

ZEITREIHEN IN DER METEOROLOGIE

Ein Praktikumsversuch

von

Prof. Dr. Volkmar Wirth
(Universität Mainz)

April 2013

1 Einleitung

Meteorologische Felder sind im allgemeinen Funktionen von Ort und Zeit, also z.B. $T = T(\mathbf{x}, t) = T(x, y, z, t)$ für die Temperatur, wobei $\mathbf{x} = (x, y, z)$ der dreidimensionale Ortsvektor und t die Zeit ist. Formal mathematisch ist T also eine Funktion von vier Veränderlichen, aber zumindest in der klassischen Physik spielt die Zeit eine besondere Rolle, und es ist sinnvoll, die t -Abhängigkeit gesondert zu betrachten. Dies soll im vorliegenden Praktikumsversuch geschehen.

Betrachten wir zunächst den Wert von T an einer festen Stelle \mathbf{x}_0 , etwa an der Wetterhütte im Botanischen Garten der Universität Mainz. Man erhält eine einfache Zeitreihe (siehe Abb. 1)

$$f(t) = T(\mathbf{x}_0, t) . \quad (1)$$

Im allgemeinen liegen nur an endlichen vielen diskreten Zeitpunkten Messungen vor, also

$$f_n = f(t_n) , \quad n = 1, 2, \dots N . \quad (2)$$

Einfache Kenngrößen, welche so eine Zeitreihe charakterisieren, sind ihr Mittelwert sowie die Varianz und Standardabweichung. Will man wissen, ob bestimmte Frequenzen dominieren, so führt man eine Spektralanalyse durch. Dominante und gut verstandene Frequenzen in der Meteorologie sind der Tages- und Jahresgang. Interessanter aber schwieriger ist die Analyse für den elfjährigen Sonnenfleckenzyklus oder die mit den Eiszeiten verbundenen quasi-periodischen Schwankungen. In verrauschten Zeitreihen ist es gar nicht so einfach, das Spektrum abzuschätzen.

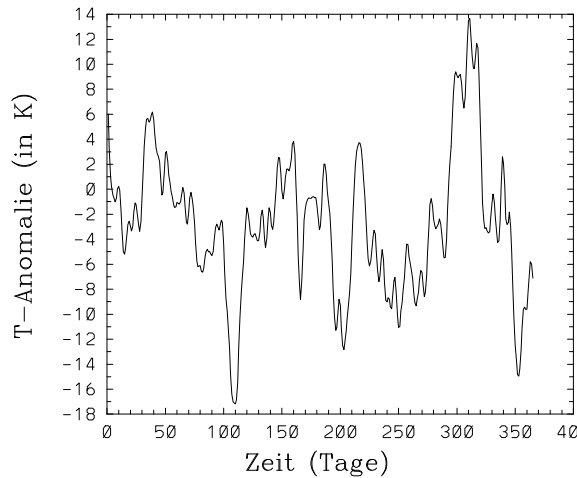


Abbildung 1: Beispiel für eine einfache (univariate) Zeitreihe.

Etwas interessanter wird es, wenn man nicht nur die zeitliche Variation an einem Raumpunkt, sondern wirklich raumzeitliche Variationen untersuchen möchte. Genau dies ist in der Meteorologie oft der Fall. Beispielsweise werden an Radiosondenstationen ein oder zweimal täglich Temperaturprofile gemessen. Man erhält eine Funktion von zwei Veränderlichen

$$g(z, t) = T(x_0, y_0, z, t) , \quad (3)$$

wobei (x_0, y_0) die horizontale Lage der Radiosondenstation beschreibt (siehe Abb. 2). In der Praxis hat man wieder nur Messungen an diskreten Raum- und Zeitpunkten, also

$$g_{kn} = g(z_k, t_n) , \quad k = 1, 2, \dots K , \quad n = 1, 2, \dots N . \quad (4)$$

Liegen kontinuierliche Messungen beispielsweise der bodennahen Temperatur über einem ganzen Gebiet vor, so entspricht das einer Funktion

$$h(x, y, t) = T(x, y, z_0, t) , \quad (5)$$

was sich im diskreten Fall wie $h_{ijn} = h(x_i, y_j, t_n)$ schreiben lässt.

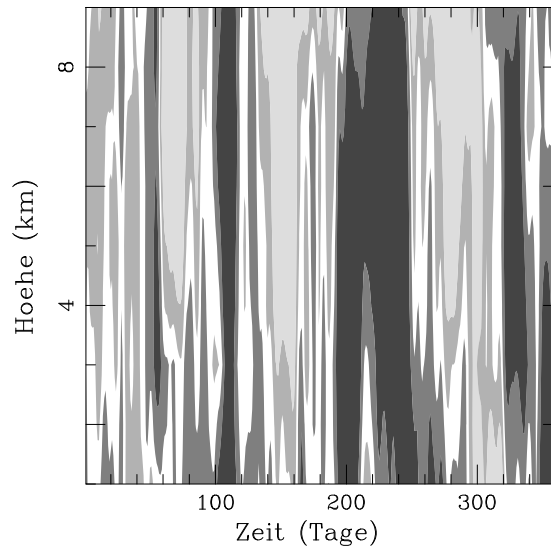


Abbildung 2: Beispiel für eine multivariate Zeitreihe. Die Stärke der Schattierung repräsentiert den Wert der betrachteten Variablen (z.B. Temperatur); die Variable ihrerseits hängt sowohl von der Zeit t als auch von der Höhe z ab.

Solche raumzeitlichen Strukturen in der Meteorologie fasst man am besten als räumliche Strukturen auf (z.B. eindimensionale Profile oder zweidimensionale Karten), welche sich mit der Zeit verändern. Dabei verhalten sich benachbarte Punkte oft recht ähnlich; erst wenn zwei betrachtete Punkte weit genug auseinander liegen, kann man erwarten, dass die Zeitreihen unkorreliert sind. Eine wichtige Aufgabe der meteorologischen Diagnose besteht darin, typische Hauptmuster der Variation zu bestimmen, da diese vermutlich (wenn auch nicht notwendigerweise) mit physikalischen Prozessen verknüpft sind. In einem der vorliegenden Praktikumsversuche wird eine solche Methode (nämlich die der EOF's oder "Empirischen Orthogonalfunktionen") angewandt, wobei auch auf die Schwierigkeit bei der Interpretation hingewiesen wird.

2 Vorbereitung

Zur Vorbereitung auf die Praktikumsaufgaben sollten die folgenden Begriffe nachgeschlagen und verstanden werden:

- Zufallsvariable (diskret und kontinuierlich), Wahrscheinlichkeitsdichtefunktion, Stichprobe, Erwartungswert, Mittelwert, Varianz, Standardabweichung, Kovarianz, Korrelation, Kovarianzmatrix (all dies sowohl für uni- als auch multivariate Zufallsvariablen). *Empfohlene Literatur:* Standardlehrbuch wie z.B. Brandt (1974), für Meteorologen besonders interessant das Buch von von Storch und Zwiers (1999); ferner in Press et al. (1992) die Kapitel 14.0, 14.1 und 14.5 sowie die Routine `pearsn`.

- Telekonnektionen.
Empfohlene Literatur: Kapitel 2.8.7, 2.8.8 und 17.4.1 in Storch und Zwiers (1999).
- Fourierreihe, Fouriertransformation, diskrete Fouriertransformation, Fast Fouriertransformation, Endlichkeit einer Zeitreihe, Nyquistfrequenz.
Empfohlene Literatur: Press et al. (1992), Kapitel 12.0 bis 12.3 (bis zur Subroutine `realft`), speziell die Abbildung 12.2.2 sowie die Routine `realft`.
- Periodogramm, Spektrum, Energiespektrum (engl. “power spectrum”), Energiespektraldichte, gesamte Energy (engl. “total power”), Parseval, Abschätzung der Energiespektraldichte aus endlichen Daten.
Empfohlene Literatur: Kapitel 6 aus dem Onlineskript von D. Hartmann; Press et al. (1992), Kapitel 13.0 und 13.4, speziell die Routine `spctrm`.
- Empirische Orthogonalfunktionen.
Empfohlene Literatur: Kapitel 4 aus dem Onlineskript von D. Hartmann; von Storch und Zwiers (1999), Kapitel 13.0 bis 13.2
- Eigensysteme, Eigenwerte und Eigenvektoren einer symmetrischen Matrix.
Empfohlene Literatur: Press et al. (1992), Kapitel 11.0 und 11.1, besonders die Routine `jacobi`.

Die Vorbereitung auf diesen Versuch wird in der Regel mindestens eine Woche intensiven Studiums in Anspruch nehmen, gegebenenfalls auch mehr. Es ist dabei vor allem wichtig, die grundlegenden Konzepte zu verstehen. Es geht weniger darum, algorithmische Details (z.B. Bestimmung von Eigenwerten, diskrete Fouriertransformation) zu studieren. Insbesondere ist es nicht erforderlich, die Programmcodes in Press et al. im Detail nachzuvollziehen. Online-Ressourcen (Wikipedia) sind erfahrungsgemäß nicht ausreichend für das Bestehen des Vortests. Wenn die vorgeschlagenen Quellen (Lehrbücher) nicht in dem Studentenfach unserer Bibliothek zu finden sind, dann geben Sie bitte umgehend Bescheid.

Es wird vorgeschlagen, dass die numerischen Praktikumsaufgaben mit Hilfe von Routinen aus den “Numerical Recipes“ gelöst werden.¹ Daher empfiehlt sich besonders die Lektüre der oben genannten Kapitel aus diesem Buch.

3 Spektralanalyse einer verrauschten Zeitreihe

In der Datei `timeser1.dat` ist eine Zeitreihe² der Länge $N = 4100$ gegeben. In der ersten Zeile dieser Datei steht die Anzahl N der Zeitpunkte (d.h. die Länge der Zeitreihe), die folgenden Zeilen stellen dann die Zeitreihe selber dar. Beachte, dass die Werte der Zeitreihe *zeilenweise* eingelesen werden müssen. Eine Möglichkeit, diese Zeitreihe in das Fortranprogramm einzulesen sieht wie folgt aus:

```
open(unit=10,file='timeser1.dat',status='old')
read(10,*) nn
allocate( ser(nn) )
read(10,*) ser
close(10)
```

¹Die benötigten Routinen liegen im Institut in elektronischer Form vor (bitte wenden Sie sich an Dr. Joachim Eichhorn).

²Es handelt sich um synthetische, d.h. künstlich erzeugte Daten.

- (a) Wie groß sind Mittelwert und Varianz dieser Zeitreihe?
- (b) Stelle die gesamte Zeitreihe bzw. nur ihr erstes Zehntel grafisch dar. Offenbar handelt es sich um ein stark verrauschtes Signal, bei dem keine spezielle Frequenz ins Auge springt. Ziel der Aufgabe ist es, herauszufinden, ob dem kontinuierlichen Untergrund bestimmte diskrete Frequenzen überlagert sind und, wenn ja, wieviele und bei welcher Frequenz.
- (c) Führe eine schnelle Fouriertransformation (kurz: FFT, für “Fast Fourier Transform“) durch (Routine `realft` aus den “Numerical Recipes“) und stelle das Energiespektraldichte $P(\nu)$ grafisch dar. In welchem Bereich ist die Frequenz ν definiert? Verifiziere explizit, dass das Integral über die Energiespektraldichte gleich der gesamten Varianz der Zeitreihe ist (dies ist der Test ob alle Normierungen korrekt durchgeführt wurden).
- (d) Trage $\nu P(\nu)$ gegen $\ln \nu$ auf. Diese Darstellung hat den Vorteil, dass den (meist interessierenden) niedrigen Frequenzen “mehr Raum zugestanden wird“ und gleichzeitig die ursprüngliche Eigenschaft erhalten bleibt, dass das Integral unter der Fläche die Varianz im entsprechenden Frequenzintervall beschreibt. Letzteres ist der Fall, weil

$$\int_{\nu_1}^{\nu_2} P(\nu) d\nu = \int_{\ln \nu_1}^{\ln \nu_2} \nu P(\nu) d \ln \nu \quad (6)$$

gilt (zeige dies!).

In beiden Darstellungen hat das Spektrum einige Spitzen, die nahelegen, dass es sich um die gesuchten Frequenzen handelt. Doch Vorsicht!

(e) In den Dateien `timeser2.dat` und `timeser3.dat` finden sich zwei weitere *Realisierungen der gleichen Zeitreihe*; das bedeutet, dass in diesen beiden Zeitreihen das Signal (d.h. die gesuchte(n) Frequenz(en)) unverändert bleibt, während das zufällige Rauschen durch eine neue Realisierung gegeben ist. Untersuche die beiden neuen Zeitreihen nach den gleichen Methoden wie oben. Beachte, dass die entsprechenden Spektren wieder einige Spitzen enthalten, teilweise jedoch an anderen Stellen als vorher. Es ist also nicht ganz klar, welche der Spitzen ein Signal darstellen und welche nur durch den verrauschten und schlecht geschätzten Untergrund zustandekommen.

(f) Führe eine Spektralanalyse mit Hilfe der Routine `spctrm` aus den “Numerical Recipes“ durch. Die Quintessenz bei dieser Routine besteht darin, dass die betrachtete Zeitreihe in mehrere kleinere Segmente unterteilt wird. Diese Segmente dürfen teilweise überlappen, wenn man gleichzeitig ein Datenfenster verwendet. Faktisch schätzt man das Spektrum als Mittelwert aus den Spektren der einzelnen Segmente. Diese Mittelung erhöht die Signifikanz der Werte, was insbesondere beim kontinuierlichen Untergrund notwendig ist. (Hinweis: wähle für die Anzahl der überlappenden Segmente $2k = 30$ gemäß der Notation der Routine `spctrm`.) Es bewährt sich auch hier wieder die Auftragung von $\nu P(\nu)$ gegen $\ln \nu$. Verifiziere wieder, dass das Integral des so geschätzten Spektrums gleich der mittleren Varianz der Zeitreihe ist (gemäß der Konvention am Beginn des Kapitels 13.4 der “Numerical Recipes“).

Beachte, dass die Schätzung der Energiespektraldichte mit Hilfe von `spctrm` bei den drei gegebenen Zeitreihen konsistentere Ergebnisse liefert als die schlichte FFT, und es erscheint jetzt leichter, den kontinuierlichen Untergrund vom Signal (d.h. einzelnen Spitzen) zu unterscheiden. Welche diskreten Frequenzen kannst Du schließlich identifizieren?

4 EOF-Analyse (Hauptkomponentenanalyse)

Die EOF-Analyse ist ein wichtiges Hilfsmittel in der Meteorologie. In der physikalischen Literatur ist diese Methode üblicherweise unter der Bezeichnung Hauptkomponentenanalyse bekannt (engl. “principal component analysis“, PCA).

In der Datei `times2.dat` sind neun Zeitreihen³ der Länge $N = 365$ gegeben. Diese Zeitreihen können wie folgt in das Fortranprogramm eingelesen werden:

```
open(unit=10,file='times2.dat',status='old')
do  itime=1,365
  read(10,'(9f9.4)') (times(itime,iz),iz=1,9)
enddo
close(10)
```

- (a) Stelle diese neun Zeitreihen je einzeln als Funktionsplot dar.
- (b) Fasse nun den zweiten Index (d.h. `iz`) als Höhenkoordinate auf. Das soll bedeuten, dass die neun Zeitreihen den zeitlichen Verlauf der gleichen meteorologischen Größe T in unterschiedlicher Höhe darstellen, also etwa die Temperatur in $z = 0$ km, $z = 1$ km, ... bis $z = 8$ km Höhe. Wir haben es also mit 365 Profilmessungen zu tun, wobei jedes Profil aus neun Werten besteht. Im kontinuierlichen Fall entspräche dies $T(z, t)$.

Stelle diese Zeitreihen als Konturenplot dar, wobei die Zeit t nach rechts und die Höhe z nach oben aufzutragen ist.

- (c) Untersuche zunächst Telekonnektionen. Fange mit der Telekonnektion des Punktes $z_1 = 0$ km an. Berechne dazu den Korrelationskoeffizienten von z_1 mit allen z_i ($i = 1, 2, \dots, 9$) und stelle das Ergebnis als Funktion von z in Form eines Funktionsplots dar. Wiederhole dies für alle z_k ($k = 2, 3, \dots, n$). Insgesamt sind also neun solcher Telekonnektionsplots gefragt.

Mache Dir klar, dass der oberste und der unterste Punkt des Profils praktisch unkorreliert sind (dies wird im Folgenden noch wichtig).

- (d) Führe eine EOF-Analyse durch. Bestimme insbesondere die neun EOF's und stelle sie grafisch als Funktion von z dar. Sortiere dabei die EOF's nach der durch sie beschriebenen Varianz. Bestimme auch die Zeitreihen der dazugehörigen Entwicklungskoeffizienten und stelle diese grafisch dar. Welcher Anteil der Varianz wird durch die einzelnen EOF's beschrieben? Zeige, dass im vorliegenden Beispiel ein großer Teil der gesamten Varianz durch die beiden führenden EOF's beschrieben wird.

Hinweis: Zum Berechnen der Eigenwerte und Eigenvektoren der Korrelationsmatrix eignet die die Routine `jacobi` aus den “Numerical Recipes“.

- (e) Erstelle eine reduzierte Zeitreihe $\hat{T}(z, t)$, die durch die beiden führenden EOF's und ihre Entwicklungskoeffizienten beschrieben wird. Stelle diese reduzierte Zeitreihe wieder in Form eines Konturenplots $\hat{T}(z, t)$ dar. Vergleiche mit $T(z, t)$ aus Teilaufgabe (b): Die beiden Plots sind sehr ähnlich! Beachte, dass die ursprüngliche Zeitreihe 9×365 Zahlen enthält, während die reduzierte Zeitreihe mit Hilfe von 2×365 Zahlen (plus der einmal berechneten EOF's) beschrieben werden kann. Man hat im vorliegenden Beispiel eine erhebliche und nahezu verlustfreie Datenreduktion erreicht!

- (f) Erstelle eine noch stärker reduzierte Zeitreihe $\hat{T}(z, t)$, die jetzt durch einen einzigen, und zwar den führenden EOF (und ihren Entwicklungskoeffizienten) beschrieben wird. Stelle diese

³Auch hier handelt es sich wieder um synthetische Daten.

reduzierte Zeitreihe wieder in Form eines Konturenplots $\hat{T}(z, t)$ dar. Vergleiche mit $T(z, t)$ aus Teilaufgabe (b): jetzt sind schon gewisse Abweichungen in den Details erkennbar, wenn auch die wesentlichen Elemente nach wie vor ganz gut wiedergegeben werden.

(g) Es ist verlockend, bei der Interpretation so zu tun als ob der führende EOF die wesentlichen Elemente des raumzeitlichen Verhalten wiedergibt. Für manche Zwecke mag dies richtig sein, aber es birgt auch Gefahren. Beispiel: Berechne den Korrelationskoeffizienten zwischen z_1 und z_9 für die reduzierte Zeitreihe $\hat{T}(z, t)$ aus Teilaufgabe (f). Die Schlussfolgerung: Boden (z_1) und Tropopause (z_9) sind perfekt korreliert?! Offensichtlich ist dies ein reines Artefakt der EOF-Analyse und der Reduktion auf den führenden EOF, weil in der ursprünglichen Zeitreihe diese beiden Punkte praktisch unkorreliert waren! Diese Angelegenheit hat in den vergangenen Jahren die Meteorologen an der Nase herumgeführt. Es ging dabei um die Frage der Nordatlantikoszillation (NAO) und der Arktischen Oszillation (AO). Mehr dazu ist in dem Artikel von Ambaum et al. (2001) nachzulesen.

5 Anmerkung zur Auswertung

Die Auswertung darf gerne in Zweiergruppen durchgeführt werden. Dabei sollen möglichst beide Studierende gemeinsam alle Aufgaben bearbeiten. Allerdings reicht es aus, wenn nur eine Auswertung angefertigt wird. In diesem Fall soll am Ende der Auswertung angegeben werden, wer welchen Teil aufgeschrieben hat.

6 Literatur

Ambaum, M. H. P., B. J. Hoskins, and D. B. Stephenson, 2001: Arctic Oscillation or North Atlantic Oscillation? *J. Climate*, **14**, 3495–3507.

Brandt, S., 1974: *Datenanalyse*. Bibliographisches Institut, 464 pp.

Encyclopedia of Atmospheric Sciences: Stichwort “Data Analysis“, bestehend aus den beiden Abschnitten “Empirical Orthogonal Functions and Singular Vectors“ und “Time Series Analysis“.

Hartmann, D: *Objective Analysis*, Course Notes by Prof. D. Hartmann, University of Washington, erhältlich online unter <http://www.atmos.washington.edu/~dennis/> (→ classes taught → ATMS552: Objective Analysis).

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1992: *Numerical Recipes. The Art of Scientific Computing*. Cambridge University Press, 2nd edition, 818 pp.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1996: *Numerical Recipes in Fortran 90. The Art of Parallel Scientific Computing*. Cambridge University Press, 1486 pp.

H. von Storch and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 513 pp.