

# Bachelorarbeit: Modellselektion und Identifikation

basierend auf mehreren Datensätzen mit Anwendung auf  
Luftpfad-Modellbildung

Betreuung: DI Patrick Schrangl

Student: Philipp Fleischanderl

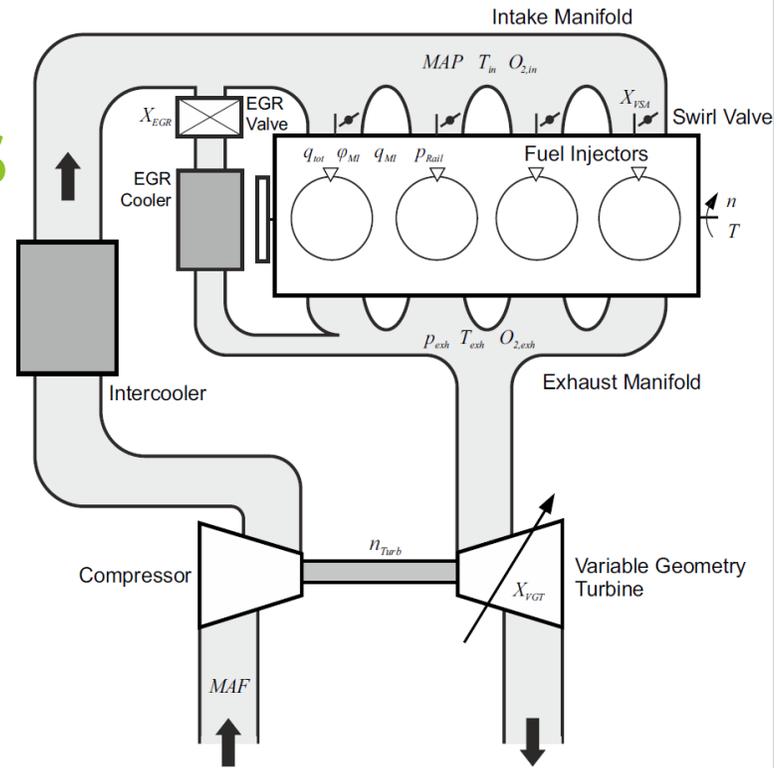
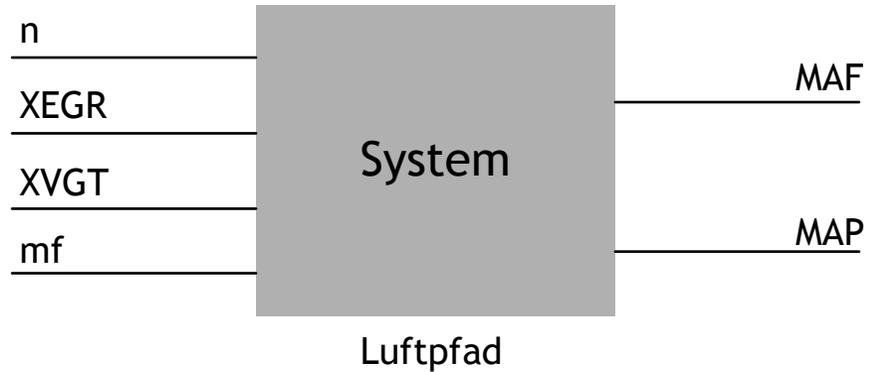
# Übersicht

- ▶ Beschreibung der Ausgangslage
- ▶ Bewertungskriterium für geeignete Modellgüte
- ▶ Wie eignet sich Cross-Validation zur Bewertung der Modellgüte
- ▶ Haben unterschiedliche Datensätze Einfluss auf die Modellgüte
- ▶ Gibt es Unterschiede zwischen Prädiktions- und Simulationsfehler

# Ausgangslage

- ▶ Beliebiges System von den man ein Modell haben möchte
- ▶ Identifikations - und Validationsdatensatz
- ▶ Festlegen einer Modellstruktur, z.B. NARX-Modelle
- ▶ Bewertung der Güte, z.B. FIT-Wert
- ▶ Problem: Es sind nur zwei Realisierungen der Datensätze

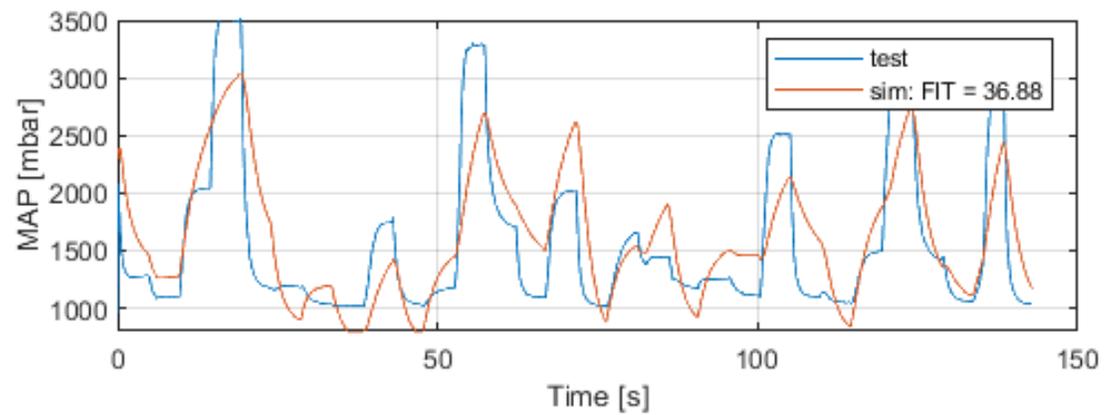
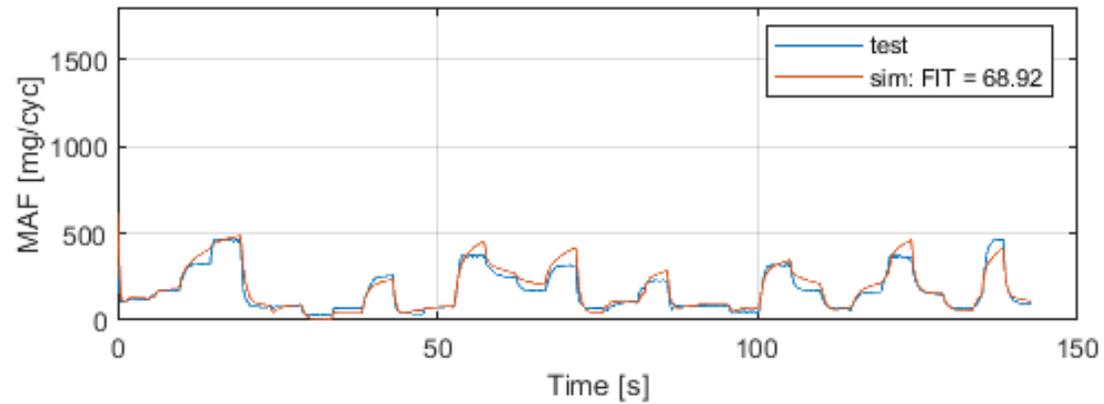
# Darstellung des Systems



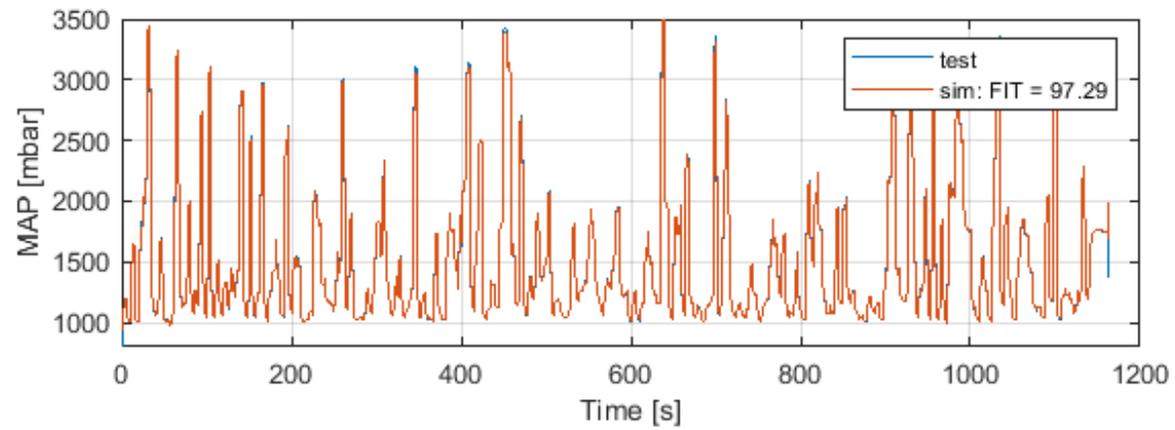
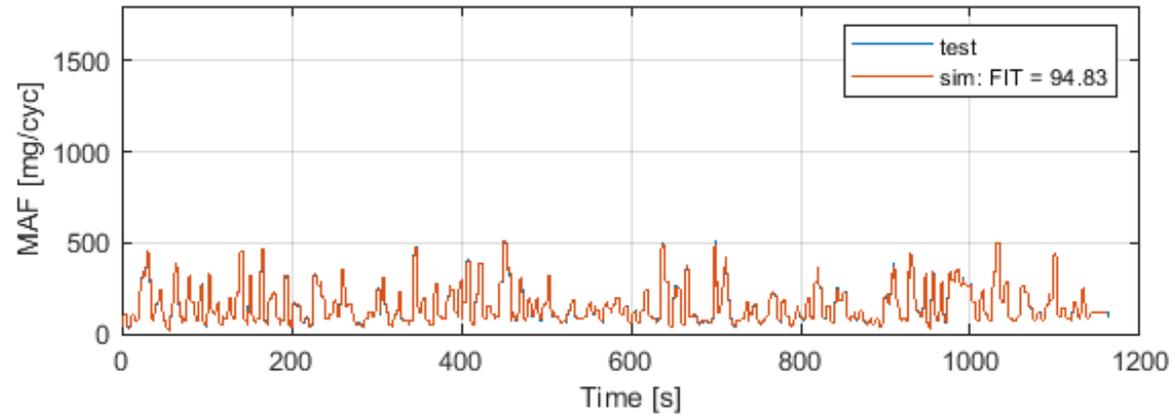
Input	Symbol	Unit
Engine speed	$n$	min-1
Exhaust gas recirculation value position	$X_{EGR}$	%
Variable geometry turbine guide vane position	$X_{VGT}$	%
Fuel injection amount per cycle	$mf$	mg

Output	Symbol	Unit
Intake manifold air mass flow per cycle	$MAF$	mg
Intake manifold air pressure	$MAP$	mbar

# Ausgangslage

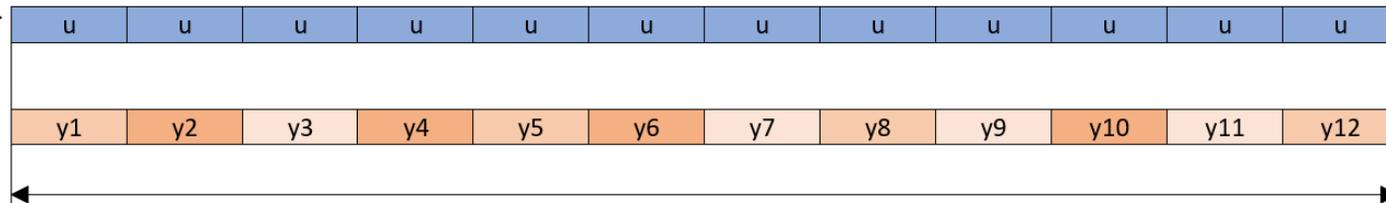


# Ausgangslage

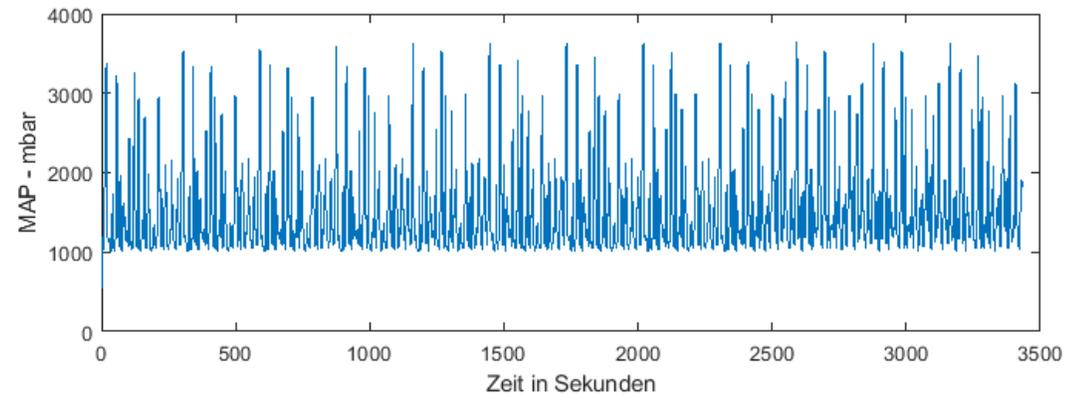
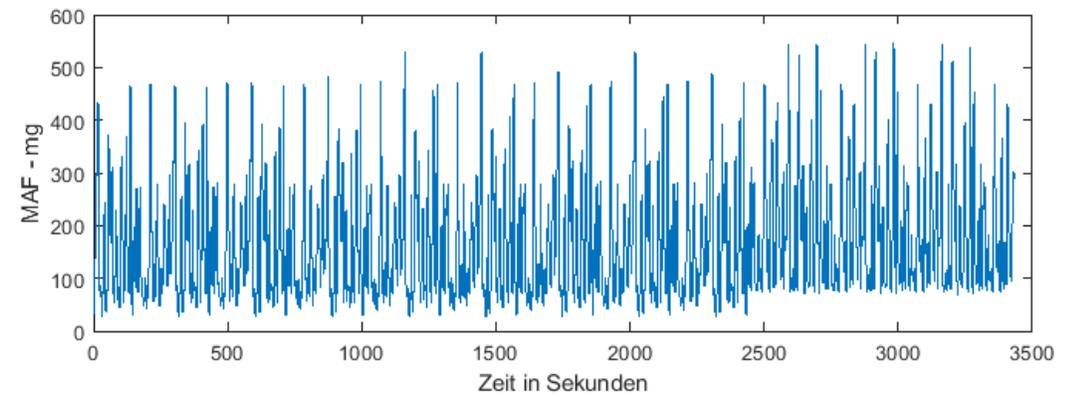
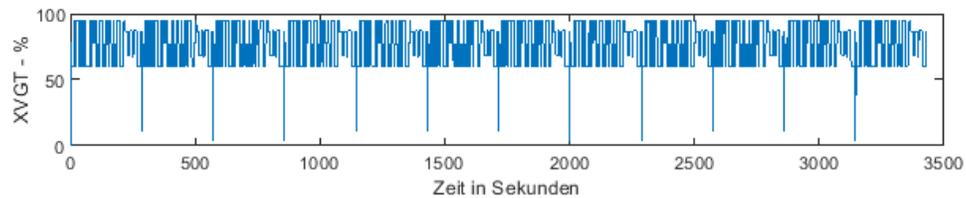
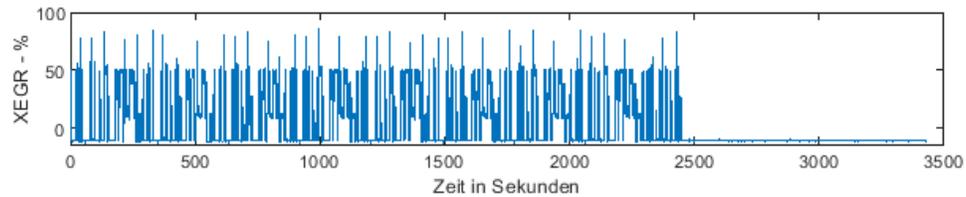
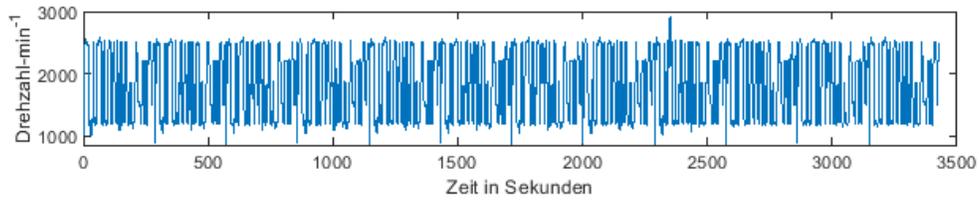


# Abhilfe

- ▶ Abschätzung des stochastischen Anteils
- ▶ Durch mehrmaliges Anlegen desselben Eingangs



# Aufgenommene Messdaten - wiederholender DS



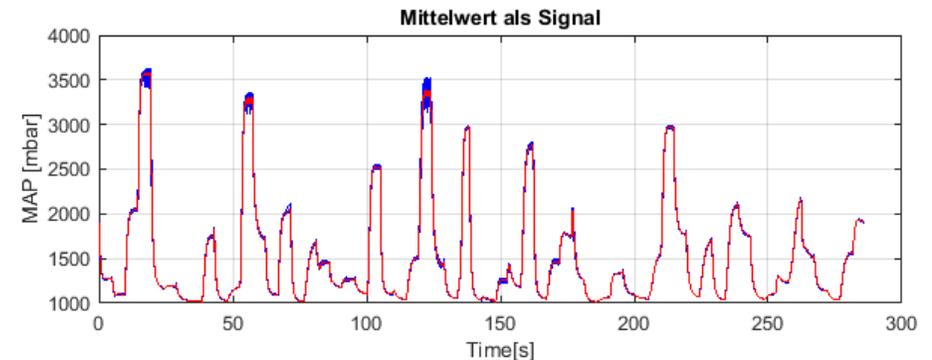
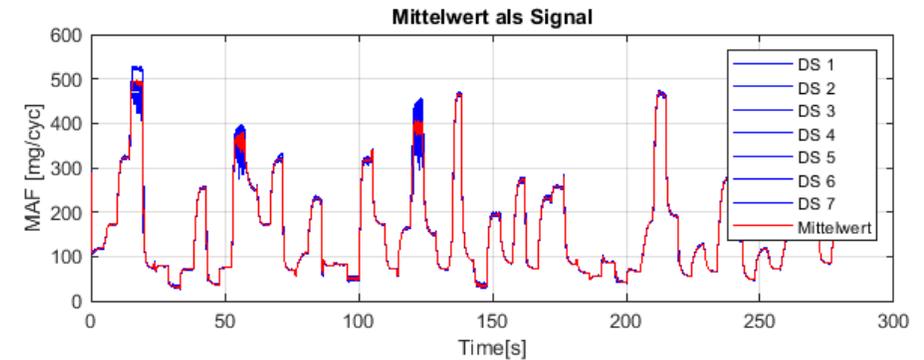
# Ermittlung des stochastischen Anteils

- Schätzung einer Realisierung:

$$\hat{E}_i = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_i)^2}}$$

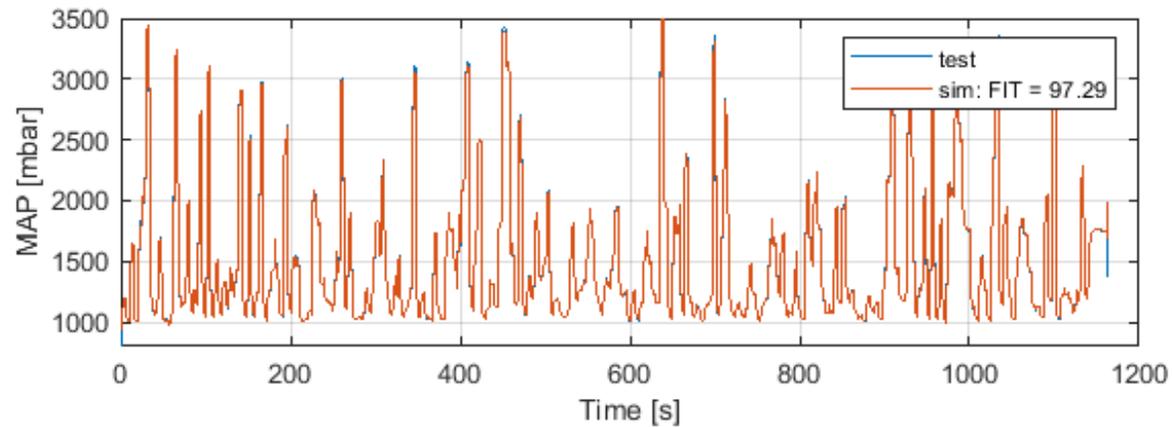
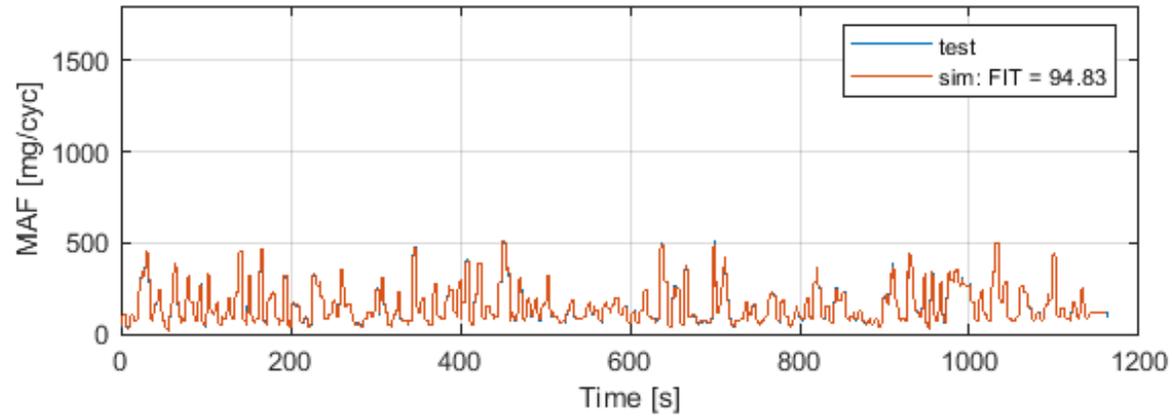
- Schätzung des stochastischen Anteils:

$$\hat{E} = \frac{1}{n} \sum_{i=1}^n \hat{E}_i$$

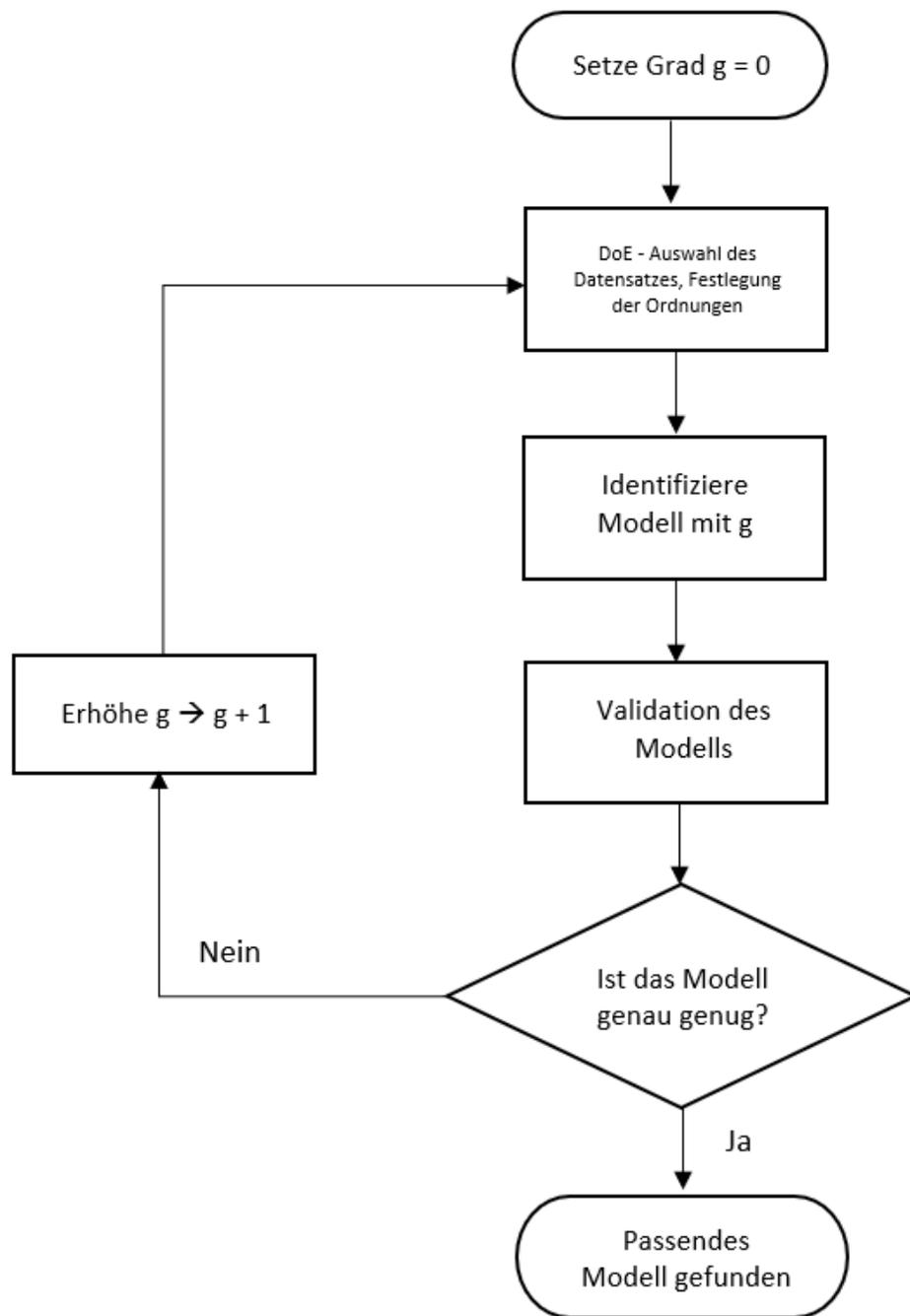


MAF $\hat{E}_i$ [%]	MAP $\hat{E}_i$ [%]
6,1	3,4
5,7	2,8
4,9	2,5
6,6	3,2
4,9	2,4
4,1	2,6
5,9	3,4
Ø5,5	Ø2,9

# Vergleich

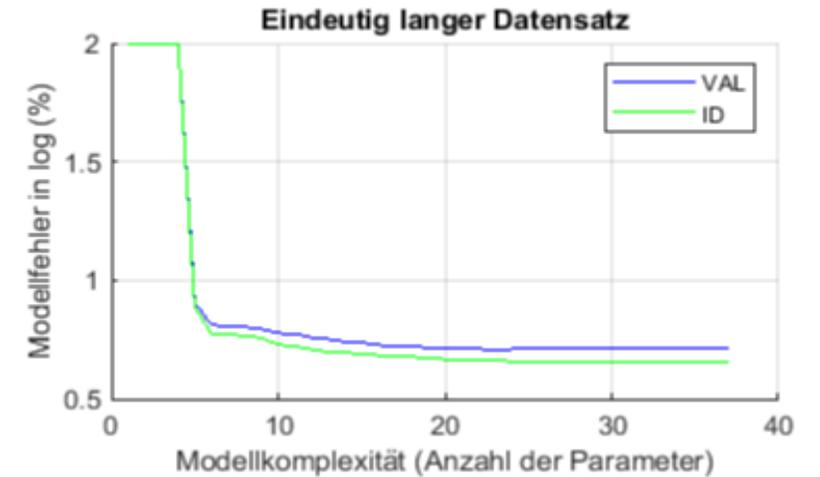


MAF $\bar{E}_i$ [%]	MAP $\bar{E}_i$ [%]
6,1	3,4
5,7	2,8
4,9	2,5
6,6	3,2
4,9	2,4
4,1	2,6
5,9	3,4
Ø5,5	Ø2,9

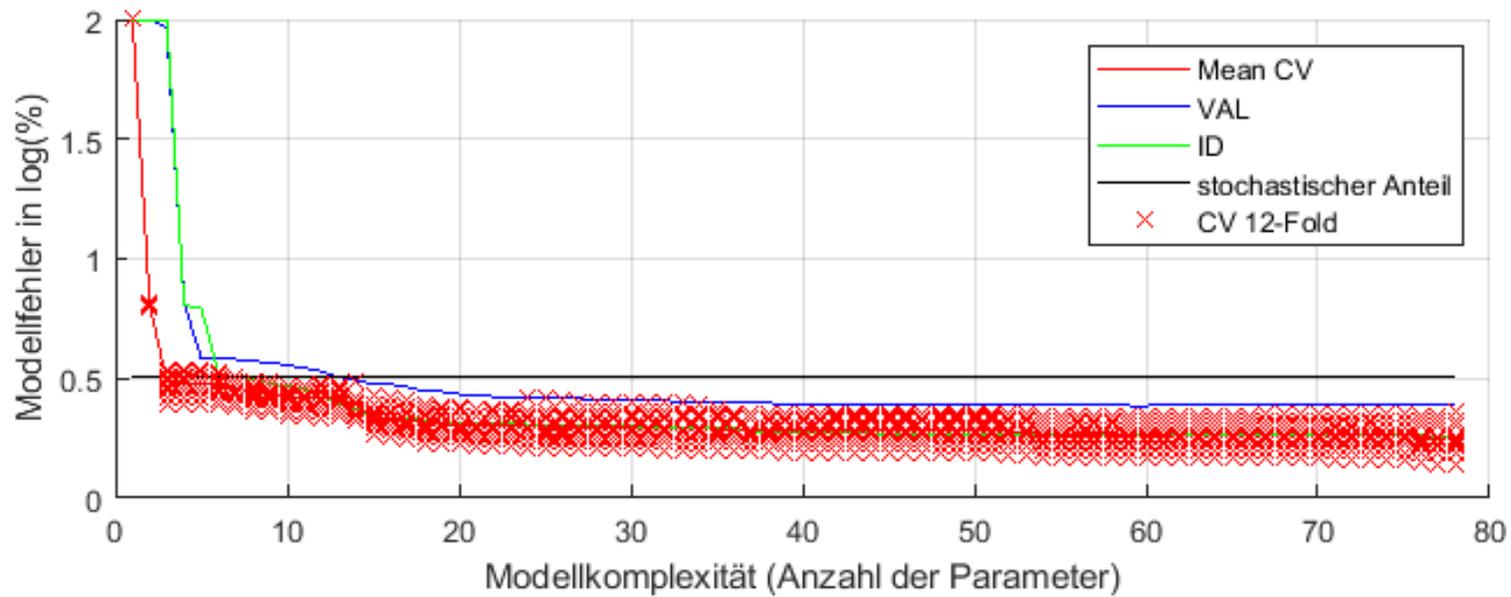
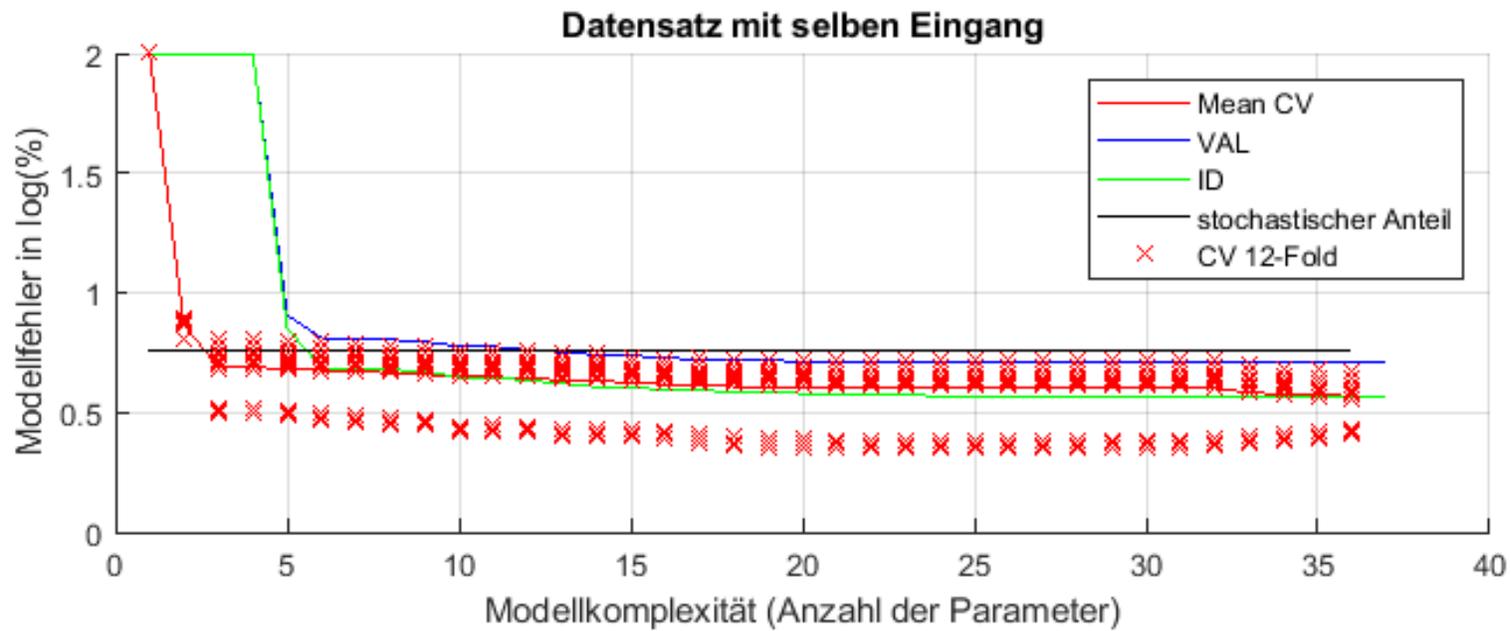


# Kann man Cross-Validation für die Regressorselektion verwenden?

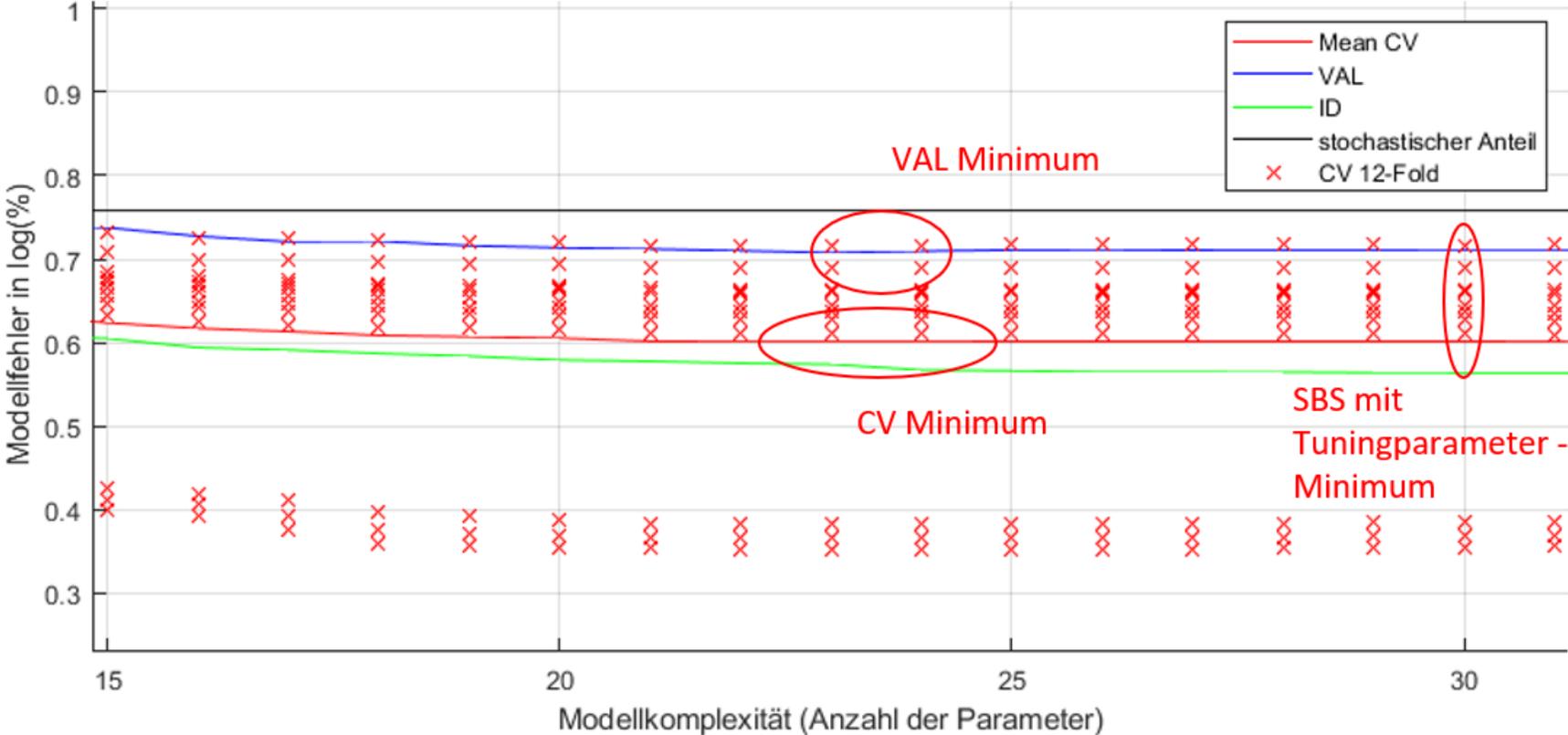
- ▶ Datensatz gute Struktur für CV?
- ▶ Liefert CV eine gute Abschätzung des Modellfehlers
- ▶ Vergleich mithilfe Stepwise Backward Selection



Val.	Id.										
Id.	Val.	Id.									
Id.	Id.	Val.	Id.								
Id.	Id.	Id.	Val.	Id.							
Id.	Id.	Id.	Id.	Val.	Id.						
Id.	Id.	Id.	Id.	Id.	Val.	Id.	Id.	Id.	Id.	Id.	Id.
Id.	Id.	Id.	Id.	Id.	Id.	Val.	Id.	Id.	Id.	Id.	Id.
Id.	Val.	Id.	Id.	Id.	Id.						
Id.	Val.	Id.	Id.	Id.							
Id.	Val.	Id.	Id.								
Id.	Val.	Id.									
Id.	Val.										

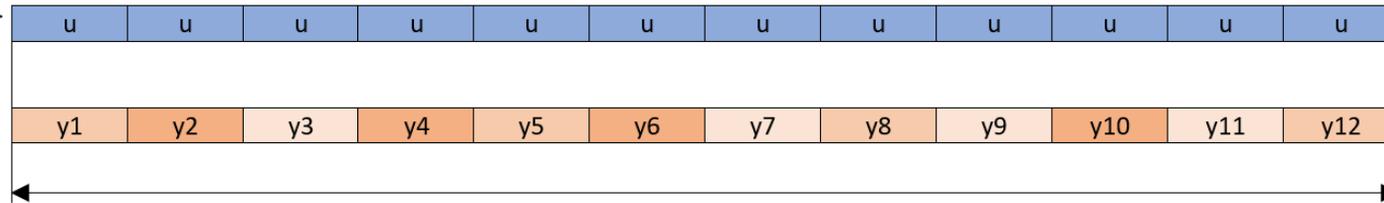


### Datensatz mit selben Eingang

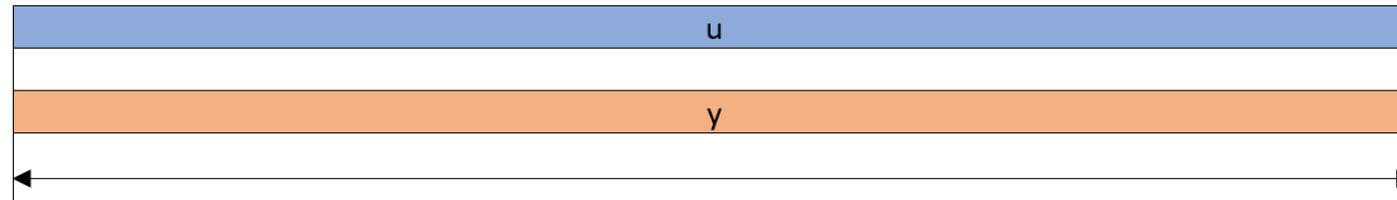


# Gibt es bessere Datensätze?

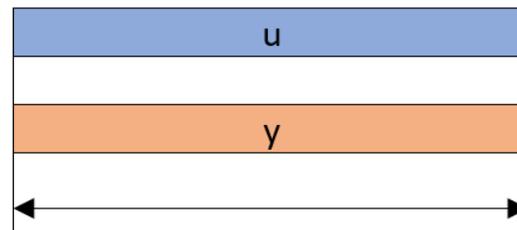
- ▶ Datensatz mit wiederholende Eingangssequenzen:



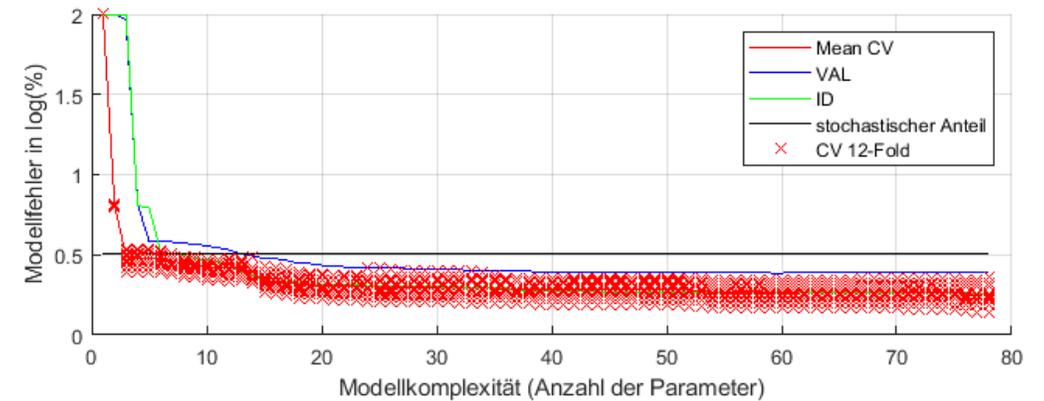
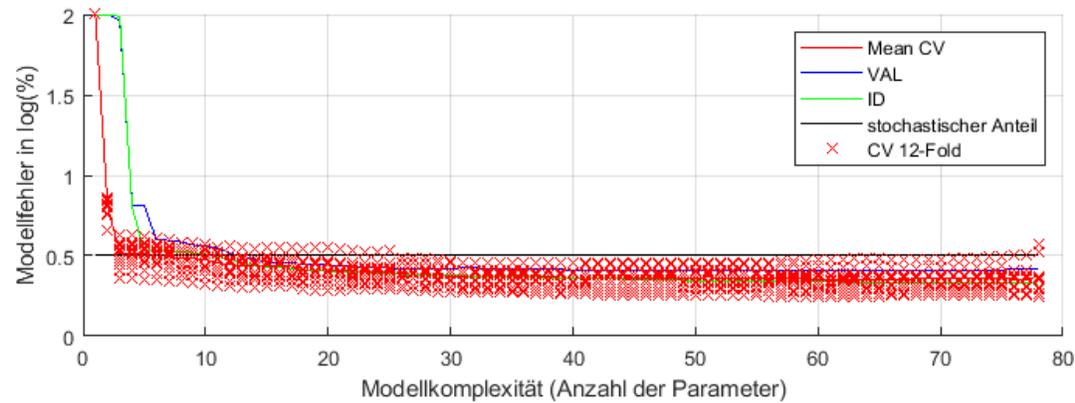
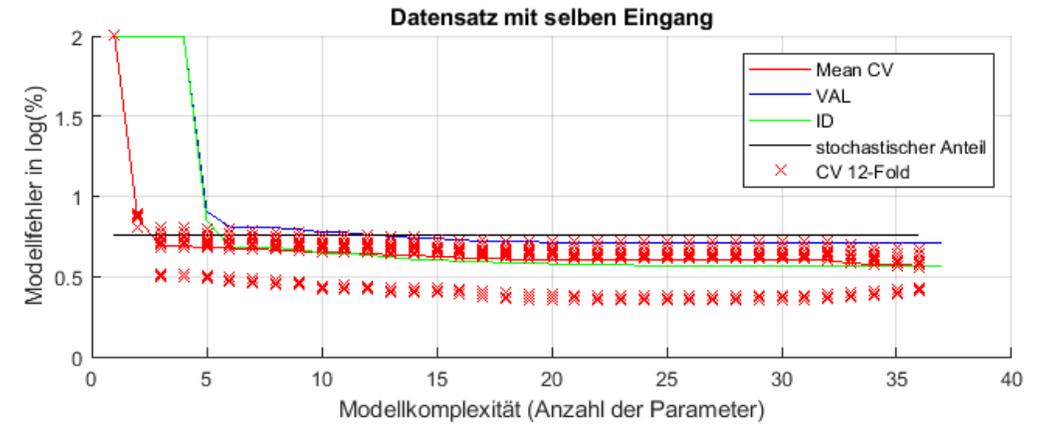
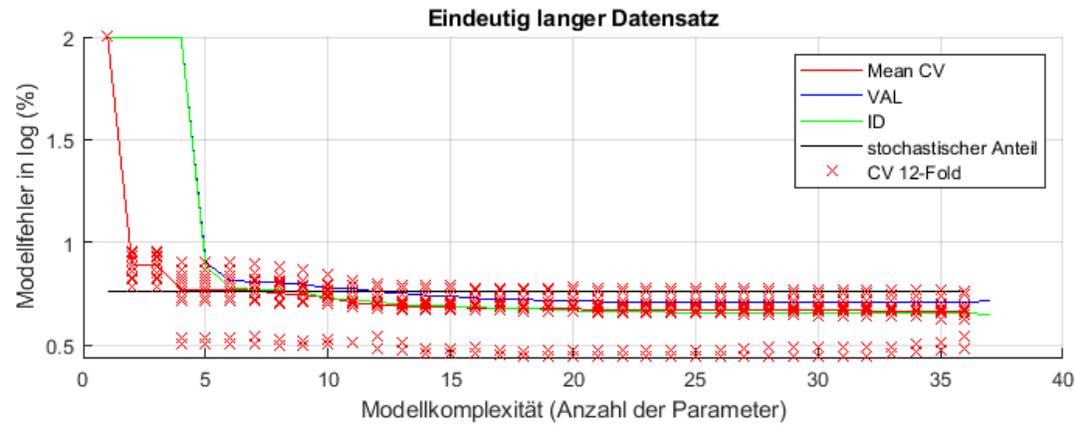
- ▶ Eindeutig langer Datensatz:



- ▶ Validationsdatensatz:



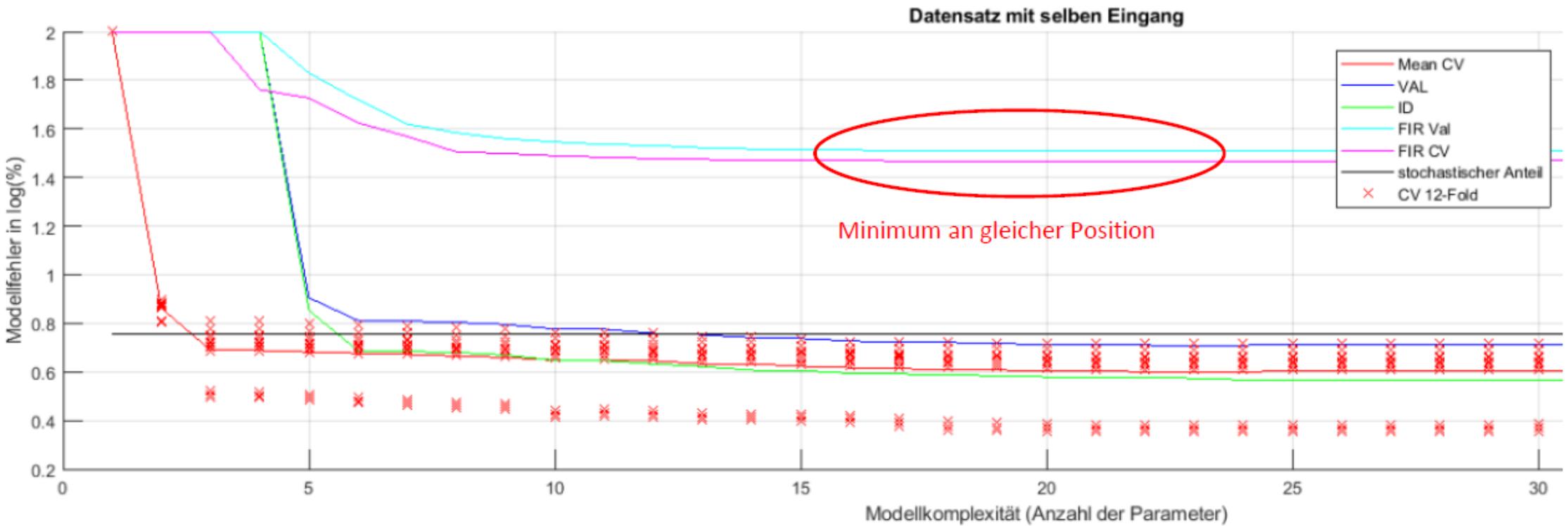
# Vergleich



# Macht es einen Unterschied ob man Prädiktions- oder Simulationsfehler für die Bewertung der Modellgüte verwendet?

- ▶ Wegen Rechenaufwand wurde auf FIR - Modelle ausgewichen
- ▶ FIR-Modelle sind eine spezielle Form von ARX Modellen
- ▶ Der vergangene Ausgang wird nicht verwendet
- ▶ Prädiktionsfehler = Simulationsfehler

Modellstruktur	Prädiktions-FIT MAF/MAP[%]	Simulations-FIT MAF/MAP [%]
FIR	70,8124/69,6521	70,5618/69,1725
PNARX	94,8315/97,2901	70,8954/72,9369



# Zusammenfassung

- ▶ Der Stochastischer Anteil ist nützlich für Bewertung einer geeigneten Modellgüte, Ermittlung durch mehrmaliges Anlegen des selben Eingangs
- ▶ CV liefert eine gute Abschätzung der wahren Validation
- ▶ Die Arten der Datensätze beeinflussen die Modellgüte nicht
- ▶ Bei der Regressorselektion ist es unerheblich ob man Simulations- oder Prädiktionsfehler verwendet