# ML-unterstützte Entwurfsmethoden zum Finden erlaubter Eingabeparameter Am Beispiel von Crashtesting

Verteidigung Masterarbeit – Julian Haluska

24.03.2020



#### Beispiel: Crashsimulationen

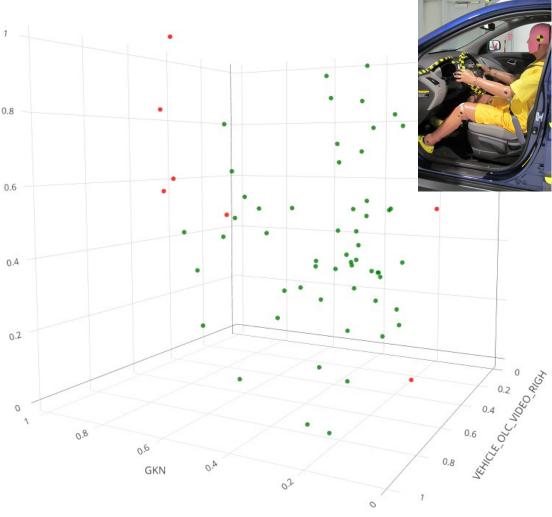
Ingenieure optimieren Designs bzgl. Gewicht, Kosten etc.

Dabei werden Werte für Designparameter festgelegt

Falls die Systemantwort über einem bestimmten Grenzwert ist:

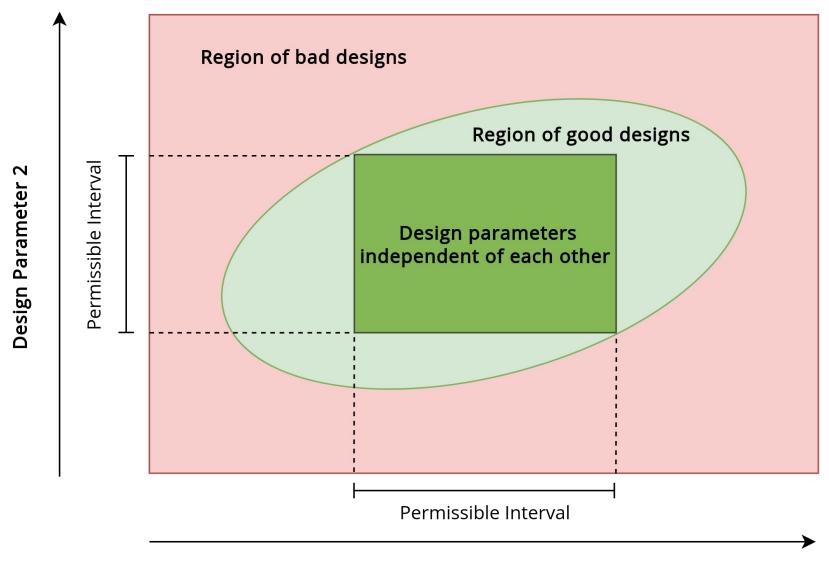
Dummy verletzt

Dummy sicher





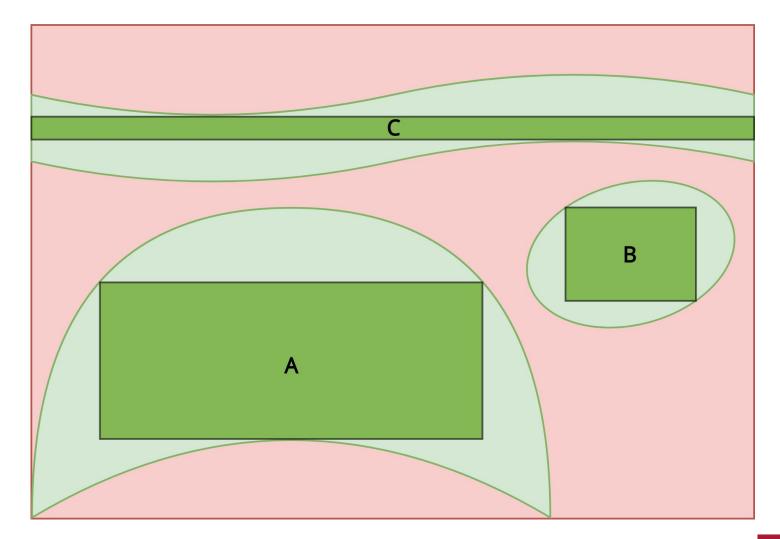
# Unabhängige, erlaubte Intervalle = Lösungsbox





### Mehrere Lösungsboxen

Clustering des Designraums, um mehr als eine Lösung anzubieten





#### **Box: Definition, Fitness & Maximierung**

#### Boxdefinition

$$\Omega = I_1 \times I_2 \times ... \times I_p \qquad I_i = [x_i^l, x_i^u], i \in [1, p]$$

#### Boxfitness (Volumen)

$$\mu(\Omega) = \begin{cases} -\infty & \text{, if } b = 0, g = 0 \\ -b & \text{, if } b > 0 \end{cases}$$
 
$$100(x_1^u - x_1^l)(x_2^u - x_2^l)...(x_p^u - x_p^l) & \text{else}$$

b = # schlechte Designs

g = # gute Designs

#### Box-Maximierungsproblem

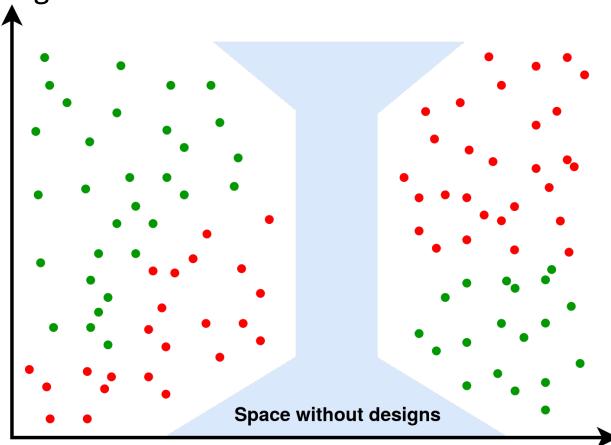
$$\mu(\Omega) \rightarrow max$$



#### Praktisches Problem: Schwach besetzter Designraum

Echte Crashtests sind teuer und Crashsimulation zeitaufwendig zu berechen

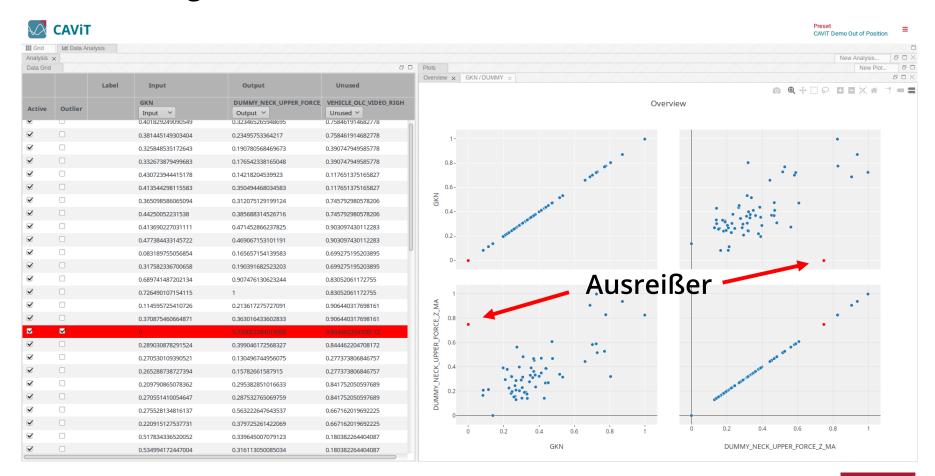
→ der Designraum weist sichtbare Lücken auf





#### Praktisches Problem 2: Ausreißer

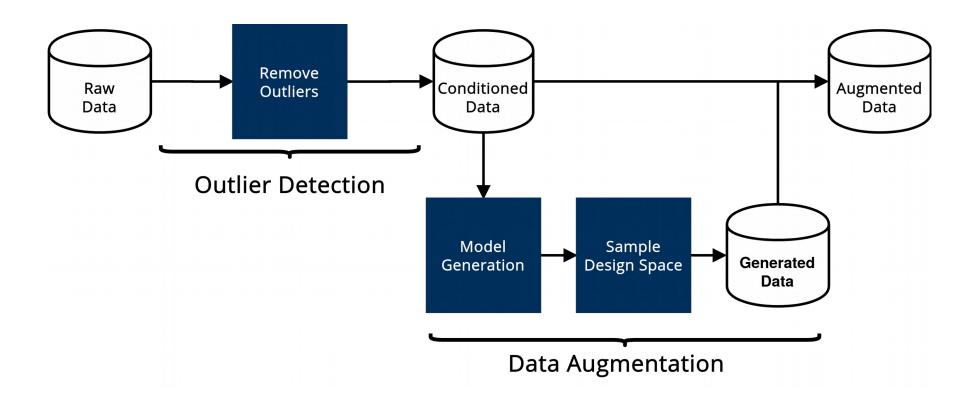
- Daten können Ausreißer enthalten
  - → müssen gefunden und entfernt werden





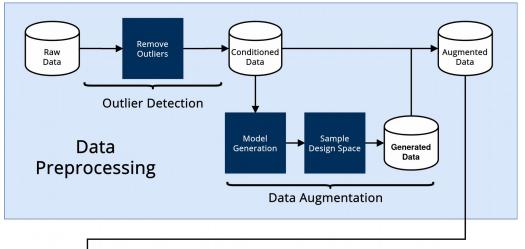
#### Datenvorverarbeitung (Data Preprocessing)

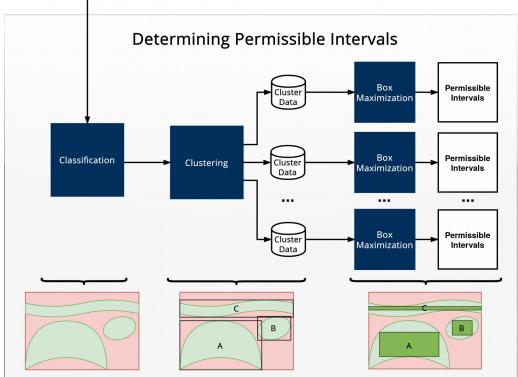
Praktische Probleme werden durch einen Datenvorverarbeitungsschritt gelöst





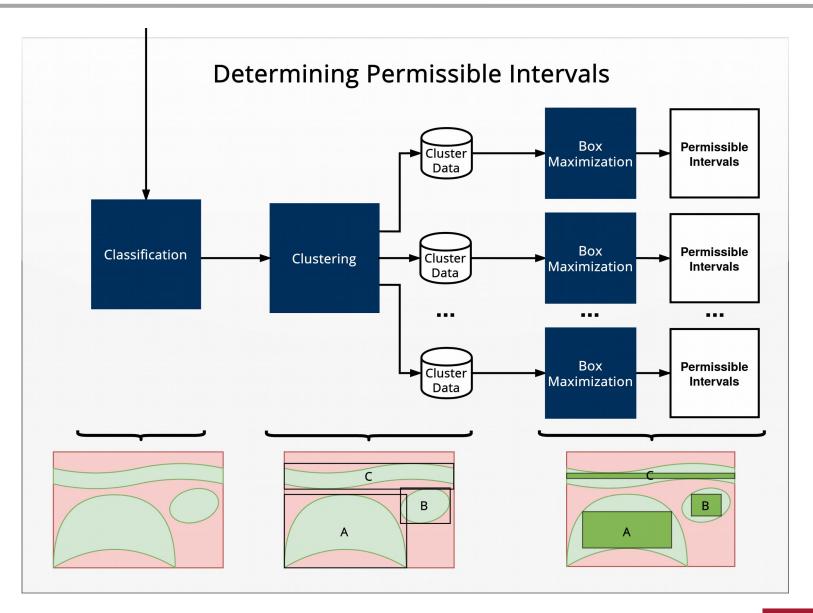
# Überblick Konzept







# Überblick Konzept





#### **Implementierung**

- Ausreißerentdeckung
  - Isolation Forest
  - k Nächste Nachbarn
  - DBSCAN
- Clustering
  - DBSCAN
- Boxmaximierung
  - Exact Max Box
  - Anneal Max Box
  - Random Max Box

Software-Prototyp



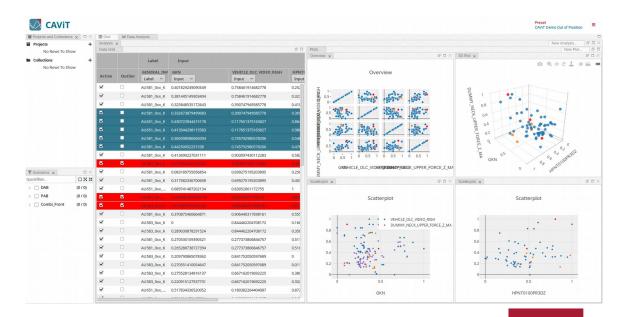






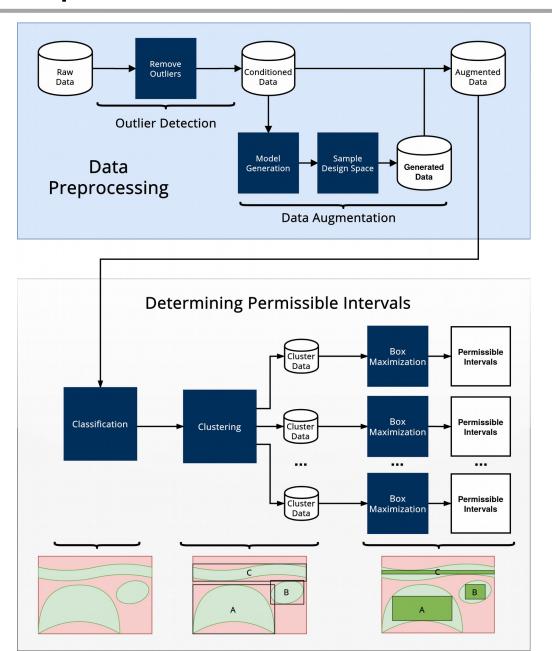
- Modelgenerierung
  - Polynomielle Regression
  - Neuronale Netze





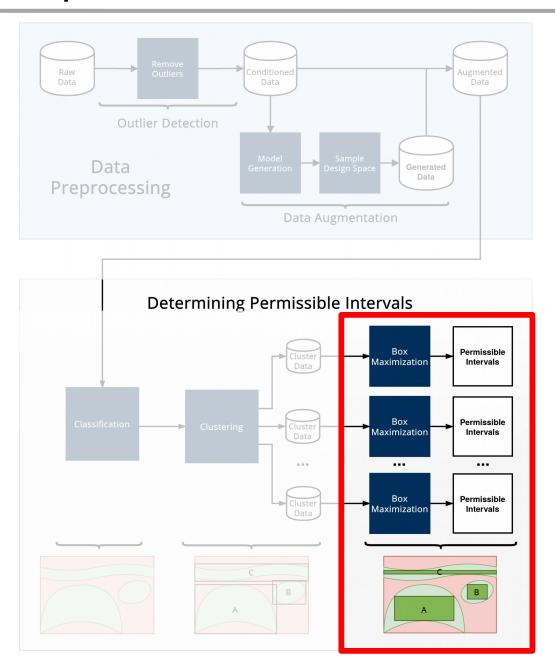


# Überblick - Komponentenweise Evaluation





# Überblick - Komponentenweise Evaluation

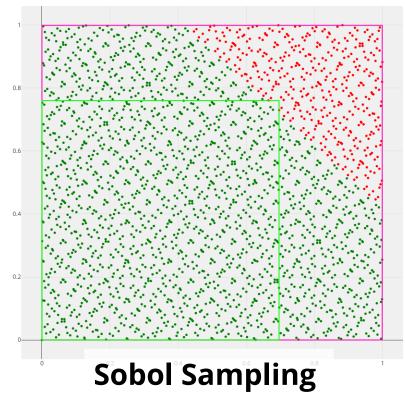


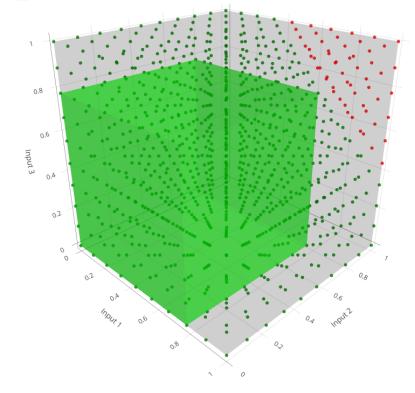


#### **Boxmaximierung - Komponentenweise Evaluation**

Synthetische Funktion für die Evaluation

Summenfunktion 
$$f(\mathbf{x}) = \sum_{i=1}^{i \leq n} x_i \;\; x_i \in [0,1], n \in \mathbb{N}_{\neq 0}$$

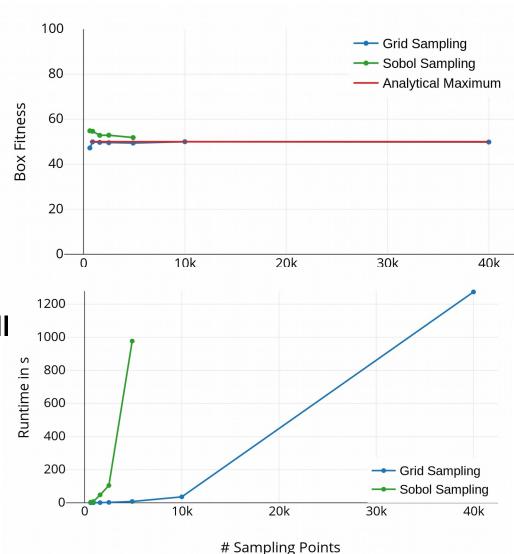




**Grid Sampling** 

#### 2D Summenfunktion – Exact Max Box

- Boxfitness:
  - Sobol Sampling:
     Über dem Maximum
  - Grid Sampling:
     Unter dem Maximum
- Laufzeit wächst exponentiell mit der Anzahl der Samplingpunkte
- Unterschied zwischen Sobol und Grid Sampling



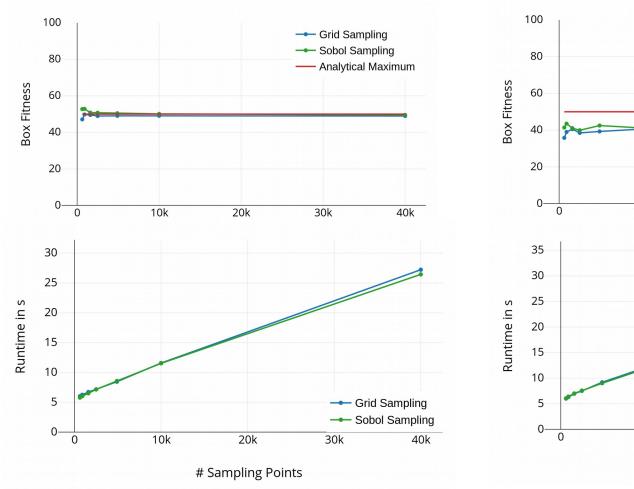


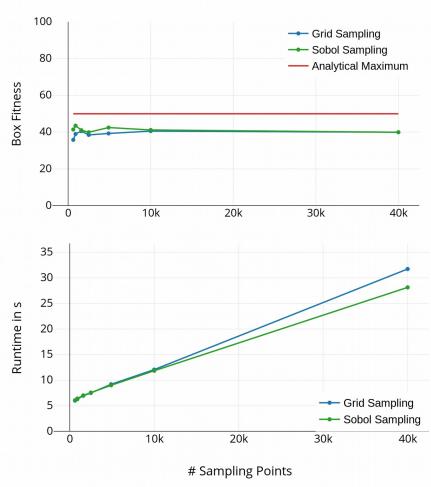
#### 2D Summenfunktion – Heuristische Algorithmen

#### Lineares Wachstsum mit der Anzahl der Samplingpunkte



#### **Random Max Box**

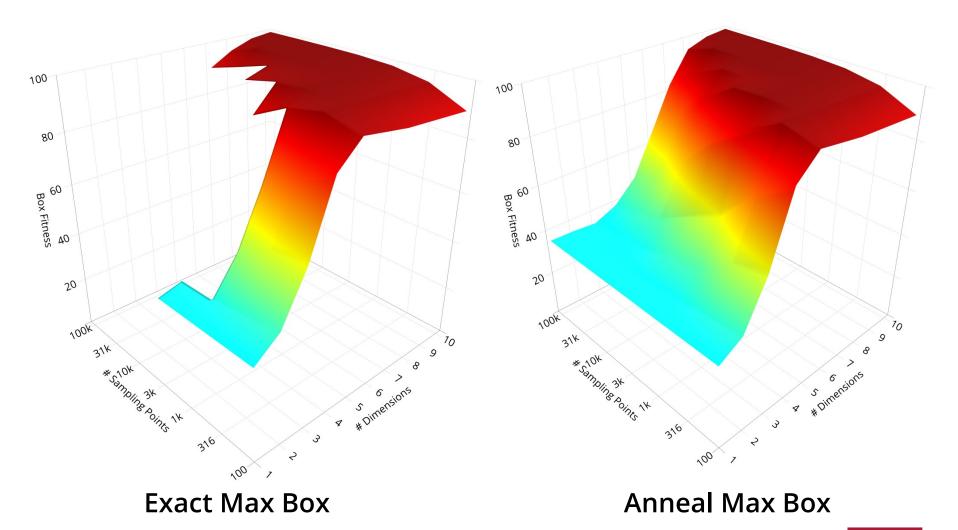






#### 1D - 10D Summenfunktion - Boxmaximierung

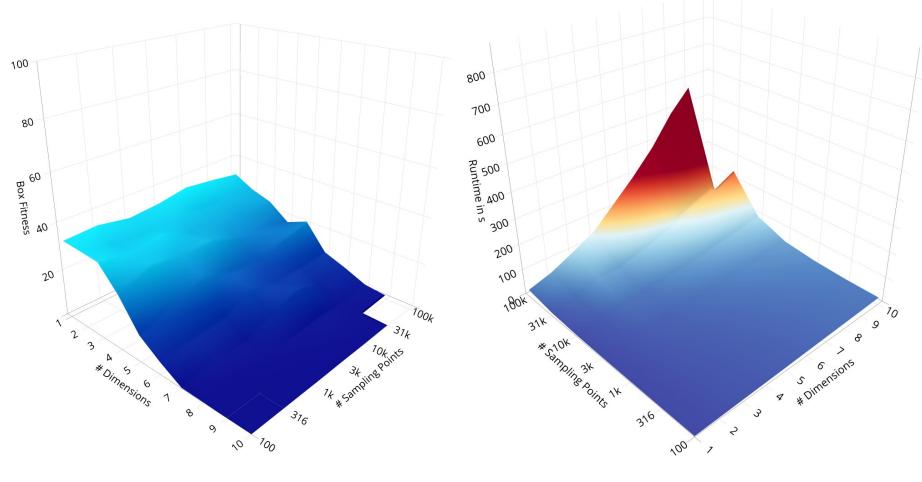
Boxfitness hängt von der #Designs & #Dimensions ab





### 1 - 10D Summenfunktion - Boxmaximierung

#### Random Max Box



**Boxfitness** 

Laufzeit



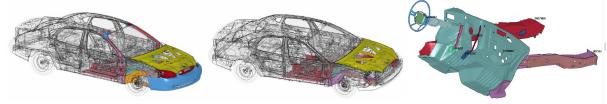
#### **Ende-Zu-Ende Evaluation**

- Drei Teildatensätze vom US-NCAP-Datensatz
- 3D US-NCAP3 Designparameter

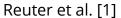
10D US-NCAP10 Designparameter



Reuter et al. [1]

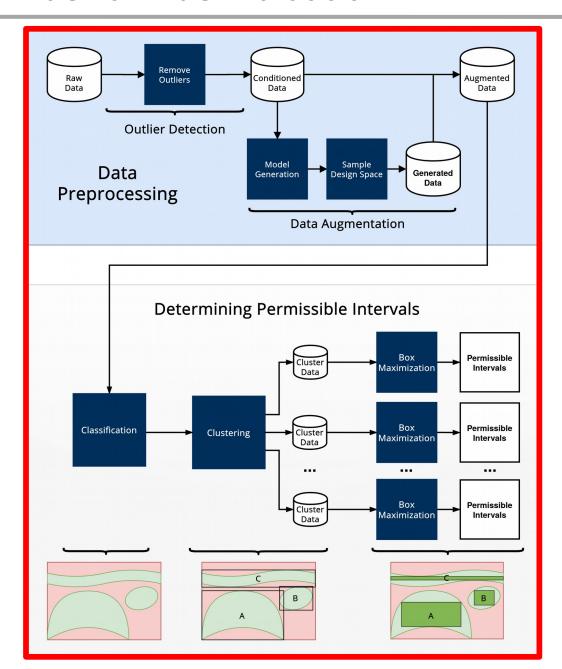


27D US-NCAP27 Designparameter





#### Überblick - Ende-zu-Ende Evaluation

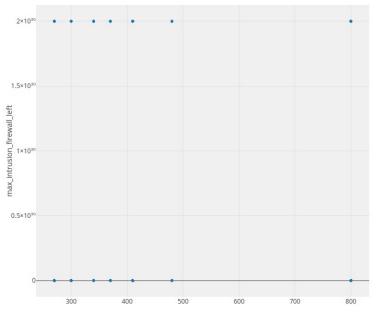




# Ausreißer & Modellgenerierung

- 118 Ausreißer im US-NCAP Datensatz
- Ausreißer sind fehlgeschlagene Berechnungen
- Modellerstellung mit Polynomieller Regression
- Ausreißerentfernung erhöht die Modellqualität signifikant

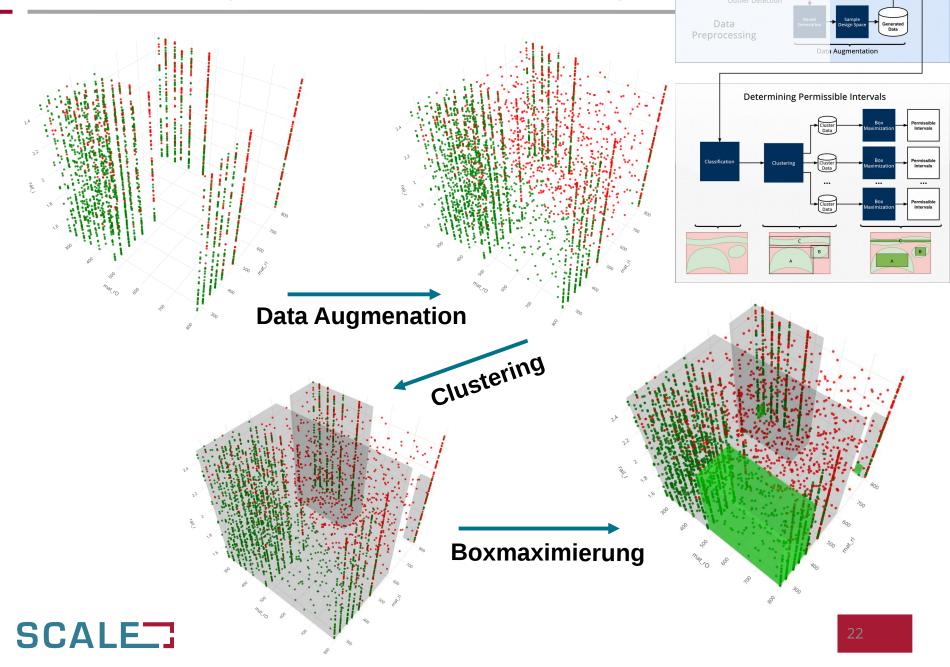
Raw Data		Conditioned Data
Pre	Data eprocessing	Model Generation Sample Generated Data  Data Augmentation



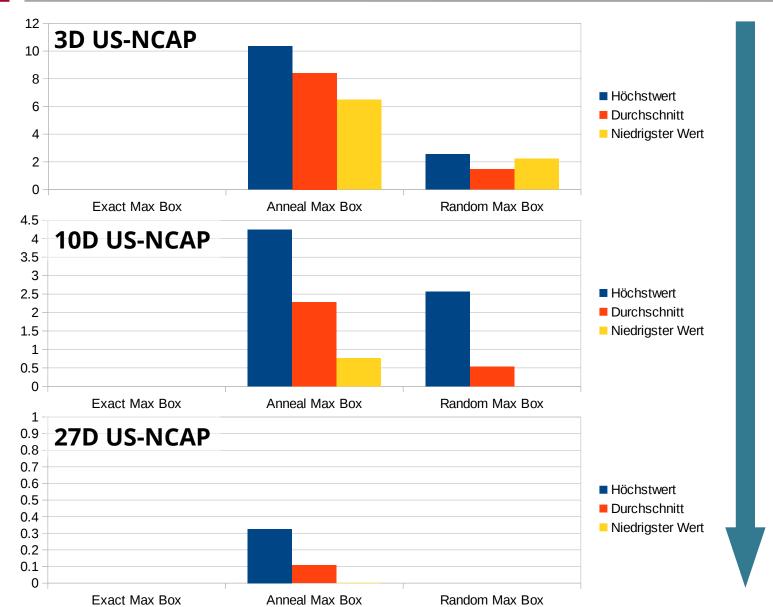
	Outlier		No Outlier	
# Design Parameters	R <sup>2</sup> Test	R <sup>2</sup> Training	$R^2$ Test	R <sup>2</sup> Training
3D US-NCAP	0.00	0.00	0.41	0.44
10D US-NCAP	-0.01	0.00	0.53	0.56
27D US-NCAP	0.02	0.05	0.72	0.80



#### 3D US-NCAP (Ende-Zu-Ende Evaluation)



#### **US-NCAP Tests - Box Fitness (End-To-End Evaluation)**



Steigende Varianz & Geringere Boxfitness

#### Diskussion / Zukünftige Arbeiten

- Synthetische Daten: Anneal Max Box funktioniert gut mit den Testproblemen
- Echte Daten: Über die Ergebnisse kann keine quantitative Aussage gemacht werden, da Benchmarkwerte nicht bekannt sind
- Ausreißerentfernung notwendig vor Modelgenerierung
- Exact & Random Max Box sind mit mehr als 2 Dimensionen praktisch nicht anwendbar
- Auflösungsproblem: Anzahl der Designs approximiert Funktion
   → in hohen Dimensionen immens viele Punkte notwendig
   Daher: Dimensionsreduktion mit Sensitivitätsanalyse denkbar



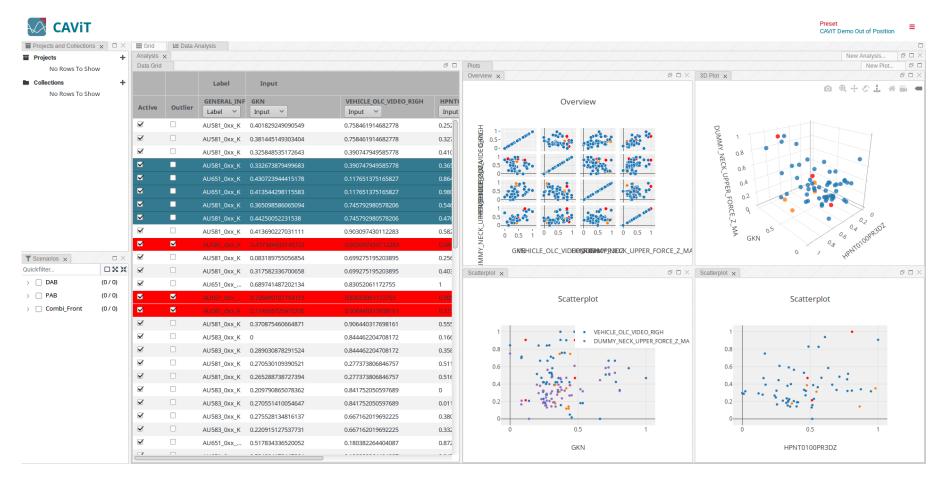
#### Zusammenfassung der Arbeit

- Software Prototyp wurde entwickelt, der unabhängige, erlaubte Eingabeparameter für Designs findet
- Prototype kann grundsätzlich beliebig viele Dimensionen und Datenpunkten verarbeiten
- Drei Algorithmen zur Boxmaximierung: Exakter Algorithmus und zwei Heuristiken
- Metaheuristik Simulated Annealing zur Erstellung eines heuristischen Algorithmus
- Prototyp wurde in einer komponentenweisen Evaluation und in einer Ende-zu-Ende Evaluation getestet



#### Prototyp-Präsentation

#### Jetzt: Live-Präsenation





# Vielen Dank!

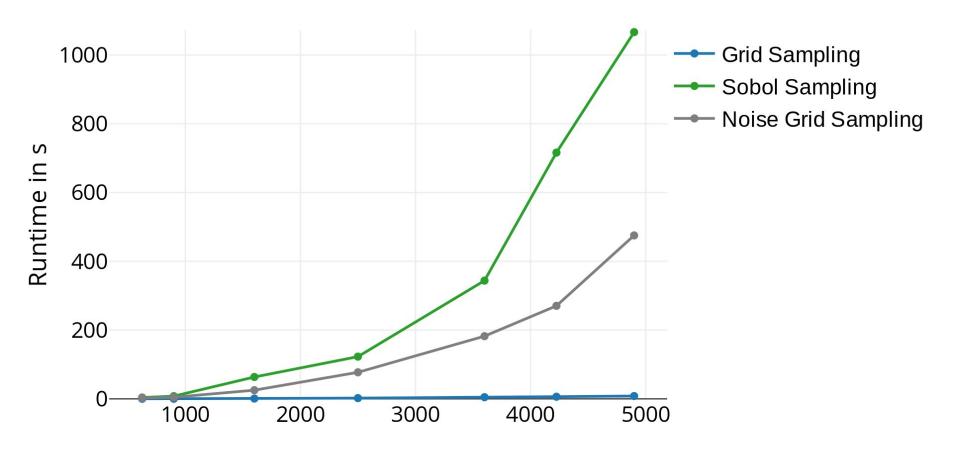


#### Bildquellen

1. Uwe Reuter, Zeeshan Mehmood, Clemens Gebhardt, Martin Liebscher, Heiner Müllerschön and Ingolf G. Lepenies. Using Is-opt for meta-model based global sensitivity analysis. 01/2011.



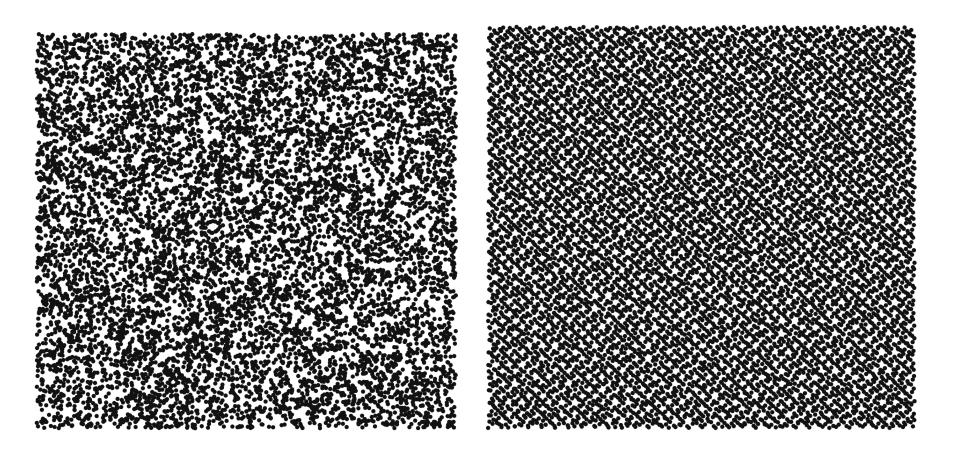
# Laufzeitunterschied Grid vs. Sobol Sampling



# Sampling Points

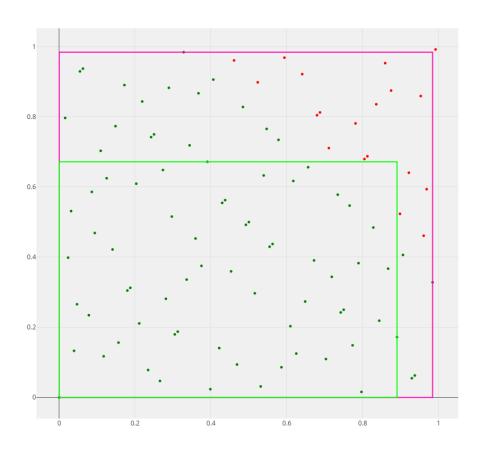


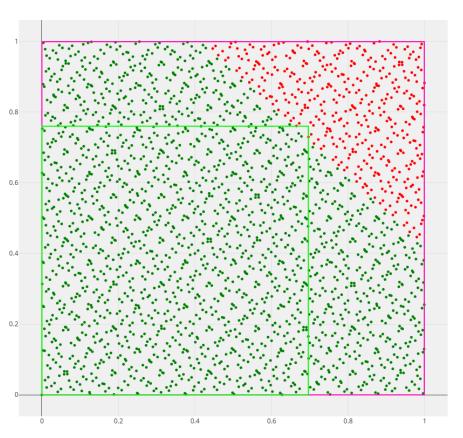
# Random Sampling vs. Sobol Sampling





# Sobol Überschätzung / Grid Unterschätzung







#### Verwandte Arbeit von Zimmermann et al.

