

# Optimizing Industrial Ventilation Systems by Clustering Workstations into Subsystems

Krystof Litomisky

Spring 2011

## Abstract

I applied well-known hierarchical clustering algorithms to datasets of workstation activity during five-minute intervals recorded over periods of time ranging from 9 days to one month. I grouped together workstations into ventilation subsystems based on the usage data as well as static pressure values measured in the factories. In simulations, I achieved significant energy savings even when the number of ventilation subsystems was small. I provide a procedure to guide factory management in dividing the ventilation system into smaller systems in order to achieve energy savings.

## 1. Background & Related Work

### 1.1. Industrial Ventilation

In any factory where waste is produced, ventilation is necessary for health, safety, and technological reasons. This is the case in a broad range of industries, including woodworking, metalworking, printing, chemical, or the automotive industry. For simplicity, in the following I will talk specifically about removing dust in the woodworking industry. However, all concepts in this report apply equally to all other industries where ventilation is necessary.

Clearly, dust only needs to be removed from workstations that are in operation at any given time.

Statistics indicate that in the woodworking industry, a typical workstation is in use only 11% of the total shift time [1]. However, in practice most factories vacuum dust from all workstations at all times. The advantage of such a system is that it is very easy to set up and maintain. On the other hand, such systems are also extremely wasteful.

In the early 2000s, on-demand systems for industrial ventilation have emerged. These systems vacuum dust only from workstations that are in use at any given time. To do this, they close the ducts to workstations that are not in use using electronic gates. This allows the system to slow down the ventilation fan, which saves electricity.

Energy savings resulting from this can be significant. Hypothetically, if all factories in the United States were upgraded to on-demand systems, 26 billion kWh annually could be saved. This represents a pollution offset of 66 billion lbs. of Carbon dioxide gas annually [1].

## 1.2. Ventilation System Design

Clearly, workstation usage data can only be used when redesigning the ventilation system for a factory has already been in operation for some time. This is not uncommon. Factory managers regularly seek to improve production efficiency and minimize energy use, and the ventilation system plays a significant role in this process. I focus on the case of redesigning existing ventilation systems in this project.

Typically, a ventilation system for a factory is designed by someone closely familiar with the factory. This process usually relies on experience rather than workstation usage data. When workstation usage data is used, it is only in an aggregate form. For instance, a designer might use the percentage of the time that each workstation was in operation over the period of a week. This project is the first to use workstation utilization data at a much finer granularity.

## 1.3. Ventilation Calculations background

Ventilation fan power is given by

$$PWR = \frac{(Q)(P)}{(CF)(\eta)}$$

where Q is the total air volume in cubic feet per minute (CFM), P is fan static pressure in inches water column,  $\eta$  is the fan mechanical efficiency, and CF is a constant conversion factor from (CFM)(inches water column) to horsepower [2]. The value of CF is 6362. The fan efficiency varies minimally (within 1-

2%) with the air volume and static pressure. I consider it as constant in this project. The value of fan efficiency is between 55% and 75%, depending on the type of fan used in a particular system.

The goal of this project is to come up with an optimal clustering of all workstations in the factory into  $k$  clusters, or subsystems. Here, an optimal system is one which uses the least power. Let the set of all workstations in the factory be denoted by  $S = \{x_1, \dots, x_n\}$ , where  $x_j$  is the  $j^{\text{th}}$  workstation, and the workstations are in no particular order. We then divide the system into  $k$  disjoint subsystems,

$$S_i \subseteq S, \text{ such that}$$

$$S_i \cap S_j = \emptyset \quad \forall i \neq j$$

$$\bigcup_i S_i = S$$

Note that any  $S_i$  can only be equal to  $S$  if  $k$  is equal to one. In the sequel, the terms “cluster” and “subsystem” are used interchangeably.

The total power required by the entire system at given point of time is given by

$$PWR = \sum_{i=1}^k \frac{Q_i P_i}{(CF)\eta} \quad (1)$$

where  $k$  is the total number of subsystems (clusters),  $Q_i$  is the required airflow for subsystem  $i$ , and  $P_i$  is the fan static pressure for subsystem  $i$ .

To calculate the total power used by the system over a period of time, the result at each timepoint is first converted to kilowatts (by multiplying it by 0.74). Then, the result is divided by the efficiency of the fan motor (typically about 95%). Finally, the result is multiplied by appropriate time units to calculate the total power consumption in kilowatt hours (kWh).

### 1.3.1. Calculating the required airflow for subsystem $i$ , $Q_i$ , in CFM:

Let the required airflow for subsystem  $S_i$  be denoted by  $Q_i$ . Then

$$Q_i = \begin{cases} 0 & \text{if no workstations are active in subsystem } i \\ \max(CFM_{MIN}, Q_{req}) & \text{otherwise} \end{cases} \quad (2)$$

Where  $CFM_{MIN}$  is the minimum air volume required to transport dust properly in the duct system. I approximate  $CFM_{MIN}$  by

$$CFM_{MIN} = r_{CFM} CFM_{MAX}$$

where  $CFM_{MAX}$  is the maximum theoretically possible air volume in the subsystem.  $CFM_{MAX}$  can be calculated by summing the air volumes required by each workstation in the subsystem when it is on:

$$CFM_{MAX} = \sum_{j=1}^N V_j$$

where  $V_j$  is the air volume required by workstation  $x_j$  when it is in operation.

$r_{CFM}$  is the ratio of maximum and minimum waste transport velocities, which depends on the material transported. The minimum transport velocity is the lowest air velocity at which waste particles are reliably transported in the duct system. The maximum air velocity is given by the physical limitations of the ventilation fan. For the woodworking industry,  $r_{CFM} = \frac{3800 \text{ feet per minute}}{6500 \text{ feet per minute}} \approx 0.6$ . In practice, if not enough workstations are on for minimum required airflow, additional gates are opened to workstations that are not in use. This ensures that waste particles do not accumulate in the ducting.

$Q_{req}$  is the air volume required at any given time by currently running workstations in the subsystem. For any subsystem  $S_i$ ,  $Q_{req}$  is given by

$$Q_{req} = \sum_{x_j \in S_i} g_j V_j$$

Where  $g_j$  is an indicator function, taking value 1 when workstation  $x_j$  is on and value 0 otherwise.

### 1.3.2. Calculating the fan static pressure for a subsystem $i$ , $P_i$ :

The static pressure at a workstation is a function of how far the workstation is from the system fan, the diameter of the duct to the workstation, and the air velocity in the system. The precise calculation is complex, and depends on the minute layout of the workstations as well as the ventilation system. I use the following approximation:

Assign each workstation  $x_j$  a default static pressure value,  $p_j$ . This needs to be done by someone familiar with the layout of the factory and the ventilation system. Pressure values in certain places of the duct system are measured empirically when the system is inspected. In other places, the pressure values must be extrapolated from the existing measurements and the layout of the duct system.

Let  $P_i$  be the fan static pressure for subsystem  $S_i$ . Then

$$P_i = \max(p_j) \quad \forall j \in S_i \quad (3)$$

## 1.4. Optimal Number of Clusters

It is straightforward to see that the selected power equation (equation 1) will be minimized when  $k = n$ , i.e. each workstation is in a cluster by itself. Intuitively, a subsystem with only one workstation never has to increase airflow to maintain minimum required transport velocities (see Equation 2). Also, for a subsystem with only one workstation, the pressure used in the power calculation is the pressure of that workstation, rather than the maximum of the pressures of all workstations currently running on the subsystem (see Equation 3).

However, this is not a realistic situation in an industrial setting. Having a single workstation per fan would require a very large initial investment in the fans and other equipment. Furthermore, by health and safety regulations, ventilation fans must be outdoors. Having a duct from each workstation leading outdoors could be cumbersome and would again require a large investment. Finally, more fans require more maintenance, which is time-consuming as well as expensive.

This means that the power computation function alone is not sufficient to find the optimal clustering of the system. Instead, the value of this function for a particular number of clusters – and, in fact, all the results of the algorithms in this project – should serve as a tool to guide the decision of factory management, which is strongly system-dependent and ultimately subjective. I revisit this problem later in this paper, after I report some experimental results.

## 2. Approach

### 2.1. Hierarchical Agglomerative Clustering

I used the hierarchical agglomerative clustering algorithm. Hierarchical clustering was a natural choice for this application for several reasons. First, the optimal number of clusters is not always known beforehand. Second, the problem lends itself well to a hierarchical structure. For instance, a factory may make two different kinds of products. Machines used to make one kind of product could form one large cluster, while the machines used to make the other kind of product could form another large cluster. For each product, there may be several manufacturing steps, where the workstations used in each step might typically operate at the same time. These could then form smaller clusters within the two large clusters.

In addition, the number of workstations in a factory will always be relatively low, typically less than 100. This means that finding the distance between each pair of workstations, which is necessary for hierarchical clustering algorithms, is not a computational burden.

I used the Manhattan distance,  $\|x - y\|_1 = \sum_i |x_i - y_i|$ , as the distance metric. I made this decision because the Manhattan distance is easily interpretable for this application: my usage features at each timepoint are binary (either on or off). Therefore, the Manhattan distance between two workstations over all timepoints is just the number of timepoints where one workstation is on while the other is off.

Let  $A$  and  $B$  be two clusters of workstations. I tried the following linkage criteria [3]:

- single linkage:  $d(A, B) = \min\{ d(a, b): a \in A, b \in B \}$
- complete linkage:  $d(A, B) = \max\{ d(a, b): a \in A, b \in B \}$
- average linkage:  $d(A, B) = \frac{1}{|A||B|} \sum_{a \in A} \sum_{b \in B} d(a, b)$

Typically, the complete linkage and average linkage criteria produced better results than single linkage.

See section 4.2 for more details.

## 2.2. Power Evaluation

For a particular clustering, the power used by the entire system was evaluated at each timepoint according to equation 1. The total power is simply the sum of these results over all timepoints.

## 2.3. Pressure Factor

In addition to workstation usage data, pressure values can also be used to compute the distances between workstations. As stated in section 1.3.2, the pressure at a workstation is a function of the length of the duct from the workstation to the system fan, the diameter of the duct to the workstation, and the air velocity in the system. As a result, workstations that are close to each other typically have similar pressure values.

To determine how useful this data can be, I gave the pressure values different weights in relation to usage data and evaluated the performance of the resulting clusterings.

## 3. Dataset

There were certain complications getting the workstation usage data. Furthermore, getting reliable values for the workstation static pressures required more input from an expert knowledgeable with the factories than I had anticipated. Therefore, I was only able to test my algorithms on data from 4 different factories, referred to as “R”, “S”, “V”, and “W”, respectively. All factories were in the woodworking industry,

although their specific products are different. The factories have 47, 33, 33, and 10 workstations, respectively.

For each factory, I have:

- Timestamps: usage data collected at five minute intervals over periods of time ranging from one week to one month. At each timepoint, the value is binary: the workstation was either on or off at a particular point of time.
- Pressures: for each workstation, I have either the measured pressure in the duct at that workstation or a value extrapolated from nearby measurements. The values are from systems where all workstations are connected to the same fan.
- Air volumes: the amount of air that needs to be vacuumed from a workstation every unit of time that the workstation is on. This is used in computing the fan power.
- $r_{CFM}$ , fan efficiency, and motor efficiency. These are constants that scale the result of the power computation appropriately. They are the same for all workstations in a particular factory and do not affect the relative performance of different clusterings.

## 4. Experimental Results

### 4.1. The Effect of Pressure

The purpose of this is to evaluate the relative usefulness of the pressure values as compared to the usefulness of the workstation usage data. Let the pressure weight,  $p_w$ , be given by

$$p_w = \frac{\max(p_j, j \in \{1 \dots n\}) p_{mult}}{m}$$

where  $\max(p_j, j \in \{1 \dots n\})$  is the maximum default static pressure value of all workstations in the factory,  $m$  is the number of timepoints, and  $p_{mult}$  is a multiplication factor. I ran the clustering algorithm for different values of  $p_w$ . In each case,  $p_{mult}$  is calculated such that the above equation holds. Figures 1

and 2 show the performance of various clusterings for different values of the pressure weight for two different factories.

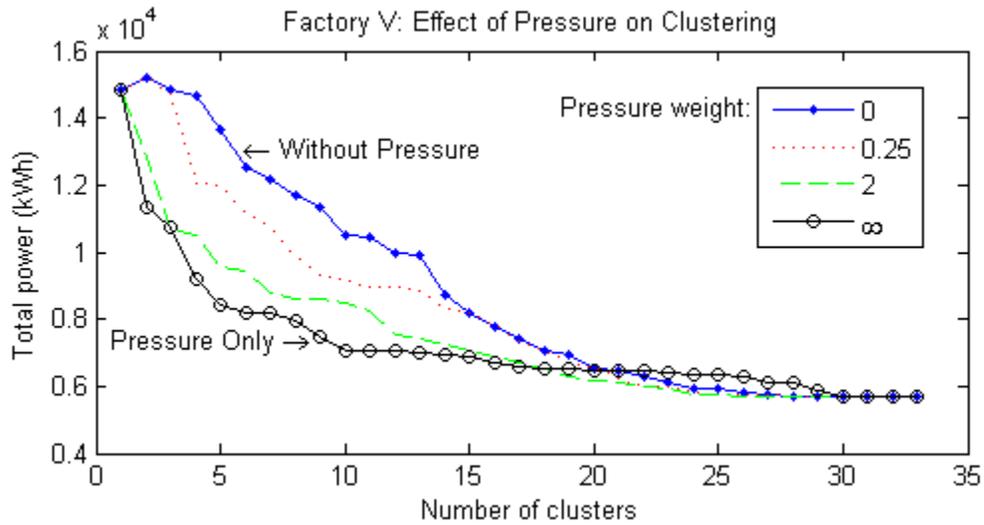


Figure 1: Effect of Pressure on the clustering for Factory V.

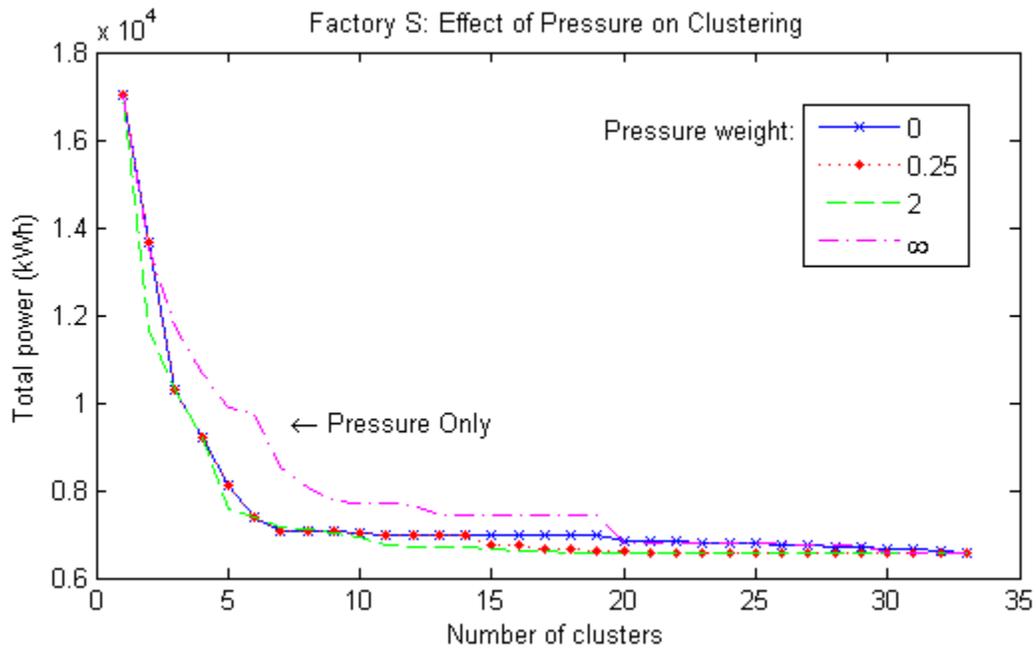


Figure 2: Effect of pressure on the clustering for Factory S.

Figures 1 and 2 indicate that relative importance of the pressure and workstation usage data is system dependent. Factory S is much more automated than Factory V, which is one reason why the timestamp data may be more useful for Factory S.

## 4.2. Linkage Methods

I tried the single, complete, and average linkage criteria, as described in section 2.1. Figure 3 shows their respective performance on one dataset.

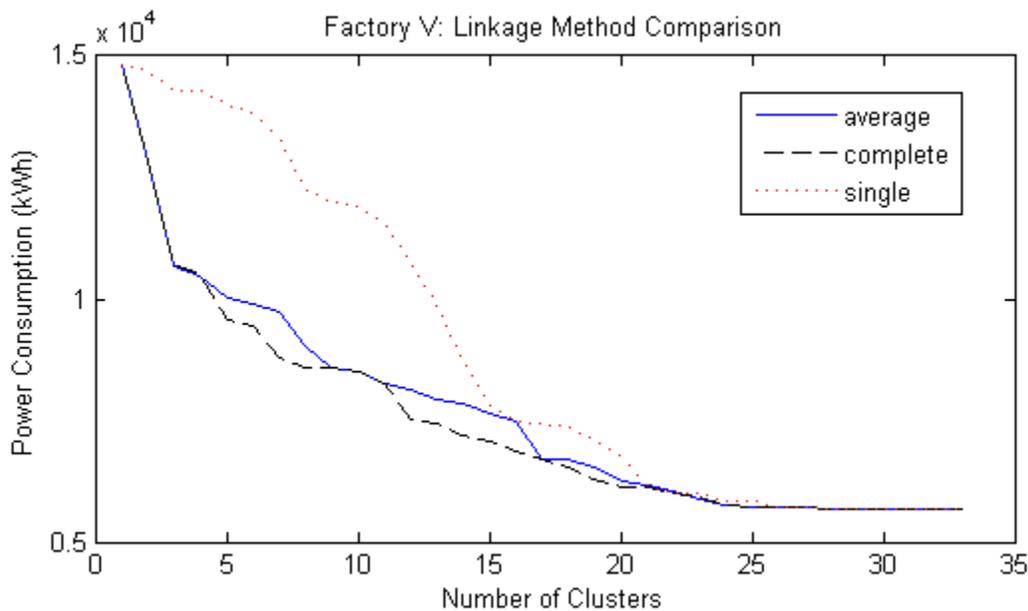


Figure 3: Comparison of linkage criteria.

The result of figure 3 is representative of the entire dataset. On all tested systems, complete and average linkage outperformed single linkage. Differences between the performance of average and complete linkage were always small. I chose to use complete linkage as the default because the algorithm runs faster.

Figure 4 shows the dendrogram (hierarchical clustering) of a particular factory. This clustering used the complete link linkage criterion and pressure weight  $p_w = 2$ .

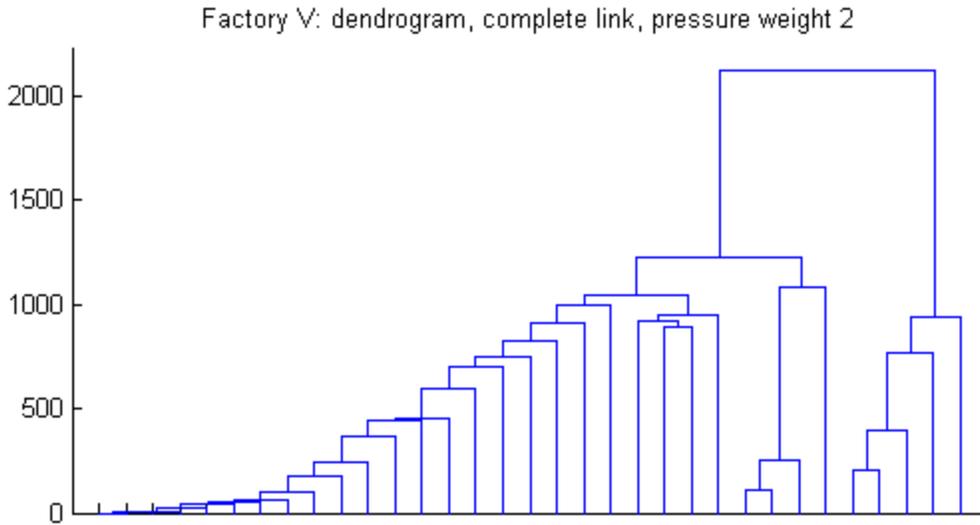


Figure 4: Workstation dendrogram.

### 4.3. Cluster Validity

There is no best way to verify the validity of any particular clustering for this application. Considering how the total power function changes is a good way to see whether adding one more cluster improves the system performance significantly.

It can also be informative to consider how different the clusterings are using different parameters. For instance, we may be interested in how increasing the pressure weight  $p_w$  changes the clustering. Even more importantly, it would be informative to see how the clustering changes based on using workstation usage data from a different time period.

To do this I use the Rand statistic. The Rand statistic can be used to compare a pair of different clusterings, say C and P. The Rand statistic is given by [3]

$$R = \frac{a + b}{a + b + c + d}$$

where (a) is the number of pairs of points that are in the same cluster in C and in P, (b) is the number of pairs of points that are in different clusters in C and in P, and (c+d) is the number of points that are in the

same cluster in one clustering but in different clusters in the other clustering. Figure 5 shows the Rand statistic for different pressure ratios (left) and for clusters generated using data from different time periods (right).

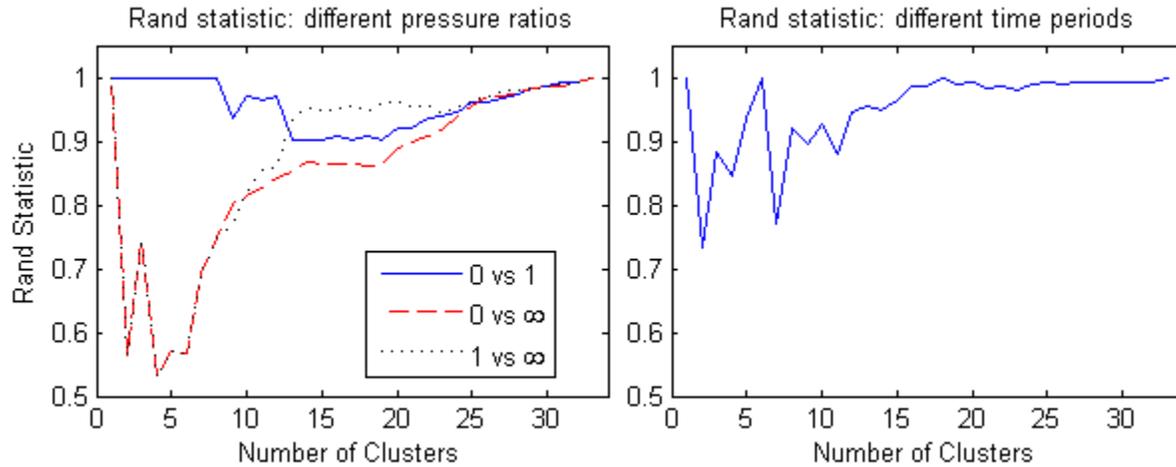


Figure 5: the Rand statistic for different pressure ratios  $p_r$  (left) and for timestamps over different time periods (right). Both results are obtained on the dataset for Factory S. On the graph on the right, the pressure weight  $p_w = 2$ .

Figure 5 (left) shows the effect of pressure on the clustering. While the clusterings with the pressure weight  $p_w$  equal to 0 and 1 match closely for all values of  $k$ , there is a large difference between those two clusterings and that created by using the pressure only. Figure 5 (right) shows that the clusterings created using data from two different time intervals match reasonably well.

Figure 6 shows the layout of the workstations in factory V for different numbers of clusters. It is beyond the scope of this project to comment on how desirable any particular clustering is in terms of the physical layout.

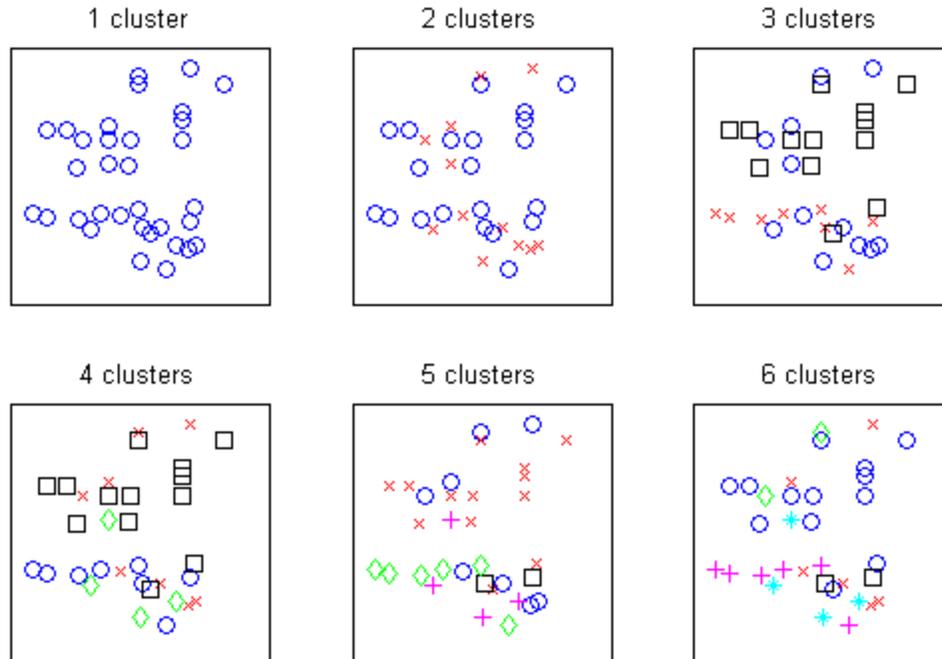


Figure 6: clustered layout of the workstations in factory V. Workstations on the same ventilation subsystem are represented by the same symbol.

## 5. Results

The experimental results indicate that the optimal number of clusters to group workstations into is not only different for different factories but also a matter of subjective opinion. Furthermore, the value of the pressure weight  $p_w$ , which leads to optimal clusterings also depends on the system. Neither of the above conclusions is surprising.

However, the experimental results showed significant savings for all of the datasets with even a few clusters. This ultimately leads to a heuristic algorithm to guide factory managers in deciding an optimal clustering of workstations into ventilation subsystems:

1. Evaluate the effect of including pressure in the cluster computation. To do this, include the pressure values multiplied by representative values of the pressure weight  $p_w$  in the computation of distances from each workstation to every other workstation, as described in

section 4.1. Some good candidates to try for  $p_w$  include 0, 0.25, 1, and 2. Also try the clustering algorithm considering only the pressures.

2. From the resulting plots (of the same type as those in figures 1 and 2), choose values for the pressure weight  $p_w$  and for the number of clusters  $k$ . Select  $p_w$  such that the total power at the particular  $k$  of interest is the lowest of all  $p_w$ s for that  $k$ .

The selection of  $k$  is highly subjective. When making this decision, consider the following:

- $k$  should be a small number. Adding an extra cluster means adding an extra ventilation fan, which is expensive. This also increases ducting costs, maintenance costs, etc.
  - The presence of a “knee” in the graph is an indicator of a good  $k$ . This is a value of  $k$  such that for the next several higher values of  $k$ , the total power consumption does not improve significantly.
  - Future plans. For instance, if a company plans to expand in the near future, it might be acceptable to choose a  $k$  somewhat larger than would otherwise be recommended.
3. Run the clustering algorithm for the selected value of  $p_w$  and use the clustering for the chosen value of  $k$ .
  4. **IMPORTANT:** Simulate the resulting system in detail. This must involve a specific layout of the entire ventilation system, including the ducting, the locations of the workstations, and the locations of the fans. If satisfied, done. Otherwise, go to step 2 and repeat the procedure for different  $k$  and/or  $p_w$ .

## 6. Future Work

This project has demonstrated that well-known clustering algorithms can be applied to detailed workstation usage data to produce a number of smaller ventilation systems that minimize total energy use.

However, this project used estimations for workstation static pressure obtained from measurements of a system with only one fan. These pressures could be different in smaller systems, because the layout of the ventilation system would change. If different, the pressure values should be lower in a smaller system, since in a smaller system the fan could be placed closer to its workstations. However, this should be verified either by physical experiments or by computer simulations using the precise layout of the entire ventilation system.

Furthermore, the algorithms described in this project should also be run on a significantly larger number of datasets. This would allow more refinement to the methods and would also give greater significance to the results.

## 7. Conclusions

I applied well-known hierarchical clustering algorithms to industrial workstation usage data. This allowed me to create a number of smaller ventilation systems. For all factories considered in this project, clustering the workstations into subsystems decreased the total energy use of the system even for a small total number of clusters. This represents a good energy-saving opportunity. In order for the results of this project to be verified, the clusterings that result from the algorithms used here should be tested in simulations which involve the precise layout of the ventilation system.

## References

- [1] A. Litomisky. Innovative Energy Efficient Industrial Ventilation. Energy Systems Laboratory (<http://esl.tamu.edu>), 2005.
- [2] Industrial Ventilation: A Manual of Recommended Practice, ACGIH ([www.acgih.org](http://www.acgih.org)). Page 6-15, formula 6.4.
- [3] S. Theodoridis and K. Koutroumbas. Pattern Recognition. Academic Press, 4<sup>th</sup> Ed. 2008.