Sales planning

The set of retailers problems:

- custom inventory,
- calculation of optimal insurance stocks,
- consumer demand forecasting.

The initial problem statement

There given:

- time-scale,
- historical time series,
- additional time series;

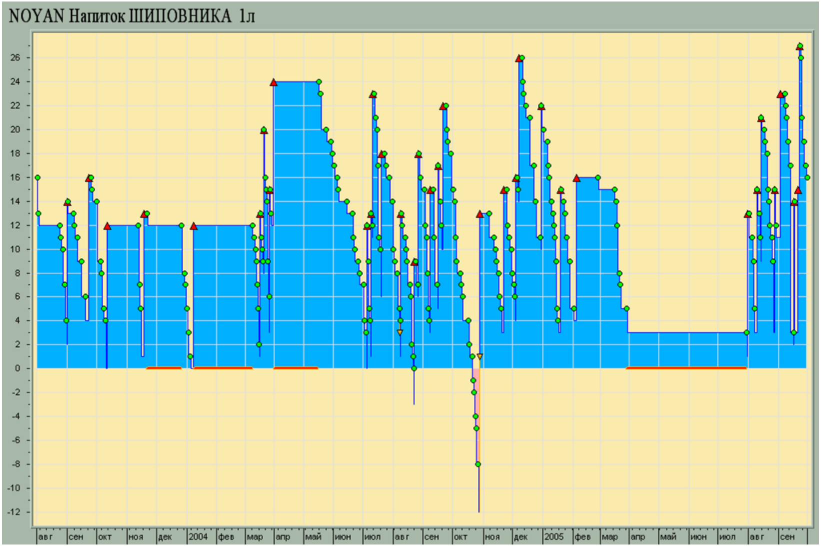
the quality of forecasting:

- minimum loss of money;

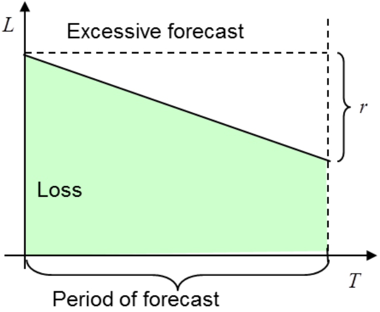
we must:

- forecast the time series.

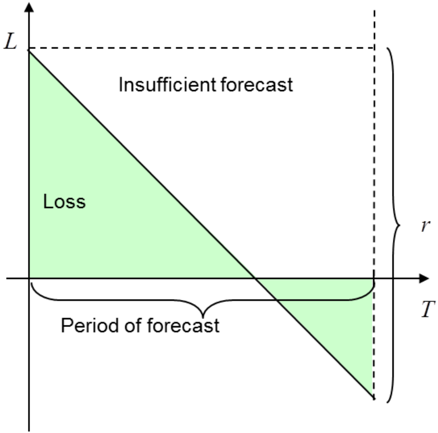
Custom Inventory



Excessive forecast



Insufficient forecast

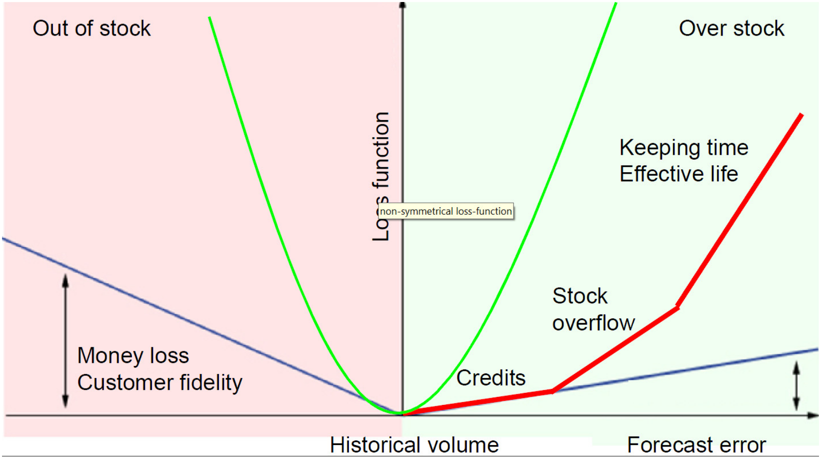


Loss function for the forecast error

reflects the sales process and is depends on the basis of the features of a particular trading network

- Symmetric quadratic function
- Module function
- Asymmetric function

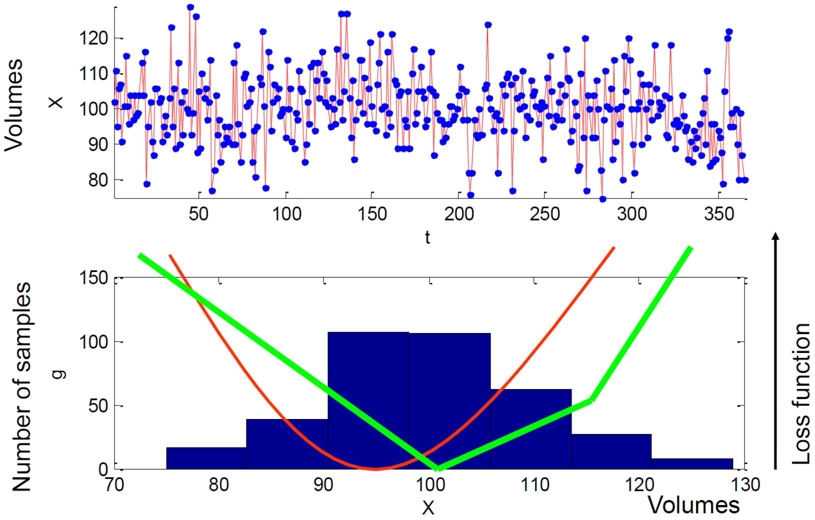
Asymmetric loss function



Noisy time series forecasting

- There is a historical time series of the volume off-takes (i.e. foodstuff).
- Let the time series be homoscedastic (its variance is the time-constant).
- Using the loss function one must forecast the next sample.

The time series and the histogram



The forecasting algorithm

Let there be given:

$$\text{the historgam} \qquad H = \{X_i, g_i\}, i = 1, \dots, m;$$

$$\text{the loss function} \quad L = L(Z, X);$$

for example, $L = |Z - X|$ or $L = (Z - X)^2$.

The problem:

For given H and L, one must find the optimal forecast value $\tilde{X}$.

Solution:

$$\tilde{X} = \arg \min_{Z \in \{X_1, \dots, X_m\}} \sum_{i=1}^{m} g_i L(Z, X_i).$$

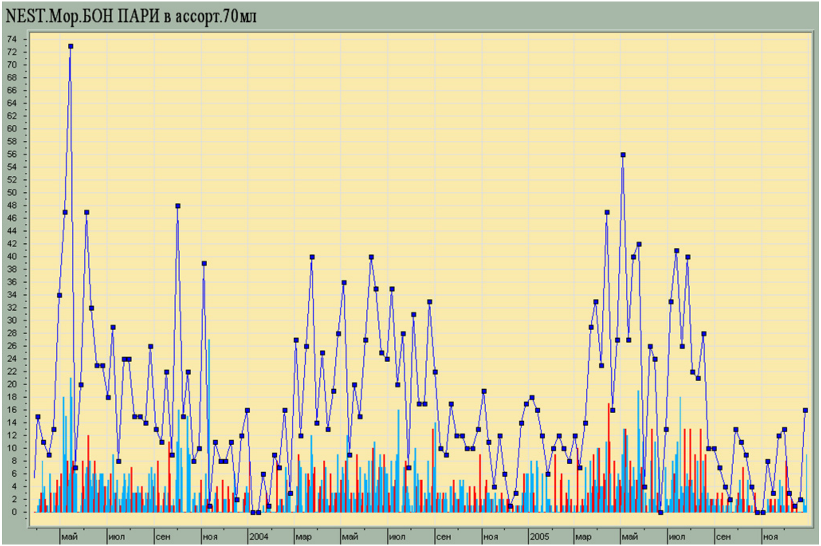
Result:

$\tilde{X}$ is the optimal forecast of the time series.

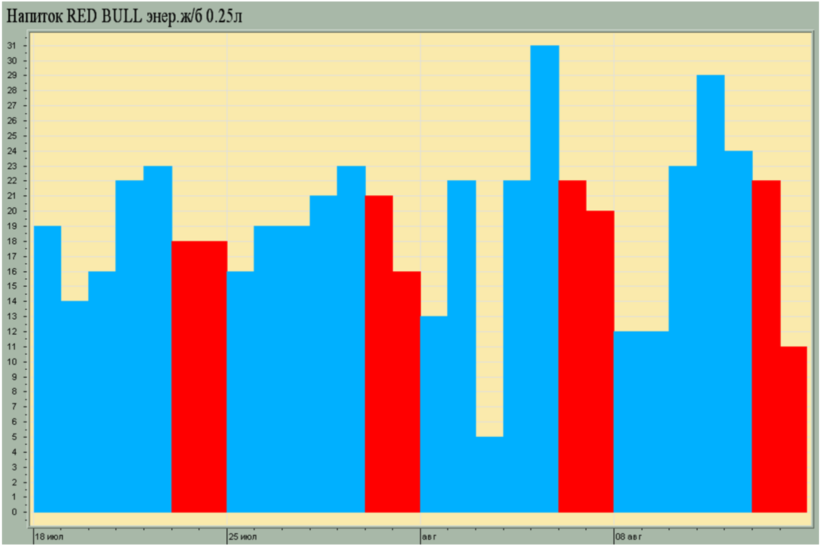
The sales time series is non-stationary

- There is a trend — total increase or decrease in sales volume,
- periodic component — week and year cycles,
- aperiodic component — promotional actions and holidays,
- life cycle of goods — mobile phones.

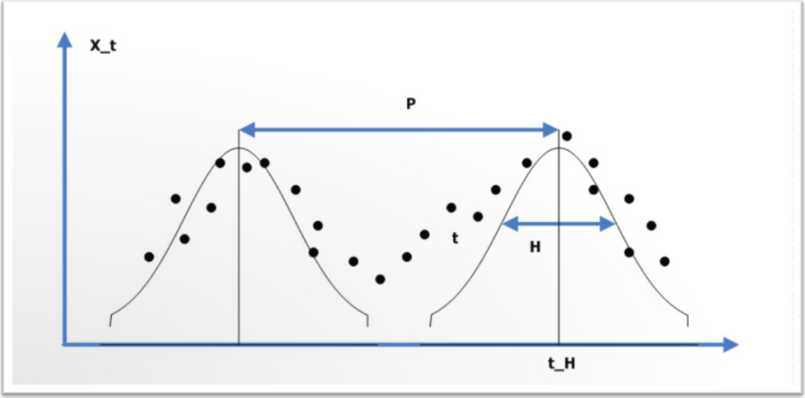
Year seasonality



Week seasonality

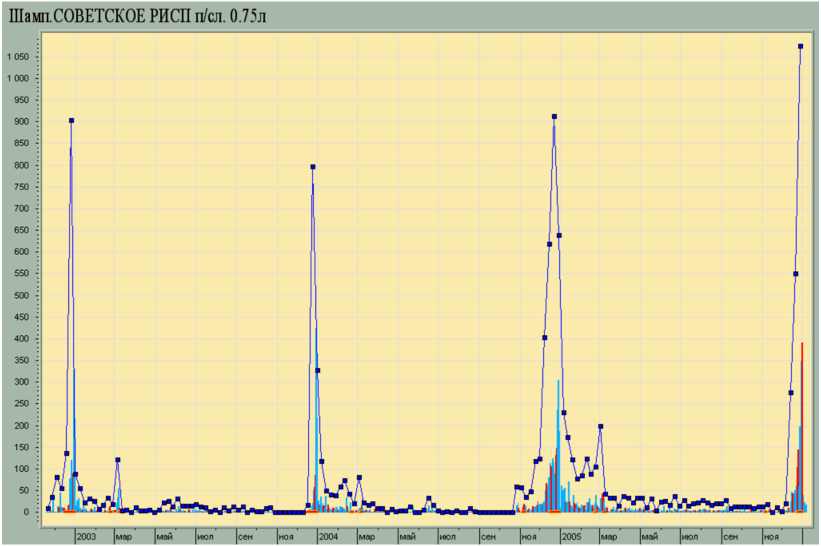


How to create sample set of periodical series

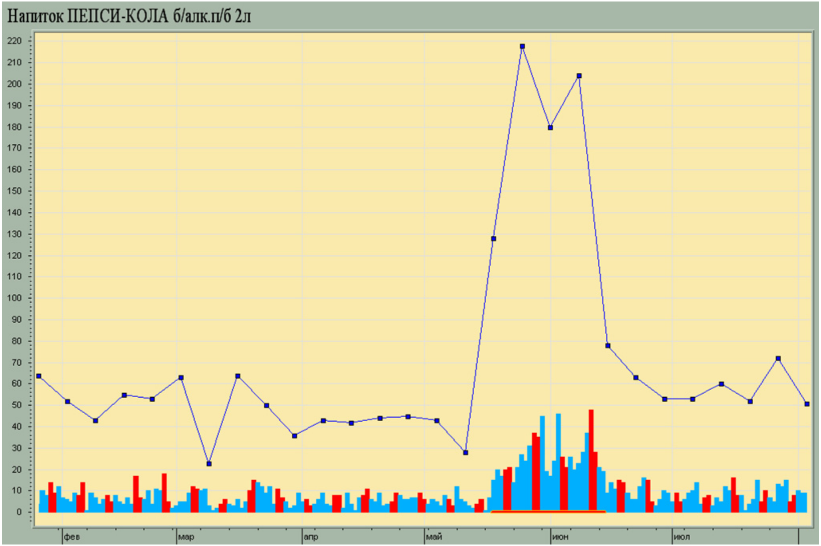


Weighting function includes neighborhood points to the sample set.

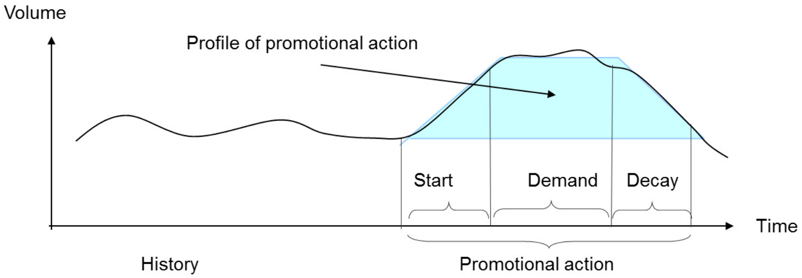
Holidays and week-ends



Promotional actions



Promotional profile extraction



- Hypothesis: the shape of the profile (excluding the profile parameters) does not depend of duration of the action.
- Problem: to forecast the customer demand during the promotional action.

Algorithmic composition

On forecasting one must consider:

- trend of time series,
- periodical components of time series,
- aperiodical components,

as well as the fact that the time series contain

- empty values — there is no information about the stock,
- empty values — new position on the stock,
- outliers and errors.

Hour by Hour Energy Forecasting

Data:

- historical consumption and prices, multivariate time series.

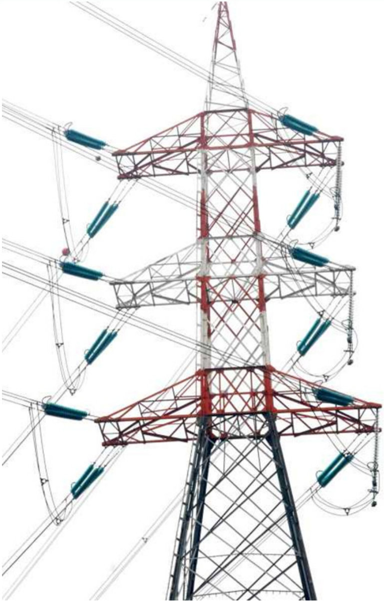
To forecast:

- hour-by-hour, the next day

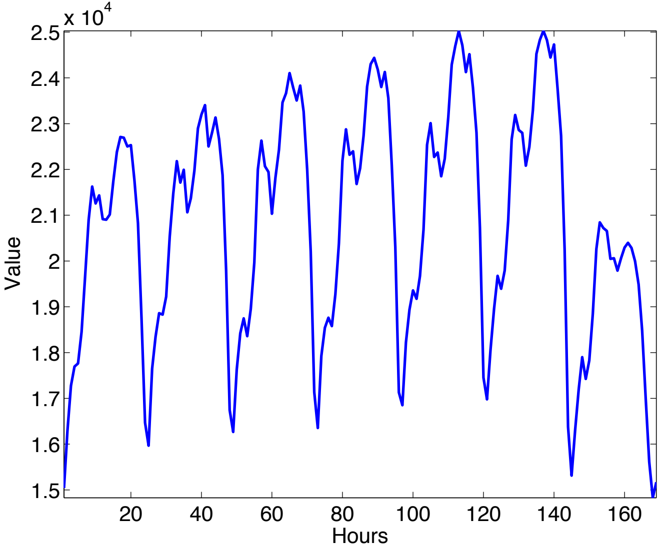
  ✓ consumption and
  ✓ price.

Solution:

- the autoregressive model generation and model selection.



Source time series, one week



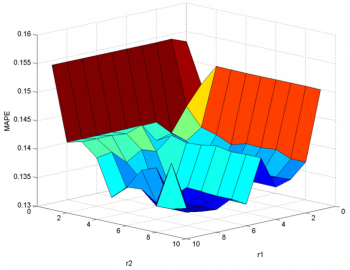
Quality of the forecasting model

Mean absolute percentage error

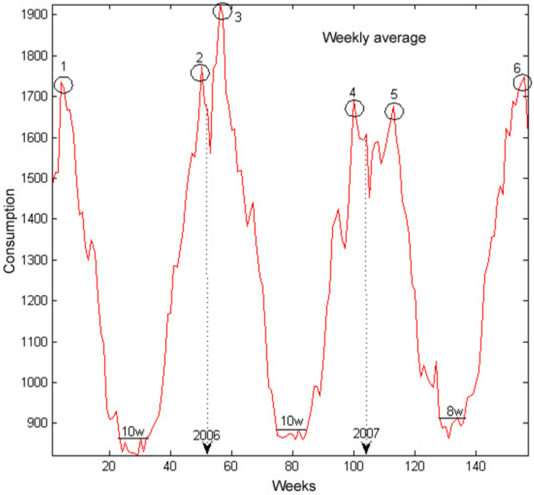
$$MAPE = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{f_i - y_i}{y_i}\right|.$$

Mean squared error

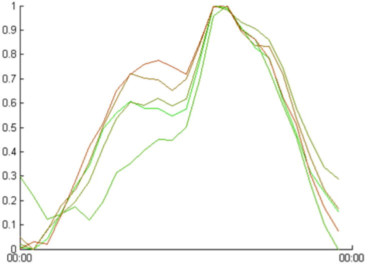
$$MSE = \frac{1}{m}\sum_{i=1}^{m}(f_i - y_i)^2.$$



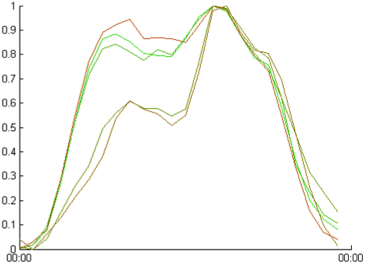
Structure of energy consumption



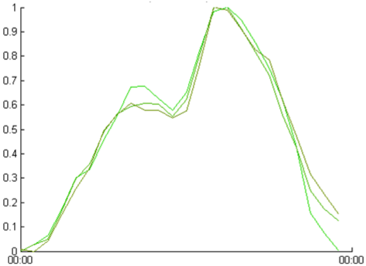
Similarity of daily consumption



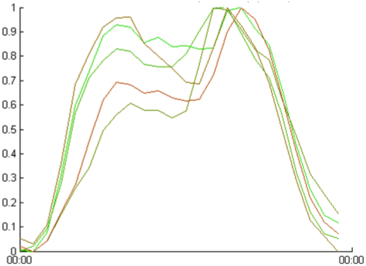
Five days of one week



One day of five consequent weeks

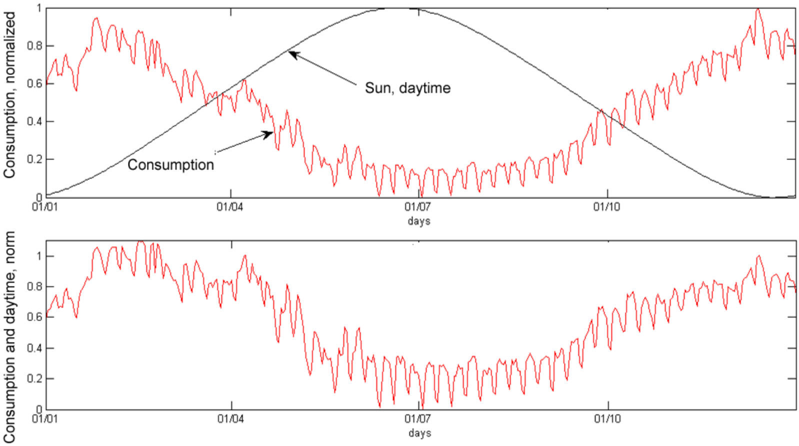


The same weekday and month of three years

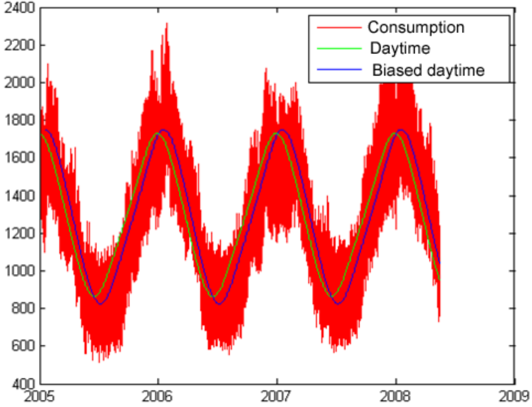


The same weekday of five months

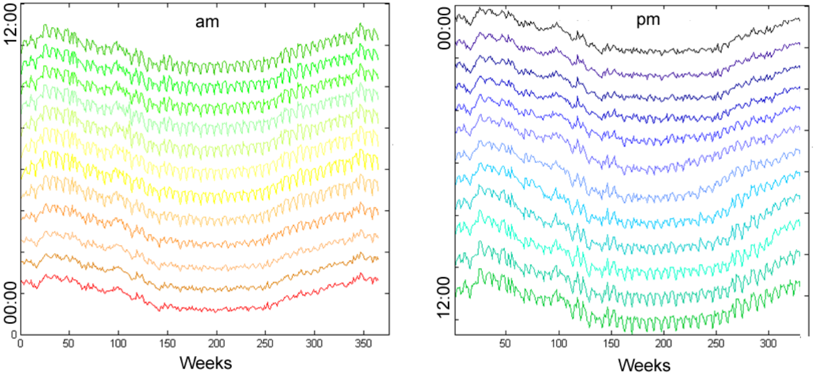
Sunrise bias: one-year daytime and consumption



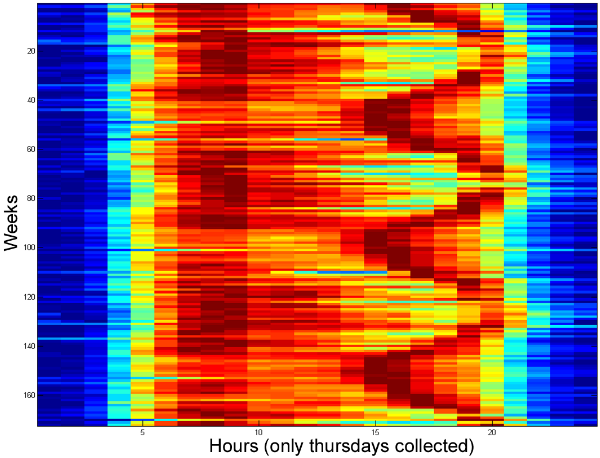
Biased and original daytime to fit consumption over years



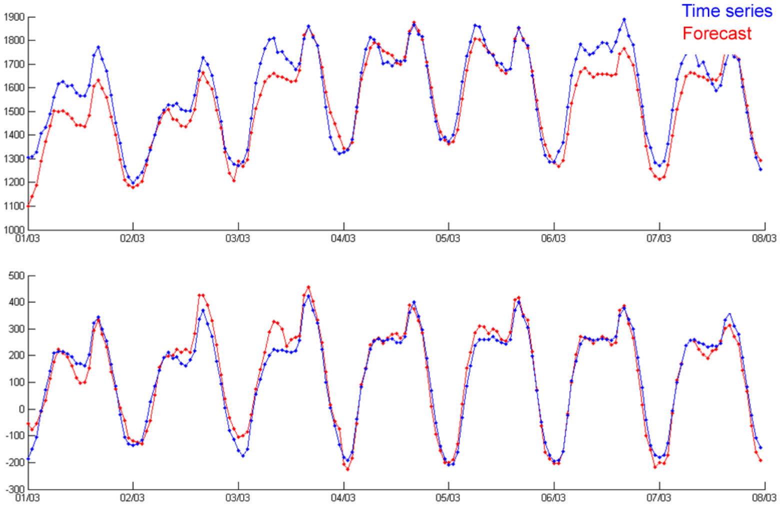
One-hour line, day-by-day during a year: autoregressive analysis



Daily similarity



Energy consumption one-week forecast



The periodic components of the multivariate time series

The time series:

- energy price,
- consumption,
- daytime,
- temperature,
- humidity,
- wind force,
- holiday schedule.

Periods:

- one year seasons (temperature, daytime),
- one week,
- one day (working day, week-end),
- a holiday,
- aperiodic events.

The autoregressive matrix to forecast periodic time series

- There given the time series $\{s_1, \dots, s_\tau, \dots, s_{T-1}\}$, the length of a period is κ.
- One must to forecast the next sample T.
- The autoregressive matrix:
  - its i-th row is a period of samples,
  - its j-th column is a phase of the period and
  - they map into the time series sample number such that $(i-1)\kappa \mapsto \tau$; let $\mathrm{mod}\frac{T}{\kappa} = 0$;

$$\underset{(m+1)\times(n+1)}{X^*} = \begin{pmatrix} s_T & s_{T-1} & \dots & s_{T-\kappa+1} \\ s_{(m-1)\kappa} & s_{(m-1)\kappa-1} & \dots & s_{(m-2)\kappa+1} \\ \dots & \dots & \dots & \dots \\ s_{n\kappa} & s_{n\kappa-1} & \dots & s_{n(\kappa-1)+1} \\ \dots & \dots & \dots & \dots \\ s_\kappa & s_{\kappa-1} & \dots & s_1 \end{pmatrix}.$$

The autoregressive matrix and the linear model

$$\underset{(m+1)\times(n+1)}{X^*} = \left(\begin{array}{c|cccc} s_T & s_{T-1} & \dots & s_{T-\kappa+1} \\ \hline & & & \\ s_{(m-1)\kappa} & s_{(m-1)\kappa-1} & \dots & s_{(m-2)\kappa+1} \\ \dots & \dots & \dots & \dots \\ s_{n\kappa} & s_{n\kappa-1} & \dots & s_{n(\kappa-1)+1} \\ \dots & \dots & \dots & \dots \\ s_\kappa & s_{\kappa-1} & \dots & s_1 \end{array}\right).$$

In a nutshell,

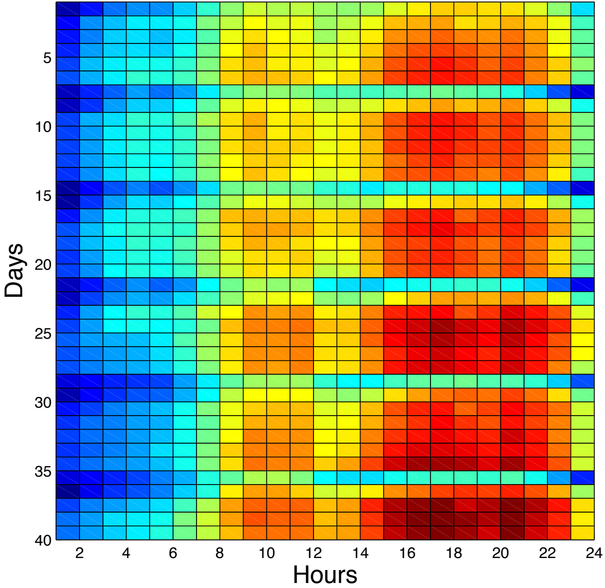
$$X^* = \left[\begin{array}{c|c} \underset{1\times 1}{s_T} & \underset{1\times n}{\mathbf{x}_{m+1}} \\ \hline \underset{m\times 1}{\mathbf{y}} & \underset{m\times n}{X} \end{array}\right].$$

In terms of linear regression:

$$\mathbf{y} = X\mathbf{w},$$

$$y_{m+1} = s_T = \mathbf{w}^\mathsf{T}\mathbf{x}_{m+1}^\mathsf{T}.$$

The autoregressive matrix, five week-ends



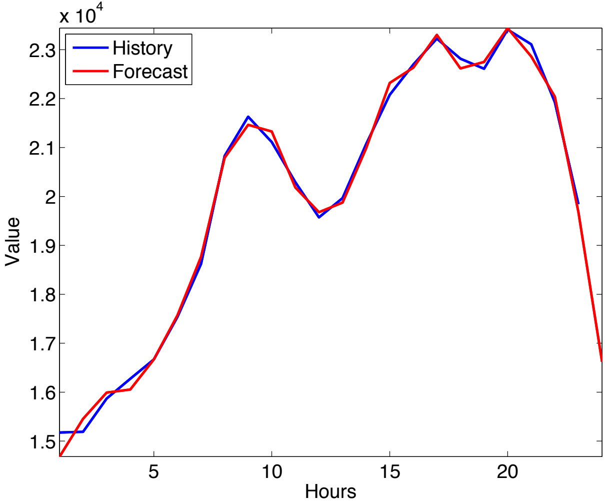
Model generation

Introduce a set of the primitive functions $G = \{g_1, \dots, g_r\}$, for example $g_1 = 1$, $g_2 = \sqrt{x}$, $g_3 = x$, $g_4 = x\sqrt{x}$, etc.

The generated set of features $X =$

$$\left(\begin{array}{lll|l|lll}
g_1 \circ s_{T-1} & \dots & g_r \circ s_{T-1} & \dots & g_1 \circ s_{T-\kappa+1} & \dots & g_r \circ s_{T-\kappa+1} \\
\hline
g_1 \circ s_{(m-1)\kappa-1} & \dots & g_r \circ s_{(m-1)\kappa-1} & \dots & g_1 \circ s_{(m-2)\kappa+1} & \dots & g_r \circ s_{(m-2)\kappa+1} \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots \\
g_1 \circ s_{n\kappa-1} & \dots & g_r \circ s_{n\kappa-1} & \dots & g_1 \circ s_{n(\kappa-1)+1} & \dots & g_r \circ s_{n(\kappa-1)+1} \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots \\
g_1 \circ s_{\kappa-1} & \dots & g_r \circ s_{\kappa-1} & \dots & g_1 \circ s_1 & \dots & g_r \circ s_1
\end{array}\right).$$

The one-day forecast, an example



Ill-conditioned matrix, or curse of dimensionality

Assume we have hourly data on price/consumption for three years. Then the matrix $\underset{(m+1)\times(n+1)}{X^*}$ is

$$156 \times 168, \text{ in details: } 52\text{w} \cdot 3\text{y} \times 24\text{h} \cdot 7\text{d};$$

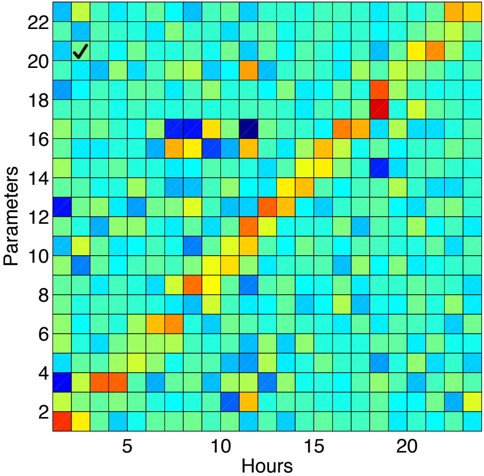
- for 6 time series the matrix X is 156×1008,
- for 4 primitive functions it is 156×4032,

$$m << n.$$

The autoregressive matrix could be considered as *ill-conditioned* and *multi-correlated*. The model selection procedure is required.

How many parameters must be used to forecast?

The color shows the value of a parameter for each hour.



Estimate parameters $\mathbf{w}(\tau) = (X^\mathsf{T}X)^{-1}X^\mathsf{T}\mathbf{y}$, then calculate the sample $s(\tau) = \mathbf{w}^\mathsf{T}(\tau)\mathbf{x}_{m+1}$ for each τ of the next ($m+1$-th) period.

The sample set and its indexes

There are given:

- the design matrix $X = [\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_m]^\mathsf{T}$,
- the target vector $\mathbf{y}$ and the data generation hypothesis $\mathbf{y} \sim \mathcal{N}(\mathbf{f}, \mathcal{B})$,
- the type of models $\{\mathbf{f} = X_\mathcal{A}\mathbf{w}_\mathcal{A} | \mathcal{A} \subseteq \mathcal{J}\}$.

The indexes of

- objects are $\{1, \dots, i, \dots, m\} = \mathcal{I}$, the split $\mathcal{I} = \mathcal{B}_1 \sqcup \dots \sqcup \mathcal{B}_K$;
- features are $\{1, \dots, j, \dots, n\} = \mathcal{J}$, the active set $\mathcal{A} \subseteq \mathcal{J}$.

Conclusion

1. The autoregressive forecasting technique appears to be effective in comparison to Singular Structiure Analysis and Neural Networks.
2. The analysis of multi-correlated features in the autoregressive matrix is absolutely a must; the Bayesian model selection could solve this problem.
3. The set of objects (periods of time series, time-segments) could be separated in several subsets and forecasted using several models; it boost the forecast precision.