Instituto Milenio
Sistemas Complejos de Ingeniería

Dynamic Data Mining for Improved Forecasting in Logistics and Supply Chain Management

José Guajardo, Richard Weber
Department of Industrial Engineering
University of Chile
{jguajard, rweber}@dii.uchile.cl

Support by Millenium Science Institute
"Complex Engineering Systems" (www.sistemasdeingenieria.cl)
and Fondecyt 1040926

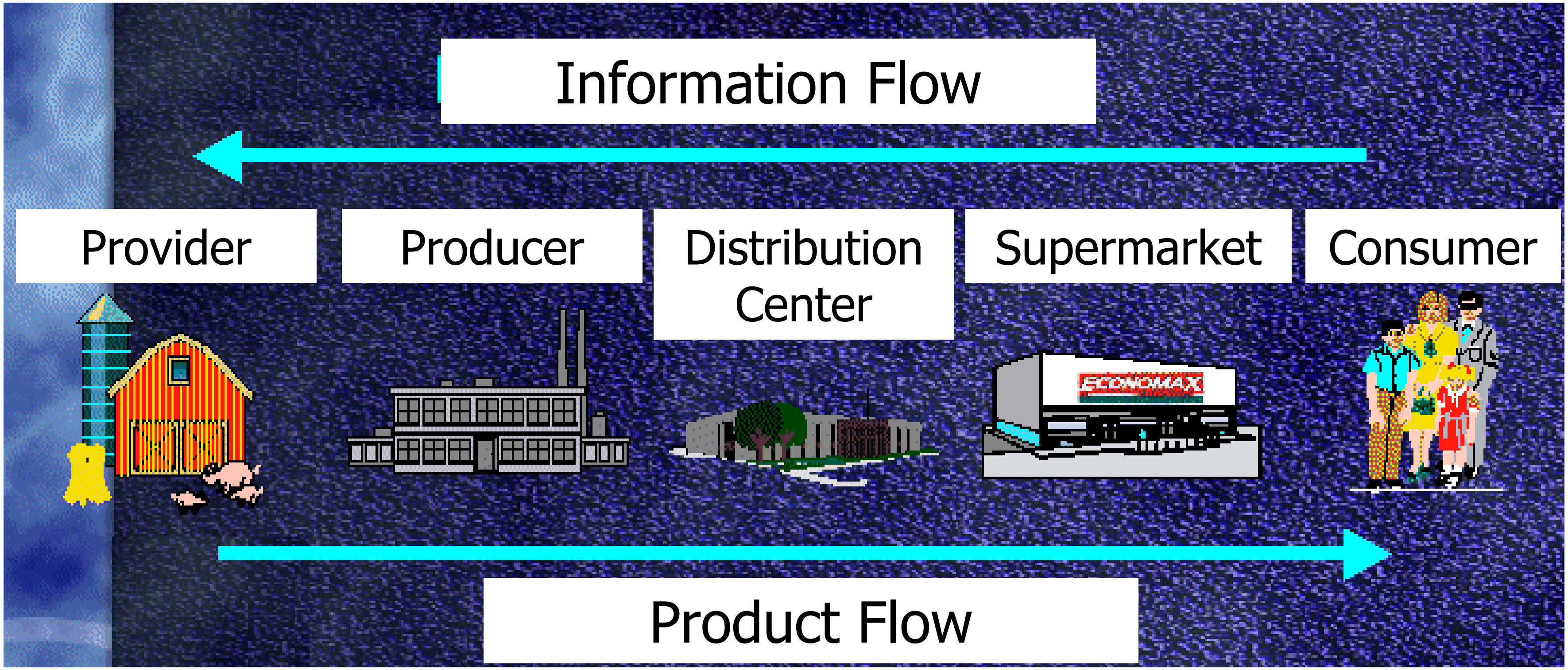
Outline

- Introduction

- Proposed Methodology
  - Forecasting approaches
  - Regression models
  - Proposed methodology using SVR
  - Application to real data set
  - Results

- Future work

Forecasting in Supply Chains



Forecasting approaches

Forecasting

- Time Series Analysis
  - Exponential smoothing
  - ARIMA models
- Regression approaches
  - Linear regression
  - Neural networks
  - Support Vector Regression

SVM – Basics

Statistical Learning Theory (Vapnik)

CLASSIFICATION	REGRESSION
-Fraud detection -Churn prediction -Risk assessment -… … …	-Demand forecasting -Stock price forecasting -Regression in biotechnology -… … …

Regression model

$$Y_t = f\,(Y_{t-j}, \ldots, Y_{t-n}; X_{1t-i}, X_{1t-j}, \ldots; X_{2t-k}, X_{2t-h} \ldots; \ldots)$$

y_t: analyzed time series

X: External variables

SVR: General model description

Data $(x_1, y_1), \ldots, (x_n, y_n), x_i \in R^m, y \in R$

Target $f(x) = (\omega \cdot \phi(x)) + b$

Objective function: Minimize $R_{reg}(f) = C\sum_{i=1}^{n} L(f(x_i) - y_i) + \frac{1}{2}(\omega \cdot \omega)$

Loss function

$\frac{1}{\text{Generalization ability}}$

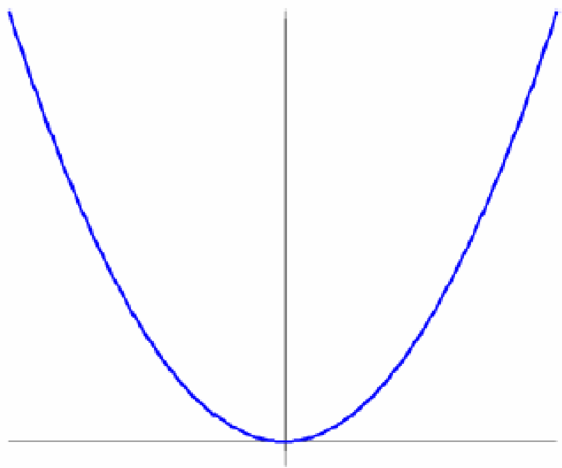
SVR PARAMETERS

- Loss function
- Cost of errors (C)
- Kernel function ($\Phi(x_i) \cdot \Phi(x)$)

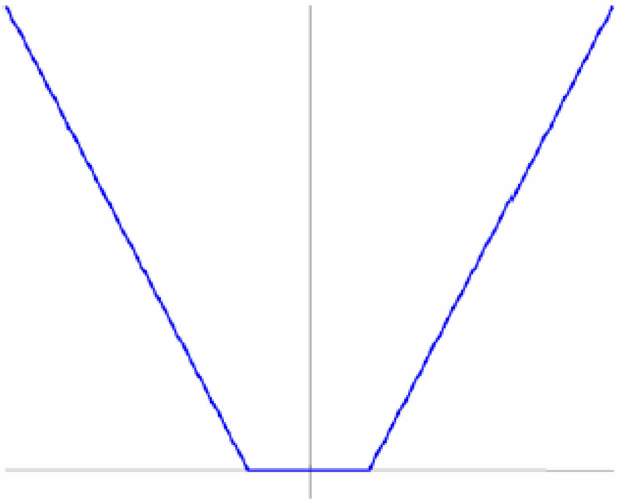
Loss functions

Cuadratic : $L(y) = (f(x) - y)^2$

ε-Insensitive: $$L(y) = \begin{cases} 0, & if \left| f(x) - y \right| < \varepsilon \\ \left| f(x) - y \right| - \varepsilon, & \text{else} \end{cases}$$

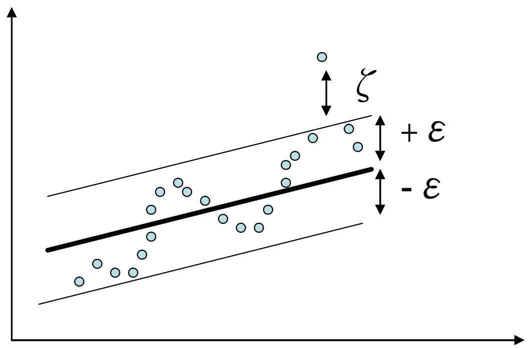


(a) Quadratic



(d) ϵ-insensitive

Loss function ε-insensitive



Trade Off: ε

- larger ε → + Generalization, - exactness
- smaller ε → - Generalization, + exactness

SVR Model: Cost of errors (C)

$$R_{reg}(f) = C\sum_{i=1}^{l} L(f(x_i) - y_i) + \frac{1}{2}(\omega \cdot \omega)$$

Trade Off: C

- larger C → - Generalization, + exactness
- Smaller C → + Generalization, - exactness

SVR Model: ε-Insensitive loss function

OPTIMIZATION PROBLEM

Min $\frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$

s.t. $\begin{cases} y_i - f(\mathbf{x}_i,\omega) \le \varepsilon + \xi_i^* \\ f(\mathbf{x}_i,\omega) - y_i \le \varepsilon + \xi_i \\ \xi_i, \xi_i^* \ge 0, i = 1,\dots,n \end{cases}$

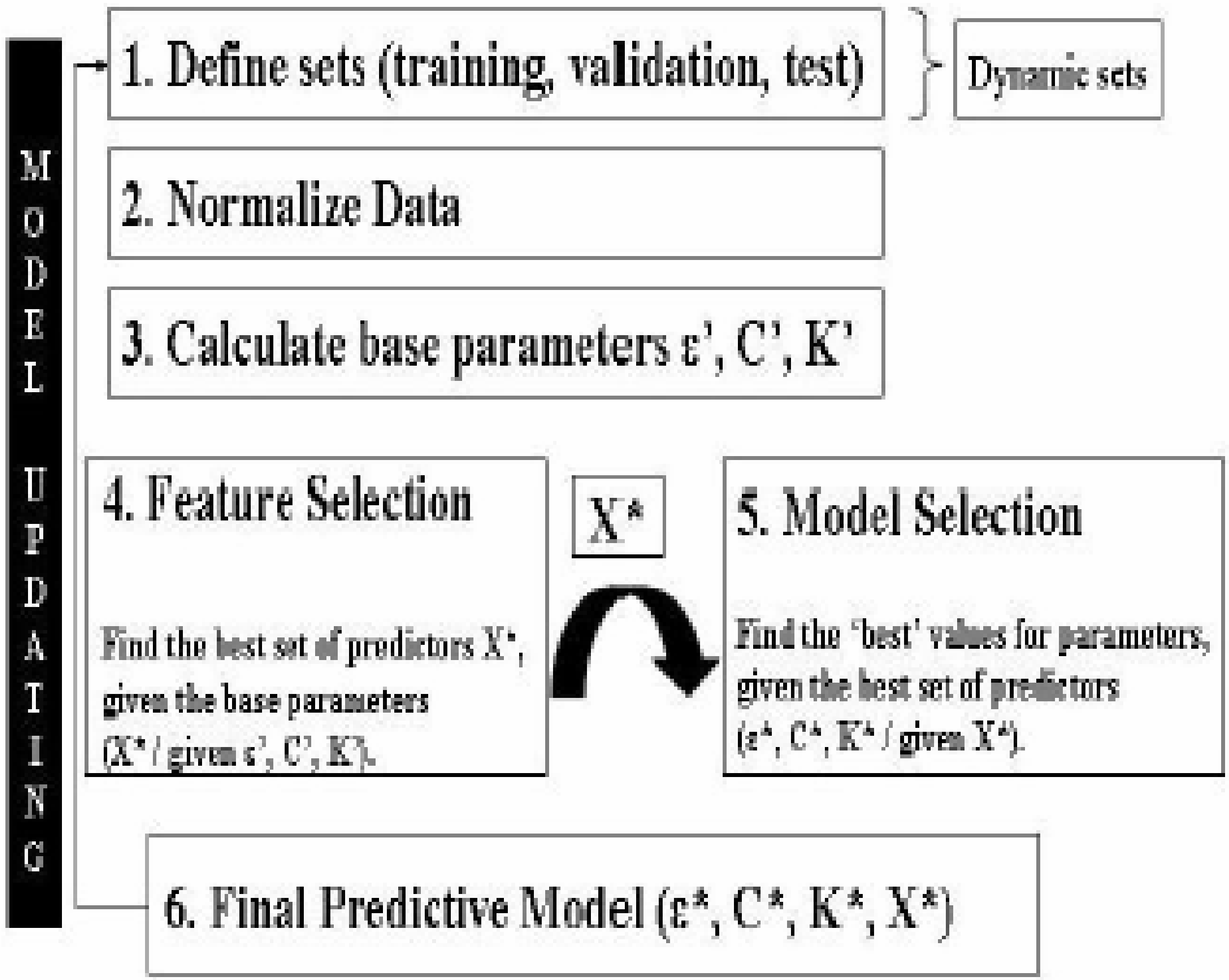
SOLUTION

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i{*})K(x_i, x) + b$$

$$b = average_k\{\delta_k + y_k - \sum_i(\alpha_i - \alpha_i{*})K(x_i, x_k)\}$$

$$\delta_k = \varepsilon * sign(\alpha_k - \alpha_k{*})$$

Proposed Methodology: Dynamic Data Mining for Forecasting



Step 1: Define Sets (Training, Validation, Test)

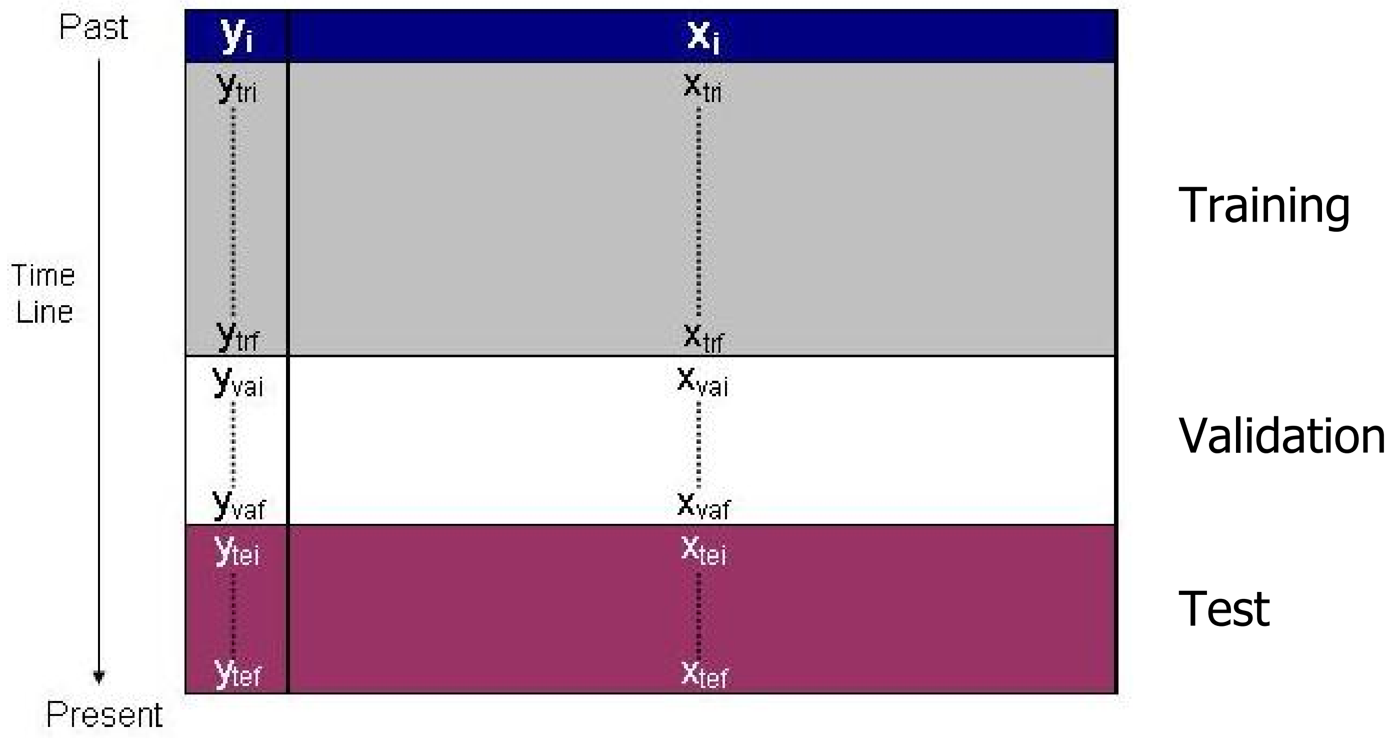


Figure 2: Static Configuration

Step 3: Calculate base parameters

Base Parameters

$$C' = |\bar{y}| + |3\sigma_y|$$

$$\varepsilon' = 3 * \sigma / \sqrt{\ln(n) / n}$$

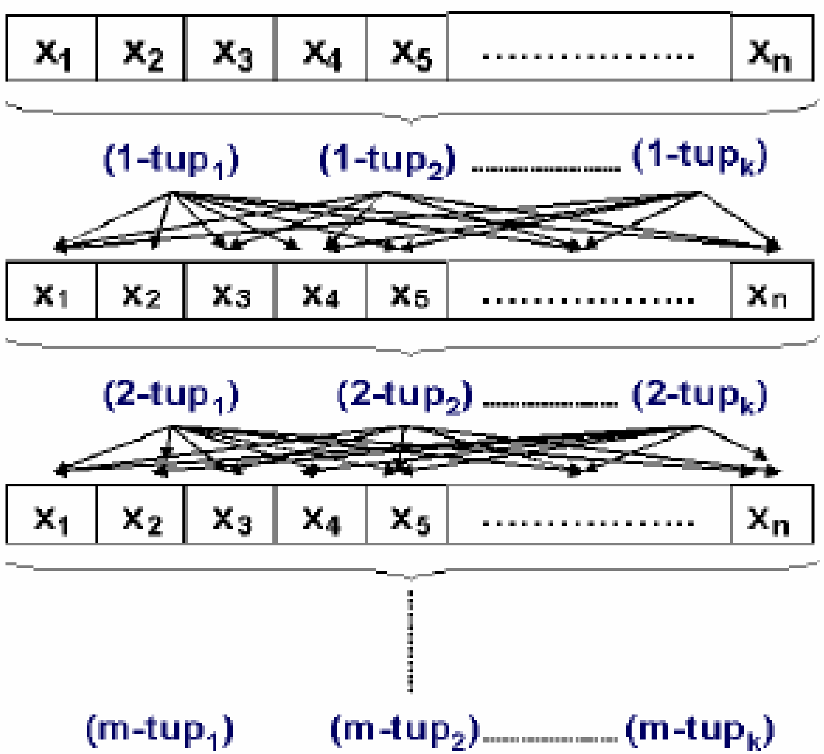
$$K' = \text{Experiments}$$

As proposed in:
V. Cherkassky and Y. Ma.: Practical selection of SVM parameters and noise estimation for SVM regression. Neural Networks 17(1):113-126, 2004.

Step 4: Feature Selection

3 approaches for feature selection: Filter, wrapper, embedded methods

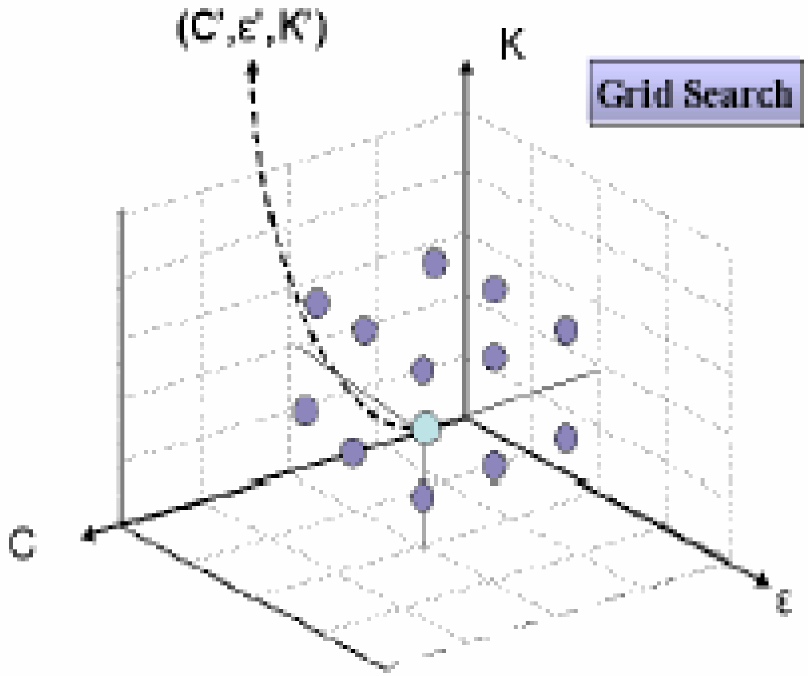
Here: wrapper using initial model with base parameters



Kohavi, R., John, G.H. (1997): Wrappers for feature subset selection. Artificial Intelligence 97 (1-2), 273-324

Step 5: Find best model

Grid search around base parameters using selected features



Problem description

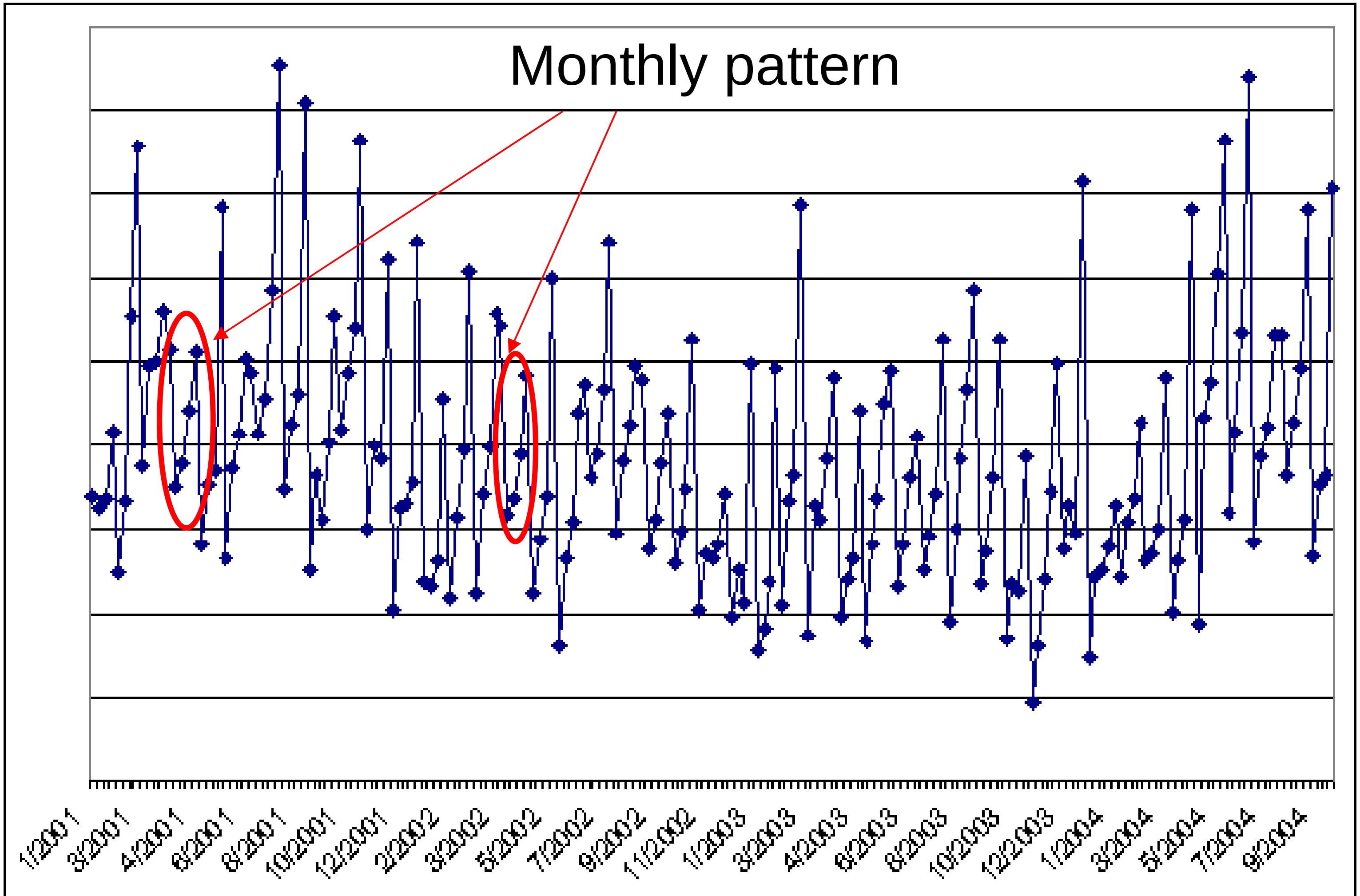
- US-company
  - Products
  - States

Sales Forecasting
[Product, State]

- Data sources
  - Web-based quotations
  - Stock levels
  - Reservations
  - … … …

Sales

Time Series: weekly sales during 2 years



Variables used (m=23)

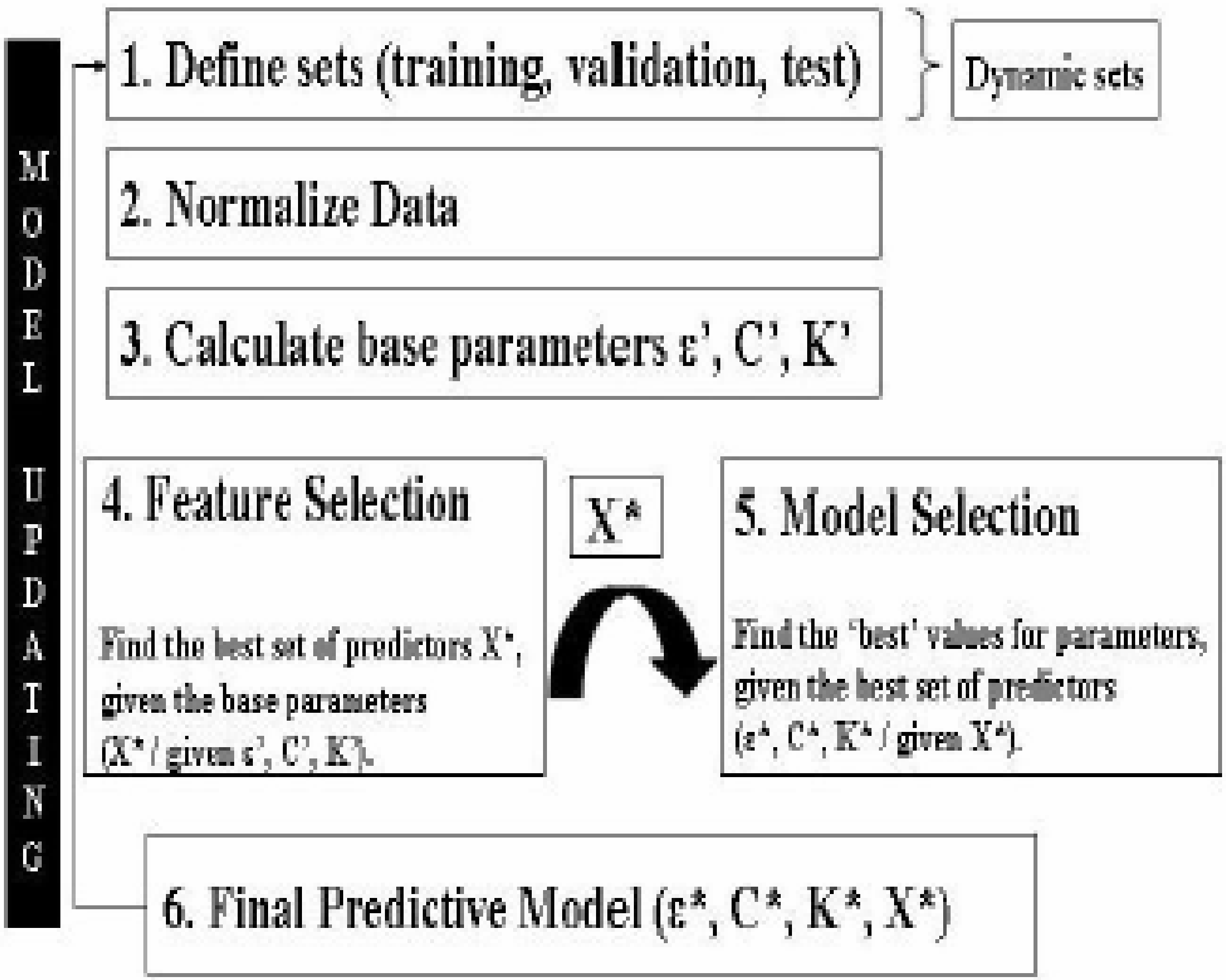
- Sales in 14 weeks previous to prediction (y_{t-j})
- Normalized number of week within a month
- Binary variable indicating if the month under consideration has 4 or 5 weeks
- Categorical variable indicating if the week under consideration contains holidays of certain categories
- Ordinal variable indicating the year under consideration (2001, ..., 2004)

- Number of the week under consideration
- Number of month within a quarter (taking values 1,2,3)
- Number of week within a year (taking values 1,... , 52)
- Number of month within a year (taking values 1,... , 12)
- Number of quarter within a year (taking values 1,... , 4)

Variable ranking after feature selection

- Normalized number of week within a month
- Sales one week before
- Sales two weeks before
- Binary variable indicating if the month under consideration has 4 or 5 weeks
- Sales eight weeks before
- Categorical variable indicating if the week under consideration contains holidays of certain categories
- Sales 13 weeks before
- Sales 14 weeks before
- Ordinal variable indicating the year under consideration (2001, ..., 2004)
- Sales seven weeks before
- Sales twelve weeks before
- Number of the week under consideration
- Number of month within a quarter (taking values 1,2,3)
- Sales three weeks before
- Sales six weeks before
- Sales ten weeks before
- Sales four weeks before
- Sales five weeks before
- Number of week within a year (taking values 1,... , 52)
- Number of month within a year (taking values 1,... , 12)
- Sales nine weeks before
- Number of quarter within a year (taking values 1,... , 4)
- Sales eleven weeks before

Proposed Methodology



Model Updating (First Cycle)

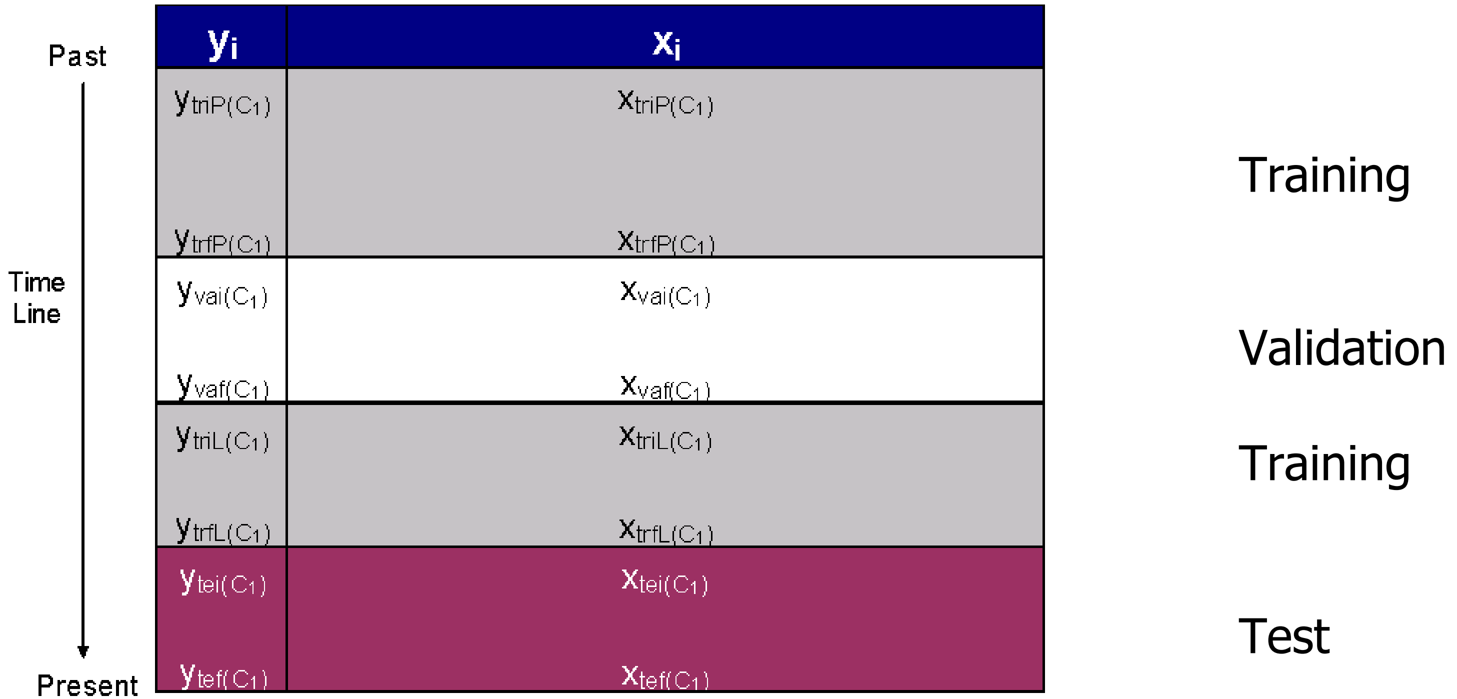


Figure 3: Predicting the first cycle (C1)

Model Updating (Second Cycle)

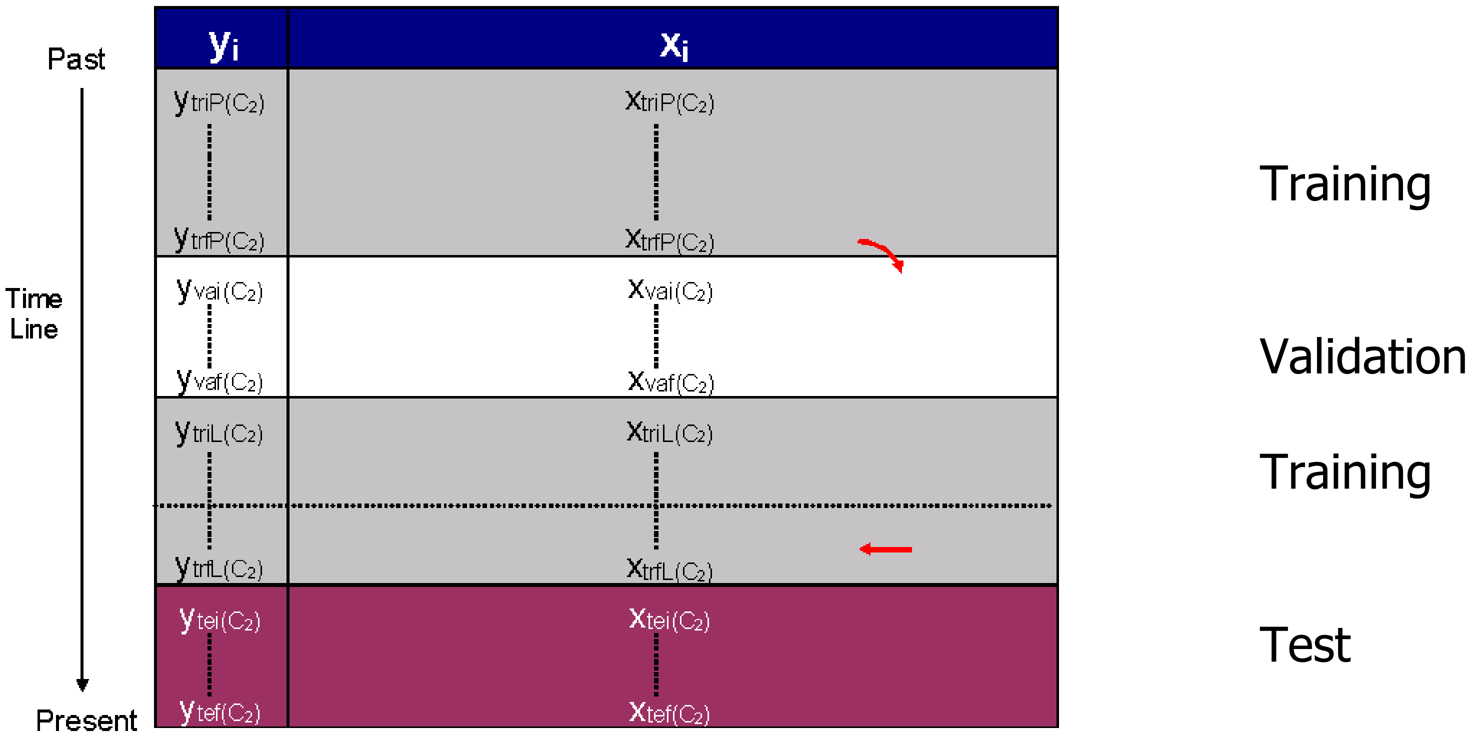


Figure 4: Predicting the second cycle (C2)

Results

5 different combinations product/state (P1, …, P5);
Training and Validation with data from January 2001 to March 2004;
Test with data from April 2004 to September 2004

Table 1: Mean absolute error (MAE) in test set (underlined: best result for each row).

Product	ARMAX	NN-UP	SVM-UP
P1	<u>292</u>	350	342
P2	347	368	<u>283</u>
P3	103	<u>89</u>	96
P4	<u>268</u>	288	284
P5	328	280	<u>264</u>
Average	275	275	<u>254</u>

Results

- Updating methodology improves forecasting results

- Proposed methodology includes dynamic feature selection thus provides interpretation of behavior.

Future work

- Further applications
- Feature selection “embedded” into SVR instead of wrapper approach
- Updating of alternative regression models
- Integration