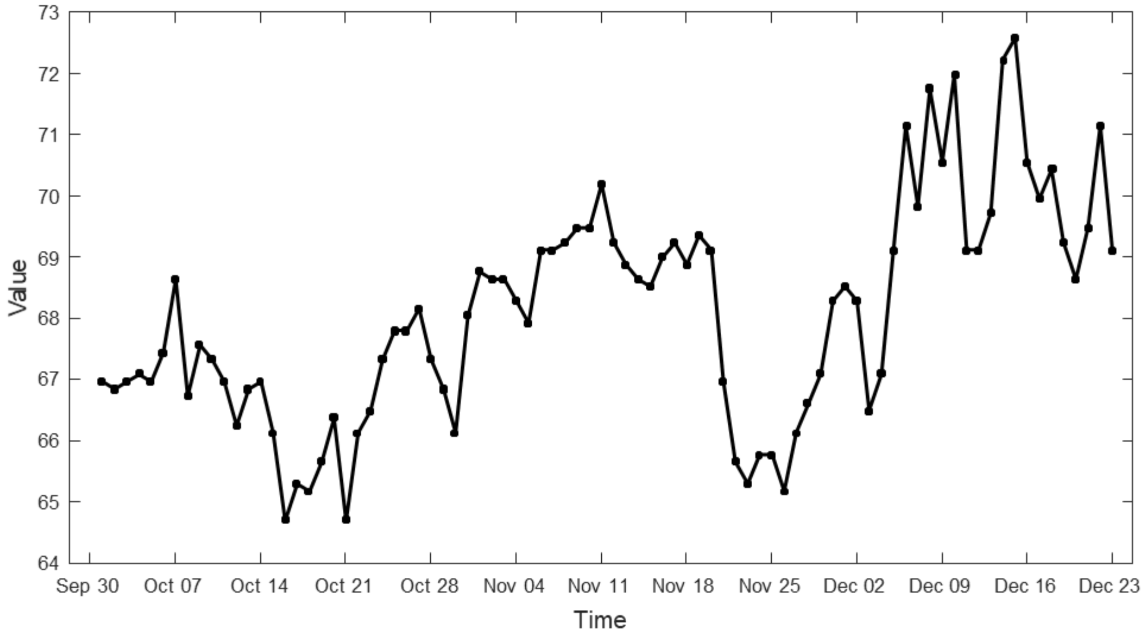
Lecture 6
Time Series Analysis

Phayung Meesad, Ph.D.
King Mongkut's University of Technology
North Bangkok (KMUTNB)
Bangkok Thailand

Outlines

- What is time series?
- Application of time series
- Components of time series
- Stationarity
- Time series prediction

What is time series



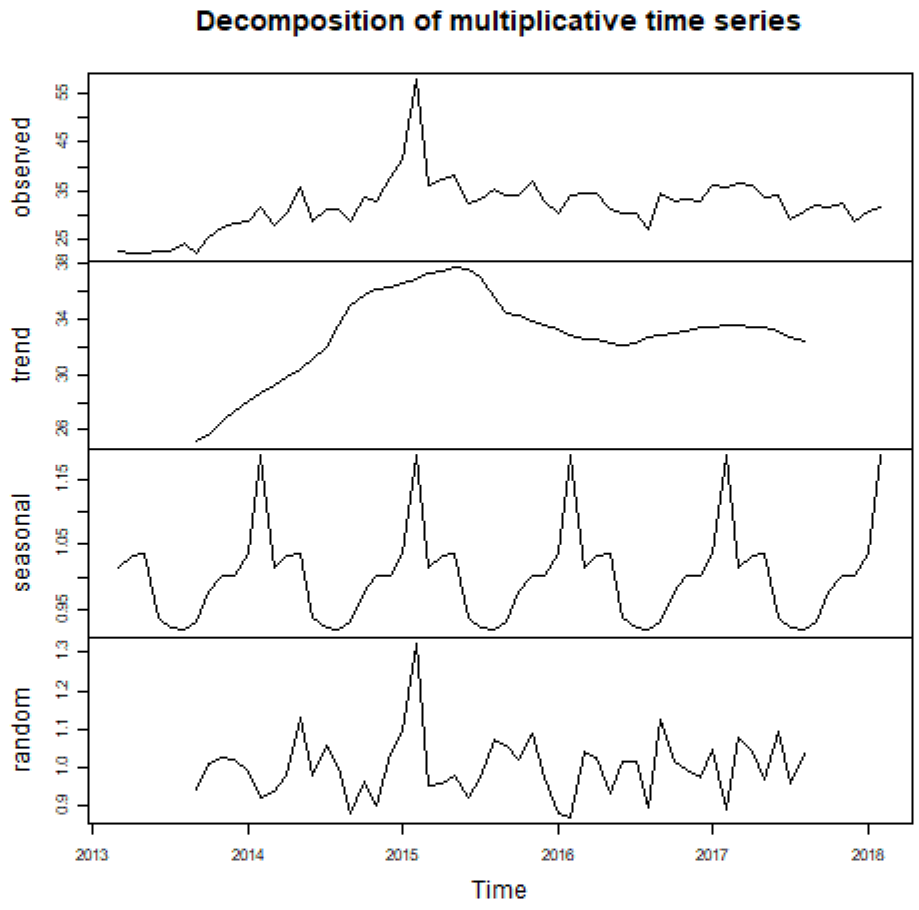
- Time Series is a sequence of measured values related to time.
- Time series data play very important roles in large domain of applications.

Time Series Applications

- The applications of time series data appear in many fields: science, engineering, business, finance, etc.
- Scientific workflow systems, scientific exploitation of operational missions, and fluid-structure simulation.
- Road traffic forecasting, wind power prediction, and quality control.
- Economic forecasting, sales forecasting, budgetary analysis, and stock price prediction.

Components of Time Series

- Trend
- Seasonality
- Irregularity
- Cyclic



Analysis of Time Series

- To identify the magnitude and direction of trends
- To estimate the effect of seasonal and cyclical variations
- To estimate the size of the residual component.
- This implies the decomposition of a time series into its several components.

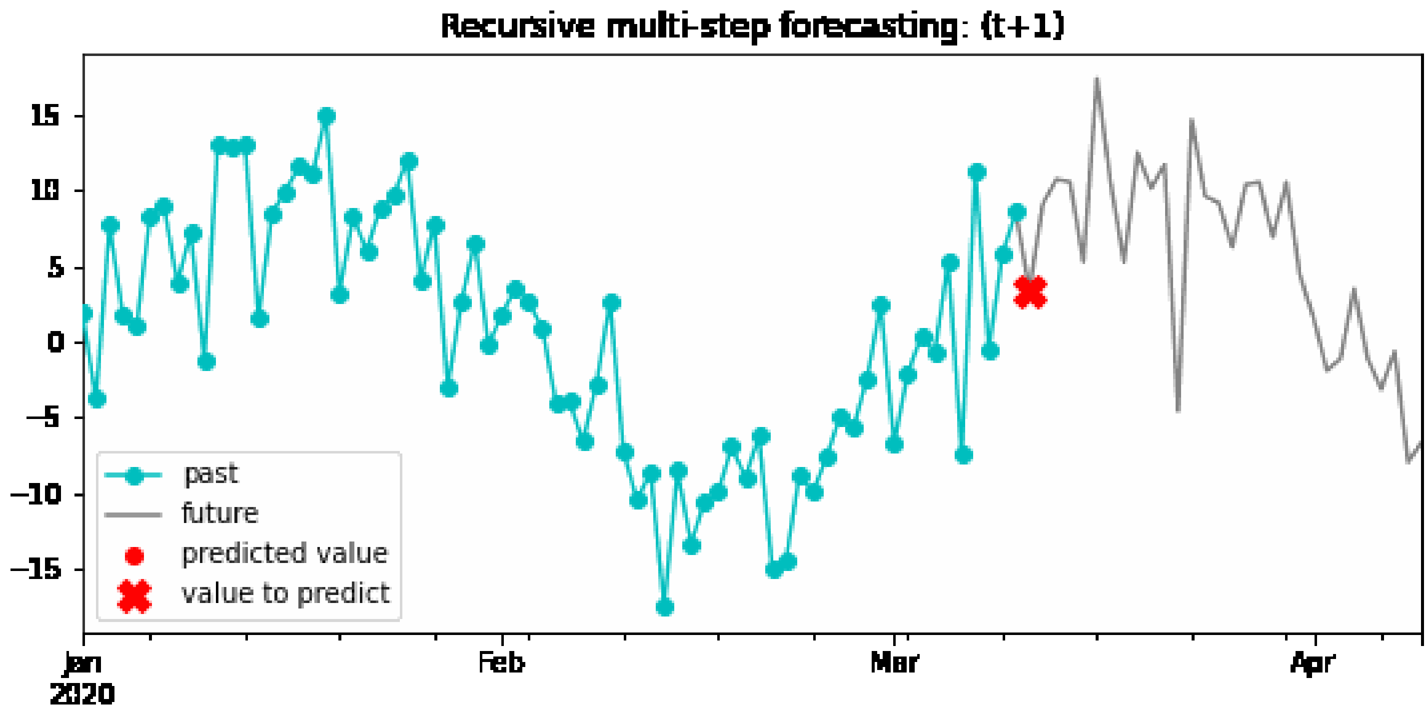
Time Series Models

- Two approaches are usually adopted in analyzing a given time series:

- Additive model:
  $Y(t) = T(t) + S(t) + C(t) + R(t)$

- Multiplicative model:
  $Y(t)=T(t) \times S(t) \times C(t) \times R(t)$

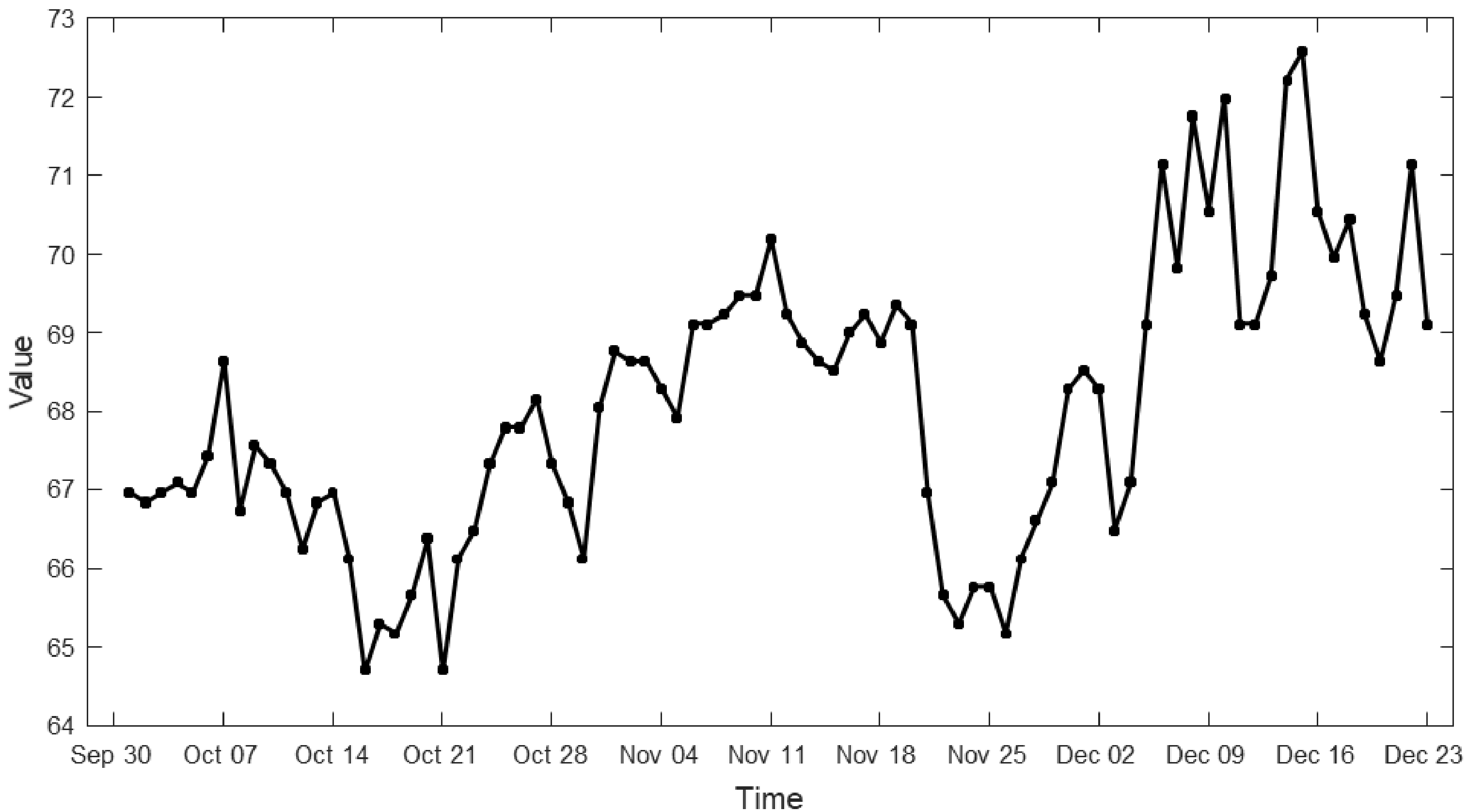
Time Series Forcasting



Time Series Trransformation

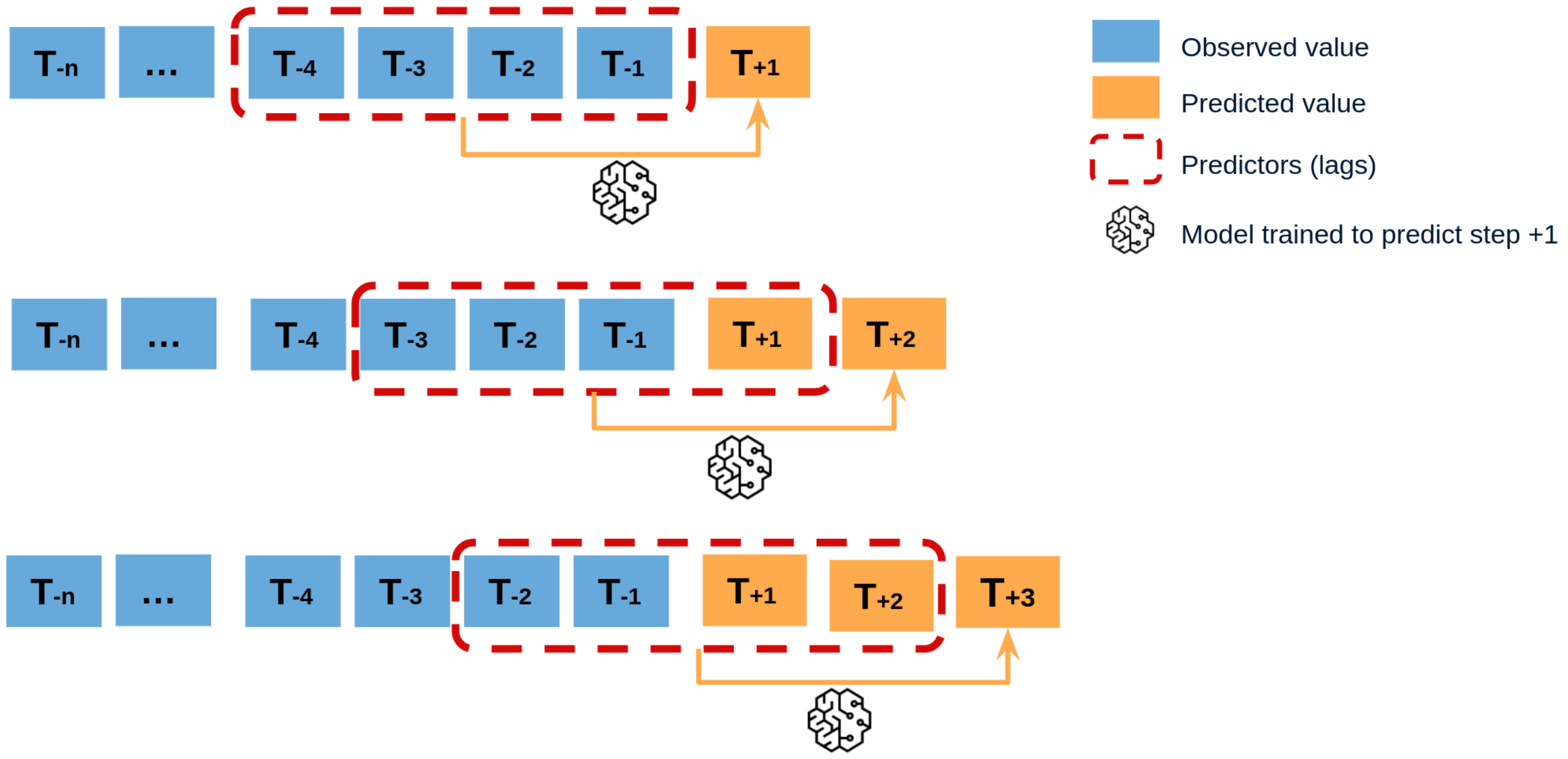
1	2	3	4	5	6	7	8	9	10

Mathematical Representation

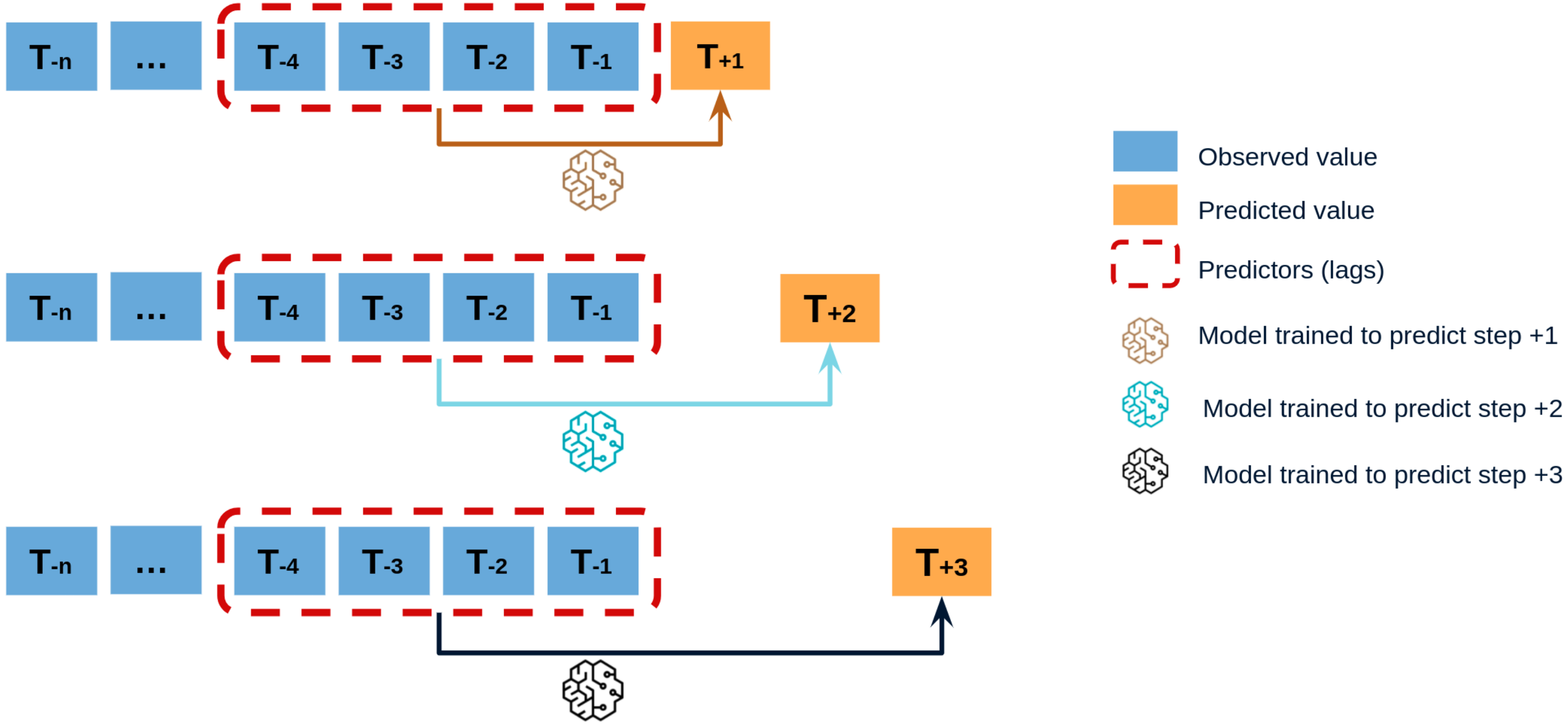


$$Y(t) = \{y(t-\infty), \ldots, y(t-1), y(t), y(t+1) \ldots, y(t+\infty)\}$$

Recursive multi-step forecasting



Direct multi-step forecasting



Direct multi-step forecasting

y	1	2	3	4	5	6	7	8	9
exog 1	A	B	C	D	E	F	G	H	I
exog 2	a	b	c	d	e	f	g	h	I

`lags=3, steps=2, create_train_X_y()`

y_train	
step 1	step 2
4	5
5	6
6	7
7	8
8	9

X_train						
X_lags			X_exog_1		X_exog_2	
lag 1	lag 2	lag 3	step 1	step 2	step 1	step 2
3	2	1	D	E	d	e
4	3	2	E	F	e	f
5	4	3	F	G	f	g
6	5	4	G	H	g	h
7	6	5	H	I	h	i

`filter_train_X_y_for_step(step=1)`

`filter_train_X_y_for_step(step=2)`

y_train
step 1
4
5
6
7
8

X_train				
X_lags			X_exog_1	X_exog_2
lag 1	lag 2	lag 3	step 1	step 1
3	2	1	D	d
4	3	2	E	e
5	4	3	F	f
6	5	4	G	g
7	6	5	H	h

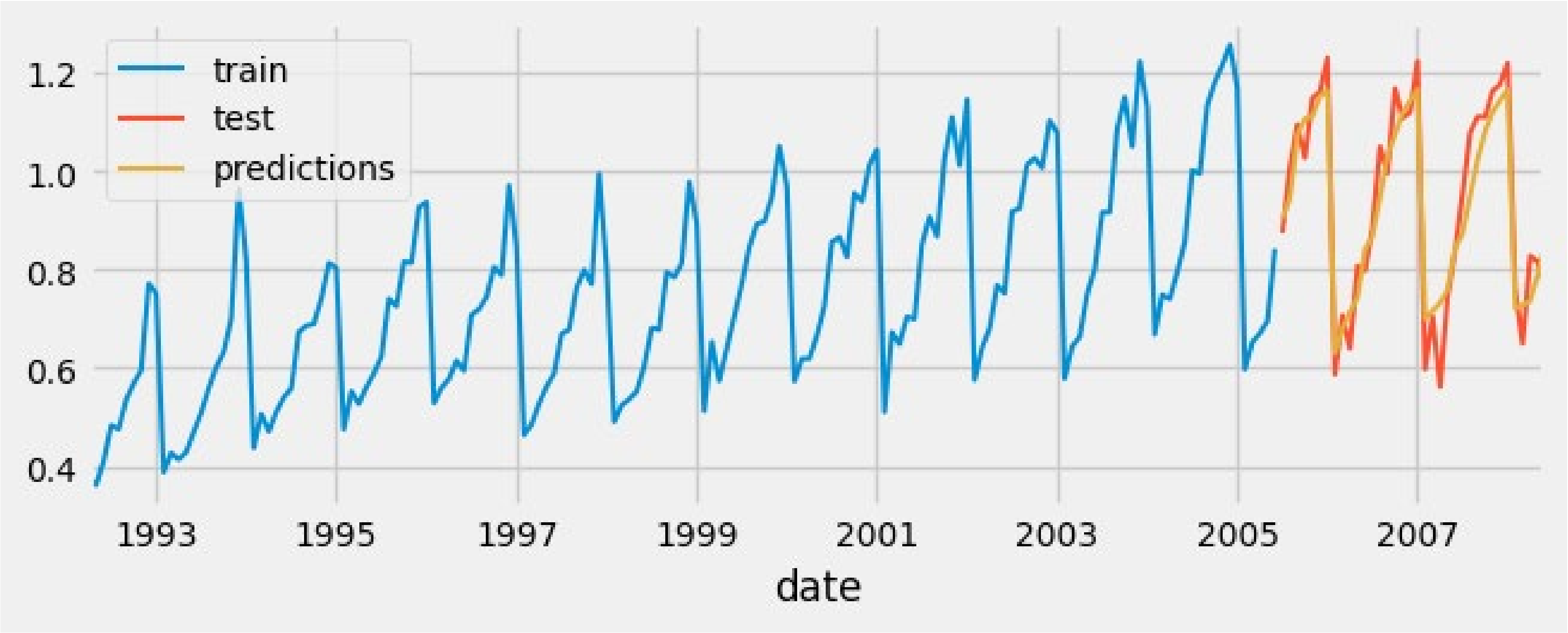
`Training data for model of step 1`

y_train
step 2
5
6
7
8
9

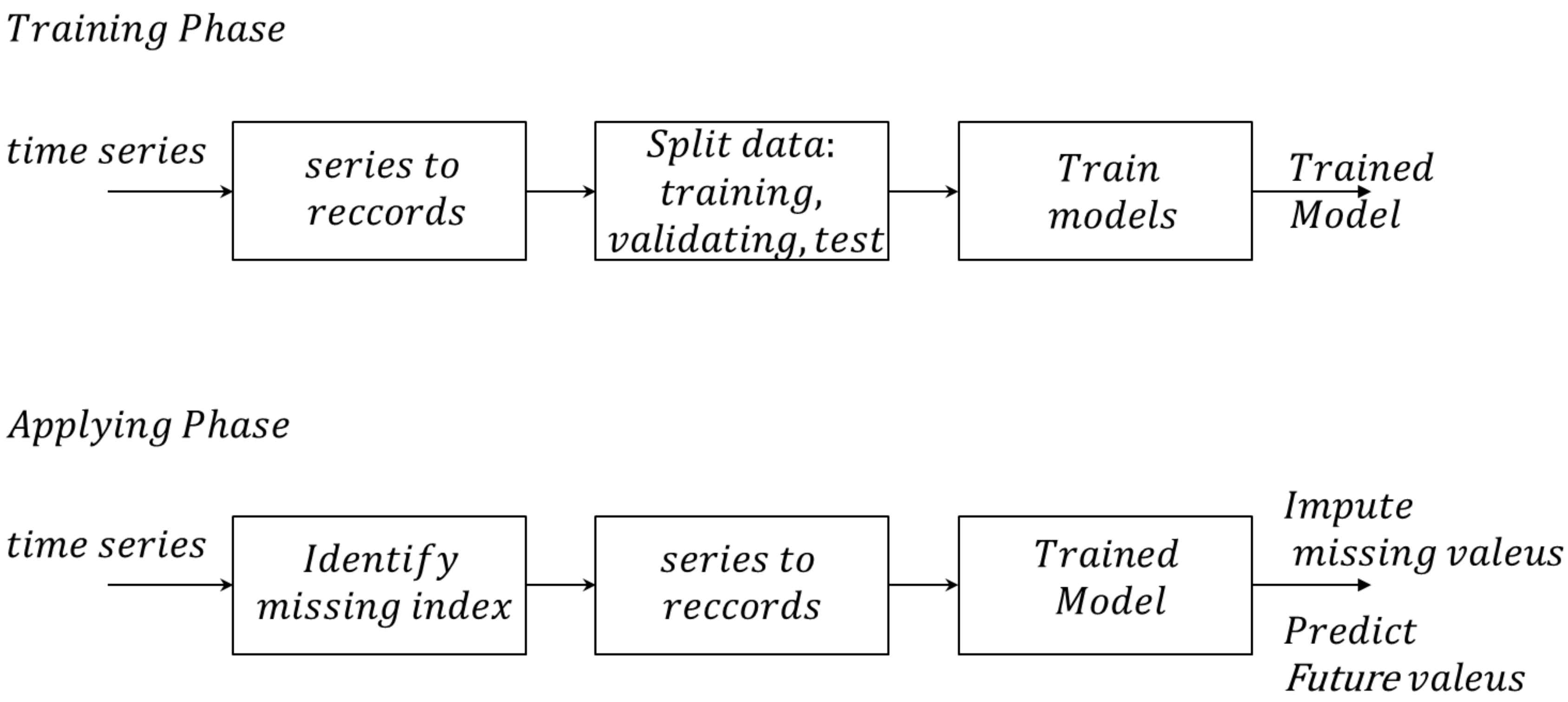
X_train				
X_lags			X_exog_1	X_exog_2
lag 1	lag 2	lag 3	step 2	step 2
3	2	1	E	e
4	3	2	F	f
5	4	3	G	g
6	5	4	H	h
7	6	5	I	i

`Training data for model of step 2`

Train and Test model



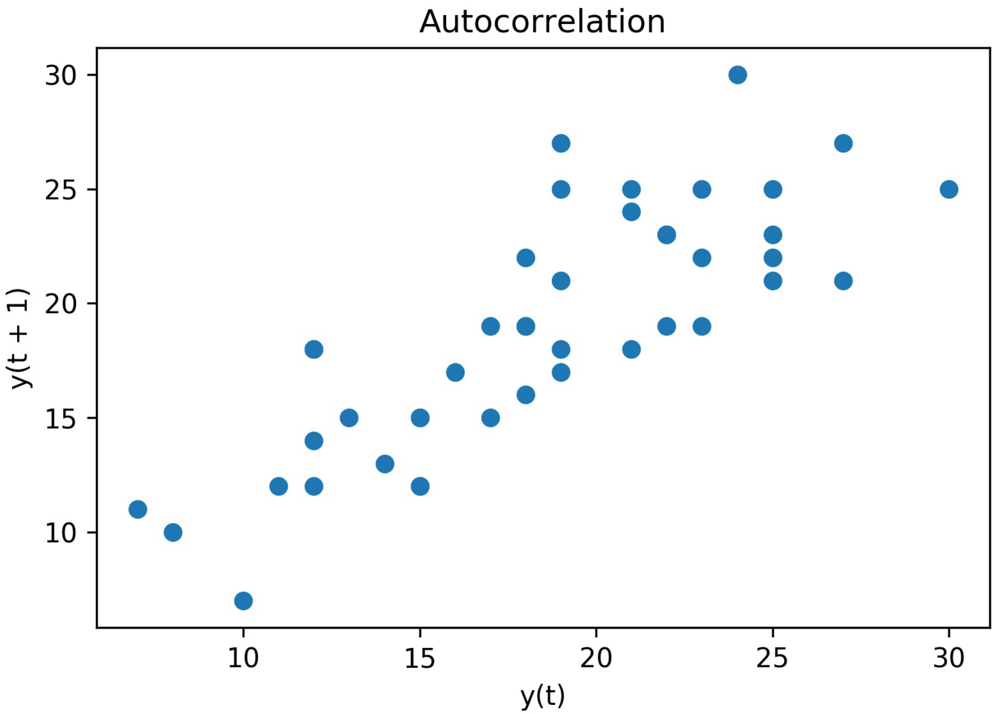
Machine Learning Modelling



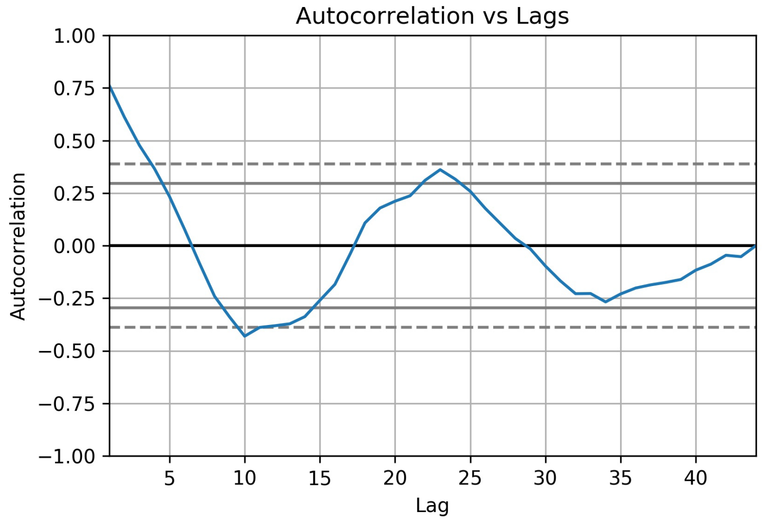
Autocorrelation

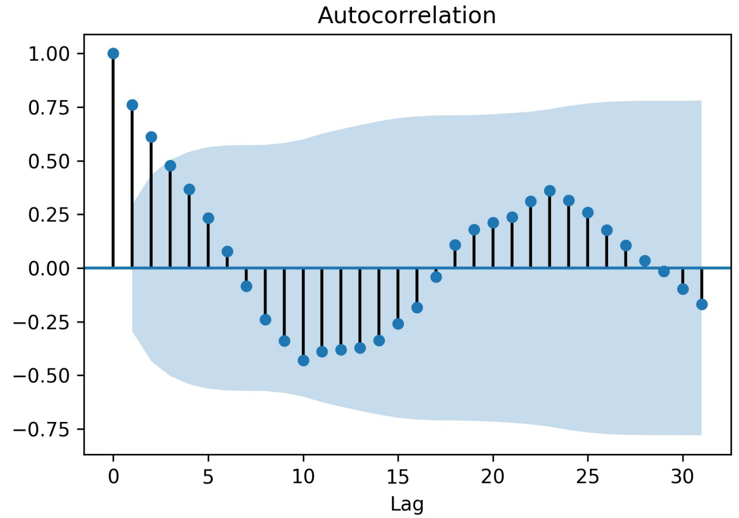
- An autoregression model makes an assumption that the observations at previous time steps are useful to predict the value at the next time step.
- This relationship between variables is called correlation.
- If both variables change in the same direction this is called a positive correlation. If the variables move in opposite directions as values change, then this is called negative correlation.

Autocorrelation



Autocorrelation vs Lags Plot

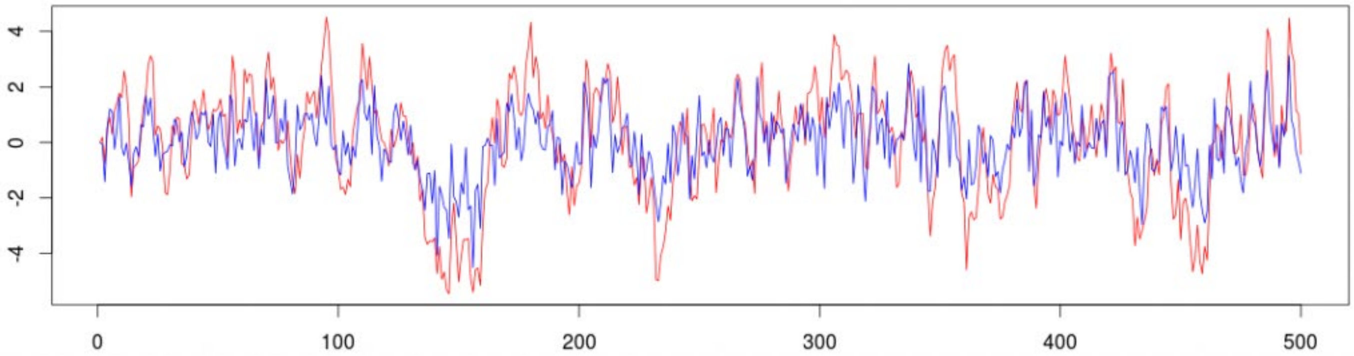




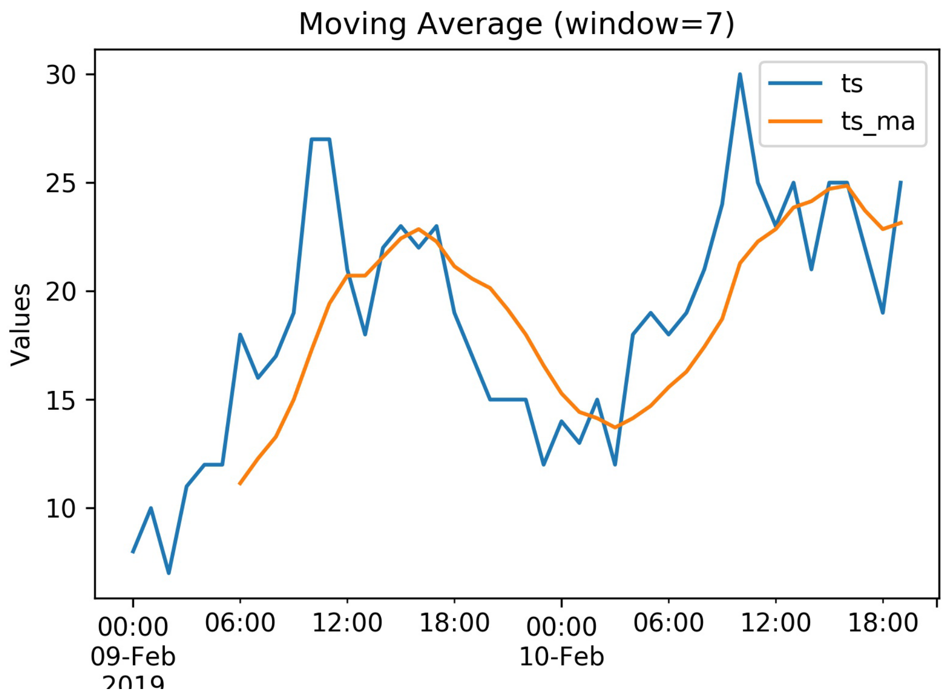
AutoRegression (AR)

- The notation AR(*p*) refers to the autoregressive model of order *p*. The AR(*p*) model is written

$$x_t = c + \sum_{i=1}^{p} \varphi_i x_{t-i} + \varepsilon_t$$



Moving Average (MA)



$MA(q)$

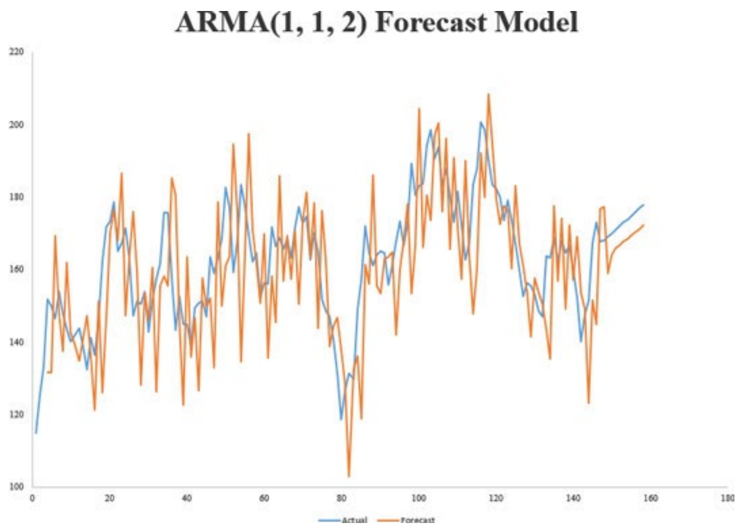
The notation MA(q) refers to the moving average model of order q:

$$x_t = \frac{1}{q}\sum_{i=t-q}^{q} x_i + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

Auto Regression Moving Average

The notation ARMA(*p*, *q*) refers to the model with *p* autoregressive terms and *q* moving-average terms. This model contains the AR(*p*) and MA(*q*) models,

$$x_t = \frac{1}{q}\sum_{i=t-q}^{q} x_i + \sum_{i=1}^{p} \varphi_i x_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$$



ARIMA

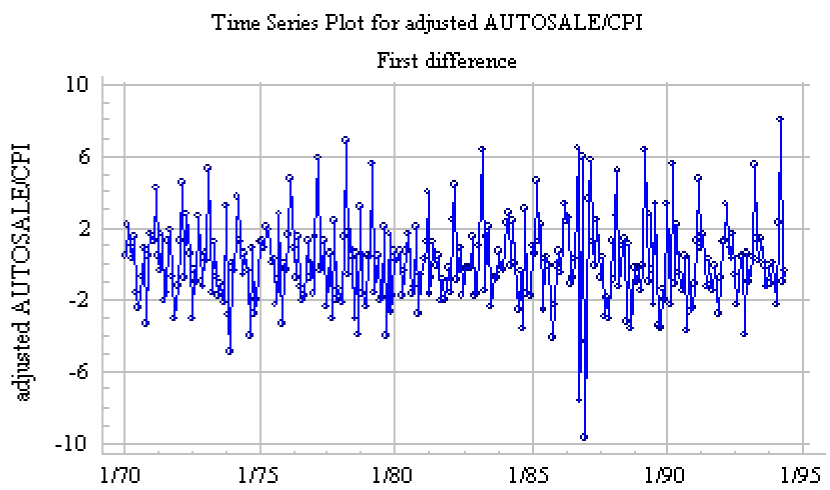
- An autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model.
- ARIMA models are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied one or more times to eliminate the non-stationarity.

$$x_t = \underbrace{\mu + \varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p}}_{\text{AR}} \underbrace{- \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}}_{\text{MA}}$$

Differencing

- Differencing is a transformation applied to time-series data in order to make it stationary.
- A stationary time series' properties do not depend on the time at which the series is observed.
- The difference between consecutive observations is computed as

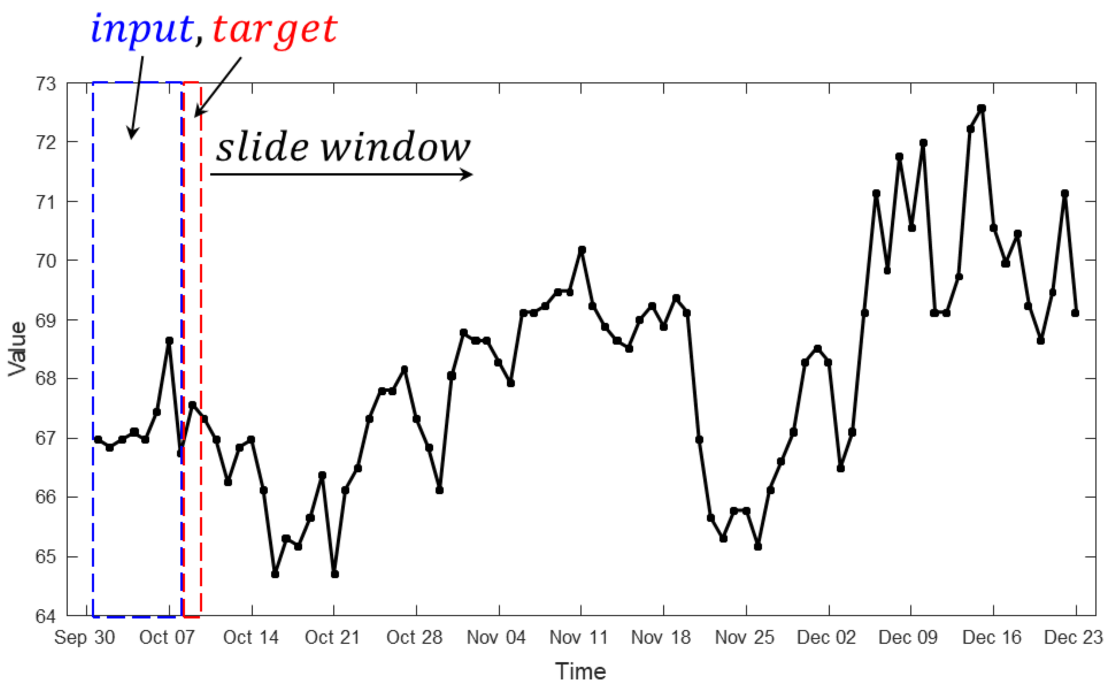
$$\hat{x}_t = x_t - x_{t-1}$$



Characteristics of Time Series

- Time series data have **seasonal recurrence values**.
- A time series may have **trend**, **seasonal effect**, and remaining variability assumptions: **stationarity**, **uncorrelated random error**, and **no outliers**.
- Since the time series repeats itself, the future values can be predicted by using the **historical values of the same series**.
- The **cross-related multivariate time series** can be influenced in the prediction of other time series.

Data Manipulation



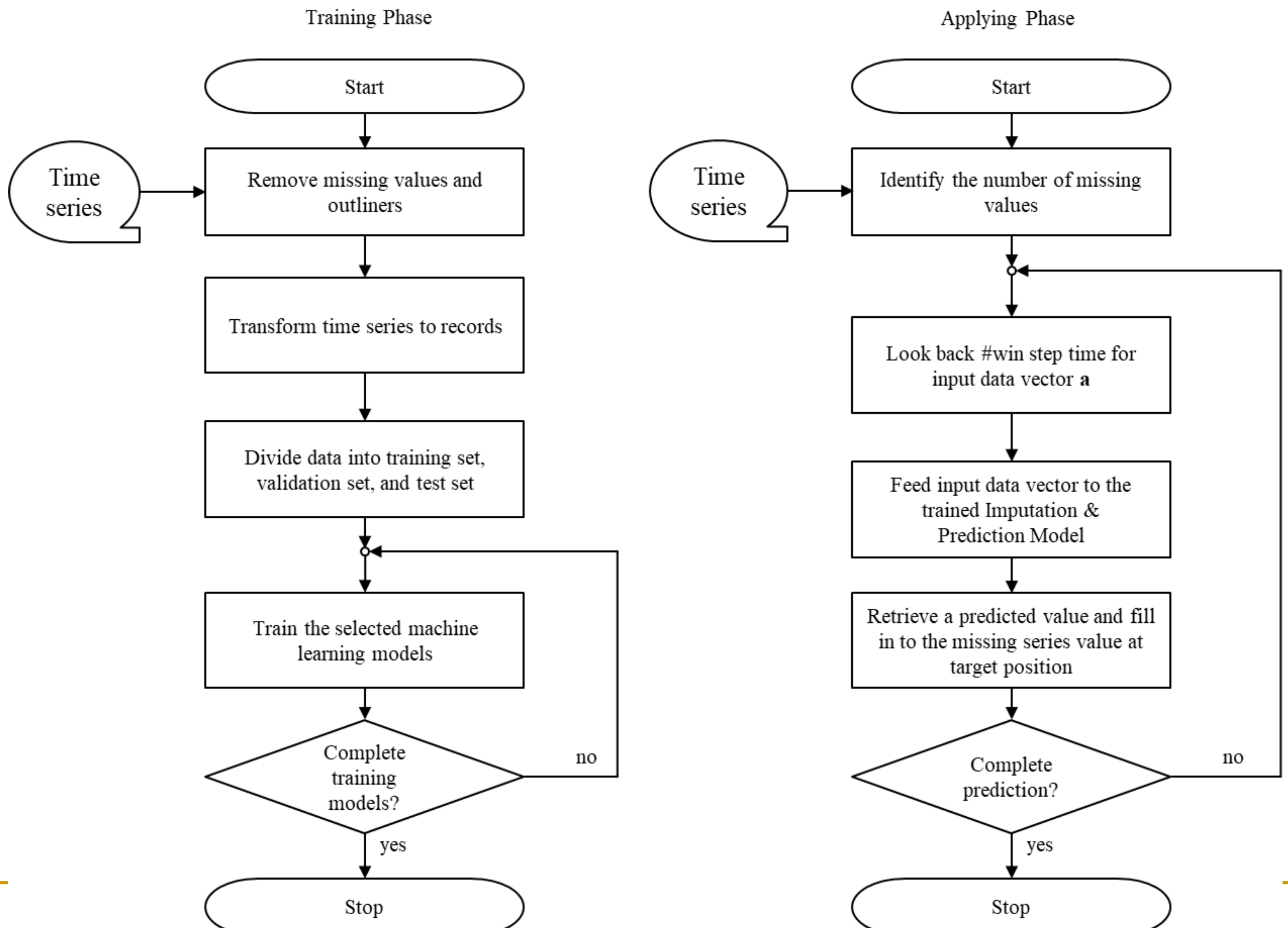
$$s(t) = \{x(t-R), \ldots, x(t-2, x(t-1), x(t), x(t+1)\}$$

$$x(t) = \{x(t-R), \ldots, x(t-2, x(t-1), x(t)\}$$

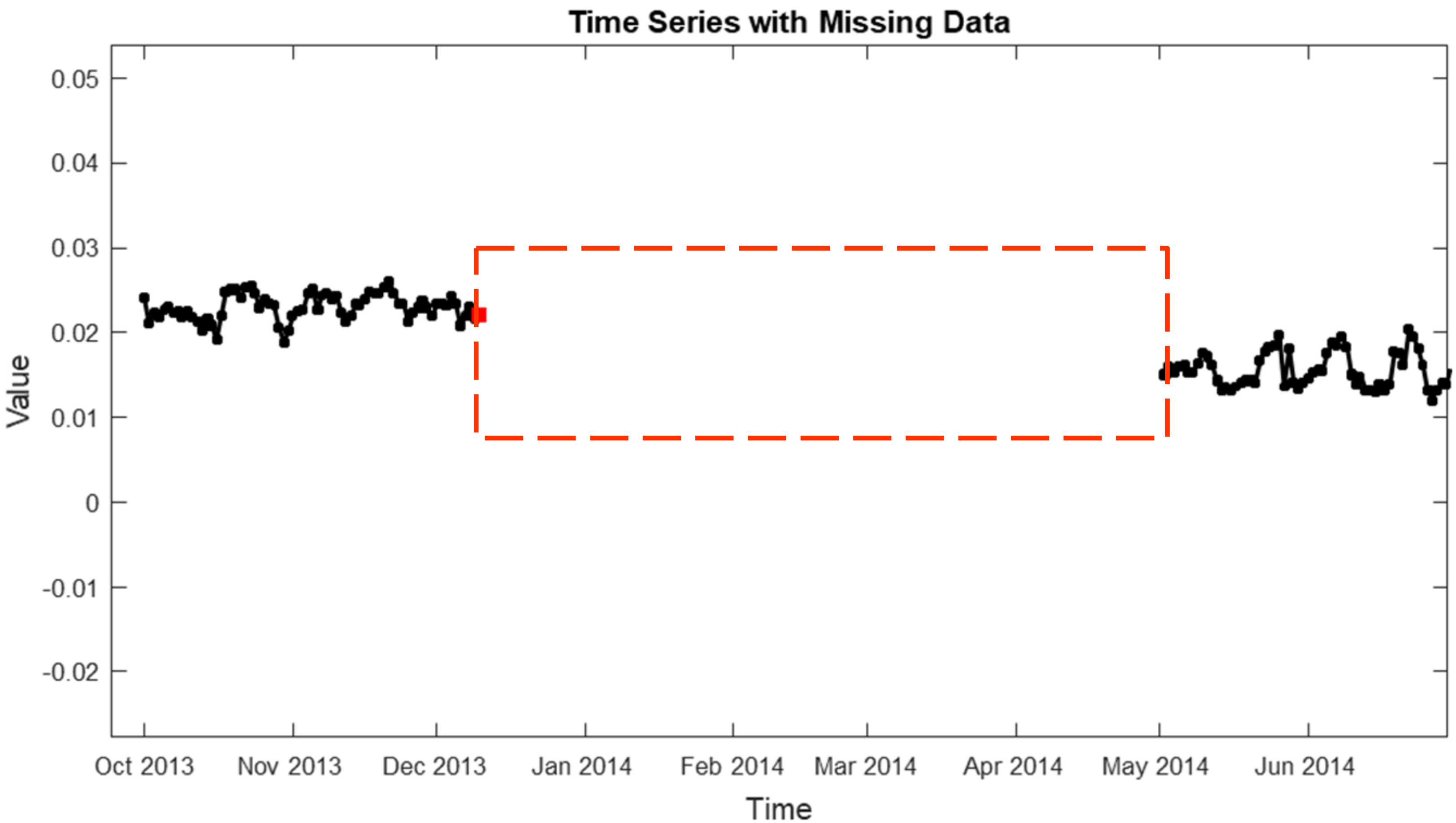
$$y(t) = x(t+1)$$

$$s(t) = \{x_1(t-R), \ldots, x_1(t-1), x_1(t), x_2(t-R), \ldots, x_2(t-1), x_2(t), y(t)\}$$

Flowchart


Training Phase
Start
Time series
Remove missing values and outliners
Transform time series to records
Divide data into training set, validation set, and test set
Train the selected machine learning models
Complete training models?
no
yes
Stop
Applying Phase
Start
Time series
Identify the number of missing values
Look back #win step time for input data vector **a**
Feed input data vector to the trained Imputation & Prediction Model
Retrieve a predicted value and fill in to the missing series value at target position
Complete prediction?
no
yes
Stop

Missing Data



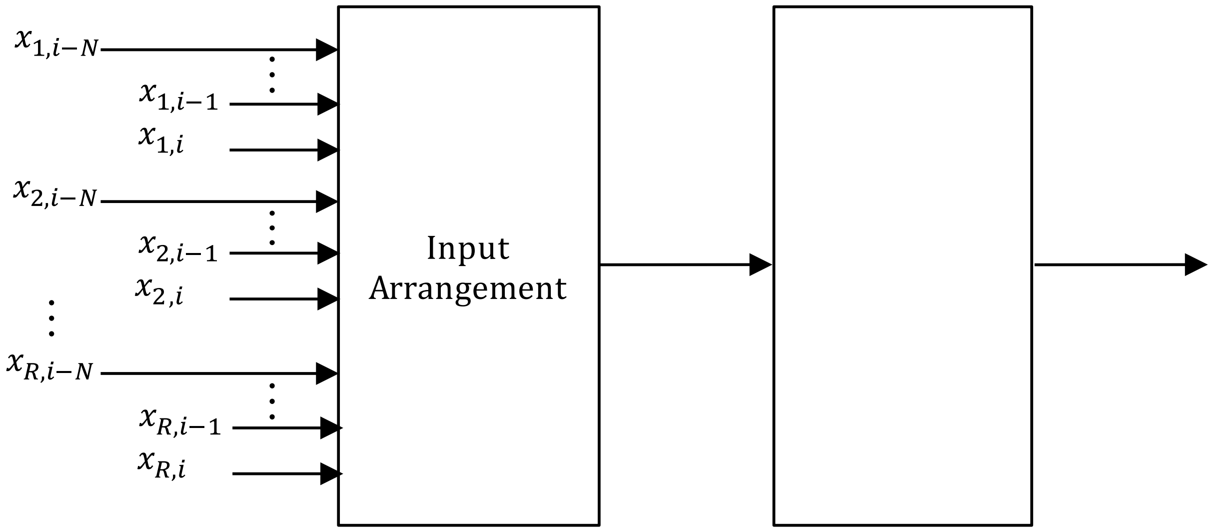
- Missing data in time series have the consequent analysis such as time series prediction cannot be done efficiently.

Time Series Imputation

- To build imputation and prediction model, firstly leaving out the missing, the available time series data are used for training, valuating, and testing.
- The trained model is later used to produce the new replacing values in each time series.
- The same model can be used for prediction of future values in the same series in which they will be used to predict unseen and new coming data.

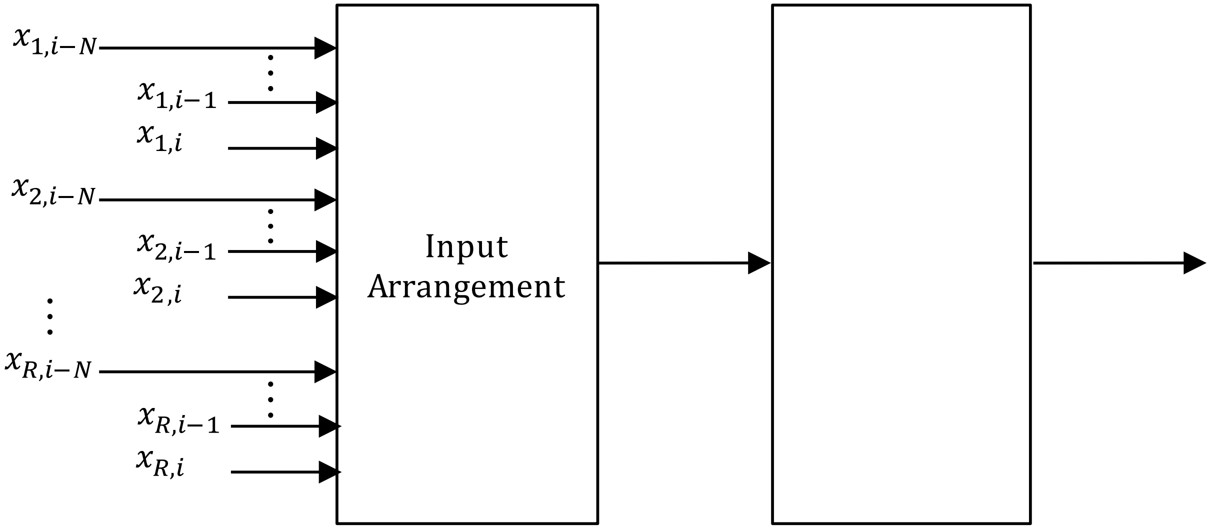
Machine Learning Approach

- The framework takes advantages of machine learning to perform both imputation of missing values and prediction for future values.
  - The model is trained and validated using available data in the time series.
  - The trained model is applied to perform missing values imputation as well as prediction of the future values.

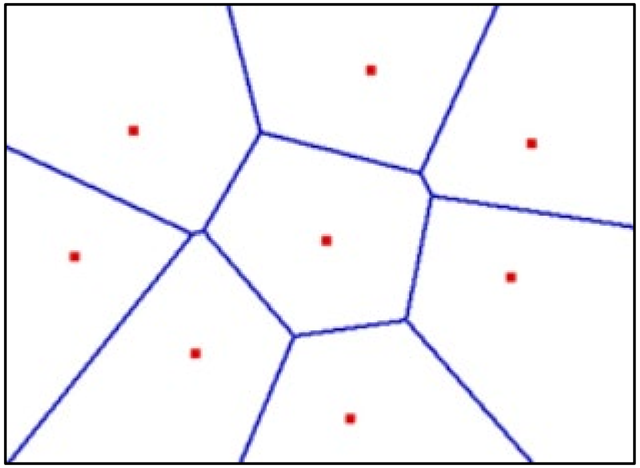


Machine Learning Imputation

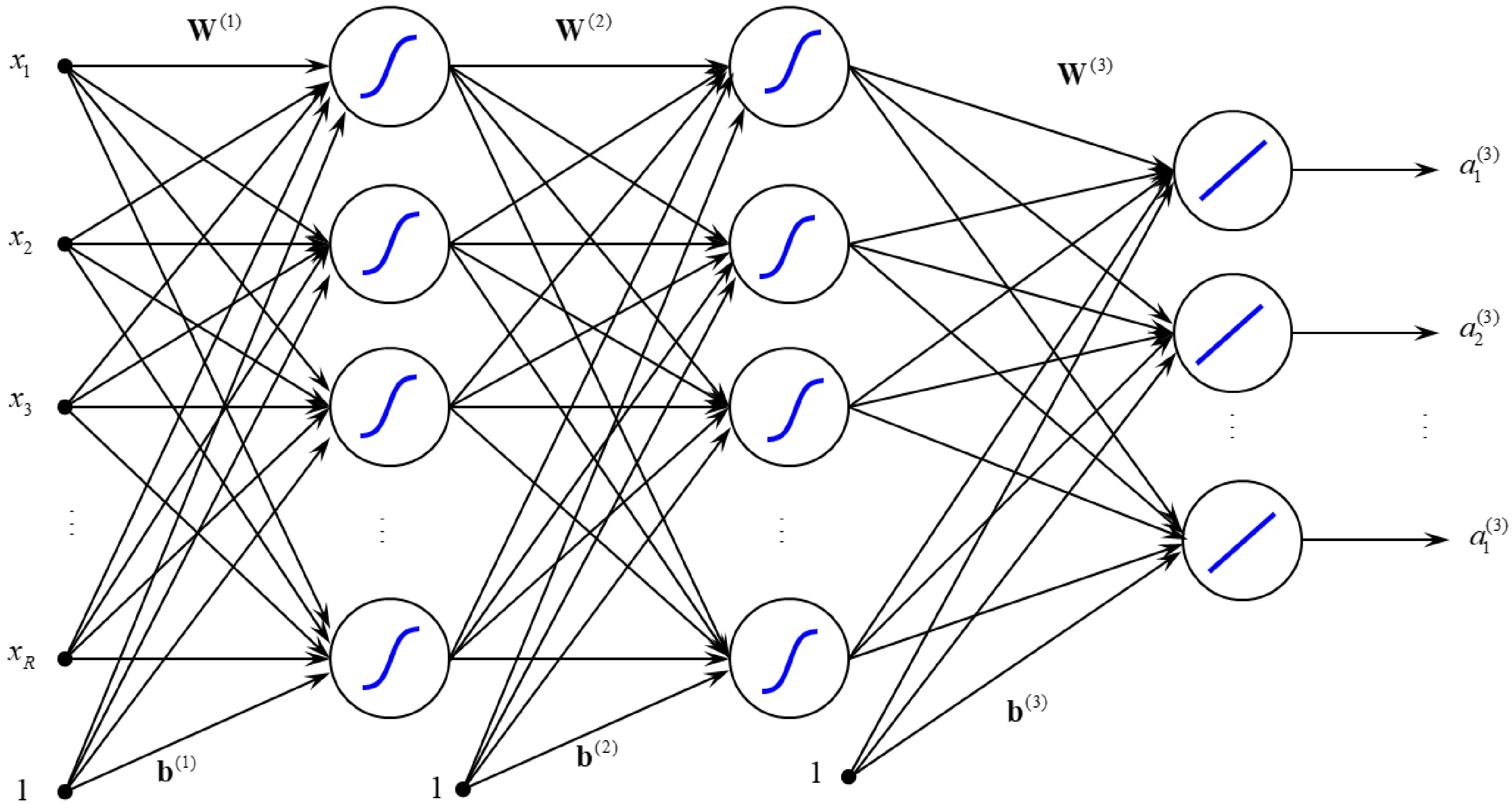
- K-Nearest Neighbour (KNN)
- Multilayer Perceptron (MLP),
- Adaptive Neuro Fuzzy Inference System (ANFIS),
- Support Vector Machine (SVR).



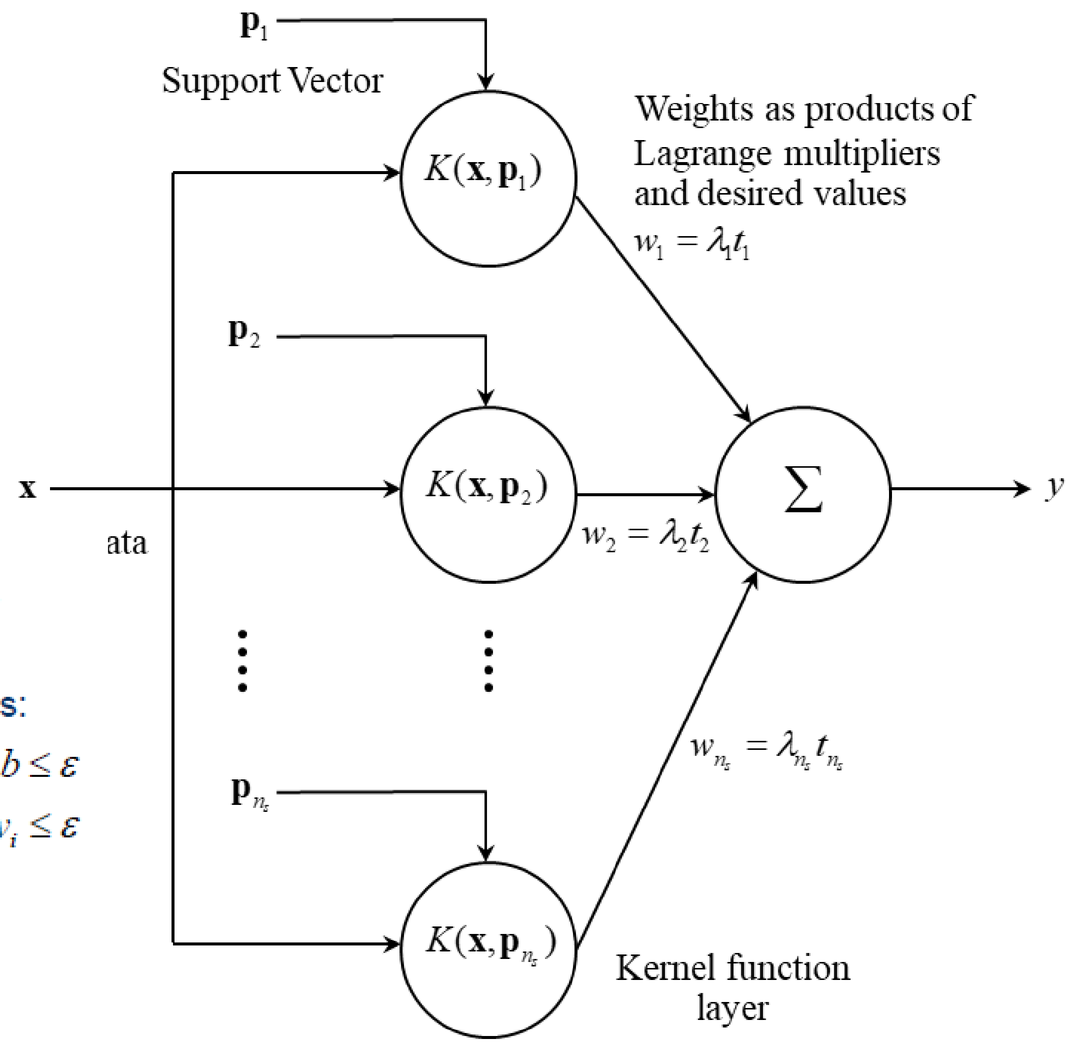
K-Nearest Neighbor: KNN

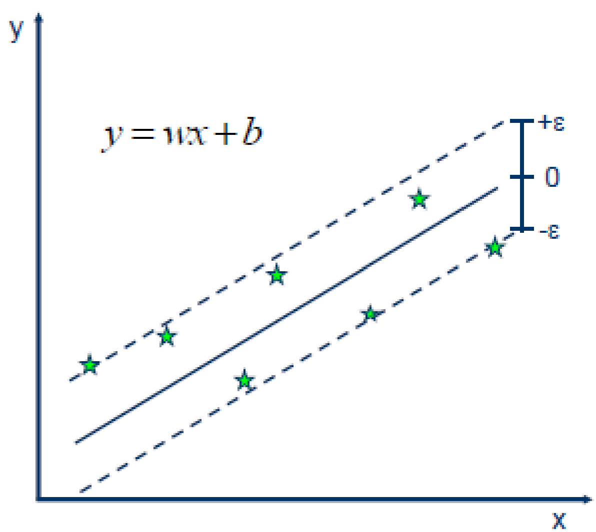


Multilayer Perceptron NN



Support Vector Machine Regression





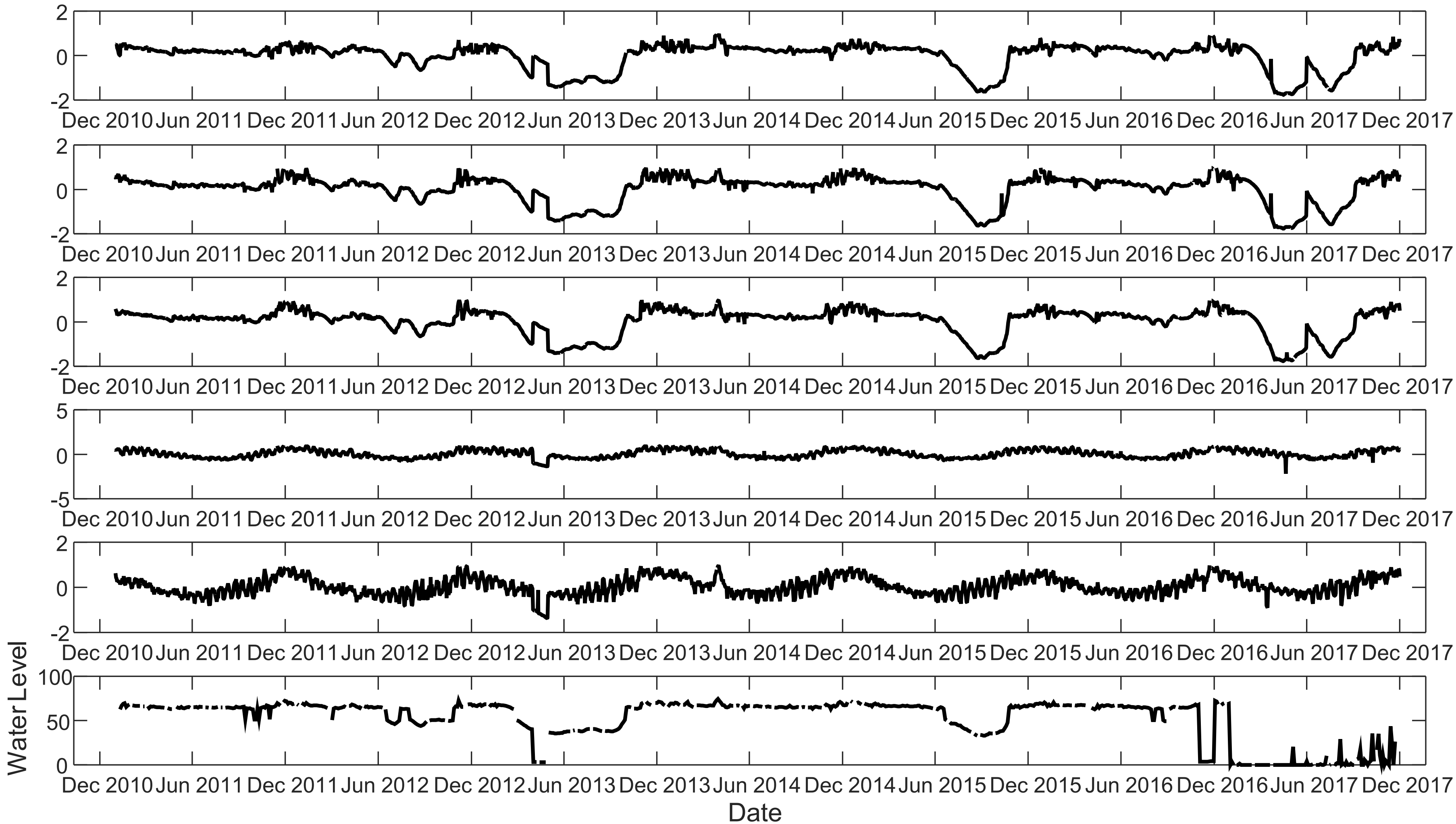
- Solution:

$$\min \frac{1}{2}\|w\|^2$$

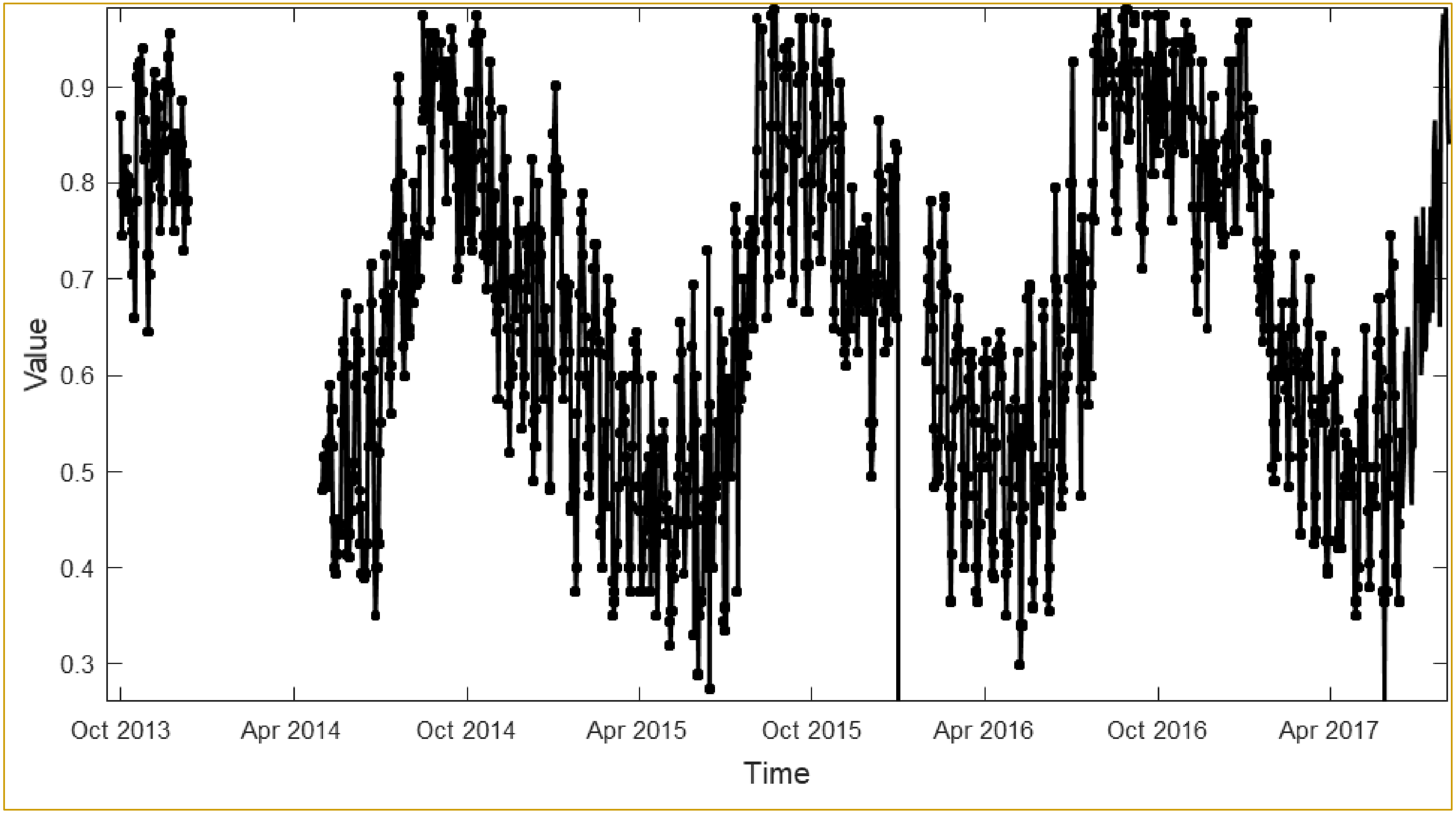
- Constraints:

$$y_i - wx_i - b \le \varepsilon$$
$$wx_i + b - y_i \le \varepsilon$$

Multivariate Time Series



Missing Data in Time Series



Imputation Example