# AUTOMATED FAULT DETECTION AND DIAGNOSIS OF HVAC SUBSYSTEMS USING STATISTICAL MACHINE LEARNING

Samuel R. West<sup>1</sup>, Ying Guo<sup>2</sup>, X. Rosalind Wang<sup>2</sup>, & Joshua Wall<sup>1</sup> <sup>1</sup>CSIRO Energy Technology, Newcastle, Australia <sup>2</sup>CSIRO ICT Centre, Sydney, Australia

## ABSTRACT

The faulty operation of Heating Ventilation and Air Conditioning (HVAC) systems in commercial buildings can waste vast amounts of energy, cause unnecessary  $CO_2$  emissions and decrease occupant thermal comfort, reducing productivity.

We propose a new method of automating Fault Detection and Diagnosis (FDD), based on the modelling of operational faults in HVAC subsystems, using techniques from statistical machine learning and information theory. Discovery of interrelationships between groups of sensors by analysing the level of Information Transfer present can help fine tune the simulation inputs and improve model accuracy.

We present results of the detection and diagnosis of faults from an occupied commercial office building in Newcastle, Australia and using data from the ASHRAE 1020 fault detection project (Norford, Wright et al. 2000).

## **INTRODUCTION**

HVAC in Australian commercial buildings accounts for 84% of the sector's annual greenhouse gas emissions (AGO 2010). A range of case studies show that significant energy savings: 5-15% (Gregerson 1999); 2-20% (Claridge, Culp et al. 2000); 13% (Mills and Mathew 2009); can be obtained by diagnosing and repairing faults in HVAC systems.

Non-catastrophic faults can be very difficult for operators to detect due to the increasing complexity of Building Management Systems and the large number of sensors, set-points and zones in modern commercial buildings. Equipment failure and degradation often goes unnoticed until it results in a direct impact on occupant comfort, triggers an equipment-level alarm, or results in excessive energy consumption. The large amount of data available from modern Building Management Systems (BMSs) often captures the information necessary to detect and diagnose such failures, but it can be difficult and time consuming for even a domain expert to analyse manually, and in practice, this rarely happens.

Most existing fault detection or diagnosis techniques fall into one of the following categories; rule-based

threshold approaches; or physical building models (Katipamula and Brambley 2003).

Rule based approaches perform well for simple faults that can be detected with only a few sensors, but it is often difficult to set thresholds for good performance. Operators are often found to have adjusted the threshold level until the alarms are effectively disabled, defeating the advantage of having the system in place.

Although FDD systems based on physical models can be useful for complex HVAC subsystems, they are generally time consuming to configure, requiring detailed information - about equipment, components, building dimensions and construction materials - and are most often used for very specific subsystems and equipment models. These systems can be successful in detecting faults, however it is difficult to translate the system for use in other buildings or with unfamiliar equipment, once developed and configured, due to the low-level operation.

With advances in statistical machine learning and increasing storage and availability of real-time and historical BMS sensor data, it has become possible to develop advanced FDD approaches that can address the shortcomings of rule-based and physical modelling approaches while delivering high detection accuracy and ease of deployment to a wide range of new and existing building and equipment stock.

## **DETECTION AND DIAGNOSIS**

The ability to detect and diagnose faults in complex HVAC systems means that abrupt failures can be quickly remedied, gradual equipment degradation addressed and preventative equipment upgrades or servicing scheduled efficiently.

The presented technique makes use of Hidden Markov Models (HMMs) to simulate a network of time-varying interdependent relationships between sensors in an HVAC subsystem, such as in an air-handling unit (AHU).

To detect faults, a model is taught the inter-sensor relationships from historical data representing *normal* operation. It is then able to infer the likelihood that a stream of real-time data matches the learned historical behaviour; enabling detection of deviations from the norm.

Similarly, to diagnose faults, a model is trained using operational sensor data recorded during the *faulty* operation of a specific subsystem, such as a water valve, air damper, filter, heat exchanger, chiller, pump or fan. Incoming real-time data with a high likelihood of matching one or more of these fault models causes the system to notify the user that a fault has occurred, as well as identifying which model, or models, it most closely matches - giving a diagnosis. The data used for building fault models are recorded during experiments where the system responds to a range of induced fault conditions. These experiments are run in a separate test system in a lab, or online in the monitored building - by forcing control signals in the BMS to induce faulty operation after-hours, or for short periods during the working day where faults will not affect occupant comfort.

It is common for flow-on effects from a single fault to cause a number of fault models to trigger with varying degrees or certainty. Data fusion is employed to resolve the uncertainty in these conflicting diagnoses and determine the most likely cause of failure.

One deployment challenge is identifying which set of points is useful for detecting a particular type of fault. It is detrimental to prediction accuracy to omit a relevant sensor, and performance and accuracy are reduced by the inclusion of many irrelevant sensors.

To assist in the process of choosing which data to include we have used a technique from Information Theory that analyses the relationships between all sensors over a representative period of operation. This allows only the system components that have a measureable effect on other components to be included in each fault model, maximising the accuracy of the fault prediction.

Later we show results of these FDD algorithms from a large, operational commercial building in Newcastle, Australia, and using standard experimental fault data from the ASHRAE 1020 Air Handling Unit FDD project (Norford, Wright et al. 2000). We are able to identify relevant sensor combinations using information transfer theory, detect a wide range of faults in air-handling unit subsystems and HVAC system components with a high degree of accuracy, and successfully resolve conflicting detection results using data fusion.

## RESULTS

To save on computational resources, the first stage in the FDD process – detection – is performed using only a model trained to recognise normal operation of the overall system. We use 3 months of data at 5minute intervals (which is typical of the resolution of data from a BMS database), representing normal operation, to train an HMM model. This model is then tested on several weeks of new data in which were manually induced a number of faults. Faults were recreated by manually overriding actuators, physically interfering with the system, or purposely decalibrating sensors and de-tuning feedback loops. Faults that have been identified as having the highest impact on emissions, energy loss and occupant discomfort were given priority.

Two datasets were used for these results. The first is from an operational office building in Newcastle, Australia with 4 floors, 15 zones and about 100 occupants. A range of faults were created during spring and summer. The second data set was obtained from the ASHRAE-1020 FDD project, which recorded 162 separate sensor readings during faults generated in one of a pair of AHUs in a special test facility over four seasons in 2007. Fault types considered in one or both of these data sets include:

- 1) Exhaust air damper stuck
- 2) Return air fan fault
- 3) Cooling coil valve control fault
- 4) Cooling coil valve stuck
- 5) Heating coil valve leaking
- 6) Outside air damper leaking
- 7) Outside air damper stuck
- 8) Supply air filter fouled

For fault detection, several days of fault-free data were used to train an HMM before testing. Five days of normal data were used for the Newcastle dataset, and only two for the ASHRAE data, due to the limited number of fault-free days available. Data for the period containing each fault was then used for testing. A number of fault-free days were also tested to check for false positives.

Fault diagnosis models were trained using data from each fault run. Faults generated at Newcastle were repeated several times, and all except one used for training. Each repetition reproduced the fault using a range of parameters, to minimise the chance of overfitting. Testing was performed on the faulty data that was omitted from training. The ASHRAE dataset could, unfortunately, not be used for diagnosis testing as each fault was only produced once making it impossible to test on different data sets.

When recreating faults via setpoint manipulation (such as forcing a valve or damper open), data from these points was not used as an input to detection or diagnosis models because the real fault would not cause these changes.

**Figure 1** shows a typical output from the detection of a fault - a stuck hot water valve from a Newcastle AHU. The circled points during Day 2 correspond to the time of the induced fault, and the likelihood of matching the model of 'normality' is an order of magnitude lower than the fault free operation, making it trivial to detect the fault period.

Stuck hot water valves are a relatively common fault and lead to a large waste of energy as hot water leaks through the exchanger. In summer, this is a particularly nasty fault because the control system opens the chilled water valve to compensate for the extra heat added to the supply air by the leak, wasting even more energy but maintaining the supply air temperature at expected levels. Because of this, these faults often goes unnoticed by occupants (who would otherwise be the most direct form of feedback), and is usually only noticed in the next energy bill, if at all.

**Figure 2** illustrates the detection model testing for a fault in the ASHRAE dataset. The fault was induced by reducing the proportional band of the PID controller to cause large fluctuations in the cooling coil water valve. This recreates the response from a control loop that is badly tuned during commissioning or maintenance. The lower chart shows the affected sensors - the chilled water valve (*CHWV*), hot water valve (*HWV*) and supply air temperature (*SATemp*) - and the fault which begins just before 15:00. The upper chart shows the output from the detection process after the clustering has been run. The samples with a class ID of 2 have fallen into the second, abnormal cluster, indicating a fault.

Initially only the detection model, trained on normal fault-free historical data, is used for detection of faults. This is done so that not all specific fault diagnosis models need to be run on every sample of incoming data, and computational resources are freed accordingly. If a fault is detected by this *normality model*, the more specific *fault models* are run against the data to determine what sub-groups of components are causing the fault. If a single fault model determines, with high likelihood, that its learnt relationships match the incoming data, the fault is considered diagnosed. Because fault models only use data from a small number of targetted components, identifying the model is tantamount to identifying the cause of the faulty behaviour.

If multiple models match the data, we use Dempster-Shafer data fusion (see **Figure 3**) to resolve the uncertainty in the diagnosis, based on pre-trained knowledge of the performance of each fault model under similar conditions. The final output of the system is a single fault diagnosis which indicates that the modelled fault has occurred, and which piece of equipment is responsible.

**Figure 4** shows the diagnosis result from testing a network of 9 sensors from an AHU during a fault where the outside air damper was forced open. This recreates a common fault where the damper becomes

stuck open and the AHU uses unconditioned air unecessarily. This experiment was performed in Newcastle during summer when the AHU was in cooling mode and the outside air relatively warm, meaning that the stuck open damper had a detrimental effect on supply air cooling and so the control loop opened the chilled water valve further to maintain the cool air temperature. The AHU supply air fan power is also charted for reference.

The diagnosis model for this experiment has been trained on data from previous experiments where the same fault was created with a range of severity levels and operating modes. The trained model was then tested by forcing several similar, but previously unseen faults in the outside air damper position. The fault, occurring after 12:00 is reliably detected at the times indicated by the red dots at the top of the chart. The two fault indications prior to 12:00 are false positives, probably caused by sharp changes in the data at these times. Our experimental results show that these false positives generally occur only for short periods of time, and are thus simple to distinguish from the real faults, which tend to persist for longer.

One difficulty with the Hidden Markov Model training algorithm, which learns the probabilistic associations between sets of sensors, is that the process is non-deterministic, and is prone to getting 'stuck' in local optima. This can result in the magnitudes of the tested likelihood curves differing between models trained on the same historical data, making it difficult to determine a robust threshold for identifying faults. To overcome this, we train a number of models with randomised sets of starting weights, discarding those that do not converge, and test new data with each trained model to produce a number of likelihood curves, as seen in Figure 5. Each point on this set of curves is then clustered by distance with a simple K-mean algorithm. То determine how many clusters in the datasets, we used the silhouette metric. The silhouette of a datum is a measure of how closely it is matched to data within its cluster and how loosely it is matched to data of the neighbouring cluster. Intuitively, if no fault is present in the tested data, the silhouette is low a single cluster results. If one or more faults are present, there is a high silhouette and two or more clusters result, with the majority of points falling into the 'normal' cluster - allowing the faulty data from other clusters to be easily identified and appropriate action taken.



Figure 1- Detection results for the stuck hot water valve fault using the Newcastle dataset.



Figure 2– Fault detection results for Cooling coil valve control unstable fault using the ASHRAE dataset.



*Figure 3 – Using Data Fusion to resolve and improve multiple conflicting diagnoses.* 

#### SENSOR SELECTION

Currently, a domain expert is responsible for identifying which sensors and actuators within the HVAC system should be considered by a given fault or normality model. For example, an expert will know from experience that a drifting supply-air pressure sensor can be detected by looking at the supply air fan speed and setpoint, the fan power, and the outside and return air duct positions - so we build a Hidden Markov Model to represent the flow of information between these sensors.

For the system to be easily deployed into an new building with a unique design, layout and set of equipment, the process of determining these relationships can become difficult. Mapping physical knowledge of the system functionality to a set of point names in the BMS database can also be tedious and error-prone thanks to the dramatic variation present in data precision, update rates, protocols, naming conventions and sensor availability. To overcome these difficulties, we employ a method for automatically determining the directional flow of information between sets of sensors. This analysis technique from Information Theory measures the Local Transfer Entropy (Lizier, Prokopenko et al. 2008) between the source and destination. This measures how much information about the next state of the destination is determined by the current state of the source but is not contained in the destination's past states (Wang, Lizier et al. 2011). Information transfer is not sufficient to directly determine causal effect, but is a useful metric to suggest some level of causality, direct or otherwise (Lizier and Prokopenko. 2010).

Results were generated by analysing a month of data, recorded at 5 second intervals, from 11 sensors in a single AHU and calculating the apparent Local Transfer Entropy between every pair of points in Figure 6 shows a graphical either direction. representation of these results; high values show strong transfer entropy from the x-axis point (source) and the y-axis point (destination) indicating a probable causal relationship. For example, the highest level of transfer discovered in the data indicates that fan power (sensor 2) depends strongly on the fan speed setpoint (7). This is an obvious relationship, and it makes sense that there is the highest transfer entropy between these two points, as the fan power is directly determined the fastest responding control loop considered by these sensors.

It can also be seen that the position of the chilled water valve (10) depends, weakly, on the average zone temperature (11). Again, this makes sense intuitively, as the chilled water valve position *is* determined by the zone temperature; a warm zone requires more cooling. The relationship is not strong

though, because other factors contribute to the valve's position (such as the supply air temperature), causing more frequent valve changes, so this explains the lower degree of information transfer between these two points.

A third and final example is the strong apparent dependency of fan power (2) on the outside air damper (3). At first glance this is not an obvious relationship appears, but further experimentation showed that a dependency did actually exist; the air return path distance from the zone is much longer when the outside air damper is closed, forcing the fan to consume more power sucking the air back to maintain supply duct pressure. However, the power level decreases when the damper opens, as the air path becomes much shorter, so the air-resistance is decreased and less fan power needs to be consumed.

Analysis by domain experts verifies that the other relationships shown in **Figure 6** also make sense with respect to this system's normal operation, as do the pairs of non-related points and the indicated direction of information transfer, showing this to be a useful technique for point relationship discovery in such sensor networks. The discovery of non-obvious relationships adds additional weight to the usefulness of Local Transfer Entropy analysis for fault detection.

#### CONCLUSION

We have presented a novel system capable of automating detection and diagnosis of faults in commercial building HVAC systems.

This system is able to detect faults accurately and robustly and in real-time using data from an operational building in Newcastle, Australia, and in a standard ASHRAE 1020 project's FDD dataset. A new FDD technique employing Hidden Markov Models has been developed to learn probabilistic relationships between groups of points during both normal and faulty operation. This can passively infer the likelihood of similar patterns in the data during future operation with a high degree of accuracy.

Multiple parallel models and clustering were used to overcome issues with training state being stuck in local optima, and Data Fusion was employed to resolve conflicting diagnoses from multiple related models.

A novel Local Transfer Entropy was used to confirm and discover relationships between pairs of points, and the process was successfully demonstrated using data from an operating AHU. The discovered correlations between sensors built upon existing knowledge of the system, as well as confirming previous assumptions.

Successful operation has been demonstrated for a number of fault types on a real building and using data from a specialised test facility.



Figure 4 – Diagnosis results for a stuck outside air damper fault using the Newcastle dataset.



Figure 5- A family of likelihood curves from a detected fault prior to clustering.



Figure 6 – The Apparent Transfer Entropy between pairs of points from an Air Handling Unit.

#### ACKNOWLEDGEMENT

This work was undertaken with the support of the CSIRO Energy Transformed Flagship and with help from staff at the Division of Energy Technology, Newcastle.

Thanks to Daniel Rowe for reviewing assistance.

#### **REFERENCES**

- AGO (2010). Australian Commercial Building Sector Greenhouse Gas Emissions 1990-2010. Canberra, ACT, Australian Greenhouse Office.
- Claridge, D., C. Culp, et al. (2000). <u>Campus-wide</u> <u>Continuous Commissioning of University</u> <u>Buildings</u>. ACEEE 2000 Summer Study on Energy Efficiency in Buildings, Washington D.C.
- Gregerson, J. (1999). "Commissioning Existing Buildings." <u>E Source Tech Update TU-97-3</u>.
- Katipamula, S. and M. R. Brambley (2003).
  "Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems A Review, Part II." <u>HVAC&R RESEARCH</u> 11(2): 169-187.
- Lizier, J. T., M. Prokopenko, et al. (2008). "Local information transfer as a spatiotemporal filter for complex systems. ." <u>Phys. Rev. E.</u> **77**(2).
- Lizier, J. T. and M. Prokopenko. (2010). "Differentiating information transfer and causal effect. ." <u>The European Physical</u> <u>Journal B</u> **73**(4): 605-615.
- Mills, E., H. Friedman, et al. (2004). The Costeffectiveness of Commercial-Buildings Commissioning - A Meta-Analysis of Energy and Non-Energy Impacts in Existing Buildings andNew Construction in the United States, Lawrence Berkeley National Laboratory. LBNL – 56637.
- Mills, E. and P. Mathew (2009). Monitoring-Based Commissioning: Benchmarking Analysis of 24 UC/CSU/IOU Projects. <u>Report number</u> <u>1972E.</u>, Lawrence Berkeley National Laboratory

- Norford, L. K., J. A. Wright, et al. (2000). ASHRAE Research Project 1020-RP - Demonstration of Fault Detection and Diagnosis Methods in a Real Building.
- Wang, R., J. T. Lizier, et al. (2011). Information Transfer Analysis for Ongoing Commissioning of Buildings. <u>Autonomous</u> <u>Agents and Multiagent Systems</u>. Taipei, Taiwan.