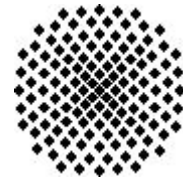




Universität Stuttgart
Geodätisches Institut



Modellierung stationärer Prozesse zur Analyse von Residuen

Studienarbeit im Studiengang
Geodäsie und Geoinformatik
an der Universität Stuttgart

Julian Daubrawa

Stuttgart, Februar 2007

Betreuer:

Dipl.-Ing. Oliver Baur
Universität Stuttgart

Prüfer:

Prof. Dr.-Ing. Nico Sneeuw
Universität Stuttgart

Selbstständigkeitserklärung

Hiermit erkläre ich, Julian Daubrawa, dass ich die von mir eingereichte Studienarbeit zum Thema

Modellierung stationärer Prozesse zur Analyse von Residuen

selbstständig verfasst und ausschließlich die angegebenen Hilfsmittel und Quellen verwendet habe.

Datum, Ort: _____

Unterschrift: _____

(Julian Daubrawa)

Vorwort

Auf dieser Seite möchte ich allen danken, die durch ihre Unterstützung, fachlicher und persönlicher Art, zum Gelingen dieser Studienarbeit beigetragen haben.

Herrn Dipl.-Ing Christian Siemes (Institut für Theoretische Geodäsie, Universität Bonn) für die Hilfe bei der Behebung algorithmischer Schwierigkeiten.

Herrn Prof. Dr.-Ing. Nico Sneeuw (Geodätisches Institut, Universität Stuttgart) für die Unterstützung und Betreuung dieser Studienarbeit.

Ganz besonders möchte ich Herrn Dipl.-Ing. Oliver Baur danken, der immer Zeit für mich hatte und mit dem ich manch aufschlussreiche Diskussion geführt habe.

Zusammenfassung

Modellierung stationäre Prozesse zur Analyse von Residuen

Die vorliegende Studienarbeit beschäftigt sich mit der Identifizierung und Berechnung stationärer Prozesse und Filter zur Analyse von Zeitreihen. Dazu werden Eigenschaften wie die Autokorrelationsfunktion (ACF), die partielle Autokorrelationsfunktion (PACF) und die spektrale Leistungsdichte (PSD) verwendet.

Es werden sowohl auto-regressive (AR) und moving average (MA) Prozesse, als auch deren Kombination (ARMA) behandelt. Mit diesen Prozessen wird jeweils eine Sequenz zeitlich korrelierter Werte generiert, welche als Beobachtungsrauschen in die GOCE (Gravity and Steady-State Ocean Circulation Explorer) Datenanalyse einfließen. Für die Datenanalyse wird nur die radiale SGG (Satellite Gravity Gradiometry) Komponente verwendet. Mit diesen Daten wird eine sphärische harmonische Analyse zur Bestimmung der Koeffizienten des Erdschwerefeldes bis zu Grad und Ordnung 50 durchgeführt. Das Simulationsszenario überdeckt eine Periode von 30 Tagen mit einer Abtastrate von 5 Sekunden. GOCE ist ein Projekt der ESA (European Space Agency), das dazu dient, das Schwerefeld der Erde hochgenau zu bestimmen. Vertiefende Ausführungen finden sich beispielsweise in ESA [1999] und Müller [2001].

Die erhaltenen Verbesserungen, die aus einer Ausgleichung nach der Methode der kleinsten Quadrate resultieren, werden auf ihre stochastischen Eigenschaften hin untersucht. Die zentrale Frage lautet, ob sich die Prozesse, mit denen die fehlerfreien Beobachtungen verrauscht wurden, in den Verbesserungen wiederfinden. Hierfür werden verschiedene Algorithmen zur Berechnung der Prozesse aus Zeitreihen vorgestellt und implementiert.

Sind die stochastischen Eigenschaften der Residuen modelliert, so kann mit ihnen eine Dekorrelation der Beobachtungen durchgeführt werden. Die dekorrelierten Beobachtungen sind dann im Idealfall nur noch mit einem weißen Rauschen verfälscht.

Die Aufgabe dieser Studienarbeit bestand nicht darin, die Ausgleichung der Erdschwerefeldkoeffizienten zu verwirklichen.

Schlüsselwörter:

- stationäre Prozesse
- Dekorrelation
- spektrale Leistungsdichte
- Autokorrelationsfunktion
- partielle Autokorrelationsfunktion

Abstract

Modelling of Stationary Processes for the Analysis of Residuals

This study report deals with the identification and computation of stationary processes and filters for time series analysis. For this purpose, some properties of time series such as the autocorrelation function (ACF), the partial autocorrelation function (PACF) and the power spectral density (PSD) are addressed.

Both autoregressive (AR) and moving average (MA) processes are discussed as well as their combination (ARMA). For each of these processes, a sequence of time-correlated values is computed, which serve as observation noise models for GOCE (Gravity and Steady State Ocean Circulation Explorer). With this data, the spherical harmonic analysis of the Earth gravity field coefficients up to degree and order 50 is computed. The simulation scenario covers 30 days with 5 seconds sampling. GOCE is a satellitemission of the ESA (European Space Agency), providing high-accurate and high-resolution gravitational measurements. Further information can be found in e.g. ESA [1999] and Müller [2001].

The prior objective is to model stochastic properties of least-squares residuals. In this context, it has to be pointed out, whether these properties fit to that ones of the observation noise models. For this task, several algorithms for the modelling of stationary processes are presented and implemented.

Well-known stochastic properties allow the decorrelation of observations. Ideally, the decorrelated observations are contaminated with white noise only.

Keywords:

- stationary processes
- decorrelation
- power spectral density
- autocorrelation
- partial autocorrelation

Inhaltsverzeichnis

Abbildungsverzeichnis	VII
Tabellenverzeichnis	IX
1 Einführung	1
2 Signaltheoretische Grundlagen	2
2.1 Prozess	2
2.2 Stationarität	2
2.3 Autokorrelationsfunktion (ACF)	3
2.4 Partielle Autokorrelationsfunktion (PACF)	4
2.5 Leistungsspektraldichte (PSD)	6
3 Stationäre Prozesse	9
3.1 AR-Prozesse	9
3.1.1 Algorithmus	9
3.1.2 Berechnung der Prozessparameter	10
3.1.3 Beispiel	10
3.2 MA-Prozesse	12
3.2.1 Algorithmus	12
3.2.2 Berechnung der Prozessparameter	13
3.2.3 Beispiel	14
3.3 ARMA-Prozesse	14
3.3.1 Algorithmus	14
3.3.2 Berechnung der Prozessparameter	16
3.3.2.1 Erstes Durbin-Verfahren	16
3.3.2.2 Zweites Durbin-Verfahren	17
3.3.3 Beispiel	18
4 Residualanalyse	20

4.1	Grundlagen	20
4.2	Weißes Rauschen als Fehlermodell	21
4.2.1	Ergebnisse	21
4.3	AR[15]-Prozess als Fehlermodell	23
4.3.1	Ergebnisse	23
4.4	MA[15]-Prozess als Fehlermodell	27
4.4.1	Ergebnisse	27
4.5	ARMA[15,15]-Prozess als Fehlermodell	30
4.5.1	Ergebnisse	30
5	Dekorrelation	34
5.1	Grundlagen	34
5.2	AR-Prozess	36
5.3	MA-Prozess	37
5.4	ARMA-Prozess	38
6	Fazit	39
	Literaturverzeichnis	40

Abbildungsverzeichnis

2.1	ACF von farbigem Rauschen	4
2.2	ACF von weißem Rauschen	4
2.3	PACF von farbigem Rauschen	6
2.4	PACF von weißem Rauschen	6
2.5	PSD von weißem Rauschen	7
2.6	PSD von farbigem Rauschen	8
3.1	PACF eines AR[15]-Prozesses	10
3.2	ACF eines MA[15]-Prozesses	12
3.3	ACF eines ARMA[15,15]-Prozesses	15
3.4	PACF eines ARMA[15,15]-Prozesses	16
4.1	PSD der Residuen (weißes Rauschen)	22
4.2	ACF der Residuen (weißes Rauschen)	22
4.3	PSD der Residuen (AR)	23
4.4	ACF der Residuen (AR)	24
4.5	PACF der Residuen (AR)	25
4.6	Differenzen (Soll-Ist) in den Koeffizienten (AR)	25
4.7	PSD der verbleibenden Sequenz (AR)	26
4.8	ACF der verbleibenden Sequenz (AR)	26
4.9	PSD der Residuen (MA)	27
4.10	ACF der Residuen (MA)	28
4.11	PACF der Residuen (MA)	28
4.12	Differenzen (Soll-Ist) in den Koeffizienten (MA)	29
4.13	PSD der verbliebenen Sequenz (MA)	29
4.14	ACF der verbliebenen Sequenz (MA)	30
4.15	PSD der Residuen (ARMA)	31
4.16	ACF der Residuen (ARMA)	31
4.17	PACF der Residuen (ARMA)	32

4.18	Differenzen (Soll-Ist) in den Koeffizienten (ARMA)	32
4.19	PSD der verbleibenden Sequenz (ARMA)	33
4.20	ACF der verbleibenden Sequenz (ARMA)	33
5.1	Fehlerkurven ohne Dekorrelation	36
5.2	Fehlerkurven (AR)	36
5.3	Fehlerkurven (MA)	37
5.4	Fehlerkurven (ARMA)	38

Tabellenverzeichnis

3.1	Ergebnisse der AR-Koeffizientenrückrechnung mit Zeitreihenlänge 1000	11
3.2	Ergebnisse der AR-Koeffizientenrückrechnung mit Zeitreihenlänge 10000	11
3.3	Ergebnisse der AR-Koeffizientenrückrechnung mit Zeitreihenlänge 100000	11
3.4	Vergleich der stochastischen Eigenschaften der Rauschen (AR)	12
3.5	Ergebnis der MA-Koeffizientenrückrechnung	14
3.6	Vergleich der stochastischen Eigenschaften der Rauschen (MA)	14
3.7	Ergebnisse der ARMA-Koeffizientenberechnung	18
3.8	Ergebnisse der ARMA-Koeffizientenberechnung	18
3.9	Vergleich der stochastischen Eigenschaften der Rauschsequenzen	19
3.10	Vergleich der stochastischen Eigenschaften der Rauschsequenzen	19
4.1	Koeffizienten AR[15]-Prozess	23
4.2	Koeffizienten MA[15]-Prozess	27

Kapitel 1

Einführung

Diese Studienarbeit hat die Berechnung digitaler Filter bzw. die Berechnung verschiedener Prozesse zur Analyse der stochastischen Eigenschaften von Zeitreihen zur Aufgabe. Beobachtete Zeitreihen weisen im Allgemeinen farbiges (oder korreliertes) Rauschen auf. Unter bestimmten Voraussetzungen können Prozesse definiert werden, mit denen sich diese Korrelationen herausfiltern lassen, wodurch das Rauschverhalten auf weißes (oder unkorreliertes) Rauschen zurückgeführt werden kann. Eine ausführliche Beschreibung findet sich in Schuh [1996] und Schuh [2003].

Das folgende Kapitel befasst sich mit signaltheoretischen Grundlagen. Im speziellen mit den Voraussetzungen für die Formulierung stationärer Prozesse und deren Eigenschaften, welche die Identifizierung der Prozesse unterstützen. Kapitel 3 widmet sich der Definition stationärer Prozesse. Es werden Verfahren zur Berechnung der Prozesse aus Zeitreihen vorgestellt und auf kleine Beispiele angewendet. In Kapitel 4 werden diese Verfahren angewendet auf Residuen, die aus einer Ausgleichung der Erdschwerefeldkoeffizienten nach der Methode der kleinsten Quadrate resultieren. In diese Ausgleichung fließen simulierte fehlerfreie Daten ein. Diese werden überlagert mit Beobachtungsruschen, das vorgegebenen stochastischen Eigenschaften unterliegt. Nach der Ausgleichung werden die stochastischen Eigenschaften der Residuen untersucht. Kapitel 5 beschäftigt sich mit der Dekorrelation. Mit diesem Verfahren kann die Schätzung der Erdschwerefeldkoeffizienten verbessert werden.

Die zentrale Frage ist, ob die Residuen die gleichen stochastischen Eigenschaften wie die Beobachtungen aufweisen. Eine weitere Frage ist, ob sich die Ergebnisse der Schätzung der Erdschwerefeldkoeffizienten verbessert, wenn mit den stochastischen Eigenschaften der Residuen eine Dekorrelation der Beobachtungen durchgeführt wird.

Kapitel 2

Signaltheoretische Grundlagen

2.1 Prozess

Ein Prozess ist eine Abfolge von Zuständen. Die Abfolge kann definiert bzw. mathematisch eindeutig beschreibbar oder zufällig sein. Ein Prozess entspricht folglich einer Zeitreihe. Eine Reihe von Beobachtungen ist ein Prozess, da die Werte verschiedene Zustände der Messgröße zu verschiedenen Zeitpunkten sind.

2.2 Stationarität

Im Folgenden bezeichnet t den Zeitpunkt, x_t den Wert der Zeitreihe zum Zeitpunkt t und ϵ_t den Wert des weißen Rauschens zum Zeitpunkt t .

Ein Prozess heißt stationär, wenn folgende Bedingungen erfüllt sind

- (I) $E(x_t^2) < \infty \quad \forall t \in Z$
- (II) $E(x_t) = \mu \quad \forall t \in Z$
- (III) $Cov(x_{t+h}, x_t) = Cov(x_h, x_0) \quad \forall t, h \in Z.$

Die erste Bedingung bedeutet, dass der Erwartungswert der Quadrate der Prozesszustände eine endliche Zahl ist. Die zweite Bedingung sagt aus, dass der Erwartungswert des Prozesses konstant über die Zeit ist. Und die dritte Bedingung sagt aus, dass die Autokovarianz zwischen zwei Zeitpunkten nur von deren Abstand, nicht aber von den Zeitpunkten selbst abhängt. Darüber hinaus ist ein Prozess stationär, wenn seine Wahrscheinlichkeitsdichteverteilung zeitunabhängig ist.

Den Abstand zwischen zwei Zuständen, in der dritten Bedingung mit h bezeichnet, nennt man auch Lag.

2.3 Autokorrelationsfunktion (ACF)

Die ACF zum Lag h beschreibt die Korrelation zwischen den Zeitreihenwerten x_t und x_{t+h} . Ein Lag macht demnach deutlich um wie viele Stellen eine Zeitreihe gegenüber sich selbst verschoben ist. Sei die Zeitreihe $x_t = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, dann sehen die beiden Zeitreihen x_t und x_{t+h} mit Lag $h = 3$ folgendermaßen aus:

$$\begin{array}{rcccccccccccc} x_t & : & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ x_{t+3} & : & & & & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10, \end{array}$$

Ist die Zeitreihe nicht verschoben, wird dies auch als Lag 0 bezeichnet.

Zur Berechnung der ACF wird zunächst die Autokovarianzfunktion (Teusch [2006]) berechnet mit

$$\text{Cov}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}) \quad -n < h < n,$$

wobei $\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t$ den Mittelwert der Zeitreihe bezeichnet.

Die ACF ergibt sich dann aus

$$\text{Cor}(h) = \frac{\text{Cov}(h)}{\text{Cov}(0)}.$$

Die Autokorrelationsfunktion ist demnach die normierte Autokovarianzfunktion. Für weißes Rauschen gibt es sowohl bei der Autokovarianzfunktion wie auch bei der Autokorrelationsfunktion nur einen Peak beim Lag 0. Für alle anderen Lags beträgt der Wert der beiden Funktionen 0. Dies ist bei farbigem Rauschen anders. Da die Werte bei farbigem Rauschen untereinander korreliert sind, ergeben sich auch bei Lags ungleich 0 signifikante Peaks.

In Matlab existiert für die Berechnung die Funktion `xcorr`. Damit läßt sich sowohl die Autokovarianzfunktion als auch die Autokorrelationsfunktion berechnen.

Abbildung 2.1 zeigt den typischen Verlauf der ACF für eine korrelierte Zeitreihe. Hier treten für Lags ungleich Null signifikante Peaks auf. Die ACF einer unkorrelierten Zeitreihe hingegen zeigt keine signifikanten Nebenpeaks, vgl. Abbildung 2.2.

Die Grenzen wurden mit der Formel $\pm 1.96n^{\frac{1}{2}}$ mit n der Länge der Zeitreihe (Teusch [2006]). Sie entsprechen dem 95%-Konfidenzintervall. Dies bedeutet, dass die Grenzen mit 95%-iger Wahrscheinlichkeit die wahren Signifikanzgrenzen sind. Die Länge der Zeitreihe ist 2000, es wurde aber aus Gründen der Übersichtlichkeit nur ein Teil davon dargestellt.

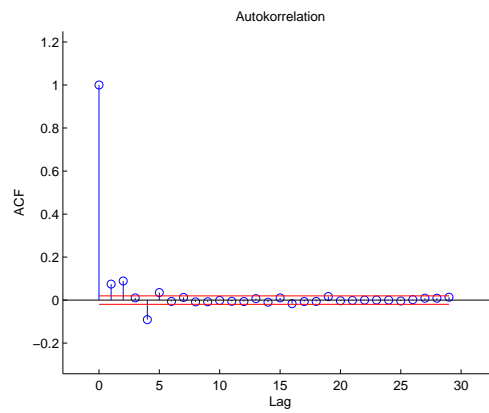


Abbildung 2.1: ACF von farbigem Rauschen

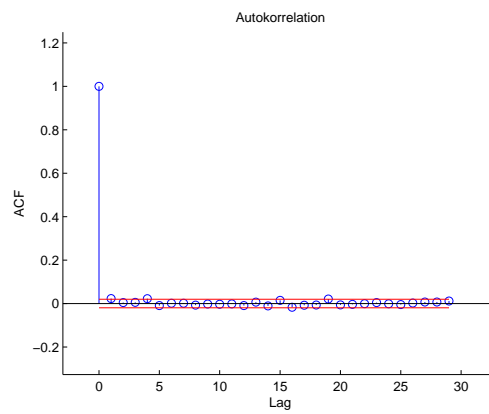


Abbildung 2.2: ACF von weißem Rauschen

2.4 Partielle Autokorrelationsfunktion (PACF)

Neben der ACF gibt es auch die PACF. Diese beschreibt die Korrelation zwischen den Werten x_t und x_{t+h} , wobei der Einfluss der dazwischenliegenden Werte eliminiert wird. Sei

$$x_t = \phi_1 x_{t-1} + \epsilon_t,$$

wobei x_t, x_{t-1} die Werte einer korrelierten Zeitreihe und ϵ_t den Wert des weißen Rauschens beschreiben. Dann ist für $\phi_1 \neq 0$ x_t mit x_{t-1} korreliert. Für x_{t-1} gilt wiederum

$$x_{t-1} = \phi_1 x_{t-2} + \epsilon_{t-1}.$$

Es lässt sich also die Gleichung für x_t umschreiben in

$$x_t = \phi_1 (\phi_1 x_{t-2} + \epsilon_{t-1}) + \epsilon_t.$$

Aus der Gleichung lässt sich eine Korrelation von x_t und x_{t-2} erkennen. Für die Werte x_t und x_{t-2} existieren aber zusätzlich indirekte Korrelationen, die aus der Gleichung nicht unbedingt ersichtlich sind. Diese Art von Korrelationen werden in der ACF berücksichtigt. Die PACF hingegen beschreibt nur die Korrelation zwischen x_t und x_{t-2} ohne den Einfluss von x_{t-1} . Genauer beschrieben ist dies in Teusch [2006].

Die PACF wird mit Hilfe der folgenden Gleichungen berechnet

$$\underbrace{\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_{p-1} \\ r_p \end{pmatrix}}_{\mathbf{r}} = \underbrace{\begin{pmatrix} 1 & r_1 & r_2 & \cdots & r_{p-2} & r_{p-1} \\ r_1 & 1 & r_1 & \cdots & r_{p-3} & r_{p-2} \\ & \vdots & & & \vdots & \\ r_{p-2} & r_{p-3} & r_{p-4} & \cdots & 1 & r_1 \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & r_1 & 1 \end{pmatrix}}_{\mathbf{R}} \underbrace{\begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{p-1} \\ \psi_p \end{pmatrix}}_{\boldsymbol{\Psi}},$$

wobei ψ_p der gesuchte Wert der PACF beim Lag p , \mathbf{r} der Vektor der Autokorrelationen und \mathbf{R} die sogenannte Autokorrelationsmatrix ist. Diese Gleichungen heißen Yule-Walker-Gleichungen. Die Autokorrelationsmatrix hat Toeplitzstruktur. Das Gleichungssystem lässt sich mittels

$$\hat{\boldsymbol{\Psi}} = \mathbf{R}^{-1} \mathbf{r},$$

lösen.

Die Werte der PACF für farbiges und weißes Rauschen sind in den Abbildungen 2.3 und 2.4 dargestellt. Bei weißem Rauschen kann es nur einen Peak geben, da die ACF keine Korrelationen aufweist. Im Gegensatz dazu gibt es bei farbigem Rauschen auch Peaks bei Lags ungleich 0.

Die Berechnung der PACF erfolgt in der Funktion `pacf`, wobei auch automatisch ein Plot der PACF mit ausgegeben wird.

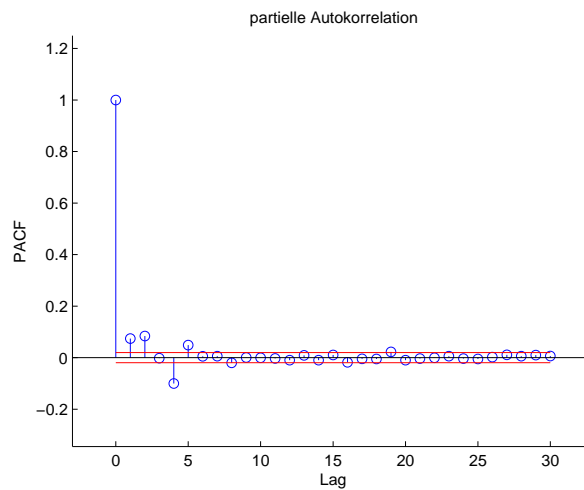


Abbildung 2.3: PACF von farbigem Rauschen

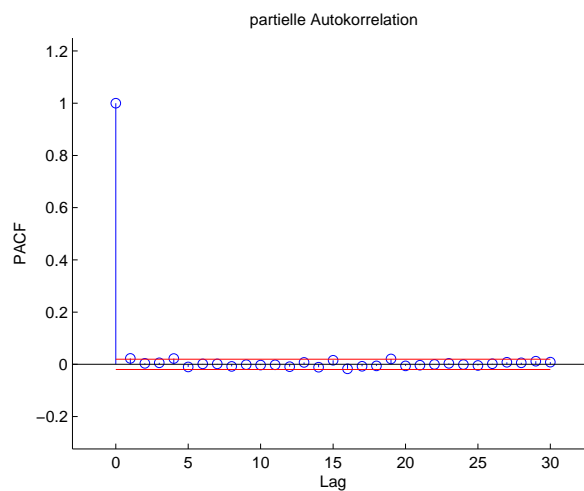


Abbildung 2.4: PACF von weißem Rauschen

2.5 Leistungsspektraldichte (PSD)

Die PSD-Funktion (Power Spectral Density oder Leistungsspektraldichte) gibt an wieviel Energie in einem bestimmten (Frequenz-)Spektrum enthalten ist. Oder anders ausgedrückt, mit welcher Amplitude die einzelnen Frequenzen in dem betrachteten Spektrum auftreten. Die PSD-Funktion wird im Allgemeinen für das Spektrum von $-f_{max}$ bis $+f_{max}$ berechnet. Die entsprechende Darstellung wird als das zweiseitige PSD-Spektrum bezeichnet.

Eine wichtige Eigenschaft der PSD-Funktion ist aber ihre Symmetrie, wodurch nur das Spektrum von 0 bis $+f_{max}$ betrachtet werden muss. Diese Darstellung wird als einseitiges PSD-

Spektrum bezeichnet. Hierbei ist zu beachten, dass den einzelnen Frequenzen in der einseitigen Darstellung die doppelte Energie zugeordnet ist. Dies ist äquivalent zur Verdoppelung der Amplituden. Die Frequenz f_{max} ist gleichzusetzen mit der Nyquist-Frequenz. Die Nyquist-Frequenz ist die halbe Abtastfrequenz gemäß

$$f_{max} = f_{Nyquist} = \frac{1}{2} f_{Abtastung}.$$

Nach dem Abtasttheorem kann ein Signal nur dann beliebig genau rekonstruiert werden, wenn alle Anteile in diesem Signal kleinere Frequenzen als die Nyquist-Frequenz haben, ansonsten treten Aliasing-Effekte auf.

Die PSD-Funktion ist die Fouriertransformierte der Autokovarianzfunktion:

$$PSD(f) = \int_{-\infty}^{+\infty} Cov(\tau) \cdot e^{-i \cdot 2\pi f \tau} d\tau.$$

Darin bezeichnet $Cov(\tau)$ die Autokovarianzfunktion mit Lag τ .

Die PSD-Funktion von weißem Rauschen ist optimalerweise ein Konstante mit dem Wert der Varianz des weißen Rauschens. In der Realität nimmt sie für gewöhnlich die Form eines Trichters an. Dies ist beispielhaft dargestellt in Abbildung 2.5

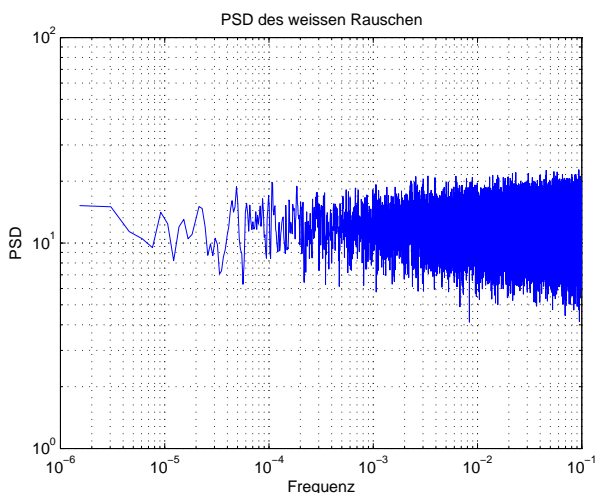


Abbildung 2.5: PSD von weißem Rauschen

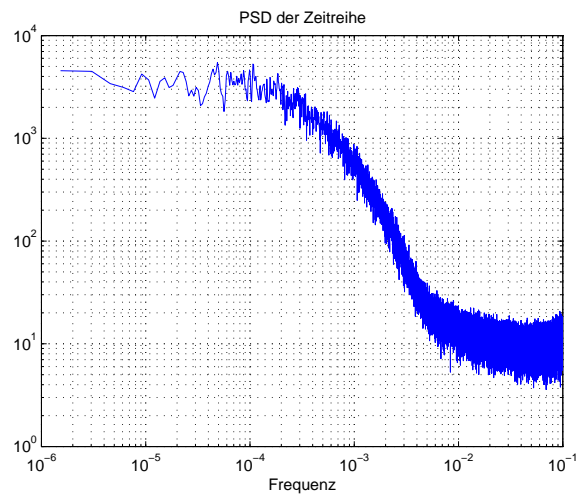


Abbildung 2.6: PSD von farbigem Rauschen

Bei farbigem Rauschen ist die PSD-Funktion kein Trichter oder eine Konstante, sondern hat einen anderen Verlauf. Abbildung 2.6 beispielsweise stellt die PSD-Funktion eines Prozesses dar, der auch bei der Residualanalyse in Kapitel 4 Anwendung findet.

Kapitel 3

Stationäre Prozesse

3.1 AR-Prozesse

3.1.1 Algorithmus

Ein AR-Prozess ist ein Auto-Regressiver Prozess. Dies bedeutet, dass jeder Wert der korrelierten Zeitreihe von einer bestimmten Anzahl von vorhergehenden Werten der Zeitreihe und dem Wert des weißen Rauschens zum Zeitpunkt t abhängt. Die Prozessgleichung lautet somit:

$$x_t + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_3 x_{t-3} + \dots + \phi_p x_{t-p} = \epsilon_t.$$

Die Koeffizienten ϕ_i definieren den Prozess und werden auch Linearfaktoren oder Prozessparameter genannt.

Die Ordnung des AR-Prozesses wird mit dem Wert p angegeben. Er wird auch Korrelationslänge genannt. Er gibt somit an, wie viele vorhergehende Werte der Zeitreihe Einfluss haben auf den aktuellen Wert der Zeitreihe. Allgemein wird auch AR[p] geschrieben. Ein unendlich langer AR-Prozess kann auch als AR[∞] geschrieben werden. Charakteristisch für einen AR[p]-Prozess ist, dass die PACF nach dem Wert p abbricht. Ein Beispiel hierfür findet sich in Abbildung 3.1.

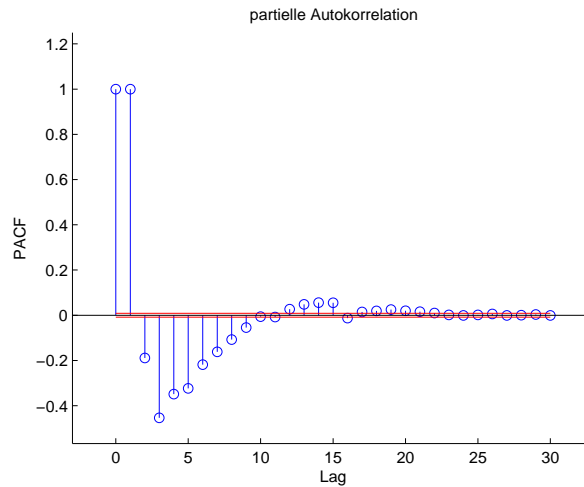


Abbildung 3.1: PACF eines AR[15]-Prozesses

3.1.2 Berechnung der Prozessparameter

Für die Berechnung der Koeffizienten eines AR-Prozesses sind in Matlab mehrere Funktionen implementiert: *lpc*, *arburg*, *aryule*, *arcov*. Die Ergebnisse für die geschätzten Koeffizienten werden besser, je länger die Zeitreihen sind. Der numerische Aufwand wächst entsprechend an.

Die Funktionen *lpc* und *aryule* liefern die gleichen Ergebnisse, da beide die Yule-Walker-Gleichungen für die Schätzung der Koeffizienten verwenden. Der Unterschied besteht nur im Zeitaufwand der Berechnungen. Die Funktion *lpc* läuft schneller ab, da sie eine schnellere Berechnungsart der Yule-Walker-Gleichungen, genannt Levinson-Durbin-Algorithmus, verwendet.

Die Funktionen *arburg*, *arcov* und *armcov* verwenden jeweils andere Methoden zur Berechnung der Koeffizienten eines AR-Prozesses. Für weitergehende Information sei auf die Matlab-Hilfe und die dort angegebenen Referenzen verwiesen.

3.1.3 Beispiel

Dem Beispiel liegt ein AR[2]-Prozess mit den Koeffizienten

$$\begin{aligned}\phi_1 &= 0.50, \\ \phi_2 &= 0.10\end{aligned}$$

zugrunde. Nach der Erzeugung der korrelierten Zeitreihe wurde mit den Routinen *lpc*, *aryule*, *arburg* und *arcov* die Berechnung der Koeffizienten durchgeführt, wobei die Ordnung des Prozesses als bekannt vorausgesetzt wurde. Zur Identifikation der Modellierung kann die PACF

verwendet werden. Die Länge der Zeitreihe betrug 500000, wobei jedoch nur ein Teil für die Bestimmung der Prozessparameter Verwendung fand. Die Ergebnisse fassen die folgenden Tabellen zusammen.

Funktion	ϕ_1	ϕ_2	Zeit [s]
lpc	0.4318	0.0669	0.080
aryule	0.4318	0.0669	0.541
arburg	0.4322	0.0677	0.110
arcov	0.4328	0.0677	0.180

Tabelle 3.1: Ergebnisse der AR-Koeffizientenrückrechnung mit Zeitreihenlänge 1000

Funktion	ϕ_1	ϕ_2	Zeit [s]
lpc	0.4889	0.0910	0.190
aryule	0.4889	0.0910	0.240
arburg	0.4889	0.0910	0.070
arcov	0.4890	0.0910	0.130

Tabelle 3.2: Ergebnisse der AR-Koeffizientenrückrechnung mit Zeitreihenlänge 10000

Funktion	ϕ_1	ϕ_2	Zeit [s]
lpc	0.4992	0.1016	0.371
aryule	0.4992	0.1016	0.460
arburg	0.4992	0.1016	0.140
arcov	0.4992	0.1016	0.221

Tabelle 3.3: Ergebnisse der AR-Koeffizientenrückrechnung mit Zeitreihenlänge 100000

Die Ergebnisse in den Tabellen 3.1, 3.2 und 3.3 zeigen, dass die Funktionen besser arbeiten, je länger die Zeitreihe ist. Ist die Ordnung des AR-Prozesses höher, dann muss die Zeitreihe automatisch länger werden, um gute Ergebnisse zu erzielen. Die Funktionen *lpc* und *aryule* liefern die gleichen Ergebnisse, da beide fast die gleichen mathematischen Grundlagen verwenden. Die Ergebnisse unterscheiden sich nur unerheblich. Ab einer ausreichenden Länge der Zeitreihe spielt es für die numerischen Ergebnisse keine Rolle mehr, welches Verfahren verwendet wird. Ein Unterschied besteht nur im Zeitbedarf.

In Tabelle 3.4 werden die stochastischen Eigenschaften des ursprünglichen Rauschens mit denen des Rauschens, das mit den Koeffizienten, die mit Hilfe der Funktion *lpc* berechnet wurden, zurückgerechnet wurde, verglichen. Letzteres wird als generiertes Rauschen bezeichnet. Die stochastischen Eigenschaften unterscheiden sich kaum. Das ursprüngliche Rauschen wird gut rekonstruiert. Man kann somit schlussfolgern, dass die Funktionen für die Rückrechnung von AR-Koeffizienten sehr gut arbeiten.

	Mittelwert	Minimalwert	Maximalwert	Standardabweichung
ursprüngliches Rauschen	0.001335	-4.852094	4.636508	0.999488
generiertes Rauschen	0.001337	-4.853390	4.634733	0.999486

Tabelle 3.4: Vergleich der stochastischen Eigenschaften der Rauschen (AR)

3.2 MA-Prozesse

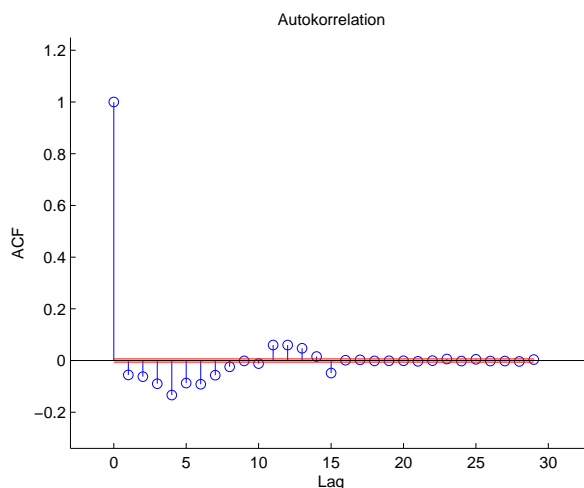
3.2.1 Algorithmus

Ein MA-Prozess ist ein Moving-Average Prozess. Dies bedeutet, dass jeder Wert der korrelierten Zeitreihe von dem aktuellen Wert des weißen Rauschens sowie von einer gewissen Anzahl vorhergehender Werte des weißen Rauschens abhängt. Die Prozessgleichung lautet demzufolge:

$$x_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \theta_3 \epsilon_{t-3} + \dots + \theta_q \epsilon_{t-q}.$$

Die Koeffizienten θ_i definieren den Prozess und werden auch Linearfaktoren oder Prozessparameter genannt.

Die Ordnung des MA-Prozesses wird mit dem Wert q angegeben. Er wird auch als Korrelationslänge bezeichnet. Er gibt an wie viele vorhergehende Werte des weißen Rauschens den Wert der Zeitreihe zum Zeitpunkt t beeinflussen. Allgemein wird auch $\text{MA}[q]$ geschrieben. Ein unendlich langer MA-Prozess wird $\text{MA}[\infty]$ geschrieben. Charakteristisch für einen $\text{MA}[q]$ -Prozess ist, dass die ACF nach dem Wert q abbricht. Ein Beispiel hierfür ist in Abbildung 3.2 dargestellt.

Abbildung 3.2: ACF eines $\text{MA}[15]$ -Prozesses

3.2.2 Berechnung der Prozessparameter

Für die Berechnung der Koeffizienten eines MA-Prozesses gibt es in Matlab keine vorimplementierten Funktionen.

Aus einer Zeitreihe eines MA-Prozesses können die MA-Koeffizienten $\theta_1, \dots, \theta_q$ mithilfe des Wilson-Algorithmus (Wilson [1969]) berechnet werden. Zuerst werden die Autokovarianzen \mathbf{c}^* der Zeitreihe berechnet. Es wird danach der Vektor der Anfangswerte der MA-Koeffizienten aufgestellt:

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_q \end{pmatrix}.$$

Für die Berechnung der Koeffizienten ist die Kenntnis der Prozessordnung notwendig. Diese kann über den Abbruch der ACF ermittelt werden. Die Anfangswerte werden in zwei Matrizen verarbeitet,

$$\mathbf{T}_1 = \begin{pmatrix} \theta_0 & \theta_1 & \cdots & \theta_q \\ \theta_1 & \cdots & \theta_q & \\ \vdots & & & \\ \theta_q & & & 0 \end{pmatrix}, \quad \mathbf{T}_2 = \begin{pmatrix} \theta_0 & \theta_1 & \cdots & \theta_q \\ & \theta_0 & \cdots & \theta_{q-1} \\ & & \ddots & \\ 0 & & & \theta_0 \end{pmatrix}.$$

Damit können nun Näherungskovarianzen berechnet werden mit

$$\mathbf{T}_1 \boldsymbol{\theta} = \mathbf{T}_2 \boldsymbol{\theta} = \mathbf{c}.$$

Dann werden die beiden Matrizen zusammengeführt durch

$$\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2.$$

Nun kann der iterative Prozess zur Bestimmung der MA-Koeffizienten gestartet werden.

$$\boldsymbol{\theta}^{(t+1)} = \mathbf{T}^{(t)-1} \left(\mathbf{c}^{(t)} + \mathbf{c}^* \right),$$

mit den Anfangswerten $\theta_0^{(0)} = \{c_0^* + 2 \sum_{i=1}^q c_i^*\}^{\frac{1}{2}}$ und $\theta_1^{(0)} = \dots = \theta_q^{(0)} = 0$. Je größer die Ordnung q des Prozesses, umso länger dauert es bis sich Konvergenz einstellt. Mit den heutigen Rechnerkapazitäten stellt dies allerdings kein Problem mehr dar.

Der Wilson-Algorithmus wurde in der Funktion *wilson* implementiert.

3.2.3 Beispiel

Es wurde ein MA[2]-Prozess erzeugt mit den Koeffizienten

$$\theta_1 = 0.45,$$

$$\theta_2 = 0.15,$$

und mittels der Funktion *wilson* die Rückrechnung der Koeffizienten durchgeführt. Die Ergebnisse für verschieden lange Zeitreihenausschnitte sind in Tabelle 3.5 aufgeführt.

Länge der Zeitreihe	θ_1	θ_2	Zeit [s]
1000	0.4223	0.1155	0.100
10000	0.4575	0.1536	0.130
100000	0.4492	0.1478	0.491

Tabelle 3.5: Ergebnis der MA-Koeffizientenrückrechnung

Wie aus den Ergebnissen ersichtlich, werden die Ergebnisse besser, je länger die Zeitreihe ist. Analog zu Kapitel 3.1 gilt hier auch, dass mit wachsender Ordnung des MA-Prozesses die Zeitreihe zur Bestimmung der Prozessparameter entsprechend länger werden muss. Es wurde, wie für den AR-Prozess, eine Zeitreihe der Länge 500000 generiert. Für die Berechnung wurden wieder nur Ausschnitte verwendet. In Tabelle 3.6 sind die stochastischen Eigenschaften von zwei Rauschsequenzen gegenüber gestellt. In der oberen Zeile stehen die Eigenschaften des originalen Rauschens, in der unteren Zeile diejenigen des Rauschens, das mit den geschätzten Koeffizienten generiert wurde. Hierfür finden die Ergebnisse der Zeitreihe mit einer Länge von 100000 Verwendung. Da sich die stochastischen Eigenschaften der Rauschsequenzen nur geringfügig unterscheiden, lässt sich feststellen, dass der Wilson-Algorithmus sehr gut arbeitet.

	Mittelwert	Minimalwert	Maximalwert	Standardabweichung
ursprüngliches Rauschen	0.002194	-4.391580	4.592952	1.000725
generiertes Rauschen	0.002202	-4.395466	4.596739	1.000738

Tabelle 3.6: Vergleich der stochastischen Eigenschaften der Rauschen (MA)

3.3 ARMA-Prozesse

3.3.1 Algorithmus

Ein ARMA-Prozess besteht aus einem AR- und einem MA-Anteil. Das heißt, dass jeder Wert der Zeitreihe vom aktuellen Wert des weißen Rauschens, von den vorhergehenden q Werten des weißen Rauschens und von den vorhergehenden p Werten der korrelierten Zeitreihe abhängt.

Die Prozessgleichung lautet demnach:

$$x_t + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}.$$

Entsprechend dem AR- und MA-Prozess definieren die Koeffizienten ϕ_i und θ_i den ARMA-Prozess. Sie werden auch als Prozessparameter bezeichnet.

Man spricht hier auch von einem ARMA[p,q]-Prozess. Ein ARMA[$p,0$]-Prozess ist gleichbedeutend mit einem AR[p]-Prozess und ein ARMA[$0,q$]-Prozess ist gleichbedeutend mit einem MA[q]-Prozess. Ein ARMA[$p,0$]- bzw. AR[p]-Prozess heißt auch "recursive Filter", während ein ARMA[$0,q$]- bzw. MA[q]-Prozess auch "convolution Filter" genannt wird (Schuh [2003]). Für einen ARMA-Prozess bricht weder die ACF noch die PACF ab, sondern es schwingen beide aus.

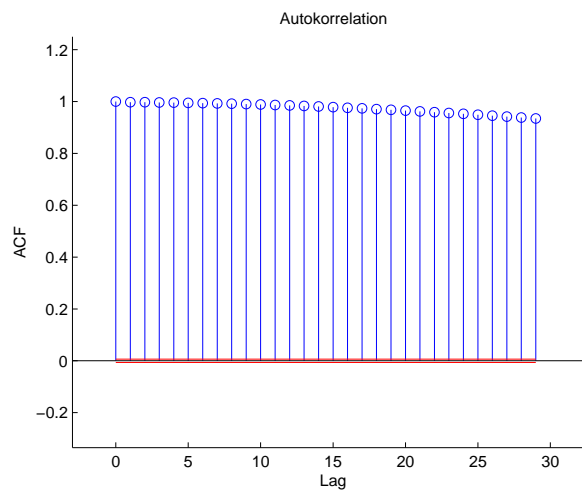


Abbildung 3.3: ACF eines ARMA[15,15]-Prozesses

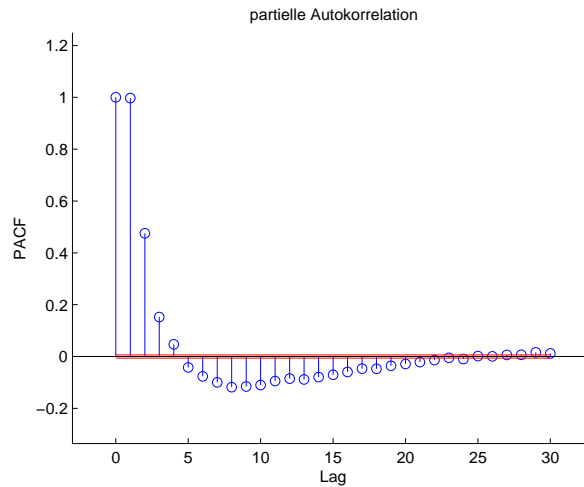


Abbildung 3.4: PACF eines ARMA[15,15]-Prozesses

Man sieht in den Abbildungen 3.3 und 3.4, dass die ACF und die PACF kein vorhersagbares Verhalten für einen ARMA-Prozess aufweisen.

3.3.2 Berechnung der Prozessparameter

Für die Berechnung der ARMA-Koeffizienten aus einer gegebenen Zeitreihe gibt es verschiedene Verfahren. Im Rahmen dieser Studienarbeit wurden das erste und das zweite Durbin Verfahren verwendet und implementiert. Für Informationen zu weiteren Berechnungsverfahren für ARMA-Koeffizienten wird auf Ries [2003] verwiesen.

Für die hier vorgestellten und angewendeten Verfahren ist die Kenntnis der Modellordnung p und q Voraussetzung. In dieser Arbeit werden p und q als bekannt vorausgesetzt. In Ries [2003] werden Verfahren zur Modellidentifikation von ARMA-Prozessen vorgestellt.

3.3.2.1 Erstes Durbin-Verfahren

Diese Verfahren benutzt ein Gauß-Markov-Modell (GMM, vgl. Kapitel 4) zur Bestimmung der Parameter. Es ergibt sich

$$\sum_t [x_t - (\epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} - \phi_1 x_{t-1} - \dots - \phi_p x_{t-p})]^2 \rightarrow \min.$$

In dieser Darstellung sind sowohl die Parameter $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ als auch das Rauschen ϵ unbekannt.

Die Zeitreihe wird zunächst durch einen langen AR-Prozess angenähert. Die Länge des AR-Prozesses kann beispielsweise $l = \sqrt{N}$ sein, wobei N die Länge der originalen Zeitreihe ist. Mit den erhaltenen Hilfskoeffizienten kann nun der AR-Prozess invers berechnet werden. Als

Ergebnis erhält man eine Näherung ϵ' für das Rauschen. Es wird angenommen, dass der Prozess durch den langen AR-Prozess so gut angenähert wird, dass man die erhaltene Näherung für das Rauschen anstelle des unbekanntes ursprünglichen Rauschens verwenden kann. Damit kann die Designmatrix des Minimierungsproblems aufgestellt werden:

$$\mathbf{A} = \begin{bmatrix} \vdots & & & \vdots & \vdots & & & \vdots \\ -x_{k-1} & -x_{k-2} & \cdots & -x_{k-p} & \epsilon'_{k-1} & \epsilon'_{k-2} & \cdots & \epsilon'_{k-q} \\ -x_k & -x_{k-1} & \cdots & -x_{k-p+1} & \epsilon'_k & \epsilon_{k-1} & \cdots & \epsilon'_{k-q+1} \\ \vdots & & & \vdots & \vdots & & & \vdots \end{bmatrix}.$$

Der Zähler k läuft von $h+1$ bis N , wobei $h = \max(p, q)$ und N gleich der Länge der Zeitreihe \mathbf{x} ist. Es ist also offensichtlich, dass die Designmatrix $p+q$ Spalten und $N-h$ Zeilen hat. Der Beobachtungsvektor des GMM ergibt sich zu

$$\mathbf{1} = \begin{pmatrix} x_{h+1} \\ x_{h+2} \\ \vdots \\ x_N \end{pmatrix}.$$

Es werden demnach alle bis auf die ersten h Beobachtungen verwendet. Mit dem gebräuchlichen Ausgleichungsalgorithmus erhält man den Unbekanntenvektor

$$\hat{\mathbf{x}} = \begin{pmatrix} \hat{\phi}_1 \\ \vdots \\ \hat{\phi}_p \\ \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_q \end{pmatrix}.$$

Die Genauigkeit der resultierenden Koeffizienten ist abhängig von der Genauigkeit des geschätzten Rauschens ϵ' . Je genauer diese Näherung ist, umso besser approximieren die geschätzten Koeffizienten die wahren Koeffizienten.

Diese Verfahren wurde in der Funktion *durbin* implementiert.

3.3.2.2 Zweites Durbin-Verfahren

Dieses Verfahren benutzt die Werte des ersten Durbin-Verfahrens als Start- bzw. Näherungswerte. Es wird angenommen, dass die berechneten AR-Koeffizienten die wahren Koeffizienten sind. Mit diesen Koeffizienten wird der AR-Anteil aus der Zeitreihe heraus gerechnet.

In der verbleibenden Zeitreihe dürfte nur noch der MA-Anteil enthalten sein, wenn die AR-Koeffizienten die richtigen waren. Mit der verbleibenden Zeitreihe werden die MA-Koeffizienten mittels des Wilson-Algorithmus berechnet. Mit diesen wird aus der ursprünglichen Zeitreihe der MA-Anteil heraus gerechnet, somit müsste ausschließlich der AR-Anteil in der Zeitreihe verbleiben. Aus der verbleibenden Zeitreihe werden danach die AR-Koeffizienten berechnet. Mit den berechneten AR- und MA-Koeffizienten kann nun wieder am Beginn des Verfahrens eingestiegen werden. Dies ist ein iterativer Prozess, der mit der Zeit konvergiert. Über die Konvergenz können keine Aussagen getroffen werden, vgl. Ries [2003].

Dieses Verfahren wurde in der Funktion *durbin2* umgesetzt.

3.3.3 Beispiel

Für das folgende Beispiel wurden der AR[2]- und der MA[2]-Prozess von oben zusammengeführt in einen ARMA[2,2]-Prozess. Die Koeffizienten wurden einmal mit dem ersten Durbin-Verfahren und einmal mit dem zweiten Durbin-Verfahren berechnet.

Länge der Zeitreihe	ϕ_1	ϕ_2	θ_1	θ_2	Zeit [s]
1000	0.7543	0.0463	0.7428	0.0958	0.481
10000	0.4326	0.1826	0.3994	0.2462	4.356
100000	0.5182	0.1248	0.4671	0.1722	40.959
500000	0.5092	0.1078	0.4586	0.1557	185.898

Tabelle 3.7: Ergebnisse der ARMA-Koeffizientenberechnung mit dem ersten Durbin-Verfahren

Tabelle 3.7 verdeutlicht, dass die Ergebnisse der Koeffizienten mit längerer Zeitreihe besser werden.

Länge der Zeitreihe	ϕ_1	ϕ_2	θ_1	θ_2	Zeit [s]
1000	0.6147	0.2090	0.6286	0.3535	0.611
10000	0.4245	0.0874	0.3885	0.1372	5.227
100000	0.5243	0.1014	0.4746	0.1515	43.162
500000	0.5053	0.1043	0.4559	0.1538	190.023

Tabelle 3.8: Ergebnisse der ARMA-Koeffizientenberechnung mit dem zweiten Durbin-Verfahren

Auch für das zweite Durbin-Verfahren, vgl. Tabelle 3.8, verbessern sich die Ergebnisse der Koeffizienten mit wachsender Zeitreihe.

Wie schon in Kapitel 3.1 und 3.2, werden auch hier die stochastischen Eigenschaften des ursprünglichen weißen Rauschens mit den stochastischen Eigenschaften des generierten weißen Rauschens verglichen. Das generierte Rauschen entsteht durch die Rückrechnung aus der originalen korrelierten Zeitreihe unter Zuhilfenahme der geschätzten Koeffizienten. Die Ergebnisse

finden sich in den Tabellen 3.9 und 3.10.

	Mittelwert	Minimalwert	Maximalwert	Standardabweichung
ursprüngliches Rauschen	-0.001019	-4.360188	5.002409	0.998551
generiertes Rauschen	-0.001016	-4.359605	5.000115	0.998548

Tabelle 3.9: Vergleich der stochastischen Eigenschaften der Rauschsequenzen für das erste Durbin-Verfahren

	Mittelwert	Minimalwert	Maximalwert	Standardabweichung
ursprüngliches Rauschen	-0.001019	-4.360188	5.002409	0.998551
generiertes Rauschen	-0.001016	-4.359544	5.000189	0.998548

Tabelle 3.10: Vergleich der stochastischen Eigenschaften der Rauschsequenzen für das zweite Durbin-Verfahren

Beide Verfahren sind für die Bestimmung der Prozessparameter geeignet. Das zweite Durbin-Verfahren liefert eine geringfügige Verbesserung der Ergebnisse des ersten Durbin-Verfahrens. Dies ist in den Tabellen 3.7 und 3.8 an einem Vergleich der jeweils letzten Zeile zu erkennen.

Kapitel 4

Residualanalyse

4.1 Grundlagen

Für diskrete Messungen können korrelierte Zeitreihen als fehlerbehaftete Beobachtungen interpretiert werden. In der Regel liegen mehr Beobachtungen als zu berechnende Unbekannte vor. Das resultierende Gleichungssystem ist somit überbestimmt. Dies ermöglicht uns die Anwendung eines Ausgleichsalgorithmus, der Ergebnisse liefert, die am besten die Forderung einer minimierten Quadratsumme der Verbesserungen erfüllen. Hierfür wird in der Regel das Gauß-Markov-Modell (GMM) verwendet

$$\mathbf{l} + \mathbf{v} = \mathbf{A}\mathbf{x} \quad , \quad E\{\mathbf{l}\} = \mathbf{A}\mathbf{x} \quad , \quad \mathbf{D}\{\mathbf{l}\} = \boldsymbol{\Sigma}_l = \sigma_0^2 \mathbf{Q}_l.$$

Darin bezeichnet \mathbf{l} den Beobachtungsvektor, \mathbf{v} den Verbesserungsvektor oder Residuenvektor, \mathbf{A} die Designmatrix, \mathbf{x} den Unbekanntenvektor, $\boldsymbol{\Sigma}_l$ die Kovarianzmatrix, \mathbf{Q}_l die Kofaktormatrix und σ_0^2 die Varianz der Standardeinheit. In der Kovarianzmatrix stehen die Varianzen der Beobachtungen und die Kovarianzen zwischen den Beobachtungen. Die Minimierung der Verbesserungsquadratsumme liefert die beste lineare erwartungstreue Schätzung des Unbekanntenvektors zu

$$\hat{\mathbf{x}} = (\mathbf{A}^T \boldsymbol{\Sigma}_l^{-1} \mathbf{A})^{-1} \mathbf{A}^T \boldsymbol{\Sigma}_l^{-1} \mathbf{l}.$$

Die Residuen ergeben sich aus dem Zusammenhang

$$\mathbf{v} = \mathbf{A}\hat{\mathbf{x}} - \mathbf{l}.$$

In die Ausgleichung fließen als Beobachtung unverfälschte, simulierte GOCE-Daten ein, die mit verschiedenen Rauschsequenzen verfälscht werden. Das Simulationsszenario deckt einen Zeitraum von 30 Tagen ab mit einer Abtastrate von 5 Sekunden. Für die folgenden Untersuchungen wird die Analyse der radialen SGG(Satellite Gravity Gradiometrie)-Tensorkomponente heran-

gezogen.

Im Unbekanntenvektor stehen die Koeffizienten des Erdschwerefeldes c_{lm} und s_{lm} . Die Designmatrix enthält die partiellen Ableitungen der Beobachtungsgleichung nach den Unbekannten. Wie oben beschrieben ergeben sich die Verbesserungen aus der Ausgleichung. Dies sind die Zeitreihen der Residuen, die im weiteren Verlauf untersucht werden sollen. Die Berechnung der Schwerefeldkoeffizienten und damit der Residuen ist nicht Bestandteil dieser Studienarbeit und wird deshalb extern ausgeführt.

Als Rauschsequenzen werden weißes Rauschen, ein AR[15]-, ein MA[15]- und ein ARMA[15,15]-Prozess verwendet. Die Residualanalyse untersucht, ob die Residuen die gleichen stochastischen Eigenschaften wie die Beobachtungen haben. Oder anders ausgedrückt, ob sich die Prozesse, mit denen die Beobachtungen verrauscht wurden, in den Residuen wiederfinden. Die aus den Residuen berechneten Prozesse werden aus der Zeitreihe der Residuen herausgerechnet. Die verbleibende Sequenz sollte im Idealfall weißes Rauschen aufweisen. Zur Überprüfung wird die ACF dieser Sequenz berechnet.

Die Prozessparameter, die aus den Residuen berechnet werden, kommen bei der Dekorrelation in Kapitel 5 zum Einsatz.

Im folgenden werden die ACF und PACF nur ausschnittsweise betrachtet.

4.2 Weißes Rauschen als Fehlermodell

In diesem Fall wurden die Beobachtungen mit einer Sequenz unkorrelierter Beobachtungsfehler versehen. Das verwendete weiße Rauschen ist dasjenige, das bei der Erzeugung der korrelierten Zeitreihen (AR[15], MA[15], ARMA[15,15]) zum Einsatz kam.

4.2.1 Ergebnisse

Die Eigenschaften der Residuen sind in den Abbildungen 4.1 und 4.2 zu sehen.

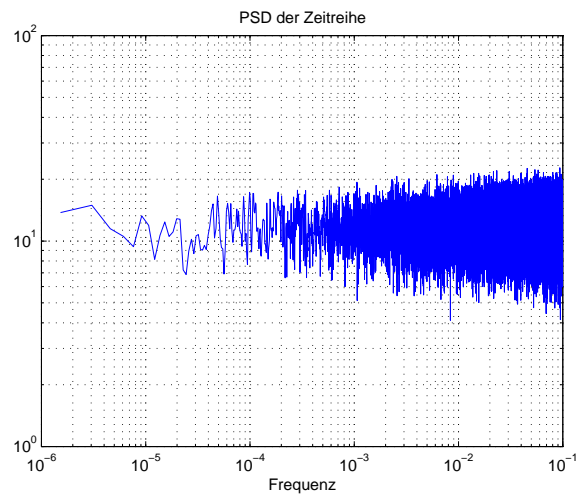


Abbildung 4.1: PSD der Residuen (weißes Rauschen)

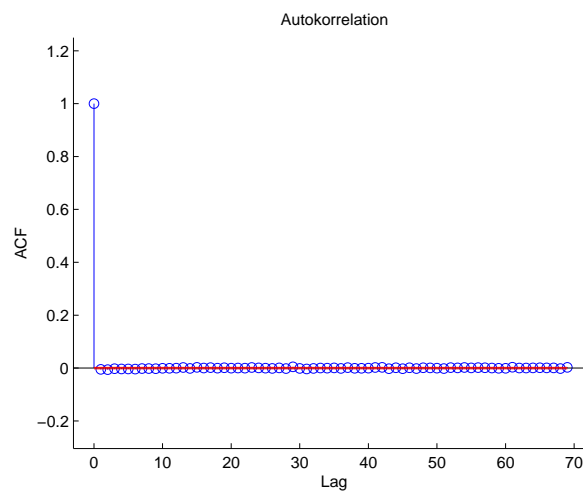


Abbildung 4.2: ACF der Residuen (weißes Rauschen)

Die PSD-Funktion legt den Schluss nahe, dass es sich bei den stochastischen Eigenschaften der Zeitreihe um die Eigenschaften von weißem Rauschen handelt. Zum Beweis lässt sich die ACF heranziehen.

Die ACF zeigt deutlich, dass es sich bei den stochastischen Eigenschaften um die von weißem Rauschen handelt. Im Falle, dass die Beobachtungen mit weißem Rauschen verfälscht wurden, haben die Residuen die gleichen stochastischen Eigenschaften wie die Beobachtungen.

4.3 AR[15]-Prozess als Fehlermodell

Der AR[15]-Prozess wurde mit den Koeffizienten aus Tabelle 4.1 und einem weißen Rauschen mit der Varianz $\sigma^2 = 15.1398$ erzeugt.

Koeffizienten	Wert
ϕ_1	-0.57313
ϕ_2	-0.45197
ϕ_3	-0.19581
ϕ_4	-0.14918
ϕ_5	0.00866
ϕ_6	0.02901
ϕ_7	0.07921
ϕ_8	0.11307
ϕ_9	0.09797
ϕ_{10}	0.05196
ϕ_{11}	0.08552
ϕ_{12}	0.03564
ϕ_{13}	0.00130
ϕ_{14}	-0.04161
ϕ_{15}	-0.08990

Tabelle 4.1: Koeffizienten AR[15]-Prozess

4.3.1 Ergebnisse

Wie in der PSD-Funktion zu sehen ist, sind die stochastischen Eigenschaften der Residuen nicht die von weißem Rauschen.

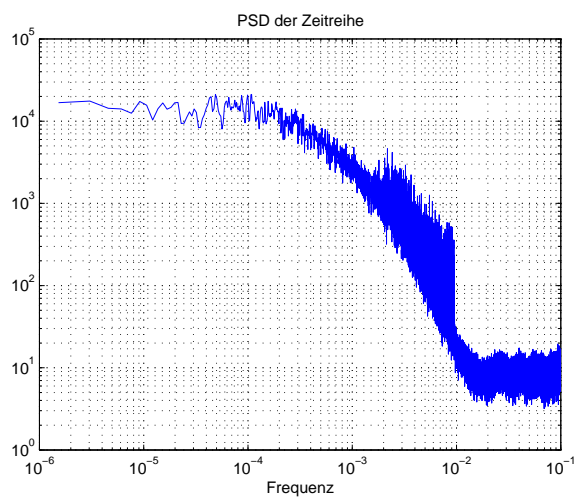


Abbildung 4.3: PSD der Residuen (AR)

In der Abbildung 4.3 ist deutlich eine Auffälligkeit der PSD bei einer Frequenz von ungefähr 0.01 Hz zu sehen. Dieser resultiert aus der Bahnfrequenz von GOCE und dem Entwicklungsgrad des Schwerefeldes. Eine Faustformel sagt aus, dass die maximal erfassbare Frequenz aus dem Produkt von Bahnfrequenz und Entwicklungsgrad abgeschätzt werden kann. GOCE besitzt eine Umlaufzeit von 5370 Sekunden. Daraus ergibt sich eine Bahnfrequenz von $1.86 \cdot 10^{-4}$ Hz. Wird diese nun zur Abschätzung mit dem Entwicklungsgrad des Schwerefeldes multipliziert, in unserem Fall ist dieser 50, so ergibt sich die maximal auflösbare Frequenz zu $9.3 \cdot 10^{-3}$ Hz. In den höheren Frequenzen ist kein Gravitationssignal enthalten, sondern ausschließlich das simulierte Messrauschen.

Nun werden zur Identifizierung des Prozesses die ACF und PACF der Residuen betrachtet. Es liegt die Vermutung nahe, dass die Residuen die gleichen stochastischen Eigenschaften wie die Beobachtungen haben. In diesem Fall müsste die ACF eine gedämpfte Schwingung aufweisen und die PACF beim Lag 15 abbrechen. Die beiden Funktionen sind in den Abbildungen 4.4 und 4.5 dargestellt.

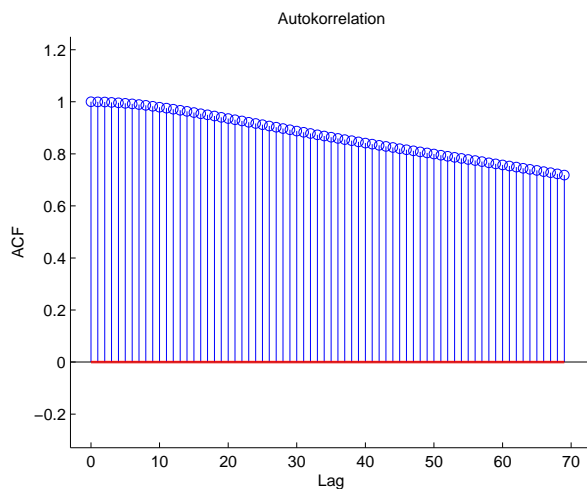


Abbildung 4.4: ACF der Residuen (AR)

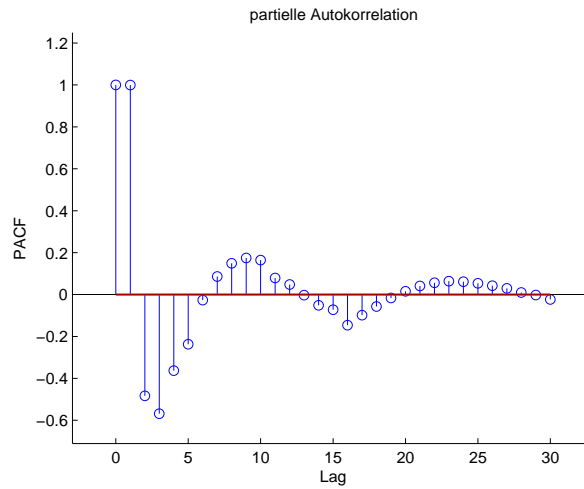


Abbildung 4.5: PACF der Residuen (AR)

Da die ACF nur als Ausschnitt betrachtet wird, ist das Ausschwingen nicht zu erkennen. Man sieht aber, dass die ACF nicht abbricht. Für die PACF ist ein Schwingen erkennbar, wodurch die Charakteristik des AR[15]-Prozesses nicht gestützt wird. Man würde eher einen ARMA-Prozess annehmen. Dennoch werden für die weiteren Ausführungen die stochastischen Eigenschaften der Residuen durch einen AR[15]-Prozess modelliert. Daraufhin werden die Differenzen zu den Koeffizienten, die zur Erzeugung des Fehlermodells verwendet wurden, betrachtet.

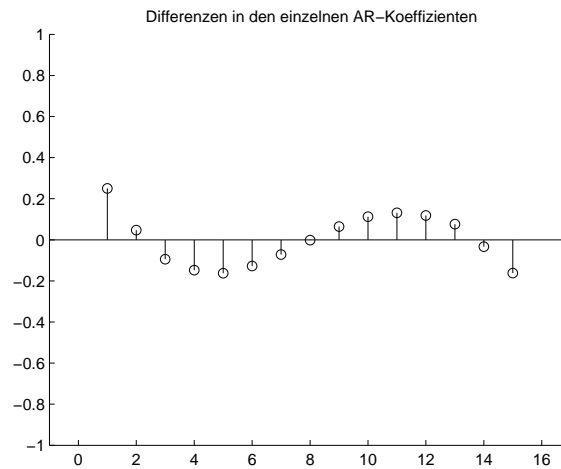


Abbildung 4.6: Differenzen (Soll-Ist) in den Koeffizienten (AR)

Es treten Differenzen von bis zu 0.3 auf, vgl. Abbildung 4.6. Das erweckt den Verdacht, dass es kein AR[15]-Prozess ist, oder dass der Prozess auch durch eine anderen AR[15]-Prozess

angenähert werden kann. Mit den geschätzten Koeffizienten wird nun der inverse Prozess berechnet. Die Residuen werden also mit dem geschätzten AR[15]-Prozess gefiltert. Als Ergebnis sollte weißes Rauschen zurückbleiben. Weißes Rauschen müsste aus der PSD-Funktion und der ACF der verbleibenden Sequenz erkennbar sein.

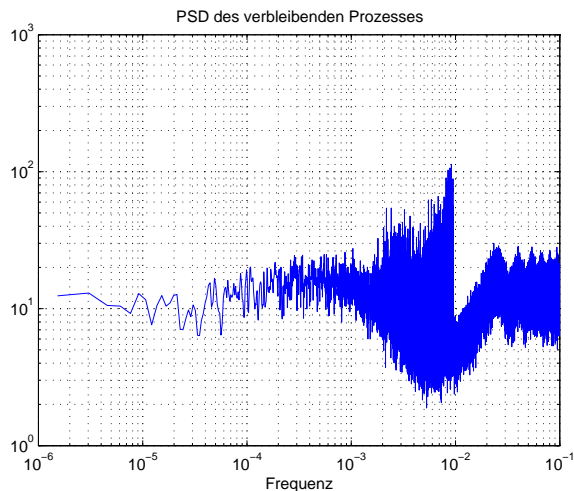


Abbildung 4.7: PSD der verbleibenden Sequenz (AR)

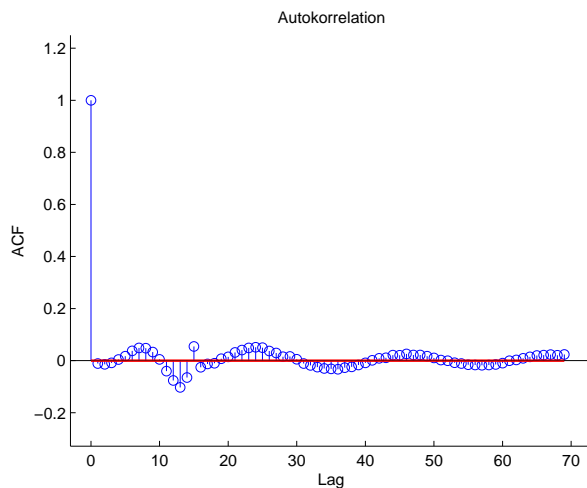


Abbildung 4.8: ACF der verbleibenden Sequenz (AR)

Was aus der PSD-Funktion in Abbildung 4.7 nicht so ohne weiteres erkennbar ist, wird aus der ACF in Abbildung 4.8 deutlich. Es bleibt kein weißes Rauschen übrig. Der AR[15]-Prozess gibt den enthaltenen Prozess nicht korrekt wieder. Erst die Dekorrelation kann klären, ob der berechnete AR[15]-Prozess die gleichen stochastischen Eigenschaften wie das Fehlermodell aufweist.

4.4 MA[15]-Prozess als Fehlermodell

Der MA[15]-Prozess wurde mit den Koeffizienten aus Tabelle 4.2 und einem weißen Rauschen mit der Varianz $\sigma^2 = 15.1398$ erzeugt.

Koeffizienten	Wert
θ_1	-0.234031
θ_2	-0.199527
θ_3	-0.190076
θ_4	-0.1916907
θ_5	-0.100385
θ_6	-0.081473
θ_7	-0.019226
θ_8	0.017724
θ_9	0.041253
θ_{10}	0.030212
θ_{11}	0.083834
θ_{12}	0.0710934
θ_{13}	0.047784
θ_{14}	0.006651
θ_{15}	-0.060133

Tabelle 4.2: Koeffizienten MA[15]-Prozess

4.4.1 Ergebnisse

Anhand der PSD-Funktion in Abbildung 4.9 wird deutlich, dass in den Residuen ein Prozess enthalten ist. Man sieht aber nicht, welcher Art der Prozess ist.

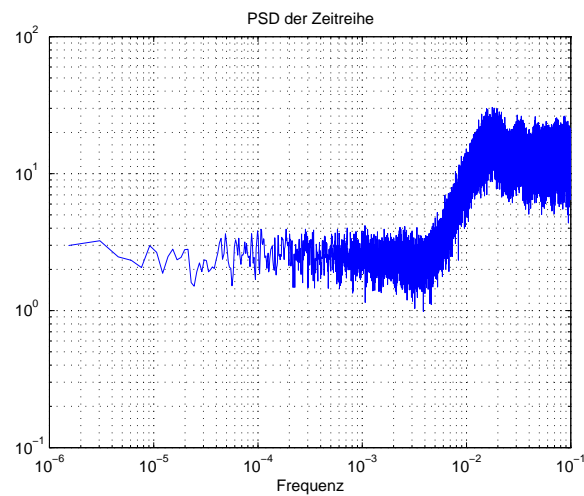


Abbildung 4.9: PSD der Residuen (MA)

Ist es ein MA[15]-Prozess, müsste dies in der ACF sichtbar sein. Diese müsste beim Lag 15 abbrechen. In der PACF sollte ein gedämpftes Schwingen zu beobachten sein.

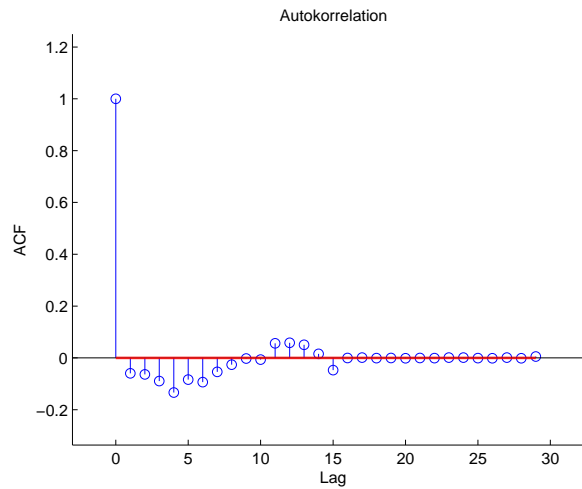


Abbildung 4.10: ACF der Residuen (MA)

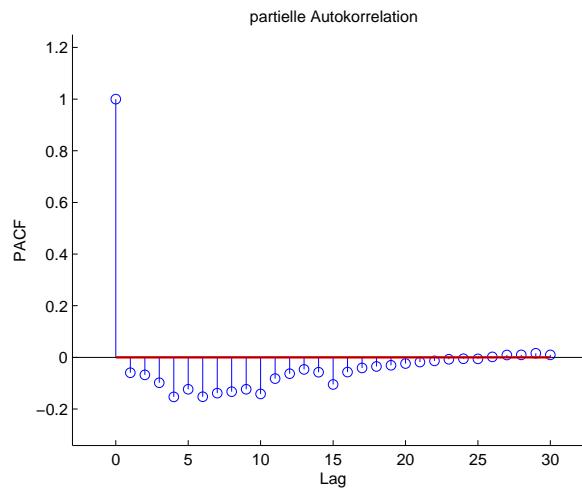


Abbildung 4.11: PACF der Residuen (MA)

Es ist deutlich zu erkennen, dass die ACF in Abbildung 4.10 beim Lag 15 abbricht. Das Schwingen in der PACF in Abbildung 4.11 ist nicht ganz so deutlich erkennbar, aber es tritt auch kein Abbruch auf. Der Prozess, der in den Residuen enthalten ist, wird als MA[15]-Prozess identifiziert. Mit diesen Erkenntnissen geht es in die Analyse der Zeitreihe. Da es sich um einen MA-Prozess handelt, werden die Koeffizienten mit Hilfe des Wilson-Algorithmuses berechnet. Diese Koeffizienten werden mit den Koeffizienten, die bei der Erstellung des Fehlermodells zum Einsatz kamen, verglichen. Man beachte, dass die Ordinate in Abbildung 4.12 mit dem

Faktor 10^{-3} versehen ist.

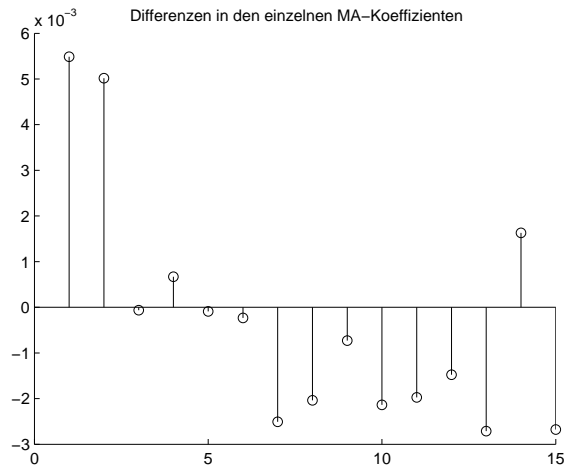


Abbildung 4.12: Differenzen (Soll-Ist) in den Koeffizienten (MA)

Mit diesen Koeffizienten wird nun der inverse Prozess berechnet. Die Residuen werden mit dem berechneten MA-Prozess gefiltert. Als Ergebnis sollte weißes Rauschen zurückbleiben. Die PSD-Funktion in Abbildung 4.13 stützt diese Aussage. Zum Beweis wird wiederum die ACF, die in Abbildung 4.14 dargestellt ist, der verbliebenen Sequenz berechnet.

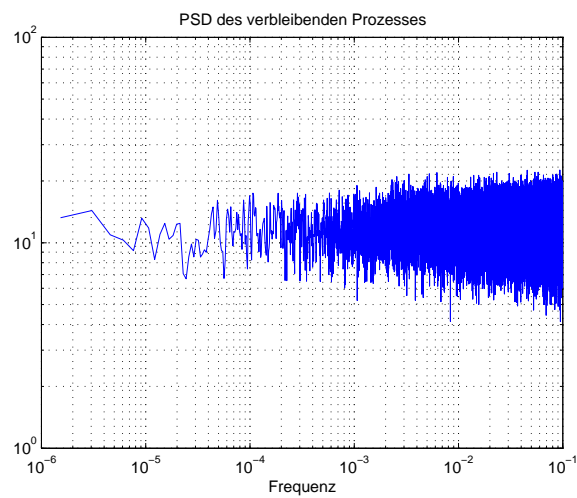


Abbildung 4.13: PSD der verbliebenen Sequenz (MA)

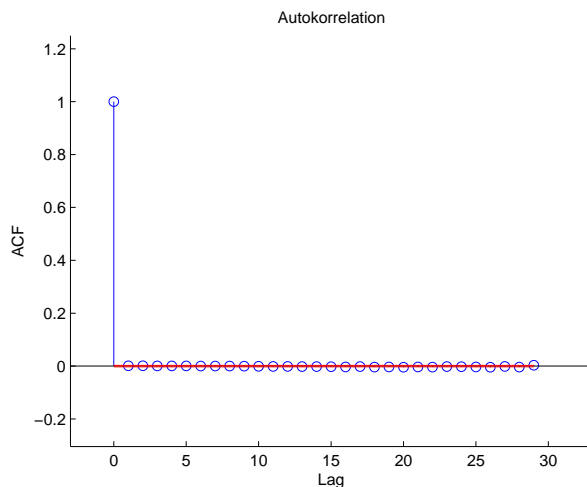


Abbildung 4.14: ACF der verbliebenen Sequenz (MA)

Die ACF bestätigt, dass die verbleibende Sequenz weißes Rauschen ist. Da die berechneten Koeffizienten sehr nahe an den ursprünglichen Koeffizienten liegen, kann gefolgert werden, dass die Residuen die gleichen stochastischen Eigenschaften wie die Beobachtungen haben.

4.5 ARMA[15,15]-Prozess als Fehlermodell

Der ARMA[15,15]-Prozess wurde mit den Koeffizienten aus den Tabellen 4.1 und 4.2 und einem weißen Rauschen mit der Varianz $\sigma^2 = 15.1398$ erzeugt.

4.5.1 Ergebnisse

Auch in diesem Fall weist die PSD-Funktion der Residuen in Abbildung 4.15 auf das Vorhandensein eines Prozesses hin.

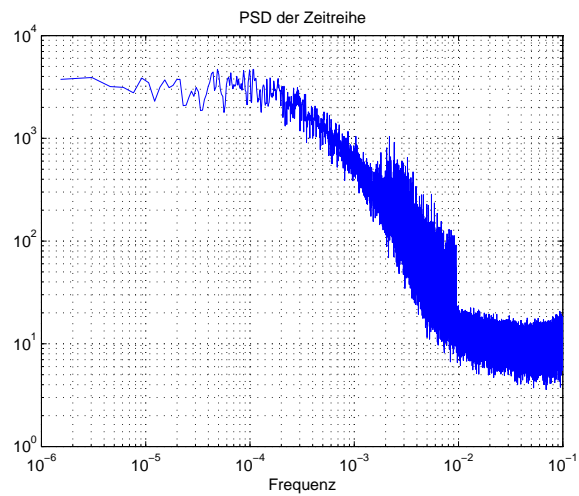


Abbildung 4.15: PSD der Residuen (ARMA)

Wiederum ist die Auffälligkeit bei ungefähr 0.01 Hz erkennbar. Dies wurde schon in Kapitel 4.3.1 erläutert. Für die Prozessidentifizierung müssen die ACF (Abbildung 4.16) und PACF (Abbildung 4.17) betrachtet werden.

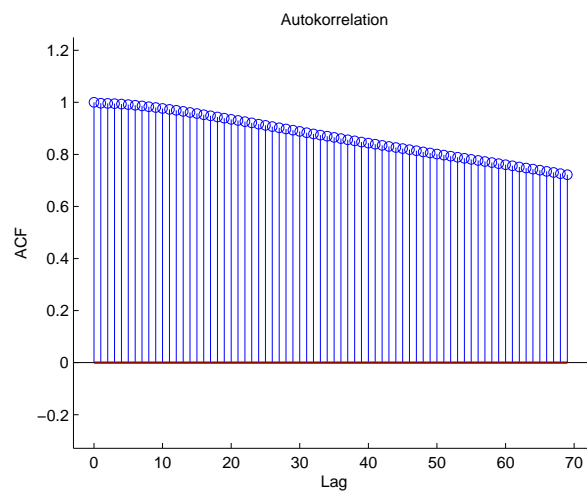


Abbildung 4.16: ACF der Residuen (ARMA)

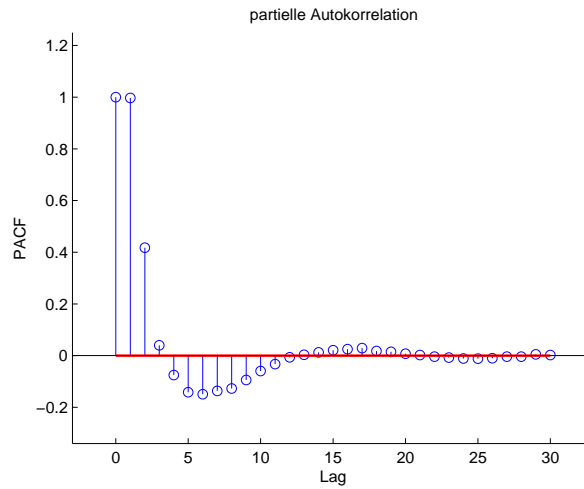


Abbildung 4.17: PACF der Residuen (ARMA)

Da in beiden Fällen kein Abbruchverhalten zu erkennen ist, wird der Prozess als ARMA-Prozess erkannt. Es wird im weiteren von einem ARMA[15,15]-Prozess ausgegangen. Es existieren verschiedene Verfahren zur Modellidentifikation, deren Beschreibung jedoch den Rahmen dieser Arbeit sprengen würden. In unserem Fall wird ein ARMA[15,15]-Prozess aus den Residuen geschätzt und die resultierenden Koeffizienten werden mit den Soll-Koeffizienten zur Erzeugung des Fehlermodells verglichen.

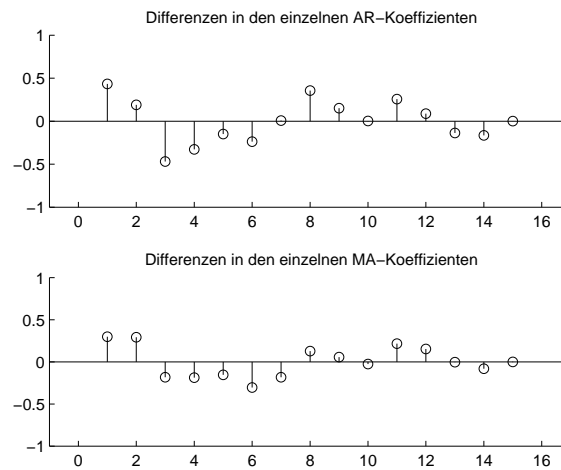


Abbildung 4.18: Differenzen (Soll-Ist) in den Koeffizienten (ARMA)

Die Differenzen in Abbildung 4.18 ergeben sich zwar bis zu 0.5, aber wichtiger ist, ob nach der Anwendung des inversen Prozesses, also nach der Herausrechnung des Prozesses aus den Residuen, weißes Rauschen übrig bleibt, oder noch immer ein Prozess enthalten ist. Hierzu

müssen wieder die PSD-Funktion und die ACF der verbleibenden Sequenz betrachtet werden.

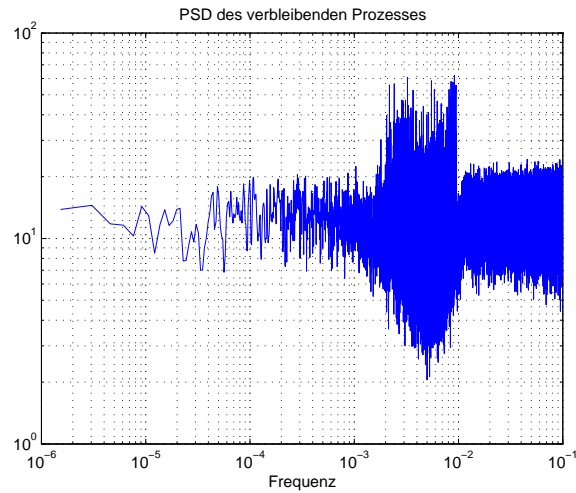


Abbildung 4.19: PSD der verbleibenden Sequenz (ARMA)

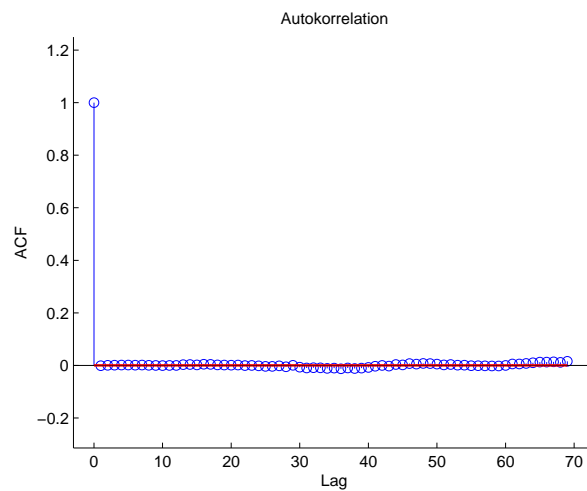


Abbildung 4.20: ACF der verbleibenden Sequenz (ARMA)

Ob verbleibende Sequenz die stochastischen Eigenschaften von weißem Rauschen hat, ist zwar mittels der PSD-Funktion in Abbildung 4.19 nicht eindeutig erkennbar, aber die ACF in Abbildung 4.20 bestätigt, dass dies der Fall ist. Der ursprüngliche ARMA[15,15]-Prozess kann laut diesen Ergebnissen auch durch einen anderen ARMA[15,15]-Prozess beschrieben werden. Ob die Residuen die gleichen stochastischen Eigenschaften wie die Beobachtungen haben, kann erst die Dekorrelation in Kapitel 5 klären.

Kapitel 5

Dekorrelation

5.1 Grundlagen

Bei der Lösung überbestimmter linearer Gleichungssysteme, wie sie in Kapitel 4.1 schon erwähnt worden sind, kann eine Dekorrelation durchgeführt werden. Die Dekorrelation hat zum Ziel, die korrelierten Beobachtungen in unkorrelierte Beobachtungen zu überführen. Dazu wird die Kovarianzmatrix durch eine Choleskyzerlegung aufgespalten in

$$\Sigma_{ll} = \mathbf{R}^T \mathbf{R},$$

wobei \mathbf{R} eine obere Dreiecksmatrix ist. Die Voraussetzung hierfür ist, dass Σ symmetrisch und positiv definit ist.

Damit kann ein neues GMM aufgestellt werden

$$\bar{\mathbf{I}} + \bar{\mathbf{v}} = \bar{\mathbf{A}}\mathbf{x} \quad , \quad E\{\bar{\mathbf{I}}\} = \bar{\mathbf{A}}\mathbf{x} \quad , \quad \mathbf{D}\{\bar{\mathbf{I}}\} = \bar{\Sigma}_{ll},$$

mit $\bar{\mathbf{I}} = (\mathbf{R}^{-1})^T \mathbf{I}$, $\bar{\mathbf{v}} = (\mathbf{R}^{-1})^T \mathbf{v}$, $\bar{\mathbf{A}} = (\mathbf{R}^{-1})^T \mathbf{A}$ und $\bar{\Sigma}_{ll} = (\mathbf{R}^{-1})^T \Sigma_{ll} \mathbf{R}^{-1} = \mathbf{I}$.

Die Choleskyzerlegung transformiert das korrelierte System in ein unkorreliertes System, denn aus der Beziehung $\bar{\Sigma}_{ll} = \mathbf{I}$ folgt, dass die Beobachtungen $\bar{\mathbf{I}}$ unkorreliert sind. Setzt man nun

$$\mathbf{F} = (\mathbf{R}^{-1})^T ,$$

so folgt unmittelbar $\bar{\mathbf{I}} = \mathbf{F}\mathbf{I}$, $\bar{\mathbf{v}} = \mathbf{F}\mathbf{v}$, $\bar{\mathbf{A}} = \mathbf{F}\mathbf{A}$.

Die Bestimmung von \mathbf{F} erfolgt mit Hilfe stationärer Prozesse. Diese Prozesse werden auch als (digitale) Filter bezeichnet. Für endliche Korrelationslängen ist \mathbf{F} eine Matrix mit Bandstruktur. In dieser Matrix stehen in jeder Zeile die Koeffizienten des Prozesses. Werden die Beobachtungen mit \mathbf{F} dekorreliert, muss aus Konsistenzgründen auch \mathbf{A} zu $\bar{\mathbf{A}}$ gefiltert werden. Die Matrix \mathbf{A} wird dabei spaltenweise gefiltert. Für diese Filterung müssen die dekorrelierten

Werte entsprechend der Korrelationslänge vorliegen.

Dies findet Anwendung bei sehr langen Zeitreihen, die zu sehr großen Matrizen führen. Durch diese Filterung kann eine Minimierung des Speicherbedarfs und der Rechengeschwindigkeit erreicht werden.

In der Geodäsie ergeben sich zum Beispiel in der Satellitengeodäsie sehr schnell sehr große Datenmengen bzw. Zeitreihen. Dadurch entstehen sehr große Gleichungssysteme, die als Ganzes nicht mehr effizient berechnet werden können. Für die Berechnung dieser Gleichungssysteme wird auf oben genanntes Verfahren zurückgegriffen. Dies läuft schneller und effizienter ab, da der Ausdruck $\mathbf{A}^T \Sigma_l^{-1} \mathbf{A}$ nicht mehr explizit ausgeführt werden muss.

Mit den Ergebnissen aus Kapitel 4 wurde eine Dekorrelation durchgeführt. Zusätzlich wurde eine Dekorrelation mit den Originalkoeffizienten aus den Tabellen 4.1 und 4.2 durchgeführt. Die entsprechenden Gravitationsfeldlösungen werden als Degree Error-RMS (DE-RMS) Kurven dargestellt gemäß

$$\text{DE-RMS}_l = \sqrt{\frac{\sum_{m=0}^l \left[(c_{lm}^{EGM96} - \hat{c}_{lm})^2 + (s_{lm}^{EGM96} - \hat{s}_{lm})^2 \right]}{2l + 1}}.$$

In dieser Formel bezeichnet l den Entwicklungsgrad des Erdschwerefeldes, m die Entwicklungsordnung des Erdschwerefeldes, c_{lm}^{EGM96} und s_{lm}^{EGM96} die Koeffizienten des EGM96-Modells und \hat{c}_{lm} und \hat{s}_{lm} die geschätzten Koeffizienten des Erdschwerefeldes. Das EGM96-Modell wird als Soll-Lösung verwendet, da es Grundlage der GOCE Simulation ist.

In Abbildung 5.1 sind die DE-RMS-Werte für die verschiedenen Lösungen aufgezeigt. Zu den simulierten Beobachtungen wird das jeweilige Fehlermodell addiert. Darauf folgt die Ausgleichung ohne Dekorrelation. Die Kurven stellen also den Einfluss der jeweiligen Beobachtungsfehler auf die Erdschwerefeldlösung ohne Berücksichtigung der Korrelationen zwischen den Beobachtungen dar. Die Kurven sind benannt nach den jeweiligen Fehlermodellen. Ausnahme hiervon ist die schwarze Kurve, welche den Signalgehalt des EGM96-Modells pro Grad wiedergibt.

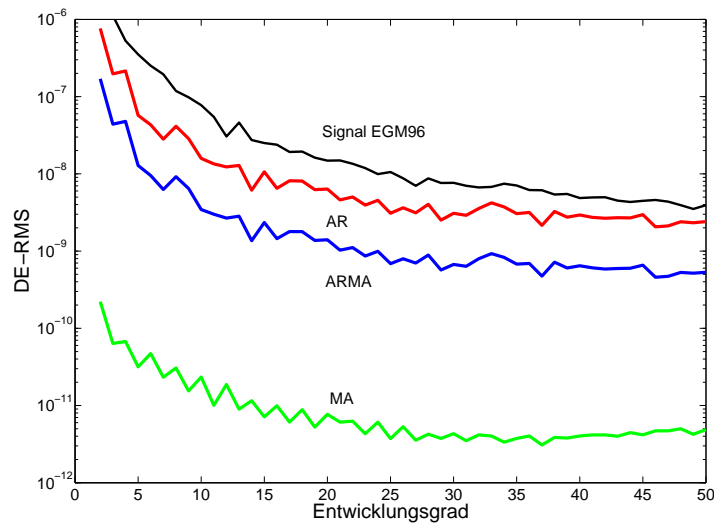


Abbildung 5.1: Fehlerkurven ohne Dekorrelation

5.2 AR-Prozess

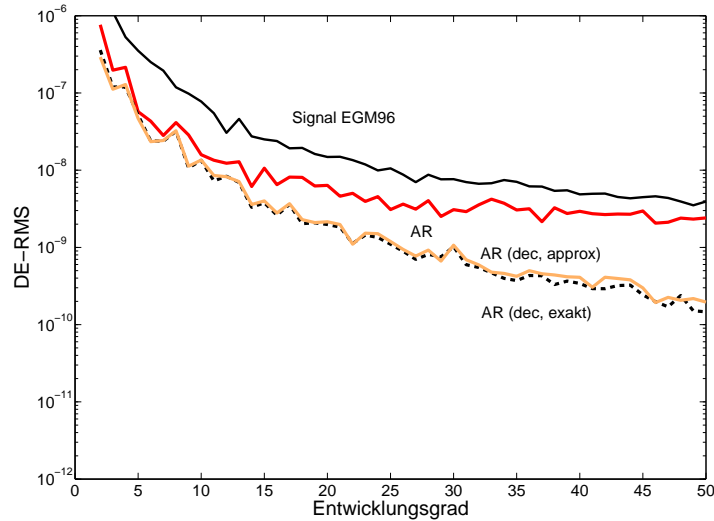


Abbildung 5.2: Fehlerkurven (AR)

Die orange Kurve in Abbildung 5.2 ergibt sich aus der Dekorrelation mit den geschätzten Koeffizienten. Die gestrichelte Kurve ergibt sich aus der Dekorrelation mit den exakten Koeffizienten. Dies ist der Idealfall. Die rote Kurve ist das Ergebnis ohne Verwendung der Dekorrelation. Es ist zu erkennen, dass sich die Ergebnisse der Schätzung der Erdschwerefeldkoeffizienten ver-

bessern. Da das Ergebnis für die Dekorrelation mit den geschätzten AR-Prozessparametern nahe am Idealfall liegt, kann man folgern, dass der geschätzte AR-Prozess den originalen AR-Prozess gut wiedergibt. Wie in Kapitel 4.3 gezeigt, modelliert der geschätzte AR-Prozess die stochastischen Eigenschaften des originalen AR-Prozesses nicht exakt wieder. Doch die Dekorrelationsergebnisse zeigen, dass die stochastischen Eigenschaften des geschätzten AR-Prozesses den stochastischen Eigenschaften des originalen AR-Prozesses sehr ähnlich sind. Dies zeigt, dass die Residuen die gleichen stochastischen Eigenschaften wie die Beobachtungen haben.

5.3 MA-Prozess

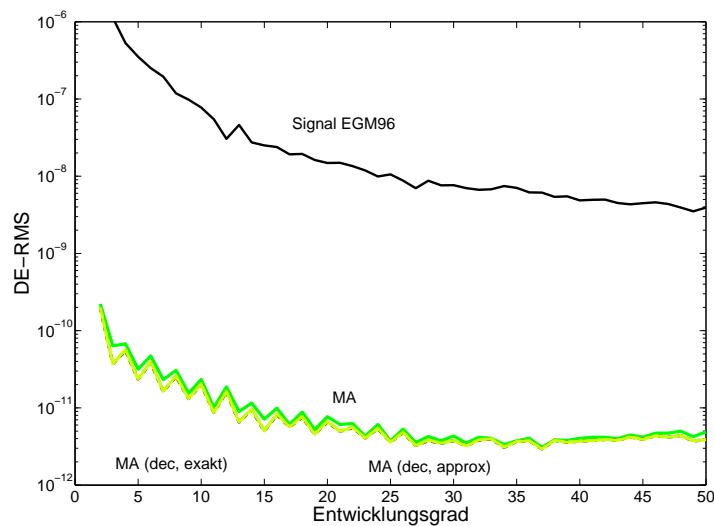


Abbildung 5.3: Fehlerkurven (MA)

Die hellgrüne Kurve in Abbildung 5.3 ergibt sich aus der Dekorrelation mit den geschätzten Koeffizienten. Die gestrichelte Kurve ergibt sich aus der Dekorrelation mit den exakten Koeffizienten. Die dunkelgrüne Kurve ist das Ergebnis ohne Verwendung der Dekorrelation. Hier liefern die Dekorrelationen keine nennenswerten Verbesserungen. Es gibt keinen erkennbaren Unterschied zwischen der Dekorrelation mit den geschätzten und den genauen Koeffizienten. In diesem Fall hat die Dekorrelation das Ergebnis der Analyse in Kapitel 4.4 bestätigt. Die Residuen besitzen die gleichen stochastischen Eigenschaften wie die Beobachtungen.

5.4 ARMA-Prozess

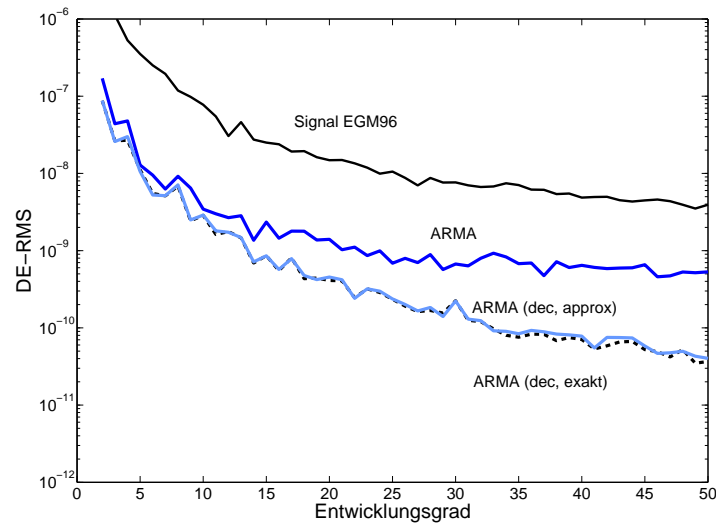


Abbildung 5.4: Fehlerkurven (ARMA)

Die hellblaue Linie in Abbildung 5.4 ergibt sich aus der Dekorrelation mit den geschätzten Koeffizienten. Die gestrichelte Linie ergibt sich aus der Dekorrelation mit den exakten Koeffizienten. Die dunkelblaue Kurve ist das Ergebnis ohne Verwendung der Dekorrelation. Beide Dekorrelationen liefern eine deutliche Verbesserung der Ergebnisse. Der Unterschied zwischen den entsprechenden Erdschwerefeldlösungen ist sehr gering. Die Residuen weisen die gleichen stochastischen Eigenschaften wie die Beobachtungen auf. Dieses Ergebnis zeigt, dass der in Kapitel 4.5 geschätzte ARMA-Prozess die gleichen stochastischen Eigenschaften wie der originale ARMA-Prozess aufweist. Daraus lässt sich folgern, dass verschiedene Prozesse die gleichen stochastischen Eigenschaften beschreiben. Die Prozessdefinition ist also nicht eindeutig.

Kapitel 6

Fazit

Das Hauptergebnis der vorliegenden Arbeit ist, dass sich (für den AR-Prozess in reduziertem Maße) die stochastischen Eigenschaften, mit denen die Beobachtungen für die Ausgleichung der Erdschwerefeldkoeffizienten versehen wurden, auch in den Verbesserungen wiederfinden. Dies macht die Residualanalyse zu einem Instrument, das stochastische Verhalten von Messzeitreihen zu modellieren. Das wichtigste Ergebnis in diesem Zusammenhang ist, dass sich der ARMA-Prozess in den Verbesserungen wiederfindet. Denn üblicherweise wird man von einem ARMA-Prozess als Rauschprozess ausgehen, da dies die allgemeine Form der Definition eines stationären Prozesses ist.

Ein weiteres Ergebnis dieser Studienarbeit ist, dass die Programme bzw. Algorithmen zur Modellierung stationärer Prozesse implementiert wurden und verlässlich arbeiten. Darüber hinaus wurden elementare signaltheoretische Grundlagen erläutert.

Abschließend ist zu sagen, dass es in dieser Studienarbeit mehrere Ansatzpunkte für weitere Arbeiten gibt. Als Beispiel sei die Bestimmung der Ordnung von AR-, MA- und ARMA-Prozessen genannt, oder aber die Erweiterungen der Prozessdefinitionen mit Hilfe von ARIMA (Autoregressive Integrated Moving Average)-Modellen.

Literaturverzeichnis

- ESA. Gravity field and steady-state ocean circulation mission. *Reports for Mission Selection of the four candidate Earth explorer missions*, European Space Agency SP-1233(1), 1999.
- J. Müller. Die Satellitengradiometriemission Goce. *Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften*, Reihe C, 541, 2001.
- J. Ries. Vergleichende Untersuchungen von Verfahren zur Schätzung von ARMA-Modellen zur Modellierung farbigen Rauschens. *Diplomarbeit an der Rheinische Friedrich-Wilhelms-Universität Bonn, Institut für Theoretische Geodäsie*, 2003.
- W.-D. Schuh. The Processing of Band-Limited Measurements; Filtering Techniques in the Least Squares Context and in the Presence of Data Gaps. *Space Science Reviews*, 108 (1): 67–78, 2003.
- W.-D. Schuh. Tailored numerical solution strategies for the global determination of the Earth's gravity field. *Mitteilungen der Universität Graz*, 81, 1996.
- A. Teusch. Einführung in die Spektral- und Zeitreihenanalyse mit Beispielen aus der Geodäsie. *Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften*, Reihe A, 120, 2006.
- G. Wilson. Factorization of the Covariance Generating Function of a Pure Moving Average Process. *SIAM Journal on Numerical Analysis*, Vol.6, No.1:1–7, 1969.