Evaluating the performance of fault detection and diagnostics protocols applied to air-cooled unitary air-conditioning equipment

David P Yuill, University of Nebraska-Lincoln
James E Braun, Purdue University

Available at: https://works.bepress.com/david_yuill/2/
Evaluating the performance of FDD protocols applied to air-cooled unitary air-conditioning equipment

David P. Yuill and James E. Braun

Ray W. Herrick Laboratories, School of Mechanical Engineering, Purdue University, 140 S. Martin Jischke Drive, West Lafayette, IN, 47907-2031, USA

To cite this article: David P. Yuill & James E. Braun, (2013), Evaluating the performance of fault detection and diagnostics protocols applied to air-cooled unitary air-conditioning equipment, HVAC&R Research, 19:7, 882-891.

ABSTRACT

Fault detection and diagnostics (FDD) tools are increasingly being applied to air-cooled unitary air-conditioning systems. However, it is not known how well these tools work because there is no standard method of measuring or evaluating the performance of FDD. In the current paper the authors describe the common faults that FDD is applied to in unitary systems, and propose a method of evaluating the performance of FDD protocols. The method involves feeding measurement data through a candidate protocol and collecting and organizing the responses based upon the fault’s impacts on performance. A library of faulted and unfaulted measurement data has been built up, and is described. Standard definitions are proposed for several pertinent terms and quantities. A case study is used to demonstrate the evaluation method, using a publicly available FDD protocol that is evaluated using the complete library of data. This protocol is found to perform poorly.

David P. Yuill, PE, Student Member ASHRAE, is Graduate Research Assistant. James E. Braun, PhD, PE, Fellow ASHRAE, is Herrick Professor of Engineering.

*Corresponding author e-mail: dyuill@purdue.edu
INTRODUCTION

Fault detection and diagnostics (FDD) is an area of supervisory control that was introduced in the 1970s (Himmelblau 1978; Isermann 1984) for use in life-critical processes such as nuclear power, aerospace, and military applications, in which early detection of a fault may prevent catastrophic failure. The FDD process compares sensed data to the expected values of these data under given operating conditions to determine whether the data are within the expected ranges, and to determine what might cause them to be out of range. As the cost of sensors and controllers has decreased, FDD has been applied to many other engineering processes such as heating, ventilation and air-conditioning (HVAC), (Breuker and Braun 1998a; Katipamula & Brambley 2005). The objectives of applying FDD to HVAC are generally to sense subtle faults that degrade performance or reduce the expected equipment life, since such faults may go unnoticed by equipment operators until they cause outright equipment failure.

FDD has been applied to many kinds of HVAC equipment, such as chillers (Comstock and Braun 1999; Reddy 2007; Zhao et al. 2011), cooling coils (Veronica 2010), VAV air handling systems (Norford et al. 2002; House et al. 1999, Wang et al. 2011) and packaged air-cooled vapor compression air-conditioning equipment (Rossi and Braun 1997; Li and Braun 2003, Kim et al. 2008, Armstrong et al. 2006). The latter type of system, which includes rooftop units and split systems, is the focus of the current paper and will be referred to as a “unitary system”. FDD on systems of this type is of particular interest for three reasons. The first is that these systems are very widely deployed, used in most houses and responsible for about 60% of the cooling energy used by commercial buildings in the US (Feng et al. 2005). The second is that these systems are often deployed in applications in which the operator (e.g. homeowner or property manager) does not provide regular maintenance and may not have the capability to recognize the presence of faults until the system fails (Roth et al. 2006). Finally, these systems likely have a higher incidence of faults because of the lack of maintenance and because of installation issues related to lower-cost and less sophisticated systems (Downey and Proctor 2002, Wiggins and Brodrick 2012). As a result there are currently several companies that market FDD for unitary systems and there are equipment manufacturers that are including FDD capabilities in some of their unitary equipment product lines. Furthermore, the 2013 California Title 24 energy code (CEC 2012) requires FDD on newly-installed unitary equipment.
When considering which FDD approach to use, or when considering whether a particular FDD approach meets a code requirement, the obvious question to ask is: how well does it work? Answering this question is not simple. There is currently no standard method of evaluating the performance of FDD applied to unitary equipment, and “there are currently no available military or commercial standards to support a systematic and consistent approach to assessing the performance and effectiveness” of FDD applied to engineered systems in general (Vachtsevanos et al. 2006).

There has been some previous research that considered evaluation of FDD for vapor compression air-conditioning equipment. Breuker & Braun (1998b) studied the accuracy of an FDD tool developed by Rossi and Braun (1997) when applied to a specific rooftop unit, and methods of tuning parameters within the tool to achieve optimal performance. Reddy (2007) discusses generic evaluation methodologies for assessing different FDD protocols applied to large chillers. None of the previous research proposed a standard method of test or evaluation and rating system.

FDD has the potential to provide significant benefits. Surveys of air-conditioning systems have found a large fraction to be operating with a fault (Rossi 2004; Downey and Proctor 2002; Breuker et al. 2000) that can have significant effects on capacity, efficiency and equipment life. For example, if refrigerant undercharge faults were eliminated from only the currently deployed residential air conditioners in the US, it is estimated that residential cooling energy consumption would be reduced by 0.1 to 0.2 quad per year, i.e. 5 to 10% (Roth et al. 2006). However, FDD in unitary equipment is still a somewhat immature technology, as evidenced by the widely varying approaches used and by the low rate of adoption. Developing a method to test and evaluate FDD protocols is expected to help advance the technology in three ways. First, it will allow regulatory bodies to give meaningful specifications for FDD requirements. Second, it will allow users of FDD – including equipment manufacturers, facilities operators, utility incentive managers or equipment owners – to make informed decisions about whether to use FDD and which protocol will work best for them. Finally, it will aid the development and improvement of FDD algorithms by providing a measure by which improvements can be tracked.

This paper describes the first stage of development of FDD evaluation methodologies that are intended to advance the technology in the three ways discussed above. An evaluation method is described first, followed by definition and discussion of fault types that are included in the method. Then the library of
input data used in the evaluation is described. A case study is used to demonstrate the application of the evaluation method. Finally, there is a brief discussion of the path forward for development and application of FDD evaluation.

EVALUATION METHOD

To evaluate an FDD protocol, inputs are fed to the protocol and the protocol’s responses are observed. The term “protocol” in this paper refers to the input specification, procedure, algorithm, and data formatting for an FDD tool. This excludes the hardware, such as sensors, and the interface between hardware and algorithm. The inputs must come from, or effectively mimic, the sensed conditions in both faulted and unfaulted systems subjected to typical operating conditions. These operating conditions include combinations of: (a) outdoor air temperature; (b) indoor air temperature; (c) indoor air humidity; (d) type of fault present, if any; (e) intensity of the fault; (f) specific unitary system. Each input scenario gives one test result. There are five possible test results, with respect to fault detection and isolation:\(^2\): (i) No Response – the FDD can’t be applied, or can’t give a diagnosis; (ii) Correct; (iii) False Alarm – a fault is indicated when no fault is present; (iv) Misdiagnosis – the wrong fault is diagnosed; (v) Missed Detection – the FDD indicates no fault when a fault is present. More rigorous definitions are given below.

The evaluation process is summarized in Figure 1. The raw results that are generated from a complete set of input scenarios are collected and organized as percentages to summarize the performance of the protocol.

\(^2\) Fault detection refers to detection of an abnormal operating condition without specification of the type or severity of the fault. Fault diagnosis consists of two processes: 1) fault isolation, in which the type or location of the fault is identified (e.g. “low evaporator airflow”, or simply “evaporator fault”); 2) fault assessment, in which the magnitude of the fault is indicated (e.g. “slightly low evaporator airflow”, or “10% low evaporator airflow”).
There are two ways used to characterize fault levels in this paper. The first is Fault Intensity (FI), which is directly related to measureable quantities (e.g., 20% undercharge). In this paper, FI is used to describe the fault levels applied in the laboratory experiments that represent the fault input scenarios. For characterizing performance of an FDD algorithm in this paper, frequency-of-occurrence rates (e.g., percent false alarms, percent misdiagnosed, etc.) are used as performance indices and presented in terms of a fault level index termed the Fault Impact Ratio (FIR). FIR is related to equipment performance, and is tied to either capacity or COP. For example, when \( \text{FIR}_{\text{COP}} = 95\% \), it means that the equipment is operating at 95% of its maximum efficiency under a given set of driving conditions. Strict definitions of FIR and of FI for different faults are presented in later sections.

Using FIR as the independent variable in characterizing FDD performance represents an improvement over the use of FI that was employed by Yuill and Braun (2012) because fault impact is what the owner or operator of an air-conditioning system is concerned about. The use of FIR allows a potential user to understand the performance of an FDD protocol for a fault impact level that is appropriate for their application. FIR is used in different ways within the current paper to present results. For instance, FDD false alarm rates are presented as a function of FIR thresholds that draw the distinction between acceptable faulted and unfaulted cases. If the FIR threshold for COP is 95%, it means that test cases with COP within 5% of the maximum for that operating condition are considered to be unfaulted, regardless of the FI. For missed detection and misdiagnosis performance, frequencies of occurrence are presented within fixed bins of FIR.

The process depicted in Figure 1 is repeated over a range of input scenarios, and the results are aggregated to generate frequency-of-occurrence rates for No Response, False Alarm, Missed Detection and Misdiagnosis. The rates are calculated using the ratios of the number of occurrences of the test result relative to the potential maximum number of occurrences, as described in the following subsections. The % Correct is not presented since it is implied by the percentages of incorrect results. The rates for Misdiagnosis and Missed Detection are calculated in two different ways, denoted (a) and (b). The (a) rates consider all faults in the evaluation method’s input data, regardless of whether the protocol is intended to diagnose all of these faults. This is an important performance indicator, because a protocol may be sensitive to the presence of faults it is not intended to diagnose. For example, if a protocol is intended only
for determination of correct refrigerant charge, its usefulness is greatly reduced if it performs poorly when other faults are present. The (b) rates consider only those faults for which a protocol is intended. These rates give additional insight into the protocol’s performance, but are less important in evaluating the overall utility of the protocol.

No Response

Numerator: number of cases in which the protocol cannot be applied or gives no response because of excessive uncertainty.

Denominator: total number of test cases.

False Alarm

Numerator: the number of cases in which: i) the protocol indicates the presence of a fault, ii) the fault impact is below a given threshold, and iii) the system isn’t overcharged with refrigerant by more than 5%. The last criterion is important because a system that is overcharged by more than 5% may not suffer significant degradation in efficiency or capacity, but could damage the compressor, so an operator would want to know about this fault. Therefore, significantly overcharged cases are not classified as False Alarms regardless of their energy or capacity impacts (they can be classified as Correct, Misdiagnosis or Missed Detection).

Denominator: the number of cases in which the protocol gives a response and the fault impact is below a specified threshold.

Misdiagnosis (a)

Numerator: the number of cases in which: i) the fault impact is within a specified range, ii) the experimenter indicated the presence and intensity of a fault, and iii) the protocol indicates that the system has a fault different from the type of fault indicated by the experimenter.

Denominator: the number of cases in which: i) the fault impact is within a specified range, ii) the experimenter indicated the presence and intensity of a fault, and iii) the protocol indicates that the system has a fault.

Misdiagnosis (b)
Numerator: the number of cases in which: i) the fault impact is within a specified range, ii) the experimenter indicated the presence and intensity of a fault type that the protocol is intended to diagnose, and iii) the protocol indicates that the system has a fault different from the type of fault indicated by the experimenter.

Denominator: the number of cases in which: i) the fault impact is within a specified range, ii) the experimenter indicated the presence and intensity of a fault type that the protocol is intended to diagnose, and iii) the protocol indicates that the system has a fault.

**Missed Detection (a)**

Numerator: the number of cases in which: i) the fault impact is within a specified range, ii) the experimenter indicated the presence and intensity of a fault, and iii) the protocol indicates that the system has no fault.

Denominator: the number of cases in which: i) the fault impact is within a specified range, ii) the experimenter indicated the presence and intensity of a fault, and iii) the protocol gives a response.

**Missed Detection (b)**

Numerator: the number of cases in which: i) the fault impact is within a specified range, ii) the experimenter indicated the presence and intensity of a fault type that the protocol is intended to diagnose, and iii) the protocol indicates that the system has no fault.

Denominator: the number of cases in which: i) the fault impact is within a specified range, ii) the experimenter indicated the presence and intensity of a fault type that the protocol is intended to diagnose, and iii) the protocol gives a response.

**FAULT TYPES AND CHARACTERIZATION**

In unitary air conditioners there are six degradation faults that are commonly detected with FDD protocols, listed in Table 1. The effects of these faults on the system’s operation are described in Breuker & Braun (1998a).

**Table 1: Faults common to air-cooled unitary equipment**
<table>
<thead>
<tr>
<th>Fault</th>
<th>Abbr.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under- or overcharge</td>
<td>UC, OC</td>
<td>A mass of refrigerant that is less or more than either i) the manufacturer specification, or ii) the mass that gives maximal efficiency at a given operating condition</td>
</tr>
<tr>
<td>Low-side heat transfer</td>
<td>EA</td>
<td>Faults in the evaporator coil, such as coil fouling or insufficient airflow</td>
</tr>
<tr>
<td>High-side heat transfer</td>
<td>CA</td>
<td>Faults in the condenser coil, such as coil fouling or insufficient airflow</td>
</tr>
<tr>
<td>Liquid line restriction</td>
<td>LL</td>
<td>Flow restrictions such as crimps or fouled filter/drier in the liquid line (Figure 2)</td>
</tr>
<tr>
<td>Non-condensables</td>
<td>NC</td>
<td>Gases that do not condense (e.g. air or nitrogen) in the refrigerant</td>
</tr>
<tr>
<td>Compressor valve leakage</td>
<td>VL</td>
<td>Leaks in the compressor from high to low pressure regions, reducing mass flow</td>
</tr>
</tbody>
</table>

**Figure 2: The major components of a vapor compression air-conditioner**

**Implementing faults in a laboratory test**

Faults are implemented during laboratory testing over a range of driving conditions (temperature and humidity of air entering the indoor coil, and ambient temperature) and with a range of fault intensities (FI).

The FI, defined in the following sections, is proposed as standard terminology for describing faults.

1. **Charge**: To impose a refrigerant under- or overcharge fault, charge is removed from or added to the system. The fault intensity is:
\[ FL_{\text{charge}} = \frac{m_{\text{actual}} - m_{\text{nominal}}}{m_{\text{nominal}}} \]  

where \( m_{\text{actual}} \) is the measured mass of refrigerant in the system  
\( m_{\text{nominal}} \) is the correct mass of refrigerant

Thus a system designed for 2 kg of charge that had 1.8 kg would be referred to as 10% undercharged or having \( FL_{\text{charge}} = -10\% \). \( m_{\text{nominal}} \) can be defined using the manufacturer specification or the mass of refrigerant that provides the maximum efficiency at a given operating condition.

2. **Low-side heat transfer faults**: In a typical laboratory setup the airflow across the evaporator coil can be modulated using a variable speed booster fan or dampers. Reducing the airflow accurately duplicates the effect of most faults in this category: airflow reduction from fan or distribution system design problems, obstructions, or filter fouling. The effect of evaporator coil air-side fouling has also been shown by Bell et al. (2012) to be well represented by reducing airflow, particularly if the fouling is assumed to be evenly distributed across the face of the heat exchanger. The fault intensity is defined using either mass flow rate or volumetric airflow rate.

\[ FL_{EA} = \frac{V_{\text{actual}} - V_{\text{nominal}}}{V_{\text{nominal}}} \]  

3. **High-side heat transfer faults**: Similarly to low-side faults, a reduction in airflow is used to implement high-side heat transfer faults. Some experimenters have simulated blockage by large scale debris, such as leaves, by covering the face of the condenser coil with paper or mesh. Although this may, in some cases, more realistically represent the fault physically, it is not repeatable nor easily quantified as a fault intensity. Furthermore, the general effect – to increase the refrigerant’s enthalpy in the liquid line – is the same as with reduced airflow. Therefore, reduced airflow is proposed as the standard means of imposing this fault in the laboratory. Accordingly, the fault intensity is defined with airflow rates in the same manner as with low-side heat transfer faults.
4. Liquid line restriction: A liquid line restriction is implemented by using one or more valves to impose the desired pressure loss. The fault intensity is defined using the ratio of the increase in pressure drop through the liquid line caused by the faulted condition, to the liquid line pressure drop under unfaulted operation and at the same operating condition.

\[ F_{LCA} = \frac{\dot{V}_{actual} - \dot{V}_{nominal}}{\dot{V}_{nominal}} \]  

\[ F_{LL} = \frac{\Delta P_{LL, fault} - \Delta P_{LL, unfaulted}}{\Delta P_{LL, unfaulted}} \]  

5. Non-condensables in the refrigerant: A non-condensables fault is imposed by introducing nitrogen into the refrigerant line. The maximum amount of non-condensables to be expected is in the case where a system has been open to the atmosphere and not evacuated prior to charging. Therefore the fault is defined with a mass of nitrogen compared to the mass of nitrogen that would fill the system at atmospheric pressure.

\[ F_{NC} = \frac{m_{N2, fault}}{m_{N2, ref}} \]  

6. Compressor valve leakage: A compressor valve leakage is simulated with the use of a hot gas bypass – a pipe carrying refrigerant from the discharge to the inlet of the compressor (from point 1 to point 5 in Figure 2). The fault intensity for this fault is defined as the change in mass flow rate (at a given operating condition) to the original mass flow rate.

\[ F_{VL} = \frac{\dot{m}_{faulted} - \dot{m}_{unfaulted}}{\dot{m}_{unfaulted}} \]

APPLYING THE EVALUATION METHOD

To apply the evaluation method described above requires input data from unitary systems that are operating with and without faults under a range of driving conditions. These input data must include any inputs that are required by candidate FDD protocols, and they must also include information on fault
impacts. It is very important that these input data are accurate. There are two general approaches to building a library of input data. The first is to use experimental data. This approach has the challenge that measurements of this kind are very difficult to conduct, hence the cost and experimental uncertainty can be quite high. The second approach is to use data from simulations. The main advantage of this approach is that with a reliable model a large number of data can be generated to evenly blanket the input space of interest (operating conditions, fault types, fault intensities, and air-conditioning units). Other advantages are that random error can be eliminated and physical constraints, such as energy and mass balances, can be imposed. The main disadvantage of simulation is the significant technical challenges of simulating systems operating with faults. The development of a physics-based model can be very difficult and time consuming, requiring a complete set of physical dimensions and properties of the system and a set of validation data for each unit modeled. An inverse model, on the other hand, is comparatively simple to develop, but has very limited use.

The current paper focuses on evaluation using experimental data. However, modeling methods currently in development (Cheung and Braun 2013) will be used in the future to enable evaluation using simulation-generated input data. The current evaluation approach also is limited in the following ways:

1. **Steady state performance** – The most common approach for FDD in unitary systems is to wait for the system to reach steady state within some tolerance. Some methods use a steady-state detector to do this, some require a system to run for a given period of time, and some require a user to determine whether steady operation has been achieved. All of the input data in the current data library represent steady-state operation.

2. **Cooling mode** – The methods in this paper could be applied to heat pumps operating in heating mode, but currently they are being applied to FDD on units operating in cooling mode only.

3. **Single Fault** – In some systems operating in the field there will be more than one fault occurring simultaneously. Diagnosing multiple simultaneous faults is very challenging because the symptoms of faults can be similar for some variables, or can have opposing
effects on some variables. Almost all FDD protocols limit their diagnoses to single faults (Li and Braun 2007c). The current evaluation method is limited to single faults.

4. **Single speed systems** – Some unitary systems have variable or multiple speed motors driving the fans or compressor. The FDD evaluation method considers only single-speed unitary systems.

5. **Diagnosis evaluation of fault isolation only** – Some FDD tools provide a diagnosis that includes both fault isolation and fault assessment. The current evaluation method evaluates FDD protocols’ ability to correctly isolate faults, but not the ability to provide fault assessment.

**Library of experimental data**

To provide the largest possible range of inputs to an FDD algorithm being evaluated, we have gathered all of the available sets of laboratory experimental results for fault testing of unitary air-conditioning equipment and organized them into a standardized format developed for the evaluations. Laboratory measurements are used because it is vital that each sensed quantity be accurately measured, and controlling field conditions to get accurate measurements is impractical.

The data are from 9 different unitary systems, including rooftop units and split systems, with a total of 607 test cases. The data sets are from previous research carried out by university and government laboratories. The experimenters followed the same standards that equipment performance rating experiments follow, such as AHRI Standard 210/240 (AHRI 2008) and ASHRAE Standard 37 (ASHRAE 2009). More details on the experimental approaches of the data included in the data library can be found in the following references: Breuker (1997), Kim et al. (2006), Shen et al. (2006), Palmiter et al. (2011), and Braun et al. (2012). Since the testing had different objectives, the data are dissimilar in many ways. For example, they include different types of faults, different operating conditions and different types of measurements. Some properties of the test data and the units tested are summarized in Table 2. In this table “RTU” refers to a rooftop unit, “FXO” refers to fixed orifice expansion devices, and “TXV” refers to thermostatic expansion valves.
The testing typically is conducted for units with faults imposed at various combinations of driving conditions (ambient temperature and indoor air humidity and temperature) and with no fault imposed at similar driving conditions. As part of the effort of gathering, organizing and analyzing the data we have generated a normal model for each unit – a model that uses driving conditions to predict coefficient of performance (COP) and capacity for the system when no faults are present. The general modeling approach is described by Brandemuehl (1993) and the specific approach is described in Yuill and Braun (2012) and Braun et al. (2012). The models were developed using multiple regression techniques, so they are only applied within the region of the unfaulted test conditions. These models and careful scrutiny of the test data were used to vet the data and eliminate tests in which there may have been experimental problems. For example, energy imbalances, refrigerant pressure increases (except across the compressor), and humidity increases all highlighted potential measurement problems. About 40% of the original test cases were eliminated because of apparent problems with the data, leaving the 607 resulting cases summarized in Table 2.

Since the normal model tells us what the performance would be under a set of conditions if no fault were present, it can be used to calculate the performance impact of a fault for a given fault test. This impact is quantified with the fault impact ratio (FIR), defined as:

\[
FIR_{\text{COP}} = \frac{\text{COP}_{\text{faulted}}}{\text{COP}_{\text{unfaulted}}} \quad FIR_{\text{capacity}} = \frac{\text{capacity}_{\text{faulted}}}{\text{capacity}_{\text{unfaulted}}}
\]  

(7)
for the impact on COP or capacity, respectively. The “faulted” value is measured and the “unfaulted” value comes from the model. FIR is proposed for standard use when describing the impact of faults. FIR can be expressed as a percentage.

**Determination of nominally correct refrigerant charge**

Manufacturers may specify a mass of charge that a unitary system should contain, or a charging procedure based on superheat or subcooling (superheat is the difference between the refrigerant suction temperature – point 5 on Figure 2 – and the saturation temperature at the suction pressure; subcooling is the temperature difference between the liquid line temperature – point 2 on Figure 2 – and the saturation temperature at the liquid line pressure). The system’s nominally correct charge level is typically based on maximizing capacity or efficiency at a particular operating condition, such as the common rating condition of 95°F ambient dry-bulb temperature, 80°F indoor dry-bulb temperature, and 67°F indoor wet-bulb temperature (often denoted 95/80/67). In order to have a consistent approach to defining the nominally correct charge in the input data library, the relationship between charge and COP has been studied at the 95/80/67 condition, and the charge level that gives the maximum COP at this condition has been defined as the correct charge level. In most cases this level agrees with the charge level defined by the experimenter as correct, but in a few cases the experimenter’s correct charge gave the maximum capacity, but not the maximum COP. These cases have been adjusted so that the input data library consistently follows the definition that the nominally correct charge gives the maximum COP at the 95/80/67 condition (or the nearest measured condition). Additional detail and plots of COP vs. charge are given in Braun et al. (2012).

**CASE STUDY: EVALUATION OF THE RCA PROTOCOL**

The evaluation method outlined in this paper is demonstrated here with an evaluation of a widely used protocol. The RCA protocol (RCA is an abbreviation of refrigerant charge and airflow) is specified in Appendix RA of California’s Title 24 – 2008 code of regulations (CEC 2008). This protocol is intended to detect and diagnose only refrigerant charge or evaporator airflow faults. It was developed for use in
residential systems but is also applied on commercial unitary equipment. There are four versions of this protocol. Two are in the current (2008) Title 24: the installer version and the HERS version. The installer version is for use by service personnel that have the capability to add or remove refrigerant charge. This is the version that is described and evaluated in the case study presented here. The HERS version is for the Home Energy Rating System (HERS) program, which rates a house’s energy efficiency. In Title 24 – 2013 (CEC 2012), which comes into effect in 2014, there are also installer and HERS versions. The 2013 protocols remove an option for diagnosis of low evaporator airflow using the temperature split across the evaporator and air handler, relying instead on more direct measurement of airflow to determine whether it is in an acceptable range. Braun et al. (2012) give a more detailed description of these four protocols, and the results of evaluating all four. Only the evaluation results for the 2008 installer version are given in the current paper. This version was chosen because it is currently in effect.

On systems with FXO expansion devices, the RCA uses return air dry bulb and wet bulb temperatures, supply air temperature, ambient air temperature and refrigerant superheat as inputs. Lookup tables are used to determine whether the superheat and temperature split (the change in airside condition across the indoor unit) are within an acceptable range. The protocol is first used to determine whether there is an evaporator airflow fault. The airflow fault diagnostic is only required if direct measurement of airflow is not conducted, but for this case study it is assumed that direct airflow measurement is not conducted. If there is no evaporator airflow fault, then the charge diagnostic can be applied.

A slightly modified approach is used for systems with a TXV expansion device. In this case subcooling is used, rather than superheat. The target subcooling value is obtained from the equipment manufacturer.

The inputs from the data library discussed above were fed to the RCA protocol. In cases where the manufacturer’s recommended subcooling target could not be obtained, the subcooling was averaged for all available unfaulted tests done under typical operating conditions and this average was used as a proxy for the manufacturer’s target.

As discussed above, the results are presented as percentages for four categories: No Response, False Alarm, Misdiagnosis and Missed Detection. In the latter three categories, results are presented as a function of fault impact ratio thresholds, using both FIR_{capacity} and FIR_{COP}. 
No Response

134 of the 607 tests (22%) gave no response. The reason is that the index ranges for the lookup tables in the protocol are limited, presumably because the protocol is assumed to perform poorly beyond these ranges. This result is difficult to interpret on its own; a higher No Response rate indicates a less useful protocol, but it is entirely dependent on the test case conditions. How well the test case conditions match the conditions under which the protocol will be deployed is unknown. However, comparing the rates for several protocols that have been evaluated will give potential FDD users a useful insight.

False Alarm

The False Alarm rates are presented as a function of FIR thresholds in Figure 3. The “threshold” is the FIR above which a system is not considered to have a fault, which is analogous to tolerating the fault. For example, at 95% FIR, the test cases in that category may have faults imposed, but the faults are trivial enough to cause performance reductions of 5% or less.

The 100% threshold includes only test cases with FIR of 100% or greater. Since the data are from experiments, they contain scatter, so that half of the unfaulted test cases can be expected to have FIR values above 100% and half below. Note that the number of test cases at each plotted point accumulate as one moves to the right across the graph; the 99.5% point includes all test cases above 100% and also those between 99.5 and 100%, and so on.
The numerical data that form the basis of Figure 3 are presented in Table 3.

**Figure 3: False alarm rates for RCA protocol as a function of FIR threshold**

These False Alarm results are surprisingly high. A False Alarm may be the most serious error an FDD protocol can make, since it could trigger service being done on a properly working system. Applying the results to an example: if a user wishes to tolerate faults that cause COP degradation of 2.5% or less (FIR\textsubscript{COP} threshold of 97.5%), he or she will receive False Alarms in about half of the test case scenarios (48%). The False Alarm rate stays fairly constant, which is surprising. It was expected that lower FIR thresholds (i.e.
larger-impact cases being considered unfaulted) would lead to increasing rates of False Alarm. Note that the False Alarm rate at the 100% FIR threshold is not 0%, since the 100% threshold includes all test cases for which the FIR is above 100%. If cases above 100% FIR were not included, then the 100% point would be undefined (0/0), since there are no cases with FIR between 100% and 100%.

**Misdiagnosis**

Misdiagnosis results are presented within FIR bins. This is slightly different than the presentation of False Alarm rates, which uses a threshold value to draw the line between what is considered faulted and unfaulted. For Misdiagnosis, all test cases fed to the protocol have had a fault deliberately imposed by the experimenter, so no such distinction is required. Figure 4 shows the results for Misdiagnosis (a). As discussed above, a Misdiagnosis (a) result is one in which the protocol diagnoses a fault for a test case that has a different fault imposed, regardless of whether the protocol is intended to diagnose the fault that is imposed. At the base of each bar the number of test cases in each bin is shown (this number is the denominator in the rate calculation).

The results for the 85-95% FIR bin are important because they represent a significant degradation in performance and are probably quite commonly occurring. In this bin roughly one-quarter of faults are misdiagnosed. In application, a Misdiagnosis can result in the wrong corrective action being taken by a service technician, which may cause greater negative impact than taking no action. The RCA results for Misdiagnosis (a) overall are poor, with a 26% Misdiagnosis rate aggregated across all FIR bins.

![Figure 4: Misdiagnosis (a) results for RCA protocol as a function of FIR category](image-url)
The Misdiagnosis (b) results, shown in Figure 5, use only test case inputs that have faults that the protocol is intended to diagnose. In the case of the RCA protocol, the faults are refrigerant charge and evaporator airflow. The Misdiagnosis (b) rates are significantly lower than the Misdiagnosis (a) results. This gives some insight into the workings of the protocol, but doesn’t reflect its overall utility because there are few, if any, scenarios in which it is known that no faults could exist except refrigerant charge or evaporator airflow.

![Figure 5: Misdiagnosis (b) results for RCA protocol as a function of FIR category](image)

**Missed Detection**

Missed Detections are cases in which a fault has been imposed but the protocol reports that there are no faults. As with the Misdiagnosis results, the Missed Detection results are divided into (a) and (b) categories and presented in FIR bins. Missed Detection could be considered the least serious error for a protocol to make, since it doesn’t result in unnecessary and potentially detrimental service. The results for Missed Detection (a) are shown in Figure 6, and for Missed Detection (b) in Figure 7.
Figure 6: Missed Detection (a) results for RCA protocol as a function of FIR category

The RCA misses a significant number of faults. Many of the faults with FIR above 100% are overcharge faults, which often increase capacity but are nevertheless undesirable because they increase the likelihood of compressor damage. In the 85-95\% FIR bin, where faults are fairly significant and likely, the RCA misses 30 to 40\% of the faults. Aggregated across all FIR the Missed Detection (a) rate is 33\%.

The Missed Detection (b) results are shown in Figure 7. The (b) results are more meaningful, generally, for Missed Detection than for Misdiagnosis because a user might deploy this protocol to check only for charge or evaporator airflow faults, and would want the protocol to ignore (i.e. miss detection of) other faults. In this scenario, the Missed Detection (b) results would be of greater interest. The aggregated Missed Detection (b) rate is 19\%.
**Figure 7: Missed Detection (b) results for RCA protocol as a function of FIR category**

**FUTURE WORK**

The case study demonstrates the evaluation method, and shows that the evaluated protocol does not perform very well. However, the results are not easily interpreted. How good or how bad, for example, is a 26% Misdiagnosis rate? Comparison from one protocol to the next will help clarify the results, but the authors are also working on developing performance metrics to more easily tie the results to cost-benefit analyses, such as the approach described by Li and Braun (2007b).

Another important issue is how well the distribution of fault types and fault intensities (fault prevalence) in the input cases match the distribution that would be found in the field. Fault prevalence is very difficult to determine accurately, and the authors are unaware of any reliable prevalence data for the faults considered in this evaluation approach. Some anecdotal evidence suggests that the distribution of faults in the data library, which has a heavy distribution of undercharge faults, provides a reasonable match to the field distribution, but reliable field studies are needed to know with any certainty how faults should be distributed in an FDD evaluation. Once this information is available, the models developed by Cheung and Braun (2013) can be used to generate an input library that matches the fault prevalence found in the field.
Finally, if a fixed set of input data is used for evaluation, developers of FDD protocols will be able to code their protocols to recognize these inputs and give the correct responses. This would allow the protocols to get perfect evaluation results regardless of their true performance. Since the methods described in this paper are expected to be used by regulatory bodies, there will be an incentive for such gaming of the evaluation. This is a further reason that model-generated input data are required for future evaluations. Models can be used to generate sets of inputs that are similar enough to provide fair and repeatable evaluations, but dissimilar enough that they can’t be learned.

CONCLUSIONS

Evaluating the performance of FDD protocols for air-cooled unitary air conditioners is an important task, but no standard evaluation methods previously existed. Data from laboratory measurements of systems operating with faults have been gathered, vetted and organized. A method of evaluating FDD protocols has been developed that is based on the performance impact of the faults. This approach was demonstrated in a case study of the RCA protocol of California’s Title 24. This protocol is found to perform poorly, flagging faults in up to 51% of the unfaulted cases, misdiagnosing 26% of cases with faults, and not detecting faults in 32% of the cases with faults present. However, it’s difficult to qualify how poor this performance is without comprehensive figures of merit or comparison to other protocols. Ongoing work will focus on development of such figures of merit, and will also include evaluations of other protocols.

Accurate and rapid simulation of unitary systems with and without faults imposed will be needed for improved evaluations. One reason is that the present library of experimental results may not accurately represent the true fault prevalence in the field. Another reason is that if the limited set of experimental results is used, it will be possible for the evaluation to be gamed, so that protocols can get perfect evaluations regardless of their true efficacy. Ongoing work to develop gray-box models, as described in Cheung and Braun (2013), will fill this need.

ACKNOWLEDGEMENTS
This work was supported by the National Institute of Standards and Technology (NIST) and by the New Buildings Institute (NBI) under funding from the California Energy Commission. We gratefully acknowledge project oversight by Drs. Piotr Domanski and Vance Payne of NIST, and by Mark Cherniack of NBI.

REFERENCES


