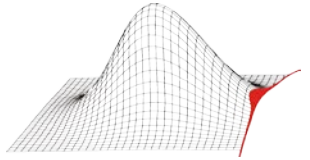
TECHNISCHE UNIVERSITÄT DRESDEN

Faculty of Mechanical Science and Engineering, Institute of Fluid Mechanics, Chair of Turbomachinery and Jet-Propulsion

Monte Carlo cross validation for response surface benchmark

André Beschorner, Matthias Voigt, Konrad Vogeler

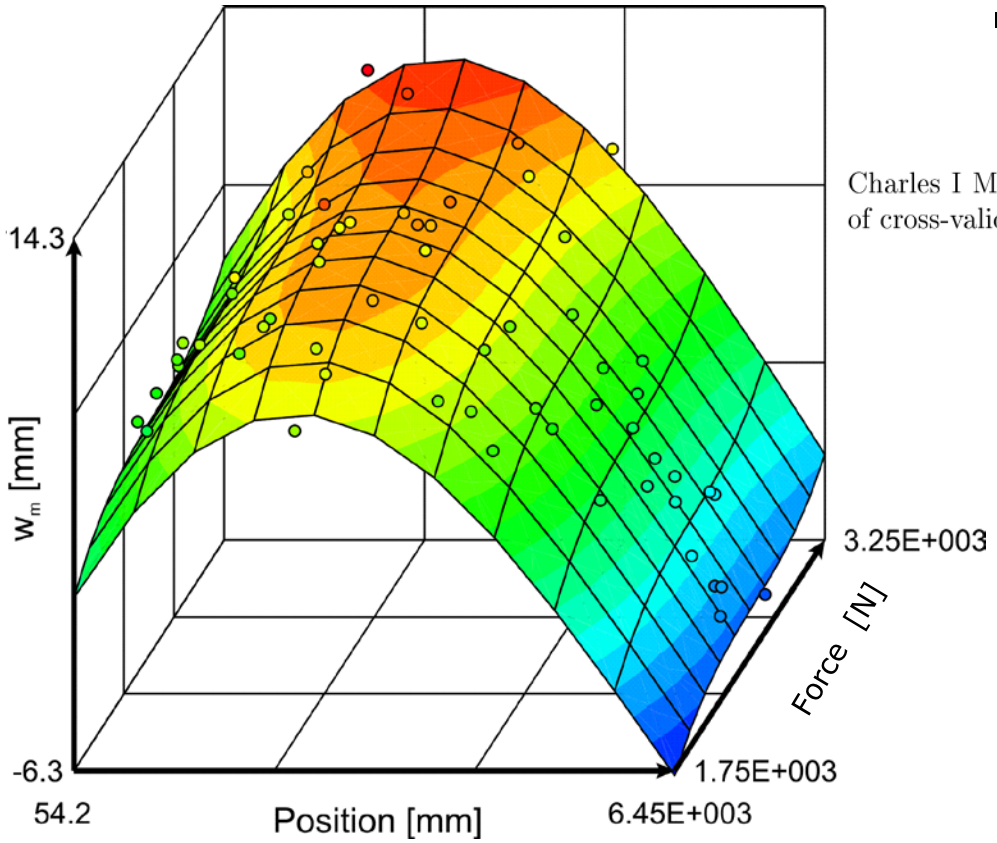
Outline



- Introduction
- Variants of cross validation
- Results of Monte Carlo cross validation
- Summary

What is cross validation?

- Cross validation is a method to estimate the prediction quality of a response surface



- Already mentioned in 1951 by Charles Mosier

Charles I Mosier. The need and means of cross-validation. i. problems and designs of cross-validation. *Educational and Psychological Measurement*, 1951.

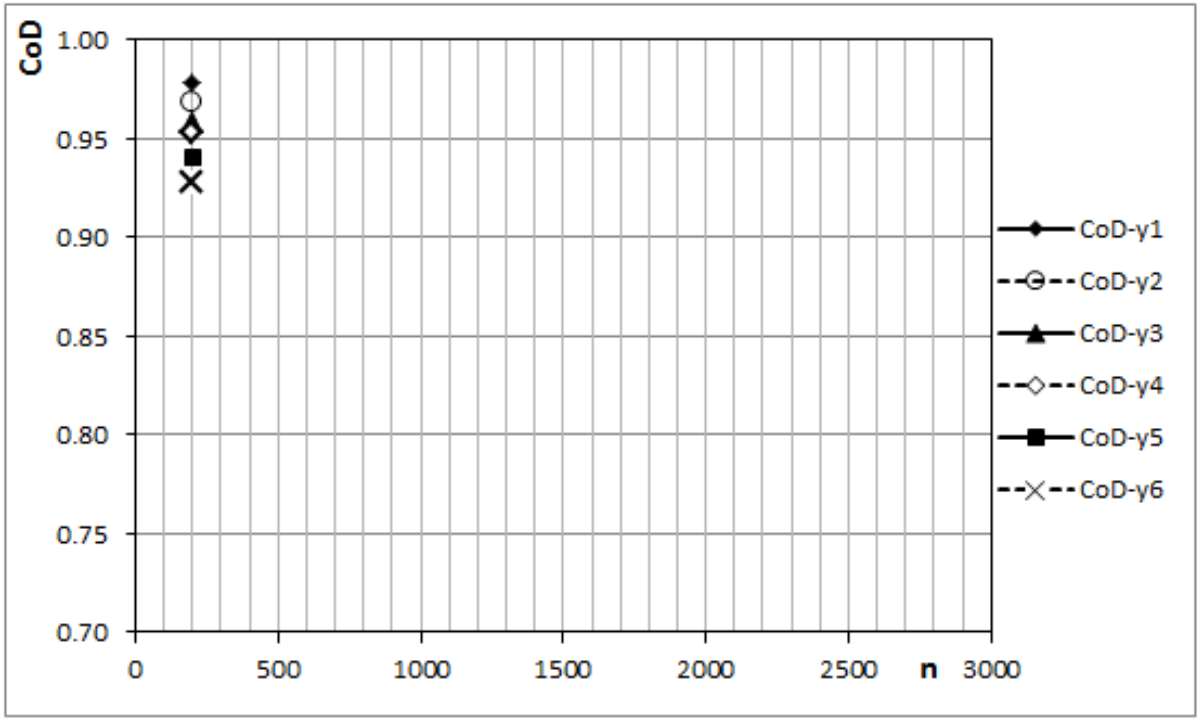
- Used more intensively since the 70's

What is cross validation?

We have R^2(CoD), MSE, ... ?

Computation and validation of model use the same database

Describes fit between y and $\tilde{y}$

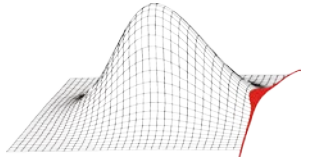


This results in an overestimation of the model quality as long as n is ”low”.

Model:

140 input variables

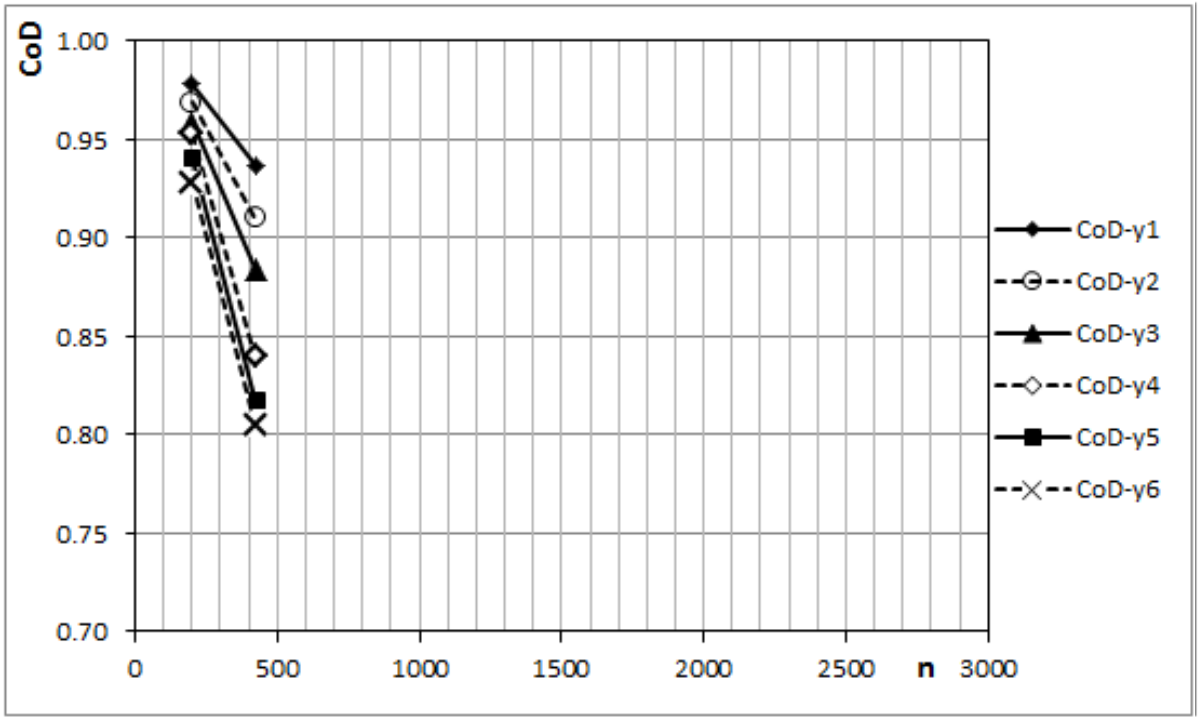
What is cross validation?



We have R^2(CoD), MSE, ... ?

Computation and validation of model use the same database

Describes fit between y and $\tilde{y}$



This results in an overestimation of the model quality as long as n is ”low”.

Model:

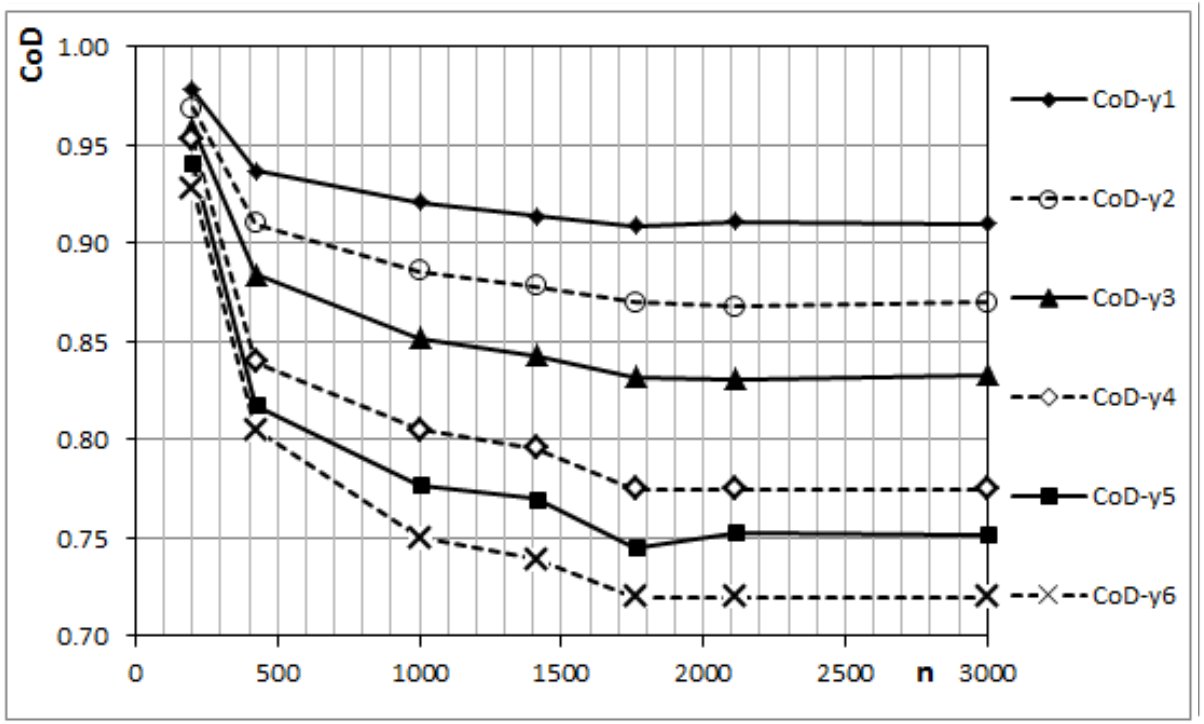
140 input variables

What is cross validation?

We have R^2(CoD), MSE, ... ?

Computation and validation of model use the same database

Describes fit between y and $\tilde{y}$

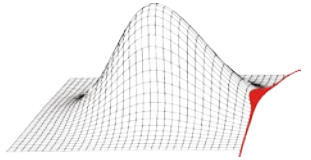


This results in an overestimation of the model quality as long as n is ”low”.

Model:

140 input variables

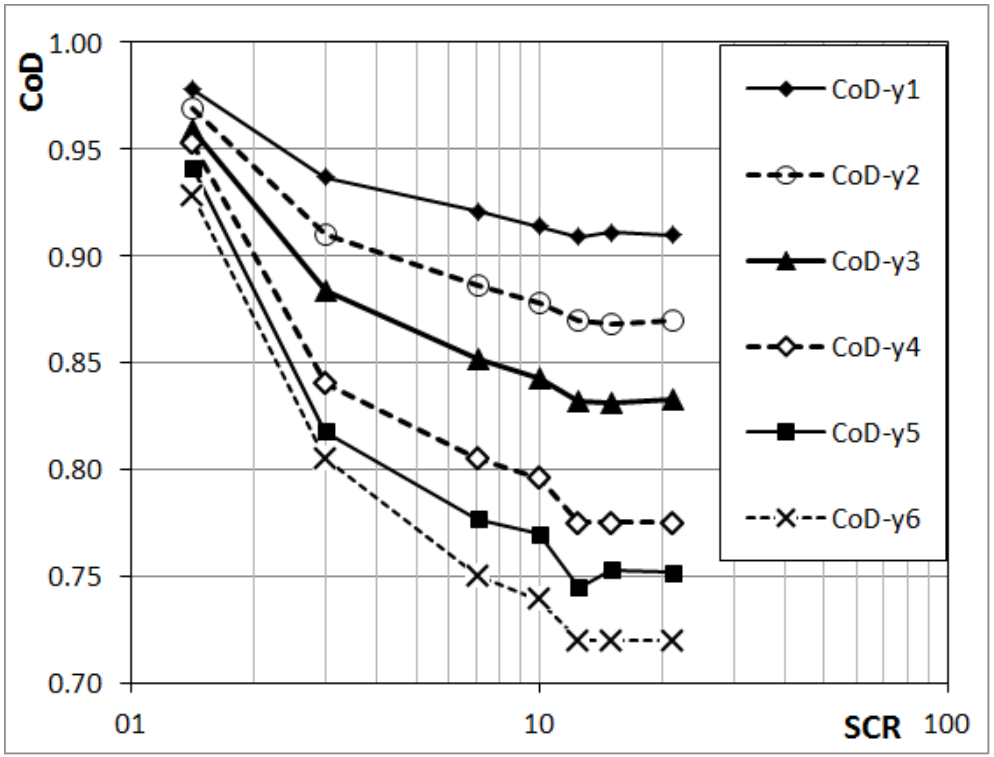
What is cross validation?



We have R^2(CoD), MSE, ... ?

Computation and validation of model use the same database

Describes fit between y and $\tilde{y}$



This results in an overestimation of the model quality as long as n is ”low”.

Model:

140 input variables

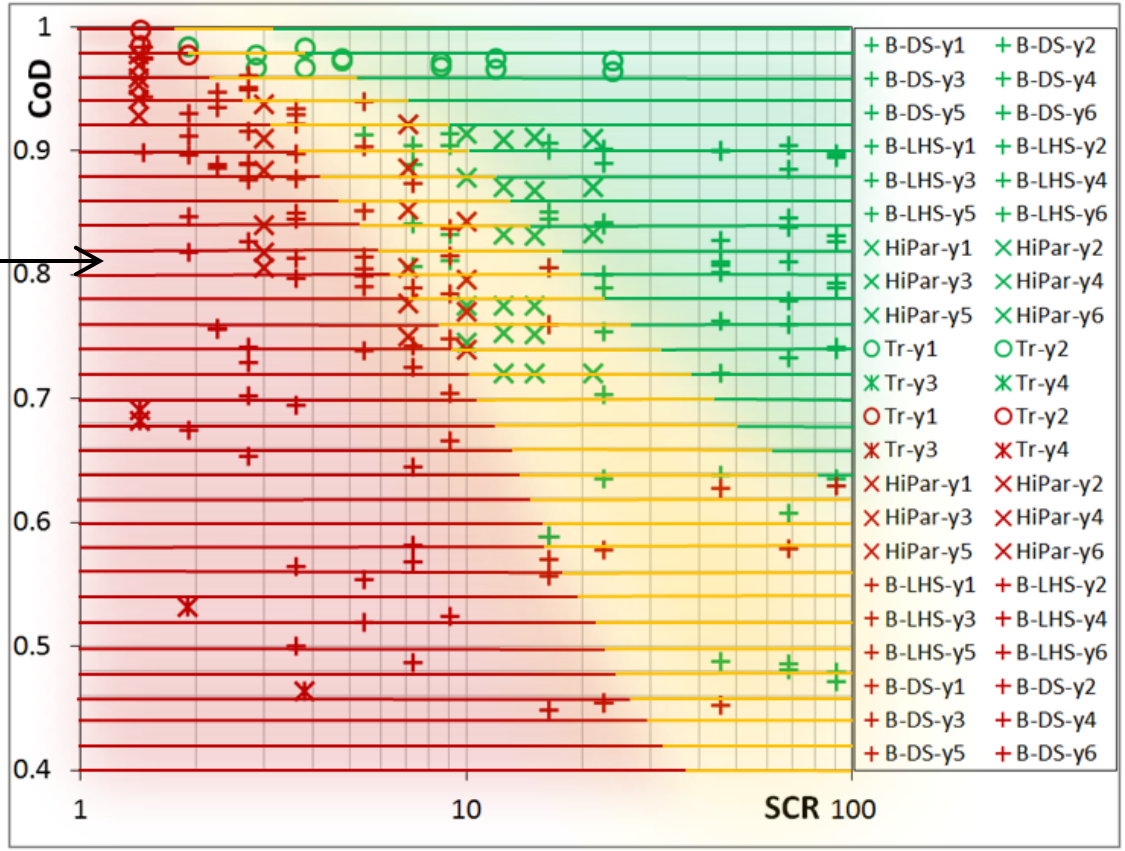
SCR = n/c

What is cross validation?

Overestimation of model quality with R^2(CoD)

Marginal changes of CoD value by enlarging database

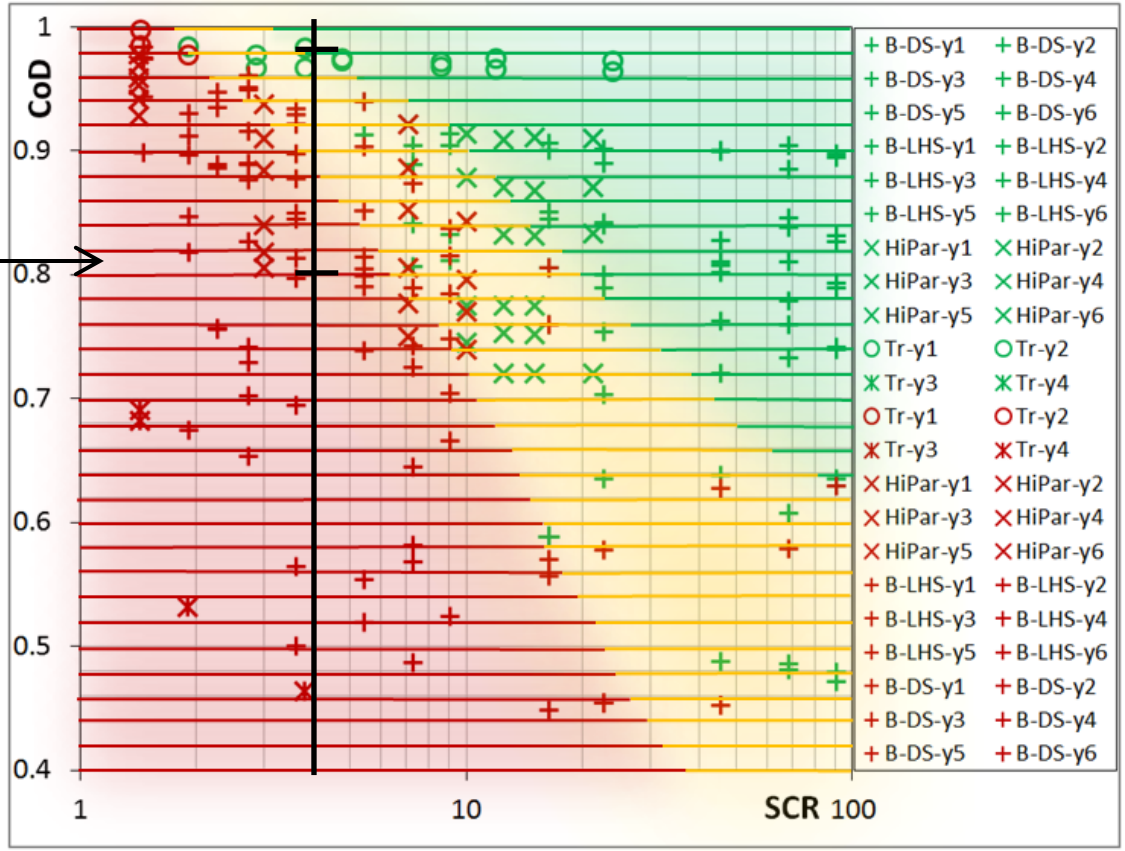
Strong changes of CoD value while enlarging database



Overestimation of model quality with R^2(CoD)

Marginal changes of CoD value by enlarging database

Strong changes of CoD value while enlarging database



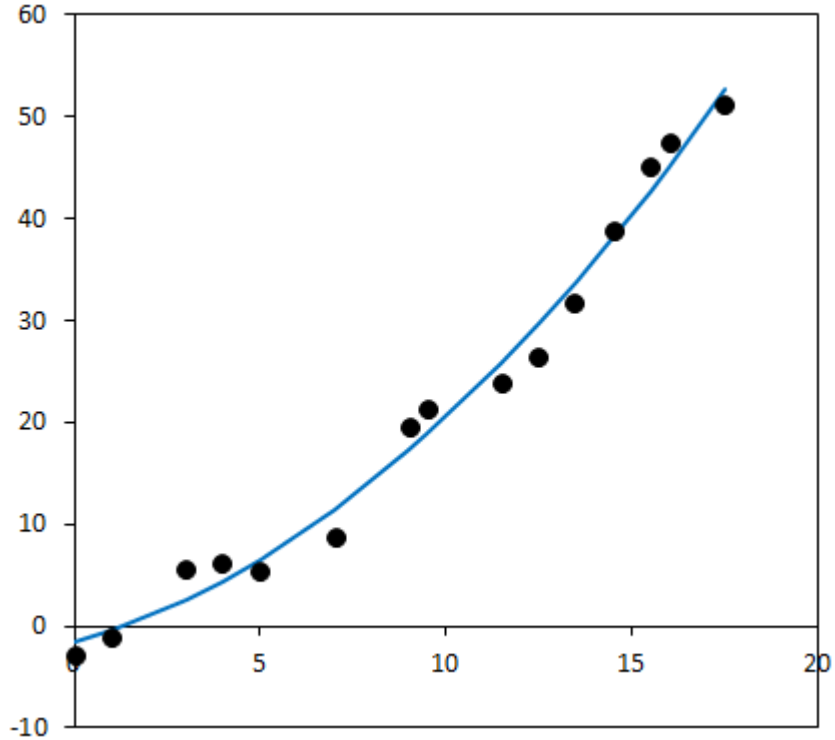
Example:
- with 25 coefficients
- 100 deterministic Support points

→ SCR = 4
→ CoD of 0.8 is unsaturated
→ CoD of 0.98 is saturated

What is cross validation?

Aim of cross validation

- Should avoid quality overestimation for small datasets
- Show the prediction quality of the model



For estimation of prediction quality, it is necessary validate the model against different samples.

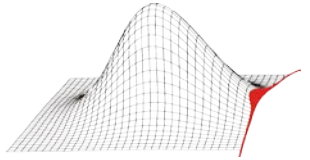
New data is seldom available.

→ Split the data

into *training samples*

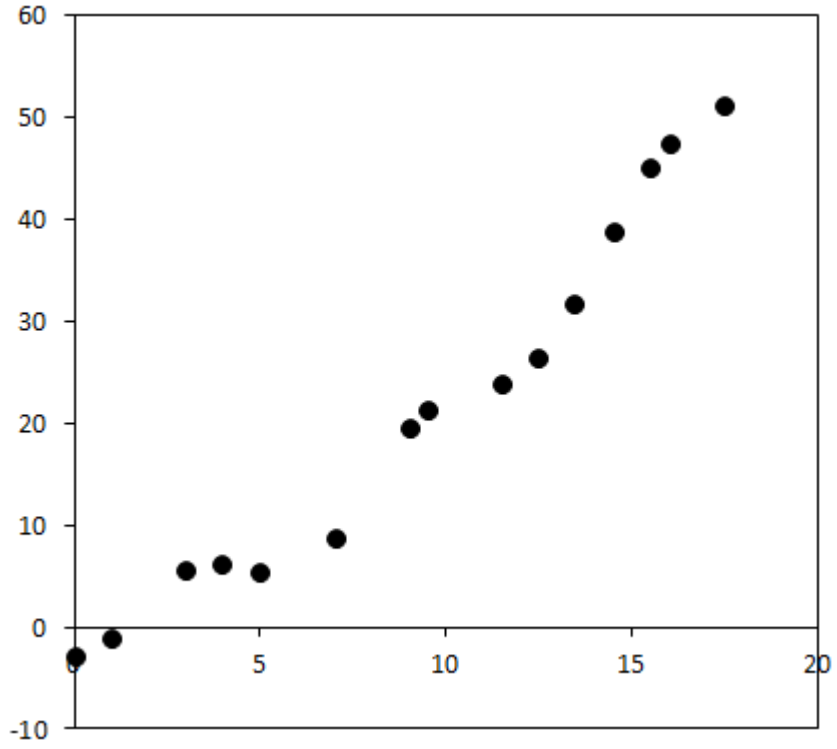
and *validation samples*

What is cross validation?

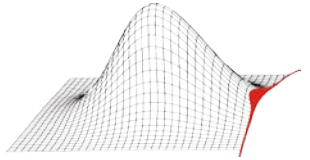


Example:

1. Leave out some (at least one) samples for the data fitting

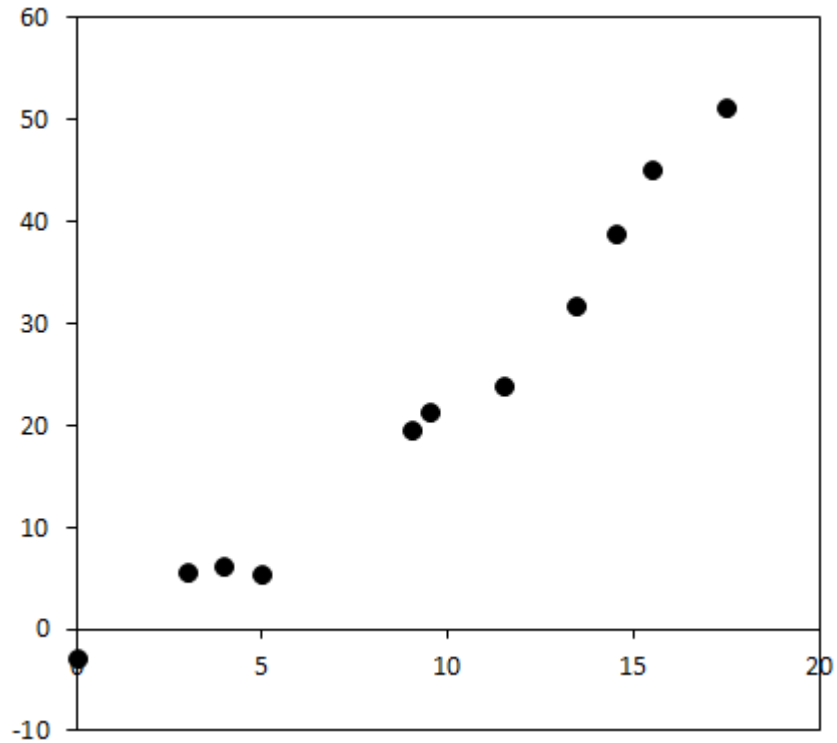


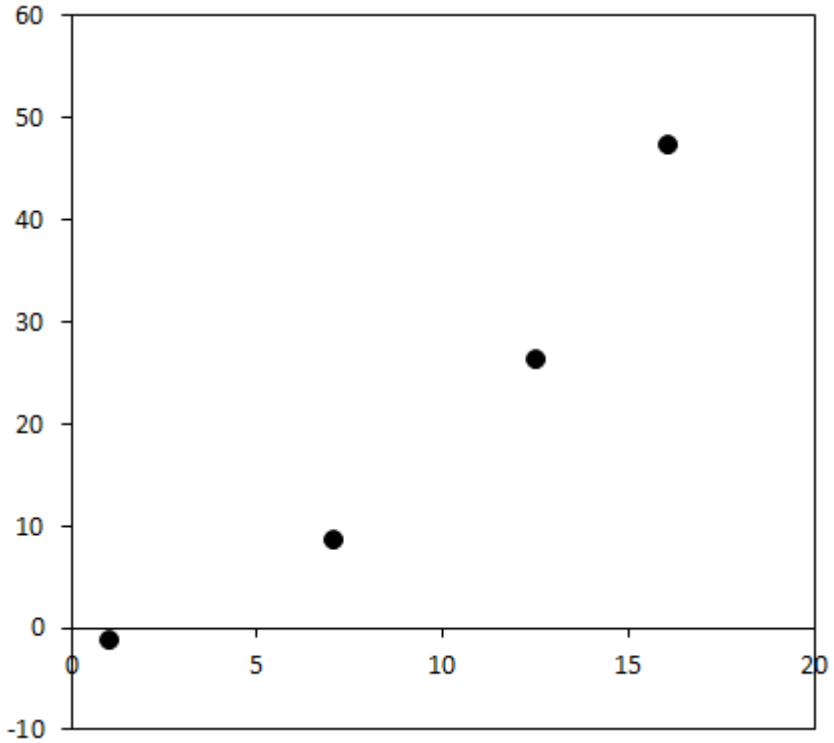
What is cross validation?



Example:

1. Leave out some (at least one) samples for the data fitting

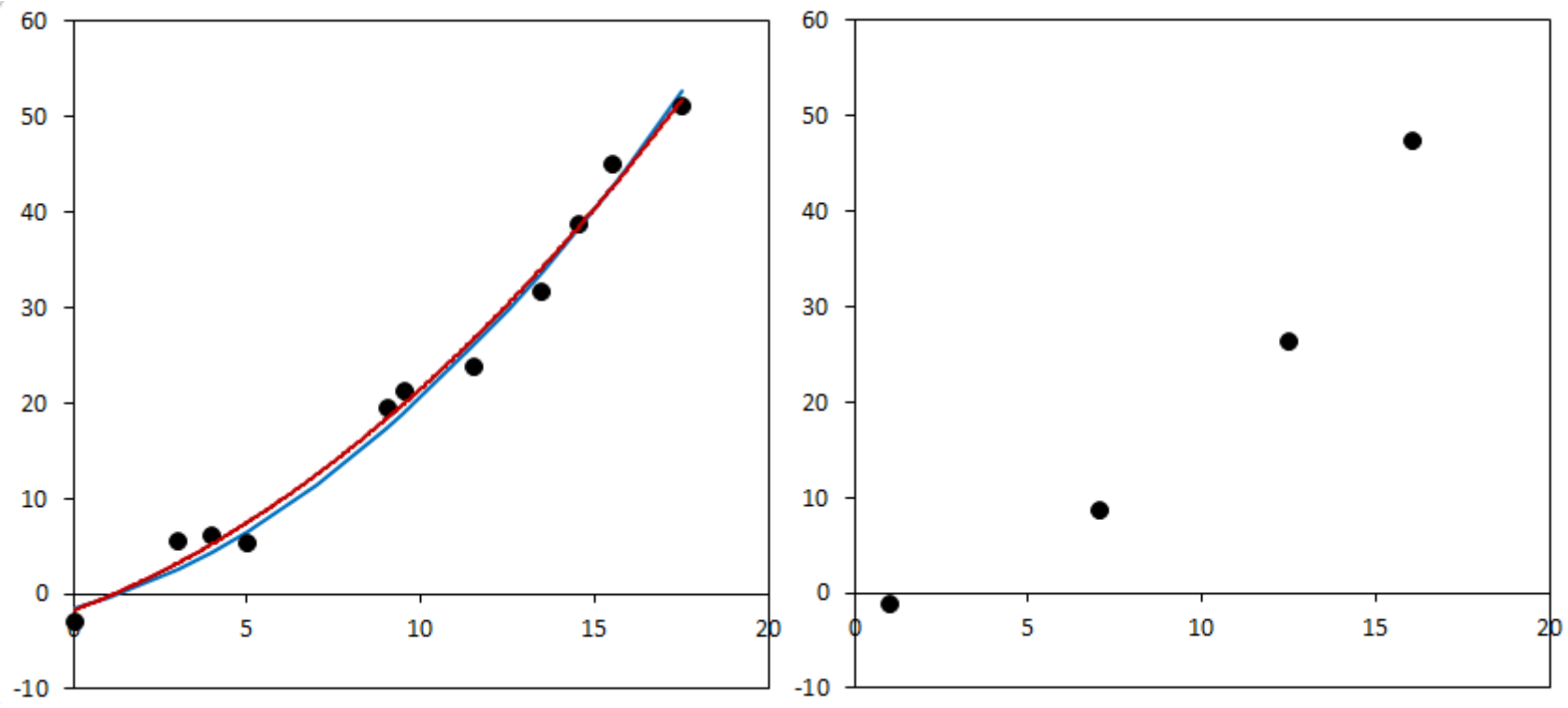




What is cross validation?

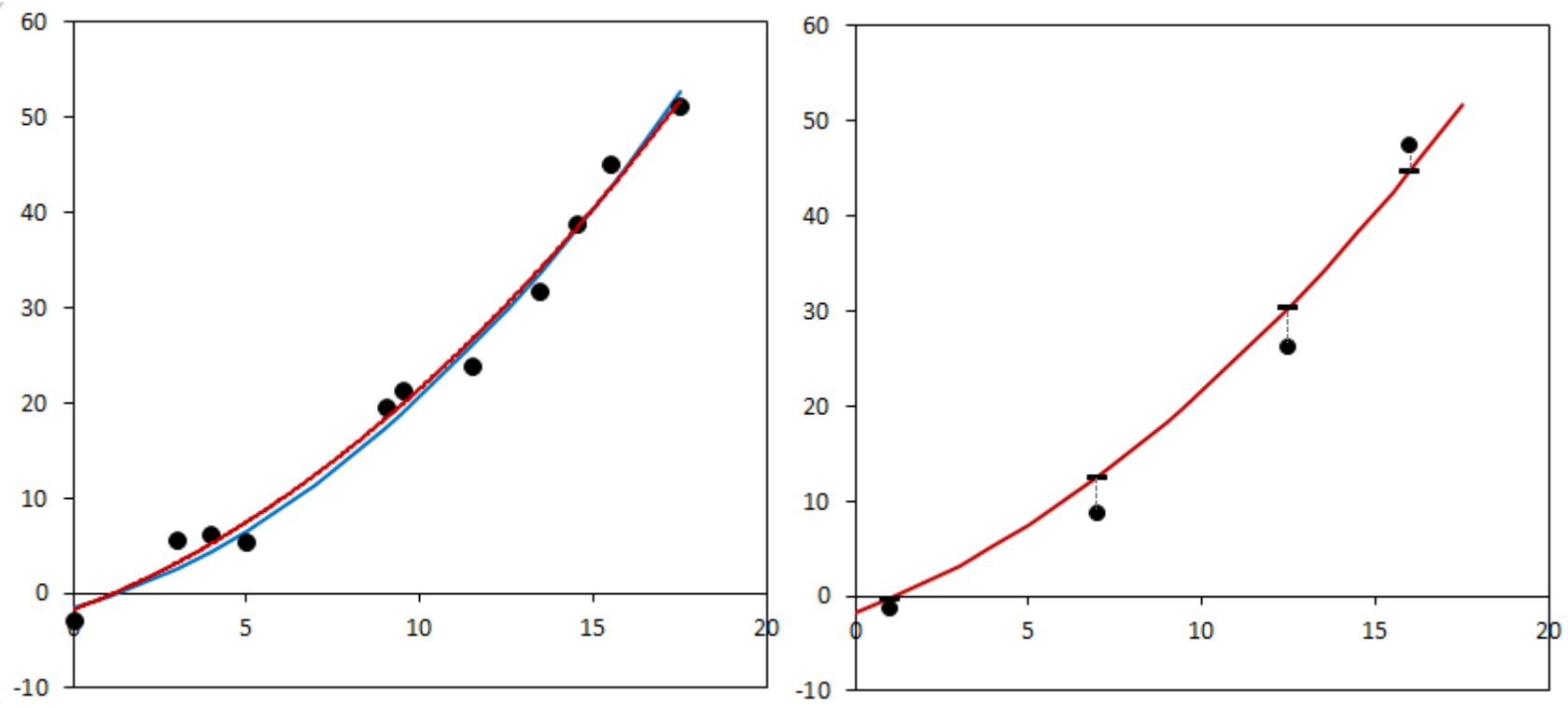
Example:

2. Compute response surface with reduced data set

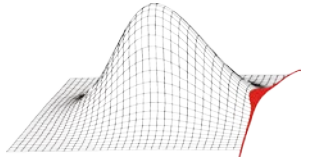


Example:

3. Use reduced model and left out samples to compute quality criteria

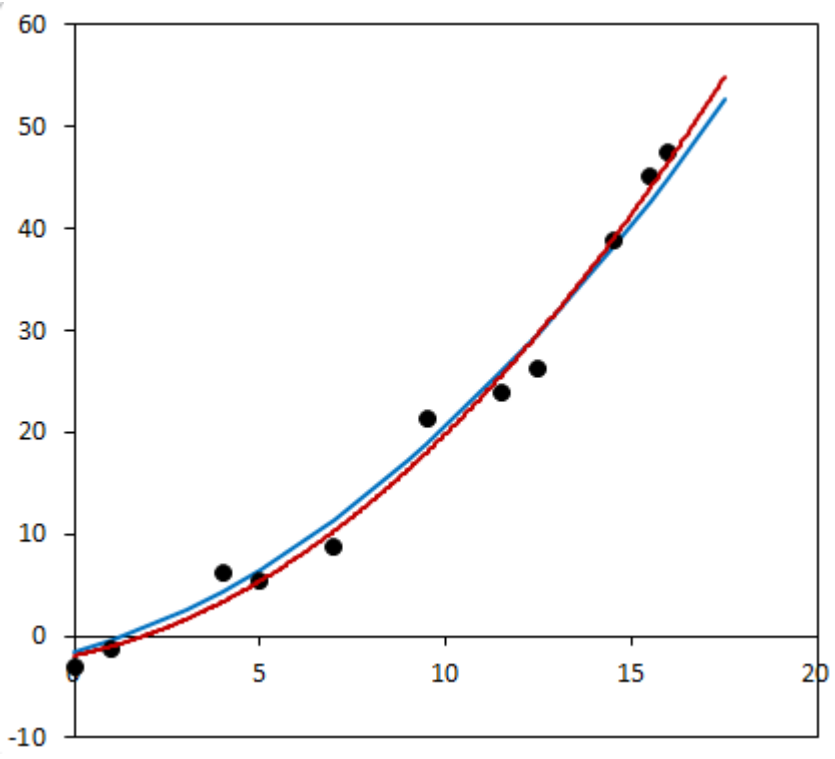


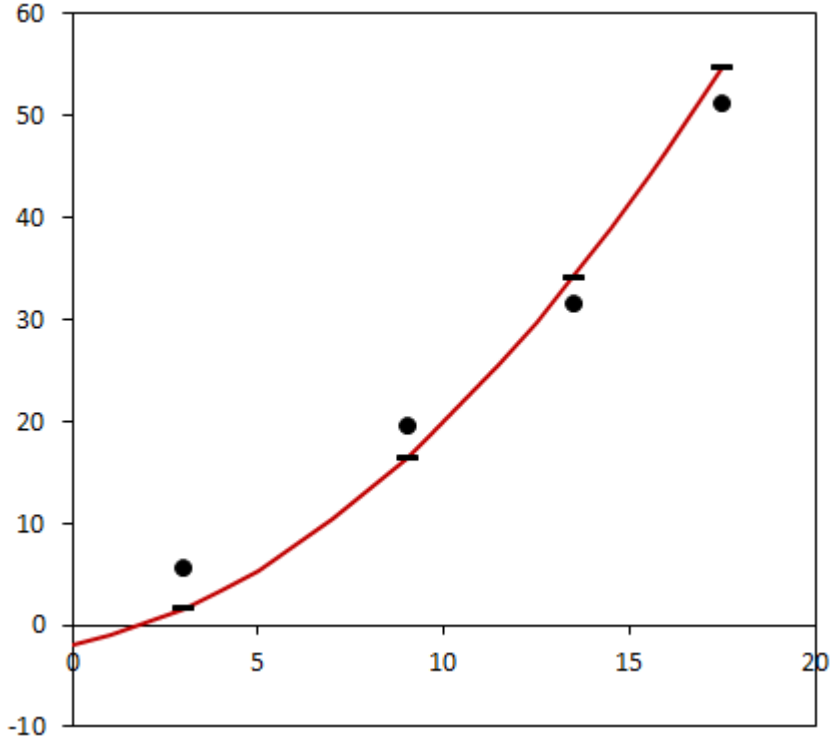
What is cross validation?



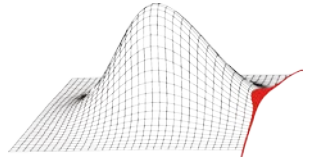
Example:

4. Repeat this procedure





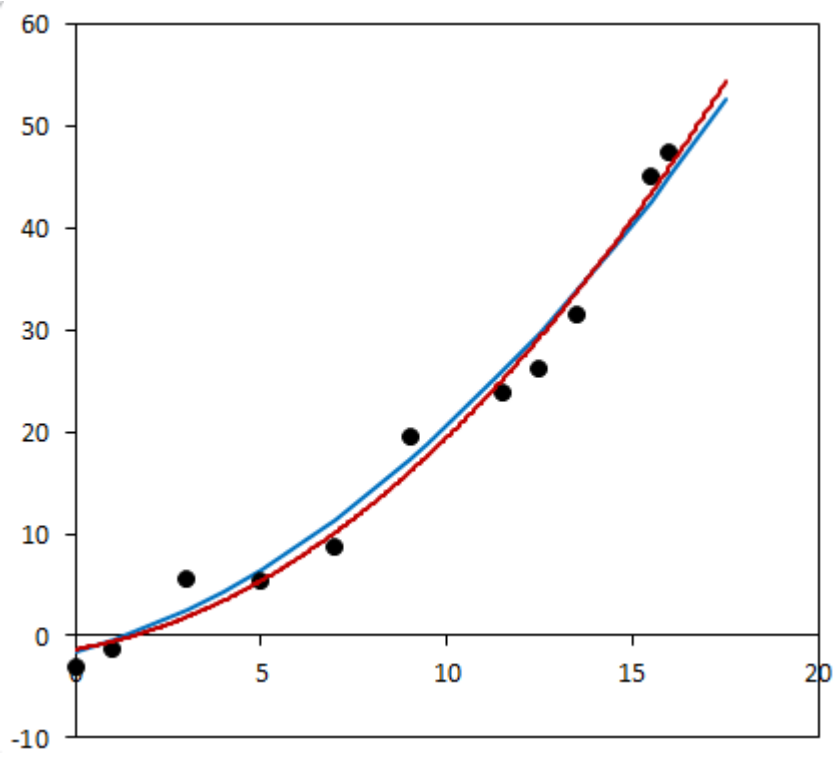
What is cross validation?

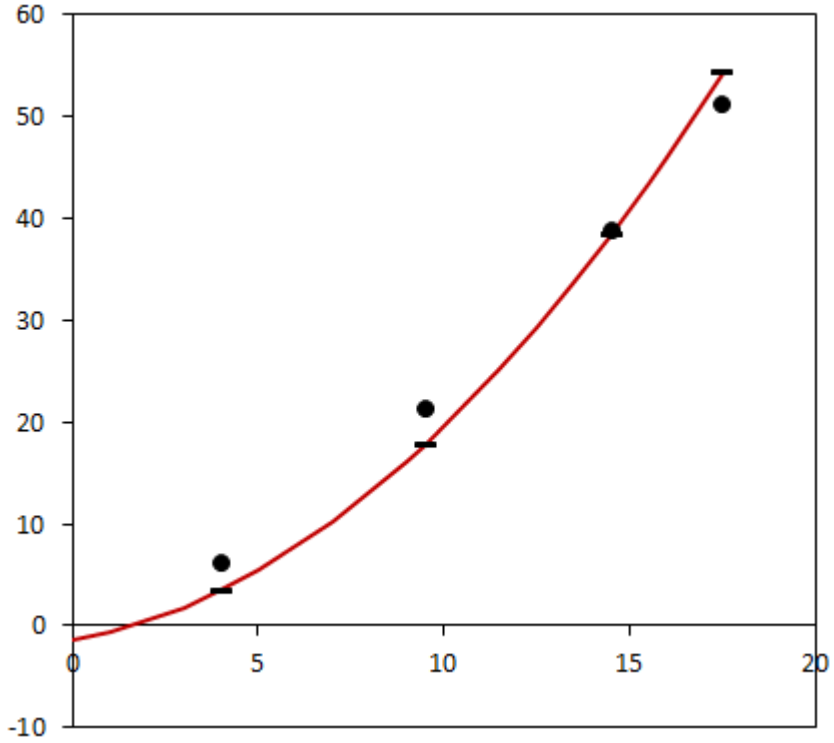


Example:

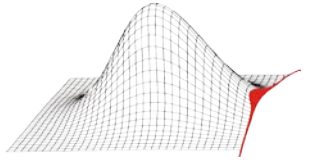
4. Repeat this procedure

   and average the validation results



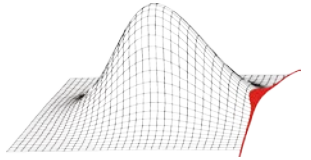


Outline



- Introduction
- Variants of cross validation
- Results of Monte Carlo cross validation
- Summary

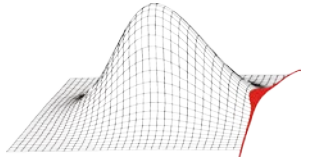
Variants of cross validation



Different kinds of data splitting:

- Leave one out cross validation (LOOCV)
  - Every single sample is used once for validation
  - Number of runs equals number of samples
- K-fold cross validation
  - Data is split into k groups
  - k validation runs that are averaged
- Monte Carlo cross validation (MCCV)
  - Samples are selected randomly
  - Number of repetitions and splitting ratio is independent selectable

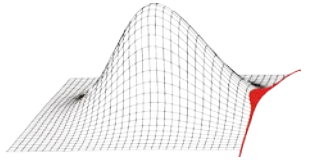
Variants of cross validation



Different kinds of quality criteria

- Error based criteria
  - e.g. PRESS – predictive error sum of squares
    - Not scaled
    - Not comparable between different data sets
- R^2 type criteria
  - Same calculation as CoD but using only validation samples
  - For LOOCV, R^2 can be computed from PRESS
  - For MCCV it is called CoD_{MCCV}

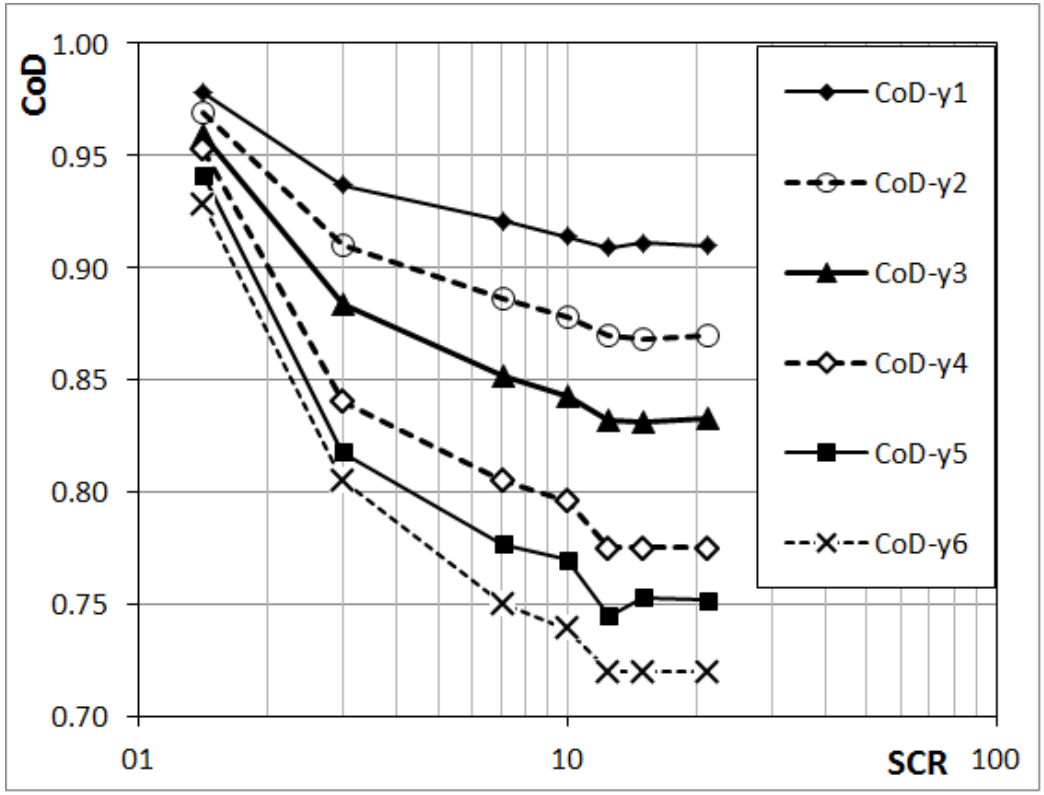
Outline



- Introduction
- Variants of cross validation
- Results of Monte Carlo cross validation
- Summary

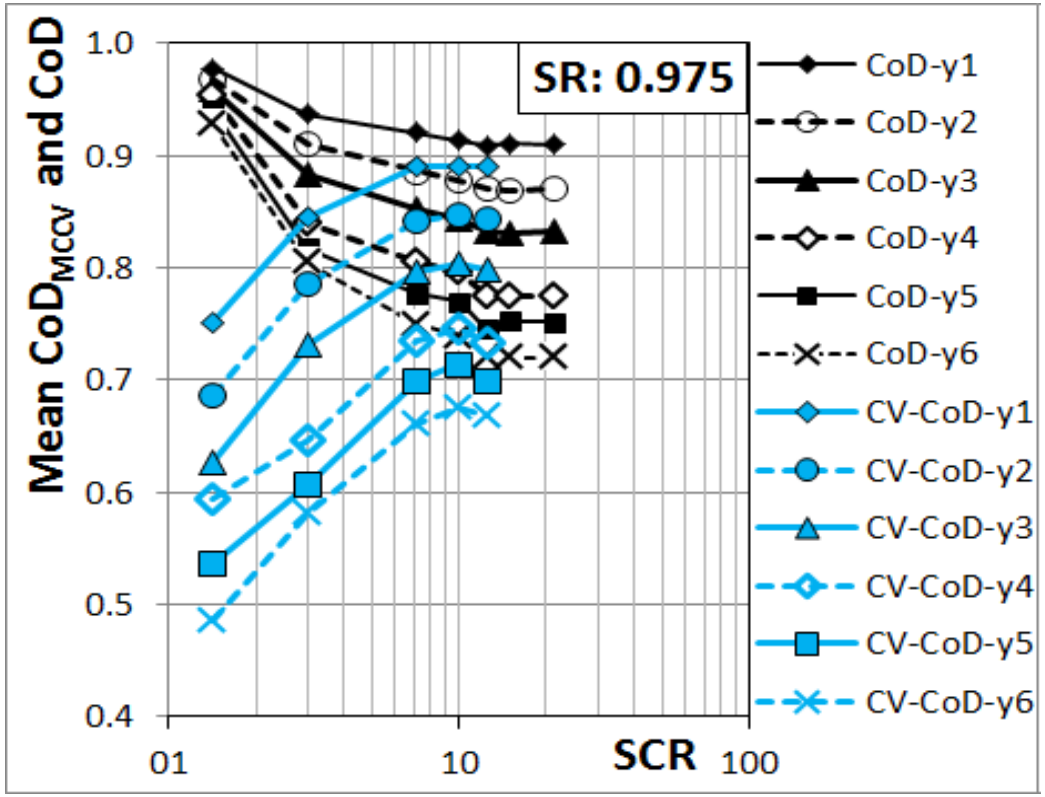
1. Quality estimation for low SCR

CoD



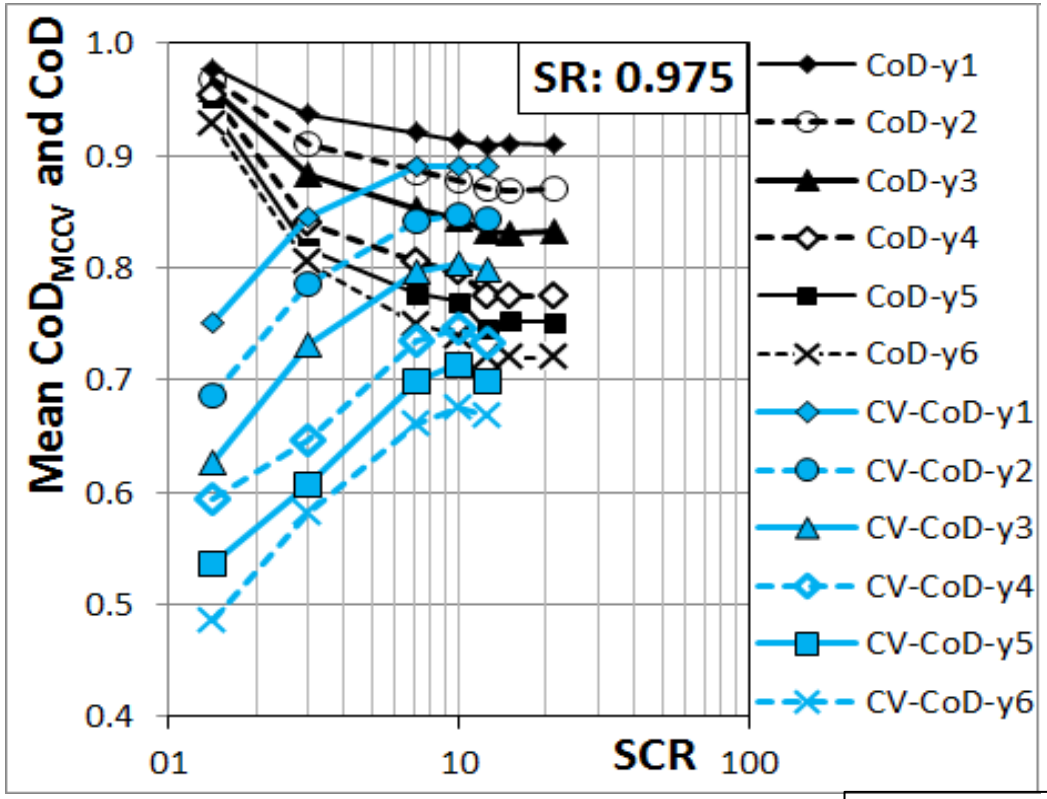
1. Quality estimation for low SCR

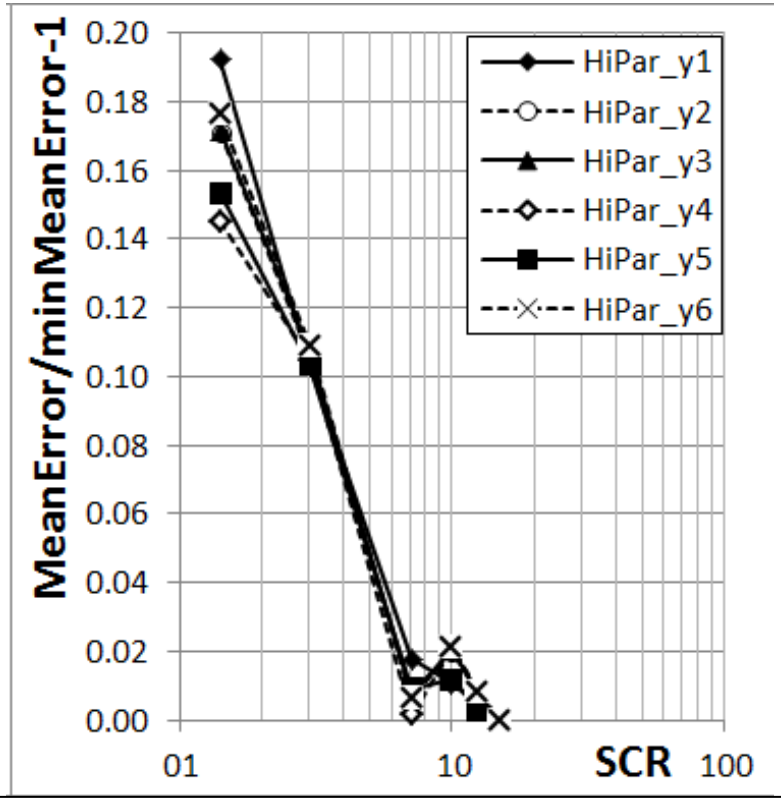
CoD vs. CoD_{MCCV}



1. Quality estimation for low SCR

CoD vs. CoD_{MCCV}

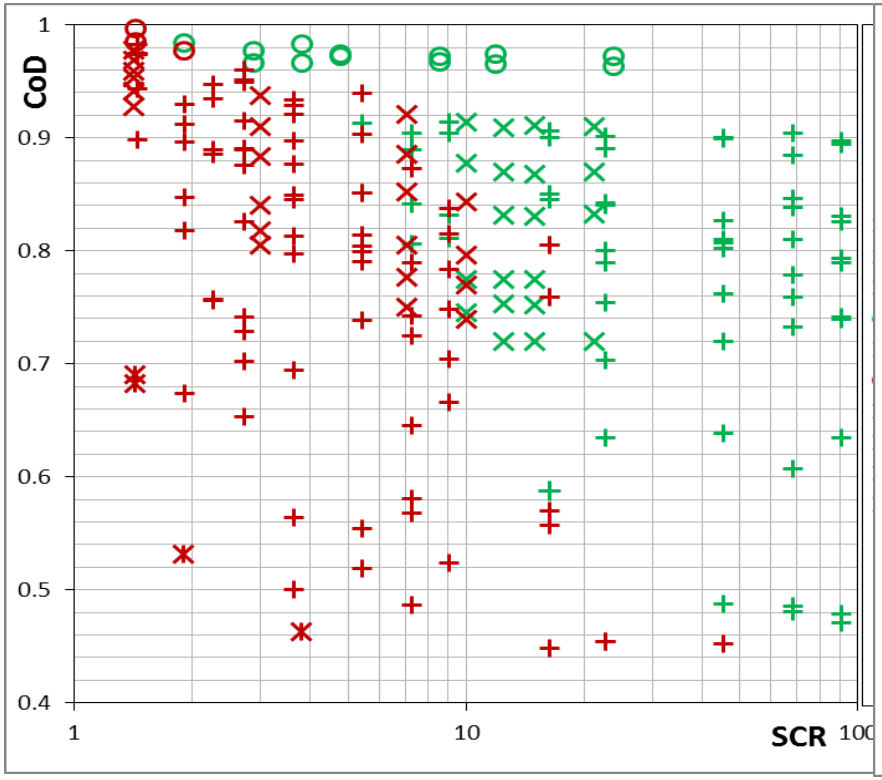


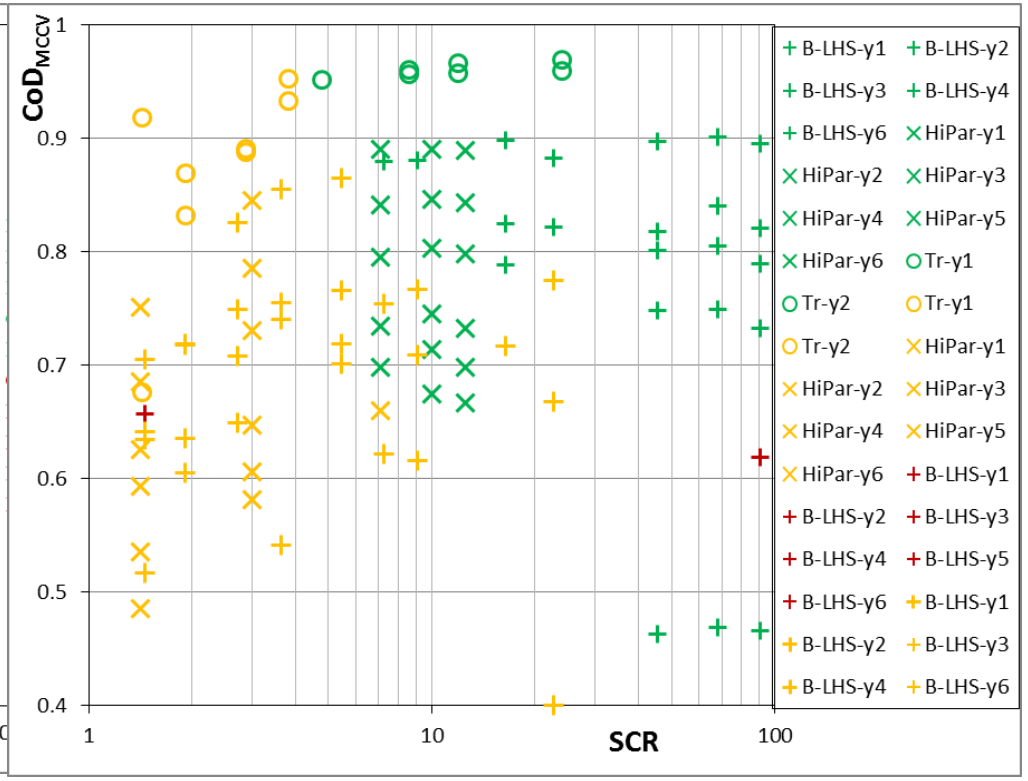


The response surfaces computed with 200 Samples are 15-19% worse than the best RS@ high SCR

1. Quality estimation for low SCR

CoD vs. CoD_{MCCV} – complete test case comparison

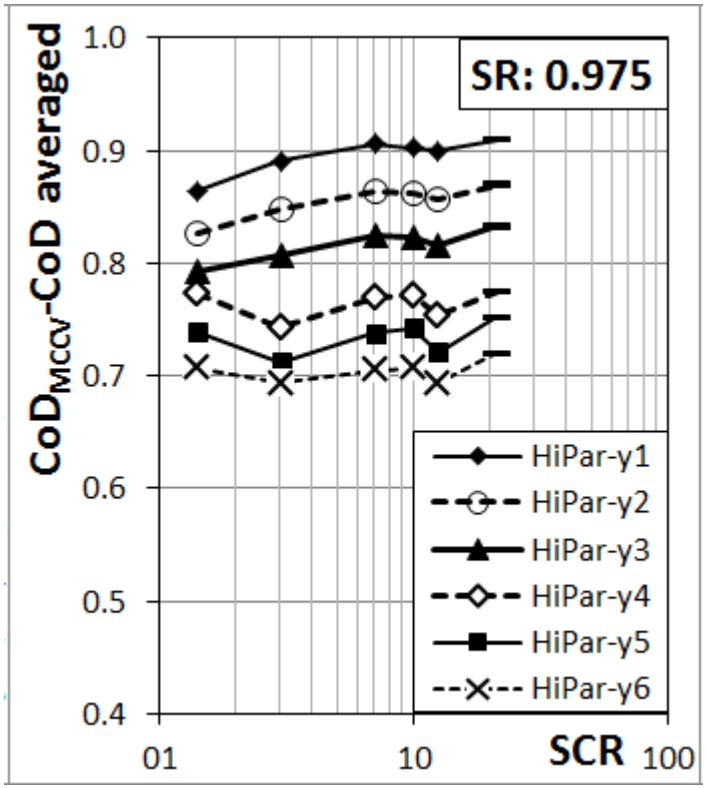




→ The CoD_{MCCV} showed no overestimation in 19/22 test cases (86%)

MCCV results

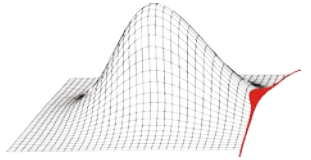
2. Ability to estimate target CoD - By averaging CoD and CoD_{MCCV}



→ Difference between this value and the current CoD_{MCCV} shows the possible improvement, if the database would be increased

→ For 91% of the test cases the target CoD was correctly predicted in the complete SCR range

MCCV results



3. Correctness in model selection

→ Comparison of 4 model quality criteria

CoD, CoD_{MCCV}, CoD_{MCCV_AV}, rme_{cv}

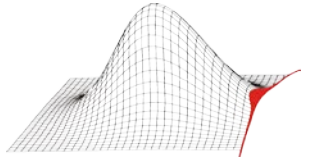
to a benchmark criteria: mes_xk

mes_xk = mean error / standard deviation, computed with x-thousand independent samples

Examples:

H5 c	CoD 60	CoD_{MCCV} 60	CoD_{MCCV_AV} 60	rme 60		CoD 100	CoD_{MCCV} 100	CoD_{MCCV_AV} 100	rme 100
1 21	0.853	0.746	0.800	0.105		0.915	0.841	0.878	0.06138
2 41	0.940	0.572	0.756	0.168		0.948	0.840	0.894	0.06140
3 61						0.962	0.748	0.855	0.085

MCCV results



3. Correctness in model selection

→ Comparison of 4 model quality criteria

CoD, CoD_{MCCV}, CoD_{MCCV_AV}, rme_{cv}

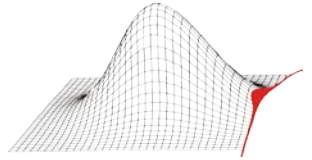
to a benchmark criteria: mes_xk

mes_xk = mean error / standard deviation, computed with x-thousand independent samples

Examples:

H5		CoD	CoD_{MCCV}	CoD_{MCCV_AV}	rme	mes_5k	CoD	CoD_{MCCV}	CoD_{MCCV_AV}	rme	mes_5k
	c	60	60	60	60	60	100	100	100	100	100
1	21	0.853	0.746	0.800	0.105	0.376	0.915	0.841	0.878	0.06138	0.366
2	41	0.940	0.572	0.756	0.168	0.396	0.948	0.840	0.894	0.06140	0.332
3	61						0.962	0.748	0.855	0.085	0.359

MCCV results



3. Correctness in model selection

→ Comparison of 4 model quality criteria

CoD, CoD_{MCCV}, CoD_{MCCV_AV}, rme_{cv}

to a benchmark criteria: mes_xk

mes_xk = mean error / standard deviation, computed with x-thousand independent samples

Examples:

H5	c	CoD 60	CoD_{MCCV} 60	CoD_{MCCV_AV} 60	rme 60	mes_5k 60	CoD 100	CoD_{MCCV} 100	CoD_{MCCV_AV} 100	rme 100	mes_5k 100
1	21	0.853	0.746	0.800	0.105	0.376	0.915	0.841	0.878	0.06138	0.366
2	41	0.940	0.572	0.756	0.168	0.396	0.948	0.840	0.894	0.06140	0.332
3	61						0.962	0.748	0.855	0.085	0.359

H5	c	CoD 300	CoD_{MCCV} 300	CoD_{MCCV_AV} 300	rme 300	mes_5k 300	CoD 2200	CoD_{MCCV} 2200	CoD_{MCCV_AV} 2200	rme 2200	mes_5k 2200
1	21	0.841	0.817	0.829	0.066	0.356	0.827	0.824	0.826	0.04673	0.343
2	41	0.905	0.857	0.881	0.057	0.287	0.900	0.896	0.898	0.03617	0.259
3	61	0.912	0.851	0.882	0.058	0.290	0.903	0.897	0.900	0.03600	0.258
2mT	231	0.977	0.569	0.773	0.117	0.426	0.941	0.925	0.933	0.03112	0.223

3. Summary of Model selection results

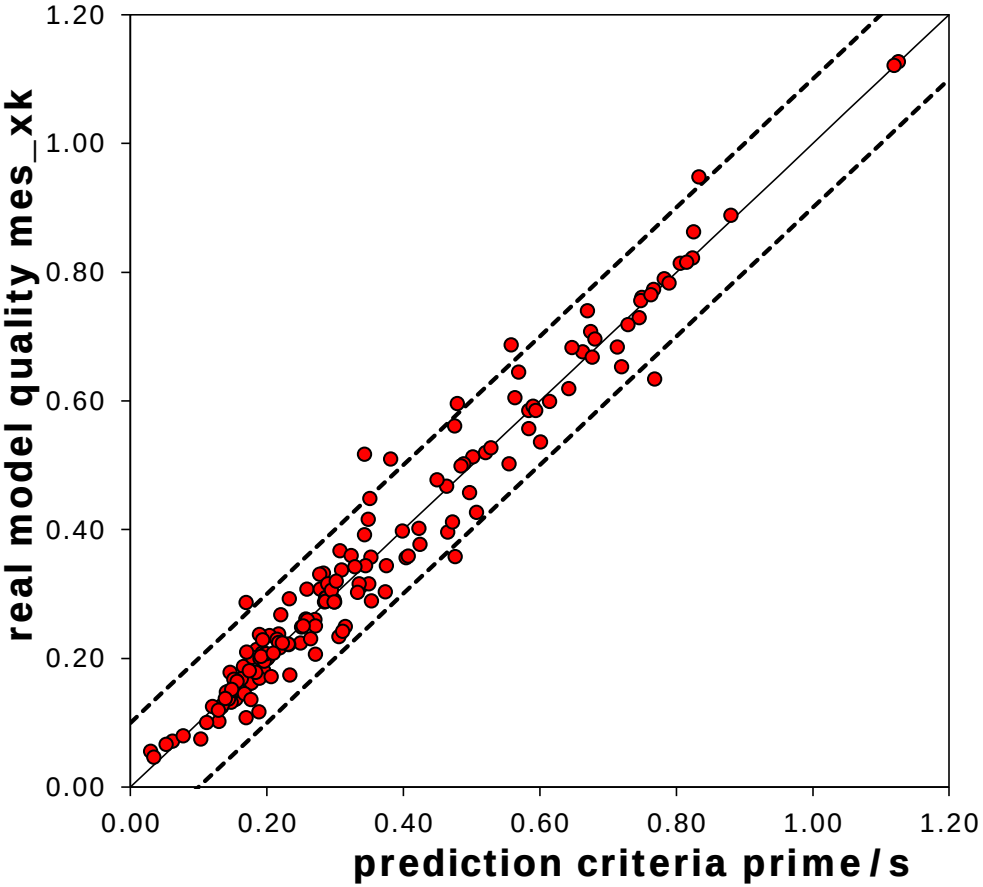
Total	Total %
CoD	0.59
CoD$_{MCCV}$	0.68
CoD$_{MCCV_AV}$	0.75
rme $_{CV}$	0.75

LOW SCR	T1	T2	T3	T4	B1	B2	B3	B4	B5	B6	H1	H2	H3	H4	H5	H6	H1	H2	H3	H4	H5	H6	H1	H2	H3	H4	H5	H6	sum	%
CoD	1	0	1	0	0	0	0	1	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0	0	0	0	0	0	12	0.43
CoD$_{MCCV}$	1	0	1	0	0	1	0	1	1	0	1	1	0	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1	17	0.61
CoD$_{MCCV_AV}$	1	1	1	0	1	1	0	1	0	0	0	1	1	1	0	1	1	0	0	0	1	0	1	1	1	1	1	1	18	0.64
rme $_{CV}$	1	1	1	0	1	1	0	1	1	0	1	1	0	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1	19	0.68
High SCR	T1	T2	T3	T4	B1	B2	B3	B4	B5	B6	H1	H2	H3	H4	H5	H6													sum	%
CoD	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1													14	0.88
CoD$_{MCCV}$	1	1	1	1	1	0	1	0	0	1	1	1	1	1	1	1													13	0.81
CoD$_{MCCV_AV}$	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1													15	0.94
rme $_{CV}$	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1													14	0.88

MCCV results

3. Improvements in model selection ability with a refined criteria

prime/s = **pr**edi**c**tive **m**ean **e**rror / **s**tandard deviation



Interpretation:

prime/s = 0.4

While predicting new data the meta model error will be about 40% of the standard deviation of the result vector

Computation:

Uses different MCCV criteria

And the internal distribution of these criteria

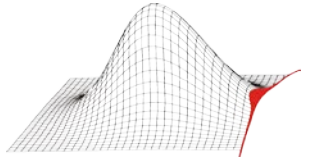
Details tbp...

3. Summary of Model selection results including prime/s

Total	%
CoD	0.59
CoD_{MCCV}	0.68
CoD_{MCCV_AV}	0.75
rme $_{CV}$	0.75
prime/s	0.89

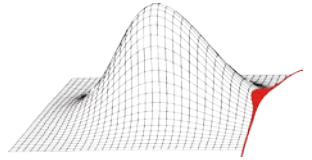
LOW SCR	T1	T2	T3	T4	B1	B2	B3	B4	B5	B6	H1	H2	H3	H4	H5	H6	H1	H2	H3	H4	H5	H6	H1	H2	H3	H4	H5	H6	sum	%
CoD	1	0	1	0	0	0	0	1	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0	0	0	0	0	0	12	0.43
CoD_{MCCV}	1	0	1	0	0	1	0	1	1	0	1	1	0	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1	17	0.61
CoD_{MCCV_AV}	1	1	1	0	1	1	0	1	0	0	0	1	1	1	0	1	1	0	0	0	1	0	1	1	1	1	1	1	18	0.64
rme $_{CV}$	1	1	1	0	1	1	0	1	1	0	1	1	0	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1	19	0.68
prime/s	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0	0	1	1	0	1	1	1	1	1	1	23	0.82
High SCR	T1	T2	T3	T4	B1	B2	B3	B4	B5	B6	H1	H2	H3	H4	H5	H6													sum	%
CoD	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1													14	0.88
CoD_{MCCV}	1	1	1	1	1	0	1	0	0	1	1	1	1	1	1	1													13	0.81
CoD_{MCCV_AV}	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1													15	0.94
rme $_{CV}$	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1													14	0.88
prime/s	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1													16	1.00

Outline



- Introduction
- Variants of cross validation
- Results of Monte Carlo cross validation
- Summary

Summary



Validation with CoD without CV

Use SCR to show reliability of CoD

CV

+ overestimation of RS quality is unusual

+ better in model selection

+ results are conform to the real model error

- Requires additional computations

MCCV

+ simple implementation

+ number of repetitions and splitting ratio is independent

- Result variance

- increasing computation time with model size and number of runs

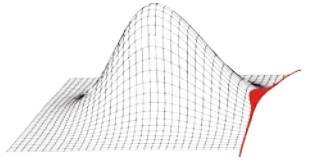
TECHNISCHE UNIVERSITÄT DRESDEN

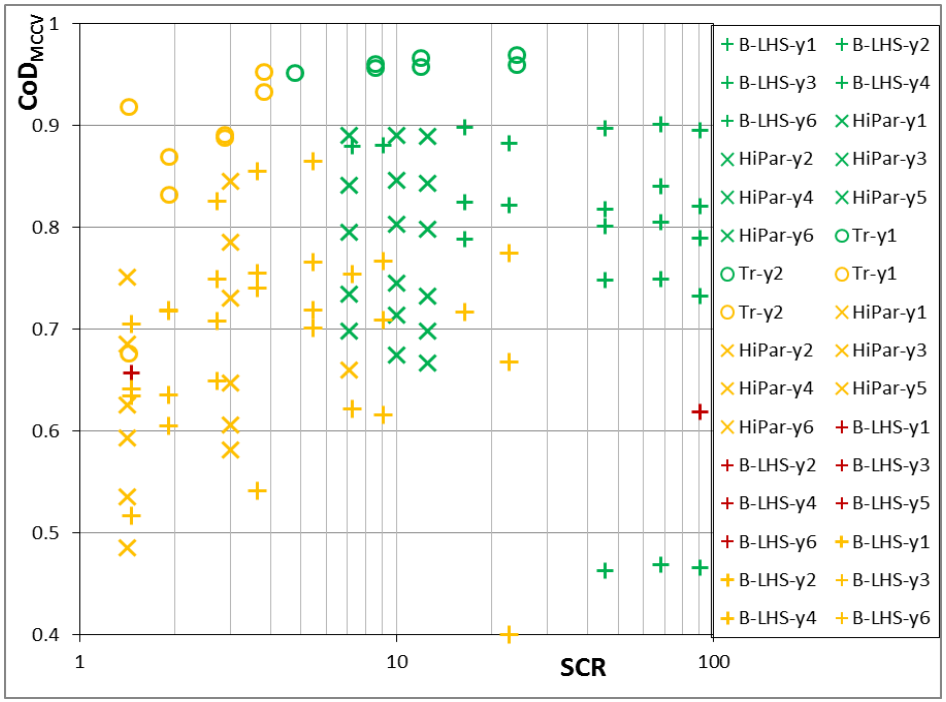
Faculty of Mechanical Science and Engineering, Institute of Fluid Mechanics, Chair of Turbomachinery and Jet-Propulsion

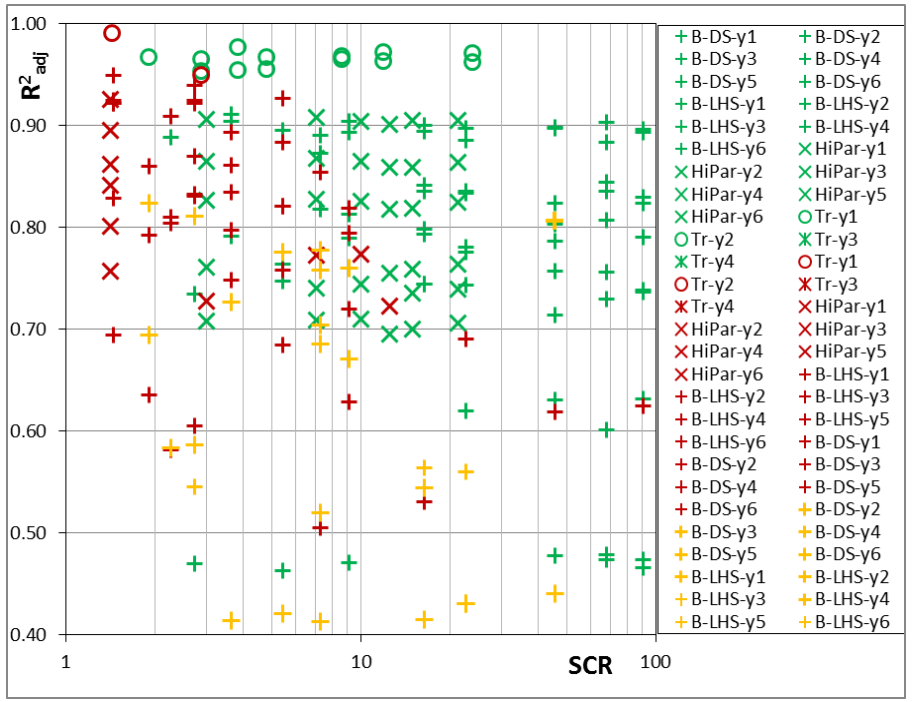
Monte Carlo cross validation for response surface benchmark

André Beschorner*, Matthias Voigt, Konrad Vogeler
*andre.beschorner@mailbox.tu-dresden.de

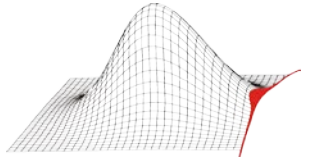
Backup – R^2_{adj}

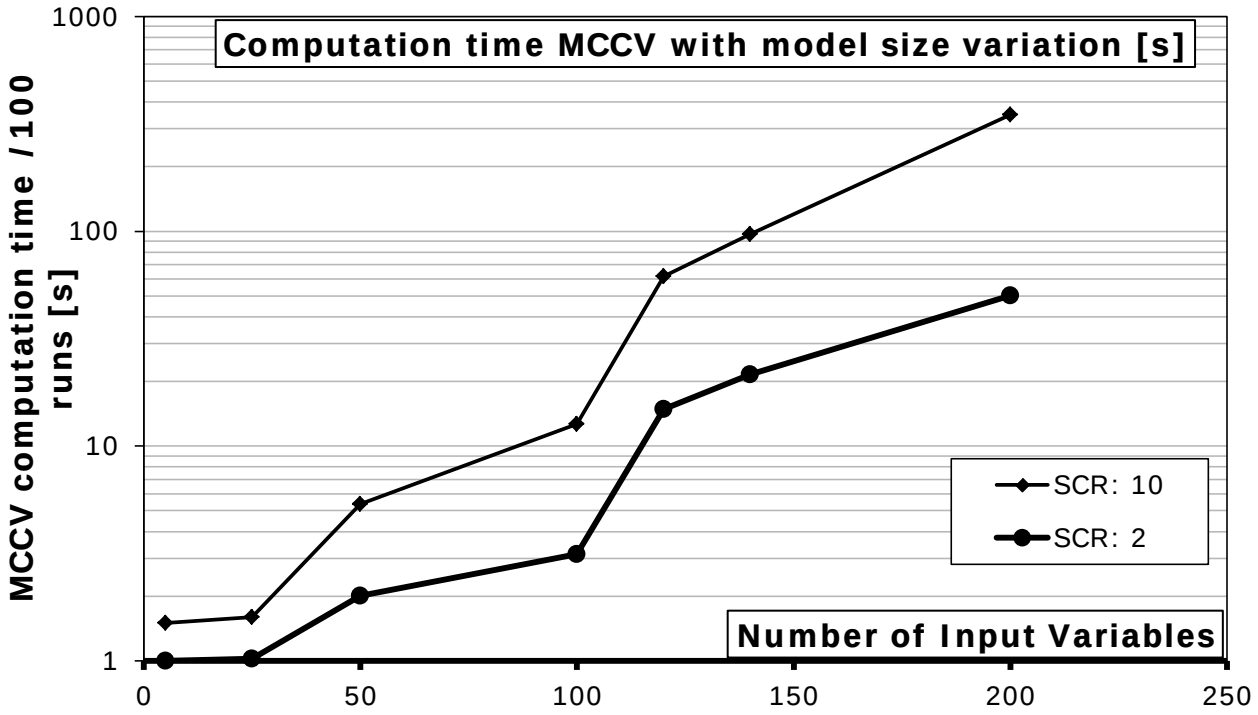




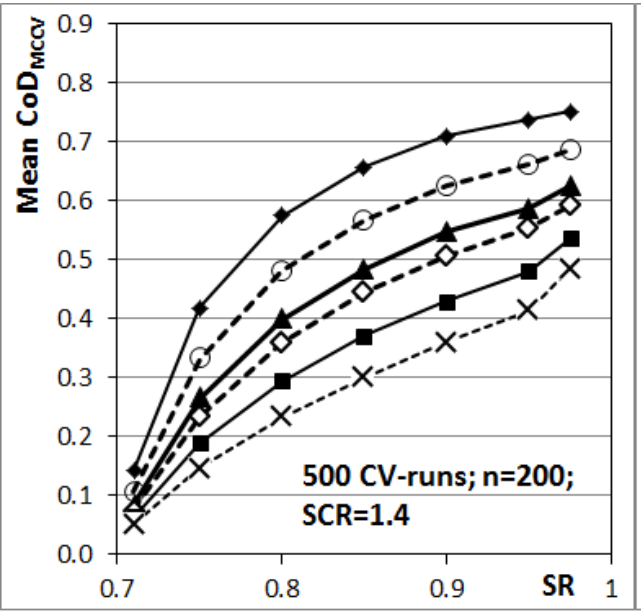


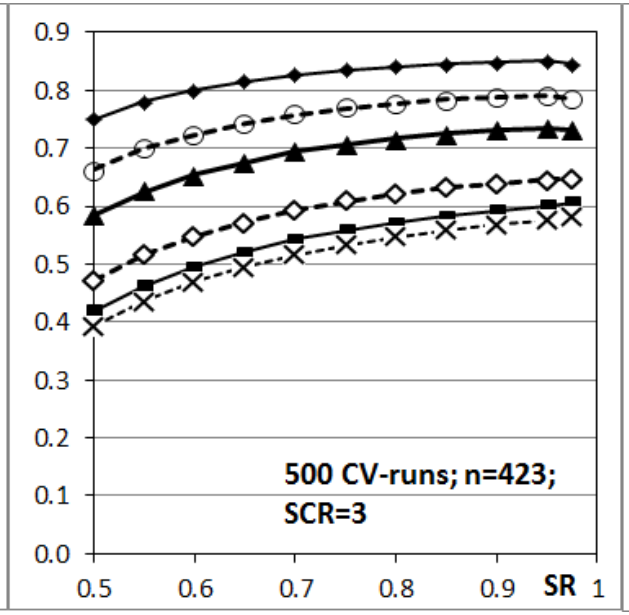
Backup – computation time

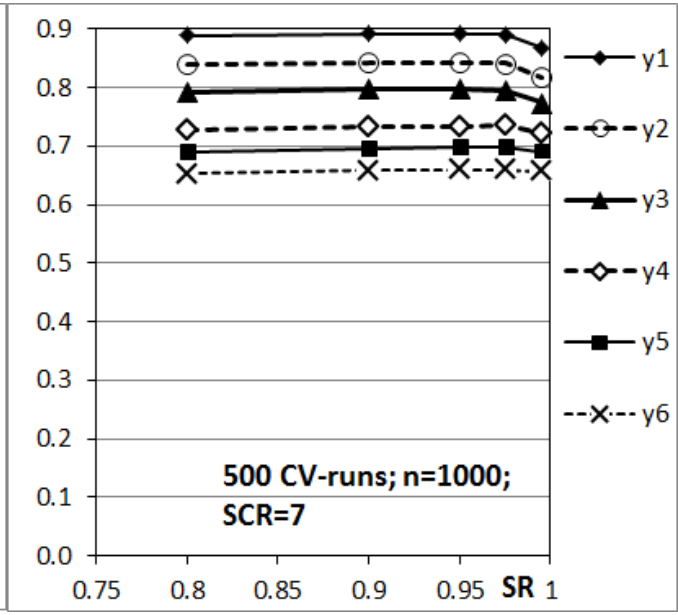




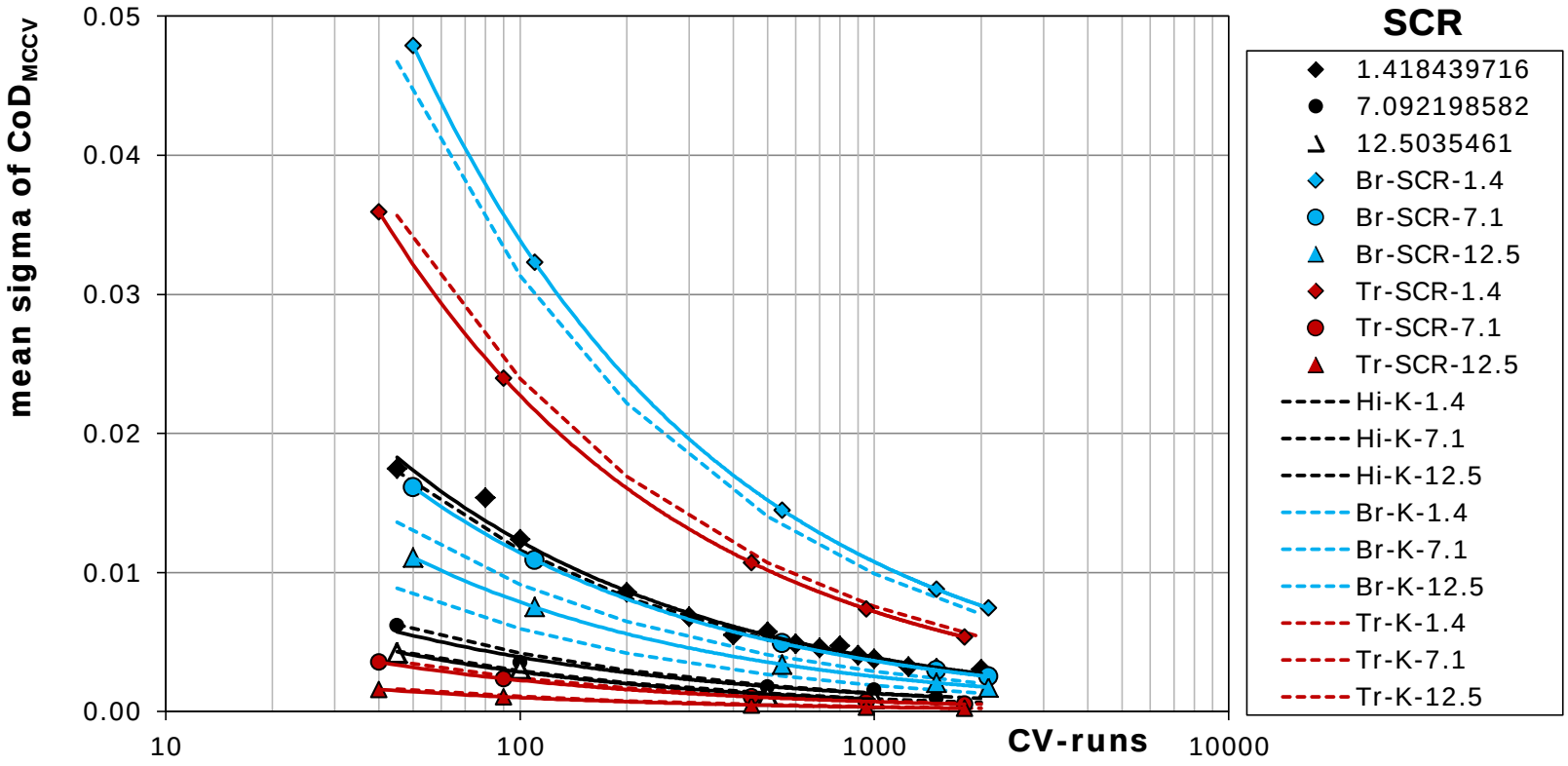
Backup – splitting ratio







Backup – variance of result



$$\sigma = \frac{[-0.002 \cdot C + 0.4265] \cdot SCR^{[1.5063 \cdot \frac{E \cdot O}{C} - 2.1283]}}{\sqrt{\text{CV-runs}}}$$