

Katja Gutmair, Stella Akouete, Noah Hurmer und Anne Gritto

Weather Frog

- Abschlusspräsentation am 01. März 2021
- Institut: Statistik
- Veranstaltung: Statistisches Praktikum
- Projektpartner: M.Sc. Maximilian Weigert und
M.Sc. Magdalena Mittermeier
- Betreuer: Prof. Dr. Helmut Küchenhoff





Gliederung

1. Einführung

- i. Vorstellen des Projekts
- ii. Datensätze

2. Methodik

- i. Preprocessing
- ii. Wahl des Clusterverfahrens
- iii. Bewertungskriterien für Cluster

3. Deskriptive Analyse

- i. Verteilung über die Zeit
- ii. Unterschiede und Ähnlichkeiten in den Clustern
- iii. Vergleich zur gegebenen GWL-Einteilung

4. Ausblick

5. Fazit



1. Einführung

i. Vorstellen des Projekts



Vorstellen des Projekts

- Übergeordnete Fragestellung:
Wie verändert sich das Auftreten verschiedener Großwetterlagen (GWL) unter dem Einfluss des Klimawandels?
- Unsere Fragestellung:
Lassen sich Tage anhand von ihren Wettermesswerten sinnvoll clustern?
Wie unterscheiden sich die entstandenen Cluster voneinander?



Vorstellen des Projekts

Definition Großwetterlage

- Atmosphärischer Wetterzustand
- Definiert über ganz Europa
- Dauer: ≥ 3 Tage
- Kategorisierung nach dem Katalog von Hess und Brezowsky
- 29 GWL nach Hess und Brezowsky



Großwetterlagen Beispiele

	Abkürzung	Großwetterlage
1	WA	Westlage, antizyklonal
2	WZ	Westlage, zyklonal
3	WS	Südliche Westlage
4	WW	Winkelförmige Westlage
5	SWA	Südwestlage, antizyklonal
6	SWZ	Südwestlage, zyklonal
...		
29	TRW	Trog Westeuropa
	U	Übergang/Unbestimmt



Ziele des Projekts

Clustereinteilung der Tage anhand beobachteter Wetterdaten

- Anzahl Cluster < Anzahl GWLs
- Berücksichtigung der räumlichen Datenstruktur
- Tage als Beobachtungseinheit
- Ohne Vorinformation der herrschenden GWL

→ Mit welcher Methode ist dies sinnvoll möglich?



Ziele des Projekts

Vergleich der Cluster

- Verteilung von GWL in den Clustern
- Vergleich der Zusammensetzung der einzelnen Cluster:
Wie scheinen sie sich auffällig zu unterscheiden?



1. Einführung

- i. Vorstellen des Projekts
- ii. Datensätze

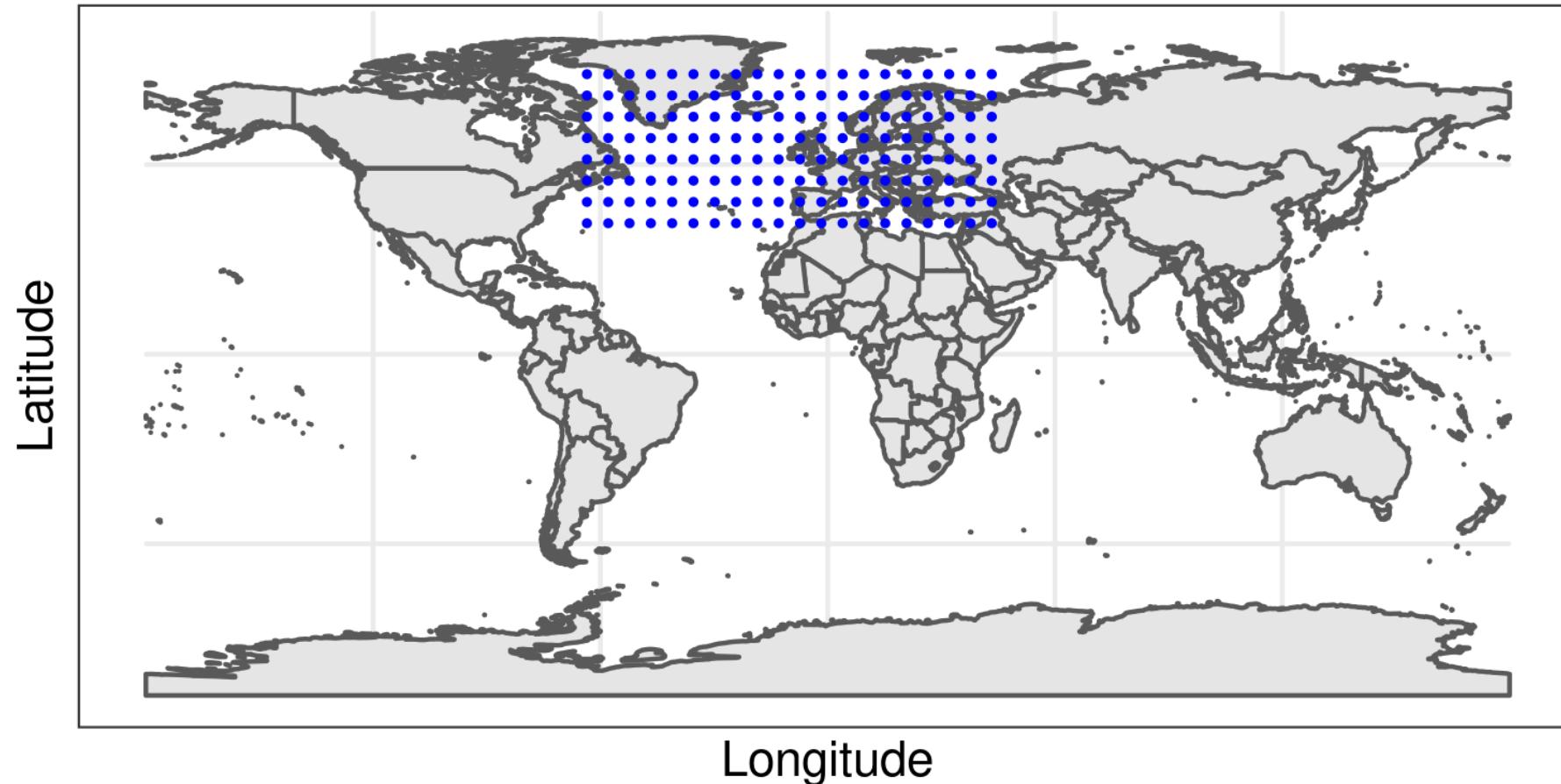


Reanalyse Datensatz

- Pro Tag Messungen an 160 Standorten zu 4 Zeitpunkten
 - Luftdruck in Pa auf Meeresspiegelhöhe (mslp)
 - Geopotential auf 500 hPa in $\frac{m^2}{s^2} = \frac{1}{9.80665} gpm$ (geopot)
- Standorte im 8x20 Grid über Europa und dem Nordatlantik
- Für die Jahre 1900 bis 2010
 - Beschränkung auf eine Klimaperiode: Jahre 1971 bis 2000



Messpunkte auf einer Weltkarte





Auszug aus dem Reanalyse Datensatz

	time	longitude	latitude	mslp	geopotential
1	1900-01-01 00:00:00	-63.56287	73.85311	100428.99	48268.86
2	1900-01-01 00:00:00	-63.56287	68.23695	100553.77	48770.82
3	1900-01-01 00:00:00	-63.56287	62.62077	99920.18	49171.14
4	1900-01-01 00:00:00	-63.56287	57.00457	100049.80	49487.83
...					
640	1900-01-01 18:00:00	43.31280	34.53973	102281.97	55097.32
641	1900-01-02 00:00:00	-63.56287	73.85311	99886.71	47843.04
...					
25946239	2010-12-31 18:00:00	43.31280	40.15595	101758.62	54154.39
25946240	2010-12-31 18:00:00	43.31280	34.53973	101400.51	54491.94



Auszug aus dem Reanalyse Datensatz

	time	longitude	latitude	mslp	geopotential
1	1900-01-01 00:00:00	-63.56287	73.85311	100428.99	48268.86
2	1900-01-01 00:00:00	-63.56287	68.23695	100553.77	48770.82
3	1900-01-01 00:00:00	-63.56287	62.62077	99920.18	49171.14
4	1900-01-01 00:00:00	-63.56287	57.00457	100049.80	49487.83
...					
640	1900-01-01 18:00:00	43.31280	34.53973	102281.97	55097.32
641	1900-01-02 00:00:00	-63.56287	73.85311	99886.71	47843.04
...					
25946239	2010-12-31 18:00:00	43.31280	40.15595	101758.62	54154.39
25946240	2010-12-31 18:00:00	43.31280	34.53973	101400.51	54491.94

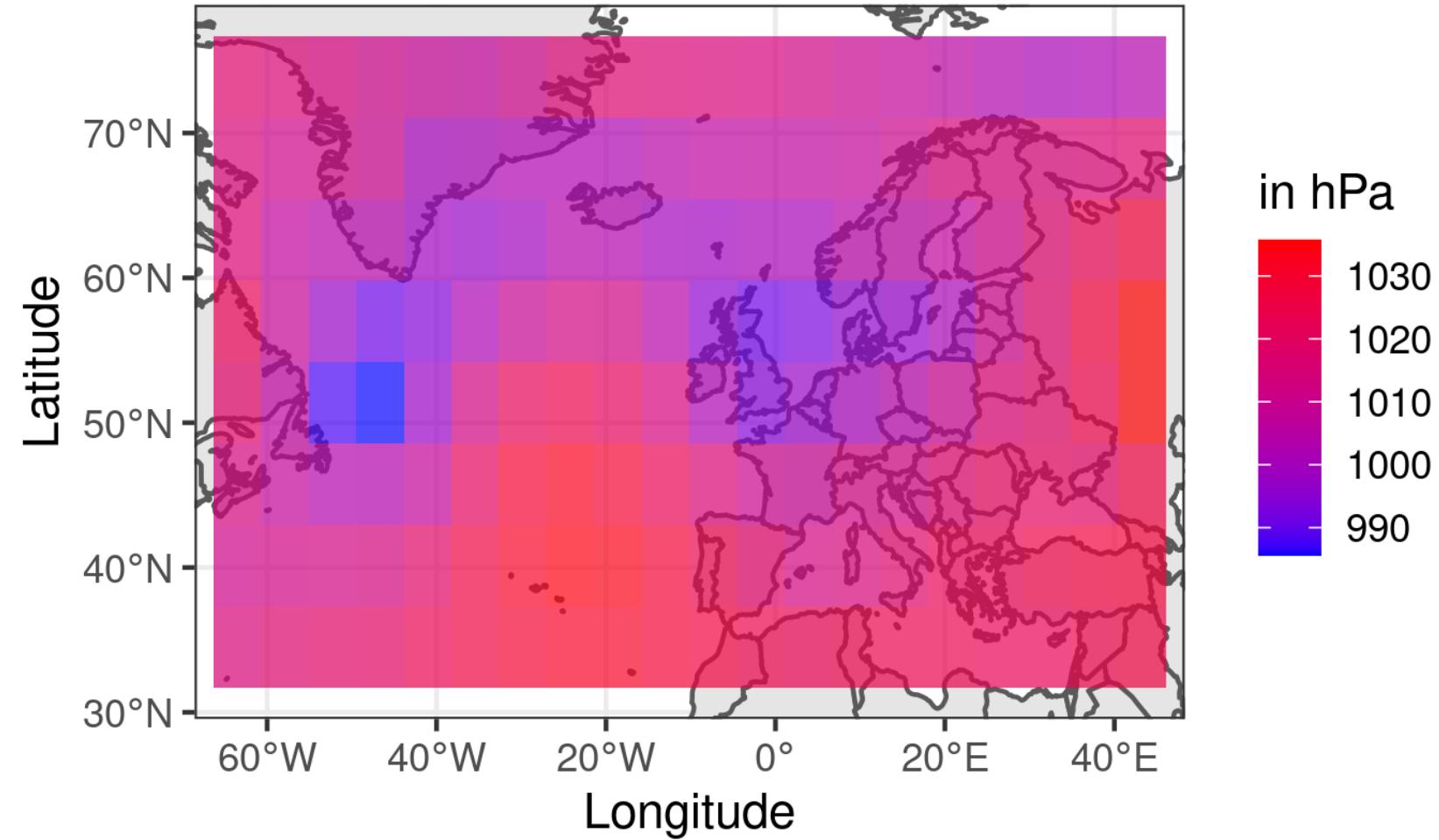


Auszug aus dem Reanalyse Datensatz

	time	longitude	latitude	mslp	geopotential
1	1900-01-01 00:00:00	-63.56287	73.85311	100428.99	48268.86
2	1900-01-01 00:00:00	-63.56287	68.23695	100553.77	48770.82
3	1900-01-01 00:00:00	-63.56287	62.62077	99920.18	49171.14
4	1900-01-01 00:00:00	-63.56287	57.00457	100049.80	49487.83
...					
640	1900-01-01 18:00:00	43.31280	34.53973	102281.97	55097.32
641	1900-01-02 00:00:00	-63.56287	73.85311	99886.71	47843.04
...					
25946239	2010-12-31 18:00:00	43.31280	40.15595	101758.62	54154.39
25946240	2010-12-31 18:00:00	43.31280	34.53973	101400.51	54491.94

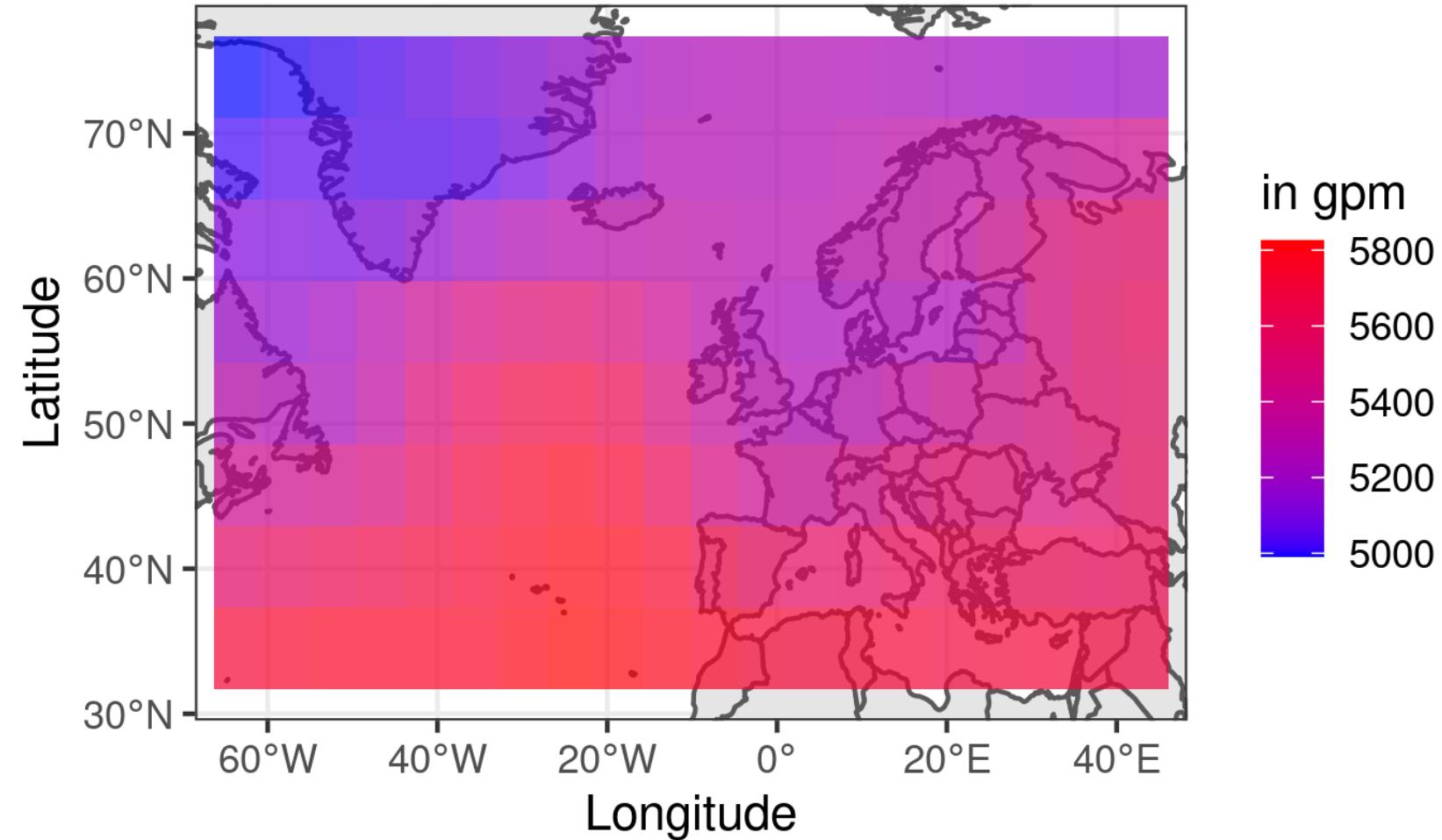


Mslp am 01-01-2006 um 0 Uhr





Geopot am 01-01-2006 um 0 Uhr





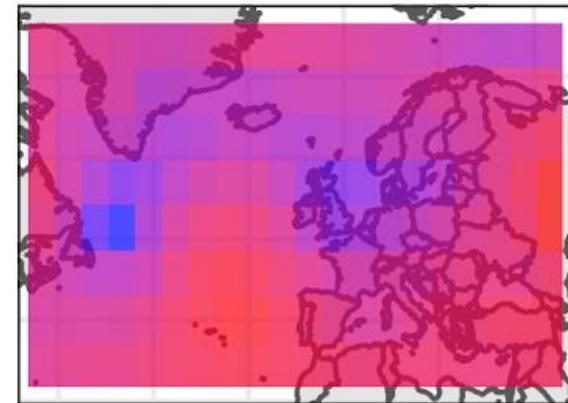
Daten pro Tag

Der Tag ist die Beobachtungseinheit

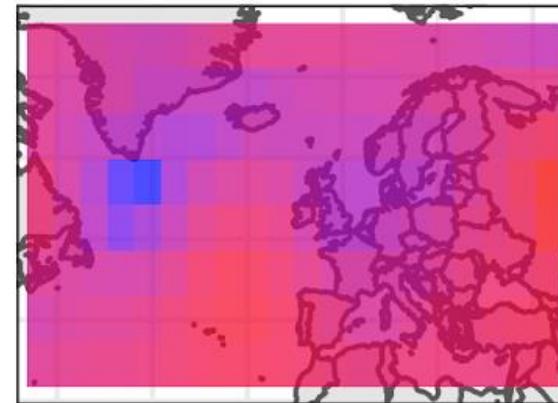
- 2 Parameter * 4 Zeitpunkte * 160 Messpunkte = 1280 Dimensionen
- 8 Bilder pro Tag



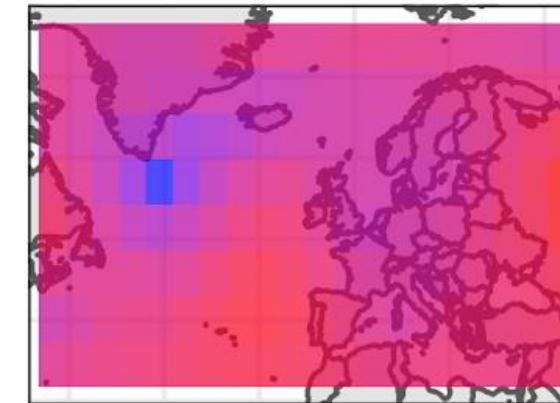
0 Uhr



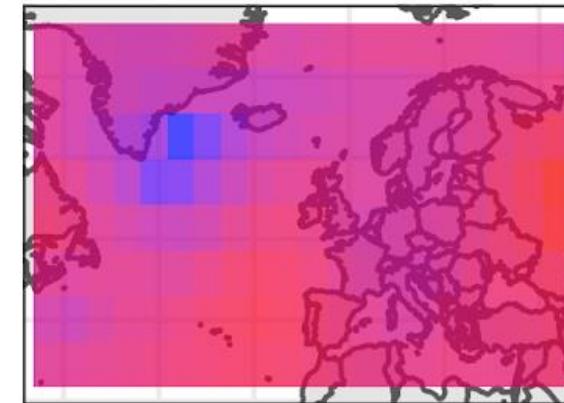
6 Uhr



12 Uhr

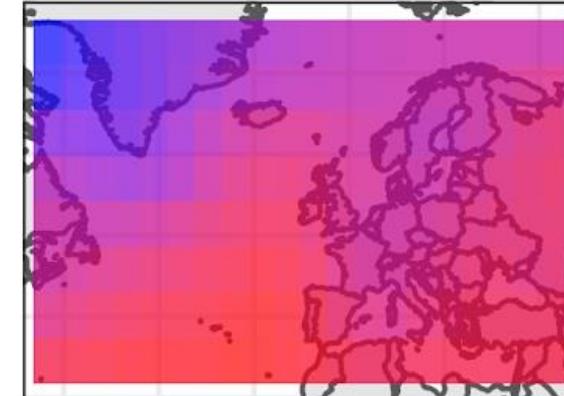
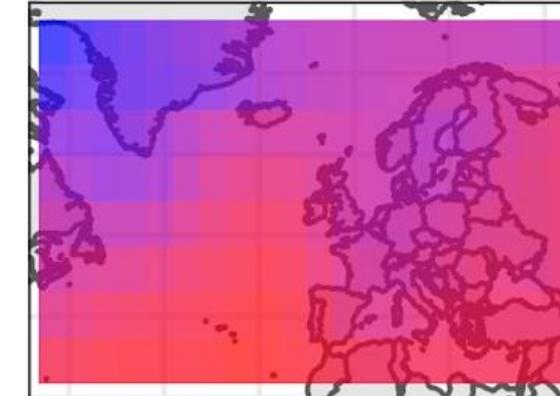
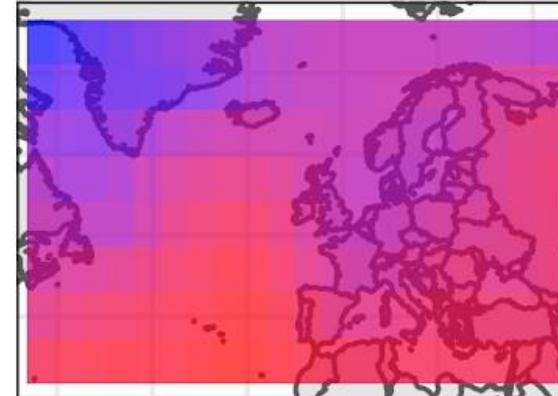
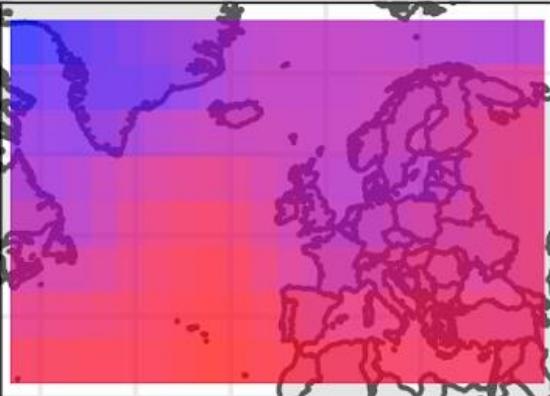


18 Uhr



Mslp

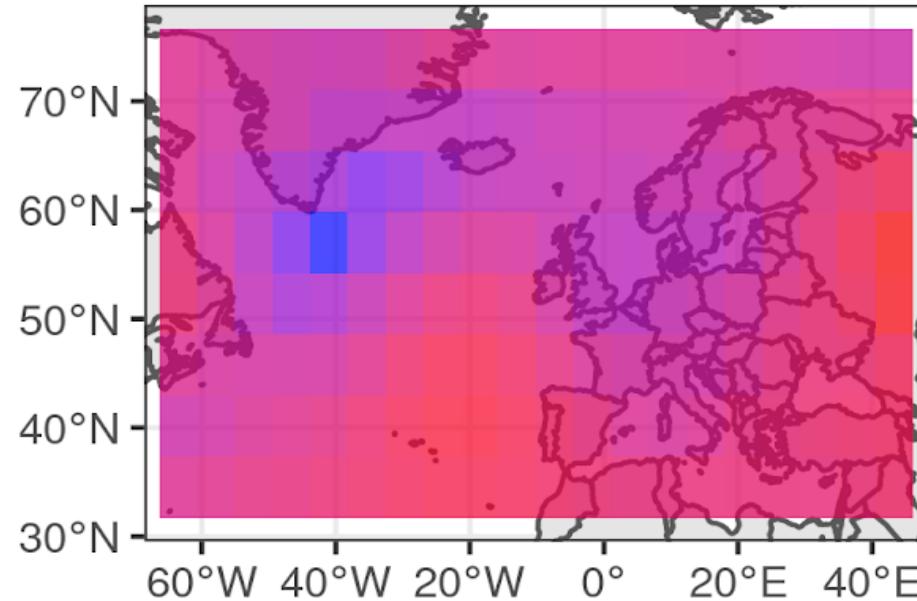
Geopotential



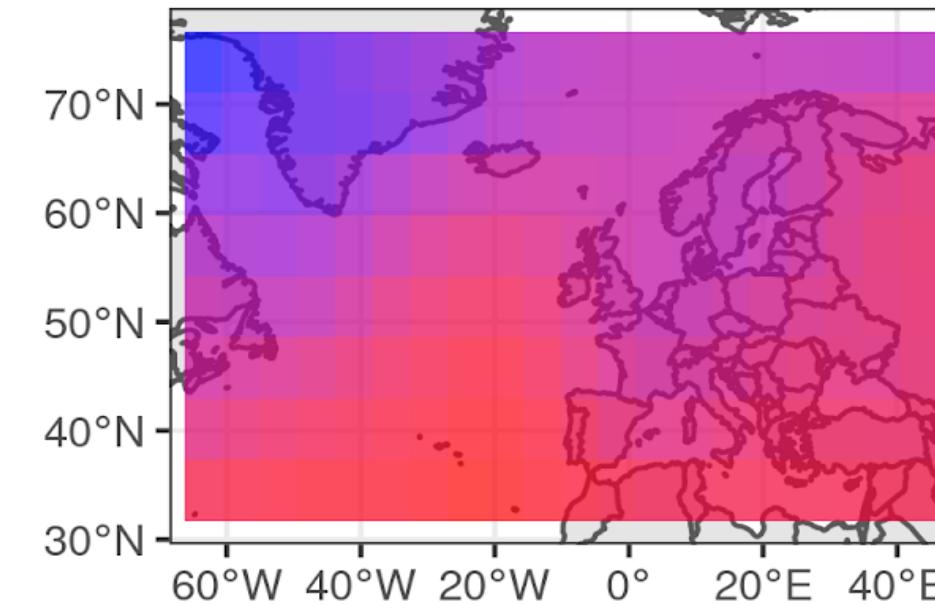


Mittelwerte am 01.01.2006

Mslp



Geopot





Daten pro Tag

Der Tag ist die Beobachtungseinheit

- 2 Parameter * 4 Zeitpunkte * 160 Messpunkte = 1280 Dimensionen
- 8 Bilder pro Tag

Reduzierung der Dimensionen

- Mittelwert über 4 Messzeiten pro Messpunkt
- 10958 Tage mit jeweils 320 Dimensionen



1. Einführung

- i. Vorstellen des Projekts
- ii. Datensätze

2. Methodik

- i. Preprocessing



Datensatz Mutation

- Idee: Erstellen eines Datensatzes durch Extrahieren gezielter Information
- Gezielte Informationen
 - Verteilung der Parameter (im Vergleich zu anderen Tagen)
 - Räumliche Lage und Form der „Hoch-“ und „Tiefgebiete“
 - Veränderung über den Tag



Datensatz Mutation

- Idee: Erstellen eines Datensatzes durch Extrahieren gezielter Information
- Gezielte Informationen
 - Verteilung der Parameter (im Vergleich zu anderen Tagen)
 - Räumliche Lage und Form der „Hoch-“ und „Tiefgebiete“
 - Veränderung über den Tag
- Erhoffte Wirkung
 - Dimensionen weiter reduzieren
 - Spezifische Gewichtung wichtiger Größen
 - Verbesserte Interpretierbarkeit



Vorgehen

- Ausgangslage: Datensatz mit 320 Dimensionen roher Messdaten
 - Transformation zu Variablen, die jeweils eine interessierende Größe über alle Standorte zusammengefasst verkörpern
 - Beispiel: Mittelwert des Luftdrucks über alle Standorte am Tag
 - Insgesamt 48 Variablen
- Beobachtungseinheit bleibt der Tag



Extrahierte Variablen

Variable	Erklärung
Minimum/Maximum	Minimaler/Maximaler Wert am Tag
Mittelwert	Mittelwert für beide Variablen pro Tag
Median/Quartile	Median und Quartile für beide Variablen pro Tag
Intensität	Anzahl der Messpunkte von beiden Variablen pro Tag die über/unter den Quartilen liegen
Differenz am Tag	Summierte Differenzen von 4 Messzeitpunkten am Tag an allen Standorten



Extrahierte Variablen

Variable	Erklärung
Minimum/Maximum	Minimaler/Maximaler Wert am Tag
Mittelwert	Mittelwert für beide Variablen pro Tag
Median/Quartile	Median und Quartile für beide Variablen pro Tag
Intensität	Anzahl der Messpunkte von beiden Variablen pro Tag die über/unter den Quartilen liegen
Differenz am Tag	Summierte Differenzen von 4 Messzeitpunkten am Tag an allen Standorten



Verteilungsvariablen



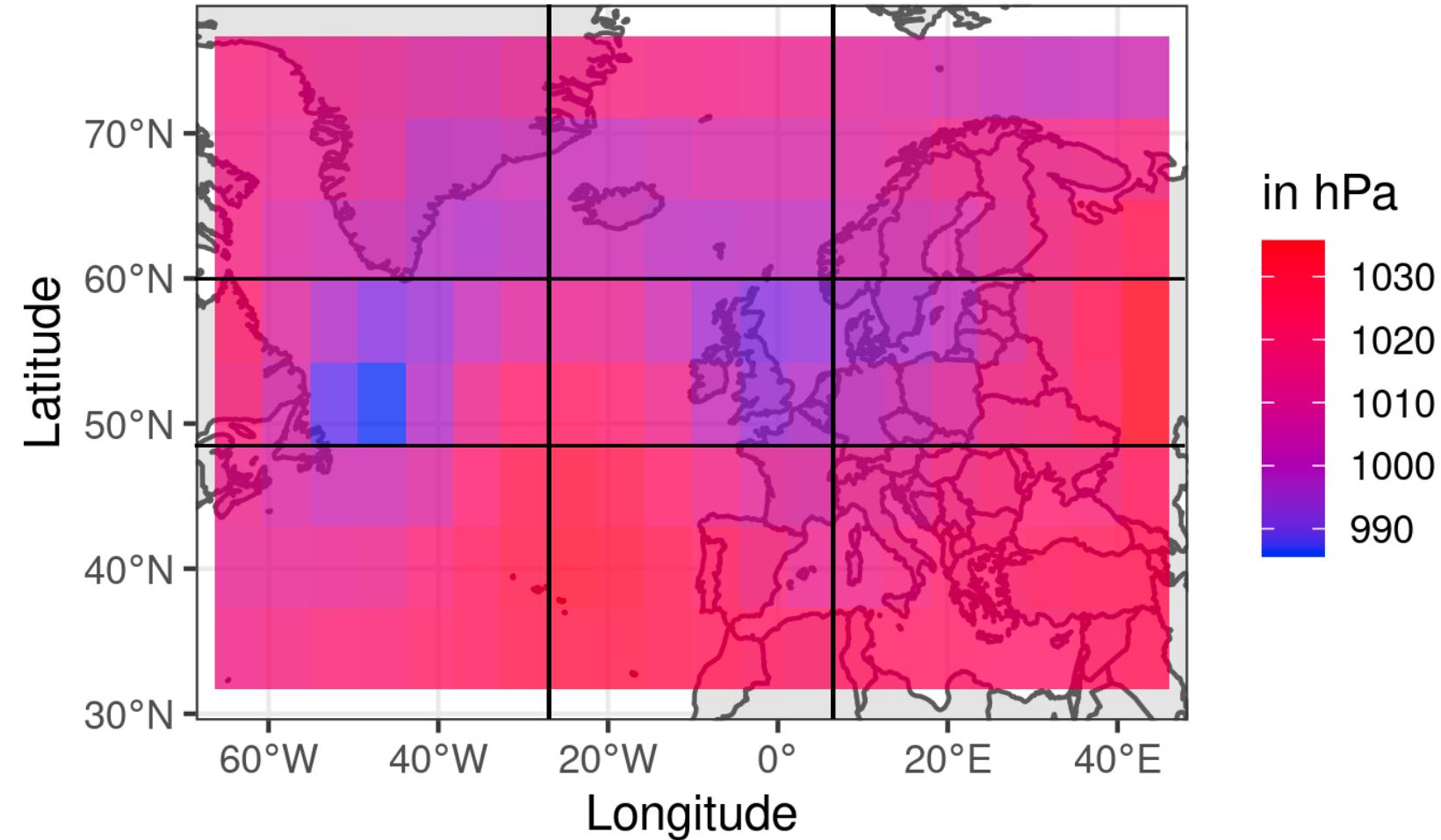
Extrahierte Variablen

Variable	Erklärung
Spalte vom Minimum/Maximum	In welchem Bereich liegt das Minimum/Maximum? Karte aufgeteilt in 3 Spalten
Zeile vom Minimum/Maximum	In welchem Bereich liegt das Minimum/Maximum? Karte aufgeteilt in 3 Zeilen
Distanz zwischen Extrema	Euklidische Distanz
Distanz der beiden Minima und Maxima	Euklidischer Abstand vom Minimum/Maximum der Parameter Geopotential zu Mslp
Mittelwerte in den Quadranten	Mittelwerte in allen 9 Quadranten von beiden Variablen

} Lage der Extrema



Mslp am 01-01-2006 um 0 Uhr





Extrahierte Variablen

Variable	Erklärung
Spalte vom Minimum/Maximum	In welchem Bereich liegt das Minimum/Maximum? Karte aufgeteilt in 3 Spalten
Zeile vom Minimum/Maximum	In welchem Bereich liegt das Minimum/Maximum? Karte aufgeteilt in 3 Zeilen
Distanz zwischen Extrema	Euklidische Distanz
Distanz der beiden Minima und Maxima	Euklidischer Abstand vom Minimum/Maximum der Parameter Geopotential zu Mspl
Mittelwerte in den Quadranten	Mittelwerte in allen 9 Quadranten von beiden Variablen



Räumliche Variablen



Skalierung und Gewichtung

- Datensatz wird standardisiert, da die Skalen der einzelnen Variablen unterschiedlich sind

$$x_{i,\text{neu}} = \frac{x_i - \mu_i}{\sigma_i} \quad \text{mit } i = 1, \dots, 48$$

- Variablen werden zudem gewichtet
 - Aufgeteilt in Kategorien, die jeweils in Summe gleich gewichtet sind



Skalierung und Gewichtung

Variablen	Gewichte
Minimum, Maximum, Mittelwert	$\frac{1}{3}$
Median, Quartile, Intensität und Differenz am Tag	$\frac{1}{6}$
Euklidische Distanzen, Spalten und Zeilen vom Minimum/Maximum	$\frac{1}{6}$
Mittelwert in den Quadranten	$\frac{1}{9}$



1. Einführung

- i. Vorstellen des Projekts
- ii. Datensätze

2. Methodik

- i. Preprocessing
- ii. Wahl des Clusterverfahrens



Clusteranalyse

- Verfahren des "unsupervised learning" (kein Target)
- Grundidee: Bildung von möglichst homogenen Gruppen, Cluster untereinander möglichst heterogen
- Betrachten von n Objekten a_1, \dots, a_n mit zugehörigen Merkmalsvektoren x_1, \dots, x_p
 - Suchen einer Partition C_1, \dots, C_k mit $\bigcup_{i=1}^k C_i = \{a_1, \dots, a_n\}$ wobei $C_i \cap C_j = \emptyset \quad \forall i \neq j$
- Verschiedene Ansätze für Clustering
 - Optimale Partitionen
- Distanz zwischen Objekten durch Ähnlichkeits- bzw. Distanzmaß



Clusteralgorithmus PAM

- PAM steht für Partitioning Around Medoids
- Gehört zu den Partitionierenden Verfahren
- Vorgehen:
 1. Anzahl k an Cluster festlegen
 2. Wahl von k repräsentativen Objekten (Medoids) aus allen Beobachtungen
 3. Für jeden Medoid m und jeden restlichen Datenpunkt o :
 - i. Entscheiden, ob ein Datenpunkt o einen Medoid m ersetzen soll anhand der Summe S der Distanzen von allen Datenpunkten zu deren jeweiligen Medoid
 - ii. Durchführen für alle Datenpunkte
 - iii. Auswahl der Datenpunkte als Medoids, die die Summe S am stärksten minimieren
 4. Datenpunkte dem Cluster zuteilen, dessen Medoid am nächsten zu o liegt



Distanzmaß

- Manhattan-Metrik
 - die Distanz d zwischen zwei Objekten a und b definiert ist als

$$d(a, b) = \sum_{i=1}^p |a_i - b_i|$$

wobei $a = (a_1, \dots, a_p)$, $b = (b_1, \dots, b_p)$



1. Einführung

- i. Vorstellen des Projekts
- ii. Datensätze

2. Methodik

- i. Preprocessing
- ii. Wahl des Algorithmus
- iii. Bewertungskriterien für Cluster



Bewertungskriterien für Clustering

- Silhouettenkoeffizient
 - Maßzahl für die Qualität eines Clusterings
 - Unabhängig von der Anzahl der Cluster
 - Gehört das Objekt o zum Cluster A , so ist die Silhouette von o definiert als

$$S(o) = \begin{cases} 0 & \text{Wenn } x \text{ einziges Element von } A, \text{ ist} \\ \frac{dist(B, o) - dist(A, o)}{\max\{dist(A, o), dist(B, o)\}} & \text{sonst,} \end{cases}$$

wobei $dist(A, o)$ die durchschnittliche Distanz eines Objektes o zu anderen Punkten des Clusters A
 $dist(B, o)$ die Distanz eines Objektes o zum nächstgelegenen Objekt des Clusters B



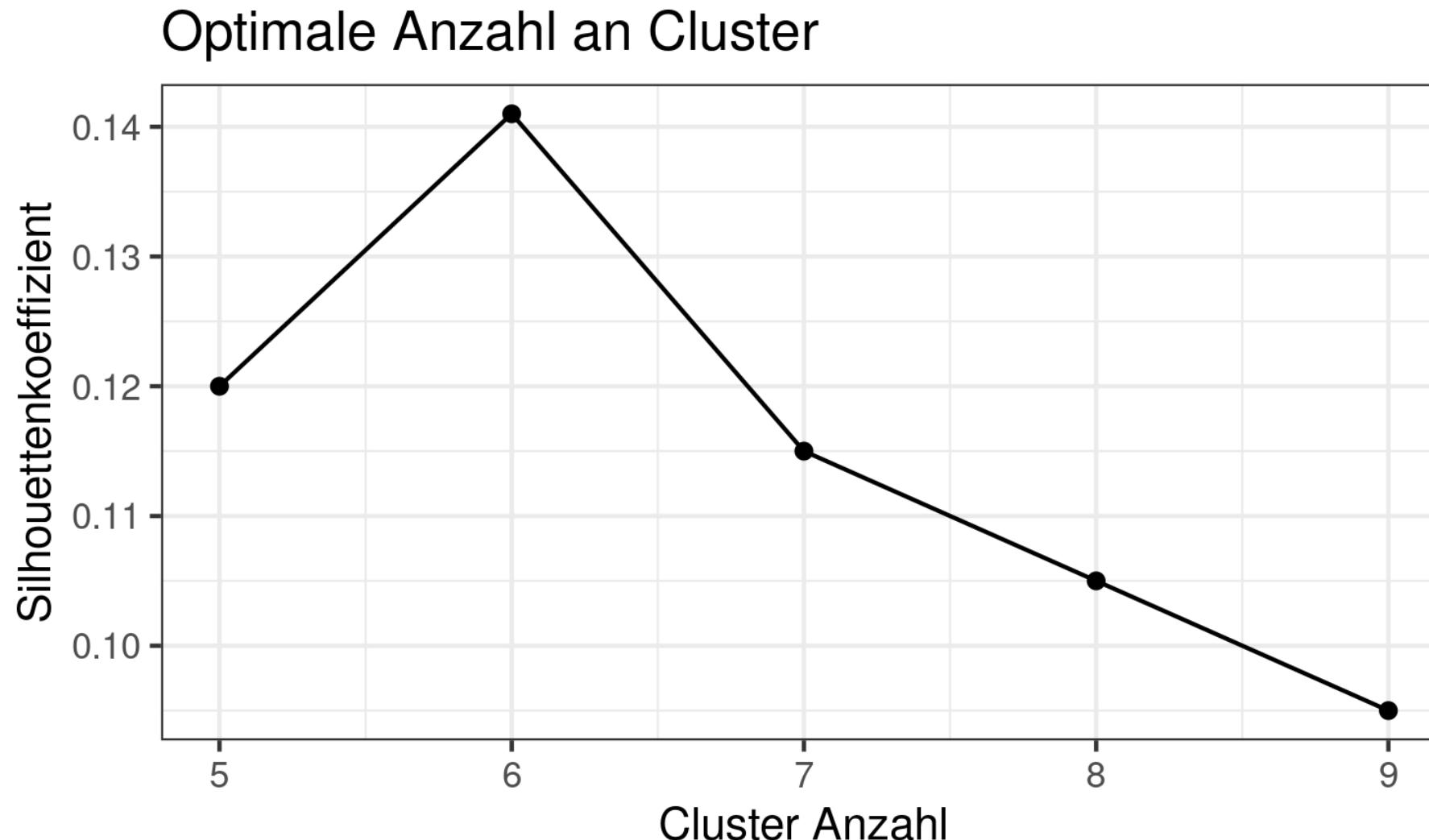
Bewertungskriterien für Clustering

- Silhouettenkoeffizient
 - Sei k die Anzahl an Cluster, dann ist der Silhouettenkoeffizient definiert durch

$$s = \frac{1}{n} \sum_{o \in N} S(o)$$

wobei

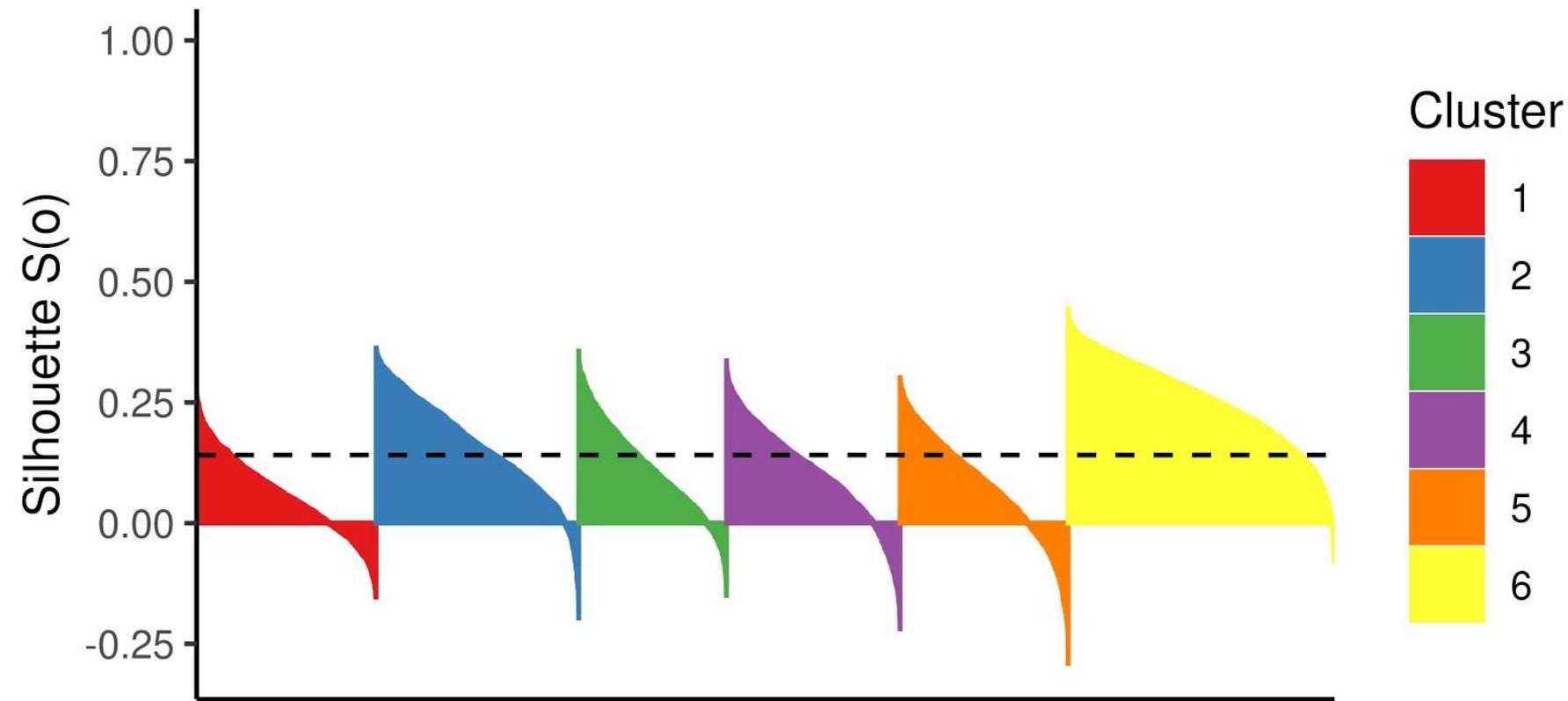
$$S(o) = \begin{cases} 0 & \text{Wenn } x \text{ einziges Element von } A, \text{ ist} \\ \frac{dist(B, o) - dist(A, o)}{\max\{dist(A, o), dist(B, o)\}} & \text{sonst,} \end{cases}$$





Silhouettenplot

Silhouettenkoeffizient: 0.141



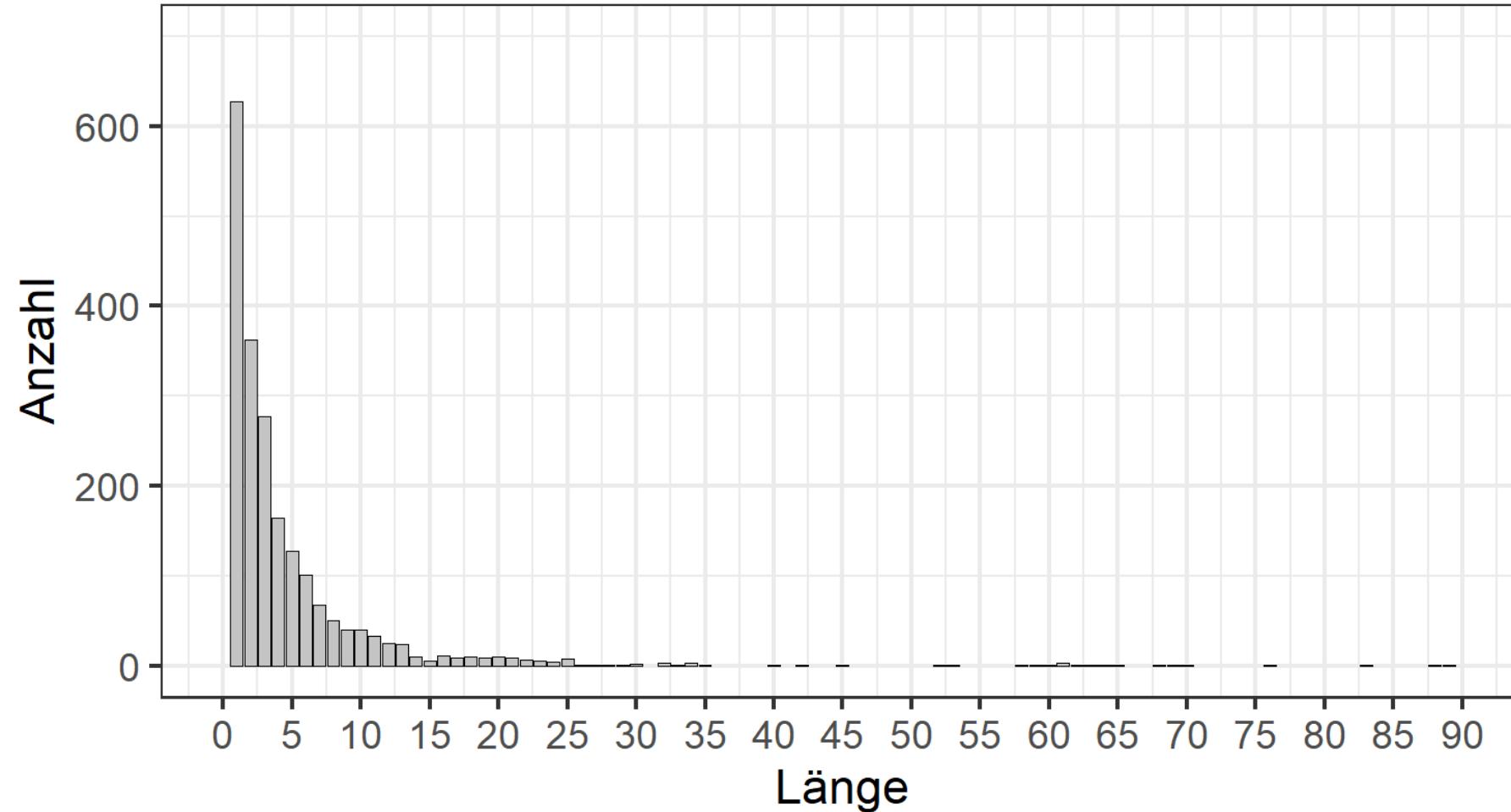


Bewertungskriterien für Clustering

- Timeline
 - Häufigkeiten bestimmter Längen an aufeinanderfolgenden Tagen im selben Cluster
 - Erwünscht:
 - Längen ab 3 Tagen
 - Nach oben limitiert



Timeline





Beispiele

	date	cluster
1	1971-04-26	1
2	1971-04-27	1
3	1971-04-28	1
4	1971-04-29	4
5	1971-04-30	1
6	1971-05-01	4
7	1971-05-02	4
8	1971-05-03	4

→ Übergang zwischen Clustern oft nicht sauber



1. Einführung

- i. Vorstellen des Projekts
- ii. Datensätze

2. Methodik

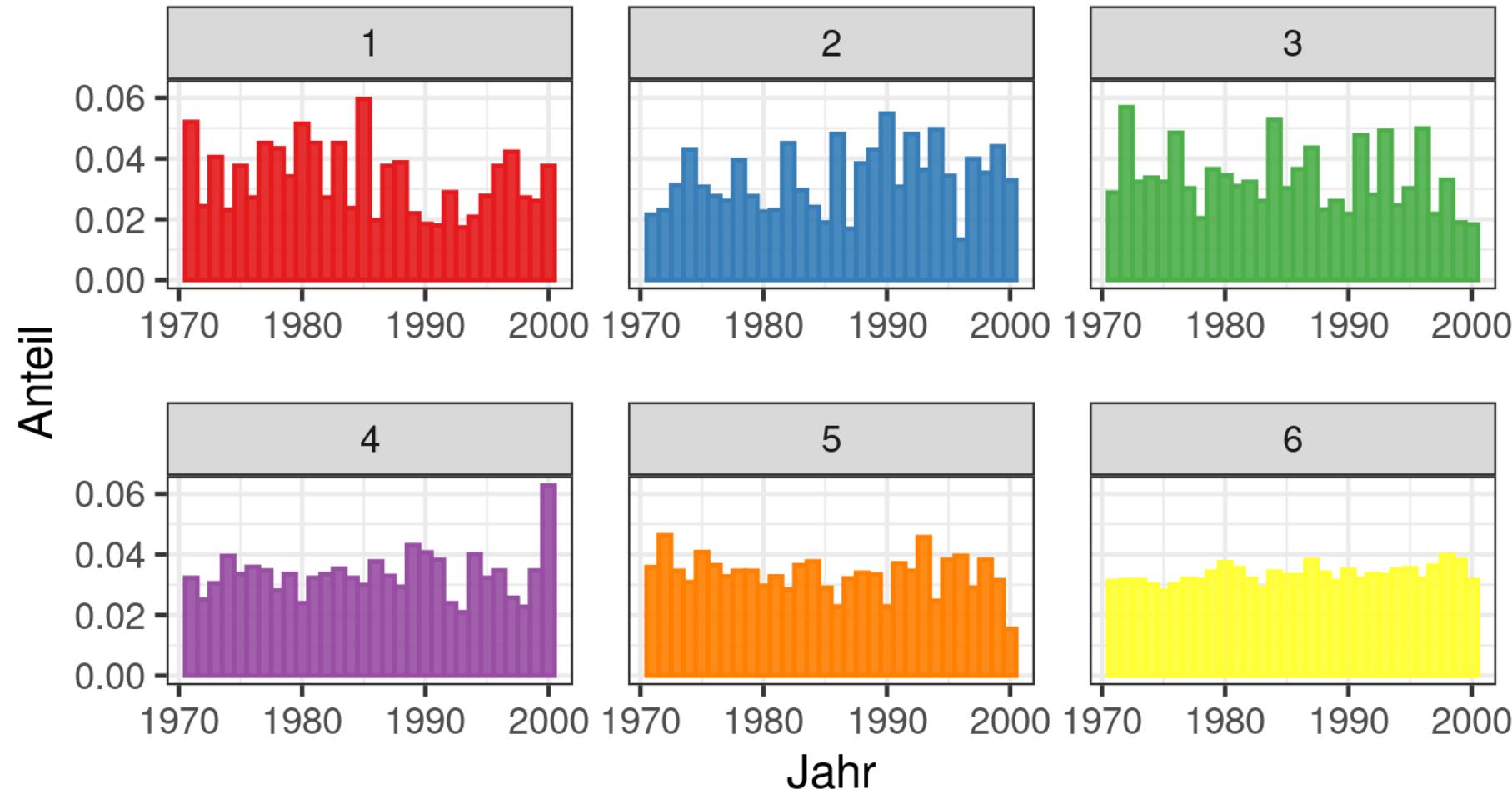
- i. Preprocessing
- ii. Wahl des Algorithmus
- iii. Bewertungskriterien für Cluster

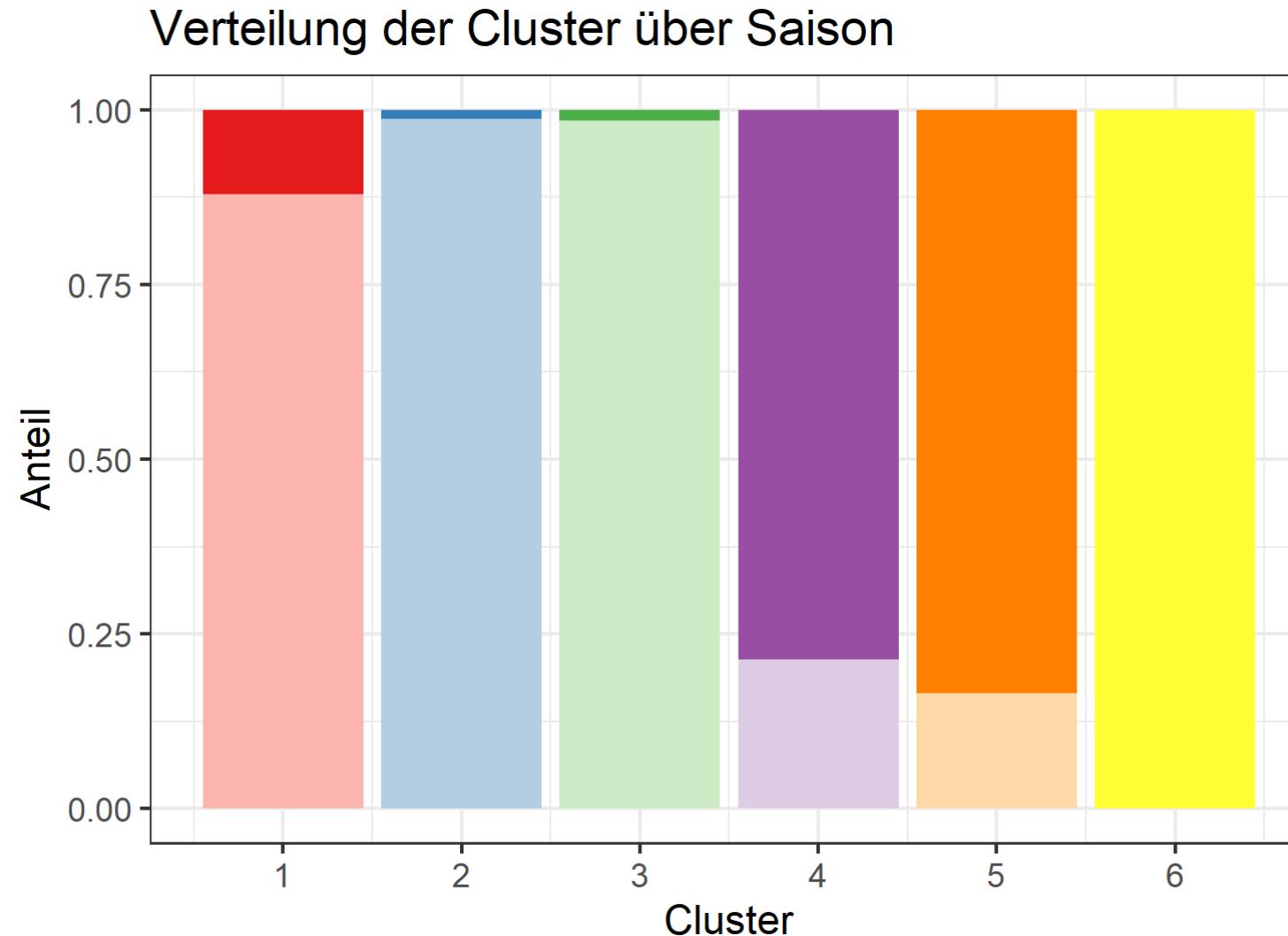
3. Deskriptive Analyse

- i. Verteilung über die Zeit



Verteilung der Cluster über die Jahre





Cluster nach Sommer

1	2
3	4
5	6

Cluster nach Winter

1	2
3	4
5	6



1. Einführung

- i. Vorstellen des Projekts
- ii. Datensätze

2. Methodik

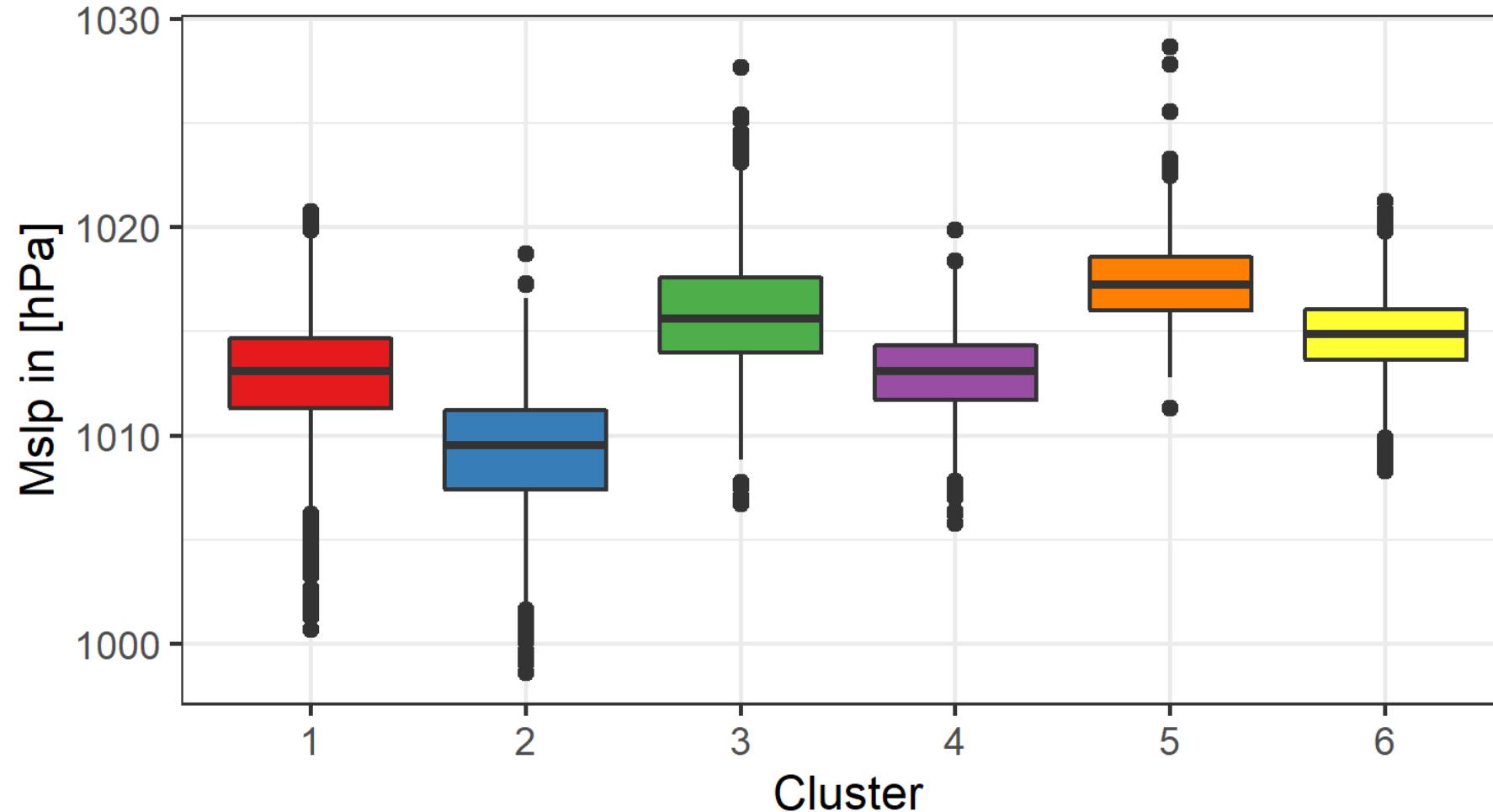
- i. Preprocessing
- ii. Wahl des Algorithmus
- iii. Bewertungskriterien für Cluster

3. Deskriptive Analyse

- i. Verteilung über die Zeit
- ii. Unterschiede und Ähnlichkeiten in den Clustern

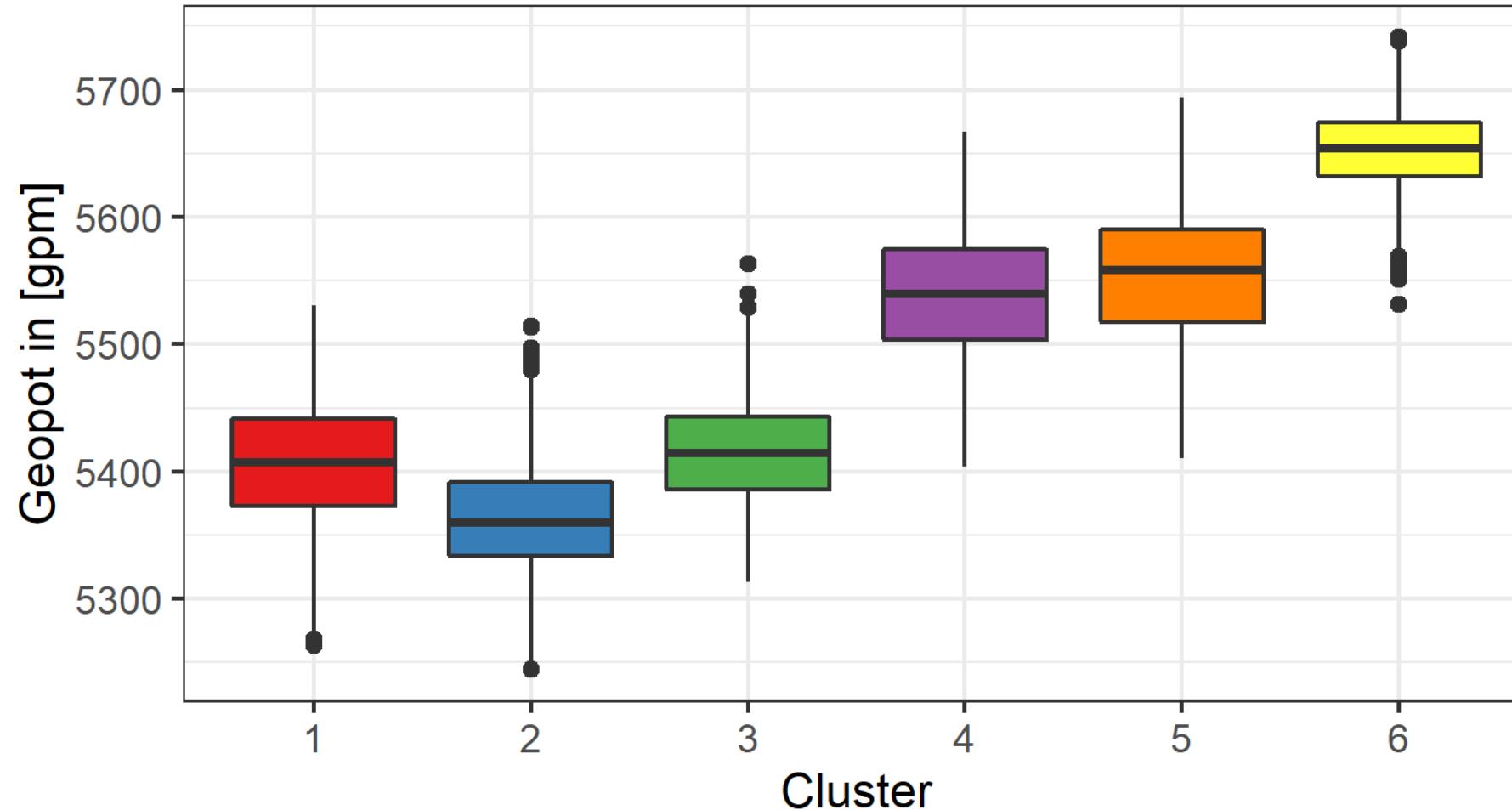


Mittelwert des Mslp in jedem Cluster



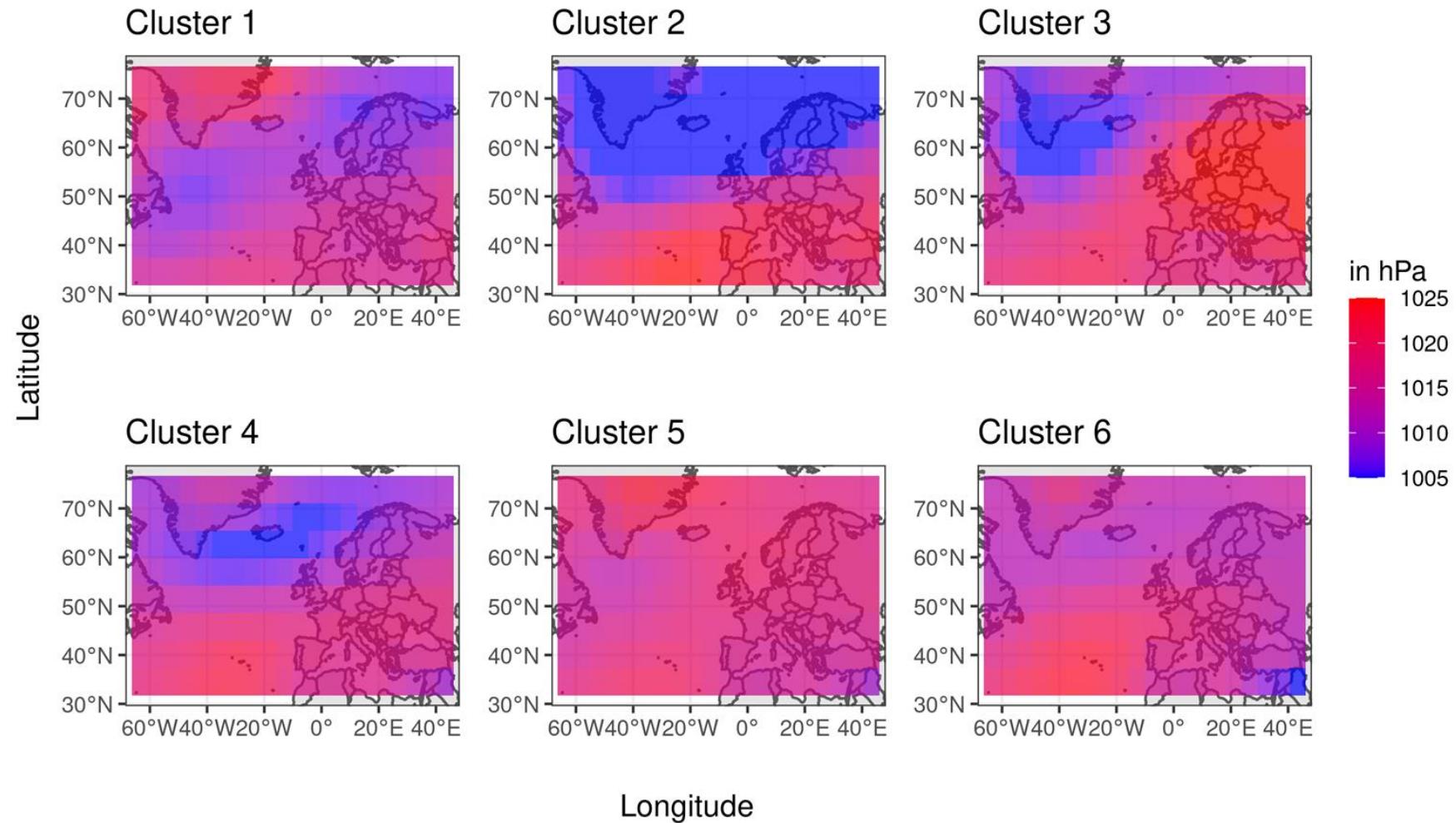


Mittelwert des Geopot in jedem Cluster



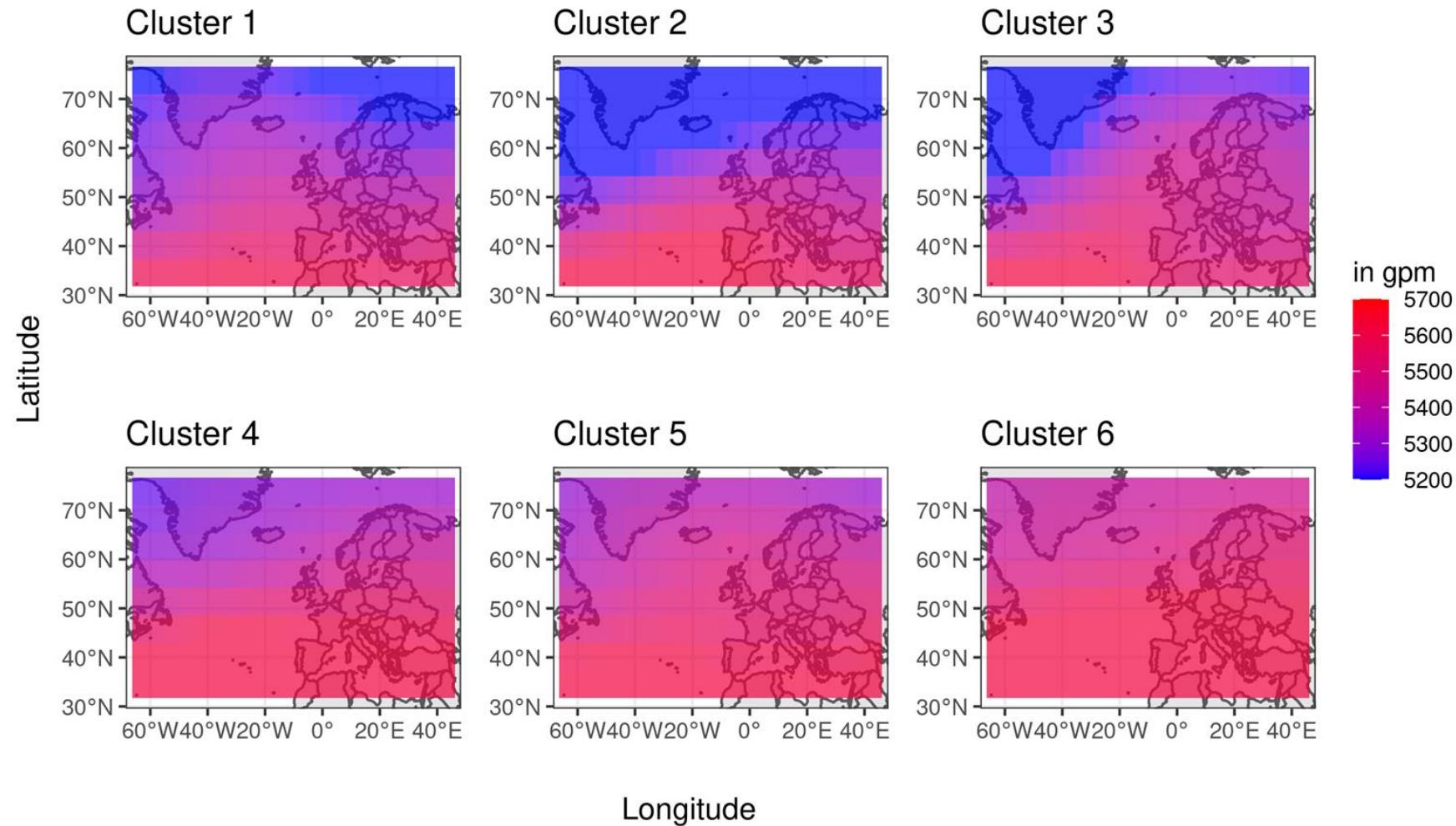


Mslp im Mittel über Messpunkte





Geopot im Mittel über Messpunkte





1. Einführung

- i. Vorstellen des Projekts
- ii. Datensätze

2. Methodik

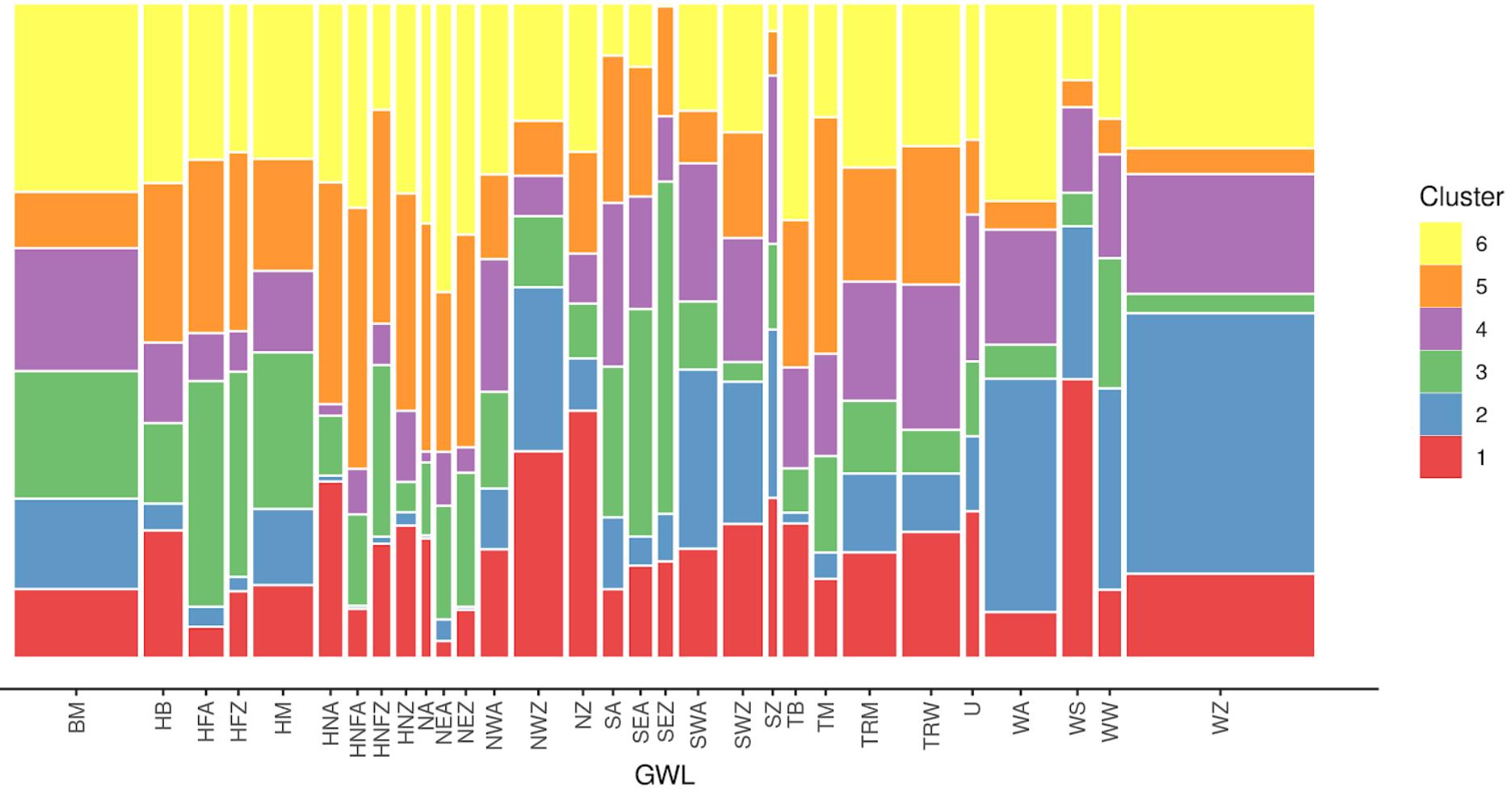
- i. Preprocessing
- ii. Wahl des Algorithmus
- iii. Bewertungskriterien für Cluster

3. Deskriptive Analyse

- i. Verteilung über die Zeit
- ii. Unterschiede und Ähnlichkeiten in den Clustern
- iii. Vergleich zur gegebenen GWL-Einteilung



Mosaikplot für Cluster ~ GWL





Beispiele

	date	cluster	gwl
1	1971-04-22	5	SA
2	1971-04-23	5	SA
3	1971-04-24	5	SA
4	1971-04-25	1	HNZ
5	1971-04-26	1	HNZ
6	1971-04-27	1	HNZ



Zum Teil wechseln Cluster passend mit den GWL am Tag



Gliederung

1. Einführung

- i. Vorstellen des Projekts
- ii. Datensätze

2. Methodik

- i. Preprocessing
- ii. Wahl des Algorithmus
- iii. Bewertungskriterien für Cluster

3. Deskriptive Analyse

- i. Verteilung über die Zeit
- ii. Unterschiede und Ähnlichkeiten in den Clustern
- iii. Vergleich zur gegebenen GWL-Einteilung

4. Ausblick

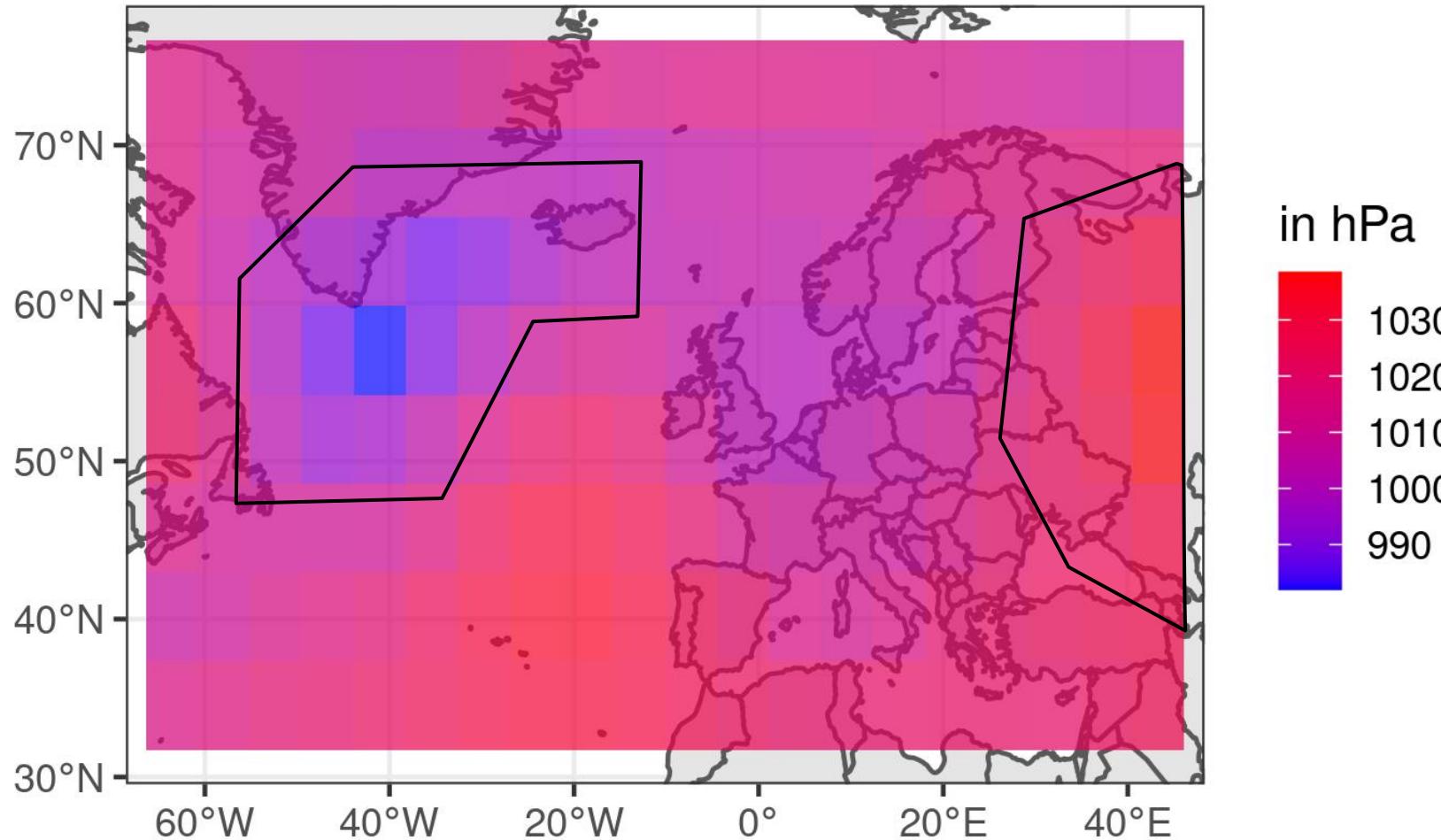


Anderer Ansatz

- Muster-Erkennung in den Bildern der Tage
 - Vorfiltern der Daten pro
 - Verwandlung Messdaten/Standort zu “Gebietszugehörigkeit”/Standor
- ➡ Clustern mit dem Standort als Beobachtungseinheit



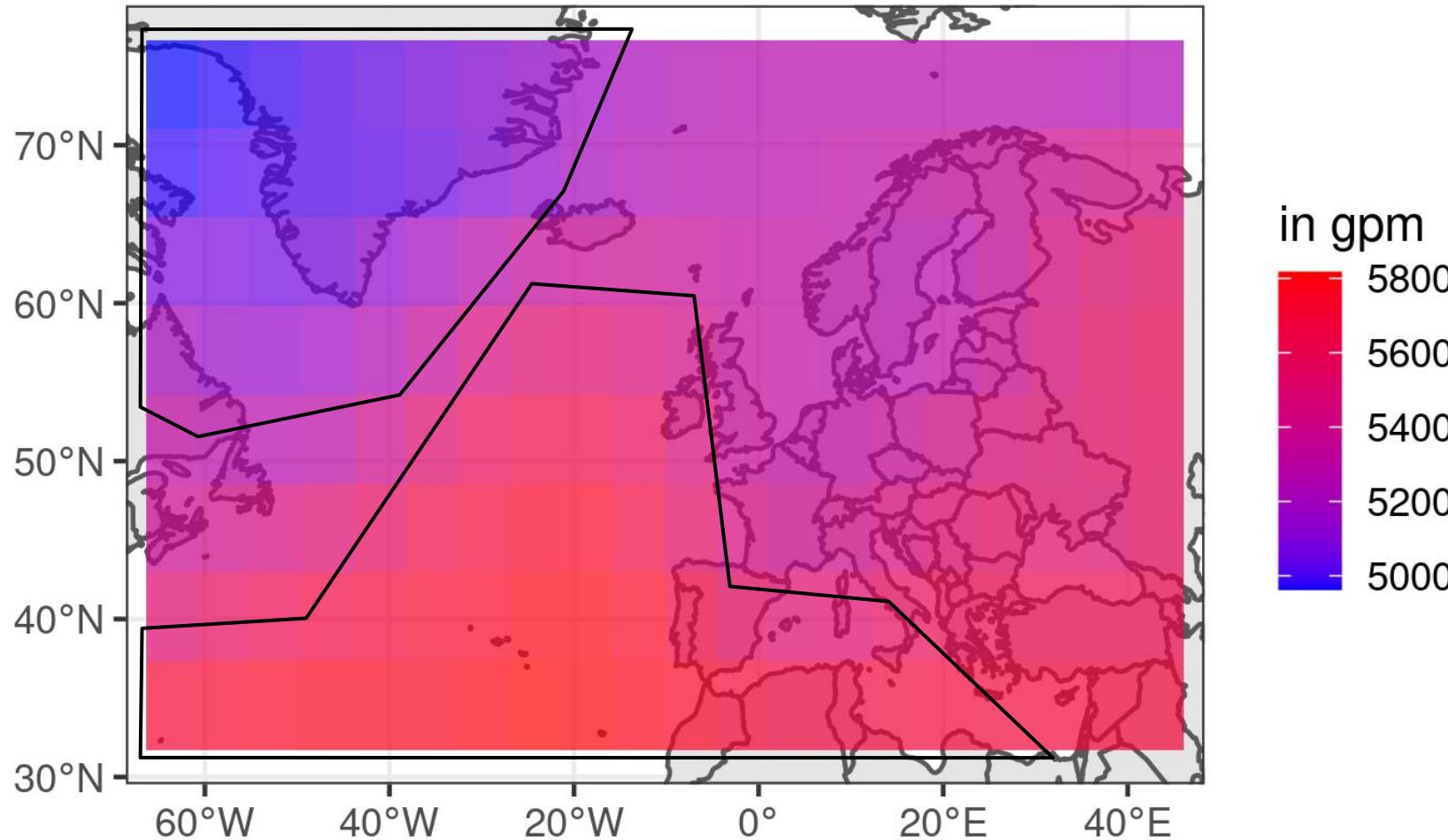
Gemittelter Mslp am 01.01.2006



- Position und Form der “Hoch-” und “Tiefgebiete”



Gemitteltes Geopot am 01.01.2006



- Position und Form der “Hoch-” und “Tiefgebiete”

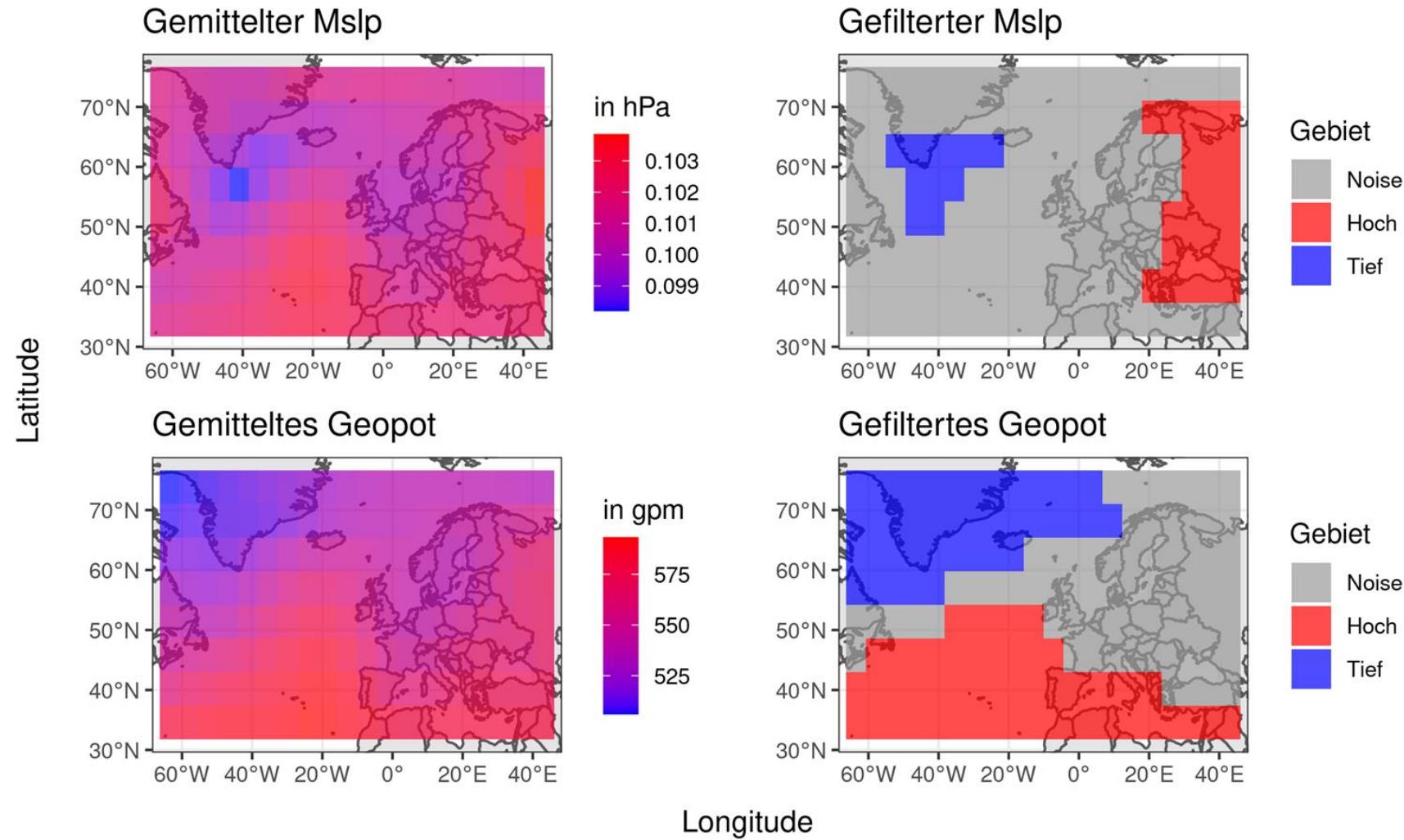


Anderer Ansatz

- Muster-Erkennung in den Bildern der Tage
 - Vorfiltern der Daten pro
 - Verwandlung Messdaten/Standort zu “Gebietszugehörigkeit”/Standor
 - ➡ Clustern mit dem Standort als Beobachtungseinheit
- Clusterverfahren
 - Dichtebasierter Clustern mit Noise
 - Startpunkte der Cluster fix
 - Cluster iterierend wachsen lassen mit zunehmend strengerem Nachbarschaftsparameter



Filtern des 01.01.2006





Anderer Ansatz

- Distanzberechnung zwischen Tagen

$$d(a, b) = 1 - \left(\frac{\sum I(a_i = b_i)}{\sum I(a_i)} \right)$$

wobei: $\sum I(a_i) :=$ Anzahl der Standorte nicht in Noise

- Weiteres Clustern auf Tagesebene mit erhaltener Distanzmatrix



Anderer Ansatz

- Probleme
 - Instabil durch Hyperparameter eps und dessen Verkleinerung
 - Sehr teuer
 - Starkes Reduzieren der gegebenen Information



Ausblick

- Wahl des Gewichtsvektors und der Variablen
 - Ausschlaggebend auf die Clusterbewertungskriterien
 - Fachlich sinnvoll
 - Evtl durch mit mehr Vorinformation über die Daten entscheiden
- Saison
 - Saisonbereinigung
 - Datensatz aufteilen und getrennt analysieren



Ausblick

- Einbeziehen der zeitlichen Struktur
 - Einführen einer 3-Tage-Regel beim Clusterverfahren
 - Datenformat als Video betrachten statt Ansammlung von Bildern
- Einbeziehen weiterer Variablen
 - Anderer vorhandenen Messdaten (z.B. Temperatur)
 - Berechnung der Störmungsrichtung anhand des Bewegens bestimmter Gebiete über den Tag



Gliederung

1. Einführung

- i. Vorstellen des Projekts
- ii. Datensätze

2. Methodik

- i. Preprocessing
- ii. Wahl des Algorithmus
- iii. Ergebnisse

3. Deskriptive Analyse

- i. Verteilung über die Zeit
- ii. Unterschiede und Ähnlichkeiten in den Clustern
- iii. Vergleich zur gegebenen GWL-Einteilung

4. Ausblick

5. Fazit



Fazit

Lassen sich Tage anhand von ihren Wettermesswerten sinnvoll clustern?

- Silhouettenkoeffizient von 0.14
- Instabil

Wie unterscheiden sich die entstandenen Cluster voneinander?

- Starke Unterteilung in Sommer- und Wintertage
- räumliche Unterscheidung auf Mspl Ebene erkennbar, beim Geopotential eher nicht
- GWL Unterteilung spiegelt sich nicht sehr deutlich wieder



Referenzen

- Fattouh, L. & Alharbi, M. Using Modified Partitioning Around Medoids Clustering Technique in Mobile Network Planning. *International Journal of Computer Science Issues* **9** (2013).
- Hoyer, A. (ed Ludwig-Maximilians-Universität München) (Sommersemester 2020).
- James, P. M. An objective classification method for Hess and Brezowsky Grosswetterlagen over Europe. *Theoretical and Applied Climatology* **88**, 17-42, doi:10.1007/s00704-006-0239-3 (2007).
- Neuen, A. *Grosswetterlagen: Die antizyklonale Westlage (WA)*,
<https://wetterkanal.kachelmannwetter.com/grosswetterlagen-die-antizyklonale-westlage-wa/> (11.11.2015).
- Schwarzer. *SKlima.de, private Wetterstation Peißenberg*, <http://sklima.de/impressum.php> (2021).



Anhang

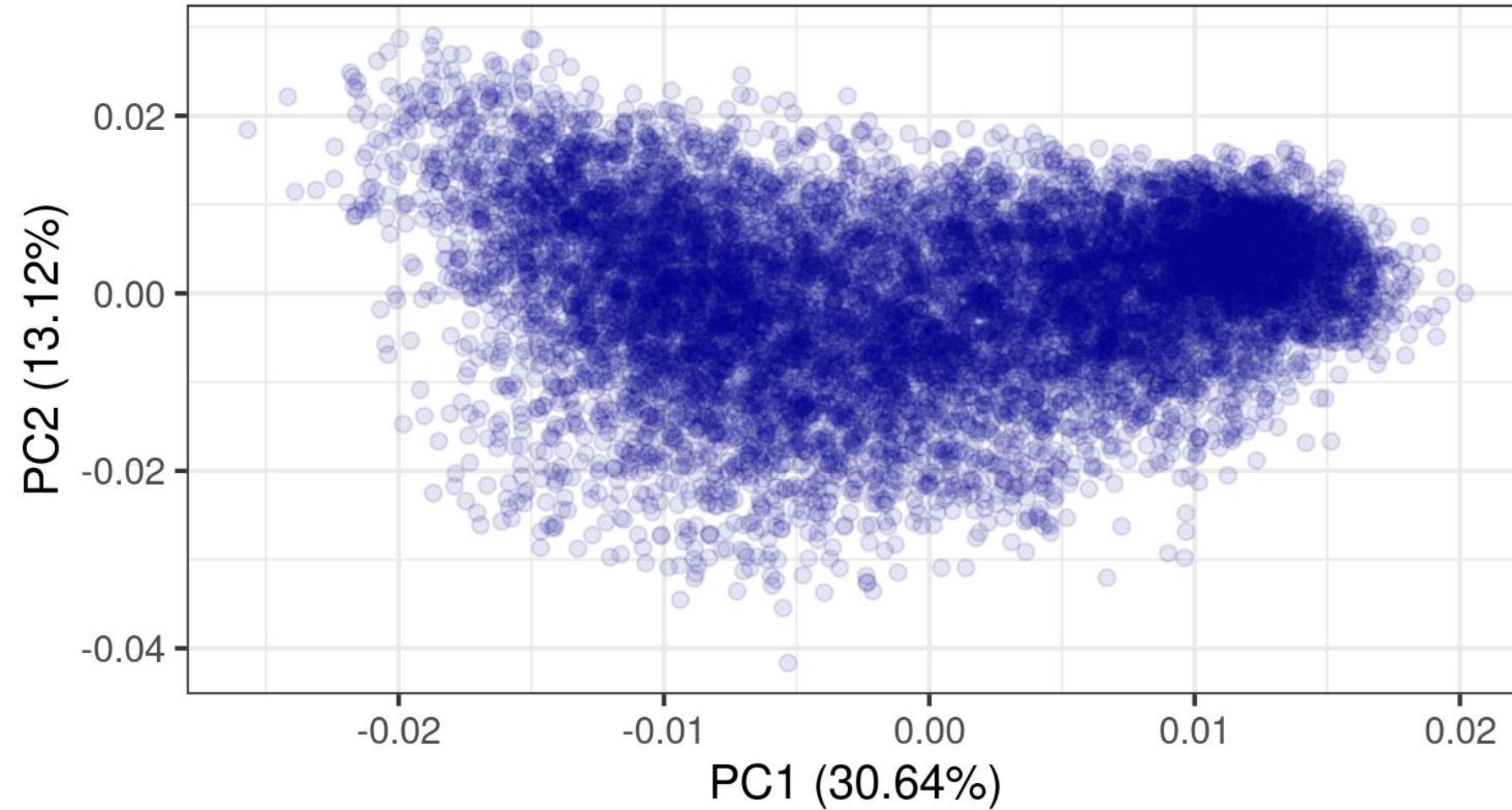


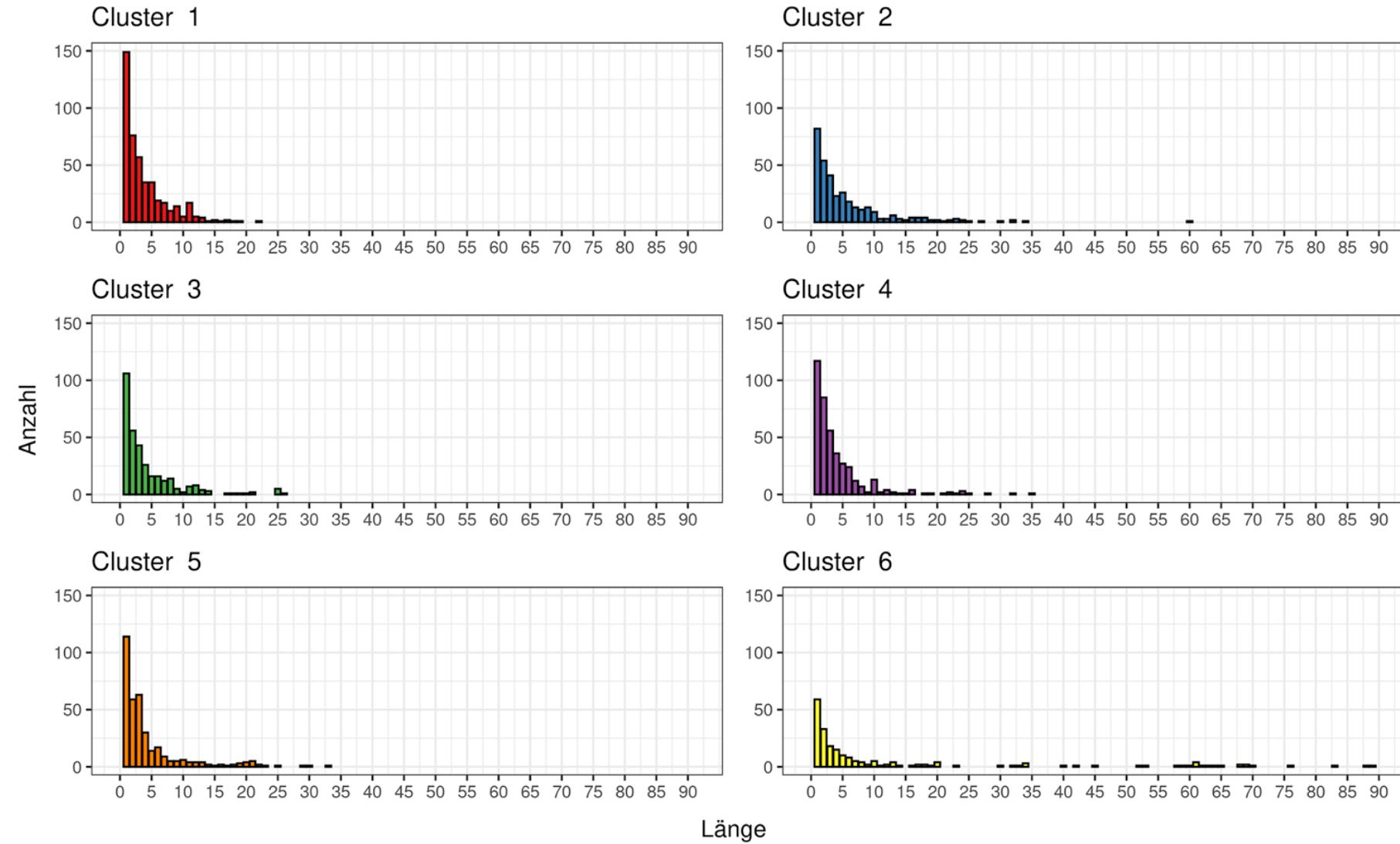
Versuchte Algorithmen/Metriken

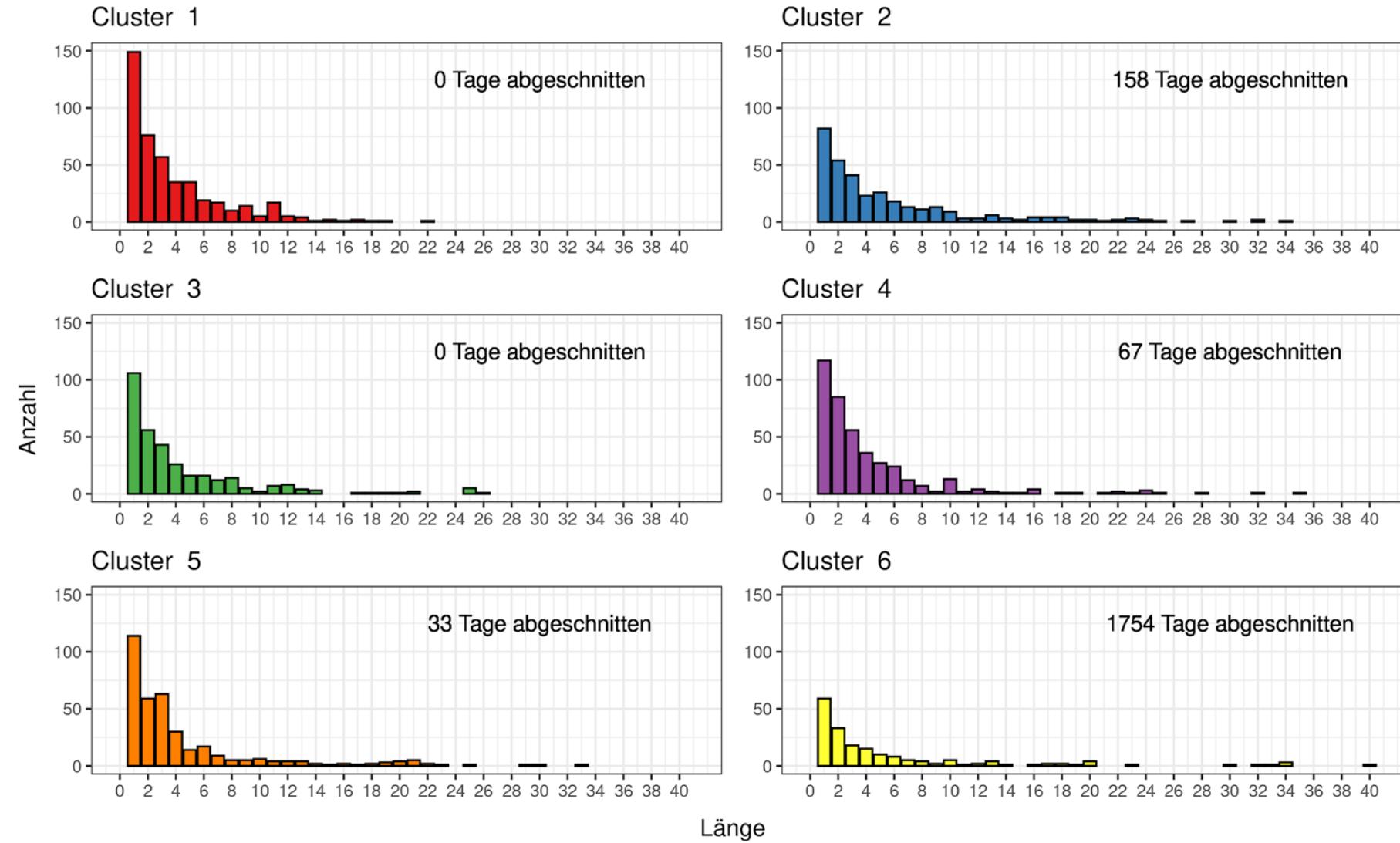
- Cluster Algorithmen:
 - PAM
 - K-means
 - Fuzzy
 - GMM
 - DBSCAN
- Metriken
 - Euklidisch
 - Manhattan
 - Mahalanobis
 - gower



Visualisierung der Daten mit PCA

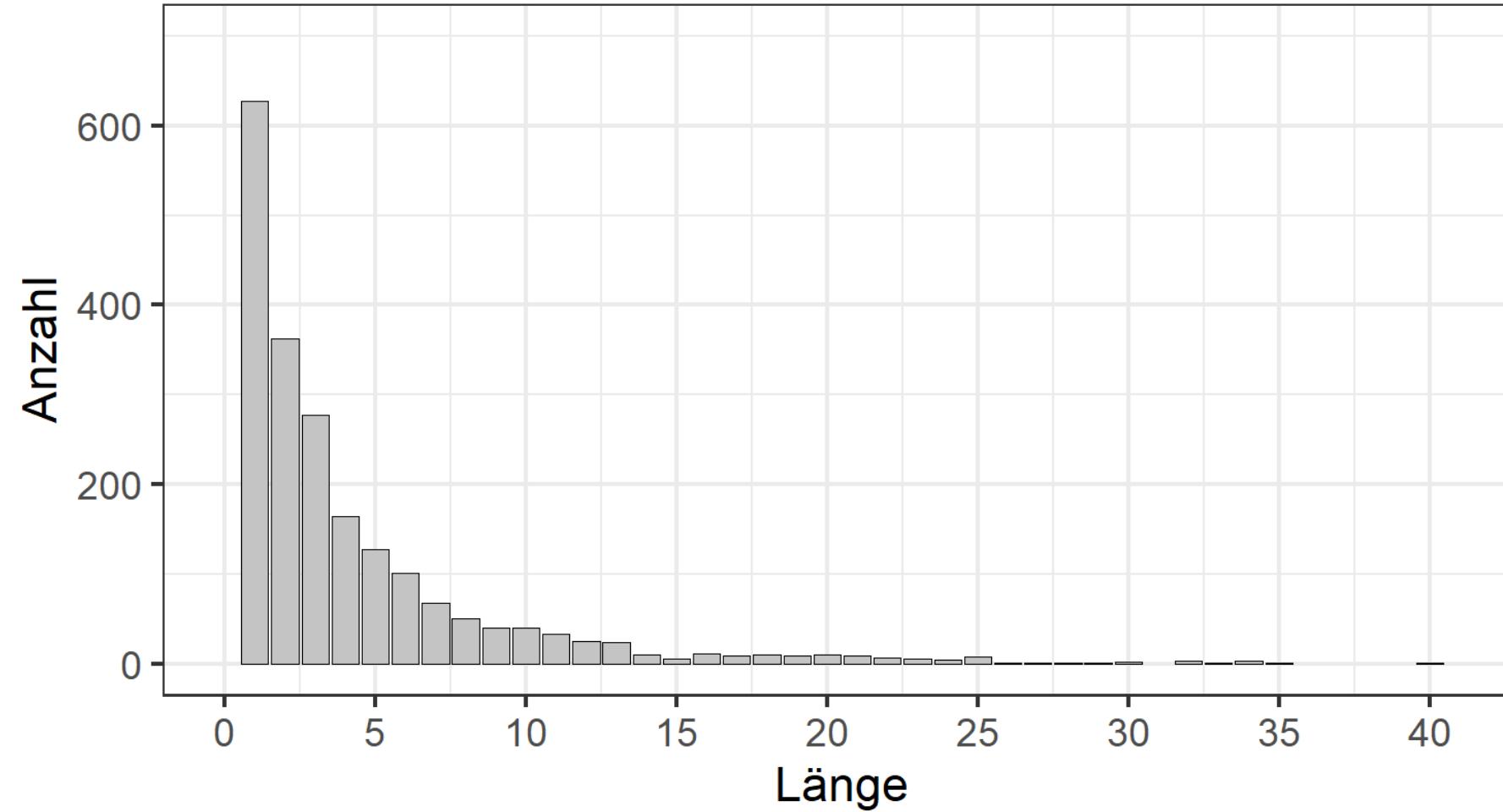






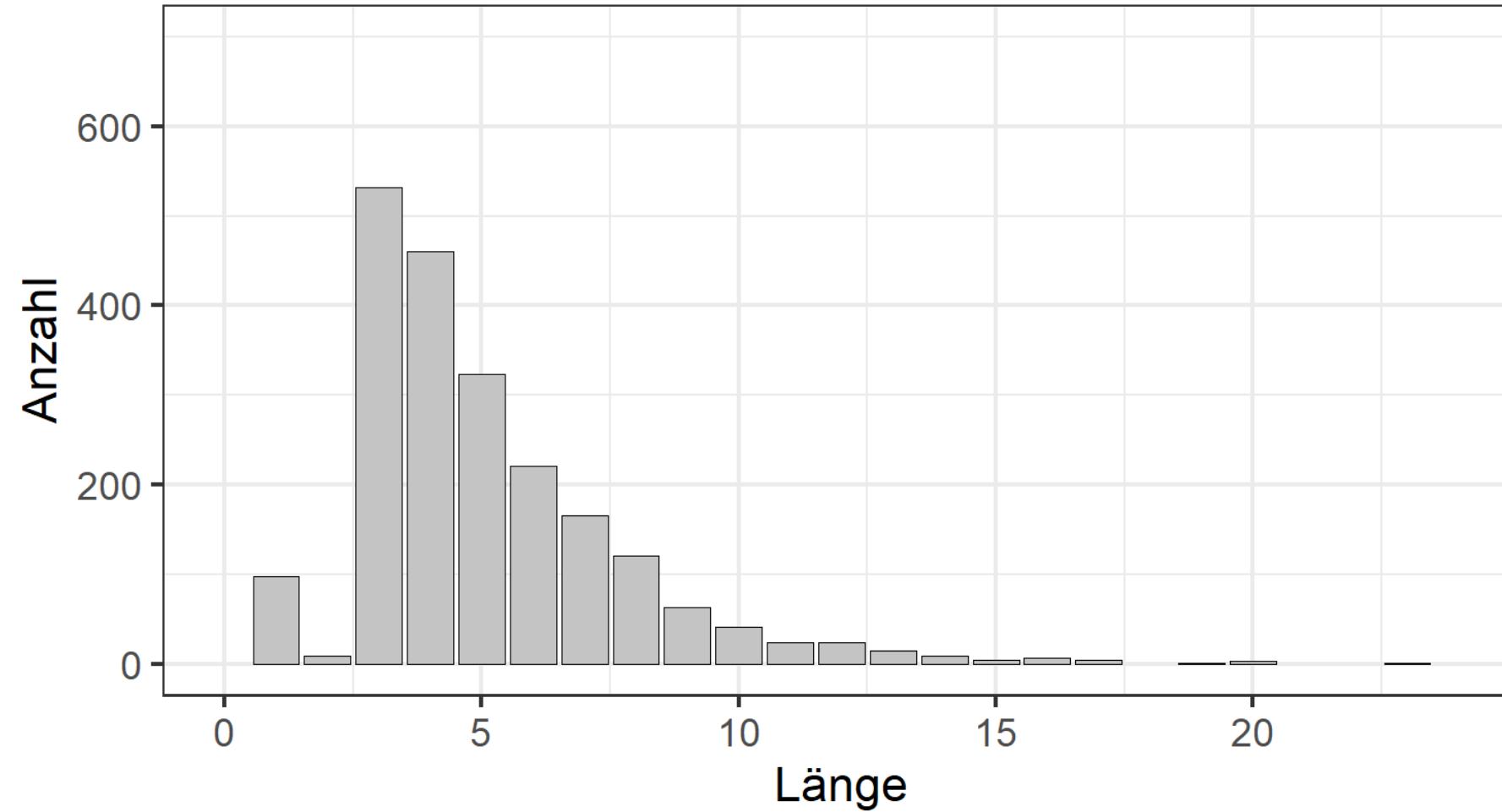


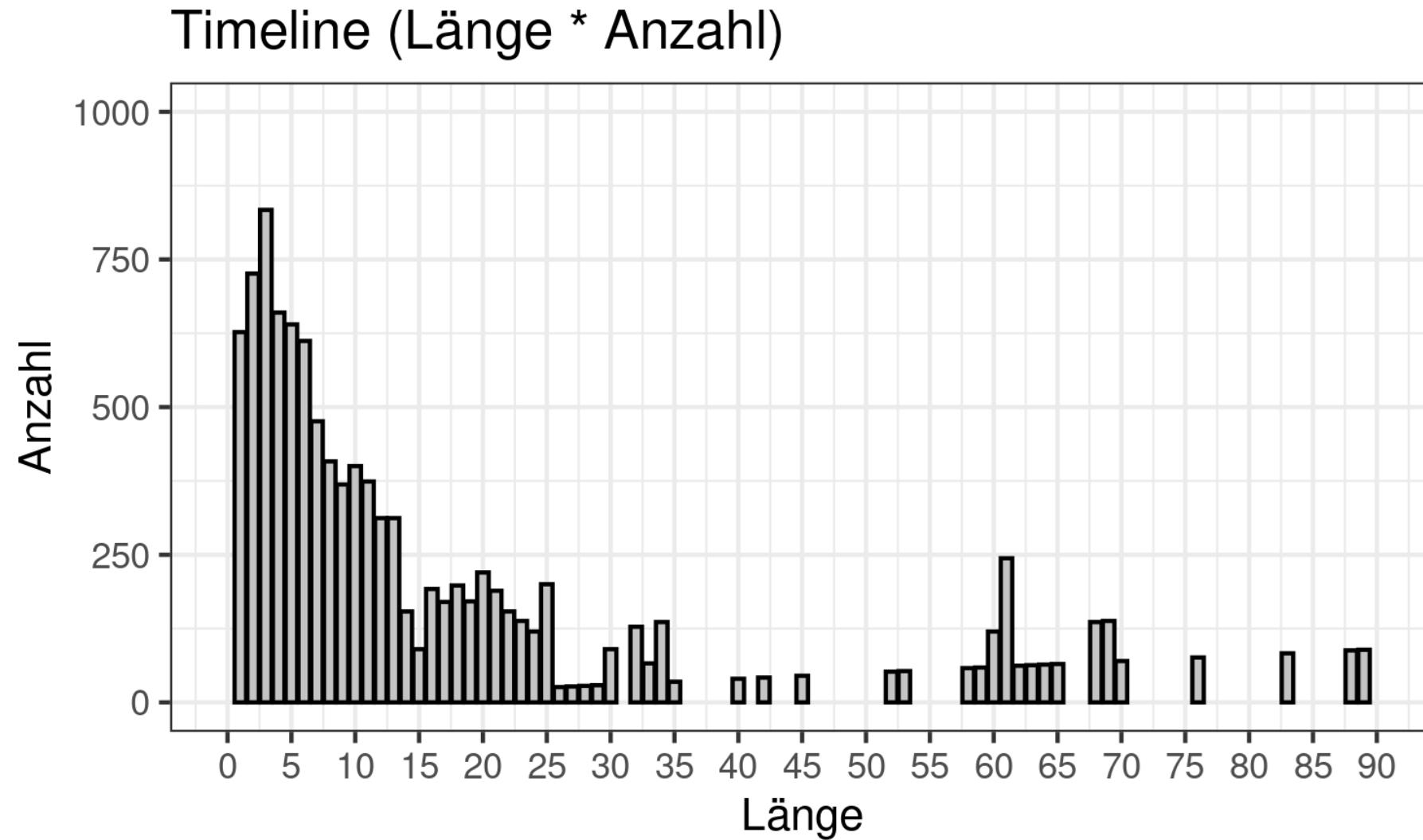
Länge der aufeinanderfolgenden, gleichen Cluster





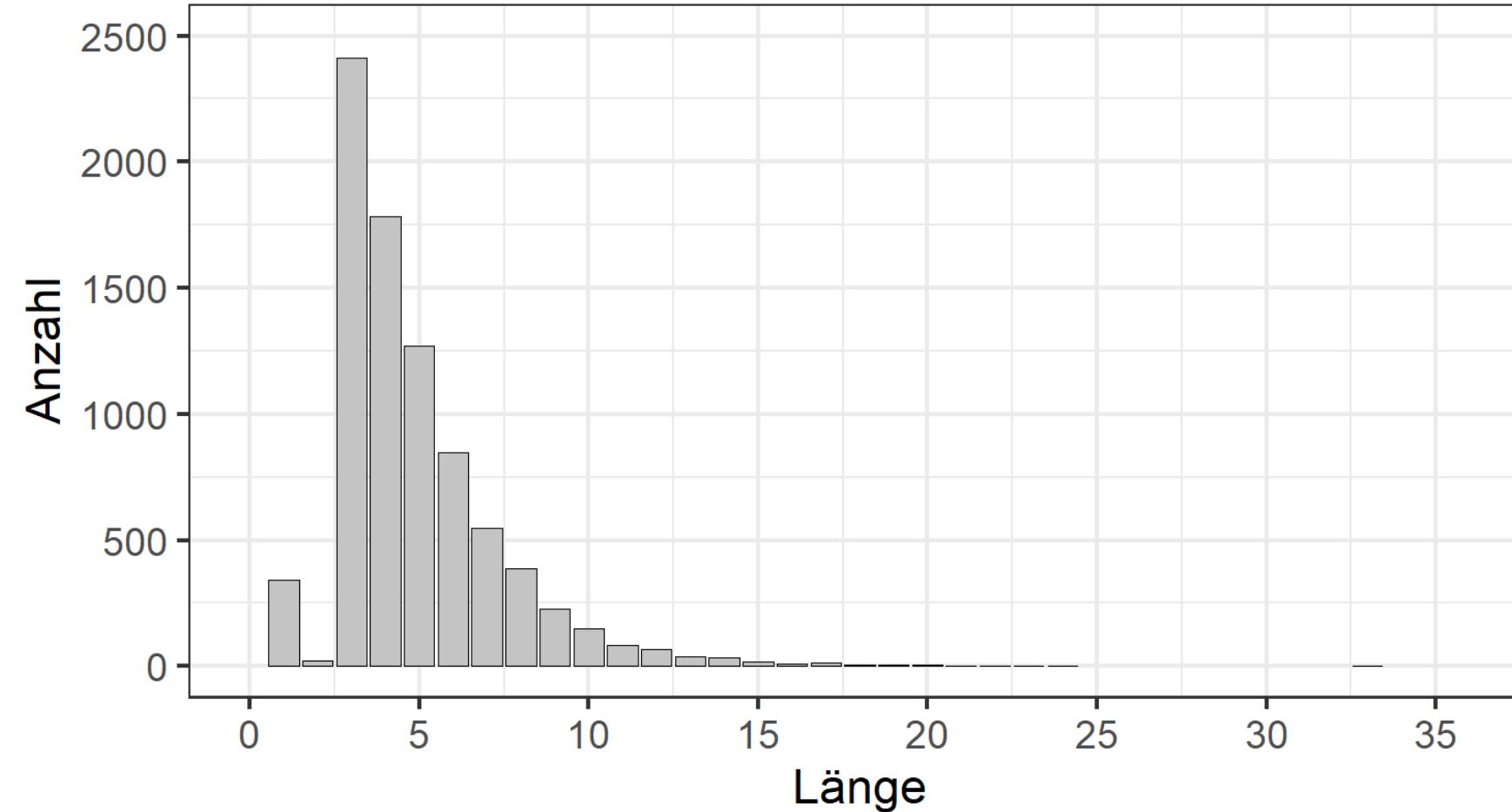
Länge der aufeinanderfolgenden, gleichen GWL





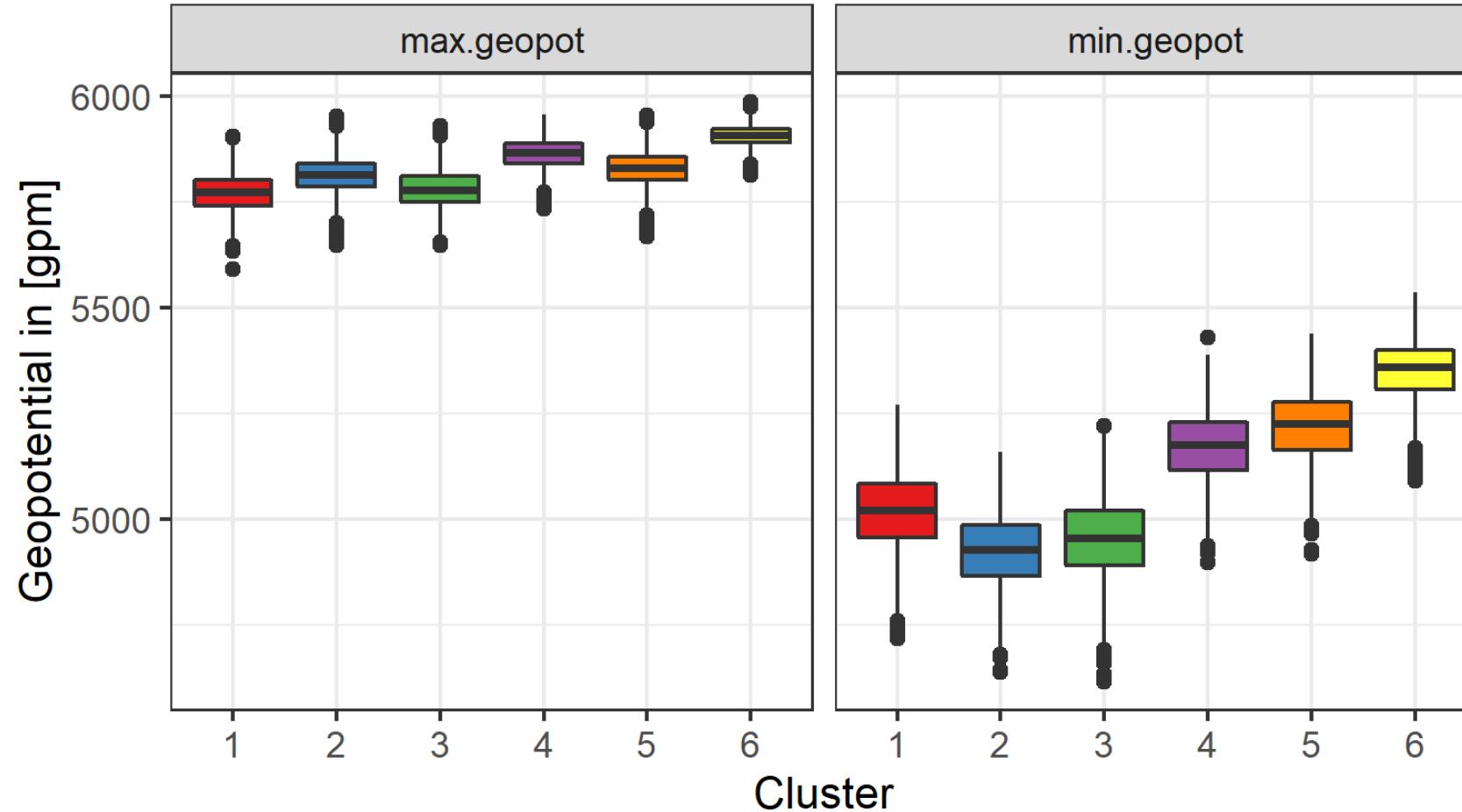


Länge der aufeinanderfolgenden, gleichen GWL





Minimaler und maximaler Luftdruck in jedem Cluster





Cluster Boosting

- insert Formula
- Pro Iteration sample ziehen
 - 3 mal zufällige 5-Jahresperioden

$X :=$ skalerter Datensatz

$W := (w_1, \dots, w_k)^T$ Gewichtsvektor

$c :=$ Schrittweite

$f :=$ Bewegungsfunktion

$$w^{(t)} = \underset{w_j, j=1, \dots, k}{\operatorname{argmax}} (f(\operatorname{cluster}(w_j^{(t-1)} X)))$$

wobei: $w_j^{(t-1)} = w^{(t-1)} + c_j$



Cluster Boosting

- Bewertungsfunktion f
 - Silhouettenkoeffizient
 - Timeline - Wert
 - Stabilität

insert formula

$$f(x) = \text{sil}(x) + \text{tl}(x) + \text{stab}(x)$$

wobei:

$$\text{sil}(x) = \frac{1}{k} \sum_{i=1}^k \text{Silhouettenkoeffizient}(x_i)$$
$$\text{tl}(x) = \frac{1}{k} \sum_{i=1}^k \text{TimelineWert}(x_i)$$

$$\begin{aligned} \text{stab}(x) = 1 &- \left(\frac{1}{2} \max_{\substack{i=1, \dots, k \\ j=1, \dots, k}} (|S(x_i) - S(x_j)|) \right. \\ &\left. + \frac{1}{2} \max_{\substack{i=1, \dots, k \\ j=1, \dots, k}} (|\tau(x_i) - \tau(x_j)|) \right) \end{aligned}$$



Cluster Boosting

- Probleme:
 - fachliche Sinnhaftigkeit der Gewichte
 - Instabilität des Algorithmus
 - sehr teuer



Filter-Ansatz Clusteralgorithmus

Pro Startpunkt:

Solang keine neuen Punkte mehr gefunden werden:

Für alle Punkte im Cluster

Finde alle Punkte, die $< \text{eps}$ entfernt sind

Füge sie dem Cluster hinzu

$\text{eps} = \text{eps} / x$



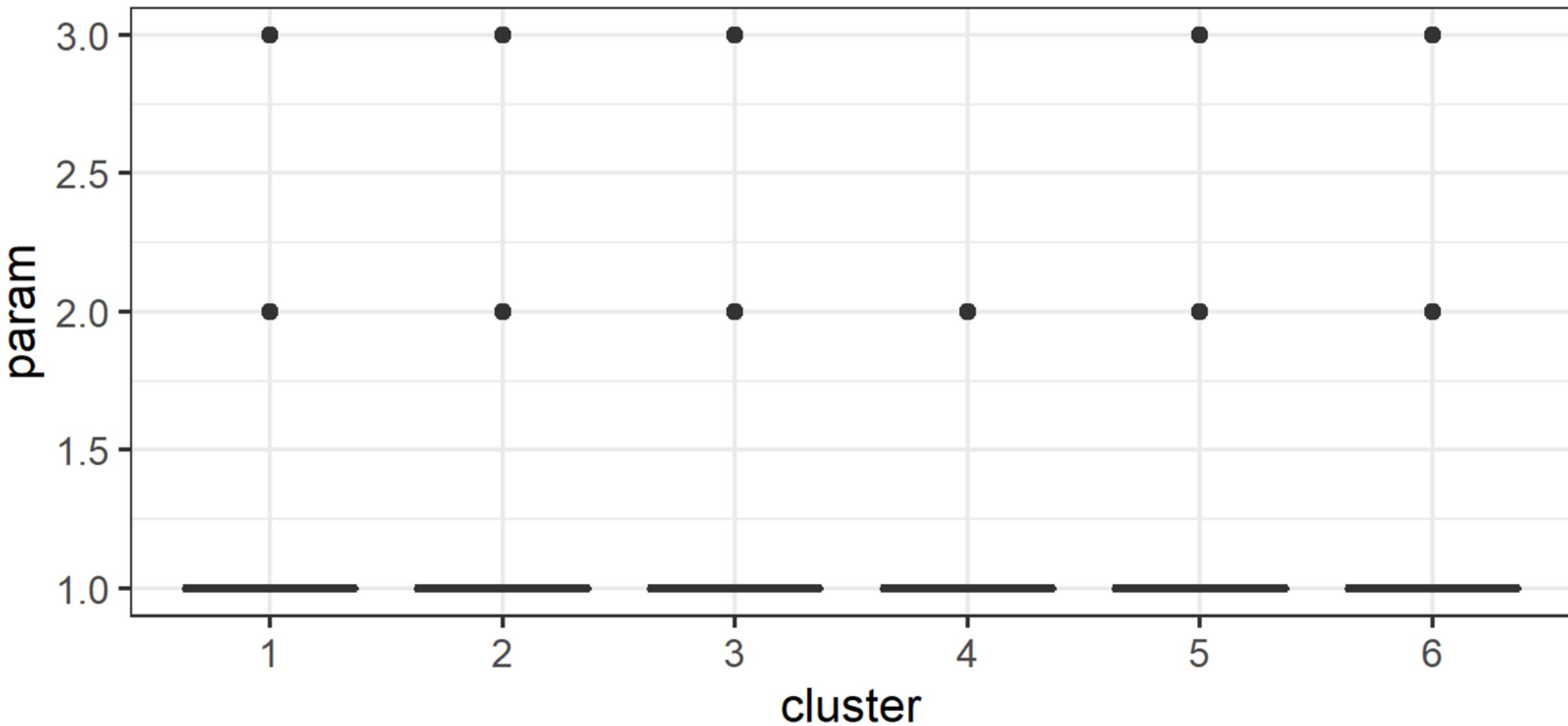
Beispiele

	date	cluster	gwl
1	1971-11-21	4	TM
2	1971-11-22	5	TM
3	1971-11-23	4	TM
4	1971-11-24	3	TM

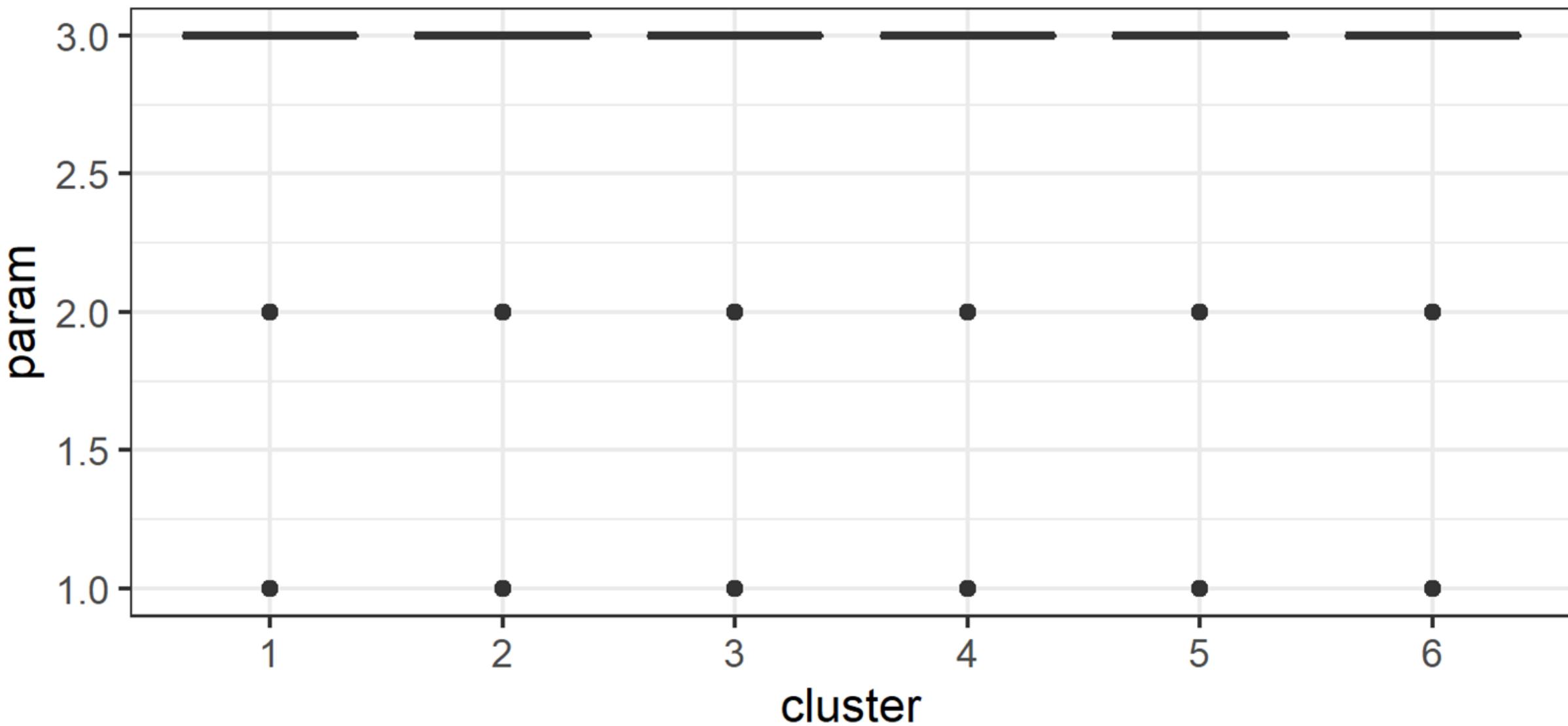


Zum Teil aber auch sehr unterschiedlich

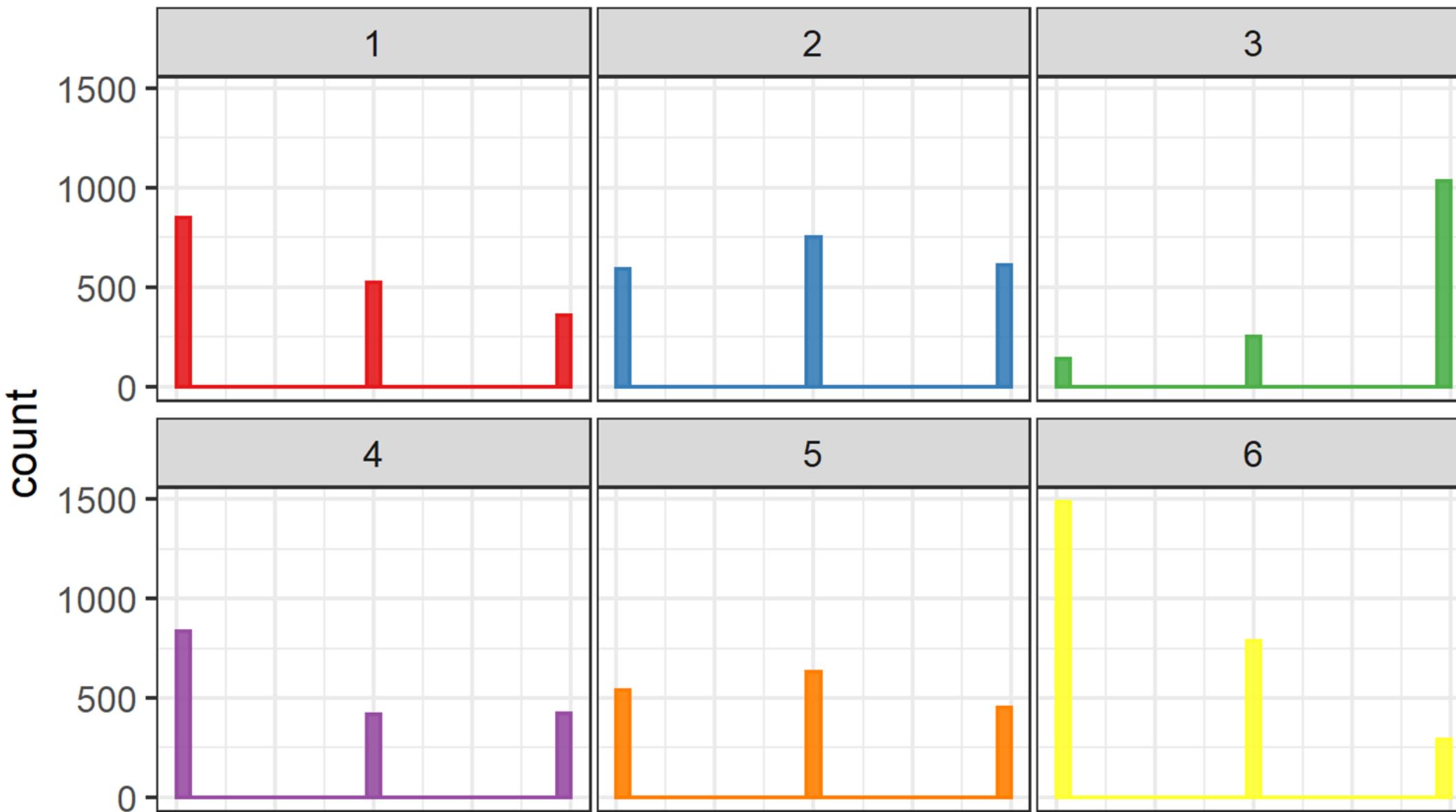
Verteilung maxGeopot.verID in jedem Cluster



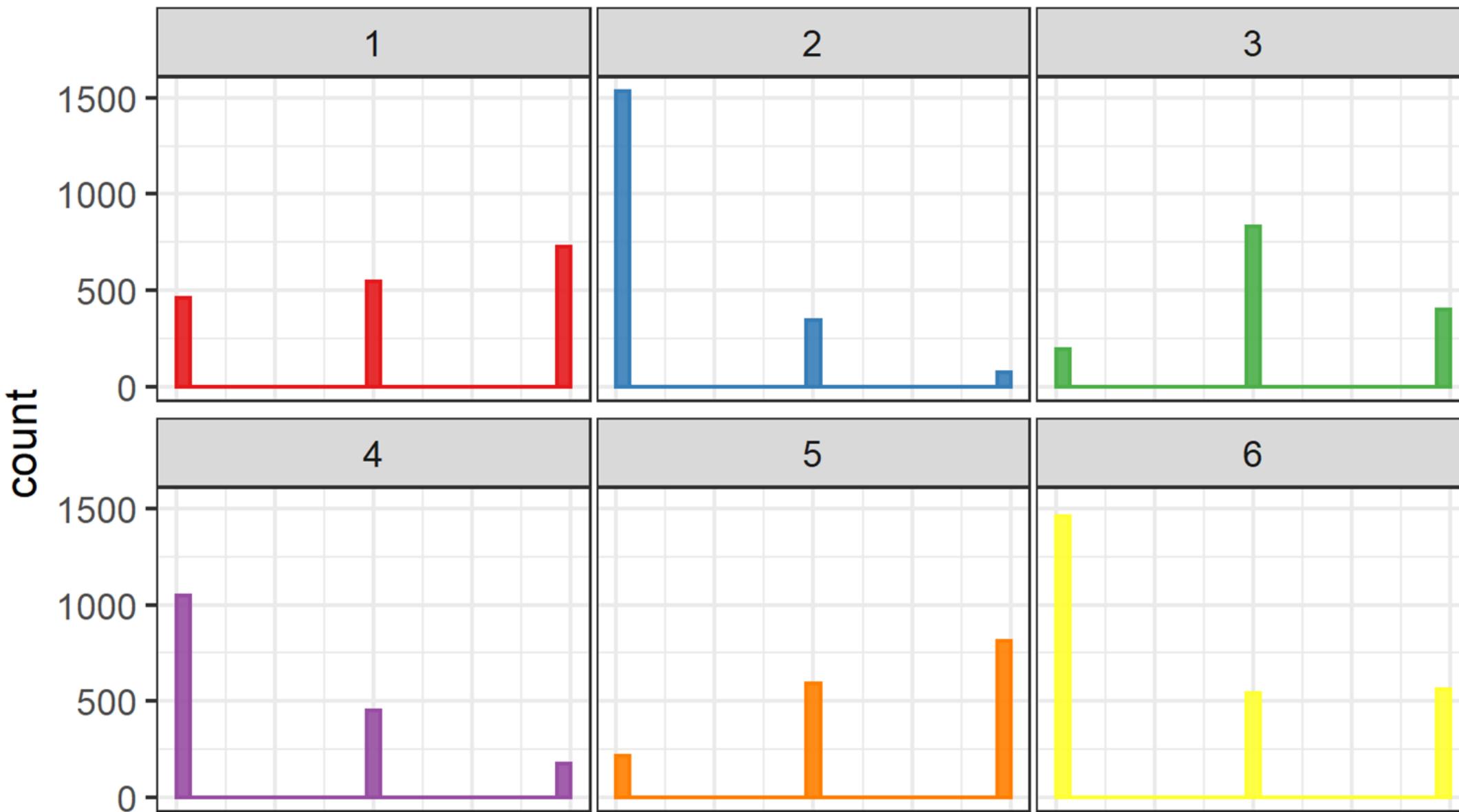
Verteilung minGeopot.verID in jedem Cluster



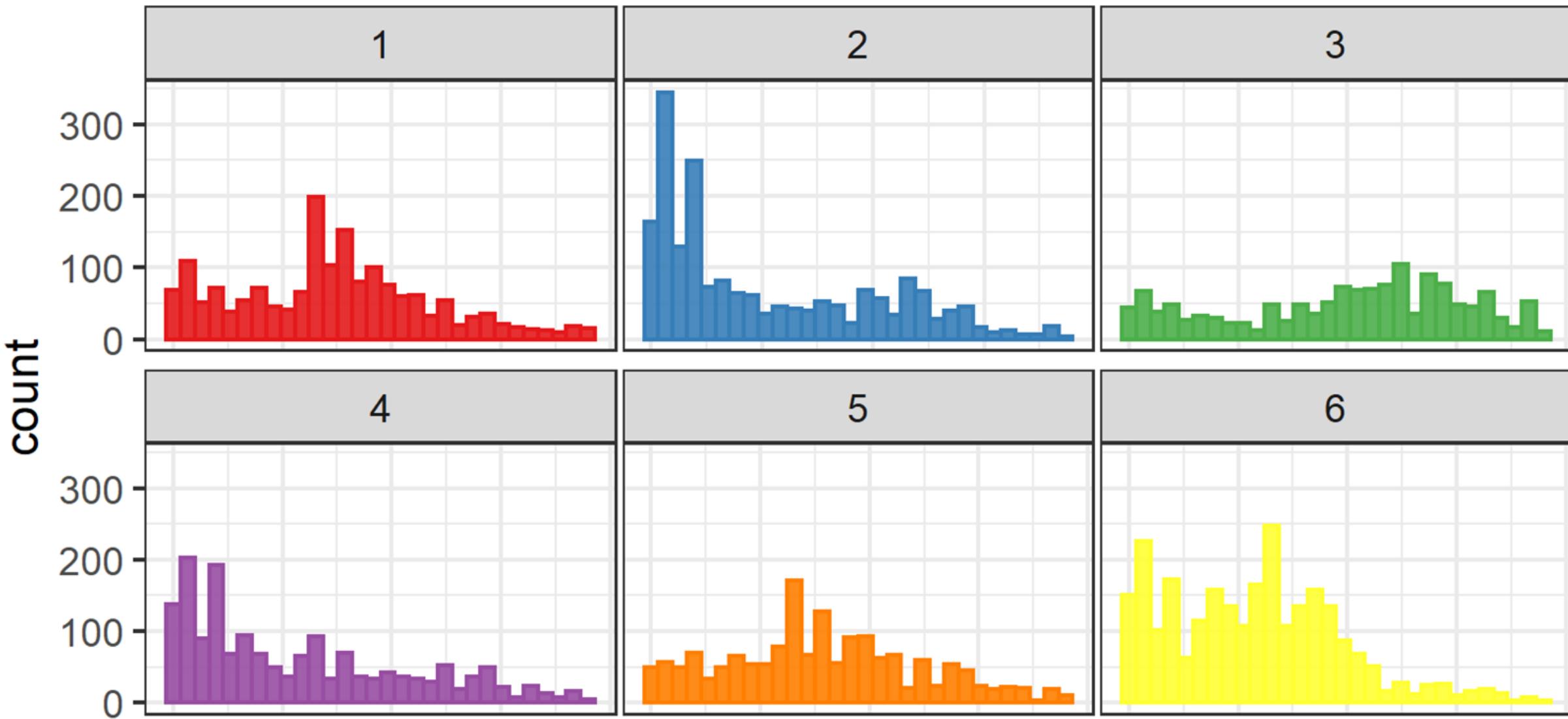
Verteilung maxMslp.horlD in jedem Cluster



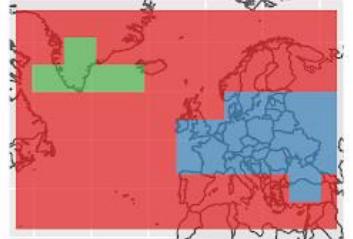
Verteilung maxMslp.verID in jedem Cluster



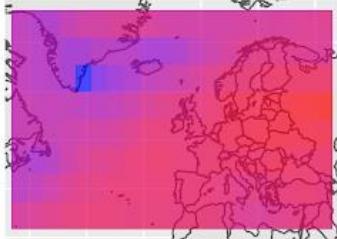
Verteilung euclidean.maxDiff in jedem Cluster



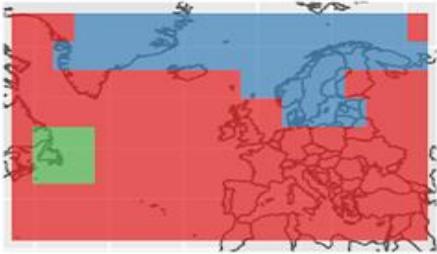
cluster



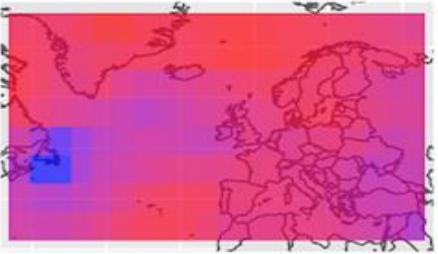
avg_mslp



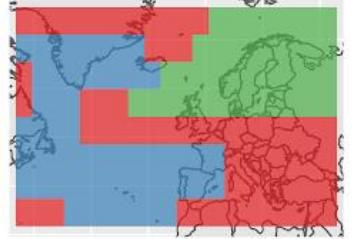
cluster



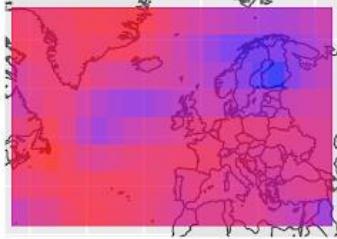
avg_mslp



cluster



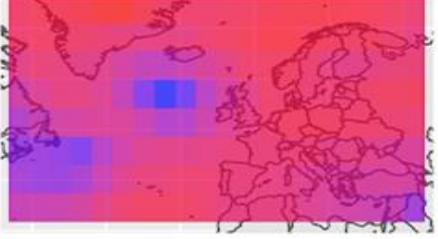
avg_mslp



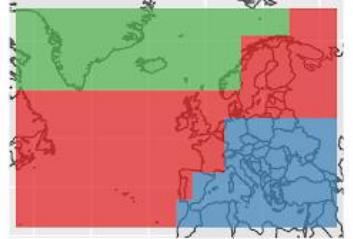
cluster



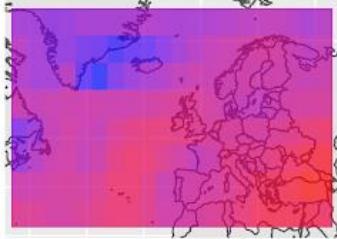
avg_mslp



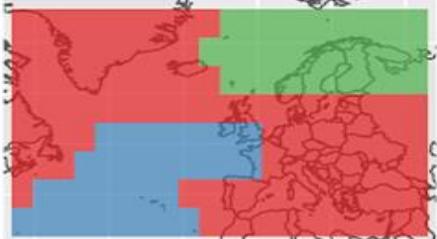
cluster



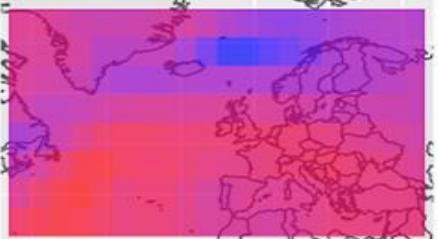
avg_mslp



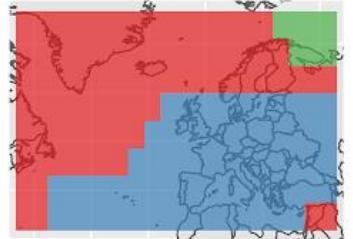
cluster



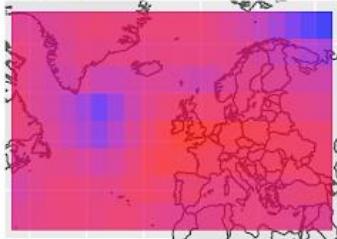
avg_mslp



cluster



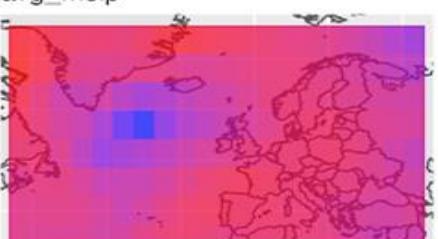
avg_mslp



cluster

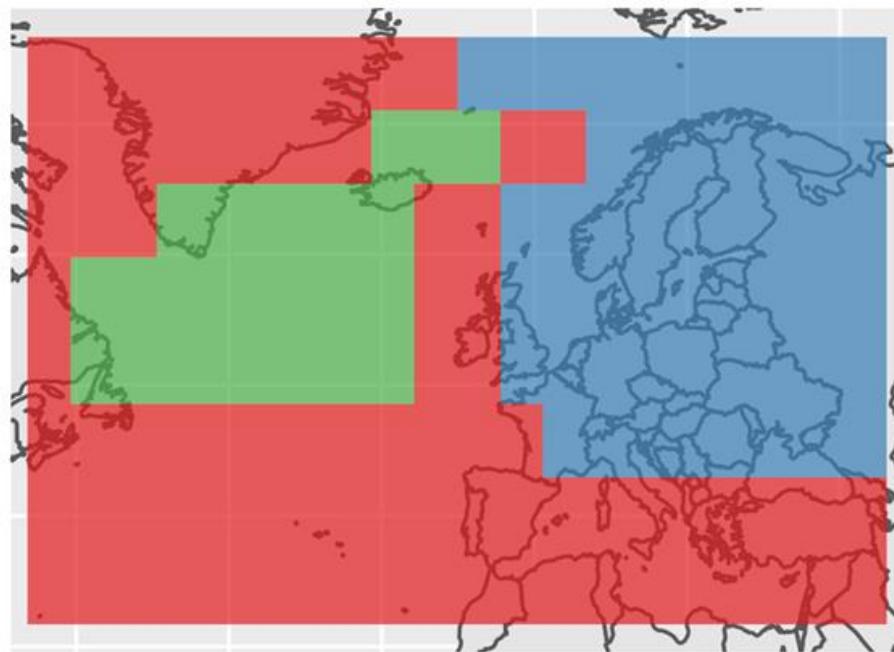


avg_mslp

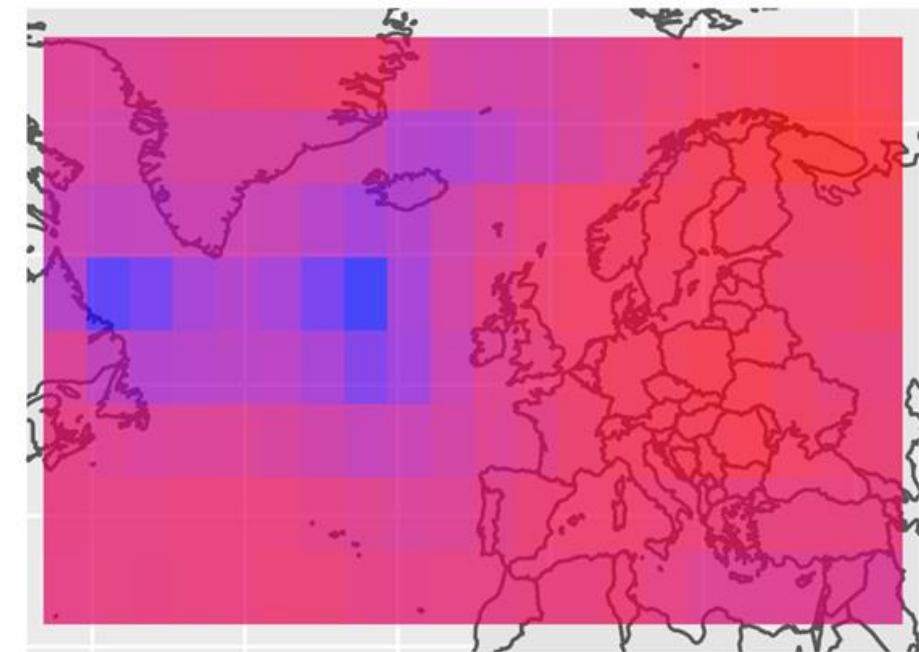


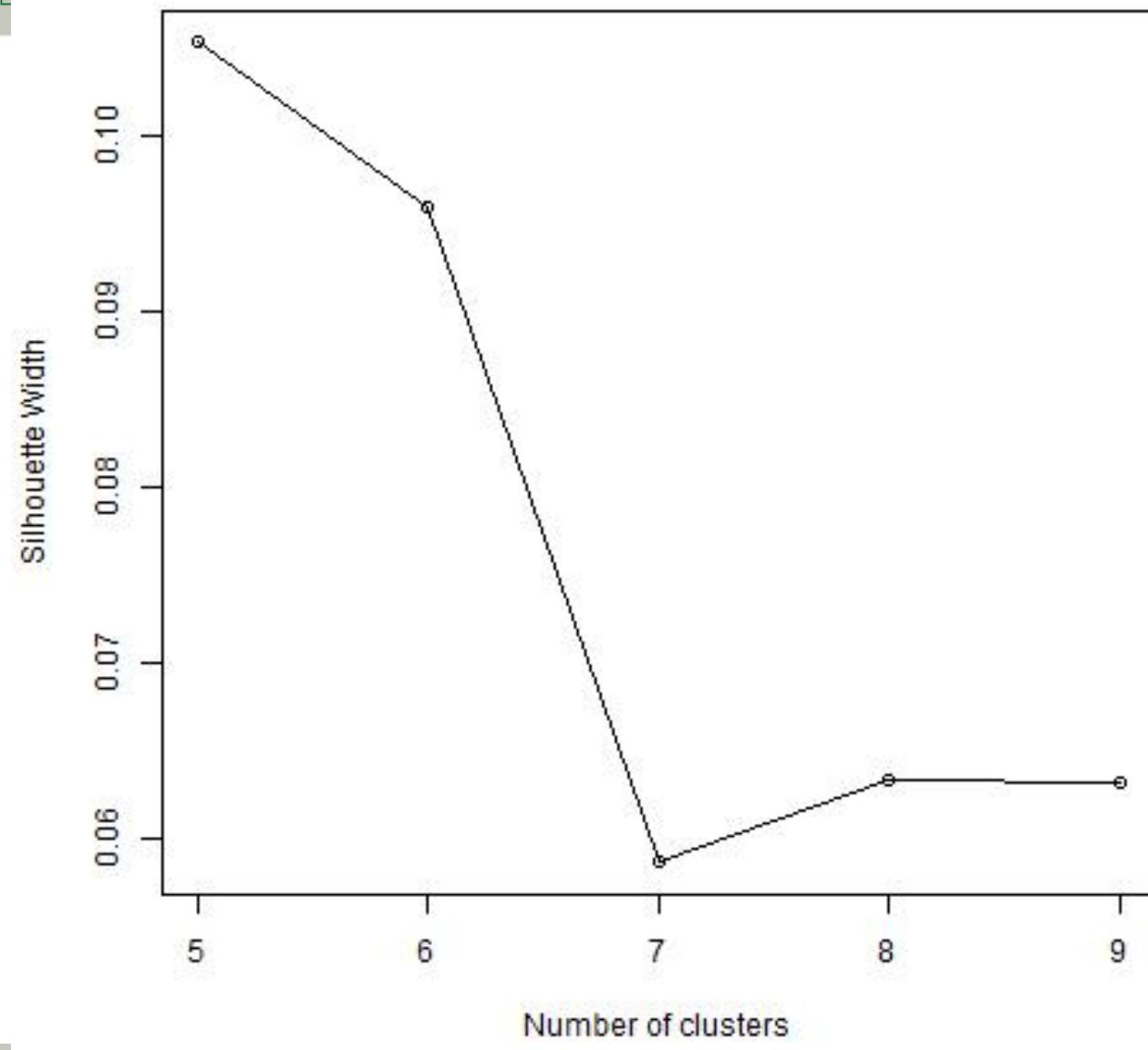
The GWL on 2006-10-10 is HM

cluster



avg_mslp

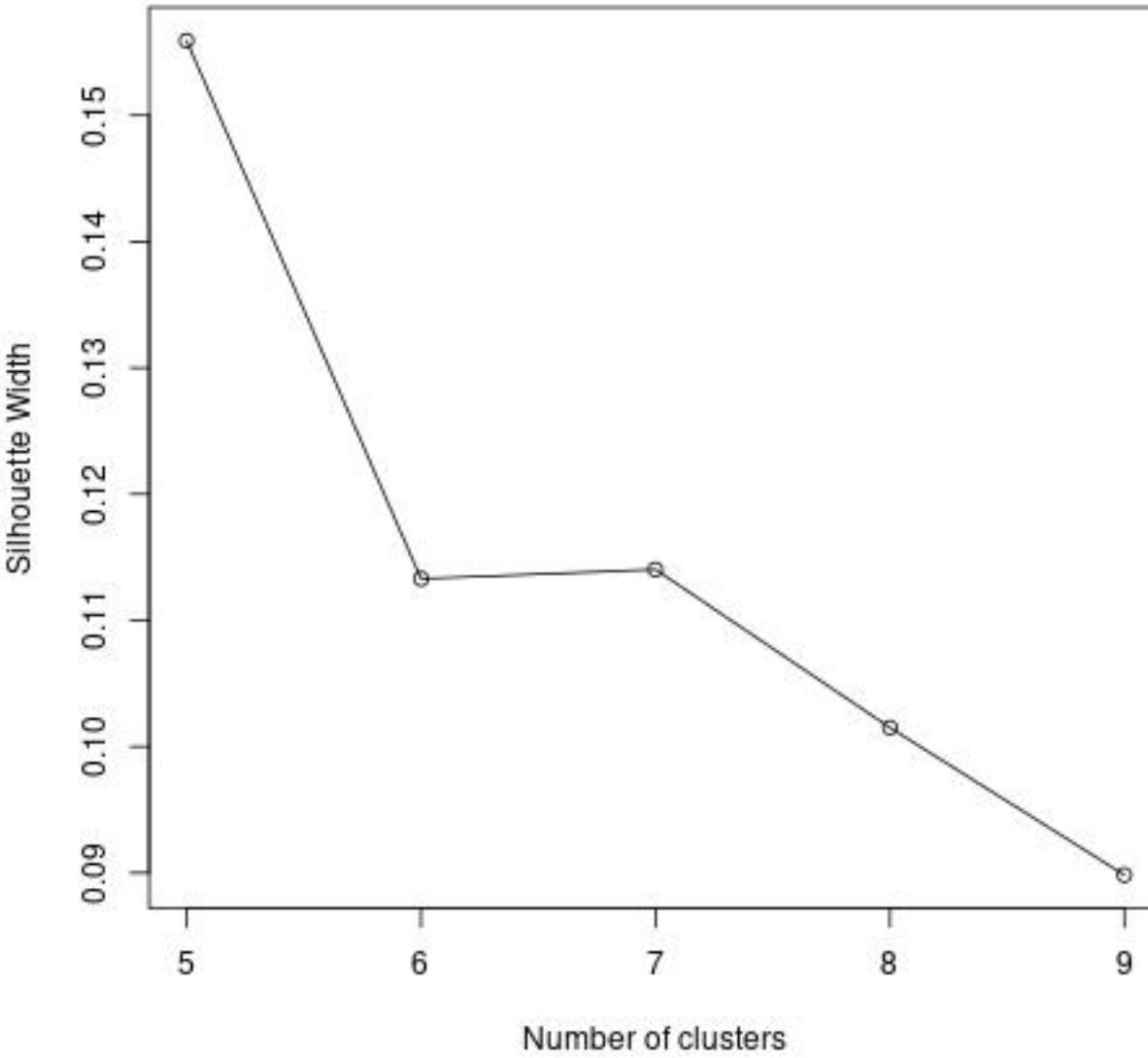


PAM manhattan

Original wide
Datensatz



PAM manhattan



Plus filter 1984

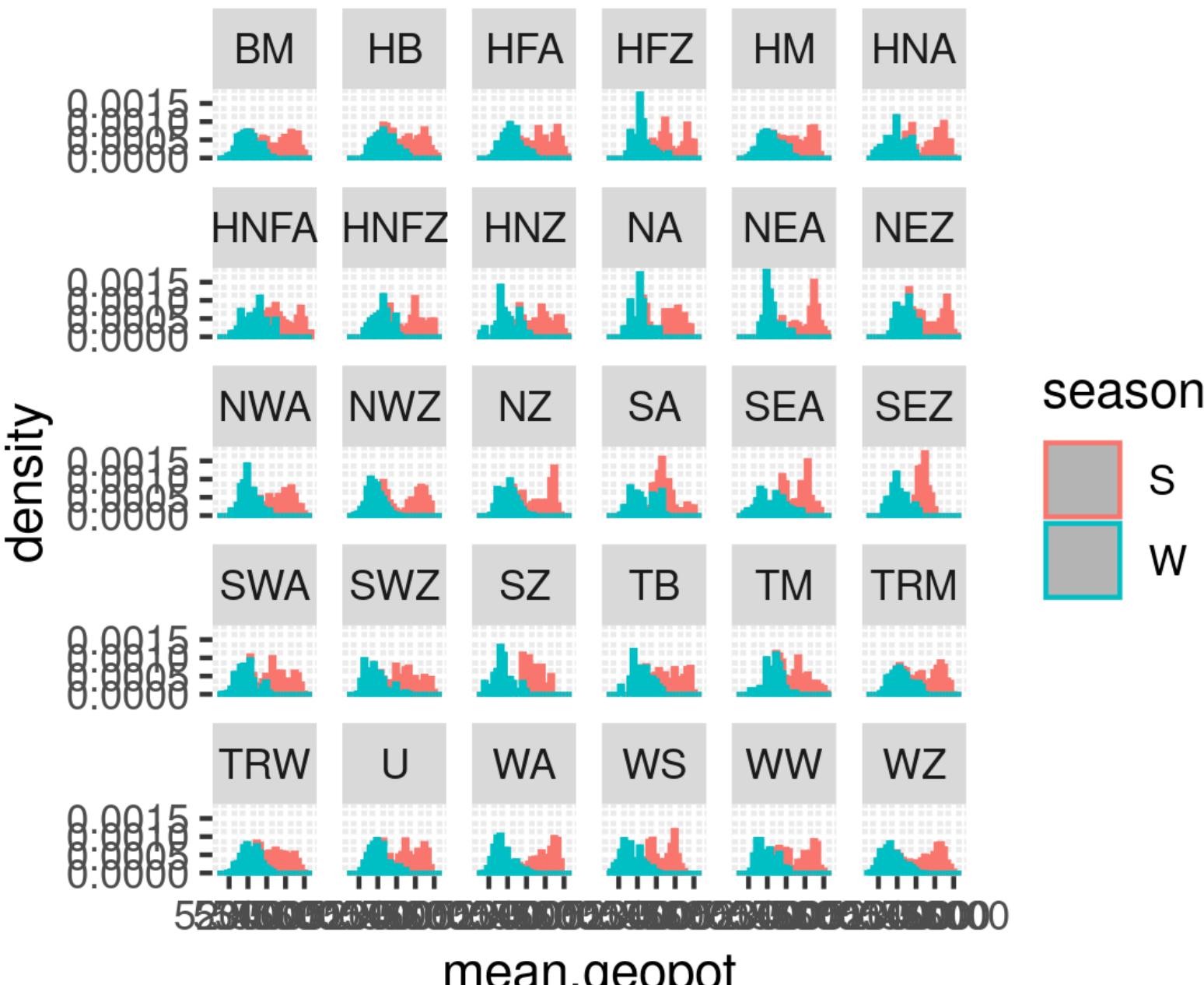
MUNICH NETWORK MANAGEMENT TEAM



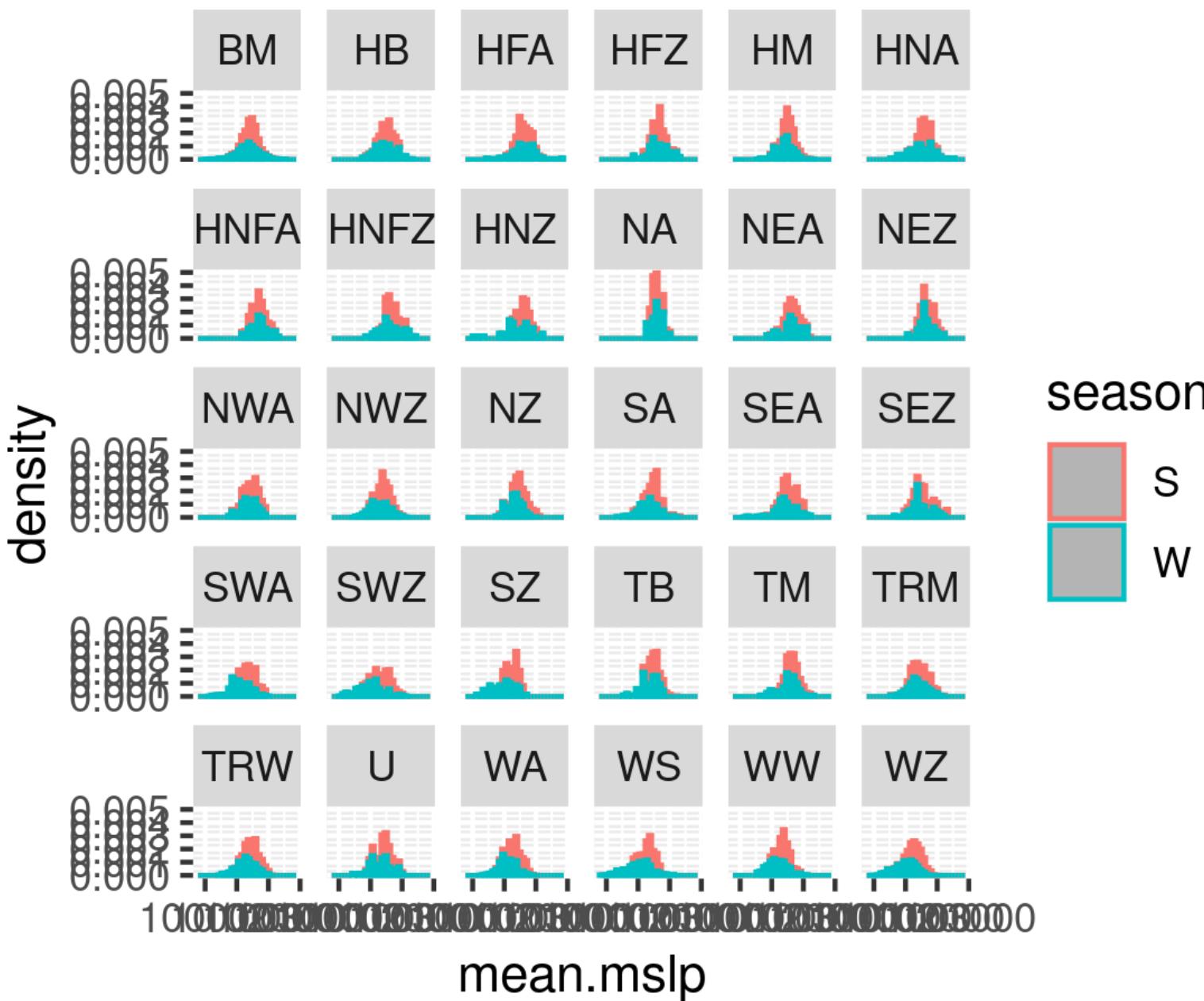
Verteilung Änderung über den Tag pro GWL pro Saison



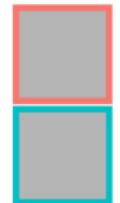
Verteilung mean.geopot pro GWL



Verteilung mean.mslp pro GWL



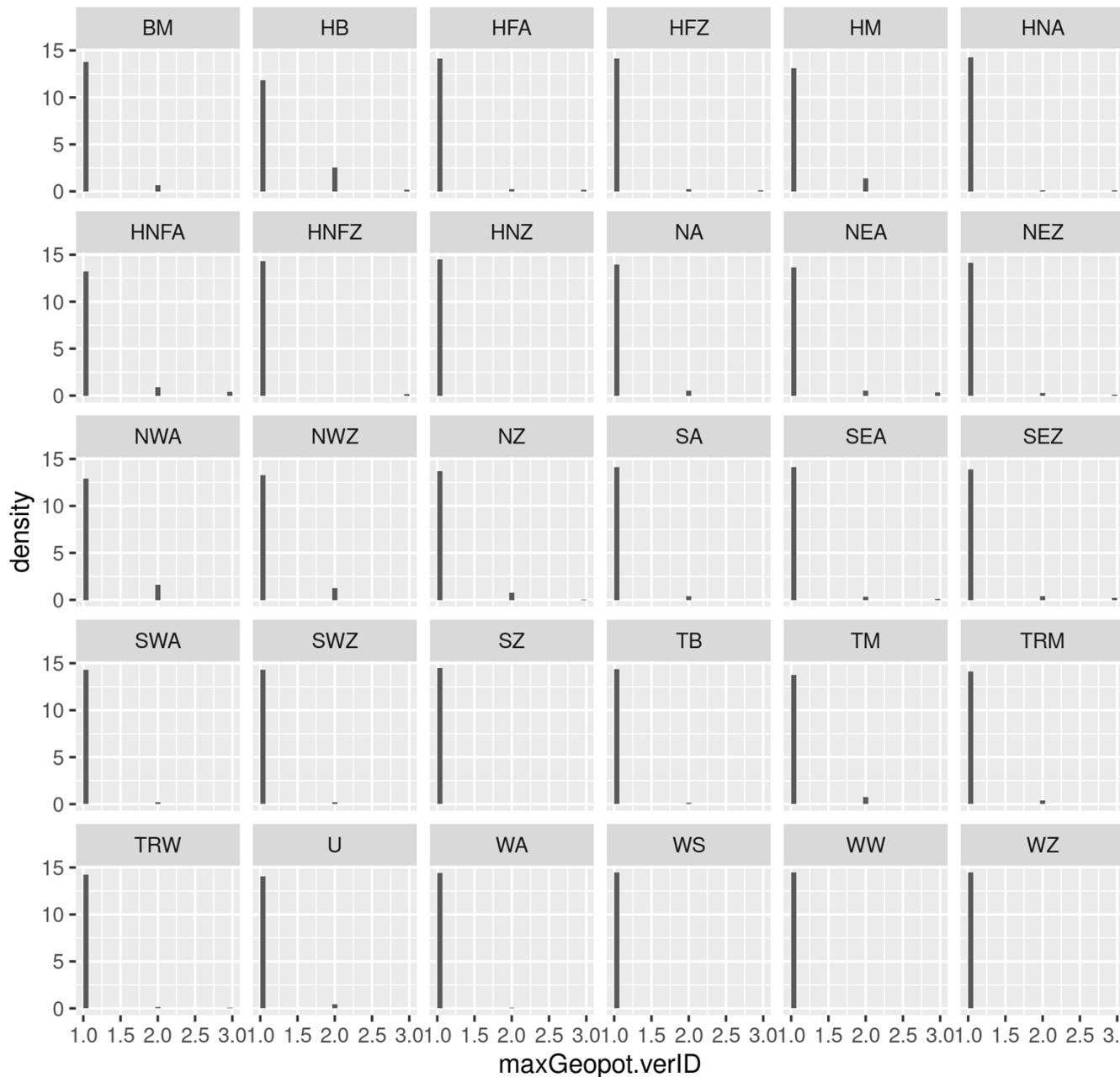
season



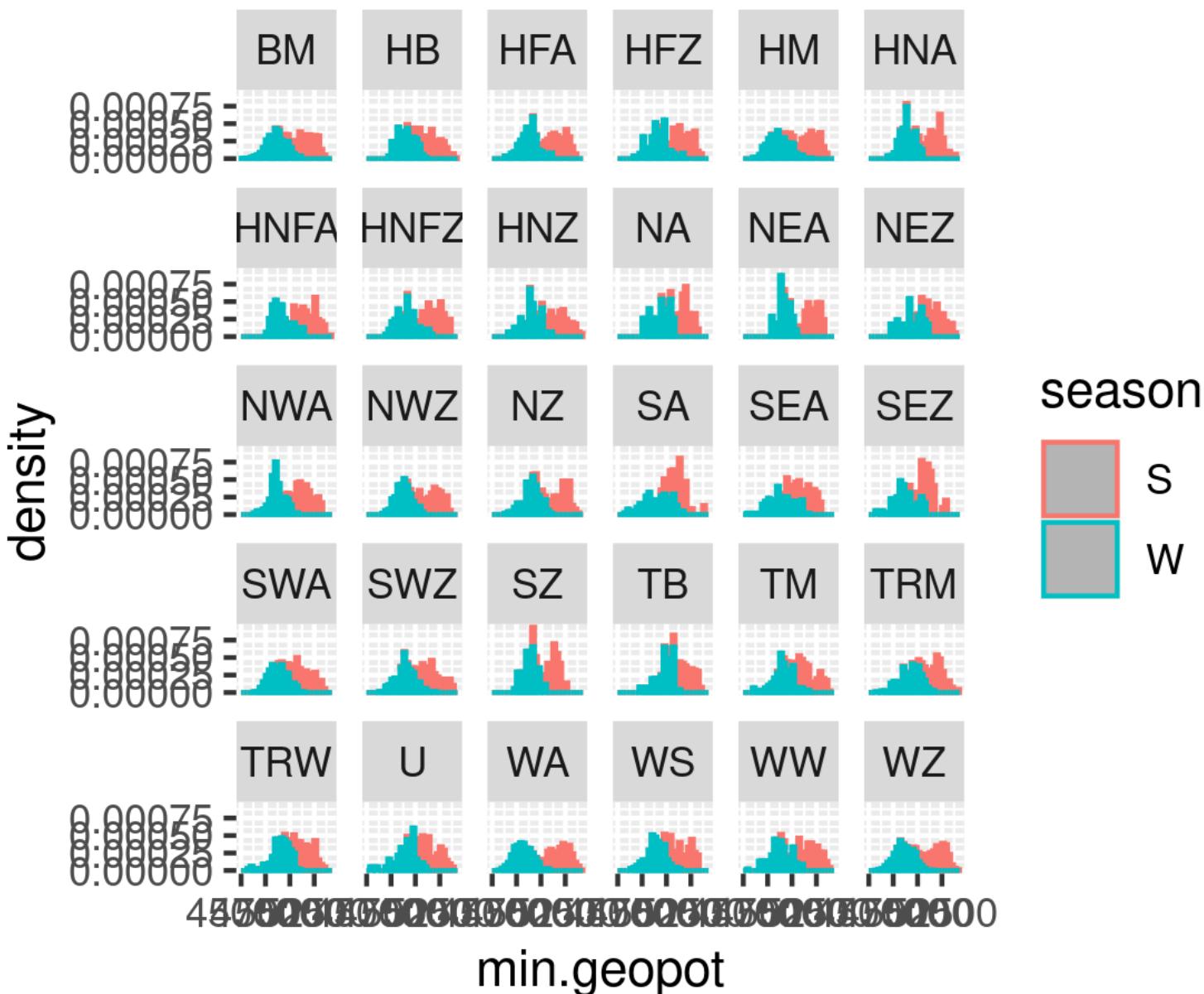
S

W

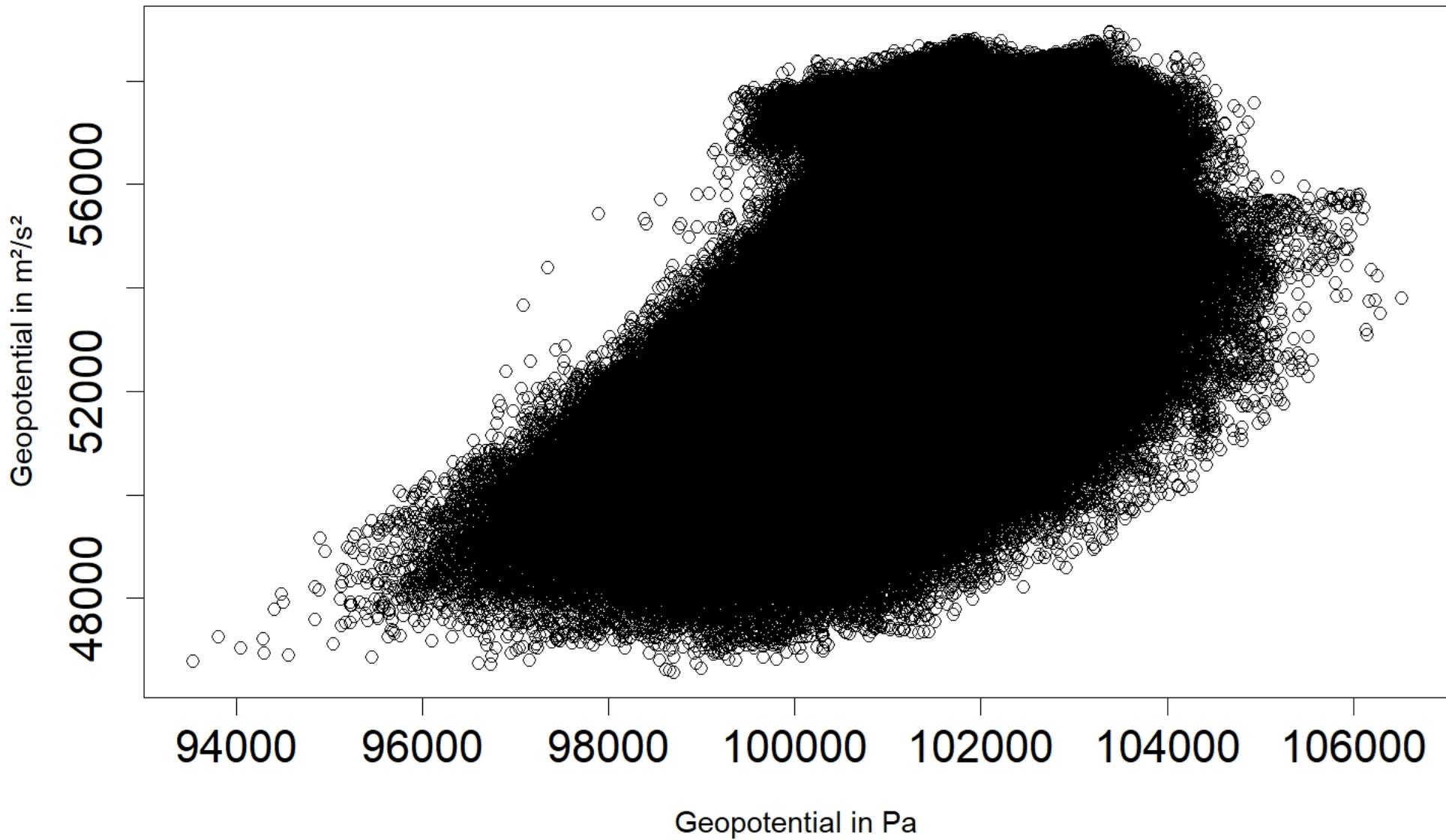
Verteilung maxGeopot.verID pro GWL



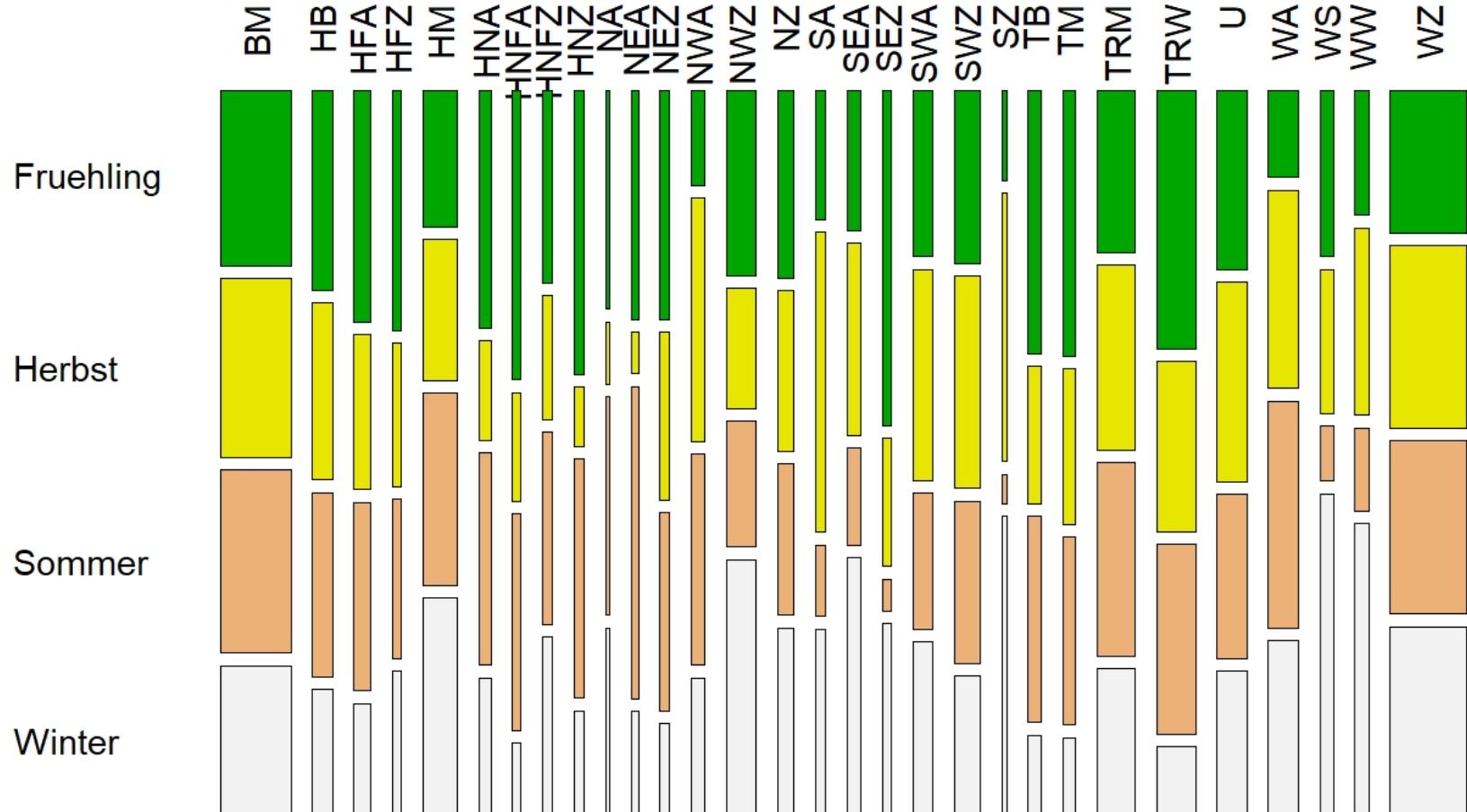
Verteilung min.geopot pro GWL



Korrelation zwischen Luftdruck und Geopotential (2006 bis 2010)



Anzahl der GWLs in Abhängigkeit der Jahreszeiten



Tageswertvergleiche der Jahre 2006 bis 2010

