University of Applied Sciences
MERSEBURG

A principle of structure selection for models of dynamic systems

Tatjana Lange
University of Applied Sciences Merseburg
Germany

Content:

University of Applied Sciences
MERSEBURG

1. Historical Background
2. Historical Terms
3. Setting up a task for the automatic selection of models
4. Problems arising when selecting models
5. Regularization operator for automatic structure selection and the estimation of parameters
6. LDUC – the Local Uncertainty Criterion
7. Summary and Outlook

Tatjana Lange

University of Applied Sciences
MERSEBURG

1. Historical Background

University of Applied Sciences
MERSEBURG

→ **on mathematical terms:**

- Stability of the solution
  → Hadamard J.: Sur les problèmes aux dérivées partielles et leur signification physique, Bull, Univ. Princeton, 1902
- Correct / Incorrect Setting up of Tasks
  → Hadamard J.: Le problème de Cauchy et les equation aux dérivées partielles linéaires hyperboliques, Hermann P., 1932
- Regularization Methods / Regularization Operator
  → Tichonoff A.N. (Тихонов, А.Н.): О регуларизации некорректно поставленных задач. ДАН СССР, 1963

→ **on the restoration of functional dependencies with the help of empirical data**

- Vapnik, V.N. (Вапник, В.Н.): Востановление зависимостей по эмпирическим данным. Москва, Наука, 1979

→ **on the application of the regularization method used for creating criteria for automatic model selection:**

→ Lange, T: Anwendung der Regularisierung bei der Bildung von Modellen in Polynomform nach MGEA. Wiss. Kolloquium Ilmenau, 1982

→ Lange, T: New Structure Criteria in Group Method of Data Handling. Kluwer Academic Publishers, 1994

2. Historical Terms

→ “Direct and Indirect” tasks

→ “Stability” and “Correctness” according to Hadamard

→ “Regularization” according to Tichonoff

Direct and Inverse Tasks:

Example: $Av = y$

whereby ν and *y* are vectors
A is an operator (e.g. a matrix)

Direct task:

given: A and ν
searched for: y

Indirect task:

given: A and vector y measured (maybe inexactly)
searched for: vector $v = A^{-1}y$

Operator A^{-1} may be unstable and discontinuous, respectively ! ?

How to find a quantitative solution for ?

$$v = R(y)$$

The stability of a quantitative solution and the correctness of the solution according to Hadamard.

University of Applied Sciences MERSEBURG

*) Finding the solution to a quantitative task usually involves determining a solution 'ν' from the given sets of data: y.

In other words: $\nu=R(y)$

**) Let us assume that y and ν are elements of the metric spaces Y and F. The distances between two elements y_1 and y_2 in the metric space Y, and two elements ν_1 and ν_2 in the metric space F, are:

$$\rho_Y(y_1, y_2) \text{ and } \rho_F(\nu_1, \nu_2)$$

Normally, metric is determined by practical conditions.

***) Let us assume that the term "solution" is defined and that every element (y Є Y) corresponds to the one and only solution $\nu=R(y)$ from the space F.

The task of searching for the solution $\nu = R(y)$ from the space F using the primary data y Є Y is denoted **stable** in the spaces F and Y if for any number, $\varepsilon > 0$, it is possible to find such a number $\delta(\varepsilon) > 0$ so that the inequation

$$\rho_Y(y_1, y_2) \leq \delta(\varepsilon)$$

follows, so that

$$\rho_F(\nu_1, \nu_2) \leq \varepsilon$$

whereby ν_1, ν_2 Є Y and $\nu_1 = R(y_1)$, $\nu_2 = R(y_2)$.

The task of determining a solution ν, ν Є F, from the space F, using the primary data, y Є Y , is regarded as **correct** for the pair of metric spaces F and Y if the following conditions are met:

1) For any element, y Є Y, there is a solution ν from the space F.
2) The solution is clearly determined.
3) The solution is stable in spaces F and Y.

Points 1 and 2 characterize mathematical certainty and point 3 is connected to physical determinism and the possibility of the application of numerical methods for solving the task from inexact data.

Regularization Method according to Tichonoff:

$$R[y, A, v] = \rho_y^2 (Av, y) + \alpha \cdot \Omega(v)$$

$$\rho_Y \text{ - distance in Y (metric)}$$

will be minimized.

Whereby

$\Omega(\nu)$ - stabilisation operator

metric ρ and operator $\Omega(\nu)$ are determined by practical conditions.

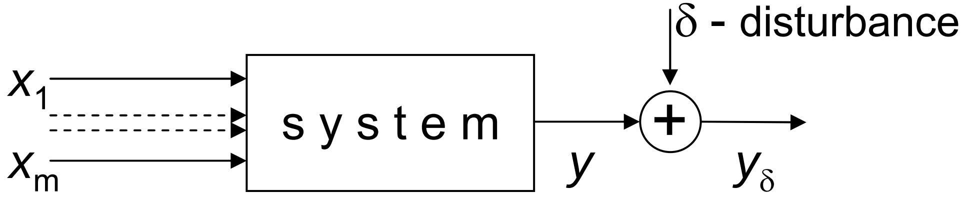
University of Applied Sciences
MERSEBURG

3. Setting up a task for the automatic selection of models

Tatjana Lange

Selection of Model from Measured Data

1. System:



2. Data table (data set):

	$x_0 \equiv 1$	x_1	x_2	...	x_j	...	x_m	y_δ
1				...		...		
2				...		...		
.				.		.		
i				.		.		
.				.		.		
l				...		...		

3. Setting up of Model (Kolmogorov-Gabor polynomial):

$$y = a_0 + \sum_k a_k x_k + \sum_k \sum_j a_{kj} x_k x_j + \ldots + \sum_k \ldots \sum_q a_{k\ldots q} x_k \ldots x_q$$

m - stands for the number of input values
(M+1) - stands for the number of coefficients to be estimated
K - stands for the maximum degree of polynomial K=(s+t+…+u)

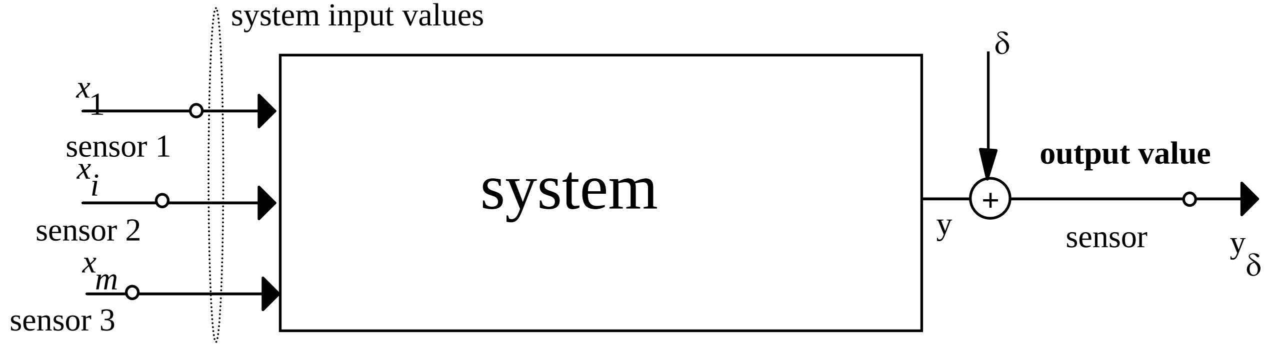
<u>The following values have to be found:</u>

$\vec{a}*$ - the optimal value of the coefficient vector $\vec{a}* = (a_0, a_1, \ldots, a_M)$

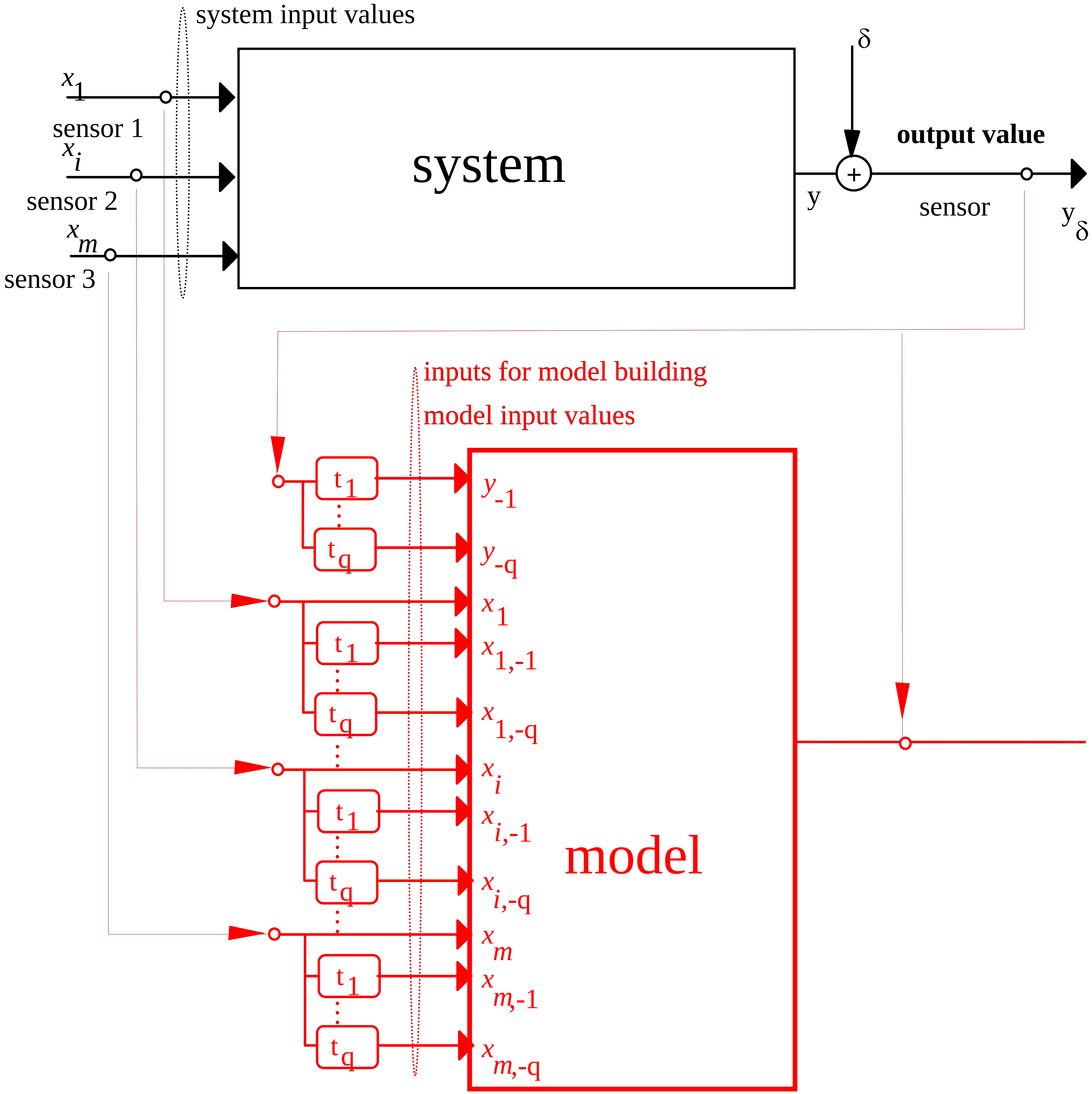
$M*, K*$ - the number of relevant inputs M* and the polynomial's optimal power K*

The task of structure modelling

Connection between System Interfaces and Model Interfaces



Connection between System Interfaces and Model Interfaces



First Group:

Static characteristic:

$$y = f(x_1, x_2, \ldots, x_m)$$

Second Group:

Dynamic characteristic:

$$y_k = f(y_{k-1}, y_{k-2}, \ldots, y_{k-q}, x_{1,k}, x_{1,k-1}, \ldots, x_{1,k-q}, x_{2,k-1}, \ldots, x_{2,k-q}, \ldots, x_{m,k-1}, \ldots, x_{m,k-q})$$

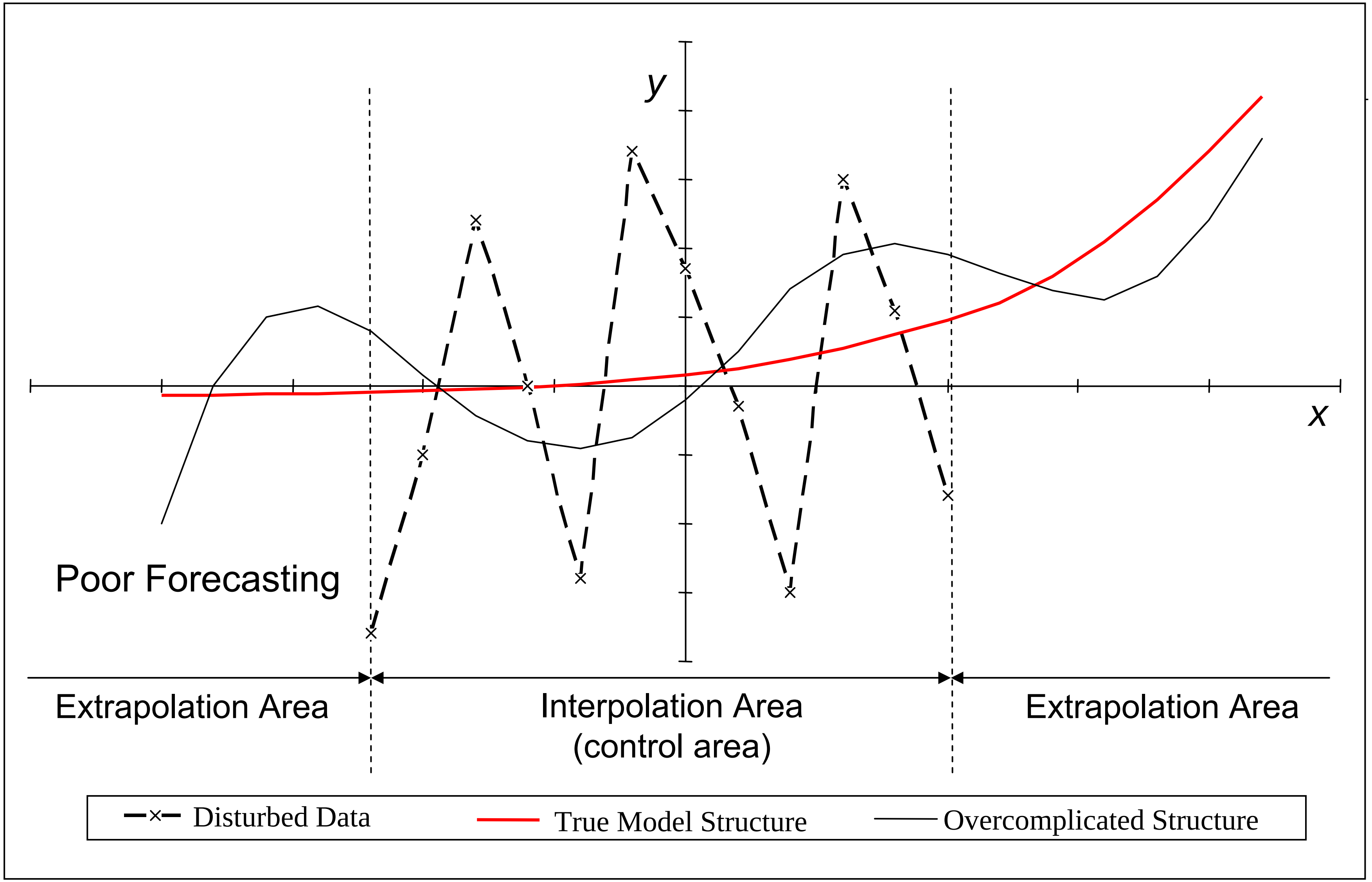
Third Group:

Autoregression:

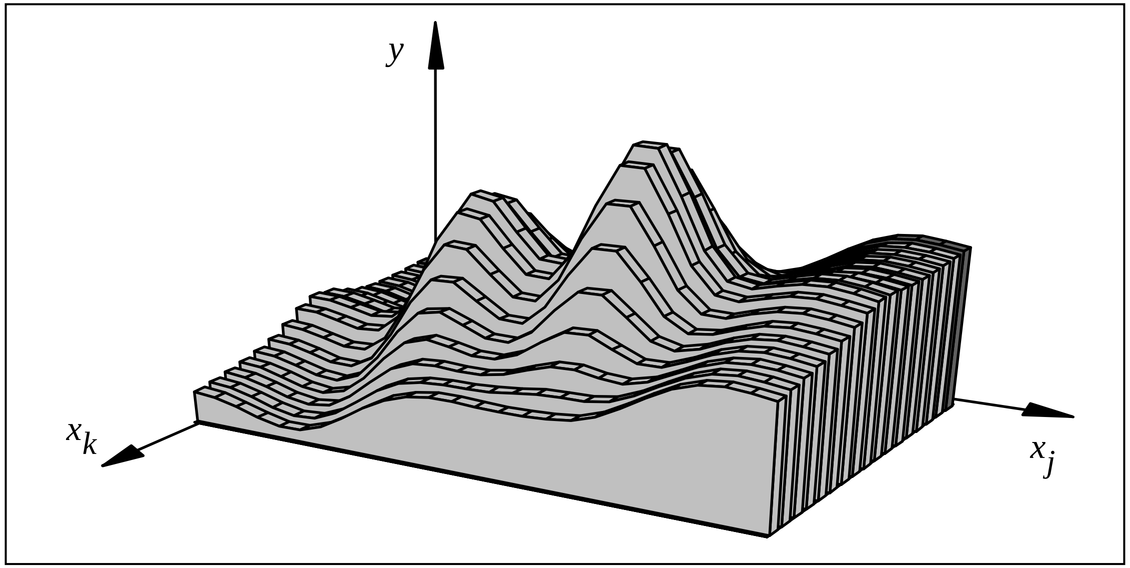
$$y_k = f(y_{k-1}, y_{k-2}, \ldots, y_{k-q})$$

University of Applied Sciences
MERSEBURG

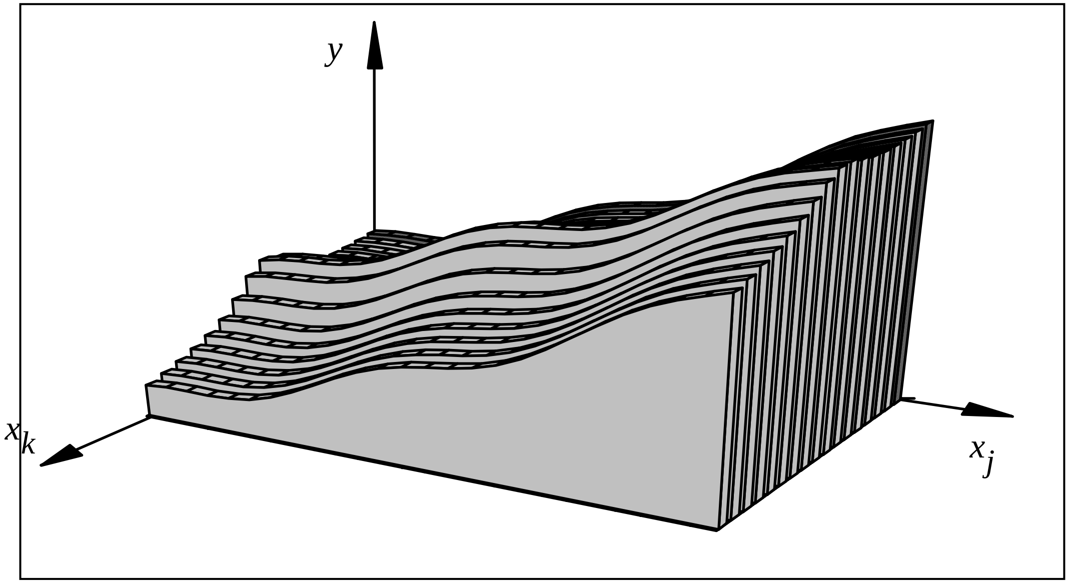
4. Problems arising when selecting models



Comparison of a True and an Overcomplicated Model Structure



Regression Surface



Example of a Regression Surface in the case of a simplified structure

University of Applied Sciences MERSEBURG

Table of Possible Solutions

	$v_1 = \hat{a}_1$	...	$v_j = \hat{a}_i$	...	$v_k = \hat{a}_k$	...	$v_L = \hat{a}_L$	$\vec{z}$
$\hat{a}_0$	2.43							1
$\hat{a}_1$	0							x_1
. .	. .							. .
$\hat{a}_j$	1,7							x_r^j
$\hat{a}_{j+1}$	0							.
. . .	. . .							. $x_m^2 x_k^3$.
$\hat{a}_M$	0,5							x_m^K
Q_I corresponds to data set $\vec{y}_I$	Q_{I_1}	...	$Q_{I_i}^*$	...	Q_{I_k}	...	Q_{I_L}	
Q_{II} corresponds to data set $\vec{y}_{II}$	Q_{II_1}	...	Q_{II_i}	...	$Q_{II_k}^*$	...	Q_{II_L}	

Discreet Solution to Structure Selection and Stability according to Hadamard.

Here are

$v_1, v_2, ..., v_L$ - different possible solutions in polynomial which are determined by vectors of the coefficients

$\vec{z}$ - a design vector

$Q_1, Q_2, ..., Q_L$ - the values of the criterion for structure selection

In the case of a Euclid Metric the distance between the solutions (models) is determined as follows:

$$\rho(\nu_i, \nu_k) = \sqrt{\sum_{j=1}^{M} (\hat{a}_{ij} - \hat{a}_{kj})^2}$$

or for two different data sets:

$$\rho(\nu_I, \nu_{II}) = \sqrt{\sum_{j=1}^{M} (\hat{a}_{Ij} - \hat{a}_{IIj})^2}$$

University of Applied Sciences
MERSEBURG

Penalty is necessary for overcomplicated model structures !

Main criterion:	Penalty criterion:
$Q_1 = \rho_y(v_{Measured}, \hat{v}) \to \min$ e.g.: $Q_1 = \sum_{i=1}^{l}(y_i - \hat{a}_0 - \hat{a}_1 x_{1i} - \hat{a}_2 x_{2i} ......)^2 \to \min$	$Q_2 = \rho_F(v_{Measured}, \hat{v}) \to \min$ e.g.: $Q_1 = \sum_{i=1}^{M} \hat{a}_i^2 \to \min$
• Improves the Interpolation Capabilities of the Model	• Improves the Forecast Capabilities of the Model
• May lead to Overcomplicated Model Structure	• May lead to Simplified Model Structure
• Reduces the Dispersion of the Model	• Reduces the Bias of the estimated result
• Improves the Convergence of the Model according to the Output	• Improves the Convergence of the Model according to the Coefficients
• Improves the Efficiency of Estimation	• Improves the Robustness of Estimation
• Corresponds to the Minimization of the Error of First Kind	•Corresponds to the Minimization of the Error of Second Kind
• If the task is assigned **correctly** according to Hadamard: → Q_1 can be used on its own.	• If the task is assigned **incorrectly** according to Hadamard: → The stabilizing penalty functional Q_2 should be introduced. → Q_2 cannot be used without Q_1

Comparison of Main and Penalty Criteria

5. Introduction of a Regularization operator for automatic structure selection and the estimation of parameters

(1) With regard to the output *y*, two models $\mathbf{M_I}$ and $\mathbf{M_{II}}$ are near to each other (that means $\rho_{R^l}(M_I, M_{II}) \le \varepsilon$, where ε is small). This corresponds to the convergence according to the output, when

$$\rho_l(y_I, y_{II}) \le \gamma \quad \text{and } \gamma \text{ is small,}$$

where y_I and y_{II} are output vectors of the models $\mathbf{M_I}$ and $\mathbf{M_{II}}$.

If ρ_l represents an *l*-dimensional Euclid Metric, then $\rho_l(y_I, y_{II}) \le \gamma$ can be replaced by the equivalent postulation

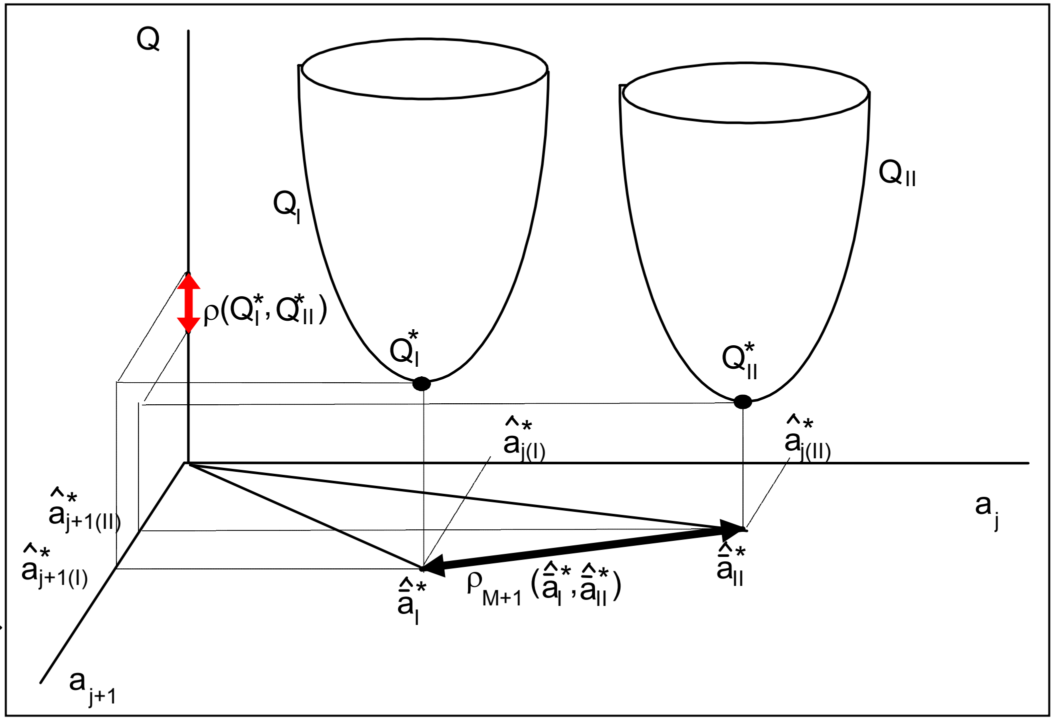
$$\rho_l(Q_I^*, Q_{II}^*) \le \gamma$$

where

Q_I and Q_{II} are functionals (criteria) for estimating parameters by means of Least Square Methods

Q_I^* and Q_{II}^* are the values of *RSS* of the first and second model.

Proximity of Models according to the output *y*



Distance between models estimated on different data sets

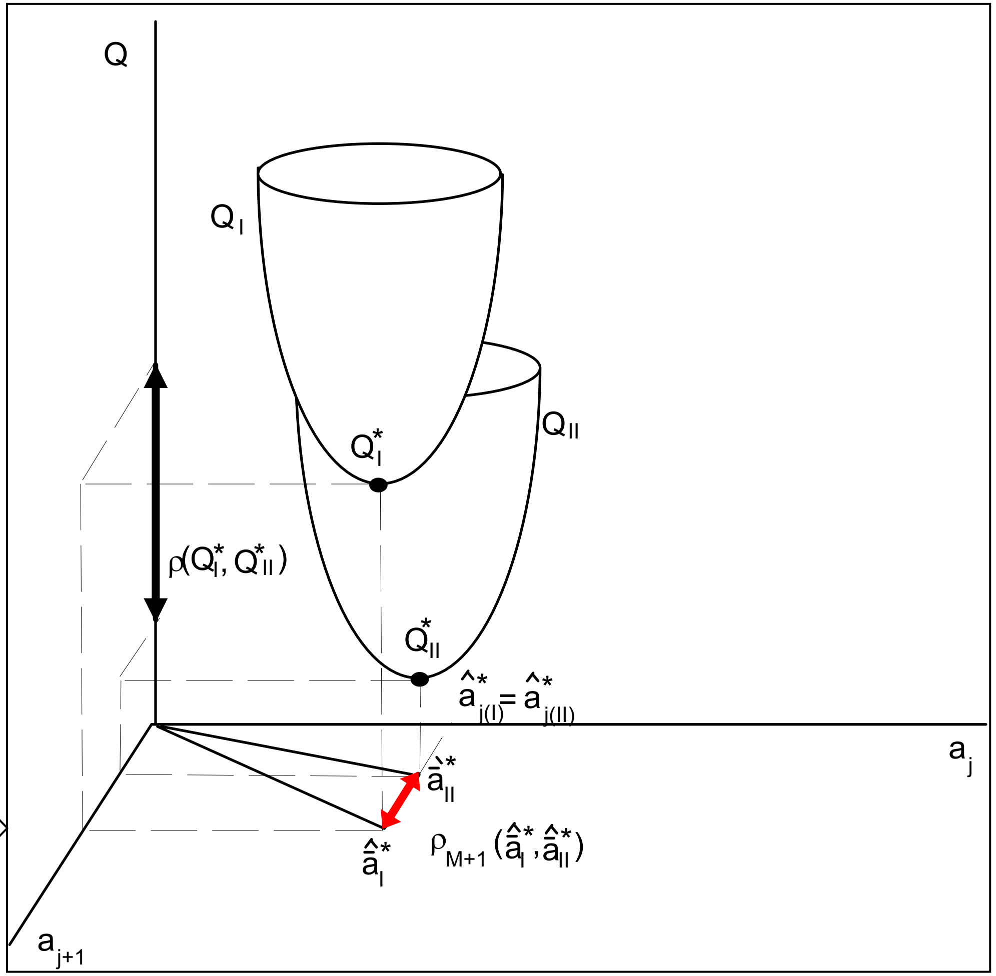
(2) Two models $\mathbf{M_I}$ and $\mathbf{M_{II}}$ are near to each other with regard to the coefficient vector $\vec{a}$ (that means $\rho_{R^l}(M_I, M_{II}) \le \varepsilon$, where ε is small). This corresponds to the convergence according to the parameters, when

$$\rho_{M+1}\left(\hat{\vec{a}}^*_{(I)}, \hat{\vec{a}}^*_{(II)}\right) \le \xi$$ and ξ is small,

where $\hat{\vec{a}}^*_I$ and $\hat{\vec{a}}^*_{II}$ are the vectors of the estimated parameters of the models $\mathbf{M_I}$ and $\mathbf{M_{II}}$.

$\hat{\vec{a}}^*_I$ and $\hat{\vec{a}}^*_{II}$ correspond to the estimated values of the parameter functional.

Proximity of Models according to the coefficient vector $\vec{a}$

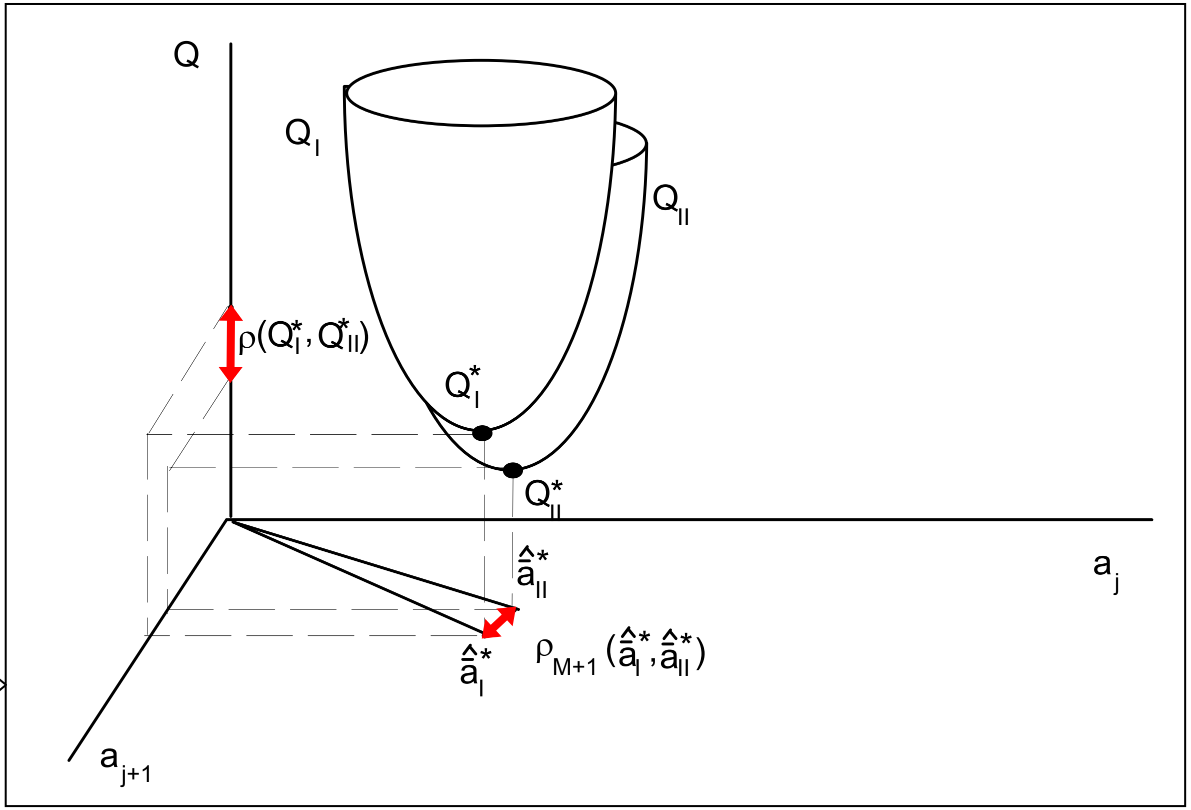


Distance between Models estimated on different data sets

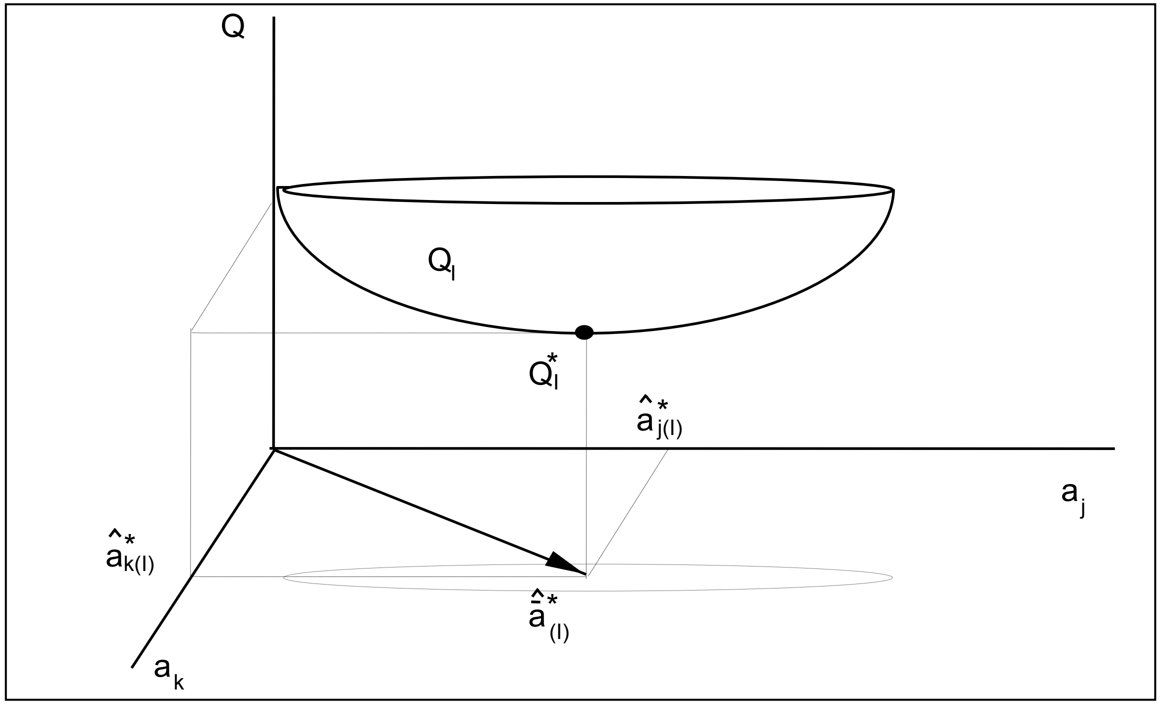
(3) The two models are really near to each other if the proximity of outputs *y* and parameters $\vec{a}$ is guaranteed. The distance between two models is defined by the sum of distances between outputs and parameters:

$$\rho_l(y_I, y_{II}) < \gamma \text{ and } \rho_{M+1}(\hat{\vec{a}}_I^*, \hat{\vec{a}}_{II}^*) < \xi \quad \Leftrightarrow \quad \rho_{R^l}(\mathrm{M}_I, \mathrm{M}_{II}) < \varepsilon$$

Complete Proximity of Models

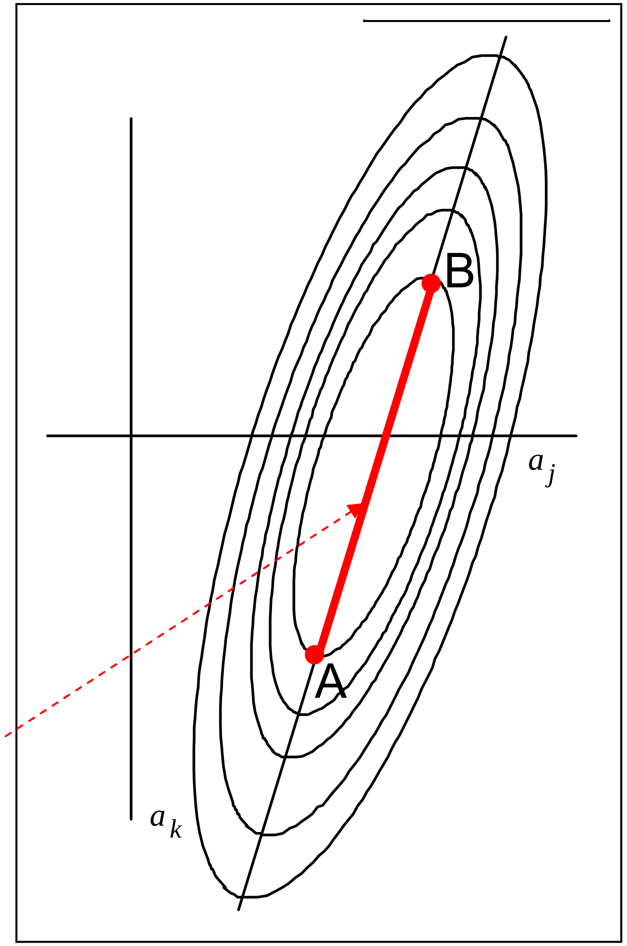


Distance between models estimated on different data sets

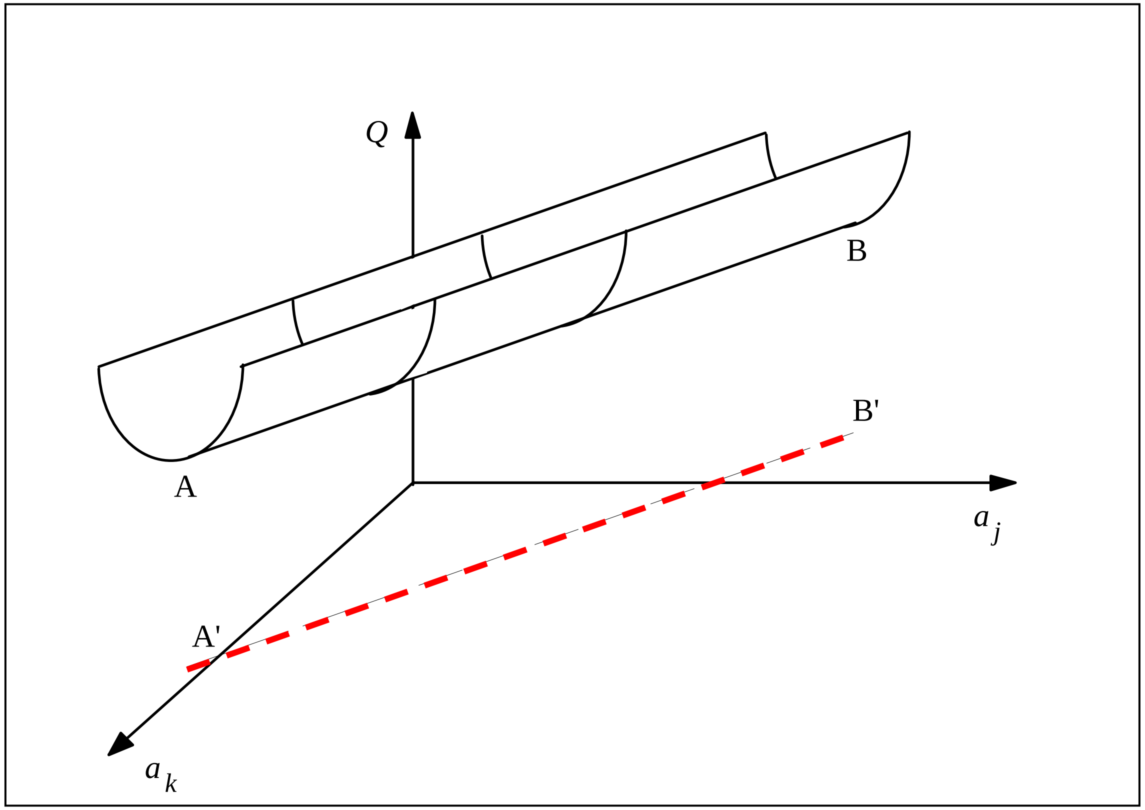


Virtual Instability of the Solution in the case of very flat functionals

The segment AB represents the area of possible solutions.

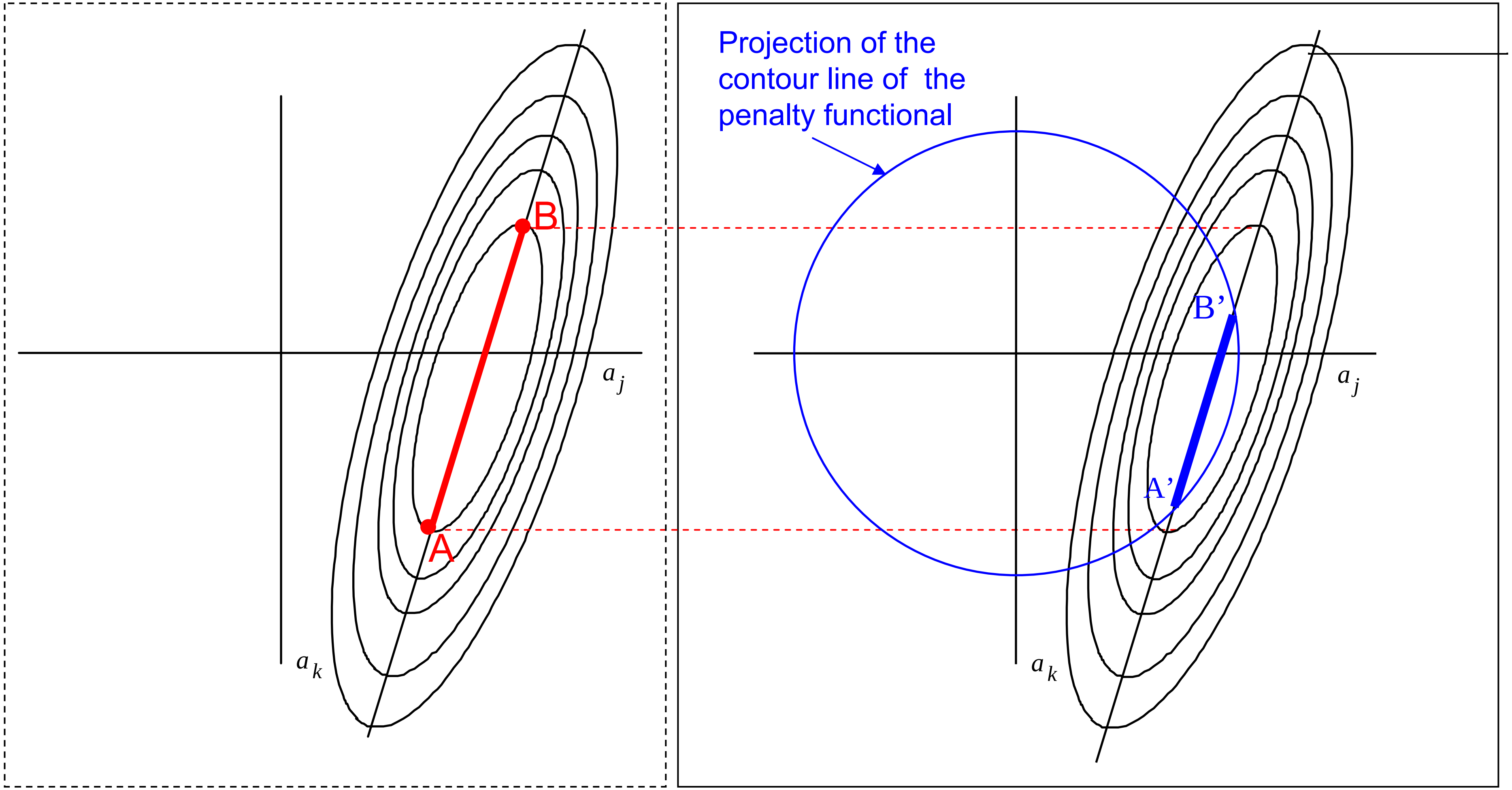


Projection of the contour lines of the functional in the case of an almost unstable solution.



A really unstable functional

The entire segment A’B’ represents the area of possible solutions.



Limitation of the solution area using a penalty functional.
The segment A'B' represents the new (reduced) solution area.

University of Applied Sciences
MERSEBURG

6. LDUC – the Local Uncertainty Criterion

Tatjana Lange

University of Applied Sciences
MERSEBURG

$$\text{LDUC} = Q = Q_1 + Q_2 = tr\left[\rho_Y\left(\hat{\vec{y}}, \vec{y}\right)\left(x^T x\right)^{-1}\right] + \rho_F\left(\vec{a}_I, \vec{a}_{II}\right)$$

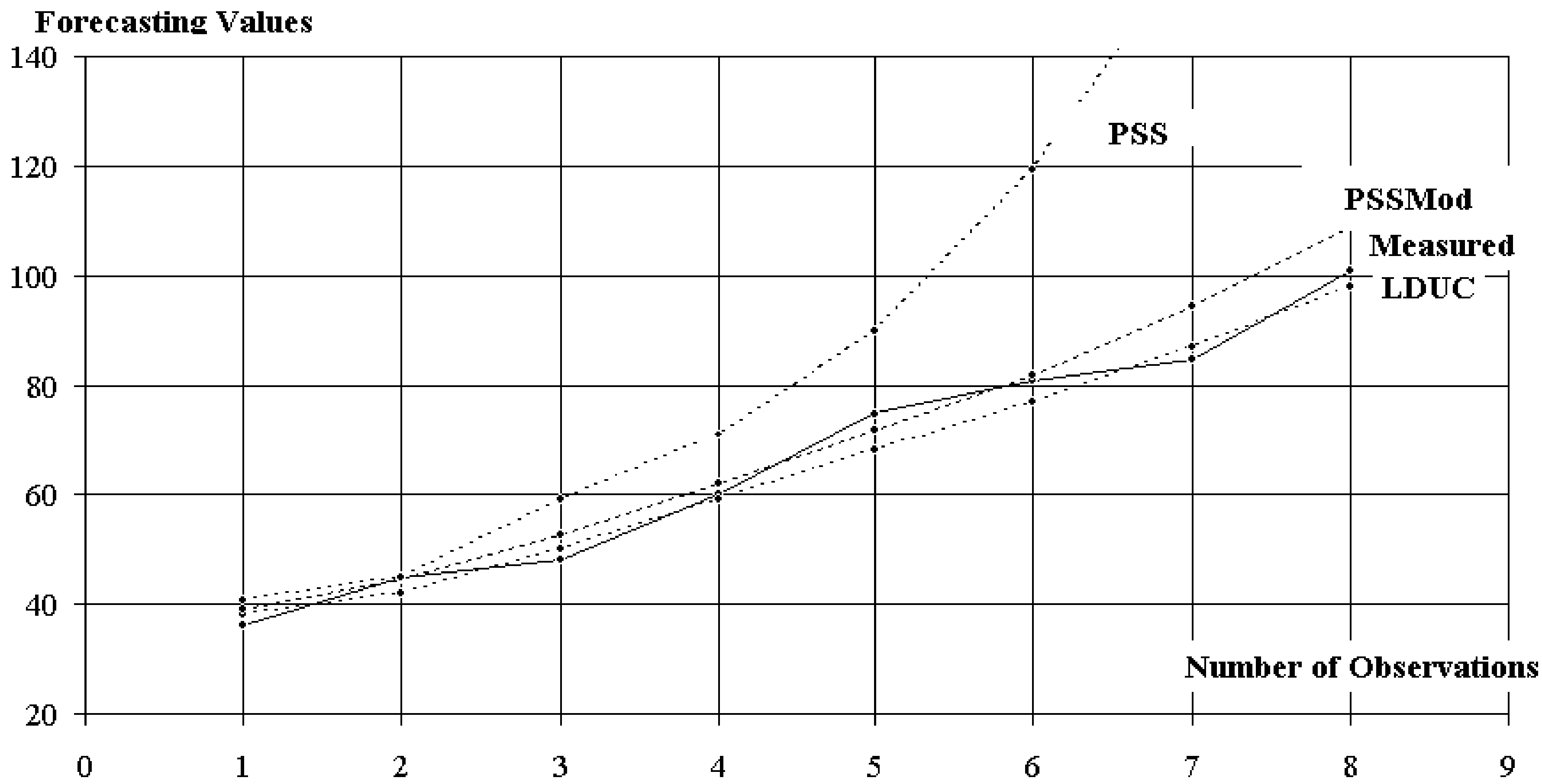
$Q_1 = Q_m$	- is the main criterion
$\left(x^T x\right)^{-1}$	- is a correlation matrix
$\rho_Y\left(\hat{\vec{y}}, \vec{y}\right)$	- is a metric in space Y but it transforms RRS into the space of coefficient F
$\rho_F\left(\vec{a}_I, \vec{a}_{II}\right)$	- is a measure of the stability of modelling in the space F

University of Applied Sciences
MERSEBURG

Example 1:
LDUC used for a Euclid Metric and a simple Data Validation

$$\mathrm{LDUC} = tr\left[\frac{\sum_{i}^{B}\left(\hat{y}_{Bi} - y_{Bi}\right)^2}{B-h}\left(\mathbf{X}_B^T \mathbf{X}_B\right)^{-1}\right] + \sum_{j=1}^{h}\left(\hat{a}_{Aj} - \hat{a}_{Bj}\right)^2$$

$A + B = l$	- is the sum of two data sets of different size (the original data set l has been subdivided into two parts A and B)
$\hat{y}_B$	- is the calculated response of model A (determined with the help of data set A) to the input data described by matrix $\mathbf{X}_B$
y_B	- is the measured response to the input data described by matrix $\mathbf{X}_B$
h	- is the number of coefficients within the model
$\hat{a}_A, \hat{a}_B$	- are the coefficients estimated with the help of the two different data sets A and B for a given model structure



- Model: $y = x^3 - x^2 + x - 1$ for $x = -2,7.....+3,0$; 20 Data Points
- multiplikativ additive Disturbance: $P = 100\%$; $\hat{y} = y(1+\theta P)$
- Forecasting in the area $x = +3,6\ldots+5,0$

Example of the Forecast Properties of Different Structure Selection Criteria

Example 2:

LDUC used for a Gnostic Metric, Bootstrap and a Ridge Correction

$$LDUC_{Boot} = tr\left[\sqrt[4]{\frac{\sum_{i=1}^{l}\left(\hat{E}_{Boot}(\hat{y}) - y_i\right)^2}{\sum_{i=1}^{l}\frac{1}{\left(\hat{E}_{Boot}(\hat{y}) - y_i\right)^2}}}(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \alpha\mathbf{I})^{-1}\right] + \sqrt[4]{\frac{\sum_{j=1}^{M}\left(\hat{E}_{Boot}(\hat{a}_j) - \hat{a}_j\right)^2}{\sum_{j=0}^{M}\frac{1}{\left(\hat{E}_{Boot}(\hat{a}_j) - \hat{a}_j\right)^2}}}$$

whereby

l - is the number of measured data points

M - is the number of coefficients within the model

$\hat{E}_{Boot}$ - is the estimation of the mathematical expectation of the evaluation value y

University of Applied Sciences
MERSEBURG

7. Summary and Outlook

Summary:

What is special about LDUC ?

1) $Q_2 = Q_p$ does not have a weighting coefficient.

   Instead of using a weighting coefficient, additional information is obtained, for example, by **“Bootstrap”**.

2) In contrast to RSS criterion, $Q_2 = Q_m$ is transferred to the coefficient space F.

3) Information for a kind of metric is obtained from the pracrical circumstances of the modelling. (a Gnostic metric may also be applied.)

University of Applied Sciences MERSEBURG

Outlook:

Comparison of the three ways of creating criteria of a “**Penalty** nature” for modelling, clustering and recognition purposes:

1) Additive criteria (refer to **Akaike**, Broerson, Schwarz, *et al.*)

$$Q = Q_m + Q_p(l, h)$$

whereby

l - is the number of measured data points

h - is the number of terms (coefficients) of the model

2) Multiplicative criteria

→ for modelling (refer to Fisher, Kondo, Mallow, Tamura, *et al.*)

→ for clustering

→ for the restoration of functions (refer to **Vapnik**)

$$Q = Q_m \cdot Q_p\left(\frac{l}{h}\right)$$

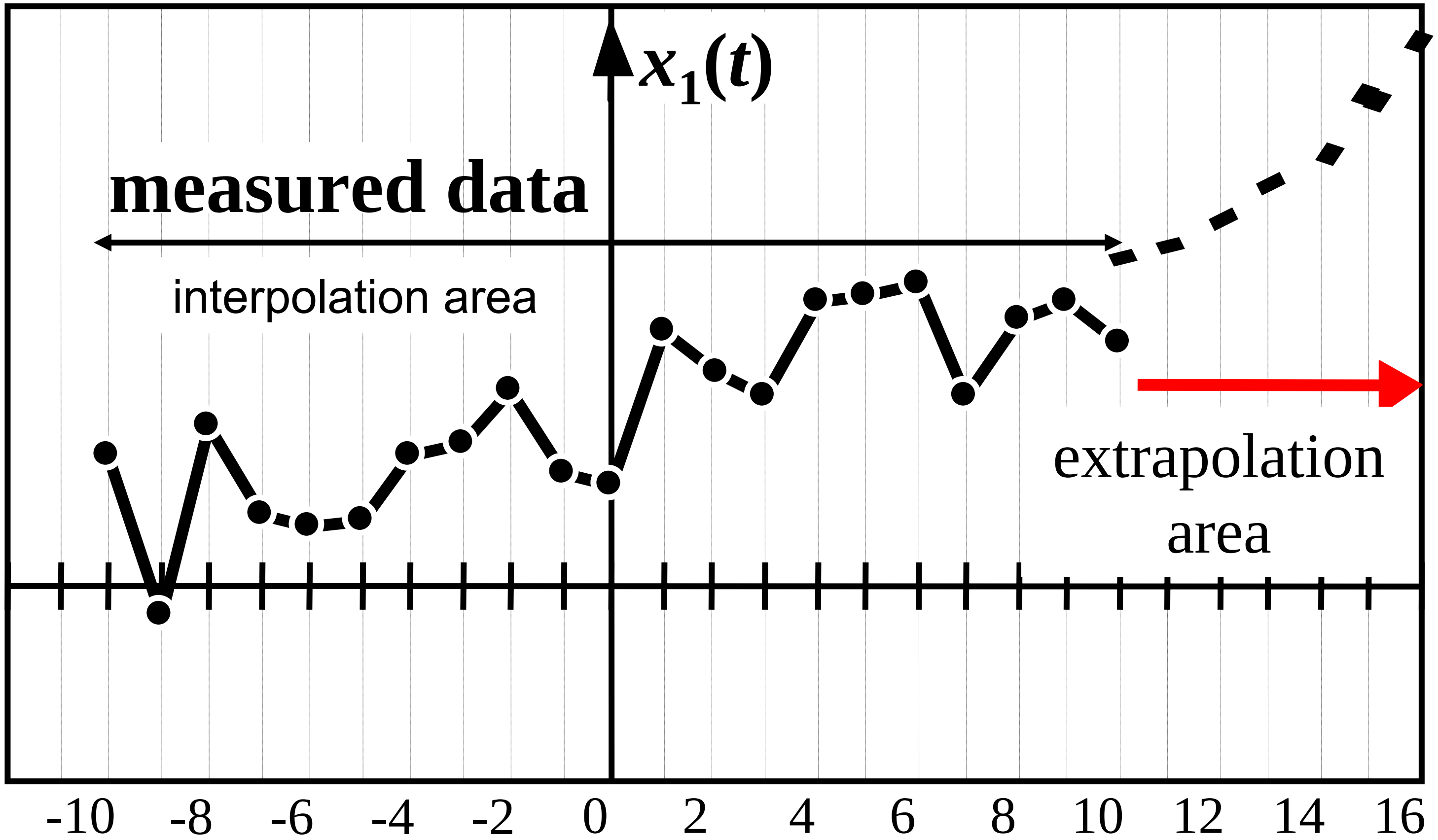
3) Additive criteria (refer to Lange)

$$Q = Q_m + Q_p$$

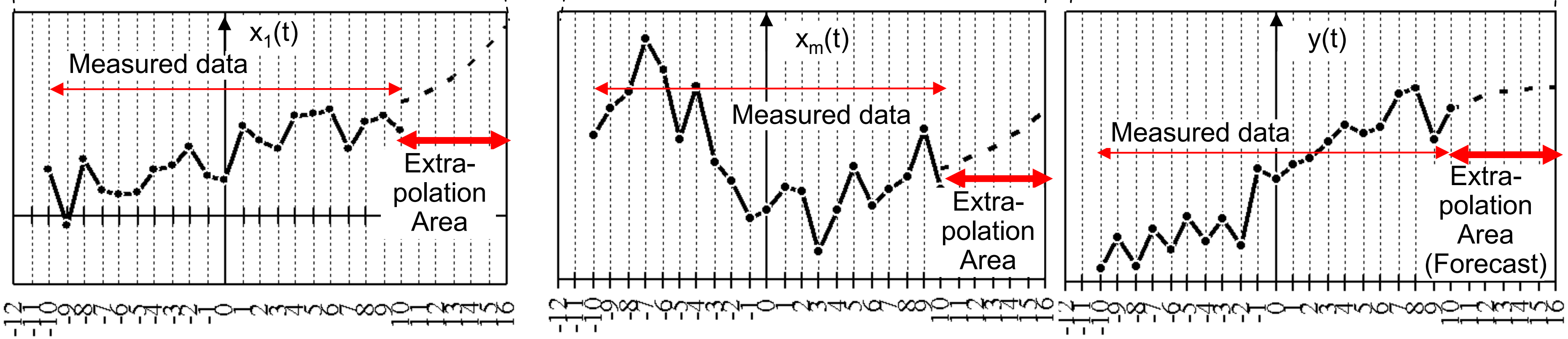
whereby

Q_p - is a measure of the stability of modelling

Back up


$x_1(t)$
measured data
interpolation area
extrapolation area
-10
-8
-6
-4
-2
0
2
4
6
8
10
12
14
16

$x_{1,k}$	$x_{1,k-1}$	$x_{1,k-2}$	…	$x_{1,k-q}$	……	$x_{m,k}$	$x_{m,k-1}$	$x_{m,k-2}$	…	$x_{m,k-q}$	Nr	y_k	y_{k-1}	y_{k-2}	…	y_{k-q}
$x_1(0)$	$x_1(-1)$	$x_1(-2)$			……	$x_m(0)$	$x_m(-1)$	$x_m(-2)$			0	y(0)	y(-1)	y(-2)		
$x_1(1)$	$x_1(0)$	$x_1(-1)$			……	$x_m(1)$	$x_m(0)$	$x_m(-1)$			1	y(1)	y(0)	y(-1)		
$x_1(2)$	$x_1(1)$	$x_1(0)$			……	$x_m(2)$	$x_m(1)$	$x_m(0)$			2	y(2)	y(1)	y(0)		
.	.	.	.	.	……	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	……	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	……	.	.	.	.	.	.	.	.	.	.	.
$x_1(10)$	$x_1(9)$	$x_1(8)$			……	$x_m(10)$	$x_m(9)$	$x_m(8)$			10	y(10)	y(9)	y(8)		

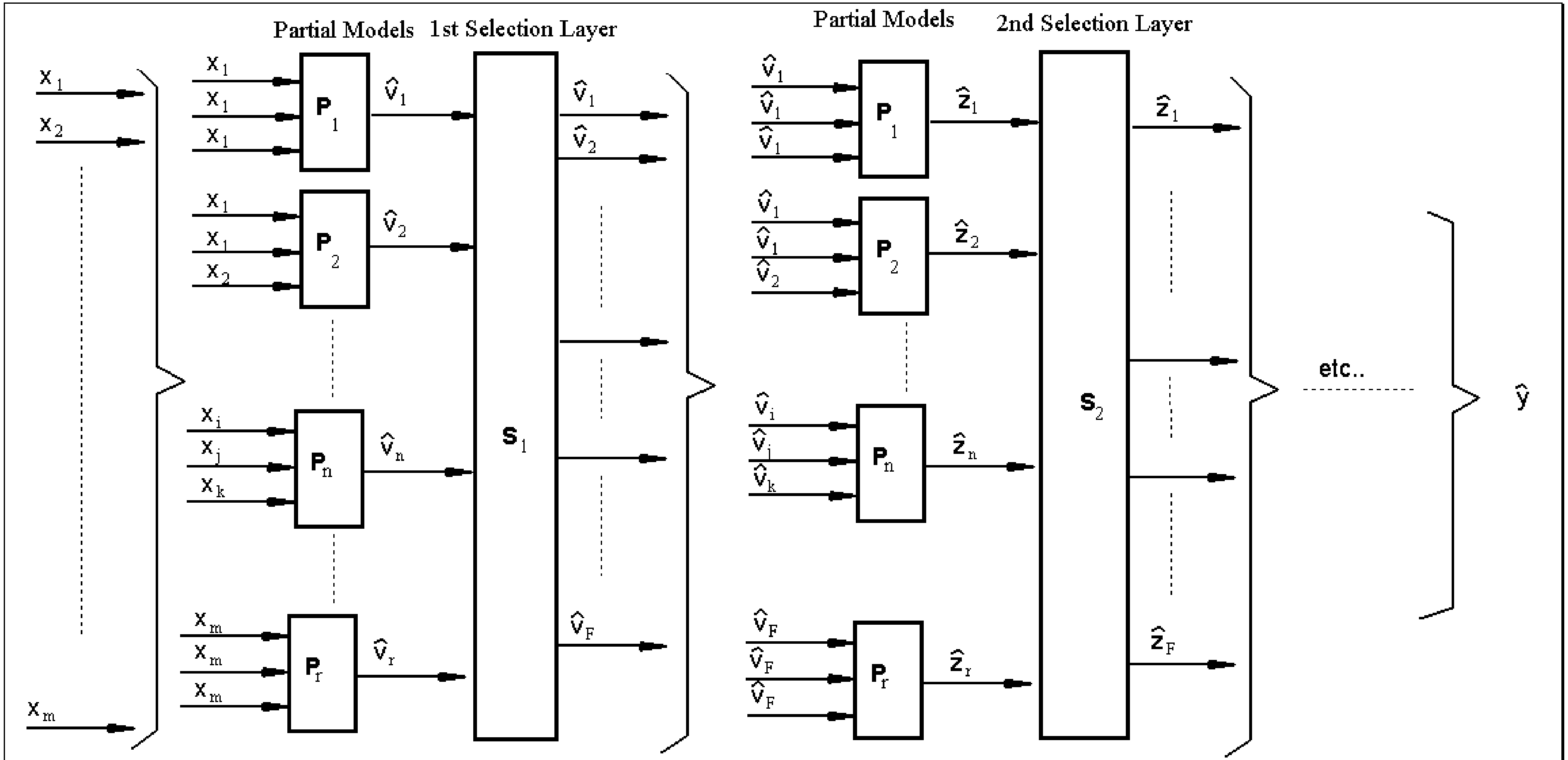


Measured data (Example)

	Field of Application		
	Static characteristic curves as a function $y = f(x_1, x_2, \dots, x_m)$	Determination of models ⇨ Dynamic characteristic curves of non-linear systems ⇨ Multi-dimensional time series as a function $y_k = f(y_{k-1}, y_{k-2}, \dots, y_{k-q}, x_{1,k}, x_{1,k-1}, \dots, x_{1,k-q}, x_{2,k-1}, \dots, x_{2,k-q}, \dots, x_{m,k-1}, \dots, x_{m,k-q})$	Autoregression as a function $y_k = f(y_{k-1}, y_{k-2}, \dots, y_{k-q})$
Number of terms of complete polynomial	$v = \frac{(1+m+K)!}{K!(1+m)!} - 1$	$v = \frac{(1+m+q+K)!}{K!(1+m+q)!} - 1$	$v = \frac{(1+q+K)!}{K!(1+q)!} - 1$
Construction of input matrix X	$\mathbf{X} = \begin{bmatrix} 1 & x_1(0) & \dots & x_m(0) \\ 1 & x_1(1) & \dots & x_m(1) \\ . & . & \dots & . \\ . & . & \dots & . \\ 1 & x_1(l) & \dots & x_m(l) \end{bmatrix}$	$\mathbf{X} = \begin{bmatrix} y(-1) & y(-2) & \dots & x_1(0) & x_1(-1) & \dots & x_m(-q) \\ y(0) & y(-1) & \dots & x_1(1) & x_1(0) & \dots & x_m(1-q) \\ y(1) & y(0) & \dots & x_1(2) & x_1(1) & \dots & . \\ . & . & \dots & . & . & \dots & . \\ y(l) & y(l-1) & \dots & x_1(l) & x_1(l-1) & \dots & x_m(l-q) \end{bmatrix}$	$\mathbf{X} = \begin{bmatrix} y(-1) & y(-2) & \dots & y(-q) \\ y(0) & y(-1) & \dots & y(1-q) \\ y(1) & y(0) & \dots & . \\ . & . & \dots & . \\ y(l) & y(l-1) & \dots & y(l+1-q) \end{bmatrix}$
Output vector	$\vec{y} = \vec{y}_k = \begin{bmatrix} y(0) \\ y(1) \\ . \\ y(l) \end{bmatrix}$	$\vec{y}_k = \begin{bmatrix} y(0) \\ y(1) \\ . \\ y(l) \end{bmatrix}$	$\vec{y}_k = \begin{bmatrix} y(0) \\ y(1) \\ . \\ y(l) \end{bmatrix}$

m – Number of variables	q – Number of delay steps	K – Power of polynomial

Fields of Application of Structure Modelling



P_n - n-th partial model on any layer; basic form: $v = Ax_i + Bx_jx_k$ or. $z = Cv_i + Dv_jv_k$

$\hat{v}_n$ - Output evaluated with the help of the partial model P_n

S_I - Selection of the best intermediate models F from partial models r with the help of a structure criterion, I - number of layer

$\hat{y}$ - Evaluated output of the final models (output of the best Model on the last layer)

Simplified block diagram of the multi-layered GMDH algorithm

Subdivision of data into a “learning data set" and a “test data set":

	x_1	x_2	....	x_m	y
1					
2					
l					

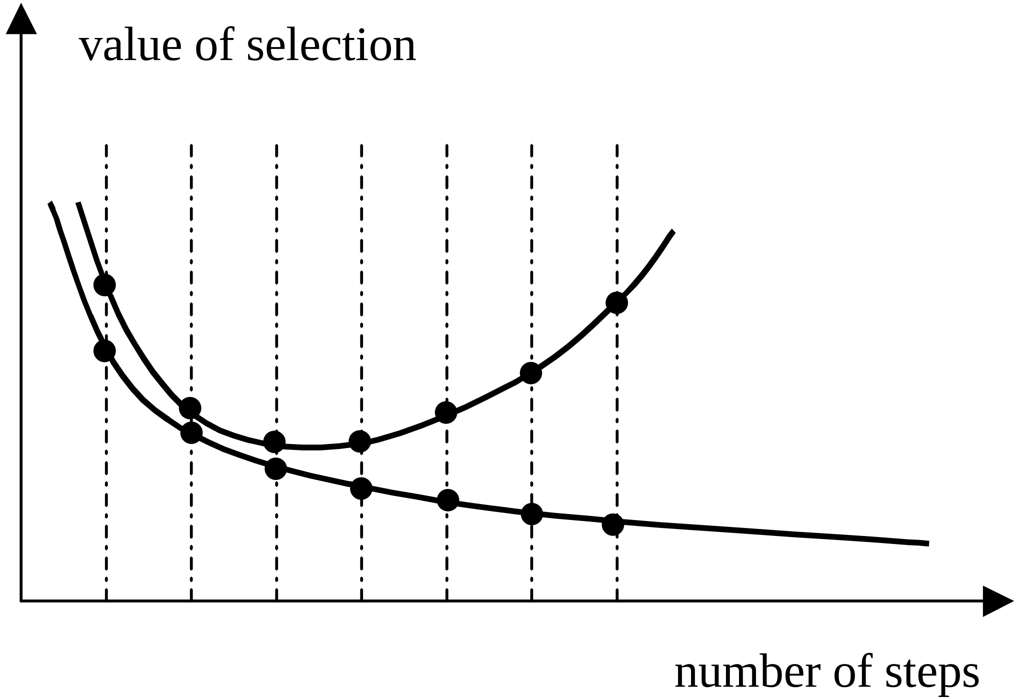
learning data set

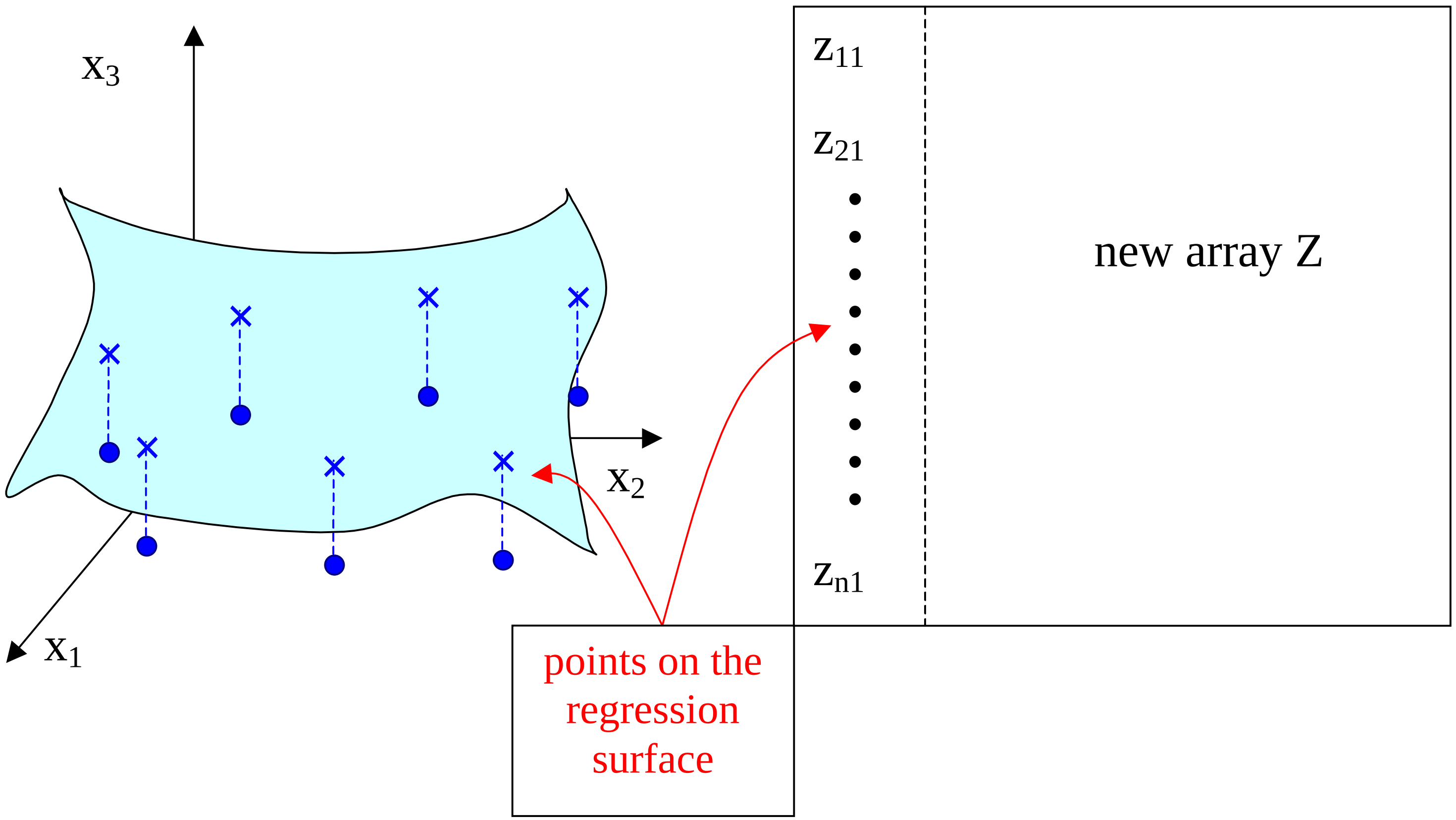
test data set

Models become increasingly complicated until they reach the minimum of the selection criterion.

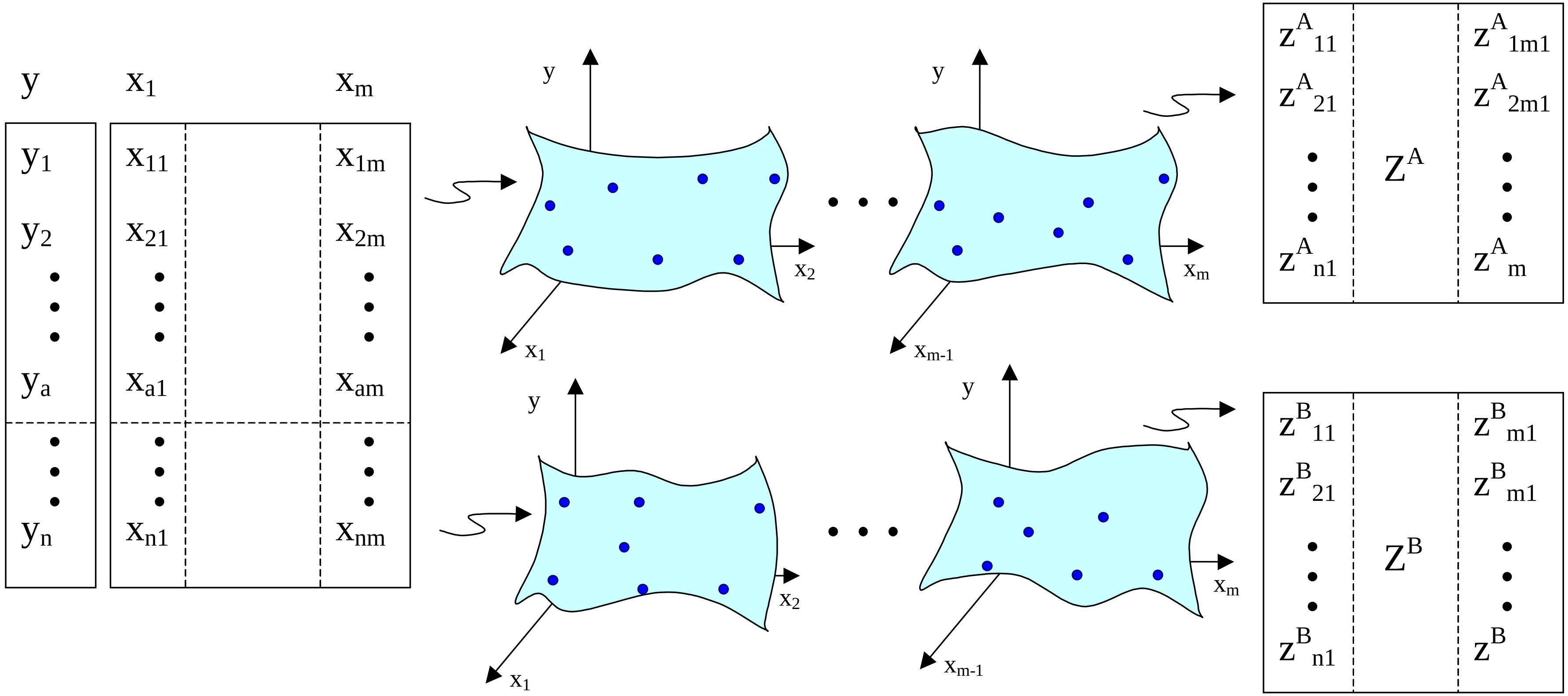
Example:

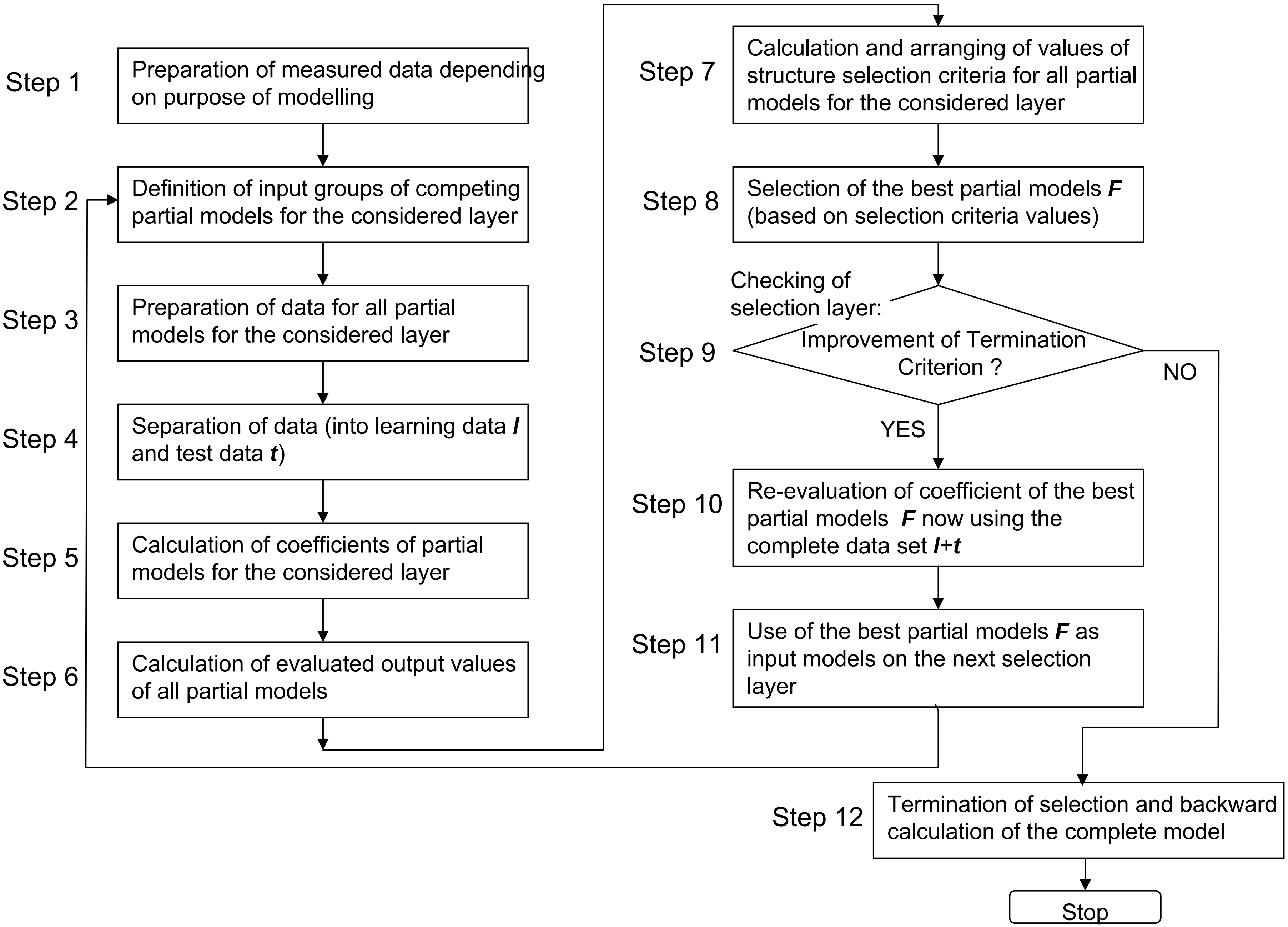
$$y = A_0 + A_1 x_{17} \cdot A_2 x_6 x_9 + ... + A_i x_1^3 x_2^5 x_3 + ...$$


value of selection
number of steps


x3
x2
x1
points on the regression surface
z11
z21
zn1
new array Z

Data Division




Step 1
Preparation of measured data depending on purpose of modelling
Step 2
Definition of input groups of competing partial models for the considered layer
Step 3
Preparation of data for all partial models for the considered layer
Step 4
Separation of data (into learning data ***l*** and test data ***t***)
Step 5
Calculation of coefficients of partial models for the considered layer
Step 6
Calculation of evaluated output values of all partial models
Step 7
Calculation and arranging of values of structure selection criteria for all partial models for the considered layer
Step 8
Selection of the best partial models ***F*** (based on selection criteria values)
Checking of selection layer:
Step 9
Improvement of Termination Criterion ?
NO
YES
Step 10
Re-evaluation of coefficient of the best partial models ***F*** now using the complete data set ***l+t***
Step 11
Use of the best partial models ***F*** as input models on the next selection layer
Step 12
Termination of selection and backward calculation of the complete model
Stop

1. Für den Fall, daß keine qualitative Information über die Daten vorhanden ist (z. B. über die Verteilung der Störung), wird die s.g. **quadratische Variante des LDUC** ($LDUC_q$) vorgeschlagen.

$$LDUC_q = Q = tr\left[\frac{\sum_{i=1}^{B}\left(\hat{y}_{B(A)_i} - y_{B_i}\right)^2}{B-M} \cdot (\mathbf{X}_B{}^T \mathbf{X}_B + \alpha \mathbf{I})^{-1}\right] + \sum_{j=0}^{M} (\hat{a}_{A_j} - \hat{a}_{B_j})^2$$	(6.2.b)

Hier sind

A, B	-	Anzahl der Datenpunkte in zwei Stichproben A und B, bzw. Indizes der Stichprobe A und B, in die eine Datentabelle geteilt wird.
$\hat{\vec{a}}_A, \hat{\vec{a}}_B$	-	Koeffizientenvektoren,, die aus den verschiedenen Datenstichproben berechnet wurden,
$\vec{y}_B$	-	gemessener Ausgang aus der Stichprobe B,
$\hat{\vec{y}}_{B(A)}$	-	geschätzter Ausgangsvektor mit den Eingangsdaten der Stichprobe B, wobei die Koeffizienten des Teilmodells aus der Stichprobe A berechnet wurden, so, daß z.B. für ein Punkt i der Stichprobe B wird $\hat{y}_{B(A)_i} = \vec{x}_{B_i}^T * \hat{a}_A$.
ρ_l, ρ_M	-	metrische (z.B. euklidische) Abstände im l- und M-Raum,
l	-	Anzahl der Datenpunkte in der jeweiligen Stichprobe,
M	-	Anzahl der Koeffizienten (ohne $\hat{a}_0$),

2. Für den Fall, daß aus der Information über die Messungen zu erwarten ist, daß Ausreißer (outlier) in den Meßreihen vorhanden sein können, wird die **robuste Variante des LDUC** (**$LDUC_r$**) vorgeschlagen.

$$LDUC_r = Q = tr\left[\frac{\sum_{i=1}^{B}\left\|y_{B_i} - \vec{x}_{B_i}^{\mathrm{T}}\cdot\hat{\hat{a}}_A\right\|^r}{B-M}\cdot(\mathbf{X}_B{}^{\mathrm{T}}\mathbf{X}_B + \alpha\cdot\mathbf{I})^{-1}\right] + \sum_{j=0}^{M}\left\|\hat{a}_{A_j} - \hat{a}_{B_j}\right\|^r$$	(6.2.c.)

mit

$1 \le r \le 2$		
$\hat{\hat{a}}_A, \hat{\hat{a}}_B$	-	(aus den Stichproben A und B) robust geschätzte Koeffizienten.
		(weitere Formelzeichen - s. 6.2.b.)

3. Für den Fall, daß aus der Information über die Messungen zu erwarten ist, daß Ausreißer (outlier) in den Meßreihen vorhanden sein können, wird auch die **gnostische Variante des LDUC** ($LDUC_g$) vorgeschlagen. Diese Variante hat im Vergleich zu robusten Variante den Vorteil, daß es nicht notwendig ist, die Größe ***r*** zu wählen.

$LDUC_g = Q = tr\left[\sqrt[4]{\dfrac{\sum_{i=1}^{B}(\hat{y}_{B(A)_i} - y_{B_i})^2}{\sum_{i=1}^{B}\dfrac{1}{(\hat{y}_{B(A)_i} - y_{B_i})^2}}}\right](\mathbf{X}_B{}^{\mathrm{T}}\mathbf{X}_B + \alpha\mathbf{I})^{-1} + \sqrt[4]{\dfrac{\sum_{j=0}^{M}(\hat{a}_{B_j} - \hat{a}_{A_j})^2}{\sum_{j=0}^{M}\dfrac{1}{(\hat{a}_{B_j} - \hat{a}_{A_j})^2}}}$	(6.2.d.)

		(Formelzeichen - s. 6.2.b.)

4. Wenn die Unsicherheit der Daten besonders groß ist und es die rechnerischen Möglichkeiten erlauben, wird die **Bootstrap-Variante des LDUC** (**$LDUC_b$**) vorgeschlagen:

$$LDUC_b = Q = tr\left[\frac{\sum_{i=1}^{B}\left(\hat{E}_{boot}(\hat{y}_i) - y_{B_i}\right)^2}{l-M}\left(\mathbf{X}_B^{\mathrm{T}}\mathbf{X}_B + \alpha\mathbf{I}\right)^{-1}\right] + \sum_{j=0}^{M}\left(\hat{E}_{boot}(\hat{a}_j) - \hat{a}_j\right)^2$$	(6.2e.)

Hier ist	$$\hat{E}_{boot}(\hat{y}_i) = \frac{\sum_{nb=1}^{NB}\hat{y}_{i_{nb}}}{NB-1}$$

die Schätzung der mathematischen Erwartung des geschätzten Ausgangs im Punkt *i*, $\hat{y}_i = \vec{x}_i^T * \hat{\vec{a}}$, nach Bootstrap. Die Schätzung der Koeffitienten wird *NB*-mal durchgeführt (*NB* ist Anzahl der Bootstrap-Datentabellen; der Index ***nb*** entspricht der laufenden Bootstrap-Datentabelle).

Analog dazu ist	$$\hat{E}_{boot}(\hat{a}_j) = \frac{\sum_{nb=1}^{NB}\hat{a}_{j_{nb}}}{NB-1}$$

die Schätzung der mathematischen Erwartung des *j*-ten Koeffizienten nach Bootstrap. $\hat{a}_j$ ist der aus der ursprünglichen Datentabelle errechnete Koeffizient.