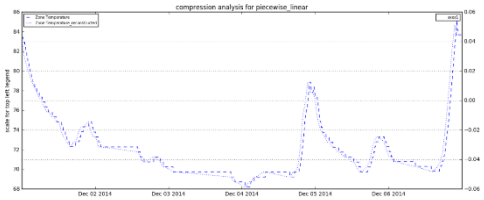
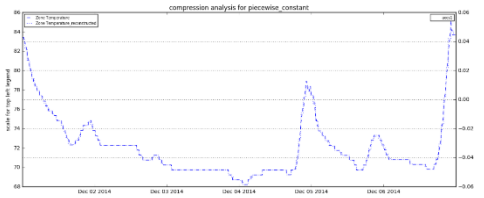
# Compression analysis on Building Sensor Data


## Abstract

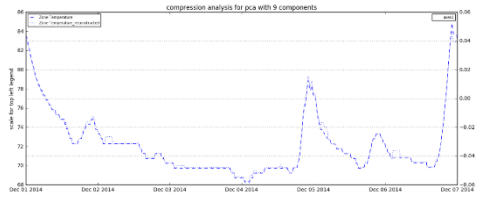
Lot of data from building sensors is being recorded and used in analysis for effective power consumption and detecting faulty sensors. Analysis often get complex due to the availability of huge data. It is important to perform analysis on non-erroneous data. Compression plays an important role in reducing the data size while also maintaining low error. Piecewise linear, piecewise constant and PCA are discussed below.


**Piecewise linear:** In piecewise linear model, n-dimensional data is approximated to several 2-dimensional data. Tolerance or maximum gap parameter can be used to tune / control the error. This is performed on single signal and is not dependent on other related signals.




**Piecewise constant:** In piecewise constant model n-dimensional data is approximated to several threshold values. Like piecewise linear model this is also performed on signal and it not dependent on other signals.




**PCA based compression:** In PCA based compression all the signals are taken into consideration. Number of features that explain the most variance is taken into consideration and used in reconstruction.

## Comparative analysis

| | | | rmse | compression_ratio |
|---|---|---|---|---|
| room | signal | model | | |
| Rm-4226 | Zone Temperature | piecewise_linear | 0.475475 | 0.940912 |
| | | piecewise_constant | 0.047159 | 0.941474 |
| | | PCA_9 | 0.054394 | 0.500000 |
| | | PCA_7 | 11.061083 | 0.611111 |
| | | PCA_4 | 11.074878 | 0.777778 |
| | Supply Airflow | piecewise_linear | 5670.052855 | 0.895329 |
| | | piecewise_constant | 586.808404 | 0.899831 |
| | | PCA_9 | 0.054394 | 0.500000 |
| | | PCA_7 | 11.061083 | 0.611111 |
| | | PCA_4 | 11.074878 | 0.777778 |

Comparative analysis is made on two signals (Zone Temperature, Supply airflow) from two rooms (Rm-4226 and Rm-3114) with 5 models. Piecewise linear, Piecewise constant, PCA with 4, 7 and 9 components. X-axis show the compression ratio and the y-axis show the log value of RMSE between original and reconstructed data.