Bugs in the Freezer: Detecting Faults in Supermarket Refrigeration Systems Using Energy Signals

Shravan Srinivasan, Arunchandar Vasan, Venkatesh Sarangan
Innovation Labs – Chennai
Tata Consultancy Services
IIT Madras Research Park, Chennai, India
s.shravan1@tcs.com

Anand Sivasubramaniam
Dept. of Comp. Sci. & Eng.
Pennsylvania State University
University Park, PA 16802, USA
anand@cse.psu.edu

ABSTRACT

Refrigeration is a major component of supermarket energy consumption. Ensuring faultless operation of refrigeration systems is essential from both economic and sustainability perspectives. Present day industry practises of monitoring refrigeration systems to detect operational anomalies have several drawbacks: (i) Over-dependence on human skills; (ii) Limited help in identifying the root-cause of the anomaly; and (iii) Presumption about high degree of instrumentation – which prevents their usage in supermarkets in developing economies. Existing approaches in literature to detect anomalies in refrigeration systems either are done in controlled laboratory settings or assume the availability of sensory information other than energy. In this paper, we present an approach to detect anomalous behavior in the operation of refrigeration systems by monitoring their energy signals alone. We test the performance of our approach using data collected from refrigeration systems across 25 stores of a real world supermarket chain. We find that using energy signal, we can not only detect anomalies but also narrow down the possible root-cause of the anomaly to a reduced set. Further, using energy signal along with data collected from other sensors (if available) allows us to reduce the false positive rate while identifying the root-cause of the anomaly.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Time series analysis; J.2 [**Physical Sciences and Engineering**]: Engineering

Keywords

Refrigeration; Energy; Anomaly detection; Root-cause analysis

1. INTRODUCTION

Refrigeration in supermarkets: The annual energy intensity of supermarkets is around 50 kWh per sqft., which is more than twice the intensity of office buildings [2]. Reducing electricity costs by 10% can boost the (typically thin) profit margins of supermarkets by 16% [15]. Therefore, it is important to discover ways to reduce energy consumption of supermarkets. Refrigeration is a key contributor to supermarket energy consumption. Refrigerated systems (RS) are used to maintain produce, dairy, and meat that are perishable at a temperature well below the HVAC conditioned temperature. Unlike space cooling, RS consumes significant energy even during winter; for instance, frozen food needs to be maintained typically at $-6^{\circ}F$ even when the space (i.e., indoor ambient) temperature around refrigeration cases is maintained $70^{\circ}F$ in winter. Thus energy consumed by RS can even surpass that of the HVAC systems for space cooling of the overall supermarket [1, 8]. Consequently, any optimization of supermarket electricity consumption necessarily needs to look at RS and that too throughout the year.

RS are subject to maintenance routines scheduled at regular intervals to address any wear and tear. As in any other physical asset, despite regular maintenance faults do occur. If a fault is critical, the entire RS can become unfit for use leading to a complete shut down necessitating an emergency maintenance. If the fault is non-critical, it will degrade the performance of the overall system while allowing it to be functional. An unattended non-critical fault may affect the energy consumption and can become a future critical fault.

Current maintenance practices and challenges: Refrigeration systems in large supermarket chains are increasingly being equipped with sensors to monitor various parameters. This sensory data is relayed to a centralized remote monitoring center. In these centers, human operators analyze the data either online or on demand basis when the sensor values exceed preset thresholds.

This present model of remote monitoring has several drawbacks: (i) It relies heavily on the ability of human operators to detect anomalies. If the operators do not have the necessary skill-sets, detection of anomalous behavior can get delayed, or even worse not happen at all; (ii) Typically, the operators pick up only the symptoms to trigger a maintenance work-order. It is up to the visiting technician to do the root-cause analysis on the ground and do the necessary fixes. Any guidance that can be given to help the technician's diagnostic process will reduce the time and hence the costs associated with the fix. (iii) It assumes the existence of a high degree of instrumentation. Refrigeration assets in small convenience stores and supermarkets in developing economies do not have such a high degree of instrumentation. At best, they may have only energy meters for the refrigeration assets. Hence, the present remote monitoring model cannot be used for these stores.

Given these drawbacks in terms of limited instrumentation and the need to reduce the time taken to identify anomalies even in such a scenario, the following questions arise:

- Is it possible to detect anomalous behavior in refrigeration systems by monitoring their energy signals alone (which are likely to be available only at coarse temporal resolutions)?

e-Energy'15, July 14–17, 2015, Bangalore, India.

DOI: http://dx.doi.org/10.1145/2768510.2768536.

- To what extent would it be possible to identify the possible components or reasons for the anomalous behavior using the energy signal alone?

- If additional sensors are available, can these be used in conjunction with the energy signal to detect faults and identify the root-cause automatically?

Existing models for detecting anomalies in refrigeration systems and diagnosing the root causes have mostly been done in controlled laboratory settings [3, 6, 7, 11–14, 17]. These may not be applicable in real world scenarios since the operating conditions may be different from a controlled environment. Consequently, the rules derived in laboratory may not be applicable. Further, not all sensors required by the laboratory models may be available in practise. There are also works which attempt to detect anomalies in refrigeration cases from the energy consumption signal [4]. However, these approaches require information pertaining to indoor and outdoor temperatures, indoor and outdoor humidity, and the loads imposed by the occupants to develop their energy models – all of which may not be available in all the stores. Further, the extent to which the root cause of the problem can be diagnosed has not been analyzed in these works.

In this paper, we attempt to address the above mentioned questions using data collected from refrigeration systems in real-world supermarket stores. Addressing these questions is non-trivial for the following reasons: (i) Typically, energy consumption of refrigeration units is measured at an aggregate level – i.e., several compressors and condenser units are grouped together in racks and it is this rack consumption that gets measured. In addition, this rack of compressors and condensers serve several refrigeration cases. Hence, a fault in one of the cases or compressor/condenser units may not manifest in the measured aggregate energy signal. (ii) Since energy consumption results from the aggregate performance of all refrigeration components, a single component's anomaly may not affect the overall energy significantly enough (or uniquely enough) to be traced back.

We overcome the challenges through **these contributions**.

- We develop an energy model that is sensitive enough to pick up even short-lived anomalies in the energy signal so that we maximize our likelihood to detect an anomaly. Simultaneously, our model also allows us place a bound on the false positives. We do this by integrating a statistical model and a physical model thereby overcoming the limitations of both.

- We use the direction of the anomaly as a signature to identify the anomaly root-cause as a group of components (rather than an individual component).

- If additional sensors are made available, we develop statistical models for the sensor data. By leveraging the anomalies detected by these statistical data models, we identify the individual component that is likely to be at fault.

These models have been developed and tested on real world sensory data and work-order logs obtained over a period of five months from refrigeration systems deployed across 25 stores from a supermarket chain[1]. The data was collected over every 15 minutes and included the energy consumption of the refrigeration systems and sensory information wherever available. In this period, we observed four different work-orders occurring across these 25 stores, at times even repeatedly. Our findings from this study include:

- The classification rules based on anomalies in the energy signal can detect between $80 - 95\%$ of all anomalies. The false positive rates in most stores is about 0.2%.
- Detecting anomalies using a set of sensors can classify individual anomaly types with a detection likelihood between $66 - 100\%$ across all types. The false positive ratio is again bounded by 0.2%.
- We note that if the supplementary sensors alone are used for detecting faults in individual components, the resulting false positive rate is higher than what would be possible when using the sensors along with the overall energy signal. This is due to the correlated nature of the energy and individual sensory signals.
- For the detected anomalies, the average detection time is $2.8 - 21$ days before the work-order is actually logged in the system. The model is able to pick up steady deterioration from the baseline over a period of time, well before the human operators notice the symptoms and log it.

To the best of our knowledge, ours is one of the first few papers to report on detecting and identifying faults in supermarket refrigeration systems using real-world data logs from an ensemble of stores. The rest of this paper is organized as follows. Section 2 presents a survey of related work. A background on supermarket refrigeration is presented in 3. We develop statistical and physical models for the energy consumption in Section 4. Section 5 gives a short introduction to the faults we had observed in the real world stores. Using the energy model as the baseline we predict if an observed energy sample can be classified as anomalous and if so of what type in Section 6. Section 7 discusses how additional sensor based information can help in reducing false positive rates of anomaly detection and identify the sub-type of the anomalies. Section 8 concludes the paper.

2. RELATED WORK

Work related to ours can be broadly categorized into two groups.
Fault diagnosis in refrigeration and air-conditioning systems: Works on detecting anomalies in refrigeration systems and diagnosing the root causes [3, 5–7, 12, 17] typically use simple sensor value based thresholding or techniques such as PCA and SVM to detect the anomalies. However, these have mostly been done in laboratory settings or as controlled one-off experiments. In real-world scenarios, the operating conditions and availability of sensors may be different from a controlled environment. Consequently, the rules derived in laboratory may not always be applicable in real world.

Some works attempt to detect anomalies (but not identify the root-cause) in refrigeration cases from the energy consumption signal [4]. However, unlike us, these approaches require additional information pertaining to indoor and outdoor temperature, indoor and outdoor humidity, and loads imposed by the occupants to develop the energy model – all of which may not be available in all the stores. Further, the extent to which the root cause of the problem can be diagnosed has not been quantified in these works.

Our work complements existing works in that we use time-series models such as ARIMA/SARIMA (which have not been reported earlier) with reasonable success to detect anomalies and identify the root-cause in refrigeration systems. Further, our approach works with limited existing sensor and metering infrastructure.

Domain model based approaches to detect faults and identify the root-causes have also been suggested in the literature [11, 14]. These approaches rely on the ability to develop a well-calibrated domain model that mimics the operations of a real world refrigeration system. While such an approach can be quite powerful,

[1]The identity of the chain is withheld due to confidentiality requirements.

the challenge is to calibrate the model to accurately reflect the real world behavior – this can become unwieldy due to two reasons: (i) the set of refrigeration systems across various stores can be different necessitating not one but several domain models; (ii) these models need to be re-calibrated periodically to keep pace with component aging.

Fault diagnosis in HVAC systems and operations: Though the underlying principle of refrigeration remains the same as that of a refrigerator/simple air-conditioner, centralized HVAC systems have a larger set of components and involve multiple heat exchangers. There have been several works that focus on detecting faults not only in HVAC systems but also in their operations (or control settings). Narayanaswamy et. al. [9] adopt a data-driven approach to detect faults in HVAC usage. Their focus is on detecting faults in variable air volume (VAV) control settings and use parameters which are very HVAC specific. Similarly, Zhou et. al [18] propose a regression based approach based on HVAC specific parameters to detect faults in HVAC sub-systems. On the other hand, the focus of our paper is on refrigeration systems and we use a different set of parameters. Consequently, the methodologies and classification rules proposed using HVAC specific parameters may not be applicable in our context. Reference [13] discusses about detecting refrigerant leaks in large chillers using artificial neural networks. They do a simulation based study and use temperatures gathered at various points in the refrigeration loop as feature vectors. Unlike them, we use the liquid level in the receiver as a feature, which is relatively simpler, to detect leaks and test their efficacy in real world systems. We also focus on identifying other kinds of faults.

Some work has been done to demonstrate that anomalous behavior of window air-conditioners in buildings can be detected from the aggregate energy consumption (more precisely, the current drawn) [10]. The anomaly is detected using the presence of high frequency current spikes drawn by a faulty air-conditioner. Consequently, these works assume that the energy/current signal is sampled at a finer temporal resolution (every few seconds). However, the energy data in real-world stores is typically available at a coarser granularity and hence such methods may not be applicable.

3. BACKGROUND

The refrigeration load in a typical supermarket consists of display cases that house beverages, produce, or perishable items in a refrigerated environment. These cases are typically at various internal temperatures depending on the kind of product: (1) Beverage coolers and dairy cases are typically maintained at around $36^\circ F$ - we refer to these as **coolers**; and 2) Frozen food needs to be maintained at around $-6^\circ F$ - we refer to these cases as **freezers.** Because there is an economy of scale in cooling systems (in terms of watts required to remove a unit heat-load), the heat loads of multiple display cases are typically aggregated into one compressor/condenser system.

3.1 System setup and operation

Figure 1 shows the configuration of a typical supermarket RS. It consists of a compressor rack and a condenser rack. A compressor/condenser rack may have one or more units of same or different capacities with a common inlet and outlet. Each compressor rack and its condenser unit serves multiple display cases where the set-point temperature is identical or similar. Typically, all freezer cases are served by one compressor-condenser system and all cooler cases are served by a different compressor-condenser system. Each display case has an individual evaporator with an evaporator coil through which cold refrigerant flows through picking up heat inside the case. The cases have fans that blow air across the cold evaporator coil to enhance heat exchange.

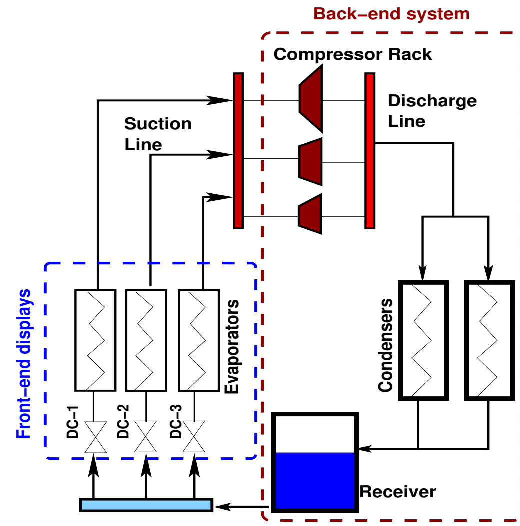


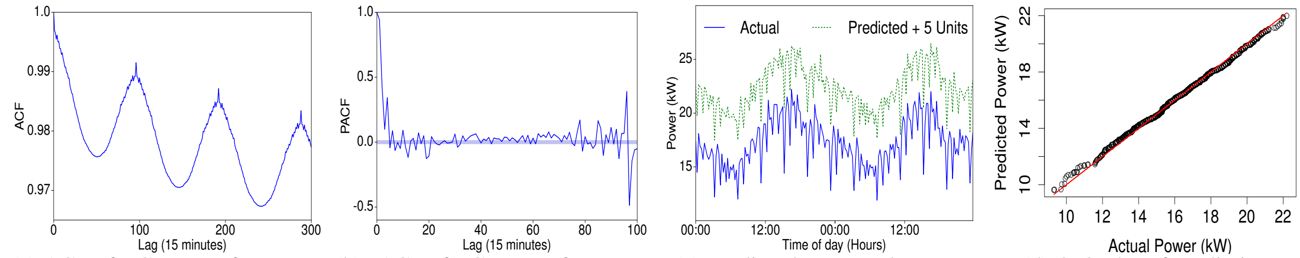
Figure 1: Schematic of a simplified Refrigeration System configuration indicating front-end and back-end components

The system operates in a typical refrigeration cycle as follows. One or more compressors are staged (i.e., scheduled ON) by the control system. These compressors pull in evaporated refrigerant gas from a common suction line. The hot gas is compressed by the compressors resulting in super-heated gas at very high pressure. This super-heated gas is discharged by the compressors on to a common discharge line. This discharge line feeds to the condenser(s) where because of the high pressure, the condensing temperature of the refrigerant is high. Because the condensing temperature of the gas is higher than the ambient temperature of the air outside the condenser (in air-cooled systems), the refrigerant rapidly loses heat (picked up from the evaporators) and undergoes a phase transition to become a high-pressure liquid.

The high-pressure liquid is stored in a receiver, from which the liquid enters the evaporators of the multiple display cases. Each evaporator has a locally controlled valve that regulates the flow of refrigerant through the evaporator. This expansion valve constricts the flow and thereby reduces the pressure, forcing the liquid refrigerant to evaporate to a gas. This evaporation causes cooling in the evaporator coil, which cools the refrigeration cases. Higher the heat-load of the evaporator, more would be the flow of the liquid refrigerant through the valve. If sufficient liquid refrigerant is not available, the pressure of the hot gas leaving the evaporator increases. The evaporator exit feeds the compressor's suction line. When the pressure on the suction line increases beyond a control threshold, the compressors are switched on (if they are off), and the entire cycle repeats.

4. ENERGY MODEL FOR ANOMALY DETECTION

Most supermarkets in developing economies are at best, likely to have energy meters. Typically, the back-end components including the compressor and condenser systems are physically separated from the front-end components in the store. Due to this spatial separation, they are on different electrical circuits. Typically, the front-end circuits (supplying to the display case lights, evaporator fans, and door heaters) are connected to the store's lighting supply, while the back-end (supplying to the compressors and condensers) has a standalone circuit which is metered. In this section, we discuss our approach to develop a model for the back-end refrigeration



(a) ACF of RC energy for a store (b) PACF of RC energy for a store (c) Predicted vs. actual energy (d) Q-Q plot of prediction

Figure 2: SARIMA model for predicting energy consumption (best viewed in color)

energy consumption. We develop a baseline model for the energy consumption using energy data gathered from fault-free durations of the system. The actual consumption observed in the refrigeration system is then compared against this model in an on-line manner. Any significant deviations are flagged as anomalies.

The underlying hypothesis behind this approach is that a fault in any of the refrigeration system component would present itself in the energy signal. As per this hypothesis, we expect that problems in the backend system would result in increase in energy consumption of the backend system. In addition, any problem in the frontend system could *also* cause an increase in the backend energy consumption. For example, consider a refrigerated case (RC) door that is left open permanently; the compressor system would see an additional heat-load and consequently consume more energy. In sum, both front-end and back-end systems are expected to result in anomalies in the back-end energy consumption. Therefore, we focus on developing energy models for the back-end energy consumption so as to be able to detect a wider range of faults.

Ideally, to model the energy consumption, we need to capture the dynamics of the control system(s) of the refrigeration system, track the evolution of the compressor and condenser system states, and map the system operating states to the energy consumption. However, the control choices in the compressor and condenser systems are typically proprietary and tuned locally to a site by the installation vendor. Also, this approach would be prone to the drawbacks associated with a domain-based model discussed in Section 2. Further, even if one were to develop a model that tracks the system state in terms of the controlled variables, it is difficult to calibrate such a control system model (which runs every few seconds) from the sensory data that is typically logged every few minutes Therefore, we do not model the back-end at a control system level.

4.1 Statistical model - SARIMA

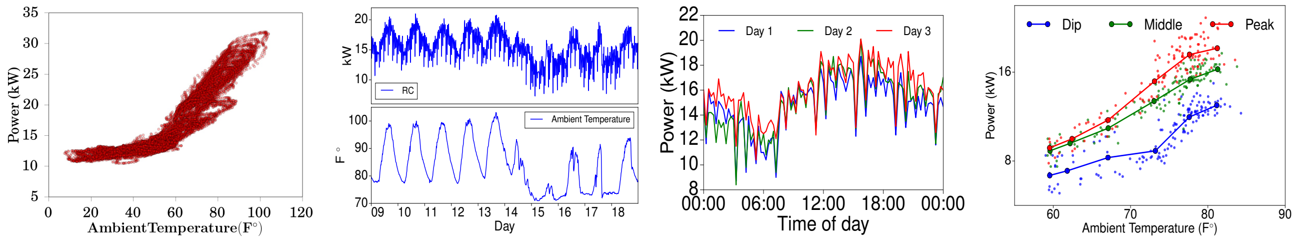
In the absence of an explicit control system model, our approach is to use a statistically-based SARIMA (Seasonal-ARIMA) model for predicting the energy consumption and implicitly tracking the system state. Essentially, the SARIMA model estimates the energy consumption as a regressed function of temporally adjacent samples as well as temporally well-separated samples. We adopted a SARIMA based approach to model the energy for the following reasons: (i) Because the SARIMA model uses temporally adjacent past energy samples to predict the next sample, any effect due to the control system actions are captured. (ii) Because the SARIMA model uses temporally well-separated samples as well (in the seasonal component), it implicitly captures the trends that may exist in the ambient weather conditions.

The energy data is available to us over every 15 minutes as the average power consumed over the interval [2] . Figure 2(a) shows the autocorrelation function (ACF) of the energy timeseries for a typical store, while Figure 2(b) shows the corresponding partial ACF. The X-axis shows the lag of the correlation function where each lag corresponds to 15 minutes, and the Y-axis shows the correlation coefficient at that lag. The figures indicate that a seasonal parameter is required to account for the trend in the ACF. We note the periodicity at every 96 samples (corresponding to one day). We used the Box–Jenkins methodology to arrive at the appropriate orders for the SARIMA model. For a typical store, the order of the parameters in the SARIMA model are as follows: Autoregressive (3), Moving Average (2), Seasonal Auto-regressive (1), Seasonal Moving Average (1), period of 96, and order of differencing 1. This indicates that to predict an power value given the past history, we need to look into local samples (order of hours) to non-local samples (order of days).

Figure 2(c) compares the predicted and actual power consumption time-series with prediction one-sample ahead with all past history known. The X-axis shows time in days and the Y-axis shows the power consumption over 15 minute intervals. The match is very good between the predicted and actual with a mean relative error of 6.7%. Figure 2(d) shows the Q-Q plot that compares the quantiles of the predicted and the actual time-series. We note that except for the upper and lower tails, the predicted matches very well the actual. The upper and lower tails can be used by one for detecting anomalous operations as explained later in Section 6. This SARIMA model is used as the baseline model for energy.

Limitation of the SARIMA model: Recall that the main goal of building a SARIMA model is to use it for detecting anomalous energy samples. Once an anomalous sample is detected say at t, that sample value should not be used for future predictions starting at $t + 1$. This is to ensure that an anomalous sample does not pollute the subsequent predictions. The standard practice is to typically use the predicted estimate of the anomalous sample – which is typically the maximum likelihood estimate (MLE) of the sample's mean value, as the actual sample value for future predictions. When we correct an anomalous sample with the MLE of the sample mean expected at that time, the ensuing sample values predicted from the SARIMA model could diverge from the actual energy series. This is true especially when there is a burst of anomalous samples detected within a small time interval. SARIMA predicted energy values can diverge from the actual energy series because the SARIMA model would converge to the stationary mean of the time-series. Figure 4 shows a case when a series of anomalous samples are omitted, and the corresponding SARIMA mean MLE is used to continue prediction. As we can see, the predicted time-series diverges significantly from the actual series when a series of anomalous samples are detected. Therefore, we need a mechanism that can help us correct anomalous samples for a burst of errors, which is likely to happen when a malfunctioning equipment has not been repaired for a period of time.

[2]**Therefore, neglecting the semantics and abusing the terminology, we use energy and power interchangeably in this paper.**



(a) Energy vs. ambient temperature (b) Time-series of RC energy and ambient temperature (c) Energy consumption on multiple days vs. time-of-day (d) Energy as function of time-of-day and ambient temperature

Figure 3: Development of physical model for back-end energy consumption as a function of time-of-day and ambient temperaure (best viewed in color)

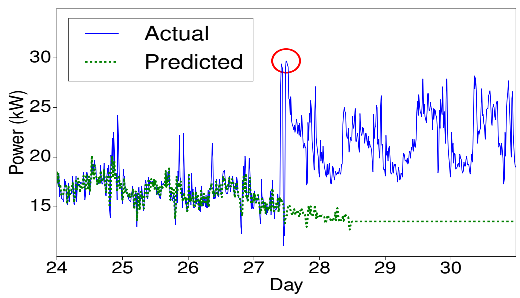


Figure 4: Correcting anomalous samples with MLE predicted samples for a burst of errors can cause the SARIMA model to diverge from the actual data. (best viewed in color).

4.2 Physical model

To create an alternate model that can be used to correct anomalous samples for SARIMA, we identify the physical parameters influencing energy and then fit a model.

As in the HVAC systems, ambient temperature is likely to a key driver for the energy consumption of a refrigeration system's back-end. Figure 3(a) shows a scatter plot of a store's average back-end consumption over 15 minutes in the Y-axis against the average ambient temperature over the same 15 minute interval on the X-axis. From the scatter plot, we see two distinct regimes of operation. Specifically, there is a knee in the trend at some critical value of the ambient temperature τ_C^* of $62^\circ F$. When the ambient temperature T_A is greater than τ_C^*, the curve shows a clear increasing linear behavior, while for $T_A < \tau_C^*$ the curve is near flat to linear with a small slope. We explain these trends as follows.

Linear trend for $T_A > \tau_C^*$: Suppose the set-point temperature of the refrigeration cases is T_S, then the compressor-condenser system is moving heat from a low-temperature of T_S to an ambient temperature of T_A. Due to the physics of the heat flow, the work done is at least given by $H \times \frac{T_A - T_S}{T_S}$, where H is the freezer's heat load to be removed [16]. In the case of supermarket freezers (and coolers), the internal set-point, T_S, remains constant. Further, the heat loads induced in the freezers would also remain constant – since the freezers are in a space conditioned environment and the items that are stocked inside afresh will already be at a temperature comparable to T_S. Therefore, we expect the energy required to remove H from inside the freezer case to outside the store to vary as $H\left(\frac{T_A - T_S}{T_S}\right)$. This explains the linearly increasing trend with T_A when the ambient temperature $T_A > \tau_C^*$.

Flat trend for $T_A < \tau_C^*$: We now want to understand why the system does not behave as expected ideally, when the ambient temperature is less than some critical value. For any cooling to happen in the refrigeration cases, one needs a refrigeration cycle to take place (unless one is letting in air at ambient outdoor temperature through separate piping). For at least some compressor(s) to be working during the refrigeration cycle, we need to maintain some minimum load on the compressor. A compressor being driven by a motor cannot work with no pressure being maintained at the exit (this is akin to short-circuiting a battery by offering zero resistance). Therefore, the condenser system offers a minimum load to the compressor by switching the cooling fans off when the ambient temperature floats below a certain critical value. Consequently, the energy consumption trend remains close to flat when $T_A < \tau_C^*$, and this is true in most stores which have mostly one compressor turned on during low ambient temperatures. We confirmed this with the manufacturer's design specification for the compressors. The minimum design discharge pressure against which the compressor is supposed to operate is around 148 *psig* for the refrigerant type R-404A (which is used in the store under study); this corresponds to a condensing temperature of $70^\circ F$. Therefore, we expect the condenser to maintain this minimum condensing temperature load at any ambient temperature below $60^\circ F$, after accounting for the temperature differential of $10^\circ F$ with respect to the ambient. Indeed, the control system for the fans maintains a differential of about $10^\circ F$ between the condensing and the ambient temperatures.

Hidden variable: While the trends are quite clear, the spread in the graph for the same ambient condition indicates a potential hidden variable that needs to be accounted for. Consider Figure 3(b) which shows the energy consumption as a function of time along with the temperature for the same store. The X-axis is in days for both energy and temperature. Although the energy follows the ambient temperature as a trend, the instantaneous values show high frequency components that do not depend on the ambient temperature alone. Instead, these high frequency components occur at deterministic times-of-the-day. This is confirmed by the time-of-day aligned peaks and troughs of the energy signal at different days as seen in Figure 3(c). The X-axis in Figure 3(c) is the time-of-day over a 24 hour period. The Y-axis shows the energy consumption across three days. We see that across all three days the peaks and dips are aligned at the same times-of-the-day.

These sharp dips and spikes in the energy signal are due to defrost cycles. Specifically, RCs are defrosted according to a schedule fixed by the store. During a defrost, the energy consumption first dips because several refrigerated cases are taken offline and thus the compressor bank sees a steep reduction in the heat load. Post the defrost, the compressors see a sharp peak in the heat-load as the cases at room temperature need to be quickly cooled to the freezing temperature for maintaining food quality.

Figure 3(d) shows the RC energy consumption for a store as a function of the ambient temperature for varying times-of-day. The X-axis is the ambient temperature. The Y-axis shows a family of three curves. Each curve shows the energy consumption for varying ambient temperatures at a specific time-of-day. The three curves

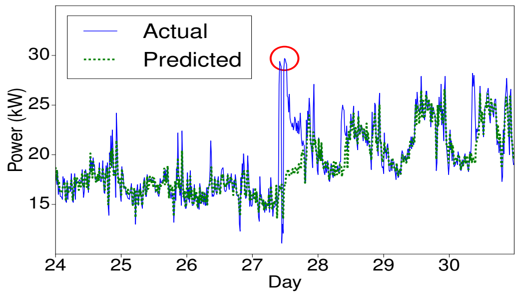


Figure 5: Time-series of RC energy and ambient temperature with an anomalous sample being substituted with the physical model prediction for that time-of-day and ambient temperature. Note that the model does not diverge from actual sample points unlike Figure 4.

are for three sequential times-of-day corresponding to normal operation, defrost dip, and defrost peak. The curve in the middle corresponds to the normal operation, while the other two curves correspond to the peak and the dip of the defrost cycle. We see that when parametrized by time-of-day the scatter plot seen in Figure 3(a) resolves into a family of curves as shown in Figure 3(d). Therefore, any physical model should consider the time-of-day as a parameter, in addition to ambient temperature, for predicting the instantaneous energy consumption of the compressors and condensers. Accordingly, to estimate the average energy consumption ($\hat{E}$) corresponding to a given ambient temperature (T), we first index into the time-of-day (t) for the prediction; and then use the ambient temperature to estimate the energy consumption. In other words, our model is of the form $\hat{E}(t, T) = \phi_t(T)$, where $\phi_t(.)$ is a regressor that relates the energy consumption and external temperature T at the time-of-day t.

We note here that the mean energy value predicted using $\phi_t(T)$ on any given day for a t will not factor the effects of any shock or disturbance seen earlier during preceding t's. In other words, the $\phi_t(T)$ model does not capture the temporal correlations between the energy samples as well as SARIMA. It is for this reason, we do not use $\phi_t(T)$ as the primary model for detecting energy anomalies. As an aside, it is also difficult to obtain a tight bound on the false positive rate of the $\phi_t(T)$ model while (as we will see shortly) it is possible to do so with a SARIMA model.

Overcoming SARIMA limitation: Figure 5 shows the effect of correcting the SARIMA sample using the average physical model prediction value for a given ambient temperature. This overcomes the limitation highlighted in Figure 4. While the SARIMA model starts diverging around the set of anomalous samples, because we use the physical model, it reconverges to normal operations and the predicted values once again match the normal values.

5. WORK-ORDERS: BACKGROUND

When a malfunctioning equipment is noticed (e.g., a condenser fan making noise) or an anomalous operation is observed (e.g., an RC not maintaining a cold environment), the store personnel log a *work-order*. A work-order log basically describes the problem symptoms, categorization, the day it was observed, the day it was fixed, and details of an eventual fix. Ideally, work-orders should be avoided through proactive maintenance. This is because a work-order indicates that there was disruption of service and food could potentially go bad entailing other losses.

Trivial work-orders are the ones that are readily detected by some sensors tripping the pre-determined level. For example, a typical compressor system would trip the compressor off if the discharge pressure becomes excessive and raises an alarm. In this paper, we focus on non-trivial work-orders. The non-trivial work-orders occurring in the system can be broadly categorized into those occurring in the front-end and back-end systems. Over the study period of five months, four types of work-orders occurred across our study sample of 25 stores. The work-orders we observed are as follows:

Work-order type	Symptoms
Leaky refrigerant	Poor cooling in the coolers and freezers.
EPR valve malfunctioning	Excessive cooling in cases. Items are too cold.
Iced evaporator	Ice buildup across expansion coils in coolers and freezers. Poor cooling.
Iced door	Excessive condensation, frosting in case doors. Poor visibility into the case.

Table 1: Work-order types and their associated symptoms

- **Leaky refrigerant:** As the name indicates, this occurs when the refrigerant leaks from system due to fatigue in pipe walls or valves. If left unattended, can lead to a complete system shut-down and no refrigeration (cooling) will take place. In certain instances, even the compressors can get damaged.
- **Malfunctioning EPR valve:** EPR valve is the element which throttles the refrigerant flow into the evaporator coils. A malfunctioning valve can flood the evaporator coils with more refrigerant than what is necessary. This can result in over-cooling of the stored items which can damage the quality. In extreme cases, even liquid refrigerant can enter the compressors permanently damaging them.
- **Iced evaporator:** This is the work-order in which thick ice or frost builds around the evaporator coils which decrease the refrigeration effect. Consequently, the refrigerated case becomes warm which can damage the food items.
- **Iced door:** This work-order results if the refrigerated case's glass door is all covered in frost/water vapour. This results in poor visibility of the case items which can annoy the end consumers. This work-order affects the sales.

These work-orders and the associated physical symptoms are summarized in Table 1. Depending on the work-order, the time to fix could vary from 9 to 83 days with an average of 29 days.

The work-orders are logged as soon as the symptoms become visible and are observed by the store personnel. Our goal, however, is to detect these work-orders as early as possible, or even anticipate them before these effects become visible thereby reducing/eliminating repair downtime through pro-active maintenance. It is also likely that sensory instrumentation associated for detecting some/all of these work-orders may not be available in smaller stores and stores in developing economies. Therefore, we are interested in early-detection/prediction of these work-orders by observing how the energy signal behaves during work-orders.

6. DETECTING WORK ORDERS USING ENERGY SIGNALS

As mentioned earlier, our underlying hypothesis is that any fault or a work-order in any refrigeration system component would present itself in the energy signal. We expect problems in both frontend and backend systems to create anomalies in energy consumption of the backend system.

Over the study period of 5 months, for each store, we identify a consecutive duration of 30 days in which no work-orders occur. To ensure that the data period corresponds to "normal" operations, we allow for an additional buffer window of three days around the end-points of the identified 30-day window. While this method cannot guarantee that no work-order points occur in the selection, we believe this to be a reasonable approximation of the same. Using the energy data over these 30 days, we train individual SARIMA and Physical energy models for each store. We then use this model the predict the energy samples for the rest of the days in our study period (which becomes our test set). We compare the predicted energy sample(s) with the actual energy sample(s) over an interval to flag something anomalous. Because we flag samples over an interval, any anomalous real sample cannot be used in future predictions. So we use the physical model for the average energy consumption to replace the anomalous energy sample in SARIMA prediction as explained in Section 4. Using the prediction model, we can identify anomalous samples and the associated deviations of the actual energy samples from the predicted.

6.1 Anomaly detection rule

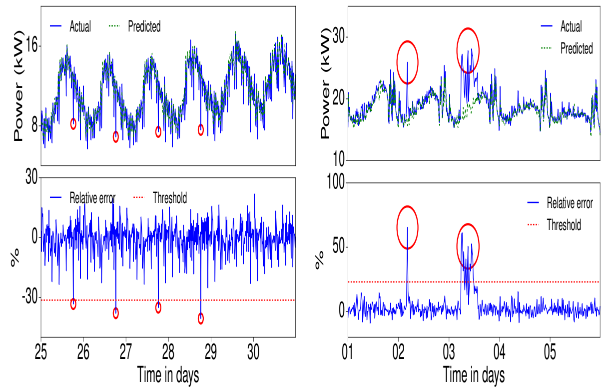
Let $E_A(t)$ denote the actual energy consumption at t and $E_P(t)$, the predicted energy consumption. Let α be such that $0 \leq \alpha \leq 1$. Define $\epsilon(t) = \frac{(E_A(t)-E_P(t))}{E_P(t)}$. Over all t in the training set, let ϵ_H^* and ϵ_L^* respectively denote the α-th and $(1-\alpha)$-th percentile points of $\epsilon(t)$. The value of α is chosen such that ϵ_H^* corresponds to outliers that are positive (i.e., indicate increased energy consumption) and ϵ_L^* corresponds to outliers that are negative (i.e., indicate highly reduced energy consumption). Our rule for classifying an energy sample as anomalous and hence indicate the onset of a work-order is as follows:

- **Positive outliers:** If $\epsilon(t) > \epsilon_H^*$, then the sample at t is anomalous and work-order is flagged.
- **Negative outliers:** If $\epsilon(t) < \epsilon_L^*$, then the sample at t is anomalous and work-order is flagged.

Note that as per the above rules, the occurrence of even one outlier would be detected as a work-order. While one could potentially use two or more consecutive outliers to flag a work-order, we do not do so for the following reason. Consider a malfunctioning refrigerated case (RC) that presents a higher (lower) heat-load to the compressor. The impact of this RC on back-end energy would not be uniform throughout. It would depend on the relative loads of other RCs served by the same compressor. Therefore, if there is any front-end work-order in an RC, its effect on compressor energy may be quite short-lived (but not necessarily). Since energy is sampled every 15 minutes at our study stores, to capture such short-lived events, we have defined the classification rule to accommodate only one outlier. We note that, since the classification rules are based on the error percentiles observed, the false-positive rate of our classification is $(1-\alpha)$.

6.2 Error direction as work-order signatures

Depending on the impact of the anomaly on the operations, the energy consumption could increase or decrease. Specifically, if the anomaly increases (reduces) the heat load seen by the compressor, it results in higher (lower) energy consumption. Among the work-orders we observed, a refrigerant leak shows an anomaly in the lower tail. Because refrigerant leaks from the system, lesser pressure builds up at the compressors' suction inlet, and so lesser number of compressors are scheduled to reduce the built-up pressure. This is demonstrated in Figure 6(a). Note that this will likely be accompanied by a loss in cooling capacity. Therefore, if there is a negative outlier in the energy signal, it can be construed as the onset of refrigerant leak. The other three work-orders, however, had anomalies which were all in the upper tail. *Thus, the direction of the anomaly can be used as a signature to identify the anomaly root-cause to be a subset of the refrigeration system components (if not to the degree of an individual component).*



(a) Lower tail anomaly (b) Upper tail anomaly

Figure 6: Error direction based classification of anomalies

6.3 Performance

Choice of α: A low value of α would increase the likelihood of detecting a work order but also would increase the false positive rate. Suppose we have N samples per day from M data sources. Then, for upper-tail anomalies, the number of samples that would be classified as potentially anomalous is approximately $(1-\alpha) \times N \times M$. Note that human checks or interventions are typically dispatched by an external contracted agency in response to store personnel complaints. Thus to keep human intervention to a reasonable frequency, we assume a conservative estimate of the number of checks or visits by a technician to be one per week. For a typical store, $N = 96$ and $M >= 2$, and thus, the value of $1-\alpha$ should be at most 0.00074. As a more conservative choice, we use $1-\alpha = 0.002$.

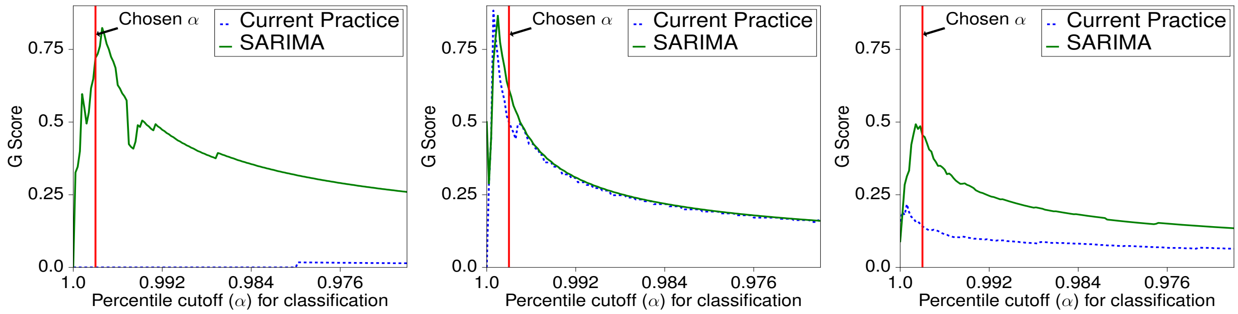
Table 2 gives the detection likelihood and false positive rates of the energy based detection. While the detection likelihood is calculated over stores with work-orders as seen in the ground truth, the false positive rate is calculated over all stores. We note that the average false positive rates for the stores with work-orders is 0.06%, which is lesser than the method's expected 0.2%. For the stores with no work-orders, the false positive rates are around 0.2%. In sum, the detection likelihood is reasonable, and the false positive rates are low. We also note that using energy anomalies, we are able to detect the work-orders well in advance which can be quite useful in practice.

Tail	Work-orders	Detection Likelihood (%)	False positive (%)	Early detection (days)
Upper	Iced door, Iced evaporator, EPR valve	80	0.17	2.8
Lower	Leaky refrigerant	95	0.12	20.1

Table 2: Performance of energy based detection

6.4 Comparison with baseline

As a baseline for comparison, we use the current practice of setting a (static) threshold for the signal being measured; and using any violation of the threshold as a potential outlier. We proceed with the comparison as follows. For a given α, we identify the $(1-\alpha)$-th percentile point as the (static) threshold for the baseline approach. For the same α, we use the $(1-\alpha)$-th percentile point



(a) Maximum improvement (b) No improvement (c) Average improvement

Figure 7: Comparison with a static threshold based approach across stores

of the error between our predicted value of the energy and the actual sample, as the (dynamic) threshold. For both approaches of classifying samples as outliers, we calculate the standard metrics of classification, i.e., precision and recall, and compose them using the G-score, which is defined as $\sqrt{(Precision \times Recall)}$.

Figure 7(a) compares the performance of our approach and the baseline approach for the best-performing store. The X-axis shows the cut-off percentile α for the classification. The Y-axis shows the G-score as a function of α as evaluated on the data-set for that store. As can be seen, our approach performs significantly better than the baseline approach for most values of α in this store. Further, for the value of α chosen to minimize the number of human interventions, the performance is close to the best performance.

Figure 7(b) shows a store where our approach performs very similar to a statically chosen threshold. In this store, the variation in energy is very low across time and temperature, that there is hardly any need to predict the current consumption as a function of time-of-day and temperature. Therefore, a simple static threshold identifies the same anomalies that any prediction-based approach identifies. The average improvement of our approach (average across all stores for each value of α) is shown in Figure 7(c). We find that the average improvement is significant enough to merit a model that accounts for modeling temporal variations in the energy signal.

7. IDENTIFYING WORK-ORDERS USING OTHER SENSORS

Having detected work-orders using anomalies in the energy signal, we now explore if the root-cause of the problem or the faulty component can be identified. This would help in stores where a reasonable set of sensors have been deployed. Typically, among others, sensors are deployed for: number of active compressors, number of active condenser fans, refrigerant temperature at condenser exit, level of the refrigerant in the receiver, temperature and pressure of the refrigerant at the suction end of the compressor, temperature and pressure of the refrigerant at the discharge side, and energy consumed by door heaters and evaporator fans.

7.1 Model fitting for sensory data

As with the energy time series, we fit a suitable statistical model for each individual sensor and identify anomalous samples. We hypothesize that a work-order will be identified by anomalies in one or more sensors. The goal is to identify one or more sensory streams that can be used as a feature vector. The training and testing set is taken from the same duration as done for the energy models. The value of α is taken to be 99.8.

Since the ambient temperature and energy consumption have daily patterns, several of these sensory variables (e.g., discharge pressure of a compressor)also follow a similar pattern. Thus, for all those sensors which have a statistically tractable model, we identify upper and lower tails as seen before for energy.

Some sensed parameters, on the other hand, are maintained at a steady level and can be treated as stationary processes. For example, the suction pressure at the inlet of compressors is maintained within a band typically by using PID based scheduling of compressors. If the heat load increases, the suction pressure goes up, and compressors are turned on to reduce the pressure by pulling the hot refrigerant in. For such sensors, the data can be treated as a stationary ARMA time-series and upper and lower tails of this stationary distribution can be obtained to see if there are any anomalies given the local history. As with the energy time series, anomalous samples are substituted with estimates obtained from appropriate alternative regression based models.

7.2 Mapping sensor readings to work-orders

Let K be the number of sensors deployed to monitor a refrigeration system in a store. Let $\mathbf{S}(t) =< S_1(t), S_2(t), \cdots, S_K(t) >$ be the array of sensor readings at time t corresponding to each of the k sensors deployed. Given the past history up to $t-1$, using either SARIMA or stationary models, we flag an anomaly at t if at least one of the sensors $S_i(t)$ is anomalous according to the individual time-series model. Admittedly, this approach does not consider combinations of sensor readings and does not explicitly use information that may be available in the joint distributions of the sensors. Nevertheless, we find that this approach is effective and can yield reasonably good results.

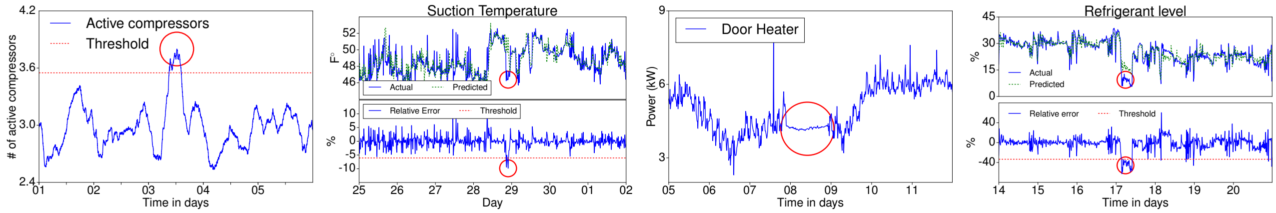
For the work-orders that we have observed in our dataset, we find that there is good correlation between the observed anomalies in the sensor data and the work-order being logged in the system. For each of the four work-orders observed in our study stores, there was one sensory signal whose anomaly uniquely identifies the work-order. In other words, the anomaly of the unique sensory signal acts as the feature vector for classifying the work-order.

The sensory parameter used as the feature for each work-order type is shown in Table 3. Figures 8(b), 8(a), 8(d), 8(c) show the sensor anomaly samples respectively for work orders corresponding to excessively cold refrigerated case, ice buildup in the evaporator, leaky refrigerant, and ice buildup on doors.

7.3 Sensors as features for work-orders

Sensor	Work-order
Liquid level	Leaky refrigerant
Suction temperature	Faulty EPR value
Active compressors	Evaporator ice buildup
Door heater energy	Frosting/condensation in doors

Table 3: Sensors features for work-orders



(a) Ice buildup in evoporator (b) Excessive cold in case (c) Ice buildup in door (d) Leaky refrigerant

Figure 8: Sensor based classification of work orders

We now explain why each type of work order manifests as an anomaly in the corresponding sensor reading. Recall Figure 1 that shows the refrigeration cycle in a RC. We expect that an anomaly in each component would show up as an anomaly in the sensor reading either in the inlet or the outlet of the component within a short period of time; the anomaly could also show up in another component after a significant period of time. For ease of presentation, we first highlight an example with anomalous sensor readings for each work-order type and summarize aggregate performance statistics later.

Ice buildup in evaporator: When ice builds up in the evaporator, the ice thermally insulates the evaporator from the rest of the refrigerated case. Therefore, the normal heat flow **rate** from the refrigerated case to the refrigerant (which carries it away as it flows) is reduced. Because the liquid refrigerant continues to flow, to maintain a set-point, the liquid refrigerant flow rate through the evaporator needs to increase. Thus, additional compressors need to be scheduled to increase the refrigerant flow rate as the compressors act in parallel. This also leads to an increase in the energy consumption. We expect to see this picked up in the number of compressors that are scheduled.

Figure 8(a) shows a smoothed (over 1 hour) version of the number of active compressors in the system. Between days 3 and 4, we see than the number of compressors exceeds the 99.9% percentile of all points during operation. The energy also picks up during this interval as seen in the corresponding upper tail anomaly graph in Figure 6(b). As explained later, using both the energy and sensor based thresholds could be used to reduce the false positive rate.

Excessive cooling in refrigerated cases: Excessive cooling in an RC implies that the flow rate of the refrigerant through the evaporator is higher than the heat load offered by the evaporator. This in turn would result in the refrigerant leaving the evaporator without picking up much-heat, i.e., a refrigerant with less super-heat. Consequently, this gets detected in the monitored suction temperature being lower than expected for that ambient condition.

Figure 8(b) shows the monitored suction temperature as a function of time for a store with a excessively cold refrigerated case. Between days 28 and 29, the monitored suction temperature dips significantly. The dip is observed with a corresponding increase in the energy consumption (i.e., an upper-tail energy anomaly).

Ice build up on case door: Whenever there is condensation or frosting in the door of a refrigeration case, the door heater gets turned on (by sensing the humidity level) and evaporates the condensate. Since this is a state dependent activity, it is non-periodic. When the door heater fails to get turned on, frost build-up happens. This non-turning of door heater can be detected as an anomaly in the front end energy consumption signal. Specifically, we expect to see that the signal has lesser "noise" than average, which we confirm by using a rolling deviation of the signal. Figure 8(c) shows that the *front-end* energy channel showing a flat consumption well before the period the work-order is actually noticed and logged.

Leaky refrigerant: Consider the receiver in Figure 1. The refrigeration loop tries to produce and maintain some level of liquid refrigerant in the receiver. Because the loop is closed, for a given ambient condition, we expect the time-average liquid level to be roughly constant, barring transient oscillations due to compressors and condensers turning on and off. However, when the refrigerant leaks from the system, we expect the average liquid level maintained in the receiver to go down significantly from the expected value. We have developed a statistical model for the liquid level, and we see that the average liquid level does go down when the work-order of refrigerant leak is logged in the system. Note that the work-order is logged in the system only when the liquid level is critically low and trips some sensor. Figure 8(d) shows the liquid level for a store with leak refrigerant. There is a progressive degradation is the maintained liquid level, and there is a sharp perturbation just before the work-order itself is logged around day 17-18.

7.4 Overall performance across work-orders

While a sensor reading can detect a particular work-order, it does so with some false positive rate. We find that while using both energy and sensor readings to identify and classify an anomaly, the false positive rate decreases in comparison to using the sensor alone for detection. This is because the energy signal is correlated with the sensory signals, especially around true anomalies which allows combining the two information streams with beneficial results. Table 4 shows the results of the detection using both the sensor and energy signals. On an average, the detection using energy and sensors has lesser false positives than the expected tail probability mass of 0.2%, indicating the robustness of the method. Further, the time of early detection is good and very significant in the case of liquid level work-orders. This is because the degradation in liquid level due to a refrigerant leak is a progressive event, the actual work order being logged only when the receiver fails to function due to to critically low refrigerant during operation. The impact on the detection likelihood is very little for three of the four work-orders. The detection likelihood decreases for the evaporator ice build-up work-order alone. This can be overcome by using a lower threshold value α for detecting the energy anomalies. The increase in false positives in energy anomalies will be toned down in the overall detection due to the supplementary sensory input (viz., # of active compressors).

Table 5 presents the confusion matrix using data from both energy and sensors. Each metric is first calculated for each store. For the sake of brevity, we present the average (across all stores) value of the metric averaged in each of the rows. While the precision and recall values for three types of work-orders are acceptable, the precision for the frosting work-order is low. This is because in our data-set we could detect only one true positive, while significant number of false positives were triggered due to a conservative choice of $\alpha = 0.98$ across all work-order types, whereas a higher α would improve the precision.

Features used	Energy		Sensor & Energy				
Work-order	Detection Likelihood (%)	False Positives (%)	Detection Likelihood (%)	False Positives (%)	# stores	# Events	Avg Early Detection (days)
Faulty EPR valve	100	0.13	66	0.05	4	6	4.9
Evaporator ice buildup	71	0.18	71	0.02	7	7	1.43
Leaky refrigerant	95	0.12	95	0.03	11	12	20.2
Frosting in doors	100	0.19	100	0.17	1	1	0.33

Table 4: Combining detection from energy and sensors improves the classification accuracy

Work-order	False Positive	False Negative	True Positive	True Negative	Precision	Recall
Faulty EPR valve	0.0005	0.0002	0.0007	0.9985	0.53	0.75
Evaporator ice buildup	0.0002	0.0005	0.0015	0.9979	0.90	0.75
Leaky refrigerant	0.0003	0.0011	0.0033	0.9953	0.82	0.77
Frosting in doors	0.0018	0.0001	0.0001	0.9981	0.06	0.67

Table 5: Details of the classification accuracy when using both energy and sensors. For each work-order type, the metrics are averaged across all the stores where that work-order occurs.

8. CONCLUSIONS

Refrigeration systems consume a significant portion of supermarket aggregate energy consumption. Because failure of such systems can have significant impact on store operation, it is important to identify any potential faults in their operation before critical failure. We presented a data-driven study of fault detection in 25 stores. Using the energy signal alone, we were able to identify faults with good detection likelihood and low false positive rates. The method could detect faults 2.8-21 days before the fault is noted in the database. Using additional sensors, we were able to further classify the faults into the four types that occurred during the duration of the study using a union of faults shown up by all sensors. Future directions of work include modeling dependencies between sensors to identify frequently occurring short-term anomalies that remain undetected with long-term modeling.

9. ACKNOWLEDGEMENTS

We thank our shepherd Mario Bergés and the anonymous reviewers; their comments have significantly improved the paper.

10. REFERENCES

[1] Commercial Buildings Energy Consumption Survey. `http://www.eia.doe.gov/emeu/cbecs/cbecs2003`.

[2] Energy Star Building Energy Manual - Facility Type Supermarkets. `http://www.energystar.gov/buildings/sites/default/uploads/tools/EPA_BUM_CH11_Supermarkets.pdf`.

[3] M. S. Breuker and J. Braun. Evaluating the performance of a fault detection and diagnostic system for vapor compression equipment. *International Journal of Heating, Ventilating, and Air Conditioning and Refrigerating Research*, 4(4):401–425, 1998.

[4] R. Fisera and P. Stluka. Performance monitoring of the refrigeration system with minimum set of sensors. *World Academy of Science, Engineering and Technology*, 6(7):396–401, 2012.

[5] H. Grimmelius, J. K. Woud, and G. Been. On-line failure diagnosis for compression refrigeration plants. *International Journal of Refrigeration*, 18(1):31–41, 1995.

[6] H. Han, Z. Cao, B. Gu, and N. Ren. Pca-svm-based automated fault detection and diagnosis (afdd) for vapor-compression refrigeration systems. *HVAC&R Research*, 16(3):295–313, 2010.

[7] M. Kim, S. H. Yoon, P. A. Domanski, and W. V. Payne. Design of a steady-state detector for fault detection and diagnosis of a residential air conditioner. *International Journal of Refrigeration*, 31:790–799, 2008.

[8] M. Leach, C. Lobato, A. Hirsch, S. Pless, and P. Torcellini. *Technical Support Document: Strategies for 50% Energy Savings in Large Office Buildings*. National Renewable Energy Laboratory, 2010.

[9] B. Narayanaswamy, B. Balaji, R. Gupta, and Y. Agarwal. Data driven investigation of faults in hvac systems with model, cluster and compare (mcc). In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, BuildSys '14, pages 50–59, 2014.

[10] K. Palani, N. Nasir, V. C. Prakash, A. Chugh, R. Gupta, and K. Ramamritham. Putting smart meters to work: Beyond the usual. In *Proceedings of the 5th International Conference on Future Energy Systems*, e-Energy '14. ACM, 2014.

[11] K. B. Rasmussen, A. Kieu, and Z. Yang. Fault detection and isolation for a supermarket refrigeration system. Master's thesis, Aalborg University, Denmark, 2009.

[12] N. Rena, J. Liang, B. Gua, and H. Han. Fault diagnosis strategy for incompletely described samples and its application to refrigeration system. *Mechanical Systems and Signal Processing*, 22:436–450, 2008.

[13] S. Tassou and I. Grace. Fault diagnosis and refrigeration leak detection in compression refrigeration systems. *International Journal of Refrigeration*, 28:680–688, 2005.

[14] C. Thybo and R. Izadi-Zamanabadi. Development of fault detection and diagnosis schemes for industrial refrigeration systems lessons learned. In *Proceedings ofthe 2004 IEEE International Conference on Control Applications*, 2004.

[15] U.S. Environmental Protection Agency. ENERGY STAR Certified Buildings and Plants. `http://www.energystar.gov/buildings/`.

[16] G. J. Van Wylen and R. E. Sonntag. *Fundamentals of classical thermodynamics*. Wiley Singapore, 1985.

[17] A. Wichman and J. E. Braun. Fault detection and diagnostics for commercial coolers and freezers. In *Proceedings of the International Refrigeration and Air Conditioning Conference*, 2008.

[18] Q. Zhou, S. Wang, and Z. Ma. A model-based fault detection and diagnosis strategy for hvac systems. *International Journal of Energy Research*, 33(10):903–918, 2009.