Persistenz und saisonale Abhängigkeiten in Abflüssen des Rheins

von

Michael Lohre und Philipp Sibbertsen
Fachbereich Statistik, Universität Dortmund, D-44221 Dortmund

Version Juli 2001

Zusammenfassung

Das Abflussverhalten des Rheins wird mittels flexibler saisonaler Modelle mit langem Gedächtnis modelliert. Zur Schätzung der Persistenz wird für jede Saisonfrequenz separat eine Log–Periodogramm Regression durchgeführt. Verglichen mit Standard–ARMA–Prozessen liefern diese Modelle eine gute Anpassung an das Langfristverhalten des Rheins. Langfristabhängigkeiten werden signifikant für die Jahres– und Halbjahresfrequenz geschätzt.

KEYWORDS: Saisonales langes Gedächtnis, Log-Periodogramm Regression, Rheinabfluss

1 Einleitung

Mitte der fünfziger Jahre entdeckte der Hydrologe Hurst (1951) anhand der jährlichen minimalen Wasserstände des Nils das Phänomen, dass der minimale Wert eines Jahres noch stark von weit zuvor beobachteten Werten beeinflusst wird. Er entdeckte, dass die Wasserstände nicht beliebig um den Mittelwert oszillieren, sondern dass auf einen hohen Wert weitere hohe Werte folgten und auf einen niedrigen Wert eine Phase mit niedrigen Wasserständen beobachtet wurde. Diese Persistenzeigenschaft von Abflüssen kann statistisch mit Zeitreihen mit langem Gedächtnis beschrieben werden. Mandelbrot/van Ness (1968) führten mit der fraktionalen Brownschen Bewegung ein erstes Modell zur Beschreibung derartiger Zeitreihen mit langem Gedächtnis ein. Die fraktionale Brownsche Bewegung ist eine Verallgemeinerung der Standard Brownschen Bewegung mittels selbstähnlicher Prozesse.

Allgemein wird ein stationärer Prozess X_t als Zeitreihe mit langem Gedächtnis bezeichnet, wenn die Bedingung

$$Cov(X_t, X_{t+k}) \quad \sim_{k\to\infty} \quad L_1(k)|k|^{-2d}, \quad d \in (0, 1/2),$$

gilt, wobei $L_1(k)$ eine für $|k| \longrightarrow \infty$ langsam variierende Funktion ist. Diese Bedingung charakterisiert die Haupteigenschaft von Zeitreihen mit langem Gedächtnis: der langsame Abfall der Korrelationen. Die Korrelationen sind nicht absolut summierbar. Dies ist ein entscheidender Unterschied zu der Standardannahme der Unabhängigkeit und zu Prozessen mit kurzem Gedächtnis, deren Korrelationen absolut summierbar sind. Dies kann in der Anwendung zu großen Schwierigkeiten führen, da kleine, aber schwach abfallende Korrelationen oft von den Standardtests nicht erkannt werden, aber trotzdem einen großen Einfluss auf die Datenanalyse haben können. So ist zum Beispiel der Standardfehler des arithmetischen Mittels nicht mehr von der Ordnung $n^{-\frac{1}{2}}$, sondern er muss durch $n^{-\alpha}$, $0 < \alpha < \frac{1}{2}$, ersetzt werden.

Äquivalent zu der obigen Definition können Zeitreihen mit langem Gedächtnis auch mittels Eigenschaften der Spektraldichte f definiert werden. Dabei wird ausgenutzt, dass die kleinen Frequenzen der Spektraldichte das Langfristverhalten des Prozesses bestimmen. Demnach heißt ein stationärer Prozess X_t Zeitreihe mit langem Gedächtnis, falls

$$f(\lambda) \quad \sim_{|\lambda|\to 0} \quad L_2(\lambda)|\lambda|^{-2d}, \quad d \in (0, 1/2),$$

gilt, wobei $L_2(\lambda)$ wiederum langsam variierend ist für $|\lambda| \longrightarrow 0$. Aus dieser Definition erhält man eine weitere Eigenschaft von Zeitreihen mit langem Gedächtnis: Die Spektraldichte hat einen Pol am Ursprung, d.h. für Frequenzen gegen Null geht die Spektraldichte gegen unendlich.

Neben der fraktionalen Brownschen Bewegung fanden Granger/Joyeux (1980) und unabhängig davon Hosking (1981) mit einer Verallgemeinerung der klassischen ARMA/ARIMA–Prozessklasse (Box/Jenkins (1976)) zu sogenannten ARFIMA–Prozessen einen weiteren Zugang zur Modellierung von Zeitreihen mit langem Gedächtnis. Ein ARFIMA–Prozess ist definiert als Lösung der Gleichung

$$\Phi(B)(1-B)^d X_t = \Psi(B)\varepsilon_t, \tag{1}$$

wobei B der Backshift Operator ist, $\Phi(x)$ und $\Psi(x)$ die AR– und MA– Polynome sind und ε_t ist Weißes Rauschen (vgl. Box/Jenkins (1976)). Im Gegensatz zur Standardannahme mit ganzzahligem d nimmt im Falle von langem Gedächtnis der Parameter d Werte aus dem Intervall $(0, 1/2)$ an. Der Term $(1-B)^d$ ist für fraktionales d dabei wie folgt definiert:

$$(1-B)^d = \sum_{k=0}^{d} c_k (-1)^k B^{d-k},$$

mit den aus der Gammafunktion abgeleiteten Koeffizienten

$$c_k := \binom{d}{k} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}. \tag{2}$$

Geweke/Porter-Hudak (1983) konnten die Äquivalenz der beiden Ansätze zeigen. In der Literatur sind diverse Schätzverfahren für die Stärke des langen Gedächtnisses (die Größe des Parameters d im ARFIMA–Modell) vorgeschlagen worden.

Mandelbrot/Wallis (1969) zeigten die gute Anpassung dieses Modells an die minimalen Wasserstände des Nils. Auch für andere Flüsse (u.a. Sankt–Lorenz–Strom, Loire, Mississippi) konnte eine sehr gute Modellanpassung nachgewiesen werden. Mit Hilfe der adjustierten Spannweite, einem heuristischen Schätzverfahren für den Gedächtnisparameter, schätzten Mandelbrot/Wallis langes Gedächtnis in den monatlichen bzw. jährlichen Abflüssen der obigen Flüsse. Zeitreihen mit langem Gedächtnis scheinen daher ein geeignetes Modell für Abflussdaten zu sein.

Seit der mathematischen Formulierung durch Mandelbrot/van Ness (1968) konnten zahlreiche weitere Zeitreihen mit langem Gedächtnis in der Praxis modelliert werden. Neben der Hydrologie spielen diese Modelle vor allem in der empirischen Kapitalmarktforschung eine große Rolle. Eine Übersicht findet man z.B. bei Beran (1994) und Sibbertsen (1999). Das Problem, Abflussdaten geeignet zu modellieren, wurde von Lawrence/Kottegoda (1977) intensiv diskutiert. Sie fassen die statistischen Eigenschaften hydrologischer Zeitreihen zusammen und diskutieren die Vorteile von Modellen mit kurzem und langem Gedächtnis. Hosking (1984) beschreibt die Modellierung hydrologischer Zeitreihen mittels fraktionaler ARIMA–Modelle. Einen allgemeinen Überblick über statistische Methoden in der Hydrologie findet sich in Hipel/McLeod (1994), wobei dort der Schwerpunkt auf modellgestützten Techniken liegt, darunter Modelle mit langem Gedächtnis oder saisonale Modelle. Bedingt durch die Entwicklung verbesserter Schätzverfahren hat in den letzten Jahren eine Wiederbelebung dieses Gebiets stattgefunden (z.B. Montanari et al. 1997, 2000).

Im Vergleich zu anderen Flüssen konnten Mandelbrot/Wallis (1969) für den Rhein bei der Analyse der monatlichen Abflüsse am Pegel Basel kein persistentes Verhalten nachweisen. Dies kann in Eigenschaften des Pegels Basel begründet liegen. So wird der Rhein bei Basel im Wesentlichen durch den Hochgebirgseinfluss bestimmt mit geringen Abflüssen im Winter und hohen Abflüssen in den Sommermonaten verursacht durch die Schneeschmelze im späten Frühjahr und Frühsommer. Außerdem verursachen sowohl der Bodensee als auch die Schweizer Seen wie der Bieler See oder der Zürichsee durch Seeretention eine Veränderung des Abflussverhaltens. Jedoch sollte gerade die Seeretention aus hydrologischen Gesichtspunkten das Phänomen eines persistenten Abflusses auch beim Rhein bewirken. Demnach können zwei Gründe für das negative Ergebnis in Mandelbrot/Wallis (1969) ausschlaggebend sein: zum einen mag dies in den Eigenschaften des Pegels Basel begründet liegen. Ein anderer Grund könnte auch in dem verwendeten Verfahren liegen: bei der Analyse von Monatsmitteln muss die Saisonstruktur vorab eliminiert werden.

In der vorliegenden Arbeit wird der Pegel Kaub vor der Moselmündung und der Pegel Köln hinter der Moselmündung betrachtet. Gerade für den Pegel Köln ist zu erwarten, dass der Hochgebirgseinfluss durch den Mittelgebirgscharakter überlagert wird. Dieser bringt insbesondere im Winter hohe, durch Regen– und Schneefälle verursachte Abflüsse sowie niedrige Abflüsse im Sommer. Einen erwartet starken Einfluss ist ferner auch der Mosel und ihrem Einzugsgebiet zuzuschreiben. Der Untergrund ist hier eher steiniger Natur mit einer nur geringen Bodenauflage, in der Regenwasser gespeichert werden kann. Hohe Abflüsse der Mosel können damit auch noch den Pegel Köln beeinflussen. Auch für den Pegel Kaub wäre es nach den Ergebnissen dieser Arbeit möglich, dass der Mittelgebirgscharakter den Hochgebirgseinfluss bereits dominiert. Auch hierdurch können abweichende Ergebnisse zu Mandelbrot/Wallis (1969) zu erklären sein. Für einen detaillierten Überblick über das Abflussverhalten an den einzelnen Rheinpegeln sei auch auf die Rhein-Monographie verwiesen (Gerhard (1976)).

Es wird daher erneut das Abflussverhalten des Rheins untersucht, diesmal allerdings an den Pegeln Kaub und Köln. Es werden monatliche maximale, mittlere und minimale Abflüsse betrachtet. Die Abhängigkeitsstruktur wird mittels einer Periodogramm-basierten Schätzmethode geschätzt. Vorteil dieses Verfahrens ist, dass weder Kenntnisse über das Kurzfristverhalten noch über die genaue Verteilung der Abflüsse erforderlich sind. Allein anhand der Struktur des Periodogramms können Langfristabhängigkeiten erkannt werden. Ein weiterer Vorteil dieses Verfahren besteht darin, dass eine Eliminierung der Saisonstruktur vorab nicht erfolgen muss. Das scheint insbesondere sinnvoll vor dem Hintergrund, dass verschiedene Einflüsse (Hochgebirgseinfluss vs. Mittelgebirgscharakter) das Regime in ihrer Struktur unterschiedlich beeinflussen. Die Arbeit zeigt, dass in der Tat ein saisonales Modell mit langem Gedächtnis eine gute Anpassung an alle betrachteten Abflussdaten liefert.

2 Verwendete Daten

Die Bundesanstalt für Gewässerkunde in Koblenz archiviert die täglichen Abflüsse des Rheins an verschiedenen Pegeln. Um Vergleiche zwischen den Pegeln treffen zu können, wird für beide Pegel derselbe Untersuchungszeitraum gewählt: sowohl für den Pegel Kaub als auch für den Pegel Köln liegen Daten von Januar 1900 bis Dezember 1998 zugrunde. Bei beiden Pegeln wurden jeweils die tatsächlichen Monatsspitzen sowie aus Tageswerten berechnete Monatsmittel und Monatsminima verwendet. Somit ergeben sich Reihenlängen von 1188 Beobachtungen für die Pegel Kaub und Köln. Anstelle der Originalreihen werden jedoch die logarithmierten Reihen betrachtet, um eventuelle Einflüsse, die sich durch eine hohe Prozessvarianz ergeben können, auszuschließen. Durch den Übergang zu logarithmierten Daten wird die Varianz der betrachteten Zeitreihen verkleinert, einen Einfluss auf die Schätzung der Abhängigkeitsstruktur hat diese Transformation nicht. Abbildung 1 zeigt die logarithmierten Reihen.

Der Hauptgrund, sowohl den Pegel Kaub als auch den Pegel Köln zu betrachten, ist der erwartete starke Einfluss der Mosel auf das Abflussverhalten des Rheins. Durch die weitgehend geringe Bodenauflage fließt der Niederschlag aus dem Einzugsgebiet der Mosel schnell in den Rhein ab. Dieser Effekt sollte den Abfluss, der aus Hochgebirgseinflüssen entstanden ist, überlagern, d.h. der Moseleinfluss sollte am Pegel Köln sichtbar sein. Auch der Pegel Kaub, der vor der Moselmündung liegt, wird durch Zuflüsse wie Neckar und Main beeinflusst. Allerdings ist bei Betrachtung der jahresdurchschnittlichen Maxima ersichtlich, dass auch der Hochgebirgseinfluss noch vorhanden ist (Gerhard (1976)), d.h. im Jahresverlauf treten zwei lokale Maxima auf. Es muss also davon ausgegangen werden, dass die Struktur der durchschnittlichen Maxima bzw. Minima an den beiden Pegeln unterschiedlich sein wird. Aufgrund dieser Vermutungen erscheint es sinnvoll ein saisonales Modell an die Daten anzupassen. Dies wird auch durch die Autokorrelationsfunktionen der Zeitreihen unterstützt, die das (für Monatsdaten) typische Bild einer saisonalen Struktur mit langem Gedächtnis aufweisen (Abbildung 2). Sie haben zum einen einen sinusoidalen

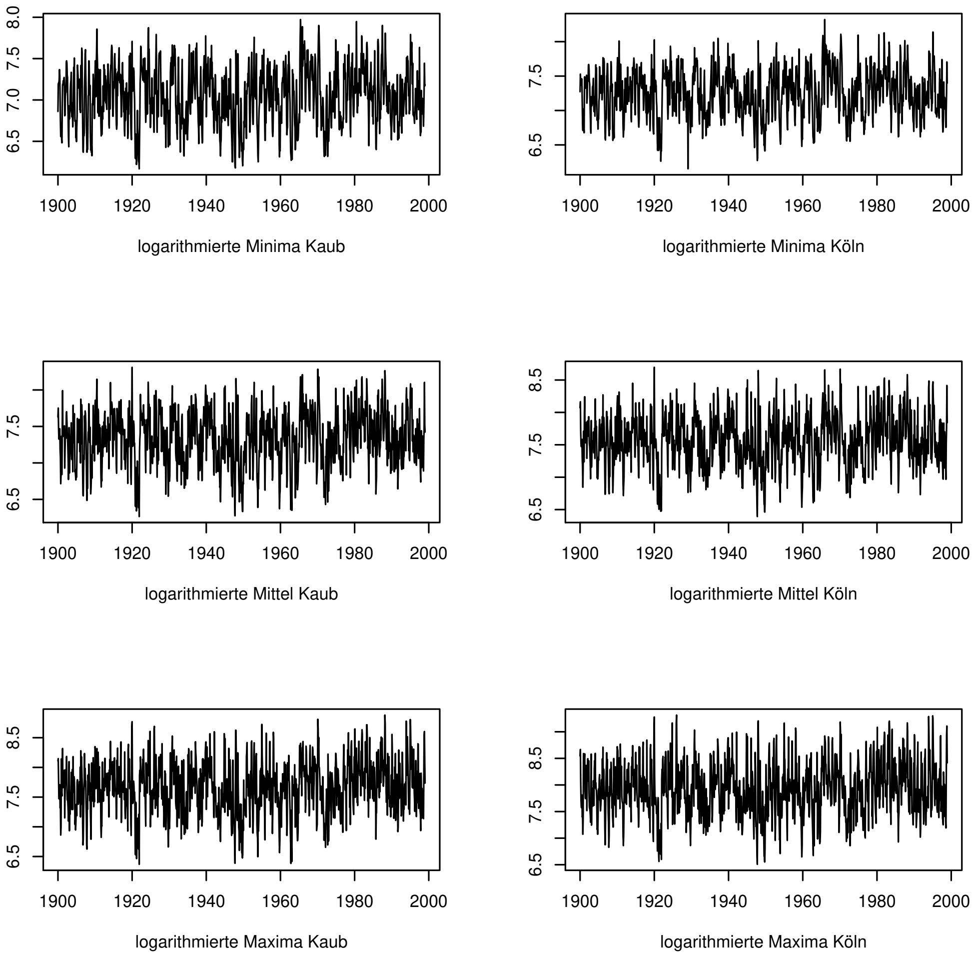

Abbildung 1: Logarithmierte Reihen des minimalen, mittleren und maximalen Abflusses am Pegel Kaub (linke Spalte) und Köln (rechte Spalte)

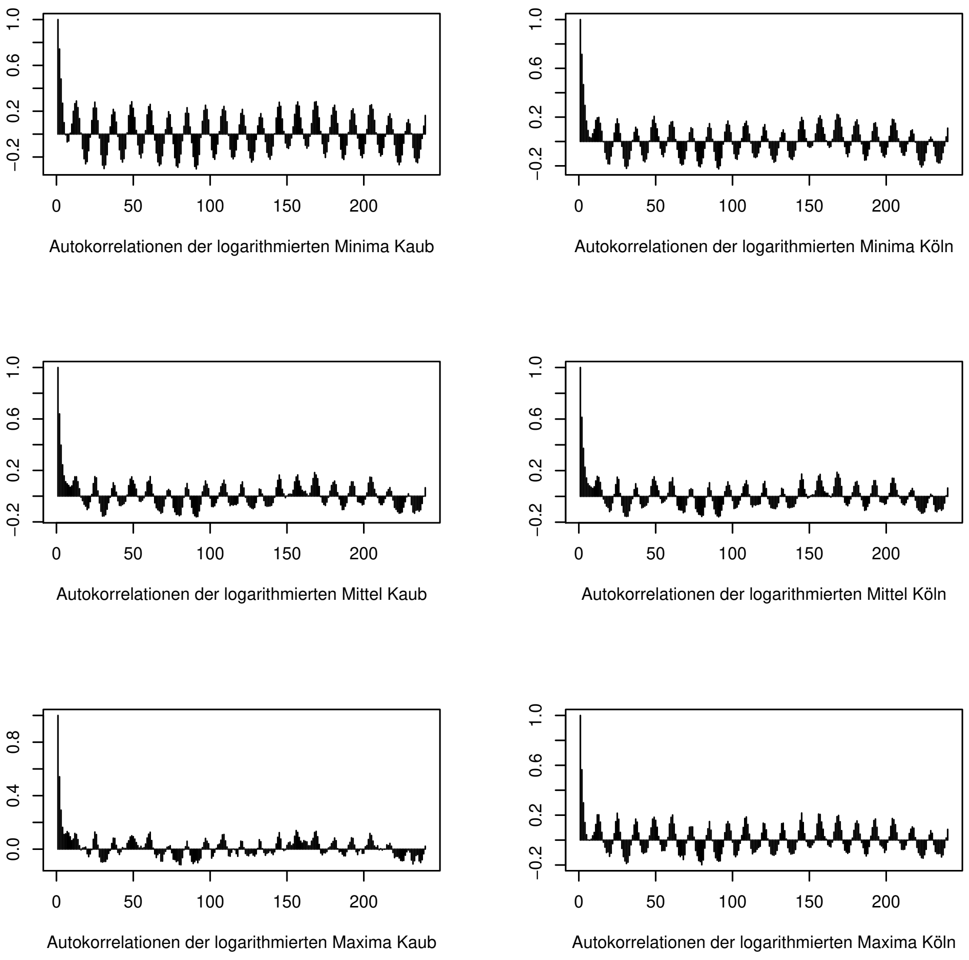

Abbildung 2: Autokorrelationen der minimalen, mittleren und maximalen Abflüsse an den Pegeln Kaub (links) und Köln (rechts)

Verlauf und zum anderen langsam abfallende Autokorrelationen, also typische Charakteristika von saisonalen Reihen mit langem Gedächtnis.

3 Bestimmung von Saison und Persistenz

Als Ausgangspunkt einer statistischen Analyse werden in der Regel kovarianzstationäre Zeitreihen betrachtet, d.h. Zeitreihen mit konstantem Mittelwert und einer Kovarianz, die nicht von den Werten selbst sondern nur vom Zeitabstand zweier Beobachtungen abhängt. Derartige Zeitreihen enthalten weder einen Trend noch eine Saisonstruktur. Um Stationarität in Zeitreihen mit Monatsdaten zu erhalten, wird von jeder Beobachtung das zugehörige Monatsmittel abgezogen und durch die Monatsstandardabweichung dividiert. Eine derartige Saisoneliminierung wurde auch von Mandelbrot/Wallis (1969) vorgenommen. Eine weitere Möglichkeit, Saisonstrukturen aus einer Zeitreihe mit Monatsdaten zu eliminieren, besteht in der Differenzierung der Zeitreihe, das heißt es werden die zwölften Differenzen gebildet:

$$(1 - B^{12}) \ln X_t = \ln X_t - \ln X_{t-12}.$$

Diese Differenzenbildung ist aber nicht immer sinnvoll, wie beispielhaft das Periodogramm der differenzierten Reihe der mittleren Abflüsse am Pegel Köln zeigt (Abbildung 3): Um die Saisonfrequenzen $\pi/6$ (Jahresfrequenz), $\pi/3$ (Halbjahresfrequenz), $\pi/2$ (Quartalsfrequenz), $2\pi/3$ (Dreimonatsfrequenz) und $5\pi/6$ (Zweimonatsfrequenz) herum erkennt man ,,Löcher" im Periodogramm. Das deutet auf eine Überdifferenzierung der Reihe hin, mit anderen Worten der Grad der Differentiation wurde mit $d = 1$ zu hoch angesetzt. Anhand des Periodogramms erscheint also ein fraktionaler Differenzenfilter mit $d \in (0; 1)$ geeigneter zu sein, um die Saisonstruktur zu eliminieren. Auf eine Vorab–Transformation der Reihe wird daher verzichtet, um eine möglicherweise falsche Modifizierung zu vermeiden. Lawrance/Kottegoda (1977) weisen ferner daraufhin, dass Trend bzw. Saison klassische Eigenschaften einer hydrologischen Zeitreihe sind und es daher fragwürdig ist, ohne genaues Wissen über die Struktur diese a priori zu eliminieren. Aus diesem Grund

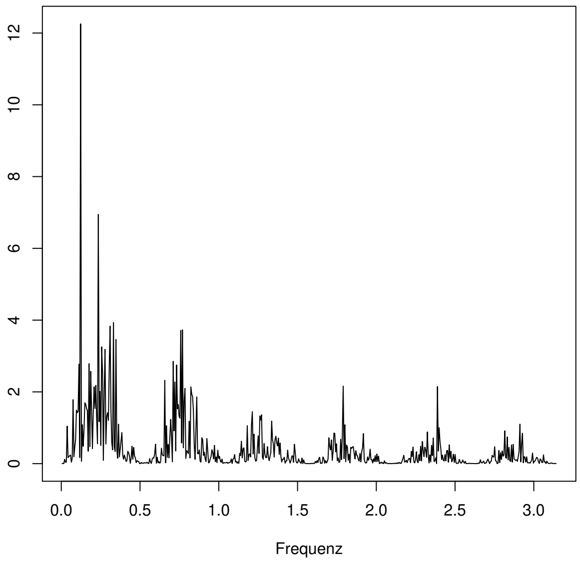

Abbildung 3: Periodogramm der Reihe $\ln X_t - \ln X_{t-12}$ für mittlere Abflüsse am Pegel Köln

wird ein fraktional integriertes saisonales ARMA–Modell gewählt, eine Erweiterung des Modells (1) um eine saisonale Komponente. Ein sogenanntes ARFISMA$(0, d, 0)$–Modell ist Lösung der Gleichung

$$(1 - B^s)^d \ln X_t = u_t, \tag{3}$$

wobei s die Saison, d den Grad der Differentiation bezeichnet und u_t ein stationärer Prozess mit beschränkter Spektraldichte (z.B. ein gewöhnlicher ARMA(p, q)–Prozess) ist. Bei dieser Art der Modellierung handelt es sich allerdings noch um ein starres Modell, da der Einfluss der verschiedenen saisonalen Abhängigkeiten und des Langzeitgedächtnisses lediglich durch einen einzigen Parameter d modelliert wird. Porter–Hudak (1990) nutzt ein solches Modell zur Modellierung von Geldmengen und Ray (1993) prognostiziert den monatlichen Absatz eines bestimmtes Produkts bei Zugrundelegung eines erweiterten Modells der Form (3). Für diese Zwecke wird ein alternativer Ansatz betrachtet, der eine flexiblere Gedächtnismodellierung ermöglicht und im Wesentlichen auf Hassler (1994) zurückgeht. Betrachten wir zunächst eine Faktorisierung des Filters

$$(1 - B^s) = (1 - B)(1 + B) \prod_{j=1}^{s/2-1} (1 - e^{i2\pi j/s} B)(1 - e^{-i2\pi j/s} B).$$

Dann kann ein flexibler saisonaler Filter mittels Faltung definiert werden durch

$$\Delta_{\delta_s} := (1 - B)^{d_0}(1 + B)^{d_{s/2}} \prod_{j=1}^{s/2-1} \{(1 - e^{i2\pi j/s} B)(1 - e^{-i2\pi j/s} B)\}^{d_j}. \tag{4}$$

Andel (1986) zeigt, dass jeder Faktor aus (4) durch

$$\begin{aligned} (1 - e^{i\omega} B)^d (1 - e^{-i\omega} B)^d &= \sum_{l=0}^{\infty} \beta_l B^l \\ \beta_l &= \sum_{k=0}^{l} c_{l-k} c_k \cos\{(l - 2k)\omega\}, \end{aligned} \tag{5}$$

dargestellt werden kann, wobei die Koeffizienten c_k gerade dem Binomialkoeffizienten (2) entsprechen. Über die Koeffizienten c_k sind demnach auch die Gedächtnisparameter festgelegt und man gelangt somit zu einem flexiblen ARFISMA$(0, \delta_s, 0)$–Prozess der Form

$$\Delta_{\delta_s} \ln X_t = u_t. \tag{6}$$

Der Einfluss jeder Saisonfrequenz kann demnach durch verschiedene Gedächtnisparameter modelliert werden. Das starre Modell ergibt sich als Spezialfall gerade dann, wenn alle Parameter d_j in (4) gleich sind. In Anlehnung an die Eigenschaften stochastischer Prozesse mit langem Gedächtnis gilt für derartige ARFISMA–Modelle:

- Der Prozess ist stationär dann und nur dann, wenn $d_j < 0,5$, $j = 0, 1, \ldots, s/2$.

- Der Prozess hat langes Gedächtnis, wenn $d_j > 0$ ist für mindestens ein $j \in [0; s/2]$.

- Für $d_j > 0$ hat das Spektrum an der Saisonfrequenz $2\pi j/s$ einen Pol, für $d_j < 0$ eine Nullstelle, $j = 0, 1, \ldots, s/2$.

Eine zentrale Bedeutung hat demnach die Spektraldichte eines solchen Prozesses, da eine Polstelle an einer Saisonfrequenz auf langes Gedächtnis hindeutet. Für den zuvor beschriebenen Prozess lässt sich die Spektraldichte f allgemein darstellen durch (siehe Hassler (1994))

$$\begin{aligned} f_X(\lambda) &= \left\{4\sin^2(\frac{\lambda}{2})\right\}^{-d_0} \left\{4\sin^2(\frac{\lambda}{2} - \frac{\pi}{2})\right\}^{-d_{s/2}} \\ &\quad \times \prod_{j=1}^{s/2-1} \left\{4\sin^2(\frac{\lambda}{2} + \frac{\pi j}{s}) 4\sin^2(\frac{\lambda}{2} - \frac{\pi j}{s})\right\}^{-d_j} f_u(\lambda). \end{aligned} \tag{7}$$

In der Umgebung der j–ten Saisonfrequenz lässt sich dann Gleichung (7) approximieren durch

$$f_X(2\pi j/s + \lambda) \quad \sim_{|\lambda|\to 0} \quad \lambda^{-2d_j} K_j \, f_u(2\pi j/s), \qquad j = 0, 1, \ldots, s/2, \tag{8}$$

d.h. die Spektraldichte um die j–te Saisonfrequenz hängt von der gegen Null konvergierenden Frequenz λ, einer Konstanten K_j und der beschränkten Restspektraldichte f_u ab. Für positives d_j ist die rechte Seite von (8) unbeschränkt. Wir wollen diese Eigenschaft von Prozessen mit saisonalem langen Gedächtnis nutzen, um mit Hilfe des empirischen Analogons der Spektraldichte die Gedächtnisparameter zu schätzen. Dazu verwenden wir das Periodogramm einer Zeitreihe, das definiert ist durch

$$I_X(\lambda_j) = \frac{1}{n} |\sum_{t=1}^{n} X_t e^{i\lambda_j t}|^2, \qquad \lambda_j = \frac{2\pi j}{n}, \quad j = 1, \ldots, n/2.$$

Eine auf dem Periodogramm basierende Schätzprozedur wurde zuerst von Geweke/Porter–Hudak (1983) für nicht saisonale Prozesse vorgeschlagen, Hassler (1994) und später Arteche/Robinson (2000) übertragen das Verfahren auf saisonale Prozesse. Dieses, auch Log–Periodogramm–Regression genannte Verfahren basiert im Wesentlichen auf dem Logarithmus der Gleichung (8) sowie der beidseitigen Addition des logarithmierten Periodogramms und resultiert bei saisonalen Prozessen in folgender Regressionsgleichung:

$$\ln(I_X(\lambda_{j;i})) = a_j + d_j R_i + \epsilon_{j;i}, \quad j = 0, 1 \ldots, s/2, \quad i = 1, \ldots, m. \tag{9}$$

Dabei ist das logarithmierte Periodogramm die abhängige Variable, a_j eine Konstante, die von der Restspektraldichte f_u an der j–ten Saisonfrequenz abhängt, d_j ist der zur j–ten Saisonfrequenz zugehörige Gedächtnisparameter und $R_i = -\ln(4\sin^2(\pi i/n))$ sind die Regressoren. Mit $\epsilon_{j;i}$ ist ferner der Fehlerterm bezeichnet. Mit Hilfe der Gleichung (9) kann nun jeder Gedächtnisparameter d_j einzeln geschätzt werden.

Abbildung 4 zeigt die Periodogramme der verschiedenen Reihen. Alle Abbildungen weisen einen Peak am Ursprung und an den ersten beiden Saisonfrequenzen auf. Diese Peaks können möglicherweise auf eine Gedächtnisstruktur hinweisen. Hinzu kommt, dass die Höhen der einzelnen Peaks unterschiedlich sind und demnach ein flexibles saisonales Modell, das unterschiedliche Saisoneinflüsse durch variable Modellparameter ermöglicht, wie oben beschrieben eine geeignete Anpassung liefern sollte. Anhand der Regressionsgleichung (9) kann der Gedächtnisparameter d_j somit als Steigungsparameter einer linearen Regression interpretiert werden, die an m aufeinanderfolgenden Fourierfrequenzen ausgewertet wird. Die Länge des Frequenzbandes ist dabei keineswegs fest vorgeben sondern vom Anwender festzulegen. Da die Regressionsgleichung auf der Approximation (8) beruht und diese nur in einer kleinen Umgebung um die Saisonfrequenz gültig ist, führt ein großes Frequenzband zu einem Bias in der Schätzung von d_j. Andererseits resultiert die Verwendung nur weniger Fourierfrequenzen in einer Vergrößerung der Varianz. Eine ,,optimale" Bandbreitenwahl ist also nur in dem Sinne möglich, dass man die Summe aus Verzerrung und Varianz minimiert. Eine Zusammenfassung der theoretischen Aspekte findet man bei Hurvich et al. (1998). Arteche/Robinson (2000) zeigen, dass unter zusätzlichen Annahmen für Gaussprozesse gilt:

$$\sqrt{m}(\hat{d}_j - d_j) \quad \to_d \quad \mathcal{N}(0, \pi^2/24),$$

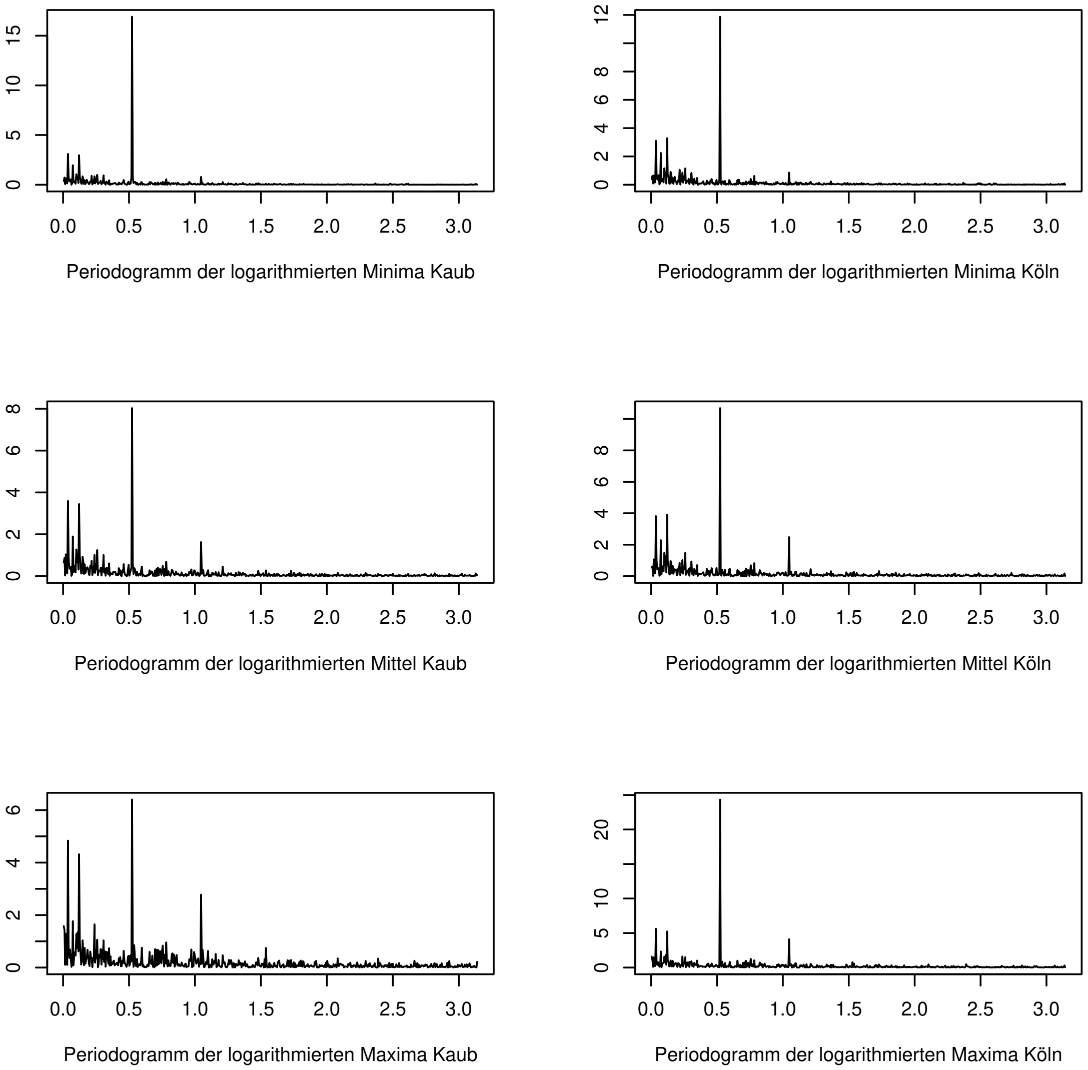

Abbildung 4: Periodogramm der minimalen, mittleren und maximalen Abflüsse an den Pegeln Kaub (linke Spalte) und Köln (rechte Spalte)

d.h. der Log-Periodogramm-Schätzer konvergiert gegen eine Normalverteilung. Trotz der Trade–Off–Situation bei der Bandbreitenwahl ist die semiparametrische Log–Periodogramm–Regression ein weit verbreitetes Verfahren, da es im Vergleich zu den parametrischen Alternativen ohne weitere Annahmen an den zugrundeliegenden Prozess auskommt. So ist z.B. bei parametrischen Verfahren immer eine Vorab–Spezifikation der autoregressiven und moving–average Ordnungen notwendig. Eine Fehlspezifikation wiederum resultiert in verzerrten Schätzern. Die Log–Periodogramm-Regression benötigt im Vergleich lediglich die Beschränktheit der Restspektraldichte. Zudem wird für die Log-Periodogram-Regression keine Verteilungsannahme an den zugrundeliegenden Prozess benötigt. Parametrische Verfahren wie der Maximum-Likelihood Schätzer gehen von normalverteilten Daten aus. Es ist aber bekannt, dass Abflussdaten einer schiefen Verteilung folgen und z.B. hinreichend gut durch verschiedene Formen einer Gamma–Verteilung modelliert werden können. Diese Verteilungen sind nicht symmetrisch und haben somit eine andere Form als die Normalverteilung. Diese Tatsache spricht ebenfalls für eine semiparametrische Modellierung im Vergleich zu parametrischen Alternativen.

Für die Log–Periodogramm–Regression wurde in unserem Fall die maximale Zahl an möglichen Fourierfrequenzen gewählt. Für die Pegel Kaub und Köln hat jede Reihe $n = 1188$ Beobachtungen, das heißt es können insgesamt $n/2 = 594$ Fourierfrequenzen berechnet werden. Da zwischen zwei Saisonfrequenzen jeweils $n/(2 * 6) = 99$ Fourierfrequenzen liegen, wählen wir für jede Regression $m = 49$ Fourierfrequenzen, um den Einfluss benachbarter Saisonfrequenzen zu vermeiden. Die Ergebnisse der einzelnen Log–Periodogramm–Regressionen sind in Tabelle 1 zusammengefasst. Der Standardfehler für die Schätzungen ergibt sich aus der obigen Verteilungseigenschaft des Log-Periodogramm-Schätzers und liegt in der Größenordnung von 0,0916. Auch im Falle nicht normalverteilter Daten liefert die Verteilungsaussage zumindest eine gute Approximation. Dies wird von Montanari et al. (2000) für einen alternativen Schätzansatz gezeigt. Die Argumente sind auf den hier vorgestellten Ansatz übertragbar.

Pegel				Saisonfrequenzen			
Kaub	0	$\pi/6$	$\pi/3$	$\pi/2$	$2\pi/3$	$5\pi/6$	π
Minima	0.0336	0.2748	0.0969	0.0569	-0.0022	0.0714	-0.0670
Mittel	0.0313	0.2195	0.1824	0.0254	-0.0321	-0.1639	0.0334
Maximum	0.1074	0.0912	0.2443	-0.0633	0.1180	-0.0908	-0.0840
Köln	0	$\pi/6$	$\pi/3$	$\pi/2$	$2\pi/3$	$5\pi/6$	π
Minimum	0.0512	0.2093	0.0768	0.0518	0.0468	0.0904	0.0846
Mittel	-0.0102	0.3000	0.1883	-0.0132	0.0179	0.0151	0.0111
Maximum	0.0865	0.2245	0.2175	-0.0711	0.1328	-0.0608	0.1561

Tabelle 1: Geschätzte Gedächtnisparameter für die verschiedenen Reihen und Saisonfrequenzen

Die Schätzung liefert zwei Ergebnisse: zunächst deuten die Schätzwerte auf das Vorliegen einer Gedächtnisstruktur hin, da für jede betrachtete Reihe mindestens ein Parameter signifikant größer Null ist. Für die minimalen und mittleren Abflüsse beider Pegel ist der Wert an der Jahresfrequenz $\pi/6$ maximal, das heißt die Ergebnisse deuten auf ein saisonales langes Gedächtnis hin. Das bedeutet, dass z.B. ein geringer Abfluss in einem bestimmten Monat eines Jahres (verglichen mit dem mittleren Abfluss dieses Monats) mit hoher Wahrscheinlichkeit auch einen geringen Abfluss in dem gleichen Monat der folgenden Jahre induziert. Die Struktur der mittleren Abflüsse kann genauso interpretiert werden, wobei Unterschiede zwischen den Pegeln nur in der Größe der geschätzten Parameter (und damit in der unterschiedlichen Ausgeprägtheit der Persistenz) liegen.

Ein anderes Bild liefert die Analyse der maximalen Abflüsse beider Pegel, hier taucht am Pegel Kaub der größte Wert bei der Halbjahresfrequenz $\pi/3$ auf. Das deutet wiederum daraufhin, dass hier der maximale Abfluss einer Halbjahresschwingung zugrunde liegt. Für den Pegel Köln liegen die Schätzungen für die Jahres– und Halbjahresfrequenz in etwa derselben Größenordnung. Insbesondere zeigt die Existenz der Halbjahresschwingung, dass die Maxima beider Pegel im Jahresverlauf zunächst von Winterniederschlägen und dann vom

Hochgebirgseinfluss beeinflusst werden. Allerdings ist aufgrund der gleichzeitigen starken Dominanz der Jahresschwingung am Pegel Köln der Hochgebirgseinfluss nicht mehr stark ausgeprägt, mit anderen Worten: während der Pegel Kaub noch von beiden Regimen beeinflusst wird, ist der Hochgebirgseinfluss am Pegel Köln nur wenig dominant. Dieses Resultat wird überdies durch die Betrachtung der durchschnittlichen Maxima im Jahresverlauf gestützt (vgl. Gerhard (1976)).

Die getroffenen Schlussfolgerungen hängen natürlich von der Genauigkeit der Schätzungen ab. Es muss daher an dieser Stelle überprüft werden, ob das gewählte Modell mit den geschätzten Parametern überhaupt eine gute Anpassung an die Daten liefert. Dazu berechnen wir den Residualprozess $\hat{u}_t$, indem wir den Filter (4) mittels der geschätzten Koeffizienten berechnen und auf den Prozess $\{\ln X_t\}$ anwenden. Bei einer akzeptablen Anpassung sollte der Residualprozess $\hat{u}_t$ weder ein Gedächtnis noch eine saisonale Struktur aufweisen. Für diese Prozesse betrachten wir erneut die empirische Autokorrelationsfunktion und das Periodogramm in den Abbildungen 5 und 6.

Betrachtet man die verschiedenen Residualreihen des Pegels Kaub, so erkennt man, dass die Saisonstruktur verglichen mit dem Ausgangsprozess (Abbildung 2) hinreichend gut eliminiert wurde, da eine sinusoidale und auf einen Jahresgang hindeutende Struktur nahezu nicht mehr erkennbar ist. Auch das Gedächtnis konnte durch den zugrundeliegenden Prozess adequat beschrieben werden, da die Peaks im Periodogramm der Residuen zwar nicht eliminiert, wohl aber in ihrer Größe reduziert wurden. Dieselbe Schlussfolgerung lässt sich auch für die Residuen der minimalen, mittleren und maximalen Abflüsse am Pegel Köln ziehen. Für die untersuchten Reihen liefert das Modell (6) augenscheinlich eine „gute" Anpassung an die Daten. Bei einer Eliminierung von Saison und Gedächtnis müsste der Residualprozess demnach durch einen ARMA(p,q)–Prozess niedriger Ordnung darstellbar sein. Wir wol-

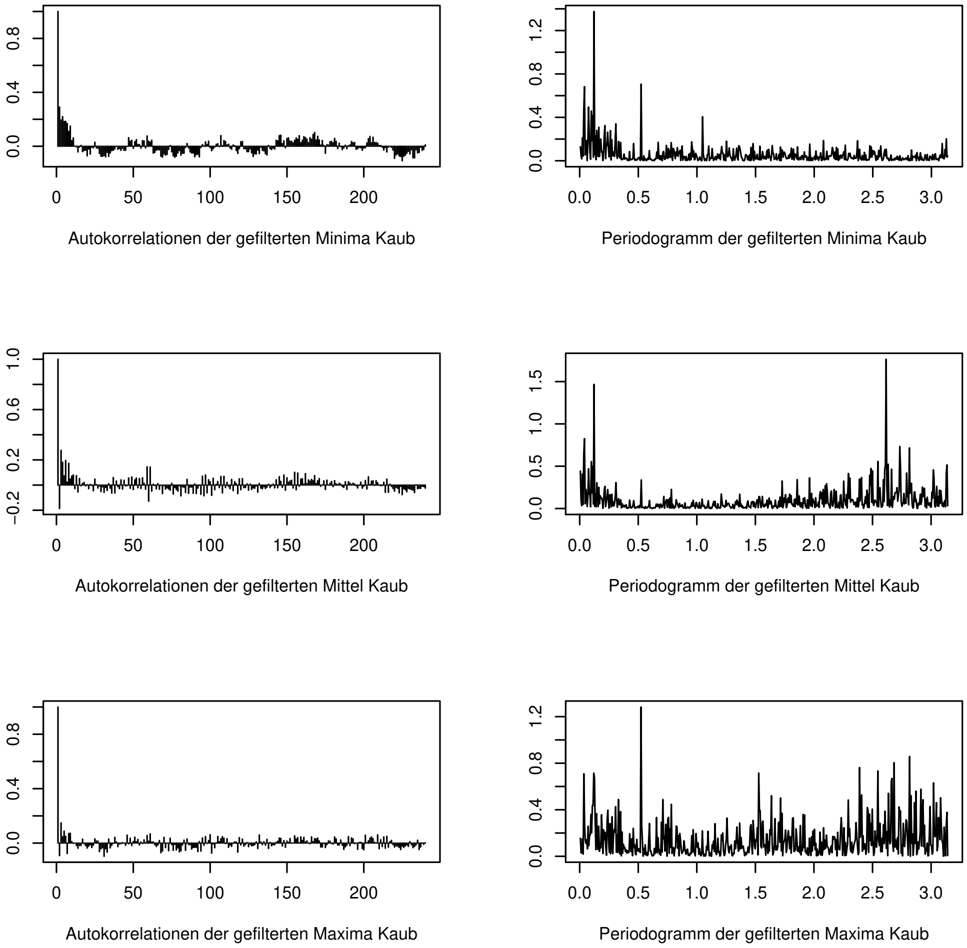

Abbildung 5: Autokorrelation und Periodogramm der geschätzten Residuen der minimalen, mittleren und maximalen Abflüsse am Pegel Kaub

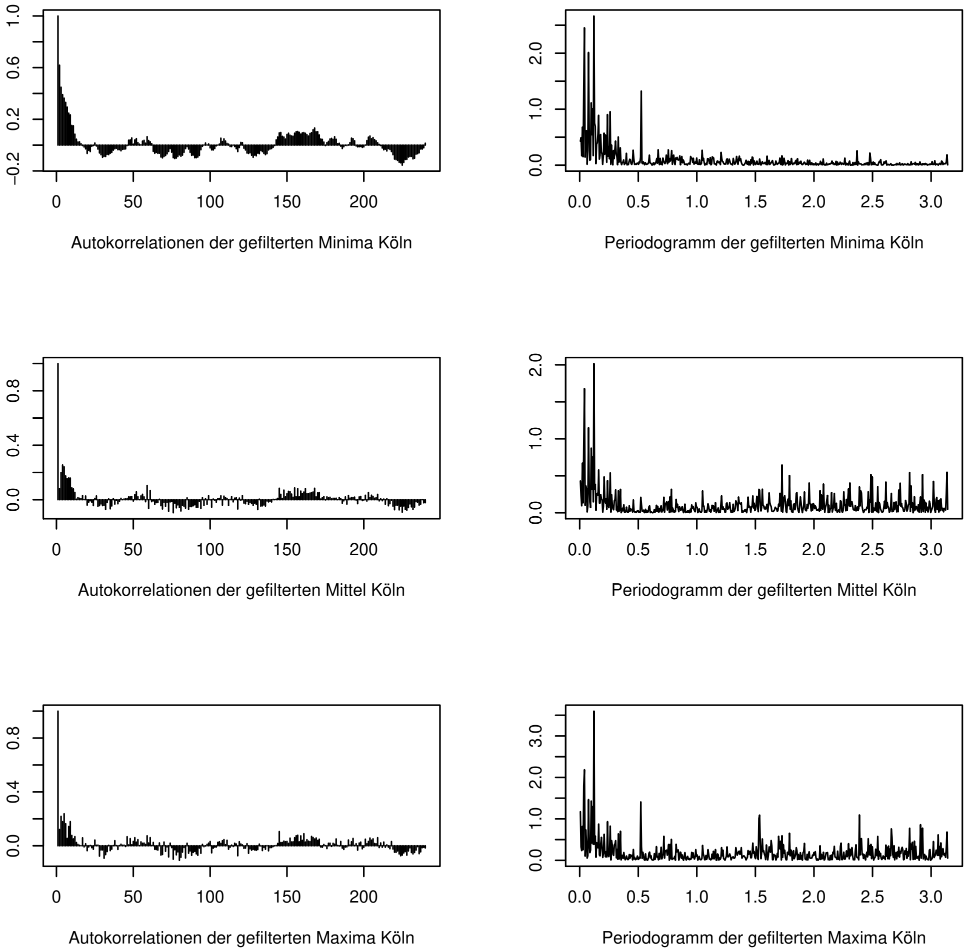

Abbildung 6: Autokorrelation und Periodogramm der geschätzten Residuen der minimalen, mittleren und maximalen Abflüsse am Pegel Köln

len daher mittels eines Informationskriteriums prüfen, ob eine Anpassung von ARMA–Modellen niedriger Ordnung möglich ist. Dazu wählen wir das Bayes–Schwarz–Kriterium (BIC), das neben der Zahl der gewählten Parameter auch die Reihenlänge berücksichtigt (siehe z.B. Schlittgen/Streitberg (1995)). Die geeignetste Anpassung liefert diejenige Modellordnung, für die das Kriterium minimal wird. Tabelle 2 zeigt die resultierenden Ordnungen für die verschiedenen Reihen.

Pegel	Ordnung	BIC
Kaub		
Minima	ARMA(2,1)	-123,21
Mittel	ARMA(3,2)	413,91
Maximum	ARMA(2,2)	957,04
Köln		
Minimum	ARMA(1,2)	5,95
Mittel	ARMA(1,2)	599,82
Maximum	ARMA(2,1)	1214,95

Tabelle 2: Modellordnung der Residualreihen bei ARMA–Anpassung mittels BIC

Für jede Residualreihe werden demnach AR– und MA–Ordnungen kleiner gleich 3 ausgewählt, was auf eine gute Anpassung des gewählten Modells schliessen lässt. Vergleicht man hingegen die BIC–Werte untereinander, so ergeben sich grosse Unterschiede. Kleinen Werten bei den Residuen der Minima stehen grosse BIC–Werte bei den Residuen der Mittel und vor allem Maxima gegenüber. Die minimalen Abflüsse lassen sich somit gut durch das Modell erklären, die mittleren und maximalen Abflüsse erfordern gegebenfalls noch eine Erweiterung durch Modellierung von kurzfristigen Abhängigkeiten. Dies muss bei der Interpretation der Ergebnisse berücksichtigt werden.

4 Zusammenfassung

Aus hydrologischer Sicht ist es plausibel, dass die Abflüsse des Rheins einem Modell mit langfristigen Korrelationen folgen. Dies liegt insbesondere in der Seeretention des Bodensees und der Schweizer Seen wie zum Beispiel des Bieler Sees und des Zürichsees begründet. Im Vergleich zu Pegeln in Bodenseenähe, die ausschließlich unter dem Hochgebirgseinfluss stehen, sollten sich rheinaufwärts zum Pegel Köln Unterschiede zeigen. Der Pegel Köln wird stark von Winterregenfällen beeinflusst und wird damit von zwei verschiedenen saisonalen Strukturen dominiert. Ebenso spielt aus hydrologischer Sicht auch der Einfluss von Zuflüssen, insbesondere der Mosel, eine große Rolle .

Die Arbeit von Mandelbrot/Wallis (1969) unterstützt die These von langem Gedächtnis in den Rheinabflüssen nicht. Die vorliegende Arbeit hat die minimalen, mittleren und maximalen Abflüsse an den Pegeln Kaub und Köln erneut untersucht. Im Vergleich zur Arbeit von Mandelbrot/Wallis (1969) wurde auf eine Vorab–Eliminierung der Saison verzichtet und stattdessen ein flexibles saisonales Modell angepasst, mit dessen Hilfe langes Gedächtnis modelliert werden kann. Saisonales langes Gedächtnis trat bei den minimalen und mittleren Abflüssen an der jährlichen Saisonfrequenz auf. Der maximale Abfluss wird am Pegel Kaub durch einen Halbjahreszyklus bestimmt, am Pegel Köln gibt es aufgrund der Überlagerung der Regime sowohl persistente Jahres– als auch Halbjahresstrukturen. Weitere wesentliche Unterschiede zwischen den Pegeln Kaub und Köln konnten dabei in Bezug auf ihr langfristiges Verhalten nicht festgestellt werden. Die Modellanpassung an die Daten erwies sich als akzeptabel. Allerdings wurden auch deutliche Unterschiede in der Güte der Anpassung festgestellt: so scheint die Modellierung der minimalen Abflüsse vielversprechender zu sein als die der maximalen Abflüsse. Es bleibt zu prüfen, inwieweit alternative Verfahren (z.B. eine simultane Modellierung von Kurzfristabhängigkeiten) eine noch bessere Anpassung leisten können.

Nachwort

Zunächst möchten wir Hans Jürgen Liebscher von der Bundesanstalt für Gewässerkunde danken, der die bereinigten Daten zur Verfügung stellte und mit hydrologischer Beratung zur Seite stand. Die Autoren danken ferner Christoph Helwig und Tamara Könning für ihre Unterstützung bei den Simulationen. Die Unterstützung der Deutschen Forschungsgemeinschaft im Rahmen des SFB 475 „Komplexitätsreduktion in multivariaten Datenstrukturen" sei dankend erwähnt.

Literatur

Andel, J. (1986): Long Memory Time Series Models. *Kybernetika* 22, 105–123.

Arteche, J., Robinson, P.M. (2000): Semiparametric Inference in Seasonal and Cyclical Long Memory Processes. *Journal of Time Series Analysis* 21, 1–25.

Beran, J. (1994): *Statistics for Long-memory Processes.* Chapman & Hall, New York.

Box, G.E.P., Jenkins, G.M. (1976): *Time Series Analysis. Forecasting and Control.* Holden-Day, San Francisco.

Gerhard, H. (1976): Abflussregime. *in: Das Rheingebiet.* Herausgeber: Internationale Kommission für die Hydrologie des Rheingebietes (KHR), 107–160.

Geweke, J., Porter–Hudak, S. (1983): The Estimation and Application of Long Memory Time Series Models. *Journal of Time Series Analysis* 4, 221–238.

Granger, C.W.J., Joyeux, R. (1980): An Introduction to Long-range Time Series Models and Fractional Differencing. *Journal of Time Series Analysis* 1, 15–30.

Hassler, U. (1994): (Mis)specification of Long Memory in Seasonal Time Series. *Journal of Time Series Analysis* 15, 19–30.

Hipel, K.W., McLeod, A.I. (1994): *Time Series Modelling of Water Resources and Environmental Systems.* Elsevier, Amsterdam.

Hosking, J.R.M. (1981): Fractional Differencing. *Biometrika* 68, 165–176.

Hosking, J.R.M. (1984): Modeling Persistence in Hydrological Time Series Using Fractional Differencing. *Water Resources Research* 20, 1898–1908.

Hurst, H. E. (1951): Long-term Storage of Capacity of Reservoirs. *Transactions of the American Society of Civil Engineers* 116, 770–799.

Hurvich, C.M., Deo, R., Brodsky, J. (1998): The Mean Squared Error of Geweke and Porter–Hudak's Estimator of the Memory Parameter of a Long–Memory Time Series. *Journal of Time Series Analysis* 19, 19–46.

Lawrance, A.J., Kottegoda, N.T. (1977): Stochastic Modelling of Riverflow Time Series. *Journal of the Royal Statistical Society* A 140, 1–31.

Mandelbrot, B. B., van Ness, J. W. (1968): Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Review* 10, 422–437.

Mandelbrot, B. B., Wallis, J. R. (1969): Some Long–run Properties of Geophysical Records. *Water resources research* 5, 321–340.

Montanari, A., Rosso, R., Taqqu, M. (1997): Fractionally Differenced ARIMA Models Applied to Hydrologic Time Series: Identification, Estimation and Simulation. *Water Resources Research* 33, 1035–1044.

Montanari, A., Rosso, R., Taqqu, M. (2000): A Seasonal Fractional ARIMA Model Applied to the Nile River Monthly Flow at Aswan. *Water Resources Research* 36, 1249–1259.

Porter–Hudak, S. (1990): An Application of the Seasonal Fractionally Differenced Model to the Monetary Aggregates. *Journal of the American Statistical Association* 85, 338–344.

Ray, B.K. (1993): Long–Range Forecasting of IBM Product Revenues Using a Seasonal Fractionally Differenced ARMA model. *International Journal of Forecasting* 9, 255–269.

Schlittgen, R., Streitberg, B. (1995): *Zeitreihenanalyse*. 6. Auflage, Oldenbourg, München.

Sibbertsen, P. (1999): *Robuste Parameterschätzung im linearen Regressionsmodell*, VWF, Berlin.