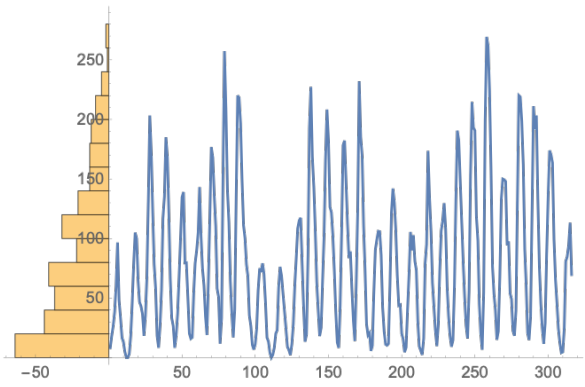
CIS 422/522
PROJECT 1
TIME SERIES CONCEPTS

Professor: Juan J. Flores
jflore10@uoregon.edu
UNIVERSITY OF OREGON

1

TIME SERIES CONCEPTS

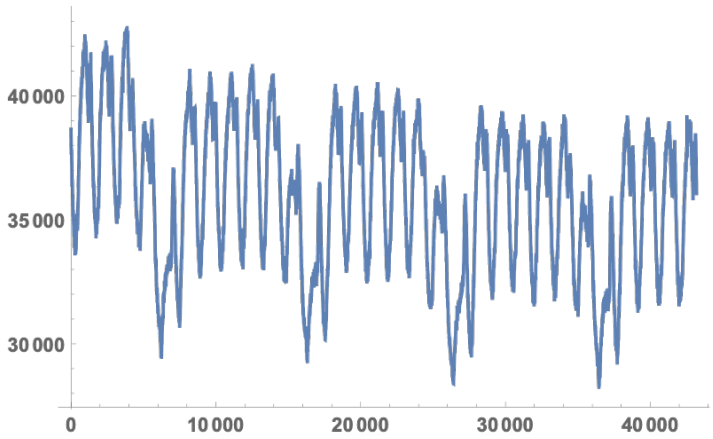
- Time Series
- Preprocessing
- Forecasting Models
- Forecasting
- TS2DB



2

TIME SERIES

$(x_t)\ t \in [1,T]$



3

Time Series Forecasting

Time Series Forecasting Methods

- Cualitative
  - Subjetive
  - No data available (new products)
  - Experts Opinion
- Cuantitative
  - Use historical data
  - Mathematical/Statistical Model
  - Models behavior patterns
  - Proyects those patterns to the future

4

Time Series Forecasting

Quantitative Methods

- Regression
- Smoothing
- ARIMA
- Artificial Intelligence

5

Preprocessing

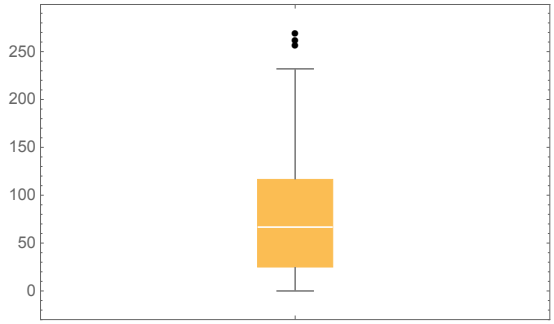
- Statistical Analysis
- Noise
- Outliers
- Missing Data
- Autocorrelation
- Chaos

6

Statistical Analysis

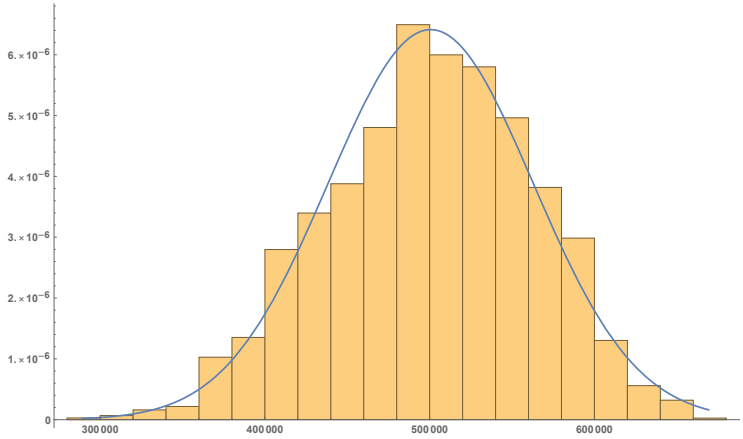
5-number summary
{0.,24.9,66.7,116.3,269.3}

Mean = 79.5032
SD = 62.0572



7

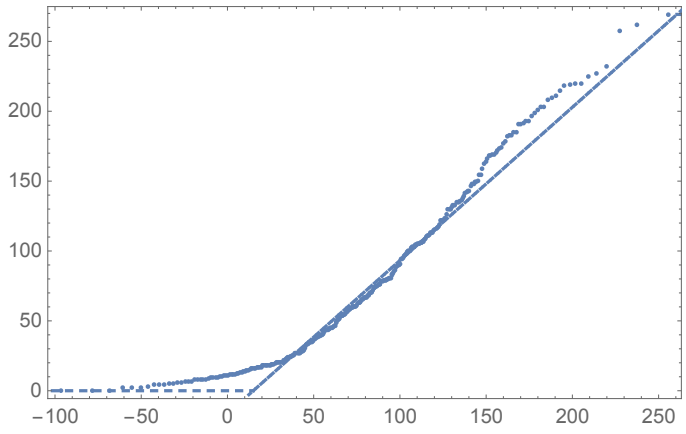
Statistical Analysis



8

Statistical Analysis

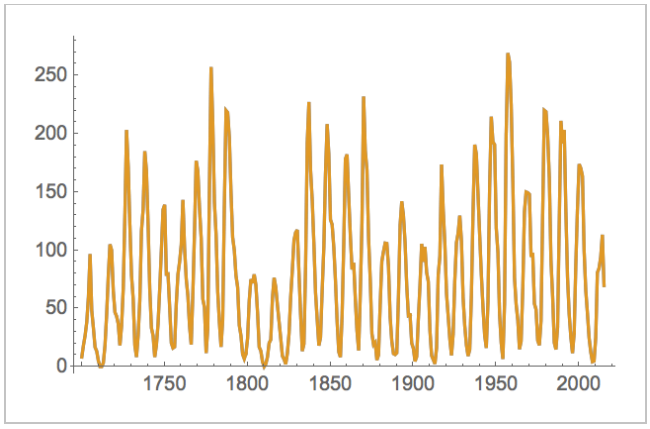
The null hypothesis that the data is distributed according to the NormalDistribution[79.5032,62.0572] is rejected at the 5% level, based on the Cramér-von Mises test.



9

Noise Removal

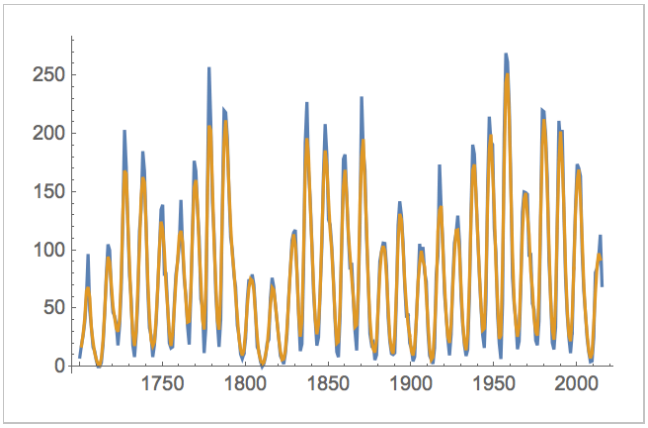
- Moving average
- Moving medians



10

Noise Removal

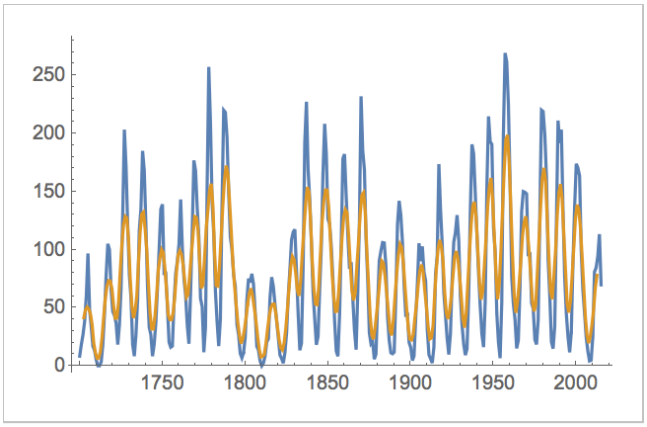
- Moving average
- Moving medians



11

Noise Removal

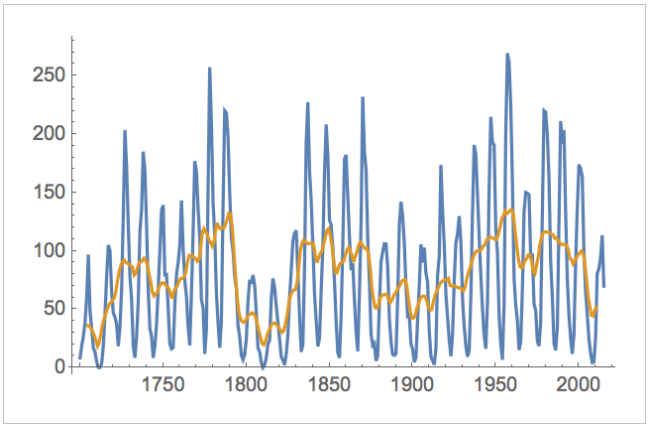
- Moving average
- Moving medians



12

Noise Removal

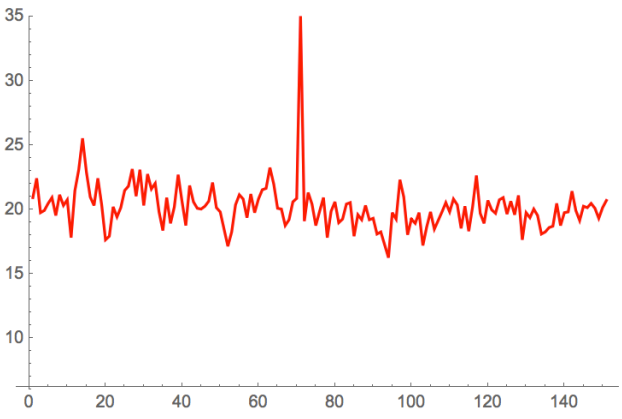
- Moving average
- Moving medians



13

Outliers

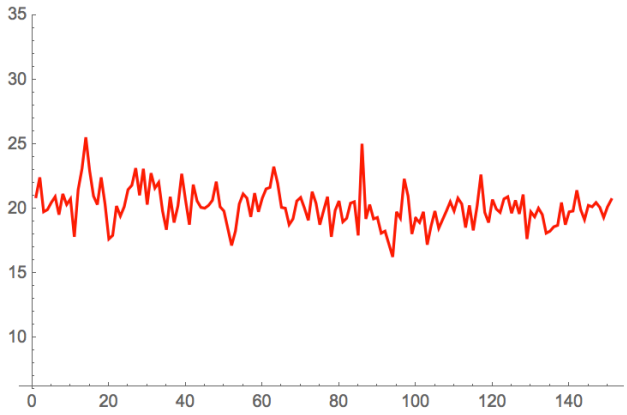
- Context sensitive
- Air Temperature TS
- Synthetic Outlier



14

Outliers

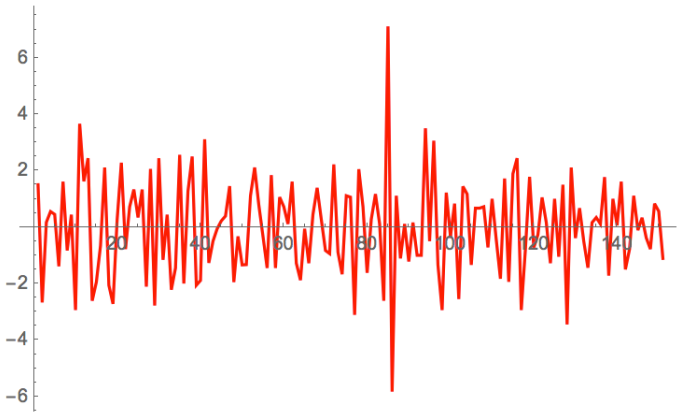
- Synthetic outlier
- Can you spot it?



15

Outliers

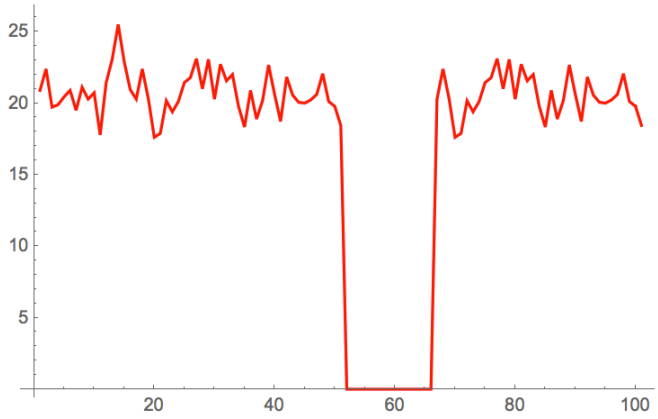
- Difference $T_i - T_{i-1}$



16

Missing Data

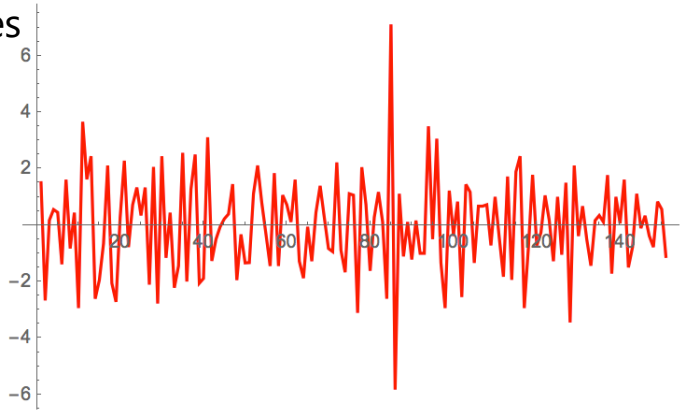
- Blackout
- Transmission failure
- Sick day



17

Differencing

- Difference also eliminates trend
- Discrete derivative



18

Other Transformations

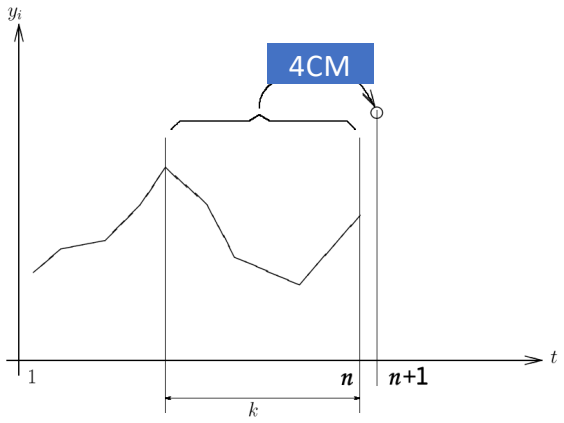
- Scaling/Normalization
- Standardization
- Logarithm
- Cubic Root

- y = (x - min) / (max - min)
- y = (x - μ) / σ
- Log(x)
- $\sqrt[3]{x}$

19

Forecasting Models

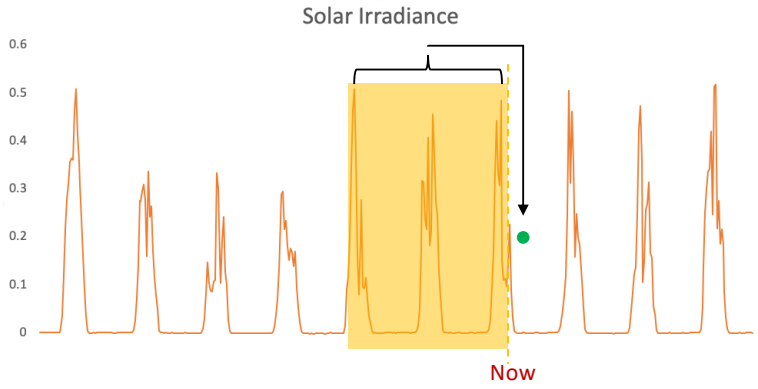
- Based on the TS history
- Determine the value at the next time instant



20

TS Forecasting

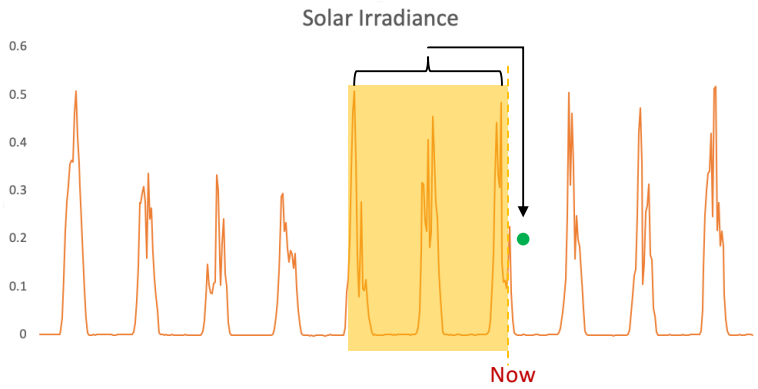
- Based on TS history
- Determine y_{n+1}



21

TS Forecasting

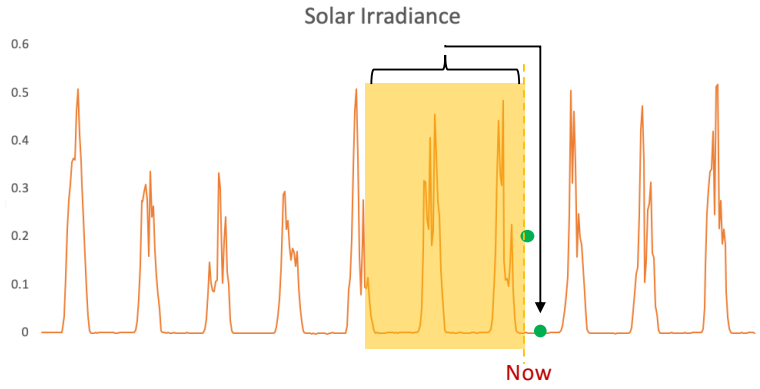
- OSA – One Step Ahead Forecasting
- Determine $<y_{n+1}, y_{n+2}, \dots, y_{n+k}>$
- One at a time



22

TS Forecasting

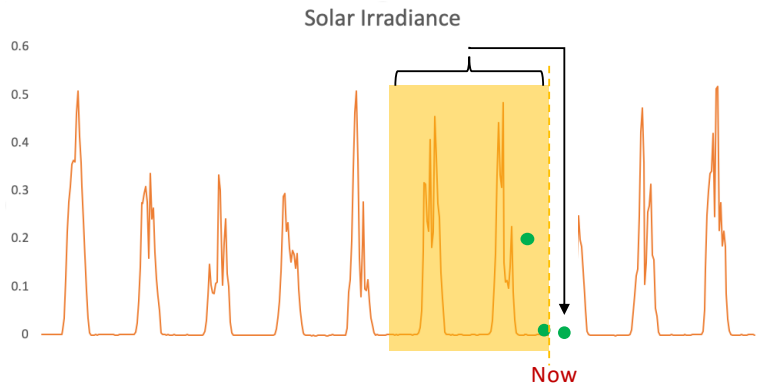
- OSA – One Step Ahead Forecasting
- Determine $<y_{n+1}, y_{n+2}, \dots, y_{n+k}>$
- One at a time



23

TS Forecasting

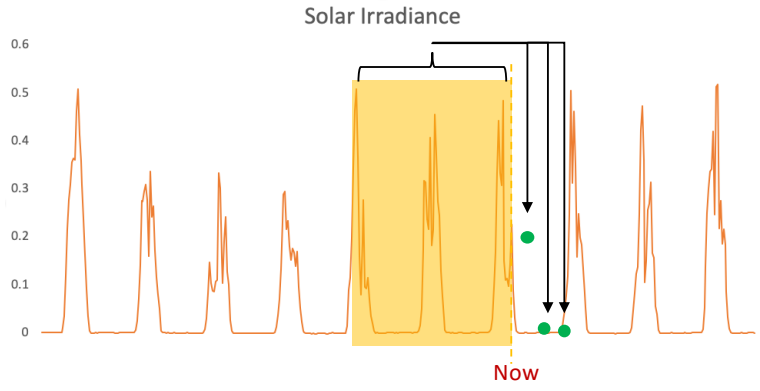
- OSA – One Step Ahead Forecasting
- Determine $<y_{n+1}, y_{n+2}, \dots, y_{n+k}>$
- One at a time



24

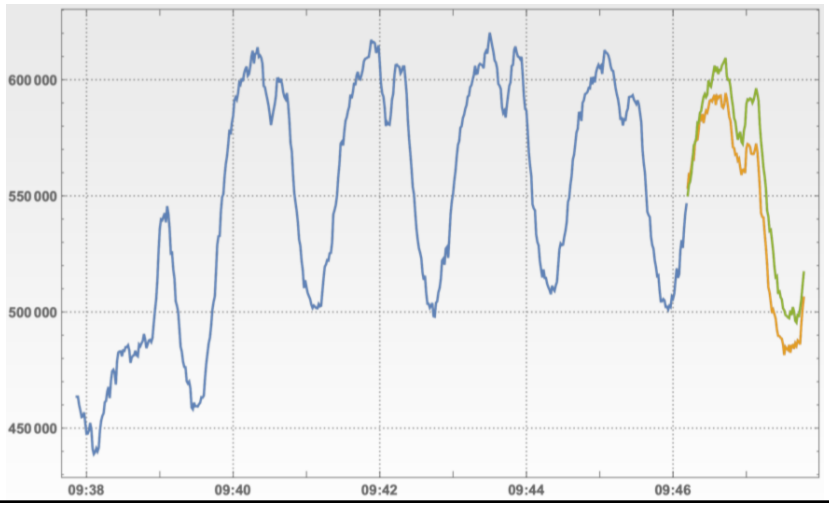
TS Forecasting

- MSS – Multiple Step Simultaneous
- Determine Determine <y_{n+1}, y_{n+2}, … , y_{n+k}>
- Simultaneously



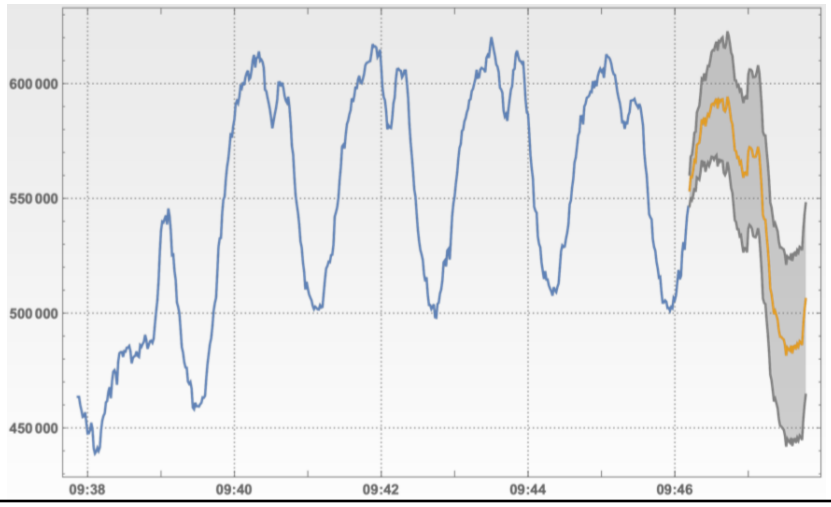
25

Point Forecasting



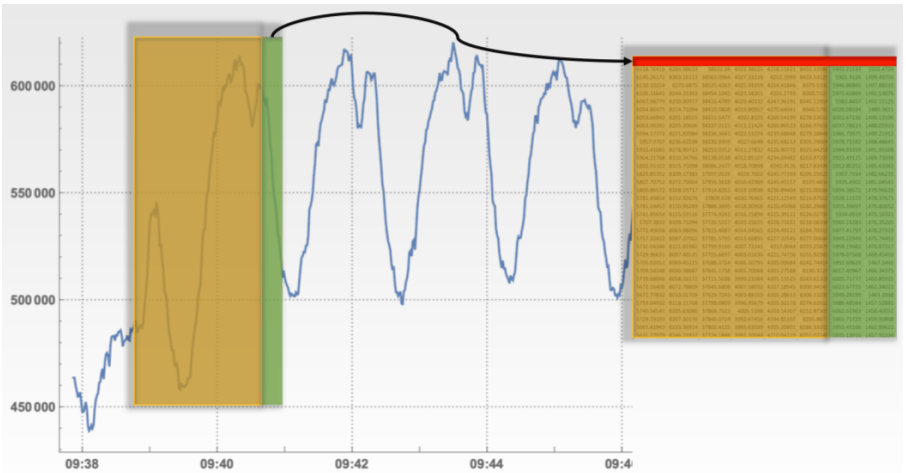
26

Interval Forecasting



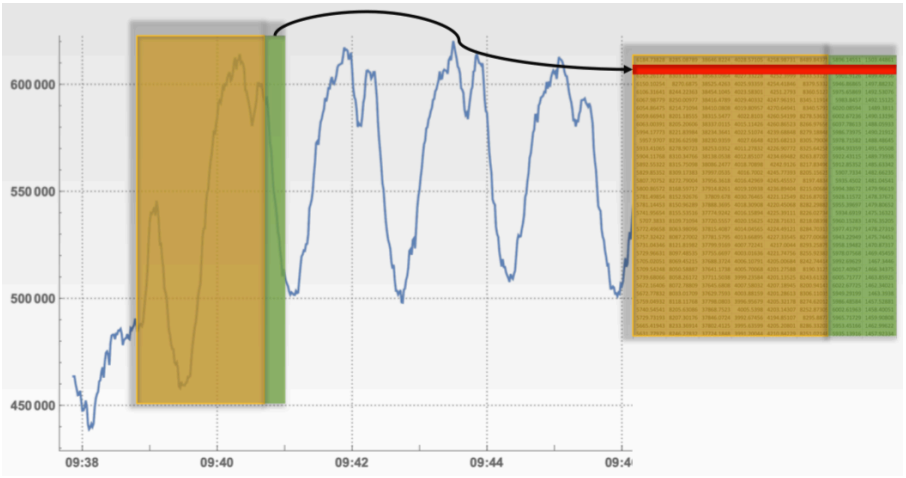
27

TS to DB



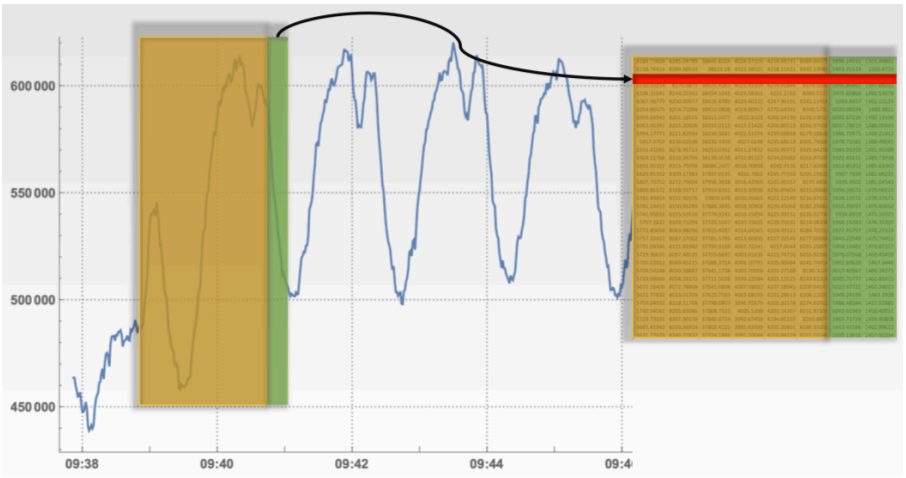
28

TS to DB



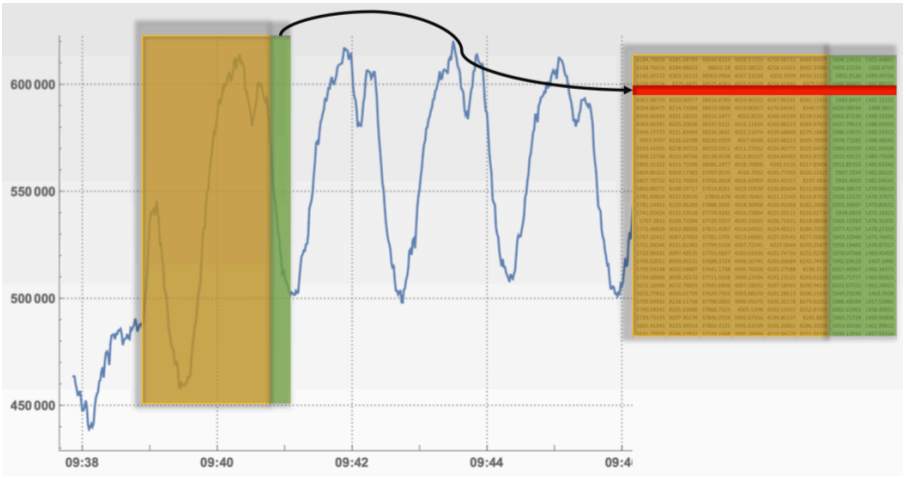
29

TS to DB



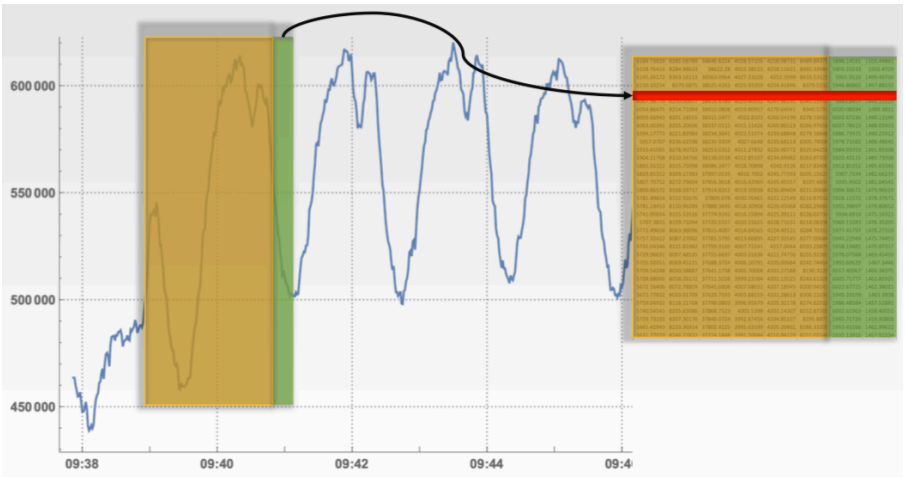
30

TS to DB



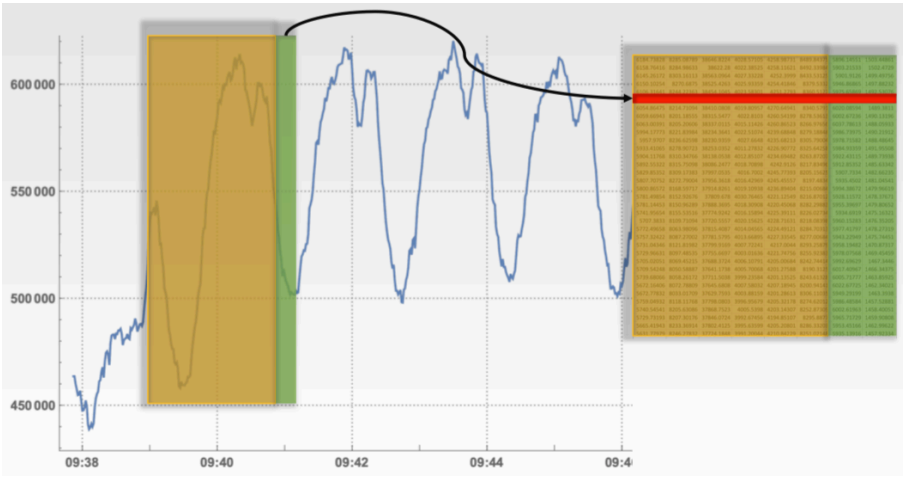
31

TS to DB



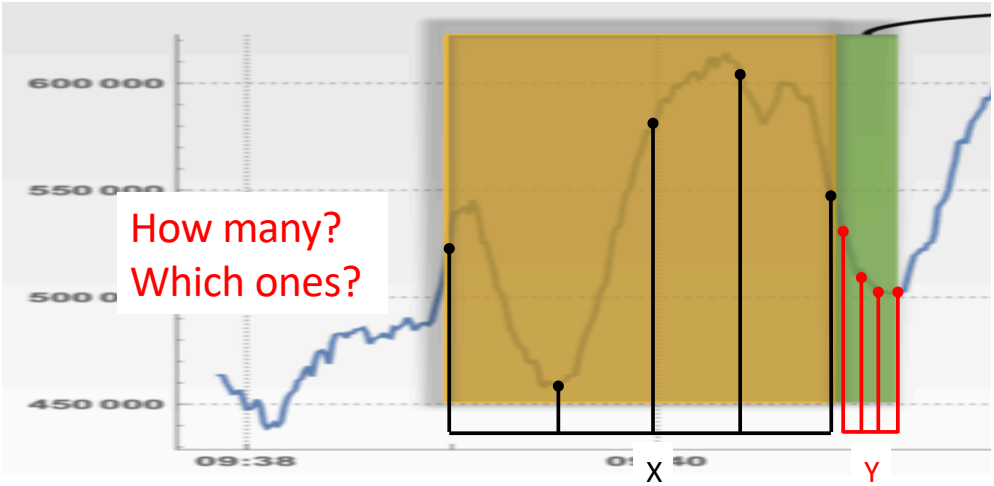
32

TS to DB



33

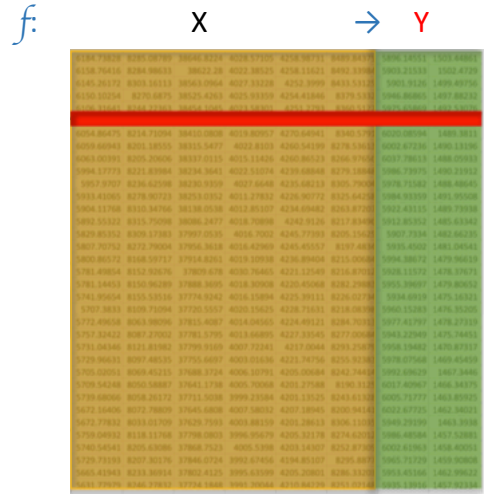
TS to DB



34

TS to DB

- Design Matrix
- Maps Forecasting to Regression
- Eliminates time (sequentiality)
- Shuffle records
- Be careful when splitting Training and Test sets



35

Error

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

$$\text{SMAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\frac{|F_t - A_t|}{|A_t| + |F_t|}$$

36