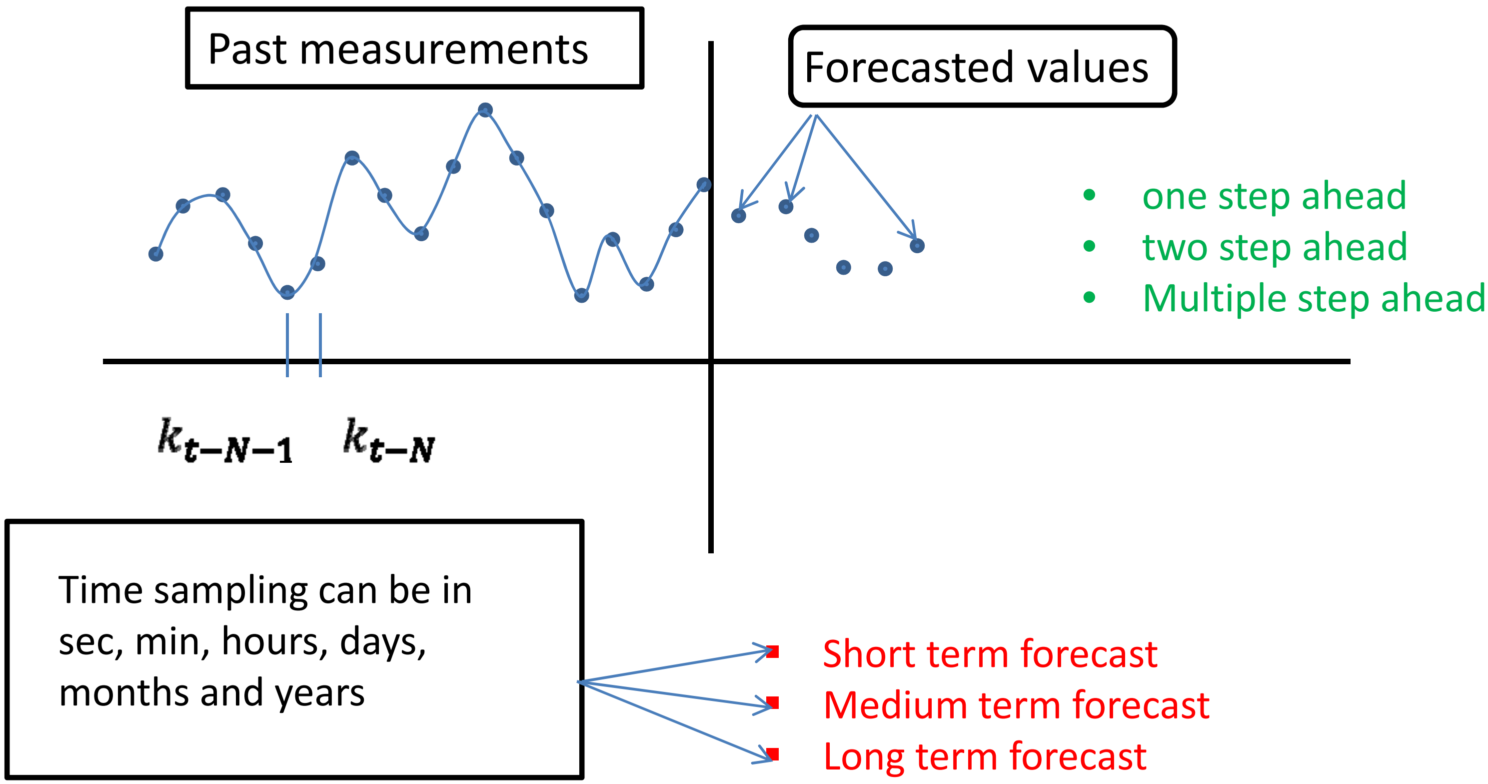
Short-Term Load Forecasting

Dr SN Singh, Professor
Department of Electrical Engineering
Indian Institute of Technology Kanpur
Email: snsingh@iitk.ac.in

Basic Definition of Forecasting

Forecasting is a problem of determining the future values of a time series from current and past values.



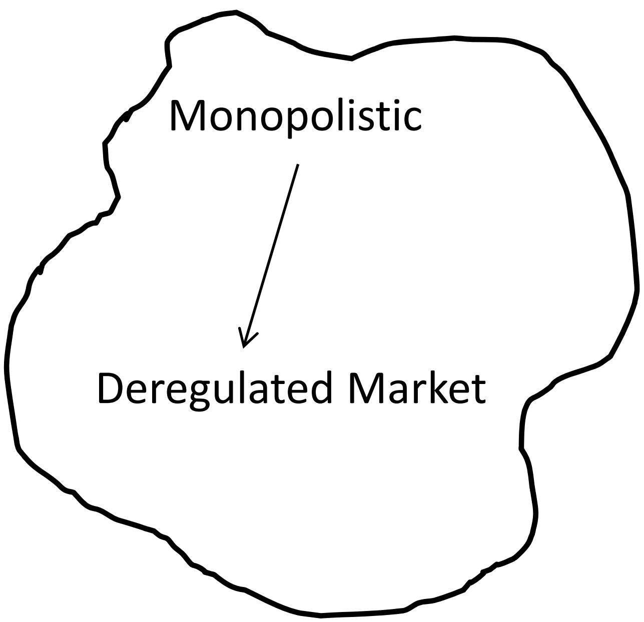
Role of Forecasting in Electric Power System

Before Deregulation of Electric Power System

Only Load Demand forecasting is carried

- Economic operation and Unit commitment
- Maintenance and planning of power system

After Deregulation of Electric Power System



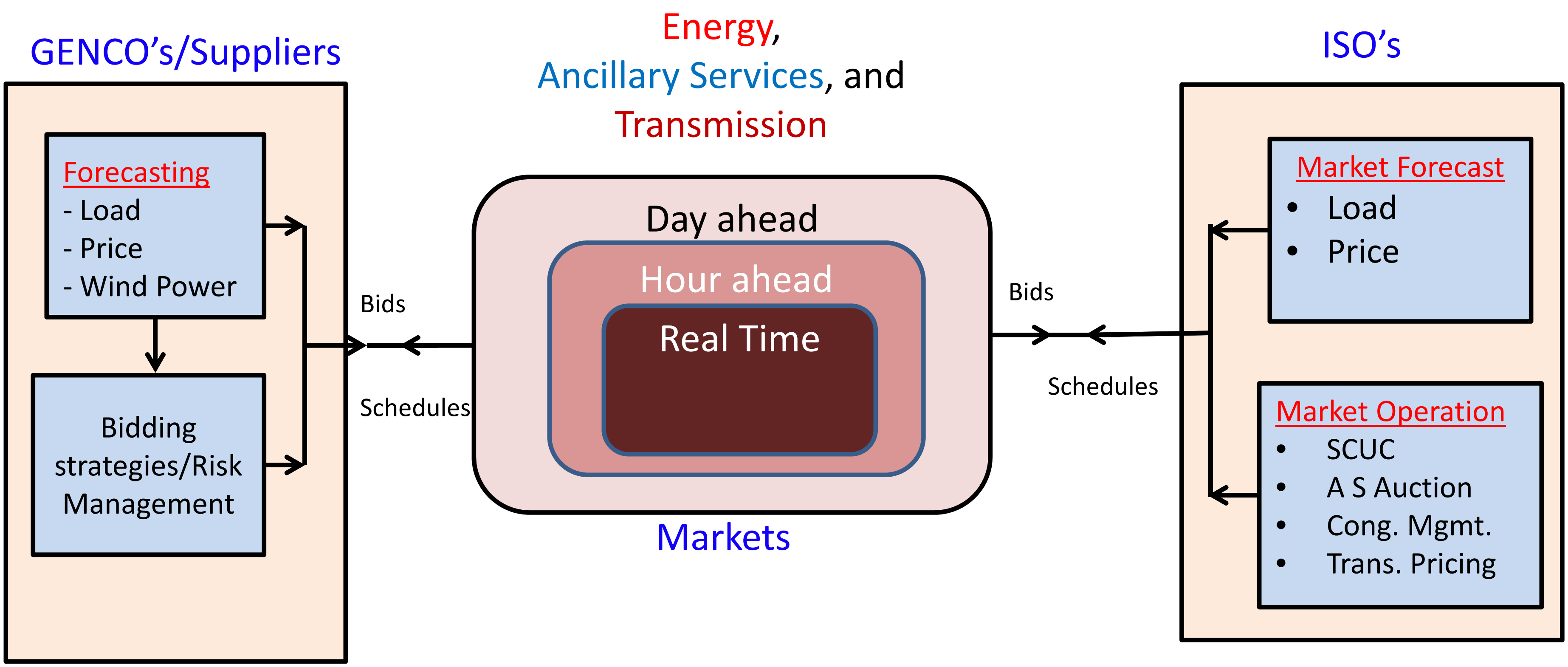
Motivation:

- Creating competition among the suppliers
- Leaving choices to the buyers

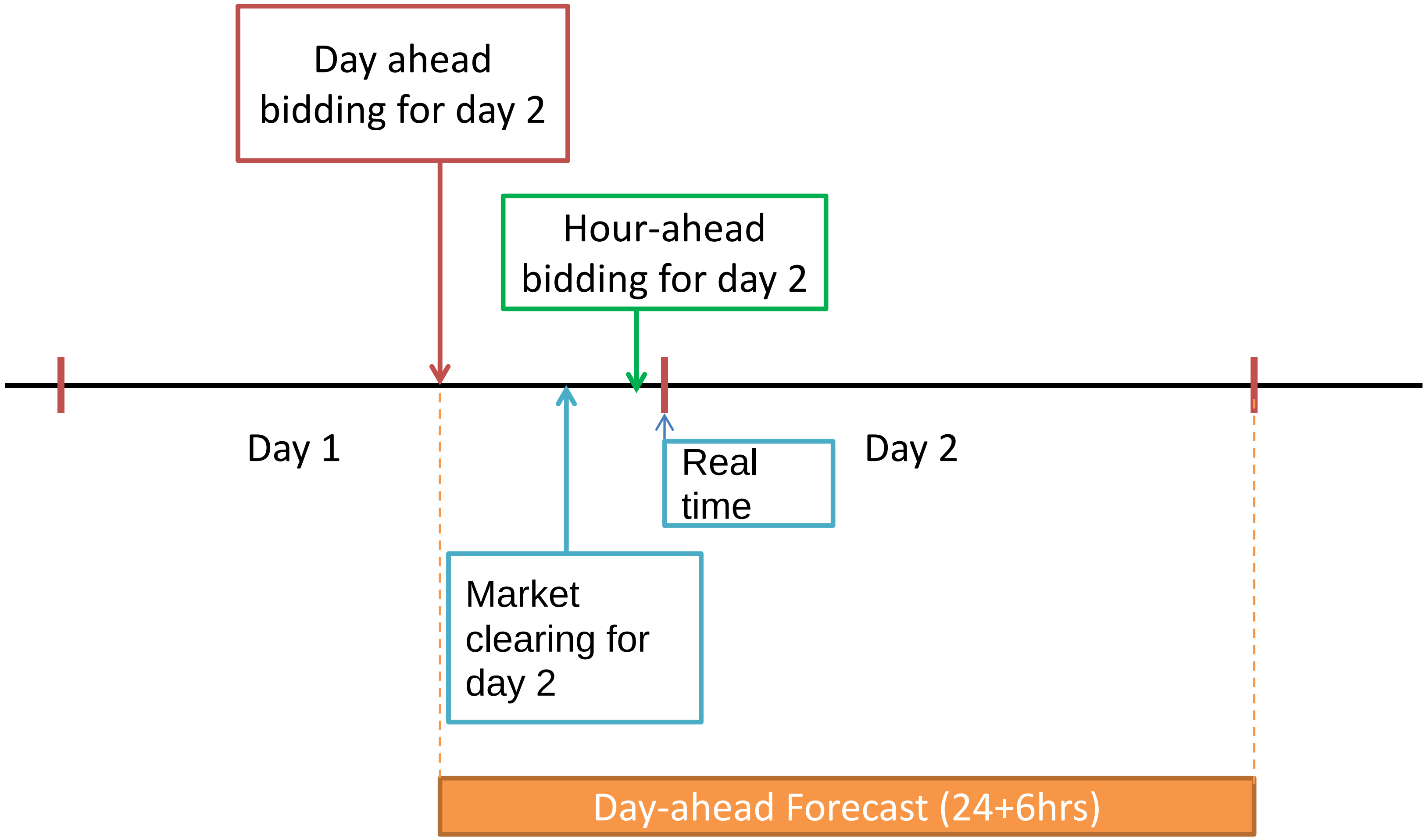
Role of Forecasting in Electric Power System

....Contd.

Electricity Market Operation



Market Bidding Process



Important Tools

1. Load Forecast
2. Price Forecast
3. Operating Reserve Margin Forecast
4. **Wind Forecast**

System Operator Point of View:

Planning Problems:

Due to uncertainty, unlike conventional generators, wind power generation cannot be included into ELD and UC problems.

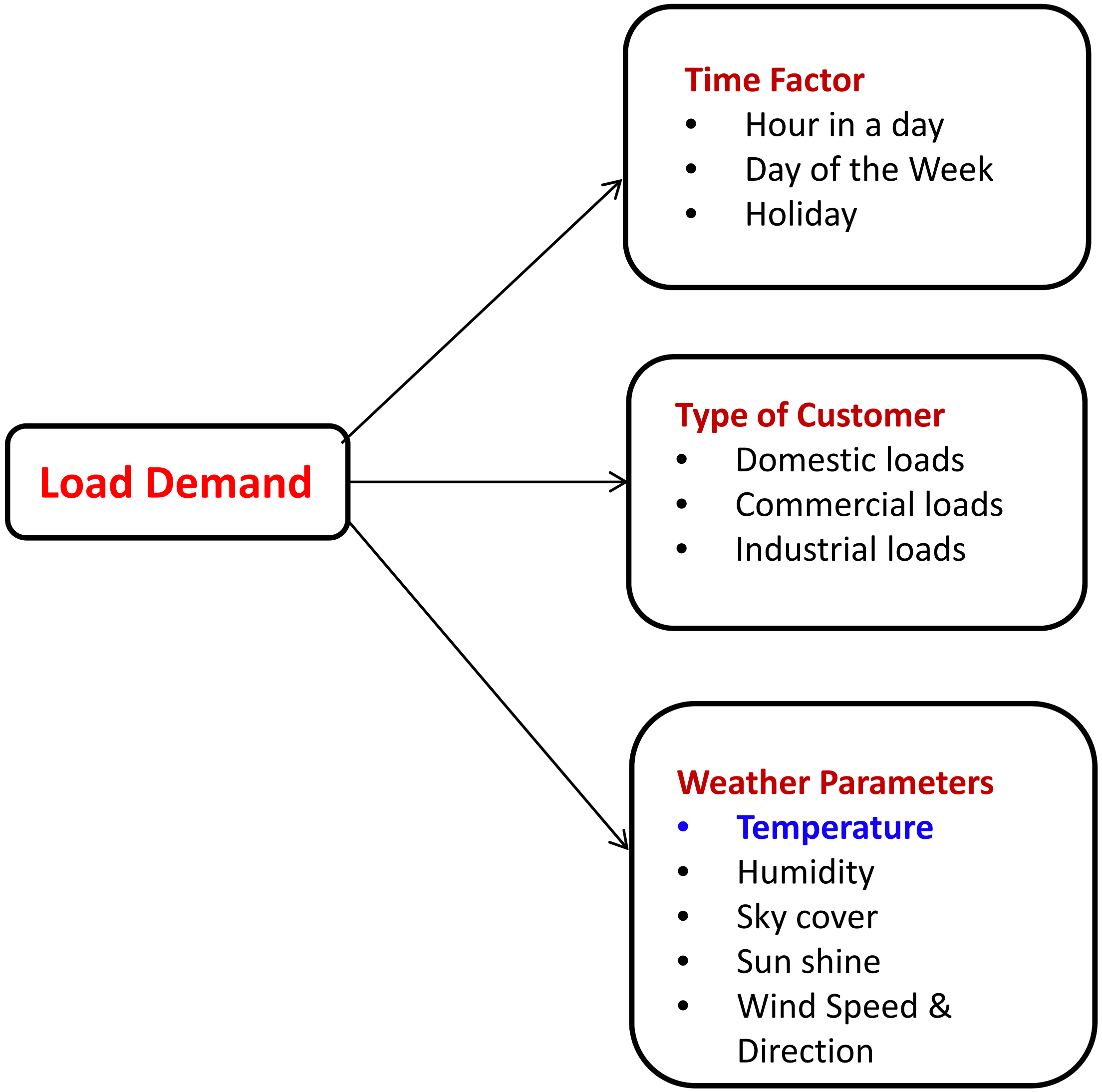
Operational:

Frequency control, Voltage control, Power Quality, Ancillary services provision.

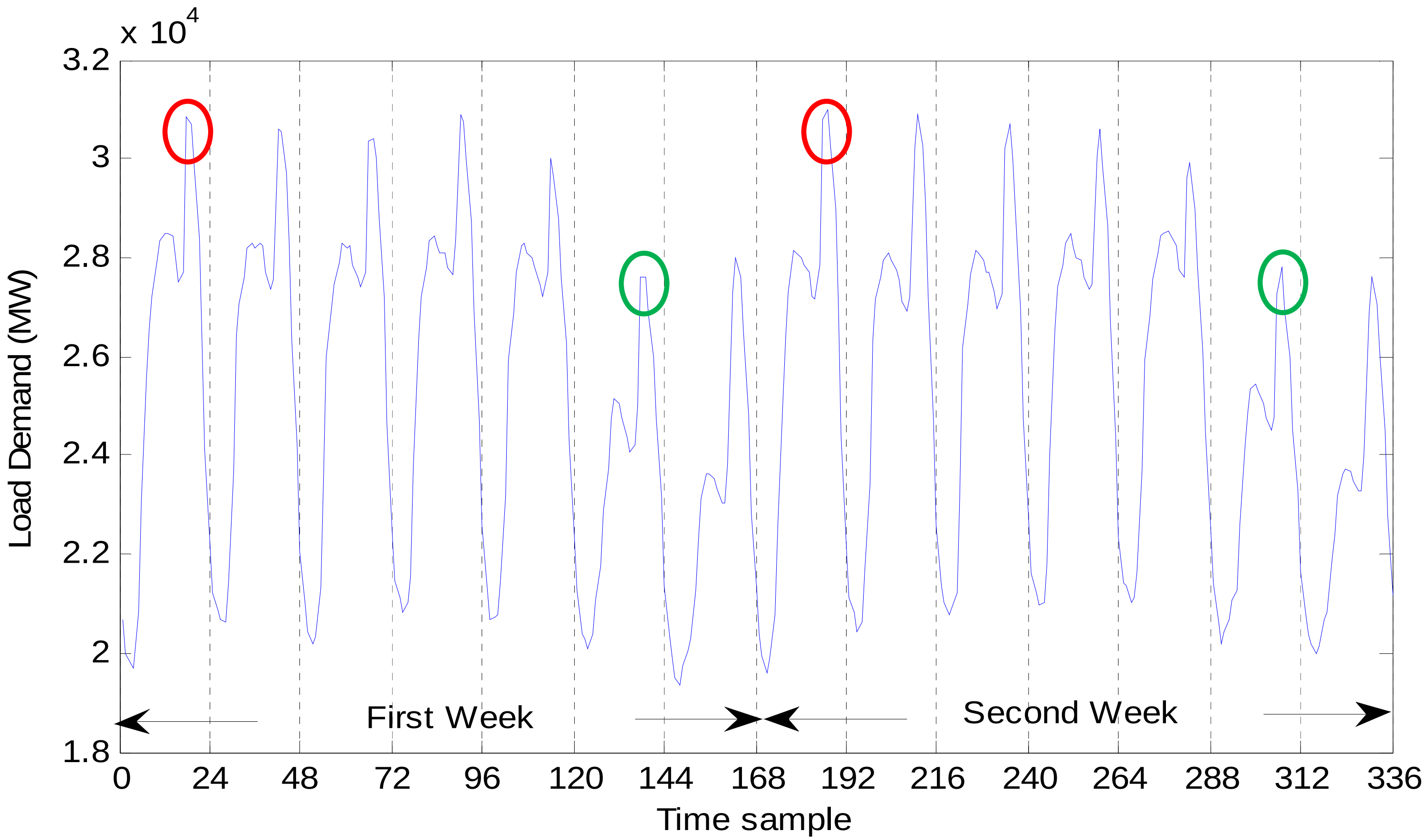
Wind power producer point of view:

Bidding in day ahead, adjustment and settling Electricity Markets to maximize profits/minimize their imbalance costs.

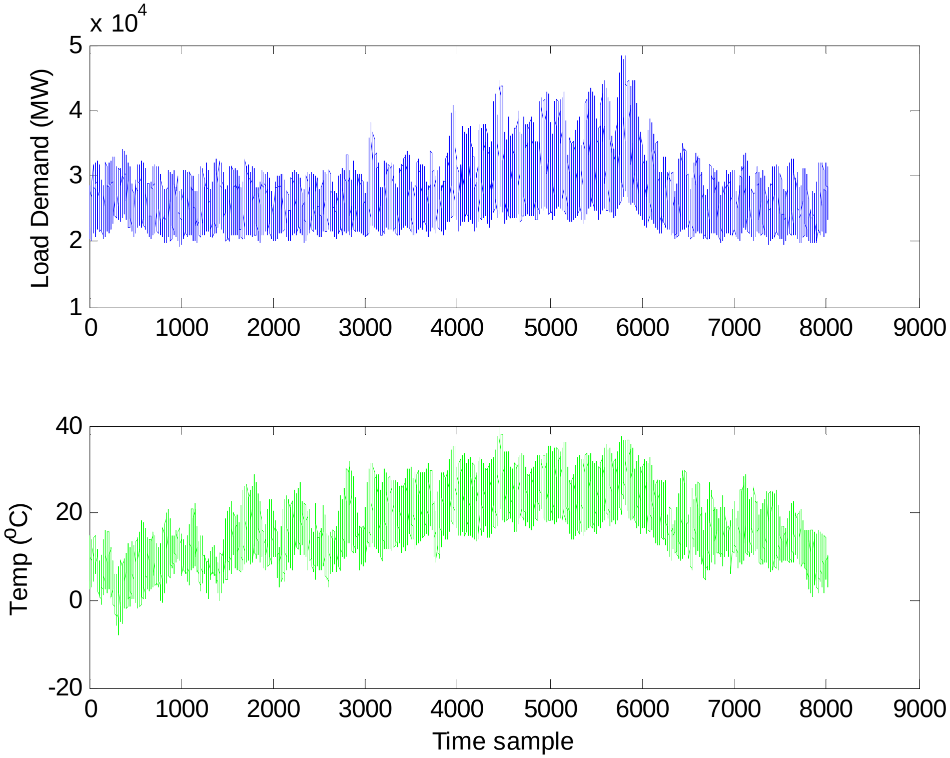
Factors Influencing the Forecast Variable

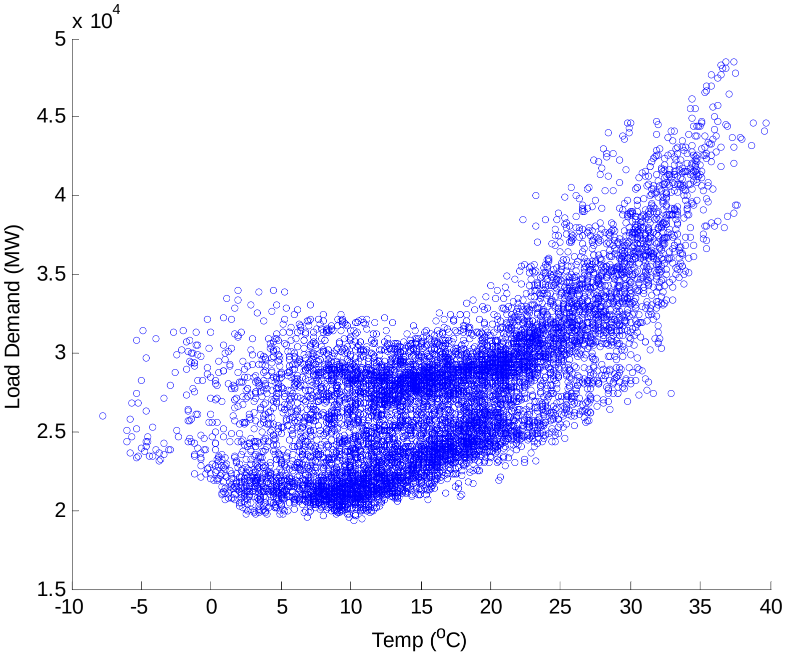


Effect of Time Factor

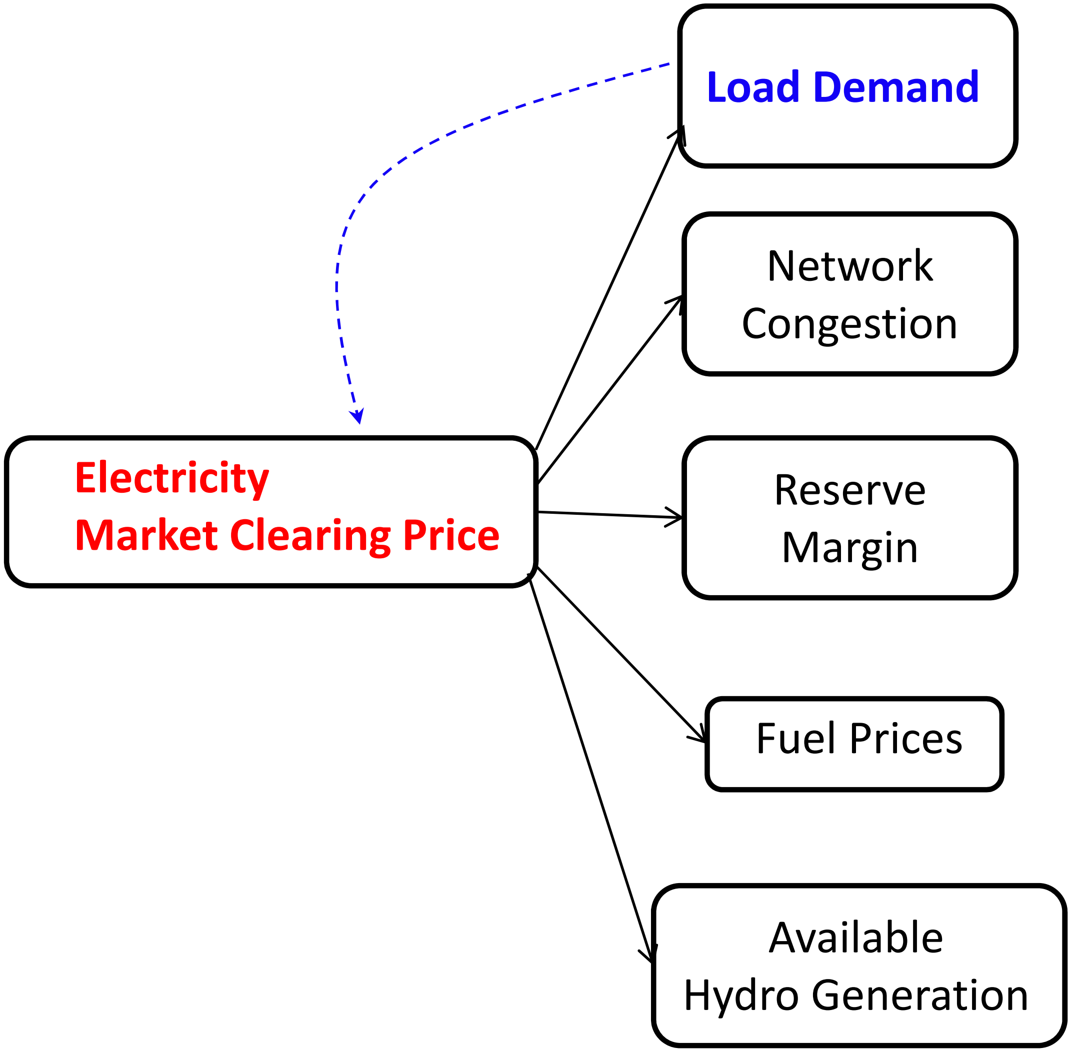


Dependency on Weather Parameter: Temperature

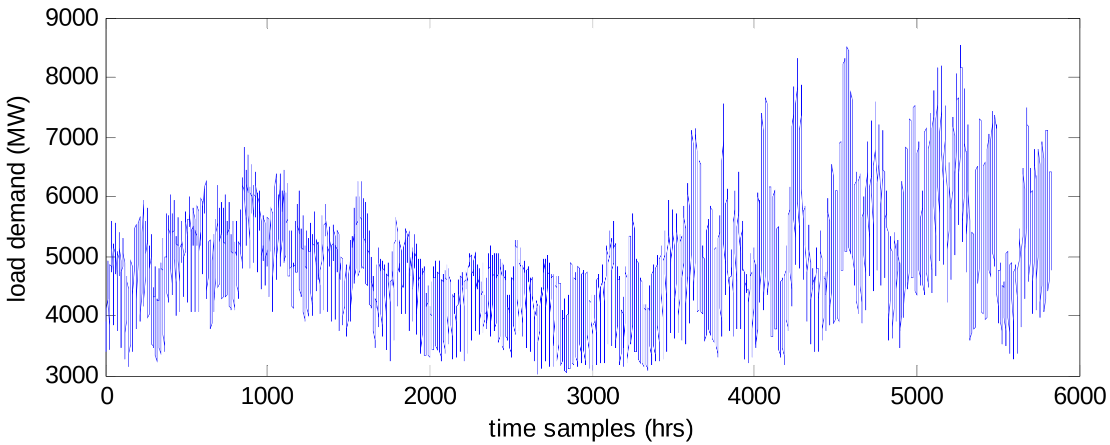


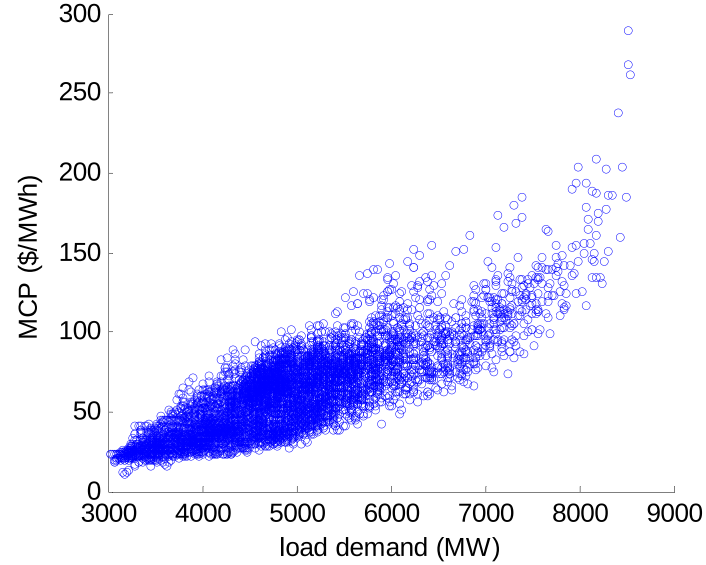


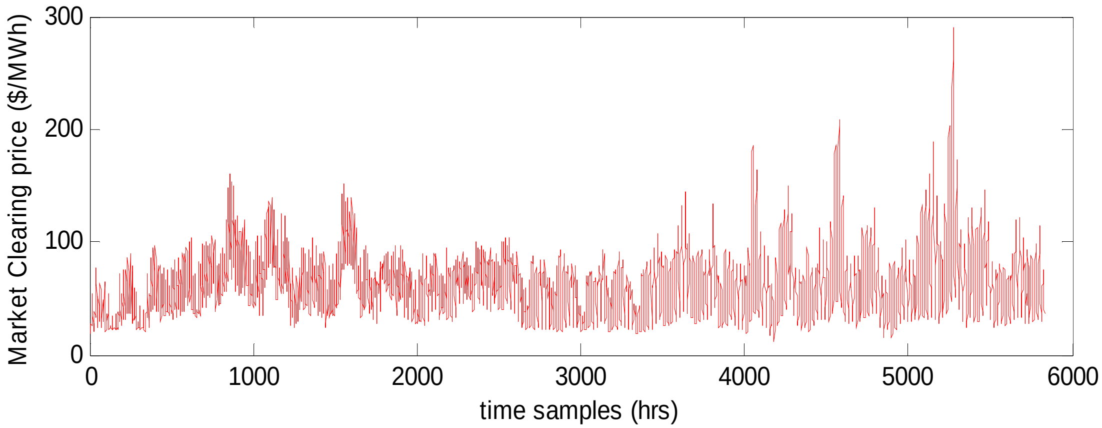
Factors Influencing Electricity Market Price

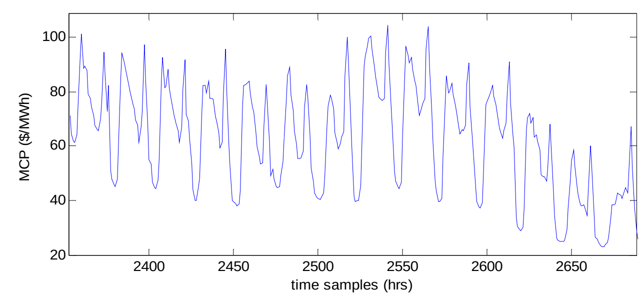


Price vs. Load Demand

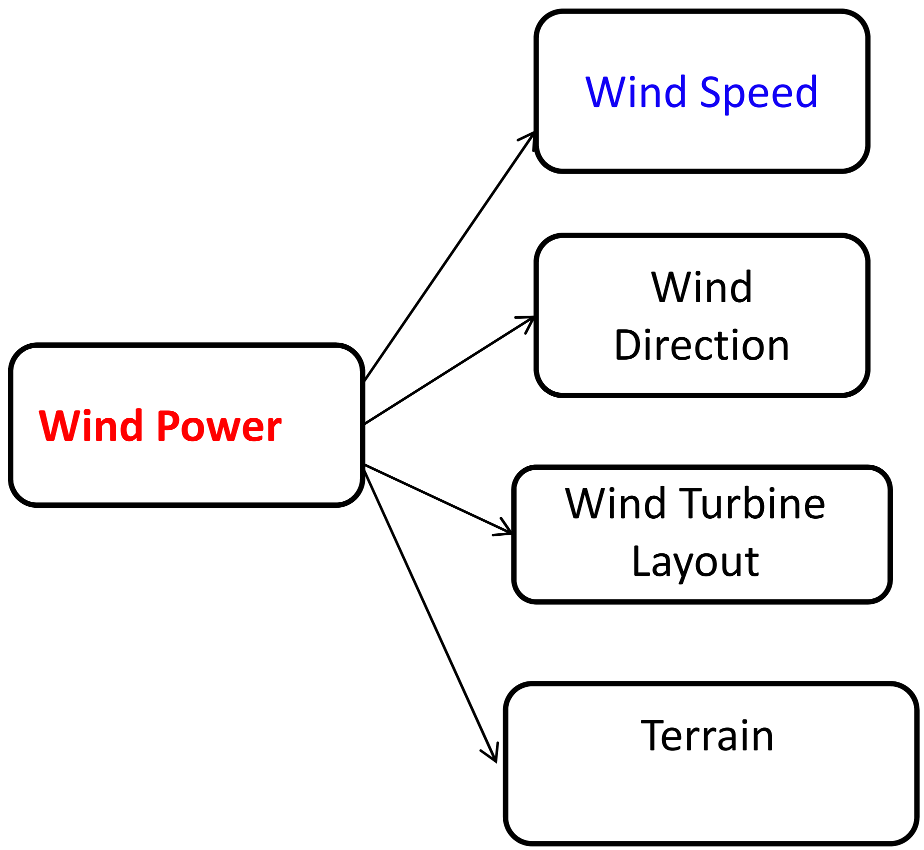




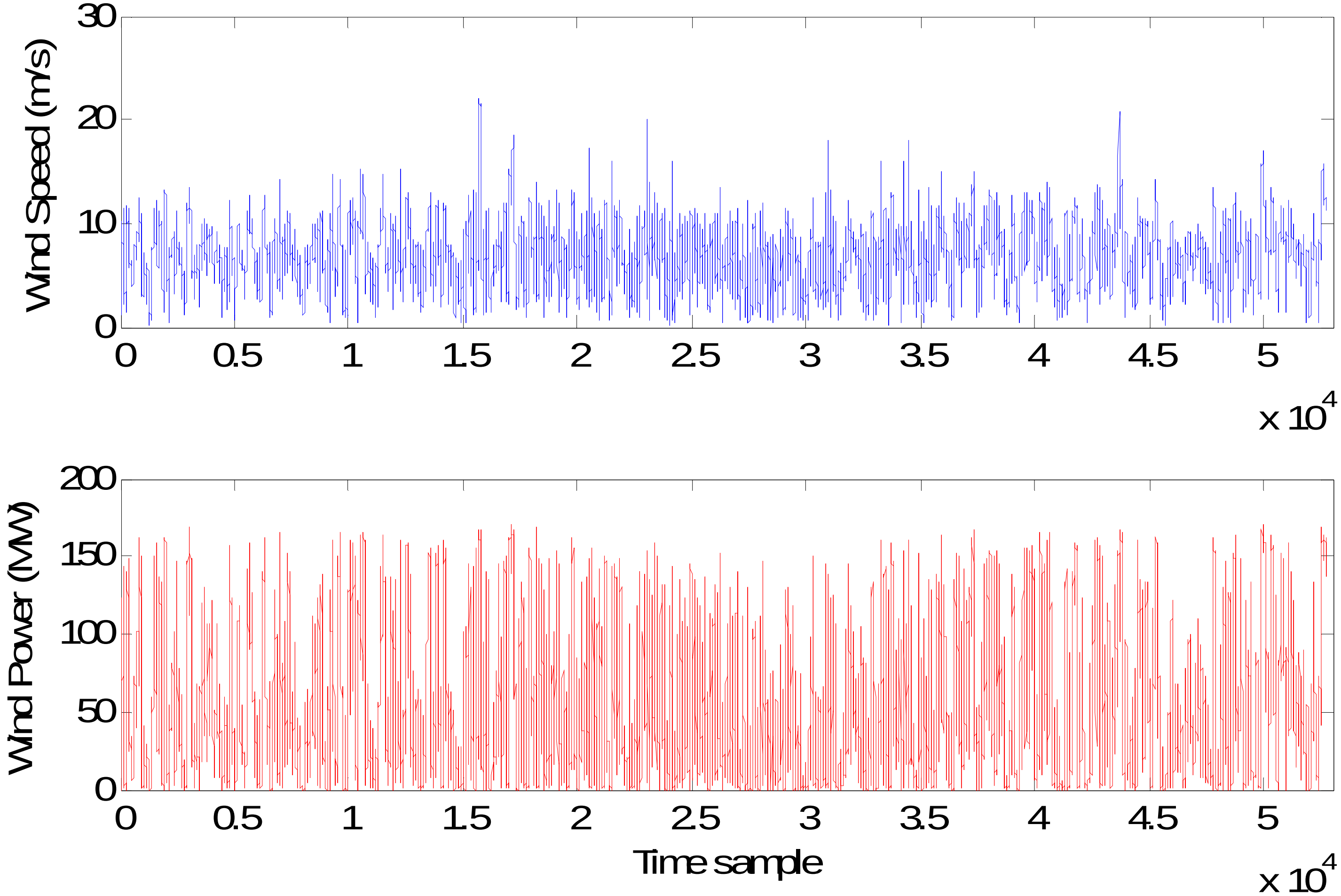




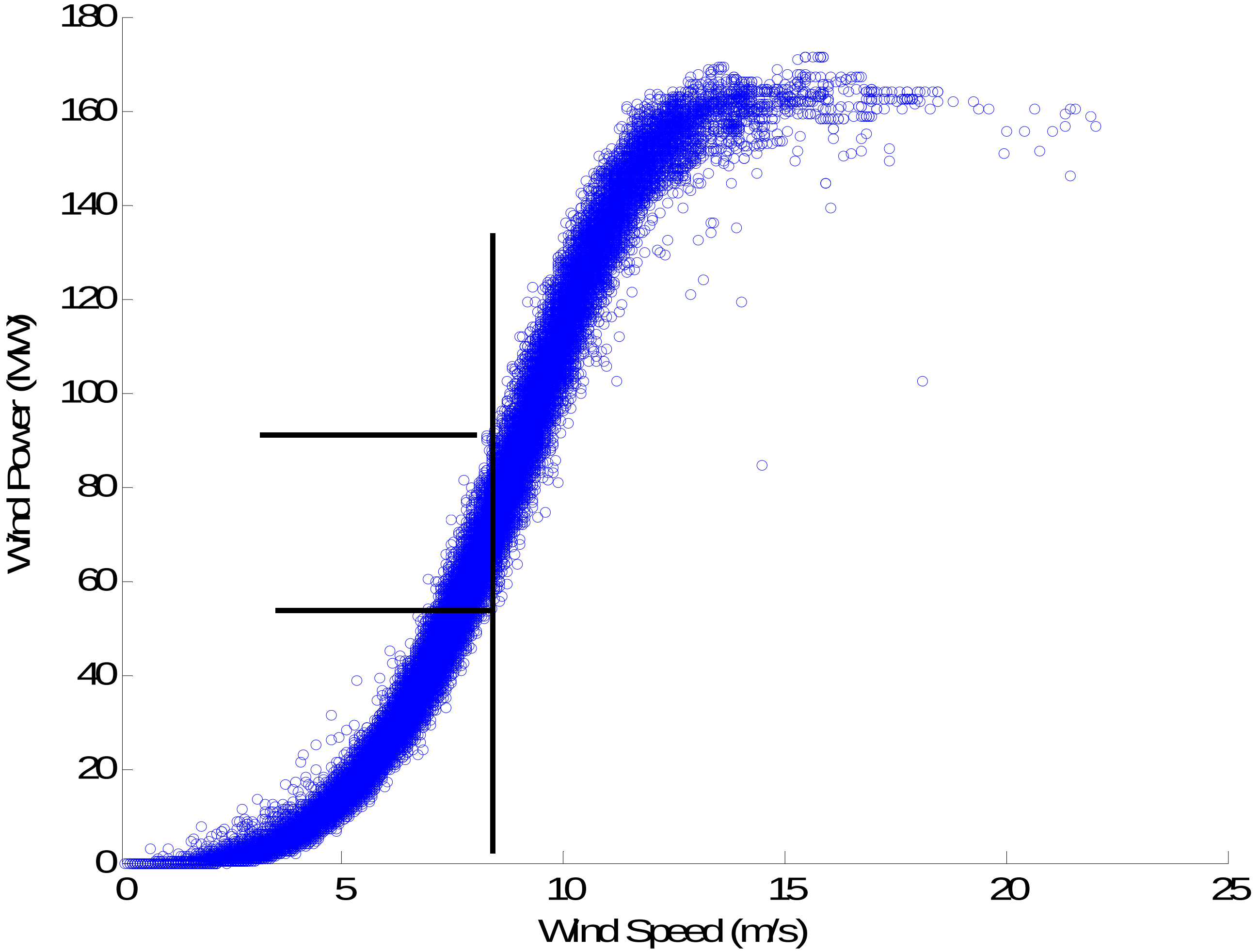
Factors Influencing Wind Power Generation



Wind speed and wind power time series



Wind Speed vs. Wind Power scatter plot



Forecasting Approaches

Linear Regression Models : The forecast value is linearly dependent on the past historical values of the time series. (AR, ARMA, ARIMA, GARCH, etc.)

AR(p): Autoregressive model of order p

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

$$\varphi(B)X_t = \varepsilon_t$$

$$\varphi(B) = 1 - \varphi_1 B_1 - \varphi_2 B_2 \cdots \varphi_p B_p \cdot$$

where B is called as backshift operator;

$$B^m = X_{t-m} \cdot$$

MA(q): Moving Average model of order q

$$X_t = \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}; \qquad X_t = \theta(B)\varepsilon_t$$

$$\theta(B) = 1 - \theta_1 B_1 - \theta_2 B_2 \cdots \theta_q B_q \cdot$$

where ε is white noise error series with mean zero and variance σ^2.

ARMA(p,q): Autoregressive Moving Average model of order (p,q).

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i};$$

$$\varphi(B)X_t = \theta(B)\varepsilon_t$$

Generalized AutoRegressive Conditional Heteroskedasticity (GARCH)

The general process for a GARCH model involves three steps.

- The first is to estimate a best-fitting autoregressive model;
- secondly, compute autocorrelations of the error term and
- lastly, test for significance.

GARCH models are used by financial professionals in several arenas including trading, investing, hedging and dealing.

Forecasting Approaches (Contd..)

ARIMA: Autoregressive Integrated Moving Average Model

ARIMA is a generalized form of ARMA model. It is applied when time series has some non-stationary behavior (do not vary about a fixed mean).

Such a time series can be represented by a generalized autoregressive operator $\emptyset(B)$, in which one or more zeros of the polynomial is unity.

$$\emptyset(B) = \varphi(B)(1 - B)^d \text{, thus}$$

$$\emptyset(B)X_t = \varphi(B)(1 - B)^d X_t = \theta(B)\varepsilon_t$$

$$\varphi(B)W_t = \theta(B)\varepsilon_t$$

where $W_t = \nabla^d X_t$; and $\nabla X_t = X_t - X_{t-1}$ is called as differencing operator.

Fractional-ARIMA:

This type of models are employed when time series exhibits long memory.

F-ARIMA model is a special case of ARIMA(p,d,q) process.

Where parameter *d* assumes fractionally continuous values in the range (-0.5, 0.5).

Since d assumes fractional value, $(1 - B)^d$ is defined through power series expansion as

$$(1 - B)^d = \sum_{j=0}^{\infty} \pi_j B^j$$

where $\pi_0 = 1$, and

$$\pi_j = \prod_{k=1}^{j} (k - 1 - d)/k \text{, j=1,2.....}$$

Correlation Analysis

Finding the appropriate values of *p* and *q* can be facilitated by Partial Auto Correlation Functions (PACF) and Auto-Correlation Functions (ACF).

If Mean of a Time series is given : $$\mu = E[X_t] = \frac{1}{N}\sum_{t=1}^{N} X_t$$

and Variance is given by : $$\sigma^2 = E[(X_t - \mu)^2] = \frac{1}{N}\sum_{t=1}^{N}(X_t - \mu)^2$$

Then auto-correlation at lag *'k'* is given by : $$\rho_k = \frac{E[(X_t - \mu)(X_{t+k} - \mu)]}{\sigma^2}$$

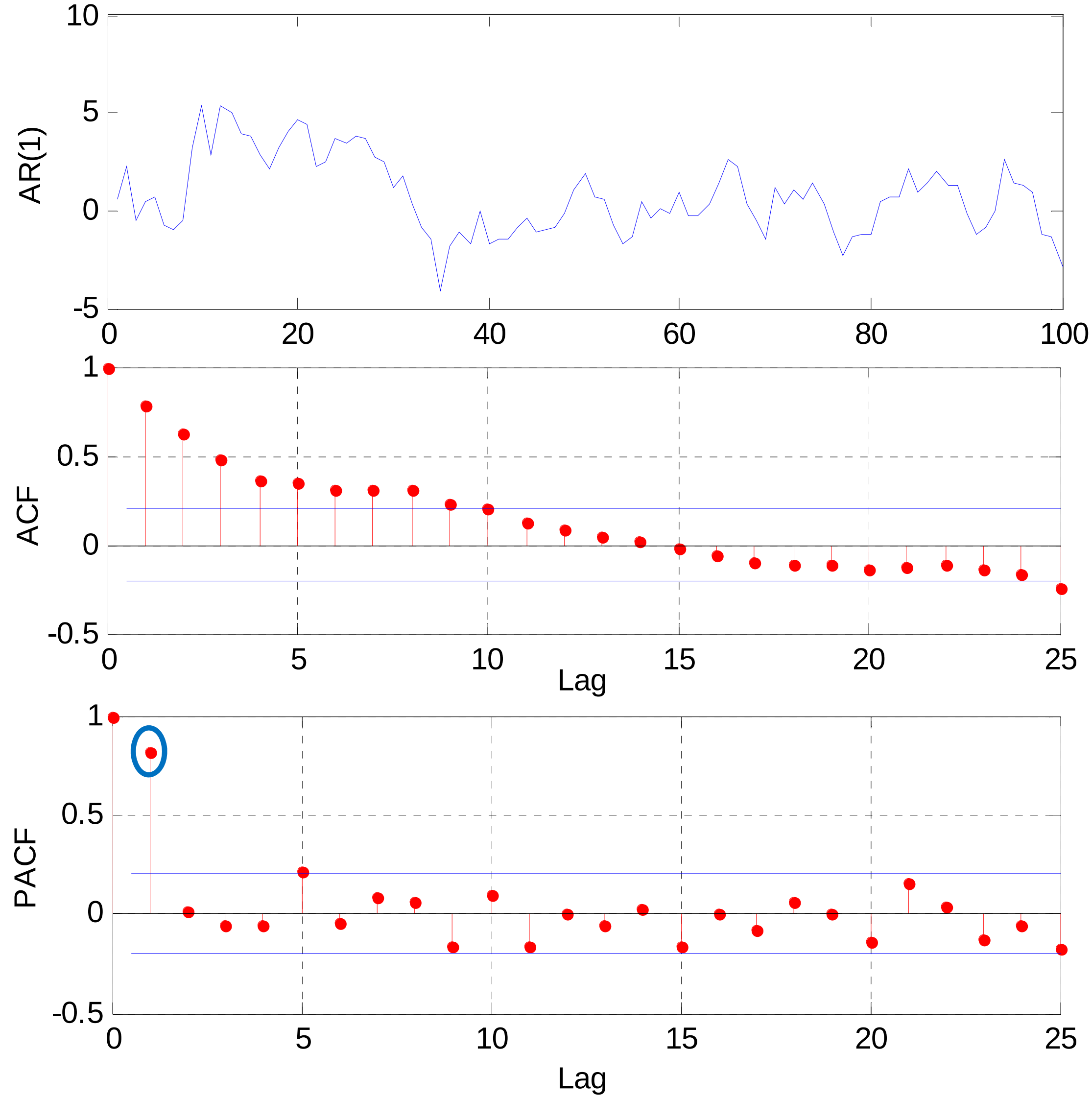
ACF gives the linear relationship (correlation) between two random variable of a time series at different points in time.

Another way to measure the connection between X_t and X_{t+k} is to filter out the linear influence of the random variables that lie in between X_{t+1} and X_{t+k-1}, and then calculate the correlation of the transformed random variables. This is called the *partial* autocorrelation.

Example in Matlab : AR(1)

$$x_t = 0.8x_{t-1} + e_t$$

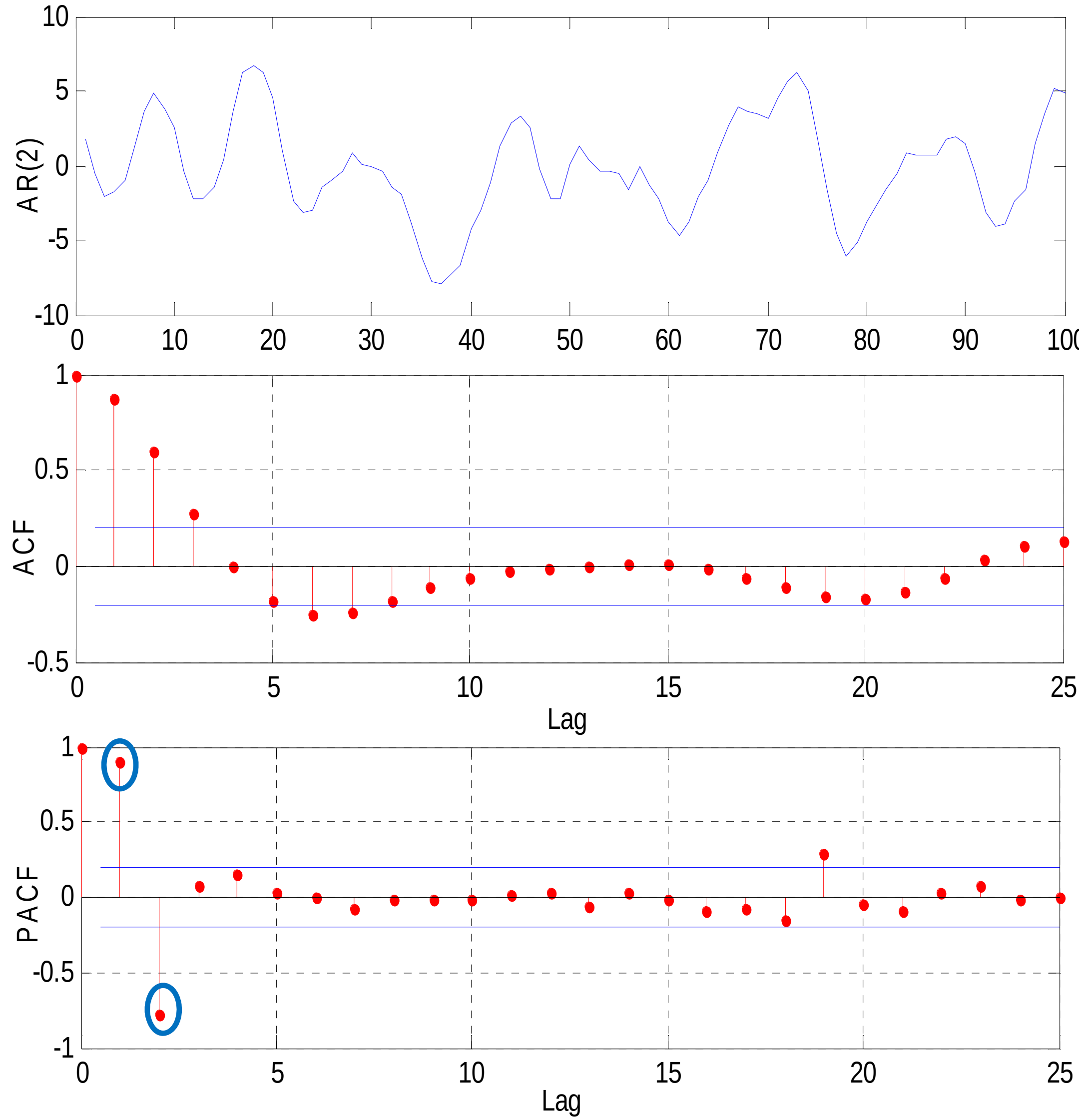
```
e=randn(100,1);
x=zeros(100,1);
x(1)=e(1);
alpha=0.8;
for i=2:100
 x(i)=alpha*x(i-1)+e(i);
 end
subplot(3,1,1)
plot(x);
title('AR(1): x_{t}=\alpha{x_{t-1}+e}, with \alpha{=0.8}');
subplot(3,1,2);
autocorr(x,25,[],2);
subplot(3,1,3);
parcorr(x,25,[],2);
```



Example in Matlab : AR(2)

$$x_t = 1.5x_{t-1} - 0.75x_{t-1} + e_t$$

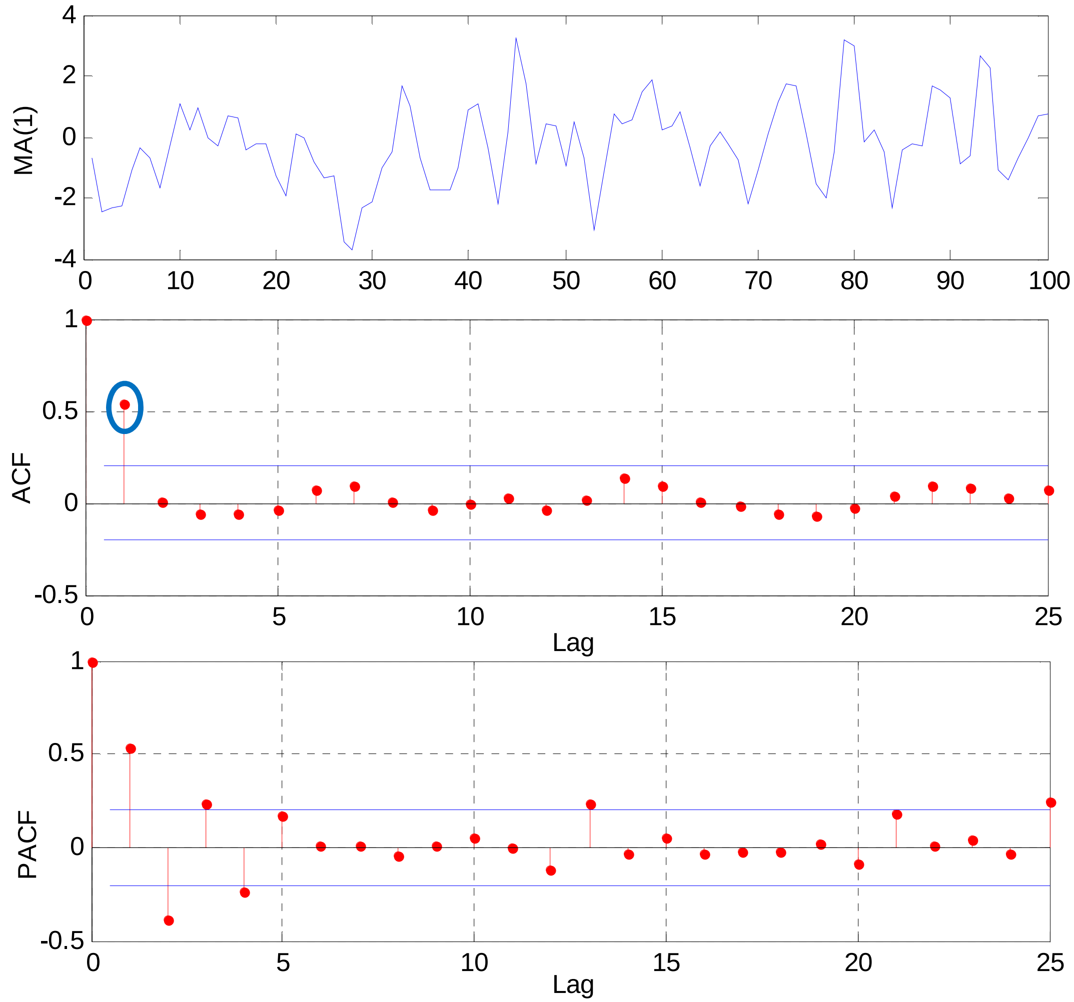
```
e=randn(100,1);
x=zeros(100,1);
x(1)=e(1); x(2)=e(2);
alpha1=1.5; alpha2 = -0.75;
for i=3:100, x(i)=alpha1*x(i-1)+alpha2*x(i-2)+e(i); end
subplot(3,1,1)
plot(x);
title('AR(2): x_{t}=\alpha_{1}x_{t-1}+\alpha_{2}x_{t-2}+e, with \alpha{1}=1.5, \alpha_{2}=-0.75');
subplot(3,1,2);
autocorr(x,25,[],2);
subplot(3,1,3);
parcorr(x,25,[],2);
```



Example in Matlab : MA(1)

```
e=randn(101,1);
theta=-0.8;
x=e(2:101,1)-theta*e(1:100,1);
subplot(3,1,1)
plot(x);
title('MA(1): x_{t}=\theta{x_{t-1}+e}, with \theta{=0.8}');
subplot(3,1,2);
autocorr(x,25,[],2);
subplot(3,1,3);
parcorr(x,25,[],2);
```

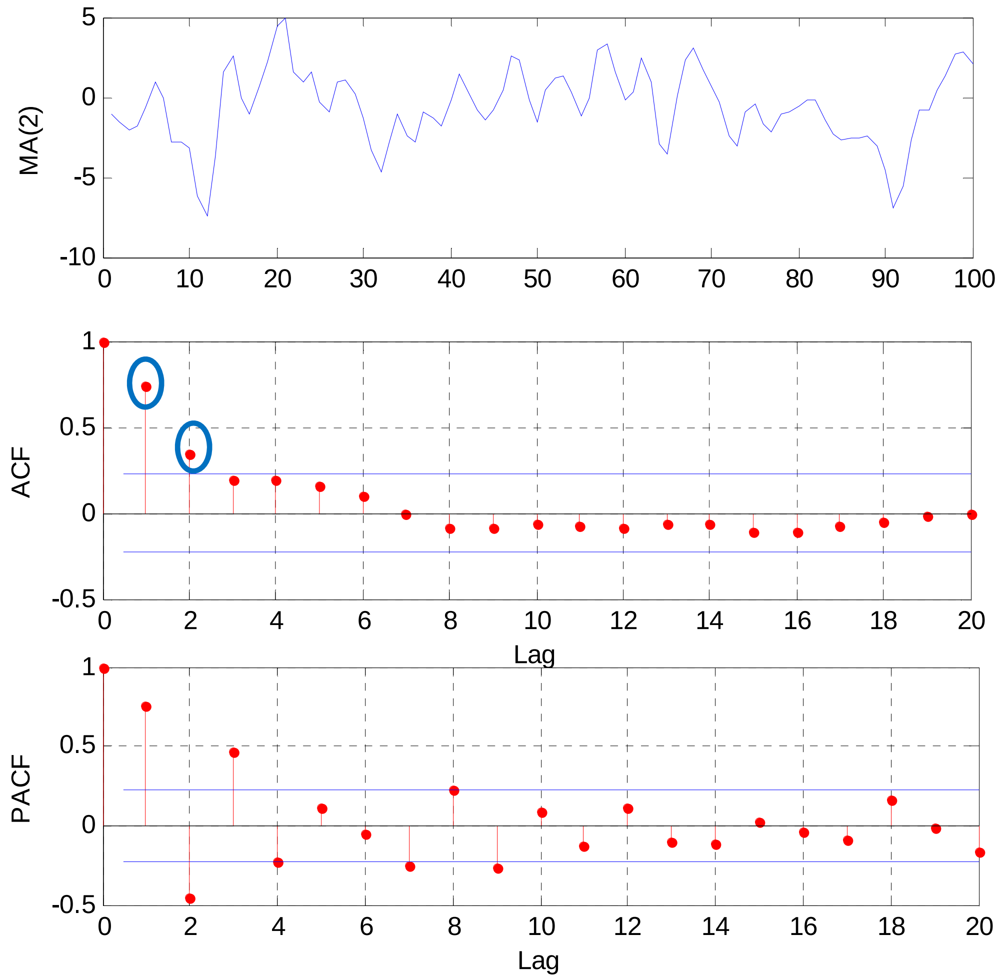
$$x_t = e_t + 0.8e_{t-1}$$



Example in Matlab : MA(2)

```
e=randn(102,1);
theta1=-1.5; theta2=-0.8;
x=e(3:102,1) - theta1*e(2:101,1) -
theta2*e(1:100,1);
subplot(3,1,1);
plot(x);
title('MA(2): x_{t}=e_{t}-
\theta_{1}e_{t-1}-\theta_{2}e_{t-2} );
subplot(3,1,2);
autocorr(x(20:end),20,[],2);
subplot(3,1,3);
parcorr(x(20:end),20,[],2);
```

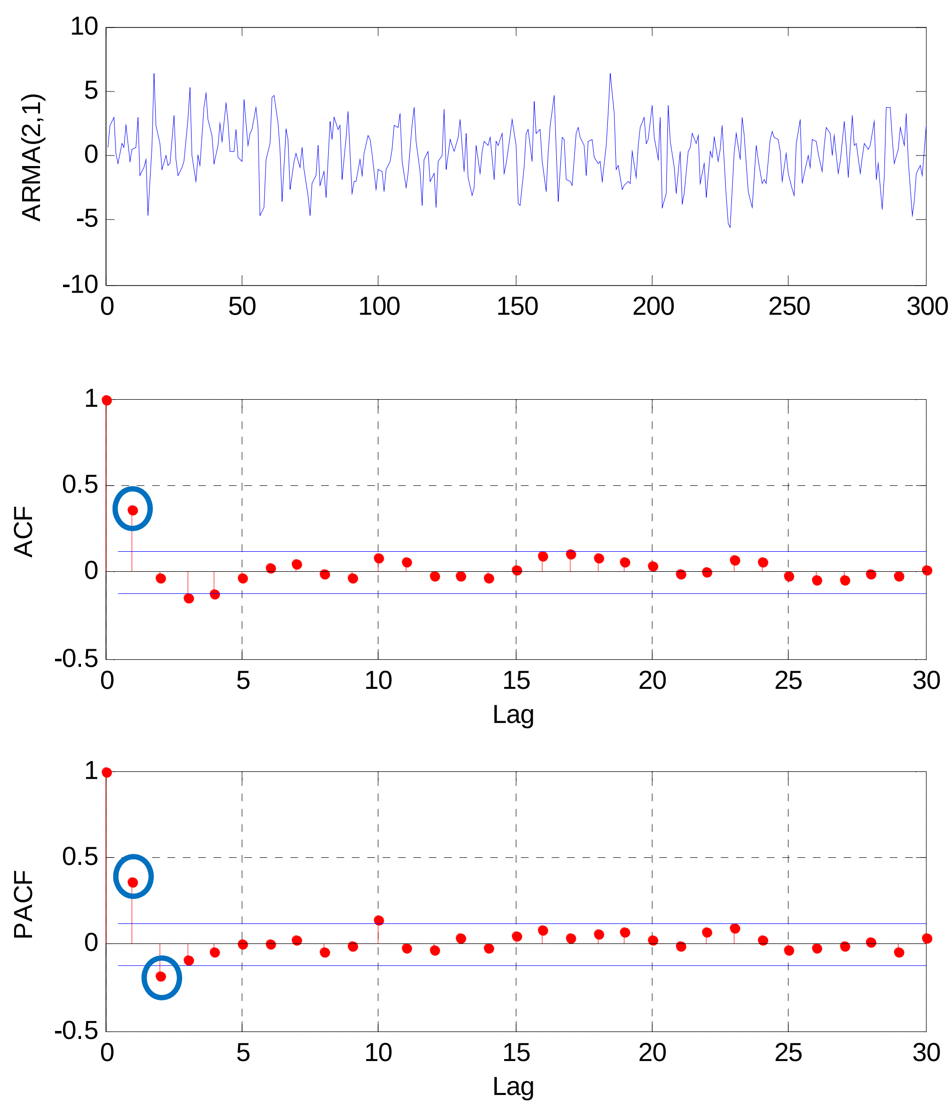
$$x_t = e_t + 1.5e_{t-1} + 0.8e_{t-2}$$



Example in Matlab : ARMA(2,1)

```
clear all;
close all;
e=2*randn(300,1);
x=zeros(300,1);
x(1)=e(1); x(2)=e(2);
alpha1=0.9; alpha2 = -0.3;
theta1= 0.5;
for i=3:300
 x(i)=alpha1*x(i-1)+
 alpha2*x(i-2)+
e(i)-theta1*e(i-1) ;
end
subplot(3,1,1)
plot(x);
subplot(3,1,2);
autocorr(x(25:end),30,[],2);
subplot(3,1,3);
parcorr(x(25:end),30,[],2);
```

$$x(t) = 0.9x(t-1) - 0.3x(t-2) + e_t - 0.5e(t-1)$$



ACF and PACF for casual Time series models

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

Estimation of Model Parameters

After choosing *p* and *q* the model can be fitted by linear least squares regression to find the model parameters which minimize the error terms.

Let X be the $(n \times p)$ data matrix of n observations on the p variables

$$\boldsymbol{X} = \begin{pmatrix} x_{1,1} & \cdots & x_{1,p} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,p} \end{pmatrix} = (\boldsymbol{x_1}, \cdots, \boldsymbol{x_p})$$

and $\boldsymbol{y} = (y_1, \cdots, y_n)^T$ be the output vector.

least square estimator denoted by $\hat{\beta}$ is that value of b which minimizes the

$$||\boldsymbol{y} - \boldsymbol{X}b||^2.$$

Which is the squared distance between the vector $\boldsymbol{y}$ and the linear combination b of the columns of the matrix $\boldsymbol{X}$. The distance is minimized by taking the projection of y on the space spanned by columns of $\boldsymbol{X}$.

The least squares estimator $\hat{\beta}$ is given by $\hat{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$.

State-Space Models

Consider a univariate *AR(p)* process:

$$(y_{t+1} - \mu) = \phi_1(y_t - \mu) + \phi_2(y_{t-1} - \mu) + \cdots + \phi_p(y_{t-p+1} - \mu) + \varepsilon_{t+1}, \qquad \varepsilon_t \sim \text{i.i.d. } N(0, \sigma^2).$$

This could be written in state-space form as,

$$\begin{bmatrix} y_{t+1} - \mu \\ y_t - \mu \\ \vdots \\ y_{t-p+2} - \mu \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} y_t - \mu \\ y_{t-1} - \mu \\ \vdots \\ y_{t-p+1} - \mu \end{bmatrix} + \begin{bmatrix} \varepsilon_{t+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

state equation $$\boldsymbol{\xi}_{t+1} = \mathbf{F}\boldsymbol{\xi}_t + \mathbf{v}_{t+1}$$

$$y_t = \mu + \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} y_t - \mu \\ y_{t-1} - \mu \\ \vdots \\ y_{t-p+1} - \mu \end{bmatrix}$$

observation equation $$\mathbf{y}_t = \mathbf{A}'\mathbf{x}_t + \mathbf{H}'\boldsymbol{\xi}_t + \mathbf{w}_t$$

$$\mathbf{Q} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$\mathbf{R} = 0.$$

Limitations of Linear Regression Models

As they are linear models, they cannot capture the non-linear relation between the independent and dependent variable.

The forecasting error increases rapidly with the increase in look-ahead time.

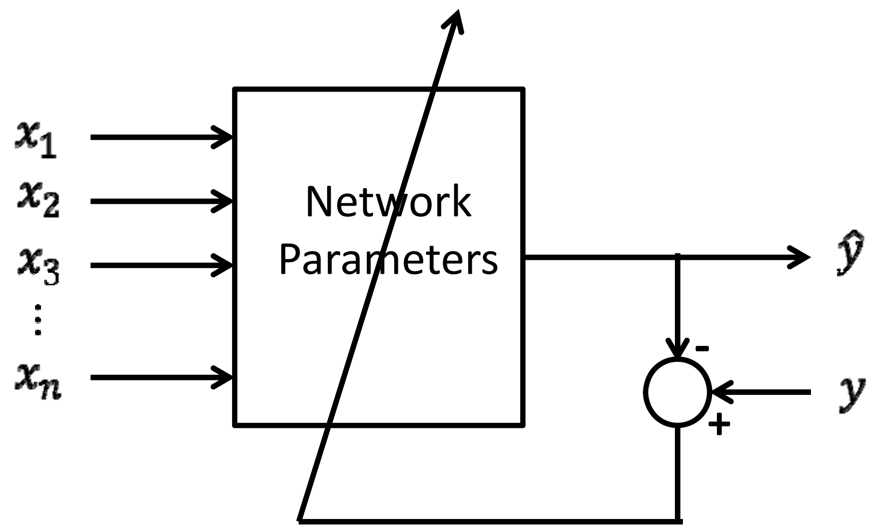
The model parameters have to be updated very frequently.

Forecasting Approaches .....contd

Non-Linear Regression models:

$$X_t = F(X_{t-1}, X_{t-2}, X_{t-3}, \cdots, u_t, u_{t-1}, u_{t-2}, \cdots) + \varepsilon_t$$

Artificial Neural Networks (ANN): are well established in function approximation, many variants of NNs are employed in the field of forecasting problem. Like FFNN, RNN, RBF, WNN.



Back-Propagation Algorithm, Evolutionary based Optimization methods like GA, PSO are also applied for network training.

Input variables are selected using ACF and PACF.

Other Methods..

- Fuzzy Logic
- Adaptive Neuro-Fuzzy Inference System (ANFIS)
- Data Mining techniques like clustering and Support Vector Machines (SVM) based classification and Regression models
- Wavelet pre-filtering based ANN and Fuzzy models.

Benchmark Models

Persistence Model:

Also called as naive predictor, the most common benchmark model, which states that future wind production remains the same as the last measured value of the power;

$$\hat{P}(t+k|t) = P(t).$$

Drawback: forecast error increases rapidly with the increase in look-ahead time

New Reference Model: $\hat{P}(t+k|t) = a_k P(t) + (1 - a_k)\bar{P}(t)$

The constant a_k is defined as the correlation coefficient between $P(t)$ and $P(t+k)$.

Measure of Errors

If error is given as; $e(t+k|t) = P(t+k) - \hat{P}(t+k|t).$

Then, $$BIAS(k) = \bar{e}_k = \frac{1}{N}\sum_{t=1}^{N} e(t+k|t)$$

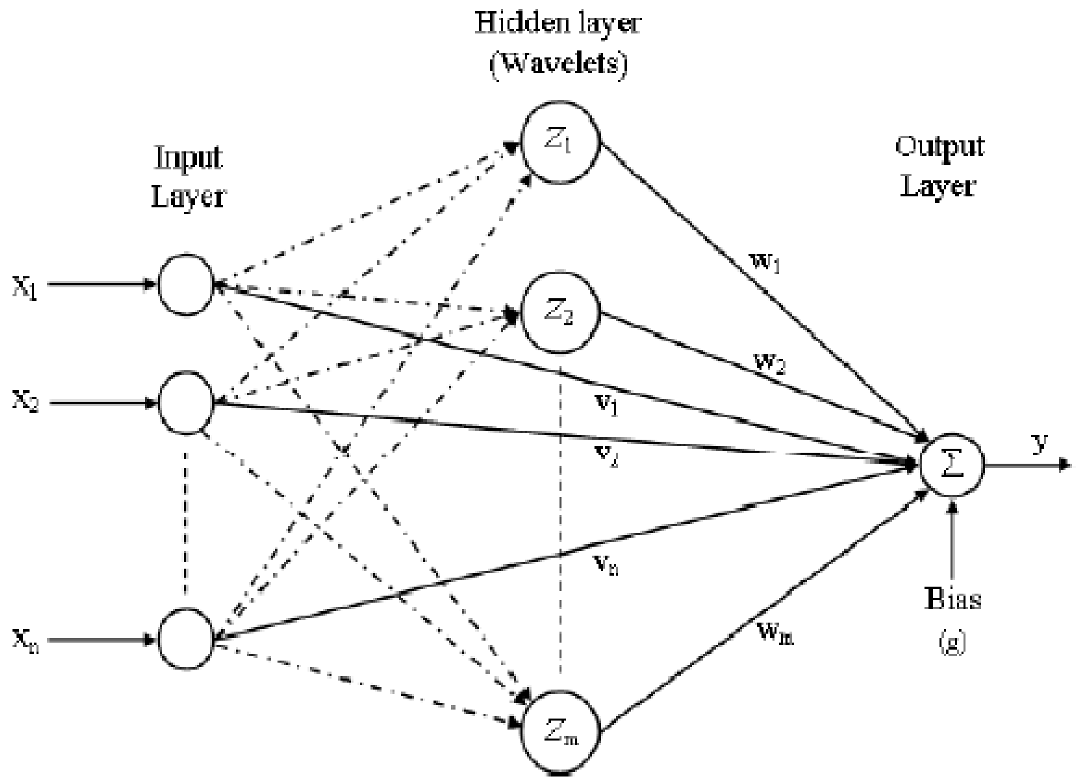
$$MAE(k) = \frac{1}{N}\sum_{t=1}^{N} |e(t+k|t)|$$

$$RMSE(k) = \left[\frac{1}{N}\sum_{t=1}^{N} e^2(t+k|t)\right]^{1/2}$$

AWNN : Architecture

$$\psi_{a,b}(x_i) = \left(1 - \left(\frac{x_i - b}{a}\right)^2\right) e^{-0.5\left(\frac{x_i - b}{a}\right)^2}$$

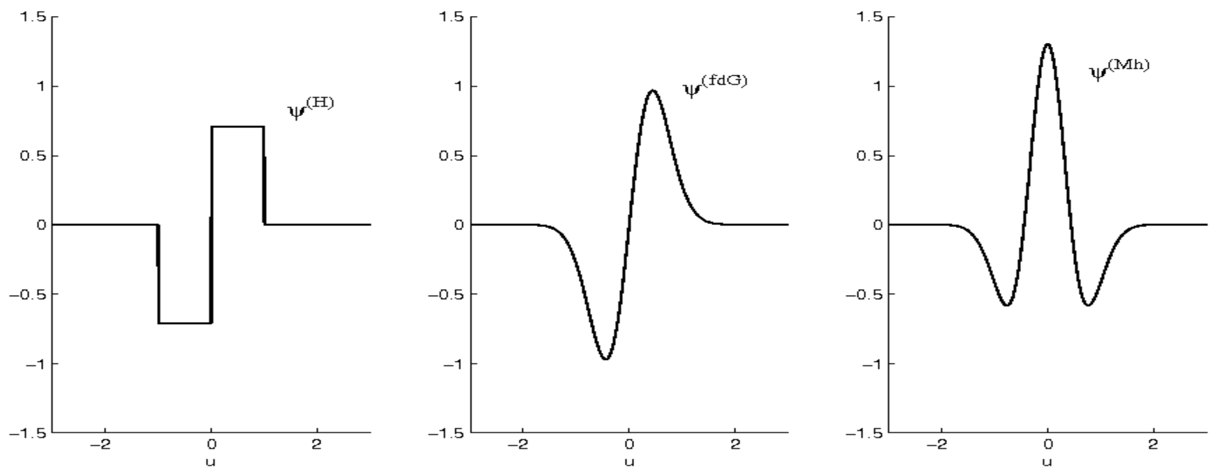
$$z_j = \prod_{i=1}^{n} \psi_{a_{ij},b_{ij}}(x_i) \quad j \in m$$



$$y = \sum_{j=1}^{m} w_j z_j + \sum_{i=1}^{n} v_i x_i + g$$

Continuous Wavelet Transforms

- A wavelet is a small wave which grow and decays essentially in a limited time period.



- Should satisfy two basic properties

$$\int_{-\infty}^{\infty} \psi(\cdot).du = 0 \qquad \text{and} \qquad \int_{-\infty}^{\infty} \psi^2(\cdot).du = 1.$$

Training Algorithm

$$E = \frac{1}{2N}\sum_{p=1}^{N}[e(p)]^2 \quad \text{and} \quad e(p) = y^d(p) - y(p)$$

$$a(k+1) = a(k) + \eta\Delta a(k) + \alpha\Delta a(k-1)$$

$$\Delta w_j = -\frac{\partial E}{\partial w_j} = -\frac{\partial E}{\partial y}\frac{\partial y}{\partial w_j} = e z_j \quad j \in m$$

$$\eta(n+1) = \begin{cases} 1.05\,\eta(n) & , \Delta E_n > 0 \\ 0.7\,\eta(n) & , otherwise \end{cases}$$

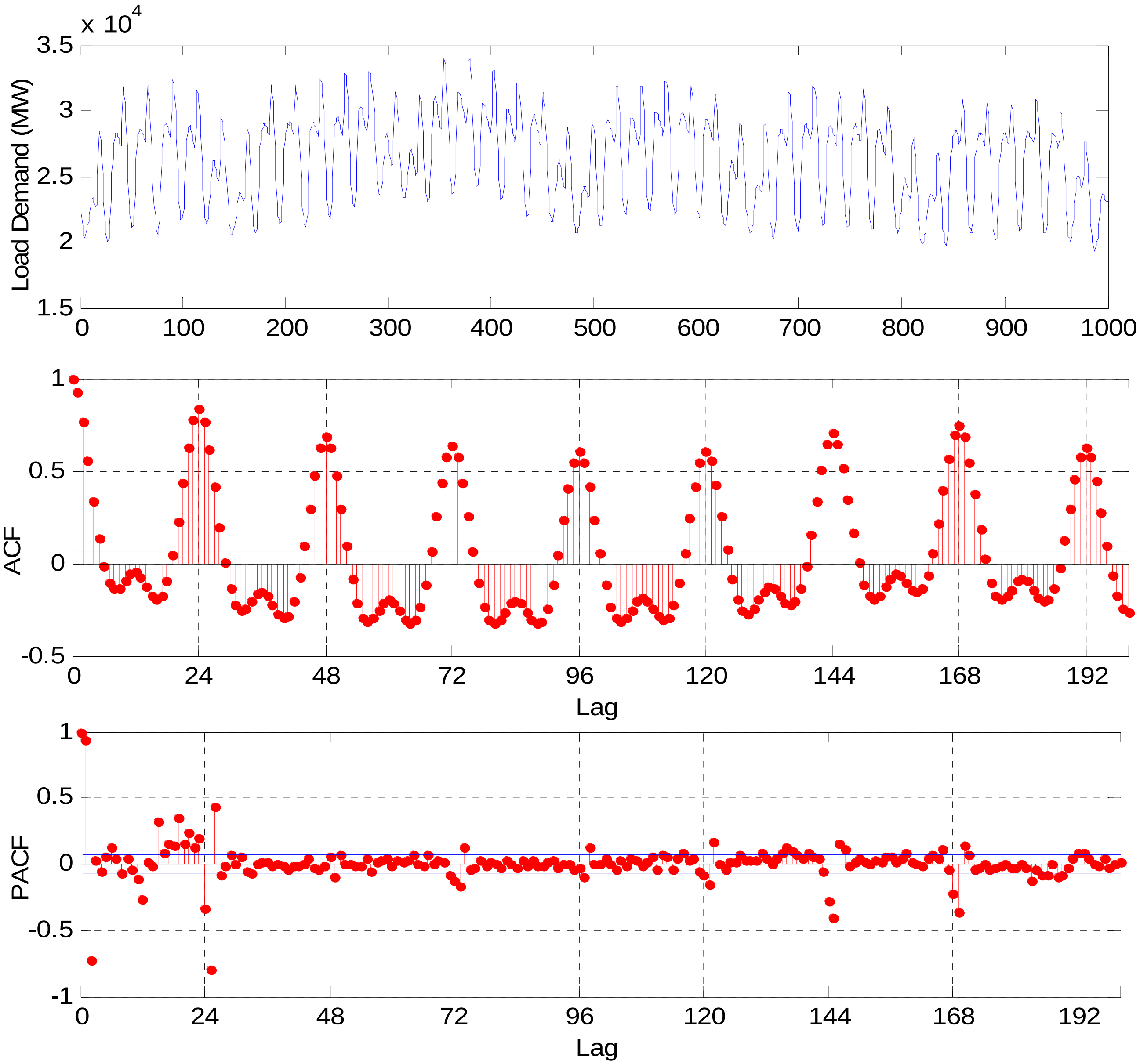
$$\Delta v_i = -\frac{\partial E}{\partial v_i} = -\frac{\partial E}{\partial y}\frac{\partial y}{\partial v_i} = e x_i \quad i \in n$$

$$\Delta g = -\frac{\partial E}{\partial g} = -\frac{\partial E}{\partial y}\frac{\partial y}{\partial g} = e$$

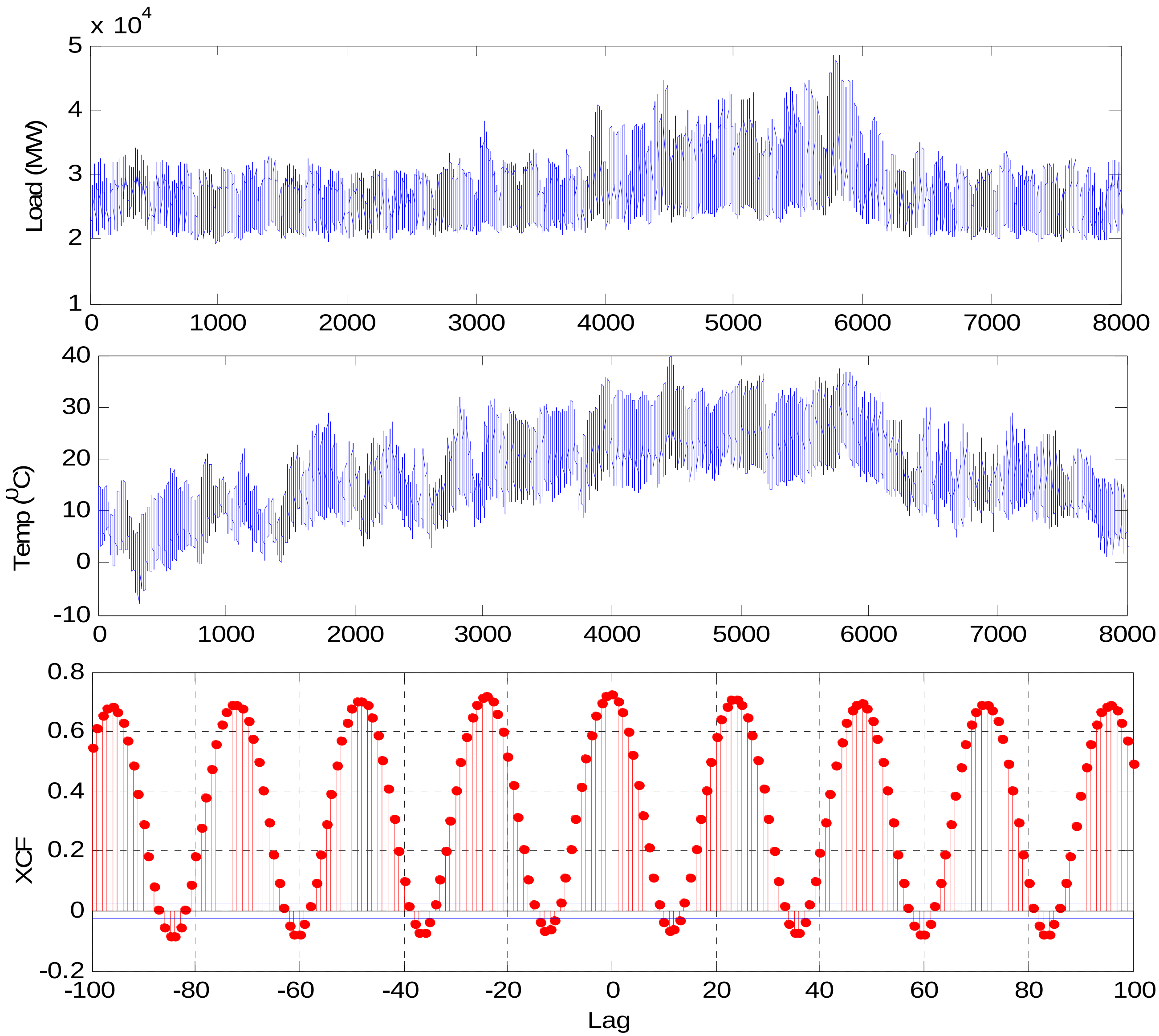
$$\Delta a_{ij} = -\frac{\partial E}{\partial a_{ij}} = -\frac{\partial E}{\partial y}\frac{\partial y}{\partial z_j}\frac{\partial z_j}{\partial \psi_{ij}}\frac{\partial \psi_{ij}}{\partial a_{ij}} = e w_j z_j \left[\frac{1}{a_{ij}}\right]\left[\frac{x_i - b_{ij}}{a_{ij}}\right]^2\left[3 - \left[\frac{x_i - b_{ij}}{a_{ij}}\right]^2\right] e^{-0.5\left[\frac{x_i - b_{ij}}{a_{ij}}\right]^2}$$

$$\Delta b_{ij} = -\frac{\partial E}{\partial b_{ij}} = -\frac{\partial E}{\partial y}\frac{\partial y}{\partial z_j}\frac{\partial z_j}{\partial \psi_{ij}}\frac{\partial \psi_{ij}}{\partial b_{ij}} = e w_j z_j \left[\frac{1}{a_{ij}}\right]\left[\frac{x_i - b_{ij}}{a_{ij}}\right]\left[3 - \left[\frac{x_i - b_{ij}}{a_{ij}}\right]^2\right] e^{-0.5\left[\frac{x_i - b_{ij}}{a_{ij}}\right]^2}$$

Case study: Load Forecasting

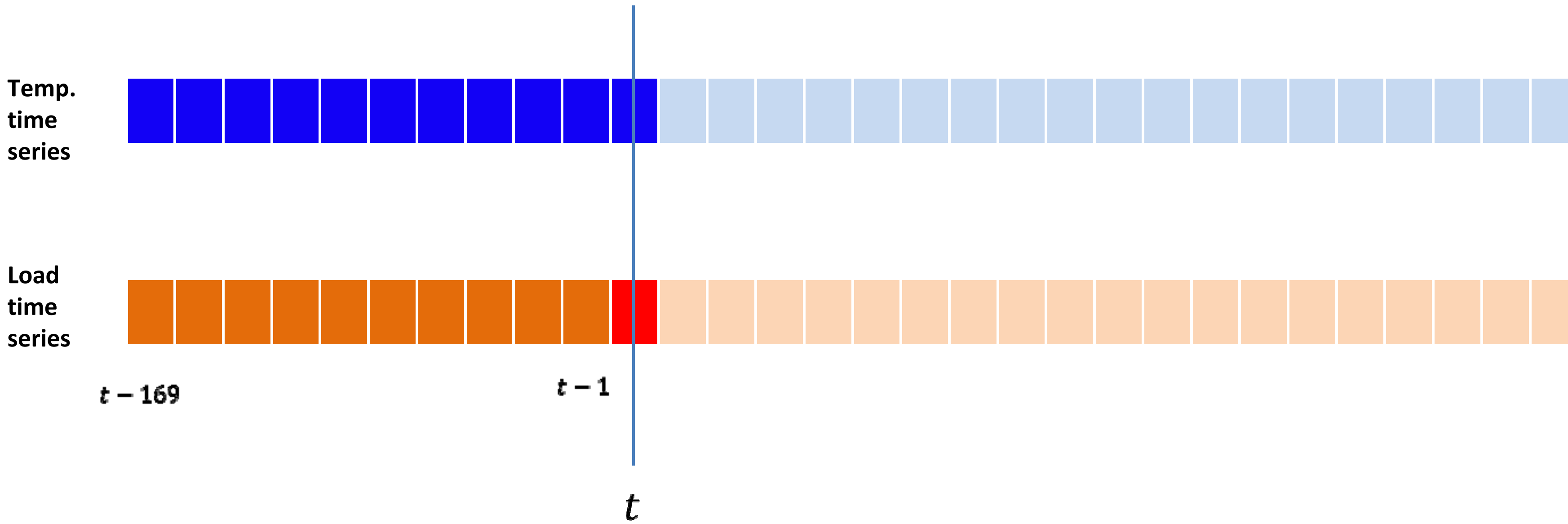


Cross correlation between Load and Temp.

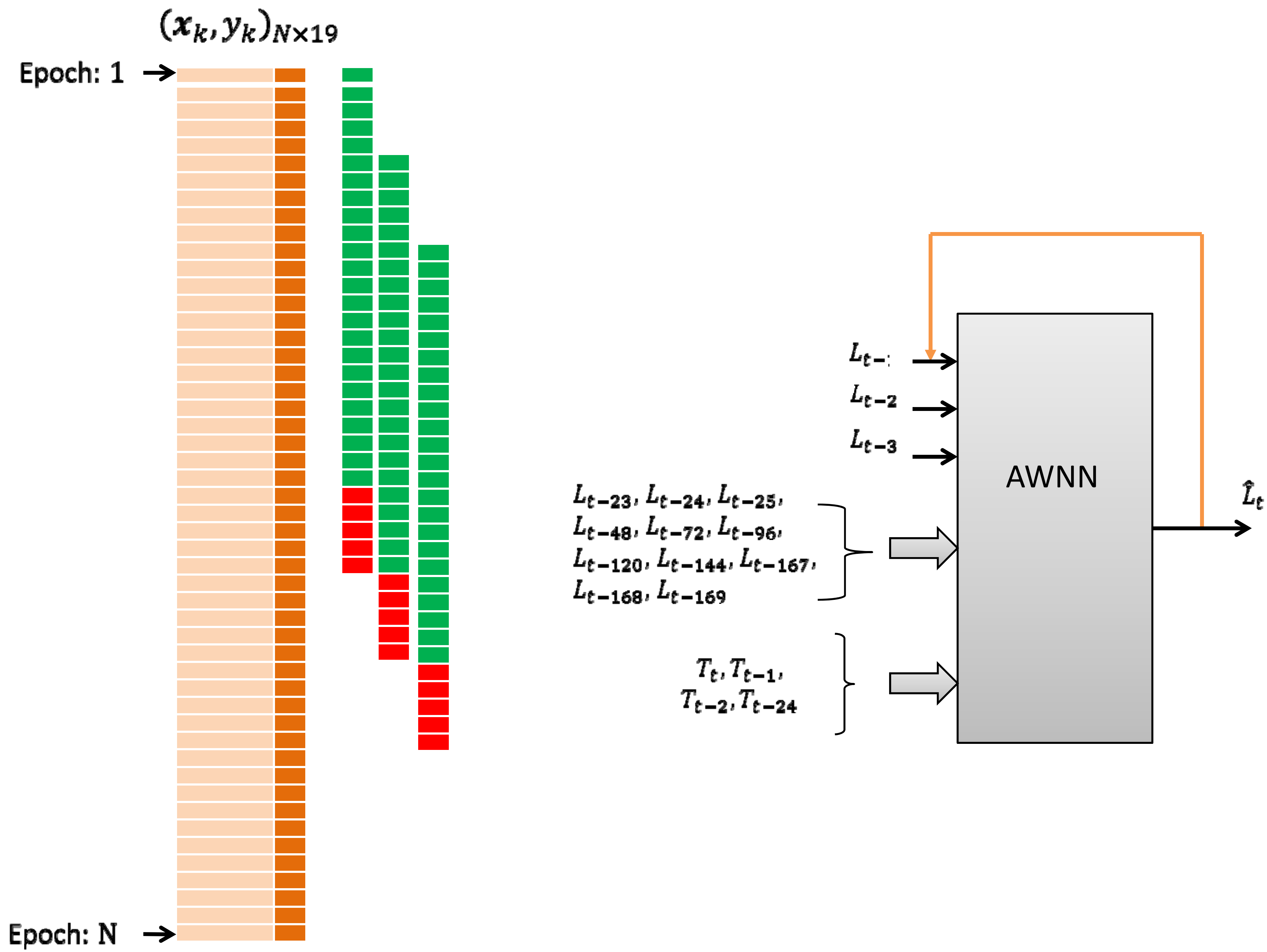


Input lag hours selected for Load Forecasting

Load Terms	$L_{t-1}, L_{t-2}, L_{t-3}, L_{t-23}, L_{t-24}, L_{t-25}, L_{t-48}, L_{t-72}, L_{t-96}, L_{t-120}, L_{t-144}, L_{t-167}, L_{t-168}, L_{t-169}$
Temp. Terms	$T_t, T_{t-1}, T_{t-2}, T_{t-24}$

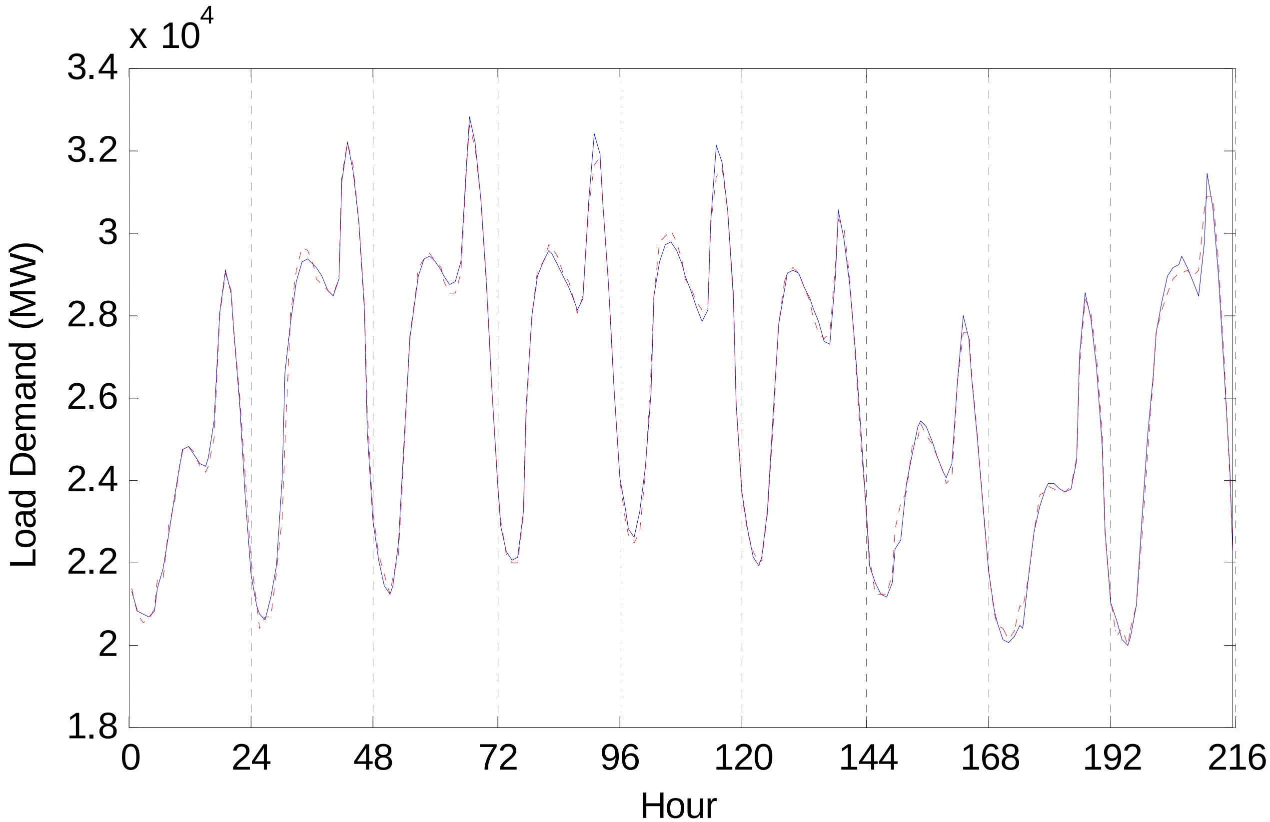


Epochs: (input, output) pairs ➡ $(x_k, y_k) \quad for \;\; k = 1,2,\cdots,N$



$(x_k, y_k)_{N\times19}$
Epoch: 1
Epoch: N
L_{t-1}
L_{t-2}
L_{t-3}
$L_{t-23}, L_{t-24}, L_{t-25},$
$L_{t-48}, L_{t-72}, L_{t-96},$
$L_{t-120}, L_{t-144}, L_{t-167},$
L_{t-168}, L_{t-169}
$T_t, T_{t-1},$
T_{t-2}, T_{t-24}
AWNN
$\hat{L}_t$

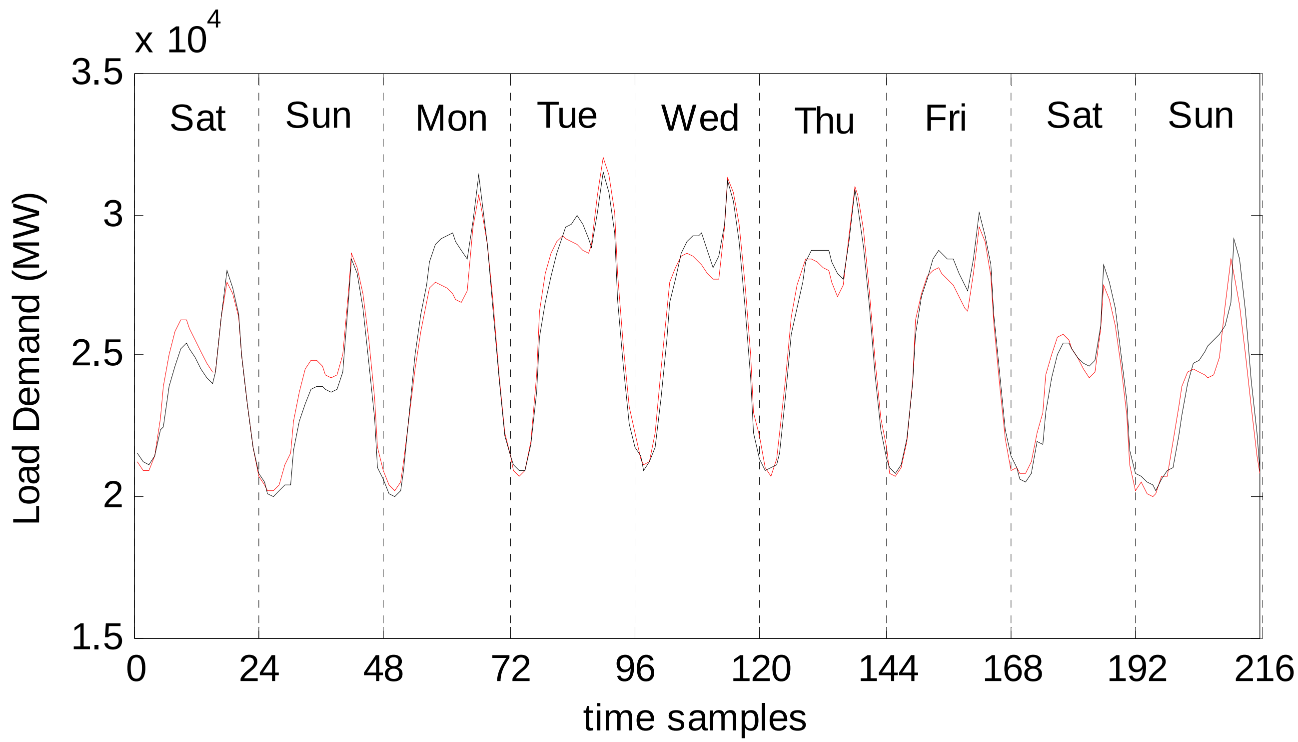
Hour ahead load forecast



$$e(t+k|t) = x(t+k) - \hat{x}(t+k|t)$$

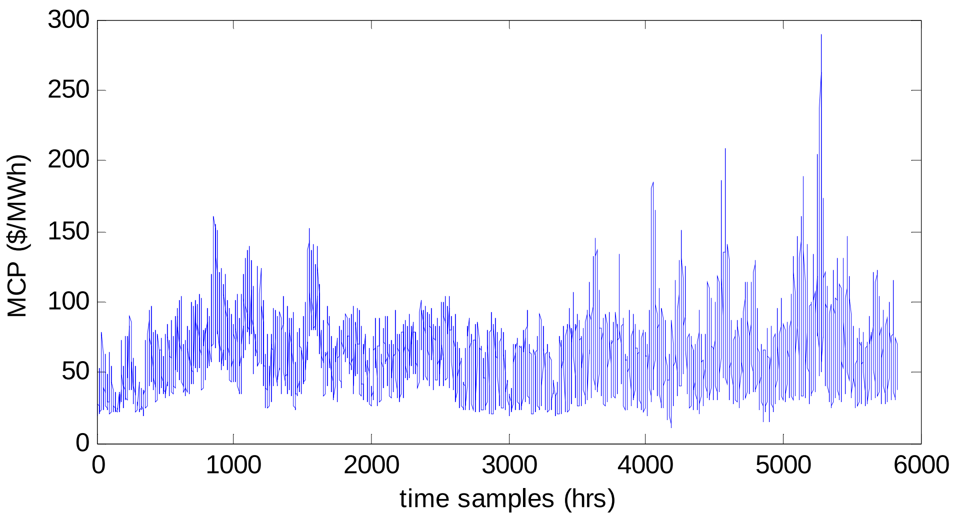
$$MAPE(k)\% = \frac{1}{N}\sum_{t=1}^{N}\frac{|e(t+k|t)|}{x(t+k)} \times 100$$ **= 0.6905**

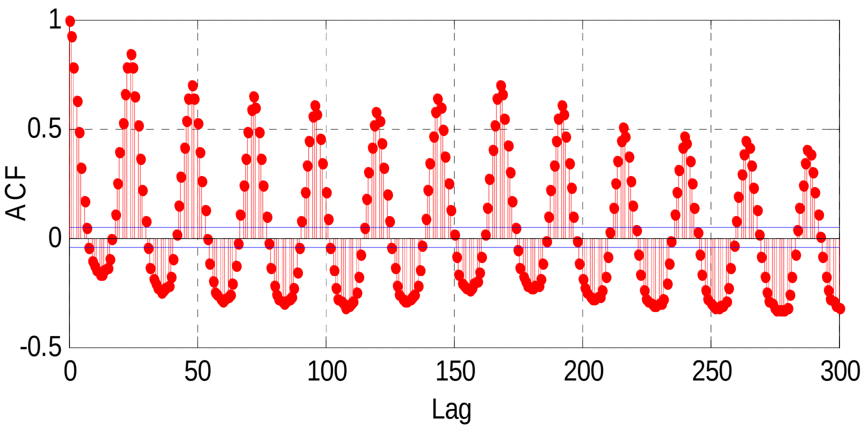
24-hours ahead Forecast

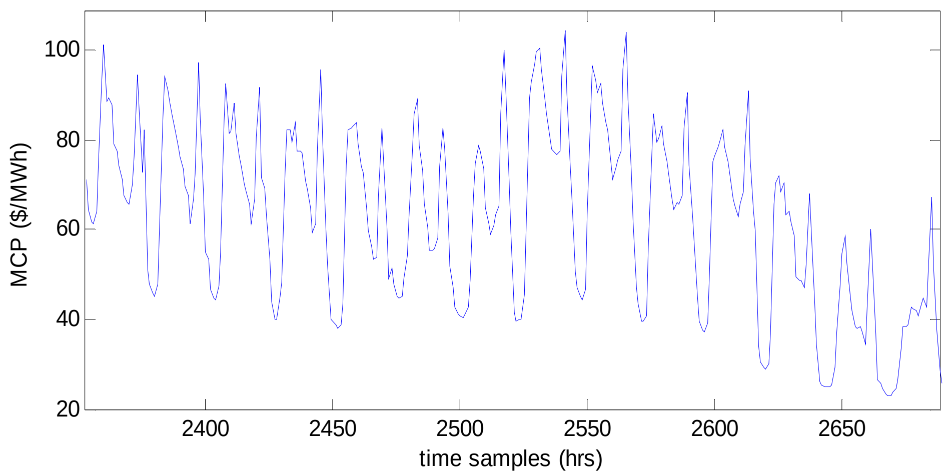


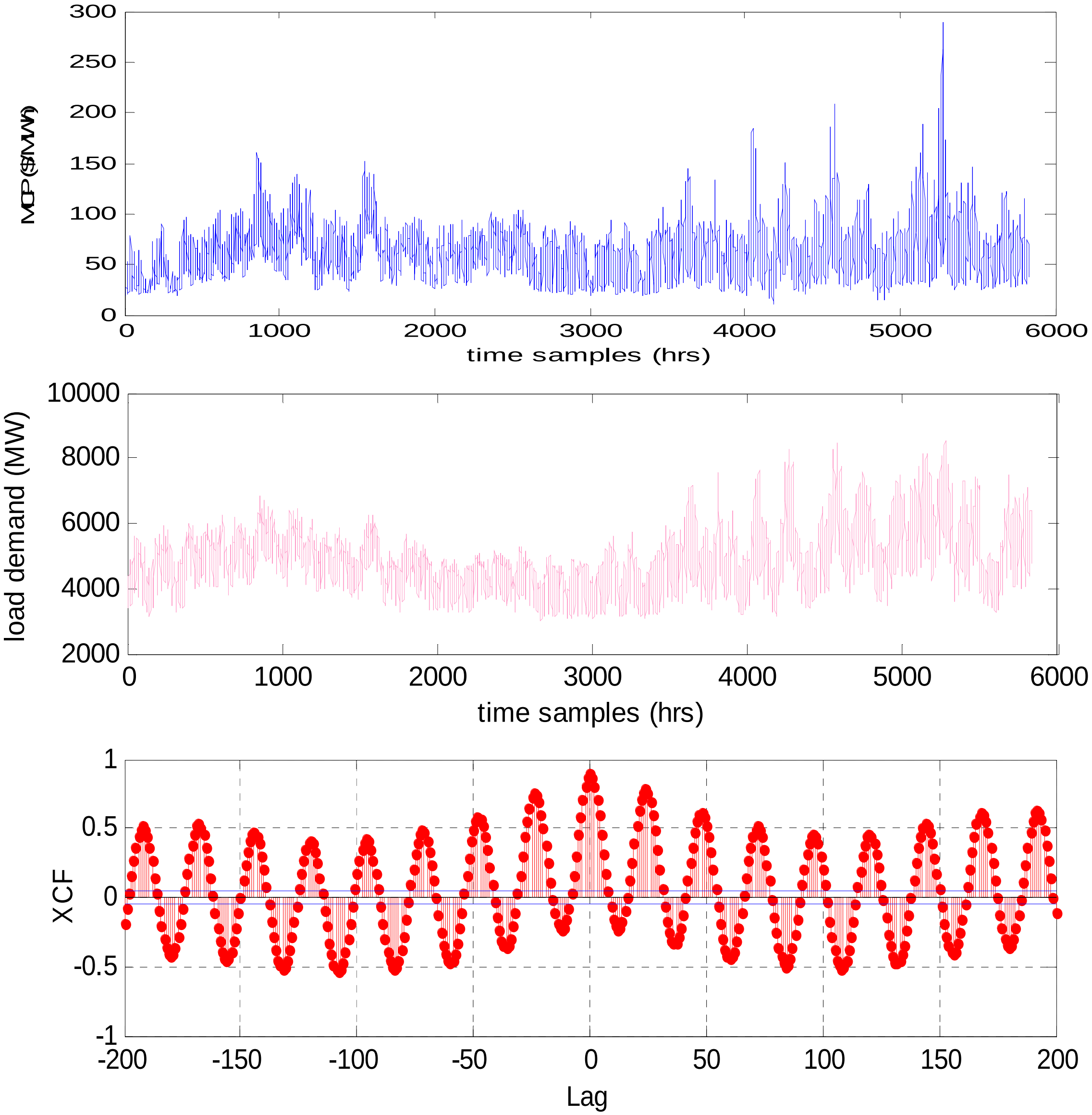
Forecast hour	MAPE
1	0.9551
2	1.5191
3	1.9664
4	2.2592
5	2.4768
6	2.5779
7	2.6639
8	2.6883
9	2.6556
10	2.6816
11	2.7565
12	2.8146
13	2.8640
14	2.8843
15	2.8949
16	2.9113
17	2.9190
18	2.9096
19	2.8514
20	2.8290
21	2.8039
22	2.7618
23	2.7373
24	2.7447
Average	**2.5886**

Case study: Price Forecasting



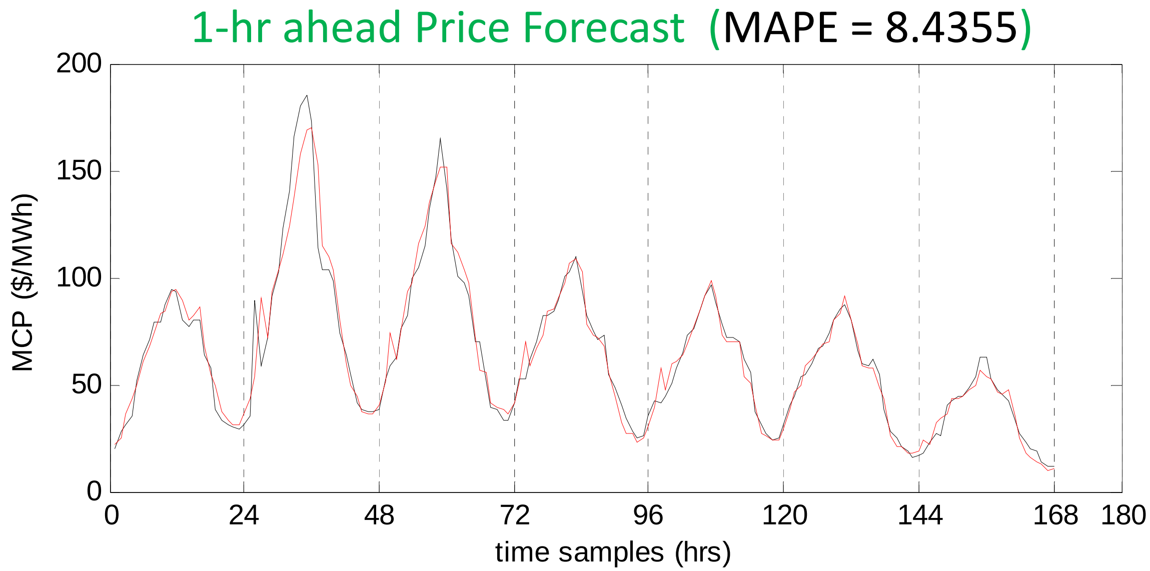


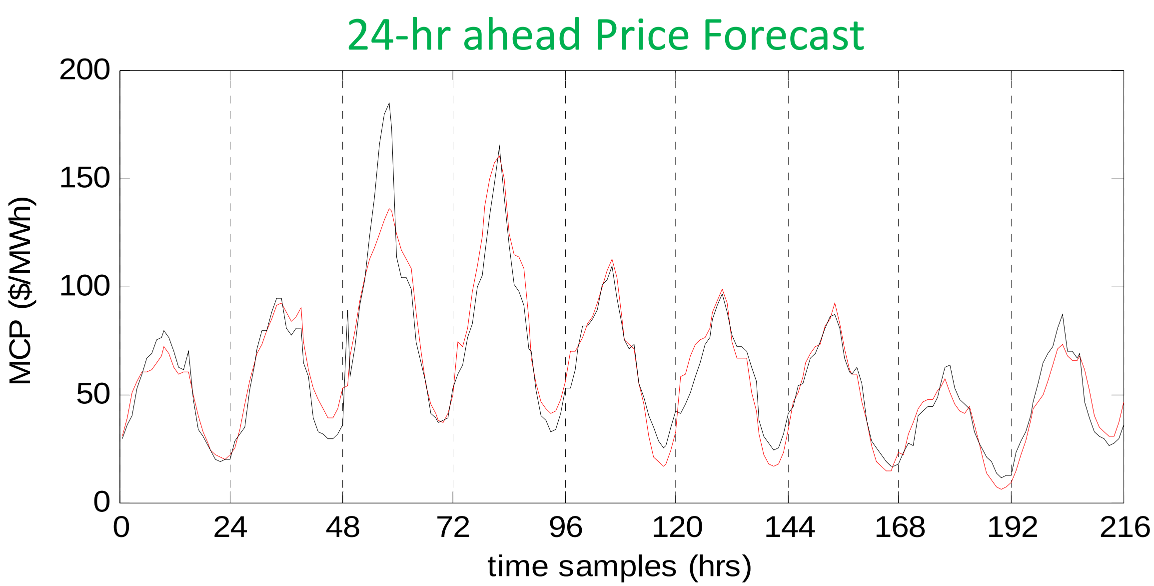




Price Terms	$P_{t-1}, P_{t-2}, P_{t-3}, P_{t-23}, P_{t-24}, P_{t-25}, P_{t-48}, P_{t-72}, P_{t-96}, P_{t-120}, P_{t-144}, P_{t-167}, P_{t-168}, P_{t-169}, P_{t-193}$
Load Terms	$L_t, L_{t-1}, L_{t-2}, L_{t-3}, L_{t-24}, L_{t-48}, L_{t-72}, L_{t-96}, L_{t-120}, L_{t-144}, L_{t-168}$

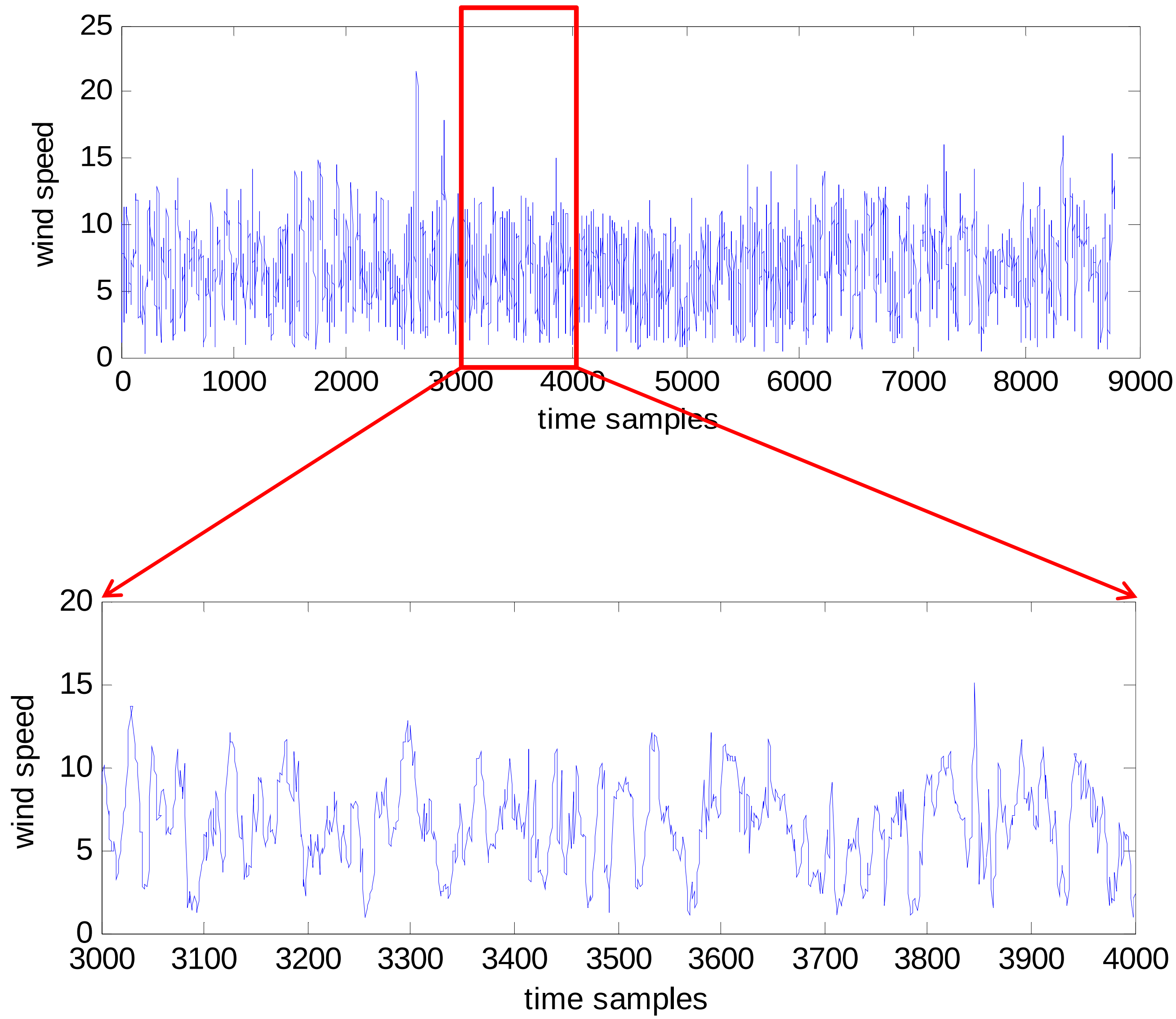
Price Forecasting Results



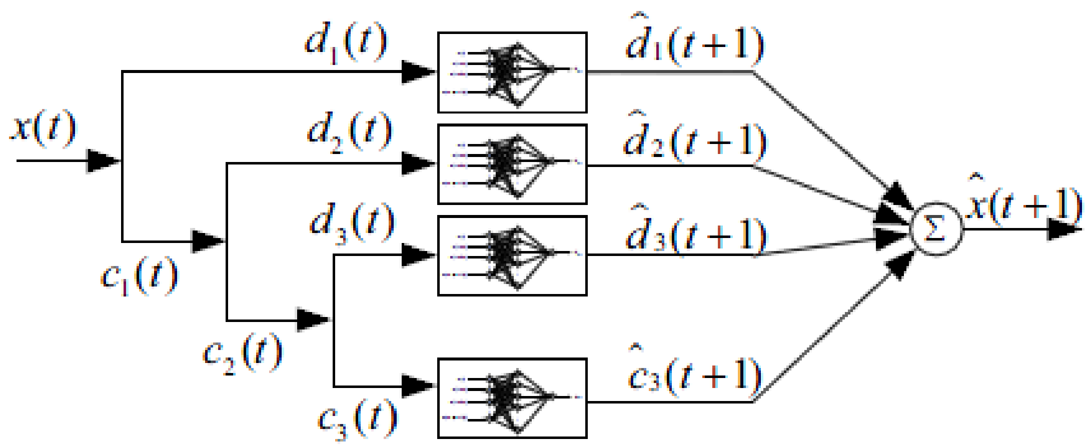


Forecast hour	MAPE
1	8.1130
2	10.6029
3	11.9206
4	12.7485
5	13.2603
6	13.8258
7	14.0576
8	14.2544
9	14.2845
10	14.2799
11	14.2896
12	14.2922
13	14.2737
14	14.3819
15	14.5195
16	14.4968
17	14.5853
18	14.7405
19	14.9456
20	15.1473
21	15.3574
22	15.5360
23	15.7343
24	15.9714
Average	**13.9841**

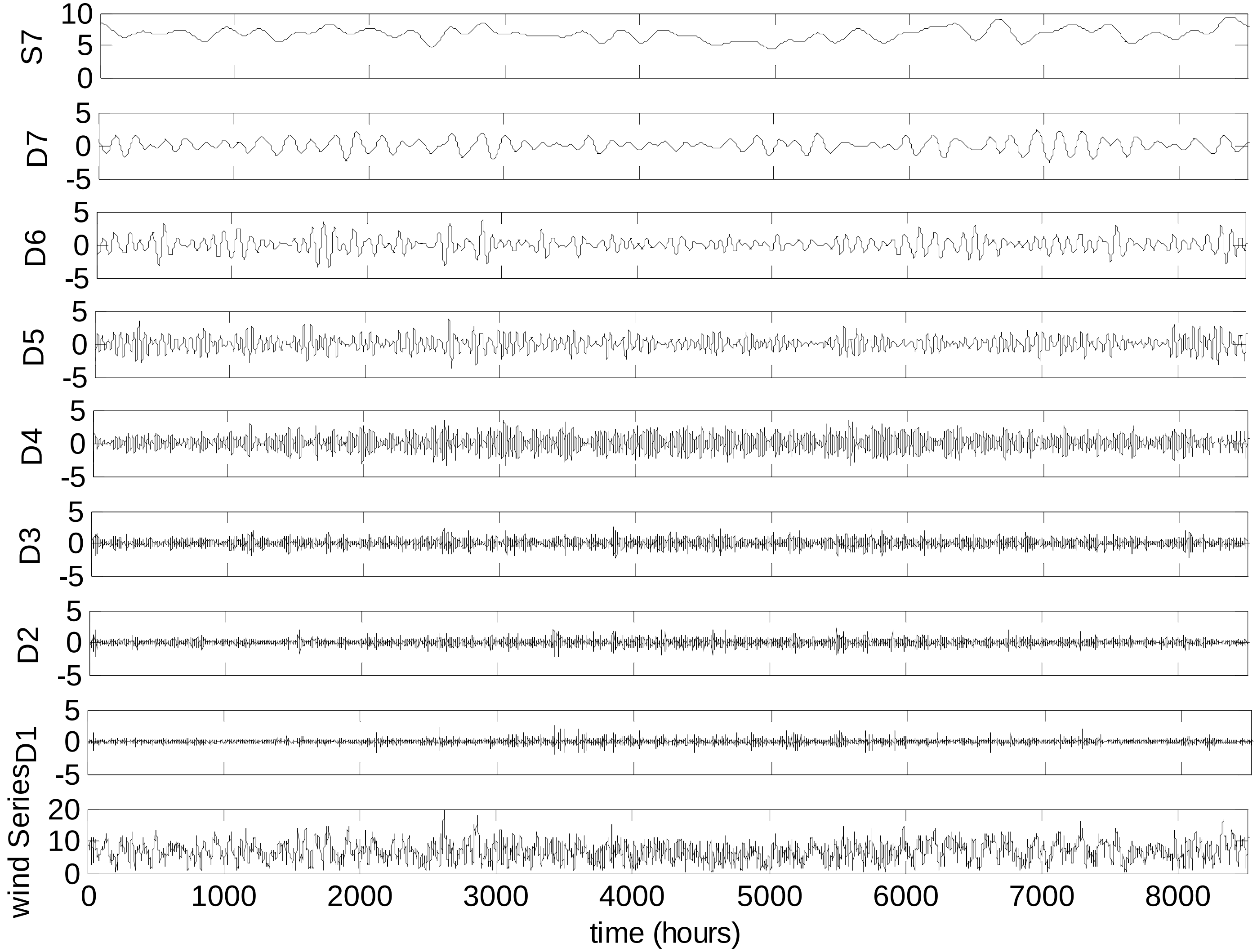
Wind Speed Time Series



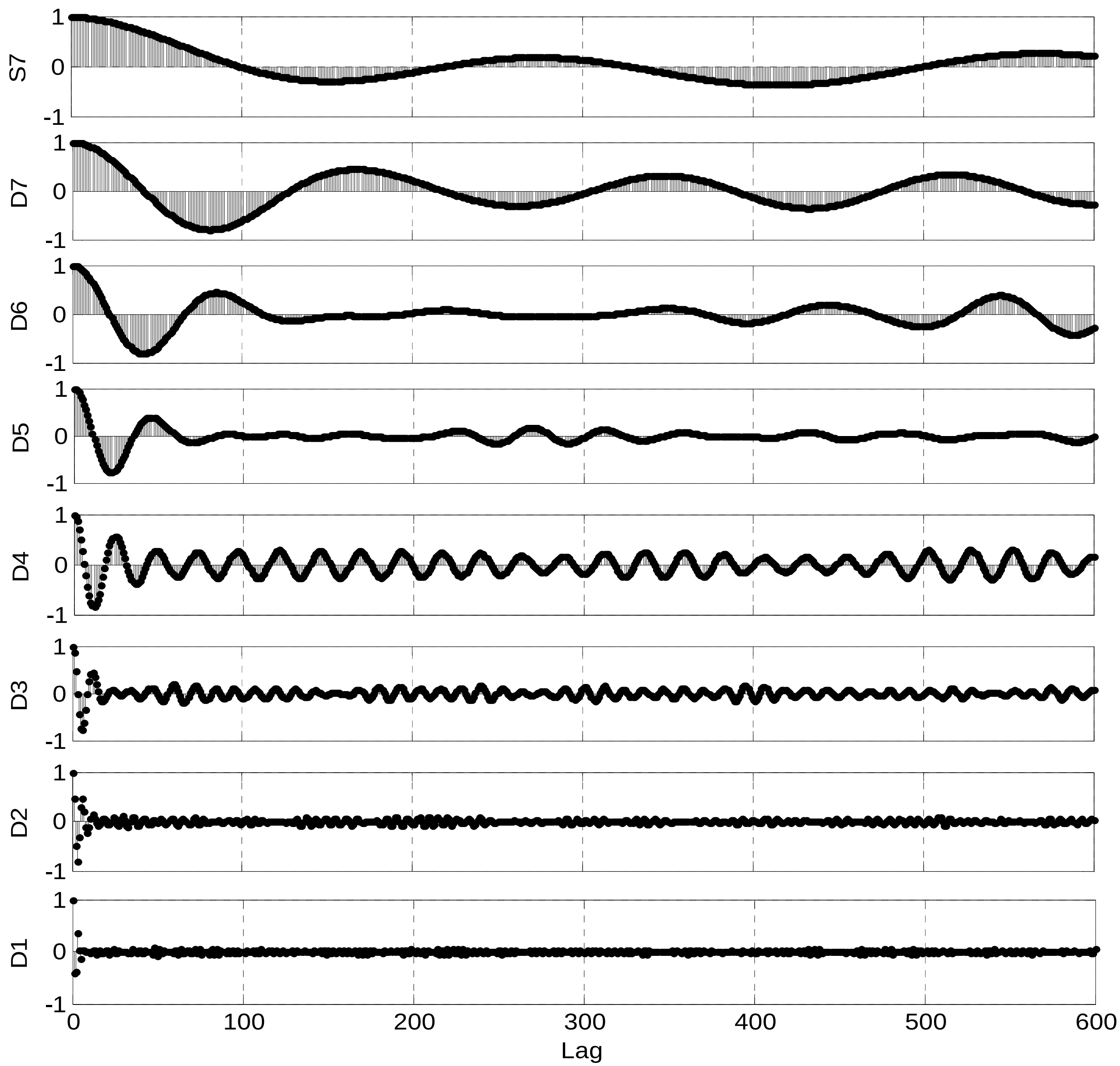
Schematic Block Diagram for Wind Speed Forecasting



Multiresolution Analysis of Wind Speed Time Series



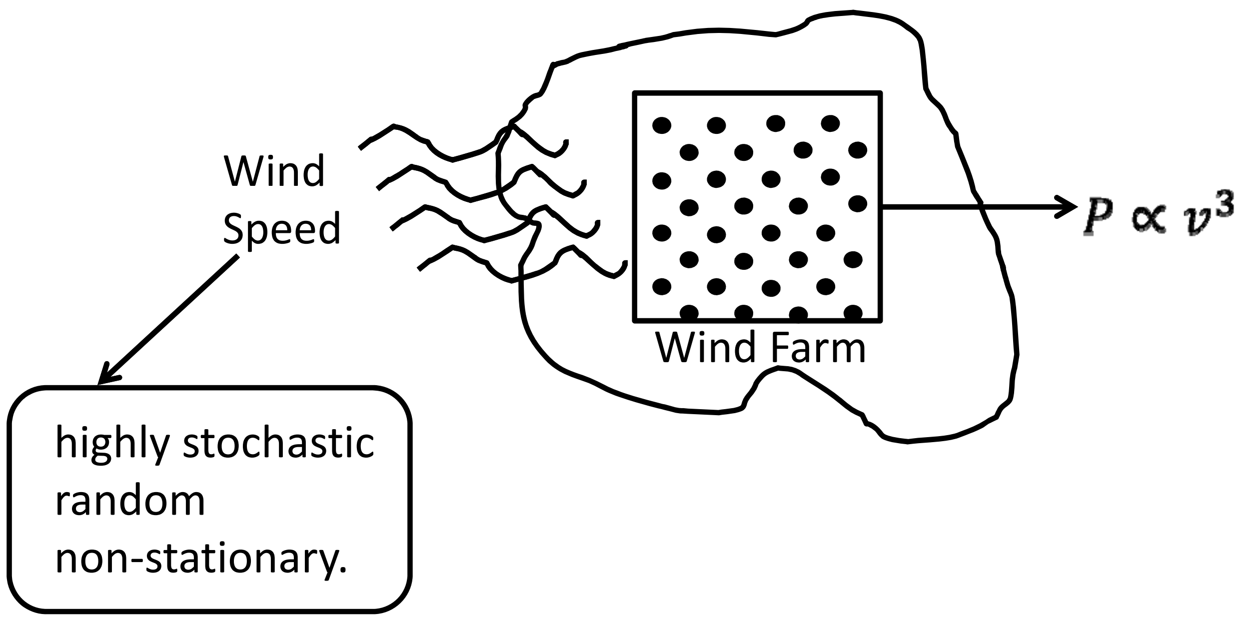
ACF’s of Decomposed Wind Speed Time Series

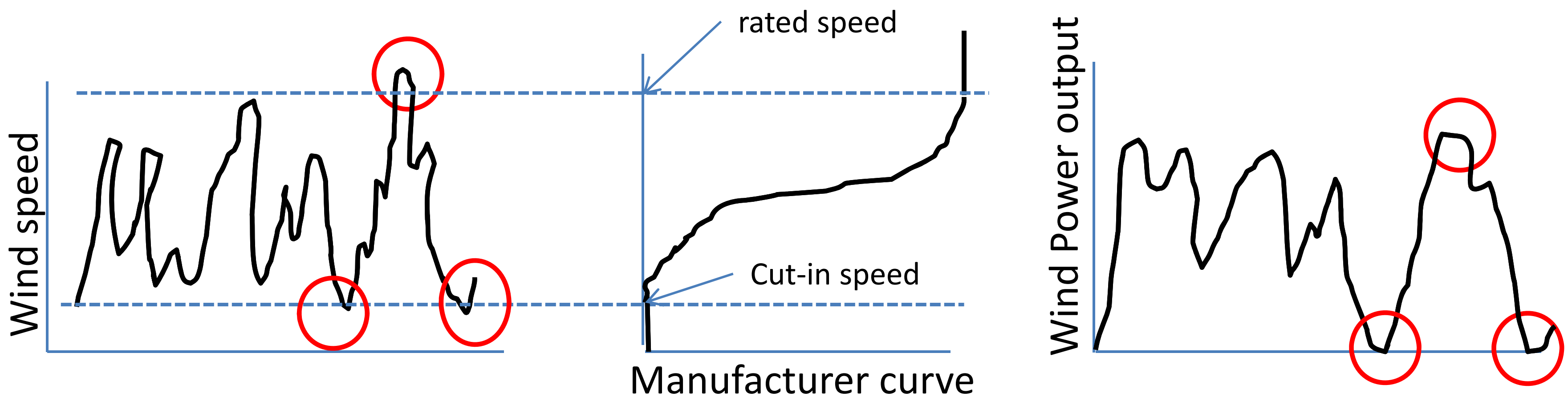


Network Architecture and Input Lag Hours

Decomposed Signal	Input Lag-hours	Network Architecture	
		AWNN	FFNN
S7	1-14,157-159,285-287	20-2-1	20-3-1
D7	1-12,76-83,167-169	19-2-1	19-3-1
D6	1-10,41-44,84-86	17-2-1	17-3-1
D5	1-6,21-23,44-47	13-2-1	13-3-1
D4	1-3,11-13,23-25,48,72	11-2-1	11-3-1
D3	1,2,5,6,12,60,72	7-2-1	7-3-1
D2	3,6,9,15	4-2-1	4-3-1
D1	1,2,5,22	4-2-1	4-3-1

Wind Power Forecast





Wind Power Forecasting: Time Horizon

Very Short-Term Forecasting : Up to 2-3 h in steps of 10min. or 15min.

-Turbine control

-Real time participation in electricity market.

Short-Term Forecasting : Up to 24h in steps of 1h.

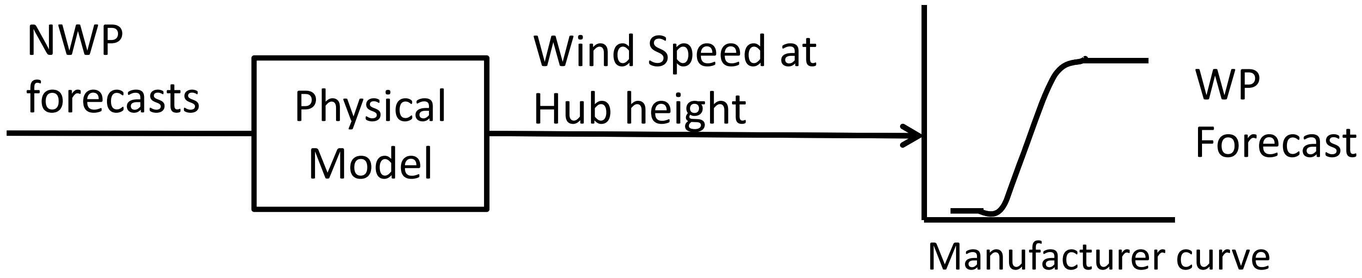
- Hour ahead bidding
- Intraday market
- Day-ahead market
- Unit commitment and Economic dispatch
- Ancillary services management
- Day-ahead reserve setting

Medium Term Forecasting (With NWP inputs): up to 72h in steps of 1h.

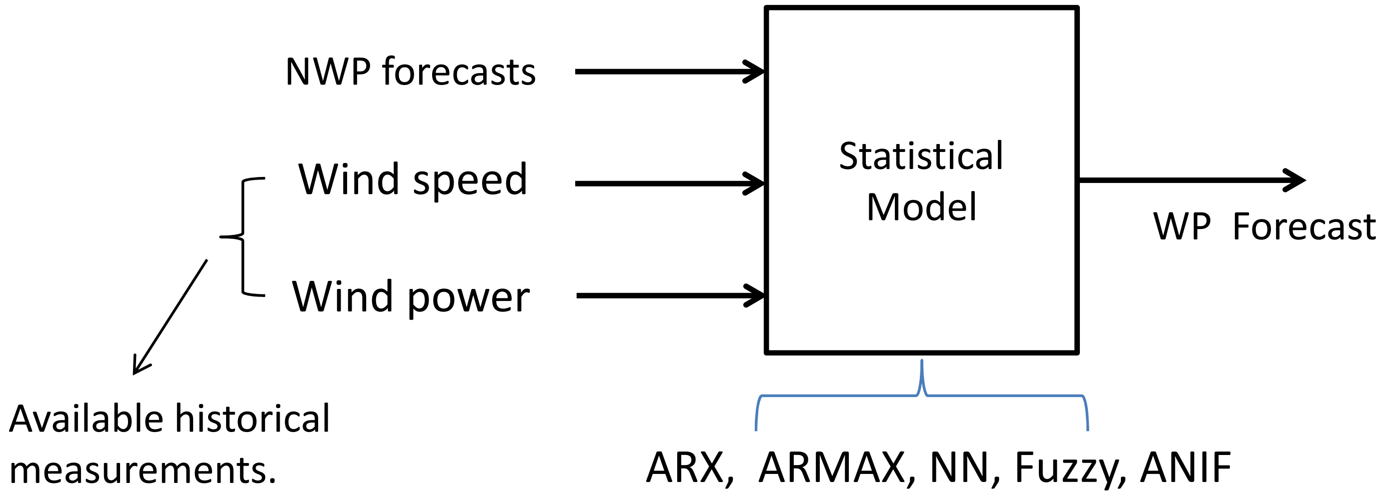
- In addition to the above mentioned benefits
- Maintenance planning of wind farms
- Wind farm and storage device Coordination
- Congestion management
- Maintenance planning of network lines.

Wind Power Forecasting: Approaches

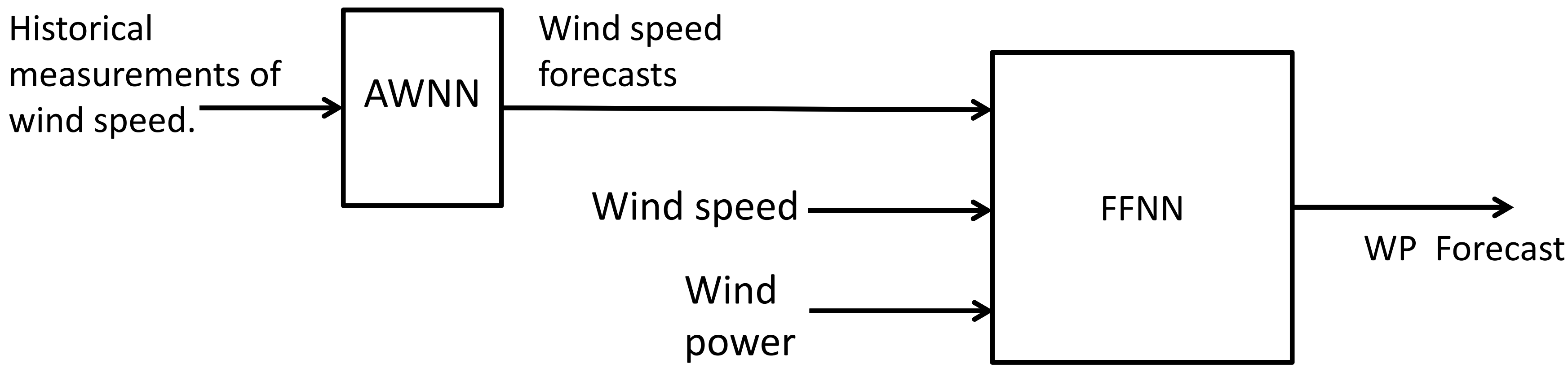
1)



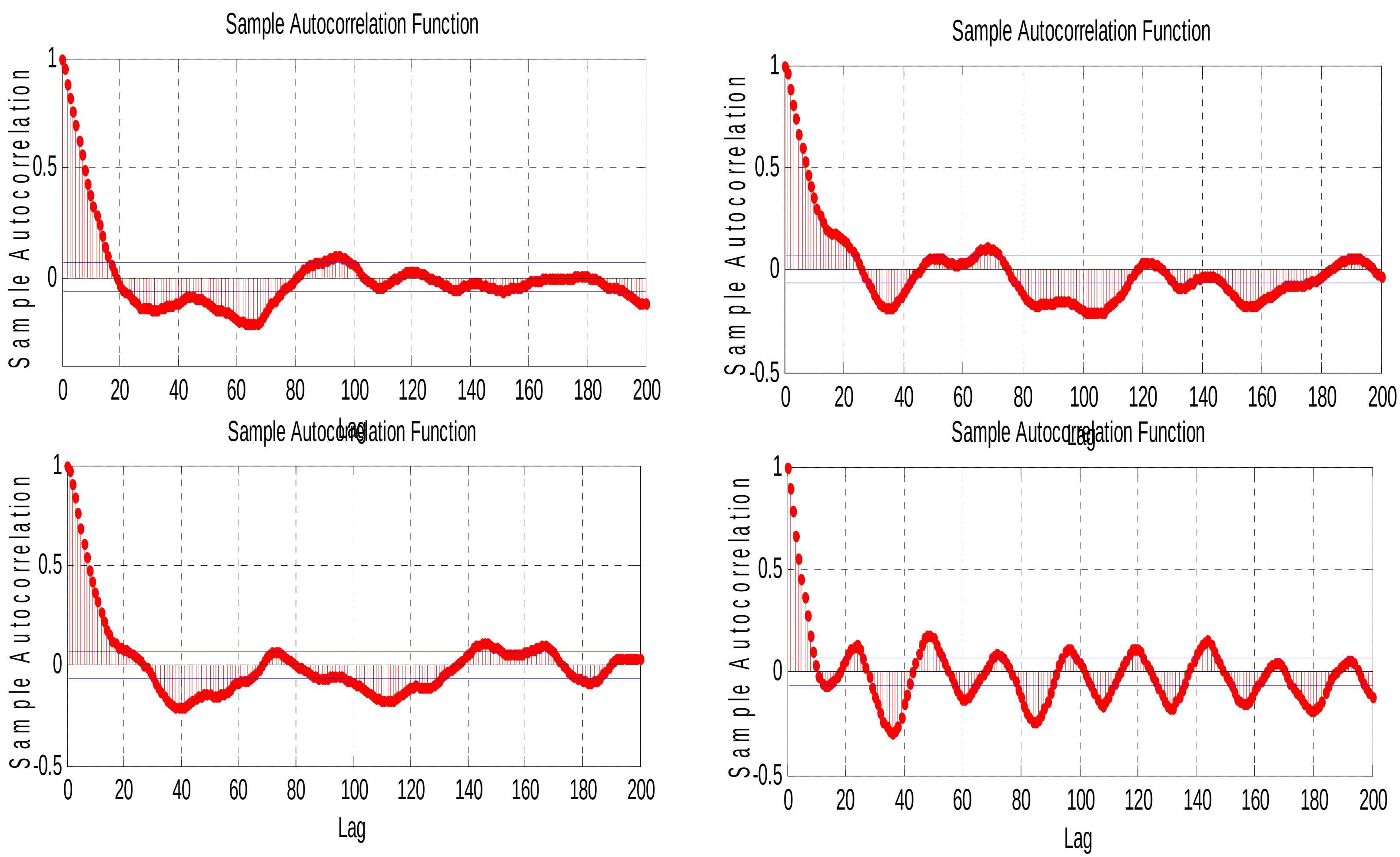
2)



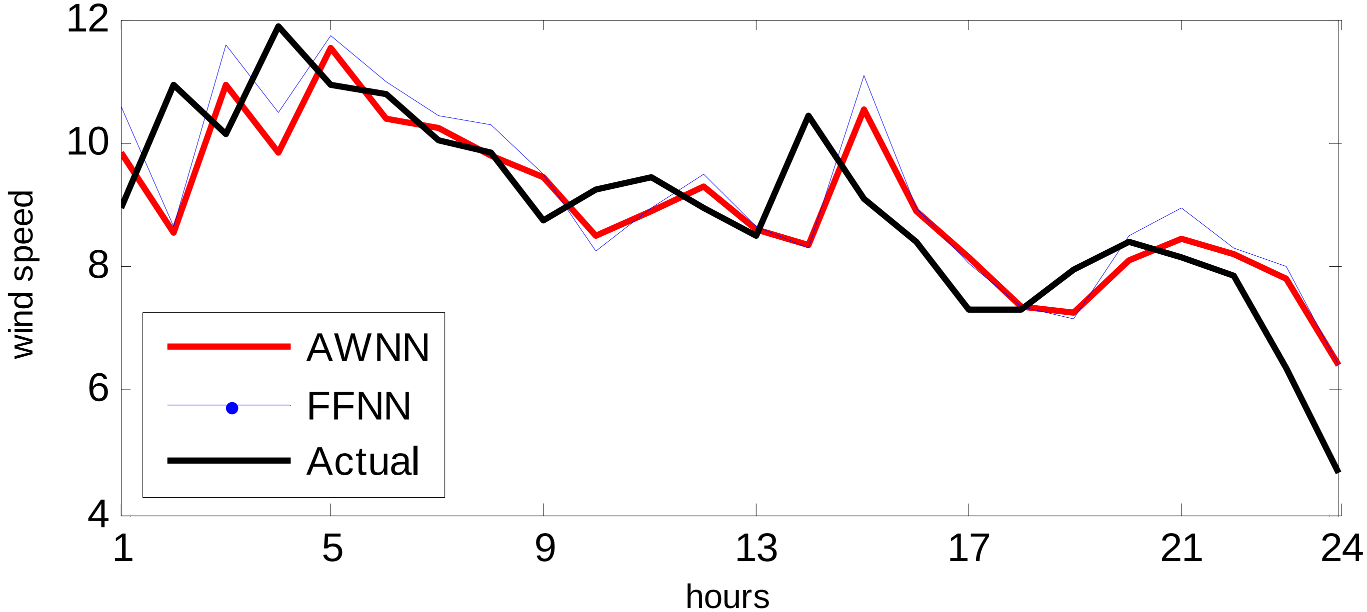
A Two stage approach for Wind Power Forecast



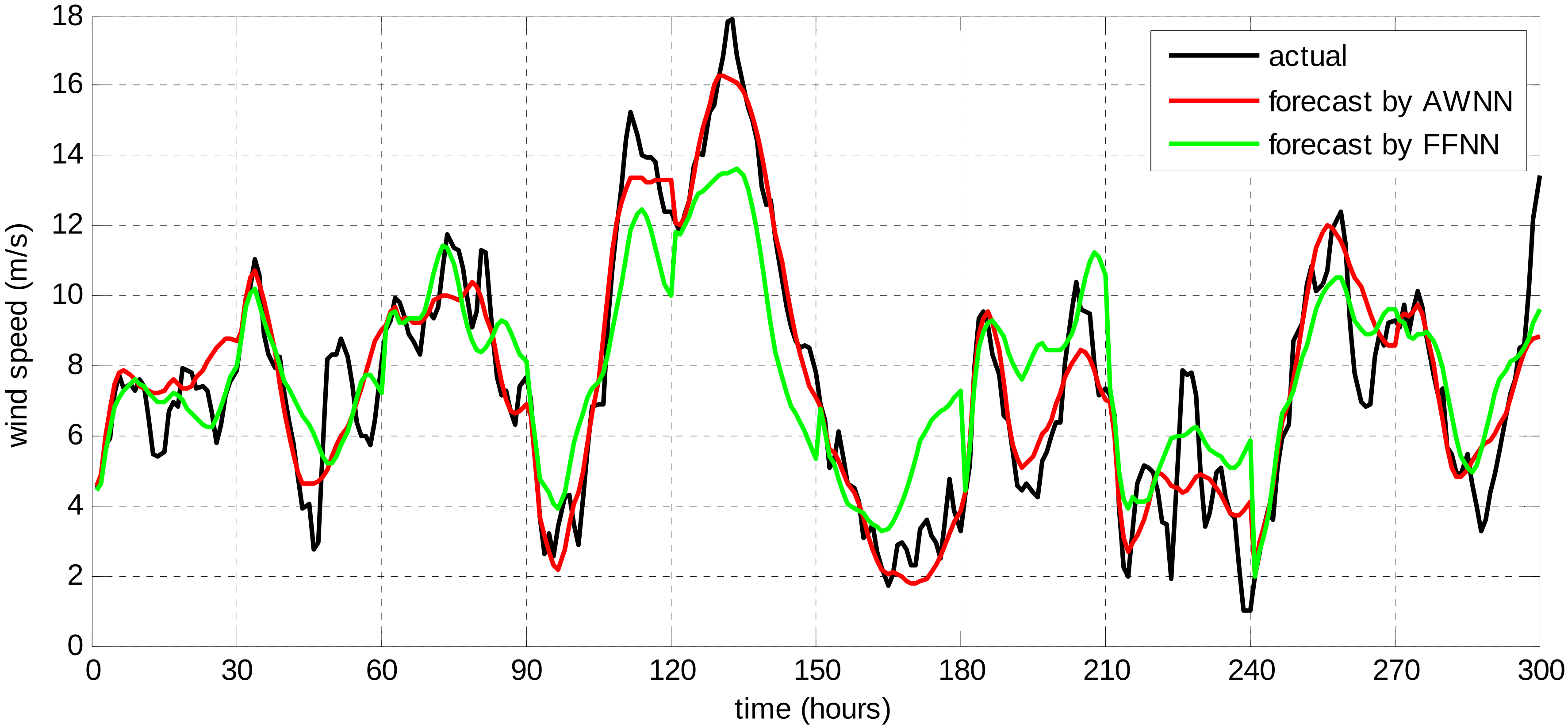
Autocorrelation Analysis of Wind Series



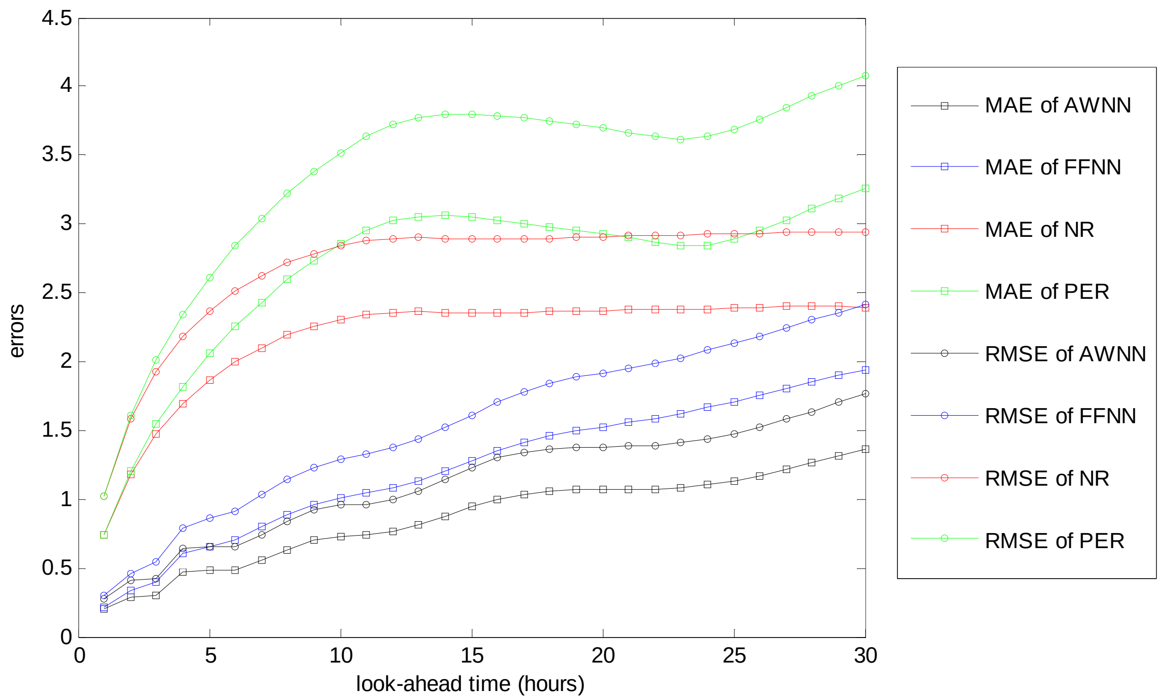
Hour Ahead Forecast of Wind Speed



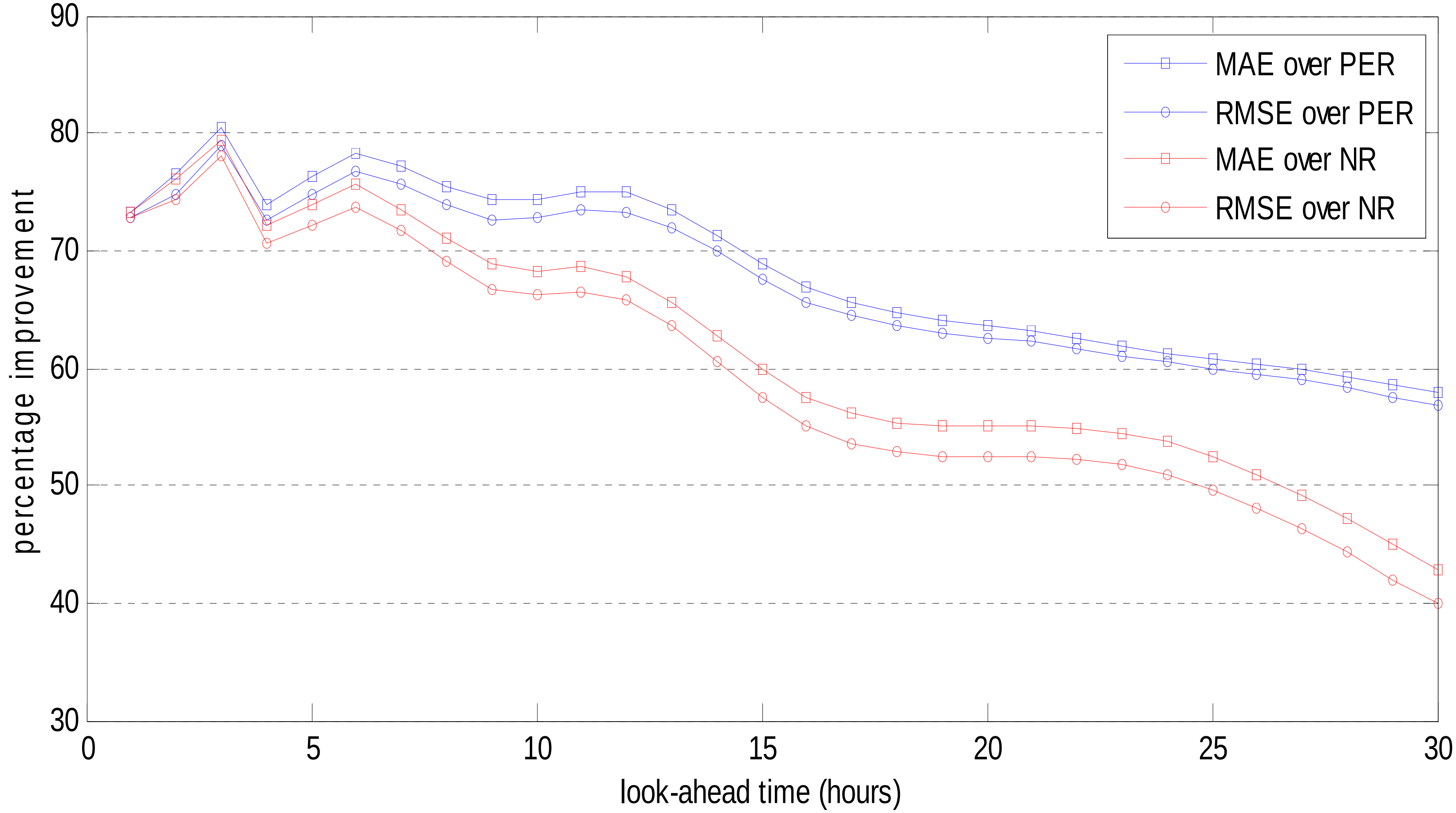
30-hours ahead Wind Speed Forecast



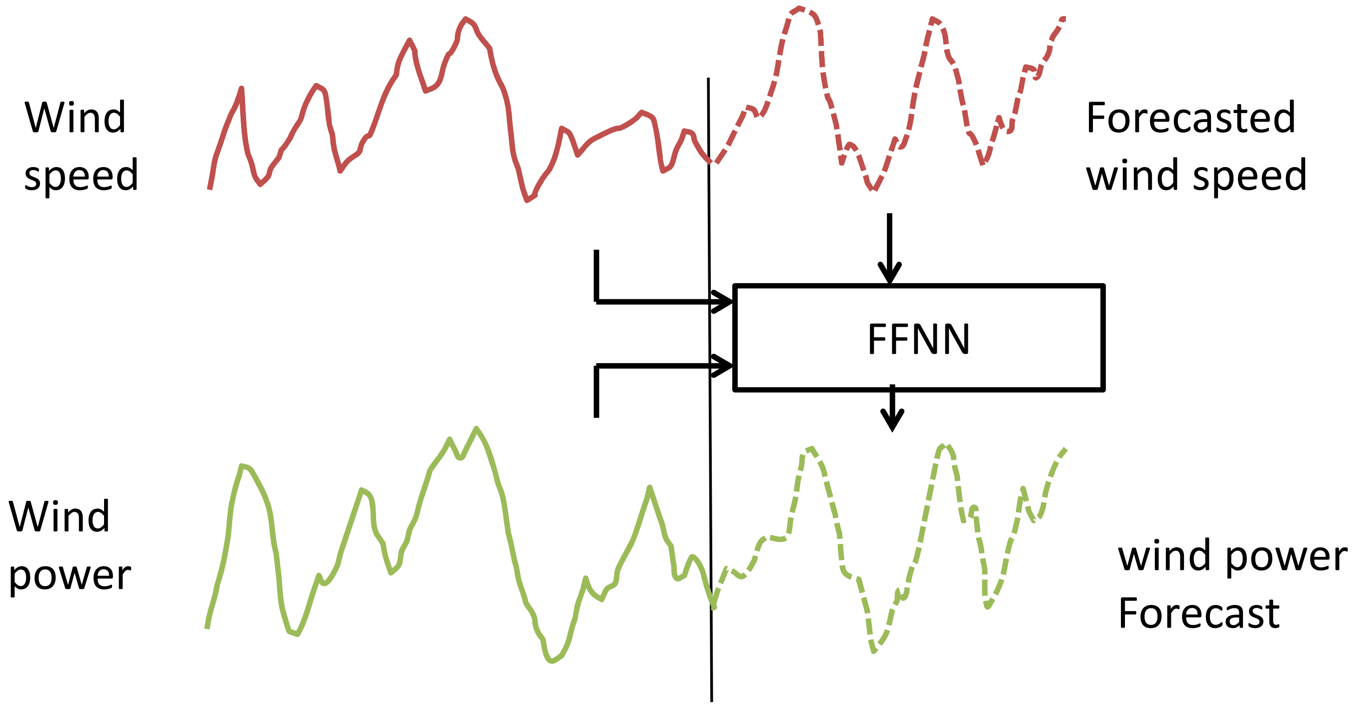
Comparative Performance



Percentage Improvement



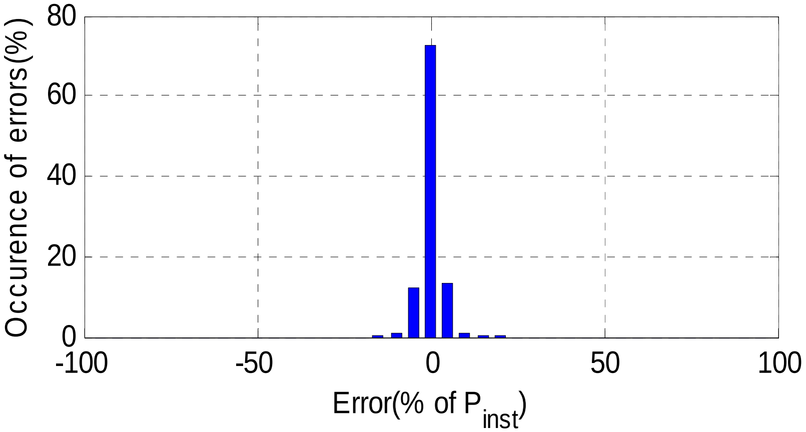
Wind speed to Wind Power Transformation



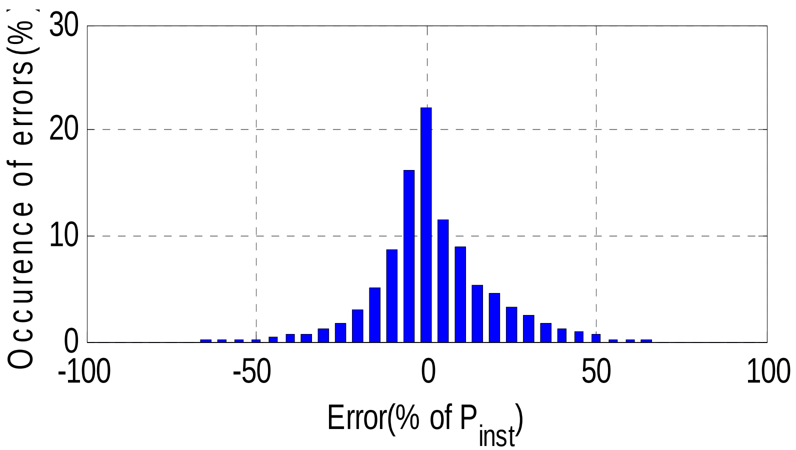
FFNN Inputs:

wind speed {0, 1, 2} lag hours and from
wind power series {1, 2, 3, 4, 5, 6} lag hours.

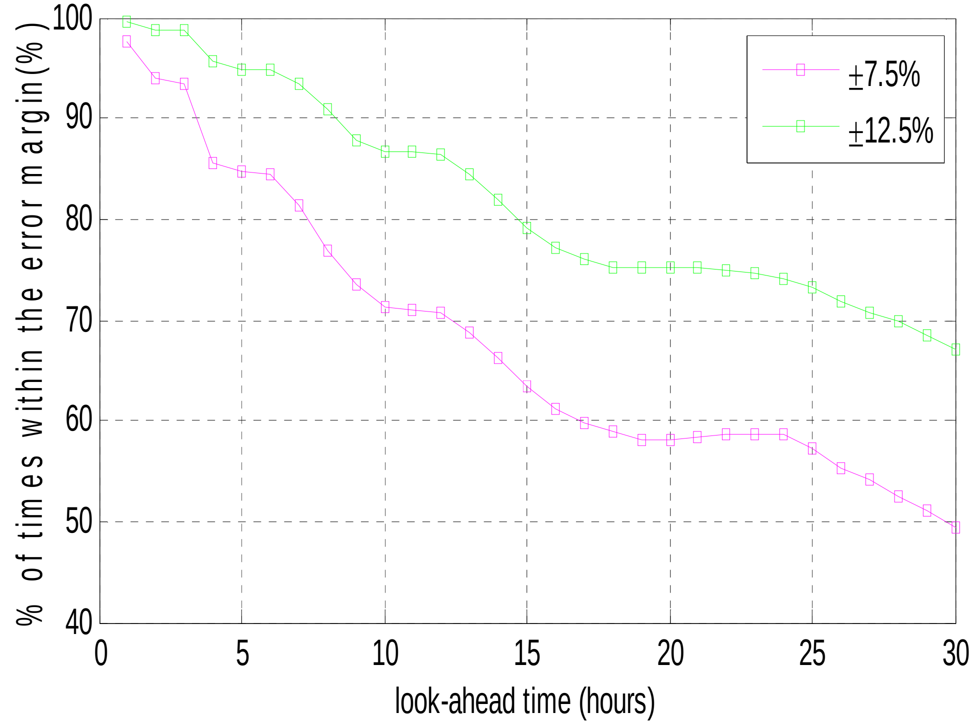
Error Distributions and Forecasting Ability



1-hr ahead forecast error distributions



30th -hr ahead forecast error distributions



Thank You