

# Persönlichkeiten aus Politik, Kunst und Wissenschaft

## Alliierte Besetzung (sowj. Zone) 1948 DDR 1952 / 53

8 Pf. – Karl Marx



Eine Untersuchung zur Unterscheidung der Farbe

Farbkatalogisierung nach Michel Katalog Deutschland

1948 – Nr. 214: schwärzlichbräunlichrot - schwarzrot

1952 – Nr. 329: (dunkel) bräunlichrot

## Die Methode

Die Daten liegen als Datensatz in einer Textdatei (\*.txt) in nachfolgend beschriebener Form vor:

*d*; 838; *d*3; 7; *f*; *vX*1; 41,77; 36,57; 22,83; *sign Paul*; -1000,00; -1000,00; -1000,00; -1000,00; -1000,00; 6,46; 6,07; 5,79; 5,63; 5,85; 5,81; 5,57; 5,36; 5,45; 5,37; 5,35; 5,44; 5,56; 5,62; 5,87; 6,27; 7,25; 9,42; 13,55; 20,44; 28,12; 33,58; 36,18; 37,46; 38,08; 38,68; 39,18; 39,71; 40,23; 40,73; 41,21; -1000,00; -1000,00; -1000,00; -1000,00; -1000,00; -1000,00; -1000,00; -1000,00; 2008; 9; 29; 23; 13; 25;; *Burkhardt*; *ff*2&

Für die geplante Analyse werden aus dem Datensatz folgende Elemente ausgewählt:

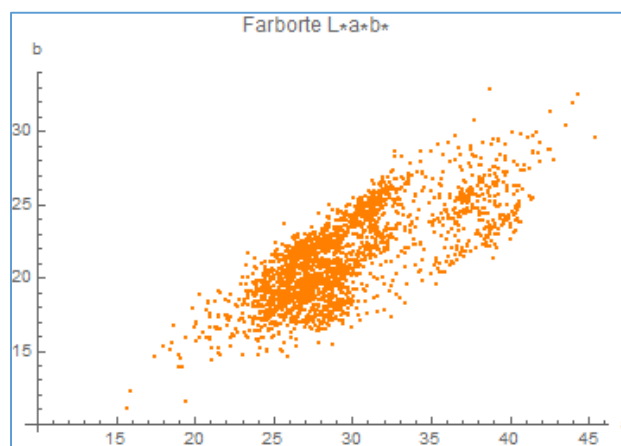
- die Nummer der Messung,
- die Farbwerte  $L^*a^*b^*$  und
- die Messwerte des Remissionsspektrums

Für die Analyse wird die Software „*Mathematica V10*“ verwendet. Die Daten werden aufbereitet (Umwandlung von „Komma“ in „Punkt“ und „Semikolon“ in „Komma“) und als Listenelemente einer Variablen zugeordnet, so dass sie mathematisch verarbeitet werden können.

$M838 = \{\{838\}, \{41.77, 36.57, 22.83\}, \{6.46, 6.07, 5.79, 5.63, 5.85, 5.81, 5.57, 5.36, 5.45, 5.37, 5.35, 5.44, 5.56, 5.62, 5.87, 6.27, 7.25, 9.42, 13.55, 20.44, 28.12, 33.58, 36.18, 37.46, 38.08, 38.68, 39.18, 39.71, 40.23, 40.73, 41.21\}\};$

Aus dem vorliegenden Datensatz wurden die Daten von 2180 Messungen aufbereitet und für die Berechnung verwendet. Über die mitgeführte Nummer der Messung können einzelne Ergebnisse direkt den Farb- und Messwerte zugeordnet werden. So haben die  $L^*a^*b^*$  Farbwerte in der Berechnung die Form  $\{41.77, 36.57, 22.83, 838\}$ . Ähnlich wird mit den Messwerten verfahren.

Die  $L^*a^*b^*$  Farbwerte können als Punktwolke in einem dreidimensionalen Koordinatensystem dargestellt werden.



Mit den  $L \times a \times b$  Farbwerten wird eine Clusteranalyse durchgeführt.

Eine Clusteranalyse ist ein mathematisches Verfahren der Multivariaten Datenanalyse aus der Statistik, mit dem Ziel, eine beliebige Anzahl von Objekten (hier die  $L \times a \times b$  Farbwerte) in möglichst homogene Gruppen (Cluster) zusammenzufassen. Der Begriff *homogene Gruppe* wird wie folgt definiert [Bacher, S.16]:

1. Die Objekte, die einer homogenen Gruppe angehören, sollen einander *ähnlich* sein.
2. Die Objekte, die unterschiedlichen homogenen Gruppen sollen verschieden sein.

Die *Ähnlichkeit* wird z.B. über *Distanzmaße* definiert:

1. Zwei Objekte sind einander umso ähnlicher, je geringer ihr Abstand zueinander ist.
2. Zwei Objekte sind umso verschiedener, je größer ihr Abstand zueinander ist.

Die Distanz zwischen zwei Objekten wird nach der L2-Norm als *euklidischer Abstand* berechnet. Die Abstände der  $L \times a \times b$  Farbwerte werden also nach der Formel

$$d_{k,j} = \sqrt{(L_k - L_j)^2 + (a_k - a_j)^2 + (b_k - b_j)^2}$$

berechnet und in einer *Distanzmatrix* gesammelt.

Für die Gruppierung von Objekten wird ein *hierarchisch-agglomeratives Verfahren* verwendet, bei dem die Clusterzentren bestimmt werden sollen. In den Berechnungsschritten wird nun genau das Datenpaar gesucht, bei dem der euklidische Abstand am kleinsten ist. Die beiden Objekte werden zu einem Cluster verschmolzen. Das Zentrum des Clusters wird als Mittelwert der beteiligten Objekte berechnet.

Die Berechnung wird beendet, wenn alle Objekte eines Datensatzes einem Cluster angehören, eine bestimmte vorher definierte Clusteranzahl erreicht wurde oder in diesem Fall der Abstand zwischen zwei Objekten (Clustern) größer als 5 wird.

Über die Nummer des Messwertes in den Objekten eines Clusters kann die Verbindung zu den jeweiligen Messwertreihen der Remissionsspektren hergestellt werden. Zur Verbesserung der Übersichtlichkeit werden die Kurvenscharen in einem *Boxplot* dargestellt.

Ein *Boxplot* ist eine Zusammensetzung verschiedener Streuungs- und Lagemaße statistischer Daten unter der Voraussetzung, dass diese einer Normalverteilung unterliegen. Der rechteckige Bereich zeigt die Lage der mittleren 50% der Daten an. Die Grenzen des Rechtecks bilden die oberen und unteren *Quartile*, die Mitte wird vom *Median* gebildet. Die oberen und unteren Antennen enden an den jeweiligen *Maximum* oder *Minimum* des Datensatzes.

## Ergebnisse der Analyse

Im Ergebnis entstehen 6 Cluster. Zusätzlich bleiben viele unbeteiligte Farbwerte bestehen.

Die Cluster sind:

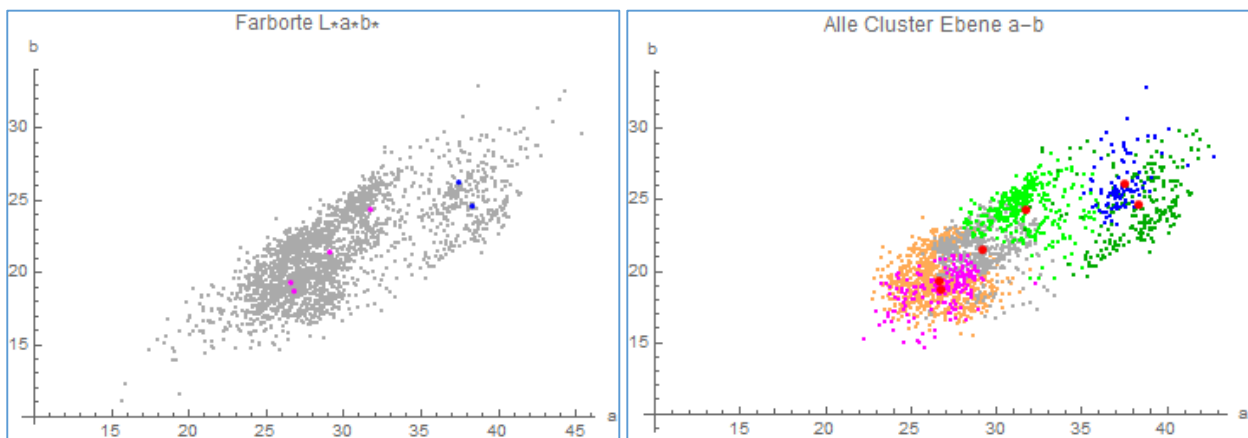
- A – Zentrum: {43.39, 26.60, 19.35}, Anzahl Punkte: 724, Abstände: {Max 6.28, Min 0.16}  
Intervall  $38.80 < L > 48.55$ ,  $22.75 < a > 31.97$ ,  $15.45 < b > 23.67$
- B – Zentrum: {40.80, 29.12, 21.47}, Anzahl Punkte: 495, Abstände: {Max 7.34, Min 0.29}  
Intervall  $35.09 < L > 44.29$ ,  $25.75 < a > 33.4$ ,  $16.63 < b > 25.41$
- C – Zentrum: {43.61, 31.67, 24.37}, Anzahl Punkte: 347, Abstände: {Max 5.95, Min 0.36}  
Intervall:  $39.08 < L > 47.27$ ,  $27.91 < a > 36.13$ ,  $19.44 < b > 28.27$
- D – Zentrum: {40.08, 38.30, 24.64}, Anzahl Punkte: 223, Abstände: {Max 7.58, Min 0.33}  
Intervall:  $34.59 < L > 44.09$ ,  $31.74 < a > 42.46$ ,  $19.53 < b > 29.86$
- E – Zentrum: {43.38, 37.43, 26.21}, Anzahl Punkte: 96, Abstände: {Max 6.76, Min 0.16}  
Intervall:  $40.66 < L > 46.16$ ,  $35.17 < a > 42.79$ ,  $23.25 < b > 32.80$
- F – Zentrum: {38.21, 26.75, 18.69}, Anzahl Punkte: 180, Abstände: {Max 7.09, Min 0.28}  
Intervall:  $33.01 < L > 40.14$ ,  $22.31 < a > 32.32$ ,  $14.63 < b > 21.09$

Es gibt 128 Farbwerte ohne Zuordnung zu einem Cluster. In der farbigen Darstellung erkennt man, dass diese fehlen.

Die Lage der Clusterzentren im Raum auf die Gesamtheit der Messwerte bezogen zeigt eine Trennung in zwei Bereiche:

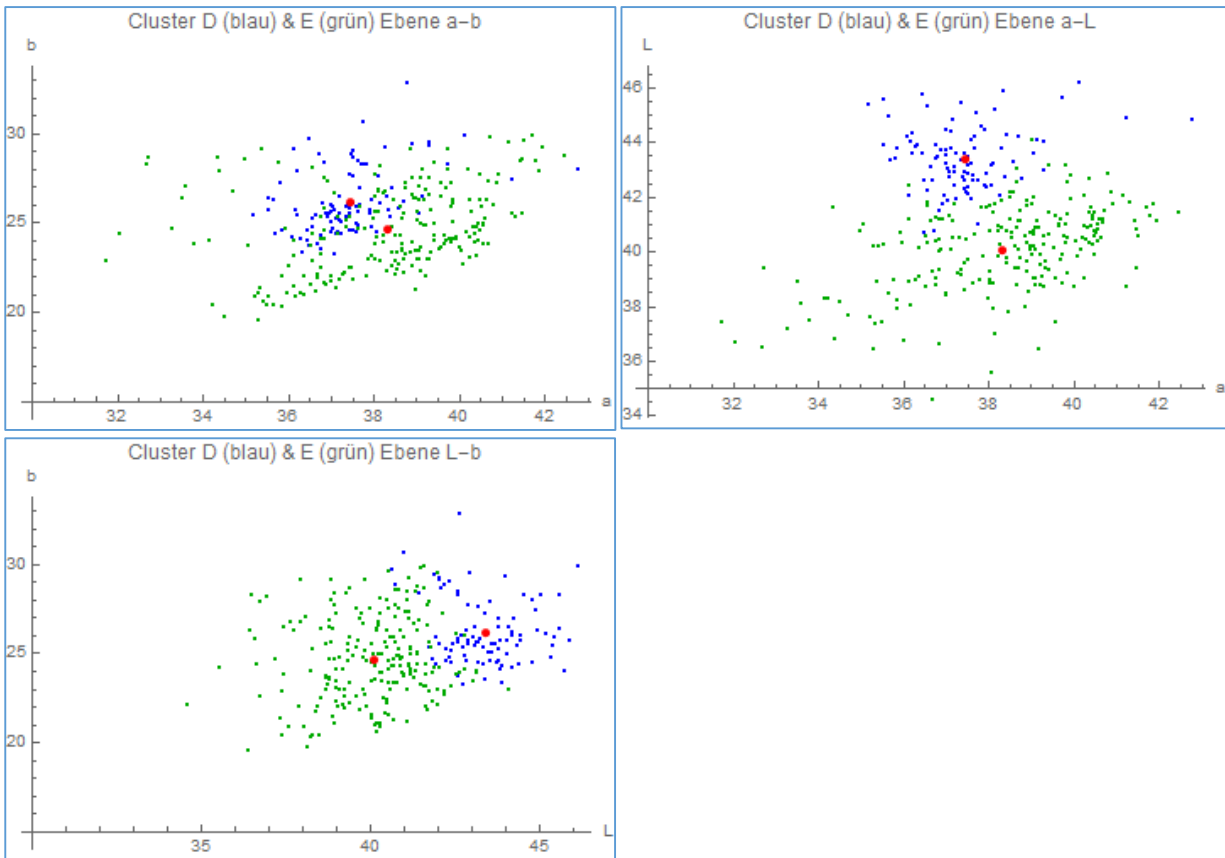
- $a < \sim 34$  : Cluster A (orange), B (grau), F (magenta)
- $a > \sim 34$  : Cluster D (blau), E (grün)

Cluster C (grün) hat eine besondere Position zwischen den beiden Hauptgruppen.



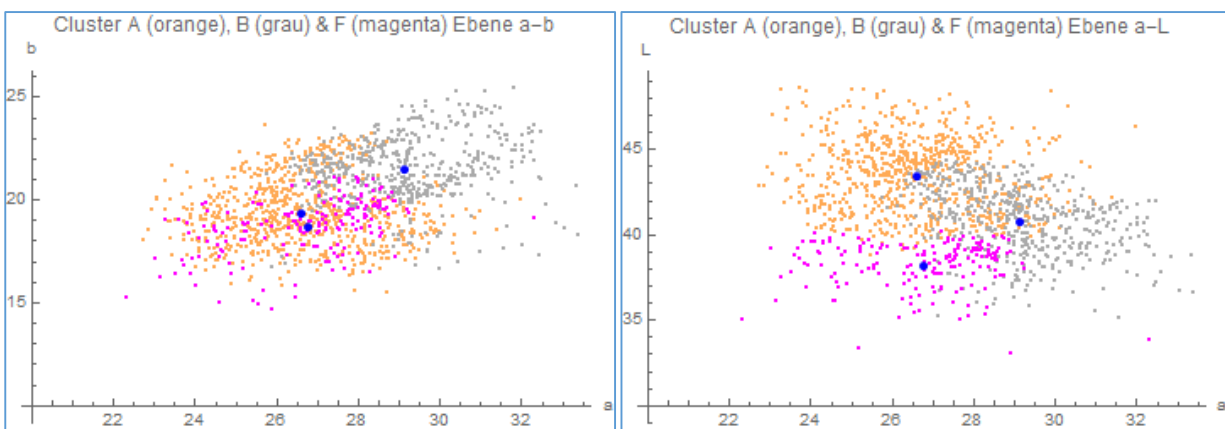
### Der Bereich $a > 34$

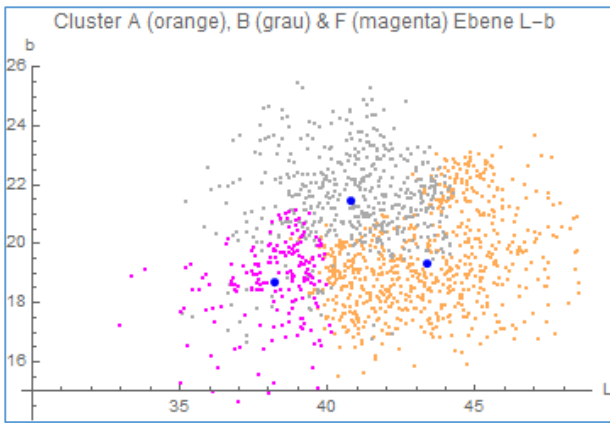
Die Cluster D (blau) und E (grün) liegen relativ dicht beieinander. Sie unterscheiden sich in den Parametern „a“ und „b“ kaum, jedoch deutlich in der Helligkeit.



### Bereich $a < 34$

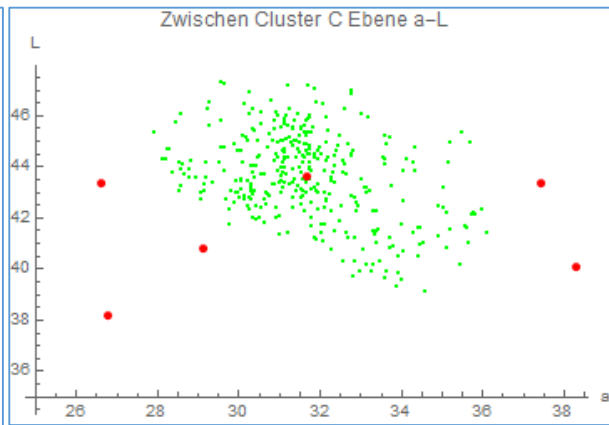
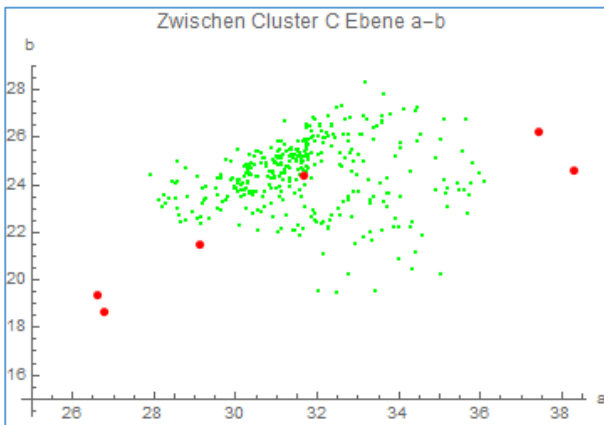
Die Cluster A (orange) und F (magenta) liegen in den Parametern „a“ und „b“ dicht zusammen, sie unterscheiden sich nur in der Helligkeit. Der Cluster B (grau) aus dieser Gruppe weicht hingegen in allen 3 Parametern etwas ab. Die Verschiebung in Richtung rot und blau wird jedoch durch die Helligkeit ausgeglichen.





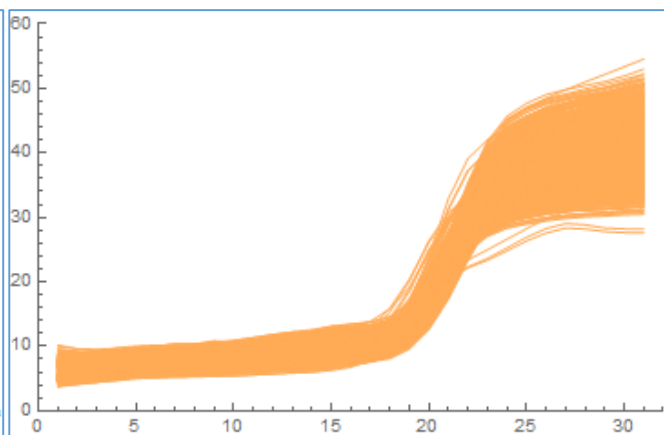
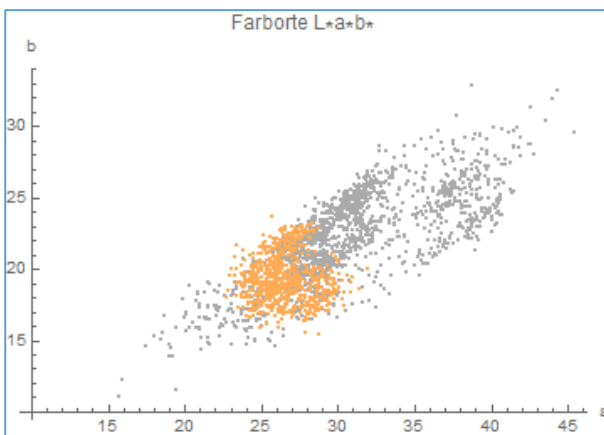
### Zwischencluster C

Der Cluster C (grün) liegt zwischen den beiden Hauptgruppen und streut in deren Randbereiche. Es erscheint zweckmäßig, diesen Cluster aufzulösen.

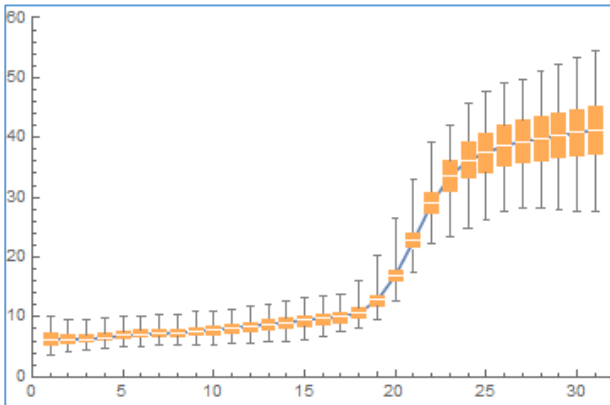


### Ergebnisse der Analyse im Einzelnen

Die Punkteanordnung hellorange mit den grauen Punkten der übrigen Werte zeigt die Einordnung des Clusters „A“ im Raum. Die zugehörigen Messwertreihen zeigt das danebenstehende Diagramm. Die Darstellung der Reihen übereinander gibt allerdings keinen guten Überblick.

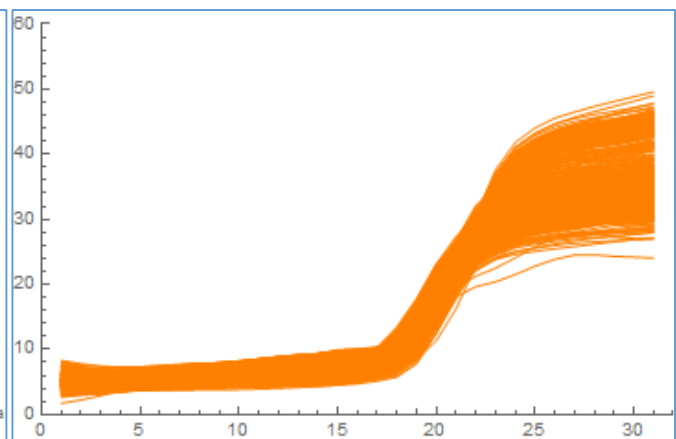
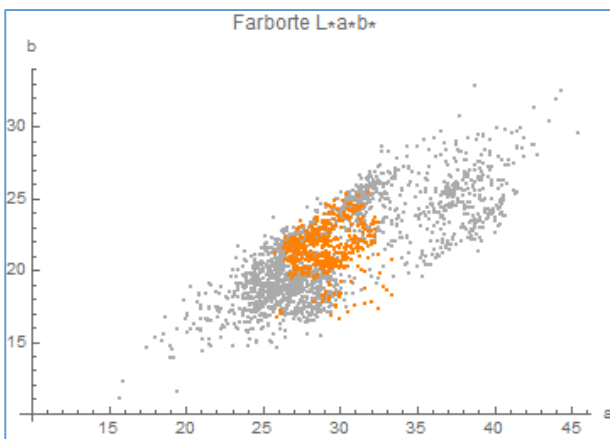


Eine Verbesserung bringt hier die Darstellung der Messwertreihe „A“ in einem Boxplot.

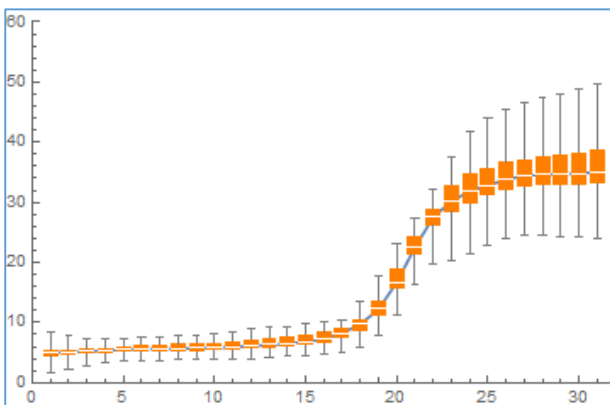


Im Folgenden die Darstellungen der weiteren Cluster.

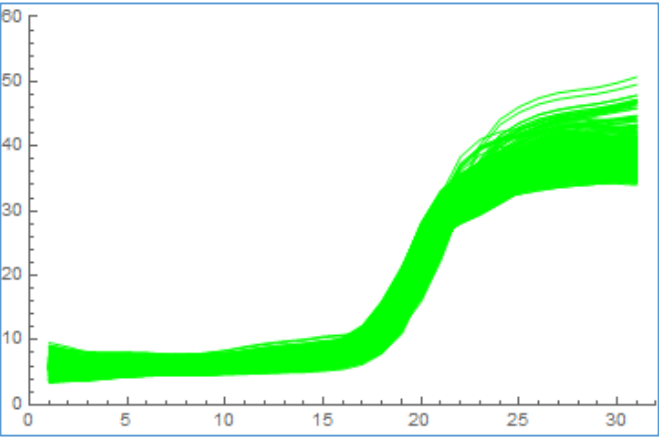
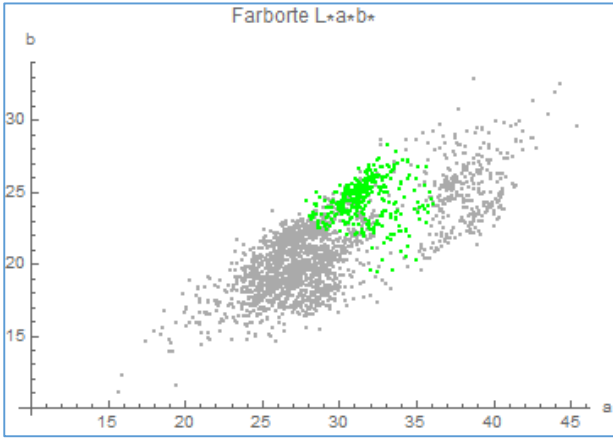
Cluster „B“



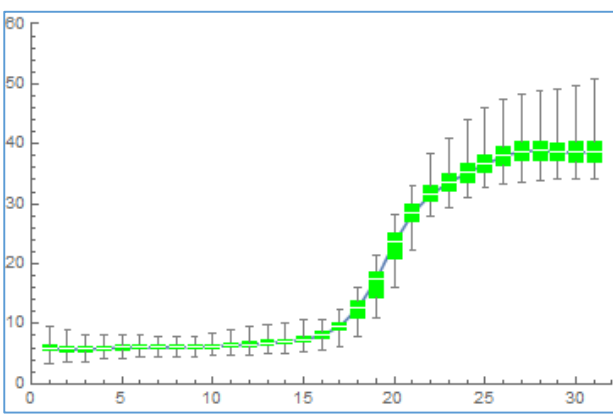
Boxplot Cluster „B“



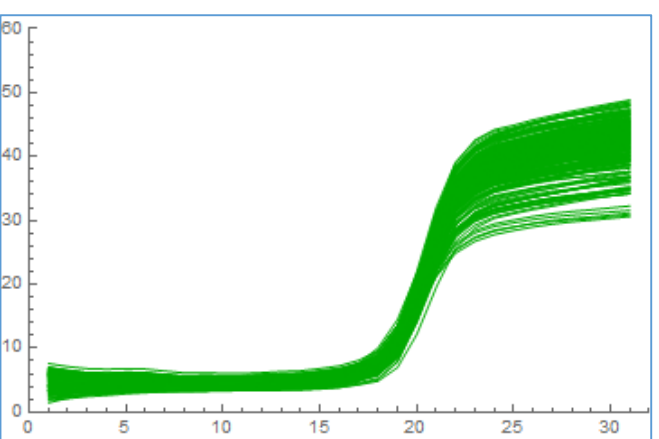
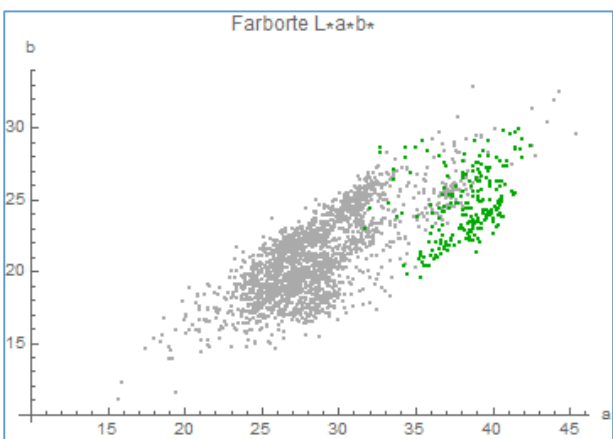
### Cluster „C“



### Boxplot „C“

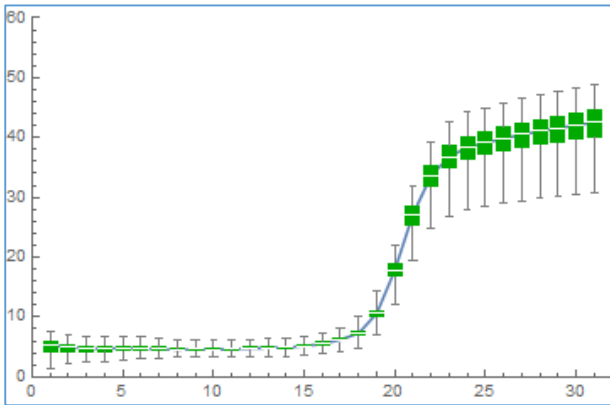


### Cluster „D“

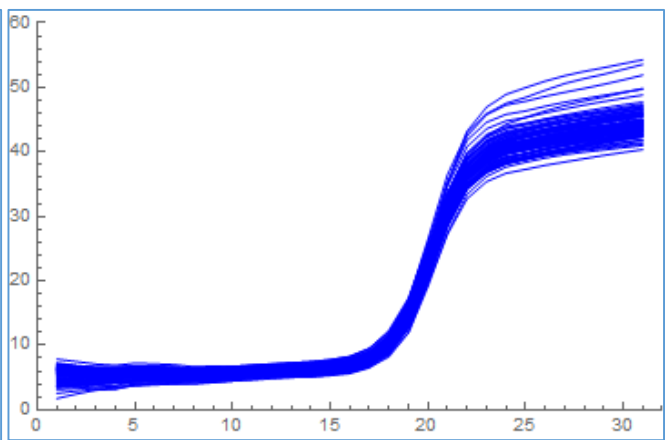
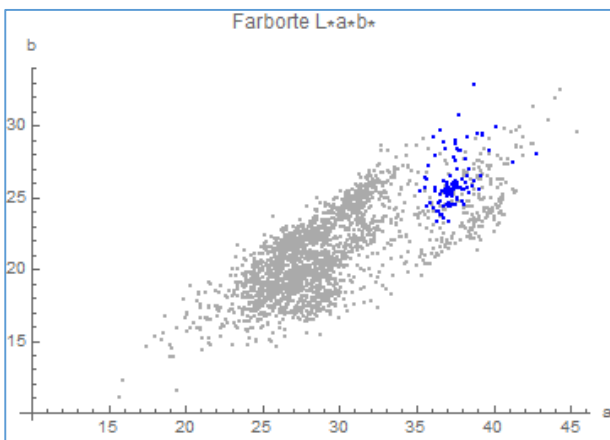




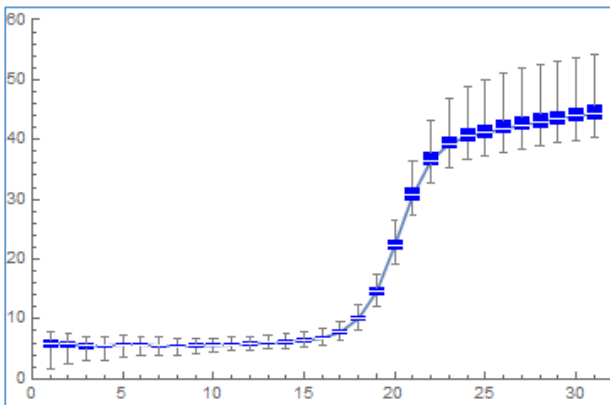
Boxplot „D“



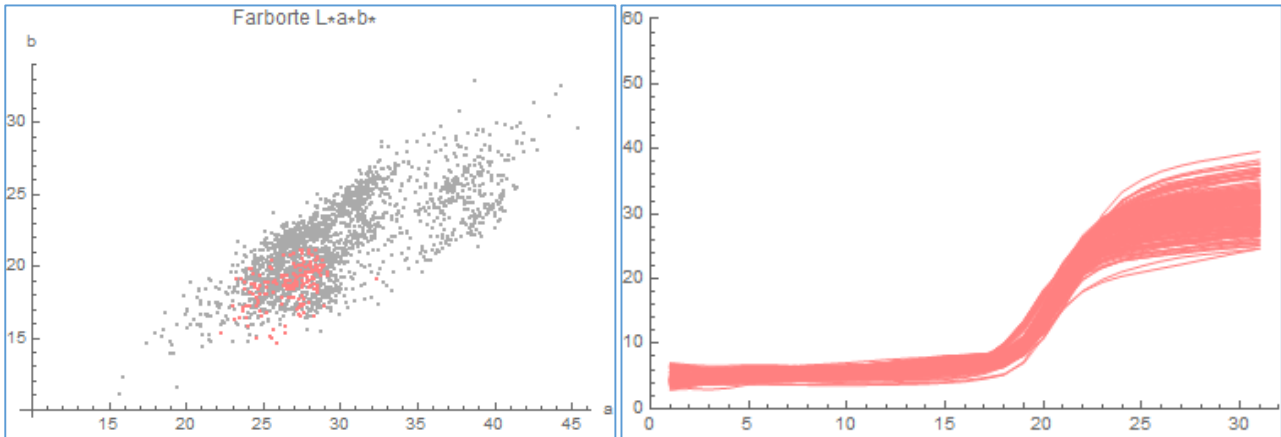
Cluster „E“



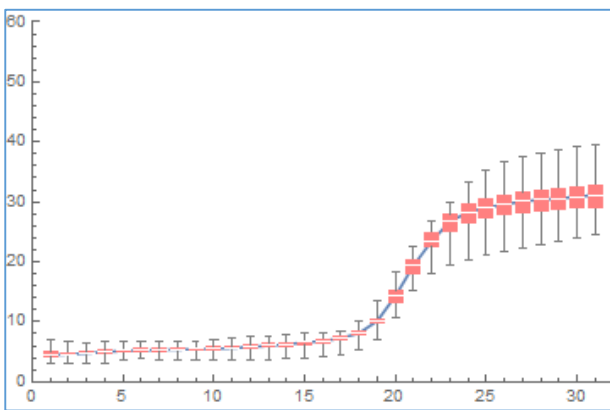
Boxplot „E“



## Cluster „F“



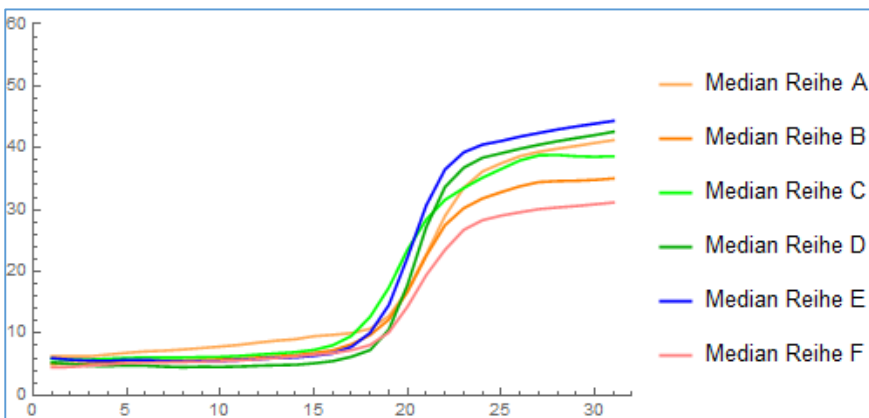
## Boxplot „F“



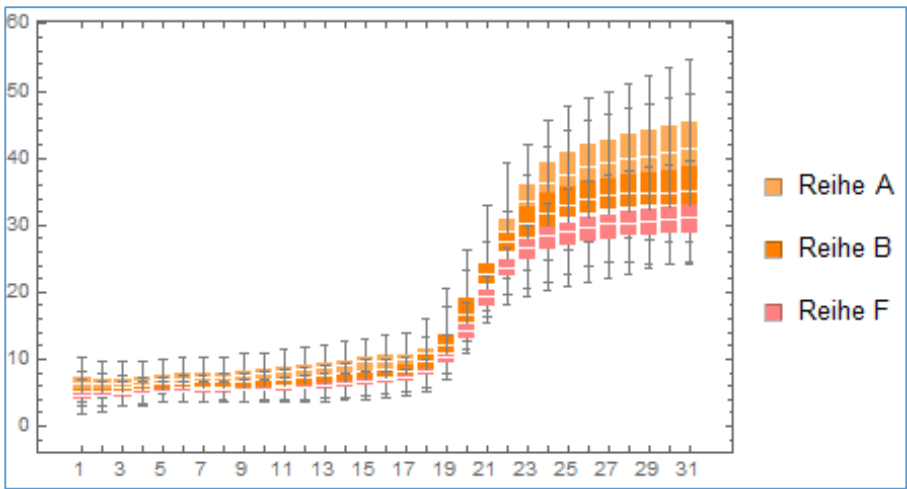
Einen guten Überblick die Lage der Messwertereihen zueinander gibt der Median der Messwertereihen.

Der Median ist ein Mittelwert einer statistischen Verteilung eines Datensatzes. Das Besondere am Median ist, dass er genau die Mitte des Datensatzes darstellt. Er ist relativ unempfindlich gegenüber weit vom Zentrum entfernten Datenpunkten.

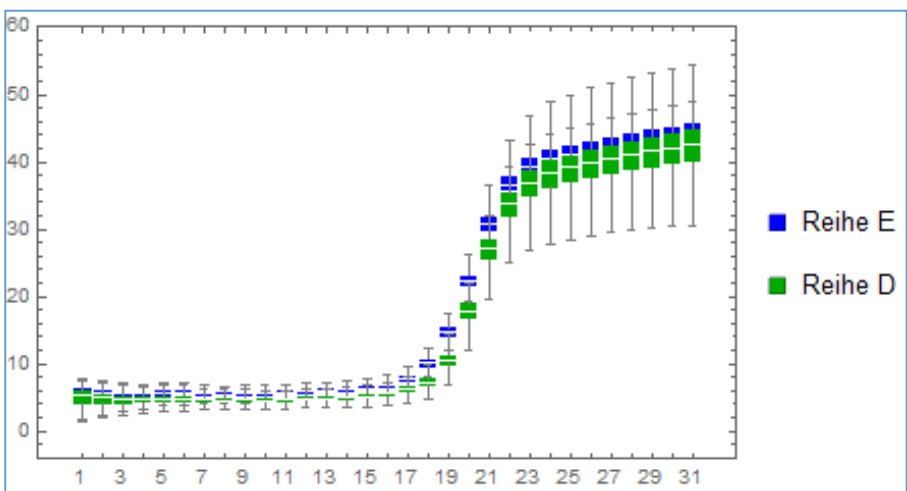
Zusammenhänge der Messwertereihen sind nicht ersichtlich, da sich die Kurven häufig überlagern und schneiden.



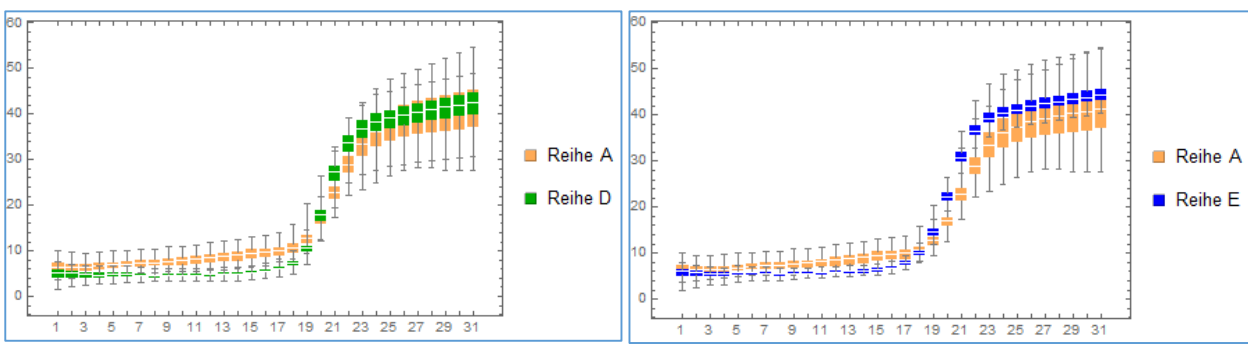
Aus der räumlichen Darstellung der Cluster „A“, „B“ und „F“ ist ein möglicher Zusammenhang bereits bekannt. Die Darstellung im Boxplot bestätigt das deutlich. Die 50% Rechtecke liegen gut erkennbar untereinander, während sich die Antennen durchaus überlagern. Die Lage der Cluster entspricht einer Abstufung in der Helligkeit „L“ und des Anteils „b“.



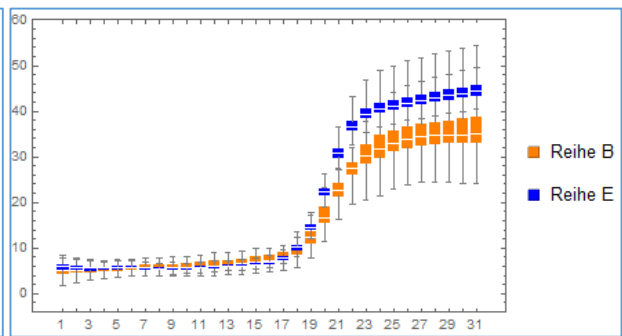
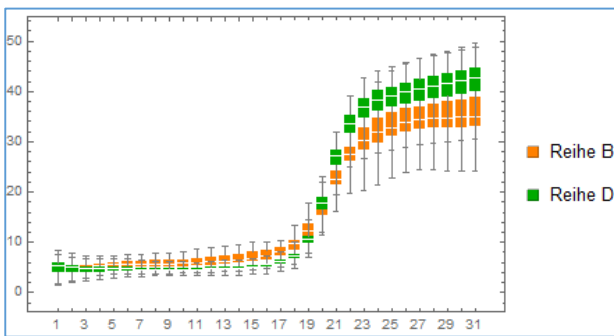
Ähnliche Eigenschaften finden sich in der Darstellung der Messwertreihen „E“ und „D“.



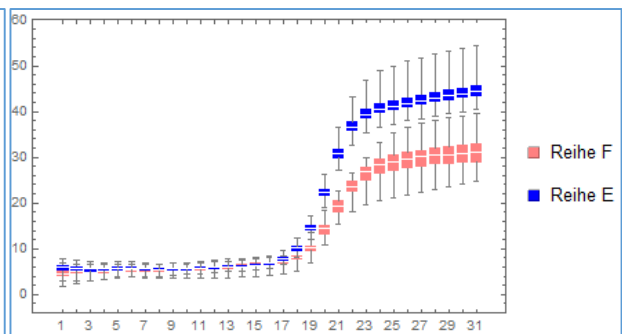
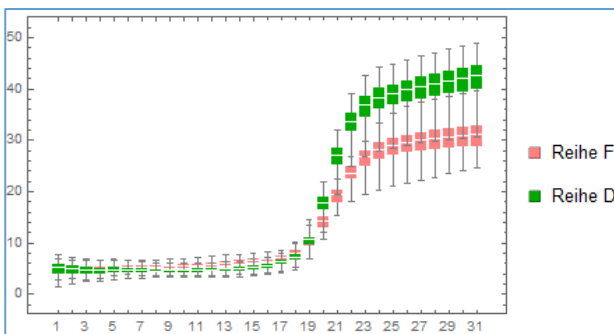
Stellt man die Reihen „A – „D“ sowie „A – „E“ gegenüber, erkennt man die Unterschiede in den Messwertreihen.



Ebenso für die Reihen „B“ – „D“ und „B“ – „E“ .



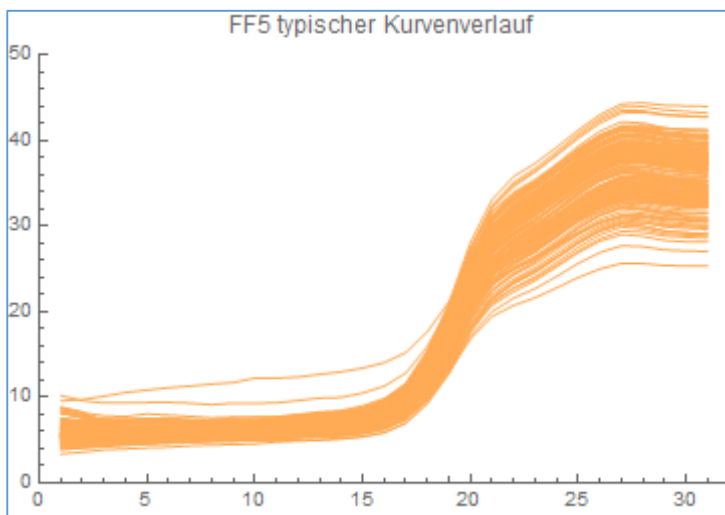
Und für die Reihen „F“ – „D“ und „F“ – „E“ .



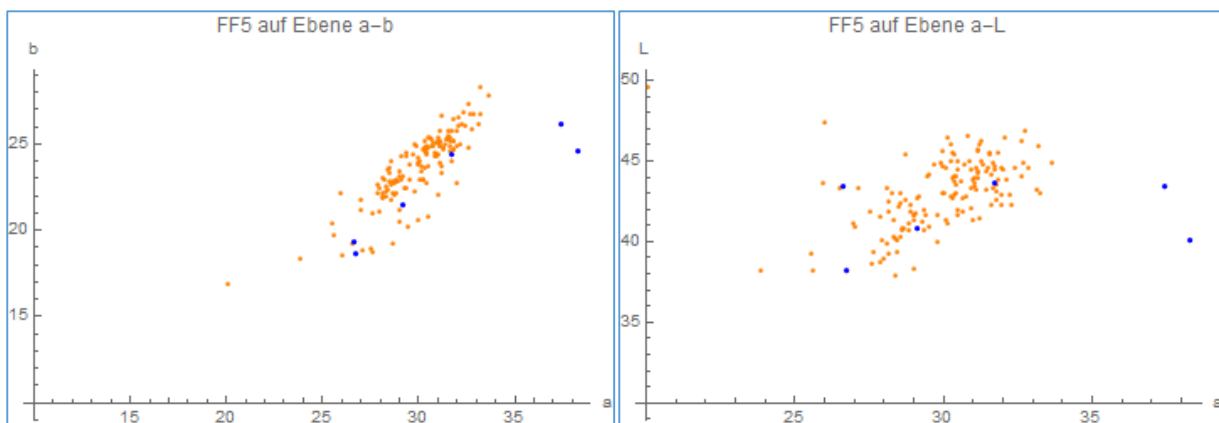
## Forschungsfarbe 5 „FF5“

„Kurvenbeschreibung: [...] FF5 (rot) und FF6 (lila) zeigen davon abweichende sehr typische Kurvenverläufe, die sie damit eindeutig vom Rest zu unterscheiden. FF5 steigt früher an, schon ab 550nm, hat dann einen scharfen Knick bei 600nm und einen weiteren Knick bei 660nm, um dann wieder zu fallen.“ [Burkhardt, S.39]

Der typische Kurvenverlauf bildet also ein Maximum bei 660nm an den Messwerten 27 und 28 des Datensatzes. Die ermittelten Kurven der Messwerte lassen sich damit leicht charakterisieren. Anhand dieses Kriteriums wurden 146 Datensätze aus der Gesamtheit der Daten extrahiert.



Die folgenden Diagramme zeigen die Lage der FF5 im Raum mit Bezug zur Gesamtheit der Messpunkte und den Zentren der berechneten Cluster.



Aus der Darstellung erkennt man die Zuordnung der Messpunkte zu den Clusterzentren, die sich wie folgt berechnen lassen:

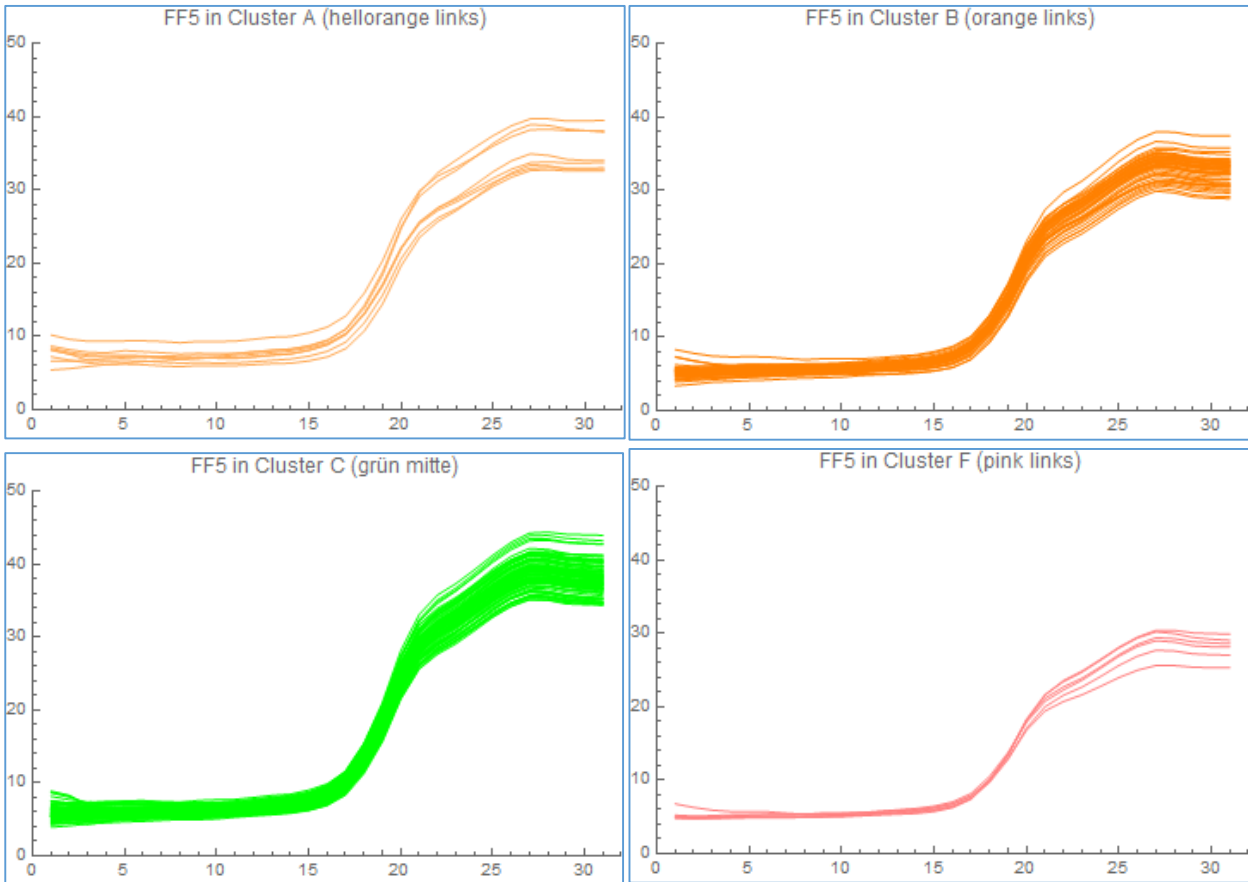
Cluster D und E (rechts): keine

Cluster C (Mitte): 88

Cluster A, B und F (links): 8, 43, 6

Lediglich der Messwert 3264 mit  $L^*a^*b^* = \{49.54, 20.07, 16.88\}$  konnte keinem Cluster zugeordnet werden.

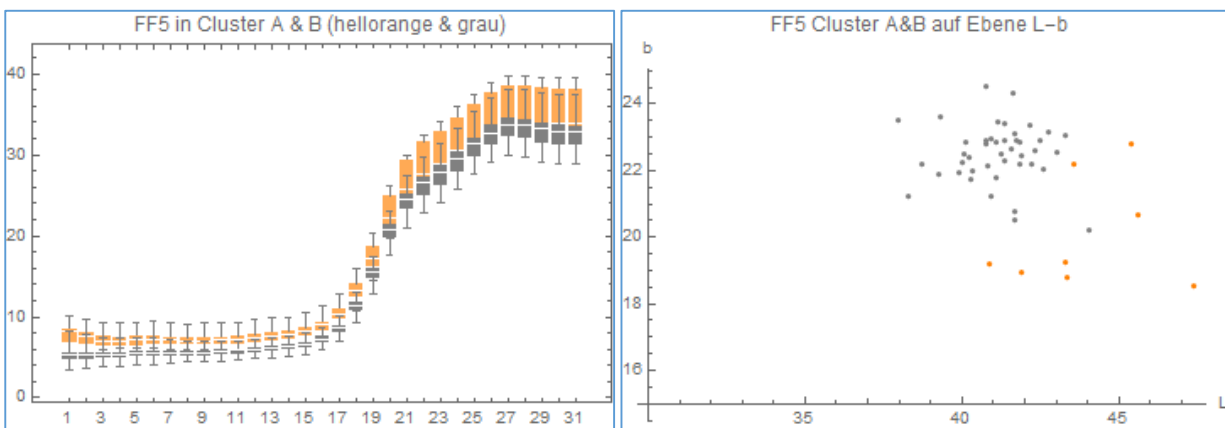
Die Zuordnung der Kurvenverläufe zu den einzelnen Clustern im Detail.



### Die Zusammenhänge im Detail

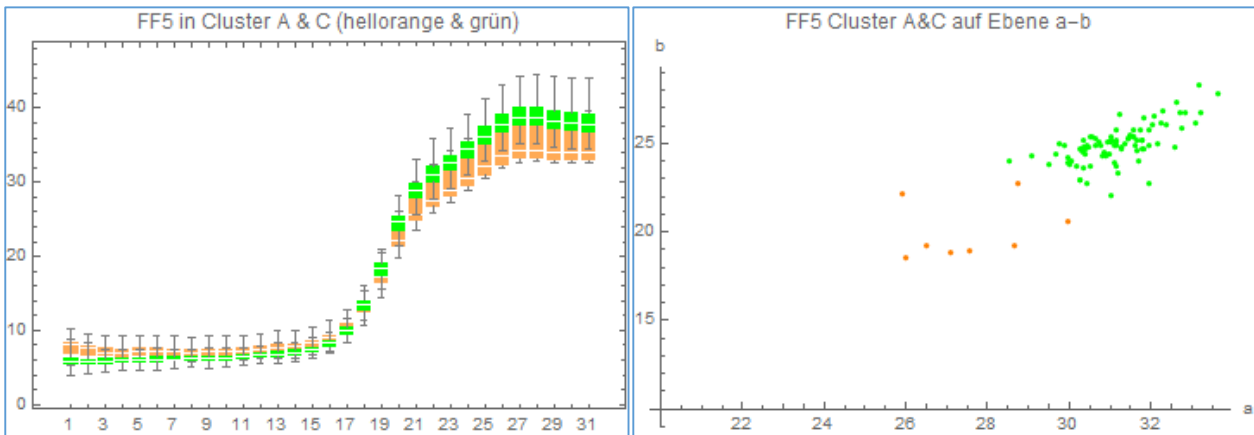
#### Cluster A und B

Die im Boxplot dargestellten Wertebereiche des Clusters A (Orange) liegen über dem Cluster B (grau). Die Unterschiede liegen im wesentlichen in der Helligkeit „L“ und im Parameter „b“. Die Punkte, die zu Überschneidungen im Bosplot führen, befinden sich im rechten Bereich des Punktediagramms  $\{L \sim 45, b \sim 22\}$ .



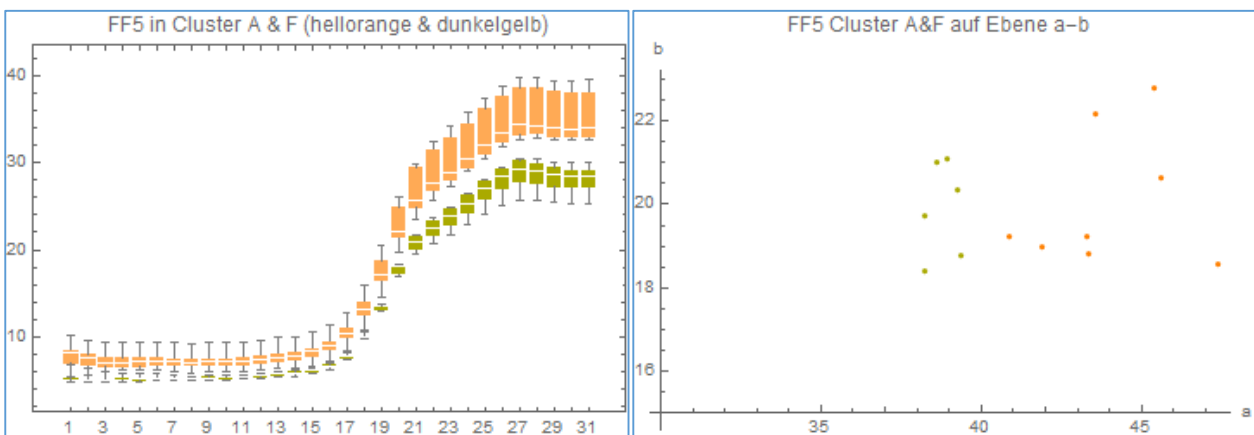
## Cluster A und C

Die im Boxplot dargestellten Wertebereiche des Clusters A (Orange) wechseln ihre Lage. Bei den niedrigen Messwerten liegen sie über dem Cluster C und bei den höheren Messwerten unter ihnen. Die Unterscheidung im Punktediagramm findet bei den Parametern „a“ und „b“ statt. Die Überschneidung aus den Kurven wird in der räumlichen Darstellung nicht abgebildet.



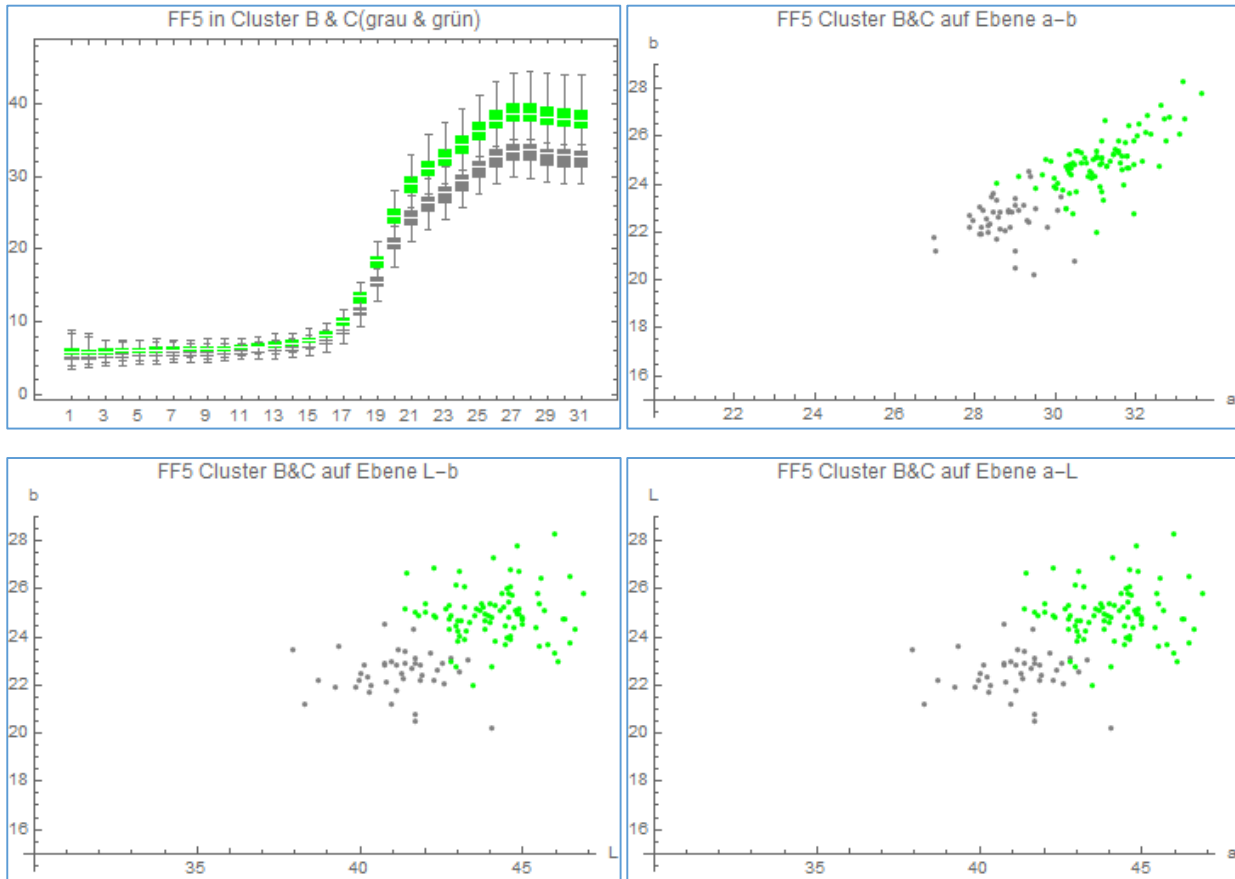
## Cluster A und F

Die im Boxplot dargestellten Wertebereiche des Clusters A (Orange) liegen über dem Cluster F (dunkelgelb). Die Unterschiede liegen nur im Parameter „a“.



## Cluster B und C

Die im Boxplot dargestellten Wertebereiche des Clusters B (Grau) liegen unter denen des Clusters C (grün). Die Unterschiede liegen in den Parametern „a“ und „b“, sowie in der Helligkeit „L“. In den Punktediagrammen sind diese zu erkennen. Deutliche Überschneidungen gibt es bei den niedrigen Messwerten.



Die Merkmale der Cluster B und C in der Gegenüberstellung

### Cluster B

Verschieben in Richtung rot „+a“  
Verschieben in Richtung gelb „b“  
Etwas heller „> L“

### Cluster C

Verschieben in Richtung grün „-a“  
Verschieben in Richtung blau  
Etwas dunkler „< L“

## Zusammenfassung

„FF5 ist die als ‚c‘ katalogisierte Farbe, gekennzeichnet durch einzigartigen Kurvenverlauf und visuell eigenständig.“ [Burkhardt, S.39]

Die FF5 bildet keine eigenständige Farbe, die sich mit der besonderen Form des Kurvenverlaufs begründen lässt. Vielmehr ordnen sich ein großer Teil der Messwerte in die Cluster B und C, sowie ein geringerer Teil in die Cluster A und F ein. Sie sind damit in der Farbwahrnehmung deutlich voneinander zu unterscheiden.



Literatur:

C. Burghardt, W. Podien, Die Köpfeserie 1948 – 1954, Handbuch und Spezialkatalog

J. Bacher, A. Pöge, K. Wenzig, Clusteranalyse – Anwendungsorientierte Einführung in Klassifikationsverfahren, Oldenbourg Verlag München.

Backhaus, Erichson, Plinke, Weiber, Multivariate Analysemethoden – eine Anwendungsorientierte Einführung, Springer Verlag.