

# Pattern mining and fault detection via $COP_{therm}$ -based profiling with correlation analysis of circuit variables in chiller systems

Jasmine Malinao · Florian Judex · Tim Selke ·  
Gerhard Zucker · Jaime Caro · Walter Kropatsch

Published online: 25 January 2015  
© Springer-Verlag Berlin Heidelberg 2015

**Abstract** In this paper, we propose methods of handling, analyzing, and profiling monitoring data of energy systems using their thermal coefficient of performance seen in uneven segmentations in their time series databases. Aside from assessing the performance of chillers using this parameter, we dealt with pinpointing different trends that this parameter undergoes through while the systems operate. From these results, we identified and cross-validated with domain experts outlier behavior which were ultimately identified as faulty operation of the chiller. Finally, we establish correlations of the parameter with the other independent variables across the different circuits of the machine with or without the observed faulty behavior.

**Keywords** Data mining · Energy efficiency · Building automation · HVAC · Adsorption chiller

## 1 Introduction

For many decades now, there had been different environmental measures introduced and maintained to mitigate global warming and climate change and protect the planet and its inhabitants from its disastrous effects. Since the Montreal and Kyoto protocol had taken effect and/or adopted in many parts of the world from 1989 and 1997, respectively, and adhered to until now, the reduction and ultimate phasing out of the use environmentally-harmful chemicals for household and commercial equipments most especially for refrigeration, air conditioning, insulation, etc. Examples of such are those categorized as Chlorofluorocarbons (CFCs), Hydrochlorofluorocarbons (HCFCs), and others that are considered greenhouse gases that have potential contributions to global warming. A report in [1] mentions that “just replacing some 25 % of the HCFCs with substances that have zero global warming potential will have a reduction in the global warming contribution that is equivalent to the reduction achieved by the Kyoto Protocol during its first commitment period from 2008 to 2012”.

In the field of Heating, Ventilation, Air Conditioning (HVAC) systems, some of the innovations already had slowly started to shift adhering and supporting the initiatives of the two Protocols. In some HVAC machines such as adsorption chillers, refrigerants now come in a very Earth-friendly form that is water. Water is continually circulated in the chiller to provide cooling to many different spaces with the aid of high-capacity adsorption materials embedded in the chiller itself. The entire cooling process is a closed cycle with a supply of high pressure values to effectively vaporize the refrigerant at

---

J. Malinao (✉) · F. Judex · T. Selke · G. Zucker  
Energy Department, AIT Austrian Institute of Technology,  
Vienna, Austria  
e-mail: jasmine.malinao@ait.ac.at

F. Judex  
e-mail: florian.judex@ait.ac.at

T. Selke  
e-mail: tim.selke@ait.ac.at

G. Zucker  
e-mail: gerhard.zucker@ait.ac.at

J. Caro  
University of the Philippines, Diliman, Quezon City, Philippines  
e-mail: jdlcaro@dcs.upd.edu.ph

W. Kropatsch  
Vienna University of Technology, Vienna, Austria  
e-mail: krw@prip.tuwien.ac.at

low temperature values. The driving temperature may be supplied partially or fully by environment-friendly and/or renewable energy resource such as solar energy. Adsorption chillers could also provide heating when necessary through the heat rejection process of the machine. This framework of heating and cooling technology is also coherent with what the German Federal Government's climate initiative states to promote emissions reduction, namely, (a) Improving energy efficiency; (b) Developing renewable energies; and (c) Reducing F-gases (fluorocarbons) which have a harmful impact on climate [2].

However, it is notable that even when the concepts and benefits of adsorption technologies had long been known [3], it is just fairly recent that people from the HVAC community had taken them more seriously. Recent innovations to support better and cheaper production, utility and management of such machines, aside from felt alarming effects of global warming as mentioned above, had continued to encourage its technological growth in recent years. The reintroduction of focus on this technology is fairly young such that the current literature still focuses on either designing and/or optimizing prototypes of adsorption chillers to yield better efficiency and energy-saving performance [4–8]. Unfortunately, only a few Asian and European manufacturers build adsorption chillers to date, e.g. Japan and Germany [9], hence there is but a small pool of related work in studying and profiling the static and dynamic behavior of such machines.

Alongside with a small number of manufacturers of adsorption chillers, the (re)design and (re)optimization of their structures and processes largely remain under the following pursuits: (a) experimentally-determined schemes, configurations, and solutions that are usually component specific [10–14]; (b) benchmarked results from systems under some specific environmental setting [15–18]; (c) experimental parameter tweaking to achieve improvements in coefficients of performance, reduction of heat losses, energy consumption, greenhouse emission, etc. [19,20]. Under these frameworks, it is difficult to obtain optimality of the structure and processes of chillers, moreso in diagnosing, and describing faults that occur in them. A few data mining based-frameworks for energy systems' fault detection and diagnosis (although component-specific too) can be seen in [21–23].

In this research, we focus on these type of adsorption chillers and profile their energy efficiency through the historical data of its operation. We extract different scenarios of the behaviour of its thermal coefficient of performance as it operates through many duty cycles to address cooling demands. From this profiling mechanism, scenarios of faulty behaviour are discovered and validated by domain experts. For all these range of faulty and non-faulty operations, protocols and control strategies could be developed for better building automation technologies.

## 2 Basic definitions and notations

### 2.1 Machine and data specifications

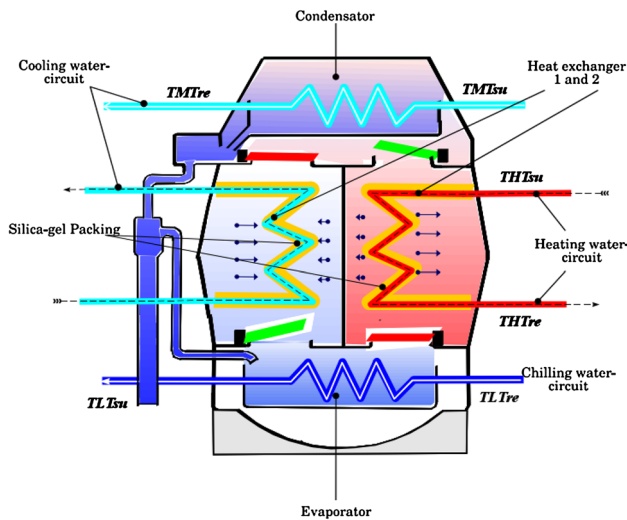
Adsorption chillers are mainly designed with an evaporator, adsorber/desorber reactor beds, and a condenser that altogether circulate a refrigerant to produce cold supply by using driving heat as seen in Fig. 1. There are three main loops that support this adsorption cycle namely the heating water circuit (*HT*), chilling water circuit (*LT*), and cooling water circuit (*MT*). At the midsection of the figure of the chiller are the two reactors that alternate in the adsorbing and desorbing of the refrigerant through the aid of the silica-gel adsorbent found inside them. As the machine operates, the left-bottom check valve opens up so vaporised refrigerant moves from the evaporator towards the adsorber reactor where the vapour is adsorbed by silica-gel granules with the aid of cooling water supply. This vaporization produces the chilling capability of the machine. Meanwhile, the right-bottom valve closes as the other reactor is provided with high heat by the *HT* circuit to desorb the existing refrigerant off from the silica gel. The desorbed refrigerant moves towards the condenser and is liquified again before finally being pumped back to the evaporator for the next cycle of operation to begin. Sensors can be used to monitor the return and supply temperature values, i.e.  $T_{LTre}$  and  $T_{LTsu}$  of the *LT* circuit, and similarly,  $T_{MTre}$  and  $T_{MTsu}$  of the *MT* circuit, and  $T_{HTre}$  and  $T_{HTsu}$  of *HT* circuit, as well as the circuits' mass flow rates and other related resource utilization during the refrigeration cycle. Please refer to the annotations provided in Fig. 1.

### 2.2 Flow rates and energy

Let  $F(c, l) = [f_{i,c,l}]$  and  $T(c, l) = [t_{i,c,l}]$  be the time series data sets collected by the chiller's energy meters where  $f_{i,c,l}$  is the flow rate of the water (or refrigerant) passing through a chiller's circuit  $c$  and  $t_{i,c,l}$  is the corresponding recorded temperature value in the circuit component  $l$  at time instance  $i$  on regular intervals  $k$  (in minutes),  $c \in \{LT, MT, HT\}$ , and  $l \in \{return, supply\}$ ,  $i, k \in \mathbb{N}$ . Let  $Q = [q_{i,c}]$  be the amount of energy (in kWh) spent by the chiller in  $c$  to aid in performing the entire cooling process during an operation of the adsorption chiller, as follows,

$$\forall i, q_{i,c} = \frac{|t_{i,c,l} - t_{i,c,\{return,supply\}}| * m_{i,c,k} * \delta}{360,000},$$

where  $m_{i,c,k} = f_{i,c,l} * (k * 60) * \rho$ , is the mass of the refrigerant on time  $i$  in  $c$ , and  $\rho$  and  $\delta$  are the density and the specific heat of the refrigerant, respectively (e.g. water has  $\rho = 1 \text{ kg}/0.001\text{m}^3$  and  $\delta = 4,185.5 \text{ J}/\text{kg}/\text{K}$ ). We use the  $(k * 60)$  and the denominator for proper SI unit conversion purposes where  $360,000 \text{ kWh} = 1 \text{ J}$ . Therefore  $Q$  is in kWh.



**Fig. 1** The adsorption chiller cooling specifications. (Image source: [26])

### 2.3 Coefficient of performance

For this study, we focus on describing and analyzing chiller performance using the thermal coefficient of performance  $COP_{therm}$ , where  $COP_{therm} = Q_{LT}/Q_{HT}$ , where  $Q_{LT}$  and  $Q_{HT}$  are the values of the spent energy per hour of the  $LT$  and  $HT$  circuits, respectively. For this study, we derive these energy values as indicated above using the flow rates (in  $m^3/h$ ) recorded from their corresponding circuits.

### 2.4 X-Means clustering algorithm

The X-Means clustering algorithm [24] is an offshoot of the conventional K-Means clustering algorithm that addresses three major drawbacks of the latter.

Firstly, K-Means is computationally expensive as it performs exhaustive computations for each data point to determine its cluster membership. When the dimensionality of the input increases, K-Means meets scalability issues in computing models. Meanwhile, X-Means provide a solution to this by employing the concept of *blacklisting*. Blacklisting ensures that a pre-specified list of centroids are the only ones considered for (re)clustering, thus, the number of computations is bounded by this list and minimizing the iterations done in K-Means.

Secondly, K-Means is dependent on user input to specify the number of clusters to be formed for the cluster model. The decision of the user bounds K-Means to finding information by partitioning the data set at this fixed number of clusters. This is problematic when some sets of behaviour are unknown to the user and therefore promotes constrictive choices in discovery of new, sometimes, important behaviour/s that could actually appear and/or develop through time

and variations in settings. X-Means however makes local decisions in each current cluster (Voronoi region) at an iteration as to how the region should be partitioned based on Bayesian Information Criterion (BIC) values, and chooses the best clustering models obtained from all the computations. X-Means therefore quickly approximates the ideal global partitioning of the data set. This eliminates the weakness we observe in K-Means wherein the users themselves decide on how we subdivide the data set to perform cluster analysis.

Finally, K-means tends to find worse local optima when the number of clusters is fixed rather than dynamically changing this parameter in its computations. The iterations done in X-Means will likely avoid being trapped in a worse local optima for a cluster model as it dynamically approximates optimal values of for the number of clusters of the data set.

X-Means is comprised of the following steps:

1. **Improve-params** This step executes the conventional K-Means until convergence. The average square distance is used as a similarity measure to determine cluster membership in each iteration of K-Means.
2. **Improve-structure** This step determines when and where new centroids will appear. Such appearance is determined by computing the BIC of the clusters formed.
3. If the number of clusters formed in an iteration exceeds  $k_{max}$ , the algorithm stops and reports the best-scoring model (i.e. the cluster model where the BIC is maximized) found throughout the search. Otherwise, proceed to the first step. Note that  $k_{max}$  can be set to at most the number of points in the data set.

## 3 Methodology

### 3.1 Data segmentation and sanitation

The adsorption chiller is under operation whenever we detect that the refrigerant is passed through the different compartments of the chiller, i.e.  $f_{i,c,l} > 0, \exists i, c$ . Using this information, a segment  $S = [s_j, s_{j+k}, s_{j+2k}, \dots, s_{j+mk}]$  is constructed whenever we find its corresponding set of positive flow rates  $[f_{j,c,l}, f_{j+k,c,l}, \dots, f_{j+mk,c,l}]$ ,  $j, m \in \mathbb{N}$ , for any of the time series of the chiller's historical data.

To initiate the discovery of the different sets of behaviour and efficiency states of the chiller, we segment the time series  $COP_{therm}$  as mentioned. For segments that do not have at least one value that is greater than or equal to the chiller's known nominal thermal coefficient of performance, we remove these segments from the consideration of further study.

In the segmentation phase, it would be notable that there are possibly very small values of the flow rates on both ends of a segment since the chiller is either near or on its startup or

terminal stage of operation, thus the derivation of the values of the  $COP_{therm}$  segments yields relatively large values near and/or on their endpoints compared with the values when the chiller achieves its steady state of operation. These large values could be wrongly interpreted as abnormal and/or erroneous operation, therefore, we perform the following steps for data sanitation locally implemented on each segment,

1. determine the first and third quartiles of the segment, i.e.  $Q1$  and  $Q3$ , respectively, and let the interquartile range  $IQR = Q3 - Q1$ . Remove all the values of the segment that not in  $[Q1 - 1.5IQR, Q3 + 1.5IQR]$ . In removing the extreme values, we retain around 99.3 % of the segment's original set of values. We shall consider the retained values of the segments as *valid*.
2. For each deleted extreme value of the element  $s_i \in S$ ,  $i \in \{1, 2, \dots, |S|\}$ , update  $s_i$  as follows,

$$s_i = \begin{cases} \frac{s_{i-1} + s_{i+1}}{2}, & \text{if } s_{i-1} \text{ and } s_{i+1} \text{ are valid,} \\ & 1 < i < |S|, \\ s_{i-1}, & \text{if } s_{i-1} \text{ is valid \& } s_{i+1} \text{ is not,} \\ & 1 < i \leq |S| \\ s_{i+1}, & \text{if } s_{i+1} \text{ is valid \& } s_{i-1} \text{ is not,} \\ & 1 \leq i < |S| \\ s_{|i \pm j|}, & \text{if } s_{|i \pm j|} \text{ s.t. } s_{|i \pm j|} \text{ is valid and} \\ & \min\{|i \pm j|\}, \\ & j = 1, 2, \dots, |S| - 1, \\ & 1 \leq i \pm j \leq |S|, \text{ otherwise.} \end{cases}$$

### 3.2 Segment scaling and clustering

As typical as any other machine, chiller's operate on the basis of its need for utility, i.e. when there is a cooling demand from its users. This implies that the segmentation phase would extract segments of varying lengths of operation. Therefore, analysis is not immediately done since there is a difficulty of realizing points of comparisons to obtain a rational result. The discovery of steady states is difficult for short segments due to high variations in values as the chiller enters different phases of operation. However, these short segments and other short subsegments with short-lived behaviour variations are keys to discovering faulty behaviour whenever they do not simulate the expected proper transitions of operational states of the chiller. On another aspect, the prolonged steady state observed in a segment for which there is no abnormal disruption in between enlarges its length and therefore, an idea of partitioning it comes to mind. However, these partitions do not model the startup, steady state, and terminal states of the chiller operation. Comparing a short segment (with a non-faulty operation) which has the same length of one of these partitions is unreasonable.

Bearing in mind the continuity of operation from the evaporation, adsorption, desorption, and condensation cycle that

the chiller undergoes, adjacent values within each segment are related to each other. Therefore, we could easily scale up or scale down the lengths of the segments by linear interpolation (and when necessary, linear extrapolation) of points to expand or compress the segments. The use of linear interpolation and extrapolation stems from the fact that they provide (interpolated) values that are coherent/consistent with the thermodynamic activities that the chiller undergoes throughout its operation. In this research, we use the median length *medLength* (measured in minutes) with respect to the lengths of all the retained segments to individually scale their original lengths either up or down.

Since all the segments extracted from the time series are already in uniform length after the said interpolations, we will obtain an  $m \times nd$  data matrix of segments as input to the X-Means clustering algorithm,  $m \in \mathbb{N}$  is the number of segments retained after all the preprocessing steps. For this study, these segments are from the values of  $COP_{therm}$ . We shall then look at the efficiency of the chiller in providing cooling demand by intercluster, intracluster, and outlier analysis of the  $COP_{therm}$ -based cluster model. We note the following justifications as to the use of clustering for this research, as follows, (a) the database used in this study (like other databases of energy systems) does not contain any record (nor variable) pertaining to faulty or non faulty behaviour; and (b) the currently-available literature, inclusive of the manual, for the adsorption chiller under study only provide theoretically-sound (i.e. conforming to thermodynamic rules, etc) and nominal values for temperature, pressure, energy consumption, etc., while the actual recorded data can show unexpected values due to many other external factors such as the loss of the vacuum while the chiller is in operation, user misuse of the machine, erroneous settings caused either by machine malfunction or failure, among others. For these scenarios, unsupervised learning such as the use of clustering is optimal so as to discover sets of behaviour of the chiller using its historical data.

Finally, different visualizations of the cluster model are produced and presented to domain experts to identify and validate the discovered sets of faulty and nonfaulty behavior of operation. We also extend the analysis to the other time series recorded in the variables of the  $LT$ ,  $MT$ ,  $HT$  circuits, e.g. return and supply temperature values, by using the same cluster model to check the influence of these variables to the efficiency, performance and behavior of the chiller.

## 4 Results and discussion

### 4.1 Implementation of methods on actual dataset

The adsorption chiller [25] that is used in this study also uses water as its refrigerant. Its monitoring data comprises



**Table 1**  $COP_{therm}$  X-Means cluster model

Cluster	Size	%	Cluster	Size	%	Cluster	Size	%
C0	27	20.77	C8	1	0.77	C16	1	0.77
C1	5	3.85	C9	1	0.77	C17	1	0.77
C2	7	5.38	C10	1	0.77	C18	1	0.77
C3	31	23.85	C11	6	4.62	C19	1	0.77
C4	1	0.77	C12	1	0.77	C20	1	0.77
C5	12	9.23	C13	1		C21	1	
C6	12	9.23	C14	1				
C7	16	12.31	C15	1				

data collected by meters from all of the circuits and pumps involved in the entire cooling process. The data we used in this study was collected in every 4-min interval for the year 2011. The preprocessing, segmentation, scaling, clustering and visualization proceeded with the following results,

1. From input data set, we extracted segments using the flow rate information of the  $LT$  circuit and obtained 176 segments with variable lengths. The timestamps corresponding to the first and last data points of the segments are recorded. We then construct the segments of all other variables to have their values corresponding to all the values within the same timestamps of the segments of  $Q_{LT}$ . These variables are the return and supply temperature values, as well as the energy and power and utilization of the three circuits. The algorithms in this research could be used in any of the variables available in the chiller database.
2. Derive the values of  $COP_{therm}$  by first computing for  $Q_{LT}$  and  $Q_{HT}$  using the segments of the temperature variables of both  $LT$  and  $HT$  circuits (as indicated in Sect. 2.2) setting  $k = 4$ .
3. From the computed values of  $COP_{therm}$  of each segment and the chiller's nominal thermal COP [25], i.e. 0.6, we reduced the number of segments to be considered for further study to 130 segments. Further data sanitation of the  $COP_{therm}$  segments was performed as specified in Sect. 3.1.
4. The remaining  $COP_{therm}$  segments were then scaled using their median length of 27 (number of dimensions) or 104 min in duration. Therefore, we obtain the input matrix of  $COP_{therm}$  segments with a dimension of  $130 \times 27$ .
5. Use the  $130 \times 27$  data matrix of the scaled  $COP_{therm}$  segments as input to the X-Means clustering to obtain the different sets of behaviour during the chiller operations. Project the line graphs (or heat maps when needed), box and scatter plots to enable intercluster and intracluster analysis from the X-Means cluster model.

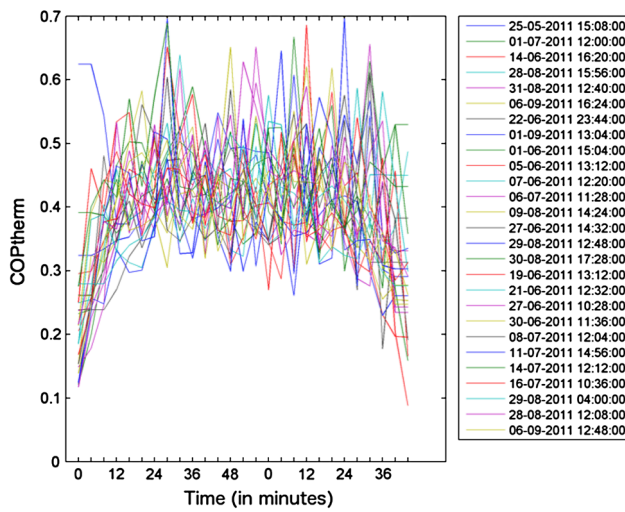
## 4.2 Cluster and outlier analysis

Using the X-Means clustering algorithm, we were able to obtain 8 (nonsingleton) clusters and 13 outliers describing different profiles and temporal variations of the  $COP_{therm}$  as the chiller undergoes its operations. Shown in Table 1 is the summary of the clusters and their sizes, i.e. number of segments belonging to them.

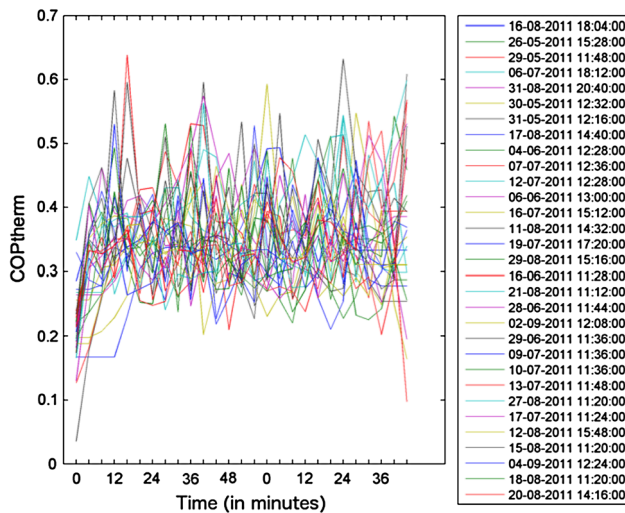
For the sake of brevity, we present in the main body of this paper the intercluster and intracluster analysis of the  $COP_{therm}$  segments belonging to clusters C0, C1, C3, and some of the outliers C9, C12, C17, and C21. Note that the same methods for analysis could be replicated on all other clusters and outliers in the model.

Cluster C3 is composed of the most number of segments in the model, i.e.  $\approx 24\%$ , followed by C0 with  $\approx 21\%$ . Shown in Figs. 2 and 3 are the line graphs of clusters C0 and C3. The x-axis and the y-axis show the time elapsed (where the tick marks are in 4-min intervals) and the  $COP_{therm}$  values observed on each segment (projected as 1 time continuous time series) for these two clusters. The legend on the right side of the figures correspond to the timestamp of the first data point in each of the segments.

From Figs. 2 and 3, it can be observed that majority of these segments follow the regular or the expected (non-faulty) trend of the machine execution. Due to the two-bed reactor structure where the beds alternately function as either adsorber or desorber on each cycle in adsorption chillers, the oscillations of the temperature values throughout majority of the chiller's operation are expected and/or imposed in the process as either cooling or heating is performed on these reactors. Furthermore, both tail ends of each segment would also imply either the short time allocated for the start-up or terminal stage of the chiller's execution as the driving heat is supplied/sustained to meet the cooling demand of the users or or weakened/suspended when this demand is already stopped. These tail ends are more apparent for cluster C0 compared with C3 as seen in the figures.

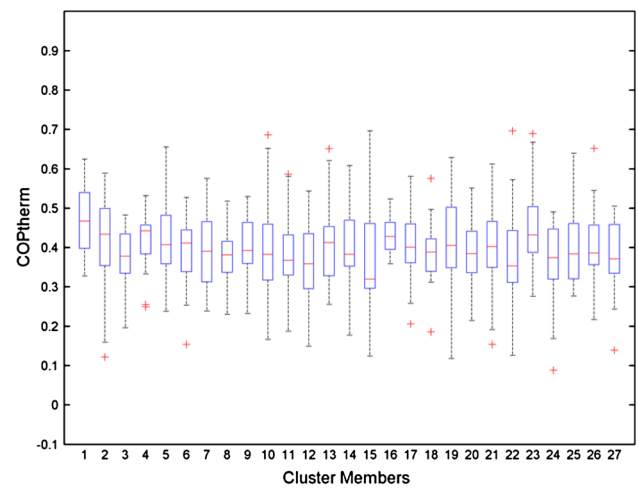


**Fig. 2** Line graphs for cluster  $C0$   $COP_{therm}$  values

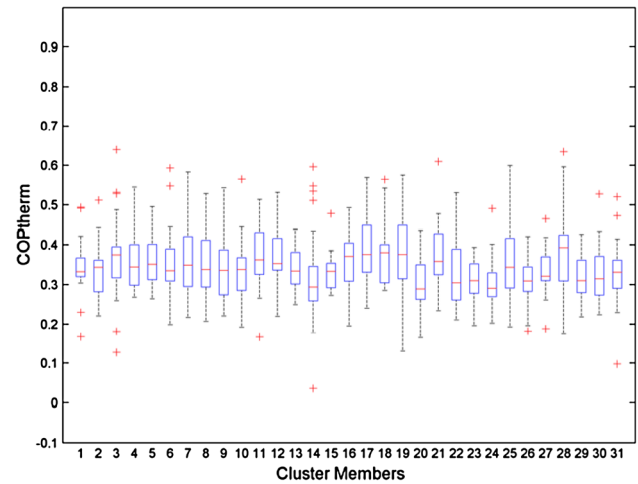


**Fig. 3** Line graphs for cluster  $C3$   $COP_{therm}$  values

Furthermore, the aforementioned figures also show that majority of the  $C3$ 's values are relatively smaller than  $C0$ 's. Notice that the midsection along the y-axis of  $C3$ 's values is relatively lower than that of  $C0$ 's. In fact, the average values of the two clusters are  $\bar{x}_{C3} = 0.3455$  and  $\bar{x}_{C0} = 0.399$ . Moreover, in terms of the general shape of the clusters from the point of start-up to the terminal stage,  $C0$ 's members show a more defined trend on these stages. Another big difference of the values and behavior of the cluster's COPs are the number of outlier values in each of the segments, wherein  $C3$ 's has more compared to  $C0$  as can be seen in Figs. 4 and 5. This implies that even when the high peaks and troughs of the segments in  $C0$  are very apparent in the behavior, these values and behavior are sustained by the chiller for a more prolonged period of time and with regular frequency compared with  $C3$ 's for the entirety of the chiller's operation. The



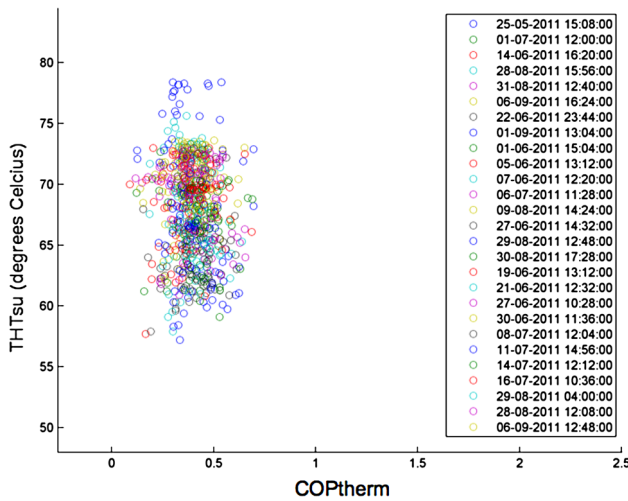
**Fig. 4** Boxplot visualization for cluster  $C0$   $COP_{therm}$  values



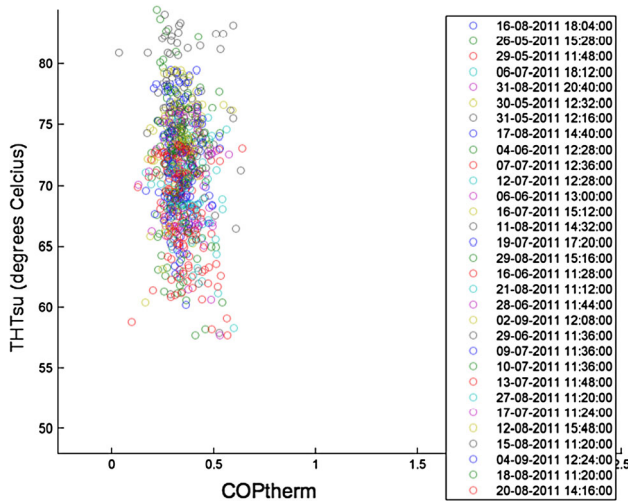
**Fig. 5** Boxplot visualization for cluster  $C3$   $COP_{therm}$  values

larger gap of the values of the troughs and peaks for  $C0$  compared with  $C3$  is highlighted in the scatterplot visualizations in Figs. 6 and 7, respectively. These differences evidently justify the groupings of the segments in their respective clusters. It is notable that the values and trend of both clusters are also relatively similar to each other compared with all other clusters in the model. The fact that these clusters have the majority of the input segments as their cluster members, they could be considered to determine non-faulty operations with COPs observed to be within the mean value and behavior within a year or with slightly larger yet valid COP values.

The scatterplot visualizations in Figs. 6 and 7 provides visual cues with how  $COP_{therm}$  values correlate with the driving heat temperature values of the  $HT$  variable  $THT_{su}$ . Shown in Table 2 are the values of the Pearson product moment correlation  $r$  for  $THT_{su}$  and  $COP_{therm}$  computed for each of the clusters of the model. The values that are



**Fig. 6** Scatterplot for cluster  $C0$   $COP_{therm}$  values against  $THTsu$  values



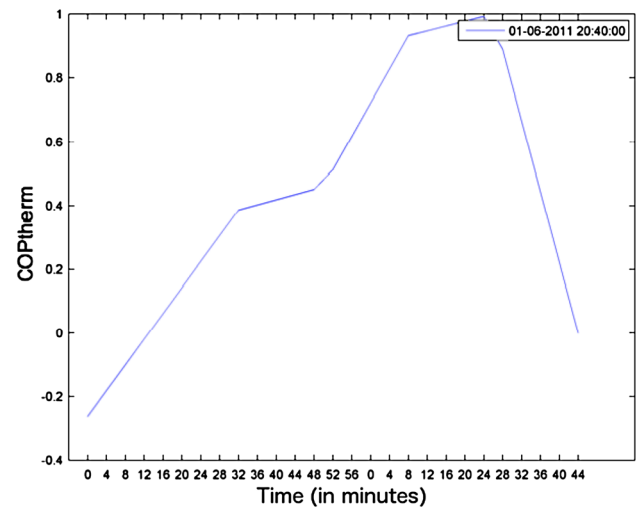
**Fig. 7** Scatterplot for cluster  $C3$   $COP_{therm}$  values against  $THTsu$  values

underlined correspond to outliers. The first two columns of the table are non singleton clusters.

Notice that most of the (nonsingleton) clusters in the first column show negative correlation for  $COP_{therm}$  and  $THTsu$ . For example, these negative correlations are perceivable for  $C0$  and  $C3$  as in Figs. 6 and 7. There are only two nonsingleton clusters that showed positive, though weak, correlation, i.e.  $C2$  and  $C5$ . These positive, yet stronger correlation of the two variables are readily apparent for 9 out of 14 outliers. From among these 9 outliers, there were 4 of them that were identified and validated by domain experts as faults of operation of the chiller, i.e.  $C9$ ,  $C10$ ,  $C12$ , and  $C19$ . Additionally, another outlier was also identified as faulty operation, i.e.  $C21$ , had a negative value for  $r$  with  $COP_{therm}$ .

**Table 2** Correlation values  $r$  for  $COP_{therm}$  and  $THTsu$  values computed per cluster

Cluster	$r$	Cluster	$r$	Cluster	$r$
$C3$	$-0.117181$	$C4$	<u><math>0.129856</math></u>	$C16$	$-0.380267$
$C0$	$-0.063631$	$C8$	<u><math>0.036321</math></u>	$C17$	<u><math>0.846726</math></u>
$C7$	$-0.250765$	$C9$	<u><math>0.725417</math></u>	$C18$	$-0.307854$
$C6$	$-0.409584$	$C10$	<u><math>0.391405</math></u>	$C19$	<u><math>0.866002</math></u>
$C11$	$-0.301475$	$C12$	<u><math>0.437844</math></u>	$C20$	$-0.037083$
$C1$	$-0.155649$	$C13$	$-0.362165$	$C21$	$-0.403705$
$C5$	$0.088842$	$C14$	<u><math>0.168064</math></u>		
$C2$	$0.050553$	$C15$	<u><math>0.467124</math></u>		



**Fig. 8** Line graphs for cluster  $C9$   $COP_{therm}$  values

Shown in Figs. 8 and 9 are the values and trend of the aforementioned identified faults. It can be seen that the typical prolonged oscillating pattern of adsorption chiller cycles are not observed for both of these faults even when  $C9$  and  $C21$  have their original length (i.e. unscaled) of 20 and 12 min, respectively, and continuous flow of the refrigerant is already present. Also notice how these outliers' set of values as shown in Figs. 10 and 11 differ from those of the non-singleton clusters  $C0$  and  $C3$ . It is clear that the COP values are not just abnormal in magnitude but in their trend too as the chiller goes through with its operation for the duration of these outlier segments.

We also refer the reader to the Appendix containing Tables 3 and 4 showing the correlations  $r$  of  $COP_{therm}$  compared independently with the variables of the  $MT$  and  $LT$  circuits. One could replicate the same process of analysis as performed previously on these circuits.

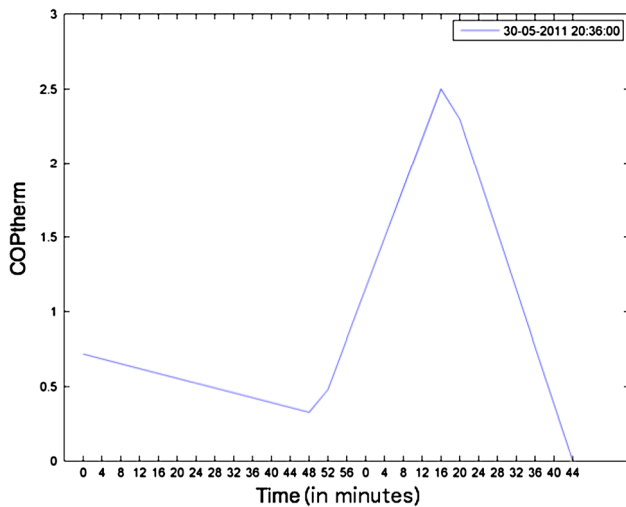


Fig. 9 Line graphs for cluster C21  $COP_{therm}$  values

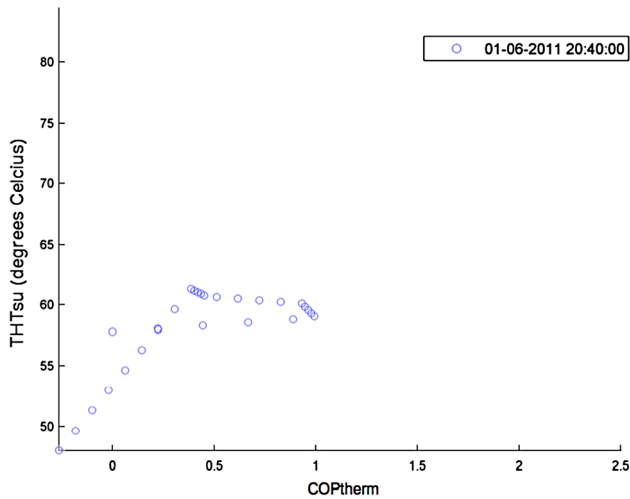


Fig. 10 Scatterplot for cluster C9  $COP_{therm}$  against  $THTsu$  values

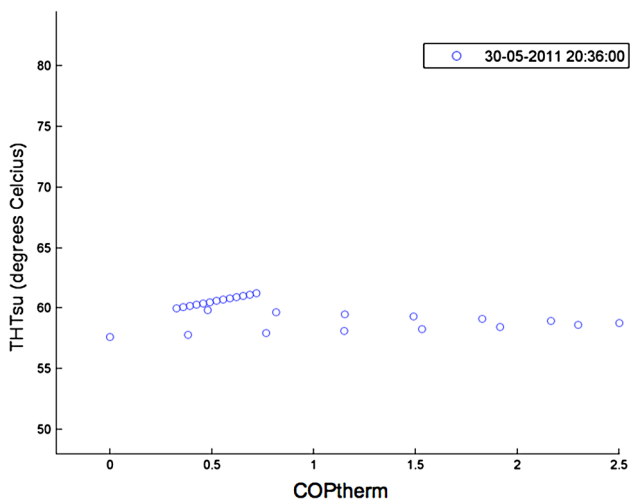


Fig. 11 Scatterplot for cluster C21  $COP_{therm}$  against  $THTsu$  values

## 5 Conclusions and future work

In this research, we were able to provide a framework in dealing with data sanitation, segmentation, clustering and analysis of time series data sets for adsorption chillers whose time and duration of operation are irregular, unequal in length, and with different sets of behavior. We were able to uncover faulty and nonfaulty modes of operation based that were validated by domain experts. We were able to profile these behaviours and establish the correlation of the thermal coefficient of performance with the variables of the  $LT$ ,  $MT$ , and  $HT$  circuits of the chiller. Future work shall focus on identifying states, duration, and values of machine start up, steady states, and terminal stages. We shall also use the proposed framework to analyze electrical coefficients of performance of chillers and compute for energy savings and losses due to faulty operations.

**Acknowledgments** The authors would like to thank Dan Pelleg and the Auton Lab of Carnegie Mellon University's School of Computer Science for the implementation of X-Means used in this research.

## 6 Appendix

Tables 3 and 4.

**Table 3** Correlation values  $r$  for  $COP_{therm}$  and  $TMTsu$  values computed per cluster

Cluster	$r$	Cluster	$r$	Cluster	$r$
C3	-0.186347	C4	-0.746703	C16	-0.681503
C0	-0.116684	C8	0.358187	C17	-0.196419
C7	-0.048101	C9	-0.487606	C18	-0.558702
C6	-0.233954	C10	0.127804	C19	0.094979
C11	-0.291695	C12	0.861950	C20	-0.329952
C1	0.053145	C13	-0.376835	C21	-0.384007
C5	-0.236404	C14	-0.498099		
C2	-0.139959	C15	-0.105373		

**Table 4** Correlation values  $r$  for  $COP_{therm}$  and  $TLTre$  values computed per cluster

Cluster	$r$	Cluster	$r$	Cluster	$r$
C3	-0.209934	C4	-0.388970	C16	-0.230884
C0	-0.076414	C8	-0.757722	C17	-0.269248
C7	-0.423012	C9	-0.869854	C18	-0.573007
C6	-0.686900	C10	-0.617938	C19	-0.536233
C11	-0.394420	C12	-0.285265	C20	0.186395
C1	-0.449397	C13	-0.513003	C21	0.132058
C5	0.029275	C14	-0.461682		
C2	-0.205804	C15	-0.554011		



## References

1. Proklima International, Natural refrigerants: sustainable Ozone- and climate-friendly alternatives to HCFCs", Technical Report, Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ) GmbH—German Technical Cooperation—Programme Proklima Dag-Hammarskjöld-Weg 1-5 65760 Eschborn, Germany (2008)
2. Federal Ministry for the Environment, Nature conservation and nuclear safety, the international climate initiative of the Federal Republic of Germany, Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU) Public Relations Division 11055 Berlin, Germany (2009)
3. Dabrowski A (2001) Adsorption—from theory to practice. *Adv Colloid Interface Sci* 93:135–224
4. Jakob U, Mittelbach W (2008) Development and investigation of a compact silica gel/water adsorption chiller integrated in solar cooling system, VII Minsk international seminar: heat pipes, heat pumps, refrigerators, power sources, minks, Belarus, pp 8–11
5. Nunez T, Mittelbach W, Henning HM (2007) Development of an adsorption chiller and heat pump for domestic heating and air-conditioning applications. *Appl Therm Eng* 27:22052212
6. Liu YL, Wang RZ, Xia ZZ (2005) Experimental study on a continuous adsorption water chiller with novel design. *Int J Refrig* 28:218230
7. Liu YL, Wang RZ, Xia ZZ (2005) Experimental performance of a silica gel–water adsorption chiller. *Appl Therm Eng* 25(2–3):359–375
8. Gong LX, Wang RZ (2012) Experimental study on an adsorption chiller employing lithium chloride in silica gel and methanol. *Int J Refrig* 35(7):19501957
9. <http://www.solair-project.eu/142.0.html>
10. Ayadi O, Aprile M, Motta M (2012) Solar cooling systems utilising concentrating solar collectors—an overview. *Energy Procedia* 30:875–883
11. Fong KF, Lee CK, Chow TT (2012) Comparative study of solar cooling systems with building-integrated solar collectors for use in sub-tropical regions like Hong Kong. *Appl Energy* 90:189–195
12. Bermejo P, Pino FJ, Rosa F (2010) Solar absorption cooling plant in Seville. *Solar Energy* 84:1503–1512
13. Zhai XQ, Wang RZ (2010) Experimental investigation and performance analysis on a solar adsorption cooling system with/without heat storage. *Appl Energy* 87:824–835
14. Helm M, Keil C, Hiebler S, Mehling H, Schweigler G (2009) Solar heating and cooling system with absorption chiller and low temperature latent heat storage: energetic performance and operational experience. *Int J Refrig* 32:596–606
15. Ayadi O, Mauro A, Aprile M, Motta M (2012) Performance assessment for solar heating and cooling system for office building in Italy. *Energy Procedia* 30:490–494
16. Baldwin C, Cruickshank C (2012) A review of solar cooling technologies for residential applications in Canada. *Energy Procedia* 30:495–504
17. Fong KF, Chow TT, Lee CK, Lin Z, Chan LS (2010) Comparative study of different solar cooling systems for buildings in subtropical city. *Solar Energy* 84:227–44
18. Tsoutsos T, Aloumpi E, Gkouskos Z, Karagiorgas M (2010) Design of a solar absorption cooling system in a Greek hospital. *Energy Build* 42(2):265–272
19. Ketjoy N, Yongphayoon R, Mansiri K (2013) Performance evaluation of 35 kW LiBr–H<sub>2</sub>O solar absorption cooling system in Thailand. *Energy Procedia* 34:198–210
20. Boopathi Raja V, Shanmugam V (2012) A review and new approach to minimise the cost of solar assisted absorption cooling system. *Renew Sustain Energy Rev* 16:6725–6731
21. Hang Y, Qu M, Ukkusuri S (2011) Optimizing design of a solar cooling system using central composite design techniques. *Energy Build* 43:988–994
22. Joudi K, Abdul-Ghafour Q (2003) Development of design charts for solar cooling systems. Part I: computer simulation for a solar cooling system and development of solar cooling system design charts. *Energy Convers Manag* 44:313–339
23. Praene JP, Marc O, Lucas F, Miranville F (2011) Simulation and experimental investigation of solar absorption cooling system in Reunion Island. *Appl Energy* 88:831–839
24. Pelleg D, Moore A (2000) X-means: extending K-means with efficient estimation of the number of clusters. In: Proceedings of the seventeenth international conference on machine learning. Morgan Kaufmann, San Francisco, pp 723–734
25. <http://www.kpwsolutions.com.au/media/brochures/KPW-Solutions-Cooling-Data-Sheets>
26. <http://www.gbunet.de/outgoing/nak-prospect>