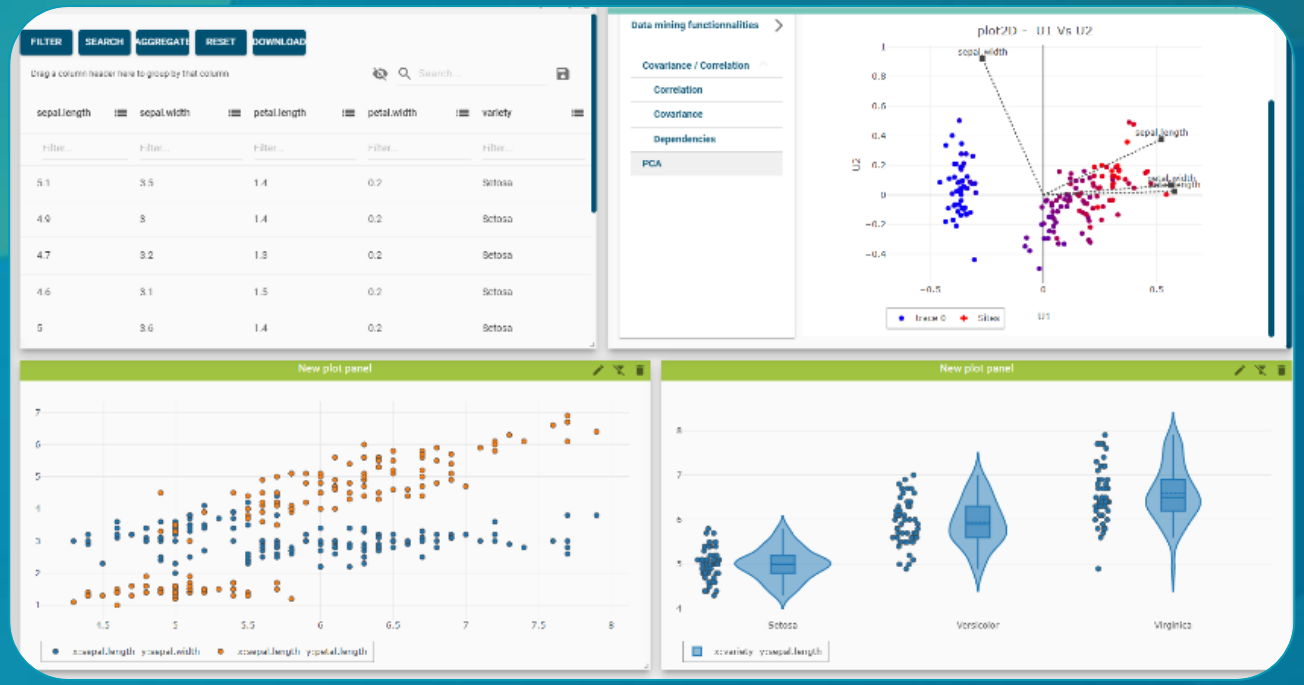


Datenwelten entdecken

Bayreuther Konstrukteurstag 13.9.2023

Cornelia Thieme

Hexagon

HEXAGON

Angenommen, Sie erhalten einen Datensatz zum Auswerten…

- Ergebnisse verschiedener Varianten einer FEM-Berechnung
- Testergebnisse
- Fertigungsdaten und -ergebnisse
- Daten aus dem Betrieb einer Maschine
- Auswertung einer Kundenumfrage
- Daten aus dem Internet zu einem relevanten Thema

Wie können Sie effizient Informationen aus diesem Datensatz ziehen?

	A	B	C	D	E	F	G
1	Drehzahl	Drehmoment	Vorschub	Materialhärte	Bohrer-Qualität	Motortemperatur	Geräusch
2	2104	209.7642	4.5228	12.2109	2.435	72.555	52.1122
3	2070	215.0593	4.5338	11.8989	2.582	73.4269	51.5143
4	2055.25	208.419	4.5739	11.8304	2.484	72.5837	51.1017
5	2025.75	215.5681	4.4701	11.9682	2.447	73.1607	50.3682
6	2064.75	211.8095	4.5355	11.9551	2.518	73.0113	51.398
7	2067.75	212.3537	4.4689	12.0357	2.532	72.7587	51.4321
8	2049	212.3897	4.4858	11.961	2.519	72.9657	51.0224
9	2064.75	209.12	4.4543	12.1252	2.482	72.4065	51.2968
10	2023.75	208.7497	4.4401	11.9	2.525	72.5756	50.3307
11	2032.25	213.1715	4.5816	11.6498	2.519	73.2682	50.6136
12	2032.5	213.2748	4.5188	11.9178	2.428	72.9208	50.5052
13	2038.25	210.3289	4.5096	11.9035	2.456	72.5823	50.6695
14	2058.5	212.0269	4.5629	11.8859	2.46	72.943	51.1304
15	2042.75	207.7407	4.4815	11.8366	2.513	72.1826	50.833
16	2054.5	209.311	4.5451	11.6545	2.571	72.3887	51.2009
17	2047	205.1366	4.4536	12.0942	2.454	72.0336	50.8443
18	2015.5	209.9634	4.4311	11.9248	2.546	72.4149	50.2328
19	2034.75	210.0638	4.5315	11.9229	2.455	72.5921	50.6218
20	2040.75	205.8472	4.4954	11.9191	2.459	72.1903	50.6404
21	2109.5	208.3838	4.555	12.0221	2.537	72.3974	52.3793
22	2032.25	213.6052	4.4875	12.1622	2.434	72.9187	50.535
23	2023.5	207.6246	4.5127	11.7377	2.475	72.2092	50.3246
24	2043.25	211.					50.8219
25	2046.75	204.					50.8411
26	2040	212.					50.8239
27	2053.75	209.					51.0307
28	2081.75	207.					51.6782
29	2053.75	211.					51.0979
30	2071.25	206.					51.3693
31	2047	204.					50.9115
32	2034.25	208.					50.6923
33	2041	212.					50.8263
34	2030	212.					50.505
35	2047.5	207.					50.8949
36	2020.75	215.					50.3154
37	2041.5	209.					50.7223
38	2063.5	208.9976	4.6138	11.7532	2.508	72.5207	51.3172
39	2017	211.7559	4.4817	11.7646	2.547	72.7074	50.3404

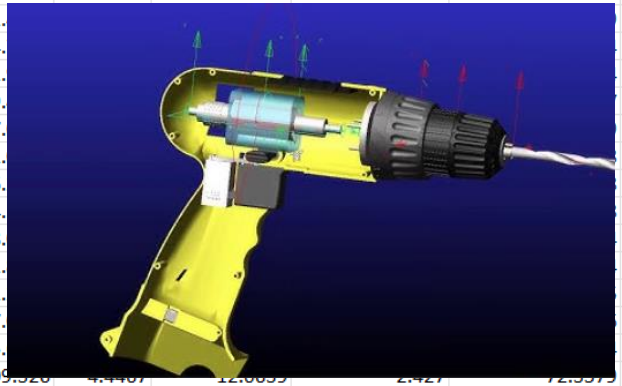


Bild: https://www.youtube.com/watch?app=desktop&v=fRDM4zsdgeA
Daten: erfunden

HEXAGON

„90% der Zeit beim Machine Learning werden für die Datenaufbereitung aufgewendet.”

- **Datenbereinigung:** Entfernen von fehlenden Werten, Ausreißern und inkonsistenten Datenpunkten.
- **Feature Engineering:** Erstellen neuer, aussagekräftiger Merkmale aus den vorhandenen Daten.
- **Datenintegration:** Zusammenführen von Daten aus verschiedenen Quellen oder Dateien.
- **Skalierung und Normalisierung:** Anpassen der Merkmale, um sicherzustellen, dass sie in einem vergleichbaren Bereich liegen.
- **Kodierung von Kategorien:** Umwandeln von kategorialen Daten oder Bildern in Zahlenwerte.
- **Aufteilung der Daten:** Trennen der Daten in Trainings- und Testsets für Modelltraining und Validierung.
- **Einführung von Datenqualitätsprüfungen:** Implementieren von Validierungen und Überprüfungen, um sicherzustellen, dass die Daten den Erwartungen entsprechen.

Datenbereinigung / Datenprüfung

Drehzahl	Drehmoment	Vorschub	Materialhärte	Bohrer-Qualität	Motortemperatur
3000	172.8	3.2	17	3	65
3000	281.7		16,5	3	76.3
3000	274.8	4.6	17.1	2	?
3000	81.6	7	6.3	2	56.2
3000	122.3	4.3	5.7	5	Unbekannt
3000	57.2	7	3.2	1	52.8
3000	361	2	18.2	1	78.3
3000	453	7	13.2	3	106.3
3000	95.3	2	9.8	2	58.2
3000	125.3	5.1	13.2	2	69.5
3000	343	2	5.8	5	89.3
3000	151	5	17	1	72

Formatprobleme finden:

- leere Felder
- Sonderzeichen
- Buchstaben, wo Zahlen erwartet werden
- nicht akzeptierte Dezimaltrennzeichen
- Parameter, die immer den gleichen Wert haben und daher keinen Einfluss haben

Betreffende Zeilen / Spalten löschen?

HEXAGON

Datenbereinigung / Datenprüfung

Drehzahl	Drehmoment	Vorschub	Materialhärte	Bohrer-Qualität	Motortemperatur	Geräusch
3000	172.8	3.2	17	3	65	73.4
3000	281.7	4.7	16.5	3	76.3	74.6
2250	274.8	4.6	17.1	2	81.2	56.4
1750	81.6	7	6.3	2	56.2	44.5
1750	122.3	4.3	5.7	5	65.3	45
1750	57.2	7	3.2	1	52.8	42.6
500	361	2	18.2	1	78.3	15.6
3000	453	7	13.2	3	106.3	74.9
1750	95.3	2	9.8	2	58.2	40.9
2250	125.3	5.1	13.2	2	69.5	55.4
500	343	2	5.8	5	89.3	17.8
3000	151	5	17	1	72	68.4
3000	151	5	17	1	72	68.4

- Datensätze, die doppelt vorkommen

3000	151	5	17	1	72	68.4
3000	151	5	17	1	72	80

- Datensätze mit gleichen Parametern, aber unterschiedlichem Ergebnis

HEXAGON

Datentyp

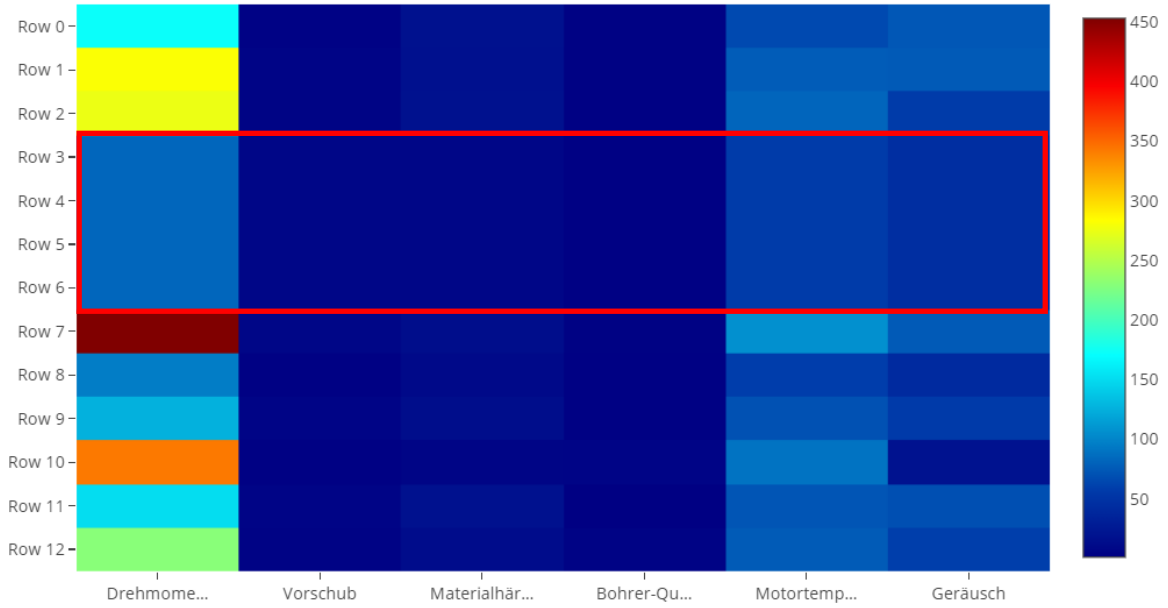
- Integer z.B. 2, 4, 15
- Real z.B. 2.3, 3.44, 20.
- Binär z.B. 0,1
- Kategorial / Name Tags z.B. blau, grün, rot, gelb oder Alu, Stahl, Kupfer
- Bilder z.B. jpg, png
- CAD-Dateien z.B. step

Alles als Zahlenwerte darstellen?

HEXAGON

Ausreißer finden per Color Map

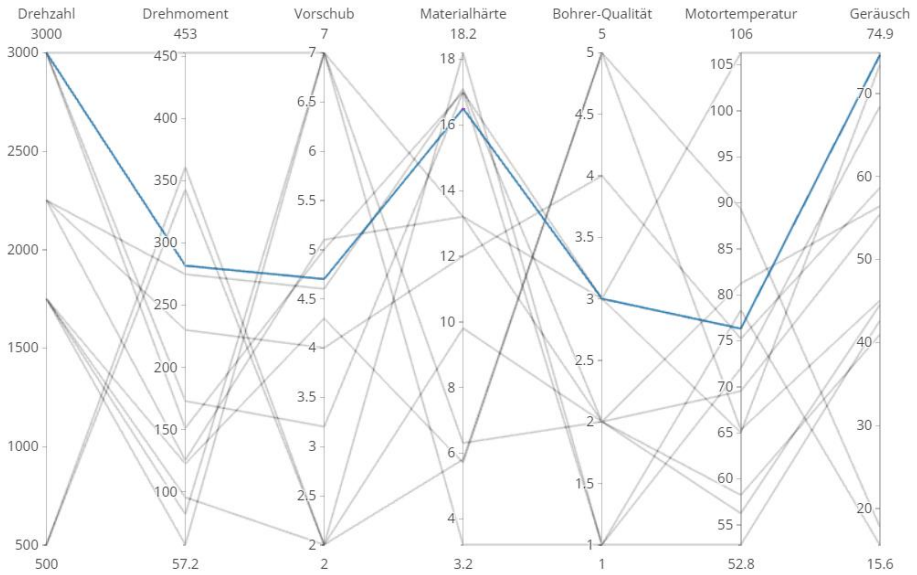
Auffällig hohe Werte



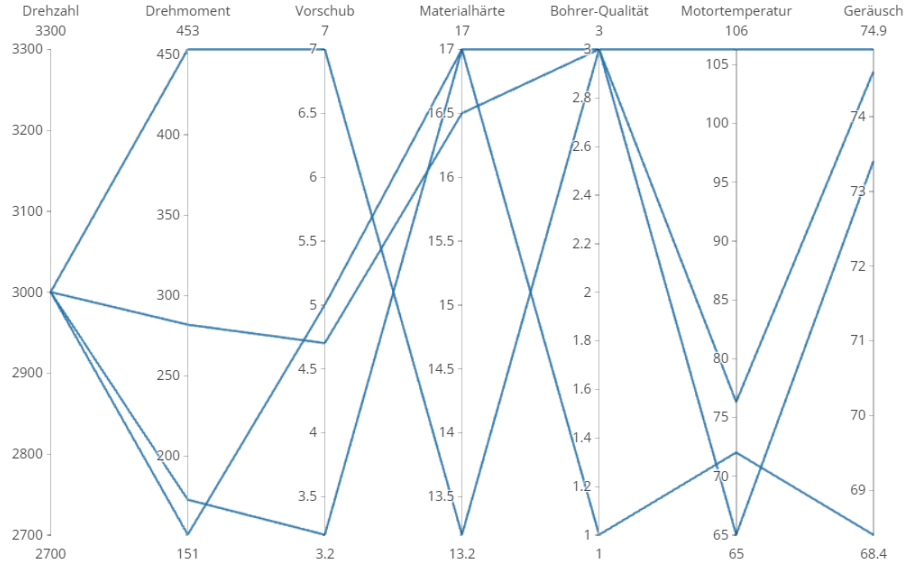
Mehrere identische Zeilen können auf Copy-Paste-Fehler hinweisen

HEXAGON

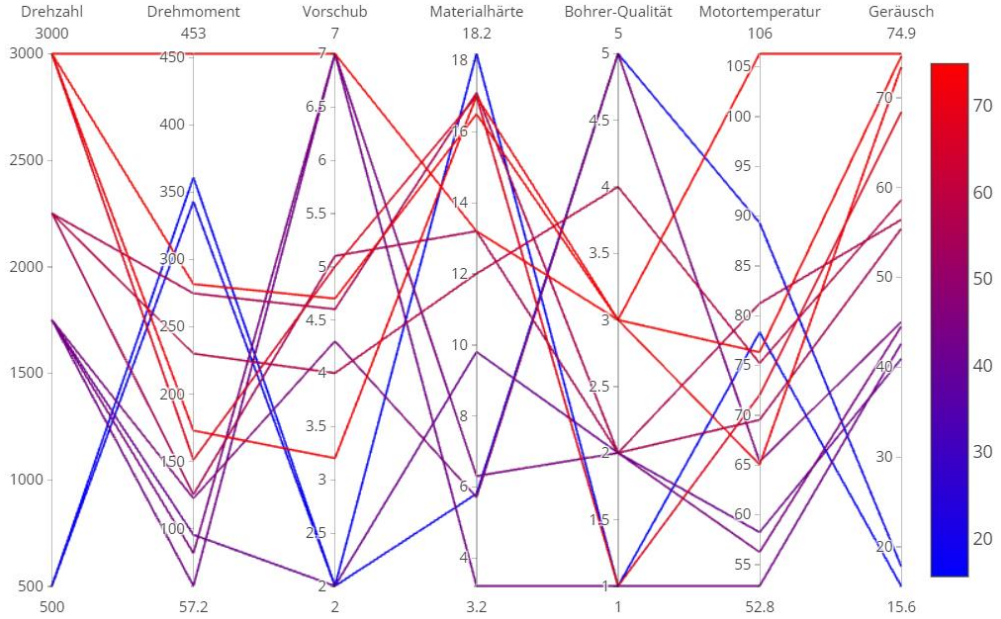
Ausreißer, Abhängigkeiten, Datenverteilung: Parallelkoordinatenplot



Jeder Datensatz ist eine Kurve



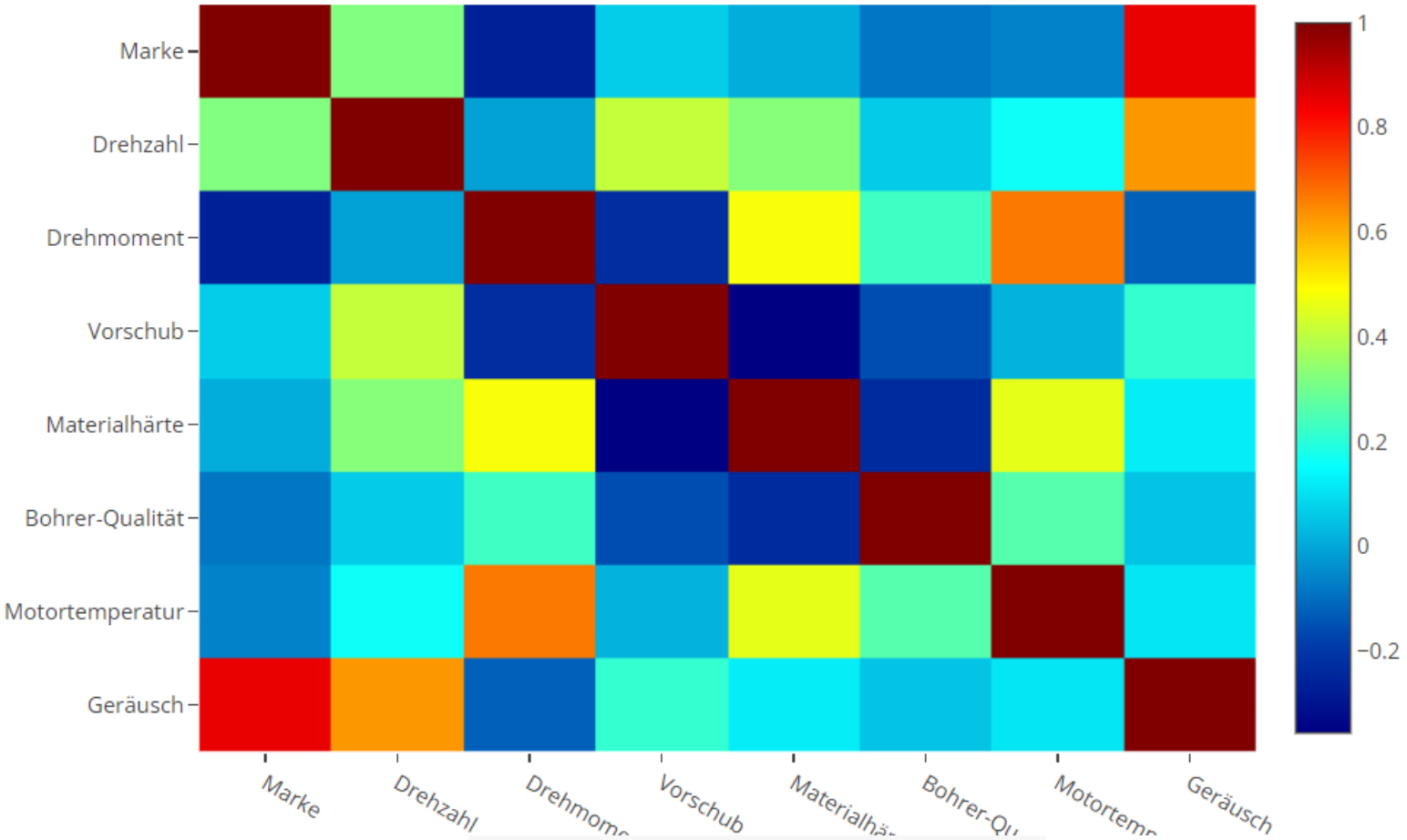
Skalierung, z.B. einen bestimmten Drehzahlbereich einblenden



Farbskala für einen Datenbereich: hohe Geräuschwerte in Rot

HEXAGON

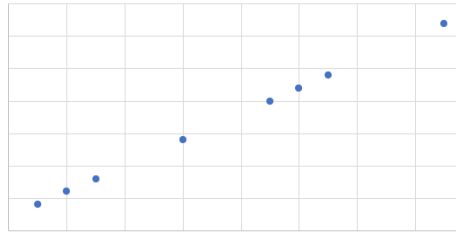
Korrelationsplot – Abhängigkeiten zwischen Daten erkennen



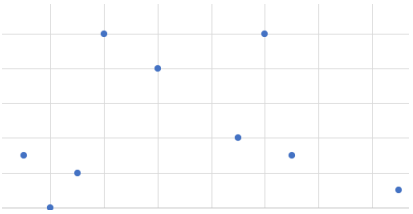
$$\rho(X,Y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

- x_i sind die Werte der Variable X.
- y_i sind die Werte der Variable Y.
- $\bar{x}$ ist der Durchschnitt der Werte von X.
- $\bar{y}$ ist der Durchschnitt der Werte von Y.
- n ist die Anzahl der Datenpunkte in der Stichprobe.

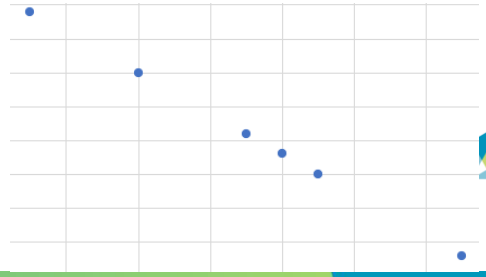
Korrelationskoeffizient = 1
Vollständige Abhängigkeit, z.B.



Korrelationskoeffizient = 0
Kein Zusammenhang, z.B.



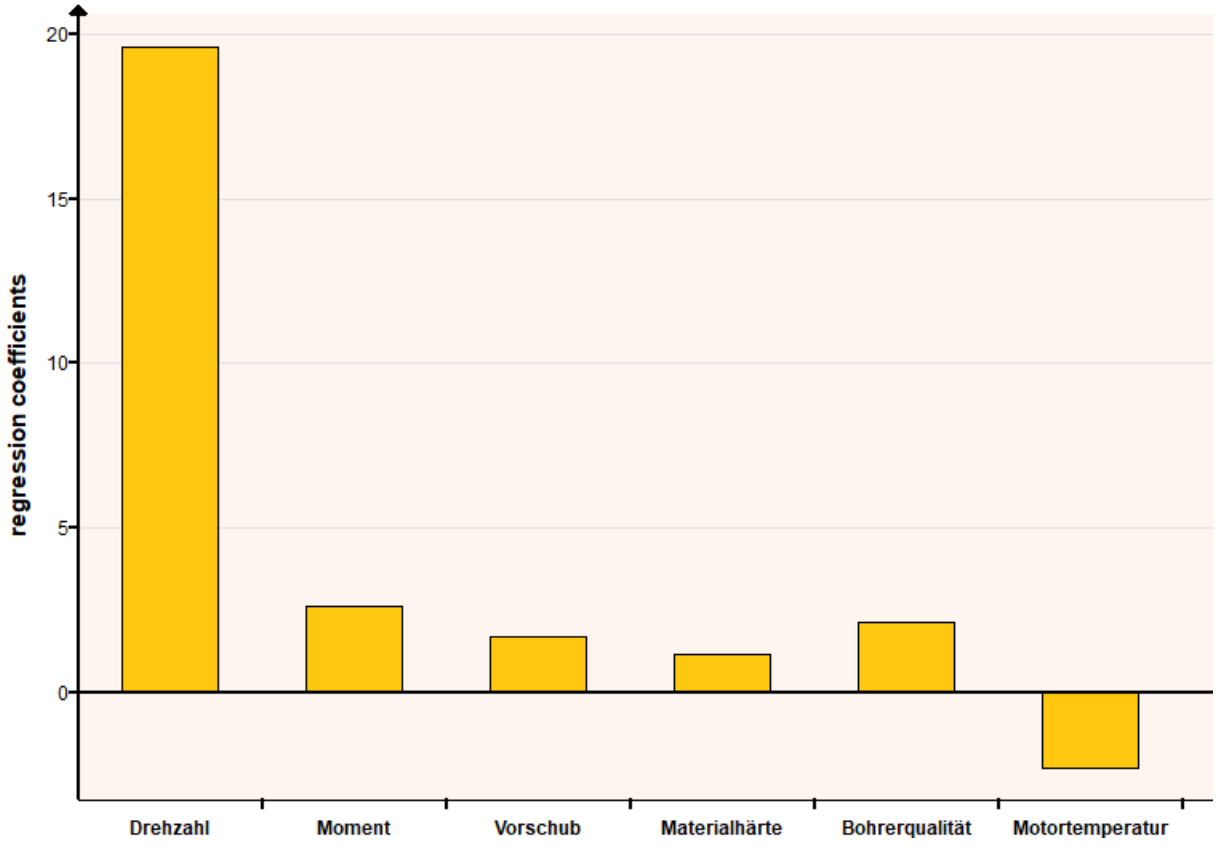
Korrelationskoeffizient = -1
Vollständige gegenläufige Abhängigkeit, z.B.



HEXAGON

Korrelation – Abhängigkeiten zwischen Daten erkennen

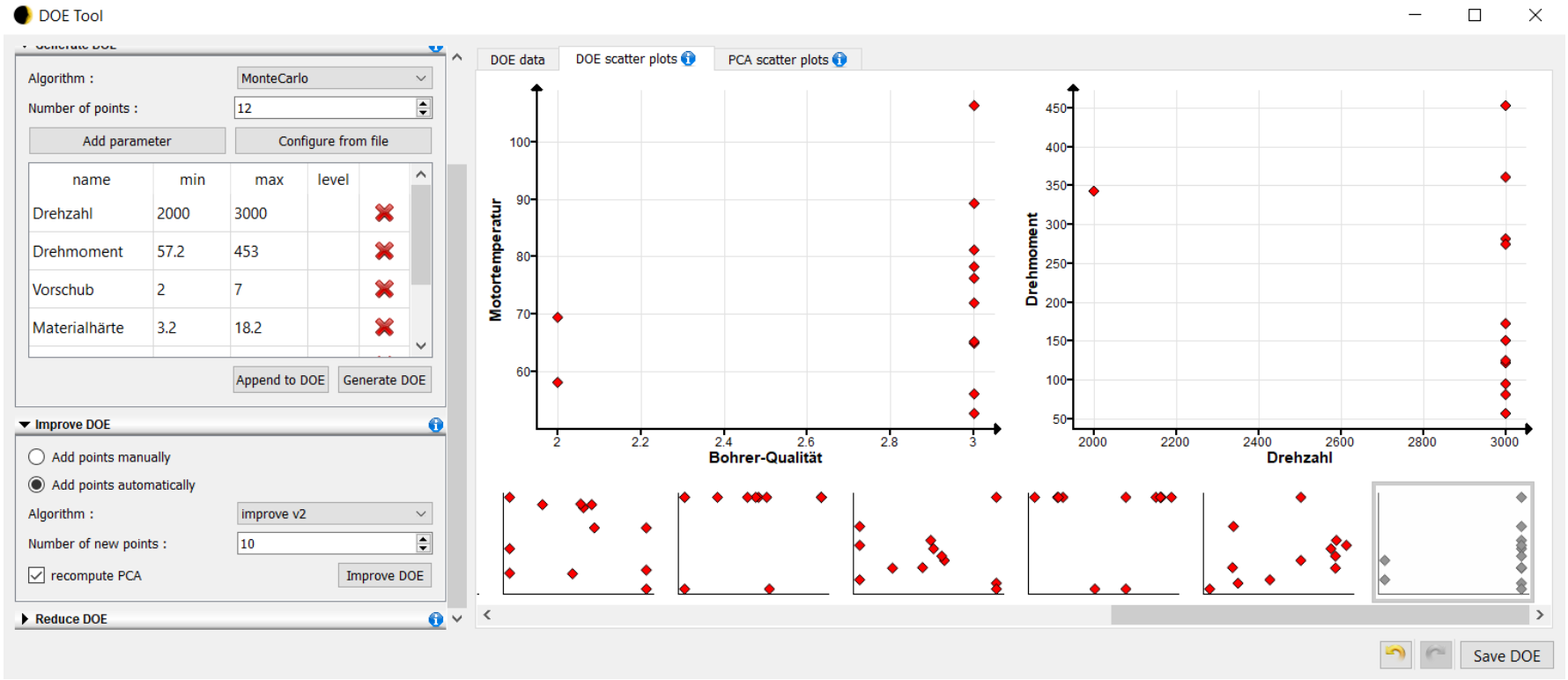
Vorhersage des Geräusches aus den Parametern



Einfluss der Parameter

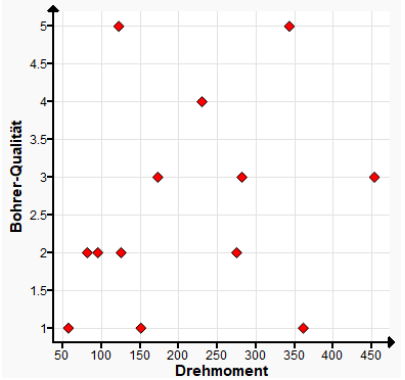
HEXAGON

DOE Plot: Jedes Parameterpaar darstellen

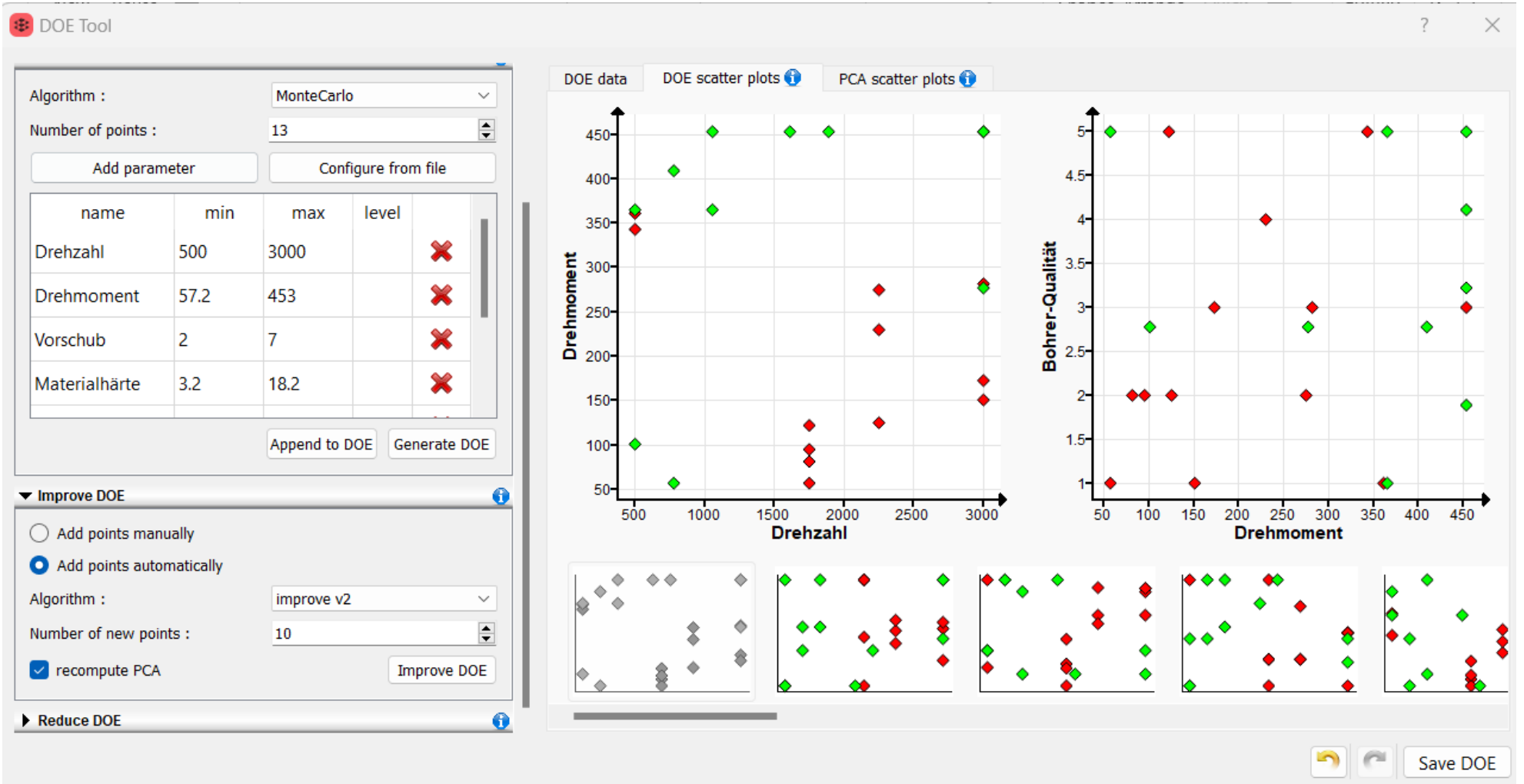


- Für jedes Parameterpaar sollte es eine Verteilung geben, die den ganzen Parameterraum abdeckt, Korrelation nahe 0

- Dies ist physikalisch nicht immer möglich. Nicht immer sind auch „schlechte“ Daten zum Lernen vorhanden, z.B. bei Fertigungsprozessen



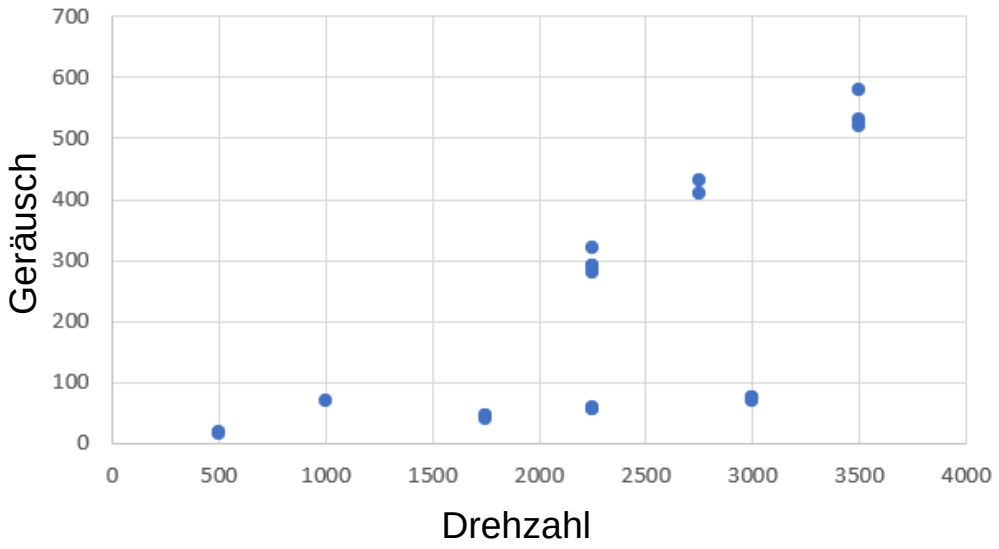
DOE ergänzen

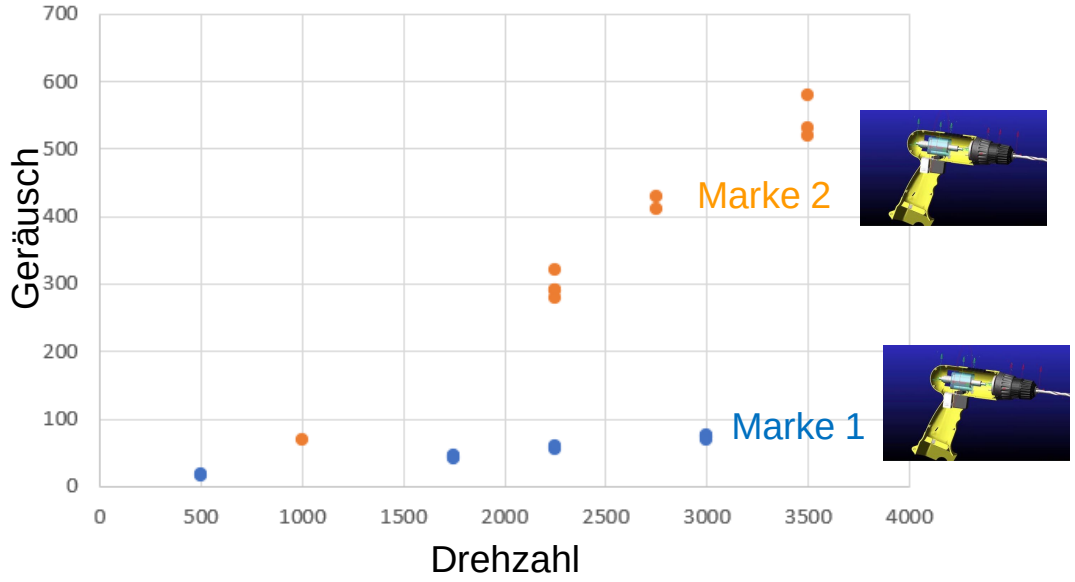


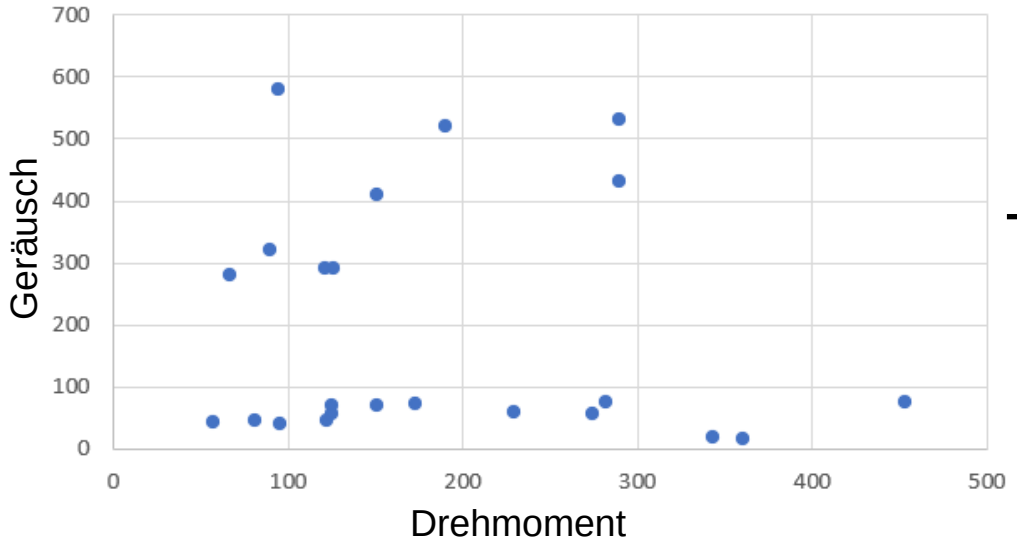
Es können automatisch und manuell Punkte (Parameterkombinationen) hinzugefügt werden, um den Parameterraum besser abzudecken. Für diese Parameterkombinationen sollten dann zusätzliche Versuche / Berechnungen durchgeführt werden.

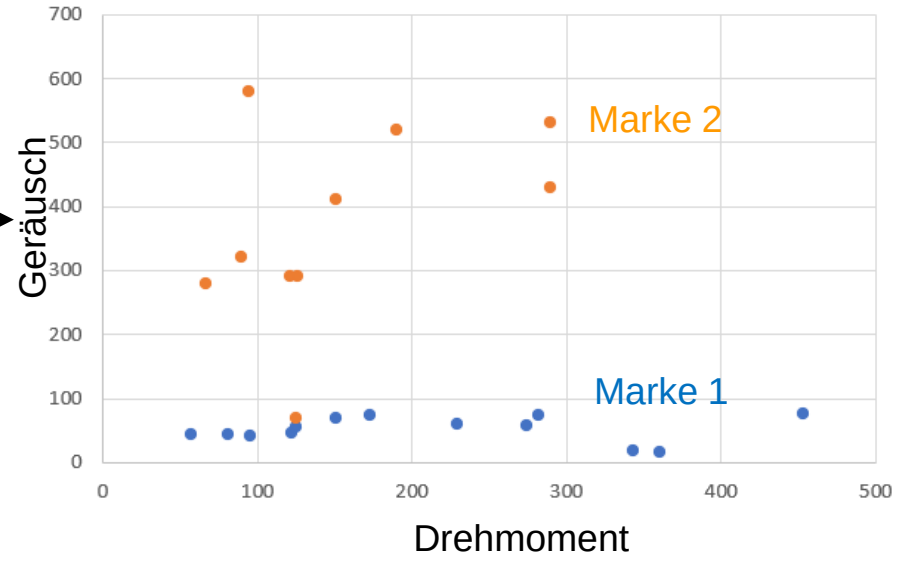
HEXAGON

Clustering





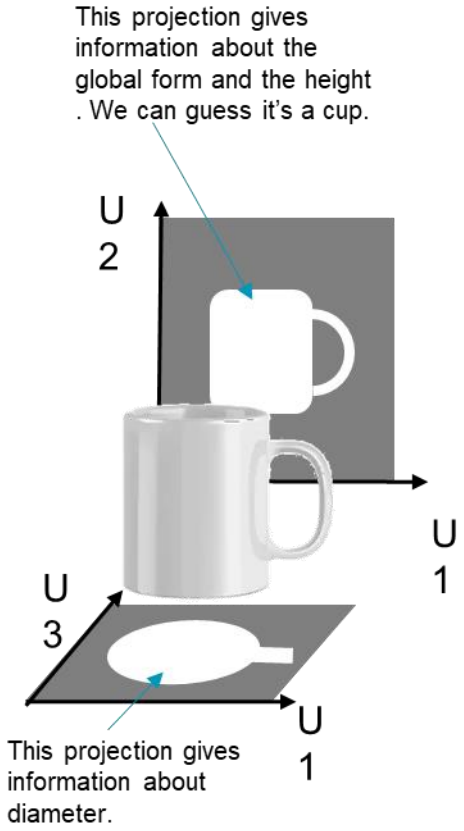


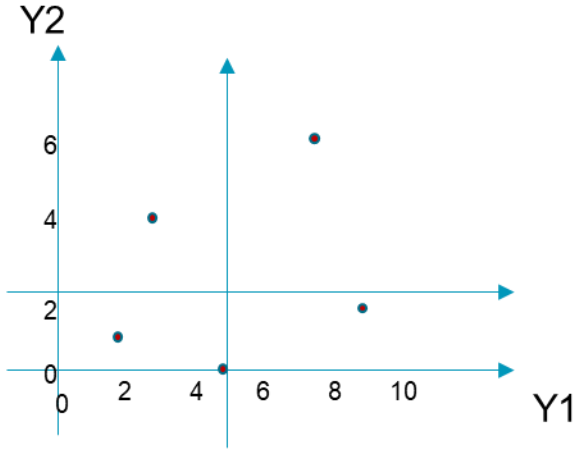


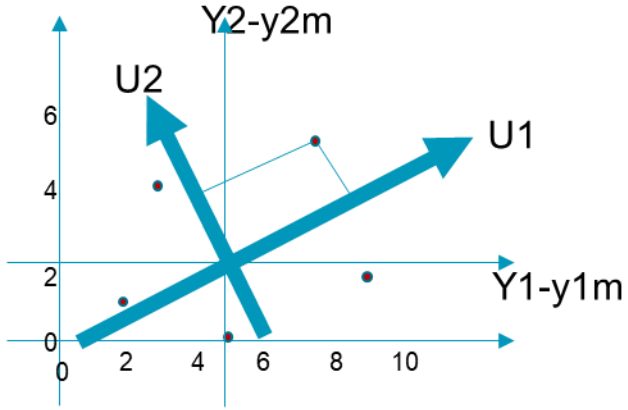
Manchmal ist es sinnvoll, die Daten in verschiedene Cluster zu teilen und getrennt zu betrachten.

HEXAGON

Principal Component Analysis (PCA)





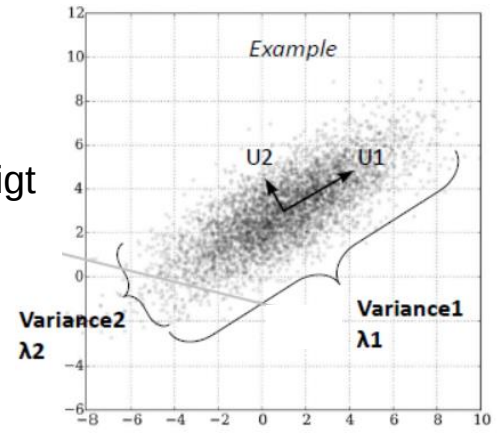


Y=
2 ; 1
3; 4
5 ; 0
7 ; 6
9 ; 2

Die Daten werden in ein Hauptachsensystem projiziert, in dem sie leichter dargestellt werden können

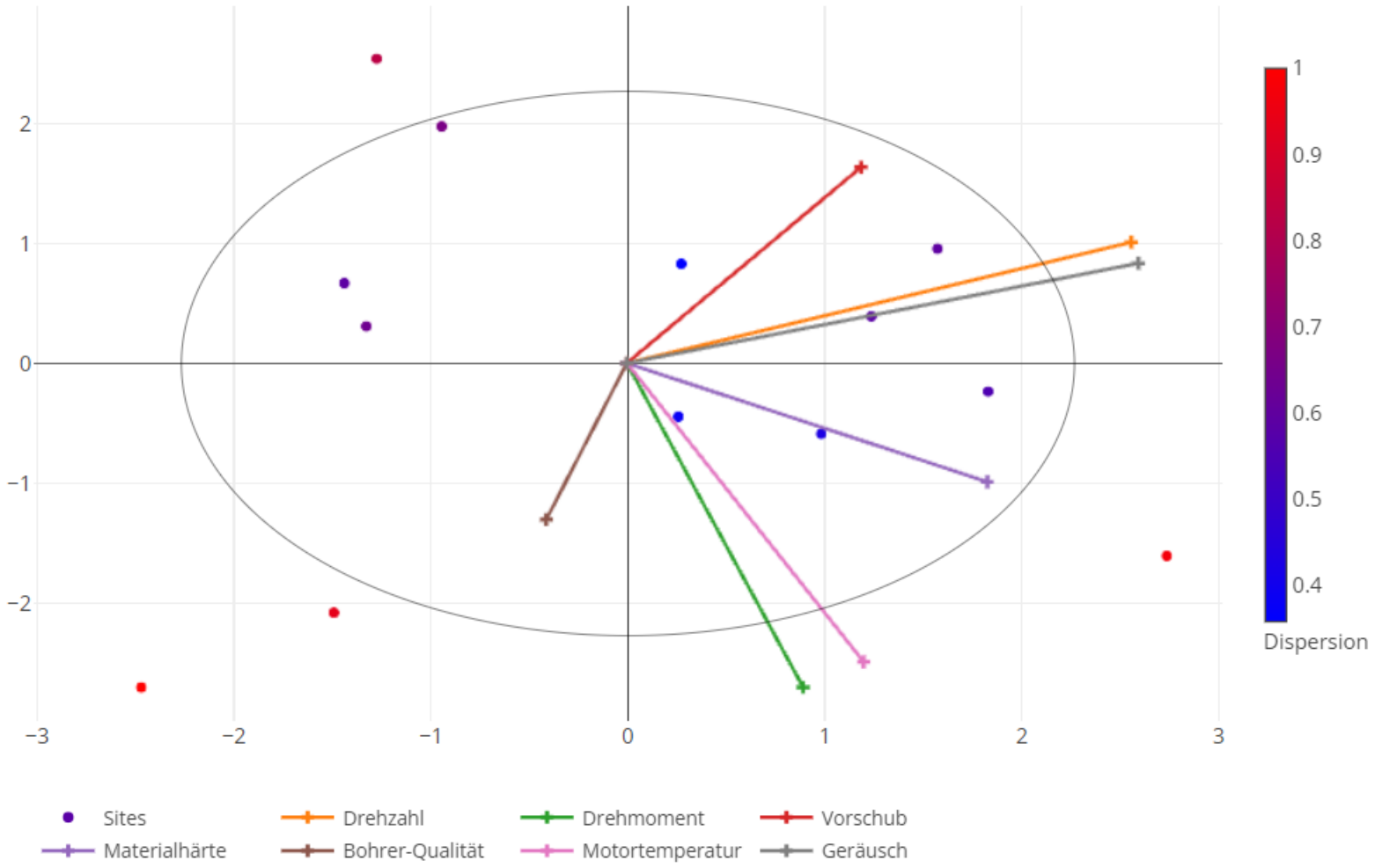
Y=
-3.578 ; 0
-1.342 ; 2.236
-1.342 ; -2.236
3.130 ; 2.236
3.130 ; -2.236

Die erste Hauptachse zeigt in Richtung der größten Varianz



HEXAGON

Principal Component Analysis (PCA) Biplot

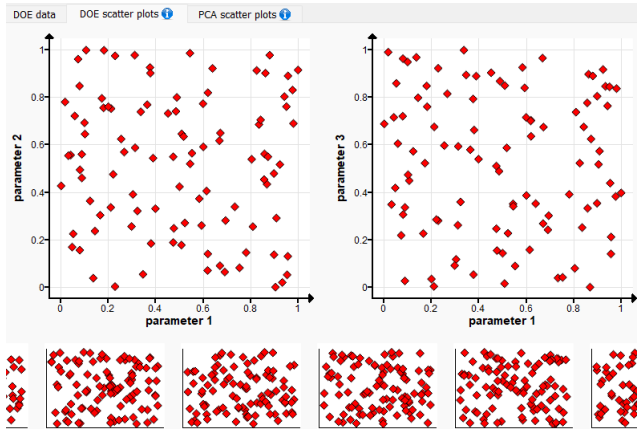


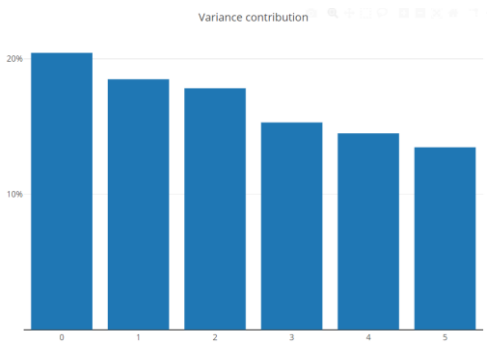
Die Daten werden in ein Hauptachsensystem projiziert.

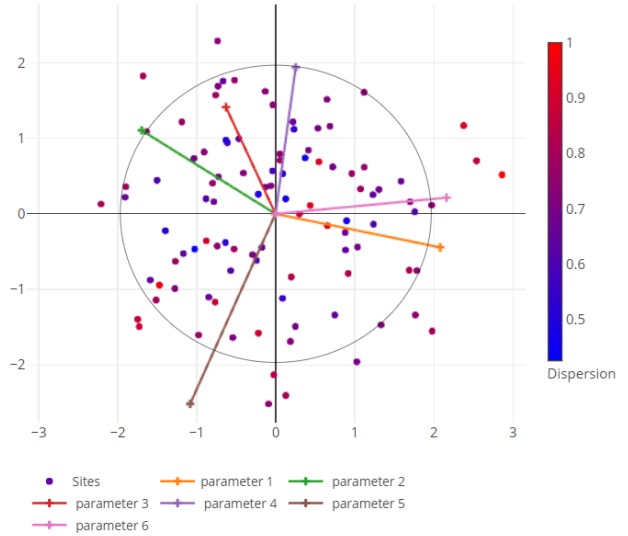
Zusammenhänge erkennen: Drehzahl und Geräusch zeigen in dieselbe Richtung

HEXAGON

Gut verteilte DOE



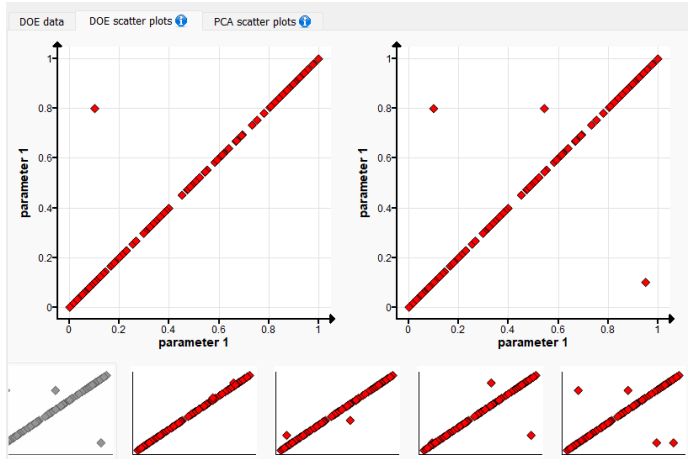


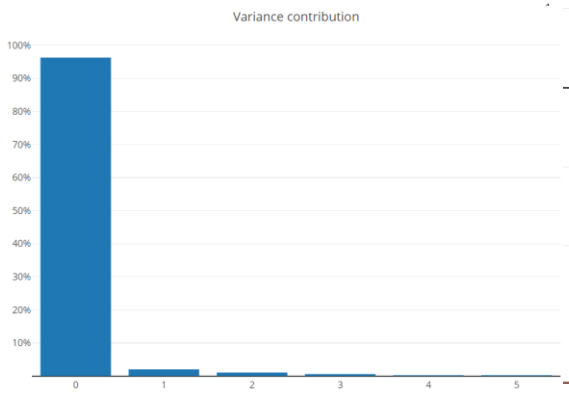


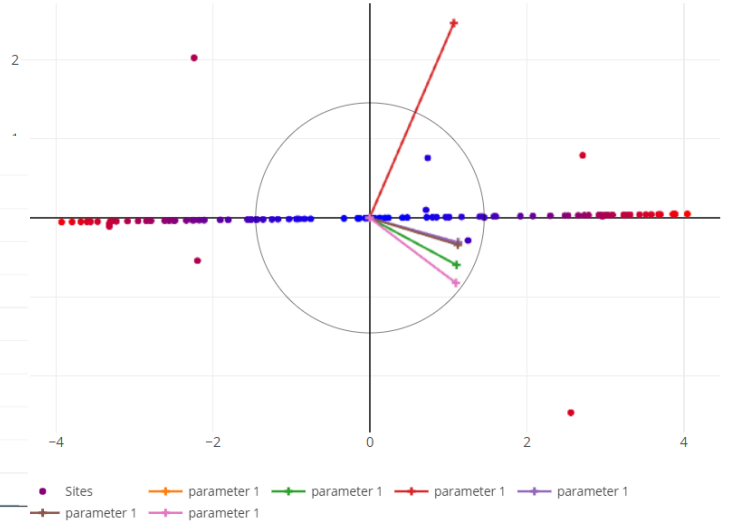
Alle Parameter tragen zur Varianz bei

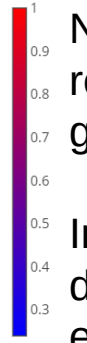
Im PCA Biplot liegen die Punkte in einem Kreis

Schlecht verteilte DOE









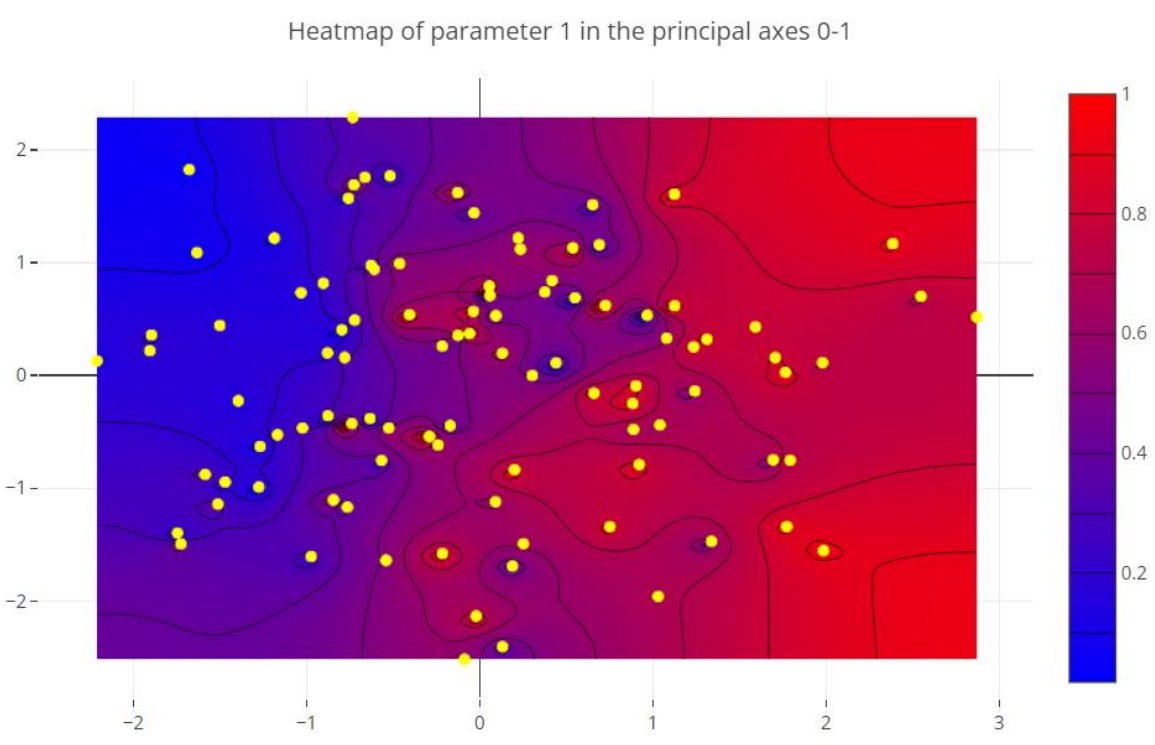
Nur ein Parameter repräsentiert fast die gesamte Varianz

Im PCA Biplot liegen die Punkte nicht in einem Kreis

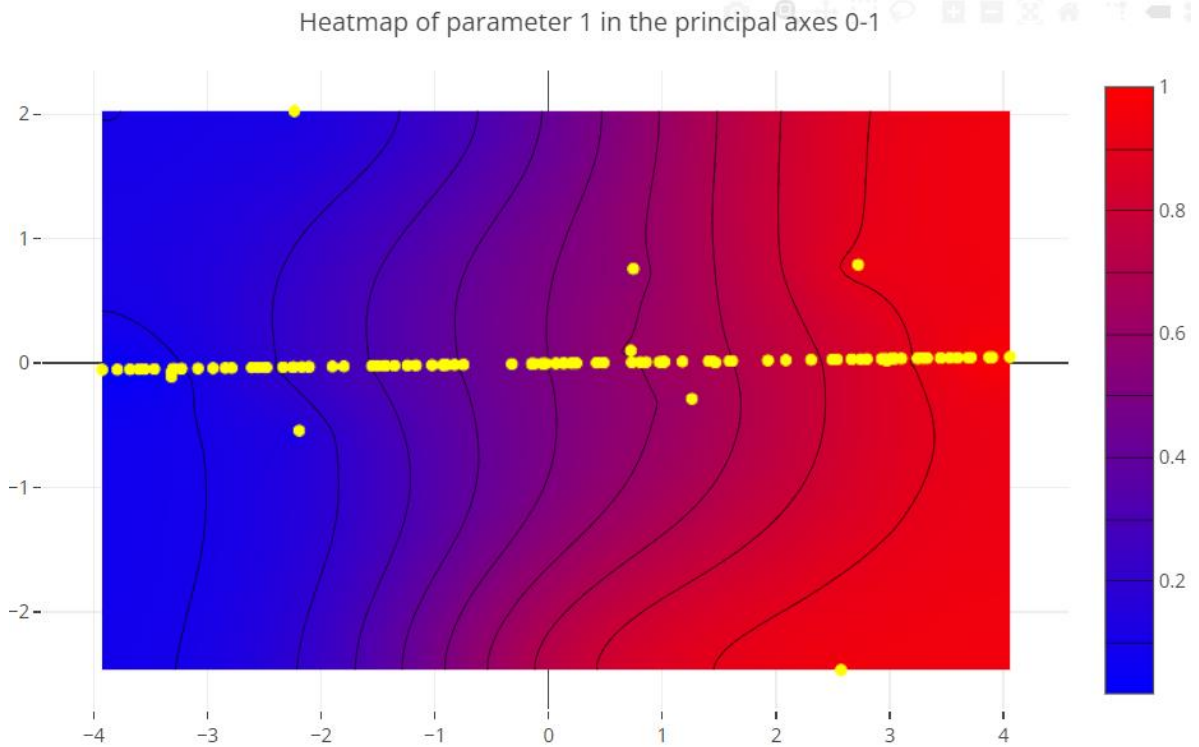
HEXAGON

PCA Heatmap

Gut verteilte DOE



Schlecht verteilte DOE



- Einfache Heatmap: z.B. Um beim Fußball Bewegungsmuster von Spielern dazustellen
- Um multidimensionale Infos in einem Bild unterzubringen: Darstellung im Hauptachsensystem
- Farbdarstellung: ein ausgewählter Parameter
- Die Farben repräsentieren die Gewichtungen der Beiträge der Parameter zu den Hauptkomponenten
- Punkte, die weit außerhalb oder allein innerhalb vieler Änderungen von Höhenlinien liegen, sind evtl. nicht gut aus anderen vorherzusagen -> DOE in diesem Bereich erweitern

HEXAGON

ODYSSEE
Explore new industrial horizons

Ariane 5 flight recordings (Aerospace) - Health monitoring

arianeGROUP

22000 records of Ariane-5 past flights sensors: compression/fusion, analysis, modeling

- **Sort information using clustering and projection plot proposed in Nebular module from ODYSSEE**
- **Virtual testing using the machine learning proposed in Lunar and Quasar modules from ODYSSEE**

« ODYSSEE's artificial intelligence solutions open up new perspectives in Ariane 5's in-flight data analysis » *B. Troclet, Structural analysis senior expert, Ariane Group*

Reduced cost:
Conduct a global analysis for fault detection and predictive modeling in a short time

New horizons:
Improve design based on flights sensors

Saving time
Analyze sensor data on-board and in real-time

From: Machine Learning & Data Mining applied to Ariane 5 Flight Recordings
Bernard Troclet, Stéphane Muller, Kambiz Kayvantash
NAFEMS France 2018

HEXAGON

ODYSSEE
Explore new industrial horizons

Was kann man mit dieser Datenauswertung anfangen?

- ✓ Zusammenhänge grafisch aufzeigen - Storytelling
- ✓ Fehler in den Daten finden
- ✓ Datenlücken identifizieren
- ✓ Envelopes darstellen
- ✓ Einflussreiche Parameter erkennen und damit das Produkt effektiv optimieren
- ✓ Machine Learning anwenden: für neue Parameterkombinationen die Ergebnisse vorhersagen

HEXAGON

Vielen Dank furs Zuhören!

Kontakt:

Cornelia Thieme
Manager Presales DACH
Design and Engineering
Manufacturing Intelligence Division
Hexagon
T: +49 89 2109 3224 ext. 4518
E: cornelia.thieme@hexagon.com

HEXAGON