

2014

Evaluating Fault Detection and Diagnostics Tools with Simulations of Multiple Vapor Compression Systems

David P. Yuill

Purdue University, United States of America, dyuill@purdue.edu

Howard Cheung

Purdue University, United States of America, cheung@purdue.edu

James E. Braun

Purdue University, United States of America, jbraun@purdue.edu

Follow this and additional works at: <http://docs.lib.purdue.edu/iracc>

Yuill, David P.; Cheung, Howard; and Braun, James E., "Evaluating Fault Detection and Diagnostics Tools with Simulations of Multiple Vapor Compression Systems" (2014). *International Refrigeration and Air Conditioning Conference*. Paper 1543.
<http://docs.lib.purdue.edu/iracc/1543>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Complete proceedings may be acquired in print and on CD-ROM directly from the Ray W. Herrick Laboratories at <https://engineering.purdue.edu/Herrick/Events/orderlit.html>

Evaluation of Fault Detection and Diagnostics Tools by Simulation Results of Multiple Vapor Compression Systems

David P. YUILL*, Howard CHEUNG and James E. BRAUN

Ray W. Herrick Laboratories, Purdue University, Mechanical Engineering
West Lafayette, IN, USA
(765) 494-2132, dyuill@purdue.edu

* Corresponding Author

ABSTRACT

A methodology for evaluating the performance of fault detection and diagnostics (FDD) tools applied to unitary air conditioners has been developed by Yuill and Braun (2013). Data from faulted and unfaulted systems operating over a range of driving conditions are fed to the FDD tools, and the FDD responses are compared to the known operating conditions. The methodology originally relied upon experimental measurement data, but the amount of available data is limited, and evaluations can be far more meaningful if the operating conditions of the inputs can be controlled. Furthermore, a finite input set can be learned by an FDD algorithm, and the evaluation can be thereby gamed. To solve these problems, a large library of data from multiple systems under a wide range of conditions, with and without faults of varying magnitude, was generated with simulations from a novel gray-box modeling approach (Cheung and Braun 2013a, 2013b). The simulation outputs are being used to train neural network models, which can be coupled to software that executes the evaluation method. The neural network models are simpler than the semi-empirical approach, so they can produce evaluation inputs very quickly. This will facilitate the evaluator generating semi-random conditions to provide a unique set of evaluation data that are sufficiently accurate and numerous to provide repeatable results. Some evaluation results from three FDD protocols are used to demonstrate the advantages of simulation data over measurement data.

1. INTRODUCTION

Fault detection and diagnostics (FDD) tools are increasingly being used in air-conditioning applications, with particular emphasis on unitary equipment (air cooled direct-expansion systems, such as rooftop units and split systems). These tools are intended to *detect* abnormal operation in the system, and to *diagnose* the type or location of the fault. Unitary systems are a particularly good application for FDD because (a) they are believed to be fault-prone, due to their low-cost manufacture and limited maintenance; and (b) because they are so widely deployed. Most houses are cooled by unitary systems, and about 60% of cooling energy for commercial buildings is consumed by unitary systems (Feng *et al.*, 2005). Although the true prevalence of faults in the field is not well understood, it is estimated that finding and addressing all faults in current buildings could reduce total building energy consumption by 5-20% (Roth *et al.*, 2005). Thus, FDD has significant potential if widely deployed.

An important consideration when selecting an FDD tool is how effective the tool is. A methodology for evaluating the performance of FDD tools for unitary systems was developed by Yuill and Braun (2013). They define a taxonomy for FDD evaluation, and a specific method for characterizing performance by using results from a large number of tests of the FDD tool. Each test consists of pushing data from a specific operating scenario through the FDD tool, then observing and categorizing the response. An example of such a scenario is a specific air-conditioner with 80% of its nominal condenser airflow, operating in 86°F (30°C) ambient temperature, with 77°F (25°C) indoor air at 50% relative humidity. A data library with 607 scenarios and the corresponding data was collected by the

authors from experiments carried out by multiple researchers at multiple laboratories. These data represent tests on nine air-conditioners conducted under steady-state operation with either no fault, or one the following faults:

Table 1: Fault type in FDD evaluation input library

Fault type	Abbr.	Description
Under- or overcharge	UC, OC	A mass of refrigerant less or more than the correct mass
Low-side heat transfer	EA	Faults in the evaporator coil such as coil fouling or insufficient airflow
High-side heat transfer	CA	Faults in the condenser coil such as coil fouling or insufficient airflow
Liquid line restriction	LL	Flow restrictions such as crimps or fouled filter/drier in the liquid line
Non-condensables	NC	The presence of gases that do not condense (e.g. air) in the refrigerant
Compressor valve leakage	VL	Leaks in the compressor from high to low pressure regions

For each scenario in the data library that is fed to a candidate FDD tool, there are six possible outcomes:

1. No Response – the protocol is uncertain or unable to provide a response under the given conditions
2. Correct – the protocol correctly identifies the state of the system
3. False Alarm – the protocol indicates the presence of a fault when no significant fault is present
4. Misdiagnosis – a fault is detected on a system with a fault, but the wrong fault is diagnosed
5. Missed Detection – the protocol indicates no fault is present on a system with a fault
6. No Diagnosis – the protocol correctly detects the presence of a fault, but does not diagnose the fault type

These outcomes are organized according to the fault impact – the reduction in performance that is caused by the fault. The fault impact is characterized with a fault impact ratio (FIR) for the capacity and for the coefficient of performance (COP), as shown in Equation 1.

$$FIR_{COP} = \frac{COP_{faulted}}{COP_{unfaulted}} \quad FIR_{capacity} = \frac{capacity_{faulted}}{capacity_{unfaulted}} \quad (1)$$

Eight commonly used FDD protocols have been evaluated using this method. A case study showing the results of one of these is presented in Yuill & Braun (2013). In studying the results of the additional evaluations, some shortcomings of the method have been identified. The first is that FDD performance is often quite dependent on operating conditions and on fault level. Since the data library is derived from previous experiments, the conditions within the tests are not well distributed. For example, there is a particular emphasis on charge fault tests (216 of 494 fault tests), and a large concentration of tests at or near a common rating condition: 95/80/67°F (35/26.5/19.5°C) ($T_{amb}/T_{ra}/WB_{ra}$). The distribution of input data affects the protocols' performance differently.

A second shortcoming is that there are too few data in many FIR categories, so that the sample size in some categories gives an evaluation result that isn't meaningful. Finally, an additional problem with the measurement data library is that it could be learned. If the evaluation method is adopted as a performance standard for compliance with a code or standard, the potential exists for a developer of FDD tools to game the system by programming the FDD to output the correct diagnosis for each case where the input set is recognized.

To address these issues a gray-box modeling approach has been developed by Cheung and Braun (2013a, 2013b) to accurately simulate the performance of eight air-conditioners. The models are used to provide large sets of evenly-distributed data for use by the evaluation method. The models have been validated (Yuill *et al.* 2014), and have been used to generate a set of 14,074 data points. The eight protocols previously evaluated with measurement data have been evaluated again, using simulation data. The current paper summarizes the modeling approach, with a description of a specific approach used to find singular points in the dependent variables, and shows the results of evaluating three protocols using both measurement and simulation data, to illustrate why simulation data should be used. A future approach, currently in progress, for generating input data using a meta-model trained with the gray-box model's outputs is also described.

2. MODEL CONSTRUCTION

2.1 Gray-box model

The vapor compression system model in Cheung and Braun (2013a, 2013b) consists of several connected component models, as shown in Figure 1.

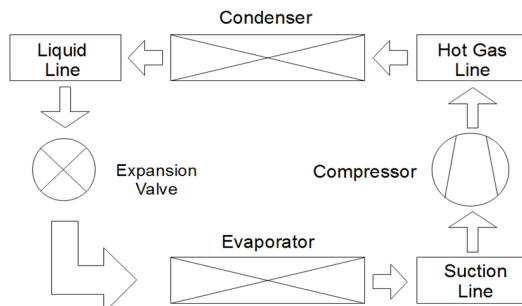


Figure 1: Schematic of component models in the gray-box system model

These gray-box component models are based on physical laws gathered from the literature, such as the polytropic compression model, heat transfer and pressure drop correlations, and heat exchanger effectiveness models, so that each model contains only a small number of empirical parameters, and can be easily trained with experimental data. The accuracy of the component models was discussed in Cheung and Braun (2013a).

The estimation of refrigerant mass in the system model deviates from the measurement (Cheung and Braun, 2013b). Shen *et al.* (2009) found that charge estimation by void fraction models always deviates from the measured charge levels because it cannot appropriately describe the effect of factors such as refrigerant dissolved in lubricating oil or maldistribution of refrigerant flow in the heat exchangers on the charge level. The deviation in the model was removed with the use of a charge tuning method. After tuning, fault impacts on system performance are modeled by applying the fault models shown in Table 2.

Table 2: Fault models

Fault type	Description
Under- or overcharge	Change the mass of charge in the system
Low-side heat transfer	Reduce the airflow across the evaporator
High-side heat transfer	Reduce the airflow across the condenser
Liquid line restriction	Reduce pressure at expansion valve inlet
Non-condensables	Use Dalton's law to estimate the refrigerant partial pressure
Compressor valve leakage	Add a compressor discharge bypass around the compressor

With these fault models, the gray-box model can simulate operation under user-defined environmental conditions and fault levels. To evaluate its accuracy and speed, simulations of eight different systems operating at their experimental conditions were conducted on a computer with a 2.7GHz CPU and 6GB memory. The details of the systems, the experimental conditions and the validation results were published in Cheung and Braun (2013a, 2013b). The average central processing unit (CPU) time required for each simulated operating condition is 28.8s, with a standard deviation of 84.4s. The standard deviation is much larger than the average because it takes a long time to solve the model at high fault levels.

To evaluate FDD tools under a wide variety of conditions, automated neural network (ANN) models will be used. The purpose for introducing this additional level of modeling is to reduce the amount of time required to generate inputs to the evaluation. If thousands of scenarios are needed and they require an average of 29 seconds, an evaluation would take many hours or days. One solution to this problem is to generate a large and static set of points for the evaluator to use. However, this set could be learned and pattern-recognition routines could be applied so that it would be possible to game the evaluator. But with an ANN model, the input conditions can be varied for each evaluation, and the ANNs have been found to run very quickly.

The size of the experimental data set is insufficient to train ANN models, but the gray-box models can provide sufficient data sets. The training data have been generated for the conditions listed in the test matrix in Table 3.

Table 3: Conditions in the test matrix

Variable	Values
Condenser air inlet temperature (T_{amb}) [°C]	18.3, 23.9, 29.4, 35, 40.6, 46.1
Evaporator air inlet temperature (T_{ra}) [°C]	21.1, 25, 28.9
Evaporator air inlet wet-bulb temperature (WB_{ra}) [°C]	12.8, 18.3, 23.9
Charge level [%]	70, 80, 90, 100, 110, 120, 130
Reduction of evaporator airflow [%]	0, 10, 25, 40, 55
Reduction of condenser airflow [%]	0, 10, 23, 37, 50, 60
Liquid line restriction level [%]	0, 50, 100, 300, 600, 1200, 2000, 3500
Non-condensable level [%]	0, 10, 30, 45, 80, 100
Compressor valve leakage level [%]	0, 10, 20, 35, 50

Only scenarios with a single fault (or no fault) are included, resulting in 1539 scenarios in the test matrix. However, to create accurate ANN models for system performance, additional data points in the test matrix are needed.

2.2 Generation of additional data points for ANN model construction

By design, vapor compression systems operate with positive subcooling (SC) and suction superheat (SH). However, when faults are present in these systems, SC or SH may fall. As they fall to zero, two-phase refrigerant enters either the compressor or the expansion valve. There is a point of non-linear system behavior as SC and SH reach zero (negative values of SC and SH are non-physical), referred to as a cusp. To predict system behavior near cusps accurately, the training data set for the ANN models must have data points at the cusps.

To simulate these cases with the gray-box models, the environmental conditions and fault levels that lead to zero SC and SH are needed. To find the cusp points of zero SH, for each environmental condition in the test matrix, the fault level leading to saturated refrigerant vapor at the compressor suction is calculated. A similar iterative solver is used to find the cusp points of zero SC by calculating the fault level leading to saturated refrigerant liquid at the expansion valve inlet. The flowchart of the iterative solver is shown in Figure 2.

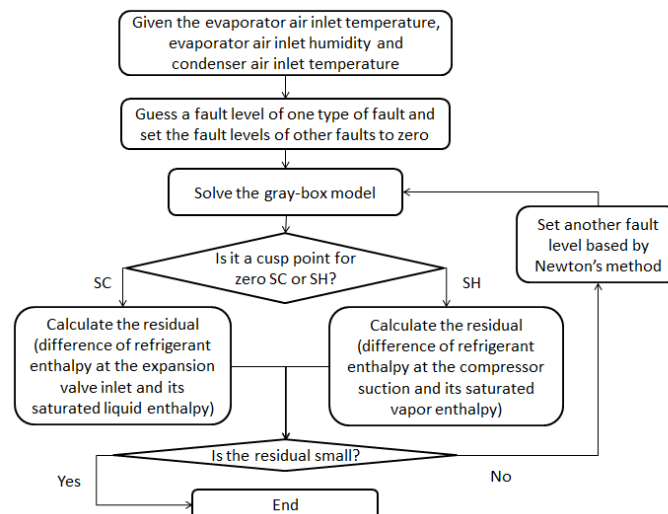


Figure 2: Iterative solver to find cusp points of zero SC or SH

To allow the ANN model to approximately fit a corner at the cusp points, the training data for the ANN model must also include simulation results from points near the cusp points. These points have the same fault levels as the cusp, but slightly varied environmental conditions. For instance, additional data points in Table 4 will be added to the test

matrix if a cusp point is found with the following condition: $T_{amb} = 18.3^{\circ}\text{C}$, $T_{ra} = 21.1^{\circ}\text{C}$, $WB_{ra} = 12.8^{\circ}\text{C}$ and valve leakage level at 25%.

Table 4: Example conditions near a cusp point to be added in the test matrix

Variable	Values
Condenser air inlet temperature [$^{\circ}\text{C}$]	18.3, 23.9
Evaporator air inlet temperature [$^{\circ}\text{C}$]	21.1, 25
Evaporator air inlet wet-bulb temperature [$^{\circ}\text{C}$]	12.8, 18.3
Charge level [%]	100
Reduction of evaporator airflow [%]	0
Reduction of condenser airflow [%]	0
Liquid line restriction level [%]	0
Non-condensable level [%]	0
Compressor valve leakage level [%]	25

The ANN model for each of the 8 unitary air conditioners will be trained using the set of 1539 conditions described in Table 3, along with a few hundred cusp and cusp-adjacent points.

2.3 Automated Neural Network models

ANNs have the advantage of being very fast, because they are purely inverse models. Typically, they have the disadvantage that they need a large and well-distributed set of training data, but in the current case, this set is available.

At the time of writing, ANNs have been generated and the approach has been found to be successful. The ANNs run quickly, and they have very good agreement with the gray-box model outputs. Comparison of ANN outputs with validation data sets (i.e. groups of data not used in training the ANN) give a coefficient of determination, R^2 , greater than 0.999 in all cases.

Some minor adjustments to the gray-box models have been made recently. The changes address a few cases in which outliers could be generated, and include an updated range of input scenarios required by the FDD evaluation method. Generation of ANNs for each dependent variable and for each modeled air-conditioner is currently in progress, using the updated gray-box models.

3. EVALUATION RESULTS USING EXPERIMENTAL AND SIMULATION DATA

3.1 Example FDD Protocols

There are FDD performance evaluation results for three FDD protocols presented in this section, identified as A, B and C. These three protocols are widely used. FDD A is part of California's energy code, Title 24 (CEC 2012), FDD B is from a commercial vendor, and FDD C was developed for use in utility-sponsored maintenance programs. These protocols use inputs such as SH, SC, T_{amb} , T_{ra} , WB_{ra} , condensing temperature, and equipment-specific information such as target SC, refrigerant type, system configuration, expansion device type, and rated SEER. Each requires that the unit be in steady operation before the protocol can be applied.

3.2 Evaluation methodology

The evaluation methodology was described in detail in Yuill and Braun (2013) and is summarized above. Improvements continue to be made as experience is gained evaluating protocols. These improvements will be described in more detail in this section.

False Alarm rates are calculated based upon threshold values of FIR. The threshold is the dividing line between what is considered faulted and what is considered unfaulted. This addresses the problem of whether a very small fault (a

1% reduction in airflow, for example) should be classified as a fault, even though it has no measureable impact on performance. It similarly allays concerns about defining the correct amount of refrigerant charge or airflow.

If a system has a fault imposed, and that fault causes a 4% reduction in FIR, the scenario is considered faulted at the 97.5% threshold, but unfaulted at the 95% threshold. A user can, therefore, choose the threshold she feels is significant, and use the evaluation results at that threshold.

A False Alarm requires that both $FIR_{capacity}$ and FIR_{COP} are above the threshold. It also requires that the refrigerant charge is less than 105% of nominal. Overcharged systems often give better performance, but risk damage to the compressor, so a user would want to know about an overcharge regardless of FIR. Aggregated rates of False Alarms are plotted at several FIR thresholds. The rates are calculated as the number actual false alarms at a given threshold divided by the number of potential false alarms at that threshold. Cases of No Response are not included in the rate calculation.

Misdiagnoses, Missed Detections, and No Diagnosis cases are presented in FIR bins. Since $FIR_{capacity}$ and FIR_{COP} often fall into different bins, they are plotted separately. Faults can often improve system performance under certain operating conditions, so FIR values up to 105% are considered. In the results presented below, only plots of FIR_{COP} are presented, due to space limitations.

3.3 False Alarm Results

Figure 3 and Figure 4 show the False Alarm rates for all three protocols from evaluations using measurement and simulation data, respectively.

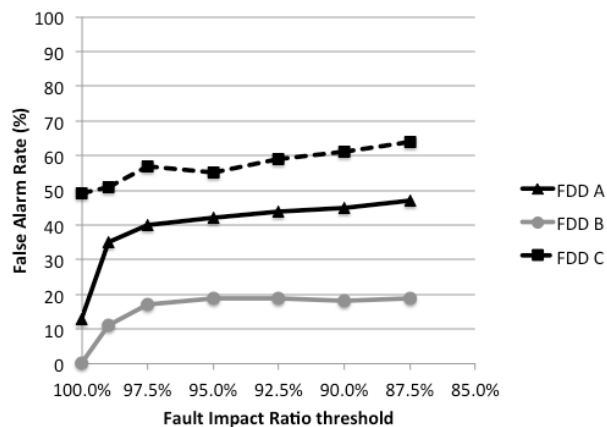


Figure 3: False Alarm rates from measurement data

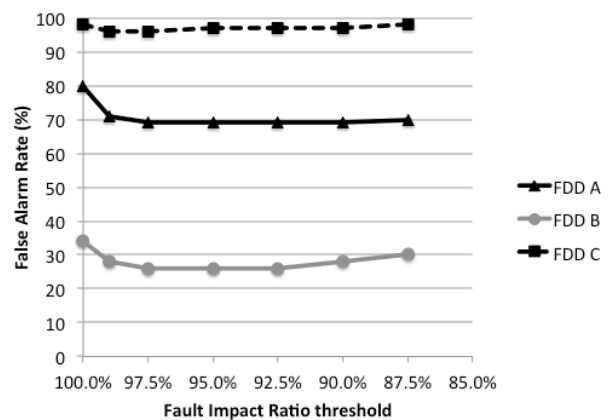


Figure 4: False Alarm rates from simulation data

The rates, overall, are surprisingly high. They are also surprisingly flat, particularly in the simulation-based results. One would expect that as the criterion for the fault level is relaxed (hence larger and larger faults are present in the nominally unfaulted set), the protocols would make increasingly more detections, so the rates increase.

The discrepancy between measurement-based and simulation-based results is also surprisingly large. The rank order of performance for the three protocols is preserved, and the comparative performance is also preserved (the improvement from C to A is slightly smaller than the improvement from A to B in both plots), but the absolute magnitudes are quite different. For FIR thresholds of 97.5% and below, the False Alarm rates in Figure 4 are 1.33 to 1.75 times the corresponding rates in Figure 3. FDD B is the best performer and is also least affected by the shift to simulation data, suggesting that it is more robust in terms of its performance under varying fault and operating conditions.

One explanation for the increased False Alarm rates with the simulation data set is that the measurement data set is heavily weighted with refrigerant charge fault tests, and the developers of the FDD protocols probably emphasized diagnosis of this particular fault. Another cause for the differences is the distribution of operating conditions included in the data sets. The simulation data are evenly distributed across a wide range of operating conditions,

whereas the measurement data are more clustered around the common rating condition 35/26.7/19.4 ($T_{amb}/T_{ra}/WB_{ra}$ in °C).

3.4 Misdiagnosis Results

Misdiagnosis rates are calculated separately for bins of $FIR_{capacity}$ and FIR_{COP} . In Figure 5 and Figure 6 only the FIR_{COP} rates are shown, because of limitations on paper length. In the measurement-based results, there are no results in the bin for $FIR_{COP} > 105\%$. This is because any bin with less than five potential Misdiagnosis cases is not shown, to avoid cluttering the plot with low-significance results.

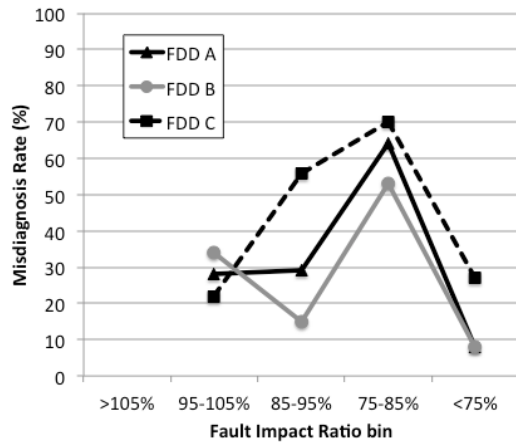


Figure 5: Misdiagnosis rates from measurement data

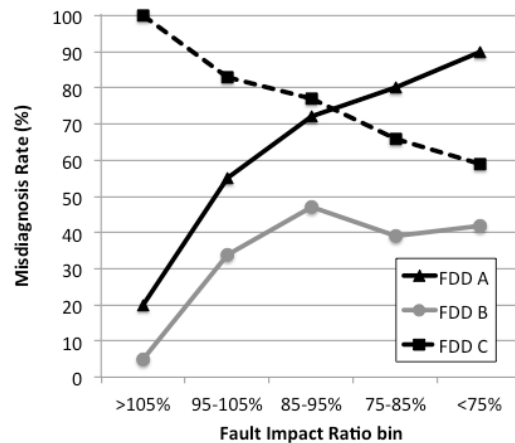


Figure 6: Misdiagnosis rates from simulation data

The lack of obvious trends in Figure 6 (where the fault type is evenly distributed) suggests that Misdiagnosis rate is not strongly correlated with fault level. This observation is consistent with other protocols that have been evaluated. In the measurement data set, Figure 5, most instances of $FIR < 75\%$ (severe faults) are caused by undercharge. As discussed above, these protocols tend to give a lot of undercharge diagnoses, so these are correct for most of the $FIR < 75\%$ bin in Figure 5. In the simulation set, however, they are subjected to a more evenly distributed range of fault types, so it's less common for undercharge to be the correct diagnosis.

3.5 Missed Detection Results

The Missed Detection rates, shown in Figure 7 and Figure 8, are presented in the same FIR_{COP} bins as Misdiagnosis rates. Again, results from any bins with less than five potential cases are not shown.

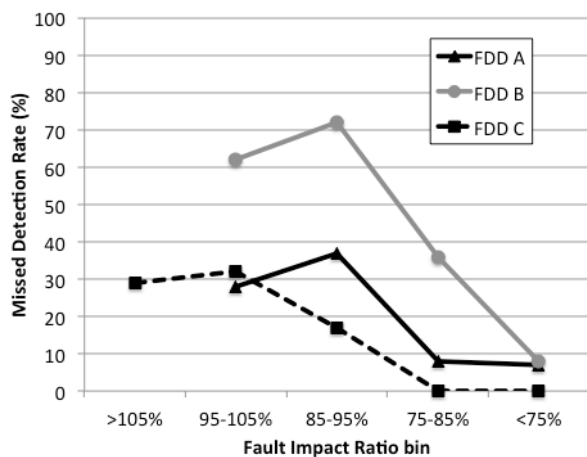


Figure 7: Missed Detection rates from measurement data

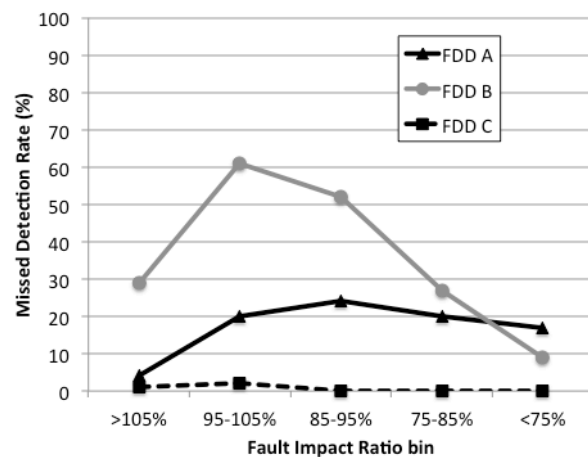


Figure 8: Missed Detection rates from simulation data

The measurement data results for all three protocols show a trend that we would expect: the more severe a fault is, the lower the rate of Missed Detection. This trend is also true in the simulation results in Figure 8, except that FDD A again behaves slightly differently because of its charge-diagnosis-only limitation. The rates for FDD C in Figure 8 are very low, suggesting an overly sensitive detection algorithm. Of 8726 responses from FDD C, only 108 assert that there's no fault present (99 of these actually are faulted cases, all of them falling in the three highest bins).

3.6 No Diagnosis Results

A No Diagnosis result is a case in which a fault is detected, but no diagnosis is provided. A No Diagnosis detection can be helpful, in that it can alert a technician to a problem so that further investigation can be made, but it is less helpful than a diagnosis, which would tell the technician what problem to address. FDD A gives a diagnosis for every case of a detected fault, so its No Diagnosis rate is always 0%.

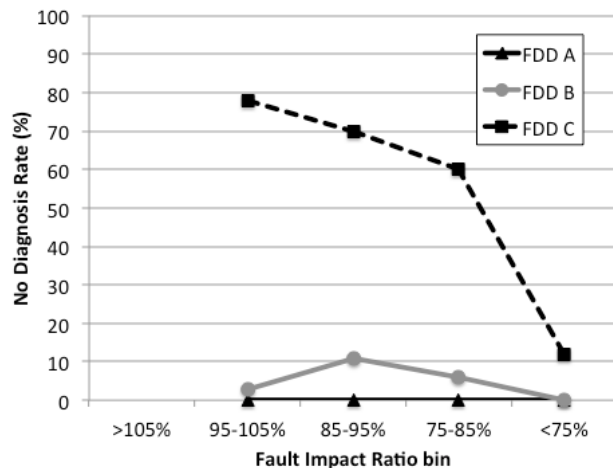


Figure 9: No Diagnosis rates from measurement data

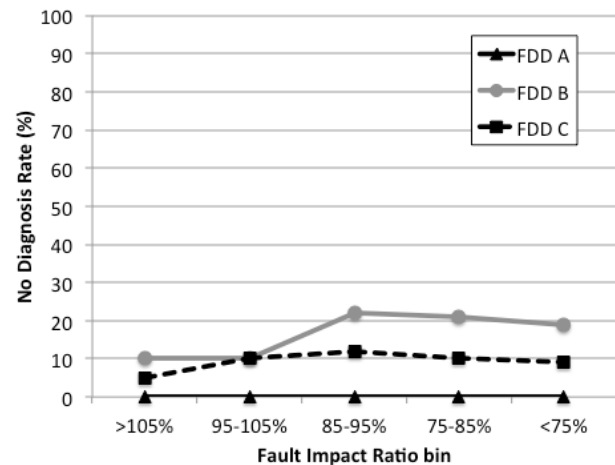


Figure 10: No Diagnosis rates from simulation data

Cases of No Diagnosis in rule-based diagnostic protocols typically stem from operating conditions that push the input variables outside of their expected boundaries. The reason for the significantly improved performance for FDD C with the simulation data as inputs (Figure 9) compared to the measurement data (Figure 10) is unknown. It may be a result of measurement error in one variable within the data, which is not present in the simulation data and which isn't used by the other protocols.

4. CONCLUSIONS

Gray-box models of eight air-conditioning systems have been developed to simulate performance with and without common faults. These models provide data for use in evaluating the performance of FDD protocols that might be deployed on the air-conditioning systems. The simulation data have several advantages over use of a measurement data library:

- The models can provide a large and wide-ranging set of input data, which can be used to train a faster ANN model. This ANN model can provide input data for evaluations in real-time, and with random input conditions. The gray-box model is specifically tailored to provide the necessary training points for consistently accurate performance of an ANN model.
- The models give a sufficiently large number of evaluation inputs to bolster the statistical significance of the results in each sub-category of the evaluation.
- Simulation results allow control over the distribution of inputs, including operating conditions, fault type and fault level. With a more even distribution of inputs, the evaluation results provide a fairer characterization of the true performance of an FDD protocols under a wide range of conditions.
- Control over conditions also facilitates more specific evaluation using a range of conditions of particular interest. For example, evaluations could be carried out using only selected fault types, selected equipment types, or using a narrower range of operating conditions.

NOMENCLATURE

ANN	automated neural network	
COP	coefficient of performance	(–)
CPU	central processing unit	
FIR	fault impact ratio	(–)
SC	subcooling	(°F or °C)
SEER	seasonal energy efficiency ratio	
SH	suction superheat	(°F or °C)
T	drybulb temperature	(°F or °C)
WB	wetbulb temperature	(°F or °C)

Subscript

amb	ambient air
ra	return air

REFERENCES

- California Energy Commission (CEC), 2012. Building Energy Efficiency Standards for residential and nonresidential Buildings. Title 24, Part 6. CEC-400-2012-004- CMF-REV2. California Energy Commission, Sacramento, CA.
- Cheung, H. and Braun, J.E., 2013a, Simulation of Fault Impacts for Vapor Compression Systems by Inverse Modeling Part I: Component Modeling and Validation, *HVAC&R Research*, vol. 19, no. 7: p. 892-906.
- Cheung, H. and Braun, J.E., 2013b, Simulation of Fault Impacts for Vapor Compression Systems by Inverse Modeling Part I: System Modeling and Validation, *HVAC&R Research*, vol. 19, no. 7: p. 907-921.
- Feng, M.Y, Roth, K.W., Westphalen, D. and Brodrick, J., 2005. Packaged Rooftop Units: Automated Fault Detection and Diagnostics. *ASHRAE Journal* vol. 47, no. 4: p. 68-70.
- Roth, K., Llana, P., Westphalen, D. and Brodrick, J., 2005, Automated Whole Building Diagnostics, *ASHRAE Journal*, vol. 47, no. 5: p. 82-84.
- Shen, B., Braun, J.E., Groll, E.A., 2009, Improved Methodologies for Simulating Unitary Air Conditioners at Off-design Conditions, *Int. J. Refrigeration*, vol. 32, p. 1837-1849.
- Yuill, D.P. and Braun, J.E., 2013, Evaluating the performance of FDD protocols applied to air-cooled unitary air-conditioning equipment. *HVAC&R Research*, vol. 19, no. 7: p. 882-891.
- Yuill, D.P., Cheung, H. and Braun, J.E., 2014, Validation of a Fault-Modeling Equipped Vapor Compression System Model Using a Fault Detection and Diagnostics Evaluation Tool. Paper 2606, *Proceedings of the 15th International Refrigeration and Air Conditioning Conference at Purdue, July 14-17*.

ACKNOWLEDGEMENT

This work was supported by the National Institute of Standards and Technology (NIST), Southern California Edison, 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.