



universität
wien

DIPLOMARBEIT

Titel der Diplomarbeit

Distanzmaße beim statistischen Downscaling
*Evaluierung ausgewählter Distanzmaße mittels der
Analog-Methode*

angestrebter akademischer Grad
Magistra der Naturwissenschaften (Mag.rer.nat)

- Verfasserin: Elsa Maria Dos Santos Cardoso
- Matrikelnummer: 9803248
- Studienrichtung (lt. Studienblatt): A 415 Meteorologie
- Betreuer: ao. Univ. Prof. Dr. Georg Skoda

Wien, Juni 2008

*Those among us who are unwilling to expose their ideas to the
hazard of refutation do not take part in the scientific game*
Sir Karl Popper (1902-1994)

Zusammenfassung

Um Aussagen über die Entwicklung von Klimavariablen zu ermöglichen ist es notwendig, eine sinnvolle Projektion von großskaligen Klimainformationen auf kleinere Skalen umzusetzen. Einen möglichen Zugang bietet das Statistische Downscaling, welches sich einerseits als kostengünstige Möglichkeit erweist im Vergleich zu anderen komplexeren Verfahren. Andererseits macht das Verfahren von den empirischen Zusammenhängen zwischen einem meteorologischen Parameter auf der groben Skala und den Beobachtungen eines anderen Parameters auf der lokalen Skala Gebrauch. In der vorliegenden Arbeit wird die *Analog-Methode* verwendet, um den Einfluss von Distanzmaßen auf der Suche nach dem „ähnlichsten Tag“ zu evaluieren. Distanzmaße werden oft in der Mustererkennung eingesetzt, um Ähnlichkeiten zwischen Objekten zu bestimmen. Die *Analog-Methode* wurde erstmals von Storch(1999) [61] mit anderen statistischen Downscaling Verfahren verglichen. Um dies umzusetzen werden die Startclim Stationen als lokale Beobachtungen und die ERA40 Daten als grob aufgelöste Felder herangezogen, welche einer Hauptkomponentenanalyse (PCA) unterworfen werden. Die hier eingesetzten Distanzmaße sind das Euklidische Distanzmaß, die *Gewichtung der Euklidischen Distanz mit den Eigenwerten*, das City-Block-Distanzmaß und anschliessend wird eine Version der Mahalanobis Distanz eingesetzt. Letztere wurde von Yambor [59] auf ein PCA-basierendes Verfahren angewandt. Es zeigt sich, dass die *Analog-Methode* auf saisonaler Basis zufriedenstellend die statistischen Eigenschaften des Niederschlags erkennt.

Abstract

In order to predict the behavior of regional climate variables, an appropriate projection of coarse scale information to finer scales is required. In the past, several statistically motivated methods for downscaling climate information to scales suitable for climate impact studies have been proposed. Compared to other approaches, statistical downscaling methods are easy to implement and offer low computational complexity. Additionally, statistical downscaling methods exploit relationships between large and fine scale meteorologic parameters, identified in the observed climate. As well established statistical downscaling algorithm is the *Analog Method*, which has been shown to compare favorably with other approaches [61]. The present work is based on the *Analog Method*, evaluating the impact of various distance measures on the performance of queries for the "most similar day". This idea has been adopted from the field of pattern recognition, where the application of specialized distance measures for assessing pattern similarity is a common technique (see e.g. [10] und [59]). In this thesis, we combined large scale ERA-40 data with local Startclim station data and performed a subsequent Principal Component Analysis (PCA) in order to reduce the dimensionality of the input data by "filtering" out irrelevant (low-variance) information. The distance measures utilized in our tests are: (a) Euclidean distance, (b) eigenvalue-weighted Euclidean distance [61], (c) City-block, and (d) a Mahalanobis distance-based measure, applied in the context of PCA-based face recognition by [59].

Evaluations of the Analog Method were carried out on a daily, as well as on a seasonal Basis. In our experiments we found that the Analog Method was able to capture the statistical properties of seasonal precipitation satisfactorily and that the eigenvalue-weighted Euclidean Distance performed best among the tested measures.

Inhaltsverzeichnis

1	Motivation und Problemstellung	11
2	Einleitung	13
2.1	Probleme der Klimamodellierung	14
2.1.1	Skaligkeit atmosphärischer Phänomene	14
2.2	Downscalingansätze	16
2.3	Forschungsstand	21
3	Die Variabilität des Klimas	31
3.1	Klimaschwankungen und Klimaänderung	31
3.2	Klimavorhersage	34
4	Niederschlag in Österreich	36
4.1	Die Grosswetterlagen in Mitteleuropa	36
4.2	Niederschlag	41
4.2.1	Vb-Wetterlage	41
4.2.2	Niederschlagsverhalten und Grosswetterlage	42
4.2.3	Niederschlagsarten	44
4.3	Lokale Zeitreihen des Niederschlags	46
4.4	Reanalyse-Datensätze	47
4.4.1	Mittelung der Reanalyse Felder	52
5	Statistisches Downscaling	57
5.1	Methodik	57
5.1.1	Die Hauptkomponentenanalyse	57
5.1.2	Die Distanzmaße	67
5.1.3	Distanzmaße bei quantitativen Attributen	70
5.2	<i>Die Analog-Methode</i>	83
6	Verifikation	87
6.1	Statistische Methoden	87

6.1.1	Prüfmaße für kontinuierliche Variablen	89
6.1.2	Prüfmaße für kategorische Variable	91
6.1.3	Genauigkeitsmaße	91
6.1.4	Qualitätsmaße	93
6.1.5	Statistische Tests	95
7	Ergebnisse	99
7.1	Downscaling auf Tagesbasis	101
7.2	Downscaling auf saisonaler Basis	109
8	Diskussion und Ausblick	127
8.1	Diskussion der Ergebnisse	127
8.2	Ausblick	129
A	Tabellen	130
B	Abbildungen	146
C	Danksagung	150
D	Persönliche Daten	151

Abbildungsverzeichnis

2.1	Größenordnungen und Zeitskalen einiger atmosphärischer Vorgänge.	14
2.2	Horizontale Diskretisierung des Nordatlantiks.	16
2.3	Mittlere Temperaturverteilung im Januar im Mackenzie Becken in Nordkanada.	17
2.4	Statistischer und Dynamisches Downscaling.	18
2.5	Vergleich von statistischen Downscaling Verfahren.	21
2.6	Schematische Struktur eines Zirkulationsmodells.	22
2.7	Skaligkeit	24
3.1	Klimatologische Messreihe.	32
3.2	Schematische Darstellung der Klimavariationstypen	33
3.3	Klimadimensionen	34
4.1	Standardabweichung des mittleren Druckes in Meereshöhe für Europa.	38
4.2	WZ-Lage und WS-Lage	39
4.3	SWZ-lage	40
4.4	NWZ-Lage und NZ-Lage	42
4.5	TRM-Lage (Vb-Lage)	43
4.6	Homogene Niederschlagsregionen	44
4.7	Orographische Niederschläge	45
4.8	Frontale Niederschläge	46
4.9	Geographischer Ausschnitt der ERA40 Datenfelder.	47
4.10	Zonales Mittel der spezifischen Feuchte für das Jahr, die Wintermonate (DJF) und die Sommermonate (JJA) in Abhängigkeit der Meridionalkomponente und des Drucks (in DB).	50
4.11	Sättigungsdampfdruck als Funktion der Temperatur.	51
4.12	Saisonale Mittelung (1958-2000) des mittleren Luftdrucks in Meereshöhe und der spezifischen Feuchte in 700 hPa für Winter (DJF) und Frühling (MAM).	53

4.13	Saisonale Mittelung (1958-2000) des mittleren Luftdrucks in Meereshöhe und der spezifischen Feuchte in 700 hPa für Sommer (JJA) und Herbst (SON).	55
5.1	ERA40 Feld des mittleren Luftdrucks auf Meereshöhe für den Tag 01.01.1958 (a) und entsprechendes ermitteltes Anomaliefeld (b).	62
5.2	Erste und zweite empirische Orthogonalfunktion (1.EOF-Muster und 2.EOF-Muster) des mittleren Luftdrucks auf Meereshöhe und der spezifischen Feuchte in 700 hPa.	63
5.3	Erste und dritte Hauptkomponentenreihe (PC1 und PC3) des <i>kombinierten</i> mittleren Luftdrucks in Meereshöhe und spezifische Feuchte in 700hPa Niveau.	64
5.4	Anteil der Gesamtvarianz der ersten 20 Hauptkomponenten.	65
5.5	Logarithmischer Eigenwertverlauf der ersten 20 Eigenwerte.	66
5.6	Bildung von Clustern(Gruppen) mittels euklidischem Distanzmaß(a) und City-Block-Distanz(b).	68
5.7	City-Block-Distanzmaß und Euklidische Distanzmaß	71
5.8	Gewichtetes Euklidisches Distanzmaß	72
5.9	Graphische Darstellung der Lp Metrik für den 2-Dim Raum: Isodistanzen für verschiedene p-Werte.	74
5.10	Zur Begründung der Anwendung der Mahalanobis Distanz.	75
5.11	Zur Erläuterung der Mahalanobis Distanz.	76
5.12	Dargestellt sind die simulierten Punktwolken für n=2000 Datenpunkte bei bekannter Kovarianzmatrix jeweils für eine Korrelation von $\rho = 0.0$ (a) und $\rho = 0.75$ (b), die entsprechenden Höhenlinien der Dichtefunktion (d) und entsprechende 3D Darstellung.	80
5.13	Ergebnisse der Berechnung des kleinsten Abstandes unter Heranziehung von 20 PCs für den Winter (DJF) mit City-Block-Distanz (a), Euklidischer Distanz(b), gewichteter Euklidischer Distanz(c) und Mahalanobis Distanz (d) am Beispiel der Station Feldkirch. Analog stellen die Abbildungen (e), (f), (g) und (h) die Ergebnisse für 7 PCS dar.	81
5.14	Schema der <i>Analog-Methode</i>	83
5.15	Feldklich	84
5.16	Euklidische (quadratische) Distanzen der Merkmale bzw. Hauptkomponenten u_k	85
5.17	Darstellung der Paare Tag - ähnlichster Tag (Analog), wenn die Euklidische Distanzmaß verwendet wird.	85

6.1	Kontingenztafel zur Berechnung von Genauigkeitsmaßzahlen.	92
6.2	Saisonale Häufigkeitsverteilung (DJF) für die Station Innsbruck-Flughafen für die Periode 1958-2000.	96
6.3	Saisonale Häufigkeitsverteilung für die modellierten Zeitreihe der Station Innsbruck Flughafen für die Periode 1958-2000. . .	97
6.4	Signifikanztest	98
7.1	Orographische Lage der Startclim-Stationen.	100
7.2	Wien-Hohe Warte: Darstellung der Beobachtungszeitreihe und modellierte Tageszeitreihe für Winter (DJF) und Sommer(JJA). Zur Suche des ähnlichsten Tages wurden die City Block Norm, die euklidische gewichtete Norm verwendet.	102
7.3	Wien-Hohe Warte: Darstellung der modellierten Tageszeitreihe für Winter(DJF)und Sommer (JJA). Zur Suche des ähnlichsten Tages wurden die Euklidische Distanz und die Mahalanobis Distanz verwendet.	103
7.4	Feldkirch: Darstellung der Beobachtungszeitreihe und modellierten Tageszeitreihe für Winter (DJF)und Sommer(JJA). Zur Suche des ähnlichsten Tages wurden die City Block Distanz und die Euklidische Distanz verwendet.	104
7.5	Feldkirch: Darstellung der modellierten Tageszeitreihe für Winter (DJF) und Sommer (JJA). Zur Suche des ähnlichsten Tages wurden die Euklidische Distanz und die Mahalanobis Distanz verwendet.	105
7.6	Klagenfurt: Darstellung der Beobachtungszeitreihe und modellierten Tageszeitreihe für Winter (DJF) und Sommer(JJA). Zur Suche des ähnlichsten Tages wurden die City Block Distanz und die Euklidische Distanz verwendet.	106
7.7	Klagenfurt: Darstellung der modellierten Tageszeitreihe für Winter (DJF) und Sommer(JJA). Zur Suche des ähnlichsten Tages wurden die Euklidische Distanz und die Mahalanobis Distanz verwendet.	107
7.8	Feuerkogel: Darstellung der modellierten Tageszeitreihe für Winter (DJF) und Sommer(JJA). Zur Suche des ähnlichsten Tages wurden die City Block Distanz und die Euklidische Distanz verwendet.	108
7.9	Feuerkogel: Darstellung der modellierten Tageszeitreihe für Winter (DJF) und Sommer(JJA). Zur Suche des ähnlichsten Tages wurden die gewichtete Euklidische Distanz und die Mahalanobis Distanz verwendet.	109

7.10	Feldkirch: Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihe für Winter(DJF) und Sommer(JJA). Für die Suche des analogen Tages wurden die City Block Distanz (CB) und die gewichtete Euklidische Distanz (eukg) verwendet.	110
7.11	Feldkirch: Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihen für Winter (DJF) und Sommer (JJA). Für die Suche des analogen Tages wurden die Euklidische Distanz (eukng) und die Mahalanobis Distanz (mahal) verwendet.	111
7.12	Wien Hohe Warte (WienHW): Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihe für Winter und Sommer (JJA). Für die Suche des analogen Tages wurden die City Block Distanz (CB) und die gewichtete Euklidische Distanz (eukg) verwendet.	113
7.13	Wien Hohe Warte (WienHW): Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihe für Winter (DJF) und Sommer(JJA). Für die Suche des analogen Tages wurden die Euklidische Distanz (eukng) und die Mahalanobis Distanz (mahal) verwendet.	114
7.14	Klagenfurt: Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihe für Winter (DJF) und Sommer(JJA). Für die Suche des analogen Tages wurden die City-Block-Distanz (CB) und die gewichtete Euklidische Distanz (eukg) verwendet.	115
7.15	Klagenfurt: Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihe für Winter (DJF) und Sommer(JJA). Für die Suche des analogen Tages wurden die Euklidische Distanz (eukng) und die Mahalanobis Distanz (mahal) verwendet.	116
7.16	Feuerkogel: Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihe für Winter (DJF) und Sommer (JJA). Für die Suche des analogen Tages wurden die City-Block-Distanz(CB) und die gewichtete Euklidische Distanz (eukg) verwendet.	117
7.17	Feuerkogel: Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihe für Winter (DJF) und Sommer (JJA). Für die Suche des analogen Tages wurden die Euklidische Distanz (eukng) und die Mahalanobis Distanz (mahal) verwendet.	118
7.18	Landeck: Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihe für Winter (DJF) und Sommer (JJA). Für die Suche des analogen Tages wurden die Euklidische Distanz (eukng) und die Mahalanobis Distanz verwendet (mahal).	119

7.19	Landeck: Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihe für Winter (DJF) und Sommer (JJA). Für die Suche des analogen Tages wurden die Euklidische Distanzmaß (eukng) und die Mahalanobis Distanz verwendet (mahal).	120
7.20	saisonale Korrelation: DJF-7PCs	121
7.21	saisonale Korrelationen: DJF-11PCs	122
7.22	saisonale Korrelationen: DJF-20PCs	123
7.23	saisonale Korrelationen: JJA-7PCs	124
7.24	saisonale Korrelationen: JJA-11PCs	125
7.25	saisonale Korrelationen: JJA-20PCs	126
B.1	saisonale Korrelationen: MAM-7PCS, MAM-11PCs	147
B.2	saisonale Korrelationen: MAM-20PCS, SON-7PCS	148
B.3	saisonale Korrelationen: SON-11PCs, SON-20PCs	149

Tabellenverzeichnis

4.1	Messstellen der Startclim-Datensätze.	48
A.1	Anzahl und Prozentsatz an Stationen mit signifikanter Korrelation (DJF) bei einem Signifikanzniveau von 5 %	130
A.2	Anzahl und Prozentsatz an Stationen mit signifikanter Korrelation (MAM) bei einem Signifikanzniveau von 5 %	131
A.3	Anzahl und Prozentsatz an Stationen mit signifikanter Korrelation (JJA) bei einem Signifikanzniveau von 5 %	131
A.4	Anzahl und Prozentsatz an Stationen mit signifikanter Korrelation (SON) bei einem Signifikanzniveau von 5 %	131
A.5	Zusammenfassung (Minimum, Mittelwert, Maximum) der Reduktion der Varianz (RV) für die Tagesreihen in den Jahreszeiten DJF, MAM, JJA, SON bei 7, 11 und 20 PCs für die 4 verschiedenen Distanzmaße: City-Block-Distanz (CB), gewichtete Euklidische Distanz (eukg), Euklidische Distanz (eukng) und Mahalanobis Distanz (mahal).	132
A.6	Zusammenfassung (Minimum, Mittelwert, Maximum) der Reduktion der Varianz (RV) für die saisonalen Zeitreihen in den Jahreszeiten DJF, MAM, JJA, SON bei 7, 11 und 20 PCs für die 4 verschiedenen Distanzmaße: City-Block-Distanz (CB), gewichtete Euklidische Distanz (eukg), Euklidische Distanz (eukng) und Mahalanobis Distanz (mahal).	133
A.7	Zusammenfassung (Minimum, Mittelwert, Maximum) des Anomaliekorrelationskoeffizienten (ACC) für die Tageszeitreihe in den Jahreszeiten DJF, MAM, JJA, SON bei 7, 11 und 20 PCs für die 4 verschiedenen Distanzmaße: City-Block-Distanz (CB), gewichtete Euklidische Distanz (eukg), Euklidische Distanz (eukng) und Mahalanobis Distanz (mahal).	134

A.8	Zusammenfassung (Minimum, Mittelwert, Maximum) des Anomaliekorrelationskoeffizienten für die saisonalen Zeitreihen in den Jahreszeiten DJF, MAM, JJA, SON bei 7, 11 und 20 PCs für die 4 verschiedenen Distanzmaße: City-Block-Distanz (CB), gewichtete Euklidische Distanz (eukg), Euklidische Distanz (eukng) und Mahalanobis Distanz (mahal).	135
A.9	Ergebnisse der statistischen Maße: Korrelation (r), RMSE (rmse), Mittelwert der Tageszeitreihe (m.b) und modellierten Tagesreihe (m.an), systematischer Fehler (bias), Varianz (var.b und var.an), Standardabweichung (sd.b und sd.an) und Interquantile (inter.b und inter.an) für DJF bei 7, 11 und 20 Hauptkomponenten.	136
A.10	Ergebnisse der statistischen Maße: Korrelation (r), RMSE (rmse), Mittelwert der Tageszeitreihe (m.b) und modellierten Tagesreihe (m.an), systematischer Fehler (bias), Varianz (var.b und var.an), Standardabweichung (sd.b und sd.an) und Interquantile (inter.b und inter.an) für MAM bei 7, 11 und 20 Hauptkomponenten.	137
A.11	Ergebnisse der statistischen Maße: Korrelation (r), RMSE (rmse), Mittelwert der Tageszeitreihe (m.b) und modellierten Tagesreihe (m.an), systematischer Fehler (bias), Varianz (var.b und var.an), Standardabweichung (sd.b und sd.an) und Interquantile (inter.b und inter.an) für JJA bei 7, 11 und 20 Hauptkomponenten.	138
A.12	Ergebnisse der statistischen Maße: Korrelation (r), RMSE (rmse), Mittelwert der Tageszeitreihe (m.b) und modellierten Tagesreihe (m.an), systematischer Fehler (bias), Varianz (var.b und var.an), Standardabweichung (sd.b und sd.an) und Interquantile (inter.b und inter.an) für SON bei 7, 11 und 20 Hauptkomponenten	139
A.13	Ergebnisse der statistischen Maße: Korrelation (r), RMSE (rmse), Mittelwert der saisonaler Zeitreihe (m.b) und modellierten Tagesreihe (m.an), systematischer Fehler (bias), Varianz (var.b und var.an), Standardabweichung (sd.b und sd.an) und Interquantile (inter.b und inter.an) für DJF bei 7, 11 und 20 Hauptkomponenten.	140

A.14 Ergebnisse der statistischen Maße: Korrelation (r), RMSE (rmse), Mittelwert der saisonaler Zeitreihe(m.b) und modellierten Tagesreihe (m.an), systematischer Fehler (bias), Varianz (var.b und var.an), Standardabweichung (sd.b und sd.an) und Interquantile (inter.b und inter.an) für MAM bei 7, 11 und 20 Hauptkomponenten.	141
A.15 Ergebnisse der statistischen Maße: Korrelation (r), RMSE (rmse), Mittelwert der saisonalen Zeitreihe (m.b) und modellierten Tagesreihe (m.an), systematischer Fehler (bias), Varianz (var.b und var.an), Standardabweichung (sd.b und sd.an) und Interquantile (inter.b und inter.an) für JJA bei 7, 11 und 20 Hauptkomponenten.	142
A.16 Ergebnisse der statistischen Maße: Korrelation (r), RMSE (rmse), Mittelwert der saisonalen Zeitreihe (m.b) und modellierten Tagesreihe (m.an), systematischer Fehler (bias), Varianz (var.b und var.an), Standardabweichung (sd.b und sd.an) und Interquantile (inter.b und inter.an) für SON bei 7, 11 und 20 Hauptkomponenten.	143
A.17 Skill Scores DJF bei 7 Hauptkomponenten.	144
A.18 Skill Scores DJF bei 20 Hauptkomponenten.	144
A.19 Skill Scores JJA bei 7 Hauptkomponenten.	145
A.20 Skill Scores JJA bei 20 Hauptkomponenten.	145

Kapitel 1

Motivation und Problemstellung

Um Ähnlichkeiten zwischen Datenpunkten, bzw. Feldern zu bestimmen, werden in der Multivariate Statistik üblicherweise Distanzmaße benutzt.

Um die Ähnlichkeit zwischen großskaliger Zirkulation und beobachteten Zirkulationsmustern zu bestimmen, haben Zorita und Storch (1999) [61] das Euklidische Distanzmaß benutzt. In ihrer Arbeit, in der sie die *Analog-Methode* mit anderen komplexeren statistischen Dowscaling-Verfahren vergleichen, wird suggeriert, die einzelnen Summanden (bei der Euklidischen Distanzmaß) mit verschiedenen Gewichten zu multiplizieren. Dabei sollen die Gewichte die Bedeutung der einzelnen empirischen Orthogonalfunktionen (EOF) reflektieren. In diesem Sinne können die Gewichte beispielsweise zu den erklärten Varianzen proportional sein. Dieser Idee wird in dieser Diplomarbeit nachgegangen werden.

Die *Analog-Methode* ist ein statistisches Verfahren, welches darin besteht, großskalige atmosphärische Felder, wie sie von GCMs simuliert werden, miteinander zu vergleichen um lokalskalige Felder zu generieren. Diese Methode kann auch für Perioden, in denen keine Beobachtungen vorliegen eingesetzt werden. In der hier vorgestellten Anwendung werden nur großskalige atmosphärische Felder miteinander verglichen, die als „historische“ Beobachtung vorliegen, um auf dieser Weise aufgrund eines Distanzmaßes zwei in der Vergangenheit liegende Tage mit ähnlichen Witterungsverhältnisse zu finden.

Ein weiteres Vorhaben in dieser Arbeit soll die Anwendung von Distanzmaßen zur Bestimmung von Ähnlichkeiten zwischen Feldern sein. Vor allem soll untersucht werden, ob andere Distanzmaße (z.B. City Block, Mahalanobis) gegenüber der Euklidischen Distanz zu verbesserten Resultaten führen. Die Motivation zu dieser Fragestellung stammt aus den Muster- und

Bilderkennungs-Verfahren (siehe W. Yambor, B. Draper, and R. Beveridge (2000)). Zahlreiche Verfahren zur Mustererkennung basieren - wie beispielsweise bei Storch und Zorita - auf der *Analog-Methode* mit vorgeschalteter Hauptkomponentenanalyse (Principal Component Analysis, PCA).

In dieser Diplomarbeit soll untersucht werden wie sich die Wahl des Distanzmaßes auf das statistische Downscaling, in Speziellen auf die *Analog-Methode* auswirkt.

Kapitel 2

Einleitung



Das Klima wurde bis in den 70er Jahren hauptsächlich mit der Atmosphäre in Verbindung gebracht. Mittlerweile weiß man, dass es aus einer komplexen Wechselwirkung zwischen den einzelnen Sphären, den Komponenten des Klimasystems, zu verstehen ist.

Der Begriff Klima wird ursprünglich auf Aristoteles zurückgeführt. Er setzte Klima mit der Neigung der Sonneneinstrahlung gleich, um Klimazonen zu unterscheiden. Hann [22] beschrieb das Klima als

"...die Gesamtheit aller meteorologischen Erscheinungen, die den mittleren Zustand der Atmosphäre an irgend einer Stelle der Erdoberfläche charakterisieren."

Im letzten Jahrhundert wurde der Klima Begriff erweitert, sodass er ausser dem „mittleren Zustand“ auch die höheren statistischen Momente miteinbezieht. Hantel et al. [23] beschreiben Klima als das

"...statistische Verhalten der Atmosphäre, das für eine relativ große zeitliche Größenordnung charakteristisch ist."

In diesem Sinn werden die Klimaparameter als statistische Variablen verstanden, die nicht nur zeitlich, sondern auch räumlich variieren können.

In den nächsten Abschnitten wird auf einige wichtigen Probleme, die in Zusammenhang mit der Modellierung des Klimas stehen, eingegangen. Anschließend werden spezielle Modellierungsansätze erläutert und im Detail näher besprochen.

2.1 Probleme der Klimamodellierung

2.1.1 Skaligkeit atmosphärischer Phänomene

Ein wichtiger Punkt bei der Klimamodellierung ist die Skaligkeit atmosphärischer Phänomene. Änderungen der atmosphärischen Phänomene vollziehen sich in Zeit und Raum, besitzen also sowohl eine räumliche Ausdehnung (horizontal und vertikal) als auch eine gewisse zeitliche Dauer. Die Abbildung 2.1 illustriert wie sich atmosphärische Prozesse räumlich und zeitlich von einander unterscheiden.

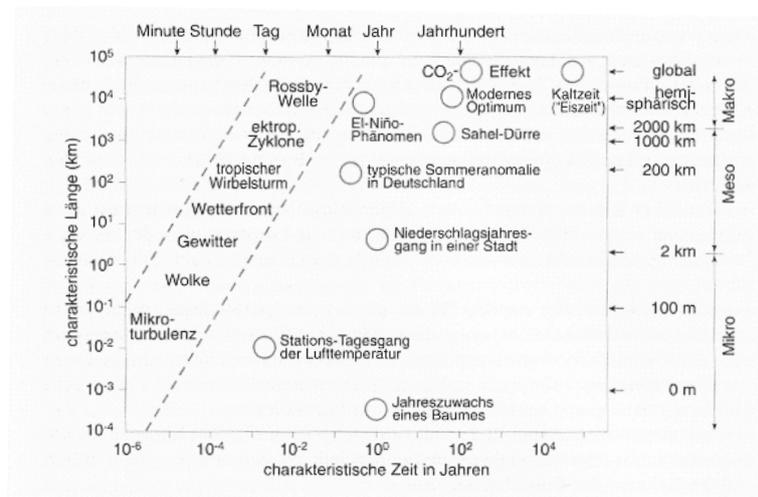


Abbildung 2.1: Größenordnungen und Zeitskala einiger atmosphärischer Vorgänge (*Quelle: Kraus, H.*).

Um die räumliche Ausdehnung der atmosphärischen Phänomene in die Modellierung einbeziehen zu können wird ein Maßstab gewählt, der typischerweise einem regelmäßigen Gitter von Beobachtungspunkten entspricht.

Dieses Gitternetz ist dreidimensional, wodurch die Bewegung der Gase in allen Höhen der Atmosphäre vorausgesagt werden kann.

Der typische Gitterpunktabstand früherer Atmosphärenmodelle beträgt 550km mal 550km in der horizontalen mal 1km in der vertikalen Ebene. Bei Ozeanmodellen dagegen beträgt die Auflösung 125-250km in der horizontalen mal 200-400km in der vertikalen Ebene.¹ Auf diese Weise lassen sich die Begriffe *regional* und *lokal* besser erläutern: Der Begriff *regional* bezieht sich immer auf ein vier-acht faches des Gitterpunktabstandes eines Klimamodells während man sich mit *lokal* immer auf einen Gitterpunkt bezieht. Zusätzlich gibt es noch die *globale* Skala 2.1, die sich auf eine horizontale Ausdehnung von mindestens 2000km bezieht.

Abbildung 2.2 zeigt ein Beispiel einer Diskretisierung anhand eines den Nordatlantik und Europa unterteilenden Gitternetzes (horizontale Diskretisierung). Während T21 (Auflösung ~ 500 km) Land und Meer in sehr rudimentärer Weise wiedergibt, kann mit T42 (Auflösung ~ 250 km) eine bessere Darstellung erreicht werden. Noch höhere Auflösungen wie T63 und T106, welche als Vorläufer moderner Klimamodelle gelten, werden derzeit in Wettervorhersagemodellen benutzt.

Die Grenze der zeitlichen Vorhersagbarkeit atmosphärischer Phänomene liegt bei zwei Wochen. Dies liegt in der Tatsache begründet, dass Unsicherheiten in den Anfangsbedingungen der Modelle über einen längeren Zeitraum zu großen Veränderungen in den Ergebnissen führen. Im Gegensatz zu Wettervorhersagemodellen berechnen Klimamodelle statistische Werte und erlauben dadurch den Forschern längerfristige Auswirkungen, ausgelöst von sich ändernden Rahmenbedingungen, zu studieren. Dabei ist es wichtig zwischen dem statistischen und deterministischen Charakter eines Modells zu unterscheiden. Während erster z.B. Mittelwerte längerfristiger Prognosen für ganze Regionen liefern, treffen deterministische Wettervorhersagen Aussagen über bestimmte Ereignisse für eine konkrete Zeit an einem bestimmten Ort.

Abbildung 2.3 zeigt die Simulationsergebnisse dreier verschiedener Klimamodelle, berechnet für die mittlere Temperaturverteilung des Januars im Mackenzie-Becken, Nordkanada.

Ein Vergleich zwischen den Simulationen mit der Beobachtung (rechts unten) lässt den Schluss zu, dass keines der Modelle in der Lage ist den Klimazustand des Mackenzies-Beckens befriedigend zu simulieren. Eine detailliertere Besprechung dieser Ergebnisse kann in [49] nachgelesen werden.

Es gibt mehrere Studien, wie in Abschnitt 2.3 beschrieben, die belegen,

¹<http://lbs.hh.schule.de/welcome.phtml?unten=/klima/klimawandel/klimamodelle/modelle2.html>

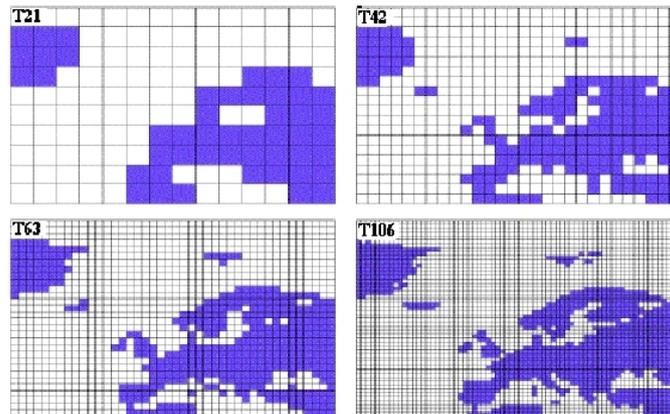


Abbildung 2.2: Horizontale Aufteilung von Europa und Nordatlantik in Gitterelemente (Quelle: Cubasch et al. 1995).

dass globale Klimamodelle jene atmosphärischen Bewegungsvorgänge, die sich auf der regionalen Skala abspielen (z.B. Temperatur) unzureichend erfassen. Jedoch sind diese in der Lage, Prozesse, die sich auf globalen Skalen (Abb. 2.2) ereignen, zufriedenstellend wiederzugeben.

2.2 Downscalingansätze

Aus Vorhersagen auf der globalen Skala kann wenig Information über Klimaänderungen für eine bestimmte Region abgeleitet werden. Sowohl gute regionale wie auch lokale Klimaprognosen sind notwendig, um die Folgen der klimatischen Schwankungen (z.B. lokale Wetterextreme) bestimmter Regionen realistisch abzuschätzen. Dafür müssen Klimamodelle in der Lage sein, die Einflüsse sowohl der Land-See-Verteilung, als auch einer detaillierten Orographie wiederzugeben.

Regionales Klima entsteht aus der Wechselwirkung zwischen dem großskaligen Zustand des Klimasystems und regionalen Faktoren (z.B. Topographie). Das Verfahren, welches den Transfer von großskaliger Information auf kleinere Skalen ermöglicht wird als *Down-scaling* [49] bezeichnet.

Im folgende Abschnitt werden gängige Downscalingverfahren vorgestellt und deren Vorteile und Nachteile gegenübergestellt. Grundsätzlich gibt es zwei Ansätze: Das dynamische und das statistische Downscaling. Abbildung 2.4 stellt vereinfacht die Prinzipien des dynamischen und statistischen Downscalings dar.

Downscaling besteht darin, hochaufgelöste Klimainformation i.a. von ei-

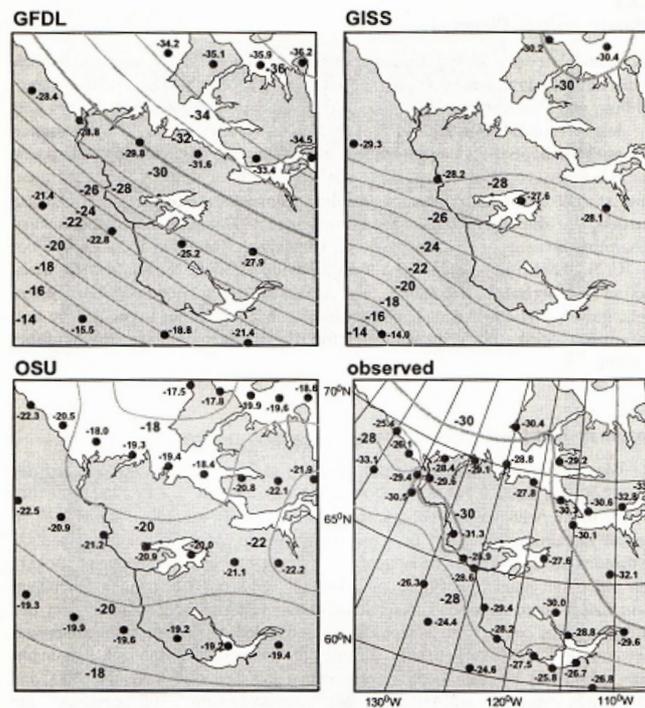


Abbildung 2.3: Mittlere Temperaturverteilung im Januar im Mackenzie Becken in Nordkanada (*Quelle: von Storch, Goss und Heimann 1999*).

ner grob aufgelösten Skala, wie sie aus globalen Klimamodellen (GCMs) bekannt ist, auf die gewünschte feinere Skala zu transferieren. Während sich das dynamische Downscaling eines physikalischen Modells (RGM, Regionales Modell) bedient, werden beim statistischen Downscaling meteorologische Größen von denen ein empirischer Zusammenhang bekannt ist zueinander in Beziehung gesetzt. Im meteorologischen Sinn bedeutet dies, dass regionales Klima in erster Linie von atmosphärischen Prozessen auf der globalen Skala konditioniert wird.

Im Vergleich zum statistischen Downscaling beansprucht das dynamische Downscaling eine höhere Rechenleistung und die Simulationszeit zur Produktion äquivalenter Ergebnisse ist länger.

Da das dynamische Downscaling nicht Gegenstand dieser Diplomarbeit ist, wird hier auf eine detaillierte Erläuterung auf [17] verwiesen. Zum besseren Verständnis des statistischen Downscalings wird ein Beispiel aus der Literatur [46] gegeben. Die Grundidee besteht darin, ein statistisches Modell

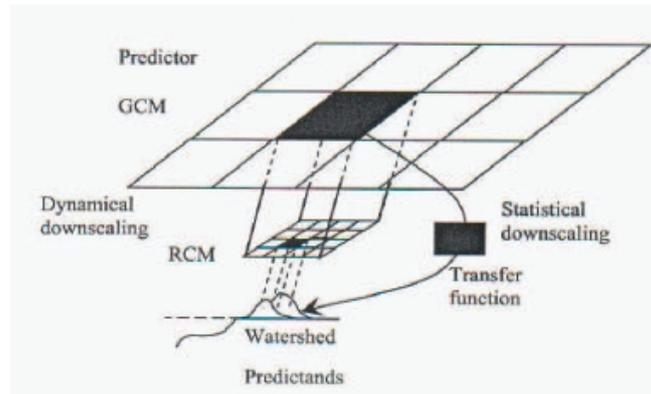


Abbildung 2.4: Schematische Darstellung der statistischen und dynamischen Downscaling Ansätze (*Quelle: Wetterhall [53]*).

aufzustellen, bei dem nur historische Beobachtungen eingesetzt werden: großskalige Daten eines meteorologischen Parameters, die realistisch von einem GCM wiedergegeben werden [48] und Daten in der gewünschten regionalen Skala, die nicht von GCMs simuliert werden können. Zwischen den beiden verschieden aufgelösten Datensets soll eine empirische Beziehung bestehen. Schließlich wird dieses Modell auf den großskaligen Modell-Output angewandt. Die Vorgehensweise lautet:

- Zuerst wird eine von Interesse regionale Klimavariablen \vec{R} gewählt.
- Es soll eine Klimavariablen \vec{S} gefunden werden, die
 1. \vec{R} steuert im Sinne einer statistischen Beziehung zwischen \vec{R} und \vec{S} der Form:

$$\vec{R} = G(\vec{S}; \vec{\alpha}) + \epsilon \quad (2.1)$$

wobei $G(\vec{S}; \vec{\alpha})$ den substantiellen Anteil der gesamten Varianz von \vec{R} darstellt. Der Vektor $\vec{\alpha}$ enthält Parameter, die geeignet sind, das *Fitten* der Gleichung zu optimieren. Die Komponenten des Vektor $\vec{\alpha}$ werden bestimmt als diejenige Werte, welche

$$\left\| \vec{R} - G(\vec{S}; \vec{\alpha}) \right\| = \epsilon(\vec{\alpha}) \quad (2.2)$$

minimieren. ϵ stellt das *Rauschen* (Residuum) dar (siehe 5.1.1).

2. von einem Klimamodell realistisch simuliert wird

- es sollen historische Beziehungen (\vec{r}_t, \vec{s}_t) von (\vec{R}, \vec{S}) verwendet werden um $\vec{\alpha}$ zu schätzen
- Validierung des gefitteten Modell mit unabhängigen Daten
- das validierte Modell wird auf Simulationen der Variablen \vec{S} von GCMs angewandt

Ein statistisches Modell lässt sich in einfacher Form ausdrücken: [51]

$$y = F(x) \quad (2.3)$$

Bei der kleinskaligen Variablen y handelt es sich um den Prädiktand, also um eine Variable, die nicht adäquat (z.B. Tagestemperatur oder akkumulierter monatlicher Niederschlag) von GCMs simuliert werden kann. Der Prädiktor x steht für die großskaligen Merkmale. Dabei wird angenommen, dass x von einem Klimamodell befriedigend simuliert wird. Als x werden oft die Eigenschaften der Zirkulation in Form von Hauptkomponenten der Luftdruck- oder Temperaturfelder oder auch deren Kombination eingesetzt [21]. Mit Gleichung (2.3) wird ein Zusammenhang zwischen synoptischen Situationen und lokalem Wetter hergestellt [51].

Das statistische Downscaling ist formell identisch zur *model output statistics*(MOS) ² und *perfect prog*(PP)Methode ³. Der Hauptunterschied zwischen der Methode des *perfect prog* und dem statistischen Downscaling besteht in der Annahme, dass beim Downscaling die Klimavariablen von GCMs befriedigend simuliert werden.

Aus der synoptischen Erfahrung ist bekannt [40], dass verschiedene y mit dem gleichen x konsistent sein können. Aus diesem Grund muss die Gleichung (2.3) als eine stochastische Gleichung verstanden werden: das Merkmal y ist ein Zufallsvektor und F eine Zufallsfunktion, die durch die Umsetzung von x abhängig ist.

In vielen Anwendungen wird F als linear angenommen. Der Prototyp ist also die Lineare Regression:

$$F = \alpha x + \epsilon \quad (2.4)$$

²dieses statistische Verfahren wird in der Wettervorhersage benutzt. Grundlage sind multilineare Regressionsgleichungen, die durch Mithilfe bestimmter Algorithmen diejenigen Modellparameter (Prädiktanden) herausfiltert, die für das Wettergeschehen von Bedeutung sind; <http://www.top-wetter.de/lexikon/m/mos.htm>

³ähnlich zum MOS, allerdings bietet die Methode keine Alternative zur Korrektur möglicher systematischer Fehler

Der Wert für ϵ stammt aus der Normalverteilung: Fehler mit Mittelwert Null und Standardabweichung σ mit $0 < \alpha < 1$. Die möglichen Schwankungen von ϵ hängen nicht mit x zusammen. In diesem Fall stammt die Zufälligkeit von y aus der Zufälligkeit von ϵ . Ein häufig benutzter Zugang ist der deterministische, bei dem y ein erwartungstreuer Schätzer von x ist.

Das SDS Verfahren kann in 3 Klassen [47] unterteilt werden:

- Wettergeneratoren: Grundsätzlich werden die statistischen Eigenschaften von Klimavariablen simuliert, welche die Informationen über die Klimavariabilität auf der lokalen Skala liefern.
- Transferfunktionen: Im wesentlichen wird dabei ein direkter Zusammenhang abgeleitet, z.B. durch eine Regression
- Klassifikationsverfahren: haben ihre Grundlage im Konzept der synoptischen Klimatologie, von dem ein Zusammenhang zwischen atmosphärischen Zustand und lokalen Variablen abgeleitet wird. Die in dieser Diplomarbeit verwendete *Analog-Methode* gehört dieser Klasse an.

Abbildung 2.5 stellt die Approximationen für die Funktion $y = F(x)$ dar. Dabei repräsentiert y den Prädiktand (regionale Klimavariablen), x den Prädiktor (großskalige Klimavariablen) und F stellt eine stochastische oder deterministische Funktion dar, welche von x bestimmt wird. Im allgemeinen ist F nicht bekannt, sie wird entweder dynamisch (z.B. durch regionale Klimavariablen) oder empirisch aus Beobachtungsdaten modelliert. In manchen Fällen ist y und x die gleiche Klimavariablen, die sich aber in der räumliche Auflösung unterscheiden [57]. Für die Approximationen der Funktion $y = F(x)$ wurden verschiedene statistische Downscaling-Verfahren verwendet: Kanonische Korrelation, Analog-Methode, neuronale Netze (wird in Kapitel 2.3 genauer erläutert) und Kriging⁴. In y wurde der monatliche mittlere Winter Niederschlag an einem Ort in Nordspanien eingesetzt. x stellt die Hauptkomponenten des Nordatlantischen mittleren monatlichen Luftdruckfeldes dar.

In den meisten Downscaling Anwendungen wurden Niederschlag und Temperatur verwendet. Es gibt auch Studien, bei denen andere Variablen eingesetzt worden sind. Im Kapitel 2.3 werden einige Arbeiten genauer diskutiert.

⁴statistisches Interpolationsverfahren, das im Gegensatz zu anderen einfachen Methoden die räumliche Varianz berücksichtigt ; <http://de.wikipedia.org/wiki/Kriging>

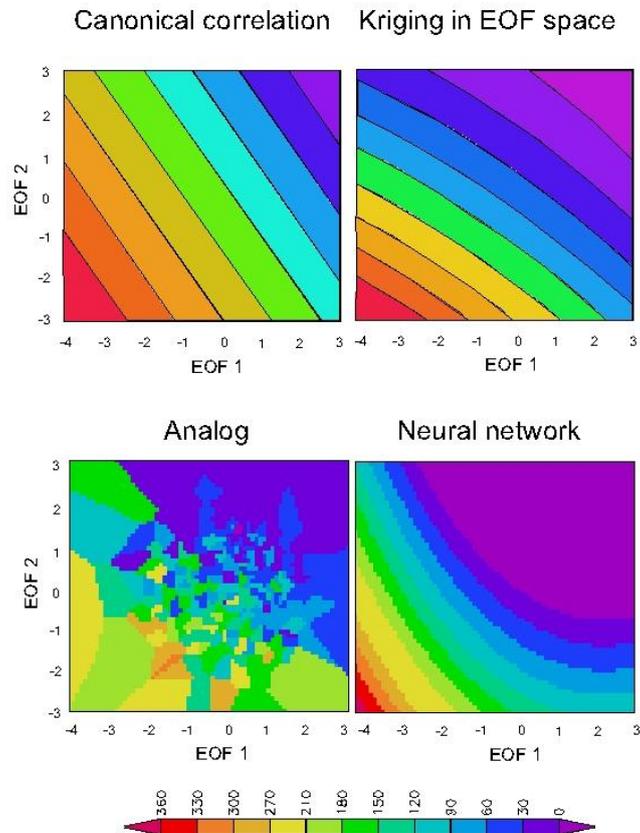


Abbildung 2.5: Verschiedene Approximationen für die Funktion $y = F(x)$ (Quelle: Storch 2000 [47]).

2.3 Forschungsstand

GCMs sind wichtige Werkzeuge zum Studium der Klimavariabilität. Sie können auf der planetaren Skala im allgemeinen die mittleren Eigenschaften des globalen Klimas, wie etwa Lage der Jetstreams oder der Innertropischen Konvergenzzone realistisch wiedergeben. Wie im vorigen Kapitel bereits erwähnt, ist die Fähigkeit der GCMs lokalskalige Klimavariablen (Temperatur, Niederschlag) zu simulieren begrenzt [20].

In Abbildung 2.6 wird der schematische Aufbau eines Zirkulationsmodells dargestellt. Die Beschreibung der physikalischen Prozesse, die mehrere Variable einbeziehen, wird in einem System von Differentialgleichungen beschrieben und berechnet.

Einen Vergleich verschiedener hochauflösender GCMs findet man in [8].

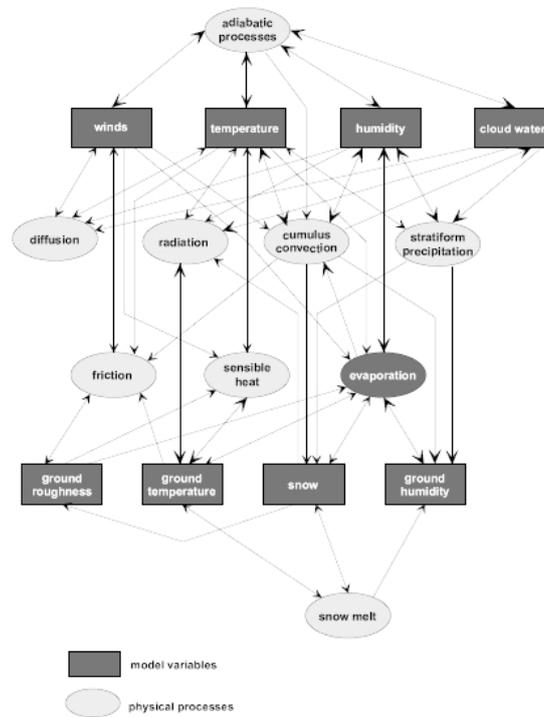


Abbildung 2.6: Physikalische Prozesse in einem typischen Globalen Zirkulationsmodell (Quelle: *Fraedrich, K., A..A. Aigner, E. Kirk, and F. Lunkeit 2005 [14]*).

Darin wurde der Schluss gezogen, dass i.a. die GCMs die Temperatur besser simulieren als den Niederschlag. In Verbindung mit der Frage nach einer möglichen Klimaänderung konnte man aus den Ergebnissen der Modelle keine eindeutige Klarheit bezüglich einer möglichen Verdoppelung des CO₂ Ausstoßes feststellen [61]. In manchen Fällen wurden zwei Versionen des gleichen atmosphärischen Modells mit verschiedenen Ozean-Modellen gekoppelt. Das Ergebnis waren verschiedene Temperaturmuster, die manchmal miteinander sogar negativ korrelieren. In Anbetracht dessen wird in [61] argumentiert, dass im Fall des Niederschlages verschiedene Modelle nicht zu gleichen, sondern zu unterschiedlichen Ergebnissen führen. Ein anderes Beispiel kann in [35] gefunden werden, in dem die Performance des Modells CCM2 mit den Auflösungen T42 und T106 verwendet wurde, um das Klima des Sacramento River Basin in Kalifornien zu simulieren. Es konnte beobachtet werden, dass obwohl der mittlere jährliche Niederschlag befriedigend simuliert werden konnte, die Wahrscheinlichkeitsverteilungen der Simulation und

der Beobachtungen sich von einander unterscheiden. Während der simulierte Niederschlag in Form von Nieselregen, verteilt über viele Niederschlagstage vorkam, handelt es sich tatsächlich um auf wenige Tage konzentrierten starken Niederschlag.

Die Tatsache, dass Klimamodelle befriedigende Ergebnisse im großskaligen Bereich liefern, aber auf der regionalen Skala versagen, scheint widersprüchlich. Storch führt dies auf folgende Gründe zurück [61]:

Die verwendete räumliche Auflösung bietet eine inadäquate Beschreibung der Erdoberfläche: z.B. ist die Land-Meer Verteilung verwischt und die Berge im Mittel ausgedehnte flache Hügel. Auf diese Weise kann z.B. der beobachtete jährliche Zyklus des Niederschlags in den Alpen nicht realistisch simuliert werden welche im Sommer auf der Nordseite ein Niederschlagsmaximum aufweisen, während dies auf der Südseite im Winter auftritt [13].

Im Fall der spektralen Modelle stellt die Topographie eine zusätzliche Quelle von Schwierigkeiten dar, die sich auf der lokalen Skala problematisch erweist [28]. Diejenigen Prozesse, die aufgrund der Gitterweite des Klimamodells nicht erfasst werden und deshalb nicht berechnet werden können, weil sie in sehr kleinen Skalen (z.B. Turbulenz, Wolkenmikrophysik) stattfinden, können nur mittels Parametrisierung berücksichtigt werden. Die Wahl der Parametrisierung hängt mit dem gewählten Modell zusammen. Manche Prozesse finden in äußerst kleinen Skalen statt, sodass sich die Parametrisierung gleichfalls als schwierig erweist. Eine gute Parametrisierung subskaliger Prozesse erscheint sinnvoll um das regionale Klima realistisch wiederzugeben, denn es sind gerade die subskaligen Prozesse, die einen grossen Einfluss darauf haben [61].

Durch Betrachtung von Abbildung 2.7 kann der Begriff der Subskaligkeit besser verstanden werden. Ein Prozess ist dann subskalig, wenn er kleiner ist als die Auflösung des Modells und deshalb nicht erfasst werden kann. Andererseits sind jene Prozesse supraskalig die in ihrer räumlicher Ausbreitung größer sind als die Fläche.

Strategie zur Überbrückung de Skalenproblems

Beim Anstreben einer höheren Auflösung der GCMs ergaben sich 2 Richtungen: Durch den Fortschritt der Rechenleistung wurde es möglich GCMs mit zunehmend feinerer horizontaler Auflösung zu entwickeln. Anfang der 90er Jahre galt T21 für ein Submodell ($5.6^\circ \times 5.6^\circ$) als die beste mögliche Auflösung. Gegen Ende der 90er Jahre konnten die Simulationen schon mit T106 (1.2°) durchgeführt werden. Diese Auflösung war sehr rechenintensiv und infolge dessen beschränkte man sich auf GCMs mit hoher räumlicher Auflösung, wel-

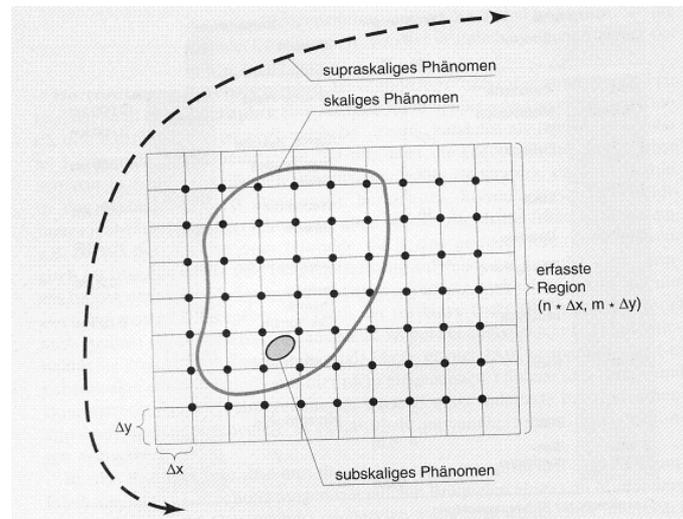


Abbildung 2.7: Hilfestellung zur Erläuterung des Begriffes Skaligkeit (räumlich-horizontal) (*Quelle: Schönwiese 1994 [38]*).

che aber nur über eine relativ kurze Zeit errechnet wurden (Zeitscheibe) [9]. Andererseits wurden die sogenannten „limited Area“- Modelle (LAM) [17] immer häufiger eingesetzt, dies geschah auch um die Ozean-Komponente des Klimamodells zu simulieren. Die LAMS sind aufwendige Atmosphären- oder Ozean-Modelle [26], gültig für ein begrenztes geographischen Gebiet (um die $10^7 km^2$) dessen räumliche Auflösung zwischen 20-50 km liegt. Die nötigen Randbedingungen werden von den GCMs geliefert. Das Einbeziehen topographischer Faktoren wurde durch diese neue Art von Klimamodellen möglich gemacht. Zusätzlich ist es gelungen zyklonale Aktivität und orographisch induzierten Niederschlag in mittleren Breiten besser zu simulieren.

In Allgemeinen wurde erwartet, dass eine zunehmende feinere Auflösung des Modells zu besseren Ergebnisse in der regionalen Simulation führen würde [31]. Bei manchen der vom UK Meteorological Office (UKMO) und von der Meteo-France entwickelten LAMs, die auf dem UKMO GCM und ARPEGE GCM beruhen, wurden jedoch systematische Fehler festgestellt, die nicht durch eine feinere Auflösung zu beseitigen waren. Es wurde angenommen, dass diese Fehlerquellen mit der Parametrisierung subskaliger Prozesse zusammenhängen. Derartige Parametrisierungen wurden von den GCM Vorgängern übernommen [30]. Auf dies Weise wird nicht nur eine feinere Auflösung benötigt sondern auch eine bessere subskalige Parametrisierung notwendig.

Statistisches Downscaling

Als alternativer Ansatz, um das Missverhältnis der Skalen zwischen Klimasimulationen und der Notwendigkeit von Aussagen über die Entwicklung des regionalen Klimas aufzulösen, bietet sich das Statistische Downscaling [17, 24] an. Aufgrund der relativ leichten Implementierung und der benötigten geringen Rechenleistung (verglichen mit LAMs) hat diese Technik weit verbreitet Anwendung gefunden. In [55] findet man einen Überblick über verschiedene statistische Downscaling Techniken. In [61] wird jeweils ein Verfahren aus einer Klasse von statistischen Downscaling Methoden gewählt und mit der *Analog-Methode* in ihrer Leistung verglichen. Für diesen Zweck wurden großskalige historische Beobachtungen herangezogen, um die statistischen Modelle zu fitten. Anschließend wurde das Ergebnis für eine unabhängige Periode mit den Beobachtungen verglichen. Im Vergleich zu den anderen statistischen Methoden zeichnet sich die *Analog-Methode* durch geringeren Umsetzungsaufwand aus, da es sich dabei in ihrer Struktur um ein simples Verfahren handelt. Bei allen statistischen Downscaling-Verfahren wird von folgenden Voraussetzungen ausgegangen [61]:

1. Der Zusammenhang zwischen den großskaligen Prädiktoren und den regionalskaligen Prädiktanden besteht auch unter veränderlichen Klimabedingungen.
2. Die Prädiktoren sind Variablen, die von den GCMs realistisch simuliert werden .
3. Mögliche Klimaänderungen müssen von den Prädiktoren bereits erfasst worden sein.

Aus einer genaueren Analyse dieser Voraussetzungen folgt, dass Punkt 1 in der Praxis nicht gewährleistet sein muß. Trotzdem steigt die Wahrscheinlichkeit, dass die erste Voraussetzung erfüllt wird, sobald lange Zeitreihen zur Verfügung stehen (mindestens einige Dekaden) um das Modell zu „trainieren“. Es sollte sich um repräsentative Zeitreihen handeln, die charakteristische Wettersituationen des betrachteten regionalen Klimas beinhalten. Das bedeutet, dass die Änderung der großskaligen Variablen in einem veränderten Klima innerhalb des natürlichen Variabilitätsbereichs des aktuellen Klimas liegen müssen. Dies impliziert jedoch nicht, dass auch zukünftige Trends richtig vorhergesagt werden können, da die statistischen Zusammenhänge unter anderen klimatischen Bedingungen nicht mehr notwendigerweise gelten. Wenn die neue Klimavariabilität außerhalb des Variabilitätsbereichs des Modells liegt, kann man nur mit einiger Vorsicht sinnvolle Informationen aus den resultierenden Ergebnissen ableiten [61].

Ähnliches gilt auch für GCMs, hinsichtlich der Aussagen über Klimaänderungen, wenn diese Parametrisierungen enthalten, die nur für das aktuelle Klima gelten. Allerdings gründet sich die funktionale Form dieser Parametrisierung auf den physikalischen Zusammenhang. Dabei wird erwartet, dass diese Parametrisierungen auch unter einem veränderten Klima ihre Gültigkeit behalten [61].

Statistische Downscaling Verfahren

Im folgenden wird auf die verschiedenen Verfahren beim statistischen Downscaling näher eingegangen:

Transferfunktionen, Wetterklassifikation und Wettergeneratoren

Die Transferfunktionen werden sehr häufig zum Downscaling verwendet. Unter Ihnen kann man zwischen linearen und nicht linearen Methoden unterscheiden. Sie umfassen die einfache lineare Regression bis hin zu den multiplen linearen Regressionsverfahren (MLR). Auch das Konzept der sogenannten Teleconnections [62] kann benutzt werden, um ein lineares Modell aufzustellen.

Die linearen Methoden bieten im Vergleich zu den anderen die Möglichkeit einer physikalischen Interpretation. Als Prädiktor werden entweder die Hauptkomponenten von atmosphärischen Feldern [61] oder Reihen atmosphärischer Variable an den Gitterpunkten gewählt [36].

Eine weiteres Verfahren stellt die Kanonische Korrelationsanalyse(CCA) dar. Diese sucht charakteristische Variabilitätsmuster und setzt sie in Beziehung zueinander. Zorita [63] bewies mittels CCA das Vorhandensein von kohärenten simultanen Variationen zwischen dem Nordatlantischen SLP Feld und dem monatlichen Niederschlag für den Winter auf der Iberischen Halbinsel [25]. CCA fand auch beim Downscaling von Temperatur und Niederschlag Anwendung [48, 21]

Der Nachteil linearer Methoden besteht daran, dass sie optimal nur auf normalverteilte regionale Variable angewandt werden können. Allerdings können auch Zusammenhänge zwischen großskaligen Feldern und der Wahrscheinlichkeitsverteilung einer nicht normalverteilten Variablen, z.B. beim lokalen Wind, hergestellt werden. Hierbei wird angenommen, dass die interannuellen Variabilitäten dieser Wahrscheinlichkeiten normal verteilt sind. Andere Varianten sind die Singularwertzerlegung (SVD) [7] und die Redundanzanalyse (RDA) [44]. Während die CCA die Korrelation der Koeffizienten zwischen zwei Musterpaaren maximiert, sucht die SVD Richtungen mit ma-

ximaler Varianz.

Im meteorologischen Kontext wurde SVD das erste Mal von [32] benutzt, um simultane Beziehungen zwischen der mittleren monatlichen Lufttemperatur über den Vereinigten Staaten und den hemisphärischen Meeresluftdruckmustern zu finden. Ein Vergleich der Methoden zur Identifizierung von gekoppelten Variabilitätsmustern kann in [7] gefunden werden.

Im Gegensatz zur CCA maximiert die Redundanzanalyse die erklärte Varianz der Koeffizienten der lokalskaligen Muster. Dieses Verfahren wird jedoch seltener verwendet, obwohl in [52] gezeigt werden konnte, dass die Redundanzanalyse ähnlich gute Ergebnisse wie CCA liefert.

Das Prinzip von Klassifikationsverfahren kann anhand des folgenden Beispiels erläutert werden [61]. Zuerst werden Wetterlagen für ein bestimmtes Gebiet definiert und eine Anzahl von historischen Beobachtungen den vorher definierten Wetterklassen zugeordnet. Die Kriterien zur Klassifikation werden in den Simulationen der GCMs berücksichtigt, sodass die atmosphärische Zirkulation zu einer bestimmten Klasse als zugehörig gilt. Zu jeder beobachteten Zirkulation gibt es eine simultane Beobachtung einer lokalskaligen Variable. Der Wert der lokalskaligen Variablen kann entweder als Mittelwert aller regionalen Beobachtungen simultan zu den Elementen der entsprechenden Klasse oder als Stichprobenwert der Klasse gewählt werden. Für Downscaling-Zwecke wurde in [61] die Großwetterlagen-Klassifikation von Hess und Brezowsky [6] verwendet.

Während für den monatlichen Niederschlag vorzugsweise lineare Methoden eingesetzt werden, eignen sich für den täglichen Niederschlag die Klassifikationsmethoden besser. Unter die Klassifikationsverfahren fallen das Statistisch-Dynamische Downscaling und die Analog-Methode.

Das statistisch-dynamische Downscaling ist eine Kombination beider Methoden, bei dem die Simulationen von GCMs und regionaler Modelle gemeinsam verwendet werden. Folgende Vorgangsweise wurde eingesetzt:

- Die Häufigkeit des Auftretens großskaliger Wetterlagen wird benutzt, auf das regionale Klima zu schließen. Dabei werden die Ergebnisse der Simulationen der GCMs über mehrere Jahre in Gruppen von Wetterlagen unterteilt, und für jede Klasse werden Simulationen mit einem regionalen Modell durchgeführt.
- Die Outputfelder des regionalen Modells werden je nach Häufigkeit der Wetterlage gewichtet und statistisch evaluiert, um die Verteilung von

klimatologischen Parametern (Mittelwerte, Varianzen), die der Simulation des GCM entsprechen, zu bestimmen [16].

Die Tatsache, dass die räumliche Auflösung der resultierenden regionalen Verteilungen von der begrenzten Auflösung des großskaligen Modells abhängt, erweist sich als Nachteil. Das gleiche gilt auch für das dynamische Downscaling. Die Qualität des Ergebnisses hängt sehr davon ab, ob die synoptische Variabilität der Wetterlagen erfasst wird. Dennoch bietet das statistisch-dynamische Downscaling eine effiziente Möglichkeit, regionale Trends von Klimaparametern, welche durch die Änderung von Häufigkeitsverteilungen der großskaligen Wetterlagen verursacht werden, zu erfassen [15].

Die *Analog-Methode*(AM) wurde 1999 von Zorita und Storch [61] eingeführt. Es handelt sich um ein einfaches Verfahren, bei dem das lokale Wettergeschehen aus dem großskaligen Zirkulationszustand der Atmosphäre ermittelt wird. Dies geschieht durch einen zeitlichen, sukzessiven Vergleich der großskaligen Zirkulation, wie sie von einem GCM simuliert wird, mit der in der Vergangenheit beobachteten Zirkulation. Dabei werden die ähnlichsten beobachteten Zirkulationsmuster dem simulierten Muster zugeordnet. Damit ein ähnliches Zirkulationsmuster erkannt werden kann, müssen Beobachtungen über einen ausreichend langen Zeitraum vorhanden sein. Van den Dool [45] behauptet in seiner Arbeit, dass für Wettervorhersagezwecke auf globaler Basis mehrere tausend Jahre notwendig wären, um zwei ähnliche Atmosphärezustände zu finden. Storch geht aber davon aus, dass viele der Freiheitsgrade „background noise“ repräsentieren und deshalb mittels empirischer Orthogonalfunktionen gefiltert werden können. Zusätzlich können die Freiheitsgrade weiter reduziert werden indem man sich auf ein begrenztes Gebiet, etwa in der Größe eines Kontinents, beschränkt [61]. Auf diese Annahme wird die in dieser Diplomarbeit verwendete Methode zurückgegriffen.

Zorita und Storch führten einen Vergleich der AM mit anderen, komplexeren Methoden durch [61]. Der bereits bekannte statistische Zusammenhang zwischen dem nordatlantischen SLP Feld und dem Niederschlag auf der Iberischen Halbinsel wurde verwendet [48], um die Leistung der einzelnen Methoden (lineare Methoden, CCA, Klassifikationsmethoden, CART und Neuronale Netze) zu evaluieren. Es wurde festgestellt, dass die *Analog-Methode* ähnlich gute Leistungen wie andere Methoden erbringt.

Klassifikationsmethoden, zu denen die *Analog-Methode* gehört, können im Allgemeinen den richtigen Variabilitätsgrad einer lokalen Variablen wiedergeben, wenn die Methode von historischen großskaligen Beobachtungen angetrieben wird. Sie eignen sich also nicht nur zur Darstellung Verhaltensverlaufs einer lokalen Klimavariablen, sondern darüber hinaus zur Wiederga-

be statistischer Eigenschaften.

Das Multi-layer Perceptron (MLP) ist ein Verfahren, welches zu den Transferfunktionen gezählt wird. Es fand in den letzten Jahren einige Anwendungen im Prozess des Downscalings [43]. Dieser neuronale Netzwerk-Typus kann als ein Algorithmus verstanden werden, der einen Eingabevektor durch einen schrittweisen nicht-linearen Prozess in den Ausgabevektor transformiert. Die Transformation erfolgt in zwei Schritten: Die einzelnen Komponenten des Eingabevektors werden durch eine nicht-lineare Funktion transformiert; danach erfolgt eine lineare Transformation. Der resultierende Vektor fungiert als Eingabe für die nächste nichtlineare Transformation. Dieser zweistufige Prozess kann als eine Schicht des neuronalen Netzes aufgefasst werden. Das gesamte neuronale Netz kann aus mehreren Schichten bestehen, wobei die Ausgabevektoren als Ergebnis der letzten Schicht entstehen. Durch die Verwendung von im Vorhinein bekannte Ein- und Ausgabevektorraum ist es möglich, ein Modell zu „lernen“ anzuregen. Dabei werden die Parameter der linearen Transformation durch die Minimierung des quadratischen Fehlers von bekannten und berechneten Ausgabevektoren adaptiert. Damit das Problem des *overfitting* von *Rauschen* nicht entsteht, muss die Anzahl der Parameter möglichst klein gehalten werden. Die Wahl der nichtlinearen Funktion, der sogenannten Filterfunktion hängt mit der Fragestellung zusammen. Sie sollte die Wahrscheinlichkeitsverteilung der lokalen Klimavariablen realistisch wiedergeben.

Zorita und Storch [61] haben diese Methode mit täglichen SLP Anomalien als Prediktor und täglichen Niederschlag als Prädiktor angewandt. Sie benutzten ein 3-Schichten MLP, um einen statistischen Zusammenhang zwischen den unterschiedlichen Feldern abzuleiten. Es konnte festgestellt werden, dass das 3-Schichten neuronale Netz-Modell die Häufigkeit und Intensität von starkem Niederschlag und die Häufigkeit von trockenen Tagen und somit den Variabilitätsgrad der lokalen Klimavariablen Niederschlag im Vergleich zur *Analog-Methode* schlechter wiedergibt. Dieselbe Methode wurde mit größerem Erfolg von Hewitson und Craine [24] für die genaue Beschreibung von lokalem Niederschlag aus der großräumigen Zirkulation monatlicher Werte eingesetzt. Eine weitere Anwendung für die tägliche Temperatur kann in [43] nachgelesen werden.

Wettergeneratoren sind Modelle, die statistische Eigenschaften von lokalskaligen Variablen modellieren. Sie ziehen Stichproben aus Zeitreihen und generieren auf diese Weise Tagesdaten. Es gibt zwei Arten von Wettergeneratoren: Die Markov-Kette [34] und den Verweilzeit-Ansatz [33]. Beim ersten wird bei der Simulation von aufeinander folgenden Zuständen eine Markov-

kette benutzt. Dabei sind im einfachsten Fall die Übergänge der Zustände voneinander unabhängig und die Übergangswahrscheinlichkeiten hängen nur vom aktuellen und vom vorhergehenden Zustand ab. Es wird also ein stochastischer Prozess modelliert, der nach den gegebenen Wahrscheinlichkeiten und Wettergeschehen des letzten Tages entscheidet, ob ein gewisser Tag als Niederschlagstag oder trockener Tag gilt. Entscheidet der Prozess, dass es sich um ein Niederschlagstag handelt, wird die Niederschlagsmenge aus der Wahrscheinlichkeitsverteilung gezogen [60]. Im statistischen Downscaling werden die Parameter der Wettergeneratoren vom großskaligen Zustand der Atmosphäre gesteuert und damit der lokalskalige Zustand beeinflusst [54].

Zur Leistung des dynamischen und empirischen Downscalings schreibt Storch in [47] folgendes:

"...Overall, the above comparative studies indicate that under the present climate both the dynamical and empirical techniques have similar skill. The question arises as to which is „more correct“ under future climates. While the dynamical model should clearly provide a better physical basis for change, it is still unclear whether different regional models generate similar downscaled changes, and whether the computational cost relative to statistical/empirical techniques is merited."

Kapitel 3

Die Variabilität des Klimas

Wie in den letzten Kapiteln bereits erwähnt, erwartet man mittels Downscaling sinnvolle Aussagen über die Entwicklung von Klimavariablen, wie Temperatur oder Niederschlag auf der lokalen Skala machen zu können. Dabei liegt der Schwerpunkt in der Vorhersagbarkeit der Variabilität des betrachteten Parameters. In diesem Kapitel wird deshalb auf einige methodische Aspekte des Klimaverhaltens näher eingegangen. Gleichzeitig wird das Konzept der Klimavariabilität erläutert. Im Anschluss werden Strategien, die entwickelt worden sind, um Aussagen über Klimaentwicklungen zu ermöglichen resümierend besprochen.

3.1 Klimaschwankungen und Klimaänderung

Aus der aktuellen Klimadefinition ist die Tatsache zu entnehmen, dass Klimaparameter ihrer Natur nach variierende Größen sind (siehe Kapitel 2). Aufgrund dessen ist für die Klimaforschung von besonderem Interesse, das große Variationsspektrum der betrachteten Parameter, von mehrjährigen interannuellen Schwankungen bis hin zum Alter der Erde, zu beachten.

Um das „aktuelle Klima,“ zu untersuchen, müssen die statistischen Charakteristika der Klimaparameter bezogen auf ein zeitliches Intervall so genannte Klimanormalperioden (engl.:climatic normals) bestimmt werden. Wird das zeitliche Intervall auf eine längere Dauer bezogen spricht man von Klimazustand.

Dabei zeigt ein Klimazustand unterschiedliche Variationen auf, die statistisch gesehen nicht *stationär* sind, höchstens *quasistationären* Charakter aufweisen [38]¹.

Unter dem Begriff *Variation* ist jede beliebige zeitliche Änderung eines Klimaelements zu verstehen, wie sie in den Zeitreihen² vorkommt [38]. Gates (1981) definiert Klimavariation als Übergang von einem Klimazustand zum nächsten. Von manchen Autoren wird *Variation* mit Schwankung gleichgesetzt. Die Variabilität des Klimas hängt in erster Linie mit internen Wechselwirkungen der rasch reagierenden Komponenten des Klimasystems [19] zusammen. Natürliche Klimaänderung und zugleich Änderung der Klimaparameter wird ursächlich auf externe Faktoren zurückgeführt [19]. Zu diesen Faktoren zählen z.B. die Änderung der der Erdbahn um die Sonne oder die wechselnde Strahlkraft der Sonne auf die Erde. In Abbildung 3.1 werden Beispiele einer Schwankung, Variabilität, Trends und Schwingung erläutert. Dargestellt ist eine klimatologische Temperaturzeitreihe. Eine Klimaänderung wäre vorhanden (laut Graßl [19]), wenn sich für große Zeitabschnitte die Mittelwerte stets ändern würden.

In den Zeitserien 1b-d ist die stationäre Reihe 1a durch einen einheitlichen Trend (1b), eine periodische Schwankung (1c) oder eine Sprungfunktion versehen (1d). Nähme die Variabilität in den Zeitreihen 2a, b und c zu, so gäbe es außer Trends (2b-c) auch noch systematische zeitliche Änderungen der Variabilität.

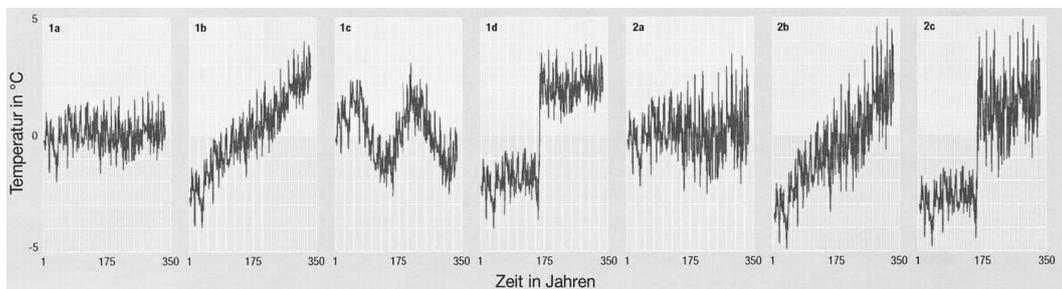


Abbildung 3.1: Klimatologische Messreihen als Zeitreihen dargestellt
(Quelle: Graßl 1993).

In Abbildung 3.2 wird eine Klassifikation zeitlicher Klimavariationstypen

¹Als *stationär* bezeichnet man jene statistischen Prozesse, dessen Momente $M_k = \frac{1}{n} \sum a_i^n$ und $ZM_k = \frac{1}{n} \sum (a_i - \bar{a})^k$ zeitlich invariant sind, wobei auch die Autokorrelationsstationarität eine Rolle spielt. Wird eine Folge von Stichprobennahmen betrachtet, so ist das im Allgemeinen nur näherungsweise der Fall, was dann als ein *quasistationärer* Prozess verstanden wird

²Zeitreihenanalyse

schematisch dargestellt. Als Trend ist im allgemeinen eine negative oder positive monotone Steigung zu verstehen. Der Sprung kann als eine abrupte Klimaänderung gedeutet werden, der auch als ein Trend verstanden wird, der zeitlich schnell stattgefunden. Weitere Klimavariationen sind die Wende oder Trendumkehrung, die einer Schwankung ähnelt und schließlich die Schwankung, die in ihrem Verhalten periodisch, zyklisch oder stochastisch sein kann. Die Schwankung ist hauptsächlich durch abwechselnde Maxima und Minima charakterisiert. Man spricht von periodischen Schwankungen (Tages- und Jahresgang), die extern erzwungen werden (etwa durch solare Strahlung) und nicht-periodischen Schwankungen, die durch interne Prozesse entstehen, oder nicht-periodische externe Einflussgrößen (z.B. Vulkanausbrüche) verursacht werden.

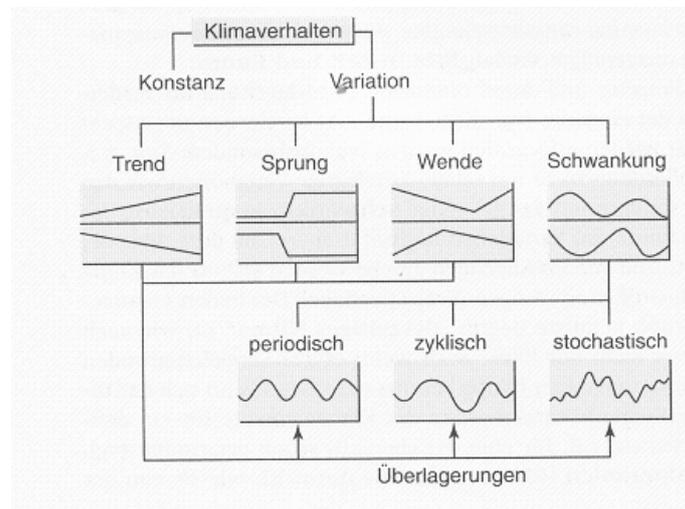


Abbildung 3.2: Schematische Darstellung der Klimavariationstypen
(Quelle: Schönwiese 1994).

Die zweite Dimension der Variationen bezieht sich auf die räumliche Auflösung, welche in Abbildung 3.3 dargestellt ist. Man unterscheidet dabei zwischen lokal („Stationsklima“) und regional bis hin zu global (siehe Kapitel 2.1.1). Schließlich wird noch eine dritte Dimension 3.3 betrachtet, welche als Ziel genannt wird. Diese Dimension steht in Zusammenhang mit der zeitlichen Dimension. Sie bezieht sich u.a. auf das Erscheinungsbild des Klimas und auf Ursachen und Auswirkungen (z.B. ökologischer Art).

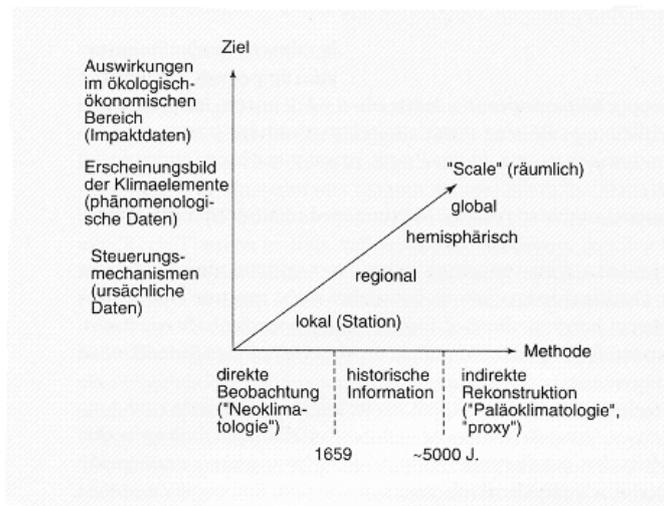


Abbildung 3.3: Zur Erläuterung der 3 Dimensionen der Klimadaten
(Quelle: Schönwiese 1994).

3.2 Klimavorhersage

In Anbetracht der komplexen Interaktionen zwischen den einzelnen Komponenten des Klimasystems stellt sich die Frage, inwieweit es möglich ist, eine Klimaänderung vorherzusagen. Bevor dies erläutert wird, muss zuerst definiert werden, was unter Klimavorhersage zu verstehen ist.

Die Definition der Klimavorhersage wird auf Ed Lorenz [29] zurückgeführt, der zwischen zwei Typen unterschieden hat. Dabei geht man von einem gewissen Zeitpunkt aus und berechnet daraus eine Klimaentwicklung. Diese Vorgehensweise ist der Wetterprognose ähnlich. Der Unterschied besteht jedoch darin, dass hier statistische Eigenschaften für eine gewisse Zeitspanne in die Zukunft zu bestimmen sind, während bei der Wetterprognose der aktuelle Zustand der Atmosphäre berechnet wird. In dieser Form der Klimavorhersage wird erwartet, dass die Randbedingungen bekannt sind. Sie entspricht der Vorhersage der internen Klimavariabilität (z.B. Modellierung des Klimasystems bei konstantem Antrieb oder unter verschiedenen Randbedingungen).

Beim zweiten Typ geht es um die Erforschung der Reaktion des Klimasystems auf externe Faktoren (z.B. Treibhausgasemissionen, Vulkanismus). In diesem Fall kann nicht von einer Vorhersage im Sinn der Wettervorhersage gesprochen werden, da auf den Entwurf sogenannter Klimaszenarien abgezielt wird. Ein Klimaszenarium bezeichnet man hierbei als einen Klimazustand, der sich aufgrund einer Änderung gewisser Einflussgrößen über einen bestimmten Zeitraum einstellt. Für den Entwurf von Klimaszenarien

sind Klimamodelle von grosser Wichtigkeit. Ursprünglich wurden Klimamodelle entwickelt um die Dynamik des Klimasystems zu erforschen. Als Ausgangspunkt für Szenarien gelten Annahmen über die zukünftige Entwicklung externer Faktoren (z.B. den Anstieg der atmosphärischen CO₂-Konzentration). Damit wird erhofft, realitätsnahe Modelle zu betreiben, die in der Lage sind zukünftige Klimaänderungen zu prognostizieren.

Von grosser Wichtigkeit ist die Konkretisierung der Szenarien auf der regionalen Skala. Da sich das regionale Klima aus der Wechselwirkung von großskaligem Zustand der Atmosphäre und regionalen Gegebenheiten ergibt, ist zu erwarten, dass durch sinnvolle Methoden ein Zugang zu regionalen Szenarien geschaffen wird (Downscaling) (siehe Abschnitt 2.1.1).

Bei der Entwicklung von Szenarien gibt es verschiedene Ansätze:

1. Die von einem Klimamodell simulierten großräumigen Änderungen werden im Mittel für lokale Klimaparameter als richtig angenommen. Unter dieser Annahme werden langjährige Beobachtungsreihen mit bestimmten statistischen Methoden aufbereitet, sodass sie die entsprechenden Änderungen in Form von Szenarien realistisch wiedergeben. Der Vorteil dieser Methode liegt in der Reduzierung des Klimamodellfehlers bei der Szenarienbildung. Als Nachteil gilt der nicht eindeutige physikalische Zusammenhang zwischen Modellergebnissen und Szenarien.
2. Ein regionales Klimamodell wird in einem globalen Modell eingebettet („nesten“). Dabei dienen die Resultate des globalen Modells als Randbedingungen für das regionale Modell [30]. Diese Vorgehensweise hat den Vorteil einer physikalischen Verknüpfung zwischen großskaligen und kleinskaligen Prozessen.
3. Statistische Verfahren werden benutzt, Resultate von Klimamodellen auf einen kleinskaligen Bereich zu transferieren [61], wodurch direkt auf die Resultate der Klimamodelle zugegriffen werden kann. Eine Fehlerfortpflanzung, wie sie üblicherweise durch die Kopplung verschiedener Klimakomponentenmodelle entsteht, wird in diesem Fall ausgeschlossen. Tatsache ist ferner, dass die statistischen Ansätze den Fehler des Klimamodells nicht beseitigen können, was sich nachteilig auf die Szenarien auswirken kann.

Kapitel 4

Niederschlag in Österreich

In diesem Kapitel werden jene Großwetterlagen beschrieben, die mit dem Entstehen von Niederschlag in Mitteleuropa zusammenhängen¹. Wenn man bedenkt, dass der Niederschlag in unseren Breiten meistens aufgrund frontaler Systeme entsteht [41], ist es sinnvoll, die wichtigsten atmosphärischen Konstellationen zu beschreiben, welche in großem Maß zur Niederschlagsentstehung beitragen. Dadurch soll ein Überblick über die klimatischen Bedingungen, die mit den Wetterlagen in Verbindung stehen und somit das Wettergeschehen in Mitteleuropa stark beeinflussen, gegeben werden. Eine Großwetterlage wird als ein atmosphärischer Zustand verstanden, in dem die großskaligen Prozesse sich nur sehr langsam verändern. Die *Analog-Methode*, so wie sie hier eingesetzt wird, setzt gerade diesen großskaligen Anteil der Atmosphäre mit dem lokalen Klima statistisch in Verbindung.

4.1 Die Grosswetterlagen in Mitteleuropa

Das Klima in einer Region wird einerseits durch ihre Lage in der Klimazone und andererseits durch den charakteristischen Jahresverlauf der Witterung bestimmt. Aufgrund der Änderungen der Sonneneinstrahlung, welche jahreszeitlich bedingt sind, ändert sich die Verlagerung der Hoch- und Tiefdruckgebiete. Dadurch entsteht für jede Region ein charakteristischer Jahresverlauf der Witterung. Auf diese Weise besteht der Jahresverlauf aus einer Abfolge von Witterungssituationen, welche als Großwetterlagen bezeichnet werden. Die genaue Definition einer Großwetterlage lautet nach Baur (1963) [2] folgendermaßen:

¹die hier herangezogenen Großwetterlagen werden in Anlehnung an den Katalog der Großwetterlagen Europas (1881-1998) von Paul Hess und Helmuth Brezowsky beschrieben. Diese Klassifikation ist auf Mitteleuropa bezogen.

"...die mittlere Luftdruckverteilung eines Großraumes, mindestens von der Größe Europa während eines mehrtägigen Zeitraumes, in welchem gewisse Züge aufeinanderfolgender Wetterlagen gleichbleiben, eben jene Züge, welche die Witterung in den einzelnen Teilgebieten des Großraums bedingen"

Die Idee, dass bestimmte ähnliche Strukturen des atmosphärischen Wind- und Massenfeldes sich wiederholen könnten, ließ verschiedene Großwetterlagen-Klassifikationen entstehen. Die bekannteste für den europäischen Raum ist jene nach Hess/Brezowsky, dessen Großwetterlagen-Gruppierung auf die Definition von Baur gründet. Für jeden Tag wurde aufgrund der Position von herrschenden Druckgebilden und der Erstreckung von Frontalsystemen über Europa eine „Wetterlage“ ausgewählt. Speziell für den Alpenraum und den Ostalpenraum wurden andere Klassifikationen entwickelt: die ostalpine Klassifikation von Lauscher [27], die Klassifikation der Witterungslagen nach Schüepp [39] und die Ostalpine Klassifikation von Steinacker [41]. Im Gegensatz zu der Hess/Brezowsky- und Lauscher-Klassifikation, die in der Klassifikation die Bodendruckverhältnisse heranziehen, verwendet Schüepp die 500-hpa Fläche.

Im Katalog der Großwetterlagen von Hess/Brezowsky wird unterschieden zwischen drei Zirkulationsformen: eine breitenparallele (zonal), kreisparallele (meridional) und gemischte Formen. Innerhalb dieser Zirkulationsformen wird zusätzlich nach vorherrschenden Winddrehung (zyklonal und antizyklonal) unterschieden. Insgesamt wurden nach diesen Kriterien 29 Großwetterlagen identifiziert. Dabei ist die am häufigsten auftretende Großwetterlage mit 15% die *Westlage*, die zu den zonalen Zirkulationstypen gehört überwiegend *zyklonal*. Sie ist für das Ziehen von Zyklonen in östliche Richtung verantwortlich, was folglich zu Niederschlägen in Mitteleuropa führt. Mit 9.8% tritt die Großwetterlage *Hoch über Mitteleuropa* auf, welche zu den gemischten Zirkulationstypen gehört.

Eine Großwetterlage manifestiert sich durch eine beinahe stationäre Zugbahn der Druckanstiegs- oder Druckfallgebiete. Das großräumige Druckfeld in Europa selbst wird von dem Azorenhoch in Südwesteuropa, den quasistationären Islandtief im Nordwesten, Hitzetiefs in Südwesteuropa und aus Sibirien stammenden kontinentalen Antizyklonen beeinflusst. Das Azorenhoch gehört zum subtropischen Hochdruckgürtel und sorgt meistens im Sommer für Schönwetter und trockene Phasen. Zusammen mit dem Islandtief bewirkt es Westwetterlagen, die für Sturmsituationen in unseren Breiten verantwortlich sind.

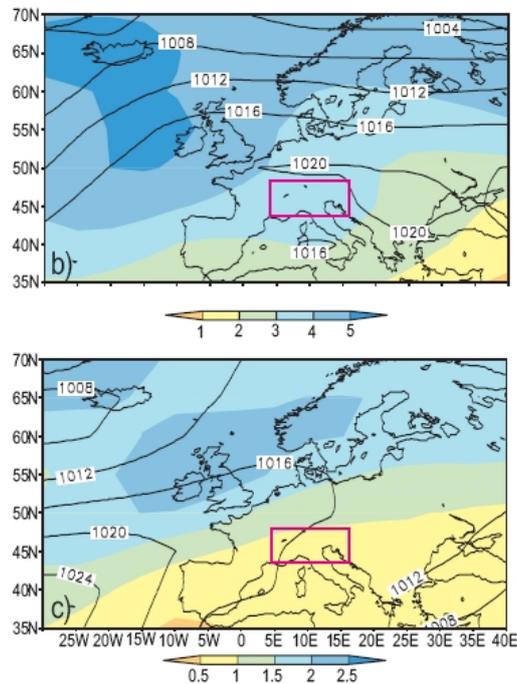


Abbildung 4.1: Isobaren (linien) und Standardabweichung (Flächen) des mittleren Druckes in Meereshöhe für Europa im Winter(DJF)(oben) und im Sommer(JJA)(unten) für die Jahre 1901-1999 (*Quelle: Wanner 2003*).

In der Regel sorgen Westwetterlagen im Winter mit dem Durchzug atlantischer Tiefs für starke Niederschläge über Mitteleuropa. Abbildung 4.1 zeigt anhand der Standardabweichung des Luftdruckes, dass die Zahl der Tiefs im Winter höher ist als im Sommer. Gleichzeitig überdecken die Zugbahnstrassen der Tiefdruckgebiete in der kalten Jahreszeit ein viel größeres Gebiet, welches bis zum Mittelmeerraum reicht, als in der warmen Jahreszeit. Im folgenden werden einige Grosswetterlagen beschrieben, die für das Auftreten von Niederschlag in unseren Breiten verantwortlich sind.

- Dazu zählt die *Westlage* überwiegend *zyklonal* 4.2, die sich durch stationäre subtropische Antizyklone, das Azorenhoch und eine Zyklone zwischen Island und Skandinavien, das Islandtief, charakterisiert. Diese Lage zeigt eine ausgeprägte Westdrift, in dessen normaler Lage eine Frontalzone erkennbar ist. Diese befindet sich zwischen 50 und 60N, verläuft also vom Seegebiet westlich von Irland über Großbritannien und die Nordsee bis hin nach Osteuropa. Das in normaler Lage be-

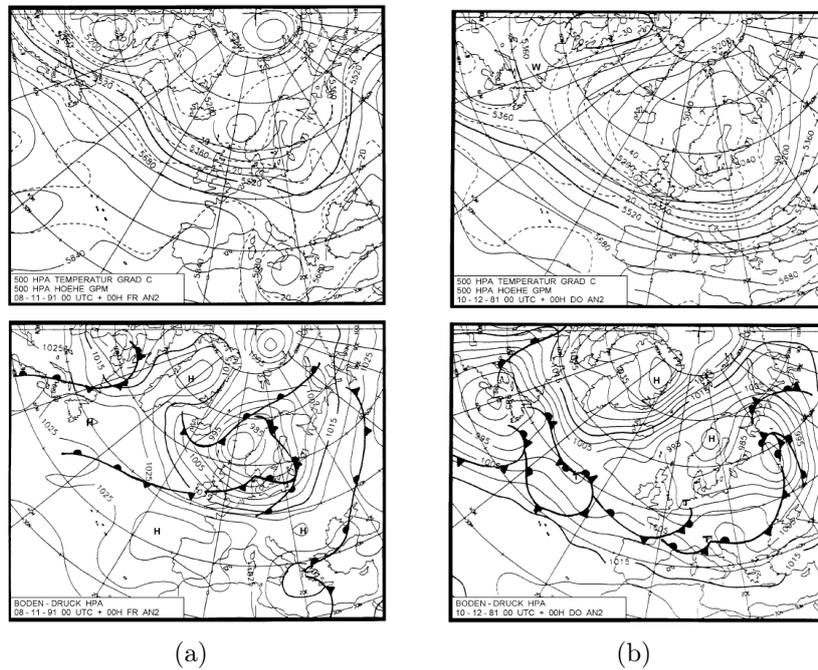


Abbildung 4.2: Westlage über Mitteleuropa überwiegend zyklonal(a) und Südlicher Westlage(b) (Quelle: Hess und Brezowsky 1999).

findliche Azorenhoch kann Süditalien und den alpinen Raum erreichen. In weiterer zeitlicher Entwicklung sind für diese Wetterlage charakteristisch: rasch in östlicher Richtung abziehende Tiefs, begleitet von den zugehörigen Fronten, Wolkenfeldern und zeitweiligem Niederschlag. Meistens wechseln anhaltende, schauerartige Niederschläge mit Aufhellungen, die einen ganzen Tag andauern können. Eine weitere Eigenheit sind Westwinde, welche maritime Luftmassen nach Mitteleuropa verfrachten. Im Sommer entstehen kühle Phasen, im Winter bewirken sie milde Temperaturen, wobei sie häufig Niederschläge mit sich bringen. Der Wind weht aus westlicher Richtung und kann auch sturmisch sein. Diese Wetterlage gilt als die häufigste auftretende Wetterlage in Mitteleuropa und tritt besonders häufig im August, selten im April auf.

- Eine weitere Großwetterlage, die in unseren Breiten für reichen Niederschlag sorgt, ist die *Südliche Westlage* (siehe Abbildung 4.2). Am Musterbeispiel dieser Wetterlage sind Einzelstörungen erkennbar, die sich südwestlich von Irland über Frankreich, Deutschland, nach Osteuropa mit anschließender Biegung nach Norden ausbreiten. Der zyklonale

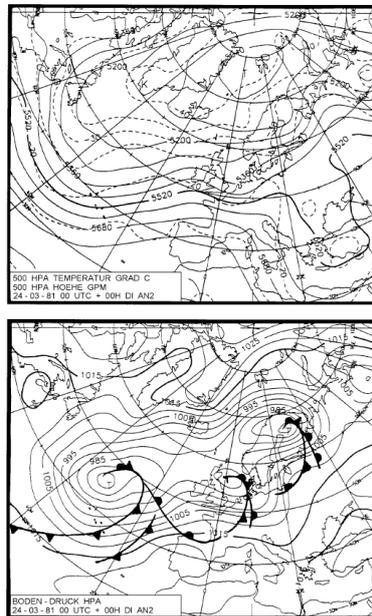


Abbildung 4.3: Südwestlage überwiegend zyklonal (Quelle: Hess und Brezowsky, 1999).

Einfluss erreicht oft den nordöstlichen Bereich des Mittelmeers. Ein ausgeprägtes Bodentief befindet sich südlich des 60°N Breitengrades, und führt dazu, dass der nördliche Teil des Atlantiks und Teilbereiche des Nordmeeres von einem kalten Polarhoch mit einer östlichen Strömung beeinflusst werden. Das Azorenhoch erreicht meistens nur Nordafrika. Diese Wetterlage tritt im Dezember und weniger oft im September auf.

- Die *Südwestlage*, überwiegend *zyklonal* (siehe Abbildung 4.3) ist charakterisiert durch hohen Luftdruck über der Ukraine mit Ausläufern zum Mittelmeer bei gleichzeitig tiefem Luftdruck über dem mittleren Nordatlantik bis hin nach Irland. Dazwischen verläuft eine nordwärts gerichtete Frontalzone, die sich vom Seegebiet nördlich der Azoren über das südliche Nordmeer bis zum Baltikum erstreckt. Darin eingebettet sind Einzelstörungen erkennbar, welche von der Biskaya über Skandinavien bis hin zum Eismeer wandern. Diese Wetterlage ist häufig im Januar anzutreffen und sehr selten im Juli.
- Eine weitere Großwetterlage, welche in jeder Jahreszeit für überdurchschnittlichen Niederschlag in Mitteleuropa sorgt, ist die *Nordwestlage*,

überwiegend *zyklonal*, 4.4. In dem hier gezeigten Beispiel befindet sich eine ausgeprägte Frontalzone zwischen einem ausgedehnten Tiefdrucksystem, das über Schottland bis hin nach Skandinavien reicht und einem nordwärts gerichteten Subtrophenhoch, welches sich bis zur westlichen Biskaya ausdehnt. Diese Frontalzone verläuft über Großbritannien, der Nordsee, dem östlichen Mittelmeer bis hin nach Südosteuropa in einem osteuropäischen Trog. In ihr wandern Fronten mit den dazugehörigen Zyklonen wodurch Niederschlagsgebiete in östlicher Richtung verfrachtet werden. Über Norditalien kommt es häufig zur Bildung von Teilstörungen. Diese Wetterlage kommt vor allem im Sommer vor, häufiger im Juli und weniger häufig im Mai.

- Im Fall der *Nordlage*, überwiegend *zyklonal*, liegt ein mächtiges blockierendes Hoch über dem östlichen Nordatlantik und ein ausgedehntes Tiefdrucksystem über Skandinavien und dem Baltikum. Die vorhandene Frontalzone über dem Atlantik ist nordostwärts gerichtet. Erkennbar ist auch, dass in der Frontalzone, die sich vom Mittelmeer zum Nordmeer erstreckt, Einzelstörungen nach Mitteleuropa gelangen. Das Vorstoßen von Kaltluft nach Mitteleuropa führt oft zur Entstehung von Störungen, die in ihrem Verlauf nach Nordosten ziehen. Daraus resultiert eine wechselhafte Witterung, die oft von böigen nördlichen Winden begleitet wird. Diese Wetterlage tritt wie die vorige im Sommer (Juni) auf und kommt im Winter weniger häufig (Dezember) vor.

Abgesehen von den Wetterlagen, die mit ihren spezifischen Zirkulationsmustern eine wichtige Einflussgröße sind (in der Praxis der synoptischen Meteorologie ist der Zusammenhang zwischen Großwetterlage und den Klimaparametern Temperatur und Niederschlag bekannt), beeinflussen regionale Gegebenheiten die Entstehung von Niederschlag ebenfalls.

4.2 Niederschlag

4.2.1 Vb-Wetterlage

Eine Wetterlage, die in erster Linie mit starken Niederschlägen in Mitteleuropa in Verbindung gebracht wird, ist die nach Van Bebber benannte *Vb-Lage*. Sie ist gekennzeichnet durch die Zugbahn eines Tiefes von Genua über die Poebene, um die Alpen über Österreich und Ungarn nach Polen. In Abbildung 4.5 wird die *Vb-Lage* durch das Umbiegen einer aus Nordwesten kommende Frontalzone über das südliches Mitteleuropa nach Nordosten

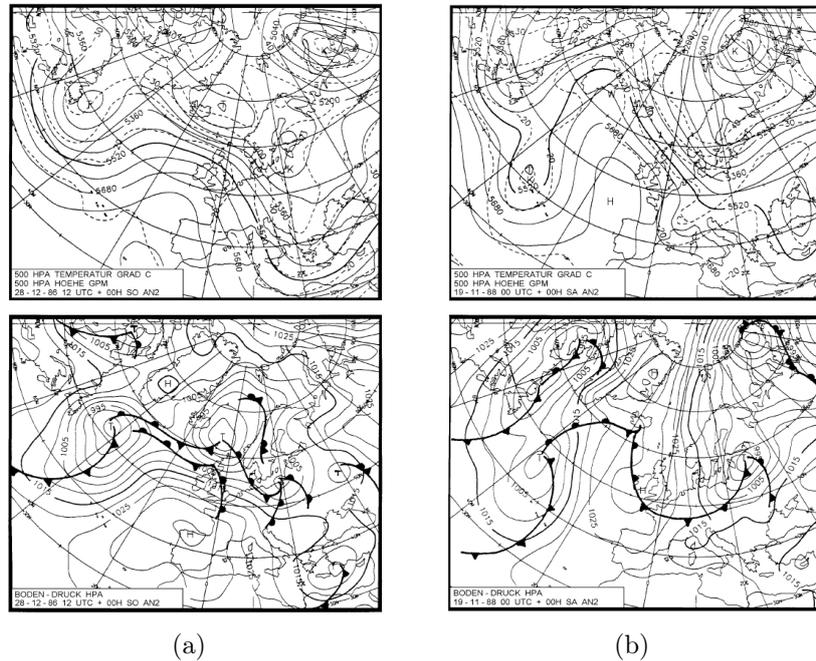


Abbildung 4.4: Nordwestlage überwiegend Zyklonal(a) und Nordlage überwiegend Zyklonal(b) (Quelle: Hess und Brezowsky, 1999).

dargestellt. Einzelstörungen werden zunächst abgeschwächt, über dem Mittelmeer gewinnen sie wieder an Intensität. Das *Vb-Tief* entsteht über dem Golf von Genua und wird durch einen Kaltluftvorstoß über dem westlichen Mittelmeer, oft in Verbindung mit einer *Norwestwetterlage*, ausgelöst. Diese Wetterlage bewirkt vor allem nördlich der Alpen und im Alpenraum ergiebige Niederschläge. Das größte Hochwasserereignis im August 2002, von dem Österreich und weitere Teile Mitteleuropas getroffen wurden, wurde durch ein sich nordwärts verlagerndes Adriatief verursacht. Die *Vb-Lage* kann in jeder Jahreszeit vorkommen, aber sie ist besonders häufig im Frühling und Herbst. Im Sommer kann sie für Österreich aufgrund der hohen Temperaturen besonders niederschlagsreich sein.

4.2.2 Niederschlagsverhalten und Grosswetterlage

Die Niederschlagsbildung in einer bestimmten Region hängt zunächst grossräumig gesehen von der herrschenden Großwetterlage ab. In [12] werden Mithilfe der Hauptkomponentenanalyse (Principal Component Analysis, PCA) anhand von Niederschlagsdaten (Stationsdaten) für die Periode 1951-1980,

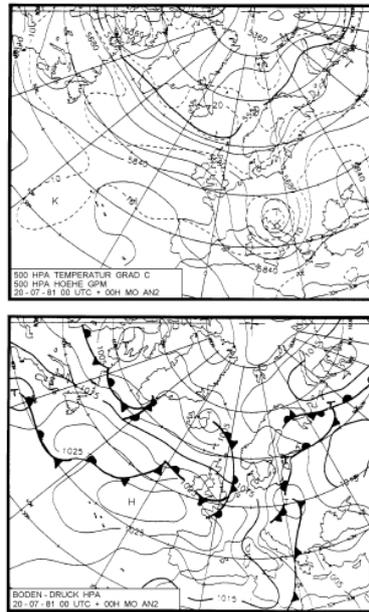


Abbildung 4.5: Trog Mitteleuropa (Vb-Lage)(Quelle: Hess und Brezowsky 1999).

3 homogene Niederschlagsregionen in Österreich gefunden. In Abbildung 4.6 wird schematisch die Einteilung Österreichs für das Winterhalbjahr und das Sommerhalbjahr dargestellt. Ehrendorfer wies den gefundenen homogenen Niederschlagsregionen den Einfluss bestimmter Wetterlagen² zu. Dabei wird die gleiche Wetterlage mehreren Regionen zugewiesen. Als Grund wird der gleichzeitige Einfluss einer Wetterlage auf die identifizierte homogene Niederschlagsregione angenommen.

Im Folgenden wird ein Überblick über die erwähnte Zuweisung gegeben werden:

- der West-Nordwest-Region: wurden die Nordlage, Nordwestlage, Westlage, Tief über Kontinent+Tiefdruckrinne, Tief über den Britischen Inseln+Tief im Suedwesten, *Vb-Lage* zugeordnet.
- der Ost-Nordost-Region: die Nordlage, Nordwestlage, *Vb-Lage* zugeordnet

²dafür wurde die Ostalpine Wetterklassifikation von Lauscher(1972) benutzt. Sowohl diese Klassifikation wie die von Hess-Brezowsky betrachtet nur die Bodendruckverteilung. Der Unterschied zwischen den beiden Varianten besteht darin, dass die Klassifikation von Lauster auf den Ostalpinen Raum zugeschnitten ist

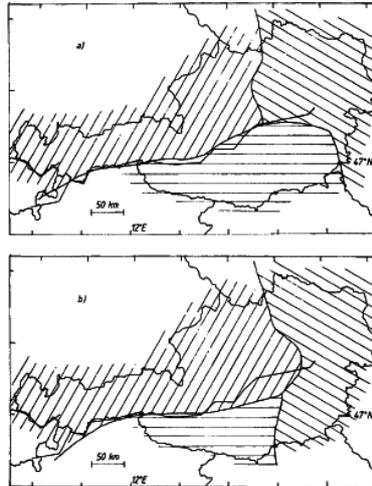


Abbildung 4.6: Unterteilung von Österreich in drei homogene Niederschlagsregionen, durch Anwendung von PCA S-Mode für den Winter(a) und Sommer(b). Die Isolinien würden aus den Linien gleicher PC Loadings ermittelt (Quelle: Ehrendorfer 1985).

- der Süd-Südwest-Region: wurden die Südwestlage+Südlage, Tief über Britischen Inseln+Tief im Südwesten, Tief im Schweden ostwärts ziehend + Tief über dem Westlichen Mittelmeer, *Vb-Lage* zugeordnet.

Diese Zuweisung wird in Anlehnung an die Arbeit von Steinhauser [42], der die Auswirkungen von Wetterlagen auf Niederschlagsverhältnisse untersuchte und dafür die Klassifikation der Grosswetterlagen im Ostalpenraum von Lauscher(1972) herangezogen hat, durchgeführt. Steinhauser konnte aufgrund der Ergebnisse von Niederschlagswahrscheinlichkeiten bei unterschiedlichen Wetterlagen herausfinden wo bestimmte Wetterlagen besonders einflussreich waren und wo ihr Einfluss auf den Niederschlag am schwächsten waren.

4.2.3 Niederschlagsarten

Folgende Arten der Niederschlagsentstehung können unterschieden werden:

Orographische Niederschläge findet statt wenn feuchte Luft, stammend aus dem Meer oder Flachland durch Gebirgsketten und andere Orographische Erhebungen zum Aufsteigen gezwungen werden. Mit aufsteigender Höhe wird die Luft weiter abgekühlt.

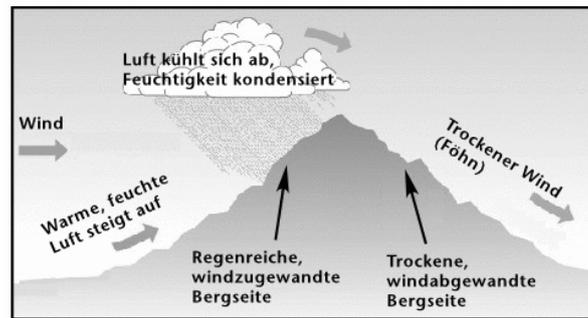


Abbildung 4.7: Orographische Niederschläge

Gleichzeitig nehmen die Temperatur und die Wasserdampfkapazität ab, bis der Taupunkt erreicht wird. Die Abkühlung der Luft erfolgt zuerst trockenadiabatisch, d.h. die Abkühlung erfolgt mit die Abnahme eines Grad Celcius pro hundert Meter. Erreicht die relative Luftfeuchtigkeit 100%, wird die Luft feuchtadiabatisch abgekühlt. Das entspricht einer Abnahme der Temperatur um 0.6°C pro 100 Meter. Dadurch kondensiert die Feuchtigkeit vom gasförmigen zu flüssigen Zustand. Dieser Prozess führt zur Freisetzung von latenter Wärme und bewirkt die Wolkenbildung. Je nach vorhandene Feuchtigkeit und Aufgleitbewegung können starke Niederschläge auftreten. Die kondensierte Feuchte bleibt für eine gewisse Zeit an der Luv-Hang Seite des Gebirges hängen. An der Lee-Seite eines Gebirges kommt es im Durchschnitt zu geringeren Niederschlag; nach der Abregnung erfährt die Luft eine trockenadiabatische Aufwärmung, was zu Föhn führen kann.

Konvektive Niederschläge: entstehen aufgrund Konvektion in der Atmosphäre; durch eine Erwärmung der bodennahen Luftschichten wird die Luft instabil geschichtet. Das führt zu vertikalem Auftrieb, und damit zum Aufsteigen von Warmluft. Enthält die aufsteigende Warmluft Feuchte, so kommt es zur Kondensation und Wolkenbildung. Im Sommer kommt diese Art von Niederschlag in unseren Breiten sehr häufig vor, meistens als Schauer oder Gewitterregen, der oft intensiv aber von kurzer Dauer ist. In Österreich kommen konvektive Niederschläge mehr in den Zentralalpen als in der alpine Randzone auf, meistens kann der Regen über einige Stunden andauern. Die Menge des konvektiven Niederschlags ist im Schnitt fünf mal größer als die des zyklonalen Niederschlags.

Zyklonale Niederschläge: umfassen die Niederschläge die durch die

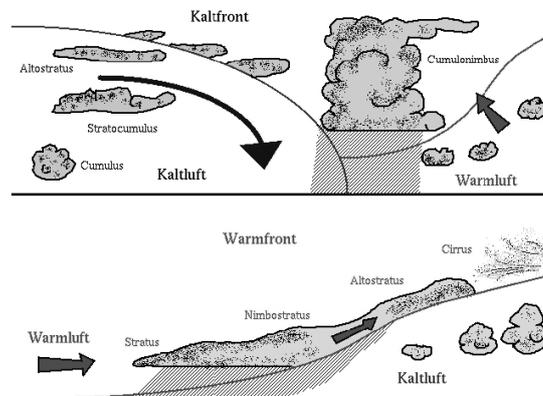


Abbildung 4.8: Frontale Niederschläge

Abfolge von Kalt- und Warmfronten ausgelöst werden und mit den Grosswetterlagen in engen Zusammenhang stehen. Handelt es sich um eine vorbeiziehende Warmfront, gleitet die warme Luft langsam über die Kaltluftmasse. Die Niederschläge sind meistens schwach, können aber ausgedehnte Gebiete umfassen und verweilen kurz. Wenn eine Kaltfront durchzieht, wird die Warmluftmasse zu schnellem Aufsteigen gezwungen. Die Front „begleitet“ hier die Niederschläge. In diesem Fall kommen Frontengewitter zustande, die wie beim Wärmegewitter zu starken Niederschlägen führen. Diese Art dauert meistens einige Tage an und umfasst größere Gebiete als der konvektive Niederschlag.

Niederschläge durch nächtliche Abkühlung Aufgrund des Wärmeverlust durch Abstrahlung und der fehlenden Sonneneinstrahlung vor allem dann, wenn die Luft sehr feucht ist, kondensiert sie, was zu Tau oder Nebel führen kann. Dadurch kann es zu Niederschlägen kommen, vor allem an den Luvseiten mancher Gebirge.

4.3 Lokale Zeitreihen des Niederschlags

Als lokale Beobachtungen werden die StartClim Stationsdatensätze von Schöner(2003) [37] verwendet. Diese bestehen aus der täglichen mittleren Temperatur und dem täglichen akkumulierten Niederschlag von 70 Stationen und umfassen die Periode von 1948 - 2002. Die Tagesmittelwerte der Lufttemperatur werden aus dem arithmetischen Mittel des Lufttemperaturmaximums

und -minimums ermittelt. Der tägliche akkumulierte Niederschlag errechnet sich aus der Summe der täglichen Messungen zwischen 19 Uhr und 7 Uhr. Die Temperatur wie auch der Niederschlag wurden einer Qualitätskontrolle unterzogen. Ein Vergleich von StartClim Datensätzen zu den homogenisierten Austrian Long-Term Climate(ALOCLIM) -Monatsdaten (Auer und Böhm et al 2001) wurde von Haas (2004) durchgeführt. Für die Homogenisierung der Temperatur wurde von Schöner eine Methode zur Prüfung der Datenqualität entwickelt. In dieser Arbeit werden nur die Niederschlags-Stationsreihen für das Downscaling mit der *Analog-Methode* benutzt. Tabelle 4.1 zeigt die geographische Lage der zum Startclim-Stationsdatensatz gehörenden Stationen.

4.4 Reanalyse-Datensätze

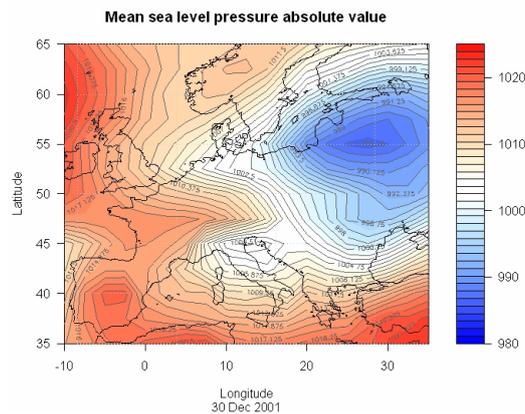


Abbildung 4.9: Ausgewählter geographischer Ausschnitt von -10°W bis 35°E und 35°N bis 65°N für die ERA40 Datenfelder.

In dieser Arbeit werden Reanalyse-Daten des ECMWF(European Centre for Medium-Range Weather Forecasts) herangezogen. Die Reanalyse-Daten entstammen einem Projekt, bei dem Messungen und Beobachtungen (Messtationen, Bojen, Flugzeumessungen, Satellitendaten. . .) unterschiedlicher Art von einem Modell assimiliert werden. Es handelt sich um ein zwei-Phasen-Projekt, bei dem als erstes Re-Analysen generiert werden (von 1957 bis 1994 (ERA15) und anschließend von 1957 bis 2001 (ERA40)).

Durch die Assimilation werden Inhomogenitäten in den Daten beseitigt. Aus der Assimilation entsteht die Analyse. Reanalyse bedeutet die Berechnung von Analysen aus der Vergangenheit durch Einsetzung von Interpolat-

Name	lon	lat	alt	Name	lon	lat	alt
Feldkirch	9.61	47.28	439	Kollerschlag	13.84	48.61	725
Bregenz	9.76	47.51	436	Kanzelhöhe	13.91	46.68	1526
Schopperrau	10.02	47.31	835	Irdning	14.10	47.51	710
Schröcken	10.09	47.26	1263	Kremsmünster	14.13	48.06	383
Galtuer	10.19	46.98	1583	Hörsching	14.19	48.24	298
Holzgau	10.35	47.26	1100	Stolzalpe	14.21	47.12	1305
Landeck	10.57	47.14	818	Loibl	14.25	46.45	1098
Reutte	10.76	47.51	870	Klagenfurt	14.34	46.65	447
Obergurgl	11.03	46.87	1938	Freistadt	14.50	48.52	548
Innsbruck Flughafen	11.36	47.26	579	Grossraming	14.52	47.89	379
Innsbruck Uni	11.38	47.26	577	Hiefrau	14.76	47.61	492
Patscherkofel	11.46	47.21	2247	St.Michael Bleiburg	14.76	46.58	500
Jenbach	11.76	47.39	530	Seckau	14.78	47.28	874
Mayrhofen	11.85	47.15	643	Zeltweg	14.78	47.21	669
Kufstein	12.16	47.57	492	Pabneukirschen	14.82	48.32	595
St.Jakob-Def	12.35	46.92	1400	Preittenegg	14.92	46.93	1055
Mooserboden	12.72	47.15	2036	Lobming	15.19	47.05	400
Lienz	12.78	46.82	668	Stift Zwetl	15.21	48.63	506
Zell am See	12.78	47.30	753	Bruck Mur	15.27	47.42	489
Sonnblick	12.96	47.05	3105	Mariazell	15.32	47.77	875
Rauris	13.00	47.22	945	Graz Flughafen	15.44	46.99	340
Salzburg Flughafen	13.00	47.80	430	Graz uni	15.45	47.08	366
Badgastein	13.13	47.12	1100	Schöckl	15.47	47.20	1436
Reisach	13.15	46.65	646	Krems	15.61	48.42	223
Kolbnitz	13.31	46.87	603	St.Pölten	15.63	48.20	273
Mondsee	13.37	47.85	491	Mürzzuschlag	15.69	47.60	700
Reichersberg	13.37	48.34	350	Gleisdorf	15.71	47.11	375
Bad Ischl	13.63	47.72	469	Reichenau	15.85	47.70	485
Villacher Alpe	13.67	46.60	2140	Bad Gleichenberg	15.91	46.87	303
Krippenstein	13.70	47.52	2050	Retz	15.95	48.77	242
Feuerkogel	13.73	47.82	1618	Wörterberg	16.10	47.23	400
Bad Aussee	13.78	47.63	675	Mariaabrunn	16.23	48.21	226
Tamsweg	13.81	47.12	1012	Hohe Warte	16.36	48.26	203
Oberleis	16.37	49.56	420	Eisenstadt	16.56	47.86	159
Laa Thaya	16.39	48.73	187	Schwechat	16.57	48.11	184
				Hohenau	16.90	48.62	155

Tabelle 4.1: Messstellen der Startclim-Datensätze.

tionstechniken und Vorhersagemodellen und ermöglicht dadurch eine räumliche und zeitliche Konsistenz der Daten.

Gewählt wurden die 12-stündigen ERA40-Datensätze für die Periode von 1958 bis 2000. Die ERA40-Daten werden mit einem spektralen Modell (T159) mit einer horizontalen Auflösung von $2.5^\circ \times 2.5^\circ (\approx 120\text{km})$ erzeugt.

Um die Evaluierung der Distanzmaßen mit der Analog-Methode umsetzen zu können, wurde überlegt, welche meteorologische Parameter als Prädiktoren auf der GCM-Skala sich aus physikalischer Sicht als geeignet erweisen, Niederschlag zu modellieren. Außerdem sollte der gesuchte Zusammenhang aus der Synoptik bekannt sein. Der mittlere Luftdruck auf Meereshöhe und die Spezifische Feuchte in 700hPa Niveau werden als **Prediktoren auf der GCM-Skala** gewählt.

Spezifische Feuchte: Diese wird definiert als die Masse des Wasserdampfes zur Masse der feuchten Luft im gleichen Volumen, oder als das Verhältnis der Wasserdampfdichte zur Dichte der feuchten Luft. q wird auch als niederschlagsfähiges Wasser (precipitable water) bezeichnet [1]

$$q = \frac{M_W}{M_L + M_W} = \frac{M_W}{M} = \frac{\rho_W}{\rho} \quad (4.1)$$

wobei:

$$\rho = \rho_L + \rho_W \quad (4.2)$$

Da die Luft ein Gasgemisch ist und sowohl feuchte wie auch trockene Anteile hat, gilt für jeden Luftbestandteil der Partialdruck:

$$p_i = R^* \rho_i^* T \quad (4.3)$$

Der Gesamtdruck p lässt sich dann schreiben als:

$$p = \sum_i p_i = R^* \left(\sum_i \rho_i^* \right) T \quad (4.4)$$

wobei gilt:

$$\sum_i \rho_i^* = \rho^* \quad (4.5)$$

Durch die Einsetzung der Spezifischen Feuchte in die Gas-Gleichung ergibt sich:

$$p = R\rho T \quad (4.6)$$

folgt:

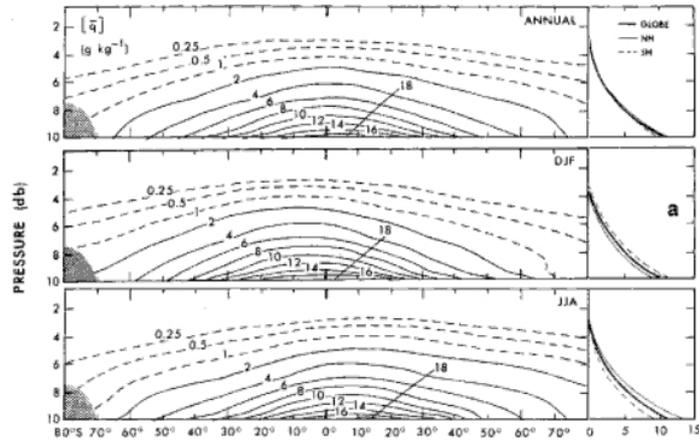


Abbildung 4.10: Zonales Mittel der spezifischen Feuchte für das Jahr, die Wintermonate DJF und die Sommermonate JJA in Abhängigkeit der Meridionalen Koordinate und des Drucks in dB (Quelle: Peixoto 1992).

$$p = p_W + p_L = \left(1 + \frac{R_W + R_L}{R_L} q\right) R_L \rho T \quad (4.7)$$

Mit dem Feuchtemaß Dampfdruck e lässt sich die spezifische Feuchte auch schreiben:

$$e = R_W \rho_W T \quad (4.8)$$

$$q = 0.622 \frac{e}{p - 0.377e} \quad (4.9)$$

wobei ρ_W die Dichte des reinen Wasserdampfes (absolute Feuchte) und ρ_L die Dichte der reinen Luft ist. Die spezifische Feuchte q hat die Dimension eins. Als Einheit wird $[q] = \text{kg kg}^{-1}$ (ERA40) verwendet.

Auf der **700 hPa**-Fläche kann q als Hinweis auf Frontalsysteme sein. Abbildung 4.10 gibt ein Bild der Verteilung der mittleren spezifischen Feuchte in Abhängigkeit von den Breitenkreisen und der Höhe bezüglich Jahr, Winter- und Sommermonate. Im Vergleich zu anderen Feuchtemaßen ist q ein *konervatives Feuchtemaß*, welches bei Vertikalbewegungen von Luftmassen und damit verbundenen Druckänderungen konstant bleibt, bis der Taupunkt eintritt und deswegen Kondensation stattfinden muss. Die spezifische Feuchte

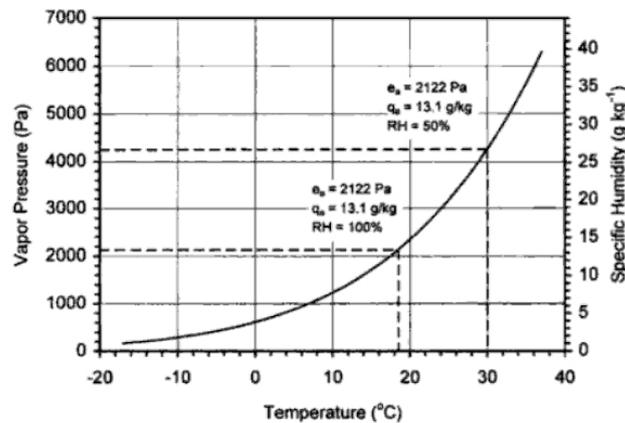


Abbildung 4.11: Sättigungsdampfdruck als Funktion der Temperatur. Auf der linken y-Achse ist der Dampfdruck und auf der rechten y-Achse die spezifische Feuchte aufgetragen. Die gestrichelte Linien zeigt den Dampfdruck bei 18,5°C und 30°C. In diesem Beispiel hat das Luftpaket mit dem Dampfdruck $e = 2122 Pa$ eine relative Feuchte(RH) von 100% bei 18,5°C und 50% bei 30°C (Quelle: Bonan 2002 [5]).

zusammen mit dem Mischungsverhältnis spielen eine große Rolle bei thermodynamischen Zustandsänderungen der Luft aufgrund vertikaler Austauschvorgänge.

In Abbildung 4.11 ist der Dampfdruck als Funktion der Temperatur dargestellt. Sowohl die Spezifische Feuchte wie der Dampfdruck sind ineinander umrechenbar. Der Zusammenhang zwischen relativer Feuchte, Temperatur und der Dampfdruck ist ersichtlich: ein Luftpaket mit einem Dampfdruck von 2122 Pa und einer spezifischen Feuchte von 13,1 kg/kg erreicht bei einer Temperatur von 18,5°C eine relative Feuchte von 100% ; steigt die Temperatur weiter auf 30°C, reduziert sich für das gleiche Luftpaket ihre relative Feuchte um 50%.

Mittlerer Luftdruck auf Meereshöhe (mean sea level pressure (mslp)). Der Luftdruck p wird als die Gewichtskraft pro Flächeneinheit, die durch Atmosphärendampfdruck auf den Boden ausgeübt wird, definiert. p wird üblicherweise auf Bodenkarten mittels Isobaren dargestellt. Damit werden die Druckverhältnisse auf Meereshöhe sichtbar. Ferner können Rückschlüsse auf Windrichtungen und Windgeschwindigkeiten sowie auf den Verlauf von Fronten gezogen werden. Folgende Faktoren sind Hinweise auf den

Übergangsbereiche von Luftmassen:

- relatives Druckminimum
- horizontale Gradienten (Feuchte, Temperatur)
- Wolken, Niederschlag
- ausgeprägte vertikale Scherung des Horizontalwindes (Windrichtungsänderung und/oder Geschwindigkeitsänderungen mit der Höhe)
- Konfluenz.

Als Einheit gilt [p]=hPa

Wie im Kapitel 4 schon erläutert, ist das Klima in Österreich durch Wettersysteme vom Atlantik, dem Mittelmeerraum und dem osteuropäischen Raum bestimmt. Bei der Wahl des geographischen Ausschnitts müssen zwei Punkte betrachtet werden: Einerseits sollten bei den gewählten Daten möglichst charakteristische Datenreihen herangezogen werden, sodass möglichst viele Zirkulationszustände erfasst werden. Auf diese Weise kann gesichert werden, dass die *Analog-Methode* möglichst ähnliche Zirkulationsfelder findet. Das wirft das Problem der grossen Anzahl der Freiheitsgrade [45] auf. In [61] wird als Lösungsansatz zur Reduktion der Zahl der Freiheitsgrade eine Hauptkomponentenanalyse vorgeschlagen. Damit kann ein *signal* von *noise* getrennt werden, da viele der Freiheitsgrade *noise* repräsentieren. Der zweite Punkt besteht darin, den geographischen Ausschnitt so zu wählen, dass jene atmosphärischen Prozesse erfasst werden, welche das Wetter in Österreich entscheidend beeinflussen. Es wird ein geographischer Ausschnitt von -10°W bis 35°E und 35°N bis 65°N gewählt (Abb. 4.9).

4.4.1 Mittelung der Reanalyse Felder

Für zwei meteorologische Felder wurde an allen Gitterpunkten das arithmetische Mittel über die Zeit gebildet. Das Ergebnis ist in den Abbildungen 4.12 und 4.13 für den Luftdruck auf Meereshöhe (links) und die Spezifische Feuchte (*precipitable water*) in 700hPa (rechts) dargestellt. Gemittelt wurde für die Periode 1958 bis 2000 für jede Jahreszeit. Der Abstand zwischen den Isobaren beträgt 1hPa.

Mittlerer Luftdruck auf Meereshöhe (mslp)

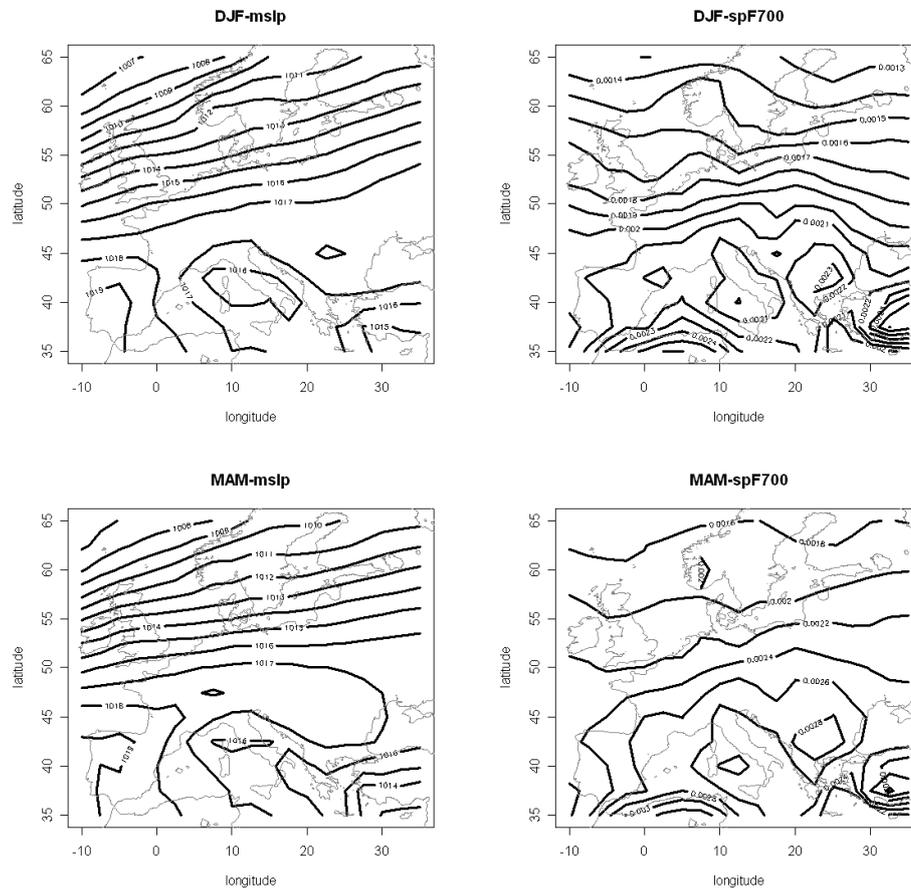


Abbildung 4.12: Saisonale Mittelung (1958-2000) des mittleren Luftdrucks in Meereshöhe und der spezifischen Feuchte in 700 hPa für Winter (DJF) und Frühling (MAM).

DJF Der erste Blick vermittelt eine ausgeprägte zonale Zirkulation: grossräumig gesehen verlaufen die Isobaren von Süd-West nach Nord-Ost. Dieser Zirkulationstyp kommt im Winter oft vor und ist nach dem Verlauf der Isobaren einer Südlage mit westlicher Komponente zugeordnet. Erkennbar ist ein Hoch über der Iberischen Halbinsel mit 1023 hPa und Rumänien mit 1022 hPa. Die Feldmittelung lässt ein Tief nordwestlich von England vermuten.

- MAM Südwestlich der Iberischen Halbinsel ist das Vorrücken eines Hochs (1020) bis zum Seegebiet südlich von England erkennbar. Im Gegensatz dazu befinden sich über Mittel- und Südeuropa niedrige Druckverhältnisse (1014-1016 hpa), die sich durch geringe Luftdruckunterschiede charakterisieren. Die deutliche Abnahme des mittleren Luftdrucks auf Meereshöhe nach Nord und Nordwesten hin ist deutlich erkennbar.
- JJA Das Hochdruckgebiet, das schon im Frühling erkennbar war, scheint hier ausgeprägter und in seiner Ausdehnung weiter nach Norden vorgezogen zu sein. Seine Ausdehnung reicht vom Golf von Biskaya bis hin zum mitteleuropäischen Raum. Im Vergleich dazu zeichnet sich sowohl der Norden als auch der Nordosten und den östliche Mittelmeerraum durch einen schwachen Tiefdruck aus. Gleichzeitig liegt eine schwach ausgeprägte zonale Zone über dem gesamten nordeuropäischen Raum.
- SON Die gemittelte Zirkulation ist ähnlich dem Zirkulationsmuster der Sommermonate (JJA). Erkennbar ist ein breites, stark ausgeprägtes zonales Zirkulationsband über Nordeuropa, wobei die Isobaren eine Südwest-Nordost Neigung aufweisen. Im Gegensatz dazu werden die gesamte Iberische Halbinsel und der Mittelmeerraum von höheren Druckverhältnissen beherrscht, stark beeinflusst von einem zumutbaren Hoch im Südwesten durch den Verlauf der Isobaren .

Spezifische Feuchte in 700hpa

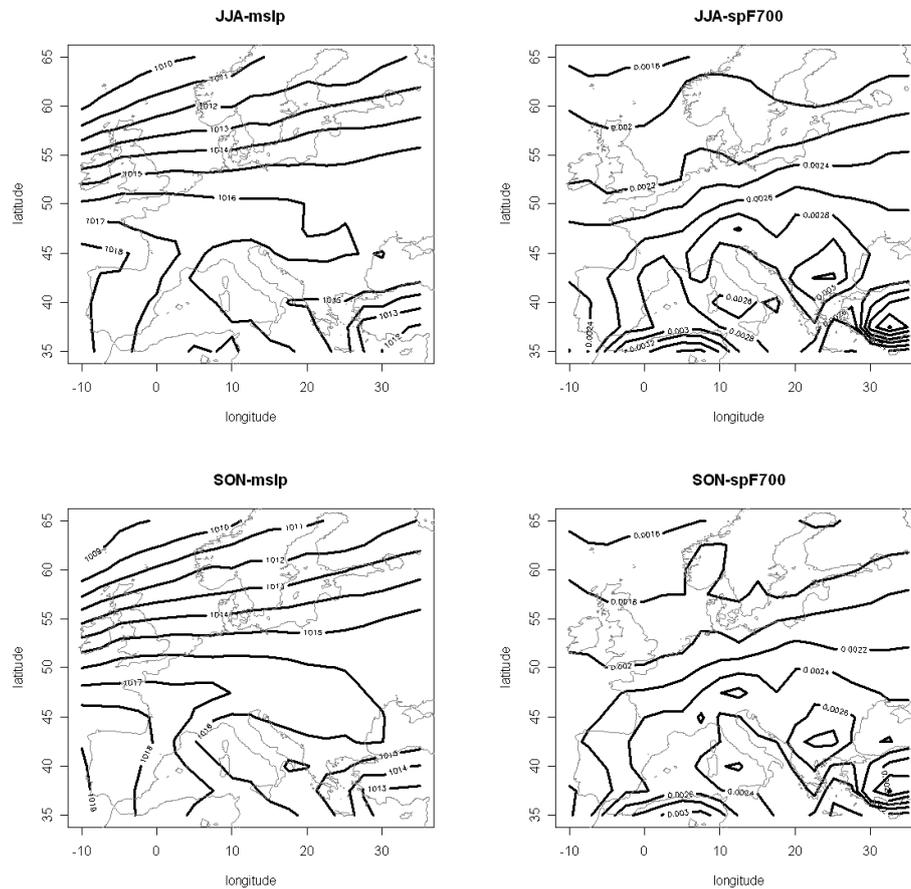


Abbildung 4.13: Saisonale Mittelung (1958-2000) des mittleren Luftdrucks in Meereshöhe und der spezifischen Feuchte in 700 hPa für Sommer (JJA) und Herbst (SON).

DJF Auf der 700 hPa Fläche bewegen sich die aus der Mittelung der spezifischen Feuchte Feld resultierenden Werte zwischen dem Maximum von $2,3 \times 10^{-3} \text{ kg/kg}$ über Rumänien, dem nordafrikanischen Kontinent, Rumänien und der Türkei und dem Minimum $1,3 \times 10^{-3} \text{ kg/kg}$ nördlich des 65° Breitengrades. Es lässt sich klar erkennen, dass die Schicht mit den höheren Werten vom nordafrikanischen Kontinent bis hin zum Mitteleuropäischen Raum reicht. Gleichzeitig ist eine Abnahme der Spezifischen Feuchte in 700 hPa von den niedrigen zu den hö-

heren Breiten feststellbar, wobei die Isolinien eine leichte Neigung von Nordwest nach Südost aufzeigen.

MAM Die höchsten q -Werte befinden sich über dem Mittelmeerraum bis zum östlichen Europa, wobei zwei Maxima von 2.8×10^{-3} über Rumänien und der östlichen Türkei leicht erkennbar sind. Im Vergleich zu den Wintermonaten reicht die Feuchte Schicht weit nach Nordeuropa. Ähnlich wie bei den Wintermonaten wird hier tendentiell auch eine Abnahme der spezifischen Feuchte zu höheren Breite hin beobachtet.

JJA In den Sommermonaten ist die mittlere Spezifische Feuchte in 700hPa weit höher als in den anderen Jahreszeiten, vor allem in Mitteleuropa und dem Mittelmeerraum. Über dem östlichen Mitteleuropa, bzw. über Bulgarien und Rumänien liegt ein Maximum von 3.1×10^{-3} , über der östlichen Türkei lässt der Isolinienverlauf einen höheren Wert vermuten.

SON Der Verlauf der Isolinien ist ähnlich der eben beschriebenen Sommermonate. Allerdings sind die mittleren q -Werte etwas geringer. Trotzdem gibt es keine Verschiebung der Maximums über Polen.

Kapitel 5

Statistisches Downscaling

5.1 Methodik

5.1.1 Die Hauptkomponentenanalyse

Um die *Analog-Methode* umsetzen zu können, wird eine Hauptkomponentenanalyse [58](Principal Components Analysis, PCA) benötigt. Die PCA wird auf die ausgewählten Reanalyse-Felder bzw. auf die großskaligen Daten angewandt, welche als Prediktoren in der Umsetzung der *Analog-Methode* dienen. Dieses Verfahren wurde erstmals im meteorologischen Kontext von Lorenz(1956) [11] eingesetzt.

Die Hauptkomponentenanalyse (PCA), von manchen Autoren auch EOF-Analyse (EOF=Empirische Orthogonal Funktion) genannt, besteht darin, meteorologische Felder in lineare unabhängige Eigenvektoren zu zerlegen und eignet sich zur Erfassung von Variabilitätsstrukturen in den Feldern. Dadurch wird eine Reduktion der Dimensionalität eines Datensatzes erreicht. Es geht darum, aus einem Datensatz bestehend aus X Variablen, neue Variablen Y zu bekommen, welche den größten Anteil der Varianz des Eingangsdatensatzes reproduzieren. In der Literatur werden verschiedene Wege beschrieben, die sich hauptsächlich im *preprocessing* der Daten unterscheiden, welche die Berechnung der Eigenvektoren und der Hauptkomponenten zulassen. In dieser Arbeit werden die EOFs und die Hauptkomponenten mittels Singularwertzerlegung (SVD) [18, 58] berechnet.

mixed-Field EOF und CPCA

Die PCA kann auf mehr als auf ein meteorologisches Feld angewandt werden. Die in dieser Arbeit eingesetzte und von Benestad *mixed-Field EOF* genann-

te [3] Methode beschreibt, wie zwei verschiedene Felder, z.B Temperatur und mittlerer Luftdruck auf Meereshöhe, zeitlich kovariieren und ist dadurch in ihrer Struktur identisch zu der von Bretherton (1992) beschriebenen *CPCA* („combined PCA“)[7] [58]. Auf diese Weise ist die *mixed EOF* Methode darauf gerichtet, das Kovariieren von Signalen verschiedener Datensets hervorzuheben.

Der Einfachheit halber möge dieses Feld als ein Vektor \mathbf{z} aufgefasst werden. In dieser Arbeit wird die PCA auf die Felder zweier verschiedener Variablen gleichzeitig angewandt: mittlerer Luftdruck auf Meereshöhe (mslp) und spezifische Feuchte in 700 hPa Niveau. Da die Anzahl der Variablen L , deren Beobachtungen auf K Gitterpunkten liegen, gleich zwei ist, ist die gesamte Dimension des $\mathbf{z}(\mathbf{t})$ Vektors gleich KL . Dabei sind die ersten K Elemente von $\mathbf{z}(\mathbf{t})$ die Beobachtungen der ersten Variablen $\mathbf{x}(\mathbf{t})$ und die zweiten K Elemente sind die Beobachtungen der zweiten Variablen $\mathbf{y}(\mathbf{t})$. Gäbe es mehr Variable, würden die letzten Elemente der letzten L -ten Variable angehören. Liegen K Beobachtungswerte vor, bzw. liegen die Beobachtungen, die zu einer bestimmten Zeit t gemessen wurden, auf K Gitterpunkten, so lassen sie sich vektoriell darstellen durch:

$$\mathbf{z}(\mathbf{t}) = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k] \quad (5.1)$$

Zwei verschiedene Vektoren $\mathbf{x}(\mathbf{t})$ und $\mathbf{y}(\mathbf{t})$ mit den entsprechenden Längen n und m lassen sie sich folgendermaßen *kombinieren*.

$$\mathbf{z}(\mathbf{t}) = [\mathbf{x}(\mathbf{t}), \mathbf{y}(\mathbf{t})] = [\mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{y}_1, \dots, \mathbf{y}_k] \quad (5.2)$$

Dabei wird die Information, die in den Vektoren enthalten ist, durch die Kombination der beiden Vektoren umgesetzt. Die Länge des Vektors \mathbf{z} entspricht dann der Summe beider Vektoren \mathbf{x} und \mathbf{y} . Die Umsetzung erfolgt durch den Algorithmus *mixedField.R* der R Package ¹*clim.pact* ²

Da die Variablen meistens verschiedene Einheiten haben, ist es sinnvoll, die PCA auf die Korrelationsmatrix anzuwenden. Die Dimension der Korrelationsmatrix und der Eigenvektormatrix ist $(KL \times KL)$. Wendet man die PCA auf die Korrelationsmatrix an, entstehen die Hauptkomponenten, in denen die gemeinsame Varianz der Variablen maximiert wird. Gleichzeitig werden die Korrelationen zwischen den beiden auf den K Gitterpunkten berücksichtigt. Die Umsetzung der Hauptkomponentenanalyse erfolgt mittels der Singulärwertzerlegung (SVD) auf dem resultierenden Feld durch die *Kombination* der Vektoren.

¹<http://www.r-project.org/>

²<http://cran.r-project.org/doc/packages/clim.pact.pdf>

Mathematisch-Statistische Grundlagen

Die Singularwertzerlegung (SVD) ermöglicht die Zerlegung einer allgemeinen ($m \times n$) rechteckigen Matrix und ist deshalb wie andere Verfahren (Eigenwertproblem) nicht auf quadratische Matrizen beschränkt. Jede beliebige rechteckige Matrix A mit der Dimension ($n \times m$), die mindestens so viele Zeilen wie auch Spalten ($n \geq m$) besitzt, lässt sich mittels SVD folgendermaßen zerlegen:

$$[A] = [L] [\Omega] [V]^* \quad (5.3)$$

Die Matrix L stellt eine Unitäre Matrix³. Die Matrix V^* ist die zur reellen oder komplexen quadratischen V adjungierte Matrix⁴. Die Matrix Ω ist eine reelle Matrix deren Einträge in der Diagonale die Singularwerte σ_i ⁵ von A sind.

Diese Matrizen haben folgende Dimension:

$$[L] = (m \times m) \quad [\Omega] = (m \times n) \quad [V]^T = (n \times n)$$

Die m Spaltenvektoren von L sind die linken singulären Vektoren und die m Spaltenvektoren von $[V]$ sind die rechten singulären Vektoren. Sowohl $[V]$, als auch $[L]$ sind orthonormale Matrizen d.h:

$$[L]^T [L] = [V]^T [V] = I \quad (5.4)$$

wobei I die Einheitsmatrix mit der Dimension $m \times m$ ist. Falls A eine quadratische und symmetrische Matrix ist, wird Gleichung 5.4 vereinfacht zu:

$$[A] = [E] [\Lambda] [E]^T \quad (5.5)$$

Dabei ist $[\Lambda]$ die Diagonalmatrix, deren nicht negative Elemente die K Eigenwerte von $[A]$ sind.

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_K \end{pmatrix}$$

³ L ist eine komplexe quadratische Matrix, deren Spalten zueinander orthonormal sind

⁴Ist V eine reelle Matrix dann ist die zu V^* adjungierte Matrix die Transponierte von V ($=V^T$). Das trifft dann genau zu wenn $V^* \cdot V = I$. I ist die Einheitsmatrix

⁵Die Singulärwerte der Matrix A sind gleich den Quadratwurzeln aus den Eigenwerten von $A^T \cdot A$.

Schreibt man die Gleichung 5.5 in Summen-Notation, folgt:

$$[A] = \sum_{k=1}^K \lambda_k \mathbf{e}_k \mathbf{e}_k^T \quad (5.6)$$

$$= \sum_{k=1}^K \lambda_k [E_k] \quad (5.7)$$

$[E]$ ist hier eine $(K \times K)$ Matrix, deren Spaltenvektoren die Eigenvektoren \mathbf{e}_k sind, d.h:

$$[E] = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_K] \quad (5.8)$$

Die Eigenvektoren besitzen folgende Eigenschaften:

- $\|\mathbf{e}\| = 1$. Die Eigenvektoren haben die Länge eins
-

$$\mathbf{e}_i^T \mathbf{e}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

In Gleichung 5.7 wird gezeigt, dass die Matrix $[A]$ wieder erlangt werden kann, indem die Summe der $[E_k]$ Matrizen mit den entsprechenden Eigenvektoren gewichtet wird. Demzufolge ist die Spektralzerlegung einer Matrix ähnlich der Fourierzerlegung. Dabei entsprechen die Eigenwerte den Fourier Amplituden und die Matrizen $[E_k]$ den Cosinus Funktionen.

Aus Gleichung 5.5 folgt:

$$tr(A) = \sum_{k=1}^K a_{k,k} = \sum_{k=1}^K \lambda_k = tr(\Lambda) \quad (5.9)$$

Der Zusammenhang gilt dann, wenn $[A]$ die Kovarianzmatrix ist, deren Elemente $a_{k,k}$ die K Varianzen sind. Gleichung 5.9 sagt aus, dass die Summe der Varianzen gleich der Summe der Eigenwerte der Kovarianzmatrix ist. Aus der Gleichung 5.3 folgt dann weiter für die Eigenwerte:

$$det [A] = \prod_{k=1}^K \lambda_k = det [\Lambda] \quad (5.10)$$

Das ist konsistent mit der Eigenschaft, dass mindestens einer der Eigenwerte einer singulären Matrix ⁶ gleich Null ist. Sind die Eigenwerte positive

⁶Ist $[A]$ eine singuläre Matrix, so ist ihre Determinante gleich Null $|A| = 0$

Werte, so nennt man die Matrix semi-definit.

Die Eigenvektormatrix $[E]$ hat die Eigenschaft, die symmetrische Matrix $[A]$ zu diagonalisieren, d.h. die Multiplikation von $[A]$ auf der linken Seite mit $[E]^{-1}$ und auf der rechten Seite mit $[E]$ diagonalisiert $[A]$:

$$[E]^{-1}[A][E] = [\Lambda] \quad (5.11)$$

Es besteht also ein starker Zusammenhang zwischen den Eigenvektoren und den Eigenwerten einer nicht singulären Matrix und den entsprechenden λ_k^* und \mathbf{e}_k^* der inversen Matrix. Die Eigenvektoren der inversen Matrix sind die gleichen, d.h. $\mathbf{e}_k^* = \mathbf{e}_k$. Die entsprechenden Eigenwerte sind reziprok $\lambda_k^* = \lambda^{-1}$. Demzufolge ist der Eigenvektor von $[A]$ mit dem größten Eigenwert der gleiche wie jener der inversen Matrix $[A]^{-1}$ mit dem zugehörigen kleinsten Eigenwert. Durch die Lösung des charakteristischen Polynoms K -ter Ordnung erhält man die K Eigenwerte und die entsprechenden Eigenvektoren:

$$\det([A] - \lambda [I]) = 0 \quad (5.12)$$

Wenn die Dimension des Problems grösser ist bzw. $K > 4$, wird bevorzugt der Rechner eingesetzt. Die Zerlegung von $[A]$ in 5.3 kann als Summe der gewichteten äußeren Produkte der linken und rechten singulären Vektoren ausgedrückt werden:

$$[A] = \sum_{i=1}^m \omega_i \mathbf{l}_i \mathbf{v}_i^T \quad (5.13)$$

Ist $[A]$ nicht symmetrisch, gibt es einen Zusammenhang zwischen der SVD und den Eigenvektoren und Eigenwerten von $[A][A]^T$ und $[A]^T[A]$. Das Produkt ergibt eine symmetrische und quadratische Matrix mit der Dimension $(n \times n)$ und $(m \times m)$. Insbesondere sind die Spalten von $[V]$ die $(m \times 1)$ Eigenvektoren von $[A]^T[A]$, und Spalten von $[L]$ die $(n \times 1)$ Eigenvektoren von $[A][A]^T$. Die singulären Werte sind die Wurzeln aus den entsprechenden Eigenwerten, d.h.:

$$\omega_i^2 = \lambda_i \quad (5.14)$$

Berechnung der Hauptkomponenten(PCs) mittels SVD

Vor der Anwendung des SVD müssen die meteorologischen Felder zentriert werden, d.h. der Mittelwert wird den entsprechenden Beobachtungen

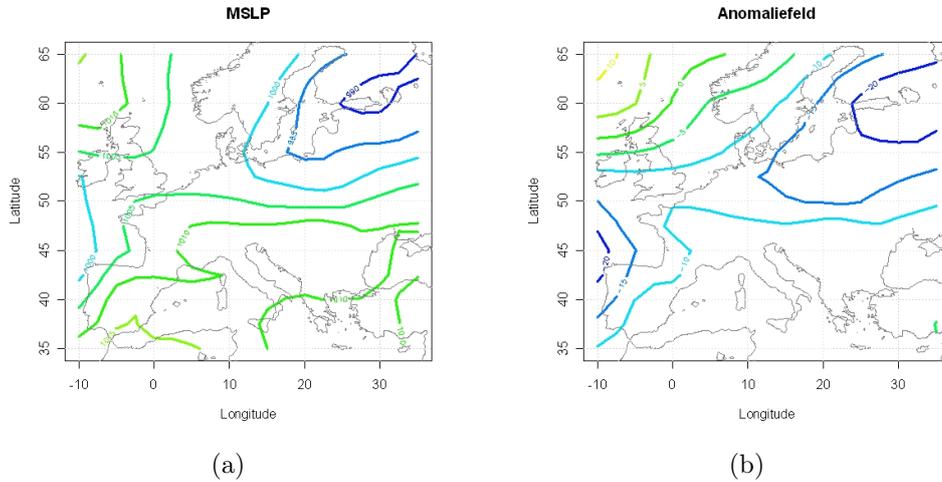


Abbildung 5.1: ERA40 Feld des mittleren Luftdrucks in Meereshöhe für den Tag 01.01.1958 (a) und entsprechendes ermitteltes Anomaliefeld (b).

abgezogen: Sei $[1]$ eine $(n \times n)$ Matrix, deren Elemente alle gleich 1 sind. Die entsprechende $(n \times K)$ Matrix der Anomalien $[Z']$ ergibt sich aus:

$$[Z'] = [Z] - \frac{1}{n}[1][Z] \quad (5.15)$$

Die Matrix $[Z]$ entspricht der $(n \times K)$ Beobachtungsmatrix, wobei n die (zeitlich) aufeinanderfolgenden Beobachtungen sind und K der Anzahl der Gitterpunkte eines meteorologischen Feldes entspricht. Die Bildung des Mittelwertes eines Feldes erfolgt durch die Division der Beobachtungen an dem jeweiligen Gitterpunkt durch deren Anzahl n . Abbildung 5.1 zeigt ein Beispiel des mittleren Luftdruckfeldes im Meeresniveau (a) für den 01.01.1958 und das ermittelte Anomaliefeld (b).

Die Matrix der Anomalien $[Z']$ hängt mit der Varianz-Kovarianz-Matrix $[S']$ zusammen durch:

$$[S] = \frac{1}{n-1} [Z'] [Z']^T \quad (5.16)$$

Die hier eingesetzte Methode erfolgt durch die Anwendung der SVD auf die zentrierte $(n \times K)$ Matrix $[Z']$. Dabei sind die Eigenwerte gleich dem Quadrat der singulären Werte d.h. es gilt Gleichung 5.14 und die Eigenvektoren von $[S]$ sind die gleichen wie die rechten singulären Vektoren von $[Z']$, d.h. $[E] = [R]$ oder $\mathbf{e}_k = \mathbf{v}_k$.

Der Vorteil SVD zu benutzen, um die Hauptkomponenten zu berechnen liegt darin, dass die singulären Vektoren, d.h die Spalten der $(n \times K)$ $[L]$ in Gleichung 5.3 proportional zu den Hauptkomponenten sind, d.h. proportional zu den Projektionen der zentrierten Beobachtungen auf die Eigenvektoren \mathbf{e}_k . Es gilt:

$$u_{i,k} = \mathbf{e}_k^T \mathbf{z}_i' = \sqrt{n-1} \ell_{i,k} \sqrt{\lambda_k} \quad (5.17)$$

mit:

$$i = 1, \dots, n \quad k = 1, \dots, K$$

bzw:

$$[U] = \sqrt{n-1} [L] [\Lambda]^{1/2} \quad (5.18)$$

Die Spalten von $[U]$ entsprechen den Hauptkomponenten u_k der (zeitlich) aufeinanderfolgenden n Beobachtungswerte x_i , mit $i = 1, \dots, n$. Dementsprechend entspricht $[U]$ der Matrix $[V]$ in 5.3, deren Spalten die sogenannten *normalized principal components*, bzw. die Eigenvektoren von $[A'] [A]$ sind, während die Spalten von $[L]$ den *spatial eigenvectors* (EOFs) von $[A] [A']$ entsprechen.

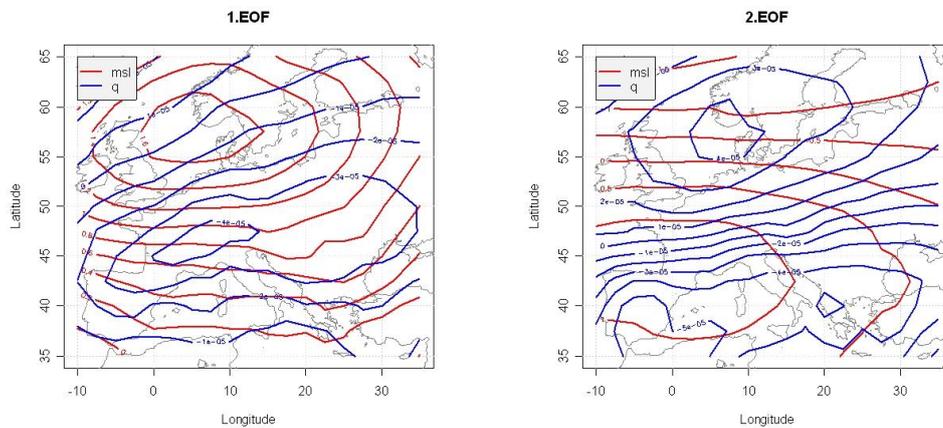


Abbildung 5.2: Erste und zweite empirische Orthogonalfunktion (1.EOF-Muster und 2.EOF-Muster) des mittleren Luftdrucks in Meereshöhe und der spezifischen Feuchte in 700 hPa.

In 5.2 sind die ersten EOF- und die zweiten EOF-Muster für den Winter (DJF) der bereits besprochenen Parameter dargestellt. Die EOFs sind rein

mathematische Konstrukte, welche eine physikalische Interpretation ohne das Einbeziehen zusätzlicher Informationen nicht gestatten. Vor allem gilt dies für die zweiten EOF-Muster. Der Grund dafür liegt in der Orthogonalitätsbedingung [50]. In den meisten Fällen ist die physikalische Interpretation nur auf das erste EOF-Muster beschränkt. In Abbildung 5.3 sind die dazugehörigen Hauptkomponenten der ersten und dritten EOF des *kombinierten* Feldes graphisch dargestellt. Die strichlierten Linien stellen den zeitlichen Sprung zwischen der Jahreszeit (z.B. DJF) eines Jahres und der gleichen Jahreszeit im darauffolgenden Jahr dar.

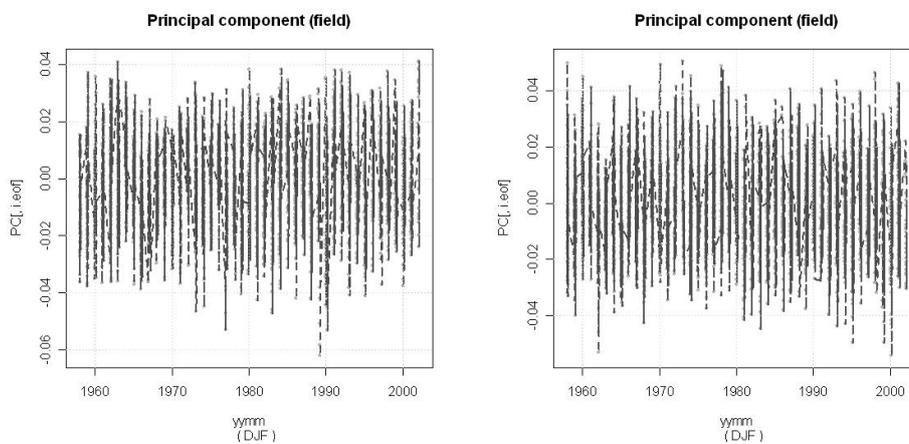


Abbildung 5.3: Erste und dritte Hauptkomponentenreihe (PC1 und PC3) des *kombinierten* Feldes des mittleren Luftdrucks in Meereshöhe und spezifische Feuchte in 700hPa.

Die Hauptkomponenten

Weisen die Beobachtungsdaten, aus denen die empirischen orthogonal Funktionen berechnet werden, eine hohe Korrelation auf, dann wird in der Regel eine rasche Abnahme der Varianz der Hauptkomponenten verifiziert. Die Hauptkomponentenanalyse ermöglicht die Berechnung der Eigenvektoren der aufgestellten Kovarianzmatrix. Dabei wird die erste Hauptkomponente durch jenen Eigenvektor mit dem größten Eigenwert dargestellt. Auf dieser ist der größte Anteil der Datenvarianz abgebildet. Anders ausgedrückt: der erste berechnete Eigenvektor \mathbf{e}_1 zeigt in die Richtung der größten Variabilität des ursprünglichen Datensatzes. Der Zweite Eigenvektor \mathbf{e}_2 mit dem zweitgrößten Eigenwert zeigt in die Richtung der zweitgrößten Variabilität usw. Die restlichen Eigenvektoren folgen diesem Verhalten. Allerdings unter der

Bedingung der Orthogonalität. Die gesamten Eigenvektoren bilden die so genannten Empirischen Orthogonal Funktionen (EOF). Man bedient sich dieses Effekts, wenn es darum geht, eine Dimensionalitätsreduktion der Daten zu erreichen. Dadurch kann die Variabilität der ursprünglichen Daten anhand von wenigen Hauptkomponenten erklärt werden.

Aus der Singularwertzerlegung (SVD) erhält man, wie schon im letzten Abschnitt erklärt, die Eigenwerte. Um den Anteil der Gesamtvariation, die durch jede m -te Hauptkomponente (PC) erfasst wird zu bestimmen, gilt:

$$\nu_d = \frac{\lambda_m}{\sum_{j=1}^k \lambda_k} \times 100\% \quad (5.19)$$

Der Plot 5.4 zeigt den Anteil der Gesamtvarianz der ersten 20 Hauptkomponenten, die aus CPCA („combined PCA“) ERA40 Feldern berechnet wurden:

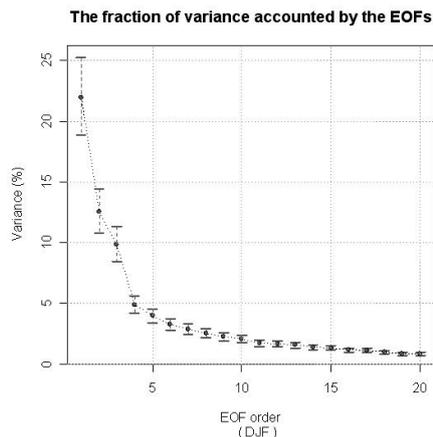


Abbildung 5.4: Anteil der Gesamtvarianz der ersten 20 Hauptkomponenten.

Es gibt verschiedene Regeln, die erlauben eine Entscheidung zu treffen, was als *Rauschen* zu verstehen ist. Dadurch kann entschieden werden, wieviele Hauptkomponenten zur Repräsentation der Varianz der ursprünglichen Daten beibehalten werden kann. Eine oft angewandte Regel ist die *Kaisers Rule*, die besagt, dass soviele Hauptkomponenten beizubehalten sind, wie λ -Werte größer sind als der Mittelwert aller Eigenwerte, das entspricht:

$$\lambda_l > \frac{1}{k} \sum_{i=1}^k \lambda_k \quad (5.20)$$

Der Mittelwert beträgt im Beispiel 21.08. Damit mussten nach Kaisers Rule mehr Hauptkomponenten einbezogen werden als üblich. Der Auswirkung der Distanzmaße auf die *Analog-Methode* wird jeweils einmal durch Einbeziehen von 7, 11 und 20 Hauptkomponenten Rechnung getragen. In früheren Arbeiten wurden mindestens so viele Hauptkomponente extrahiert bis mindestens 80 % [61] der Gesamtvarianz erklärbar waren.

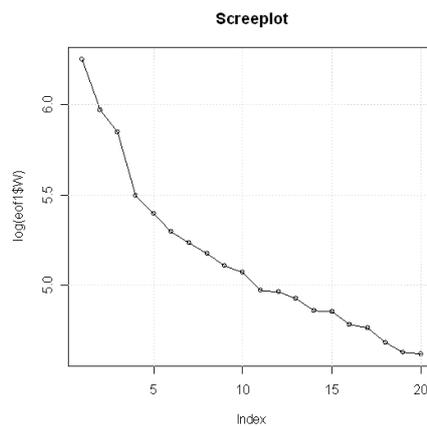


Abbildung 5.5: Logarithmischer Eigenwertverlauf der ersten 20 Eigenwerte der PCA auf die *kombinierten* ERA40 gewählte Datenfelder angewandt.

Aus dem Plot in 5.4 ist zu folgern, dass aufgrund der langsamen Abnahme der Varianz die Beobachtungsdaten eine niedrige Korrelation untereinander aufweisen. Es werden jeweils die ersten 7, 11 und 20 Hauptkomponenten extrahiert:

- **7 PCs** Die kumulierte Summe der sieben ersten Hauptkomponenten erklärt insgesamt 59% der gesamten Varianz
- **11 PCs** Die ersten 11 Hauptkomponenten machen insgesamt 70% der gesamten Varianz aus
- **20 PCS** werden die ersten 20 Hauptkomponenten extrahiert, so können 80% der Gesamtvarianz erfasst werden.

In 5.5 ist der logarithmische Eigenwertverlauf der Hauptkomponentenanalyse auf die *kombinierten* ERA40 Datenfelder des mittleren Luftdruckes auf Meereshöhe und der spezifischen Feuchte in 700 hPa Niveau dargestellt. Die x-Koordinate gibt die Anzahl der Eigenwerte an. Dieses Verfahren gilt

zusätzlich zur Kaiser's Rule als Kriterium bei der Auswahl der Hauptkomponenten. Der Screenplot ist ein reines „optisches“ Verfahren, bei dem es darum geht einen „Knick“ im Verlauf festzustellen, der die Grenze zwischen *signal* und *noise* repräsentiert. In diesem Fall ist der „Knick“ nicht feststellbar, was das Ergebnis mit der Kaiser Rule bestätigt, d.h es müssen viele Hauptkomponenten einbezogen werden.

5.1.2 Die Distanzmaße

Für die Suche nach dem ähnlichsten Tag werden verschiedene Distanzmaße eingesetzt.

Um den Grad der Übereinstimmung zwischen Punkten oder Vektoren zu bestimmen, werden in der multivariaten Statistik üblicherweise Distanzfunktionen oder auch Ähnlichkeitsmaße eingesetzt. In den meisten Anwendungen stellen die Vektoren eine Folge von Datenwerten dar. Solche Vektoren werden üblicherweise Merkmalsvektoren genannt. Sie fassen die Merkmalsausprägungen zusammen. Eine Anwendung der Ähnlichkeits- und Distanzmaße ist das *clustering*, bei dem Ähnlichkeiten zwischen den Objekten gesucht werden. Dieses Verfahren ermöglicht die Gruppierung von *ähnlichen* Objekten oder Datenpunkten. Abbildung 5.6 zeigt die Gruppierung von *scores*⁷ nach ihrer Ähnlichkeit, einmal mit der City-Block-Distanz (b), einmal mit dem Euklidischen Distanzmaß (a). Obwohl die Gruppierung Ähnlichkeiten aufweist, gibt es kleine Unterschiede, die bei genauerer Betrachtung auffallen. Dieses Beispiel deutet darauf hin, dass die Wahl des Distanzmaßes sehr wohl das Ergebnis der Gruppierung beeinflussen kann. Im folgenden werden die Begriffe Ähnlichkeitsmaß und Distanzfunktion (=Distanzmaß)[4] definiert. Letztes ist als Oberbegriff zu verstehen und sollte nicht mit der *Metrik* verwechselt werden; metrische Distanzmaße sind Distanzfunktionen, aber nicht alle Distanzfunktionen sind Metriken.

Definitionen und Eigenschaften

Ähnlichkeitsmaß: Sei $M = \{1, \dots, n\}$ eine endliche Menge. Eine reellwertige Funktion $s: M \times M \rightarrow \mathfrak{R}$ mit $s_{i,j} = s(i, j)$ heißt *Ähnlichkeitsmaß* auf der Menge M , wenn $\forall i, j \in M$ folgende axiomatische Bedingungen erfüllt sind:

⁷die einzelne Elemente einer Hauptkomponente werden *scores* genannt, während die einzelnen Elemente eines Eigenvektors als *loading* bezeichnet werden. Der Eigenvektor selbst wird als *Empirische Orthogonal Funktion* (EOF) bezeichnet

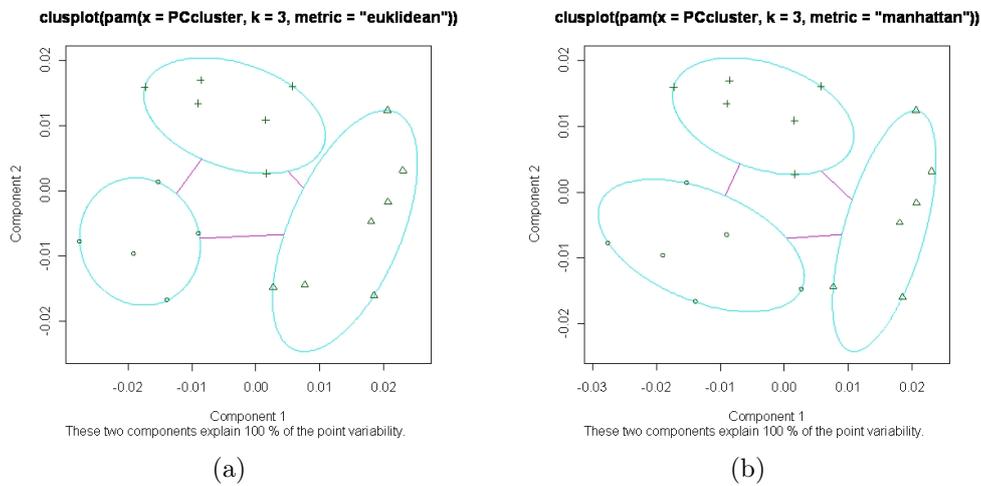


Abbildung 5.6: Bildung von Clustern(Gruppen) mittels euklidischem Distanzmaß(a)und City-Block-Distanz(b).

1. $s_{i,j} = s_{j,i}$
2. $s_{i,j} \leq s_{i,i}$
3. $s_{i,j} = 1$
4. $s_{i,i} \geq 0$

Dabei gibt der Zahlenwert $s_{j,k}$ den Ähnlichkeitsgrad zwischen zwei Merkmalsvektoren, der zwischen 0 (kleine Ähnlichkeit) und 1 (große Ähnlichkeit) variieren kann, an.

Distanzmaß: Sei $M = \{1, \dots, n\}$ eine endliche Menge. Eine reellwertige Funktion $d: M \times M \rightarrow \mathfrak{R}$ mit $d_{i,j} = d(i, j)$ heißt Distanzmaß auf der Menge M , wenn $\forall i, j \in M$ folgende axiomatische Bedingungen erfüllt sind:

1. $d_{i,j} \geq 0$
2. $d_{i,i} = 0$ (Reflexivität)
3. $d_{i,j} = d_{j,i}$ (Symmetrie)

Wenn folgende Bedingungen zusätzlich erfüllt sind:

4. $d_{i,j} = 0 \Rightarrow \mathbf{x}_i = \mathbf{x}_j$

5. $d_{i,k} \leq d_{i,j} + d_{j,k} \quad \forall i, j, k \in M$ (Dreiecksungleichung)

dann handelt es sich um metrische Distanzfunktionen. Diese entsprechen dem räumlichen Abstandsbegriff und lassen sich geometrisch gut darstellen. Zur Erläuterung der axiomatische Bedingungen:

1. Im Allgemeinen wird der Abstand zwischen zwei Datenpunkten als positive reelle Zahl angegeben
2. **Reflexivität:** Der Abstand zwischen zwei Datenpunkten ist dann Null, wenn sie identisch sind.
3. **Symmetrie** bedeutet, dass der Hinweg- und Rückweg- Abstand zwischen zwei bestimmten Datenpunkten immer gleich ist.
4. **Dreiecksungleichung** Der Länge einer Seite eines im Raum aufgespannten Dreieck kann nicht grösser sein als die Summe der anderen beiden Seiten des Dreiecks.

Das Ergebnis der Distanzberechnung lässt sich durch eine Matrix darstellen. Eine $(k \times k)$ Matrix D

$$D = \begin{pmatrix} 0 & d_{1,2} & \cdots & d_{1,k} \\ d_{2,1} & 0 & \cdots & d_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n,1} & d_{n,2} & \cdots & 0 \end{pmatrix}$$

heißt eine **Distanzmatrix**. Dabei sind die Diagonalelemente Null wegen der 2. Bedingung und symmetrisch wegen der 3. Bedingung. In 5.2, einer Tabelle der gefundenen analogen Tage für einen Teil einer Niederschlagsreihe einer Station, wird dies verdeutlicht.

Transformation von Ähnlichkeiten in Distanzmaßen

Distanzmaße und Ähnlichkeitsmaße lassen sich gegenseitig transformieren. Da der Abstand zwischen \mathbf{x}_j und \mathbf{x}_i , möglichst klein sein sollte, d.h $d_{i,j} \rightarrow \min!$, damit die größte Ähnlichkeit vorliegt, wird eine nicht negative (streng) monotone fallende Transformationsfunktion $f(\cdot)$ gewählt, wobei $f(1) = 0$. Man setzt:

$$d_{i,j} = f(s_{i,j}) \quad 1 \leq i, j \leq K$$

Für $0 \leq s_{i,j} \leq 1$ gelten die Transformationen:

1. $d_{i,j} = 1 - s_{j,k}$
2. $d_{i,j} = \sqrt{1 - s_{i,j}}$
3. $d_{i,j} = -\log s_{i,j}$

5.1.3 Distanzmaße bei quantitativen Attributen

Im Gegensatz zu Datensätzen in Form von binären Attributen, sind meteorologische Felder als Datensätze zu verstehen, welche nur quantitative Attribute besitzen. Mit quantitativen Attributen werden Datenpunkte im Raum \mathfrak{R}_n aufgefasst. n ist die Anzahl der Attribute.

Die Abstände werden zwischen den Elementen (=Merkmalsausprägungen) der Spalten der Matrix $[U]$ bestimmt. Aus diesem Grund entspricht $x_{i,k}$ dem $u_{i,k}$, bzw $x_{j,k}$ dem $u_{j,k}$, also den Hauptkomponenten der Gleichung 5.17.

Folgende Merkmalsvektoren sind definiert durch:

$$\mathbf{x}_i = \begin{pmatrix} x_{i,1} \\ \vdots \\ x_{i,n} \end{pmatrix}$$

und

$$\mathbf{x}_j = \begin{pmatrix} x_{j,1} \\ \vdots \\ x_{j,n} \end{pmatrix}$$

damit lassen sich folgende Distanzmaße definieren:

City-Block-Distanzmaß: Dieses bestimmt die Entfernung zwischen zwei Punkten und wird definiert bezüglich eines rechtwinkligen Abstands in jeder Richtung

$$L_1 : d_{i,j} = |\mathbf{x}_i - \mathbf{x}_j| = \sum_{k=1}^n |x_{i,k} - x_{j,k}| \quad (5.21)$$

Auswirkung: Bei dieser Metrik gehen die Unterschiede linear, d.h gleichgewichtig in die Formel ein. Im Vergleich zur quadrierten euklidischen Distanz werden starke Abweichungen nicht stärker gewichtet als kleine. Anders ausgedrückt: kleine Abweichungen sind genauso bedeutsam wie große Abweichungen. Auf diese Weise sind Datenpunkte, die sich weit entfernt befinden

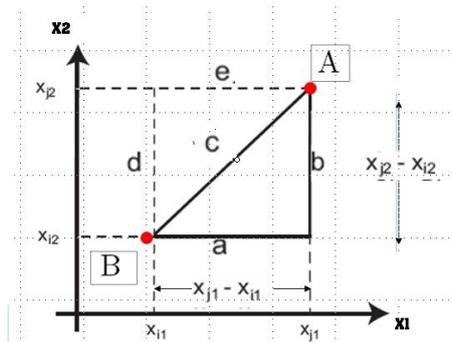


Abbildung 5.7: City-Block-Distanzmaß und Euklidische Distanzmaß: Beispiel eines Abstandsmessung im 2D-Raum: Abstand von Punkt B nach Punkt A mittels *CityBlock* = $a + b$ und *Euklidischer Distanz* = c (direkter Weg).

(= Ausreißer) im Vergleich zur Euklidischen Distanz weniger stark gewichtet.

Euklidisches Distanzmaß Im 2D- oder 3D- Zahlenraum ist die direkte Distanz auch die kürzeste Distanz im geometrischen Raum zwischen 2 Datenpunkten eines „gedachten“ Dreiecks. Sie ist also gleich der Aufsummierung der quadrierten Differenzwerte (ähnlich der least-square-methode in der Regressionsanalyse). In der Klimatologie [61] wurde dieses Maß benutzt, um den ähnlichsten Tag mittels der *Analog-Methode* zu finden.

$$L2 : d_{i,j} = \|\mathbf{x}_i - \mathbf{x}_j\|^2 = \sum_{k=1}^n (x_{i,k} - x_{j,k})^2 \quad (5.22)$$

Auswirkung: Im Vergleich zu anderen Distanzmaßen werden Abweichungen in der Datenmenge stärker gewichtet, d.h. eine starke Abweichung zwischen den Datenpunkten erhält mehr Gewicht als kleine Abweichungen. Das quadrierte Euklidische Distanzmaß bewirkt, dass benachbarte Datenpunkte im Vergleich zu weit entfernten Datenpunkten weiter zusammenrücken. Der Unterschied zwischen euklidischer und City-Block-Distanz wird am einfachen Beispiel in 5.7 demonstriert.

Gewichtetes Euklidisches Distanzmaß: hier werden die quadrierten Differenzwerte gewichtet. Grundsätzlich gilt zu überlegen ob eine Gewichtung der Merkmale sich als sinnvoll erweist oder nicht. Dies hängt meistens davon ab wie die Gewichte w_i bestimmt werden. Wird die Gewichtung so eingesetzt, dass die interne Struktur der Datenpunkte berücksichtigt wird, kann sie sich in der Bestimmung der Ähnlichkeitsmessung als sehr sinnvoll erweisen. Die

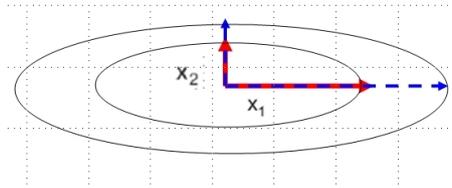


Abbildung 5.8: Eine Gewichtung führt zu einer Streckung der Eigenvektoren, die sich je nach Wichtigkeit der Hauptkomponenten auswirkt.

gewichtete euklidische Distanz wird durch

$$wL2 : d_{i,j} = \sum_{k=1}^n \{w_k (x_{i,k} - y_{j,k})^2\} \quad (5.23)$$

definiert, wobei gilt:

$$\sum_{k=1}^n w_k = 1 \quad (5.24)$$

Hier sind w_1, \dots, w_k , mit $w_i \in [0, 1]$. Die Gewichtung erfolgt hier mit den berechneten Eigenwerten, was gleichbedeutend ist, mit den Varianzen 5.4 aufgrund der Gleichung 5.9 zu gewichten. Dabei sollen die Gewichte die „Bedeutung“ der einzelnen empirischen Orthogonalfunktionen (EOF) reflektieren. Die Gewichte sind hier den erklärten Varianzen proportional.

Auswirkung Die Gewichtung bewirkt im allgemeinen eine Streckung des Raumes in einer bestimmten Dimension k bzw. bewirkt eine Stauchung entlang derjenigen Dimension. Eine bessere Vorstellung davon bekommt man, wenn dies auf die Koordinatenachsen übertragen wird. Das wird in Abbildung 5.8 schematisch dargestellt.

Zum Auffinden „sinnvoller“ Gewichte kann iterativ vorgegangen werden, indem von einer ungewichteten Ähnlichkeitsdefinition ausgegangen wird, um Datenpunkte, bzw Merkmale zu gruppieren. Aus dieser können dann „günstigere“ Gewichte bestimmt werden (z.B durch Klassen). Dadurch würden sich neue Gruppierungen und Distanzen ergeben [4].

Minkowski-Metrik:Die Verallgemeinerung der euklidischen Distanz stel-

len die L_r Metriken dar

$$L_{i,j}^p = d_{i,j}^r = \left[\sum_{k=1}^n (x_{i,k} - x_{j,k})^p \right]^{\frac{1}{p}} \quad (5.25)$$

- p entspricht dem Metrik-Parameter mit $p \geq 1$, $1 < p < \infty$ und $p \in \mathbb{N}$
- $x_{i,k}$ und $x_{j,k}$ stellen die Koordinatenwerte des Vektors \mathbf{x}_i und \mathbf{x}_j auf der Achse k dar.
- n bezeichnet die Dimension des Raumes
- p stellt die Metrik-Konstante dar
- Ist $p = 1$, erhält man die City-Block Distanz
- für $p = 2$ folgt die euklidische Distanz

Mit wachsenden p , fallen Differenzen zwischen den Werten eines Merkmales grösser aus, wohingegen kleine Differenzen weniger ins Gewicht fallen. In Abbildung 5.9 sind Isodistanzen für den 2D Raum ($n=2$) dargestellt. Folgende Fälle sind zu unterscheiden:

- Ist $p=2$, so nimmt die Kontur die Form eines Kreises an, d.h. in einer Ebene sind dann alle Orte mit gleicher Distanz zu einem beliebigen Punkt im Raum durch einen Kreis gekennzeichnet, und zur Messung dieses Abstandes ist jede Richtung möglich.
- Ist $p=1$, nimmt die Kontur die Form einer Raute an, was bedeutet, dass alle Datenpunkte mit der gleichen Distanz zum Zentrum des Koordinatensystems im Raum durch eine Raute beschrieben werden. Zwischen Messung des Abstandes können nur vertikale oder horizontale Schritte gemacht werden.

Für $p \rightarrow \infty$ gilt:

$$d_{i,j}^\infty = \max |\mathbf{x}_i - \mathbf{x}_j| \quad (5.26)$$

- Im Fall $p = \infty$ („Maximums Distanz“) nimmt die Kontur die Form eines Quadrats an, wobei nur diagonale Schritte gegangen werden können.

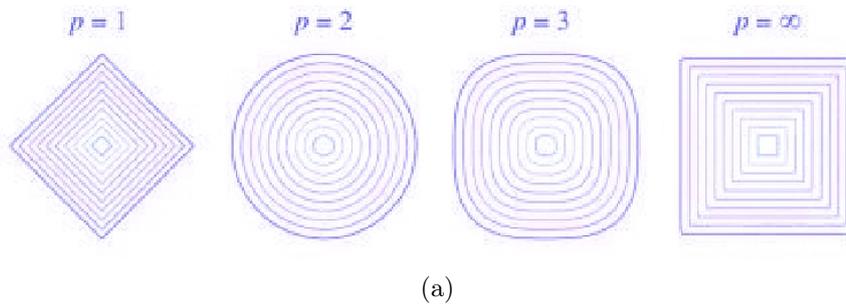


Abbildung 5.9: Graphische Darstellung der L_p Metrik für den 2-D Raum: Isodistanzen für verschiedene p -Werte (*Quelle: Wilkinson 2005 [56]*).

- Im Fall einer Gewichtung der quadratischen Differenzen, wie etwa bei der gewichteten Euklidischen Distanz 5.8 werden alle Datenpunkte mit gleichen Abständen zum Zentrum des Koordinatensystems (mit Zentrum $(0,0)$) durch eine Ellipse gekennzeichnet. Dabei werden sowohl vertikale wie auch horizontale Bewegungen unterschiedlich gewichtet, je nach vorhandener Varianz.

Mahalanobis Distanzmaß

Es besteht ein allgemeines Problem: Angenommen, man möchte die Distanz zwischen zwei beliebigen Datenpunkten 5.10 der Punktwolke us Abb.(a) berechnen. Üblicherweise würde man die euklidische Distanz benutzen. Die Kreiskonturen um das Zentrum der Wolke würden die Wahl dieses Distanzmaßes unterstützen. Weiters angenommen, die Form der Punktwolke sei nicht kreisförmig, sondern elliptisch(b). In diesem Fall wäre die euklidische Distanz möglicherweise nicht die beste Wahl; In Abbildung (c) haben die Datenpunkte $\mathbf{x}_{i,k}$ und $\mathbf{x}_{j,k}$ den gleichen Abstand vom Zentrum, wenn das Euklidische Distanzmaß verwendet würde. Bezüglich der Art, wie die Datenpunkte im Raum liegen, befindet sich der Punkt $\mathbf{x}_{j,k}$ weiter entfernt von der Punktwolke als $\mathbf{x}_{i,k}$. Als sinnvoll gewähltes Distanzmaß würde sich die Mahalanobis Distanz anbieten da diese ebensolche Besonderheiten berücksichtigt⁸. Ein weiteres Beispiel [58] erläutert die statistische Bedeutung dieses Maßes. Abbildung 5.11 zeigt für den 2D-Raum die Streuung der Datenpunkte zweier Vektoren x_i und x_j in Form eines Biplots. Die Verteilung hat ihr Zentrum im Ursprung des Koordinatensystems, wobei die Standardabweichung von \mathbf{x}_i dreimal grösser ist als von \mathbf{x}_j , d.h. $s_i \approx 3s_j$.

Die Orientation der Punktwolke entlang der Achsen lässt den Schluss zu,

⁸<http://zoonek2.free.fr/UNIX/48R/06.html>

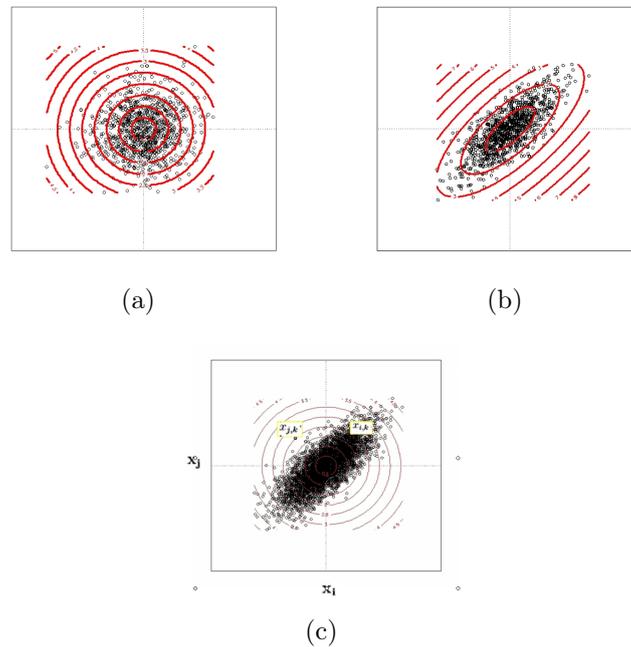


Abbildung 5.10: Zur Begründung der Anwendung der Mahalanobis Distanz: Um die Distanz zwischen 2 beliebigen Punkte zu berechnen, kann die euklidische Distanz verwendet werden (a). Hat die Punktvolke eine elliptische Form, ist die euklidische Distanz nicht die beste Wahl. In Abbildung (c) haben die Punkte $\mathbf{x}_{j,k}$ und $\mathbf{x}_{i,k}$ die gleiche Euklidische Distanz, aber Punkt $\mathbf{x}_{j,k}$ ist weiter von der Punktvolke entfernt. Mit der Mahalanobis Distanz haben beide erwähnte Punkte die gleiche Distanz (b).

dass beide Variablen (=Vektoren) \mathbf{x}_i und \mathbf{x}_j , welche aus einer Bivariaten Gaußschen Kurve stammen, unkorreliert sind. Als Beispiel einer Bivariaten Gaußschen Verteilung dient Abbildung 5.12, deren Punktvolke (a) aus einer Zufallsvariablen mit $n = 2000$ Datenpunkten erzeugt wurde und deren Korrelation von $\rho = 0$ mittels einer bekannten Kovarianzmatrix festgelegt wurde. Die Abbildung (c) stellt die Dichteschätzung im 2D Raum dar und Abbildung (e) die entsprechende Bivariate Normal Verteilung.

Die Abbildungen (b),(d) und (f) stellen die gleichen Funktionen für eine Zufallsvariable dar, deren Korrelation $\rho = 0.75$ beträgt. Obwohl Punkt $\mathbf{x}_{i,k}$ dem Zentrum näher ist als Punkt $\mathbf{x}_{j,k}$, wenn der euklidische Abstand berechnet wird, ist er aber gleichzeitig statistisch gesehen von der Punktvolke weiter entfernt. Unterscheiden sich die Varianzen der Variablen bzw. Vektoren \mathbf{x}_i und \mathbf{x}_j , erweist sich die Division durch die entsprechende Standardabweichung

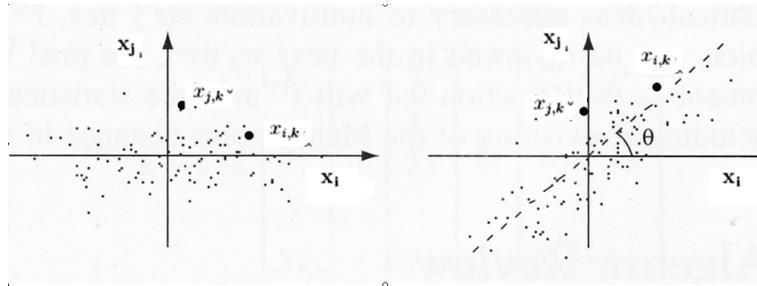


Abbildung 5.11: Zur Erläuterung der Mahalanobis Distanz (*Quelle: Wilks 2006*).

chungen nach der Zentrierung der Vektoren als sinnvoll. Die Zentrierung der Daten wird meistens dann eingesetzt, wenn es sich um Variablen handelt, die verschiedene Einheiten haben.

$$d_{i,j}^2 = \frac{(x_{i,1} - \bar{x}_i)^2}{s_{1,1}} + \frac{(x_{j,2} - \bar{x}_j)^2}{s_{2,2}} \quad (5.27)$$

Dies ist ein Spezialfall der Mahalanobis-Distanz für die Punktwolke in Abbildung 5.11 im 2D Raum, d.h. $k = 2$. Das Zentrum (=Ursprung des Koordinatensystems) ist gleich dem Mittelwert, welcher Null ist. Die Gleichung 5.27 kann weiter für bereits zentrierte Datenpunkte geschrieben werden:

$$d_{i,k} = \left[\sum_{k=1}^n \frac{(x_{i,k} - x_{j,k})^2}{s_{k,k}^2} \right]^{\frac{1}{2}} \quad (5.28)$$

Durch die Division mit der entsprechenden Standardabweichung wird die Streuung berücksichtigt. Dadurch ist die des Abstandes entlang einer der Achsen, umgekehrt proportional zur Streuung der Datenpunkte. Infolgedessen ist Punkt $x_{j,k}$ weiter entfernt vom Zentrum als Punkt $x_{i,j}$, wenn die Mahalanobis Distanz benutzt wird.

Für eine fixe Mahalanobis Distanz $d_{i,j}^2$, stellt die Gleichung eine Ellipse im 2D Raum mit konstantem Abstand im statistischen Sinn dar. Ist $s_{1,1} = s_{2,2}$, ist die Ellipse ein Kreis. Damit die Datenpunkte korrelieren, müsste die Punktwolke 5.11 um den Winkel θ rotiert werden. Im Allgemeinen weisen aber multivariate Datenfelder Korrelationen zwischen den Datenpunkten auf. Aus Gleichung folgt weiter:

$$d_{i,j} = \left[\sum_{k=1}^n \left(\frac{x_{i,k}}{s_{k,k}} - \frac{x_{j,k}}{s_{k,k}} \right)^2 \right]^{\frac{1}{2}} \quad (5.29)$$

Wendet man die Matrix-Schreibweise an, so existiert eine Matrix $[V]$ deren diagonal Elemente die Stichprobenvarianzen sind:

$$V = \begin{pmatrix} s_{1,1} & 0 & \cdots & 0 \\ 0 & s_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_{k,n} \end{pmatrix}$$

Gleichung 5.29 entspricht der skalierten Euklidischen Distanz. Weisen die Datenpunkte (=Merkmale) Korrelationen auf, werden die Datenpunkte so rotiert, dass sie hinsichtlich der neuen Koordinatenachsen unkorreliert sind.

Die Rotationsmatrix liefert die Spektralzerlegung der Kovarianzmatrix, so wie in Kapitel 5.1 beschrieben. Die Rotationsmatrix entspricht der Eigenvektormatrix. Seien die $x_1 \cdots x_n$ die Beobachtungen, so entspricht $[E]^T x_i$ der rotierten Beobachtungsmatrix, deren Elemente unkorreliert sind:

$$\text{var}([E]^T x_i) = [E]^T \text{var}(\mathbf{x}_i) [\mathbf{E}] = [\mathbf{E}]^T [\mathbf{S}] [\mathbf{E}] \quad (5.30)$$

Unter der Bedingung, dass $S = [E] [\Lambda] [E]^T$ und $[E]^T [E] = [E][E]^T = I$, folgt:

$$[E]^T [E] [\Lambda] [E]^T [E] = [\Lambda] \quad (5.31)$$

Diese Gleichung zeigt die Spektralzerlegung, gibt aber auch Information über die Skalierung. In diesem Fall sind die Eigenwerte die Varianzen der rotierten Matrix. Für alle Beobachtungen $x_i \cdots x_n$ gilt dann:

$$\begin{aligned} \text{var}([\Lambda]^{\frac{1}{2}} [E]^T x_i) &= [\Lambda]^{\frac{1}{2}} [E]^T \text{var}(x_i) ([\Lambda]^{\frac{1}{2}} [E]^T)^T \\ &= [\Lambda]^{\frac{1}{2}} [E]^T [S] [E] [\Lambda]^{\frac{1}{2}} \\ &= [\Lambda]^{\frac{1}{2}} [\Lambda] [\Lambda]^{\frac{1}{2}} \\ &= [I] \end{aligned} \quad (5.32)$$

wobei $[I]$ die Einheitsmatrix ist. Der Abstand zwischen den skalierten und rotierten Beobachtungen lautet also:

$$\begin{aligned}
 d_{s,r} &= \left[([\Lambda]^{-\frac{1}{2}} [E]^T \mathbf{x}_i - [\Lambda]^{-\frac{1}{2}} [\mathbf{E}]^T \mathbf{x}_j)^T ([\Lambda]^{-\frac{1}{2}} [\mathbf{E}]^T \mathbf{x}_i - [\Lambda]^{-\frac{1}{2}} [\mathbf{E}]^T \mathbf{x}_j) \right]^{\frac{1}{2}} \\
 &= \left[([\Lambda]^{-\frac{1}{2}} [E]^T (\mathbf{x}_i - \mathbf{x}_j))^T [\Lambda]^{-\frac{1}{2}} [\mathbf{E}]^T (\mathbf{x}_i - \mathbf{x}_j) \right]^{\frac{1}{2}} \\
 &= \left[(\mathbf{x}_i - \mathbf{x}_j)^T ([\Lambda]^{-\frac{1}{2}} \mathbf{T}^T)^T [\Lambda]^{-\frac{1}{2}} [\mathbf{E}]^T (\mathbf{x}_i - \mathbf{x}_j) \right]^{\frac{1}{2}} \\
 &= \left[(\mathbf{x}_i - \mathbf{x}_j)^T [\mathbf{E}] [\Lambda]^{-\frac{1}{2}} [\Lambda]^{-\frac{1}{2}} [\mathbf{E}]^T (\mathbf{x}_i - \mathbf{x}_j) \right]^{\frac{1}{2}} \\
 &= \left[(\mathbf{x}_i - \mathbf{x}_j)^T [\mathbf{E}] [\Lambda]^{-1} [\mathbf{E}]^T (\mathbf{x}_i - \mathbf{x}_j) \right]^{\frac{1}{2}}
 \end{aligned} \tag{5.33}$$

Der Ausdruck $[\mathbf{E}] [\Lambda]^{-1} [\mathbf{E}]^T$ entspricht der Spektralzerlegung von $[S]^{-1}$. Falls $[S]$ eine reguläre Matrix ist, gilt $[\mathbf{E}] [\Lambda] [\mathbf{E}]^T$ und deswegen gilt auch $[S]^{-1} = [\mathbf{E}] [\Lambda]^{-1} [\mathbf{E}]^T$. Auf diese Weise folgt dann für die Mahalanobis Distanz $d_{i,j}^M$:

$$d_{i,j}^M = \left[(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{x}_j) \right]^{\frac{1}{2}} \tag{5.34}$$

- S^{-1} ist die inverse Kovarianzmatrix
- Ist die Kovarianzmatrix gleich der Einheitsmatrix ($S = I$) d.h wenn die einzelnen Komponenten des Zufallsvektors \mathbf{x} paarweise unabhängig sind und jeweils Varianz 1 haben, dann ist die Mahalanobis Distanz gleich der Euklidischen Distanz. Anders ausgedrückt: für standardisierte und unkorrelierte Daten sind beide Distanzmaße identisch.
- $S \neq I$ Ist die Kovarianzmatrix ungleich der Einheitsmatrix, stellen Datenpunkte mit dem gleichen Mahalanobis-Abstand vom Zentrum im 2D Raum eine gedrehte und verzerrte Ellipse dar .

Die Mahalanobis Distanz bietet Vorteile gegenüber der Euklidischen Distanz: Sie berücksichtigt bei der Berechnung des Abstandes im Gegensatz zum Euklidischen Distanzmaß die statistischen Eigenschaften einer Datenpunktmenge, d.h Mittelwert, Varianz und Kovarianz der Datenpunkte. Auf

diese Weise wird die Korrelation zwischen den Merkmalen von der Mahalanobis Distanz berücksichtigt. Bei der euklidischen Distanz führt die Einbeziehung von stark korrelierenden Merkmalen zu einer Überbewertung der hinter den Variablen stehenden Dimensionen. Inhaltlich haben die vorhandenen Korrelationen zwischen den Merkmalen keinen Einfluss auf die Abstandsrechnung. Dadurch eignet sich die Mahalanobis-Distanz für Fälle, bei denen starke Korrelationen zwischen den Merkmalen auftreten.

In dieser Arbeit wird die Version von Yambor (Yambor 2000) [59] verwendet. Diese ist eine Vereinfachung der Moons Definition (siehe [59]), die im Vergleich bessere Ergebnisse bringt.

$$d_{i,j} = - \sum_{k=1}^n \frac{1}{\sqrt{\lambda_k}} \mathbf{x}_i \mathbf{x}_j \quad (5.35)$$

Dabei ist λ_i der entsprechende Eigenvektor, berechnet mittels Hauptkomponentenanalyse. Es wird also mit der Matrix $\Lambda^{\frac{1}{2}}$ gewichtet:
Es gilt:

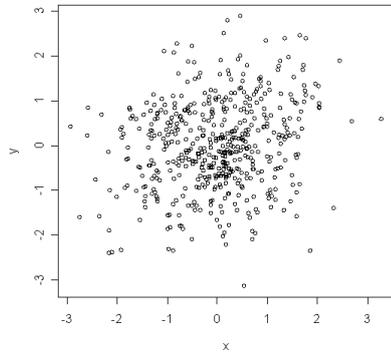
$$[\Lambda]^{\frac{1}{2}} ([\Lambda]^{\frac{1}{2}})^T = ([\Lambda]^{\frac{1}{2}})^T [\Lambda]^{\frac{1}{2}} = [\Lambda] \quad (5.36)$$

mit:

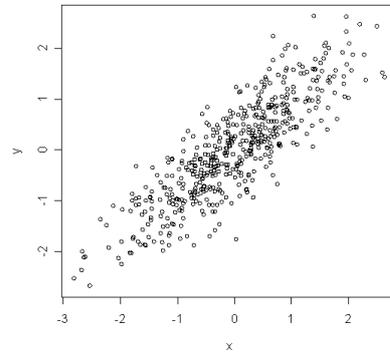
$$\Lambda^{\frac{1}{2}} = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\lambda_n}} \end{pmatrix}$$

Das negative Vorzeichen ändert lediglich den Grenzbereich von $[0, \infty]$ auf $[-\infty, 0]$. Der Charakter eines Ähnlichkeitsmaßes bleibt bestehen. Statistisch gesehen werden alle Richtungen der Eigenvektoren, die aus der Hauptkomponentenanalyse entstehen *gleich wichtig*. Das steht in Gegensatz zur gewichteten euklidischen Distanz. Anders ausgedrückt: Alle Richtungen werden gleich gewichtet.

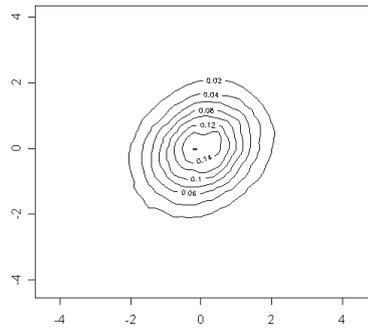
In 5.13 werden die Ergebnisse der Berechnungen des kleinsten Abstandes zwischen den *scores* unter Heranziehung von 20 PCs für den Winter (DJF) mit City-Block-Distanz (a), euklidischer(b), euklidisch gewichteter(c) und Mahalanobis Distanz (d) für die Station Feldkirch gegenübergestellt. Analog stellen die Abbildungen (e), (f), (g) und (h) dies für 7 PCs für die gleiche



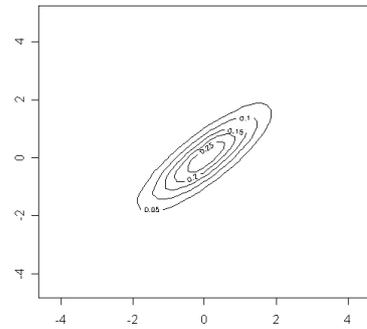
(a)



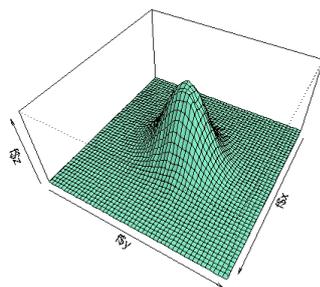
(b)



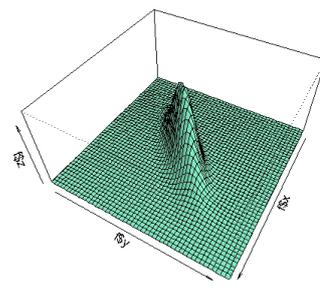
(c)



(d)



(e)



(f)

Abbildung 5.12: Dargestellt sind die simulierten Punktwolken für $n=2000$ Datenpunkte bei bekannter Kovarianzmatrix jeweils für eine Korrelation von $\rho = 0.0$ (a) und $\rho = 0.75$ (b), die entsprechenden Höhenlinien der Dichtefunktion (d) und entsprechende 3D Darstellung.

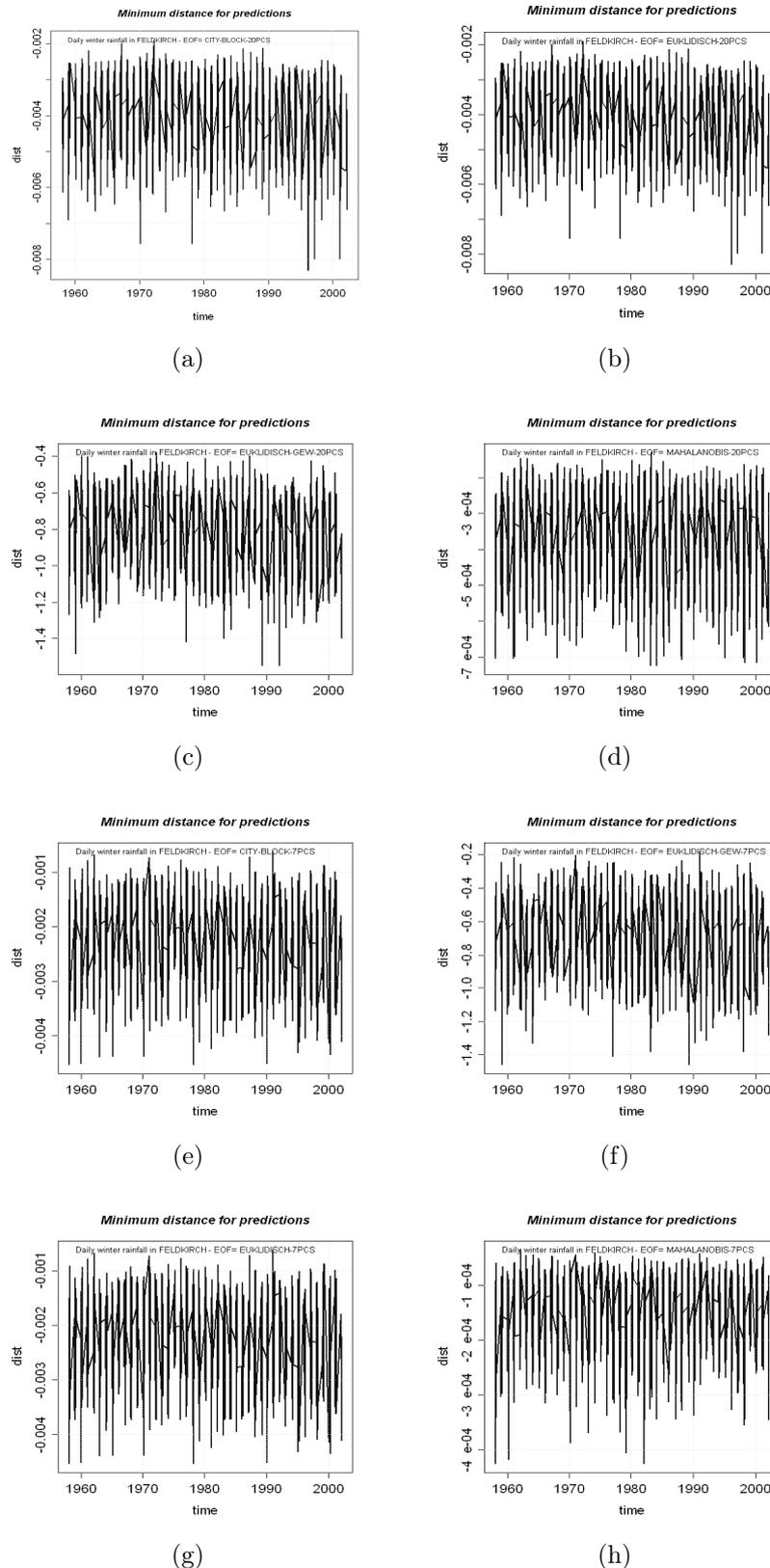


Abbildung 5.13: Ergebnisse der Berechnung des kleinsten Abstandes unter Heranziehung von 20 PCs für den Winter (DJF) mit City-Block-Distanz (a), Euklidischer Distanz (b), gewichteter Euklidischer Distanz (c) und Mahalanobis Distanz (d) am Beispiel der Station Feldkirch. Analog stellen die Abbildungen (e), (f), (g) und (h) die Ergebnisse für 7 PCs dar.

Station und Jahreszeit dar. Die Berechnung der Abstände wurde zwischen den Elementen der Spalten von der Matrix $[U]$ 5.18 für den Winter(DJF) bei 20 PCs und 7 PCs durchgeführt: zuerst wurde der Abstand zwischen den Hauptkomponenten berechnet und anschließend der kleinste numerische Wert herausgehoben. Das Programm wurde für alle 70 Stationen für jeweils 7, 11 und 20 Hauptkomponenten und für die vier Distanzmaße eingesetzt.

5.2 Die Analog-Methode

Die *Analog-Methode* ist ein statistisches Downscaling-Verfahren, welches mit anderen empirischen Downscaling-Verfahren [61] verglichen wurde. Auf Basis einer Hauptkomponentenanalyse werden ähnliche Verteilungsmuster in großräumigen meteorologischen Feldern (ERA40) gesucht. Das ähnlichste Feld wird durch die Minimierung der gewählten Distanz der Hauptkomponenten (PCs) festgelegt. Es wird davon ausgegangen, dass ähnliche großräumige Bedingungen lokal zu den selben Wetterbedingungen (d.h. ähnlichen Niederschlagsverhältnissen) führen.

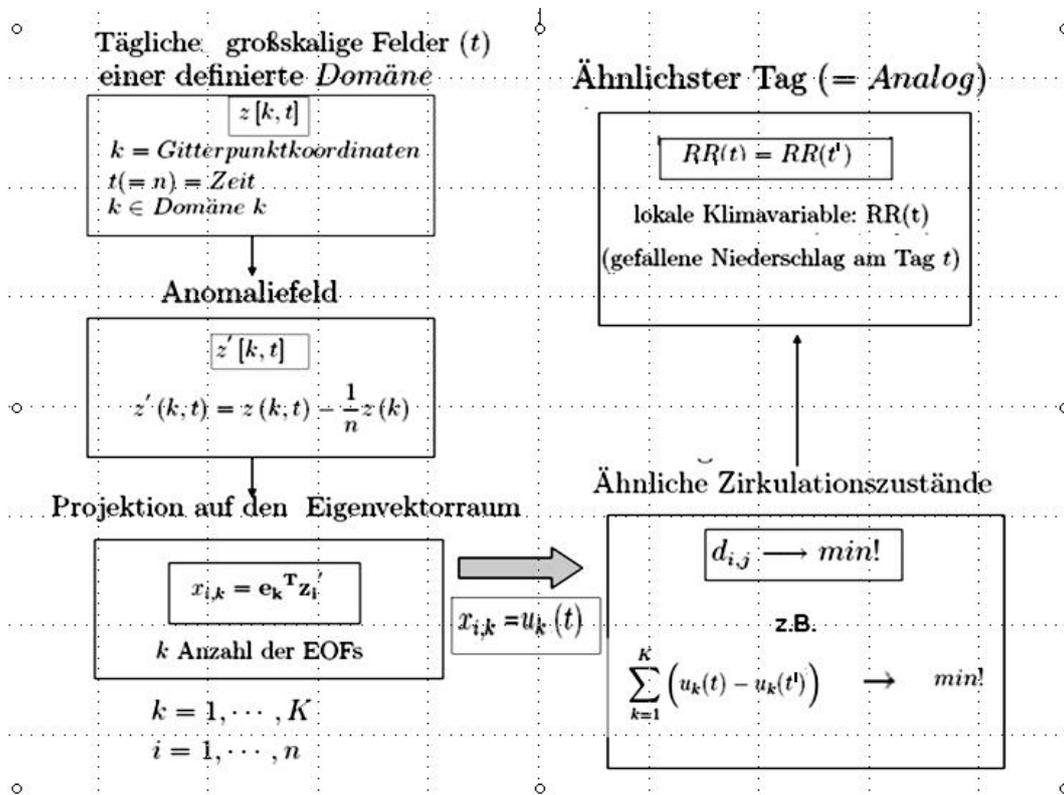


Abbildung 5.14: Schema der *Analog-Methode*.

Im Schema 5.14 wird die *Analog-Methode* in Schritten erläutert. Jeder dieser Schritte wurde in den bisherigen Kapiteln definiert und erläutert. Wie wird das ähnlichste großräumige Feld gefunden? Durch ein Distanzmaß. Folgendes konkrete Beispiel stellt die Vorgehensweise dar.

Beob.	yy	mm	dd	PC1	PC2	PC3	PC4	PC5	PC6	PC7
0	1958	1	1	-0,015302399	-0,009645228	-0,015431095	0,008425801	-0,016607004	-0,002390363	0,001239652
0	1958	1	2	-0,00287529	-0,016718457	-0,014250657	0,009174398	-0,016294824	-0,006681245	0,005015542
0	1958	1	3	0,012939134	-0,016365364	-0,000937747	0,011378873	-0,0220139	0,003682021	-0,007633483
0	1958	1	4	0,013731697	-0,001936872	0,016020911	-0,0019639	-0,029496615	-0,011683141	0,005943224
3	1958	1	5	-0,006499453	0,008706349	0,03487307	-0,014809021	-0,011203999	-0,038487315	0,004431004
8,1	1958	1	6	-0,022331777	-0,002405748	0,027644539	-0,00952459	-0,002714178	-0,007397465	-0,012505389
5,5	1958	1	7	-0,02498182	-0,013738639	-0,020747417	-0,015109978	0,002228035	0,006151749	0,002372696
3,5	1958	1	8	-0,005595538	0,000574549	-0,031909581	-0,012029009	-0,027528047	0,001199565	0,014072432
14,8	1958	1	9	-0,01723586	0,01378333	-0,00262492	0,003990311	-0,013507436	-0,033601189	0,01480578
0	1958	1	10	-0,025811071	0,011875355	0,008718462	-0,01583323	-0,017365386	-0,018372037	-0,005991246

Abbildung 5.15: Modell

Abbildung 5.15 zeigt eine lokale Zeitreihe (Taglich akkumulierter Niederschlag (Beob), der an einem bestimmten Tag (dd) des Monats (mm) und Jahres (yy) gemessen wurde). An jedem dieser Tage hat ein gewisser groskaliger atmospharische Zustand geherrscht. Es wird davon ausgegangen, dass die grosskalige Information in den 7 PCs kodiert ist. Fur die hier als Beispiel dienenden 10 Tage hat es 10 Zirkulationsmuster gegeben, welche durch die 7 PCs dargestellt werden.

Die Distanz wird zwischen den Elementen der Spalten ($PC1 \dots PC7$) berechnet. Die Ergebnisse werden mittels einer symmetrischen Distanzmatrix in 5.16 dargestellt. Die Nullwerte ergeben sich aus der Distanzbestimmung eines Datenpunktes mit sich selber. In der *Analog-Methode* wird auf diese Weise fur den Tag X der ahnlichste Tag Y gesucht und ausgeschaltet. Dadurch wird verhindert, dass der ahnlichste Tag wieder sich selbst bestimmt. Im Algorithmus wurde das gesamte Jahr, zu dem der Tag X gehort ausgeschaltet. Dadurch wird die Suche nach dem ahnlichsten Tag Y in den restlichen Jahren forciert.

Bei genauerer Betrachtung kann der kleinste berechnete Abstand leicht erkannt werden. Sucht man fur den 1. Tag den ahnlichsten Tag so erkennt man dies an dem kleinsten berechneten Wert ($d_{i,j} = d_{1,2}$). Damit wird angenommen, dass am 1. Tag die gleichen Niederschlagsverhaltnisse an dieser Station geherrscht haben wie am 2. Tag. Der Plot 5.17 stellt fur dieses Beispiel die Paare (Beobachtung-ahnlichster Tag(Analog)) dar.

Dabei kann es schon vorkommen, dass mehrere verschiedene Tage laut Distanzmatrix den gleichen Tag als ahnlichsten Tag aufweisen. Betrachtet

	[.1]	[.2]	[.3]	[.4]	[.5]	[.6]	[.7]	[.8]	[.9]	[.10]
[.1]	0	0,00024	0,00121	0,00227	0,00483	0,00269	0,00112	0,00119	0,0019	0,00205
[.2]	0,00024	0	0,00073	0,00173	0,00469	0,00318	0,00164	0,00134	0,00213	0,00276
[.3]	0,00121	0,00073	0	0,00115	0,00502	0,00321	0,00323	0,00265	0,00384	0,00364
[.4]	0,00227	0,00173	0,00115	0	0,0021	0,00257	0,0045	0,00301	0,0024	0,00233
[.5]	0,00483	0,00469	0,00502	0,0021	0	0,00178	0,00612	0,00647	0,00204	0,00162
[.6]	0,00269	0,00318	0,00321	0,00257	0,00178	0	0,00294	0,00524	0,00294	0,00099
[.7]	0,00112	0,00164	0,00323	0,0045	0,00612	0,00294	0	0,00176	0,00349	0,00258
[.8]	0,00119	0,00134	0,00265	0,00301	0,00647	0,00524	0,00176	0	0,00283	0,00309
[.9]	0,0019	0,00213	0,00384	0,0024	0,00204	0,00294	0,00349	0,00283	0	0,00128
[.10]	0,00205	0,00276	0,00364	0,00233	0,00162	0,00099	0,00258	0,00309	0,00128	0

Abbildung 5.16: Euklidische (quadratische) Distanzen der Merkmale bzw. Hauptkomponenten u_k .

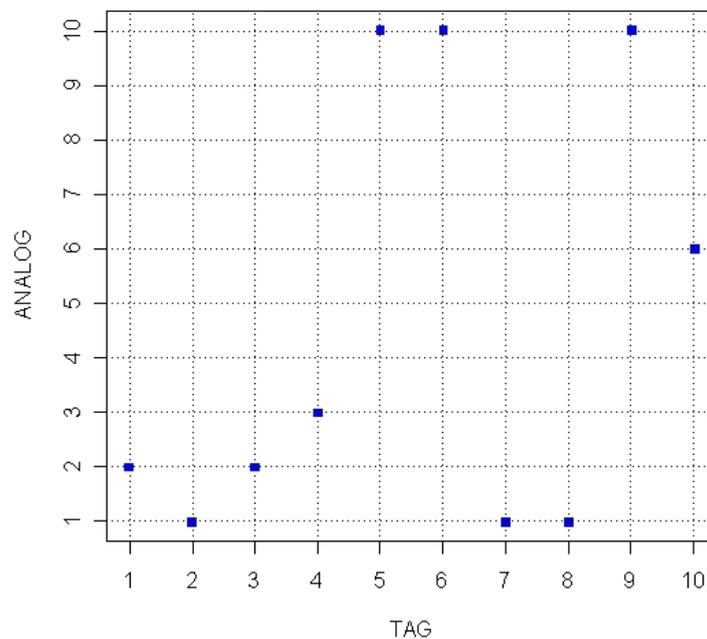


Abbildung 5.17: Darstellung der Paare Tag - ähnlichster Tag (Analog), wenn die Euklidische Distanz verwendet wird.

man die berechneten Werte genau, so wird festgestellt, dass die Unterschiede untereinander äußerst klein sind, so dass in diesem konkreten Beispiel für einen bestimmten Tag weitere Tage in Frage kommen würden. Das erscheint logisch, wenn man bedenkt, dass die „Witterung“ meistens einige Tage andauert.

Die Algorithmen Um das Verfahren umsetzen zu können, wurden die R-Packages *clim.pact*, *anm*, *cluster*, *stats* und *base* benutzt. In der Routine *predict.anm* wurden zusätzlich City-Block-Distanzmaß, Mahalanobis und die gewichtete euklidische Distanz implementiert und gleichzeitig die Suche nach dem ähnlichsten Tag im gleichen Jahr verhindert. Weitere notwendige Programme, etwa zur Berechnung bestimmter statistischer Größen, wurden mit der R-Programmiersprache ⁹ umgesetzt.

⁹<http://www.r-project.org/>

Kapitel 6

Verifikation

6.1 Statistische Methoden

Korrelationsanalyse: Um die Modelleistung zu beurteilen werden hier folgende Arten von Korrelationen eingesetzt. Der Korrelationskoeffizient ist zwar eine notwendige aber keine ausreichende Bedingung, um die Modelleistung zu beurteilen. Es müssen daher zusätzliche Größen berechnet werden.

Der Produkt-Moment-Korrelationskoeffizienten nach PEARSON gibt den Grad an linearem Zusammenhang zwischen Beobachtung und Vorhersage an. Um die Ergebnisse des Downscaling auf saisonaler Basis beurteilen zu können, wurde statt der Korrelation nach Spearman, die Korrelation nach Pearson angewandt. Das wird damit gerechtfertigt, dass in diesem Fall saisonale Summen vorliegen (Niederschlagssummen), die in ihrer Verteilung normal approximiert sind.

$$r_s = \frac{\sum_{i=1}^N (F_i - \bar{F})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^N (F_i - \bar{F})^2} \sqrt{\sum_{i=1}^N (O_i - \bar{O})^2}} \quad (6.1)$$

\bar{F} ... arithmetisches Mittel der Vorhersagen

F_i ... i - Werte der Vorhersage

O_i ... i - Werte der Beobachtungen

O ... arithmetisches Mittel der Beobachtungen

N ... Anzahl der Beobachtungen

Der Korrelationskoeffizient kann Werte zwischen -1 und +1 annehmen,

Rangkorrelationsrechnung nach SPEARMAN Um die Stärke zwischen zwei nicht normalverteilten Variable zu bestimmen, im aktuellen Fall zwischen den täglichen Niederschlagszeitreihen, meistens rechtsschiefe Verteilungen, kann die parameterfreie Rangordnung der Korrelationskoeffizienten nach Spearman benutzt werden. Dabei werden jede Beobachtung und jeder modellierte Wert durch einen Rang ersetzt je nach Gesamtheit aller Beobachtungen und modellierter Werte. Man erhält dabei eine Verteilungsfunktion der Ränge. Ein Vergleich zwischen den Ränge stellt diese Art von Korrelation dar.

$$r_s = \frac{\sum_{i=1}^N (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^N (R_i - \bar{R})^2} \sqrt{\sum_{i=1}^N (S_i - \bar{S})^2}} \quad (6.2)$$

S_i ... Rangzahlen der Modellwerte

R_i ... Rangzahlen der Beobachtungen

Für die Berechnung der Rangkorrelationskoeffizienten wird aber noch die sum squared difference of ranks benötigt:

$$D = \sum (R_i - S_i)^2 \quad (6.3)$$

Damit folgt:

$$r_s = 1 - \frac{6 \sum (D_i^2)}{n^3 - n} \quad (6.4)$$

dabei werden sowohl Beobachtungen wie Vorhersage in aufsteigende Folge gebracht. Im Fall gleicher Rangplätze (wenn B_i und V_i numerisch gleich sind, so genannte „Bindungen“) werden gleiche Rangplätze vergeben. Ist die Anzahl der Bindungen höher als 25 Prozent, muss die obige Formel korrigiert werden:

$$r_s = 1 - \frac{D_i^2}{n^3 - n - T_O - T_F} \quad (6.5)$$

$i = 1 \dots n$

$D_i \dots$ Differenzen zwischen den Rangplätzen

mit:

$$T_O = \frac{1}{2} \sum_{k=1}^J [h_k^3(O) - h_k(O)] \quad (6.6)$$

$$T_F = \frac{1}{2} \sum_{k=1}^J [h_k^3(F) - h_k(F)] \quad (6.7)$$

hier ist h_k , die Anzahl der gleichen Rangplätze bei den Beobachtungen und Vorhersagen, angegeben durch k .

6.1.1 Prüfmaße für kontinuierliche Variablen

In der Literatur findet man eine große Anzahl von Prüfmaßen. Diese dienen dazu, die Vorhersagegüte besser einzuschätzen:

- **Mittlerer Fehler (Mean error, Bias)**

Durch den Bias wird die mittlere Richtung der Abweichung der Beobachtungen ausgedrückt. Ist der mittlere Fehler negativ, deutet dies darauf hin, dass das Modell die beobachteten Werte im Mittel unterschätzt. Im Fall eines positiven Fehlers ist das gleichbedeutend mit einer Überschätzung der Beobachtungen durch das Modell.

$$ME = \frac{1}{N} \sum_{i=1}^N (F_i - O_i) \quad (6.8)$$

- **Mittlerer absoluter Fehler (Mean absolute error, (MAE))**

Dieses lineare Maß gibt das Mittel aller Fehler an. Über dessen Richtung kann keine Aussage getroffen werden.

$$MAE = \frac{1}{N} \sum_{i=1}^N |F_i - O_i| \quad (6.9)$$

Sowohl der MAE wie der ME würden im Fall einer guten Vorhersage den Wert 0 annehmen. Der Unterschied zwischen den beiden Maßen besteht darin, dass im Vergleich zum linearen MAE der ME empfindlich

auf große Fehler ist. Der MAE zusammen mit der Wurzel des quadratischen Mittelfehlers stellen ein Maß für den üblichen Vorhersagefehler dar.

- **Wurzel aus dem quadratischen Mittelfehler (root mean square error, RMSE)** eine andere Fehlerangabe ist der RMSE, ein lineares Maß, durch das der mittlere Betrag des Fehlers ausgedrückt wird, ist aber mit dem Quadrat des Fehlers gewichtet. Aus einem Vergleich mit dem mittleren absoluten Fehler (MAE) ergibt sich, dass größere Fehler im quadratischen Mittel eine höhere Gewichtung bekommen. Der RMS gibt keine Richtung des Fehlers an.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (F_i - O_i)^2} \quad (6.10)$$

Zusätzlich zur Größenordnung des systematischen und des allgemeinen Vorhersagefehlers ist das Feststellen der Leistungsfähigkeit einer Methode wichtig.

- **Reduktion der Varianz, RV** dient als Maß zur Festlegung der Vorhersageleistung bedient man sich „skill Scores“. Es handelt sich um ein Maß für die relative Genauigkeit einer Vorhersage relativ zur Referenzprognose. Für die Referenzprognose kann der klimatologische Mittelwert benutzt werden.

$$SS = 1 - \frac{MSE}{MSE_{Klima}} \quad (6.11)$$

Wird SS so definiert, wird auch von Reduktion der Varianz (RV) gesprochen. Dieses Maß ist vollständig dem in manchen Arbeiten genannten Coefficient of Efficiency E äquivalent.

$$E = 1 - \frac{\sum_{i=1}^N (O_i - F_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \quad (6.12)$$

- **Anomaliekorrelationskoeffizient, ACC**

Mit den ACC kann ein Zusammenhang zwischen Beobachtungs- und Vorhersagefeldern, bei dem beide in Anomalien umgewandelt werden, hergestellt werden, d.h. die Abweichungen vom klimatologischen Mittel des Beobachtungs- und Vorhersagefelds.

$$ACC = \frac{\sum_{i=1}^N (F_i - C_i)(O_i - C_i)}{\left[\sum_{i=1}^N (F_i - C_i)^2 (O_i - C_i)^2 \right]^{1/2}} \quad (6.13)$$

Der ACC kann Werte zwischen -1 und 1 annehmen. In der Literatur gilt ACC=0.6 als Angabe für eine sinnvolle Vorhersage. Der ACC ist dann positiv wenn die beide Anomalien gleichsinnig sind, negativ im umgekehrten Fall.

6.1.2 Prüfmaße für kategoriale Variable

Um die Ergebnisse, bzw. Prognosen auf Tagesbasis genauer zu untersuchen, wurden Kontingenztabellen erzeugt, mit dem Ziel, die Anzahl an „estimates“ und Beobachtungen besser abschätzen zu können. Auf diese Weise werden die „hits“ und „misses“ genau quantifiziert. Das ist deshalb interessant weil damit festgestellt werden kann ob für einen Tag an dem Niederschlag gefallen ist, das Modell in der Lage ist einen Tag (den Ähnlichsten) zu liefern, an dem ebenfalls Niederschlag gefallen ist.

In Abbildung 6.1 wird eine Kontingenztafel dargestellt. Mit Hilfe der Kontingenztabellen können so genannte Genauigkeits- („Accuracy measures“) und Qualitätsmaße („Skill Scores“) berechnet werden .

Abbildung 6.1 erläutert das Prinzip der Kontingenztafel zur Berechnung von Genauigkeitsmaßzahlen.

6.1.3 Genauigkeitsmaße

Entdeckungswahrscheinlichkeit-Probability of Detection oder Pre-figurance(POD): Das POD gibt das Verhältnis von positiven Ereignissen zur Anzahl an positiven Beobachtungen an, oder anders gesagt, den Prozentanteil der vorhergesagten Ereignisse.

$$POD = \frac{a}{a + d} \quad (6.14)$$

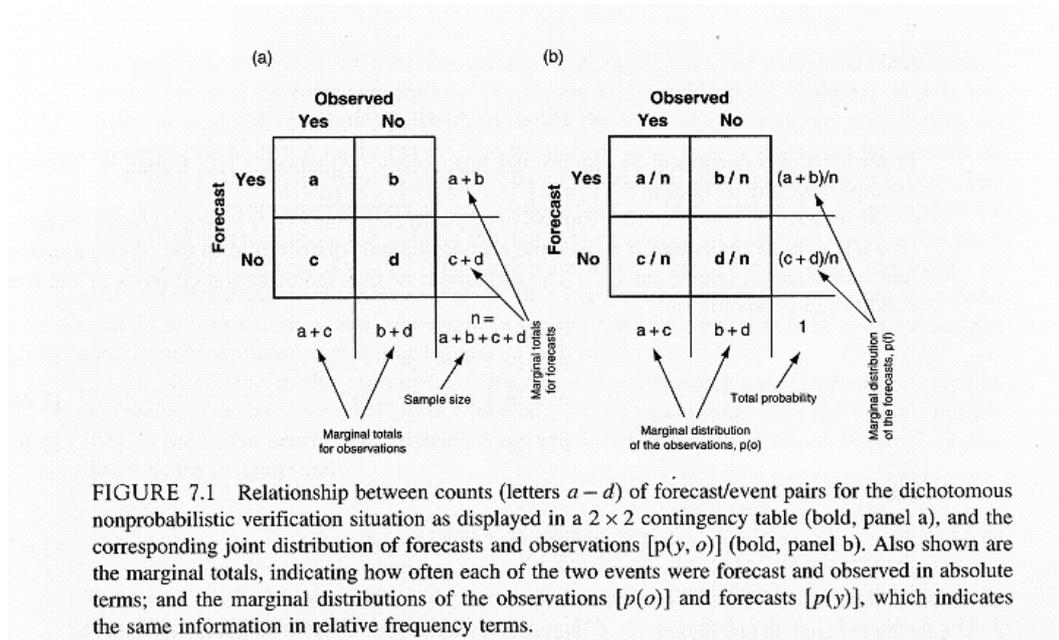


Abbildung 6.1: (Quelle: Wilks 2006)

Das POD erreicht Werte von 0 bis 1. Mit 1 hat man eine perfekte Prognose. Wenn die Anzahl der missglückten Ereignisse höher ist als die Zahl der positiven Ereignisse, anders gesagt, wenn die Zahl der Niederschlagsereignisse überschätzt wurde entspricht eine gute Vorhersage nicht der Wirklichkeit. In diesem Fall wird die **Rate Falschen Alarms zusätzlich berechnet (False Alarm Ration), FAR** .

$$FAR = \frac{c}{a + c} \quad (6.15)$$

Der POFD stellt eine perfekte Prognose dar, wenn er den Wert 0 annimmt; er kann Werte zwischen 0 und 1 annehmen.

Bias Score: der Bias Score liefert das Verhältnis von Beobachtungen und abgeschätzten Werten, ohne jegliche Information über die Verteilung der Größen. Er gibt das Verhältnis von positiver Beobachtung zu positiver Vorhersage, genauer gesagt zwischen den rain estimates und rain observation wieder.

$$BiasScore = \frac{a + b}{a + c} \quad (6.16)$$

Der Bias kann Werte zwischen 0 und unendlich annehmen und ist derart zu interpretieren, dass im Fall von 1 der vorhergesagte Niederschlag gleich dem beobachteten Niederschlag ist. Wenn der Bias Wert höher als 1 ist, dann bedeutet dies, dass eine Überschätzung vorliegt, bei einem Wert, der niedriger als 1 ist handelt es sich um eine Unterschätzung.

6.1.4 Qualitätsmaße

Bei den so genannten Qualitätsmaßen oder Skill Scores geht es darum, die Vorhersagen mit Referenzprognosen zu vergleichen. Die Referenzprognosen können etwa klimatologische Mittel sein. Wenn die Vorhersage nur auf Zufall aufgebaut ist, kann das mithilfe der Qualitätsmaße aufgedeckt werden. Qualitätsmaße basieren auf folgender Gleichung:

$$SS = \frac{A_{estimated} - A_{reference}}{A_{perfect} - A_{reference}} \quad (6.17)$$

wobei A ein Genauigkeitsmaß ist. **Heidke Skill Score-HSS** bedient sich als Grundgenauigkeitsmaß der Trefferrate, die sich aus der Verwendung der Zufallsvorhersage ergibt, folgend der vorigen Gleichung, das heißt der üblichen Definition der Skill Scores. Die richtige Vorhersage wird definiert durch

$$p(y_1)p(o_1) = \frac{a+b}{n} \frac{c+d}{n} = \frac{(a+b)(c+d)}{n^2} \quad (6.18)$$

und die korrekte „nicht“ Vorhersage durch

$$p(y_2)p(o_2) = \frac{b+d}{n} \frac{c+d}{n} = \frac{(b+d)(c+d)}{n^2} \quad (6.19)$$

damit ist der Heidke Skill Score definiert durch:

$$HSS = \frac{(a+d)n - [(a+b)(c+d) + (b+d)(c+d)]}{1 - [(a+b)(c+d) + (b+d)(c+d)]} \quad (6.20)$$

$$= \frac{2(ad - bc)}{(a+c)(c+d) + (a+b)(b+d)} \quad (6.21)$$

Der HSS kann Werte zwischen -1 und 1 annehmen. Wenn er den Wert 1 annimmt liegt eine ideale Prognose vor, bei Wert 0 ist die Vorhersage gleich der Referenzprognose, negative Werte sind ein Zeichen dafür, dass die Vorhersage schlechter ist als die Referenzprognose.

True Skill Statistics-TSS Zur Referenzprognose wird auch hier die Trefferrate benutzt - mit dem Unterschied, dass die Wahrscheinlichkeit einer richtigen Vorhersage und die Wahrscheinlichkeit einer „Nicht“- Vorhersage im Nenner ohne systematische Abweichungen angenommen wird. Damit ist die Wahrscheinlichkeit einer korrekten Vorhersage ohne systematische Abweichung:

$$TSS = \frac{ad - bc}{(a + c)(b + d)} \quad (6.22)$$

Aus der Erweiterung:

$$a + b = 0 \quad (6.23)$$

lässt sich die Gleichung so schreiben:

$$TSS = \frac{a}{a + d} - \frac{b}{b + d} \quad (6.24)$$

was gleichbedeutend ist mit:

$$TSS = POD - POFD \quad (6.25)$$

Ähnlich wie bei der anderen Maßen kann der TSS Werte zwischen -1 und 1 annehmen, wobei ein negatives Ergebnis bedeutet, dass die Referenzprognose besser ist. Ist das Ergebnis gleich 1, bedeutet dies eine ideale Prognose, bei 0 ist die Prognose gleich der Referenzprognose: Im Vergleich zu den anderen Skill Scores bietet der TSS eine Verifikation an, die frei von systematischen Abweichungen ist.

Die Ergebnisse der hier vorgestellten und erklärte Verifikationsmaße können in Kapitel A nachgelesen werden. Für die Evaluierung der verschiedenen Distanzmodelle wird die übliche Zusammenfassung in Form von Minimum, Mittelwert und Maximum präsentiert.

6.1.5 Statistische Tests

Manche statistische Tests setzen eine Normalverteilung der Zufallsvariablen der Grundgesamtheit voraus. Es ist deshalb sinnvoll diese Voraussetzung auf ihre Gültigkeit zu überprüfen. Zu diesem Zweck wurde ein Verteilungstest ausgewählt um festzustellen, ob die saisonalen Niederschlagsbeobachtungen und die modellierten saisonalen Niederschlagsreihen einer Normalverteilung folgen. Aufgrund des *Gesetz der großen Zahlen*¹ und durch die Ergebnisse des Shapiro-Wilk Tests wird die Annahme der Normalverteilung unterstützt. In diesem Abschnitt werden die Grundlagen des benutzen statistischen Tests erklärt und deren Interpretation erläutert.

Shapiro Wilk Test: Der Shapiro Wilk Test W prüft die Nullhypothese ob die Beobachtungen $x_1 \dots x_n$ einer Zufallsstichprobe einer normalverteilten Zufallsvariable zugeordnet werden können. Damit die Prüfgröße W berechnet werden kann, muss zuerst die Verteilung des Quotienten aus zwei Schätzungen betrachtet werden: das Quadrat einer kleinsten Fehlerquadratschätzung der Regressionsgeraden und die Stichprobenvarianz. Liegt eine Normalverteilung vor, sind diese Schätzungen identisch und ergeben im Verhältnis 1. Kleine Werte von W weisen auf kleine Abweichungen von der Normalverteilung hin. Der W -Wert ist definiert als:

$$W = \frac{(\sum_{i=1}^N a_i x_i)^2}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (6.26)$$

Dabei sind x_i die Beobachtungen und die a_i konstante Werte. Die a_i werden aus Maßzahlen der Ordnungsstatistik einer normalverteilten Zufallsvariable in Abhängigkeit vom Stichprobenumfang n festgelegt. Der Shapiro Wilk Test hat im Vergleich zu anderen Testverfahren die höchste Treffsicherheit. Das Signifikanzniveau wurde mit 0.05 festgelegt.

Abbildung 6.2 zeigt die Häufigkeitsverteilung der saisonalen Niederschlagssummen für die Station Innsbruck Flughafen. Die entsprechenden Häufigkeitsverteilungen der modellierten Reihen werden in Abbildung 6.3 dargestellt. Ein Vergleich zwischen den Abbildungen lässt leicht erkennen, dass sich die Häufigkeit der verschiedenen Niederschlagsklassen unterscheiden, je

¹welches besagt, je öfter ein Zufallsexperiment durchgeführt worden ist, desto näher kommt die relative Häufigkeit eines Zufallsergebnisses an die theoretische Wahrscheinlichkeit heran

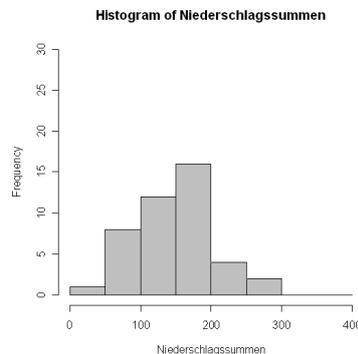


Abbildung 6.2: Saisonale Häufigkeitsverteilung (DJF) für die Station Innsbruck-Flughafen für die Periode 1958-2000.

nach „Distanzmodell“.

Um zu überprüfen wie signifikant die saisonal berechneten Korrelationskoeffizienten sind wird ein Signifikanztest angewandt. Dieser Test wird jedoch nur für die ermittelten Korrelationen zwischen den saisonalen Zeitreihen gemacht. Für die berechneten Korrelationen zwischen den täglichen Beobachtungszeitreihen und den modellierten akkumulierten Niederschlag wird auf Signifikanzprüfung verzichtet, da sich die Korrelationen als sehr niedrig erweisen.

Wie bereits im vorigen Kapitel erklärt gibt die Korrelation einen Hinweis auf Stärke und Richtung der Abhängigkeit zwischen Beobachtungen und Modell. Mit einem sinnvollen Test kann geprüft werden, ob dieser Wert tatsächlich signifikant von 0 verschieden ist. Zur Bestimmung des Korrelationskoeffizienten wird ein **Korrelationstest der auf die t-Verteilung** beruht, eingesetzt. Der t-Wert lässt sich folgendermaßen berechnen:

$$t = \frac{n - 2}{1 - r_{xy}^2} * r_{xy} \quad df = n - 2 \quad (6.27)$$

Der t-Wert gibt die Grenze an, bei der und oberhalb welcher ein Unterschied auf dem 100 σ %-Niveau als statistisch signifikant anzuerkennen ist. Unterhalb dieser Grenze gilt es, den Unterschied als zufällig einzustufen. Der r_{xy} ist hier die Korrelation, während n die Anzahl der Messungen angibt. Die Anzahl der Freiheitsgrade lässt sich durch „df“ berechnen.

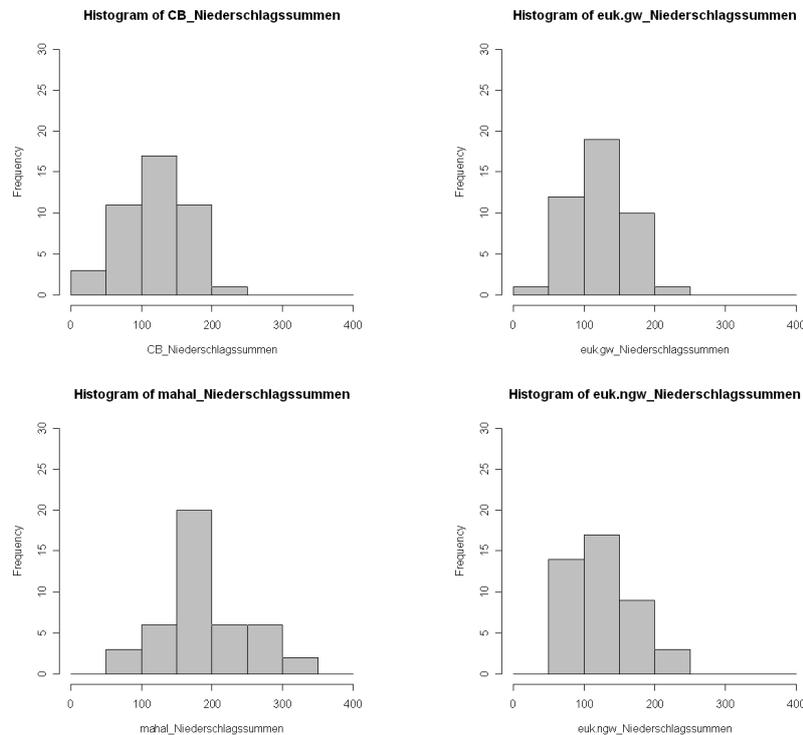


Abbildung 6.3: Saisonale Häufigkeitsverteilung für die modellierte Zeitreihe der Station Innsbruck Flughafen für die Periode 1958-2000.

Es wird in ähnlicher Weise vorgegangen wie beim Shapiro Wilks Test. Die aufgestellten Hypothesen für den Korrelationstest (zweiseitigen Test) lautet:

$$H_0 : r = 0 \quad (6.28)$$

$$H_A : r \neq 0 \quad (6.29)$$

Abbildung 6.4 ist ein Beispiel eines Tests zur Überprüfung der Korrelation auf ihrer Signifikanz zwischen saisonalen Niederschlagssummen der Station Innsbruck Flughafen und die modellierte Zeitreihe des „City-Blocks-Distanzmodell“. Dieser Test ergibt 0.51 als Korrelation, der Wert der Prüfgrösse ist 3.8 bei 41 Freiheitsgraden (Die Anzahl der Beobachtungen entspricht der Anzahl der Jahre, d.h $n = 43$). Die Ergebnisse dieses Tests werden folgendermaßen interpretiert: die Wahrscheinlichkeit einer t-Verteilung mit 41 Freiheitsgrade einen grösseren Wert als 3.8 zu erreichen beträgt 0

```
Pearson's product-moment correlation

data: beo and ci
t = 3.7969, df = 41, p-value = 0.000476
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2476327 0.7027354
sample estimates:
   cor
0.5100422
```

Abbildung 6.4:

(p-Wert). Da diese Wahrscheinlichkeit gering ist, d.h die festgelegte Irrtumswahrscheinlichkeit von $\alpha = 0.05$ ist in diesem Fall grösser als der berechnete p-Wert von 0, darf die Nullhypothese verworfen werden.

Eine andere Möglichkeit der Überprüfung ist der Vergleich des berechneten t-Wertes mit dem tabellierten t-Wert. Die Ablehnung der Nullhypothese erfolgt, wenn:

$$|t| > T_{1-\frac{\alpha}{2}}(n-2) \quad (6.30)$$

Die Ergebnisse der Signifikanztests für die Korrelationen der saisonalen Zeitreihen für die jeweilige Jahreszeit befinden sich in Kapitel A. In der ersten Zeile der Tabellen findet man die Anzahl der Stationen, deren Korrelationskoeffizient vom Test als signifikant eingestuft worden sind. In der zweiten Zeile wird diese Zahl in Prozent angegeben.

Kapitel 7

Ergebnisse

In diesem Kapitel werden die Ergebnisse der Downscaling-Methode auf Tages- und saisonaler Basis in zwei Hauptteilen dargestellt. Im ersten Hauptteil werden einzelne Stationen nach festgelegten Kriterien ausgewählt und deren Ergebnisse dargestellt. Im zweiten Hauptteil wird die Evaluierung der Distanzmaße bezüglich der *Analog-Methode* diskutiert. Der erste Hauptteil wird seinerseits in einem Teil als Tageszeitreihen, sowohl von Seite der Beobachtung wie auch von den Distanzmodellen her dargestellt. Im zweiten Teil werden die Ergebnisse der gleichen Stationen auf saisonaler Basis in gleicher Form präsentiert. Die StartClim Stationen werden nach geographischer Lage in 7.1 aufgelistet :

- **Bergstationen:** Patscherkofel, Mooserboden, Sonnblick, Villacher-Alpe, Krippenstein, Feuerkogel, Kanzelhöhe, Preitenegg und Schoeckl; insgesamt 9 Stationen (13 %), wobei im Projekt StartClim für den Sonnblick keine Niederschlagsreihen vorliegen
- **nicht-alpine oder Talstationen:** vorwiegend im Osten befindliche Stationen; insgesamt 9 Stationen (13%)
- **Stationen, die über über 1200m Seehöhe liegen:** vorwiegend im Westen; insgesamt 52 Stationen (74%)

Zur Darstellung der Ergebnisse einiger ausgewählter Stationen werden folgende Kriterien gewählt:

- **Seehöhe:** Bergstationen, Stationen über 1200m und nichtalpine oder Talstationen

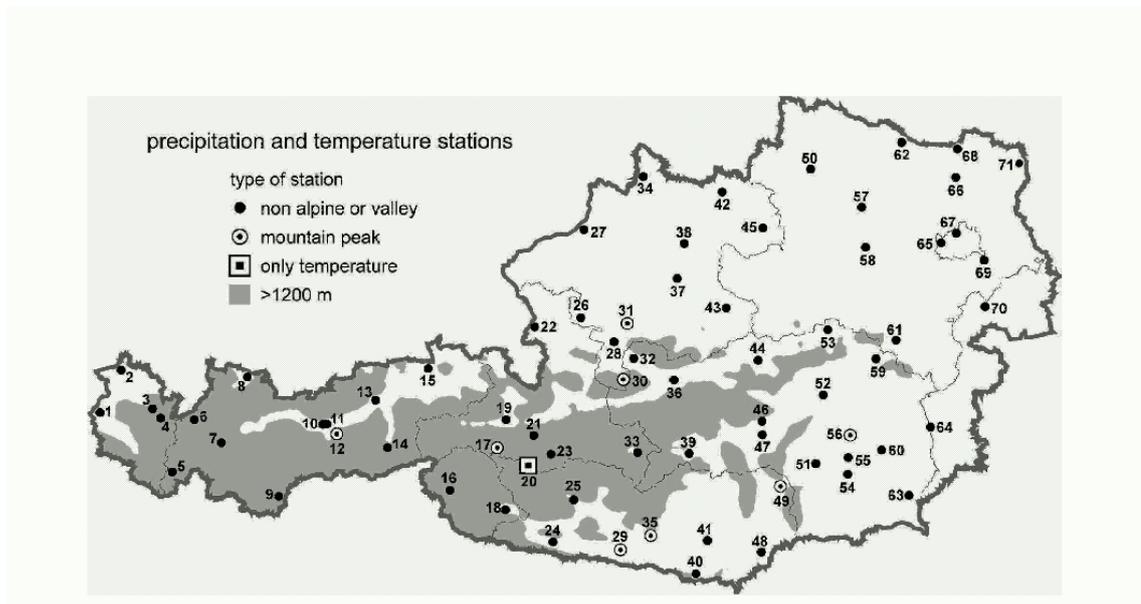


Abbildung 7.1: Orographische Lage der StartClim-Stationen: Unterteilung in nicht-alpine oder Talstationen und Bergstationen. Die graue Fläche repräsentiert Gebiete über 1200 m (*Quelle: Matulla et al. 1995*).

- **geographische Lage:** Unterscheidung in Anlehnung an die Unterteilung Österreichs in homogene Niederschlagsregionen nach Ehrendorfer in 4.6 für den Winter und Sommer: Zur besseren Handhabung der Regionen werden sie Nord- Süd- und Westregion genannt.

Gleichzeitig werden als Zwischenergebnisse in der Evaluierung der Distanzmaße bei der *Analog-Methode* statistische Maßzahlen offengelegt. Diese sind

- *Korrelation nach Spearman:* da die Niederschlagsreihen nicht normalverteilt sind, wird hier die Korrelation nach Spearman berechnet, um den Grad an linearem Zusammenhang zwischen Beobachtung und Modell zu bestimmen
- *mittlere Niederschlagshöhe:* berechnet aus der Division der Summe des gesamten Niederschlags über alle 43 Jahre (1958-2000) durch die Anzahl der Tage der jeweiligen Jahreszeit DJF und die Anzahl der Jahre. Für einen Vergleich wurde das auch für die modellierten Reihen be-

rechnet.

- *mittlere Anzahl der Niederschlagstage*: damit ist die Zahl der Tage mit einem Niederschlag von mindestens 0.1 mm/Tag gemeint: Diese Anzahl ist natürlich bezogen auf die jeweilige Jahreszeit und kann deshalb die Anzahl ihrer Tage nie überschreiten.

7.1 Downscaling auf Tagesbasis

Von den 70 Stationen wurden einige Stationen nach den vorher festgelegten Kriterien ausgewählt. Dargestellt werden die modellierten Tageszeitreihen für die vier Distanzmaße (euklidische-, City-Block-, euklidische gewichtete- und Mahalanobis Distanz) für Winter (DJF) und Sommer (JJA). Für die Modellierung wurden die ersten 20 PCs herangezogen. Es handelt sich um Zwischenergebnisse, um die *Analog-Methode* auf Tagesbasis zu beurteilen.

Nichtalpine oder Talstationen (unterhalb von 1200m) Wien Hohe Warte (Osten), Klagenfurt (Süden) und Feldkirch (Westen) werden gewählt.

Wien Hohe Warte:(Abbildungen 7.2 und 7.3) Die Auswertung für diese Station ist repräsentativ für die Stationen im Osten Österreichs. Die Korrelation zwischen der Beobachtung und den modellierten Reihen ist sehr niedrig und weist auf kaum vorhandene Zusammenhänge hin. Die Unterschiede zwischen den Korrelationen sind minimal (0.10 bis 0.25). Gleichzeitig ist kein Unterschied zwischen Winter (DJF) und Sommer(JJA) in der Performance des Modells erkennbar. Das Modell gibt annähernd den gleichen Mittelwert für jedes der „Distanzmodelle“ und auch eine ähnliche Anzahl von Niederschlagstagen, wobei die Mahalanobis-Distanz etwas mehr Niederschlagstage verzeichnet

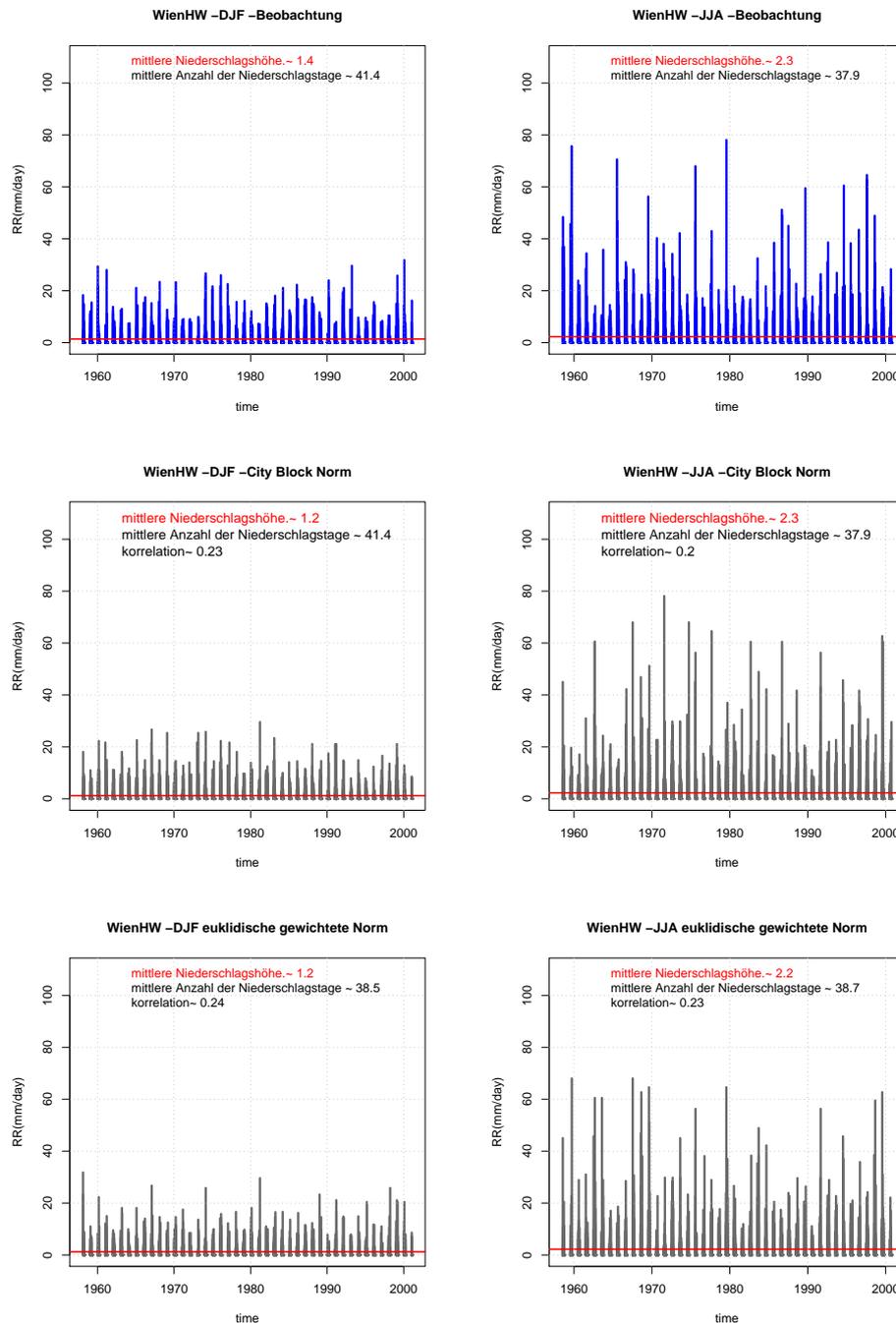


Abbildung 7.2: Wien-Hohe Werte: Darstellung der Beobachtungszeitreihe und modellierte Tageszeitreihe für Winter (DJF) und Sommer(JJA). Zur Suche des ähnlichsten Tages wurden die City Block Norm, die euklidische gewichtete Norm verwendet.

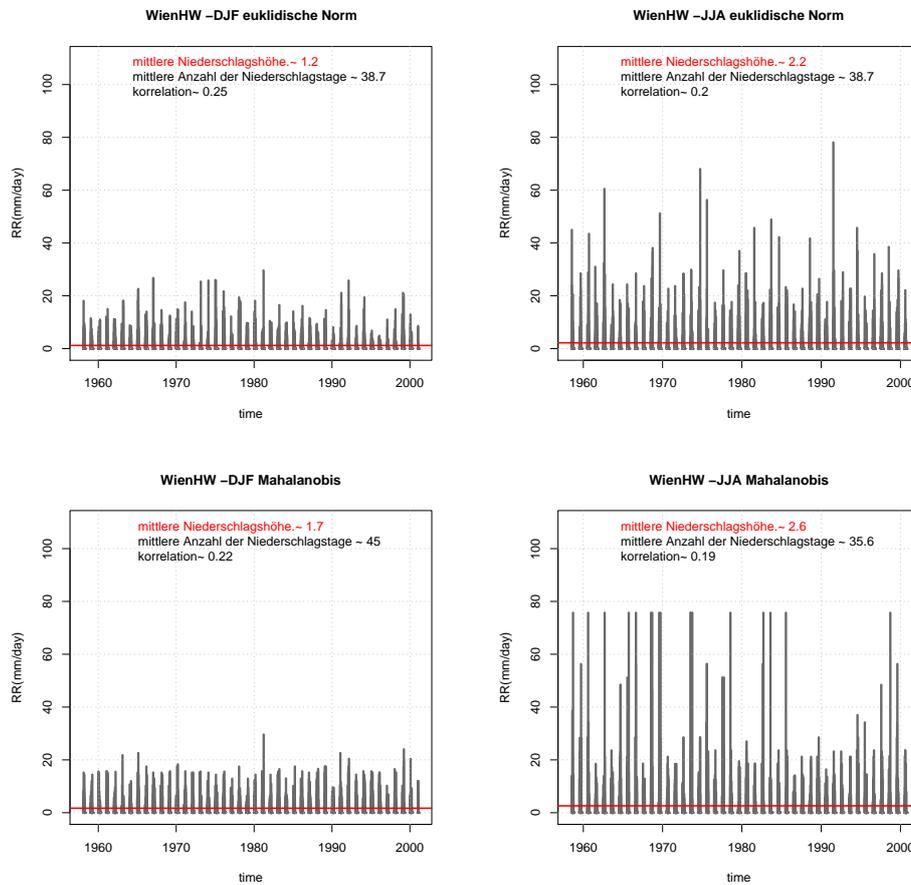


Abbildung 7.3: Wien-Hohe Warte: Darstellung der modellierten Tageszeitreihe für Winter(DJF)und Sommer (JJA). Zur Suche des ähnlichsten Tages wurden die Euklidische Distanz und die Mahalanobis Distanz verwendet.

Feldkirch(Abbildungen 7.4 und 7.5) Diese Station befindet sich im äußersten Westen Österreichs und weist im Vergleich zum Osten und Süden eine höhere Korrelation im Winter auf. Die beste Korrelation liefert die euklidische gewichtete Norm mit 0.46. Die anderen Distanzmaße schneiden in der Korrelation schlechter ab. Das schlechteste Ergebnis findet man bei der Mahalanobis-Distanz. Im Vergleich zum Winter sind die Korrelationen im Schnitt um ein Drittel höher als im Sommer. Die mittlere Niederschlagshöhe bei der City-Block-Distanz ist mit 40.8 Tagen gleich dem tatsächlichen Wert. Im Schnitt werden von der Mahalanobis-Distanz mehr Niederschlagstage als Analoga herangezogen als von den anderen Distanzmaßen. Vergleichsweise liegt die mittlere Anzahl der Niederschlagstage bei der Beobachtungsreihe

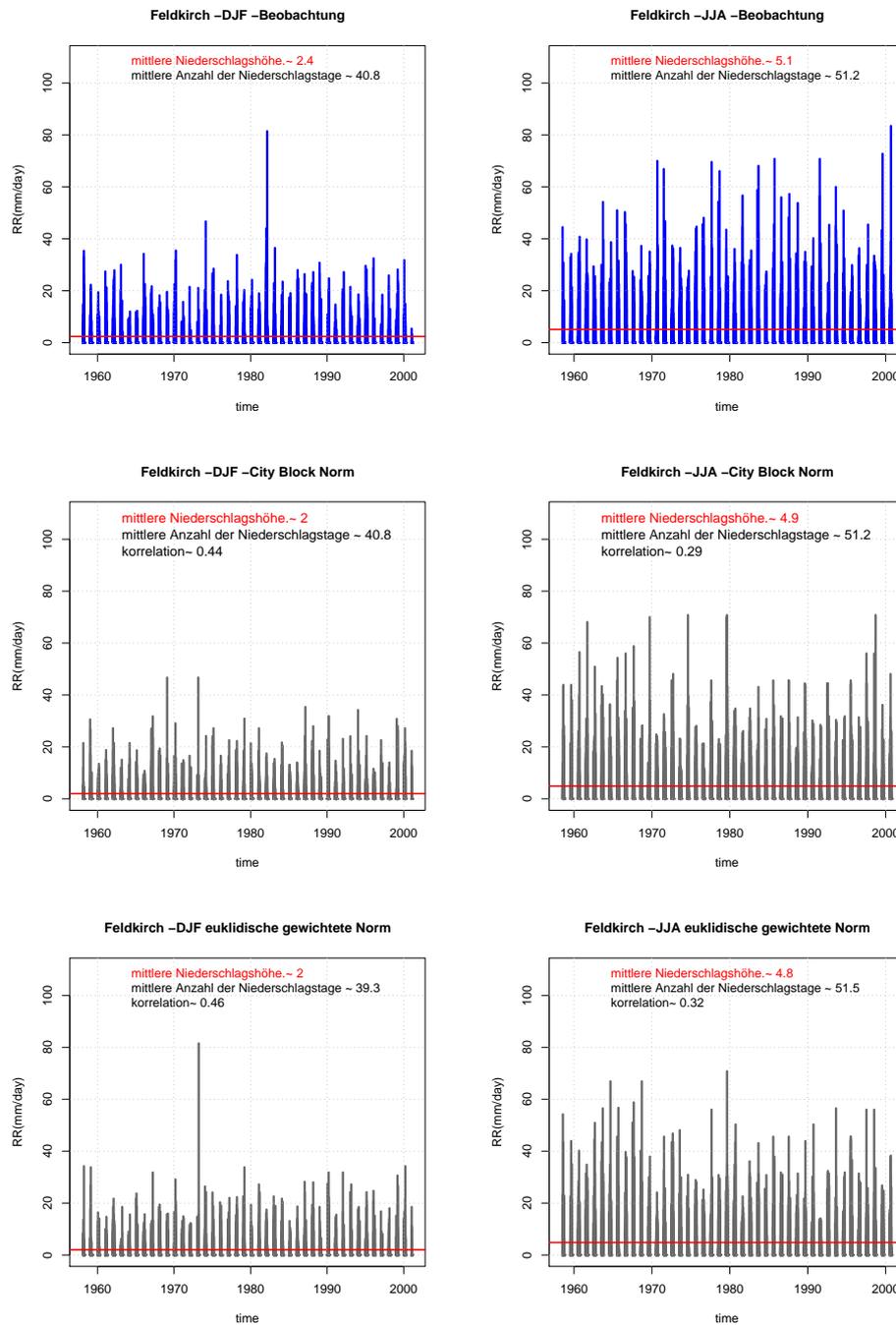


Abbildung 7.4: Feldkirch: Darstellung der Beobachtungszeitreihe und modellierten Tageszeitreihe für Winter (DJF) und Sommer (JJA). Zur Suche des ähnlichsten Tages wurden die City Block Distanz und die Euklidische Distanz verwendet.

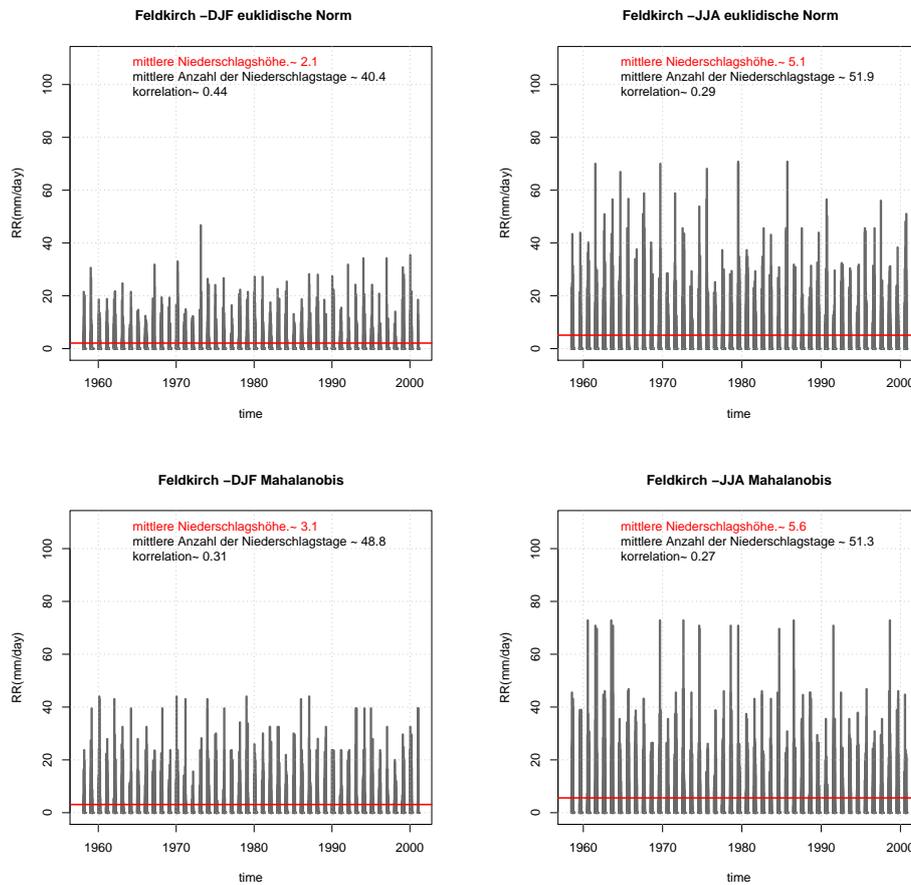


Abbildung 7.5: Feldkirch: Darstellung der modellierten Tageszeitreihe für Winter (DJF) und Sommer (JJA). Zur Suche des ähnlichsten Tages wurden die Euklidische Distanz und die Mahalanobis Distanz verwendet.

bei 40.8, gefolgt von den anderen Distanzmaßen mit ähnlicher Anzahl. An letzter Stelle findet man die Mahalanobis Distanz mit 48.8.

Klagenfurt Bei dieser Station in Süden Österreichs bewegen sich die Korrelationen für den Winter zwischen 0.29 (Mahalanobis) und 0.39. Bessere Werte werden von der Euklidischen gewichteten Distanz geliefert. Genauso wie bei den anderen Stationen können die mittlere Anzahl der Niederschlags-tage und die mittlere Niederschlagshöhe bis auf Werte bei Verwendung die Mahalanobis-Distanz gut wiedergeben werden. Die Korrelationen bleiben sowohl im Winter als auch im Sommer gedämpft.

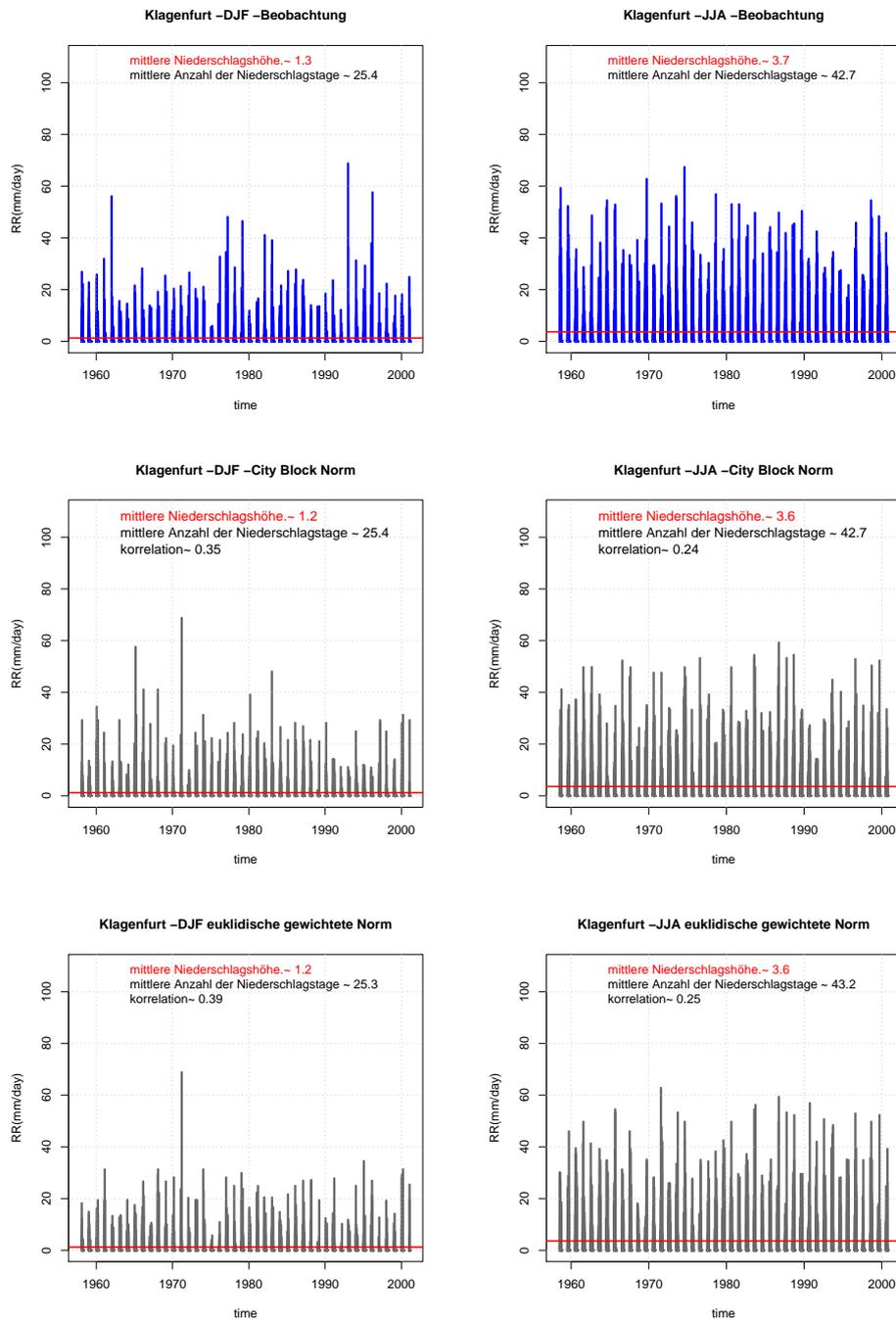


Abbildung 7.6: Klagenfurt: Darstellung der Beobachtungszeitreihe und modellierten Tageszeitreihe für Winter (DJF) und Sommer (JJA). Zur Suche des ähnlichsten Tages wurden die City Block Distanz und die Euklidische Distanz verwendet.

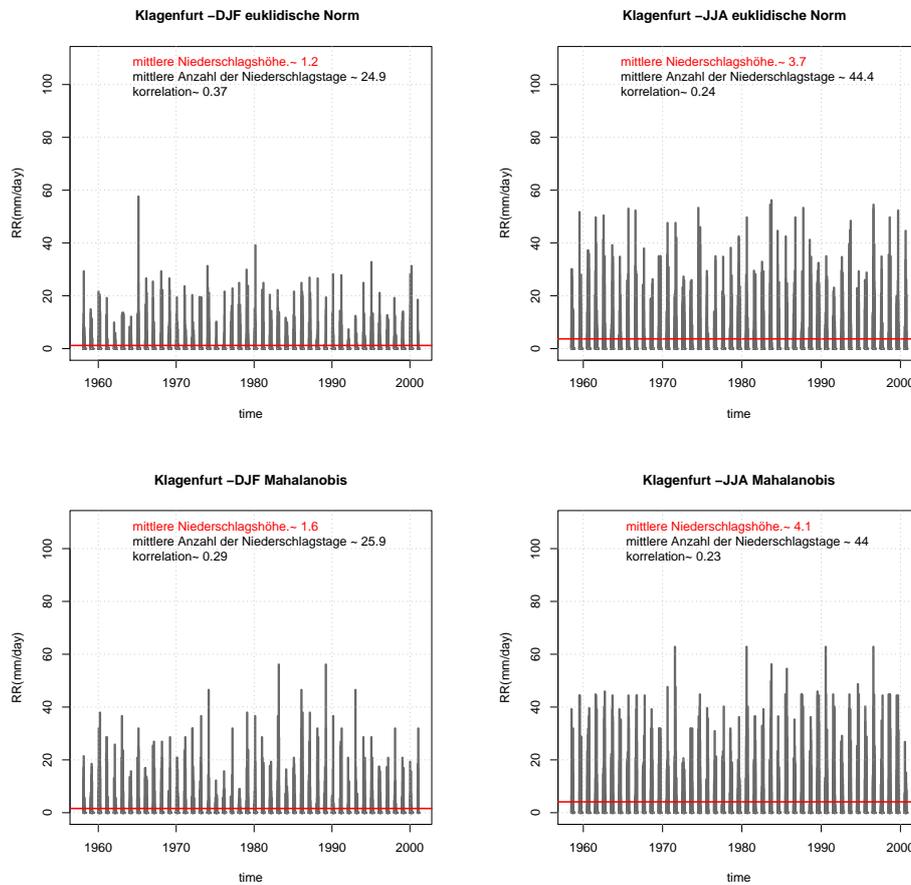


Abbildung 7.7: Klagenfurt: Darstellung der modellierten Tageszeitreihe für Winter (DJF) und Sommer (JJA). Zur Suche des ähnlichsten Tages wurden die Euklidische Distanz und die Mahalanobis Distanz verwendet.

Stationen über 1200m: Feuerkogel (Westen)

Feuerkogel: (Abbildungen 7.8 und 7.9) Diese Bergstation liegt wie Feldkirch im gleichen Korrelationsniveau. Hier zeigt das euklidische gewichtete Distanzmaß den besseren Wert (0.47). Auch hier bleiben die Korrelationen bei allen Distanzmodellen unter 0.50. Im Vergleich zum Sommer sind die Korrelationen bei dieser Station ebenfalls höher. Die mittlere Anzahl der Tage mit Niederschlag mit mindestens 0.1 ist bei allen Modellen annähernd gleich. Dasselbe gilt für den mittleren Niederschlagswert.

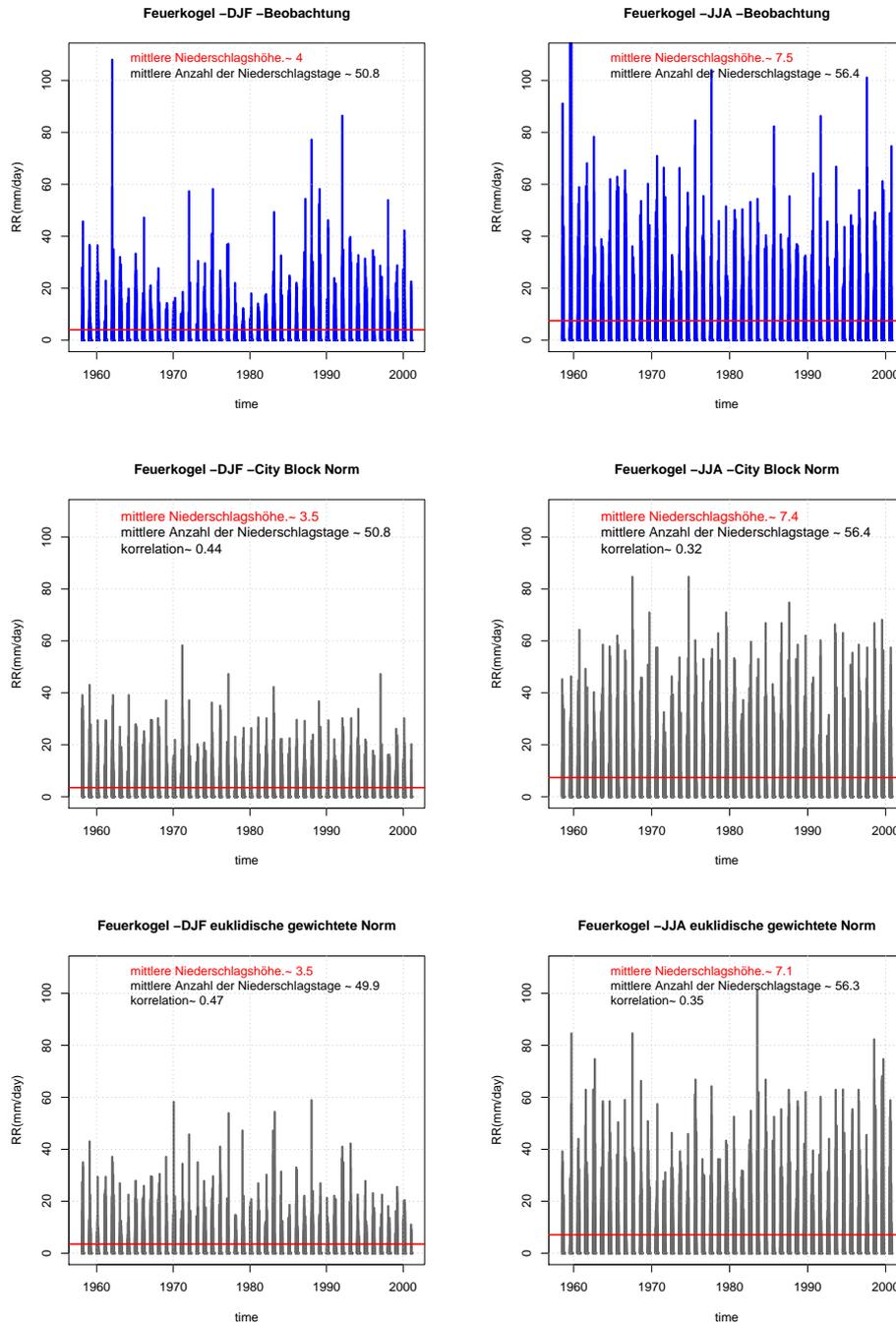


Abbildung 7.8: Feuerkogel: Darstellung der modellierten Tageszeitreihe für Winter (DJF) und Sommer(JJA). Zur Suche des ähnlichsten Tages wurden die City Block Distanz und die Euklidische Distanz verwendet.

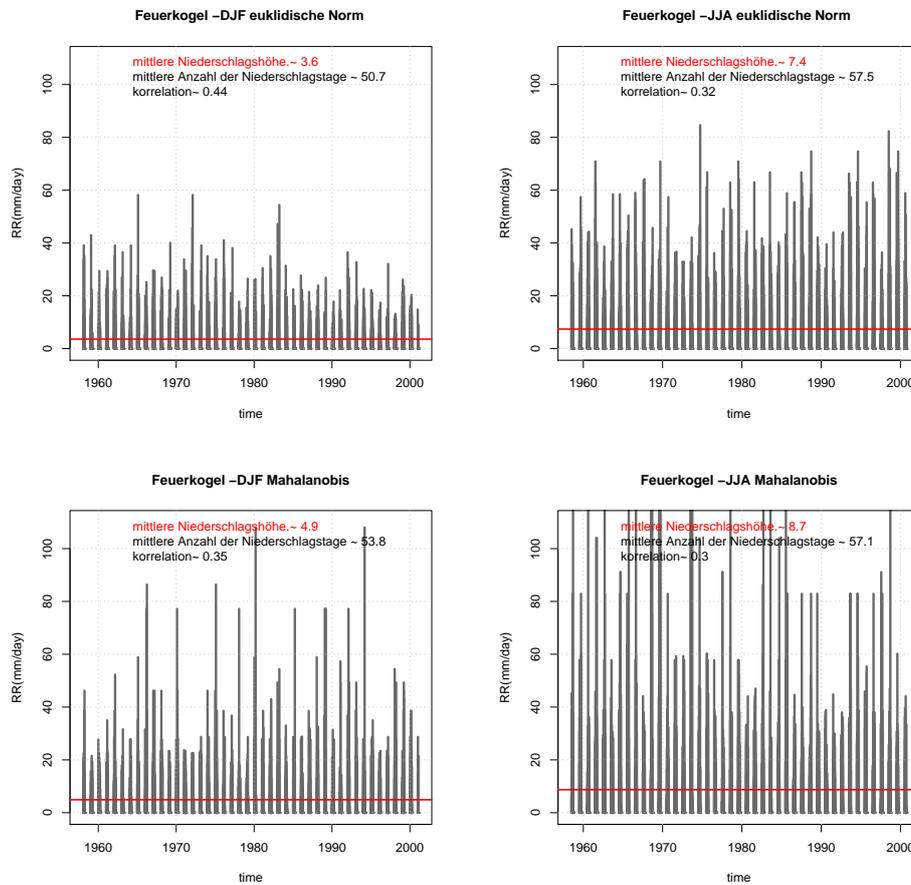


Abbildung 7.9: Feuerkogel: Darstellung der modellierten Tageszeitreihe für Winter (DJF) und Sommer (JJA). Zur Suche des ähnlichsten Tages wurden die gewichtete Euklidische Distanz und die Mahalanobis Distanz verwendet.

7.2 Downscaling auf saisonaler Basis

Beim Downscaling auf saisonaler Basis wurde für jedes „Distanzmodell“ die Niederschlagssumme über jede Jahreszeit, insgesamt 43 Werte für die Periode 1958-2000 (DJF, MAM, JJA und SON) gebildet und für die statistische Auswertung mit den jahreszeitlichen modellierten Niederschlagssummen verglichen. In diesem Abschnitt wird der Einfluss der Distanzmaße auf die *Analog-Methode* auf saisonaler Basis evaluiert.

Um einen Einblick in die erwarteten Ergebnisse zu gewähren wurden auf ähnliche Weise wie im vorigen Abschnitt die gleichen Stationen herangezogen und mit den modellierten Reihen verglichen. Es wurden nochmals die

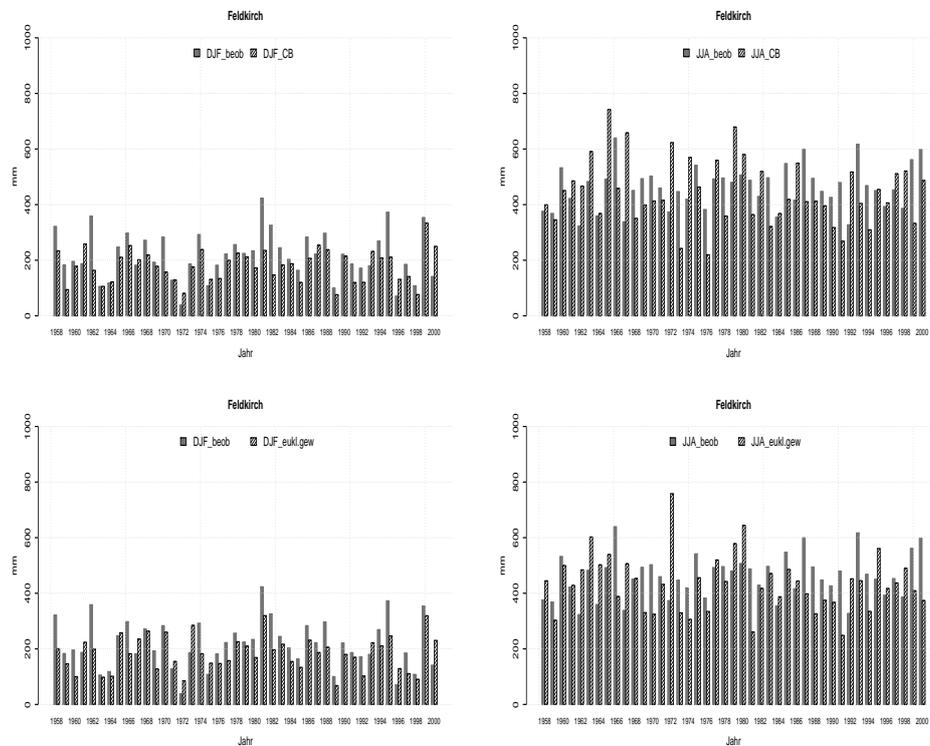


Abbildung 7.10: Feldkirch: Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihe für Winter(DJF) und Sommer(JJA). Für die Suche des analogen Tages wurden die City Block Distanz (CB) und die gewichtete Euklidische Distanz (eukg) verwendet.

Jahreszeiten Winter und Sommer gewählt und die ersten 20 PCs herangezogen. Die zum Vergleich herangezogenen statistischen Maße sind Mittelwert, RMSE, RV und Bias. Diese wurden im Kapitel Verifikation beschrieben und erläutert.

Nichtalpine oder Talstationen (unterhalb von 1200m) Es werden Wien Hohe Warte (Osten), Klagenfurt (Süden) und Feldkirch (Westen) gewählt

Feldkirch:

Winter Die Korrelationen zwischen der saisonalen Zeitreihe und der saisonalen modellierten Reihe ergab 0.71, wenn die euklidische Distanz verwendet

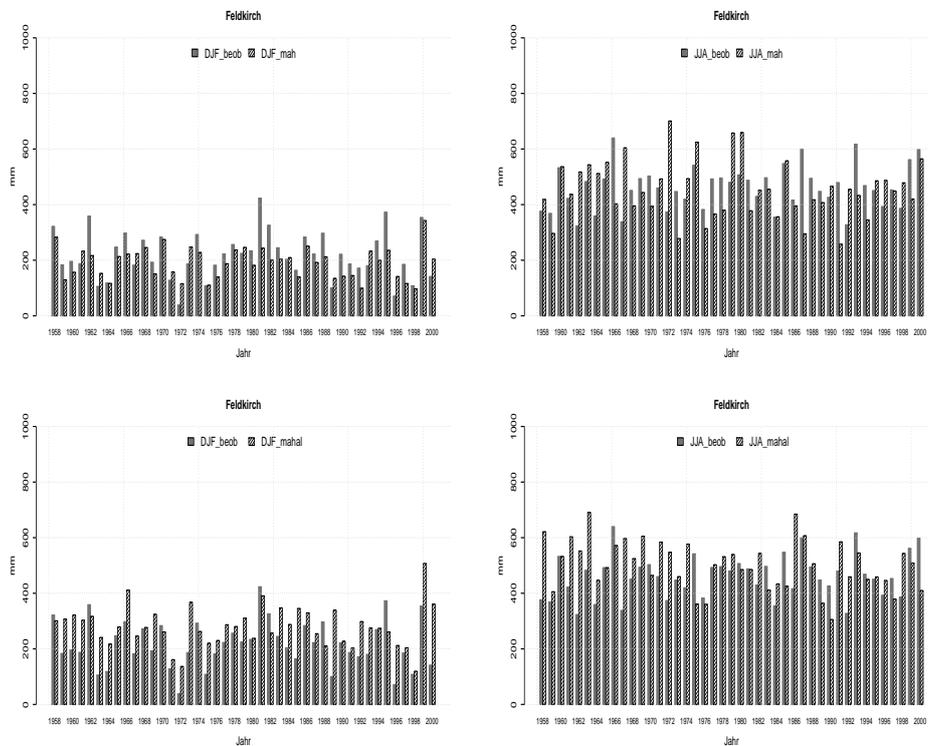


Abbildung 7.11: Feldkirch: Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihen für Winter (DJF) und Sommer (JJA). Für die Suche des analogen Tages wurden die Euklidische Distanz (eukng) und die Mahalanobis Distanz (mahal) verwendet.

wurde, gefolgt von der euklidisch gewichteten mit 0.72, der City-Block-Distanz mit 0.65 und schließlich der Mahalanobis Distanz mit 0.49. Dem ermittelten Mittelwert der saisonalen Zeitreihe kommt die euklidische Distanz näher mit 190.9 mm, gefolgt von der City-Block-Distanz mit 183.3 mm, der gewichteten euklidischen Distanz mit 180.4 mm und schließlich der Mahalanobis-Distanz mit 279.1mm. Der mittlere quadratische Fehler (RMSE) ist am kleinsten bei der euklidischen Distanz (64,12 mm) und am größten bei der Mahalanobis-Distanz (100,74). Die Distanzmodelle bringen bezüglich der Leistung der Vorhersage (RV) für die euklidische Distanz 0.42 gefolgt von der gewichteten Euklidischen (0.34), City-Block Verfahren (0.22) und im negativen Bereich (-0.42) die Mahalanobis Distanz. Beim mittleren Fehler (Bias) ergaben sich negative Werte (-26 mm, -34 mm, -36,9 mm bei der euklidischen, eukli-

disch gewichteten und City-Block-Distanz) was ein Zeichen dafür ist, dass in diesen Fällen die Modelle die beobachteten Werte unterschätzen. Wie zu erwarten ist der Bias bei der Mahalanobis-Distanz positiv, was auf eine Überschätzung der akkumulierten saisonalen Niederschläge der Modelle hindeutet.

Sommer Für Feldkirch wurde für jedes Distanzmaß eine Korrelation von 0.0 berechnet, was darauf hindeutet, dass die *Analog-Methode* im Sommer für diese Station nicht funktioniert. Erwartungsgemäß ist der mittlere quadratische Fehler ungefähr ein Viertel (City Block Distanz: 143 mm, euklidisch gewichtete Distanz 135 mm, euklidische Distanz 130 mm und Mahalanobis-Distanz 121.8 mm) des saisonalen Mittelwerts (Beobachtung: 460.8mm, City-Block-Distanz 444.8 mm, euklidische Distanz 455.2 mm, euklidisch gewichtete Distanz 435.1 mm und Mahalanobis Distanz 502.4 mm). Die Reduktion der Varianz ist bei allen Distanzmodellen negativ, bzw alle RV- Werte sind kleiner als 1.

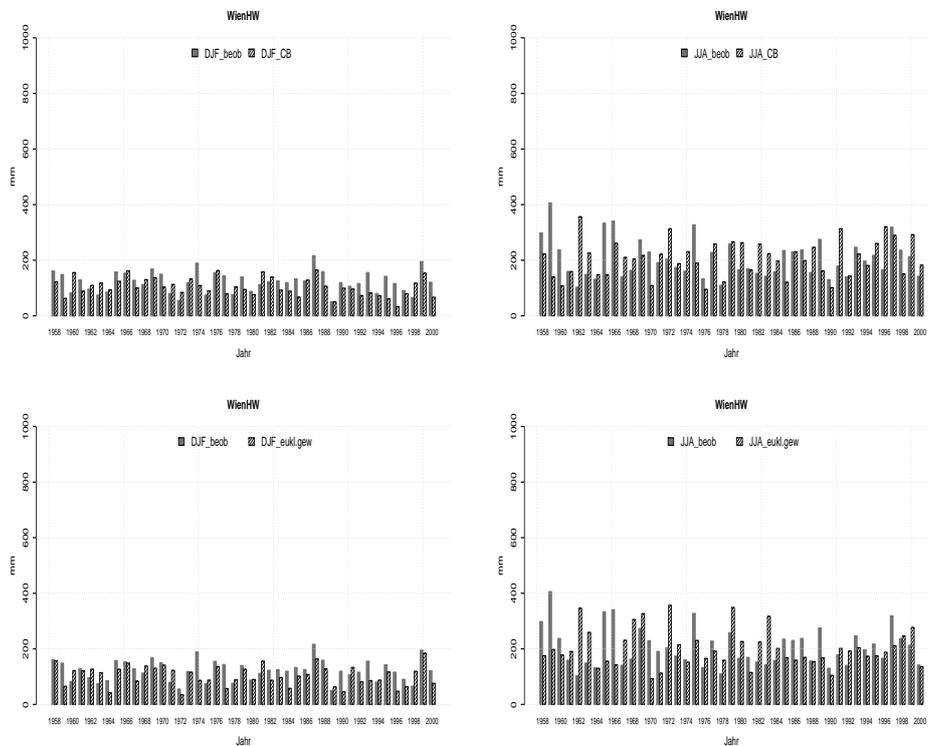


Abbildung 7.12: Wien Hohe Warte (WienHW): Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihe für Winter und Sommer (JJA). Für die Suche des analogen Tages wurden die City Block Distanz (CB) und die gewichtete Euklidische Distanz (eukg) verwendet.

WienHW

Winter Im Gegensatz zu den bisherigen Ergebnissen bezüglich eines linearen Zusammenhangs liefert die Mahalanobis Distanz mit 0.54 für diese Station den besten Wert. Alle anderen Distanzmodelle liefern kleinere Werte (euklidische Distanz: 0.45, euklidische gewichtete: 0.46, City-Block-Distanzmaß: 0.35). Die Wurzel aus dem mittleren quadratischen Fehler (RMSE) ist identisch bei allen Distanzmodellen, also zwischen 44.5mm und 41.5mm. Für den mittleren saisonalen Niederschlagswert ergaben sich 121.7 mm, wobei die anderen Distanzmodelle bis auf die Mahalanobis Distanz bezüglich dieses Wertes im Schnitt mit 15mm wenig dahinter blieben. Alle ermittelten RV-Werte sind negativ, bzw. bewegen sich zwischen 0 und -1 .

Sommer Die ermittelten Korrelationen sind für alle bis auf die Mahalanobis Di-

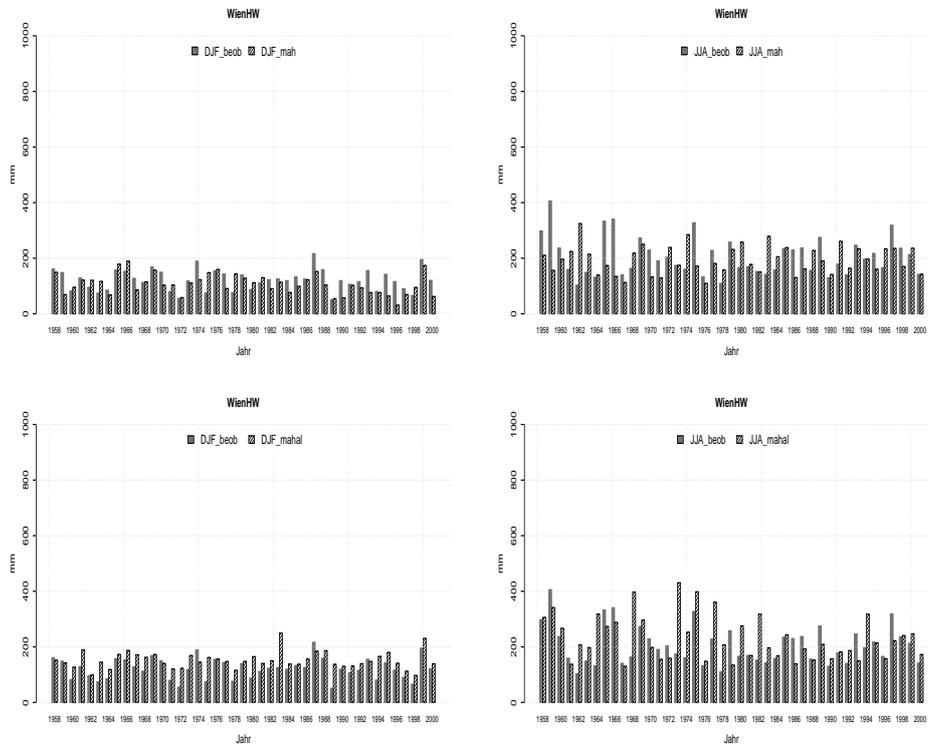


Abbildung 7.13: Wien Hohe Warte (WienHW): Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihe für Winter (DJF) und Sommer(JJA). Für die Suche des analogen Tages wurden die Euklidische Distanz (eukng) und die Mahalanobis Distanz (mahal) verwendet.

stanz gleich 0. Die beste Korrelation wird mit der Mahalanobis-Distanz mit 0.34 erreicht. Betrachtet man den mittleren Fehler, so liegt dieser am höchsten bei der Mahalanobis-Distanz mit 24.8 mm, während die anderen Distanzmodelle im Schnitt 7 mm hinter dem saisonalen Mittelwert (=205.5 mm) liegen. Auch hier liegen die RV-Werte im negativen Bereich, bzw. zwischen -1 und 0 .

Klagenfurt

Winter Die höchste Korrelation wird von der euklidisch gewichteten mit 0.70 geliefert. Die zweithöchste Korrelation von 0.50 wird bei der euklidischen Distanz errungen. Die restlichen Korrelationen sind niedriger. Der RMSE bewegt sich zwischen 60.81 mm (Mahalanobis) und 41.93 mm. Der Bias ist bei allen relativ klein. Die Abweichung ist negativ

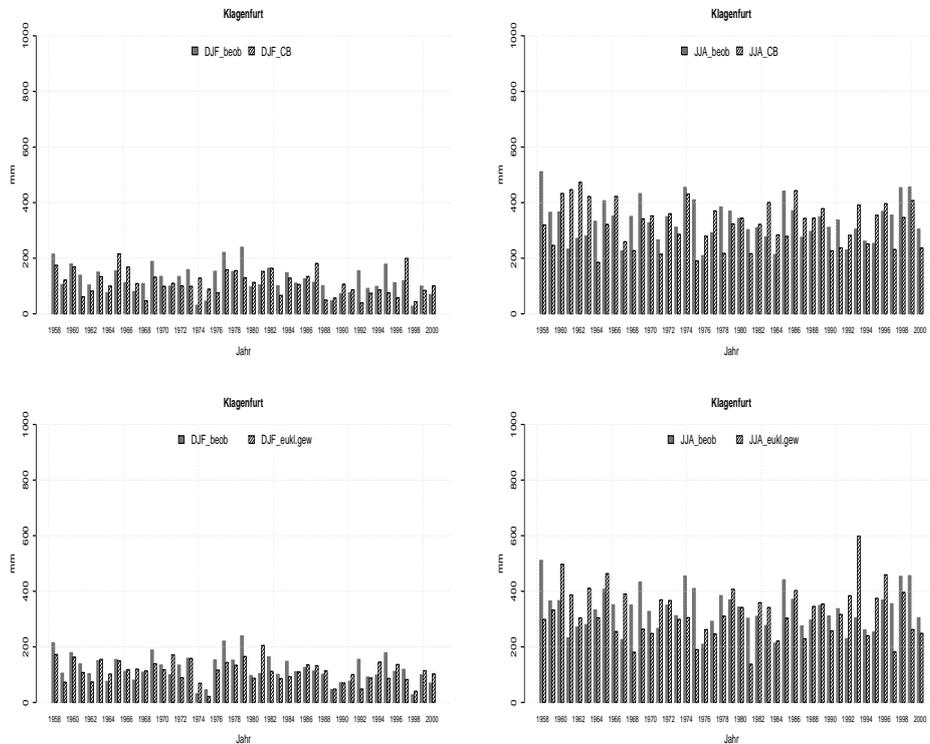


Abbildung 7.14: Klagenfurt: Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihe für Winter (DJF) und Sommer(JJA). Für die Suche des analogen Tages wurden die City-Block-Distanz (CB) und die gewichtete Euklidische Distanz (eukg) verwendet.

($\tilde{10}$ mm). Bei der Reduktion der Varianz sind die Werte bei beiden euklidischen Distanzmaßen mit 0.22 und 0.17 positiv, während bei den restlichen die RV-Werte negativ sind. Der saisonale Mittelwert liegt bei 120.7 mm.

Sommer Die ermittelten Korrelationen für alle Distanzmodelle betragen 0. Damit ist ein linearer Zusammenhang zwischen Beobachtungen und modellierten Werten nicht vorhanden. Der RMSE ist dementsprechend hoch und bewegt sich maximal zwischen 115.41 mm und minimal 102.83 mm. Die RV sind alle negativ. Diese Station bringt noch einmal einen Hinweis, dass die *Analog-Methode* unabhängig davon welches Distanzmaß eingesetzt wird, für die warme Jahreszeit unbrauchbar ist.

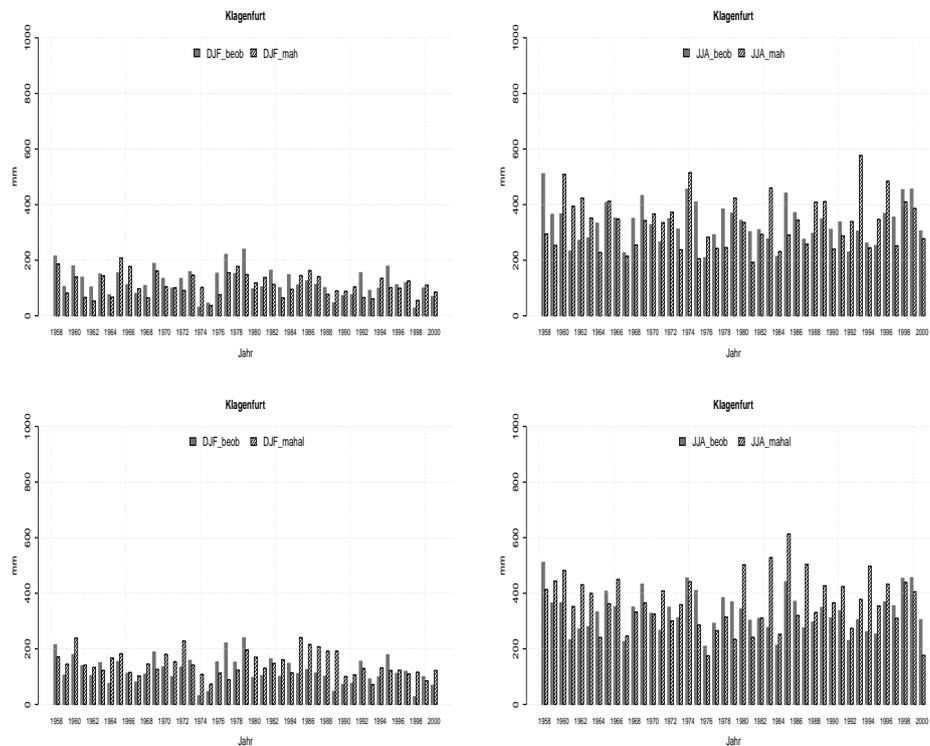


Abbildung 7.15: Klagenfurt: Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihe für Winter (DJF) und Sommer(JJA). Für die Suche des analogen Tages wurden die Euklidische Distanz (eukng) und die Mahalanobis Distanz (mahal) verwendet.

Stationen über 1200m:

Feuerkogel

Winter Für diese Bergstation wird die höchste Korrelation im Winter mit 0.59 bei der City-Block-Distanz erreicht. Die übrigen Korrelationen bewegen sich zwischen 0.49 und 0.40. Die RV-Werte sind positiv für die gewichtete Euklidische Distanz und die City Block Distanz, nämlich 0.21 und 0.014, während bei der euklidischen und der Mahalanobis Distanz negative Werte (-0.04 und -0.45) verzeichnet wurden.

Sommer Der lineare Zusammenhang ist mit der City-Block-Distanz und der gewichteten euklidischen Distanz gleich Null, während bei der Mahalanobis Distanz 0.27 und bei der Euklidischen Distanz 0.11 berechnet wurden. Die RV-Werte aller Distanzmodelle weisen auf eine unrichtige Vor-

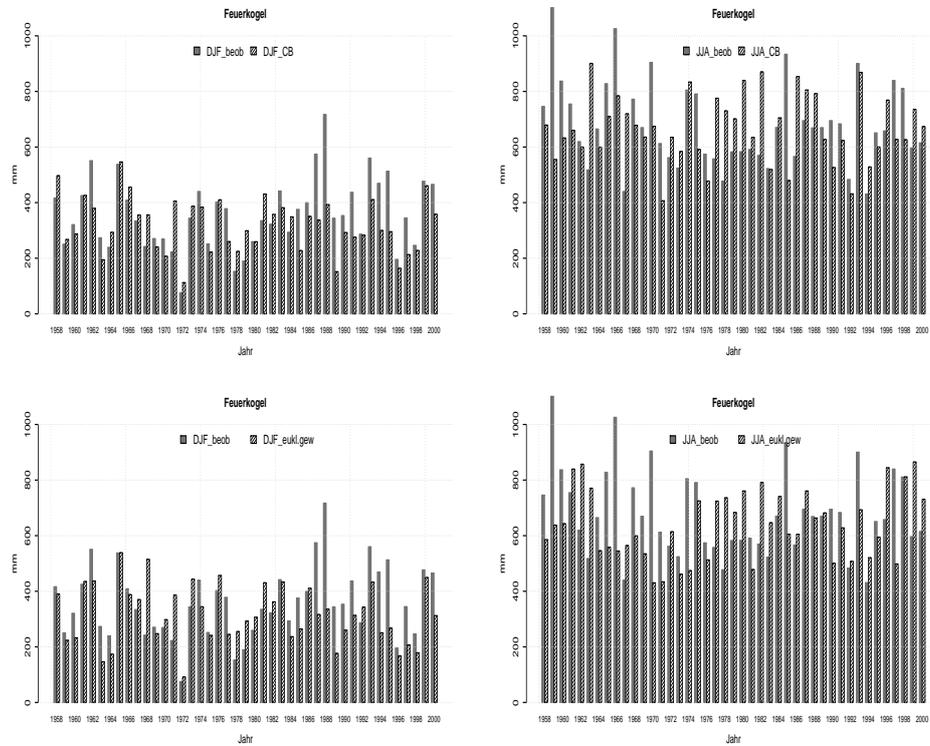


Abbildung 7.16: Feuerkogel: Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihe für Winter (DJF) und Sommer (JJA). Für die Suche des analogen Tages wurden die City-Block-Distanz(CB) und die gewichtete Euklidische Distanz (eukg) verwendet.

hersage hin, da alle Werte negativ sind. Die ermittelten RMSE-Werte sind relativ hoch bzw. bewegen sich zwischen 216 mm (Mahalanobis Distanz) und 186 mm (Euklidische Distanz).

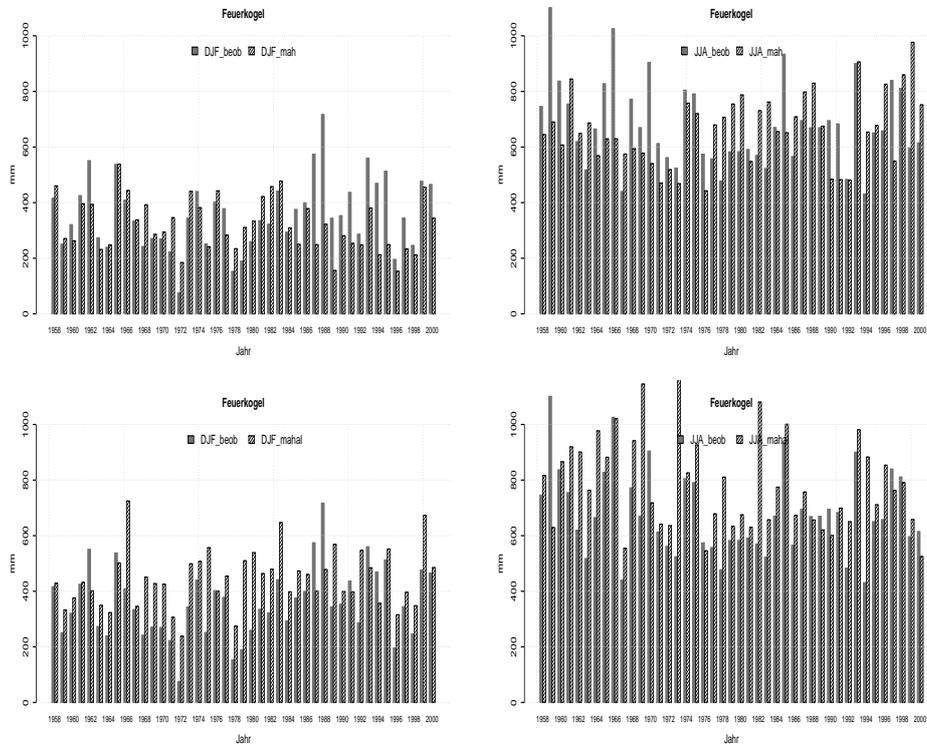


Abbildung 7.17: Feuerkogel: Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihe für Winter (DJF) und Sommer (JJA). Für die Suche des analogen Tages wurden die Euklidische Distanz (eukng) und die Mahalanobis Distanz (mahal) verwendet.

Stationen unter 1200m: Landeck

Winter Die für diese Station errechneten Korrelation beträgt 0.65 für die euklidische gewichtete Distanz, gefolgt von der euklidischen Distanz mit 0.62 und mit dem gleichen Wert von 0.51 für die anderen zwei Distanzmaße. Bis auf die Mahalanobis Distanz sind alle RV-werte positiv und bewegen sich zwischen 0.32 bis 0.09, was darauf hindeutet, dass bis zu einem gewissen Punkt die Methode für eine Vorhersage auf saisonale Basis eignet ist. Der Mittelwert wird vom Distanzmodell relativ gut reproduziert und wird um 25mm unterschätzt. Mit der Mahalanobis Distanz wird der Mittelwert um 50 mm überschätzt.

Sommer Von den bisherigen analysierten Stationen zeigt Landeck für den Sommer, auch wenn sehr gering, einen korrelativen Zusammenhang. Die

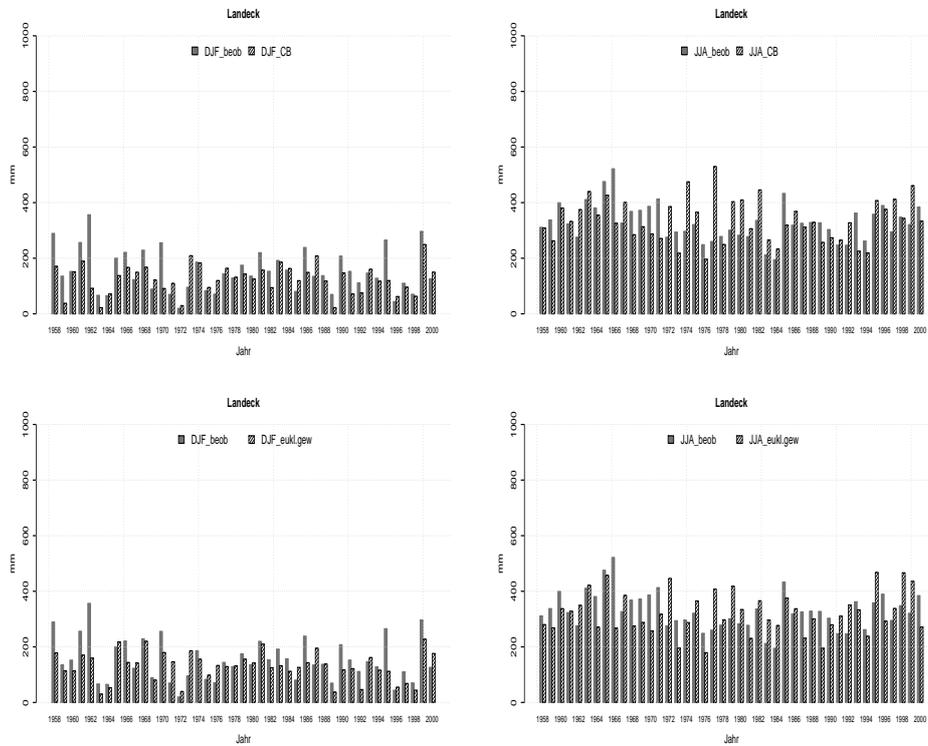


Abbildung 7.18: Landeck: Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihe für Winter (DJF) und Sommer (JJA). Für die Suche des analogen Tages wurden die Euklidische Distanz (eukng) und die Mahalanobis Distanz verwendet (mahal).

höchste Korrelation erbringt die euklidische Distanz mit 0.28, während die anderen Distanzmaße knapp unter 0.20 liegen. Der Mittelwert wird aber auch hier befriedigend wiedergegeben. Der RMSE ist erwartungsgemäß hoch und liegt bei 90 mm bei allem Distanzmaßen. Wie erwartet und passend zu den anderen statistischen Maßen für diese Station liegen die RV-Werte im negativen Bereich.

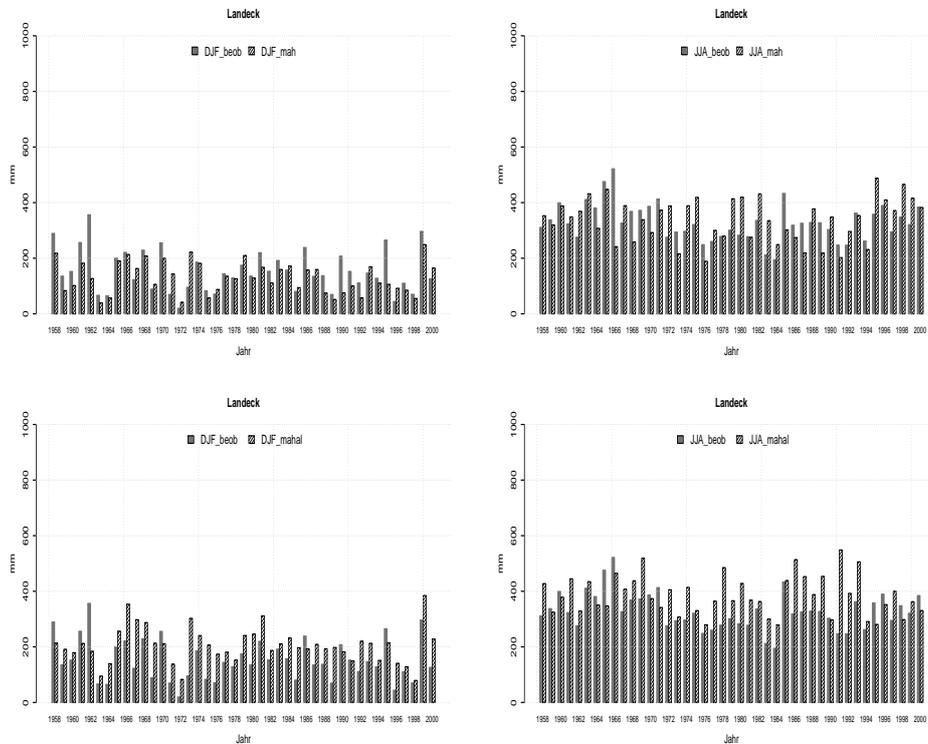


Abbildung 7.19: Landeck: Vergleich der Beobachtungs- und modellierten saisonalen Zeitreihe für Winter (DJF) und Sommer (JJA). Für die Suche des analogen Tages wurden die Euklidische Distanzmaß (eukng) und die Mahalanobis Distanz verwendet (mahal).

Darstellung der saisonalen Korrelationen bezüglich der geographischen Lage der StartClim Stationen

Die folgenden Abbildungen stellen die Korrelationen des Downscalings auf saisonaler Basis aller 70 Stationen für den Parameter Niederschlag bei Verwendung der City-Block-Distanz, Euklidische Distanz, gewichtete Euklidische Distanz und Mahalanobis Distanz dar. Dadurch kann eine mögliche Information über den linearen Zusammenhang zwischen Beobachtung und Modell bezüglich der von Ehrendorfer festgestellten Regionen erhalten werden. Die Interpretation der Ergebnisse wird auf die erwähnten definierten Regionen bezogen. Zur Farbskala: Die Farben Dunkelblau bis Weissblau stellen Korrelationen von 0.09 bis 0.45 und weissorange bis dunkelrot Korrelationen für den Bereich 0.45 bis über 0.85 dar.

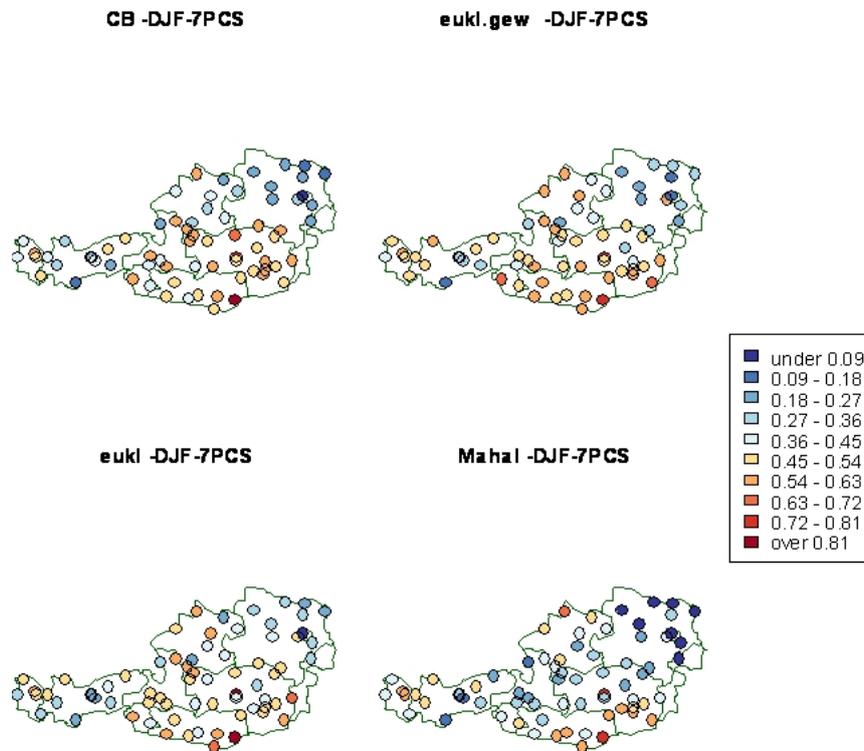


Abbildung 7.20: DJF-7PCS: Korrelationen zwischen saisonaler und modellierter Zeitreihe bei Anwendung der City-Block-Distanz, der gewichteten Euklidischen Distanz, der Euklidischen Distanz und der Mahalanobis Distanz.

Dezember-Januar(DJF)

DJF-7PCS: Beim Vergleich der vier Darstellungen erkennt man, dass höhere Korrelationen für die City-Block-Distanz und die gewichtete Euklidische Distanz gefunden werden. Die meisten Stationen zeigen eine Korrelation im Bereich von 0.54 bis 0.63, wobei die gewichtete euklidische im Schnitt etwas bessere Ergebnisse liefert. Eine gemeinsame Eigenschaft aller 4 Darstellungen sind die sehr niedrigen Korrelationen im Osten (Wien und Niederösterreich), die unter 0.27 liegen. Die höchsten Korrelationen (ueber 0.63) werden in Südkärnten mit Ausnahme vom Mahalanobis Distanzmaß gefunden. Die Station St. Michael bei Bleiburg zeigt in allen Darstellungen eine Korrelation von 0.81.

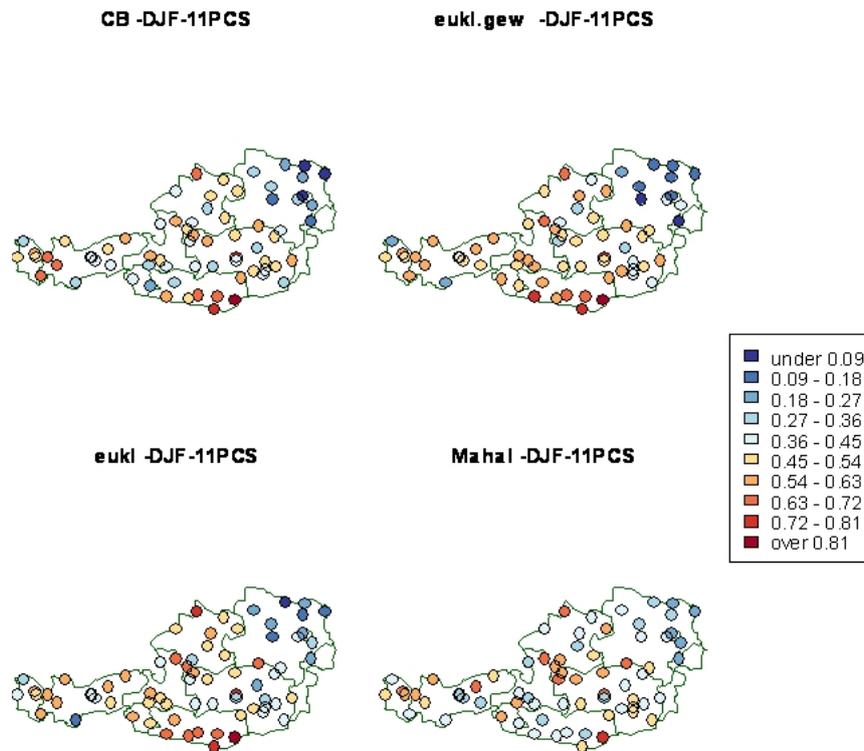


Abbildung 7.21: DJF-11PCs: Korrelationen zwischen saisonaler und modellierte Zeitreihe bei Anwendung der City-Block-Distanz, gewichtete Euklidische Distanz, Euklidische Distanz und der Mahalanobis Distanz.

DJF-11PCS: Bei genauer Betrachtung zeigt das Modell mit der gewichteten Euklidischen Distanz eine höhere Anzahl von Stationen mit guten Korrelationen in Mittel- und Westösterreich. Bei den anderen Distanzmaßen zeigen manche Stationen geringe Verbesserungen, wobei andere schlechtere Korrelationswerte aufweisen. Signifikant ist der Unterschied zwischen der Mahalanobis Distanz bei Anwendung von 7 PCs (vorige Abbildung) und 11 PCs.

DJF-20PCS: Für die Wintermonate werden im Vergleich zu den anderen Jahreszeiten die besten Ergebnisse erreicht wenn 20 Hauptkomponenten herangezogen werden. Für manche Station liegt die Korrelation bei über 0.81. Solche guten Korrelationen werden sowohl bei der gewichtete Euklidischen Distanz als auch bei der Euklidischen Distanz erlangt. Die Häufigkeit der besseren Korrelationen werden vor allem im Westen und Süden (gewichte-

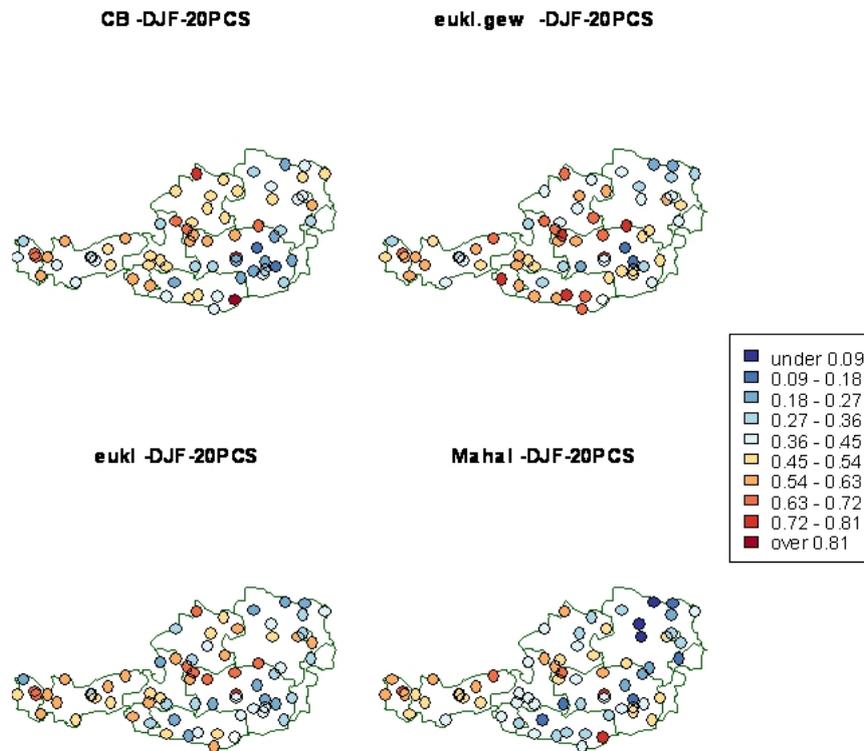


Abbildung 7.22: DJF-20PCS: Korrelationen zwischen saisonaler und modellierte Zeitreihe bei Anwendung der City-Block, euklidische gewichtete, euklidische und der Mahalanobis Distanz.

te Euklidische Distanz) erkennbar. Die niedrigsten Werte findet man über dem Südburgenland und der Steiermark bei dem City-Block-Distanzmodell. Bei der euklidischen und Mahalanobis Distanz sind die Korrelationen über Kärnten, Steiermark und Burgenland am niedrigsten. Die Gewichtung der euklidischen Distanz im Vergleich zu den anderen eingesetzten Distanzmaßen bringt deutlich bessere Ergebnisse.

März-April(MAM)

(siehe Anhang B)

MAM-7PCS, MAM-11PCS, MAM-20PCS: Für den Frühling (MAM) sind die Korrelationen bei allen drei Anzahlen von Hauptkomponenten kaum vorhanden, bei minimal höheren Korrelationswerten im Fall von 20 Hauptkomponenten. Die Korrelationen sind nicht höher als 0.45, mit Ausnahme der

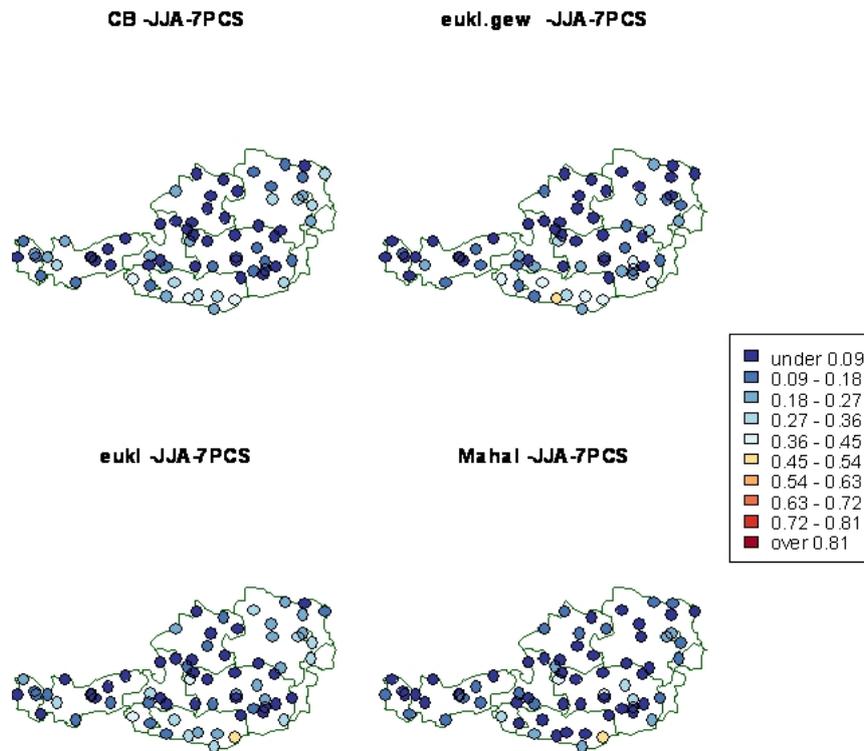


Abbildung 7.23: JJA-7PCS: Korrelationen zwischen saisonaler und modellierter Zeitreihe bei Anwendung der City-Block-Distanz, gewichtete Euklidische Distanz, Euklidische Distanz und der Mahalanobis Distanz.

Station St. Michael bei Bleiburg, die bei allen Distanzmodellen höhere Korrelationen aufweist. Erkennbar sind größere Korrelationen einzelner Stationen – die meisten korrelieren mit Werten zwischen 0.45 und 0.54, bei 20 PCs, wobei sich die Stationen je nach Distanzmodell unterscheiden.

Juni-August (JJA)

JJA-7PCS, JJA-11PCS, JJA-20PCS (Abbildungen B.3, 7.24, 7.25):

Der erste Blick auf die Darstellung der Korrelationen für die Sommermonate lässt den Schluss zu, dass der lineare Zusammenhang in allen Regionen Österreichs von sehr gering bis nicht vorhanden ist. Vergleicht man die 3 Abbildungen bezüglich des Einbeziehen der Hauptkomponenten (PCs) so wird ersichtlich, dass die Anzahl der PCs kaum Einfluss auf die Performance der *Analog-Methode* hat. Abgesehen von kleinen Unterschieden, die in diesem Kontext nicht relevant sind, kann festgestellt werden, dass das Einsetzen an-

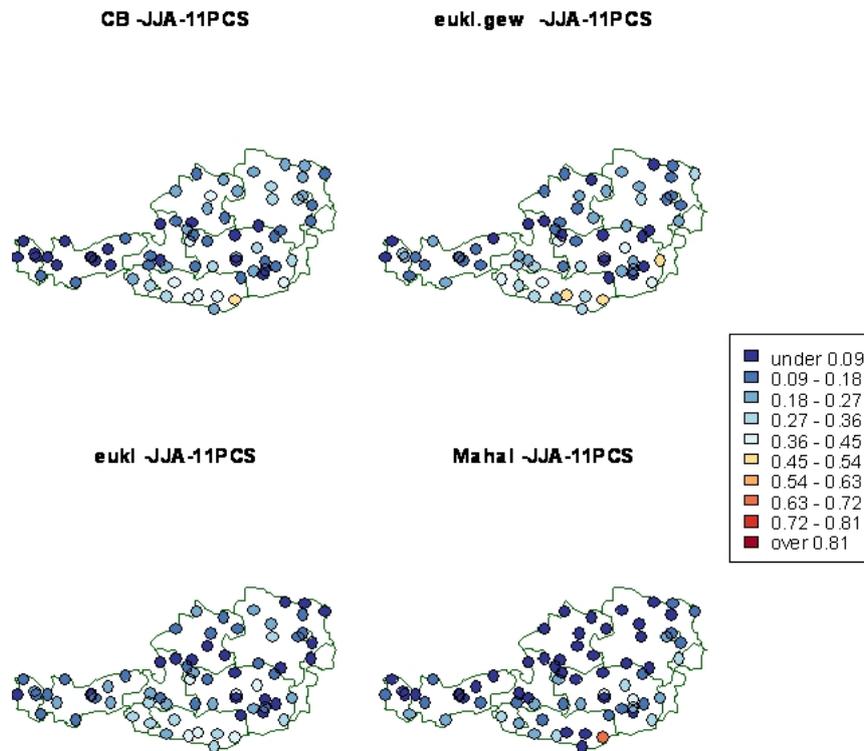


Abbildung 7.24: JJA-11PCS: Korrelationen zwischen saisonaler und modellierte Zeitreihe bei Anwendung der City-Block, euklidische gewichtete, euklidische und der Mahalanobis Distanz.

derer Distanzmaße keine Verbesserungen bringt. Die höchsten Korrelationen werden mit 0.45 im Süden Österreichs erreicht.

September-November(SON)

(siehe Anhang B)

SON-7PCS Als gemeinsame Eigenschaft haben die 4 Distanzmaße im Süden (Kärnten) ähnliche Korrelationen zwischen 0.54 und 0.72. Erkennbar ist, dass die gewichtete Euklidische Distanz im Schnitt mehr Stationen hat mit höheren Korrelationen aufweist. Bei der Mahalanobis Distanz findet man im Westen (Tirol, Vorarlberg) keine Korrelation (>0.09).

SON-11PCS Die höchsten Korrelationswerte, welche über 0.63 liegen,

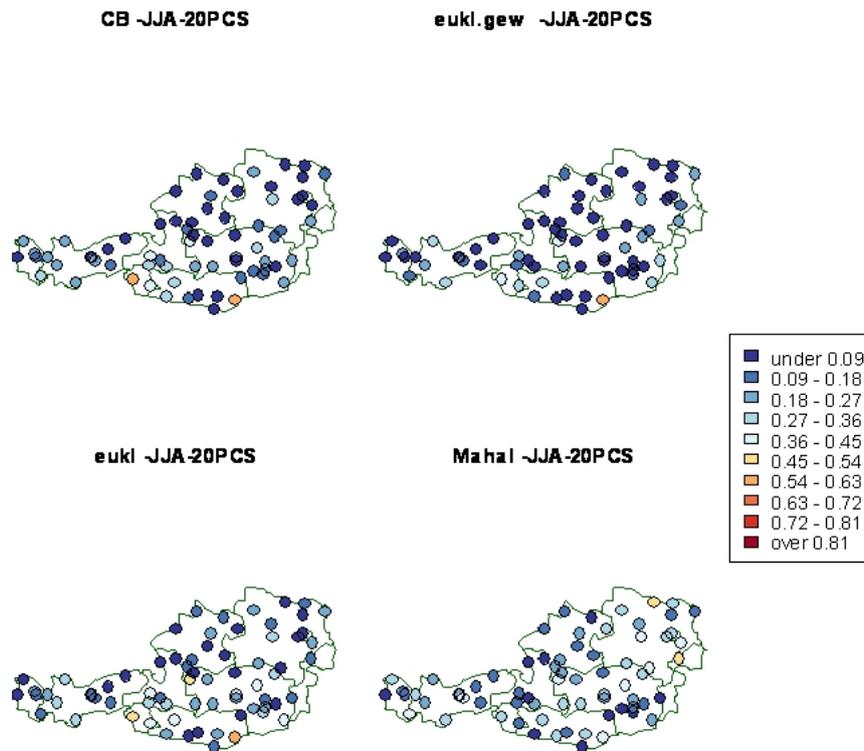


Abbildung 7.25: JJA-20PCS: Korrelationen zwischen saisonaler und modellierte Zeitreihe bei Anwendung der City-Block-Distanz, gewichtete Euklidische Distanz, Euklidische Distanz und der Mahalanobis Distanz.

findet man mit der gewichtete Euklidischen Distanz im Süden (Steiermark und Kärnten) und im Westen (Grenze Oberösterreich zu Kärnten). Mit der City Block Distanz werden die höchsten Korrelationswerte (über 0.72) im Süden (Kärnten) erlangt. Die anderen Distanzmodelle weisen auch tendentiell die besseren Werte im Süden auf, diese sind jedoch nicht höher als 0.54.

SON-20PCS: Durch Vergleich mit den vorigen Abbildungen (SON) erkennt man, dass das Einbeziehen von 20 Komponenten für alle Distanzmodelle schwächere Korrelationen bringt, wobei die City-Block Distanz besser abschneidet. Die besseren Korrelationen findet man wieder im Süden (Kärnten). Diese sind nicht höher als 0.72. Im Osten weisen alle Distanzmodelle geringe Korrelation, nicht höher als 0.36, auf.

Kapitel 8

Diskussion und Ausblick

8.1 Diskussion der Ergebnisse

Um über die Leistung der Methode Aussagen treffen zu können, wurden die im Kapitel 6.1 besprochenen statistischen Größen berechnet und in Tabellen (im Anhang) in Form einer Zusammenfassung (Minimum, Mittelwert und Maximum) zusammengefasst. Dadurch kann u.a. evaluiert werden, wie sich die Distanzmaße auf die Methode auswirken. Grundsätzlich wird folgendes festgestellt:

- Ein erster Vergleich zwischen den Ergebnissen für das Downscaling auf Tages- und saisonaler Basis lässt erkennen, dass die Methode für die saisonalen Zeitreihen besser funktioniert. Das wird aufgrund der für Niederschlag relativ hohen Korrelationen erkannt, bzw. bei den saisonalen Zeitreihen ist die Wurzel des mittleren quadratischen Fehlers (RMSE) im Schnitt weit kleiner.
- Wenn die Jahreszeiten untereinander verglichen werden, wird klar, dass die Methode im allgemeinen sowohl für das Downscaling auf Tages- wie auch auf saisonaler Basis für die Wintermonate (DJF) eine bessere Leistung liefert. Am schlechtesten funktioniert die Methode für die Sommermonate. Teilweise funktioniert sie auch im Herbst (SON) für die saisonalen Zeitreihen.
- Die Gewichtung der Euklidischen Distanz mit dem Eigenwert liefert im Mittel ohne Zweifel die besten Ergebnisse, wobei bedacht werden soll, dass hier nur 69 Stationen zum Testen benutzt worden sind und für jede einzelne Station in manchen Fällen mit einem anderen Distanzmaß im statistischen Kontext bessere Werte erlangt werden. Vergleicht man die Ergebnisse der Distanzmaße untereinander, kommt man zu

dem Schluss (mit Ausnahme der Mahalanobis), dass die City-Block-Distanzmaß oder die Euklidische Distanz im Mittel von der gewichteten Euklidischen Distanz nicht sehr entfernt sind.

- Was die Leistung der Methode bezüglich geographischer Lage der Stationen betrifft konnte beobachtet werden, dass die Methode im Mittel für alle Distanzmaße besser für den Westen geeignet ist als für den östlichen Teil Österreichs.
- Die *Analog-Methode* gehört zu den sogenannten Klassifikationsmethoden, von denen nicht erwartet wird, dass sie die lokalen Variablen zeitlich exakt wiedergeben wird aber sehr wohl deren Mittelwert und Varianz. Dies gilt auch aufgrund der errechnete Werte in den Tabellen im Appendix.
- In dieser Arbeit wurden zusätzlich Verifikationsmaße für Kategorische Größen berechnet um theoretisch den Erfolg der *Analog-Methode*, sobald sie für Prognose-Zwecke angewandt wird, zu testen. Dadurch kann der errechneten Maße eine (theoretische) Aussage getroffen werden. Die Ergebnisse liegen nur für den Winter (DJF) und für den Sommer(JJA) vor, wobei ein Vergleich gemacht werden kann zwischen den Fällen mit 7 PCs und mit 20 PCs. Auch wenn nur sehr kleine Unterschiede vorhanden sind, können diese analysiert werden. Bezüglich des Erfolgs einer echten Vorhersage kann beobachtet werden, dass der POD im Mittel im Sommer höher ist als im Winter. Beim FARA werden die analogen Tage, an denen Niederschlag gefallen ist mit dem entsprechen „Partner“ verglichen, an dem kein Niederschlag gefallen ist. In diesem Fall liegt der Winter etwas schlechter, wobei die Unterschiede minimal sind. Für den Erfolg oder Mißerfolg einer Vorhersage (POFDa), wobei der Vergleich immer zwischen einem bestimmten Tag und dem ähnlichsten Tag gemacht wird, werden die niedrigsten Werte im Schnitt in den Wintermonaten beobachtet. Beim Bias-Score können keine Unterschiede festgestellt werden. Beim HSS und TSS bewegen sich die Werte um den Wert 0, was gleichbedeutend damit ist, dass eine Vorhersage gleich der Referenzprognose ist.

8.2 Ausblick

In dieser Arbeit konnte festgestellt werden, dass die *Analog-Methode* auf saisonaler Basis zufriedenstellend funktioniert, und dass eine Gewichtung mittels euklidischen Abstandes sich als sinnvoll erweist. Auf diese Weise eignet sich die *Analog-Methode* gut für die Modellierung statistischer Eigenschaften meteorologischer Parameter, vor allem auch dann, wenn diese nicht normalverteilte Größen sind.

Um zusätzlich zu verifizieren, wie die Paare Tag-Ähnlichster Tag zusammenhängen, könnte anhand einer Wetterlagenklassifikation überprüft werden ob es sich um ähnliche Wetterlagen handelt. Auf diese Weise könnte man eine weitere Evaluierung erreichen.

Grundsätzlich können weitere Prädiktorenfelder herangezogen werden, um zu überprüfen, ob weitere meteorologische Parameter als Prädiktoren besser geeignet für die Modellierung des Niederschlages mit der *Analog-Methode* sind.

Anhang A

Tabellen

Die folgende Tabellen stellen die numerischen Ergebnisse der in Kapitel 6.1 gewählten statistischen Maße dar. Die hervorgehobenen Zahlen (in fetter Schrift) sollen auf die besten Ergebnisse aufmerksam machen.

sig.kor	CB	eukng	eukg	mahal
DJF7no	54	55	50	31
prozent	77	79	71	44
DJF11no	53.0	60.0	60.0	54.0
prozent	76	86	86	77
DJF20no	53.0	59.0	60.0	47.0
prozent	76	84	86	67

Tabelle A.1: Anzahl und Prozentsatz an Stationen mit signifikanter Korrelation (DJF) bei einem Signifikanzniveau von 5 %

sig.kor	CB	eukng	eukg	mahal
MAM7no	15	12	13	7
prozent	21	17	19	10
MAM11no	14.0	12.0	13.0	28.0
prozent	20	17	19	40
MAM20no	20.0	26.0	12.0	8.0
prozent	29	37	17	11

Tabelle A.2: Anzahl und Prozentsatz an Stationen mit signifikanter Korrelation (MAM) bei einem Signifikanzniveau von 5 %

sig.kor	CB	eukng	eukg	mahal
JJA7no	8	5	6	2
prozent	11	7	9	3
JJA11no	8.0	6.0	5.0	2.0
prozent	11	9	7	3
JJA20no	7.0	2.0	10.0	16.0
prozent	10	3	14	23

Tabelle A.3: Anzahl und Prozentsatz an Stationen mit signifikanter Korrelation (JJA) bei einem Signifikanzniveau von 5 %

sig.kor	CB	eukng	eukg	mahal
SON7no	56	55	37	43
prozent	80	79	53	61
SON11no	44.0	57.0	47.0	30.0
prozent	63	81	67	43
SON20no	52.0	49.0	53.0	31.0
prozent	74	70	76	44

Tabelle A.4: Anzahl und Prozentsatz an Stationen mit signifikanter Korrelation (SON) bei einem Signifikanzniveau von 5 %

RV	CB	eukg	eukng	mahal
DJF7	-0.96	-0.88	-4.65	-0.84
	-0.57	-0.61	-1.21	-0.60
	0.00	0.00	0.00	0.00
DJF11	-0.83	-0.82	-2.02	-0.70
	-0.45	-0.47	-1.08	-0.40
	0.00	0.00	0.00	0.00
DJF20	-0.81	-0.80	-1.62	-0.81
	-0.37	-0.34	-0.89	-0.36
	0.00	0.00	0.00	0.00
MAM7	-0.99	-0.96	-2.41	-1.07
	-0.77	-0.74	-1.06	-0.78
	0.00	0.00	0.00	0.00
MAM11	-1.03	-0.96	-3.87	-1.00
	-0.67	-0.65	-1.45	-0.68
	0.00	0.00	0.00	0.00
MAM20	-0.94	-0.80	-2.34	-0.78
	-0.53	-0.50	-1.18	-0.53
	0.00	0.00	0.00	0.00
JJA7	-0.88	-0.93	-5.13	-0.88
	-0.68	-0.67	-1.49	-0.65
	0.00	0.00	0.00	0.00
JJA11	-0.92	-0.96	-2.83	-0.97
	-0.64	-0.62	-0.94	-0.65
	0.00	0.00	0.00	0.00
JJA20	-0.98	-0.96	-1.57	-0.92
	-0.62	-0.59	-1.00	-0.63
	0.00	0.00	0.00	0.00
SON7	-0.95	-0.98	-3.51	-0.93
	-0.64	-0.60	-1.02	-0.65
	0.00	0.00	0.00	0.00
SON11	-0.92	-0.96	-2.83	-0.97
	-0.64	-0.62	-0.94	-0.65
	0.00	0.00	0.00	0.00
SON20	-0.64	-0.69	-1.88	-0.70
	-0.44	-0.42	-0.97	-0.46
	0.00	0.00	0.00	0.00

Tabelle A.5: Zusammenfassung (Minimum, Mittelwert, Maximum) der Reduktion der Varianz (RV) für die Tagesreihen in den Jahreszeiten DJF, MAM, JJA, SON bei 7, 11 und 20 PCs für die 4 verschiedenen Distanzmaße: City-Block-Distanz (CB), gewichtete Euklidische Distanz (eukg), Euklidische Distanz (eukng) und Mahalanobis Distanz (mahal).

RV	CB	eukg	eukng	mahal
DJF7	-0.53	-0.72	-0.62	-12.64
	-0.03	-0.04	-0.06	-2.30
	0.39	0.44	0.39	0.15
DJF11	-0.91	-0.66	-0.68	-3.36
	-0.02	0.07	0.03	-1.30
	0.48	0.43	0.49	0.41
DJF20	-0.96	-0.77	-0.69	-1.59
	-0.02	0.05	0.06	-0.41
	0.56	0.59	0.62	0.45
MAM7	-0.80	-1.38	-1.13	-8.55
	-0.29	-0.37	-0.38	-2.22
	0.10	0.28	0.21	0.00
MAM11	-1.20	-1.03	-1.11	-14.07
	-0.38	-0.38	-0.40	-3.95
	0.31	0.31	0.19	0.00
MAM20	-1.95	-1.64	-1.68	-17.12
	-0.80	-0.71	-0.80	-5.19
	0.05	0.00	0.00	0.00
JJA7	-1.68	-1.50	-1.47	-5.28
	-0.71	-0.61	-0.64	-1.74
	0.00	0.08	0.00	0.00
JJA11	-2.43	-2.07	-1.86	-3.38
	-0.81	-0.93	-0.79	-1.09
	0.21	0.00	0.13	0.04
JJA20	-0.80	-1.07	-1.25	-8.05
	-0.26	-0.29	-0.35	-2.59
	0.23	0.30	0.18	0.00
SON7	-1.24	-1.04	-1.29	-6.30
	-0.20	-0.12	-0.27	-1.50
	0.46	0.46	0.39	0.27
SON11	-1.05	-1.10	-1.31	-6.38
	-0.27	-0.10	-0.23	-1.29
	0.44	0.50	0.29	0.09
SON20	-1.44	-1.97	-1.63	-6.99
	-0.29	-0.36	-0.28	-1.84
	0.32	0.31	0.39	0.00

Tabelle A.6: Zusammenfassung (Minimum, Mittelwert, Maximum) der Reduktion der Varianz (RV) für die saisonalen Zeitreihen in den Jahreszeiten DJF, MAM, JJA, SON bei 7, 11 und 20 PCs für die 4 verschiedenen Distanzmaße: City-Block-Distanz (CB), gewichtete Euklidische Distanz (eukg), Euklidische Distanz (eukng) und Mahalanobis Distanz (mahal).

ACC	CiBl	eukg	eukng	mahal
DJF7	0.04	0.05	0.05	0.04
	0.17	0.17	0.16	0.16
	0.36	0.34	0.36	0.33
DJF11	0.06	0.08	0.10	0.08
	0.21	0.23	0.24	0.20
	0.35	0.37	0.39	0.32
DJF20	0.11	0.13	0.12	0.08
	0.26	0.27	0.26	0.19
	0.41	0.43	0.42	0.32
MAM7	0.06	0.06	0.06	0.08
	0.11	0.12	0.11	0.13
	0.23	0.22	0.22	0.25
MAM11	0.06	0.07	0.07	0.07
	0.15	0.16	0.15	0.16
	0.27	0.26	0.25	0.24
MAM20	0.08	0.06	0.06	0.07
	0.18	0.18	0.17	0.14
	0.32	0.31	0.29	0.25
JJA7	0.06	0.06	0.07	0.10
	0.14	0.14	0.14	0.17
	0.21	0.19	0.20	0.24
JJA11	0.06	0.08	0.08	0.06
	0.15	0.17	0.16	0.15
	0.25	0.25	0.25	0.26
JJA20	0.06	0.09	0.08	0.06
	0.15	0.17	0.16	0.14
	0.25	0.26	0.26	0.22
SON7	0.08	0.06	0.06	0.05
	0.17	0.18	0.16	0.17
	0.32	0.35	0.33	0.34
SON11	0.06	0.08	0.08	0.06
	0.15	0.17	0.16	0.15
	0.25	0.25	0.25	0.26
SON20	0.10	0.12	0.11	0.08
	0.20	0.23	0.21	0.18
	0.31	0.38	0.34	0.28

Tabelle A.7: Zusammenfassung (Minimum, Mittelwert, Maximum) des Anomaliekorrelationskoeffizienten (ACC) für die Tageszeitreihe in den Jahreszeiten DJF, MAM, JJA, SON bei 7, 11 und 20 PCs für die 4 verschiedenen Distanzmaße: City-Block-Distanz (CB), gewichtete Euklidische Distanz (eukg), Euklidische Distanz (eukng) und Mahalanobis Distanz (mahal).

ACC	CB	eukg	eukng	mahal
DJF7	0.13	-0.07	0.08	-0.04
	0.39	0.43	0.38	0.22
	0.63	0.67	0.63	0.51
DJF11	-0.07	-0.09	0.04	0.00
	0.39	0.46	0.45	0.32
	0.69	0.68	0.71	0.65
DJF20	0.04	0.13	0.18	-0.01
	0.43	0.48	0.48	0.32
	0.76	0.78	0.79	0.69
MAM7	-0.09	-0.14	-0.23	-0.31
	0.19	0.17	0.15	0.05
	0.48	0.55	0.52	0.37
MAM11	0.14	-0.08	-0.15	-0.13
	0.15	0.18	0.15	0.15
	0.56	0.55	0.48	0.38
MAM20	-0.06	-0.05	-0.05	-0.16
	0.22	0.24	0.19	0.08
	0.52	0.55	0.49	0.31
JJA7	-0.20	-0.28	-0.20	-0.11
	0.08	0.10	0.08	0.05
	0.39	0.39	0.46	0.34
JJA11	-0.11	-0.01	-0.07	-0.24
	0.13	0.18	0.14	0.04
	0.44	0.51	0.40	0.34
JJA20	-0.10	-0.20	-0.10	-0.10
	0.12	0.06	0.15	0.19
	0.58	0.44	0.54	0.48
SON7	-0.01	0.19	-0.06	-0.01
	0.38	0.41	0.34	0.31
	0.71	0.69	0.68	0.66
SON11	0.06	0.17	0.08	-0.03
	0.34	0.45	0.37	0.23
	0.69	0.71	0.63	0.56
SON20	0.03	-0.05	0.05	0.01
	0.36	0.34	0.36	0.22
	0.68	0.67	0.68	0.48

Tabelle A.8: Zusammenfassung (Minimum, Mittelwert, Maximum) des Anomaliekorrelationskoeffizienten für die saisonalen Zeitreihen in den Jahreszeiten DJF, MAM, JJA, SON bei 7, 11 und 20 PCs für die 4 verschiedenen Distanzmaße: City-Block-Distanz (CB), gewichtete Euklidische Distanz (eukg), Euklidische Distanz (eukng) und Mahalanobis Distanz (mahal).

DJF7	r	rmse	m.b	m.an	bias	var.b	var.an	sd.b	sd.an	int.b	int.an
eukl	0.03	2.55	1.94	1.92	-1.12	7.58	8.32	2.75	2.88	2.10	2.10
	0.25	5.97	4.92	4.73	-0.19	50.94	46.09	6.57	6.28	5.52	5.33
	0.40	14.79	13.01	12.13	0.18	414.70	359.18	20.36	18.95	15.00	14.00
eukl.gew	0.06	2.49	1.94	1.93	-1.37	7.58	7.78	2.75	2.79	2.10	2.10
	0.25	5.97	4.92	4.77	-0.15	50.94	47.08	6.57	6.35	5.52	5.35
	0.40	15.00	13.01	12.49	0.19	414.70	358.08	20.36	18.92	15.00	14.70
CB	0.05	2.51	1.94	1.93	-1.35	7.58	8.15	2.75	2.86	2.10	2.10
	0.24	5.93	4.92	4.71	-0.21	50.94	44.88	6.57	6.20	5.52	5.27
	0.40	14.72	13.01	12.27	0.07	414.70	345.59	20.36	18.59	15.00	13.90
mahal	0.10	2.79	1.94	2.49	-0.93	7.58	8.25	2.75	2.87	2.10	2.60
	0.25	6.84	4.92	6.10	1.18	50.94	76.56	6.57	7.54	5.52	7.50
	0.36	19.44	13.01	19.28	8.79	414.70	657.55	20.36	25.64	15.00	26.70
DJF11											
eukl	0.09	2.30	1.94	1.76	-0.81	7.58	6.00	2.75	2.45	2.10	1.90
	0.29	5.53	4.92	4.65	-0.27	50.94	42.78	6.57	6.02	5.52	5.29
	0.41	14.61	13.01	12.51	0.04	414.70	352.55	20.36	18.78	15.00	15.00
eukl.gew	0.12	2.55	1.94	1.89	-1.11	7.58	8.95	2.75	2.99	2.10	1.80
	0.29	5.44	4.92	4.57	-0.35	50.94	43.32	6.57	6.05	5.52	5.05
	0.42	15.80	13.01	12.84	0.12	414.70	392.43	20.36	19.81	15.00	13.98
CB	0.07	2.32	1.94	1.74	-1.51	7.58	5.91	2.75	2.43	2.10	1.80
	0.27	5.67	4.92	4.67	-0.25	50.94	42.96	6.57	6.04	5.52	5.28
	0.40	14.63	13.01	12.71	0.21	414.70	316.55	20.36	17.79	15.00	15.03
mahal	0.16	2.61	1.94	2.15	-0.14	7.58	7.17	2.75	2.68	2.10	2.18
	0.27	6.80	4.92	5.99	1.08	50.94	72.89	6.57	7.76	5.52	6.94
	0.37	18.74	13.01	17.53	4.52	414.70	584.61	20.36	24.18	15.00	22.90
DJF20											
eukl	0.16	2.51	1.94	1.88	-0.85	7.58	7.59	2.75	2.76	2.10	2.00
	0.35	5.43	4.92	4.63	-0.28	50.94	42.32	6.57	5.97	5.52	5.24
	0.49	14.96	13.01	12.60	0.18	414.70	373.29	20.36	19.32	15.00	14.40
eukl.gew	0.15	2.52	1.94	1.89	-0.96	7.58	7.71	2.75	2.78	2.10	1.90
	0.36	5.38	4.92	4.58	-0.34	50.94	42.12	6.57	5.97	5.52	5.14
	0.51	14.97	13.01	12.22	0.12	414.70	360.98	20.36	19.00	15.00	14.00
CB	0.16	2.36	1.94	1.86	-0.67	7.58	6.65	2.75	2.58	2.10	2.10
	0.33	5.70	4.92	4.76	-0.16	50.94	45.65	6.57	6.25	5.52	5.36
	0.48	14.91	13.01	12.48	0.16	414.70	368.52	20.36	19.20	15.00	14.90
mahal	0.17	2.62	1.94	2.18	-0.65	7.58	7.29	2.75	2.70	2.10	2.20
	0.26	6.51	4.92	5.41	0.49	50.94	65.54	6.57	7.42	5.52	6.07
	0.35	16.10	13.01	14.31	2.14	414.70	429.77	20.36	20.73	15.00	18.00

Tabelle A.9: Ergebnisse der statistischen Maße: Korrelation (r), RMSE (rmse), Mittelwert der Tageszeitreihe (m.b) und modellierten Tagesreihe (m.an), systematischer Fehler (bias), Varianz (var.b und var.an), Standardabweichung (sd.b und sd.an) und Interquantile (inter.b und inter.an) für DJF bei 7, 11 und 20 Hauptkomponenten.

MAM7	r	rmse	m.b	m.an	bias	var.b	var.an	sd.b	sd.an	int.b	int.an
eukl	0.09	4.99	3.37	3.58	-0.58	21.83	21.35	4.67	4.62	3.60	3.50
	0.18	7.73	5.72	5.71	-0.01	61.21	59.48	7.52	7.44	6.52	6.49
	0.28	18.15	13.35	12.77	0.27	373.12	352.76	19.32	18.78	15.08	14.40
eukl	0.09	4.74	3.37	3.50	-0.58	21.83	19.39	4.67	4.40	3.60	3.40
	0.19	7.61	5.70	5.65	-0.04	60.71	58.43	7.48	7.35	6.48	6.43
	0.28	18.22	13.35	12.77	0.23	373.12	361.69	19.32	19.02	15.08	14.25
CB	0.10	4.81	3.37	3.31	-0.52	21.83	20.87	4.67	4.57	3.60	3.40
	0.18	7.69	5.70	5.69	-0.01	60.71	60.10	7.48	7.45	6.48	6.46
	0.29	18.12	13.35	12.83	0.37	373.12	365.70	19.32	19.12	15.08	14.38
mahal	0.10	5.14	3.37	4.35	-1.56	21.83	24.08	4.67	4.91	3.60	5.10
	0.20	8.20	5.70	6.98	1.28	60.71	81.97	7.48	8.60	6.48	9.11
	0.29	20.84	13.35	18.33	4.98	373.12	551.37	19.32	23.48	15.08	21.50
MAM11											
eukl	0.14	4.71	3.37	3.45	-0.64	21.83	20.90	4.67	4.57	3.60	3.60
	0.24	7.47	5.70	5.52	-0.18	60.71	59.04	7.48	7.38	6.48	6.21
	0.34	18.42	13.35	12.71	0.11	373.12	384.00	19.32	19.60	15.08	14.40
eukl.gew	0.15	4.66	3.37	3.30	-0.87	21.83	21.67	4.67	4.66	3.60	3.50
	0.25	7.41	5.70	5.53	-0.17	60.71	59.66	7.48	7.39	6.48	6.18
	0.35	17.53	13.35	12.48	0.25	373.12	348.41	19.32	18.67	15.08	13.65
CB	0.13	4.63	3.37	3.41	-1.13	21.83	22.98	4.67	4.79	3.60	3.50
	0.23	7.44	5.70	5.52	-0.18	60.71	58.21	7.48	7.37	6.48	6.17
	0.34	17.81	13.35	12.22	0.22	373.12	331.78	19.32	18.21	15.08	13.68
mahal	0.17	4.90	3.37	4.40	0.24	21.83	21.97	4.67	4.69	3.60	5.20
	0.25	8.92	5.70	7.98	2.28	60.71	97.16	7.48	9.53	6.48	10.20
	0.35	19.76	13.35	17.86	4.51	373.12	424.08	19.32	20.59	15.08	23.00
MAM20											
eukl	0.14	4.49	3.37	3.19	-0.81	21.83	19.93	4.67	4.46	3.60	3.33
	0.25	7.15	5.70	5.36	-0.34	60.71	52.61	7.48	6.94	6.48	6.18
	0.35	17.73	13.35	13.23	0.37	373.12	349.20	19.32	18.69	15.08	16.15
eukl.gew	0.13	4.40	3.37	3.02	-1.25	21.83	19.63	4.67	4.43	3.60	3.00
	0.26	7.06	5.70	5.26	-0.44	60.71	51.24	7.48	6.84	6.48	6.03
	0.36	17.25	13.35	13.30	0.07	373.12	343.13	19.32	18.52	15.08	16.08
CB	0.16	4.47	3.37	3.21	-0.88	21.83	21.49	4.67	4.64	3.60	2.90
	0.26	7.14	5.70	5.38	-0.32	60.71	53.15	7.48	6.97	6.48	6.18
	0.37	17.91	13.35	13.38	0.49	373.12	352.25	19.32	18.77	15.08	16.35
mahal	0.14	5.23	3.37	4.11	-0.33	21.83	27.68	4.67	5.26	3.60	4.50
	0.22	8.44	5.70	7.20	1.50	60.71	78.03	7.48	8.59	6.48	8.70
	0.28	18.49	13.35	15.29	3.28	373.12	355.28	19.32	18.85	15.08	17.70

Tabelle A.10: Ergebnisse der statistischen Maße: Korrelation (r), RMSE (rmse), Mittelwert der Tageszeitreihe (m.b) und modellierten Tagesreihe (m.an), systematischer Fehler (bias), Varianz (var.b und var.an), Standardabweichung (sd.b und sd.an) und Interquantile (inter.b und inter.an) für MAM bei 7, 11 und 20 Hauptkomponenten.

JJA7	r	rmse	m.b	m.an	bias	var.b	var.an	sd.b	sd.an	int.b	int.an
eukl	0.13	6.38	4.59	4.45	-1.09	42.34	42.41	6.51	6.51	5.60	5.30
	0.22	10.90	8.32	8.04	-0.28	109.51	102.62	10.29	9.95	9.86	9.44
	0.30	16.64	13.55	12.97	0.11	260.55	251.97	16.14	15.87	16.10	15.35
eukl.gew	0.15	6.38	4.59	4.34	-0.86	42.34	41.16	6.51	6.42	5.60	5.10
	0.23	10.76	8.28	7.95	-0.33	108.73	100.91	10.25	9.86	9.80	9.29
	0.30	16.36	13.55	13.12	0.21	260.55	247.79	16.14	15.74	16.10	15.30
CB	0.11	6.41	4.59	4.45	-0.97	42.34	41.94	6.51	6.48	5.60	5.30
	0.22	10.88	8.28	8.00	-0.28	108.73	103.13	10.25	9.99	9.80	9.37
	0.29	16.30	13.55	13.09	0.25	260.55	238.85	16.14	15.45	16.10	15.05
mahal	0.18	7.22	4.59	7.07	0.14	42.34	63.95	6.51	8.00	5.60	7.00
	0.26	12.95	8.28	11.67	3.39	108.73	201.94	10.25	13.74	9.80	14.28
	0.33	21.03	13.55	19.31	8.52	260.55	634.29	16.14	25.19	16.10	26.90
JJA11											
eukl	0.14	4.71	3.37	3.45	-0.64	21.83	20.90	4.67	4.57	3.60	3.60
	0.24	7.47	5.70	5.52	-0.18	60.71	59.04	7.48	7.38	6.48	6.21
	0.34	18.42	13.35	12.71	0.11	373.12	384.00	19.32	19.60	15.08	14.40
eukl.gew	0.15	4.66	3.37	3.30	-0.87	21.83	21.67	4.67	4.66	3.60	3.50
	0.25	7.41	5.70	5.53	-0.17	60.71	59.66	7.48	7.39	6.48	6.18
	0.35	17.53	13.35	12.48	0.25	373.12	348.41	19.32	18.67	15.08	13.65
CB	0.13	4.63	3.37	3.41	-1.13	21.83	22.98	4.67	4.79	3.60	3.50
	0.23	7.44	5.70	5.52	-0.18	60.71	58.21	7.48	7.37	6.48	6.17
	0.34	17.81	13.35	12.22	0.22	373.12	331.78	19.32	18.21	15.08	13.68
mahal	0.17	4.90	3.37	4.40	0.24	21.83	21.97	4.67	4.69	3.60	5.20
	0.25	8.92	5.70	7.98	2.28	60.71	97.16	7.48	9.53	6.48	10.20
	0.35	19.76	13.35	17.86	4.51	373.12	424.08	19.32	20.59	15.08	23.00
JJA20											
eukl	0.08	6.52	4.59	4.57	-1.01	42.34	45.95	6.51	6.78	5.60	5.60
	0.16	10.70	8.28	8.07	-0.21	108.73	100.13	10.25	9.82	9.80	9.54
	0.27	16.71	13.55	13.57	0.48	260.55	272.78	16.14	16.52	16.10	16.20
euk.gew	0.09	6.58	4.59	4.70	-1.37	42.34	49.00	6.51	7.00	5.60	5.40
	0.18	10.53	8.28	7.92	-0.36	108.73	97.92	10.25	9.73	9.80	9.32
	0.27	16.19	13.55	13.35	0.32	260.55	236.77	16.14	15.39	16.10	16.30
CB	0.06	6.62	4.59	4.67	-0.73	42.34	46.22	6.51	6.80	5.60	5.60
	0.16	10.66	8.28	7.98	-0.30	108.73	96.45	10.25	9.66	9.80	9.38
	0.26	16.34	13.55	13.37	0.39	260.55	224.95	16.14	15.00	16.10	15.70
Ymahal	0.06	6.84	4.59	5.21	-0.10	42.34	55.81	6.51	7.47	5.60	6.00
	0.14	11.85	8.28	9.34	1.06	108.73	147.37	10.25	11.84	9.80	10.88
	0.23	18.83	13.55	15.32	2.21	260.55	389.43	16.14	19.73	16.10	19.00

Tabelle A.11: Ergebnisse der statistischen Maße: Korrelation (r), RMSE (rmse), Mittelwert der Tageszeitreihe (m.b) und modellierten Tagesreihe (m.an), systematischer Fehler (bias), Varianz (var.b und var.an), Standardabweichung (sd.b und sd.an) und Interquantile (inter.b und inter.an) für JJA bei 7, 11 und 20 Hauptkomponenten.

SON7	r	rmse	m.b	m.an	bias	var.b	var.an	sd.b	sd.an	int.b	int.an
euk.ngw	0.14	4.16	3.16	3.19	-1.35	22.24	22.02	4.72	4.69	3.50	3.68
	0.29	8.32	6.85	6.73	-0.12	93.94	92.84	9.04	9.01	8.10	7.89
	0.42	21.90	20.22	18.88	0.28	726.54	664.30	26.95	25.77	25.60	23.65
euk.gw	0.16	4.24	3.16	3.26	-0.99	22.24	23.60	4.72	4.86	3.50	3.60
	0.30	8.17	6.85	6.67	-0.17	93.94	90.83	9.04	8.90	8.10	7.78
	0.42	22.00	20.22	19.80	0.16	726.54	701.69	26.95	26.49	25.60	24.13
CB	0.15	4.14	3.16	3.12	-0.56	22.24	22.82	4.72	4.78	3.50	3.60
	0.29	8.29	6.85	6.77	-0.08	93.94	91.68	9.04	8.97	8.10	8.01
	0.42	21.82	20.22	19.76	0.52	726.54	660.73	26.95	25.70	25.60	25.10
mahal	0.16	4.30	3.16	3.05	-2.72	22.24	13.76	4.72	3.71	3.50	3.00
	0.27	9.02	6.85	7.48	0.63	93.94	123.12	9.04	10.10	8.10	8.18
	0.40	26.19	20.22	25.52	5.29	726.54	1151.82	26.95	33.94	25.60	29.30
SON11											
euk.ngw	0.16	4.15	3.16	3.13	-1.30	22.24	21.19	4.72	4.60	3.50	3.60
	0.30	8.23	6.85	6.72	-0.12	93.94	92.64	9.04	9.02	8.10	7.92
	0.41	21.95	20.22	19.03	0.35	726.54	696.35	26.95	26.39	25.60	23.60
euk.gw	0.17	4.25	3.16	3.10	-0.74	22.24	22.55	4.72	22.55	3.50	4.75
	0.31	8.15	6.85	6.70	-0.14	93.94	90.06	9.04	90.06	8.10	8.86
	0.42	22.10	20.22	19.68	0.23	726.54	691.60	26.95	691.60	25.60	26.30
CB	0.17	4.11	3.16	3.15	-0.85	22.24	21.67	4.72	4.65	3.50	3.60
	0.30	8.16	6.85	6.63	-0.22	93.94	90.71	9.04	8.87	8.10	7.78
	0.39	22.25	20.22	19.43	0.25	726.54	713.20	26.95	26.71	25.60	24.20
mahal	0.18	4.09	3.16	3.21	-0.46	22.24	16.28	4.72	4.04	3.50	3.60
	0.27	8.92	6.85	7.74	0.90	93.94	104.42	9.04	9.51	8.10	9.44
	0.37	24.69	20.22	22.04	2.65	726.54	809.57	26.95	28.45	25.60	29.20
SON20											
euk.ngw	0.11	3.96	3.16	2.84	-0.87	22.24	18.60	4.72	4.31	3.50	3.10
	0.20	7.82	6.85	6.44	-0.41	93.94	81.05	9.04	8.37	8.10	7.69
	0.33	21.74	20.22	20.19	0.29	726.54	684.48	26.95	26.16	25.60	26.10
euk.gw	0.11	3.95	3.16	2.87	-1.12	22.24	20.93	4.72	4.58	3.50	3.30
	0.23	7.72	6.85	6.35	-0.49	93.94	83.99	9.04	8.46	8.10	7.45
	0.37	21.46	20.22	20.08	0.12	726.54	757.08	26.95	27.52	25.60	24.80
CB	0.10	3.97	3.16	2.78	-1.14	22.24	18.47	4.72	4.30	3.50	3.00
	0.20	7.82	6.85	6.40	-0.45	93.94	82.58	9.04	8.34	8.10	7.63
	0.31	23.09	20.22	21.50	1.28	726.54	802.44	26.95	28.33	25.60	26.60
mahal	0.08	4.28	3.16	3.51	-0.22	22.24	20.63	4.72	4.54	3.50	4.70
	0.18	9.10	6.85	7.87	1.03	93.94	109.37	9.04	9.86	8.10	9.21
	0.27	23.34	20.22	21.13	2.59	726.54	659.85	26.95	25.69	25.60	23.50

Tabelle A.12: Ergebnisse der statistischen Maße: Korrelation (r), RMSE (rmse), Mittelwert der Tageszeitreihe (m.b) und modellierten Tagesreihe (m.an), systematischer Fehler (bias), Varianz (var.b und var.an), Standardabweichung (sd.b und sd.an) und Interquantile (inter.b und inter.an) für SON bei 7, 11 und 20 Hauptkomponenten

DJF7	r	rmse	m.b	m.an	bias	var.b	var.an	sd.b	sd.an	int.b	int.an
euk.ngw	0.08	0.29	0.69	0.67	-0.34	0.06	0.04	0.24	0.21	0.27	0.26
	0.44	0.72	1.98	1.90	-0.08	0.73	0.49	0.76	0.63	1.04	0.81
	0.86	1.68	5.68	5.34	0.01	4.55	3.34	2.13	1.83	2.91	2.47
euk.gw	0.10	0.29	0.69	0.66	-0.41	0.06	0.05	0.24	0.23	0.27	0.30
	0.46	0.73	1.98	1.92	-0.06	0.73	0.56	0.76	0.67	1.04	0.82
	0.80	1.82	5.68	5.55	0.05	4.55	3.36	2.13	1.83	2.91	2.80
CB	0.03	0.32	0.69	0.66	-0.43	0.06	0.05	0.24	0.23	0.27	0.23
	0.42	0.72	1.98	1.88	-0.10	0.73	0.47	0.76	0.62	1.04	0.80
	0.84	1.96	5.68	5.43	0.00	4.55	2.99	2.13	1.73	2.91	2.86
mahal	-0.05	0.41	0.69	0.85	-0.31	0.06	0.10	0.24	0.32	0.27	0.40
	0.36	1.16	1.98	2.64	0.66	0.73	1.02	0.76	0.88	1.04	1.17
	0.79	3.87	5.68	7.76	3.19	4.55	7.63	2.13	2.76	2.91	4.86
DJF11											
euk.ngw	0.09	0.30	0.69	0.60	-0.45	0.06	0.04	0.24	0.21	0.27	0.23
	0.46	0.70	1.98	1.84	-0.14	0.73	0.49	0.76	0.61	1.04	0.77
	0.87	1.81	5.68	5.49	-0.02	4.55	3.79	2.13	1.95	2.91	2.79
euk.gw	0.06	0.29	0.69	0.64	-0.28	0.06	0.04	0.24	0.20	0.27	0.21
	0.47	0.70	1.98	1.89	-0.09	0.73	0.56	0.76	0.65	1.04	0.82
	0.87	1.83	5.68	5.47	0.03	4.55	4.32	2.13	2.08	2.91	3.08
CB	-0.01	0.25	0.69	0.58	-0.59	0.06	0.02	0.24	0.15	0.27	0.19
	0.44	0.71	1.98	1.87	-0.11	0.73	0.52	0.76	0.62	1.04	0.82
	0.87	1.86	5.68	5.66	0.09	4.55	3.95	2.13	1.99	2.91	2.73
mahal	0.18	0.33	0.69	0.88	0.14	0.06	0.07	0.24	0.27	0.27	0.34
	0.45	1.14	1.98	2.76	0.78	0.73	0.88	0.76	0.84	1.04	1.09
	0.73	3.68	5.68	7.50	2.68	4.55	7.75	2.13	2.78	2.91	4.28
DJF20											
euk.ngw	0.18	0.27	0.69	0.64	-0.43	0.06	0.06	0.24	0.24	0.27	0.34
	0.46	0.70	1.98	1.82	-0.15	0.73	0.47	0.76	0.61	1.04	0.80
	0.91	1.73	5.68	5.38	0.07	4.55	3.06	2.13	1.75	2.91	2.37
euk.gw	0.13	0.29	0.69	0.64	-0.47	0.06	0.06	0.24	0.24	0.27	0.31
	0.50	0.69	1.98	1.79	-0.18	0.73	0.52	0.76	0.63	1.04	0.83
	0.90	1.65	5.68	5.24	-0.01	4.55	3.78	2.13	1.94	2.91	2.18
CB	0.14	0.28	0.69	0.63	-0.57	0.06	0.05	0.24	0.23	0.27	0.32
	0.45	0.72	1.98	1.76	-0.22	0.73	0.42	0.76	0.59	1.04	0.77
	0.85	2.13	5.68	5.13	-0.01	4.55	2.75	2.13	1.66	2.91	2.40
mahal	-0.01	0.35	0.69	0.83	-0.15	0.06	0.06	0.24	0.25	0.27	0.22
	0.41	0.89	1.98	2.34	0.36	0.73	0.70	0.76	0.76	1.04	1.01
	0.77	2.17	5.68	6.79	1.35	4.55	3.63	2.13	1.91	2.91	2.19

Tabelle A.13: Ergebnisse der statistischen Maße: Korrelation (r), RMSE (rmse), Mittelwert der saisonaler Zeitreihe (m.b) und modellierten Tagesreihe (m.an), systematischer Fehler (bias), Varianz (var.b und var.an), Standardabweichung (sd.b und sd.an) und Interquantile (inter.b und inter.an) für DJF bei 7, 11 und 20 Hauptkomponenten.

MAM7	r	rmse	m.b	m.an	bias	var.b	var.an	sd.b	sd.an	int.b	int.an
euk.ngw	-0.18	0.51	1.23	1.31	-0.13	0.14	0.14	0.37	0.37	0.41	0.47
	0.18	0.94	2.64	2.68	0.05	0.82	0.46	0.84	0.64	1.03	0.81
	0.71	2.56	6.12	5.99	0.22	5.23	2.36	2.29	1.54	2.99	1.78
euk.gw	-0.08	0.55	1.23	1.28	-0.23	0.14	0.10	0.37	0.31	0.41	0.39
	0.20	0.94	2.64	2.63	-0.01	0.82	0.46	0.84	0.65	1.03	0.85
	0.68	2.64	6.12	5.97	0.15	5.23	2.56	2.29	1.60	2.99	2.06
CB	-0.10	0.45	1.23	1.20	-0.19	0.14	0.11	0.37	0.33	0.41	0.39
	0.21	0.92	2.64	2.68	0.05	0.82	0.45	0.84	0.63	1.03	0.81
	0.68	2.24	6.12	6.00	0.28	5.23	2.41	2.29	1.55	2.99	2.24
mahal	-0.31	0.41	1.23	1.48	-1.02	0.14	0.28	0.37	0.53	0.41	0.64
	0.07	2.21	2.64	2.97	0.33	0.82	1.14	0.84	1.00	1.03	1.32
	0.62	18.66	6.12	8.97	2.85	5.23	7.98	2.29	2.83	2.99	3.31
MAM11											
euk.ngw	-0.15	0.46	1.23	1.26	-0.44	0.14	0.14	0.37	0.38	0.41	0.39
	0.17	0.95	2.64	2.55	-0.09	0.82	0.43	0.84	0.63	1.03	0.80
	0.58	2.37	6.12	5.70	0.06	5.23	2.23	2.29	1.49	2.99	2.12
euk.gw	-0.09	0.43	1.23	1.17	-0.65	0.14	0.11	0.37	0.33	0.41	0.44
	0.20	0.94	2.64	2.53	-0.10	0.82	0.47	0.84	0.65	1.03	0.82
	0.56	2.54	6.12	5.73	0.08	5.23	2.63	2.29	1.62	2.99	1.83
CB	-0.13	0.46	1.23	1.24	-0.63	0.14	0.12	0.37	0.35	0.41	0.42
	0.18	0.93	2.64	2.54	-0.09	0.82	0.41	0.84	0.61	1.03	0.78
	0.58	2.32	6.12	5.59	0.11	5.23	2.21	2.29	1.49	2.99	2.23
mahal	-0.16	0.64	1.23	1.98	0.28	0.14	0.20	0.37	0.44	0.41	0.57
	0.26	3.23	2.64	3.84	1.21	0.82	1.19	0.84	1.04	1.03	1.22
	0.66	13.19	6.12	8.41	2.54	5.23	5.32	2.29	2.31	2.99	2.01
MAM20											
euk.ngw	-0.05	0.22	1.23	1.11	-0.39	0.14	0.15	0.37	0.38	0.41	0.45
	0.22	0.95	2.64	2.48	-0.16	0.82	0.45	0.84	0.63	1.03	0.80
	0.75	4.93	6.12	6.03	0.01	5.23	3.40	2.29	1.84	2.99	2.64
euk.gw	-0.04	0.42	1.23	1.06	-0.59	0.14	0.11	0.37	0.33	0.41	0.28
	0.27	0.90	2.64	2.42	-0.22	0.82	0.43	0.84	0.61	1.03	0.76
	0.78	1.94	6.12	6.01	-0.02	5.23	2.57	2.29	1.60	2.99	2.10
CB	-0.05	0.39	1.23	1.13	-0.41	0.14	0.13	0.37	0.36	0.41	0.40
	0.26	0.89	2.64	2.47	-0.16	0.82	0.40	0.84	0.60	1.03	0.82
	0.70	2.23	6.12	5.93	0.02	5.23	2.41	2.29	1.55	2.99	2.24
mahal	-0.24	0.66	1.23	2.00	0.09	0.14	0.17	0.37	0.41	0.41	0.50
	0.12	1.44	2.64	3.50	0.87	0.82	0.81	0.84	0.86	1.03	1.03
	0.59	2.62	6.12	6.91	2.08	5.23	4.18	2.29	2.04	2.99	1.93

Tabelle A.14: Ergebnisse der statistischen Maße: Korrelation (r), RMSE (rmse), Mittelwert der saisonaler Zeitreihe(m.b) und modellierten Tagesreihe (m.an), systematischer Fehler (bias), Varianz (var.b und var.an), Standardabweichung (sd.b und sd.an) und Interquantile (inter.b und inter.an) für MAM bei 7, 11 und 20 Hauptkomponenten.

JJA7	r	rmse	m.b	m.an	bias	var.b	var.an	sd.b	sd.an	int.b	int.an
euk.ngw	-0.28	0.76	1.87	1.82	-0.45	0.30	0.26	0.54	0.51	0.59	0.54
	0.13	1.25	4.31	4.19	-0.11	1.06	0.87	1.00	0.90	1.23	1.11
	0.47	2.25	8.22	7.95	0.08	3.03	2.78	1.74	1.67	2.31	2.02
euk.gw	-0.21	0.67	1.87	1.76	-0.47	0.30	0.25	0.54	0.50	0.59	0.53
	0.11	1.29	4.31	4.14	-0.17	1.06	0.94	1.00	0.94	1.23	1.14
	0.46	2.18	8.22	8.01	0.03	3.03	3.10	1.74	1.76	2.31	2.14
CB	-0.21	0.76	1.87	1.81	-0.54	0.30	0.24	0.54	0.49	0.59	0.59
	0.10	1.29	4.31	4.20	-0.11	1.06	0.92	1.00	0.93	1.23	1.20
	0.42	2.14	8.22	7.98	0.13	3.03	2.54	1.74	1.59	2.31	2.09
mahal	-0.16	1.19	1.87	2.95	-0.34	0.30	0.92	0.54	0.96	0.59	0.88
	0.09	2.29	4.31	5.68	1.37	1.06	2.63	1.00	1.56	1.23	1.86
	0.49	5.00	8.22	10.74	3.93	3.03	8.51	1.74	2.92	2.31	3.45
JJA11											
euk.ngw	-0.05	0.67	1.87	1.71	-0.31	0.30	0.16	0.54	0.40	0.59	0.48
	0.17	1.23	4.31	4.29	-0.02	1.06	0.90	1.00	0.92	1.23	1.17
	0.44	1.94	8.22	8.32	0.35	3.03	2.63	1.74	1.62	2.31	1.89
euk.gw	-0.01	0.71	1.87	1.71	-0.28	0.30	0.21	0.54	0.46	0.59	0.58
	0.20	1.22	4.31	4.27	-0.03	1.06	0.95	1.00	0.95	1.23	1.21
	0.54	1.92	8.22	8.51	0.28	3.03	2.70	1.74	1.64	2.31	1.99
CB	-0.05	0.67	1.87	1.71	-0.31	0.30	0.16	0.54	0.40	0.59	0.48
	0.17	1.23	4.31	4.29	-0.02	1.06	0.90	1.00	0.92	1.23	1.17
	0.44	1.94	8.22	8.32	0.35	3.03	2.63	1.74	1.62	2.31	1.89
mahal	-0.24	0.92	1.87	2.05	-0.60	0.30	0.34	0.54	0.58	0.59	0.73
	0.08	1.56	4.31	4.55	0.24	1.06	1.54	1.00	1.19	1.23	1.48
	0.64	2.81	8.22	8.39	1.46	3.03	4.20	1.74	2.05	2.31	2.60
JJA20											
eukl.ngw	-0.09	0.83	1.87	1.91	-0.26	0.30	0.30	0.54	0.55	0.59	0.65
	0.17	1.28	4.31	4.32	0.01	1.06	1.09	1.00	1.01	1.23	1.30
	0.56	2.07	8.22	8.39	0.34	3.03	3.07	1.74	1.75	2.31	2.24
eul.gw	-0.20	0.85	1.87	1.87	-0.46	0.30	0.28	0.54	0.52	0.59	0.57
	0.08	1.33	4.31	4.19	-0.12	1.06	0.99	1.00	0.97	1.23	1.23
	0.58	2.22	8.22	8.15	0.15	3.03	2.42	1.74	1.56	2.31	2.57
CB	-0.10	0.78	1.87	1.95	-0.31	0.30	0.31	0.54	0.56	0.59	0.68
	0.14	1.28	4.31	4.28	-0.03	1.06	1.04	1.00	0.99	1.23	1.38
	0.62	2.18	8.22	8.32	0.23	3.03	3.75	1.74	1.94	2.31	3.08
mahal	-0.10	0.66	1.87	2.19	-0.05	0.30	0.24	0.54	0.49	0.59	0.61
	0.22	1.40	4.31	4.79	0.49	1.06	1.24	1.00	1.07	1.23	1.42
	0.49	2.52	8.22	9.21	1.10	3.03	4.14	1.74	2.04	2.31	2.88

Tabelle A.15: Ergebnisse der statistischen Maße: Korrelation (r), RMSE (rmse), Mittelwert der saisonalen Zeitreihe (m.b) und modellierten Tagesreihe (m.an), systematischer Fehler (bias), Varianz (var.b und var.an), Standardabweichung (sd.b und sd.an) und Interquantile (inter.b und inter.an) für JJA bei 7, 11 und 20 Hauptkomponenten.

SON7	r	rmse	m.b	m.an	bias	var.b	var.an	sd.b	sd.an	int.b	int.an
euk.ngw	0.09	0.40	1.06	1.01	-0.43	0.13	0.15	0.36	0.39	0.47	0.36
	0.34	0.98	2.68	2.61	-0.07	1.08	0.80	0.94	0.83	1.24	1.06
	0.67	2.12	7.66	7.23	0.03	7.37	5.78	2.71	2.40	4.26	2.99
euk.gw	0.19	0.39	1.06	1.05	-0.29	0.13	0.16	0.36	0.41	0.47	0.46
	0.42	0.93	2.68	2.58	-0.10	1.08	0.79	0.94	0.82	1.24	1.06
	0.72	2.27	7.66	7.53	0.02	7.37	6.89	2.71	2.63	4.26	3.02
CB	0.13	0.37	1.06	1.00	-0.55	0.13	0.15	0.36	0.38	0.47	0.41
	0.39	0.96	2.68	2.59	-0.09	1.08	0.83	0.94	0.84	1.24	1.02
	0.70	2.29	7.66	7.12	0.06	7.37	7.06	2.71	2.66	4.26	2.83
mahal	-0.01	0.50	1.06	1.35	-0.73	0.13	0.17	0.36	0.41	0.47	0.52
	0.35	1.30	2.68	3.18	0.50	1.08	1.46	0.94	1.10	1.24	1.33
	0.71	3.95	7.66	10.58	2.92	7.37	14.62	2.71	3.82	4.26	5.38
SON11											
euk.ngw	0.08	0.46	1.06	1.04	-0.46	0.13	0.13	0.36	0.37	0.47	0.43
	0.38	0.97	2.68	2.63	-0.05	1.08	0.86	0.94	0.85	1.24	1.18
	0.75	2.36	7.66	7.25	0.20	7.37	7.75	2.71	2.78	4.26	4.38
euk.gw	0.18	0.48	1.06	1.00	-0.32	0.13	0.15	0.36	0.39	0.47	0.45
	0.46	0.91	2.68	2.59	-0.09	1.08	0.87	0.94	0.86	1.24	1.09
	0.72	2.19	7.66	7.41	0.11	7.37	7.46	2.71	2.73	4.26	4.37
CB	0.06	0.42	1.06	1.04	-0.37	0.13	0.11	0.36	0.33	0.47	0.36
	0.36	0.99	2.68	2.59	-0.09	1.08	0.86	0.94	0.85	1.24	1.12
	0.71	2.70	7.66	7.33	0.12	7.37	8.03	2.71	2.83	4.26	3.97
mahal	-0.04	0.54	1.06	1.30	0.14	0.13	0.16	0.36	0.40	0.47	0.53
	0.27	1.27	2.68	3.29	0.61	1.08	0.97	0.94	0.92	1.24	1.19
	0.66	2.95	7.66	9.55	1.89	7.37	7.95	2.71	2.82	4.26	3.61
SON20											
eukl.ngew	0.06	0.50	1.06	0.91	-0.55	0.13	0.18	0.36	0.43	0.47	0.32
	0.39	0.98	2.68	2.40	-0.28	1.08	0.75	0.94	0.79	1.24	1.00
	0.68	2.19	7.66	7.17	-0.02	7.37	7.33	2.71	2.71	4.26	4.36
eul.gew	-0.05	0.52	1.06	0.93	-0.65	0.13	0.18	0.36	0.42	0.47	0.45
	0.37	1.00	2.68	2.38	-0.30	1.08	0.76	0.94	0.80	1.24	1.08
	0.68	2.23	7.66	7.19	-0.03	7.37	7.60	2.71	2.76	4.26	4.50
CB	0.03	0.48	1.06	0.91	-0.62	0.13	0.14	0.36	0.38	0.47	0.43
	0.40	0.98	2.68	2.39	-0.29	1.08	0.77	0.94	0.80	1.24	1.06
	0.65	2.35	7.66	7.83	0.17	7.37	8.52	2.71	2.92	4.26	4.08
Ymahal	0.05	0.62	1.06	1.39	0.28	0.13	0.20	0.36	0.44	0.47	0.39
	0.27	1.45	2.68	3.47	0.79	1.08	1.31	0.94	1.07	1.24	1.30
	0.51	3.11	7.66	8.61	1.79	7.37	7.76	2.71	2.79	4.26	3.38

Tabelle A.16: Ergebnisse der statistischen Maße: Korrelation (r), RMSE (rmse), Mittelwert der saisonalen Zeitreihe (m.b) und modellierten Tagesreihe (m.an), systematischer Fehler (bias), Varianz (var.b und var.an), Standardabweichung (sd.b und sd.an) und Interquartile (inter.b und inter.an) für SON bei 7, 11 und 20 Hauptkomponenten.

DJF7	PODa	FARa	POFDa	bias.score	HSSa	TSSa
CB	0.35	0.30	0.17	0.95	0.01	0.04
	0.52	0.47	0.30	0.99	0.05	0.22
	0.70	0.64	0.43	1.02	0.08	0.32
euk.gw	0.33	0.31	0.17	0.95	0.01	0.04
	0.53	0.47	0.30	0.99	0.06	0.22
	0.69	0.65	0.42	1.01	0.09	0.34
euk.ngw	0.34	0.30	0.17	0.94	0.00	0.01
	0.52	0.47	0.30	0.99	0.06	0.22
	0.70	0.65	0.43	1.03	0.08	0.34
mahl	0.34	0.32	0.20	0.91	0.02	0.09
	0.57	0.48	0.35	1.10	0.05	0.22
	0.73	0.64	0.47	1.29	0.08	0.31

Tabelle A.17: Skill Scores DJF bei 7 Hauptkomponenten.

DJF20	PODa	FARa	POFDa	bias.score	HSSa	TSSa
CB	0.37	0.27	0.16	0.91	0.03	0.14
	0.55	0.43	0.26	0.95	0.07	0.28
	0.71	0.61	0.38	1.00	0.10	0.40
euk.gw	0.40	0.26	0.17	0.90	0.03	0.12
	0.57	0.42	0.26	0.97	0.08	0.31
	0.72	0.59	0.37	1.03	0.11	0.43
euk.ngw	0.39	0.27	0.17	0.91	0.03	0.13
	0.56	0.42	0.27	0.98	0.07	0.30
	0.73	0.61	0.38	1.04	0.10	0.41
mahal	0.36	0.31	0.18	0.93	0.03	0.14
	0.56	0.48	0.33	1.07	0.06	0.23
	0.73	0.62	0.45	1.21	0.08	0.31

Tabelle A.18: Skill Scores DJF bei 20 Hauptkomponenten.

JJA7	PODa	FARa	POFDa	bias.score	HSSa	TSSa
CB	0.43	0.27	0.31	0.98	0.03	0.11
	0.61	0.40	0.42	1.01	0.05	0.19
	0.72	0.57	0.54	1.05	0.07	0.26
euk.gw	0.45	0.27	0.30	0.98	0.03	0.14
	0.61	0.39	0.41	1.00	0.05	0.20
	0.74	0.55	0.53	1.03	0.07	0.27
euk.ngw	0.44	0.26	0.29	0.97	0.03	0.13
	0.61	0.39	0.41	1.00	0.05	0.20
	0.74	0.55	0.52	1.02	0.07	0.27
mahal	0.46	0.27	0.28	0.78	0.04	0.15
	0.59	0.37	0.37	0.95	0.06	0.23
	0.71	0.53	0.51	1.13	0.07	0.30

Tabelle A.19: Skill Scores JJA bei 7 Hauptkomponenten.

JJA20	PODa	FARa	POFDa	bias.score	HSSa	TSSa
CB	0.47	0.26	0.31	1.00	0.04	0.15
	0.63	0.39	0.42	1.03	0.05	0.21
	0.76	0.54	0.54	1.08	0.07	0.28
euk.gw	0.45	0.26	0.28	0.96	0.03	0.13
	0.63	0.38	0.40	1.01	0.06	0.22
	0.76	0.54	0.52	1.06	0.08	0.32
euk.ngw	0.45	0.26	0.29	0.98	0.04	0.15
	0.63	0.39	0.42	1.03	0.05	0.21
	0.77	0.55	0.55	1.07	0.07	0.29
mahal	0.43	0.26	0.29	0.90	0.03	0.14
	0.60	0.39	0.40	0.98	0.05	0.20
	0.74	0.55	0.51	1.10	0.07	0.26

Tabelle A.20: Skill Scores JJA bei 20 Hauptkomponenten.

Anhang B

Abbildungen

Die folgenden Abbildungen stellen die Korrelationen zwischen den saisonalen Beobachtungsreihen und den entsprechenden aufgrund der Distanzmaße modellierten Zeitreihen dar. In Abschnitt 7.2 wird auf diese Abbildungen eingegangen.

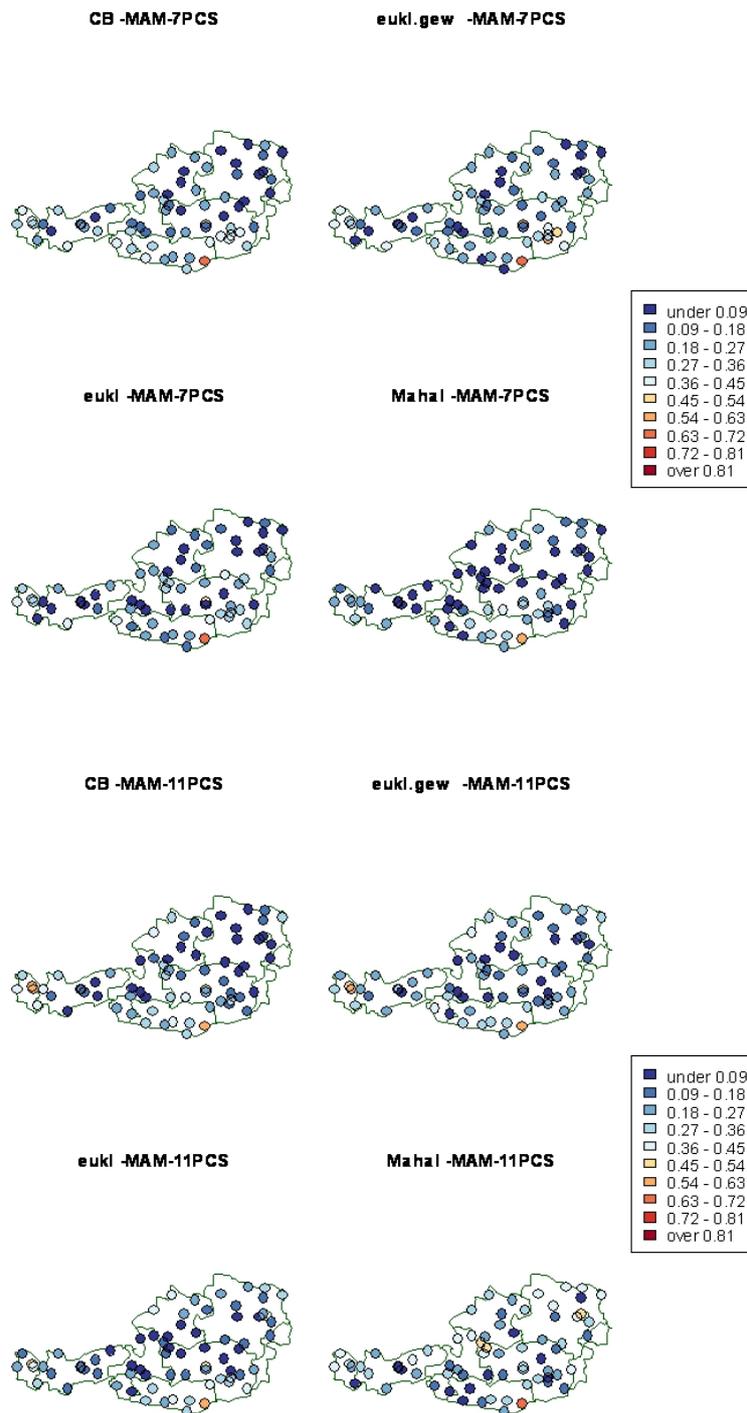


Abbildung B.1: MAM-7PCS, MAM-11PCS: Korrelationen zwischen saisonaler und modellierter Zeitreihe bei Anwendung der City-Block-Distanz, der gewichteten Euklidischen Distanz, der Euklidischen Distanz und der Mahalanobis Distanz.

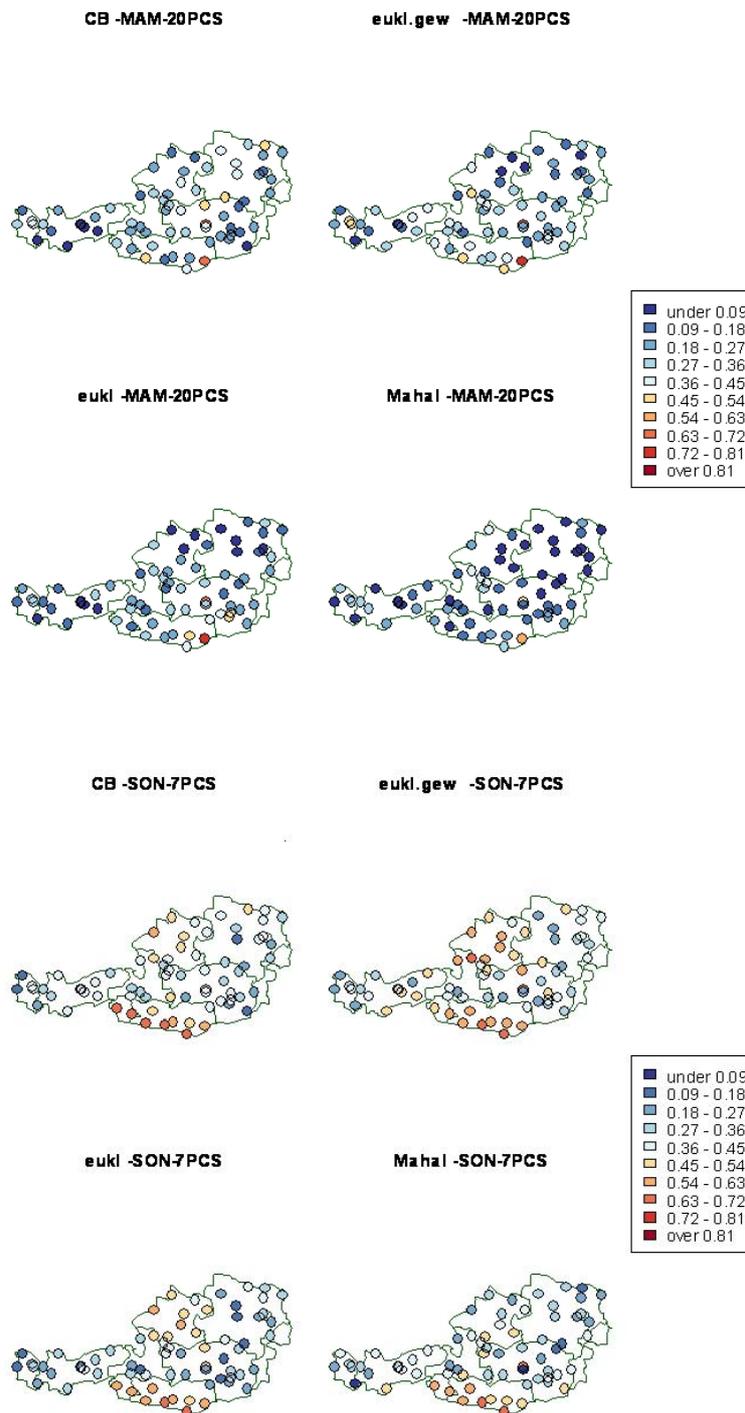


Abbildung B.2: MAM-20PCS, SON-7PCS: Korrelationen zwischen saisonaler und modellierter Zeitreihe bei Anwendung der City-Block-Distanz, der gewichteten Euklidischen Distanz, der Euklidischen Distanz und der Mahalanobis Distanz.

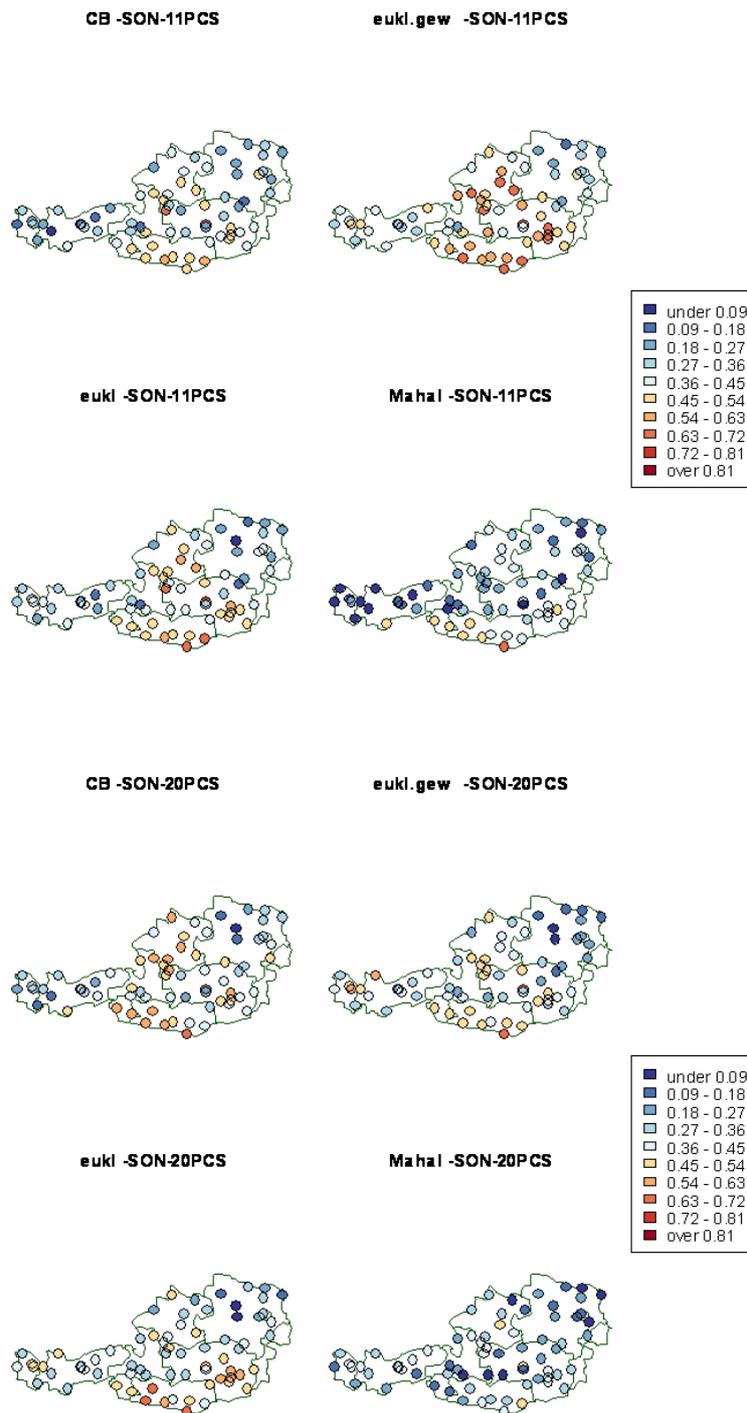


Abbildung B.3: SON-11PCS, SON-20PCS: Korrelationen zwischen saisonaler und modellierter Zeitreihe bei Anwendung der City-Block-Distanz, der gewichteten Euklidischen Distanz, der Euklidischen Distanz und der Mahalanobis Distanz.

Anhang C

Danksagung

Ich möchte mich an dieser Stelle bei meiner Familie und allen jenen Personen bedanken, die mich sowohl in guten, als auch in schlechten Zeiten tatkräftig unterstützt haben und mir verständnisvoll zur Seite gestanden sind.

Als erstes möchte mich bei Univ. Prof. Dr. Georg Skoda sowohl für die Betreuung meines Diplomarbeitsthemas, als auch für die vielen motivierenden und hilfreichen Gespräche bedanken. Weiters danke ich Univ. Prof. Dr. Michael Hantel und der gesamten Arbeitsgruppe der Theoretischen Meteorologie der Universität Wien für die vielen fruchtbaren Diskussionen und die Hilfe bei fachlichen Problemen.

Mein besonderer Dank gilt DI Horst Wildenauer vom Institut für Automatisierungs- und Regelungstechnik der Technischen Universität Wien. Durch unsere Gespräche wurde die Idee geboren, die im Bereich der statistischen Mustererkennung gebräuchlichen Distanzmaße für Statistisches Downscaling zu adaptieren. Er richtete meine Aufmerksamkeit auf die Publikationen von Kelly Yambor, welche letztendlich den Grundstein für meine in dieser Arbeit beschriebenen Analysen legten. Gleichzeitig stand er mir sowohl mit seinem Wissen im Bereich der PCA basierten explorativen Datenanalyse als auch bei Fragen zur Implementierung zur Seite.

Schließlich möchte ich Mag. Ursula Riegl dankend erwähnen, die sich prompt bereit erklärte mein Skriptum auf noch vorhandene orthographische Fehler zu korrigieren.

Anhang D

Persönliche Daten



- Name: Elsa Maria Dos Santos Cardoso
- Geburtsdatum: 11.06.72 Lissabon, Portugal
- e-mail: elsa.m@gmx.net

- Mutter: Maria Isabel Dos Santos
- Vater: Joao Ferreira Cardoso
- Geschwichter: Maria de Fatima Dos Santos Cardoso,
Joao Carlos Dos Santos Cardoso

Schule und Studium

- 1984-87 Escola Secundaria (AHS) in Palmela, Portugal
- 1988-90 Escola Secundaria D. Joao in Setubal, Portugal - AHS-Schulabschluss 1992
- 1994-97 Externistenprüfung zur Nostrifikation des „Oberstufenrealgymnasiums mit ergänzendem Unterricht in Biologie, Umweltkunde, Physik und Chemie“ am Bundesrealgymnasium Kundmannngasse, Wien
- 06/1998 Österreichische Matura am Bundesrealgymnasium Kundmannngasse, Wien
- WS 98/99 Beginn des Diplomstudiums der Meteorologie und Geophysik an der Uni Wien
- Mai/2001 Abschluss des 1. Studienabschnittes in Meteorologie und Geophysik

Berufserfahrung in Zusammenhang mit dem Studium:

- 02/2004-03/2007: Projektmitarbeiterin bei RECLIP:more (REsearch for CLImate Protection: model run evaluation), Abteilung für Klimatologie, Zentralanstalt für Meteorologie und Geodynamik (ZAMG)
- 07/2003: Flugmeteorologisches Praktikum bei der Firma Austrocontrol, Abteilung Meteorologie, Flughafen Wien-Schwechat
- 06/2001: Wissenschaftliche Recherche zum Thema Klimawandel, Institut für Meteorologie und Physik, Universität für Bodenkultur Wien

Literaturverzeichnis

- [1] S. J. Bauer, R. Gutdeutsch, M. Hantel, H. Reuter, H.O. Rucker, Siedler G., T. Spohn, R. Steinacker, and W. Zenk. *Erde und Planeten*, volume Band 7 of *Lehrbuch der Experimentalphysik*. Wilhelm Raith, 1997.
- [2] F. Baur. *Grosswetterkunde und langfristige Witterungsvorhersage*. Frankfurt a.M. : Akadem. Verl.-Ges., 1963.
- [3] R.E. Benestad, D. Chen, and I. Hanssen-Bauer. *Empirical-Statistical Downscaling*. World Scientific Publishers In Press., 2007.
- [4] H.H. Bock. *Automatische Klassifikation. Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten*. Vandenhoeck and Ruprecht, Göttingen, 1974.
- [5] G.B. Bonan. *Ecological Climatology: Concepts and Applications*. Cambridge University Press, 2002.
- [6] A. Bárdossy and E. J. Plate. Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resour. Res.*, 28:1247–1259, 1992.
- [7] C.S. Bretherton, C. Smith, and J.M. Wallace. An intercomparison of methods for finding coupled patterns in climate data. *J. Climate*, 5:541–560, 1992.
- [8] H. Cubasch, U. von Storch. Estimates of climate change in southern europe using different downscaling technique. *Clim. Res.*, 7:129–149, 1996.
- [9] U. Cubasch, H. von Storch, J. Waskewitz, and E. Zorita. Estimates of climate change in southern europe using different downscaling techniques. *Clim. Res.*, 7:129–149, 1996.
- [10] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley Interscience, 2001.
- [11] E. N. Lorenz. Empirical Orthogonal Functions and Statistic Weather Prediction . *Sci. Rept. No. 1, Statistical Forecasting Project, Mass. Inst. Tech., Dept. of Meteorology, Cambridge, Mass.*, page 49pp, 1956.

-
- [12] M. Ehrendorfer. *Hauptkomponentenanalyse des Niederschlages*. Diplomarbeit, Universität Wien, 1985.
- [13] F. Fliri. Niederschlag und lufttemperatur im alpenraum. *Wissenschaftliche Alpenvereinshefte*, 24:1–100, 1996.
- [14] K. Fraedrich, A. A. Aigner, E. Kirk, and F. Lunkeit. General circulation models of the atmosphere new york. In *Encyclopedia of Nonlinear Science*, 2005.
- [15] F. Frey-Buness, D. Heimann, and R. Sausen. A statistical-dynamical downscaling procedure for global climate simulations. *Theor. Appl. Climatol.*, 50:117–131, 1995.
- [16] Heimann D. Fuentes U. Verification of statistical-dynamical downscaling in the alpine region. *Climate Res.*, 7:151–168, 1996.
- [17] F. Giorgi and L. Mearns. Approaches to the simulation of regional climate change: a review. *J Geophys Res*, 29:191–216, 1991.
- [18] G.H. Golub and C.F. Van Loan. *Matrix Computations*. JohnsHopkinsPress, Baltimore, MD, second edition, 1989.
- [19] H. Graßl. Natürliche klimaschwankungen–eine einföhrung. *Promet*, 29(1-4):1–2, 1998.
- [20] SL Grotch and MC MacCracken. The use of general circulation models to predict regional climate change. *J. Climate*, 4:286–303, 1991.
- [21] D. Gyalistras, H. von Storch, A. Fischlin, and M. Beniston. Linking gcm-simulated climatic changes to ecosystem models: case studies of statistical downscaling in the alps. *Clim. Res.*, 4:117–189, 1994.
- [22] J. Hann. *Handbuch der Klimatologie*. Engelhorn, Stuttgart, 1883.
- [23] H. Hantel, M. Kraus and C.D. Schoenwiese. *Climate definition*. Fischer G, 1987.
- [24] B. Hewitson and R. G. Crane. Large-scale atmospheric controls on local precipitation in tropical mexico. *Geophys. Res. Lett.*, 19:1835–1838, 1992.
- [25] H. Hotelling. Relations between two sets of variants. *Biometrika*, 28:321–377, 1936.
- [26] F. Kauker and J.M. Oberhuber. A regional version of the ocean general circulation model opyc with open boundaries and tides. *Tellus*, in press, 1998.
- [27] F. Lauscher. 25 jahre mit täglicher klassifikation der wetterlage in den ostalpenländern. *Wetter und Leben*, 24:185–189, 1972.

- [28] C. Lindberg and A. J. Broccoli. Representation of topography in spectral climate models and its effect on simulated precipitation. *J. Climate*, 9(11):2641–2659, 1996.
- [29] E. N. Lorenz. Climate predictability: The physical basis of climate and climate modeling. global atmosphere prediction programme. Report 16, WMO, Geneva, 1996.
- [30] B. Machenauer, M. Windelband, M. and Botzet, R. Jones, and M. Déqué. Validation of present-day regional climate simulations over europe: Nested lam and variable resolution global model simulations with observed or mixed layer ocean boundary conditions. Technical Report"191, Max-Planck-Institut für Meteorologie (MPI), 1996.
- [31] L. O. Mearns, F. Giorgi, L. McDaniel, and C. Shields. Analysis of daily variability of precipitation in a nested regional climate model: Comparisons with observations and doubled co2 results. *Global Planet. Change*, 10:55–78, 1995.
- [32] J. Prohaska. A technique for analyzing the linear relationships between two meteorological fields. *Mon. Wea. Rev.*, 104:1345–1353, 1976.
- [33] P. Racksko, L. Szeidl, and M. Semenov. A serial approach to local stochastic weather models. *Ecological modelling*, 57:27–41, 1991.
- [34] C. Richardson. Stochastic simulation of daily precipitation, temperature and solar radiation. *Water Resources Research*, 17:182–190, 1981.
- [35] James S. Risbey and Peter H. Stone. A case study of the adequacy of gcm simulations for input to regional climate change assessments. *J. Climate* 9, 7:1441–1467, 1996.
- [36] D.J. Sailor and X. Li. A semiempirical downscaling approach for predicting regional temperature impacts associated with climatic change. *J. Climate*, 12:103–114, 1999.
- [37] W. Schoener. Startclim 2003; teil a: Extreme wetterereignisse und ihre auswirkungen; startclim.1: Qualitätskontrolle und statistische eigenschaften ausgewählter klimaparameter (temperatur, niederschlag, schneehöhe) auf tageswertbasis im hinblick auf extremwertanalysen, 2003.
- [38] Schönwiese, Christian-Dietrich. *Klimatologie*. UTB für Wissenschaft, Nr. 1793. Verlag Eugen Ulmer, Stuttgart, 1994.
- [39] M. Schuepp. Kalender der wetter- und witterungslagen von 1955 bis 1967. *Veröffentlichungen der Schweizerischen Meteorologischen Zentralanstalt*, 11:43, 1968.

- [40] V. P. Starr. *Basic Principles of Weather Forecasting*. Harper Brothers Publishers, 1942.
- [41] R. Steinacker. Eine ostalpine strömungslagenklassifikation. Technical report, Institut für Meteorologie und Geophysik, Universität Innsbruck, 1990.
- [42] F. Steinhauser. Die auswirkung der verschiedene wetterlagen in Österreich. *5.Intern.tagung f. Alp. Met. Bled, Beograd*, pages 99–107, 1962.
- [43] R. Trigo and P. Palutikof. Simulation of daily temperatures for climate change scenarios over portugal: a neural network model approach. *Clim. Res.*, 13:45–59, 1999.
- [44] D. Tyler. On the optimality of the simultaneous redundancy transformations. *Psychometrika*, 47:77–86, 1982.
- [45] H. van den Dool. Searching for analogs, how long must we wait? *Tellus*, A(46):314–324, 1994.
- [46] F. W. von Storch, H. Zwiers. *Statistical Analysis in Climate Research*. Cambridge University Press, 1999.
- [47] H. von Storch, Hewitson B., and L. Mearns. Review of empirical downscaling techniques. In *Regional climate development under global warming. Conference proceedings regclim spring meeting Jevnaker Torbjørnrud Norway*, volume 4, pages 29–46. T. Iversen and B. A. K. Hø and B. A. K. Høøiskar, 2000.
- [48] Hans von Storch, Eduardo Zorita, and Ulrich Cubasch. Downscaling of global climate change estimates to regional scales: An application to iberian rainfall in wintertime. *J. Climate*, 6:11161 – 11171, 1993.
- [49] S. und M. Heimann von Storch, H. Güss. *Das Klimasystem und seine Modellierung. Eine Einführung*. Springer Verlag, 1999.
- [50] von Storch, H., and A. Navarra, editor. *Analysis of Climate Variability: Applications of Statistical Techniques*. Springer Verlag, 1995.
- [51] von Storch, Hans. On the use of "Inflation" in Statistical Downscaling. *J. Climate*, pages 3505–3506, 1999.
- [52] Wasa. Changing waves and storms in the northeast atlantic. *Bull.Amer.Meteorol.Soc.*, 79:741–760, 1998.
- [53] F. Wetterhall. *Statistical Downscaling of Precipitation from Large-scale Atmospheric Circulation: Comparison of Methods and Climate Regions*. Dissertation, Universität Uppsala, 2004.

-
- [54] R. Wilby, T. Wigley, D. Conway, P. Jones, B. Hewitson, J. Main, and D. Wilks. Statistical downscaling of general circulation model output: a comparison of methods. *Water Resources Research*, 34:2995–3008, 1998.
- [55] R.L. Wilby and T.M.L. Wigley. Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography*, 21:530–548, 1997.
- [56] Wilkinson, L., editor. *The Grammar of Graphics*. Springer Verlag, 2005.
- [57] D. Wilks. Multisite downscaling of daily precipitation with a stochastic weather generator. *Clim. Res*, 11:125–136, 1999.
- [58] Daniel S. Wilks. *Statistical methods in the atmospheric sciences*. Academic Press, 2006.
- [59] B. Yambor, W. Draper and R. Beveridge. Analyzing pca-based face recognition algorithms: Eigenvector selection and distance measures. In *Proc. Second Workshop Empirical Evaluation in Computer Vision*, 2000.
- [60] E. Zorita, J. Hughes, D. Lettenmaier, and H. von Storch. Stochastic downscaling of regional circulation patterns for climate model diagnosis and estimation of local precipitation. *J. Climate*, 8:1023–1042, 1991.
- [61] E. Zorita and H. von Storch. The analog method - a simple statistical downscaling. technique: comparison with more complicated methods. *Journal of Climate*, 20:2474–2489, 1999.
- [62] Eduardo Zorita and Fidel González-Rouco. Are temperature-sensitive proxies adequate for north atlantic oscillation reconstructions? *eophysical Research Letters*, 29 (14):48–1 – 48–4, 2002.
- [63] Kharin V. Zorita E. and von Storch H. The atmospheric circulation and sea surface temperature in the north atlantic area in winter: their interaction and relevance for iberian precipitation. *Journal of Climate*, 5:1097–1108, 1992.