# 2017 AEIC WLRA

## Inference from Smart Meter Data Using the Fourier Transform

Prasenjit Shil, PhD, Ameren

Tom Anderson, SAS Institute Inc.

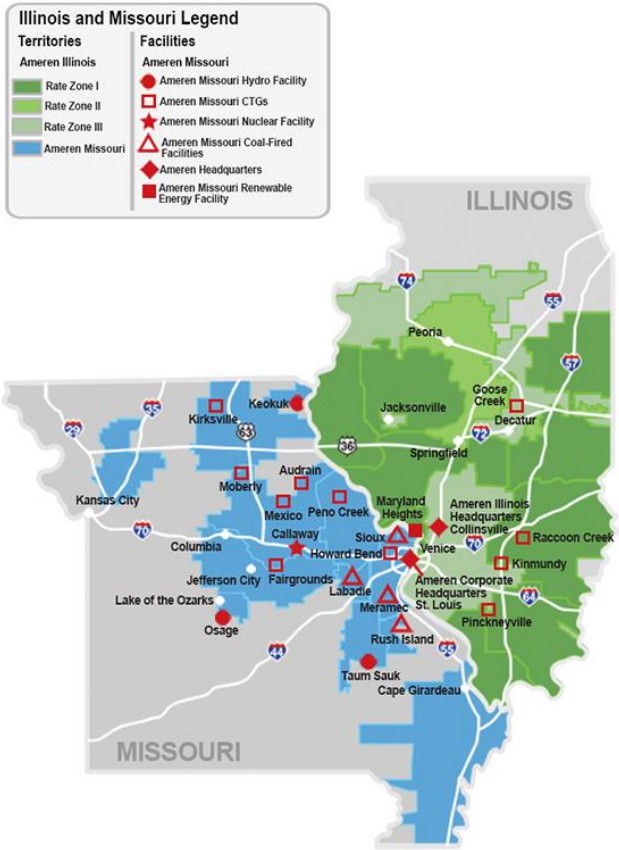
# Presenter

**Prasenjit Shil, Sr Forecasting and Load Research Specialist, Ameren**

**Tom Anderson, SAS**

- **Prasenjit Shil**, PhD is a Sr. Forecasting and Load Research Specialist at Ameren, a St Louis based utility company. Prasenjit's work involves short term and long load forecast, analytics to support rate case and strategic projects, quarterly sales variance analysis for CFO, customer segmentation for energy efficiency marketing, call center analytics, and process improvements through data analytics. He was part of the analytics strategy team and data governance team at Ameren. Prasenjit has been using SAS since 2002.
- **Tom Anderson,** Principal Systems Engineer with the SAS US Energy Division. 20+ years of analytical experience includes 15 years with SAS concentrating on advanced analytics and data management in both Utilities and Oil and Gas. As a leader for developing technology solutions to smart grid initiatives, Tom addresses the challenges that utilities are having around data management. The challenges result from new infrastructures that provide more data points than ever before. Recent achievements include solutions in Asset Performance Analytics in both upstream and downstream O&G applications as well as Advanced Meter Infrastructure analysis and application development within electric utilities.

# Ameren at a Glance



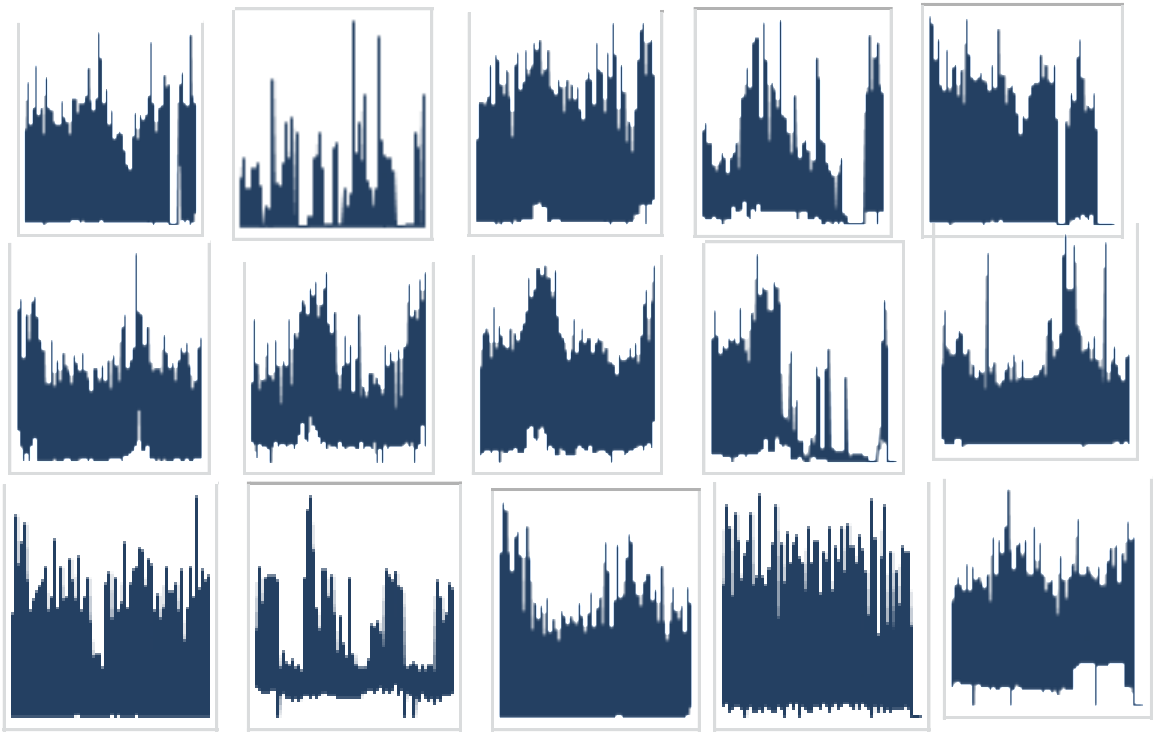

- Ameren Missouri, Ameren Illinois and Ameren Transmission Company
- 2.4 million electric customers and more than 900,000 natural gas customers across 64,000-square-mile area
- Ameren Missouri ranks as the largest electric power provider in Missouri, and Ameren Illinois ranks as Illinois' third largest natural gas distribution operation in total number of customers
- Ameren Missouri has 10,200 MWs of Generation capacity.
- Ameren's rates are some of the lowest in the nation

# Introduction

## Fast Fourier Transformation (FFT) on Smart Meter Data

- This presentation will demonstrate how Fast Fourier Transformation (FFT) can be applied on smart meter data for enhanced customer segmentation and discovery.
- Smart meters or Advanced Metering Infrastructure (AMI), provides utilities with an explosion of data creating vast analytical opportunities to improve reliability, customer satisfaction, and safety.
  - However, the data explosion comes with its own challenges.
    - **Volume:** Consider that just 20,000 customers with AMI data can reach over 300GB of data per year. Simply aggregating the data from minutes to hours or even days can skew results and not provide accurate segmentations.
    - **Bad data**: Outliers caused by missing or incorrect reads, outages or other factors must be addressed.
- FFT can help identifying such “noise”.
  - The proposed framework is expected to identify various customer segments
    - Use: Demand response programs, marketing programs, error or bad data detection, outage detection, diversion or fraud or failing meters detection (revenue protection)

# Smart meter data provides volumes of “good” and “bad” quality interval data



- These charts represent 15 minute interval data between Jan 1,2015 and July 31, 2016 (load research data from randomly sampled customers).
- Indicate interval data quality can be perceived as “good” or “bad”
  - Without investigating each customer profile, how do we identify whether those spikes, dips or flat portion in the chart represent correct profile?
- Imagine identifying such outliers from millions of rows of data everyday

# Energy consumption follows various cycles

- Example of four different types of consumers
  - These segments of profiles highlight that customers can be clustered into various profile segments depending on numerous criterion, including usage pattern
  - Energy consumption follow seasonality, and are cyclical in nature

**Figure A: Working family**

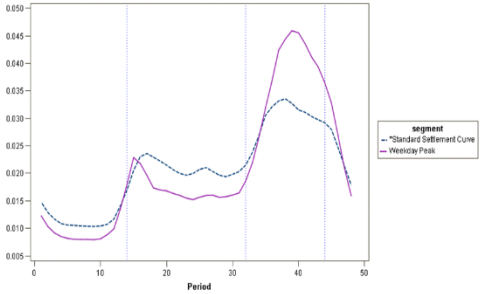



**Figure B: School Age Family**

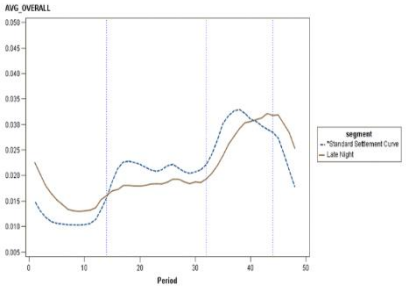



**Figure C: Late Night Gamer**

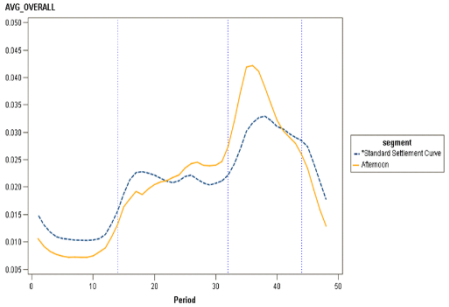
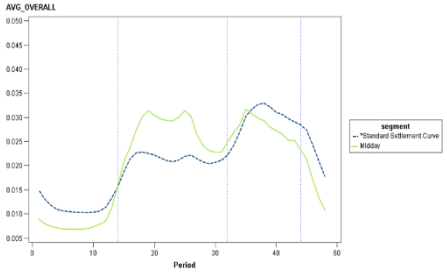
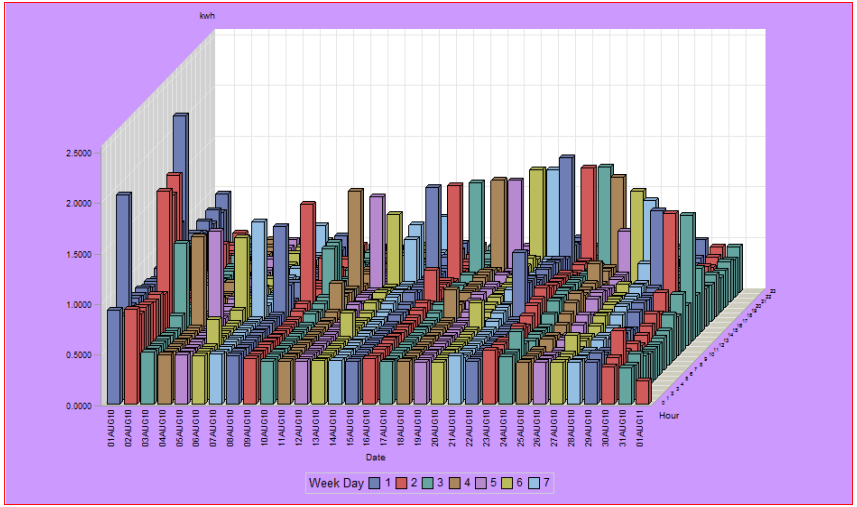
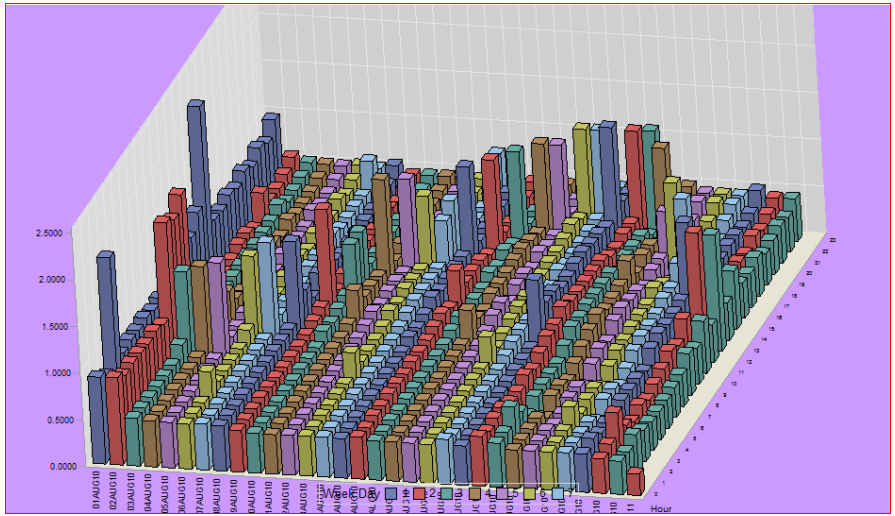



**Figure D: Home Worker**



# Despite load profiles following certain cycles, uncovering the pattern is NOT an easy job!!



## Uncovering the pattern is even more difficult when the data granularity is small (5 minute or 15 minute interval data) and record size is huge

| Meters | Total # of Records | Total Record Size (GB) |
|---|---|---|
| 10,000 | 1,314,000,000 | 10.51 |
| 20,000 | 2,628,000,000 | 21.02 |
| 40,000 | 5,256,000,000 | 42.05 |
| 80,000 | 10,512,000,000 | 84.10 |
| 160,000 | 21,024,000,000 | 168.19 |
| 320,000 | 42,048,000,000 | 336.38 |
| 640,000 | 84,096,000,000 | 672.77 |
| 1,280,000 | 168,192,000,000 | 1,345.54 |
| 2,560,000 | 336,384,000,000 | 2,691.07 |
| 5,120,000 | 672,768,000,000 | 5,382.14 |

10,000 Meters with 15 minute of interval data collection could accumulate 10.51 GB record

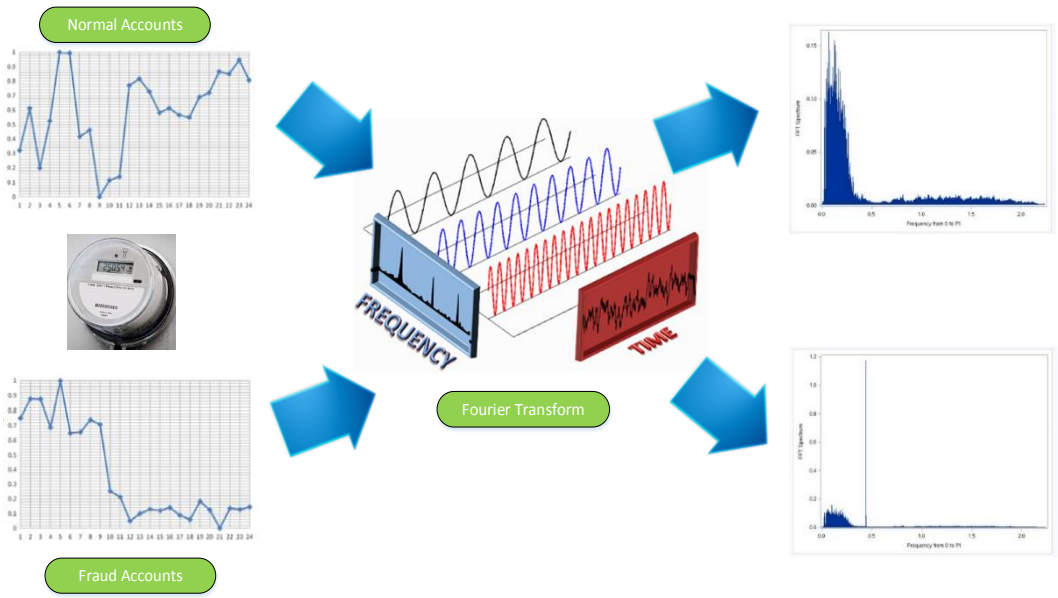
# Fast Fourier Transformation (FFT)

- The Fourier Transform (FT) is a mathematical method for transforming a function of time into a function of frequency.
- The FT basically segments the time series function into its frequency components.
- Fast Fourier Transform (FFT) is an efficient way of computing discrete Fourier Transform (DFT).
- FFT functions basically returns periodic functions of sine and cosine.
- FFT was originated with its application on partial differential equations and sound waves, it has found applications on various other fields such as fingerprint classification, imaging analysis, pattern recognition, texture analysis, sound analysis, or electrical power system.

# Proposed Methodology

## Signal Processing to Load Analytics

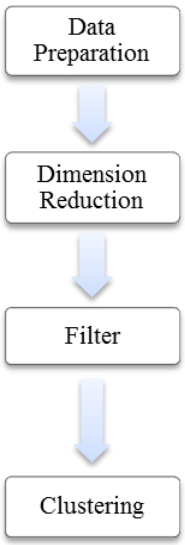
- Proposed methodology applies the concept of signal processing to Load Analytics
  - E.g.: Think of mp3 music
- Use spectral analysis procedure in SAS to create clusters.
  - Hardcore programming or
  - Using SAS Enterprise Miner

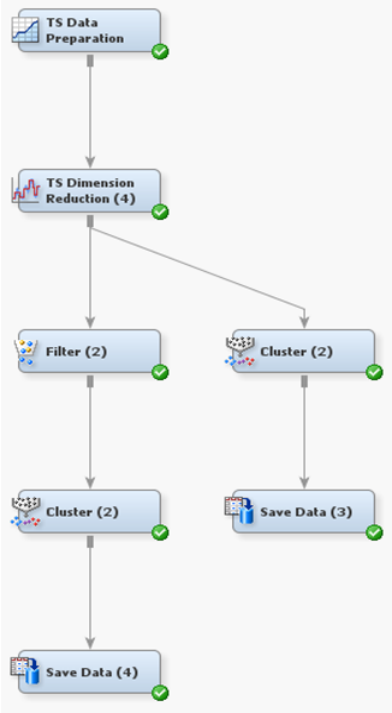


# Proposed Methodology

## Using SAS Enterprise Miner

- SAS Enterprise Miner was used to perform Fast Fourier Transformation on 15 minute interval data collected from over 2,000 smart AMI meters during June 2014.
- Steps using SAS EM:
  1. Data Preparation
  2. Dimension reduction
  3. Filter
  4. Clustering

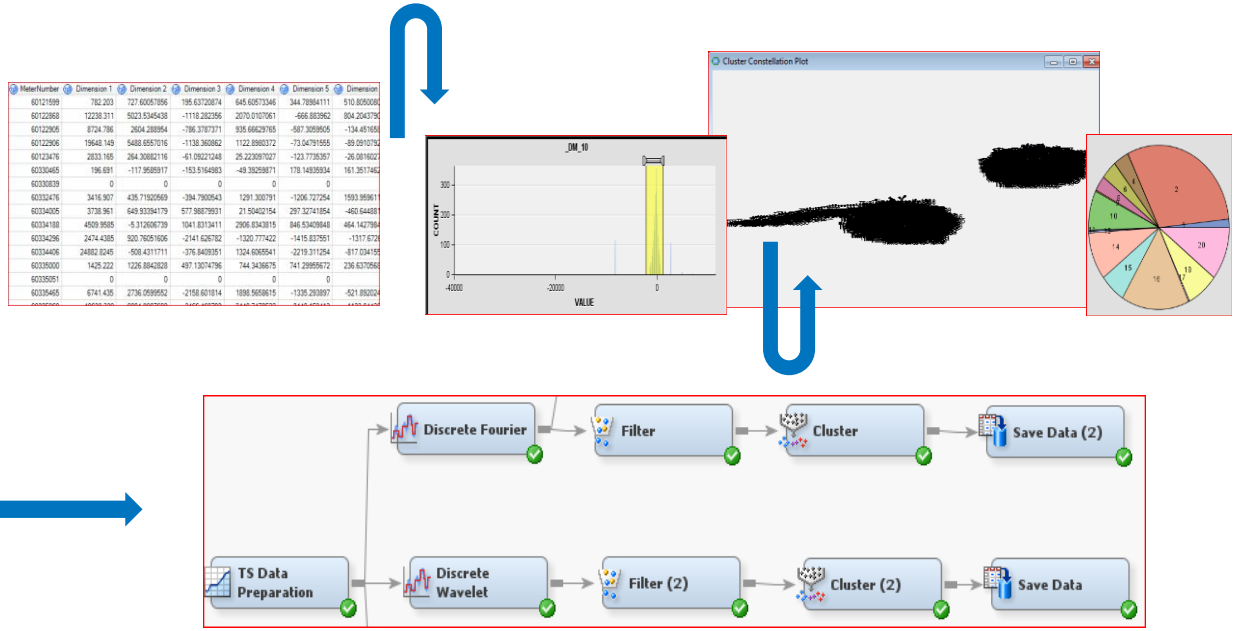


# The process…

## Fourier Transformation is performed by the TS Dimension Reduction Node

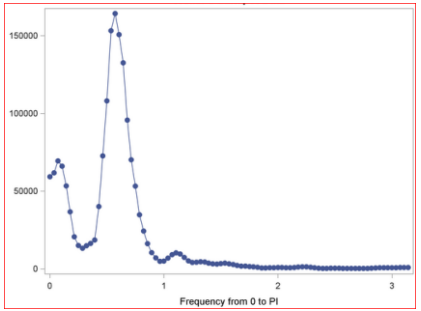
| | |
|---|---|
| **General** | |
| Node ID | TSDR4 |
| Imported Data | |
| Exported Data | |
| Notes | |
| **Train** | |
| Variables | |
| Time Interval | Hour |
| Normalization | |
| Apply Normalization | No |
| Normalization Method | Global Normalization |
| Piecewise Normalization B | 10 |
| Dimension Reduction Metl | |
| Reduction Method | Discrete Fourier Transfor |
| Dimension Number | 20 |
| Output Data Type | Clustering Input Data |
| Max Number of Time Serie | 100 |
| **Status** | |
| Create Time | 4/11/16 1:30 PM |
| Run ID | 9821acc0-1468-4c63-b63 |
| Last Error | |
| Last Status | Complete |
| Last Run Time | 3/3/17 10:33 AM |
| Run Duration | 0 Hr. 0 Min. 11.29 Sec. |
| Grid Host | |
| User-Added Node | No |

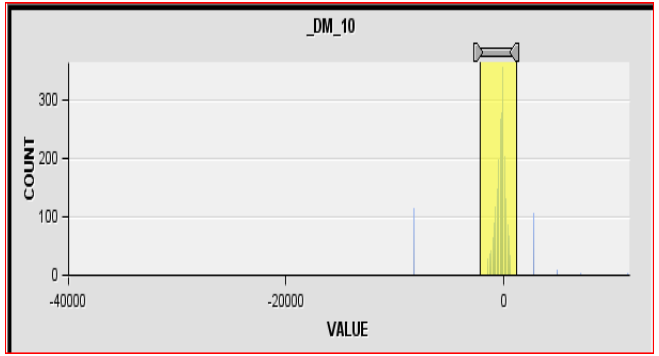


# The process…

## Dimension reduction and Spectral analysis

| MeterNumber | Dimension 1 | Dimension 2 | Dimension 3 | Dimension 4 | Dimension 5 | Dimension |
|---|---|---|---|---|---|---|
| 60121599 | 782.203 | 727.60057856 | 195.63720874 | 645.60573346 | 344.78984111 | 510.8050080 |
| 60122868 | 12238.311 | 5023.5345438 | -1118.282356 | 2070.0107061 | -666.883962 | 804.2043790 |
| 60122905 | 8724.786 | 2604.288954 | -786.3787371 | 935.66629765 | -587.3059505 | -134.451658 |
| 60122906 | 19648.149 | 5488.6557016 | -1138.360862 | 1122.8980372 | -73.04791555 | -89.0910792 |
| 60123476 | 2833.165 | 264.30882116 | -51.09221248 | 25.223097027 | -123.7735357 | -26.0816027 |
| 60330465 | 196.691 | -117.9585917 | -153.5164983 | -49.39259871 | 178.14935934 | 161.3517462 |
| 60330839 | 0 | 0 | 0 | 0 | 0 | |
| 60332476 | 3416.907 | 435.71920569 | -394.7900543 | 1291.300791 | -1206.727254 | 1593.959611 |
| 60334005 | 3738.961 | 649.93394179 | 577.98879931 | 21.50402154 | 297.32741854 | -460.644881 |
| 60334188 | 4509.9585 | -5.312606739 | 1041.8313411 | 2906.8343815 | 846.53409848 | 464.1427984 |
| 60334296 | 2474.4385 | 920.76051606 | -2141.626782 | -1320.777422 | -1415.837551 | -1317.6726 |
| 60334406 | 24882.8245 | -508.4311711 | -376.8409351 | 1324.6065541 | -2219.311254 | -817.034155 |
| 60335000 | 1425.222 | 1226.8842828 | 497.13074796 | 744.3436675 | 741.29955672 | 236.6370568 |
| 60335051 | 0 | 0 | 0 | 0 | 0 | |
| 60335465 | 6741.435 | 2736.0599552 | -2158.601814 | 1898.5658615 | -1335.293897 | -521.892024 |

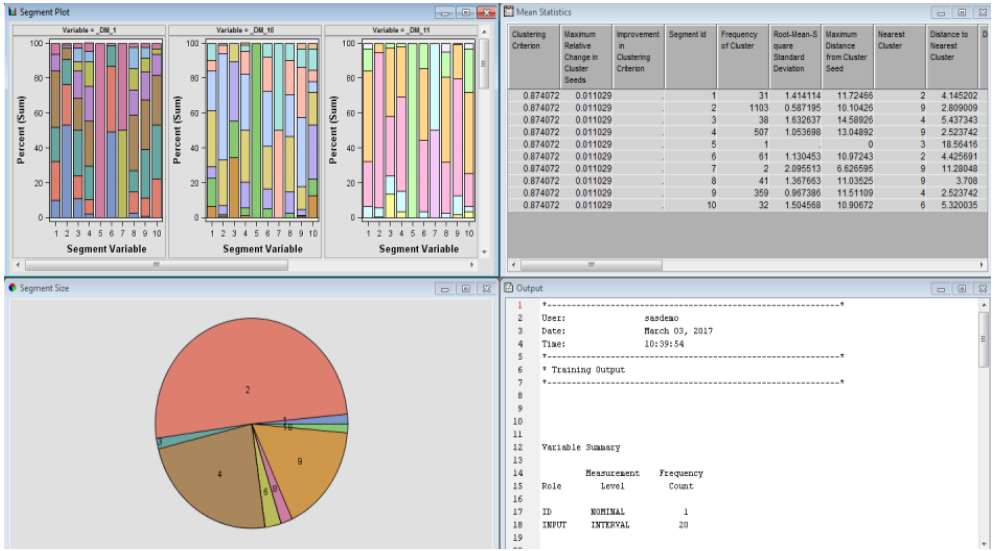


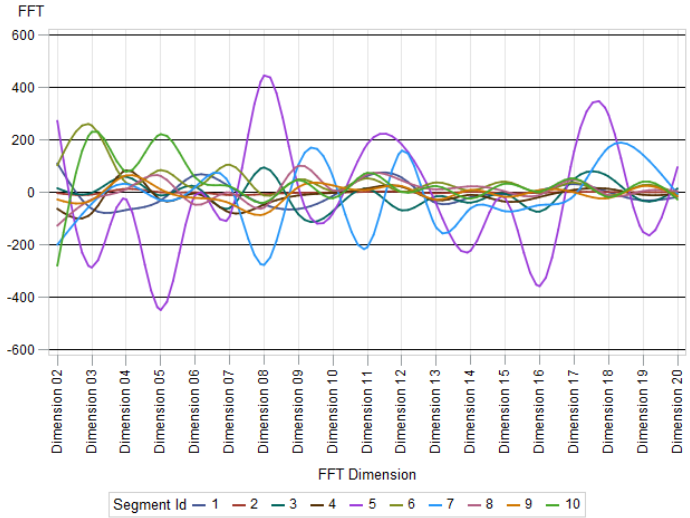

- Use spectral analysis techniques to identify periodicities or cyclical patterns in data.
  - Produces estimates of the spectral and cross-spectral densities of a multivariate time series using a finite Fourier transform
  - Filtering allows to eliminate outliers thereby improving the accuracy of segmentation
  - Periodogram ordinates are smoothed by a moving average
    - Produces estimated spectral and cross-spectral densities
  - Can also test whether or not the data are white noise

# Output from the FFT clustering model in SAS EM

10 different  load shapes represent 10 different segments




| Clustering Criterion | Maximum Relative Change in Cluster Seeds | Improvement in Clustering Criterion | Segment Id | Frequency of Cluster | Root-Mean-Square Standard Deviation | Maximum Distance from Cluster Seed | Nearest Cluster | Distance to Nearest Cluster |
|---|---|---|---|---|---|---|---|---|
| 0.874072 | 0.011029 | . | 1 | 31 | 1.414114 | 11.72466 | 2 | 4.145202 |
| 0.874072 | 0.011029 | . | 2 | 1103 | 0.587195 | 10.10426 | 9 | 2.809009 |
| 0.874072 | 0.011029 | . | 3 | 38 | 1.632637 | 14.58926 | 4 | 5.437343 |
| 0.874072 | 0.011029 | . | 4 | 507 | 1.053698 | 13.04892 | 9 | 2.523742 |
| 0.874072 | 0.011029 | . | 5 | 1 | . | 0 | 3 | 18.56416 |
| 0.874072 | 0.011029 | . | 6 | 61 | 1.130453 | 10.97243 | 2 | 4.425691 |
| 0.874072 | 0.011029 | . | 7 | 2 | 2.095513 | 6.626595 | 9 | 11.28048 |
| 0.874072 | 0.011029 | . | 8 | 41 | 1.367663 | 11.03525 | 9 | 3.708 |
| 0.874072 | 0.011029 | . | 9 | 359 | 0.967386 | 11.51109 | 4 | 2.523742 |
| 0.874072 | 0.011029 | . | 10 | 32 | 1.504568 | 10.90672 | 6 | 5.320035 |




The customers were segmented into 10 different clusters over the study period

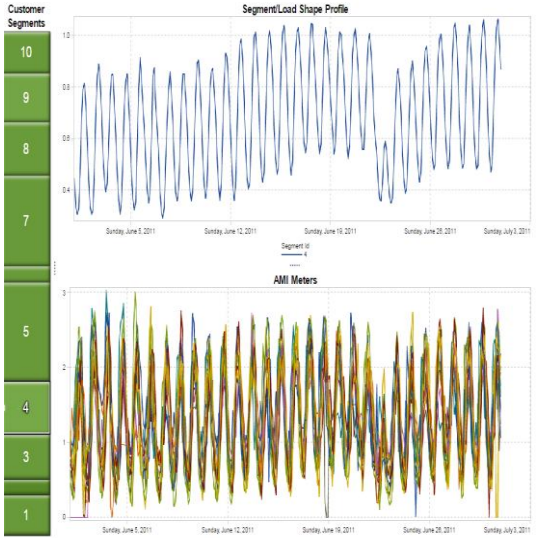
# Interference from the clusters

Applying these 10 segments to the original data, we can gather meaningful information about the customer class

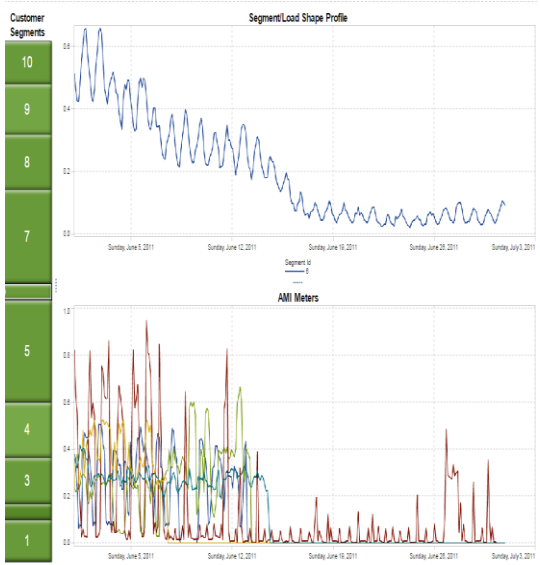
Example

- **Segment 4:** This segment displays a load shape that suddenly dropped during the month (around June 22) and then slowly increased back to the original level. Smart metered customers belonging to this segment might have experienced outage in service resulting a lower usage.
- **Segment 6:** These AMI meters began the month with a stable usage. However, the usage patterns dropped off significantly or completely stopped after June 14.
- Clearly in these two cases, there could be several factors such as true outage or fraudulent activities or system reliability that may have caused such drops.
- With an easy identification of such clusters of use profile, utility would investigate only a specific group of customers therefore reducing the possibility of revenue erosion.

**Load profile for Segment 4**




**Load profile for Segment 6**



# Questions??

thank you
danke
謝謝
ngiyabonga
teşekkür ederim
gracias
спасибо
dank je
tapadh leat
mochchakkeram
go raibh maith agat
dziękuję
obrigado
sukriya
kop khun krap
grazie
arigatō
dakujem
merci
감사합니다
ευχαριστώ
мерси
terima kasih
bedankt
hvala
mauruuru
saagolun
taiku
takk
рахмат
Баярлалаа
хвала

# References

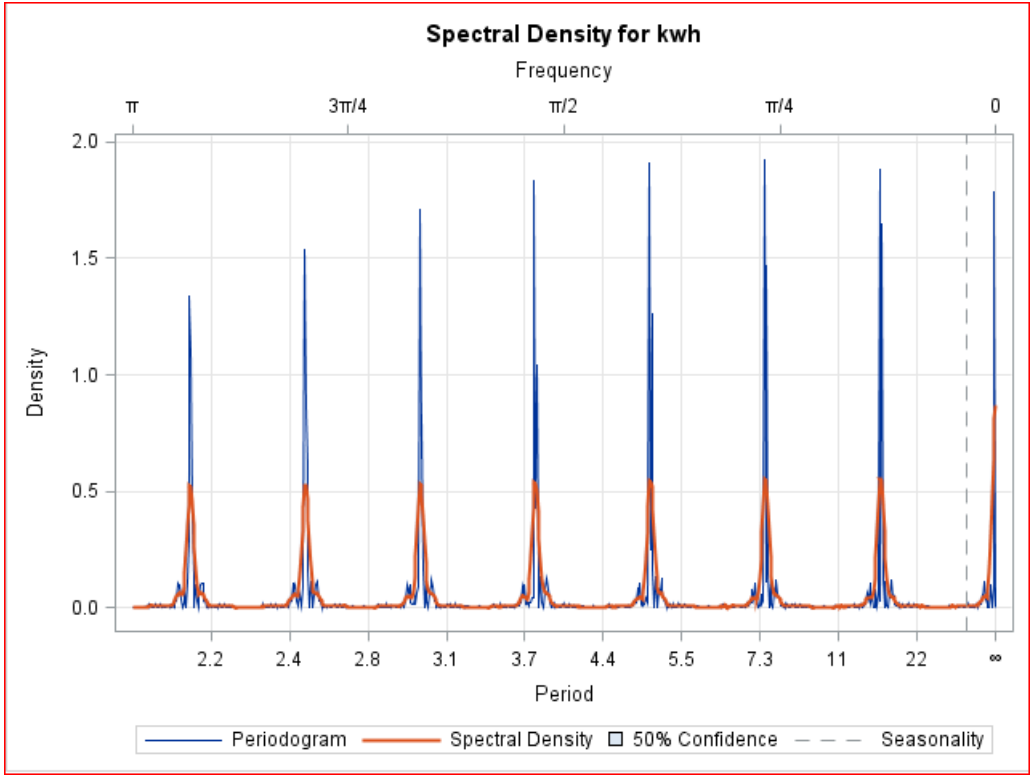
**Reference:**


1. Cooley, James W.; Tukey, John W. (1965). "An algorithm for the machine calculation of complex Fourier series". Mathematics of Computation. 19 (90): 297–301.
2. Gentleman, W. M.; Sande, G. (1966). "Fast Fourier transforms—for fun and profit". Proc. AFIPS. 29: 563–578.
3. Arfken, G. (1985). Development of the Fourier Integral, Fourier Transforms–Inversion Theorem, and Fourier Transform of Derivatives, 15.2-15.4 in Mathematical Methods for Physicists, 3rd ed. Orlando, FL: Academic Press, pp. 794-810.
4. Krantz, S. G. (1999). The Fourier Transform. 15.2 in Handbook of Complex Variables. Boston, MA: Birkhuser, pp. 202-212.
5. Trott, M. (2004). The Mathematica GuideBook for Programming. New York: Springer-Verlag.
6. http://v8doc.sas.com/sashtml/stat/chap33/sect14.htm
7. Bracewell, R.N., (2001), The Fourier Transform and Its Applications, third edition, McGraw-Hill Book Co., New York, 1999.
8. V. Puliafito, S. Vergura, M. Carpentieri, “Fourier, Wavelet, and Hilbert-Huang Transforms for Studying Electrical Users in the Time and Frequency Domain”, Energies 2017(10 188) · February 2017
9. Brandenburg, K. and Bosi, M. (1997). Overview of MPEG audio: Current and future standards for low-bit-rate audio coding, Journal of the Audio Engineering Society, vol. 45, pp. 4-21.
10. Chowning, J.M. (1973). The synthesis of complex audio spectra bymeans of frequency modulation, Journal of the Audio Engineering Society, vol. 21, no. 7, pp. 526-534
11. Vergura, S.; Carpentieri, M.; Puliafito, V. A time-frequency analysis of electrical users by means of Fourier and Wavelet transforms. In Proceedings of the 16th International Conference on Environment and Electrical Engineering (EEEIC 2016), Florence, Italy, 7–10 June 2016.
12. Zou, W. and Wang, D. (2001). Texture identification and image segmentation via Fourier transform. Proc. SPIE, Vol. 4550, pp. 34-39, Image Extraction, Segmentation, and Recognition, Tianxu Zhang; Bir Bhanu; Ning Shu; Eds.
13. M. Beiraghi, A. M. Ranjbar, "Discrete Fourier Transform Based Approach to Forecast Monthly Peak Load", Asia-Pacific Power and Energy Engineering Conference, pp. 1-5, 2011.
14. Shu Fan, Luonan Chen, "Short Term Load Forecasting Based on an Adaptive Hybrid Method", IEEE Transaction on Power Systems, vol. 21, no. 1, February 2006.

# Proc Timeseries

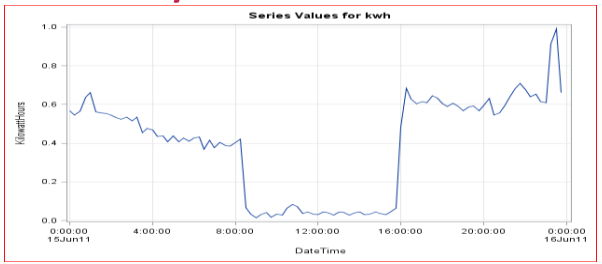
## Spectral Analysis

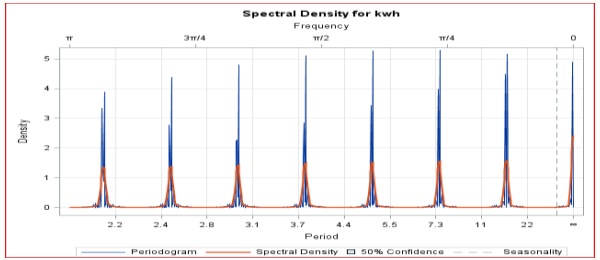
**Peridogram –**
**Spectral Density**
**Confidence**
**Seasonality**

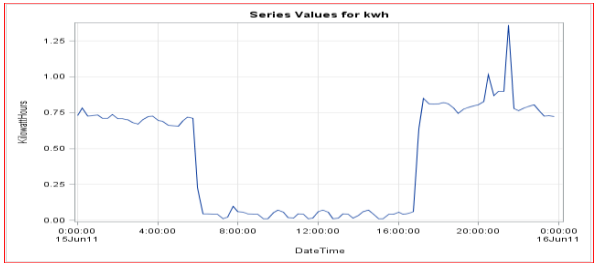


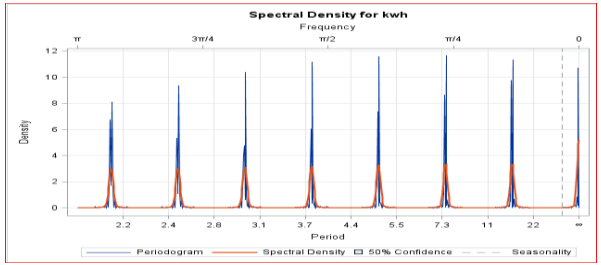
# Proc timeseries

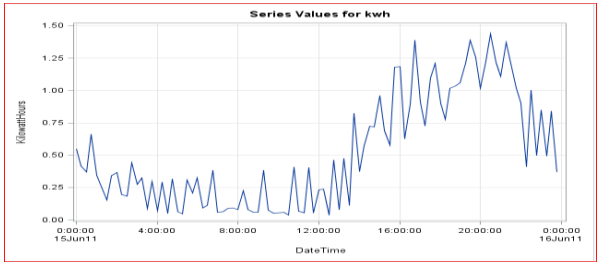
## Spectral analysis

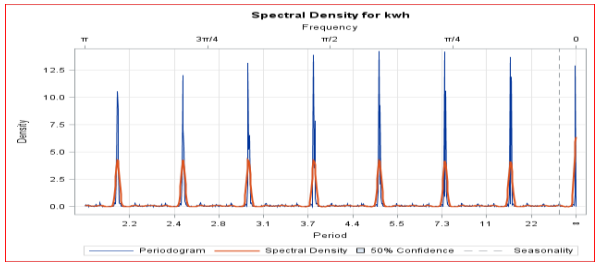


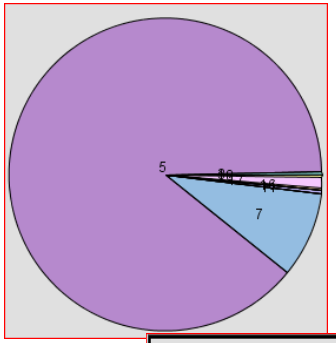
# Results

## Clustering from Spectral Analysis: Filtered vs. Non-filtered approach

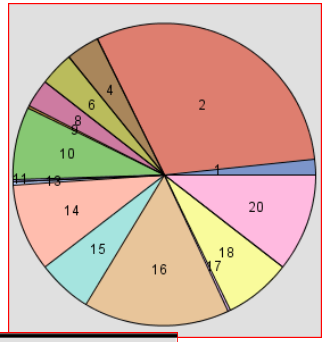
| Segment Id | Frequency of Cluster |
|---|---|
| 1 | 1 |
| 2 | 1 |
| 3 | 2 |
| 4 | 1 |
| 5 | 2115 |
| 6 | 2 |
| 7 | 211 |
| 8 | 1 |
| 9 | 2 |
| 10 | 1 |
| 11 | 3 |
| 12 | 1 |
| 13 | 1 |
| 14 | 1 |
| 15 | 1 |
| 16 | 1 |
| 17 | 32 |
| 18 | 2 |
| 19 | 1 |
| 20 | 1 |

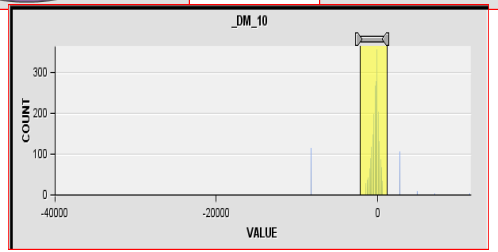
Using 100 dimensions

Non-filtered




Filtered



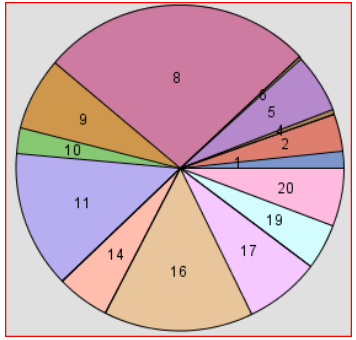

| Segment Id | Frequency of Cluster |
|---|---|
| 1 | 40 |
| 2 | 673 |
| 3 | 2 |
| 4 | 78 |
| 5 | 5 |
| 6 | 77 |
| 7 | 1 |
| 8 | 70 |
| 9 | 3 |
| 10 | 175 |
| 11 | 2 |
| 12 | 2 |
| 13 | 5 |
| 14 | 214 |
| 15 | 127 |
| 16 | 346 |
| 17 | 8 |
| 18 | 160 |
| 19 | 1 |
| 20 | 232 |

# Results

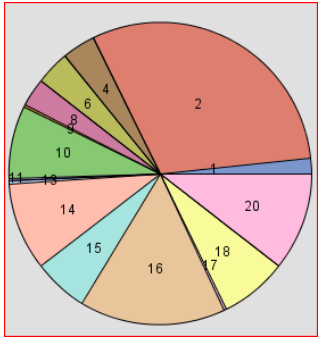
## Clustering with spectral analysis

| Segment Id | Frequency of Cluster |
|---|---|
| 1 | 37 |
| 2 | 81 |
| 3 | 1 |
| 4 | 12 |
| 5 | 130 |
| 6 | 6 |
| 7 | 1 |
| 8 | 608 |
| 9 | 163 |
| 10 | 54 |
| 11 | 306 |
| 12 | 1 |
| 13 | 2 |
| 14 | 117 |
| 15 | 2 |
| 16 | 328 |
| 17 | 168 |
| 18 | 1 |
| 19 | 100 |
| 20 | 131 |

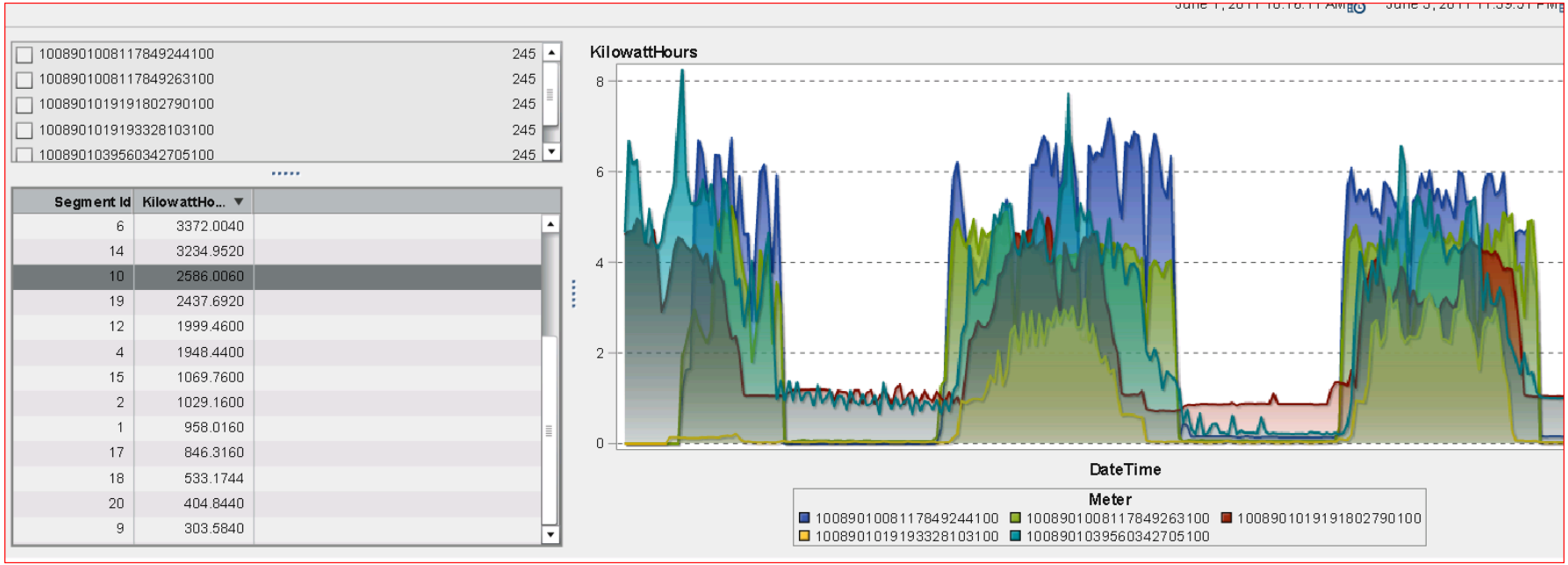
10 Dims




100 Dims




| Segment Id | Frequency of Cluster |
|---|---|
| 1 | 40 |
| 2 | 673 |
| 3 | 2 |
| 4 | 78 |
| 5 | 5 |
| 6 | 77 |
| 7 | 1 |
| 8 | 70 |
| 9 | 3 |
| 10 | 175 |
| 11 | 2 |
| 12 | 2 |
| 13 | 5 |
| 14 | 214 |
| 15 | 127 |
| 16 | 346 |
| 17 | 8 |
| 18 | 160 |
| 19 | 1 |
| 20 | 232 |

# Results

## Load shapes