

Persönlichkeiten aus Politik, Kunst und Wissenschaft Alliierte Besetzung (sowj. Zone) 1948 DDR 1952 / 53

2 Pf. – Käthe Kollwitz



Eine Untersuchung zur Unterscheidung der Farbe

Farbkatalogisierung nach Michel Katalog Deutschland

1948 – Nr. 212: dunkel- bis schwarz (blau) grau, grauschwarz

1952 – Nr. 327: dunkelgrau bis grauschwarz

Die Methode

Datenstruktur

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

d; 8; 3a; 7; f; zX2; 42,04; -1,67; 4,39; gepr_Schönherr; -1000,00; -1000,00; -1000,00; -1000,00; -1000,00; 8,97; 9,15; 9,70; 10,23; 10,64; 10,98; 11,30; 11,50; 11,78; 11,96; 12,26; 12,55; 12,76; 12,87; 13,00; 12,96; 12,77; 12,48; 12,24; 12,19; 12,16; 12,16; 12,41; 12,97; 13,69; 14,37; 14,91; 15,02; 15,01; 15,11; 15,26; -1000,00; -1000,00; -1000,00; -1000,00; -1000,00; -1000,00; -1000,00; -1000,00; 2008; 9; 24; 18; 23; 38; ; Burkhardt; ff2&

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.

$M8 = \{\{8\}, \{42.04, -1.67, 4.39\}, \{8.97, 9.15, 9.70, 10.23, 10.64, 10.98, 11.30, 11.50, 11.78, 11.96, 12.26, 12.55, 12.76, 12.87, 13.00, 12.96, 12.77, 12.48, 12.24, 12.19, 12.16, 12.16, 12.41, 12.97, 13.69, 14.37, 14.91, 15.02, 15.01, 15.11, 15.26\}\};$

Aus dem vorliegenden Datensatz wurden die Daten von 3376 Messungen aufbereitet und für die Berechnung verwendet. Über die mitgeführte Nummer der Messung können einzelne Ergebnisse direkt den Farb- und Messwerten zugeordnet werden. So haben die L*a*b* Farbwerte in der Berechnung die Form $\{42.04, -1.67, 4.39, 8\}$. Ähnlich wird mit den Messwerten verfahren.

Clusteranalyse

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.

$$\begin{bmatrix} & L_1 & L_2 & \dots & L_n \\ L_1 & 0 & d_{1,2} & \dots & d_{1,n} \\ L_2 & d_{2,1} & 0 & \dots & d_{2,n} \\ \vdots & & \vdots & & \vdots \\ L_n & d_{n,1} & d_{n,2} & \dots & 0 \end{bmatrix}$$

Für die Gruppierung von Objekten wird ein *hierarchisch-agglomeratives Verfahren* verwendet, bei dem die Clusterzentren bestimmt werden sollen. In den einzelnen 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.

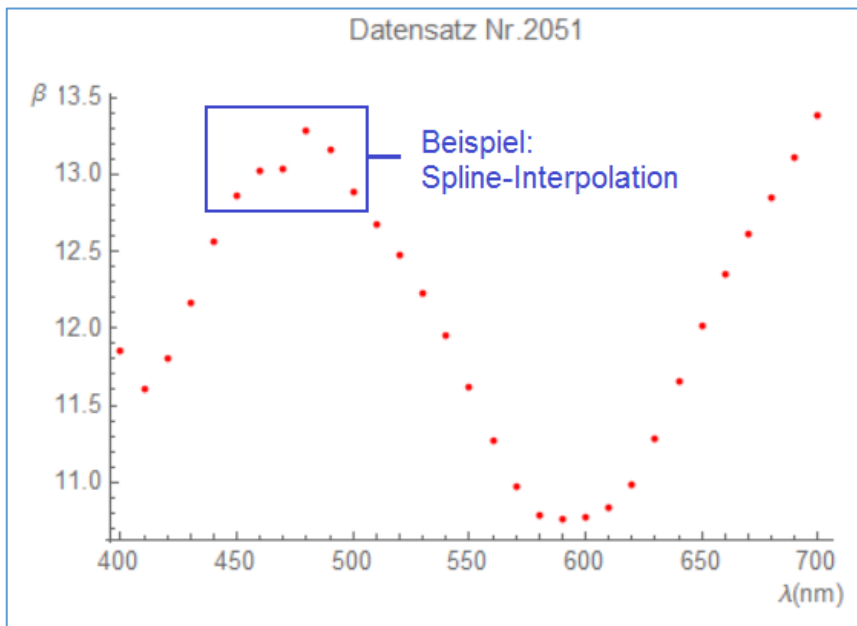
Die Berechnung wird beendet, wenn alle Objekte eines Datensatzes einem Cluster angehören, eine bestimmte vorher definierte Clusteranzahl erreicht wurde oder in diesem Fall der Abstand zwischen zwei Objekten (Clustern) größer als 5 wird.

Spline – Interpolation

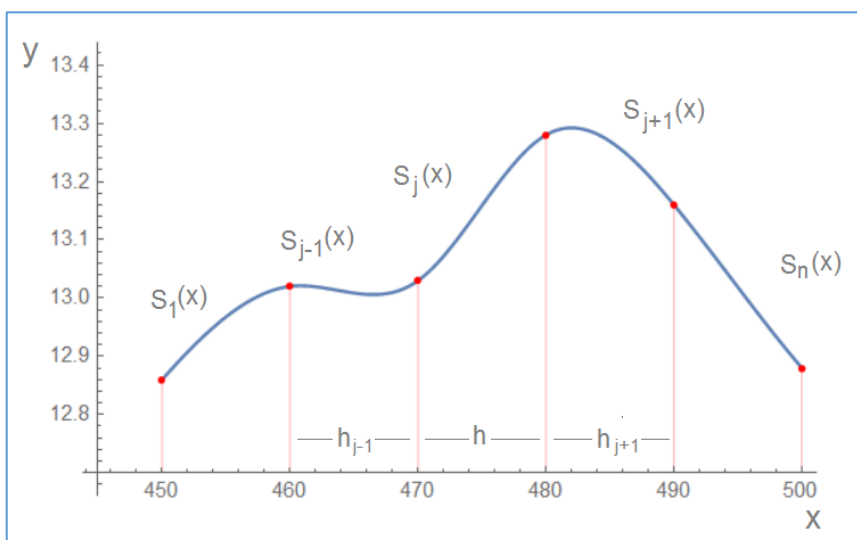
Zur Erklärung der Spline-Interpolation wird aus den Messwerten der Datensatz Nr. 2051 verwendet.

$M_{2051} = \{ \{2051\}, \{40.8, -2.02, -2.59\}, \{11.86, 11.6, 11.81, 12.17, 12.56, 12.86, 13.02, 13.03, 13.28, 13.16, 12.88, 12.67, 12.48, 12.23, 11.95, 11.62, 11.27, 10.97, 10.78, 10.76, 10.77, 10.83, 10.98, 11.28, 11.65, 12.01, 12.35, 12.61, 12.85, 13.11, 13.38\} \}$

Die dargestellten Messpunkte bilden nachfolgende Kurve, aus der ein Abschnitt vorgestellt werden soll.



Zwischen je zwei benachbarte Punkte $\{x_j, y_j\}$ und $\{x_{j+1}, y_{j+1}\}$ wird eine Funktion 3. Grades $S(x) = d(x - x_j)^3 + c(x - x_j)^2 + b(x - x_j) + a$ gesucht, so dass die Übergangsstellen zwischen benachbarten Funktionen tangential sind. [1]



Für die Funktionen gelten folgende Bedingungen:

- Am linken Rand der Funktion erhält man für x_{j-1} den Funktionswert y_{j-1} .

$$S_j(x_{j-1}) = y_{j-1}$$
- Die Gleichheit benachbarter Funktionen am linken Rand wird über die 1. Ableitung gewährleistet.

$$S'_{j-1}(x_{j-1}) = S'_j(x_{j-1})$$
- Am rechten Rand der Funktion erhält man für x_j den Funktionswert y_j .

$$S_j(x_j) = y_j$$
- Die Gleichheit benachbarter Funktionen am rechten Rand wird über die 2. Ableitung gewährleistet.

$$S''_j(x_j) = S''_{j+1}(x_j)$$

Aus den 4 Beziehungen erhält man die Variablen a_j, b_j, d_j, g_j . Weiter gelten an den linken und rechten Funktionen ohne Nachbarn die Bedingungen

$$S''_1(x_0) = 0 \quad \text{und} \quad S''_n(x_n) = 0$$

Für den Fall gleichabständiger x- Werte gilt zudem

$$h = \text{konst.} = 10$$

Für den Parameter c_j wird folgendes Gleichungssystem aufgestellt und mit Hilfe des Cholesky – Verfahrens, sowie Vorwärts- und Rückwärtseinsetzen gelöst.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 40 & 10 & 0 & & 0 & 0 \\ 0 & 10 & 40 & 10 & & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & & 10 & 0 \\ 0 & 0 & 0 & 0 & & 40 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ c_2 \\ c_3 \\ \vdots \\ c_4 \\ c_n \\ 0 \end{bmatrix} = \frac{3}{h^2} \begin{bmatrix} y_0 - 2y_1 + y_2 \\ y_1 - 2y_2 + y_3 \\ \vdots \\ \vdots \\ \vdots \\ y_{n-2} - 2y_{n-1} + y_n \end{bmatrix}$$

Die Parameter a_j, b_j, d_j werden nach den Gleichungen

$$a_j = y_{j-1}$$

$$b_j = \frac{(y_j - y_{j-1})}{h} - \frac{h(2c_j + c_{j+1})}{3}$$

$$d_j = \frac{(c_{j+1} - c_j)}{3h}$$

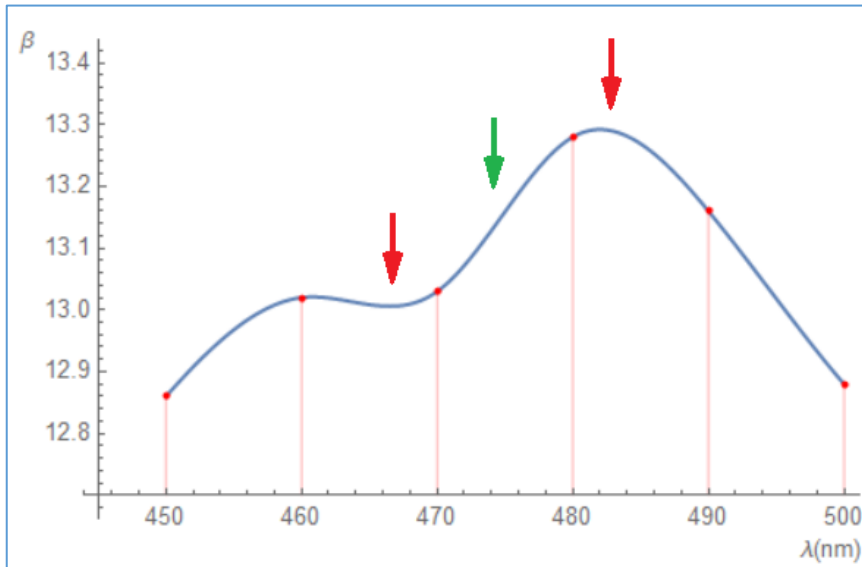
berechnet.

Für 31 Messwerte aus den Datensätzen erhält man eine Funktionenschar über 30 Funktionen.

$$S(x) = \begin{cases} S_1(x) = d(x - x_1)^3 + c(x - x_1)^2 + b(x - x_1) + a \\ S_2(x) = d(x - x_2)^3 + c(x - x_2)^2 + b(x - x_2) + a \\ \vdots \\ S_{29}(x) = d(x - x_{29})^3 + c(x - x_{29})^2 + b(x - x_{29}) + a \\ S_{30}(x) = d(x - x_{30})^3 + c(x - x_{30})^2 + b(x - x_{30}) + a \end{cases}$$

Differentialrechnung

An der Detaildarstellung erkennt man sehr gut, dass aus dem Datensatz der Messpunkte abgeleitete Extremwerte (Maxima, Minima) nicht mit dem interpolierten Funktionsverlauf übereinstimmen (rote Markierung). Der Datensatz ist für diese Auswertung demzufolge nicht geeignet.



Da die Extremwerte (Minimum, Maximum) der ursprünglichen Datensätze und auch der gebildeten Funktionen bei jedem Datensatz immer an anderer Stelle liegen, also λ und β immer unterschiedlich sind, ist eine Verwendung dieser Eigenschaften zum Filtern der Datensätze nicht sinnvoll.

Vielmehr soll zum Beispiel die Funktion im Bereich $470 - 480$, und im speziellen die Ableitungen von ihr, untersucht werden (grüne Markierung).

Es gilt dabei folgende Theorie [2]:

Die Funktion $f(x)$ ist eine differenzierbare Funktion, bei der die Kurventangente an der Stelle eines Extremwertes waagrecht verläuft.

Die Bedingung $f'(x_0) = 0$ ist eine notwendige und die Bedingung $f''(x_0) \neq 0$ eine hinreichende Voraussetzung für die Existenz eines relativen Extremwertes an der Stelle x_0 . Die Lösung $f''(x_0) < 0$ stellt ein Maximum dar und die Lösung $f''(x_0) > 0$ ein Minimum.

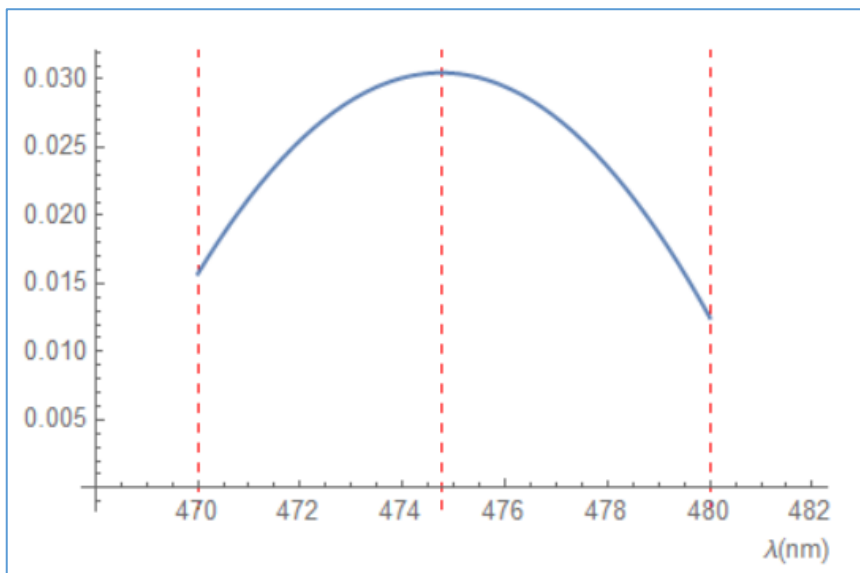
Im Bereich $470 - 480$ nm wird die folgende Funktion verwendet

$$f(x) = -2.14 \cdot 10^{-4} x^3 + 0.31 x^2 - 147.17 x + 23294.5$$

Die Funktion hat die 1. Ableitung in der Form

$$f'(x) = -6.5 \cdot 10^{-4} x^2 + 0.62 x - 147.17$$

Die Darstellung bildet eine quadratische, bei der man bereits einen Extremwert erkennt.



Die 2. Ableitung hat die Form

$$f''(x) = -1.3 \cdot 10^{-3}x + 0.62$$

Sie wird für x gelöst.

$$x_1 = 474.75$$

Die 3. Ableitung der Funktion ergibt

$$f'''(x) = -0.0013$$

Die 3. Ableitung ist negativ und damit ein Maximum an der Stelle $x_1 = 474.75$.

Für die 1. Ableitungen der Funktionsgleichungen gilt also, dass das Vorhandensein eines Extremwertes (Minimum, Maximum) oder eine bestimmte Steigung (positiv, negativ) an einer beliebigen Stelle in einem definierten Intervall als Filterkriterium der Datensätze geeignet ist.

Darstellung

Farborte werden als Punkte mit den Koordinaten $\{L, a\}$, $\{L, b\}$ oder $\{a, b\}$ in einer Listenform zusammengefasst und als *Listplot* in einem 2D Diagramm dargestellt.

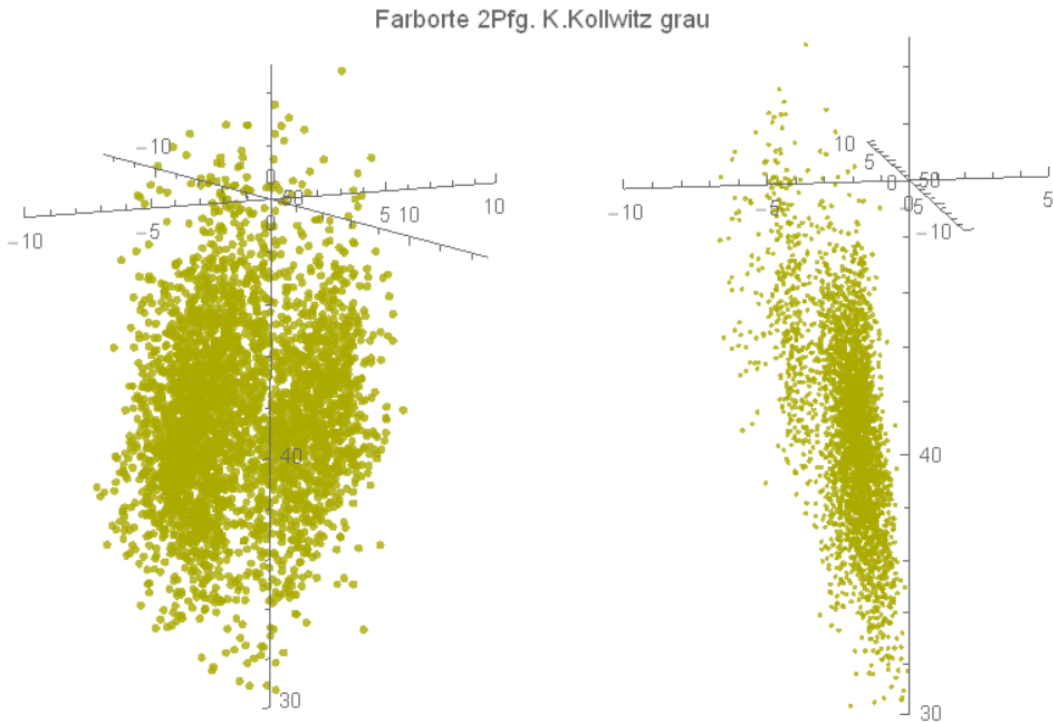
Histogramme zeigen die Häufigkeitsverteilung eines bestimmten Parameters der Farbortkoordinaten einer Gruppe. Der entsprechende Parameter wird in einer Listezusammengefasst und die Werte der Größe nach sortiert.

Die Messwerte (Remission β bei einer Wellenlänge λ) liegen in Listenform $\{\lambda, \beta\}$ in einem Datensatz vor. Die Darstellung der Messwerte erfolgt wieder als *Listplot*. Die einzelnen Punkte können mit Linien verbunden werden.

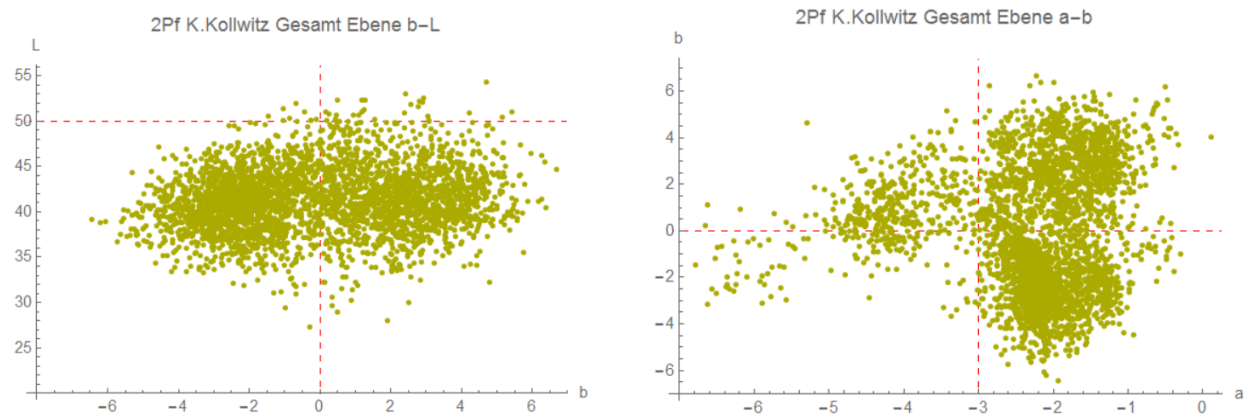
Aus den Messwerten interpolierte *Kurven* liegen als Funktionsgleichung $f(x)$ vor und werden als durchgehender Kurvenverlauf im entsprechend gültigen Intervall $\{x_0, \dots, x_n\}$ dargestellt.

Überblick über den Farbraum

Der Überblick über den Farbraum der vorhandenen Messwerte zeigt einen langgezogenen Helligkeitsbereich L , mit ausschließlich negativen a -Werten und zusätzlich eine relativ symmetrische Anordnung um den Wert $b = 0$.



Aufgrund der symmetrischen Anordnung um den Wert $b = 0$ wäre eine Trennung an dieser Stelle vorstellbar. Des Weiteren erfolgt eine deutliche Abtrennung bei $a \cong -3.2$.



Ergebnisse der Analyse

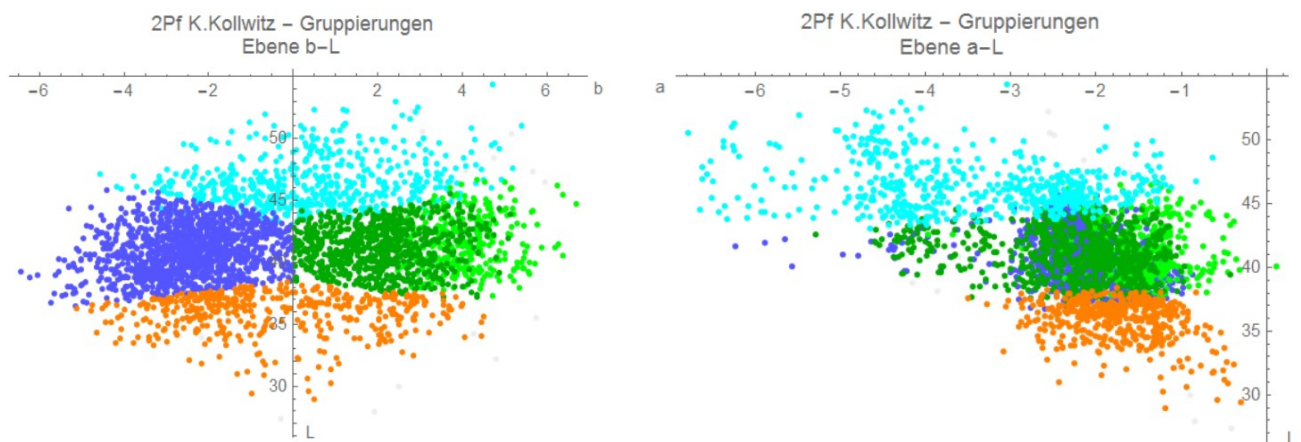
Im Ergebnis der Clusteranalyse entstehen 5 Cluster. Zusätzlich bleiben unbeteiligte Farbwerte als Einzelwerte bestehen.

Die Cluster sind:

- Cluster 2 (Farbe Blau)
Zentrum: $\{40.86, -2.16, -2.46\}$, Anzahl Punkte: 1252
Intervall $36.49 < L > 45.83, -0.87 < a > -6.23, -0.02 < b > -6.44$
- Cluster 3 (Farbe Hellgrün)
Zentrum: $\{41.68, -1.45, 4.01\}$, Anzahl Punkte: 328
Intervall $37.27 < L > 46.67, -3.98 < a > 0.11, 1.14 < b > 6.68$
- Cluster 4 (Farbe Grün)
Zentrum: $\{41.13, -2.27, 2.05\}$, Anzahl Punkte: 750
Intervall: $37.18 < L > 45.54, -0.43 < a > -5.3, 0.01 < b > 5.14$
- Cluster 5 (Farbe Cyan)
Zentrum: $\{46.46, -3.44, 0.40\}$, Anzahl Punkte: 568
Intervall: $43.08 < L > 54.37, -0.63 < a > -6.79, -4.57 < b > 5.41$
- Cluster 6 (Farbe Orange)
Zentrum: $\{35.90, -1.74, -0.44\}$, Anzahl Punkte: 454
Intervall: $29.00 < L > 38.45, -0.30 < a > -3.51, -5.13 < b > 4.51$

Es gibt zwei kleine Gruppen (11 Werte), die weit am Rand liegen, und weitere 13 Farbwerte ohne Zuordnung zu einem Cluster.

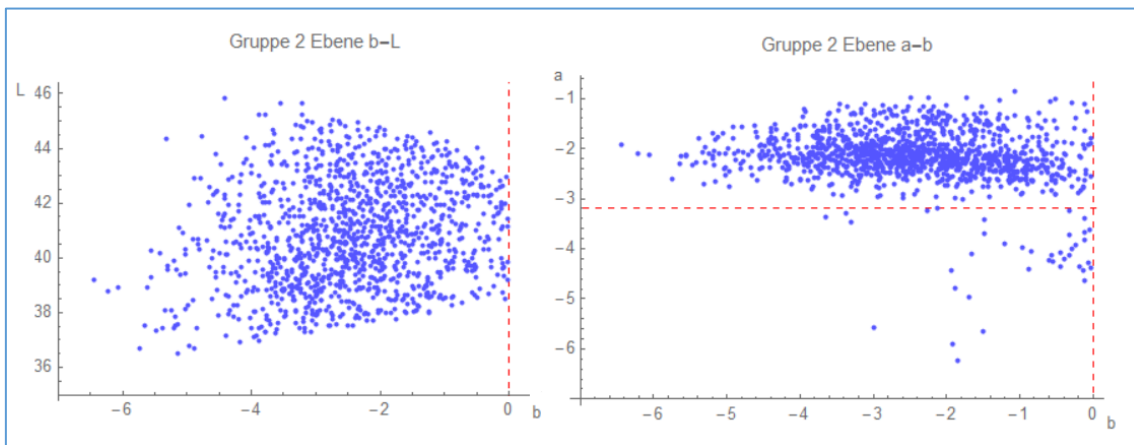
Aufgrund der räumlichen Nähe der Cluster 3 (hellgrün) und 4 (grün) können diese auch als „zusammengehörend“ betrachtet werden.



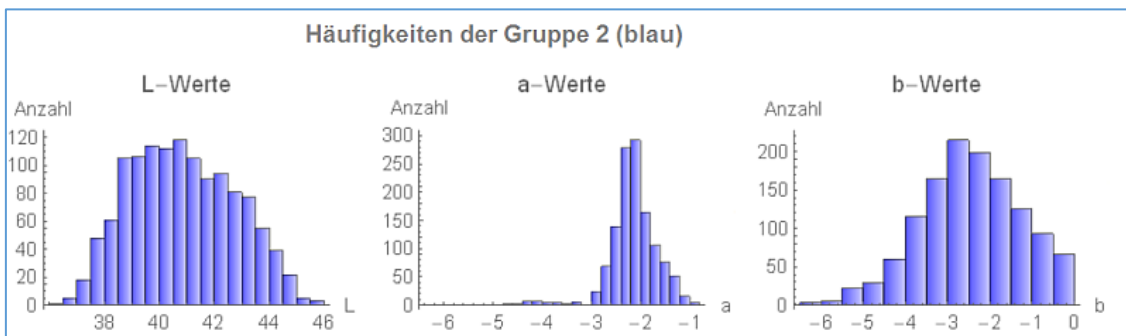
Eine Trennung bei $a \cong -3.2$ wurde mit der Analyse nicht erzielt. In den Gruppen bleibt ein Spalt bei diesem Wert sichtbar, den man überprüfen könnte. Deshalb wird die Gruppe 5 später beispielhaft auf Werte mit $a < -3.2$ korrigiert und neu bewertet.

Ergebnisse Cluster 2 (blau) $b < 0$

Im Einzelergebnis kann für die Gruppe 2 mit Zentrum $\{40.86, -2.16, -2.46\}$ die Trennungsebene $b = 0$ bestätigt werden. Die Anordnung zeigt sich langgestreckt in Richtung $-b$ und eng begrenzt zwischen $-0.5 < a < 3.5$, so dass überlegt werden sollte, Werte mit $a < -3.5$ aus der Gruppe herauszulösen (die Werte gehen später in eine korrigierte Berechnung der Gruppe 5 ein).

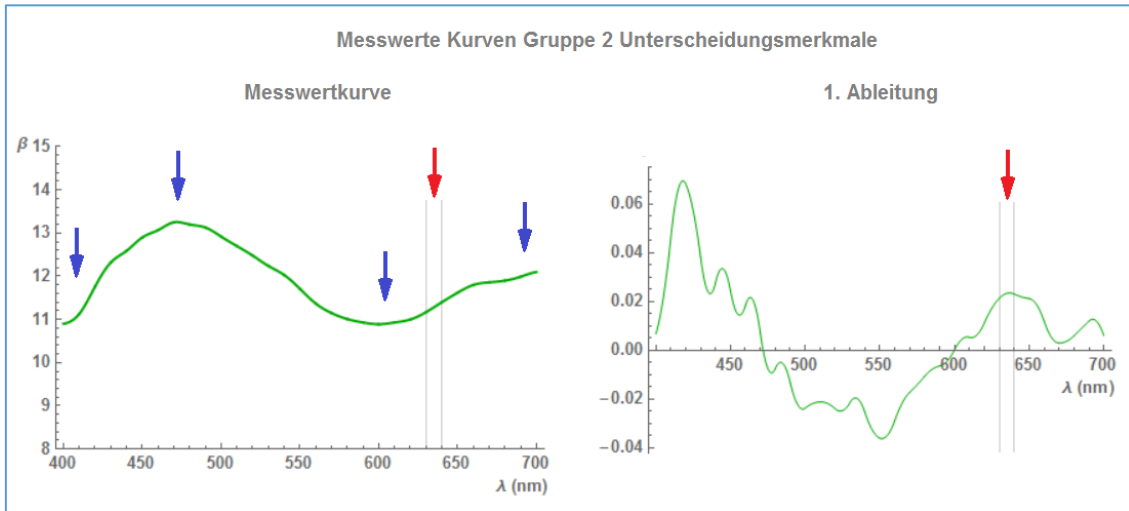


Während die Verteilung der Häufigkeiten für L und b annähernd symmetrisch ist, zeigt die Verteilung bei a eine Schiefe, die diese Überlegungen unterstützt.



Die Farbortkoordinaten werden aus zwei verschiedenen Formen von Remissionskurven gebildet, deren Funktionen mithilfe der Spline-Interpolation bestimmt wurden. Anhand bestimmter Merkmale können diese Kurvenverläufe unterschieden werden.

Form A: 1201 Messwerte



Merkmal 1: 400 – 410 nm – ansteigender Kurvenbeginn.

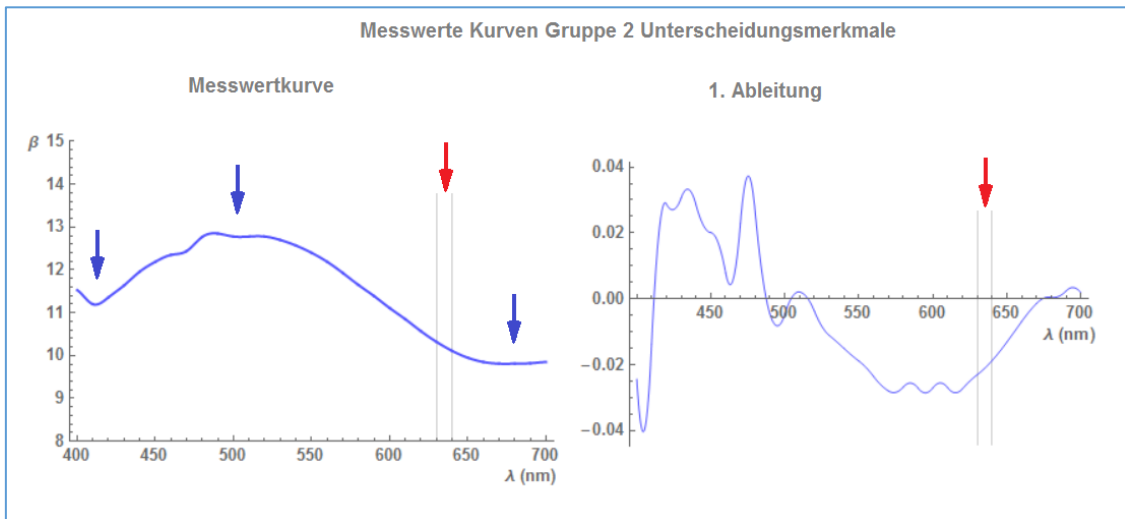
Merkmal 2: 430 – 440 nm – Maximum im Kurvenverlauf.

Merkmal 3: 580 – 620 nm – Minimum im Kurvenverlauf.

Merkmal 4: 630 – 640 nm – ansteigender Kurvenverlauf.

Merkmal 5: 670 – 700 nm – steigend auslaufender Kurvenschluss.

Form B: 51 Messwerte



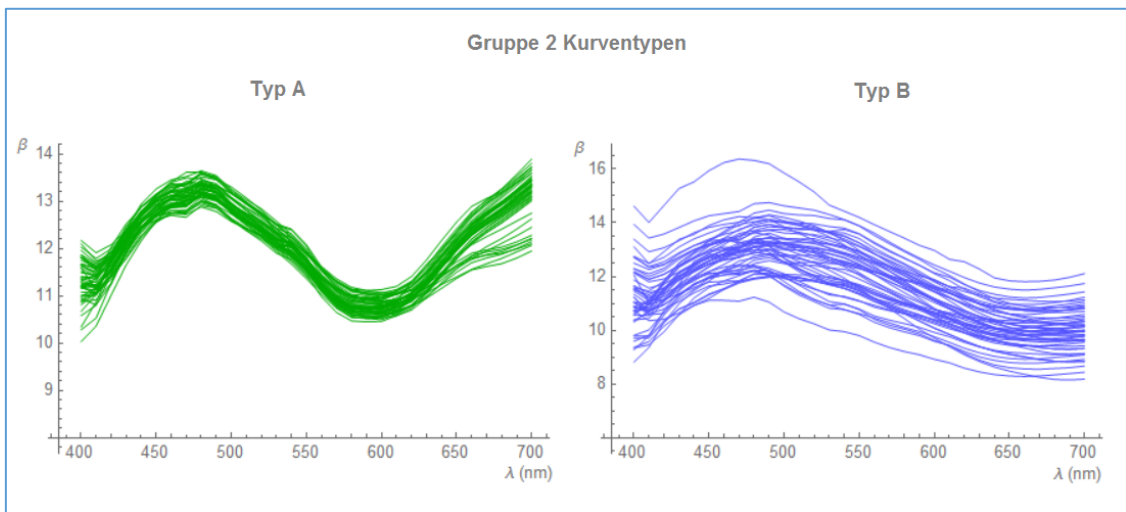
Merkmal 1: 400 – 410 nm – fallender Kurvenbeginn.

Merkmal 2: 490 – 510 nm – Maximum im Kurvenverlauf und Plateaubildung.

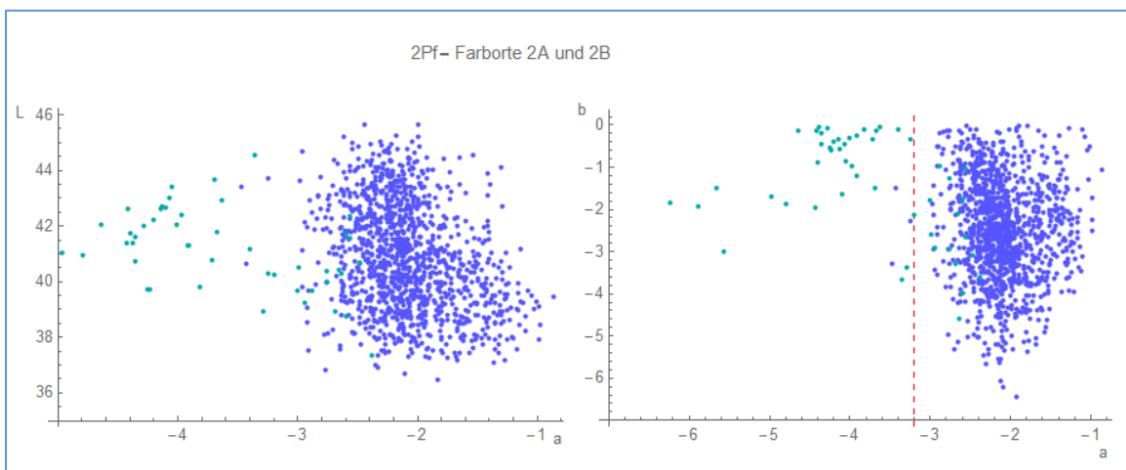
Merkmal 3: 630 – 640 nm – fallender Kurvenverlauf.

Merkmal 4: 670 – 700 nm – flach auslaufender Kurvenschluss.

Als sicheres Unterscheidungsmerkmal der Typen und Filter in den Datensätzen wird jeweils die *positive / negative Steigung der 1. Ableitung* der Kurvenfunktion im Bereich 630 – 640 nm verwendet. Das Auswahlkriterium wird am Beispiel von je 20 Datensätzen bestätigt.



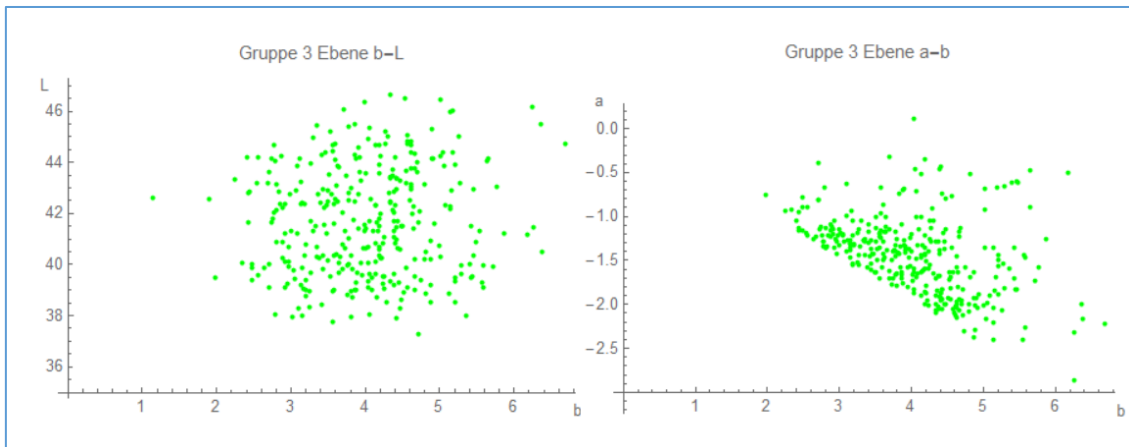
Die Unterscheidung der Farbortkoordinaten nach den Kurventypen zeigt auf einen Unterschied im a- Wert. Farborte mit der Kurvenform B befinden sich hauptsächlich im Bereich $a < -3.5$.



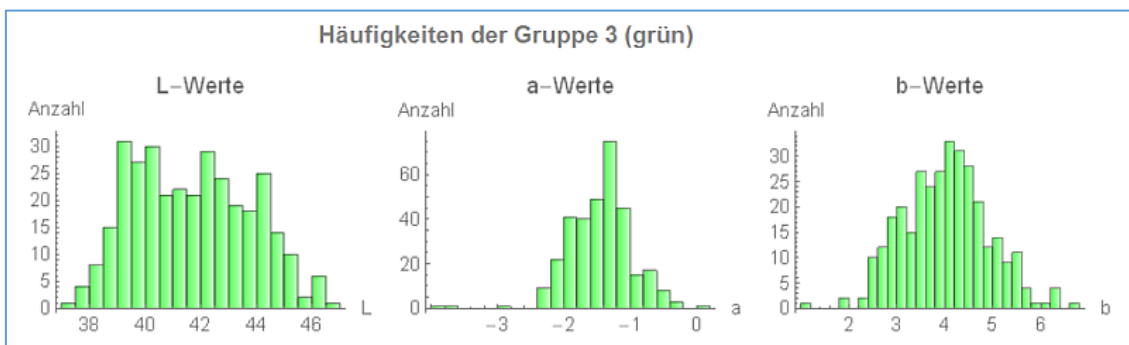
In der Wahrnehmung haben die Farben dieser Marken aufgrund des Bereiches $b < 0$ eine Tendenz zum blau. Die Kurvenverläufe bestätigen diesen Trend mit hohen Remissionswerten im Bereich um 450 nm.

Ergebnisse Cluster 3 (hellgrün)

Die Darstellung der Farborte der Gruppe 3 mit Zentrum $\{41.68, -1.45, 4.01\}$ zeigt im Einzelergebnis keine Auffälligkeiten.

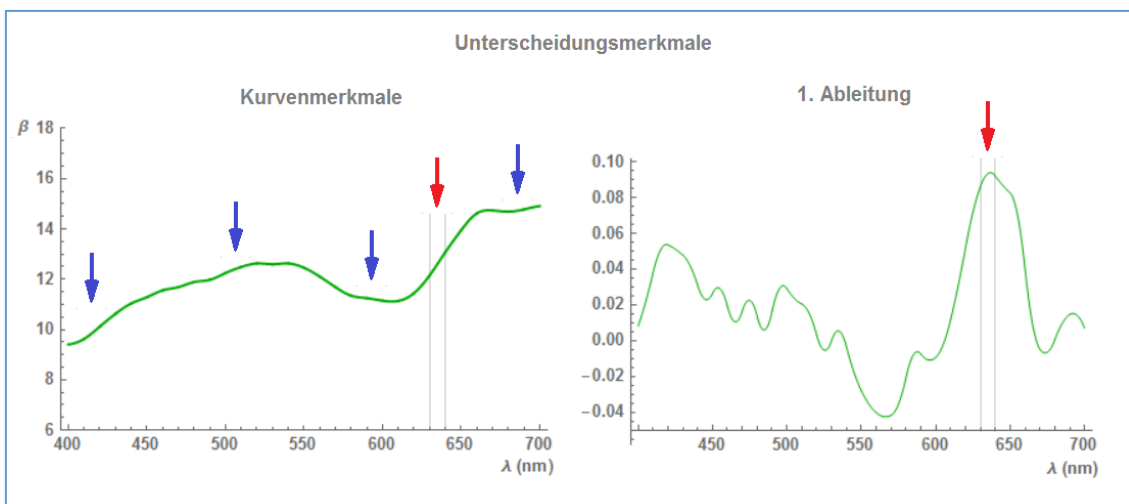


Die Häufigkeiten für L zeigen eine größere Anzahl an Farborten in $L = 39$ und $L = 42$, eine schiefe Verteilung in a und eine symmetrische Verteilung in b .



Auch in dieser Gruppe bilden sich die Farborte aus zwei unterschiedlichen Formen der Kurvenverläufe der Messwerte.

Form A: 322 Messwerte



Merkmal 1: 400 – 410 nm – ansteigender Kurvenbeginn.

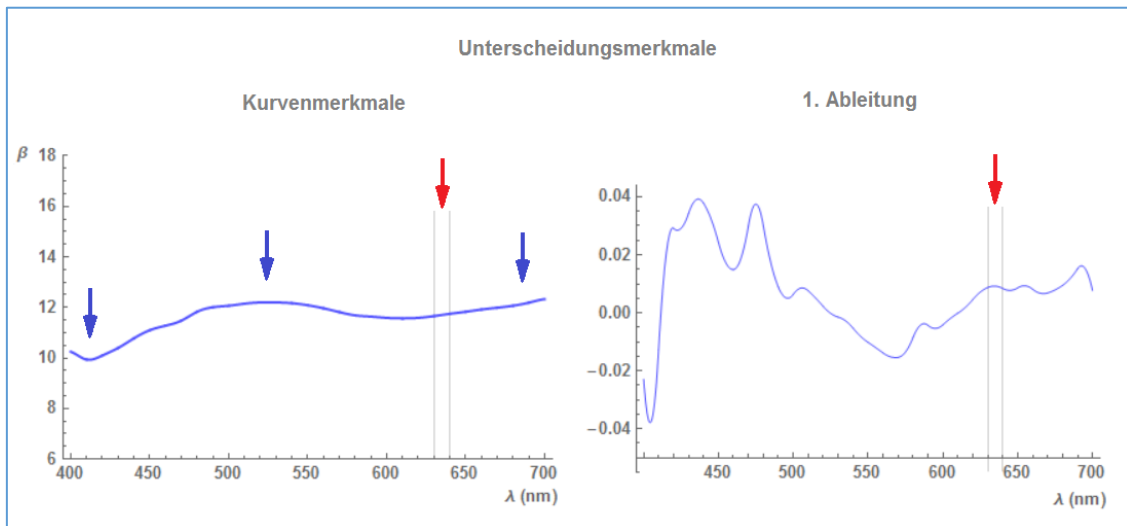
Merkmal 2: 520 – 530 nm – Maximum im Kurvenverlauf.

Merkmal 3: 580 – 620 nm – Minimum im Kurvenverlauf.

Merkmal 4: 630 – 640 nm – stark ansteigender Kurvenverlauf.

Merkmal 5: 670 – 700 nm – auf hohem Niveau flach auslaufender Kurvenschluss.

Form B: 6 Messwerte



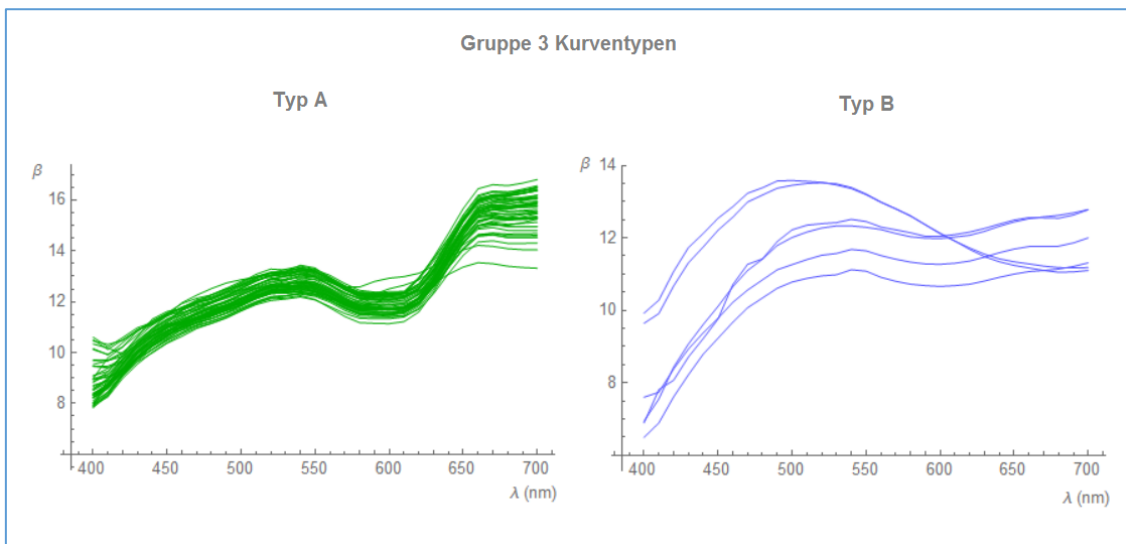
Merkmal 1: 400 – 410 nm – fallender Kurvenbeginn.

Merkmal 2: 510 – 550 nm – Maximum im Kurvenverlauf und Plateaubildung.

Merkmal 3: 630 – 640 nm – flach steigender Kurvenverlauf.

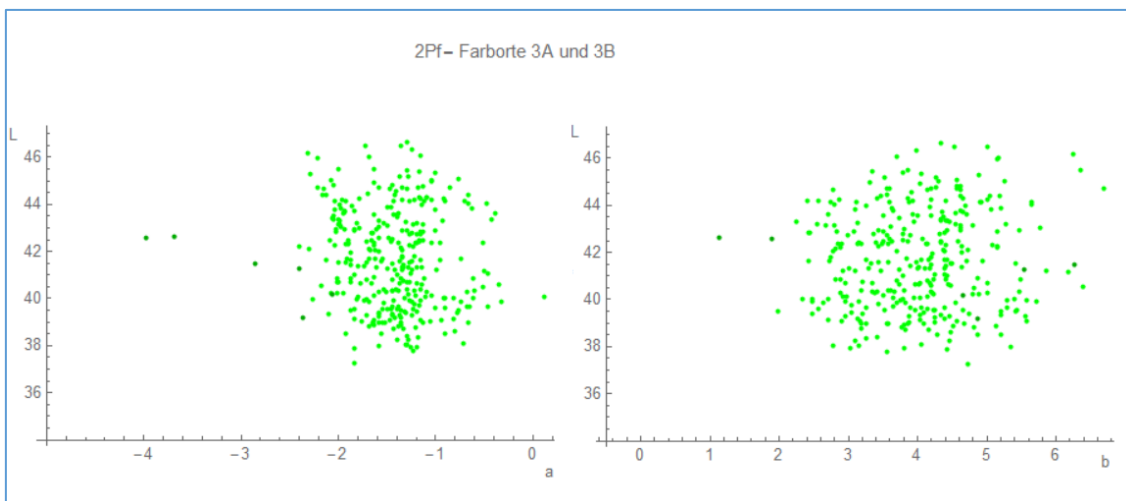
Merkmal 4: 670 – 700 nm – flach auslaufender Kurvenschluss.

Als sicheres Unterscheidungsmerkmal der Typen und Filter in den Datensätzen wird jeweils die *positive / negative Steigung der 1. Ableitung* der Kurvenfunktion im Bereich 630 – 640 nm verwendet.



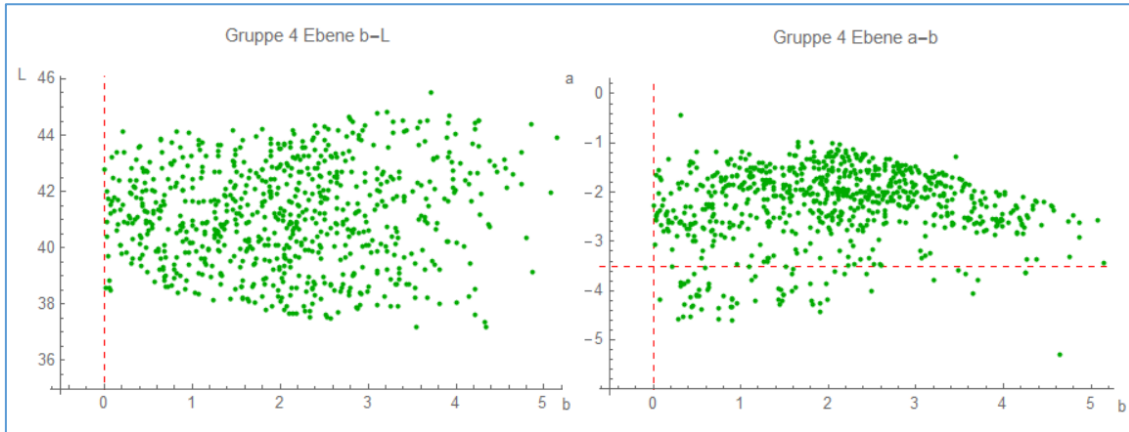
Das Auswahlkriterium wird am Beispiel von 20 Datensätzen für die Form A bestätigt, wobei für die Form B nur 6 Datensätze vorhanden sind, sich diese wegen ihrer besonderen Form noch einmal unterteilen lassen, bzw. die abweichenden beiden Kurven mit flach fallendem Kurvenschluss eher der Gruppe 5 mit dem Kriterium $a < -3.2$ zuzuordnen sind.

Die Darstellung der Farborte der Formen bestätigt die Einteilung.

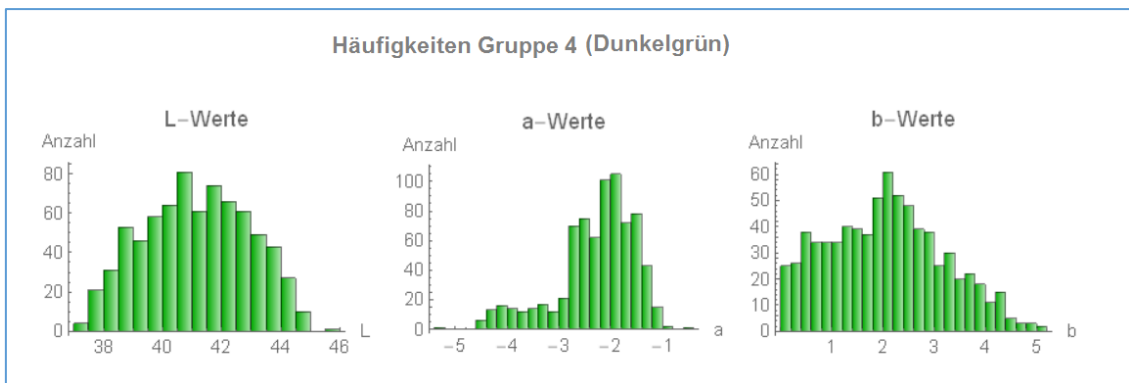


Ergebnisse Cluster 4 (grün)

Im Einzelergebnis kann die Trennungsebene $b = 0$ für die Gruppe 4 mit Zentrum $\{41.13, -2.27, 2.05\}$ wieder bestätigt werden. Die Anordnung zeigt sich langgestreckt in Richtung $-b$ und der Bereich $a < -3.2$ wird wieder angedeutet.

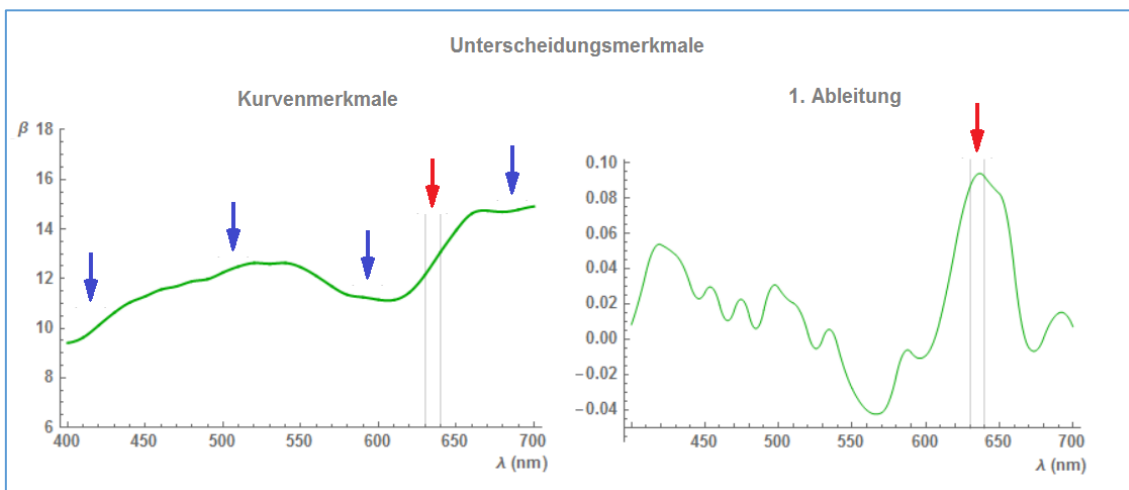


Die Verteilungen der Werte L und b können wieder als nahezu symmetrisch angesehen werden, während a die schon bekannte Schiefe aufweist.



Die Farborte dieser Gruppe bilden sich wieder aus zwei unterschiedlichen Formen der Kurvenverläufe der Messwerte heraus.

Form A: 388 Messwerte



Merkmal 1: 400 – 410 nm – ansteigender Kurvenbeginn.

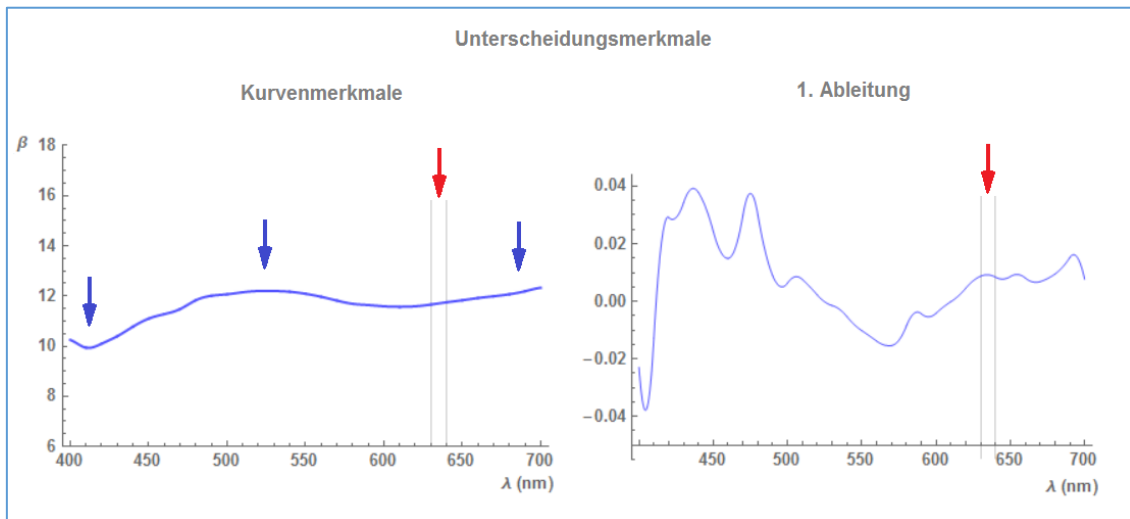
Merkmal 2: 500 – 550 nm – Maximum im Kurvenverlauf.

Merkmal 3: 580 – 620 nm – Minimum im Kurvenverlauf.

Merkmal 4: 630 – 640 nm – stark ansteigender Kurvenverlauf.

Merkmal 5: 670 – 700 nm – auf hohem Niveau flach auslaufender Kurvenschluss.

Form B: 362 Messwerte



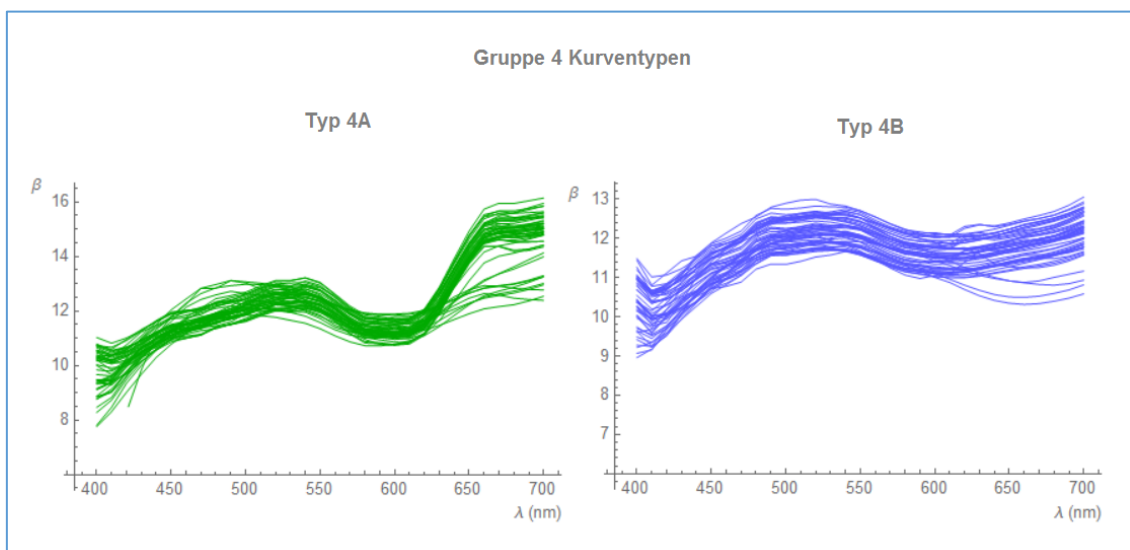
Merkmal 1: 400 – 410 nm – fallender Kurvenbeginn.

Merkmal 2: 510 – 550 nm – Plateaubildung im Kurvenverlauf.

Merkmal 3: 630 – 640 nm – flach steigender Kurvenverlauf.

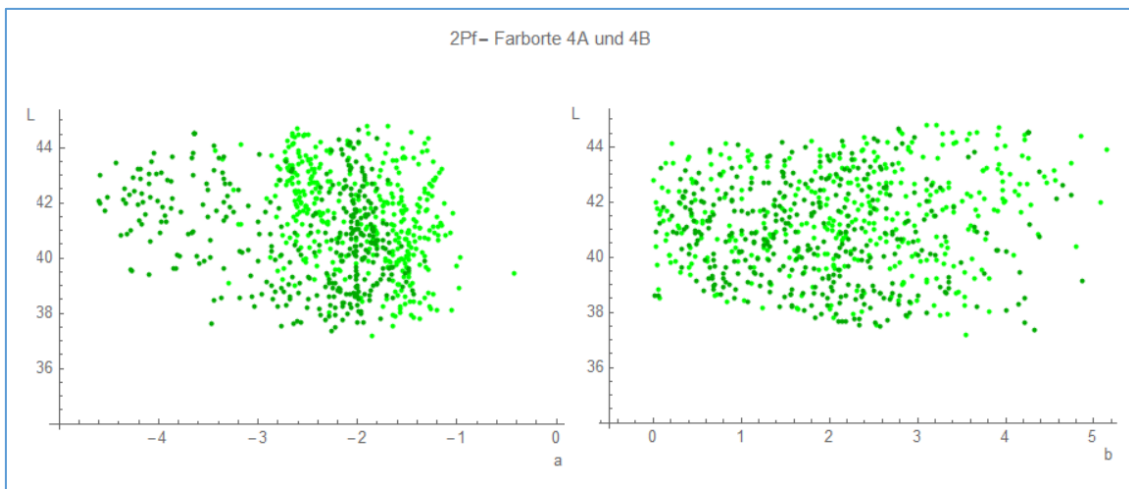
Merkmal 4: 670 – 700 nm – flach ansteigender Kurvenschluss.

Mit dem Merkmal 4 als Filter können zwei Formen von Kurven unterschieden werden.



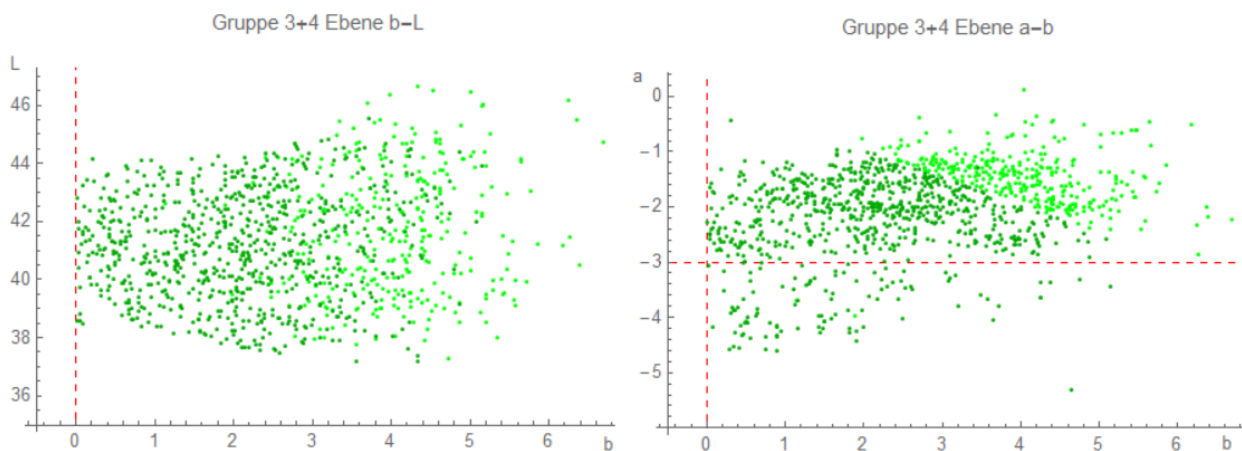
Das Auswahlkriterium *positive / negative Steigung der 1. Ableitung der Kurvenfunktion* wird am Beispiel von je 20 Datensätzen bestätigt.

Auf die Darstellung der Farborte haben die unterschiedlichen Kurvenformen keinen Einfluss, es bilden sich keine eigenständigen Bereiche heraus.



Ergebnisse Cluster 3+4 (zusammengefasst)

Es ist sinnvoll, die Gruppen 3 + 4 zu einer Gruppe zusammenzufassen.

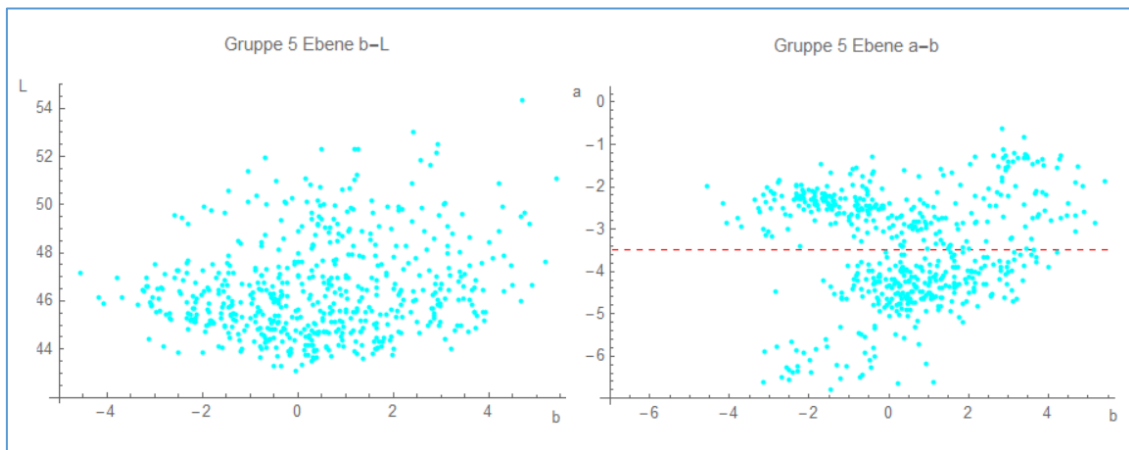


Kurvenformen 3A und 4A identisch. Ebenso 3B und 4B unter Beachtung der besonderen Form bei $a < -3.2$. Diese Werte werden in einer korrigierten Berechnung der Gruppe 5 aus dem Cluster 3 + 4 herausgelöst.

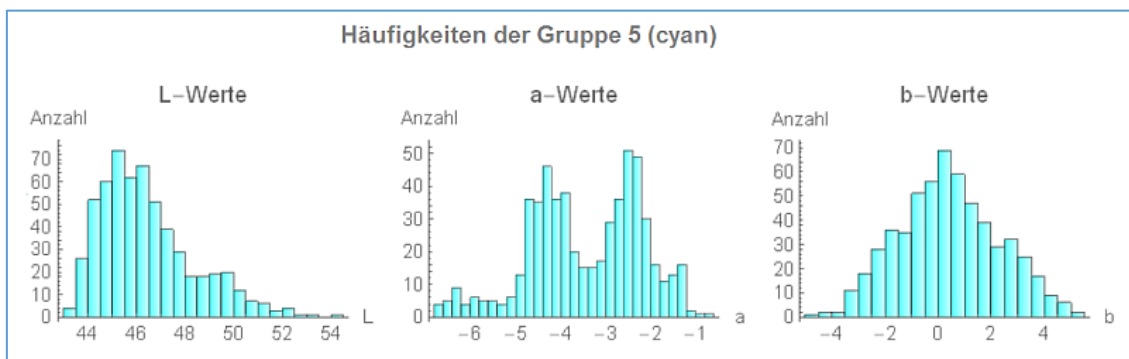
In der Wahrnehmung haben die Farben dieser Marken aufgrund des Bereiches $b > 0$ eine Tendenz zum gelb. In den Messwertekurven wird das bestätigt, da hier die Remissionswerte für die Wellenlängen 500 nm und 700 nm sehr hoch sind. Es kommt zu einer additiven Farbmischung von grün und rot.

Ergebnisse Cluster 5 (cyan)

Im Einzelergebnis ist die Gruppe 5 der Analyse mit dem Zentrum $\{46.46, -3.44, 0.40\}$ eine sehr interessante Gruppe. Im Bereich $L - b$ zeigen sich noch keine Auffälligkeiten. Auch die bisherige Trennungsebene $b = 0$ ist nicht ersichtlich. Mit Blick auf die Ebene $a - b$ werden jedoch zwei unterschiedlich stark verdichtete Bereiche deutlich, die sich, wie weiter oben schon beschrieben, durch $a < -3.2$ trennen lassen.



Die Verteilung für L ist ausgeprägt schief, aber plausibel, da die Werte mit ansteigender Helligkeit immer weniger werden. Die Verteilung für b zeigt eine sehr deutliche Symmetrie um $b = 0$. Zwei separate Häufungen zeigt aber die Verteilung von a an der Stelle $a < -3.2$ und unterstützt den Gedanken der Trennung an dieser Stelle.



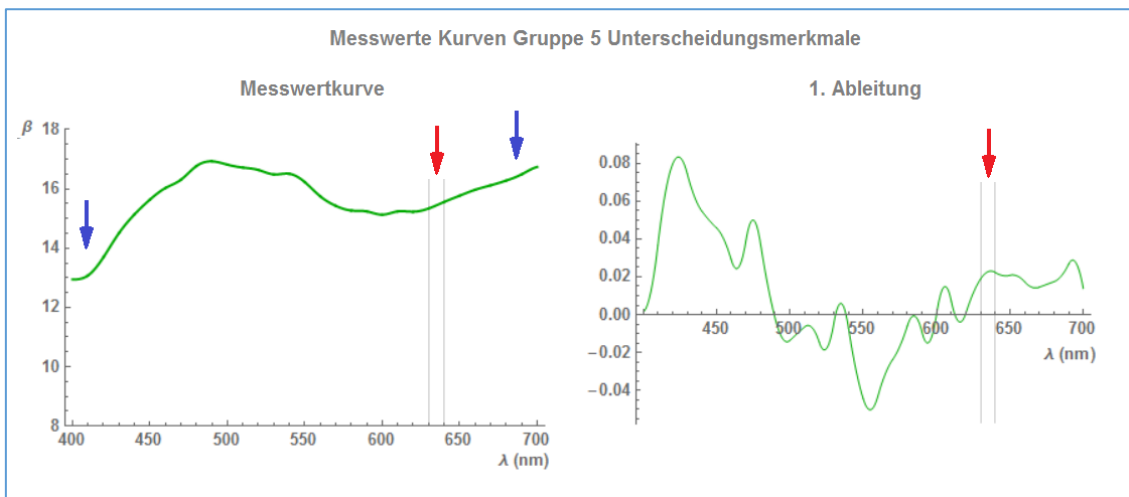
Form A: 267 Messwerte Form B: 301

Merkmal 1: 400 – 410 nm – ansteigender Kurvenbeginn.

Merkmal 2: 480 – 500 nm – Maximum im Kurvenverlauf.

Merkmal 3: 630 – 640 nm – ansteigender Kurvenverlauf.

Merkmal 4: 650 – 700 nm – ansteigender Kurvenauslauf.



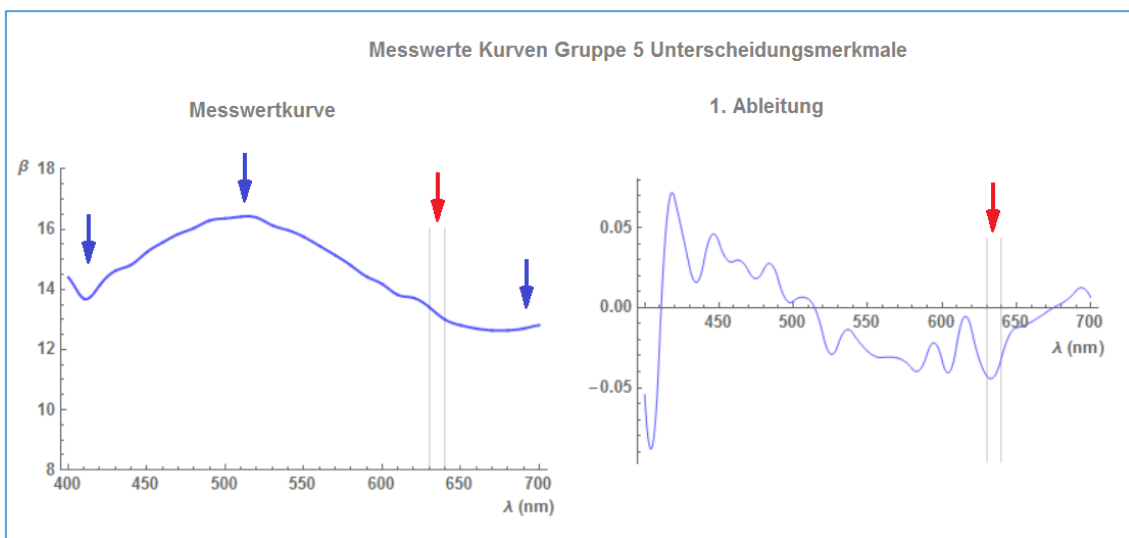
Form B: 301 Messwerte

Merkmal 1: 400 – 410 nm – abfallender Kurvenbeginn.

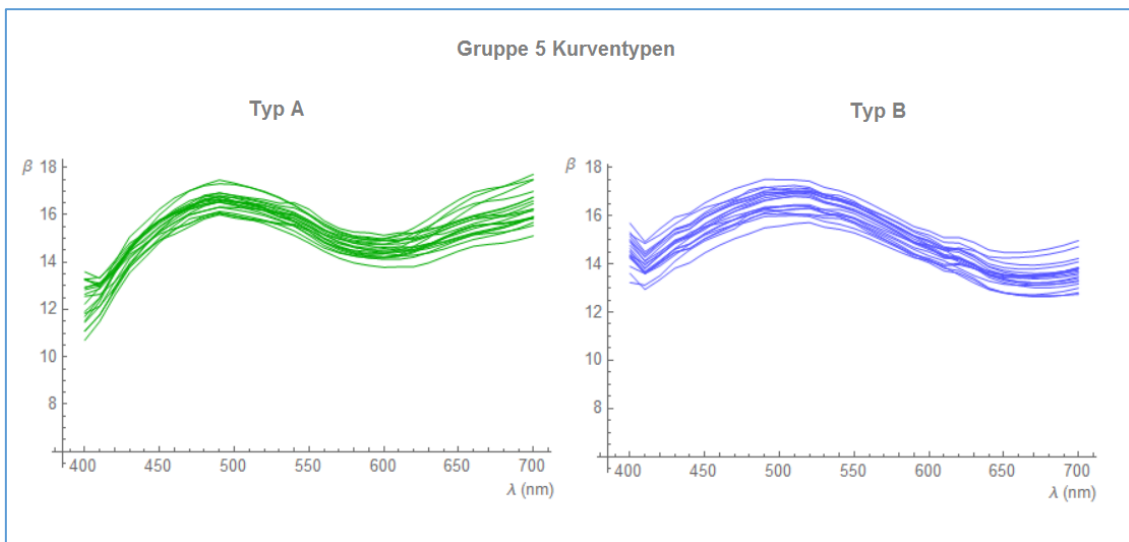
Merkmal 2: 490 – 520 nm – Maximum im Kurvenverlauf.

Merkmal 3: 630 – 640 nm – fallender Kurvenverlauf.

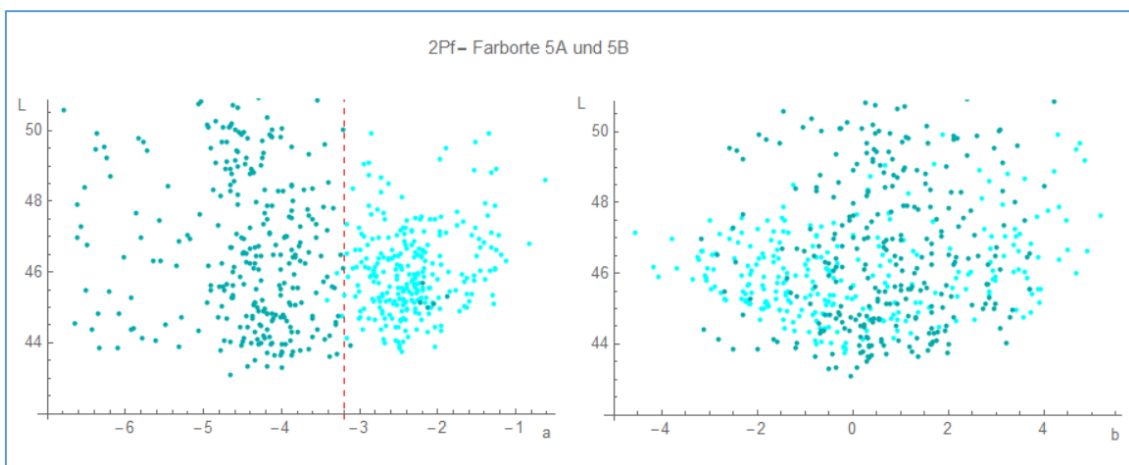
Merkmal 4: 650 – 700 nm – mit einem Minimum und sanft ansteigenden Kurvenauslauf.



Als Unterscheidungsmerkmal der Typen und Filter in den Datensätzen wird wieder jeweils die *positive / negative Steigung der 1. Ableitung* der Kurvenfunktion im Bereich 630 – 640 nm verwendet. Am Beispiel von je 20 Messwerten können die Kurvenformen in ihrem Verlauf auch bestätigt werden.



Während sich in der Darstellung der Farborte in der Ansicht $L - b$ keine Besonderheiten zeigen, wird die Trennung in der Ansicht $L - a$ wieder sehr deutlich bei $a < -3.2$ erkennbar.



korrigiertes Cluster 5 (cyan)

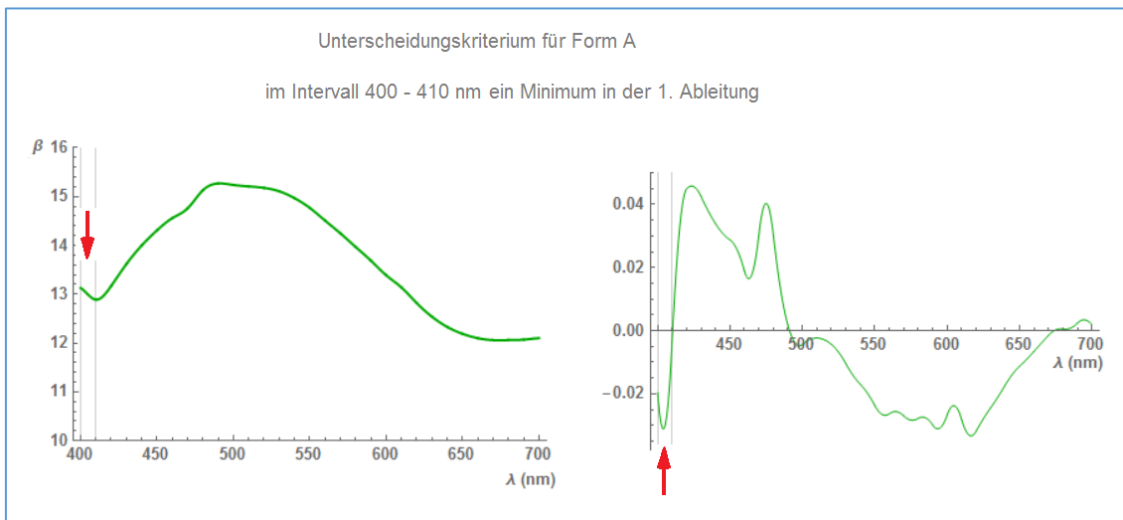
Wird die Überlegung zur Gruppierung der Farborte für die Gruppe 5 auf Werte $a < -3.2$ umgesetzt, ergibt sich ein sehr klares Bild der Farborte in Übereinstimmung mit den Kurven der Messwerte.

Es entstehen wieder zwei Formen an Messwertekurven, die in ihrem Kurvenverlauf sehr ähnlich sind und sich wesentlich in einem Kriterium unterscheiden: dem Kurvenverlauf im Bereich $400 - 410 \text{ nm}$.

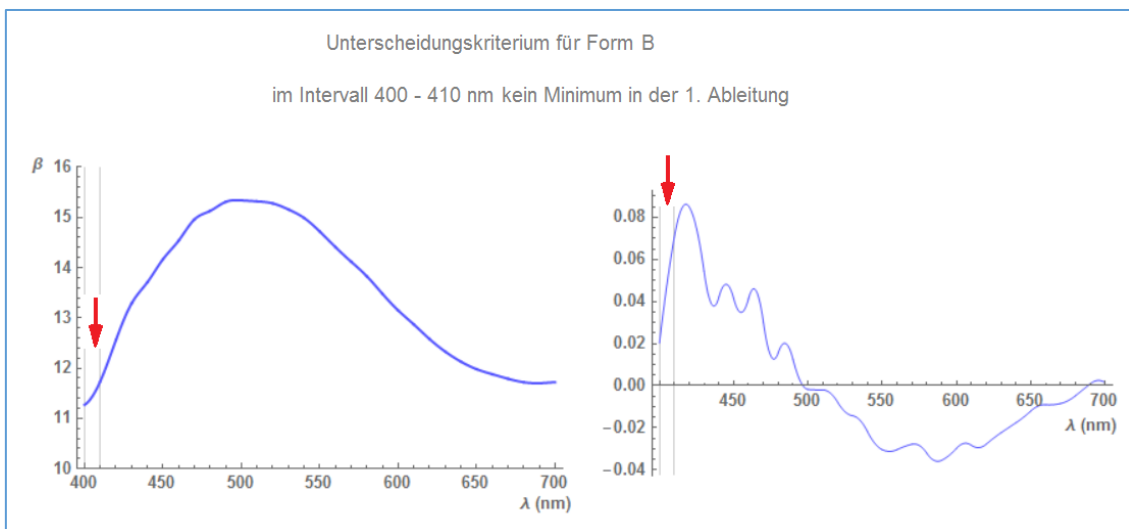
Form A hat in diesem Bereich eine fallende Kurve und bildet in der 1. Ableitung ein Minimum aus, während die Form B in diesem Bereich ansteigt und kein Minimum bildet.

Im gesamten Verlauf der Messwertekurve ist die Form A mit $12.5 < \beta < 15.5$ zudem deutlich flacher gestaltet, als Form B mit $11 < \beta < 16$.

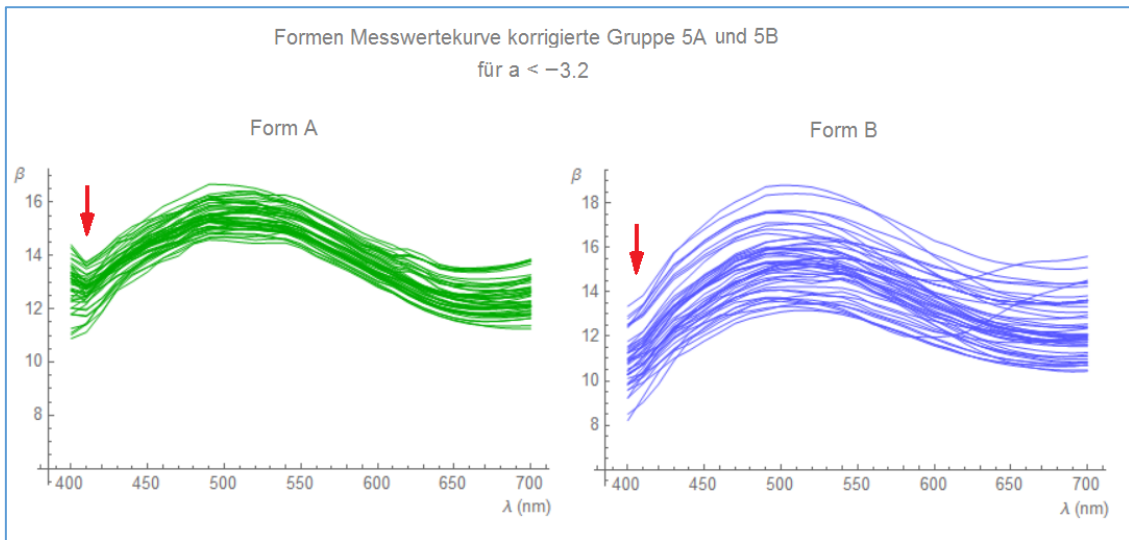
Form A: 362 Messwerte



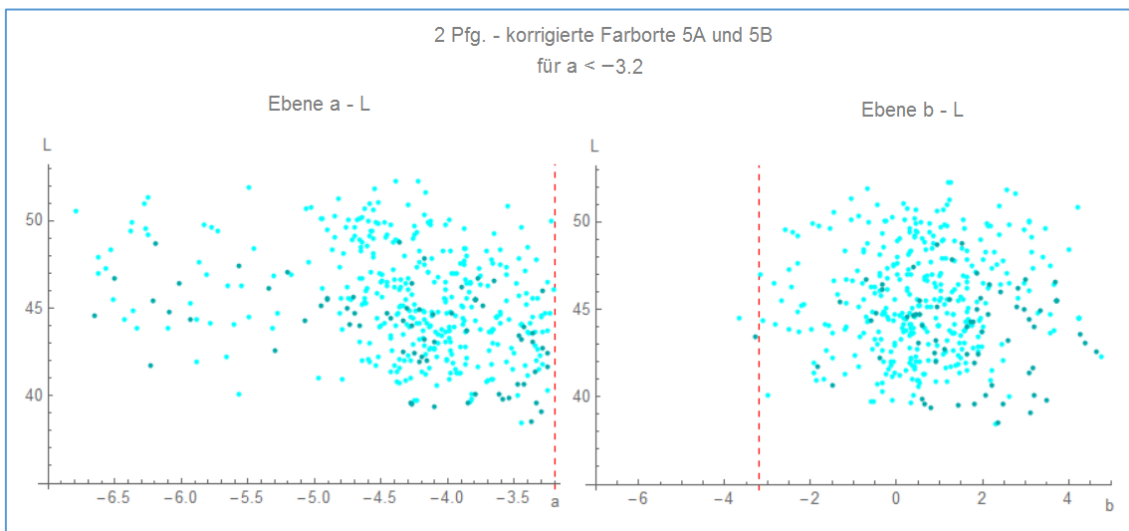
Form B: 71 Messwerte



Das Kriterium „Minimum in der 1. Ableitung“ wird wieder als Filter für die Datensätze verwendet und findet seine Bestätigung im Beispielplot mit je 50 Messwertkurven.



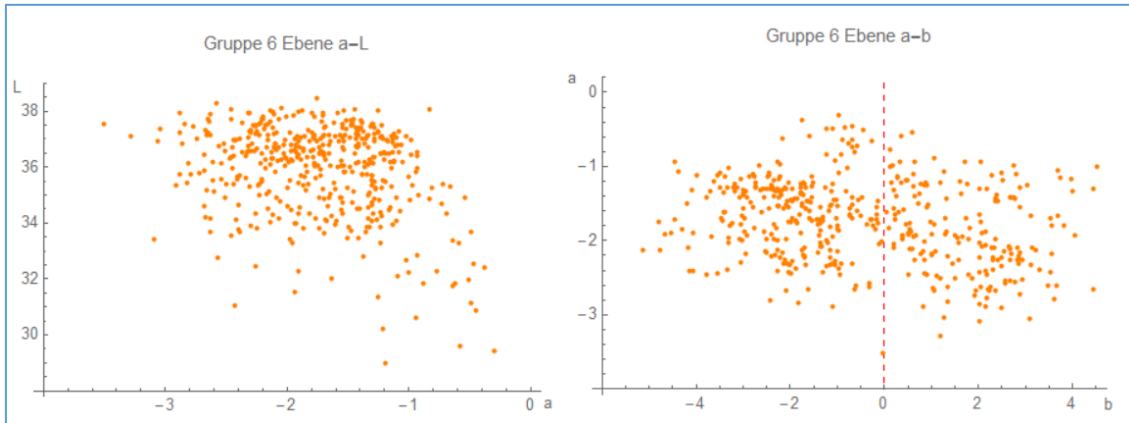
In der Darstellung der Farborte wird die Unterscheidung der Messwertkurven nicht getrennt.



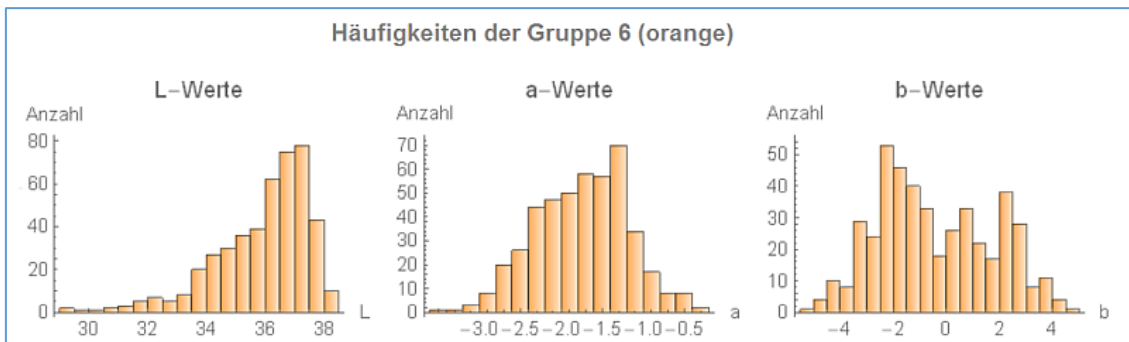
In der Wahrnehmung haben die Farben dieser Marken aufgrund des Bereiches $a < -3.2$ eine Tendenz zum grün, was auch in den Messwertkurven mit hohen Remissionswerten im Bereich 520 – 550 nm bestätigt wird.

Ergebnisse Cluster 6 (orange)

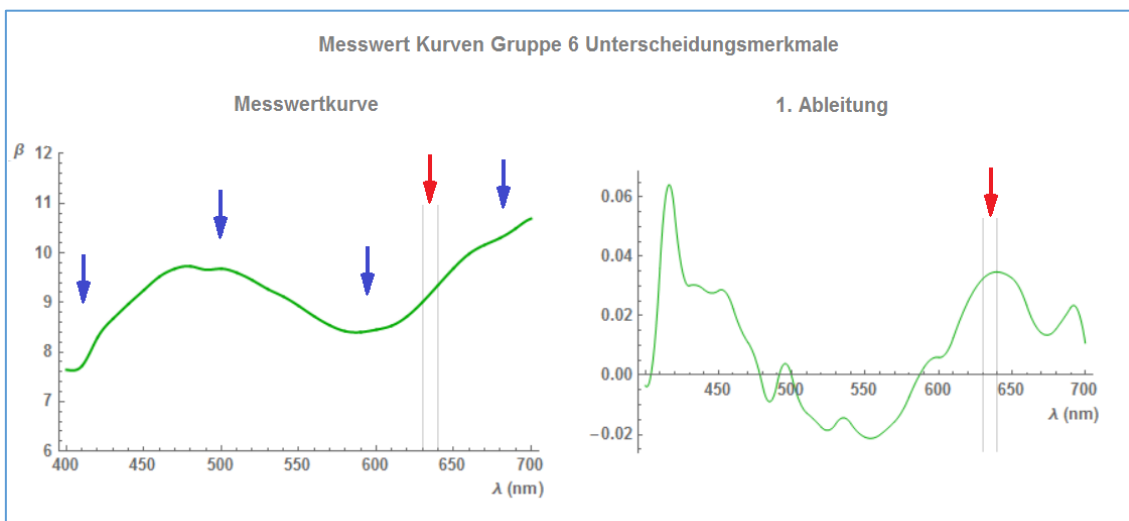
Die Gruppe 6 der Analyse mit Zentrum $\{35.90, -1.74, -0.44\}$ zeigt in der Farbortdarstellung eine Besonderheiten. Interessehalber wurde die Ebene $b = 0$ in der Darstellung markiert.



Bei den L – Werten wird wieder eine schiefe Verteilung erkennbar, was aber plausibel ist, da die Werte abnehmen, je dunkler sie werden. Die Verteilung bei a ist annähernd symmetrisch. Lediglich bei b deutet sich eine Unterscheidung am Wert $b = 0$ an.



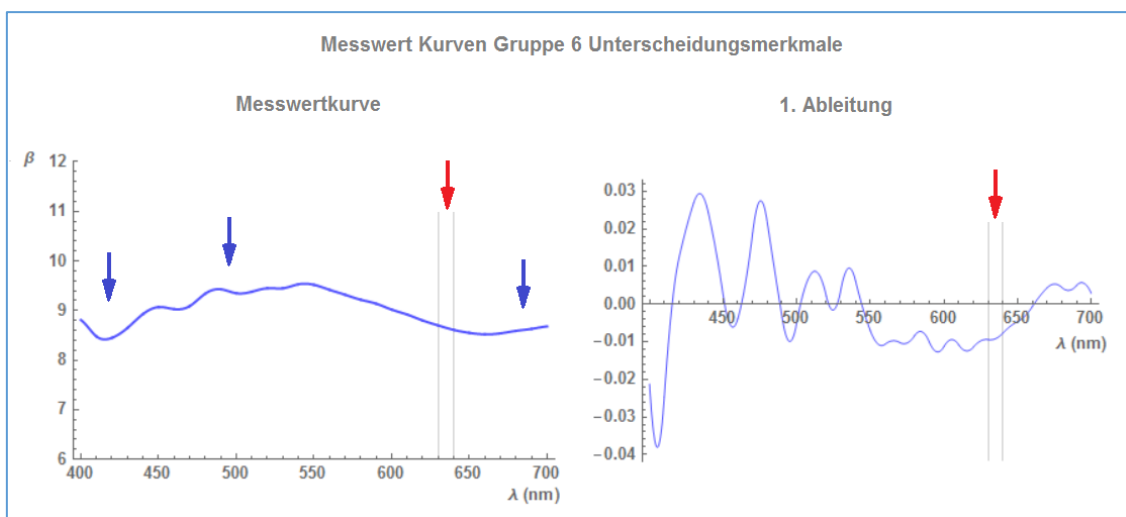
Form A: 374 Messwerte



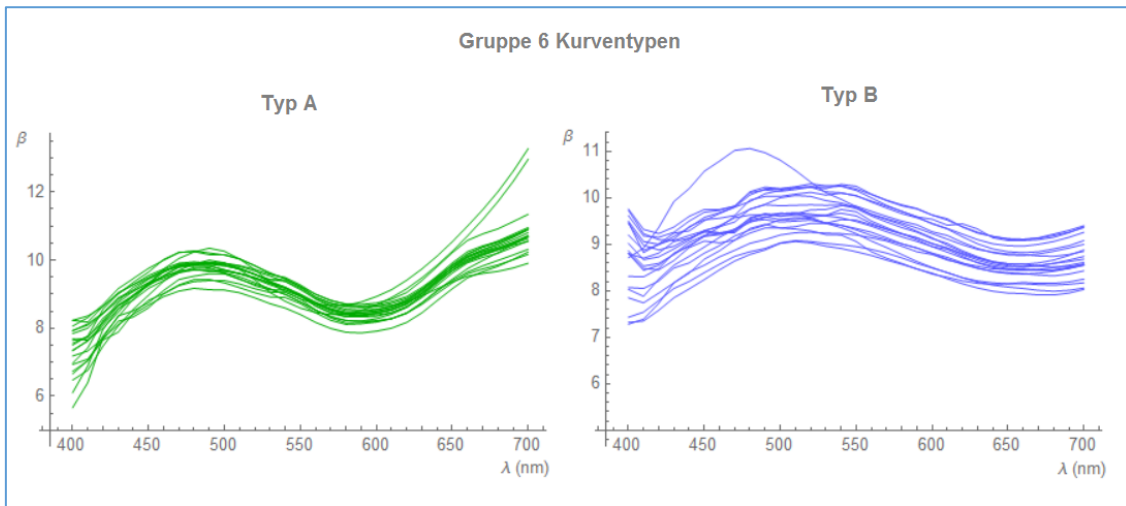
- Merkmal 1: 400 – 410 nm – ansteigender Kurvenbeginn.
- Merkmal 2: 480 – 510 nm – Maximum im Kurvenverlauf.
- Merkmal 3: 580 – 610 nm – Minimum im Kurvenverlauf.
- Merkmal 4: 630 – 640 nm – ansteigender Kurvenverlauf.
- Merkmal 5: 650 – 700 nm – stark ansteigender Kurvenauslauf.

Form B: 80 Messwerte

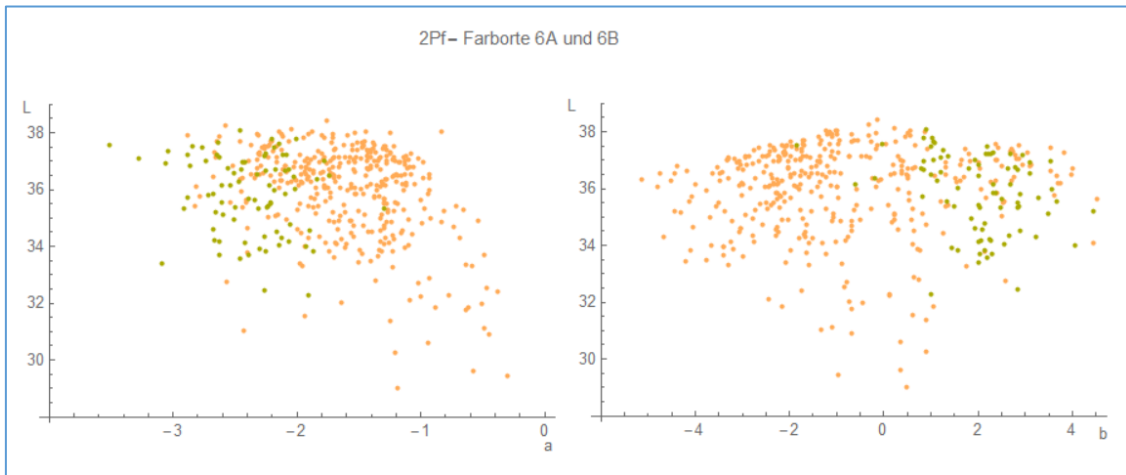
- Merkmal 1: 400 – 410 nm – abfallender Kurvenbeginn.
- Merkmal 2: 480 – 520 nm – mehrere kleinere Maxima im Kurvenverlauf.
- Merkmal 3: 630 – 640 nm – fallender Kurvenverlauf.
- Merkmal 4: 670 – 700 nm – ansteigender Kurvenauslauf.



Als Unterscheidungsmerkmal der Typen und Filter in den Datensätzen wird wieder jeweils die *positive / negative Steigung der 1. Ableitung* der Kurvenfunktion im Bereich 630 – 640 nm verwendet. Am Beispiel von je 20 Messwerten können die Kurvenformen wieder bestätigt werden. Bei der Form B erkennt man mit dem Vorkommen einer anderen Kurvenform eine ähnliche Besonderheit, wie in Gruppe 3.



Die unterschiedlichen Kurvenformen widerspiegeln sich auch in der Darstellung der Farborte.



Fazit

Eine Clusteranalyse mit den Farbwerten L^* , a^* , b^* hatte zum Ziel, mögliche Gruppierungen, die sich aufgrund bestimmter Eigenschaften ergeben, aufzuzeigen. Im Ergebnis der Analyse wurden 5 Gruppierungen gefunden. Die Gruppen 3 + 4 wurden aufgrund ihrer Nähe im Farbraum zusammengeschlossen und die Gruppe 5 mit der Eigenschaft $a < -3.2$ neu bewertet.

Die Farben der Marken können daher in 4 Gruppen unterschieden werden:

- Farbe a – mit $b < 0$ zum Blau tendierend.
- Farbe b – mit $b > 0$ zum Gelb tendierend.
- Farbe c – mit $a < -3.2$ zum Grün tendierend.
- Farbe d – mit L zwischen 29 und 38 sehr dunkel

Alle 4 Farbgruppen zeigen wesentlich zwei unterscheidbare Kurvenverläufe, die jeweils in der Form A und B beschrieben wurden. Es gibt mehrere Produktionszeiträume, die sich mit den unterschiedlichen Kurvenformen nachweisen lassen.

Da die Messungen unter Normlicht D65 durchgeführt wurden, deutet dieser Sachverhalt der zwei Kurvenformen auf *bedingt gleiche Farben* hin. Praktisch bedeutet das, dass sich die Farbgruppen mit einer anderen Normlichtart (z.B. A oder F) noch einmal trennen lassen, da sie dort anders aussehen. Eine Überprüfung wird deshalb empfohlen.

Literatur

- [1] J. Dankert, Informatik 2
- [2] L. Papula, Mathematik
- [3] J. Bacher, e.a., Clusteranalyse, Anwendungsorientierte Einführung in Klassifikationsverfahren, Oldenbourg

- [4] A. Eichler, Farbabstände in der industriellen Farbmessung, Sun Chemical
- [5] H.G. Völz, Industrielle Farbprüfung, Grundlagen und Methoden, Wiley-VCH

- [6] DIN 5033 Teile 1-9, Farbmessung
- [7] DIN 6174, Farbmetrische Bestimmung von Körperfarben nach der CIELAB Formel