

Projektreport
Institut für Meteorologie und Geophysik

"Analogmethode"

Thomas Rupnig B.Sc. (1507989),
Clemens Bauer B.Sc. (1502136)
und Nikolaus Suppan B.Sc. (1204101)

Übungen zu Klimamodelle
Sommersemester 2019

Contents

1	Einführung	2
1.1	Aufgabenstellung	2
1.2	Theorie	2
1.3	Datensätze	3
2	Daten Download	4
3	Preprocessing	4
4	Anwendung der Analog-Methode	5
5	Validierung	8
6	Diskussion	10

1 Einführung

1.1 Aufgabenstellung

Ziel der Übung zu Klimamodellen in diesem Semester war es, numerische Konzepte aus der Klimaforschung zu verstehen, ihren Nutzen zu erkennen und in Folge selbst auf vorhandene Datensets anzuwenden um deren Konzepte und Benutzbarkeit für uns selbst sichtbar und nachvollziehbar zu machen. Im Konkreten war dies die Analogmethode, eine Methode für statistisches Downscaling im Bereich der Klimaprognose.

1.2 Theorie

Das Konzept des **Downscaling** in der Klimaprognose zielt auf eine Übertragung einer großskaligen Klimaprognose (oder Analyse) eines "global circulation model" (large scale - LASC) auf eine kleinskaligere lokale Domäne mit geringeren Gitterpunktabständen (local scale - LOSC) ab. Es ist nicht sinnvoll mit GCM Daten Rückschlüsse auf lokale Klimaveränderungen vorzunehmen, da die Skalengröße dieser Modelle einen "skillful scale" Grenzwert besitzen, welcher auf Größenskalen kleinerer geographischer Räume nicht angewandt werden kann. Dies gilt nicht exklusiv, aber vor allem für inneralpine Räume, stark inhomogene Oberflächen sowie Inseln.

Für dieses Downscaling gibt es zwei grundlegende große Konzepte:

Das **dynamische Downscaling**, bei welchem mit Hilfe physikalischer Grundgleichungen und Parametrisierungen eine Domäne in das globale Modell genestet werden kann und mit aus den großskaligen Modellen entnommenen Randbedingungen, eine Berechnung der lokalen Verhältnisse vorgenommen werden kann. Diese Methode benötigt je nach gewählter Domänengröße sowie Anzahl an regionalen und globalen Modellen eine enorme Rechenleistung und ist sehr abhängig von den gewählten physikalischen Rahmenbedingungen. Weiters ist ein gewisser Bias, welcher von dem Modell das die Randbedingungen stellt übernommen werden kann, nicht auszuschliessen (Ribalaygua et al. 2013).

Die zweite Methode, das **statistische Downscaling** wird oft durch die bereits genannte **Analogmethode** realisiert. Für eine Umsetzung sind einige Voraussetzungen nötig. So sollten die Prediktor-Variablen und die Predikant-Variablen in einer starken Beziehung zueinander stehen (Kovarianz und Zeitstruktur), dafür gibt es standardisierte Methoden. Außerdem sollte der gefundene Zusammenhang auch unter veränderten klimatischen Bedingungen seine Gültigkeit behalten (Schoof 2013). Um den Rechenaufwand zu minimieren können verschiedene Verfahren auf die Reanalysedatensätze aus den GCMs angewandt werden. Eine der meistgenutzten Methoden ist die EOF (empirical orthogonal function) oder PC (principal component) Analyse, bei welcher ein Datensatz durch eine Eigenwertanalyse seiner Kovarianzmatrix in orthogonale Basisfunktionen zerlegt wird, deren Gesamtheit wiederum auf einige wenige beschränkt werden kann, welche den Gesamtdatensatz durch Linearkombinationen wiederum ausreichend umfassend darstellen können. Besagte Basisfunktionen können sowohl Zeitserien

als auch räumliche Muster darstellen, sie sollten die maximale Variabilität des Ursprungsdatensatzes beinhalten (Wilks 2011). Über statistische Methoden kann dann der ähnlichste (analoge) Zustand in einem Vergleichszeitraum in der Vergangenheit gefunden werden. Die zu diesem Zeitpunkt passenden Messdaten eines regionalen Modells werden schlussendlich als kleinräumige Daten zu den ursprünglichen GCM Daten festgelegt und schließen so den Weg von globaler Klimaprognose zum regionalskaligen Zustand (Zorita and Von Storch 1999).

1.3 Datensätze

Um die statistisch klimatologische Analogmethode zum Downscaling zu verwenden, wurde jeder Arbeitsgruppe ein Datensatz eines GCM ("Global Climate Model") zugeordnet.

Der GCM-Reanalyse-Datensatz welcher von uns verwendet wird ist der NCEP-Datensatz, welcher von der US-amerikanischen NOAA (National Oceanic and Atmospheric Administration) betreut wird. Er umfasst globale klimatologische Daten von 1948 bis heute. Die Daten sind mit 4 täglichen Werten und als tägliche Mittelwerte vorhanden. Das mit einem Gitterpunktabstand von 2.5 Grad und 17 Drucklevels.

Zur Validierung wurden für dieses Projekt die Spartacus Daten der ZAMG zur Verfügung gestellt. Dieser beinhaltet Temperatur- und Niederschlagswerte, gegittert und die Fläche Österreichs umfassend.

2 Daten Download

Der NCEP-Datensatz ist über den ftp-Server "ftp.cdc.noaa.gov" der NOAA ESRL Physical Sciences Division (PSD) erhältlich. Für dieses Projekt wurden die Variablen Sea Level Pressure (SLP), relative Humidity (rHum) und specific Humidity (sHum) heruntergeladen. Auf dem ftp-Server werden diese Variablen bereits als tägliche Mittelwerte unter "/Projects/Datasets/ncep.reanalysis.dailyavgs/" angeboten. Für die hier verwendete Analog Methode werden tägliche Mittelwerte benötigt. Für jedes Jahr ist ein netCDF file vorhanden.

3 Preprocessing

Bevor die Analogmethode durchgeführt werden kann, muss der NCEP Datensatz noch bearbeitet werden. Die Beiden Variablen rHum und sHum beinhalten verschiedene Druckniveaus. Um die Datenmenge zu reduzieren, wurde als erstes die hier benötigte 700hPa Fläche mit dem Linux Tool "ncks" ausgeschnitten.

Von hier an wurde das Python Package Xarray für das weitere Preprocessing verwendet. Zu Beginn wurde aus dem Dataset die relevante geographische Region (10°W - 25°E, 32.5°N - 67.5°N) ausgeschnitten. Die Longitude des Datensatzes ist von 0 - 360° definiert. Die Koordinaten müssen also vor dem Zuschneiden um 180° gedreht werden.

```
ds.coords['lon'].data = (ds.coords['lon'] + 180) % 360 - 180
```

Dies wurde von uns zuerst nicht berücksichtigt, was zu einer falsch ausgewählten Region und damit zu falschen Anomalien führte.

Da es für die beiden Variablen rHum und sHum eine vertikale Dimension gibt, muss diese noch, um Inkonsistenzen zu vermeiden, entfernt werden. Hierzu muss die dimension "level" für die einzelnen Variablen und anschließend für das gesamte Dataset entfernt werden.

```
ds['rhum'] = ds['rhum'].squeeze(dim='level')  
ds = ds.drop('level')
```

In den neueren netCDF files (ca. ab dem Jahr 2000) befindet sich eine weitere Variable "time_bnds" in dem Datensatz. Diese muss ebenfalls wie zuvor entfernt werden.

Da diese Variable nicht immer vorhanden ist, wurde ein try/except um den drop-Befehl gelegt.

4 Anwendung der Analog-Methode

Anomalien

Um die Analog Methode anwenden zu können, müssen als erstes genormte Anomalien der Daten berechnet werden. Hierzu wird ein sogenanntes "Rolling Window" verwendet um den Mittelwert und die Standardabweichung zu berechnen. Bei dem "Rolling Window" wird eine neue Dimension im Datensatz erzeugt, welche für jeden Tag aus diesem Tag und jeweils 10 Tage davor und danach besteht.

```
ds_roll = ds.rolling(time=window_size, center=True).construct('window_dim')
```

Es wird für jeden "day of year" (doy) ein Mittelwert über die Räumliche dimension, die window_dim und alle Jahre berechnet.

```
ds_mean = ds_roll.groupby('time.dayofyear').mean(dim=['window_dim','time'])
```

Die Standardabweichung, welche für die Normierung der Anomalien benötigt wird, wird über alle Dimensionen berechnet.

```
ds_std = ds_roll.groupby('time.dayofyear').std(dim=xr.ALL_DIMS)
```

Die Anomalien werden anschließend aus Mittelwert und Standardabweichung folgend berechnet.

```
ds = ds.groupby('time.dayofyear') - ds_mean  
ds = ds.groupby('time.dayofyear') / ds_std
```

Beim bilden der Anomalien sind zwei Probleme aufgetreten. Zum ersten bietet xarray keine vernünftige Methode um Schalttage zu entfernen. Bei dem sortieren nach den day of years gibt es an Schaltjahren eine verschiebung um einen Tag ab dem 29.02. Von uns wurde kein Weg gefunden, den Kalender, welcher der Sortierung zugrunde liegt, zu verändern. Dies führt entsprechend zu Fehlern.

Ein zweites Problem trat beim berechnen der Mittelwerte auf. Wird der Mittelwert direkt von einem Rolling Window Objekt berechnet (noch vor dem construct), so werden alle Windows verworfen, in denen ein Wert fehlt. Wird das Mittel erst danach berechnet, so werden nur die konkreten fehlenden Werte verworfen. Dies hat in erster linie Auswirkungen auf den Beginn und das Ende der Zeitreihe, da dort die Tage vor dem 1. Jänner und nach dem 31.12 nicht existieren. An dieser Stelle soll erwähnt sein, dass Xarray sich in der Version 0.12.3 befindet. Mit solchen Inkonsistenzen ist also allgemein zu rechnen.

In Abbildung 1 sind die Anomalien für einen Beispieltag (5.2.2004) dargestellt.

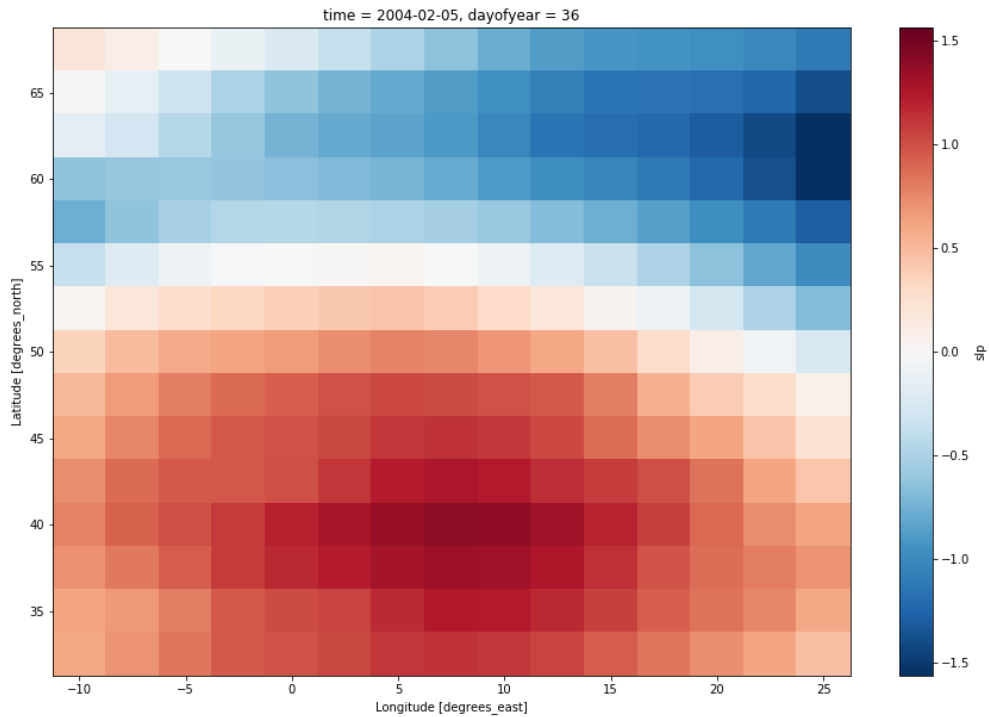


Figure 1: Berechnete Anomalien

Auffinden der Analoga

Zum Auffinden des Analogtages wurde der Zeitraum 1.1.1979 - 31.12.2017 gewählt. Dieser Zeitraum sollte ausreichend groß für die Analogmethode sein und wird später für die Validierung auch vollständig vom Spartakus Datensatz abgedeckt. Nach den Anomalien werden die EOFs berechnet. Mithilfe dieser EOFs können die Analogtage gefunden werden. Zur Berechnung dieser EOFs wurde das python package "eofs" verwendet. Um die Multivariaten EOFs zu berechnen, werden zuerst für jeden doy sogenannte solver-objekte erzeugt. Hierzu wird wieder eine Rolling-Window Dimension erstellt und über alle doys iteriert. Die time und die Rolling-Window Dimension werden anschließend zu einer neuen time Dimension zusammen gefasst.

```
ds_doy = ds_doy.rename("time":"time_old")
ds_doy = ds_doy.stack(time=('time_old','window_dim')).transpose('time', 'lat', 'lon')
ds_doy.coords['time'].attrs['axis'] = 'T'
```

Die Solver werden folgend erzeugt.

```
msolver = MultivariateEof([SLP.values, rHum.values, sHum.values])
```

Der Befehl SLP.values macht aus der variable SLP ein numpy array.

Im Anschluss wird ein konkreter Tag ausgewählt und es wird über alle Tage des betrachteten

Zeitraums iteriert. Um zu vergleichen, wie ähnlich sich die beiden betrachteten Tage sind, wird die Norm für diese Tage berechnet. Hierzu werden die PCs des vorgegebenen Tages und die pseudo-PCs des zu vergleichenden Tages berechnet.

```
pcs = msolver.pcs(npcs=N)
pseudo_pcs = msolver.projectField([SLP, RHUM, SHUM],neofs=N)
norm = np.sum(np.sqrt((pcs - pseudo_pcs)**2),axis=1)
```

Die Variablen SLP, RHUM und SHUM sind wieder numpy arrays und sind die Werte der Variablen an dem zu vergleichenden Tag. Betrachtet werden für den Vergleich der doy selbst für jedes Jahr und die 20 weiteren Tage jedes Jahres, welche sich im Rolling-Window befinden. Der Tag an dem die Norm minimal ist, ist der analoge Tag. Das N gibt an wie viele PCs bzw. pseudo_PCs berechnet werden sollen. Es werden so viele berechnet, sodass die erklärte Varianz 90% beträgt.

```
N = 1
var = 0
while var < .9:
    var = np.sum(my_solver.varianceFraction(neigs=N).data)
    N = N+1
```

Anstatt diese Berechnung mit numpy arrays durchzuführen, wäre es auch möglich das package iris zu verwenden. Der Vorteil von iris wäre, dass man die Dimensionsinformationen nicht verliert. Beim verwenden von numpy muss sehr genau auf die Dimensionen geachtet werden und das Datum des Analog-Tages muss händisch aus einer Liste von Normen rekonstruiert werden.

5 Validierung

Für die Validierung wurde der Spartacus-Datensatz der ZAMG verwendet. Konkret wurden die Variablen Niederschlag (RR), Maximal Temperatur (Tx), Minimal Temperatur (Tn) und Standardized Precipitation Evapotraspiration Index (SPEI) verwendet. Zum Vergleich von zwei Tagen wurde der Pearson Korrelationskoeffizient zwischen zwei Tagen berechnet. Hierzu wurden die Werte eines Tages und einer Variable in eine 1D Reihe gebracht. Im Anschluss wurde der Korrelationskoeffizient berechnet. Die beiden 1D Reihen wurden x und y genannt.

```
cor = np.corrcoef(x,y)
```

Zuerst wurde ein Beispieltag (5.2.2004) ausgesucht. Als Analog-Tag wurde der 6.2.2000 gefunden.

In Abbildung 2 ist die Maximaltemperatur dieser beiden Tage dargestellt. Bereits mit freiem Auge ist eine gewisse Korrelation erkennbar.

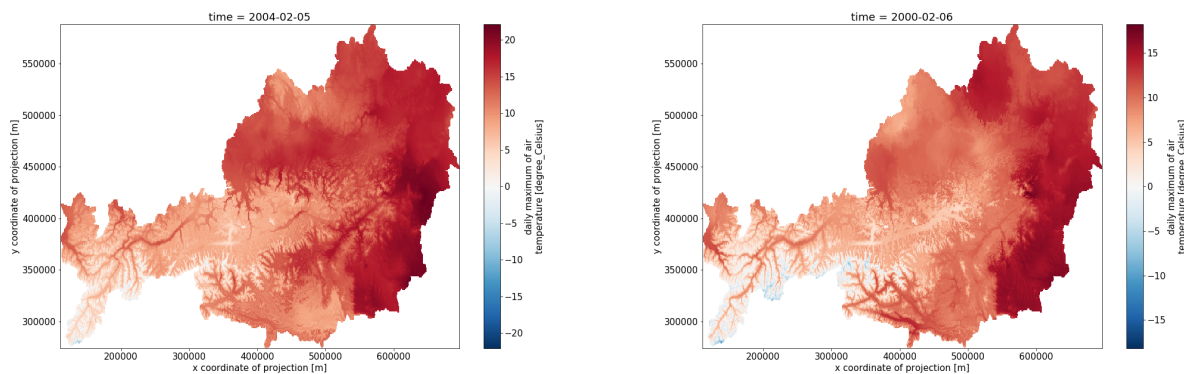


Figure 2: Maximal Temperatur des Spartacus-Datensatzes für den 5.2.2004 (links) und dessen Analogtag 6.2.2000 (rechts)

In Tabelle 1 ist der Korrelationskoeffizient für jede Variable zwischen den beiden Tagen und der mittlere Korrelationskoeffizient für alle Tage, welche mit dem Target Day zum finden des Analogons, verglichen wurden.

Variable	Mittlere Korrelationskoeffizient	Korrelationskoeffizient Analogtag
RR	-0.004	-0.008
Tx	0.162	0.893
Tn	0.278	0.746
SPEI	-0.005	0.347

Table 1: Korrelationskoeffizient zwischen 4.2.2004 und 6.2.2000 und mittlerer Korrelationskoeffizient für gesamten verglichenen Zeitraum

Als nächstes wurden die Korrelationskoeffizienten für jeden Tag mit dessen Analogtag berechnet. Diese Korrelationskoeffizienten wurden für jeden day gemittelt und für alle vorhandenen variablen geplottet. Der Niederschlag ist in Abbildung 3 nicht dargestellt, da es zu viele Fehlwerte bei der Berechnung gab.

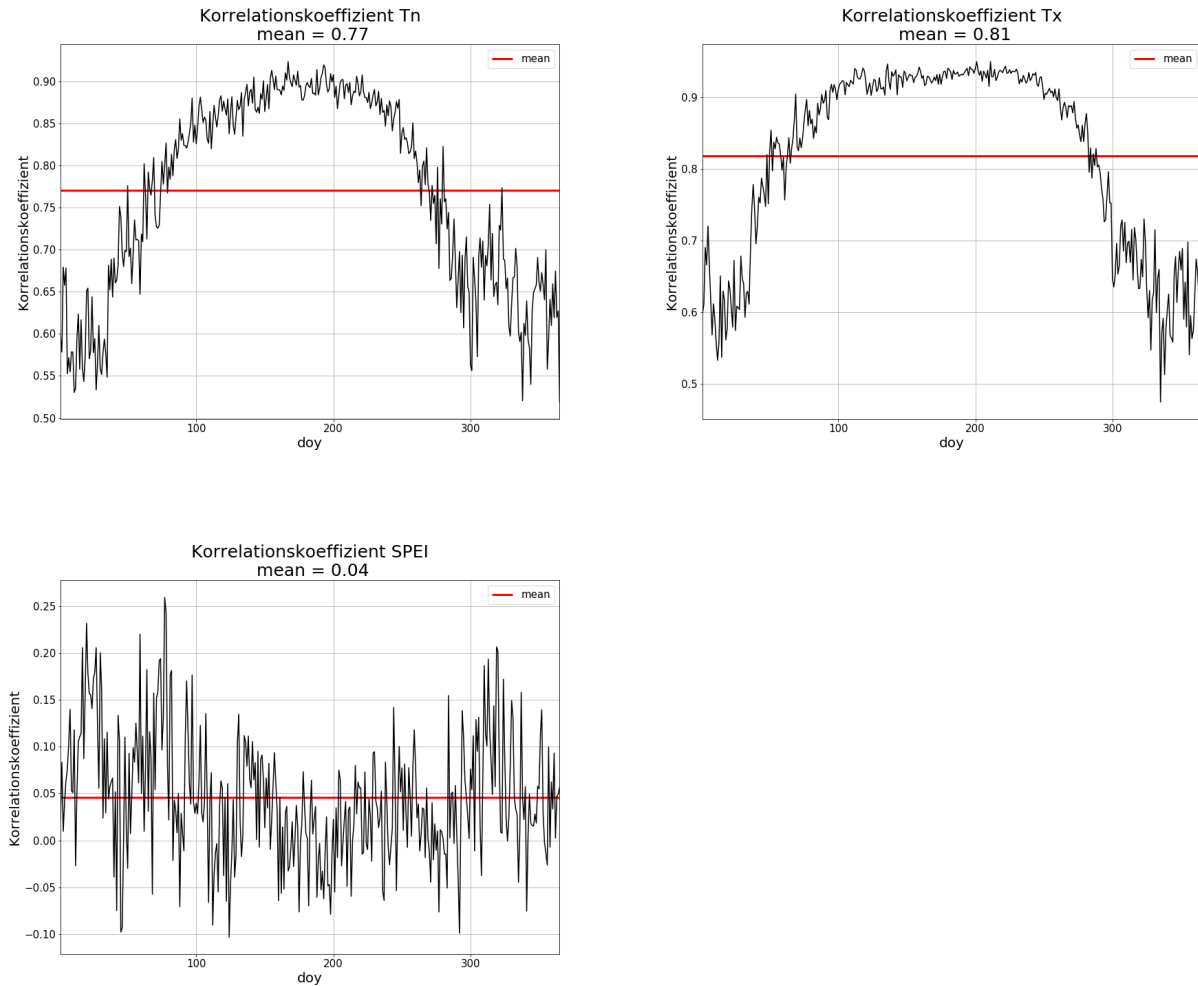


Figure 3: Korrelationskoeffizienten für Minimal Temperatur (links oben), Maximal Temperatur (rechts oben) und SPEI (links unten)

6 Diskussion

Die Validierung hat gezeigt, dass die Analogmethode grundsätzlich funktioniert. Bei den Temperaturen zeigt Tabelle 1 einen klar höheren Korrelationskoeffizienten als bei allen anderen Tagen.

Für den Niederschlag konnte am Analogtag kein besseres Ergebnis erzielt werden. Wobei hier der Korrelationskoeffizient wahrscheinlich keine geeignete Validierungsmethode ist. Sobald sich die Niederschlagsfelder leicht in der Örtlichkeit unterscheiden, kann keine Korrelation mehr gefunden werden. Das selbe gilt für die Variable SPEI. Das der Korrelationskoeffizient für den SPEI am Analogtag so hoch ist, dürfte eher dem Zufall geschuldet sein. In Abbildung 3 erkennt man, dass der SPEI im Mittel an den Analogtagen keine Korrelation aufweist.

Weiters auffallend ist der Abfall der Korrelationskoeffizienten im Winter für die Temperaturen in Abbildung 3. Sichtet man vorangegangene Untersuchungen im Bereich der Analogmethode sind aufgrund des hohen Einflusses der Advektion in den Wintermonaten oft die besten Korrelationen zu dieser Jahreszeit zu finden (bsp. (Ribalaygua et al. 2013)). Wir sind uns bis zuletzt bezüglich dieser Diskrepanz im unklaren. Der gewählte geringe Umfang an Prediktoren könnte ein Grund dafür sein. Für eine exakte Fehlerfindung ist ein mehr an Erfahrung unumgänglich.

Die Anwendung dieser Methode birgt noch einige Verbesserungsmöglichkeiten. Wie bereits oben erwähnt, müssten die Schalttage noch entfernt werden, um Inkonsistenzen zu vermeiden. Außerdem könnte die Analogmethode selbst, wie auch die Validierung mit noch weiteren Variablen durchgeführt werden.

References

- Ribalaygua, J et al. (2013). “Description and validation of a two-step analogue/regression downscaling method”. In: *Theoretical and Applied Climatology* 114.1-2, pp. 253–269.
- Schoof, Justin T (2013). “Statistical downscaling in climatology”. In: *Geography Compass* 7.4, pp. 249–265.
- Wilks, Daniel S (2011). “Principal component (EOF) analysis”. In: *International Geophysics*. Vol. 100. Elsevier, pp. 519–562.
- Zorita, Eduardo and Hans Von Storch (1999). “The analog method as a simple statistical downscaling technique: comparison with more complicated methods”. In: *Journal of climate* 12.8, pp. 2474–2489.

List of Figures

1	Berechnete Anomalien	6
2	Maximal Temperatur des Spartacus-Datensatzes für den 5.2.2004 (links) und dessen Analogtag 6.2.2000 (rechts)	8
3	Korrelationskoeffizienten für Minimal Temperatur (links oben), Maximal Temperatur (rechts oben) und SPEI (links unten)	9