An Open, Cloud-Based Platform for Whole-Building Fault Detection and Diagnostics

Stephen Frank, NREL
Jason Nichols, GE Global Research Center
Intelligent Building Operations Workshop
July 8, 2018
DOE AFDD Project Portfolio

- Active AFDD and Adaptive Controls

Model-Based, Whole-Building AFDD Platform & Algorithms (This Project)

AFDD Market Deployment & Algorithm Evaluation

Smaller Buildings

Larger Buildings

Controls

Analytics
Project Goal: Develop cost-effective automated fault detection and diagnosis algorithms for small commercial buildings
U.S. Small Commercial Buildings

Number of Buildings
- 72% ≤ 10,000 ft²
- 19% 10,001–50,000 ft²
- 20% > 50,000 ft²

Floor Area
- 19% ≤ 10,000 ft²
- 19% 10,001–50,000 ft²
- 20% > 50,000 ft²

Energy Use
- 72% ≤ 10,000 ft²
- 19% 10,001–50,000 ft²
- 20% > 50,000 ft²

Source: EIA (2012)
Automation Systems In Small Commercial Buildings

≤ 10,000 ft²

Source: EIA (2012)
### Small Buildings

**HVAC&R Equipment**

#### Cooling Equipment

<table>
<thead>
<tr>
<th>Equipment Type</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residential-type central AC</td>
<td>30%</td>
</tr>
<tr>
<td>Packaged AC units</td>
<td>28%</td>
</tr>
<tr>
<td>Individual AC units</td>
<td>12%</td>
</tr>
<tr>
<td>Heat pumps</td>
<td>12%</td>
</tr>
<tr>
<td>Swamp coolers</td>
<td>1%</td>
</tr>
<tr>
<td>Central chillers</td>
<td>NA*</td>
</tr>
<tr>
<td>District chilled water</td>
<td>NA</td>
</tr>
<tr>
<td>Other</td>
<td>NA</td>
</tr>
</tbody>
</table>

#### Heating Equipment

- **Packaged heating units**: 47%
- **Individual space heaters**: 22%
- **Furnaces**: 14%
- **Heat pumps**: 11%
- **Boilers**: 6%
- **District heat**: NA
- **Other**: NA

#### Refrigeration Equipment

- **Residential-type/compact units**: 58%
- **Cases or cabinets**: 14%
- **Walk-in units**: 12%
- **Commercial ice makers**: 11%
- **Vending machines**: 9%
- **Large cold storage areas**: NA

---

*NA values = No data or no estimate due to small sample size

**Source:** EIA (2012)
Net Present Value Analysis: 5,000 ft² Building

Annual Cost of Capital: 10% | Cost of Energy: $1.80/ft² | Annual Subscription Cost: 15% of Purchase Price

Source: Frank et al. (2018)
Research Approach

AUTOMATED MODELING & CALIBRATION

DETECTION

DIAGNOSIS

Prioritization

Fault Model Library

Measured

Fault

Predicted

Faults

+ 

= 

With Fault

Baseline

$
Case Study: ORNL FRP #2

Flexible Research Platform (2-Story)
Oak Ridge National Laboratory

- Dedicated experimental facility (not occupied)
- 3,200 ft²
- Resembles 80’s era office building
- RTU + VAV system (10 zone)

We modeled this facility in OpenStudio / E+
Baseline Results Highlight Challenges

AUC = 0.92

CDDR << 0.8
Automated AFDD Pipeline

Data Generation
- Observability

Feature Extraction
- Genetic Algorithm

Classifier Training
- Random Forest
Data Generation

Training Data
25 Fault Measures
6 climate zones (TMY3)
1 year

Testing Data
25 Fault Measures
1 climate zone (AMY3)
1 year

Poor Fault Observability
Good Fault Observability
Uninformative Sensor
Informative Sensor
Recommend that ground truth label of faulted state be based on observability criterion, rather than an energy/performance threshold.
Feature Extraction

Observable Faults

Sensor Selection by Genetic Algorithm

- Encode: 1001010110
- Keep Winners
- Mutate: 1101010110
- Fitness function: 
  \[- w_1 \times (CDDR) + w_2 \times (\text{# of sensors})\]
- Evaluate: Fitness Function

0 – Sensor not selected
1 – Sensor selected
### Multiclass Classification Challenge

#### Features

<table>
<thead>
<tr>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>room.204.zone_air_relative_humidity[%].mean</td>
</tr>
<tr>
<td>room.204.zone_air_relative_humidity[%].sum</td>
</tr>
<tr>
<td>rooftop_supply_fan_fan_electric_energy[WI].mean</td>
</tr>
<tr>
<td>rooftop_supply_fan_fan_electric_energy[WI].sum</td>
</tr>
<tr>
<td>rooftop_supply_fan_fan_electric_energy[WI].max</td>
</tr>
<tr>
<td>rooftop_supply_fan_fan_electric_energy[WI].std</td>
</tr>
<tr>
<td>rooftop_mixed_air_outlet_system_node_temperature[CL].mean</td>
</tr>
<tr>
<td>rooftop_mixed_air_outlet_system_node_temperature[CL].sum</td>
</tr>
<tr>
<td>rooftop_mixed_air_outlet_system_node_temperature[CL].std</td>
</tr>
<tr>
<td>rooftop_mixed_air_outlet_system_node_temperature[CL].min</td>
</tr>
<tr>
<td>room.201.zone_mean_air_temperature[CL].min</td>
</tr>
<tr>
<td>2f.plenum_zone_mean_air_temperature[CL].min</td>
</tr>
<tr>
<td>gas_facility[WI].sum</td>
</tr>
<tr>
<td>gas_facility[WI].max</td>
</tr>
<tr>
<td>gas_facility[WI].std</td>
</tr>
<tr>
<td>gas_facility[WI].mean</td>
</tr>
<tr>
<td>heating_gas[WI].mean</td>
</tr>
<tr>
<td>heating_gas[WI].max</td>
</tr>
<tr>
<td>heating_gas[WI].sum</td>
</tr>
<tr>
<td>heating_gas[WI].std</td>
</tr>
</tbody>
</table>

25 informative sensors
7 maximally informative sensors
20 input features
# Remaining Challenges & Paths Forward

<table>
<thead>
<tr>
<th>Goal</th>
<th>Challenge</th>
<th>Potential Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automated data generation</td>
<td>Generating data with enough load variation</td>
<td>Building load normalization methods</td>
</tr>
<tr>
<td>Improved diagnostic capability</td>
<td>Automatically generating informative features</td>
<td>Include feature types for:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Building hysteresis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Fault sensitivity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Cyclic pattern recognition</td>
</tr>
<tr>
<td>Fast feature generation</td>
<td>Compute for observability and genetic algorithm is intensive</td>
<td>• Code optimization</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Run-time parallelization</td>
</tr>
</tbody>
</table>
Thank you

www.nrel.gov

Evaluation: Protocol Outcomes

- **Input Samples**
  - Predicted as fault-free sample
    - False Negative
    - True Negative
  - Predicted as faulty sample
    - No Detection
    - False Positive
    - True Positive

- **Fault Detection**
  - Fault Diagnosis
    - Correct Diagnosis
    - Misdiagnosis
    - No Diagnosis
# Evaluation: Confusion Matrix

<table>
<thead>
<tr>
<th>Ground Truth (True Condition)</th>
<th>Protocol Output (Prediction Condition)</th>
<th>Prediction Condition Negative (no fault detected)</th>
<th>No Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Condition Positive</strong></td>
<td>Fault 1: Correct diagnosis (CD₁)</td>
<td>Misdiagnosis (MD₁₂)</td>
<td>False negative (FN₁)</td>
</tr>
<tr>
<td></td>
<td>Fault 2: Misdiagnosis (MD₂₁)</td>
<td>Correct diagnosis (CD₂)</td>
<td>False negative (FN₂)</td>
</tr>
<tr>
<td></td>
<td>Fault 3: Misdiagnosis (MD₃₁)</td>
<td>Misdiagnosis (MD₃₂)</td>
<td>False negative (FN₃)</td>
</tr>
<tr>
<td><strong>Condition Negative (No Fault)</strong></td>
<td>False positive (FP)</td>
<td>True negative (TN)</td>
<td>No detection negative (NDₙ)</td>
</tr>
</tbody>
</table>
Evaluation: Combined Detection & Diagnosis Rate (CDDR)

Combined Detection & Diagnosis Rate:Fault $i$

$$CDDR_{i} = \frac{CD_{i}}{CD_{i} + \sum_{j=1, i \neq j}^{N} MD_{i,j} + ND_{i} + FN_{i} + NDP_{i}}$$

Combined Detection & Diagnosis Rate: Overall

$$CDDR = \frac{TN + \sum_{i=1}^{N} CD_{i}}{\sum_{i=1}^{N} \left( CD_{i} + \sum_{j=1, i \neq j}^{N} MD_{i,j} + ND_{i} + FN_{i} + NDP_{i} \right)}$$