

### **3. Is it really true that the change between days is larger than the the changes within a day? That is surprising, and suggest that the pattern has more to do with some longer term trends.**

It seems to depend on the day. In an effort to summarize 24 hour coefficient segments, I generated the below plots with this process:

Using Model 2 10-minute data

- Drop sparse packages without any complete days
- Drop engine starts and engine hour tags
- Drop records with any nulls
- PCA on remaining sensor data
- For each of the top 4 eigenvectors
  - Compile all complete 24 hour data samples, vectors of length 144
  - Run KMeans on the samples for 5 to 225 clusters (step 5)

For each of the top 4 eigenvectors there is a cluster distribution histogram and a 15x15 grid of sparklines showing the general 24-hour shape of each of the 225 clusters. The sparkline grid is ordered left to right, top to bottom from largest cluster to smallest. All plots share an X and Y axis.

Model 2 was chosen because they all have the same general use-case as power generators and the data set is more complete than Model 1. The amount of 10 minute data was reasonable to work with on a laptop.

Starting with Eigenvector 1, there appears to be three main groups of patterns.

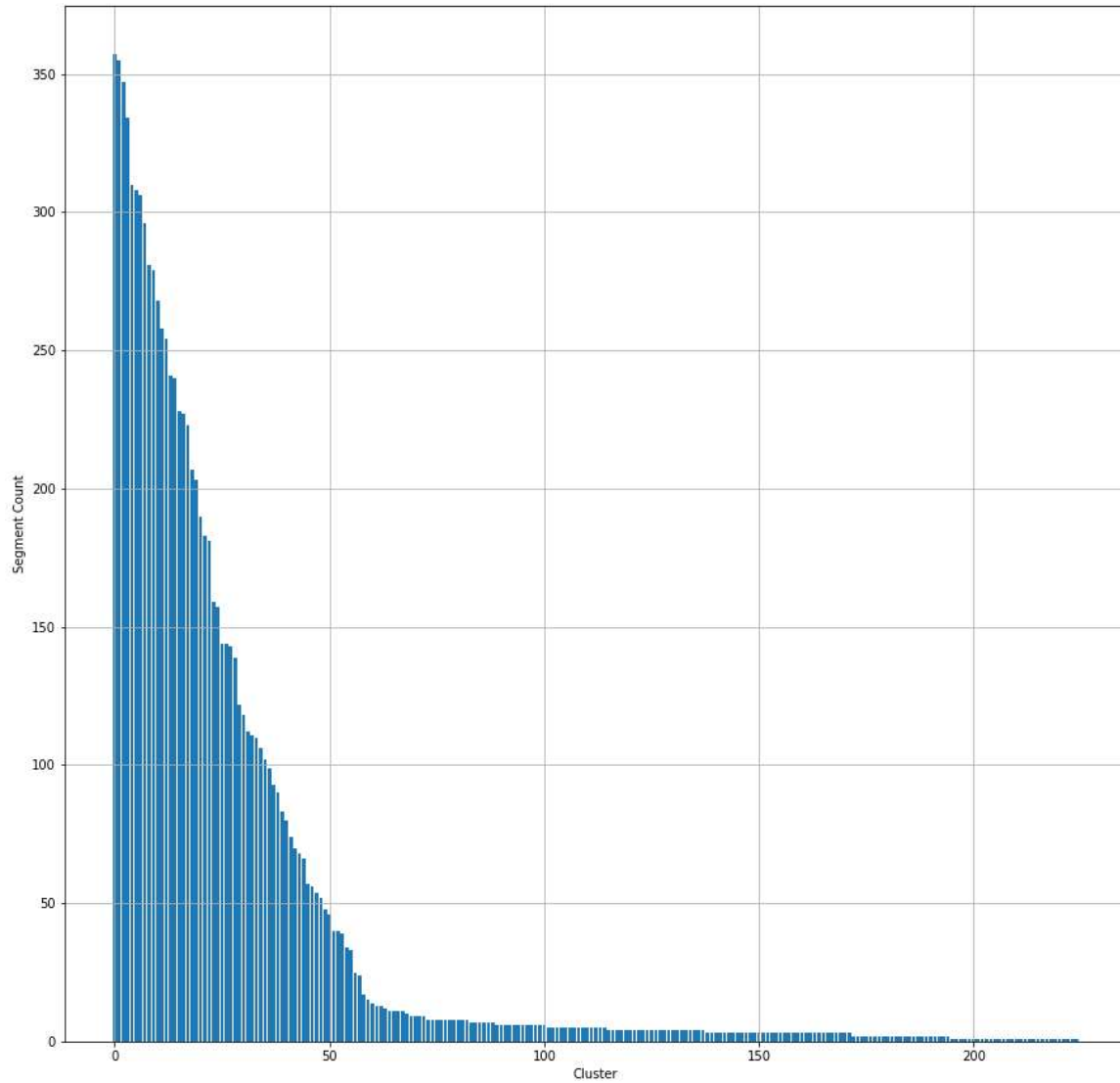
- The largest one dominates the top 3 rows of clusters and likely represents normal operating. Some of these groups show regular 24-hour patterns, visible but not highlighted in these plots. Others are more flat, staying in the same range all day.
- The second group shows abrupt shifts in the coefficient. These are the long tail of the cluster distribution. Both drops and rises in value all seem to share similar shape, just occurring at different times of day and lasting different lengths. We're hypothesizing these represent transient states (ie, the package moving from low operational mode to high).
- The third group is of highly erratic values and only just starts showing up in the last 3 rows of clusters. Many segments fitting into this pattern are still clustered in the top 3 rows and chaotic enough to not get grouped together.

Eigenvectors 2 through 4 show the same patterns but with the second, abrupt pattern group showing less and less influence.

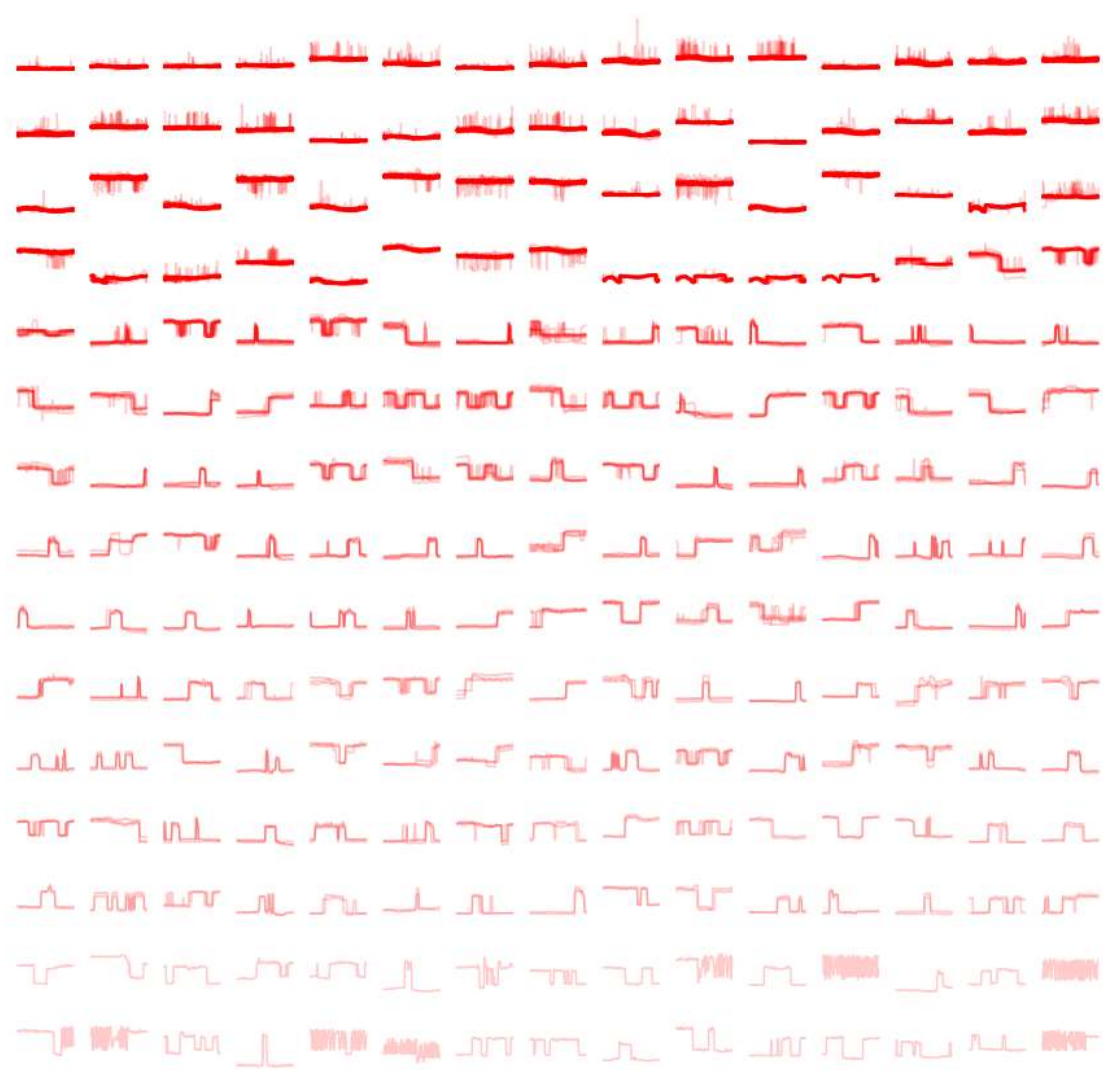
We propose each type of pattern be modeled separately.

## Eigenvector 1, Clustering 24-hour coefficient segments

Model 2 Cluster Distributions for Eigenvector 1, 24 Hour Profiles

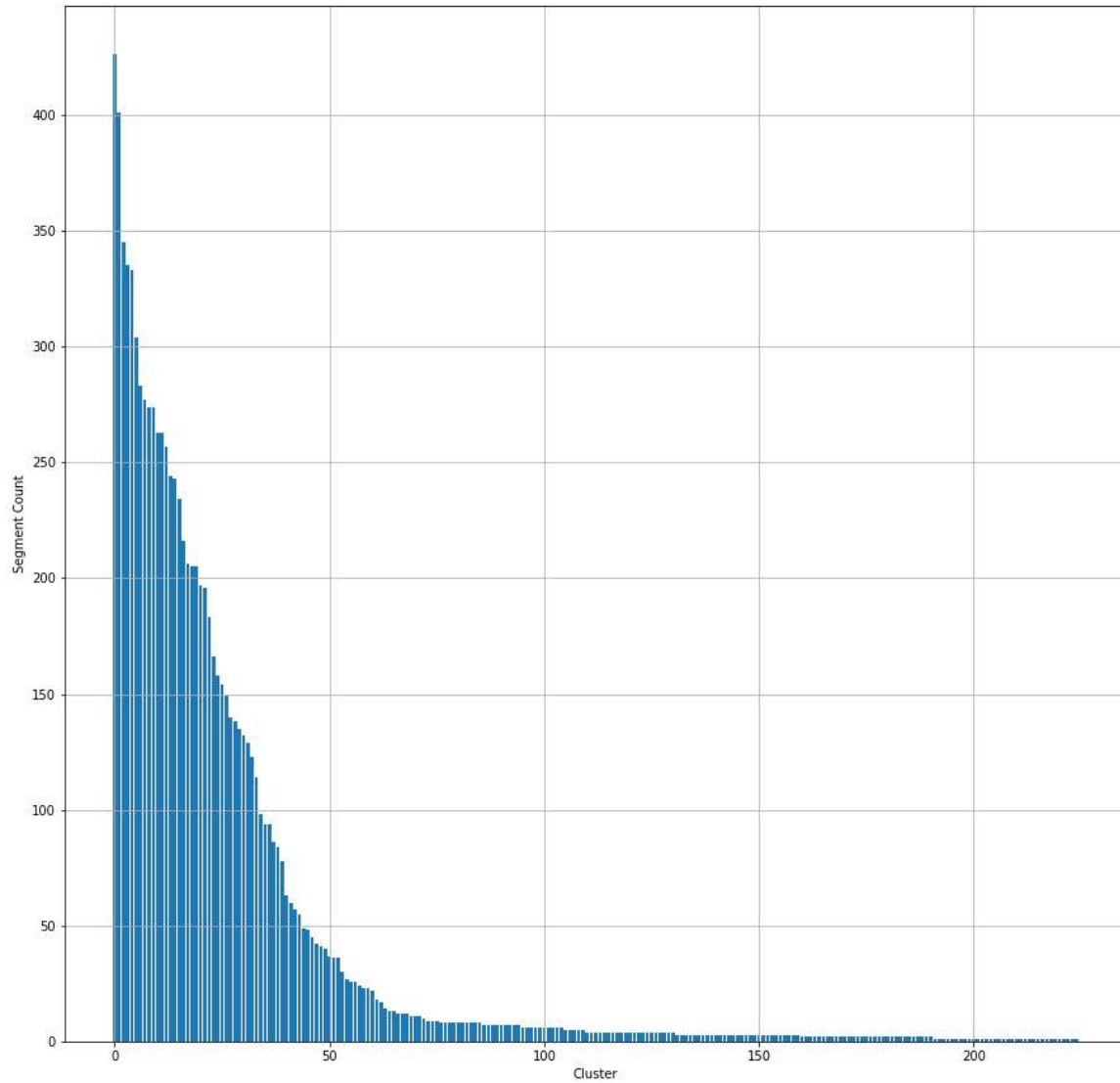


Model 2 Eigenvector 1, 24 Hour Profiles

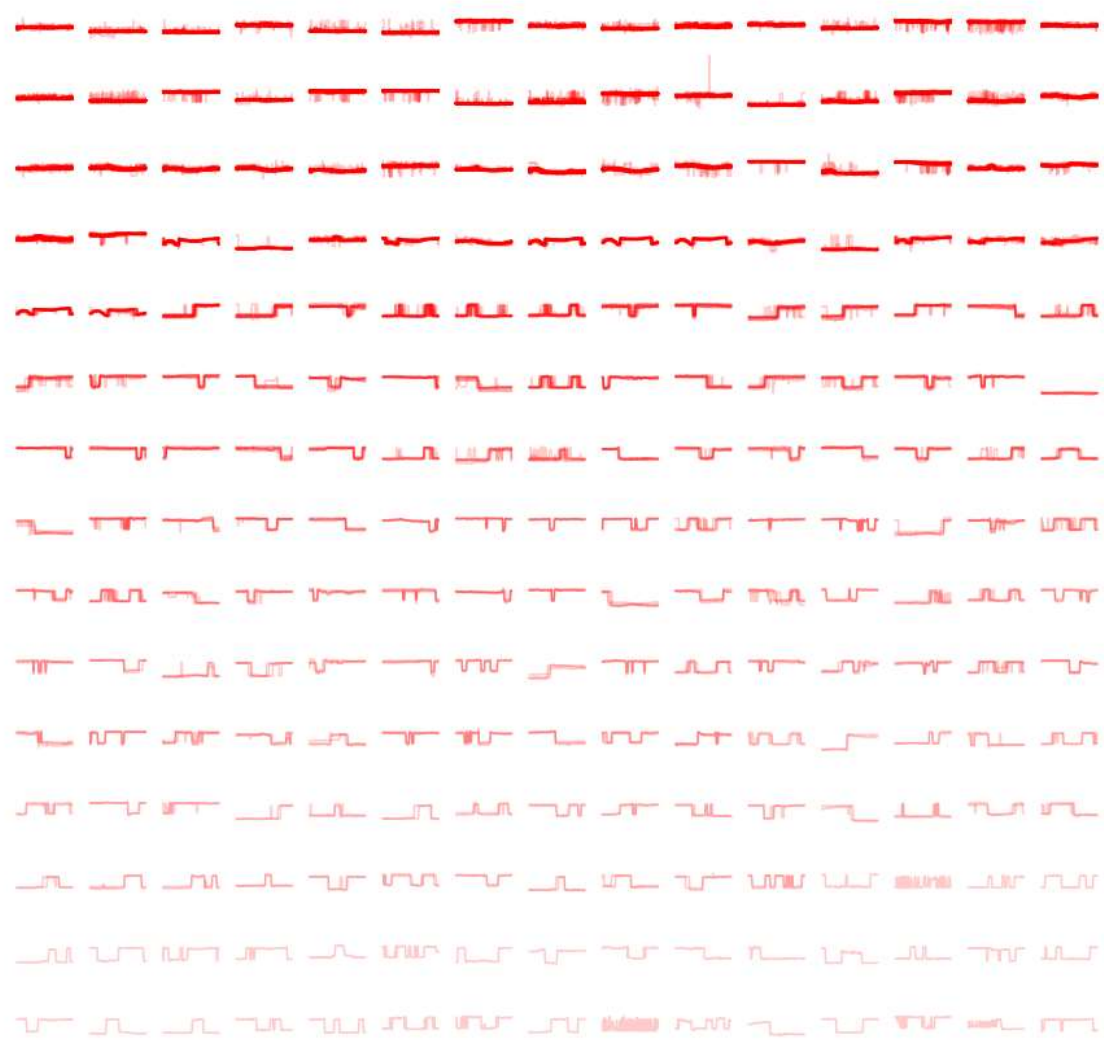


## Eigenvector 2, Clustering 24-hour coefficient segments

Model 2 Cluster Distributions for Eigenvector 2, 24 Hour Profiles

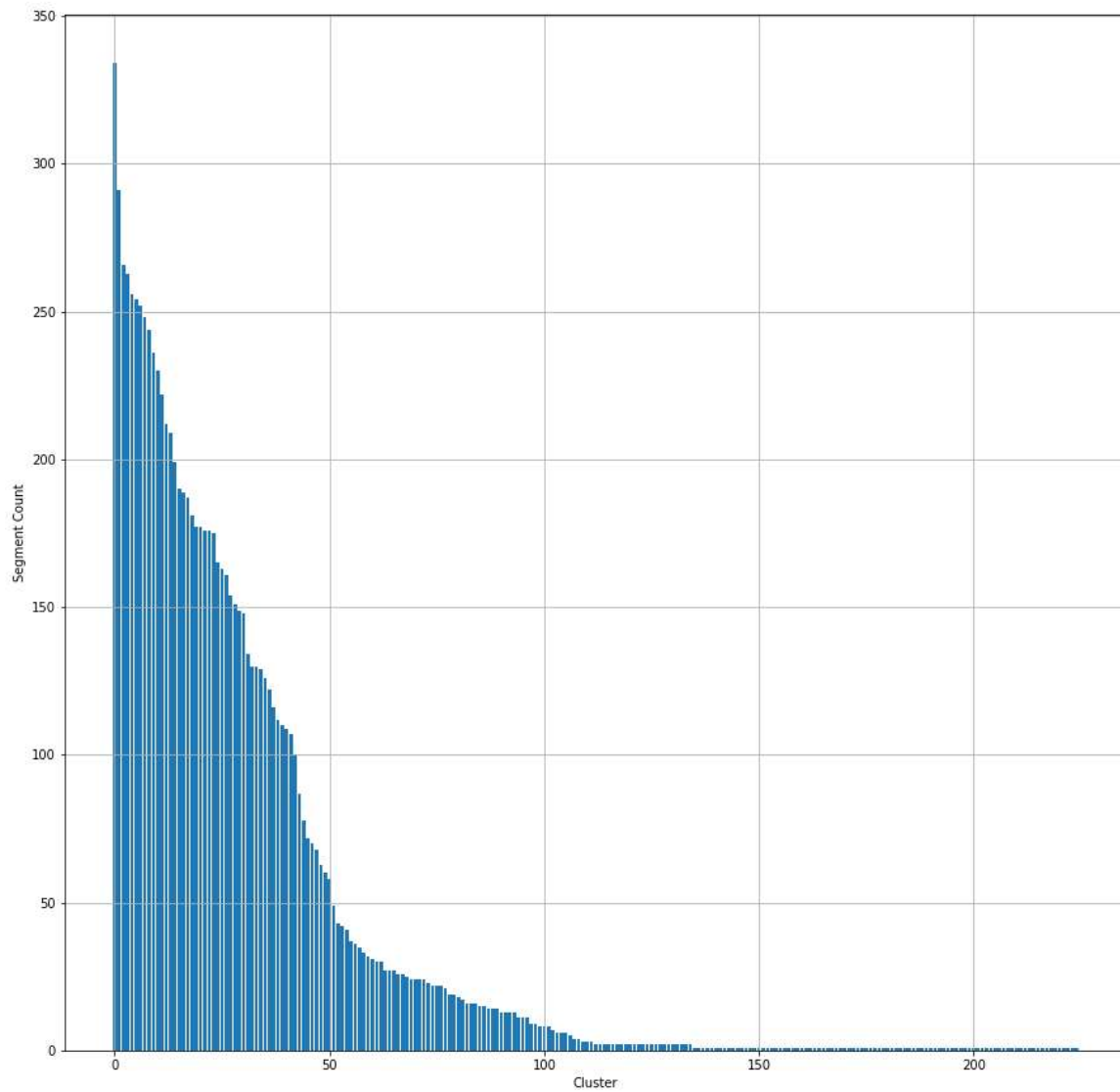


Model 2 Eigenvector 2, 24 Hour Profiles

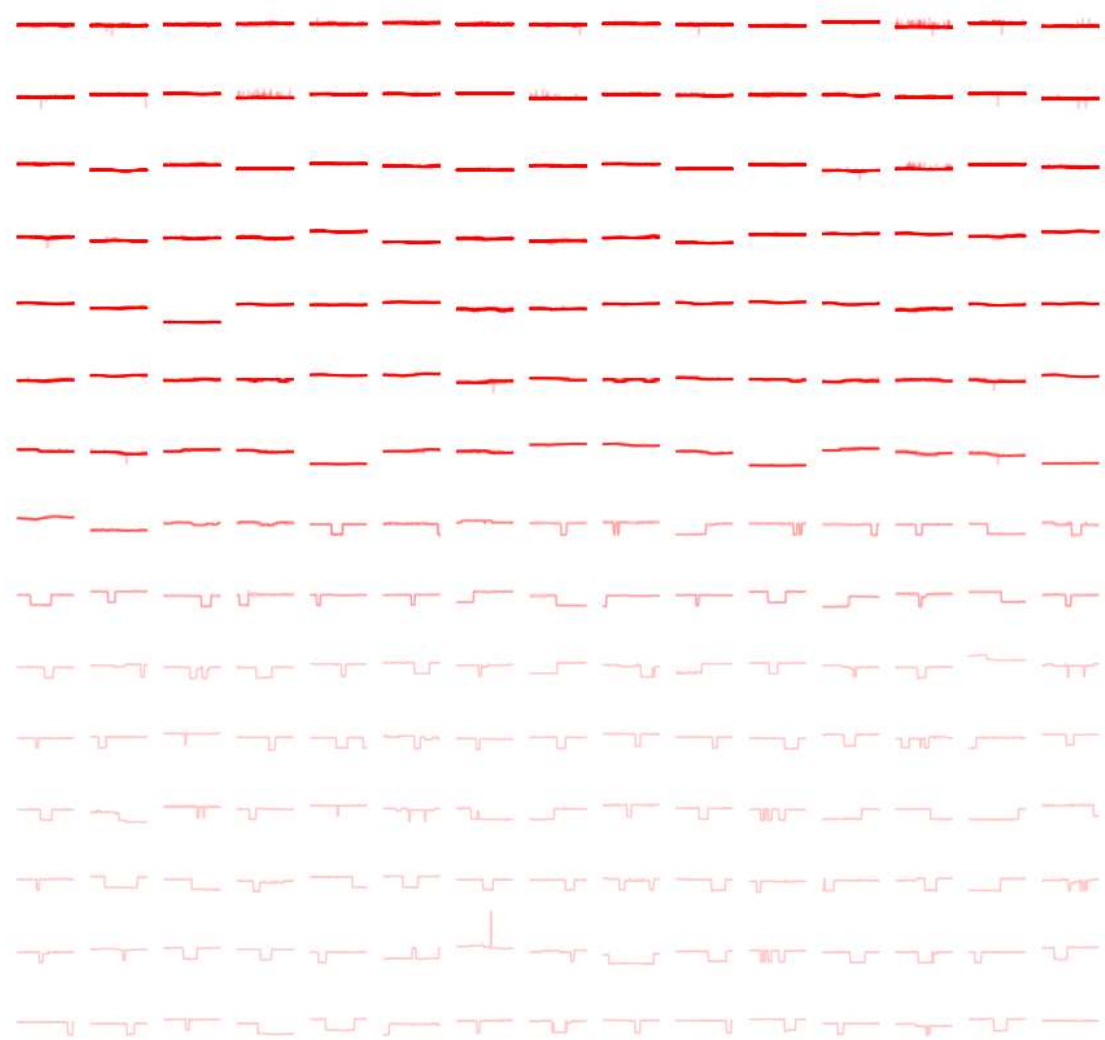


## Eigenvector 3, Clustering 24-hour coefficient segments

Model 2 Cluster Distributions for Eigenvector 3, 24 Hour Profiles

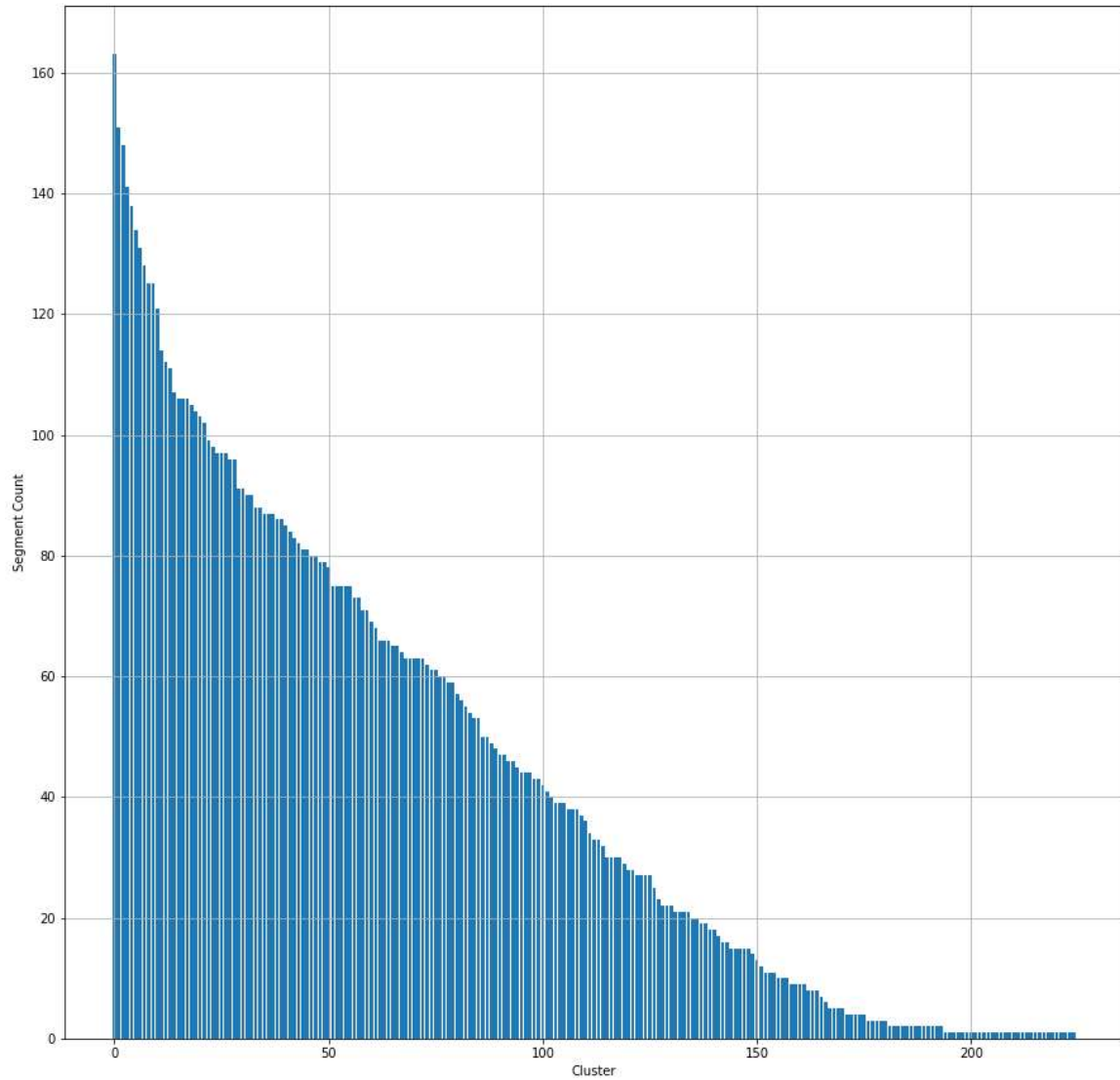


Model 2 Eigenvector 3, 24 Hour Profiles



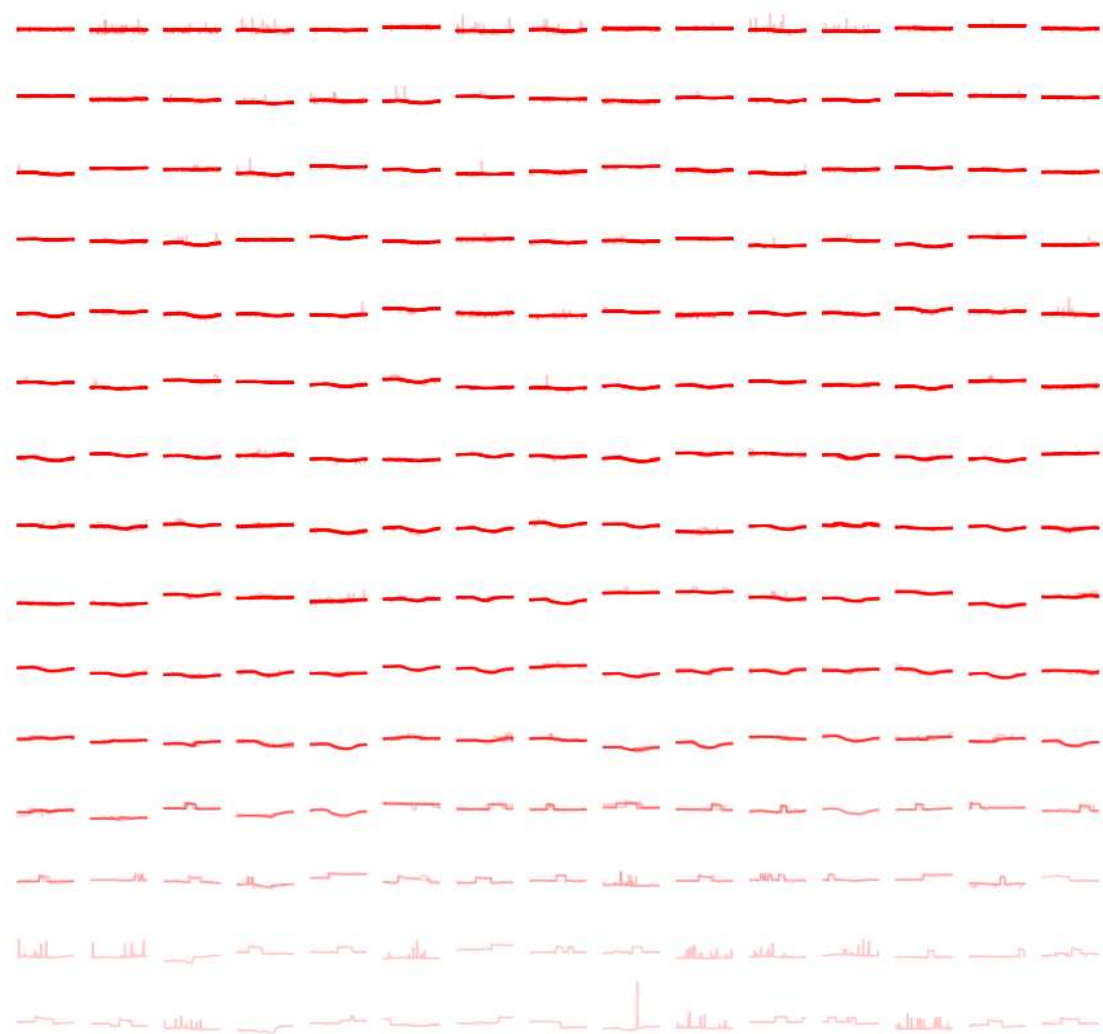
## Eigenvector 4, Clustering 24-hour coefficient segments

Model 2 Cluster Distributions for Eigenvector 4, 24 Hour Profiles





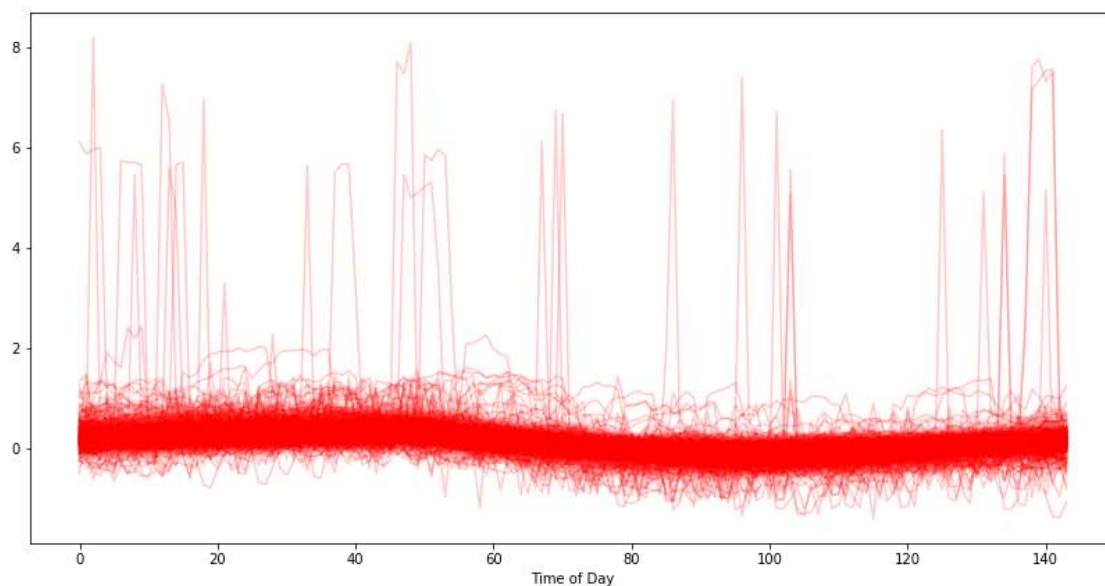
Model 2 Eigenvector 4, 24 Hour Profiles



## Sample 24 hour Clusters from Eigenvector 1

These plots are full-sized examples of clusters from Eigenvector 1. The first shows what we consider normal operating mode. The second shows a cluster with four mode changes.

Model 2 Eigenvector 1, 24 Hour Cluster 4



Model 2 Eigenvector 1, 24 Hour Cluster 63

