivity of the method at specific fault levels. Two prototype systems, designated the low-cost system (two temperature inputs, five outputs) and the high-performance system (three inputs, seven outputs), were postulated, and fault detection (without excessive false alarm rates) was generally found to be effective when capacity and COP dropped by 10% to 20% for five common mechanical faults:

- Undercharge (refrigerant leakage) or overcharge
- Liquid line restriction
- Loss of volumetric efficiency (leaky valves, seals)
- Fouled condenser coil
- Dirty supply air filter

FAULT TESTING AND ANALYSIS

Of the five faults examined by Rossi and Braun (1997) and Breuker and Braun (1998a, 1998b), NILM detection was attempted for all but the liquid line restriction fault. Based on Breuker and Braun's (1998a) review of repair costs and fault frequencies, two new mechanical faults—rotor imbalance and flooded start/liquid ingestion—were added to the test suite. Purely electrical faults were not tested. Detection of an unbalanced three-phase voltage condition (failure of any phase to maintain a preset threshold voltage for a preset number of cycles while the other two hold up) is completely straightforward. Results of contact erosion and bounce testing, currently underway, will be reported in the future.

Observation of start transients under fault conditions requires a special procedure for each type of fault. Baseline start transients were collected by starting each motor, letting it run for

five to ten seconds, and waiting for it to stop spinning after shutdown. This process was repeated five times while the NILM wrote the data to a single file.

Evaporator and condenser blockage faults were artificially introduced by inserting strips of cardboard over the heat exchanger inlet face. This achieved a quasi-uniform flow distribution of known blockage factor. Refrigerant over- and undercharge conditions were achieved by evacuating the system, then adding a known mass of refrigerant to obtain the desired level of charge. Leaky valves or seals were mimicked by opening the hot bypass valve a specific amount and making repeated compressor starts. Fan imbalance was artificially introduced by adding small weights of known mass near the rotor tip at one point on the circumference. For the supply fan, binder clips were fastened to a vane inside the squirrel cage. For the condenser, a known mass was affixed at the tip of one blade.

To detect small changes in an electrical load signal, it was essential to suppress the 120 Hz component and consider only the "dc" component of power (P) and reactive power (Q). This was accomplished by a special digital filter implemented in the NILM software (Leeb 1992; Shaw and Laughman 2005). This work makes use of real and reactive power at the fundamental frequency, P1 and Q1; the magnitudes of harmonics P3, P5, P7, Q3, Q5, and Q7 were recorded but not used. In subsequent sections we show plots of the start trajectories of P1 or Q1 for each motor load. In some cases the apparent power, R1, and phase angle, A1, are also of interest. The repeatability of these transients is important to the FDD application. The trajectories are time-shifted and superimposed to highlight similarities or differences between repeated transient shapes. Additional start transients, recorded after fault introduction, are presented and the sensitivity to the change is characterized.

Short Cycling

Compressor start transients with no pressure relief between starts are shown in Figure 1. There is an increase in steady-state power, at the end of the transients, compared to the transients with pressure relief between starts shown in Figure 2. The fans were off in both of these tests. The compressor cycling frequency of about three starts per minute used during the test is much higher than the shortest cycling frequency of normal operation and is considered to be representative of a short-cycling fault.

Refrigerant Charge

Plant capacity is adversely affected by incorrect refrigerant charge. The system can be overcharged or undercharged. In either case, excessive portions of the evaporator or condenser are relegated to sensible, rather than boiling or condensing, modes of heat transfer. With undercharge, there is too much vapor being superheated in the evaporator and desuperheated in the condenser. With overcharge, too much liquid collects in the condenser. Increased pressure ratio results in both cases and, thus, increased compressor power is a symptom of both undercharge and overcharge.

The increase in compressor power is illustrated in Figure 3 for overcharge. With undercharge, however, power does not quickly settle to a steady value. Two start transients representative of undercharge are compared to the normal start transients in Figure 4.

Detection can be based on P, Q, or one of the derived parameters, apparent power, $R = (P^2 + Q^2)^{1/2}$, or phase angle, $A = \tan^{-1}(P,Q)$. Based on the sample mean and variance for each parameter (P1, Q1, R1, A1) and each condition (deviations of −20%, 0%, and +20% from normal charge) shown in Table 1, two of the parameters, power (P1) and phase (A1), are useful for detecting refrigerant-charge faults. The estimated population distributions for both, assumed to be Gaussian, are plotted in Figures 5 and 6.

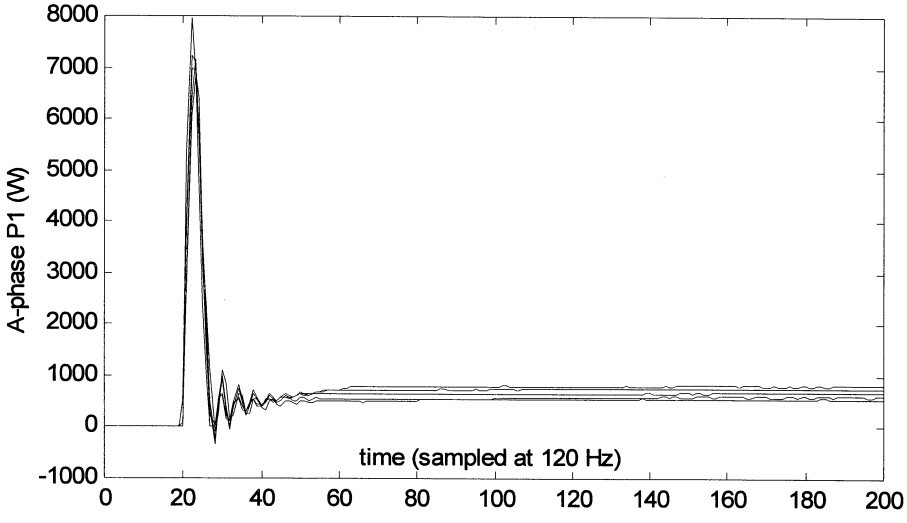

Figure 1. Five compressor starts with no pressure relief. The main difference between start traces is that the final value of P1 (time index >50) increases with successive starts.

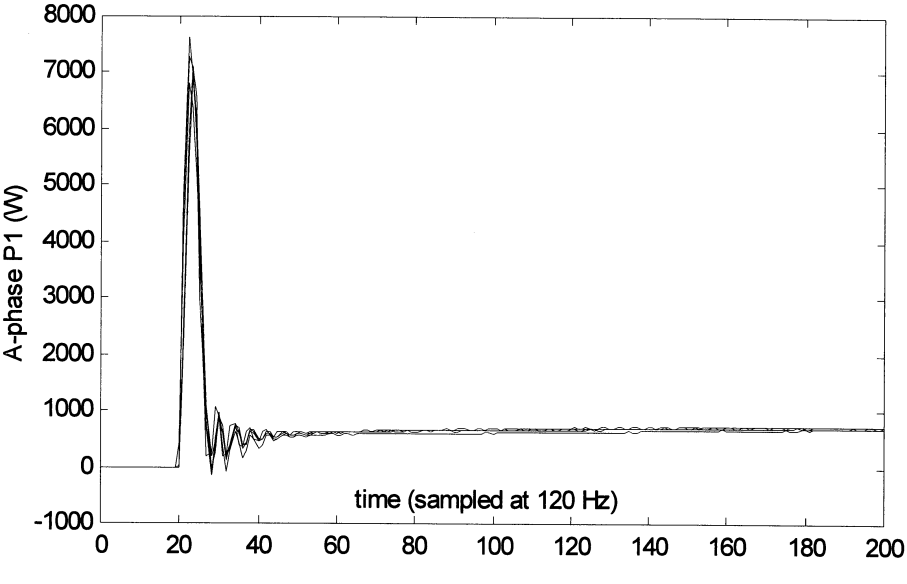

Figure 2. Five compressor starts with pressure relief (zero initial head).

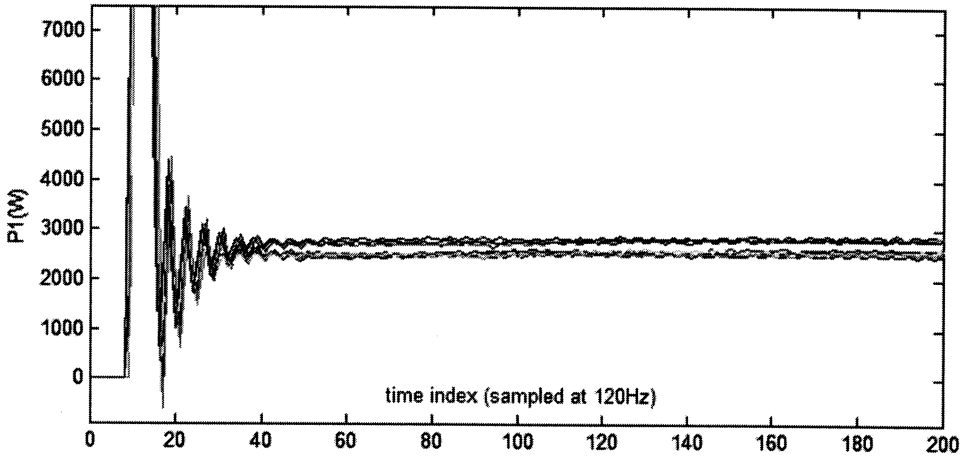

Figure 3. Compressor start transients: lower traces normal, upper with 20% overcharge.

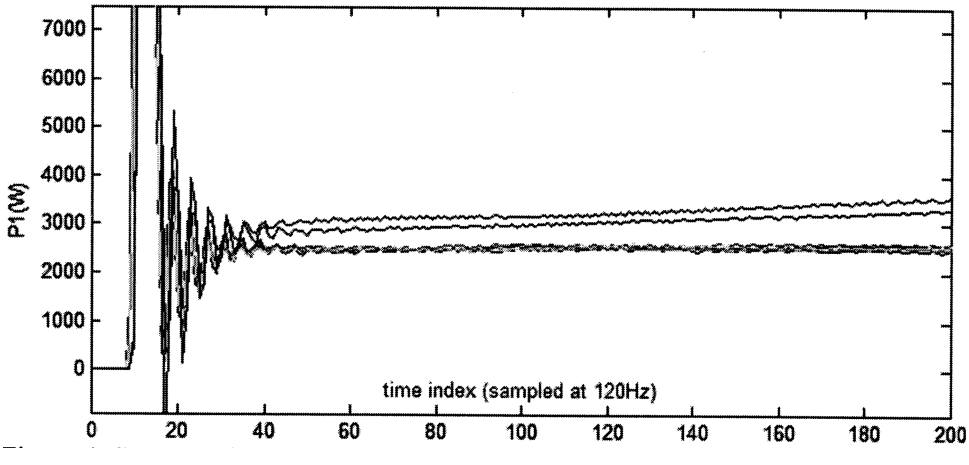

Figure 4. Start transients: lower seven traces normal, upper two traces 20% undercharge. For clarity, only two of the four undercharge traces are shown.

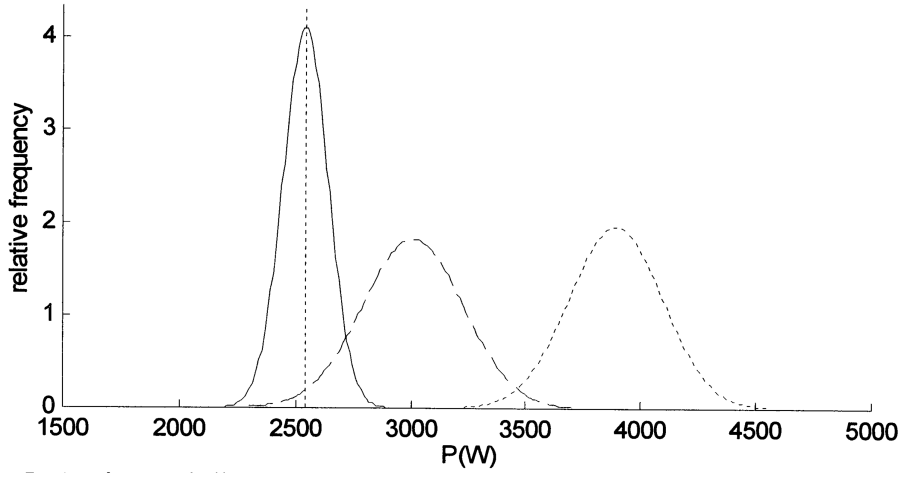

Figure 5. Estimated distributions of steady-state power. Because only a small portion of the tail of the 20% overcharge envelope (centered at 3009 W) crosses the no-blockage mean (2542 W), detection at 20% overcharge will be almost immediate (~10 starts) once the normal-charge mean has been well established.

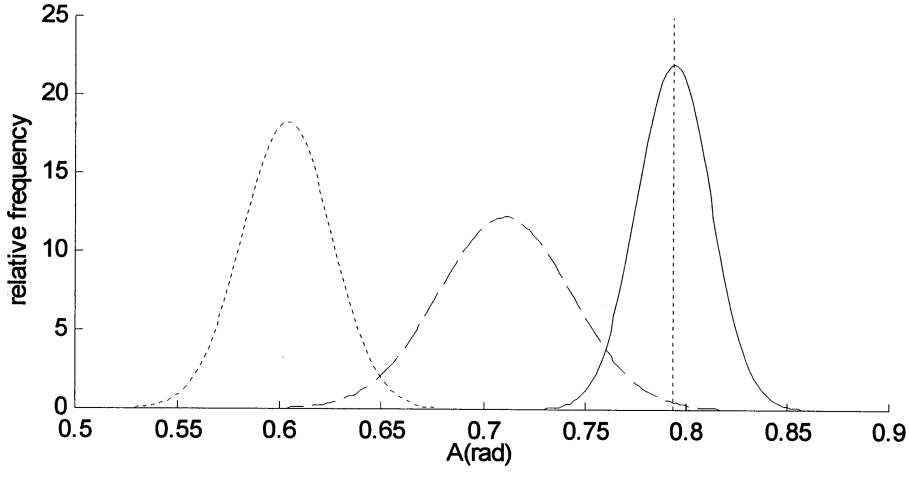

Figure 6. Estimated distributions of steady-state phase. Because only a small portion of the tail of the 20% overcharge envelope (centered at 0.711 radians) crosses the no-blockage mean (0.794 radians), detection at 20% overcharge will be almost immediate (~10 starts) once the normal-charge mean has been well established.

Table 1. Detection Sensitivity for Refrigerant Undercharge (–20%) and Overcharge (+20%)

Deviation from Normal Charge		–20%	0%	+20%
Number of Observations, *N*		4	7	7
Real power, P1, W	Sample average	3891	2542	3009
	Standard deviation	203	97.5	219
Reactive power, Q1, VAR	Sample average	2675	2585	2583
	Standard deviation	24	16.3	22.8
Apparent power, R1, VA	Sample average	4723	3626	3967
	Standard deviation	177	74.1	177
Phase angle, A1, rad	Sample average	0.6033	0.7941	0.7107
	Standard deviation	0.0219	0.0181	0.0327

Sample size is an important factor in detection. One advantage of online fault detection is that the parameters for the no-fault condition can be established with relatively high confidence because in a new RTU there will usually be a large number of observations, N, before a fault occurs and uncertainty is proportional to $N^{-1/2}$. The vertical dashed line in each plot indicates the sample mean for the no-fault condition. If this mean value were confirmed after many fault-free observations, the probability that the new mean indicated for a fault condition (based on a given number of post-fault observations) truly differs from the fault-free mean is given by the area under the distribution minus the area of the tail on the other side of the vertical line. The NILM can be programmed to implement standard tests (Wild and Seber 2000) for change of sample mean given actual sample sizes for the fault-free and post-fault observations that have accumulated at any given time.

The tests for improper charge did not account for the variations of steady-state compressor power that occur with operating conditions. The observed changes in compressor power alone cannot, therefore, be attributed to a particular fault (overcharge versus undercharge) or distinguished from normal variations in capacity. Discharge-suction pressure difference or condensing and evaporating temperatures could be used to normalize operating conditions.[1]

Compressor Back Leakage

Start transients were recorded (four repetitions) for the compressor/condenser fan with hot bypass and two more start transients were recorded with the bypass closed, as shown in Figure 7. The transients with no bypass are the top two. The amplitude difference when the motor is developing peak torque (0.2-0.5 s from initial contact) is much greater than the steady-state (time index >200) amplitude difference, and, moreover, the shapes of the faulty and fault-free start transients are distinctly different. The reactive power, not shown, reveals no such distinct change in response to the fault.

The mechanism for a behavior that is distinctive during the start transient, yet results in little or no change in steady-state compressor power, is not obvious. However, it turns out that a simple model, in which evaporator and condenser vapor masses are the only state variables, explains the observed behavior (Armstrong 2004). We assume that evaporator and condenser vapor volumes are constant, that the leak rate follows a power law, and that the compressor flow rate is smooth, not pulsing, with zero clearance volume. Under these assumptions, the rate of

[1] Further testing, completed after review of the manuscript, has shown that both overcharge and undercharge can be detected by start transient shape if the signal is normalized to the minimum instantaneous load observed after the inrush phase. With this normalization, post-inrush undercharge power is consistently less than that oberved for the overcharge fault.

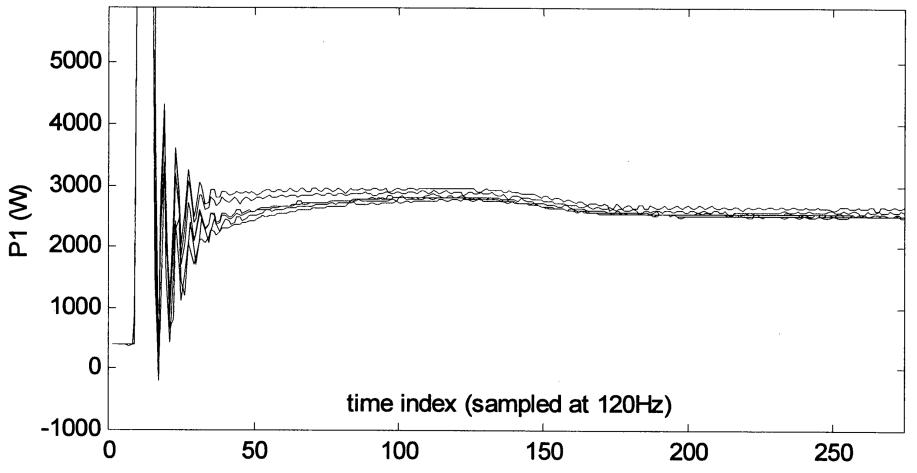

Figure 7. Compressor/condenser active power with bypass leakage (bottom four traces) and without (top two traces).

change in evaporator vapor mass M_e is the rate of evaporation plus the back leakage minus the compressor suction rate, given approximately by the following mass balance equation:

$$\frac{dM_e}{dt} = \frac{\Delta T_e u_e}{i_{fg}(T_{sat}(P_e))} + \sqrt{\rho_e \rho_c} C (P_c - P_e)^n - \rho_e V \frac{\omega}{2\pi}, \quad (1)$$

where

V = $V_{actual}\eta_{displacement}$ = effective compressor displacement
ω = compressor shaft speed
ρ_e = M_e/V_e is the evaporator (suction) vapor density
ρ_c = M_c/V_c is the condenser (discharge) vapor density
P_e = $P_{sat}(\rho_e)$, $T_e = T_{sat}(P_e)$ = evaporator pressure and temperature
P_c = $P_{sat}(\rho_c)$, $T_c = T_{sat}(P_c)$ = condenser pressure and temperature
ΔT_e = $T_{ra} - T_e$ = return air minus refrigerant temperature difference at the evaporator
u_e = effective evaporator conductance-area product
C,n = leak parameters with leakage flow modeled as a power law
$i_{fg}(T)$ = heat of vaporization as a function of temperature
$P_{sat}(\rho)$, $T_{sat}(P)$ = the other saturation properties

The rate of change in condenser vapor mass, M_c, equals the suction rate minus the condensation rate minus the leakage rate,

$$\frac{dM_c}{dt} = \rho_e V \frac{\omega}{2\pi} + \frac{\Delta T_c u_c}{i_{fg}(T_{sat}(P_c))} + \sqrt{\rho_e \rho_c} C (P_c - P_e)^n, \quad (2)$$

where

$\Delta T_c = T_c - T_{amb}$ = refrigerant minus ambient air temperature difference at the condenser and
u_c = effective condenser conductance-area product.

Compressor power (averaged over one shaft revolution) is given by the approximate energy balance:[2]

Simulated power transients with two different leak levels are shown in Figure 8.[3] With a leak fault, condenser pressure builds more slowly until the vapor reaches the saturated condition. Evaporator vapor is already saturated, so pressure is initially almost constant; thus, initial suction density and flow rate are the same for both cases. With similar mass flow rates, less head translates to lower power initially. At steady state, suction flow rate is higher for the faulty case due to higher evaporator pressure. Despite loss of capacity, with higher flow rate and lower head there is little change in steady-state compressor input power.

Flow Blockage

On the air side, reduced flow (blockage or restriction) is one of the main faults of interest. Condenser fan start transients were measured at three levels of blockage: 0%, 14%, and 39% of coil face area. The compressor was locked out, and six to eight repetitions were made at each fault level.[4] Figure 9 shows the mean P1 transients for each fault level. Figure 10 gives the mean static pressure transient for each fault level; distinct transitions in fan flow regime, clearly evident in the static pressure trajectory, are weakly echoed in P1. Only one of the parameters, P1, is useful for detecting blockage; variances for the others are large relative to their change in mean with fault level.The estimated population distributions for P1, assumed to be Gaussian, are plotted in Figure 11.

Condenser fan power is a good fault detection criterion because it is normally quite constant. Use of temperature-rise-based detection, on the other hand, is problematic for the condenser fan blockage fault. Both inlet and outlet sensors are exposed to the elements.The inlet (ambient) temperature sensor in particular is susceptible to radiation, icing, and wet-bulb bias errors.

Supply fan start transients were measured with four levels of blockage: 0%, 10%, 50%, and 100% of coil face area. Six to eight repetitions were made at each fault level. Figures 12 and 13 show the mean P1 and Q1 transient responses for no fault and each of three fault levels.

Two of the parameters, R1 and A1, are not useful for detecting supply-airflow blockage because their variances are large relative to the change in mean. The other two parameters, P1 and Q1, are useful and their estimated population distributions are plotted in Figures 14 and 15.

In contrast to the condenser fan response, supply fan power and reactive power both *decrease* with increasing blockage. This is a consequence of differences in fan curves and the points on those curves that correspond to the respective no-blockage conditions. Moreover, in the case of the supply fan, detection of a blockage is best accomplished by testing the hypothesis that *both* the reactive and real components of steady-state electrical load have deviated from the established no-fault mean values.

From the test results it appears that a 10% blockage can be readily detected. However, variations in system resistance resulting from air density changes (Armstrong 1983) upstream or

[2] Energy dissipated by the leak is assumed to add to discharge superheat, which is recognized but not explicitly modeled; the ideal gas work is given by

$$\dot{W} = \frac{\omega}{2\pi} V P_e \frac{k}{k-1}\left(\left(\frac{P_c}{P_e}\right)^{(k-1)/k} - 1\right) = \frac{\omega}{2\pi} V \rho_e c_p (T_c - T_e) .$$

[3] The final (t = 250/120 = 2.1 s) simulated power in both cases is much lower than the measured power because superheating, flow losses, and mechanical losses are not modeled.

[4] The compressor would not be locked out in practice, requiring the NILM to detect the fan in the presence of an operating compressor. The NILM's signal-processing algorithms are capable of decomposing start transient and spectral data based on a priori knowledge of connected loads (Leeb 1992; Laughman 2003). Since an RTU typically has 3 to 5 main connected loads (1 or 2 compressors, 1 or 2 condenser fans, and a supply fan), this task is relatively straightforward.

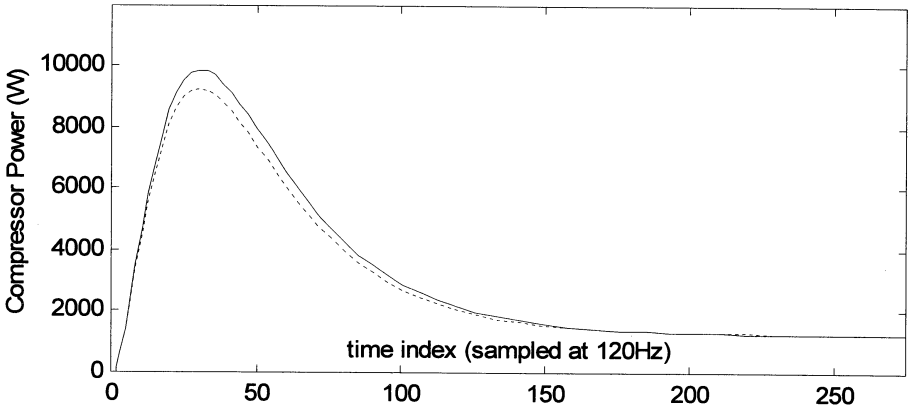

Figure 8. Simulated transient compressor power with leak fault (dashed) and no fault.

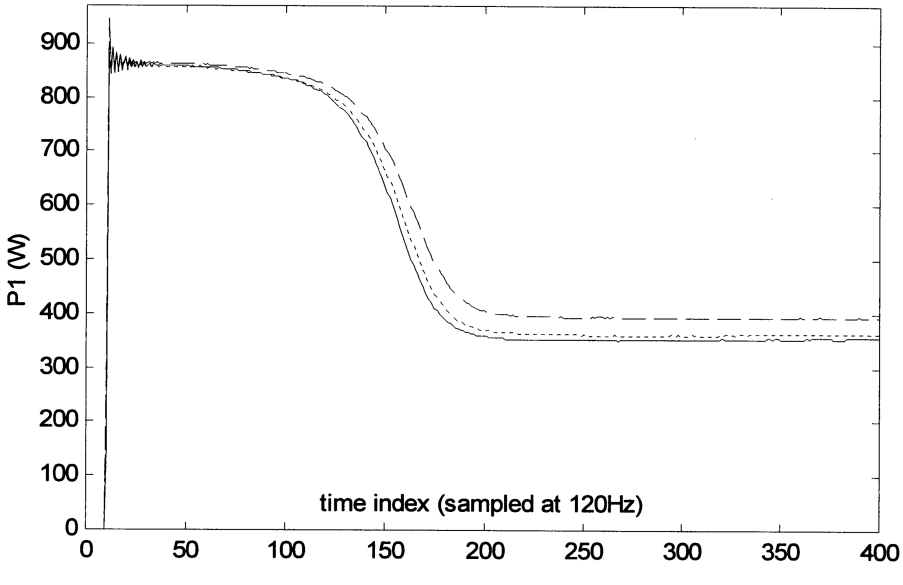

Figure 9. Condenser fan real power start transients, mean of repetitions for each fault level: 0% blockage (solid), 14% blockage (dotted), and 39% blockage (dashed).

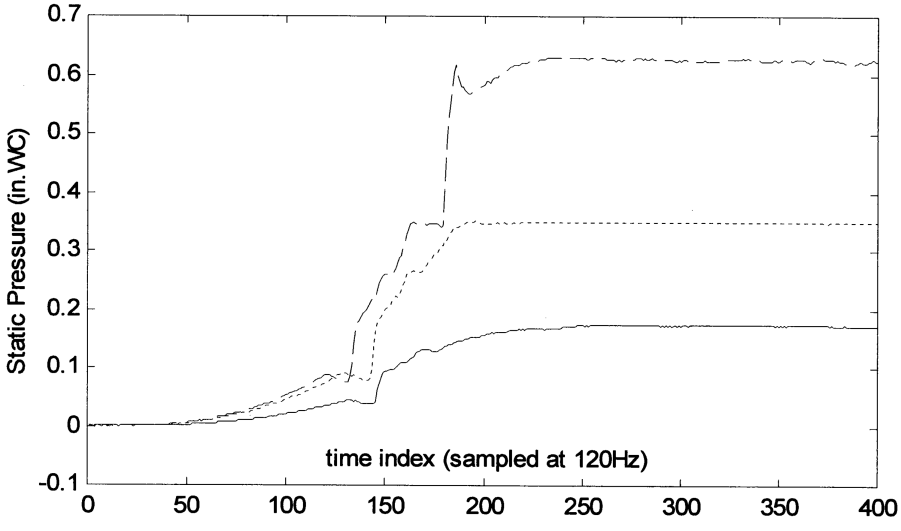

Figure 10. Condenser fan pressure start transients, mean of repetitions for each fault level: 0% blockage (solid), 14% blockage (dotted), and 39% blockage (dashed).

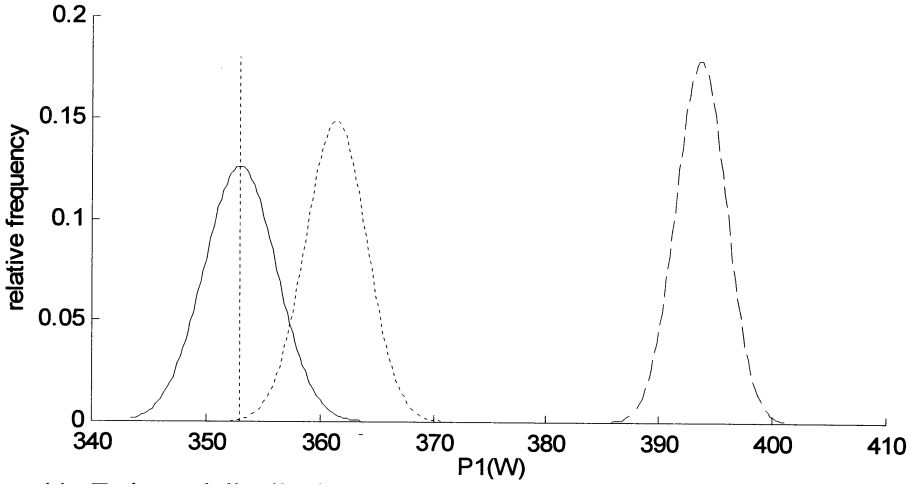

Figure 11. Estimated distributions of steady-state power. Because only a small portion of the tail of the 14% blockage envelope (centered at 361 W) crosses the no-blockage mean, 353 W, detection at 14% blockage will be almost immediate (~10 starts) once the no-blockage mean has been well established.

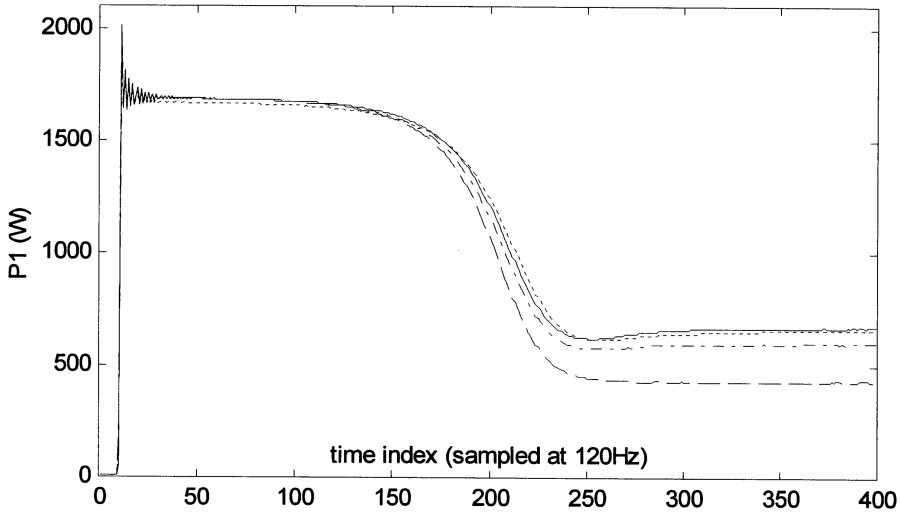

Figure 12. Supply fan real power start transients, mean of repetitions for each blockage level: 0% (solid), 10% (dotted), 50% (dot-dash), and 100% blockage (dashed trace).

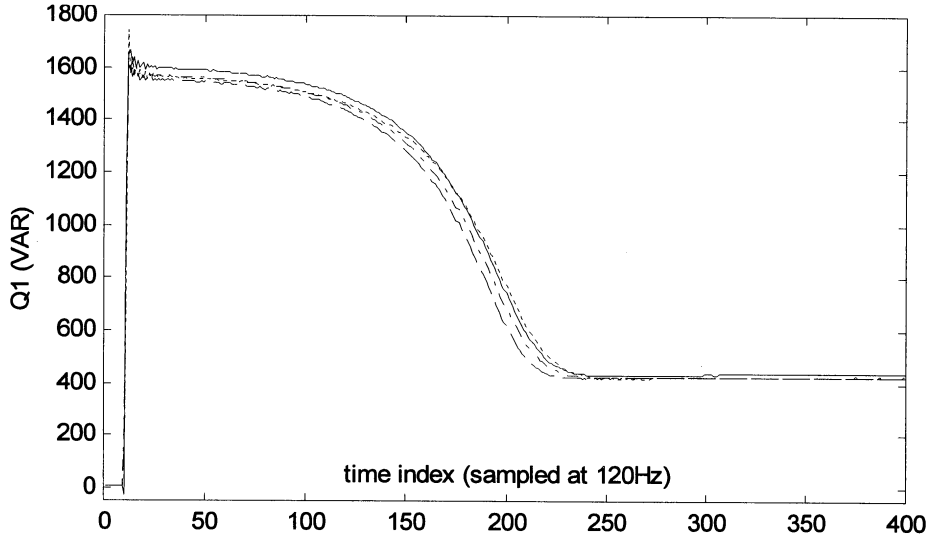

Figure 13. Supply fan reactive power start transients, mean of repetitions for each blockage level: 0% (solid), 10% (dotted), 50% (dot-dash), and 100% blockage (dashed trace).

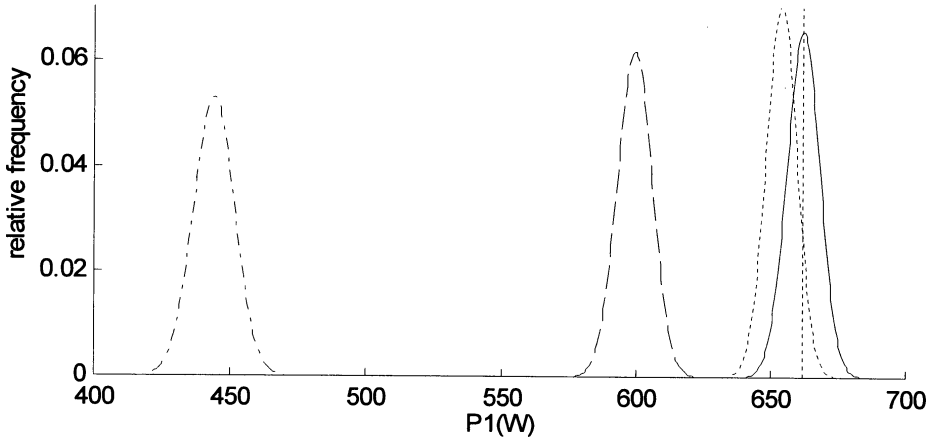

Figure 14. Estimated distributions of steady-state power. Because the 10% blockage envelope, centered at 654 W, overlaps the no-blockage mean (662 W), it would take several hundred start-transient observations before a 10% blockage could be reliably detected based on P alone.

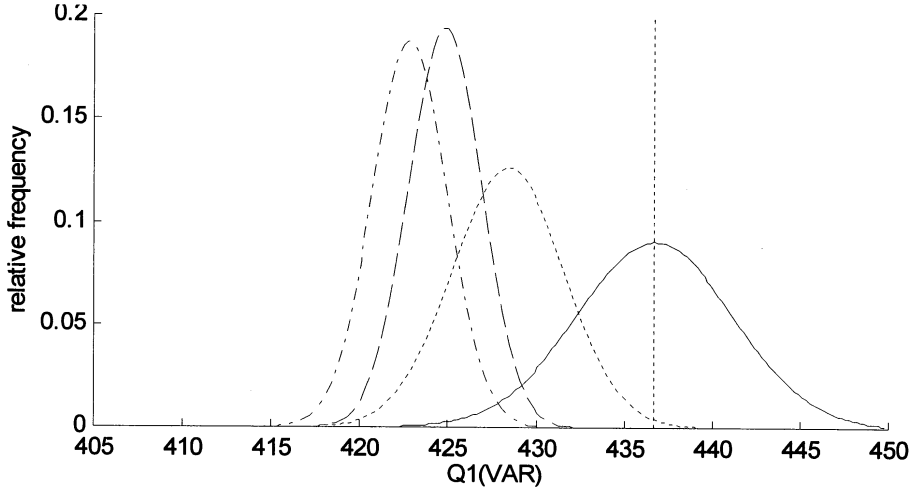

Figure 15. Estimated distributions of steady-state reactive power. Because only a small portion of the tail of the 10% blockage envelope (centered at 428 VAR) crosses the no-blockage mean, 437 VAR, detection at 10% blockage will be almost immediate (~10 starts) once the no-blockage mean has been well established.

downstream of the fan and changing of damper settings may result in substantial uncertainty in the no-fault means. Further work is needed to understand if the supply fan blockage fault is truly amenable to detection by NILM in the face of typical system disturbances.

Load Spectra with Artificially Introduced Imbalance Faults

The condenser fan and supply fan are susceptible to faults that result in impeller or rotor imbalance, vibrations, and associated periodic variations in motor load. Such load variations can be detected by the NILM. Supply fans are typically of the squirrel cage type with direct or belt drive. Imbalance can result from a bent impeller air foil, dirt accumulation, or from a foreign object, e.g., piece of filter material becoming lodged inside the impeller cage. Although not explored here, a bent shaft or damaged sheave will result in periodic motor load variations that can be easily detected, and it is likely that serious belt wear can also be detected. Condenser fans are typically of the direct-drive axial type with a three- or four-bladed impeller. A cracked blade or blade root is the most common serious fault. A slightly bent blade or shaft or the accretion of dirt will also cause imbalance and may, eventually, lead to mechanical failure. Early detection is important because the cost to repair collateral damage to a motor or condenser is much higher than the cost to replace a fan impeller.

Periodic load variations may be detected by Fourier decomposition of the power signal. To test this fault detection method, we set the NILM to record (P,Q) data for ten minutes of steady operation. The magnitudes and frequencies of peaks in the amplitude spectrum provide a useful description of shaft or impeller imbalance and associated vibrations. For the purpose of demonstrating basic detection feasibility, we subtracted the mean value (dc component) and collected very long time sequences so that the Fourier transform gives good estimates of the amplitude spectra.[5] Discrete changes in rotor imbalance were induced by attaching small masses (a few grams) to the fan impellers.

Condenser Fan. Shaft harmonics data for the condenser fan are shown in Figure 16. The peaks at 59.5 Hz (Figure 17) and 29.7 Hz coincide with the fundamental frequency and half-frequency of shaft speed (nominally 3550 rpm); their magnitudes change little with fault level. However, the interaction of unbalanced impeller rotation with one of the lower frequency modes in the RTU structure results in a peak at 18.4 to 18.7 Hz that is very sensitive to fault level, as shown in Figure 18. Note that the frequency of this amplitude spectrum peak drops moderately as imbalance is increased.

Supply Fan. Shaft harmonics data for the supply fan are shown in Figures 19 and 20. There are peaks, barely visible on the scale of Figure 19, at 59.6 Hz and 29.8 Hz coinciding with the fundamental frequency and second harmonic[6] of shaft speed (nominally 1770 rpm). As with the condenser fan, the interaction of unbalanced impeller rotation with one of the lower frequency modes in the RTU structure results in a peak that is very sensitive to fault level. Figure 20 shows the peak for fault-free operation at around 19 Hz; the frequency of the peak drops to below 18.9 Hz, and its magnitude increases many times as fault level is increased.

The results for both the supply and condenser fans show that the existence and relative level of imbalance faults can be obtained from signals measured by the standard NILM platform. Appropriate scaling factors for reporting actual fault levels can be determined for a given RTU design by exactly the type of simple test performed here.

[5] Use of long sequences may be costly, in practice; a moving average of P can be used for short sequences or an appropriate windowing or multi-taper technique may be useful on low-cost embedded NILM platforms.

[6] "...a component…twice the rotational speed is generally linked with rotor anisotropy" Genta (1998), p. 410.

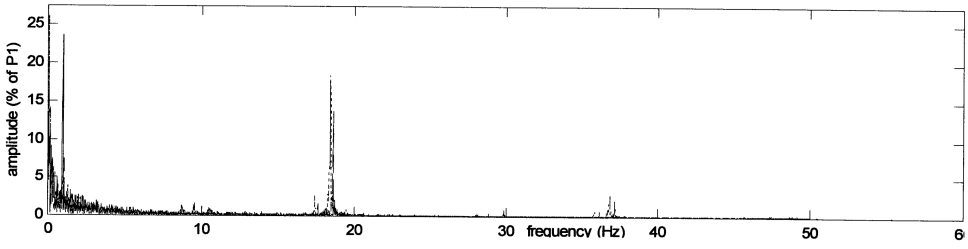

Figure 16. Amplitude spectrum (0-60 Hz, $N = 34260$) of condenser fan power (P1) with 0g (solid), 2.6g (dashed), and 5.3g (dotted) imbalance fault levels.

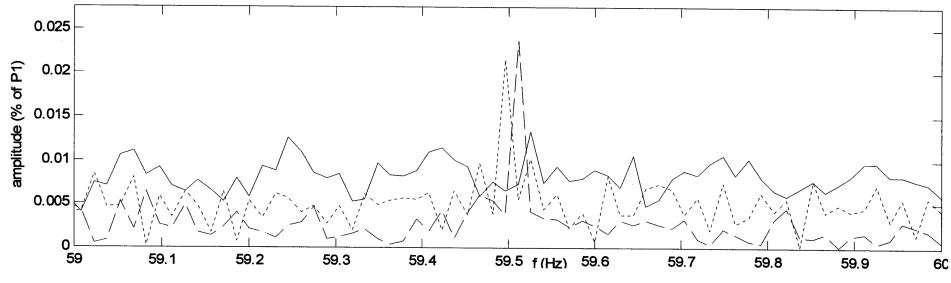

Figure 17. Detail of the amplitude spectrum (59-60 Hz band) of condenser fan power (P1) with 0g (solid), 2.6g (dashed), and 5.3g (dotted) imbalance fault levels.

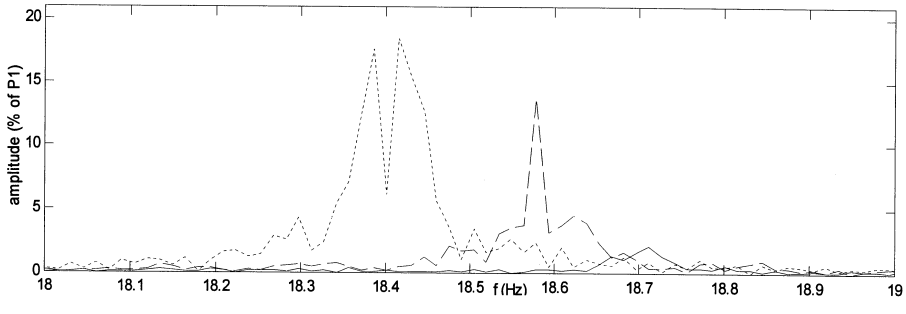

Figure 18. Detail of the amplitude spectrum (18-19 Hz band) of condenser fan power (P1) with 0g (solid), 2.6g (dashed), and 5.3g (dotted) imbalance fault levels.

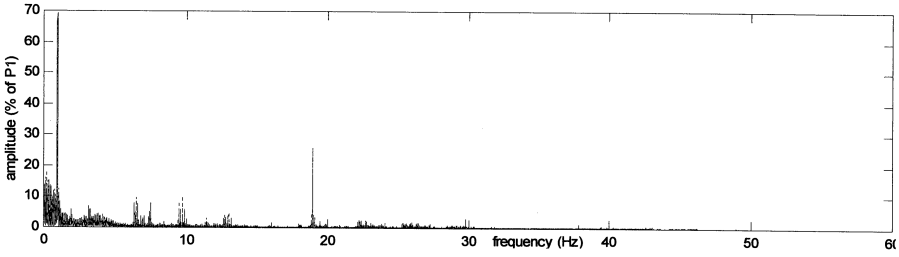

Figure 19. Amplitude spectrum (0-60 Hz, N = 37300) of supply fan power (P1) with 0g (solid), 8g (dashed), and 16g (dotted) imbalance fault levels.

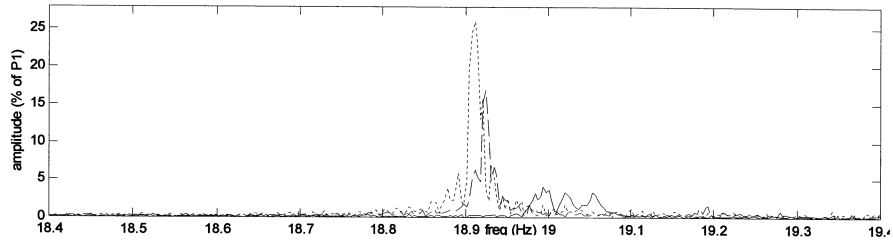

Figure 20. Detail of the amplitude spectrum (18.4-19.4 Hz band) of supply fan power (P1) with 0g (solid), 8g (dashed), and 16g (dotted) imbalance fault levels.

Compressor Liquid Ingestion

Compressor damage by liquid ingestion (Liu and Soedel 1994; Tomczyk 1995)[7] is likely to be preceded by one or more incidents in which small, relatively inconsequential, amounts of liquid[8] enter the machine. The NILM's ability to detect flooded starts and liquid slugging was tested. To introduce a known, repeatable mass of liquid at the suction port, it was necessary to perform the tests in the lab using a small semi-hermetic compressor.

To determine the effect of liquid ingestion on start transients, pairs of successive starts were monitored. Liquid was injected in the first start of each pair and not in the second. By alternating between wet and dry starts, certain confounding effects, such as change in winding temperatures and buildup of liquid residue on the valves, were minimized. A series of five pairs of wet-dry starts was observed with 1.0 cc oil injected above the intake valve ports immediately before each wet start. R-30 was injected after about 4 s and the compressor allowed to run another 10-20 s to clear oil out of the head and valve assemblies. The dry and wet start transients are shown in Figures 21 and 22. Figure 23 shows that there is a systematic difference in start transient shape. The three peaks indicate that wet start power increases significantly during compression strokes,

[7] Sean Mannion of United Refrigeration, Somerville, MA, provided additional information on March 1, 2004.

[8] An amount of liquid that at least fills the clearance volume, about 2% of displacement, can cause damage.

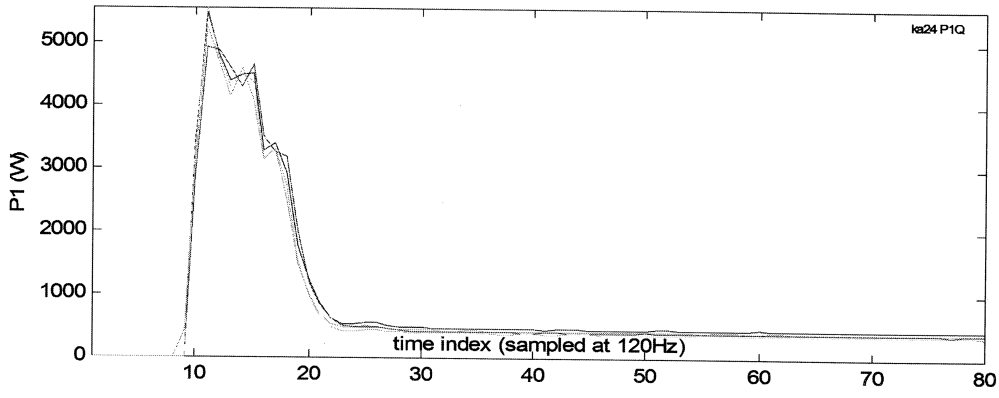

Figure 21. Reciprocating compressor start transients with no liquid ingestion, five repetitions.

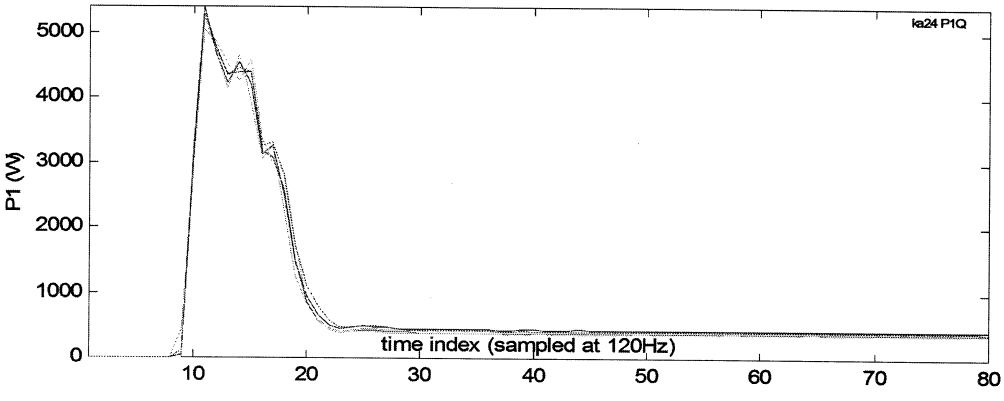

Figure 22. Compressor start transients with 1.0 cc liquid ingestion, five repetitions.

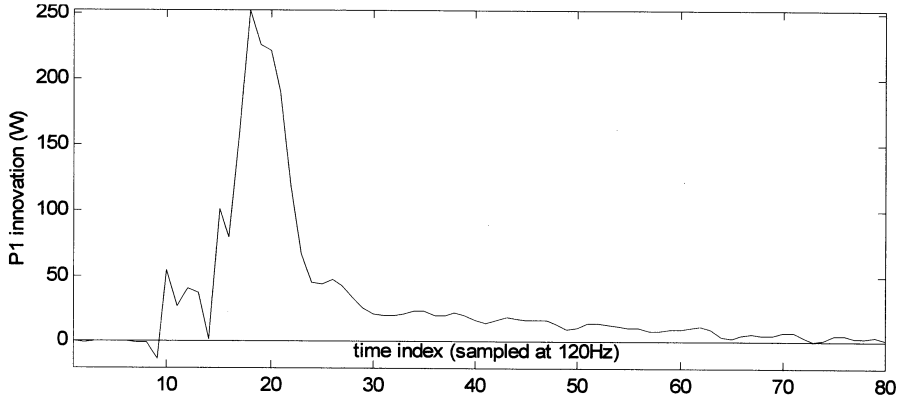

Figure 23. Mean innovation of P1 for starts with 1.0 cc liquid (deviation from fault free starts).

resulting in a distinct pattern that can be detected, even if the amplitude of the pattern is small relative to process noise.

A second series of alternating wet and dry starts was observed with 3.0 cc R-30 injected above the intake valve ports immediately before each wet start. The shape of the difference, Figure 24, is very similar to that obtained from the wet starts involving injection of 1.0 cc of oil.

Another liquid ingestion fault, most often described as *liquid slugging*, involves liquid entering the compressor during steady operation. To detect liquid slugging, we must first understand the steady-state compressor load under fault-free conditions. A reciprocating compressor presents a cyclic load component corresponding to shaft speed. More torque is required during suction and compression than when the crank is taking the pistons through their top and bottom positions. The frequency of the shaft speed cycle (twice the actual shaft speed, or about 58 Hz, for a symmetric two-cylinder machine) interacts with the second harmonic of line frequency to produce strong beats, evident in Figure 25 at 3.6 Hz. A model used to detect the small power transient associated with liquid slugging must account for the periodic variations in compressor power that are normal in steady operation. There is a further difficulty, not apparent in the 1 s time frame plotted in Figure 25, that the dc component of P1 varies gradually with refrigeration load, ambient temperature, and motor temperature (stator and rotor resistances).

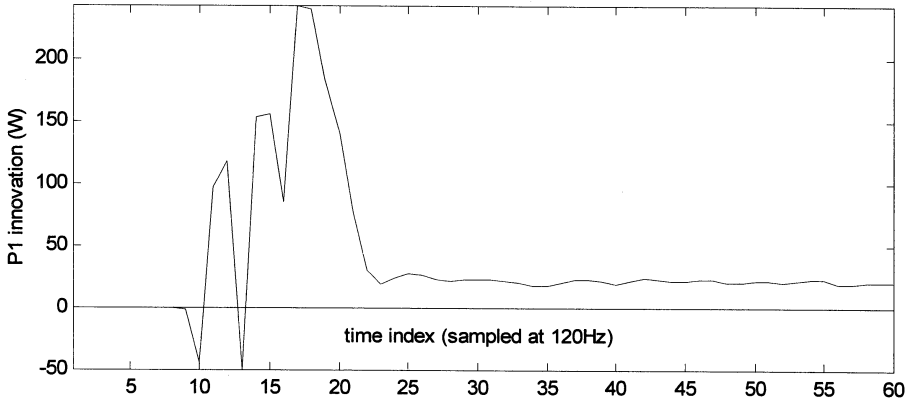

Figure 24. Mean innovation of P1 for starts with 3.0 cc liquid (deviation from fault-free starts).

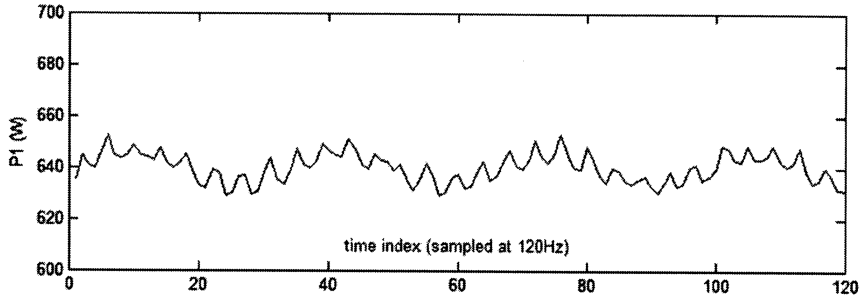

Figure 25. Compressor power observed during steady operation.

To address these difficulties, we model compressor power as a gradually changing autoregressive process (Seem and Braun 1991; Ljung 1999) with non-zero mean:

$$y(t) = \Phi(q)y(t) + \Delta + e(t), \tag{3}$$

where

$\Phi(q)$ = nth-order polynomial in the back-shift operator, q:[9]
$\Phi(q) = \varphi_1 q^{-1} + \varphi_2 q^{-2} + \ldots + \varphi_n q^{-n}$
Δ = the gradually changing mean compressor power (detrending parameter)
$e(t)$ = the estimation error, a combination of model error and observation noise

The model coefficients are assumed to vary slowly with time. In vector notation, let

$$\mathbf{x}(t) = [y(t-1)\ y(t-2)\ \ldots\ y(t-n)\ 1] \text{ and}$$

$$\mathbf{b} = [\varphi_1 \varphi_2 \ldots \varphi_n \Delta]^{\mathrm{T}}; \text{ thus,}$$

$$y(t) = \Phi(q)y(t) + \Delta + e(t) = \mathbf{x}(t)\mathbf{b} + e(t). \tag{4}$$

The model coefficients are estimated recursively by exponentially weighted least squares:

$$\mathbf{b} = \mathbf{b}_{t-1} + \mathbf{A}^{-1}\mathbf{x}^{\mathrm{T}}(t)(y(t) - \mathbf{x}(t)\mathbf{b}_{t-1}), \tag{5}$$

where

$\mathbf{A} = \lambda\mathbf{A}_{t-1} + \mathbf{x}^{\mathrm{T}}(t)\mathbf{x}(t)$.

The forgetting factor, λ, results in a weight sequence, $w = \lambda^p$, where p is the number of past time steps that decays exponentially with time. The efficient recursive least-squares algorithm that avoids matrix inversion ($\mathbf{A}^{-1}$) at every time step, based on the work of Kalman and Bierman is given accessible presentation in, e.g., Ljung (1999) and Wunsch (1996).

A fifth-order autoregressive model with one exogenous input (ARX(5,1) model, i.e., Equation 4 with $n = 5$) was applied to the 120 Hz signal. The forgetting factor was set to $\lambda = 0.995$ and the exogenous input was set to unity so that the corresponding model coefficient is a slowly varying estimate (i.e., exponentially weighted, moving average) of the "dc" component of P1 denoted in Equation 4 as Δ.

The foregoing model gives an estimate of compressor power for the next time step; i.e., it is the one-step-ahead forecast that is optimal for a quasi-stationary (slowly varying) fifth-order autoregressive process. A process disturbance can be detected because it will generally result in model forecast errors (innovations) of increasing absolute value, e.g., an increase in root-mean-square error norm. The one-step-ahead innovation is a very good indicator of an abrupt disturbance.

Liquid slugging, the disturbance of interest, might be caused by a faulty thermostatic expansion valve allowing too much liquid to accumulate in the evaporator. Instead of pure vapor, a mixture of vapor and liquid may exit the evaporator and some of the refrigerant may still be present as liquid at the compressor suction port. To observe this effect, equal doses of oil were injected, at the rate of about one injection every 20 seconds, during steady operation. A series of five events was observed with 1.0 cc oil injected at the suction port for each event. The P1 transients observed in tests are superimposed in Figure 26. Compressor power increases briefly (by about 15 W [2.5%] for about 0.5 s) and then returns gradually to normal. At higher time resolu-

[9] Note that $\Phi(q)y(t) = \varphi_1 y(t-1) + \varphi_2 y(t-2) + \ldots + \varphi_n y(t-n)$.

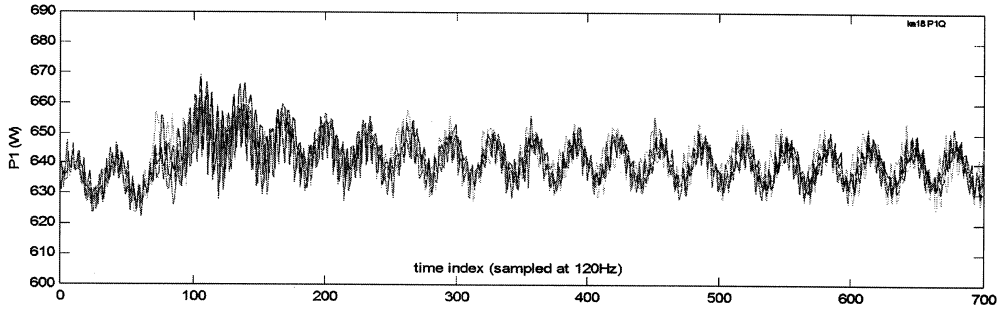

Figure 26. Real power during steady run liquid slugging, five repetitions.

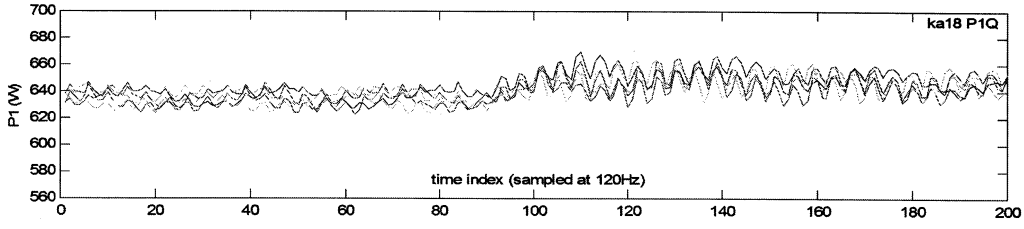

Figure 27. Detail of Figure 26 showing increase in amplitude of shaft-speed cycle. Note that beat cycles have been aligned for clarity, and the time reference of ingestion thus varies ± half a beat cycle among the five repetitions.

tion (Figure 27) the amplitude of the shaft cycle power (P1) fluctuations is seen to increase substantially when liquid is injected. The corresponding Q1 traces showed little or no sign of the disturbance.

An increase in one-step-ahead forecast error may be used to detect such a disturbance. However, this simple detection scheme is far from optimal because the liquid slugging disturbance is not abrupt. Compressor power rises over several time steps, persists, and then gradually returns to normal. A single slug of liquid that passes the suction flange in 0.1 s or less has a longer effect on compressor power—on the order of seconds. The adaptive one-step-ahead forecaster loses sensitivity because it starts to adapt to the new compressor behavior while the disturbance is still in progress. A more sensitive and reliable detector can be obtained by combining several forecasts that reach farther into the future and by using an appropriately weighted sum of the deviations of measured power from such multiple forecasts of power. We define the k-step-ahead forecast as

$$y_k(t + k) = \Phi_k(q)y(t) + \Delta_k + e_k(t). \tag{6}$$

Note that the adaptive model provides not only k-step-ahead predictions but also estimates of the variances, σ_k^2, of the prediction errors, e_k. The increase of error variance with number of forecast time steps is shown in Figure 28 and it approaches, as expected, the rms error of a simple adaptive model with just the gradually changing power parameter.

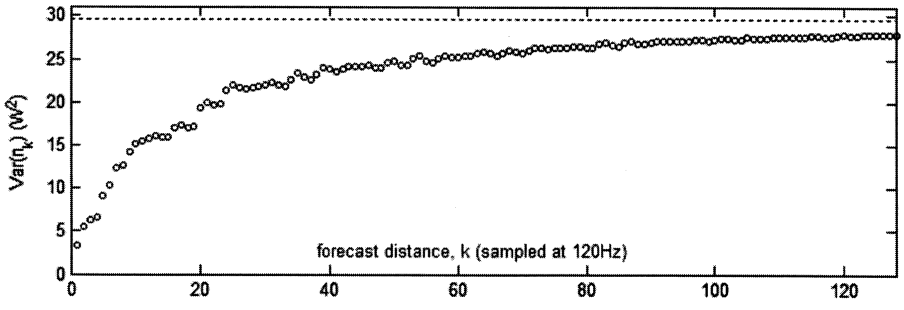

Figure 28. Variance for k-step-ahead prediction errors of the ARX(5,1) model during a randomly selected 10-second (N = 1200) period of steady operation. Dashed line is the variance about the mean of all 1200 observations.

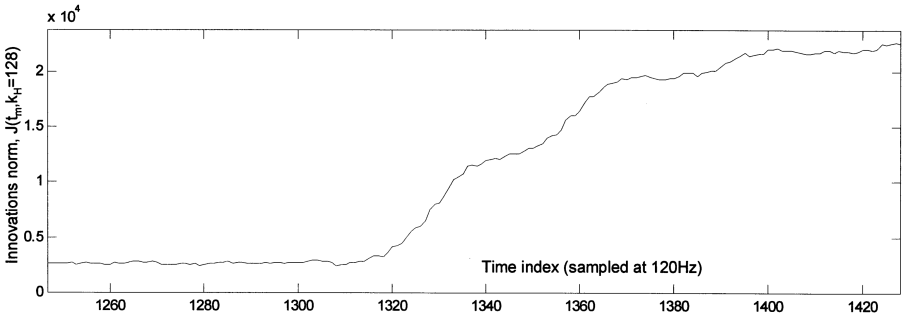

Figure 29. Liquid slugging at t = 1315 as detected by the weighted norm of ARX(5,1) k-step-ahead innovations (forecast errors) for k = 1:128.

Together these numbers may be used to reduce detection to an adaptive scalar threshold test based on the norm of 1/variance-weighted squared innovations:

$$J(t_m,k_H) = \mathbf{e}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{e} = \frac{e_1^2}{\sigma_1^2} + \frac{e_2^2}{\sigma_2^2} + \ldots + \frac{e_{k_H}^2}{\sigma_{k_H}^2} \tag{7}$$

where $e_k = y_k(t_m + k) - \mathrm{P1}(t_m + k)$ is the innovation at time $t_m + k$ according to the k-step-ahead forecast of time t_m.

Note that in the compact vector form, **e** is the vector of innovations estimated from the time (t_m) forecast models over a prediction horizon ($t_m + 1$: $t_m + k_H$) and **R** is the diagonal matrix of corresponding variance (mean-square-error) estimates (Figure 28) for the set of k-step-ahead forecast models defined as of time t_m. Thus, $J(t_m,k_H)$ is a scalar by which we can test, at every timestep t, the hypothesis that there has been a significant process disturbance since time $t_m = t - k_H$ based on (1) the set of $k = 1{:}k_H$ k-step-ahead forecast models available at t_m and (2) the set of measured process outputs, $y(t_m + k)$ $k = 1{:}k_H$, observed since t_m. The time evolution of this norm out to $k_H = 128$ is shown in Figure 29. This result is for a single liquid ingestion test. The smoothness of J is the result of ensemble averaging.

In summary we find that, as in the case of flooded starts, the effect on P1 of ingesting non-damaging amounts (5%-10% of displacement for liquid slugging) of liquid is small, making reliable detection tricky. There are four aspects of the difficulty: (1) small magnitude of the change relative to normal short-time variations in P1, (2) lack of abrupt transient features, (3) lack of a distinct transient shape, and (4) gradual changes in mean power that occur with motor temperature (rotor and stator resistance) and refrigeration load. Use of the weighted innovations norm based on an adaptive ARX model addresses these difficulties in a rigorous, yet practical, way.

SUMMARY AND DISCUSSION

We conclude with a summary of the fault detection methods presented, a brief description of the hardware implementation and its cost, and a semi-quantitative assessment of benefits.

Fault Detection Architecture. The NILM uses five generic detection methods:

- Change in mean of steady-state power (P1) and reactive power (Q1) (Hart 1992)
- Frequency analysis of the real-power amplitude spectrum
- Pattern matching of start transients (Leeb 1992)
- Detection of anomalous transients
- Voltage (threshold) and current (ratio) monitoring to detect phase imbalance

The data collected in RTU fault-free modes indicate that the start transients of small single-phase fans and small three-phase compressors are just as distinct and repeatable as for large three-phase fans and compressors (Laughman et al. 2003). Air-side blockage faults can be detected by a shift in steady-state P1 and Q1 and, to some extent, by changes in start transient shape. However, the change magnitudes and directions are related to blockage magnitude via the interaction of fan curve, air-side flow-pressure curve, motor torque curve, and no-fault operating point. Thus, diagnosis requires more information than that needed for detection, e.g., addition of static pressure. Flow transitions clearly evident in the pressure start transient are also evident in the P and Q start transients.

Short cycling can be detected in the obvious way—by logging time between start transients. The advantage over detection by a temperature-sensor-based FDD system is that very short—even incomplete—start transients are detected.This is important because just a few inrush current events in quick succession will raise stator winding temperature to the point of insulation damage. The second method of short-cycling detection relies on the shape of individual start transients being sensitive to residual head. The two methods should be implemented together because they are complementary. A time-stamp and the thermal conditions at the time of each short-cycle occurrence can be saved to aid in diagnosing a root cause.

Refrigerant charge faults result in elevated head pressures, which translate to higher steady-state compressor loads. Detection of undercharge is possible based only on the distinct shape of its start transient. Detection of overcharge, on the other hand, requires normalizing compressor load over the range of thermal operating conditions.

The refrigerant vapor bypass fault has a distinct effect on the start transient shape: the change of magnitude in the latter half of the transient is two to five times the change in steady-state power and can be scaled by the steady-state power to improve the detection statistics.

Shaft harmonics have been used to detect faults in belt-driven machines (Lee 2003). Belt-driven supply fans are used in some package equipment, but condenser fans are almost always direct-drive machines. In the present work, tests using small weights to unbalance direct-drive impellers indicated that the NILM can identify degradation of impeller balance well before catastrophic failure and associated additional repair costs are incurred.

Flooded start is detected by a change in start transient shape that exhibits distinct compression stroke features. The shapes of liquid ingestion transients during steady operation are not repeatable in the usual sense, but they always result in increases in power that appear clearly as positive deviations from k-step-ahead forecasts. Thus, any transient occurs when all motors are running and that exceeds the adaptive model innovation norm threshold indicates liquid ingestion. Conditions recorded at the time of occurrence can be saved to aid in diagnosing a root cause.

Hardware Implementation. The NILM requires only current and voltage sensors at the RTU feed, all located safely in the electrical box. A single-phase NILM requires one current sensor and one voltage sensor, while a three-phase NILM requires two of each sensor. There may be a niche for a NILM with one current and three voltage sensors so that phase imbalance or phase loss can be immediately detected. The single-phase NILM can be implemented in a low-cost 150 MHz PC-on-a-chip with a two-channel A/D converter. The incremental RTU manufacturing cost could be as low as $200. Integration with RTU controls is a natural extension that reduces the incremental cost of FDD capability and may be especially attractive for high SEER designs with variable-speed fans or compressors. Such designs typically require an embedded processor to compute control actions.

Benefits. The suite of faults that are best detected by the NILM complements, to a large extent, the faults that are best detected by temperature-based FDD systems. Faults that are primarily electrical or electromechanical are difficult to detect by thermal measurements but can be detected by the NILM using the methods indicated in Table 2. The NILM can also play a role in detecting some of the more interesting non-electrical faults as summarized in Table 3. A comprehensive FDD system could employ an extension of the models developed by Rossi and Braun (1997) and Breuker and Braun (1998a, 1998b), which currently use return air and ambient air conditions as inputs, by adding compressor power as an input or output (innovations approach).

Finally, Table 4 presents our reinterpretation of Breuker and Braun's data (1998a) to estimate the cost-savings potential of the NILM in detecting RTU faults. An estimated 33% of repair costs are associated with faults that can be detected by the NILM with near certainty and an additional 26% of the costs are due to faults that the NILM can probably detect. Not all of the resulting 59% in repair costs can be avoided by automated fault detection. But early detection by the NILM can be expected eliminate a substantial fraction of those costs and translate to savings on the order of 10% to 30% of total repair costs.

Table 2. Summary of Electrical and Electromechanical RTU Faults Detectable by the NILM Alone

Fault	NILM Method
Short cycling	Event sequence
Fan rotor faults that result in imbalance	Amplitude spectrum in steady operation
Loss of phase	Current and voltage
Locked rotor	Start transient
Slow-starting motor	Start transient
Unbalanced voltage	Voltage
Motor disconnect/failure to start	Event sequence
Incorrect control sequence	Event sequence
Contactor (improper contact closure)	Phase current interruption transient
Compressor mechanical faults (not tested)	Amplitude spectrum in steady operation

Table 3. Summary of Non-Electrical Faults [1]

Fault	NILM Method	7T[2]	10T[2]
Refrigerant leakage, undercharge, or overcharge	Change of mean	✔	✔
Loss of volumetric efficiency (leaky valves, seals)	Start transient	[3]	✔
Fouled condenser coil (flow blockage)	Change of mean	✔	✔
Dirty supply air filter (flow blockage)	Change of mean [4]	✔	✔
Flooded start	Start transient		
Liquid slugging	Anomalous transient		
Liquid line restriction	[5]	✔	✔
Vapor line restriction	[5]		
Noncondensable gas	[5]		
COP degradation (any of the above root causes)	[6]	[7]	[7]

[1] Some of these can only be detected by thermal measurements and some can only be detected by the NILM, but detection of most can benefit by having both types of measurements.
[2] 7T FDD measures T_{wb}, T_{amb}, T_{evp}, T_{sh}, T_{hg}, T_{cnd}, and T_{sc}; 10T FDD also measures T_{ra}, ΔT_{cndair}, and ΔT_{evpair} (Breuker and Braun 1998b).
[3] Low sensitivity
[4] System flow-pressure-temperature model may be needed for reliable detection
[5] Not tested
[6] With addition of refrigerant flow (or compressor map), T_{evp}, $P_{suction}$, and $P_{discharge}$
[7] With addition of compressor power and refrigerant mass flow rate

Table 4. Cost Breakdown for Repair of Faults*

	%	Detection by NILM Certain	%	Detection by NILM Probable	%
Compressor	24	Shorted, open, locked rotor	66	Broken internals, vapor bypass	11
Condenser	9	Motor, belt failure	51	Fouling, cycle control	36
Evaporator	6		0	Fouling, loss of capacity	75
Air handler	5	Motor, impeller, belt, electrical	67	Damper motor, duct leaks	20
Controls	10		0	Compressor interlocks, fan sequence	37
Electrical	7	Contactor, loose connection	44	Fuse/breaker, short/open	27
Other	21	Water loop/pump, fan belt	29	Refrigerant leak, installation/start-up	42
Unclassified	18	NA		NA	
Total %	**100**	**(Dot product with left column)**	**33**	**(Dot product with left column)**	**26**

* The first column lists repair cost as percent of total; the second and third give percents of column 1 based on Tables 2-8 of Breuker and Braun (1998a). Assignment of faults to columns 2 and 3 represent the authors' qualitative assessment of NILM FDD potential for the fault subcategories tabulated by Breuker and Braun.

ACKNOWLEDGMENTS

This research was sponsored by the California Energy Commission with support from the US Navy and the Grainger Foundation. Pacific Northwest National Laboratory provided tuition during two semesters and logistical support throughout Armstrong's educational leave. Technical assistance was generously and expertly provided at the Purdue field test site by Jim Braun, Haorong Li, and Frank Lee.

REFERENCES

Armstrong, P.R. 1983. Instrumentation of solar heating and cooling systems at CSU. *Proc. 2nd Workshop on Performance Monitoring, FSEC, Cape Canaveral*.

Armstrong, P.R. 2004. Model identification with applications to building control and fault detection. PhD thesis, Massachusetts Institute of Technology.

Breuker, M.S., and J.E. Braun. 1998a. Common faults and their impacts for rooftop air conditioners. *HVAC&R Research* 4(3):303–318.

Breuker, M.S., and J.E. Braun. 1998b. Evaluating the performance of a fault detection and diagnostic system for vapor compression equipment. *HVAC&R Research* 4(4):401–425.

Genta, G. 1998. *Vibration of Structures and Machines—Practical Aspects*, 3d ed. Springer-Verlag.

Hart, G.W. 1992. Nonintrusive appliance load monitoring. *Proc. IEEE* 80(12):1870–1891.

Laughman, C.R., et al. 2003. Power signature analysis. *IEEE Power and Energy*, March 2003.

Lee, D.K. 2003. Electric load information system based on non-intrusive power monitoring. PhD thesis, Massachusetts Institute of Technology.

Leeb, S.B. 1992. A conjoint pattern recognition approach to non-intrusive load monitoring. PhD thesis, Massachusetts Institute of Technology.

Li, H., and J.E. Braun. 2004. An economic evaluation of fault detection and diagnosis for rooftop air conditioners. *Tenth Int'l Refrigeration and Air Conditioning Conference at Purdue.*

Liu, Z., and W Soedel. 1994. An investigation of compressor slugging problems. *XII Int'l Compressor Engineering Conference at Purdue*, pp. 433–440.

Ljung, L. 1999. *System Identification: Theory for the User*. Englewood Cliffs, N.J.: Prentice Hall.

Rossi, T., and J.E. Braun. 1997. A statistical, rule-based fault detection and diagnostic method for vapor compression air conditioners. *HVAC&R Research* 3(1):19–37.

Seem, J.E., and J. Braun. 1991. EWMA forecast of building electrical load. *ASHRAE Trans.* 97(1):710–721.

Shaw, S.R., C.B. Abler, R.F. Lepard, D. Luo, S.B. Leeb, and L.K. Norford. 1998. Instrumentation for high performance nonintrusive electrical load monitoring. *ASME J. Solar Energy Engineering* 120:224–229.

Shaw, S.R., and C.R. Laughman. 2005. A Kalman-filter spectral envelope preprocessor. *IEEE Trans. Instrumentation and Measurement* (accepted for publication).

Stouppe, D.E., and T.Y Lau. 1989. Air conditioning and refrigeration equipment failures. *National Engineer* 93(9):14–17.

Tomczyk, J. 1995. *Troubleshooting and Servicing Modern Air Conditioning and Refrigeration Systems*. Business New Publishers (BNP).

Wild, C.J., and G.A.F. Seber. 2000. *Chance Encounters: A First Course in Data Analysis and Inference*. Wiley.

Wunsch, C. 1996. *The Ocean Circulation Inverse Problem*. Cambridge U. Press.