

UNIVERSITÄT WIEN

UE KLIMAMODELLE

PROJEKTARBEIT

Dokumentation - Analogmethode

Author:

Gloria FALTL

Sabrina MARTH

29. Juli 2019

Inhaltsverzeichnis

1	Einleitung	2
2	Preprocessing	2
2.1	Berechnung der Anomalien	3
3	Empirische Orthogonalfunktionen (EOFs)	4
3.1	Berechnung der EOFs	5
3.2	Berechnung der Norm	6
4	Analogmethode	7
5	Validierung	7
6	Anmerkungen und Zusammenfassung	9

1 Einleitung

Das Ziel dieser Projektarbeit ist es eine Analyse mit verschiedenen Reanalysedatensätzen durchzuführen und Unterschiede aufzuzeigen. Dabei soll die sogenannte Analogmethode durchgeführt werden, auf welche in den folgenden Kapiteln näher eingegangen wird.

Zuallererst werden die Reanalysedaten heruntergeladen, danach erfolgt das Preprocessing des Datensatzes. Anschließend werden die Anomalien der vorhandenen Daten berechnet, was für die Ermittlung der EOFs, der PCs und Pseudo-PCs sowie das Finden des Analogons (durch das Minimum der Norm) im weiteren Schritt notwendig ist. Bei der Analogmethode soll zu einem bestimmten Tag der ähnlichste Tag gefunden werden.

Nach der Anwendung dieser Methode werden die Ergebnisse validiert. In diesem Fall wird zur Verifikation der Root Mean Square Error (RMSE) herangezogen.

2 Preprocessing

Die gewünschten Reanalysedaten werden zuallererst heruntergeladen. In diesem Fall handelt es sich um den ERA5-Datensatz von ECMWF, welcher sich von 1979 bis 2018 erstreckt [1]. Da die Daten des globalen Beobachtungsnetzes stündlich zur Verfügung stehen, werden diese auf eine tägliche Auflösung 'gemittelt', danach erfolgt das Preprocessing des Datensatzes.

Für die Analyse werden folgende Parameter verwendet:

- mean sea level pressure (msl)
- relative humidity (r) in 700 hPa
- specific humidity (q) in 700 hPa

Im Preprocessing wird außerdem eine bestimmte geografische Region ausgeschnitten, die folgende Koordinaten aufweist:

$$\lambda = [10^{\circ}W, 25^{\circ}E], \phi = [32.5^{\circ}N, 67.5^{\circ}N]$$

2.1 Berechnung der Anomalien

Die Berechnung der Anomalien erfolgt auf täglicher Basis. Außerdem soll für die weitere Analyse ein größeres Intervall (mit 21 Tagen, 10 Tage vor dem ausgewählten Tag (= Target Day, TD) und 10 Tage danach) erzeugt werden, um mehrere Zustände in die Analyse miteinzubeziehen. Dieser Pool an Tagen wird speziell für die Mittelwertberechnung berücksichtigt. Bei der Ermittlung des Mittelwerts ergeben sich anschließend 365 Mittelwerte, dh. für jeden Tag des Jahres (=Day of year, DOY) wird ein Mittelwert ermittelt. Auch zu erwähnen ist, dass die Methode für jeden Gitterpunkt durchgeführt wird. Die Ergebnisse sind dann mittlere Felder des bestimmten Tages bzw. wird hier der gesamte Pool an Tagen ($DOY \pm d$) berücksichtigt.

Die berechneten Anomalien müssen für die anschließende EOF-Analyse normiert werden, um numerische oder physikalische Probleme zu vermeiden. Für die Normierung kommt die Standardabweichung ins Spiel. Sie wird über jeden Day of Year (DOY) über das Intervall $DOY \pm d$, sowie über alle Gitterpunkte berechnet. Bei der Normierung der Anomalien selbst gilt die Standardabweichung nicht für jeden Gitterpunkt, sondern für die gesamte räumliche Dimension. Zuletzt werden die Anomalien dann durch die Standardabweichung des jeweiligen Tages dividiert.

Abbildung 1 zeigt einen Ausschnitt der Anomalien für einen ausgewählten Gitterpunkt über einen kurzen Zeitraum (1980-1981).

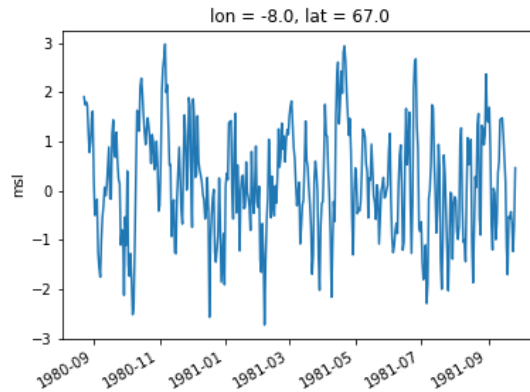


Abbildung 1: Ausschnitt der berechneten Anomalien

3 Empirische Orthogonalfunktionen (EOFs)

Die Idee hinter Empirical Orthogonal Functions (=EOFs) ist die Vereinfachung von Informationen. Mit der EOF-Analyse kann man die räumlich und zeitliche Variation mit nur wenigen EOFs übersichtlich darstellen und eine bestimmte erklärte Varianz ermitteln. Bei der Anwendung wird nach einer Linearkombination der EOFs (welche die maximale Varianz enthält) gesucht. Ein EOF stellt den Eigenvektor eines Eigenwertproblems dar, wobei der zugehörige Eigenwert die Varianz enthält und die Principal Components (PCs) die zeitliche Entwicklung.

Mathematisch gesehen sind EOFs stationäre Funktionen, die die Richtung der Schwankungen ohne eine spezifische Amplitude darstellen, eine Haupteigenschaft davon ist die gegenseitige Orthogonalität. Vorzeichen und Amplitude werden von der PC bestimmt, welche außerdem auch die zeitliche Komponenten darstellt. Mit dieser Überlegung kann der Zustand des Feldes von einer bestimmten EOF zu einer bestimmten Zeit auf einfache Weise beschrieben werden [2].

Im Großen und Ganzen kann man daher den zeitlichen Verlauf der beobachteten Klimaveränderung an N Gitterpunkten von 1979-2018 (ERA5 Datensatz) mit Hilfe von nur wenigen Zeitreihen (Time-Coefficients/Principal Components), die in die EOF-Analyse eingehen, abbilden.

3.1 Berechnung der EOFs

Die Berechnung der EOFs erfolgt unter gleichem Schema wie die der Anomalien. Dabei wird für jeden DOY (mit dem 21 Tage Pool) eine multivariate EOF-Analyse durchgeführt. Aufgabe ist es weiters, die Anzahl der EOFs über eine erklärte Varianz von 90% zu ermitteln. Nach der Berechnung werden die Ergebnisse der EOFs, PCs, Eigenwerte und der erklärten Varianzen für jeden DOY abgespeichert. Die Berechnung erfolgt nach diesem Prinzip:

$$PCS, solver = calculate_EOF(data_inside_time_window) \quad (1)$$

Das Ergebnis der GCM Daten (ERA5- Reanalyse) wird nun in den EOF-Raum projiziert, dies erfolgt über den 'solver' als input. Die resultierenden Zeitkoeffizienten der EOF-Muster, auch Pseudo-PCs genannt, werden ebenfalls für jeden DOY getrennt abgespeichert:

$$Pseudo_PCS = project_field(PCS, solver) \quad (2)$$

Mittels Linearkombination der EOFs mit den zugehörigen Pseudo-PCs kann nun jedes Feld (Gitterpunkte) ermittelt und dargestellt werden. Zur Veranschaulichung sind in Abbildung 2 und 3 EOF-Darstellungen ersichtlich. Es handelt sich hier um jeweils den ersten EOF der Parameter r (relative Feuchte) und msl (mean sea level pressure) an einem beliebig ausgewählten Day of Year (hier: 230. Tag des Jahres).

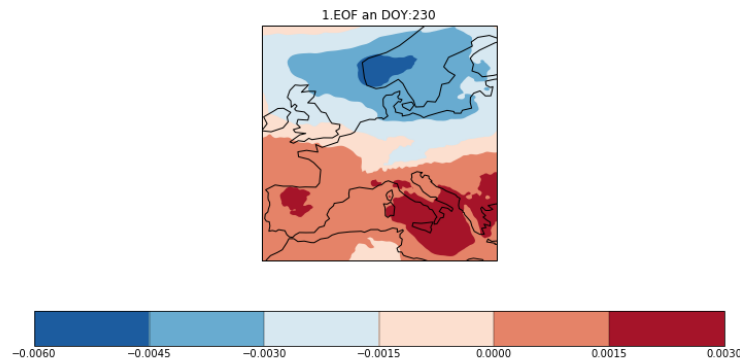


Abbildung 2: Erster EOF von DOY 230 (r)

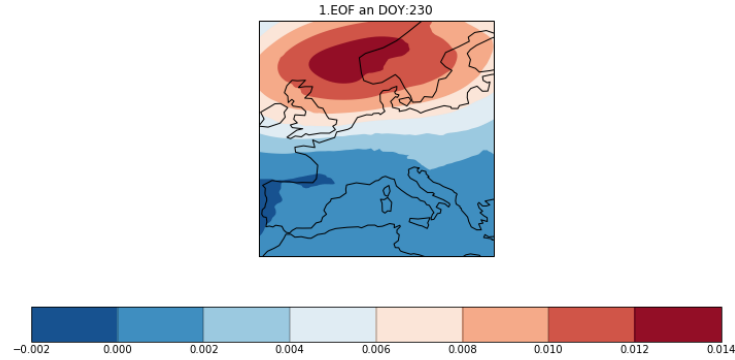


Abbildung 3: Erster EOF von DOY 230 (msl)

Die Ergebnisse der EOF Berechnungen vom 230. Tag des Jahres stellen sehr gut die Verteilung der relativen Feuchte sowie den Druck des Meeresspiegels dar. Der ausgewählte Tag findet im Monat Juli statt, zeigt also ein Sommerbeispiel für den zu betrachteten Ausschnitt. Zu diesem Zeitpunkt ist eine hohe relative Feuchte über Südeuropa und eine geringe Feuchte über Skandinavien vorherrschend. Die Druckverteilung ist genau umgekehrt, hierbei ist hoher Druck im Norden Europas zu finden, während tieferer Druck im südlichen Teil vorliegt.

3.2 Berechnung der Norm

Der weitere Schritt führt zur Berechnung der Norm im EOF-Raum. Diese wird für jeden Target Day (TD) durchgeführt, dh. für jeden Tag wird die Norm zwischen den entsprechenden Pseudo-PCs und den PCs ermittelt. Werden nun mit der erklärten Varianz zB. 10 EOFs berechnet, so ergeben sich für einen bestimmten TD 10 Pseudo-PCs. Diese werden von den PCs abgezogen, quadriert und über alle Terme wird die Summe gebildet, was der Normberechnung entspricht. Es resultiert also für jeden Zeitpunkt, der in den PCs enthalten ($\text{DOY} \pm d \times \text{Anzahl Jahre}$) ist, eine Norm.

$$Norm = calculate_Norm(Pseudo_PCs, PCs) \quad (3)$$

4 Analogmethode

Das Ziel der Analogmethode ist das Finden des ähnlichsten beobachteten großskaligen Musters. Dabei wird die geringste Abweichung der jeweiligen Parameter gesucht. Dies erfolgt über die Suche nach dem Minimum über die berechneten Normen.

$$Analoga = \operatorname{argmin}(Norm)) \quad (4)$$

Die Analoga werden für den besseren Überblick separat als Liste abgespeichert. Ein Ausschnitt davon ist in Abbildung 4 ersichtlich.

init_day	corresponding_day	daynum
1979-01-01	2003-01-04	0
1979-01-02	1995-01-12	1
1979-01-03	1981-01-05	2
1979-01-04	1986-01-07	3
1979-01-05	2010-01-07	4
1979-01-06	1990-01-07	5
1979-01-07	2008-01-08	6
1979-01-08	1988-01-17	7
1979-01-09	2015-01-16	8
1979-01-10	1994-01-05	9
1979-01-11	1994-01-04	10
1979-01-12	2013-01-15	11
1979-01-13	1981-01-05	12
1979-01-14	1986-01-07	13

Abbildung 4: Ausschnitt aus der Liste der gefundenen Analoga

5 Validierung

Um eine gewisse Vorhersagegenauigkeit zu garantieren muss überprüft werden, wie gut die gewählten Transferfunktionen sind. Dies geschieht mit Hilfe der Validierung.

Für die Validierung werden der „Initial day“ und der „Corresponding day“ miteinander verglichen. Damit soll veranschaulicht werden, wie stark die (von der Analogmethode ausgewählten) Tage voneinander abweichen. Als Referenz (Initial Day) werden die Anomalien herangezogen, als Corresponding Day (Analog) jene der PCs. Um starke bzw. weniger starke Abweichungen übersichtlich darzustellen, wird von diesen beiden Parametern der RMSE berechnet. Dieser wird für jeden Tag ermittelt und ist ein Maß für die Genauigkeit. Die Formel dafür lautet:

$$RMSE = \sqrt{\frac{1}{N} \sum day_{corresp} - day_{init}} \quad (5)$$

Der RMSE wichtet große Fehler sehr stark, hierbei ist schnell zu erkennen, an welchen Tagen die Differenz besonders groß bzw. besonders klein ist. Ein RMSE von 0 bedeutet, dass der Unterschied gering ist.

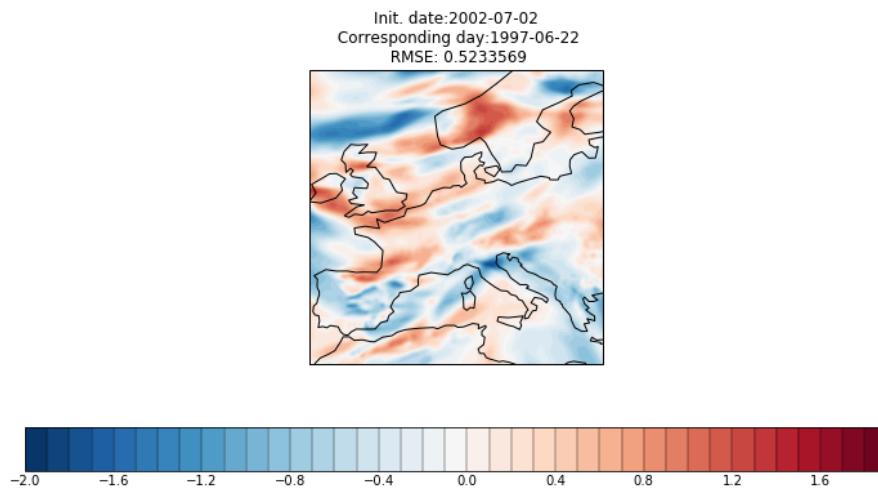


Abbildung 5: Analoga mit dem geringsten RMSE für den Parameter r

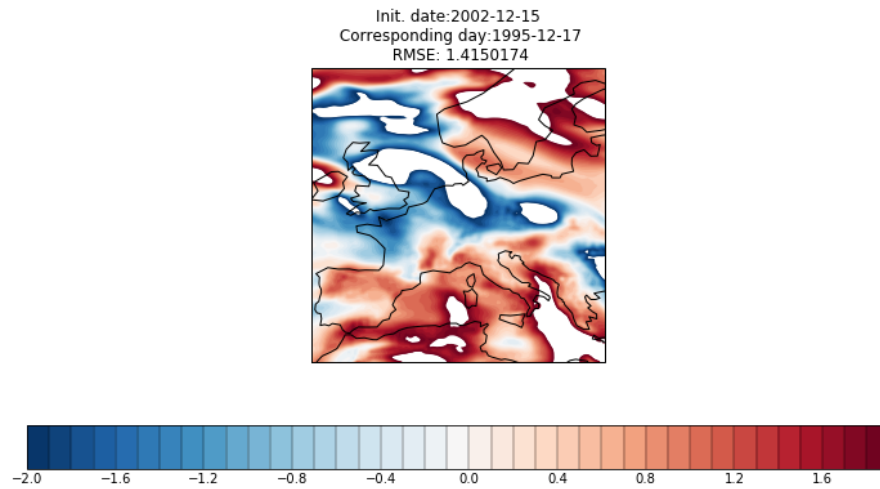


Abbildung 6: Analoga mit dem höchsten RMSE für den Parameter r

In Abbildung 5 und 6 wird jeweils der kleinste sowie der größte RMSE für den Parameter der relativen Feuchte dargestellt. Der kleinste RMSE ergibt sich mit einem mittleren Wert von rund 0.52 aus dem Initial Day des 02.07.2002 und dem Corresponding Day des 10.11.2003. Der größte RMSE mit einem Wert von 1.42 kommt für den Initial Day des 15.12.2002 und für einen Corresponding Day des 17.12.1995 zustande. Die in Abbildung 6 (mit den hohen Fehlerwerten) befindlichen weißen Regionen sind möglicherweise auf lückenhafte Datensätze zurückzuführen.

Interessanterweise ist der größte RMSE für den Wintermonat Dezember berechnet worden, der geringste RMSE für Juni/Juli – dies stellt eine Vermutung auf jahreszeitliche Schwankungen dar.

6 Anmerkungen und Zusammenfassung

Nach anfänglichen Schwierigkeiten die Aufgabenstellung richtig zu interpretieren, stellte das Hauptproblem dieser Arbeit tatsächlich den große Datensatz dar. Beim Anwenden der Analogmethode über den gesamten Zeitraum (1979-2018, tägliche Auflösung) sowie beim Validierungsvorgang wurde die Rechenkapazität des Computers ziemlich unter Beschuss genommen, sodass etliche Stunden vergingen um alle Berechnungen durchzuführen.

In Folge dieser Arbeit wurde mittels der Analogmethode und der Berechnung des RMSE zur Evaluierung erfolgreich eine Gegenüberstellung der ähnlichsten Tage gemacht. Die Validierung hätte auch mit anderen Verifikationsmaße durchgeführt werden können (zB Korrelation,..) , allerdings ist der RMSE bereits ein sehr guter Indikator für die Feststellung der Genauigkeit.

Literatur

- [1] URL: <https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5>
- [2] C. Rasch (2015): Analyse der täglichen Variation von polaren Temperaturprofilen der mittleren Atmosphäre im Winter sowie deren Zusammenhang mit geophysikalischen Feldern der mittleren Breiten, Bachelorarbeit, S. 12