Memory-Based Prediction in Control and Optimization

R. Kulhavý and P. Ivanova

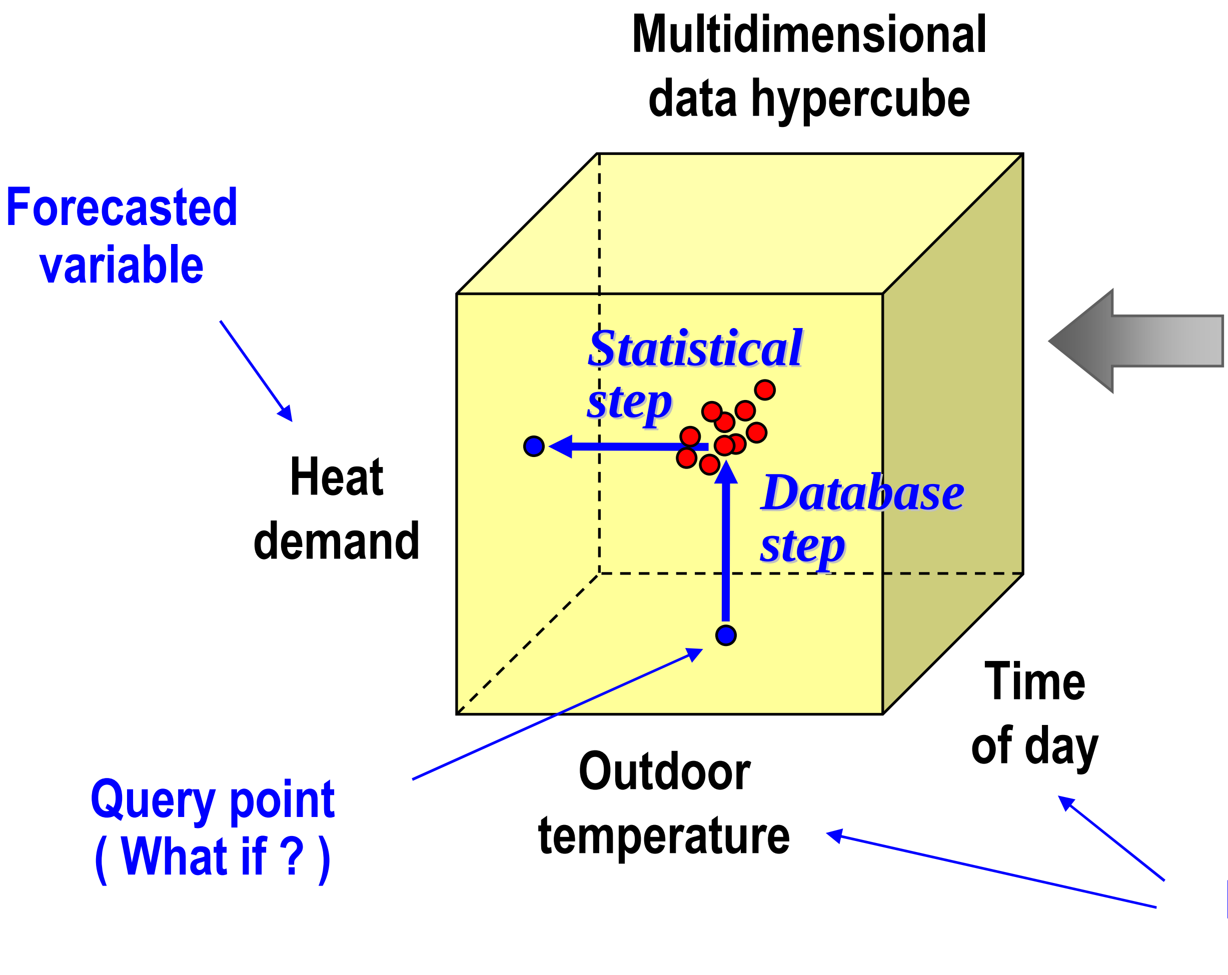
***Honeywell Technology Center Europe* and**

Institute of Information Theory and Automation

Outline

- Memory-based approach in a nutshell
- Locally weighted regression
- From classical to Bayesian perspective
- Regularization by prior information
- Practical aspects

Memory-Based Prediction in a Nutshell



Relational database

Microsoft Excel - Forecast.xls

	A	B	C	D	E	F
1						
2		Serial Time	Time of Day	Holiday Fuction	Outdoor Temper.	Steam Production
3		861.40	.39931	1	2.8434	480.15
4		861.41	.40972	1	2.9617	474.27
5		861.42	.42014	1	3.2116	460.8
6		861.43	.43056	1	3.5268	457.48
7		861.44	.44097	1	3.4794	461.82
8		861.45	.45139	1	3.441	465.45
9		861.46	.46181	1	3.4978	463.03
10		861.47	.47222	1	3.4317	465.11
11		861.48	.48264	1	3.4038	465.26
12		861.49	.49306	1	3.4367	464.33
13		861.50	.50347	.99653	3.5004	463.34
14		861.51	.51389	.98611	3.4643	465.09
15		861.52	.52431	.97569	3.3145	466.96
16		861.53	.53472	.96528	3.2887	471.44
17		861.55	.54514	.95486	3.3118	476.46
18		861.56	.55556	.94444	3.2966	473.80
19		861.57	.56597	.93403	3.2634	473.91
20		861.58	.57639	.92361	3.203	474.66
21		861.59	.58681	.91319	3.1758	467.91
22		861.60	.59722	.90278	3.2848	466.26
23		861.61	.60764	.89236	3.2977	464.83
24		861.62	.61806	.88194	3.1362	465.57
25		861.63	.62847	.87153	2.9414	470.77
26		861.64	.63889	.86111	2.8038	475.56

Locally Weighted Regression

- Forecasted variable x prognostic variables

$$y_k \times z_k = z(u^k, y^{k-1}), \quad k = 1, \ldots, N$$

- General regression model

$$y = f_\theta(z) + e, \quad \theta \in \Theta$$

- Locally weighted least-square fit

$$\min_\theta \frac{\sum_{k=1}^{N} K(\|z - z_k\|_H)\,(y_k - f_\theta(z_k))^2}{\sum_{k=1}^{N} K(\|z - z_k\|_H)}$$

- Kernel function

$K(x) = 1$ for $x = 0$ and decreasing to 0, e.g., $K(x) = \exp(-x^2)$

Popular Approach

- *locally-weighted smoothing*
  [Cleveland, 1978] [Cleveland, Devlin and Grosse, 1988]
- *non-parametric regression*
  [Härdle, 1990] [Hastie and Tibshirani, 1990]
- *local learning*
  [Bottou and Vapnik, 1992]
- *memory-based learning*
  [Schaal and Atkeson, 1994]
- *local learning*
  [Deng and Moore, 1994]
- *just-in-time estimation*
  [Cybenko, 1996] [Stenman *et al*, 1997]

Likelihood Based Inference

- **General regression**

$$y_k, z_k = z(u^k, y^{k-1}), \quad k = 1, \ldots, N$$

- **Model**

$$s_\theta(y|z), \quad \theta \in \Theta$$

- **Likelihood function**

$$l_N(\theta) = q_\theta(y^N, u^N) = c \prod_{k=1}^{N} s_\theta(y_k | z_k)$$

- **Posterior density**

$$p_N(\theta) \propto p_0(\theta)\, l_N(\theta)$$

Information Based Inference

- Empirical density

$$r_N(y,z) = \frac{1}{N} \sum_{k=m+1}^{N+m} \delta(y - y_k, z - z_k)$$

$$-\frac{1}{N} \log \prod_{k=m+1}^{N+m} s_\theta(y_k \mid z_k)$$

- Conditional inaccuracy

$$K(r_N : s_\theta) = \iint r_N(y,z) \log \frac{1}{s_\theta(y \mid z)} \, dy \, dz$$

- Likelihood

$$l_N(\theta) \propto \exp(-N K(r_N : s_\theta))$$

- Posterior density

$$p_N(\theta) \propto p_0(\theta) \exp(-N K(r_N : s_\theta))$$

‘Robust AR’ Example

$$y_k = \boxed{(\mu + v_k)}\,\boxed{y_{k-1}} + e_k$$

random fluctuation of AR(1) coefficient $= z_k$ *regressor*

μ is constant

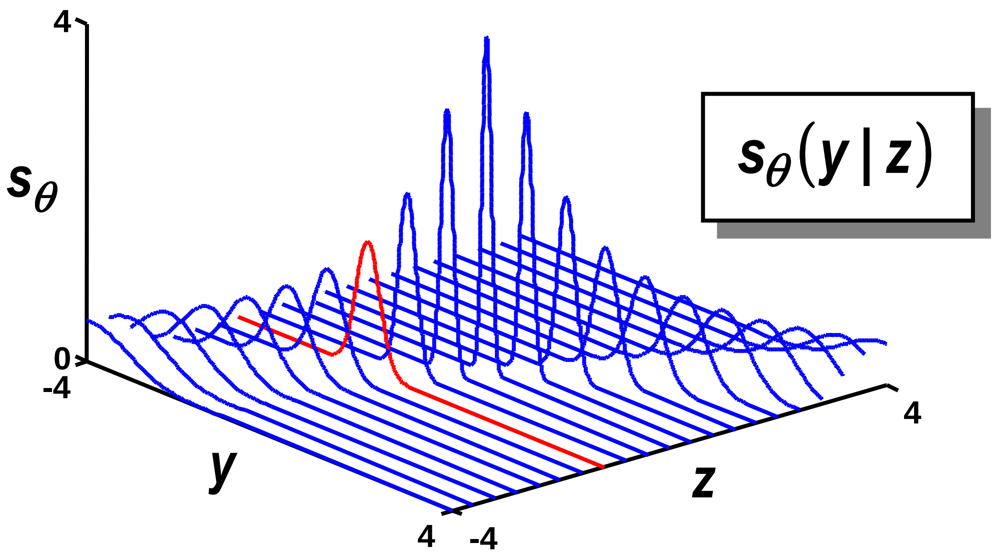
v_k is $N(0, \sigma_v^2)$ distributed

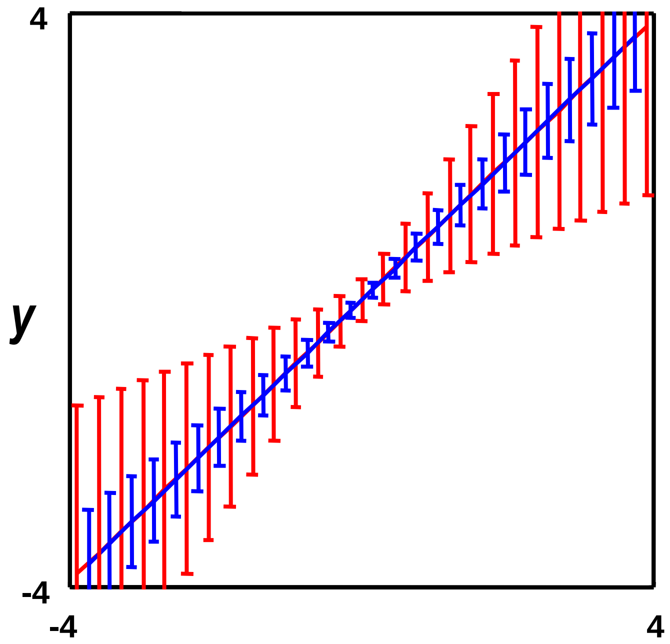
e_k is $N(0, \sigma_e^2)$ distributed

unknown parameters

$$\theta = (\mu, \sigma_e, \sigma_v)$$

Model Density

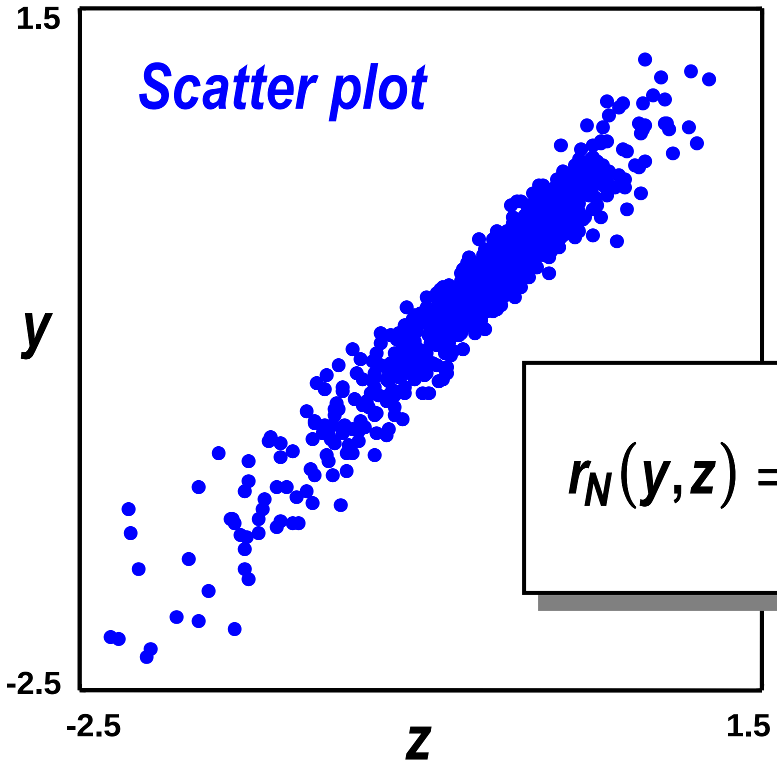




z-dependent variance: $\sigma^2(z) = \sigma_e^2 + z^2\,\sigma_v^2$

$$s_\theta(y\,|\,z) = \frac{1}{\sqrt{2\pi\sigma^2(z)}}\,\exp\left(-\frac{1}{2\sigma^2(z)}(y-\mu z)^2\right)$$

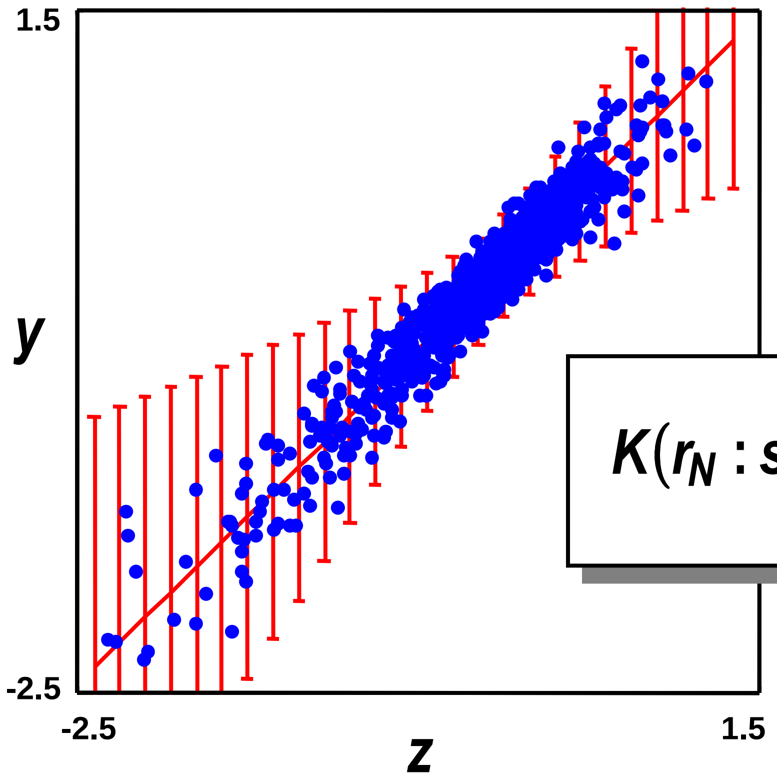
Empirical Density



Mixture of
Dirac functions

$$r_N(y,z) = \frac{1}{N}\sum_{k=1}^{N}\delta(y-y_k, z-z_k)$$

Information Measure



Conditional inaccuracy

$$K(r_N : s_\theta) = \iint r_N(y,z)\ \log \frac{1}{s_\theta(y\,|\,z)}\ \mathrm{d}y\,\mathrm{d}z$$

Conjugate Prior

$$p_0(\theta) \propto \exp(-\nu_0 K(\rho_0 : s_\theta))$$

$\nu_0 > 0$

Degree of belief in the prior

$\rho_0(y,z)$

"Prior" density of data

Posterior with Conjugate Prior

$$p_0(\theta) \propto \exp(-\nu_0 K(\rho_0 : s_\theta)) \quad \text{... prior}$$

$$l_N(\theta) \propto \exp(-N K(r_N : s_\theta)) \quad \text{... likelihood}$$

$$\nu_N = \nu_0 + N$$

$$p_N(\theta) \propto \exp(-\nu_N K(\rho_N : s_\theta)) \quad \text{... posterior}$$

$$\rho_N(y,z) = \frac{\nu_0}{\nu_N}\rho_0(y,z) + \frac{N}{\nu_N} r_N(y,z)$$

Bayesian Locally Weighted Regression

Query point

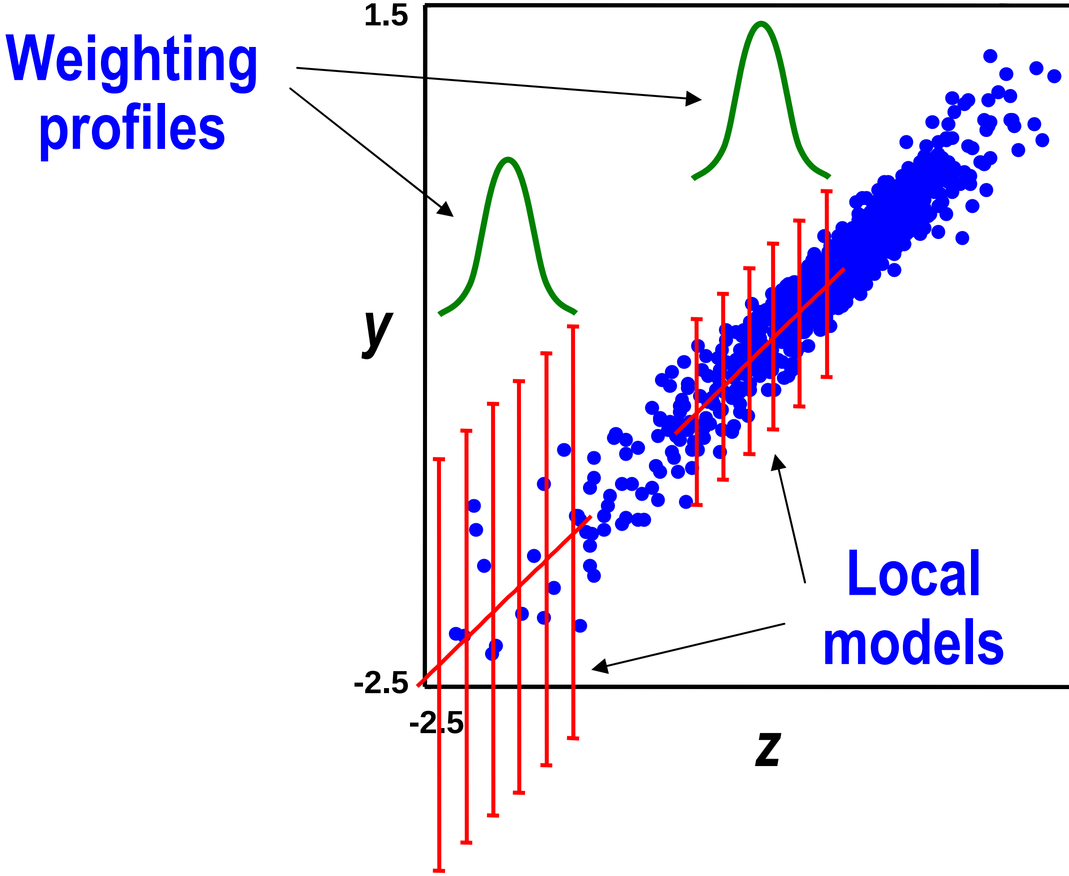
$$\nu_N = \nu_0 + \sum_{k=1}^{N} w_k$$

$$w_k = K(\| z^* - z_k \|_H)$$

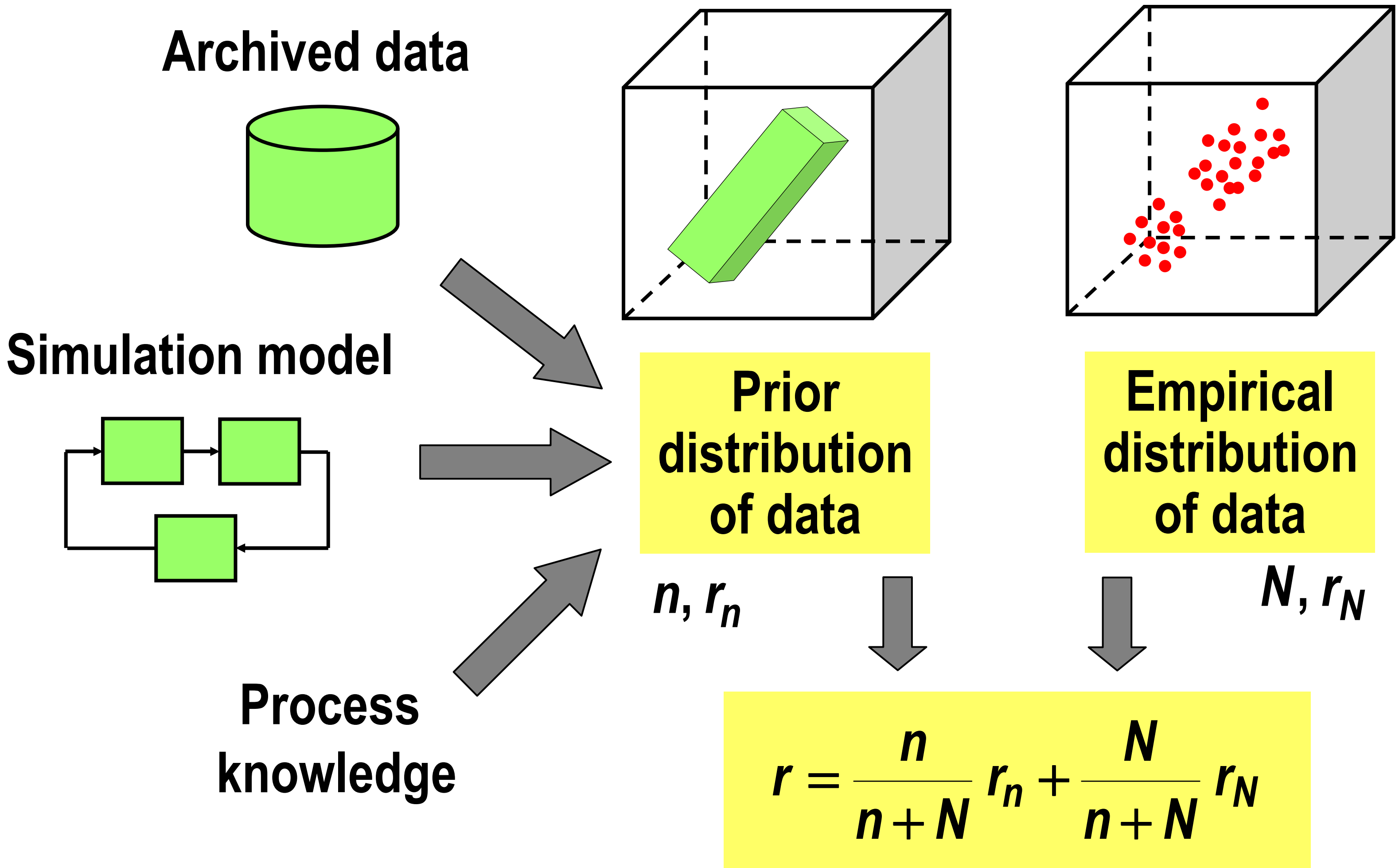
$$p_N(\theta) \propto \exp(-\nu_N K(\rho_N : s_\theta))$$

$$\rho_N(y,z) = \frac{\nu_0}{\nu_N}\rho_0(y,z) + \frac{\nu_N - \nu_0}{\nu_N}\frac{\sum_{k=1}^{N} w_k \delta(y - y_k, z - z_k)}{\sum_{k=1}^{N} w_k}$$

‘Robust AR’ Example

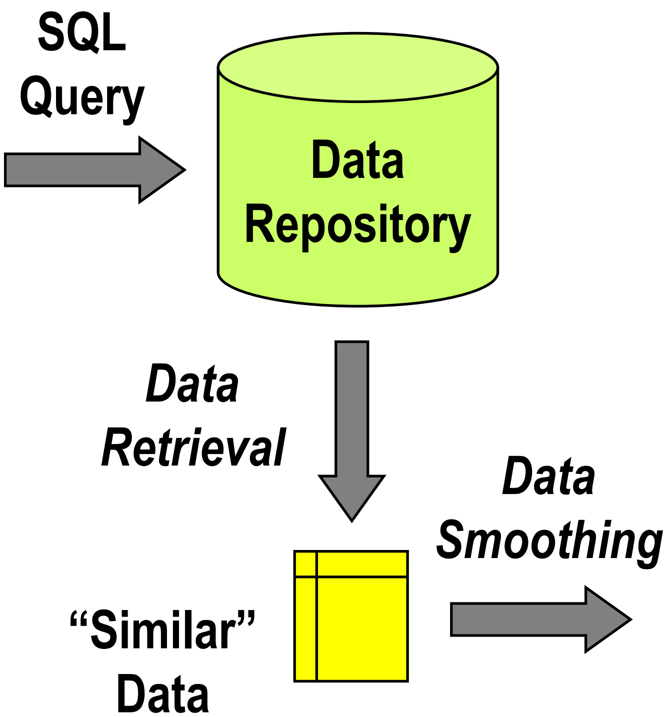


Use of Prior Knowledge

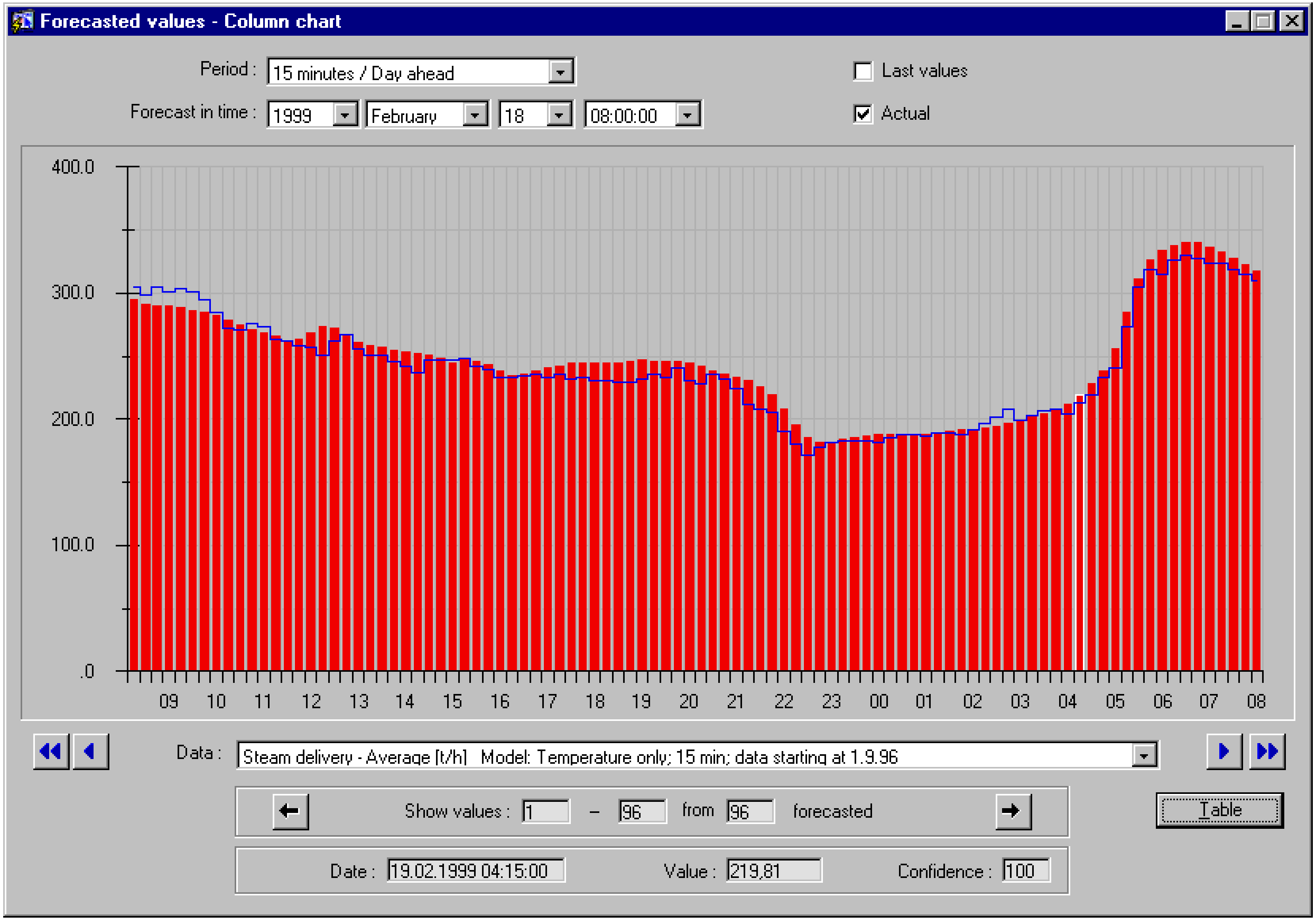


$$r = \frac{n}{n+N} r_n + \frac{N}{n+N} r_N$$

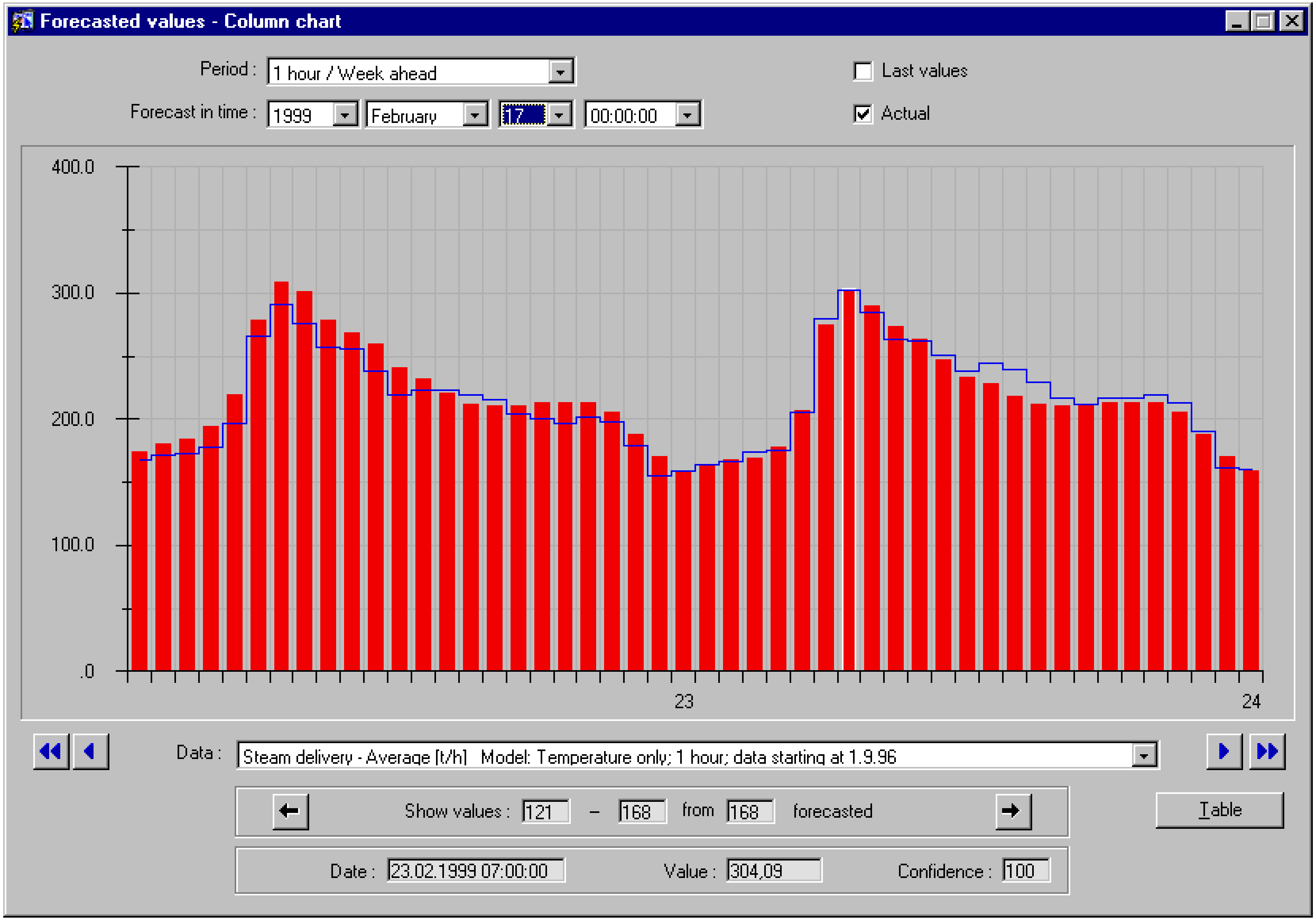
Practical Aspects



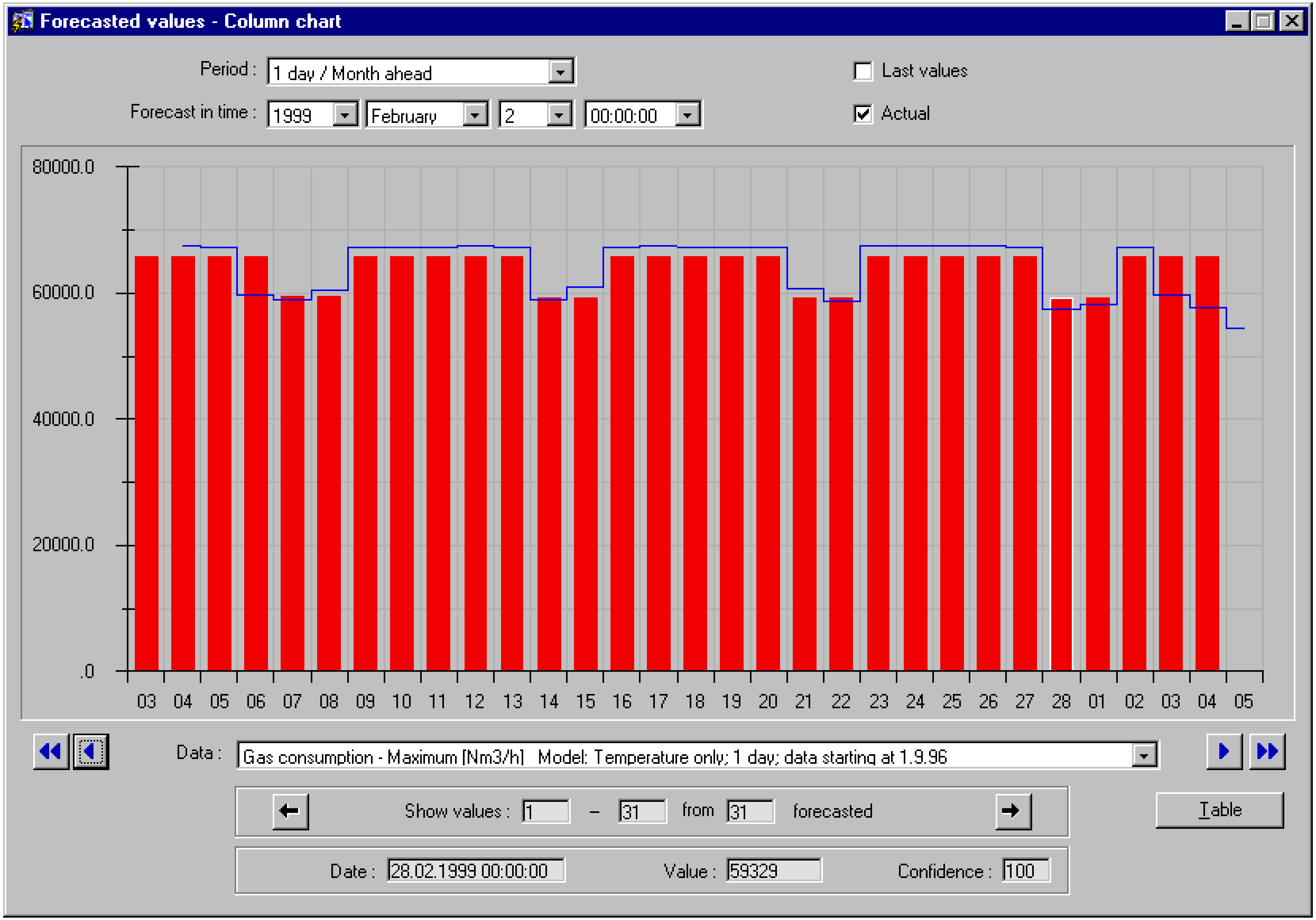
- The quality of data is crucial; validation of all stored data is a must.
- An operational database cannot guarantee the performance required; a data warehouse and/or data marts have to be built on top of it.
- Only cube-shaped neighborhoods can be used in the data retrieval step.
- Only a "soft" variant of the nearest neighbor method is feasible.
- *Dynamic* SQL may be required to ensure both performance and flexibility.



Forecasted values - Column chart
Period : 15 minutes / Day ahead
Forecast in time : 1999 February 18 08:00:00
Last values
Actual
400.0
300.0
200.0
100.0
.0
09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 00 01 02 03 04 05 06 07 08
Data : Steam delivery - Average [t/h] Model: Temperature only; 15 min; data starting at 1.9.96
Show values : 1 – 96 from 96 forecasted
Table
Date : 19.02.1999 04:15:00
Value : 219,81
Confidence : 100



Forecasted values - Column chart
Period :
1 hour / Week ahead
Last values
Forecast in time :
1999
February
17
00:00:00
Actual
400.0
300.0
200.0
100.0
.0
23
24
Data :
Steam delivery - Average [t/h]   Model: Temperature only; 1 hour; data starting at 1.9.96
Show values :
121
–
168
from
168
forecasted
Table
Date :
23.02.1999 07:00:00
Value :
304,09
Confidence :
100


Forecasted values - Column chart
Period : 1 day / Month ahead
Last values
Forecast in time : 1999 February 2 00:00:00
Actual
80000.0
60000.0
40000.0
20000.0
.0
03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 01 02 03 04 05
Data : Gas consumption - Maximum [Nm3/h] Model: Temperature only; 1 day; data starting at 1.9.96
Show values : 1 – 31 from 31 forecasted
Table
Date : 28.02.1999 00:00:00
Value : 59329
Confidence : 100

Three Approaches to Approximation

❶ **Empirical density**

$r_N(y,z)$

❷ **Model family**

$\{ s_\theta(y \mid z) : \theta \in \Theta \}$

❸ **Inaccuracy**

$K(r_N : s_\theta)$

- Replacement of the empirical distribution with its locally-weighted version
- Reduction of the model family to a finite model set or explicit approximation
- Restoration of the information divergence from compressed data

Memory-Based Versus Model-Based