Energy Load Mining Using Univariate Time Series Analysis

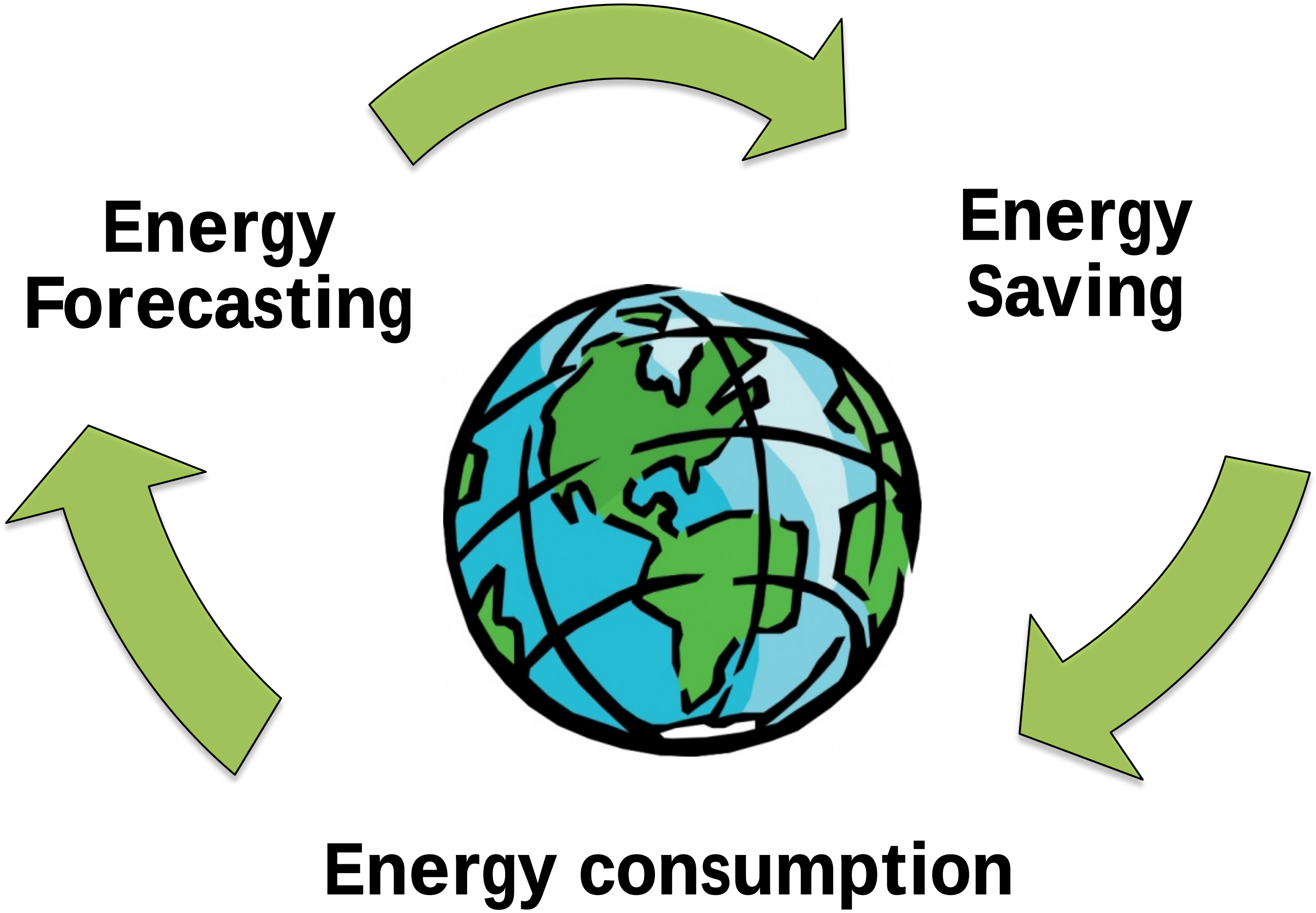
By: Taghreed Alghamdi &

Ali Almadan

03/02/2015

Caruth Hall 0184

SMU | LYLE SCHOOL OF ENGINEERING



Introduction:

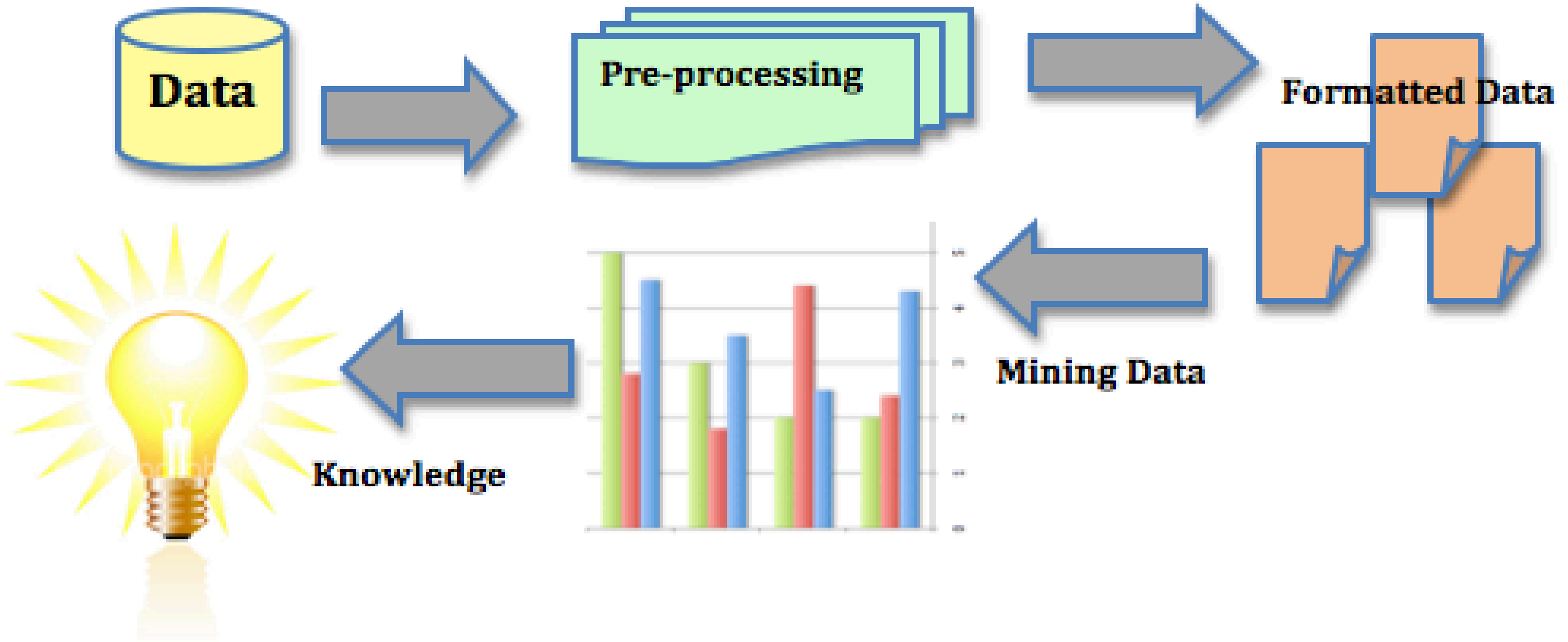
- ◆Energy consumption.
- ◆Energy Forecasting.
- ◆Energy Saving.

SMU | LYLE SCHOOL OF ENGINEERING

Part I
Energy Load Consumption Analysis

By: Taghreed Alghamdi

Energy Load Consumption Analysis



▪ Data Preparation:

✓Detect missing values.

Many methods deal with missing values have been developed to detect missing values

1. Pre-replacing methods:

They replace missing values before the data mining process. They use statistical methods and machine learning methods.

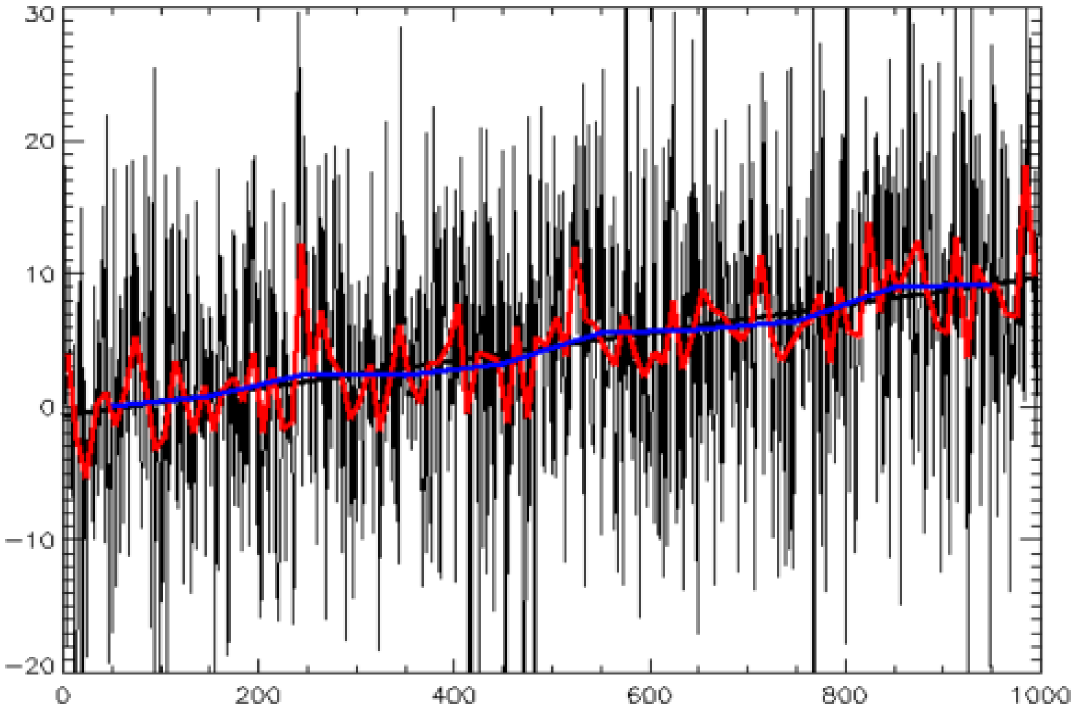
2. Embedded methods:

They deal with missing values while doing data mining itself.

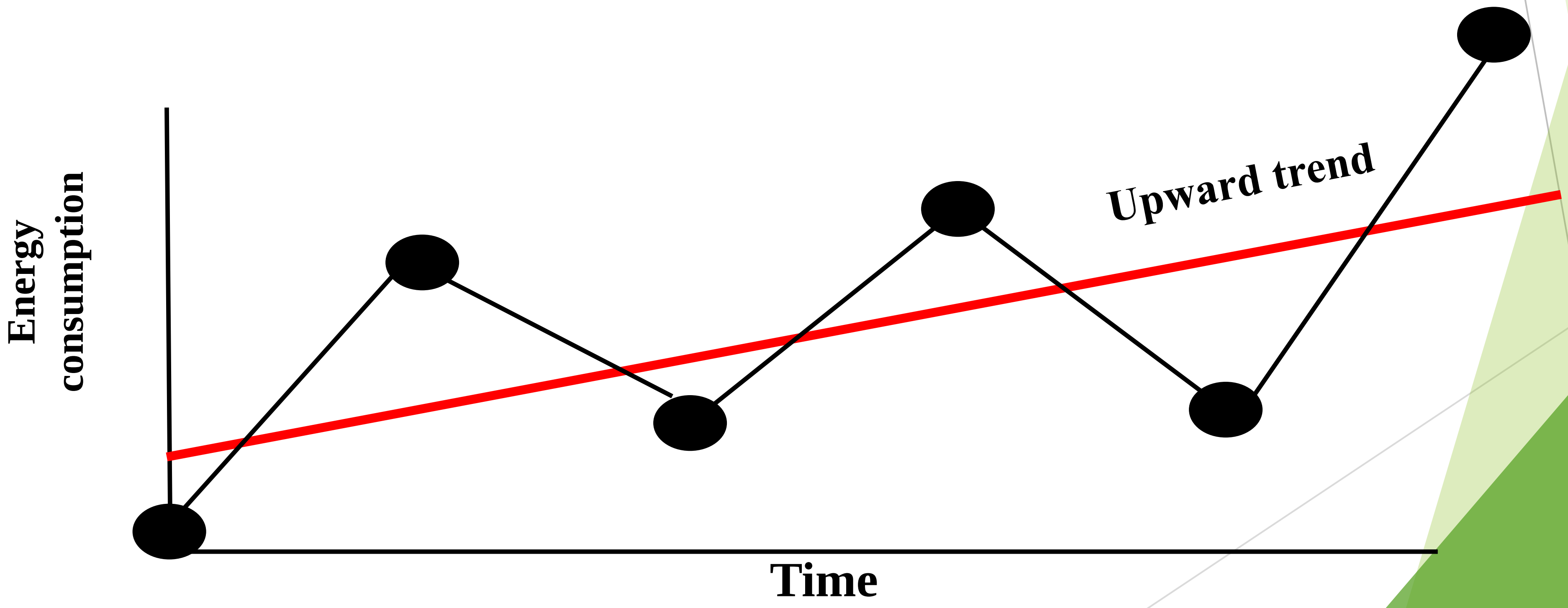
Method	Group	Cost
Mean-and-mode method	Pre-replacing	Low
Linear regression	Pre-replacing	Low
Standard deviation method	Pre-replacing	Low
Nearest neighbor estimator	Pre-replacing	High
Decision tree imputation	Pre-replacing	Middle
Autoassociative neural network	Pre-replacing	High
Casewise deletion	Embedded	Low
Lazy decision tree	Embedded	High
Dynamic path generation	Embedded	High
C4.5	Embedded	Middle
Surrogate split	Embedded	Middle

<u>Time series:</u> **is a sequence of observations, which are ordered in time. There're methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data.**

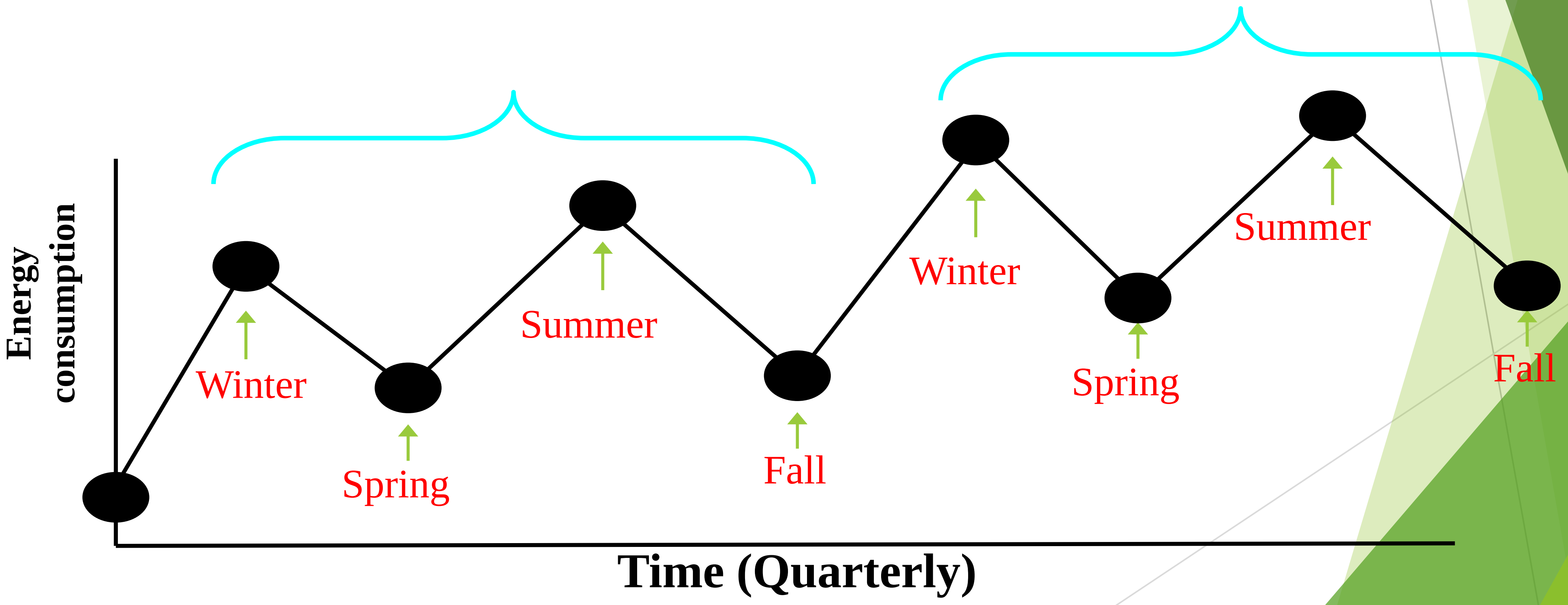
1. **Trend**
2. **Seasonal**
3. **Cyclical**
4. **Irregular**



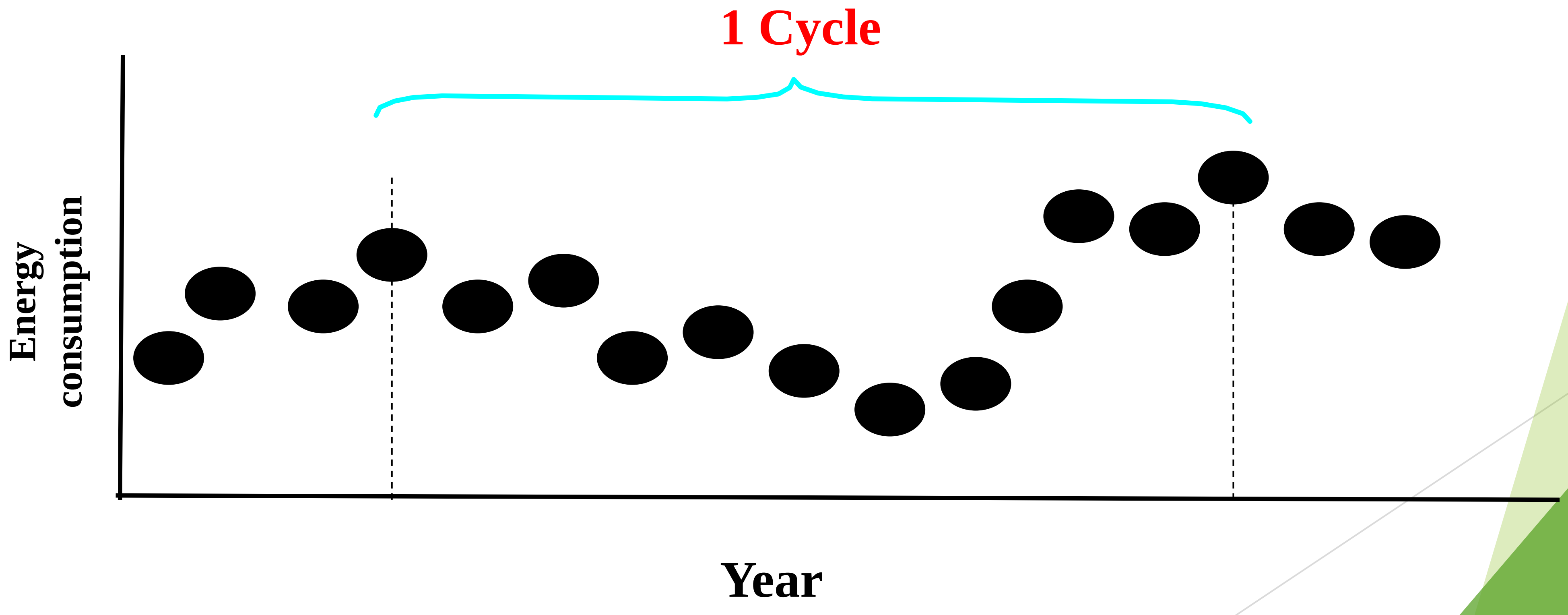
▪ **Trend**: A trend exists when there is a long-term increase or decrease in the data. Trend can be linear or nonlinear such as exponential growth.



- **Seasonal:** A seasonal exists when a series is influenced by seasonal factors (e.g., the quarter of the year, the month, or day of the week). Seasonality is always of a fixed and known period.



- **Cyclical:** A cyclical exists when data exhibit rises and falls that are not of fixed period. The duration of these fluctuations is usually of at least 2 years.



How do we capture the time series we have ?!

- **Visual inspection.**
- **Augmented Dickey-Fuller (ADF) test.**
- **Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test.**

Seasonality:

- Sequence plot.
- Seasonal subseries plot.
- Multiple box plot.
- Autocorrelation plot

What if the time series data contain considerable error?

Then, the first step in the process is using filtering and smoothing techniques .

Smoothing techniques :

Smoothing techniques used for reducing of canceling the effect due to random variation in time series data. This technique, when properly applied, provide a clearer view of the true underlying behavior of the time series we have.

The most common techniques:

1- Exponential smoothing methods.

2- Moving average smoothing methods.

Exponential smoothing methods:

- Single Exponential Smoothing.
  - Data show no trend or seasonality.
- Double Exponential Smoothing (Holt method).
  - Data show only trend.
- Triple Exponential Smoothing (Holt-Winters method).
  - Data show trend and seasonality.

Smoothing Objective:

Select smoothing parameters that minimize the error over the historical data.

Single Exponential Smoothing:

$$S_t = \alpha X_t + (1-\alpha)S_{t-1}$$

smoothing constant value $0 \le \alpha \le 1$

The closer smoothing constant value to 1, the more strongly the forecast depends upon recent values.

0.05 to 0.30 ➔ works well

Trend Analysis:

Double Exponential Smoothing:

- $F_{t+1}=\alpha A_t+(1-\alpha)(F_t+T_t)$
- $T_{t+1}=\beta(F_{t+1}-F_t)+(1-\beta)T_t$
- $HF_{t+m}=F_{t+1}+mT_{t+1}$

- Where
  - F_{t+1} = forecasted value for next period
  - α = the smoothing constant ($0\leq\alpha\leq1$)
  - A_t = actual value of time series now (in period t)
  - F_t = forecasted value for time period t
  - Tt+1= trend value for next period
  - Tt = actual value of trend now (in period t)
  - β = the trend smoothing constant
  - m = number of periods into the future to forecast from the last actual level and trend values

Trend & Seasonality Analysis:

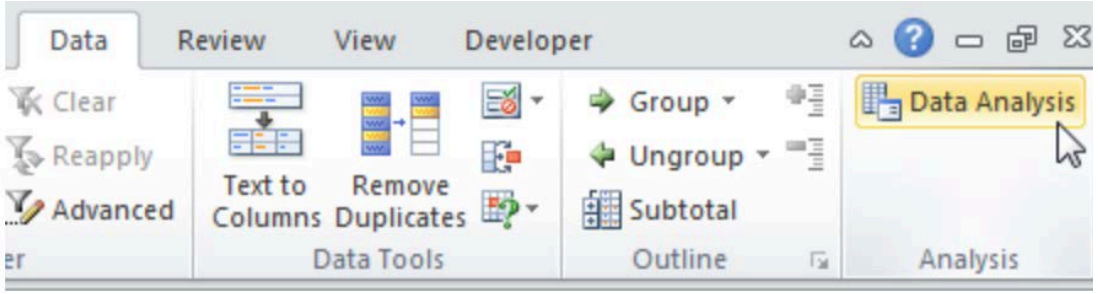
1- Triple Exponential Smoothing (Holt-Winters) :

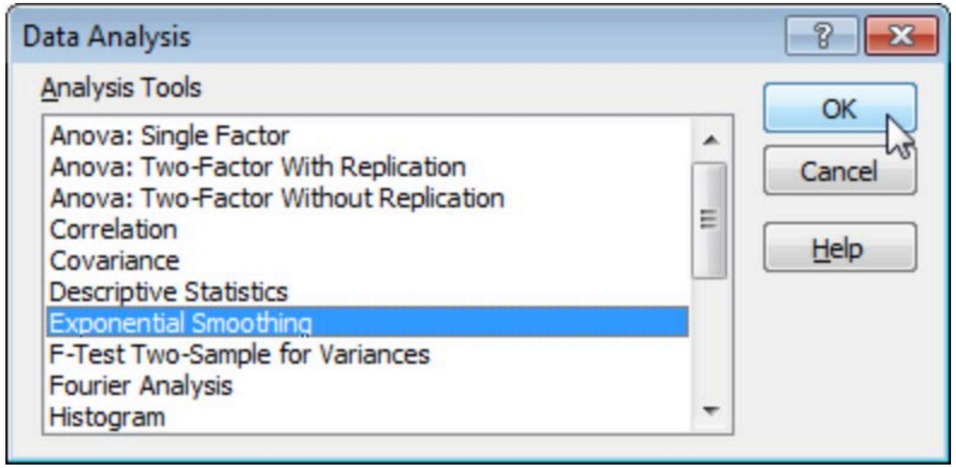
Seasonality Analysis:

1- Autocorrelation correlogram function (ACF).

2- Partial autocorrelations function (PACF).

Applying exponential smoothing in Excel:





Conclusion:

Although nobody can really look into the future, modern statistical methods and data mining methods go along way in helping us to analysis the consumption and forecasting of energy. Exponential smoothing has proven through the years to be very useful in many forecasting situations, and a tool used to observe the data clearly.

SMU | LYLE SCHOOL OF ENGINEERING

Ready

For

Advanced

Modeling

Part II
Energy Load Forecasting

By: Ali Almadan

Outline:

- *Load Forecasting Categories.*
- *Techniques.*
- *ARIMA Model.*
    - *Identification.*
    - *Estimation and Diagnostic Checks.*
    - *Forecasting.*
- *ARIMA Support.*
- *Conclusion.*

Load Forecasting Categories:

Short-Term Forecasting: The short term load forecasting predicts the load demand from *one day to several weeks.* This forecasting is needed to avoid overloading and to ensure security.

Medium-Term Forecasting: The medium term load forecasting predicts the load demand from *a month to several years*. This forecasting is used when a company is planning to expand or shrink its operations in the next years.

Long-Term Forecasting: The long term load forecasting predicts the demand from a *year up to 20 years*. This is usually used to plan for the power system network.

SMU | LYLE
SCHOOL OF ENGINEERING

Techniques & Models:

There are different models and techniques can be used for forecasting electricity load. Some of those are:

- Neural Networks.
- Holt-Winters Forecasting Model.
- Autoregressive Integrated Moving Average ***ARIMA***:
    - Non-seasonal ARIMA.
    - Seasonal ARIMA (Also called ***SARIMA***)
    - Vector ARIMA (Also called ***VARIMA***).

Models Classification:

Univariate Models:

Univariate Models are those forecasting models that are based on the description of only one variable. For example, using just energy consumption in MWh.

Multivariate Models:

Multivariate Models are those forecasting models that are based on the description of many variables to forecast energy consumption and demand.

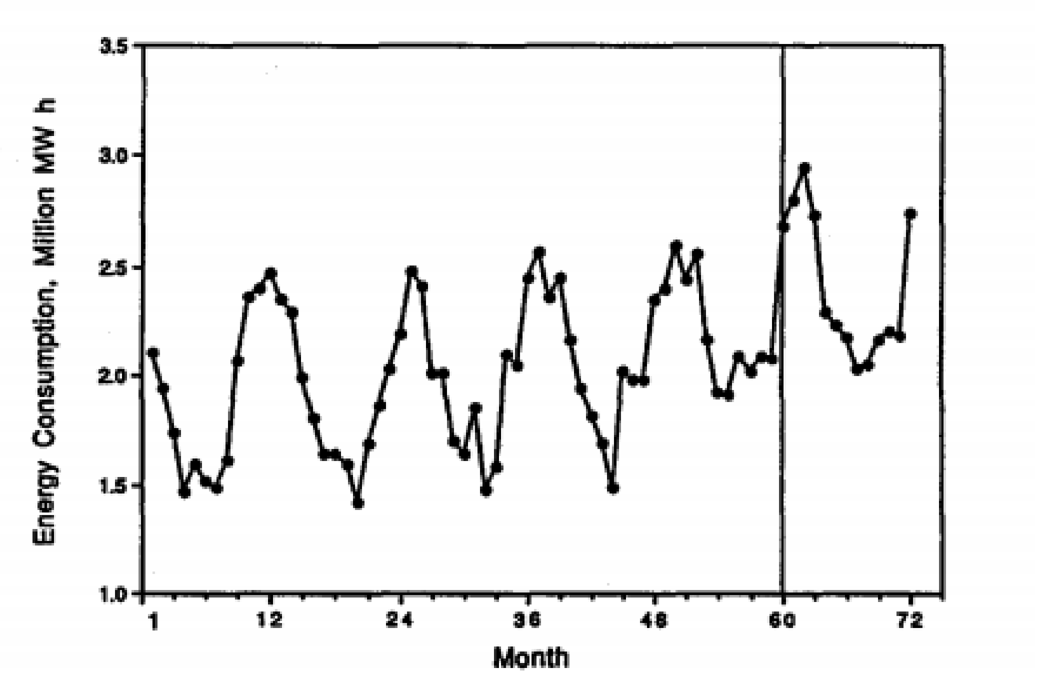
Example:

ARIMA:

Autoregressive Integrated Moving Average model is a univariate time series model that is used to:

1) Understand the data.

2) Predict points in the future.



ARIMA:

Representation:

ARIMA is represented as ARIMA(p,d,q).

ARIMA consists of three parts:

$$\underbrace{AR}\; I\; \underbrace{MA}$$

AR: Autoregressive (*p*)

I: Integrated (*d*)

MA: Moving Average (*q*)

SMU LYLE SCHOOL OF ENGINEERING

ARIMA: “AR” Model

AR: Autoregressive.

Autoregressive (AR) models are models in which the value of a variable in one period is related to its values in previous periods.

AR(q) is an autoregressive model with q lags: $y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \epsilon_t$

ARIMA: "MA" Model

MA: Moving Average.

Moving Average (MA) models account for the possibility of a relationship between a variable and the residual from previous periods.

MA(q) is a moving average model with q lags: $y_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$

ARIMA: “I” Model

To understand the I part, we need to explain a stationary process first.

What is a stationary process?

A stationary process has a mean and variance that do not change over time

AND the process does not have trend.

ARIMA model needs the data as stationary.

What if we need to apply ARIMA and the data is not stationary?

We can use *differencing variable*

$$\Delta y_t = y_t - y_{t-1}$$

ARIMA Stages: Box-Jenkins Methodology

ARIMA modeling has three stages. Those stages are:

- Identification.
- Estimation and diagnostic checking.
- Forecasting.

ARIMA Stages: Box-Jenkins Methodology

ARIMA modeling has three stages. Those stages are:

- ***Identification.***
- Estimation and diagnostic checking.
- Forecasting.

SMU | LYLE SCHOOL OF ENGINEERING

ARIMA Stages: Identification

Looking at the time series and try to identify its characteristics:

Plot examination:

- Are there outliers?
- Are there missing values?
- Is the time series stationary?
- Do we need to perform any transformation?

Autocorrelation Examination:

- What kind of ACF do we have?
- What kind of PACF do we have?

ARIMA Stages: Identification

Looking at the time series and try to identify its characteristics:

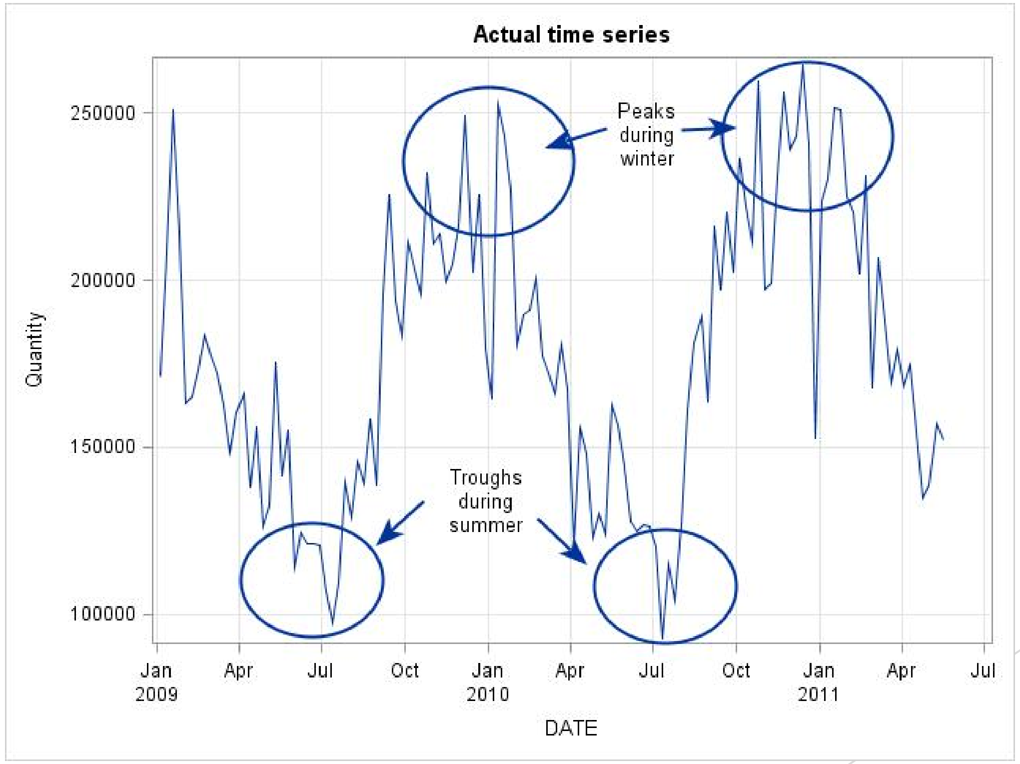
- Is it seasonal

without a trned?

Then use the

moving average

part to detrend it.

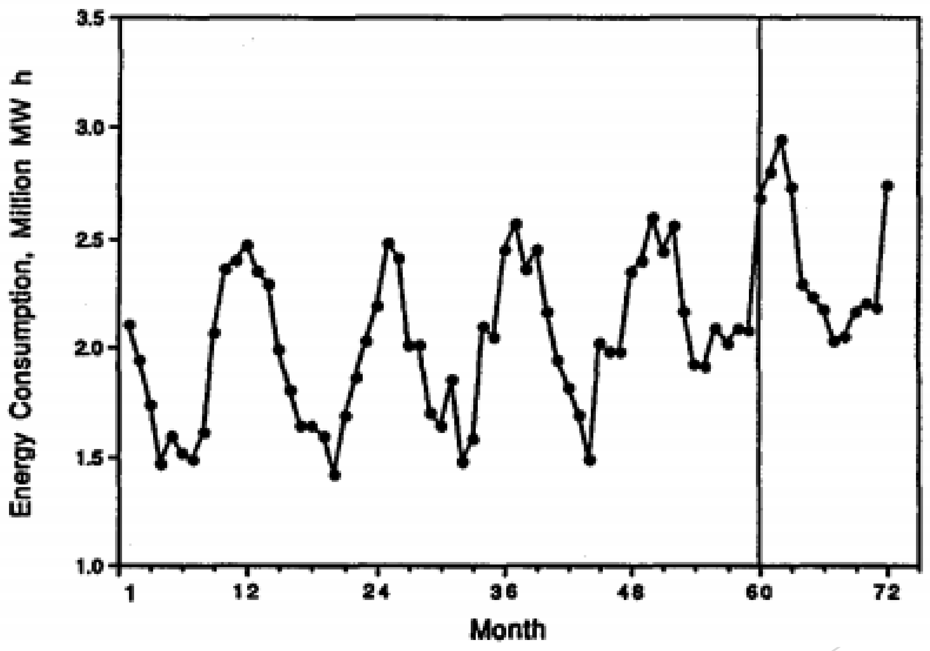


ARIMA Stages: Identification

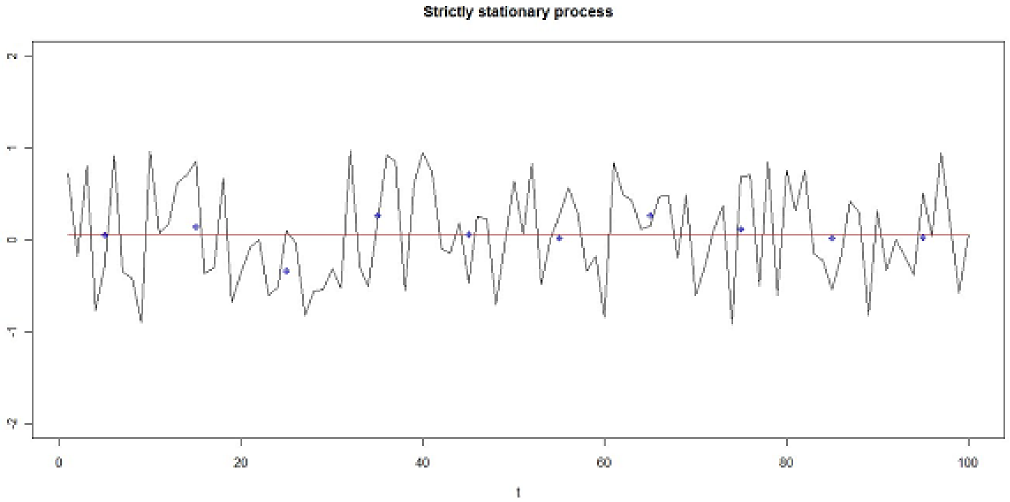
Looking at the time series and try to identify its characteristics:

- Is it stationary?

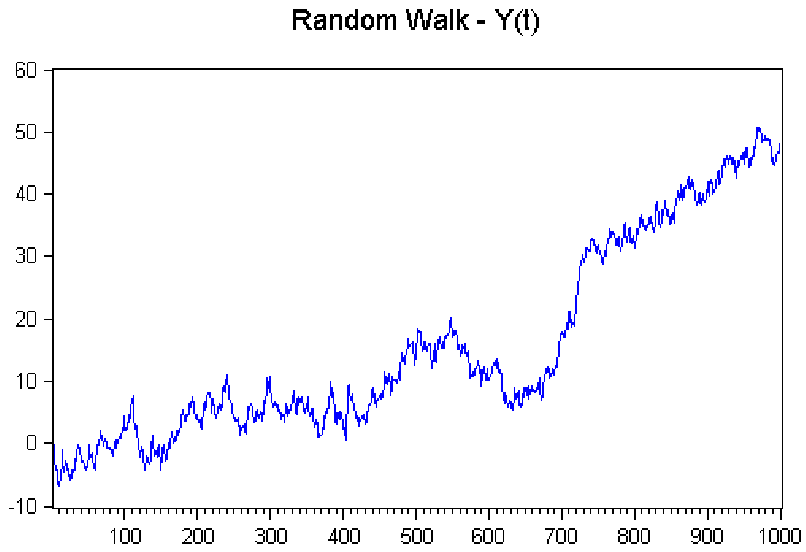
A stationary process has a mean and variance that do not change over time AND the process does not have trend.



ARIMA: Type of Inputs



Stationary



Non-stationary

ARIMA Stages: Box-Jenkins Methodology

ARIMA modeling has three stages. Those stages are:

- Identification.

- ***Estimation and diagnostic checking.***

- Forecasting.

ARIMA Stages: Estimation

In this stage, we estimate the different orders for different ARIMA models and we examine them.

ARIMA(?,?,?)

We need to estimate the autoregressive order, differencing order, and the moving average order.

Rules:

- High positive autocorrelation for a high number of lags often mean we need to use a differencing order more than 1.

ARIMA Stages: Estimation

Rules:

- If the first lag has a negative or zero autocorrelation, then we probably do not need a high differencing order.
- If we use zero as differencing order, then we are dealing with stationary time series (e.g. ARIMA (1,0,1))
- If we use ***one*** as differencing order, then we are dealing with a time series that has a constant average trend (e.g. ARIMA (1,1,1))
- If we use ***two*** as differencing order, then we are dealing with a time series that has a time-varying trend (e.g. ARIMA (1,2,1))

ARIMA Stages: Diagnostic Checking

Diagnostic Checking:

We check if the model fits well. We used the autocorrelation function **ACF** and the partial autocorrelation function **PACF** in the checking process.

If yes, then the residuals of the model should resemble a white noise process.

If no, we need to come up with better estimation to come up with a better model.

ARIMA Stages: Box-Jenkins Methodology

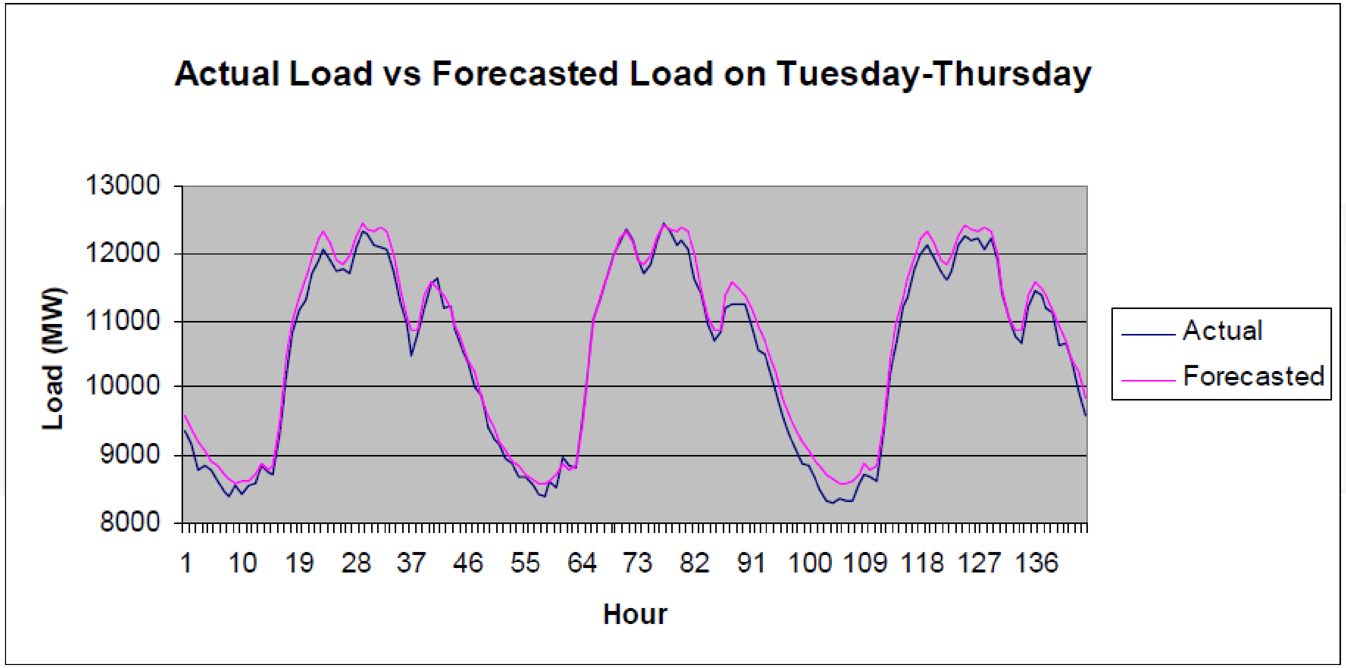
ARIMA modeling has three stages. Those stages are:

- Identification.
- Estimation and diagnostic checking.
- ***Forecasting.***

ARIMA Stages: Forecasting

The final part? ALMOST THERE!

When we get a model that fits, it’s time do the forecasting by running the model:



Evaluation

The final model can be evaluated using one year testing set and ***Mean Absolute Percentage Error***

$$\text{MAPE}\ (\%) = \frac{1}{N}\left[\frac{\left|Z'_t - x_t\right|}{x_t}\right] \times 100\%$$

Where Z′t = Forecasted Load,

Xt = Actual Load

N = Forecasting number

Conclusion

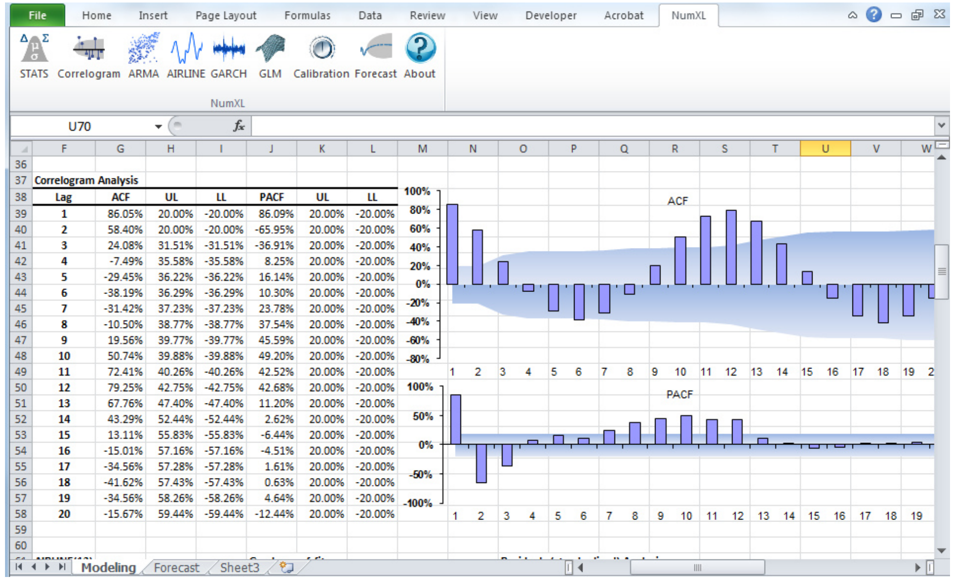
How does that help us save energy?

Is it just saving energy?

Support

Different programs support ARIMA:

- R (tseries Package)
- SAS (PROC ARIMA)
- Microsoft Excel (With NumXL)



Correlogram Analysis

Lag	ACF	UL	LL	PACF	UL	LL
1	86.05%	20.00%	-20.00%	86.09%	20.00%	-20.00%
2	58.40%	20.00%	-20.00%	-65.95%	20.00%	-20.00%
3	24.08%	31.51%	-31.51%	-36.91%	20.00%	-20.00%
4	-7.49%	35.58%	-35.58%	8.25%	20.00%	-20.00%
5	-29.45%	36.22%	-36.22%	16.14%	20.00%	-20.00%
6	-38.19%	36.29%	-36.29%	10.30%	20.00%	-20.00%
7	-31.42%	37.23%	-37.23%	23.78%	20.00%	-20.00%
8	-10.50%	38.77%	-38.77%	37.54%	20.00%	-20.00%
9	19.56%	39.77%	-39.77%	45.59%	20.00%	-20.00%
10	50.74%	39.88%	-39.88%	49.20%	20.00%	-20.00%
11	72.41%	40.26%	-40.26%	42.52%	20.00%	-20.00%
12	79.25%	42.75%	-42.75%	42.68%	20.00%	-20.00%
13	67.76%	47.40%	-47.40%	11.20%	20.00%	-20.00%
14	43.29%	52.44%	-52.44%	2.62%	20.00%	-20.00%
15	13.11%	55.83%	-55.83%	-6.44%	20.00%	-20.00%
16	-15.01%	57.16%	-57.16%	-4.51%	20.00%	-20.00%
17	-34.56%	57.28%	-57.28%	1.61%	20.00%	-20.00%
18	-41.62%	57.43%	-57.43%	0.63%	20.00%	-20.00%
19	-34.56%	58.26%	-58.26%	4.64%	20.00%	-20.00%
20	-15.67%	59.44%	-59.44%	-12.44%	20.00%	-20.00%

References:

Derby, Nate. "Time Series Forecasting Methods - SAS | Business Analytics and." *By Nate Derby. Calgary SAS Users Group, n.d. Web. 02 Mar. 2015.*

"FACTORS AFFECTING ELECTRICITY CONSUMPTION IN THE U.S." *Innovation Electricity Efficiency (2013): 1-72. Web.*

"FACTORS INFLUENCING ENERGY CONSUMPTION AND COSTS IN BROILER PROCESSING PLANTS IN THE SOUTH." *SOUTHERN JOURNAL OF AGRICULTURAL ECONOMICS (1978): 1-6. Web.*

Fomby, Thomas B. "Exponential Smoothing Models1." *Southern Methodist University. Department of Economics, n.d. Web.*

Gotham, Douglas J. "Energy Forecasting Methods." *Energy Center. Purdue University, 15 Nov. 2007. Web.*

Mayilvahanan, M., and M. Sabitha. "Opportunities and Challenges in Using Renewable Resources in India: A Data Mining Approach." *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) 1.2 (2012): 1-4. Web.*

Mohamed*,, Hoda K., Soliman M. El-Debeiky, Hassan M. Mahmoud, and Khaled M. Destawy. "Data Mining for Electrical Load Forecasting In Egyptian Electrical Network." *Ain Shams U., Faculty of Eng., Computer & System Eng. Dept., Cairo, Egypt (n.d.): 1-12. Web*

Piasecki, Dave. "Exponential Smoothing Explained." *InventoryOps. N.p., n.d. Web*

Zhu, Bo, Min-You Chen, Neal Wade, and Li Ran. "A Prediction Model for Wind Farm Power Generation Based on Fuzzy Modeling." *Procedia Environmental Sciences 12 (2012): 122-29.*

Questions?